Criminal Justice Administrative Records System (CJARS) 2026q2 Vintage

2026-09-29

Abstract

The Criminal Justice Administrative Records System (CJARS) is a nationally integrated data repository designed to transform research and policymaking on the United States criminal justice system. At the University of Michigan, CJARS collects longitudinal electronic records from criminal justice agencies and harmonizes these records to track a criminal episode across all stages of the system. The data are accessible for research and statistical reporting through the Virtual Data Enclave (VDE) at the University of Michigan and through the Census Bureau's Federal Statistical Research Data Center (FSRDC) network. The purpose of this document is to describe how the CJARS database is constructed, database structure, available variables, and other information necessary to use the database for research and statistical reporting.

The project website is https://cjars.org. The latest version of the data documentation can always be found there. Data users who have questions about CJARS can contact cjars-data-users@umich.edu.

Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. *All statistics in this document are based on CJARS data at the University of Michigan except for those explicitly labeled as sourced from Census Bureau data protected by 13 USC §9a.* This document meets all of the U.S. Census Bureau Disclosure Review Board (DRB) standards and has been assigned DRB approval number CBDRB-FY19-371 (approved 2019-06-03).

Table 1: Project change log

Release	Data change log	Documentation change log
2026Q2	– New data: incarceration (MO, WY), parole (MO, WY), probation (MO, WY), jail (counties from CA, OR), arrest (counties from CA, OR) – Limits on values were removed from the following variables: `adj_sent_inc`, `adj_sent_pro`, `adj_sent_rest`, `adj_sent_fine`, `adj_sent_inc_min`, `adj_sent_inc_max`	– Nothing to report
2025Q4	– New data: courts (AK, IA, RI, VT), historical incarceration (NJ, OR) – Added a new jail table populated with counties from the following states: AL, AZ, AR, CA, CO, FL, GA, IL, IN, KS, KY, LA, ME, MN, MS, MO, NJ, MN, NC, OH, OK, OR, PA, SC, TN, TX, VA, WA, WV, WI – Revised arrest table to contain only arrests; bookings have been moved to jail table – Renamed `arr_st_ori_fips` to `arr_st_juris_fips`, `arr_cnty_ori_fips` to `arr_cnty_juris_fips`, `adj_st_ori_fips` to `adj_st_juris_fips` and `adj_cnty_ori_fips` to `adj_cnty_juris_fips` – Broke out the `_rec_src_le` variable into `_rec_src_arr` for arrests and `_rec_src_bkg` for bookings – Reduced over-deduplication of charges and added out-of-coverage observations that were being dropped in the adjudication table – Added a new experimental offense flag `_wpn_off` which indicates weapon involvement in an offense – Revised offense classification algorithm (version 2.0.0)	– Added variable codebook for jail table – Removed deprecated arrest variables from codebook – Renamed `arr_st_ori_fips` to `arr_st_juris_fips`, `arr_cnty_ori_fips` to `arr_cnty_juris_fips`, `adj_st_ori_fips` to `adj_st_juris_fips` and `adj_cnty_ori_fips` to `adj_cnty_juris_fips` in codebook – Removed deprecated `_rec_src_le` variable, added `_rec_src_arr` for arrests and `_rec_src_bkg` for bookings – Added appendix detailing coverage determination process and coverage usage guidance
2025Q1	– New data: incarceration (OH)	– Nothing to report

Table 1: Project change log

Release	Data change log	Documentation change log
2024Q4	– New data: courts (IN), incarceration (ND), probation (ND), parole (ND) – Revised coverage for NY parole data – Updates were made to improve processing of the following: courts (AZ, MI, MN), incarceration (IL), probation (AR, TX), parole (IL) – A bug in the scripts that create the coverage table was fixed which impacts some counties with FIPS codes less than 100	– Added appendix that lists missing variables by state and data table – Added appendix that includes disclaimers for agencies where our data agreement with the agency requires us to provide a disclaimer about their data – Removed obsolete appendix that described structure of state data management systems
2023Q3	– New data: courts (AR), incarceration (CA, CT, ID, NM, RI), probation (ID, NM), parole (CA, CT, ID, NM, RI) – Revised entity resolution algorithm	– Revised data coverage table
2022Q4	– New data: courts (SC), incarceration (IA, KS, MN, NV, NY, OK), probation (IA, MN), parole (IA, KS, MN, OK)	– Added domestic violence variables (`arr_dv_off`, `adj_dv_off`) to codebook – Added data coverage table to codebook
2022Q2	– New data: courts (OK), incarceration (GA, IL, IN, MT, UT, VT), probation (MT, UT, WA), parole (IL, MT, UT) – Revised offense classification schema: Uniform Crime Classification Standard (UCCS) – Revised offense classification algorithm (version 1.0.2)	– Revised methods of determining data coverage – Revised discussions of data coverage, offense classification schema and algorithm
2021Q1	– New data: courts (FL, VA)	– Expanded discussion of Census Bureau record linkage, ML offense classifier – Additional demographic variables in roster codebook
2020Q4	– First harmonized, research-ready data delivered to the Census Bureau	– First version of data documentation released

Contents

1 Project overview **9**
1.1 Project description and objectives . . . 9
1.2 Project scope . . . 9
1.3 Data collection . . . 9
1.4 Current data coverage . . . 10
1.5 Data access . . . 10
1.6 Data privacy . . . 11
1.7 Project leadership . . . 12
1.8 Project funding . . . 12
1.9 Project institutional review board compliance . . . 12
1.10 Acknowledgments . . . 13
1.11 Data citation . . . 14
1.12 Documentation structure . . . 14

2 Data acquisition and coverage **15**
2.1 Sources of criminal justice data . . . 15
2.2 Mechanisms used to acquire criminal justice administrative data . . . 15
2.3 Data acquisition strategy . . . 15
2.4 Acquired data by state, criminal justice domain, and temporal coverage . . . 16

3 Data integration, harmonization, and linkage at the University of Michigan **21**
3.1 Overview of CJARS data integration, harmonization, and linkage . . . 21
3.2 CJARS schema is an abstraction of the sequence of criminal justice processes . . . 21
3.3 CJARS schema supports linkage across agencies at the person or case levels . . . 21
3.3.1 Person-level linkage using the `cjars_id` . . . 22
3.3.2 Data linkage via episode stage unique identifiers . . . 24
3.4 Probabilistic entity resolution algorithms are used to link records . . . 24
3.4.1 Entity resolution algorithms . . . 24
3.4.2 Episode resolution algorithm . . . 25

4 Table and variable codebook **26**
4.1 Variables by relational table . . . 26
4.2 Variable codebook . . . 29
4.2.1 Demographics . . . 31
4.2.2 Arrest . . . 39
4.2.3 Jail . . . 52
4.2.4 Adjudication . . . 74
4.2.5 Probation . . . 115
4.2.6 Incarceration . . . 133
4.2.7 Parole . . . 153
4.2.8 Coverage . . . 169
4.3 ID variables used to link tables . . . 174

Appendices **175**

A Data sources **176**

B CJARS Coverage table details **194**
B.1 What is CJARS Coverage? . . . 194
B.1.1 Coverage Table Structure . . . 194
B.2 Coverage determination examples . . . 195
B.2.1 Adjudication Example . . . 195

B.2.2 Arrest Example . . . 195

C Variable harmonization **196**
C.1 Process . . . 196

D Offense classification **200**

E Code schemes including offense classifications **202**
E.1 Geographic and demographic codes . . . 202
E.1.1 State FIPS and abbreviations . . . 202
E.1.2 County FIPS . . . 203
E.1.3 Sex codes . . . 204
E.1.4 Race and ethnicity codes . . . 204
E.1.5 Demographic imputation codes . . . 204
E.2 Criminal justice event codes . . . 205
E.2.1 Offense classification . . . 205
E.2.2 Offense charge grade . . . 210
E.2.3 Offense legal code . . . 211
E.2.4 Court disposition . . . 211
E.2.5 Probation conditions . . . 211
E.2.6 Probation exit . . . 212
E.2.7 Incarceration entry . . . 212
E.2.8 Incarceration facility type . . . 213
E.2.9 Incarceration exit . . . 213
E.2.10 Parole exit . . . 214
E.2.11 Coverage type . . . 214
E.2.12 Coverage source . . . 214

F Notes on record linkage **216**
F.1 Entity resolution to identify individuals . . . 216
F.2 Episode resolution to link sequences of events . . . 217

G Data notes by jurisdiction **218**
G.1 National data notes . . . 218
G.2 Alabama data notes . . . 219
G.3 Alaska data notes . . . 219
G.4 Arizona data notes . . . 220
G.5 Arkansas data notes . . . 221
G.6 California data notes . . . 221
G.7 Connecticut data notes . . . 222
G.8 Delaware data notes . . . 222
G.9 Florida data notes . . . 222
G.10 Georgia data notes . . . 224
G.11 Idaho data notes . . . 225
G.12 Illinois data notes . . . 225
G.13 Indiana data notes . . . 226
G.14 Kentucky data notes . . . 226
G.15 Maryland data notes . . . 227
G.16 Michigan data notes . . . 228
G.17 Minnesota data notes . . . 228
G.18 Montana data notes . . . 229
G.19 Nebraska data notes . . . 229
G.20 Nevada data notes . . . 230
G.21 New Jersey data notes . . . 230
G.22 New York data notes . . . 231

G.23 North Carolina data notes . . . 231
G.24 Oklahoma data notes . . . 232
G.25 Oregon data notes . . . 233
G.26 Pennsylvania data notes . . . 233
G.27 Rhode Island data notes . . . 234
G.28 South Carolina data notes . . . 234
G.29 Texas data notes . . . 235
G.30 Vermont data notes . . . 235
G.31 Virginia data notes . . . 236
G.32 Washington data notes . . . 236
G.33 Wisconsin data notes . . . 237
G.34 Wyoming data notes . . . 238

H Data disclaimers from agency data providers **239**
H.1 Oregon Judicial Department . . . 239
H.2 Indiana Office of Judicial Administration . . . 239
H.3 Minnesota State Court Administrator's Office . . . 239

I Missing data **240**
I.1 List of missing variables by CJARS table, state, and county . . . 240
I.1.1 Missing Arrest variables . . . 240
I.1.2 Missing Jail variables . . . 240
I.1.3 Missing Adjudication variables . . . 244
I.1.4 Missing Probation variables . . . 245
I.1.5 Missing Incarceration variables . . . 245
I.1.6 Missing Parole variables . . . 246

Bibliography **248**

Index **251**

List of Tables

1 Project change log . . . 2
1 Project change log . . . 3
2 CJARS records acquired by acquisition method (as of 2026-06-24) . . . 15
3 Agencies whose data have been integrated into CJARS . . . 176
4 Event Dates . . . 194
5 Primary Record Sources . . . 195
6 Variable harmonization . . . 196
7 State FIPS codes and abbreviations . . . 202
8 Sex codes . . . 204
9 Race and ethnicity codes . . . 204
10 Demographic imputation codes . . . 204
11 Offense classification scheme . . . 205
12 Charge grade classification scheme . . . 211
13 Charge legal code classification scheme . . . 211
14 Court disposition classification scheme . . . 211
15 Probation conditions classification scheme . . . 212
16 Probation exit classification scheme . . . 212
17 Incarceration entry classification scheme . . . 212
18 Incarceration facility type classification scheme . . . 213
19 Incarceration exit classification scheme . . . 213
20 Parole exit classification scheme . . . 214
21 Coverage type classification scheme . . . 214
22 Coverage source classification scheme . . . 215

List of Figures

1 Map of CJARS geographic and criminal justice domain coverage . . . 10
2 CJARS events . . . 11
3 Estimated CJARS statewide coverage (as of 2026-06-25) . . . 17
4 CJARS data exchange and processing . . . 22
5 CJARS data schema . . . 23
6 Person-level linkage with the `cjars_id` . . . 23
7 Data linkage via unique identifiers of episodes . . . 24
8 CJARS variable harmonization process . . . 199
9 Offense hierarchical classification . . . 200
10 Offense charge flag classification . . . 201

1 Project overview

1.1 Project description and objectives

The Criminal Justice Administrative Records System (CJARS) is a nationally integrated data repository designed to transform research and policymaking on the United States criminal justice system. At the University of Michigan, CJARS collects longitudinal electronic records from criminal justice agencies and harmonizes these records to track a criminal episode across all stages of the system.

CJARS has the following primary objectives:

- Create a dataset that follows criminal cases across criminal justice agencies, from arrest through discharge from the justice system.
- Capture the progression of criminal activities over the life course by constructing longitudinal records of contact that individuals have with the justice system.
- Acquire, store, and use criminal justice data securely and ethically so that the identities of individuals in the CJARS data are kept confidential.
- Provide a national research platform in perpetuity enabling users to produce cutting-edge research not previously possible. The Federal Statistical Research Data Center (FSRDC) network gives qualified researchers access to integrate de-identified criminal justice data at the person-level with Census Bureau survey and administrative records. The CJARS Virtual Data Enclave (VDE) is an environment for researchers to access CJARS micro-data with lower access barriers than the FSRDC. Aggregate statistical information is also available publicly through the Justice Outcome Explorer (JOE).
- Enable the federal statistical system to produce timely and innovative statistical products about the criminal justice system and the justice-involved population.

1.2 Project scope

Many types of agencies are involved in the administration of the U.S. criminal justice system. Police agencies record crime reports and suspect arrests. Sheriff's offices record suspect bookings and periods of confinement before and during court proceedings. Courts record judicial proceedings from initial hearings through case dispositions. Prosecutor and defense attorney offices record other aspects of case processing. State departments of corrections record periods of incarceration. And community supervision agencies records periods of probation and parole. In the federal system, the process is similar but with roles played by federal agencies like the U.S. Marshals Service and the U.S. Bureau of Prisons.

Of these justice processes, CJARS captures records of criminal cases such as those of arrests and bookings, criminal court case filings (including expungements when the data is available), and terms of probation, prison, and parole. In general, a prerequisite for inclusion in the database is the existence of adequate identifying information, which makes collecting de-identified and aggregated criminal justice records out of scope. Crime reports are difficult to link to criminal cases and so are generally considered out of scope. Many of the agencies above also conduct activities that are considered out of scope for the project, including evictions, civil court cases, traffic cases that do not rise to the level of a misdemeanor, and juvenile records except those of cases transferred to adult court.

1.3 Data collection

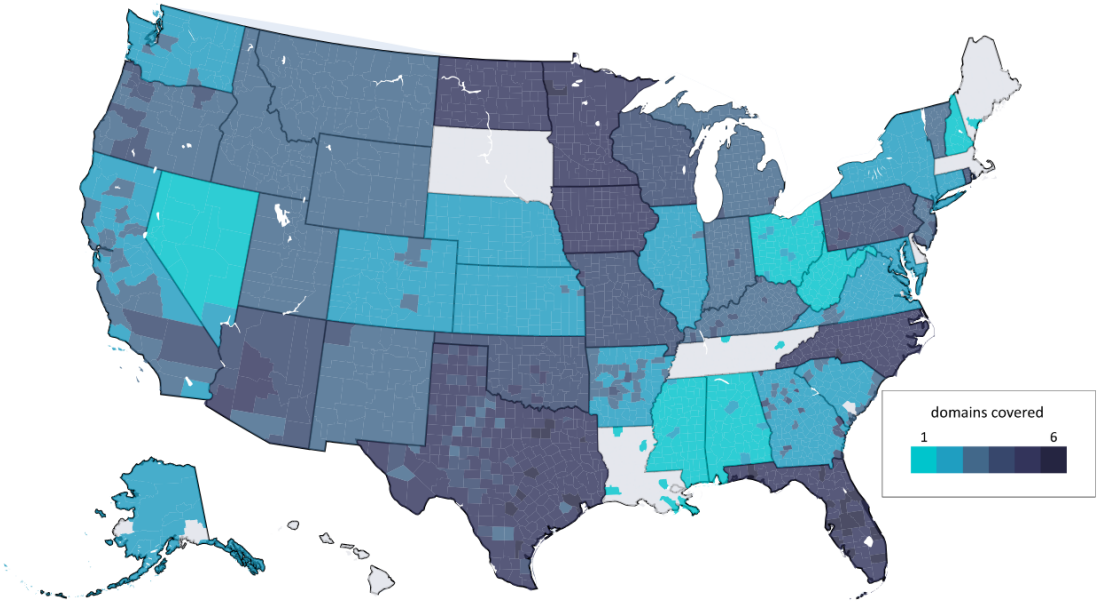
There is substantial variation across federal, state, and local criminal justice agencies in the extent to which criminal justice records are considered public, the degree of decentralization in record processing and storage, and the means by which agencies make records available to the public. The CJARS team at the University of Michigan tries to be entrepreneurial in its approach to data collection, using three different channels: (1) data-use agreements, (2) public records requests, and (3) web scraping or bulk downloads. CJARS prefers data-use agreements where possible since they establish a relationship through which CJARS can share statistical products. For every acquisition, the team maintains high standards of data provenance. CJARS also accepts data with well-documented provenance from individual researchers.

1.4 Current data coverage

The CJARS team has been successful in acquiring data from state and local agencies across the U.S. Our data acquisition strategy has prioritized state agencies and agencies that make their data easily available over the internet, while making steady inroads on data-use agreements and expanding procedural coverage of criminal justice events. Ultimately, our data coverage reflects the resources available to the team.

CJARS has collected data from across the U.S. Figure 1 identifies geographies for which CJARS has collected data and indicates the number of domains of the justice system covered in a given jurisdiction. An interactive version of Figure 1 is available on the CJARS website and can be used to explore more details about data coverage. Based on 2019 state population estimates, states for which CJARS has longitudinal, statewide coverage in at least one criminal justice domain constitute 87.6% of the U.S. population (as of 2026-06-25).

Figure 1: Map of CJARS geographic and criminal justice domain coverage (as of 2026-06-25)



It is important to note that data coverage varies substantially from state to state. Specifically, there are differences in coverage of event type, time frame, and geography. Section 2.4 discusses how state coverage varies over time and by criminal justice domain. Appendix A lists the agencies whose data have been integrated into CJARS.

Many of the data sets acquired by the CJARS team include coverage of criminal justice events that occurred many decades ago. Figure 2 shows the number of records collected by the CJARS project, by year and by criminal justice event type.

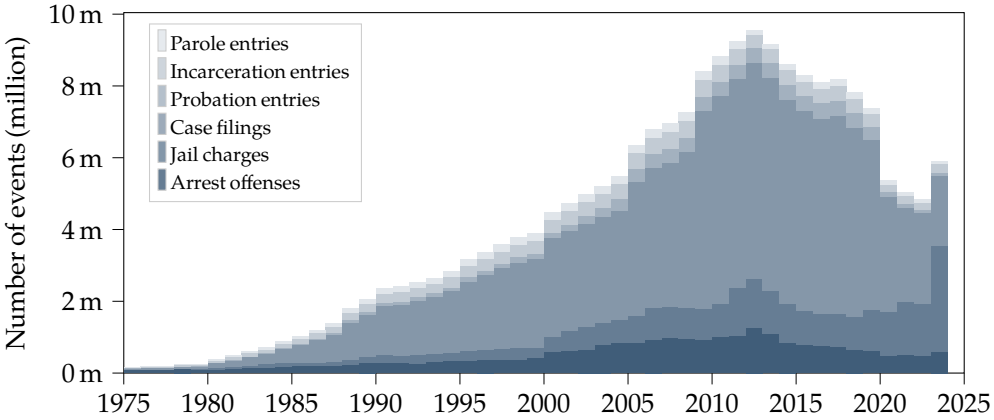
Finally, there are three domains where CJARS currently has no or very limited coverage but which are still considered in-scope: records from prosecutors, records on confinement before and during trial, and records from the federal criminal justice system.

> USER NOTE
>
> Data coverage may differ for the version of the CJARS database within the Virtual Data Enclave (VDE). Please refer to the CJARS VDE Data Documentation Supplement for jurisdictions excluded from the version of the CJARS database available on the VDE.

1.5 Data access

A major contribution of CJARS is the creation of a broadly accessible repository that will permanently enhance research infrastructure in the United States. Data are collected, cleaned, and harmonized at the University

Figure 2: Events in the CJARS database, by year and event type (as of 2026-06-26)



of Michigan. The CJARS database is made available through the VDE and FSRDC network.

The CJARS VDE is a highly secure virtualized server where the CJARS database can be used for research and statistical reporting. The CJARS VDE is hosted at the University of Michigan and managed by CJARS project staff. The environment is designed to minimize barriers to getting data access while still maintaining strong security over the highly sensitive data that can be accessed. The data environment provides a suite of auxiliary data resources and services to support a wide range of research needs. Users are welcome to bring their own data into the research environment. For more information on the VDE, please visit: https://www.cjars.org/vde.

Qualified researchers can use the standard Census Bureau FSRDC proposal process to request use of the restricted-access CJARS data. The FSRDC network is broadly available to researchers and provides excellent scientific computing resources, while maintaining the highest levels of information security and confidentiality protection. Researchers working in the FSRDCs only have access to de-identified data, and can only request release of aggregate statistical material. The FSRDCs allow qualified researchers on approved projects to link CJARS records at the person-level to an abundance of socioeconomic survey and administrative data held by the Census Bureau. For more information on the FSRDC system and other available data, please visit: https://www.census.gov/fsrdc.

Researchers interested in working with the CJARS data within the Federal Statistical Research Data Center network should contact the CJARS team or their closest FSRDC administrator. The CJARS team has prepared a proposal development guide to assist researchers in the FSRDC proposal process.

Additionally, the CJARS team developed the Justice Outcomes Explorer (JOE) which is a web-based interactive data dashboard that hosts a curated set of aggregate statistics that were produced by linking the CJARS database to other survey and administrative records available through the FSRDC network. JOE was created to provide a low-barrier access mechanism to CJARS-produced statistics. Available statistics cover criminal justice caseloads and various outcomes of justice-involved populations, such as recidivism, employment, earnings, take-up of social benefits, and health. Details about the statistical information available through JOE can be found on the site's methodology page.

1.6 Data privacy

The CJARS project is built around data protection and security, in order to ensure that the identities of all individuals in CJARS records are kept confidential.

The University of Michigan processes all of the criminal justice records that it collects on a data system that was built to be compliant with FBI Criminal Justice Information Services (CJIS) standards. The Michigan State Police has reviewed this system and determined that it is "substantially compliant" with all FBI CJIS policies that are applicable to an entity like CJARS.

University of Michigan IT staff are responsible for the technical security of the CJARS VDE. Security protocols

for the CJARS VDE are built to comply with NIST 800-171 standards. University of Michigan IT staff regularly perform security assessments of the CJARS VDE environment.

The CJARS VDE is hosted on servers within an environment isolated behind dedicated hardware firewalls. The VDE has an internal Active Directory domain, and only authorized parties are granted access. The network is monitored by a University of Michigan maintained IPS logging system that feeds into a Snort event monitoring system. Access to the CJARS VDE terminal server is protected with DUO multi-factor authentication and strong passwords enforced by Group Policy Settings. Security groups are used to maintain least privileged role based access controls.

The Census Bureau and its associated FSRDCs comply with all current National Institute of Standards and Technology (NIST) standards and publications in accordance with Title III of the E-Government Act of 2002 (PL107-347). All systems are fully assessed against NIST Special Publication 800-53r3 and Special Publication 800-37r1. The Census Bureau IT Security Program is reviewed annually by the Department of Commerce Office of the Chief Information Officer and Inspector General.

At the Census Bureau, data are protected by 13 USC §9a. The University of Michigan transfers data securely to Census, whereupon received data are accessible only by a small staff responsible for inventorying the contents of the data and removing sensitive identifying information. Next, a data linkage team uses a probabilistic linkage process to assign a unique identifier called a Protected Identification Key (PIK) that can be used to link records to other databases held at the Census Bureau.

Finally, de-identified CJARS records are moved to secure research servers. There, qualified researchers on approved projects can work in the FSRDCs to link CJARS records with other data held by the Census Bureau. Only de-identified information is available for researchers in order to maintain the privacy of individuals.

1.7 Project leadership

Michael Mueller-Smith. CJARS Co-founder and Director. Associate Professor, Department of Economics, University of Michigan; Faculty Associate, Population Studies Center, University of Michigan.

Keith Finlay. CJARS Co-founder. Research Economist, U.S. Census Bureau.

Samuel Marvin. CJARS Associate Director. University of Michigan.

Holly Champagne. CJARS Director of Data Partnerships & Acquisition. University of Michigan.

Chandler Rombes. CJARS Director of Data Production. University of Michigan.

Jordan Papp. CJARS Director of Product Development. University of Michigan.

1.8 Project funding

The CJARS project has been supported by National Science Foundation, as well as grants from Arnold Ventures, the Bill & Melinda Gates Foundation, the Robert Wood Johnson Foundation, and The Annie E. Casey Foundation. The University of Michigan has supported the CJARS project through the following programs: Michigan Institute for Teaching and Research (MITRE), Populations Studies Center (PSC), and Poverty Solutions. Finlay's work at the Census Bureau has been supported by funding from the Evidence-Based Policymaking Commission Act (P.L. 114-140) and the Foundations for Evidence-Based Policymaking Act (P.L. 115-435).

1.9 Project institutional review board compliance

CJARS has received approval from the University of Michigan Institutional Review Board (repository application reference number REP00000094). In addition to initial review and approval, the University of Michigan requires that data repository projects undergo annual reviews to secure continued approval. CJARS completes these reviews and remains in compliance. The Census Bureau does not have an institutional review board, but the project has been vetted by a project review process that verifies compliance with legal, ethical, and scientific standards. Research users should consult with their respective institutional review boards to determine if they must apply for approval to cover specific research projects using the CJARS data.

1.10 Acknowledgments

This project would not be possible without the hard work of a great group of people. Jordan Papp, Holly Champagne, Diana Sutton, Ashley Gross, and Matt Van Eseltine have supported the project as project managers. Jay Choi developed the CJARS web scraping system and a lot of the harmonization processes. Matt Gross developed entity resolution algorithms and harmonized a lot of data. Ben Pyle, James Reeves, Susan Parker, Arkey Barnett, and Paula Lopez contributed to data harmonization and research as graduate student researchers. Josh Kim and Brian Miller helped develop automated quality control systems. Chandler Rombes has contributed to the codebase as a programmer and has lead improvements to entity resolution algorithms. Ethan Leonello has contributed to the development of the codebase and to the web scraping system. Katherine Godwin and Ying Tang have contributed to data processing and production. Francis Fiore, David Smith, Neel Shah, Manav Bhatia, Luke Reuter, and Hasika Sridhar turned the web scraping system into an industrial-strength platform. Shaeq Ahmed, Luis Baldomero-Quintana, Madeleine Danes, Kenna Garrison, Lauren Lee, Lyllian Simerly, Ellen Stuart, Peixin Yang, Tyler Shea, Konstantine Wade, Daniel Luce, and Merin McCallum have also supported the project as research assistants. Brittany Street, Elizabeth Luh, and Hina Usman have furthered CJARS research as postdoctoral research fellows.

Catherine Seay-Ostrowski, Johnathon Phillips, Debra Bourque, Melissa Wallace, Connie Hallom, Melissa Milligan, Carly Hughes, Sarah Smith, Charles Mattison, Carol Bowen, Kerri Cross, Violet Elder, Lauren Tingwall, Lisa Blumenauer, Matt Toaz, Daniel Gallegos, Kyle McCallum, and Joe Braun have supported the project as research administrators and IT staff. The following advisors have guided the project as part of the CJARS Board of Directors: Shawn Bushway, Julia Lane, Jens Ludwig, Jim Lynch, Justin McCrary, Jeff Morenoff, Amy O'Hara, Anne Piehl, Steve Raphael, Bill Sabol. The project also received guidance from members of its Scientific Advisory Council: Maggie Levenstein, Jeff Morenoff, and JJ Prescott.

At the Census Bureau, Carla Medalia, Katie Genadek, and Kevin Deardorff have offered a lot of support for the project. As Civic Digital Fellows, Daniel Schnelbach, Tony Goss, Jane Boettcher, Jessica Shi, and Carl Xu helped develop the statistical reporting system for CJARS data providers.

We would also like to thank a number of individuals and organizations who have generously donated their research data to CJARS:

- Amanda Agan, Crystal Yang, and Alex Mas (New Jersey Superior Court records),
- Ben Grunwald (North Carolina Judicial Branch records),
- Ben Schoenfeld (Virginia Circuit and General District Court records), and
- The NYU Public Safety Lab Jail Data Initiative

CJARS has also benefited from the assistance many others have provided for our data collection efforts by facilitating our outreach with agencies. While not all of these efforts have translated into successful data acquisitions, we are grateful for the community support that we have received from:

- Milda Aksamitauskas (Wisconsin criminal justice records),
- Beth Allman (Florida court records),
- Tiffany Bergin (New York criminal justice records),
- Kwan-Lamar Blount-Hill (New York City court records),
- Mark Cohen (Tennessee court records),
- Susan Dewey (Alabama court and incarceration records),
- Timothy G. Edgemon (Alabama court records),
- Zachary Hamilton (Nebraska court records),
- Margaret Henn (Maryland court records),
- Amy O'Hara (Utah court records),
- Brittany T. Martin (Kentucky court records),
- Annie Rexford-Bowen (Oregon incarceration and court records),
- Ashley Richards (Maryland court records),
- William Sabol (Georgia incarceration records),
- Christopher Schreck (New York City incarceration records),
- Sarah Shannon (Georgia court records),

- Megan Stevenson (Virginia incarceration records),
- Sara Stewart (Oregon incarceration and court records),
- Andy Yuan (Arizona court records), and
- Jeffrey Zuback (Maryland court records).

1.11 Data citation

Publications and research reports based on the CJARS database should cite it appropriately. The citation should include the following:

> Keith Finlay and Michael Mueller-Smith. 2026. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. https://cjars.org.

1.12 Documentation structure

- Section 2 provides an overview of how the CJARS project acquires data and gives a summary of data coverage at state level and by criminal justice domain.
- Section 3 describes how raw criminal justice data is integrated and harmonized into a single national format at the University of Michigan.
- Section 4 has the detailed table and variable codebook.
- Appendices cover the project in much more detail.
  - Appendix A identifies the agency sources of CJARS administrative data.
  - Appendix B describes how to use the coverage table, as well as the coverage determination process.
  - Appendix C explains the variable harmonization process in detail.
  - Appendix D describes the machine learning used to classify offenses.
  - Appendix E lists code schemes used in the variables, including criminal classification scheme (and explains how those relate to BJS schemes).
  - Appendix F explains the methods used to identify individuals and procedural episodes in the CJARS data.
  - Appendix G has jurisdiction-specific data notes.
  - Appendix H has disclaimer language provided at the request of agencies.
  - Appendix I includes a table that describes missing information broken down by state and data type.

2 Data acquisition and coverage

2.1 Sources of criminal justice data

CJARS acquires records of criminal cases such as those of arrests and bookings, criminal court case filings, and terms of probation, prison, and parole. These data are produced and maintained by many different types of agencies that together constitute the criminal justice system. For example, post-conviction incarceration records are typically held and managed by a state's Department of Corrections. Conversely, arrest records are often held and managed by local police departments. To provide researchers with a clear understanding of the typical sources of CJARS data, the types of events held in this repository are described below.

- Booking and arrests. This information is typically recorded by police departments and sheriff's offices.
- Trial, dispositions, sentences, and appeals. Such information may be held by local and state courts, as well as district attorney's offices and county clerks.
- Incarceration, probation, parole, and special programs. Most of this information may be recorded by a state Department of Corrections (or its equivalent). Nevertheless, it can be also recorded by some counties or regional offices.

The sources of all data held within the CJARS repository can be found in Appendix A. These data have been collected from a mix of state, county, and municipal agencies. Some data have also been acquired from third parties who manage data on behalf of criminal justice agencies or who otherwise have access to those records.

2.2 Mechanisms used to acquire criminal justice administrative data

The CJARS team at the University of Michigan uses several approaches to acquire criminal justice administrative records:

1. **Data use agreements.** The University of Michigan enters into formal legal agreements with criminal justice agencies to acquire data. These agreements delineate the specific responsibilities of the CJARS project and the allowable uses of the acquired data. These agreements must allow secure transmission of the data from the University of Michigan to the Census Bureau in order for linkage-based research to occur.
2. **Public records requests.** A request for public records is filed in jurisdictions where legal statutes allow for such requests.
3. **Web scraping or bulk downloads.** Data that are publicly available online are collected using web scrapers or downloaded in bulk when available. Scraping is only conducted using methods that comply with an agency's terms of use.
4. **Data donations.** Data are provided by individuals not associated with a criminal justice agency, but who have the ability to share administrative criminal justice records with the CJARS project, usually because the records were a collection of publicly available information.

Table 2 shows the number of records that each of these methods has contributed to the CJARS repository (records collected from data donations fall into the category that the original requestor used to collect the data).

Table 2: CJARS records acquired by acquisition method (as of 2026-06-24)

Acquisition source	Records	Individuals
Data use agreements	1,916,537,739	27,024,773
Public records requests	470,770,406	22,572,189
Web scraping or bulk downloads	1,055,144,992	11,589,000

2.3 Data acquisition strategy

The ultimate goal of CJARS is to provide data infrastructure with nationwide coverage of the major events that occur at each stage of the justice system. CJARS is focused on expanding the geographic and procedural

coverage of its data holdings. Expanding geographic coverage means expanding the number of states covered at the statewide level. Expanding procedural coverage means filling in data sources from the sequence of criminal justice processes, so CJARS can track cases and events from charge to final sanction. Expanding data coverage along these two dimensions will facilitate broader cross-jurisdictional research, comparative analysis, and research that requires tracking a criminal episode though the different stages of the justice system.

2.4 Acquired data by state, criminal justice domain, and temporal coverage

CJARS collects data on a continual basis, so its data holdings will change over time. Therefore, the current data holdings are constrained by both geography and time. In many cases, it is possible to acquire data regarding incarceration, probation and parole at a state level, but information regarding other parts of the criminal justice process (e.g. arrests and bookings or dispositions) might not be available at the moment due to different circumstances (e.g., FOIA requests take longer than expected or jurisdictions do not have electronic records for a specific period).

Figure 3 summarizes the availability of data by state (where statewide coverage is available), time frame, and event type. In this table, markings designate coverage of criminal justice domains in each state as one of: statewide coverage, partial geographic coverage in a state, and "caseload snapshots" of all offenders under supervision at a single point in time. The thickness of the bars is roughly proportional to how much of a state's population is covered. Please see the legend for further detail.

Note that Figure 3 only provides a general overview of CJARS data holdings. A more detailed list of jurisdictions whose data have been integrated into CJARS can be found in Appendix A.

> **USER NOTE**
>
> Data coverage may differ for the version of the CJARS database within the VDE. Please refer to the CJARS VDE Data Documentation Supplement for more information.

Figure 3: Estimated CJARS statewide coverage (as of 2026-06-25)

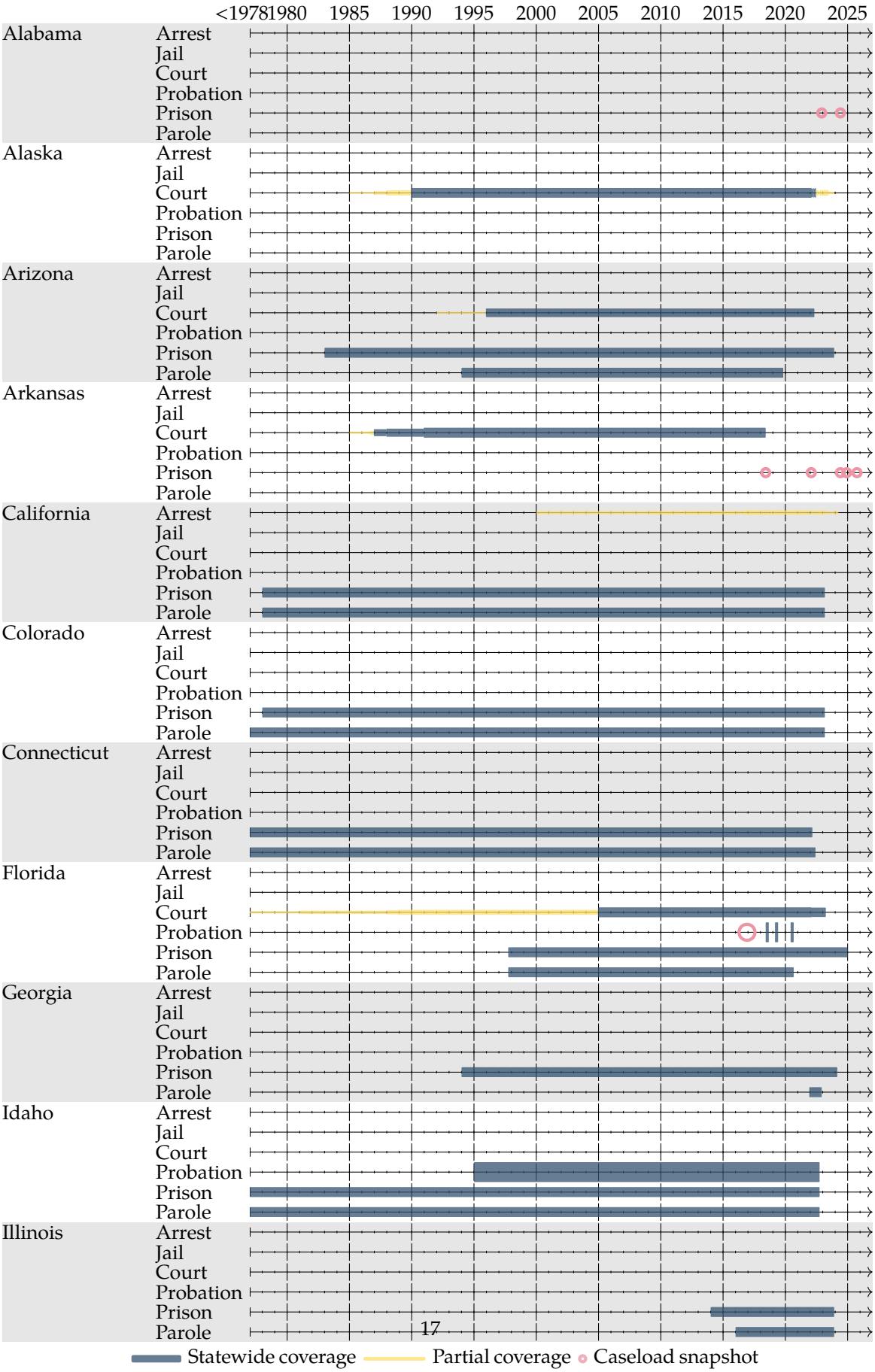


Figure 3: CJARS statewide coverage (cont'd)

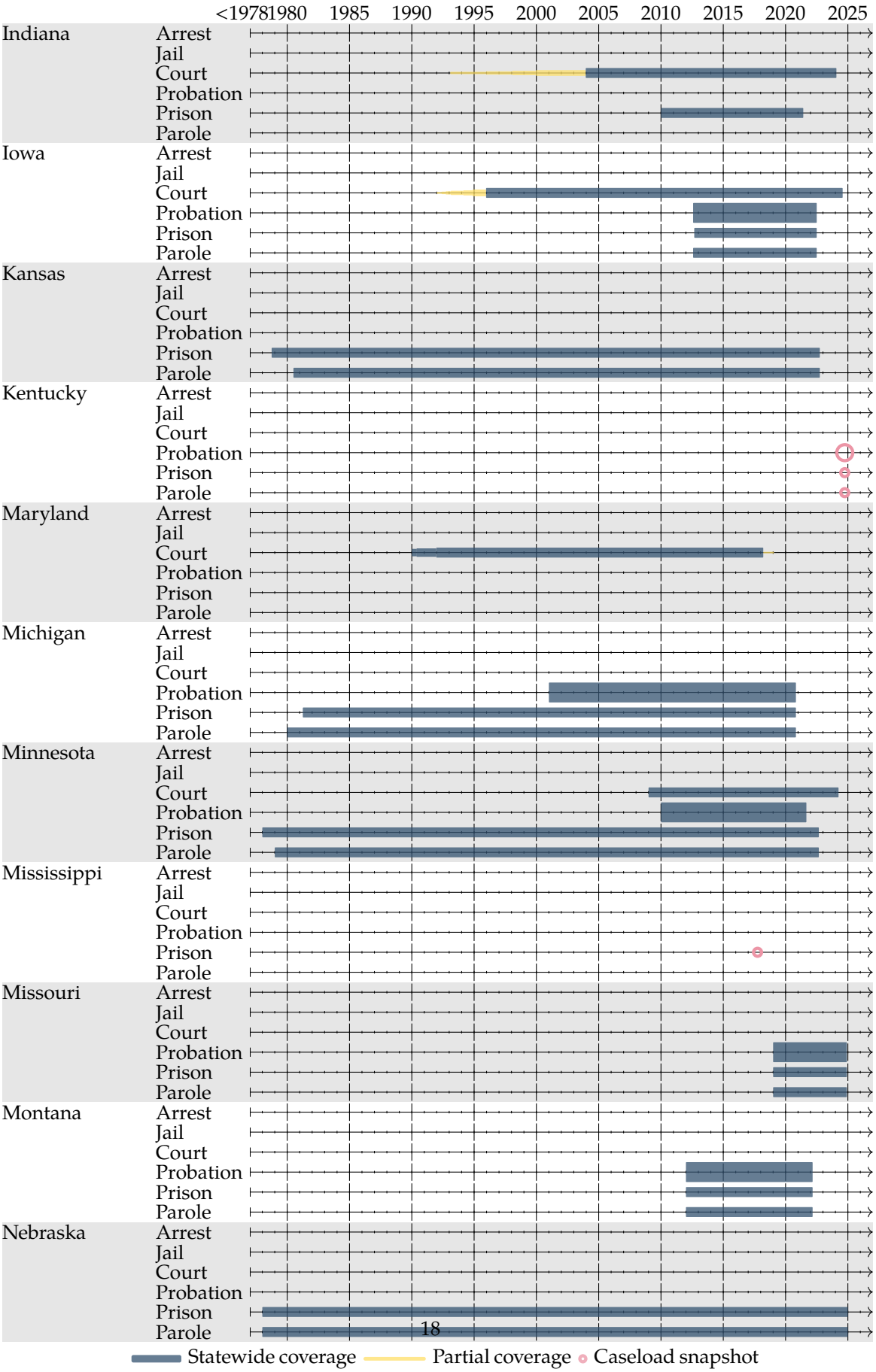
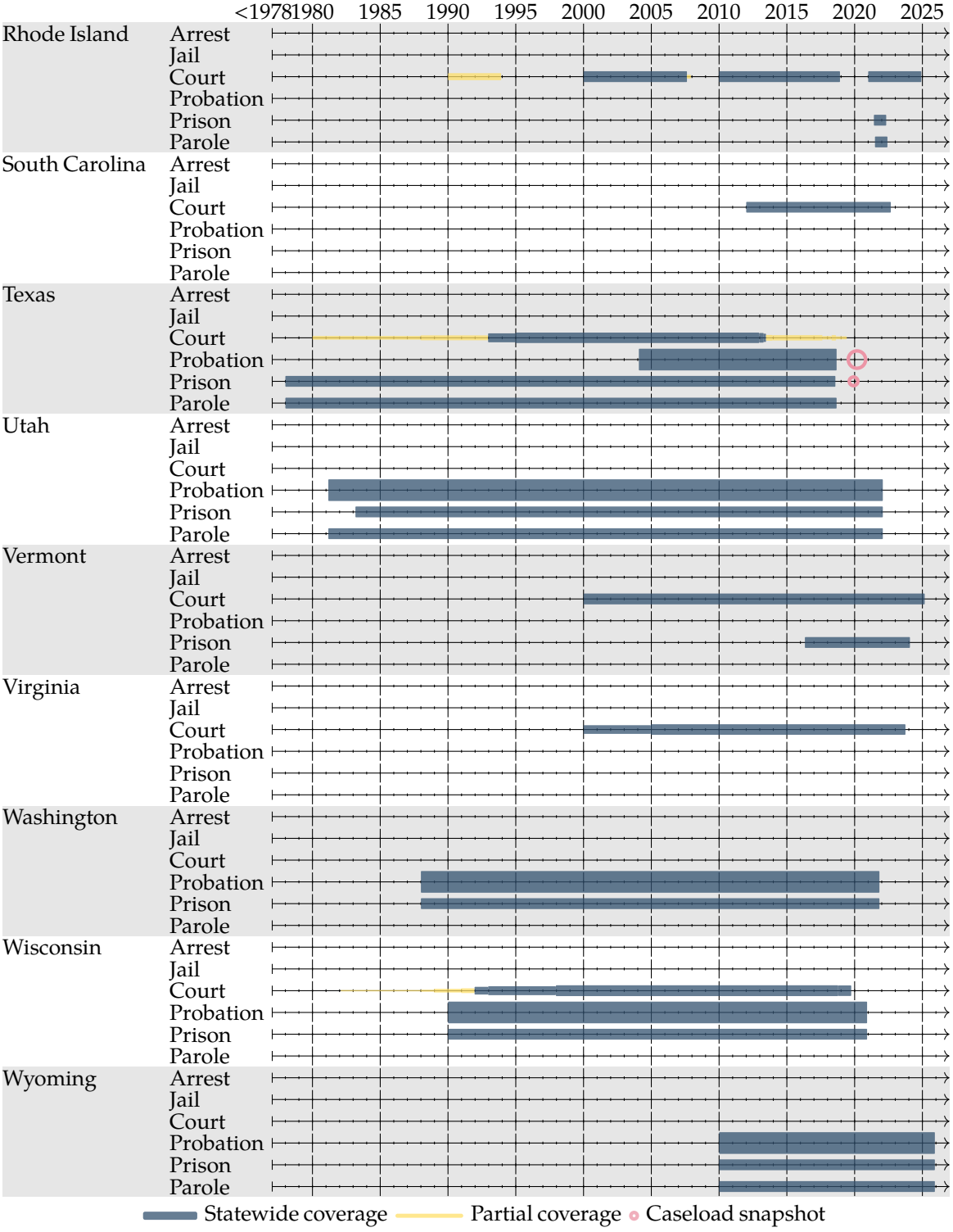


Figure 3: CJARS statewide coverage (cont'd)

Figure 3: CJARS statewide coverage (cont'd)



3 Data integration, harmonization, and linkage at the University of Michigan

3.1 Overview of CJARS data integration, harmonization, and linkage

One of the major barriers to research on the criminal justice system is a lack of integration of records across agencies. For example, arrest records are rarely linkable to information that occurs at later points in the system (e.g., incarceration). The CJARS team at the University of Michigan implements the following steps in order to create a single integrated data platform:

- **Data requests and collection**. CJARS collects data using the three-pronged approach that is described above. All data and other relevant information describing the data are kept in a secure data enclave.
- **Data cleaning**. Original data are cleaned and harmonized to fit a common CJARS schema (described in more detail in later sections).
- **Merge**. The harmonized data are appended and merged with other criminal justice records in CJARS to create a single dataset spanning agencies and jurisdictions.
- **Person-level linkage**. CJARS uses direct and indirect identifiers in a roster file to link records that all belong to the same individual. The information is used to find the criminal justice records that belong to the same person. A `cjars_id` study id is created for each unique individual identified during this process.

Figure 4 provides a visual representation of the data integration process. This figure shows how data that is collected from data providers (first panel) is added to the UM CJARS Repository (second panel). Then once the data is processed, the data are integrated in the Census Bureau (third panel). Finally, once processed by the Census Bureau, the data are made available to external researchers (fourth panel).

3.2 CJARS schema is an abstraction of the sequence of criminal justice processes

The goal of CJARS is to create a national database that tracks each criminal justice episode for an individual from arrest through discharge from the justice system. An episode refers to the complete chain of events as a case is processed through the justice system and can include information from numerous agencies. Due to differences across states and agencies, CJARS developed a national data schema as a standard to which disparate administrative records are linked and harmonized. The events that are included in the national data schema are shown in Figure 5. The arrows in this figure depict the expected progression of events that make up an episode through the system.

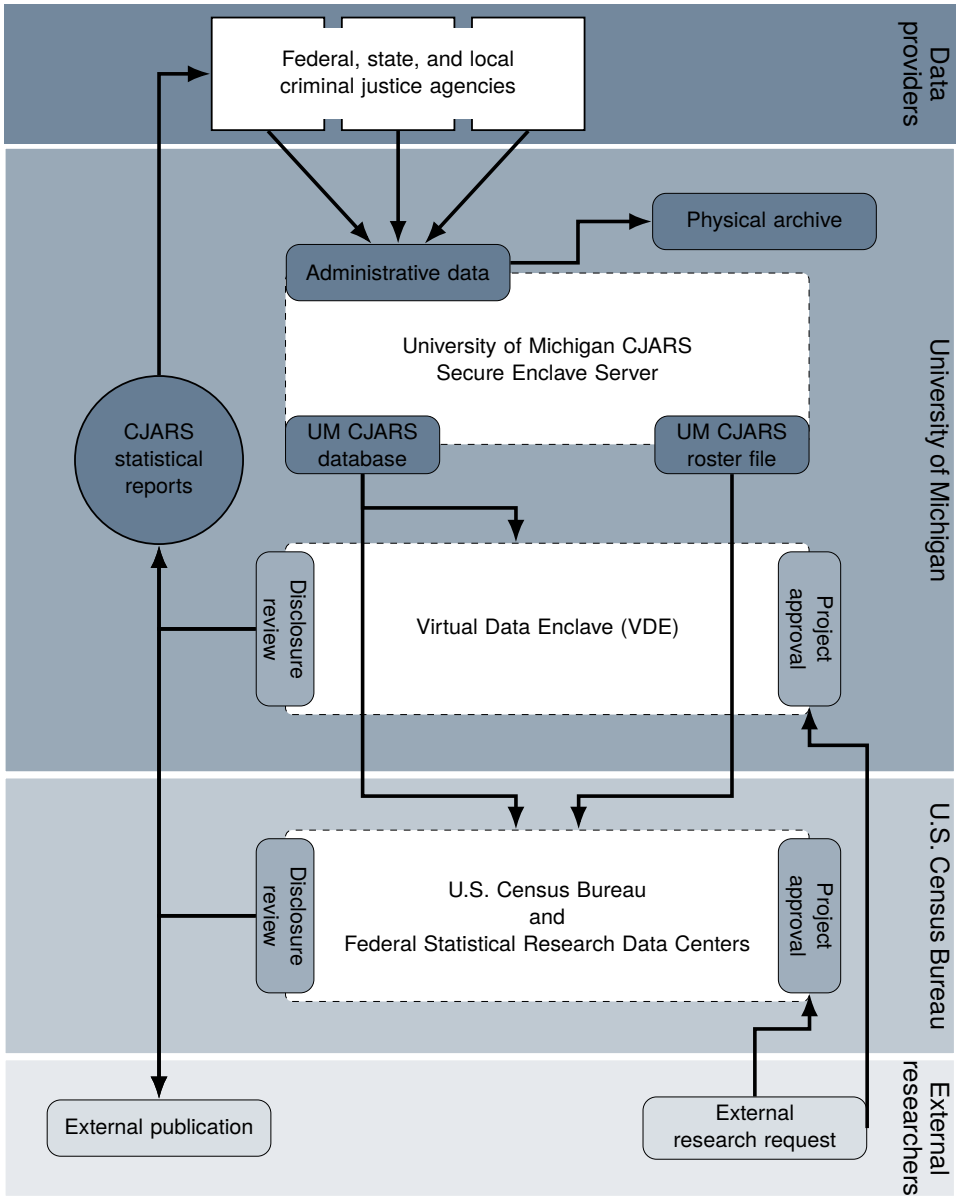
This figure of events was used to develop a data schema which facilitates the linkage of records across disparate systems. The CJARS national schema was designed in a way that strikes a balance between capturing the complexity of the way that events related to a criminal episode are processed through the system while also structuring the data so that it is usable for research purposes. More information about these linkages and design of the national data schema is described below.

3.3 CJARS schema supports linkage across agencies at the person or case levels

A major barrier to research on the criminal justice system is a lack of integration of data across criminal justice agencies or with data on non-criminal justice outcomes. For example, unique individual identifiers are often not present that allow a person's records to be linked across different agency data systems. In addition, there are often no identifiers to indicate which records belong to a single criminal justice episode. CJARS has addressed these issues by leveraging machine learning models that probabilistically match records to individuals and events to episodes.

The following subsections describe more about the methods used to develop these linkages and the linkages themselves.

Figure 4: CJARS data exchange and processing

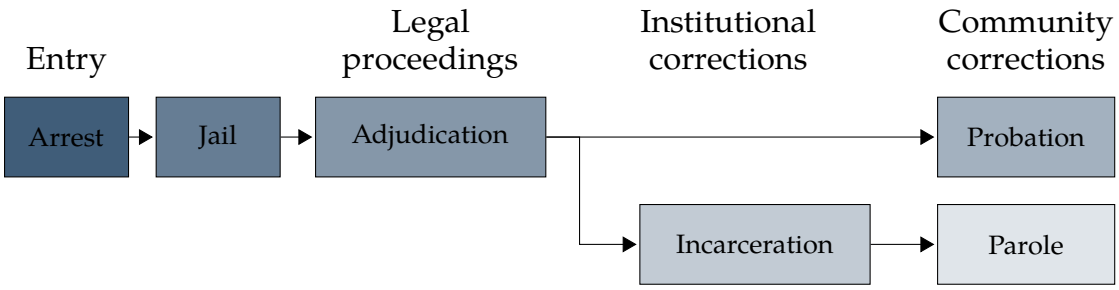


3.3.1 Person-level linkage using the `cjars_id`

CJARS has developed an algorithm that probabilistically matches records to individuals when no unique identifier is available by using other identifying information in the data. Once an individual has been identified, he or she is assigned a unique individual identifier (`cjars_id`). This is done so the data can be used in de-identified form to protect individual identities.

Assignment of a `cjars_id` allows researchers to link all records in the CJARS database that belong to an individual, to that individual. This data linkage facilitates analyzing the data to determine how many events are seen for an individual, how many offense he or she has committed, recidivism, etc.

Figure 5: CJARS data schema

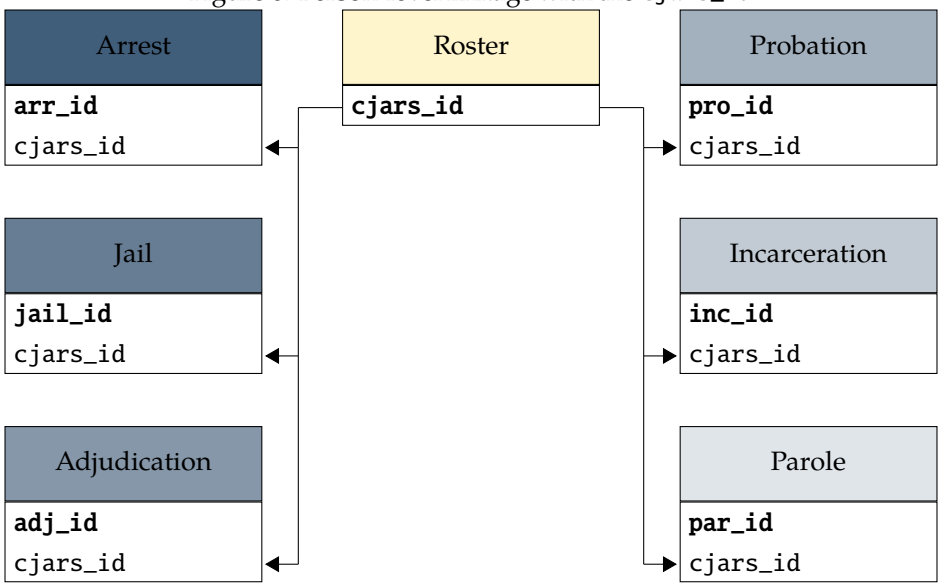


USER NOTE

The CJARS roster is the product of entity resolution algorithms that make use of information from a variety of data sources. Membership in the CJARS roster should *never* be used as a proxy for involvement in the criminal justice system.

The CJARS data is comprised of seven separate databases. The seven databases include a demographics table and one database for each of the six types of events that are covered in Figure 5. Collectively, the six databases containing the criminal justice events are referred to as the CJARS relational databases. The roster uses the `cjars_id` identifier to uniquely identify every individual included in the CJARS repository. Additionally, each CJARS relational database has a unique identifier that identifies each event in the file. Specifically, the `arr_id` uniquely identifies arresting offenses, `jail_id` uniquely identifies offenses associated with jail bookings, `adj_id` is for charges filed in criminal court, `pro_id` is for probation events, `inc_id` is for incarceration events, and `par_id` is for parole events. As can be seen in Figure 6, the `cjars_id` can be used to link all events in each CJARS relational database to the individual that they belong to based on the roster.

Figure 6: Person-level linkage with the `cjars_id`



Note: Arrows indicate linking variables between tables. Variable names listed in **bold** represent unique variables in the dataset. Variables that are not bold may have multiple entries.

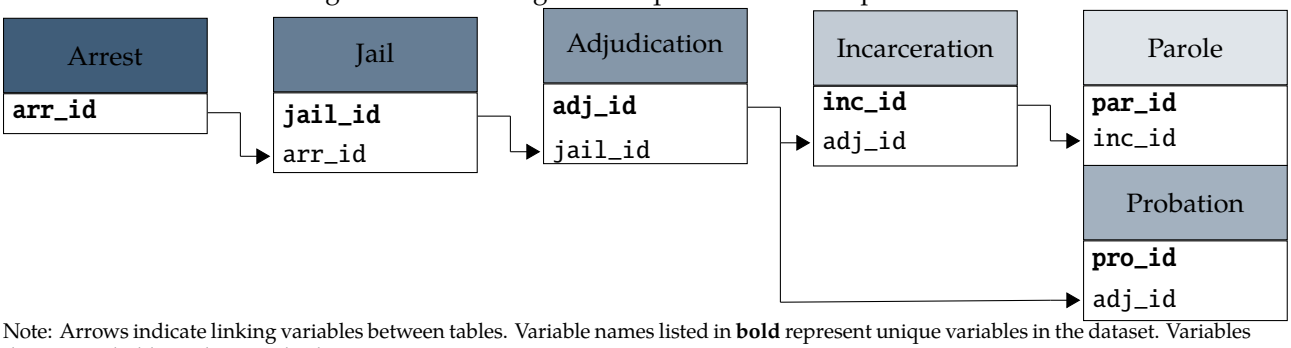
3.3.2 Data linkage via episode stage unique identifiers

> USER NOTE
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense. Therefore, topical relational tables may contain the primary keys that uniquely identify events but will not contain the foreign keys described below that allow event-level links with other tables.

CJARS has also developed a method of probabilistically matching criminal justice events to an episode. This linkage is created so that researchers can trace every event associated with a single criminal justice episode.

Figure 7 summarizes how the data linkage is structured based on unique identifiers of events in each CJARS relational database. Each CJARS relational database contains an identifier for that respective type of event contained in the table, but also an identifier that can be used to link to the event that led to the event described in the table. For example, a jail booking offense in the jail relational database will contain the `jail_id` which uniquely identifies that booking offense, but also the `arr_id` when it is known that an arrest within the arrest relational database preceeded that booking offense. This allows for the reconstruction of a chain of events which can ultimately be linked back to an individual.

Figure 7: Data linkage via unique identifiers of episodes



Note: Arrows indicate linking variables between tables. Variable names listed in **bold** represent unique variables in the dataset. Variables that are not bold may have multiple entries.

It is important to note that reconstruction of a criminal justice episode is constrained by the availability of data. There are limitations in both geographic coverage and coverage across time. Please refer to Appendix G for more information about the constraints of the CJARS database.

3.4 Probabilistic entity resolution algorithms are used to link records

The data linkages that were described above were established using probabilistic linkage algorithms. The development, validation, and nature of these algorithms are described in the following two sections. Further detail can be found in Appendix F.

3.4.1 Entity resolution algorithms

A common issue in linking administrative criminal justice records across disparate sources is the absence of a unique individual identifier. As a result, researchers have to turn to other information that identifies individuals, such as names and dates of birth. In "big data" applications, such as CJARS, this requires an algorithmic approach to make matching feasible.

There are two broad classes of entity resolution algorithms: deterministic and probabilistic. Deterministic algorithms focus on the variables common to both sets of data being matched. In some examples, paired cases must match on all common variables to be classified as a match. In other cases, with richer sets of matching variables, some flexibility can be built into the matching process. Conversely, probabilistic algorithms attempt to predict the probability that any two observations should be linked based on the relative agreement of their

matching variables. This approach has benefits over deterministic models in that it more flexibly sets a decision rule that optimizes the trade-off between making more matches and limiting false matches. The decision of which method to apply is dependent upon the available set of matching variables and the type/quality of information available in those matching variables. CJARS developed a probabilistic algorithm based on the identifiable information that is usually available in criminal justice records and the quality of said data.

Development and validation of a probabilistic algorithm requires training data. This was available in data from two jurisdictions in the form of agency identifiers those agencies have validated biometrically using fingerprints. A total of 2.7 million records were available in these two jurisdictions and allowed for the data to be split into two large training and validation samples. The final algorithm that was developed and implemented makes use of full name and date of birth to identify individuals.

3.4.2 Episode resolution algorithm

Another common issue that researchers using administrative criminal justice data face is longitudinally tracking events that are linked to a single criminal justice episode. This is due in part to the siloing of records across different parts of the criminal justice system. As such, the CJARS project has developed a set of algorithms that can reconstruct criminal justice episodes based on the events linked to each episode.

Similar to entity resolution, CJARS used training data to develop the algorithm. There were a few instances where data that CJARS had acquired contained identifiers that allowed for linkages of events to be made (e.g., case number) so that an event and any subsequent events that occurred in a different part of the system could be tracked. Data sets that were linkable via an identifier were first merged and then split into training and validation samples.

The training data were used to develop a model using other available variables in the data to predict the likelihood that events within an individual are connected to a single event. A few examples of variables used to match events to an episode include event date and similarity of offense codes. After model development, a threshold was set to determine what events should be linked to a single episode based on predicted probability. After development of the model using the training sample and the performance of the model was checked against the true match status of events to episodes based on the available identifying variable in the training data, the model was validated on the validation sample.

4 Table and variable codebook

4.1 Variables by relational table

The CJARS relational databases contain detailed information about arrests, jail stays, adjudications, incarcerations, and terms of parole and probation. Additionally, CJARS provides tables detailing demographic information associated with each `cjars_id`, as well as coverage guidance. Primary key variables are shown in bold.

> USER NOTE
>
> A relational table containing information on DUIs is included in the CJARS distribution available on the FSRDCs, but it is experimental, unprocessed, and not research-ready.

- **Demographics**. The demographics table contains the unique list of person-level `cjars_ids`. Data is at the person-level.
  - **`cjars_id`**: CJARS person identifier
  - `dob_yyyy`: Date of birth, year
  - `sex`: Sex
  - `sex_raw`: Sex, raw
  - `sex_imputed`: Sex imputation indicator
  - `race`: Race and ethnicity
  - `race_raw`: Race and ethnicity, raw
  - `race_imputed`: Race imputation flag
- **Arrest**. The arrest table contains information on arrest dates and associated offenses. Data is at the offense-level.
  - `cjars_id`: CJARS person identifier
  - **`arr_id`**: Arrest identifier
  - `arr_dt_yyyy`: Arrest date, year
  - `arr_dt_mm`: Arrest date, month
  - `arr_dt_dd`: Arrest date, day of month
  - `arr_off_cd`: CJARS standardized offense code - arresting offense
  - `arr_off_cd_src`: Raw offense code from source - arresting offense
  - `arr_dv_off`: Domestic violence offense flag
  - `arr_wpn_off`: Weapon offense flag
  - `arr_st_juris_fips`: State FIPS code for location of arrest
  - `arr_cnty_juris_fips`: County FIPS code for location of arrest
  - `arr_rec_src_bkg`: Record source - booking agency
  - `arr_rec_src_arr`: Record source - arresting agency
  - `arr_rec_src_crt`: Record source - courts
  - `arr_rec_src_doc`: Record source - department of corrections
  - `arr_rec_src_rep`: Record source - criminal history repository
  - `arr_rec_src_cc`: Record source - community corrections agency
- **Jail**. The jail table contains information on jail entry and exit dates, associated bail dates and amounts, and related charge information. Data is at the charge-level.
  - `cjars_id`: CJARS person identifier
  - **`jail_id`**: Booking offense identifier
  - `jail_entry_dt_yyyy`: Jail entry date, year
  - `jail_entry_dt_mm`: Jail entry date, month
  - `jail_entry_dt_dd`: Jail entry date, day of month
  - `jail_exit_dt_yyyy`: Jail exit date, year
  - `jail_exit_dt_mm`: Jail exit date, month
  - `jail_exit_dt_dd`: Jail exit date, day of month
  - `jail_bail_amt`: Bail amount in dollars
  - `jail_off_cd`: CJARS standardized offense code - booking offense
  - `jail_off_cd_src`: Raw offense code from source - booking offense

  - jail_dv_off: Domestic violence offense flag
  - jail_wpn_off: Weapon offense flag
  - jail_grd_cd: CJARS standardized offense grade
  - jail_grd_cd_src: Raw offense grade from source (e.g., felony, misdemeanor)
  - jail_st_ori_fips: State FIPS code for location of originating event
  - jail_cnty_ori_fips: County FIPS code for location of originating event
  - jail_st_juris_fips: FIPS code of state responsible for physical supervision of individual
  - jail_cnty_juris_fips: FIPS code of county responsible for physical supervision of individual
  - jail_rec_src_bkg: Record source - booking agency
  - jail_rec_src_arr: Record source - arresting agency
  - jail_rec_src_crt: Record source - courts
  - jail_rec_src_doc: Record source - department of corrections
  - jail_rec_src_rep: Record source - criminal history repository
  - jail_rec_src_cc: Record source - community corrections agency
- **Adjudication**. The adjudication table contains detailed information about the offense the person was charged with, disposition information, and sentencing. Data is at the charge-level.
  - cjars_id: CJARS person identifier
  - **adj_id**: Court case filing identifier
  - adj_grd_cd: CJARS standardized offense grade
  - adj_grd_cd_src: Raw offense grade from source (e.g., felony, misdemeanor)
  - adj_off_lgl_cd: CJARS standardized offense legal code
  - adj_off_lgl_cd_src: Raw legal code from source (e.g., ordinance violation)
  - adj_file_dt_yyyy: Case filing date, year
  - adj_file_dt_mm: Case filing date, month
  - adj_file_dt_dd: Case filing date, day of month
  - adj_chrg_off_cd: CJARS standardized charge offense - offense charged at case filing
  - adj_chrg_off_cd_src: Raw charge offense description from source - offense charged at case filing
  - adj_disp_dt_yyyy: Case disposition date, year
  - adj_disp_dt_mm: Case disposition date, month
  - adj_disp_dt_dd: Case disposition date, day of month
  - adj_disp_cd: CJARS standardized disposition
  - adj_disp_cd_src: Raw disposition description from source
  - adj_disp_off_cd: CJARS standardized disposition offense - offense recorded at disposition
  - adj_disp_off_cd_src: Raw disposition offense from source - offense recorded at disposition
  - adj_dv_off: Domestic violence offense flag
  - adj_wpn_off: Weapon offense flag
  - adj_off_dt_yyyy: Date offense was committed, year
  - adj_off_dt_mm: Date offense was committed, month
  - adj_off_dt_dd: Date offense was committed, day of month
  - adj_sent_dt_yyyy: Sentencing date, year
  - adj_sent_dt_mm: Sentencing date, month
  - adj_sent_dt_dd: Sentencing date, day of month
  - adj_sent_serv: Community service sentence
  - adj_sent_dth: Death sentence
  - adj_sent_inc: Incarceration length in months
  - adj_sent_pro: Probation length in months
  - adj_sent_rest: Restitution amount in dollars
  - adj_sent_sus: Suspended sentence
  - adj_sent_trt: Treatment sentence
  - adj_sent_fine: Fine amount in dollars
  - adj_sent_inc_min: Minimum incarceration term in months
  - adj_sent_inc_max: Maximum incarceration term in months
  - adj_sent_src: Raw sentence from source
  - adj_st_juris_fips: State FIPS code for location of filing

  - adj_cnty_juris_fips: County FIPS code for location of filing
  - adj_rec_src_bkg: Record source - booking agency
  - adj_rec_src_arr: Record source - arresting agency
  - adj_rec_src_crt: Record source - courts
  - adj_rec_src_doc: Record source - department of corrections
  - adj_rec_src_rep: Record source - criminal history repository
  - adj_rec_src_cc: Record source - community corrections agency
- **Probation**. The probation table contains information on probation conditions, probation begin status and date, and probation end status and date. Data is at the level of a probation term.
  - cjars_id: CJARS person identifier
  - **pro_id**: Probation term identifier
  - pro_cond_cd: CJARS standardized probation conditions
  - pro_cond_cd_src: Raw description of probation conditions from source
  - pro_bgn_dt_yyyy: Probation start date, year
  - pro_bgn_dt_mm: Probation start date, month
  - pro_bgn_dt_dd: Probation start date, day of month
  - pro_end_dt_yyyy: Probation end date, year
  - pro_end_dt_mm: Probation end date, month
  - pro_end_dt_dd: Probation end date, day of month
  - pro_end_cd: CJARS standardized probation end status
  - pro_end_cd_src: Raw description of probation end status from source
  - pro_st_ori_fips: State FIPS code for location of sentencing
  - pro_cnty_ori_fips: County FIPS code for location of sentencing
  - pro_st_juris_fips: FIPS code of state with jurisdiction over supervision of individual
  - pro_cnty_juris_fips: FIPS code of county with jurisdiction over supervision of individual
  - pro_rec_src_bkg: Record source - booking agency
  - pro_rec_src_arr: Record source - arresting agency
  - pro_rec_src_crt: Record source - courts
  - pro_rec_src_doc: Record source - department of corrections
  - pro_rec_src_rep: Record source - criminal history repository
  - pro_rec_src_cc: Record source - community corrections agency
- **Incarceration**. The incarceration table contains information about the facility where an individual is/was housed, entry and exit dates, as well as the current status of the person. Data is at the level of an incarceration term.
  - cjars_id: CJARS person identifier
  - **inc_id**: Incarceration term identifier
  - inc_fcl_cd: CJARS standardized facility type
  - inc_fcl_cd_src: Raw description of facility from source
  - inc_entry_dt_yyyy: Incarceration entry date, year
  - inc_entry_dt_mm: Incarceration entry date, month
  - inc_entry_dt_dd: Incarceration entry date, day of month
  - inc_entry_cd: CJARS standardized entry status
  - inc_entry_cd_src: Raw description of entry type into incarceration
  - inc_exit_dt_yyyy: Incarceration exit date, year
  - inc_exit_dt_mm: Incarceration exit date, month
  - inc_exit_dt_dd: Incarceration exit date, day of month
  - inc_exit_cd: CJARS standardized exit status
  - inc_exit_cd_src: Raw description of exit type from incarceration
  - inc_st_ori_fips: State FIPS code for location of sentencing
  - inc_cnty_ori_fips: County FIPS code for location of sentencing
  - inc_st_juris_fips: FIPS code of state responsible for physical supervision of individual
  - inc_rec_src_bkg: Record source - booking agency
  - inc_rec_src_arr: Record source - arresting agency
  - inc_rec_src_crt: Record source - courts

  - inc_rec_src_doc: Record source - department of corrections
  - inc_rec_src_rep: Record source - criminal history repository
  - inc_rec_src_cc: Record source - community corrections agency
- **Parole**. The parole table contains information on parole begin/end dates and exit status when available. Data is at the level of a parole term.
  - cjars_id: CJARS person identifier
  - **par_id**: Parole term identifier
  - par_bgn_dt_yyyy: Parole start date, year
  - par_bgn_dt_mm: Parole start date, month
  - par_bgn_dt_dd: Parole start date, day of month
  - par_end_dt_yyyy: Parole end date, year
  - par_end_dt_mm: Parole end date, month
  - par_end_dt_dd: Parole end date, day of month
  - par_end_cd: Parole end status classification
  - par_end_cd_src: Raw description of parole end status from source
  - par_st_ori_fips: State FIPS code for location of sentencing
  - par_cnty_ori_fips: County FIPS code for location of sentencing
  - par_st_juris_fips: FIPS code of state with jurisdiction over supervision of individual
  - par_rec_src_bkg: Record source - booking agency
  - par_rec_src_arr: Record source - arresting agency
  - par_rec_src_crt: Record source - courts
  - par_rec_src_doc: Record source - department of corrections
  - par_rec_src_rep: Record source - criminal history repository
  - par_rec_src_cc: Record source - community corrections agency
- **Coverage**. The coverage table contains information on where and when jurisdictions are covered in the CJARS data. Data is at the level of a county-month. Data coverage is determined by a combination of automated and manual processes. First, an algorithm is run that was designed to determine the beginning and end of data coverage in a jurisdiction by identifying relatively large increases/decreases from month-to-month of events that are observed in the data. Next, CJARS staff with expertise in data collection and processing compare coverage beginning and end dates that were initially algorithmically determined, and compare them against basic frequency distributions of events observed in the data. Finally, CJARS staff make a final determination based on their informed approximation of data coverage. For additional details on coverage use, please refer to Appendix B
  - st_fips: State FIPS code
  - cnty_fips: County FIPS code
  - cjars_table: Record type
  - month: Year/month of coverage
  - dt_yyyy: Coverage month, year
  - dt_mm: Coverage date, month
  - coverage: Primary vs. secondary source

4.2 Variable codebook

Each of the tables that comprise the CJARS data contains a unique set of variables that capture the core variables describing the events that are contained in each data table. The large amount of variation in data from jurisdiction to jurisdiction has led to a significant amount of effort that was put into harmonizing data from all of the disparate sources into one common format. The following tables provide a list of the harmonized variables in each table.

These tables provide basic information about the variables such as variable names, labels, formats, a description of each variable, and basic descriptive statistics. This descriptive information includes the total number of records for each variable and counts of valid, invalid, and missing values. For the sake of clarity, these terms can be defined as:

- Valid values: values within an acceptable numeric range for numeric variables or that fall within a code scheme for qualitative variables.

- Invalid values: values that are outside of the range of what should be possible (e.g., a value of 32 for the day of the month of an event) or values that are not present in the coding schema of qualitative variables.
- Missing: in numerical data, values are marked as missing if they were missing in the original, raw data. In most qualitative code schemes, there is an explicit category for missing values, and these are enumerated as missing in the variable-level missing count.

All statistical material in this section is based on CJARS data at the University of Michigan. None is based on Census Bureau data protected by 13 USC §9a.

> **USER NOTE**
>
> The summary statistics reported for each variable are representative of the complete CJARS database. They may differ for the version of the CJARS database within the VDE due to difference in VDE data coverage.

4.2.1 Demographics

USER NOTE

The CJARS demographics table is delivered with duplicate rows containing distinct identifiable information to maximize record linkage success. Thus, `cjars_ids` can be included more than once when an individual's information varies across entries even though we have probabilistically linked the entries to one person.

USER NOTE

The CJARS demographics table is the product of entity resolution algorithms that make use of information from a variety of data sources. Membership in the CJARS demographics table should *never* be used as a proxy for involvement in the criminal justice system.

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.3.1 and Figure 6. Table keys, including the `cjars_id`, should not be assumed to be stable across CJARS vintages. They should only be used in the context of a single vintage.
Table	Demographics
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Unique values	47,814,328	

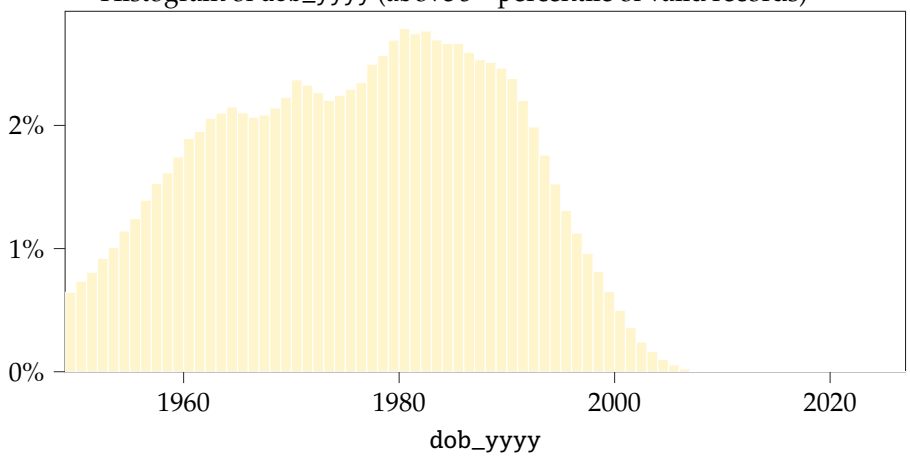
dob_yyyy

Label	Date of birth, year
Description	The year when the individual was born.
Table	Demographics
Format	numeric

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	77,192,121	90.0
Invalid values	11,850	0.0
Missing values	8,558,743	10.0

Statistic	*Value*
Mean	1,974.0
Minimum	1,900
25th percentile	1,964
Median	1,976
75th percentile	1,985
Maximum	2,026

Histogram of dob_yyyy (above 5th percentile of valid records)

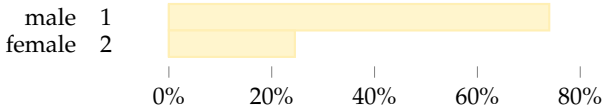


`sex`

Label	Sex
Description	A variable indicating the individual's sex/gender created by CJARS and based on sex/gender in the source data if available. Please note that it is not always clear in source data that is collected whether sex, gender, or some combination of both was provided due to a lack of available documentation that defines source data fields. Therefore, this variable may contain some combination of sex and/or gender, depending on what information was collected. If missing, sex is imputed based on the frequency of the first name in the CJARS data by Census region by sex. If there are insufficient observations of first name by Census region, sex is imputed based on frequency of first name in Census 2000 by state.
Table	Demographics
Format	numeric
Code scheme	sex code

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	84,458,342	98.5
Missing values	1,304,372	1.5

Histogram of `sex` (valid records)

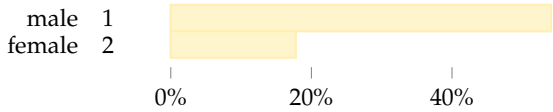


`sex_raw`

Label	Sex, raw
Description	The individual's sex/gender as recorded by the agency. Please note that it is not always clear in source data that is collected whether sex, gender, or some combination of both was provided due to a lack of available documentation that defines source data fields. Therefore, this variable may contain some combination of sex and/or gender, depending on what information was collected.
Table	Demographics
Format	numeric
Code scheme	sex code

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	61,649,911	71.9
Missing values	24,112,803	28.1

Histogram of `sex_raw` (valid records)

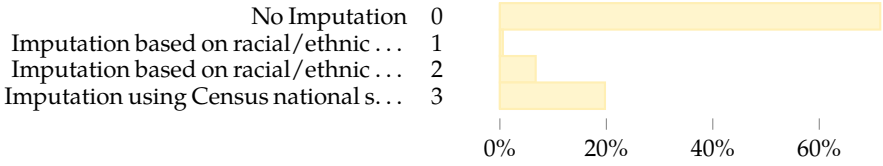


`sex_imputed`

Label	Sex imputation indicator
Description	An indicator for whether sex was imputed.
Table	Demographics
Format	numeric
Code scheme	demographic imputation code

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	84,458,342	98.5
Missing values	1,304,372	1.5

Histogram of `sex_imputed` (valid records)



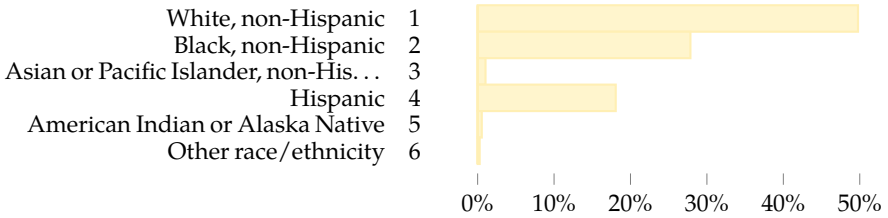
race

Label	Race and ethnicity
Description	A composite race and ethnicity variable imputed by CJARS and standardized to the CJARS race codes. The imputation is based on the frequency of observing the last name with a certain race in the CJARS data combined by Census region, first name with a certain race in the CJARS data combined by Census region, and on Census surnames data. The imputation prioritizes smaller race groups over large race groups (order of priority: AIAN, Asian, Hispanic, Black, White, Other) and differs by sex. Imputation thresholds vary based on the sex of the individual and whether the raw race was missing or being overwritten. If overwriting for males, then the frequency of the individual's first name or last name being associated with that race in the combined CJARS data by Census region must exceed 95% or the frequency of the surname being associated with that race in the Census surnames data must exceed 75%, or the joint probability of the first and last name being observed with that race in the combined CJARS data by Census region must exceed 75%. If imputing missing race for males, then either the last name or first name observed with a certain race must exceed 75% in the CJARS' data combined by Census region or the last name must exceed 75% in the Census surname data. If overwriting race for females, either the frequency of observing a first name with a certain race in a given Census region in the CJARS combined data must exceed 95% or the joint probability of observing a last and first name combination had to exceed 60%. If imputing missing race for females, either the frequency of a first name and race combination in the combined CJARS data in a given Census region must exceed 75% or the joint probability of observing the first and last name combination had to exceed 35%. For researchers interested in using information about race, we recommend requesting data resources available in the FSRDC network that can be linked to CJARS to determine individual race. Other FSRDC data resource on race are considered to be the higher quality measures of race as compared to administrative criminal justice records.
Table	Demographics
Format	numeric
Code scheme	race and ethnicity code

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	83,639,923	97.5
Missing values	2,122,791	2.5

`race` (continued)

Histogram of `race` (valid records)

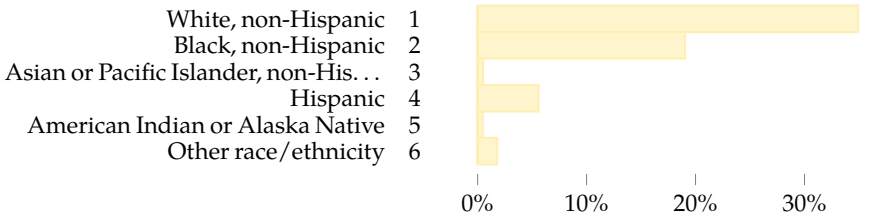


`race_raw`

Label	Race and ethnicity, raw
Description	A composite race and ethnicity variable standardized to the CJARS race codes and based on race and ethnicity in the source data if available.
Table	Demographics
Format	numeric
Code scheme	race and ethnicity code

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	53,421,699	62.3
Missing values	32,341,015	37.7

Histogram of `race_raw` (valid records)

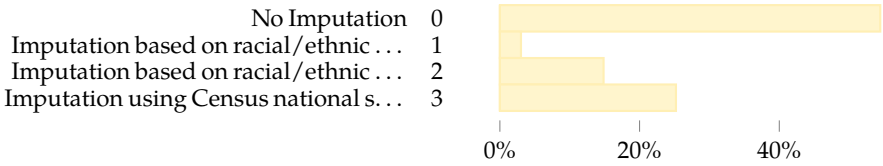


`race_imputed`

Label	Race imputation flag
Description	An indicator for whether the individual's race was imputed.
Table	Demographics
Format	numeric
Code scheme	demographic imputation code

Set	*Count*	*Percent (%)*
All records	85,762,714	100.0
Valid records	83,639,923	97.5
Missing values	2,122,791	2.5

Histogram of `race_imputed` (valid records)



4.2.2 Arrest

The arrest table contains information on arrest dates and associated offenses. Data is at the offense-level.

Table-level data notes

- All states: `US003`, `US004`
- Delaware: `DE001`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.3.1 and Figure 6.
Table	Arrest
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Unique values	9,689,789	

`arr_id`

Label	Arrest identifier
Description	Uniquely identifies arrest offenses. For more details on use of `arr_id` for data linkage, refer to Section 3.3.2 and Figure 7.
Table	Arrest
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Unique values	26,711,726	

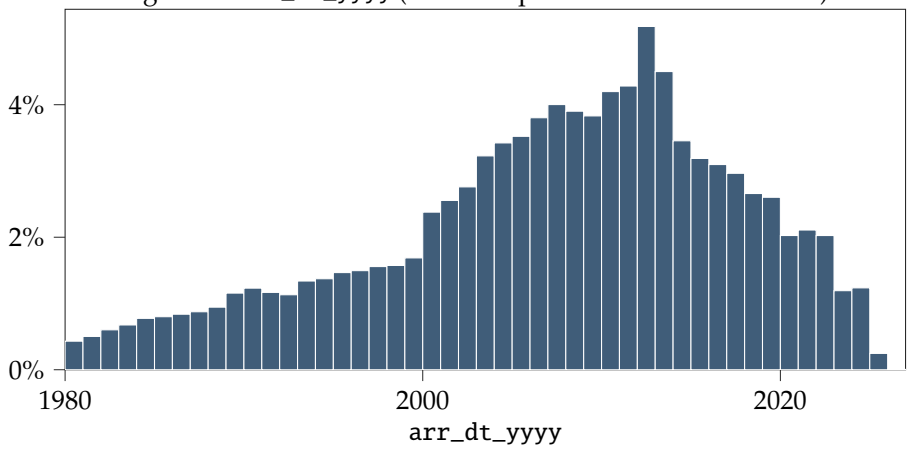
arr_dt_yyyy

Label	Arrest date, year
Description	The year when the individual was arrested.
Table	Arrest
Format	numeric

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	25,444,831	95.3
Invalid values	3	0.0
Missing values	1,266,892	4.7

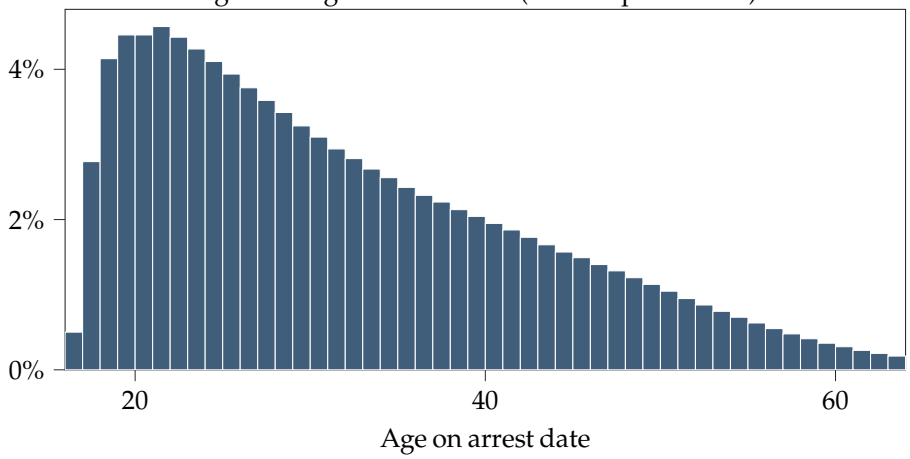
Statistic	*Value*
Mean	2,005.1
Minimum	1,900
25th percentile	1,999
Median	2,008
75th percentile	2,014
Maximum	2,026

Histogram of arr_dt_yyyy (above 5th percentile of valid records)



arr_dt_yyyy (continued)

Histogram of age on arrest date (1st–99th percentiles)

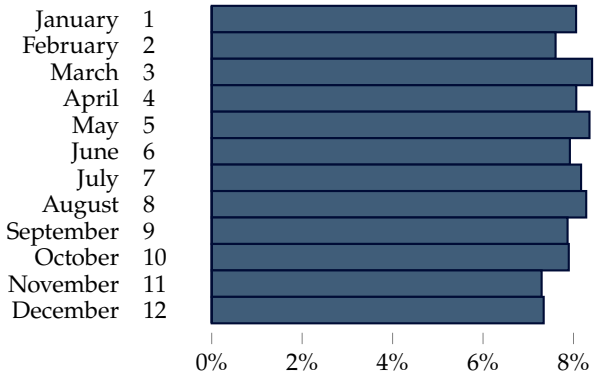


arr_dt_mm

Label	Arrest date, month
Description	The month when the individual was arrested.
Table	Arrest
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	25,444,834	95.3
Missing values	1,266,892	4.7

Histogram of `arr_dt_mm` (valid records)

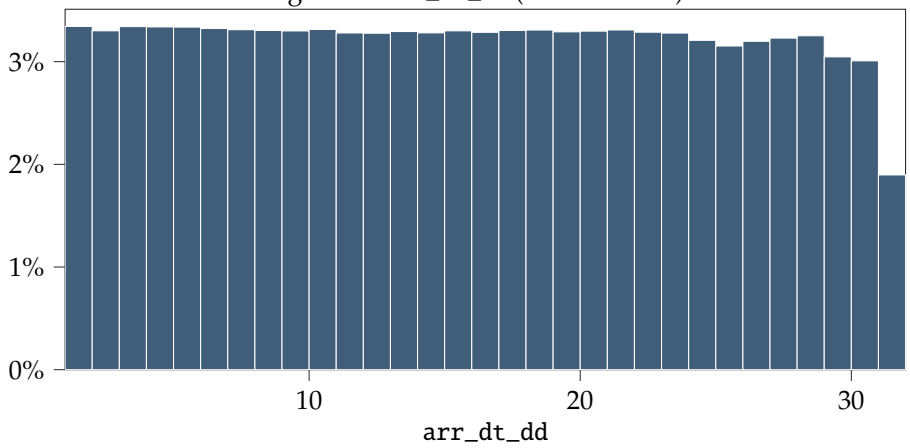


arr_dt_dd

Label	Arrest date, day of month
Description	The day of the month when the individual was arrested.
Table	Arrest
Format	numeric

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	25,444,834	95.3
Missing values	1,266,892	4.7

Histogram of `arr_dt_dd` (valid records)



arr_off_cd

Label	CJARS standardized offense code - arresting offense
Description	CJARS standardized offense code classification for arresting offense. Used when consistent offense coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize offense codes, please refer to Appendices C and E.2.1.
Table	Arrest
Format	string
Code scheme	offense classification

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Data notes

- California: `CA001`

arr_off_cd_src

Label	Raw offense code from source - arresting offense
Description	Original description of the arresting offense used by the agency where data was collected. Offense descriptions are agency-specific and thus inconsistent across jurisdictions. For harmonized offense code scheme please see `arr_off_cd`. Even though a harmonized offense code is provided, this raw description is retained for research that is not well-suited by the harmonized offense codes and leaves the opportunity for researchers to recode the original offense descriptions as needed. Note that original descriptions have been truncated to maximum 30 characters due to excessive memory string variables take up.
Table	Arrest
Format	string

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0

Data notes

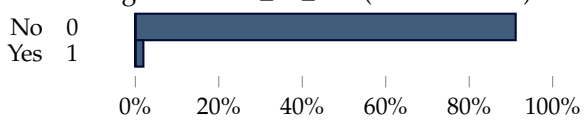
- All states: `US001`

arr_dv_off

Label	Domestic violence offense flag
Description	Whether the type of offense was flagged as domestic violence. It should be noted that this variable is experimental due to a lack of sufficient training data and should be used with some caution. Raw offense descriptions provided in the variable arr_off_cd_src can be inspected to evaluate appropriateness of offenses being flagged as domestic violence, especially for violent offenses (1st digit of arr_off_cd is 1) and family-related offenses (arr_off_cd is 5100, 5101, or 5102).
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	24,861,215	93.1
Missing values	1,850,511	6.9

Histogram of arr_dv_off (valid records)

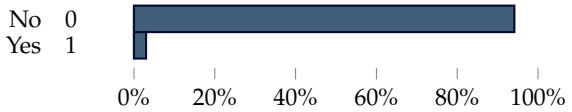


arr_wpn_off

Label	Weapon offense flag
Description	Whether the type of offense was flagged as involving a weapon. It should be noted that this variable is experimental due to a lack of sufficient training data and should be used with some caution. Raw offense descriptions provided in the variable arr_off_cd_src can be inspected to evaluate appropriateness of offenses being flagged as involving a weapon.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	25,946,168	97.1
Missing values	765,558	2.9

Histogram of arr_wpn_off (valid records)



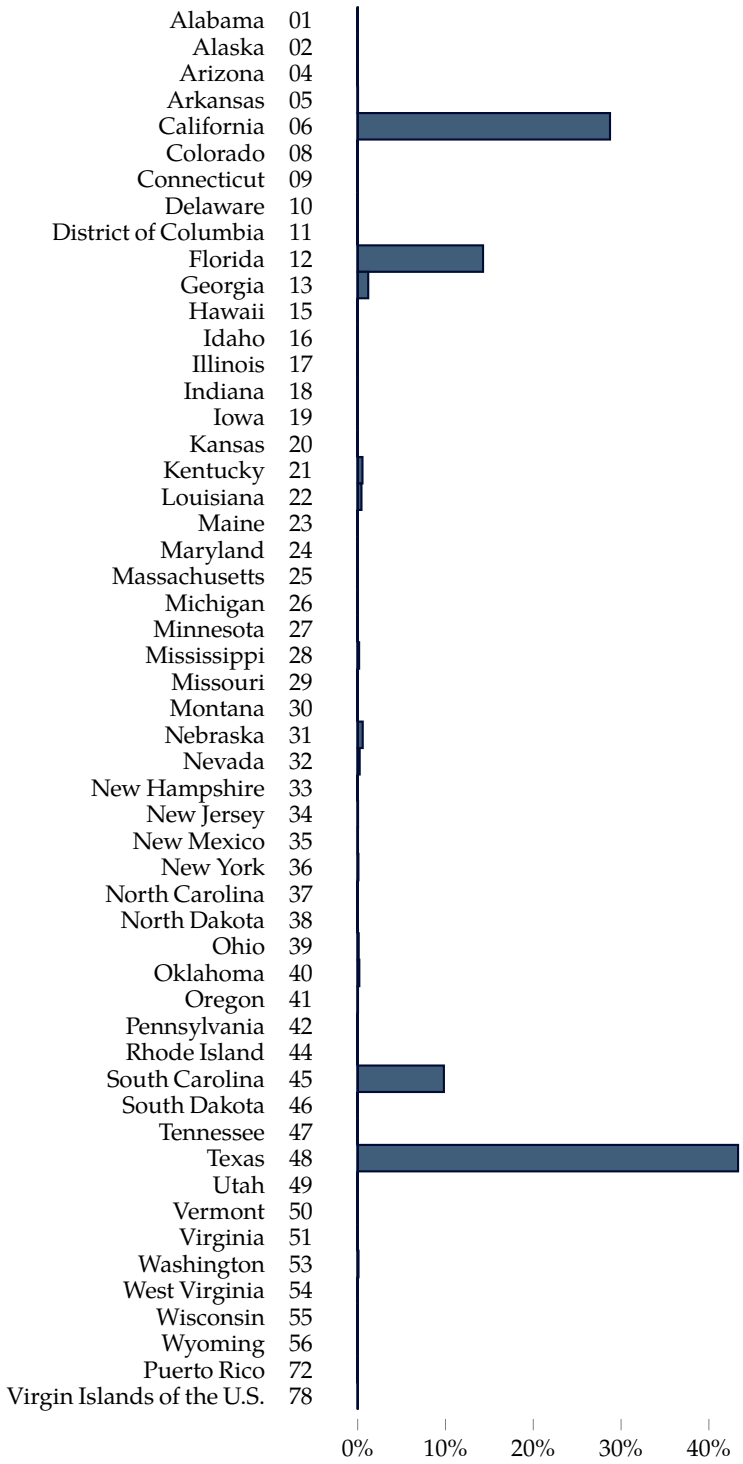
arr_st_juris_fips

Label	State FIPS code for location of arrest
Description	State-level Federal Information Processing Standards (FIPS) code where arrest occurred. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Arrest
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

`arr_st_juris_fips` (continued)

Histogram of `arr_st_juris_fips` (valid records)



arr_cnty_juris_fips

Label	County FIPS code for location of arrest
Description	County-level Federal Information Processing Standards (FIPS) code where occurred. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Arrest
Format	string
Code scheme	county FIPS code

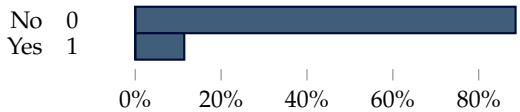
Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	24,007,903	89.9
Missing values	2,703,823	10.1

arr_rec_src_bkg

Label	Record source - booking agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a booking agency.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Histogram of `arr_rec_src_bkg` (valid records)

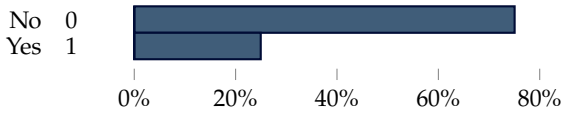


arr_rec_src_arr

Label	Record source - arresting agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from an arresting agency.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Histogram of `arr_rec_src_arr` (valid records)

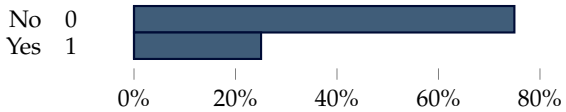


arr_rec_src_crt

Label	Record source - courts
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a court system.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Histogram of `arr_rec_src_crt` (valid records)

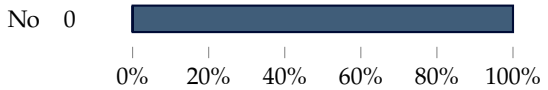


arr_rec_src_doc

Label	Record source - department of corrections
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a department of corrections.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Histogram of `arr_rec_src_doc` (valid records)

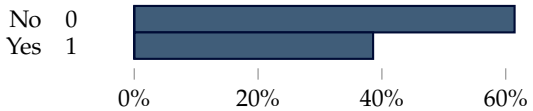


arr_rec_src_rep

Label	Record source - criminal history repository
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a criminal history repository.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Histogram of `arr_rec_src_rep` (valid records)

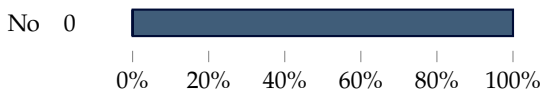


arr_rec_src_cc

Label	Record source - community corrections agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a community corrections agency.
Table	Arrest
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	26,711,726	100.0
Valid records	26,711,726	100.0

Histogram of `arr_rec_src_cc` (valid records)



4.2.3 Jail

The jail table contains information on jail entry and exit dates, associated bail dates and amounts, and related charge information. Data is at the charge-level.

Table-level data notes

- All states: US003, US004, US005
- Florida: FL006

cjars_id

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of cjars_id for data linkage, refer to Section 3.3.1 and Figure 6.
Table	Jail
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Unique values	7,258,049	

jail_id

Label	Booking offense identifier
Description	Uniquely identifies booking offenses. For more details on use of jail_id for data linkage, refer to Section 3.3.2 and Figure 7.
Table	Jail
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Unique values	29,576,717	

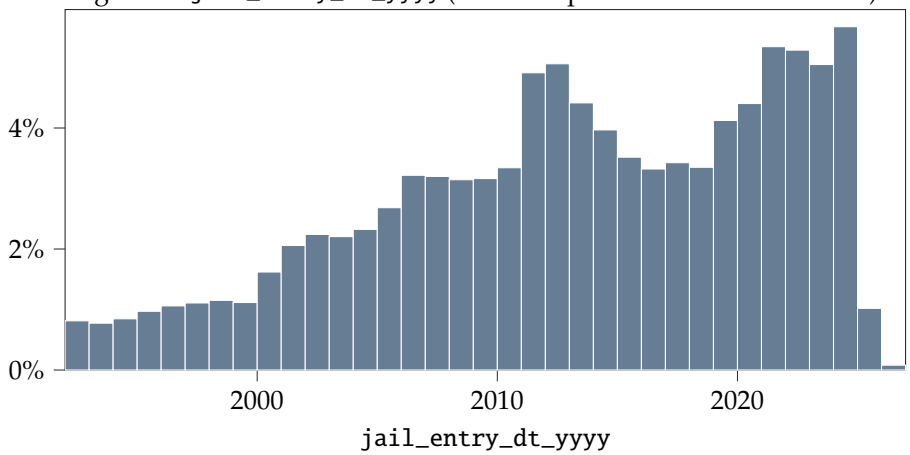
jail_entry_dt_yyyy

Label	Jail entry date, year
Description	The year when the individual entered into jail.
Table	Jail
Format	numeric

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	28,828,811	97.5
Invalid values	9	0.0
Missing values	747,897	2.5

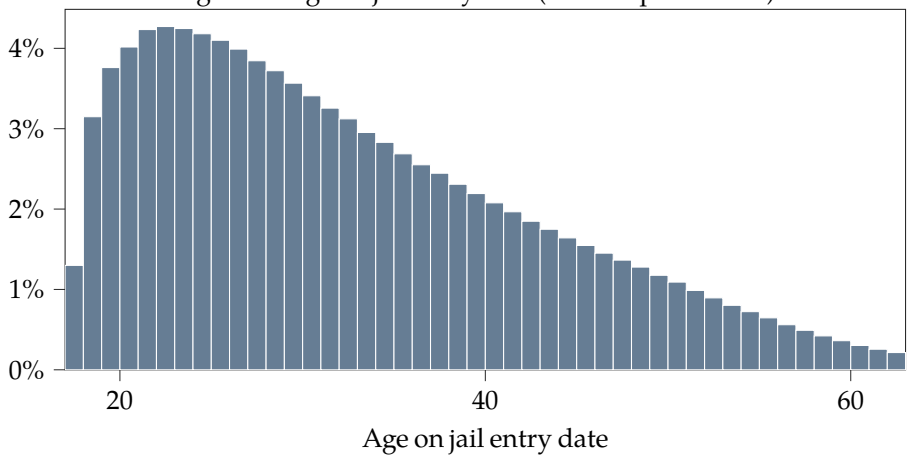
Statistic	*Value*
Mean	2,011.4
Minimum	1,900
25th percentile	2,006
Median	2,013
75th percentile	2,020
Maximum	2,026

Histogram of `jail_entry_dt_yyyy` (above 5th percentile of valid records)



`jail_entry_dt_yyyy` (continued)

Histogram of age on jail entry date (1st–99th percentiles)

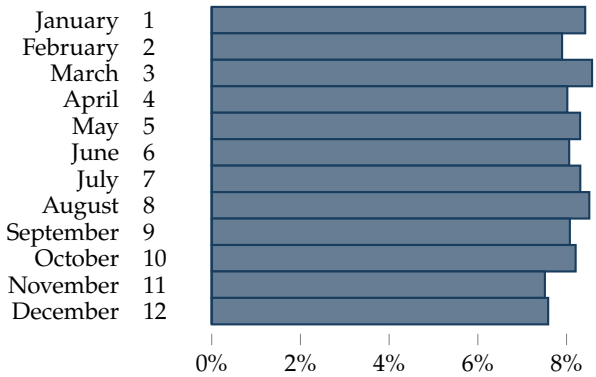


`jail_entry_dt_mm`

Label	Jail entry date, month
Description	The month when the individual entered into jail.
Table	Jail
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	28,828,820	97.5
Missing values	747,897	2.5

Histogram of `jail_entry_dt_mm` (valid records)



jail_entry_dt_dd

Label	Jail entry date, day of month
Description	The day of the month when the individual entered into jail.
Table	Jail
Format	numeric

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	28,828,820	97.5
Missing values	747,897	2.5

Histogram of `jail_entry_dt_dd` (valid records)

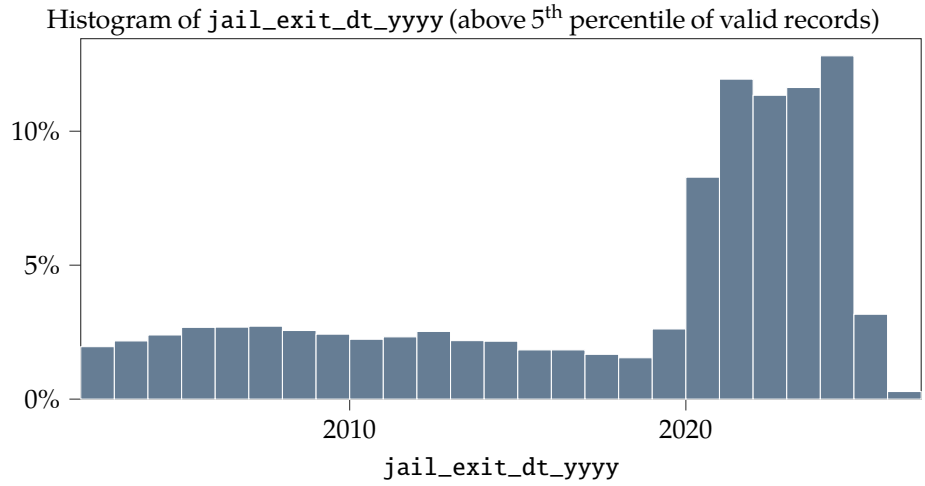
jail_exit_dt_yyyy

Label	Jail exit date, year
Description	The year when the individual exited jail.
Table	Jail
Format	numeric

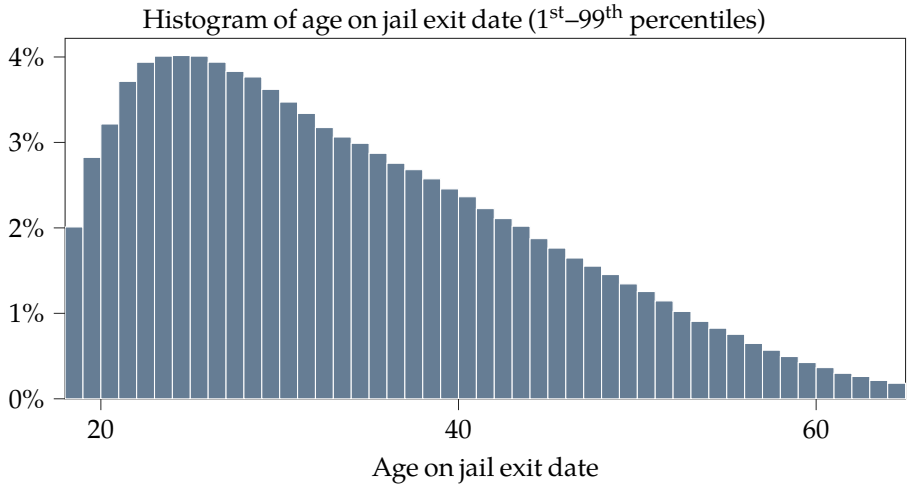
Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	7,894,192	26.7
Missing values	21,682,525	73.3

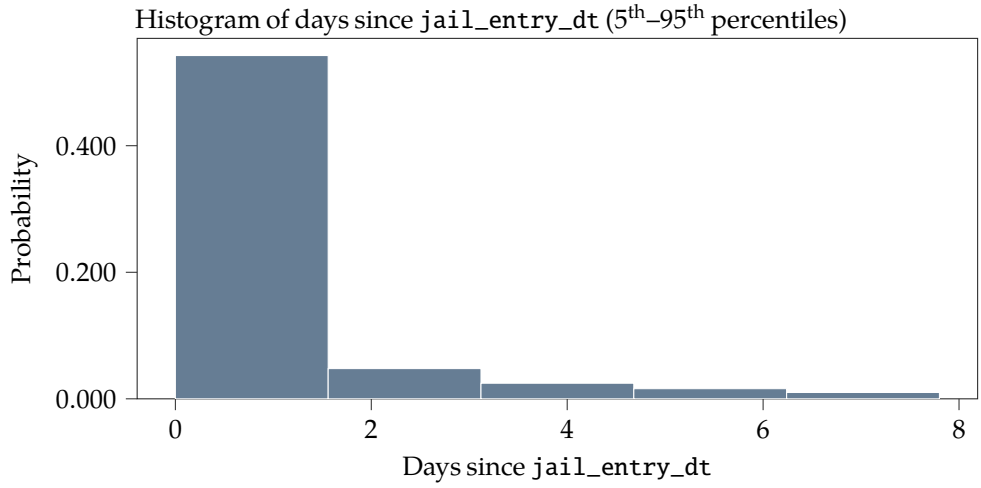
Statistic	*Value*
Mean	2,016.6
Minimum	1,900
25th percentile	2,011
Median	2,020
75th percentile	2,023
Maximum	2,026

Histogram of `jail_exit_dt_yyyy` (above 5th percentile of valid records)



`jail_exit_dt_yyyy` (continued)



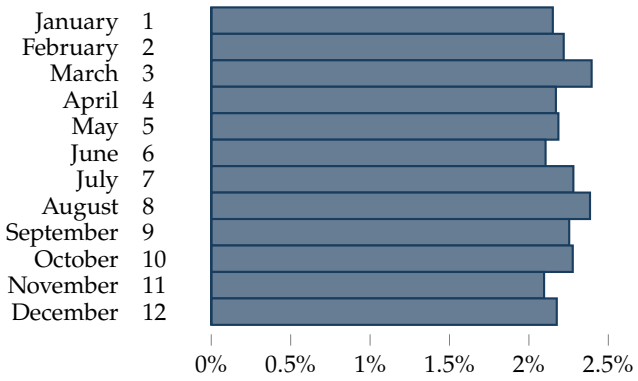


`jail_exit_dt_mm`

Label	Jail exit date, month
Description	The month when the individual exited jail.
Table	Jail
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	7,894,192	26.7
Missing values	21,682,525	73.3

Histogram of `jail_exit_dt_mm` (valid records)

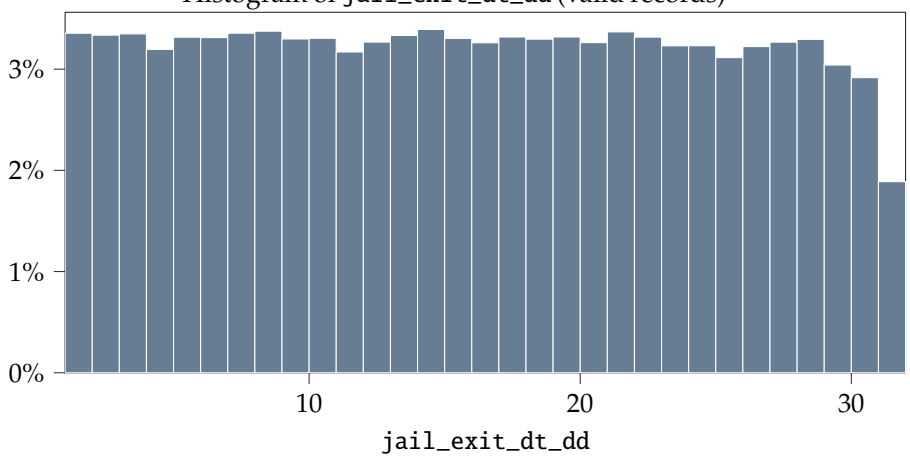


jail_exit_dt_dd

Label	Jail exit date, day of month
Description	The day of the month when the individual exited jail.
Table	Jail
Format	numeric

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	7,894,192	26.7
Missing values	21,682,525	73.3

Histogram of `jail_exit_dt_dd` (valid records)



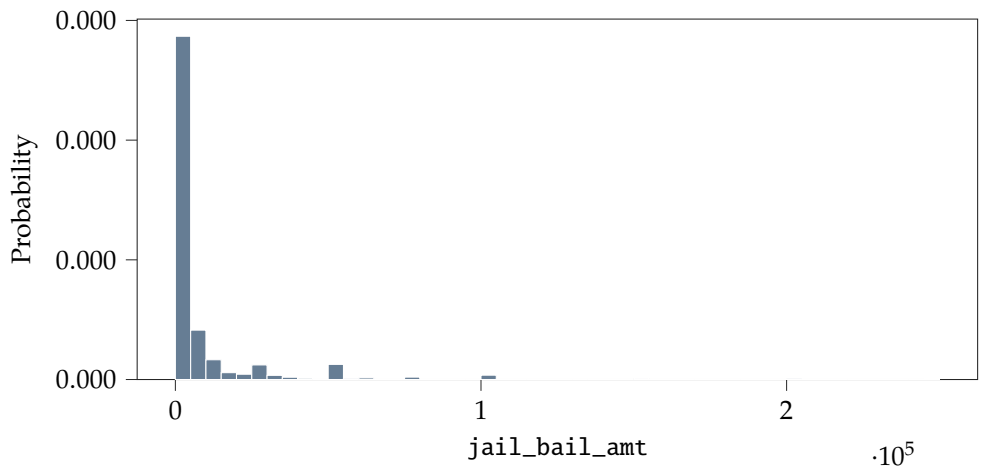
jail_bail_amt

Label	Bail amount in dollars
Description	The amount in dollars that the individual's bail had been set to.
Table	Jail
Format	numeric

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	2,995,133	10.1
Missing values	26,581,584	89.9

Statistic	*Value*
Mean	9,769.9
Minimum	0.0
25th percentile	0.0
Median	500.0
75th percentile	5,000.0
Maximum	500,000.0

Histogram of `jail_bail_amt` (1st–99th percentiles of valid records)



Data notes

- Georgia: GA005
- Indiana: IN002
- Kentucky: KY001, KY002, KY003, KY004, KY005
- Oklahoma: OK002
- Oregon: OR002

jail_off_cd

Label	CJARS standardized offense code - booking offense
Description	CJARS standardized offense code classification of the offense for which the individual was booked. Used when consistent offense coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize offense codes, please refer to Appendices C and E.2.1.
Table	Jail
Format	string
Code scheme	offense classification

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Data notes

- Alabama: `AL001`

jail_off_cd_src

Label	Raw offense code from source - booking offense
Description	Original description of the booking offense used by the agency where data was collected. Offense descriptions are agency-specific and thus inconsistent across jurisdictions. For harmonized offense code scheme please see `jail_off_cd`. Even though a harmonized offense code is provided, this raw description is retained for research that is not well-suited by the harmonized offense codes and leaves the opportunity for researchers to recode the original offense descriptions as needed. Note that original descriptions have been truncated to maximum 30 characters due to excessive memory string variables take up.
Table	Jail
Format	string

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0

Data notes

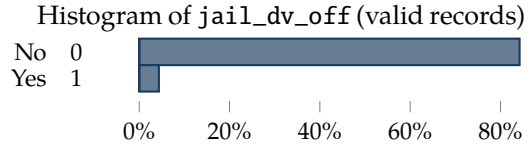
- All states: `US001`
- Alabama: `AL001`

jail_dv_off

Label	Domestic violence offense flag
Description	Whether the type of offense was flagged as domestic violence. It should be noted that this variable is experimental due to a lack of sufficient training data and should be used with some caution. Raw offense descriptions provided in the variable jail_off_cd_src can be inspected to evaluate appropriateness of offenses being flagged as domestic violence, especially for violent offenses (1st digit of jail_off_cd is 1) and family-related offenses (jail_off_cd is 5100, 5101, or 5102).
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	26,199,921	88.6
Missing values	3,376,796	11.4

Histogram of jail_dv_off (valid records)

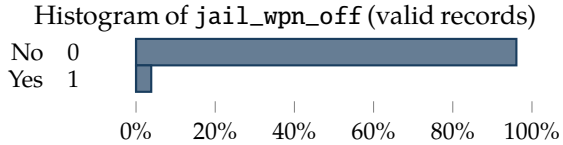


jail_wpn_off

Label	Weapon offense flag
Description	Whether the type of offense was flagged as involving a weapon. It should be noted that this variable is experimental due to a lack of sufficient training data and should be used with some caution. Raw offense descriptions provided in the variable `jail_off_cd_src` can be inspected to evaluate appropriateness of offenses being flagged as involving a weapon.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,549,470	99.9
Missing values	27,247	0.1

Histogram of `jail_wpn_off` (valid records)

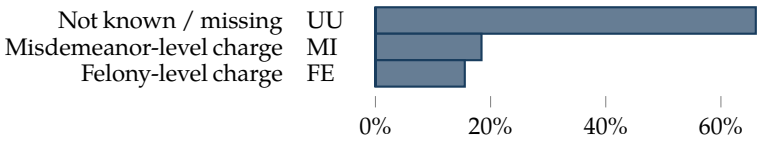


`jail_grd_cd`

Label	CJARS standardized offense grade
Description	CJARS standardized classification of an offense by its severity. For a full description of the coding categories and method used to harmonize offense grade, please refer to Appendices C and E.2.2.
Table	Jail
Format	string
Code scheme	charge grade code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_grd_cd` (valid records)



jail_grd_cd_src

Label	Raw offense grade from source (e.g., felony, misdemeanor)
Description	Original description of the crime severity used by the agency where data was collected. Offense grades are agency-specific and thus inconsistent across jurisdictions. For a harmonized offense grade please see jail_grd_cd. Even though a harmonized offense grade is provided, this raw description is retained for research that is not well-suited by the harmonized offense grade and leaves the opportunity for researchers to recode the original offense grade as needed. Note that original descriptions have been truncated to maximum 10 characters due to excessive memory string variables take up.
Table	Jail
Format	string

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0

Data notes

- All states: US001

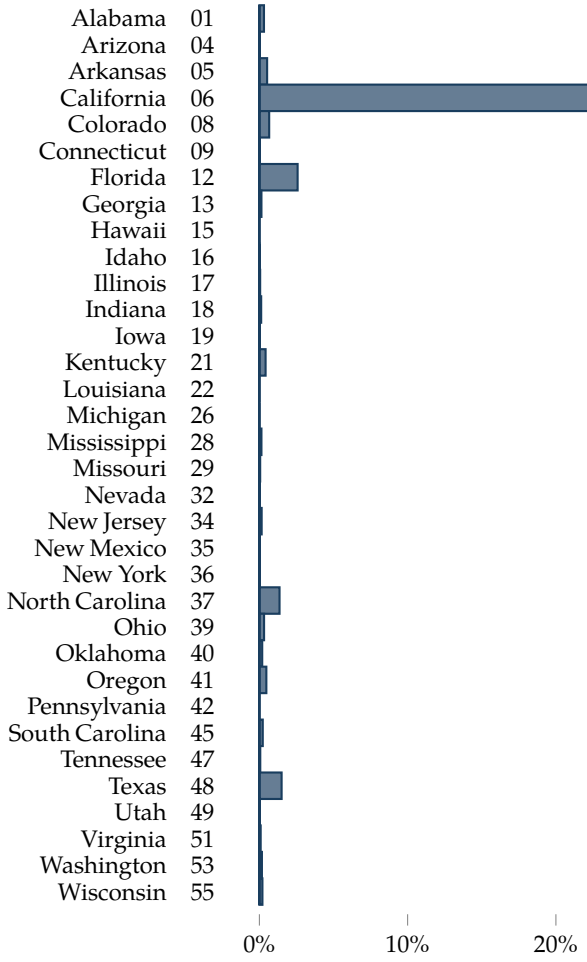
jail_st_ori_fips

Label	State FIPS code for location of originating event
Description	State-level Federal Information Processing Standards (FIPS) code indicating where the event (e.g., arrest, court sentence) leading to the individual's booking occurred. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Jail
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	10,479,926	35.4
Missing values	19,096,791	64.6

`jail_st_ori_fips` (continued)

Histogram of `jail_st_ori_fips` (valid records)



jail_cnty_ori_fips

Label	County FIPS code for location of originating event
Description	County-level Federal Information Processing Standards (FIPS) code indicating where the event (e.g., arrest, court sentence) leading to the individual's booking occurred. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Jail
Format	string
Code scheme	county FIPS code

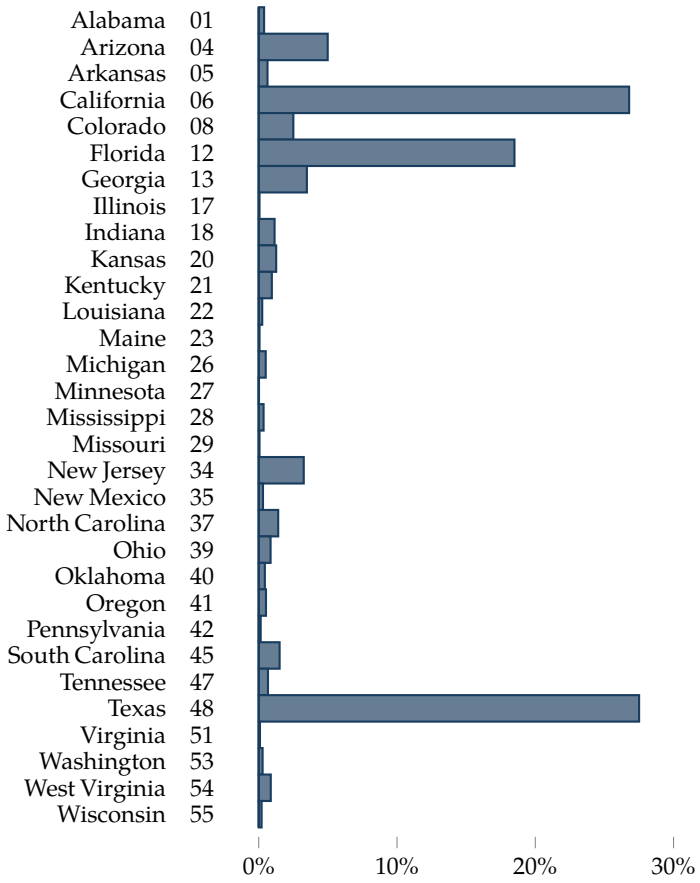
Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	9,744,531	32.9
Invalid values	1,756	0.0
Missing values	735,384	2.5

jail_st_juris_fips

Label	FIPS code of state responsible for physical supervision of individual
Description	State-level Federal Information Processing Standards (FIPS) code where the individual was booked. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Jail
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_st_juris_fips` (valid records)



jail_cnty_juris_fips

Label	FIPS code of county responsible for physical supervision of individual
Description	County-level Federal Information Processing Standards (FIPS) code where the individual was booked. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Jail
Format	string
Code scheme	county FIPS code

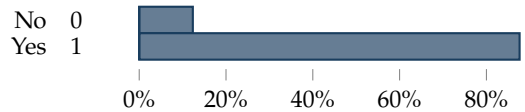
Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,319,444	99.1
Missing values	257,273	0.9

jail_rec_src_bkg

Label	Record source - booking agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a booking agency.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_rec_src_bkg` (valid records)

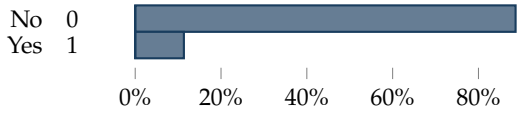


jail_rec_src_arr

Label	Record source - arresting agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from an arresting agency.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_rec_src_arr` (valid records)

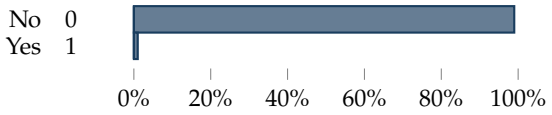


jail_rec_src_crt

Label	Record source - courts
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a court system.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_rec_src_crt` (valid records)

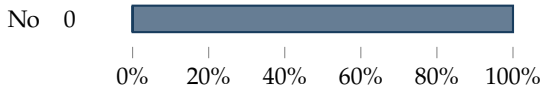


`jail_rec_src_doc`

Label	Record source - department of corrections
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a department of corrections.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_rec_src_doc` (valid records)

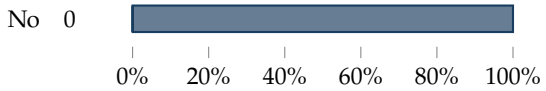


`jail_rec_src_rep`

Label	Record source - criminal history repository
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a criminal history repository.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_rec_src_rep` (valid records)

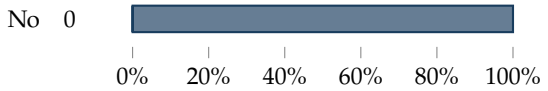


`jail_rec_src_cc`

Label	Record source - community corrections agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a community corrections agency.
Table	Jail
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	29,576,717	100.0
Valid records	29,576,717	100.0

Histogram of `jail_rec_src_cc` (valid records)



4.2.4 Adjudication

The adjudication table contains detailed information about the offense the person was charged with, disposition information, and sentencing. Data is at the charge-level.

> USER NOTE
>
> Data sourced from agencies that include offense information is used to populate records in the arrest, jail, and adjudication tables. Multiple charges for a given criminal justice event are typically kept as unique observations. However, there may be a loss of repeat charges for the same offense description in a case when deduplication occurs.

Table-level data notes

- All states: `US003`, `US004`
- Alaska: `AK001`
- Arizona: `AZ002`
- Arkansas: `AR001`
- Delaware: `DE001`
- Florida: `FL007`
- Maryland: `MD001`
- Minnesota: `MN001`
- New Jersey: `NJ001`, `NJ002`
- North Carolina: `NC002`, `NC004`
- Oregon: `OR001`
- South Carolina: `SC001`
- Texas: `TX003`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.3.1 and Figure 6.
Table	Adjudication
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Unique values	37,780,875	

adj_id

Label	Court case filing identifier
Description	Uniquely identifies charge filings. For more details on use of adj_id for data linkage, refer to Section 3.3.2 and Figure 7.
Table	Adjudication
Table key	primary, unique
Format	string

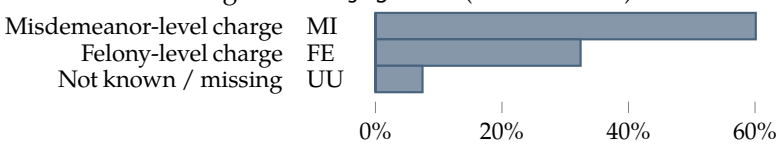
Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Unique values	162,034,905	

adj_grd_cd

Label	CJARS standardized offense grade
Description	CJARS standardized classification of an offense by its severity. For a full description of the coding categories and method used to harmonize offense grade, please refer to Appendices C and E.2.2.
Table	Adjudication
Format	string
Code scheme	charge grade code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of adj_grd_cd (valid records)



Data notes

- Florida: FL002
- Minnesota: MN003
- Nebraska: NE001
- Nevada: NV002

adj_grd_cd_src

Label	Raw offense grade from source (e.g., felony, misdemeanor)
Description	Original description of the crime severity used by the agency where data was collected. Offense grades are agency-specific and thus inconsistent across jurisdictions. For a harmonized offense grade please see adj_grd_cd. Even though a harmonized offense grade is provided, this raw description is retained for research that is not well-suited by the harmonized offense grade and leaves the opportunity for researchers to recode the original offense grade as needed. Note that original descriptions have been truncated to maximum 10 characters due to excessive memory string variables take up.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0

Data notes

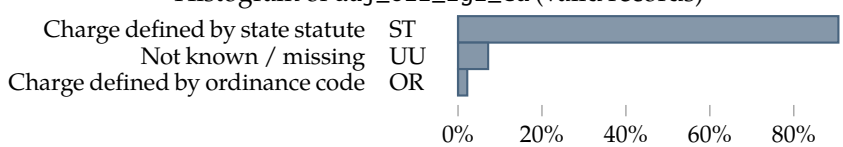
- All states: US001
- Minnesota: MN003
- Nebraska: NE001

adj_off_lgl_cd

Label	CJARS standardized offense legal code
Description	CJARS standardized classification of whether an offense was charged under state statute or an ordinance violation. For a full description of the coding categories and method used to harmonize offense legal code, please refer to Appendices C and E.2.3.
Table	Adjudication
Format	string
Code scheme	legal code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_off_lgl_cd` (valid records)



adj_off_lgl_cd_src

Label	Raw legal code from source (e.g., ordinance violation)
Description	Original description of the legal code used by the agency where data was collected. Legal codes are agency-specific and thus inconsistent across jurisdictions. For a harmonized legal code please see `adj_off_lgl_cd`. Even though a harmonized legal code is provided, this raw description is retained for research that is not well-suited by the harmonized legal code and leaves the opportunity for researchers to recode the original legal code as needed.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0

Data notes

- All states: `US001`

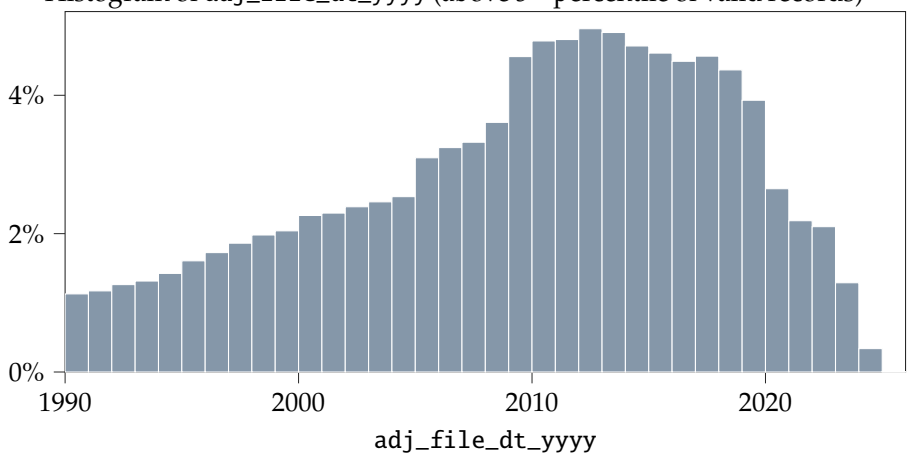
adj_file_dt_yyyy

Label	Case filing date, year
Description	The year when the individual's case was filed.
Table	Adjudication
Format	numeric

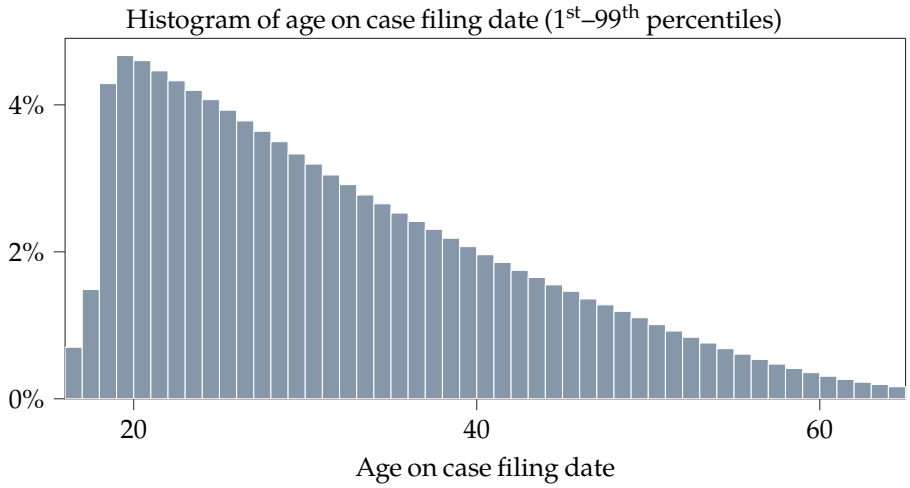
Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	126,396,083	78.0
Invalid values	1,194	0.0
Missing values	35,637,628	22.0

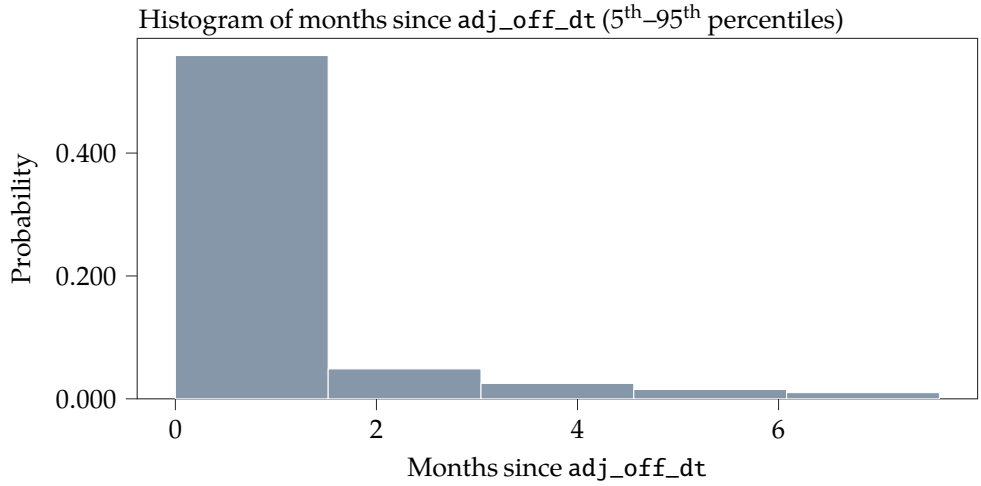
Statistic	*Value*
Mean	2,008.1
Minimum	1,900
25th percentile	2,002
Median	2,010
75th percentile	2,015
Maximum	2,025

Histogram of adj_file_dt_yyyy (above 5th percentile of valid records)



`adj_file_dt_yyyy` (continued)



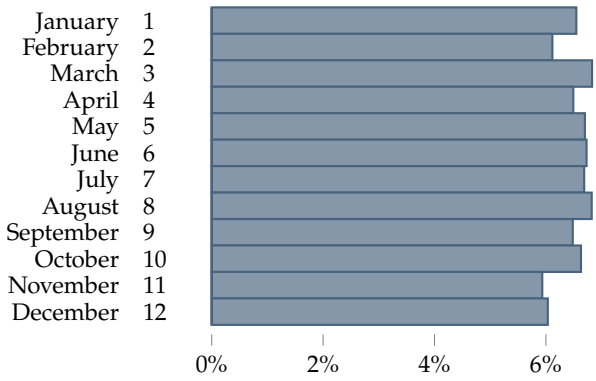


adj_file_dt_mm

Label	Case filing date, month
Description	The month when the individual's case was filed.
Table	Adjudication
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	126,397,277	78.0
Missing values	35,637,628	22.0

Histogram of `adj_file_dt_mm` (valid records)

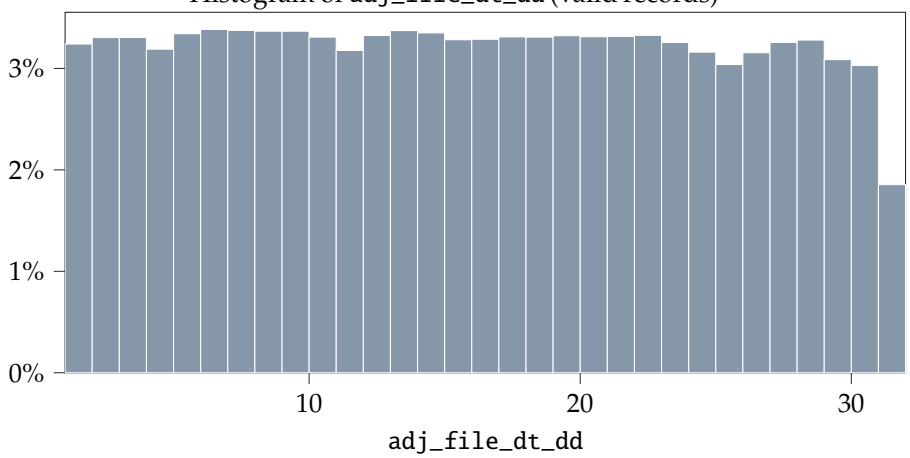


adj_file_dt_dd

Label	Case filing date, day of month
Description	The day of the month when the individual's case was filed.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	126,397,277	78.0
Missing values	35,637,628	22.0

Histogram of `adj_file_dt_dd` (valid records)



adj_chrg_off_cd

Label	CJARS standardized charge offense - offense charged at case filing
Description	CJARS standardized offense code classification for offense charged at case filing. Used when consistent offense coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize offense codes, please refer to Appendices C and E.2.1.
Table	Adjudication
Format	string
Code scheme	offense code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Data notes

- Florida: `FL002`, `FL004`

adj_chrg_off_cd_src

Label	Raw charge offense description from source - offense charged at case filing
Description	Original description of the offense charged at case filing used by the agency where data was collected. Offense descriptions are agency-specific and thus inconsistent across jurisdictions. For harmonized offense code scheme please see `adj_chrg_off_cd`. Even though a harmonized offense code is provided, this raw description is retained for research that is not well-suited by the harmonized offense codes and leaves the opportunity for researchers to recode the original offense descriptions as needed. Note that original offense descriptions have been truncated to maximum 30 characters due to excessive memory string variables take up.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0

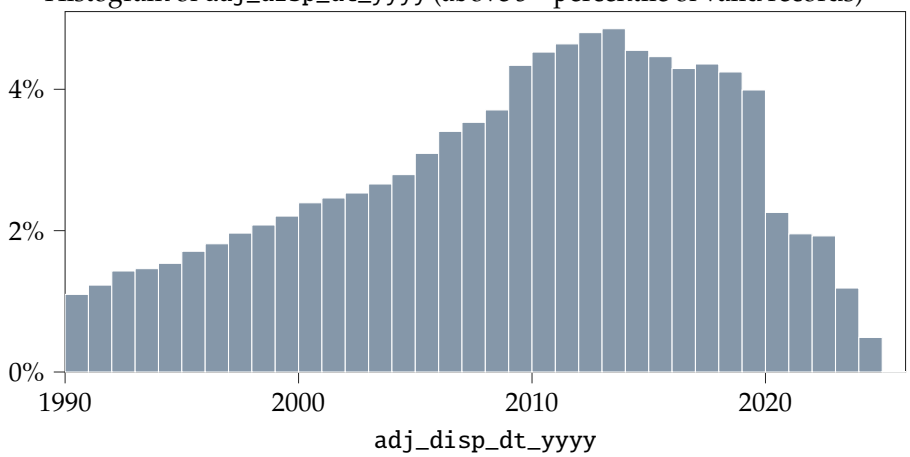
adj_disp_dt_yyyy

Label	Case disposition date, year
Description	The year when the individual's case was disposed.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	134,861,597	83.2
Invalid values	1,209	0.0
Missing values	27,172,099	16.8

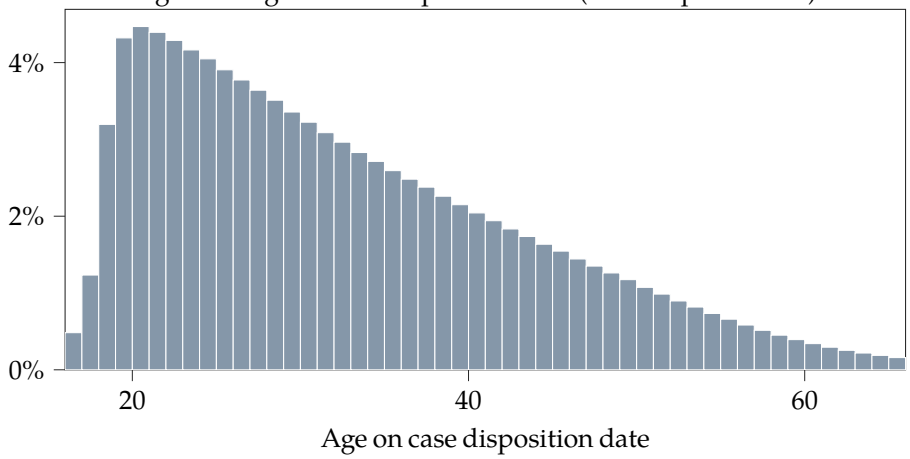
Statistic	*Value*
Mean	2,007.7
Minimum	1,900
25th percentile	2,001
Median	2,010
75th percentile	2,015
Maximum	2,025

Histogram of adj_disp_dt_yyyy (above 5th percentile of valid records)

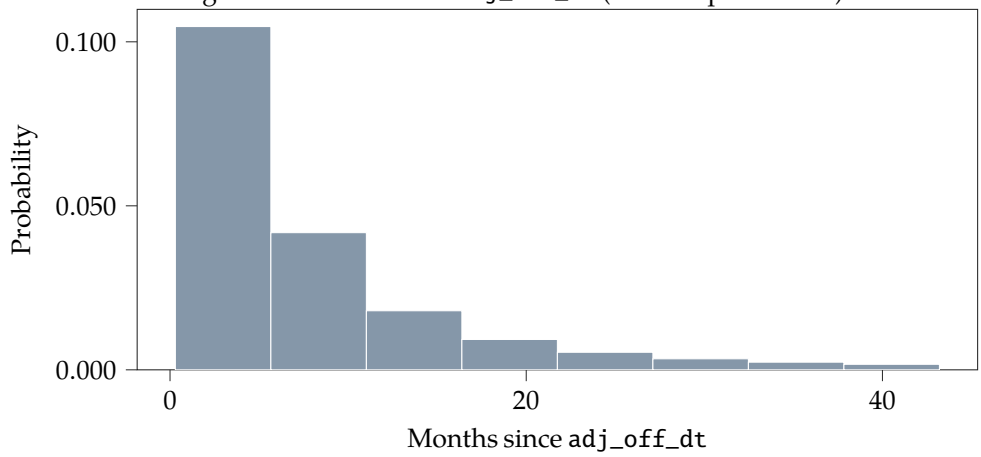


`adj_disp_dt_yyyy` (continued)

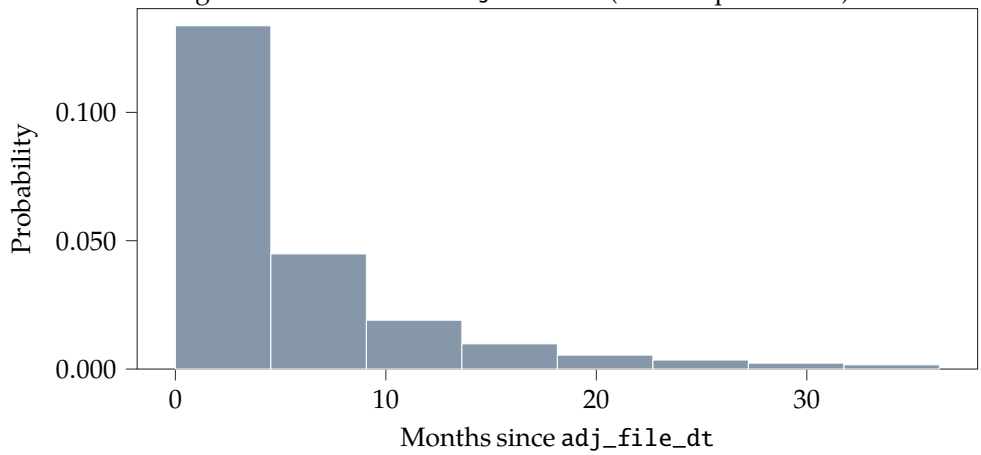
Histogram of age on case disposition date (1st–99th percentiles)



Histogram of months since `adj_off_dt` (5th–95th percentiles)



Histogram of months since `adj_file_dt` (5th–95th percentiles)

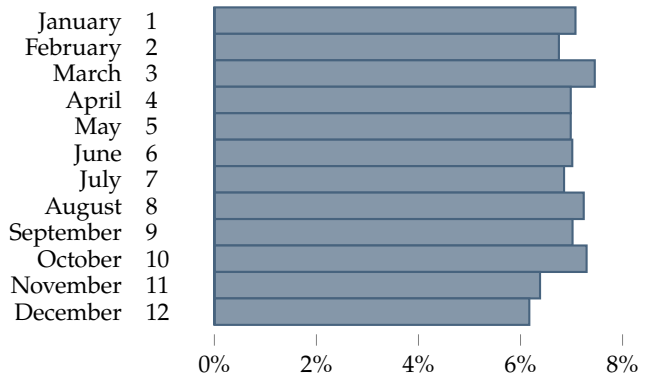


adj_disp_dt_mm

Label	Case disposition date, month
Description	The month when the individual's case was disposed.
Table	Adjudication
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	134,862,806	83.2
Missing values	27,172,099	16.8

Histogram of `adj_disp_dt_mm` (valid records)

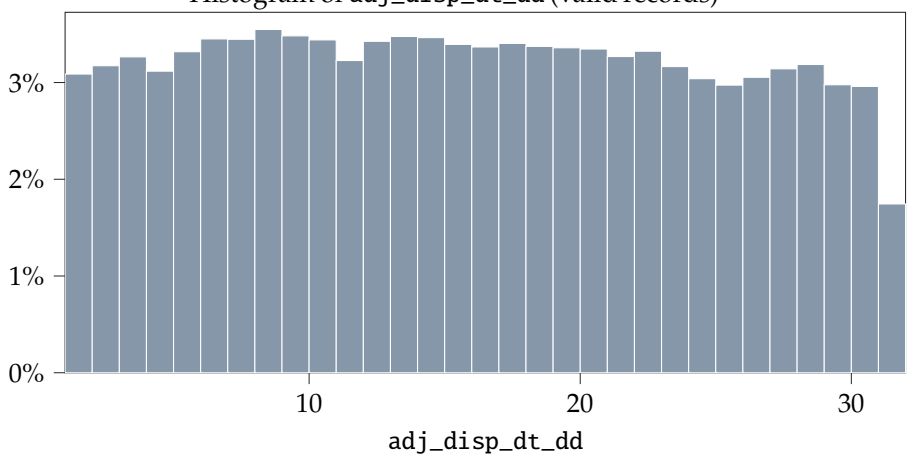


adj_disp_dt_dd

Label	Case disposition date, day of month
Description	The day of the month when the individual's case was disposed.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	134,862,806	83.2
Missing values	27,172,099	16.8

Histogram of `adj_disp_dt_dd` (valid records)

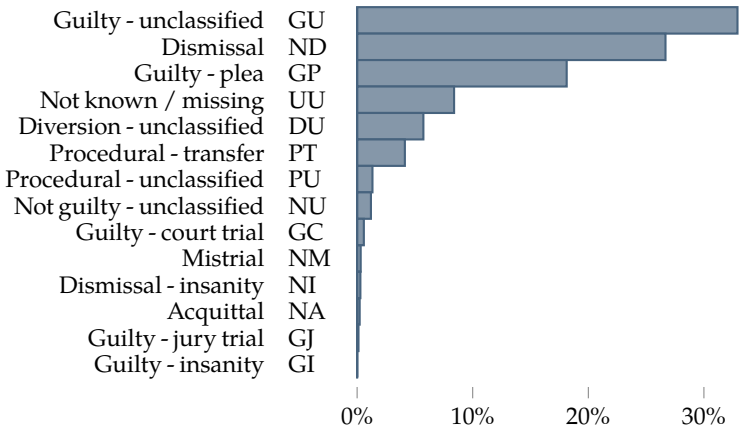


adj_disp_cd

Label	CJARS standardized disposition
Description	CJARS standardized disposition classification. Used when consistent disposition coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize disposition, please refer to Appendices C and E.2.4.
Table	Adjudication
Format	string
Code scheme	disposition code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_disp_cd` (valid records)



adj_disp_cd_src

Label	Raw disposition description from source
Description	Original description of the disposition used by the agency where data was collected. Dispositions are agency-specific and thus inconsistent across jurisdictions. For a harmonized disposition code please see adj_disp_cd. Even though a harmonized disposition code is provided, this raw description is retained for research that is not well-suited by the harmonized disposition code and leaves the opportunity for researchers to recode the original disposition descriptions as needed. Note that original descriptions have been truncated to maximum 10 characters due to excessive memory string variables take up.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0

Data notes

- All states: US001

adj_disp_off_cd

Label	CJARS standardized disposition offense - offense recorded at disposition
Description	CJARS standardized offense code for offense recorded at disposition. For a full description of the coding categories and method used to harmonize offense codes, please refer to Appendices C and E.2.1.
Table	Adjudication
Format	string
Code scheme	offense classification

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Data notes

- Florida: FL002
- Minnesota: MN002
- North Carolina: NC003

adj_disp_off_cd_src

Label	Raw disposition offense from source - offense recorded at disposition
Description	Original description of the offense recorded at disposition used by the agency where data was collected. For a harmonized disposition offense please see the following: adj_disp_off_cd, Disposition code descriptions are agency-specific and thus inconsistent across jurisdictions. Note that original offense descriptions have been truncated to maximum 30 characters due to excessive memory string variables take up.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0

Data notes

- North Carolina: NC003

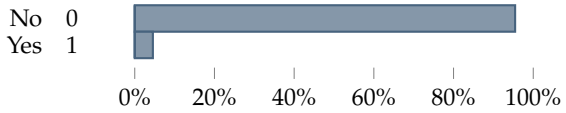
adj_dv_off

Label	Domestic violence offense flag
Description	Whether the type of offense was flagged as domestic violence. It should be noted that this variable is experimental due to a lack of sufficient training data and should be used with some caution. Raw offense descriptions provided in the variables adj_chrg_off_cd_src and/or adj_disp_off_cd_src can be inspected to evaluate appropriateness of offenses being flagged as domestic violence, especially for violent offenses (1st digit of adj_chrg_off_cd and/or adj_disp_off_cd is 1) and family-related offenses (adj_chrg_off_cd and/or adj_disp_off_cd is 5100, 5101, or 5102). This variable is equal to 1 if either the variables adj_chrg_off_cd_src and/or adj_disp_off_cd_src were identified as involving domestic violence.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,027,477	100.0
Missing values	7,428	0.0

adj_dv_off (continued)

Histogram of adj_dv_off (valid records)

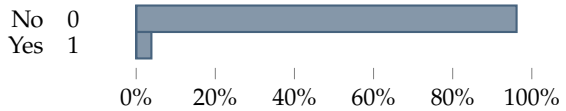


adj_wpn_off

Label	Weapon offense flag
Description	Whether the type of offense was flagged as involving a weapon. It should be noted that this variable is experimental due to a lack of sufficient training data and should be used with some caution. Raw offense descriptions provided in the variables adj_chrg_off_cd_src and/or adj_disp_off_cd_src can be inspected to evaluate appropriateness of offenses being flagged as involving a weapon.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,027,477	100.0
Missing values	7,428	0.0

Histogram of adj_wpn_off (valid records)



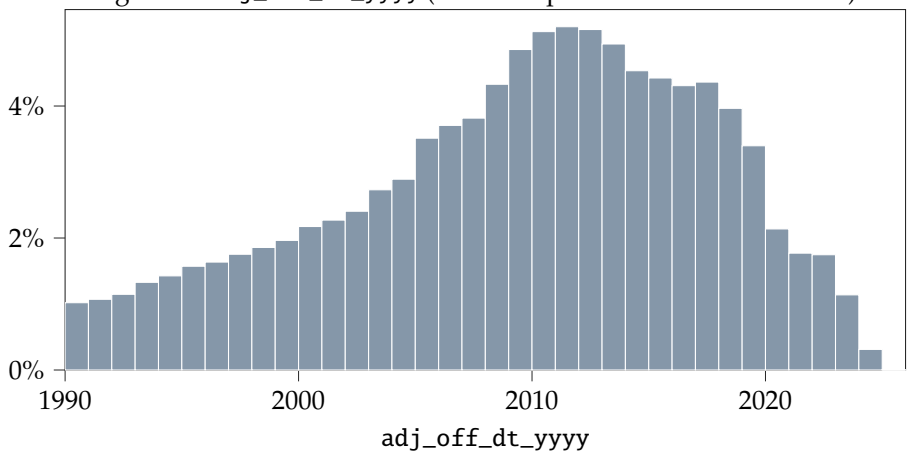
adj_off_dt_yyyy

Label	Date offense was committed, year
Description	The year when the individual committed the offense.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	106,683,782	65.8
Invalid values	6,535	0.0
Missing values	55,344,588	34.2

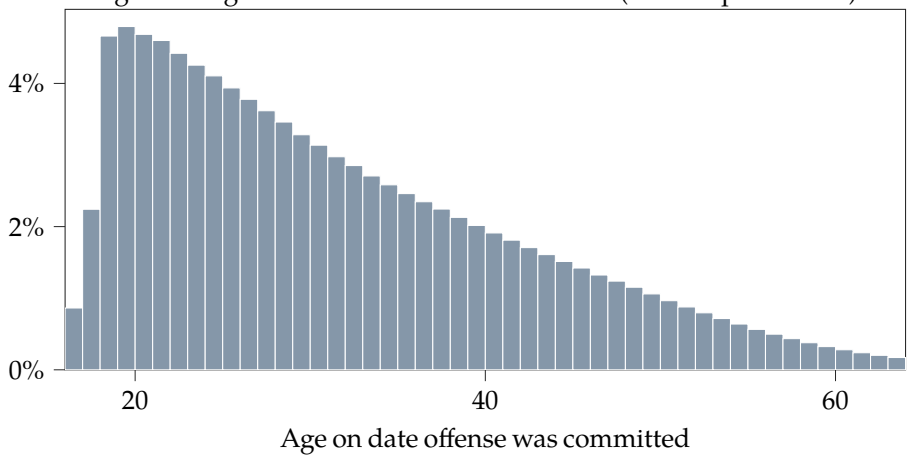
Statistic	*Value*
Mean	2,007.7
Minimum	1,900
25th percentile	2,002
Median	2,010
75th percentile	2,015
Maximum	2,025

Histogram of adj_off_dt_yyyy (above 5th percentile of valid records)



adj_off_dt_yyyy (continued)

Histogram of age on date offense was committed (1st–99th percentiles)

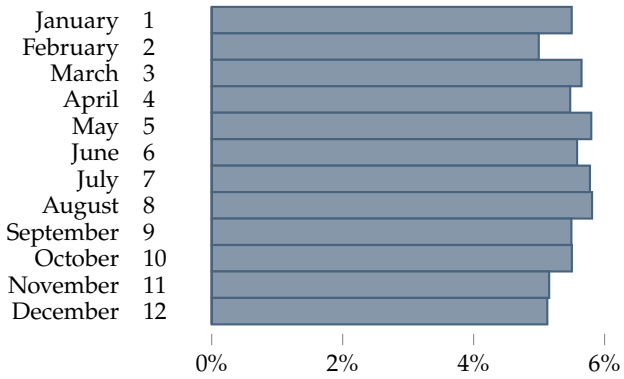


adj_off_dt_mm

Label	Date offense was committed, month
Description	The month when the individual committed the offense.
Table	Adjudication
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	106,690,317	65.8
Missing values	55,344,588	34.2

Histogram of `adj_off_dt_mm` (valid records)

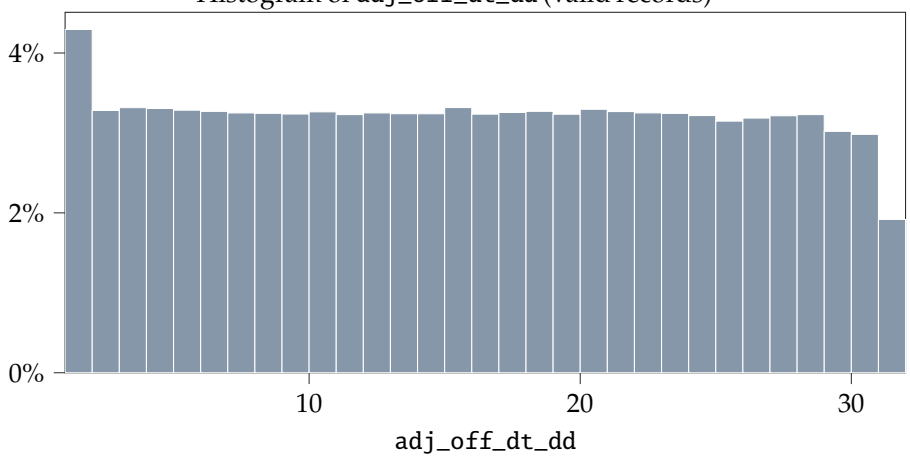


adj_off_dt_dd

Label	Date offense was committed, day of month
Description	The day of the month when the individual committed the offense.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	106,690,317	65.8
Missing values	55,344,588	34.2

Histogram of `adj_off_dt_dd` (valid records)



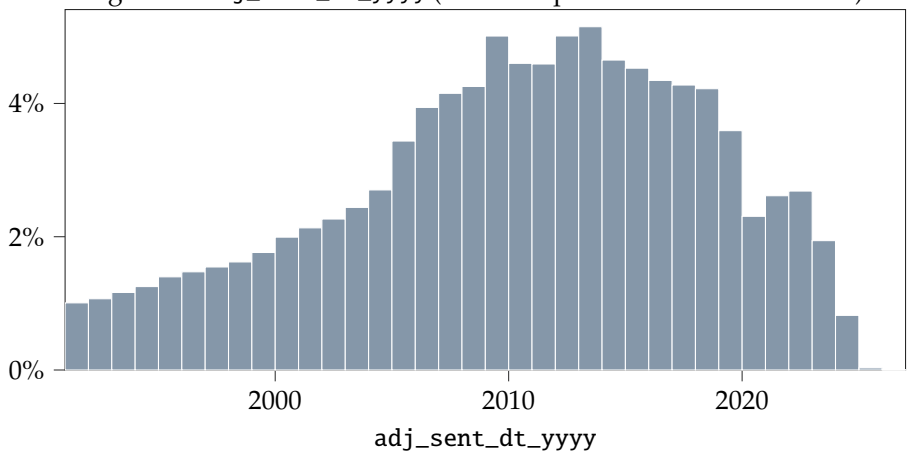
adj_sent_dt_yyyy

Label	Sentencing date, year
Description	The year when the individual was sentenced.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	58,688,114	36.2
Invalid values	584	0.0
Missing values	103,346,207	63.8

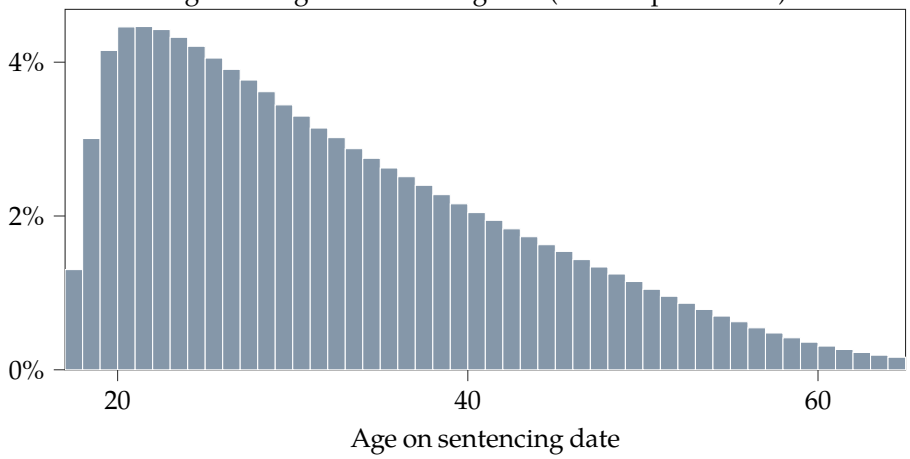
Statistic	*Value*
Mean	2,008.6
Minimum	1,900
25th percentile	2,004
Median	2,010
75th percentile	2,016
Maximum	2,026

Histogram of `adj_sent_dt_yyyy` (above 5th percentile of valid records)

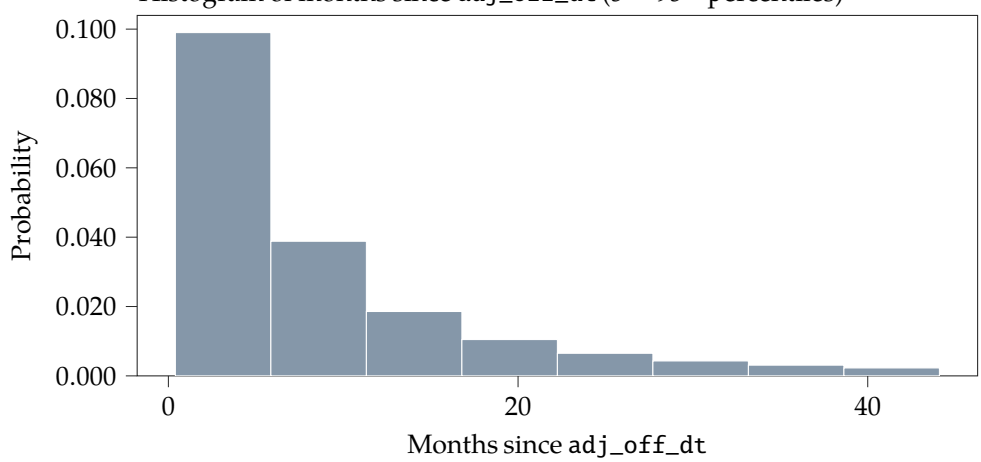


`adj_sent_dt_yyyy` (continued)

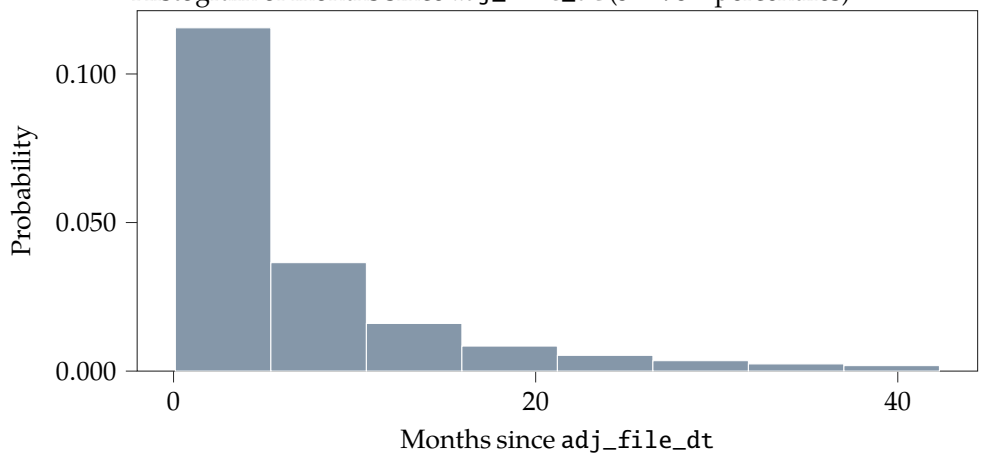
Histogram of age on sentencing date (1st–99th percentiles)



Histogram of months since `adj_off_dt` (5th–95th percentiles)

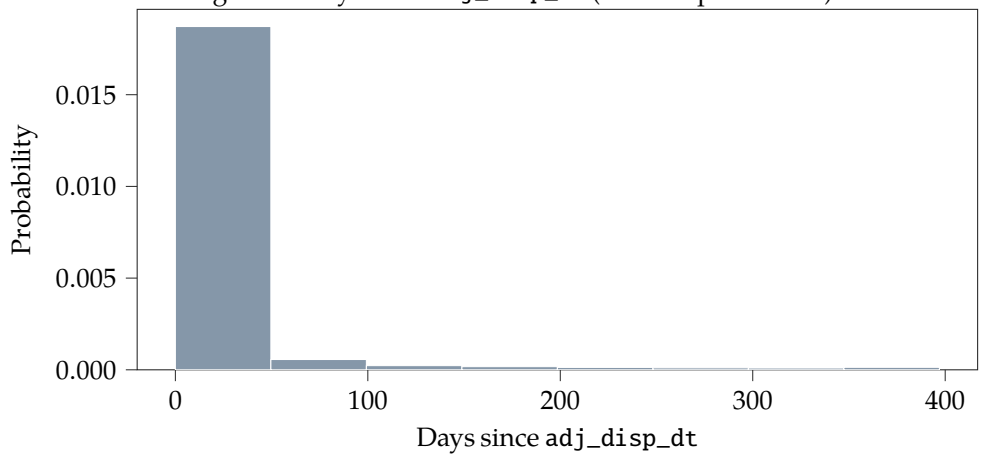


Histogram of months since `adj_file_dt` (5th–95th percentiles)



adj_sent_dt_yyyy (continued)

Histogram of days since adj_disp_dt (5th–95th percentiles)

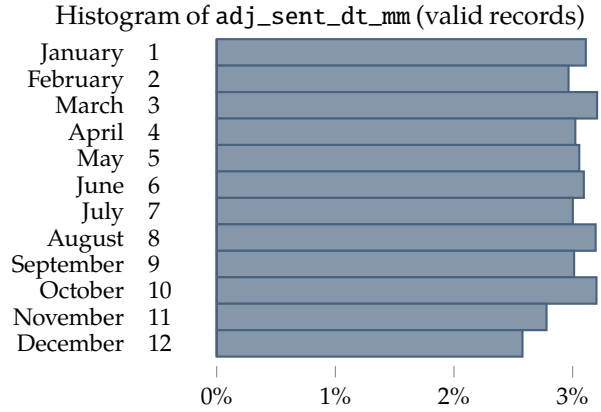


adj_sent_dt_mm

Label	Sentencing date, month
Description	The month when the individual was sentenced.
Table	Adjudication
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	58,688,698	36.2
Missing values	103,346,207	63.8

Histogram of adj_sent_dt_mm (valid records)

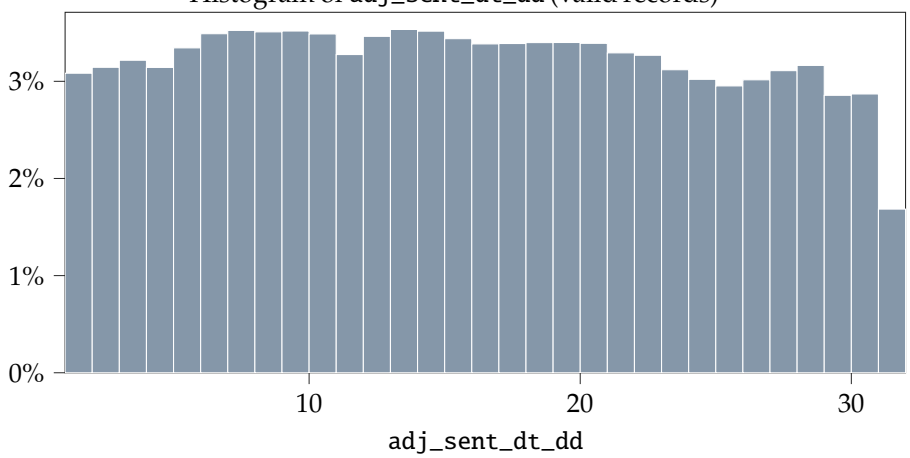


adj_sent_dt_dd

Label	Sentencing date, day of month
Description	The day of the month when the individual was sentenced.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	58,688,698	36.2
Missing values	103,346,207	63.8

Histogram of `adj_sent_dt_dd` (valid records)

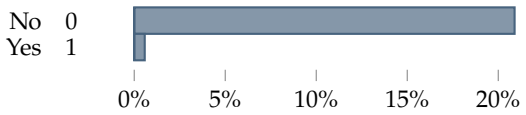


adj_sent_serv

Label	Community service sentence
Description	Whether the individual's sentence involved community service.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	34,799,252	21.5
Missing values	127,235,653	78.5

Histogram of `adj_sent_serv` (valid records)



Data notes

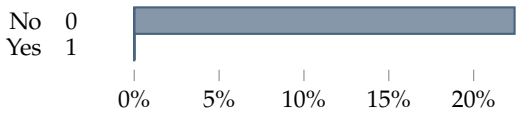
- Florida: FL002
- Pennsylvania: PA001

adj_sent_dth

Label	Death sentence
Description	Whether the individual received a death sentence.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	36,311,741	22.4
Missing values	125,723,164	77.6

Histogram of `adj_sent_dth` (valid records)



Data notes

- Florida: FL002
- Pennsylvania: PA001

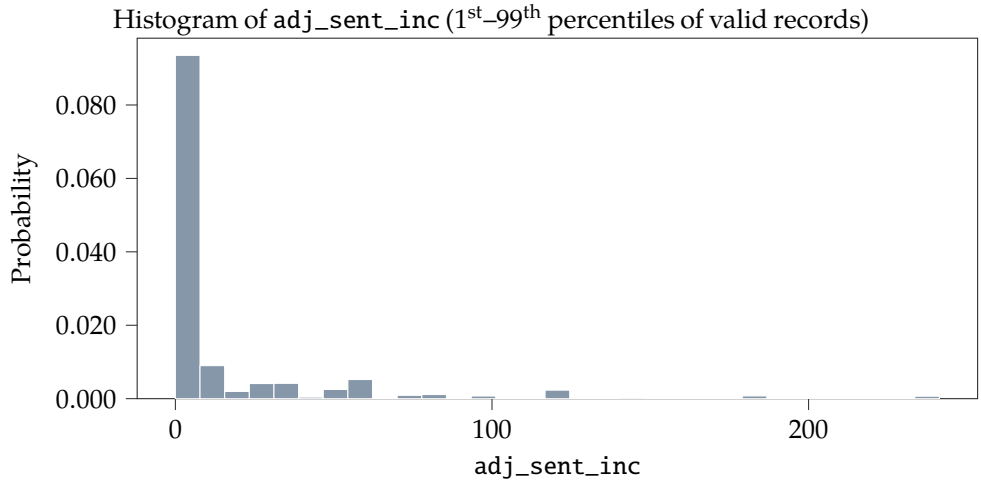
adj_sent_inc

Label	Incarceration length in months
Description	The number of months the individual was sentenced to serve in prison.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	46,980,267	29.0
Invalid values	5,169	0.0
Missing values	115,049,469	71.0
Special values		
Life sentence (-88888)	62,952	0.0
Death sentence (-99999)	14,189	0.0

Statistic	*Value*
Mean	17.3
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	12.1
Maximum	6,000.0

Histogram of `adj_sent_inc` (1st–99th percentiles of valid records)



Data notes

- All states: `US002`
- Florida: `FL002`
- Pennsylvania: `PA001`
- Virginia: `VA001`

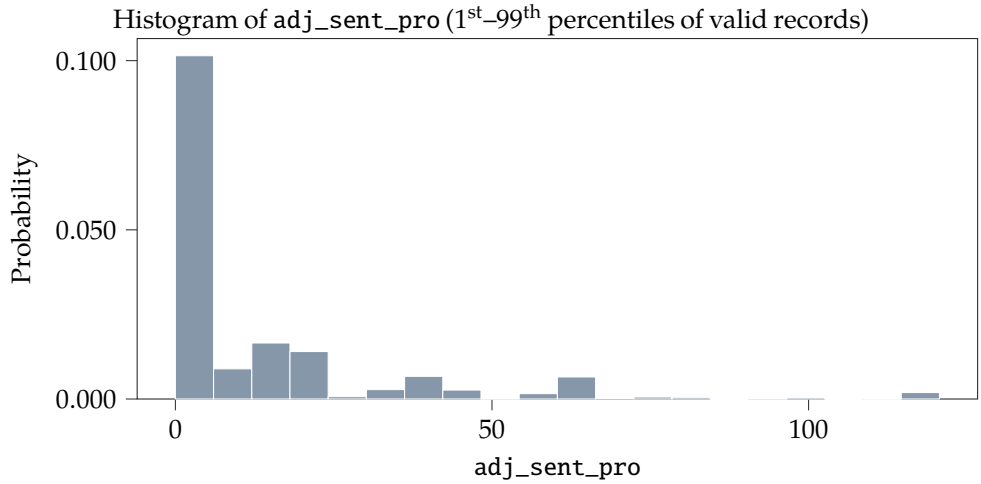
adj_sent_pro

Label	Probation length in months
Description	The number of months the individual was sentenced to serve on probation.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	41,276,522	25.5
Invalid values	2,765	0.0
Missing values	120,755,618	74.5

Statistic	*Value*
Mean	13.7
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	18.0
Maximum	6,000.0

Histogram of `adj_sent_pro` (1st–99th percentiles of valid records)



Data notes

- All states: US002
- Florida: FL002, FL004
- Pennsylvania: PA001
- Virginia: VA001

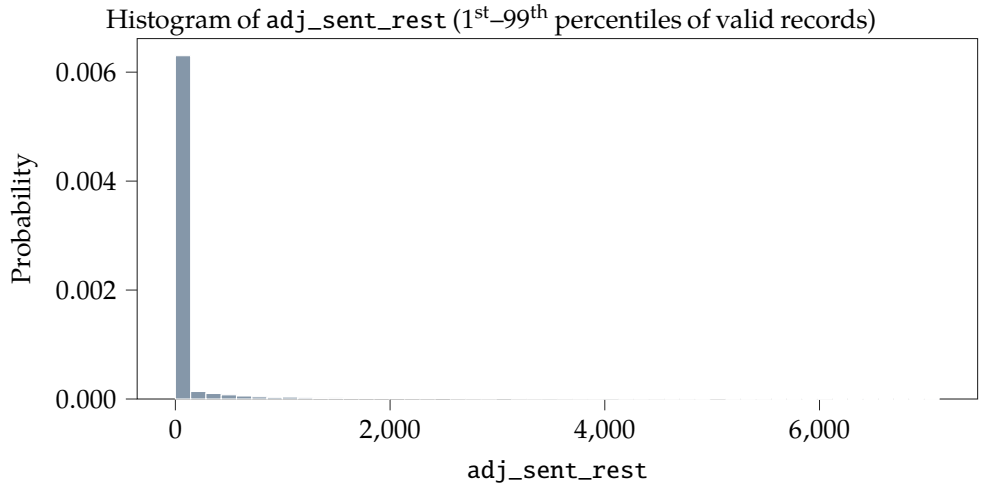
adj_sent_rest

Label	Restitution amount in dollars
Description	The amount in dollars of restitution the individual had to pay.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	18,980,804	11.7
Invalid values	3,320	0.0
Missing values	143,050,781	88.3

Statistic	*Value*
Mean	474.1
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	0.0
Maximum	500,000.0

Histogram of `adj_sent_rest` (1st–99th percentiles of valid records)



Data notes

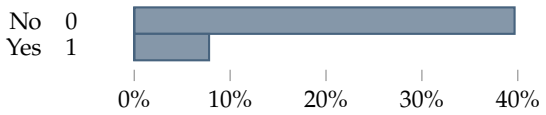
- All states: US002
- Florida: FL002, FL004
- Pennsylvania: PA001

adj_sent_sus

Label	Suspended sentence
Description	Whether the individual received a suspended sentence.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	76,915,838	47.5
Missing values	85,119,067	52.5

Histogram of `adj_sent_sus` (valid records)



Data notes

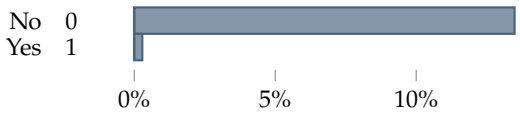
- Florida: `FL002`, `FL004`
- Pennsylvania: `PA001`

adj_sent_trt

Label	Treatment sentence
Description	Whether the individual received a treatment-oriented sentence.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	22,323,462	13.8
Missing values	139,711,443	86.2

Histogram of `adj_sent_trt` (valid records)



Data notes

- Florida: `FL002`, `FL004`
- Pennsylvania: `PA001`

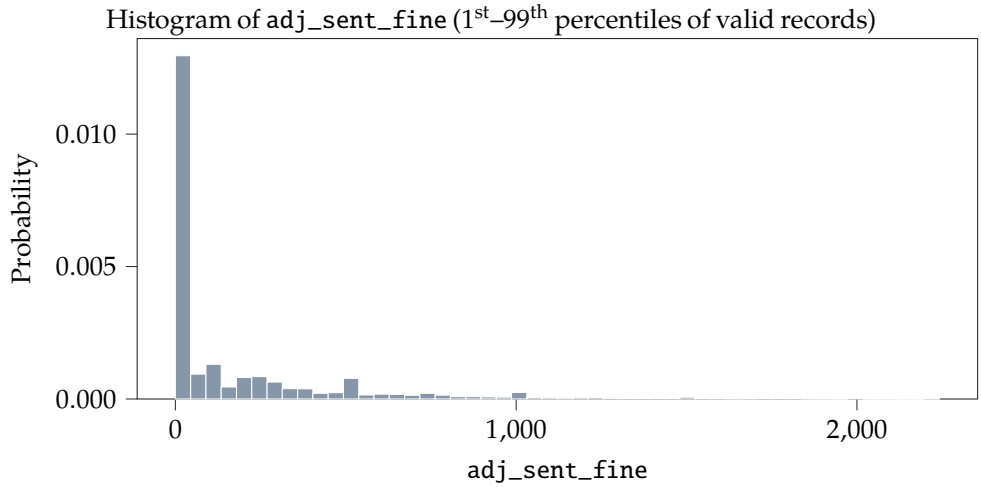
adj_sent_fine

Label	Fine amount in dollars
Description	The amount in dollars the individual was fined.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	42,718,071	26.4
Invalid values	306,550	0.2
Missing values	119,010,284	73.4

Statistic	*Value*
Mean	303.3
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	250.0
Maximum	500,000.0

Histogram of `adj_sent_fine` (1st–99th percentiles of valid records)



Data notes

- All states: `US002`
- Florida: `FL003`, `FL004`
- Pennsylvania: `PA001`

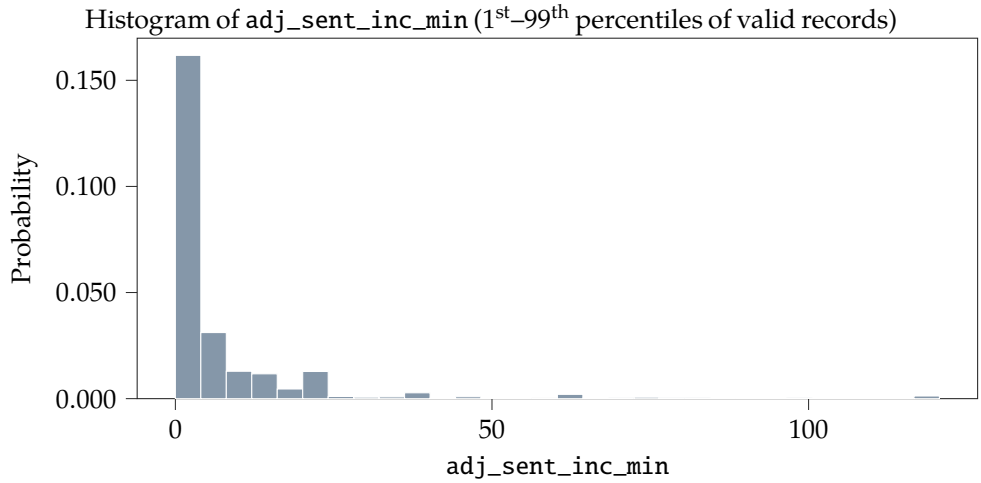
adj_sent_inc_min

Label	Minimum incarceration term in months
Description	The minimum number of months the individual was sentenced to serve in prison.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	9,988,949	6.2
Invalid values	2,359	0.0
Missing values	152,043,597	93.8
Special values		
Life sentence (-88888)	3,651	0.0
Death sentence (-99999)	818	0.0

Statistic	*Value*
Mean	8.7
Minimum	0.0
25th percentile	0.0
Median	1.5
75th percentile	8.0
Maximum	2,895.9

Histogram of adj_sent_inc_min (1st–99th percentiles of valid records)



Data notes

- All states: US002
- Florida: FL004
- Pennsylvania: PA001

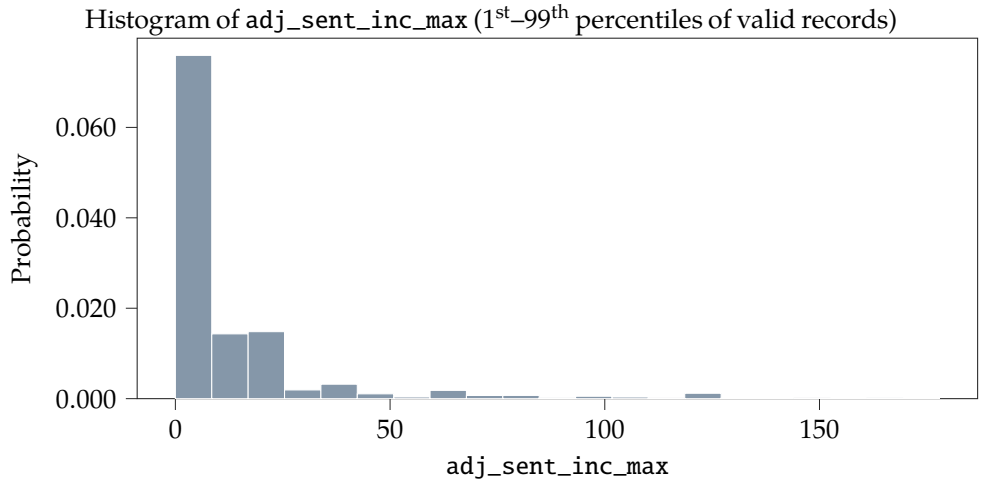
adj_sent_inc_max

Label	Maximum incarceration term in months
Description	The maximum number of months the individual was sentenced to serve in prison.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	9,041,647	5.6
Invalid values	6,184	0.0
Missing values	152,987,074	94.4
Special values		
Life sentence (-88888)	16,433	0.0
Death sentence (-99999)	852	0.0

Statistic	*Value*
Mean	15.1
Minimum	0.0
25th percentile	0.0
Median	3.2
75th percentile	15.1
Maximum	3,620.8

Histogram of `adj_sent_inc_max` (1st–99th percentiles of valid records)



Data notes

- All states: US002
- Pennsylvania: PA001

adj_sent_src

Label	Raw sentence from source
Description	Original description of the sentence given by the agency where data was collected. Sentencing is agency-specific and thus inconsistent across jurisdictions. For a harmonized sentence please see the following: adj_sent_serv, adj_sent_dth, adj_sent_inc, adj_sent_pro, adj_sent_rest, adj_sent_sus, adj_sent_trt, adj_sent_fine, adj_sent_inc_min, adj_sent_inc_max. Even though harmonized sentencing information is provided, this raw description is retained for research that is not well-suited by the harmonized sentence and leaves the opportunity for researchers to recode the original descriptions as needed. Note that original descriptions have been truncated to maximum 30 characters due to excessive memory string variables take up.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0

Data notes

- All states: US001
- Florida: FL004
- Georgia: GA002, GA003, GA004

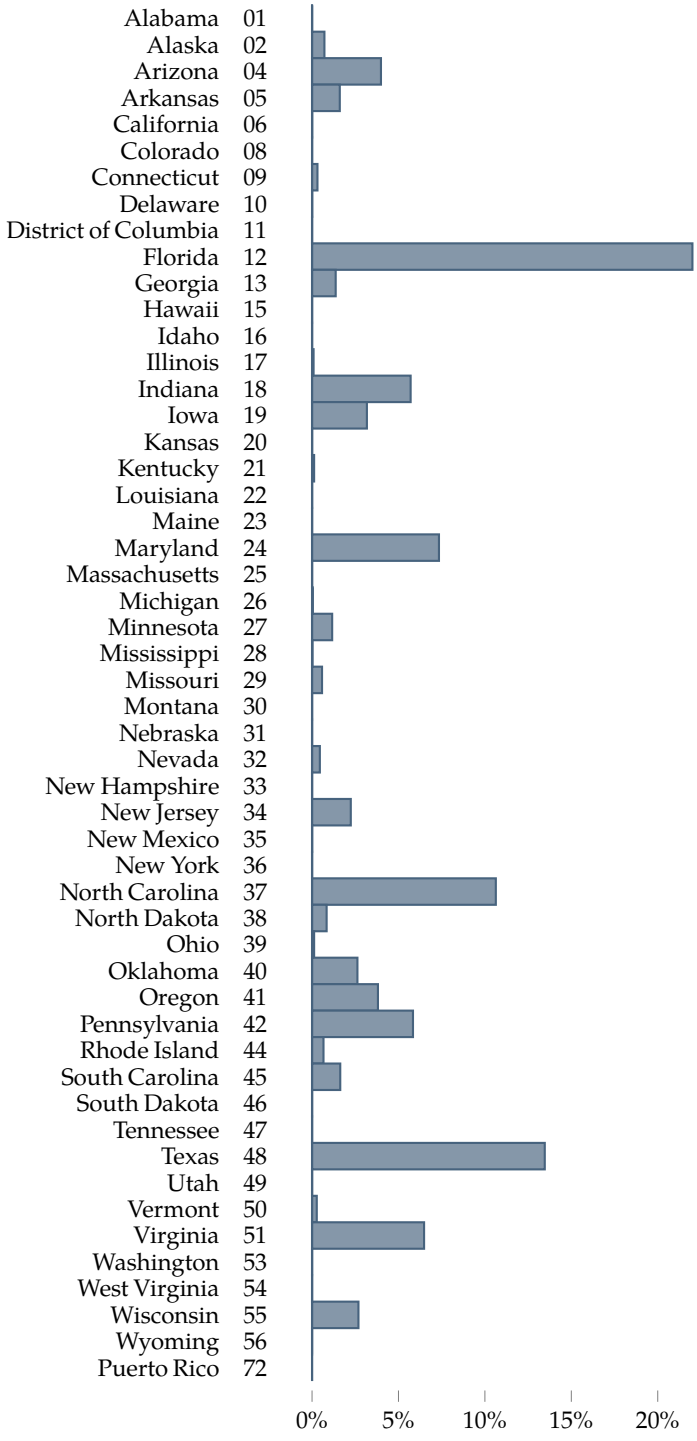
adj_st_juris_fips

Label	State FIPS code for location of filing
Description	State-level Federal Information Processing Standards (FIPS) code where case was filed. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Adjudication
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

`adj_st_juris_fips` (continued)

Histogram of `adj_st_juris_fips` (valid records)



adj_cnty_juris_fips

Label	County FIPS code for location of filing
Description	County-level Federal Information Processing Standards (FIPS) code where case was filed. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Adjudication
Format	string
Code scheme	county FIPS code

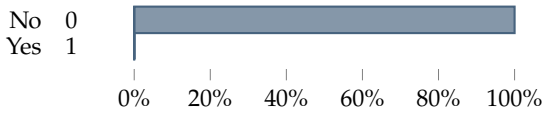
Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	160,081,558	98.8
Invalid values	32,304	0.0
Missing values	1,921,043	1.2

adj_rec_src_bkg

Label	Record source - booking agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a booking agency.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_rec_src_bkg` (valid records)

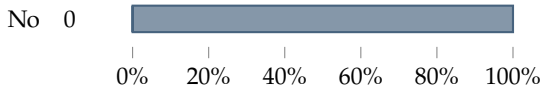


adj_rec_src_arr

Label	Record source - arresting agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from an arresting agency.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_rec_src_arr` (valid records)

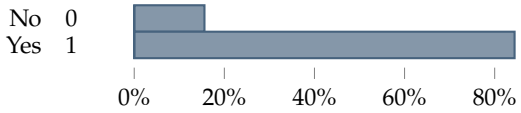


adj_rec_src_crt

Label	Record source - courts
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a court system.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_rec_src_crt` (valid records)

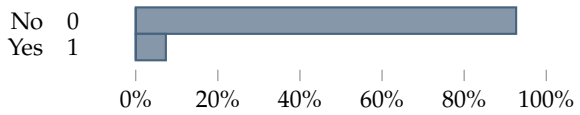


adj_rec_src_doc

Label	Record source - department of corrections
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a department of corrections.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_rec_src_doc` (valid records)

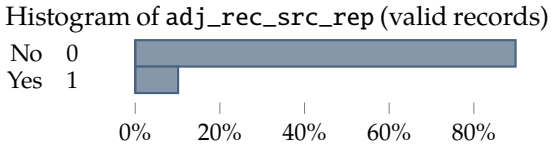


adj_rec_src_rep

Label	Record source - criminal history repository
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a criminal history repository.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_rec_src_rep` (valid records)

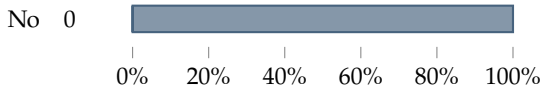


adj_rec_src_cc

Label	Record source - community corrections agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a community corrections agency.
Table	Adjudication
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	162,034,905	100.0
Valid records	162,034,905	100.0

Histogram of `adj_rec_src_cc` (valid records)



4.2.5 Probation

The probation table contains information on probation conditions, probation begin status and date, and probation end status and date. Data is at the level of a probation term.

Table-level data notes

- All states: `US003`
- Idaho: `ID001`
- Michigan: `MI001`
- Montana: `MT001`, `MT002`
- Washington: `WA001`
- Wisconsin: `WI003`
- Wyoming: `WY001`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.3.1 and Figure 6.
Table	Probation
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Unique values	7,808,097	

pro_id

Label	Probation term identifier
Description	Uniquely identifies terms of probation. For more details on use of pro_id for data linkage, refer to Section 3.3.2 and Figure 7.
Table	Probation
Table key	primary, unique
Format	string

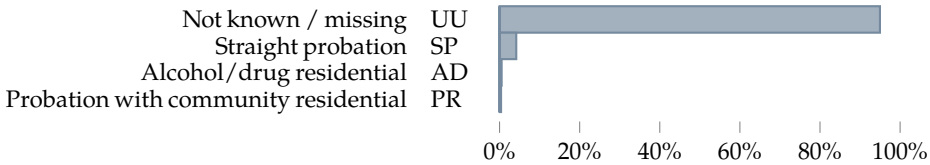
Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Unique values	10,645,065	

pro_cond_cd

Label	CJARS standardized probation conditions
Description	CJARS standardized probation conditions classification. Used when consistent coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize probation conditions descriptions, please refer to Appendices C and E.2.5.
Table	Probation
Format	string
Code scheme	probation condition code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of pro_cond_cd (valid records)



pro_cond_cd_src

Label	Raw description of probation conditions from source
Description	A description of the conditions of probation for an individual. This variable provides the raw description that was provided from the agency. For harmonized probation conditions please see pro_cond_cd. Even though a harmonized description of probation conditions is provided, this raw description is retained for research that is not well-suited by the harmonized description of probation conditions and leaves the opportunity for researchers to recode the original descriptions as needed.
Table	Probation
Format	string

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0

Data notes

- All states: US001

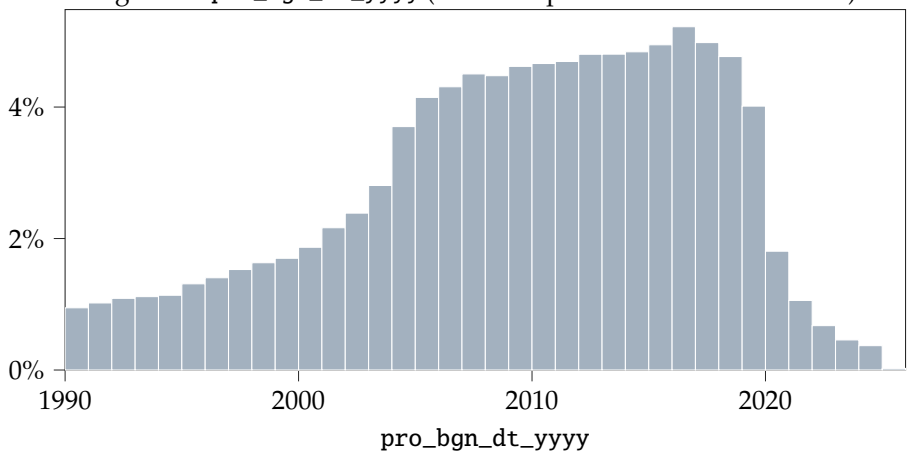
pro_bgn_dt_yyyy

Label	Probation start date, year
Description	The year when the individual began probation.
Table	Probation
Format	numeric

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	9,042,696	84.9
Invalid values	33	0.0
Missing values	1,602,336	15.1

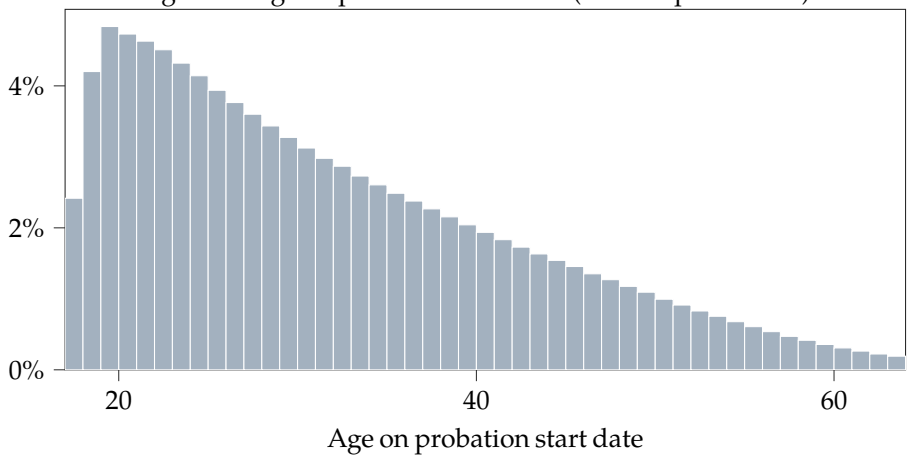
Statistic	*Value*
Mean	2,008.0
Minimum	1,900
25th percentile	2,003
Median	2,009
75th percentile	2,015
Maximum	2,025

Histogram of pro_bgn_dt_yyyy (above 5th percentile of valid records)



pro_bgn_dt_yyyy (continued)

Histogram of age on probation start date (1st–99th percentiles)

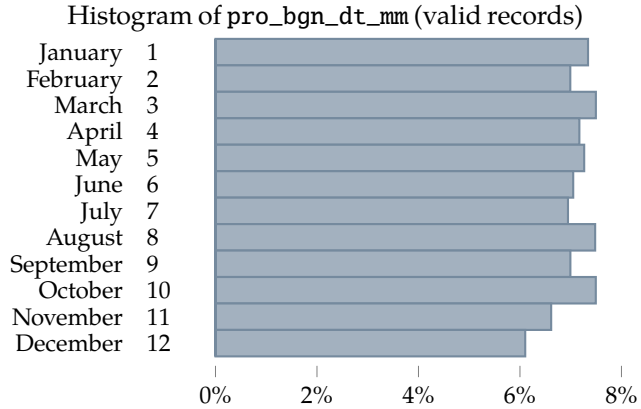


pro_bgn_dt_mm

Label	Probation start date, month
Description	The month when the individual began probation.
Table	Probation
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	9,042,729	84.9
Missing values	1,602,336	15.1

Histogram of `pro_bgn_dt_mm` (valid records)

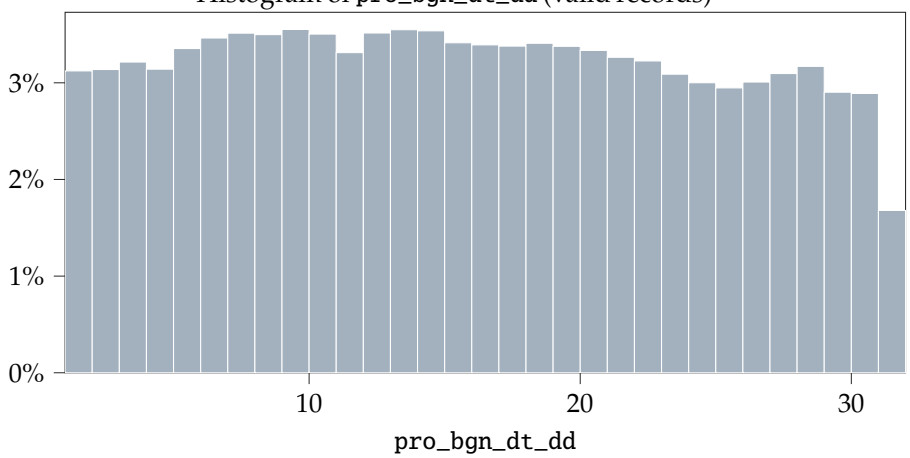


pro_bgn_dt_dd

Label	Probation start date, day of month
Description	The day of the month when the individual began probation.
Table	Probation
Format	numeric

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	9,042,729	84.9
Missing values	1,602,336	15.1

Histogram of pro_bgn_dt_dd (valid records)



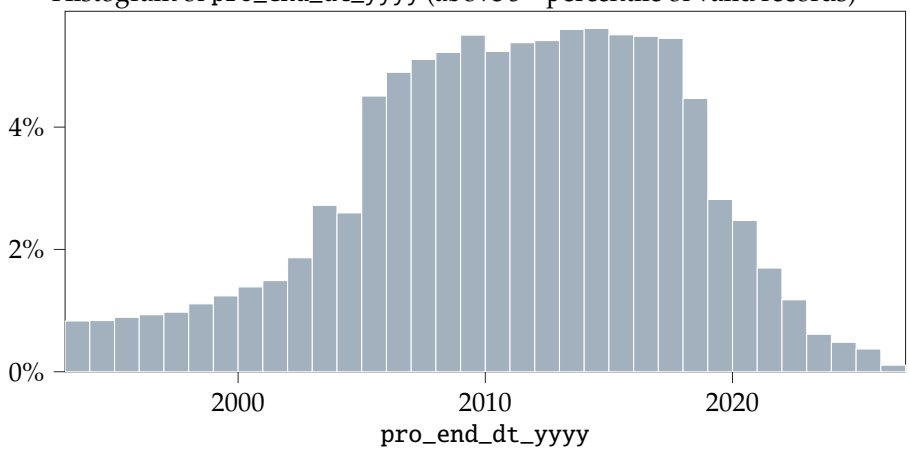
pro_end_dt_yyyy

Label	Probation end date, year
Description	The year when the individual's probation ended.
Table	Probation
Format	numeric

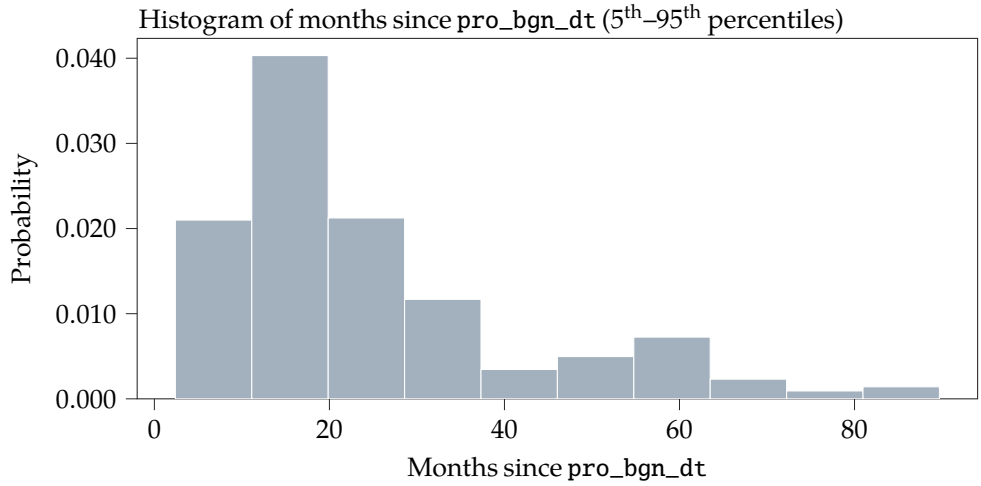
Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	5,143,658	48.3
Invalid values	26	0.0
Missing values	5,501,381	51.7

Statistic	*Value*
Mean	2,009.5
Minimum	1,900
25th percentile	2,006
Median	2,011
75th percentile	2,015
Maximum	2,026

Histogram of pro_end_dt_yyyy (above 5th percentile of valid records)



`pro_end_dt_yyyy` (continued)

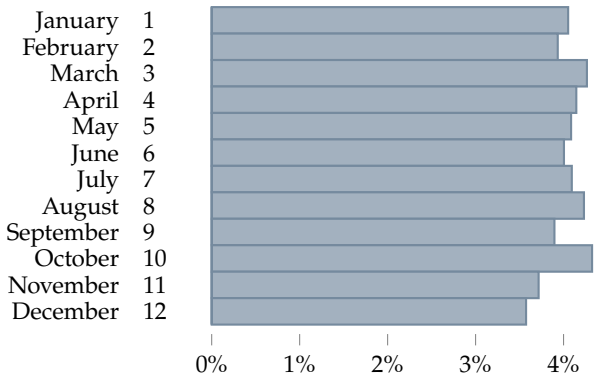


pro_end_dt_mm

Label	Probation end date, month
Description	The month when the individual's probation ended.
Table	Probation
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	5,143,684	48.3
Missing values	5,501,381	51.7

Histogram of `pro_end_dt_mm` (valid records)

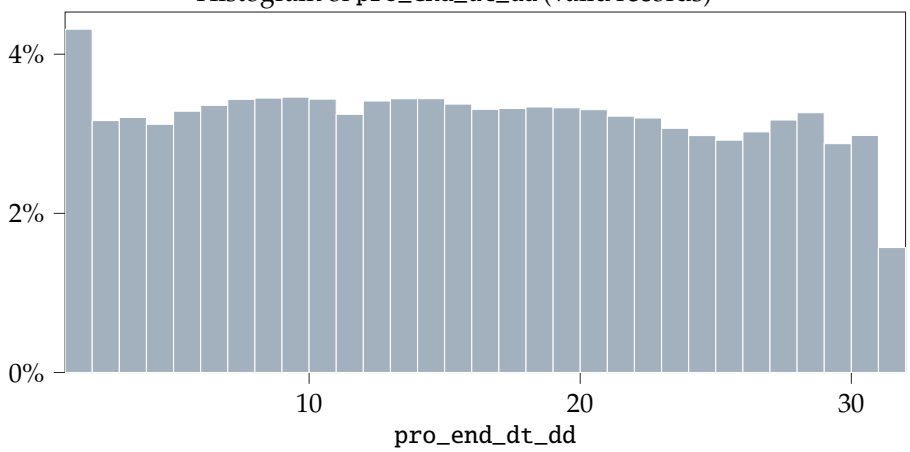


pro_end_dt_dd

Label	Probation end date, day of month
Description	The day of the month when the individual's probation ended.
Table	Probation
Format	numeric

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	5,143,684	48.3
Missing values	5,501,381	51.7

Histogram of pro_end_dt_dd (valid records)

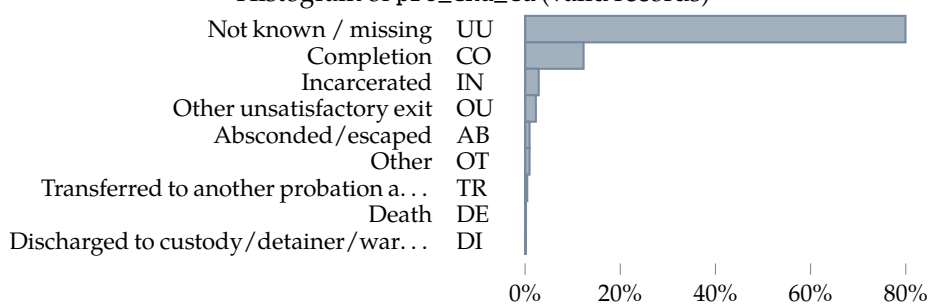


pro_end_cd

Label	CJARS standardized probation end status
Description	CJARS standardized probation end status classification. Used when consistent coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize probation end status, please refer to Appendices C and E.2.6.
Table	Probation
Format	string
Code scheme	probation exit code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_end_cd` (valid records)



Data notes

- Florida: `FL004`

pro_end_cd_src

Label	Raw description of probation end status from source
Description	A description of the probation end status for an individual. This variable provides the raw description that was provided from the agency. For harmonized probation end status please see pro_end_cd. Even though a harmonized description of probation end status is provided, this raw description is retained for research that is not well-suited by the harmonized description of probation end status and leaves the opportunity for researchers to recode the original descriptions as needed.
Table	Probation
Format	string

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0

Data notes

- All states: US001

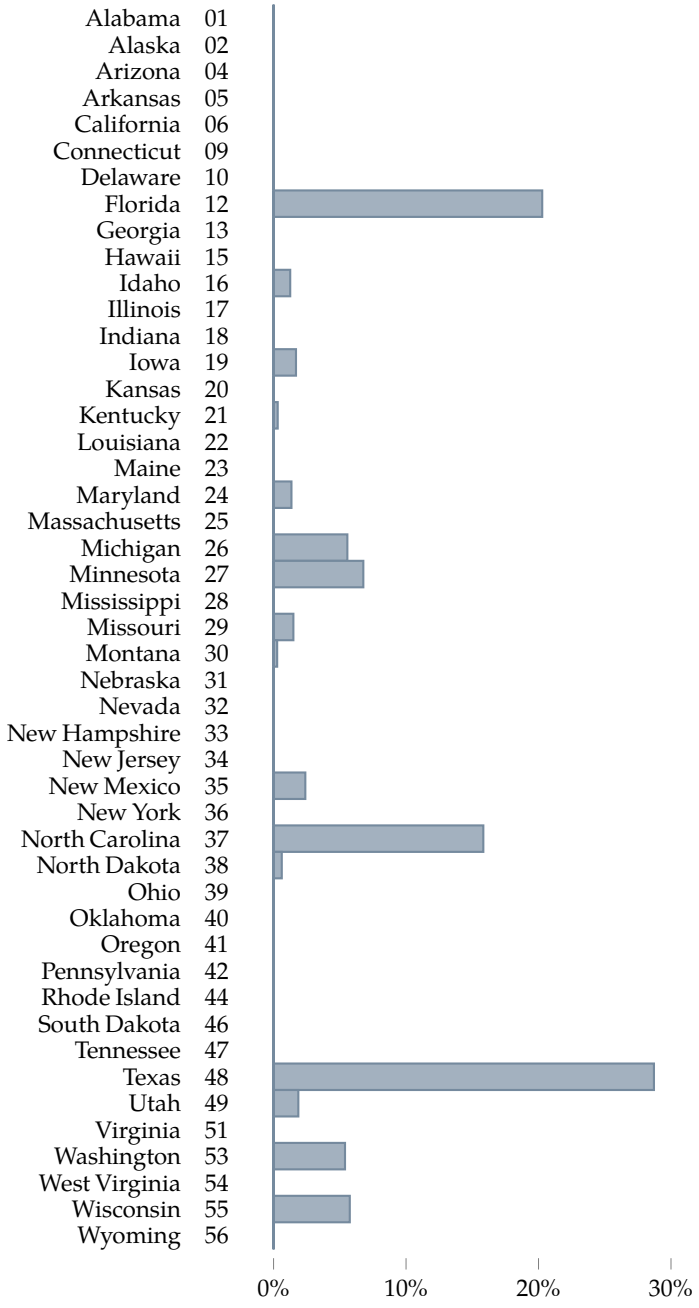
pro_st_ori_fips

Label	State FIPS code for location of sentencing
Description	State-level Federal Information Processing Standards (FIPS) code for location where the individual was sentenced. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Probation
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,601,629	99.6
Missing values	43,436	0.4

`pro_st_ori_fips` (continued)

Histogram of `pro_st_ori_fips` (valid records)



pro_cnty_ori_fips

Label	County FIPS code for location of sentencing
Description	County-level Federal Information Processing Standards (FIPS) code for location where the individual was sentenced. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Probation
Format	string
Code scheme	county FIPS code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	8,207,273	77.1
Invalid values	2,271	0.0
Missing values	2,394,356	22.5

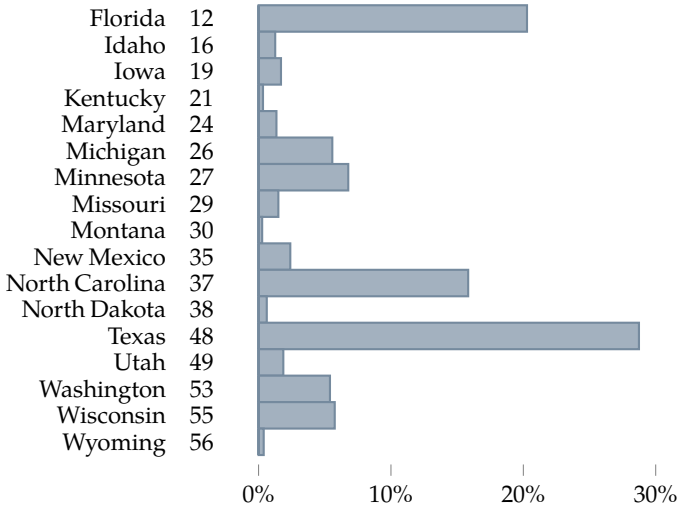
pro_st_juris_fips

Label	FIPS code of state with jurisdiction over supervision of individual
Description	State-level Federal Information Processing Standards (FIPS) code for the state with jurisdiction over the individual's term of probation. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Probation
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

`pro_st_juris_fips` (continued)

Histogram of `pro_st_juris_fips` (valid records)



`pro_cnty_juris_fips`

Label	FIPS code of county with jurisdiction over supervision of individual
Description	County-level Federal Information Processing Standards (FIPS) code for the county with jurisdiction over the individual's term of probation. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Probation
Format	string
Code scheme	county FIPS code

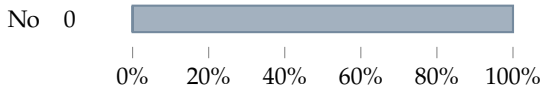
Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Missing values	10,645,065	100.0

pro_rec_src_bkg

Label	Record source - booking agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a booking agency.
Table	Probation
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_rec_src_bkg` (valid records)

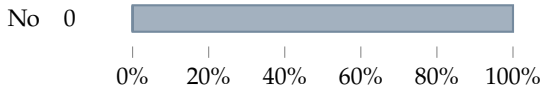


pro_rec_src_arr

Label	Record source - arresting agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from an arresting agency.
Table	Probation
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_rec_src_arr` (valid records)

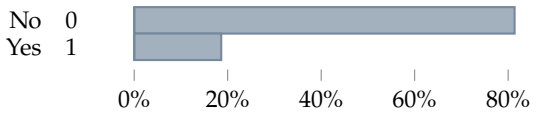


pro_rec_src_crt

Label	Record source - courts
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a court system.
Table	Probation
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_rec_src_crt` (valid records)

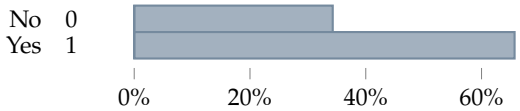


pro_rec_src_doc

Label	Record source - department of corrections
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a department of corrections.
Table	Probation
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_rec_src_doc` (valid records)

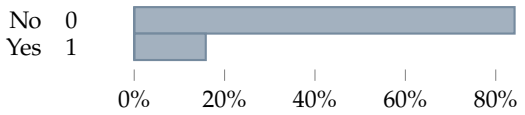


pro_rec_src_rep

Label	Record source - criminal history repository
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a criminal history repository.
Table	Probation
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_rec_src_rep` (valid records)

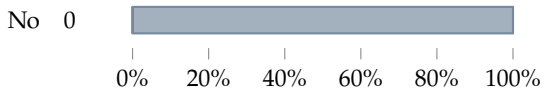


pro_rec_src_cc

Label	Record source - community corrections agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a community corrections agency.
Table	Probation
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	10,645,065	100.0
Valid records	10,645,065	100.0

Histogram of `pro_rec_src_cc` (valid records)



4.2.6 Incarceration

The incarceration table contains information about the facility where an individual is/was housed, entry and exit dates, as well as the current status of the person. Data is at the level of an incarceration term.

Table-level data notes

- All states: `US003`
- Arizona: `AZ001`
- Arkansas: `AR002`
- Connecticut: `CT001`
- Illinois: `IL002`
- Montana: `MT002`
- Rhode Island: `RI001`
- Vermont: `VT001`, `VT002`
- Wyoming: `WY001`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.3.1 and Figure 6.
Table	Incarceration
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Unique values	6,879,376	

inc_id

Label	Incarceration term identifier
Description	Uniquely identifies term of incarceration. For more details on use of `inc_id` for data linkage, refer to Section 3.3.2 and Figure 7.
Table	Incarceration
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Unique values	14,067,898	

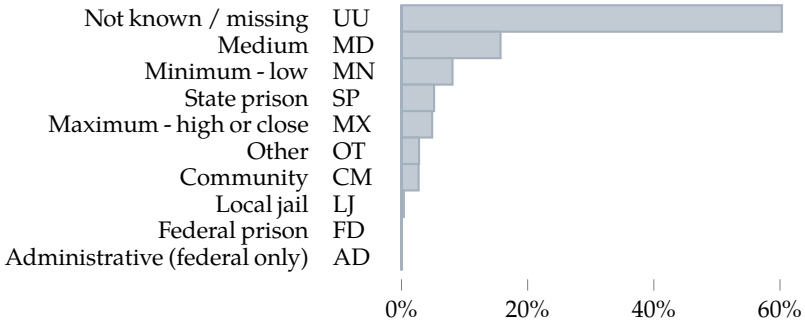
inc_fcl_cd

Label	CJARS standardized facility type
Description	CJARS standardized facility type where individual is/was housed. Used when consistent offense coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize offense facility type, please refer to Appendices C and E.2.8.
Table	Incarceration
Format	string
Code scheme	incarceration facility type

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

inc_fcl_cd (continued)

Histogram of inc_fcl_cd (valid records)



inc_fcl_cd_src

Label	Raw description of facility from source
Description	The agency's description of the facility that the individual is/was housed. For harmonized facility type please see inc_fcl_cd. Even though a harmonized facility type is provided, this raw description is retained for research that is not well-suited by the harmonized facility types and leaves the opportunity for researchers to recode the original descriptions as needed.
Table	Incarceration
Format	string

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0

Data notes

- All states: US001

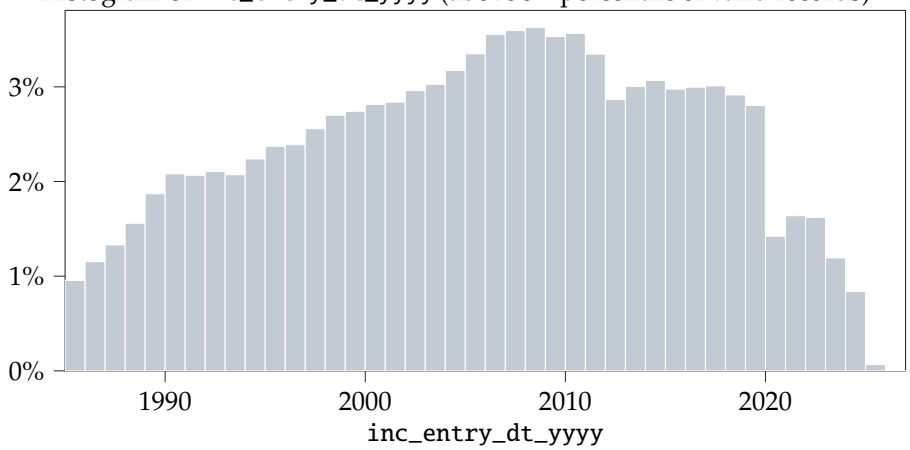
inc_entry_dt_yyyy

Label	Incarceration entry date, year
Description	The year when the individual entered incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	13,803,202	98.1
Missing values	264,696	1.9

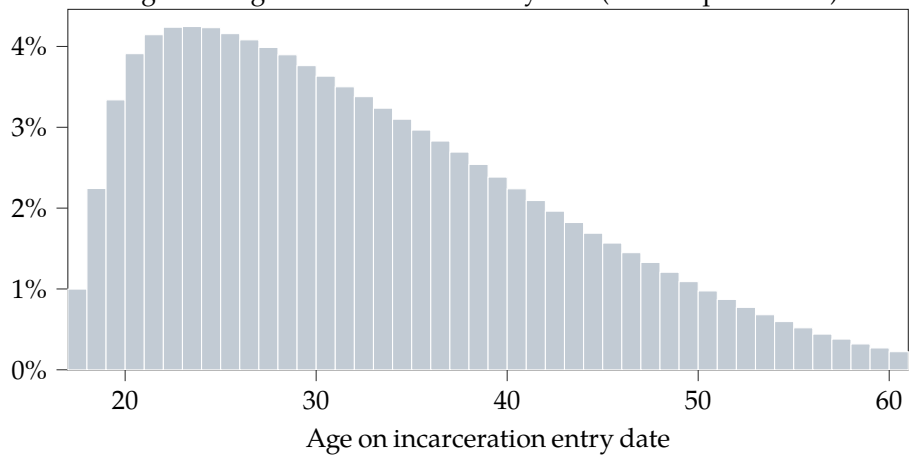
Statistic	*Value*
Mean	2,004.1
Minimum	1,900
25th percentile	1,996
Median	2,005
75th percentile	2,013
Maximum	2,026

Histogram of `inc_entry_dt_yyyy` (above 5th percentile of valid records)



`inc_entry_dt_yyyy` (continued)

Histogram of age on incarceration entry date (1st–99th percentiles)

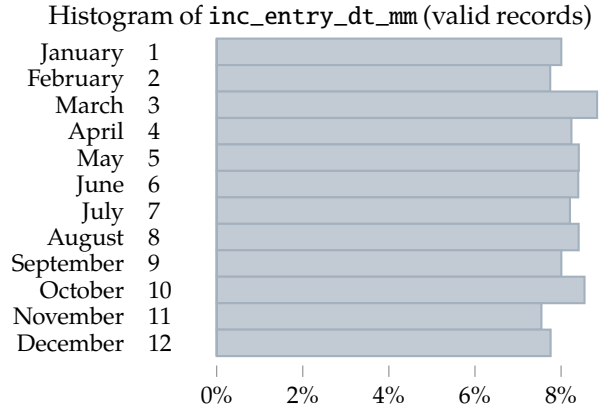


`inc_entry_dt_mm`

Label	Incarceration entry date, month
Description	The month when the individual entered incarceration.
Table	Incarceration
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	13,801,797	98.1
Missing values	266,101	1.9

Histogram of `inc_entry_dt_mm` (valid records)

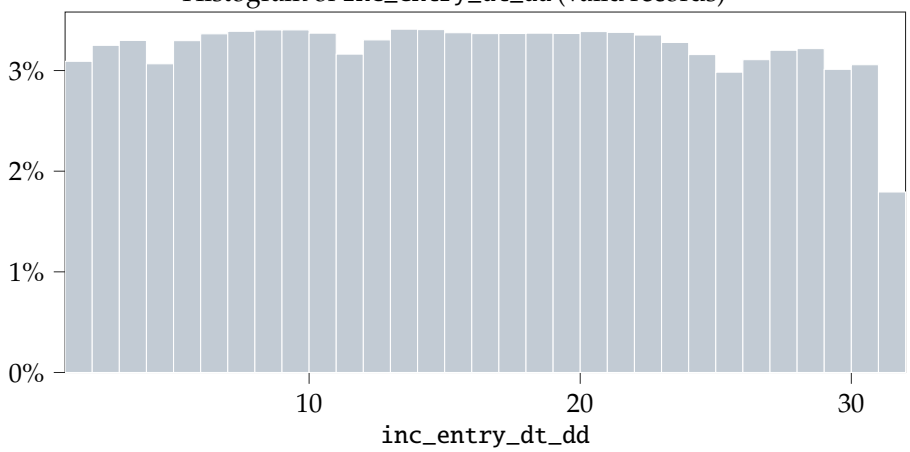


inc_entry_dt_dd

Label	Incarceration entry date, day of month
Description	The day of the month when the individual entered incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	13,801,797	98.1
Missing values	266,101	1.9

Histogram of `inc_entry_dt_dd` (valid records)

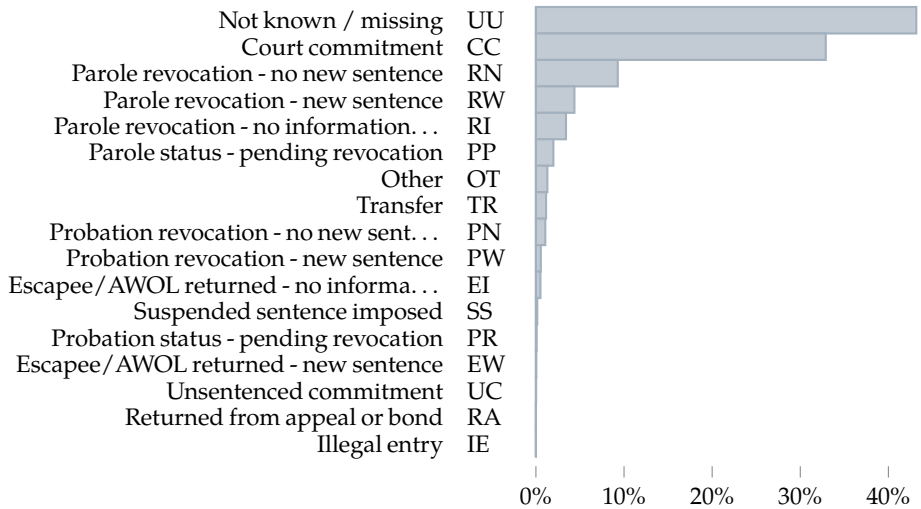


`inc_entry_cd`

Label	CJARS standardized entry status
Description	CJARS standardized entry status into incarceration. Used when consistent coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize entry status, please refer to Appendices C and E.2.7.
Table	Incarceration
Format	string
Code scheme	incarceration entry code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_entry_cd` (valid records)



Data notes

- Florida: `FL004`

inc_entry_cd_src

Label	Raw description of entry type into incarceration
Description	The agency's description of the entry status into incarceration. For harmonized entry status please see inc_entry_cd. Even though a harmonized entry status is provided, this raw description is retained for research that is not well-suited by the harmonized entry status and leaves the opportunity for researchers to recode the original descriptions as needed.
Table	Incarceration
Format	string

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0

Data notes

- All states: US001

inc_exit_dt_yyyy

Label	Incarceration exit date, year
Description	Year when the individual exited incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	12,211,775	86.8
Invalid values	1	0.0
Missing values	1,856,122	13.2

Statistic	*Value*
Mean	2,004.5
Minimum	1,900
25th percentile	1,997
Median	2,006
75th percentile	2,013
Maximum	2,026

`inc_exit_dt_yyyy` (continued)

Histogram of `inc_exit_dt_yyyy` (above 5th percentile of valid records)

Histogram of age on incarceration exit date (1st–99th percentiles)

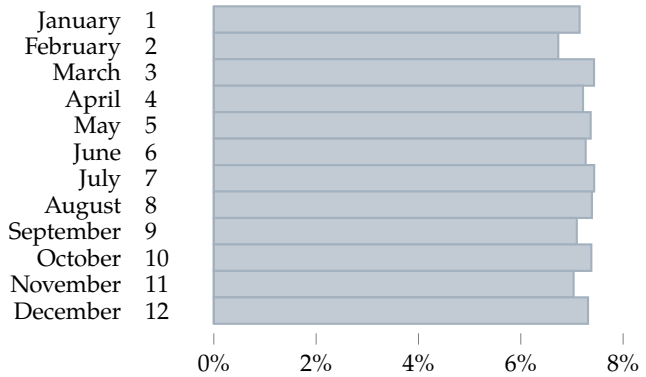
Histogram of months since `inc_entry_dt` (5th–95th percentiles)

`inc_exit_dt_mm`

Label	Incarceration exit date, month
Description	Month when the individual exited incarceration.
Table	Incarceration
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	12,211,776	86.8
Missing values	1,856,122	13.2

Histogram of `inc_exit_dt_mm` (valid records)

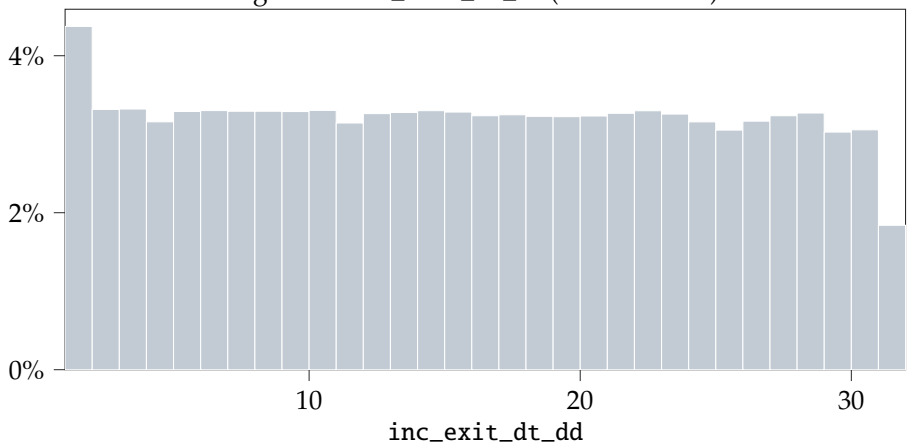


inc_exit_dt_dd

Label	Incarceration exit date, day of month
Description	Day of the month when the individual exited incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	12,211,776	86.8
Missing values	1,856,122	13.2

Histogram of `inc_exit_dt_dd` (valid records)



Data notes

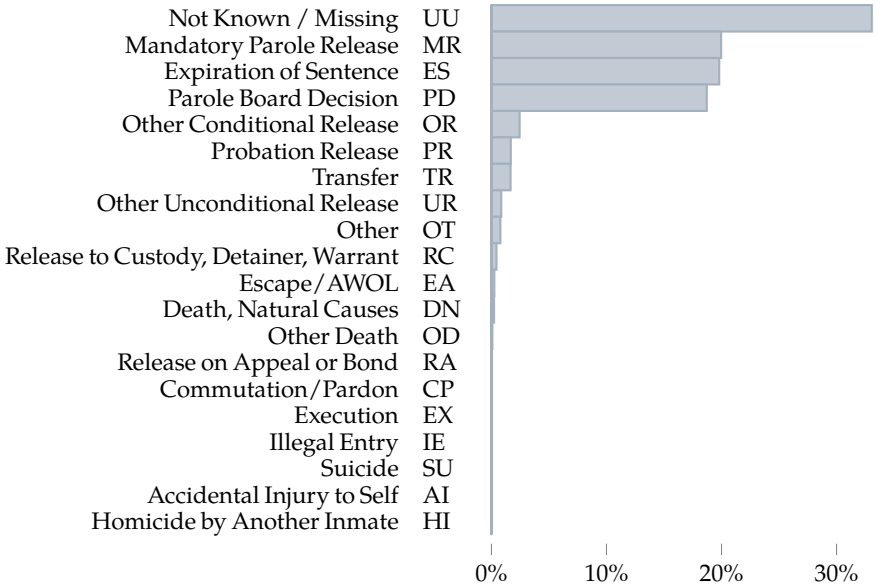
- Florida: FL001

`inc_exit_cd`

Label	CJARS standardized exit status
Description	CJARS standardized exit status from incarceration. Used when consistent offense coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize exit status, please refer to Appendices C and E.2.9.
Table	Incarceration
Format	string
Code scheme	incarceration exit code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_exit_cd` (valid records)



inc_exit_cd_src

Label	Raw description of exit type from incarceration
Description	The agency's description of the exit status from incarceration. For harmonized exit status please see inc_exit_cd. Even though a harmonized exit status is provided, this raw description is retained for research that is not well-suited by the harmonized exit status and leaves the opportunity for researchers to recode the original descriptions as needed.
Table	Incarceration
Format	string

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0

Data notes

- All states: US001

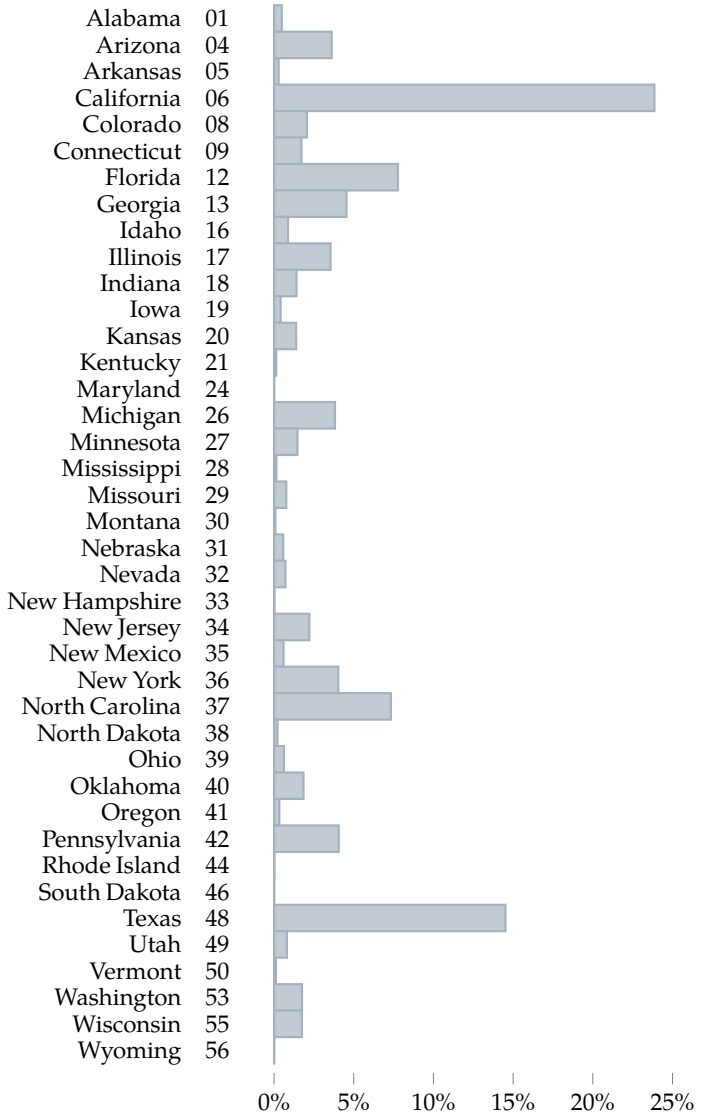
inc_st_ori_fips

Label	State FIPS code for location of sentencing
Description	State-level Federal Information Processing Standards (FIPS) code for the location where the individual was sentenced. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Incarceration
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,056,028	99.9
Invalid values	11,680	0.1
Missing values	190	0.0

`inc_st_ori_fips` (continued)

Histogram of `inc_st_ori_fips` (valid records)



inc_cnty_ori_fips

Label	County FIPS code for location of sentencing
Description	County-level Federal Information Processing Standards (FIPS) code for the location where the individual was sentenced. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Incarceration
Format	string
Code scheme	county FIPS code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	7,821,615	55.6
Invalid values	11,680	0.1
Missing values	6,234,413	44.3

Data notes

- Wisconsin: `WI001`

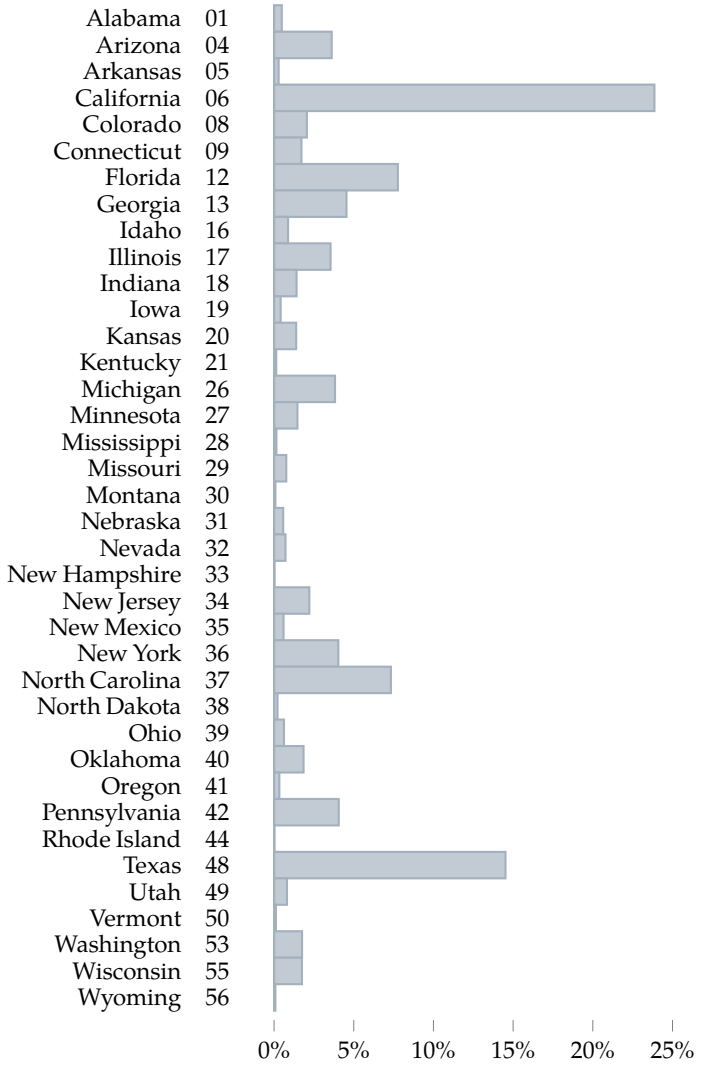
inc_st_juris_fips

Label	FIPS code of state responsible for physical supervision of individual
Description	State-level Federal Information Processing Standards (FIPS) code for the state with responsibility of physical supervision over individual's term of incarceration. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Incarceration
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

`inc_st_juris_fips` (continued)

Histogram of `inc_st_juris_fips` (valid records)

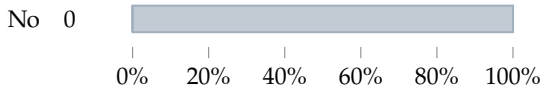


`inc_rec_src_bkg`

Label	Record source - booking agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a booking agency.
Table	Incarceration
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_rec_src_bkg` (valid records)

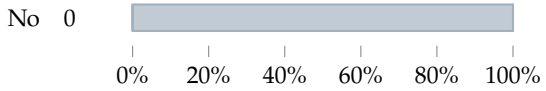


`inc_rec_src_arr`

Label	Record source - arresting agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from an arresting agency.
Table	Incarceration
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_rec_src_arr` (valid records)

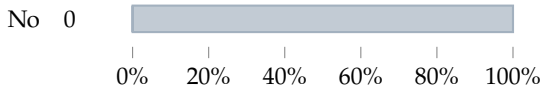


inc_rec_src_crt

Label	Record source - courts
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a court system.
Table	Incarceration
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_rec_src_crt` (valid records)

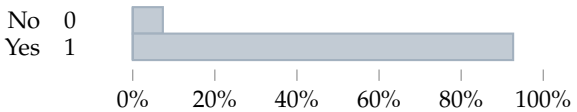


inc_rec_src_doc

Label	Record source - department of corrections
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a department of corrections.
Table	Incarceration
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_rec_src_doc` (valid records)

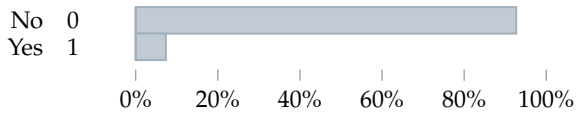


inc_rec_src_rep

Label	Record source - criminal history repository
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a criminal history repository.
Table	Incarceration
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_rec_src_rep` (valid records)

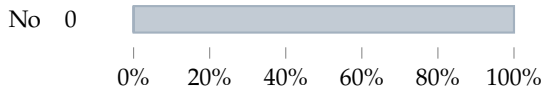


inc_rec_src_cc

Label	Record source - community corrections agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a community corrections agency.
Table	Incarceration
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	14,067,898	100.0
Valid records	14,067,898	100.0

Histogram of `inc_rec_src_cc` (valid records)



4.2.7 Parole

The parole table contains information on parole begin/end dates and exit status when available. Data is at the level of a parole term.

Table-level data notes

- All states: `US003`
- Connecticut: `CT001`
- Illinois: `IL001`
- Montana: `MT002`
- Nebraska: `NE003`
- New Jersey: `NJ004`
- New York: `NY001`
- Pennsylvania: `PA003`
- Wyoming: `WY001`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.3.1 and Figure 6.
Table	Parole
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Unique values	3,380,614	

par_id

Label	Parole term identifier
Description	Uniquely identifies terms of parole. For more details on use of par_id for data linkage, refer to Section 3.3.2 and Figure 7.
Table	Parole
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Unique values	7,008,581	

par_bgn_dt_yyyy

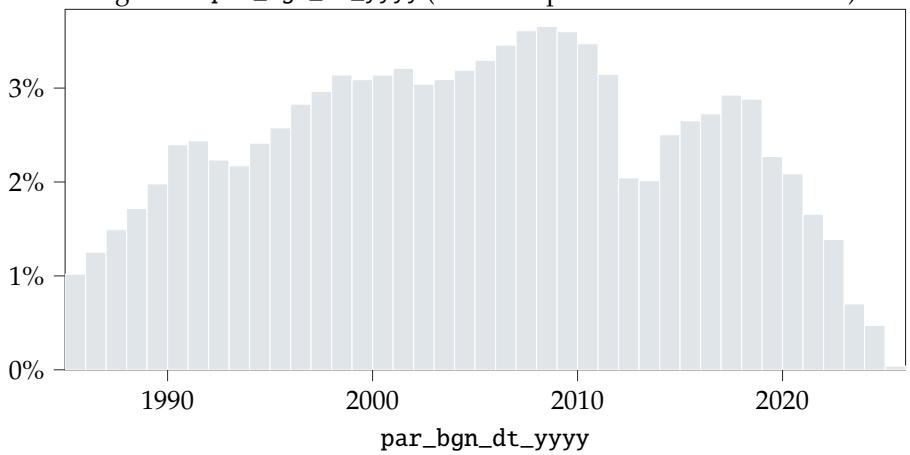
Label	Parole start date, year
Description	The year when the individual began parole.
Table	Parole
Format	numeric

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	6,927,635	98.8
Invalid values	1	0.0
Missing values	80,945	1.2

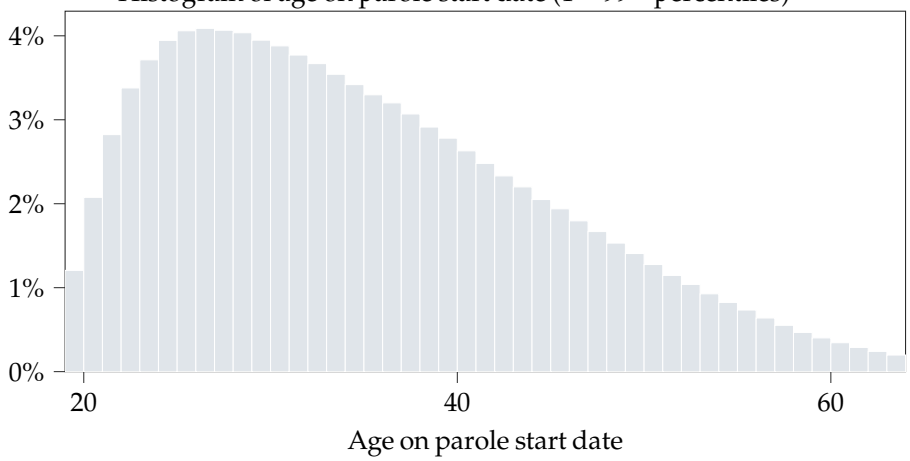
Statistic	*Value*
Mean	2,003.5
Minimum	1,900
25th percentile	1,996
Median	2,004
75th percentile	2,012
Maximum	2,025

`par_bgn_dt_yyyy` (continued)

Histogram of `par_bgn_dt_yyyy` (above 5th percentile of valid records)



Histogram of age on parole start date (1st–99th percentiles)

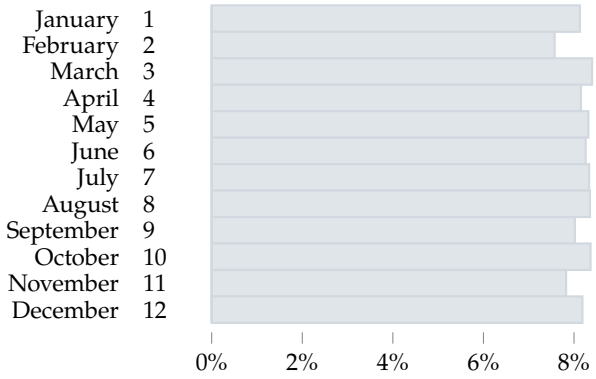


par_bgn_dt_mm

Label	Parole start date, month
Description	The month when the individual began parole.
Table	Parole
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	6,862,280	97.9
Missing values	146,301	2.1

Histogram of `par_bgn_dt_mm` (valid records)

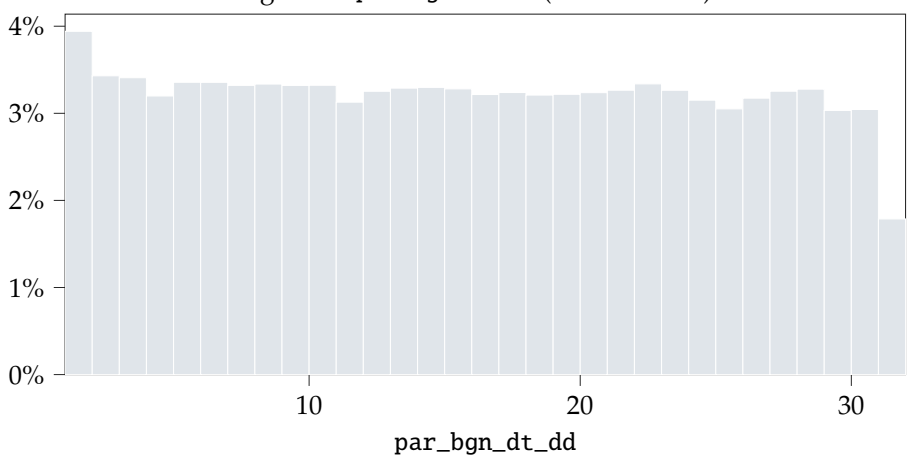


par_bgn_dt_dd

Label	Parole start date, day of month
Description	The day of the month when the individual began parole.
Table	Parole
Format	numeric

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	6,862,280	97.9
Missing values	146,301	2.1

Histogram of `par_bgn_dt_dd` (valid records)



Data notes

- Nebraska: NE002

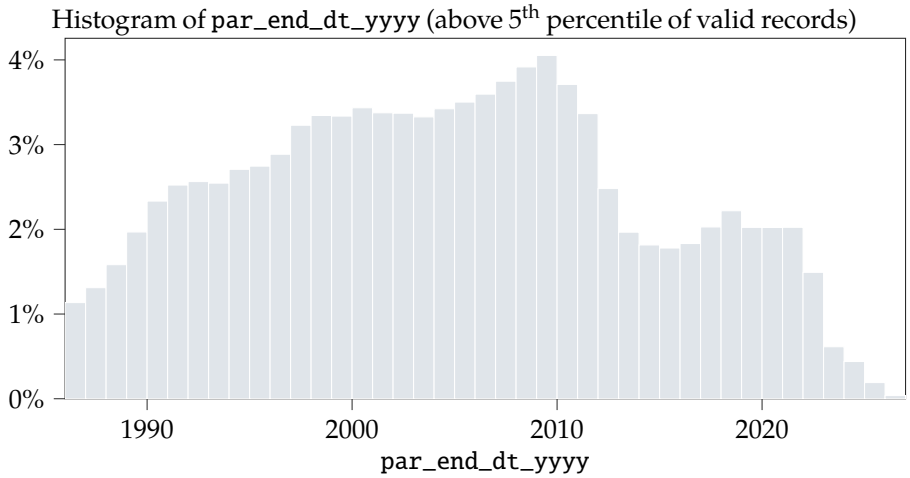
par_end_dt_yyyy

Label	Parole end date, year
Description	The year when the individual's parole ended.
Table	Parole
Format	numeric

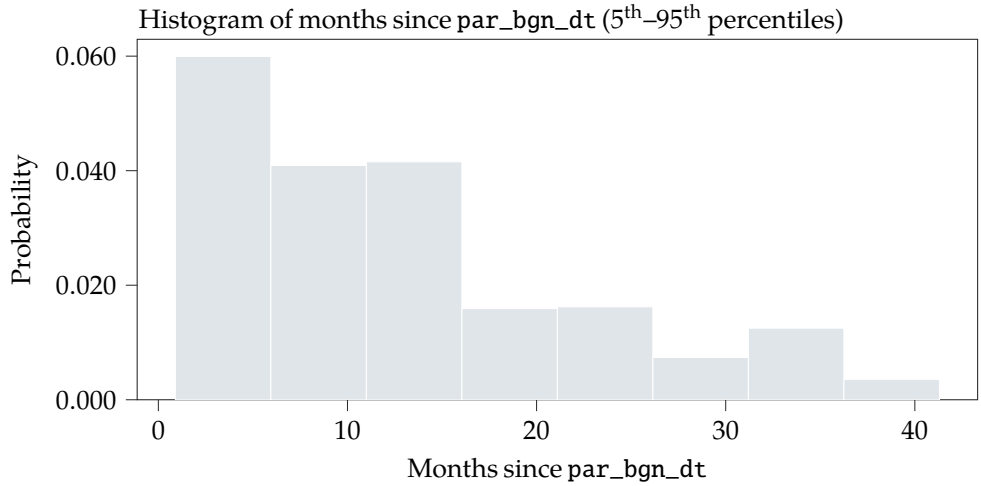
Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	5,374,894	76.7
Invalid values	1	0.0
Missing values	1,633,686	23.3

Statistic	*Value*
Mean	2,003.3
Minimum	1,900
25th percentile	1,996
Median	2,004
75th percentile	2,011
Maximum	2,026

Histogram of par_end_dt_yyyy (above 5th percentile of valid records)



`par_end_dt_yyyy` (continued)

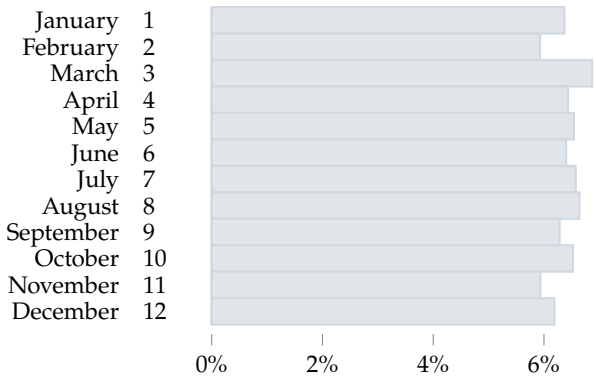


par_end_dt_mm

Label	Parole end date, month
Description	The month when the individual's parole ended.
Table	Parole
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	5,374,895	76.7
Missing values	1,633,686	23.3

Histogram of `par_end_dt_mm` (valid records)

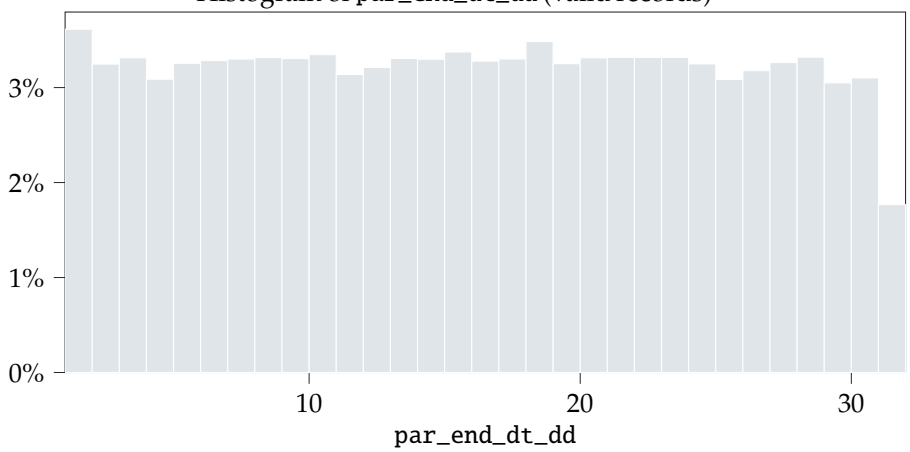


par_end_dt_dd

Label	Parole end date, day of month
Description	The day of the month when the individual's parole ended.
Table	Parole
Format	numeric

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	5,374,895	76.7
Missing values	1,633,686	23.3

Histogram of par_end_dt_dd (valid records)

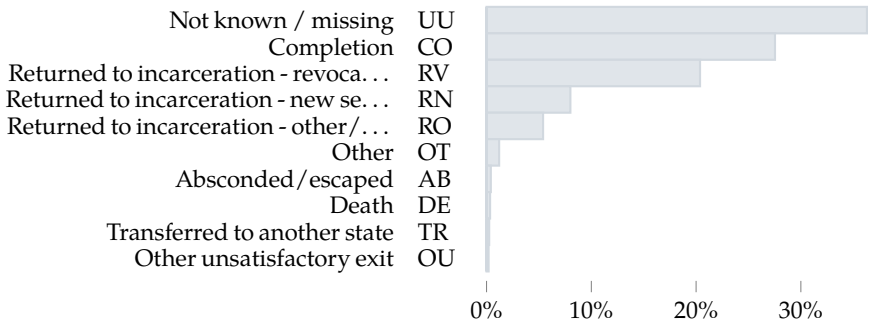


par_end_cd

Label	Parole end status classification
Description	CJARS standardized parole end status classification. Used when consistent coding is needed across jurisdictions. For a full description of the coding categories and method used to harmonize parole end descriptions, please refer to Appendices C and E.2.10.
Table	Parole
Format	string
Code scheme	parole exit code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of par_end_cd (valid records)



par_end_cd_src

Label	Raw description of parole end status from source
Description	A description of the parole end status for an individual. This variable provides the raw description that was provided from the agency. For harmonized parole end status please see par_end_cd. Even though a harmonized description of parole end status is provided, this raw description is retained for research that is not well-suited by the harmonized description of parole end status and leaves the opportunity for researchers to recode the original descriptions as needed.
Table	Parole
Format	string

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0

Data notes

- All states: US001

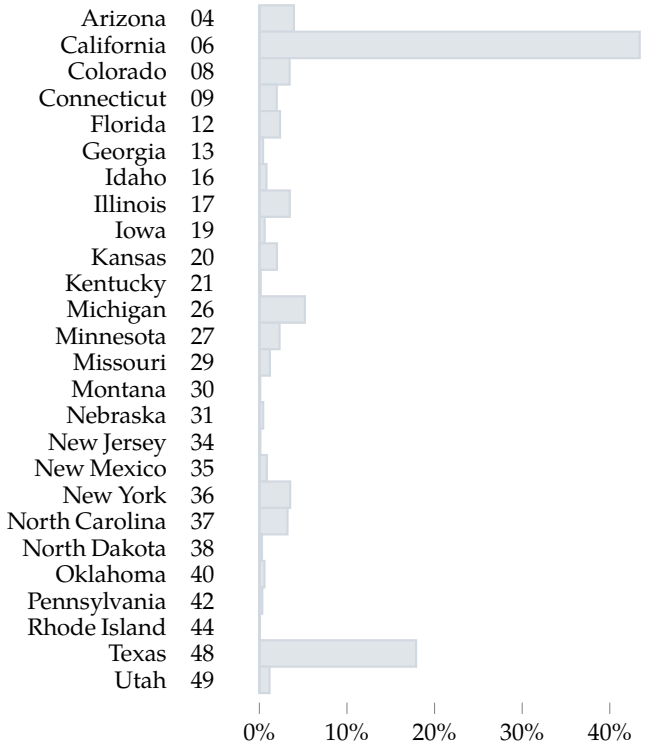
par_st_ori_fips

Label	State FIPS code for location of sentencing
Description	State-level Federal Information Processing Standards (FIPS) code for the location where the individual was sentenced. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Parole
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	6,998,415	99.9
Missing values	10,166	0.1

par_st_ori_fips (continued)

Histogram of par_st_ori_fips (valid records)



par_cnty_ori_fips

Label	County FIPS code for location of sentencing
Description	County-level Federal Information Processing Standards (FIPS) code for the location where the individual was sentenced. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Parole
Format	string
Code scheme	county FIPS code

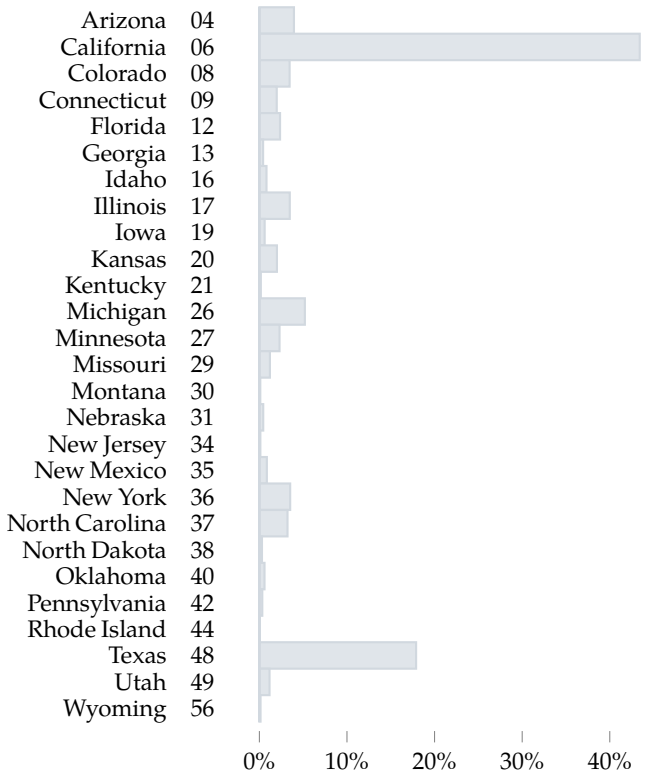
Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	1,931,974	27.6
Invalid values	812	0.0
Missing values	5,066,441	72.3

par_st_juris_fips

Label	FIPS code of state with jurisdiction over supervision of individual
Description	State-level Federal Information Processing Standards (FIPS) code for the state with jurisdiction over the individual's term of parole. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Parole
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_st_juris_fips` (valid records)

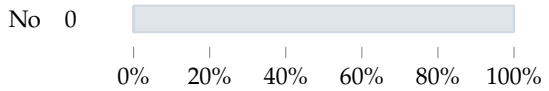


par_rec_src_bkg

Label	Record source - booking agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a booking agency.
Table	Parole
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_rec_src_bkg` (valid records)

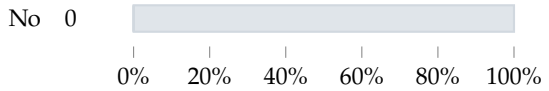


par_rec_src_arr

Label	Record source - arresting agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from an arresting agency.
Table	Parole
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_rec_src_arr` (valid records)

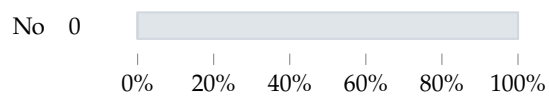


par_rec_src_crt

Label	Record source - courts
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a court system.
Table	Parole
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_rec_src_crt` (valid records)

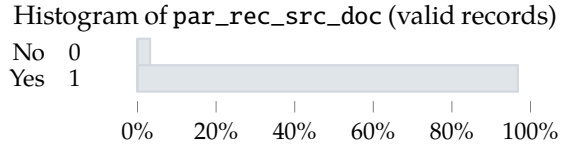


par_rec_src_doc

Label	Record source - department of corrections
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a department of corrections.
Table	Parole
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_rec_src_doc` (valid records)

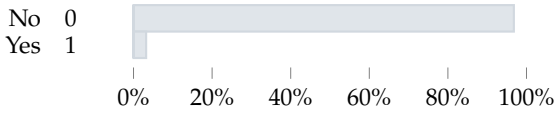


par_rec_src_rep

Label	Record source - criminal history repository
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a criminal history repository.
Table	Parole
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_rec_src_rep` (valid records)

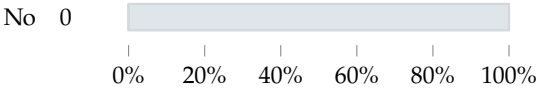


par_rec_src_cc

Label	Record source - community corrections agency
Description	A binary variable that indicates that the record was fully or partially generated using information that was sourced from a community corrections agency.
Table	Parole
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	7,008,581	100.0
Valid records	7,008,581	100.0

Histogram of `par_rec_src_cc` (valid records)



4.2.8 Coverage

The coverage table contains information on where and when jurisdictions are covered in the CJARS data. Data is at the level of a county-month. Data coverage is determined by a combination of automated and manual processes. First, an algorithm is run that was designed to determine the beginning and end of data coverage in a jurisdiction by identifying relatively large increases/decreases from month-to-month of events that are observed in the data. Next, CJARS staff with expertise in data collection and processing compare coverage beginning and end dates that were initially algorithmically determined, and compare them against basic frequency distributions of events observed in the data. Finally, CJARS staff make a final determination based on their informed approximation of data coverage. For additional details on coverage use, please refer to Appendix B

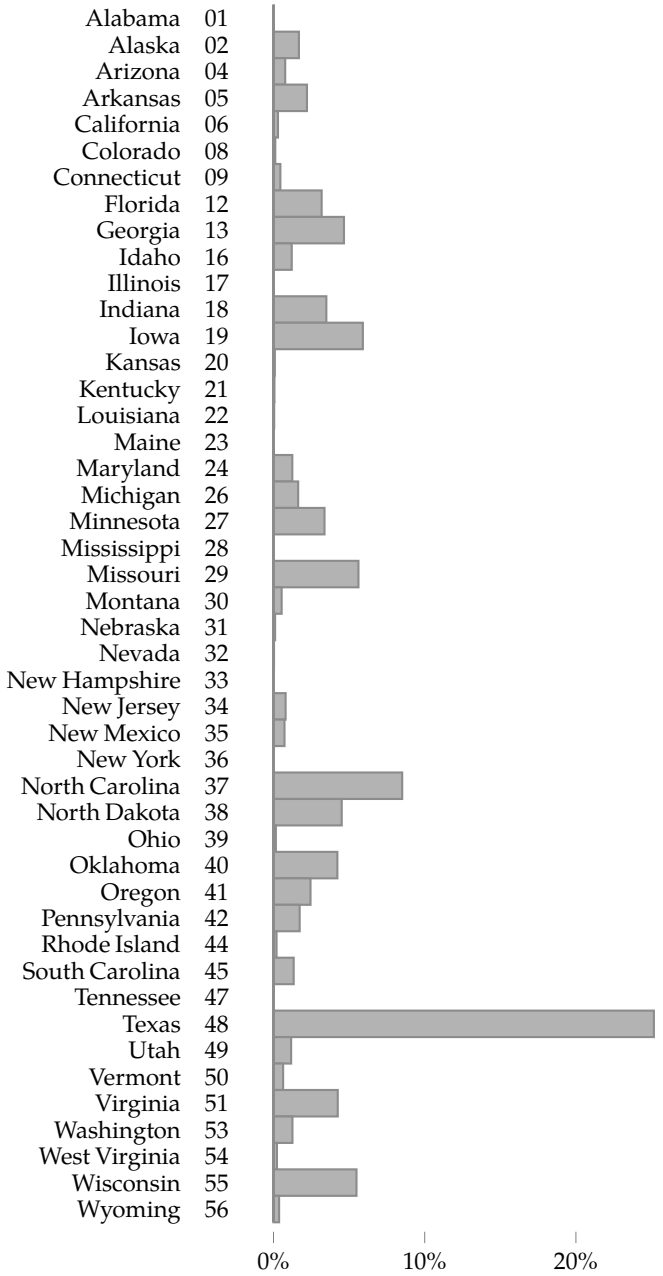
`st_fips`

Label	State FIPS code
Description	State-level Federal Information Processing Standards (FIPS) code of jurisdiction potentially covered by CJARS. Used to help uniquely identify geographic areas in the United States. Two-digit code used to identify states.
Table	Coverage
Format	string
Code scheme	state FIPS code

Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Valid records	1,348,061	100.0

st_fips (continued)

Histogram of st_fips (valid records)



cnty_fips

Label	County FIPS code
Description	County-level Federal Information Processing Standards (FIPS) code of jurisdiction potentially covered by CJARS. Used to help uniquely identify geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Coverage
Format	string
Code scheme	county FIPS code

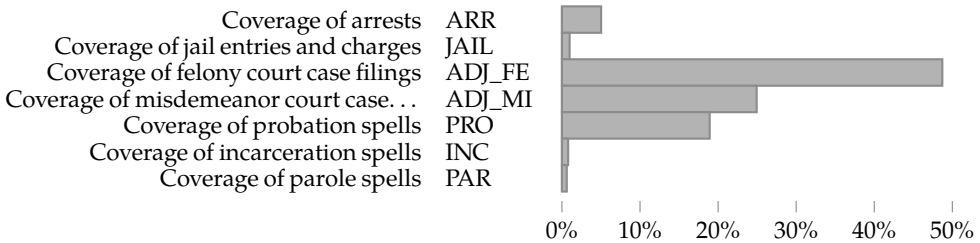
Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Valid records	1,319,027	97.8
Invalid values	2,186	0.2
Missing values	26,848	2.0

cjars_table

Label	Record type
Description	Type of record covered (i.e., arrest, misdemeanor case filings, felony case filings, incarceration spells, probation spells, or parole spells).
Table	Coverage
Format	string
Code scheme	coverage type

Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Valid records	1,348,061	100.0

Histogram of `cjars_table` (valid records)



month

Label	Year/month of coverage
Description	Year/month of coverage for a given state, county, and record type combination.
Table	Coverage
Format	numeric

Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Invalid values	1,348,061	100.0

dt_yyyy

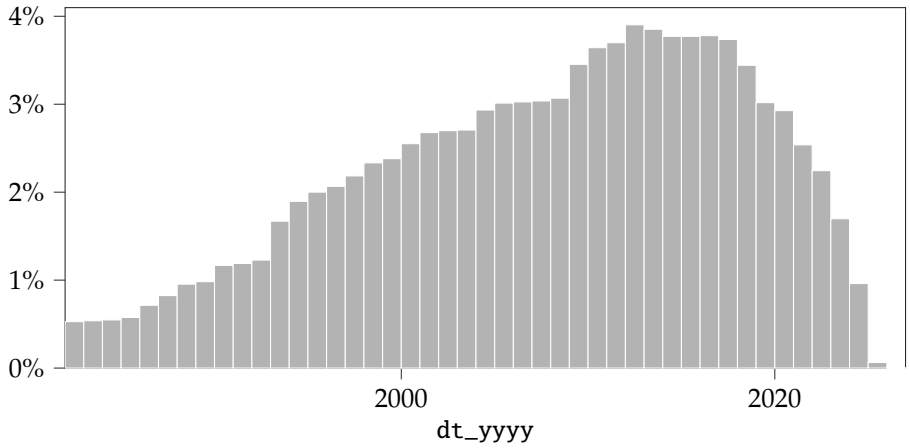
Label	Coverage month, year
Description	The year associated with a particular month of coverage for a given state, county, and record type combination.
Table	Coverage
Format	numeric

Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Valid records	1,348,061	100.0

Statistic	*Value*
Mean	2,005.7
Minimum	1,967
25th percentile	1,998
Median	2,008
75th percentile	2,015
Maximum	2,026

dt_yyyy (continued)

Histogram of dt_yyyy (above 5th percentile of valid records)

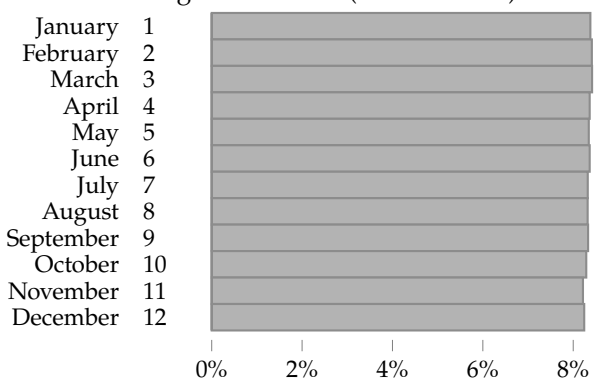


dt_mm

Label	Coverage date, month
Description	The month of the year associated with a particular month of coverage for a given state, county, and record type combination.
Table	Coverage
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Valid records	1,348,061	100.0

Histogram of dt_mm (valid records)

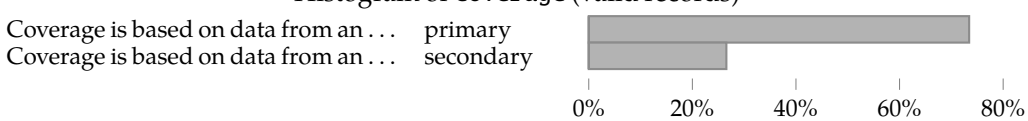


`coverage`

Label	Primary vs. secondary source
Description	Indicates whether coverage is from a primary data source (e.g., arrest data collected from a police department) or secondary source (e.g., data collected from a department of corrections that included court sentencing information) data source.
Table	Coverage
Format	string
Code scheme	coverage source

Set	*Count*	*Percent (%)*
All records	1,348,061	100.0
Valid records	1,348,061	100.0

Histogram of `coverage` (valid records)



4.3 ID variables used to link tables

The tables in the section above show that there are several ID variables contained in each of the CJARS data tables. The variables are used to identify individuals, events, episodes, and the linkages among them. Refer to Sections 3.3.1 and 3.3.2 for more information about how these linkages are structured.

> USER NOTE
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense. This means that the key identifying event variables have not yet been added to the CJARS relational tables.

Appendices

A Data sources

Table 3: Agencies whose data have been integrated into CJARS

State	Level	Domain	County	Municipality	Data provider/agency
Alabama	State	Corrections			Alabama Department of Corrections
	County	Sheriff	Covington		NYU Public Safety Lab Jail Data Initiative
			Jefferson		Jefferson County Sheriff's Office
			Lee		Lee County Sheriff's Office
Alaska	State	Judiciary			Commercial Vendor
Arizona	State	Corrections			Arizona Department of Corrections, Rehabilitation and Reentry
		Judiciary			Arizona Administrative Office of the Courts
	County	Sheriff	Maricopa		Maricopa County Sheriff's Office
			Mohave		Mohave County Sheriff's Office
			Navajo		Navajo County Sheriff's Office
			Yavapai		Yavapai County Sheriff's Office
			Yuma		Yuma County Sheriff's Office
Arkansas	State	Corrections			Arkansas Department of Correction
		Judiciary			Arkansas Administrative Office of the Courts
	County	Judiciary	Craighead		Craighead County District Court Clerk
		Sheriff	Craighead		Craighead County Sheriff's Office
			Faulkner		Faulkner County Sheriff's Office
			Garland		Garland County Sheriff's Office
			Jefferson		Jefferson County Sheriff's Office
			Lonoke		Lonoke County Sheriff's Office
			Pope		Pope County Sheriff's Office
			Pulaski		Pulaski County Sheriff's Office
			Saline		Saline County Sheriff's Office
			Sebastian		Sebastian County Sheriff's Office
			Washington		Washington County Sheriff's Office
			White		White County Arkansas Sheriff's Office
	Municipal	Judiciary	Benton	Bentonville	Bentonville District Court Clerk
California	State	Corrections			California Department of Corrections and Rehabilitation
	County	Judiciary	Amador		The Superior Court of California, County of Amador
			San Bernardino		The Superior Court of California, County of San Bernardino
			Santa Cruz		The Superior Court of California, County of Santa Cruz
			Stanislaus		The Superior Court of California, County of Stanislaus
		Sheriff	Butte		Butte County Sheriff's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Contra Costa		Contra Costa County Sheriff's Office
			Fresno		Fresno County Sheriff's Office
			Kern		Kern County Sheriff's Office
			Kings		Commercial Vendor
			Marin		Marin County Sheriff's Office
			Mendocino		Mendocino County Sheriff's Office
			Monterey		Monterey County Sheriff's Office
			Napa		NYU Public Safety Lab Jail Data Initiative
			Nevada		Nevada County Sheriff's Office
			Orange		Orange County Sheriff's Office
			Placer		Placer County Sheriff's Office
			Riverside		Riverside County Sheriff's Office
			San Bernardino		Commercial Vendor
			San Diego		San Diego County Sheriff's Office
			San Francisco		San Francisco Sheriff's Office
			San Joaquin		San Joaquin County Sheriff's Office
			San Luis Obispo		San Luis Obispo County Sheriff's Office
			San Mateo		San Mateo County Sheriff's Office
			Santa Barbara		Santa Barbara County Sheriff's Office
			Santa Cruz		Santa Cruz County Sheriff's Office
			Shasta		Shasta County Sheriff's Office
			Solano		Solano County Sheriff's Office
			Sonoma		Sonoma County Sheriff's Office
			Stanislaus		Stanislaus County Sheriff's Office
			Tulare		Tulare County Sheriff's Office
			Ventura		Ventura County Sheriff's Office
	Municipal	Police		Woodland	Woodland Police Department
			Kern	Bakersfield	Bakersfield Police Department
			Los Angeles	Long Beach	Long Beach Police Department
				Los Angeles	Los Angeles Police Department
				Santa Monica	Commercial Vendor
			Orange	Anaheim	Anaheim Police Department
			Riverside	Riverside	Riverside Police Department
			San Bernardino	San Bernardino	San Bernardino Police Department
			San Diego	San Diego	San Diego Police Department
			San Joaquin	Stockton	Stockton Police Department

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
Colorado	State	Corrections			Colorado Department of Corrections
	County	Sheriff	Adams		Adams County Sheriff's Office
			Boulder		Boulder County Sheriff's Office
			Larimer		Larimer County Sheriff's Office
			Pueblo		NYU Public Safety Lab Jail Data Initiative
			Weld		Weld County Sheriff's Office
Connecticut	State	Corrections			Connecticut Department of Correction
		Judiciary			State of Connecticut Judicial Branch
Delaware	State	Judiciary			Commercial Vendor
Florida	State	Corrections			Florida Department of Corrections
		Judiciary			Florida Court Clerks and Comptrollers
	County	Judiciary	Alachua		Alachua County Clerk of Courts
			Baker		Baker County Clerk of Courts
			Bay		Bay County Clerk of Courts
			Bradford		Bradford County Clerk of Courts
			Brevard		Brevard County Clerk of Courts
			Broward		Broward County Clerk of Courts
			Calhoun		Calhoun County Clerk of Courts
			Charlotte		Charlotte County Clerk of Courts
			Citrus		Citrus County Clerk of Courts
			Clay		Clay County Clerk of Courts
			Collier		Collier County Clerk of Courts
			Columbia		Columbia County Clerk of Courts
			DeSoto		DeSoto County Clerk of Courts
			Duval		Duval County Clerk of Courts
			Escambia		Escambia County Clerk of Courts
			Flagler		Flagler County Clerk of Courts
			Franklin		Franklin County Clerk of Courts
			Gadsden		Gadsden County Clerk of Courts
			Gilchrist		Gilchrist County Clerk of Courts
			Gulf		Gulf County Clerk of Courts
			Hamilton		Hamilton County Clerk of Courts
			Hardee		Hardee County Clerk of Courts
			Hendry		Hendry County Clerk of Courts
			Hernando		Hernando County Clerk of Courts
			Highlands		Highlands County Clerk of Courts

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Hillsborough		Hillsborough County Clerk of Courts
			Holmes		Holmes County Clerk of Courts
			Indian River		Indian River County Clerk of Courts
			Jackson		Jackson County Clerk of Courts
			Jefferson		Jefferson County Clerk of Courts
			Lafayette		Lafayette County Clerk of Courts
			Lee		Lee County Clerk of Courts
			Leon		Leon County Clerk of Courts
			Levy		Levy County Clerk of Courts
			Liberty		Liberty County Clerk of Courts
			Madison		Madison County Clerk of Courts
			Manatee		Manatee County Clerk of Courts
			Martin		Martin County Clerk of Courts
			Miami Dade		Miami-Dade County Clerk of Courts
			Monroe		Monroe County Clerk of Courts
			Nassau		Nassau County Clerk of Courts
			Okeechobee		Okeechobee County Clerk of Courts
			Orange		Orange County Clerk of Courts
			Osceola		Osceola County Clerk of Courts
			Palm Beach		Palm Beach County Clerk of Courts
			Pasco		Pasco County Clerk of Courts
			Pinellas		Pinellas County Clerk of Courts
			Polk		Polk County Clerk of Courts
			Putnam		Putnam County Clerk of Courts
			Santa Rosa		Santa Rosa County Clerk of Courts
			Sarasota		Sarasota County Clerk of Courts
			Seminole		Seminole County Clerk of Courts
			St Johns		St. Johns County Clerk of Courts
			St Lucie		St. Lucie County Clerk of Courts
			Sumter		Sumter County Clerk of Courts
			Suwannee		Suwannee County Clerk of Courts
			Taylor		Taylor County Clerk of Courts
			Union		Union County Clerk of Courts
			Volusia		Volusia County Clerk of Courts
			Wakulla		Wakulla County Clerk of Courts
			Walton		Walton County Clerk of Courts

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Washington		Washington County Clerk of Courts
		Sheriff	Broward		Commercial Vendor
			Citrus		Commercial Vendor
			Columbia		NYU Public Safety Lab Jail Data Initiative
			Hardee		Commercial Vendor
			Hillsborough		Commercial Vendor
			Lake		Commercial Vendor
			Lee		Commercial Vendor
			Martin		Commercial Vendor
			Osceola		NYU Public Safety Lab Jail Data Initiative
			Pinellas		Pinellas County Sheriff's Office
			Polk		Polk County Sheriff's Office
			Sarasota		NYU Public Safety Lab Jail Data Initiative
			St Lucie		NYU Public Safety Lab Jail Data Initiative
			Volusia		NYU Public Safety Lab Jail Data Initiative
			Walton		NYU Public Safety Lab Jail Data Initiative
Georgia	State	Corrections			Georgia Department of Corrections
	County	Judiciary	Baldwin		Baldwin County Clerk of Superior Court
			Bryan		Bryan County Clerk of Courts
			Camden		Camden County Clerk of Superior Court
			Carroll		Carroll County Clerk of Courts
			Catoosa		Catoosa County Clerk of Courts
			Clarke		Athens-Clarke County Superior Court Clerk
			Clayton		Clayton County State, Magistrate, Superior Courts
			Coweta		Coweta County Clerk of Courts
			Dougherty		Dougherty County Clerk of Superior Court
			Effingham		Effingham County Superior Court Clerk
			Houston		Houston County State Court Clerk Office
			Liberty		Libery County Office of the Clerk of Courts
			Newton		Newton County Clerk of Courts
		Sheriff	Banks		Bank County Sheriff's Office
			Bibb		Bibb County Sheriff's Office
			Camden		Camden County Sheriff's Office
			Chatham		Chatham County Sheriff's Office
			Cherokee		Cherokee County Sheriff's Office
			Clarke		Clarke County Sheriff's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Clayton		Clayton County Sheriff's Office
			Cobb		Cobb County Sheriff's Office
			Effingham		NYU Public Safety Lab Jail Data Initiative
			Gwinnett		Gwinnet County Sheriff's Office
			Houston		Houston County Sheriff's Office
			Newton		Newton County Sheriff's Office
			Rockdale		Rockdale County Sheriff's Office
			Troup		NYU Public Safety Lab Jail Data Initiative
Idaho	State	Corrections			Idaho Department of Corrections
Illinois	State	Corrections			Illinois Department of Corrections
	County	Sheriff	Kankakee		NYU Public Safety Lab Jail Data Initiative
Indiana	State	Corrections			Indiana Department of Correction
		Judiciary			Indiana Office of Judicial Administration
	County	Sheriff	Madison		NYU Public Safety Lab Jail Data Initiative
			Marion		NYU Public Safety Lab Jail Data Initiative
Iowa	State	Corrections			Iowa Department of Corrections
		Judiciary			Commercial Vendor
Kansas	State	Corrections			Kansas Department of Corrections
	County	Sheriff	Johnson		Johnson County Sheriff's Office
			Sedgwick		Sedgwick County Sheriff's Office
			Shawnee		NYU Public Safety Lab Jail Data Initiative
Kentucky	State	Corrections			Kentucky Department of Corrections
	County	Jail	Boone		Commercial Vendor
			Christian		Christian County Jail
			Fayette		Lexington-Fayette Urban County Division of Community Corrections
			Graves		Graves County Jail and Restricted Custody Center
			Grayson		Grayson County Detention Center
			Kenton		Commercial Vendor
			Letcher		Letcher County Jail
			Muhlenberg		NYU Public Safety Lab Jail Data Initiative
		Sheriff	Muhlenberg		Muhlenberg County Detention Center
Louisiana	County	Sheriff	Acadia		Acadia Parish Sheriff's Office
			Allen		Allen Parish Sheriff's Office
			Ascension		Ascension Parish Sheriff's Office
			Assumption		Assumption Parish Sheriff's Office
			Avoyelles		Avoyelles Parish Sheriff's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Beauregard		Beauregard Parish Sheriff's Office
			Bienville		Bienville Parish Sheriff's Office
			Bossier		Bossier Parish Sheriff's Office
			Caddo		Caddo Parish Sheriff's Office
			Calcasieu		Calcasieu Parish Sheriff's Office
			Caldwell		Caldwell Parish Sheriff's Office
			Cameron		Cameron Parish Sheriff's Office
			Catahoula		Catahoula Parish Sheriff's Department
			Claiborne		Claiborne Parish Sheriff's Office
			Concordia		Concordia Parish Sheriff's Office
			Desoto		DeSoto Parish Sheriff's Office
			East Baton Rouge		East Baton Rouge Parish Sheriff's Office
			East Feliciana		East Feliciana Parish Sheriff's Office
			Franklin		Franklin Parish Sheriff's Office
			Iberia		Iberia Parish Sheriff's Office
			Iberville		Iberville Parish Sheriff's Office
			Jefferson		Jefferson Parish Sheriff's Office
			Jefferson Davis		Jefferson Davis Parish Sheriff's Office
			Lafayette		Lafayette Parish Sheriff's Office
			Lafourche		NYU Public Safety Lab Jail Data Initiative
			Lasalle		LaSalle Parish Sheriff's Office
			Lincoln		Lincoln Parish Sheriff's Office
			Livingston		Livingston Parish Sheriff's Office
			Madison		Madison Parish Sheriff's Office
			Morehouse		Morehouse Parish Sheriff's Office
			Natchitoches		Natchitoches Parish Sheriff's Office
			Orleans		Orleans Parish Sheriff's Office
			Ouachita		Ouachita Parish Sheriff's Office
			Plaquemines		Plaquemines Parish Sheriff's Office
			Pointe Coupee		Pointe Coupee Parish Sheriff's Office
			Rapides		Rapides Parish Sheriff's Office
			Red River		Red River Parish Sheriff's Office
			Richland		Richland Parish Sheriff's Office
			Sabine		Sabine Parish Sheriff's Office
			St Bernard		St. Bernard Parish Sheriff's Office
			St Charles		St. Charles Parish Sheriff's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			St Helena		St. Helena Parish Sheriff's Office
			St James		St. James Parish Sheriff's Office
			St John		St. John the Baptist Parish Sheriff's Office
			St Landry		St. Landry Parish Sheriff's Office
			St Martin		St. Martin Parish Sheriff's Office
			St Mary		St. Mary Parish Sheriff's Office
			St Tammany		St. Tammany Parish Sheriff's Office
			Tangipahoa		Tangipahoa Parish Sheriff's Office
			Tensas		Tensas Parish Sheriff's Office
			Terrebonne		Terrebonne Parish Sheriff's Office
			Vermilion		Vermilion Parish Sheriff's Office
			Vernon		Vernon Parish Sheriff's Office
			Washington		Washington Parish Sheriff's Office
			Webster		Webster Parish Sheriff's Office
			West Baton Rouge		West Baton Rouge Parish Sheriff's Office
			West Carroll		West Carroll Parish Sheriff's Department
			West Feliciana		West Feliciana Parish Sheriff's Office
			Winn		Winn Parish Sheriff's Office
Maine	County	Sheriff	Cumberland		NYU Public Safety Lab Jail Data Initiative
Maryland	State	Judiciary			Maryland Judiciary
Michigan	State	Corrections			Michigan Department of Corrections
		Supplementary			Michigan Department of State
	County	Sheriff	Berrien		NYU Public Safety Lab Jail Data Initiative
			Wayne		NYU Public Safety Lab Jail Data Initiative
Minnesota	State	Corrections			Minnesota Department of Corrections
		Judiciary			Minnesota State Court Administrator's Office
	County	Sheriff	Polk		NYU Public Safety Lab Jail Data Initiative
Mississippi	State	Corrections			Mississippi Department of Corrections
	County	Sheriff	Desoto		Commercial Vendor
			Madison		NYU Public Safety Lab Jail Data Initiative
Missouri	State	Corrections			Missouri Department of Corrections
	County	Sheriff	Buchanan		NYU Public Safety Lab Jail Data Initiative
Montana	State	Corrections			Montana Department of Corrections
Nebraska	State	Corrections			Nebraska Department of Correctional Services
	County	Sheriff	Douglas		Douglas County Sheriff's Office
Nevada	State	Corrections			Nevada Department of Corrections

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
	County	Sheriff	Clark		Clark County Detention Center
	Municipal	Judiciary	Clark	Henderson	Henderson Municipal Court
				Las Vegas	City of Las Vegas Municipal Court
New Hampshire	State	Corrections			New Hampshire Department of Corrections
New Jersey	State	Corrections			New Jersey Department of Corrections
		Judiciary			Superior Court of New Jersey
	County	Sheriff	Bergen		Bergen County Sheriff's Office
			Cumberland		NYU Public Safety Lab Jail Data Initiative
			Essex		Commercial Vendor
			Ocean		Ocean County Sheriff's Office
			Passaic		Passaic County Sheriff's Office
			Salem		Commercial Vendor
			Union		Union County Sheriff's Office
New Mexico	State	Corrections			New Mexico Corrections Department
	County	Sheriff	San Juan		NYU Public Safety Lab Jail Data Initiative
New York	State	Corrections			New York State Department of Corrections and Community Supervision
	Municipal	Corrections	(5 NYC counties)	New York City	New York City Department of Correction
		Jail	(5 NYC counties)	New York City	New York City Department of Correction
North Carolina	State	Judiciary			North Carolina Administrative Office of the Courts
		Registry			NC/St/Registry: AGENCY NAME MISSING FROM CROSSWALK
		Repository			North Carolina Department of Public Safety
	County	Sheriff	Mecklenburg		Commercial Vendor
North Dakota	State	Corrections			North Dakota Department of Corrections and Rehabilitation
		Judiciary			North Dakota Court System
Ohio	State	Corrections			Ohio Department of Rehabilitation and Correction
		Registry			OH/St/Registry: AGENCY NAME MISSING FROM CROSSWALK
	County	Judiciary	Fairfield		Fairfield County Clerk of Courts
		Sheriff	Allen		NYU Public Safety Lab Jail Data Initiative
			Clermont		NYU Public Safety Lab Jail Data Initiative
			Franklin		NYU Public Safety Lab Jail Data Initiative
			Hancock		NYU Public Safety Lab Jail Data Initiative
			Highland		Commercial Vendor
			Lorain		NYU Public Safety Lab Jail Data Initiative
			Mahoning		NYU Public Safety Lab Jail Data Initiative
Oklahoma	State	Corrections			Oklahoma Department of Corrections
		Judiciary			Oklahoma Administrative Office of the Courts

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
	County	Sheriff	Carter		NYU Public Safety Lab Jail Data Initiative
			Comanche		NYU Public Safety Lab Jail Data Initiative
			Garfield		NYU Public Safety Lab Jail Data Initiative
			Oklahoma		NYU Public Safety Lab Jail Data Initiative
			Pottawatomie		NYU Public Safety Lab Jail Data Initiative
Oregon	State	Corrections			Oregon Department of Corrections
		Judiciary			Oregon Judicial Department
	County	Sheriff	Benton		Benton County Sheriff's Office
			Clackamas		Clackamas County Sheriff's Office
			Clatsop		Clatsop County Sheriff's Office
			Columbia		Columbia County Sheriff's Office
			Douglas		Douglas County Sheriff's Office
			Jackson		Jackson County Sheriff's Office
			Klamath		Klamath County Sheriff's Office
			Lane		Lane County Sheriff's Office
			Marion		Marion County Sheriff's Office
			Multnomah		Multnomah County Sheriff's Office
			Umatilla		Umatilla County Sheriff's Office
			Washington		Commercial Vendor
			Yamhill		Yamhill County Sheriff's Office
Pennsylvania	State	Corrections			Pennsylvania Department of Corrections
		Judiciary			Administrative Office of Pennsylvania Courts
	County	Sheriff	Greene		Commercial Vendor
			Lancaster		NYU Public Safety Lab Jail Data Initiative
			Westmoreland		NYU Public Safety Lab Jail Data Initiative
Rhode Island	State	Corrections			Rhode Island Department of Corrections
		Judiciary			Commercial Vendor
South Carolina	State	Judiciary			South Carolina Judicial Branch
	County	Judiciary	Greenville		Greenville County Clerk of Court
		Sheriff	Beaufort		Beaufort County Detention Center
			Charleston		Charleston County Sheriff's Office
			Cherokee		Cherokee County Sheriff's Office
			Georgetown		Georgetown County Sheriff's Office
			Horry		Horry County Sheriff's Office
			Marion		Marion County Detention Center
			Spartanburg		Spartanburg County Sheriff's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
	Municipal	Jail	Anderson	Anderson	NYU Public Safety Lab Jail Data Initiative
Tennessee	County	Sheriff	Davidson		Davidson County Sheriff's Office
			Knox		Knox County Sheriff's Office
			Robertson		NYU Public Safety Lab Jail Data Initiative
			Shelby		Shelby County Sheriff's Office
Texas	State	Corrections			Texas Department of Criminal Justice
		Judiciary			iDocket
		Registry			TX/St/Registry: AGENCY NAME MISSING FROM CROSSWALK
		Repository			Texas Department of Public Safety
	County	Judiciary	Angelina		Angelina County Clerk's Office
					Angelina County District Clerk's Office
			Aransas		Aransas County Clerk's Office
					Aransas County District Clerk's Office
			Bandera		Bandera County Clerk's Office
					Bandera County District Clerk's Office
			Bee		Bee County District Clerk's Office
			Bell		Bell County Clerk's Office
					Bell County District Clerk's Office
			Bexar		Bexar County Clerk
					Bexar County District Clerk's Office
			Bowie		Bowie County Clerk's Office
					Bowie County District Clerk's Office
			Brazoria		Brazoria County District Clerk's Office
			Briscoe		Briscoe County Clerk's Office
					Briscoe County District Clerk's Office
			Brown		Brown County Clerk's Office
			Caldwell		Caldwell County Clerk's Office
			Callahan		Callahan County District Clerk's Office
			Cameron		Cameron County Clerk's Office
					Cameron County District Clerk's Office
			Cherokee		Cherokee County District Clerk's Office
			Clay		Clay County Clerk's Office
			Coleman		Coleman County Clerk's Office
			Collin		Collin County Clerk's Office
					Collin County Courts
					Collin County District Clerk's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Colorado		Colorado County District Clerk's Office
			Coryell		Coryell County Clerk's Office
			Crane		Crane County Clerk's Office
					Crane County District Clerk's Office
			Dallam		Dallam County Clerk's Office
					Dallam County District Clerk's Office
			Dallas		Dallas County Clerk's Office
			Dickens		Dickens County Clerk's Office
					Dickens County District Clerk's Office
			Duval		Duval County District Clerk's Office
			Eastland		Eastland County Clerk's Office
			El Paso		El Paso County Clerk's Office
					El Paso District Clerk
			Floyd		Floyd County Clerk's Office
					Floyd County District Clerk's Office
			Franklin		Franklin County Clerk's Office
					Franklin County District Clerk's Office
			Galveston		Galveston County Clerk's Office
					Galveston County District Clerk's Office
			Goliad		Goliad County Clerk's Office
					Goliad County District Clerk's Office
			Harris		Harris County Clerk's Office
					Harris County District Clerk
			Hays		Hays County Clerk's Office
					Hays County Courts at Law
					Hays County District Clerk's Office
			Hidalgo		Hidalgo County Clerk's Office
					Hidalgo County District Clerk's Office
			Hopkins		Hopkins County Clerk's Office
					Hopkins County District Clerk's Office
			Houston		Houston County Clerk's Office
					Houston County District Clerk's Office
			Hudspeth		Hudspeth County Clerk's Office
					Hudspeth County District Clerk's Office
			Hunt		Hunt County Clerk's Office
			Jack		Jack County Clerk's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
					Jack County District Clerk's Office
			Jefferson		Jefferson County District Clerk's Office
			Jim Wells		Jim Wells County District Clerk's Office
			Kinney		Kinney County Clerk's Office
					Kinney County District Clerk's Office
			Liberty		Liberty County Clerk's Office
					Liberty County District Clerk's Office
			Limestone		Limestone County Clerk's Office
					Limestone County District Clerk's Office
			McLennan		McLennan County District Clerk's Office
			Montgomery		Montgomery County Clerk's Office
			Motley		Motley County Clerk's Office
					Motley County District Clerk's Office
			Nacogdoches		Nacogdoches County Clerk's Office
					Nacogdoches County District Clerk's Office
			Ochiltree		Ochiltree County Clerk's Office
					Ochiltree County District Clerk's Office
			Oldham		Oldham County Clerk's Office
					Oldham County District Clerk's Office
			Orange		Orange County Clerk's Office
			Polk		Polk County Clerk's Office
					Polk County District Clerk's Office
			Presidio		Presidio County Clerk's Office
					Presidio County District Clerk's Office
			San Patricio		San Patricio County Clerk's Office
					San Patricio County District Clerk's Office
			Schleicher		Schleicher County Clerk's Office
					Schleicher County District Clerk's Office
			Shelby		Shelby County Clerk's Office
			Starr		Starr County Clerk's Office
					Starr County District Clerk's Office
			Titus		Titus County District Clerk's Office
			Travis		Travis County District Clerk's Office
			Trinity		Trinity County Clerk's Office
			Tyler		Tyler County Clerk's Office
			Upton		Upton County Clerk's Office

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
					Upton County District Clerk's Office
			Val Verde		Val Verde County District Clerk's Office
			Victoria		Victoria County Clerk's Office
					Victoria County District Clerk's Office
			Washington		Washington County Clerk's Office
					Washington County District Clerk's Office
			Webb		Webb County Clerk's Office
					Webb County District Clerk's Office
			Wharton		Wharton County Clerk's Office
					Wharton County District Clerk's Office
			Willacy		Willacy County Clerk's Office
					Willacy County District Clerk's Office
			Young		Young County Clerk's Office
					Young County District Clerk's Office
			Zapata		Zapata County Clerk's Office
		Sheriff	Bell		NYU Public Safety Lab Jail Data Initiative
			Bexar		Bexar County Sheriff's Office
			Collin		Collin County Sheriff's Office
			Dallas		Dallas County Sheriff's Office
			El Paso		El Paso County Sheriff's Office
			Gregg		Commercial Vendor
			Harris		Harris County Sheriff's Office
			Hays		Hays County Sheriff's Office
			Jim Wells		NYU Public Safety Lab Jail Data Initiative
			Montgomery		Commercial Vendor
			Tarrant		Tarrant County Sheriff's Office
			Travis		Travis County Sheriff's Office
	Municipal	Jail	(2 TX counties)	League City	NYU Public Safety Lab Jail Data Initiative
		Police		Arlington	Arlington Police Department
				Austin	Austin Police Department
				Dallas	Dallas Police Department
				El Paso	El Paso Police Department
				Houston	Houston Police Department
				Plano	Plano Police Department
			Tarrant	Fort Worth	Fort Worth Police Department
Utah	State	Corrections			Utah Department of Corrections

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
Vermont	State	Corrections			Vermont Department of Corrections
		Judiciary			Vermont Judiciary
Virginia	State	Judiciary			Judiciary of Virginia
	Regional	Jail	(7 VA counties)		NYU Public Safety Lab Jail Data Initiative
	County	Judiciary	Accomack		Accomack Circuit Court
			Albemarle		Albemarle Circuit Court
			Alexandria		Alexandria Circuit Court
			Alleghany		Alleghany Circuit Court
			Amelia		Amelia Circuit Court
			Amherst		Amherst Circuit Court
			Appomattox		Appomattox Circuit Court
			Arlington		Arlington Circuit Court
			Augusta		Augusta Circuit Court
			Bath		Bath Circuit Court
			Bedford		Bedford Circuit Court
			Bland		Bland Circuit Court
			Botetourt		Botetourt Circuit Court
			Brunswick		Brunswick Circuit Court
			Buchanan		Buchanan Circuit Court
			Buckingham		Buckingham Circuit Court
			Buena Vista		Buena Vista Circuit Court
			Campbell		Campbell Circuit Court
			Caroline		Caroline Circuit Court
			Carroll		Carroll Circuit Court
			Charles City		Charles City Circuit Court
			Charlotte		Charlotte Circuit Court
			Charlottesville		Charlottesville Circuit Court
			Chesapeake		Chesapeake Circuit Court
			Chesterfield		Chesterfield Circuit Court
			Clarke		Clarke Circuit Court
			Colonial Heights		Colonial Heights Circuit Court
			Craig		Craig Circuit Court
			Culpeper		Culpeper Circuit Court
			Cumberland		Cumberland Circuit Court
			Danville		Danville Circuit Court
			Dickenson		Dickenson Circuit Court

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Dinwiddie		Dinwiddie Circuit Court
			Essex		Essex Circuit Court
			Fairfax		Fairfax Circuit Court
			Fauquier		Fauquier Circuit Court
			Floyd		Floyd Circuit Court
			Fluvanna		Fluvanna Circuit Court
			Franklin		Franklin Circuit Court
			Frederick		Frederick Circuit Court
			Fredericksburg		Fredericksburg Circuit Court
			Giles		Giles Circuit Court
			Gloucester		Gloucester Circuit Court
			Goochland		Goochland Circuit Court
			Grayson		Grayson Circuit Court
			Greene		Greene Circuit Court
			Greensville		Greensville Circuit Court
			Halifax		Halifax Circuit Court
			Hampton		Hampton Circuit Court
			Hanover		Hanover Circuit Court
			Henrico		Henrico Circuit Court
			Henry		Henry Circuit Court
			Highland		Highland Circuit Court
			Hopewell		Hopewell Circuit Court
			Isle of Wight		Isle of Wight Circuit Court
			King George		King George Circuit Court
			King William		King William Circuit Court
			King and Queen		King and Queen Circuit Court
			Lancaster		Lancaster Circuit Court
			Lee		Lee Circuit Court
			Loudoun		Loudoun Circuit Court
			Louisa		Louisa Circuit Court
			Lunenburg		Lunenburg Circuit Court
			Lynchburg		Lynchburg Circuit Court
			Madison		Madison Circuit Court
			Martinsville		Martinsville Circuit Court
			Mathews		Mathews Circuit Court
			Mecklenburg		Mecklenburg Circuit Court

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Middlesex		Middlesex Circuit Court
			Montgomery		Montgomery Circuit Court
			Nelson		Nelson Circuit Court
			New Kent		New Kent Circuit Court
			Newport News		Newport News Circuit Court
			Norfolk		Norfolk Circuit Court
			Northampton		Northampton Circuit Court
			Northumberland		Northumberland Circuit Court
			Nottoway		Nottoway Circuit Court
			Orange		Orange Circuit Court
			Page		Page Circuit Court
			Patrick		Patrick Circuit Court
			Petersburg		Petersburg Circuit Court
			Pittsylvania		Pittsylvania Circuit Court
			Portsmouth		Portsmouth Circuit Court
			Powhatan		Powhatan Circuit Court
			Prince Edward		Prince Edward Circuit Court
			Prince George		Prince George Circuit Court
			Prince William		Prince William Circuit Court
			Pulaski		Pulaski Circuit Court
			Radford		Radford Circuit Court
			Rappahannock		Rappahannock Circuit Court
			Richmond		Richmond County Circuit Court
			Richmond city		Richmond City Circuit Court
			Roanoke		Roanoke County Circuit Court
			Roanoke city		Roanoke City Circuit Court
			Rockbridge		Rockbridge Circuit Court
			Rockingham		Rockingham Circuit Court
			Russell		Russell Circuit Court
			Salem city		Salem Circuit Court
			Scott		Scott Circuit Court
			Shenandoah		Shenandoah Circuit Court
			Smyth		Smyth Circuit Court
			Southampton		Southampton Circuit Court
			Spotsylvania		Spotsylvania Circuit Court
			Stafford		Stafford Circuit Court

Table 3: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Staunto		Staunton Circuit Court
			Suffolk		Suffolk Circuit Court
			Surry		Surry Circuit Court
			Sussex		Sussex Circuit Court
			Tazewell		Tazewell Circuit Court
			Virginia Beach		Virginia Beach Circuit Court
			Warren		Warren Circuit Court
			Washington		Washington Circuit Court
			Waynesboro		Waynesboro Circuit Court
			Westmoreland		Westmoreland Circuit Court
			Williamsburg		Williamsburg/James City County Circuit Court
			Winchester		Winchester Circuit Court
			Wise		Wise Circuit Court
			Wythe		Wythe Circuit Court
			York		York County Poquoson Circuit Court
		Sheriff	Norfolk		Norfolk Sheriff's Office
	Municipal	Police		Dublin	Dublin Police Department
				North Prince George	Prince George County Police Department
				Saluda	Middlesex County Sheriff's Office
				Virginia Beach	Virginia Beach Police Department
				Winchester	Winchester Police Department
Washington	State	Corrections			Washington State Department of Corrections
		Registry			WA/St/Registry: AGENCY NAME MISSING FROM CROSSWALK
	County	Sheriff	Benton		NYU Public Safety Lab Jail Data Initiative
	Municipal	Jail	(2 WA counties)	Nisqually	NYU Public Safety Lab Jail Data Initiative
			King	Kent	NYU Public Safety Lab Jail Data Initiative
West Virginia	State	Jail			Commercial Vendor
Wisconsin	State	Corrections			Wisconsin Department of Corrections
		Judiciary			Wisconsin Court System
	County	Sheriff	Kenosha		Commercial Vendor
Wyoming	State	Corrections			Wyoming Department of Corrections

B CJARS Coverage table details

B.1 What is CJARS Coverage?

The coverage table provides information on which geographic units, for what months, for each criminal justice event type are reasonably represented, or covered, in the CJARS database. Coverage start and end dates for each data source are determined by combining information from the provider on expected coverage, distributions of observed event dates, and benchmarks against external sources. These are compiled into the coverage table provided to users. This compiled coverage table indicates to users when we are confident the majority of events for a given geographic unit/month/event type have been captured.

B.1.1 Coverage Table Structure

Each row in the coverage table specifies the relevant event table (`cjars_table`) (with the adjudication table subdivided into felony and misdemeanor categories). It also includes both a state FIPS code (`st_fips`) and a county FIPS code (`cnty_fips`) to indicate the geographic unit covered. The state and county FIPS code in the coverage table refer to the *_st_juris_fips and/or *_cnty_juris_fips, respectively, where the asterisk represents the event table. A missing county FIPS code indicates that the entire state is covered.

The provided covered year and month (`month`, `dt_yyyy` and `dt_mm`) correspond to the various event date variables in the event tables. Coverage is not assessed at the day level; a covered month indicates that >90% of the days of the month are covered. Coverage is primarily assessed based on the availability of event dates marking the beginning of criminal justice events. If sufficient data is also available for the end dates of these events, the reported coverage reflects the combined range of both start and end dates. End dates may be missing because an individual's criminal justice event is ongoing or because an accurate end date was not provided in the raw data. See Table 4 for a breakdown of each event table's respective beginning and end date variables.

Table 4: Event Dates

Event Table	Beginning Event Dates	End Event Dates
arrest	`arr_dt_`	
jail	`jail_entry_dt_`	`jail_exit_dt_`
adjudication	`adj_file_dt_`, (if file date not available: `adj_disp_dt_`, `adj_sent_dt_`, or `adj_off_dt_`)	
incarceration	`inc_entry_dt_`	`inc_exit_dt_`
parole	`par_bgn_dt_`	`par_end_dt_`
probation	`pro_bgn_dt_`	`pro_end_dt_`

Finally, the coverage table indicates whether the data is from a primary source or only from a secondary source via `coverage`. An entry of "primary" for a county-month indicates we expect the data source, hereafter referred to as a "primary source", to consist of a complete set of events for that period. Alternatively, an entry of "secondary" for a county-month indicates that we only expect the data source, the "secondary source" in this context, to contain a relevant subset of events for that period. If a jurisdiction has a period of overlapping coverage (primary and secondary) from available data, the primary source takes precedence. In order to appropriately merge the coverage table with an event table, it is important to factor in which rows in the event table come from primary and/or secondary sources. In each event table, a primary source is denoted with a "1" for the primary source variable within each row. Reference Table 5 for said variable names specific to each table, noting relevant exceptions.

Special attention is required when looking at data coverage at the county level due to the way data at the municipal level is represented in the table. Namely, if a county-month row in the coverage table is listed as coming from a primary source, this means at least one municipality in that county is fully covered but does not necessarily hold for other municipalities within the same county. Secondary coverage is handled as usual - indicating a relevant subset of events for that period is included. See Appendix A for a list of municipal data sources.

Table 5: Primary Record Sources

Event Table	Primary * rec src * variables	Exceptions
arrest	arr_rec_src_arr	
jail	jail_rec_src_bkg	
adjudication	adj_rec_src_crt, adj_rec_src_rep	Only adj_rec_src_crt is primary for North Carolina (State FIPS code 37)
incarceration	inc_rec_src_doc, inc_rec_src_rep	
parole	par_rec_src_doc, par_rec_src_rep	
probation	par_rec_src_doc, par_rec_src_rep	

B.2 Coverage determination examples

Below are some specific examples to illustrate how coverage is determined and represented in the coverage table.

B.2.1 Adjudication Example

For Florida there is secondary felony adjudication data from the Department of Corrections and primary felony adjudication data directly from the Court Clerks and Comptrollers, as can be seen in Appendix A. In the coverage table, for Florida felony adjudication, coverage starts with the secondary source in 1997m10 and switches to the primary source in 2005m1. This means that primary coverage from the Court source between 1997m10 and 2005m1 does not exist in this database; that data is only considered mostly covered starting in 2005m1. Secondary source data means only a subset of felony adjudication charges for people held by the Department of Corrections is represented. Because the primary source takes precedence over the secondary, it is impossible to tell if the secondary source has data that extends past 2005m1 based on the coverage table. In this example, rows in the adjudication table will have primary source variables adj_rec_src_doc set to "1" if the event comes from the Department of Corrections source and adj_rec_src_crt set to "1" if the event comes from the Court Clerks and Comptrollers source. Rows with both rec_src variables set to "1" have been identified as the same event and combined into a single row.

B.2.2 Arrest Example

For California there is secondary arrest data from the Kern County Sheriff's Office and primary arrest data from Bakersfield Police Department, a municipality within Kern County, as can be seen in Appendix A. Within the arrest table, the sheriff dataset is considered a secondary source because it only contains booking information, and the booked population is a subset of the total arrested population. In the coverage table, for California arrest coverage, the primary source is covered from 2002m12 to 2022m8. Because the Bakersfield Police Department is the only primary source municipality available to date for Kern county, the coverage denoted refers to arrests made by that police department. Coverage dates for arrests made by Kern County Sheriff's Office remain unknown, but it is possible to use arr_rec_src_bkg set to "1" to identify rows from that source. Rows from the Bakersfield source will have arr_rec_src_arr set to "1".

C Variable harmonization

Variation in legal statutes and across jurisdictions leads to substantial differences in the way that data are coded, processed, and stored from agency to agency. This results in significant barriers for researchers attempting to analyze data from multiple sources. To address these issues, the CJARS data infrastructure includes harmonized versions of key variables that describe criminal justice events. These include variables such as event dates, offense descriptions, disposition and sentencing information, and descriptions of begin/end status of probation, incarceration, and parole.

In addition, while the harmonized variables may be well-suited for research that extends across multiple jurisdictions, they may not fit all research questions well. For this reason, the original versions of variables received from the source are retained in the CJARS infrastructure so that researchers can chose to recode variables in the way that is most fitting for their research if they choose to do so.

C.1 Process

Data brought into the CJARS project goes through numerous steps in order to produce the harmonized set of variables that are available for research. Figure 8 gives an overview of the steps in the CJARS variable harmonization process. As can be seen from this figure, data processing is broken up into one of two divisions: identifiable and de-identified. The identifiable data are used for matching purposes while the de-identified data contains the information that is relevant for describing criminal justice events.

Harmonization begins when data is obtained from data providers and put onto the CJARS data system (intake). Following intake, the raw data is localized, which involves converting it into a Stata data file to prepare for processing. During localization, each record is also assigned a unique record identifier. Next, the data go through standardization. This involves processing all of the identifiable information to prepare for entity resolution (see Appendix F) and further harmonization.

The identifiable information that is processed through entity resolution is used to create a roster of all individuals in the CJARS data, which is ultimately used at the Census Bureau for matching purposes. Alternatively, the cleaned data that was de-identified goes into the de-identified division for further processing. This involves variable harmonization, episode resolution (see Appendix F), and then finally a complete CJARS research database is built that can be sent and integrated into the Census Bureau's records.

While there are many steps in data processing, variable harmonization is a key aspect of this process (harmonized variable schemes can be found in Appendix E). Variable harmonization is a complex task that employs many techniques including hand coding, employing the use of regular expression commands, and machine learning techniques (for an example of its application in offense classification, see Appendix D).

More details about the harmonization of other variables in each of the CJARS relational databases are included below in Table 6.

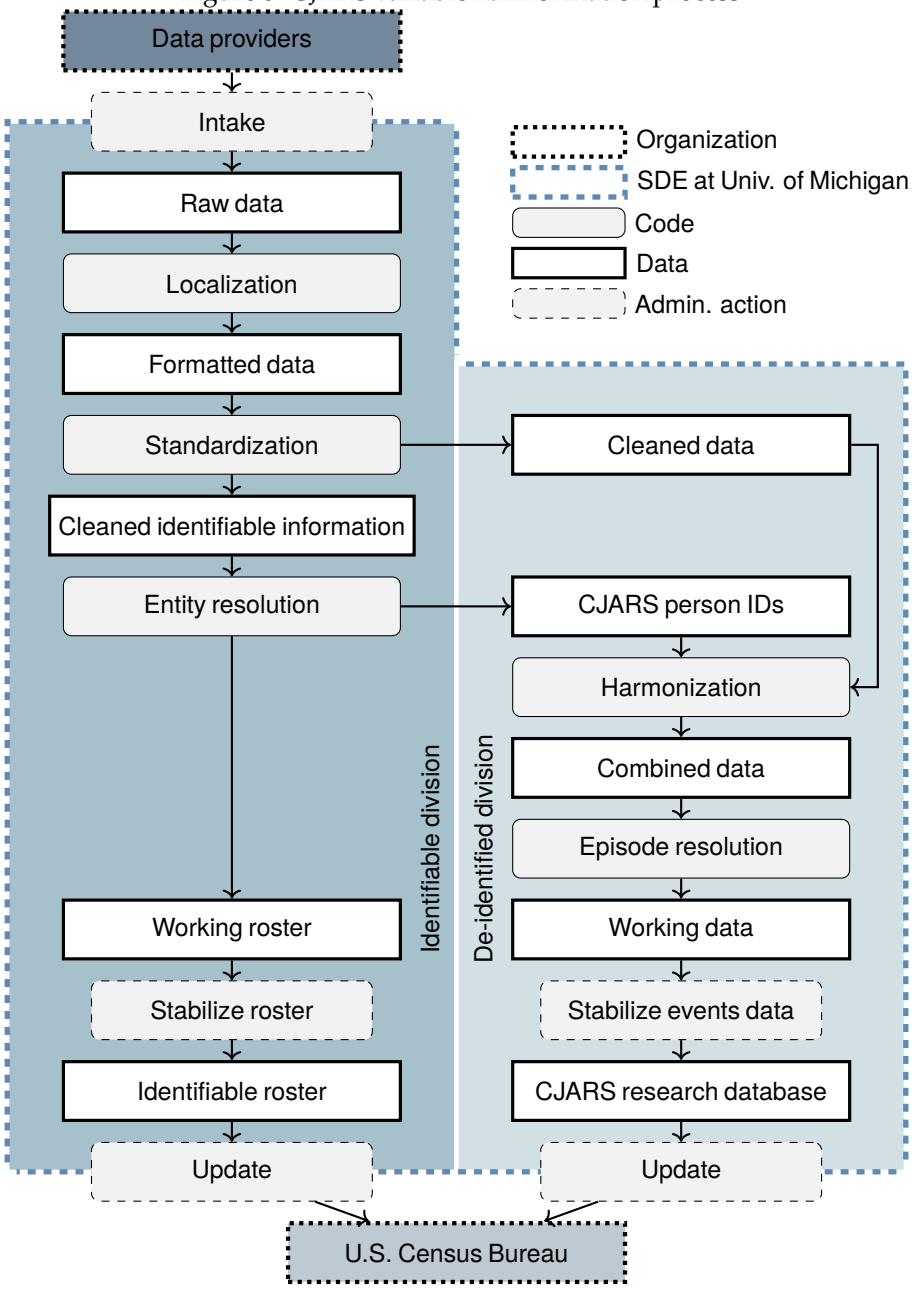
Table 6: Variable harmonization

CJARS relational database	Information	Variable(s)	Coding method(s)
arrest	arrest date	`arr_dt_yyyy`, `arr_dt_mm`, `arr_dt_dd`	recorded as separate variables for year, month, and day
arrest	offense classification	`arr_off_cd`	machine learning employed to generate a standardized offense type, (see Appendix C.1 for more details).
jail	jail entry date	`jail_entry_dt_yyyy`, `jail_entry_dt_mm`, `jail_entry_dt_dd`	recorded as separate variables for year, month, and day

CJARS relational database	Information	Variable(s)	Coding method(s)
jail	jail exit date	jail_exit_dt_yyyy, jail_exit_dt_mm, jail_exit_dt_dd	recorded as separate variables for year, month, and day
jail	jail bail amount	jail_bail_amt	hand-coded and use of regular expressions
jail	offense classification	jail_off_cd	machine learning employed to generate a standardized offense type, (see Appendix C.1 for more details).
jail	offense grade	jail_grd_cd	hand-coded and use of regular expressions
adjudication	offense grade	adj_grd_cd	hand-coded and use of regular expressions
adjudication	legal code	adj_off_lgl_cd	hand-coded and use of regular expressions
adjudication	case filing date	adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd	split into year, month, and day
adjudication	charge offense date	adj_off_dt_yyyy, adj_off_dt_mm, adj_off_dt_dd	split into year, month, and day
adjudication	charge offense classification	adj_chrg_off_cd	machine learning employed to generate a standardized offense type, (see Appendix C.1 for more details).
adjudication	disposition date	adj_disp_dt_yyyy, adj_disp_dt_mm, adj_disp_dt_dd	split into year, month, and day
adjudication	disposition description	adj_disp_cd	hand-coded and use of regular expressions
adjudication	disposition offense classification	adj_disp_off_cd	machine learning employed to generate a standardized offense type, (see Appendix C.1 for more details).
adjudication	sentence date	adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd	split into year, month, and day
adjudication	sentencing details	adj_sent_serv, adj_sent_dth, adj_sent_inc, adj_sent_pdiv, adj_sent_pro, adj_sent_rest, adj_sent_sus, adj_sent_trt, adj_sent_fine, adj_sent_inc_min, adj_sent_inc_max	hand-coded and use of regular expressions
incarceration	facility type	inc_fcl_cd	hand-coded and use of regular expressions

CJARS relational database	Information	Variable(s)	Coding method(s)
incarceration	entry date	inc_entry_dt_yyyy, inc_entry_dt_mm, inc_entry_dt_dd	split into year, month, and day
incarceration	entry status	inc_entry_cd	hand-coded and use of regular expressions
incarceration	exit date	inc_exit_dt_yyyy, inc_exit_dt_mm, inc_exit_dt_dd	split into year, month, and day
incarceration	exit status	inc_exit_cd	hand-coded and use of regular expressions
probation	conditions	pro_cond_cd	hand-coded and use of regular expressions
probation	begin date	pro_bgn_dt_yyyy, pro_bgn_dt_mm, pro_bgn_dt_dd	split into year, month, and day
probation	end date	pro_end_cd	split into year, month, and day
probation	end status	pro_end_dt_yyyy, pro_end_dt_mm, pro_end_dt_dd	hand-coded and use of regular expressions
parole	begin date	par_bgn_dt_yyyy, par_bgn_dt_mm, par_bgn_dt_dd	split into year, month, and day
parole	end date	par_end_dt_yyyy, par_end_dt_mm, par_end_dt_dd	split into year, month, and day
parole	end status	par_end_cd	hand-coded and use of regular expressions

Figure 8: CJARS variable harmonization process



D Offense classification

CJARS uses an offense classification scheme developed by Measures for Justice (MFJ). MFJ grounds its classification scheme on the codes developed for the National Corrections Reporting Program (NCRP) in the early 1990s. Those codes offered a disaggregated means to categorize offenses from state statutes. Charge descriptions vary widely across states based on statutory organization and language. MFJ's modifications include adding clarifications to the NCRP codes to ensure consistency, reclassifying DUI to its own offense type, reclassifying many of the "other" and "public order" offenses, and adding new codes for offenses including human trafficking, amphetamine drug offenses, opiate drug offenses, and other prescription drug offenses. The multi-digit scheme allows charges to be grouped easily into violent, property, drug, driving under the influence (DUI), public order, other, and unknown offense categories. The MFJ classification scheme can be found in Appendix 11.

CJARS and MFJ have developed Text-based Offense Classification (`TOC`) to classify offenses uniformly across states and jurisdictions. `TOC` is an offense classification tool that uses a hierarchical classification framework in which a multilayer perceptron classifier is trained for each local node. The data taxonomy used to create the hierarchy of classifiers is predefined in the MFJ offense classification schema as offense type code (parent class), offense category code (sub-parent class), and charge code (child class). In total, `TOC` uses 104 classifiers (one parent classifier, 12 sub-parent classifiers, and 92 child classifiers) and four independent classifiers for charge flags (domestic violence, gang, gun, and habitual offense).

Figure 9 outlines how hierarchical offense classification occurs. First, raw string descriptions are broken up into n-grams. N-grams are rolling windows of a string. The offense string "murder" can be split into the 3-gram set {"mur", "urd", "rde", "der"}. Next, the n-grams are used to classify descriptions into a parent class (e.g., violent, property). Then, offenses classified at the parent class level are further classified into a sub-parent class (e.g., murder, manslaughter). Finally, offenses are classified into a child class (e.g., attempt, conspiracy). This hierarchical approach was implemented because it substantially improved the predictive validity of the offense classification algorithm. The algorithms are tested against a test set of half a million unique offense descriptions hand validated by MFJ.

Figure 9: Offense hierarchical classification

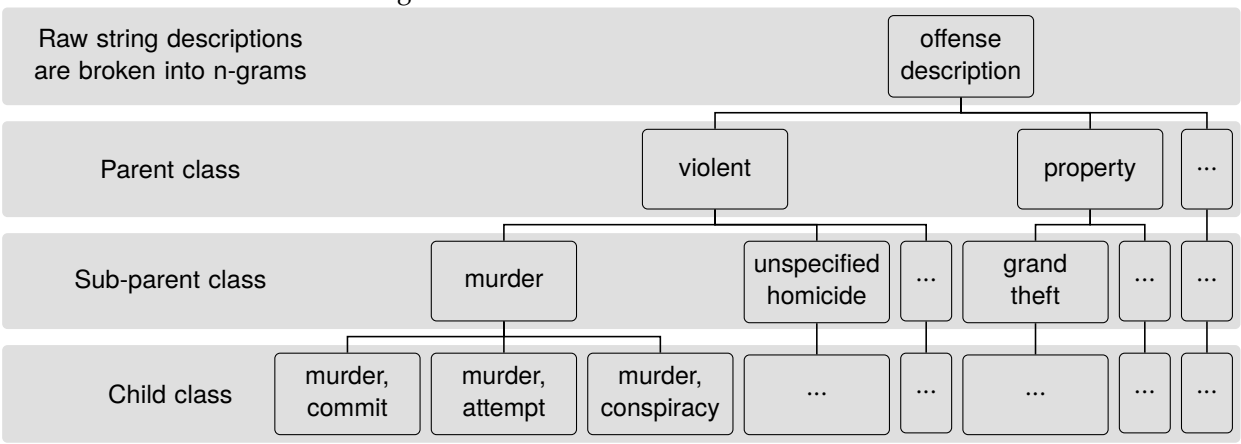
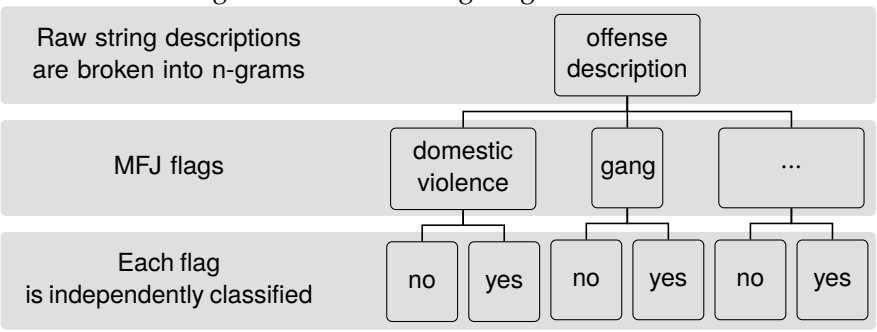


Figure 10 outlines how charge flags are assigned to offenses. Again, we start with n-grams of the offense string. The n-grams are then used to run multiple independent binary classification algorithms, one for each charge flag. Thus, offenses can be assigned multiple flags. This multi-label classification approach was used to predict charge flags in order to identify a comprehensive set of relevant labels for a given offense description.

Figure 10: Offense charge flag classification



E Code schemes including offense classifications

E.1 Geographic and demographic codes

E.1.1 State FIPS and abbreviations

This scheme is used for the following variables:

- adj_st_juris_fips
- arr_st_juris_fips
- st_fips
- inc_st_ori_fips
- inc_st_juris_fips
- jail_st_ori_fips
- jail_st_juris_fips
- par_st_ori_fips
- par_st_juris_fips
- pro_st_ori_fips
- pro_st_juris_fips

Table 7: State FIPS codes and abbreviations

State FIPS	State abbreviation	State name
01	AL	Alabama
02	AK	Alaska
04	AZ	Arizona
05	AR	Arkansas
06	CA	California
08	CO	Colorado
09	CT	Connecticut
10	DE	Delaware
11	DC	District of Columbia
12	FL	Florida
13	GA	Georgia
15	HI	Hawaii
16	ID	Idaho
17	IL	Illinois
18	IN	Indiana
19	IA	Iowa
20	KS	Kansas
21	KY	Kentucky
22	LA	Louisiana
23	ME	Maine
24	MD	Maryland
25	MA	Massachusetts
26	MI	Michigan
27	MN	Minnesota
28	MS	Mississippi
29	MO	Missouri

State FIPS	State abbreviation	State name
30	MT	Montana
31	NE	Nebraska
32	NV	Nevada
33	NH	New Hampshire
34	NJ	New Jersey
35	NM	New Mexico
36	NY	New York
37	NC	North Carolina
38	ND	North Dakota
39	OH	Ohio
40	OK	Oklahoma
41	OR	Oregon
42	PA	Pennsylvania
44	RI	Rhode Island
45	SC	South Carolina
46	SD	South Dakota
47	TN	Tennessee
48	TX	Texas
49	UT	Utah
50	VT	Vermont
51	VA	Virginia
53	WA	Washington
54	WV	West Virginia
55	WI	Wisconsin
56	WY	Wyoming
60	AS	American Samoa
66	GU	Guam
72	PR	Puerto Rico
78	VI	Virgin Islands of the U.S.

E.1.2 County FIPS

This scheme is used for the following variables:

- `adj_cnty_juris_fips`
- `arr_cnty_juris_fips`
- `cnty_fips`
- `inc_cnty_ori_fips`
- `jail_cnty_ori_fips`
- `jail_cnty_juris_fips`
- `par_cnty_ori_fips`
- `pro_cnty_ori_fips`
- `pro_cnty_juris_fips`

To see a list of all county FIPS codes, please see www.census.gov/geographies/reference-files/2018/demo/popest/2018-fips.html.

E.1.3 Sex codes

This scheme is used for the following variables:

- `sex`
- `sex_raw`

Table 8: Sex codes

Value	Label
1	male
2	female

E.1.4 Race and ethnicity codes

This scheme is used for the following variables:

- `race`
- `race_raw`

Table 9: Race and ethnicity codes

Value	Label
1	White, non-Hispanic
2	Black, non-Hispanic
3	Asian or Pacific Islander, non-Hispanic
4	Hispanic
5	American Indian or Alaska Native
6	Other race/ethnicity
9	Missing

E.1.5 Demographic imputation codes

This scheme is used for the following variables:

- `sex_imputed`
- `race_imputed`

Table 10: Demographic imputation codes

Value	Label
0	No Imputation
1	Imputation based on racial/ethnic national prevalence of last name using 2000 Decennial Census Surnames data
2	Imputation based on racial/ethnic prevalence of last name and first name within Census region among CJARS records with non-missing race/ethnicity values
3	Imputation using Census national surnames and CJARS regional first/last names resulted in the same outcome

E.2 Criminal justice event codes

E.2.1 Offense classification

CJARS uses an offense classification scheme developed by Measures for Justice (MFJ). MFJ grounds its classification scheme on the codes developed for the National Corrections Reporting Program (NCRP) in the early 1990s. Those codes offered a disaggregated means to categorize offenses from state statutes. Charge descriptions vary widely across states based on statutory organization and language. MFJ's modifications include reordering offenses to reflect seriousness, adding clarifications to the NCRP codes to ensure consistency, reclassifying DUI to its own offense type, reclassifying many of the "other" and "public order" offenses, and adding new codes for offenses including human trafficking, amphetamine drug offenses, opiate drug offenses, and other prescription drug offenses. The multi-digit scheme allows charges to be grouped easily into violent, property, drug, driving under the influence (DUI), public order, other, and unknown offense categories. Appendix D describes the machine learning algorithms CJARS has developed in partnership with MFJ to classify strings into this harmonized offense classification scheme.

This scheme is used for the following variables:

- `adj_chrg_off_cd`
- `adj_disp_off_cd`
- `arr_off_cd`
- `jail_off_cd`

Table 11: Offense classification scheme

Offense type	Offense code	Offense code description
Violent	1010	Murder
Violent	1011	Attempted murder
Violent	1012	Conspiracy to commit murder
Violent	1020	Unspecified homicide
Violent	1021	Unspecified homicide, attempted
Violent	1022	Unspecified homicide, conspiracy
Violent	1030	Voluntary manslaughter
Violent	1031	Voluntary manslaughter, attempted
Violent	1032	Voluntary manslaughter,conspiracy
Violent	1040	Vehicular manslaughter
Violent	1041	Vehicular manslaughter, attempted
Violent	1042	Vehicular manslaughter, conspiracy
Violent	1050	Involuntary manslaughter
Violent	1051	Involuntary manslaughter, attempt
Violent	1052	Involuntary manslaughter, conspiracy
Violent	1060	Kidnapping
Violent	1061	Kidnapping, attempted
Violent	1062	Kidnapping, conspiracy
Violent	1070	Rape
Violent	1071	Rape, attempted
Violent	1072	Rape, conspiracy
Violent	1080	Statutory rape
Violent	1081	Statutory rape, attempted
Violent	1082	Statutory rape, conspiracy
Violent	1090	Child molestation
Violent	1091	Child molestation, attempted
Violent	1092	Child molestation, conspiracy
Violent	1100	Sexual assault

Offense type	Offense code	Offense code description
Violent	1101	Sexual assault, attempted
Violent	1102	Sexual assault, conspiracy
Violent	1110	Human trafficking, sex - child
Violent	1111	Human trafficking, sex - child, attempted
Violent	1112	Human trafficking, sex - child, conspiracy
Violent	1120	Human trafficking, sex - adult or no age specified
Violent	1121	Human trafficking, sex - adult or no age specified, attempted
Violent	1122	Human trafficking, sex - adult or no age specified, conspiracy
Violent	1130	Human trafficking, labor - child
Violent	1131	Human trafficking, labor - child, attempted
Violent	1132	Human trafficking, labor - child, conspiracy
Violent	1140	Human trafficking, labor - adult or no age specified
Violent	1141	Human trafficking, labor - adult or no age specified, attempted
Violent	1142	Human trafficking, labor - adult or no age specified, conspiracy
Violent	1150	Human trafficking, unspecified - child
Violent	1151	Human trafficking, unspecified - child, attempted
Violent	1152	Human trafficking, unspecified - child, conspiracy
Violent	1160	Human trafficking, unspecified - adult or no age specified
Violent	1161	Human trafficking, unspecified - adult or no age specified, attempted
Violent	1162	Human trafficking, unspecified - adult or no age specified, conspiracy
Violent	1170	Human trafficking
Violent	1171	Human trafficking, attempted
Violent	1172	Human trafficking, conspiracy
Violent	1180	Armed robbery
Violent	1181	Armed robbery, attempted
Violent	1182	Armed robbery, conspiracy
Violent	1190	Unarmed robbery
Violent	1191	Unarmed robbery, attempted
Violent	1192	Unarmed robbery, conspiracy
Violent	1200	Aggravated assault
Violent	1201	Aggravated assault, attempted
Violent	1202	Aggravated assault, conspiracy
Violent	1210	Assault of an officer
Violent	1211	Assault of an officer, attempted
Violent	1212	Assault of an officer, conspiracy
Violent	1220	Child abuse
Violent	1221	Child abuse, attempted
Violent	1222	Child abuse, conspiracy
Violent	1230	Simple assault
Violent	1231	Simple assault, attempted
Violent	1232	Simple assault, conspiracy
Violent	1240	Extortion/threat
Violent	1241	Extortion/threat, attempted
Violent	1242	Extortion/threat, conspiracy
Violent	1250	Hit and run with bodily injury
Violent	1251	Hit and run with bodily injury, attempted

Offense type	Offense code	Offense code description
Violent	1252	Hit and run with bodily injury, conspiracy
Violent	1990	Violent offense, other
Violent	1991	Violent offense other, attempted
Violent	1992	Violent offense other, conspiracy
Property	2010	Burglary
Property	2011	Burglary, attempted
Property	2012	Burglary, conspiracy
Property	2020	Arson
Property	2021	Arson, attempted
Property	2022	Arson, conspiracy
Property	2030	Auto theft
Property	2031	Auto theft, attempted
Property	2032	Auto theft, conspiracy
Property	2040	Forgery/fraud
Property	2041	Forgery/fraud, attempted
Property	2042	Forgery/fraud, conspiracy
Property	2050	Grand theft (>$500)
Property	2051	Grand theft (>$500), attempted
Property	2052	Grand theft (>$500), conspiracy
Property	2060	Petty theft (=<$500)
Property	2061	Petty theft (=<$500), attempted
Property	2062	Petty theft (=<$500), conspiracy
Property	2070	Theft, value unknown
Property	2071	Theft, value unknown, attempted
Property	2072	Theft, value unknown, conspiracy
Property	2080	Financial crimes
Property	2081	Financial crimes attempted
Property	2082	Financial crimes conspiracy
Property	2090	Sale of stolen property
Property	2091	Sale of stolen property, attempted
Property	2092	Sale of stolen property, conspiracy
Property	2100	Receiving stolen property
Property	2101	Receiving stolen property, attempted
Property	2102	Receiving stolen property, conspiracy
Property	2110	Destruction of property
Property	2111	Destruction of property, attempted
Property	2112	Destruction of property, conspiracy
Property	2120	Hit and run driving with property damage
Property	2121	Hit and run driving, attempted
Property	2122	Hit and run driving, conspiracy
Property	2130	Unauthorized use of vehicle
Property	2131	Unauthorized use of vehicle, attempted
Property	2132	Unauthorized use of vehicle, conspiracy
Property	2140	Criminal trespass
Property	2141	Criminal trespass, attempted
Property	2142	Criminal trespass, conspiracy
Property	2150	Possession of property crime tools
Property	2151	Possession of property crime tools, attempted
Property	2152	Possession of property crime tools, conspiracy
Property	2990	Other property offense
Property	2991	Other property offense, attempt
Property	2992	Other property offense, conspiracy

Offense type	Offense code	Offense code description
Drug	3010	Distribution heroin
Drug	3011	Distribution, heroin, attempted
Drug	3012	Distribution, heroin, conspiracy
Drug	3020	Distribution of amphetamines
Drug	3021	Distribution of amphetamines, attempted
Drug	3022	Distribution of amphetamines, conspiracy
Drug	3030	Distribution cocaine or crack
Drug	3031	Distribution cocaine or crack, attempted
Drug	3032	Distribution cocaine or crack, conspiracy
Drug	3040	Distribution of opioids
Drug	3041	Distribution of opioids, attempted
Drug	3042	Distribution of opioids, conspiracy
Drug	3050	Distribution of prescription drugs
Drug	3051	Distribution of prescription drugs, attempted
Drug	3052	Distribution of prescription drugs, conspiracy
Drug	3060	Distribution other controlled substances
Drug	3061	Distribution other controlled substances, attempted
Drug	3062	Distribution other controlled substances, conspiracy
Drug	3070	Distribution marijuana
Drug	3071	Distribution marijuana, attempted
Drug	3072	Distribution marijuana, conspiracy
Drug	3080	Distribution, drug unspecified
Drug	3081	Distribution, drug unspecified, attempted
Drug	3082	Distribution, drug unspecified, conspiracy
Drug	3090	Possession/use of heroin
Drug	3091	Possession/use of heroin, attempted
Drug	3092	Possession/use of heroin, conspiracy
Drug	3100	Possession of amphetamines
Drug	3101	Possession of amphetamines, attempted
Drug	3102	Possession of amphetamines, conspiracy
Drug	3110	Possession/use of cocaine or crack
Drug	3111	Possession/use of cocaine or crack, attempted
Drug	3112	Possession/use of cocaine or crack, conspiracy
Drug	3120	Possession of opioids
Drug	3121	Possession of opioids, attempted
Drug	3122	Possession of opioids, conspiracy
Drug	3130	Possession of prescription drugs
Drug	3131	Possession of prescription drugs, attempted
Drug	3132	Possession of prescription drugs, conspiracy
Drug	3140	Possession/use of other controlled substance
Drug	3141	Possession/use of other controlled substance, attempted
Drug	3142	Possession/use of other controlled substance, conspiracy
Drug	3150	Possession/use of marijuana
Drug	3151	Possession/use of marijuana, attempted
Drug	3152	Possession/use of marijuana, conspiracy
Drug	3160	Possession/use of unspecified drug
Drug	3161	Possession/use, drug unspecified, attempted
Drug	3162	Possession/use, drug unspecified, conspiracy
Drug	3170	Heroin violation, offense unspecified
Drug	3180	Amphetamines, offense unspecified
Drug	3190	Cocaine/crack violation, offense unspecified
Drug	3200	Prescription of opioid drugs, offense unspecified

Offense type	Offense code	Offense code description
Drug	3210	Prescription, offense unspecified
Drug	3220	Other controlled substance violation, offense unspecified
Drug	3230	Marijuana violation, offense unspecified
Drug	3240	Fraudulent drug offense
Drug	3241	Fraudulent drug offense, attempted
Drug	3242	Fraudulent drug offense, conspiracy
Drug	3250	Drug paraphernalia
Drug	3251	Drug paraphernalia, attempted
Drug	3252	Drug paraphernalia, conspiracy
Drug	3990	Other drug offense
Drug	3991	Other drug offense, attempt
Drug	3992	Other drug offense, conspiracy
DUI offense	4010	Driving while intoxicated
DUI offense	4011	Driving while intoxicated, attempted
DUI offense	4012	Driving while intoxicated, conspiracy
DUI offense	4020	Driving under the influence of alcohol
DUI offense	4021	Driving under the influence of alcohol, attempted
DUI offense	4022	Driving under the influence of alcohol, conspiracy
DUI offense	4030	Driving under the influence of drugs
DUI offense	4031	Driving under the influence of drugs, attempted
DUI offense	4032	Driving under the influence of drugs, conspiracy
Public order	5010	Riot
Public order	5011	Riot, attempting to incite
Public order	5012	Riot, conspiracy to incite
Public order	5020	Escape from custody
Public order	5021	Escape from custody, attempted
Public order	5022	Escape from custody, conspiracy
Public order	5030	Flight to avoid prosecution
Public order	5031	Flight to avoid prosecution, attempted
Public order	5032	Flight to avoid prosecution, conspiracy
Public order	5040	Weapons offense
Public order	5041	Weapons offense, attempted
Public order	5042	Weapons offense, conspiracy
Public order	5050	Habitual offender
Public order	5060	Parole violation
Public order	5070	Probation violation
Public order	5080	Contempt of court/violate court order
Public order	5081	Contempt of court/violate court order, attempted
Public order	5082	Contempt of court/violate court order, conspiracy
Public order	5090	Other court offense
Public order	5091	Other court offense, attempted
Public order	5092	Other court offense, conspiracy
Public order	5100	Family or custody related offense
Public order	5101	Family or custody related offense, attempted
Public order	5102	Family or custody related offense, conspiracy
Public order	5110	Offense against morals/decency
Public order	5111	Offense against morals/decency, attempted
Public order	5112	Offense against morals/decency, conspiracy
Public order	5120	Immigration violation
Public order	5121	Immigration violation, attempted
Public order	5122	Immigration violation, conspiracy
Public order	5130	Obstruction/resisting

Offense type	Offense code	Offense code description
Public order	5131	Obstruction/resisting, attempted
Public order	5132	Obstruction/resisting, conspiracy
Public order	5140	Invasion of privacy
Public order	5141	Invasion of privacy, attempted
Public order	5142	Invasion of privacy, conspiracy
Public order	5150	Commercialized vice
Public order	5151	Commercialized vice, attempted
Public order	5152	Commercialized vice, conspiracy
Public order	5160	Contributing to the delinquency of a minor
Public order	5161	Contributing to the delinquency of a minor, attempted
Public order	5162	Contributing to the delinquency of a minor, conspiracy
Public order	5170	Disorderly conduct offense
Public order	5171	Disorderly conduct offense, attempted
Public order	5172	Disorderly conduct offense, conspiracy
Public order	5180	Liquor law violation
Public order	5181	Liquor law violation, attempted
Public order	5182	Liquor law violation, conspiracy
Public order	5190	Taxation offense
Public order	5191	Taxation offense, attempted
Public order	5192	Taxation offense, conspiracy
Public order	5200	Bribery/conflict of interest
Public order	5201	Bribery/conflict of interest, attempt
Public order	5202	Bribery/conflict of interest, conspiracy
Public order	5990	Public order offense, other
Public order	5991	Public order offense, other, attempted
Public order	5992	Public order offense, other, conspiracy
Criminal traffic	6010	Traffic offense, minor
Not known/missing	9010	Unspecified felony
Not known/missing	9011	Unspecified felony, attempt
Not known/missing	9012	Unspecified felony, conspiracy
Not known/missing	9020	Unspecified misdemeanor
Not known/missing	9021	Unspecified misdemeanor, attempt
Not known/missing	9022	Unspecified misdemeanor, conspiracy
Other	9990	Other offense
Exclude	8010	Juvenile offense
Flag for removal	8020	Flag for removal
Call for service	8030	Call for service
Federal charges	8040	Federal charges
Not available	8050	Variable not available in county
Not applicable	8060	Not applicable
Not known/missing	9999	Not known/missing

E.2.2 Offense charge grade

This scheme is used for the following variables:

- `adj_grd_cd`
- `jail_grd_cd`

Table 12: Charge grade classification scheme

Charge grade code	Charge grade description
FE	Felony-level charge
MI	Misdemeanor-level charge
UU	Not known / missing

E.2.3 Offense legal code

This scheme is used for the following variables:

- `adj_off_lgl_cd`

Table 13: Charge legal code classification scheme

Legal code	Legal code description
ST	Charge defined by state statute
OR	Charge defined by ordinance code
UU	Not known / missing

E.2.4 Court disposition

This scheme is used for the following variables:

- `adj_disp_cd`

Table 14: Court disposition classification scheme

Disposition code	Disposition parent code	Disposition child code	Disposition description
DU	D	U	Diversion - unclassified
GC	G	C	Guilty - court trial
GJ	G	J	Guilty - jury trial
GP	G	P	Guilty - plea
GI	G	I	Guilty - insanity
GU	G	U	Guilty - unclassified
NA	N	A	Acquittal
ND	N	D	Dismissal
NI	N	I	Dismissal - insanity
NM	N	M	Mistrial
NU	N	U	Not guilty - unclassified
PT	P	T	Procedural - transfer
PU	P	U	Procedural - unclassified
UU			Not known / missing

E.2.5 Probation conditions

This scheme is used for the following variables:

- `pro_cond_cd`

Table 15: Probation conditions classification scheme

Probation condition code	Probation condition description
PJ	Probation with jail
SP	Straight probation
AD	Alcohol/drug residential
PR	Probation with community residential
UU	Not known / missing

E.2.6 Probation exit

This scheme is used for the following variables:

- pro_end_cd

Table 16: Probation exit classification scheme

Probation exit code	Probation exit description
CO	Completion
IN	Incarcerated
AB	Absconded/escaped
DI	Discharged to custody/detainer/warrant
OU	Other unsatisfactory exit
TR	Transferred to another probation agency
DE	Death
OT	Other
UU	Not known / missing

E.2.7 Incarceration entry

This scheme is used for the following variables:

- inc_entry_cd

Table 17: Incarceration entry classification scheme

Incarceration entry code	Incarceration entry description
CC	Court commitment
RA	Returned from appeal or bond
TR	Transfer
RW	Parole revocation - new sentence
RN	Parole revocation - no new sentence
RI	Parole revocation - no information on new sentence
MW	Mandatory parole release - new sentence
MN	Mandatory parole release - no new sentence
MI	Mandatory parole release - no information on new sentence
SS	Suspended sentence imposed
EW	Escapee/AWOL returned - new sentence
EN	Escapee/AWOL returned - no new sentence
EI	Escapee/AWOL returned - no information on new sentence
PP	Parole status - pending revocation
MP	Mandatory parole release status - pending
PW	Probation revocation - new sentence

Incarceration entry code	Incarceration entry description
PN	Probation revocation - no new sentence
OT	Other
PR	Probation status - pending revocation
UC	Unsentenced commitment
IE	Illegal entry
UU	Not known / missing

E.2.8 Incarceration facility type

This scheme is used for the following variables:

- `inc_fcl_cd`

Table 18: Incarceration facility type classification scheme

Incarceration custody code	Incarceration custody description
CM	Community
MN	Minimum - low
MD	Medium
MX	Maximum - high or close
CX	Complex (federal only)
AD	Administrative (federal only)
FD	Federal prison
SP	State prison
LJ	Local jail
OT	Other
UU	Not known / missing

E.2.9 Incarceration exit

This scheme is used for the following variables:

- `inc_exit_cd`

Table 19: Incarceration exit classification scheme

Incarceration exit code	Incarceration exit description
PD	Parole Board Decision
MR	Mandatory Parole Release
PR	Probation Release
OR	Other Conditional Release
ES	Expiration of Sentence
CP	Commutation/Pardon
RC	Release to Custody, Detainer, Warrant
UR	Other Unconditional Release
DN	Death, Natural Causes
SU	Suicide
HI	Homicide by Another Inmate
OH	Other Homicide
EX	Execution
OD	Other Death
TR	Transfer

Incarceration exit code	Incarceration exit description
RA	Release on Appeal or Bond
OT	Other
EA	Escape/AWOL
AI	Accidental Injury to Self
IE	Illegal Entry
UU	Not Known / Missing

E.2.10 Parole exit

This scheme is used for the following variables:

- `par_end_cd`

Table 20: Parole exit classification scheme

Parole exit code	Parole exit description
CO	Completion
RN	Returned to incarceration - new sentence
RV	Returned to incarceration - revocation
RO	Returned to incarceration - other/unknown
AB	Absconded/escaped
OU	Other unsatisfactory exit
TR	Transferred to another state
DE	Death
OT	Other
UU	Not known / missing

E.2.11 Coverage type

This scheme is used for the following variables:

- `cjars_table`

Table 21: Coverage type classification scheme

Coverage type code	Coverage type description
ARR	Coverage of arrests
JAIL	Coverage of jail entries and charges
ADJ_FE	Coverage of felony court case filings
ADJ_MI	Coverage of misdemeanor court case filings
PRO	Coverage of probation spells
INC	Coverage of incarceration spells
PAR	Coverage of parole spells

E.2.12 Coverage source

This scheme is used for the following variables:

- `coverage`

Table 22: Coverage source classification scheme

Coverage source code	Coverage source description
primary	Coverage is based on data from an agency with primary responsibility for maintaining these variables
secondary	Coverage is based on data from an agency that is a user of these variables

F Notes on record linkage

Records from disparate criminal justice agencies most often lack identifiers that allow for linkage of records to an individual or across criminal justice episodes. These issues were overcome here by developing probabilistic matching algorithms that identify individuals and episodes across disparate sources of records. The two following sections describe in more detail the methods that were developed to accomplish entity resolution and episode resolution.

F.1 Entity resolution to identify individuals

A common issue in linking administrative criminal justice records across disparate sources is the absence of a unique individual identifier. As a result, it is usually necessary to turn to other information that identifies individuals, such as name and date of birth. In "big data" applications, this requires an algorithmic approach to make matching feasible.

There are two broad classes of entity resolution algorithms, deterministic and probabilistic. Deterministic algorithms focus on the variables common to two sets of data being matched. In some examples, paired observations must match on all common variables to be classified as a match. In other settings with a rich set of matching variables, multiple linkage rules are defined to allow for more flexibility in the matching process. The last class of deterministic models use an "iterative method" of rules to identify matches.

In contrast, probabilistic algorithms attempt to predict the probability that any two observations are the same identity based on the relative agreement of their matching variables. This approach has benefits over deterministic models in that it more flexibly sets a decision rule that optimizes the trade-off between making more matches and limiting false matches. A common method used in probabilistic matching is the implementation of a weighting system that places different value on each variable used to determine match status. Modern applications of this strategy employ the use of machine or supervised learning techniques to estimate match weights. For best implementation of this strategy, training data is used for algorithmic development.

Training data came in the form of records linked by biometrically validated identifiers from the Texas Department of Criminal Justice and the Harris County Court System. These sources of data have varying personally identifiable information which allowed for us to build a predictive model to match individuals based on agreement of their name and date of birth. The algorithm was trained by using blocking rules to partition the data and generate candidate pairs of all potential matches. True match status of the pairs is known based on a agency-validated identifier, which assisted with generating and refining a model to determine probability of true match status based on name and date of birth. The true match status was then used to set a threshold used to determine a statistical match status that maximizes precision and recall.

When CJARS data are delivered to the Census Bureau, the roster file includes all of the distinct realizations of individual identifying information available for every individual. These duplicate records are included to maximize the likelihood that individuals are matched to a PIK. But this also means that, upon delivery, the roster is not unique by `cjars_id` and PVS may assign different PIKs to a single `cjars_id`. To make the roster unique, we recommend taking the following steps:

1. Some rows in the roster are associated with known aliases. CJARS keeps these rows to maximize the likelihood of record linkage during PVS. Although the records are de-identified after PVS, these rows are still marked with an alias indicator flag. For each `cjars_id`, if at least one of the rows *not* known to be an alias has been successfully matched to a PIK, we recommend deleting all of the rows flagged as coming from aliases.
2. (Requires access to the Census Numident.) If there are still `cjars_id`s associated with multiple PIKs, keep the PIKs in the roster where the difference between the year of birth in the CJARS roster and the year of birth in the Census Numident is smallest.
3. (Requires access to the Census Numident.) If there are still `cjars_id`s associated with multiple PIKs, prioritize the PIKs in the roster where the first two characters in the `cjars_id` are the same as the state of birth variable in the Census Numident.
4. Finally, if there are still `cjars_id`s associated with multiple PIKs, retain one of the modal PIKs.

F.2 Episode resolution to link sequences of events

> USER NOTE
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense.

Another barrier to linking administrative criminal justice records is connecting criminal justice events to a single episode. For instance, often times there is no identifier in criminal justice data that allows for court records to be linked back to arrest records to determine which arrest led to a case being filed against an individual in criminal court. In addition, there is also often no information that allows for the linkage of court records to supervision outcomes whether that be in the community or in a secure facility.

To overcome this issue and to reconstruct the series of criminal justice events that are all connected to a single episode in the CJARS data, probabilistic matching techniques were employed for episode resolution. Similar to entity resolution, training data was used to create a model that predicts the likelihood that events associated to an individual are associated with a single criminal justice episode. One example of the training data that was used came from the Harris County Sheriff's Office which provided a case number which could be used to identify which arrest(s) was/were associated with which court case filing.

To generate the matching model, all criminal justice records were first linked to an individual and then data from the arrest and court records were merged to determine all possible combinations of events within an individual. Next, the court case number was used to determine true match status which generated the data that was necessary to train the matching model. Then a model was estimated using variables that would help to predict whether various events were related to a single episode. Some predictors included the date events occurred, similarity of offense type, number of arrests in arrest data, and number of cases filed in court data. This resulted in a model that estimated the likelihood that events were linked to an episode, which allowed a threshold to be set to determine statistical match status. It is important to note that this process was also iterated through for various linkages (court to incarceration) to estimate predictive models for these series of events as well.

G Data notes by jurisdiction

These data notes describe aspects of CJARS source data that may be associated with deviations from the CJARS schema, variable missingness, or case selection (e.g., charged caseload versus convicted caseload). They should not be considered exhaustive, and we encourage data users to contribute their own notes as they learn more about the data.

> USER NOTE
>
> Data coverage may differ for the version of the CJARS database within the VDE. Data notes listed in jurisdictions not available on the VDE are not relevant for the version of the data in the VDE environment.

G.1 National data notes

Note ID	Data provider	Variables	Notes
US001		arr_off_cd_src, jail_off_cd_src, jail_grd_cd_src, adj_grd_cd_src, adj_off_lgl_cd_src, adj_disp_cd_src, adj_sent_src, inc_fcl_cd_src, inc_entry_cd_src, inc_exit_cd_src, pro_cond_cd_src, pro_end_cd_src, par_end_cd_src	Data sourced from any type of agency often has information stored in open-field string variables. Variables that contain raw source information (i.e., *_src variables) identify the source of the information when the value of their sister coded variable (e.g., adj_grd_cd and adj_grd_cd_src) was generated from secondary source information. For example, if adj_grd_cd was coded as felony because the information was sourced from incarceration records, adj_grd_cd_src would be recorded as: PRISON RECORD.
US002		adj_sent_pro, adj_sent_inc, adj_sent_inc_min, adj_sent_inc_max, adj_sent_fine, adj_sent_rest	Data sourced from courts (or possibly secondary sources) has a variety of variables that sometimes include outliers in the raw data. Upper and/or lower limits are placed on these variables to avoid issues outliers cause. These outliers are caused by factors that cannot be identified (e.g., data quality issues).

Note ID	Data provider	Variables	Notes
US003		arr_, jail_, adj_, inc_, pro_, par_	Data sourced from any type of agency often has information about the dates that events (e.g., arrests, jail entries, court case filings) take place. Notably, sometimes records exist with illogical sequencing of events. An example would be a record of a parole spell with the end date of the spell occurring earlier in time than the begin date. Another example would be a court case that has a sentencing date that occurred prior to the filing date. It is important to note that the dates for these records are left as is because it is impossible to determine the cause of the error (e.g., data quality issue).
US004		arr_, jail_, adj_	Data sourced from agencies that include offense information are used to populate records in the arrest, jail, and adjudication tables. Multiple charges for a given criminal justice event are typically kept as unique observations. However, there may be a loss of repeat charges for the same offense description in a case when deduplication occurs.
US005		jail_	Data sourced from law enforcement and correctional agencies (e.g., local police departments, sheriff's offices, local jails) are used to populate booking records in the jail table. One method used to collect data from these sources is to download point-in-time snapshots of the jail population on a daily basis. Individuals that are booked and released from jail in a single day may not be recorded as being in the population at the point in time when source data are downloaded which can lead to missing bookings.

G.2 Alabama data notes

Note ID	Data provider	Variables	Notes
AL001	Lee County Sheriff's Office	jail_off_cd, jail_off_cd_src	Data sourced from the Lee County Sheriff's Office only includes information on the primary charge associated with a given jail booking.

G.3 Alaska data notes

Note ID	Data provider	Variables	Notes
AK001	Commercial Vendor	adj_	Data sourced from Alaska Administrative Office of the Courts includes information on court case filings. County FIPS are based on the most recent changes in 2019-2020 postcensal series.

G.4 Arizona data notes

Note ID	Data provider	Variables	Notes
AZ001	Arizona Department of Corrections, Rehabilitation and Reentry	inc_	Data sourced from the Arizona Department of Corrections, Rehabilitation and Reentry Inmate Datasearch system includes information on incarceration terms. This system does not record date of birth which is necessary for data processing purposes. However, date of birth was available through our court record holdings, which were linked to DOC records via exact match on name and court case number. Notably, because our data holdings from the AZ court system are known to have coverage gaps, see Arizona data note for adjudication variables, these gaps also impact the coverage of the incarceration records.
AZ002	Arizona Administrative Office of the Courts	adj_	Data sourced from the Arizona Administrative Office of the Courts does not include data from the following courts: Arizona Supreme Court, Court of Appeals – Division 1, Court of Appeals – Division 2, Superior Court in Maricopa County (Phoenix, non-criminal cases), Superior Court in Pima County (Tuscon), Justice of the Peace Courts in Maricopa County (all), Pima County Consolidated Justice Courts (cases that do not have outstanding debt owed to the court), Prescott Justice Court (Yavapai County), Chandler Municipal (cases that do not have outstanding debt owed to the court), Gilbert Municipal Court, Mesa Municipal Court, Paradise Valley Municipal Court, Prescott Municipal Court, and Tempe Municipal Court.

Note ID	Data provider	Variables	Notes
AZ003	Arizona Administrative Office of the Courts	adj_disp_dt_	Data sourced from the Arizona Administrative Office of the Courts includes case record information. This data provides a case disposition date and a charge disposition date. The charge disposition date is used. The case disposition date is used in the roughly 5% of observations where the charge disposition date is missing, except if the case disposition date is 4/3/2016, 7/19/2017, or 11/3/2017 as these appear to be catch all dates when charge disposition dates differ or one of the charge disposition dates is missing.

G.5 Arkansas data notes

Note ID	Data provider	Variables	Notes
AR001	Arkansas Administrative Office of the Courts	adj_	Data sourced from the Arkansas Administrative Office of the Courts Public CourtConnect system includes information on court case filings. This system provides District Court cases from only a select number of counties.
AR002	Arkansas Department of Correction	inc_	Data sourced from the Arkansas Department of Corrections includes information on the current population; however, the entry dates reflect the first date of entry by an individual and are not updated for re-entries.

G.6 California data notes

Note ID	Data provider	Variables	Notes
CA001	Los Angeles Police Department	arr_off_cd	Data sourced from the California Los Angeles Police Department website includes arrest information. This website does not provide arrest offense code for arrests prior to 2019.
CA002	Riverside County Sheriff's Office	jail_exit_dt_	Jail admissions data for Riverside County has entry coverage starting in 2001, but only includes exit dates starting in 2020.

G.7 Connecticut data notes

Note ID	Data provider	Variables	Notes
CT001	Connecticut Department of Correction	inc_, par_	Data sourced from the Connecticut Department of Corrections only includes data for inmates who were discharged between 1/1/1997 and 10/31/2022 and who have not returned to custody as of 12/6/2022.

G.8 Delaware data notes

Note ID	Data provider	Variables	Notes
DE001	Commercial Vendor	adj_, arr_	Data sourced from the Delaware Superior Court Calendars includes information on court case filings. However, it only includes cases heard by a grand jury.

G.9 Florida data notes

Note ID	Data provider	Variables	Notes
FL001	Florida Department of Corrections	inc_exit_dt_dd	Data sourced from the Florida Department of Corrections includes information on the date that inmates exit prison. There is a concentration of exits on the first of the month. This may reflect a standard release date, or it may reflect data warehouse processes that record some types of events as always occurring on the first of the month. See Blomberg et al. (2011, p. 20).

Note ID	Data provider	Variables	Notes
FL002	Florida Court Clerks and Comptrollers	adj_grd_cd, adj_sent_serv, adj_sent_dth, adj_sent_inc, adj_sent_pro, adj_sent_rest, adj_sent_sus, adj_sent_trt, adj_chrg_off_cd, adj_disp_off_cd	Data sourced from the Florida Court Clerks and Comptrollers does not have detailed sentencing information. Consequently, sentencing outcomes such as sentenced restitution, whether the sentence was suspended, if the individual was sentenced to death, and other are not observed. Data sourced from Florida courts are inconsistent in what types of charges have been provided. Some counties include Criminal Traffic, some do not. Some counties include municipal and county ordinances, some do not.
FL003	Florida Court Clerks and Comptrollers	adj_sent_fine	Data sourced from the Florida Court Clerks and Comptrollers includes information on assessed fines and fees. The CJARS data schema is designed only to report fines that are assessed, however fines and fees are both reported in an aggregated data field from this source. Therefore, fine information is only reported for charges associated with a conviction and may include fees, which are not generally reported in the CJARS schema.
FL004	Florida Department of Corrections	inc_entry_cd, pro_end_dt_, pro_end_cd, par_end_dt_, adj_file_dt_, adj_chrg_off_cd, adj_sent_dt_, adj_sent_pro, adj_sent_rest, adj_sent_sus, adj_sent_trt, adj_sent_fine, adj_sent_src, adj_sent_inc_min	Data sourced from the Florida Department of Corrections includes repeated snapshots of the current population under probation and parole supervision. Consequently, incarceration entry code, case filing dates and probation and parole exit dates are not always observed.
FL005	Miami-Dade County Clerk of Courts	pro_end_dt_	Data sourced from the Miami-Dade County Clerk of Courts includes information on probation terms. The data includes exit dates, which may be projected, as there were exit dates in the agency-provided data that extended beyond the data collection date. Agency-provided exits that do not extend beyond the date of data collection are used as the probation exit dates.

Note ID	Data provider	Variables	Notes
FL006	Pinellas County Sheriff's Office	jail_	Data sourced from the Pinellas County Sheriff's Office contains booking record information. Only a subset of records are available for July 2015, and no records are available for August through November of 2015.
FL007	Florida Court Clerks and Comptrollers	adj_	Data sourced from the Florida Court Clerks and Comptrollers includes information on cases that have received a disposition. For this reason, open case filings that remained undisposed by the creation date of the data extract are unobserved.

G.10 Georgia data notes

Note ID	Data provider	Variables	Notes
GA001	Houston County Sheriff's Office	jail_book_dt_	Data sourced from Houston County Sheriff's Office includes information on bookings. The records cover bookings between 1990 and 2020, with the exception of the year 2007 which is not covered.
GA002	Baldwin County Clerk of Superior Court	adj_sent_src	Data sourced from the Baldwin County County Clerk's Office includes two fields from which sentence information may be drawn. One is not provided in the source variable because of the presence of personally identifiable information. As such, content in the source variable may not match other sentence variables.
GA003	Bryan County Clerk of Courts	adj_sent_src	Data sourced from the Bryan County County Clerk's Office includes two fields from which sentence information may be drawn. One is not provided in the source variable because of the presence of personally identifiable information. As such, content in the source variable may not match other sentence variables.
GA004	Camden County Clerk of Superior Court	adj_sent_src	Data sourced from the Camden County County Clerk's Office includes two fields from which sentence information may be drawn. One is not provided in the source variable because of the presence of personally identifiable information. As such, content in the source variable may not match other sentence variables.

Note ID	Data provider	Variables	Notes
GA005	NYU Public Safety Lab Jail Data Initiative	jail_bail_amt	Data sourced from the Effingham County Sheriff's Office includes information on bail amounts. These bail amounts are given at the booking level and cannot be disaggregated to individual charges. All charges in a given booking are therefore given the same, booking-level total bail amount.

G.11 Idaho data notes

Note ID	Data provider	Variables	Notes
ID001	Idaho Department of Corrections	pro_	Data sourced from the Idaho Department of Corrections includes information on the terms of probation. The responsibility of probation supervision in Idaho is split between the Department of Corrections and courts. For this reason, data sourced from the Department of Corrections will only cover probationers supervised by the agency.

G.12 Illinois data notes

Note ID	Data provider	Variables	Notes
IL001	Illinois Department of Corrections	par_	Data sourced from the Illinois Department of Corrections inmate search includes information on parole terms. This data consists of repeated snapshots of the Illinois Department of Corrections making it sometimes impossible to identify exact entry and exit dates. Only inmates who are currently on parole at the time of the repeated snapshots are in this data.

Note ID	Data provider	Variables	Notes
IL002	Illinois Department of Corrections	inc_	Data sourced from the Illinois Department of Corrections includes information on terms of incarceration. Availability of data varies across years. For example, periodic snapshots of the inmate population are available since 2006. These snapshot are used to identify entries and exits, but exact entry and exit dates are sometimes impossible to determine based on the structure of the available data. However, data on exits from incarceration are available starting in 2014 while data on entries are available starting in 2018. The availability of this data makes identifying exact entry and exit dates possible.

G.13 Indiana data notes

Note ID	Data provider	Variables	Notes
IN001	Indiana Department of Correction	inc_entry_dt_, inc_exit_dt_	Data sourced from the Indiana Department of Corrections includes the entry and exit movements of prisoners from January 1, 2010 onward. On January 1, 2010, when the new offender management system was activated, there were approximately twenty thousand adult prisoners within the Indiana Department of Corrections and our data do not show entries for those prisoners.
IN002	NYU Public Safety Lab Jail Data Initiative	jail_bail_amt	Data sourced from the Madison County Sheriff's Department includes information on bail amounts. These bail amounts are given at the case level and cannot be disaggregated to individual charges. All charges in a given case are therefore given the same, case-level total bail amount.

G.14 Kentucky data notes

Note ID	Data provider	Variables	Notes
KY001	Christian County Jail	jail_bail_amt	Data sourced from the Christian County Jail includes information on bail amounts. These bail amounts are given at the booking level and cannot be disaggregated to individual charges. All charges in a given booking are therefore given the same, booking-level total bail amount.
KY002	Christian County Jail	jail_bail_amt	Data sourced from the Christian County Sheriff's Department include information on bail amounts. The bail amounts available cannot be linked to individual charges in the data. All charges in a given booking are therefore given the same, booking-level total bail amount.
KY003	Graves County Jail and Restricted Custody Center	jail_bail_amt	Data sourced from the Graves County Sheriff's Department include information on bail amounts. The bail amounts available cannot be linked to individual charges in the data. All charges in a given booking are therefore given the same, booking-level total bail amount.
KY004	Letcher County Jail	jail_bail_amt	Data sourced from the Letcher County Sheriff's Department include information on bail amounts. The bail amounts available cannot be linked to individual charges in the data. All charges in a given booking are therefore given the same, booking-level total bail amount.
KY005	Christian County Jail	jail_bail_amt	Data sourced from the Christian County Jail includes information on bail amounts. These bail amounts are given at the booking level and cannot be disaggregated to individual charges. All charges in a given booking are therefore given the same, booking-level total bail amount.

G.15 Maryland data notes

Note ID	Data provider	Variables	Notes
MD001	Maryland Judiciary	adj_	Data from the Maryland Judiciary, as sourced from the Maryland Volunteer Lawyers Services' Client Legal Utility Engine (CLUE) database, includes information on adjudication records. Cases from Prince George's County Circuit Court is excluded due to lack of sufficient linking variables.

G.16 Michigan data notes

Note ID	Data provider	Variables	Notes
MI001	Michigan Department of Corrections	pro_	Data sourced from the Michigan Department of Corrections includes information on terms of probation. The responsibility of probation supervision in Michigan is split between the Department of Corrections (felony convictions) and the district courts (non-felony convictions). For this reason, data sourced from the Department of Corrections will only cover felony probationers.

G.17 Minnesota data notes

Note ID	Data provider	Variables	Notes
MN001	Minnesota State Court Administrator's Office	adj_	Data sourced from the Minnesota State Court Administrator's Office includes information on court case filings. However, the only court case filings that are covered from this source include those that resulted in a conviction.
MN002	Minnesota State Court Administrator's Office	adj_disp_off_cd	Data sourced from the Minnesota State Court Administrator's Office includes information on court case filings. However, court cases with a criminal traffic offense as the controlling charge are not covered.
MN003	Minnesota State Court Administrator's Office	adj_grd_cd, adj_grd_cd_src	Data sourced from the Minnesota State Court Administrator's Office includes information on charge offense grade. However, between 2009 and 2015, some charges in Hennepin County and Ramsey County are missing offense grades at disposition, which may include cases that were originally charged at a criminal level but were later reduced to non-criminal offense at disposition. Records that are non-criminal at disposition fall outside of the CJARS data schema, but because of the lack of certainty caused by missing offense grades at disposition, some non-criminal offenses in Hennepin County and Ramsey County may be included in the CJARS database.

G.18 Montana data notes

Note ID	Data provider	Variables	Notes
MT001	Montana Department of Corrections	pro_	Data sourced from the Montana Department of Corrections includes information on terms of probation. Only records on terms of probation supervised by the Montana Department of Corrections are covered.
MT002	Montana Department of Corrections	inc_, par_, pro_	Data sourced from the Montana Department of Corrections includes information on terms of parole, probation and incarceration. The data only covers individuals that began a parole, probation or incarceration term after the start of data coverage (see data coverage file). Therefore, terms of supervision for individuals that were under supervision as of the start of coverage will not be observed in the data.

G.19 Nebraska data notes

Note ID	Data provider	Variables	Notes
NE001	Nebraska Department of Correctional Services	adj_grd_cd, adj_grd_cd_src	Data sourced from the Nebraska Department of Correctional Services includes offense information for the offenses inmates were sentenced for. The data includes information on the grade of the offense and includes some misdemeanor records. This is indicated in the adj_rec_src_doc variable. Some misdemeanor records are expected (e.g., those in a case with a felony charge). However, some of the charges classified as misdemeanors from the offense grade information provided are associated with descriptions in adj_disp_off_cd_src that appear to be felony offenses. The original grade information is retained along the with the offense description.

Note ID	Data provider	Variables	Notes
NE002	Nebraska Department of Correctional Services	par_bgn_dt_dd	Data sourced from the Nebraska Department of Correctional Services includes information on parole terms. The distribution of the beginning of parole terms is not evenly distributed across the days of the month. Specifically, more individuals begin parole during the latter part of each month, which causes the distribution to be left skewed. This is caused by the timing of Nebraska's parole board hearings which take place during the last two weeks of each month.
NE003	Nebraska Department of Correctional Services	par_	Data sourced from the Nebraska Department of Correctional Services includes information on parole terms for those under the agency's supervision. It does not include information on post-release supervision terms, which are overseen by the Administrative Office of the Courts and Probation.

G.20 Nevada data notes

Note ID	Data provider	Variables	Notes
NV001	Nevada Department of Corrections	inc_entry_dt_, inc_exit_dt_	Data sourced from the Nevada Department of Corrections include information on prison terms. Only the initial prison entry and exit for a given sentence are included; subsequent re-entries (e.g. parole revocations) are not available.
NV002	City of Las Vegas Municipal Court	adj_grd_cd	Data sourced from the City of Las Vegas Municipal Court does not include information on whether charges and dispositions are felonies or misdemeanors. We do have temporal coverage of filings and dispositions, but due to the lack of charge grade differentiation, we are unable to identify beginning and end dates of coverage in the CJARS data coverage table for this jurisdiction.

G.21 New Jersey data notes

Note ID	Data provider	Variables	Notes
NJ001	Superior Court of New Jersey	adj_	Data sourced from the Superior Court of New Jersey includes information on Superior Court case filings. The data only include cases filed in the Superior Court system of New Jersey. Therefore, all other cases filed in lower courts (e.g., municipal courts) are not included.
NJ002	Superior Court of New Jersey	adj_	Data sourced from the Superior Court of New Jersey includes information on case records and sentencing. Records filed starting July 1, 2018 are exclusively available at the case-level. Records filed prior to July 1, 2018 include information at the charge-level for each case, with the exception of sentencing information. Sentencing information is only recorded at the case-level and cannot be disaggregated to the individual charges. All charges in a given case are therefore given the same, case-level total sentencing values.
NJ003	Superior Court of New Jersey	adj_disp_dt_	Data sourced from the Superior Court of New Jersey includes information on dispostions. Coverage of disposition dates is only available for cases disposed prior to July 1, 2018.
NJ004	New Jersey Department of Corrections	par_	Data sourced from the New Jersey Department of Corrections includes information on parole. This data only includes a snapshot of people entering parole who were released from prison in the previous year.
NJ005	Bergen County Sheriff's Office	jail_exit_dt_	Data sourced from Bergen County Sheriff's Office includes information on jail admissions. Coverage starts mid-1993, but only includes exit dates starting in 2020.

G.22 New York data notes

Note ID	Data provider	Variables	Notes
NY001	New York State Department of Corrections and Community Supervision	par_	Data sourced from the New York State Department of Corrections and Community Supervision provides parole data. This data only consists of the most recent parole spell for an individual.

G.23 North Carolina data notes

Note ID	Data provider	Variables	Notes
NC001	North Carolina Department of Public Safety	par_end_dt_	Data sourced from the North Carolina Department of Public Safety includes information on parole terms. The data includes exit dates, which may be projected, as there were exit dates in the agency-provided data that extended beyond the data collection date. Agency-provided exits that do not extend beyond the date of data collection are used as the parole exit dates.
NC002	North Carolina Administrative Office of the Courts	adj_	Data sourced from the North Carolina Administrative Office of the Courts contains records on court case filings. Some of the records included in the CJARS repository from this agency were obtained from open access data that was submitted as replication data for Silveira (2017). These court data exclude records where the method of disposition included: dismissal by deferred prosecution, dismissal with leave by district attorney, or dismissal without leave by district attorney.
NC003	North Carolina Administrative Office of the Courts	adj_disp_off_cd, adj_disp_off_cd_src	Data sourced from the North Carolina Administrative Office of the Courts contains descriptions of disposition offenses. These descriptions are generally provided for convictions, but not for charges without a conviction.
NC004	North Carolina Administrative Office of the Courts,North Carolina Department of Public Safety	adj_	Data sourced from the North Carolina Administrative Office of the Courts and North Carolina Department of Public Safety contain records on court case record information. Data from the Department of Public Safety goes back further in time but is limited to coverage of court commitments for individuals under the supervision of the Department of Public Safety. Data from the North Carolina Administrative Office of the Courts does not go as far back in time but covers publicly available records of court case filings.

G.24 Oklahoma data notes

Note ID	Data provider	Variables	Notes
OK001	Oklahoma Department of Corrections	inc_entry_dt_, inc_exit_dt_	Data sourced from the Oklahoma Department of Corrections includes information on prison entries and exits. Entry and exit dates are provided separately and are resolved into unique spells during the harmonization process. Spells prior to the designated start of coverage in 2000 are included, but spell resolution in earlier years is less accurate due to data quality issues. Spell resolution is more accurate after 2000, but some unaligned entries and exits remain.
OK002	NYU Public Safety Lab Jail Data Initiative	jail_bail_amt	Data sourced from the Oklahoma County Sheriff's Office includes information on bail amounts. These bail amounts are given at the case level and cannot be disaggregated to individual charges. All charges in a given case are therefore given the same, case-level total bail amount.

G.25 Oregon data notes

Note ID	Data provider	Variables	Notes
OR001	Oregon Judicial Department	adj_	Data sourced from the Oregon Judicial Department includes information on case records and sentencing. Coverage of sentencing information is available starting on January 1, 2000. Coverage of case filings goes back further in time.
OR002	Columbia County Sheriff's Office	jail_bail_amt	Data sourced from the Columbia County Sheriff's Department includesinformation on bail amounts. These bail amounts are given at the booking level and cannot be disaggregated to individual charges. All charges in a given booking are therefore given the same, booking-level total bail amount.

G.26 Pennsylvania data notes

Note ID	Data provider	Variables	Notes
PA001	Administrative Office of Pennsylvania Courts	adj_sent_dt_, adj_sent_serv, adj_sent_dth, adj_sent_inc, adj_sent_pro, adj_sent_rest, adj_sent_sus, adj_sent_trt, adj_sent_fine, adj_sent_inc_min, adj_sent_inc_max	Data sourced from the Administrative Office of Pennsylvania Courts includes information on court case filings. However, this data only includes sentencing information for 2020 and later. Case information earlier than 2020 is limited to offense type, offense date, file date, and disposition information.
PA002	Administrative Office of Pennsylvania Courts	adj_disp_dt_	Data sourced from the Administrative Office of Pennsylvania Courts includes information on court case filings. There is a known drop in case dispositions in 2019 that is an artifact of data collection limitations.
PA003	Pennsylvania Department of Corrections	par_	Data sourced from the Pennsylvania Department of Corrections includes information on parolees. The extract of data includes the population of parolees in October of 2022. However, due to insufficiant information on individuals, we are only able to process a portion of the records.

G.27 Rhode Island data notes

Note ID	Data provider	Variables	Notes
RI001	Rhode Island Department of Corrections	inc_	Data sourced from the Rhode Island Department of Corrections includes information on terms of incarceration. Only individuals released from prison during the period of 07/01/2021 and 06/30/2022 are included.

G.28 South Carolina data notes

Note ID	Data provider	Variables	Notes
SC001	South Carolina Judicial Branch	adj_	Data sourced from the South Carolina Judicial Branch does not include Colleton county.

G.29 Texas data notes

Note ID	Data provider	Variables	Notes
TX001	Texas Department of Criminal Justice	par_end_dt_	Data sourced from the Texas Department of Criminal Justice includes information on parole terms. The data does not include descriptions for parole status codes. Entries are able to be approximated using sentencing information but exits are harder to pinpoint resulting in a large share of unknown exit dates.
TX002	Texas Department of Criminal Justice	pro_end_dt_	Data sourced from the Texas Department of Criminal Justice includes information on probation terms. The data does not include actual exit dates, but does include projected dates based on the term length. Projected exits that do not extend beyond the date of data collection are used as probation exit dates.
TX003	iDocket	adj_	Data sourced from iDocket includes information on case records provided individually by counties in Texas. Data quality and coverage varies by county. Variables often have higher rates of UU codes and sentencing information is often missing.

G.30 Vermont data notes

Note ID	Data provider	Variables	Notes
VT001	Vermont Department of Corrections	inc_	Data sourced from the Vermont Department of Corrections includes information on terms of incarceration. A shift in Offender Management Systems caused an error where entries and exits occurring prior to 2016 are not recorded reliably. Entries and exits from 2016 forward are accurately recorded.

Note ID	Data provider	Variables	Notes
VT002	Vermont Department of Corrections	inc_	Data sourced from the Vermont Department of Corrections includes information on terms of incarceration. Vermont Department of Corrections has a unified jail and prison system. Incarceration types were not identified in the data provided, therefore jail and prison observations are combined.

G.31 Virginia data notes

Note ID	Data provider	Variables	Notes
VA001	Fairfax Circuit Court	adj_off_dt_, adj_sent_inc, adj_sent_pro	Data sourced from the Fairfax Circuit Court is missing offense dates. Filing dates, disposition dates and sentencing dates, however, are available. Moreover, more than 50% of the sentenced incarceration length and probation length information are missing in the data.
VA002	Judiciary of Virginia	adj_sent_dt_, adj_disp_dt_	Data sourced from the Judiciary of Virginia includes information on court case filings. The data does not include disposition or sentencing dates, and only includes information on the last available finalized court hearings. The last available hearing is used as disposition dates for charges without sentencing information and as a sentencing date for charges with sentencing information. As a result, no charge will have both disposition and sentencing dates. Only charges that do not have a finalized hearing will be missing both disposition and sentencing dates.

G.32 Washington data notes

Note ID	Data provider	Variables	Notes
WA001	Washington State Department of Corrections	pro_	Washington's distinctive system of community supervision does not map directly onto the standard probation and parole categories used in the CJARS schema. CJARS classifies community supervision records from the Washington State Department of Corrections as probation for data-integration purposes. Given the scope of the Department of Corrections data and the structure of community supervision in Washington, many of these records likely represent supervision following a period of incarceration.

G.33 Wisconsin data notes

Note ID	Data provider	Variables	Notes
WI001	Wisconsin Department of Corrections	inc_cnty_ori_fips	Data sourced from the Wisconsin Department of Corrections includes information on incarceration terms. County of conviction is missing for most records because it was obtained separately through the Department of Correction's website only for a small number of records.
WI002	Wisconsin Department of Corrections	inc_entry_dt_, inc_exit_dt_	Data sourced from the Wisconsin Department of Corrections includes information on incarceration terms. The start date for tracking entries is January of 1990. However, the distribution of exits suggests that we are not observing all exits starting at this time, but rather, the exits of inmates that had been admitted to prison since January 1990.
WI003	Wisconsin Department of Corrections	pro_	Data sourced from the Wisconsin Department of Corrections includes information on probation terms. The agency holds point-in-time snapshots of the probation population, but these data do not directly capture admission or release information. Therefore, the agency addresses this shortcoming by combining information from population snapshots and court case information to generate estimated admission and release information. This causes limitations prior to 2010 for Wisconsin probation data.

G.34 Wyoming data notes

Note ID	Data provider	Variables	Notes
WY001	Wyoming Department of Corrections	inc_, par_, pro_	Data sourced from the Wyoming Department of Corrections includes information on terms of parole, probation and incarceration. The data only covers individuals that began a parole, probation or incarceration term after the start of data coverage (see data coverage file). Therefore, terms of supervision for individuals that were under supervision as of the start of coverage will not be observed in the data.

H Data disclaimers from agency data providers

Text below is included at the request of data providers or agency partners and is reproduced exactly as provided to CJARS. Please remember that CJARS is strictly a de-identified research data resource and nothing stated below should be interpreted as permission to attempt re-identification of any individuals or records included therein. Attempts at re-identification are strictly prohibited. Additionally, none of the information in the CJARS database should be considered official records, but rather a data product that must only be used for research and statistical reporting.

H.1 Oregon Judicial Department

The data or information provided from the Extract(s) and Sentencing Data is not the official records of the court. Data provided from the Extract(s) and Sentencing Data may not reflect the most current disposition activity.

Users should verify the data and information by consulting the official court record maintained by the court in question. The official custodian of all official circuit and appellate court records in Oregon is the clerk or court administrator of the respective circuit and appellate court. The official court records are the OECI and ACMS data repository systems themselves.

H.2 Indiana Office of Judicial Administration

The data or information provided is based on information obtained from Indiana Courts in 2024. It is the user's responsibility to verify the information by personally consulting the official record maintained by the Indiana Court in question. Indiana Office of Judicial Administration and the Indiana Courts and Clerks of Court: 1) Do not warrant that the information is accurate or complete; 2) Make no representations regarding the identity of any persons whose names appear in the information; and 3) Disclaim any liability for any damages resulting from the release or use of the information.

H.3 Minnesota State Court Administrator's Office

VERIFICATION. User understands and agrees that the Records are provided "as is" as of the preparation date indicated by the State and that recent entries made by court staff may not be immediately reflected in the Records. The STATE does not assume any liability for inaccurate or delayed data, errors, or omissions and User relieves the State and the Minnesota Judicial Branch from any and all such liability. User assumes all risk and liability for verification, use and misuse of the data. User may verify the accuracy and public status of the data by logging in to the State's public access portal at www.mncourts.gov or visiting a public access terminal at any State courthouse.

USE LIMITS. The Records shall not be used in place of a criminal history background check which is available from the State of Minnesota Bureau of Criminal Apprehension.

I Missing data

The tables in this section are included to provide details about missing variables by data table and state (or county for the arrest and jail tables). A variable listed in the tables in this section means that the information for that variable is completely missing for that variable in that state (or county) because we were not able to collect that information from the data provider. Please note that partially missing information is beyond the scope of this section. Rather, the goal of these tables is to provide researchers with a resource to understand instances when variables are 100% missing in a state.

The use of an asterisk on variables means that all variables that begin with that string of characters are missing (e.g., inc_exit_dt_* covers: inc_exit_dt_yyyy, inc_exit_dt_mm, and inc_exit_dt_dd).

> USER NOTE
>
> Data coverage may differ for the version of the CJARS database within the VDE. Missing information listed in jurisdictions not available on the VDE are not relevant for the version of the data on the VDE.

I.1 List of missing variables by CJARS table, state, and county

I.1.1 Missing Arrest variables

State FIPS	County FIPS	Missing Variables
6	37	No variables missing
6	59	No variables missing
6	71	No variables missing
6	73	No variables missing
6	77	No variables missing
6	113	No variables missing
12	17	No variables missing
48	439	No variables missing

I.1.2 Missing Jail variables

State FIPS	County FIPS	Missing Variables
1	73	No variables missing
1	81	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
4	13	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
4	25	jail_bail_amt, jail_cnty_ori_fips, jail_dv_off, jail_exit_dt_*, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips
4	27	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
5	45	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
5	51	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
5	69	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips

State FIPS	County FIPS	Missing Variables
5	115	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
5	125	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
5	131	No variables missing
5	143	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
6	7	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
6	13	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
6	19	No variables missing
6	29	jail_cnty_ori_fips, jail_st_ori_fips
6	31	jail_bail_amt, jail_exit_dt_*
6	41	No variables missing
6	45	No variables missing
6	53	jail_cnty_ori_fips, jail_exit_dt_*, jail_st_ori_fips
6	55	jail_grd_cd_src, jail_off_cd_src
6	57	No variables missing
6	61	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
6	65	No variables missing
6	71	jail_bail_amt, jail_grd_cd_src
6	79	jail_cnty_ori_fips, jail_exit_dt_*, jail_st_ori_fips
6	87	No variables missing
6	89	jail_exit_dt_*
6	95	jail_exit_dt_*
6	99	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
6	107	jail_exit_dt_*
6	111	No variables missing
8	1	jail_grd_cd_src
8	13	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
8	101	jail_bail_amt
8	123	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
12	11	jail_bail_amt, jail_exit_dt_*, jail_grd_cd_src
12	23	No variables missing
12	49	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
12	57	jail_bail_amt, jail_exit_dt_*, jail_grd_cd_src
12	69	No variables missing
12	71	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
12	85	jail_bail_amt, jail_exit_dt_*, jail_grd_cd_src
12	97	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
12	103	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
12	105	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
12	111	jail_grd_cd_src
12	115	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips

State FIPS	County FIPS	Missing Variables
12	127	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
12	131	No variables missing
13	7	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	11	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	57	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	59	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	67	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	103	No variables missing
13	135	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	153	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	247	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
13	285	No variables missing
17	91	jail_grd_cd_src
18	95	jail_grd_cd_src
18	97	jail_bail_amt, jail_cnty_ori_fips, jail_st_ori_fips
20	91	jail_bail_amt, jail_cnty_ori_fips, jail_st_ori_fips
20	177	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
21	15	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
21	47	No variables missing
21	67	jail_exit_dt_*, jail_grd_cd_src
21	83	No variables missing
21	85	jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
21	117	jail_bail_amt, jail_grd_cd_src
21	133	jail_exit_dt_*
21	177	No variables missing
22	19	jail_bail_amt, jail_cnty_ori_fips, jail_dv_off, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips, jail_wpn_off
22	27	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips
22	41	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips
22	57	jail_cnty_ori_fips, jail_st_ori_fips
22	63	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
22	75	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips
23	5	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
26	21	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
26	163	jail_cnty_ori_fips, jail_st_ori_fips
27	119	jail_bail_amt, jail_cnty_ori_fips, jail_st_ori_fips

State FIPS	County FIPS	Missing Variables
28	33	jail_bail_amt, jail_grd_cd_src
28	89	jail_grd_cd_src
29	21	No variables missing
34	3	No variables missing
34	11	jail_bail_amt
34	13	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
34	29	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
34	31	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
34	33	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
35	45	No variables missing
37	119	jail_bail_amt, jail_exit_dt_*, jail_grd_cd_src
39	3	No variables missing
39	25	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
39	49	jail_cnty_ori_fips, jail_st_ori_fips
39	63	jail_bail_amt
39	93	No variables missing
39	99	jail_grd_cd_src
40	19	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
40	31	No variables missing
40	109	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
40	125	jail_bail_amt
41	5	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
41	7	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_st_ori_fips
41	9	jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
41	29	jail_exit_dt_*, jail_grd_cd_src
41	35	No variables missing
41	39	jail_exit_dt_*
41	47	jail_exit_dt_*
41	51	No variables missing
41	59	jail_exit_dt_*
41	67	jail_bail_amt, jail_exit_dt_*
41	71	jail_bail_amt, jail_cnty_ori_fips, jail_st_ori_fips
42	59	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
42	71	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips
42	129	jail_bail_amt, jail_cnty_ori_fips, jail_grd_cd_src, jail_off_cd_src, jail_st_ori_fips
45	7	jail_bail_amt
45	13	jail_grd_cd_src
45	19	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
45	21	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
45	43	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips

State FIPS	County FIPS	Missing Variables
45	51	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
45	83	jail_grd_cd_src
47	93	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
47	147	jail_grd_cd_src
48	27	No variables missing
48	29	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
48	85	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
48	113	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
48	167	jail_grd_cd_src
48	183	jail_bail_amt, jail_grd_cd_src
48	201	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
48	209	jail_bail_amt, jail_cnty_ori_fips, jail_exit_dt_*, jail_grd_cd_src, jail_st_ori_fips
48	249	No variables missing
48	439	jail_bail_amt, jail_exit_dt_*, jail_grd_cd_src
51	155	jail_bail_amt, jail_grd_cd_src
53	5	jail_bail_amt
53	33	jail_cnty_ori_fips, jail_grd_cd_src, jail_st_ori_fips
53	67	No variables missing
55	59	jail_bail_amt, jail_exit_dt_*, jail_grd_cd_src

I.1.3 Missing Adjudication variables

State FIPS	Missing Variables
4	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_dt_*, adj_sent_fine, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
5	adj_sent_inc_max, adj_sent_inc_min
12	adj_sent_inc_max, adj_sent_inc_min
13	No variables missing
18	adj_sent_fine, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
19	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_dt_*, adj_sent_fine, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
24	adj_sent_inc_max, adj_sent_inc_min, adj_sent_serv, adj_sent_trt
27	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_fine, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
29	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_cnty_juris_fips, adj_file_dt_*, adj_grd_cd, adj_grd_cd_src, adj_sent_dt_*, adj_sent_fine, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
32	adj_sent_dt_*, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
34	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_inc_max, adj_sent_inc_min
37	adj_sent_trt
38	adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt

State FIPS	Missing Variables
39	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_fine, adj_sent_rest, adj_sent_serv, adj_sent_trt
40	adj_sent_inc_max, adj_sent_inc_min, adj_sent_trt
41	adj_sent_inc_max, adj_sent_inc_min
42	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_fine, adj_sent_rest, adj_sent_serv, adj_sent_trt
44	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
45	adj_sent_inc_max, adj_sent_inc_min, adj_sent_serv, adj_sent_trt
48	adj_sent_inc_max, adj_sent_inc_min
50	adj_chrg_off_cd, adj_chrg_off_cd_src, adj_sent_dt_*, adj_sent_fine, adj_sent_inc_max, adj_sent_inc_min, adj_sent_rest, adj_sent_serv, adj_sent_trt
51	adj_sent_inc_max, adj_sent_inc_min
55	adj_sent_inc_max, adj_sent_inc_min

I.1.4 Missing Probation variables

State FIPS	Missing Variables
16	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src
19	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src
21	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src, pro_end_cd, pro_end_cd_src, pro_end_dt_*
26	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src
27	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src, pro_end_cd, pro_end_cd_src, pro_end_dt_*
29	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src
30	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src
35	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src
37	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src, pro_end_cd, pro_end_cd_src, pro_end_dt_*
38	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src
48	pro_cnty_juris_fips, pro_cond_cd, pro_cond_cd_src, pro_end_cd, pro_end_cd_src
49	pro_cnty_juris_fips, pro_cnty_ori_fips
53	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src, pro_end_cd, pro_end_cd_src
55	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src
56	pro_cnty_juris_fips, pro_cnty_ori_fips, pro_cond_cd, pro_cond_cd_src, pro_st_ori_fips

I.1.5 Missing Incarceration variables

State FIPS	Missing Variables
4	inc_entry_cd, inc_entry_cd_src
5	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src, inc_exit_dt_*
6	inc_cnty_ori_fips, inc_fcl_cd, inc_fcl_cd_src
8	inc_fcl_cd, inc_fcl_cd_src

State FIPS	Missing Variables
9	inc_cnty_ori_fips, inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src
12	inc_entry_cd, inc_entry_cd_src
13	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src
16	inc_cnty_ori_fips
17	inc_fcl_cd, inc_fcl_cd_src
18	No variables missing
19	inc_cnty_ori_fips, inc_fcl_cd, inc_fcl_cd_src
20	No variables missing
21	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src, inc_exit_dt_*
26	No variables missing
27	No variables missing
28	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src, inc_exit_dt_*
29	inc_fcl_cd, inc_fcl_cd_src
30	inc_cnty_ori_fips
31	inc_entry_cd, inc_entry_cd_src
32	inc_entry_cd, inc_entry_cd_src
33	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src, inc_exit_dt_*
34	No variables missing
35	No variables missing
36	inc_entry_cd, inc_entry_cd_src
37	No variables missing
38	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src
39	inc_cnty_ori_fips, inc_entry_cd, inc_entry_cd_src, inc_exit_dt_*
40	No variables missing
41	inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src
42	inc_fcl_cd
44	inc_cnty_ori_fips
48	No variables missing
49	inc_cnty_ori_fips, inc_entry_cd, inc_fcl_cd, inc_fcl_cd_src
50	inc_cnty_ori_fips, inc_entry_cd, inc_entry_cd_src, inc_exit_cd, inc_exit_cd_src, inc_fcl_cd
53	inc_cnty_ori_fips, inc_fcl_cd, inc_fcl_cd_src
55	No variables missing

I.1.6 Missing Parole variables

State FIPS	Missing Variables
6	par_cnty_ori_fips
8	par_cnty_ori_fips
9	par_cnty_ori_fips, par_end_cd, par_end_cd_src
12	par_end_cd, par_end_cd_src
13	par_end_cd, par_end_cd_src, par_end_dt_*
16	par_cnty_ori_fips
17	par_end_cd, par_end_cd_src, par_end_dt_*
19	par_cnty_ori_fips
20	No variables missing
21	par_end_cd, par_end_cd_src, par_end_dt_*
26	No variables missing
27	No variables missing
29	No variables missing

State FIPS	Missing Variables
30	par_cnty_ori_fips
31	No variables missing
34	par_end_cd, par_end_cd_src, par_end_dt_*
35	No variables missing
36	No variables missing
37	par_end_cd, par_end_cd_src
38	par_cnty_ori_fips
40	par_end_cd, par_end_cd_src, par_end_dt_*
42	par_end_cd, par_end_cd_src, par_end_dt_*
44	par_cnty_ori_fips, par_end_cd, par_end_cd_src, par_end_dt_*
48	No variables missing
49	par_cnty_ori_fips
56	par_cnty_ori_fips, par_st_ori_fips

Bibliography

Abowd, John M., Joelle Abramowitz, Margaret C. Levenstein, Kristin McCue, Dhiren Patki, Trivellore Raghunathan, Ann M. Rodgers, Matthew D. Shapiro, and Nada Wasi. 2019. *Optimal Probabilistic Record Linkage: Best Practice for Linking Employers in Survey and Administrative Data.* Working Papers 19-08. Center for Economic Studies, U.S. Census Bureau. https://ideas.repec.org/p/cen/wpaper/19-08.html.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2017. *How Well Do Automated Linking Methods Perform? Lessons from U.S. Historical Data.* Working Paper, Working Paper Series 24019. National Bureau of Economic Research. http://www.nber.org/papers/w24019.

Belin, Thomas R., and Donald B. Rubin. 1995. A Method for Calibrating False-Match Rates in Record Linkage. *Journal of the American Statistical Association* 90 (430): 694–707. https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476563.

Blomberg, Thomas G., Jim Clark, Leslie Hill, Bill Bales, and Karen Mann. 2011. Correctional Operations Trend Analysis System (COTAS): An Independent Validation. Report prepared for the Florida Department of Corrections. https://criminology.fsu.edu/wp-content/uploads/Correctional-Operations-Trend-Analysis-System-An-Independent-Validation.pdf.

Bonczar, Thomas P., and Allen J. Beck. 1997. Lifetime Likelihood of Going to State or Federal Prison. Bureau of Justice Statistics, NCJ 160092. https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=160092.

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Christen, Peter. 2006. *A Comparison of Personal Name Matching: Techniques and Practical Issues.* Technical Report, Joint Computer Science Technical Report Series 06-02. The Australian National University.

Coffin, Marie, and Shashikala Sukhatme. 1997. Receiver Operating Characteristic Studies and Measurement Errors. *Biometrics* 53 (3): 823–37. http://www.jstor.org/stable/2533545.

Doherr, Thorsten. 2018. *BRAIN: Stata module to provide neural network.* Boston College Department of Economics.

Feigenbaum, James J. 2016. *A Machine Learning Approach to Census Record Linking.* Technical report. http://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf?m=1423080976.

Fellegi, Ivan P., and Alan B. Sunter. 1969. A Theory for Record Linkage. *Journal of the American Statistical Association* 64 (328): 1183–210. https://doi.org/10.1080/01621459.1969.10501049.

Ferrie, Joseph P. 1996. A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules. Last updated - 2013-02-23, *Historical Methods* 29 (4): 141. http://proxy.lib.umich.edu/login?url=https://search.proquest.com/docview/1300162525?accountid=14667.

Finlay, Keith, and Michael Mueller-Smith. 2026. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. https://cjars.org.

Fortini, Marco, Brunero Liseo, Alessandra Nuccitelli, and Mauro Scanu. 2001. On Bayesian Record Linkage. *Research in official statistics* 4:185–98.

Guenther, Nick, and Matthias Schonlau. 2016. Support Vector Machines. *The Stata Journal* 16 (4): 917–37.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2016. The Elements of Statistical Learning, Springer Series in statistics.

Hof, Michel H., Anita C. Ravelli, and Aeilko H. Zwinderman. 2017. A Probabilistic Record Linkage Model for Survival Data. *Journal of the American Statistical Association* 112 (520): 1504–15. https://doi.org/10.1080/01621459.2017.1311262.

Jaro, Matthew A. 1989. Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida. *Journal of the American Statistical Association* 84 (406): 414–20. https://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478785.

Kaplan, Andee, Brenda Betancourt, and Rebecca C. Steorts. 2018. Posterior Prototyping: Bridging the Gap between Bayesian Record Linkage and Regression. *arXiv e-prints,* arXiv:1810.01538: arXiv:1810.01538.

Lahiri, P, and Michael D Larsen. 2005. Regression Analysis With Linked Data. *Journal of the American Statistical Association* 100 (469): 222–30. https://doi.org/10.1198/016214504000001277.

Lait, A.J., and B. Randell. 1996. *An Assessment of Name Matching Algorithms.* Technical Report, Department of Computing Science Technical Report Series 550. University of Newcastle upon Tyne.

Layne, Mary, Deborah Wagner, and Cynthia Rothhaas. 2014. Estimating Record Linkage False Match Rate for the Person Identification Validation System. Center for Administrative Records Research and Applications Working Paper Series 2014-02. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-02.pdf.

National Academies of Sciences, Engineering, and Medicine. 2016. *Modernizing Crime Statistics: Report 1: Defining and Classifying Crime.* Washington, DC: National Academies Press. https://doi.org/10.17226/23492.

———. 2018. *Modernizing Crime Statistics: Report 2: New Systems for Measuring Crime.* Washington, DC: National Academies Press. https://doi.org/10.17226/25035.

Neal, Derek, and Armin Rick. 2016. The Prison Boom and Sentencing Policy. *Journal of Legal Studies* 45 (1): 1–41. https://doi.org/10.1086/684310.

Newcombe, H. B., J. M. Kennedy, S. J. Axford, and A. P. James. 1959. Automatic Linkage of Vital Records. *Science* 130 (3381): 954–59. http://www.jstor.org/stable/1756667.

NYU Public Safety Lab. 2026. *Jail Data Initiative (JDI) [dataset].* New York, NY: New York University. https://jaildatainitiative.org.

Pérez, Aritz, Pedro Larrañaga, and Iñaki Inza. 2009. Bayesian classifiers based on kernel density estimation: Flexible classifiers. Special Section on The Imprecise Dirichlet Model and Special Section on Bayesian Robustness (Issues in Imprecise Probability), *International Journal of Approximate Reasoning* 50 (2): 341–62. http://www.sciencedirect.com/science/article/pii/S0888613X08001400.

Sadinle, Mauricio. 2018. Bayesian propagation of record linkage uncertainty into population size estimation of human rights violations. *The Annals of Applied Statistics* 12 (2): 1013–38. http://dx.doi.org/10.1214/18-AOAS1178.

Sadinle, Mauricio, and Stephen E. Fienberg. 2013. A Generalized Fellegi-Sunter Framework for Multiple Record Linkage With Application to Homicide Record Systems. *Journal of the American Statistical Association* 108 (502): 385–97. https://doi.org/10.1080/01621459.2012.757231.

Sayers, Adrian, Yoav Ben-Shlomo, Ashley W Blom, and Fiona Steele. 2015. Probabilistic record linkage. *International Journal of Epidemiology* 45 (3): 954–64. https://doi.org/10.1093/ije/dyv322.

Schonlau, Matthias. 2019. *RFOREST: Stata module to implement Random Forest algorithm.* Boston College Department of Economics.

Setoguchi, Soko, Yin Zhu, Jessica Jalbert, Lauren A. Williams, and Chih-Ying Chen. 2014. Validity of Deterministic Record Linkage Using Multiple Indirect Personal Identifiers: Linking a Large Registry to Claims Data. *Circulation: Cardiovascular Quality & outcomes* 7 (3): 475–80.

Silveira, Bernardo S. 2017. Bargaining with Asymmetric Information: An Empirical Study of Plea Negotiations. *Econometrica* 85 (2): 419–52. https://doi.org/10.3982/ECTA12974.

Steorts, Rebecca C., Rob Hall, and Stephen E. Fienberg. 2016. A Bayesian Approach to Graphical Record Linkage and Deduplication. *Journal of the American Statistical Association* 111 (516): 1660–72. https://doi.org/10.1080/01621459.2015.1105807.

Tahamont, Sarah, Zubin Jelveh, Aaron Chalfin, Shi Yan, and Benjamin Hansen. 2019. *Administrative Data Linking and Statistical Power Problems in Randomized Experiments.* Working Paper, Working Paper Series 25657. National Bureau of Economic Research. http://www.nber.org/papers/w25657.

Tancredi, Andrea, and Brunero Liseo. 2011. A hierarchical Bayesian approach to record linkage and population size problems. *The Annals of Applied Statistics* 5 (2B): 1553–85. https://doi.org/10.1214/10-AOAS447.

Tran, Khoi-Nguyen, Dinusha Vatsalan, and Peter Christen. 2013. GeCo: An Online Personal Data Generator and Corruptor. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*, 2473–76. CIKM '13. San Francisco, California, USA: ACM. http://doi.acm.org/10.1145/2505515.2508207.

Tromp, Miranda, Anita C. Ravelli, Gouke J. Bonsel, Arie Hasman, and Johannes B. Reitsma. 2011. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *Journal of Clinical Epidemiology* 64 (5): 565–72. http://www.sciencedirect.com/science/article/pii/S0895435610002258.

Vick, Rebecca, and Lap Huynh. 2011. The Effects of Standardizing Names for Record Linkage: Evidence from the United States and Norway. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 44 (1): 15–24. https://doi.org/10.1080/01615440.2010.514849.

Wagner, Deborah, and Mary Layne. 2014. The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software. Center for Administrative Records Research and Applications Working Paper Series 2014-01. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf.

Winkler, William E. 1990. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage.

———. 2006. *Overview of Record Linkage and Current Research Directions.* Working Paper, Research Report Series #2006-2. Statistical Research Division, U.S. Census Bureau.

———. 2014. Matching and record linkage. *Wiley Interdisciplinary Reviews: Computational Statistics* 6 (5): 313–25. https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1317.

Yancey, William E. 2002. *BigMatch: A Program for Extracting Probable Matches from a Large File for Record Linkage.* Working Paper, Research Report Series #2002-1. Statistical Research Division, U.S. Census Bureau.

Zhu, Ying, Yutaka Matsuyama, Yasuo Ohashi, and Soko Setoguchi. 2015. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics* 56:80–86. http://www.sciencedirect.com/science/article/pii/S1532046415000921.

Index

access to data, 10
acknowledgments, 13
adj_chrg_off_cd, 82
adj_chrg_off_cd_src, 82
adj_cnty_juris_fips, 111
adj_disp_cd, 87
adj_disp_cd_src, 88
adj_disp_dt_dd, 86
adj_disp_dt_mm, 85
adj_disp_dt_yyyy, 83
adj_disp_off_cd, 88
adj_disp_off_cd_src, 89
adj_dv_off, 89
adj_file_dt_dd, 81
adj_file_dt_mm, 80
adj_file_dt_yyyy, 78
adj_grd_cd, 75
adj_grd_cd_src, 76
adj_id, 24, 75
adj_off_dt_dd, 93
adj_off_dt_mm, 92
adj_off_dt_yyyy, 91
adj_off_lgl_cd, 77
adj_off_lgl_cd_src, 77
adj_rec_src_arr, 112
adj_rec_src_bkg, 111
adj_rec_src_cc, 114
adj_rec_src_crt, 112
adj_rec_src_doc, 113
adj_rec_src_rep, 113
adj_sent_dt_dd, 97
adj_sent_dt_mm, 96
adj_sent_dt_yyyy, 94
adj_sent_dth, 99
adj_sent_fine, 105
adj_sent_inc, 100
adj_sent_inc_max, 107
adj_sent_inc_min, 106
adj_sent_pro, 101
adj_sent_rest, 102
adj_sent_serv, 98
adj_sent_src, 108
adj_sent_sus, 103
adj_sent_trt, 104
adj_st_juris_fips, 109
adj_wpn_off, 90
adjudication, 74
adjudication process, 198
Alabama
    data notes, 219
Alaska
    data notes, 219
All states
    data notes, 218
Arizona
    data notes, 220
Arkansas
    data notes, 221
arr_cnty_juris_fips, 48
arr_dt_dd, 42
arr_dt_mm, 41
arr_dt_yyyy, 40
arr_dv_off, 44
arr_id, 24, 39
jail_id, 24
arr_off_cd, 43
arr_off_cd_src, 43
arr_rec_src_arr, 49
arr_rec_src_bkg, 48
arr_rec_src_cc, 51
arr_rec_src_crt, 49
arr_rec_src_doc, 50
arr_rec_src_rep, 50
arr_st_juris_fips, 46
arr_wpn_off, 45
arrest, 39
arrest and booking process, 198

California
    data notes, 221
charge grade, 210
citing the project, 14
cjars_id, 22, 31, 39, 52, 74, 115, 133, 153
cjars_table, 171
cnty_fips, 171
code schemes, 202
confidentiality, 11
Connecticut
    data notes, 222
county FIPS, 203
court disposition, 211
Coverage, 194
coverage, 10, 16, 169
    geographic, 16
    procedural, 16
    temporal, 16
coverage, 174
coverage source, 214
coverage type, 214
Criminal Justice Information Services, 11

data acquisition, 15
data collection, 9

data disclaimers, 239
data integration, 21
data notes, 218
data schema, 21
data security, 11
data sources, 15
Delaware
    data notes, 222
demographics, 31
dob_yyyy, 32
dt_mm, 173
dt_yyyy, 172

entity resolution, 24
episode resolution, 25
ethnicity, 204
event dates, 194

Federal Statistical Research Data Center, 10
Florida
    data notes, 222
funders, 12

Georgia
    data notes, 224

harmonization process, 196

Idaho
    data notes, 225
Illinois
    data notes, 225
inc_cnty_ori_fips, 147
inc_entry_cd, 139
inc_entry_cd_src, 140
inc_entry_dt_dd, 138
inc_entry_dt_mm, 137
inc_entry_dt_yyyy, 136
inc_exit_cd, 144
inc_exit_cd_src, 145
inc_exit_dt_dd, 143
inc_exit_dt_mm, 142
inc_exit_dt_yyyy, 140
inc_fcl_cd, 134
inc_fcl_cd_src, 135
inc_id, 24, 134
inc_rec_src_arr, 150
inc_rec_src_bkg, 150
inc_rec_src_cc, 152
inc_rec_src_crt, 151
inc_rec_src_doc, 151
inc_rec_src_rep, 152
inc_st_juris_fips, 148
inc_st_ori_fips, 145
incarceration, 133
incarceration entry, 212
incarceration exit, 213
incarceration facility type, 213
incarceration process, 198
Indiana
    data notes, 226
IRB approval, 12

jail, 52
jail_bail_amt, 60
jail_cnty_juris_fips, 70
jail_cnty_ori_fips, 68
jail_dv_off, 62
jail_entry_dt_dd, 55
jail_entry_dt_mm, 54
jail_entry_dt_yyyy, 53
jail_exit_dt_dd, 59
jail_exit_dt_mm, 58
jail_exit_dt_yyyy, 56
jail_grd_cd, 64
jail_grd_cd_src, 65
jail_id, 52
jail_off_cd, 61
jail_off_cd_src, 61
jail_rec_src_arr, 71
jail_rec_src_bkg, 70
jail_rec_src_cc, 73
jail_rec_src_crt, 71
jail_rec_src_doc, 72
jail_rec_src_rep, 72
jail_st_juris_fips, 69
jail_st_ori_fips, 66
jail_wpn_off, 63

Kentucky
    data notes, 226

legal code, 211

Maryland
    data notes, 227
Michigan
    data notes, 228
Minnesota
    data notes, 228
missing data, 239
Montana
    data notes, 229
month, 172

Nebraska
    data notes, 229
Nevada
    data notes, 230
New Jersey

data notes, 230
New York
data notes, 231
North Carolina
data notes, 231

offense classification, 200, 205
Oklahoma
data notes, 232
Oregon
data notes, 233

par_bgn_dt_dd, 157
par_bgn_dt_mm, 156
par_bgn_dt_yyyy, 154
par_cnty_ori_fips, 164
par_end_cd, 162
par_end_cd_src, 163
par_end_dt_dd, 161
par_end_dt_mm, 160
par_end_dt_yyyy, 158
par_id, 24, 154
par_rec_src_arr, 166
par_rec_src_bkg, 166
par_rec_src_cc, 168
par_rec_src_crt, 167
par_rec_src_doc, 167
par_rec_src_rep, 168
par_st_juris_fips, 165
par_st_ori_fips, 163
parole, 153
parole exit, 214
parole process, 198
Pennsylvania
data notes, 233
primary record sources, 194
privacy, 11
pro_bgn_dt_dd, 120
pro_bgn_dt_mm, 119
pro_bgn_dt_yyyy, 118
pro_cnty_juris_fips, 129
pro_cnty_ori_fips, 128
pro_cond_cd, 116
pro_cond_cd_src, 117
pro_end_cd, 125
pro_end_cd_src, 126
pro_end_dt_dd, 124
pro_end_dt_mm, 123
pro_end_dt_yyyy, 121
pro_id, 24, 116
pro_rec_src_arr, 130
pro_rec_src_bkg, 130
pro_rec_src_cc, 132
pro_rec_src_crt, 131
pro_rec_src_doc, 131
pro_rec_src_rep, 132
pro_st_juris_fips, 128
pro_st_ori_fips, 126
probabilistic record linkage, 24
probation, 115
probation conditions, 211
probation exit, 212
probation process, 198
project description and objectives, 9
project leadership, 12
project scope, 9
Protected Identification Key, 12

race, 204
race, 36
race_imputed, 38
race_raw, 37
Rhode Island
data notes, 234

sex, 204
sex, 33
sex_imputed, 35
sex_raw, 34
South Carolina
data notes, 234
sponsors, 12
st_fips, 169
staff, 13
state FIPS, 202

tables, 26
Texas
data notes, 235

universe, 9

Vermont
data notes, 235
Virginia
data notes, 236

Washington
data notes, 236
Wisconsin
data notes, 237
Wyoming
data notes, 238