Criminal Justice Administrative Records System (CJARS)

Keith Finlay[1] and Michael Mueller-Smith[2]

[1]U.S. Census Bureau
[2]University of Michigan

2020-06-17

Abstract

The Criminal Justice Administrative Records System (CJARS) is an integrated data repository designed to transform research and policy-making in the United States justice system. At the University of Michigan, CJARS collects longitudinal electronic records from justice-related agencies and harmonizes these records to track a criminal episode across all stages of the system. At the Census Bureau, harmonized criminal justice records can be linked at the individual-level with extensive social, demographic, and economic information from national survey and administrative data. The CJARS project is a partnership between the Census Bureau and University of Michigan, with the goal of increasing researcher access to criminal justice data.

This is a draft document. It is incomplete and may contain errors.

Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. This document meets all of the U.S. Census Bureau Disclosure Review Board (DRB) standards and has been assigned DRB approval numbers CBDRB-FY19-371 (approved 2019-06-03).

The project website is cjars.isr.umich.edu.

Data users who have questions about the nature and use of CJARS data may contact cjars-data-users@umich.edu.

Contents

List of Tables

List of Figures

1 Project description

1.1 Project description and objectives

The Criminal Justice Administrative Records System (CJARS) is an integrated data repository designed to transform research and policy-making in the United States justice system. CJARS collects longitudinal electronic records from justice-related agencies and harmonizes these records to track a criminal episode across all stages of the system. Criminal justice records can then be linked at the individual-level with extensive social, demographic, and economic information from national survey and administrative data. The CJARS data infrastructure is being built from a partnership between the U.S. Census Bureau and University of Michigan.

CJARS has the following main 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.
- Provide a platform with nationwide scalability to facilitate research in perpetuity through the Federal Statistical Research Data Center (FSRDC) network.
- Integrate criminal justice data with existing Census Bureau individual-level surveys and administrative records to facilitate novel research not previously possible.

1.2 Data collection

There is substantial variation in the way that agencies make data available for collection by outside parties. Due to this variation, CJARS pursues various methods of data collection depending on each agency it conducts outreach to. Data is collected through three different channels: (1) data use agreements, (2) public records requests, and (3) web scraping or bulk downloads.

The CJARS database relies on individual-level electronic administrative records that include process information such as the dates and outcomes of important events as well as personally identifying information (PII), such as names and dates of birth. A range of local and state entities create and maintain these records, including police departments, sheriff's offices, criminal courts, departments of corrections, and community supervision agencies. There are also some instances where criminal justice data is available through third parties that receive data from agencies.

1.3 Project scope

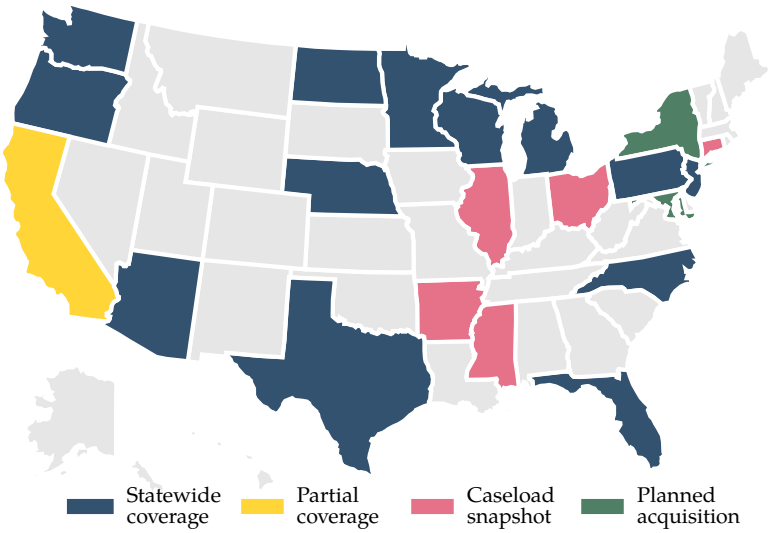
Variation also exists across states and agencies in the way that data is collected and stored. For instance, some state court agencies maintain a statewide centralized repository where all data on events that occur under the state's jurisdiction are stored. Conversely, some states have a less centralized system where data storage is the responsibility of counties, or even cities. These inter-agency differences lead to variation in the scope of data coverage across geographic region and time in CJARS.

CJARS has collected data in a number of states, which can be seen in Figure 1. Shaded states are those where CJARS has collected data in a state in at least one criminal justice domain. Using 2019 state population estimates, states with statewide CJARS coverage (in at least one criminal justice domain) constitute 48.2% of the U.S. population.

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. More detailed information on coverage within states will be described in later sections.

Finally, it is worth noting the types of criminal justice events that are captured by the CJARS data infrastructure. These include: records of arrests, criminal court case filings, and terms of probation, prison, and parole. Other events that are processed through similar institutions (i.e., courts) but do not fall under the purview of a

Figure 1: CJARS coverage summary map: data acquisition in at least one criminal justice domain



criminal justice event are excluded from the scope of the project (e.g., civil court filings). Juvenile records are also excluded from the scope of the CJARS data infrastructure.

1.4 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. The data collected, cleaned, and harmonized at the University of Michigan will be integrated into U.S. Census Bureau data systems and made anonymous and available through the Federal Statistical Research Data Center (FSRDC) network. Researchers can use the standard Census Bureau FSRDC proposal process to request use of CJARS. The data cannot be requested directly from the University of Michigan.

Distribution through the FSRDCs provides an ideal outlet. 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 have no access to sensitive PII, and can only request release of aggregate statistical material. The FSRDCs allow researchers to link CJARS records at the person-level to a wealth of socioeconomic survey and administrative data held by the Census Bureau. For more information on the FSRDC system and other available data, please visit: census.gov/fsrdc.

1.5 Data privacy

CJARS goes to great lengths to ensure the protection and security of its data, so that the identities and characteristics 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 have reviewed this system and determined that it is "substantially compliant" with all FBI CJIS policies that are applicable to an entity like CJARS. On the secure data system at the University of Michigan, PII is removed from the criminal history records at an early stage of processing. Only those individuals working on record linkage have access to the PII.

When CJARS data have been harmonized, the CJARS team at the University of Michigan transfers encrypted data files securely to the Census Bureau. 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, received data are accessible only by a small staff responsible for inventorying the contents of the data and removing sensitive PII. Next, a data linkage team uses a probabilistic linkage process to replace sensitive PII with a unique identifier called a Personal Identification Key (PIK) that can be used to link records to other databases held at the Census Bureau. More information about PIK assignment can be found in Subsection 2.4.

Finally, anonymized 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. PII is not available for researchers in order to maintain the privacy of individuals.

1.6 Principal investigators

Michael Mueller-Smith. Assistant Professor, Department of Economics, University of Michigan; Faculty Associate, Population Studies Center, University of Michigan.

Keith Finlay. Research Economist, U.S. Census Bureau.

1.7 Project funding

The CJARS project has been supported by National Science Foundation grant SES-1925563, as well as grants from the Laura and John Arnold Foundation and the Bill and Melinda Gates 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.

1.8 Acknowledgments

This project would not be possible without the hard work of a great group of people. Jordan Papp, Diana Sutton, 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 contributed to data harmonization and research. Francis Fiore and David Smith 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, and Peixin Yang have also supported the project as research assistants. Brittany Street has furthered CJARS research as a postdoctoral research fellow.

Carol Bowen, Kerri Cross, Violet Elder, and Lauren Tingwall have supported the project as research administrators. The following advisors have guided the project as part of the CJARS Board of Directors: Shawn Bushway, Jens Ludwig, Julia Lane, 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 and Katie Genadek have offered a lot of support for the project.

1.9 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. 2020. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. https://cjars.isr.umich.edu.

1.10 Documentation structure

- Section 2 provides an overview of how the CJARS project acquires data, and how it is integrated into a unified system at the University of Michigan.
- Section 3 explains the structure of the harmonized data system. Of particular interest is the variable codebook in Subsection 3.6.
- Appendices cover the project in much more detail.

- Appendix A identifies the agency sources of CJARS administrative data.
- Appendix B explains the variable harmonization process in detail.
- Appendix C lists code schemes used in the variables, including criminal classification scheme (and explains how those relate to BJS schemes).
- Appendix D explains the methods used to identify individuals and procedural episodes in the CJARS data.
- Appendix E gives an overview of how to do record linkage on Census Bureau systems using CJARS data.
- Appendix G describes a survey of state criminal history repository database designs, which provides some context for CJARS schema choices.

2 Data acquisition and integration

2.1 Mechanisms used to acquire criminal justice administrative data

CJARS utilizes a three-pronged approach to acquiring administrative criminal justice records, which include the following:

1. **Data use agreements.** Formal legal documents are signed which guide the obligations of CJARS when data is obtained from an agency.
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 is publicly available online is 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.

Table 1 shows the number of records that each of these methods has contributed to the CJARS repository.

Table 1: CJARS records acquired by acquisition method

Acquisition source	Records	Individuals
Data use agreements	1,175,100,911	8,164,073
Public records requests	261,831,295	10,330,761
Web scraping or bulk downloads	301,470,747	3,352,297

2.2 Sources of criminal justice data

Data is collected within many separate parts of the justice system. For example, 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. Data has been collected by a mixture of state, county, and local agencies. In addition, some data has also been collected by a third party with access to data from an agency.

2.3 Criminal justice data integration and harmonization process

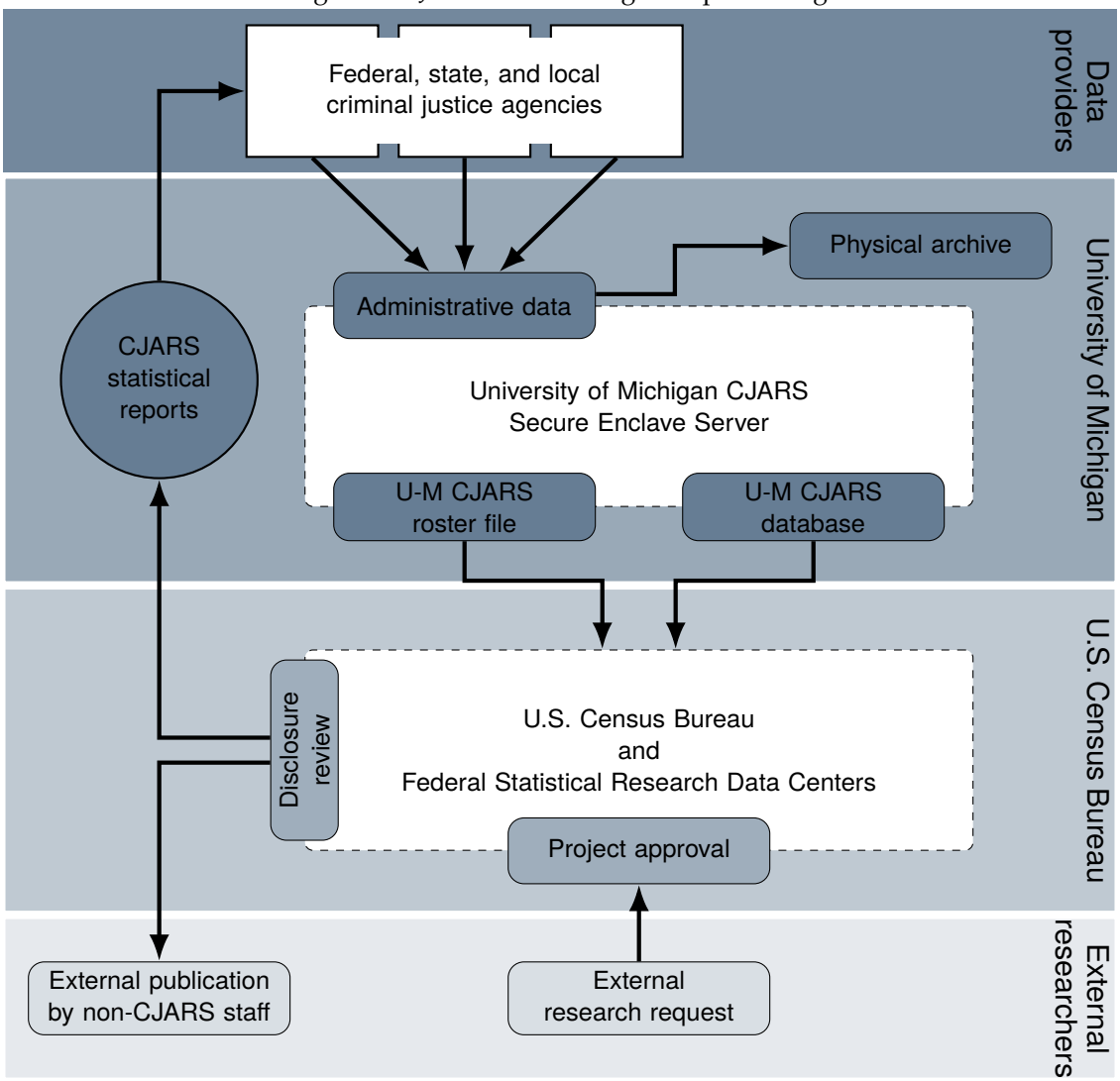
One of the major barriers to research on the criminal justice system is a lack of integration of records across agencies. For example, arrest records are rarely linkable to information that occurs at later points in the system (e.g., incarceration). CJARS implements the following steps in order to create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

- **PII data**. CJARS uses personally identifiable information to link records that all belong to the same individual. This occurs in two distinct steps. The PII data is used with two purposes. First, the data with PII information is used to find the criminal justice records that belong to the same person. Second, the PII data is used to link to demographic, economic, and government-related information held at the Census Bureau. To ensure confidentiality of individuals, all PII is segmented from other CJARS data and stored separately.

Figure 2 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).

Figure 2: CJARS data exchange and processing



2.4 Integration of criminal justice and Census Bureau records

At the Census Bureau and in the FSRDCs, CJARS data may be linked to other socioeconomic survey and administrative records using an anonymous identifier called a Personal Identification Key (PIK). Staff at the Census Bureau attempt to use all available PII to assign a PIK using a probabilistic record linkage system called the Person Identification Validation System (PVS). Once the PIK assignment process has occurred,

the anonymized files with PIKs attached are transferred to a secure computing environment that is available at the Census Bureau headquarters and in the FSRDCs. On those servers, approved data in approved projects can be linked at the person-level using the PIKs attached to each file, including the CJARS data.

Note that this process introduces a second round of entity resolution before the researcher has access to the CJARS data. This duplication has benefits and costs. The CJARS team has more control over the entity resolution process at the University of Michigan, and is able to use biometric identifiers and its own subject-matter expertise, but has no access to a population-level registry primarily based on the Social Security Agency's Numident file, which includes anyone who has ever received a Social Security Number (SSN) or an Individual Taxpayer Identification Number (ITIN). The Census Bureau entity resolution team has access to registry data not available outside of the Census Bureau, but it uses a relatively rigid process that ignores the criminal justice aspect of the data.

When researchers first use CJARS in the FSRDCs, they must decide how to resolve any discordances between CJARS IDs and PIKs. There are different types of identifier discordances:

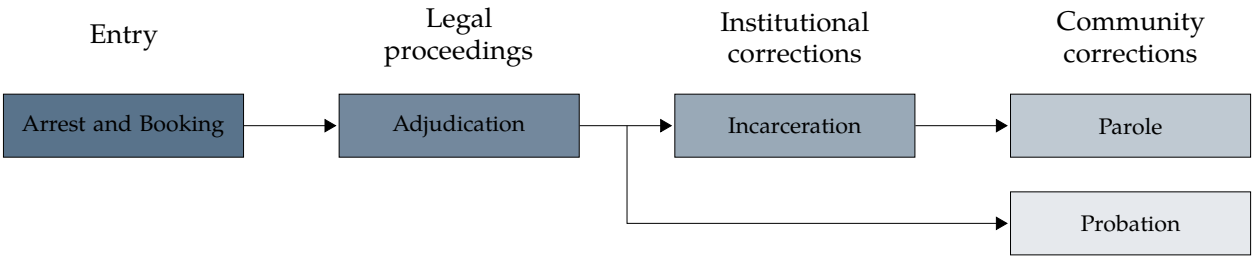
- Multiple PIKs have been linked to a single CJARS ID. In this case, there is sufficient variation in the PII within a CJARS ID, but across criminal justice records that Person Identification Validation System (PVS) has assigned different registry identities to a single CJARS ID. This could be caused by a poor linkage choice during UM entity resolution, by a poor linkage choice during PVS, or by low-quality PII for some of the records linked to a CJARS ID. The researcher can choose to use a single assigned PIK for analysis with a CJARS ID.
- Multiple CJARS IDs have been linked to a single PIK. In this case, there is sufficient similarity in the PII across CJARS IDs to allow a PIK to be assigned across different CJARS IDs. This can be caused by the same reasons as above. Here, the researcher can choose to link the associated CJARS IDs into a single identity for analysis.
- Some records within a CJARS ID have not been linked to a PIK. In this case, there are likely some CJARS records with low-quality PII, but UM entity resolution linked them together using CJ-specific information. In this case, researchers can decide which of the assigned PIKs should apply to the entire set.
- All records within a CJARS ID have not been linked to a PIK. In this case, the quality of the PII associated with a CJARS ID is too low, or the individual associated with a CJARS ID is not in the Census Bureau registry. The second case is possible if the individual is a citizen without an SSN, or an immigrant without an SSN or ITIN. There is no solution to this linkage problem. These records cannot be linked to other data at the Census Bureau.

3 Data

3.1 General description of the data

The goal for 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 3. The arrows in this figure depict the expected progression of events that make up an episode through the system.

Figure 3: CJARS national schema



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.

3.2 Data linkage

A major barrier to research on the criminal justice system is a lack of integration of data across agencies. For example, unique individual identifiers are often not present that allow linking records across different agencies to an individual. In addition, there are usually 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.

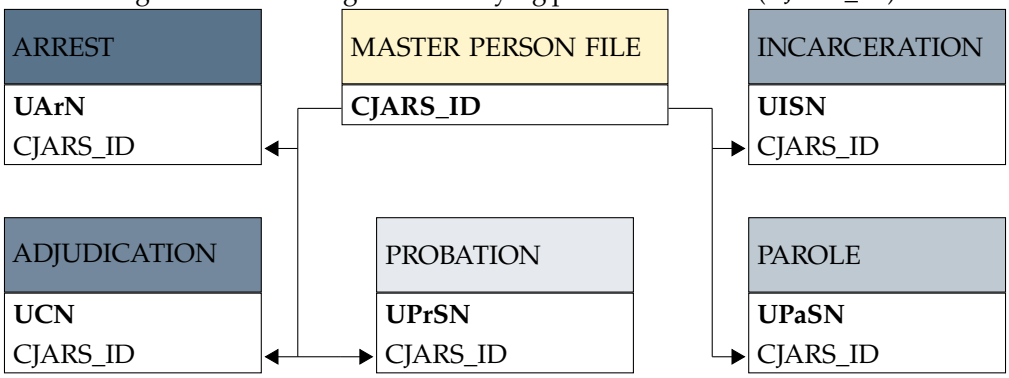
3.2.1 Data linkage via identifying personal number

CJARS has developed an algorithm that probabilistically matches records to individuals when no unique identifier is available by using names and dates of birth to identify individuals in the data. Once an individual has been identified, he or she is assigned an anonymized individual identifier (CJARS_ID). This is done so that all personally identifiable information can be removed and the data can be used in anonymized form to protect the data.

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.

The CJARS data is comprised of six separate databases. The six databases include a master person file and one database for each of the five types of events that are covered in Figure 3. Collectively, the five databases containing the criminal justice events are referred to as the CJARS relational databases. The master person file 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 UArN uniquely identifies arrests, UCN is for court filings, UPrSN is for probation events, UISN is for incarceration events, and UPaSN is for parole events. As can be seen in Figure 4, 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 master person file.

Figure 4: Data linkage via identifying personal number (CJARS_ID)



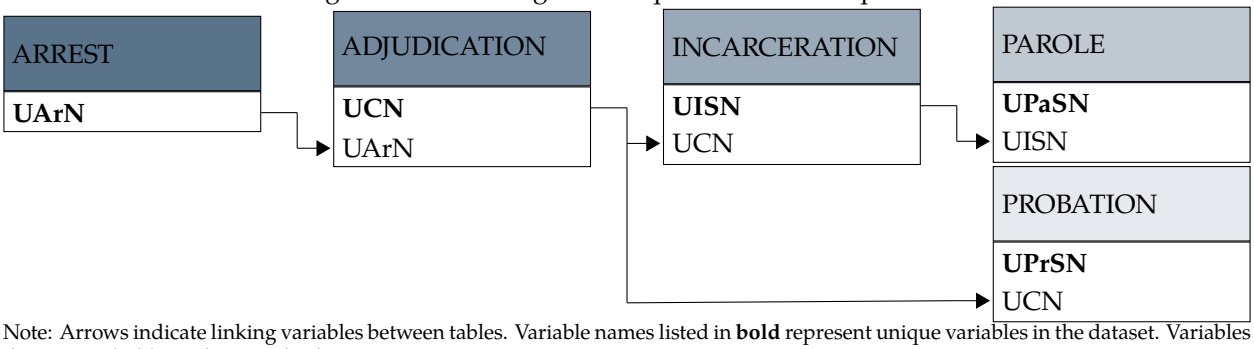
Note: Arrows indicate linking variables between tables. Variable names listed in **bold** represent unique variables in the dataset. Variables that are not bold may have multiple entries.

3.2.2 Data linkage via episode stage unique identifiers

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 5 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 court case filing in the adjudication relation database will contain the UCN which uniquely identifies that case filing, but also the UArN when it is known that an arrest led to that case filing. This allows for the reconstruction of a chain of events which can ultimately be linked back to an individual.

Figure 5: 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 F for more information about the constraints of the CJARS database.

3.3 Description of probabilistic entity resolution algorithms

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 D.

3.3.1 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 biometrically linked data (via 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.

3.3.2 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 was 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.

3.4 Availability of data by state

CJARS collects data on a continual basis and 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).

Table 6 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, "caseload snapshots" of all offenders under supervision at a single point in time, and planned coverage. Please see the legend for further detail.

Note that Table 6 only provides a general overview of CJARS data holdings. More details about the data in each jurisdiction can be found in Appendix A.

Figure 6: CJARS statewide coverage

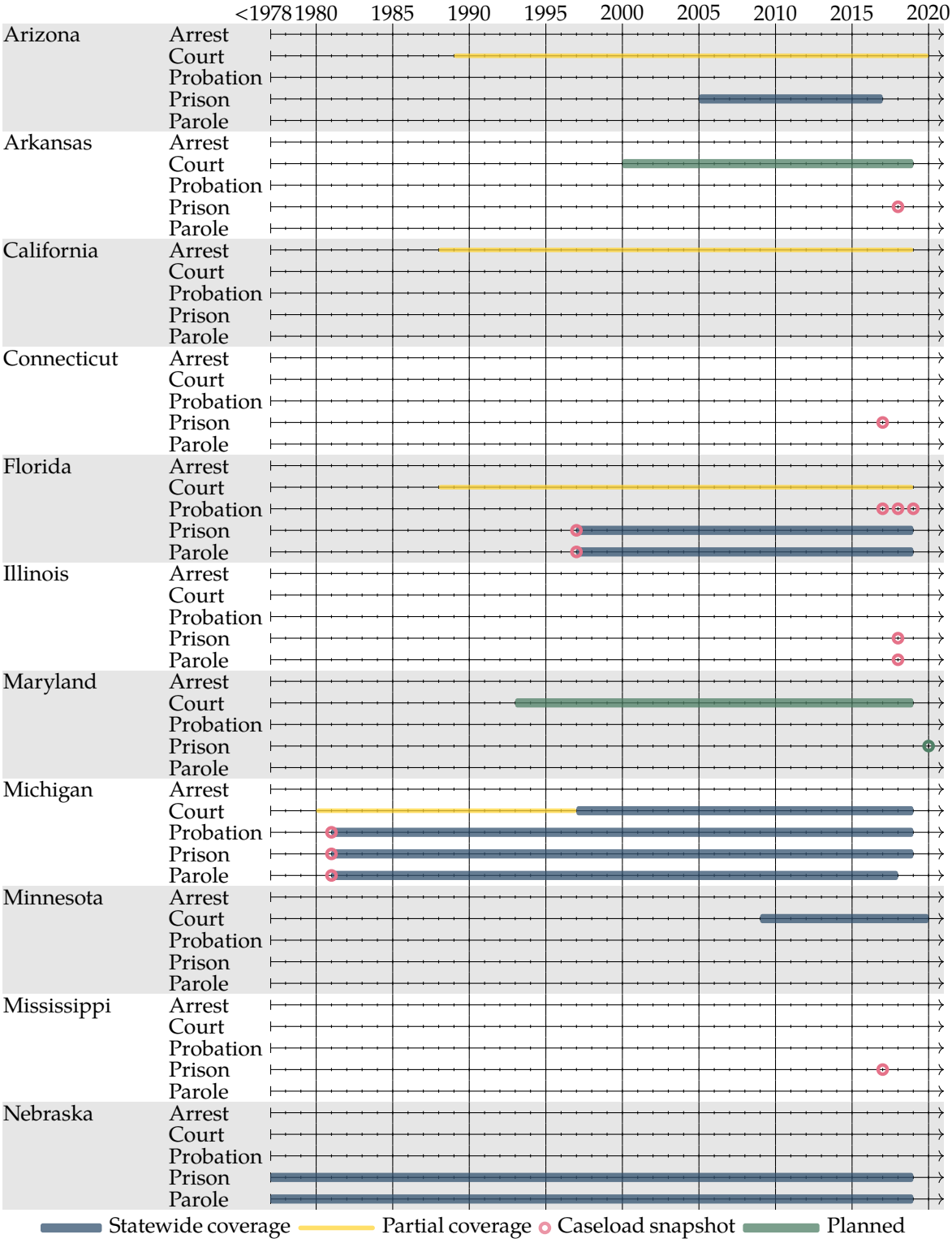
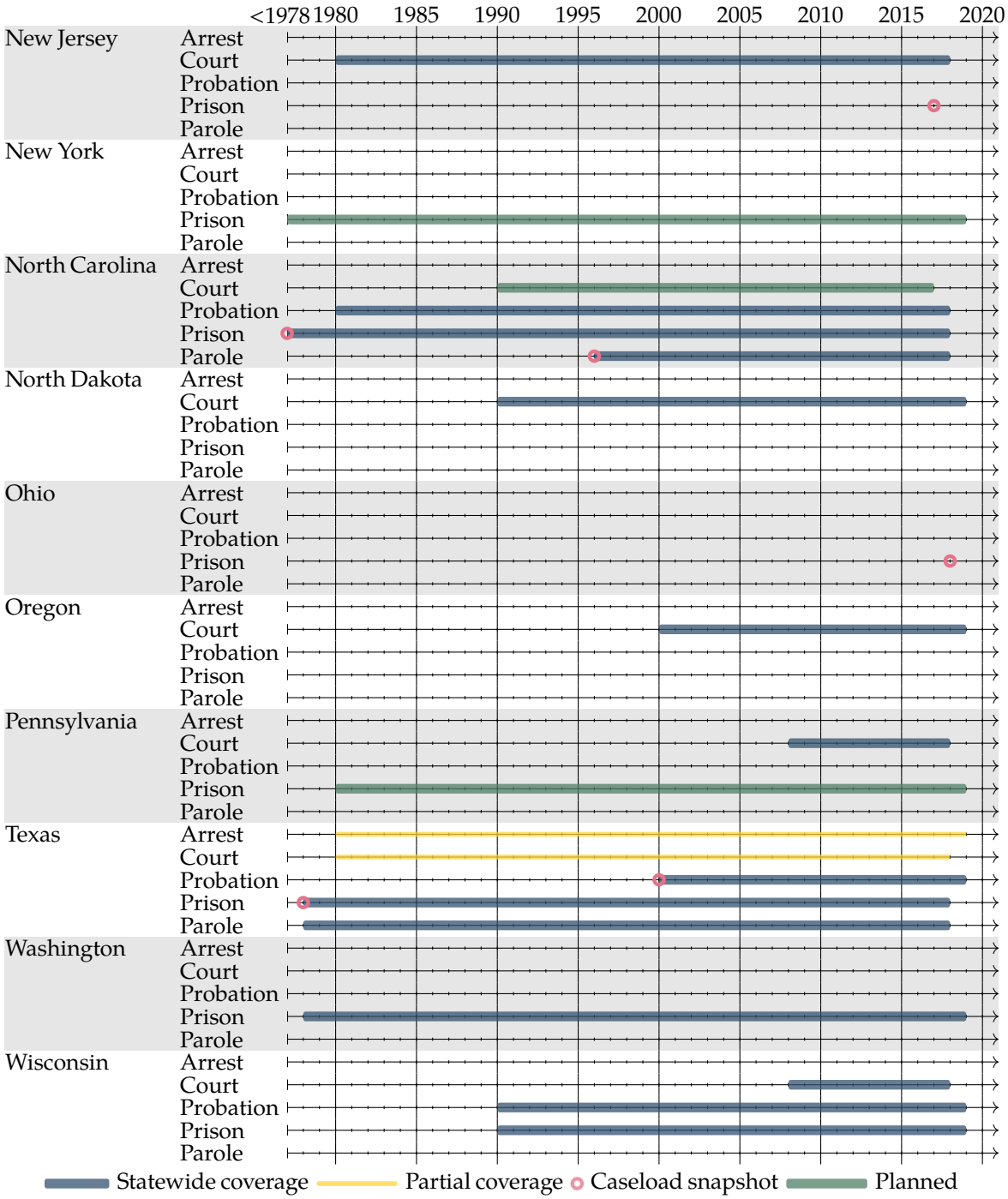


Figure 6: CJARS statewide coverage (cont'd)



3.5 Variables by relational table

The CJARS relational databases contain detailed information about arrests and bookings, adjudications, incarcerations, and terms of parole and probation. Users can also use a summary file of a minimal set of conviction and sanction events.

- **Arrest and booking.** The arrest table contains information regarding the arrest and booking date, as well as the offense that led to the arrest.
  - `cjars_id`: CJARS identifier
  - `uarn`: Arrest identifier
  - `arr_arr_dt_yyyy`: Year of arrest
  - `arr_arr_dt_mm`: Month of arrest
  - `arr_arr_dt_dd`: Day of the month of arrest
  - `arr_book_dt_yyyy`: Year of booking
  - `arr_book_dt_mm`: Month of booking
  - `arr_book_dt_dd`: Day of the month of booking
  - `arr_off_cd`: CJARS standardized offense code - arresting offense
  - `arr_off_cd_src`: Raw offense code from source - arresting offense
  - `arr_st_ori_fips`: State FIPS code
  - `arr_cnty_ori_fips`: County FIPS code
  - `arr_rec_src_le`: Record source - law enforcement 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
- **Adjudication.** The adjudication table contains detailed information about the offense the person was charged with, disposition information, and sentencing.
  - `cjars_id`: CJARS identifier
  - `ucn`: Court case filing identifier
  - `adj_grd_cd`: CJARS standardized offense grade
  - `adj_grd_cd_src`: Raw offense grade from source (e.g., felony, citation, misdemeanor)
  - `adj_file_dt_yyyy`: Year case was filed
  - `adj_file_dt_mm`: Month case was filed
  - `adj_file_dt_dd`: Day of month case was filed
  - `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`: Year of disposition
  - `adj_disp_dt_mm`: Month of disposition
  - `adj_disp_dt_dd`: Day of the month of disposition
  - `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_off_dt_yyyy`: Year offense was committed
  - `adj_off_dt_mm`: Month offense was committed
  - `adj_off_dt_dd`: Day of month offense was committed
  - `adj_sent_dt_yyyy`: Year sentenced
  - `adj_sent_dt_mm`: Month sentenced
  - `adj_sent_dt_dd`: Day of the month sentenced
  - `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_ori_fips: State FIPS code
  - adj_cnty_ori_fips: County FIPS code
  - adj_rec_src_le: Record source - law enforcement 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
- **Incarceration.** The incarceration table contains information about the facility an individual is/was housed, entry and exit dates, as well as the current status of the person.
  - cjars_id: CJARS identifier
  - uisn: Incarceration term identifier
  - inc_fcl_cd: CJARS standardized facility type
  - inc_fcl_cd_src: Raw description of facility from source
  - inc_entry_dt_yyyy: Year of entry into incarceration
  - inc_entry_dt_mm: Month of entry into incarceration
  - inc_entry_dt_dd: Day of month of entry into incarceration
  - inc_entry_cd: CJARS standardized entry status
  - inc_entry_cd_src: Raw description of entry type into incarceration
  - inc_exit_dt_yyyy: Year of exit from incarceration
  - inc_exit_dt_mm: Month of exit from incarceration
  - inc_exit_dt_dd: Day of month of exit from incarceration
  - 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: State FIPS code of state with jurisdiction over supervision of individual
  - inc_rec_src_le: Record source - law enforcement 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
- **Probation.** The probation table contains information on probation conditions, probation begin status and date, and probation end status and date.
  - cjars_id: CJARS identifier
  - uprsn: 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: Year of start of probation
  - pro_bgn_dt_mm: Month of start of probation
  - pro_bgn_dt_dd: Day of month of start of probation
  - pro_end_dt_yyyy: Year of end of probation
  - pro_end_dt_mm: Month of end of probation
  - pro_end_dt_dd: Day of month of end of probation
  - 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: State FIPS code of state with jurisdiction over supervision of individual
  - pro_rec_src_le: Record source - law enforcement 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
- **Parole.** The parole table contains information on parole begin/end dates and exit status when available.
  - cjars_id: CJARS identifier
  - upasn: Parole term identifier
  - par_bgn_dt_yyyy: Year of start of parole
  - par_bgn_dt_mm: Month of start of parole
  - par_bgn_dt_dd: Day of month of start of parole
  - par_end_dt_yyyy: Year of end of parole
  - par_end_dt_mm: Month of end of parole
  - par_end_dt_dd: Day of month of end of parole
  - 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: State FIPS code of state with jurisdiction over supervision of individual
  - par_rec_src_le: Record source - law enforcement 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

3.6 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 descriptives 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 defines as:

- Valid values: a value within range of what would reasonably be expected for the variable
- 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)
- Missing: the data was missing in the original raw data

3.6.1 Arrest and booking

cjars_id

Label	CJARS identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 4.
Table	Arrest and booking
Format	string

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0

uarn

Label	Arrest identifier
Description	Uniquely identifies arrest. For more details on use of `uarn` for data linkage, refer to Section 3.2.2 and Figure 5.
Table	Arrest and booking
Format	string

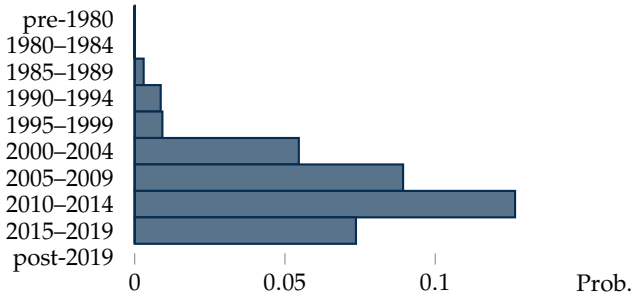
arr_arr_dt_yyyy

Label	Year of arrest
Description	The year when the individual was arrested.
Table	Arrest and booking
Format	numeric

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	6,117,927	36.5
Missing values	10,644,453	63.5

Statistic	*Value*
Mean	2009.2
Median	2010.0
Minimum	1903
Maximum	2019

Histogram of valid arr_arr_dt_yyyy records



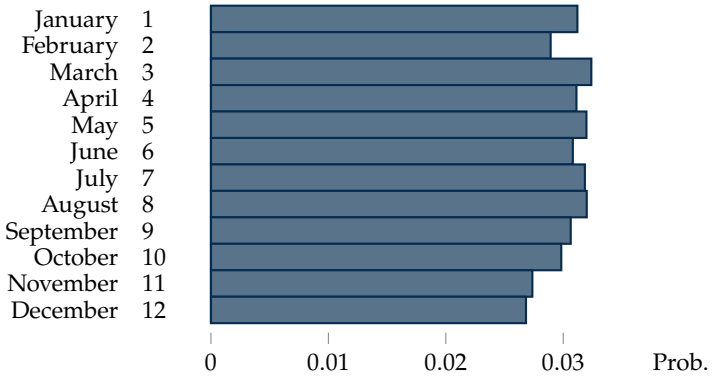
arr_arr_dt_mm

Label	Month of arrest
Description	The month when the individual was arrested.
Table	Arrest and booking
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	6,117,927	36.5
Missing values	10,644,453	63.5

Statistic	*Value*
Mean	6.4
Median	6.0
Minimum	1
Maximum	12

Histogram of valid arr_arr_dt_mm records



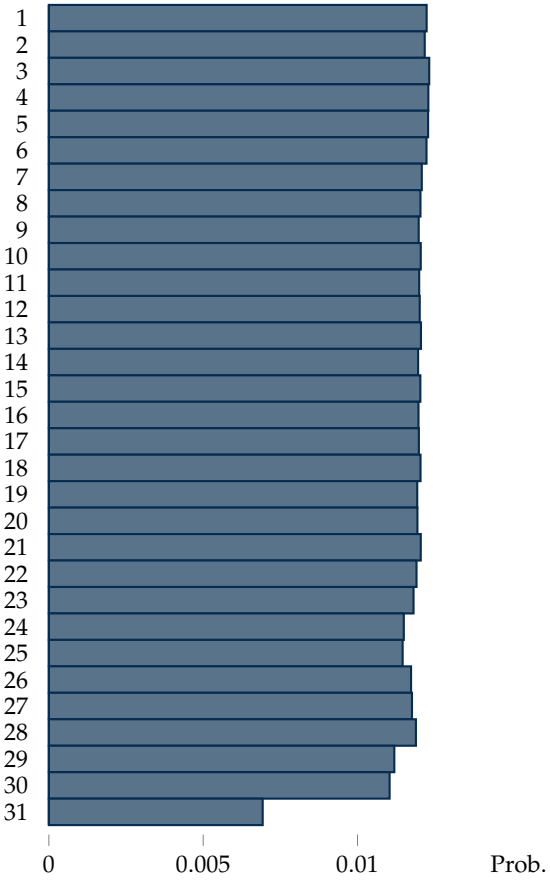
arr_arr_dt_dd

Label	Day of the month of arrest
Description	The day of the month when the individual was arrested.
Table	Arrest and booking
Format	numeric

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	6,117,927	36.5
Missing values	10,644,453	63.5

Statistic	*Value*
Mean	15.6
Median	16.0
Minimum	1
Maximum	31

Histogram of valid arr_arr_dt_dd records



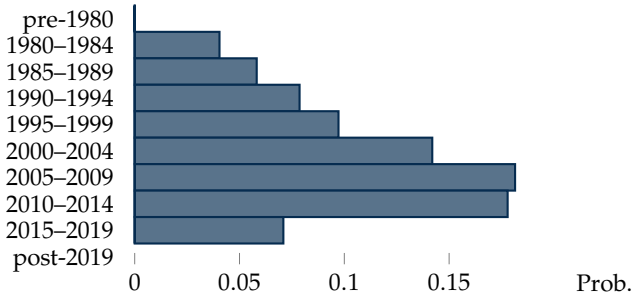
arr_book_dt_yyyy

Label	Year of booking
Description	The year when the individual was booked into jail.
Table	Arrest and booking
Format	numeric

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	14,194,209	84.7
Missing values	2,568,171	15.3

Statistic	*Value*
Mean	2002.9
Median	2005.0
Minimum	1911
Maximum	2019

Histogram of valid arr_book_dt_yyyy records



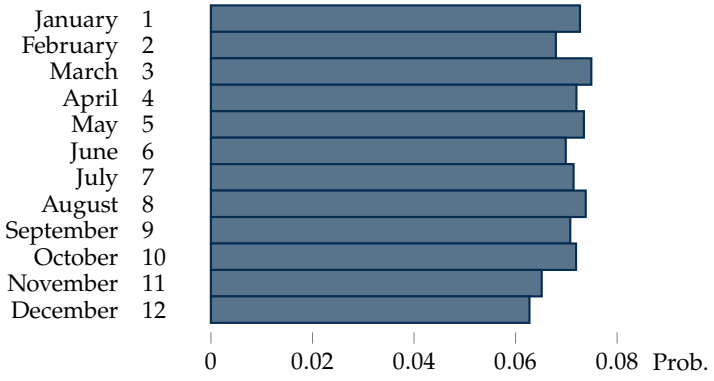
arr_book_dt_mm

Label	Month of booking
Description	The month when the individual was booked into jail.
Table	Arrest and booking
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	14,194,209	84.7
Missing values	2,568,171	15.3

Statistic	*Value*
Mean	6.4
Median	6.0
Minimum	1
Maximum	12

Histogram of valid arr_book_dt_mm records



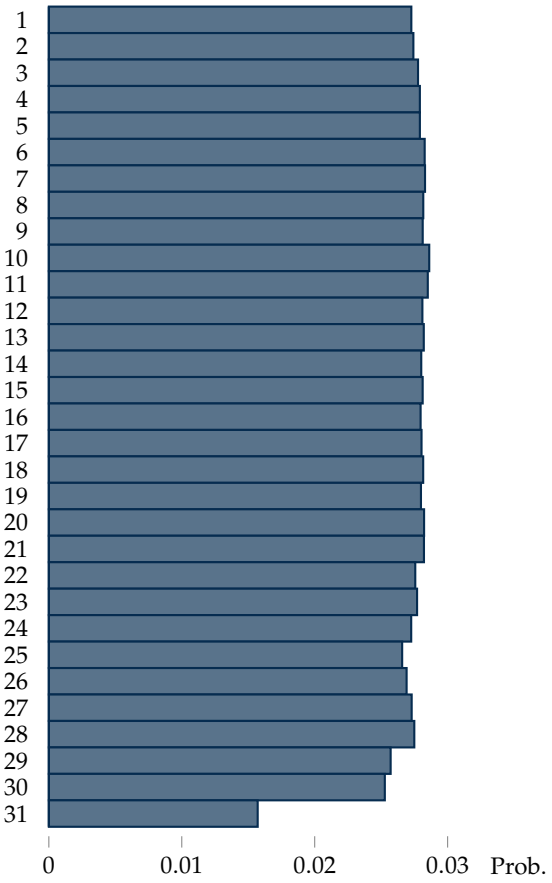
arr_book_dt_dd

Label	Day of the month of booking
Description	The day of the month when the individual was booked into jail.
Table	Arrest and booking
Format	numeric

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	14,194,209	84.7
Missing values	2,568,171	15.3

Statistic	*Value*
Mean	15.7
Median	16.0
Minimum	1
Maximum	31

Histogram of valid arr_book_dt_dd 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 B and C.2.1.
Table	Arrest and booking
Format	string
Code scheme	offense classification

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	15,524,171	92.6
Invalid values	1,006,498	6.0
Missing values	231,711	1.4

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.
Table	Arrest and booking
Format	string

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0

arr_st_ori_fips

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

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

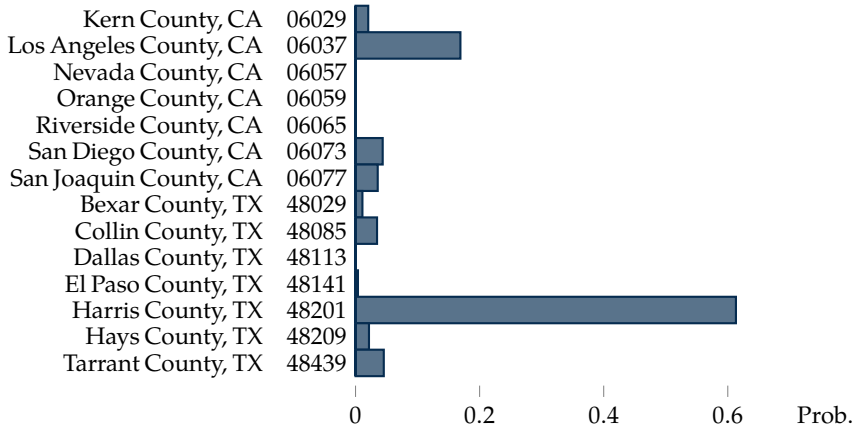
Histogram of valid `arr_st_ori_fips` records

arr_cnty_ori_fips

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

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

Histogram of valid `arr_cnty_ori_fips` records



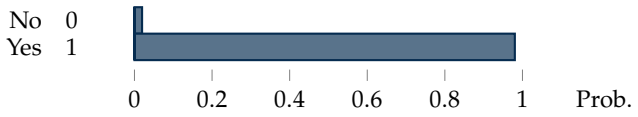
arr_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

Statistic	*Value*
Mean	1.0
Median	1.0
Maximum	1.0

Histogram of valid `arr_rec_src_le` records



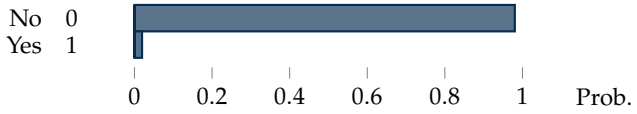
arr_rec_src_crt

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 and booking
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

Statistic	*Value*
Mean	0.0
Maximum	1.0

Histogram of valid arr_rec_src_crt records

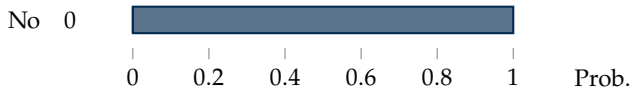


arr_rec_src_doc

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 and booking
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

Histogram of valid arr_rec_src_doc records

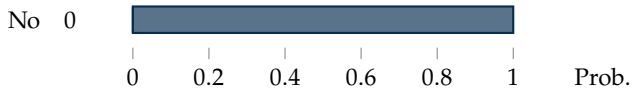


arr_rec_src_rep

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 and booking
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

Histogram of valid arr_rec_src_rep records

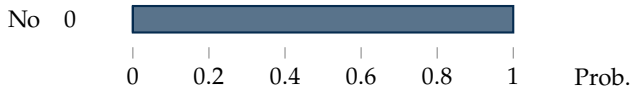


arr_rec_src_cc

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 and booking
Format	numeric
Code scheme	binary code

Set	*Count*	*Percent (%)*
All records	16,762,380	100.0
Valid records	16,762,380	100.0

Histogram of valid arr_rec_src_cc records



3.6.2 Adjudication

cjars_id

Label	CJARS identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 4.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0

ucn

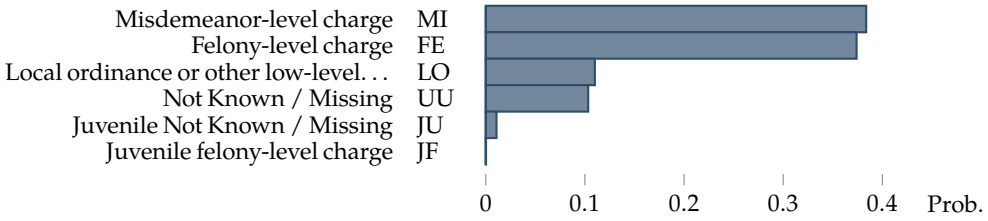
Label	Court case filing identifier
Description	Uniquely identifies court case filings. For more details on use of `ucn` for data linkage, refer to Section 3.2.2 and Figure 5.
Table	Adjudication
Format	string

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 disposition, please refer to Appendices B and C.2.2.
Table	Adjudication
Format	string
Code scheme	charge grade code

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	78,066,624	98.2
Invalid values	13,754	0.0
Missing values	1,408,759	1.8

Histogram of valid adj_grd_cd records



adj_grd_cd_src

Label	Raw offense grade from source (e.g., felony, citation, 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.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0

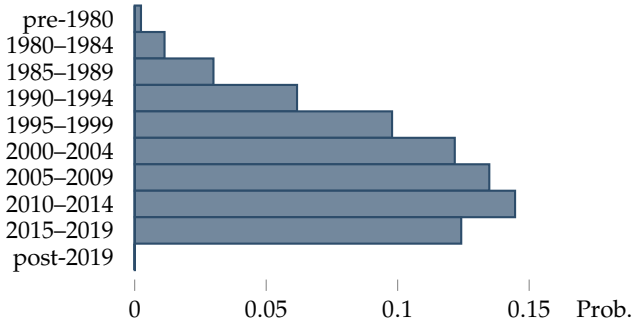
adj_file_dt_yyyy

Label	Year case was filed
Description	The year when the individual's case was filed.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	57,925,464	72.9
Invalid values	1297	0.0
Missing values	21,562,376	27.1

Statistic	*Value*
Mean	2004.9
Median	2006.0
Minimum	100
Maximum	2020

Histogram of valid adj_file_dt_yyyy records



Data notes

- Texas: 1

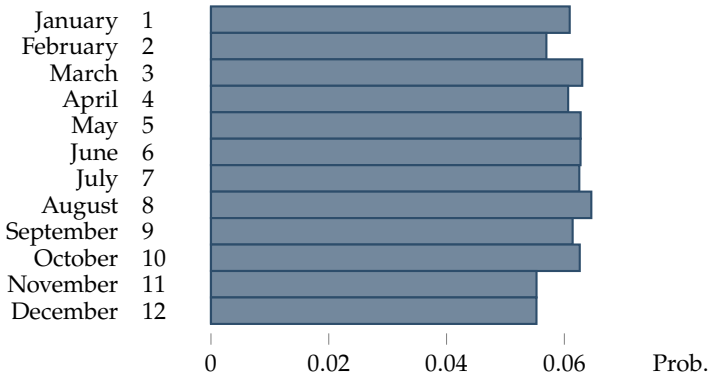
adj_file_dt_mm

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	57,926,756	72.9
Missing values	21,562,381	27.1

Statistic	*Value*
Mean	6.5
Median	6.0
Minimum	1
Maximum	12

Histogram of valid adj_file_dt_mm records



Data notes

- Texas: 1

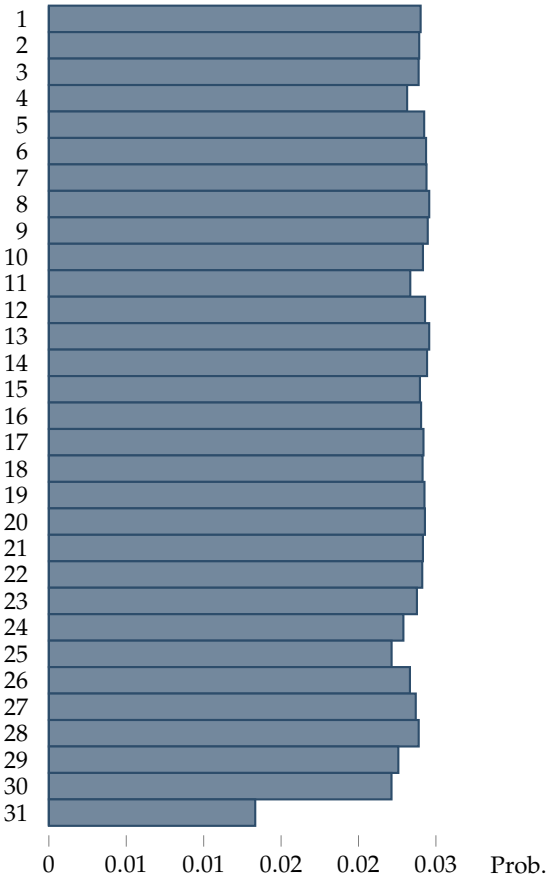
adj_file_dt_dd

Label	Day of month case was filed
Description	The day of the month when the individual's case was filed.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	57,926,756	72.9
Missing values	21,562,381	27.1

Statistic	*Value*
Mean	15.7
Median	16.0
Minimum	1
Maximum	31

Histogram of valid `adj_file_dt_dd` records



adj_file_dt_dd (continued)

Data notes

- Texas: 1

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 B and C.2.1.
Table	Adjudication
Format	string
Code scheme	offense code

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	26,924,122	33.9
Invalid values	506,770	0.6
Missing values	52,058,245	65.5

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.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0

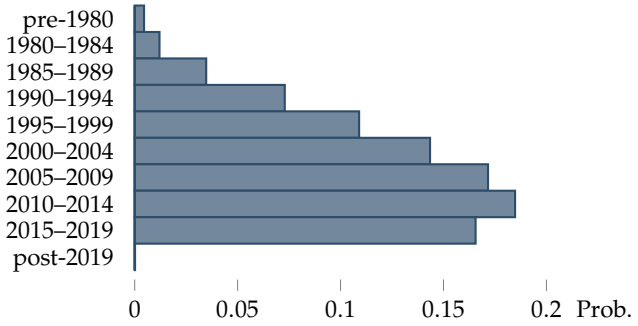
adj_disp_dt_yyyy

Label	Year of disposition
Description	The year when the individual's case was disposed.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	71,450,503	89.9
Invalid values	1092	0.0
Missing values	8,037,542	10.1

Statistic	*Value*
Mean	2005.4
Median	2007.0
Minimum	111
Maximum	2020

Histogram of valid `adj_disp_dt_yyyy` records



Data notes

- Nebraska: 1

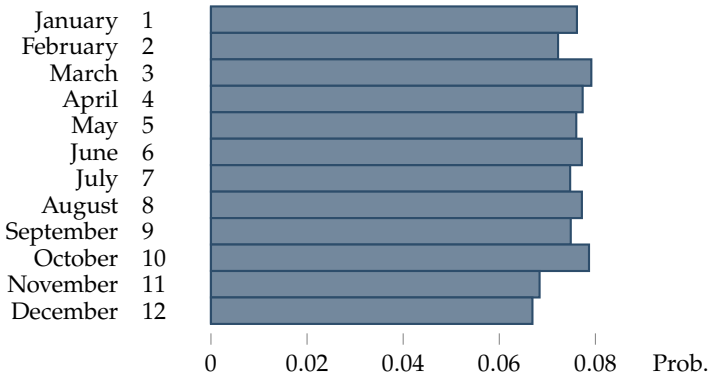
adj_disp_dt_mm

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	71,451,595	89.9
Missing values	8,037,542	10.1

Statistic	*Value*
Mean	6.4
Median	6.0
Minimum	1
Maximum	12

Histogram of valid adj_disp_dt_mm records



Data notes

- Nebraska: 1

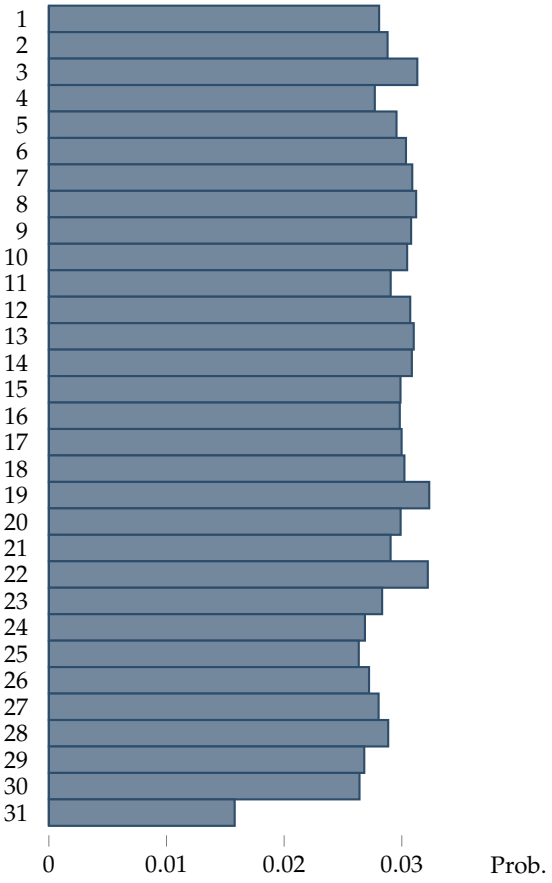
adj_disp_dt_dd

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	71,451,595	89.9
Missing values	8,037,542	10.1

Statistic	*Value*
Mean	15.6
Median	15.0
Minimum	1
Maximum	31

Histogram of valid adj_disp_dt_dd records



adj_disp_dt_dd (continued)

Data notes

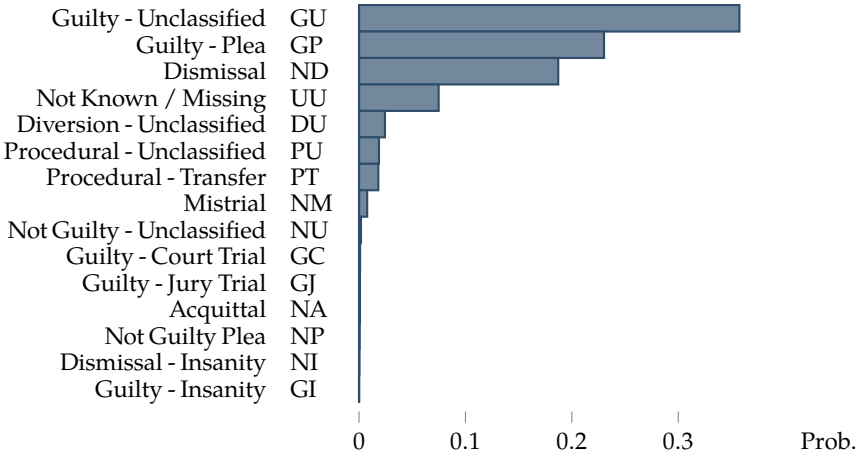
- Nebraska: 1

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 B and C.2.3.
Table	Adjudication
Format	string
Code scheme	disposition code

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	73,446,746	92.4
Missing values	6,042,391	7.6

Histogram of valid adj_disp_cd 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.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0

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 B and C.2.1.
Table	Adjudication
Format	string
Code scheme	offense classification

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	66,238,509	83.3
Invalid values	4,280,954	5.4
Missing values	8,969,674	11.3

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.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0

adj_off_dt_yyyy

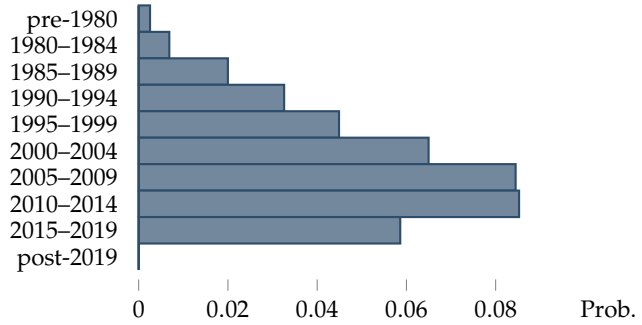
Label	Year offense was committed
Description	The year when the individual committed the offense.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	31,812,037	40.0
Invalid values	115	0.0
Missing values	47,676,985	60.0

Statistic	*Value*
Mean	2004.7
Median	2006.0
Minimum	100
Maximum	5007

adj_off_dt_yyyy (continued)

Histogram of valid adj_off_dt_yyyy records



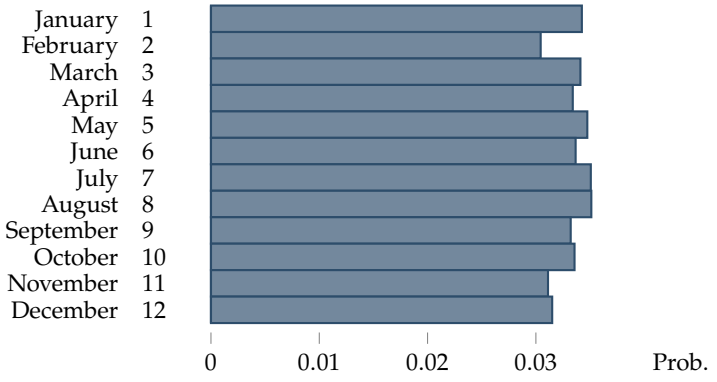
adj_off_dt_mm

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	31,812,152	40.0
Missing values	47,676,985	60.0

Statistic	*Value*
Mean	6.5
Median	6.0
Minimum	1
Maximum	12

Histogram of valid adj_off_dt_mm records



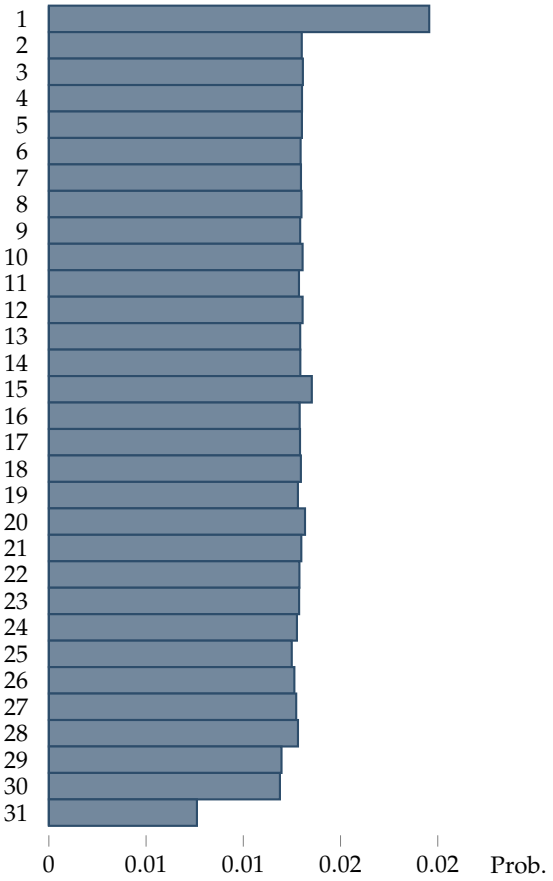
adj_off_dt_dd

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	31,812,152	40.0
Missing values	47,676,985	60.0

Statistic	*Value*
Mean	15.4
Median	15.0
Minimum	1
Maximum	31

Histogram of valid adj_off_dt_dd records



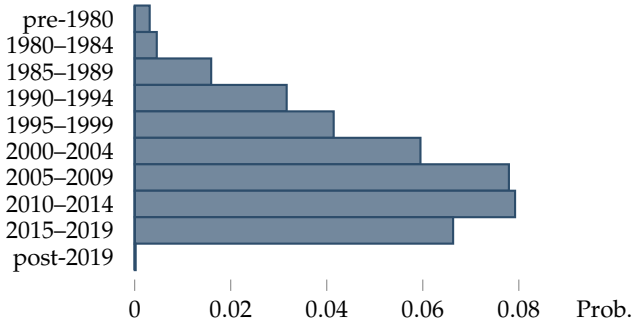
adj_sent_dt_yyyy

Label	Year sentenced
Description	The year when the individual was sentenced.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	30,198,738	38.0
Invalid values	886	0.0
Missing values	49,289,513	62.0

Statistic	*Value*
Mean	2005.3
Median	2007.0
Minimum	1883
Maximum	2020

Histogram of valid adj_sent_dt_yyyy records



Data notes

- Nebraska: 1
- New Jersey: 1
- Pennsylvania: 1

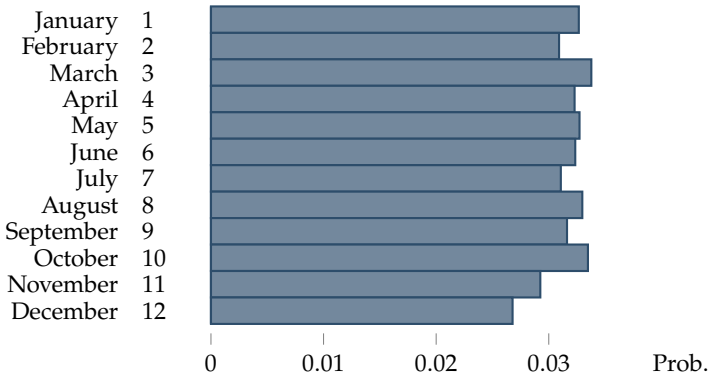
adj_sent_dt_mm

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	30,199,624	38.0
Missing values	49,289,513	62.0

Statistic	*Value*
Mean	6.4
Median	6.0
Minimum	1
Maximum	12

Histogram of valid adj_sent_dt_mm records



Data notes

- Nebraska: 1
- New Jersey: 1
- Pennsylvania: 1

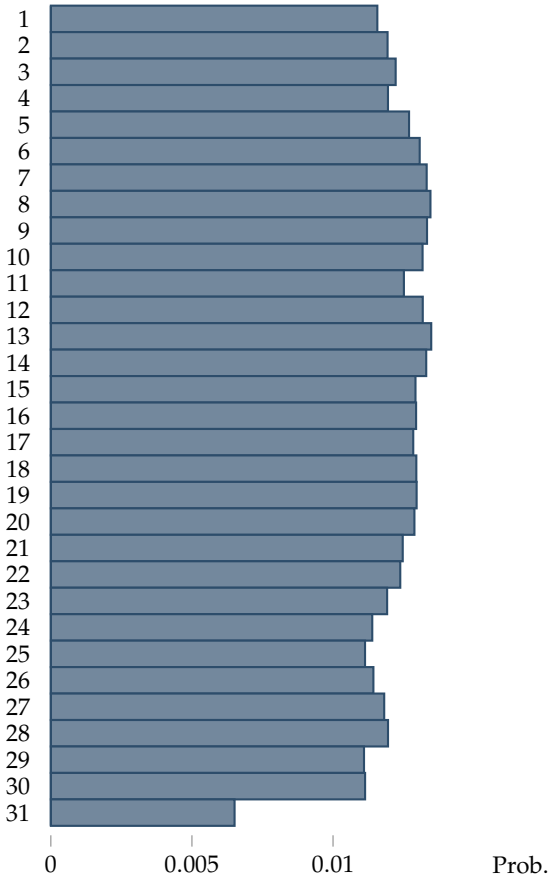
adj_sent_dt_dd

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	30,199,624	38.0
Missing values	49,289,513	62.0

Statistic	*Value*
Mean	15.5
Median	15.0
Minimum	1
Maximum	31

Histogram of valid adj_sent_dt_dd records



adj_sent_dt_dd (continued)

Data notes

- Nebraska: 1
- New Jersey: 1
- Pennsylvania: 1

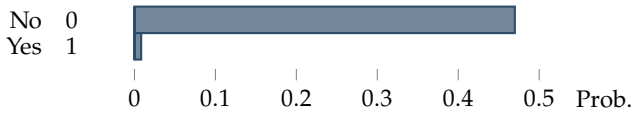
adj_sent_serv

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	79,489,137	100.0
Valid records	38,013,136	47.8
Missing values	41,476,001	52.2

Statistic	*Value*
Mean	0.0
Maximum	1

Histogram of valid adj_sent_serv records



Data notes

- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

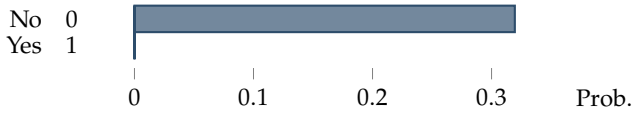
adj_sent_dth

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	25,403,918	32.0
Missing values	54,085,219	68.0

Statistic	*Value*
Mean	0.0
Maximum	1

Histogram of valid adj_sent_dth records



Data notes

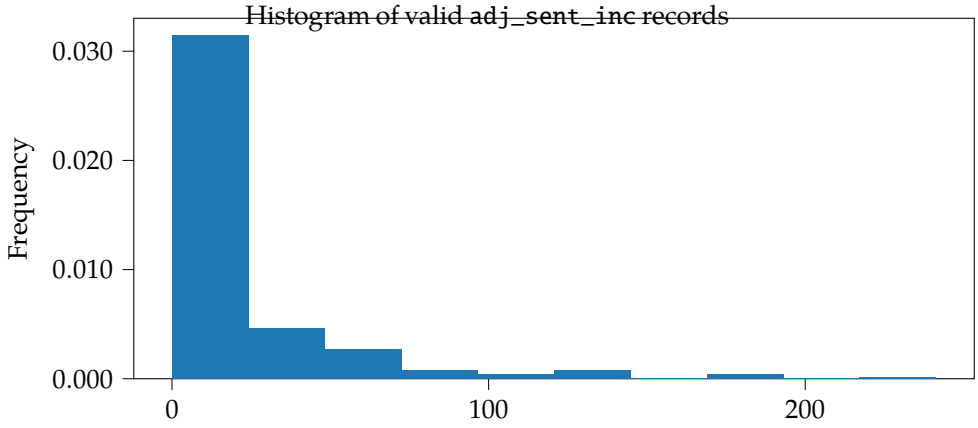
- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

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	79,489,137	100.0
Valid records	20,125,425	25.3
Invalid values	77,771	0.1
Missing values	59,285,941	74.6

Statistic	*Value*
Mean	−319.8
Median	1.0
Minimum	−99,999.0
Maximum	108,595.0



Data notes

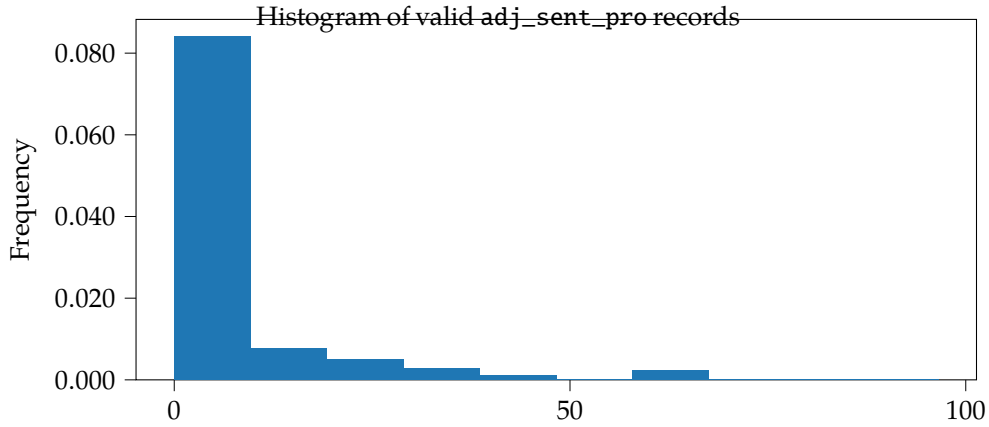
- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

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	79,489,137	100.0
Valid records	48,069,898	60.5
Invalid values	149	0.0
Missing values	31,419,090	39.5

Statistic	*Value*
Mean	8.5
Minimum	−99,999.0
Maximum	108,655.4



Data notes

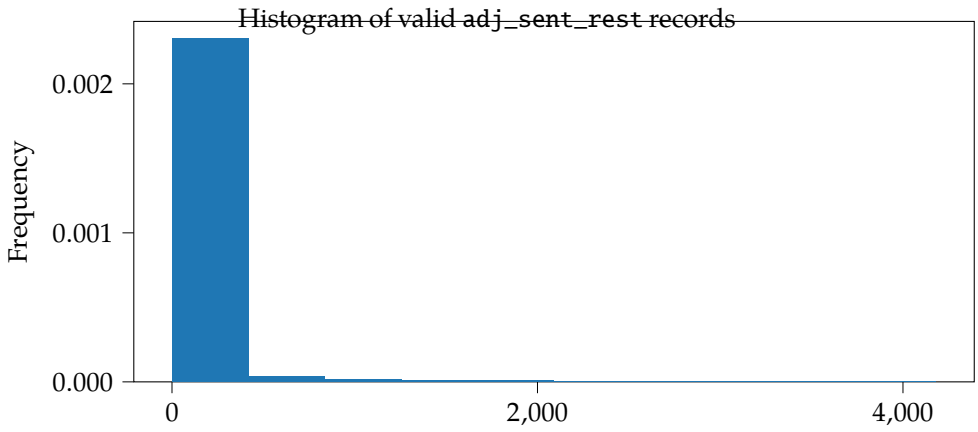
- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

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	79,489,137	100.0
Valid records	11,585,027	14.6
Invalid values	6	0.0
Missing values	67,904,104	85.4

Statistic	*Value*
Mean	441.0
Minimum	−86.7
Maximum	10,000,000.0



Data notes

- New Jersey: 1
- North Carolina: 1
- Pennsylvania: 1
- Texas: 1

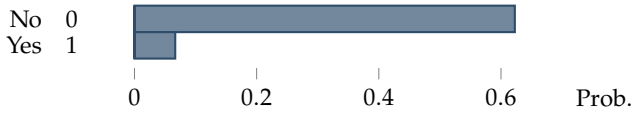
adj_sent_sus

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	54,796,901	68.9
Missing values	24,692,236	31.1

Statistic	*Value*
Mean	0.1
Maximum	1

Histogram of valid adj_sent_sus records



Data notes

- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

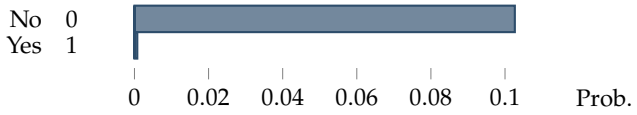
adj_sent_trt

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	79,489,137	100.0
Valid records	8,220,040	10.3
Missing values	71,269,097	89.7

Statistic	*Value*
Mean	0.0
Maximum	1

Histogram of valid adj_sent_trt records



Data notes

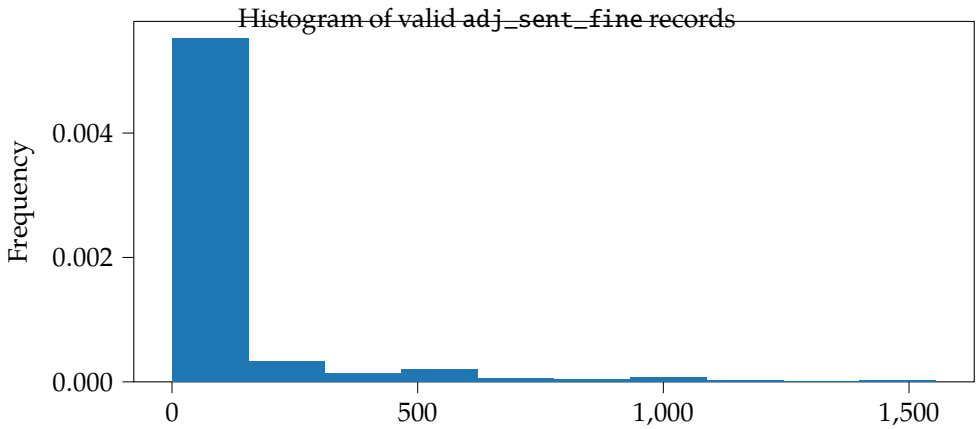
- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

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	79,489,137	100.0
Valid records	27,666,051	34.8
Invalid values	1193	0.0
Missing values	51,821,893	65.2

Statistic	*Value*
Mean	139.3
Minimum	−35,492.0
Maximum	9,999,999.0



Data notes

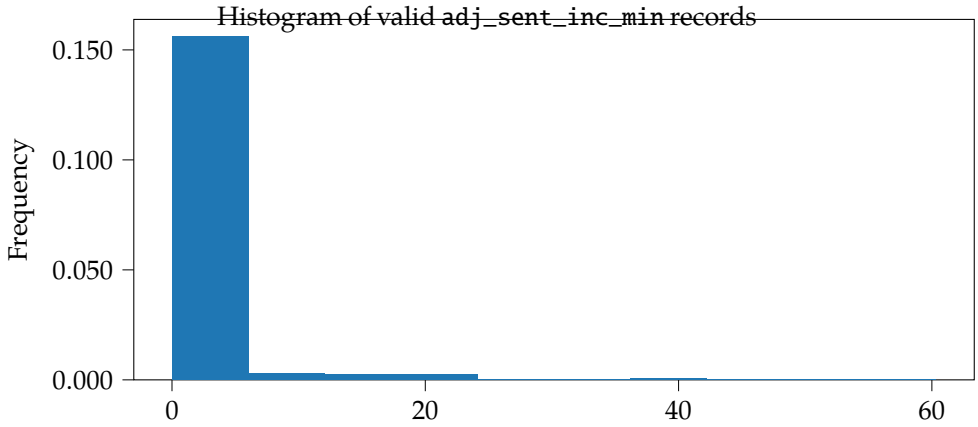
- New Jersey: 1
- North Carolina: 1
- North Dakota: 1
- Pennsylvania: 1
- Texas: 1
- Wisconsin: 1

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	79,489,137	100.0
Valid records	27,639,032	34.8
Invalid values	5459	0.0
Missing values	51,844,646	65.2

Statistic	*Value*
Mean	−14.5
Minimum	−99,999.0
Maximum	12,654.0



Data notes

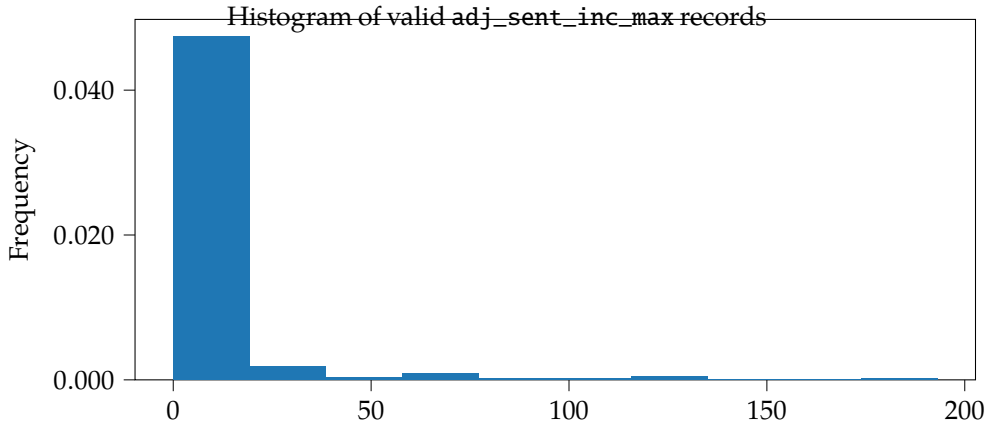
- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

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	79,489,137	100.0
Valid records	29,790,655	37.5
Invalid values	40,298	0.1
Missing values	49,658,184	62.5

Statistic	*Value*
Mean	−115.5
Minimum	−99,999.0
Maximum	67,084.1



Data notes

- New Jersey: 1
- Pennsylvania: 1
- Texas: 1

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_pdiv, 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.
Table	Adjudication
Format	string

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0

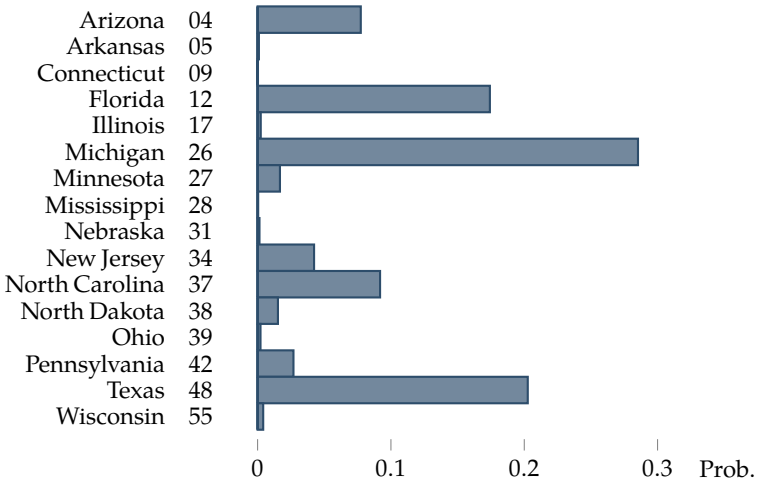
adj_st_ori_fips

Label	State FIPS code
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	79,489,137	100.0
Valid records	75,064,074	94.4
Missing values	4,425,063	5.6

adj_st_ori_fips (continued)

Histogram of valid adj_st_ori_fips records



adj_cnty_ori_fips

Label	County FIPS code
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

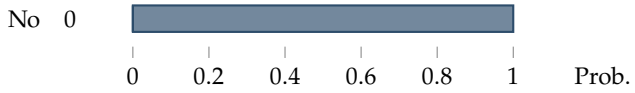
Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	73,141,014	92.0
Invalid values	1,923,060	2.4
Missing values	4,425,063	5.6

adj_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	79,489,137	100.0
Valid records	79,489,137	100.0

Histogram of valid `adj_rec_src_le` records



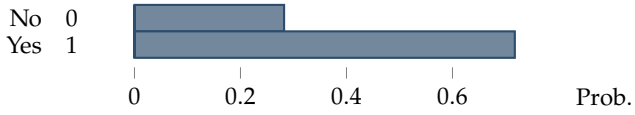
adj_rec_src_crt

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	79,489,137	100.0
Valid records	79,489,137	100.0

Statistic	*Value*
Mean	0.2
Median	1.0
Maximum	1.0

Histogram of valid `adj_rec_src_crt` records



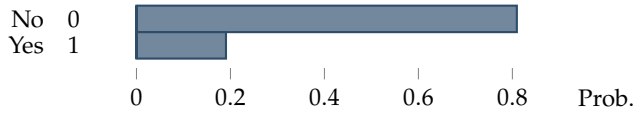
adj_rec_src_doc

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	79,489,137	100.0
Valid records	79,489,137	100.0

Statistic	*Value*
Mean	0.2
Maximum	1.0

Histogram of valid `adj_rec_src_doc` records



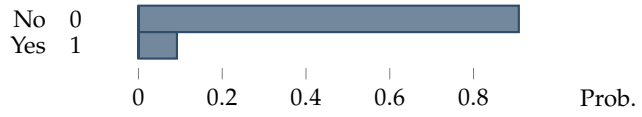
adj_rec_src_rep

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	79,489,137	100.0
Valid records	79,489,137	100.0

Statistic	*Value*
Mean	0.1
Maximum	1.0

Histogram of valid adj_rec_src_rep records

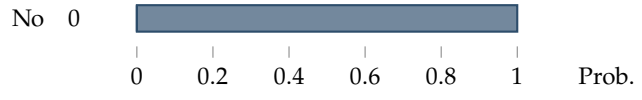


adj_rec_src_cc

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	79,489,137	100.0
Valid records	79,489,137	100.0

Histogram of valid adj_rec_src_cc records



3.6.3 Incarceration

cjars_id

Label	CJARS identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 4.
Table	Incarceration
Format	string

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0

uisn

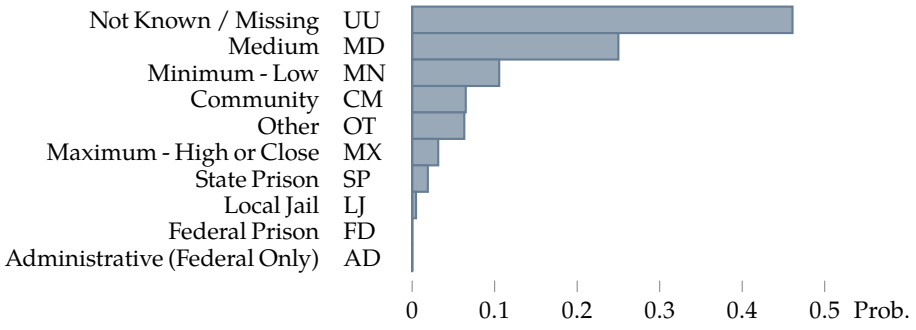
Label	Incarceration term identifier
Description	Uniquely identifies term of incarceration. For more details on use of `uisn` for data linkage, refer to Section 3.2.2 and Figure 5.
Table	Incarceration
Format	string

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 B and C.2.7.
Table	Incarceration
Format	string
Code scheme	incarceration facility type

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	10,368,704	100.0
Invalid values	311	0.0
Missing values	45	0.0

Histogram of valid `inc_fcl_cd` 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	10,369,060	100.0

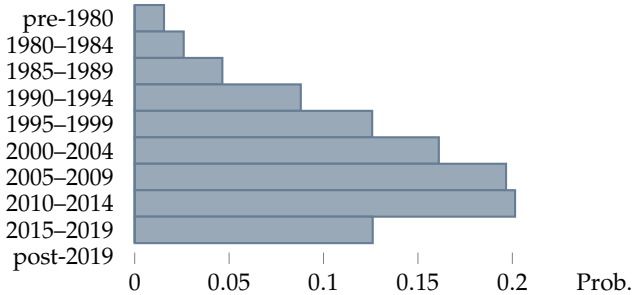
inc_entry_dt_yyyy

Label	Year of entry into incarceration
Description	The year when the individual entered incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	10,232,393	98.7
Missing values	136,667	1.3

Statistic	*Value*
Mean	2003.8
Median	2005.0
Minimum	1901
Maximum	2019

Histogram of valid inc_entry_dt_yyyy records



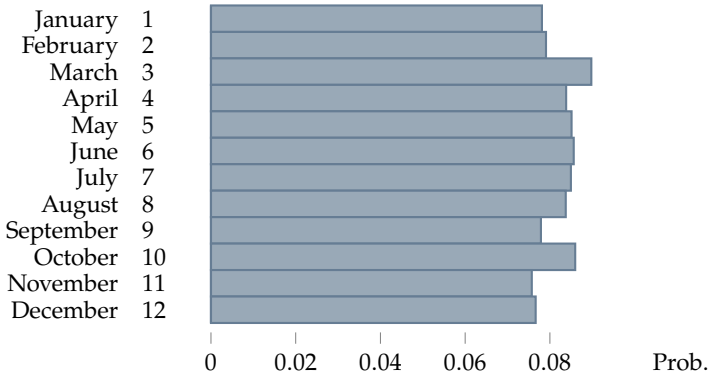
inc_entry_dt_mm

Label	Month of entry into incarceration
Description	The month when the individual entered incarceration.
Table	Incarceration
Format	numeric
Code scheme	month code

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	10,228,703	98.6
Missing values	140,357	1.4

Statistic	*Value*
Mean	6.4
Median	6.0
Minimum	1
Maximum	12

Histogram of valid inc_entry_dt_mm records



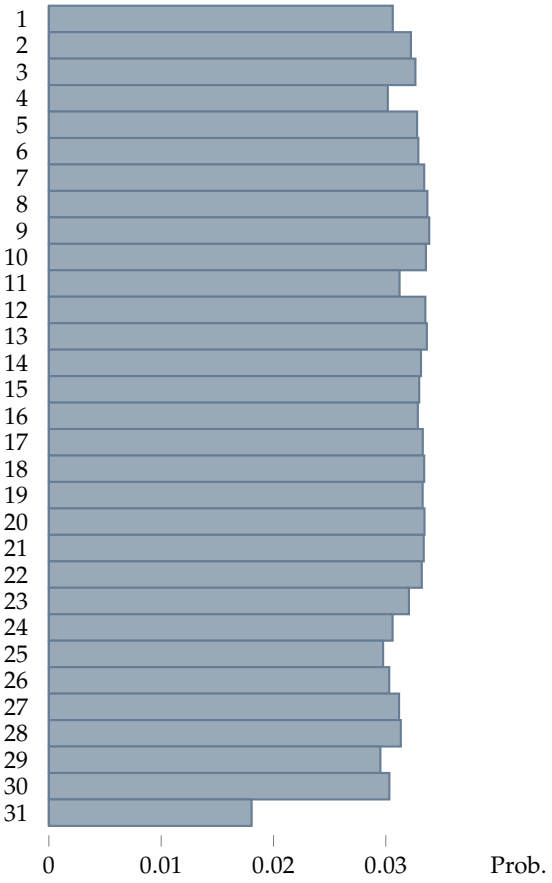
inc_entry_dt_dd

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

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	10,228,703	98.6
Missing values	140,357	1.4

Statistic	*Value*
Mean	15.6
Median	16.0
Minimum	1
Maximum	31

Histogram of valid inc_entry_dt_dd records

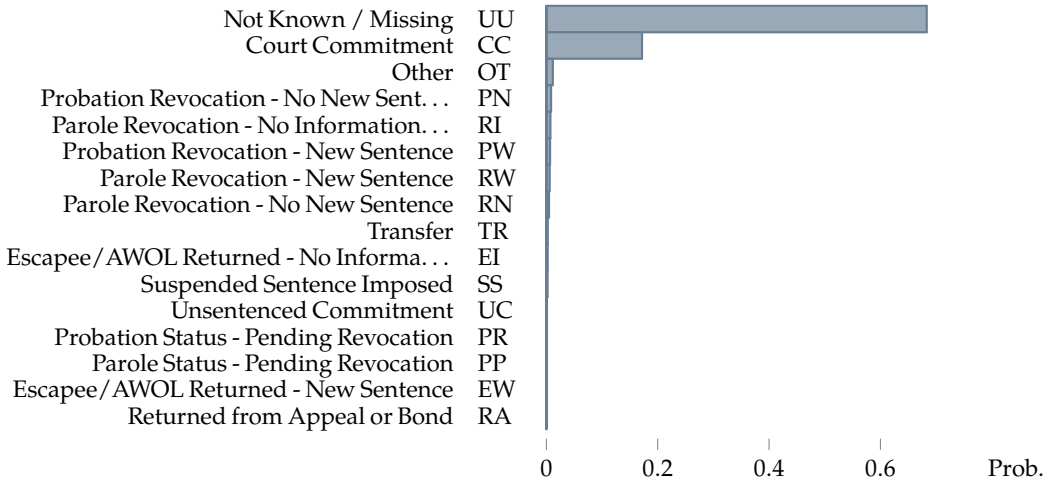


inc_entry_cd

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 B and C.2.6.
Table	Incarceration
Format	string
Code scheme	incarceration entry code

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	9,431,937	91.0
Invalid values	4486	0.0
Missing values	932,637	9.0

Histogram of valid `inc_entry_cd` records



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	10,369,060	100.0

inc_exit_dt_yyyy

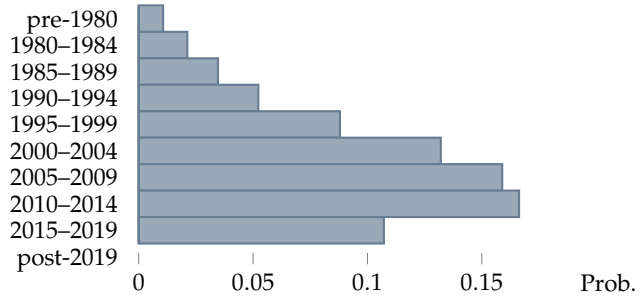
Label	Year of exit from incarceration
Description	Year when the individual exited incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	7,998,040	77.1
Invalid values	1	0.0
Missing values	2,371,019	22.9

Statistic	*Value*
Mean	2004.4
Median	2006.0
Minimum	1513
Maximum	2019

inc_exit_dt_yyyy (continued)

Histogram of valid inc_exit_dt_yyyy records



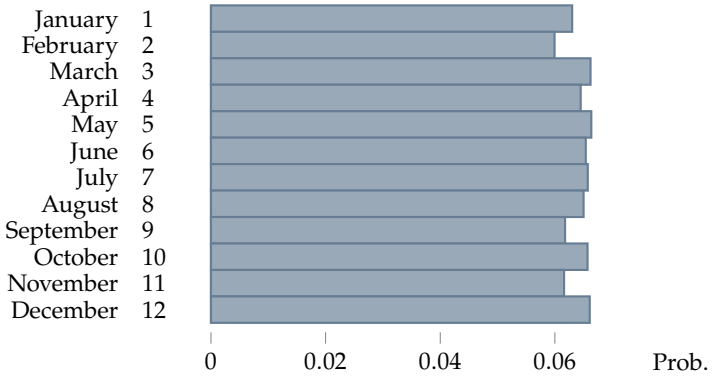
inc_exit_dt_mm

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

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	7,998,041	77.1
Missing values	2,371,019	22.9

Statistic	*Value*
Mean	6.5
Median	7.0
Minimum	1
Maximum	12

Histogram of valid inc_exit_dt_mm records



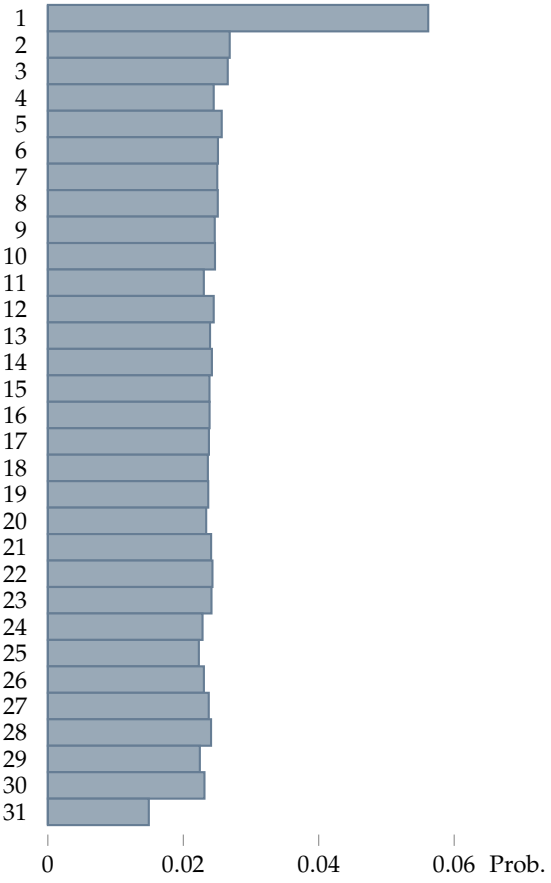
inc_exit_dt_dd

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

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	7,998,041	77.1
Missing values	2,371,019	22.9

Statistic	*Value*
Mean	14.9
Median	15.0
Minimum	1
Maximum	31

Histogram of valid `inc_exit_dt_dd` records



Data notes

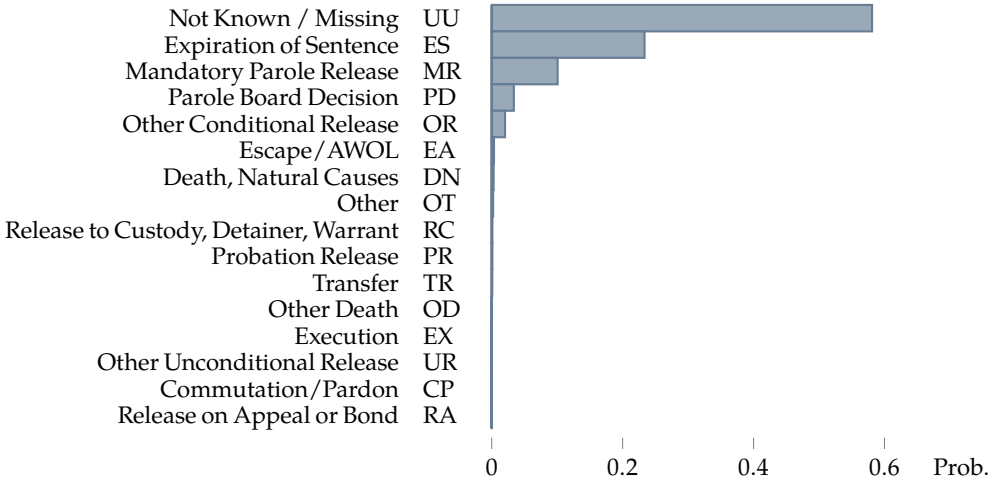
- Florida: 1

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 B and C.2.8.
Table	Incarceration
Format	string
Code scheme	incarceration exit code

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	10,177,861	98.2
Invalid values	8059	0.1
Missing values	183,140	1.8

Histogram of valid `inc_exit_cd` 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	10,369,060	100.0

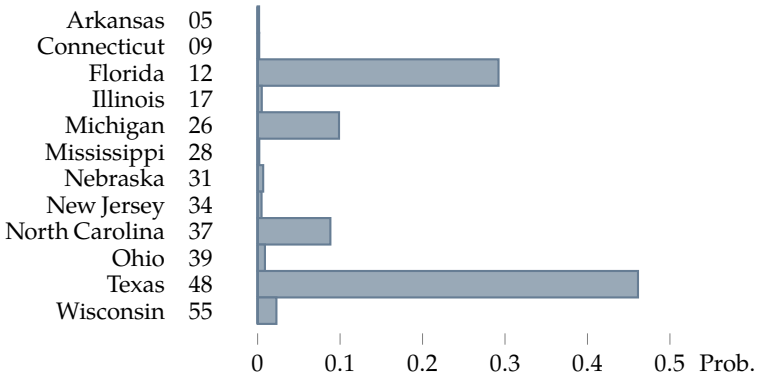
inc_st_ori_fips

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 identifying 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	10,369,060	100.0
Valid records	10,309,358	99.4
Missing values	59,702	0.6

inc_st_ori_fips (continued)

Histogram of valid inc_st_ori_fips 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 identifying 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	10,369,060	100.0
Valid records	5,713,813	55.1
Invalid values	4,595,545	44.3
Missing values	59,702	0.6

Data notes

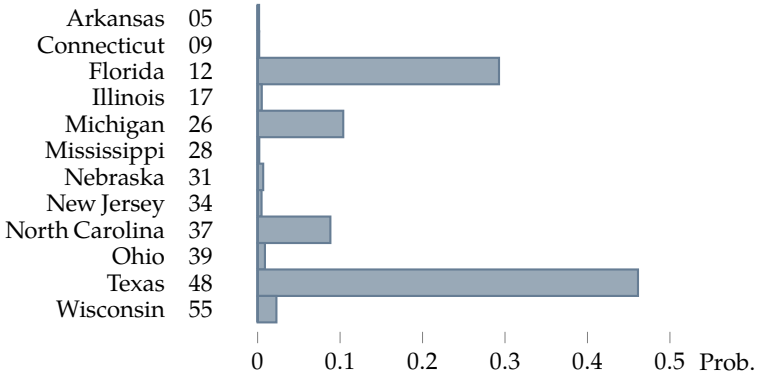
- Wisconsin: 1

inc_st_juris_fips

Label	State 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 incarceration. Used to help uniquely identifying 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	10,369,060	100.0
Valid records	10,369,060	100.0

Histogram of valid `inc_st_juris_fips` records

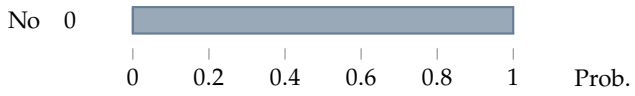


inc_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	10,369,060	100.0
Valid records	10,369,060	100.0

Histogram of valid `inc_rec_src_le` records

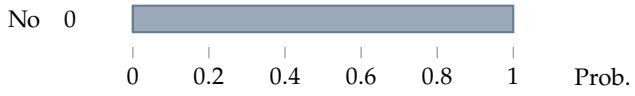


inc_rec_src_crt

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	10,369,060	100.0
Valid records	10,369,060	100.0

Histogram of valid `inc_rec_src_crt` 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	10,369,060	100.0
Valid records	10,369,060	100.0

Statistic	*Value*
Mean	0.9
Median	1.0
Maximum	1.0

Histogram of valid `inc_rec_src_doc` records

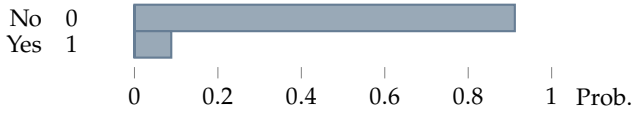
inc_rec_src_rep

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	10,369,060	100.0
Valid records	10,369,060	100.0

Statistic	*Value*
Mean	0.1
Maximum	1.0

Histogram of valid `inc_rec_src_rep` records

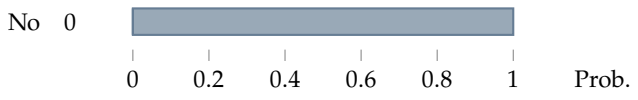


inc_rec_src_cc

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	10,369,060	100.0
Valid records	10,369,060	100.0

Histogram of valid `inc_rec_src_cc` records



3.6.4 Probation

cjars_id

Label	CJARS identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 4.
Table	Probation
Format	string

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0

uprsn

Label	Probation term identifier
Description	Uniquely identifies terms of probation. For more details on use of `uprsn` for data linkage, refer to Section 3.2.2 and Figure 5.
Table	Probation
Format	string

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 B and C.2.4.
Table	Probation
Format	string
Code scheme	probation condition code

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	9,386,602	100.0

Histogram of valid pro_cond_cd records

pro_cond_cd_src

Label	Raw description of probation conditions from source
Description	A descrption 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	9,386,602	100.0

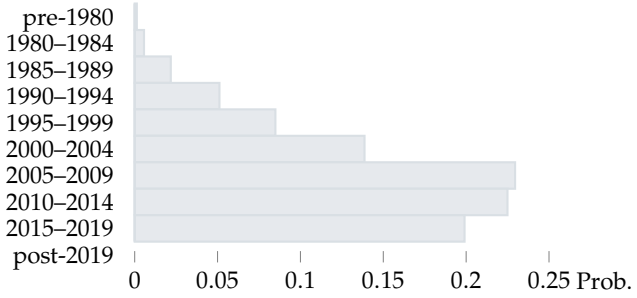
`pro_bgn_dt_yyyy`

Label	Year of start of probation
Description	The year when the individual began probation.
Table	Probation
Format	numeric

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	8,981,868	95.7
Invalid values	52	0.0
Missing values	404,682	4.3

Statistic	*Value*
Mean	2007.2
Median	2008.0
Minimum	177
Maximum	2019

Histogram of valid `pro_bgn_dt_yyyy` records



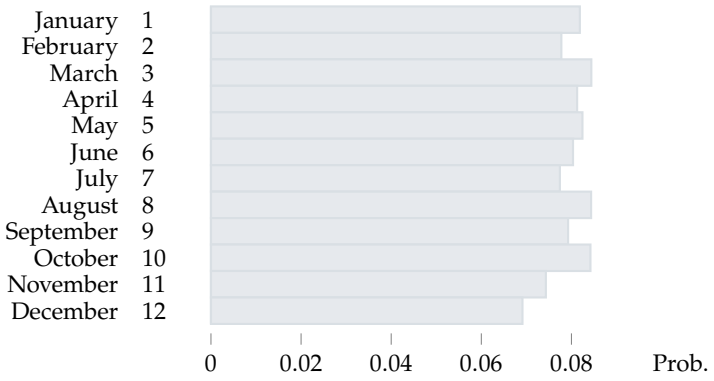
`pro_bgn_dt_mm`

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

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	8,981,920	95.7
Missing values	404,682	4.3

Statistic	*Value*
Mean	6.4
Median	6.0
Minimum	1
Maximum	12

Histogram of valid `pro_bgn_dt_mm` records



`pro_bgn_dt_dd`

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

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	8,981,920	95.7
Missing values	404,682	4.3

Statistic	*Value*
Mean	15.5
Median	15.0
Minimum	1
Maximum	31

Histogram of valid `pro_bgn_dt_dd` records

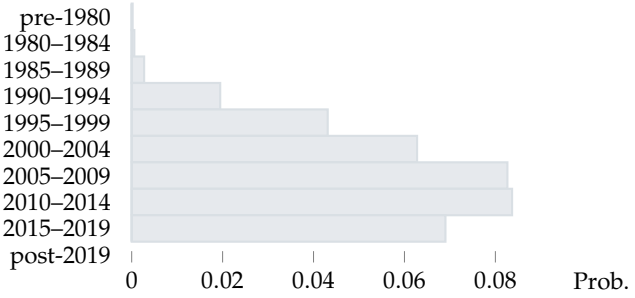
`pro_end_dt_yyyy`

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

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	3,418,822	36.4
Invalid values	1	0.0
Missing values	5,967,779	63.6

Statistic	*Value*
Mean	2007.0
Median	2008.0
Minimum	179
Maximum	2019

Histogram of valid `pro_end_dt_yyyy` records



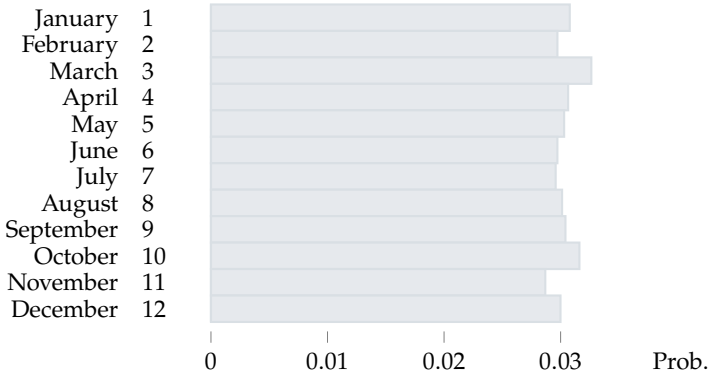
`pro_end_dt_mm`

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

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	3,418,823	36.4
Missing values	5,967,779	63.6

Statistic	*Value*
Mean	6.5
Median	6.0
Minimum	1
Maximum	12

Histogram of valid `pro_end_dt_mm` records



pro_end_dt_dd

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

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	3,418,823	36.4
Missing values	5,967,779	63.6

Statistic	*Value*
Mean	15.7
Median	16.0
Minimum	1
Maximum	31

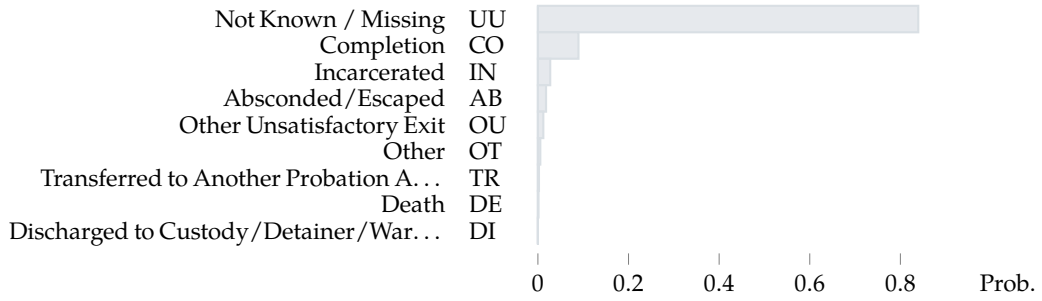
Histogram of valid pro_end_dt_dd records

`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 B and C.2.5.
Table	Probation
Format	string
Code scheme	probation exit code

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	9,341,085	99.5
Missing values	45,517	0.5

Histogram of valid `pro_end_cd` records



pro_end_cd_src

Label	Raw description of probation end status from source
Description	A descrption 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

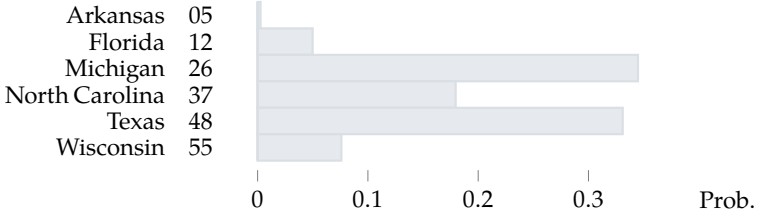
Set	*Count*	*Percent (%)*
All records	9,386,602	100.0

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 identifying 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	9,386,602	100.0
Valid records	9,236,825	98.4
Missing values	149,777	1.6

Histogram of valid pro_st_ori_fips 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 identifying geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Probation
Format	string
Code scheme	county FIPS code

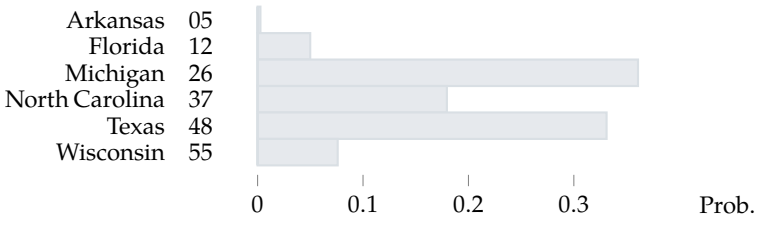
Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	8,311,547	88.5
Invalid values	925,278	9.9
Missing values	149,777	1.6

pro_st_juris_fips

Label	State 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 identifying 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	9,386,602	100.0
Valid records	9,386,602	100.0

Histogram of valid pro_st_juris_fips records

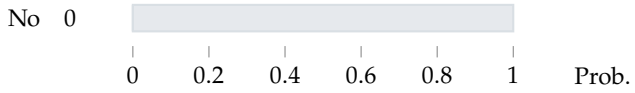


pro_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	9,386,602	100.0
Valid records	9,386,602	100.0

Histogram of valid `pro_rec_src_le` records



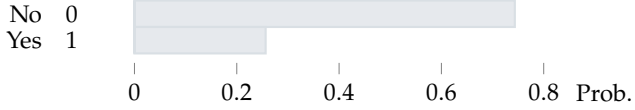
pro_rec_src_crt

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	9,386,602	100.0
Valid records	9,386,602	100.0

Statistic	*Value*
Mean	0.3
Maximum	1.0

Histogram of valid `pro_rec_src_crt` records



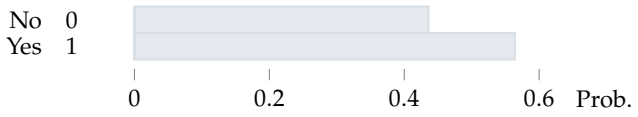
pro_rec_src_doc

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	9,386,602	100.0
Valid records	9,386,602	100.0

Statistic	*Value*
Mean	0.6
Median	1.0
Maximum	1.0

Histogram of valid pro_rec_src_doc records



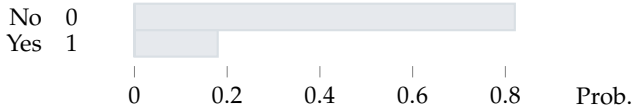
pro_rec_src_rep

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	9,386,602	100.0
Valid records	9,386,602	100.0

Statistic	*Value*
Mean	0.2
Maximum	1.0

Histogram of valid pro_rec_src_rep records

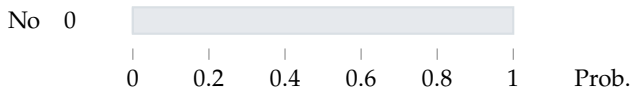


pro_rec_src_cc

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	9,386,602	100.0
Valid records	9,386,602	100.0

Histogram of valid pro_rec_src_cc records



3.6.5 Parole

cjars_id

Label	CJARS identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 4.
Table	Parole
Format	string

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0

upasn

Label	Parole term identifier
Description	Uniquely identifies terms of parole. For more details on use of `upasn` for data linkage, refer to Section 3.2.2 and Figure 5.
Table	Parole
Format	string

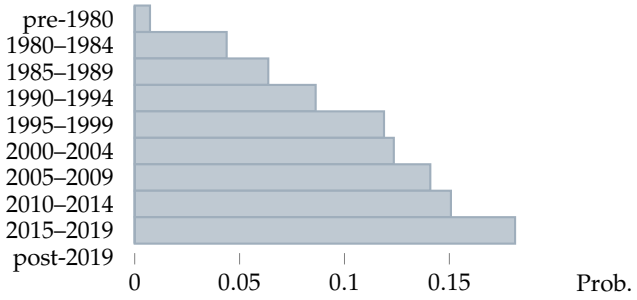
par_bgn_dt_yyyy

Label	Year of start of parole
Description	The year when the individual began parole.
Table	Parole
Format	numeric

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	2,026,872	91.6
Invalid values	1	0.0
Missing values	184,853	8.4

Statistic	*Value*
Mean	2003.6
Median	2005.0
Minimum	1013
Maximum	2019

Histogram of valid par_bgn_dt_yyyy records



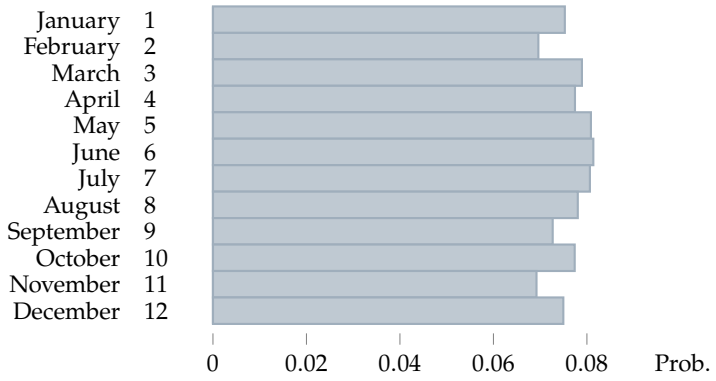
par_bgn_dt_mm

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

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	2,026,873	91.6
Missing values	184,853	8.4

Statistic	*Value*
Mean	6.5
Median	6.0
Minimum	1
Maximum	12

Histogram of valid par_bgn_dt_mm records



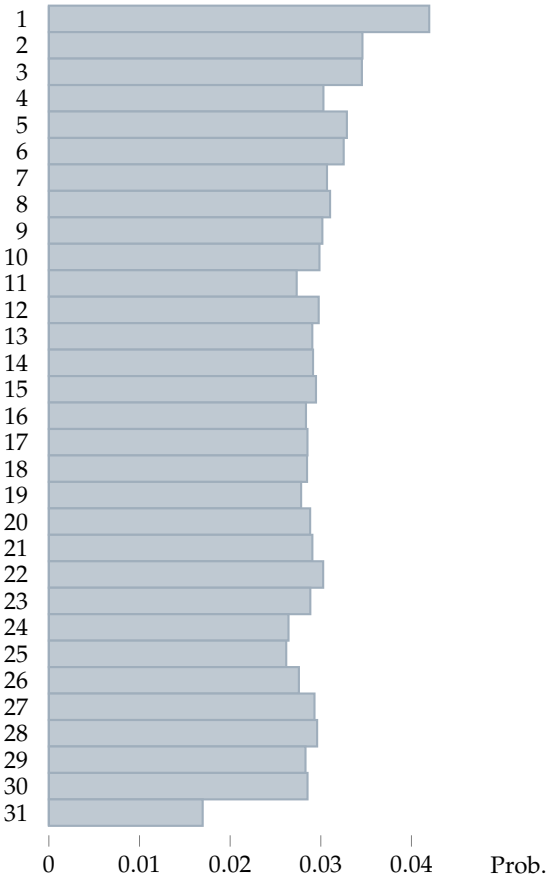
par_bgn_dt_dd

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

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	2,026,873	91.6
Missing values	184,853	8.4

Statistic	*Value*
Mean	15.2
Median	15.0
Minimum	1
Maximum	31

Histogram of valid par_bgn_dt_dd records



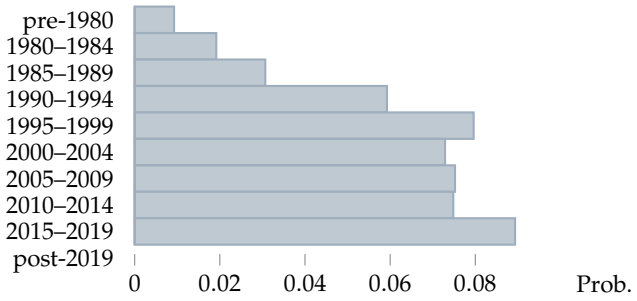
par_end_dt_yyyy

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

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	1,128,546	51.0
Invalid values	1	0.0
Missing values	1,083,179	49.0

Statistic	*Value*
Mean	2002.7
Median	2004.0
Minimum	1013
Maximum	2019

Histogram of valid par_end_dt_yyyy records



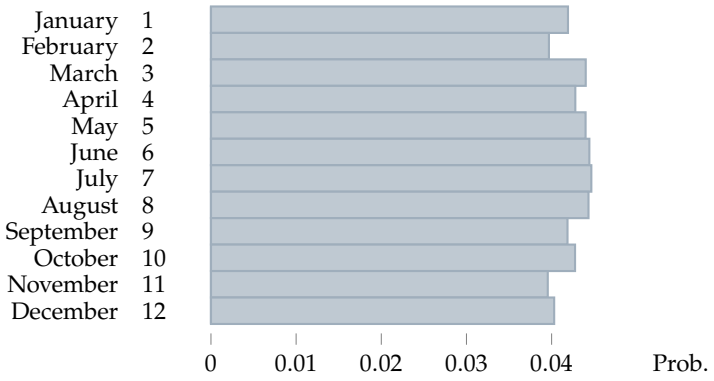
par_end_dt_mm

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

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	1,128,547	51.0
Missing values	1,083,179	49.0

Statistic	*Value*
Mean	6.5
Median	6.0
Minimum	1
Maximum	12

Histogram of valid par_end_dt_mm records



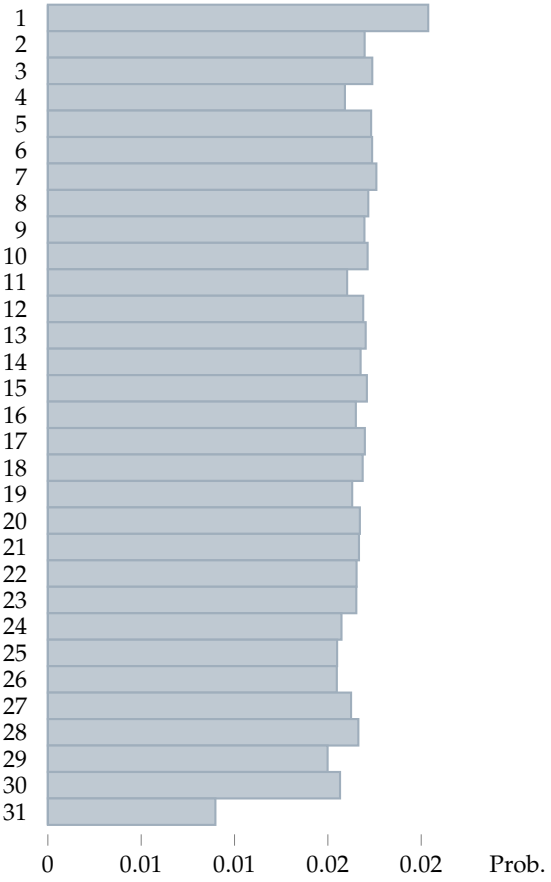
par_end_dt_dd

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

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	1,128,547	51.0
Missing values	1,083,179	49.0

Statistic	*Value*
Mean	15.5
Median	15.0
Minimum	1
Maximum	31

Histogram of valid par_end_dt_dd records

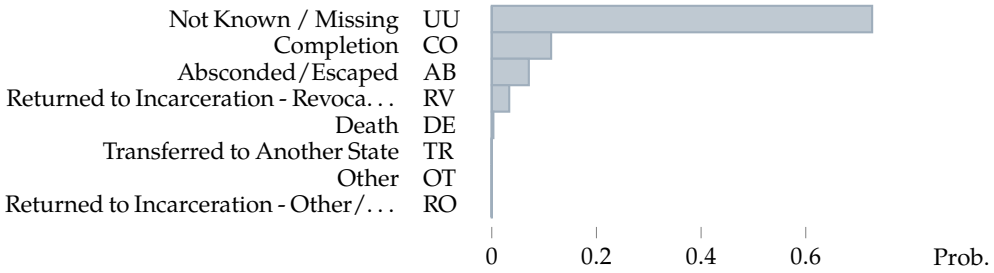


par_end_cd

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 B and C.2.9.
Table	Parole
Format	string
Code scheme	parole exit code

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	2,095,782	94.8
Missing values	115,944	5.2

Histogram of valid par_end_cd records



par_end_cd_src

Label	Raw description of parole end status from source
Description	A descrption 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

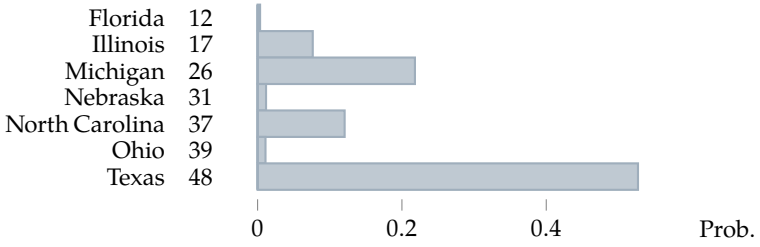
Set	*Count*	*Percent (%)*
All records	2,211,726	100.0

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 identifying 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	2,211,726	100.0
Valid records	2,141,331	96.8
Missing values	70,395	3.2

Histogram of valid `par_st_ori_fips` 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 identifying geographic areas in the United States. Three-digit code used to identify counties within states.
Table	Parole
Format	string
Code scheme	county FIPS code

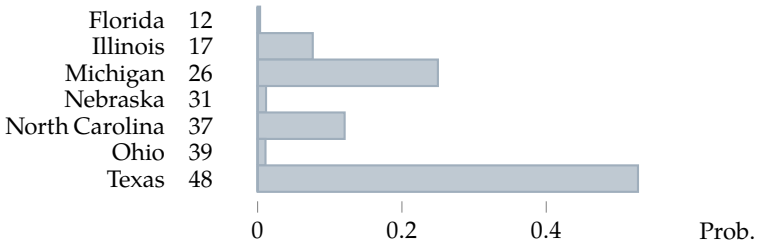
Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	949,246	42.9
Invalid values	1,192,085	53.9
Missing values	70,395	3.2

par_st_juris_fips

Label	State 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 identifying 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	2,211,726	100.0
Valid records	2,211,726	100.0

Histogram of valid `par_st_juris_fips` records

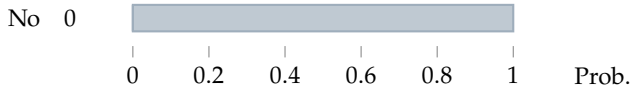


par_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	2,211,726	100.0
Valid records	2,211,726	100.0

Histogram of valid `par_rec_src_le` records

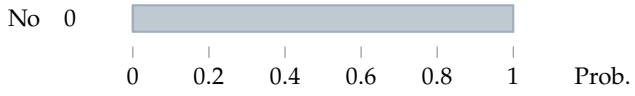


par_rec_src_crt

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	2,211,726	100.0
Valid records	2,211,726	100.0

Histogram of valid `par_rec_src_crt` 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	2,211,726	100.0
Valid records	2,211,726	100.0

Statistic	*Value*
Mean	0.9
Median	1.0
Maximum	1.0

Histogram of valid par_rec_src_doc records

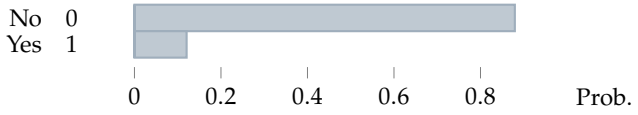
par_rec_src_rep

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	2,211,726	100.0
Valid records	2,211,726	100.0

Statistic	*Value*
Mean	0.1
Maximum	1.0

Histogram of valid `par_rec_src_rep` records

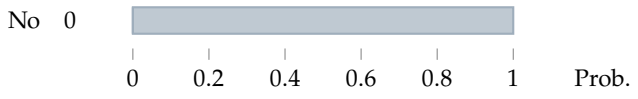


par_rec_src_cc

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	2,211,726	100.0
Valid records	2,211,726	100.0

Histogram of valid `par_rec_src_cc` records



3.7 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 these three. Refer to Sections 3.2.1 and 3.2.2 for more information about how these linkages are structured. Table 2 also provides a description of the key linking variable in each CJARS relational database and how it can be used to link with records in other CJARS relational databases.

Table 2: Variables to link tables

CJARS relational table	Key identifying event variable	Preceding event	Linking variable for preceding event	Description of linkage
arrest	`uarn`	none	none	n/a
adjudication	`ucn`	arrest	uarn	The `ucn` is a unique identifier of a court case filing that can be linked to the arrest(s) that led to the case filing via the `uarn`.
incarceration	`uisn`	adjudication	ucn	The `uisn` is a unique identifier of a term of incarceration that can be linked to the court case filing(s) that led to the incarceration term via the `ucn`.
probation	`uprsn`	adjudication	ucn	The `uprsn` is a unique identifier of a term of probation that can be linked to the court case filing(s) that led to the probation term via the `ucn`.
parole	`upasn`	incarceration	uisn	The `upasn` is a unique identifier of a term of parole that can be linked to the incarceration term(s) that led to the parole term via the `uisn`.
master person file	`cjars_id`	n/a	n/a	The `cjars_id` is a unique identifier for individuals. It can be used to connect events across the relational databases to an individual.

Appendices

A Data sources

Table 3: Agencies whose data have been integrated into CJARS

State	Level	Domain	Provider
Arkansas	State	Corrections	Arkansas Department of Corrections
Arizona	State	Corrections	Arizona Department of Corrections
Arizona	State	Judiciary	Arizona Administrative Office of the Courts
California	County	Sheriff	Nevada County Sheriff's Office
California	Municipal	Police	Anaheim Police Department
California	Municipal	Police	Bakersfield Police Department
California	Municipal	Police	Long Beach Police Department
California	Municipal	Police	Los Angeles Police Department
California	Municipal	Police	Riverside Police Department
California	Municipal	Police	San Diego Police Department
California	Municipal	Police	Stockton Police Department
Connecticut	State	Corrections	Connecticut Department of Correction
Florida	County	Judiciary	Hillsborough County Clerk of Courts
Florida	State	Corrections	Florida Department of Corrections
Illinois	State	Corrections	Illinois Department of Corrections
Michigan	State	Corrections	Michigan Department of Corrections
Michigan	State	Judiciary	Michigan State Court Administrative Office
Minnesota	State	Judiciary	Minnesota State Court Administrator's Office
Mississippi	State	Corrections	Mississippi Department of Corrections
North Carolina	State	Repository	North Carolina Department of Public Safety
North Dakota	State	Judiciary	North Dakota Court System
Nebraska	State	Corrections	Nebraska Department of Correctional Services
New Jersey	State	Corrections	New Jersey Department of Corrections
New Jersey	State	Judiciary	Superior Court of New Jersey
Ohio	State	Corrections	Ohio Department of Rehabilitation and Correction
Oregon	State	Judiciary	Oregon Judicial Department
Pennsylvania	State	Judiciary	Administrative Office of Pennsylvania Courts
Texas	County	Judiciary	Bexar County Clerk
Texas	County	Judiciary	Bexar County District Clerk
Texas	County	Sheriff	Bexar County Sheriff's Office
Texas	County	Judiciary	Collin County Courts
Texas	County	Sheriff	Collin County Sheriff's Office
Texas	County	Sheriff	Dallas County Sheriff's Office
Texas	County	Judiciary	El Paso County Clerk
Texas	County	Judiciary	Harris County District Clerk
Texas	County	Sheriff	Harris County Sheriff's Office
Texas	County	Judiciary	Hays County Courts at Law
Texas	County	Sheriff	Hays County Sheriff's Office
Texas	County	Sheriff	Tarrant County Sheriff's Office
Texas	Municipal	Police	Fort Worth Police Department
Texas	State	Corrections	Texas Department of Criminal Justice
Texas	State	Repository	Texas Department of Public Safety
Texas	State	Judiciary	iDocket
Washington	State	Corrections	Washington State Department of Corrections

State	Level	Domain	Provider
Wisconsin	State	Corrections	Wisconsin Department of Corrections
Wisconsin	State	Judiciary	Wisconsin Court System

B 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 this 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.

B.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 7 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: PII and anonymized. The PII data is used for matching purposes while the anonymized data contains the information that is relevant for describing criminal justice events.

Harmonization begins when data is obtained from data provides 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 personally identifiable information to prepare for entity resolution (see Appendix D) and further harmonization.

The PII 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 anonymized goes into the anonymized division for further processing. This involves variable harmonization, episode resolution (see Appendix D), 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 C). Variable harmonization is a complex task that employs many techniques including hand coding, employing the use of regular expression commands, and machine learning techniques.

Offense descriptions are standardized to the offense classification scheme found in Appendix C. For classifying the string offense descriptions, a multi-level classification model for predicting "parent group" (broader offense category) and for predicting "child group" (more detailed offense code). For parent group classification, `CountVectorizer` is used to generate a matrix of token counts while `TfidfVectorizer` (Term Frequency-Inverse Document Frequency) is used for generating features in child group classification. The design rationale for this multi-level model is to allow the model to focus on most common words or phrases during the parent group classification, and to give more weight to less frequent words or phrases during the child group classification. During parent group classification, only one multi-layer perceptron model is trained to predict all of the parent group classes. Afterwards, one multi-layer perceptron classifier is trained for each parent group. As a result, if there are three parent groups (e.g. Violent Offense, Drug Offense, Property Offense), then the algorithm will train three multi-layer perceptron models during the child group classification.

More details about the harmonization of other variables in each of the five CJARS relational databases are included below in Table 4.

Figure 7: CJARS variable harmonization process

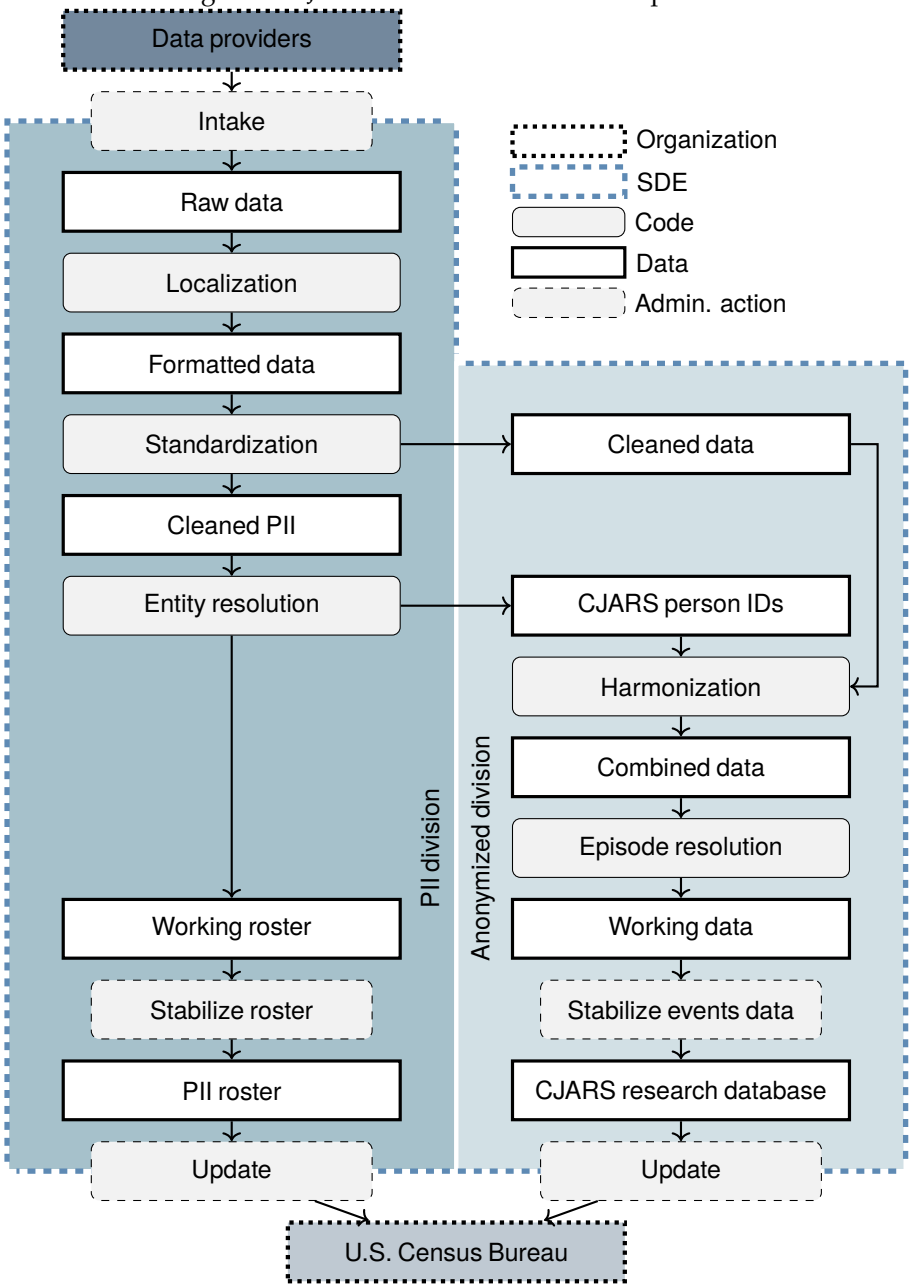


Table 4: Variable harmonization

CJARS relational database	Information	Variable(s)	Coding method(s)
arrest	arrest date	arr_arr_dt_yyyy, arr_arr_dt_mm, arr_arr_dt_dd	recorded as separate variables for year, month, and day
arrest	booking date	arr_book_dt_yyyy, arr_book_dt_mm, arr_book_dt_dd	split into year,month, and day
arrest	offense classification	arr_off_cd	machine learning employed to generate a standardized offense type, (see Appendix B.1 for more details).
adjudication	offense grade	adj_grd_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 classification	adj_chrg_off_cd	machine learning employed to generate a standardized offense type, (see Appendix B.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 B.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
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

CJARS relational database	Information	Variable(s)	Coding method(s)
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

C Code schemes including crime classifications

C.1 Geographic codes

C.1.1 State FIPS and abbreviations

This scheme is used for the following variables:

- adj_st_ori_fips
- arr_st_ori_fips
- inc_st_ori_fips
- inc_st_juris_fips
- par_st_ori_fips
- par_st_juris_fips
- pro_st_ori_fips
- pro_st_juris_fips

Table 5: 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
30	MT	Montana
31	NE	Nebraska
32	NV	Nevada
33	NH	New Hampshire
34	NJ	New Jersey

State FIPS	State abbreviation	State name
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.

C.1.2 County FIPS

This scheme is used for the following variables:

- `adj_cnty_ori_fips`
- `arr_cnty_ori_fips`
- `inc_cnty_ori_fips`
- `par_cnty_ori_fips`
- `pro_cnty_ori_fips`

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

C.2 Criminal justice event codes

C.2.1 Offense classification

This scheme is used for the following variables:

- adj_chrg_off_cd
- adj_disp_off_cd
- arr_off_cd

Table 6: Offense classification scheme

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
010	Murder	1	Murder	1	Violent
011	Attempted Murder	1	Murder	1	Violent
012	Conspiracy to Commit Murder	1	Murder	1	Violent
015	Unspecified Homicide	2	Unspecified homicide	1	Violent
016	Unspecified Homicide, Attempted	2	Unspecified homicide	1	Violent
017	Unspecified Homicide, Conspiracy	2	Unspecified homicide	1	Violent
020	Voluntary Manslaughter	3	Voluntary/nonnegligent manslaughter	1	Violent
021	Voluntary Manslaughter, Attempted	3	Voluntary/nonnegligent manslaughter	1	Violent
022	Voluntary Manslaughter,Conspiracy	3	Voluntary/nonnegligent manslaughter	1	Violent
025	Vehicular Manslaughter	3	Voluntary/nonnegligent manslaughter	1	Violent
026	Vehicular Manslaughter, Attempted	3	Voluntary/nonnegligent manslaughter	1	Violent
027	Vehicular Manslaughter, Conspiracy	3	Voluntary/nonnegligent manslaughter	1	Violent
030	Involuntary Manslaughter	4	Manslaughter - non-vehicular	1	Violent
031	Involuntary Manslaughter, Attempt	4	Manslaughter - non-vehicular	1	Violent
032	Involuntary Manslaughter, Conspiracy	4	Manslaughter - non-vehicular	1	Violent
035	Kidnapping	5	Kidnapping	1	Violent
036	Kidnapping, Attempted	5	Kidnapping	1	Violent
037	Kidnapping, Conspiracy	5	Kidnapping	1	Violent
040	Rape	6	Rape - force	1	Violent
041	Rape, Attempted	6	Rape - force	1	Violent
042	Rape, Conspiracy	6	Rape - force	1	Violent
045	Statutory Rape	7	Rape - statutory - no force	1	Violent
046	Statutory Rape, Attempted	7	Rape - statutory - no force	1	Violent
047	Statutory Rape, Conspiracy	7	Rape - statutory - no force	1	Violent
050	Child Molestation	8	Lewd act with children	1	Violent
051	Child Molestation, Attempted	8	Lewd act with children	1	Violent
052	Child Molestation, Conspiracy	8	Lewd act with children	1	Violent
055	Sexual Assault	9	Sexual assault - other	1	Violent
056	Sexual Assault, Attempted	9	Sexual assault - other	1	Violent
057	Sexual Assault, Conspiracy	9	Sexual assault - other	1	Violent
060	Human Trafficking, Sex - child	10	Human Trafficking	1	Violent
065	Human Trafficking, Sex - adult or no age specified	10	Human Trafficking	1	Violent

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
070	Human Trafficking, Labor - child	10	Human Trafficking	1	Violent
075	Human Trafficking, Labor - adult or no age specified	10	Human Trafficking	1	Violent
080	Human Trafficking, Unspecified - child	10	Human Trafficking	1	Violent
085	Human Trafficking, Unspecified - adult or no age specified	10	Human Trafficking	1	Violent
086	Human Trafficking, Attempted	10	Human Trafficking	1	Violent
087	Human Trafficking, Conspiracy	10	Human Trafficking	1	Violent
090	Armed Robbery	11	Armed robbery	1	Violent
091	Armed Robbery, Attempted	11	Armed robbery	1	Violent
092	Armed Robbery, Conspiracy	11	Armed robbery	1	Violent
095	Unarmed Robbery	12	Unarmed robbery	1	Violent
096	Unarmed Robbery, Attempted	12	Unarmed robbery	1	Violent
097	Unarmed Robbery, Conspiracy	12	Unarmed robbery	1	Violent
100	Aggravated Assault	13	Aggravated assault	1	Violent
101	Aggravated Assault, Attempted	13	Aggravated assault	1	Violent
102	Aggravated Assault, Conspiracy	13	Aggravated assault	1	Violent
105	Assault of an Officer	14	Assaulting public officer	1	Violent
106	Assault of an Officer, Attempted	14	Assaulting public officer	1	Violent
107	Assault of an Officer, Conspiracy	14	Assaulting public officer	1	Violent
110	Child Abuse	15	Child abuse	1	Violent
111	Child Abuse, Attempted	15	Child abuse	1	Violent
112	Child Abuse, Conspiracy	15	Child abuse	1	Violent
115	Simple Assault	16	Simple assault	1	Violent
116	Simple Assault, Attempted	16	Simple assault	1	Violent
117	Simple Assault, Conspiracy	16	Simple assault	1	Violent
120	Extortion/Threat	17	Blackmail/extortion/intimidation	1	Violent
121	Extortion/Threat, Attempted	17	Blackmail/extortion/intimidation	1	Violent
122	Extortion/Threat, Conspiracy	17	Blackmail/extortion/intimidation	1	Violent
125	Hit and Run with Bodily Injury	18	Hit and run driving - injury	1	Violent
126	Hit and Run with Bodily Injury, Attempted	18	Hit and run driving - injury	1	Violent
127	Hit and Run with Bodily Injury, Conspiracy	18	Hit and run driving - injury	1	Violent
130	Violent Offense, Other	19	Violent offenses - other	1	Violent
131	Violent Offense Other, Attempted	19	Violent offenses - other	1	Violent
132	Violent Offense Other, Conspiracy	19	Violent offenses - other	1	Violent
135	Burglary	20	Burglary	2	Property
136	Burglary, Attempted	20	Burglary	2	Property
137	Burglary, Conspiracy	20	Burglary	2	Property
140	Arson	21	Arson	2	Property
141	Arson, Attempted	21	Arson	2	Property
142	Arson, Conspiracy	21	Arson	2	Property
145	Auto Theft	22	Auto theft	2	Property
146	Auto Theft, Attempted	22	Auto theft	2	Property
147	Auto Theft, Conspiracy	22	Auto theft	2	Property

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
150	Forgery/Fraud	23	Forgery/fraud	2	Property
151	Forgery/Fraud, Attempted	23	Forgery/fraud	2	Property
152	Forgery/Fraud, Conspiracy	23	Forgery/fraud	2	Property
155	Grand Theft (>$500)	24	Grand larceny - theft over $500	2	Property
156	Grand Theft (>$500), Attempted	24	Grand larceny - theft over $500	2	Property
157	Grand Theft (>$500), Conspiracy	24	Grand larceny - theft over $500	2	Property
160	Petty Theft (=<$500)	25	Petty larceny - theft equal or under $500	2	Property
161	Petty Theft (=<$500), Attempted	25	Petty larceny - theft equal or under $500	2	Property
162	Petty Theft (=<$500), Conspiracy	25	Petty larceny - theft equal or under $500	2	Property
165	Theft, Value Unknown	26	Larceny/theft - value unknown	2	Property
166	Theft, Value Unknown, Attempted	26	Larceny/theft - value unknown	2	Property
167	Theft, Value Unknown, Conspiracy	26	Larceny/theft - value unknown	2	Property
170	Financial Crimes	27	Financial Crimes	2	Property
171	Financial Crimes Attempted	27	Financial Crimes	2	Property
172	Financial Crimes Conspiracy	27	Financial Crimes	2	Property
175	Sale of Stolen Property	28	Stolen property - trafficking	2	Property
176	Sale of Stolen Property, Attempted	28	Stolen property - trafficking	2	Property
177	Sale of Stolen Property, Conspiracy	28	Stolen property - trafficking	2	Property
180	Receiving Stolen Property	29	Stolen property - receiving	2	Property
181	Receiving Stolen Property, Attempted	29	Stolen property - receiving	2	Property
182	Receiving Stolen Property, Conspiracy	29	Stolen property - receiving	2	Property
185	Destruction of Property	30	Destruction of property	2	Property
186	Destruction of Property, Attempted	30	Destruction of property	2	Property
187	Destruction of Property, Conspiracy	30	Destruction of property	2	Property
190	Hit and Run Driving with Property Damage	31	Hit and run driving - property damage	2	Property
191	Hit and Run Driving, Attempted	31	Hit and run driving - property damage	2	Property
192	Hit and Run Driving, Conspiracy	31	Hit and run driving - property damage	2	Property
195	Unauthorized use of Vehicle	32	Unauthorized use of vehicle	2	Property
196	Unauthorized use of Vehicle, Attempted	32	Unauthorized use of vehicle	2	Property
197	Unauthorized use of Vehicle, Conspiracy	32	Unauthorized use of vehicle	2	Property
200	Criminal Trespass	33	Trespassing	2	Property
201	Criminal Trespass, Attempted	33	Trespassing	2	Property
202	Criminal Trespass, Conspiracy	33	Trespassing	2	Property
205	Other Property Offense	34	Property offenses - other	2	Property
206	Other Property Offense, Attempt	34	Property offenses - other	2	Property
207	Other Property Offense, Conspiracy	34	Property offenses - other	2	Property
210	Possession of Property Crime Tools	34	Property offenses - other	2	Property
211	Possession of Property Crime Tools, Attempted	34	Property offenses - other	2	Property
212	Possession of Property Crime Tools, Conspiracy	34	Property offenses - other	2	Property
215	Distribution Heroin	35	Distribution - heroin	3	Drug
216	Distribution, Heroin, Attempted	35	Distribution - heroin	3	Drug
217	Distribution, Heroin, Conspiracy	35	Distribution - heroin	3	Drug

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
220	Distribution of amphetamines	36	Distribution - amphetamines	3	Drug
221	Distribution of amphetamines, Attempted	36	Distribution - amphetamines	3	Drug
222	Distribution of amphetamines, Conspiracy	36	Distribution - amphetamines	3	Drug
225	Distribution Cocaine or Crack	37	Distribution - cocaine or crack	3	Drug
226	Distribution Cocaine or Crack, Attempted	37	Distribution - cocaine or crack	3	Drug
227	Distribution Cocaine or Crack, Conspiracy	37	Distribution - cocaine or crack	3	Drug
230	Distribution of opioids	38	Distribution of opioids	3	Drug
231	Distribution of opioids, Attempted	38	Distribution of opioids	3	Drug
232	Distribution of opioids, Conspiracy	38	Distribution of opioids	3	Drug
235	Distribution of prescription drugs	39	Distribution of prescription drugs	3	Drug
236	Distribution of prescription drugs, Attempted	39	Distribution of prescription drugs	3	Drug
237	Distribution of prescription drugs, Conspiracy	39	Distribution of prescription drugs	3	Drug
240	Distribution Other Controlled Substances	40	Distribution - other controlled substances	3	Drug
241	Distribution Other Controlled Substances, Attempted	40	Distribution - other controlled substances	3	Drug
242	Distribution Other Controlled Substances, Conspiracy	40	Distribution - other controlled substances	3	Drug
245	Distribution Marijuana	41	Distribution marijuana/hashish	3	Drug
246	Distribution Marijuana, Attempted	41	Distribution marijuana/hashish	3	Drug
247	Distribution Marijuana, Conspiracy	41	Distribution marijuana/hashish	3	Drug
250	Distribution, Drug Unspecified	42	Distribution - drug unspecified	3	Drug
251	Distribution, Drug Unspecified, Attempted	42	Distribution - drug unspecified	3	Drug
252	Distribution, Drug Unspecified, Conspiracy	42	Distribution - drug unspecified	3	Drug
255	Possession/Use of Heroin	43	Possession/use - heroin	3	Drug
256	Possession/Use of Heroin, Attempted	43	Possession/use - heroin	3	Drug
257	Possession/Use of Heroin, Conspiracy	43	Possession/use - heroin	3	Drug
260	Possession of amphetamines	44	Possession of amphetamines	3	Drug
261	Possession of amphetamines, Attempted	44	Possession of amphetamines	3	Drug
262	Possession of amphetamines, Conspiracy	44	Possession of amphetamines	3	Drug
265	Possession/Use of Cocaine or Crack	45	Possession/use - cocaine or crack	3	Drug
266	Possession/Use of Cocaine or Crack, Attempted	45	Possession/use - cocaine or crack	3	Drug
267	Possession/Use of Cocaine or Crack, Conspiracy	45	Possession/use - cocaine or crack	3	Drug
270	Possession of opioids	46	Possession of opioids	3	Drug
271	Possession of opioids, Attempted	46	Possession of opioids	3	Drug

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
272	Possession of opioids, Conspiracy	46	Possession of opioids	3	Drug
275	Possession of prescription drugs	47	Possession of prescription drugs	3	Drug
276	Possession of prescription drugs, Attempted	47	Possession of prescription drugs	3	Drug
277	Possession of prescription drugs, Conspiracy	47	Possession of prescription drugs	3	Drug
280	Possession/Use of Other Controlled Substance	48	Possession/use - other controlled substances	3	Drug
281	Possession/Use of Other Controlled Substance, Attempted	48	Possession/use - other controlled substances	3	Drug
282	Possession/Use of Other Controlled Substance, Conspiracy	48	Possession/use - other controlled substances	3	Drug
285	Possession/Use of Marijuana	49	Possession/use - marijuana/hashish	3	Drug
286	Possession/Use of Marijuana, Attempted	49	Possession/use - marijuana/hashish	3	Drug
287	Possession/Use of Marijuana, Conspiracy	49	Possession/use - marijuana/hashish	3	Drug
290	Possession/Use of Unspecified Drug	50	Possession/use - drug unspecified	3	Drug
291	Possession/Use, Drug Unspecified, Attempted	50	Possession/use - drug unspecified	3	Drug
292	Possession/Use, Drug Unspecified, Conspiracy	50	Possession/use - drug unspecified	3	Drug
295	Heroin Violation, Offense Unspecified	51	Heroin violation - offense unspecified	3	Drug
300	Amphetamines, Offense unspecified	52	Amphetamines - offense unspecified	3	Drug
305	Cocaine/Crack Violation, Offense Unspecified	53	Cocaine or crack violation offense unspecified	3	Drug
310	Prescription of opioid drugs, offense unspecified	54	Prescription of opioid drugs - offense unspecified	3	Drug
315	Prescription, offense unspecified	55	Prescription - offense unspecified	3	Drug
320	Other Controlled Substance Violation, Offense Unspecified	56	Controlled substance - offense unspecified	3	Drug
325	Marijuana Violation, Offense Unspecified	57	Marijuana/hashish violation - offense unspecified	3	Drug
330	Fraudulent Drug Offense	58	Other Drug Offense/Paraphernalia	3	Drug
331	Fraudulent Drug Offense, Attempted	58	Other Drug Offense/Paraphernalia	3	Drug
332	Fraudulent Drug Offense, Conspiracy	58	Other Drug Offense/Paraphernalia	3	Drug
335	Drug Paraphernalia	58	Other Drug Offense/Paraphernalia	3	Drug
336	Drug Paraphernalia, Attempted	58	Other Drug Offense/Paraphernalia	3	Drug
337	Drug Paraphernalia, Conspiracy	58	Other Drug Offense/Paraphernalia	3	Drug
340	Other Drug Offense	58	Other Drug Offense/Paraphernalia	3	Drug
345	Driving While Intoxicated	59	Driving while intoxicated	4	DUI Offense
346	Driving While Intoxicated, Attempted	59	Driving while intoxicated	4	DUI Offense
347	Driving While Intoxicated, Conspiracy	59	Driving while intoxicated	4	DUI Offense
350	Driving Under the Influence of Alcohol	60	Driving Under the Influence	4	DUI Offense
351	Driving Under the Influence of Alcohol, Attempted	60	Driving Under the Influence	4	DUI Offense

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
352	Driving Under the Influence of Alcohol, Conspiracy	60	Driving Under the Influence	4	DUI Offense
355	Driving Under the Influence of Drugs	61	Driving under influence - drugs	4	DUI Offense
356	Driving Under the Influence of Drugs, Attempted	61	Driving under influence - drugs	4	DUI Offense
357	Driving Under the Influence of Drugs, Conspiracy	61	Driving under influence - drugs	4	DUI Offense
360	Riot	62	Rioting	5	Public Order
361	Riot, Attempting to Incite	62	Rioting	5	Public Order
362	Riot, Conspiracy to Incite	62	Rioting	5	Public Order
365	Escape from Custody	63	Escape from custody	5	Public Order
366	Escape from Custody, Attempted	63	Escape from custody	5	Public Order
367	Escape from Custody, Conspiracy	63	Escape from custody	5	Public Order
370	Flight to Avoid Prosecution	64	Flight to avoid prosecution	5	Public Order
371	Flight to Avoid Prosecution, Attempted	64	Flight to avoid prosecution	5	Public Order
372	Flight to Avoid Prosecution, Conspiracy	64	Flight to avoid prosecution	5	Public Order
375	Weapons Offense	65	Weapon offense	5	Public Order
376	Weapons Offense, Attempted	65	Weapon offense	5	Public Order
377	Weapons Offense, Conspiracy	65	Weapon offense	5	Public Order
380	Habitual Offender	66	Habitual offender	5	Public Order
385	Parole Violation	67	Parole violation	5	Public Order
390	Probation Violation	68	Probation violation	5	Public Order
400	Contempt of Court/Violate Court Order	69	Contempt of court	5	Public Order
401	Contempt of Court/Violate Court Order, Attempted	69	Contempt of court	5	Public Order
402	Contempt of Court/Violate Court Order, Conspiracy	69	Contempt of court	5	Public Order
405	Other Court Offense	70	Offenses against courts, legislatures and commissions	5	Public Order
406	Other Court Offense, Attempted	70	Offenses against courts, legislatures and commissions	5	Public Order
407	Other Court Offense, Conspiracy	70	Offenses against courts, legislatures and commissions	5	Public Order
410	Family or Custody Related Offense	71	Family related offenses	5	Public Order
411	Family or Custody Related Offense, Attempted	71	Family related offenses	5	Public Order
412	Family or Custody Related Offense, Conspiracy	71	Family related offenses	5	Public Order
415	Offense Against Morals/Decency	72	Morals/decency - offense	5	Public Order
416	Offense Against Morals/Decency, Attempted	72	Morals/decency - offense	5	Public Order
417	Offense Against Morals/Decency, Conspiracy	72	Morals/decency - offense	5	Public Order
420	Immigration Violation	73	Immigration violations	5	Public Order
421	Immigration Violation, Attempted	73	Immigration violations	5	Public Order
422	Immigration Violation, Conspiracy	73	Immigration violations	5	Public Order

Charge code	Charge code description	Offense category code	Offense category description	Offense type code	Offense type description
425	Obstruction/Resisting	74	Obstruction - law enforcement	5	Public Order
426	Obstruction/Resisting, Attempted	74	Obstruction - law enforcement	5	Public Order
427	Obstruction/Resisting, Conspiracy	74	Obstruction - law enforcement	5	Public Order
430	Invasion of Privacy	75	Invasion of privacy	5	Public Order
431	Invasion of Privacy, Attempted	75	Invasion of privacy	5	Public Order
432	Invasion of Privacy, Conspiracy	75	Invasion of privacy	5	Public Order
435	Commercialized Vice	76	Commercialized vice	5	Public Order
436	Commercialized Vice, Attempted	76	Commercialized vice	5	Public Order
437	Commercialized Vice, Conspiracy	76	Commercialized vice	5	Public Order
440	Contributing to the Delinquency of a Minor	77	Contributing to delinquency of a minor	5	Public Order
445	Disorderly Conduct Offense	78	Drunkenness/Vagrancy/Disorderly Conduct	5	Public Order
446	Disorderly Conduct Offense, Attempted	78	Drunkenness/Vagrancy/Disorderly Conduct	5	Public Order
447	Disorderly Conduct Offense, Conspiracy	78	Drunkenness/Vagrancy/Disorderly Conduct	5	Public Order
450	Liquor Law Violation	79	Liquor law violations	5	Public Order
451	Liquor Law Violation, Attempted	79	Liquor law violations	5	Public Order
452	Liquor Law Violation, Conspiracy	79	Liquor law violations	5	Public Order
455	Taxation Offense	80	Taxation Offenses	5	Public Order
456	Taxation Offense, Attempted	80	Taxation Offenses	5	Public Order
457	Taxation Offense, Conspiracy	80	Taxation Offenses	5	Public Order
460	Bribery/Conflict of Interest	81	Bribery and conflict of interest	5	Public Order
461	Bribery/Conflict of Interest, Attempt	81	Bribery and conflict of interest	5	Public Order
462	Bribery/Conflict of Interest, Conspiracy	81	Bribery and conflict of interest	5	Public Order
465	Public Order Offense, Other	82	Public order offenses - other	5	Public Order
466	Public Order Offense, Other, Attempted	82	Public order offenses - other	5	Public Order
467	Public Order Offense, Other, Conspiracy	82	Public order offenses - other	5	Public Order
470	Juvenile Offense	83	Juvenile offenses	8	Exclude
475	Traffic Offense, Minor	84	Traffic offenses - minor	6	Criminal traffic
480	Unspecified Felony	85	Felony - unspecified	999	Not known/Missing
481	Unspecified Felony, Attempt	85	Felony - unspecified	999	Not known/Missing
482	Unspecified Felony, Conspiracy	85	Felony - unspecified	999	Not known/Missing
485	Unspecified Misdemeanor	86	Misdemeanor - unspecified	999	Not known/Missing
486	Unspecified Misdemeanor, Attempt	86	Misdemeanor - unspecified	999	Not known/Missing
487	Unspecified Misdemeanor, Conspiracy	86	Misdemeanor - unspecified	999	Not known/Missing
490	Other Offense	87	Other	7	Other
500	Flag for Removal	500	Flag for Removal	500	Flag for Removal
505	Call for service	505	Call for service	505	Call for service
510	Federal charges	510	Federal charges	510	Federal charges
777	Variable not available in county	777	Variable not available in county	777	Variable not available in county
888	Not applicable	888	Not applicable	888	Not applicable
999	Not Known/Missing	999	Not Known/Missing	999	Not Known/Missing

C.2.2 Offense charge grade

This scheme is used for the following variables:

- adj_grd_cd

Table 7: Charge grade classification scheme

Charge grade code	Charge grade description
FE	Felony-level charge
MI	Misdemeanor-level charge
LO	Local ordinance or other low-level charge
UU	Not Known / Missing
JF	Juvenile felony-level charge
JM	Juvenile misdemeanor-level charge
JL	Juvenile local ordinance or other low-level charge
JU	Juvenile Not Known / Missing

C.2.3 Court disposition

This scheme is used for the following variables:

- adj_disp_cd

Table 8: 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
NP	N	P	Not Guilty Plea
NU	N	U	Not Guilty - Unclassified
PT	P	T	Procedural - Transfer
PU	P	U	Procedural - Unclassified
UU			Not Known / Missing

C.2.4 Probation conditions

This scheme is used for the following variables:

- pro_cond_cd

Table 9: Probation conditions classification scheme

Probation condition code	Probation condition description
PJ	Probation With Jail

Probation condition code	Probation condition description
SP	Straight Probation
AD	Alcohol/Drug Residential
PR	Probation With Community Residential
UU	Not Known / Missing

C.2.5 Probation exit

This scheme is used for the following variables:

- pro_end_cd

Table 10: 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

C.2.6 Incarceration entry

This scheme is used for the following variables:

- inc_entry_cd

Table 11: 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
PN	Probation Revocation - No New Sentence
OT	Other

Incarceration entry code	Incarceration entry description
PR	Probation Status - Pending Revocation
UC	Unsentenced Commitment
IE	Illegal Entry
UU	Not Known / Missing

C.2.7 Incarceration facility type

This scheme is used for the following variables:

- inc_fcl_cd

Table 12: 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

C.2.8 Incarceration exit

This scheme is used for the following variables:

- inc_exit_cd

Table 13: 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
RA	Release on Appeal or Bond
OT	Other

Incarceration exit code	Incarceration exit description
EA	Escape/AWOL
AI	Accidental Injury to Self
IE	Illegal Entry
UU	Not Known / Missing

C.2.9 Parole exit

This scheme is used for the following variables:

- par_end_cd

Table 14: 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

D 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.

D.1 Entity resolution to identify unique 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 biometrically linked records obtained 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 biometric 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.

D.2 Episode resolution to link sequences of events

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.

E Record linkage at the U.S. Census Bureau

E.1 Record linkage rates

Completed roster files and CJARS databases are assigned a version number before distribution. The data and complete code base are permanently archived to maintain long-term reproducibility.

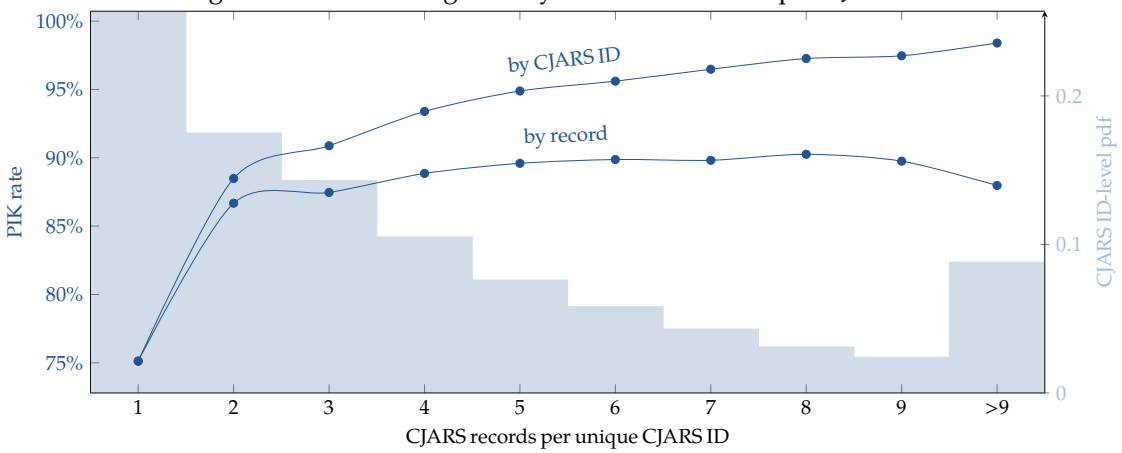
The roster file and anonymized criminal justice database are then transferred securely to the Census Bureau, where they are processed by the Person Identification Validation System (PVS), the Census Bureau's probabilistic record linkage system that assigns anonymized Personal Identification Keys (PIKs) to sensitive microdata records (Wagner and Layne 2014). When the CJARS roster file is processed by PVS, unique CJARS IDs can be linked to unique PIKs. Since PIKs are used for all record linkage throughout the Census Bureau, this allows the CJARS data to be linked to the full set of data held by the Census Bureau Data Linkage Infrastructure.

Record linkage rates at the Census Bureau are a function of the quality of Personally Identifiable Information (PII) as well as the degree to which the underlying population intersects with the reference file that the Census Bureau uses for linkage, which is based on the Social Security Administration's Numident file. For example, arrests records have lower quality PII on average, so we expect lower match rates. And in regions with a higher number of immigrants *who have not been assigned Social Security Numbers or Individual Taxpayer Identification Numbers*, we also expect lower match rates.

Conversely, we expect the match rate to increase for individuals for which CJARS has multiple event records from multiple agencies. These individuals are more likely to have biometric ids, and are more likely to have had their PII updated.

Figure 8 shows these predictions borne out in record linkage at the Census Bureau. The X-axis identifies the number of records received by CJARS for a unique individual, identified by a CJARS ID. The bars show the distribution of CJARS IDs by the number of records. The lines show the proportion of records (by CJARS ID or event record) that have been assigned a PIK at the Census Bureau. When CJARS has only a single record for an individual, the PIK rate is about 75%. When a second record is available, the PIK rate jumps about 85%. This trend increases until individuals with more than nine records have PIK rates of about 98%. We see a decline in the PIK rate when the rate is calculated by event record, which is likely caused by some superclusters of records which all contain poor PII. As CJARS grows, we expect the PIK rate to increase as indicated in the figure.

Figure 8: Record linkage rate by number of records per CJARS ID



Source: Authors' calculations from the CJARS PIK crosswalk. All results were approved for release by the U.S. Census Bureau Disclosure Review Board (DRB), authorization number CBDRB-FY19-371. All numerators and denominators were rounding according to DRB rounding rules for unweighted counts.

F Variable availability by jurisdiction

F.1 Notes on data availability from Arizona

Variables	Notes
	Data collected from the AZ DOC was scraped from records made publicly available through their Inmate Datasearch system. This system does not record date of birth. 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 (e.g., geographically), these gaps also impact the coverage of DOC records. What is more, the timeframe of data coverage of the Inmate Datasearch system is uncertain. Communications with the AZ DOC in May of 2020 indicated that they estimate the system contains historical records dating back approximately 15 years, but this is only their best approximation.

F.2 Notes on data availability from Florida

Variables	Notes
`inc_exit_dt_dd`	There is a concentration of dates on the first of the month. This may reflect a standard release date, or it may reflect data warehouse processes that record some types events as always occurring on the first of the month. See Blomberg et al. (2011, p. 20).

F.3 Notes on data availability from Michigan

Variables	Notes
	Probation records in Michigan were collected from the state's Department of Corrections. The responsibility of probation supervision in Michigan is split between the Department of Corrections (felony convictions) and the counties (non-felony convictions). For this reason, our coverage of probation records in Michigan only includes felony-level convictions.

F.4 Notes on data availability from Minnesota

Variables	Notes
	The adjudication relational table only includes cases files in the state that resulted in a conviction.

F.5 Notes on data availability from North Carolina

Variables	Notes
adj_sent_rest	There are a relatively small number of negative values. This may represent data input errors, payments, elimination of outstanding payments, etc. These negative values were left as is because their explanation is unknown.
adj_sent_fine	There are a relatively small number of negative values. This may represent data input errors, payments, elimination of outstanding payments, etc. These negative values were left as is because their explanation is unknown.

F.6 Notes on data availability from North Dakota

Variables	Notes
adj_sent_fine	There are a relatively small number of negative values. This may represent data input errors, payments, elimination of outstanding payments, etc. These negative values were left as is because their explanation is unknown.

F.7 Notes on data availability from Nebraska

Variables	Notes
adj_disp_dt_yyyy, adj_disp_dt_mm, adj_disp_dt_dd, adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd	The data acquired from the Nebraska Department of Corrections only includes the beginning and end dates of the sentence term. It does not include the disposition date or the date of sentencing.

F.8 Notes on data availability from New Jersey

Variables	Notes
adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd, 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	The data provided by the Superior Court of New Jersey includes information at the charge-level for each case, with the exception of sentencing information. Sentencing information is only recorded at the case-level. This leads to situations where cases have both relatively less serious and relatively more serious charges, but the sentencing is recorded the same for all charges, which can make it appear as though a severe sentence is associated with a low-level offense.

F.9 Notes on data availability from Pennsylvania

Variables	Notes
adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd, 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	The data acquired from the Administrative Office of Pennsylvania Courts does not currently include sentencing information. Case data are limited to offense, offense date, file date, and disposition information.

F.10 Notes on data availability from Texas

Variables	Notes
adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd, 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	Filing date and sentencing information is missing for all records.

F.11 Notes on data availability from Wisconsin

Variables	Notes
inc_cnty_ori_fips	Data acquired through public information request from the Wisconsin Department of Corrections did not include county of conviction. County of conviction was obtained from the DOC website for a small number of individuals in the active caseload.
adj_sent_fine	There are a relatively small number of negative values. This may represent data input errors, payments, elimination of outstanding payments, etc. These negative values were left as is because their expanation is unknown.

G State computerized criminal history systems

As part of the process of developing a national CJARS data schema, we surveyed the data systems that states use to track individuals and events of the criminal justice system. We received data schemas from 18 states. This section briefly summarizes those schemas and provides an overview figure of the data structures. In some cases, we did not receive an explicit data schema, but a less formal summary of the system. For these states, the models are described as "presumed" to indicate that we are attempting to reconstruct a schema from an informal description.

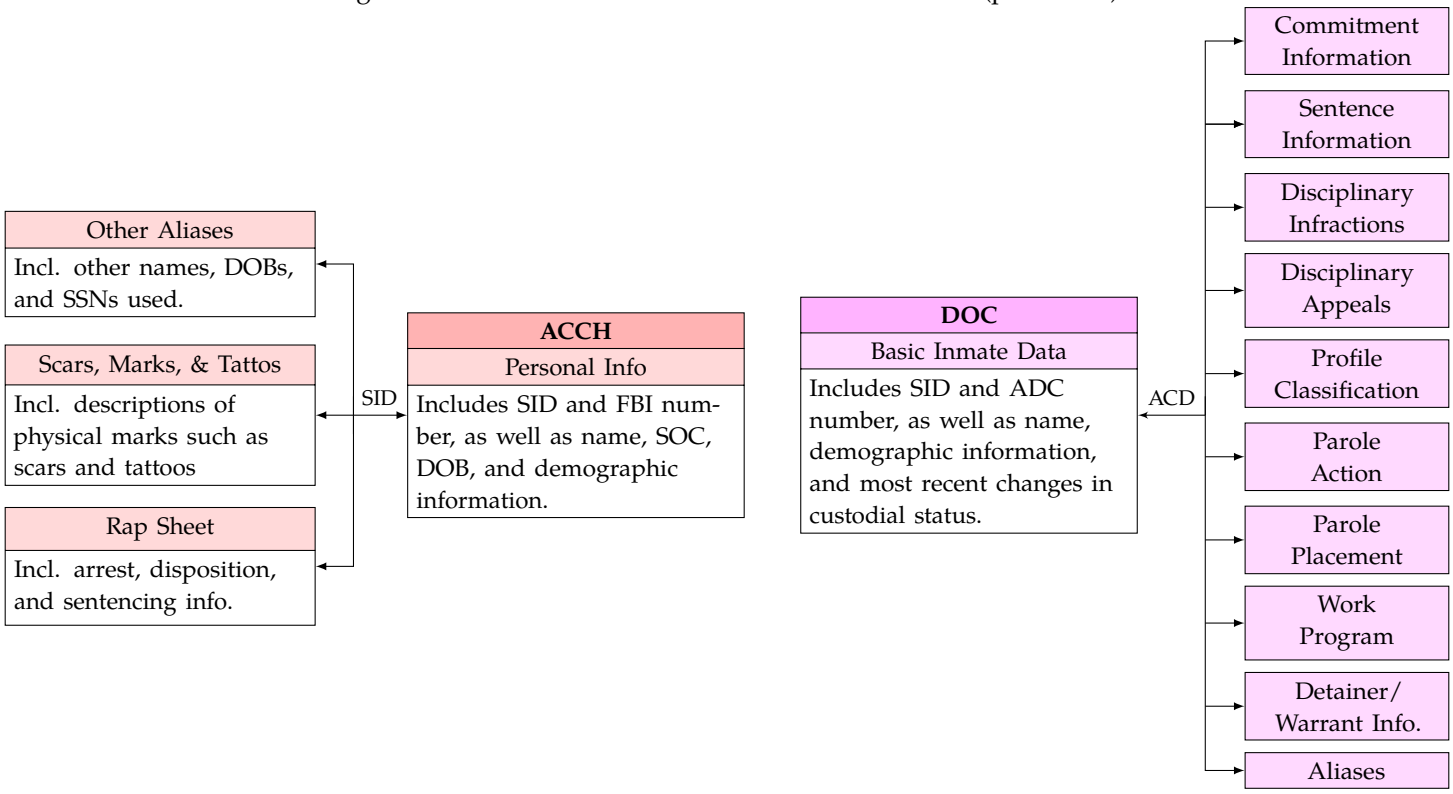
G.1 Arizona

Arizona does not have one single comprehensive criminal justice database. The Arizona Computerized Criminal History (ACCH) is housed within the Criminal History Records Section of the Arizona Department of Public Safety. The ACCH contains information about arrests (including arrest date and offense), dispositions (including date and offense), and sentence summary variables (confinement, fine, restitution, etc.). We believe that ACCH is maintained at the arrest level. Information in the ACCH files is taken directly from arrest fingerprint cards, disposition report forms, and court order information that is submitted to the Central State Repository (CSR) by law enforcement and criminal justice agencies throughout Arizona. CSR employees then enter the information into the appropriate ACCH file.

The Arizona Department of Corrections (ADC) maintains a separate database called the Automated Inmate Management System (AIMS). The online, searchable version of this database includes information on inmates who were in the custody of ADC in 1985 and anyone who has entered into the custody of ADC since then. The database includes over 112,000 inmate records, including around 26,000 records of active inmates. Inmates are identified by a 6-digit ADC number. An inmate record includes: basic inmate data, commitment information, sentence information, profile classification, infractions, parole action, parole placement, work program, detainer/warrant information, and aliases.

The ACCH and AIMS could theoretically be linked, but our understanding is that, given the current structure of these databases, a given arrest is not associated with a particular incarceration spell. Figure 9 approximates our understanding of the layout of the two databases.

Figure 9: Arizona ACCH & DOC AIMS database model (presumed)



G.2 California

California's Bureau of Criminal Information and Analysis (housed within the Department of Justice (DOJ)) maintains the Automated Criminal History System (ACHS).[1] ACHS is an Oracle database consisting of 87

[1]The California DOJ maintains additional data, including the Monthly Arrest and Citations Record (MACR) database, which covers adult (18+) arrests in California from 1980 to the present. This database has a record of individuals who were arrested and then released

"entities" (tables). These tables contain information relating to all stages of the criminal justice system. Figure 10 shows the names of some of the tables that contain information for each stage. It is not clear from the list of tables and their associated variables how the tables link together. It seems as if some tables are able to contain information about different types of events (e.g., Arrest vs. Probation) and have codes that indicate which type of even the stored information pertains to.

The ACHS includes individuals who are fingerprinted. ACHS retains misdemeanor arrests (with or without conviction) and felony arrests (without conviction) for 10 years from the date of arrest, misdemeanor conviction with a prior and felony convictions until the subject is 70 years old, and convictions of registrable sex offenses until the subject is 100 years old.

Figure 10: Description of some tables from California ACHS

Arrest
• COUNT (incl. arresting agency) • EVENT (incl. type of event, e.g. booking number) • STEP (incl. event date, if step code = Arrest)

Institutional Corrections
• INSTITUTION (incl. description of institution to which sentenced) • INSTITUTION NUMBER (incl. inmate number assigned by the institution)

Legal Proceedings
• CONVICTED OFFENSE (incl. offense code) • CONVICTED STATUS (incl. level of conviction) • DISPOSITION (incl. description of the disposition code) • DISPOSITION CATEGORY (incl. category into which the disposition falls)

Community Corrections
• SENTENCE TYPE (incl. length, if sentence type = Probation) • COUNT (incl. from and to where a subject was paroled)

G.3 Connecticut

The Computerized Criminal History (CCH) is maintained by the Connecticut State Police (a division of the Connecticut Department of Public Safety, which itself is housed in the Connecticut Department of Emergency Services and Public Protection). The database is maintained in an IBM DB2 UDB database on an IBM hardware platform using a relational database structure, which is reproduced below in Figure 11.

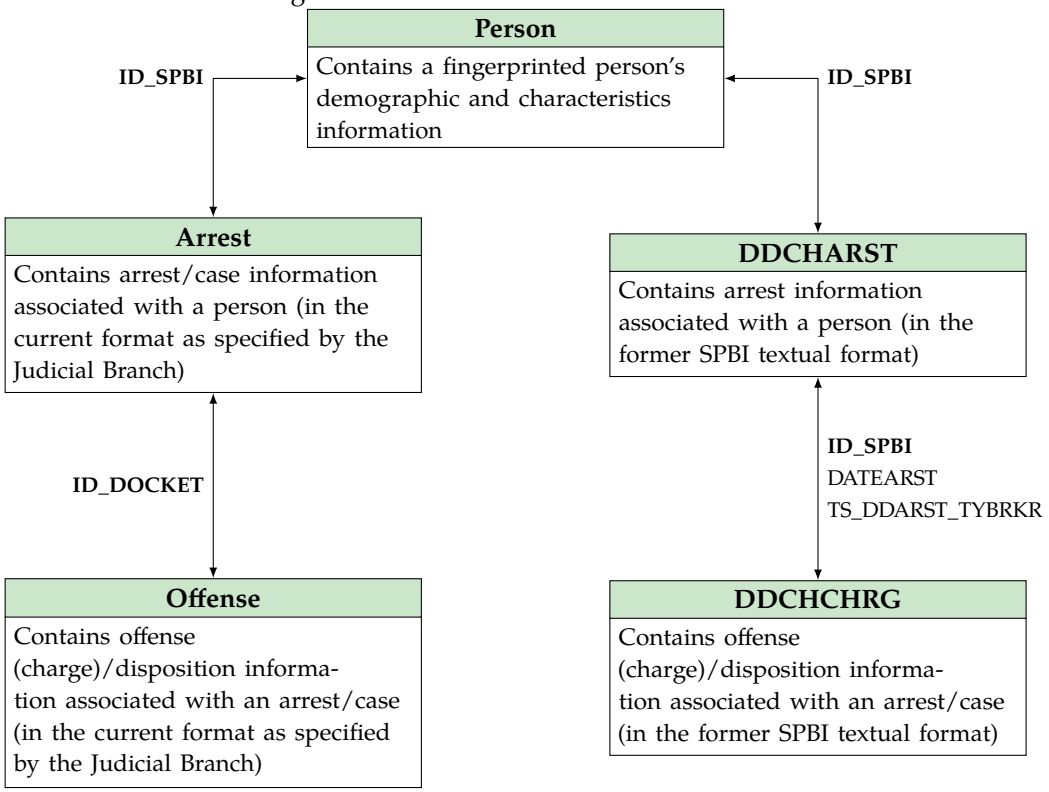
The database contains information provided to the Division of State Police by courts and criminal justice agencies. The database is not static and individual records are subject to change as new information is received. The database includes individuals who were arrested, but it is unclear if the database contains the universe of individuals who were arrested or only individuals whose case went court. We do not know how far back the database goes.

G.4 Florida

The Florida Department of Corrections (DOC) hosts a publicly available dataset for download on their website. This dataset represents a subset of the data collected by the Florida DOC, but the subset still provides a significant amount of information. The publicly available database is a Microsoft Access file with 19 data tables.

(unlike ACHS, which only includes anyone who is fingerprinted) and therefore tends to catch more misdemeanors than ACHS. The DOJ also has the JCPSS, which contains juvenile records dating back to 2003.

Figure 11: Connecticut CCH database model



As shown in Figure 12, the 19 tables can be organized by the types of people included in each table: active inmates, released inmates, and offenders currently under community supervision. Information on both current and prior offenses is included. Only individuals who were sentenced to state prison or state supervision are included in these tables. Records could be linked between datasets by DCNumber, but the records are not set-up so that it is immediate, e.g., which arrest record is associated with which incarceration spell.

G.5 Illinois

Criminal history records are maintained by the Illinois State Police in the Computerized Criminal History (CCH) database. There are currently over 1,000 Illinois policing bodies that submit arrest data to ISP. Clerk's offices submit disposition and sentencing information, while county jails and the Illinois Department of Corrections (IDOC) report custodial information. The CCH database uses fingerprint information to link arrests and custodial records for a single individual when creating a rap sheet.

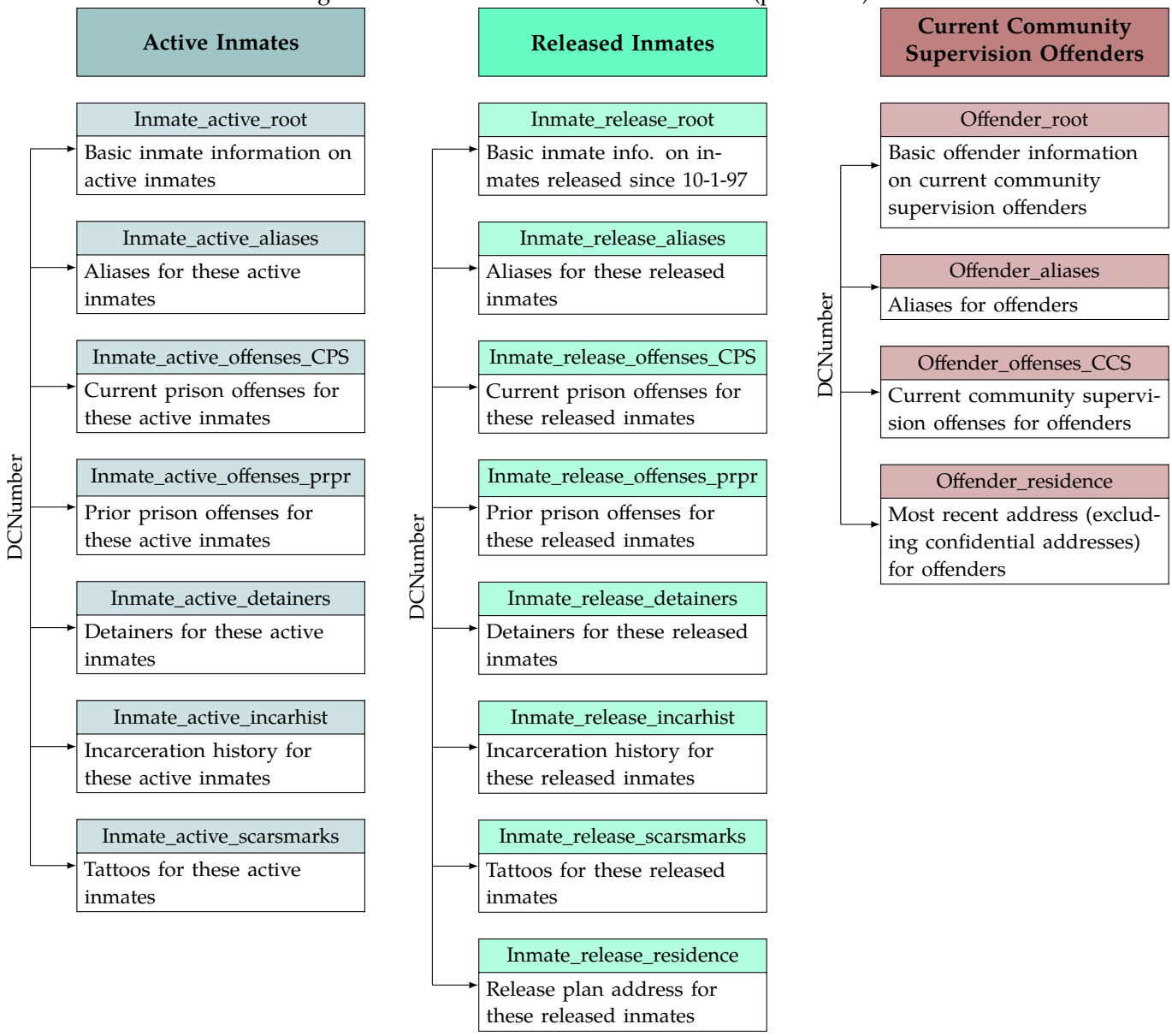
The Illinois Criminal Justice Information Authority (Authority) has access to a subset of this data called the Criminal History Record Information (CHRI) Ad Hod data (the "Ad Hoc Database").[2]. The Ad Hoc Database is stored in an Oracle Database comprised of 61 tables. The extracted CHRI Ad Hoc data is organized into five tables (see Figure 13) for each year between 1990-2005.[3] Adult and juvenile records are stored in separate yearly databases.

The Ad Hoc Database is comprised of "arrest cycles," which link arrest and disposition information for an individual based on a single arrest. The State Identification Number (SID) can be used to link all arrest events in an individual's criminal history. The Ad Hoc Database does not include the custodial information contained in CCH.

[2]Outside researchers who have signed a user's agreement with the Authority (jointly with ISP) are given access to a public version of the Ad Hoc Database for approved research purposes

[3]Databases for 1960-1989 are being developed as staff resources allow. ISP has been designated as the CHRI state central repository since 1931. At this time, only a few records from 1931-1960 have been automated.

Figure 12: Florida DOC Public Database ERD (presumed)

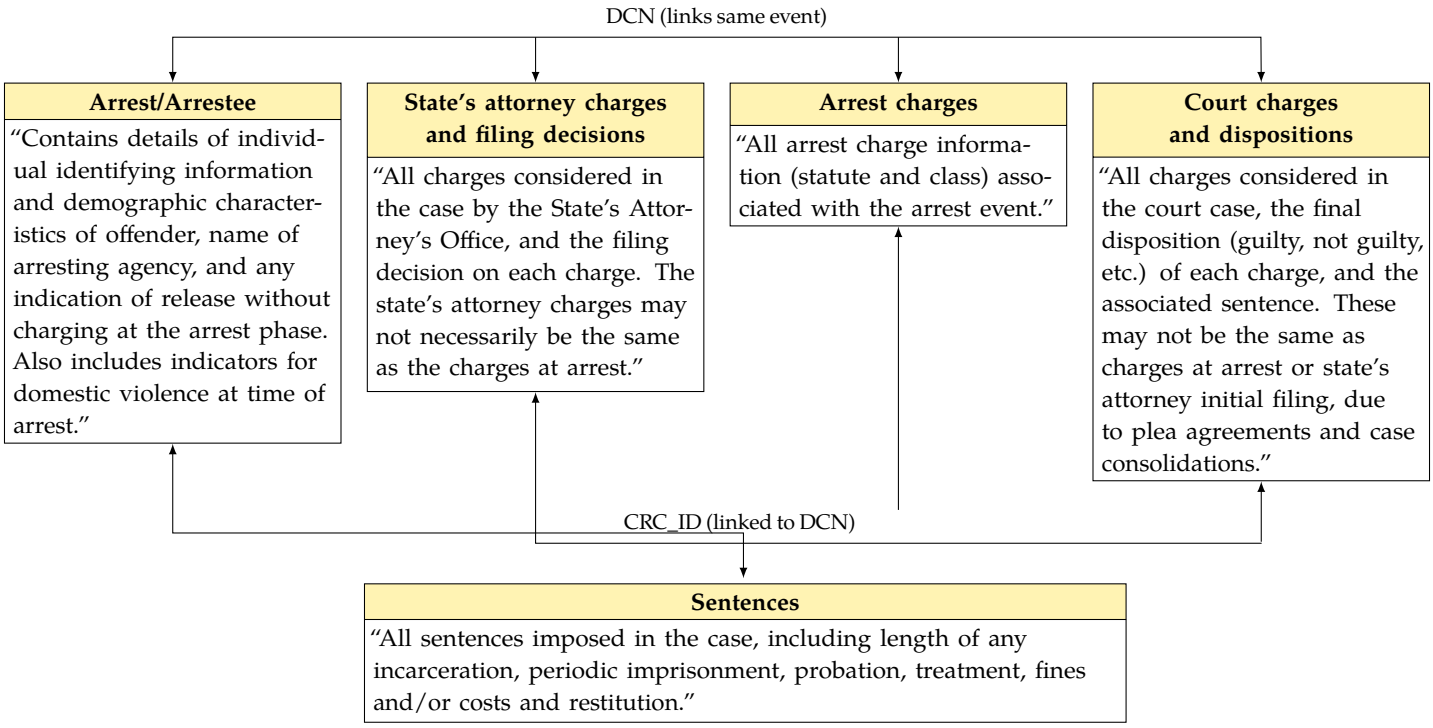


G.6 Kansas

The Kansas Bureau of Investigation (KBI) and the Kansas Highway Patrol (KHP) share the administrative and operational responsibilities for core KCJIS functions.[4] As part of this, KBI operates Kansas's Central Repository for Criminal History Record Information (CHRI). The Kansas Central Repository is a system of connected data sources that contains information about felony and misdemeanor arrests, court convictions, dispositions, and incarceration in state-operated facilities. The Central Repository database receives summary information for these events from contributing police departments, sheriff's offices, prosecutors and courts throughout the state. This information is summarized in Kansas Disposition Report (KDR; see Figure 14). KBI receives incarceration information from the Kansas Adult Supervised Population Electronic Repository

[4]Criminal history information is also collected by the Kansas Sentencing Commission (which is also the Kansas SAC). The Sentencing commission has been collecting sentencing data and probation revocation disposition data since 1998. They receive prison data from Kansas Department of Corrections (KDOC) and arresting data from Kansas Bureau of Investigation (KBI) for the prison population. The Sentencing Commission collects this information into two SPSS databases which are updated annually based on policy change.

Figure 13: Illinois CHRI Ad Hoc database model (presumed)



(KASPER), which is a website maintained by the Kansas Department of Corrections that provides current information on offenders currently incarcerated or on parole.

It is not entirely clear from the available information how the different tables link together. Figure 14 shows the Kansas Disposition Report (KDR) Class Diagrams based off of a presentation given in 2006. The ERD displayed in the presentation does not say which variables link the different tables together.

G.7 Kentucky

The Computerized Criminal History (CCH) is maintained by the Kentucky State Police. Information is contributed by arresting officers, the Administrative Office of the Courts, and prison facilities. Arrest information is originally collected by the arresting officer and a citation is created. The citation is provided to the jail, who is responsible for taking fingerprints. The information from the citation is "attached" to the fingerprints, and jointly that information makes up the arrest section of the CCH. Disposition information comes from the Administrative Office of the Courts and is listed with each arrest. When an individual is committed to a prison facility they are fingerprinted. The data from the court judgment is "attached" to those fingerprints, and jointly that data creates the Commitment section of the CCH record.

Anyone with access to the CCH system can look up an offender by SID and then flip through the data contained in the 6 tables shown in Figure 15.

G.8 Minnesota

The Criminal History System (CHS) is maintained by the Justice Information Services (MNJIS) section of the Bureau of Criminal Apprehension (BCA), which is housed within the Minnesota Department of Public Safety. In January 2014, BCA requested proposals to replace their old Computerized Criminal History (CCH) system. The old Criminal History Record information had 1,105,121 State Identification Numbers (SIDs), including 53,070 juvenile SIDs, and 1,770,195 names (74,010 juvenile).

It is not entirely clear how to follow an incident through these tables, or how these tables link together. Figure

Figure 14: Kansas Disposition Report (KDR) Class Diagrams (presumed)

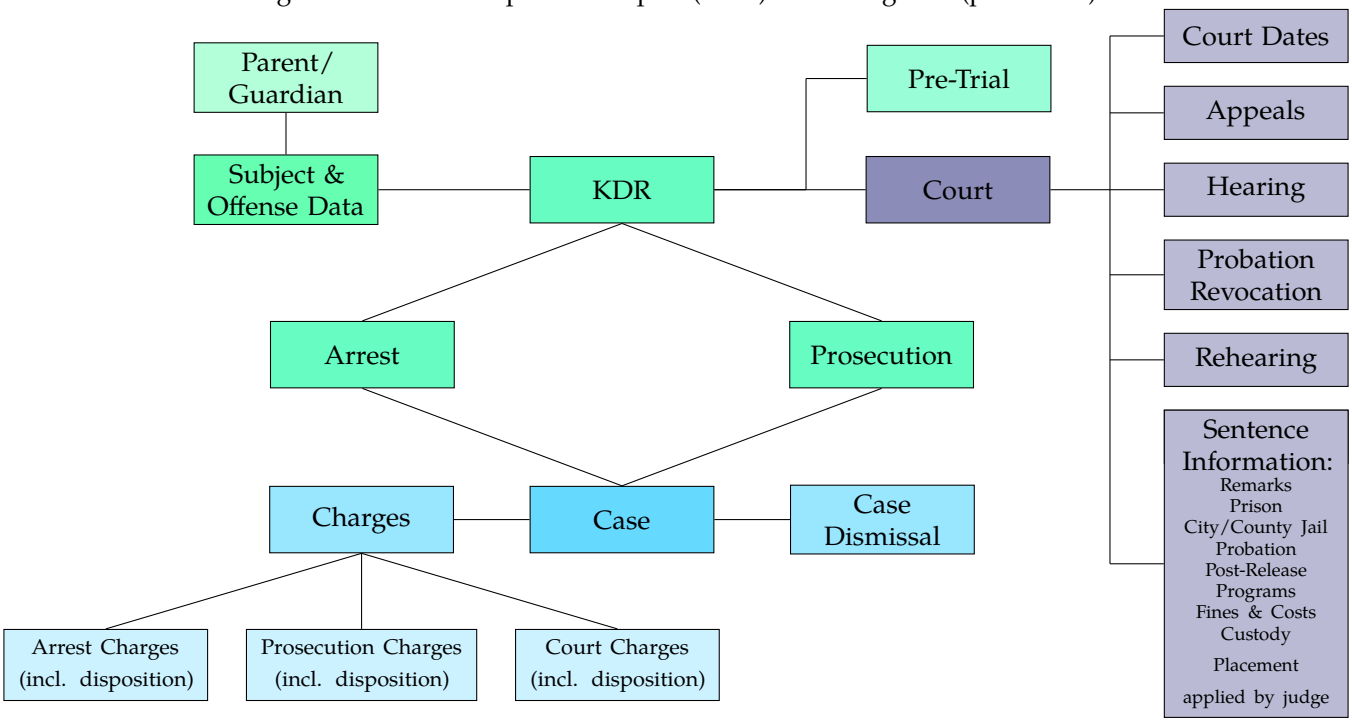
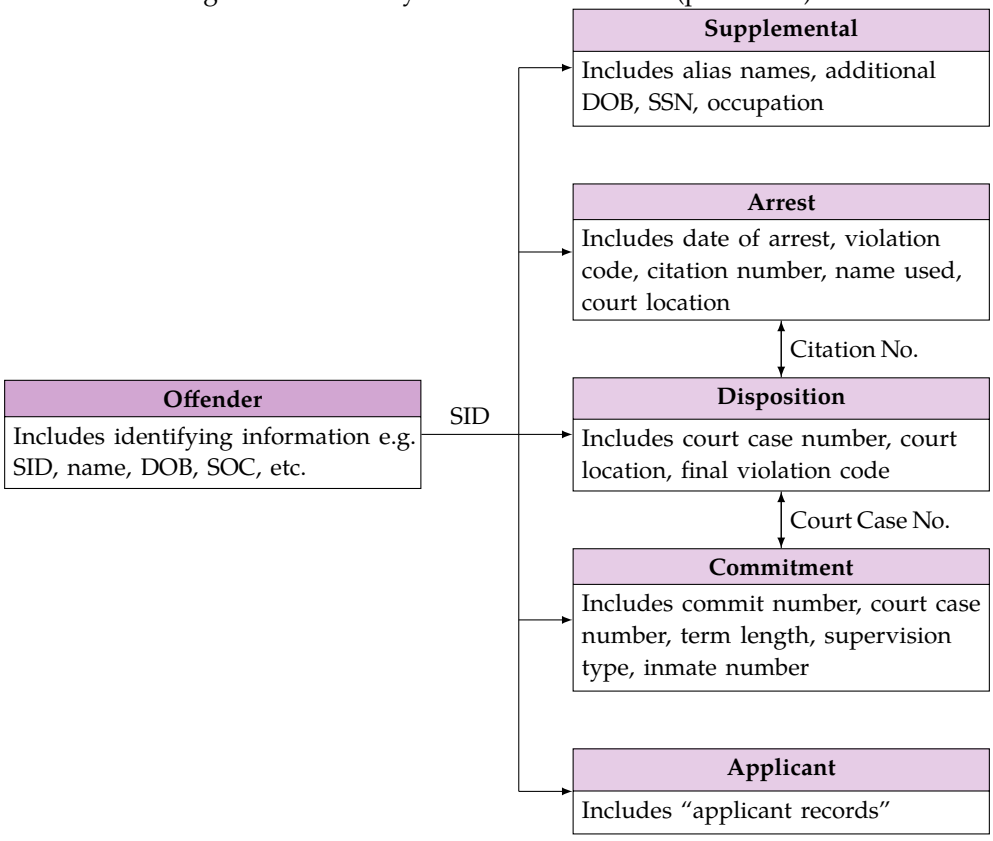
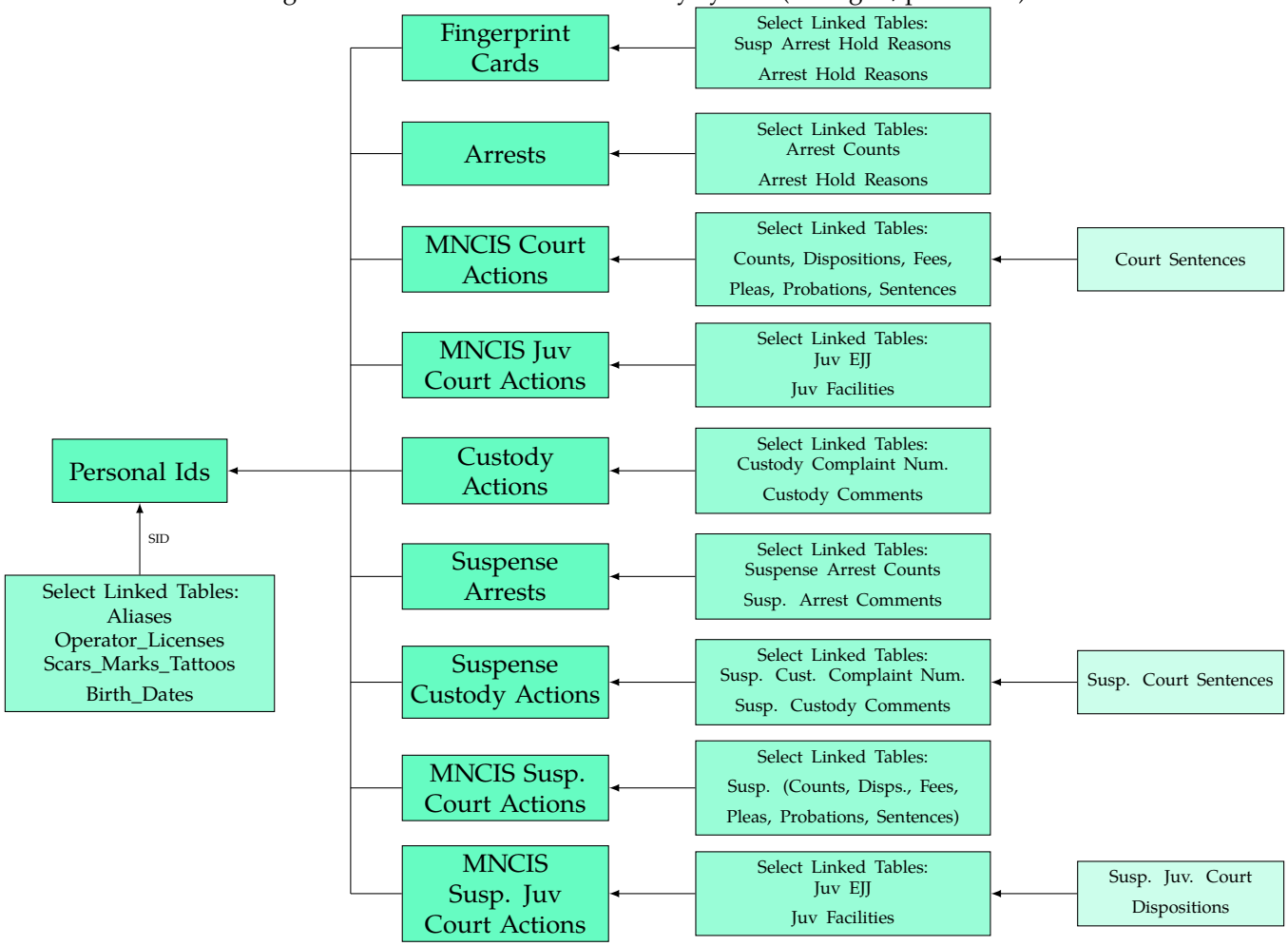


Figure 15: Kentucky CCH database model (presumed)



16 shows the Entity Relationship Diagram (ERD) for the old CCH system updated based on our understanding of the RFP documents. The old ERD diagram gave Primary and Foreign Keys for each table, but the variables listed as keys are not always included in the linked tables. The RFP documents also do not specify where BCA receives various pieces of information.

Figure 16: Minnesota Criminal History System (abridged/presumed)



G.9 Nevada

The Computerized Criminal History (CCH) is part of the Nevada Criminal Justice Information System (NCJIS) environment.[5] CCH (also referred to as the Nevada Criminal History Repository) acts as a centralized storage facility for Nevada Arrest and Disposition Records. It is maintained by the Records Bureau of the General Services Division of the Nevada Department of Public Safety on Oracle SQL Servers.

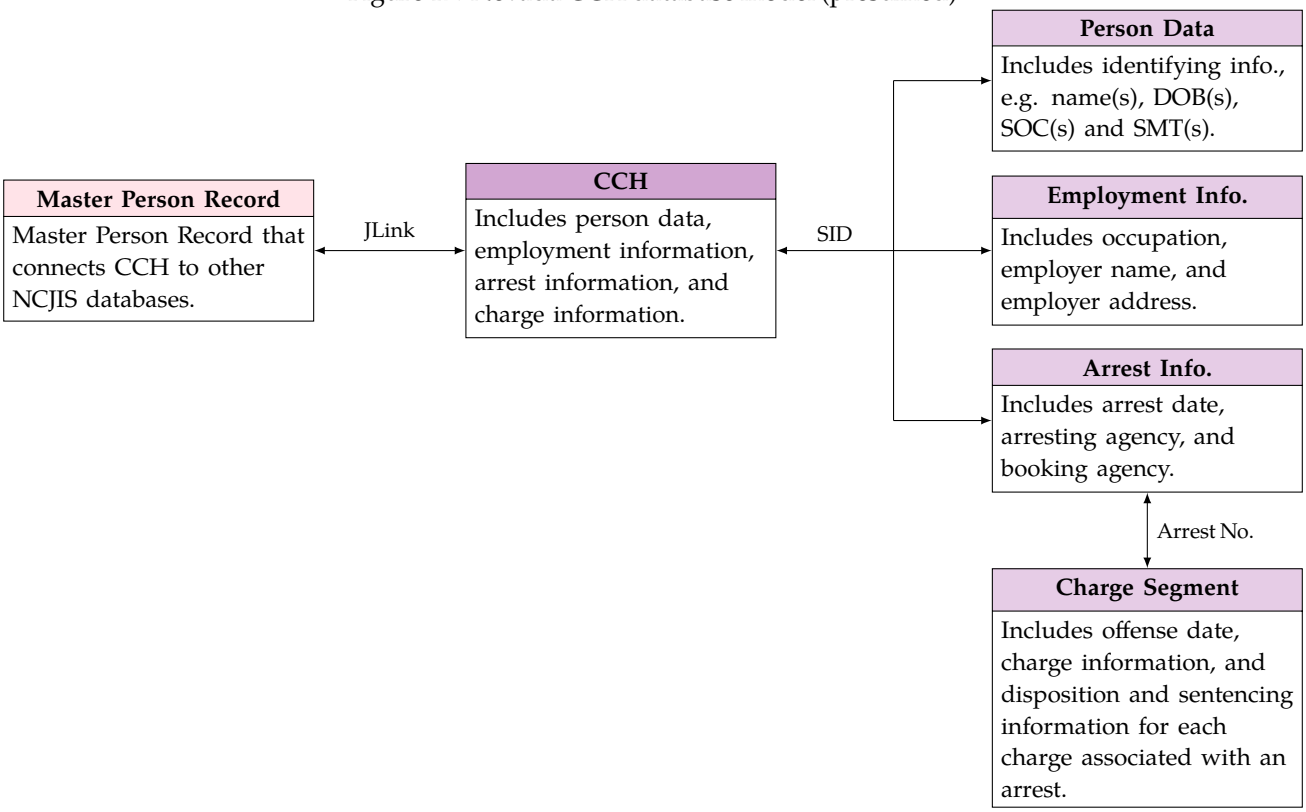
Authorized users from around the state contribute to CCH. Booking and arresting agencies collect fingerprints at the time of arrest, which are submitted to the state repository for inclusion in the criminal history. Criminal history records also contain information regarding individuals who are placed on parole or probation and supervised by the Division of Parole and Probation.

Figure 17 is based on the rap sheet and the four major categories of information represented by the data elements included in the CCH data element list. These four major categories are: Person Data, Employment

[5]NCJIS links to other systems such as the Offender Tracking Information System (OTIS) and the database of NV persons with concealed weapons permits (CCW) through a Master Person Record.

Information, Arrest Information, Charge Information. NCJIS is currently undergoing a "modernization" process that includes updates to CCH. The CCH project is scheduled to be completed by June 30, 2017.

Figure 17: Nevada CCH database model (presumed)



G.10 New Mexico

The New Mexico Department of Public Safety Information Technology Department (DPS ITD) serves all local law enforcement and criminal justice agency programs within New Mexico including state police, motor transportation division officers, and special investigations division officers. DPS ITD is responsible for a network of state-wide databases as well as maintaining links to national database systems.

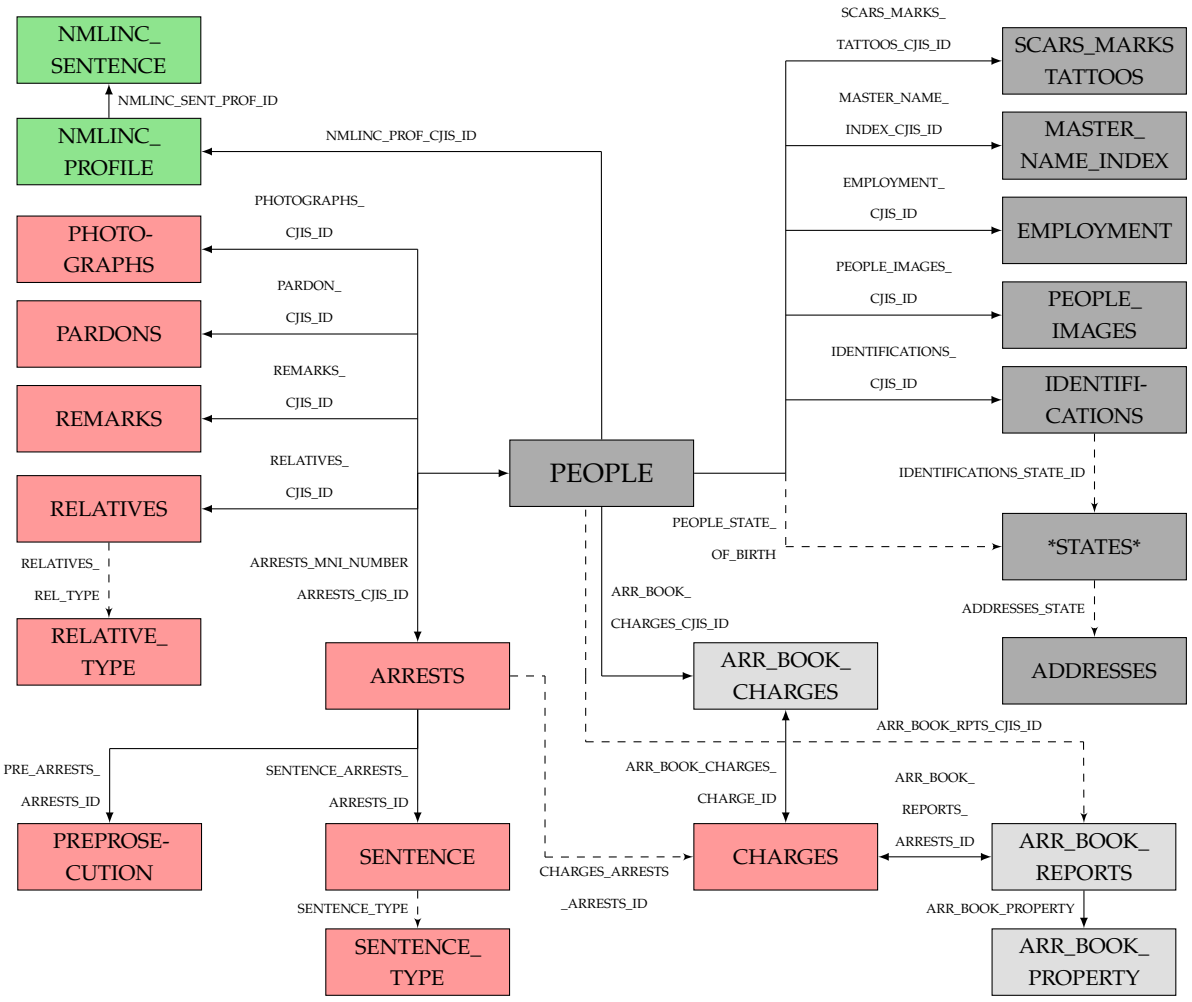
The Criminal Justice Information System (CJIS) is comprised of 10 different, interconnected databases, including the Computerized Criminal History (CCH), which includes arrest information, and Law Enforcement Network with Corrections (LINC), which includes inmate information from the New Mexico Department of Corrections. The 8 additional databases are titled: Intelligence; Arrest Booking; Unidentified Body; Core; Sexual Offender; Field Interview; NMIBRS (New Mexico Incident Based Reporting System); Missing Person.

Although each table can be connected back to the main PEOPLE table, it is not clear that an individual can easily be followed through the different databases except by their person ID (as opposed to an incident ID connecting an incarceration spell to a specific arrest). Figure 18 re-creates the relationship between the CAI/CCH and Arrest Booking tables (in red), the LINC tables (in green), and some of the Core (in gray).

G.11 New York

The Computerized Criminal History (CCH) is maintained by the Division of Criminal Justice Services. CCH contains the criminal history records of all persons arrested and prosecuted since 1970 for crimes who were fingerprintable. An individual's criminal history includes a record of all arrests for that individual who were fingerprintable, as well as the charges reported with the arrest, disposition information, and information related to sentencing (including alcohol treatment and community service) if the individual was convicted.

Figure 18: New Mexico ERD (abridged)



Arresting agencies submit fingerprints. Case disposition information is submitted throughout the processing of the case by the courts. Corrections records are submitted by various supervising agencies post-conviction

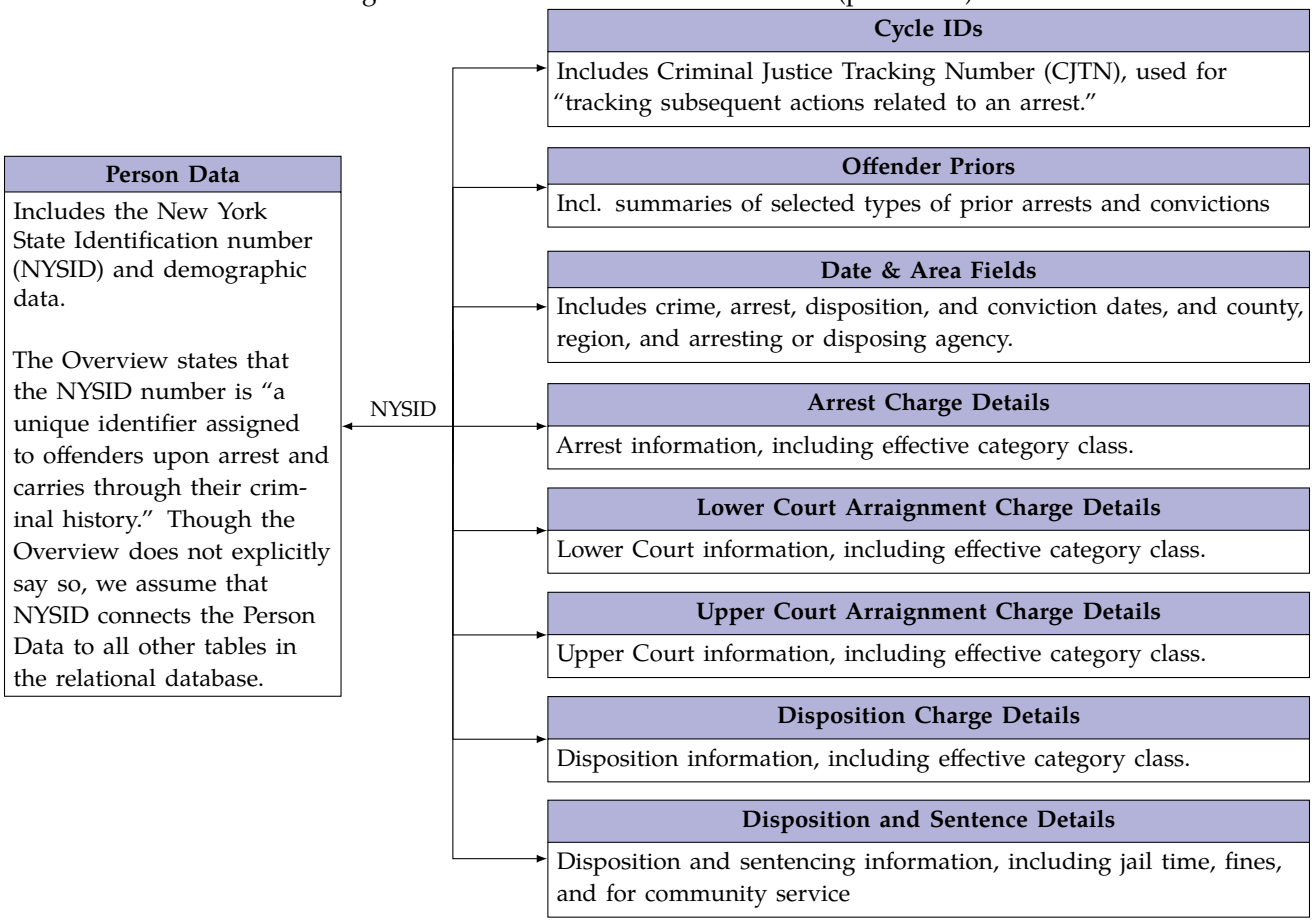
The underlying CCH data is stored in a series of Oracle tables containing approximately 21 million criminal history records. The Office of Justice Research and Performance (OJRP) extracts data from the CCH once a month and stores the data in a relational database structure called the OJRP Analytic Database. The OJRP Analytic Database is used to create specialized files for analysis, including the "CCH Top Charge file." The unit of event is a criminal event cycle, representing processing of a single case from arrest and prosecution through the court system, including sentencing information.

The CCH Top Charge file is divided into the 9 sections shown in Figure 19, except that the Arrest Charge Details, Lower and Upper Court Arraignment Charge Details, and Disposition Charge Details are called, e.g., "Top Arrest Charge Details." In the CCH Top Charge file, only the "top (most serious) charge is stored" for each of these tables. We assume that the data for every charge in a criminal even cycle is stored in the underlying Oracle tables.

G.12 North Carolina

The North Carolina State Bureau of Investigation (SBI) collects, stores, maintains, and disseminates criminal history information. SBI operates the North Carolina Computerized Criminal History (CCH) system, which includes criminal records of individuals who have been arrested with a valid criminal fingerprint card, historical misdemeanor fingerprint cards, and Juvenile Investigative Records for those adjudicated delinquent

Figure 19: New York CCH database model (presumed)



and fingerprinted.

In addition to biographical and biometric information on individuals with CCH records, CCH records normally consist of three components that are merged together and associated with these individuals electronically (presumably by State Identification Number): arrest, court disposition, and custody. Law enforcement agencies submit arrest data with records of an individual's crime for which they were fingerprinted. Court disposition data is supplied by the North Carolina Administrative Office of the Courts (AOC) after dispositions are released. Custody data includes custodial/prison information supplied by the North Carolina Department of Corrections.

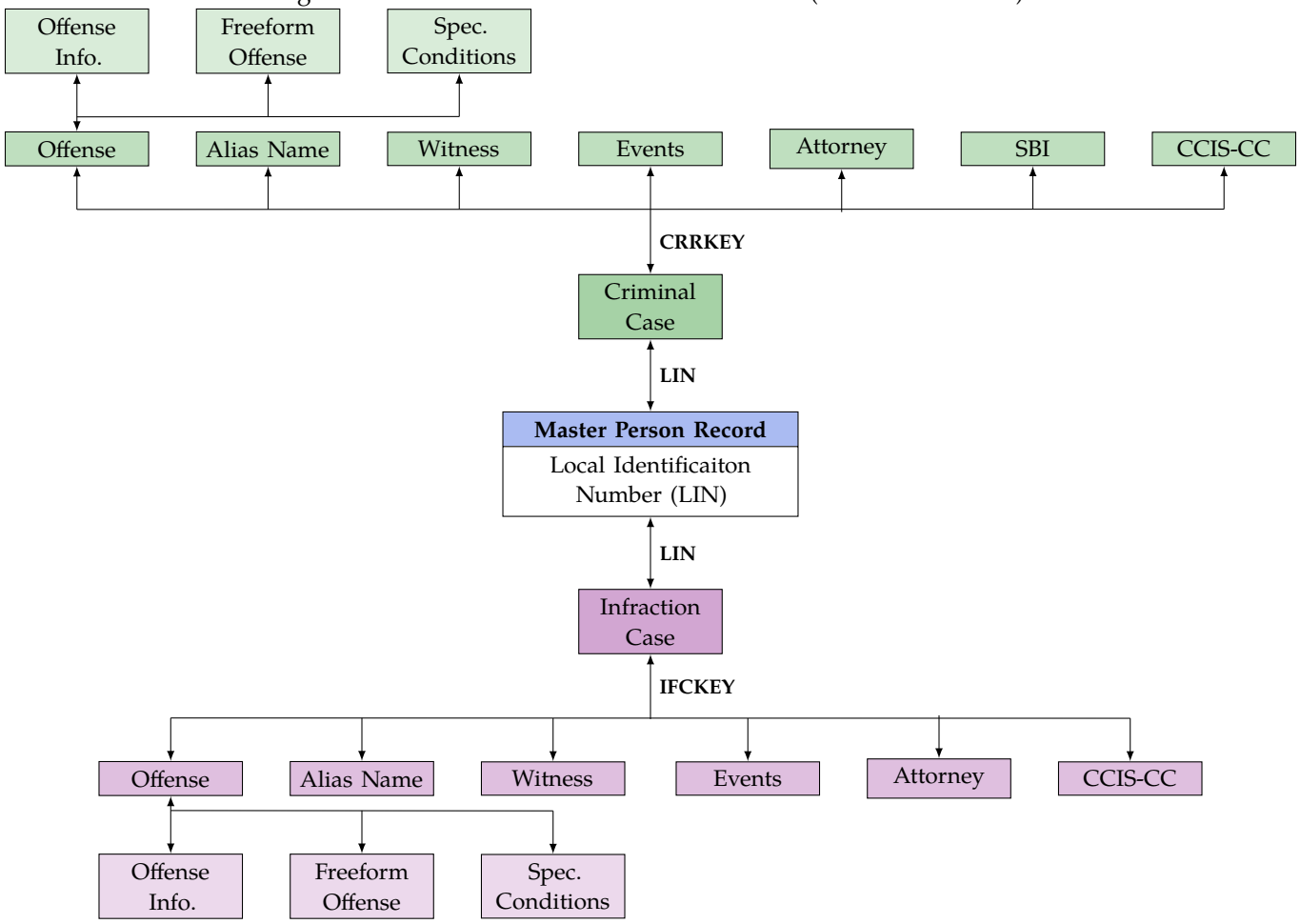
The data maintained by the AOC is called the Automated Criminal / Infractions System (ACIS). ACIS is a mainframe computer system that has been enhanced and maintained for over 30 years. It interfaces with several in-house systems as well as several outside agencies, including the Department of Motor Vehicles (DMV), the State Bureau of Investigation (SBI), and the Department of Correction (DOC). ACIS is accessible through public computer terminals that are stationed inside the clerk's office in each county. See Figure 20 for an approximation of how ACIS is organized. These data are linked to other criminal history data through SID by the SBI.

G.13 Oregon

Oregon does not have a unified criminal history database. The Oregon Statistical Analysis Center (the "Criminal Justice Commission") works to collect and merge data cross systems. The three datasets most frequently used by the Criminal Justice Commission are from the state police, courts, and the Department of Corrections.

The Oregon State Police maintain the LEDS data. Computerized Criminal History (CCH) records are kept in the LEDS system. The LEDS data starts at arrest. The LEDS data also contains disposition information

Figure 20: North Carolina ACIS database model (LIN link assumed)



entered from the court, though many dispositions are missing.

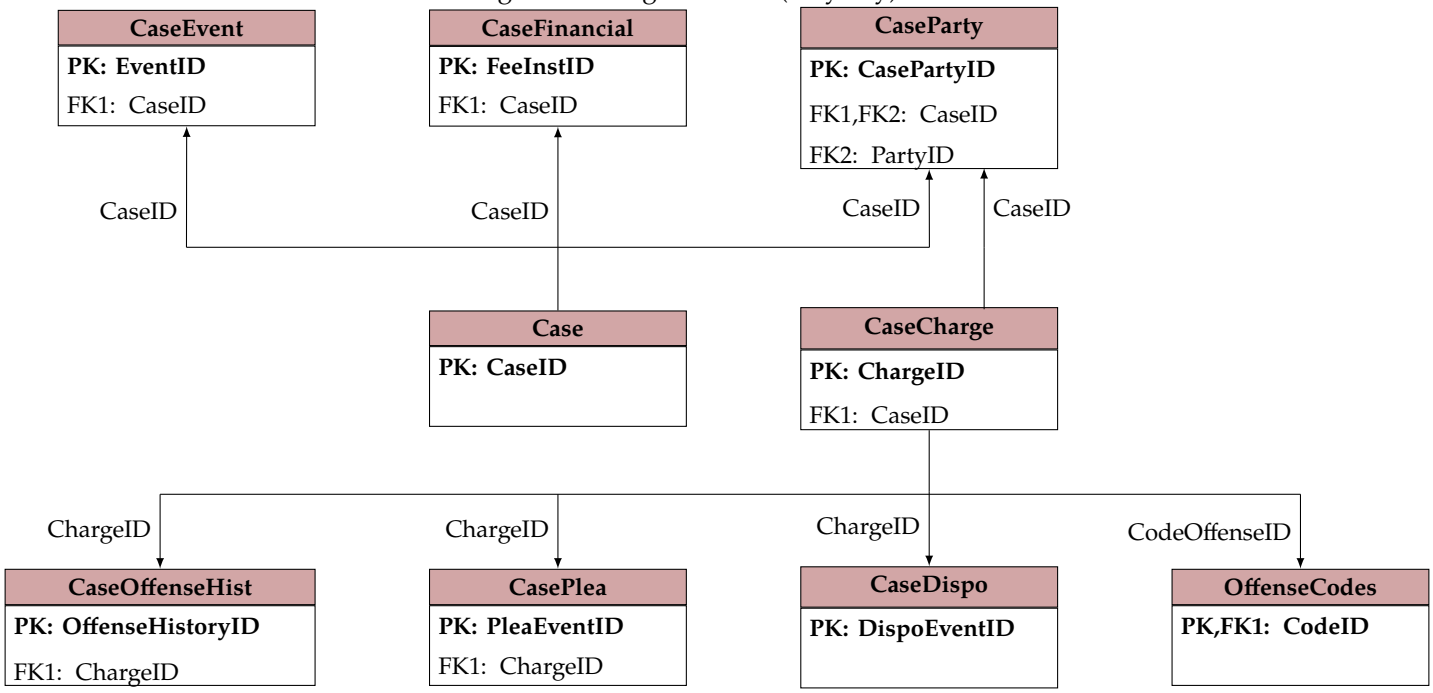
The court data has recently transferred to a new system called "Oregon eCourt" (a.k.a. "Odyssey"). This data is at the charge level. The data comes from Oregon's 36 circuit courts. The courts enter disposition information into LEDS (i.e., much of the information in Odyssey is also contained in LEDS–there just can be a lag for when information becomes available). The ERD for the Odyssey system SQL tables is reproduced in Figure 21.

The Department of Corrections data contains all felony convictions from Oregon's 36 circuit courts. The sentences include felony probation, felony local control (jail), and prison sentences. Each row in the data file is a felony sentence.

G.14 Pennsylvania

There is no unified source of criminal justice data in Pennsylvania. There has been some collaboration among several several Commonwealth adult criminal justice agencies to produce the Pennsylvania Criminal Justice Data Dictionary (CJDD), which contains data elements and their individual agency specific definitions. The participating agencies are the Administrative Office of Pennsylvania Courts (AOPC), the Department of Corrections (DOC), the Pennsylvania Board of Probation and Parole (PBPP), the Pennsylvania Commission on Crime and Delinquency (PCCD), and the Pennsylvania Commission on Sentencing (PCS). Though there is no schema that shows how individual cases can be tracked through these different datasets, we can use

Figure 21: Oregon eCourt (Odyssey) ERD



State ID data element to link these databases together (see Figure 22).[6]

One of the primary sources of criminal justice in Pennsylvania is the Pennsylvania State Police. The PA State Police is responsible for collecting all the Uniform Crime Report data for the Commonwealth and opted not to participate in the data dictionary process. In 2015, the Pennsylvania Uniform Crime Reporting Program received data from 1,925 jurisdictions. For UCR purposes, an adult arrests are counted for each adult processed by arrest, citation, or summons, including those individuals arrested and released without a formal charge being placed against them.

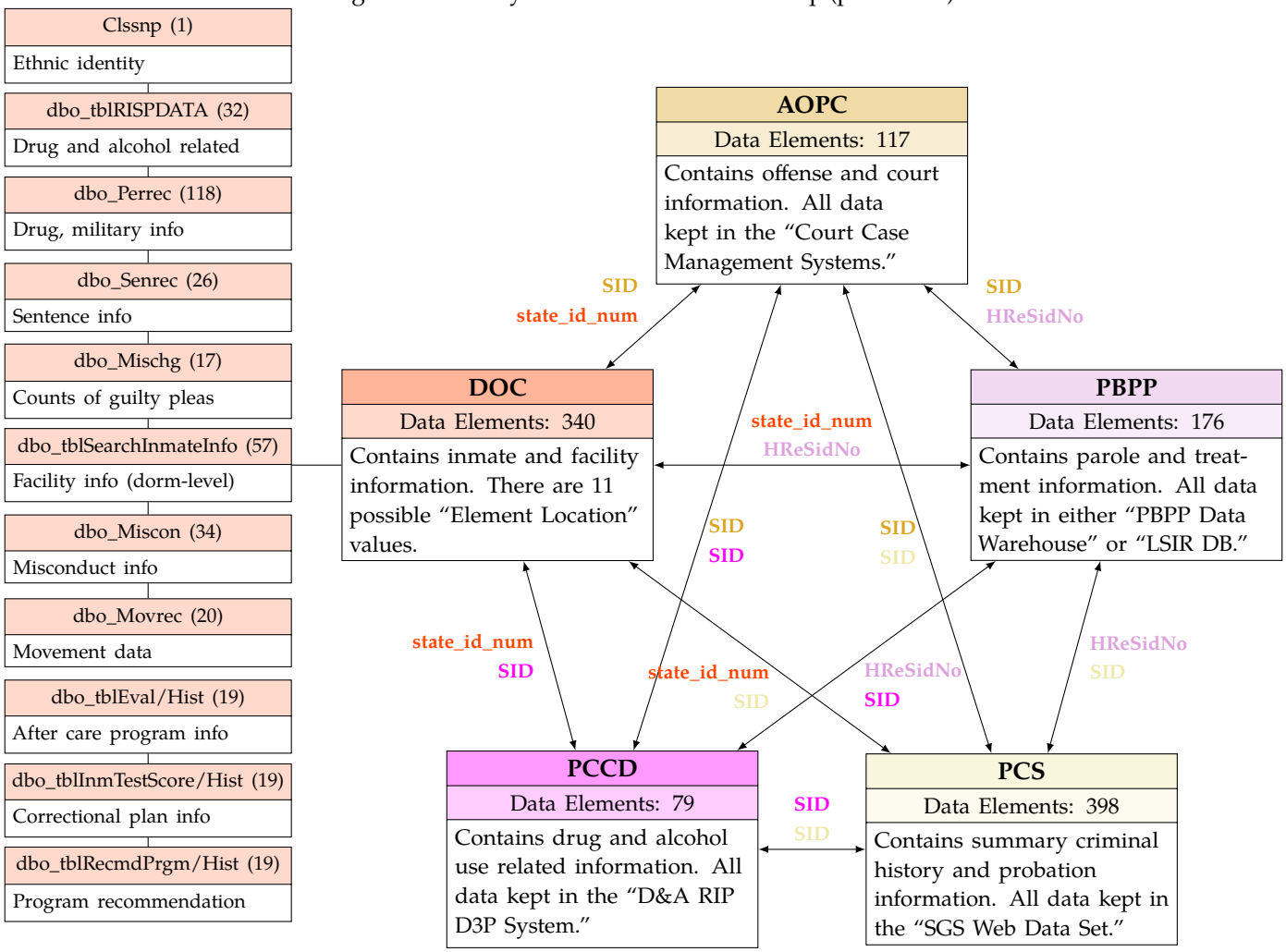
G.15 Texas

The Computerized Criminal History (CCH) is maintained by the Crime Records Service of the Texas Department of Public Safety. CCH includes information on arrests, prosecutions and the disposition of the case for persons arrested for Class B misdemeanor (or greater) violation of Texas criminal statutes. Many different agencies contribute information to CCH. Police Departments, Sheriff's Offices or any other criminal justice agency in Texas that arrests a person for a Class B misdemeanor or higher violation of a Texas statute is required to report that event to DPS within seven days. County Attorney, District Attorney or other prosecutor receiving a class B misdemeanor or greater offense must report to DPS the decision to accept, reject, change, or add to the charge for trial. Finally, County Clerks, District Clerks, or others clerks whose courts try Class B misdemeanor or greater violations of Texas statutes must report the disposition of the case to DPS. The CCH database model is reproduced below in Figure 23.

CCH only comprises one aspect of the Texas Criminal Justice Information System (CJIS). The other component of CJIS is the Corrections Tracking System (CTS) managed by the Department of Criminal Justice (DCJ). Incident Tracking Number (TRN) and Incident Tracking Number Suffix (TRS) are used as the keys for linking charges from arrest through adjudication.

[6]Generally speaking, each agency makes data requests to each other. The notable exception to that is PA Commission on Sentencing (PCS), whose Sentencing Guidelines Software (SGS) Web system interfaces with AOPC's court records.

Figure 22: Pennsylvania database relationship (presumed)



G.16 Utah

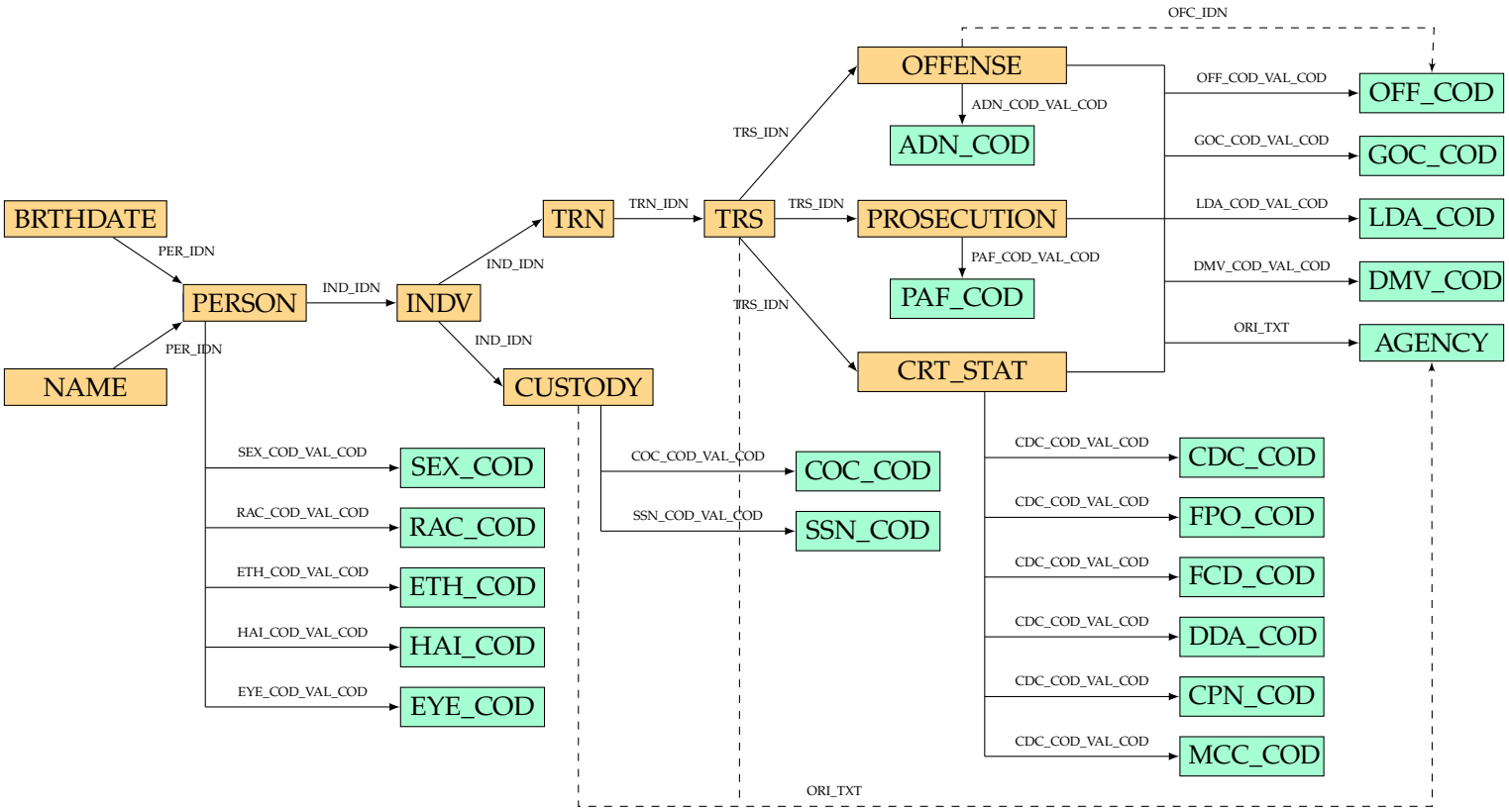
Criminal history data is kept by the Utah Department of Public Safety, Department of Technology Services division. The information we received was centered around the organization of the data (see Figure 24); we received no little information about what data is collected, who collects the data and submits it to DPS, how this data might connect with criminal justice data collected by other agencies, etc.

G.17 Vermont

The Vermont Criminal Information Center (VCIC) is the central repository of all criminal record information generated by criminal justice agencies statewide. The VCIC Criminal History Repository contains information documenting an individual's contact with the criminal justice system, including data regarding identification, arrest or citation, arraignment, judicial disposition, custody and supervision. The VCIC CCH system is provided by CPI. The CCH user interface operates from the CPI OpenFox™ Desktop Web Portal as Java applications running under the Java Runtime Environment.

The CCH system interfaces with several other systems that exchange data with CCH, including the MorphoTrak Automated Fingerprint Identification System (AFIS), the Sex Offender Registry (SOR) system, and the Department of Corrections (DOC) system (see Figure 25). The fields DOC Supervision Status and Supervising Officer in the CCH system are updated every night from the DOC system.

Figure 23: Texas CCH database model



G.18 Washington

The centralized criminal history repository in Washington is called A Central Computerized Enforcement Service System (ACCESS). It is maintained by the Information Technology Division of the Washington State Patrol (WSP). These records include criminal history on convicted criminals, persons who have been arrested but not charged with a crime, applicants fingerprinted as a result of employment with a law enforcement agency, and fingerprint data retained for compromised identification purposes. ACCESS extracts data from multiple places, including the Washington Crime Information Center (WACIC), the Department of Corrections (DOC), the National Crime Information Center (NCIC), and the Washington State Identification System (WASIS).

Most records are placed directly into the WACIC and NCIC systems by an originating agency (agency holding a warrant, missing person report, or theft report, etc.) WASIS is the Criminal History Record Information (CHRI) database maintained by the Criminal Records Division of WSP. The CHRI consists of fingerprint-based records and disposition information submitted by law enforcement agencies and courts throughout the state.

It's not entirely clear how the information from these different organizations can be linked together; figure 26 is an estimate of how the information in these files could be connected.

Figure 24: Utah database model (presumed)

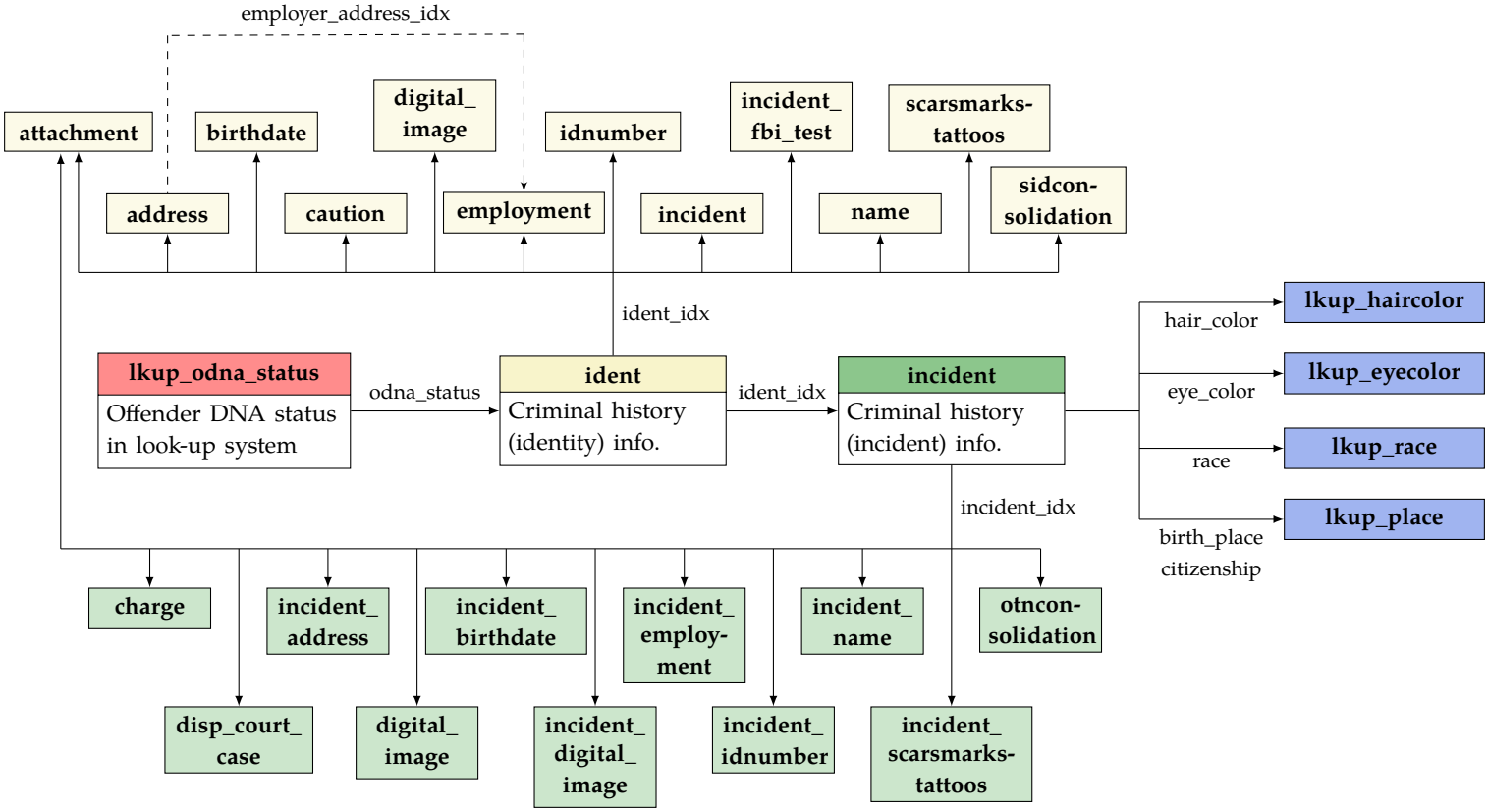


Figure 25: Vermont database model (presumed)

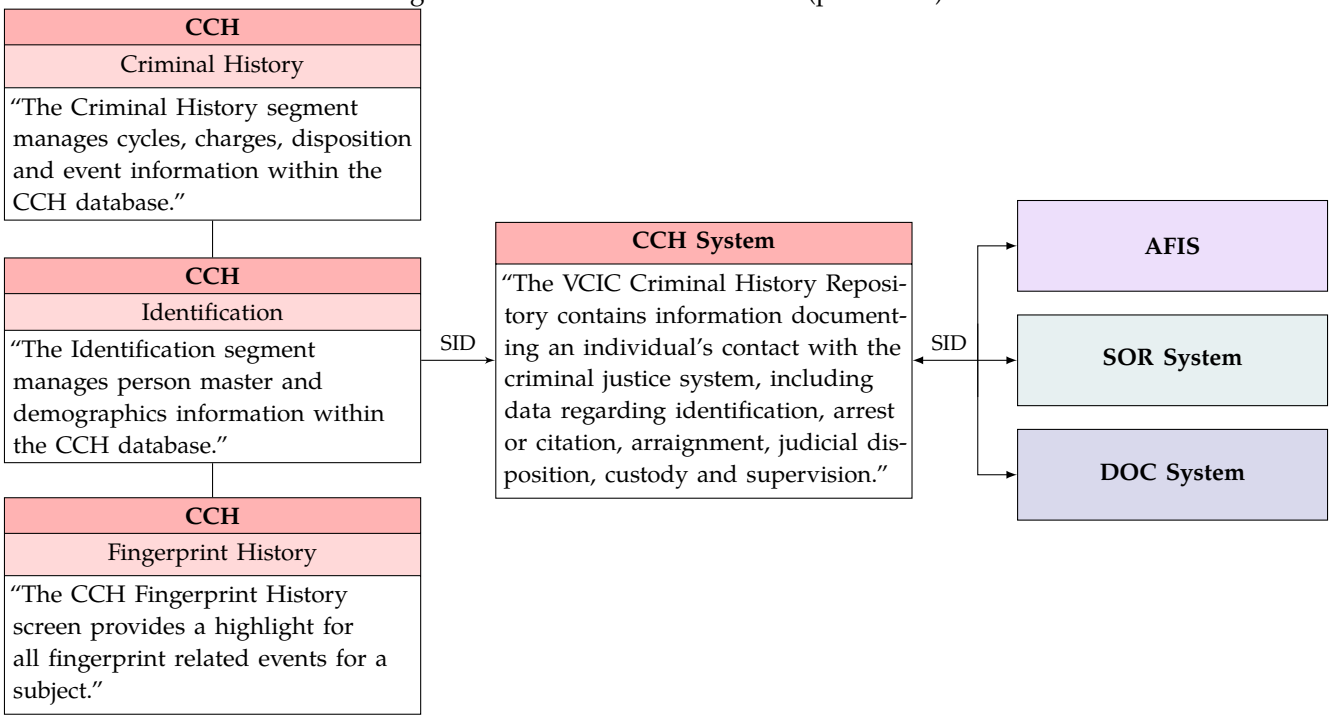
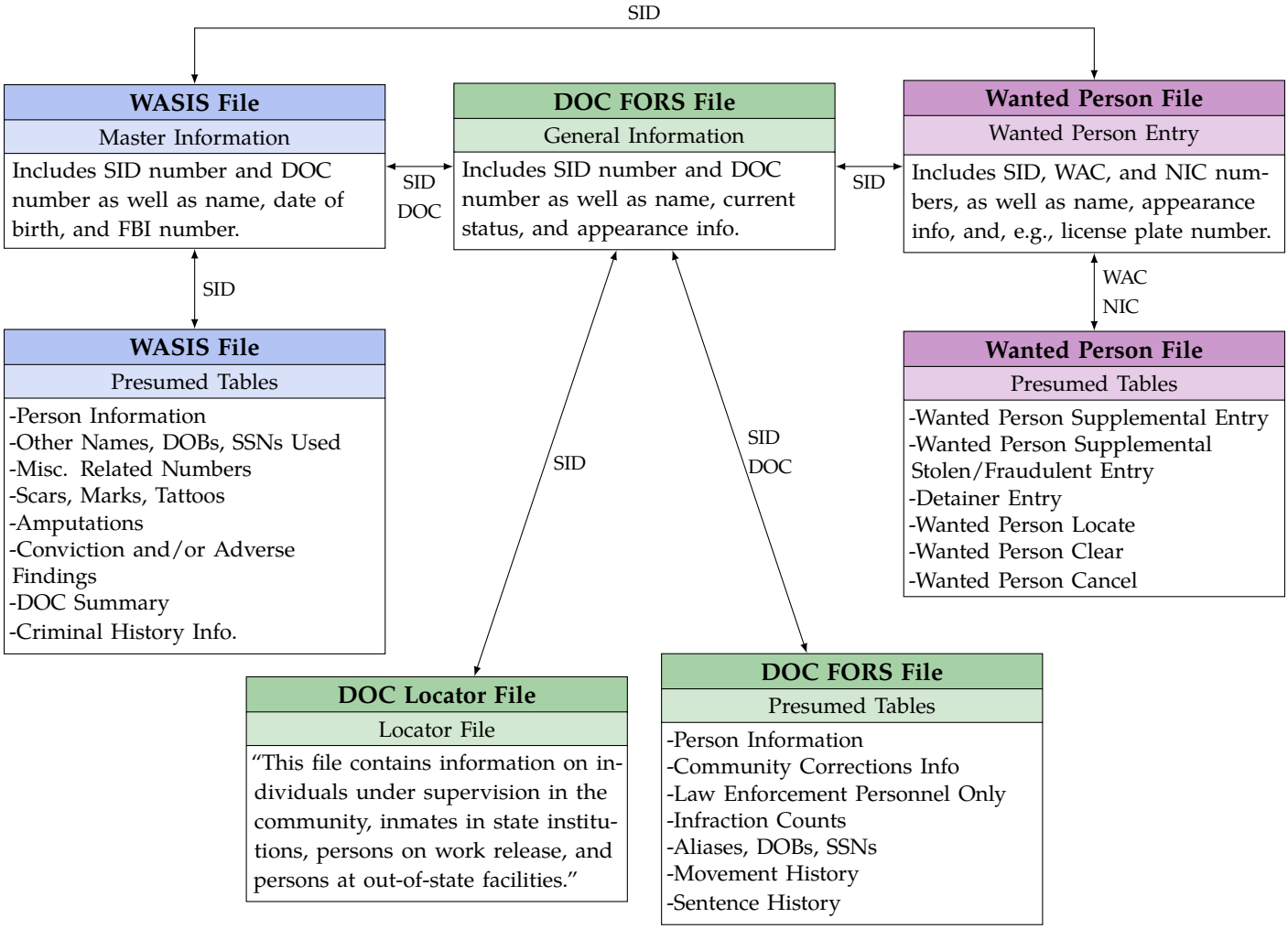


Figure 26: Washington data organization (presumed)



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. 2020. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. https://cjars.isr.umich.edu.

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.

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.

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