Criminal Justice Administrative Records System (CJARS)

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

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

2021-03-22

Abstract

The Criminal Justice Administrative Records System (CJARS) is a nationally integrated data repository designed to transform research and policymaking on the United States criminal justice system. At the University of Michigan, CJARS collects longitudinal electronic records from criminal justice agencies and harmonizes these records to track a criminal episode across all stages of the system. At the U.S. Census Bureau, harmonized criminal justice records can be linked anonymously at the person-level with extensive social, demographic, and economic information from national survey and administrative records. The CJARS project is a partnership between the Census Bureau and University of Michigan with the goal of expanding research and statistical reporting on the criminal justice system.

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

Table 1: Project change log

Release	Data change log	Documentation change log
2020Q4	– First harmonized, research-ready data delivered to the Census Bureau	– First version of data documentation released
2021Q1	– New data: FL, VA courts	– Expanded discussion of Census Bureau record linkage, ML offense classifier – Additional demographic variables in roster codebook

Contents

List of Tables

List of Figures

1 Project overview

1.1 Project description and objectives

The Criminal Justice Administrative Records System (CJARS) is a nationally integrated data repository designed to transform research and policymaking on the United States criminal justice system. At the University of Michigan, CJARS collects longitudinal electronic records from criminal justice agencies and harmonizes these records to track a criminal episode across all stages of the system. At the U.S. Census Bureau, harmonized criminal justice records can be linked anonymously at the person-level with extensive social, demographic, and economic information from national survey and administrative records. The CJARS project is a partnership between the Census Bureau and University of Michigan with the goal of expanding research and statistical reporting on the criminal justice system.

CJARS has the following primary objectives:

- Create a dataset that follows criminal cases across criminal justice agencies, from arrest through discharge from the justice system.

- Capture the progression of criminal activities over the life course by constructing longitudinal records of contact that individuals have with the justice system.

- Provide a national research platform in perpetuity through the Federal Statistical Research Data Center (FSRDC) network, allowing qualified researchers on approved projects to integrate anonymized criminal justice data at the person-level with Census Bureau survey and administrative records and produce cutting-edge research not previously possible.

- Enable the federal statistical system to produce timely and innovative statistical products about the criminal justice system and the justice-involved population.

- Acquire, store, and use criminal justice data securely and ethnically so that the identities and characteristics of individuals in the CJARS data are kept confidential.

1.2 Project scope

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

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

1.3 Data collection

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

statistical products. For every acquisition, the team maintains high standards of data provenance. CJARS also accepts data with well-documented provenance from individual researchers.

1.4 Current data coverage

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

CJARS has collected data from states across the U.S. Figure 1 identifies (1) states for which CJARS has longitudinal statewide data, (2) states with partial longitudinal coverage, (3) states with caseload snapshot data, and (4) states from which CJARS plans to acquire data. Based on 2019 state population estimates, states for which CJARS has longitudinal, statewide coverage in at least one criminal justice domain constitute 44.1% of the U.S. population (as of 2021-03-17).

Figure 1: Map of CJARS geographic and criminal justice domain coverage (as of 2021-03-17)

| Statewide coverage, 2+ domains | Statewide coverage, 1 domain | Partial geographic coverage | Caseload snapshot | Planned acquisition |
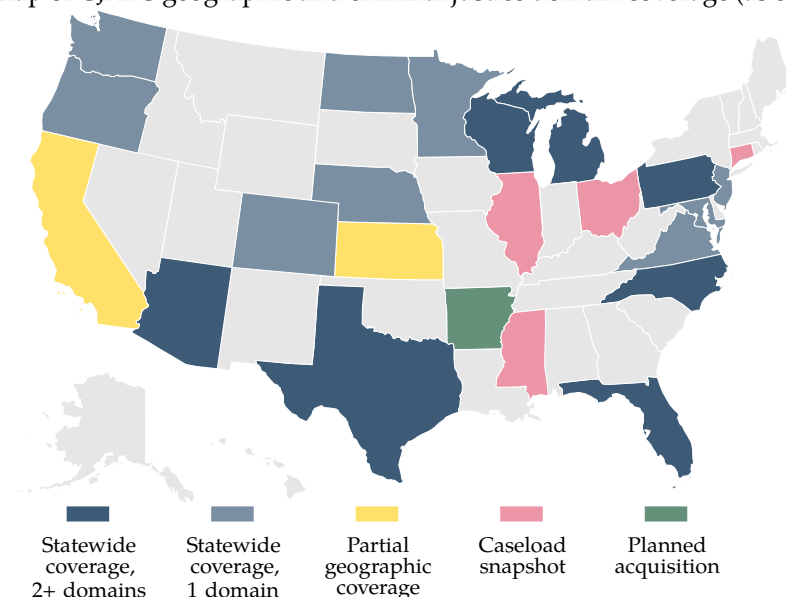
It is important to note that data coverage varies substantially from state to state. Specifically, there are differences in coverage of event type, time frame, and geography. Section 2.4 discusses how state coverage varies over time and by criminal justice domain. Appendix A lists the agencies whose data have been integrated into CJARS.

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

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

1.5 Data access

A major contribution of CJARS is the creation of a broadly accessible repository that will permanently enhance research infrastructure in the United States. Data are collected, cleaned, and harmonized at the University of Michigan and then integrated into U.S. Census Bureau data systems and made anonymous and available through the Federal Statistical Research Data Center (FSRDC) network. Qualified researchers can use the standard Census Bureau FSRDC proposal process to request use of the restricted-access CJARS data. The data cannot be requested directly from the University of Michigan.

Figure 2: Events in the CJARS database, by year and event type (as of 2021-03-20)
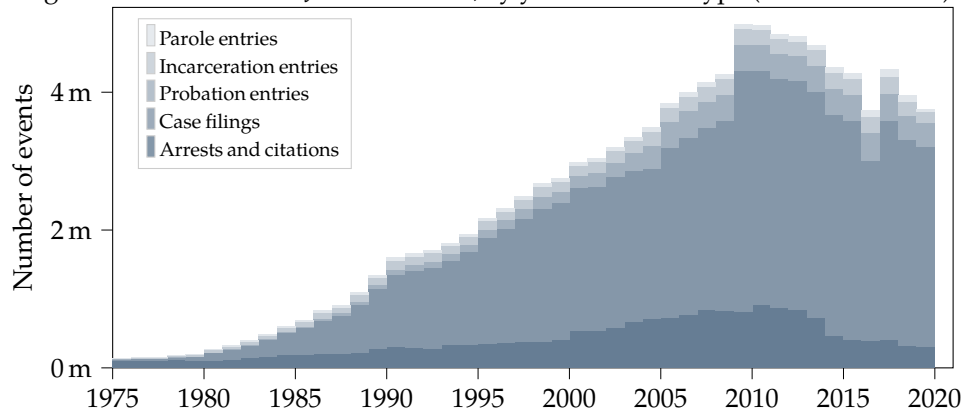
Distribution through the FSRDCs provides an ideal outlet. The FSRDC network is broadly available to researchers and provides excellent scientific computing resources, while maintaining the highest levels of information security and confidentiality protection. Researchers working in the FSRDCs have no access to sensitive PII, and can only request release of aggregate statistical material. The FSRDCs allow qualified researchers on approved projects to link CJARS records anonymously at the person-level to an abundance of socioeconomic survey and administrative data held by the Census Bureau. For more information on the FSRDC system and other available data, please visit: https://www.census.gov/fsrdc.

1.6 Data privacy

The CJARS project is built around data protection and security, in order to ensure 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 has 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, data are protected by 13 USC §9a. The University of Michigan transfers data securely to Census, whereupon received data are accessible only by a small staff responsible for inventorying the contents of the data and removing sensitive PII. Next, a data linkage team uses a probabilistic linkage process to replace sensitive PII with a unique identifier called a Protected 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 4.1.

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.

1.7 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.

1.8 Project funding

1.9 Project institutional review board compliance

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

1.10 Acknowledgments

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

1.11   Data citation

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

> Keith Finlay and Michael Mueller-Smith. 2021. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. https://cjars.isr.umich.edu.

1.12   Documentation structure

- Section 2 provides an overview of how the CJARS project acquires data and gives a summary of data coverage at state level and by criminal justice domain.
- Section 3 describes how raw criminal justice data is integrated and harmonized into a single national format at the University of Michigan.
- Section 4 explains how CJARS data, after being securely transferred to the Census Bureau and anonymized, can be linked with other socioeconomic records for research.
- Section 5 has the detailed table and variable codebook.
- Appendices cover the project in much more detail.
  - Appendix A identifies the agency sources of CJARS administrative data.
  - Appendix B explains the variable harmonization process in detail.
  - Appendix C describes the machine learning used to classify offenses.
  - Appendix D lists code schemes used in the variables, including criminal classification scheme (and explains how those relate to BJS schemes).
  - Appendix E explains the methods used to identify individuals and procedural episodes in the CJARS data.
  - Appendix F gives an overview of how to do record linkage on Census Bureau systems using CJARS data.
  - Appendix G has jurisdiction-specific data notes.
  - Appendix H describes a survey of state criminal history repository database designs, which provides some context for CJARS schema choices.

2 Data acquisition and coverage

2.1 Sources of criminal justice data

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

- Booking and arrests. This information is typically recorded by police departments and sheriff's offices.

- Trial, dispositions, sentences, and appeals. Such information may be held by local and state courts, as well as district attorney's offices and county clerks.

- Incarceration, probation, parole, and special programs. Most of this information may be recorded by a state Department of Corrections (or its equivalent). Nevertheless, it can be also recorded by some counties or regional offices.

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

2.2 Mechanisms used to acquire criminal justice administrative data

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

1. **Data use agreements.** The University of Michigan enters into formal legal agreements with criminal justice agencies to acquire data. These agreements delineate the specific responsibilities of the CJARS project and the allowable uses of the acquired data. These agreements must allow secure transmission of the data from the University of Michigan to the Census Bureau in order for linkage-based research to occur.
2. **Public records requests.** A request for public records is filed in jurisdictions where legal statutes allow for such requests.
3. **Web scraping or bulk downloads.** Data that are publicly available online are collected using web scrapers or downloaded in bulk when available. Scraping is only conducted using methods that comply with an agency's terms of use.

Table 2 shows the number of records that each of these methods has contributed to the CJARS repository.

Table 2: CJARS records acquired by acquisition method (as of 2020-02-08)

Acquisition source	Records	Individuals
Data use agreements	1,557,724,953	14,949,726
Public records requests	291,601,086	12,623,722
Web scraping or bulk downloads	400,965,342	3,309,058

2.3 Data acquisition strategy

The ultimate goal of CJARS is to provide data infrastructure with nationwide coverage of the major events that occur at each stage of the justice system. CJARS is focused on expanding the geographic and procedural coverage of its data holdings. Expanding geographic coverage means expanding the number of states covered at the statewide level. Expanding procedural coverage means filling in data sources from the sequence of criminal justice processes, so that CJARS can track cases and events from charge to final sanction. Expanding data coverage along these two dimensions will facilitate broader cross-jurisdictional research, comparative

analysis, and research that requires tracking a criminal episode though the different stages of the justice system.

The fragmented nature of record keeping in the justice system also shapes the strategy CJARS takes toward data collection. To maximize coverage given the project's resource constraints, CJARS prioritizes acquisitions from agencies that manage statewide data systems including departments of corrections, state court administrative offices, departments of public safety, etc. CJARS also seeks to acquire local agency data where the costs of doing so are low, such as where web scraping or public records requests are possible.

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

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

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

Figure 3: Estimated CJARS statewide coverage (as of 2021-03-17)

Figure 3: CJARS statewide coverage (cont'd)

Figure 3: CJARS statewide coverage (cont'd)

Wisconsin — Arrest, Court, Probation, Prison, Parole
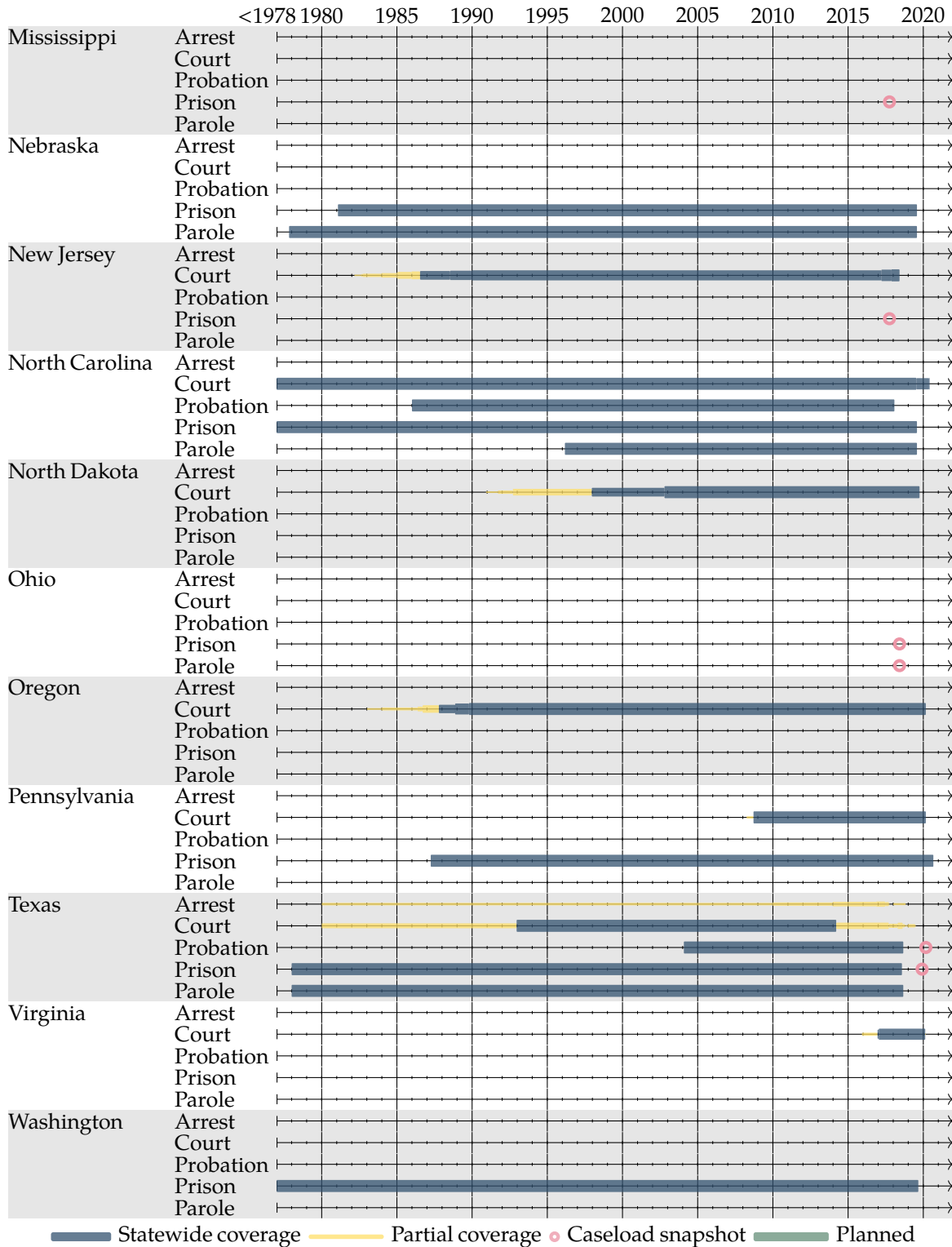
Legend: Statewide coverage, Partial coverage, Caseload snapshot, Planned
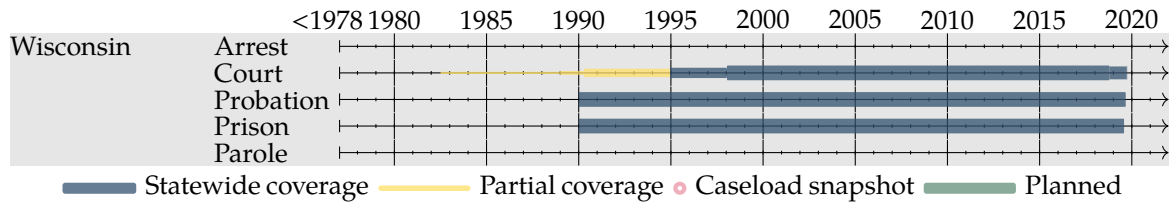
3 Data integration, harmonization, and linkage at the University of Michigan

3.1 Overview of CJARS data integration, harmonization, and linkage

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

- **Data requests and collection**. CJARS collects data using the three-pronged approach that is described above. All data and other relevant information describing the data are kept in a secure data enclave.

- **Data cleaning**. Original data are cleaned and harmonized to fit a common CJARS schema (described in more detail in later sections).

- **Merge**. The harmonized data are appended and merged with other criminal justice records in CJARS to create a single dataset spanning agencies and jurisdictions.

- **Person-level linkage**. CJARS uses 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. A `cjars_id` study id is created for each unique individual identified during this process.

Figure 4: CJARS data exchange and processing

18
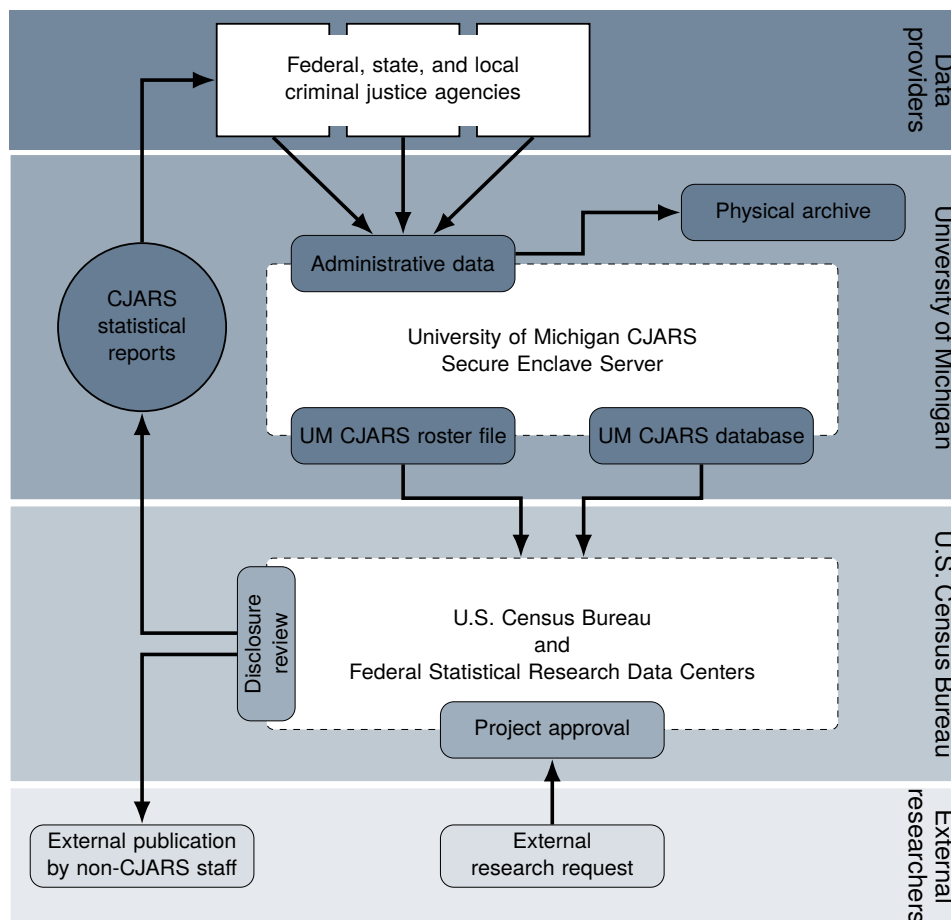
Figure 4 provides a visual representation of the data integration process. This figure shows how data that is collected from data providers (first panel) is added to the UM CJARS Repository (second panel). Then once the data is processed, the data are integrated in the Census Bureau (third panel). Finally, once processed by the Census Bureau, the data are made available to external researchers (fourth panel).

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

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

Figure 5: CJARS data schema
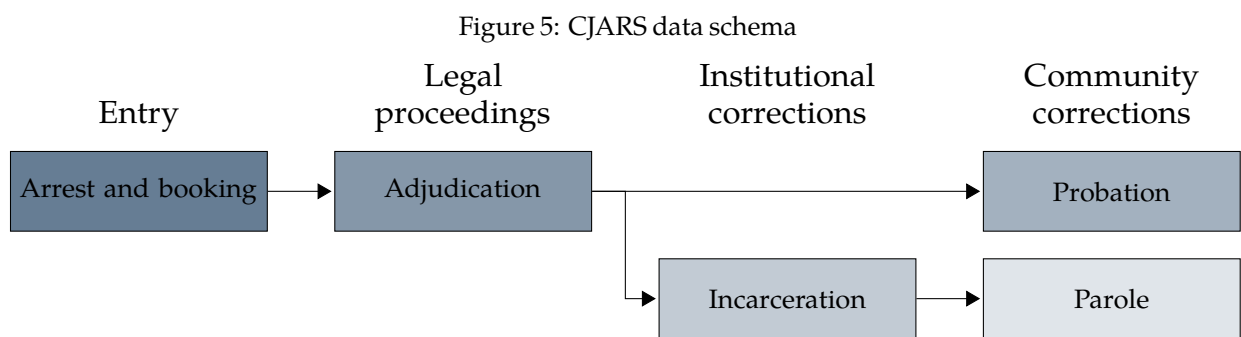
This figure of events was used to develop a data schema which facilitates the linkage of records across disparate systems. The CJARS national schema was designed in a way that strikes a balance between capturing the complexity of the way that events related to a criminal episode are processed through 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.

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

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

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

3.3.1   Person-level linkage using the `cjars_id`

CJARS has developed an algorithm that probabilistically matches records to individuals when no unique identifier is available by using 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 roster and one database for each of the five types of events that are covered in Figure 5. Collectively, the five databases containing the criminal justice events are referred to as the CJARS relational databases. The roster uses the `cjars_id` identifier to uniquely identify every individual included in the CJARS repository. Additionally, each CJARS relational database has a unique identifier that identifies each event in the file. Specifically, the `arr_id` uniquely identifies arrests, `adj_id` is for court filings, `pro_id` is for probation events, `inc_id` is for incarceration events, and `par_id` is for parole events. As can be seen in Figure 6, the `cjars_id` can be used to link all events in each CJARS relational database to the individual that they belong to based on the roster.

Figure 6: Person-level linkage with the `cjars_id`
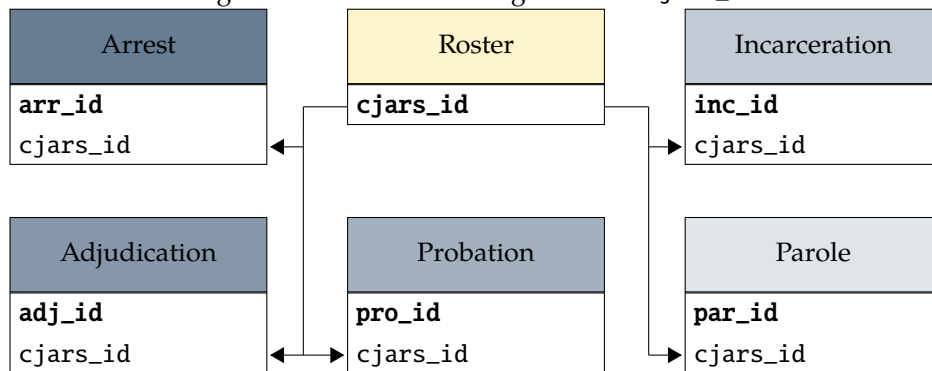
Note: Arrows indicate linking variables between tables. Variable names listed in **bold** represent unique variables in the dataset. Variables that are not bold may have multiple entries.

3.3.2 Data linkage via episode stage unique identifiers

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

Figure 7 summarizes how the data linkage is structured based on unique identifiers of events in each CJARS relational database. Each CJARS relational database contains an identifier for that respective type of event contained in the table, but also an identifier that can be used to link to the event that led to the event described in the table. For example, a court case filing in the adjudication relation database will contain the `adj_id` which uniquely identifies that case filing, but also the `arr_id` 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.

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

Figure 7: Data linkage via unique identifiers of episodes
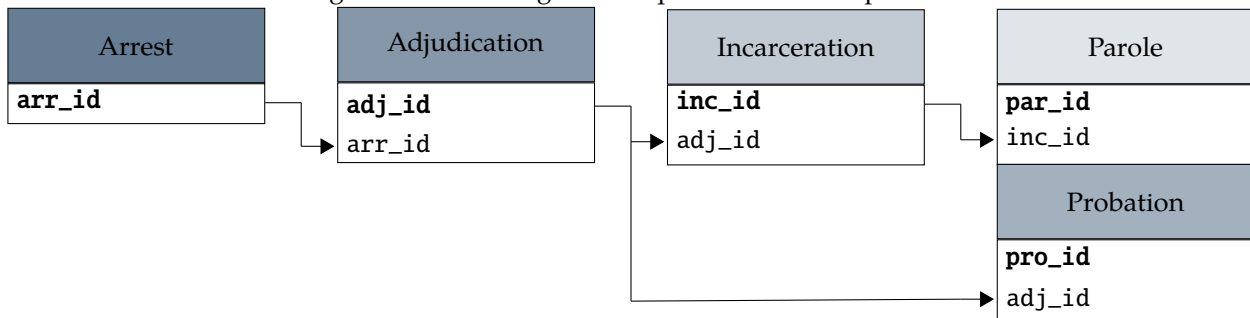
Arrest		Adjudication		Incarceration		Parole
arr_id	→	**adj_id**	→	**inc_id**	→	**par_id**
		arr_id		adj_id		inc_id

Probation
pro_id
adj_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.

3.4 Probabilistic entity resolution algorithms are used to link records

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

3.4.1 Entity resolution algorithms

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

There are two broad classes of entity resolution algorithms: deterministic and probabilistic. Deterministic algorithms focus on the variables common to both sets of data being matched. In some examples, paired cases must match on all common variables to be classified as a match. In other cases, with richer sets of matching variables, some flexibility can be built into the matching process. Conversely, probabilistic algorithms attempt to predict the probability that any two observations should be linked based on the relative agreement of their matching variables. This approach has benefits over deterministic models in that it more flexibly sets a decision rule that optimizes the trade-off between making more matches and limiting false matches. The decision of which method to apply is dependent upon the available set of matching variables and the type/quality of information available in those matching variables. CJARS developed a probabilistic algorithm based on the identifiable information that is usually available in criminal justice records and the quality of said data.

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

3.4.2 Episode resolution algorithm

> **USER NOTE**
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense. We expect to complete this work in 2021. Users can expect to see linked processes in the RDC data warehouse in late 2021.

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

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

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

4 Record linkage and research applications at the Census Bureau

4.1 Integration of criminal justice and Census Bureau records

At the Census Bureau and in the FSRDCs, CJARS data may be linked at the person-level to other socioeconomic survey and administrative records using an anonymous identifier called a Protected Identification Key (PIK). PIKs are assigned by the Census Bureau after data have been transferred securely from the University of Michigan. 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, see Wagner and Layne 2014). Once the PIK assignment process has occurred, sensitive PII is removed from the research files and 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.

Figure 8: Two rounds of entity resolution are applied to the CJARS roster
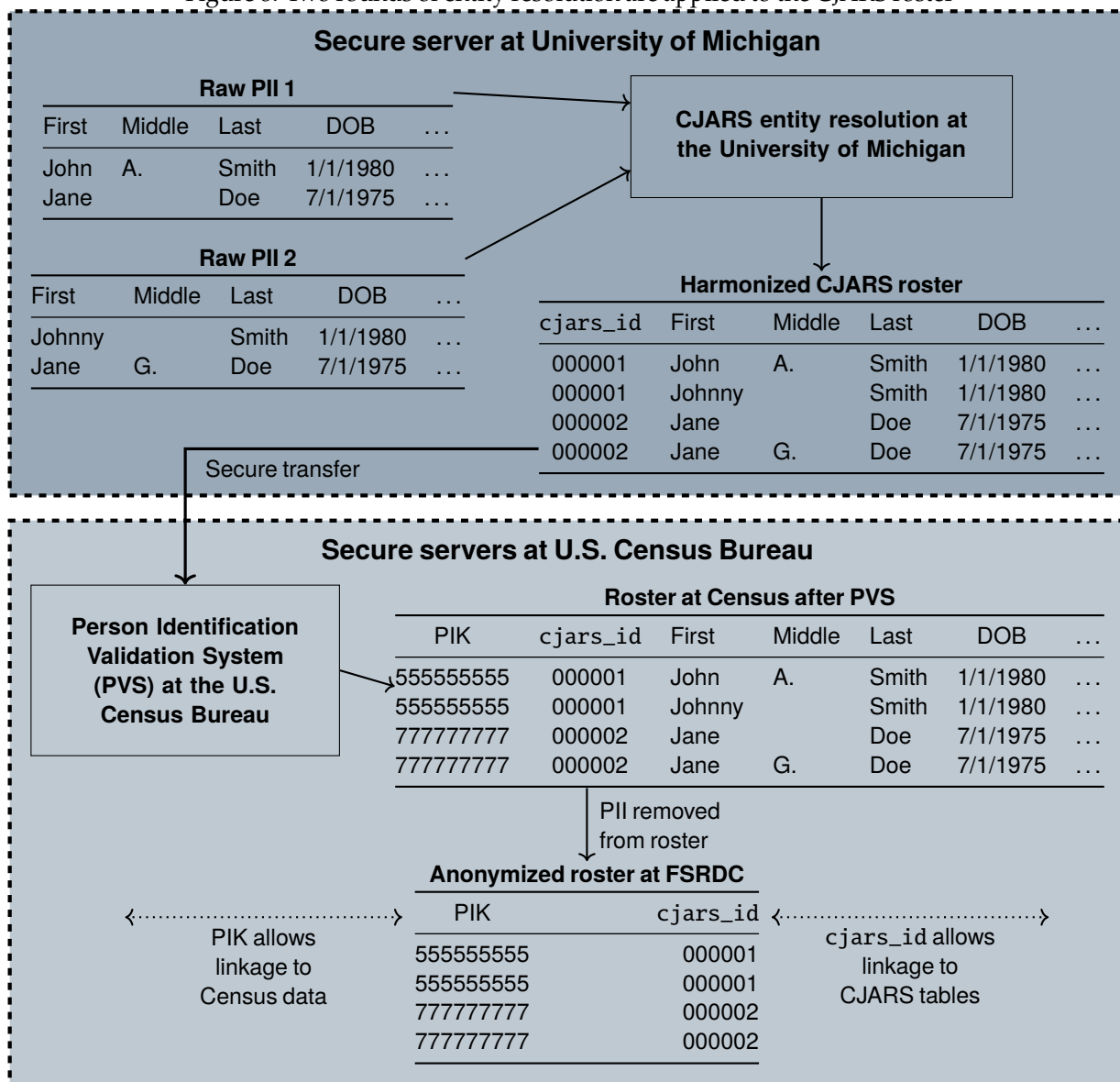
Note that this process introduces a second round of entity resolution before the researcher has access to the CJARS

23

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 agency-validated identifiers and its own subject-matter expertise, but has no access to a population-level registry. The Census Bureau entity resolution team has access to registry data not available outside of the Census Bureau, but that process cannot make use of identifying information specific to criminal justice data. (The Census Bureau's registry for record linkage is based primarily on the Census Numident, which in turn is based on the Social Security Administration's Numident file and includes anyone who has ever received a Social Security Number (SSN) or an Individual Taxpayer Identification Number (ITIN).)

When researchers first use CJARS in the FSRDCs, they must decide how to resolve any discordances between `cjars_id`s and PIKs. In Appendix F.2, we discuss techniques for doing so.

4.2   Research vignettes using record linkage at the Census Bureau

Qualified researchers on approved projects can link CJARS data at the person level with an abundance of socioeconomic survey and administrative data held by the Census Bureau. In this section, we discuss some research vignettes to demonstrate how a researcher might accomplish projects with CJARS data.

4.2.1   Labor market outcomes after a criminal justice intervention

- Research question: How does felony conviction affect labor market outcomes?
- Datasets
    - CJARS: roster and adjudication table
    - Longitudinal Employer-Household Dynamics (LEHD): quarterly employment, earnings, industry
    - Researchers can also bring their own data to link at the person level (e.g., training data from correctional facilities).
- Linkage
    - The roster contains both the `cjars_id` and the PIK, so it can serve as a crosswalk between the CJARS adjudication table and the LEHD labor market data.

4.2.2   Neighborhood environment and criminal justice involvement

- Research question: How do characteristics of one's neighborhood environment correlate with criminal justice involvement?
- Datasets
    - CJARS: roster and all topical tables
    - 2010 Decennial Census: to identify residential location for individuals who are not incapacitated on Census day
    - Other public or restricted datasets to identify other neighborhood characteristics
- Linkage
    - Use the roster to link the CJARS tables to individuals in the census data. Use the geographic identifiers in the census data to link in neighborhood characteristics.

4.2.3   Criminal justice contact as an outcome for a non-criminal justice intervention

- Research question: How does access to job training programs, for example, affect criminal justice involvement?
- Datasets required
    - CJARS: roster and all topical tables
    - Researcher provided job training data that can be linked at the person level
- Linkage
    - Use the roster to link the CJARS tables to individuals in the intervention data.

5 Table and variable codebook

5.1 Variables by relational table

The CJARS relational databases contain detailed information about arrests and bookings, adjudications, incarcerations, and terms of parole and probation. Primary key variables are shown in bold.

- **Roster**. The roster table contains the unique list of person-level `cjars_id`s. Data is at the person-level.
    - **cjars_id**: CJARS person identifier
    - dob_yyyy: Date of birth, year
    - dob_mm: Date of birth, month (no longer available - deprecated in 2024)
    - dob_dd: Date of birth, day of month (no longer available - deprecated in 2024)
    - sex: Sex
    - sex_raw: Sex, raw
    - sex_imputed: Sex imputation indicator
    - race: Race and ethnicity
    - race_raw: Race and ethnicity, raw
    - race_imputed: Race imputation flag
- **Arrest and booking**. The arrest table contains information regarding the arrest and booking date, as well as the offense that led to the arrest. Data is at the charge-level.
    - cjars_id: CJARS person identifier
    - **arr_id**: Arrest identifier
    - arr_arr_dt_yyyy: Arrest date, year
    - arr_arr_dt_mm: Arrest date, month
    - arr_arr_dt_dd: Arrest date, day of month
    - arr_book_dt_yyyy: Booking date, year
    - arr_book_dt_mm: Booking date, month
    - arr_book_dt_dd: Booking date, day of month
    - arr_off_cd: CJARS standardized offense code - arresting offense
    - arr_off_cd_src: Raw offense code from source - arresting offense
    - arr_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. Data is at the charge-level.
    - cjars_id: CJARS person identifier
    - **adj_id**: Court case filing identifier
    - adj_grd_cd: CJARS standardized offense grade
    - adj_grd_cd_src: Raw offense grade from source (e.g., felony, misdemeanor)
    - adj_off_lgl_cd: CJARS standardized offense legal code
    - adj_off_lgl_cd_src: Raw legal code from source (e.g., ordinance violation)
    - adj_file_dt_yyyy: Case filing date, year
    - adj_file_dt_mm: Case filing date, month
    - adj_file_dt_dd: Case filing date, day of month
    - adj_chrg_off_cd: CJARS standardized charge offense - offense charged at case filing
    - adj_chrg_off_cd_src: Raw charge offense description from source - offense charged at case filing
    - adj_disp_dt_yyyy: Case disposition date, year
    - adj_disp_dt_mm: Case disposition date, month
    - adj_disp_dt_dd: Case disposition date, day of month
    - adj_disp_cd: CJARS standardized disposition

- **– adj_disp_cd_src**: Raw disposition description from source
- **– adj_disp_off_cd**: CJARS standardized disposition offense - offense recorded at disposition
- **– adj_disp_off_cd_src**: Raw disposition offense from source - offense recorded at disposition
- **– adj_off_dt_yyyy**: Date offense was committed, year
- **– adj_off_dt_mm**: Date offense was committed, month
- **– adj_off_dt_dd**: Date offense was committed, day of month
- **– adj_sent_dt_yyyy**: Sentencing date, year
- **– adj_sent_dt_mm**: Sentencing date, month
- **– adj_sent_dt_dd**: Sentencing date, day of month
- **– adj_sent_serv**: Community service sentence
- **– adj_sent_dth**: Death sentence
- **– adj_sent_inc**: Incarceration length in months
- **– adj_sent_pro**: Probation length in months
- **– adj_sent_rest**: Restitution amount in dollars
- **– adj_sent_sus**: Suspended sentence
- **– adj_sent_trt**: Treatment sentence
- **– adj_sent_fine**: Fine amount in dollars
- **– adj_sent_inc_min**: Minimum incarceration term in months
- **– adj_sent_inc_max**: Maximum incarceration term in months
- **– adj_sent_src**: Raw sentence from source
- **– adj_st_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
- **Probation**. The probation table contains information on probation conditions, probation begin status and date, and probation end status and date. Data is at the level of a probation term.
  - **– cjars_id**: CJARS person identifier
  - **– pro_id**: Probation term identifier
  - **– pro_cond_cd**: CJARS standardized probation conditions
  - **– pro_cond_cd_src**: Raw description of probation conditions from source
  - **– pro_bgn_dt_yyyy**: Probation start date, year
  - **– pro_bgn_dt_mm**: Probation start date, month
  - **– pro_bgn_dt_dd**: Probation start date, day of month
  - **– pro_end_dt_yyyy**: Probation end date, year
  - **– pro_end_dt_mm**: Probation end date, month
  - **– pro_end_dt_dd**: Probation end date, day of month
  - **– pro_end_cd**: CJARS standardized probation end status
  - **– pro_end_cd_src**: Raw description of probation end status from source
  - **– pro_st_ori_fips**: State FIPS code for location of sentencing
  - **– pro_cnty_ori_fips**: County FIPS code for location of sentencing
  - **– pro_st_juris_fips**: 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
- **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. Data is at the level of an incarceration term.
  - **– cjars_id**: CJARS person identifier
  - **– inc_id**: Incarceration term identifier

- – `inc_fcl_cd`: CJARS standardized facility type
- – `inc_fcl_cd_src`: Raw description of facility from source
- – `inc_entry_dt_yyyy`: Incarceration entry date, year
- – `inc_entry_dt_mm`: Incarceration entry date, month
- – `inc_entry_dt_dd`: Incarceration entry date, day of month
- – `inc_entry_cd`: CJARS standardized entry status
- – `inc_entry_cd_src`: Raw description of entry type into incarceration
- – `inc_exit_dt_yyyy`: Incarceration exit date, year
- – `inc_exit_dt_mm`: Incarceration exit date, month
- – `inc_exit_dt_dd`: Incarceration entry date, day of month
- – `inc_exit_cd`: CJARS standardized exit status
- – `inc_exit_cd_src`: Raw description of exit type from incarceration
- – `inc_st_ori_fips`: State FIPS code for location of sentencing
- – `inc_cnty_ori_fips`: County FIPS code for location of sentencing
- – `inc_st_juris_fips`: State FIPS code of state responsibe for physical 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
- **Parole**. The parole table contains information on parole begin/end dates and exit status when available. Data is at the level of a parole term.
  - – `cjars_id`: CJARS person identifier
  - – **`par_id`**: Parole term identifier
  - – `par_bgn_dt_yyyy`: Parole start date, year
  - – `par_bgn_dt_mm`: Parole start date, month
  - – `par_bgn_dt_dd`: Parole start date, day of month
  - – `par_end_dt_yyyy`: Parole end date, year
  - – `par_end_dt_mm`: Parole end date, month
  - – `par_end_dt_dd`: Parole end date, day of month
  - – `par_end_cd`: Parole end status classification
  - – `par_end_cd_src`: Raw description of parole end status from source
  - – `par_st_ori_fips`: State FIPS code for location of sentencing
  - – `par_cnty_ori_fips`: County FIPS code for location of sentencing
  - – `par_st_juris_fips`: 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

5.2 Variable codebook

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

These tables provide basic information about the variables such as variable names, labels, formats, a description of each variable, and basic 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 defined as:

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

- Invalid values: values that are outside of the range of what should be possible (e.g., a value of 32 for the day of the month of an event) or are present in qualitative variables with code schemes.
- Missing: in numerical data, values are marked as missing if there were missing in the original, raw data. In most qualitative code schemes, there is an explicit category for missing values, and these are enumerated as missing in the variable-level missing count.

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

5.2.1 Roster

---

cjars_id

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

Set	Count	Percent (%)
All records	163,998,112	100.0
Unique values	59,465,579	

Data notes
- Virginia: VA001

---

dob_yyyy

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

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	144,696,417	88.2
Invalid values	23,409	0.0
Missing values	19,278,286	11.8

Statistic	Value
Mean	1,972.6
Minimum	1,900
25th percentile	1,963
Median	1,974
75th percentile	1,984
Maximum	2,021

Histogram of dob_yyyy (above 5th percentile of valid records)
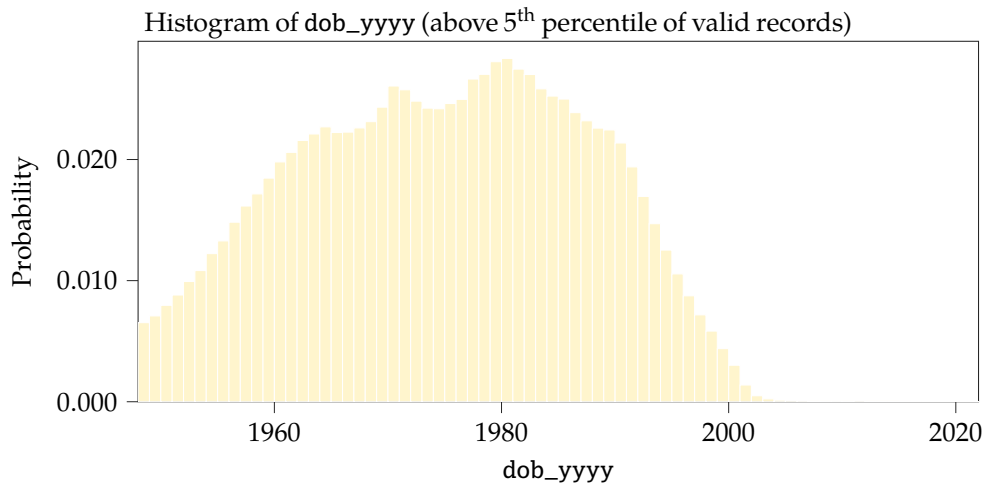
dob_mm (no longer available - deprecated in 2024)

Label	Date of birth, month
Description	The month when the individual was born.
Table	Roster
Format	numeric
Code scheme	month code

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	159,021,490	97.0
Missing values	4,976,622	3.0

Histogram of dob_mm (valid records)

January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12

0   0.02   0.04   0.06   0.08   Prob.
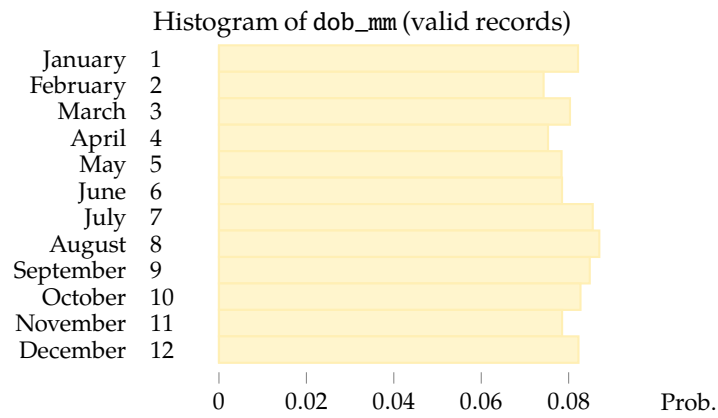
dob_dd (no longer available - deprecated in 2024)

Label   Date of birth, day of month

Description   The day of the month when the individual was born.

Table   Roster
Format   numeric

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	159,021,490	97.0
Missing values	4,976,622	3.0

Histogram of dob_dd (valid records)
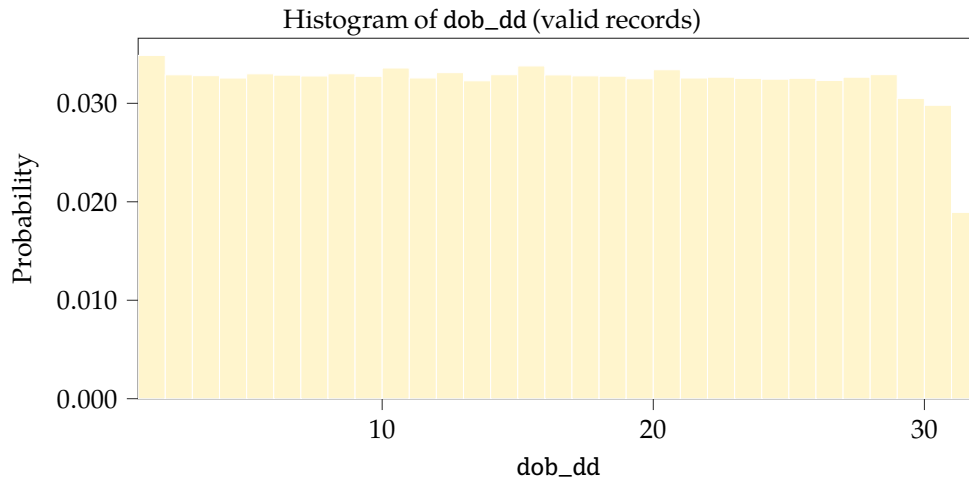
sex

Label	Sex
Description	A variable indicating the individual's sex created by CJARS and based on sex in the source data if available. If missing, sex is imputed based on the frequency of the first name in the CJARS data by Census region by sex. If there are insufficient observations of first name by Census region, sex is imputed based on frequency of first name in Census 2000 by state.
Table	Roster
Format	numeric
Code scheme	sex code

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	162,716,795	99.2
Missing values	1,281,317	0.8

Histogram of sex (valid records)

```
male    1  |██████████████████████|
female  2  |████████|
           |-------|-------|-------|-------|
           0      0.2     0.4     0.6       Prob.
```
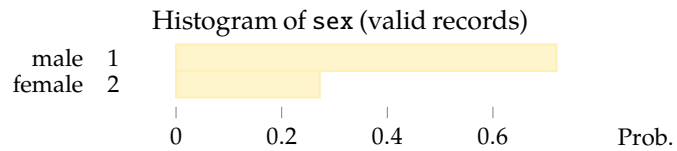
sex_raw

Label   Sex, raw

Description   The individual's sex as recorded by the agency.

Table   Roster
Format   numeric
Code scheme   sex code

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	132,942,654	81.1
Missing values	31,055,458	18.9

Histogram of sex_raw (valid records)
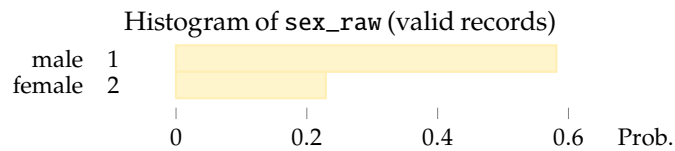
sex_imputed

Label   Sex imputation indicator

Description   An indicator for whether sex was imputed.

Table   Roster
Format   numeric
Code scheme   demographic imputation code

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	162,716,795	99.2
Missing values	1,281,317	0.8

Histogram of sex_imputed (valid records)
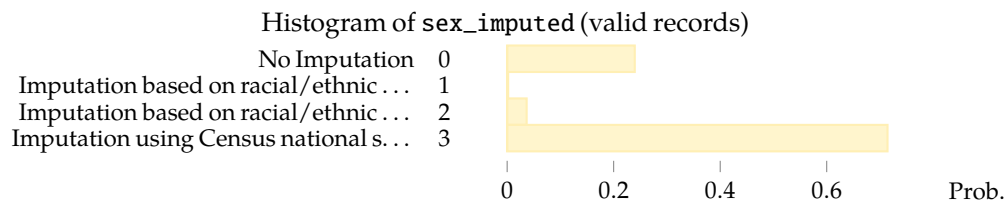
race

Label	Race and ethnicity
Description	A composite race and ethnicity variable imputed by CJARS and standardized to the CJARS race codes. The imputation is based on the frequency of observing the last name with a certain race in the CJARS data combined by Census region, first name with a certain race in the CJARS data combined by Census region, and on Census surnames data. The imputation prioritizes smaller race groups over large race groups (order of priority: AIAN, Asian, Hispanic, Black, White, Other) and differs by sex. Imputation thresholds vary based on the sex of the individual and whether the raw race was missing or being overwritten. If overwriting for males, then the frequency of the individual's first name or last name being associated with that race in the combined CJARS data by Census region must exceed 95% or the frequency of the surname being associated with that race in the Census surnames data must exceed 75%, or the joint probability of the first and last name being observed with that race in the combined CJARS data by Census region must exceed 75%. If imputing missing race for males, then either the last name or first name observed with a certain race must exceed 75% in the CJARS' data combined by Census region or the last name must exceed 75% in the Census surname data. If overwriting race for females, either the frequency of observing a first name with a certain race in a given Census region in the CJARS combined data must exceed 95% or the joint probability of observing a last and first name combination had to exceed 60%. If imputing missing race for females, either the frequency of a first name and race combination in the combined CJARS data in a given Census region must exceed 75% or the joint probability of observing the first and last name combination had to exceed 35%.
Table	Roster
Format	numeric
Code scheme	race and ethnicity code

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	154,059,393	93.9
Missing values	9,938,719	6.1

Histogram of race (valid records)

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

```
     0        0.2        0.4        0.6   Prob.
```
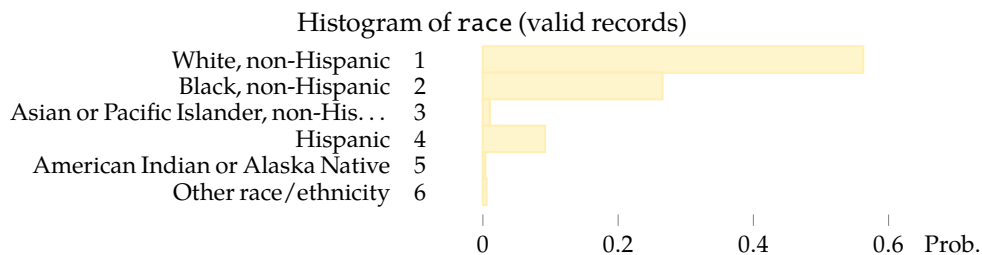
```
race_raw
```

Label   Race and ethnicity, raw

Description   A composite race and ethnicity variable standardized to the
                         CJARS race codes and based on race and ethnicity in the source data if available.

Table   Roster
Format   numeric
Code scheme   race and ethnicity code

Set	*Count*	*Percent (%)*
All records	163,998,112	100.0
Valid records	103,801,146	63.3
Missing values	60,196,966	36.7

Histogram of `race_raw` (valid records)

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

```
          0        0.1       0.2       0.3      Prob.
```
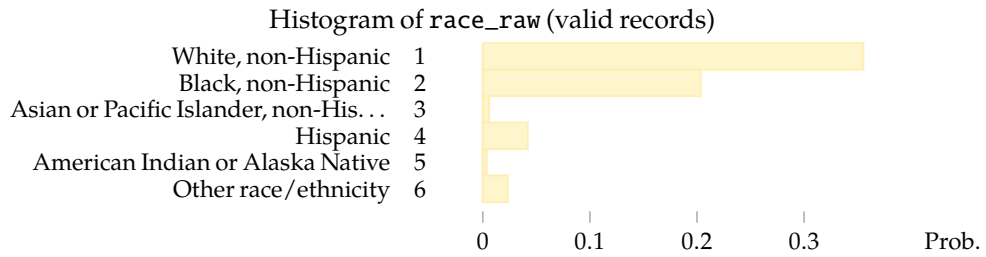
race_imputed

Label   Race imputation flag

Description   An indicator for whether the individual's race was imputed.

Table   Roster
Format   numeric
Code scheme   demographic imputation code

Set	Count	Percent (%)
All records	163,998,112	100.0
Valid records	152,833,897	93.2
Missing values	11,164,215	6.8

Histogram of race_imputed (valid records)

No Imputation	0	
Imputation based on racial/ethnic . . .	1	
Imputation based on racial/ethnic . . .	2	
Imputation using Census national s. . .	3	

```
         0     0.1    0.2    0.3    0.4    Prob.
```

37

5.2.2 Arrest and booking

The arrest table contains information regarding the arrest and booking date, as well as the offense that led to the arrest. Data is at the charge-level.

Table-level data notes
- All states: US012

---

cjars_id

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

Set	*Count*	*Percent (%)*
All records	32,459,075	100.0
Unique values	8,380,641	

Data notes
- Virginia: VA001

---

arr_id

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

arr_arr_dt_yyyy

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

Set	Count	Percent (%)
All records	32,459,075	100.0
Valid records	19,117,391	58.9
Invalid values	2	0.0
Missing values	13,341,682	41.1

Statistic	Value
Mean	2,001.6
Minimum	1,901
25th percentile	1,995
Median	2,005
75th percentile	2,010
Maximum	2,021

Histogram of arr_arr_dt_yyyy (above 5th percentile of valid records)
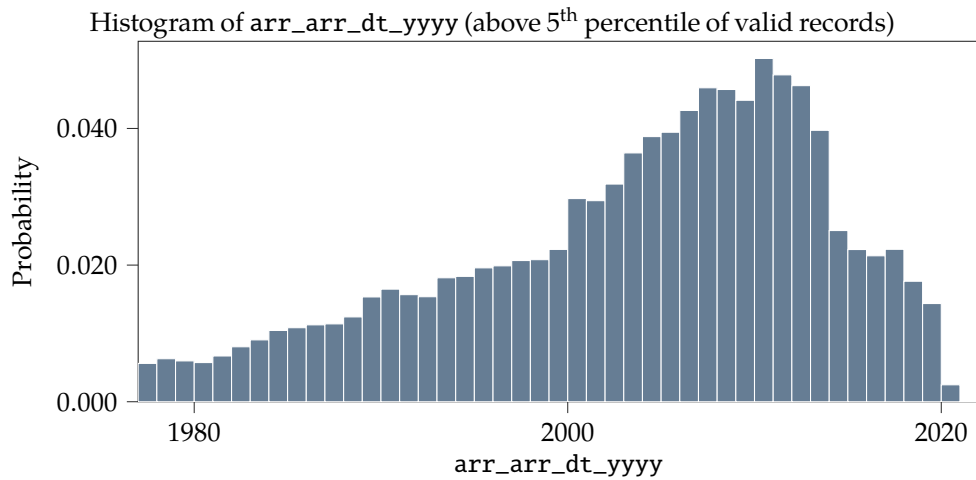
arr_arr_dt_yyyy (continued)
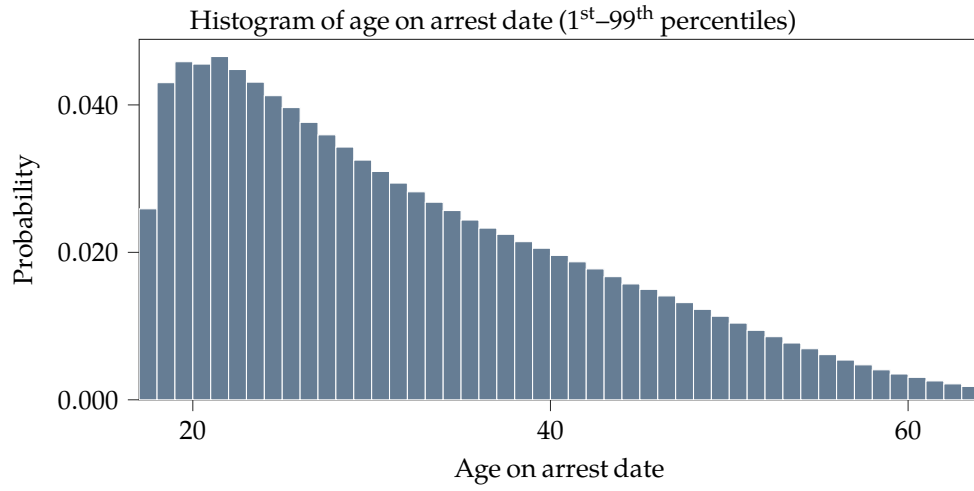
Histogram of age on arrest date (1st–99th percentiles)

Data notes

- California: CA001, CA002, CA003, CA006, CA007
- Colorado: CO001
- Florida: FL004
- Kansas: KS001
- Texas: TX010, TX021, TX023, TX024

arr_arr_dt_mm

Label   Arrest date, month

Description   The month when the individual was arrested.

Table   Arrest and booking
Format   numeric
Code scheme   month code

Set	Count	Percent (%)
All records	32,459,075	100.0
Valid records	19,117,393	58.9
Missing values	13,341,682	41.1

Histogram of arr_arr_dt_mm (valid records)
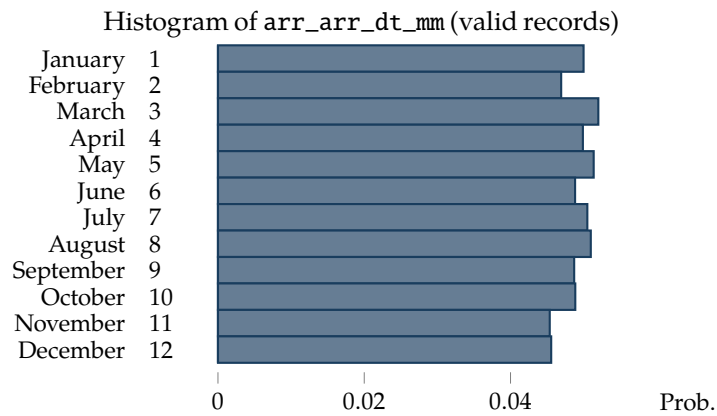
Data notes
- California: CA001, CA002, CA003, CA006, CA007
- Colorado: CO001
- Florida: FL004
- Kansas: KS001
- Texas: TX010, TX021, TX023, TX024

arr_arr_dt_dd

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

Set	Count	Percent (%)
All records	32,459,075	100.0
Valid records	19,117,393	58.9
Missing values	13,341,682	41.1

Histogram of arr_arr_dt_dd (valid records)
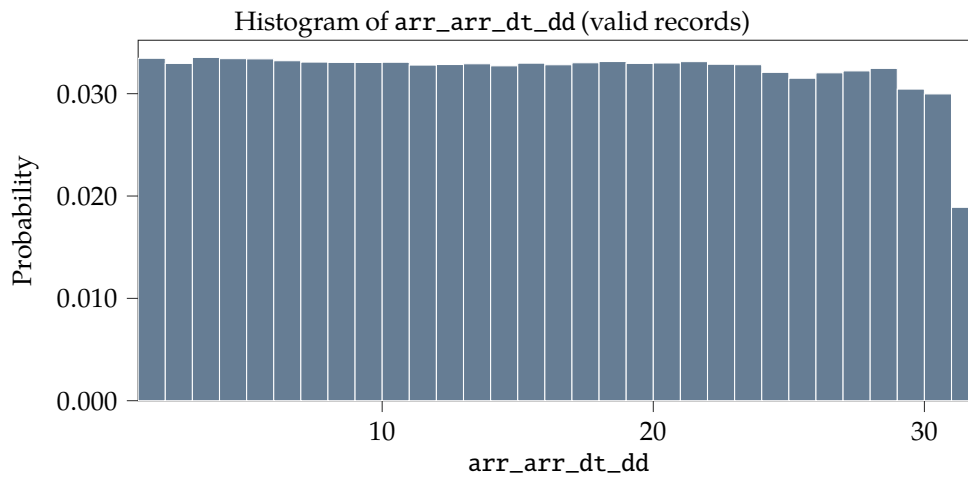
Data notes
- California: CA001, CA002, CA003, CA006, CA007
- Colorado: CO001
- Florida: FL004
- Kansas: KS001
- Texas: TX010, TX021, TX023, TX024

arr_book_dt_yyyy

Label	Booking date, year
Description	The year when the individual was booked into jail.
Table	Arrest and booking
Format	numeric

Set	Count	Percent (%)
All records	32,459,075	100.0
Valid records	15,235,051	46.9
Missing values	17,224,024	53.1

Statistic	Value
Mean	2,003.7
Minimum	1,911
25th percentile	1,997
Median	2,006
75th percentile	2,011
Maximum	2,020
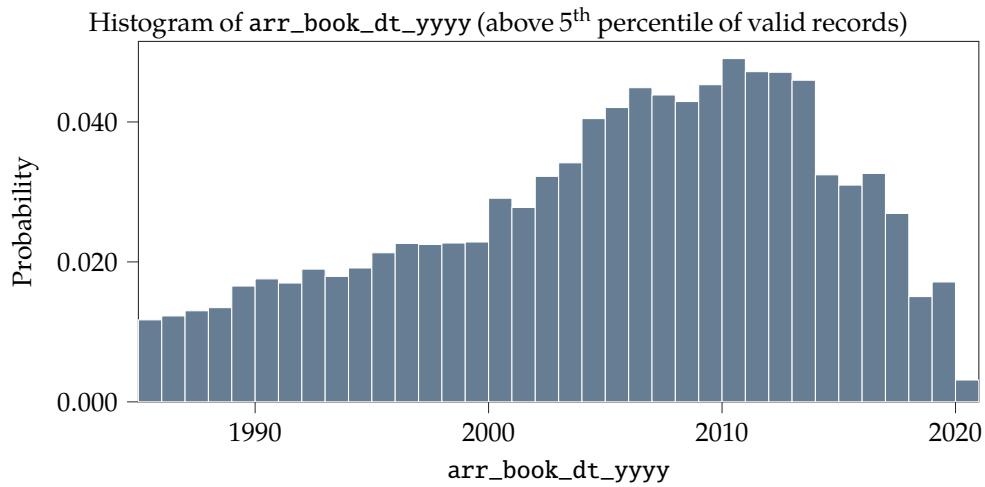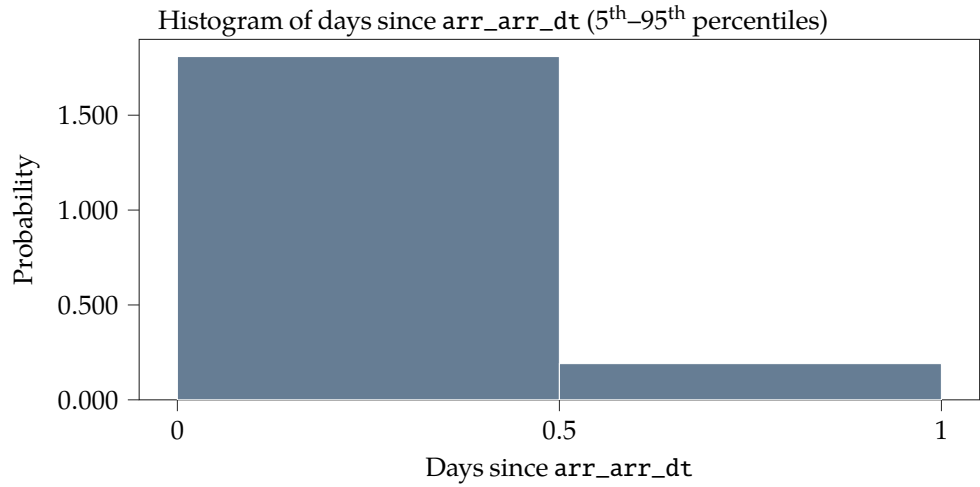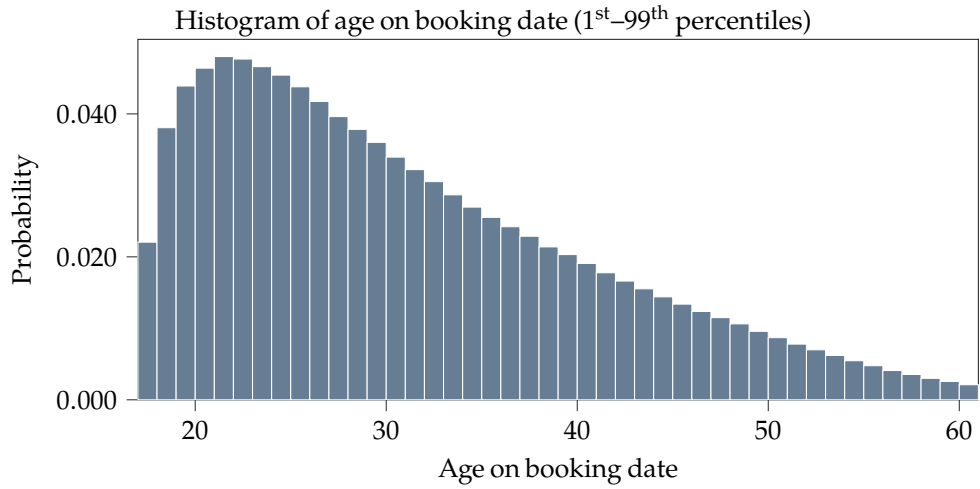
Histogram of arr_book_dt_yyyy (above 5th percentile of valid records)

arr_book_dt_yyyy (continued)

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

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

Data notes
- California: CA004, CA005, CA008, CA009
- Texas: TX002, TX004, TX005, TX007, TX008, TX012, TX019, TX020, TX022, TX025, TX026, TX027

arr_book_dt_mm

Label	Booking date, month
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	32,459,075	100.0
Valid records	15,235,051	46.9
Missing values	17,224,024	53.1

Histogram of arr_book_dt_mm (valid records)

Month	Code
January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12

0   0.01   0.02   0.03   0.04   Prob.
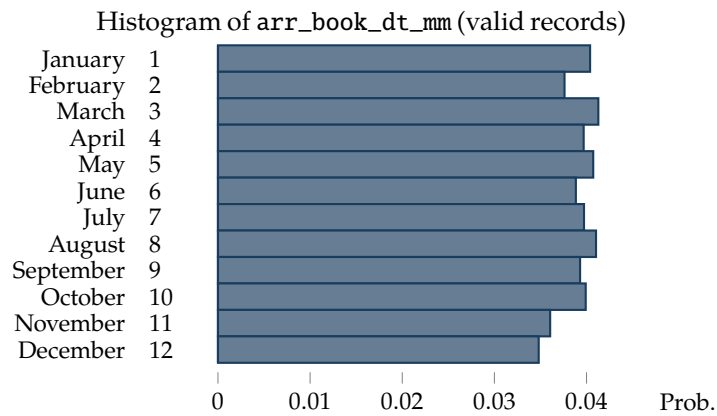
Data notes
- California: CA004, CA005, CA008, CA009
- Texas: TX002, TX004, TX005, TX007, TX008, TX012, TX019, TX020, TX022, TX025, TX026, TX027

arr_book_dt_dd

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

Set	*Count*	*Percent (%)*
All records	32,459,075	100.0
Valid records	15,235,051	46.9
Missing values	17,224,024	53.1

Histogram of arr_book_dt_dd (valid records)

Data notes
- California: CA004, CA005, CA008, CA009
- Texas: TX002, TX004, TX005, TX007, TX008, TX012, TX019, TX020, TX022, TX025, TX026, TX027

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 D.2.1.
Table	Arrest and booking
Format	string
Code scheme	offense classification

Set	*Count*	*Percent (%)*
All records	32,459,075	100.0
Invalid values	32,459,075	100.0

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	32,459,075	100.0

arr_st_ori_fips

Label	State FIPS code
Description	State-level Federal Information Processing Standards (FIPS) code where arrest occurred. 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	32,459,075	100.0
Valid records	32,459,075	100.0

Histogram of `arr_st_ori_fips` (valid records)

State	Code
Alabama	01
Alaska	02
Arizona	04
Arkansas	05
California	06
Colorado	08
Connecticut	09
Delaware	10
District of Columbia	11
Florida	12
Georgia	13
Hawaii	15
Idaho	16
Illinois	17
Indiana	18
Iowa	19
Kansas	20
Kentucky	21
Louisiana	22
Maine	23
Maryland	24
Massachusetts	25
Michigan	26
Minnesota	27
Mississippi	28
Missouri	29
Montana	30
Nevada	32
New Hampshire	33
New Jersey	34
New Mexico	35
New York	36
North Carolina	37
North Dakota	38
Ohio	39
Oklahoma	40
Oregon	41
Pennsylvania	42
Rhode Island	44
South Carolina	45
South Dakota	46
Tennessee	47
Texas	48
Utah	49
Vermont	50
Virginia	51
Washington	53
West Virginia	54
Wisconsin	55
Wyoming	56
Puerto Rico	72
Virgin Islands of the U.S.	78

0   0.2   0.4   0.6   Prob.

arr_cnty_ori_fips

Label	County FIPS code
Description	County-level Federal Information Processing Standards (FIPS) code where arrest occurred. 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	32,459,075	100.0
Valid records	29,604,122	91.2
Invalid values	168,860	0.5
Missing values	2,686,093	8.3

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	32,459,075	100.0
Valid records	32,459,075	100.0

Histogram of arr_rec_src_le (valid records)

		Prob.
No	0	
Yes	1	

0   0.2   0.4

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	32,459,075	100.0
Valid records	32,459,075	100.0

Histogram of arr_rec_src_crt (valid records)

No 0
Yes 1

0   0.2   0.4   0.6   0.8   Prob.
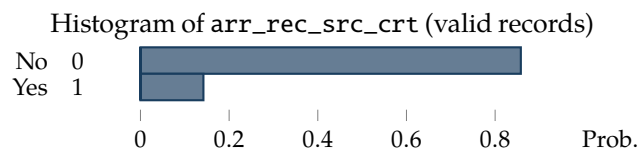
arr_rec_src_doc

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	32,459,075	100.0
Valid records	32,459,075	100.0

Histogram of arr_rec_src_doc (valid records)

No 0

0   0.2   0.4   0.6   0.8   1   Prob.

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	32,459,075	100.0
Valid records	32,459,075	100.0

Histogram of `arr_rec_src_rep` (valid records)

```
No   0  ┌──────────────────────────┐
Yes  1  │────────────┐
        │    │    │    │
        0   0.2  0.4  0.6    Prob.
```

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	32,459,075	100.0
Valid records	32,459,075	100.0

Histogram of `arr_rec_src_cc` (valid records)

```
No   0  ┌────────────────────────────────┐
        │    │    │    │    │    │
        0   0.2  0.4  0.6  0.8   1    Prob.
```

5.2.3 Adjudication

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

Table-level data notes
- All states: US012
- New Jersey: NJ002
- North Carolina: NC004
- Texas: TX018

---

cjars_id

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

Set	*Count*	*Percent (%)*
All records	100,715,876	100.0
Unique values	28,962,834	

Data notes
- Virginia: VA001

---

adj_id

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

adj_grd_cd

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

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_grd_cd (valid records)

Misdemeanor-level charge	MI	
Felony-level charge	FE	
Not known / missing	UU	

0 — 0.2 — 0.4 — 0.6 — Prob.
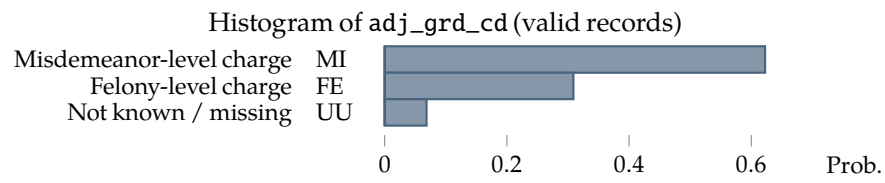
Data notes
- All states: US003, US013
- Nebraska: NE002
- New Jersey: NJ003

adj_grd_cd_src

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

Set	Count	Percent (%)
All records	100,715,876	100.0

Data notes
- All states: US002
- Nebraska: NE002
- New Jersey: NJ003

`adj_off_lgl_cd`

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

Set	*Count*	*Percent (%)*
All records	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of `adj_off_lgl_cd` (valid records)

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

0   0.2   0.4   0.6   0.8   Prob.
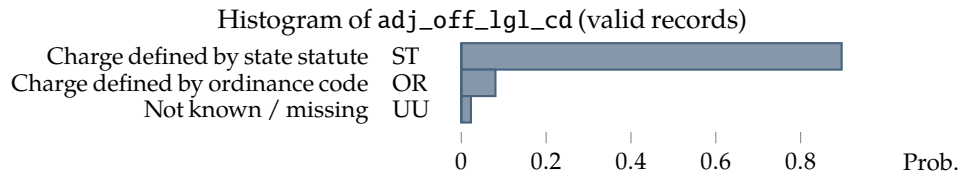
`adj_off_lgl_cd_src`

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

Set	*Count*	*Percent (%)*
All records	100,715,876	100.0

Data notes
- All states: US002

adj_file_dt_yyyy

Label   Case filing date, year

Description   The year when the individual's case was filed.

Table   Adjudication
Format   numeric

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	75,105,183	74.6
Invalid values	1,198	0.0
Missing values	25,609,495	25.4

Statistic	Value
Mean	2,005.8
Minimum	1,900.0
25th percentile	1,999.0
Median	2,008.0
75th percentile	2,013.0
Maximum	2,021.0
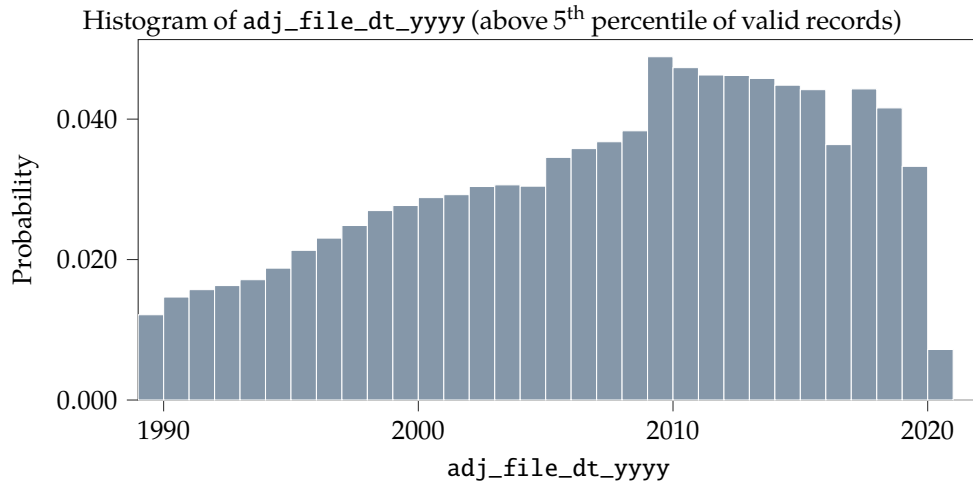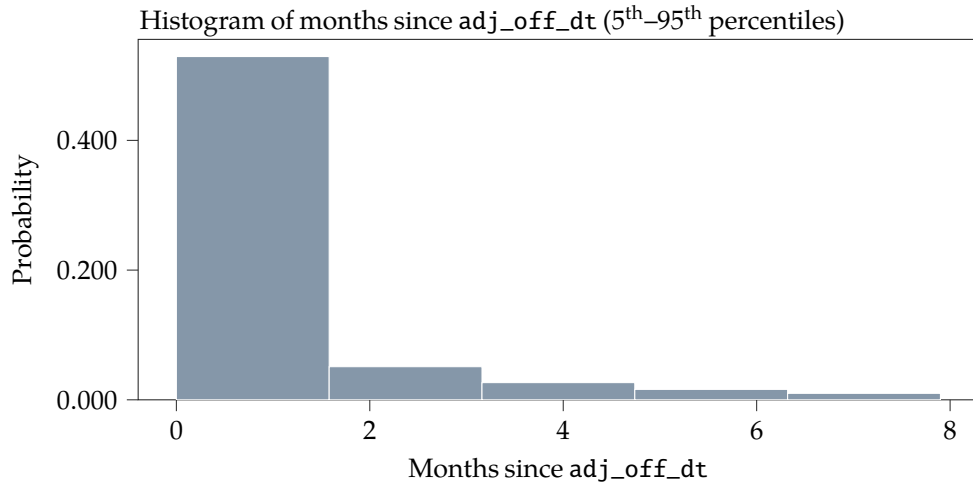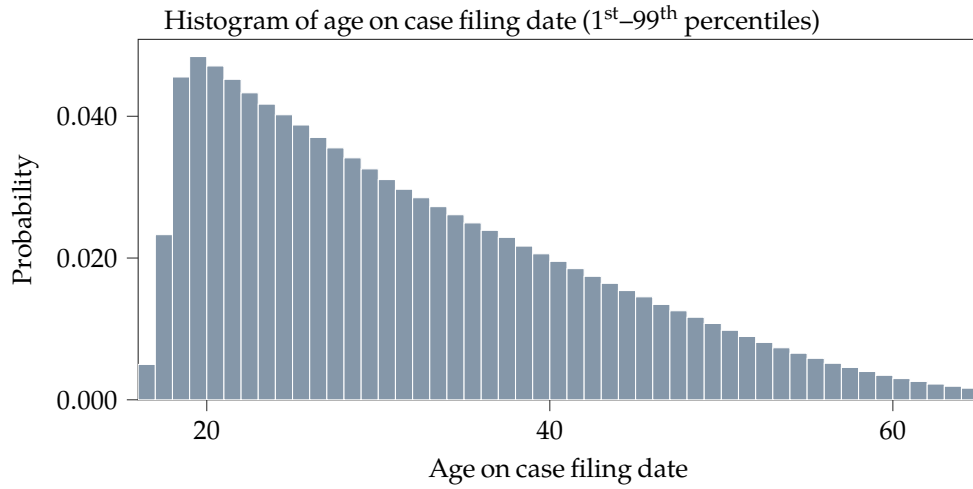
Histogram of `adj_file_dt_yyyy` (above 5th percentile of valid records)

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

Age on case filing date

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

Months since `adj_off_dt`

Data notes
- Arkansas: `AR001`
- Mississippi: `MS001`
- Nebraska: `NE001`
- North Carolina: `NC001`
- Ohio: `OH001`
- Texas: `TX008`, `TX013`, `TX028`

`adj_file_dt_mm`

Label   Case filing date, month

Description   The month when the individual's case was filed.

Table   Adjudication
Format   numeric
Code scheme   month code

Set	*Count*	*Percent (%)*
All records	100,715,876	100.0
Valid records	75,106,381	74.6
Missing values	25,609,495	25.4

Histogram of `adj_file_dt_mm` (valid records)
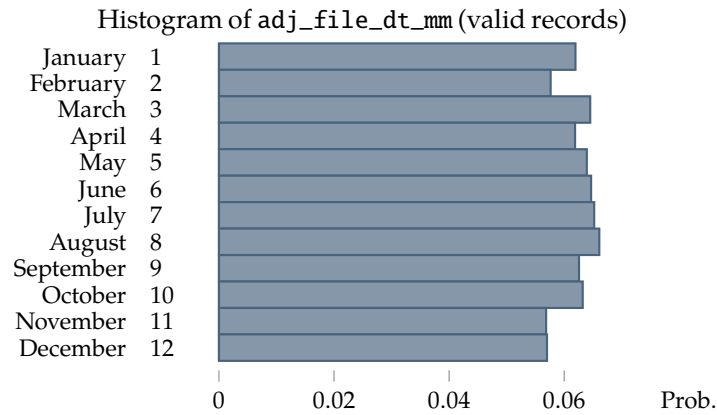
Data notes
- Arkansas: `AR001`
- Mississippi: `MS001`
- Nebraska: `NE001`
- North Carolina: `NC001`
- Ohio: `OH001`
- Texas: `TX008`, `TX013`, `TX028`

adj_file_dt_dd

Label   Case filing date, day of month

Description   The day of the month when the individual's case was filed.

Table   Adjudication
Format   numeric

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	75,106,381	74.6
Missing values	25,609,495	25.4

Histogram of adj_file_dt_dd (valid records)

Data notes
- Arkansas: AR001
- Mississippi: MS001
- Nebraska: NE001
- North Carolina: NC001
- Ohio: OH001
- Texas: TX008, TX013, TX028

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 D.2.1.
Table	Adjudication
Format	string
Code scheme	offense code

Set	Count	Percent (%)
All records	100,715,876	100.0
Invalid values	100,715,876	100.0

Data notes
- Texas: TX018, TX020, TX022, TX028, TX031

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	100,715,876	100.0

adj_disp_dt_yyyy

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

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	83,901,411	83.3
Invalid values	1,051	0.0
Missing values	16,813,414	16.7

Statistic	Value
Mean	2,005.4
Minimum	1,900.0
25th percentile	1,999.0
Median	2,007.0
75th percentile	2,013.0
Maximum	2,020.0

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

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

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

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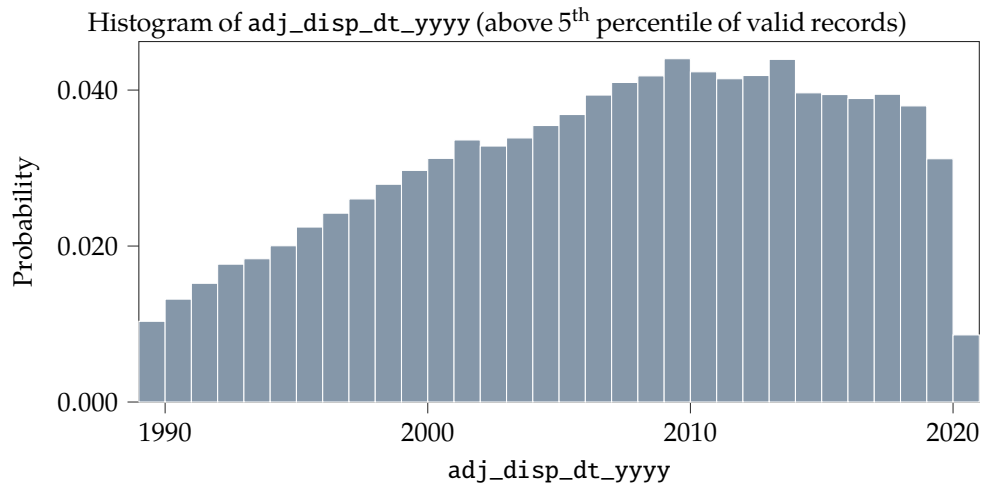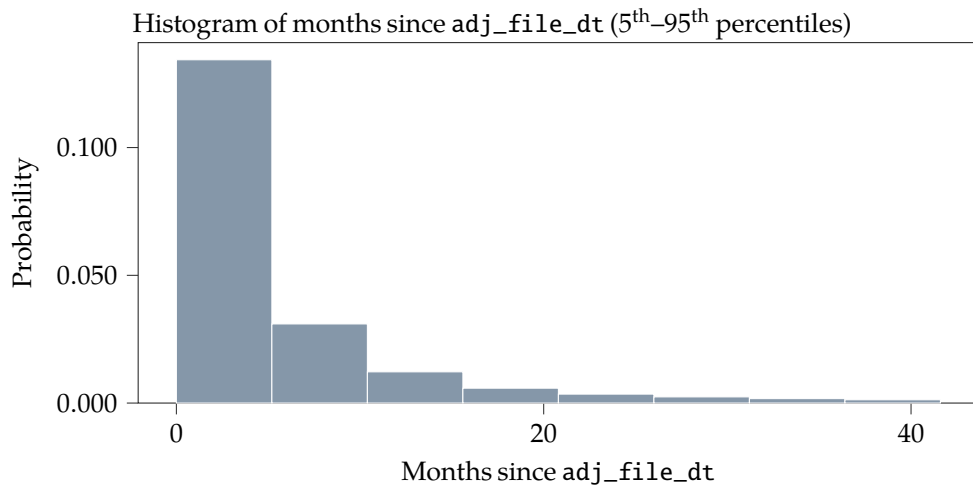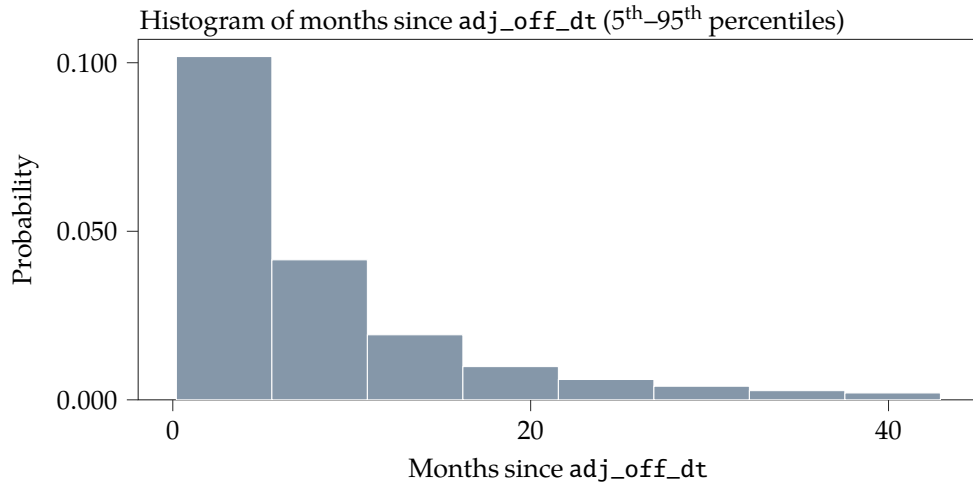
adj_disp_dt_yyyy (continued)

Data notes
- Arizona: AZ003
- Illinois: IL001
- Nebraska: NE001
- Ohio: OH001

adj_disp_dt_mm

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

Set	*Count*	*Percent (%)*
All records	100,715,876	100.0
Valid records	83,902,462	83.3
Missing values	16,813,414	16.7

Histogram of adj_disp_dt_mm (valid records)
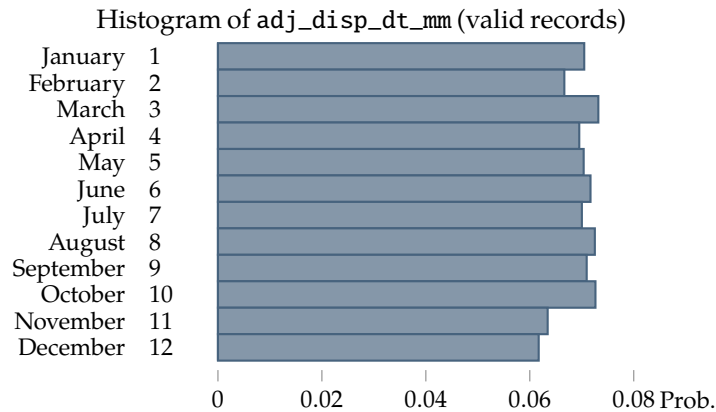
Data notes
- Arizona: AZ003
- Illinois: IL001
- Nebraska: NE001
- Ohio: OH001

adj_disp_dt_dd

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

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	83,902,462	83.3
Missing values	16,813,414	16.7

Histogram of adj_disp_dt_dd (valid records)

Data notes
- Arizona: AZ003
- Illinois: IL001
- Nebraska: NE001
- Ohio: OH001

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 D.2.4.
Table	Adjudication
Format	string
Code scheme	disposition code

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_disp_cd (valid records)

Guilty - unclassified	GU	
Guilty - plea	GP	
Dismissal	ND	
Not known / missing	UU	
Diversion - unclassified	DU	
Procedural - transfer	PT	
Procedural - unclassified	PU	
Not guilty - unclassified	NU	
Mistrial	NM	
Guilty - court trial	GC	
Guilty - jury trial	GJ	
Acquittal	NA	
Dismissal - insanity	NI	
Not guilty plea	NP	
Guilty - insanity	GI	

```
        0        0.1       0.2       0.3  Prob.
```
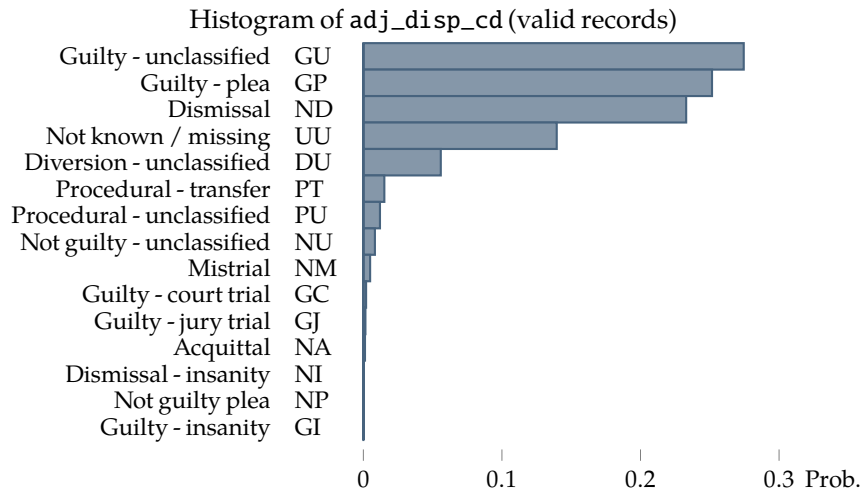
Data notes
- All states: US001, US007
- Minnesota: MN001

66

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	100,715,876	100.0

Data notes
- All states: US002, US007

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 D.2.1.
Table	Adjudication
Format	string
Code scheme	offense classification

Set	Count	Percent (%)
All records	100,715,876	100.0
Invalid values	100,715,876	100.0

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	100,715,876	100.0

Data notes
- Oregon: OR001

adj_off_dt_yyyy

Label   Date offense was committed, year

Description   The year when the individual committed the offense.

Table   Adjudication
Format   numeric

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	62,353,233	61.9
Invalid values	116	0.0
Missing values	38,362,527	38.1

Statistic	Value
Mean	2,005.6
Minimum	1,900.0
25th percentile	2,000.0
Median	2,008.0
75th percentile	2,013.0
Maximum	2,020.0
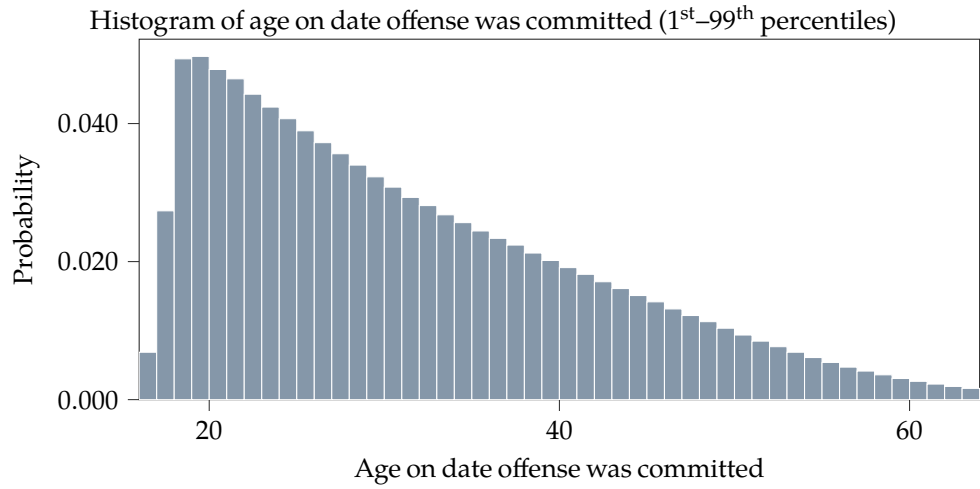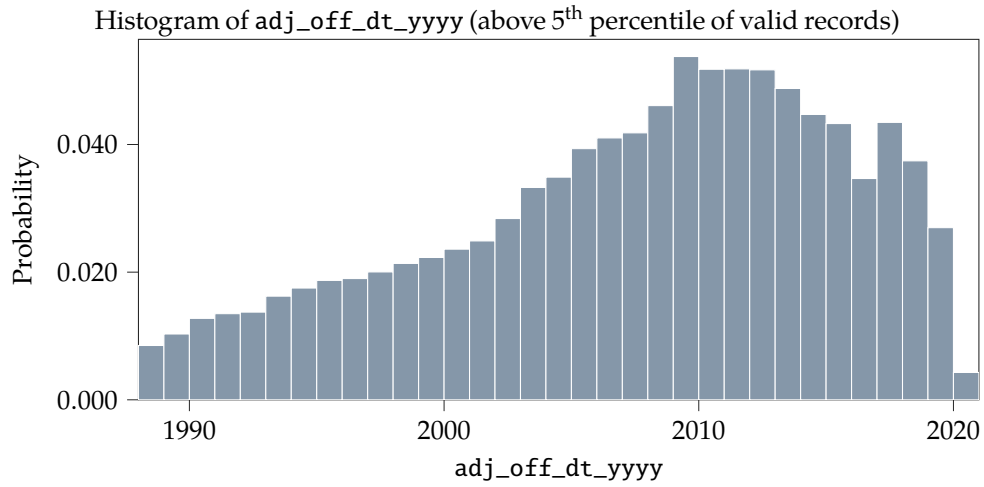
Histogram of `adj_off_dt_yyyy` (above 5th percentile of valid records)

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

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

Data notes

- Arizona: `AZ002`
- Arkansas: `AR001`
- Illinois: `IL001`
- Mississippi: `MS001`
- Nebraska: `NE001`
- Ohio: `OH001`
- Texas: `TX009`, `TX018`, `TX031`

`adj_off_dt_mm`

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

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	62,353,349	61.9
Missing values	38,362,527	38.1

Histogram of `adj_off_dt_mm` (valid records)

January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12

0   0.02   0.04   0.06 Prob.
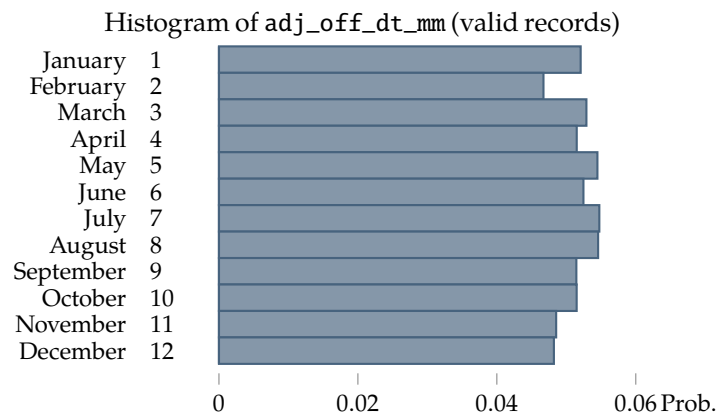
Data notes
- Arizona: `AZ002`
- Arkansas: `AR001`
- Illinois: `IL001`
- Mississippi: `MS001`
- Nebraska: `NE001`
- Ohio: `OH001`
- Texas: `TX009`, `TX018`, `TX031`

adj_off_dt_dd

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

Set	*Count*	*Percent (%)*
All records	100,715,876	100.0
Valid records	62,353,349	61.9
Missing values	38,362,527	38.1

Histogram of `adj_off_dt_dd` (valid records)

Data notes
- Arizona: `AZ002`
- Arkansas: `AR001`
- Illinois: `IL001`
- Mississippi: `MS001`
- Nebraska: `NE001`
- Ohio: `OH001`
- Texas: `TX009, TX018, TX031`

adj_sent_dt_yyyy

Label   Sentencing date, year

Description   The year when the individual was sentenced.

Table   Adjudication
Format   numeric

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	37,029,018	36.8
Invalid values	2,698	0.0
Missing values	63,684,160	63.2

Statistic	Value
Mean	2,005.8
Minimum	1,900.0
25th percentile	2,000.0
Median	2,007.0
75th percentile	2,013.0
Maximum	2,020.0
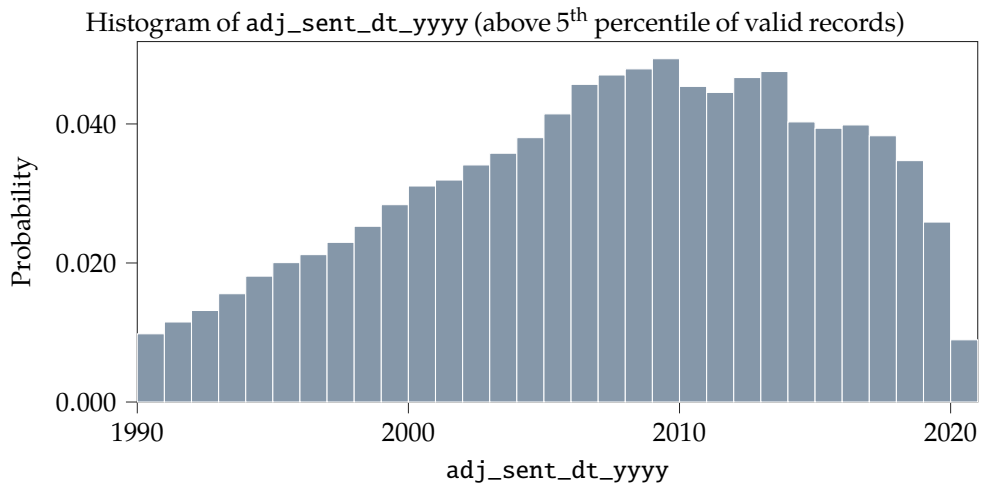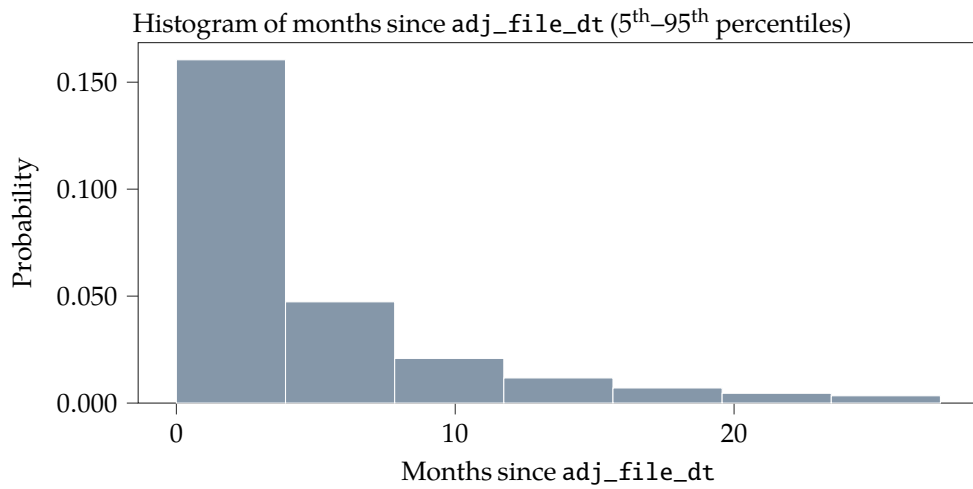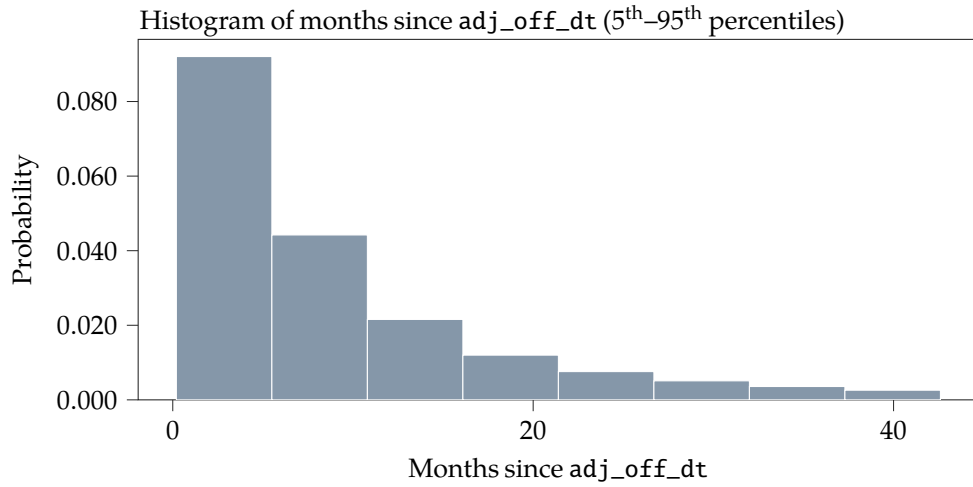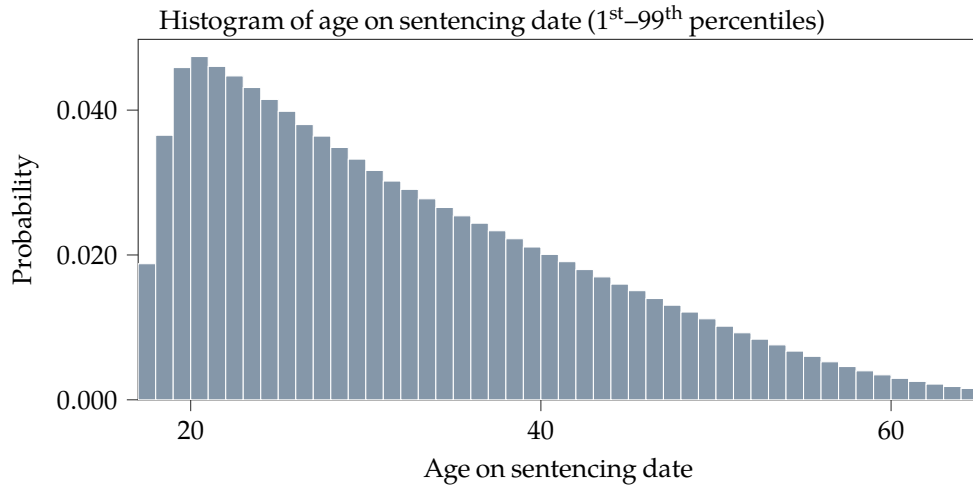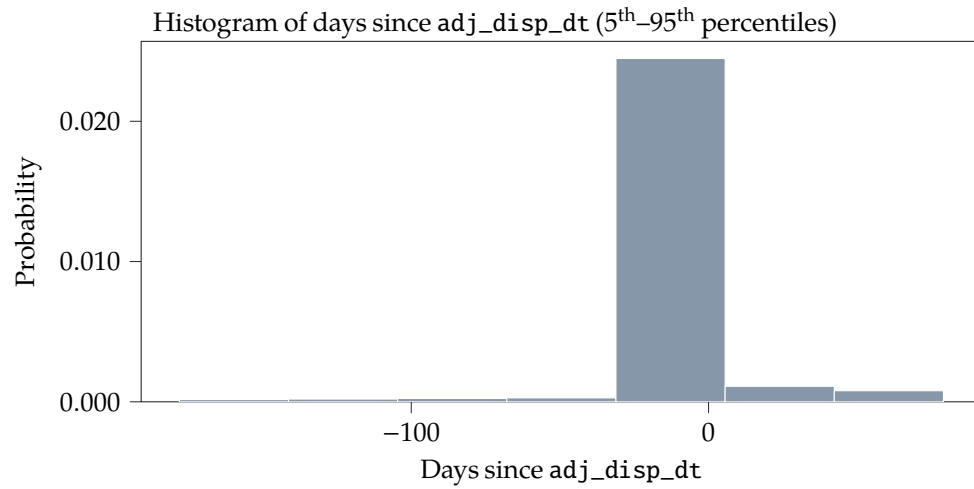
Histogram of adj_sent_dt_yyyy (above 5th percentile of valid records)

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

Probability

Age on sentencing date

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

Probability

Months since `adj_off_dt`

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

Probability

Months since `adj_file_dt`

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

Days since `adj_disp_dt`

Data notes
- Arizona: `AZ002`
- Florida: `FL001`
- Illinois: `IL001`
- Nebraska: `NE001`
- New Jersey: `NJ001`
- North Carolina: `NC002`
- Ohio: `OH001`
- Pennsylvania: `PA001`
- Texas: `TX001`, `TX003`, `TX006`, `TX009`, `TX011`, `TX020`, `TX022`, `TX029`, `TX031`, `TX032`

adj_sent_dt_mm

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

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	37,031,716	36.8
Missing values	63,684,160	63.2

Histogram of `adj_sent_dt_mm` (valid records)
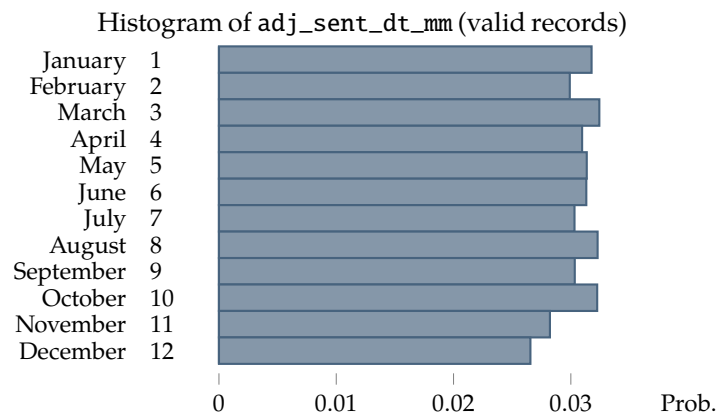
Data notes
- Arizona: `AZ002`
- Florida: `FL001`
- Illinois: `IL001`
- Nebraska: `NE001`
- New Jersey: `NJ001`
- North Carolina: `NC002`
- Ohio: `OH001`
- Pennsylvania: `PA001`
- Texas: `TX001`, `TX003`, `TX006`, `TX009`, `TX011`, `TX020`, `TX022`, `TX029`, `TX031`, `TX032`

adj_sent_dt_dd

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

Set	Count	Percent (%)
All records	100,715,876	100.0
Valid records	37,031,716	36.8
Missing values	63,684,160	63.2

Histogram of adj_sent_dt_dd (valid records)
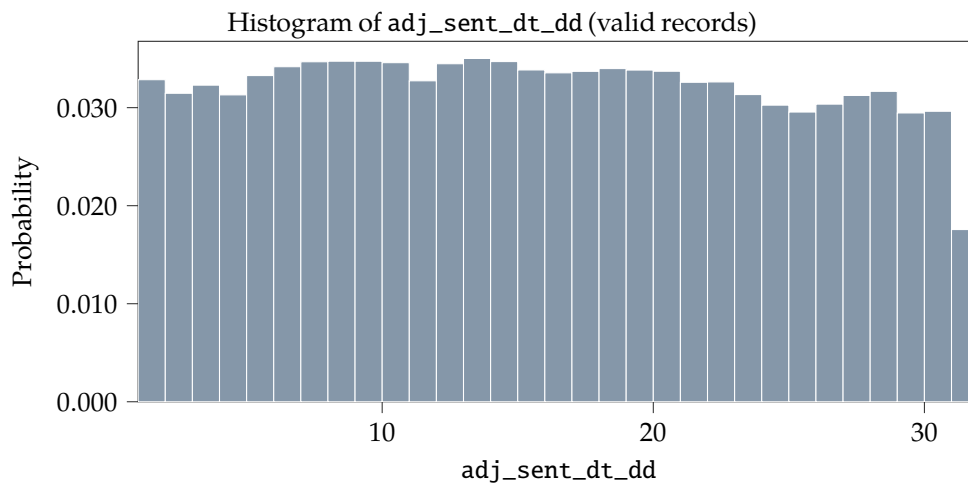
Data notes
- Arizona: AZ002
- Florida: FL001
- Illinois: IL001
- Nebraska: NE001
- New Jersey: NJ001
- North Carolina: NC002
- Ohio: OH001
- Pennsylvania: PA001
- Texas: TX001, TX003, TX006, TX009, TX011, TX020, TX022, TX029, TX031, TX032

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	100,715,876	100.0
Valid records	34,294,925	34.1
Missing values	66,420,951	65.9
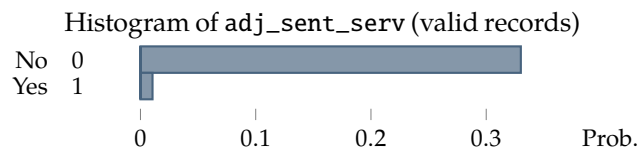
Histogram of `adj_sent_serv` (valid records)

No   0
Yes 1

0   0.1   0.2   0.3   Prob.

Data notes
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX018, TX020, TX022, TX027, TX028, TX029, TX030, TX031

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	100,715,876	100.0
Valid records	24,012,156	23.8
Missing values	76,703,720	76.2
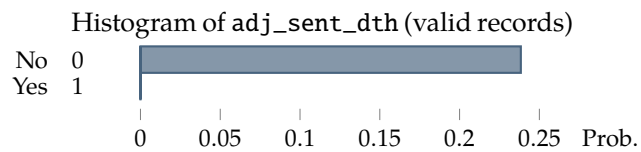
Histogram of adj_sent_dth (valid records)

No 0
Yes 1

0   0.05   0.1   0.15   0.2   0.25   Prob.

Data notes
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX018, TX020, TX022, TX028, TX032

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	100,715,876	100.0
Valid records	21,073,179	20.9
Missing values	79,642,697	79.1
Special values		
Life sentence (-88888)	49,562	0.0
Death sentence (-99999)	1,844	0.0

Statistic	*Value*
Mean	13.3
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	6.0
Maximum	1,200.0
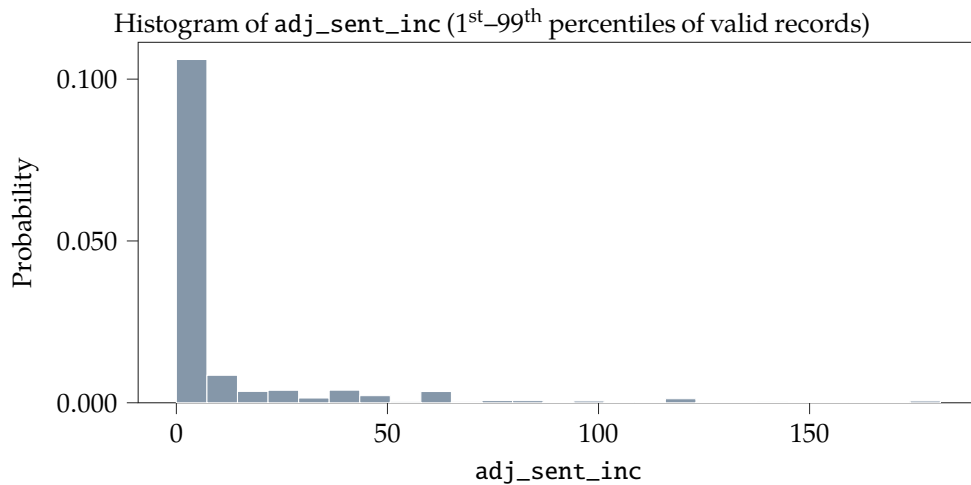
Histogram of `adj_sent_inc` (1st–99th percentiles of valid records)

Data notes
- All states: US004, US008
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX020, TX028, TX029, TX032

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	100,715,876	100.0
Valid records	29,500,935	29.3
Missing values	71,214,941	70.7

Statistic	Value
Mean	10.7
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	12.1
Maximum	120.0

Histogram of adj_sent_pro (1st–99th percentiles of valid records)
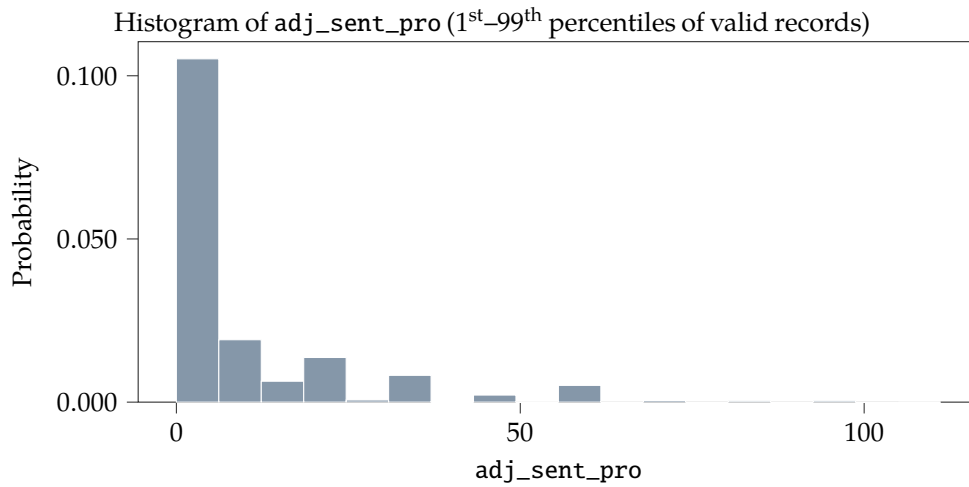
Data notes
- All states: US009
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX020, TX022, TX028, TX029, TX032

`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	100,715,876	100.0
Valid records	14,481,976	14.4
Missing values	86,233,900	85.6

Statistic	*Value*
Mean	1,032.9
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	0.0
Maximum	500,000.0

Histogram of `adj_sent_rest` (1st–99th percentiles of valid records)
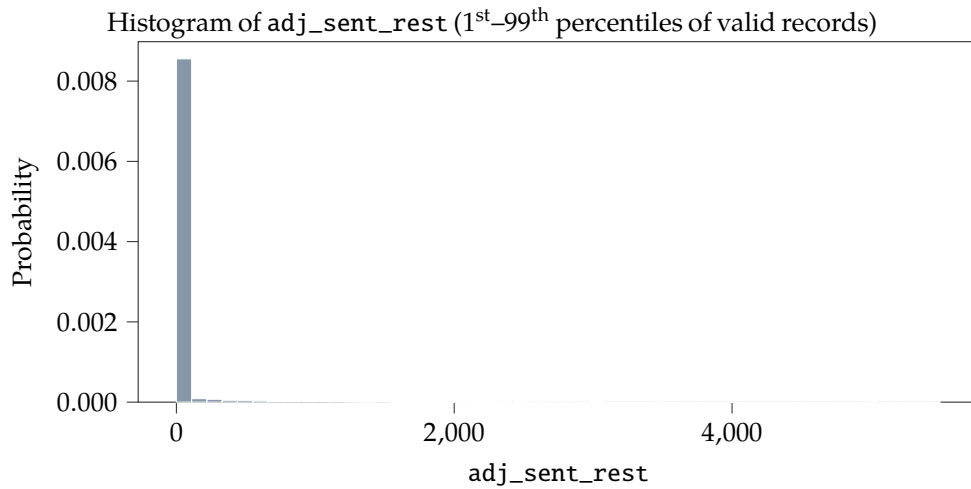
Data notes
- All states: US005, US010
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX020, TX022, TX027, TX028, TX029, TX030, TX031, TX032

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	100,715,876	100.0
Valid records	59,647,185	59.2
Missing values	41,068,691	40.8

Histogram of adj_sent_sus (valid records)

No   0
Yes 1

0   0.2   0.4   Prob.
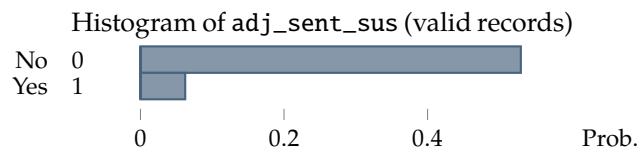
Data notes
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX028

82

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	100,715,876	100.0
Valid records	8,703,216	8.6
Missing values	92,012,660	91.4
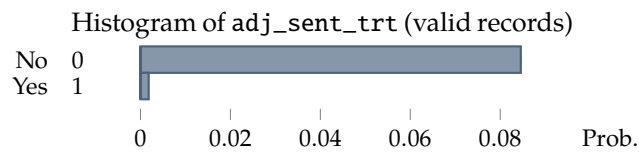
Histogram of adj_sent_trt (valid records)

No   0
Yes 1

0   0.02   0.04   0.06   0.08   Prob.

Data notes
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX018, TX020, TX022, TX027, TX028, TX029, TX030, TX031

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	100,715,876	100.0
Valid records	32,875,301	32.6
Invalid values	904	0.0
Missing values	67,839,671	67.4

Statistic	Value
Mean	443.1
Minimum	0.0
25$^{\text{th}}$ percentile	0.0
Median	0.0
75$^{\text{th}}$ percentile	200.0
Maximum	500,000.0

Histogram of adj_sent_fine (1$^{\text{st}}$–99$^{\text{th}}$ percentiles of valid records)
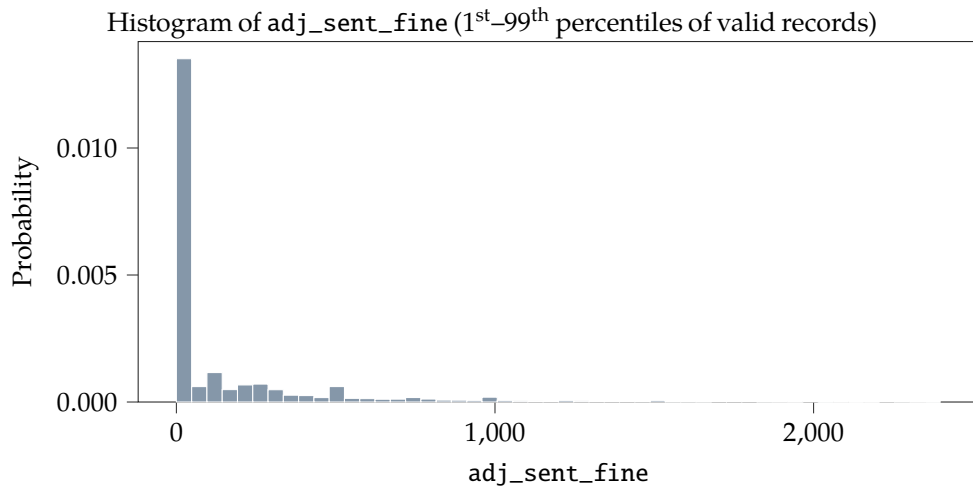
Data notes
- All states: US005, US011
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX020, TX022, TX028, TX032

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	100,715,876	100.0
Valid records	4,853,054	4.8
Missing values	95,862,822	95.2
Special values		
Life sentence (-88888)	3,306	0.0
Death sentence (-99999)	960	0.0

Statistic	Value
Mean	14.2
Minimum	0.0
25th percentile	0.0
Median	3.9
75th percentile	12.1
Maximum	1,200.0
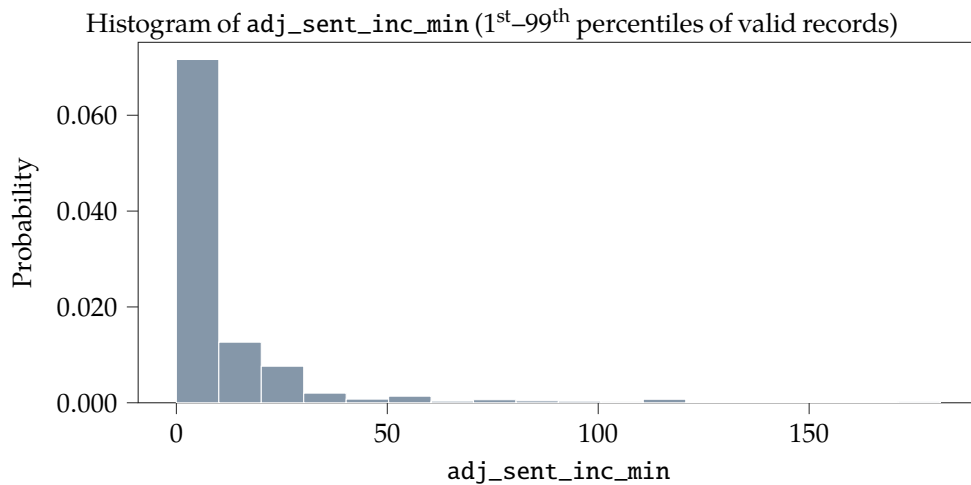
Histogram of `adj_sent_inc_min` (1st–99th percentiles of valid records)

Data notes
- All states: US004, US008
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX018, TX020, TX022, TX027, TX028, TX029, TX030, TX031, TX032

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	100,715,876	100.0
Valid records	6,724,946	6.7
Missing values	93,990,930	93.3
Special values		
Life sentence (-88888)	19,908	0.0
Death sentence (-99999)	14,155	0.0

Statistic	Value
Mean	36.7
Minimum	0.0
25th percentile	0.6
Median	10.0
75th percentile	30.1
Maximum	1,200.0
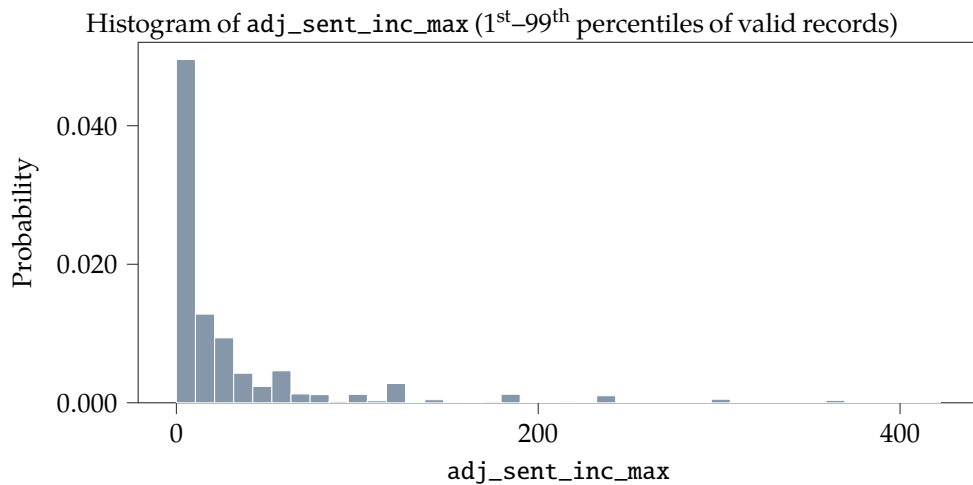
Histogram of `adj_sent_inc_max` (1st–99th percentiles of valid records)

Data notes
- All states: US004, US008
- New Jersey: NJ001
- Pennsylvania: PA001
- Texas: TX008, TX018, TX020, TX022, TX027, TX028, TX029, TX030, TX031, TX032

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	100,715,876	100.0

Data notes
- All states: `US002`
- Arizona: `AZ002`
- Oregon: `OR001`

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	100,715,876	100.0
Valid records	100,715,876	100.0

adj_st_ori_fips (continued)

Histogram of adj_st_ori_fips (valid records)

Alabama	01
Alaska	02
Arizona	04
Arkansas	05
California	06
Colorado	08
Connecticut	09
Delaware	10
District of Columbia	11
Florida	12
Georgia	13
Hawaii	15
Idaho	16
Illinois	17
Indiana	18
Iowa	19
Kansas	20
Kentucky	21
Louisiana	22
Maine	23
Maryland	24
Massachusetts	25
Michigan	26
Minnesota	27
Mississippi	28
Missouri	29
Montana	30
Nebraska	31
Nevada	32
New Hampshire	33
New Jersey	34
New Mexico	35
New York	36
North Carolina	37
North Dakota	38
Ohio	39
Oklahoma	40
Oregon	41
Pennsylvania	42
Rhode Island	44
South Carolina	45
South Dakota	46
Tennessee	47
Texas	48
Utah	49
Vermont	50
Virginia	51
Washington	53
West Virginia	54
Wisconsin	55
Wyoming	56
Puerto Rico	72

0   0.05   0.1   0.15   0.2   Prob.

Data notes
- All states: US006

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	100,715,876	100.0
Valid records	92,934,008	92.3
Invalid values	617,925	0.6
Missing values	7,163,943	7.1

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	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_rec_src_le (valid records)

No   0

0   0.2   0.4   0.6   0.8   1   Prob.

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	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_rec_src_crt (valid records)

No   0
Yes 1

0   0.2   0.4   0.6   0.8   Prob.

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	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_rec_src_doc (valid records)

No   0
Yes 1

0   0.2   0.4   0.6   0.8   1 Prob.

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	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_rec_src_rep (valid records)

No   0
Yes 1

0   0.2   0.4   0.6   0.8   Prob.

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	100,715,876	100.0
Valid records	100,715,876	100.0

Histogram of adj_rec_src_cc (valid records)

No   0

0   0.2   0.4   0.6   0.8   1   Prob.

5.2.4 Probation

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

Table-level data notes
- All states: US012
- Michigan: MI001

cjars_id		
Label	CJARS person identifier	
Description	Uniquely identifies individuals. For more details on use of cjars_id for data linkage, refer to Section 3.3.1 and Figure 6.	
Table	Probation	
Table key	foreign, not unique	
Format	string	
Set	*Count*	*Percent (%)*
All records	8,649,643	100.0
Missing values	1	0.0
Unique values	5,936,303	

Data notes
- Virginia: VA001

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

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 D.2.5.
Table	Probation
Format	string
Code scheme	probation condition code

Set	*Count*	*Percent (%)*
All records	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_cond_cd (valid records)

Not known / missing	UU
Straight probation	SP
Alcohol/drug residential	AD
Probation with community residential	PR

0   0.2   0.4   0.6   0.8   1   Prob.
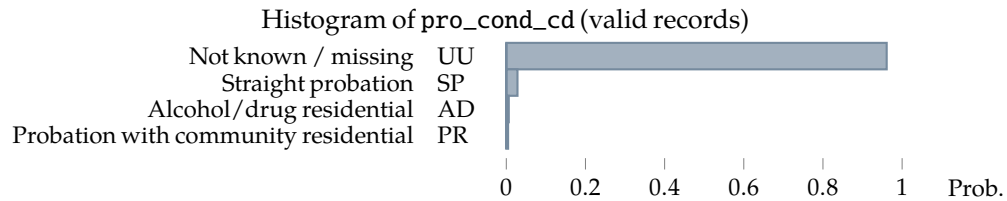
Data notes
- Texas: TX017

pro_cond_cd_src

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

Set	Count	Percent (%)
All records	8,649,643	100.0

Data notes
- All states: US002

pro_bgn_dt_yyyy

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

Set	Count	Percent (%)
All records	8,649,643	100.0
Valid records	8,223,379	95.1
Invalid values	53	0.0
Missing values	426,211	4.9

Statistic	Value
Mean	2,007.4
Minimum	1,900
25th percentile	2,003
Median	2,009
75th percentile	2,014
Maximum	2,020

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

pro_bgn_dt_yyyy (continued)

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

Probability / Age on probation start date
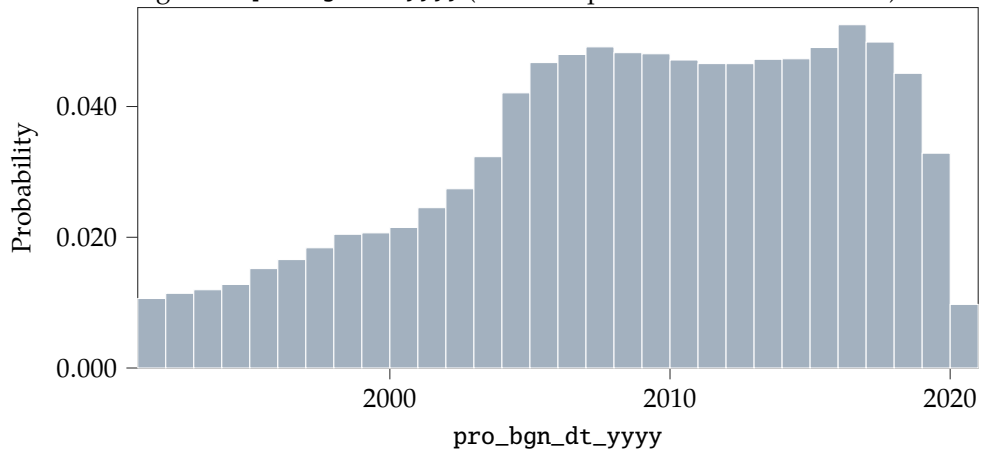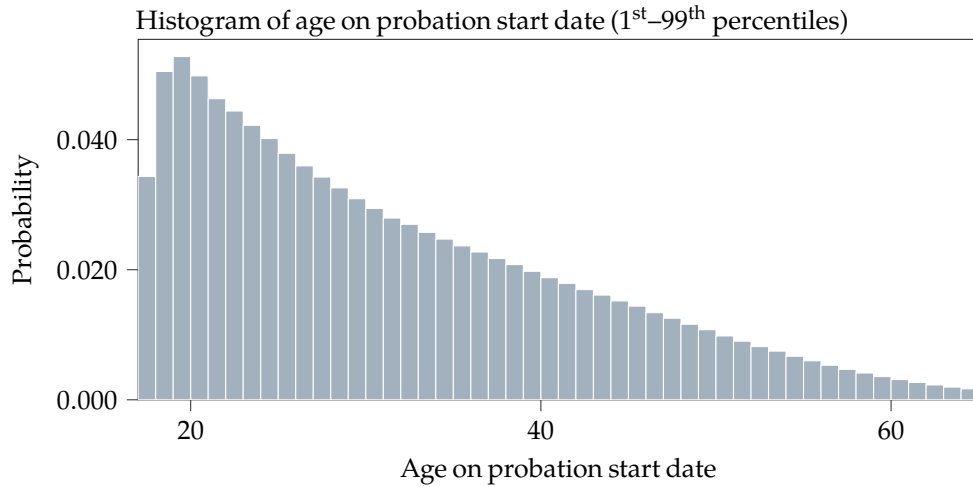
pro_bgn_dt_mm

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

Set	Count	Percent (%)
All records	8,649,643	100.0
Valid records	8,223,432	95.1
Missing values	426,211	4.9

Histogram of pro_bgn_dt_mm (valid records)

January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12

0   0.02   0.04   0.06   0.08   Prob.
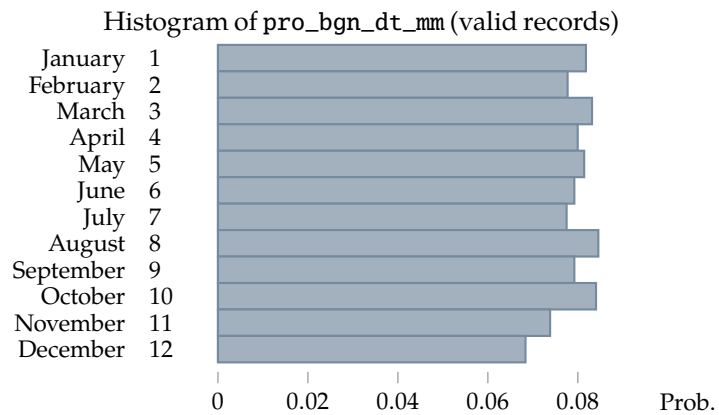
pro_bgn_dt_dd

Label   Probation start date, day of month

Description   The day of the month when the individual began probation.

Table   Probation
Format   numeric

Set	Count	Percent (%)
All records	8,649,643	100.0
Valid records	8,223,432	95.1
Missing values	426,211	4.9

Histogram of pro_bgn_dt_dd (valid records)
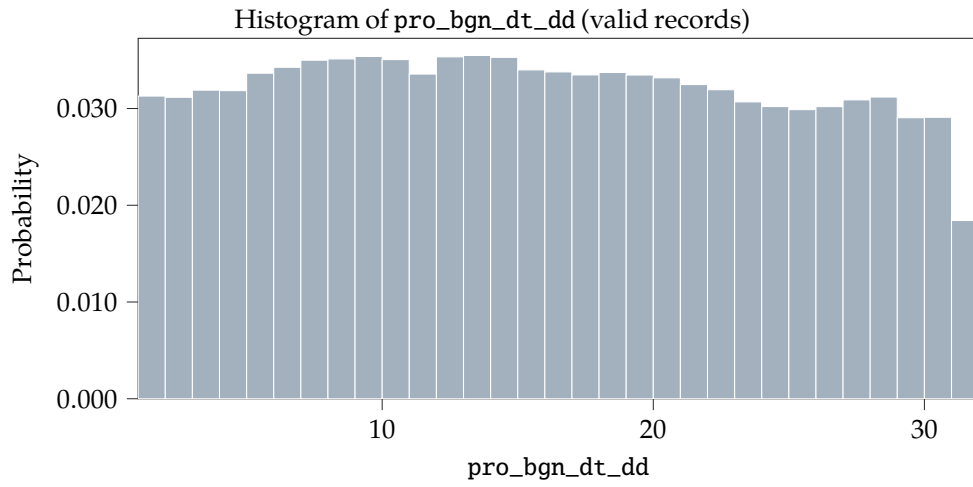
pro_end_dt_yyyy

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

Set	Count	Percent (%)
All records	8,649,643	100.0
Valid records	5,096,519	58.9
Invalid values	2	0.0
Missing values	3,553,122	41.1

Statistic	Value
Mean	2,009.0
Minimum	1,900.0
25th percentile	2,005.0
Median	2,010.0
75th percentile	2,014.0
Maximum	2,020.0

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

pro_end_dt_yyyy (continued)

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

Histogram of months since pro_bgn_dt (5th–95th percentiles)
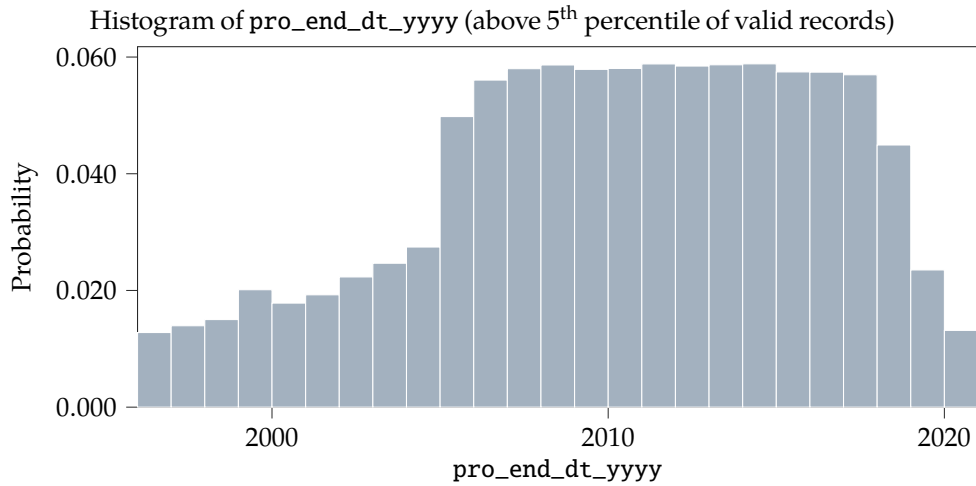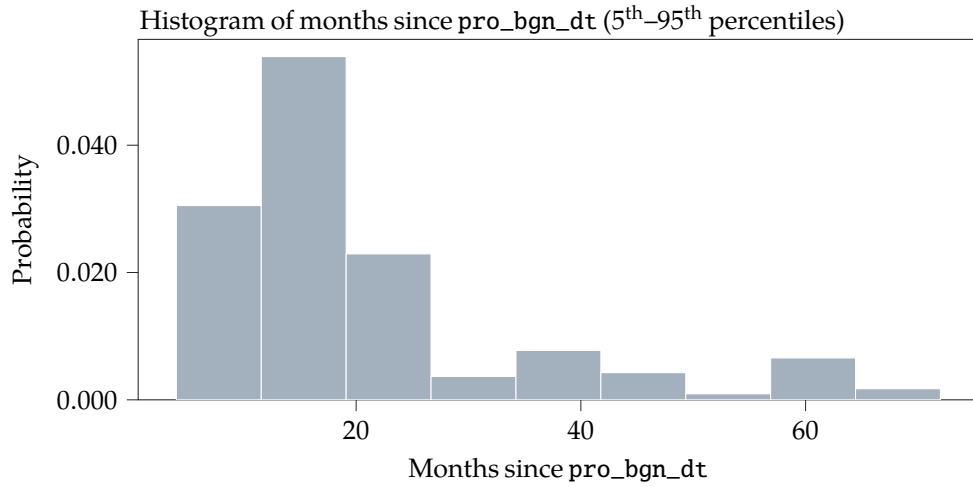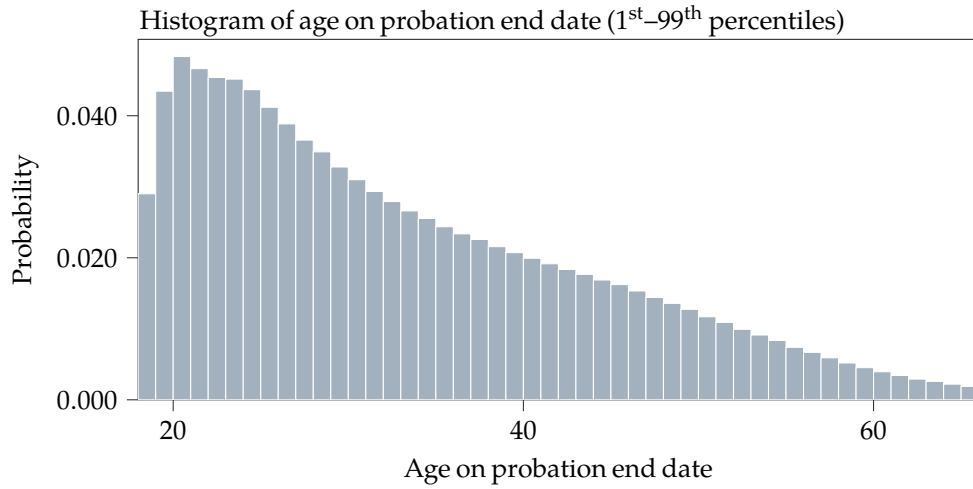
Data notes
- Arkansas: AR003
- Florida: FL003
- North Carolina: NC003
- Texas: TX017

pro_end_dt_mm

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

Set	Count	Percent (%)
All records	8,649,643	100.0
Valid records	5,096,521	58.9
Missing values	3,553,122	41.1

Histogram of pro_end_dt_mm (valid records)

Data notes
- Arkansas: AR003
- Florida: FL003
- North Carolina: NC003
- Texas: TX017

pro_end_dt_dd

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

Set	*Count*	*Percent (%)*
All records	8,649,643	100.0
Valid records	5,096,521	58.9
Missing values	3,553,122	41.1

Histogram of pro_end_dt_dd (valid records)
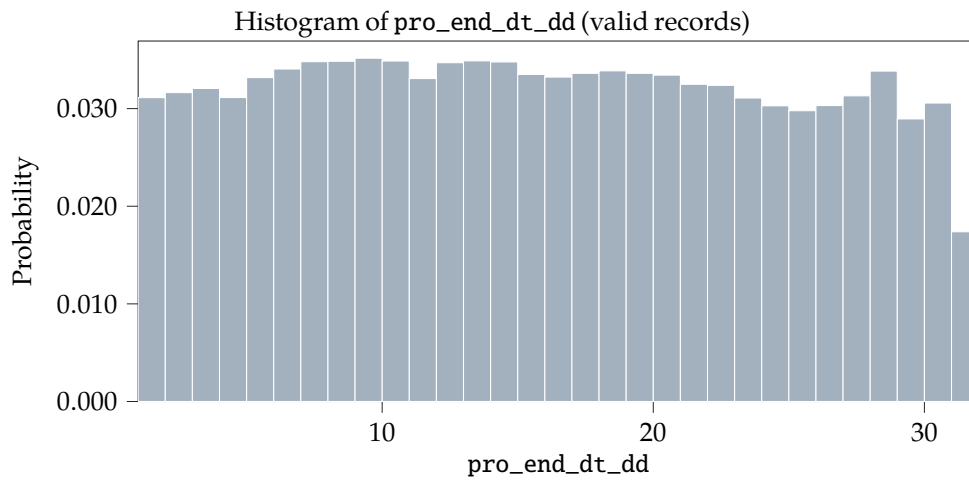
Data notes
- Arkansas: AR003
- Florida: FL003
- North Carolina: NC003
- Texas: TX017

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 D.2.6.
Table	Probation
Format	string
Code scheme	probation exit code

Set	Count	Percent (%)
All records	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_end_cd (valid records)

Not known / missing	UU	
Completion	CO	
Incarcerated	IN	
Absconded/escaped	AB	
Other unsatisfactory exit	OU	
Other	OT	
Transferred to another probation a...	TR	
Death	DE	
Discharged to custody/detainer/war...	DI	

0   0.2   0.4   0.6   0.8   Prob.

Data notes
- Texas: TX017

103

pro_end_cd_src

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

Set	*Count*	*Percent (%)*
All records	8,649,643	100.0

Data notes
- All states: US002

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	8,649,643	100.0
Valid records	8,548,234	98.8
Missing values	101,409	1.2

pro_st_ori_fips (continued)

Histogram of pro_st_ori_fips (valid records)

State	Code
Arkansas	05
Florida	12
Maryland	24
Michigan	26
North Carolina	37
Texas	48
Wisconsin	55

0   0.1   0.2   0.3   Prob.

---

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	8,649,643	100.0
Valid records	7,762,644	89.7
Missing values	785,590	9.1

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 identi-fying 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	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_st_juris_fips (valid records)

Arkansas	05	
Florida	12	
Maryland	24	
Michigan	26	
North Carolina	37	
Texas	48	
Wisconsin	55	

0   0.1   0.2   0.3   Prob.

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	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_rec_src_le (valid records)

No   0

0	0.2	0.4	0.6	0.8	1	Prob.
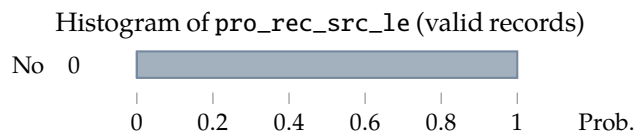
pro_rec_src_crt

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	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_rec_src_crt (valid records)

No   0
Yes   1

0	0.2	0.4	0.6	0.8	Prob.
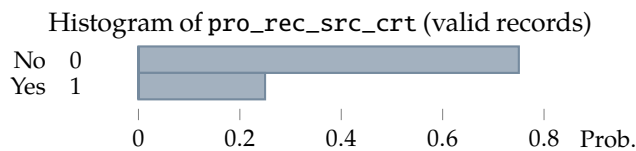
pro_rec_src_doc

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	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_rec_src_doc (valid records)

No 0
Yes 1

0   0.2   0.4   0.6 Prob.
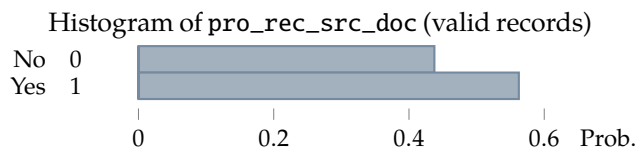
pro_rec_src_rep

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	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_rec_src_rep (valid records)

No 0
Yes 1

0   0.2   0.4   0.6   0.8   Prob.

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	8,649,643	100.0
Valid records	8,649,643	100.0

Histogram of pro_rec_src_cc (valid records)

No	0	

```
      0    0.2   0.4   0.6   0.8   1     Prob.
```

5.2.5 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. Data is at the level of an incarceration term.

Table-level data notes
- All states: US012
- Arizona: AZ001

---

cjars_id

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

Set	*Count*	*Percent (%)*
All records	6,399,324	100.0
Unique values	3,618,360	

Data notes
- Virginia: VA001

---

inc_id

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

```
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 D.2.8.

Table   Incarceration
Format   string
Code scheme   incarceration facility type

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of `inc_fcl_cd` (valid records)
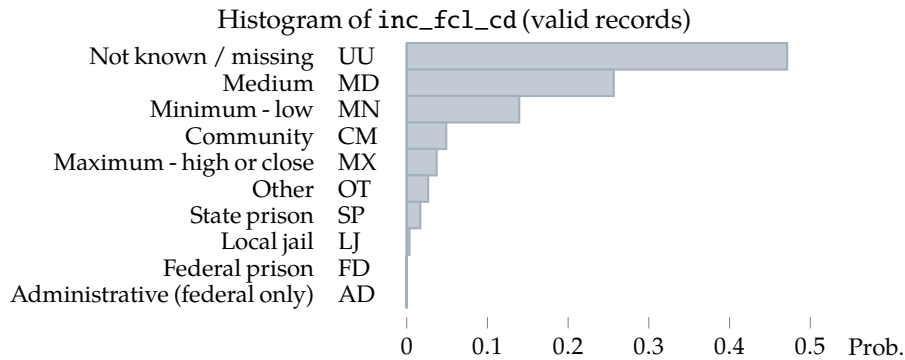
Data notes
- Texas: TX028

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	6,399,324	100.0

Data notes
- All states: US002

inc_entry_dt_yyyy

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

Set	*Count*	*Percent (%)*
All records	6,399,324	100.0
Valid records	6,176,043	96.5
Missing values	223,281	3.5

Statistic	*Value*
Mean	2,003.3
Minimum	1,901
25th percentile	1,996
Median	2,005
75th percentile	2,012
Maximum	2,020

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

Histogram of age on incarceration entry date (1st–99th percentiles)
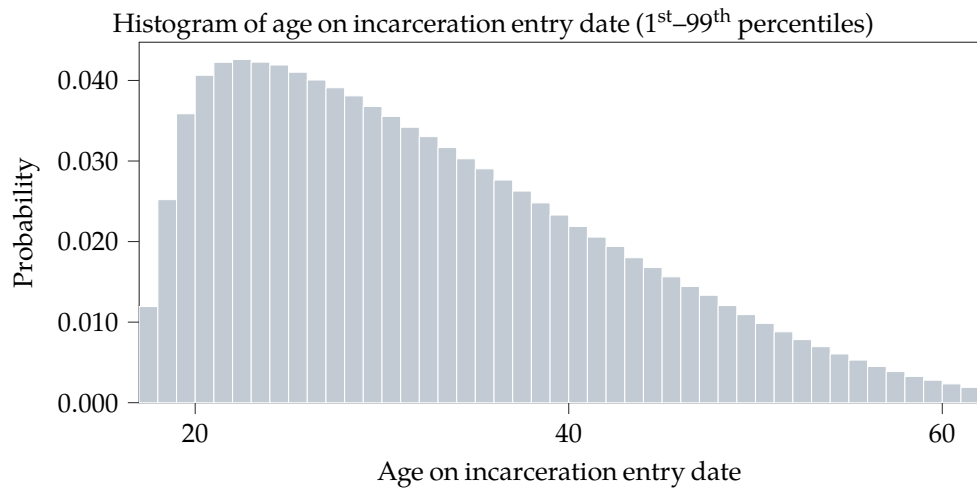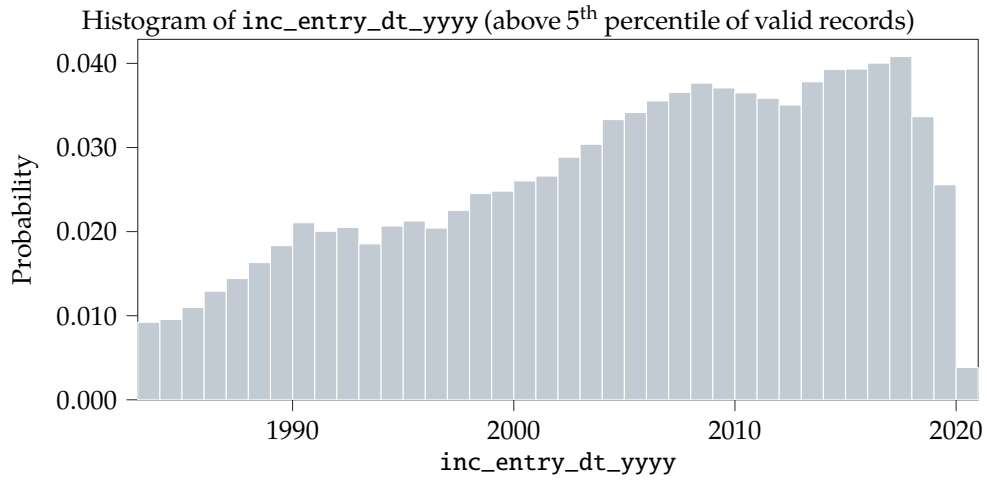
Data notes
- Texas: TX014
- Wisconsin: WI002

inc_entry_dt_mm

Label   Incarceration entry date, month

Description   The month when the individual entered incarceration.

Table   Incarceration
Format   numeric
Code scheme   month code

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	6,174,645	96.5
Missing values	224,679	3.5

Histogram of inc_entry_dt_mm (valid records)
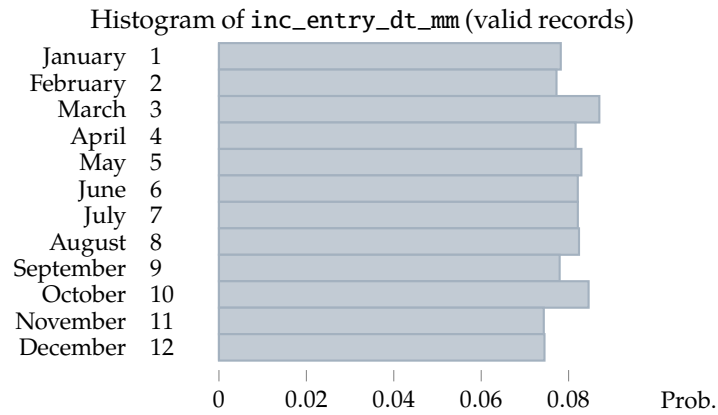
```
inc_entry_dt_dd
```

Label   Incarceration entry date, day of month

Description   The day of the month when the individual entered incarceration.

Table   Incarceration
Format   numeric

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	6,174,645	96.5
Missing values	224,679	3.5

Histogram of `inc_entry_dt_dd` (valid records)
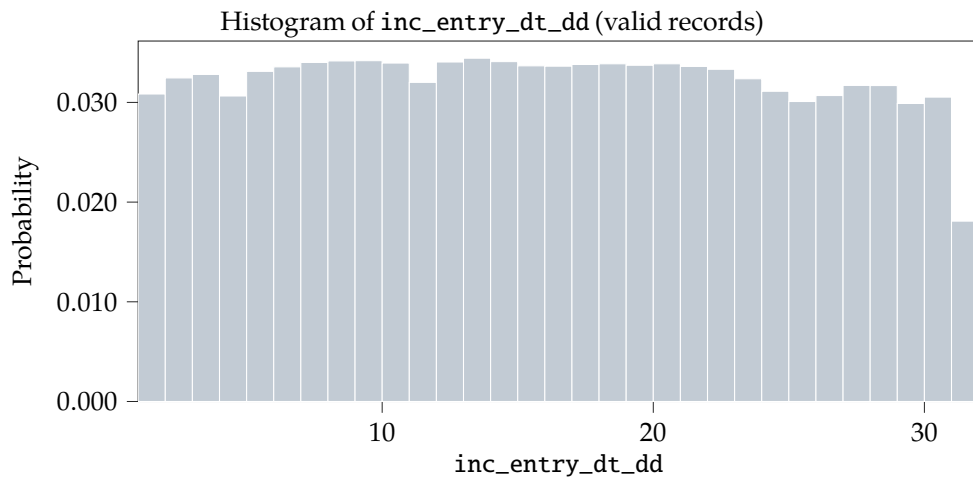
```
inc_entry_cd
```

Label   CJARS standardized entry status

Description   CJARS standardized entry status into incarceration. Used when consistent coding
                            is needed across jurisdictions. For a full description of the coding categories and
                            method used to harmonize entry status, please refer to Appendices B and D.2.7.

Table   Incarceration
Format   string
Code scheme   incarceration entry code

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of `inc_entry_cd` (valid records)
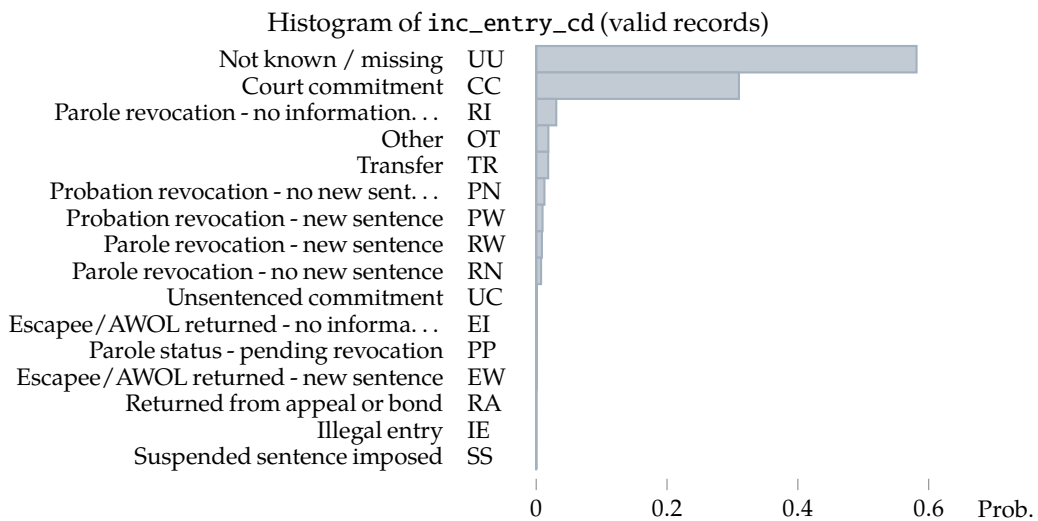
Data notes
- Texas: TX028

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	6,399,324	100.0

Data notes
- All states: US002

inc_exit_dt_yyyy

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

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	5,489,112	85.8
Invalid values	1	0.0
Missing values	910,211	14.2

Statistic	Value
Mean	2,003.7
Minimum	1,901.0
25th percentile	1,997.0
Median	2,006.0
75th percentile	2,012.0
Maximum	2,020.0

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

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

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

inc_exit_dt_yyyy (continued)

Data notes
- Arkansas: AR002
- Mississippi: MS002
- Texas: TX014
- Wisconsin: WI002

inc_exit_dt_mm

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

Set	*Count*	*Percent (%)*
All records	6,399,324	100.0
Valid records	5,489,113	85.8
Missing values	910,211	14.2

Histogram of inc_exit_dt_mm (valid records)

Month	Code
January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12
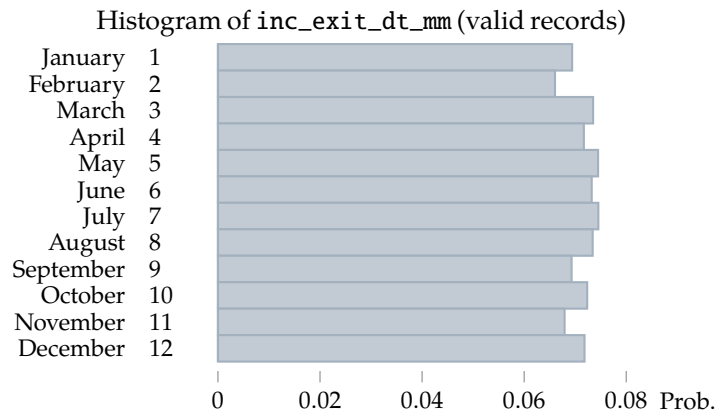
Data notes
- Arkansas: AR002
- Mississippi: MS002

inc_exit_dt_dd

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

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	5,489,113	85.8
Missing values	910,211	14.2

Histogram of inc_exit_dt_dd (valid records)
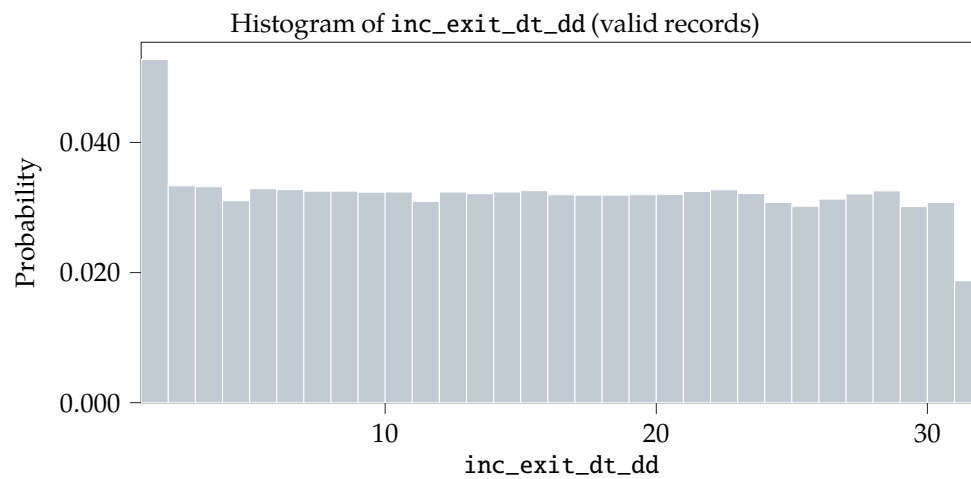
Data notes
- Arkansas: AR002
- Florida: FL002
- Mississippi: MS002

120

```
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 D.2.9.

Table   Incarceration
Format   string
Code scheme   incarceration exit code

Set	Count	Percent (%)
All records	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of `inc_exit_cd` (valid records)

Not Known / Missing	UU	
Expiration of Sentence	ES	
Mandatory Parole Release	MR	
Parole Board Decision	PD	
Other Conditional Release	OR	
Transfer	TR	
Probation Release	PR	
Other	OT	
Release to Custody, Detainer, Warrant	RC	
Escape/AWOL	EA	
Death, Natural Causes	DN	
Other Death	OD	
Other Unconditional Release	UR	
Release on Appeal or Bond	RA	
Execution	EX	
Illegal Entry	IE	
Commutation/Pardon	CP	
Suicide	SU	
Accidental Injury to Self	AI	
Homicide by Another Inmate	HI	

0   0.1   0.2   0.3   0.4   Prob.

Data notes
  ● Texas: TX028

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	6,399,324	100.0

Data notes
- All states: US002

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	6,399,324	100.0
Valid records	6,075,679	94.9
Missing values	323,645	5.1

inc_st_ori_fips (continued)

Histogram of inc_st_ori_fips (valid records)

Arizona	04
Arkansas	05
Connecticut	09
Florida	12
Illinois	17
Michigan	26
Mississippi	28
Nebraska	31
New Jersey	34
North Carolina	37
Ohio	39
Pennsylvania	42
Texas	48
Washington	53
Wisconsin	55

0   0.1   0.2   0.3   Prob.
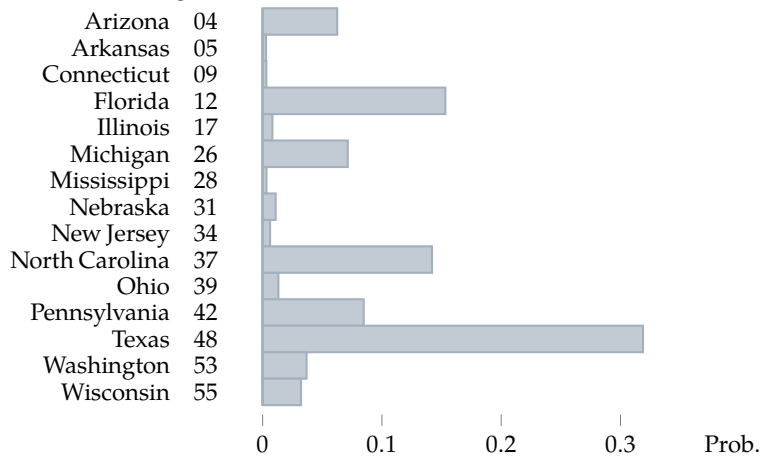
Data notes
- Colorado: CO002

---

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	6,399,324	100.0
Valid records	4,507,094	70.4
Missing values	1,568,585	24.5

Data notes
- Colorado: CO002
- Washington: WA001
- Wisconsin: WI001

```
inc_st_juris_fips
```

Label   State FIPS code of state responsibe for physical supervision of individual

Description   State-level Federal Information Processing
Standards (FIPS) code for the state with responsibility of physical supervision
over individual's term of incarceration. Used to help uniquely 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	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of `inc_st_juris_fips` (valid records)

State	Code
Arizona	04
Arkansas	05
Colorado	08
Connecticut	09
Florida	12
Illinois	17
Michigan	26
Mississippi	28
Nebraska	31
New Jersey	34
North Carolina	37
Ohio	39
Pennsylvania	42
Texas	48
Washington	53
Wisconsin	55

0   0.1   0.2   0.3   Prob.

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	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of `inc_rec_src_le` (valid records)

No   0

0   0.2   0.4   0.6   0.8   1   Prob.

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	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of `inc_rec_src_crt` (valid records)

No   0

0   0.2   0.4   0.6   0.8   1   Prob.
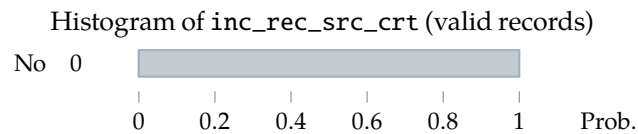
inc_rec_src_doc

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	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of inc_rec_src_doc (valid records)

No	0	
Yes	1	

0   0.2   0.4   0.6   0.8   Prob.
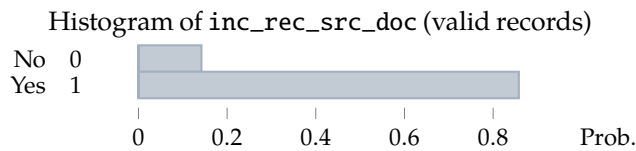
inc_rec_src_rep

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	6,399,324	100.0
Valid records	6,399,324	100.0

Histogram of inc_rec_src_rep (valid records)

No	0	
Yes	1	

0   0.2   0.4   0.6   0.8   Prob.

```
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   6,399,324   100.0
Valid records   6,399,324   100.0

Histogram of `inc_rec_src_cc` (valid records)

No   0

      0   0.2   0.4   0.6   0.8   1   Prob.

5.2.6 Parole

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

Table-level data notes
- All states: US012

---

cjars_id

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

Set	*Count*	*Percent (%)*
All records	2,231,276	100.0
Unique values	1,274,529	

Data notes
- Virginia: VA001

---

par_id

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

```
par_bgn_dt_yyyy
```

Label   Parole start date, year

Description   The year when the individual began parole.

Table   Parole
Format   numeric

Set	Count	Percent (%)
All records	2,231,276	100.0
Valid records	2,101,464	94.2
Invalid values	1	0.0
Missing values	129,811	5.8

Statistic	Value
Mean	2,003.4
Minimum	1,948
25th percentile	1,995
Median	2,005
75th percentile	2,013
Maximum	2,020

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

par_bgn_dt_yyyy (continued)

Histogram of age on parole start date (1st–99th percentiles)
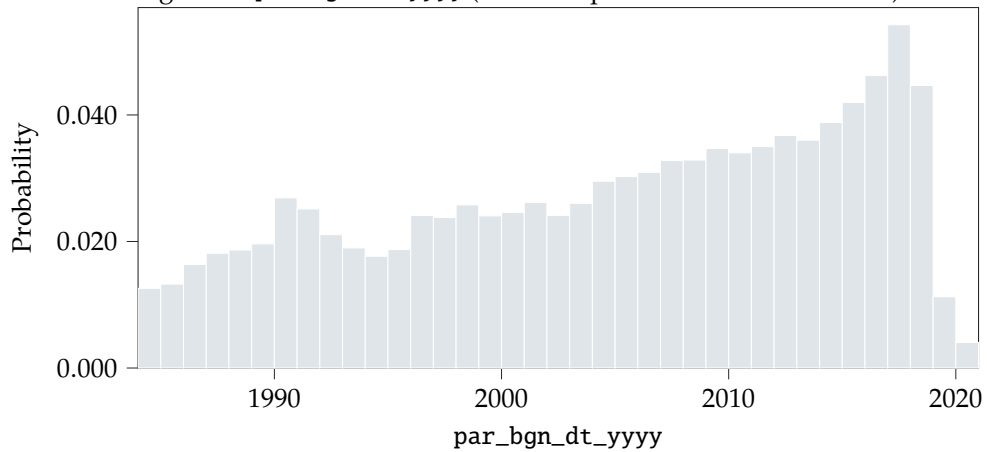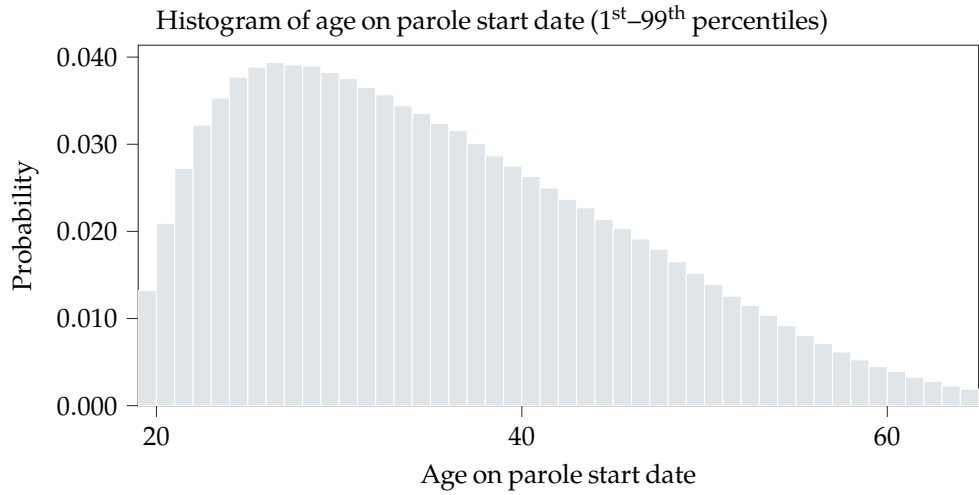
Data notes
- Texas: TX015

---

par_bgn_dt_mm

Label   Parole start date, month

Description   The month when the individual began parole.

Table   Parole
Format   numeric
Code scheme   month code

Set	Count	Percent (%)
All records	2,231,276	100.0
Valid records	2,036,235	91.3
Missing values	195,041	8.7

par_bgn_dt_mm (continued)

Histogram of par_bgn_dt_mm (valid records)

January	1	
February	2	
March	3	
April	4	
May	5	
June	6	
July	7	
August	8	
September	9	
October	10	
November	11	
December	12	

0   0.02   0.04   0.06   0.08   Prob.
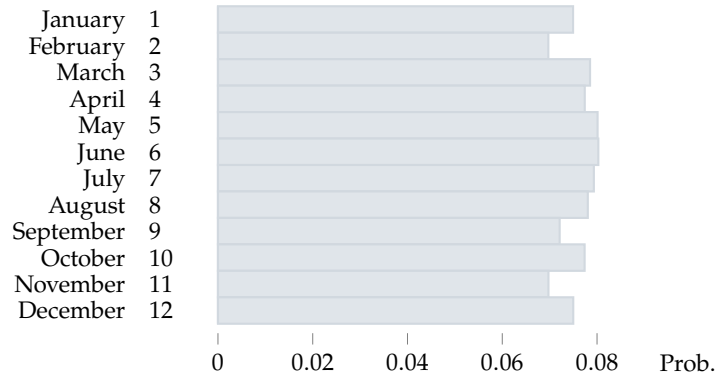
par_bgn_dt_dd

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

Set	*Count*	*Percent (%)*
All records	2,231,276	100.0
Valid records	2,036,235	91.3
Missing values	195,041	8.7

Histogram of par_bgn_dt_dd (valid records)
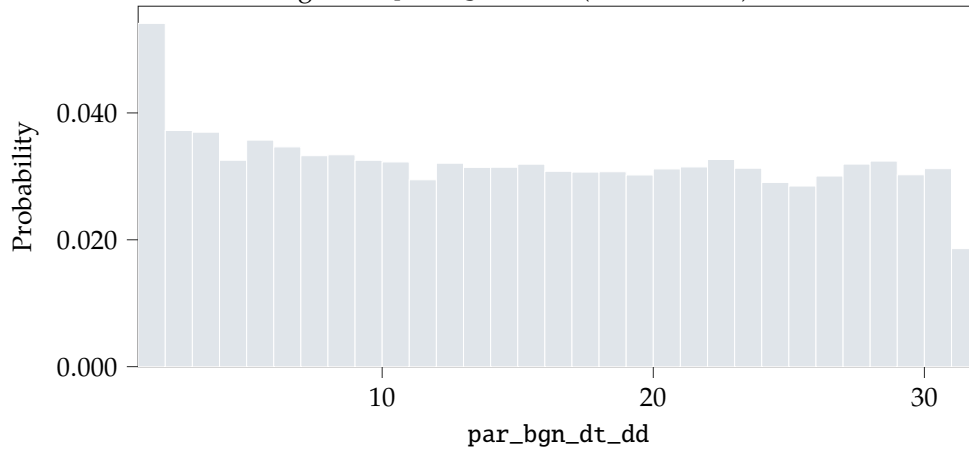
Data notes
- Nebraska: NE003

par_end_dt_yyyy

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

Set	Count	Percent (%)
All records	2,231,276	100.0
Valid records	1,197,408	53.7
Invalid values	1	0.0
Missing values	1,033,867	46.3

Statistic	Value
Mean	2,003.1
Minimum	1,900.0
25th percentile	1,995.0
Median	2,004.0
75th percentile	2,012.0
Maximum	2,020.0

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

par_end_dt_yyyy (continued)

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

Histogram of months since `par_bgn_dt` (5th–95th percentiles)
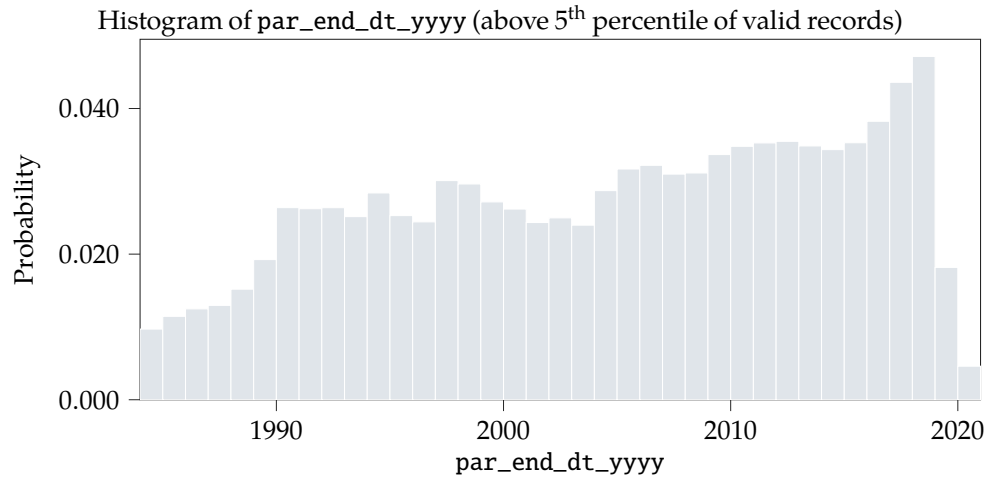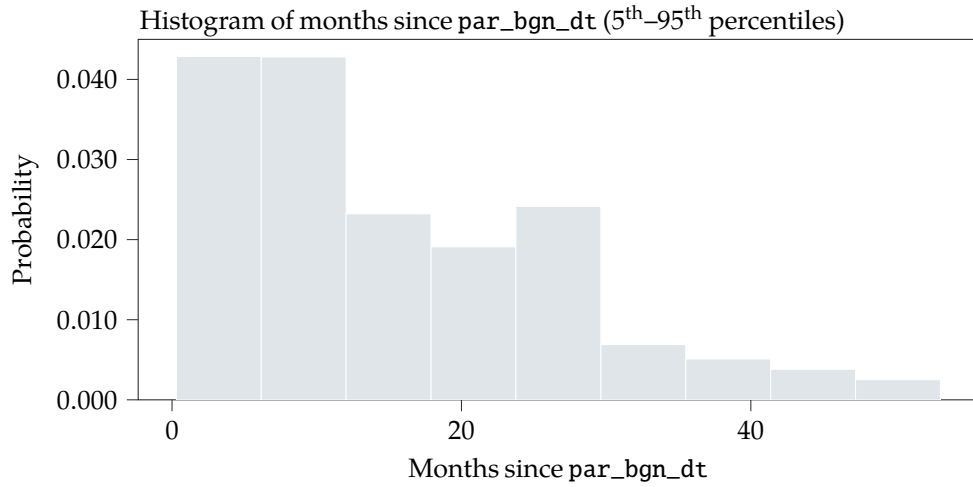
Data notes
- Florida: FL003
- Illinois: IL002
- Texas: TX015, TX016

par_end_dt_mm

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

Set	Count	Percent (%)
All records	2,231,276	100.0
Valid records	1,197,409	53.7
Missing values	1,033,867	46.3

Histogram of par_end_dt_mm (valid records)

January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12

0   0.01   0.02   0.03   0.04   0.05   Prob.

Data notes
- Florida: FL003
- Illinois: IL002
- Texas: TX016

134

```
par_end_dt_dd
```

Label   Parole end date, day of month

Description   The day of the month when the individual's parole ended.

Table   Parole
Format   numeric

Set	Count	Percent (%)
All records	2,231,276	100.0
Valid records	1,197,409	53.7
Missing values	1,033,867	46.3

Histogram of `par_end_dt_dd` (valid records)
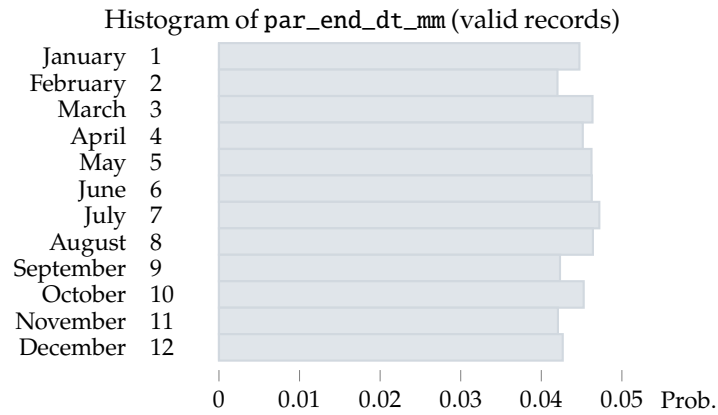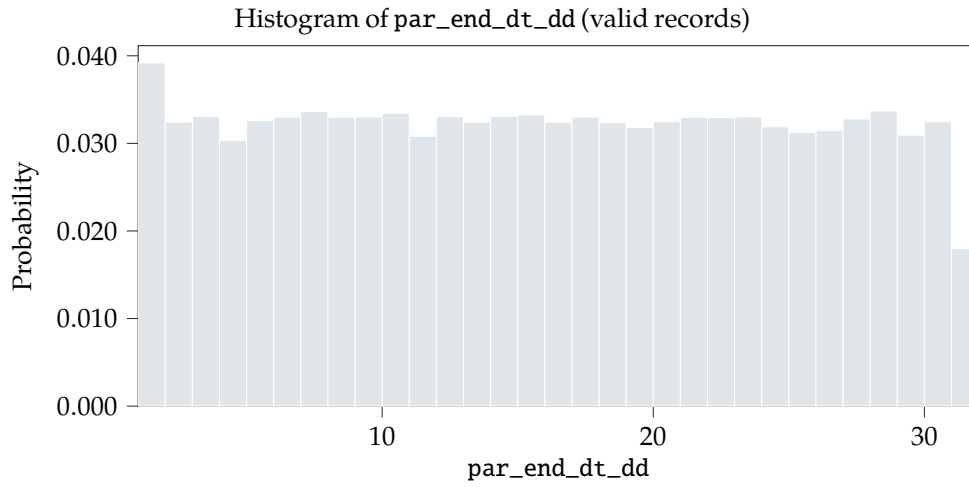
Data notes
- Florida: FL003
- Illinois: IL002
- Texas: TX016

```
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 D.2.10.

Table   Parole
Format   string
Code scheme   parole exit code

Set	*Count*	*Percent (%)*
All records	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of `par_end_cd` (valid records)

Not known / missing	UU
Completion	CO
Returned to incarceration - revoca...	RV
Absconded/escaped	AB
Returned to incarceration - other/...	RO
Death	DE
Transferred to another state	TR
Other	OT
Other unsatisfactory exit	OU

0   0.2   0.4   0.6   0.8   Prob.
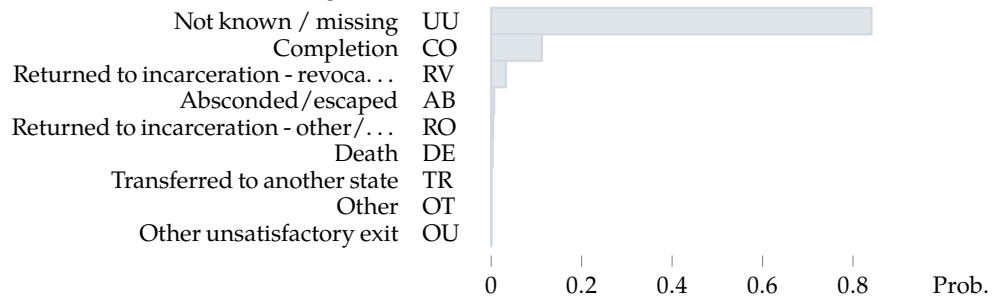
par_end_cd_src

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

Set	*Count*	*Percent (%)*
All records	2,231,276	100.0

Data notes
- All states: US002

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	2,231,276	100.0
Valid records	2,004,825	89.9
Missing values	226,451	10.1

par_st_ori_fips (continued)

Histogram of par_st_ori_fips (valid records)

		Prob.
Florida	12	
Illinois	17	
Michigan	26	
Nebraska	31	
North Carolina	37	
Ohio	39	
Texas	48	

0   0.2   0.4   0.6 Prob.
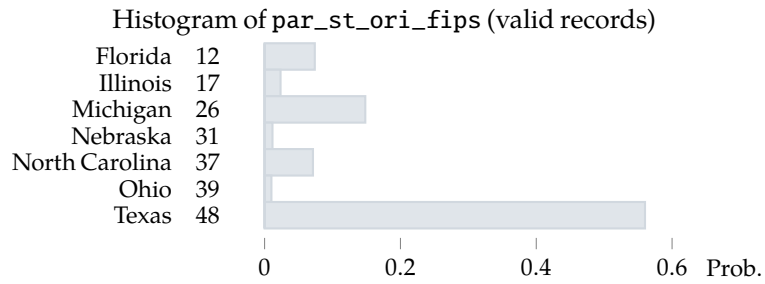
par_cnty_ori_fips

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	2,231,276	100.0
Valid records	911,228	40.8
Missing values	1,093,597	49.0

Data notes
- Colorado: CO002

138

```
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	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of `par_st_juris_fips` (valid 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	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of `par_rec_src_le` (valid records)

No   0

0   0.2   0.4   0.6   0.8   1   Prob.

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	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of `par_rec_src_crt` (valid records)

No   0

0   0.2   0.4   0.6   0.8   1   Prob.
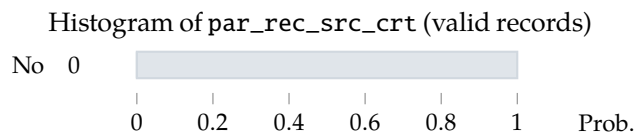
par_rec_src_doc

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	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of par_rec_src_doc (valid records)

No   0
Yes 1

0   0.2   0.4   0.6   0.8   1   Prob.
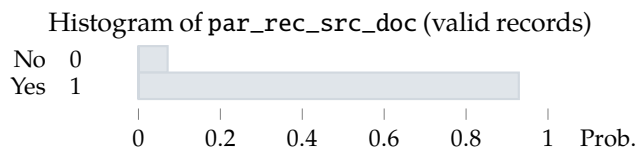
par_rec_src_rep

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	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of par_rec_src_rep (valid records)

No   0
Yes 1

0   0.2   0.4   0.6   0.8   1   Prob.

```
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	2,231,276	100.0
Valid records	2,231,276	100.0

Histogram of `par_rec_src_cc` (valid records)

No   0

      0   0.2   0.4   0.6   0.8   1   Prob.

5.3   ID variables used to link tables

The tables in the section above show that there are several ID variables contained in each of the CJARS data tables. The variables are used to identify individuals, events, episodes, and the linkages among these three. Refer to Sections 3.3.1 and 3.3.2 for more information about how these linkages are structured. Table 3 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.

> **USER NOTE**
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense. This means that the key identifying event variables have not yet been added to the CJARS relational tables. We expect to complete this work in 2021. Users can expect to see linked processes in the RDC data warehouse in late 2021.

Table 3: Variables to link tables

CJARS relational table	Key identifying event variable	Preceding event	Linking variable for preceding event	Description of linkage
roster	`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. At the Census Bureau, the PIK is added to the roster to enable linkage to other Census Bureau survey and administrative records.
arrest	`arr_id`	n/a	n/a	n/a
adjudication	`adj_id`	arrest	`arr_id`	The `adj_id` 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 `arr_id`.
incarceration	`inc_id`	adjudication	`adj_id`	The `inc_id` 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 `adj_id`.
probation	`pro_id`	adjudication	`adj_id`	The `pro_id` 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 `adj_id`.
parole	`par_id`	incarceration	`inc_id`	The `par_id` 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 `inc_id`.

Appendices

A   Data sources

Table 4: Agencies whose data have been integrated into CJARS

State	Level	Domain	County	Municipality	Data provider/agency
Arizona	State	Corrections			Arizona Department of Corrections
		Judiciary			Arizona Administrative Office of the Courts
Arkansas	State	Corrections			Arkansas Department of Corrections
California	County	Sheriff	Butte		Butte County Sheriff's Office
			Contra Costa		Contra Costa County Sheriff's Office
			Fresno		Fresno County Sheriff's Office
			Nevada		Nevada County Sheriff's Office
			San Luis Obispo		San Luis Obispo County Sheriff's Office
	Municipal	Police	Kern	Bakersfield	Bakersfield Police Department
			Los Angeles	Long Beach	Long Beach Police Department
				Los Angeles	Los Angeles Police Department
			Orange	Anaheim	Anaheim Police Department
			Riverside	Riverside	Riverside Police Department
			San Bernardino	San Bernardino	San Bernardino Police Department
			San Diego	San Diego	San Diego Police Department
			San Joaquin	Stockton	Stockton Police Department
Colorado	State	Corrections			Colorado Department of Corrections
	County	Sheriff	Adams		Adams County Sheriff's Office
			Boulder		Boulder County Sheriff's Office
			Weld		Weld County Sheriff's Office
Connecticut	State	Corrections			Connecticut Department of Correction
Florida	State	Corrections			Florida Department of Corrections
		Judiciary			Florida Court Clerks and Comptrollers
	County	Judiciary	Alachua		Alachua County Clerk of Courts
			Baker		Baker County Clerk of Courts
			Bay		Bay County Clerk of Courts
			Bradford		Bradford County Clerk of Courts
			Brevard		Brevard County Clerk of Courts
			Broward		Broward County Clerk of Courts
			Calhoun		Calhoun County Clerk of Courts
			Charlotte		Charlotte County Clerk of Courts
			Citrus		Citrus County Clerk of Courts
			Clay		Clay County Clerk of Courts

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Collier		Collier County Clerk of Courts
			Columbia		Columbia County Clerk of Courts
			DeSoto		DeSoto County Clerk of Courts
			Duval		Duval County Clerk of Courts
			Franklin		Franklin County Clerk of Courts
			Gadsden		Gadsden County Clerk of Courts
			Gilchrist		Gilchrist County Clerk of Courts
			Gulf		Gulf County Clerk of Courts
			Hamilton		Hamilton County Clerk of Courts
			Hardee		Hardee County Clerk of Courts
			Hendry		Hendry County Clerk of Courts
			Hernando		Hernando County Clerk of Courts
			Highlands		Highlands County Clerk of Courts
			Hillsborough		Hillsborough County Clerk of Courts
			Holmes		Holmes County Clerk of Courts
			Indian River		Indian River County Clerk of Courts
			Jackson		Jackson County Clerk of Courts
			Jefferson		Jefferson County Clerk of Courts
			Lafayette		Lafayette County Clerk of Courts
			Lee		Lee County Clerk of Courts
			Leon		Leon County Clerk of Courts
			Levy		Levy County Clerk of Courts
			Liberty		Liberty County Clerk of Courts
			Madison		Madison County Clerk of Courts
			Manatee		Manatee County Clerk of Courts
			Martin		Martin County Clerk of Courts
			Miami Dade		Miami-Dade County Clerk of Courts
			Monroe		Monroe County Clerk of Courts
			Nassau		Nassau County Clerk of Courts
			Okeechobee		Okeechobee County Clerk of Courts
			Orange		Orange County Clerk of Courts
			Osceola		Osceola County Clerk of Courts
			Palm Beach		Palm Beach County Clerk of Courts
			Pasco		Pasco County Clerk of Courts
			Pinellas		Pinellas County Clerk of Courts
			Polk		Polk County Clerk of Courts

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Putnam		Putnam County Clerk of Courts
			Santa Rosa		Santa Rosa County Clerk of Courts
			Seminole		Seminole County Clerk of Courts
			St Johns		St. Johns County Clerk of Courts
			St Lucie		St. Lucie County Clerk of Courts
			Sumter		Sumter County Clerk of Courts
			Suwannee		Suwannee County Clerk of Courts
			Taylor		Taylor County Clerk of Courts
			Union		Union County Clerk of Courts
			Volusia		Volusia County Clerk of Courts
			Wakulla		Wakulla County Clerk of Courts
			Walton		Walton County Clerk of Courts
			Washington		Washington County Clerk of Courts
		Sheriff	Pinellas		Pinellas County Sheriff's Office
Illinois	State	Corrections			Illinois Department of Corrections
Iowa	State	Corrections			Iowa Department of Corrections
Kansas	County	Sheriff	Johnson		Johnson County Sheriff's Office
Maryland	State	Judiciary			Maryland Judiciary
Michigan	State	Corrections			Michigan Department of Corrections
		Judiciary			Michigan State Court Administrative Office
	County	Corrections	Oakland		Oakland County Probation Department
Minnesota	State	Judiciary			Minnesota State Court Administrator's Office
Mississippi	State	Corrections			Mississippi Department of Corrections
Nebraska	State	Corrections			Nebraska Department of Correctional Services
New Jersey	State	Corrections			New Jersey Department of Corrections
		Judiciary			Superior Court of New Jersey
North Carolina	State	Judiciary			North Carolina Administrative Office of the Courts
		Repository			North Carolina Department of Public Safety
North Dakota	State	Judiciary			North Dakota Court System
Ohio	State	Corrections			Ohio Department of Rehabilitation and Correction
Oregon	State	Judiciary			Oregon Judicial Department
Pennsylvania	State	Corrections			Pennsylvania Department of Corrections
		Judiciary			Administrative Office of Pennsylvania Courts
Texas	State	Corrections			Texas Department of Criminal Justice
		Judiciary			iDocket
		Repository			Texas Department of Public Safety

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
	County	Judiciary	Angelina		Angelina County Clerk's Office
					Angelina County District Clerk's Office
			Aransas		Aransas County Clerk's Office
					Aransas County District Clerk's Office
			Bandera		Bandera County Clerk's Office
					Bandera County District Clerk's Office
			Bee		Bee County District Clerk's Office
			Bell		Bell County Clerk's Office
					Bell County District Clerk's Office
			Bexar		Bexar County Clerk
					Bexar County District Clerk
					Bexar County District Clerk's Office
			Bowie		Bowie County Clerk's Office
					Bowie County District Clerk's Office
			Brazoria		Brazoria County District Clerk's Office
			Briscoe		Briscoe County Clerk's Office
					Briscoe County District Clerk's Office
			Brown		Brown County Clerk's Office
			Caldwell		Caldwell County Clerk's Office
			Callahan		Callahan County District Clerk's Office
			Cameron		Cameron County Clerk's Office
					Cameron County District Clerk's Office
			Cherokee		Cherokee County District Clerk's Office
			Clay		Clay County Clerk's Office
			Coleman		Coleman County Clerk's Office
			Collin		Collin County Clerk's Office
					Collin County Courts
					Collin County District Clerk's Office
			Colorado		Colorado County District Clerk's Office
			Coryell		Coryell County Clerk's Office
			Crane		Crane County Clerk's Office
					Crane County District Clerk's Office
			Dallam		Dallam County Clerk's Office
					Dallam County District Clerk's Office
			Dallas		Dallas County Clerk's Office
			Dickens		Dickens County Clerk's Office

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
					Dickens County District Clerk's Office
			Duval		Duval County District Clerk's Office
			Eastland		Eastland County Clerk's Office
			El Paso		El Paso County Clerk
					El Paso County Clerk's Office
					El Paso District Clerk
			Floyd		Floyd County Clerk's Office
					Floyd County District Clerk's Office
			Franklin		Franklin County Clerk's Office
					Franklin County District Clerk's Office
			Galveston		Galveston County Clerk's Office
					Galveston County District Clerk's Office
			Goliad		Goliad County Clerk's Office
					Goliad County District Clerk's Office
			Harris		Harris County Clerk's Office
					Harris County District Clerk
			Hays		Hays County Clerk's Office
					Hays County Courts at Law
					Hays County District Clerk's Office
			Hidalgo		Hidalgo County Clerk's Office
					Hidalgo County District Clerk's Office
			Hopkins		Hopkins County Clerk's Office
					Hopkins County District Clerk's Office
			Houston		Houston County Clerk's Office
					Houston County District Clerk's Office
			Hudspeth		Hudspeth County Clerk's Office
					Hudspeth County District Clerk's Office
			Hunt		Hunt County Clerk's Office
			Jack		Jack County Clerk's Office
					Jack County District Clerk's Office
			Jefferson		Jefferson County District Clerk's Office
			Jim Wells		Jim Wells County District Clerk's Office
			Kinney		Kinney County Clerk's Office
					Kinney County District Clerk's Office
			Liberty		Liberty County Clerk's Office
					Liberty County District Clerk's Office

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Limestone		Limestone County Clerk's Office
					Limestone County District Clerk's Office
			McLennan		McLennan County District Clerk's Office
			Montgomery		Montgomery County Clerk's Office
			Motley		Motley County Clerk's Office
					Motley County District Clerk's Office
			Nacogdoches		Nacogdoches County Clerk's Office
					Nacogdoches County District Clerk's Office
			Ochiltree		Ochiltree County Clerk's Office
					Ochiltree County District Clerk's Office
			Oldham		Oldham County Clerk's Office
					Oldham County District Clerk's Office
			Orange		Orange County Clerk's Office
			Polk		Polk County Clerk's Office
					Polk County District Clerk's Office
			Presidio		Presidio County Clerk's Office
					Presidio County District Clerk's Office
			San Patricio		San Patricio County Clerk's Office
					San Patricio County District Clerk's Office
			Schleicher		Schleicher County Clerk's Office
					Schleicher County District Clerk's Office
			Shelby		Shelby County Clerk's Office
			Starr		Starr County Clerk's Office
					Starr County District Clerk's Office
			Titus		Titus County District Clerk's Office
			Travis		Travis County District Clerk's Office
			Trinity		Trinity County Clerk's Office
			Tyler		Tyler County Clerk's Office
			Upton		Upton County Clerk's Office
					Upton County District Clerk's Office
			Val Verde		Val Verde County District Clerk's Office
			Victoria		Victoria County Clerk's Office
					Victoria County District Clerk's Office
			Washington		Washington County Clerk's Office
					Washington County District Clerk's Office
			Webb		Webb County Clerk's Office

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
					Webb County District Clerk's Office
			Wharton		Wharton County Clerk's Office
					Wharton County District Clerk's Office
			Willacy		Willacy County Clerk's Office
					Willacy County District Clerk's Office
			Young		Young County Clerk's Office
					Young County District Clerk's Office
			Zapata		Zapata County Clerk's Office
		Sheriff	Bexar		Bexar County Sheriff's Office
			Collin		Collin County Sheriff's Office
			Dallas		Dallas County Sheriff's Office
			Harris		Harris County Sheriff's Office
			Hays		Hays County Sheriff's Office
			Tarrant		Tarrant County Sheriff's Office
	Municipal	Police	Tarrant	Fort Worth	Fort Worth Police Department
Virginia	State	Judiciary			Judiciary of Virginia
	County	Judiciary	Accomack		Accomack Circuit Court
			Albemarle		Albemarle Circuit Court
			Alexandria		Alexandria Circuit Court
			Alleghany		Alleghany Circuit Court
			Amelia		Amelia Circuit Court
			Amherst		Amherst Circuit Court
			Appomattox		Appomattox Circuit Court
			Arlington		Arlington Circuit Court
			Augusta		Augusta Circuit Court
			Bath		Bath Circuit Court
			Bedford		Bedford Circuit Court
			Bland		Bland Circuit Court
			Botetourt		Botetourt Circuit Court
			Brunswick		Brunswick Circuit Court
			Buchanan		Buchanan Circuit Court
			Buckingham		Buckingham Circuit Court
			Buena Vista		Buena Vista Circuit Court
			Campbell		Campbell Circuit Court
			Caroline		Caroline Circuit Court
			Carroll		Carroll Circuit Court

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Charles City		Charles City Circuit Court
			Charlotte		Charlotte Circuit Court
			Charlottesville		Charlottesville Circuit Court
			Chesapeake		Chesapeake Circuit Court
			Chesterfield		Chesterfield Circuit Court
			Clarke		Clarke Circuit Court
			Colonial Heights		Colonial Heights Circuit Court
			Craig		Craig Circuit Court
			Culpeper		Culpeper Circuit Court
			Cumberland		Cumberland Circuit Court
			Danville		Danville Circuit Court
			Dickenson		Dickenson Circuit Court
			Dinwiddie		Dinwiddie Circuit Court
			Essex		Essex Circuit Court
			Fairfax		Fairfax Circuit Court
			Fauquier		Fauquier Circuit Court
			Floyd		Floyd Circuit Court
			Fluvanna		Fluvanna Circuit Court
			Franklin		Franklin Circuit Court
			Frederick		Frederick Circuit Court
			Fredericksburg		Fredericksburg Circuit Court
			Giles		Giles Circuit Court
			Gloucester		Gloucester Circuit Court
			Goochland		Goochland Circuit Court
			Grayson		Grayson Circuit Court
			Greene		Greene Circuit Court
			Greensville		Greensville Circuit Court
			Halifax		Halifax Circuit Court
			Hampton		Hampton Circuit Court
			Hanover		Hanover Circuit Court
			Henrico		Henrico Circuit Court
			Henry		Henry Circuit Court
			Highland		Highland Circuit Court
			Hopewell		Hopewell Circuit Court
			Isle of Wight		Isle of Wight Circuit Court
			King George		King George Circuit Court

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			King William		King William Circuit Court
			King and Queen		King and Queen Circuit Court
			Lancaster		Lancaster Circuit Court
			Lee		Lee Circuit Court
			Loudoun		Loudoun Circuit Court
			Louisa		Louisa Circuit Court
			Lunenburg		Lunenburg Circuit Court
			Lynchburg		Lynchburg Circuit Court
			Madison		Madison Circuit Court
			Martinsville		Martinsville Circuit Court
			Mathews		Mathews Circuit Court
			Mecklenburg		Mecklenburg Circuit Court
			Middlesex		Middlesex Circuit Court
			Montgomery		Montgomery Circuit Court
			Nelson		Nelson Circuit Court
			New Kent		New Kent Circuit Court
			Newport News		Newport News Circuit Court
			Norfolk		Norfolk Circuit Court
			Northampton		Northampton Circuit Court
			Northumberland		Northumberland Circuit Court
			Nottoway		Nottoway Circuit Court
			Orange		Orange Circuit Court
			Page		Page Circuit Court
			Patrick		Patrick Circuit Court
			Petersburg		Petersburg Circuit Court
			Pittsylvania		Pittsylvania Circuit Court
			Portsmouth		Portsmouth Circuit Court
			Powhatan		Powhatan Circuit Court
			Prince Edward		Prince Edward Circuit Court
			Prince George		Prince George Circuit Court
			Prince William		Prince William Circuit Court
			Pulaski		Pulaski Circuit Court
			Radford		Radford Circuit Court
			Rappahannock		Rappahannock Circuit Court
			Richmond		Richmond County Circuit Court
			Richmond city		Richmond City Circuit Court

Table 4: Agencies whose data have been integrated into CJARS (cont'd)

State	Level	Domain	County	Municipality	Data provider/agency
			Roanoke		Roanoke County Circuit Court
			Roanoke city		Roanoke City Circuit Court
			Rockbridge		Rockbridge Circuit Court
			Rockingham		Rockingham Circuit Court
			Russell		Russell Circuit Court
			Salem city		Salem Circuit Court
			Scott		Scott Circuit Court
			Shenandoah		Shenandoah Circuit Court
			Smyth		Smyth Circuit Court
			Southampton		Southampton Circuit Court
			Spotsylvania		Spotsylvania Circuit Court
			Stafford		Stafford Circuit Court
			Staunto		Staunton Circuit Court
			Suffolk		Suffolk Circuit Court
			Surry		Surry Circuit Court
			Sussex		Sussex Circuit Court
			Tazewell		Tazewell Circuit Court
			Virginia Beach		Virginia Beach Circuit Court
			Warren		Warren Circuit Court
			Washington		Washington Circuit Court
			Waynesboro		Waynesboro Circuit Court
			Westmoreland		Westmoreland Circuit Court
			Williamsburg		Williamsburg/James City County Circuit Court
			Winchester		Winchester Circuit Court
			Wise		Wise Circuit Court
			Wythe		Wythe Circuit Court
			York		York County Poquoson Circuit Court
Washington	State	Corrections			Washington State Department of Corrections
Wisconsin	State	Corrections			Wisconsin Department of Corrections
		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 9 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 are 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 E) 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 E), 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 D). Variable harmonization is a complex task that employs many techniques including hand coding, employing the use of regular expression commands, and machine learning techniques (for an example of its application in offense classification, see Appendix C).

More details about the harmonization of other variables in each of the five CJARS relational databases are included below in Table 5.
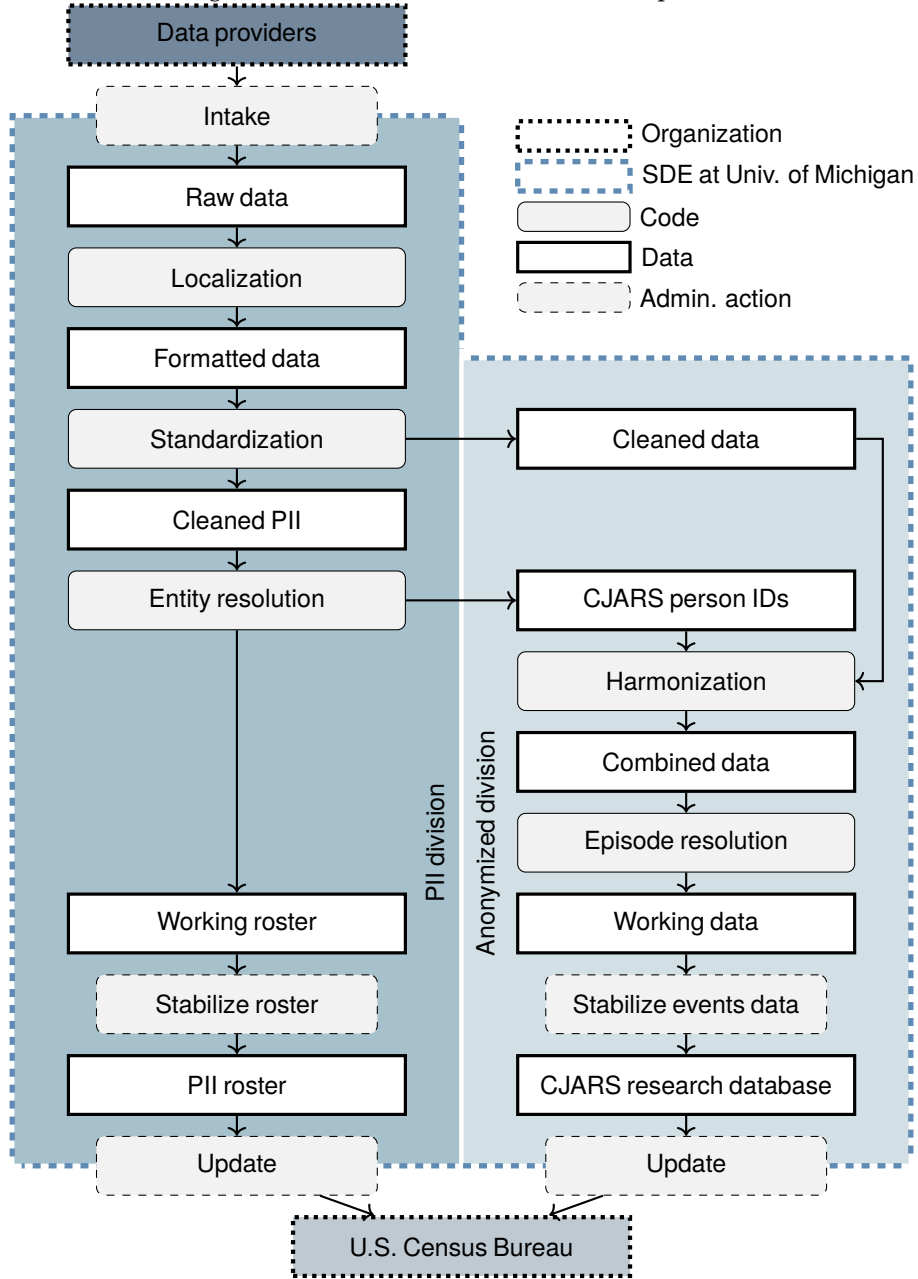
Table 5: Variable harmonization

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).

155

CJARS relational database	Information	Variable(s)	Coding method(s)
adjudication	offense grade	adj_grd_cd	hand-coded and use of regular expressions
adjudication	legal code	adj_off_lgl_cd	hand-coded and use of regular expressions
adjudication	case filing date	adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd	split into year, month, and day
adjudication	charge offense 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
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

CJARS relational database	Information	Variable(s)	Coding method(s)
probation	end date	pro_end_cd	split into year,month, and day
probation	end status	pro_end_dt_yyyy, pro_end_dt_mm, pro_end_dt_dd	hand-coded and use of regular expressions
parole	begin date	par_bgn_dt_yyyy, par_bgn_dt_mm, par_bgn_dt_dd	split into year,month, and day
parole	end date	par_end_dt_yyyy, par_end_dt_mm, par_end_dt_dd	split into year,month, and day
parole	end status	par_end_cd	hand-coded and use of regular expressions

Figure 9: CJARS variable harmonization process

C Offense classification

CJARS uses an offense classification scheme developed by Measures for Justice (MFJ). MFJ grounds its classification scheme on the codes developed for the National Corrections Reporting Program (NCRP) in the early 1990s. Those codes offered a disaggregated means to categorize offenses from state statutes. Charge descriptions vary widely across states based on statutory organization and language. MFJ's modifications include reordering offenses to reflect seriousness, adding clarifications to the NCRP codes to ensure consistency, reclassifying DUI to its own offense type, reclassifying many of the "other" and "public order" offenses, and adding new codes for offenses including human trafficking, amphetamine drug offenses, opiate drug offenses, and other prescription drug offenses. The multi-digit scheme allows charges to be grouped easily into violent, property, drug, driving under the influence (DUI), public order, other, and unknown offense categories. The MFJ classification scheme can be found in Appendix 10.

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

Figure 10 outlines how hierarchical offense classification occurs. First, raw string descriptions are broken up into n-grams. N-grams are rolling windows of a string. The offense string "murder" can be split into the 3-gram set {"mur", "urd", "rde", "der"}. Next, the n-grams are used to classify descriptions into a parent class (e.g., violent, property). Then, offenses classified at the parent class level are further classified into a sub-parent class (e.g., murder, manslaughter). Finally, offenses are classified into a child class (e.g., attempt, conspiracy). This hierarchical approach was implemented because it substantially improved the predictive validity of the offense classification algorithm. The algorithms are tested against a test set of half a million unique offense descriptions hand validated by MFJ.
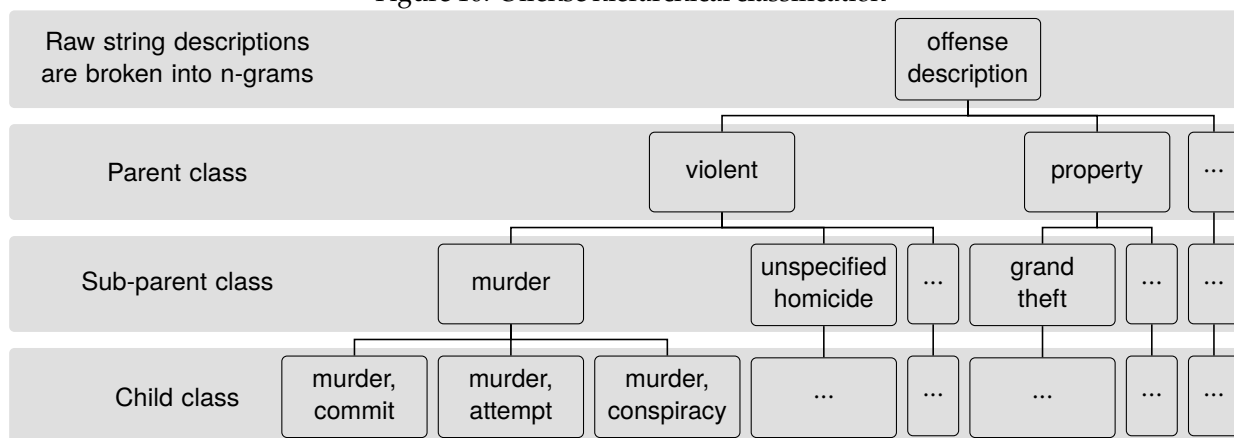
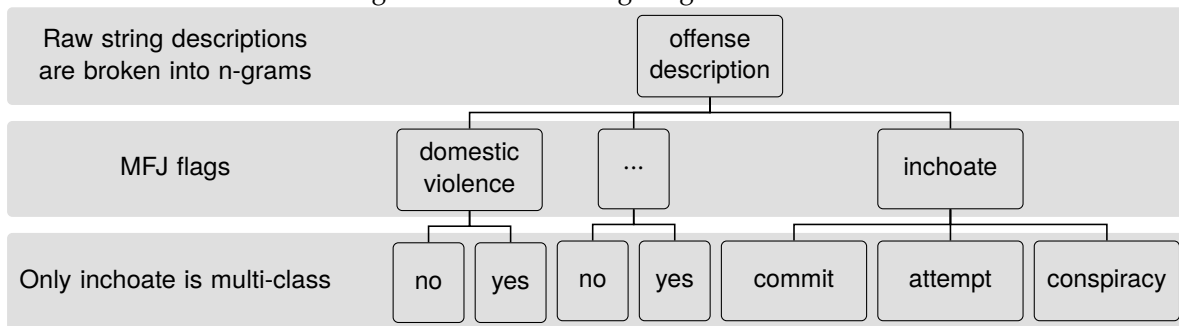Figure 10: Offense hierarchical classification

Figure 11 outlines how charge flags are assigned to offenses. Offenses can be assigned multiple classes. Again, we start with n-grams of the offense string. The n-grams are used to classify descriptions into MFJ flags (e.g., domestic violence, gang-related).

Figure 11: Offense charge flag classification

Raw string descriptions are broken into n-grams	offense description				
MFJ flags	domestic violence	...		inchoate	
Only inchoate is multi-class	no	yes	no	yes	commit	attempt	conspiracy

D Code schemes including offense classifications

D.1 Geographic and demographic codes

D.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 6: 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.

D.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.

D.1.3 Sex codes

This scheme is used for the following variables:

- `sex`
- `sex_raw`

Table 7: Sex codes

Value	Label
1	male
2	female

D.1.4 Race and ethnicity codes

This scheme is used for the following variables:

- race
- race_raw

Table 8: Race and ethnicity codes

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

D.1.5 Demographic imputation codes

This scheme is used for the following variables:

- sex_imputed
- race_imputed

Table 9: Demographic imputation codes

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

D.2 Criminal justice event codes

D.2.1 Offense classification

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

This scheme is used for the following variables:

- `adj_chrg_off_cd`

- `adj_disp_off_cd`

- `arr_off_cd`

Table 10: Offense classification scheme

Offense type	Offense code	Offense code description
Violent	1010	Murder
Violent	1011	Attempted murder
Violent	1012	Conspiracy to commit murder
Violent	1020	Unspecified homicide
Violent	1021	Unspecified homicide, attempted
Violent	1022	Unspecified homicide, conspiracy
Violent	1030	Voluntary manslaughter
Violent	1031	Voluntary manslaughter, attempted
Violent	1032	Voluntary manslaughter,conspiracy
Violent	1040	Vehicular manslaughter
Violent	1041	Vehicular manslaughter, attempted
Violent	1042	Vehicular manslaughter, conspiracy
Violent	1050	Involuntary manslaughter
Violent	1051	Involuntary manslaughter, attempt
Violent	1052	Involuntary manslaughter, conspiracy
Violent	1060	Kidnapping
Violent	1061	Kidnapping, attempted
Violent	1062	Kidnapping, conspiracy
Violent	1070	Rape
Violent	1071	Rape, attempted
Violent	1072	Rape, conspiracy
Violent	1080	Statutory rape
Violent	1081	Statutory rape, attempted
Violent	1082	Statutory rape, conspiracy
Violent	1090	Child molestation
Violent	1091	Child molestation, attempted
Violent	1092	Child molestation, conspiracy
Violent	1100	Sexual assault
Violent	1101	Sexual assault, attempted
Violent	1102	Sexual assault, conspiracy
Violent	1110	Human trafficking, sex - child
Violent	1111	Human trafficking, sex - child, attempted
Violent	1112	Human trafficking, sex - child, conspiracy
Violent	1120	Human trafficking, sex - adult or no age specified
Violent	1121	Human trafficking, sex - adult or no age specified, attempted
Violent	1122	Human trafficking, sex - adult or no age specified, conspiracy
Violent	1130	Human trafficking, labor - child
Violent	1131	Human trafficking, labor - child, attempted
Violent	1132	Human trafficking, labor - child, conspiracy
Violent	1140	Human trafficking, labor - adult or no age specified
Violent	1141	Human trafficking, labor - adult or no age specified, attempted
Violent	1142	Human trafficking, labor - adult or no age specified, conspiracy

Offense type	Offense code	Offense code description
Violent	1150	Human trafficking, unspecified - child
Violent	1151	Human trafficking, unspecified - child, attempted
Violent	1152	Human trafficking, unspecified - child, conspiracy
Violent	1160	Human trafficking, unspecified - adult or no age specified
Violent	1161	Human trafficking, unspecified - adult or no age specified, attempted
Violent	1162	Human trafficking, unspecified - adult or no age specified, conspiracy
Violent	1170	Human trafficking
Violent	1171	Human trafficking, attempted
Violent	1172	Human trafficking, conspiracy
Violent	1180	Armed robbery
Violent	1181	Armed robbery, attempted
Violent	1182	Armed robbery, conspiracy
Violent	1190	Unarmed robbery
Violent	1191	Unarmed robbery, attempted
Violent	1192	Unarmed robbery, conspiracy
Violent	1200	Aggravated assault
Violent	1201	Aggravated assault, attempted
Violent	1202	Aggravated assault, conspiracy
Violent	1210	Assault of an officer
Violent	1211	Assault of an officer, attempted
Violent	1212	Assault of an officer, conspiracy
Violent	1220	Child abuse
Violent	1221	Child abuse, attempted
Violent	1222	Child abuse, conspiracy
Violent	1230	Simple assault
Violent	1231	Simple assault, attempted
Violent	1232	Simple assault, conspiracy
Violent	1240	Extortion/threat
Violent	1241	Extortion/threat, attempted
Violent	1242	Extortion/threat, conspiracy
Violent	1250	Hit and run with bodily injury
Violent	1251	Hit and run with bodily injury, attempted
Violent	1252	Hit and run with bodily injury, conspiracy
Violent	1990	Violent offense, other
Violent	1991	Violent offense other, attempted
Violent	1992	Violent offense other, conspiracy
Property	2010	Burglary
Property	2011	Burglary, attempted
Property	2012	Burglary, conspiracy
Property	2020	Arson
Property	2021	Arson, attempted
Property	2022	Arson, conspiracy
Property	2030	Auto theft
Property	2031	Auto theft, attempted
Property	2032	Auto theft, conspiracy
Property	2040	Forgery/fraud
Property	2041	Forgery/fraud, attempted
Property	2042	Forgery/fraud, conspiracy
Property	2050	Grand theft (>$500)
Property	2051	Grand theft (>$500), attempted

Offense type	Offense code	Offense code description
Property	2052	Grand theft (>$500), conspiracy
Property	2060	Petty theft (=<$500)
Property	2061	Petty theft (=<$500), attempted
Property	2062	Petty theft (=<$500), conspiracy
Property	2070	Theft, value unknown
Property	2071	Theft, value unknown, attempted
Property	2072	Theft, value unknown, conspiracy
Property	2080	Financial crimes
Property	2081	Financial crimes attempted
Property	2082	Financial crimes conspiracy
Property	2090	Sale of stolen property
Property	2091	Sale of stolen property, attempted
Property	2092	Sale of stolen property, conspiracy
Property	2100	Receiving stolen property
Property	2101	Receiving stolen property, attempted
Property	2102	Receiving stolen property, conspiracy
Property	2110	Destruction of property
Property	2111	Destruction of property, attempted
Property	2112	Destruction of property, conspiracy
Property	2120	Hit and run driving with property damage
Property	2121	Hit and run driving, attempted
Property	2122	Hit and run driving, conspiracy
Property	2130	Unauthorized use of vehicle
Property	2131	Unauthorized use of vehicle, attempted
Property	2132	Unauthorized use of vehicle, conspiracy
Property	2140	Criminal trespass
Property	2141	Criminal trespass, attempted
Property	2142	Criminal trespass, conspiracy
Property	2150	Possession of property crime tools
Property	2151	Possession of property crime tools, attempted
Property	2152	Possession of property crime tools, conspiracy
Property	2990	Other property offense
Property	2991	Other property offense, attempt
Property	2992	Other property offense, conspiracy
Drug	3010	Distribution heroin
Drug	3011	Distribution, heroin, attempted
Drug	3012	Distribution, heroin, conspiracy
Drug	3020	Distribution of amphetamines
Drug	3021	Distribution of amphetamines, attempted
Drug	3022	Distribution of amphetamines, conspiracy
Drug	3030	Distribution cocaine or crack
Drug	3031	Distribution cocaine or crack, attempted
Drug	3032	Distribution cocaine or crack, conspiracy
Drug	3040	Distribution of opioids
Drug	3041	Distribution of opioids, attempted
Drug	3042	Distribution of opioids, conspiracy
Drug	3050	Distribution of prescription drugs
Drug	3051	Distribution of prescription drugs, attempted
Drug	3052	Distribution of prescription drugs, conspiracy
Drug	3060	Distribution other controlled substances
Drug	3061	Distribution other controlled substances, attempted
Drug	3062	Distribution other controlled substances, conspiracy

Offense type	Offense code	Offense code description
Drug	3070	Distribution marijuana
Drug	3071	Distribution marijuana, attempted
Drug	3072	Distribution marijuana, conspiracy
Drug	3080	Distribution, drug unspecified
Drug	3081	Distribution, drug unspecified, attempted
Drug	3082	Distribution, drug unspecified, conspiracy
Drug	3090	Possession/use of heroin
Drug	3091	Possession/use of heroin, attempted
Drug	3092	Possession/use of heroin, conspiracy
Drug	3100	Possession of amphetamines
Drug	3101	Possession of amphetamines, attempted
Drug	3102	Possession of amphetamines, conspiracy
Drug	3110	Possession/use of cocaine or crack
Drug	3111	Possession/use of cocaine or crack, attempted
Drug	3112	Possession/use of cocaine or crack, conspiracy
Drug	3120	Possession of opioids
Drug	3121	Possession of opioids, attempted
Drug	3122	Possession of opioids, conspiracy
Drug	3130	Possession of prescription drugs
Drug	3131	Possession of prescription drugs, attempted
Drug	3132	Possession of prescription drugs, conspiracy
Drug	3140	Possession/use of other controlled substance
Drug	3141	Possession/use of other controlled substance, attempted
Drug	3142	Possession/use of other controlled substance, conspiracy
Drug	3150	Possession/use of marijuana
Drug	3151	Possession/use of marijuana, attempted
Drug	3152	Possession/use of marijuana, conspiracy
Drug	3160	Possession/use of unspecified drug
Drug	3161	Possession/use, drug unspecified, attempted
Drug	3162	Possession/use, drug unspecified, conspiracy
Drug	3170	Heroin violation, offense unspecified
Drug	3180	Amphetamines, offense unspecified
Drug	3190	Cocaine/crack violation, offense unspecified
Drug	3200	Prescription of opioid drugs, offense unspecified
Drug	3210	Prescription, offense unspecified
Drug	3220	Other controlled substance violation, offense unspecified
Drug	3230	Marijuana violation, offense unspecified
Drug	3240	Fraudulent drug offense
Drug	3241	Fraudulent drug offense, attempted
Drug	3242	Fraudulent drug offense, conspiracy
Drug	3250	Drug paraphernalia
Drug	3251	Drug paraphernalia, attempted
Drug	3252	Drug paraphernalia, conspiracy
Drug	3990	Other drug offense
Drug	3991	Other drug offense, attempt
Drug	3992	Other drug offense, conspiracy
DUI offense	4010	Driving while intoxicated
DUI offense	4011	Driving while intoxicated, attempted
DUI offense	4012	Driving while intoxicated, conspiracy
DUI offense	4020	Driving under the influence of alcohol
DUI offense	4021	Driving under the influence of alcohol, attempted
DUI offense	4022	Driving under the influence of alcohol, conspiracy

Offense type	Offense code	Offense code description
DUI offense	4030	Driving under the influence of drugs
DUI offense	4031	Driving under the influence of drugs, attempted
DUI offense	4032	Driving under the influence of drugs, conspiracy
Public order	5010	Riot
Public order	5011	Riot, attempting to incite
Public order	5012	Riot, conspiracy to incite
Public order	5020	Escape from custody
Public order	5021	Escape from custody, attempted
Public order	5022	Escape from custody, conspiracy
Public order	5030	Flight to avoid prosecution
Public order	5031	Flight to avoid prosecution, attempted
Public order	5032	Flight to avoid prosecution, conspiracy
Public order	5040	Weapons offense
Public order	5041	Weapons offense, attempted
Public order	5042	Weapons offense, conspiracy
Public order	5050	Habitual offender
Public order	5060	Parole violation
Public order	5070	Probation violation
Public order	5080	Contempt of court/violate court order
Public order	5081	Contempt of court/violate court order, attempted
Public order	5082	Contempt of court/violate court order, conspiracy
Public order	5090	Other court offense
Public order	5091	Other court offense, attempted
Public order	5092	Other court offense, conspiracy
Public order	5100	Family or custody related offense
Public order	5101	Family or custody related offense, attempted
Public order	5102	Family or custody related offense, conspiracy
Public order	5110	Offense against morals/decency
Public order	5111	Offense against morals/decency, attempted
Public order	5112	Offense against morals/decency, conspiracy
Public order	5120	Immigration violation
Public order	5121	Immigration violation, attempted
Public order	5122	Immigration violation, conspiracy
Public order	5130	Obstruction/resisting
Public order	5131	Obstruction/resisting, attempted
Public order	5132	Obstruction/resisting, conspiracy
Public order	5140	Invasion of privacy
Public order	5141	Invasion of privacy, attempted
Public order	5142	Invasion of privacy, conspiracy
Public order	5150	Commercialized vice
Public order	5151	Commercialized vice, attempted
Public order	5152	Commercialized vice, conspiracy
Public order	5160	Contributing to the delinquency of a minor
Public order	5161	Contributing to the delinquency of a minor, attempted
Public order	5162	Contributing to the delinquency of a minor, conspiracy
Public order	5170	Disorderly conduct offense
Public order	5171	Disorderly conduct offense, attempted
Public order	5172	Disorderly conduct offense, conspiracy
Public order	5180	Liquor law violation
Public order	5181	Liquor law violation, attempted
Public order	5182	Liquor law violation, conspiracy
Public order	5190	Taxation offense

Offense type	Offense code	Offense code description
Public order	5191	Taxation offense, attempted
Public order	5192	Taxation offense, conspiracy
Public order	5200	Bribery/conflict of interest
Public order	5201	Bribery/conflict of interest, attempt
Public order	5202	Bribery/conflict of interest, conspiracy
Public order	5990	Public order offense, other
Public order	5991	Public order offense, other, attempted
Public order	5992	Public order offense, other, conspiracy
Criminal traffic	6010	Traffic offense, minor
Not known/missing	9010	Unspecified felony
Not known/missing	9011	Unspecified felony, attempt
Not known/missing	9012	Unspecified felony, conspiracy
Not known/missing	9020	Unspecified misdemeanor
Not known/missing	9021	Unspecified misdemeanor, attempt
Not known/missing	9022	Unspecified misdemeanor, conspiracy
Other	9990	Other offense
Exclude	8010	Juvenile offense
Flag for removal	8020	Flag for removal
Call for service	8030	Call for service
Federal charges	8040	Federal charges
Not available	8050	Variable not available in county
Not applicable	8060	Not applicable
Not known/missing	9999	Not known/missing

D.2.2 Offense charge grade

This scheme is used for the following variables:

- `adj_grd_cd`

Table 11: Charge grade classification scheme

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

D.2.3 Offense legal code

This scheme is used for the following variables:

- `adj_off_lgl_cd`

Table 12: Charge legal code classification scheme

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

D.2.4   Court disposition

This scheme is used for the following variables:

- `adj_disp_cd`

Table 13: 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

D.2.5   Probation conditions

This scheme is used for the following variables:

- `pro_cond_cd`

Table 14: Probation conditions classification scheme

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

D.2.6   Probation exit

This scheme is used for the following variables:

- `pro_end_cd`

Table 15: Probation exit classification scheme

Probation exit code	Probation exit description
CO	Completion
IN	Incarcerated
AB	Absconded/escaped
DI	Discharged to custody/detainer/warrant

Probation exit code	Probation exit description
OU	Other unsatisfactory exit
TR	Transferred to another probation agency
DE	Death
OT	Other
UU	Not known / missing

D.2.7 Incarceration entry

This scheme is used for the following variables:

- `inc_entry_cd`

Table 16: 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
PR	Probation status - pending revocation
UC	Unsentenced commitment
IE	Illegal entry
UU	Not known / missing

D.2.8 Incarceration facility type

This scheme is used for the following variables:

- `inc_fcl_cd`

Table 17: 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)

Incarceration custody code	Incarceration custody description
AD	Administrative (federal only)
FD	Federal prison
SP	State prison
LJ	Local jail
OT	Other
UU	Not known / missing

D.2.9   Incarceration exit

This scheme is used for the following variables:

- `inc_exit_cd`

Table 18: 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
EA	Escape/AWOL
AI	Accidental Injury to Self
IE	Illegal Entry
UU	Not Known / Missing

D.2.10   Parole exit

This scheme is used for the following variables:

- `par_end_cd`

Table 19: 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

Parole exit code	Parole exit description
OU	Other unsatisfactory exit
TR	Transferred to another state
DE	Death
OT	Other
UU	Not known / missing

E 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.

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

E.2 Episode resolution to link sequences of events

> **USER NOTE**
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense. We expect to complete this work in 2021. Users can expect to see linked processes in the RDC data warehouse in late 2021.

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.

F Record linkage at the U.S. Census Bureau

F.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_id`s 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 primarily 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 agency-validated ids, and are more likely to have had their PII updated.

Figure 12 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_id`s 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 12: Record linkage rate by number of records per `cjars_id`
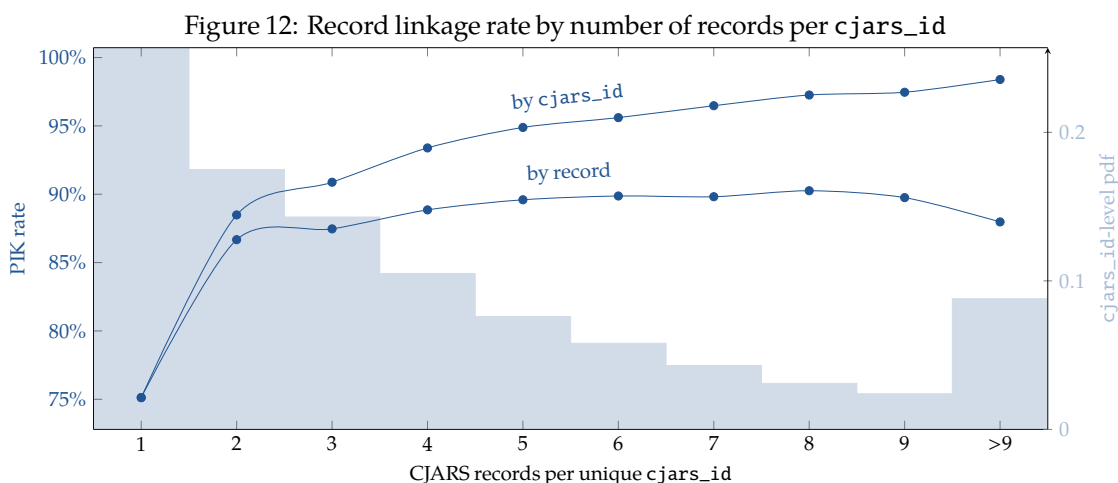
Source: Authors' calculations from the CJARS PIK crosswalk. All results were approved for release by the U.S. Census Bureau Disclosure Review Board (DRB), authorization number CBDRB-FY19-371. All numerators and denominators were rounding according to DRB rounding rules for unweighted counts.

F.2 Making the roster unique

When CJARS data are delivered to the Census Bureau, the roster file includes all of the distinct realizations of PII for every individual. These duplicate records are included to maximize the likelihood that individuals are

matched to a PIK. But this also means that, upon delivery, the roster is not unique by `cjars_id` and PVS may assign different PIKs to a single `cjars_id`. To make the roster unique, we recommend taking the following steps:

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

F.3   Proposal development

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

G   Data notes by jurisdiction

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

G.1   National data notes

Note ID	Data provider	Variables	Notes
US001		`adj_disp_cd`	Data sourced from agencies that serve as a secondary source of court case information sometimes includes disposition information. When `adj_disp_cd` is generated using secondary source data, such as incarceration, its code is assumed as "GU" (Guilty Unclassified) unless there is more information to determine a different code.
US002		`adj_grd_cd_src`, `adj_off_lgl_cd_src`, `adj_disp_cd_src`, `adj_sent_src`, `inc_fcl_cd_src`, `inc_entry_cd_src`, `inc_exit_cd_src`, `pro_cond_cd_src`, `pro_bgn_cd_src`, `pro_end_cd_src`, `par_end_cd_src`	Data sourced from any type of agency often has information stored in open-field string variables. Variables that contain raw source information (i.e., `*_src` variables) identify the source of the information when the value of their sister coded variable (e.g., `adj_grd_cd` and `adj_grd_cd_src`) was generated from secondary source information. For example, if `adj_grd_cd` was coded as felony because the information was sourced from incarceration records, `adj_grd_cd_src` would be recorded as: PRISON RECORD.
US003		`adj_grd_cd`	Data sourced from any type of agency will sometimes have information about offense grade (e.g., misdemeanor versus felony). Coding of `adj_grd_cd` is sometimes assumed based on record source. For instance, if the record was generated using incarceration records, it is assumed to be a felony. Another example of a record being assumed as a felony is when the record came from a trial court system (e.g., Circuit or Superior court system).
US004		`adj_sent_inc`, `adj_sent_inc_max`, `adj_sent_inc_min`	Data sourced from courts (or possibly secondary sources) sometimes includes information about the length of a prison sentence. Death sentences are recorded as -99999 while life sentences are recorded as -88888 for these variables.

Note ID	Data provider	Variables	Notes
US005		`adj_sent_rest`, `adj_sent_fine`	Data sourced from courts (or possibly secondary sources) sometimes includes information about restitution and fines. Notably, negative values appear in `adj_sent_rest` and `adj_sent_fine`, albeit infrequently. This may represent data input errors, payments, elimination of outstanding payments, etc. These negative values were left as is because their explanation is unknown.
US006		`adj_st_ori_fips`	Data sourced from agencies that serve as a secondary source of court case information often includes information about the state of conviction. There are a relatively small number of missing values that are caused by data being generated from secondary sourced data that did not record the location where a case was filed. An example is when court records are generated using data from a Department of Corrections that recorded the sentencing location as: OUT OF STATE.
US007		`adj_disp_cd`, `adj_disp_cd_src`	Data sourced from courts (or possibly secondary sources) sometimes includes sentencing information, but is missing disposition information. The variable `adj_disp_cd` is coded as GU when sentencing information is available for a charge, but disposition information is missing. These records are assigned a code of "GU" (Guilty Unclassified) because it can be implied from this information that an individual was found guilty of a charge. When this is the case, `adj_disp_cd_src` is recorded as: CONV. IMPLD FROM SENT. INFO
US008		`adj_sent_inc`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from courts (or possibly secondary sources) sometimes includes information about the length of a prison sentence. An upper limit of 100 years (1,200 months) is set for incarceration sentences to avoid issues outliers cause. These outliers are caused by factors that cannot be identified (e.g., data quality issues).
US009		`adj_sent_pro`	Data sourced from courts (or possibly secondary sources) sometimes includes information about the length of a probation sentence. An upper limit of 10 years (120 months) is set for probation sentences to avoid issues outliers cause. These outliers are caused by factors that cannot be identified (e.g., data quality issues).

Note ID	Data provider	Variables	Notes
US010		`adj_sent_rest`	Data sourced from courts (or possibly secondary sources) sometimes includes information about restitution. An upper limit of $500,000 dollars is set for restitution to avoid issues outliers cause. These outliers are caused by factors that cannot be identified (e.g., data quality issues).
US011		`adj_sent_fine`	Data sourced from courts (or possibly secondary sources) sometimes includes information about fines. An upper limit of $500,000 and lower limit of -$500,000 dollars is set for fines to avoid issues outliers cause. These outliers are caused by factors that cannot be identified (e.g., data quality issues).
US012		`arr_`, `adj_`, `inc_`, `pro_`, `par_`	Data sourced from any type of agency often has information about the dates that events (e.g., arrests, court case filings) take place. Notably, sometimes records exist with illogical sequencing of events. An example would be a record of a parole spell with the end date of the spell occurring earlier in time than the begin date. Another example would be a court case that has a sentencing date that occurred prior to the filing date. It is important to note that the dates for these records are left as is because it is impossible to determine the cause of the error (e.g., data quality issue).
US013		`adj_grd_cd`	Data sourced from courts (or possibly secondary sources) sometimes includes information about offense grade (.e.g., misdemeanor versus felony). Some records contain missing information about the offense grade of a charge. When this occurs, other available information is leveraged to identify the level of the charge. For example, if other information indicates that the charge was on a felony case, and there was only a single charge on the case, `adj_grd_cd` is assumed to be a felony charge for that record.

G.2   Arizona data notes

Note ID	Data provider	Variables	Notes
AZ001	Arizona Department of Corrections	inc_	Data sourced from the Arizona Department of Corrections Inmate Datasearch system includes information on incarceration terms. This system does not record date of birth which is necessary for data processing purposes. However, date of birth was available through our court record holdings, which were linked to DOC records via exact match on name and court case number. Notably, because our data holdings from the AZ court system are known to have coverage gaps (e.g., geographically), these gaps also impact the coverage of the incarceration records.
AZ002	Arizona Administrative Office of the Courts	adj_off_dt_yyyy, adj_off_dt_mm, adj_off_dt_dd, adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd, adj_sent_src	Data sourced from the Arizona Administrative Office of the Courts includes case records information. This data does not provide information on offense date or sentencing.
AZ003	Arizona Administrative Office of the Courts	adj_disp_dt_yyyy, adj_disp_dt_mm, adj_disp_dt_dd	Data sourced from the Arizona Administrative Office of the Courts includes case record information. This data provides a case disposition date and a charge disposition date. The charge disposition date is used. The case disposition date is used in the roughly 5% of observations where the charge disposition date is missing, except if the case disposition date is 4/3/2016, 7/19/2017, or 11/3/2017 as these appear to be catch all dates when charge disposition dates differ or one of the charge disposition dates is missing.

G.3 Arkansas data notes

Note ID	Data provider	Variables	Notes
AR001	Arkansas Department of Corrections	adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd, adj_off_dt_yyyy, adj_off_dt_mm, adj_off_dt_dd	Data sourced from the Arkansas Department of Corrections Inmate Data search system includes information on incarceration terms. This system does not provide information on case file dates, offense dates, and charges.

Note ID	Data provider	Variables	Notes
AR002	Arkansas Department of Corrections	`inc_exit_dt_yyyy`, `inc_exit_dt_mm`, `inc_exit_dt_dd`	Data sourced from the Arkansas Department of Corrections Inmate Data search system includes information on incarceration terms. This system does not provide information on prison exit dates.
AR003	Arkansas Department of Corrections	`pro_end_dt_yyyy`, `pro_end_dt_mm`, `pro_end_dt_dd`	Data sourced from the Arkansas Department of Corrections Inmate Data search system includes information on incarceration terms. This system does not provide information on probation end dates.

G.4 California data notes

Note ID	Data provider	Variables	Notes
CA001	Butte County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the California Butte County Sheriff's Office website includes booking information. This website does not provide arrest date information.
CA002	Nevada County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the California Nevada County Sheriff's Office website includes booking information. This website does not provide arrest date information.
CA003	San Luis Obispo County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the California San Luis Obispo County Sheriff's Office website includes booking information. This website does not provide arrest date information.
CA004	Anaheim Police Department	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the California Anaheim Police Department website includes arrest information. This website does not provide booking date information.
CA005	Bakersfield Police Department	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the California Bakersfield Police Department website includes arrest information. This website does not provide booking date information.
CA006	Riverside Police Department	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the California Riverside Police Department includes arrest information. This website does not provide arrest date information.
CA007	Riverside Police Department	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the California Riverside Police Department includes booking information. This website does not provide arrest date information.

Note ID	Data provider	Variables	Notes
CA008	San Bernardino Police Department	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the California San Bernardino Police Department website includes arrest information. This website does not provide booking date information.
CA009	San Diego Police Department	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the California San Diego Police Department website includes arrest information. This website does not provide booking date information.

G.5   Colorado data notes

Note ID	Data provider	Variables	Notes
CO001	Adams County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the Colorado Adams County Sheriff's Office website includes booking information. This website does not provide arrest date information.
CO002	Colorado Department of Corrections	`inc_cnty_ori_fips`, `inc_st_ori_fips`, `par_cnty_ori_fips`	Data sourced from the Colorado Department of Corrections includes information on incarceration and parole terms. This data does not contain information on location of (both state and county) sentencing.

G.6   Florida data notes

Note ID	Data provider	Variables	Notes
FL001	Hillsborough County Clerk of Courts	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`	Data sourced from the Florida Hillsborough County Clerk Court system includes case record information. This system does not provide sentencing information.
FL002	Florida Department of Corrections	`inc_exit_dt_dd`	Data sourced from the Florida Department of Corrections includes information on the date that inmates exit prison. There is a concentration of exits on the first of the month. This may reflect a standard release date, or it may reflect data warehouse processes that record some types of events as always occurring on the first of the month. See Blomberg et al. (2011, p. 20).

Note ID	Data provider	Variables	Notes
FL003	Florida Department of Corrections	`pro_end_dt_yyyy`, `pro_end_dt_mm`, `pro_end_dt_dd`, `par_end_dt_yyyy`, `par_end_dt_mm`, `par_end_dt_dd`	Data sourced from the Florida Department of Corrections includes repeated snapshots of the current population under probation and parole supervision. Consequently, probation and parole exit dates are not observed.
FL004	Pinellas County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the Florida Pinellas County Sheriff includes booking information. Arrest date information is not observed. Only contains booking date, offense description, offense level (grade), state FIPS and county FIPS info.

G.7 Illinois data notes

Note ID	Data provider	Variables	Notes
IL001	Illinois Department of Corrections	`adj_disp_dt_yyyy`, `adj_disp_dt_mm`, `adj_disp_dt_dd`, `adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`, `adj_off_dt_yyyy`, `adj_off_dt_mm`, `adj_off_dt_dd`	Data sourced from the Illinois Department of Corrections inmate search includes information on court case records. This website does provide not provide information on offense date, case file date, disposition date, and sentence date.
IL002	Illinois Department of Corrections	`par_end_dt_yyyy`, `par_end_dt_mm`, `par_end_dt_dd`	Data sourced from the Illinois Department of Corrections inmate search includes information on parole terms. This website does not provide information on parole end dates.

G.8 Kansas data notes

Note ID	Data provider	Variables	Notes
KS001	Johnson County Sheriff's Office	arr_arr_dt_yyyy, arr_arr_dt_mm, arr_arr_dt_dd	Data sourced from the Kansas Johnson County Sheriff's Office includes booking information. This website does not provide information on arrest date.

G.9   Maryland data notes

Note ID	Data provider	Variables	Notes

G.10   Michigan data notes

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

G.11   Minnesota data notes

Note ID	Data provider	Variables	Notes
MN001	Minnesota State Court Administrator's Office	adj_disp_cd	Data sourced from the Minnesota State Court Administrator's Office includes information on court case filings. However, the only court case filings that are covered from this source include those that resulted in a conviction.

G.12 Mississippi data notes

Note ID	Data provider	Variables	Notes
MS001	Mississippi Department of Corrections	adj_off_dt_yyyy, adj_off_dt_mm, adj_off_dt_dd, adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd	Data sourced from the Mississippi Department of Corrections inmate search includes court case record information. This website does provide not provide information on offense date and case file date.
MS002	Mississippi Department of Corrections	inc_exit_dt_yyyy, inc_exit_dt_mm, inc_exit_dt_dd	Data sourced from the Mississippi Department of Corrections inmate search includes information on incarceration terms. This website does not provide information on prison exit date.

G.13 Nebraska data notes

Note ID	Data provider	Variables	Notes
NE001	Nebraska Department of Correctional Services	adj_disp_dt_yyyy, adj_disp_dt_mm, adj_disp_dt_dd, adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd, adj_off_dt_yyyy, adj_off_dt_mm, adj_off_dt_dd, adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd	Data sourced from the Nebraska Department of Correctional Services includes sentencing information on those under the agency's supervision. The data only includes the beginning and end dates of the sentence term. It does not include the offense, case file, disposition, and sentence dates.

Note ID	Data provider	Variables	Notes
NE002	Nebraska Department of Correctional Services	`adj_grd_cd`, `adj_grd_cd_src`	Data sourced from the Nebraska Department of Correctional Services includes offense information for the offenses inmates were sentenced for. The data includes information on the grade of the offense and includes some misdemeanor records. This is indicated in the `adj_rec_src_doc` variable. Some misdemeanor records are expected (e.g., those in a case with a felony charge). However, some of the charges classified as misdemeanors from the offense grade information provided are associated with descriptions in `adj_disp_off_cd_src` that appear to be felony offenses. The original grade information is retained along the with the offense description.
NE003	Nebraska Department of Correctional Services	`par_bgn_dt_dd`	Data sourced from the Nebraska Department of Correctional Services includes information on parole terms. The distribution of the beginning of parole terms is not evenly distributed across the days of the month. Specifically, more individuals begin parole during the latter part of each month, which causes the distribution to be left skewed. This caused by the timing of Nebraska's parole board hearings which take place during the last two weeks of each month.

G.14 New Jersey data notes

Note ID	Data provider	Variables	Notes
NJ001	Superior Court of New Jersey	`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`	Data sourced from the Superior Courts of New Jersey includes information on sentencing. The data 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.

Note ID	Data provider	Variables	Notes
NJ002	Superior Court of New Jersey	adj_	Data sourced from the Superior Courts of New Jersey includes information on Superior Court case filings. The data only include cases filed in the Superior Court system of New Jersey. Therefore, all other cases filed in lower courts (e.g., municipal courts) are not included.
NJ003	Superior Court of New Jersey	adj_grd_cd, adj_grd_cd_src	Data sourced from the Superior Courts of New Jersey includes information on offense grades (e.g., misdemeanor versus felony). The data only include cases filed in the Superior Court system of New Jersey. Therefore, all other cases filed are assumed to be felonies. Source variable indicates Superior Court case filing to convey this.

G.15 North Carolina data notes

Note ID	Data provider	Variables	Notes
NC001	North Carolina Department of Public Safety	adj_file_dt_yyyy, adj_file_dt_mm, adj_file_dt_dd	Data sourced from the North Carolina Department of Public Safety includes court case record information. This data does not provide case file date.
NC002	North Carolina Department of Public Safety	adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd	Data sourced from the North Carolina Department of Public Safety offender search system includes court case record information. This data does not provide sentence date.
NC003	North Carolina Department of Public Safety	pro_end_dt_dd, pro_end_dt_mm, pro_end_dt_yyyy	Data sourced from the North Carolina Department of Public Safety offender search system includes information on probation terms. This data does not provide probation end date.
NC004	North Carolina Administrative Office of the Courts	adj_	Data sourced from the North Carolina Administrative Office of the Courts contains records on court case filings. Some of the records included in the CJARS repository from this agency were obtained from open access data that was submitted as replication data for Silveira (2017). These court data exclude records where the method of disposition included: dismissal by deferred prosecution, dismissal with leave by district attorney, or dismissal without leave by district attorney (codes DD, VL, and VD in source variable CRDMOD).

G.16   Ohio data notes

Note ID	Data provider	Variables	Notes
OH001	Ohio Department of Rehabilitation and Correction	`adj_off_dt_yyyy`, `adj_off_dt_mm`, `adj_off_dt_dd`, `adj_file_dt_yyyy`, `adj_file_dt_mm`, `adj_file_dt_dd`, `adj_disp_dt_yyyy`, `adj_disp_dt_mm`, `adj_disp_dt_dd`, `adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`	Data sourced from the Ohio Department of Corrections inmate search includes court case record information. This website does not provide information on offense date, case file date, disposition date, and sentence date.

G.17   Oregon data notes

Note ID	Data provider	Variables	Notes
OR001	Oregon Judicial Department	`adj_sent_src`, `adj_disp_off_cd_src`	Data sourced from the Oregon Judicial Department includes information on case records and sentencing. Raw sentencing data included personally identifiable information of restitution payees, which has been removed in the final version.

G.18   Pennsylvania data notes

Note ID	Data provider	Variables	Notes
PA001	Administrative Office of Pennsylvania Courts	`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`	Data sourced from the Administrative Office of Pennsylvania Courts includes information on court case filings. However, this data does not include sentencing information. Fine and restitution amount information are also missing. Case information is limited to offense type, offense date, file date, and disposition information.

G.19 Texas data notes

Note ID	Data provider	Variables	Notes
TX001	Bexar County Clerk	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`	Data sourced from the Texas Bexar County Clerk website includes case record information. This website does not provide information on sentence date.
TX002	Bexar County Sheriff's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the Texas Bexar County Sheriff's Office website includes arrest information. This website does not provide information on booking date.
TX003	Collin County Courts	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`	Data sourced from the Texas Collin County Courts Records Inquiry system includes case record information. This system does not provide sentencing information.
TX004	Collin County Courts	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the Texas Collin County Courts Records Inquiry system includes information on arrests. This system does not provide booking date.
TX005	Dallas County Sheriff's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the Texas Dallas County Sheriff's Office includes booking information. This data does not provide arrest date.
TX006	El Paso County Clerk's Office	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`	Data sourced from the Texas El Paso County County Clerk includes case record information. This data does not provide sentence date.

Note ID	Data provider	Variables	Notes
TX007	El Paso County Clerk's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the Texas El Paso County County Clerk includes information on arrests. This data does not provide booking date.
TX008	El Paso District Clerk	`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`, `arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the El Paso County District Clerk's Office includes information about court case filings and includes arrest information. Filing date and sentencing information is missing for these records, as well as booking date.
TX009	Harris County District Clerk	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`, `adj_off_dt_yyyy`, `adj_off_dt_mm`, `adj_off_dt_dd`	Data sourced from the Texas Harris County District Clerk includes case record information. This data does not provide offense date and sentence date.
TX010	Hays County Courts at Law	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from the Texas Hays County Courts Records Inquiry includes information on arrests. This system does not provide arrest date.
TX011	Hays County Courts at Law	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`	Data sourced from the Texas Hays County Courts Records Inquiry includes case record information. This system does not provide sentence date.
TX012	Tarrant County Sheriff's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from the Texas Tarrant County Sheriff's Office website includes arrest information. This website does not provide booking date.
TX013	Texas Department of Criminal Justice	`adj_file_dt_yyyy`, `adj_file_dt_mm`, `adj_file_dt_dd`	Data sourced from the Texas Department of Criminal Justice includes information on court case records. This data does not provide case file date.

Note ID	Data provider	Variables	Notes
TX014	Texas Department of Criminal Justice	`inc_entry_dt_yyyy`, `inc_exit_dt_yyyy`	Data sourced from the Texas Department of Criminal justice includes information on incarceration and parole terms. The data is missing most records from 1989 because of a data loss event.
TX015	Texas Department of Criminal Justice	`par_bgn_dt_yyyy`, `par_end_dt_yyyy`	Data sourced from the Texas Department of Criminal justice includes information on incarceration and parole terms. The data is missing most records from 1989 because of a data loss event.
TX016	Texas Department of Criminal Justice	`par_end_dt_yyyy`, `par_end_dt_mm`, `par_end_dt_dd`	Data sourced from the Texas Department of Criminal Justice includes information on parole terms. The data does not include descriptions for parole status codes. Entries are able to be approximated using sentencing information but exits are harder to pinpoint resulting in a large share of unknown exit dates.
TX017	Texas Department of Criminal Justice	`pro_end_dt_yyyy`, `pro_end_dt_mm`, `pro_end_dt_dd`, `pro_cond_cd`, `pro_end_cd`	Data sourced from the Texas Department of Criminal Justice includes information on probation terms. The data does not include actual exit dates, but does include projected exit dates based on the term length. Projected exits that do not extend beyond the date of data collection are used as probation exit dates. The data is also missing `pro_cond_cd` and `pro_end_cd`.
TX018	iDocket	`adj_`, `adj_chrg_off_cd`, `adj_off_dt_yyyy`, `adj_off_dt_mm`, `adj_off_dt_dd`, `adj_sent_serv`, `adj_sent_dth`, `adj_sent_trt`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from iDocket includes information on case records provided individually by counties in Texas. Data quality and coverage varies by county. Variables often have higher rates of UU codes and sentencing information is often missing. The variables listed are missing on all observations.
TX019	Bexar County Sheriff's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from Bexar County Sheriff's Office includes information on arrest records. There is no information on booking dates.

Note ID	Data provider	Variables	Notes
TX020	Collin County Courts	`adj_chrg_off_cd`, `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_trt`, `adj_sent_fine`, `adj_sent_inc_min`, `adj_sent_inc_max`, `arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from Collin County Courts includes information about court case filings and includes arrest information. The variables listed are missing for all observations from this source.
TX021	Dallas County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from Dallas County Sheriff's Office includes information on bookings. There is no information on arrest date.
TX022	El Paso County Clerk's Office	`adj_chrg_off_cd`, `adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`, `adj_sent_serv`, `adj_sent_dth`, `adj_sent_pro`, `adj_sent_rest`, `adj_sent_trt`, `adj_sent_fine`, `adj_sent_inc_min`, `adj_sent_inc_max`, `arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from El Paso County Clerk includes information on court case filings as well as arrest information. The variables listed are missing for all observations from this source.
TX023	Harris County Sheriff's Office	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from Harris County Sheriff's Office includes information on bookings. There is no information available on arrest dates.

Note ID	Data provider	Variables	Notes
TX024	Hays County Courts at Law	`arr_arr_dt_yyyy`, `arr_arr_dt_mm`, `arr_arr_dt_dd`	Data sourced from Hays County Courts at Law includes case records with arrest information. There is no information available on arrest dates.
TX025	Hays County Sheriff's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from Hays County Sheriff's Office includes information on arrest records. There is no information available on booking dates.
TX026	Tarrant County Sheriff's Office	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`	Data sourced from Tarrant County Sheriff's Office includes information on bookings. There is no information available on arrest dates.
TX027	Texas Department of Public Safety	`arr_book_dt_yyyy`, `arr_book_dt_mm`, `arr_book_dt_dd`, `adj_sent_serv`, `adj_sent_rest`, `adj_sent_trt`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from Texas Department of Safety includes information on court case and arrest records. There variables listed are missing on all observations.
TX028	Texas Department of Criminal Justice	`inc_fcl_cd`, `inc_entry_cd`, `inc_exit_cd`, `adj_file_dt_yyyy`, `adj_file_dt_mm`, `adj_file_dt_dd`, `adj_chrg_off_cd`, `adj_sent_serv`, `adj_sent_dth`, `adj_sent_inc`, `adj_sent_pro`, `adj_sent_rest`, `adj_sent_sus`, `adj_sent_trt`, `adj_sent_fine`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from Texas Department of Criminal of Criminal Justice includes information on incarceration records and court records. The following variables are missing for all observations in some data extracts: `inc_fcl_cd`, `inc_entry_cd`, and `inc_exit_cd`.

Note ID	Data provider	Variables	Notes
TX029	Bexar County Clerk	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`, `adj_sent_serv`, `adj_sent_inc`, `adj_sent_pro`, `adj_sent_rest`, `adj_sent_trt`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from Bexar County Clerk includes information on case records. The variables listed are missing on all observations.
TX030	Bexar County District Clerk's Office	`adj_sent_serv`, `adj_sent_rest`, `adj_sent_trt`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from Bexar District Clerk includes information on case records. The variables listed are missing on all observations.
TX031	Harris County District Clerk	`adj_chrg_off_cd`, `adj_off_dt_yyyy`, `adj_off_dt_mm`, `adj_off_dt_dd`, `adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`, `adj_sent_serv`, `adj_sent_rest`, `adj_sent_trt`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from Harris District Clerk includes information on case records. The variables listed are missing on all observations.
TX032	Hays County Courts at Law	`adj_sent_dt_yyyy`, `adj_sent_dt_mm`, `adj_sent_dt_dd`, `adj_sent_dth`, `adj_sent_inc`, `adj_sent_pro`, `adj_sent_rest`, `adj_sent_fine`, `adj_sent_inc_min`, `adj_sent_inc_max`	Data sourced from Hays District Clerk includes information on case records. The variables listed are missing on all observations.

G.20   Virginia data notes

Note ID	Data provider	Variables	Notes
VA001	Judiciary of Virginia	cjars_id	Data sourced from the Judiciary of Virginia is missing year of birth. The CJARS entity resolution algorithm was adapted to allow for matches on month and day of birth, as well as on the other PII variables, so that the records could still be integrated.

G.21   Washington data notes

Note ID	Data provider	Variables	Notes
WA001	Washington State Department of Corrections	inc_cnty_ori_fips	Data sourced from the Washington Department of Corrections includes information on incarceration terms. This data does not provide information on county of conviction.

G.22   Wisconsin data notes

Note ID	Data provider	Variables	Notes
WI001	Wisconsin Department of Corrections	inc_cnty_ori_fips	Data sourced from the Wisconsin Department of Corrections includes information on incarceration terms. County of conviction is missing for most records because it was obtained separately through the Department of Correction's website only for a small number of records.
WI002	Wisconsin Department of Corrections	inc_entry_dt_yyyy, inc_exit_dt_yyyy	Data sourced from the Wisconsin Department of Corrections includes information on incarceration terms. The start date for tracking entries is January of 1990. However, the distribution of exits suggests that we are not observing all exits starting at this time, but rather, the exits of inmates that had been admitted to prison since January 1990.

H 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.

H.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 13 approximates our understanding of the layout of the two databases.

H.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 "entities" (tables). These tables contain information relating to all stages of the criminal justice system. Figure 14 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.

H.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 15.

---

[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 (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.

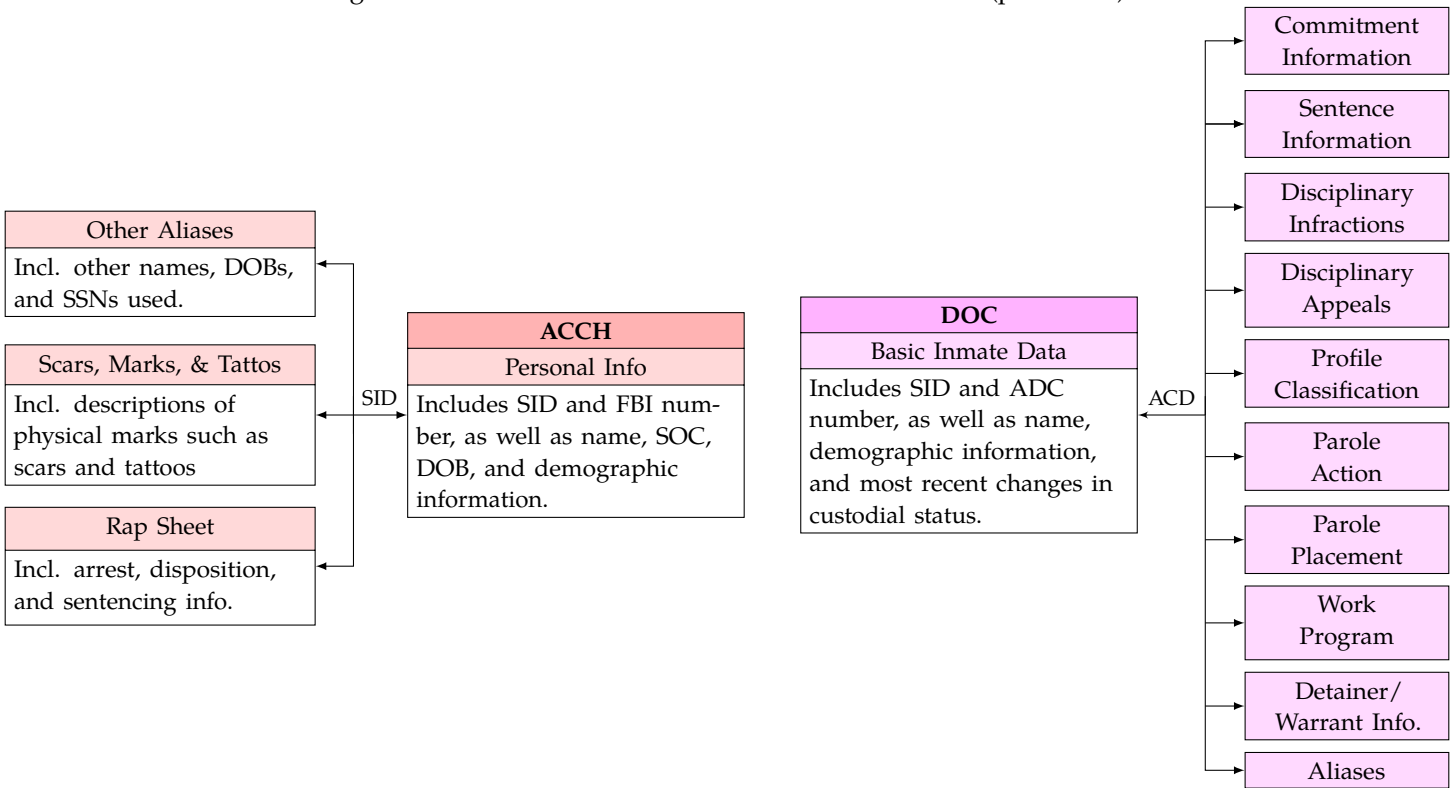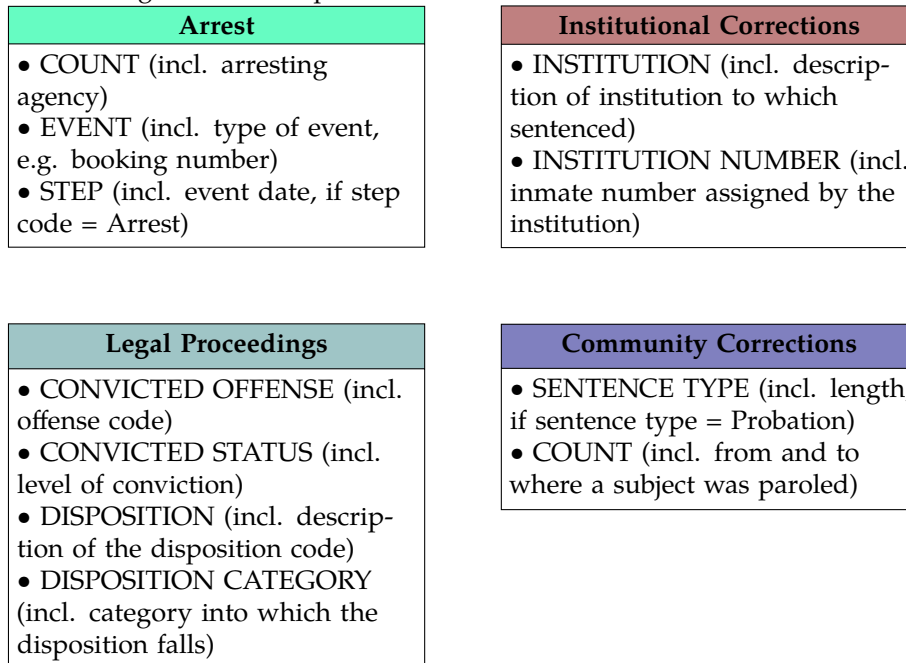Figure 13: Arizona ACCH & DOC AIMS database model (presumed)

Figure 14: 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)

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.

Figure 15: Connecticut CCH database model
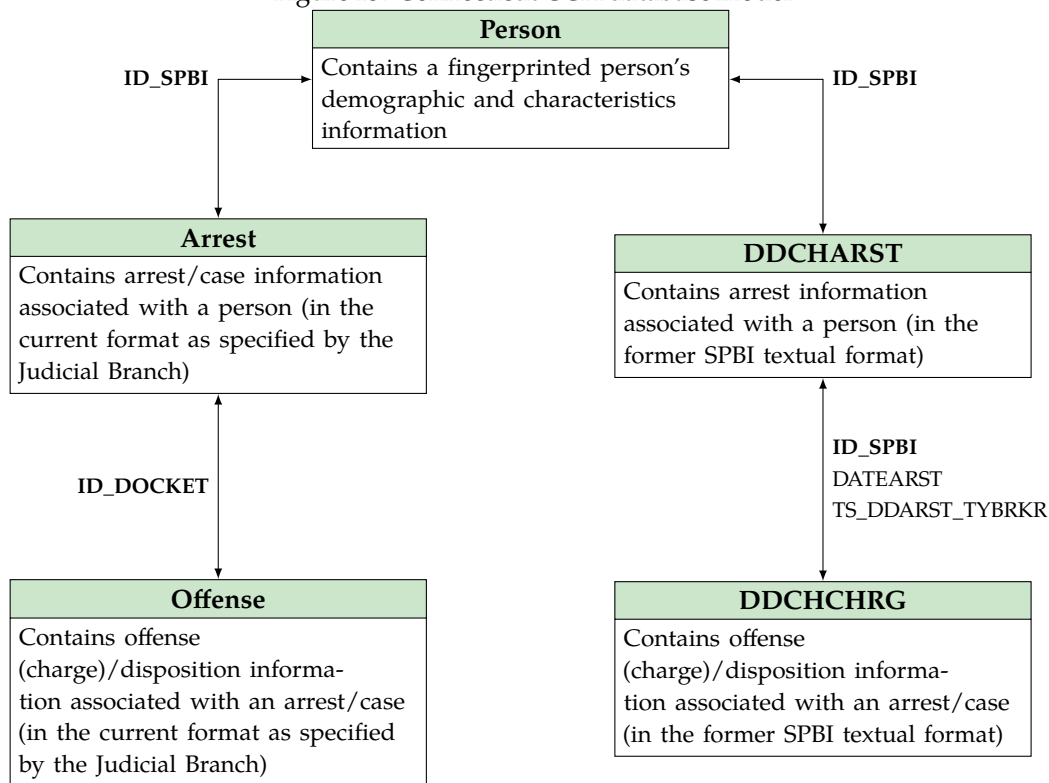
H.4 Florida

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.

As shown in Figure 16, 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.

H.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

---

[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

Figure 16: Florida DOC Public Database ERD (presumed)

Active Inmates	Released Inmates	Current Community Supervision Offenders
Inmate_active_root Basic inmate information on active inmates	**Inmate_release_root** Basic inmate info. on inmates released since 10-1-97	**Offender_root** Basic offender information on current community supervision offenders
Inmate_active_aliases Aliases for these active inmates	**Inmate_release_aliases** Aliases for these released inmates	**Offender_aliases** Aliases for offenders
Inmate_active_offenses_CPS Current prison offenses for these active inmates	**Inmate_release_offenses_CPS** Current prison offenses for these released inmates	**Offender_offenses_CCS** Current community supervision offenses for offenders
Inmate_active_offenses_prpr Prior prison offenses for these active inmates	**Inmate_release_offenses_prpr** Prior prison offenses for these released inmates	**Offender_residence** Most recent address (excluding confidential addresses) for offenders
Inmate_active_detainers Detainers for these active inmates	**Inmate_release_detainers** Detainers for these released inmates	
Inmate_active_incarhist Incarceration history for these active inmates	**Inmate_release_incarhist** Incarceration history for these released inmates	
Inmate_active_scarsmarks Tattoos for these active inmates	**Inmate_release_scarsmarks** Tattoos for these released inmates	
	Inmate_release_residence Release plan address for these released inmates	

(DCNumber links the Active Inmates, Released Inmates, and Current Community Supervision Offenders groupings.)

Figure 17) 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.

H.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

---

[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.

[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.

Figure 17: Illinois CHRI Ad Hoc database model (presumed)
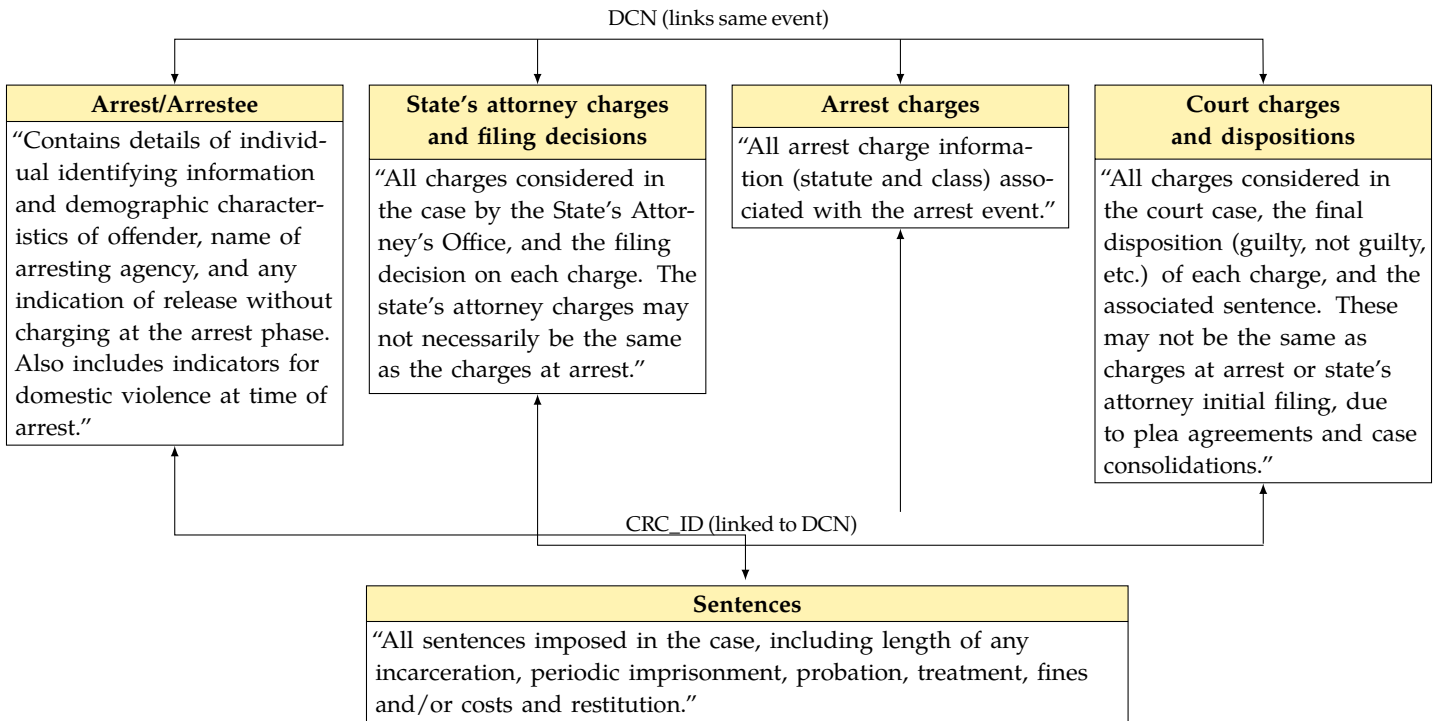
Repository for Criminal History Record Information (CHRI). The Kansas Central Repository is a system of connected data sources that contains information about felony and misdemeanor arrests, court convictions, dispositions, and incarceration in state-operated facilities. The Central Repository database 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 18). KBI receives incarceration information from the Kansas Adult Supervised Population Electronic Repository (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 18 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.

H.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 19.

The Sentencing Commission collects this information into two SPSS databases which are updated annually based on policy change.

Figure 18: Kansas Disposition Report (KDR) Class Diagrams (presumed)
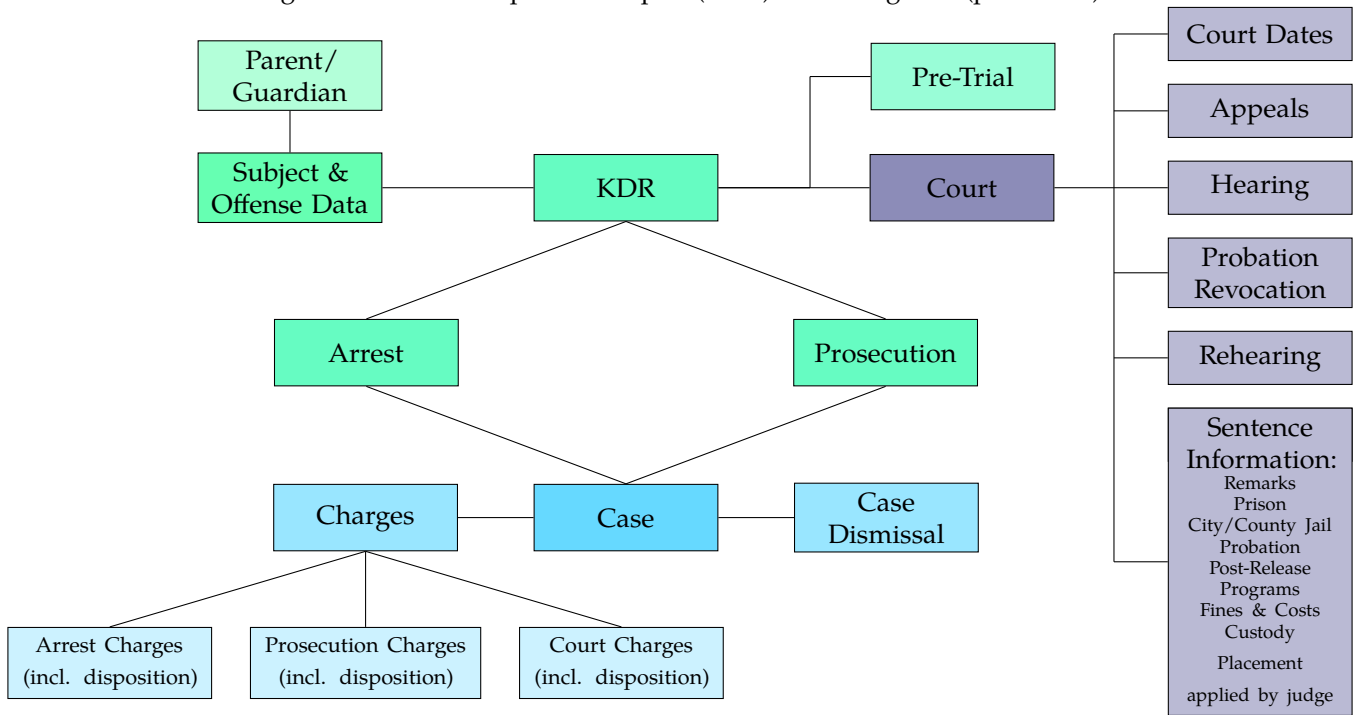
Figure 19: Kentucky CCH database model (presumed)
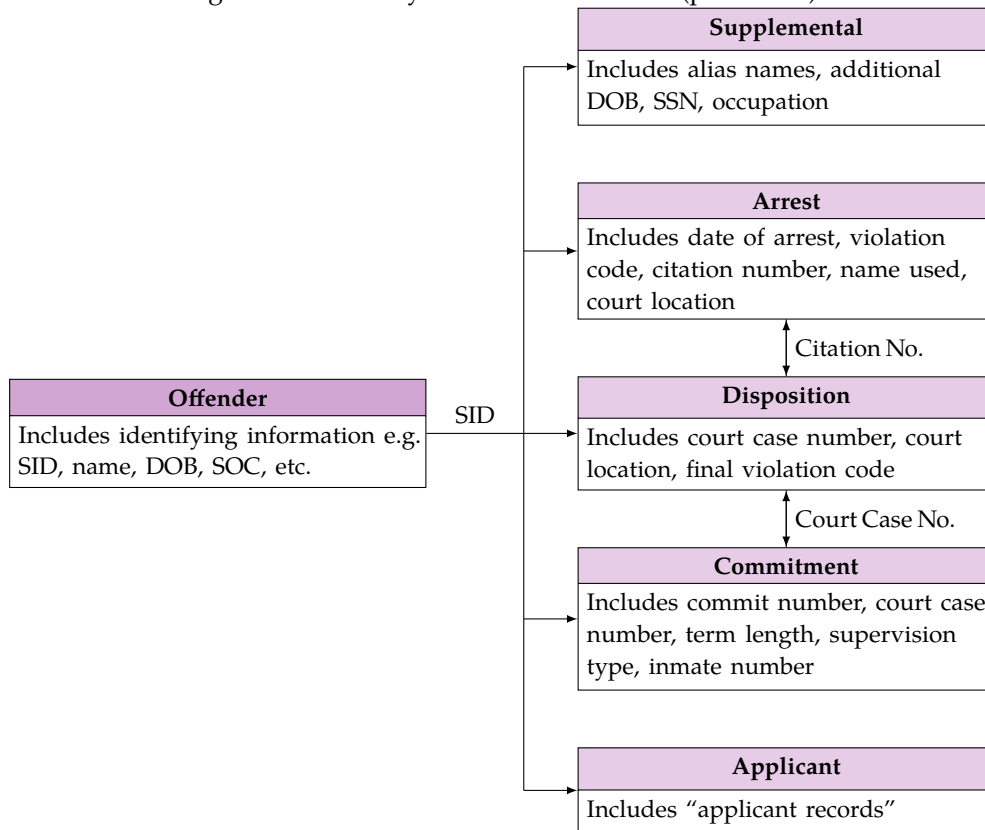
H.8   Minnesota

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 20 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 20: Minnesota Criminal History System (abridged/presumed)
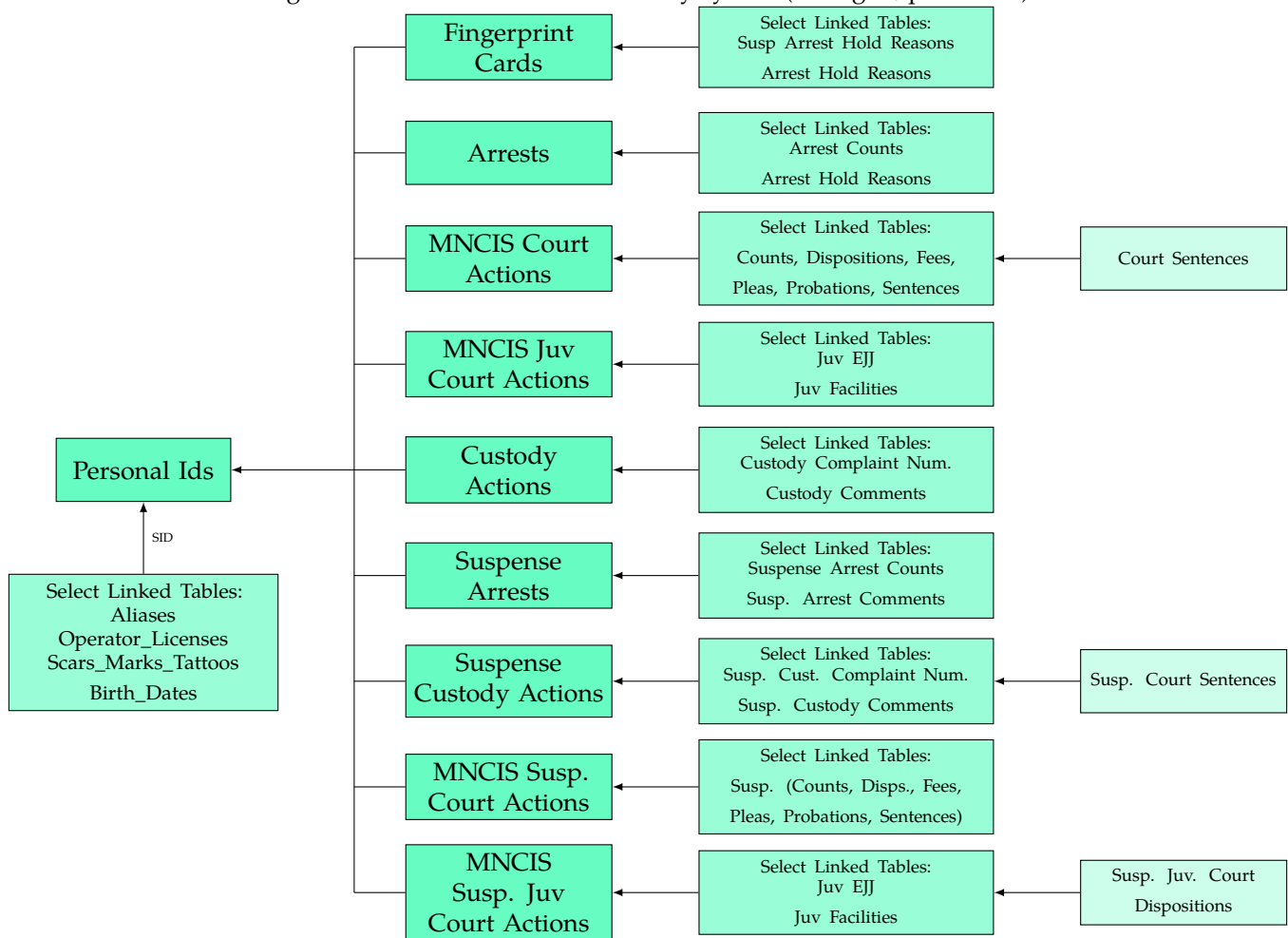
H.9   Nevada

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.

---

[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.

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 21 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 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 21: Nevada CCH database model (presumed)
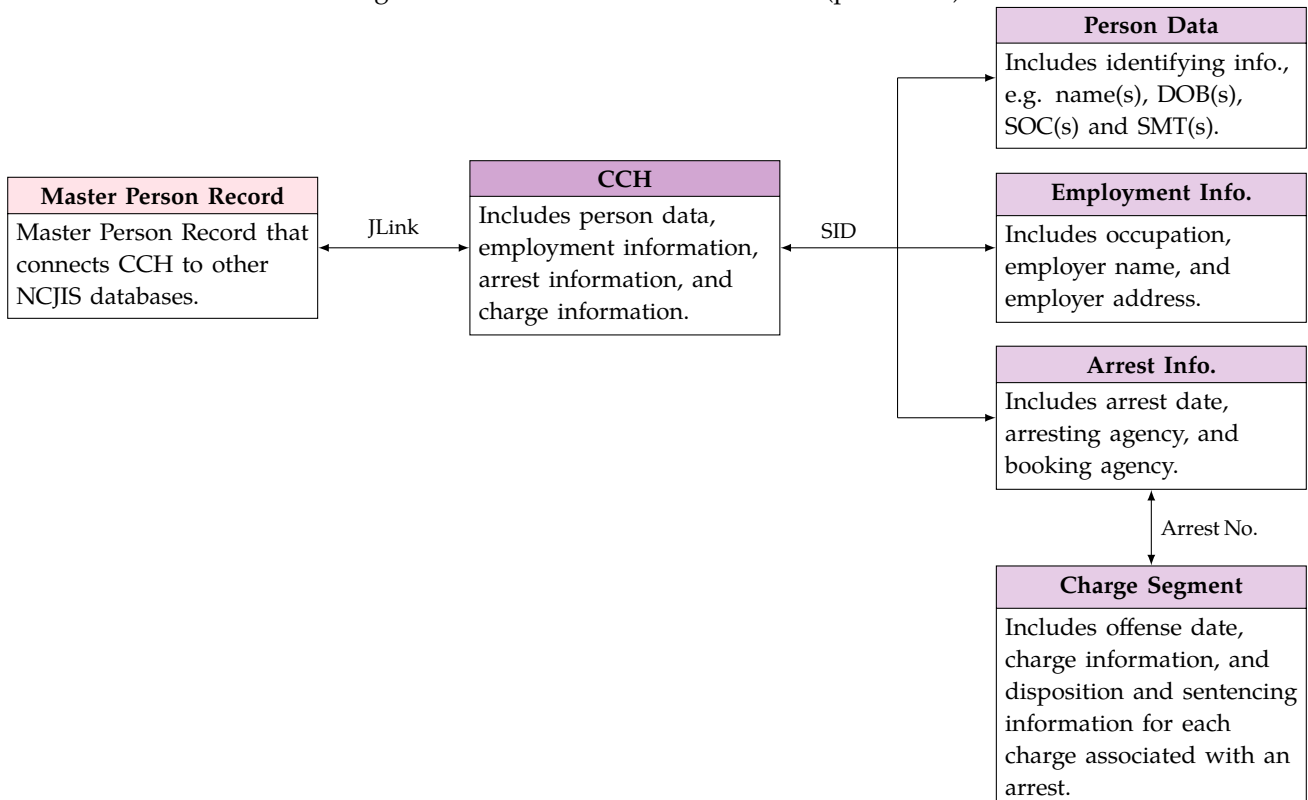
H.10   New Mexico

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 22 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).

Figure 22: New Mexico ERD (abridged)
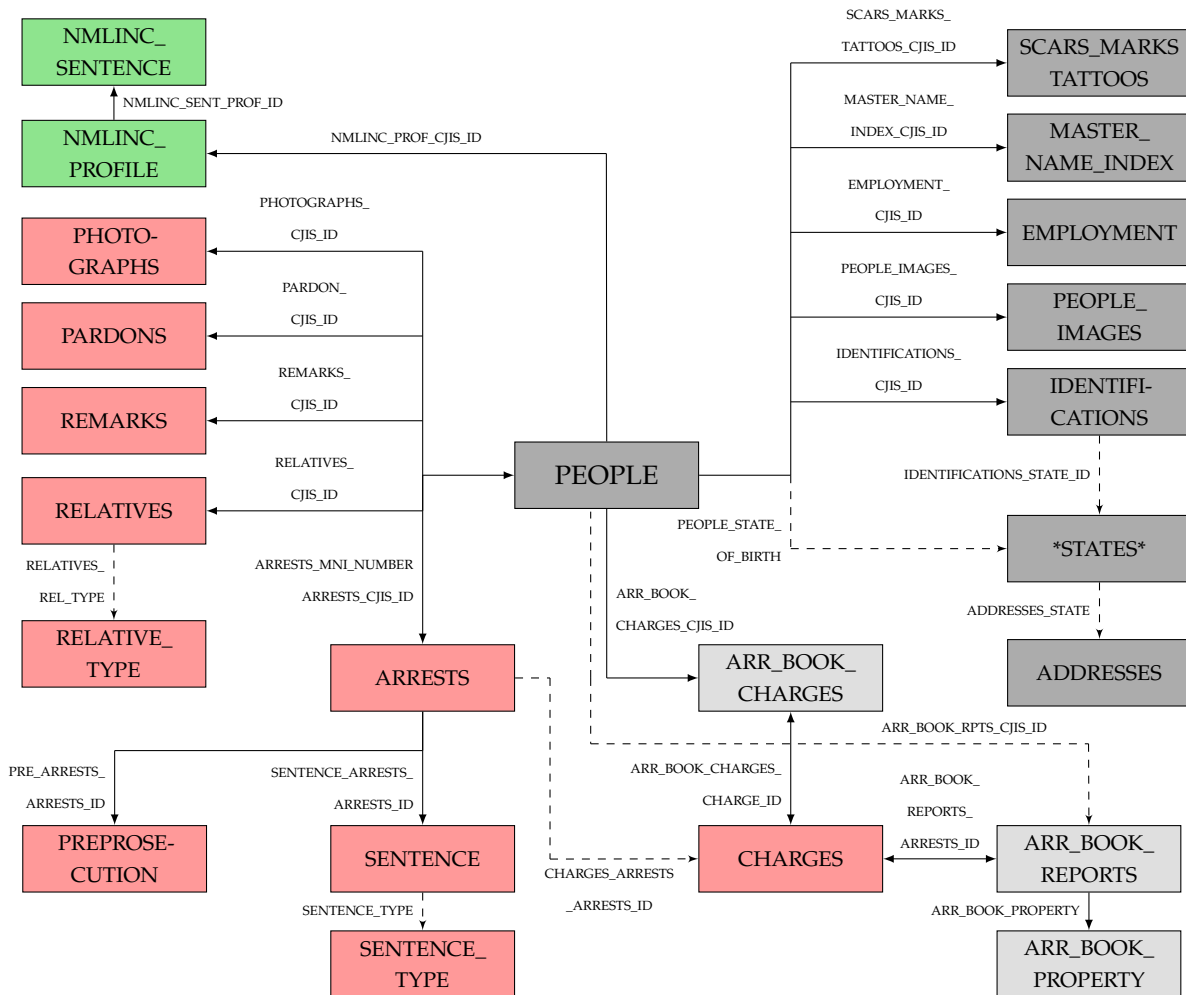
H.11 New York

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. 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 23, 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.

206

Figure 23: New York CCH database model (presumed)

Cycle IDs
Includes Criminal Justice Tracking Number (CJTN), used for "tracking subsequent actions related to an arrest."

Offender Priors
Incl. summaries of selected types of prior arrests and convictions

Date & Area Fields
Includes crime, arrest, disposition, and conviction dates, and county, region, and arresting or disposing agency.

Arrest Charge Details
Arrest information, including effective category class.

Lower Court Arraignment Charge Details
Lower Court information, including effective category class.

Upper Court Arraignment Charge Details
Upper Court information, including effective category class.

Disposition Charge Details
Disposition information, including effective category class.

Disposition and Sentence Details
Disposition and sentencing information, including jail time, fines, and for community service

Person Data
Includes the New York State Identification number (NYSID) and demographic data.

The Overview states that the NYSID number is "a unique identifier assigned to offenders upon arrest and carries through their criminal history." Though the Overview does not explicitly say so, we assume that NYSID connects the Person Data to all other tables in the relational database. |

NYSID

H.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 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 24 for an approximation of how ACIS is organized. These data are linked to other criminal history data through SID by the SBI.

Figure 24: North Carolina ACIS database model (LIN link assumed)
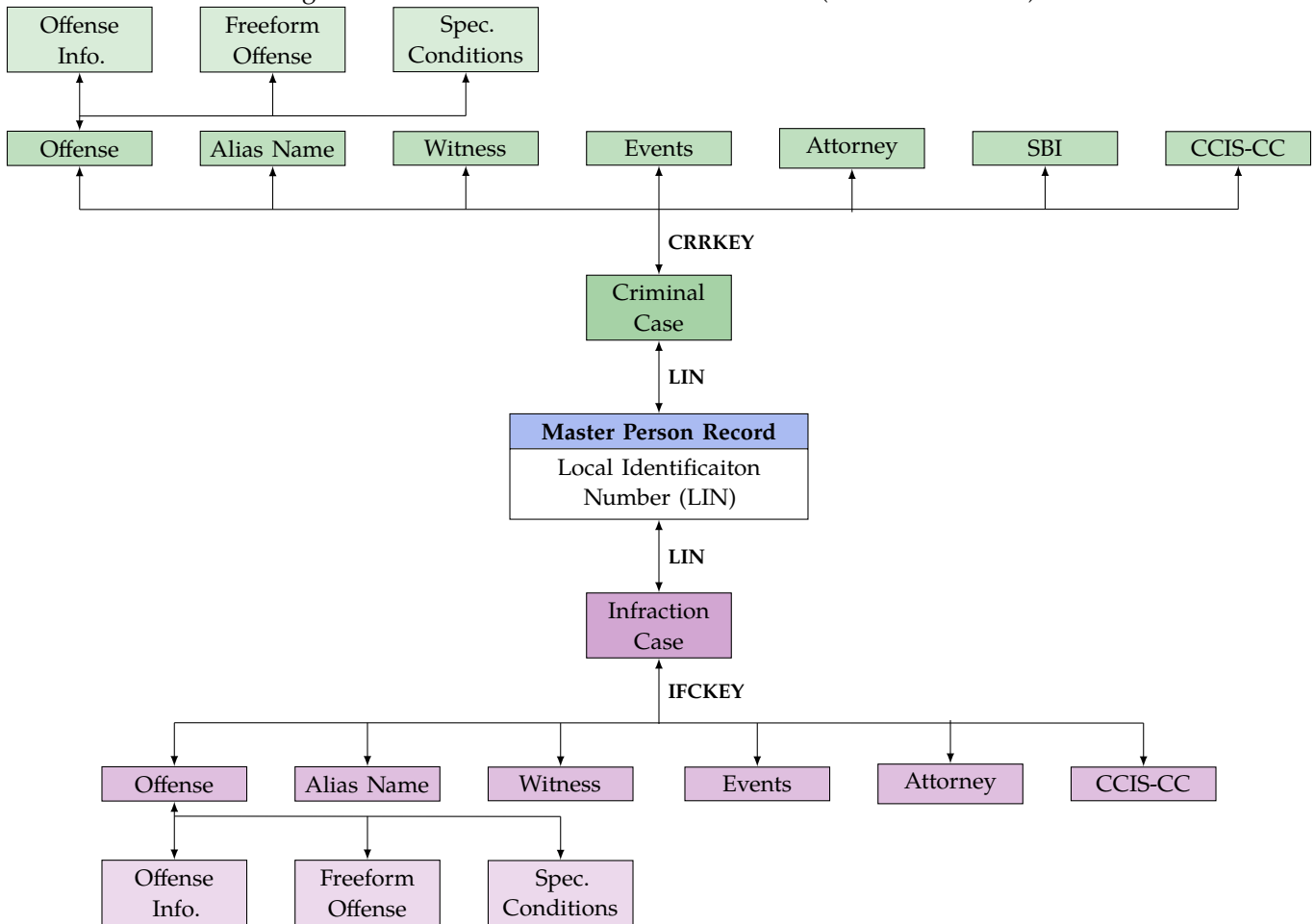
H.13   Oregon

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 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 25.

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.

H.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

Figure 25: Oregon eCourt (Odyssey) ERD
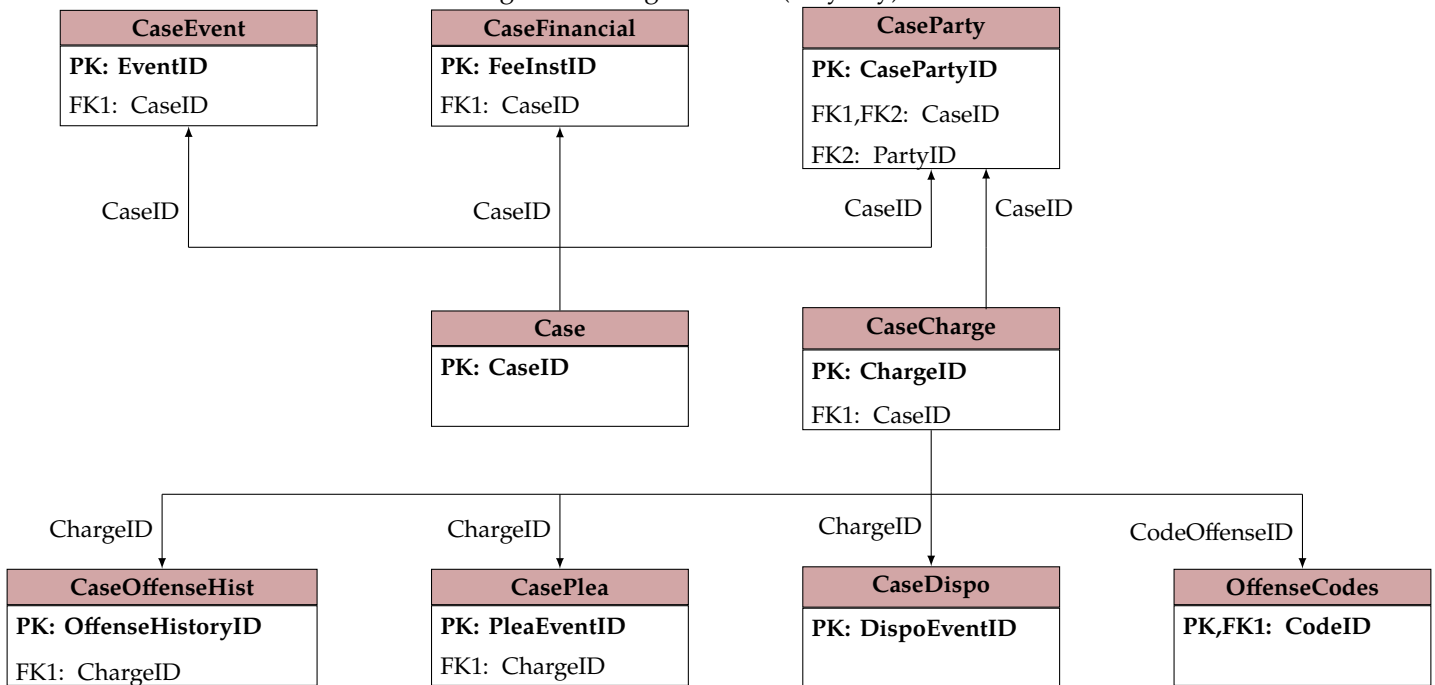
Data Dictionary (CJDD), which contains data elements and their individual agency specific definitions. The participating agencies are the Administrative Office of Pennsylvania Courts (AOPC), the Department of Corrections (DOC), the Pennsylvania Board of Probation and Parole (PBPP), the Pennsylvania Commission 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 State ID data element to link these databases together (see Figure 26).[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.

H.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 27.

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).

---

[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 26: Pennsylvania database relationship (presumed)

Left column tables:

Clssnp (1)
Ethnic identity

dbo_tblRISPDATA (32)
Drug and alcohol related

dbo_Perrec (118)
Drug, military info

dbo_Senrec (26)
Sentence info

dbo_Mischg (17)
Counts of guilty pleas

dbo_tblSearchInmateInfo (57)
Facility info (dorm-level)

dbo_Miscon (34)
Misconduct info

dbo_Movrec (20)
Movement data

dbo_tblEval/Hist (19)
After care program info

dbo_tblInmTestScore/Hist (19)
Correctional plan info

dbo_tblRecmdPrgm/Hist (19)
Program recommendation

Relationship diagram:

AOPC
Data Elements: 117
Contains offense and court information. All data kept in the "Court Case Management Systems."

DOC
Data Elements: 340
Contains inmate and facility information. There are 11 possible "Element Location" values.

PBPP
Data Elements: 176
Contains parole and treatment information. All data kept in either "PBPP Data Warehouse" or "LSIR DB."

PCCD
Data Elements: 79
Contains drug and alcohol use related information. All data kept in the "D&A RIP D3P System."

PCS
Data Elements: 398
Contains summary criminal history and probation information. All data kept in the "SGS Web Data Set."

Relationship key labels: SID, state_id_num, HReSidNo

Incident Tracking Number (TRN) and Incident Tracking Number Suffix (TRS) are used as the keys for linking charges from arrest through adjudication.

H.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 28); 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.

H.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[TM] 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

Figure 27: Texas CCH database model
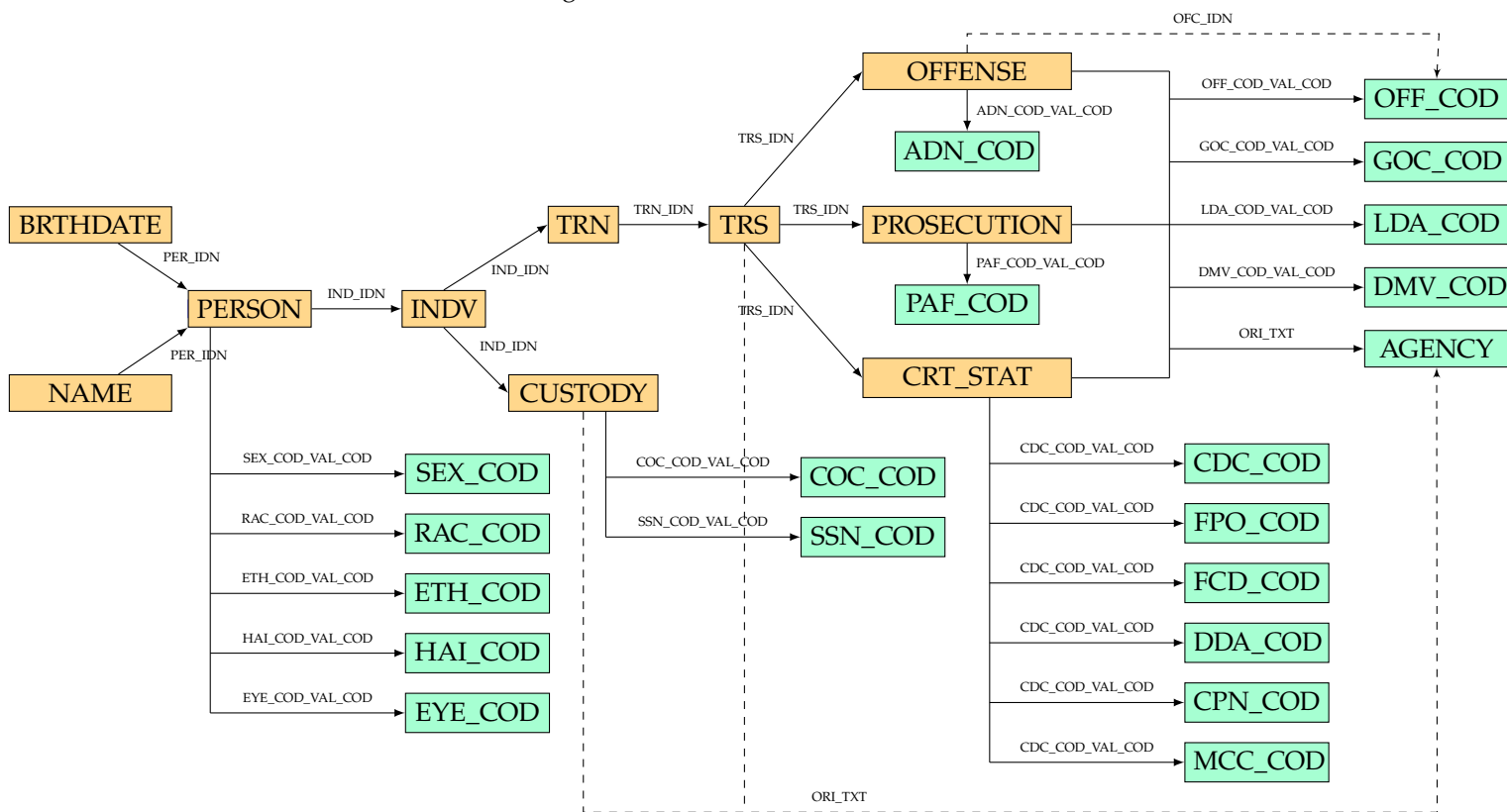
Automated Fingerprint Identification System (AFIS), the Sex Offender Registry (SOR) system, and the Department of Corrections (DOC) system (see Figure 29). The fields DOC Supervision Status and Supervising Officer in the CCH system are updated every night from the DOC system.

H.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 30 is an estimate of how the information in these files could be connected.

Figure 28: Utah database model (presumed)
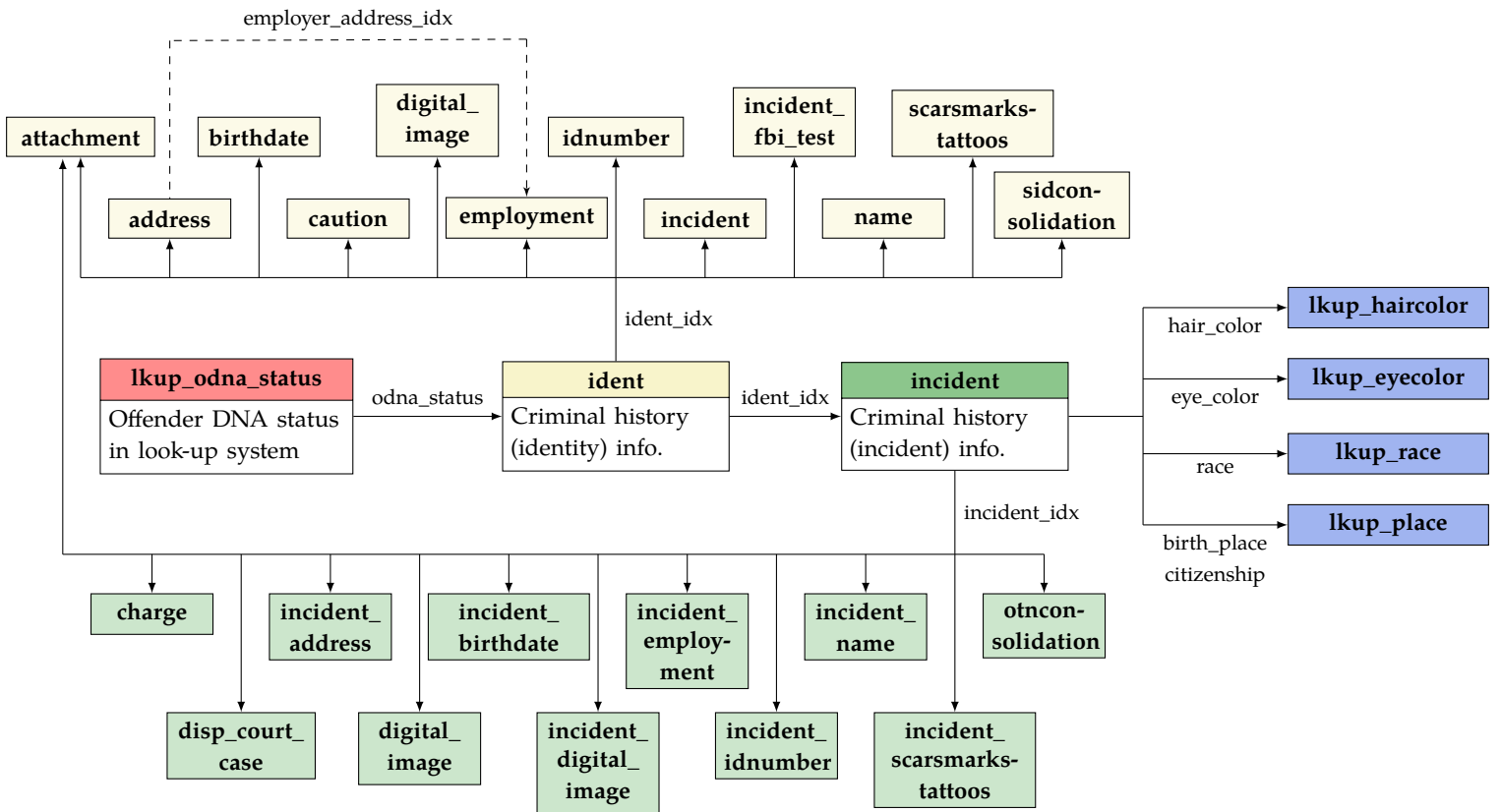
Figure 29: Vermont database model (presumed)

212
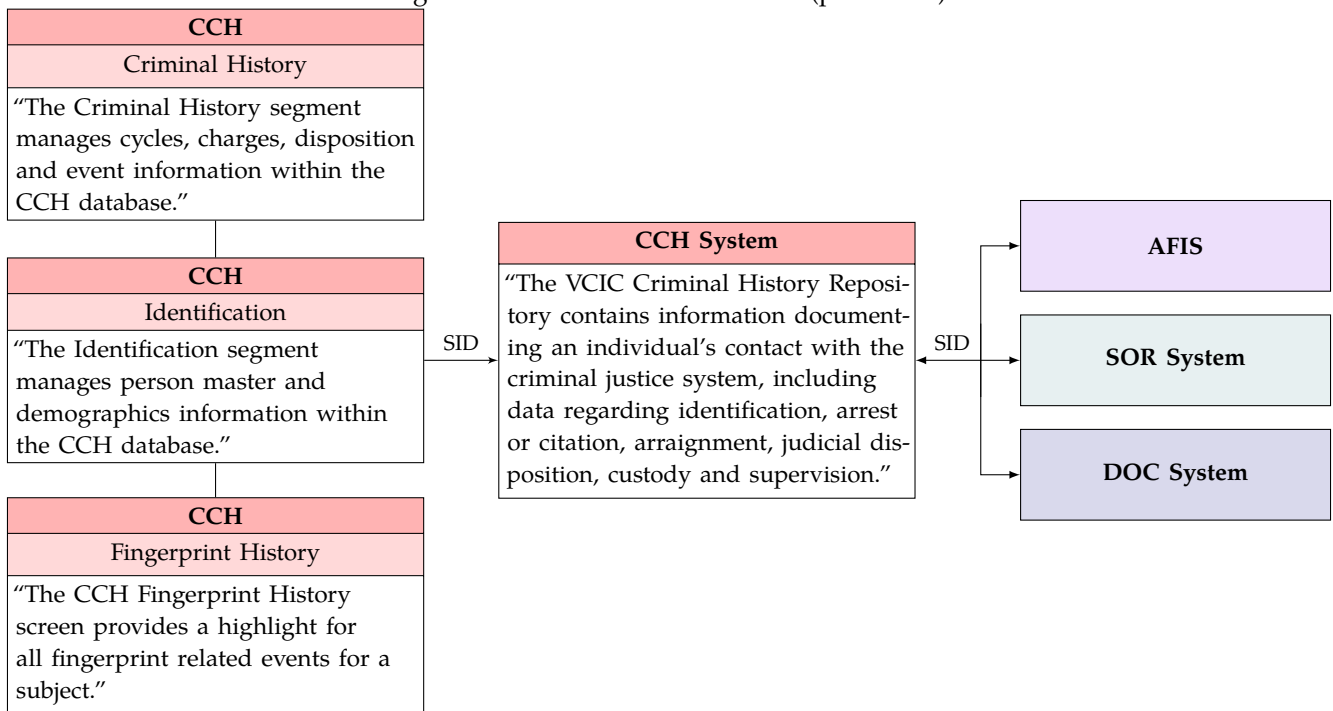
Figure 30: Washington data organization (presumed)
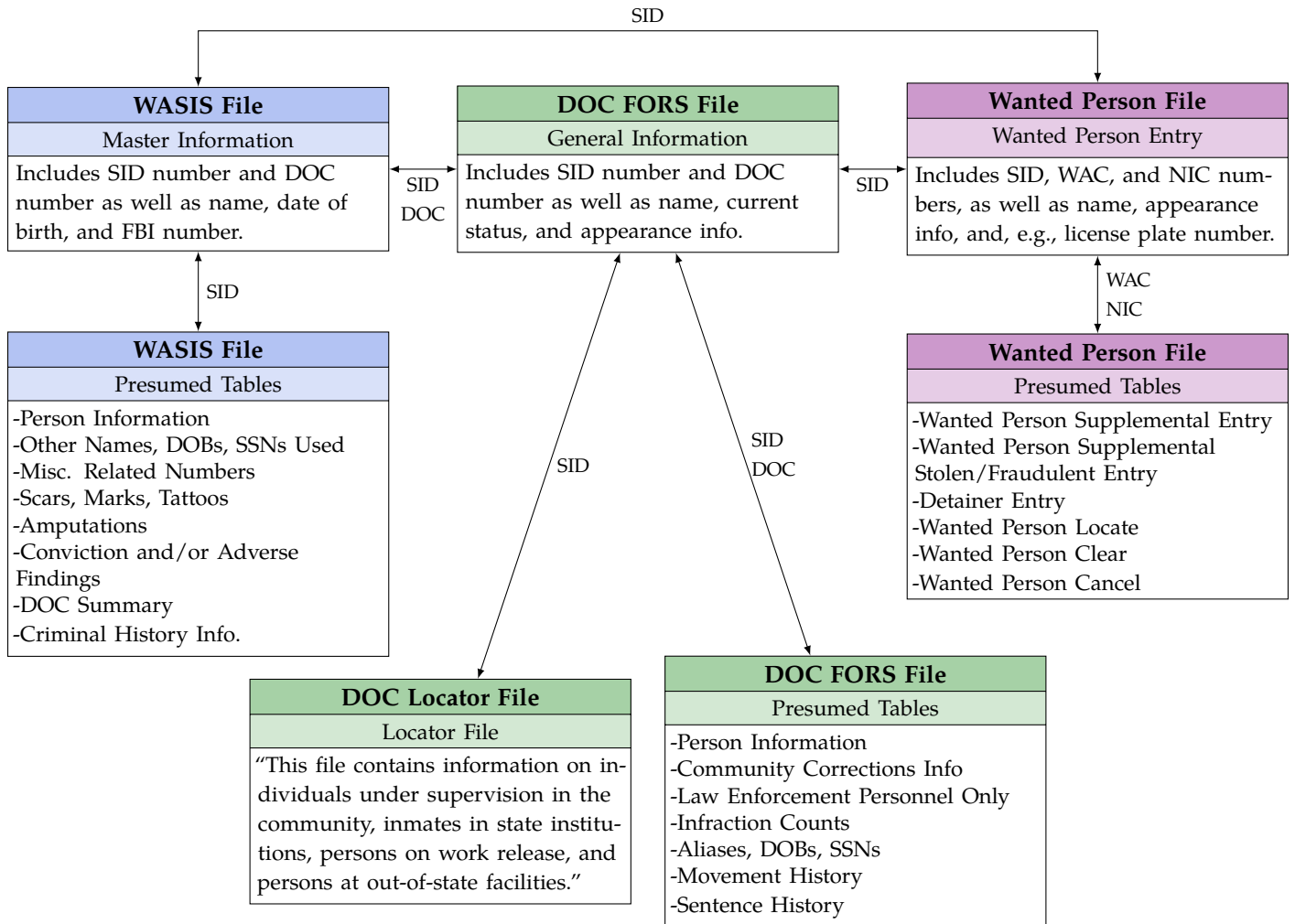
Bibliography

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. 2021. 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

Index