Criminal Justice Administrative Records System (CJARS)

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

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

2020-11-12

Abstract

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

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

The project website is https://cjars.isr.umich.edu.

Data users who have questions about the nature and use of CJARS data may 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

Contents

1 Project description **8**
- 1.1 Project description and objectives . . . 8
- 1.2 Data collection . . . 8
- 1.3 Project scope . . . 8
- 1.4 Data access . . . 9
- 1.5 Data privacy . . . 10
- 1.6 Principal investigators . . . 10
- 1.7 Project funding . . . 10
- 1.8 Project institutional review board compliance . . . 10
- 1.9 Acknowledgments . . . 11
- 1.10 Data citation . . . 11
- 1.11 Documentation structure . . . 11

2 Data acquisition and integration **12**
- 2.1 Mechanisms used to acquire criminal justice administrative data . . . 12
- 2.2 Sources of criminal justice data . . . 12
- 2.3 Criminal justice data integration and harmonization process . . . 12
- 2.4 Integration of criminal justice and Census Bureau records . . . 13

3 Data **15**
- 3.1 General description of the data . . . 15
- 3.2 Data linkage . . . 15
  - 3.2.1 Person-level linkage using the `cjars_id` . . . 15
  - 3.2.2 Data linkage via episode stage unique identifiers . . . 16
- 3.3 Description of probabilistic entity resolution algorithms . . . 16
  - 3.3.1 Entity resolution algorithms . . . 17
  - 3.3.2 Episode resolution algorithm . . . 17
- 3.4 Availability of data by state . . . 18
- 3.5 Variables by relational table . . . 22
- 3.6 Variable codebook . . . 24
  - 3.6.1 Roster . . . 25
  - 3.6.2 Arrest and booking . . . 31
  - 3.6.3 Adjudication . . . 46
  - 3.6.4 Probation . . . 86
  - 3.6.5 Incarceration . . . 103
  - 3.6.6 Parole . . . 121
- 3.7 ID variables used to link tables . . . 135

Appendices **137**

A Data sources **138**

B Variable harmonization **142**
- B.1 Process . . . 142

C Offense classification **146**

D Code schemes including offense classifications **147**
- D.1 Geographic and demographic codes . . . 147
  - D.1.1 State FIPS and abbreviations . . . 147
  - D.1.2 County FIPS . . . 148
  - D.1.3 Sex codes . . . 148
  - D.1.4 Race and ethnicity codes . . . 149

D.2 Criminal justice event codes . . . 150
D.2.1 Offense classification . . . 150
D.2.2 Offense charge grade . . . 160
D.2.3 Offense legal code . . . 160
D.2.4 Court disposition . . . 160
D.2.5 Probation conditions . . . 161
D.2.6 Probation exit . . . 161
D.2.7 Incarceration entry . . . 161
D.2.8 Incarceration facility type . . . 162
D.2.9 Incarceration exit . . . 162
D.2.10 Parole exit . . . 163

E Notes on record linkage **164**
E.1 Entity resolution to identify unique individuals . . . 164
E.2 Episode resolution to link sequences of events . . . 164

F Record linkage at the U.S. Census Bureau **166**
F.1 Record linkage rates . . . 166
F.2 Making the roster unique . . . 166
F.3 Proposal development . . . 167

G Data notes by jurisdiction **168**
G.1 National data notes . . . 168
G.2 Arizona data notes . . . 170
G.3 Arkansas data notes . . . 171
G.4 California data notes . . . 172
G.5 Colorado data notes . . . 173
G.6 Florida data notes . . . 173
G.7 Illinois data notes . . . 174
G.8 Kansas data notes . . . 174
G.9 Maryland data notes . . . 175
G.10 Michigan data notes . . . 175
G.11 Minnesota data notes . . . 175
G.12 Mississippi data notes . . . 176
G.13 Nebraska data notes . . . 176
G.14 New Jersey data notes . . . 177
G.15 North Carolina data notes . . . 178
G.16 Ohio data notes . . . 179
G.17 Oregon data notes . . . 179
G.18 Pennsylvania data notes . . . 179
G.19 Texas data notes . . . 180
G.20 Washington data notes . . . 186
G.21 Wisconsin data notes . . . 186

H State computerized criminal history systems **187**
H.1 Arizona . . . 187
H.2 California . . . 187
H.3 Connecticut . . . 187
H.4 Florida . . . 189
H.5 Illinois . . . 189
H.6 Kansas . . . 190
H.7 Kentucky . . . 191
H.8 Minnesota . . . 193
H.9 Nevada . . . 193
H.10 New Mexico . . . 194

H.11 New York . . . 195
H.12 North Carolina . . . 196
H.13 Oregon . . . 197
H.14 Pennsylvania . . . 197
H.15 Texas . . . 198
H.16 Utah . . . 199
H.17 Vermont . . . 199
H.18 Washington . . . 200

Bibliography **203**

Index **206**

List of Tables

1 Project change log . . . 2
2 CJARS records acquired by acquisition method . . . 12
3 Variables to link tables . . . 135
4 Agencies whose data have been integrated into CJARS . . . 138
5 Variable harmonization . . . 142
6 State FIPS codes and abbreviations . . . 147
7 Sex codes . . . 148
8 Race and ethnicity codes . . . 149
9 Offense classification scheme . . . 150
10 Charge grade classification scheme . . . 160
11 Charge legal code classification scheme . . . 160
12 Court disposition classification scheme . . . 160
13 Probation conditions classification scheme . . . 161
14 Probation exit classification scheme . . . 161
15 Incarceration entry classification scheme . . . 161
16 Incarceration facility type classification scheme . . . 162
17 Incarceration exit classification scheme . . . 162
18 Parole exit classification scheme . . . 163

List of Figures

1 Map of CJARS geographic and criminal justice domain coverage . . . 9
2 CJARS events . . . 9
3 CJARS data exchange and processing . . . 13
4 CJARS national schema . . . 15
5 Person-level linkage with the `cjars_id` . . . 16
6 Data linkage via unique identifiers of episodes . . . 17
7 Estimated CJARS statewide coverage . . . 19
8 CJARS variable harmonization process . . . 145
9 Record linkage rate by number of records per `cjars_id` . . . 166
10 Arizona ACCH & DOC AIMS database model (presumed) . . . 188
11 Description of some tables from California ACHS . . . 188
12 Connecticut CCH database model . . . 189
13 Florida DOC Public Database ERD (presumed) . . . 190
14 Illinois CHRI Ad Hoc database model (presumed) . . . 191
15 Kansas Disposition Report (KDR) Class Diagrams (presumed) . . . 192
16 Kentucky CCH database model (presumed) . . . 192
17 Minnesota Criminal History System (abridged/presumed) . . . 193
18 Nevada CCH database model (presumed) . . . 194
19 New Mexico ERD (abridged) . . . 195
20 New York CCH database model (presumed) . . . 196
21 North Carolina ACIS database model (LIN link assumed) . . . 197
22 Oregon eCourt (Odyssey) ERD . . . 198
23 Pennsylvania database relationship (presumed) . . . 199
24 Texas CCH database model . . . 200
25 Utah database model (presumed) . . . 201
26 Vermont database model (presumed) . . . 201
27 Washington data organization (presumed) . . . 202

1 Project description

1.1 Project description and objectives

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

CJARS has the following main objectives:

- Create a dataset that follows criminal cases across criminal justice agencies, from arrest through discharge from the justice system.
- Capture the progression of criminal activities over the life course by constructing longitudinal records of contact that individuals have with the justice system.
- Provide a platform with nationwide scalability to facilitate research in perpetuity through the Federal Statistical Research Data Center (FSRDC) network.
- Integrate criminal justice data with existing Census Bureau individual-level surveys and administrative records to facilitate novel research not previously possible.

1.2 Data collection

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

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

1.3 Project scope

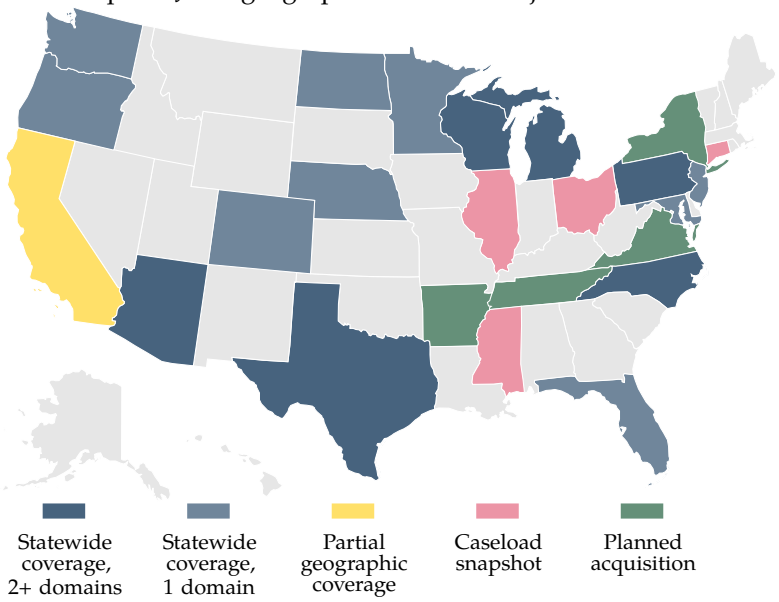
Variation also exists across states and agencies in the way that data is collected and stored. For instance, some state court agencies maintain a statewide centralized repository where all data on events that occur under the state's jurisdiction are stored. Conversely, some states have a less centralized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

CJARS has collected data in a number of states, which can be seen in Figure 1. The map delineates (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 41.5% of the U.S. population.

It is important to note that data coverage varies substantially from state to state. Specifically, there are differences in coverage of event type, time frame, and geography. More detailed information on coverage within states will be described in later sections.

Finally, it is worth noting the types of criminal justice events that are captured by the CJARS data infrastructure. These include: records of arrests, criminal court case filings, and terms of probation, prison, and parole. Other

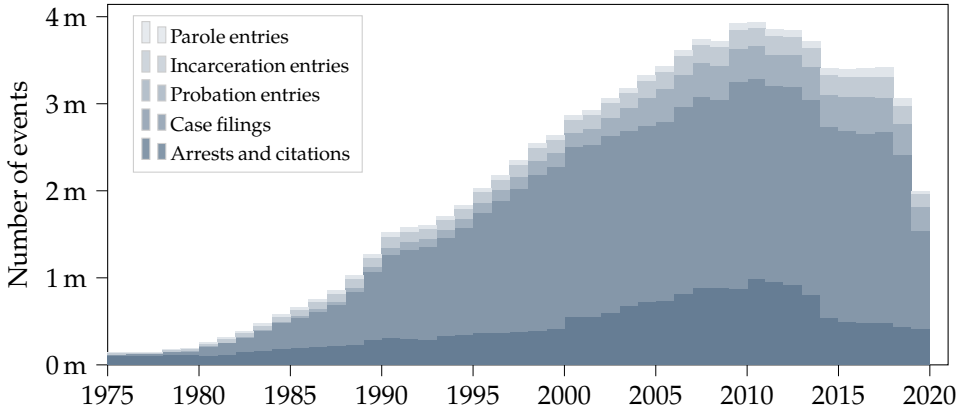
Figure 1: Map of CJARS geographic and criminal justice domain coverage



events that are processed through similar institutions (i.e., courts) but do not fall under the purview of a criminal justice event are excluded from the scope of the project (e.g., civil court filings). Juvenile records are also considered out of scope.

Figure 2 shows the number of records collected by the CJARS project, by year and by criminal justice event type.

Figure 2: Events in the CJARS database, by year and event type



1.4 Data access

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

Distribution through the FSRDCs provides an ideal outlet. The FSRDC network is broadly available to researchers and provides excellent scientific computing resources, while maintaining the highest levels of information security and confidentiality protection. Researchers working in the FSRDCs have no access to sensitive PII, and can only request release of aggregate statistical material. The FSRDCs allow researchers to link CJARS records at the person-level to a wealth of socioeconomic survey and administrative data held

by the Census Bureau. For more information on the FSRDC system and other available data, please visit: https://www.census.gov/fsrdc.

1.5 Data privacy

CJARS goes to great lengths to ensure the protection and security of its data, so that the identities and characteristics of all individuals in CJARS records are kept confidential.

The University of Michigan processes all of the criminal justice records that it collects on a data system that was built to be compliant with FBI Criminal Justice Information Services (CJIS) standards. The Michigan State Police 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, received data are accessible only by a small staff responsible for inventorying the contents of the data and removing sensitive PII. Next, a data linkage team uses a probabilistic linkage process to replace sensitive PII with a unique identifier called a Personal Identification Key (PIK) that can be used to link records to other databases held at the Census Bureau. More information about PIK assignment can be found in Subsection 2.4.

Finally, anonymized CJARS records are moved to secure research servers. There, qualified researchers on approved projects can work in the FSRDCs to link CJARS records with other data held by the Census Bureau. PII is not available for researchers in order to maintain the privacy of individuals.

1.6 Principal investigators

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

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

1.7 Project funding

The CJARS project has been supported by National Science Foundation grant SES-1925563, as well as grants from the Laura and John Arnold Foundation and the Bill and Melinda Gates Foundation. The University of Michigan has supported the CJARS project through the following programs: Michigan Institute for Teaching and Research (MITRE), Populations Studies Center (PSC), and Poverty Solutions.

1.8 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.

1.9 Acknowledgments

This project would not be possible without the hard work of a great group of people. Jordan Papp, Diana Sutton, and Matt Van Eseltine have supported the project as project managers. Jay Choi developed the CJARS web scraping system and a lot of the harmonization processes. Matt Gross developed entity resolution algorithms and harmonized a lot of data. Ben Pyle, James Reeves, and Susan Parker contributed to data harmonization and research as graduate student researchers. As a research assistant, Josh Kim helped develop automated quality control systems. Francis Fiore, David Smith, Neel Shah, and Manav Bhatia turned the web scraping system into an industrial-strength platform. Shaeq Ahmed, Luis Baldomero-Quintana, Madeleine Danes, Kenna Garrison, Lauren Lee, Lyllian Simerly, Ellen Stuart, and Peixin Yang have also supported the project as research assistants. Brittany Street and Elizabeth Luh have furthered CJARS research as postdoctoral research fellows.

Carol Bowen, Kerri Cross, Violet Elder, and Lauren Tingwall have supported the project as research administrators. The following advisors have guided the project as part of the CJARS Board of Directors: Shawn Bushway, Julia Lane, Jens Ludwig, Jim Lynch, Justin McCrary, Jeff Morenoff, Amy O'Hara, Anne Piehl, Steve Raphael, Bill Sabol. The project also received guidance from members of its Scientific Advisory Council: Maggie Levenstein, Jeff Morenoff, and JJ Prescott.

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

Amanda Agan and Ben Grunwald donated data to the project.

1.10 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.

1.11 Documentation structure

- Section 2 provides an overview of how the CJARS project acquires data, and how it is integrated into a unified system at the University of Michigan.
- Section 3 explains the structure of the harmonized data system. Of particular interest is the variable codebook in Subsection 3.6.
- Appendices cover the project in much more detail.
  - Appendix A identifies the agency sources of CJARS administrative data.
  - Appendix B explains the variable harmonization process in detail.
  - Appendix C 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 integration

2.1 Mechanisms used to acquire criminal justice administrative data

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

1. **Data use agreements.** Formal legal documents are signed which guide the obligations of CJARS when data is obtained from an agency.
2. **Public records requests.** A request for public records is filed in jurisdictions where legal statutes allow for such requests.
3. **Web scraping or bulk downloads.** Data that is publicly available online is collected using web scrapers or downloaded in bulk when available. Scraping is only conducted using methods that comply with an agency's terms of use.

Table 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

Acquisition source	Records	Individuals
Data use agreements	1,556,831,164	18,646,241
Public records requests	276,629,240	10,797,241
Web scraping or bulk downloads	342,130,380	3,712,069

2.2 Sources of criminal justice data

Data is collected within many separate parts of the justice system. For example, incarceration records are typically held and managed by a state's Department of Corrections. Conversely, arrest records are often held and managed by local police departments. To provide researchers with a clear understanding of the typical sources of CJARS data, the types of events held in this repository are described below.

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

The sources of all data held within the CJARS repository can be found in Appendix A. Data has been collected by a mixture of state, county, and local agencies. In addition, some data has also been collected by a third party with access to data from an agency.

2.3 Criminal justice data integration and harmonization process

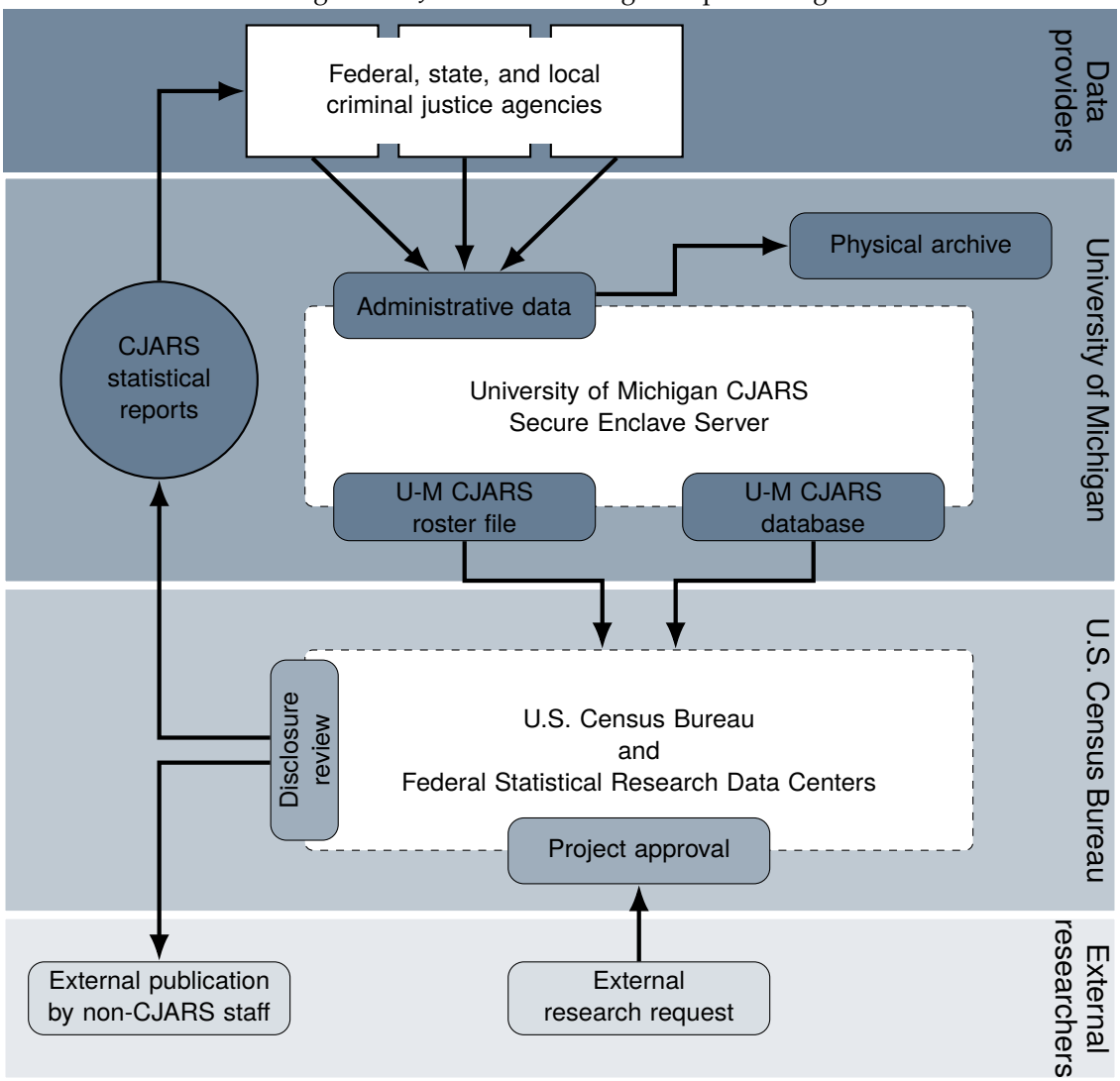
One of the major barriers to research on the criminal justice system is a lack of integration of records across agencies. For example, arrest records are rarely linkable to information that occurs at later points in the system (e.g., incarceration). CJARS implements the following steps in order to create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

- **Data requests and collection**. CJARS collects data using the three-pronged approach that is described above. All data and other relevant information describing the data are kept in a secure data enclave.
- **Data cleaning**. Original data are cleaned and harmonized to fit a common CJARS schema (described in more detail in later sections).
- **Merge**. The harmonized data are appended and merged with other criminal justice records in CJARS to create a single dataset spanning agencies and jurisdictions.

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

Figure 3 provides a visual representation of the data integration process. This figure shows how data that is collected from data providers (first panel) is added to the UM CJARS Repository (second panel). Then once the data is processed, the data are integrated in the Census Bureau (third panel). Finally, once processed by the Census Bureau, the data are made available to external researchers (fourth panel).

Figure 3: CJARS data exchange and processing



2.4 Integration of criminal justice and Census Bureau records

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

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

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

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

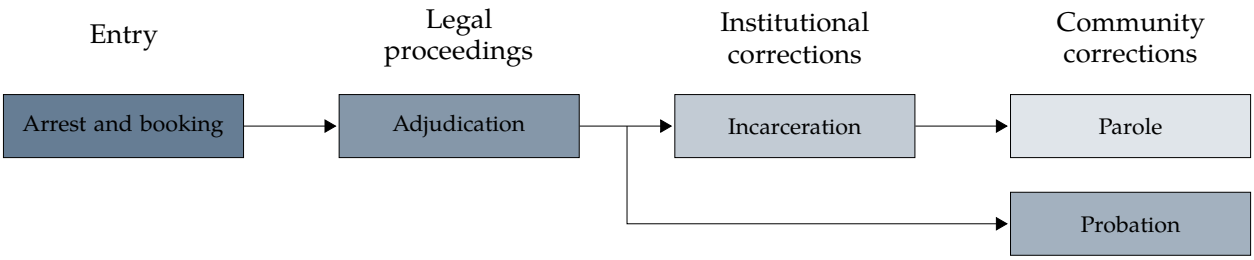
- Multiple PIKs have been linked to a single `cjars_id`. In this case, there is sufficient variation in the PII within a `cjars_id`, but across criminal justice records that Person Identification Validation System (PVS) has assigned different registry identities to a single `cjars_id`. This could be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

3 Data

3.1 General description of the data

The goal 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 4. The arrows in this figure depict the expected progression of events that make up an episode through the system.

Figure 4: CJARS national schema



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

3.2 Data linkage

A major barrier to research on the criminal justice system is a lack of integration of data across agencies. For example, unique individual identifiers are often not present that allow linking records across different agencies to an individual. In addition, there are usually no identifiers to indicate which records belong to a single criminal justice episode. CJARS has addressed these issues by leveraging machine learning models that probabilistically match records to individuals and events to episodes. The following subsections describe more about the methods used to develop these linkages and the linkages themselves.

3.2.1 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.

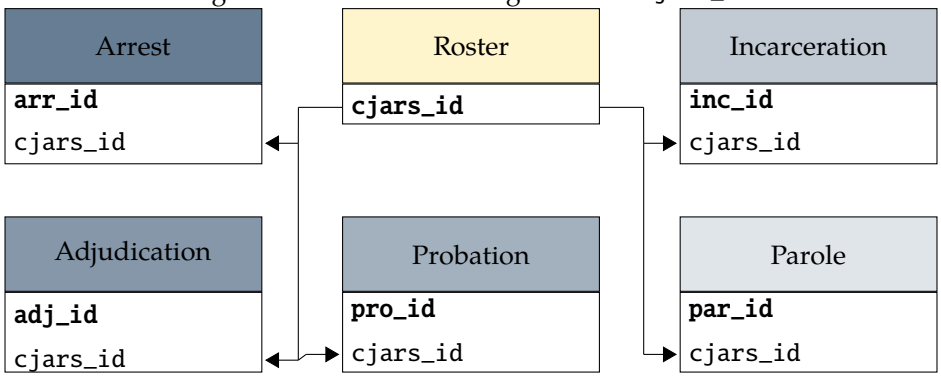
Assignment of a `cjars_id` allows researchers to link all records in the CJARS database that belong to an individual, to that individual. This data linkage facilitates analyzing the data to determine how many events are seen for an individual, how many offense he or she has committed, recidivism, etc.

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

The CJARS data is comprised of 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 4. 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 5, 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 5: 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.

3.2.2 Data linkage via episode stage unique identifiers

> USER NOTE
>
> The CJARS team has not yet completed its work on episode resolution, our algorithms for linking together related processes associated with a single criminal offense. We expect to complete this work in 2021. Users can expect to see linked processes in the RDC data warehouse in late 2021. Therefore, topical relational tables may contain primary keys but will not contain the foreign keys described below.

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.

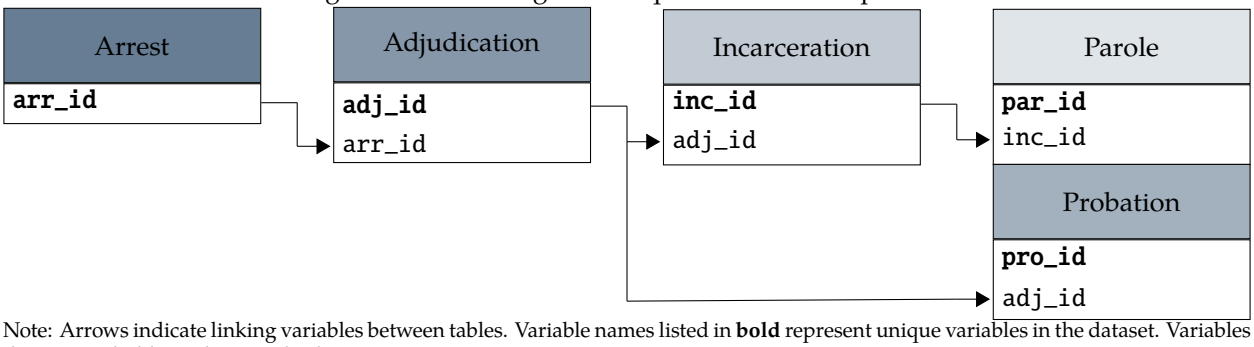
Figure 6 summarizes how the data linkage is structured based on unique identifiers of events in each CJARS relational database. Each CJARS relational database contains an identifier for that respective type of event contained in the table, but also an identifier that can be used to link to the event 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.

3.3 Description of probabilistic entity resolution algorithms

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

Figure 6: Data linkage via unique identifiers of episodes



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

3.3.1 Entity resolution algorithms

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

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

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

3.3.2 Episode resolution algorithm

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

3.4 Availability of data by state

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

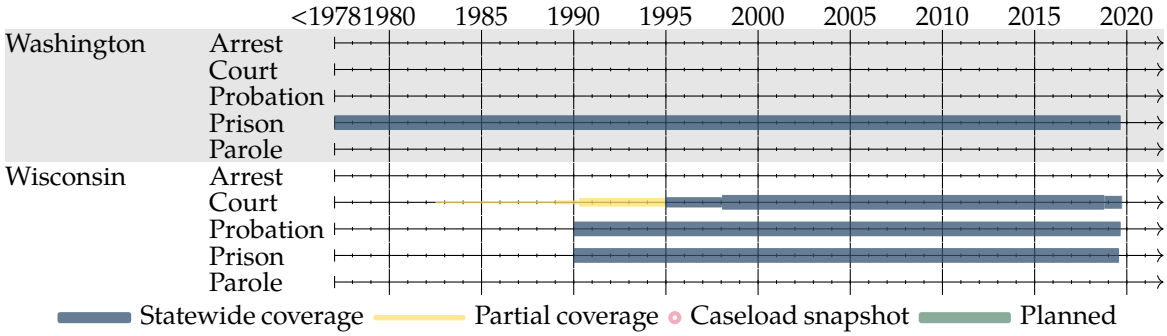
Table 7 summarizes the availability of data by state (where statewide coverage is available), time frame, and event type. In this table, markings designate coverage of criminal justice domains in each state as one of: statewide coverage, partial geographic coverage in a state, "caseload snapshots" of all offenders under supervision at a single point in time, and planned coverage. Please see the legend for further detail.

Note that Table 7 only provides a general overview of CJARS data holdings. More details about the data in each jurisdiction can be found in Appendix A.

Figure 7: Estimated CJARS statewide coverage

Figure 7: CJARS statewide coverage (cont'd)

Figure 7: CJARS statewide coverage (cont'd)



3.5 Variables by relational table

The CJARS relational databases contain detailed information about arrests and bookings, adjudications, incarcerations, and terms of parole and probation. 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
  - `race`: Race and ethnicity
- **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

3.6 Variable codebook

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

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

3.6.1 Roster

USER NOTE

The CJARS roster is delivered to the Census Bureau with duplicate rows containing distinct PII to maximize record linkage success. In order to make the roster unique by `cjars_id`, please use the algorithm described in Section F.2.

USER NOTE

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

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 5. 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	113,334,609	100.0
Unique values	40,429,123	

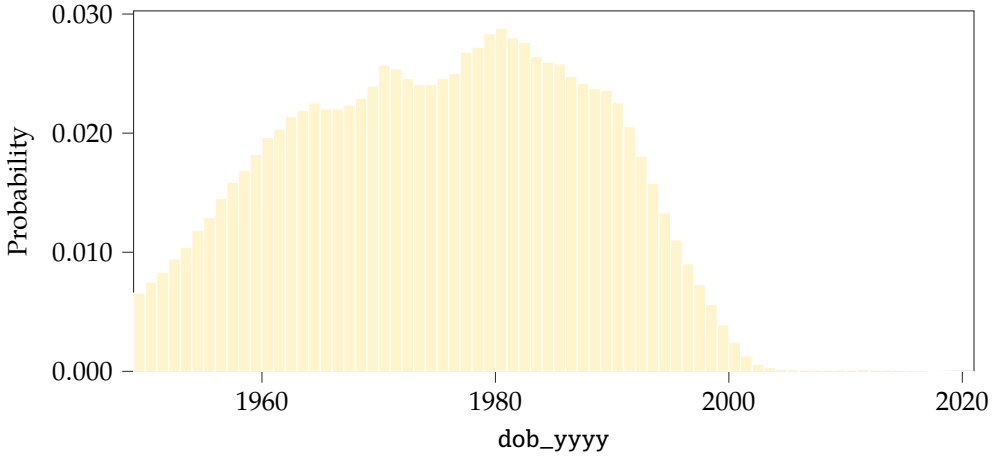
dob_yyyy

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

Set	*Count*	*Percent (%)*
All records	113,334,609	100.0
Valid records	96,757,652	85.4
Invalid values	7,528	0.0
Missing values	16,569,429	14.6

Statistic	*Value*
Mean	1,973.2
Minimum	1,900
25th percentile	1,963
Median	1,974
75th percentile	1,984
Maximum	2,020

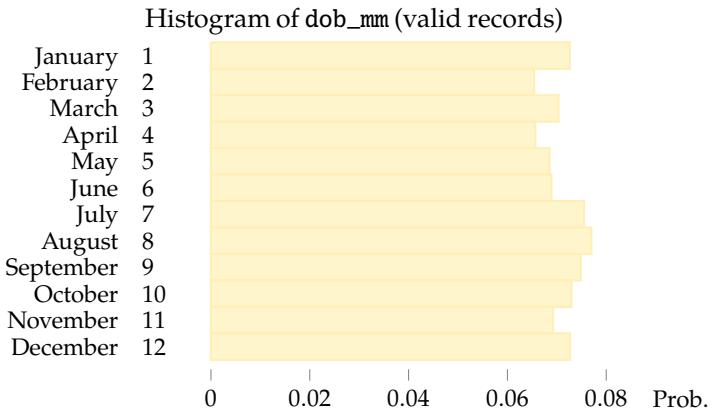
Histogram of dob_yyyy (above 5th percentile of valid records)



`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

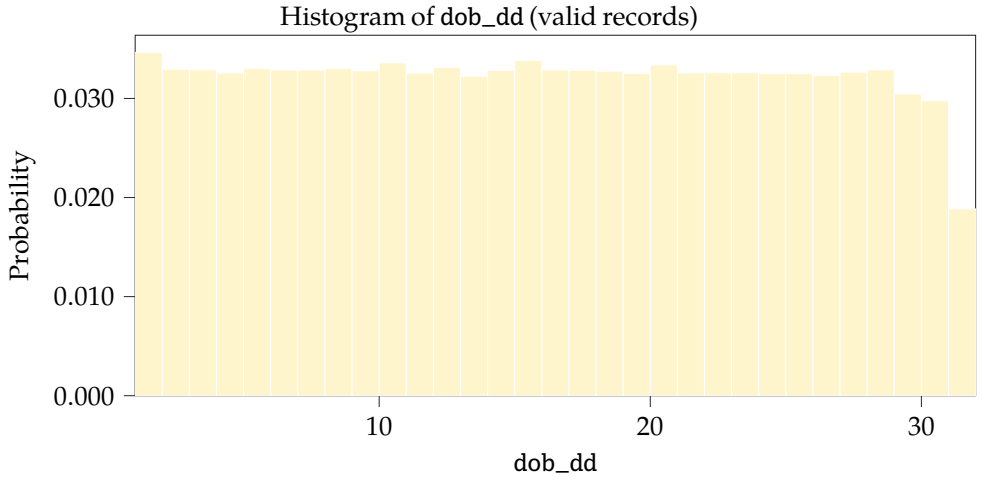
Set	*Count*	*Percent (%)*
All records	113,334,609	100.0
Valid records	96,753,310	85.4
Missing values	16,581,299	14.6



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	113,334,609	100.0
Valid records	96,753,310	85.4
Missing values	16,581,299	14.6

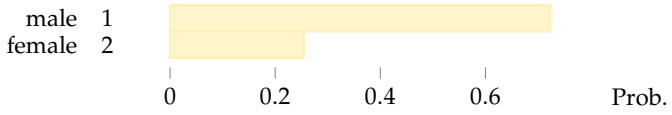


sex

Label	Sex
Description	The individual's sex.
Table	Roster
Format	numeric
Code scheme	sex code

Set	*Count*	*Percent (%)*
All records	113,334,609	100.0
Valid records	110,844,667	97.8
Missing values	2,489,942	2.2

Histogram of **sex** (valid records)

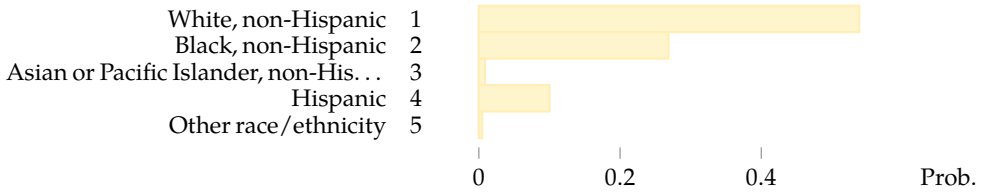


`race`

Label	Race and ethnicity
Description	A composite race and ethnicity variable created by CJARS and based on race and ethnicity in the source data if available, as well as on name and location.
Table	Roster
Format	numeric
Code scheme	race and ethnicity code

Set	*Count*	*Percent (%)*
All records	113,334,609	100.0
Valid records	104,384,748	92.1
Invalid values	700,052	0.6
Missing values	8,249,809	7.3

Histogram of `race` (valid records)



3.6.2 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: `twd`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 5.
Table	Arrest and booking
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Unique values	8,433,052	

`arr_id`

Label	Arrest identifier
Description	Uniquely identifies arrest. For more details on use of `arr_id` for data linkage, refer to Section 3.2.2 and Figure 6.
Table	Arrest and booking
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Unique values	32,504,351	

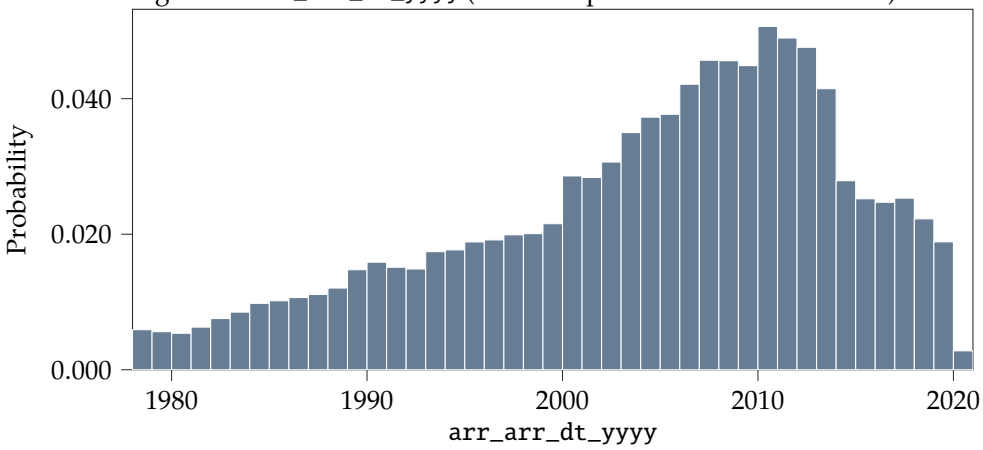
arr_arr_dt_yyyy

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

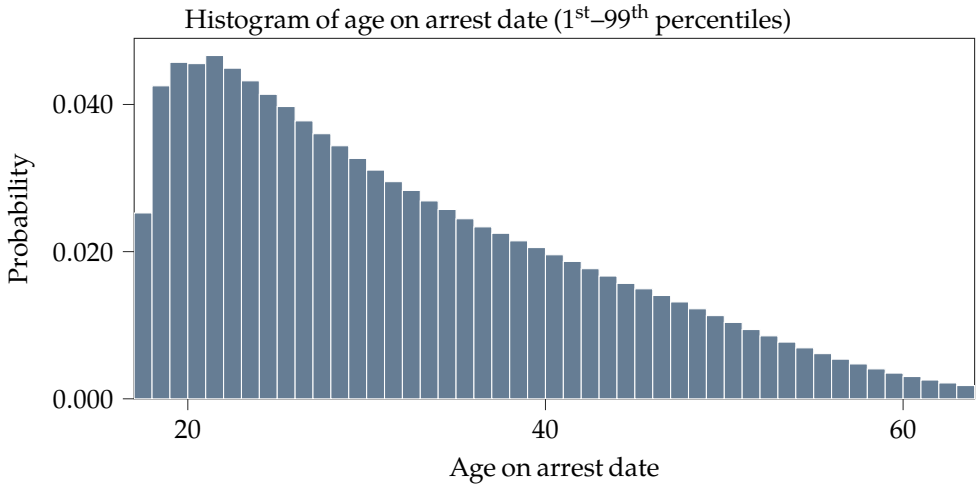
Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	20,402,688	62.8
Invalid values	2	0.0
Missing values	12,101,661	37.2

Statistic	*Value*
Mean	2,002.2
Minimum	1,901
25th percentile	1,996
Median	2,005
75th percentile	2,011
Maximum	2,020

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



arr_arr_dt_yyyy (continued)



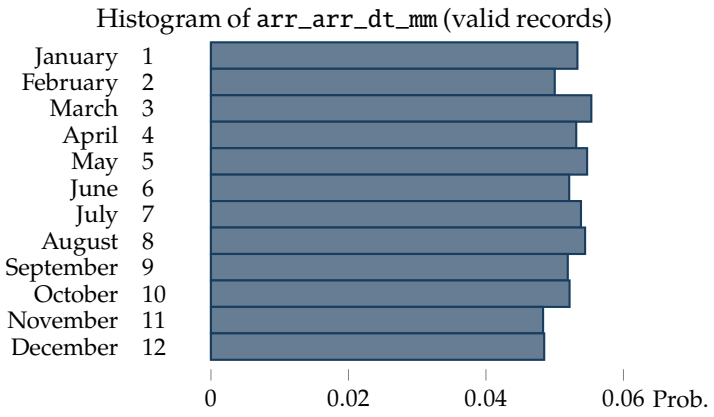
Data notes

- California: t6z, twf, lao, rsq, f41
- Colorado: wt8
- Florida: fdw
- Kansas: pya
- Texas: qfd, bgp, j6j, acs

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	32,504,351	100.0
Valid records	20,402,690	62.8
Missing values	12,101,661	37.2



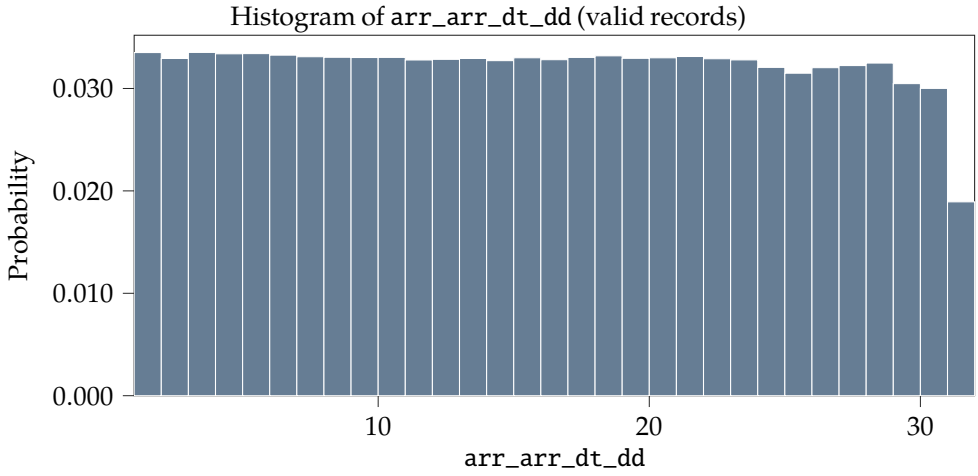
Data notes

- California: t6z, twf, lao, rsq, f41
- Colorado: wt8
- Florida: fdw
- Kansas: pya
- Texas: qfd, bgp, j6j, acs

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	32,504,351	100.0
Valid records	20,402,690	62.8
Missing values	12,101,661	37.2



Data notes

- California: t6z, twf, lao, rsq, f41
- Colorado: wt8
- Florida: fdw
- Kansas: pya
- Texas: qfd, bgp, j6j, acs

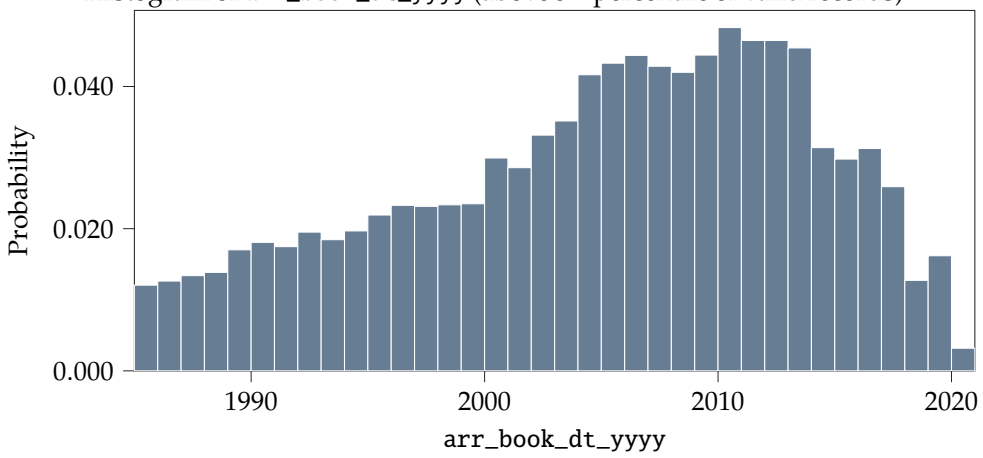
arr_book_dt_yyyy

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

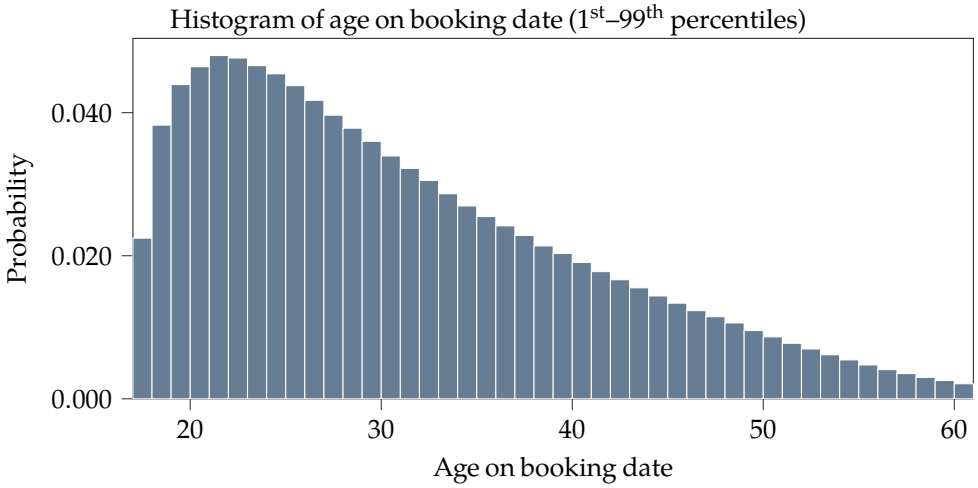
Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	14,825,119	45.6
Missing values	17,679,232	54.4

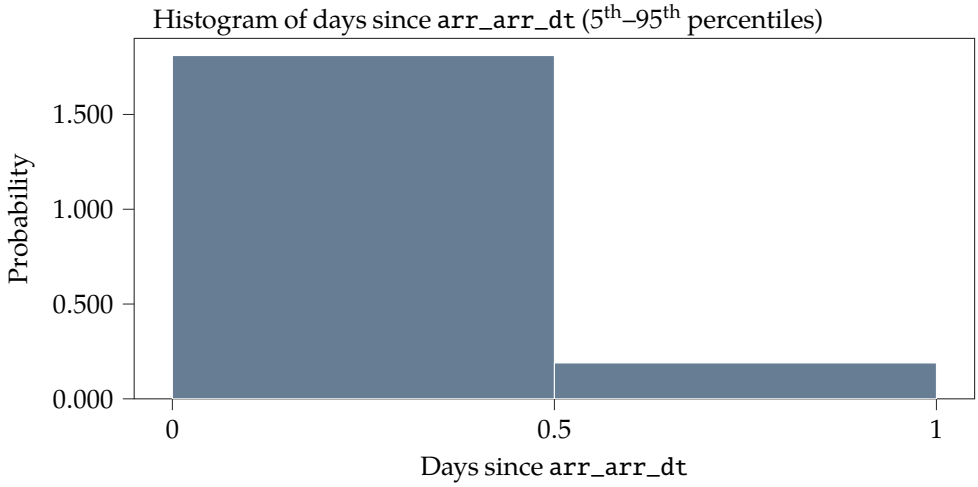
Statistic	*Value*
Mean	2,003.5
Minimum	1,911
25th percentile	1,997
Median	2,005
75th percentile	2,011
Maximum	2,020

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



`arr_book_dt_yyyy` (continued)





Data notes

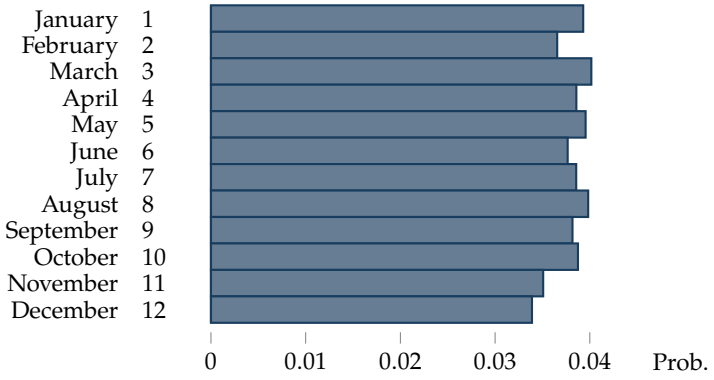
- California: `ldw`, `myf`, `fe2`, `nbh`
- Texas: `yb0`, `pzl`, `w82`, `jsg`, `g5n`, `hd0`, `ce6`, `vbm`, `y69`, `l2h`, `cro`, `e0d`

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	32,504,351	100.0
Valid records	14,825,119	45.6
Missing values	17,679,232	54.4

Histogram of arr_book_dt_mm (valid records)



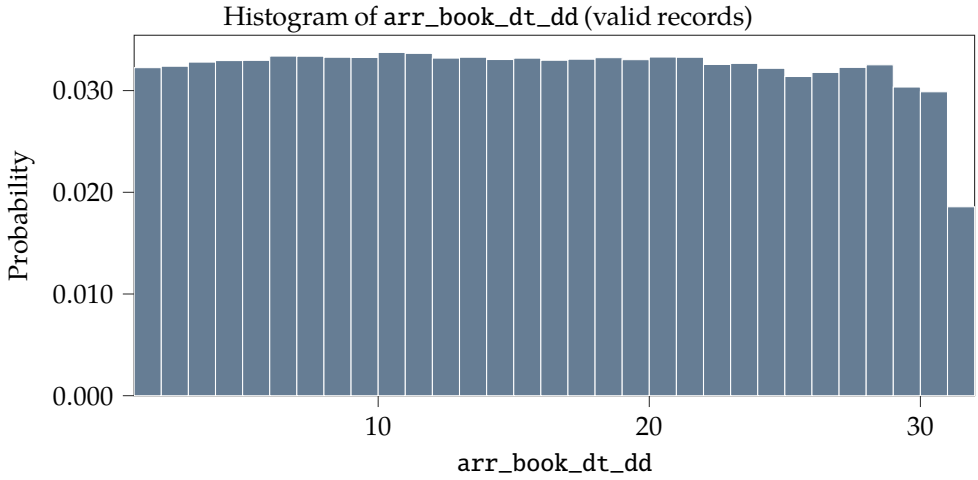
Data notes

- California: ldw, myf, fe2, nbh
- Texas: yb0, pzl, w82, jsg, g5n, hd0, ce6, vbm, y69, l2h, cro, e0d

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	32,504,351	100.0
Valid records	14,825,119	45.6
Missing values	17,679,232	54.4



Data notes

- California: ldw, myf, fe2, nbh
- Texas: yb0, pzl, w82, jsg, g5n, hd0, ce6, vbm, y69, l2h, cro, e0d

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	32,504,351	100.0
Valid records	32,504,351	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

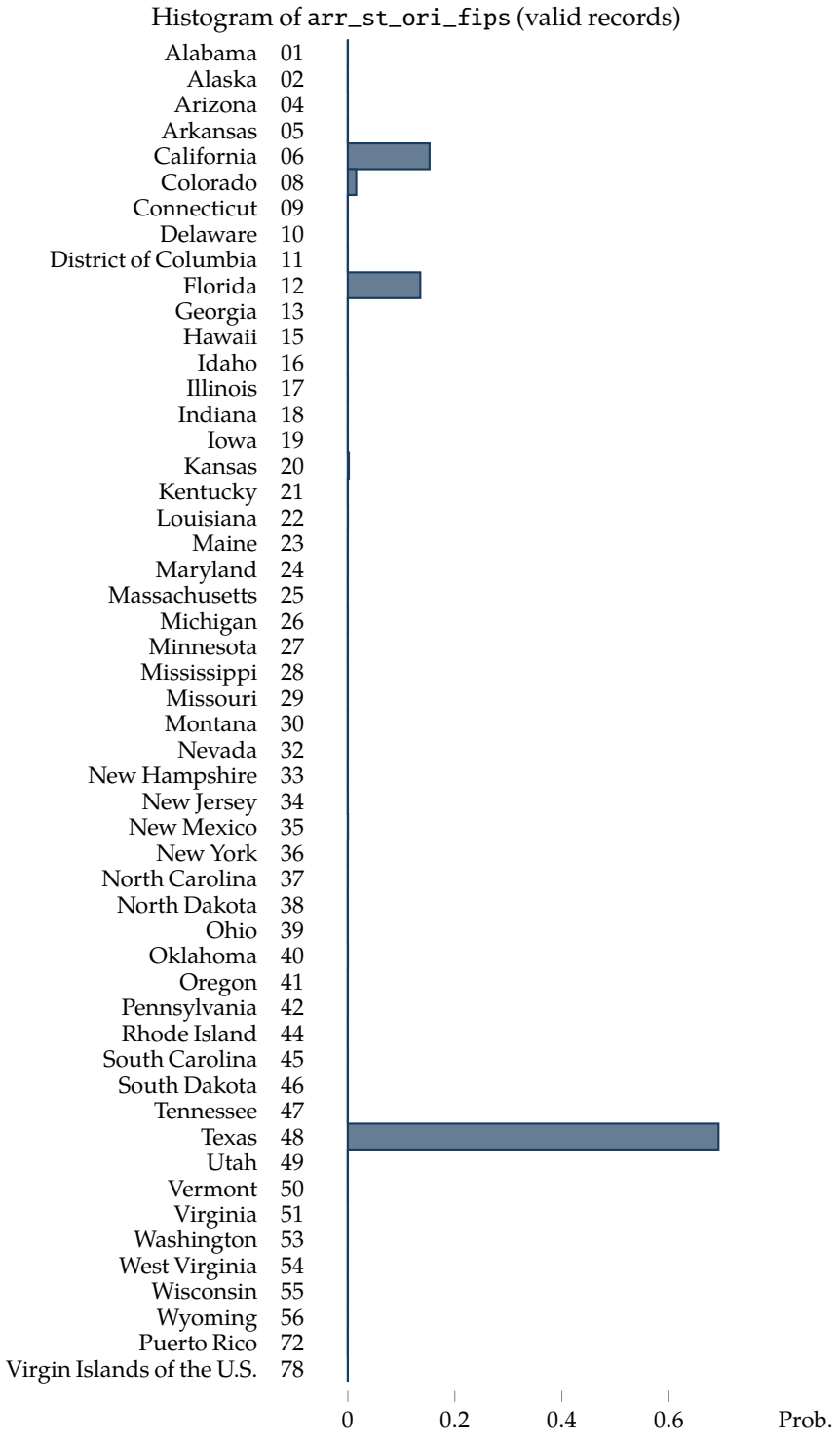
Set	*Count*	*Percent (%)*
All records	32,504,351	100.0

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	32,504,351	100.0
Valid records	32,504,351	100.0

arr_st_ori_fips (continued)



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

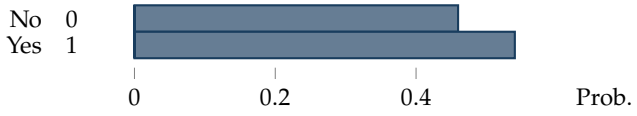
Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	29,654,776	91.2
Invalid values	168,860	0.5
Missing values	2,680,715	8.2

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	32,504,351	100.0
Valid records	32,504,351	100.0

Histogram of arr_rec_src_le (valid records)

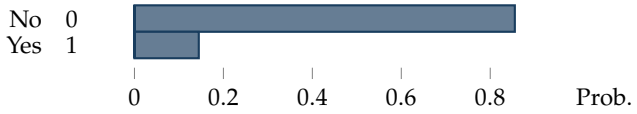


arr_rec_src_crt

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

Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	32,504,351	100.0

Histogram of `arr_rec_src_crt` (valid records)

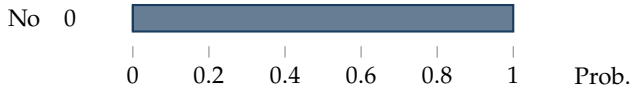


arr_rec_src_doc

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

Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	32,504,351	100.0

Histogram of `arr_rec_src_doc` (valid records)

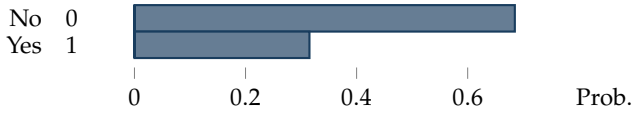


arr_rec_src_rep

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

Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	32,504,351	100.0

Histogram of `arr_rec_src_rep` (valid records)

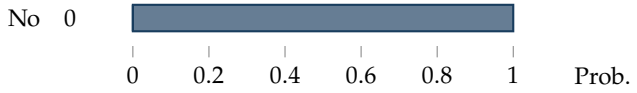


arr_rec_src_cc

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

Set	*Count*	*Percent (%)*
All records	32,504,351	100.0
Valid records	32,504,351	100.0

Histogram of `arr_rec_src_cc` (valid records)



3.6.3 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: `twd`
- New Jersey: `r92`
- North Carolina: `t4z`
- Texas: `lrk`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 5.
Table	Adjudication
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Unique values	23,253,422	

`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 3.2.2 and Figure 6.
Table	Adjudication
Table key	primary, unique
Format	string

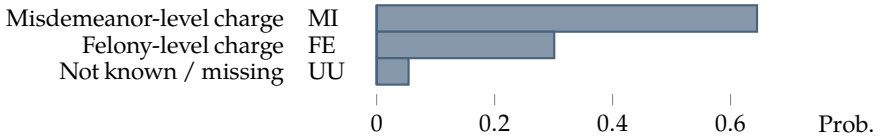
Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Unique values	83,276,047	

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	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of `adj_grd_cd` (valid records)



Data notes

- All states: `b5n`, `fei`
- Nebraska: `rl7`
- New Jersey: `q57`

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	83,276,047	100.0

Data notes

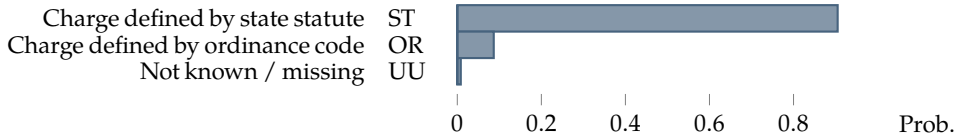
- All states: yik
- Nebraska: rl7
- New Jersey: q57

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	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of `adj_off_lgl_cd` (valid records)



adj_off_lgl_cd_src

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0

Data notes

- All states: `yik`

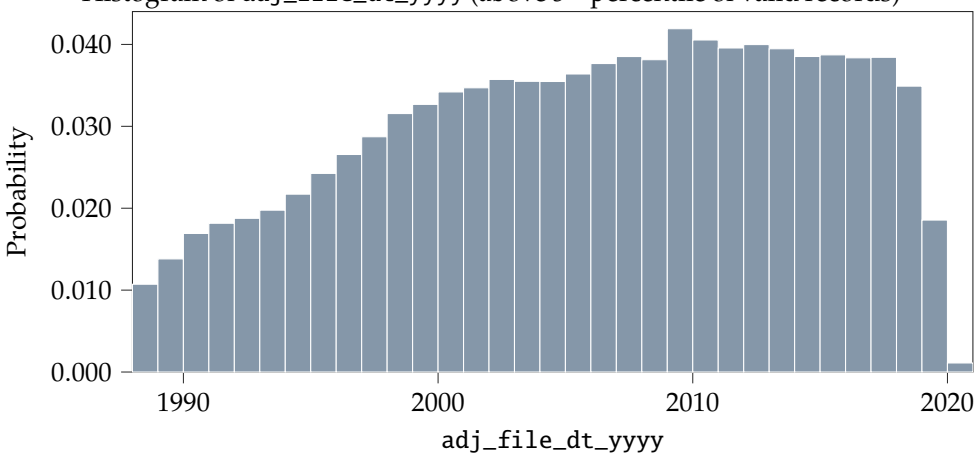
adj_file_dt_yyyy

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

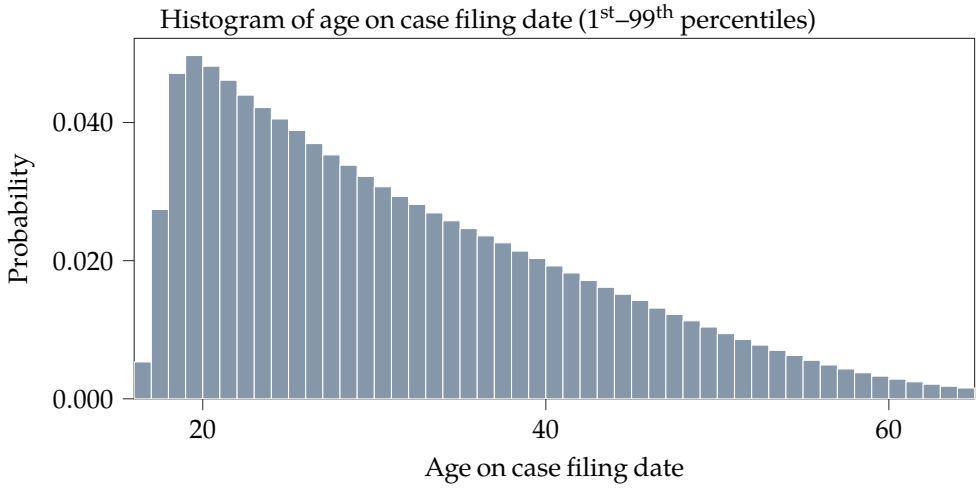
Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	59,185,149	71.1
Invalid values	1,207	0.0
Missing values	24,089,691	28.9

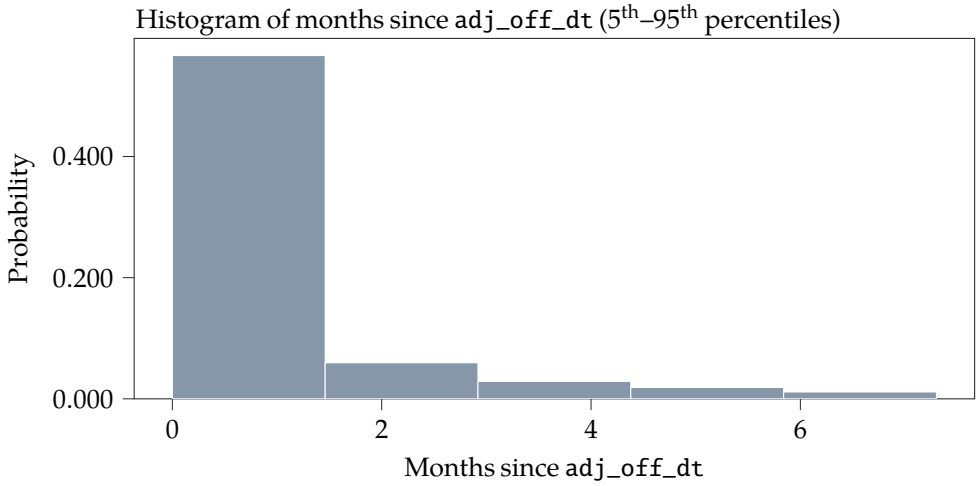
Statistic	*Value*
Mean	2,004.6
Minimum	1,900
25th percentile	1,998
Median	2,006
75th percentile	2,012
Maximum	2,020

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



`adj_file_dt_yyyy` (continued)





Data notes

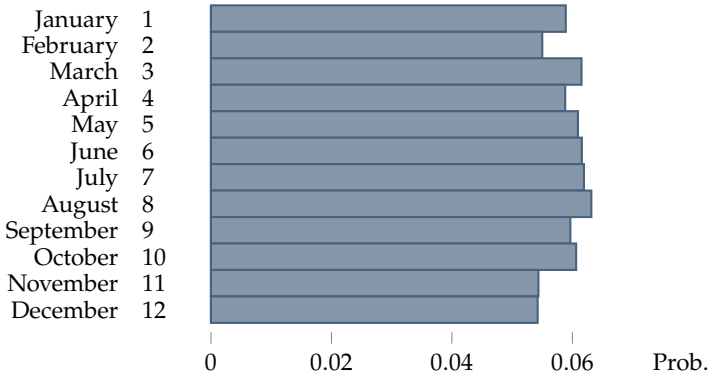
- Arkansas: `f7i`
- Mississippi: `sc5`
- Nebraska: `l8c`
- North Carolina: `xwn`
- Ohio: `xe4`
- Texas: `g5n`, `uab`, `c74`

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	83,276,047	100.0
Valid records	59,186,356	71.1
Missing values	24,089,691	28.9

Histogram of adj_file_dt_mm (valid records)



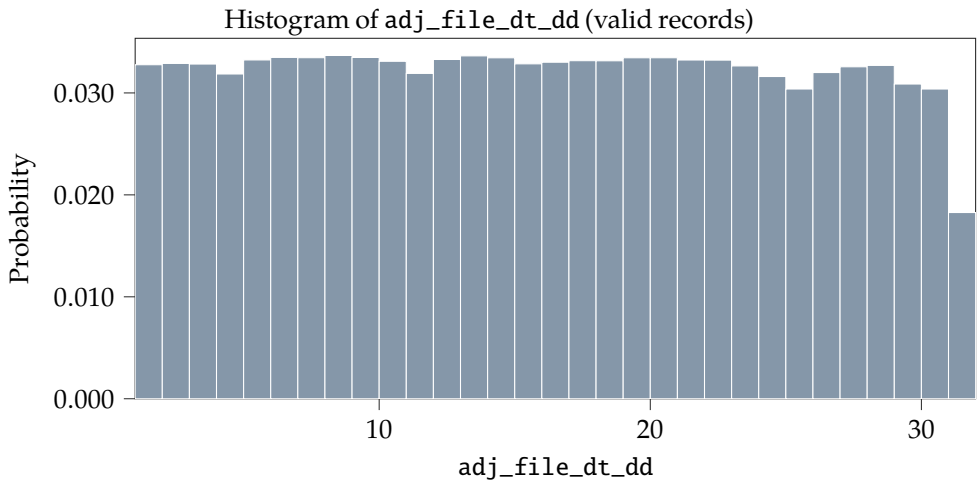
Data notes

- Arkansas: f7i
- Mississippi: sc5
- Nebraska: l8c
- North Carolina: xwn
- Ohio: xe4
- Texas: g5n, uab, c74

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	83,276,047	100.0
Valid records	59,186,356	71.1
Missing values	24,089,691	28.9



Data notes

- Arkansas: f7i
- Mississippi: sc5
- Nebraska: l8c
- North Carolina: xwn
- Ohio: xe4
- Texas: g5n, uab, c74

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	83,276,047	100.0
Valid records	83,276,047	100.0

Data notes

- Texas: `lrk`, `vbm`, `y69`, `c74`, `vsl`

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	83,276,047	100.0

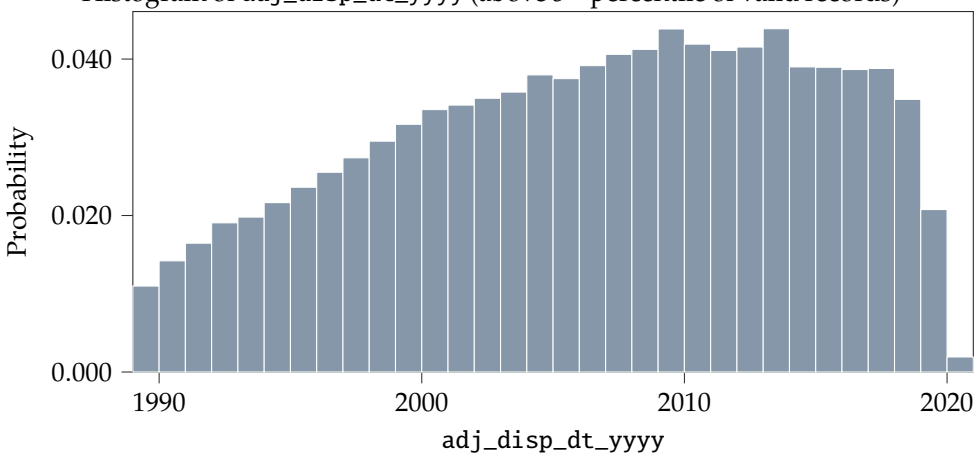
adj_disp_dt_yyyy

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	72,560,987	87.1
Invalid values	1,042	0.0
Missing values	10,714,018	12.9

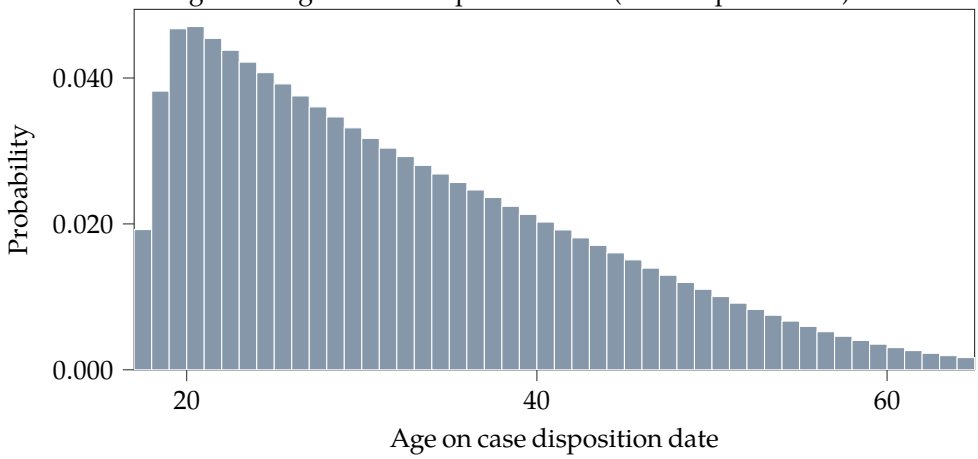
Statistic	*Value*
Mean	2,004.9
Minimum	1,900
25th percentile	1,999
Median	2,006
75th percentile	2,012
Maximum	2,020

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

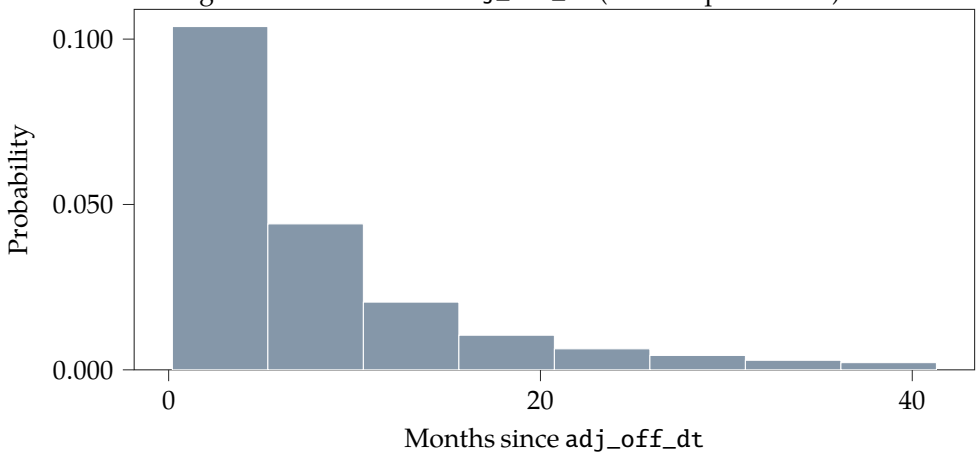


`adj_disp_dt_yyyy` (continued)

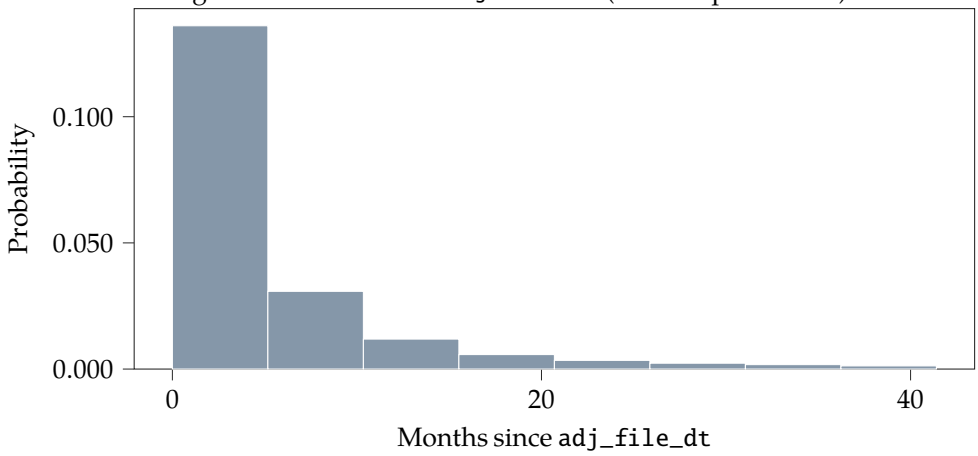
Histogram of age on case disposition date (1st–99th percentiles)



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



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



adj_disp_dt_yyyy (continued)

Data notes

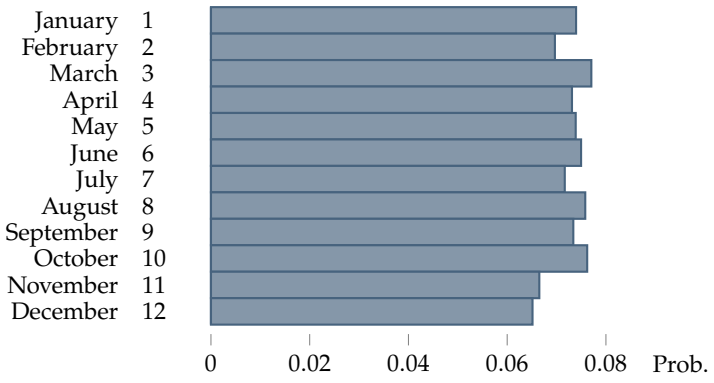
- Arizona: qlu
- Illinois: njg
- Nebraska: l8c
- Ohio: xe4

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	83,276,047	100.0
Valid records	72,562,029	87.1
Missing values	10,714,018	12.9

Histogram of adj_disp_dt_mm (valid records)



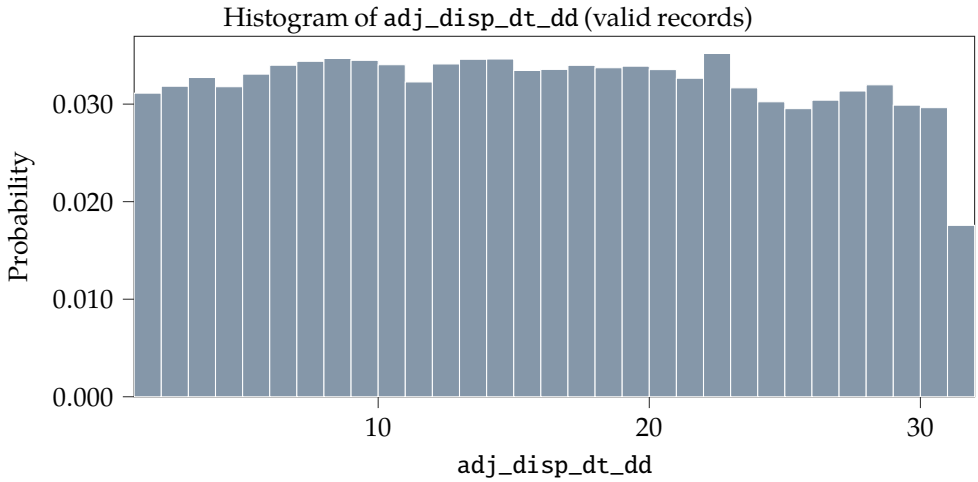
Data notes

- Arizona: qlu
- Illinois: njg
- Nebraska: l8c
- Ohio: xe4

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	83,276,047	100.0
Valid records	72,562,029	87.1
Missing values	10,714,018	12.9



Data notes

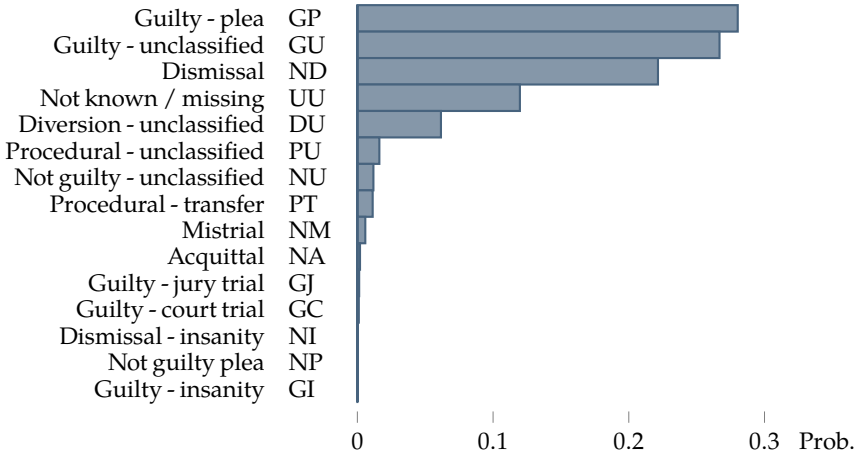
- Arizona: qlu
- Illinois: njg
- Nebraska: l8c
- Ohio: xe4

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	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of adj_disp_cd (valid records)



Data notes

- All states: qfp, mix
- Minnesota: qoz

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	83,276,047	100.0

Data notes

- All states: yik, mix

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	83,276,047	100.0
Valid records	83,276,047	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	83,276,047	100.0

Data notes

- Oregon: lxn

adj_off_dt_yyyy

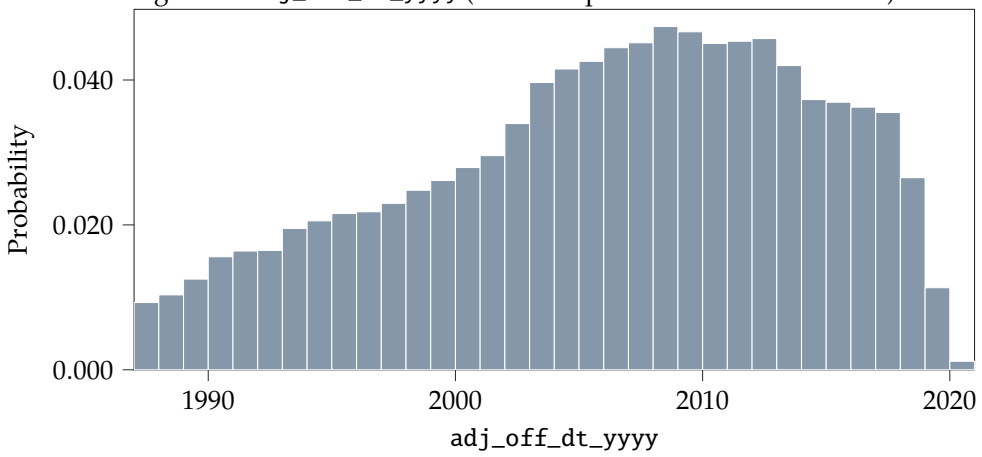
Label	Date offense was committed, year
Description	The year when the individual committed the offense.
Table	Adjudication
Format	numeric

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	47,076,540	56.5
Invalid values	120	0.0
Missing values	36,199,387	43.5

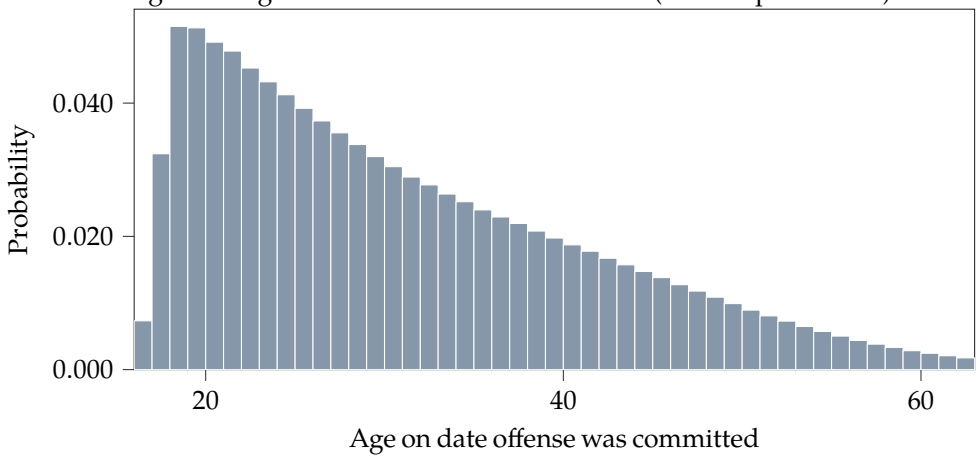
Statistic	*Value*
Mean	2,004.2
Minimum	1,900
25th percentile	1,998
Median	2,006
75th percentile	2,012
Maximum	2,020

adj_off_dt_yyyy (continued)

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

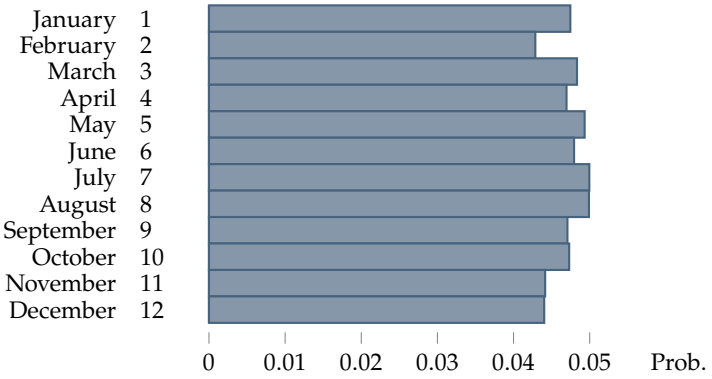
- Arizona: qho
- Arkansas: f7i
- Illinois: njg
- Mississippi: sc5
- Nebraska: l8c
- Ohio: xe4
- Texas: pr2, lrk, vsl

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	83,276,047	100.0
Valid records	47,076,660	56.5
Missing values	36,199,387	43.5

Histogram of adj_off_dt_mm (valid records)



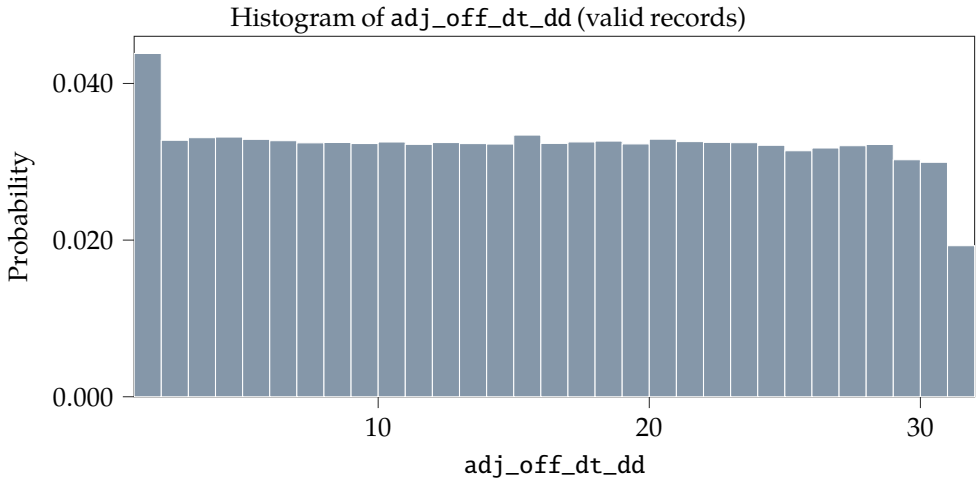
Data notes

- Arizona: qho
- Arkansas: f7i
- Illinois: njg
- Mississippi: sc5
- Nebraska: l8c
- Ohio: xe4
- Texas: pr2, lrk, vsl

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	83,276,047	100.0
Valid records	47,076,660	56.5
Missing values	36,199,387	43.5



Data notes

- Arizona: qho
- Arkansas: f7i
- Illinois: njg
- Mississippi: sc5
- Nebraska: l8c
- Ohio: xe4
- Texas: pr2, lrk, vsl

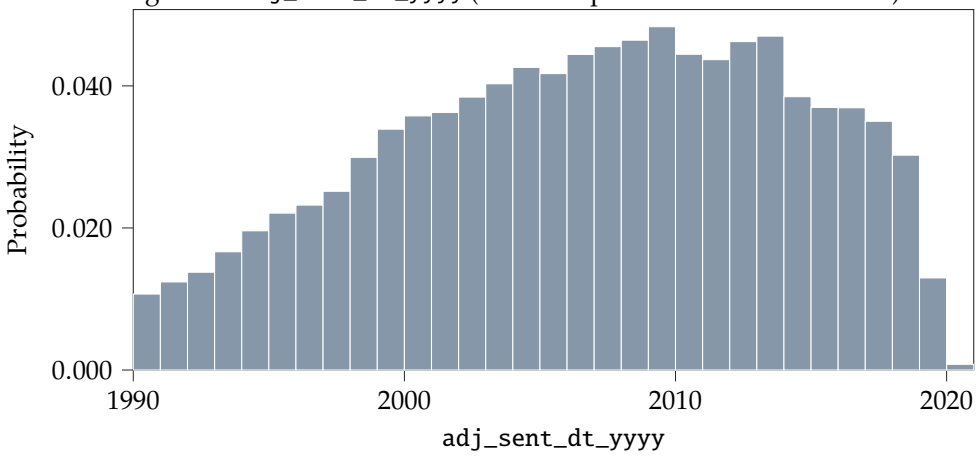
adj_sent_dt_yyyy

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

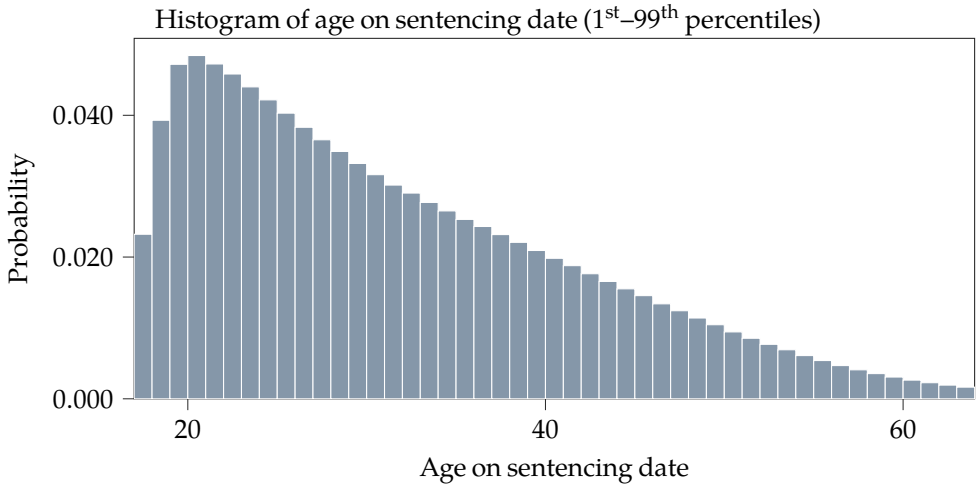
Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	31,723,538	38.1
Invalid values	2,004	0.0
Missing values	51,550,505	61.9

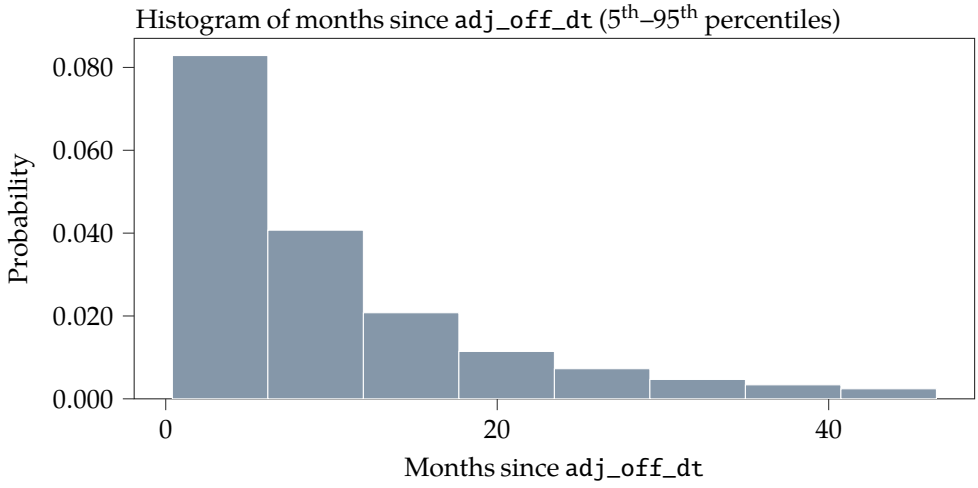
Statistic	*Value*
Mean	2,005.1
Minimum	1,900
25th percentile	2,000
Median	2,006
75th percentile	2,012
Maximum	2,020

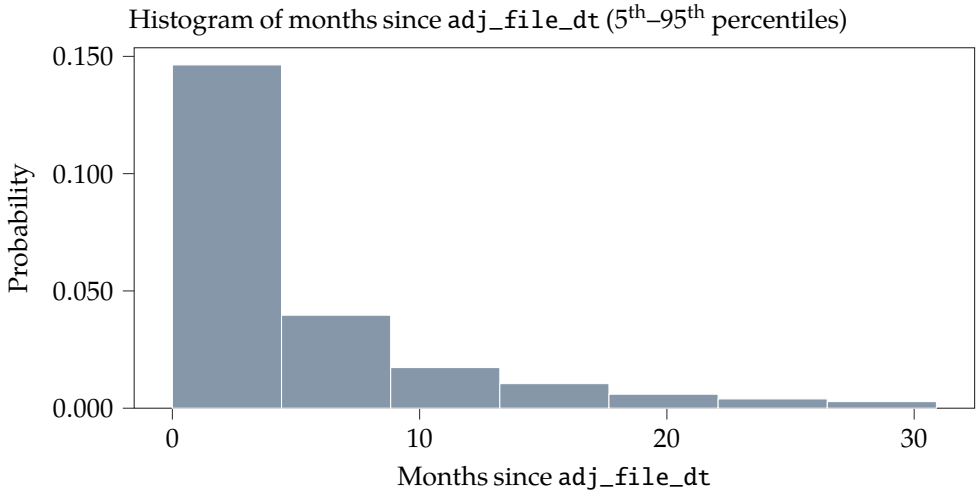
Histogram of adj_sent_dt_yyyy (above 5th percentile of valid records)



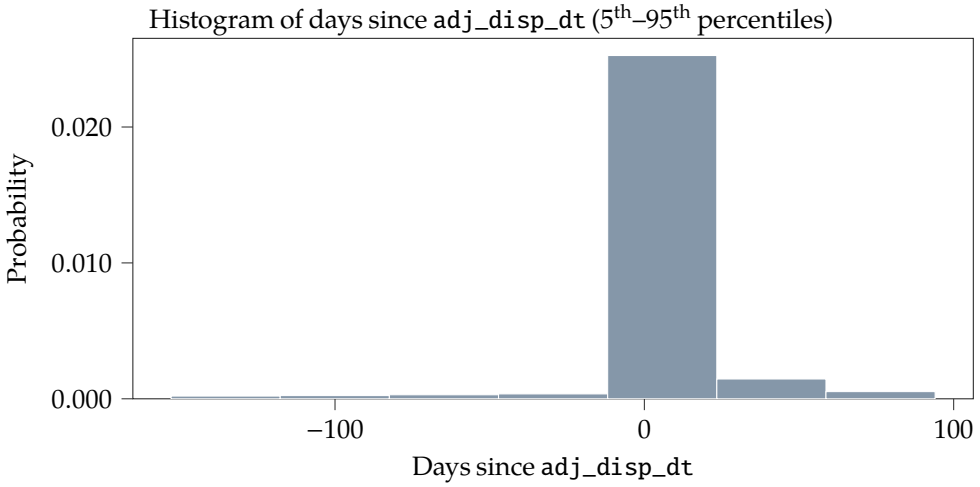
`adj_sent_dt_yyyy` (continued)







`adj_sent_dt_yyyy` (continued)



Data notes

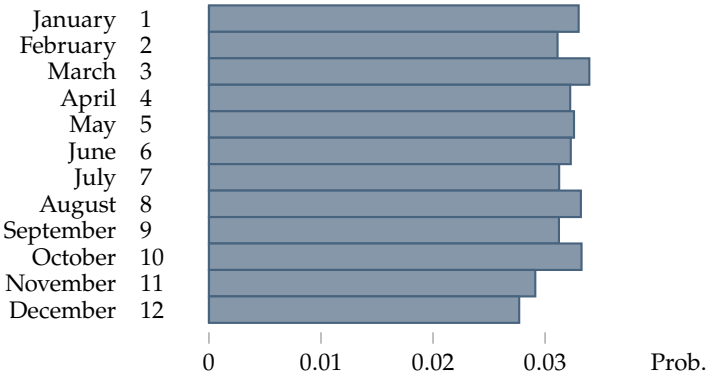
- Arizona: qho
- Florida: ny1
- Illinois: njg
- Maryland: e7v
- Nebraska: l8c
- New Jersey: se3
- North Carolina: p4u
- Ohio: xe4
- Pennsylvania: bwa
- Texas: ihz, kke, qc4, pr2, k8r, vbm, y69, q8v, vsl, ip7

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	83,276,047	100.0
Valid records	31,725,542	38.1
Missing values	51,550,505	61.9

Histogram of adj_sent_dt_mm (valid records)



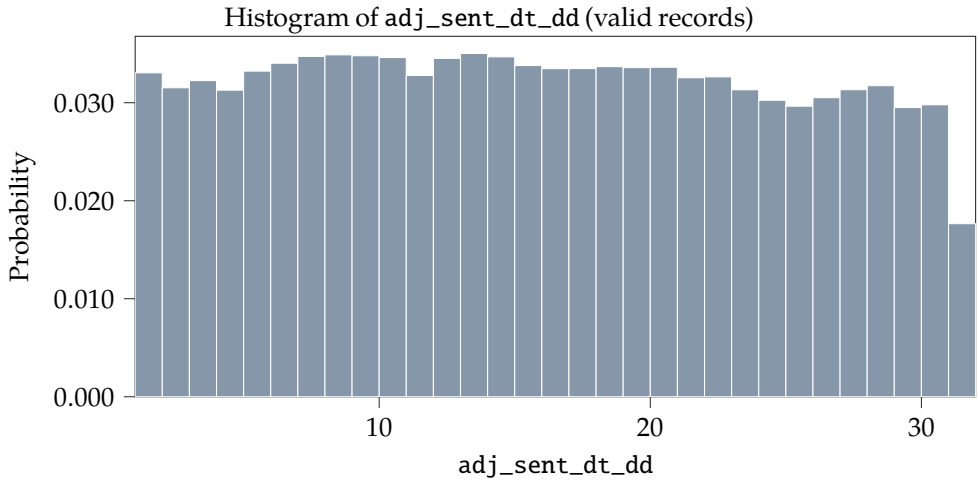
Data notes

- Arizona: qho
- Florida: ny1
- Illinois: njg
- Maryland: e7v
- Nebraska: l8c
- New Jersey: se3
- North Carolina: p4u
- Ohio: xe4
- Pennsylvania: bwa
- Texas: ihz, kke, qc4, pr2, k8r, vbm, y69, q8v, vsl, ip7

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	83,276,047	100.0
Valid records	31,725,542	38.1
Missing values	51,550,505	61.9



Data notes

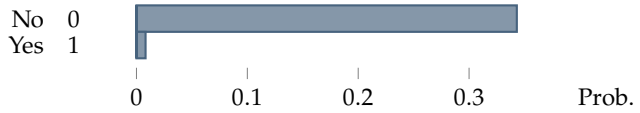
- Arizona: qho
- Florida: ny1
- Illinois: njg
- Maryland: e7v
- Nebraska: l8c
- New Jersey: se3
- North Carolina: p4u
- Ohio: xe4
- Pennsylvania: bwa
- Texas: ihz, kke, qc4, pr2, k8r, vbm, y69, q8v, vsl, ip7

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	83,276,047	100.0
Valid records	29,253,171	35.1
Missing values	54,022,876	64.9

Histogram of adj_sent_serv (valid records)



Data notes

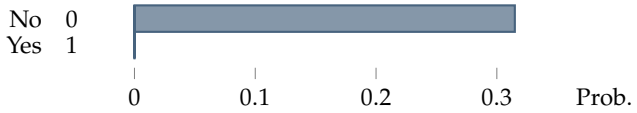
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, lrk, vbm, y69, e0d, c74, q8v, ux0, vsl

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	83,276,047	100.0
Valid records	26,226,524	31.5
Missing values	57,049,523	68.5

Histogram of `adj_sent_dth` (valid records)



Data notes

- New Jersey: `se3`
- Pennsylvania: `bwa`
- Texas: `g5n`, `lrk`, `vbm`, `y69`, `c74`, `ip7`

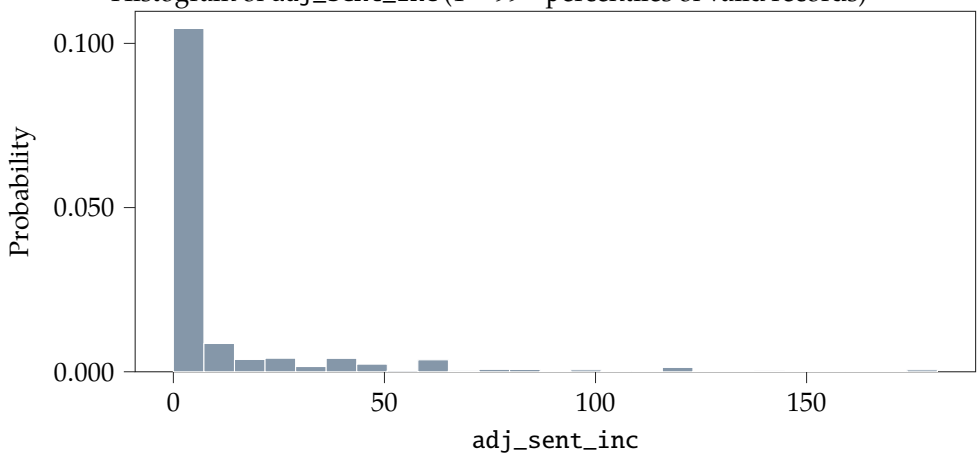
adj_sent_inc

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	83,276,047	100.0
Valid records	19,486,281	23.4
Missing values	63,789,766	76.6
Special values		
Life sentence (-88888)	32,178	0.0
Death sentence (-99999)	1,692	0.0

Statistic	*Value*
Mean	13.9
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	7.8
Maximum	1,200.0

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



Data notes

- All states: dor, tzh
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, vbm, c74, q8v, ip7

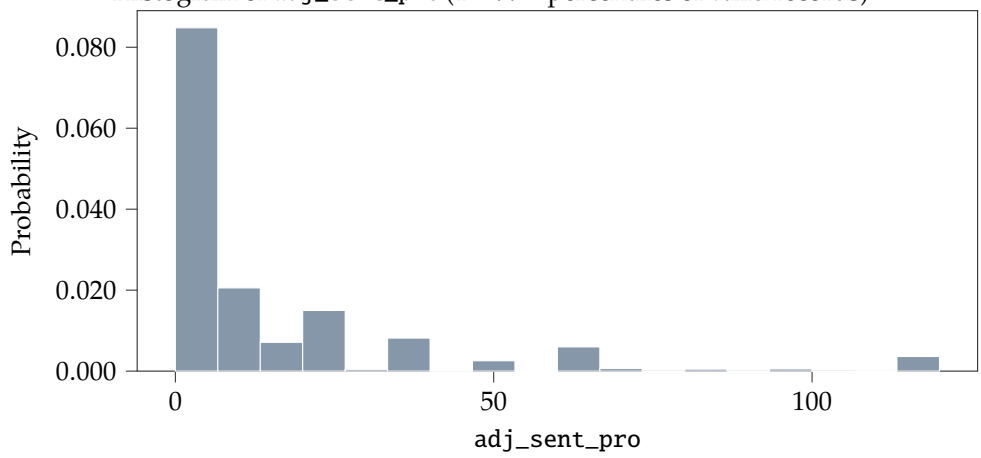
adj_sent_pro

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	83,276,047	100.0
Valid records	20,791,834	25.0
Missing values	62,484,213	75.0

Statistic	*Value*
Mean	14.3
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	20.1
Maximum	120.0

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



Data notes

- All states: j8o
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, vbm, y69, c74, q8v, ip7

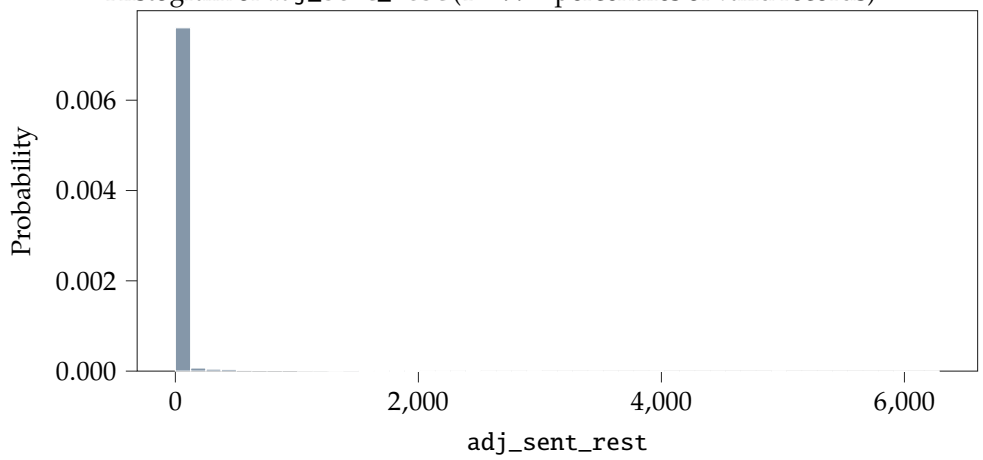
adj_sent_rest

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	83,276,047	100.0
Valid records	15,747,614	18.9
Missing values	67,528,433	81.1

Statistic	*Value*
Mean	1,185.1
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	0.0
Maximum	500,000.0

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



Data notes

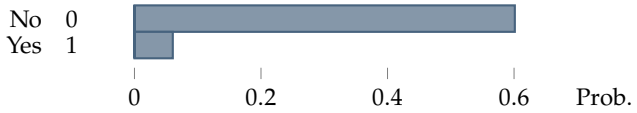
- All states: z6t, kjh
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, vbm, y69, e0d, c74, q8v, ux0, vsl, ip7

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	83,276,047	100.0
Valid records	55,114,563	66.2
Missing values	28,161,484	33.8

Histogram of adj_sent_sus (valid records)



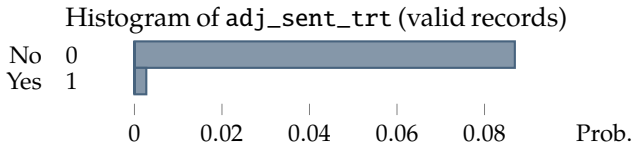
Data notes

- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, c74

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	83,276,047	100.0
Valid records	7,464,654	9.0
Missing values	75,811,393	91.0



Data notes

- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, lrk, vbm, y69, e0d, c74, q8v, ux0, vsl

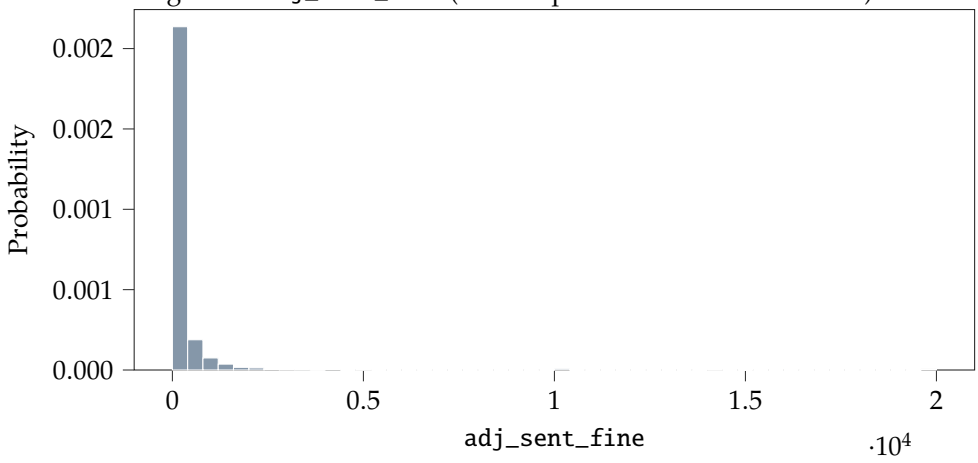
adj_sent_fine

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	29,229,364	35.1
Invalid values	10	0.0
Missing values	54,046,673	64.9

Statistic	*Value*
Mean	1,148.5
Minimum	0.0
25th percentile	0.0
Median	0.0
75th percentile	100.0
Maximum	500,000.0

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



Data notes

- All states: z6t, zr5
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, vbm, y69, c74, ip7

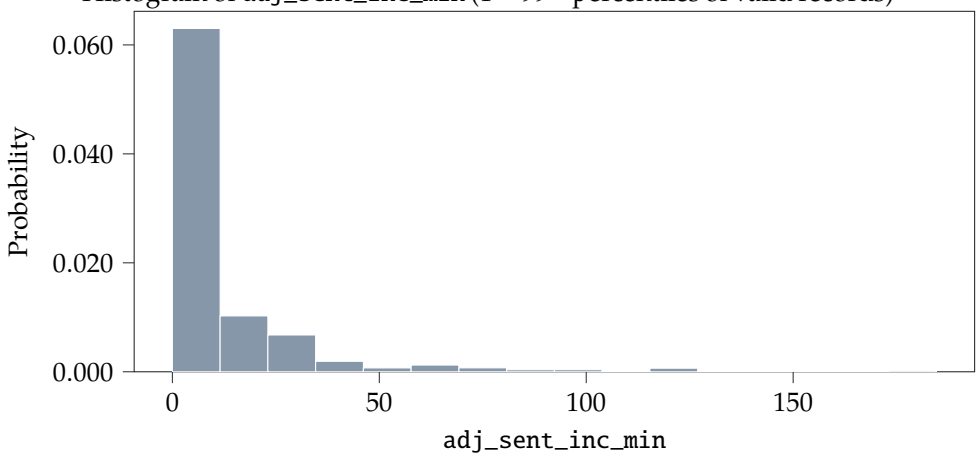
adj_sent_inc_min

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	83,276,047	100.0
Valid records	4,334,274	5.2
Missing values	78,941,773	94.8
Special values		
Life sentence (-88888)	2,519	0.0
Death sentence (-99999)	878	0.0

Statistic	*Value*
Mean	14.2
Minimum	0.0
25th percentile	0.0
Median	1.5
75th percentile	12.1
Maximum	1,200.0

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



Data notes

- All states: dor, tzh
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, lrk, vbm, y69, e0d, c74, q8v, ux0, vsl, ip7

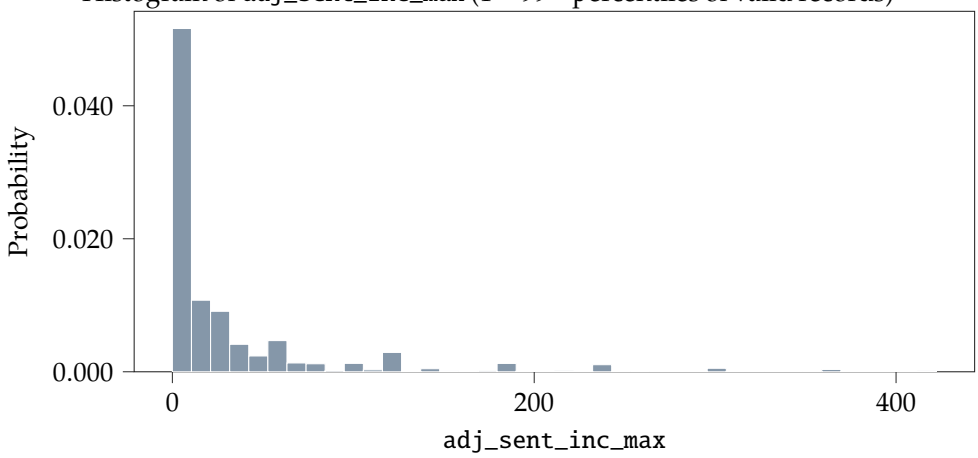
adj_sent_inc_max

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	83,276,047	100.0
Valid records	6,442,392	7.7
Missing values	76,833,655	92.3
Special values		
Life sentence (-88888)	16,898	0.0
Death sentence (-99999)	14,022	0.0

Statistic	*Value*
Mean	36.7
Minimum	0.0
25th percentile	0.0
Median	8.0
75th percentile	35.1
Maximum	1,200.0

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



Data notes

- All states: dor, tzh
- New Jersey: se3
- Pennsylvania: bwa
- Texas: g5n, lrk, vbm, y69, e0d, c74, q8v, ux0, vsl, ip7

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	83,276,047	100.0

Data notes

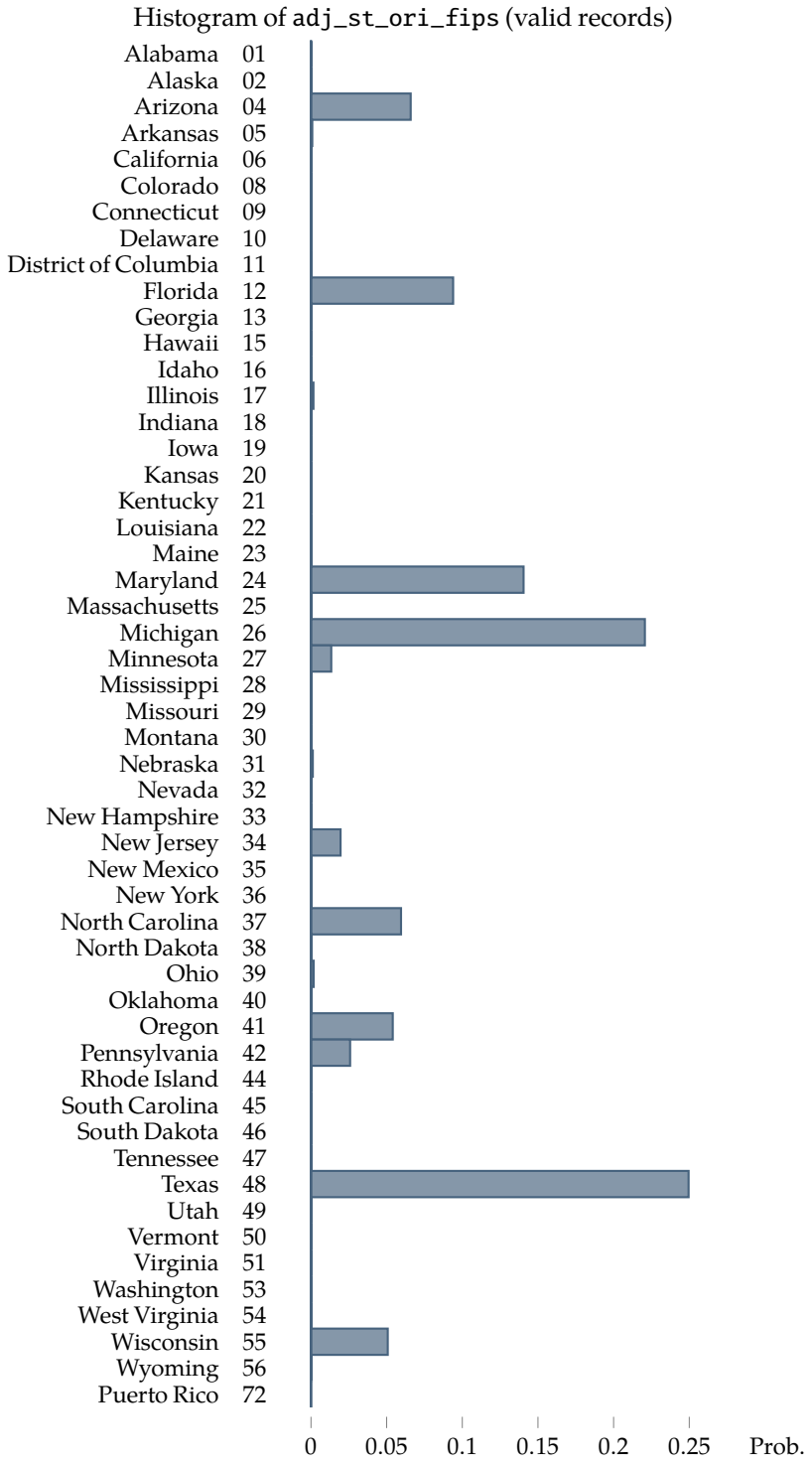
- All states: yik
- Arizona: qho
- Oregon: lxn

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	83,276,047	100.0
Valid records	83,276,047	100.0

`adj_st_ori_fips` (continued)



Data notes

- All states: `btp`

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

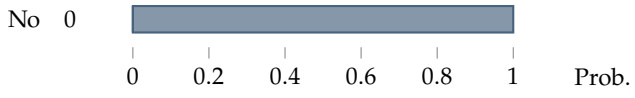
Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	76,266,138	91.6
Missing values	7,009,909	8.4

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	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of adj_rec_src_le (valid records)

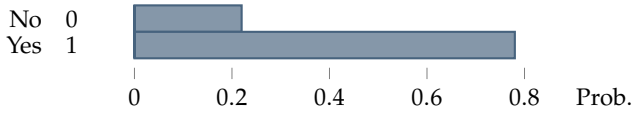


adj_rec_src_crt

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of `adj_rec_src_crt` (valid records)

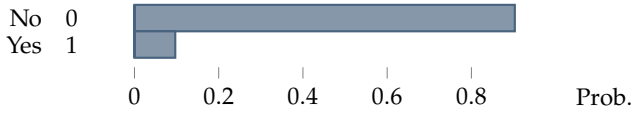


adj_rec_src_doc

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of `adj_rec_src_doc` (valid records)

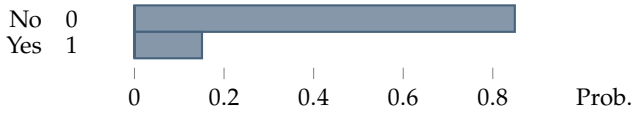


adj_rec_src_rep

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of `adj_rec_src_rep` (valid records)

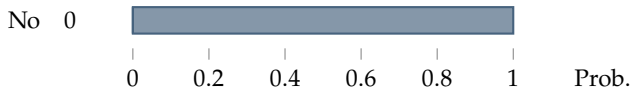


adj_rec_src_cc

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

Set	*Count*	*Percent (%)*
All records	83,276,047	100.0
Valid records	83,276,047	100.0

Histogram of `adj_rec_src_cc` (valid records)



3.6.4 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: `twd`
- Michigan: `ydo`

cjars_id

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 5.
Table	Probation
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Missing values	1	0.0
Unique values	5,893,185	

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 3.2.2 and Figure 6.
Table	Probation
Table key	primary, unique
Format	string

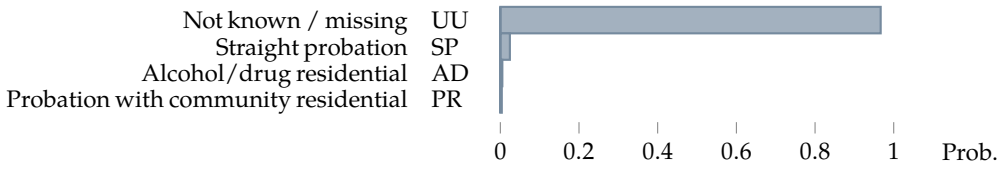
Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Unique values	8,598,938	

`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	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of `pro_cond_cd` (valid records)



Data notes

- Texas: `cjk`

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	8,598,938	100.0

Data notes

- All states: yik

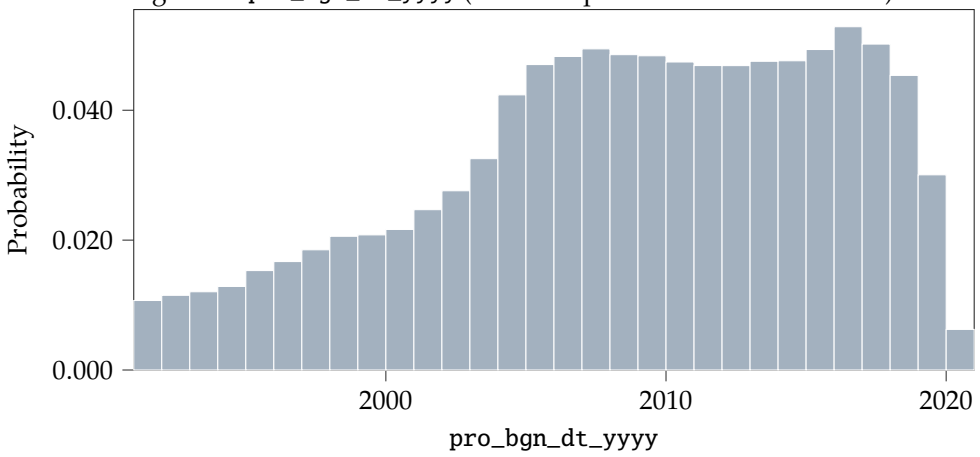
pro_bgn_dt_yyyy

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,172,432	95.0
Invalid values	53	0.0
Missing values	426,453	5.0

Statistic	*Value*
Mean	2,007.3
Minimum	1,900
25th percentile	2,003
Median	2,008
75th percentile	2,014
Maximum	2,020

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



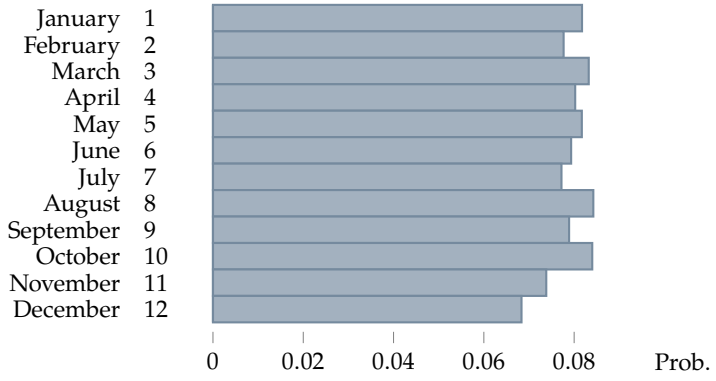
pro_bgn_dt_yyyy (continued)

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	8,598,938	100.0
Valid records	8,172,485	95.0
Missing values	426,453	5.0

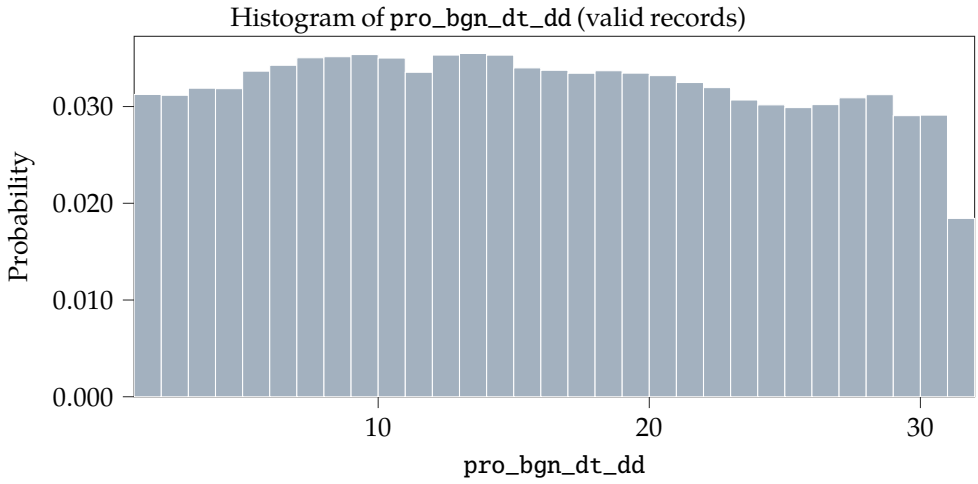
Histogram of pro_bgn_dt_mm (valid records)



pro_bgn_dt_dd

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,172,485	95.0
Missing values	426,453	5.0



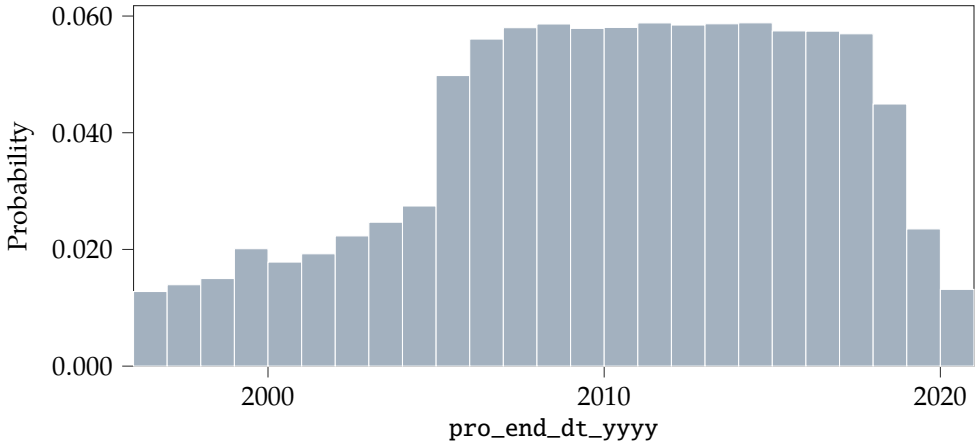
pro_end_dt_yyyy

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

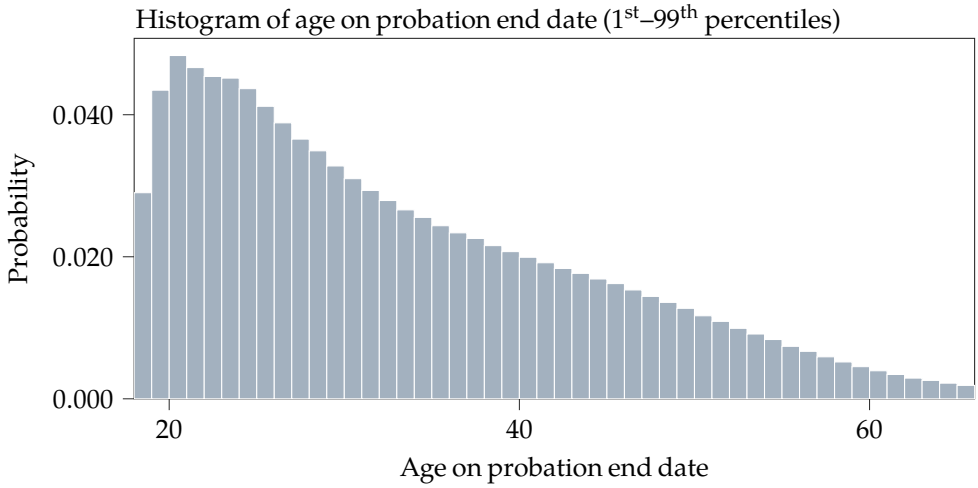
Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	5,097,249	59.3
Invalid values	2	0.0
Missing values	3,501,687	40.7

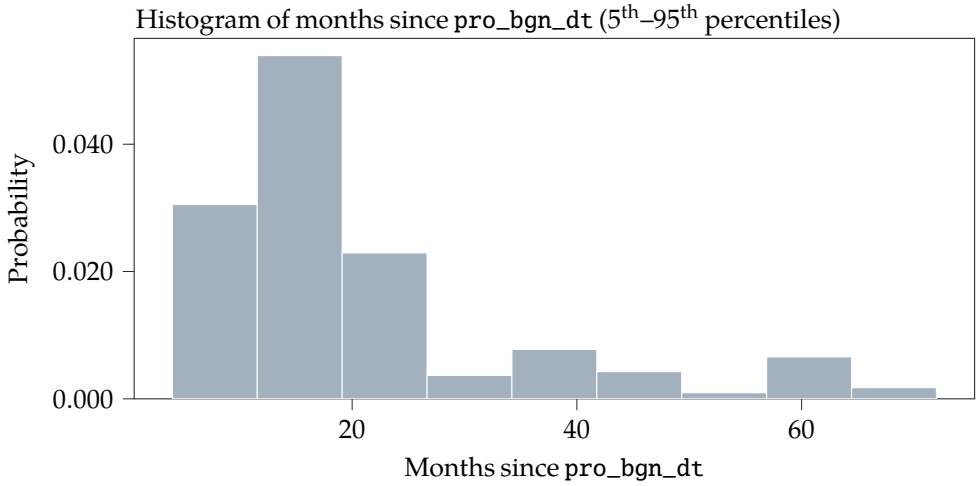
Statistic	*Value*
Mean	2,009.0
Minimum	1,900
25th percentile	2,005
Median	2,010
75th percentile	2,014
Maximum	2,020

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



`pro_end_dt_yyyy` (continued)





Data notes

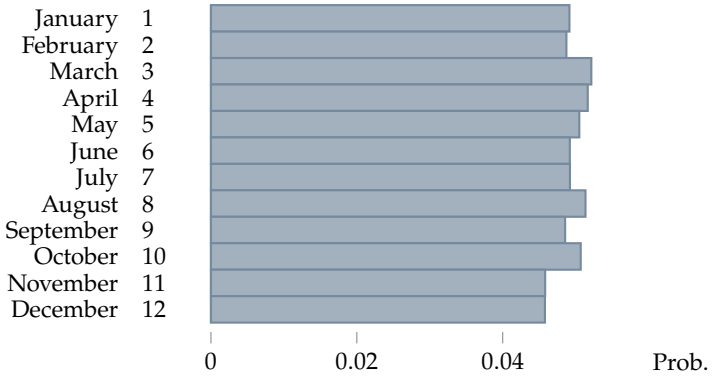
- Arkansas: `lrw`
- Florida: `u2g`
- North Carolina: `t0b`
- Texas: `cjk`

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	8,598,938	100.0
Valid records	5,097,251	59.3
Missing values	3,501,687	40.7

Histogram of pro_end_dt_mm (valid records)



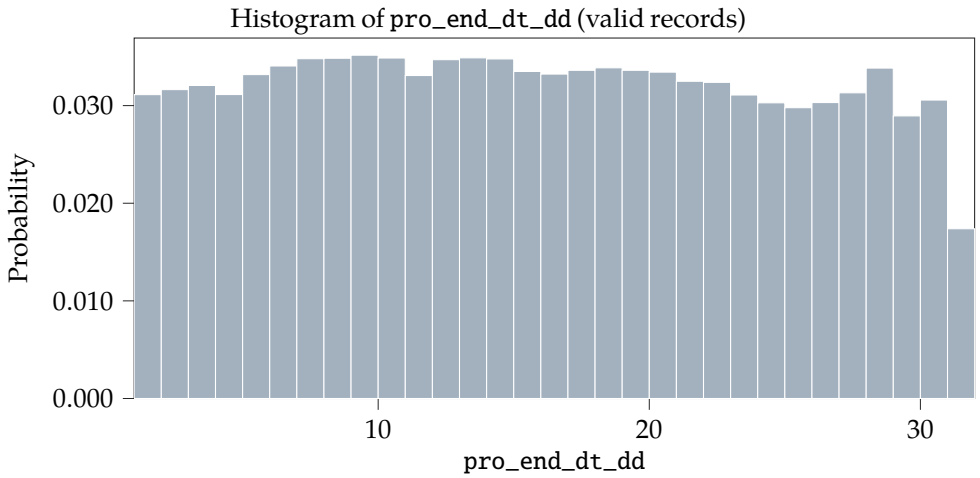
Data notes

- Arkansas: lrw
- Florida: u2g
- North Carolina: t0b
- Texas: cjk

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	8,598,938	100.0
Valid records	5,097,251	59.3
Missing values	3,501,687	40.7



Data notes

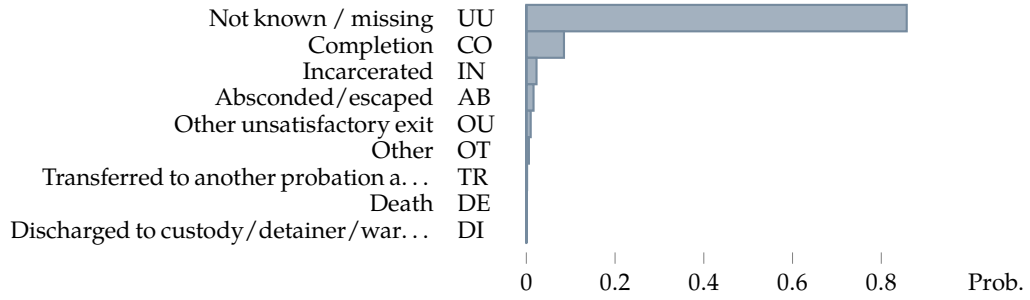
- Arkansas: lrw
- Florida: u2g
- North Carolina: t0b
- Texas: cjk

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	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of pro_end_cd (valid records)



Data notes

- Texas: cjk

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	8,598,938	100.0

Data notes

- All states: yik

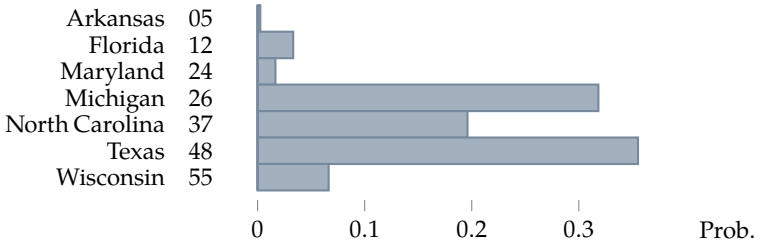
pro_st_ori_fips

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	8,598,938	100.0
Valid records	8,497,445	98.8
Missing values	101,493	1.2

pro_st_ori_fips (continued)

Histogram of pro_st_ori_fips (valid records)



pro_cnty_ori_fips

Label	County FIPS code for location of sentencing
Description	County-level Federal Information Processing Standards (FIPS) code for location where the individual was sentenced. Used to help uniquely 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

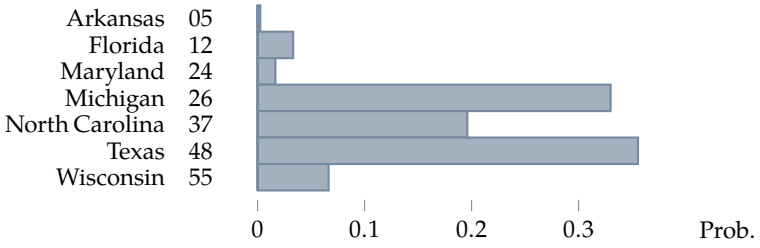
Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	7,715,697	89.7
Missing values	781,748	9.1

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 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	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of `pro_st_juris_fips` (valid records)

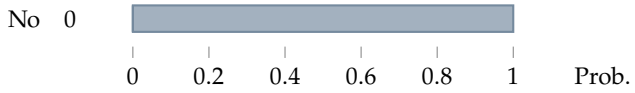


pro_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of `pro_rec_src_le` (valid records)

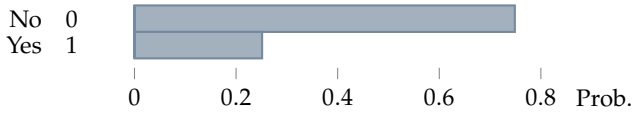


pro_rec_src_crt

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of `pro_rec_src_crt` (valid records)

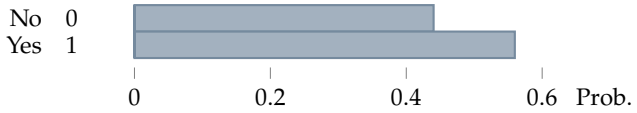


pro_rec_src_doc

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of pro_rec_src_doc (valid records)

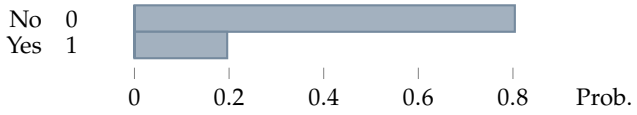


pro_rec_src_rep

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of pro_rec_src_rep (valid records)

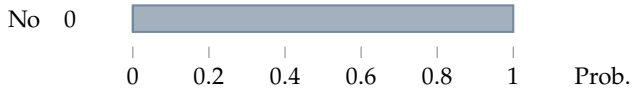


pro_rec_src_cc

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

Set	*Count*	*Percent (%)*
All records	8,598,938	100.0
Valid records	8,598,938	100.0

Histogram of pro_rec_src_cc (valid records)



3.6.5 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: `twd`
- Arizona: `jdt`

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 5.
Table	Incarceration
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Unique values	3,608,120	

`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 3.2.2 and Figure 6.
Table	Incarceration
Table key	primary, unique
Format	string

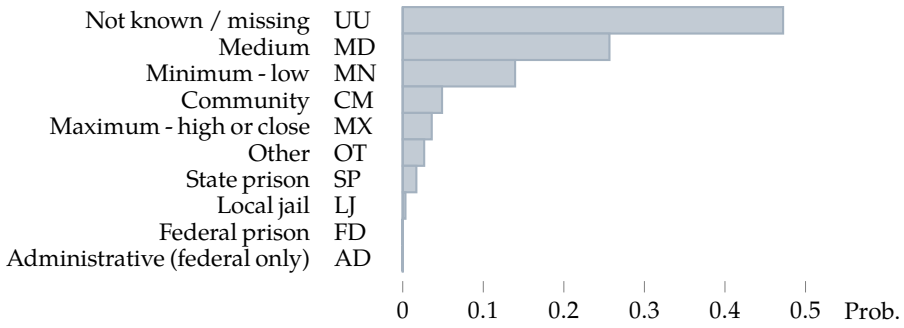
Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Unique values	6,375,505	

`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	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_fcl_cd` (valid records)



Data notes

- Texas: c74

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	6,375,505	100.0

Data notes

- All states: yik

inc_entry_dt_yyyy

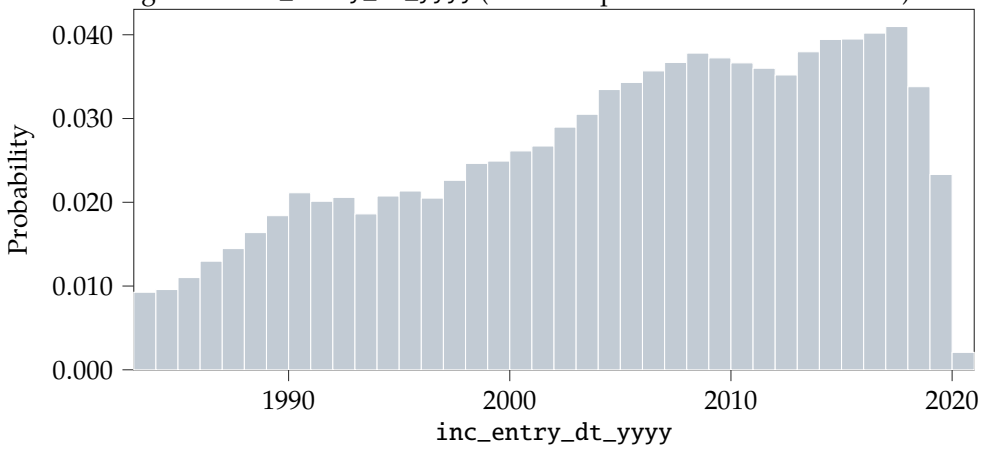
Label	Incarceration entry date, year
Description	The year when the individual entered incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	6,152,187	96.5
Missing values	223,318	3.5

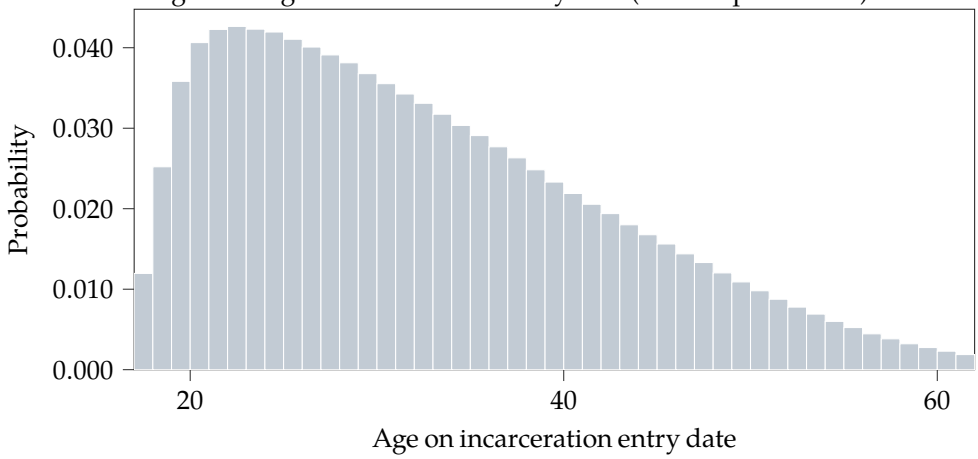
Statistic	*Value*
Mean	2,003.3
Minimum	1,901
25th percentile	1,996
Median	2,005
75th percentile	2,012
Maximum	2,020

`inc_entry_dt_yyyy` (continued)

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



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



Data notes

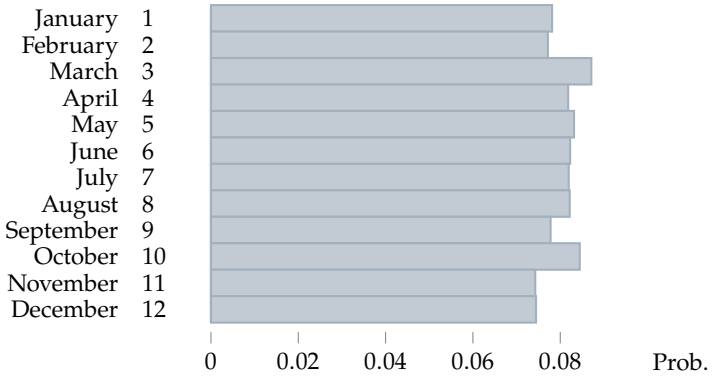
- Texas: `s9m`
- Wisconsin: `jf0`

`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	6,375,505	100.0
Valid records	6,150,789	96.5
Missing values	224,716	3.5

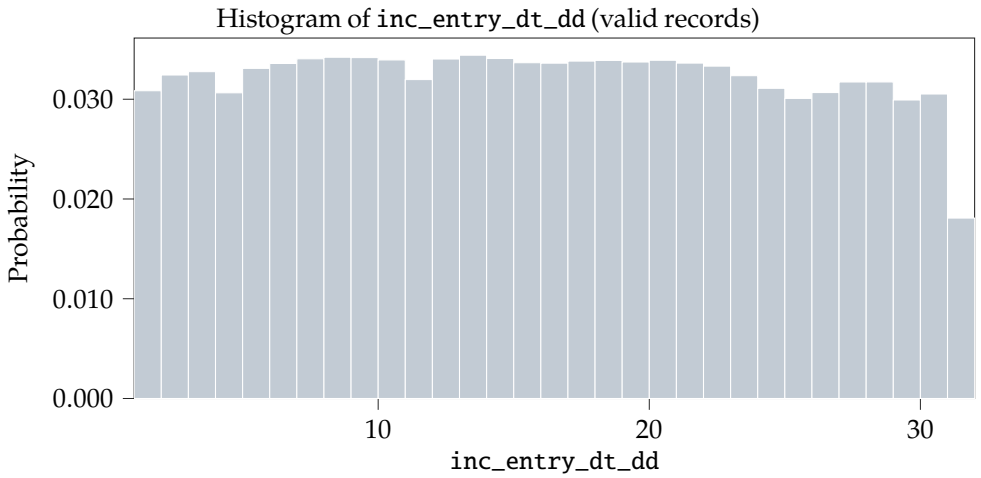
Histogram of `inc_entry_dt_mm` (valid records)



`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	6,375,505	100.0
Valid records	6,150,789	96.5
Missing values	224,716	3.5

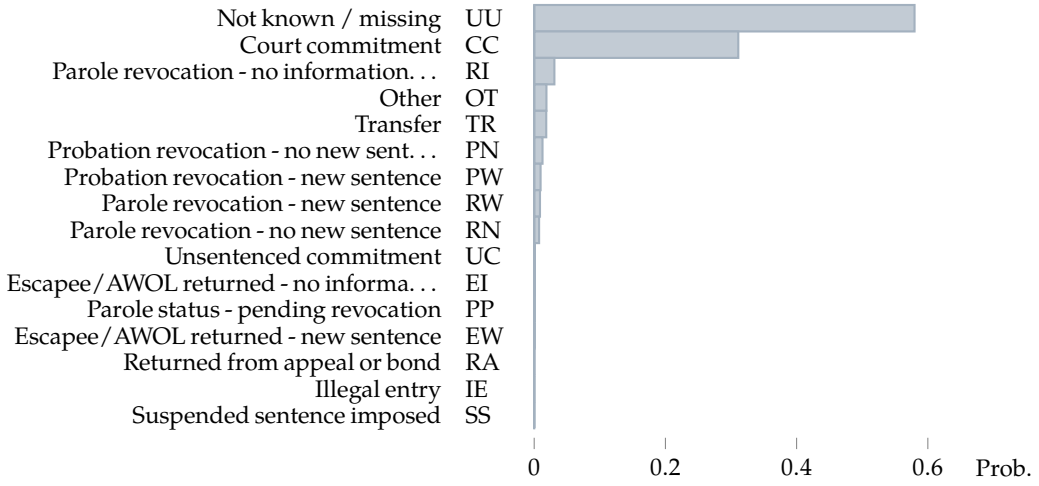


`inc_entry_cd`

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	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_entry_cd` (valid records)



Data notes

- Texas: c74

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	6,375,505	100.0

Data notes

- All states: yik

inc_exit_dt_yyyy

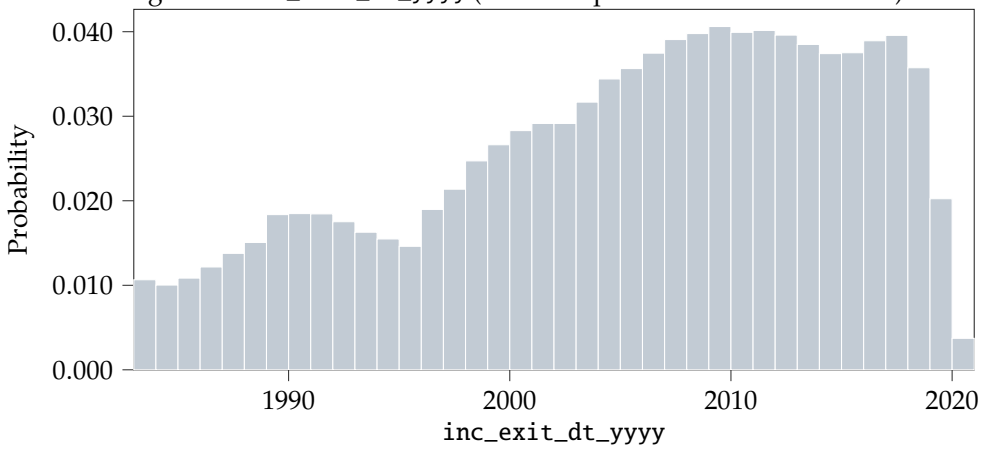
Label	Incarceration exit date, year
Description	Year when the individual exited incarceration.
Table	Incarceration
Format	numeric

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	5,452,397	85.5
Invalid values	1	0.0
Missing values	923,107	14.5

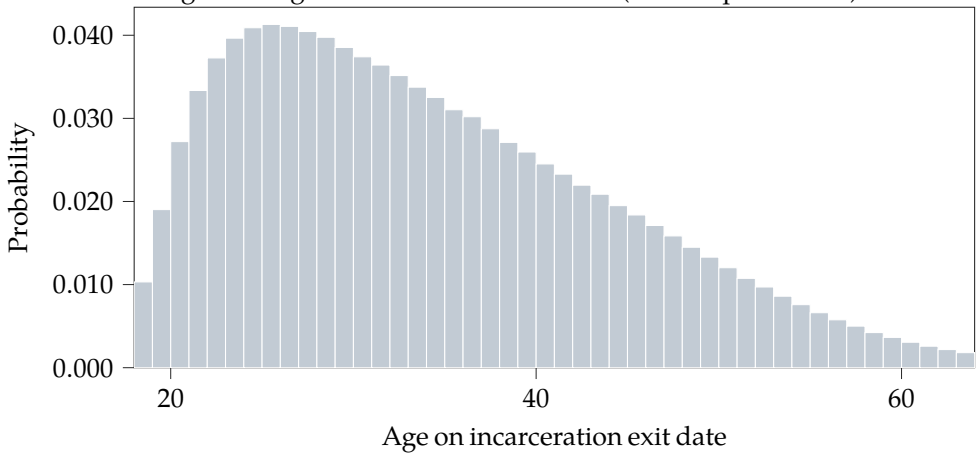
Statistic	*Value*
Mean	2,003.6
Minimum	1,901
25^{th} percentile	1,997
Median	2,006
75^{th} percentile	2,012
Maximum	2,020

`inc_exit_dt_yyyy` (continued)

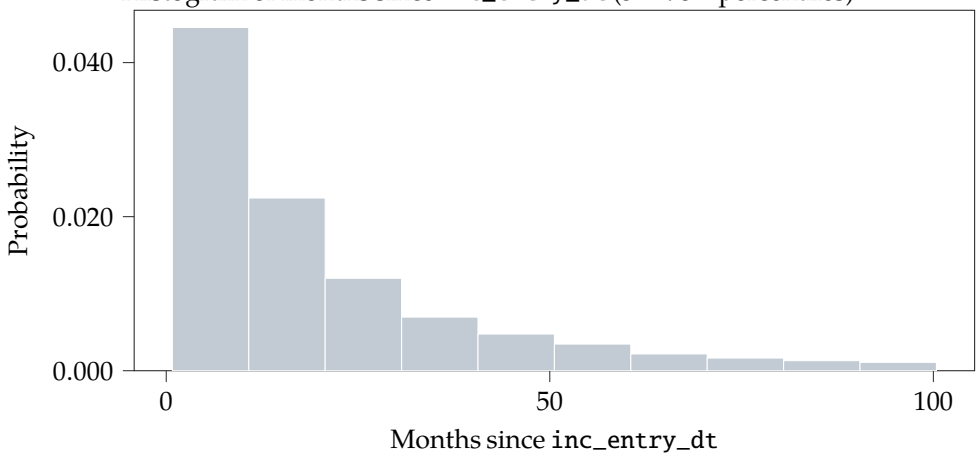
Histogram of `inc_exit_dt_yyyy` (above 5th percentile of valid records)



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



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



inc_exit_dt_yyyy (continued)

Data notes

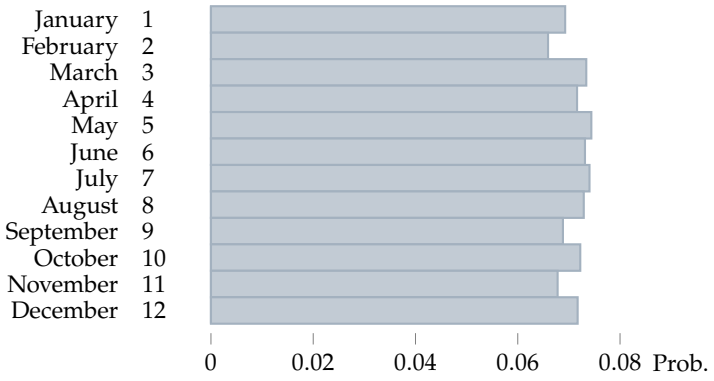
- Arkansas: dd7
- Mississippi: wl2
- Texas: s9m
- Wisconsin: jf0

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	6,375,505	100.0
Valid records	5,452,398	85.5
Missing values	923,107	14.5

Histogram of inc_exit_dt_mm (valid records)



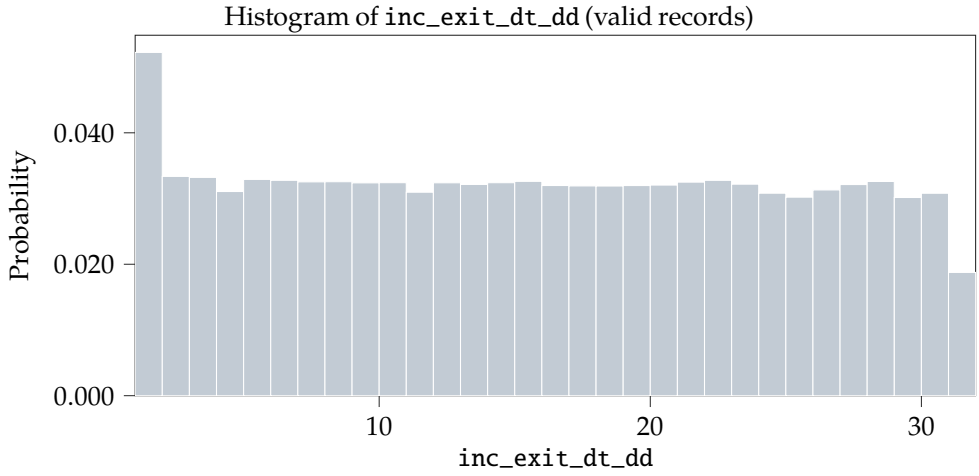
Data notes

- Arkansas: dd7
- Mississippi: wl2

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	6,375,505	100.0
Valid records	5,452,398	85.5
Missing values	923,107	14.5



Data notes

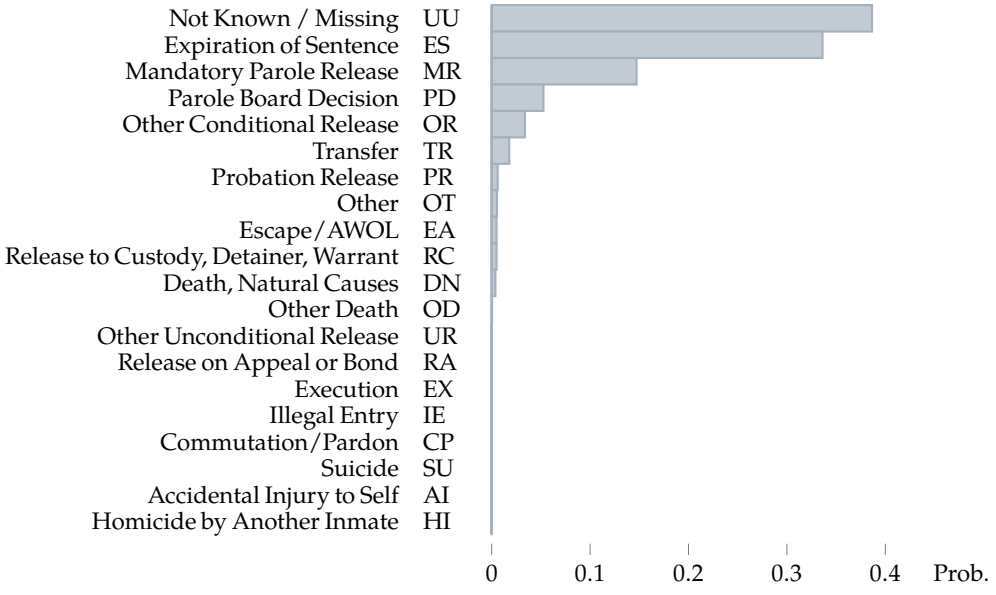
- Arkansas: dd7
- Florida: hn4
- Mississippi: wl2

`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	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_exit_cd` (valid records)



Data notes

- Texas: c74

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	6,375,505	100.0

Data notes

- All states: yik

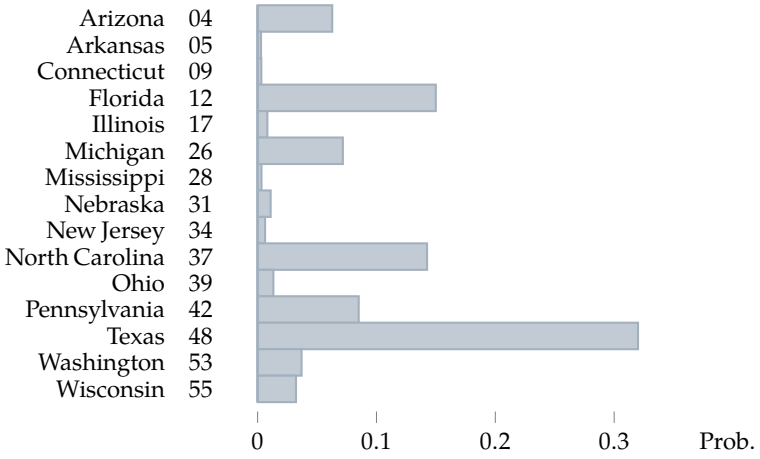
inc_st_ori_fips

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	6,375,505	100.0
Valid records	6,051,826	94.9
Missing values	323,679	5.1

inc_st_ori_fips (continued)

Histogram of inc_st_ori_fips (valid records)



Data notes

- Colorado: o6q

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	6,375,505	100.0
Valid records	4,483,695	70.3
Missing values	1,568,131	24.6

Data notes

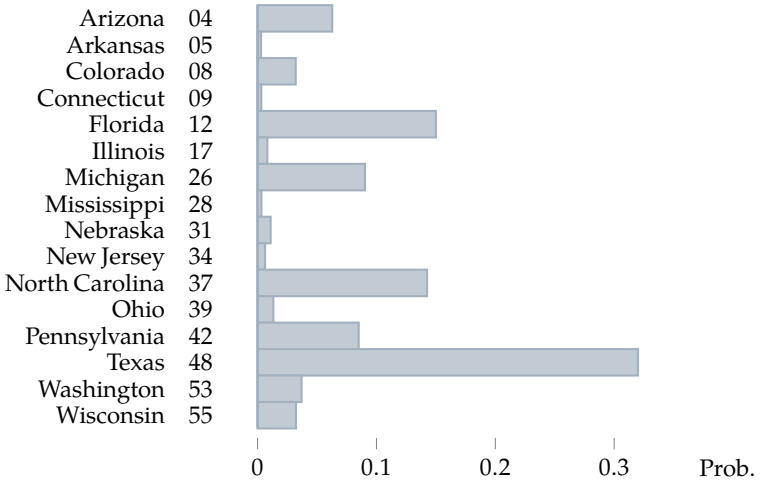
- Colorado: o6q
- Washington: fpm
- Wisconsin: ffx

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	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_st_juris_fips` (valid records)

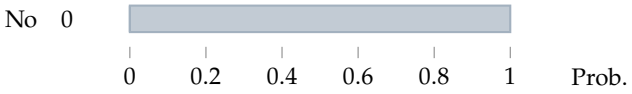


inc_rec_src_le

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

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_rec_src_le` (valid records)

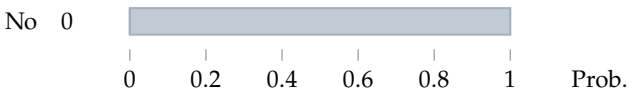


inc_rec_src_crt

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

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_rec_src_crt` (valid records)

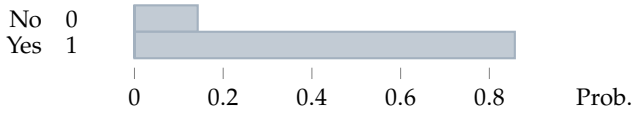


inc_rec_src_doc

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

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_rec_src_doc` (valid records)

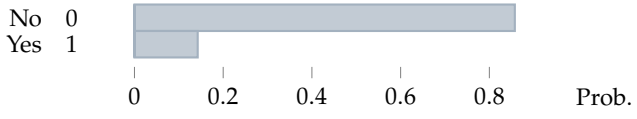


inc_rec_src_rep

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

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of `inc_rec_src_rep` (valid records)

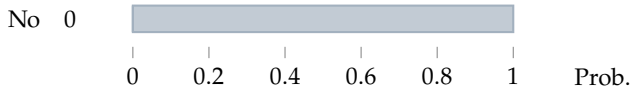


inc_rec_src_cc

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

Set	*Count*	*Percent (%)*
All records	6,375,505	100.0
Valid records	6,375,505	100.0

Histogram of inc_rec_src_cc (valid records)



3.6.6 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: twd

`cjars_id`

Label	CJARS person identifier
Description	Uniquely identifies individuals. For more details on use of `cjars_id` for data linkage, refer to Section 3.2.1 and Figure 5.
Table	Parole
Table key	foreign, not unique
Format	string

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Unique values	1,271,717	

`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 3.2.2 and Figure 6.
Table	Parole
Table key	primary, unique
Format	string

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Unique values	2,227,140	

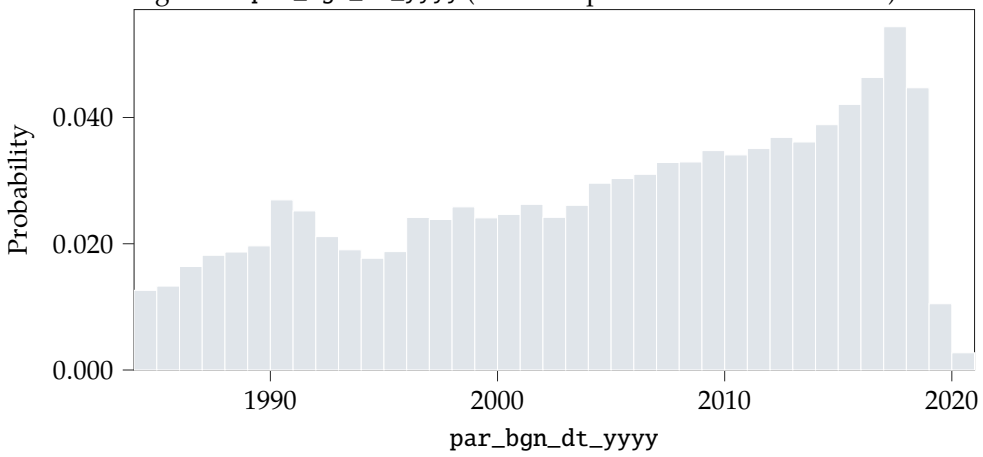
`par_bgn_dt_yyyy`

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

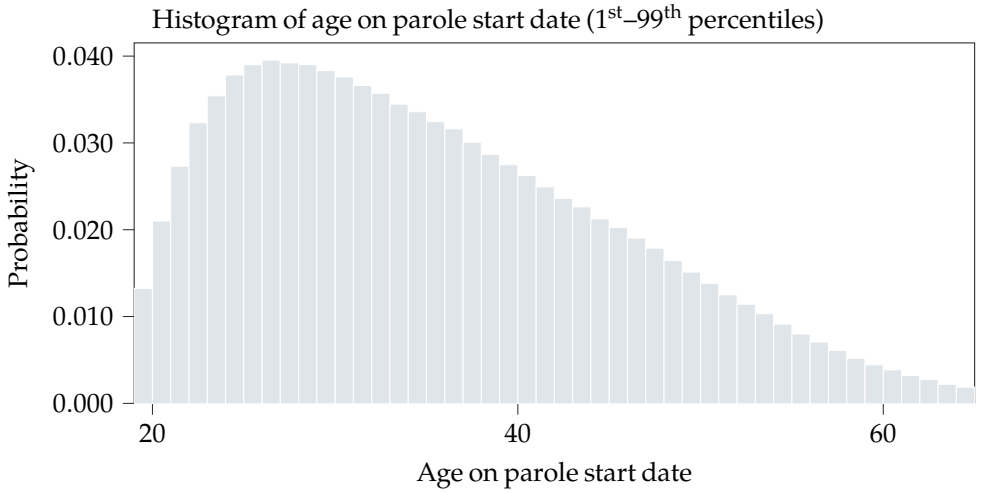
Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	2,097,304	94.2
Invalid values	1	0.0
Missing values	129,835	5.8

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)



Data notes

- Texas: s8m

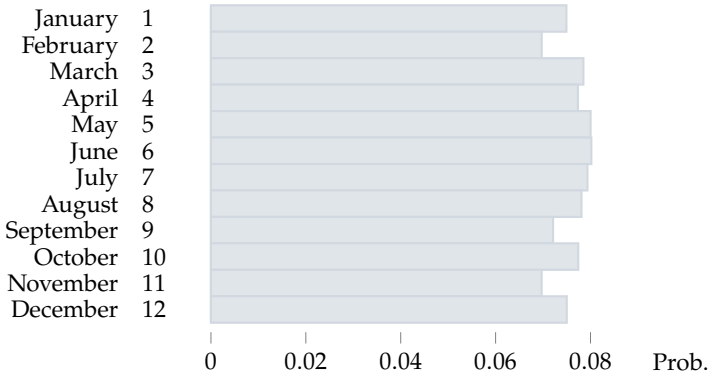
`par_bgn_dt_mm`

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	2,227,140	100.0
Valid records	2,032,061	91.2
Missing values	195,079	8.8

`par_bgn_dt_mm` (continued)

Histogram of `par_bgn_dt_mm` (valid records)

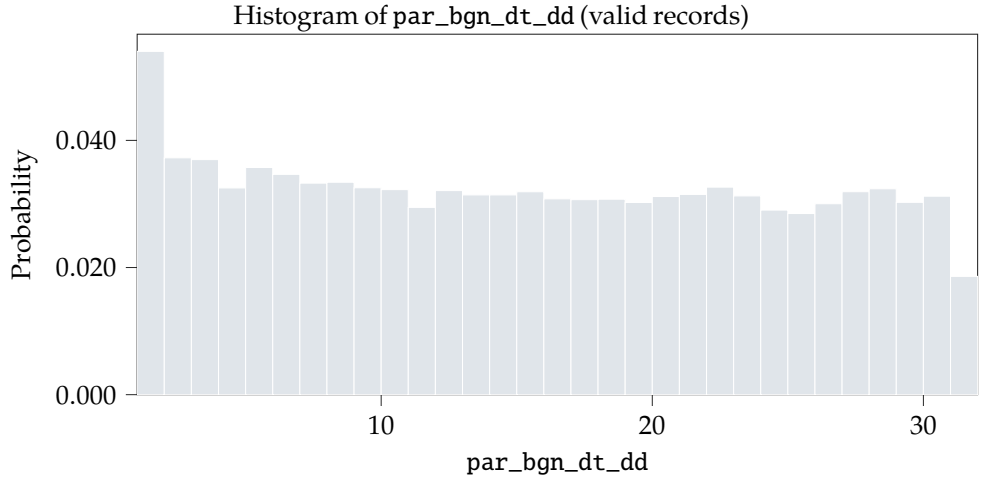


`par_bgn_dt_dd`

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

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	2,032,061	91.2
Missing values	195,079	8.8

Histogram of `par_bgn_dt_dd` (valid records)



Data notes

- Nebraska: `hsi`

`par_end_dt_yyyy`

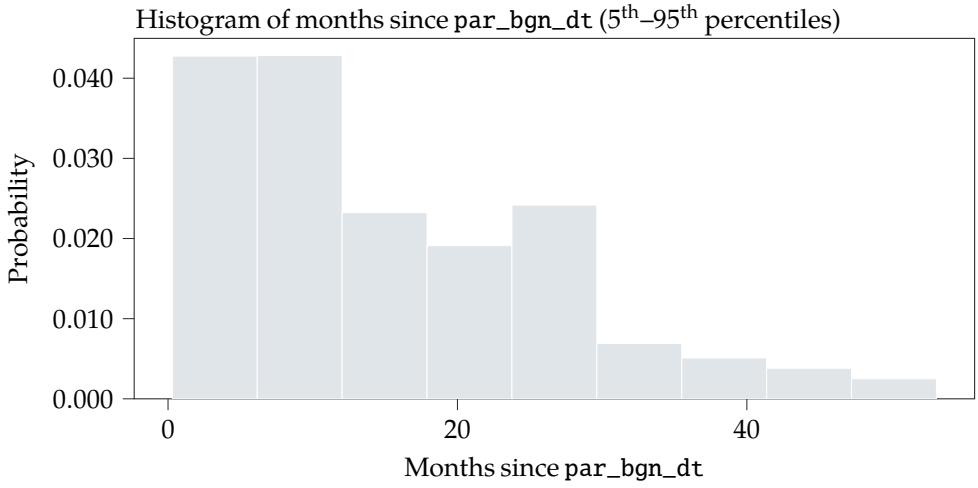
Label	Parole end date, year
Description	The year when the individual's parole ended.
Table	Parole
Format	numeric

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	1,195,806	53.7
Invalid values	1	0.0
Missing values	1,031,333	46.3

Statistic	*Value*
Mean	2,003.0
Minimum	1,900
25th percentile	1,995
Median	2,004
75th percentile	2,012
Maximum	2,020

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

`par_end_dt_yyyy` (continued)



Data notes

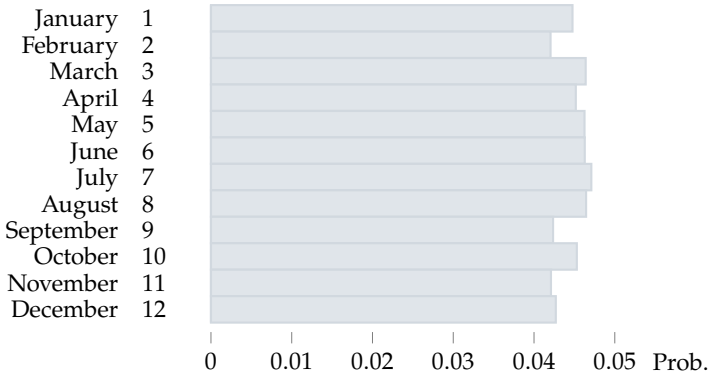
- Florida: `u2g`
- Illinois: `kr7`
- Texas: `s8m`, `qs8`

`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	2,227,140	100.0
Valid records	1,195,807	53.7
Missing values	1,031,333	46.3

Histogram of `par_end_dt_mm` (valid records)



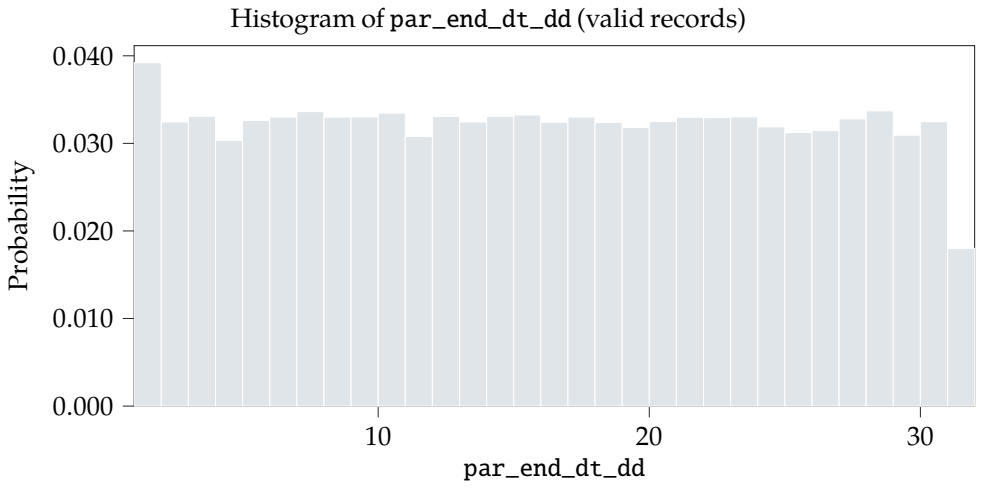
Data notes

- Florida: `u2g`
- Illinois: `kr7`
- Texas: `qs8`

`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	2,227,140	100.0
Valid records	1,195,807	53.7
Missing values	1,031,333	46.3



Data notes

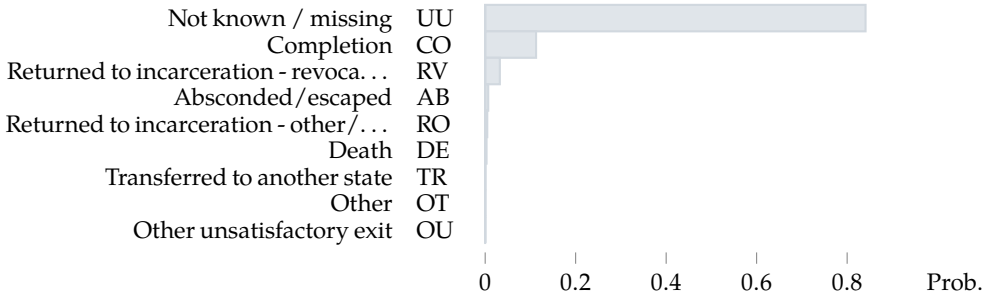
- Florida: u2g
- Illinois: kr7
- Texas: qs8

`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	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_end_cd` (valid records)



par_end_cd_src

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

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0

Data notes

- All states: yik

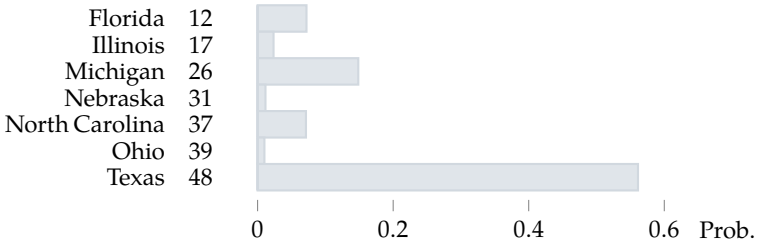
par_st_ori_fips

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	2,227,140	100.0
Valid records	2,000,680	89.8
Missing values	226,460	10.2

`par_st_ori_fips` (continued)

Histogram of `par_st_ori_fips` (valid records)



`par_cnty_ori_fips`

Label	County FIPS code for location of sentencing
Description	County-level Federal Information Processing Standards (FIPS) code for the location where the individual was sentenced. Used to help uniquely 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	2,227,140	100.0
Valid records	907,844	40.8
Missing values	1,092,836	49.1

Data notes

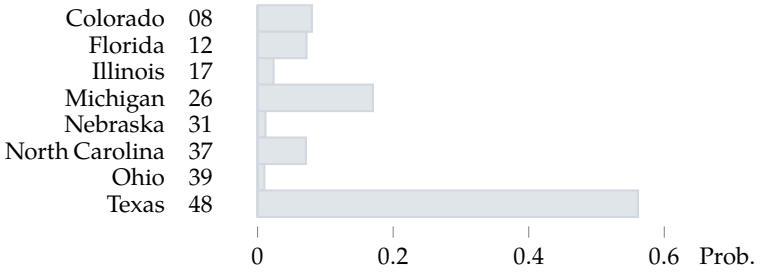
- Colorado: o6q

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	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_st_juris_fips` (valid records)

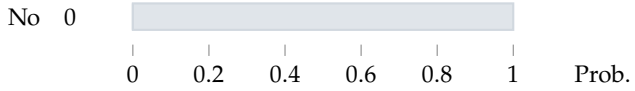


par_rec_src_le

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	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_rec_src_le` (valid records)

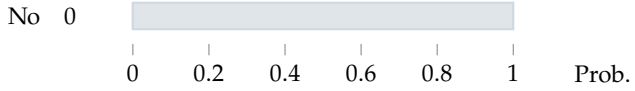


par_rec_src_crt

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

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_rec_src_crt` (valid records)

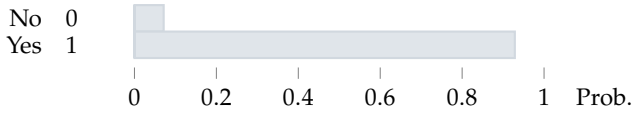


par_rec_src_doc

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

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_rec_src_doc` (valid records)

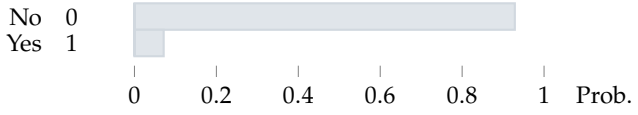


par_rec_src_rep

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

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_rec_src_rep` (valid records)

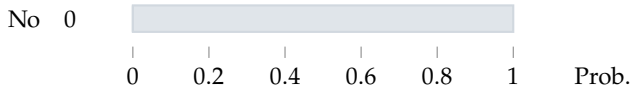


`par_rec_src_cc`

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

Set	*Count*	*Percent (%)*
All records	2,227,140	100.0
Valid records	2,227,140	100.0

Histogram of `par_rec_src_cc` (valid records)



3.7 ID variables used to link tables

The tables in the section above show that there are several ID variables contained in each of the CJARS data tables. The variables are used to identify individuals, events, episodes, and the linkages among these three. Refer to Sections 3.2.1 and 3.2.2 for more information about how these linkages are structured. Table 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.
arrest	`arr_id`	n/a	n/a	n/a

CJARS relational table	Key identifying event variable	Preceding event	Linking variable for preceding event	Description of linkage
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	Provider
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
|	County	Judiciary	Duval		Duval County Clerk of Courts
|	|	|	Hillsborough		Hillsborough County Clerk of Courts
|	|	|	Orange		Orange County Clerk of Court
|	|	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

State	Level	Domain	County	Municipality	Provider
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
|	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
|	|	|	|		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

State	Level	Domain	County	Municipality	Provider
					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
			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

State	Level	Domain	County	Municipality	Provider
					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
					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
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 8 gives an overview of the steps in the CJARS variable harmonization process. As can be seen from this figure, data processing is broken up into one of two divisions: PII and anonymized. The PII data is used for matching purposes while the anonymized data contains the information that is relevant for describing criminal justice events.

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

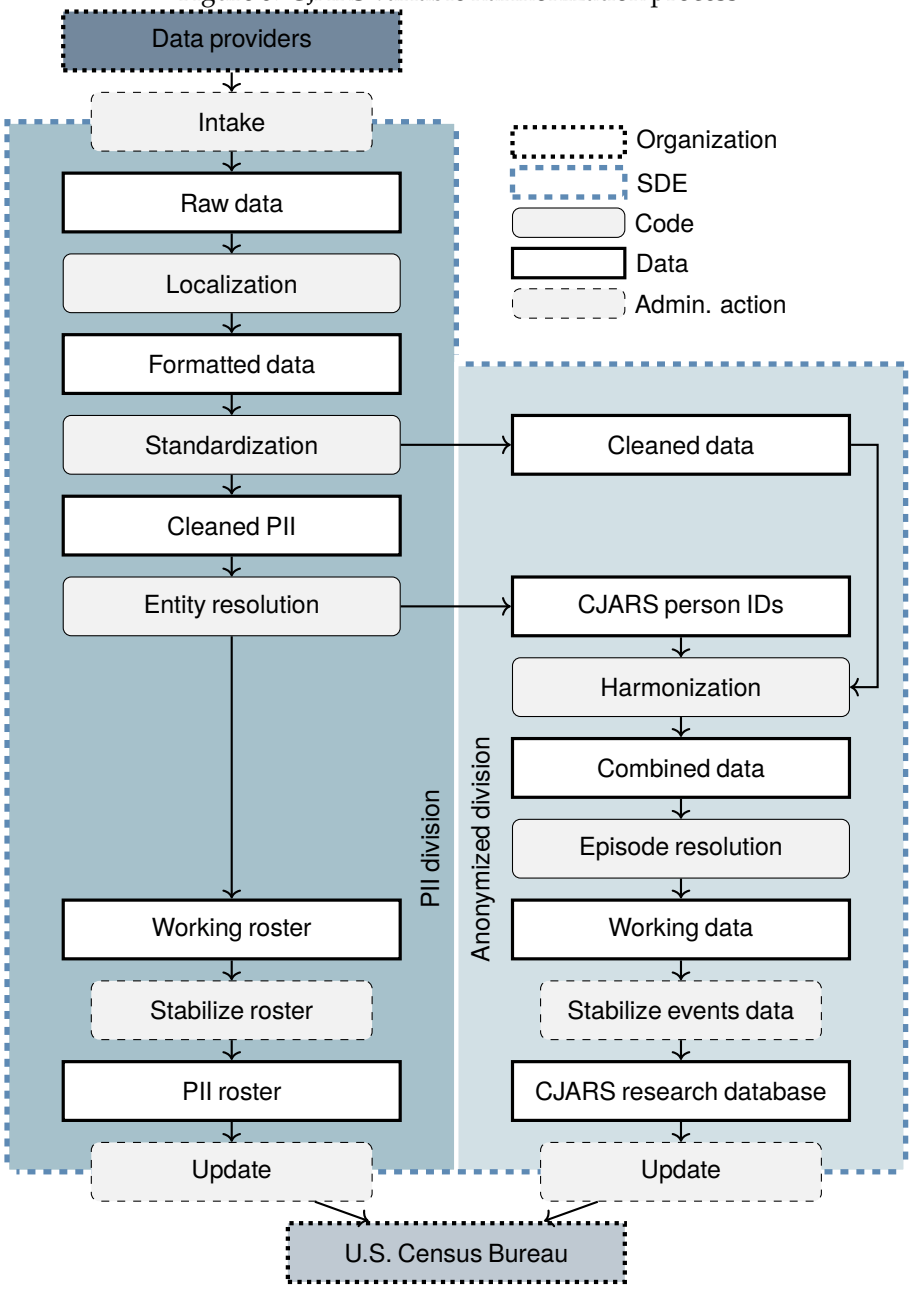
Table 5: Variable harmonization

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).

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 8: CJARS variable harmonization process



C Offense classification

There is no consistent method that criminal justice agencies use to describe the types of offenses committed that it records information about. For example, CJARS receives string descriptions of offenses, numeric or alphanumeric statute codes, charge grades, etc. These descriptors of offenses vary substantially across jurisdiction due to a number of factors, such as differences in legal statutes across jurisdictions, variation in nomenclature, differences in data input and storage, and other factors. Adding to the complexity of the issue is the variation in descriptions of offenses that exists within agencies because a common practice is to input offense descriptions as free format strings (which causes a number of barriers to classifications, such as misspellings). As a results, there are millions of unique offense descriptors in the raw data that CJARS has collected. The goal of the CJARS offense classification algorithm was to classify this huge number of unique descriptors into a common set of offense classifications (see ref{tab:app-code-schemes-offense_classification} for the offense classification scheme used by CJARS).

To make this task practical, CJARS leveraged machine learning techniques to assist with offense classification. The CJARS offense classification algorithm uses a hierarchical classification framework in which a multilayer perceptron classifier is trained for each local node. Hierarchical classification occurs in several steps. First, raw string descriptions are broken up into n=grams. Next these 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, and helped to avoid common issues that arise with classification tasks, such as the formations of superclusters.

The data taxonomy used to create the hierarchy of classifiers is predefined based on the offense classification scheme used by CJARS. This includes the following levels of classification: offense type code (parent class), offense category code (sub-parent class), and charge code (child class). In total, this schema comprises 104 classifiers (parent = 1, sub-parent = 12, child = 91) and 5 independent classifiers for charge flags (domestic violence, gang, gun, habitual offender, and inchoate).

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`

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

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	Other race/ethnicity
9	Missing

D.2 Criminal justice event codes

D.2.1 Offense classification

This scheme is used for the following variables:

- adj_chrg_off_cd
- adj_disp_off_cd
- arr_off_cd

Table 9: Offense classification scheme

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
1010	Murder	1	Murder	1	Violent
1011	Attempted murder	1	Murder	1	Violent
1012	Conspiracy to commit murder	1	Murder	1	Violent
1020	Unspecified homicide	2	Unspecified homicide	1	Violent
1021	Unspecified homicide, attempted	2	Unspecified homicide	1	Violent
1022	Unspecified homicide, conspiracy	2	Unspecified homicide	1	Violent
1030	Voluntary manslaughter	3	Voluntary/nonnegligent manslaughter	1	Violent
1031	Voluntary manslaughter, attempted	3	Voluntary/nonnegligent manslaughter	1	Violent
1032	Voluntary manslaughter,conspiracy	3	Voluntary/nonnegligent manslaughter	1	Violent
1040	Vehicular manslaughter	3	Voluntary/nonnegligent manslaughter	1	Violent
1041	Vehicular manslaughter, attempted	3	Voluntary/nonnegligent manslaughter	1	Violent
1042	Vehicular manslaughter, conspiracy	3	Voluntary/nonnegligent manslaughter	1	Violent
1050	Involuntary manslaughter	4	Manslaughter - non-vehicular	1	Violent
1051	Involuntary manslaughter, attempt	4	Manslaughter - non-vehicular	1	Violent
1052	Involuntary manslaughter, conspiracy	4	Manslaughter - non-vehicular	1	Violent
1060	Kidnapping	5	Kidnapping	1	Violent
1061	Kidnapping, attempted	5	Kidnapping	1	Violent
1062	Kidnapping, conspiracy	5	Kidnapping	1	Violent
1070	Rape	6	Rape - force	1	Violent
1071	Rape, attempted	6	Rape - force	1	Violent
1072	Rape, conspiracy	6	Rape - force	1	Violent

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
1080	Statutory rape	7	Rape - statutory - no force	1	Violent
1081	Statutory rape, attempted	7	Rape - statutory - no force	1	Violent
1082	Statutory rape, conspiracy	7	Rape - statutory - no force	1	Violent
1090	Child molestation	8	Lewd act with children	1	Violent
1091	Child molestation, attempted	8	Lewd act with children	1	Violent
1092	Child molestation, conspiracy	8	Lewd act with children	1	Violent
1100	Sexual assault	9	Sexual assault - other	1	Violent
1101	Sexual assault, attempted	9	Sexual assault - other	1	Violent
1102	Sexual assault, conspiracy	9	Sexual assault - other	1	Violent
1110	Human trafficking, sex - child	10	Human trafficking	1	Violent
1111	Human trafficking, sex - child, attempted	10	Human trafficking	1	Violent
1112	Human trafficking, sex - child, conspiracy	10	Human trafficking	1	Violent
1120	Human trafficking, sex - adult or no age specified	10	Human trafficking	1	Violent
1121	Human trafficking, sex - adult or no age specified, attempted	10	Human trafficking	1	Violent
1122	Human trafficking, sex - adult or no age specified, conspiracy	10	Human trafficking	1	Violent
1130	Human trafficking, labor - child	10	Human trafficking	1	Violent
1131	Human trafficking, labor - child, attempted	10	Human trafficking	1	Violent
1132	Human trafficking, labor - child, conspiracy	10	Human trafficking	1	Violent
1140	Human trafficking, labor - adult or no age specified	10	Human trafficking	1	Violent
1141	Human trafficking, labor - adult or no age specified, attempted	10	Human trafficking	1	Violent
1142	Human trafficking, labor - adult or no age specified, conspiracy	10	Human trafficking	1	Violent
1150	Human trafficking, unspecified - child	10	Human trafficking	1	Violent
1151	Human trafficking, unspecified - child, attempted	10	Human trafficking	1	Violent
1152	Human trafficking, unspecified - child, conspiracy	10	Human trafficking	1	Violent
1160	Human trafficking, unspecified - adult or no age specified	10	Human trafficking	1	Violent

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
1161	Human trafficking, unspecified - adult or no age specified, attempted	10	Human trafficking	1	Violent
1162	Human trafficking, unspecified - adult or no age specified, conspiracy	10	Human trafficking	1	Violent
1170	Human trafficking	10	Human trafficking	1	Violent
1171	Human trafficking, attempted	10	Human trafficking	1	Violent
1172	Human trafficking, conspiracy	10	Human trafficking	1	Violent
1180	Armed robbery	11	Armed robbery	1	Violent
1181	Armed robbery, attempted	11	Armed robbery	1	Violent
1182	Armed robbery, conspiracy	11	Armed robbery	1	Violent
1190	Unarmed robbery	12	Unarmed robbery	1	Violent
1191	Unarmed robbery, attempted	12	Unarmed robbery	1	Violent
1192	Unarmed robbery, conspiracy	12	Unarmed robbery	1	Violent
1200	Aggravated assault	13	Aggravated assault	1	Violent
1201	Aggravated assault, attempted	13	Aggravated assault	1	Violent
1202	Aggravated assault, conspiracy	13	Aggravated assault	1	Violent
1210	Assault of an officer	14	Assaulting public officer	1	Violent
1211	Assault of an officer, attempted	14	Assaulting public officer	1	Violent
1212	Assault of an officer, conspiracy	14	Assaulting public officer	1	Violent
1220	Child abuse	15	Child abuse	1	Violent
1221	Child abuse, attempted	15	Child abuse	1	Violent
1222	Child abuse, conspiracy	15	Child abuse	1	Violent
1230	Simple assault	16	Simple assault	1	Violent
1231	Simple assault, attempted	16	Simple assault	1	Violent
1232	Simple assault, conspiracy	16	Simple assault	1	Violent
1240	Extortion/threat	17	Blackmail/extortion/intimidation	1	Violent
1241	Extortion/threat, attempted	17	Blackmail/extortion/intimidation	1	Violent
1242	Extortion/threat, conspiracy	17	Blackmail/extortion/intimidation	1	Violent
1250	Hit and run with bodily injury	18	Hit and run driving - injury	1	Violent
1251	Hit and run with bodily injury, attempted	18	Hit and run driving - injury	1	Violent
1252	Hit and run with bodily injury, conspiracy	18	Hit and run driving - injury	1	Violent
1990	Violent offense, other	19	Violent offenses - other	1	Violent
1991	Violent offense other, attempted	19	Violent offenses - other	1	Violent
1992	Violent offense other, conspiracy	19	Violent offenses - other	1	Violent
2010	Burglary	20	Burglary	2	Property
2011	Burglary, attempted	20	Burglary	2	Property
2012	Burglary, conspiracy	20	Burglary	2	Property
2020	Arson	21	Arson	2	Property

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
2021	Arson, attempted	21	Arson	2	Property
2022	Arson, conspiracy	21	Arson	2	Property
2030	Auto theft	22	Auto theft	2	Property
2031	Auto theft, attempted	22	Auto theft	2	Property
2032	Auto theft, conspiracy	22	Auto theft	2	Property
2040	Forgery/fraud	23	Forgery/fraud	2	Property
2041	Forgery/fraud, attempted	23	Forgery/fraud	2	Property
2042	Forgery/fraud, conspiracy	23	Forgery/fraud	2	Property
2050	Grand theft (>$500)	24	Grand larceny - theft over $500	2	Property
2051	Grand theft (>$500), attempted	24	Grand larceny - theft over $500	2	Property
2052	Grand theft (>$500), conspiracy	24	Grand larceny - theft over $500	2	Property
2060	Petty theft (=<$500)	25	Petty larceny - theft equal or under $500	2	Property
2061	Petty theft (=<$500), attempted	25	Petty larceny - theft equal or under $500	2	Property
2062	Petty theft (=<$500), conspiracy	25	Petty larceny - theft equal or under $500	2	Property
2070	Theft, value unknown	26	Larceny/theft - value unknown	2	Property
2071	Theft, value unknown, attempted	26	Larceny/theft - value unknown	2	Property
2072	Theft, value unknown, conspiracy	26	Larceny/theft - value unknown	2	Property
2080	Financial crimes	27	Financial crimes	2	Property
2081	Financial crimes attempted	27	Financial crimes	2	Property
2082	Financial crimes conspiracy	27	Financial crimes	2	Property
2090	Sale of stolen property	28	Stolen property - trafficking	2	Property
2091	Sale of stolen property, attempted	28	Stolen property - trafficking	2	Property
2092	Sale of stolen property, conspiracy	28	Stolen property - trafficking	2	Property
2100	Receiving stolen property	29	Stolen property - receiving	2	Property
2101	Receiving stolen property, attempted	29	Stolen property - receiving	2	Property
2102	Receiving stolen property, conspiracy	29	Stolen property - receiving	2	Property
2110	Destruction of property	30	Destruction of property	2	Property
2111	Destruction of property, attempted	30	Destruction of property	2	Property
2112	Destruction of property, conspiracy	30	Destruction of property	2	Property
2120	Hit and run driving with property damage	31	Hit and run driving - property damage	2	Property
2121	Hit and run driving, attempted	31	Hit and run driving - property damage	2	Property
2122	Hit and run driving, conspiracy	31	Hit and run driving - property damage	2	Property

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
2130	Unauthorized use of vehicle	32	Unauthorized use of vehicle	2	Property
2131	Unauthorized use of vehicle, attempted	32	Unauthorized use of vehicle	2	Property
2132	Unauthorized use of vehicle, conspiracy	32	Unauthorized use of vehicle	2	Property
2140	Criminal trespass	33	Trespassing	2	Property
2141	Criminal trespass, attempted	33	Trespassing	2	Property
2142	Criminal trespass, conspiracy	33	Trespassing	2	Property
2150	Possession of property crime tools	34	Property offenses - other	2	Property
2151	Possession of property crime tools, attempted	34	Property offenses - other	2	Property
2152	Possession of property crime tools, conspiracy	34	Property offenses - other	2	Property
2990	Other property offense	34	Property offenses - other	2	Property
2991	Other property offense, attempt	34	Property offenses - other	2	Property
2992	Other property offense, conspiracy	34	Property offenses - other	2	Property
3010	Distribution heroin	35	Distribution - heroin	3	Drug
3011	Distribution, heroin, attempted	35	Distribution - heroin	3	Drug
3012	Distribution, heroin, conspiracy	35	Distribution - heroin	3	Drug
3020	Distribution of amphetamines	36	Distribution - amphetamines	3	Drug
3021	Distribution of amphetamines, attempted	36	Distribution - amphetamines	3	Drug
3022	Distribution of amphetamines, conspiracy	36	Distribution - amphetamines	3	Drug
3030	Distribution cocaine or crack	37	Distribution - cocaine or crack	3	Drug
3031	Distribution cocaine or crack, attempted	37	Distribution - cocaine or crack	3	Drug
3032	Distribution cocaine or crack, conspiracy	37	Distribution - cocaine or crack	3	Drug
3040	Distribution of opioids	38	Distribution of opioids	3	Drug
3041	Distribution of opioids, attempted	38	Distribution of opioids	3	Drug
3042	Distribution of opioids, conspiracy	38	Distribution of opioids	3	Drug
3050	Distribution of prescription drugs	39	Distribution of prescription drugs	3	Drug
3051	Distribution of prescription drugs, attempted	39	Distribution of prescription drugs	3	Drug
3052	Distribution of prescription drugs, conspiracy	39	Distribution of prescription drugs	3	Drug
3060	Distribution other controlled substances	40	Distribution - other controlled substances	3	Drug

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
3061	Distribution other controlled substances, attempted	40	Distribution - other controlled substances	3	Drug
3062	Distribution other controlled substances, conspiracy	40	Distribution - other controlled substances	3	Drug
3070	Distribution marijuana	41	Distribution marijuana/hashish	3	Drug
3071	Distribution marijuana, attempted	41	Distribution marijuana/hashish	3	Drug
3072	Distribution marijuana, conspiracy	41	Distribution marijuana/hashish	3	Drug
3080	Distribution, drug unspecified	42	Distribution - drug unspecified	3	Drug
3081	Distribution, drug unspecified, attempted	42	Distribution - drug unspecified	3	Drug
3082	Distribution, drug unspecified, conspiracy	42	Distribution - drug unspecified	3	Drug
3090	Possession/use of heroin	43	Possession/use - heroin	3	Drug
3091	Possession/use of heroin, attempted	43	Possession/use - heroin	3	Drug
3092	Possession/use of heroin, conspiracy	43	Possession/use - heroin	3	Drug
3100	Possession of amphetamines	44	Possession of amphetamines	3	Drug
3101	Possession of amphetamines, attempted	44	Possession of amphetamines	3	Drug
3102	Possession of amphetamines, conspiracy	44	Possession of amphetamines	3	Drug
3110	Possession/use of cocaine or crack	45	Possession/use - cocaine or crack	3	Drug
3111	Possession/use of cocaine or crack, attempted	45	Possession/use - cocaine or crack	3	Drug
3112	Possession/use of cocaine or crack, conspiracy	45	Possession/use - cocaine or crack	3	Drug
3120	Possession of opioids	46	Possession of opioids	3	Drug
3121	Possession of opioids, attempted	46	Possession of opioids	3	Drug
3122	Possession of opioids, conspiracy	46	Possession of opioids	3	Drug
3130	Possession of prescription drugs	47	Possession of prescription drugs	3	Drug
3131	Possession of prescription drugs, attempted	47	Possession of prescription drugs	3	Drug
3132	Possession of prescription drugs, conspiracy	47	Possession of prescription drugs	3	Drug
3140	Possession/use of other controlled substance	48	Possession/use - other controlled substances	3	Drug
3141	Possession/use of other controlled substance, attempted	48	Possession/use - other controlled substances	3	Drug

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
3142	Possession/use of other controlled substance, conspiracy	48	Possession/use - other controlled substances	3	Drug
3150	Possession/use of marijuana	49	Possession/use - marijuana/hashish	3	Drug
3151	Possession/use of marijuana, attempted	49	Possession/use - marijuana/hashish	3	Drug
3152	Possession/use of marijuana, conspiracy	49	Possession/use - marijuana/hashish	3	Drug
3160	Possession/use of unspecified drug	50	Possession/use - drug unspecified	3	Drug
3161	Possession/use, drug unspecified, attempted	50	Possession/use - drug unspecified	3	Drug
3162	Possession/use, drug unspecified, conspiracy	50	Possession/use - drug unspecified	3	Drug
3170	Heroin violation, offense unspecified	51	Heroin violation - offense unspecified	3	Drug
3180	Amphetamines, offense unspecified	52	Amphetamines - offense unspecified	3	Drug
3190	Cocaine/crack violation, offense unspecified	53	Cocaine or crack violation offense unspecified	3	Drug
3200	Prescription of opioid drugs, offense unspecified	54	Prescription of opioid drugs - offense unspecified	3	Drug
3210	Prescription, offense unspecified	55	Prescription - offense unspecified	3	Drug
3220	Other controlled substance violation, offense unspecified	56	Controlled substance - offense unspecified	3	Drug
3230	Marijuana violation, offense unspecified	57	Marijuana/hashish violation - offense unspecified	3	Drug
3240	Fraudulent drug offense	58	Other drug offense/paraphernalia	3	Drug
3241	Fraudulent drug offense, attempted	58	Other drug offense/paraphernalia	3	Drug
3242	Fraudulent drug offense, conspiracy	58	Other drug offense/paraphernalia	3	Drug
3250	Drug paraphernalia	58	Other drug offense/paraphernalia	3	Drug
3251	Drug paraphernalia, attempted	58	Other drug offense/paraphernalia	3	Drug
3252	Drug paraphernalia, conspiracy	58	Other drug offense/paraphernalia	3	Drug
3990	Other drug offense	58	Other drug offense/paraphernalia	3	Drug
3991	Other drug offense, attempt	58	Other drug offense/paraphernalia	3	Drug
3992	Other drug offense, conspiracy	58	Other drug offense/paraphernalia	3	Drug
4010	Driving while intoxicated	59	Driving while intoxicated	4	DUI offense
4011	Driving while intoxicated, attempted	59	Driving while intoxicated	4	DUI offense

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
4012	Driving while intoxicated, conspiracy	59	Driving while intoxicated	4	DUI offense
4020	Driving under the influence of alcohol	60	Driving under the influence	4	DUI offense
4021	Driving under the influence of alcohol, attempted	60	Driving under the influence	4	DUI offense
4022	Driving under the influence of alcohol, conspiracy	60	Driving under the influence	4	DUI offense
4030	Driving under the influence of drugs	61	Driving under influence - drugs	4	DUI offense
4031	Driving under the influence of drugs, attempted	61	Driving under influence - drugs	4	DUI offense
4032	Driving under the influence of drugs, conspiracy	61	Driving under influence - drugs	4	DUI offense
5010	Riot	62	Rioting	5	Public order
5011	Riot, attempting to incite	62	Rioting	5	Public order
5012	Riot, conspiracy to incite	62	Rioting	5	Public order
5020	Escape from custody	63	Escape from custody	5	Public order
5021	Escape from custody, attempted	63	Escape from custody	5	Public order
5022	Escape from custody, conspiracy	63	Escape from custody	5	Public order
5030	Flight to avoid prosecution	64	Flight to avoid prosecution	5	Public order
5031	Flight to avoid prosecution, attempted	64	Flight to avoid prosecution	5	Public order
5032	Flight to avoid prosecution, conspiracy	64	Flight to avoid prosecution	5	Public order
5040	Weapons offense	65	Weapon offense	5	Public order
5041	Weapons offense, attempted	65	Weapon offense	5	Public order
5042	Weapons offense, conspiracy	65	Weapon offense	5	Public order
5050	Habitual offender	66	Habitual offender	5	Public order
5060	Parole violation	67	Parole violation	5	Public order
5070	Probation violation	68	Probation violation	5	Public order
5080	Contempt of court/violate court order	69	Contempt of court	5	Public order
5081	Contempt of court/violate court order, attempted	69	Contempt of court	5	Public order
5082	Contempt of court/violate court order, conspiracy	69	Contempt of court	5	Public order
5090	Other court offense	70	Offenses against courts, legislatures and commissions	5	Public order

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
5091	Other court offense, attempted	70	Offenses against courts, legislatures and commissions	5	Public order
5092	Other court offense, conspiracy	70	Offenses against courts, legislatures and commissions	5	Public order
5100	Family or custody related offense	71	Family related offenses	5	Public order
5101	Family or custody related offense, attempted	71	Family related offenses	5	Public order
5102	Family or custody related offense, conspiracy	71	Family related offenses	5	Public order
5110	Offense against morals/decency	72	Morals/decency - offense	5	Public order
5111	Offense against morals/decency, attempted	72	Morals/decency - offense	5	Public order
5112	Offense against morals/decency, conspiracy	72	Morals/decency - offense	5	Public order
5120	Immigration violation	73	Immigration violations	5	Public order
5121	Immigration violation, attempted	73	Immigration violations	5	Public order
5122	Immigration violation, conspiracy	73	Immigration violations	5	Public order
5130	Obstruction/resisting	74	Obstruction - law enforcement	5	Public order
5131	Obstruction/resisting, attempted	74	Obstruction - law enforcement	5	Public order
5132	Obstruction/resisting, conspiracy	74	Obstruction - law enforcement	5	Public order
5140	Invasion of privacy	75	Invasion of privacy	5	Public order
5141	Invasion of privacy, attempted	75	Invasion of privacy	5	Public order
5142	Invasion of privacy, conspiracy	75	Invasion of privacy	5	Public order
5150	Commercialized vice	76	Commercialized vice	5	Public order
5151	Commercialized vice, attempted	76	Commercialized vice	5	Public order
5152	Commercialized vice, conspiracy	76	Commercialized vice	5	Public order
5160	Contributing to the delinquency of a minor	77	Contributing to delinquency of a minor	5	Public order
5161	Contributing to the delinquency of a minor, attempted	77	Contributing to delinquency of a minor	5	Public order
5162	Contributing to the delinquency of a minor, conspiracy	77	Contributing to delinquency of a minor	5	Public order
5170	Disorderly conduct offense	78	Drunkenness/vagrancy/disorderly conduct	5	Public order
5171	Disorderly conduct offense, attempted	78	Drunkenness/vagrancy/disorderly conduct	5	Public order
5172	Disorderly conduct offense, conspiracy	78	Drunkenness/vagrancy/disorderly conduct	5	Public order
5180	Liquor law violation	79	Liquor law violations	5	Public order
5181	Liquor law violation, attempted	79	Liquor law violations	5	Public order

Offense code	Offense code description	Offense category code	Offense category description	Offense type code	Offense type description
5182	Liquor law violation, conspiracy	79	Liquor law violations	5	Public order
5190	Taxation offense	80	Taxation offenses	5	Public order
5191	Taxation offense, attempted	80	Taxation offenses	5	Public order
5192	Taxation offense, conspiracy	80	Taxation offenses	5	Public order
5200	Bribery/conflict of interest	81	Bribery and conflict of interest	5	Public order
5201	Bribery/conflict of interest, attempt	81	Bribery and conflict of interest	5	Public order
5202	Bribery/conflict of interest, conspiracy	81	Bribery and conflict of interest	5	Public order
5990	Public order offense, other	82	Public order offenses - other	5	Public order
5991	Public order offense, other, attempted	82	Public order offenses - other	5	Public order
5992	Public order offense, other, conspiracy	82	Public order offenses - other	5	Public order
6010	Traffic offense, minor	84	Traffic offenses - minor	6	Criminal traffic
9010	Unspecified felony	85	Felony - unspecified	999	Not known/missing
9011	Unspecified felony, attempt	85	Felony - unspecified	999	Not known/missing
9012	Unspecified felony, conspiracy	85	Felony - unspecified	999	Not known/missing
9020	Unspecified misdemeanor	86	Misdemeanor - unspecified	999	Not known/missing
9021	Unspecified misdemeanor, attempt	86	Misdemeanor - unspecified	999	Not known/missing
9022	Unspecified misdemeanor, conspiracy	86	Misdemeanor - unspecified	999	Not known/missing
9990	Other offense	87	Other	7	Other
8010	Juvenile offense	83	Juvenile offenses		Exclude
8020	Flag for removal	500	Flag for removal	500	Flag for removal
8030	Call for service	505	Call for service	505	Call for service
8040	Federal charges	510	Federal charges	510	Federal charges
8050	Variable not available in county	777	Variable not available in county	777	Variable not available in county
8060	Not applicable	888	Not applicable	888	Not applicable
9999	Not known/missing	999	Not known/missing	999	Not known/missing

D.2.2 Offense charge grade

This scheme is used for the following variables:

- adj_grd_cd

Table 10: 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 11: 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 12: 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 13: 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 14: Probation exit classification scheme

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

D.2.7 Incarceration entry

This scheme is used for the following variables:

- inc_entry_cd

Table 15: 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

Incarceration entry code	Incarceration entry description
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 16: Incarceration facility type classification scheme

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

D.2.9 Incarceration exit

This scheme is used for the following variables:

- inc_exit_cd

Table 17: 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

Incarceration exit code	Incarceration exit description
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 18: Parole exit classification scheme

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

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

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.

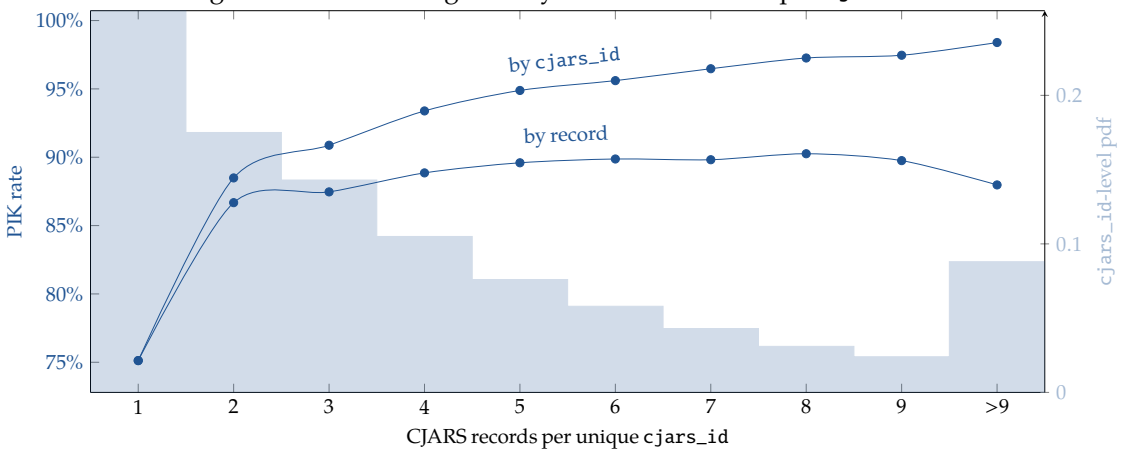
The roster file and anonymized criminal justice database are then transferred securely to the Census Bureau, where they are processed by the Person Identification Validation System (PVS), the Census Bureau's probabilistic record linkage system that assigns anonymized Personal Identification Keys (PIKs) 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 on the Social Security Administration's Numident file. For example, arrests records have lower quality PII on average, so we expect lower match rates. And in regions with a higher number of immigrants *who have not been assigned Social Security Numbers or Individual Taxpayer Identification Numbers*, we also expect lower match rates.

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

Figure 9 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 9: 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. For `cjars_id`s that have non-alias PII rows, where at least one non-alias row was assigned a PIK, delete all PII rows associated with 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 is preparing a proposal development guide that will be available on the CJARS project website by the end of 2020.

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
qfp		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.
yik		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.
b5n		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).
dor		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
z6t		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.
btp		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.
mix		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
tzh		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).
j8o		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
kjh		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).
zr5		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).
twd		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).
fei		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
jdt	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.
qho	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.
qlu	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
f7i	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
dd7	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.
lrw	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
t6z	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.
twf	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.
lao	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.
ldw	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.
myf	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.
rsq	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.
f41	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
fe2	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.
nbh	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
wt8	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.
o6q	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
ny1	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.
hn4	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
u2g	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.
fdw	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
njg	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.
kr7	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
pya	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
e7v	Maryland Judiciary	adj_sent_dt_yyyy, adj_sent_dt_mm, adj_sent_dt_dd	Data sourced from the Maryland Judiciary includes case record information. Sentencing date information is missing.

G.10 Michigan data notes

Note ID	Data provider	Variables	Notes
ydo	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
qoz	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
sc5	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.
wl2	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
18c	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
rl7	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.
hsi	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
se3	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
r92	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.
q57	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
xwn	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.
p4u	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.
t0b	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.
t4z	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
xe4	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
lxn	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
bwa	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
ihz	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.
yb0	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.
kke	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.
pzl	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.
w82	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.
qc4	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
jsg	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.
g5n	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.
pr2	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.
qfd	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.
k8r	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.
hd0	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.
uab	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
s9m	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.
s8m	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.
qs8	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.
cjk	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.
lrk	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.
ce6	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
vbm	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.
bgp	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.
y69	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.
j6j	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
acs	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.
l2h	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.
cro	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.
e0d	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.
c74	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 observarions in some data extracts: inc_fcl_cd, inc_entry_cd, and inc_exit_cd.

Note ID	Data provider	Variables	Notes
q8v	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.
ux0	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.
vsl	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.
ip7	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 Washington data notes

Note ID	Data provider	Variables	Notes
`fpm`	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.21 Wisconsin data notes

Note ID	Data provider	Variables	Notes
`ffx`	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.
`jf0`	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 10 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 11 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 12.

[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 10: Arizona ACCH & DOC AIMS database model (presumed)

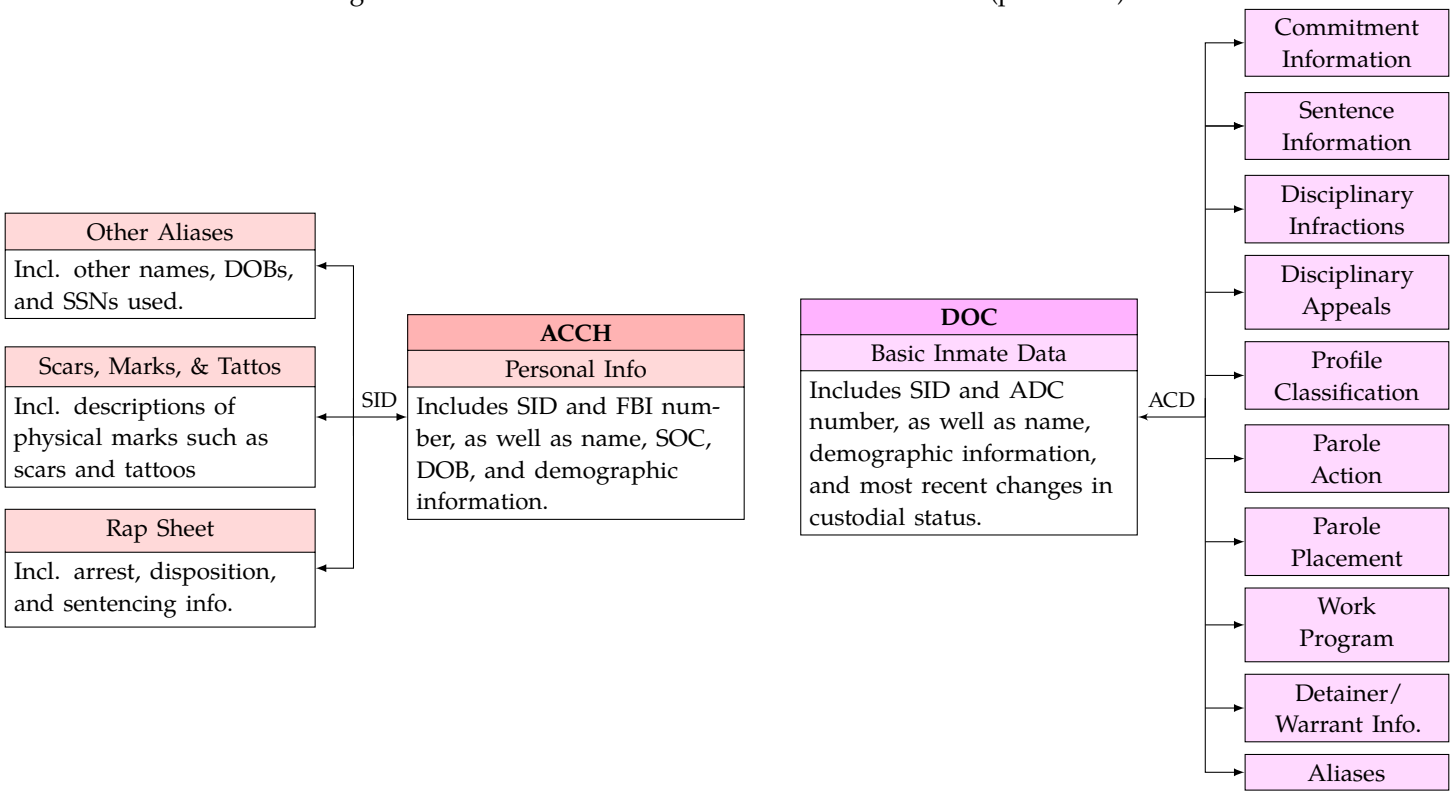


Figure 11: Description of some tables from California ACHS

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)

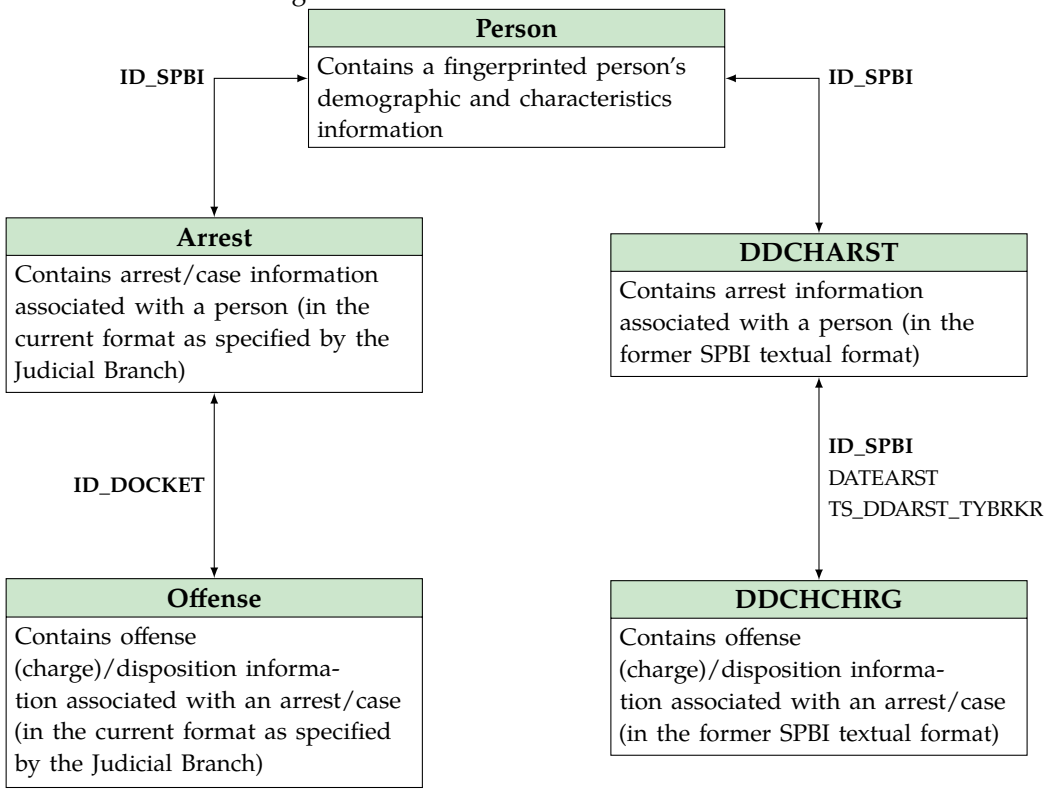
Community Corrections

- 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 12: 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 13, 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 13: Florida DOC Public Database ERD (presumed)

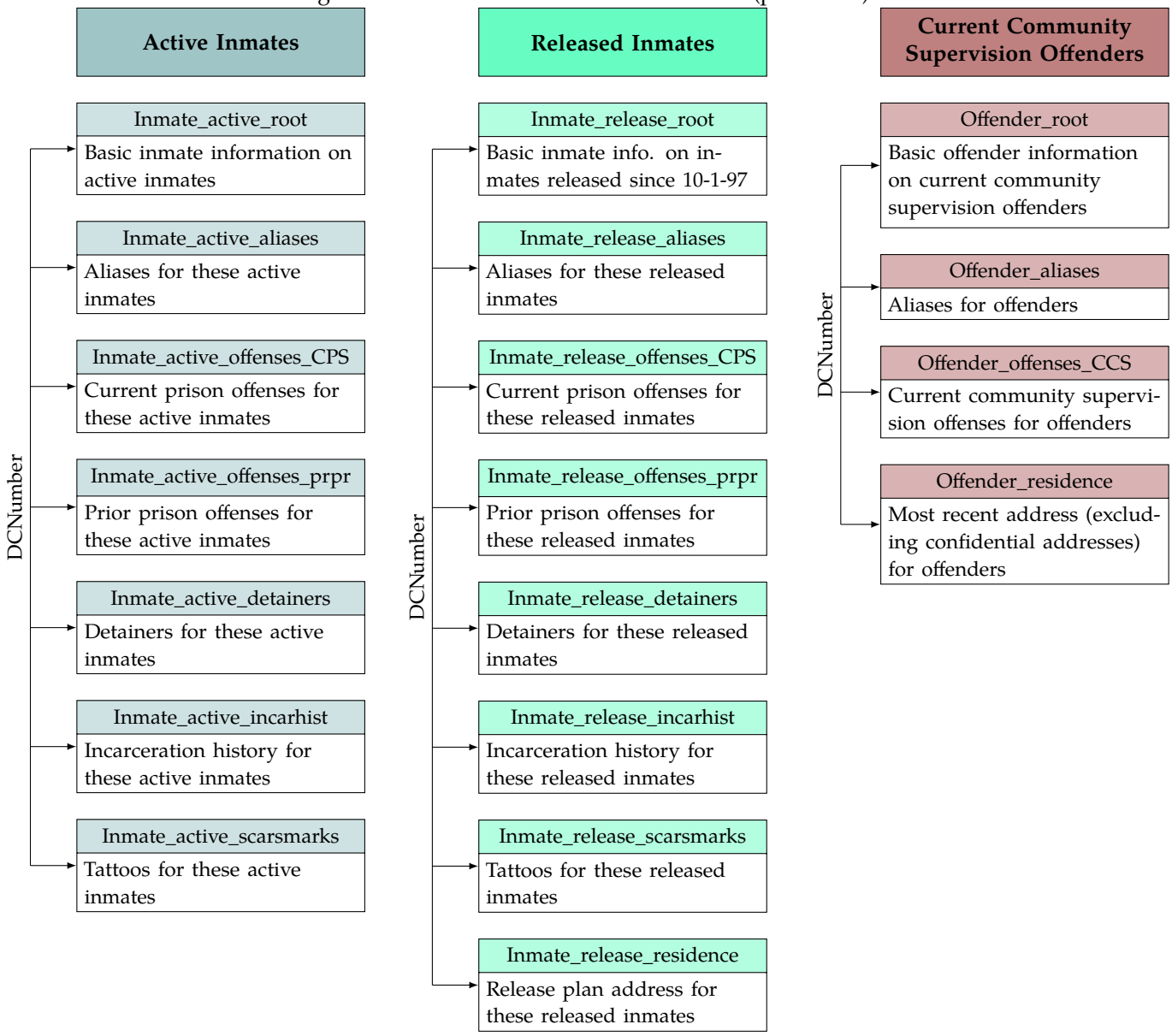


Figure 14) for each year between 1990-2005.[3] Adult and juvenile records are stored in separate yearly databases.

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.

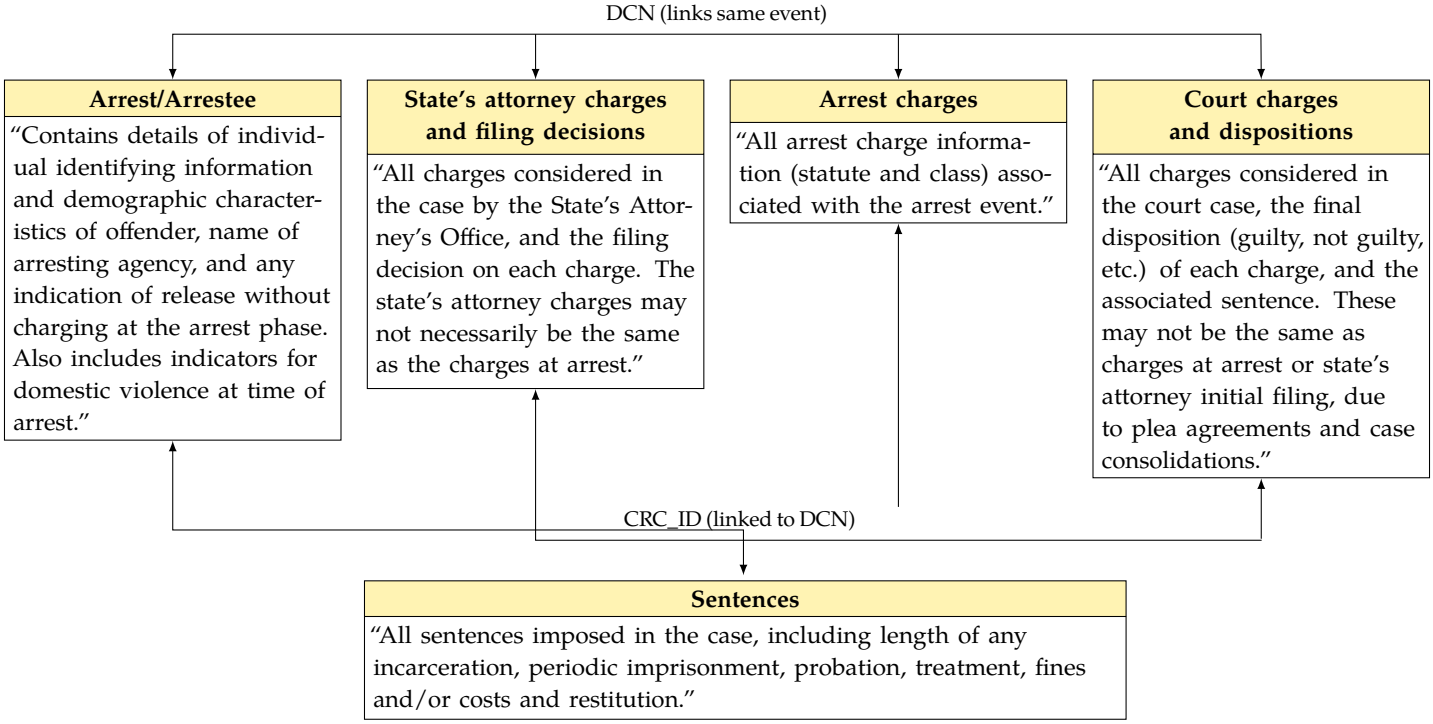
H.6 Kansas

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 14: 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 15). 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 15 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 16.

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

Figure 15: Kansas Disposition Report (KDR) Class Diagrams (presumed)

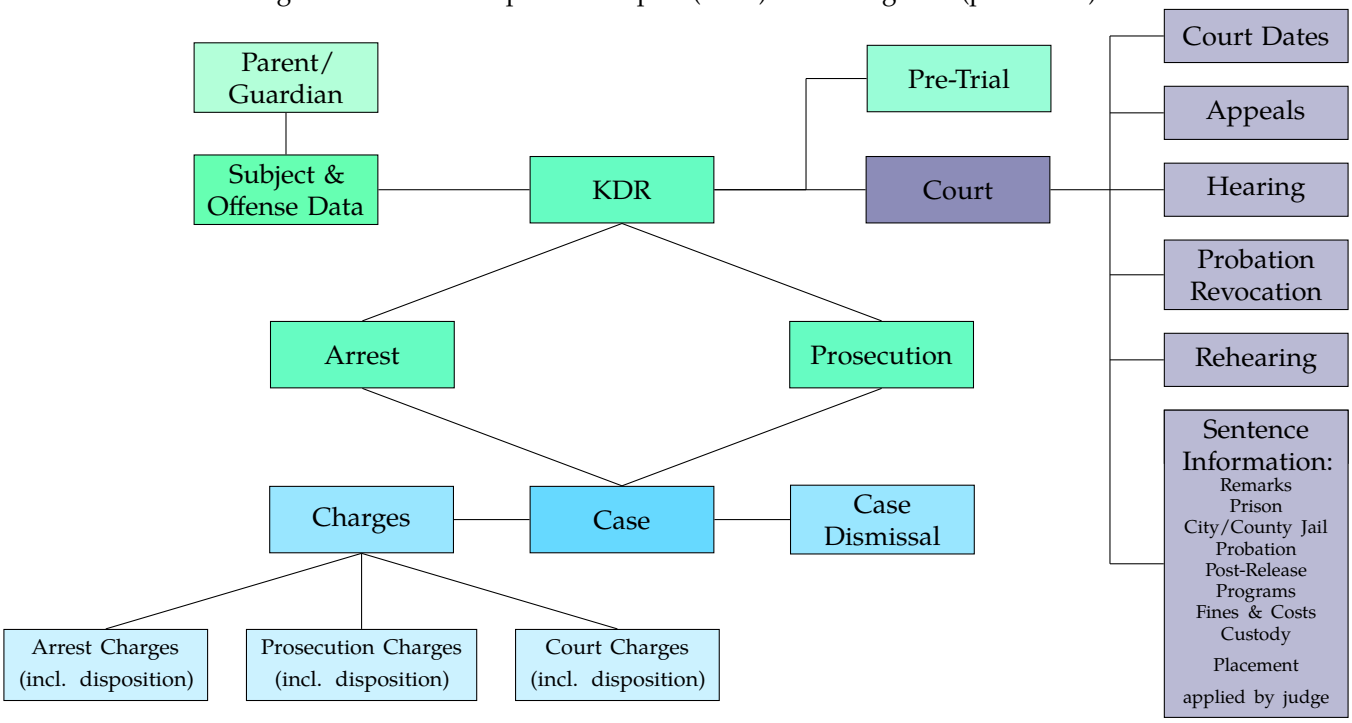
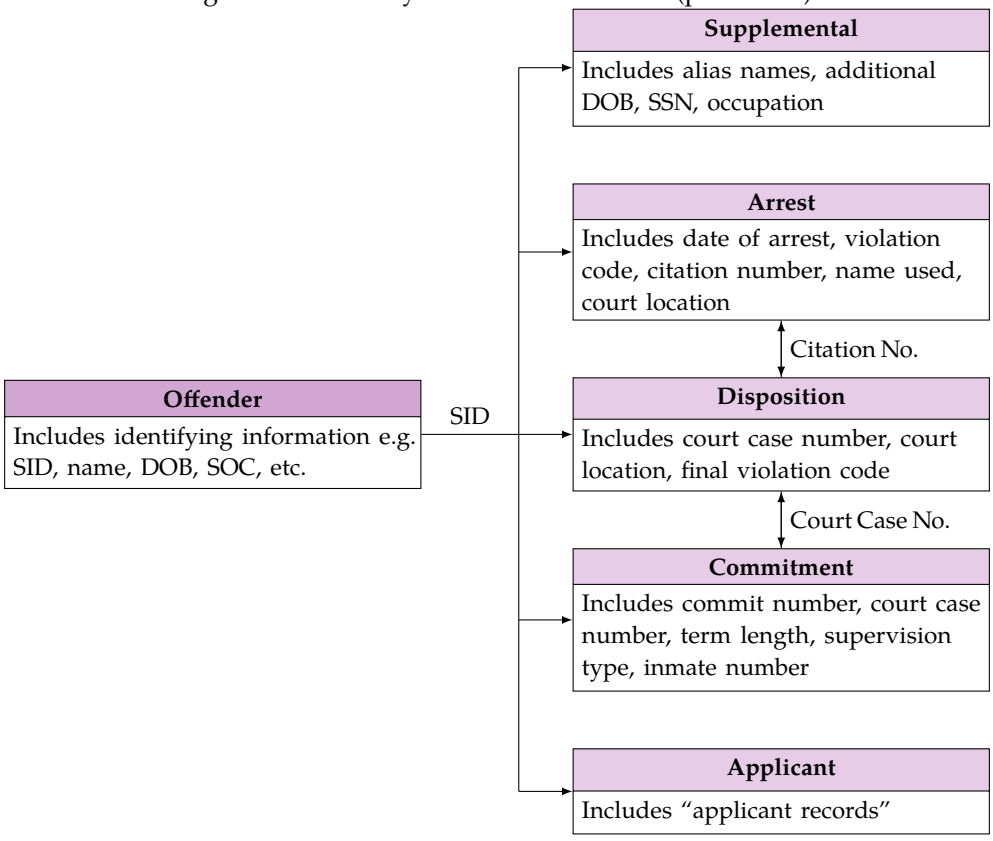


Figure 16: Kentucky CCH database model (presumed)

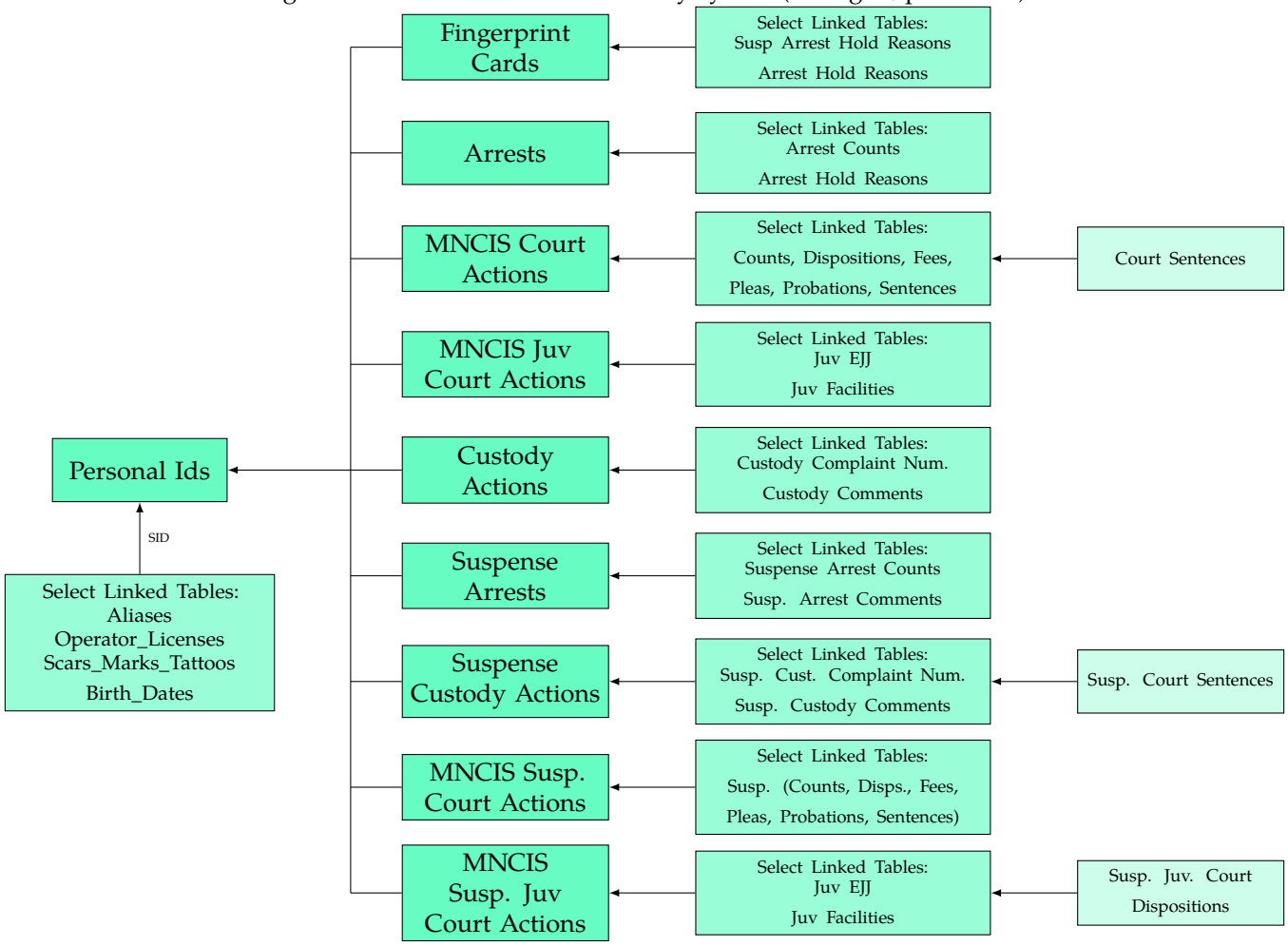


H.8 Minnesota

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 17 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 17: Minnesota Criminal History System (abridged/presumed)



H.9 Nevada

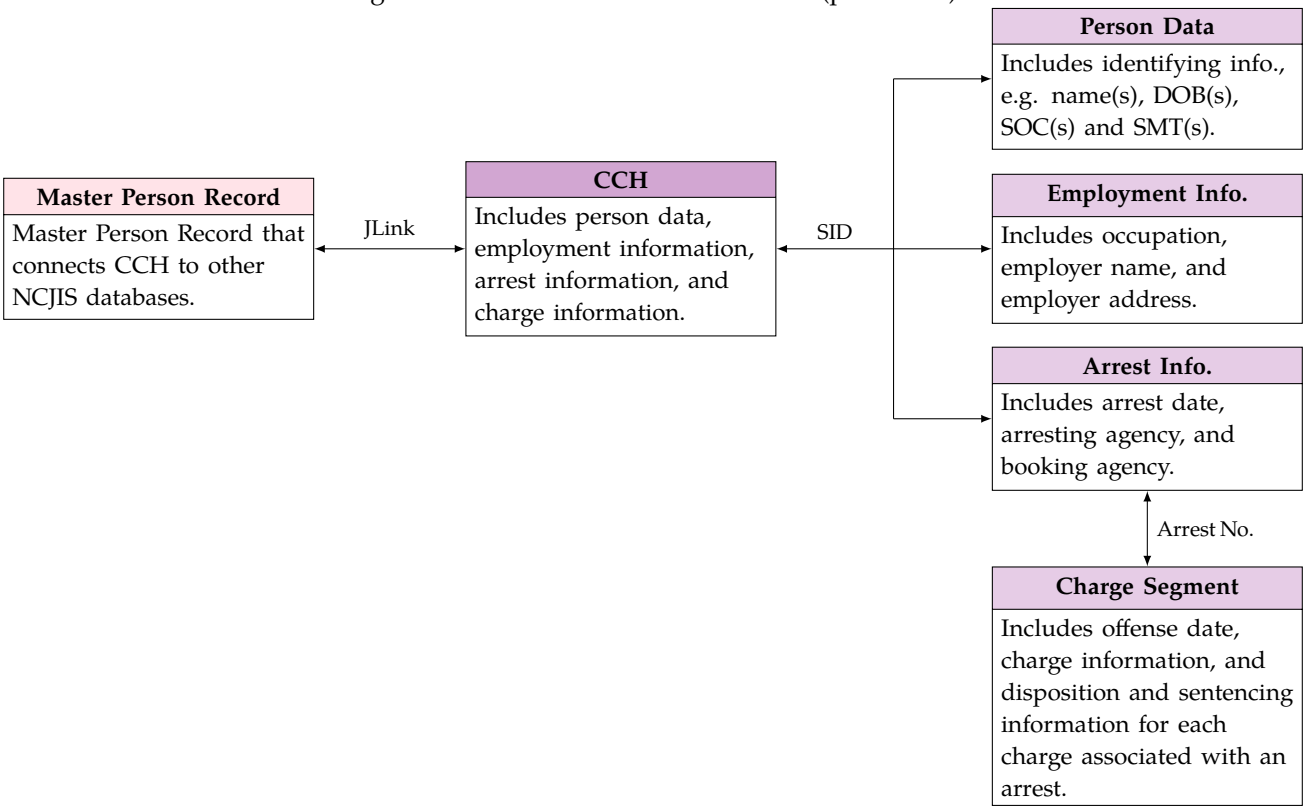
The Computerized Criminal History (CCH) is part of the Nevada Criminal Justice Information System (NCJIS) environment.[5] CCH (also referred to as the Nevada Criminal History Repository) acts as a centralized storage facility for Nevada Arrest and Disposition Records. It is maintained by the Records 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 18 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 18: Nevada CCH database model (presumed)



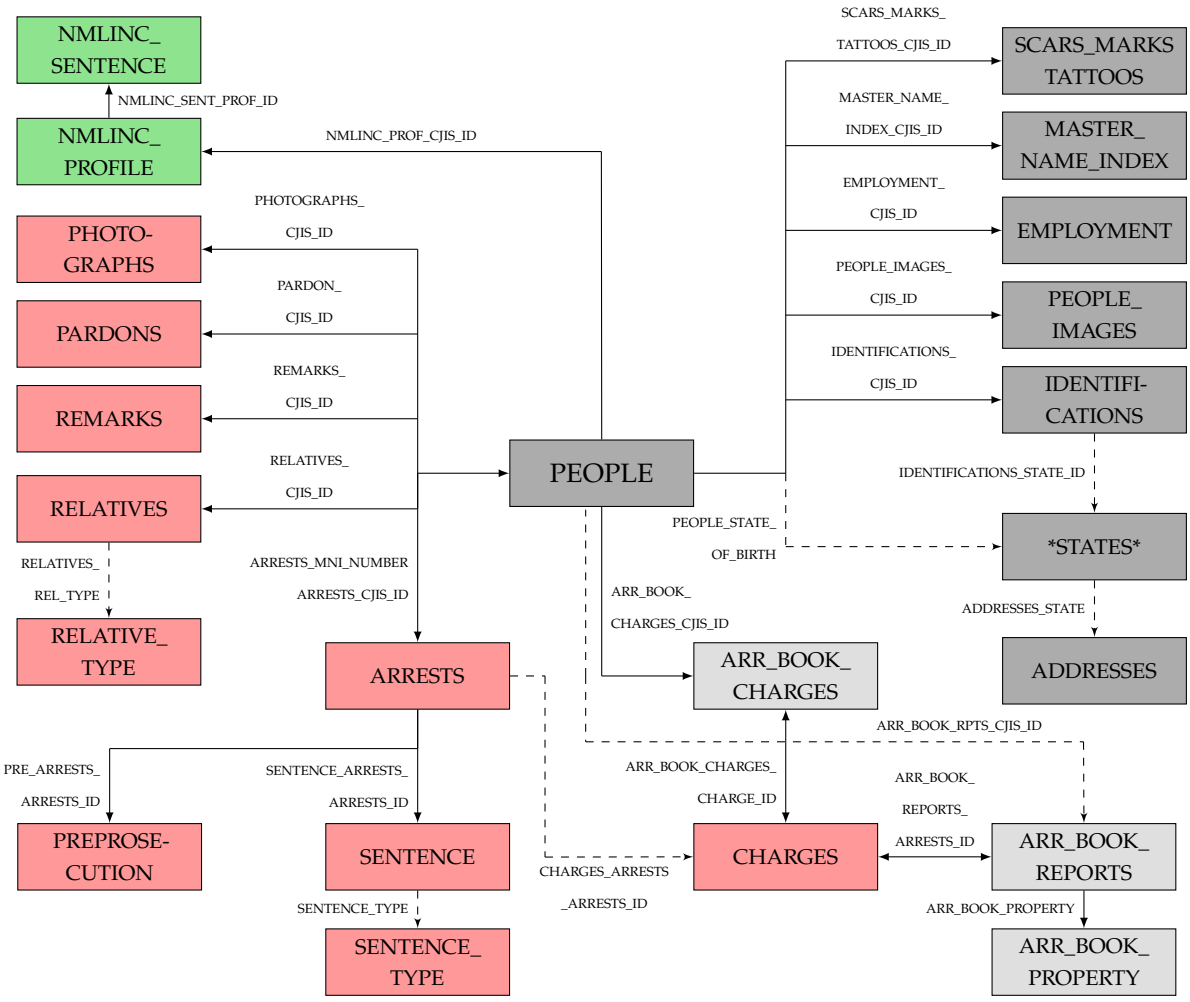
H.10 New Mexico

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 19 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 19: New Mexico ERD (abridged)



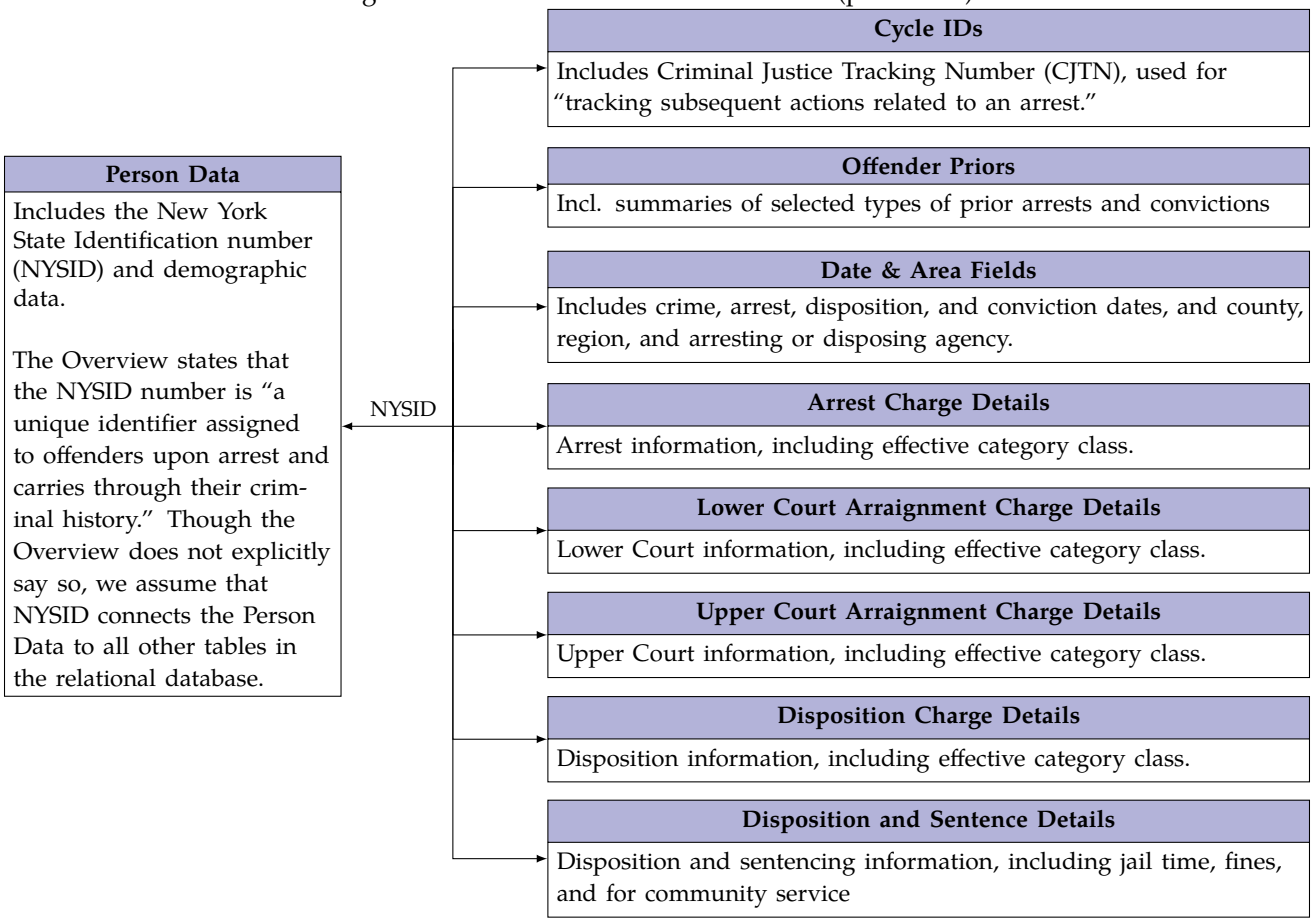
H.11 New York

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 20, 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.

Figure 20: New York CCH database model (presumed)



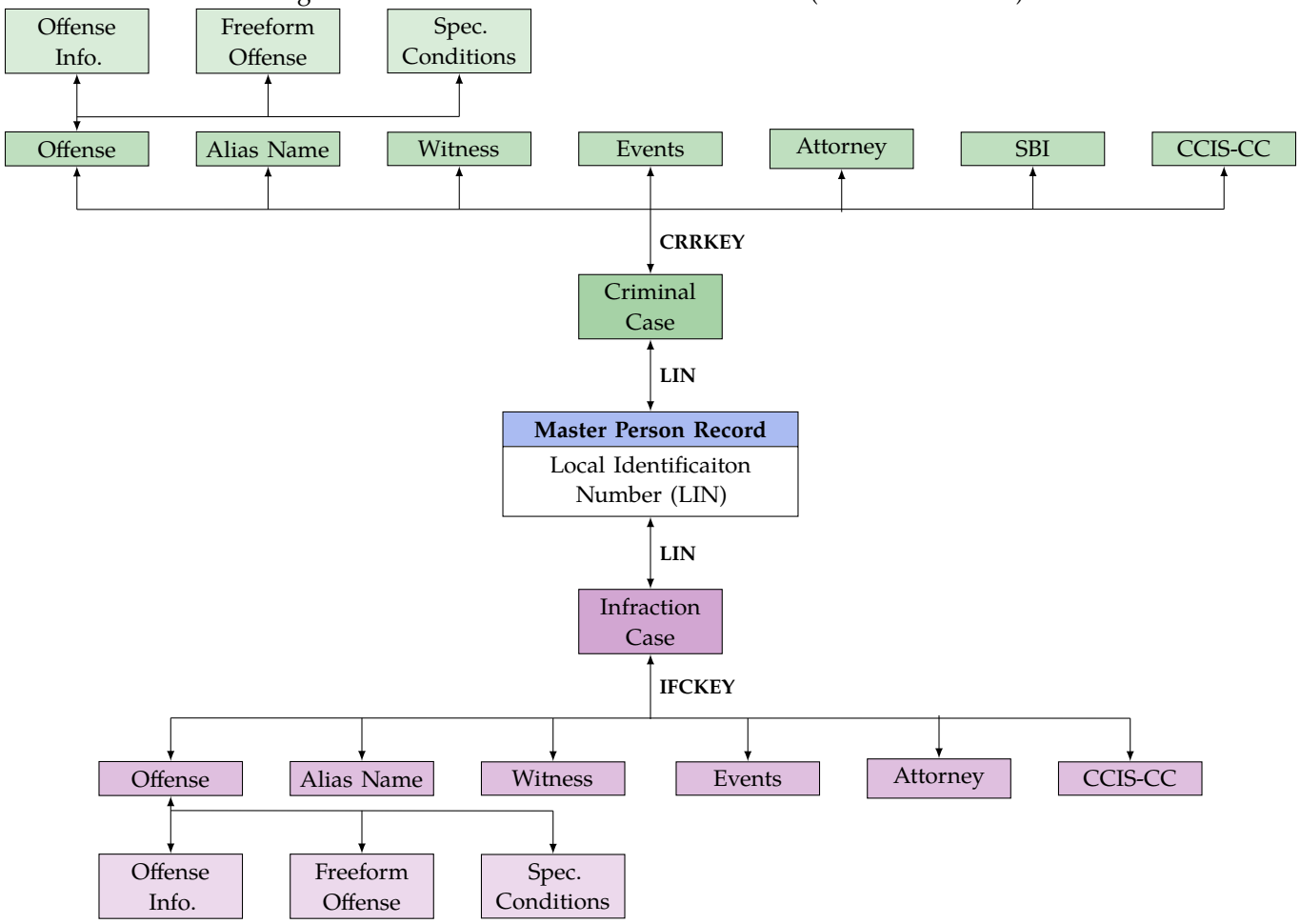
H.12 North Carolina

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

Figure 21: 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.

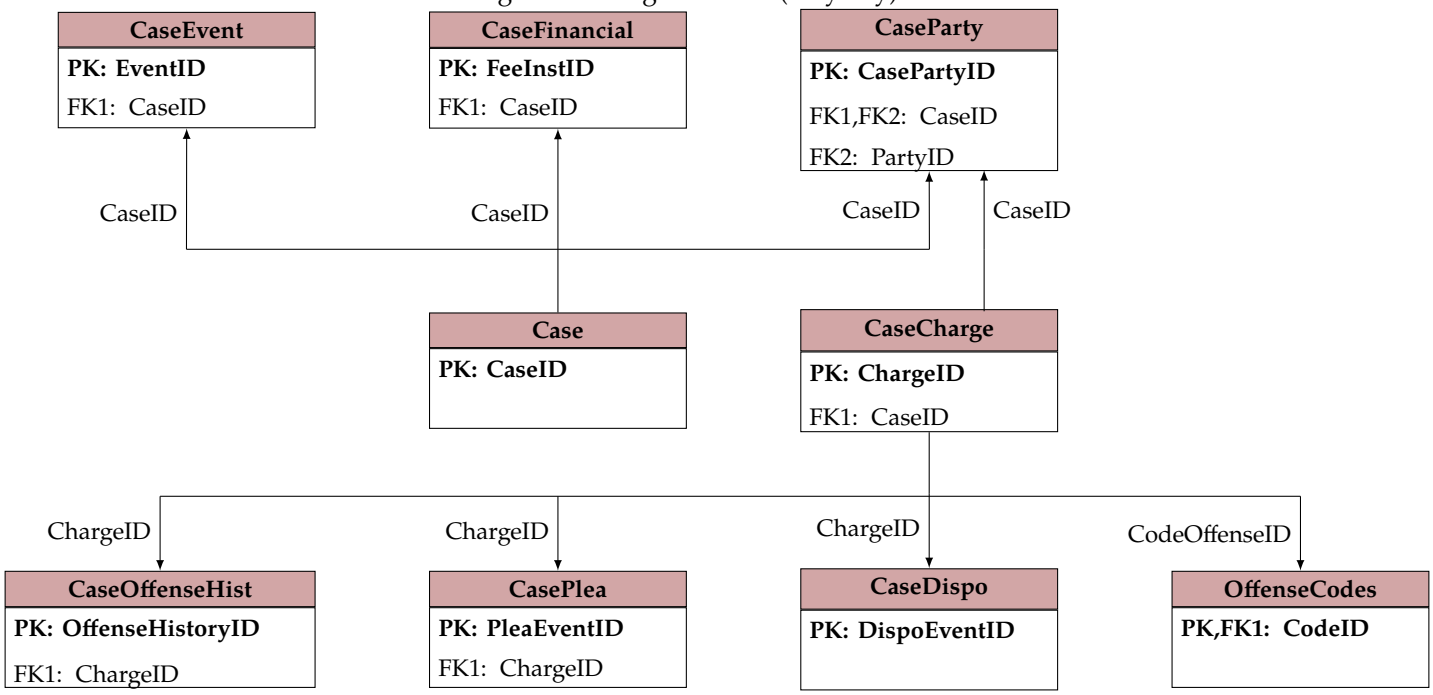
The court data has recently transferred to a new system called "Oregon eCourt" (a.k.a. "Odyssey"). This data is at the charge level. The data comes from Oregon's 36 circuit courts. The courts enter disposition information into LEDS (i.e., much of the information in Odyssey is also contained in LEDS–there just can be a lag for when information becomes available). The ERD for the Odyssey system SQL tables is reproduced in Figure 22.

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 22: 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 23).[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.

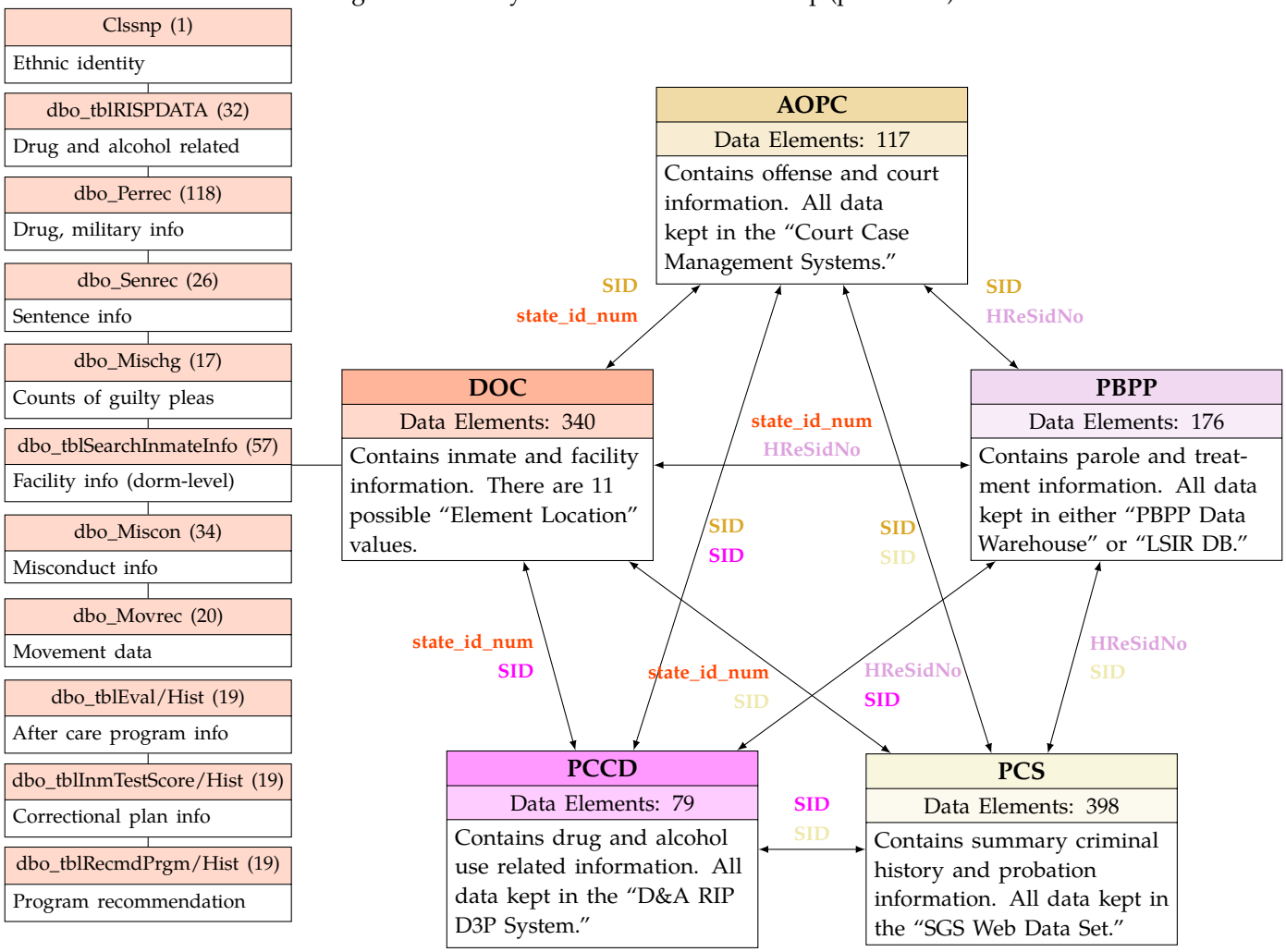
H.15 Texas

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 24.

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 23: Pennsylvania database relationship (presumed)



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

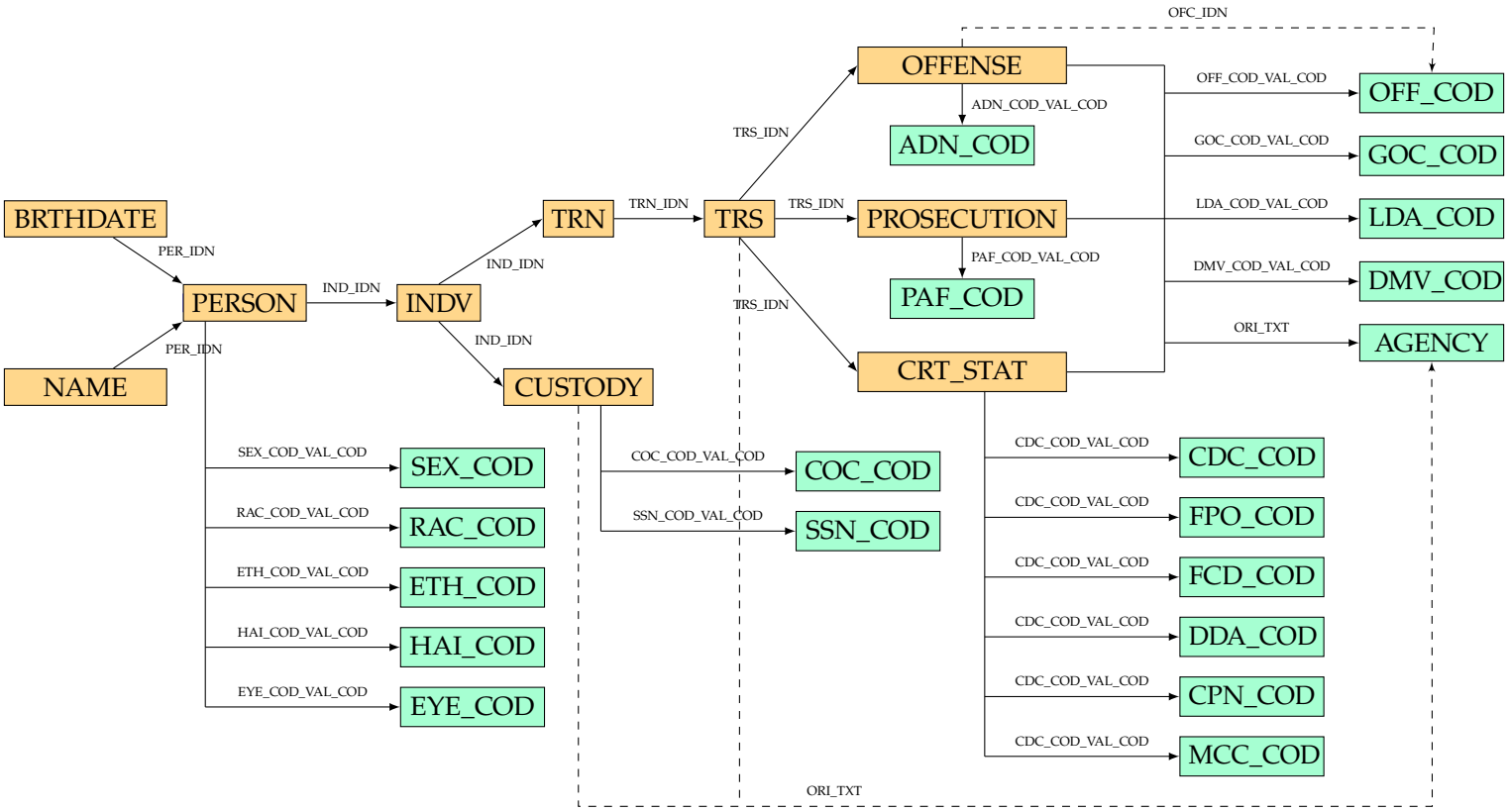
Criminal history data is kept by the Utah Department of Public Safety, Department of Technology Services division. The information we received was centered around the organization of the data (see Figure 25); we received no little information about what data is collected, who collects the data and submits 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™ 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 24: Texas CCH database model



Automated Fingerprint Identification System (AFIS), the Sex Offender Registry (SOR) system, and the Department of Corrections (DOC) system (see Figure 26). 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 27 is an estimate of how the information in these files could be connected.

Figure 25: Utah database model (presumed)

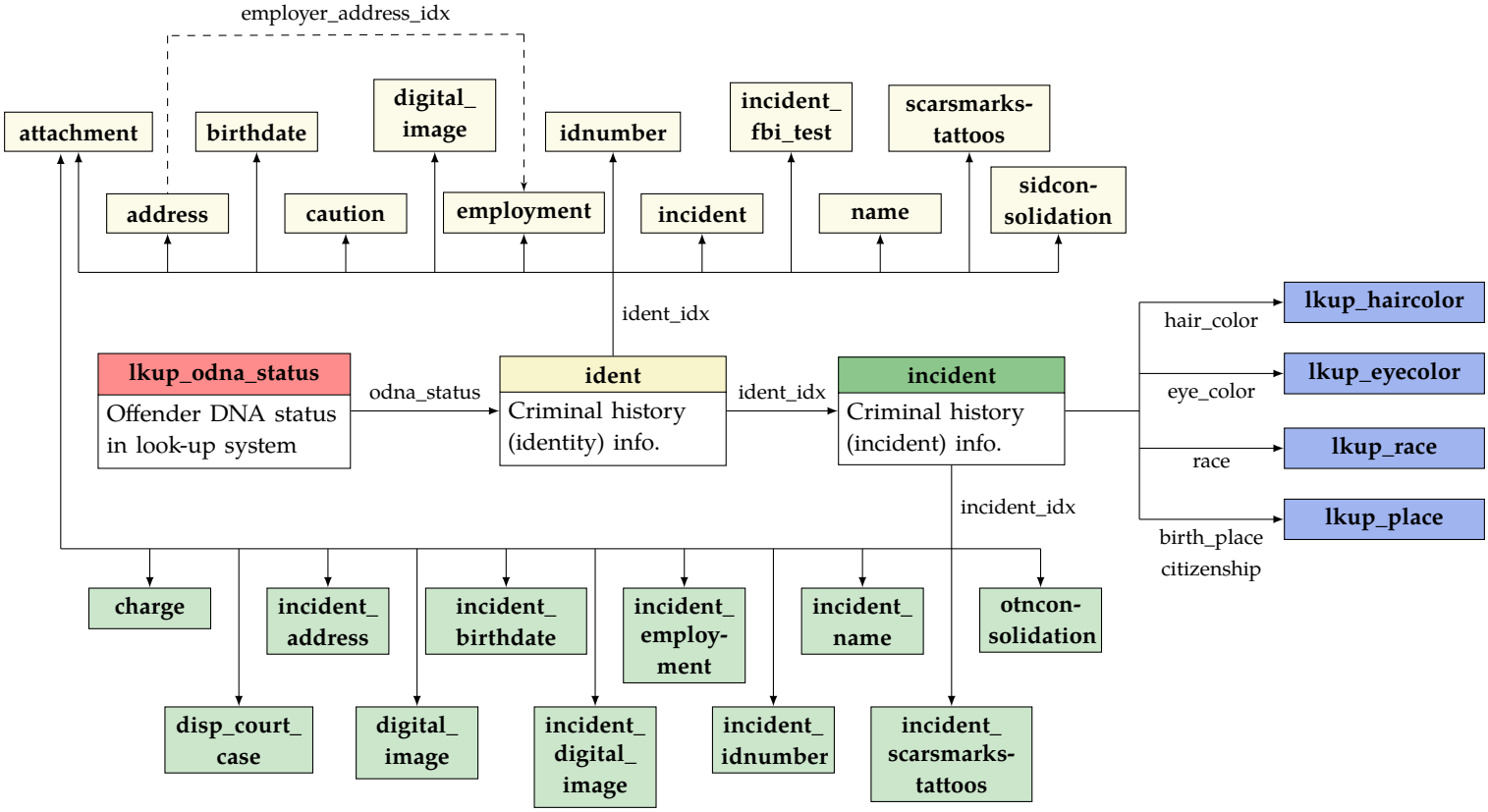


Figure 26: Vermont database model (presumed)

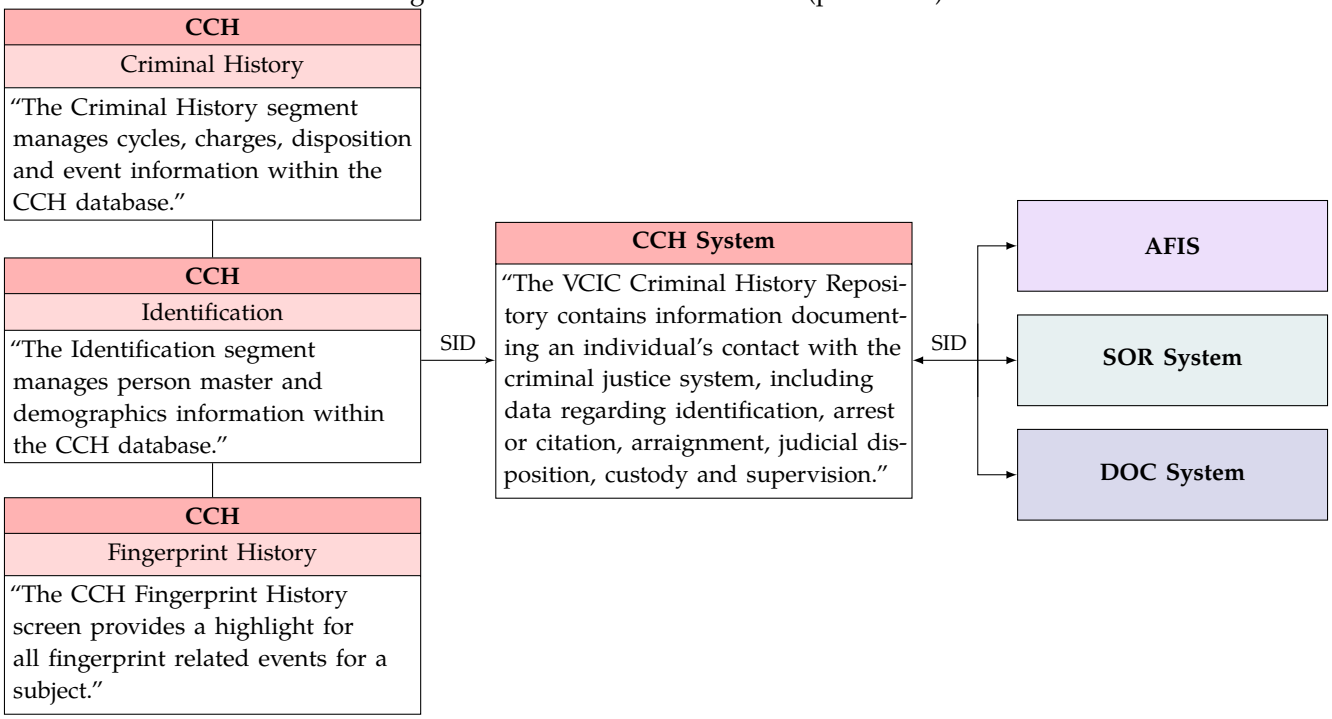
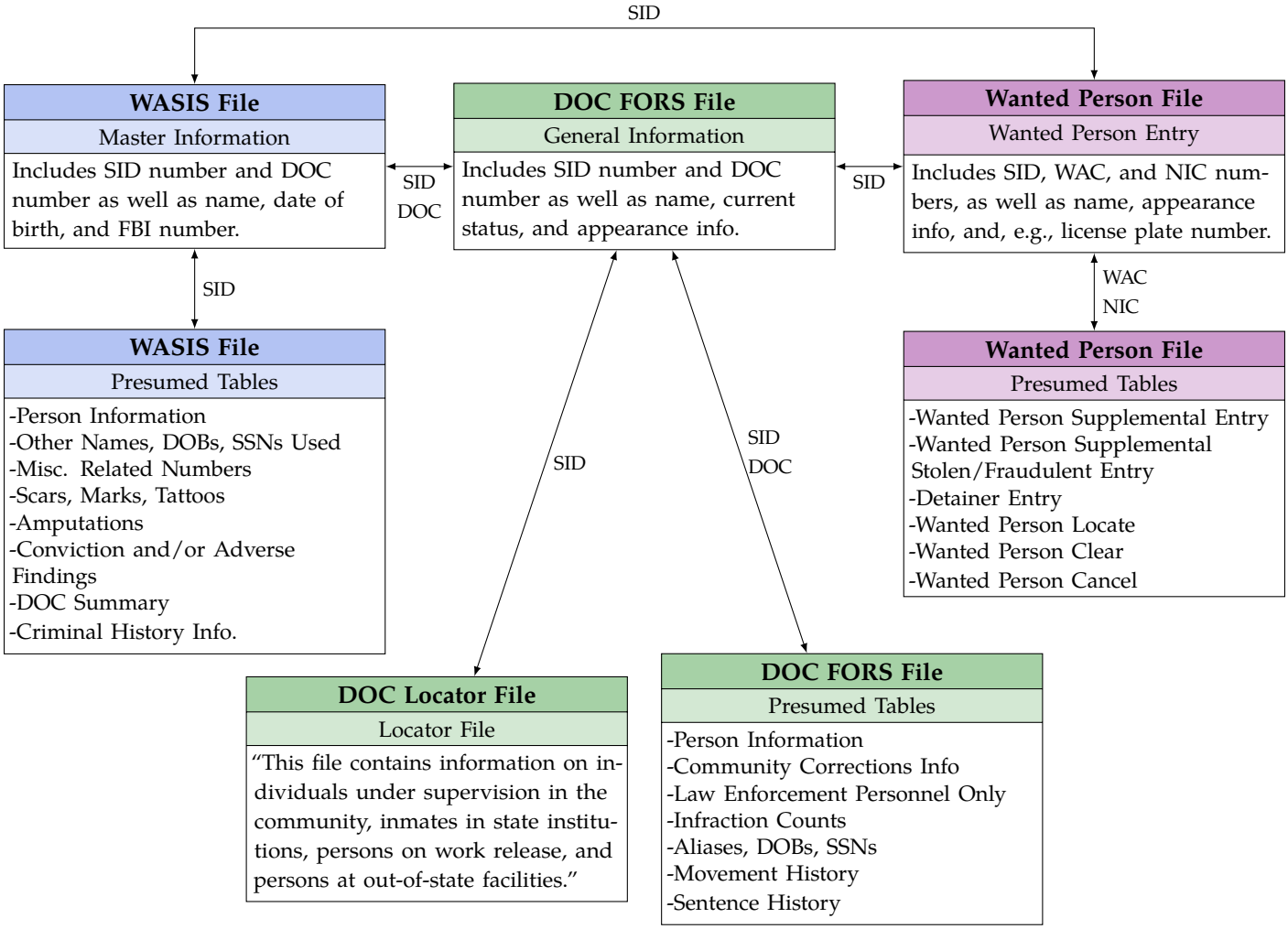


Figure 27: Washington data organization (presumed)



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

access to data, 9, 167
acknowledgments, 11
adj_chrg_off_cd, 54
adj_chrg_off_cd_src, 54
adj_cnty_ori_fips, 83
adj_disp_cd, 59
adj_disp_cd_src, 60
adj_disp_dt_dd, 58
adj_disp_dt_mm, 57
adj_disp_dt_yyyy, 55
adj_disp_off_cd, 60
adj_disp_off_cd_src, 61
adj_file_dt_dd, 53
adj_file_dt_mm, 52
adj_file_dt_yyyy, 50
adj_grd_cd, 47
adj_grd_cd_src, 48
adj_id, 16, 46
adj_off_dt_dd, 64
adj_off_dt_mm, 63
adj_off_dt_yyyy, 61
adj_off_lgl_cd, 49
adj_off_lgl_cd_src, 49
adj_rec_src_cc, 85
adj_rec_src_crt, 84
adj_rec_src_doc, 84
adj_rec_src_le, 83
adj_rec_src_rep, 85
adj_sent_dt_dd, 69
adj_sent_dt_mm, 68
adj_sent_dt_yyyy, 65
adj_sent_dth, 71
adj_sent_fine, 77
adj_sent_inc, 72
adj_sent_inc_max, 79
adj_sent_inc_min, 78
adj_sent_pro, 73
adj_sent_rest, 74
adj_sent_serv, 70
adj_sent_src, 80
adj_sent_sus, 75
adj_sent_trt, 76
adj_st_ori_fips, 81
adjudication, 46
adjudication process, 144
All states
    data notes, 168
Arizona
    data notes, 170
    data schema, 187
Arkansas
    data notes, 171
arr_arr_dt_dd, 35
arr_arr_dt_mm, 34
arr_arr_dt_yyyy, 32
arr_book_dt_dd, 39
arr_book_dt_mm, 38
arr_book_dt_yyyy, 36
arr_cnty_ori_fips, 43
arr_id, 16, 31
arr_off_cd, 40
arr_off_cd_src, 40
arr_rec_src_cc, 45
arr_rec_src_crt, 44
arr_rec_src_doc, 44
arr_rec_src_le, 43
arr_rec_src_rep, 45
arr_st_ori_fips, 41
arrest and booking, 31
arrest and booking process, 144

California
    data notes, 172
    data schema, 187
Census Bureau, 166
charge grade, 160
citing the project, 11
CJARS roster, 25, 166
cjars_id, 15, 25, 31, 46, 86, 103, 121, 166
code schemes, 147
Colorado
    data notes, 173
confidentiality, 10
Connecticut
    data schema, 187
county FIPS, 148
court disposition, 160
coverage, 8, 18
    geographic, 18
    procedural, 18
    temporal, 18
Criminal Justice Information Services, 10

data acquisition, 12
data integration, 12
data notes, 168
data schema, 15, 187
data security, 10
data sources, 12
dob_dd, 28
dob_mm, 27
dob_yyyy, 26

entity resolution, 16
episode resolution, 17
ethnicity, 149

Federal Statistical Research Data Center, 9, 167
Florida
    data notes, 173
    data schema, 189
funders, 10

harmonization process, 142

Illinois
    data notes, 174
    data schema, 189
inc_cnty_ori_fips, 116
inc_entry_cd, 109
inc_entry_cd_src, 110
inc_entry_dt_dd, 108
inc_entry_dt_mm, 107
inc_entry_dt_yyyy, 105
inc_exit_cd, 114
inc_exit_cd_src, 115
inc_exit_dt_dd, 113
inc_exit_dt_mm, 112
inc_exit_dt_yyyy, 110
inc_fcl_cd, 104
inc_fcl_cd_src, 105
inc_id, 16, 103
inc_rec_src_cc, 120
inc_rec_src_crt, 118
inc_rec_src_doc, 119
inc_rec_src_le, 118
inc_rec_src_rep, 119
inc_st_juris_fips, 117
inc_st_ori_fips, 115
incarceration, 103
incarceration entry, 161
incarceration exit, 162
incarceration facility type, 162
incarceration process, 144
IRB approval, 10

Kansas
    data notes, 174
    data schema, 190
Kentucky
    data schema, 191

legal code, 160

Maryland
    data notes, 175
Michigan
    data notes, 175
Minnesota
    data notes, 175
    data schema, 193
Mississippi
    data notes, 176

Nebraska
    data notes, 176
Nevada
    data schema, 193
New Jersey
    data notes, 177
New Mexico
    data schema, 194
New York
    data schema, 195
North Carolina
    data notes, 178
    data schema, 196
Numident, 166

offense classification, 146, 150
Ohio
    data notes, 179
Oregon
    data notes, 179
    data schema, 197

par_bgn_dt_dd, 124
par_bgn_dt_mm, 123
par_bgn_dt_yyyy, 122
par_cnty_ori_fips, 131
par_end_cd, 129
par_end_cd_src, 130
par_end_dt_dd, 128
par_end_dt_mm, 127
par_end_dt_yyyy, 125
par_id, 16, 121
par_rec_src_cc, 135
par_rec_src_crt, 133
par_rec_src_doc, 134
par_rec_src_le, 133
par_rec_src_rep, 134
par_st_juris_fips, 132
par_st_ori_fips, 130
parole, 121
parole exit, 163
parole process, 144
Pennsylvania
    data notes, 179
    data schema, 197
Person Identification Validation System, 13, 166
Personal Identification Key, 10, 13, 166
personally identifying information, 8
principal investigators, 10

privacy, 10
pro_bgn_dt_dd, 91
pro_bgn_dt_mm, 90
pro_bgn_dt_yyyy, 89
pro_cnty_ori_fips, 98
pro_cond_cd, 87
pro_cond_cd_src, 88
pro_end_cd, 96
pro_end_cd_src, 97
pro_end_dt_dd, 95
pro_end_dt_mm, 94
pro_end_dt_yyyy, 92
pro_id, 16, 86
pro_rec_src_cc, 102
pro_rec_src_crt, 100
pro_rec_src_doc, 101
pro_rec_src_le, 100
pro_rec_src_rep, 101
pro_st_juris_fips, 99
pro_st_ori_fips, 97
probabilistic record linkage, 16, 166
probation, 86
probation conditions, 161
probation exit, 161
probation process, 144
project description and objectives, 8

race, 149
race, 30
roster, 25

sex, 148
sex, 29
sponsors, 10
staff, 11
state data systems, 187
state FIPS, 147

tables, 22
Texas
    data notes, 180
    data schema, 198

Utah
    data schema, 199

Vermont
    data schema, 199

Washington
    data notes, 186
    data schema, 200
Wisconsin
    data notes, 186