MACR Data Inventory

Abstract

The Monthly Arrest and Citations Record (MACR) database describes arrests in California from 1980 to the present. Each row includes the race, age, and gender of the individual arrested, the most serious offense for which he or she was arrested, the year of the arrest, and the county in which the arrest took place. This version of the data only contains arrests of adults (18+).

Introduction

MACR Main Database

The Monthly Arrest and Citation Register compiles monthly arrest reports from all law enforcement agencies in California. The reports are supposed to contain every arrest (and citation until 2005) that an agency makes of juveniles or adults.

- Each row represents a single arrest, and includes the arrest date, the offense for which the individual was arrested, the individual's age, gender, race or ethnicity, and whether the case was referred to the local prosecutor. Each row also contains a full name and birthdate, which are excluded in these data.
- An arrest is defined as detaining an individual with the intention of seeking charges for a specific offense
- The MACR contains only the most serious offense based on the severity of possible punishment for retention.
- MACR data began to be stored in a digital format in 1980. The dataset spans 1980 to the present.

Law enforcement agencies are instructed to report all persons arrested within their jurisdiction. They are instructed to include arrests that result in a release without charges (including arrests of juveniles that only result in a warning). Departments that fail to send their data within thirty days of the due date are contacted, with increasing escalation when they reach 60 or 90 days past the due date.

Historical changes:

- The race_or_ethnicity codes for Asian/Pacific Islander expanded in 1991.
- The CA DOJ stopped collecting data about arrests or citations made for infractions in 2005.
- In 2011, the lower limit of felony theft was raised from $400 to $950, contributing to the decrease in felony theft arrests and increase in misdemeanor theft arrests.
- In 2011, some misdemeanor marijuana offenses were re-classified to infractions leading to a decrease in misdemeanor marijuana arrests.
- In 2014, California voters passed Proposition 47, which reduced numerous state statutes from felonies to misdemeanors - leading to a reduction in some types of felony arrests.

Department-specific changes:

- Bakersfield Police Department (PD) and Oakland PD did not report arrest data in 1995.
- San Francisco did not update its race_or_ethnicity codes until 2012, when it adoped the FBI's categories: white, black, American Indian, other Asian, and other. San Francisco data since 2012 does not distinguish between Hispanic and non-Hispanic whites.

Sample Rows

record_type_id	bcs_jurisdict...	ncic_jurisdic...	arrest_year	arrest_month	arrest_day	...
94	0	1900	1980	1	5	...

record_type_id	bcs_jurisdict...	ncic_jurisdic...	arrest_year	arrest_month	arrest_day	...
94	0	1900	1980	1	1	...
94	0	1900	1980	1	1	...
94	0	1900	1980	1	1	...
94	0	1900	1980	1	1	...
...	...	...	...	...	...	...

Variable Summary

name	type	value	description
age	integer	0-112	individual's age
arrest_day	integer	1-31	date of arrest
arrest_month	integer	1-12	date of arrest
arrest_year	integer	1980-2015	date of arrest
bcs_jurisdiction	factor	0/1/5/7/12/28/36/40...	deprecated
bcs_offense_code	factor	1/2/3/4/6/7/9/16/17...	groups penal codes
bcs_summary_offense_code	factor	1/2/3/4/5/6/7/8/9/1...	groups BCS codes
birth_day	integer	pii	date of birth
birth_month	integer	pii	date of birth
birth_year	integer	pii	date of birth
disposition	factor	released/turned	law enforcement disposition
fbi_offense_code	factor	01A/01B/02/03/04/05...	Uniform Crime Reporting code
gender	factor	male/female	individual's gender
name	character	pii	individual's name
id	integer	pii	local id number
ncic_jurisdiction	factor	0100/0101/0102/0103...	law enforcement agency
offense_level	factor	status .../misdeme....	severity of offense
race_or_ethnicity	factor	White/Hispanic/Blac...	individual's race or ethnicity
record_type_id	factor	14/24/94	administrative handling
status_type	factor	cited/booked/other	booked, cited, other
summary_offense_level	factor	felony/juvenile/mis...	DOJ offense level

BCS Code Table

BCS codes combine like statutes for statistical analysis. This table maps statutes to BCS codes to BCS summary codes (groups of BCS codes).

Sample Rows

offense_code	summary_offen...	summary_offen...	offense_categ...	new_2013
1	68	Truancy	Status	0
2	69	Runaway	Status	0
3	70	Curfew	Status	0
4	72	Other Stat Of...	Status	0
6	72	Other Stat Of...	Status	0
...	...	...	...	...

Variable Summary

name	type	value	description
offense_code	integer	1-998	
summary_offense_code	integer	1-76	groups offense codes
summary_offense_type	character	Truancy/Runaway/Curfew/...	
offense_category	character	Status/Misdemean.../Oth...	
new_2013	binary	0-1	law changed in 2013

NCIC Jurisdiction Table

The jurisdiction is the law enforcement agency that made the arrest. This table maps jurisdiction codes to their names and counties. It also describes when agencies began and stopped reporting, when agencies merged, and if agencies subcontracted to one another.

Sample Rows

CntyCode	County	Code	Agency	Start	End	...
1	Alameda County	0100	Alameda Co. S...			...
1	Alameda County	0101	Alameda			...
1	Alameda County	0102	Albany			...
1	Alameda County	0103	Berkeley			...
1	Alameda County	0104	Emeryville			...
...	...	...	...	...	...	...

Variable Summary

name	type	value	description
CntyCode	integer	1-58	
County	character	Alameda C.../Alpine Co....	
Code	character	0100/0101/0102/0103/010...	
Agency	character	Alameda C.../Alameda/Al...	
Start	character	/1/1/1997/1/1/2003/7/1/...	if absent, active throughout
End	character	/12/31/2003/6/30/2007/1...	
Contract	character	/C	
CJSC.Notes	character	/Name chan.../MACR only...	
Old.Juris.Code	character	20-000/20-002/20-004/20...	remove '-' to match MACR

Tables

Main Table Variables

age

Description

age of the individual arrested

Prompt

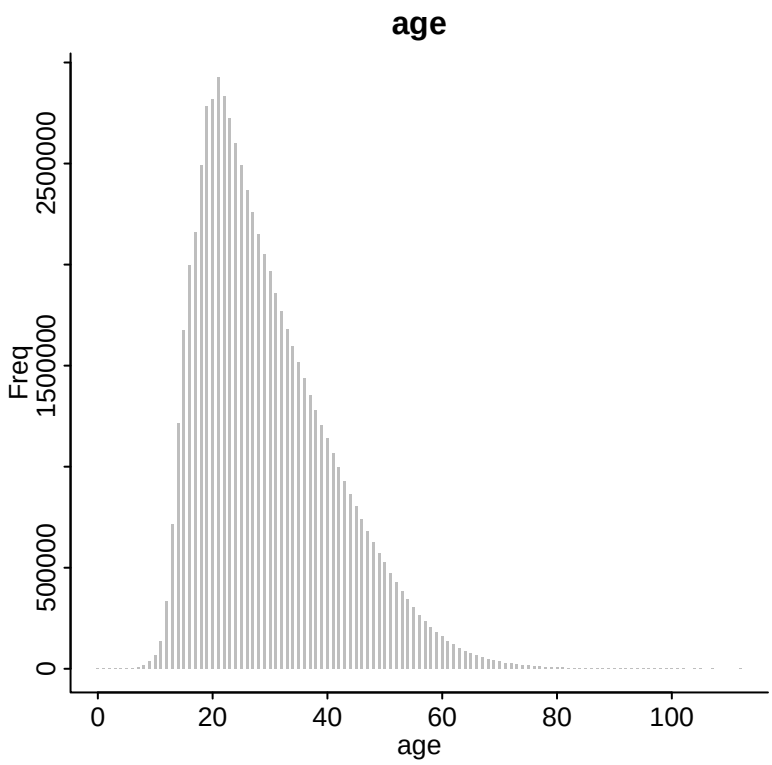
Arrest data for adults (age 18 years and older on the date of arrest) and juveniles (age 17 years or younger on the date of arrest) must be separated. Check the proper box to indicate if the data on the page submitted is adult or juvenile. If an agency has no adult or juvenile arrests for a month, "no adults to report" or "no juveniles to report" box must be checked.

Notes

if not already done so, users should consider dropping arrestees under age 5 and over age 89 as they may be data entry errors

Summary

Name	Value
Min.	0.00000
1st Qu.	21.00000
Median	27.00000
Mean	29.51242
3rd Qu.	36.00000
Max.	112.00000


age
Freq
0
500000
1500000
2500000
0
20
40
60
80
100
age

arrest_day

Description

day of the month the individual was arrested

Prompt

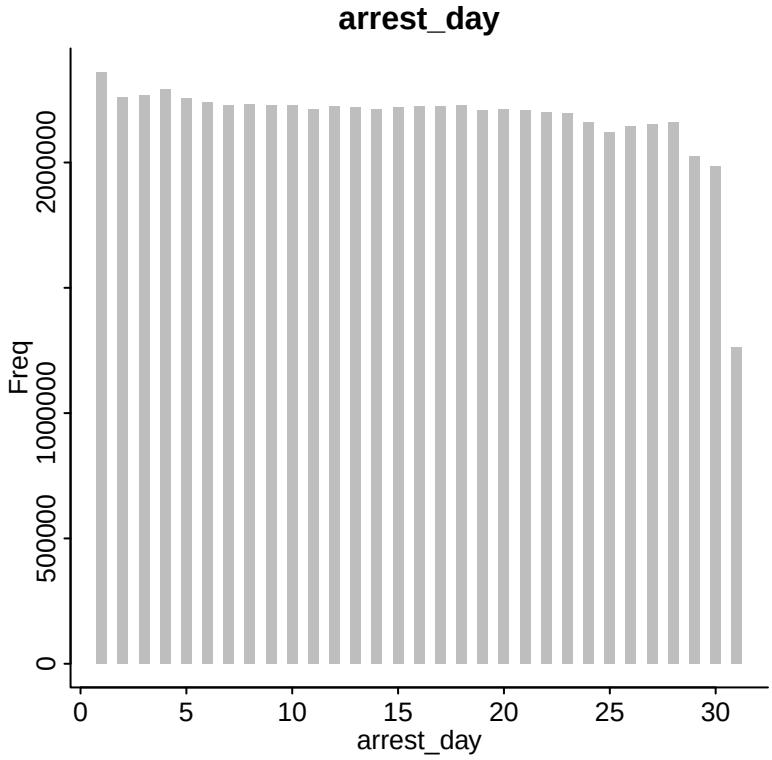
Use eight-digits: two each for the month, and day, and four for the year. For example, an arrest made on February 9, 2006 should be entered as: 02/09/2006.

Notes

The date February 30 ("02/30") was originally used to indicate a missing arrest date, these were recoded to NA

Summary

Name	Value
Min.	1.00000
1st Qu.	8.00000
Median	16.00000
Mean	15.56912
3rd Qu.	23.00000
Max.	31.00000
NA's	51.00000



arrest_month

Description

month the individual was arrested

Prompt

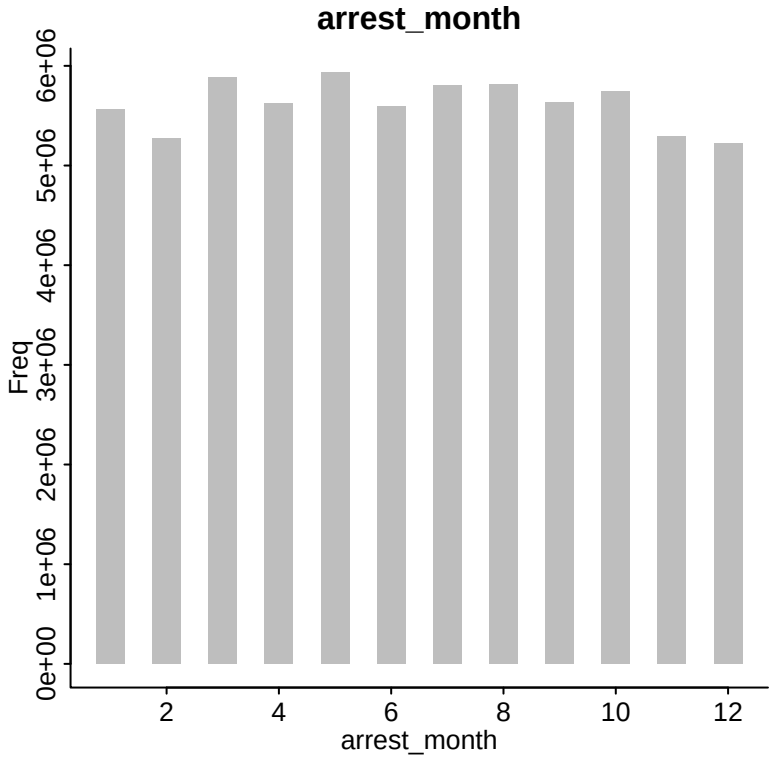
Use eight-digits: two each for the month, and day, and four for the year. For example, an arrest made on February 9, 2006 should be entered as: 02/09/2006.

Notes

The date February 30 ("02/30") was originally used to indicate a missing arrest date, these were recoded to NA

Summary

Name	Value
Min.	1.000000
1st Qu.	4.000000
Median	6.000000
Mean	6.465797
3rd Qu.	9.000000
Max.	12.000000



arrest_year

Description

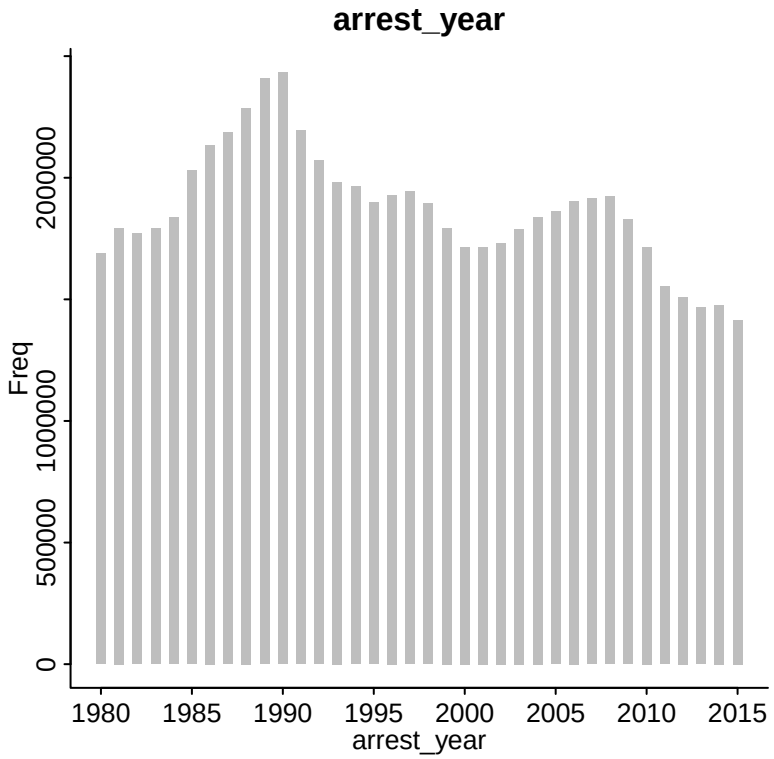
year the individual was arrested

Prompt

Use eight-digits: two each for the month, and day, and four for the year. For example, an arrest made on February 9, 2006 should be entered as: 02/09/2006.

Summary

Name	Value
Min.	1980.000
1st Qu.	1988.000
Median	1996.000
Mean	1996.735
3rd Qu.	2005.000
Max.	2015.000



bcs_jurisdiction

Description

deprecated

Prompt

The code/ID number assigned to the reporting agency (old).

Notes

[should we remove this field?]

Summary

Name	Freq
NA	8168645
411	4369571
14664	2871242
21670	1751098
22682	1388024
40281	1372841
0	1238111
408	1107957
20512	959163
1	945997
14000	672932
54637	662980
54000	662762
45768	597126
11705	574342
12000	573645
46459	526869
13000	496114
41052	444611
12625	394859
41000	394196
13652	391984
16533	376597
11019	351143
...	...

bcs_offense_code

Description

groups like penal codes

Labels

see the BCS offense codes table

Prompt

The code assigned to an offense. This code combines like statutes for statistical analysis.

Notes

This is the code for the most serious offense for which the individual was arrested. Officers and departments vary in their application of the penal code. By the time the code is entered onto the MACR, it might have been updated by detectives or by records clerks. [CJSC, is this true?: Depending on the knowledge and experience of the records clerks, coding consistency and accuracy might vary]. The BCS code system is maintanied by the Criminal Justice Statistics Center, with new penal codes being added as they gain usage. Misdemeanor traffic violations (BCS codes 086 and 087) are optional to report on the MACR

Summary

Name	Code	Freq
Drive Under the Influence	856	8449788
Misc Traffic	86	5814111
Drunk	46	5081866
Petty Theft	516	3432602
Failure to Appear/Non Traffic	98	2851950
Outside Warrant Misd	69	2780578
City/County Ordinance	97	2521050
Burglary	400	1995076
Other Drug Law Violations	836	1688291
Assault and Battery	397	1669267
Dangerous Drugs	825	1446113
Assault	372	1363754
Traffic	88	1359592
Narcotics	800	1240059
Marijuana	819	1161242
Other	96	1119441
Outside Warrant	65	903477
Assault	320	887030
Other Drug Law Violations	837	864658
Other Felony	993	851390
Liquor Laws	77	807106
Theft	530	797609
Motor Vehicle Theft	570	747147
Trespassing	68	714461
...	...	...

bcs_summary_offense_code

Description

[added by CA DOJ] a code that combines BCS codes for more general analyses

Labels

see the BCS offense codes table

Prompt

These codes are assigned to BCS codes. They combine like BCS codes for more general statistical analysis.

Notes

Because Summary Offense Codes combine BCS codes, the same caveats about potential coding errors at the individual or agency-level apply. Arrests for the following offenses are not included in publications from the California DOJ's Criminal Justice Statistics Center:

- Summary code 26 = Felony Federal offense
- Summary code 27 = Felony outside warrant
- Summary code 28 = Felony probation/parole violation
- Summary code 65 = Misdemeanor civil drunk
- Summary code 66 = Misdemeanor outside warrant
- Summary code 67 = Misdemeanor probation/parole violation
- Summary code 74 = Misdemeanor miscellaneous traffic

Summary

Name	Code	Freq
Drive Under the Influence	51	8460894
Misc Traffic	74	5808841
Drunk	43	5081866
Petty Theft	31	3436020
Assault	6	3294714
Failure to Appear/Non Traffic	59	2851950
Assault and Battery	30	2815249
Outside Warrant Misd	66	2780578
Other Drug Law Violations	36	2676626
City/County Ordinance	58	2521050
Burglary	8	2263322
Narcotics	12	2027052
Dangerous Drugs	14	1895519
Theft	9	1883046
Other	60	1831728
Traffic	53	1727569
Other Felony	25	1459465
Marijuana	34	1315304
Liquor Laws	44	1022221
Motor Vehicle Theft	10	996304
Outside Warrant	27	903477
Robbery	5	812140

Name	Code	Freq
Trespassing	49	714476
Weapons	19	706417
...	...	...

birth_day

Description

day of the month the individual was born

Prompt

Use eight-digits: two each for the month and day, and four for the year. For example, a birthdate of January 9, 1949 should be entered as: 01/09/1949. If the month and day are not known, use February 30 for the month and day and show the year of birth. For example, if the year of birth is 1945, enter the following: 02/30/1945. Do not write in the age. If the age is known, but not the date of birth, subtract the age from the present year and enter the resulting year of birth.

Notes

Because it was originally used to indicate missing values, “02/30” was recoded to NA

birth_month

Description

month the individual was born

Prompt

Use eight-digits: two each for the month and day, and four for the year. For example, a birthdate of January 9, 1949 should be entered as: 01/09/1949. If the month and day are not known, use February 30 for the month and day and show the year of birth. For example, if the year of birth is 1945, enter the following: 02/30/1945. Do not write in the age. If the age is known, but not the date of birth, subtract the age from the present year and enter the resulting year of birth.

Notes

Because it was originally used to indicate missing values, “02/30” was recoded to NA

birth_year

Description

year the individual was born

Prompt

Use eight-digits: two each for the month and day, and four for the year. For example, a birthdate of January 9, 1949 should be entered as: 01/09/1949. If the age is known, but not the date of birth, subtract the age from the present year and enter the resulting year of birth.

disposition

Description

law enforcement disposition (e.g., released, referred to district attorney)

Labels

Misdemeanor (only for adults):

- Misdemeanor complaints that are sought by the arresting agency. (not used for juveniles)

Felony (only for adults):

- Felony complaints that are sought by the arresting agency. (not used for juveniles)

Released (only for adults):

- Each arrest released under 849(B) PC, or other sections, when no further action is planned by the arresting agency.
- Civil drunk arrest (647 (G) PC) or those individuals placed on other diversion programs by the local law enforcement agency, including those deemed not to be arrested.
- A new local offense in conjunction with an outside warrant. The level, status, charge, and disposition should be related to the local offense so that statistics on the local charges are captured. If the local offense is released so the out warrant may be acted upon, then the disposition is released.
- A new local offense in conjunction with a federal offense. The level, status, charge, and disposition should relate to the local offense so that statistics on the local charges are captured. If the local offense is released so the federal out warrant may be acted upon, then the disposition is released.
- not used for juveniles

Turned Over:

- Arrests made on another law enforcement agency's warrant (out warrant), with no local charges, and the subject is being held for the other agency
- Arrests made for a federal offense with no local charges.
- Fugitives from justice with no local charges.
- When a fine is paid to the local agency on a failure to appear traffic warrant issued by an outside jurisdiction and the money is forwarded to the issuing agency.

Juvenile Court:

- A juvenile that is referred to juvenile court or turned over to the probation department, welfare agency, other police agency, criminal or adult court or juvenile hall.

Department (only for juveniles):

- A situation that has been settled by the arresting agency, no action is to be taken by the juvenile probation department or the court, and the juvenile is released to his/her parents, guardian, or the street with a warning.
- A juvenile is placed on a local diversion program including, for statistical purposes, any juvenile deemed not arrested or cited.

Prompt

This column is intended for the disposition of the agency reporting the arrest or citation. DO NOT report the district attorney or court disposition in this column. It is intended to reflect the law enforcement agency disposition of the charge, not the person. ENTER ONE DISPOSITION PER LINE ITEM.

Summary

Name	Freq
misdemeanor complaint sought	37552079
felony complaint sought	12713879
referred to juvenile probation department	6198209
turned over to other agency	4626341
released	4293862
handled within department	2009501

fbi_offense_code

Description

[added by CA DOJ] code under the FBI Uniform Crime Reporting system

Prompt

Code under the FBI's Uniform Crime Reporting system, which is used for national comparisons. These do not include all offenses and do not distinguish between felony and misdemeanor levels.

Notes

FBI codes only apply to the subset of offenses tracked in the Uniform Crime Reporting system

Summary

Name	Freq
NA	12028396
26	10649286
21	8679186
23	5081866
06	4344494
04	3269815
18E	3024070
08	2823617
18H	2541534
05	2295842
18F	1392488
15	1066020
07	1033051
22	1022221
14	953858
13	823716
03	812140
18A	793787
17	571009
16	565195
24	531802
18B	490620
18D	452562
28	438102
...	...

gender

Description

arrested individual's gender

Prompt

Enter either (1) Male or (2) Female.

Summary

Name	Freq
male	54724030
female	12669841

name

Description

arrested individual's name

Prompt

Print legibly or type the last name, middle name or initial (if known), and first name of the arrestee. If name is unknown, use "John Doe" or "Jane Doe."

id

Description

an id number assigned by the reporting agency

Prompt

Enter the most reliable number for locating the arrested person in your agency's files in case questions arise. This can be the booking, arrest, or crime report number.

ncic_jurisdiction

Description

Unique law enforcement agency code. First two digits are the county.

Labels

see the NCIC jurisdiction table

Prompt

Enter your agency ORI/NCIC number. Agencies should abbreviate the nine-character NCIC code on the MACR report by using the fourth through seventh character of the NCIC code. For example, if your NCIC number is "CA0570100," report "5701" only.

Notes

Some agencies disappear and others are created over time. From 2005-2015, about 95% of arrests were made by about 250 of the 911 agencies in the dataset.

Summary

Name	Code	Freq
Los Angeles	1942	4369571
San Diego	3711	3018802
San Francisco	3801	1747749
Los Angeles Co. Sheriff's Department	1900	1695301
San Jose	4313	1424026
Fresno	1005	1372400
Long Beach	1941	1145832
Oakland	0109	1087181
CA Highway Patrol - Los Angeles	1999	946001
Sacramento	3404	880077
LAPD - Non-San Fernando Valley	193W	869266
Sacramento Co. Sheriff's Department	3400	816401
San Diego Co. Sheriff's Department	3700	714790
Bakersfield	1502	685394
San Bernardino Co. Sheriff's Department	3600	653726
Stockton	3905	608576
Santa Ana	3019	593718
Kern Co. Sheriff's Department	1500	589065
Riverside Co. Sheriff's Department	3300	573668
Modesto	5002	550464
Anaheim	3001	459184
Riverside	3313	414829
Oxnard	5604	406605
San Bernardino	3610	391938
...	...	...

offense_level

Description

severity of offense (misdemeanor, felony, status offense)

Labels

Status offense; Misdemeanor; Felony

Prompt

Select the level (delinquent, misdemeanor or felony) that best describes the most serious offense. Enter only one level per arrest or citation. 1) Delinquent (juvenile-only; also known as a status offense), 2) Misdemeanor, 3) Felony

Notes

Status offenses only apply to juveniles.

Summary

Name	Freq
misdemeanor	47383398
felony	18999927
status offense	1010546

race_or_ethnicity

Description

arrested individual's race or ethnicity

Prompt

Record only one alpha designation that applies. Agencies submitting automated reports must verify that the appropriate codes are being entered. Do not report the race as "Unknown." Record the appropriate alpha code for race. Do not use "other" for unknown race.

Notes

The codes for Asian/Pacific Islander became more detailed in 1991. San Francisco did not change its reporting practices until 2012, when it adopted the FBI's categories for race: white, black, American Indian, other Asian, and other. Since 2012, San Francisco has not distinguished between non-Hispanic whites and Hispanic whites.

Summary

Name	Freq
White	27991111
Hispanic	24180079
Black	11763777
Other	1743674
Other Asian	391566
American Indian	376349
Filipino	293964
Vietnamese	149585
Chinese	122474
Pacific Islander	122024
Asian Indian	54884
Japanese	41810
Laotian	41506
Korean	35536
Hawaiian	29953
Samoan	29173
Cambodian	18532
Guamanian	7874

record_type_id

Description

[added by CA DOJ] type of arrest record (administrative code)

Labels

Arrest Codes:

- 14 - Add a record
- 24 - Replace a specific record
- 94 - Record sent to FBI

Records of No Arrest Codes:

- 21 - Report of no arrest
- 91 - Report of no arrest sent to FBI

Deleted Record Code:

- 32 - Specific delete action

Prompt

Flag that describes the action of the record. Codes 14, 24, and 94 represent arrest records. Codes 21, 32 and 91 represent deleted records or records of no arrest.

Summary

Name	Freq
94	65818499
14	1572148
24	3224

status_type

Description

whether the individual was booked, cited, or other

Labels

Cited:

- Cited (or summoned) to appear in court as an alternative to being jailed or cited to court and later booked as directed by the court. A cite occurs in the field, when the suspect is not physically arrested by the officer.
- Informal booking -voluntarily go in and sign a notice to appear later in court.
- When a juvenile is cited in lieu of being delivered to juvenile authorities.

Booked:

- An adult is actually booked into jail for any period of time or booked into jail and later released on a citation.
- When a juvenile is booked into a juvenile holding facility of any type or any time an arrest report is filled out.

Other:

- An adult makes bail on a warrant and is neither cited nor booked.
- Detained for civil drunk occurrences per 647 (G) PC.
- When juveniles are neither cited nor booked (e.g., detained only, sent to a diversion program, referred to the probation department, etc.). Use "other" when there was no arrest report filled out.

Prompt

The status column describes the type of apprehension (at the time of initial contact with the arrestee). It determines how many individuals are cited versus those actually delivered to jail. The arresting agency is responsible for determining if it is a "cite," "book" or "other." The arresting agency should report "book" even when the suspect is sent to another law enforcement agency for processing. For example, many police departments send suspects that have been arrested to the county jail to be booked.

Notes

Booking rates vary to an implausible extent by agency and by year. Some agencies report 100% booking rates for every year. Other agencies report low booking rates for violent felonies. We recommend not using this variable unless you have reason to believe that particular agencies have reliable data.

Summary

Name	Freq
booked	47127834
cited	16548551
other	3717486

summary_offense_level

Description

[added by CA DOJ] BCS summary offense level (used to separate juveniles)

Labels

F - Felony (Adults) J - Juvenile M - Misdemeanor (Adults)

Prompt

The level distinguishs between juvenile and adult records.

Notes

"Juvenile" should match the count for those under 18.

Summary

Name	Freq
misdemeanor	42566580
felony	16458449
juvenile	8368842

Recommendations for Data Use

The MACR data are best used for analyses of general trends, they are less reliable for point estimates of numbers of arrests or numbers of people arrested. It is important to keep in mind that these data are heavily conditioned by individual, agency, and county variation in propensity to arrest, how offenses are categorized, and how well data are captured and reported to the CA DOJ. In using these data and preparing them for release, we have come across several anomalies and inconsistencies. They may produce results that are artefacts of data collection and reporting processes. To help researchers avoid potential pitfalls, we summarize our recommendations about data use below.

age

- Very young and very old ages are suspect. We suggest dropping those 5 or younger and 89 or older.

bcs_offense_code

- Arrest numbers for certain offenses may be more reliable than others. Different jurisdictions (even different law enforcement officers within the same jurisdiction) may report the same type of arrest using different codes. Generally, it is probably safe to assume that reporting of arrests for more serious offenses is more accurate than it is for arrests of less serious offenses. In some areas with high rates of violent crimes, police may be less willing to invest the time required to make an arrest for a non-violent misdemeanor. In other words, some arrests, particularly for less serious offenses, may be missing.

bcs_summary_offense_code

- Arrest numbers for certain offenses may be more reliable than others. Some arrests, particularly for less serious offenses, may be missing. Different jurisdictions may report the same type of arrest using different codes.

county

- County totals may be affected by reporting irregularities, such as large drops in reported arrests in one jurisdiction. See the Variation in Number of Arrests section for an explanation and use the VarArrestsFlag indicator variable to keep track of jurisdictions or counties that may have been affected by reporting problems in a particular year.

disposition

- Whether a complaint was filed tends to be reliable, but whether it was a felony or a misdemeanor complaint tends to be unreliable. We suggest users combine "felony complaint filed" and "misdemeanor complaint filed" into "complaint filed". Users focused on dispositions should first check the data by juridiction and year.

ncic_jurisdiction

- Some jurisdictions have implausible data for certain years, such as a drop from a few hundred or a few thousand arrests to zero. See the Variation in Number of Arrests section for an explanation and use the VarArrestsFlag indicator variable to keep track of jurisdictions or counties that may have been affected by reporting problems in a particular year. Note that jurisdictions that report zero arrests in one year will not have any records in the data - they can be found by looking at trend data or at the List of Missing Jurisdiction-Years in the Variation in Number of Arrests section.

race_or_ethnicity

- More specific codes for Asian/Pacific Islander were added in 1991. Researchers may want to map these to a more general category.
- Post 2012, San Francisco does not count arrests of Hispanics separately. Researchers may want to treat San Francisco separately in addressing questions about race or ethnicity.

status_type

- Booking data appears to be unreliable overall. We recommend not using it.

Data Cleaning

Deleted Records

Records with a type id of 32 represent deleted rows, and as they contain no information about the kind of arrest made are deemed unusuable and omitted. A typical example would be:

reco...	bcs_...	ncic...	arre...	arre...	arre...	summ...	offe...	bcs_...	...
32	20000	0100	1996	6	14	NA	NA	NA	...
32	20000	0100	1996	6	29	NA	NA	NA	...
32	20000	0100	1996	6	30	NA	NA	NA	...
32	20000	0100	1996	6	1	NA	NA	NA	...

Arrest Date

A total of 880 records cannot be parsed into valid dates, i.e. the combination of `arrest_year`, `arrest_month`, and `arrest_day` results in a non-sensical date. Records with `arrest_day` of 0 are changed to `NA`, however in other cases it may the arrest day or month that causes the failure to parse. Some structure in the coding may allow the correct date to be recovered. For example:

- Records at the end of a month with the `arrest_month` field prematurely incremented:

ncic_jurisdiction	arrest_year	arrest_month	arrest_day	arrest_date
4900	1981	10	31	1981-10-31
4900	1981	10	31	1981-10-31
4900	1981	10	31	1981-10-31
4900	1981	11	31	NA
4900	1981	11	1	1981-11-01
4900	1981	11	1	1981-11-01
4900	1981	11	1	1981-11-01

- Records where the digits in `arrest_day` appear to be transposed:

ncic_jurisdiction	arrest_year	arrest_month	arrest_day	arrest_date
1942	1980	2	13	1980-02-13
1942	1980	2	13	1980-02-13
1942	1980	2	13	1980-02-13
1942	1980	2	31	NA
1942	1980	2	13	1980-02-13
1942	1980	2	13	1980-02-13
1942	1980	2	13	1980-02-13

- Records where days were added at the end of a month:

ncic_jurisdiction	arrest_year	arrest_month	arrest_day	arrest_date
1942	1980	2	29	1980-02-29
1942	1980	2	29	1980-02-29
1942	1980	2	29	1980-02-29
1942	1980	2	30	NA
1942	1980	2	30	NA

ncic_jurisdiction	arrest_year	arrest_month	arrest_day	arrest_date
1942	1980	2	30	NA
1942	1980	2	31	NA

As for now determining a 'correct arrest date' requires estimation, we leave those as is and augment the data with a field `arrest_date` that contains `NA` for all rows where a date cannot be parsed.

Birth Date

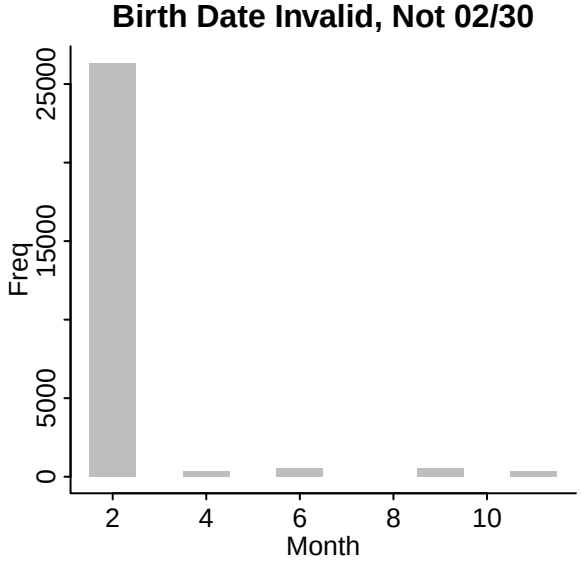
MACR includes fields for both the birth date and age, which do not always align. Errors in the birth date include:

- birth year recorded as `19xx` instead of `18xx`
- birth year recorded as `9xx` instead of `19xx`
- birth year in wrong millenia
- birth day 0
- birth month 0
- birth date invalid (e.g. `02/31/1991`)

Errors in the birth year are detected by looking for when the age and distance from arrest year to birth year exceed 1 in absolute value, and are 2372 in number. Cases where there difference is 100 or 1000 years can be resolved automatically, which leaves only 2 cases to be resolved by hand.

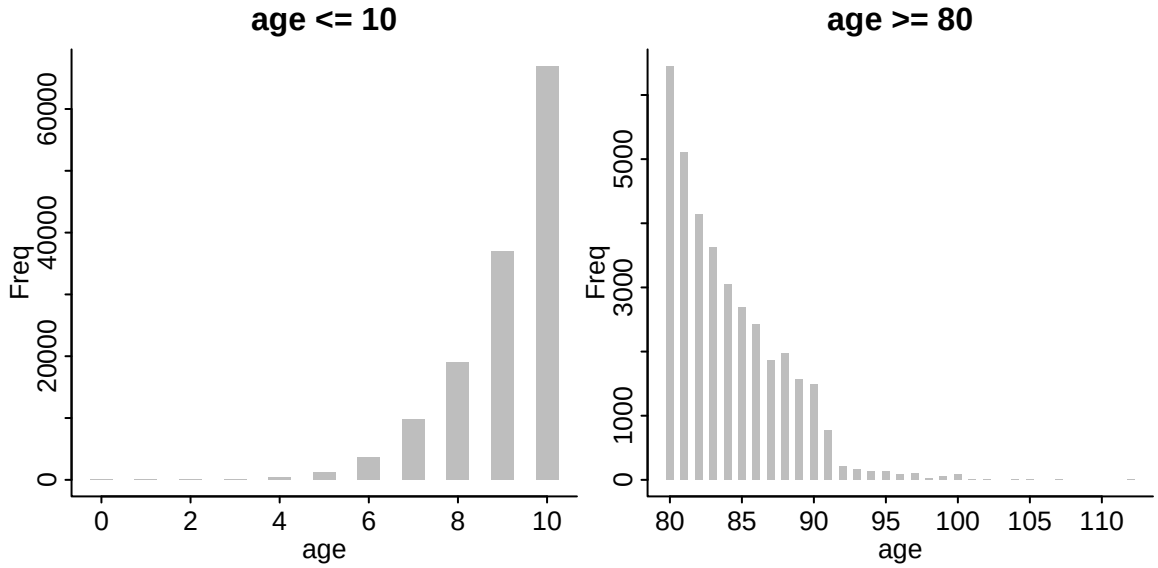
Of the remaining cases where a valid birth date can be derived from the birth year, month, and day, there are 74 cases where the distance from arrest date to birth date does not yield the recorded age.

The MACR manual states that if the specific birth date is unknown, the birth month and day should be recorded as February 30th. This accounts for 105452 of the remaining 133569 records with an invalid birth date. After excluding these records, there are still an anomalous amount of birth dates in February.



Variable Analysis

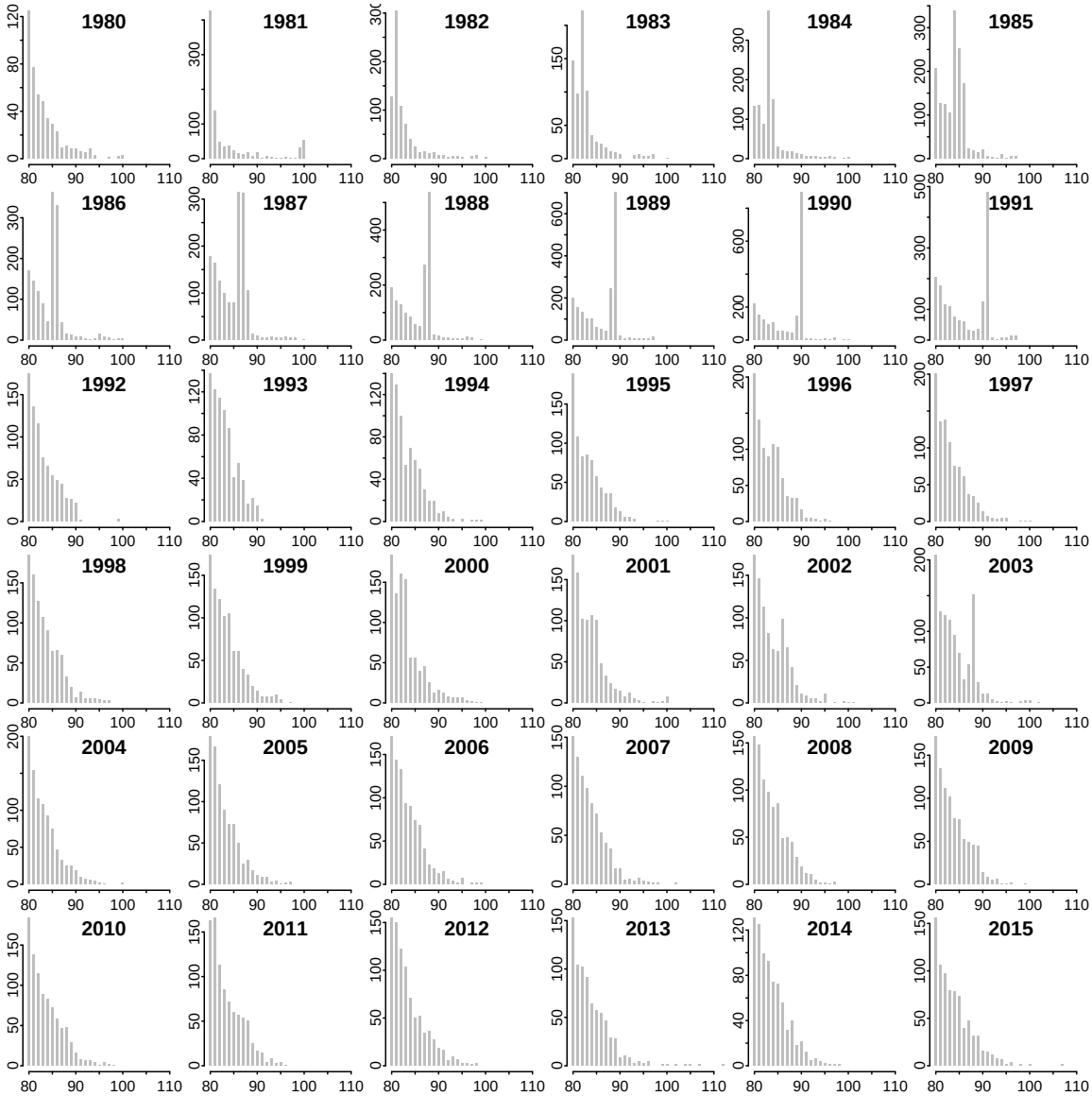
Age



As the above figure shows, there are a large number of arrests for exceptionally young children, and an odd decrease in arrests at age 90. For the young we find:

	age					
offense_level	0	1	2	3	4	5
status offense	1	11	10	36	68	139
misdemeanor	0	10	14	80	261	751
felony	3	8	3	24	113	369

Old age appears to be handled differently in different years. While the numbers are relatively small, it is difficult to believe that there were spikes in crime for, say, 91 year olds in 1991, 92 year olds in 1992 and so on.


1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015

Directly examining these rows shows another form of missingness:

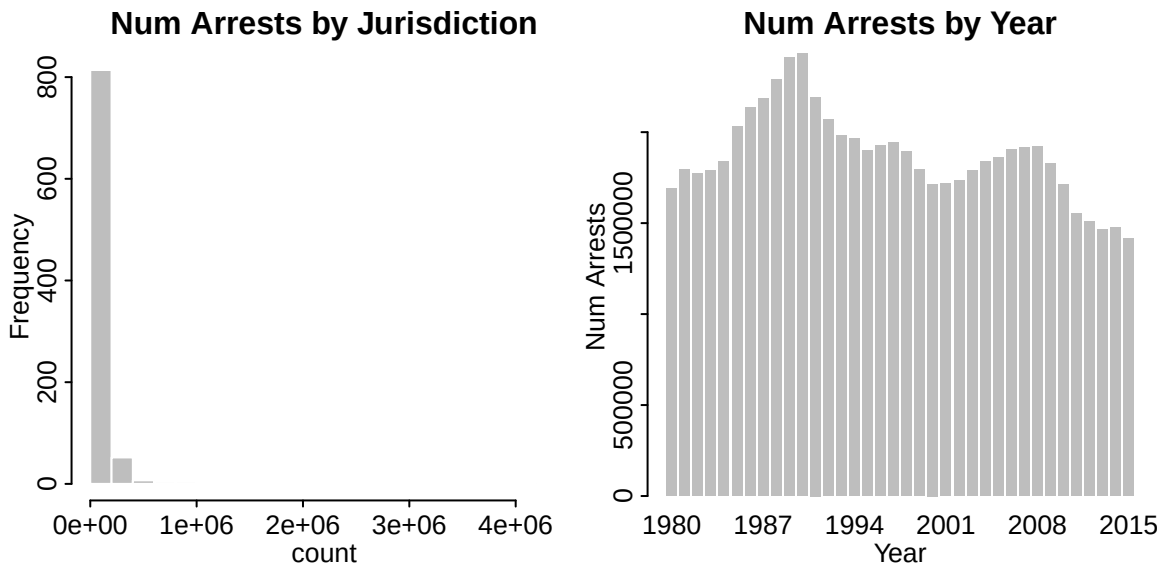
ncic_jurisdiction	arrest_date	age	birth_month	birth_day	birth_date
3711	1991-05-16	91	1	1	1900-01-01
3711	1991-01-12	91	1	1	1900-01-01
0105	1991-01-05	90	1	1	1901-01-01
3711	1991-10-01	91	1	1	1900-01-01
3710	1991-04-24	90	1	1	1901-01-01
3710	1991-06-29	90	1	1	1901-01-01
3711	1991-12-06	91	1	1	1900-01-01
3711	1991-07-18	91	1	1	1900-01-01
3710	1991-05-31	91	1	1	1900-01-01
3711	1991-04-25	91	1	1	1900-01-01
3711	1991-01-13	91	1	1	1900-01-01
3700	1991-09-14	91	1	1	1900-01-01
3710	1991-08-04	90	1	1	1901-01-01
3701	1991-11-07	91	1	1	1900-01-01
3711	1991-11-26	91	1	1	1900-01-01

This spike apparently bubbles through the population until 1992, at which point the practice ended. In 2003, a handful of jurisidctions used an age of 88, together with a birthday of 02/30 to indicate missingness.

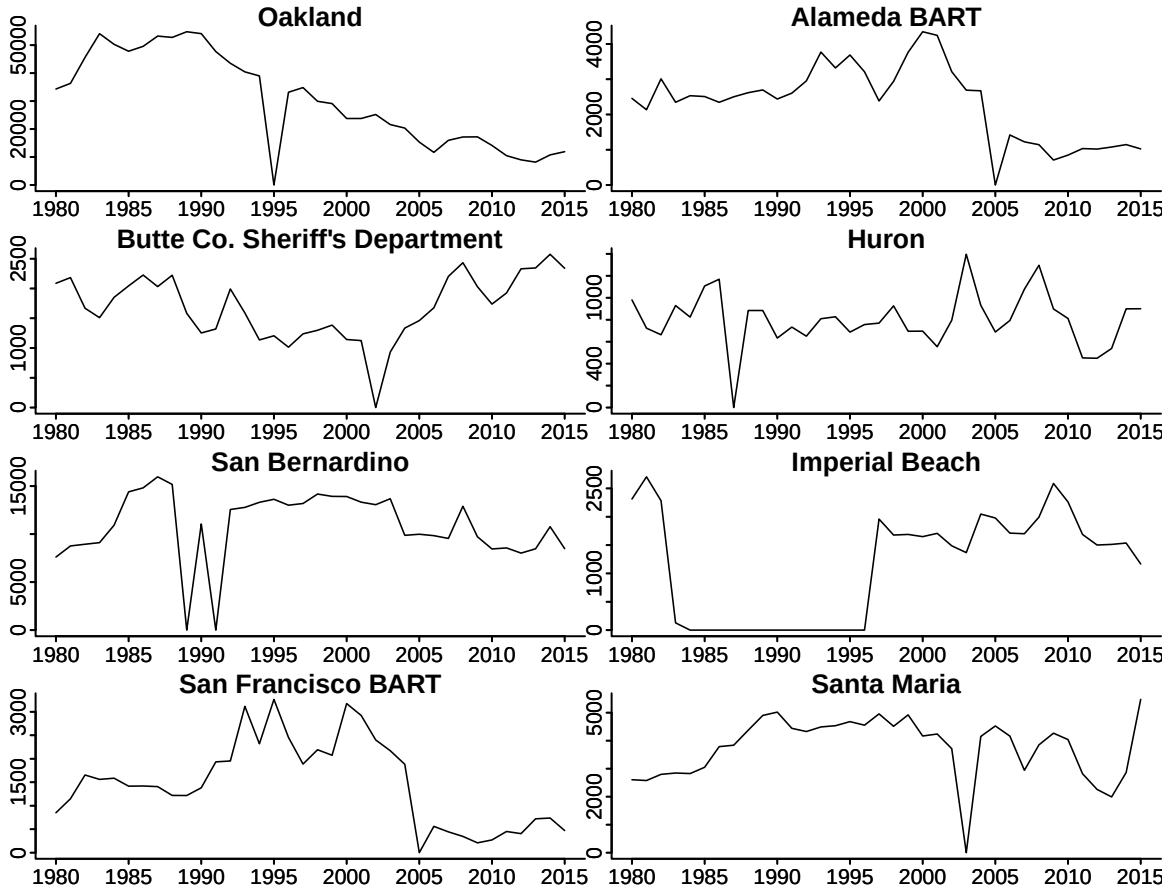
ncic_jurisdiction	arrest_date	age	birth_month	birth_day	birth_date
1005	2003-07-03	88	12	12	1914-12-12
1502	2003-08-30	88	1	20	1915-01-20
1900	2003-01-12	88	2	30	NA
1900	2003-01-26	88	2	30	NA
1900	2003-02-04	88	2	30	NA
1900	2003-02-11	88	2	30	NA
1900	2003-02-14	88	2	30	NA
1900	2003-02-14	88	2	30	NA
1900	2003-03-14	88	2	30	NA
1900	2003-03-16	88	2	30	NA

Jurisdictions

After removing deleted records, there are 889 different NCIC jurisdictions. The number of arrests in each varies wildly, from a minimum of 1 arrest across all 36 years to 4369571 arrests. The largest 396 account for 95% of the records, the smallest of which made 27464 arrests.



Furthermore, the number arrests within jurisdictions also appears to vary wildly over time. A total of 8 have unexpected years with 0 arrests, in some cases dropping from thousands of arrests to return to that rate immediately after.



Variation in Number of Arrests

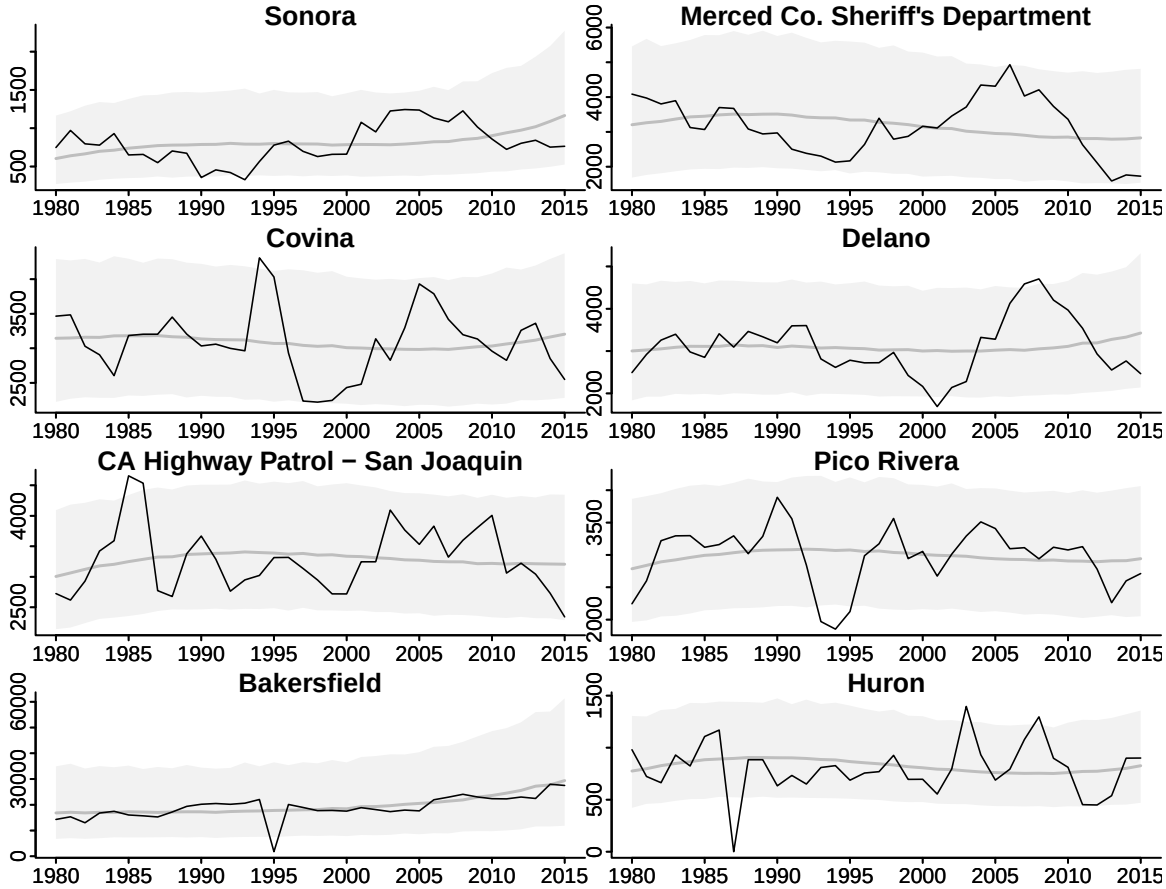
As a preliminary analysis, an hierarchical time series model was fit to the number of arrests in each jurisdiction. Within each jurisdiction, the logarithm of the number of arrests was modeled using a cubic function of time. In addition, the amount of noise in each jurisdiction was also modeled as a cubic function of the size of the jurisdiction. Jurisdiction parameters were directly modeled as "random effects", so that estimates were pooled together and information shared across groups. "Fixed effects" included terms that modeled the overall trend in number of arrests as a cubic polynomial, with contributions from jurisdictions weighted by the average number of arrests per jurisdiction. This is not an ideal model to fit to this kind of data, but should serve to capture the general trends.

Since the model was fit to logarithms, years with zero arrests were excluded.

In the plots that follow, the black lines are the observed number of arrests, while the solid gray line is what the model would predict for a given year (posterior mean). The gray shading shows point-wise 95% central intervals of the posterior predictive distribution prediction - it can be thought of as the range of new obserations that the model deems to be consistent with the observed data, and incorporates both the noise that the observations express as well as uncertainty in the model fit.

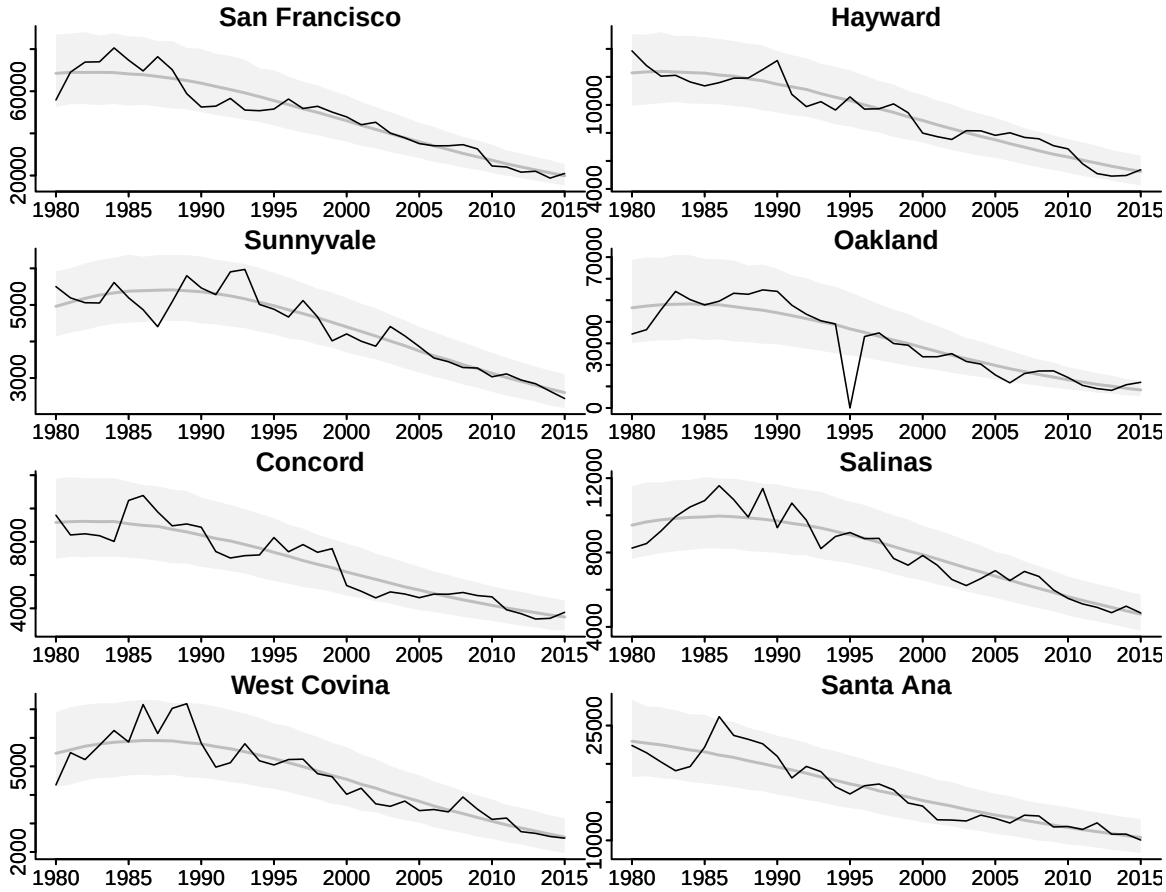
High Residual Variance

This graph shows those jursidictions with the highest amount of noise relative to their sizes. This indicates either that the underlying observations are simply noisy (i.e. there are frequent changes in the number of arrests), or that the model is a poor fit.



Low Residual Variance

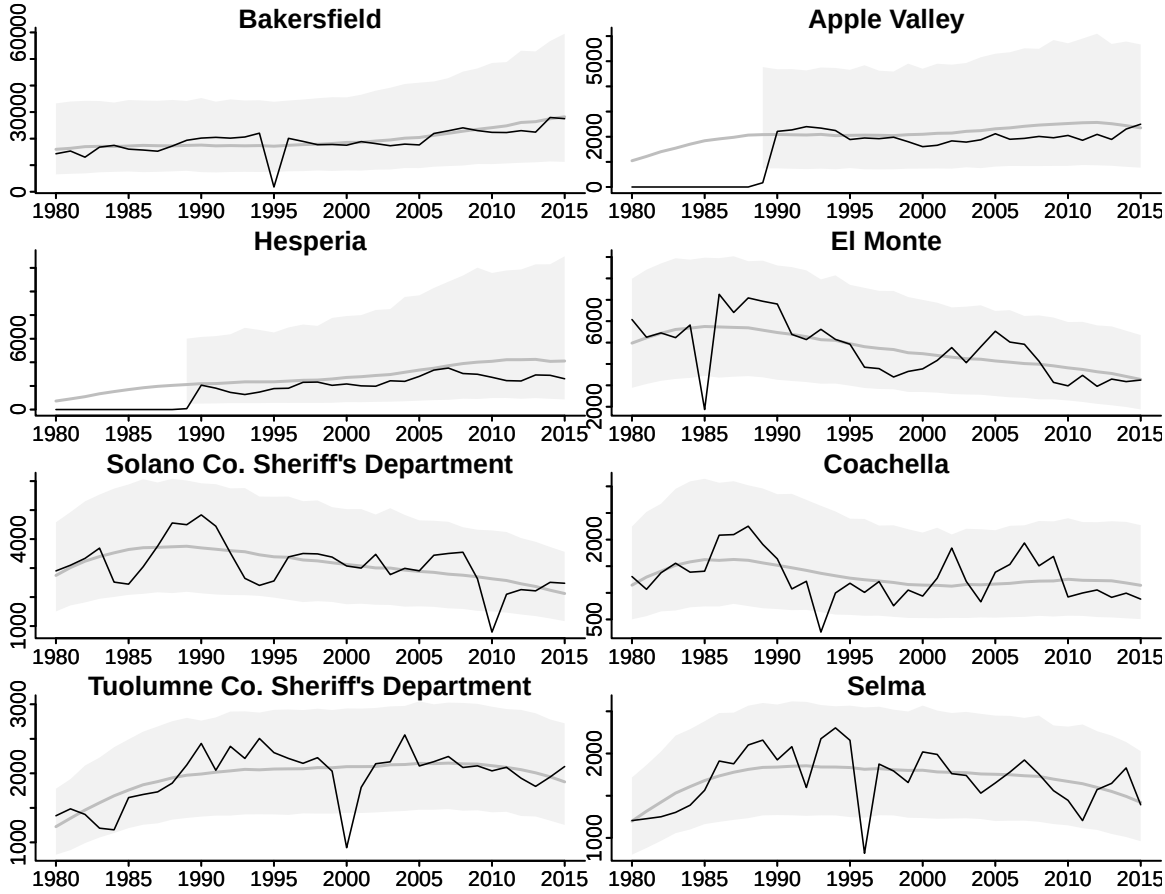
This graph shows those jursidictions with the smallest amount of noise relative to their size. A low amount of variation can either indicate that the number of arrests are too consistent, are naturally not noisy, or that in these circumstances (and not in others), the model fits the data very well. If the model was correct everywhere, then small variance could itself also be a product of chance.



Most Extreme Standardized Residuals

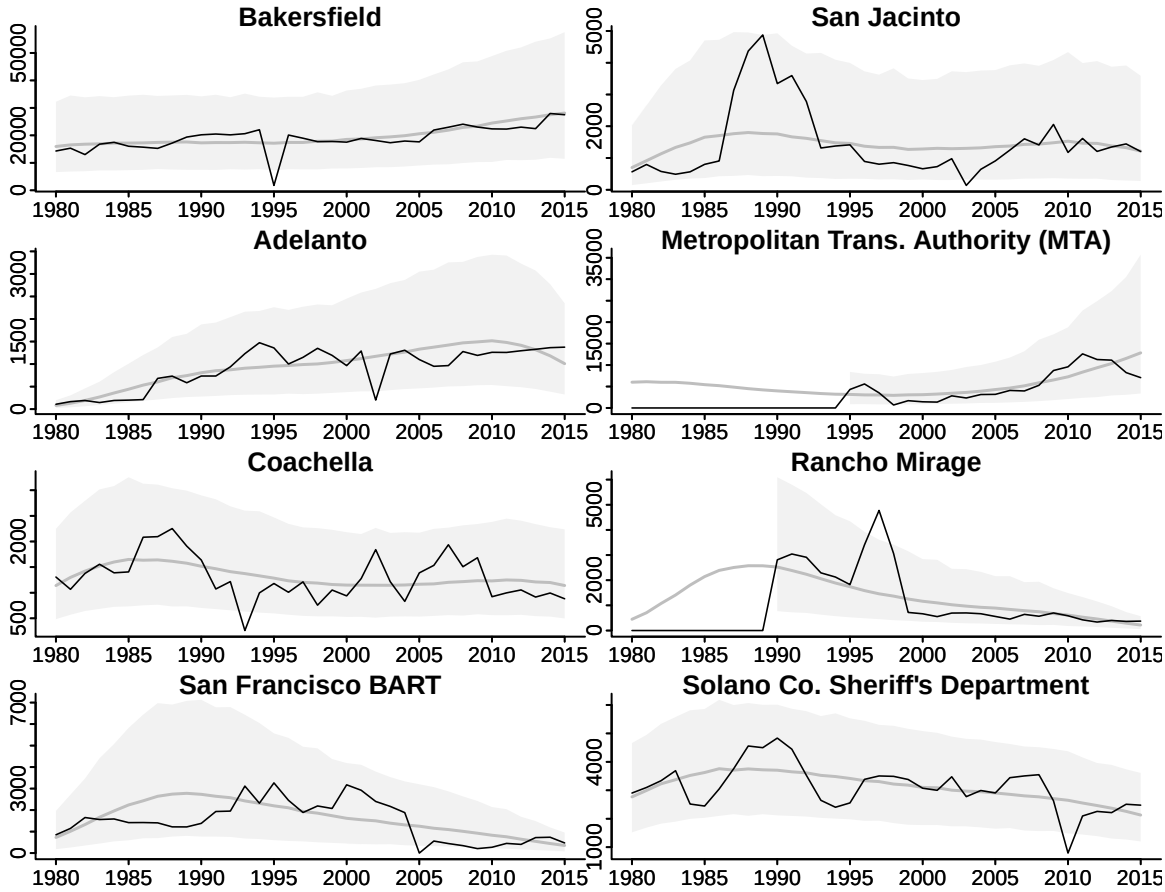
The below shows those jurisdictions with the largest standardized residuals, that is those jurisdictions which have an observation for which the model predicts something wildly different than what was observed in a given year, relative to how noisy that jurisdiction is. If the model appears a poor fit, that is either because it lacks complexity or that, for smaller jurisdictions, the jurisdiction does not follow the general trend of number arrests within the state.

Jurisidictions with small standardized residuals are not show, as they can arise by having a large estimated noise term in addition to simply not being very noisy.



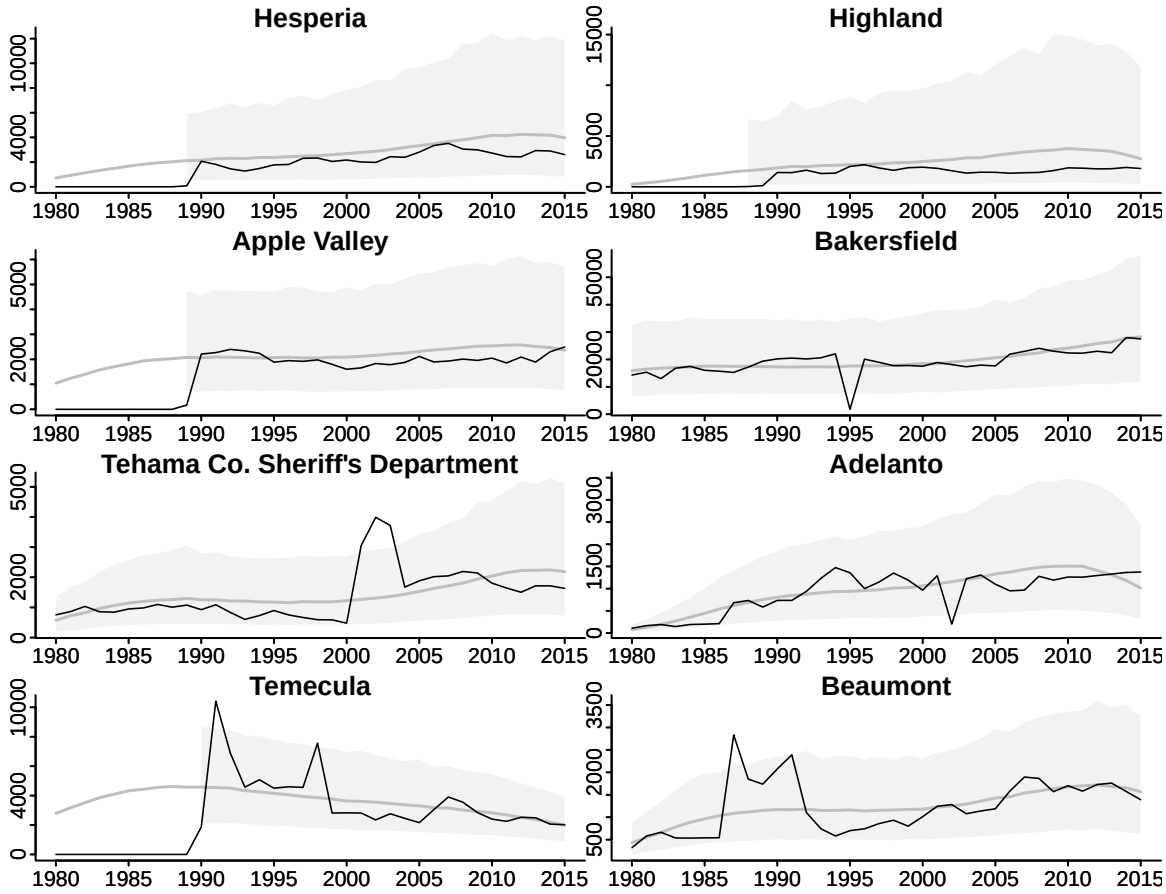
Sharpest Decreases

Finally, as the above were based on a fitted model that does not accurately describe the data in all circumstances, we include an ad-hoc approach that simply looks at those jurisdictions with the largest and smallest increases in the number of arrests going from year to year. These plots show those with years where the number of arrests are the smallest percentage-wise from the previous year, i.e. the sharpest drops.



Sharpest Increases

Conversely, these are those with years having the largest percent increases going from year to year.



Further Analysis

As indicated at the start of this section, the model used can be improved in a number of ways. Trivially, more complicated functions could be fit within jurisdictions, including higher order polynomials or change points. Errors in the models should be examined for serial correlations and lagged predictors added in response.

Finally, it should be addressed that the model is only appropriate for jurisdictions with large numbers of arrests. For the smaller jurisdictions, the most "correct" model would likely be a Poisson regression. In order to also model the larger jurisdictions simultaneously, a finer level of time granularity should be employed.

Booking Rates

This section examines the reliability of booking data from the MACR database (the status_type field). We find that booking data appears to be extremely noisy and generally biased upward. In other words, MACR data about whether an individual was cited or booked is unreliable; an implausibly high percentage of arrests are recorded as booked. We conclude that researchers who decide they want to use these data must be extremely cautious in assessing the data for themselves and in deciding how to use them. We provide a variable loosely describing the quality of booking data by department, but these quality measures are based on heavy assumptions and should not be the only basis for deciding which data to use.

The first reliability analyses focus on the plausibility of booking rates given what we know from talking with law enforcement personnel. For instance, we consider whether a jurisdiction reports booking 100% of arrests or whether a jurisdiction tends to book juveniles for misdemeanors. In discussions with law enforcement personnel (officers, IT staff, department analysts), we learned that officers typically cite and release for misdemeanors and book for felonies. Individuals we spoke with said they would be skeptical if a department claimed it booked the majority of misdemeanors (only those which pose potential for immediate harm, such as domestic violence, are candidates for booking). Booking someone takes at least 1.5-2 hours, potentially diverting an officer from responding to calls, and requiring a great deal of tedious paperwork. Individuals who are on probation or parole who are arrested for a misdemeanor must be booked (roughly 22% of individuals arrested in CA are on probation or parole). In addition, individuals who have a warrant out must be booked, and the MACR will show the new crime for which they were arrested. If an individual was stopped, but not arrested until the officer discovered the outstanding warrant, the warrant offense code will be shown on the MACR.

The second portion of the reliability analyses are based on a probabilistic matching of reported bookings to actual bookings. The reported bookings come from the MACR, while the actual bookings are drawn from the Automated Criminal History System (ACHS). Records in ACHS are created when someone is fingerprinted during jail intake (i.e., booking). Due to data quality issues and because the matching was probabilistic, there are fewer matched bookings in our analyses than there would be if the matching process matching were perfect. This has the overall effect of reducing the percentage of reports of bookings that appear accurate. Only about one in five reported bookings could be matched to an actual booking. The percentage match by department is normally distributed around this figure. In other words, for the average department, we can ony confirm that one in five reported bookings had a corresponding record in a database of actual bookings. We do not know if one in five therefore reflects "accurate" reporting or whether there are systematic biases that have reduced the match percentage beyond artefacts of the data linkage process. Both systematic biases and data linkage problems may produce the one in five figure.

One way to examine the plausibilty of booking data by department is to consider the severity of the offense. We expected that arrests for more severe crimes would have a higher match percentage because they were more likely to have actually been booked. Arrests for less serious crimes, that were in fact cited, would have a lower percentage match. Surprisingly, the booking match percentage did not vary much by the severity of offense. The booking match percentage only declines slightly from homicide to misdemeanors. One unexplored potential explanation is that there are missing data about arrests for less serious crimes (perhaps due to automatic entry of bookings into the agency's record-management system). In other words, we may not know the real denominator of total arrests for less serious crimes. This would affect studies that examine the propensity to book arrested individuals across department, crime, or demographic group.

Eliminating Jurisdictions

This section uses the MACR file to aggregate and display booking rates for felonies and misdemeanors by first removing jurisdictions with anomalous booking rates.

Sample Size

The first cut we make is to drop jurisdictions that make a small number of arrests overall. By considering the cumulative distribution of the number of arrests, we can set cut points at arbitrary percentages. To start we include only those jurisdictions responsible for 95% of all arrests.

Abnormal Booking Rates

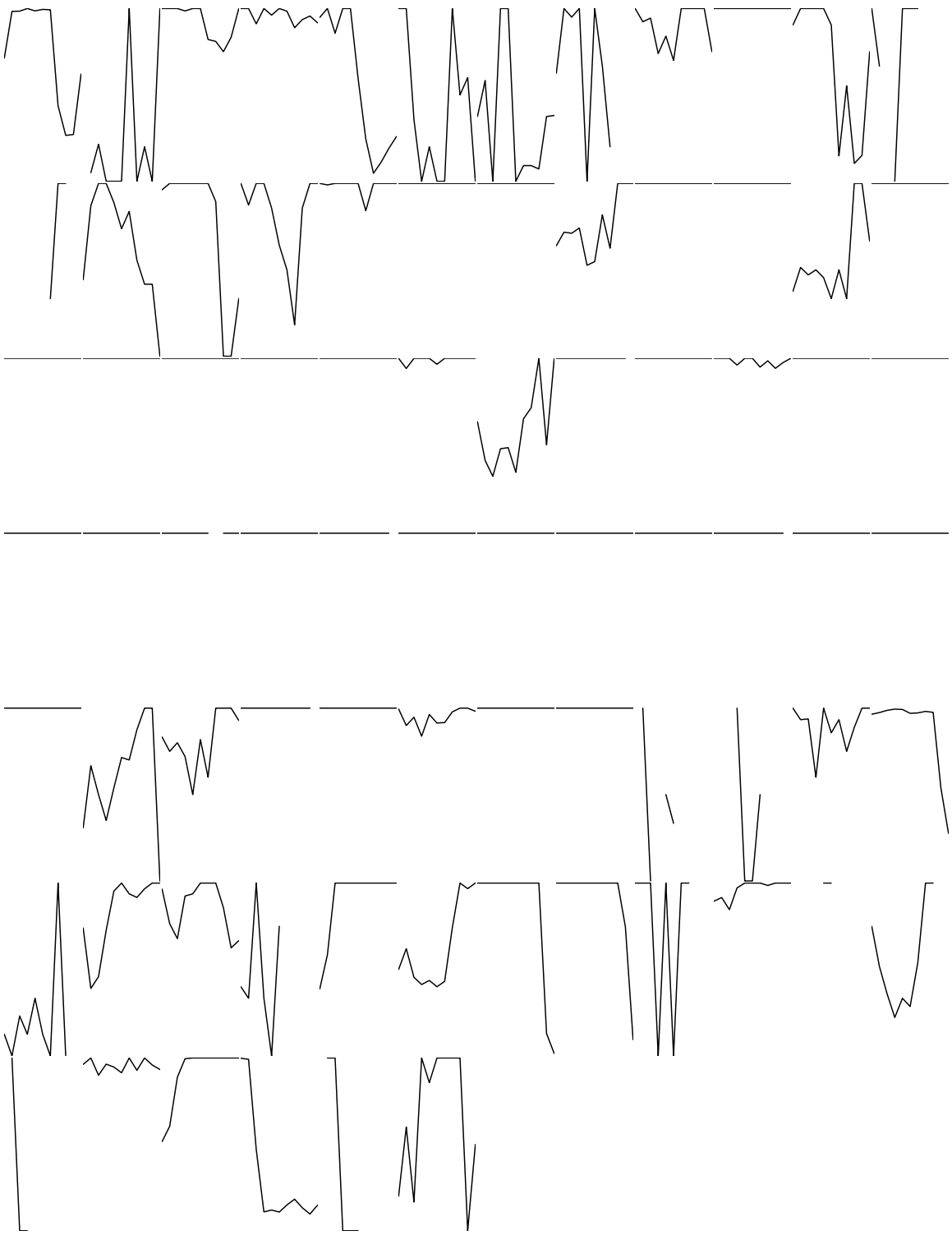
Our goal at this stage is to assess the reliability of jurisdictions in their reporting. We can look for jursidictions that are reporting abornomally high booking rates for less serious crimes or low booking rates for serious crimes. We first focus on booking rates of juveniles for less serious crimes. An analysis of booking rates for individual offense codes (not shown here) indicates a large amount of variability across years and jursidictions, so we instead consider measures that aggregate arrests *before* computing booking rates. That is, we define rates such as:

$$\text{mostly-harmless-booking-rate} = \frac{\text{num juveniles booked for selected codes}}{\text{total num juveniles arrested for selected codes}}$$

Mostly-Harmless Misdemeanors

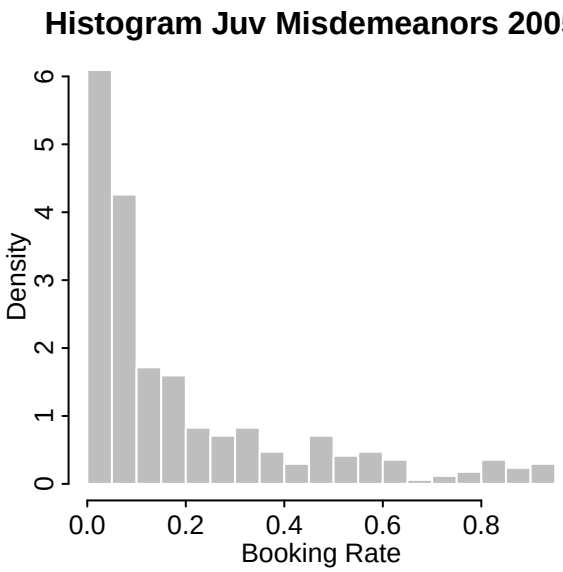
We begin by examining the booking rates for juveniles arrested for less serious offenses, which we would expect to be very low if the data are accurate and comprehensive.

The codes are: Petty Theft, Other Theft, Marijuana, Liquor Laws, and Disturbing the Peace. One immediate cause for concern is the number of jursidictions that, for some years, have reported booking rates close to 100%. Specifically,78 jursidictions that have a booking rate in excess of 0.99 for at least two years.

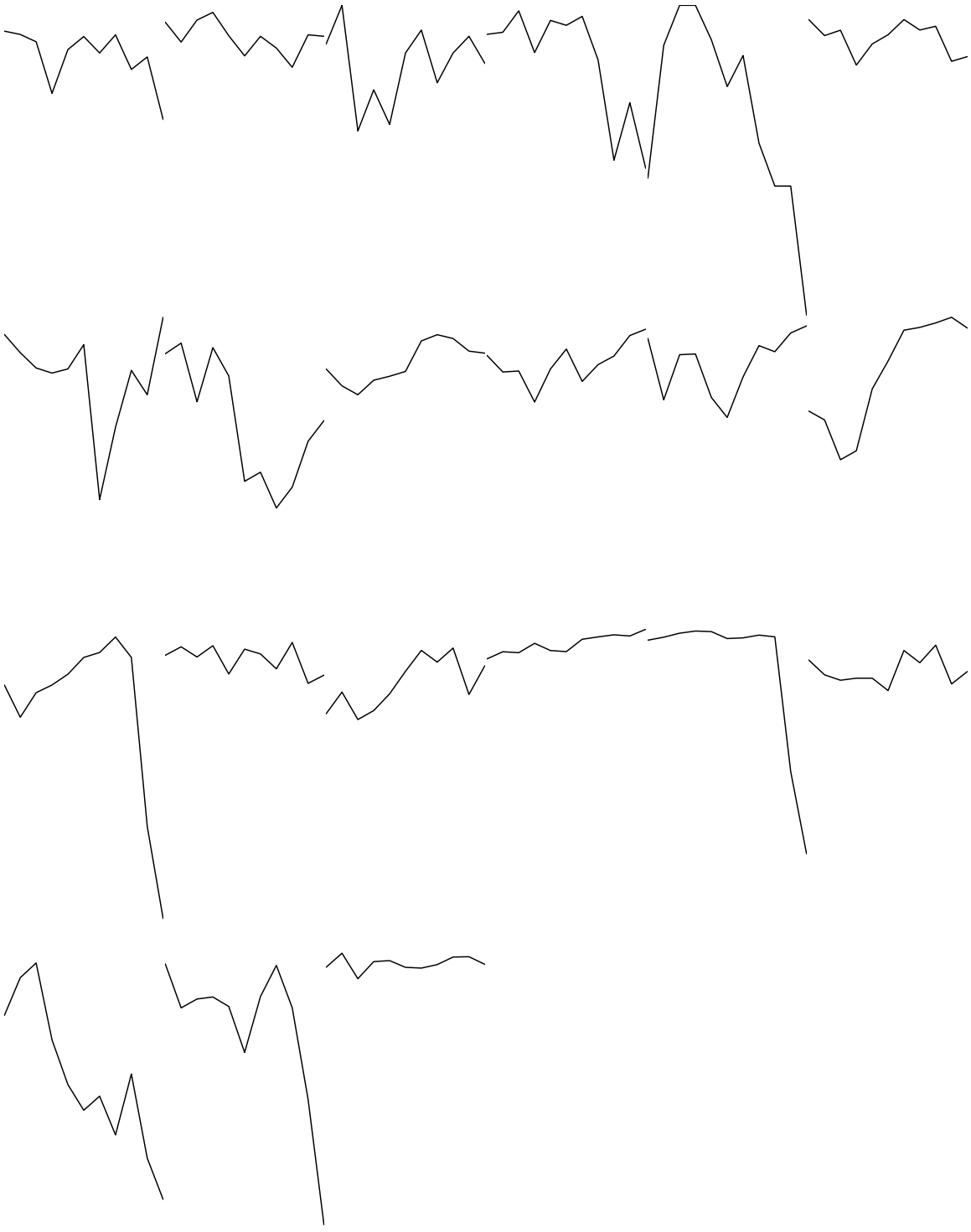


Although a 100% booking rate is implausible some smaller jurisdictions may have very high booking rates due to small sample sizes. Arbitrarily, we proceed by excluding those that approach 100% three or more times. Another simple cut can be made by excluding jursidictions that are mostly missing, such as those for which have only one recorded value.

Next, we average each jurisdiction's arrests for misdemeanors across years and look at the empirical distribution of booking rates.



If we assume that the data are relatively complete, there appears to be a plausible cutoff around 0.65, which includes 21 jurisdictions. Their yearly booking rates appear as:

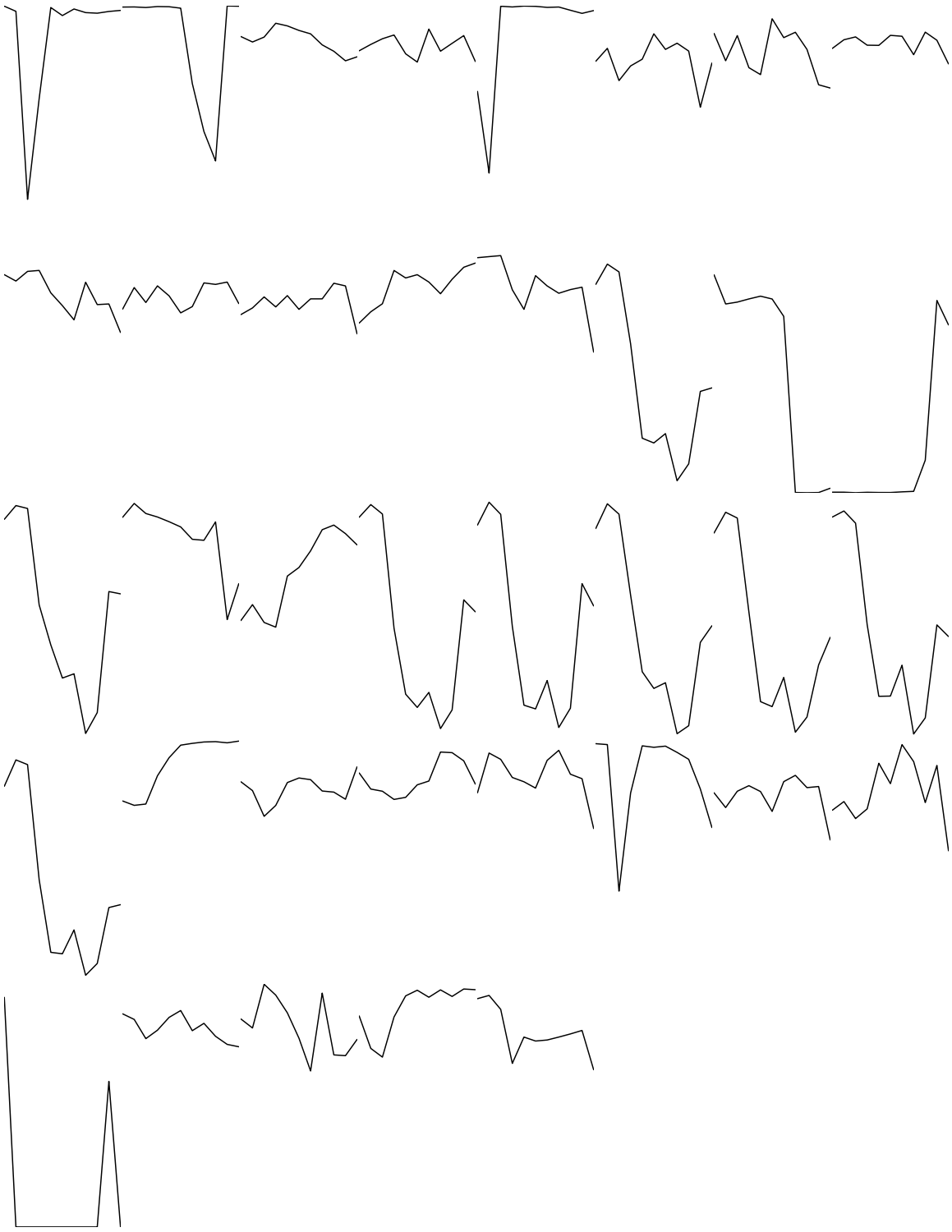


While some of these may certainly be valid, we again arbitrarily exclude them from further analysis.

Serious Felonies

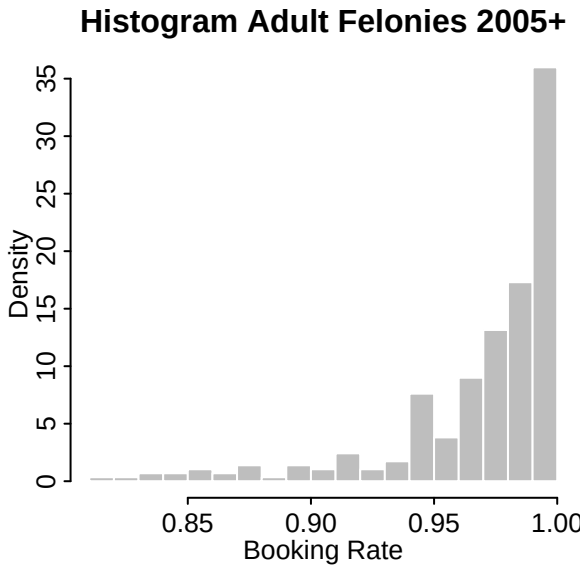
We repeat the above on the diminished set of jurisdictions, but instead looking at arrest rates for serious felonies. Our goal is to identify jurisdictions that report implausibly low booking rates for serious felonies.

The codes correspond to: Assault, Dangerous Drugs, Other Felony, Burglary, Theft, Narcotics, and Outside Warrant. Because the rates for these felonies should be high, we highlight any jursidiction that, for two or more years, had a booking rate lower than 80%. This includes 37 jursidictions:



For comparison, the typical jurisdiction has a straight line near 100%. Some of these look categorically low, including the set in the middle which all decline and increase with each other; others seem like the natural variation that might come with small sample sizes. For that reason, we set the threshold at three years.

Again we look at the empirical distribution of the aggregate across years.

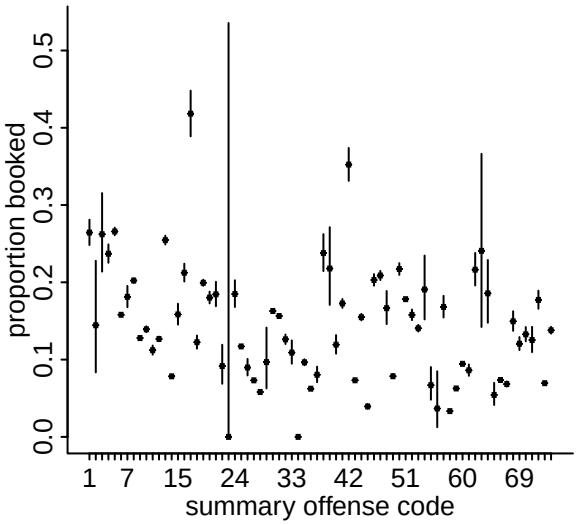


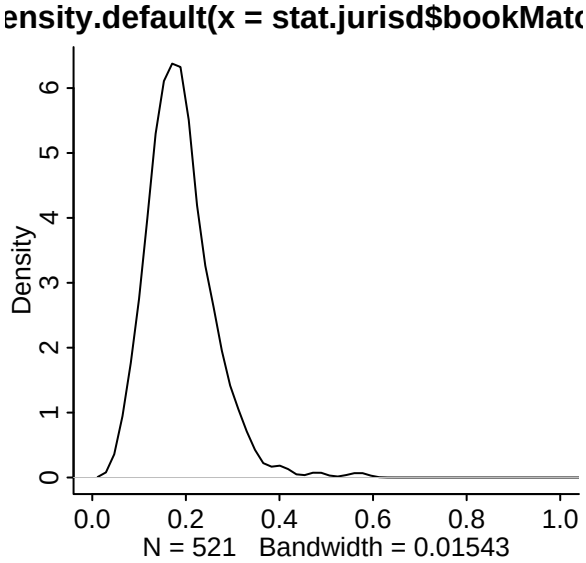
Matched ACHS-MACR data

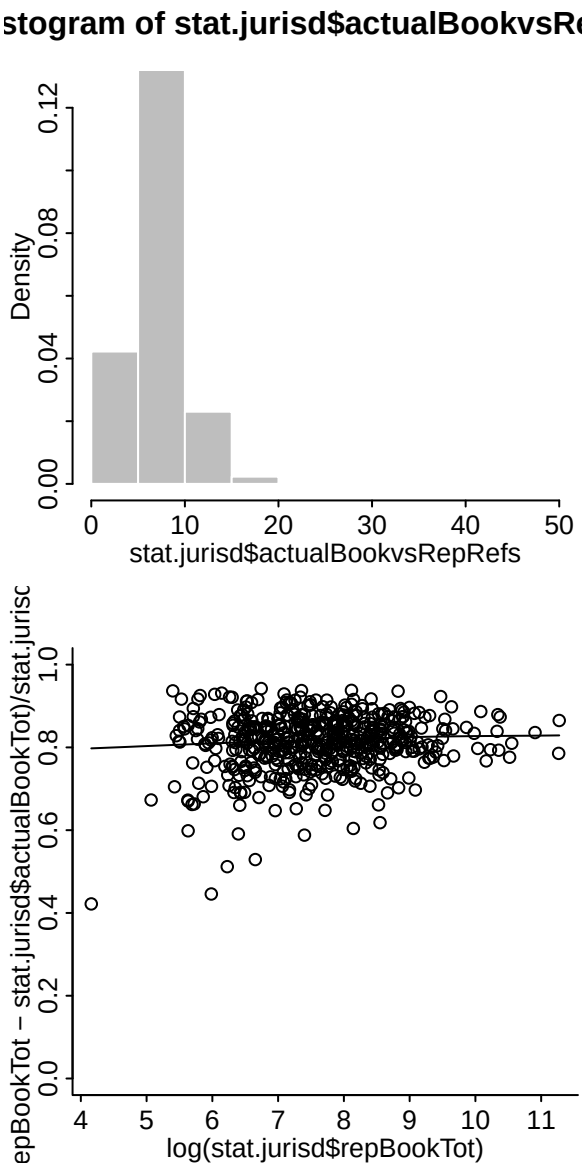
We obtained a probabilistic record linkage for 2014 and 2015 ACHS and MACR data. We use the linked records to estimate the percentage of arrests reported as booked (from MACR) that were actually booked by department (from ACHS). In other words, we use a simple metric: the number of actual bookings (MACR records linked to ACHS) divided by the number of reported bookings (total bookings in MACR).

We do not possess a full count of individuals who appear in ACHS. This would have allowed us to directly compare the total number of reported bookings to the total number of actual bookings. We would also be able to see what percentage of people who were booked did not appear in the MACR (though DOJ CJIS retroactively fills some of these in). The data we use here contain only the records that could plausibly be linked between MACR and ACHS. The total number of matches was about 374,000 and there were about 2M reported bookings over the same period (out of 2.9M arrests). We believe the matching process misses many true matches (a high false negative rate), which means that we expect that even departments that are accurately reporting bookings would fall well below 100%.

To examine the quality of booking data, we primarily focus on deparments and offenses. Departmental practices and systems may be responsible for the seemingly small percentage of MACR reported bookings that show up in ACHS. Among the 75% of departments from which we would expect enough bookings to compute statistics (>95 felony arrests), we find large variation in the fraction of reported bookings that could be matched to actual bookings. Apart from deparments reporting bookings incorrectly, this could also be due to department-level variation in the quality of personal data reported on the MACR. If some departments reported poor name and date of birth data, we would find fewer corresponding records in ACHS.


proportion booked
0.0
0.1
0.2
0.3
0.4
0.5
1
7
15
24
33
42
51
60
69
summary offense code

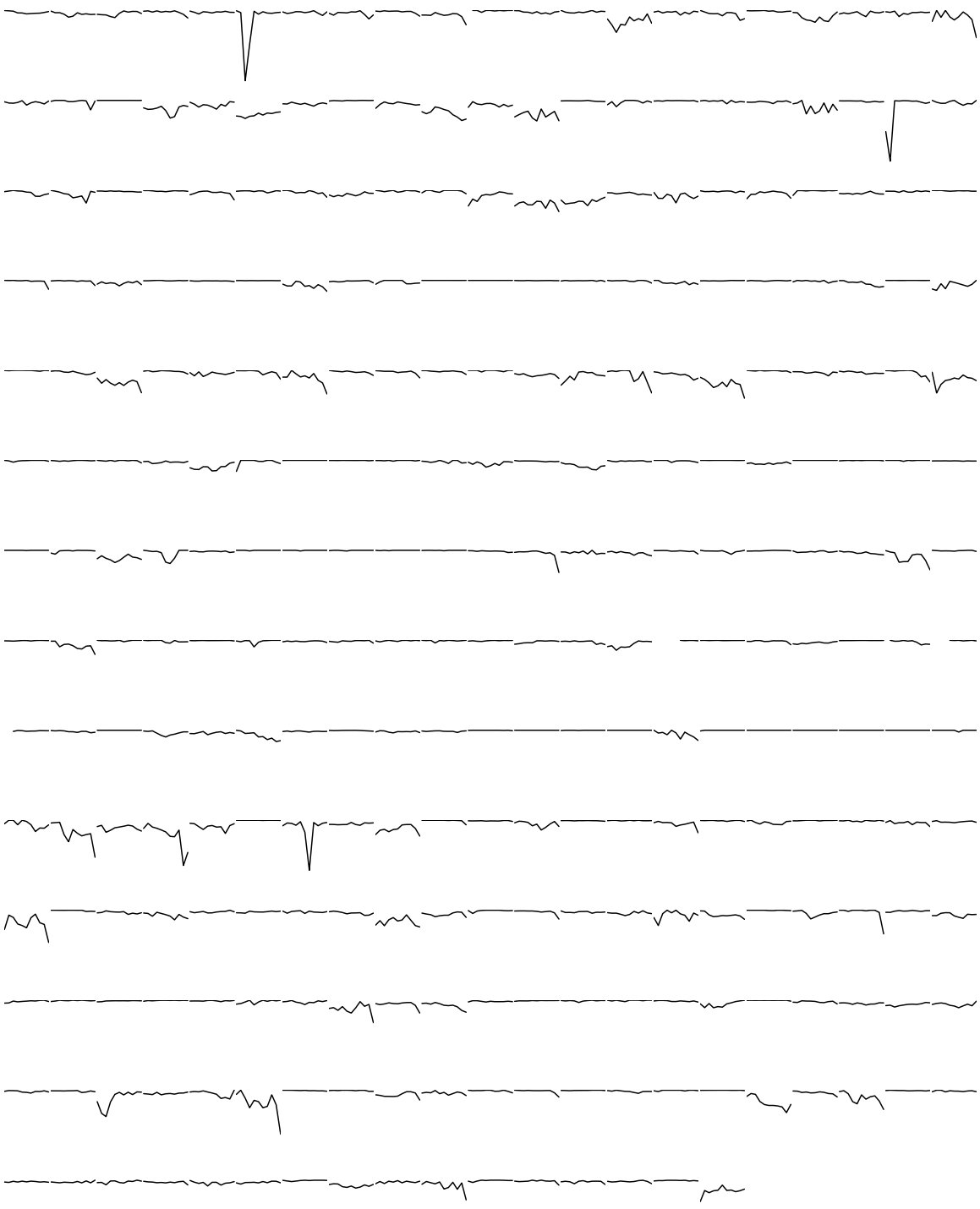

ensity.default(x = stat.jurisd$bookMatc
Density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
N = 521 Bandwidth = 0.01543



Final Visualization

At this point, the 289 jurisdictions that remain seem to be adequate in terms of variabilty. Their booking rates for juvenile misdemeanors are:

The yearly booking rates for serious felonies are:



Booking Rate Tables

Juvenile Misdemeanors

offense code	N	bkd	ctd	N wt	bkd wt	ctd wt	N bl	bkd bl	ctd bl	N hs	bkd hs	ctd hs
6-Assault	751389	0.97	0.02	244963	0.97	0.02	153184	0.97	0.02	300661	0.97	0.02
14-Dangerous	530254	0.97	0.02	237826	0.97	0.02	45117	0.97	0.01	216783	0.97	0.01
25-Other Fel	472710	0.94	0.04	179048	0.95	0.03	93191	0.97	0.02	174617	0.93	0.04
8-Burglary	383872	0.91	0.05	134287	0.91	0.05	83025	0.93	0.04	141676	0.89	0.06
9-Theft	342573	0.94	0.04	126264	0.95	0.03	70483	0.95	0.04	122417	0.93	0.04
12-Narcotics	341012	0.97	0.01	110888	0.97	0.02	122728	0.98	0.01	93183	0.97	0.01
27-Outside W	268358	0.98	0.01	104452	0.98	0.01	59744	0.99	0.01	88469	0.98	0.01
28-Probation	259211	0.97	0.00	93251	0.98	0.00	66153	0.98	0.00	89081	0.96	0.00
19-Weapons	180822	0.90	0.07	48036	0.90	0.07	36032	0.95	0.03	86521	0.88	0.09
5-Robbery	149626	0.96	0.02	27690	0.96	0.02	58792	0.95	0.02	56091	0.96	0.02
10-Motor Veh	131727	0.94	0.03	39117	0.94	0.03	24264	0.94	0.03	60636	0.93	0.04
13-Marijuana	109143	0.91	0.06	38219	0.90	0.07	27650	0.96	0.03	34270	0.89	0.07
11-Forgery/C	64354	0.96	0.02	21228	0.96	0.02	14603	0.96	0.02	23589	0.96	0.02
20-Drive Und	41364	0.94	0.04	15223	0.93	0.04	3929	0.95	0.03	19461	0.95	0.03
18-Other Sex	32969	0.93	0.05	12347	0.94	0.04	8541	0.95	0.04	9886	0.90	0.06
16-Lewd or L	23082	0.92	0.05	6150	0.90	0.06	2512	0.89	0.06	13014	0.93	0.04
4-Rape	16229	0.96	0.02	3972	0.96	0.02	3498	0.96	0.02	7696	0.97	0.01
26-Federal O	15753	0.99	0.01	1777	0.97	0.02	489	0.94	0.04	13152	0.99	0.00
1-Homicide	12996	0.98	0.00	2658	0.99	0.00	3382	0.98	0.00	6023	0.98	0.00
7-Kidnapping	12981	0.98	0.01	3081	0.98	0.01	3027	0.98	0.00	6087	0.98	0.01
15-Other Dru	11451	0.95	0.03	5302	0.96	0.03	1997	0.97	0.02	3534	0.93	0.04
21-Hit and R	9977	0.85	0.10	2769	0.83	0.14	1079	0.87	0.09	5288	0.85	0.09

Adult Felonies

offense code	N	bkd	ctd	N wt	bkd wt	ctd wt	N bl	bkd bl	ctd bl	N hs	bkd hs	ctd hs
6-Assault	751389	0.97	0.02	244963	0.97	0.02	153184	0.97	0.02	300661	0.97	0.02
14-Dangerous	530254	0.97	0.02	237826	0.97	0.02	45117	0.97	0.01	216783	0.97	0.01
25-Other Fel	472710	0.94	0.04	179048	0.95	0.03	93191	0.97	0.02	174617	0.93	0.04
8-Burglary	383872	0.91	0.05	134287	0.91	0.05	83025	0.93	0.04	141676	0.89	0.06
9-Theft	342573	0.94	0.04	126264	0.95	0.03	70483	0.95	0.04	122417	0.93	0.04
12-Narcotics	341012	0.97	0.01	110888	0.97	0.02	122728	0.98	0.01	93183	0.97	0.01
27-Outside W	268358	0.98	0.01	104452	0.98	0.01	59744	0.99	0.01	88469	0.98	0.01
28-Probation	259211	0.97	0.00	93251	0.98	0.00	66153	0.98	0.00	89081	0.96	0.00
19-Weapons	180822	0.90	0.07	48036	0.90	0.07	36032	0.95	0.03	86521	0.88	0.09
5-Robbery	149626	0.96	0.02	27690	0.96	0.02	58792	0.95	0.02	56091	0.96	0.02
10-Motor Veh	131727	0.94	0.03	39117	0.94	0.03	24264	0.94	0.03	60636	0.93	0.04
13-Marijuana	109143	0.91	0.06	38219	0.90	0.07	27650	0.96	0.03	34270	0.89	0.07
11-Forgery/C	64354	0.96	0.02	21228	0.96	0.02	14603	0.96	0.02	23589	0.96	0.02
20-Drive Und	41364	0.94	0.04	15223	0.93	0.04	3929	0.95	0.03	19461	0.95	0.03
18-Other Sex	32969	0.93	0.05	12347	0.94	0.04	8541	0.95	0.04	9886	0.90	0.06
16-Lewd or L	23082	0.92	0.05	6150	0.90	0.06	2512	0.89	0.06	13014	0.93	0.04
4-Rape	16229	0.96	0.02	3972	0.96	0.02	3498	0.96	0.02	7696	0.97	0.01
26-Federal O	15753	0.99	0.01	1777	0.97	0.02	489	0.94	0.04	13152	0.99	0.00
1-Homicide	12996	0.98	0.00	2658	0.99	0.00	3382	0.98	0.00	6023	0.98	0.00
7-Kidnapping	12981	0.98	0.01	3081	0.98	0.01	3027	0.98	0.00	6087	0.98	0.01
15-Other Dru	11451	0.95	0.03	5302	0.96	0.03	1997	0.97	0.02	3534	0.93	0.04
21-Hit and R	9977	0.85	0.10	2769	0.83	0.11	1079	0.87	0.09	5288	0.85	0.09
24-Arson	9600	0.76	0.17	4426	0.76	0.17	1468	0.82	0.12	3020	0.73	0.18

To summarize, we know that: - Due to the high false negative rate in linking records, the match percentage of a department that accurately reports bookings will be below 100% - Reported booking rates, even for less serious misdemeanors, appear to be too high - The percentage of individuals arrested who are on parole or probation (22%) is too low to explain these high booking rates - The only hypothesis we have that would explain all these issues is that the MACR reflects under-reporting of arrests that are not booked (typically, those that are less serious), thus shrinking the denominator on booking rates and making it seem that booking rates are very high across all crimes.