OBTS Data Inventory

Introduction

OBTS Main Database

The Offender Based Transaction Statistics database is an extract of California's Automated Criminal History System (ACHS), which is used, among other things, for background checks. The OBTS is intended to capture the arrest cycle for all felonies in ACHS, beginning with arrest and ending with a release, a decline to prosecute, a not guilty verdict, or a guilty verdict and a sentence. OBTS data begins in 1982 and contains 250,000 to 400,000 records per year. Each record includes the arresting officer's jurisdiction, the arrest date and BCS code for offense, the age/birth date of the arrestee, the arrestee's demographics, and, if applicable, charges pressed, charges convicted, and sentence type. The OBTS also contains direct identifiers such as name, CII, and DoB.

Sample Rows

id	cii_record_ty	cii_number	pdr_record_id	name	birth_year	
1	OBN	SEALED	unknown	SEALED	1993	
2	automated	SEALED	$\operatorname{current_offen}$	SEALED	1932	
3	automated	SEALED	$\operatorname{current_offen}$	SEALED	1933	
4	automated	SEALED	$\operatorname{current_offen}$	SEALED	1931	
5	automated	SEALED	$\operatorname{current_offen}$	SEALED	1932	
• • •	•••	• • •	• • •	• • •	• • •	

Variable Summary

name	type	value	description
id	numeric	1-9654994	
cii_record_type	factor	automat/h	ybrid/m
cii_number	character	pii	Criminal Identification and
			Information number
pdr_record_id	factor	current/ne	w_offPersonal Descriptor Record ID
name	character	pii	
birth_year	integer	1891-1999	
birth_month	integer	1-12	
birth_day	integer	1-31	
gender	factor	female/male/unknown	
race	factor	america/asian_i	
deceased	factor	deceased/deceasewhether individual is deceased	
arrest_record_id	factor	final_c/pol	lice/plast reported step in crim justice
			process
arrest_bypass	factor	no_bypa/r	not_cor
$arrest_converted_data$	factor	convert/no	t_conconverted from old format
arrest_event_year	integer	1929-2015	
arrest_event_month	integer	1-12	
arrest_event_day	integer	1-31	
arresting_agency	factor	010000/01000	1/01010.arresting or booking agency
local_file_number	character	pii	

name	type	value	description
arrest_charge_type	factor	felony/misdeme	
multiple_arrest_charges	factor	1/2/3/4/5/6/7/8/9	
arrest_disposition_year	integer	1944-2015	date of arresting officer disposition
$arrest_disposition_month$	integer	1-12	
arrest_disposition_day	integer	1-31	
$arrest_disposition_type$	factor	1/2/3/4/5/6/7/10/13	1
$arrest_multiple_dispositions$	factor	0/1/2/3/4	
$prior_record_code$	factor	length/misc.pr.	•••
$arrest_offense$	factor	1004/1005/3005/9004	4
arrest_qualifier	character	$\dots/AQ\dots$	
$arrest_summary_code$	factor	1/2/3/4/5/6/7/8/9/3	1.BCS summary code for arrest offense
$arrest_edit_error_code$	character	/C /B /C	
$\operatorname{court}\operatorname{\underline{\hspace{1pt}rec}ord}\operatorname{\underline{\hspace{1pt}id}}$	factor	$lower_c/superio$. court of final disposition
court_bypass	factor	$correct/no_bypa.$	•••
$court_converted_data$	factor	1	
court_event_year	integer	1982-2014	last date of court proceedings
$court_event_month$	integer	1-12	
court_event_day	integer	1-31	
$court_judicial_district$	factor	01****/01100J/01120	0
court_number	character	pii	
$court_disposition_type$	factor	acquited/certifi	outcome of court trial
court_multiple_dispositions factor		1/2/3/4/5/6/7/9	
$\operatorname{court_proceeding_type}$	factor	$\operatorname{certifi/felony}$	
sentence_type	factor	1/2/3/4/5/6	
sentence	factor	CRC/CYA/death/fin	
court_charge_type factor		felony/infract/moffense level of conviction crime	
court_multiple_disposition_chafagetsor		1/2/3/4/5/6/7/8/9	
$court_disposition_offense$	factor	1004/3005/4016/8001	
court _qualifier	factor	/ KA	
$court_summary_code$	factor	1/2/3/4/5/6/7/8/9/1	
court_edit_error_code	character	/D /K /F	

Tables

Main Table Variables

 ${\bf cii_record_type}$

Description

CII record type

Labels

- \bullet 1 automated
- 2- manual
- 3 OBN (no CII record on file)
- 4 Hybrid (began 1993)

Name	Freq
automated	8782819
OBN	398939
manual	316923
hybrid	100757
NA	55556

${\bf cii_number}$

${\bf Description}$

Statewide unique identification number for an individual.

${\bf Prompt}$

Enter the unique state identification number, referred to as the "CII#", assigned to the subject's criminal history record. If no CII number is available or known to the agency, leave the area blank and continue processing the form.

pdr_record_id

${\bf Description}$

ID in the Personal Descriptor Record.

Labels

- 00 new offender (no arrest before 8/23/82)
- 01 current offender (arrested before 8/23/82)
- 02 unknown
- $\bullet~$ 09 delete CII

Notes

DOES THIS DATA GO PAST 1980s?

Name	Freq
new_offender	7238025
current_offender	1701846
unknown	715123

name

${\bf Description}$

Individual's name.

Prompt

Enter the complete name of the subject as it was submitted on the arrest fingerprint submission or as indicated on the citation. Aliases, nicknames, or monikers are not required.

${\bf birth_year}$

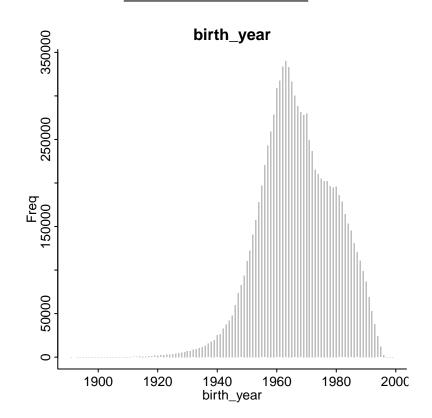
Description

Individual's birth year.

Prompt

Enter the month/day/year (MM/DD/YYYY) of the subject's birth.

Name	Value
Min.	1891.000
1st Qu.	1959.000
Median	1966.000
Mean	1967.062
3rd Qu.	1976.000
Max.	1999.000



$birth_month$

Description

Individual's birth month.

Prompt

Enter the month/day/year (MM/DD/YYYY) of the subject's birth.

Name	Value
Min.	1.000000
1st Qu.	4.000000
Median	7.000000
Mean	6.568342
3rd Qu.	10.000000
Max.	12.000000
NA's	66806.000000



birth_day

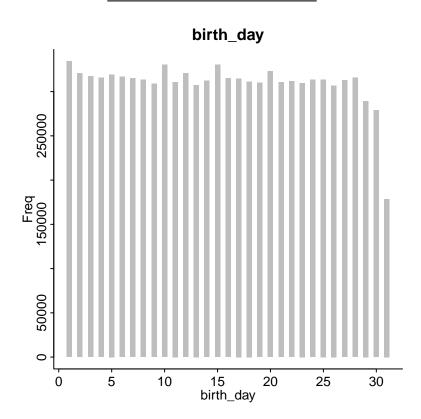
Description

Individual's birth day.

Prompt

Enter the month/day/year (MM/DD/YYYY) of the subject's birth.

Name	Value
Min.	1.00000
1st Qu.	8.00000
Median	16.00000
Mean	15.62725
3rd Qu.	23.00000
Max.	31.00000
NA's	67009.00000



gender

Description

Individual's sex/gender.

Labels

- M = Male
- F = Female
- X = Unknown

Prompt

Enter the appropriate sex code abbreviation for the subject.

Name	Freq
male female	7861805 1787784
unknown	5405

race

Description

Individual's race/ethnicity.

Labels

see race_or_ethnicity.csv

Prompt

Enter the appropriate race code abbreviation for the subject. The criteria for determining race categories are in the TABLES Section.

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.

Name	Freq
white	3504045
hispanic	3241557
black	2358065
other	164450
unknown	132547
american_indian	48894
other_asian	48371
filipino	47441
vietnamese	26384
chinese	26169
pacific_islander	19560
laotian	8834
japanese	5617
samoan	5562
asian_indian	5374
korean	5111
cambodian	2890
hawaiian	2789
guamanian	1334

deceased

${\bf Description}$

Whether individual is deceased, alive, or unknown.

Labels

- \bullet 0 not deceased or no information
- \bullet 1 deceased
- 2 deceased (homicide)

Name	Freq
not_deceased_or_no_information	9591048
NA	55556
deceased_homicide	5756
deceased	2634

${\tt arrest_record_id}$

${\bf Description}$

Last reported step in criminal justice process.

Labels

- 10 court disposition
- 11 prosecutor
- 12 police

Name	Freq
final_court_disposition prosecutor police	7426045 1701337 527612

$arrest_bypass$

Description

Internal processing code.

Name	Freq
no_bypass_needed	9618967
received_and_corrected	19947
NA	10975
$not_corrected$	5105

$arrest_converted_data$

Description

Internal processing code.

Name	Freq
not_converted	9515110
converted_1	139884

 $arrest_event_year$

Description

Arrest year.

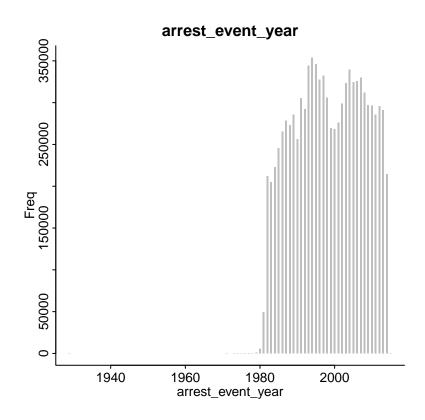
Prompt

Enter the month/day/year (MM-DD-YYYY) of: 1) the subject's arrest as it was submitted on the arrest fingerprint submission, 2) the date the complaint was sought or filed, or 3) the date of citation.

Notes

EG: MAYBE WE JUST PROVIDE YEAR HERE, BUT INTERNALLY USE THE FULL DATE AS DAY 0 IN THE PROCESS

Name	Value
Min.	1929.000
1st Qu.	1991.000
Median	1998.000
Mean	1998.368
3rd Qu.	2006.000
Max.	2015.000



$arrest_event_month$

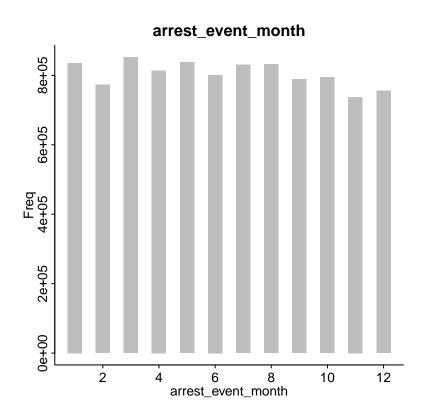
Description

Arrest month.

${\bf Prompt}$

Enter the month/day/year (MM-DD-YYYY) of: 1) the subject's arrest as it was submitted on the arrest fingerprint submission, 2) the date the complaint was sought or filed, or 3) the date of citation.

Name	Value
Min.	1.000000
1st Qu.	3.000000
Median	6.000000
Mean	6.411147
3rd Qu.	9.000000
Max.	12.000000



 $arrest_event_day$

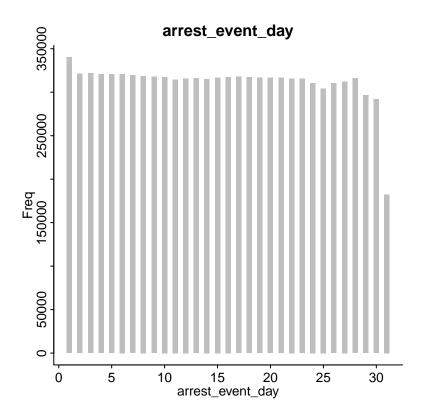
Description

Arrest day.

${\bf Prompt}$

Enter the month/day/year (MM-DD-YYYY) of: 1) the subject's arrest as it was submitted on the arrest fingerprint submission, 2) the date the complaint was sought or filed, or 3) the date of citation.

Name	Value
Min.	1.0000
1st Qu.	8.0000
Median	16.0000
Mean	15.6296
3rd Qu.	23.0000
Max.	31.0000



arresting_agency

Description

Arresting or booking agency.

Labels

- PD Police Department
- SO Sheriff's Office (Department)
- CHP California Highway Patrol
- MR Marshal
- INS United States Immigration and Naturalization Service
- FBI Federal Bureau of Investigation
- DA District Attorney
- CA City Attorney
- NTF Narcotic Task Force

Prompt

Enter the complete name of the agency that arrested, cited, or requested a complaint be filed against the subject. Abbreviates for Type of Agency are acceptable; however, the county or city name should be spelled out followed by SO, PD, etc.

Notes

may actually be booking and not arresting agency

Name	Freq
194200	829442
380100	411893
190000	405271
360000	334991
371100	271722
431300	235863
010900	226795
330000	210015
300000	197450
340000	186042
340400	164186
370000	144609
100500	143750
150000	129940
150200	125882
194100	120693
390500	90896
300100	70622
500200	69642
390000	68235
010000	67389
490001	63597

Name	Freq
100000 301900	58186 58059

 $local_file_number$

Description

Sealed.

$arrest_charge_type$

Description

Arrest offense level.

Labels

- F felony
- $\bullet~$ M misdemeanor
- I infraction
- \bullet X unknown

Name	Freq
felony	9653751
misdemeanor	1243

$multiple_arrest_charges$

Description

Number of arrest charges.

Labels

1-9 where 9 is 9+; prior to 1986 only used: 1 - one offense and 2 - more than one

Name	Freq
1	5711246
2	2269846
3	978276
4	390492
5	146942
6	61679
NA	55556
7	22838
8	10407
9	7712

$arrest_disposition_year$

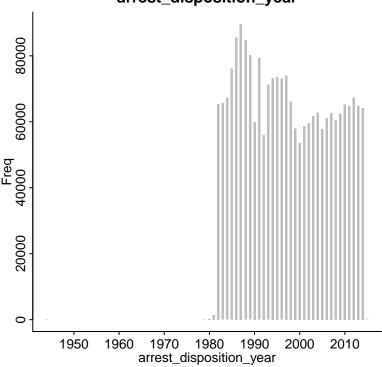
Description

Date of final disposition decided by arresting officer.

Summary

Name	Value
Min.	1944.00
1st Qu.	1989.00
Median	1997.00
Mean	1997.35
3rd Qu.	2006.00
Max.	2015.00
NA's	7426566.00

arrest_disposition_year

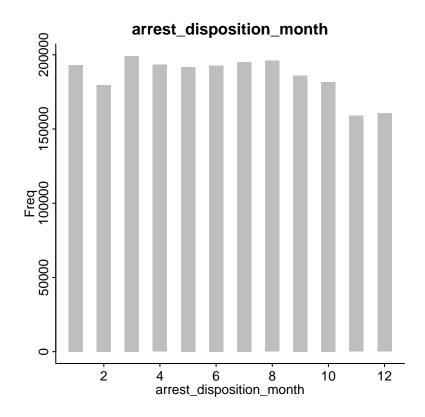


$arrest_disposition_month$

Description

Date of final disposition decided by arresting officer.

Name	Value
Min.	1.000000e+00
1st Qu.	3.000000e+00
Median	6.000000e+00
Mean	6.345454e + 00
3rd Qu.	9.000000e+00
Max.	1.200000e+01
NA's	7.426566e + 06



$arrest_disposition_day$

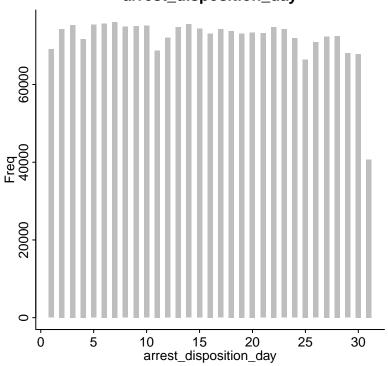
Description

Date of final disposition decided by arresting officer.

Summary

Name	Value
Min.	1.000000e+00
1st Qu.	8.000000e+00
Median	1.600000e+01
Mean	1.564913e+01
3rd Qu.	2.300000e+01
Max.	3.1000000e+01
NA's	7.426566e + 06

arrest_disposition_day



${\bf arrest_disposition_type}$

Description

Release or decline to prosecute rationale.

Labels

see dispositions.csv table

Notes

PROSECUTOR REJECTS is only the final disposition if the Court Record ID is blank EG: we should figure out a way to make this type of data easily understandable to users VD: I can just code them in the database, since there really aren't that many

Name	Freq
NA	7426045
11	720714
16	429302
6	196243
10	147449
15	112572
4	108734
13	100876
5	81336
12	54007
3	52722
19	50161
1	47626
18	37673
17	23611
7	22246
2	18701
20	13398
14	11578

$arrest_multiple_dispositions$

${\bf Description}$

Number of arrest dispositions.

Labels

- 0 subject did not receive all final dispositions at arrest level
- 1-9 where 9 is 9+;

prior to 1986 only used: 1 - one offense and 2 - more than one

Name	Freq
0	7549331
1	2016556
NA	55556
2	32456
3	1075
4	20

$prior_record_code$

Description

Number of prior prison commitments.

Labels

- 0 no prior record
- \bullet 1 misc prior record
- 7 one prior prison commitment
- ullet 8 two prior prison commitments
- ullet 9 three or more prior prison commitments
- $\bullet\,$ blank pre-1982 means lengthy record

Notes

EG: we should fix the discrepancy between documentation labels and data values VD: not sure what you mean, they seem to be the same as in the OBTS manual

Name	Freq
misc_prior_record	4548209
length_record	2305492
no_prior_record	2087482
one_prior_commitment	425620
two_prior_commitments	164598
$three_or_more_prior_commitments$	123593

$arrest_offense$

Description

Offense for which individual was arrested.

${\bf Labels}$

see cjis_codeTable.csv

Name	Freq
22004	892330
35172	722847
35152	584825
38028	519683
38025	389090
13094	367757
12004	314317
24054	209856
23097	193583
35186	183162
35127	166715
35138	160973
28025	153981
35188	143985
16001	143112
35154	122654
35254	113757
35137	108170
52096	102668
24053	101858
28012	92609
35215	92047
35276	88215
28011	86967
• • •	

$arrest_qualifier$

Description

Arrest offense qualifier.

Labels

SEE OBTS MANUAL APPX F

$arrest_summary_code$

Description

BCS summary code for arrest offense.

Labels

see bcs_offense_codes_variables.csv

Freq
2000302
1328695
1176712
1136148
1084953
419789
416760
408080
351713
345788
277054
205080
87613
81499
71178
57398
54755
50104
23614
21942
20654
19537
7487
6798
1341

 $arrest_edit_error_code$

Description

Internal processing code.

${\bf court_record_id}$

${\bf Description}$

Court of final disposition.

Labels

- $\bullet~$ 20 - final disposition in superior court
- ullet 21 final disposition in lower court

Name	Freq
superior_court	4808342
lower_court	2965568
NA	1881084

$court_bypass$

Description

Internal processing code.

Name	Freq
no_bypass_needed	9184652
$not_corrected$	275274
NA	110311
$corrections_sent$	84757

${\bf court_converted_data}$

Description

Internal processing code.

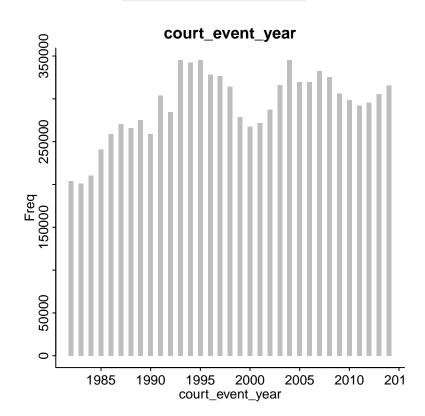
Name	Freq
NA	9599438
1	55556

court_event_year

Description

Last date of court proceedings.

Name	Value
Min.	1982.000
1st Qu.	1991.000
Median	1999.000
Mean	1998.767
3rd Qu.	2007.000
Max.	2014.000

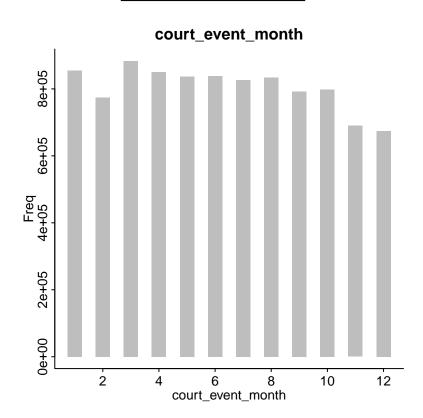


$court_event_month$

Description

Last date of court proceedings.

Name	Value
Min.	1.00000
1st Qu.	3.00000
Median	6.00000
Mean	6.31144
3rd Qu.	9.00000
Max.	12.00000



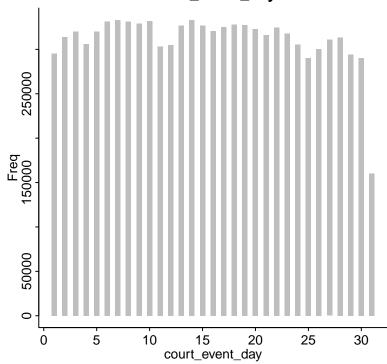
${\bf court_event_day}$

Description

Last date of court proceedings.

Name	Value
Min.	1.00000
1st Qu.	8.00000
Median	16.00000
Mean	15.62613
3rd Qu.	23.00000
Max.	31.00000





$court_judicial_district$

Description

Court judicial district.

Labels

see judicial_districts.csv

Name	Freq
19****	785494
19100J	304714
19463J	257809
37100J	211884
34470J	207425
43100J	190563
38****	177816
37****	167027
15420J	163929
10440J	150352
34100J	149181
38100J	144353
01100J	141132
01****	133451
43470J	129684
33460J	124430
30100J	119732
39460J	117005
30470J	115486
50100J	113418
30420J	111595
36470J	107659
49460J	101303
01440J	100452
• • •	• • •

$court_number$

Description

Sealed.

${\bf court_disposition_type}$

${\bf Description}$

Outcome of court trial.

Labels

- 21 convicted
- \bullet 22 acquitted
- 23 not guilty, insane
- \bullet 24 certified to juvenile court
- \bullet 25 defendant deceased
- $\bullet~26$ dismissed
- \bullet 27 diversion dismissed (begins 1988)

Name	Freq
convicted	6416370
NA	1881084
dismissed	1172645
diversion_dismissed	155270
acquited	24680
certified_to_juvenile_court	2302
defendant_deceased	1581
not_guilty_insane	1062

${\bf court_multiple_dispositions}$

${\bf Description}$

Number of court dispositions.

Labels

Pre 1986:

- 1 one disposition
- \bullet 2 more than one

post 1985:

- 1 only one
- 2 8 two to eight
- 9 9+ dispositions

Name	Freq
1	3881016
2	3387667
NA	1921751
3	425894
4	35492
5	2866
6	278
7	29
9	1

$court_proceeding_type$

Description

Court proceeding type.

Name	Freq
misdemeanor_complaint_filed	2863867
felony_complaint_filed	2801467
NA	1887273
information	1136298
certification	948677
filed_as_infraction	10257
indictment	7155

$sentence_type$

${\bf Description}$

Court sentence type.

Labels

- \bullet 01 death
- 02 prison
- 03 CYA
- 04 CRC
- 05 MDSO
- \bullet 06 probation and jail
- 07 probation
- 08 probation
- 09 fine
- 10 other: sentence suspended; work program; volunteer program; no sentence
- 11 prison term suspended (post 1984)

Name	Freq
1	3325469
NA	3308608
4	1984192
5	661199
3	253836
2	89892
6	31798

sentence

Description

Court sentence.

Name	Freq
probation_and_jail	3889758
NA	3238624
prison	1175414
probation	819844
jail	367954
other	69925
fine	67473
CRC	20534
CYA	3710
death	883
suspended	726
MSDO	149

${\bf court_charge_type}$

Description

Offense level of conviction crime.

Labels

- F felony
- M misdemeanor
- I infraction
- N place holder codes
- X unknown level of arrest

Name	Freq
felony	5249184
misdemeanor	2233766
NA	1876011
placeholder	285871
infraction	10162

$court_multiple_disposition_charges$

Description

Number of disposition charges.

Labels

Pre 1986:

- \bullet 1 one offense
- 2 more than one

post 1985:

- 1 only one
- 2 8 two to eight
- 9 9+ offenses

Notes

EG: is this always the number of conviction charges, or are there charges associated with non-conviction dispositions? Ask Alyson?

Name	Freq
1	3580934
2	2620343
NA	1921751
3	973389
4	358149
5	123652
6	45780
7	17987
8	7357
9	5652

${\bf court_disposition_offense}$

Description

Conviction offense.

Labels

see cjis_codeTable.csv

Notes

EG: todo

Name	Freq
NA	1890900
35172	544530
35152	417506
66085	285592
22004	247954
38028	195049
13094	170191
24054	149101
22022	138002
13045	137124
23097	129421
35138	120468
38064	116455
35353	97894
28025	96752
35154	96101
35127	91976
12004	90088
35186	85894
22021	77740
35212	76463
23132	72771
23099	71337
35276	68789
• • •	• • •

$court_qualifier$

Description

Conviction offense qualifiers.

Labels

see qualifier $_$ codes.csv

Notes

EG: todo

Name	Freq		
	9340440		
AB	71533		
KA	38453		
XX	35277		
MJ	30472		
ML	17468		
MC	17293		
ME	13615		
MB	10812		
MD	9223		
MM	9222		
MW	7034		
MK	6437		
NS	6131		
ON	2266		
MLMJ	1888		
MJML	1812		
MT	1766		
XXMC	1453		
MF	1163		
AE	1117		
ABME	997		
MA	936		
MU	887		

${\bf court_summary_code}$

Description

Conviction offense summary code.

Labels

see bcs_offense_codes_variables.csv

Name	Freq
NA	1890901
12	838025
14	831928
9	783351
6	722912
30	550353
8	541489
60	431875
31	310474
36	291119
25	226112
10	218266
19	207727
13	189618
11	180195
5	177660
51	130717
32	90625
50	89011
20	82303
46	80839
47	65756
34	54448
16	53885

 ${\bf court_edit_error_code}$

Description

Internal processing code.

Data Cleaning

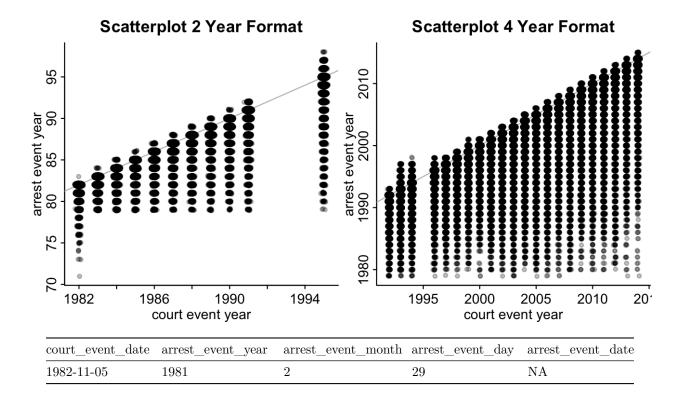
Court Event Date

Most of the single-digit birth are from prior to 1992, when years were stored as only 2 digits. For reasons unknown, 1995 is also in a 2- digit year format. The court event year match that of the corresponding file, so that they range from 82 to 2014, and the missing digits can be safely added. 4 are unable to be parsed into a correct date:

court_event_year	court_event_month	court_event_day	court_event_date
1982	2	29	NA
1982	2	29	NA
1982	11	31	NA
1982	11	31	NA

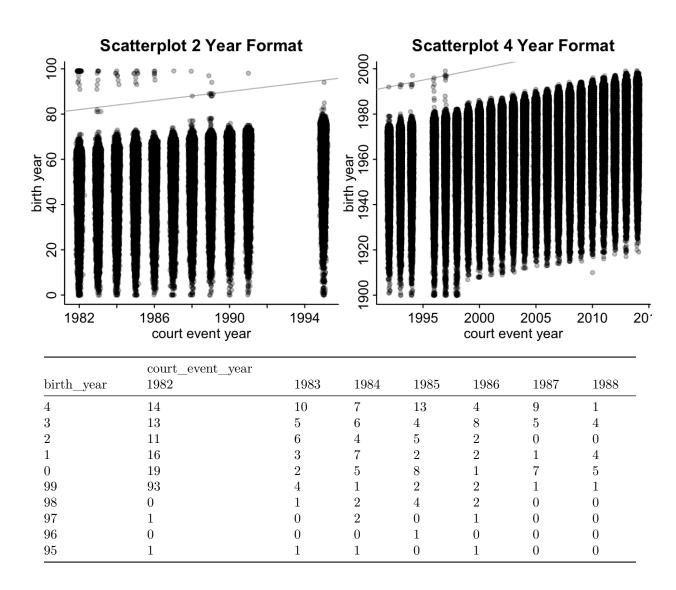
Arrest Event Date

Arrest event codings appear to be more inconsistent. For all but 1982, arrest years go back no further than 1979. 1 record lie outside of this range. Nevertheless, it appears as if the two-year coded years can safely be converted to four-year codes. After doing so, 1 record fail to parse.



Birth Date

The coding for birth years has numerous anomalies. 1982 has an abnormally large number of birth years equal to 99, while other years have clusters at 0.



In addition, there are a number of records where the birth-year is extraordinarily close to the court-event-year, including in the near-future.

birth_year	court_event_year 1983	1988	1989	1992	1993	1994	1995	1996	1997
81	3	0	0	0	0	0	0	0	0
82	1	0	0	0	0	0	0	0	0
88	0	1	3	0	0	0	0	0	0
89	0	0	5	0	0	0	0	0	0
94	0	0	0	0	0	0	1	0	0
1992	0	0	0	1	1	1	0	0	0
1993	0	0	0	0	2	3	0	1	0
1995	0	0	0	0	0	0	0	1	1
1996	0	0	0	0	0	0	0	1	1
1997	0	0	0	0	0	1	0	0	4
1999	0	0	0	0	0	0	0	0	1

To impute the century for the birth-year, we follow the rule:

- 1. If the listed 2-digit birth year is greater than 3 years from the court-event-year minus 1900, then we assume that it is in the 1800s.
- 2. Otherwise, we assume that it is in the 1900s.

This scheme allows the birth years that are close to the court years to persist, as that trend exists in the data sets for 4-digit years. These records bear further investigation.

After making this correction, a number of birth dates still cannot be successfully parsed, largely owing to a number of records that have either the birth month or birth day recorded as 0 (66931 rows). The remaining rows all seem to be coded as 99/99/99 and are in the court year 1982, indicating a particular kind of not-available.

court_event_date	birth_year	birth_month	birth_day
1982-02-24	1899	99	99
1982-01-28	1899	99	99
1982-12-02	1899	99	99
1982-12-02	1899	99	99

Arrest Disposition Date

The disposition dates can all safely be assumed to be in the same century as the court event, as it seems rare that the disposition date happens significantly after. Excluding those that are naturally missing, 2 dates cannot be parsed.

