OBTS Data Inventory

Introduction

OBTS Main Database

Words words words

Sample Rows

id	cii_record_ty...	cii_number	pdr_record_id	name	birth_year	...
1	OBN	SEALED	unknown	SEALED ...	1993	...
2	automated	SEALED	current_offen...	SEALED ...	1932	...
3	automated	SEALED	current_offen...	SEALED ...	1933	...
4	automated	SEALED	current_offen...	SEALED ...	1931	...
5	automated	SEALED	current_offen...	SEALED ...	1932	...
...	...	...	...	...	...	...

Variable Summary

name	type	value	description
id	numeric	1-9654994	
cii_record_type	factor	automat.../hybrid/m...	
cii_number	character	pii	
pdr_record_id	factor	current.../new_off....	Personal 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/decease......	
arrest_record_id	factor	final_c.../police/p...	
arrest_bypass	factor	no_bypa.../not_cor....	
arrest_converted_data	factor	convert.../not_con....	converted from old format
arrest_event_year	integer	1929-2015	
arrest_event_month	integer	1-12	
arrest_event_day	integer	1-31	
arresting_agency	factor	010000/010001/01010...	
local_file_number	character	pii	
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	
arrest_disposition_month	integer	1-12	
arrest_disposition_day	integer	1-31	
arrest_disposition_type	factor	1/2/3/4/5/6/7/10/11...	
arrest_multiple_dispositions	factor	0/1/2/3/4	
prior_record_code	factor	length_.../misc_pr....	

name	type	value	description
arrest_offense	factor	1004/1005/3005/9004...	
arrest_qualifier	character	.../AQ	
arrest_summary_code	factor	1/2/3/4/5/6/7/8/9/1...	
arrest_edit_error_code	character	/C /B /C...	
court_record_id	factor	lower_c.../superio....	
court_bypass	factor	correct.../no_bypa....	
court_converted_data	factor	1	
court_event_year	integer	1982-2014	
court_event_month	integer	1-12	
court_event_day	integer	1-31	
court_judicial_district	factor	01****/01100J/01120...	
court_number	character	pii	
court_disposition_type	factor	acquited/certifi......	
court_multiple_dispositions	factor	1/2/3/4/5/6/7/9	
court_proceeding_type	factor	certifi.../felony_....	
sentence_type	factor	1/2/3/4/5/6	
sentence	factor	CRC/CYA/death/fine/...	
court_charge_type	factor	felony/infract.../m...	
court_multiple_disposition_charges	factor	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

cii_record_type

Summary

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

cii_number

pdr_record_id

Summary

Name	Freq
new_offender	7238025
current_offender	1701846
unknown	715123

name

birth_year

Summary

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

birth_year
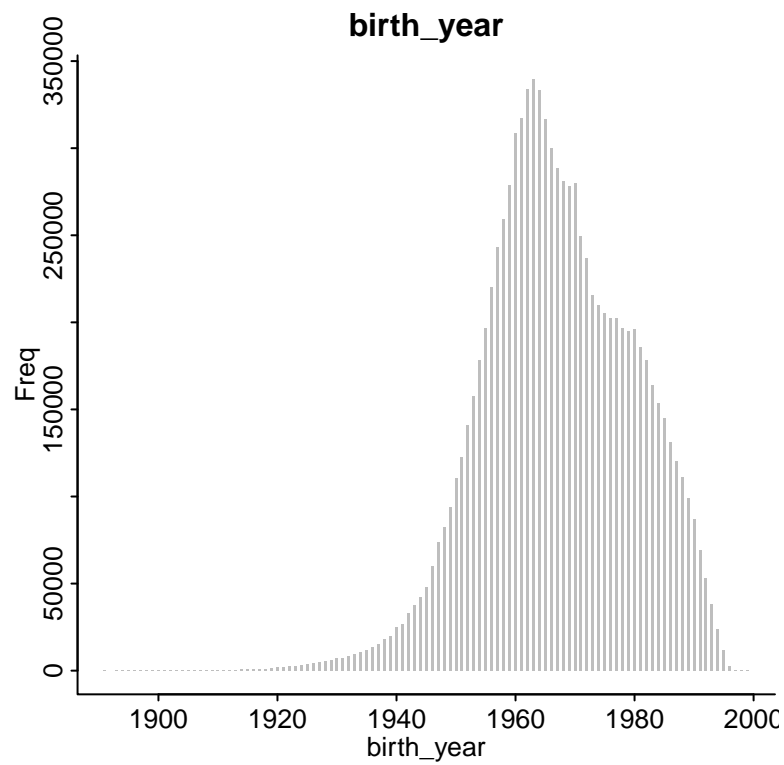
birth_month

Summary

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_month
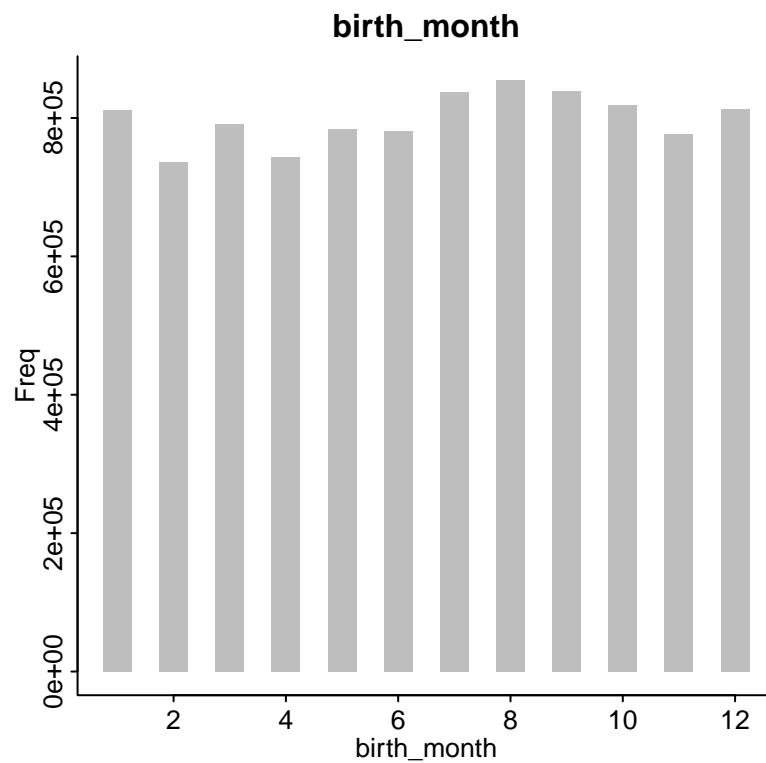
birth_day

Summary

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

birth_day
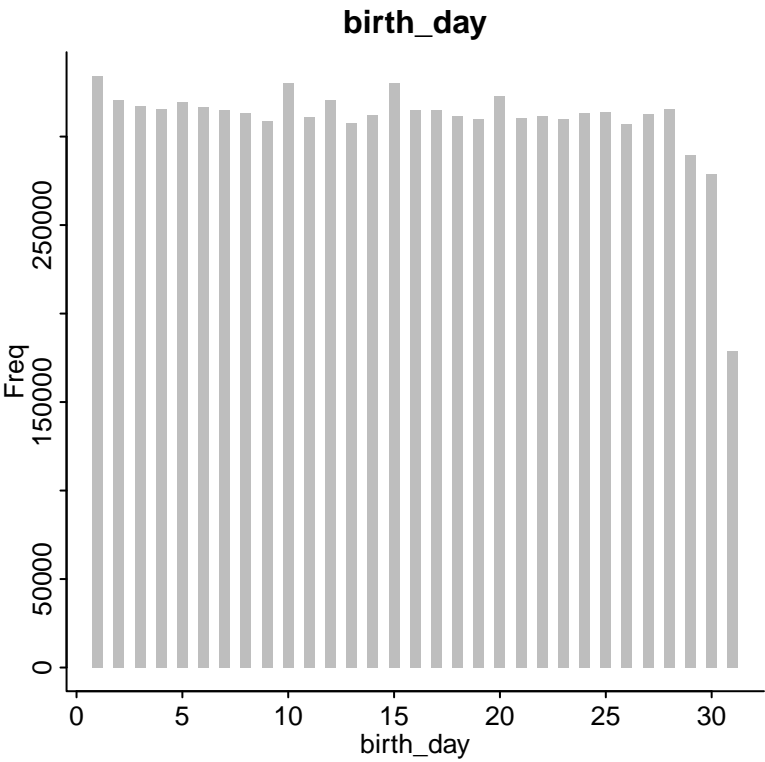
gender

Summary

Name	Freq
male	7861805
female	1787784
unknown	5405

race

Summary

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

Summary

Name	Freq
not_deceased_or_no_information	9591048
NA	55556
deceased_homicide	5756
deceased	2634

arrest_record_id

Summary

Name	Freq
final_court_disposition	7426045
prosecutor	1701337
police	527612

arrest_bypass

Summary

Name	Freq
no_bypass_needed	9618967
received_and_corrected	19947
NA	10975
not_corrected	5105

arrest_converted_data

Summary

Name	Freq
not_converted	9515110
converted_1	139884

arrest__event__year

Summary

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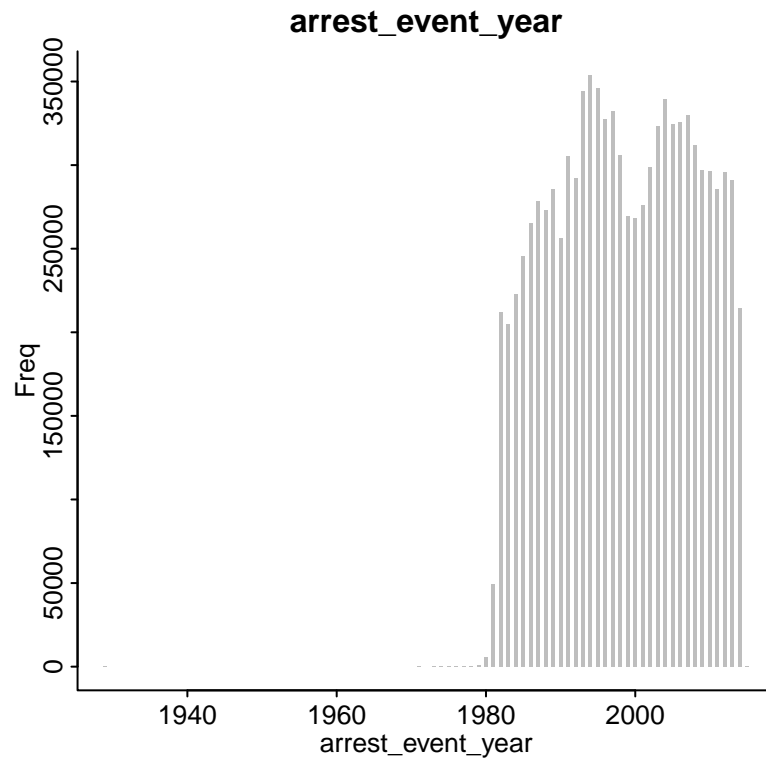
arrest_event_year

arrest__event__month

Summary

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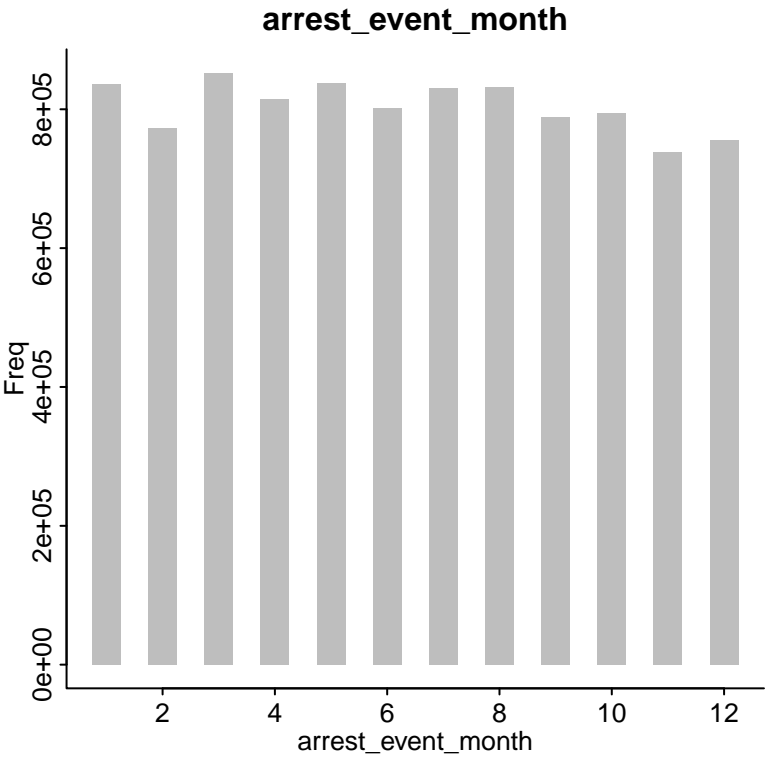
arrest_event_month

arrest__event__day

Summary

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

arrest_event_day
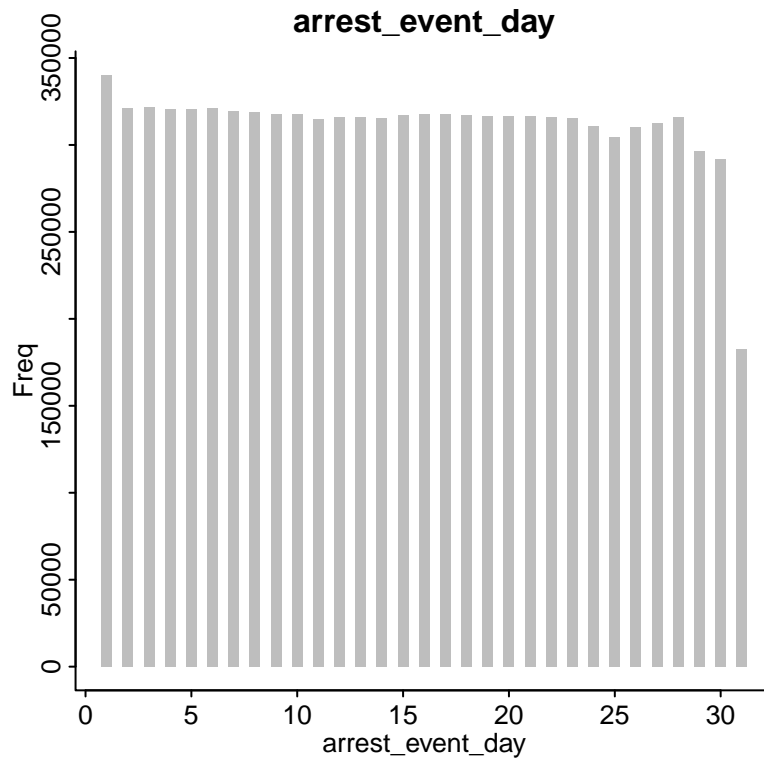
arresting_agency

Summary

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
100000	58186
301900	58059
. . .	. . .

local_file_number

arrest_charge_type

Summary

Name	Freq
felony	9653751
misdemeanor	1243

multiple_arrest_charges

Summary

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

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
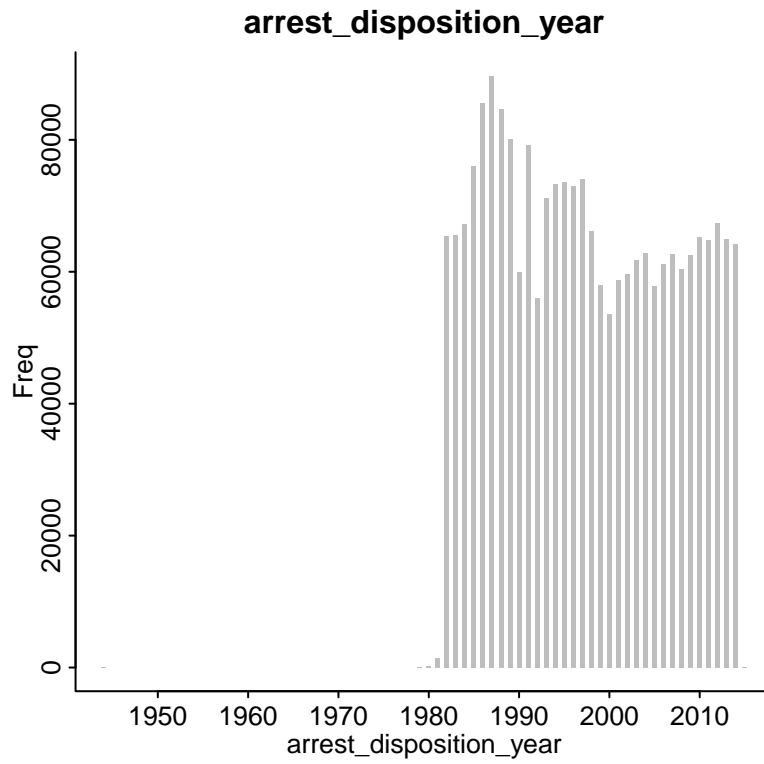
arrest__disposition__month

Summary

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_month
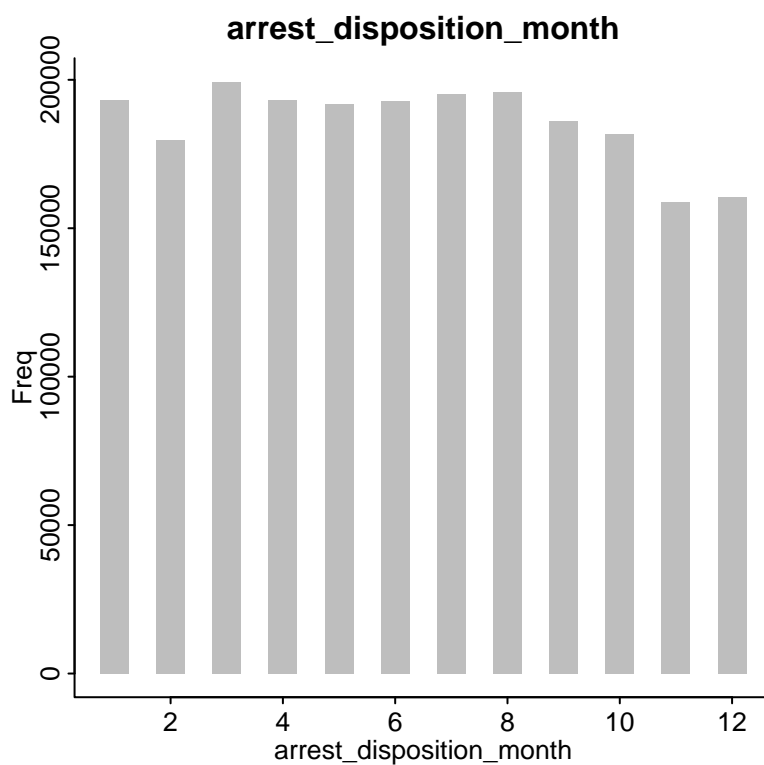
arrest__disposition__day

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.	3.100000e+01
NA's	7.426566e+06

arrest_disposition_day
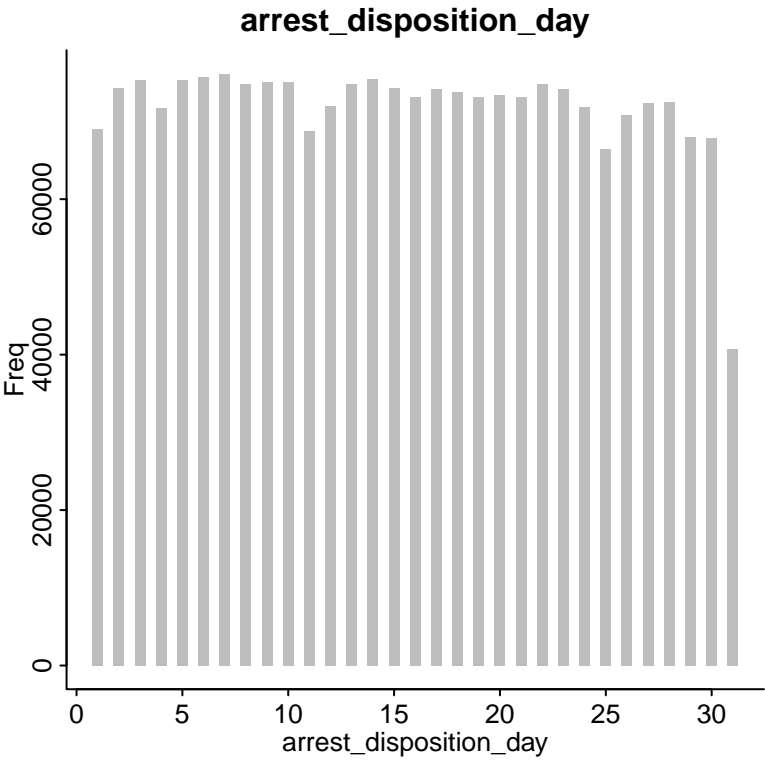
arrest__disposition__type

Summary

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

Summary

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

prior_record_code

Summary

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

Summary

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

arrest_summary_code

Summary

Name	Freq
6	2000302
12	1328695
14	1176712
8	1136148
9	1084953
10	419789
5	416760
25	408080
19	351713
13	345788
11	277054
20	205080
18	87613
16	81499
4	71178
15	57398
1	54755
7	50104
21	23614
24	21942
17	20654
22	19537
3	7487
23	6798
2	1341

arrest__edit__error__code

court_record_id

Summary

Name	Freq
superior_court	4808342
lower_court	2965568
NA	1881084

court_bypass

Summary

Name	Freq
no_bypass_needed	9184652
not_corrected	275274
NA	110311
corrections_sent	84757

court_converted_data

Summary

Name	Freq
NA	9599438
1	55556

court__event__year

Summary

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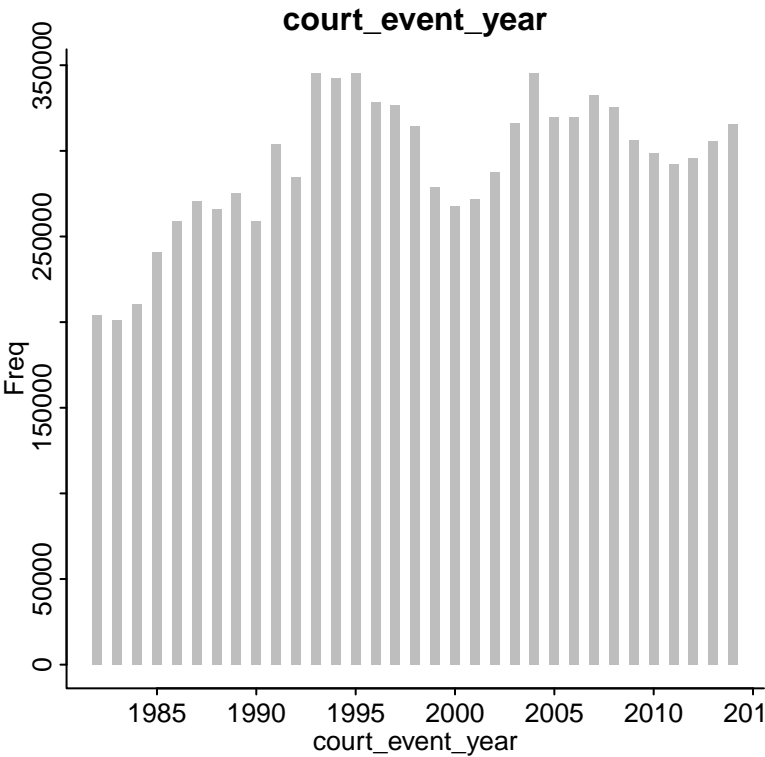
court_event_year

court_event_month

Summary

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

court_event_month
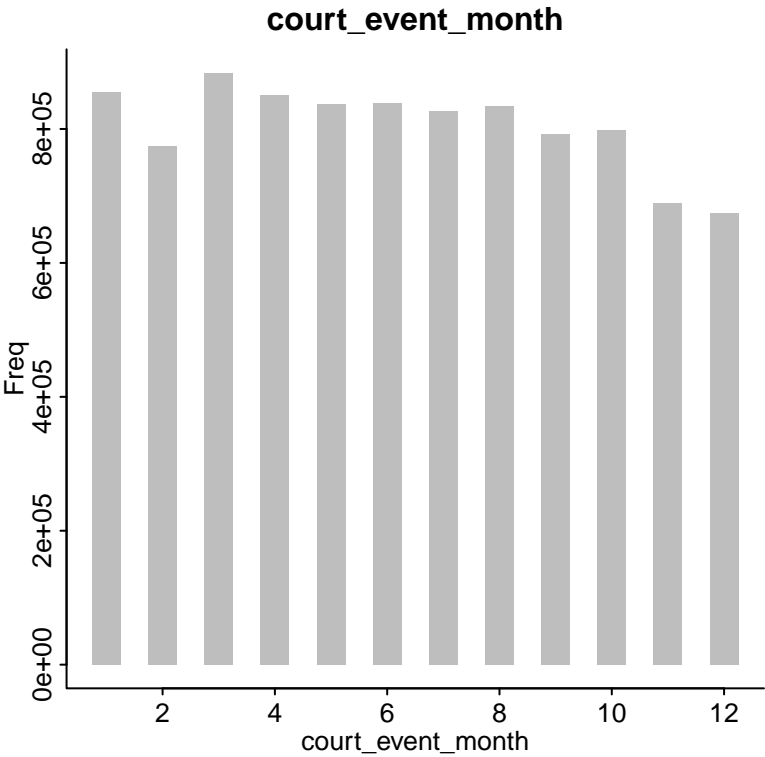
court_event_day

Summary

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

court_event_day
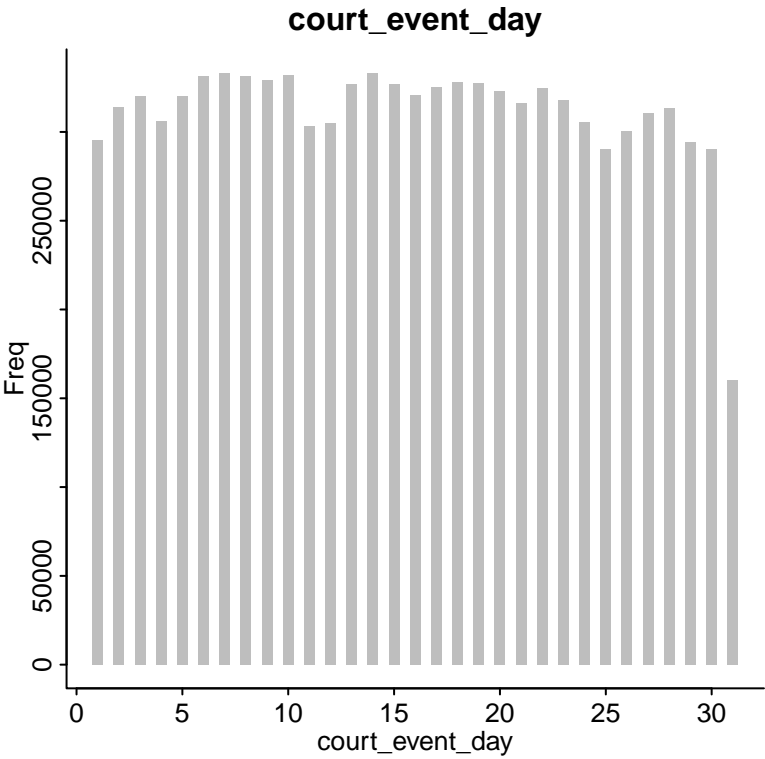
court_judicial_district

Summary

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

court_disposition_type

Summary

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

court_multiple_dispositions

Summary

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

court_proceeding_type

Summary

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

sentence_type

Summary

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

sentence

Summary

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

court_charge_type

Summary

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

court__multiple__disposition__charges

Summary

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

court_disposition_offense

Summary

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

Summary

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
. . .	. . .

court_summary_code

Summary

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
. . .	. . .

**court_edit_error_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:

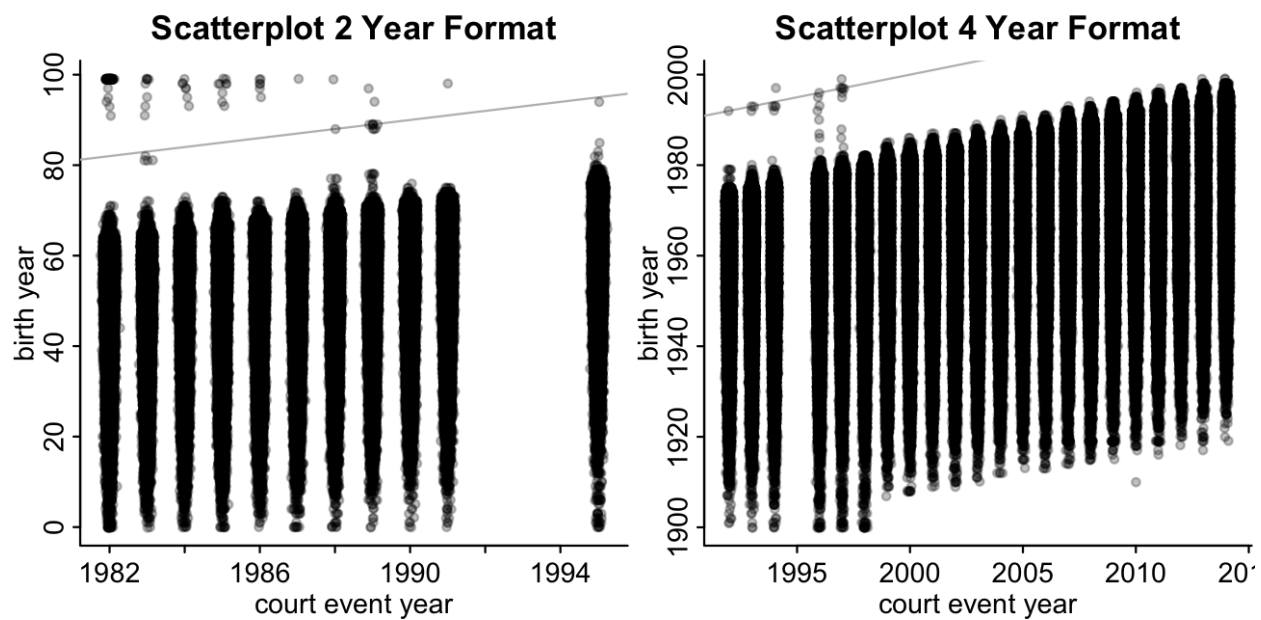| court_event_year	court_event_month	court_event_day	court_event_date
1982	2	29	NA
1982	2	29	NA
1982	11	31	NA
1982	11	31	NA

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.
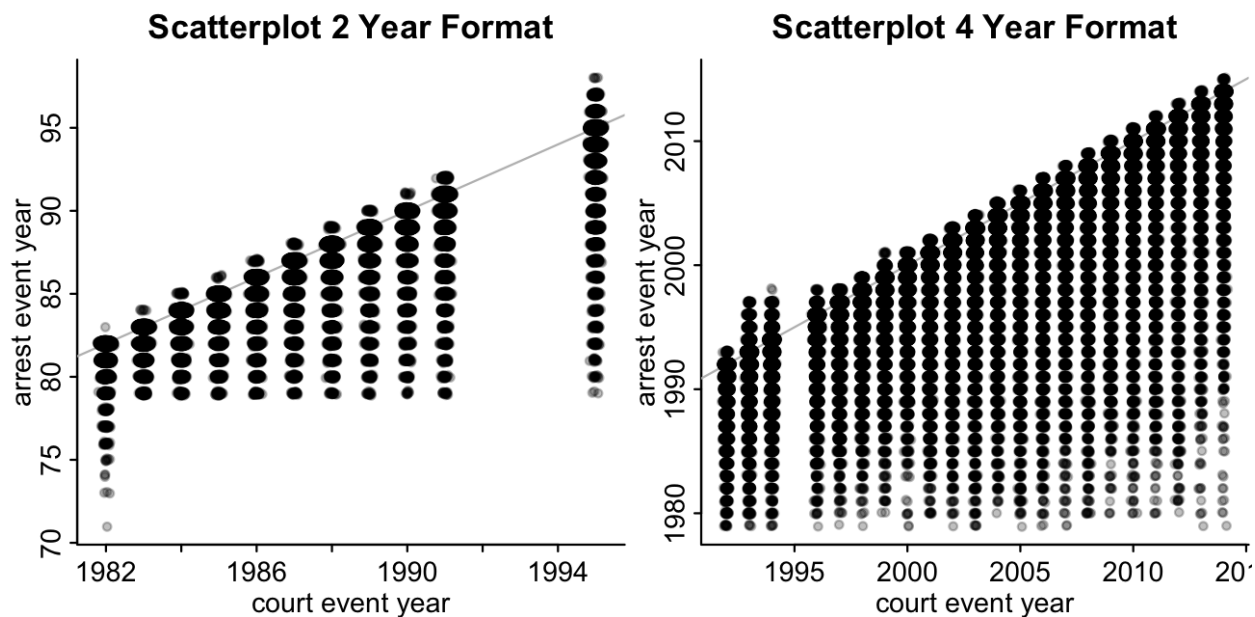
court_event_date	arrest_event_year	arrest_event_month	arrest_event_day	arrest_event_date
1982-11-05	1981	2	29	NA

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.

Scatterplot 2 Year Format

birth year vs court event year

Scatterplot 4 Year Format

birth year vs court event year

birth__year	court__event__year						
	1982	1983	1984	1985	1986	1987	1988
4	14	10	7	13	4	9	1
3	13	5	6	4	8	5	4
2	11	6	4	5	2	0	0
1	16	3	7	2	2	1	4
0	19	2	5	8	1	7	5
99	93	4	1	2	2	1	1
98	0	1	2	4	2	0	0
97	1	0	2	0	1	0	0
96	0	0	0	1	0	0	0
95	1	1	1	0	1	0	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.
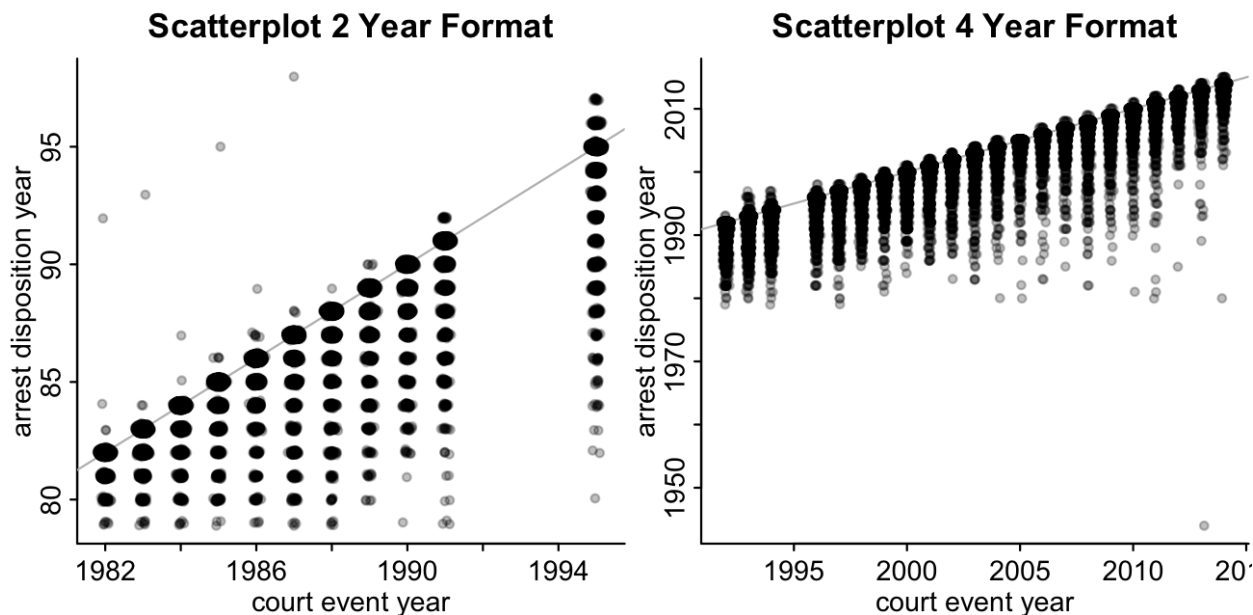
crt_evt_yr	crt_evt_mo	crt_evt_dy	arr_dsp_yr	arr_dsp_mo	arr_dsp_dy
1982	2	29	1982	2	29
1982	2	29	1982	2	29