Data Schemas

Officer Roster

- **birth_year**: Birth year of the officer.
- **appointed_month**: The month and year the officer was made an officer in YYYY-MM-DD format. The day is always the first day of the month.
- **officer_id**: Unique identifier for each officer.
- **officer_race**: Race of the officer.
- **officer_gender**: Sex of the officer.
- **spanish**: Does the officer speak Spanish or not?
- Uniquely identified by **officer_id**. The unit of observation is an officer.
- **Number of officers**: 33645

Creating the officer roster

> "The administrative data from the CPD used in this study span multiple datasets collected in collaboration with the Invisible Institute, Sam Stecklow, and Emma Herman over the course of three years (2016-2019). We obtained these records from the Chicago Police Department or Chicago Department of Human Resources via Freedom of Information Act (FOIA) or through court ordered releases stemming from requests made by Invisible Institute and Jaime Kalven. CPD provided the following data: rosters of all available current and past officers up to 2018, unit history data for individual officers from the 1930s to 2016, Tactical Response Reports from 2004 to 2018 (i.e. use of force reports), and arrest data with arresting officers and arrestee demographic information from 2001 to 2017. The Chicago Department of Human Resources provided data on officers' language skills up to 2019. We supplement our core data with data on 'Stop, Question and Frisk' (SQF) activity between 2012-2015, which was shared by the Lucy Parson's Lab. Finally, the Automated Daily Attendance and Assignment sheet data for each police district between 2012 and 2015 was obtained via a FOIA request to the CPD and shared by Rachel Ryley." pages 5-6 of Appendix Section S1.2 in Ba et al. 2021.

> "These data and others have been used to construct rich profiles of Chicago Police Officers. While no file contains a unique identifier (star numbers change over time, names are common, etc.), we constructed unique officer profiles through a successive merge process described here. Each file contains some identifying information such as of demographic data (birth year, race, gender) or other characteristics (name, start/badge number, appointed date, resignation date, current unit). We used these identifying characteristics to first de-duplicate officers within a file and to then merge to pre-existing officer data with inter-file unique identifiers. The merging process itself is an iterative-pairwise matching method, where the officers in each dataset are repeatedly merged on identifying characteristics and any successful 1-to-1 match in a round removes the matched officers from the next round of merging." page 6 of Appendix Section S1.2 in Ba et al. 2021.

Officer Race

> "We determine race/ethnicity of CPD officers based on demographic data obtained from the CPD through FOIA. The CPD usually classifies race/ethnicity in at most 7 mutually exclusive groups: White/Caucasian, White Hispanic, Black/African American, Black Hispanic, Asian/Pacific Islander, Native American/Native Alaskan, and unknown/missing. However, there are inconsistencies in how races and ethnicities are coded across files. For example, some files do not include 'Black Hispanic' as a racial category (very few officers are ever classified as Black Hispanic), and

some files contain outdated racial categories which we update to the best of our ability. For consistency, we classify 'Hispanic' and 'White Hispanic' as 'Hispanic'; 'Black' and 'Black Hispanic' (rare cases) as 'Black.' 'White' in our analysis refers to non-Hispanic White. If an officer has multiple races associated with them across different datasets, we aggregate by most common non-missing races." page 5 of Appendix Section S1.1 in Ba et al. 2021.

$numeric

col	min	first_q	med	mean	third_q	max	sd	iqr	mad	na	prcnt_na
birth_year	1916	1944	1956	1956.5	1971	1996	18.13	27	20.76	926	0.03

$factor

n	prcnt	name	value
20675	0.6145044	officer_race	officer_white
7634	0.2268985	officer_race	officer_black
4578	0.1360678	officer_race	officer_hisp
538	0.0159905	officer_race	officer_aapi
153	0.0045475	officer_race	NA
67	0.0019914	officer_race	officer_native
27869	0.8283252	officer_gender	MALE
5771	0.1715262	officer_gender	FEMALE
5	0.0001486	officer_gender	NA

$logical

col	mean	true	false	missing	prcnt_missing
spanish	0.0845296	2844	30801	0	0

$other

col	missing	nr_unique	prcnt_missing
appointed_month	70	476	0.0020805
officer_id	0	33645	0.0000000

Shift Assignments

- **officer_id**: Unique identifier for each officer.
- **month**: Month of the shift in YYYY-MM-DD format. The day is always the first day of the month.
- **rank**: Rank of the officer assigned to the shift.
- **unit**: Unit of the officer assigned to the shift.
- **date**: Date of the shift in YYYY-MM-DD format.
- **shift**: The shift the officer is assigned to.
- **start_time**: Hour start time of the shift in military time.
- **end_time**: Hour end time of the shift in military time.
- **start_datetime**: Combines **start_time** and **date** into a starting datetime for the shift in YYYY-MM-DD HH:MM:SS format. **Created by me.**
- **end_datetime**: Combines **end_time** and **date** into an ending datetime for the shift in YYYY-MM-DD HH:MM:SS format. **Created by me.**
- **weekday**: Day of the week of the shift.
- **beat_assigned**: The beat the officer is assigned to.
- **appointed_month**: The month and year the officer was made an officer.
- **months_from_start**: The number of months between the officer's appointment date and their shift date.
- **months_from_start_sq**: The number of months between the officer's appointment date and their shift date, squared.
- **duration**: Length of the shift in hours.
- **shift_id**: Unique identifier for each shift assignment. **Created by me**.

- Uniquely identified by **officer_id** and **date** or **shift_id**. The unit of observation is a specific shift for a specific officer.
- **Number of shift assignments**: 3519518

$factor

n	prcnt	name	value
3050853	0.8668383	rank	POLICE OFFICER
281965	0.0801147	rank	SERGEANT
74751	0.0212390	rank	POLICE OFFICER / FLD TRNG OFFICER
53765	0.0152762	rank	NA
28077	0.0079775	rank	POLICE OFFICER (ASSIGNED AS DETECTIVE)
16467	0.0046788	rank	COMMANDER
4678	0.0013292	rank	LIEUTENANT
3918	0.0011132	rank	POLICE OFFICER (ASSIGNED AS SECURITY SPECIALIST)
2909	0.0008265	rank	POLICE OFFICER (PER ARBITRATION AWARD)
1080	0.0003069	rank	POLICE OFFICER (ASSIGNED AS CANINE HANDLER)
350	0.0000994	rank	DEPUTY CHIEF
275	0.0000781	rank	POLICE AGENT
218	0.0000619	rank	EXPLOSIVES TECHNICIAN I
84	0.0000239	rank	POLICE OFFICER (ASGND AS MARINE OFFICER)
77	0.0000219	rank	POLICE OFFICER (ASSIGNED AS EVIDENCE TECHNICIAN)
47	0.0000134	rank	CHIEF
2	0.0000006	rank	SERGEANT (PER ARBTRN AGR)
1	0.0000003	rank	POLICE LEGAL OFFICER I
1	0.0000003	rank	POLICE LEGAL OFFICER II
210250	0.0597383	unit	7
200102	0.0568549	unit	8
197317	0.0560636	unit	11
194330	0.0552149	unit	2
191057	0.0542850	unit	12
185394	0.0526760	unit	9
184942	0.0525475	unit	19
182661	0.0518994	unit	3
179211	0.0509192	unit	6
176369	0.0501117	unit	4
169000	0.0480179	unit	25
149365	0.0424390	unit	10
144866	0.0411607	unit	24
139062	0.0395117	unit	15
137113	0.0389579	unit	18
132216	0.0375665	unit	22
131301	0.0373065	unit	5
124804	0.0354605	unit	14
117569	0.0334049	unit	1
114630	0.0325698	unit	20
114422	0.0325107	unit	17
110579	0.0314188	unit	16
23516	0.0066816	unit	13
4846	0.0013769	unit	23
4596	0.0013059	unit	21
1331644	0.3783598	shift	3
1204317	0.3421824	shift	2
983557	0.2794579	shift	1
547248	0.1554895	weekday	Tue
546066	0.1551536	weekday	Wed
538404	0.1529766	weekday	Thu
515802	0.1465547	weekday	Fri
508085	0.1443621	weekday	Mon
436811	0.1241110	weekday	Sat
427102	0.1213524	weekday	Sun

col	min	first_q	med	mean	third_q	max	sd	iqr	mad	na	prcnt_na
start_time	0.0	7	14	13.87	20.5	23.5	6.02	13.5	10.38	48210	0.01
end_time	0.5	16	23	22.80	29.5	46.0	6.10	13.5	10.38	48210	0.01
months_from_start	1.0	107	172	172.18	230.0	646.0	83.78	123.0	90.44	778	0.00
months_from_start_sq	1.0	11449	29584	36663.51	52900.0	417316.0	31136.20	41451.0	29330.28	778	0.00
duration	0.0	9	9	8.93	9.0	23.5	0.39	0.0	0.00	48210	0.01

$other

col	missing	nr_unique	prcnt_missing
officer_id	0	8502	0.0000000
month	0	49	0.0000000
date	0	1462	0.0000000
beat_assigned	0	6616	0.0000000
appointed_month	0	228	0.0000000
shift_id	0	3519518	0.0000000
start_datetime	48210	48749	0.0136979
end_datetime	48210	54031	0.0136979

Stops

- **stop_id**: Identifier for each stop.
- **time**: Time of the stop in YYYY-MM-DD HH:MM:SS format.
- **date**: Date of the stop in YYYY-MM-DD format.
- **district**: Police district where the stop took place.
- **po_first**: Was the focal officer the first to respond to the scene?
- **stop_type**: What was the type of the stop?
- **contact_type**: Collapsed version of **stop_type** (less categories).
- **civ.race**: Race of the civilian.
- **civ.gender**: Sex of the civilian.
- **civ.age**: Age of the civilian at the time of the stop.
- **lat**: Latitude of the stop.
- **lon**: Longitude of the stop.
- **officer_id**: Unique identifier for the officer.
- **month**: Month of the stop in YYYY-MM-DD format. The day is always the first day of the month.
- **civilian_race_short**: Collapsed version of **civ.race** (less categories).
- **hour**: Hour of the day when the stop took place rounded to the nearest hour in military time.
- **stop_officer_id**: Unique identifier for each entry. **Created by me**.
- The unit of analysis is a unique officer involved in a stop. Each row can be uniquely identified by **officer_id** and **stop_id** or by **stop_officer_id**.
- Each stop involves only one civilian, but they can involve multiple officers. It is possible multiple stops are all a part of one larger incident involving multiple civilians. This can be investigated by examining stops that took place in the same location at the same time involving the same officers.
- **Number of rows**: 1703158
- **Number of unique stops**: 946912

Stop Type vs. Contact Type

> "Stops for 'dispersal' and 'gang and narcotics-related loitering' are coded as loitering stops; those that are 'gang / narcotics related' are coded as drug stops; 'investigatory stops' and stops of 'suspicious persons' are coded as suspicious behavior; and stops under the 'Repeat Offender Geographic Urban Enforcement Strategy (ROGUES)' program are combined with the 'other' category." page 8 of Appendix Section S1.5 in Ba et al. 2021.

$numeric

col	min	first_q	med	mean	third_q	max	sd	iqr	mad	na	prcnt_na
civ.age	0	22	29	32.90	42	1815	13.82	20	13.34	19459	0.01
hour	0	9	16	13.61	19	23	7.24	10	7.41	0	0.00

$factor

n	prcnt	name	value
153554	0.0901584	district	7
116621	0.0684734	district	11
113954	0.0669075	district	9
109460	0.0642688	district	3
107179	0.0629296	district	8
105842	0.0621446	district	6
105072	0.0616925	district	10
97039	0.0569759	district	2
90303	0.0530209	district	4
79885	0.0469040	district	25
77759	0.0456558	district	15
73882	0.0433794	district	12
72292	0.0424459	district	5
61321	0.0360043	district	19
61013	0.0358235	district	22
52571	0.0308668	district	24
44496	0.0261256	district	18
41572	0.0244088	district	20
39832	0.0233871	district	16
37711	0.0221418	district	14
27783	0.0163126	district	17
22671	0.0133112	district	1
4844	0.0028441	district	13
4219	0.0024772	district	31
1455	0.0008543	district	23
828	0.0004862	district	21
488592	0.2868741	stop_type	Other
449868	0.2641376	stop_type	Traffic Related
326587	0.1917538	stop_type	Suspicious Person
238349	0.1399453	stop_type	Investigatory Stop
175939	0.1033016	stop_type	Gang / Narcotics Related
14370	0.0084373	stop_type	Dispersal
9314	0.0054687	stop_type	Gang and Narcotics-Related Loitering
139	0.0000816	stop_type	Repeat Offender Geog. Urban Enforcement Strategy (ROGUES)
564936	0.3316991	contact_type	suspicious
488731	0.2869558	contact_type	other
449868	0.2641376	contact_type	traffic
175939	0.1033016	contact_type	drug
23684	0.0139059	contact_type	loitering
1124118	0.6600198	civ.race	BLACK
346558	0.2034797	civ.race	HISPANIC
210545	0.1236204	civ.race	WHITE
19871	0.0116672	civ.race	ASIAN/PACIFIC ISLANDER
2066	0.0012130	civ.race	AMER IND/ALASKAN NATIVE
1322801	0.7766754	civ.gender	M
379401	0.2227632	civ.gender	F
953	0.0005595	civ.gender	X
3	0.0000018	civ.gender	NA
1124118	0.6600198	civilian_race_short	civilian_black
346558	0.2034797	civilian_race_short	civilian_hisp
210545	0.1236204	civilian_race_short	civilian_white
19871	0.0116672	civilian_race_short	civilian_aapi
2066	0.0012130	civilian_race_short	NA

$logical

col	mean	true	false	missing	prcnt_missing
po_first	0.5523228	940693	762465	0	0

$other

col	missing	nr_unique	prcnt_missing
stop_id	0	946912	0.0000000
time	0	519050	0.0000000
date	0	1461	0.0000000
lat	248657	23863	0.1459976
lon	248657	18467	0.1459976
officer_id	0	8937	0.0000000
month	0	48	0.0000000
stop_officer_id	0	1703158	0.0000000

Arrests

- **date**: Date of the arrest in YYYY-MM-DD format.
- **hour**: Hour of the day when the arrest took place rounded to the nearest hour in military time.
- **crime_code**: The suspected crime type causing the arrest.
- **statute_description**: More detailed categories describing what specific statute was suspected to have been violated.
- **lat**: Latitude of the arrest.
- **lon**: Longitude of the arrest.
- **district**: Police district where the arrest took place.
- **civ.race**: Race of the civilian.
- **civ.gender**: Sex of the civilian.
- **civ.age**: Age of the civilian at the time of the stop.
- **arrest_id**: Identifier for each arrest.
- **officer_id**: Unique identifier for each officer.
- **month**: Month of the arrest in YYYY-MM-DD format. The day is always the first day of the month.
- **civilian_race_short**: Collapsed version of **civ.race** (less categories).
- **arrest_officer_id**: Unique identifier for each entry. **Created by me**.
- The unit of analysis is a unique officer involved in an arrest. Each row can be uniquely identified by **officer_id** and **arrest_id** or by **arrest_officer_id**.
- Each arrest involves only one civilian, but they can involve multiple officers. It is possible multiple arrests are all a part of one larger incident involving multiple civilians. This can be investigated by examining arrests that took place in the same location at the same time involving the same officers.
- **Number of rows**: 321872
- **Number of unique arrests**: 164802

$numeric

col	min	first_q	med	mean	third_q	max	sd	iqr	mad	na	prcnt_na
hour	0	7	15	13.05	19	23	7.31	12	7.41	0	0
civ.age	18	23	29	32.55	41	94	11.92	18	11.86	22	0

$factor

n	prcnt	name	value
119271	0.3705541	crime_code	other
81042	0.2517833	crime_code	violent
57604	0.1789656	crime_code	property
36240	0.1125913	crime_code	drug
27715	0.0861057	crime_code	TRF
28523	0.0886160	district	11
25911	0.0805009	district	25
23109	0.0717956	district	8
22166	0.0688659	district	6
21130	0.0656472	district	7
20592	0.0639757	district	9
19144	0.0594771	district	4
19134	0.0594460	district	10
18995	0.0590141	district	15
17732	0.0550902	district	3
14253	0.0442816	district	5
11458	0.0355980	district	19
10463	0.0325067	district	18
10152	0.0315405	district	2
9786	0.0304034	district	12
8194	0.0254573	district	1
7622	0.0236802	district	14
7279	0.0226146	district	22
7143	0.0221921	district	24
7014	0.0217913	district	17
5946	0.0184732	district	16
4510	0.0140118	district	20
1274	0.0039581	district	13
193	0.0005996	district	23
149	0.0004629	district	21
214048	0.6650097	civ.race	BLACK
70132	0.2178879	civ.race	HISPANIC
34528	0.1072725	civ.race	WHITE
2167	0.0067325	civ.race	ASIAN/PACIFIC ISLANDER
670	0.0020816	civ.race	NA
327	0.0010159	civ.race	NATIVE AMERICAN/ALASKAN NATIVE
262116	0.8143486	civ.gender	MALE
59722	0.1855458	civ.gender	FEMALE
34	0.0001056	civ.gender	NA
214048	0.6650097	civilian_race_short	civilian_black
70132	0.2178879	civilian_race_short	civilian_hisp
34528	0.1072725	civilian_race_short	civilian_white
2167	0.0067325	civilian_race_short	civilian_aapi
670	0.0020816	civilian_race_short	NA
327	0.0010159	civilian_race_short	civilian_native

$other

col	missing	nr_unique	prcnt_missing
date	0	1462	0.0000000
statute_description	2	1621	0.0000062
lat	59621	18324	0.1852320
lon	59621	14987	0.1852320
arrest_id	0	164802	0.0000000
officer_id	0	8369	0.0000000
month	0	49	0.0000000
arrest_officer_id	0	321872	0.0000000

Uses of Force

- **date**: Date of the use of force in YYYY-MM-DD format.
- **time**: Time of the use of force in YYYY-MM-DD HH:MM:SS format.
- **district**: Police district where the use of force took place.
- **lat**: Latitude of the use of force.
- **lon**: Longitude of the force.
- **civ.race**: Race of the civilian.
- **civ.gender**: Sex of the civilian.
- **civ.age**: Age of the civilian at the time of the stop.
- **civ.injured**: Was the civilian injured?
- **force_id**: Unique identifier for each use of force incident.
- **officer_id**: Unique identifier for each officer.
- **month**: Month of the arrest in YYYY-MM-DD format. The day is always the first day of the month.
- **civilian_race_short**: Collapsed version of **civ.race** (less categories).
- **hour**: Hour of the day when the use of force took place rounded to the nearest hour in military time.
- The unit of analysis is a use of force incident. Only one police officer is listed for each use of force incident. Each row can be uniquely identified by **force_id**.
- **Number of uses of force**: 9293

$numeric

col	min	first_q	med	mean	third_q	max	sd	iqr	mad	na	prcnt_na
civ.age	18	23	28	30.52	36	81	10.33	13	8.9	31	0
hour	0	4	15	12.76	19	23	7.77	15	8.9	0	0

$factor

n	prcnt	name	value
935	0.1006134	district	11
782	0.0841494	district	6
780	0.0839341	district	7
647	0.0696223	district	15
627	0.0674701	district	4
559	0.0601528	district	8
525	0.0564941	district	25
465	0.0500377	district	3
443	0.0476703	district	5
442	0.0475627	district	10
431	0.0463790	district	19
427	0.0459486	district	9
255	0.0274400	district	24
252	0.0271172	district	12
249	0.0267944	district	18
248	0.0266868	district	14
216	0.0232433	district	22
195	0.0209835	district	16
183	0.0196922	district	20
178	0.0191542	district	17
173	0.0186162	district	1
161	0.0173249	district	2
50	0.0053804	district	13
34	0.0036587	district	NA
24	0.0025826	district	31
10	0.0010761	district	23
1	0.0001076	district	21
1	0.0001076	district	41
6774	0.7289358	civ.race	BLACK
1387	0.1492521	civ.race	HISPANIC
949	0.1021199	civ.race	WHITE
106	0.0114064	civ.race	NA
62	0.0066717	civ.race	ASIAN/PACIFIC ISLANDER
15	0.0016141	civ.race	NATIVE AMERICAN/ALASKAN NATIVE
7539	0.8112558	civ.gender	MALE
1746	0.1878834	civ.gender	FEMALE
8	0.0008609	civ.gender	NA
6774	0.7289358	civilian_race_short	civilian_black
1387	0.1492521	civilian_race_short	civilian_hisp
949	0.1021199	civilian_race_short	civilian_white
106	0.0114064	civilian_race_short	NA
62	0.0066717	civilian_race_short	civilian_aapi
15	0.0016141	civilian_race_short	civilian_native

$logical

col	mean	true	false	missing	prcnt_missing
civ.injured	0.2704186	2513	6780	0	0

$other

col	missing	nr_unique	prcnt_missing
date	0	1434	0.0000000
time	0	6898	0.0000000
lat	1754	3629	0.1887442
lon	1754	3475	0.1887442
force_id	0	9293	0.0000000
officer_id	0	3843	0.0000000
month	0	49	0.0000000