

# Officer Codebook

## 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