Codebooks for the Core, Cleaned Data Sets Used for "Do Women Officers Police Differently? Evidence from Traffic Stops"

Kelsey Shoub, Katelyn Stauffer, Miyeon Song
University of South Carolina

Overview

Three data sets are used to conduct the analysis. The first data set is associated with the Charlotte (NC) Police Department analysis. The second and third data sets are used for the Florida Highway Patrol analysis. The first of these is the data set of individual traffic stops, and the second of these is aggregated by context and officer.

The data originally came from three sources. The citations for these are as follows:

- Baumgartner, F. R., K. Bell, L. Beyer, T. Boldrin, L. Doyle, L. Govan, J. Halpert, J. Hicks, K. Kyriakoudes, C. Lee, M. Leger, S. McAdon, S. Michalak, C. Murphy, E. Neal, O. O'Malley, E. Payne, A. Sapirstein, S. Stanley, and K. Thacker. 2020. "Intersectional Encounters: Representative Bureaucracy and the Routine Traffic Stop." *Policy Studies Journal*, Early View, 12 February 2020. Doi: 10.1111/PSJ.12382.
- City of Charlotte [North Carolina]. Officer Traffic Stops. Charlotte [North Carolina] Open Data Portal. https://data.charlottenc.gov/datasets/officer-traffic-stops.
- E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, and S. Goel. "A large-scale analysis of racial disparities in police stops across the United States". Nature Human Behaviour, Vol. 4, 2020.

Note that throughout each data set missing information is indicated by NA values.

Charlotte (NC) Police Department Data Set

This data set (NorthCarolina.RData) is an aggregated version of the Charlotte (NC) Police Department data set, where each observation is a traffic stop. It contains 218,158 observations and has 32 variables. The updated data comes from the Charlotte's open data portal: https://data.charlottenc.gov/datasets/officer-traffic-stops. The older data comes from the replication files from Baumgartner et al (2020): https://fbaum.unc.edu/articles/OfficerEffects-ReplicationFiles.zip.

- **OBJECTID:** Unique identifier for the traffic stop.
- **Month_of_Stop:** Year and month (e.g., YEAR/MO).
- **Reason_for_Stop:** Reason that the stop was initiated. These are: CheckPoint, Driving While Impaired, Investigation, Other, Safe Movement, SeatBelt, Speeding, Stop Light/Sign, Vehicle Equipment, Vehicl Regulatory.
- **Officer_Race:** Race of the officer. These are: 2 or More, American Indian/Alaska Native, Asian/Pacific Islander, Black/African American, Hispanic/Latino, Native Hawaiian/Oth Pac Island, Not Specified, and White.
- **Officer_Gender:** Sex of officer (Female or Male).

- **Officer_Years_of_Service:** Number of years an officer has been on the force.
- **Driver_Race:** Driver race. These are: Asian, Black, Native American, Other/Unknown, and White.
- **Driver_Ethnicity:** Driver ethnicity. These are: Hispanic or Non-Hispanic.
- **Driver_Gender:** Driver sex (Female or Male).
- **Driver_Age:** Age of the driver.
- **Was_a_Search_Conducted:** Yes or No indicator of whether a search was conducted.
- **Result_of_Stop:** Indicator of the ultimate result of the stop: Arrest, Citation Issued, No Action Taken, Verbal Warning, or Written Warning.
- **CMPD_Division:** Police division within which the stop took place.
- **GlobalID:** Unique identifier for the traffic stop.
- **ObjectID:** Unique identifier for the traffic stop.
- **CreationDate:** Date stop entry was created in the system.
- **Creator:** Creator of the entry (either missing or CharlotteNC).
- **EditDate:** Date of last edit. Mostly missing.
- **Editor:** Creator of the entry (either missing or CharlotteNC).
- **driver_re:** 0 indicates a non-Hispanic White driver, 1 indicates a non-Hispanic Black driver, and 2 indicates a Hispanic driver.
- **of_rg:** 0 indicates a White male officer, 1 indicate a White female officer, 2 indicates a Black male officer, and 3 indicates a Black female officer.
- **of_race:** 0 indicates a White officer, and 1 indicates a Black officer.
- **of_gender:** 0 indicates a male officer, and 1 indicates a female officer.
- **investigatory:** 0 indicates a safety stop (Impaired, Speeding, Stop Light/Sign, or Movement stop), and 1 indicates an investigatory stop (all other stop types).
- **race_gender:** 0 indicates a White Male driver, 1 indicates a White female driver, 2 indicates a Black male driver, and 3 indicates a Black female driver.
- **search:** 0 indicates no search occurred, and 1 indicates a search occurred.
- **subject_sex:** female or male
- **subject_age:** Driver age.
- **officer_sex:** female or male
- **month:** Two-digit month indicator.
- **year:** Four-digit year.
- **arrest:** 0 indicates no arrest, and 1 indicates an arrest was made.

Florida Highway Patrol Data Set

This data set (FloridaSmall.RData) is a cleaned and cropped version of the Florida Highway Patrol data set, where each observation is by officer operating within a given context. It contains 2,731,204 observations and has 52 variables. Many of these variables come directly from the original data set downloaded from the Stanford Open Policing Project (SOPP) page: https://openpolicing.stanford.edu/data/. For those that remain from that original data set, the variable definition and code are quoted directly from the original codebook and are noted as such.

- **Year:** Year the stop was conducted.
- **county_name:** Name of the county where the stop was made.
- **raw_row_number:** "A number used to join clean data back to the raw data" – SOPP

- **date**: "The date of the stop, in YYYY-MM-DD format. Some states do not provide the exact stop date: for example, they only provide the year or quarter in which the stop occurred. For these states, stop_date is set to the date at the beginning of the period: for example, January 1 if only year is provided." – SOPP
- **time**: "The 24-hour time of the stop, in HH:MM format." -- SOPP
- **location**: "The freeform text of the location. Occasionally, this represents the concatenation of several raw fields, i.e. street_number, street_name" -- SOPP
- **subject_age**: "The age of the stopped subject. When date of birth is given, we calculate the age based on the stop date. Values outside the range of 10-110 are coerced to NA." -- SOPP
- **subject_race**: "The race of the stopped subject. Values are standardized to white, black, hispanic, asian/pacific islander, and other/unknown" -- SOPP
- **subject_sex**: "The recorded sex of the stopped subject." -- SOPP
- **officer_id_hash**: "A unique hash of the officer id used to identify individual officers within a location. This is usually just a hash of the provided officer ID or badge number. In some places, notably South Carolina Statewide data and the Seattle, WA city data, this ID is not unique, so we hash it with other officer attributes to make it unique, i.e. officer_last_name, officer_race, etc. In about half of locations, however, we get an officer ID but no other officer information, so our ability to test for uniqueness and deduplicate is limited. In other locations, a small number of IDs are duplicated because of what appears to be a data entry issue, i.e. Chicago, IL, where 0.3% of officer IDs appear to be only the prefix of their ID." – SOPP
- **officer_age**: "The age of the stopped officer. When date of birth is given, we calculate the age based on the stop date. Values outside the range of 10-100 are coerced to NA." -- SOPP
- **officer_race**: "The race of the stopped officer. Values are standardized to white, black, hispanic, asian/pacific islander, and other/unknown" – SOPP
- **officer_sex**: "The recorded sex of the stopped officer." – SOPP
- **officer_years_of_service**: "Number of years officer has been with the police department." – SOPP
- **department_name**: "Name of department or subdivision to which officer has been assigned." – SOPP
- **unit**: "Unit to which officer has been assigned." -- SOPP
- **type**: "Type of stop: vehicular or pedestrian." -- SOPP
- **violation**: "Specific violation of stop where provided. What is recorded here varies widely across police departments." – SOPP
- **arrest_made**: "Indicates whether an arrest made." -- SOPP
- **citation_issued**: "Indicates whether a citation was issued." – SOPP
- **warning_issued**: "Indicates whether a warning was issued." – SOPP
- **outcome**: "The strictest action taken among arrest, citation, warning, and summons." – SOPP
- **frisk_performed**: "Indicates whether a frisk was performed. This is technically different from a search, but departments will sometimes include frisks as a search type." – SOPP
- **search_conducted**: "Indicates whether any type of search was conducted, i.e. driver, passenger, vehicle. Frisks are excluded where the department has provided resolution on both." – SOPP
- **search_basis**: "This provides the reason for the search where provided and is categorized into k9, plain view, consent, probable cause, and other. If a search occurred but the reason wasn't listed, we assume probable cause." – SOPP
- **reason_for_search**: "A freeform text field indicating the reason for search where provided." – SOPP

- **reason_for_stop**: "A freeform text field indicating the reason for the stop where provided." – SOPP
- **vehicle_registration_state**: "A freeform text of the vehicle registration state where provided; format varies widely." – SOPP
- **notes**: "A freeform text field containing any officer notes." – SOPP
- **raw_row_number_old**: From SOPP. Original row number in uncleaned file.
- **raw_Race**: From SOPP. Original text entry for driver race.
- **raw_Ethnicity**: From SOPP. Original text entry for driver ethnicity.
- **raw_row_number_new**: From SOPP. New row number in semi-cleaned file.
- **raw_SearchType**: From SOPP. Original text entry for search type.
- **raw_EnforcementAction**: From SOPP. Original text entry for enforcement action.
- **investigatory**: 0 indicates a safety stop, and 1 indicates an investigatory stop.
- **contraband_found**: 1 indicates contraband was found, and 0 indicates no contraband was found.
- **race_gender**: 0 indicates a White male driver, 1 indicates a White female driver, 2 indicates a Black male driver, 3 indicates a Black female driver, 4 indicates a Latino male driver, and 5 indicates a Latina female driver.
- **of_rg**: 0 indicates a White male officer, 1 indicates a White female officer, 2 indicates a Black male officer, 3 indicates a Black female officer, 4 indicates a Latino male officer, and 5 indicates a Latina female officer.
- **of_race**: 0 indicates a White officer, 1 indicates a Black officer, 2 indicates a Latinx officer, 3 indicates an Asian/Pacific Islander officer, and 4 indicates an officer of another race.
- **of_gender**: 0 indicates a male officer, and 1 indicates a female officer.
- **out_of_state**: 0 indicates in state, and 1 indicates out of state.
- **hour_of_day**: Hour of day.
- **month**: Month of stop.
- **officers_include**: Made by an officer to be included.
- **county_include**: County to be included.
- **count.female**: Number of female officers that made a stop within the county.
- **tot.officer**: Total number of officers that made a stop within the county.
- **prop.female**: Proportion of officers that are women.
- **officer_exclude**: Indicator for whether officer should be excluded.
- **search_occur**: 0 indicates no search occurred, and 1 indicates a search occurred.
- **contra**: 0 indicates contraband was not found, and 1 indicates contraband was found.

Florida Officer Data Set

This data set (FL_Aggregated.RData) is an aggregated version of the Florida Highway Patrol data set, where each observation is by officer operating within a given context. It contains 748,887 observations and has 16 variables.

- **officer_id:** Unique identifier for each officer.
- **of_race**: 0 indicates White officer, 1 indicates Black officer, 2 indicates Latinx officer, 3 indicates an Asian/Pacific Islander officer officer, and 4 indicates an officer of another race.
- **of_gender**: 0 indicates that the officer is a man, and 1 indicates that the officer is a woman.
- **of_exper**: 0 indicates an officer with below the median amount of experience, and 1 indicates and officer with above the median amount of experience.

- **of_age**: 1 indicates less than 30, 2 indicates 30 through 64, and 3 indicates 64 and older.
- **race_gender**: 0 indicates a White male driver, 1 indicates a White female driver, 2 indicates a Black male driver, 3 indicates a Black female driver, 4 indicates a Latino male driver, and 5 indicates a Latina female driver.
- **driver_age**: 1 indicates less than 30, 2 indicates 30 through 64, and 3 indicates 64 and older.
- **out_of_state**: 0 indicates in state, and 1 indicates out of state.
- **investigatory**: 0 indicates a safety stop, and 1 indicates an out of state stop.
- **year**: Year of stop.
- **tod**: 1 indicates midnight until 3 am, 2 indicates 3 am until 6 am, 3 indicates 6 am until 9 am, 4 indicates 9 am until noon, 5 indicates noon until 3 pm, 6 indicates 3 pm until 6 pm, 7 indicates 6 pm until 9 pm, and 8 indicates 9 pm until midnight.
- **stops**: Number of stops made.
- **search_occur**: Number of searches conducted.
- **contra**: Number of times contraband is found.
- **contra.search.rate**: Rate of contraband found per 10 searches.
- **contra.stop.rate**: Rate of contraband found per 100 stops.