Data System Background and Methodology for City of Boulder Police Stop Data

Data System

Police stop data was extracted from the Boulder Police Record Management System (RMS) and was contained in two files. The main file contains demographic data of the stopped individuals, while the results file contains information on the corresponding reason and outcome of the stop. The 2 files can be joined by the numeric "rpmainid" field in the main file with the "appid" field in the results file.

Data Entry Methods

Police officers pull up a separate module when making a discretionary stop and enter the information contained in the stop data. The module is separate from any modules containing identifying information about the individual stopped, such as name and driver's license number. All fields, including stop time, duration, and street intersection, are entered in the system by the officer. Police procedure instructs that all information is entered by the end of an officer's shift on the same day.

Officers enter the individual's year of birth, sex, and whether they are a Boulder resident based on the information presented in the identification card and/or reported by the individual. For race and ethnicity information, officers record this information based on their perception of the individual. All Boulder Police officers complete mandatory periodic trainings on the data entry system, the FBI's race/ethnicity classifications that are used in the system, and identification methods.

Data Cleaning Methods for Annual Police Report

Data published on the Open Data Catalog is the raw data with no manipulation performed. Basic data cleaning was performed for the annual report; details for each year are included below.

2018 Data File

The main file (n=8215) was composed of demographic data of stopped individuals while the results file (n=39568) included the corresponding reason and outcome of the stop.

Data was loaded into RStudio version 1.1.442 and prepared for cleaning and analysis. Rows of data were removed that had a stop outside of calendar year 2018. This resulted in removal of 9 records in the stop file.

Individuals with more than one race, ethnicity or sex category were recoded as "Unknown". Due to data collection methods in the stop data, race and ethnicity was combined by creating two new categories out of the "White" category called: "White- Hispanic" and "White Non-Hispanic". Individuals whose year of birth was recorded as being before 1900 or 2018 or after were recoded as "N/A".

After data cleaning was complete, the stop main file and results file were joined by the unique individual id, which resulted in a loss of 18 id's due to no matches between the files. The joined stop data had a total of n=39524 rows.

All data analysis conducted by the City of Boulder's Analytics Team was validated using two different methods. The code used to clean and analyze the data was reviewed by city staff with experience in RStudio. Additionally, all final numbers included in the annual report were validated by city staff using Microsoft Excel to re-create the same numbers.