

Why Statistics is Not Data Science

Chris Malone | Tisha Hooks

www.statsclass.org/uscots2017

Goal: Identify New York City Police precincts who are most unbiased (or biased) regarding the rate in which police “Stop/Question/Frisk” people.

Data Sources

- Inspiration: <http://johnkeefe.net/nyc-police-precinct-and-census-data>
- NYC Stop/Question/Frisk Data - <http://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page>
- Census Data (2010 / Curator John Keefe): <https://fusiontables.google.com/DataSource?dsrclid=767562>

Stop/Question/Frisk Dataset

- Each row provides information gathered from one individual, e.g. age, race, reason for stop, etc.
- Precinct is identified for each row, also, several years of data is available
- Variable description for race

race	
	NOT LISTED
A	ASIAN/PACIFIC ISLANDER
B	BLACK
I	AMERICAN INDIAN/ALASKAN NATIVE
P	BLACK-HISPANIC
Q	WHITE-HISPANIC
W	WHITE
X	UNKNOWN
Z	OTHER

Census Data (2010)

- Each row provides information on the population within a particular block / neighborhood
- Population can be broken down by race / ethnicity

Variables that are available

RACE [71]

Universe: Total population

Total:

Population of one race:

White alone

Black or African American alone

American Indian and Alaska Native alone

Asian alone

Native Hawaiian and Other Pacific Islander alone

Some Other Race alone

Two or More Races:

Population of two races:

White; Black or African American

White; American Indian and Alaska Native

White; Asian

White; Native Hawaiian and Other Pacific Islander

White; Some Other Race

Black or African American; American Indian and Alaska Native

Black or African American; Asian

Black or African American; Native Hawaiian and Other Pacific Islander

Black or African American; Some Other Race

American Indian and Alaska Native; Asian

American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander

American Indian and Alaska Native; Some Other Race

Asian; Native Hawaiian and Other Pacific Islander

Asian; Some Other Race

Native Hawaiian and Other Pacific Islander; Some Other Race

Other Race

P0010001

P0010002

P0010003

P0010004

P0010005

P0010006

P0010007

P0010008

P0010009

P0010010

P0010011

P0010012

P0010013

P0010014

P0010015

P0010016

P0010017

P0010018

P0010019

P0010020

P0010021

P0010022

P0010023

P0010024

P0010025

Variables for separating

Hispanic from Non-Hispanic

HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE [73]

Universe: Total population

Total:

Hispanic or Latino

Not Hispanic or Latino:

Population of one race:

White alone

Black or African American alone

American Indian and Alaska Native alone

Asian alone

Native Hawaiian and Other Pacific Islander alone

Some Other Race alone

P002000

P002000

P002000

P002000

P002000

P002000

P002000

P002000

P002000

P002000

P002000

Why Statistics is Not Data Science

Chris Malone | Tisha Hooks

www.statsclass.org/uscots2017

Identify the procedural steps in providing an *analytics* solution for the goal stated above.

Step 1	
Step 2	
Step 3	
Step 4	
Step 5	
Step 6	
Step 7	
Step 8	
Step 9	
Step 10	