Supplementary Information:

Visualizing Police Exposure by Race, Gender and Age in New York City

Kalisha Dessources Figures and Joscha Legewie

January 2, 2019

The figure presented in this article is based on two data sources. The first is data from the Decennial Census and the second is data on Stop, Question and Frisk operations in New York City. We use these two data sources to determine the number of residents and the number of Stop, Question and Frisk operations for each subgroup defined by race, gender and age. This data allows us to calculate the rate of police stops per 1,000 residents across all subgroups, which is the central statistic shown in the figure.

Before For this purpose, we first create a working environment in R.

```r
# set working director
PATH <- "/Users/jlegewie/Documents/projects/policing-education/paper-socius"
## load libraries
library("tidyverse")
library("tidycensus")
library("gridExtra")
library("sessioninfo")
# Note: The current CRAN version of tidycensus (v 0.8.1) is broken.
# You can install a fixed version with
# `devtools::install_github("jlegewie/tidycensus")`
```

Census Data

The first data source is data from the Decennial Census, Summary File 1, Table PCT012. Table PCT012 includes population counts for Sex By Age with separate tables by race/ethnicity. We download this data through the Census API for all five New York City counties.

We continue by downloading a dataset with the variables in Summary File 1 of the 2010 Decennial Census, restrict the dataset to Table PCT012 and create indicators for female, race and age.

```r
# Variables from Census table PCT012. Sex By Age
sf1_vars <- load_variables(2010, "sf1", cache = FALSE)
PCT012 <- sf1_vars %>%
  filter(str_detect(name, "PCT012"), str_detect(concept, "(HISPANIC|Hispanic)")) %>%
mutate(
  female = case_when(
    ...
The result is a dataset with 1,672 rows where each row refers to one variable from Table PCT012 in the Summary File 1 of the 2010 Decennial Census with information on the variable name, label and which group the variable refers to.

In the next step, we download the data for all 1,672 variables for all counties in New York State using the tidycensus::get_decennial function. We also join the information from our PCT012 to the downloaded data, recode the race variable, restrict the dataset to the five counties in New York City and aggregate the population counts across NYC.
The result is a dataset with four variables that includes the 2010 population counts in New York City by race, gender and age.

## # A tibble: 1,296 x 4
## # Groups:   race, female [?]
## race     female   age   pop
## <chr>     <lgl>  <int> <dbl>
## 1 African-American FALSE      5 11568
## 2 African-American FALSE      6 11399
## 3 African-American FALSE      7 11499
## 4 African-American FALSE      8 11811
## 5 African-American FALSE      9 12146
## 6 African-American FALSE     10 12504
## 7 African-American FALSE     11 12056
## 8 African-American FALSE     12 12442
## 9 African-American FALSE     13 12606
## 10 African-American FALSE    14 13161
## # ... with 1,286 more rows

### Stop, Question and Frisk Data

The second data source are pedestrian stops based on the “Stop, Question and Frisk” program and include records on 4.6 million time- and geocoded police stops of pedestrians in New York City between 2004 and 2012. Stops are recorded by the officer on the “Stop, Question and Frisk Report Worksheet” (UF-250 form). Each record includes information on the exact timing, geographical location, the circumstances that led to the stop, details about the stopped person, the suspected crime, and the events during the stop itself such as an arrest or the use of physical force by the police officer. For this figure, the relevant information is the race, gender and age of every person stopped by the police as part of the “Stop, Question and Frisk” program.

We open this data (download from the NYPD website) and restrict the dataset to the relevant variables and racial groups.

```r
sqf <- read_rds(file.path(PATH, "data", "nyc-sqf-2004-2012.rds"))
sqf_subset <- sqf %>%
  transmute(
    id = id,
    year = year,
    female = suspect_female,
    race = suspect_race %>%
      fct_recode("Hispanic"="Black Hispanic", "Hispanic"="White Hispanic") %>%
      as.character(),
    age = suspect_age %>% as.integer()
  )
```
We aggregate the stop, question and frisk data by year, race, gender and age to obtain the number of police stops across these groups.

```r
sqf_rga <- sqf_subset %>%
  group_by(year, race, female, age) %>%
dplyr::summarize(stops = length(id))
```

**Combining Census and SQF Data**

As a final data processing step, we combine the census data on population counts with data on the number of stop and frisk operations. We also average the stop, frisk data over the years and create a variable `rate_stops` for the rate of police stops per 1,000 residents.

```r
stop_rates <- crossing(
  year = 2004:2012,
  race = c("African-American", "White", "Hispanic"),
  female = c(FALSE, TRUE),
  age = 8:85
) %>%
  left_join(sf1, by = c("race", "female", "age")) %>%
  left_join(sqf_rga, by = c("year", "race", "female", "age")) %>%
  mutate_at(
    vars(stops, starts_with("arrests"), starts_with("crime")),
    funs(coalesce(., 0L))
  ) %>%
  mutate(rate_stops = stops / (pop/1000), gender = female %>% ifelse("Girls", "Boys")
)

# average over years
stop_rates <- stop_rates %>%
  group_by(race, female, age) %>%
  summarize_at(vars(stops, pop, starts_with("rate")), funs(mean(., na.rm = TRUE)))

ungroup()
```

```r
stop_rates <- stop_rates %>%
  ungroup()
```

```r
mutate(
  gender = female %>% ifelse("Girls", "Boys"),
  race = factor(race) %>% fct_relevel(c("African-American", "Hispanic", "White")),
  group = case_when(
    race == "African-American" & !female ~ "Black Men",
    race == "African-American" & female ~ "Black Women",
    race == "Hispanic" & !female ~ "Hispanic Men",
    race == "Hispanic" & female ~ "Hispanic Women",
    race == "White" & !female ~ "White Men",
    race == "White" & female ~ "White Women"
  )
)
Finally, we create an additional variable for the stop rate of African-American boys, which is used as a reference category in the visualization.

```r
# define black stop rate for comparison
rate_stops_black_men <- stop_rates %>%
  filter(race == "African-American", !female) %>%
  pluck("rate_stops")
```

Creating the Visualization

We now create the visualization for the rate of police stops per 1,000 residents by race, gender and age.

```r
text_label <- tibble(group = unique(stop_rates$group)) %>%
  filter(group %in% c("Hispanic Men", "White Men"))
g1 <- stop_rates %>%
  filter(between(age, 8, 65), gender == "Boys") %>%
  ggplot(aes(x = age, y = rate_stops)) +
    geom_bar(aes(y = rate_stops_black_men), stat = "identity",
             fill = "#0072B2", alpha = 0.3) +
    geom_bar(stat = "identity", fill = "#0072B2") +
    facet_grid(~ group) +
    scale_x_continuous("Age", breaks = c(10, 20, 30, 40, 50, 60)) +
    scale_y_continuous("Rate of Police Stops per 1,000", expand = c(0.025, 0)) +
    geom_text(aes(y = rate_stops_black_men * 1.05, label = "")) +
    geom_text(data = text_label, x = 21, y = 540, label = "Black Men",
              size = 2.5, alpha = 0.5) +
    theme_minimal() +
    theme(
      strip.background = element_rect(color = "black", size = 0),
      strip.text.x = element_text(size = 12, hjust = 0)
    )
text_label <- tibble(group = unique(stop_rates$group)) %>%
  filter(group %in% c("Black Women", "Hispanic Women", "White Women"))
g2 <- stop_rates %>%
  filter(between(age, 8, 65), gender == "Girls") %>%
  ggplot(aes(x = age, y = rate_stops)) +
    geom_bar(aes(y = rate_stops_black_men), stat = "identity",
             fill = "#0072B2", alpha = 0.3) +
    geom_bar(stat = "identity", fill = "#0072B2") +
    facet_grid(~ group) +
    scale_x_continuous("Age", breaks = c(10, 20, 30, 40, 50, 60)) +
    scale_y_continuous("Rate of Police Stops per 1,000", expand = c(0.025, 0)) +
    geom_text(aes(y = rate_stops_black_men * 1.05, label = "")) +
    geom_text(data = text_label, x = 21, y = 540, label = "Black Men",
              size = 2.5, alpha = 0.5) +
    theme_minimal() +
    theme(
      ...)
System and Requirements

This file was created on the following system.

sessioninfo::session_info()

```r
## ─ Session info ──────────────────────────────────────────────────────────
##  setting  value
##  version  R version 3.5.1 (2018-07-02)
##  os       macOS  10.14.2
##  system   x86_64, darwin15.6.0
##  ui       X11
##  language (EN)
##  collate  en_US.UTF-8
##  ctype    en_US.UTF-8
```
| package         | version | date       | lib      | source       |
|-----------------|---------|------------|----------|--------------|
| assertthat      | 0.2.0   | 2017-04-11 | [1]      | CRAN (R 3.5.0) |
| backports       | 1.1.2   | 2017-12-13 | [1]      | CRAN (R 3.5.0) |
| bindr           | 0.1.1   | 2018-03-13 | [1]      | CRAN (R 3.5.0) |
| bindrcpp        | 0.2.2   | 2018-03-29 | [1]      | CRAN (R 3.5.0) |
| broom           | 0.5.0   | 2018-07-17 | [1]      | CRAN (R 3.5.0) |
| cellranger      | 1.1.0   | 2016-07-27 | [1]      | CRAN (R 3.5.0) |
| class           | 7.3-14  | 2015-08-30 | [2]      | CRAN (R 3.5.1) |
| classInt        | 0.2-3   | 2018-04-16 | [1]      | CRAN (R 3.5.0) |
| cli             | 1.0.1   | 2018-09-25 | [1]      | CRAN (R 3.5.0) |
| colorspace      | 1.3-2   | 2016-12-14 | [1]      | CRAN (R 3.5.0) |
| crayon          | 1.3.4   | 2017-09-16 | [1]      | CRAN (R 3.5.0) |
| curl            | 3.2     | 2018-03-28 | [1]      | CRAN (R 3.5.0) |
| DBI             | 1.0.0   | 2018-05-02 | [1]      | CRAN (R 3.5.0) |
| digest          | 0.6.17  | 2018-09-12 | [1]      | CRAN (R 3.5.0) |
| dplyr           | 0.7.8   | 2018-11-10 | [1]      | CRAN (R 3.5.0) |
| e1071           | 1.7-0   | 2018-07-28 | [1]      | CRAN (R 3.5.0) |
| evaluate        | 0.11    | 2018-07-17 | [1]      | CRAN (R 3.5.0) |
| fansi           | 0.4.0   | 2018-10-05 | [1]      | CRAN (R 3.5.0) |
| forcats         | 0.3.0   | 2018-02-19 | [1]      | CRAN (R 3.5.0) |
| foreign         | 0.8-70  | 2017-11-28 | [2]      | CRAN (R 3.5.1) |
| ggplot2         | 3.0.0   | 2018-07-03 | [1]      | CRAN (R 3.5.0) |
| glue            | 1.3.0   | 2018-07-17 | [1]      | CRAN (R 3.5.0) |
| gridExtra       | 2.3     | 2017-09-09 | [1]      | CRAN (R 3.5.0) |
| gtable          | 0.2.0   | 2016-02-26 | [1]      | CRAN (R 3.5.0) |
| haven           | 1.1.2   | 2018-06-27 | [1]      | CRAN (R 3.5.0) |
| hms             | 0.4.2   | 2018-03-10 | [1]      | CRAN (R 3.5.0) |
| htmltools       | 0.3.6   | 2017-04-28 | [1]      | CRAN (R 3.5.0) |
| httr            | 1.4.0   | 2018-12-11 | [1]      | CRAN (R 3.5.0) |
| jsonlite        | 1.6     | 2018-12-07 | [1]      | CRAN (R 3.5.0) |
| knitr           | 1.20    | 2018-02-20 | [1]      | CRAN (R 3.5.0) |
| labeling        | 0.3     | 2014-08-23 | [1]      | CRAN (R 3.5.0) |
| lattice         | 0.20-35 | 2017-03-25 | [2]      | CRAN (R 3.5.1) |
| lazyeval        | 0.2.1   | 2017-10-29 | [1]      | CRAN (R 3.5.0) |
| lubridate       | 1.7.4   | 2018-04-11 | [1]      | CRAN (R 3.5.0) |
| magrittr        | 1.5     | 2014-11-22 | [1]      | CRAN (R 3.5.0) |
| maptools        | 0.9-4   | 2018-09-19 | [1]      | CRAN (R 3.5.0) |
| modelr          | 0.1.2   | 2018-05-11 | [1]      | CRAN (R 3.5.0) |
| munsell         | 0.5.0   | 2018-06-12 | [1]      | CRAN (R 3.4.4) |
| nlme            | 3.1-137 | 2018-04-07 | [2]      | CRAN (R 3.5.1) |
| pillar          | 1.3.0   | 2018-07-14 | [1]      | CRAN (R 3.5.0) |
| pkgconfig       | 2.0.2   | 2018-08-16 | [1]      | CRAN (R 3.5.0) |
| plyr            | 1.8.4   | 2016-06-08 | [1]      | CRAN (R 3.5.0) |
| purrr           | 0.2.5   | 2018-05-29 | [1]      | CRAN (R 3.5.0) |
| R6              | 2.3.0   | 2018-10-04 | [1]      | CRAN (R 3.5.0) |
| rappdirs        | 0.3.1   | 2016-03-28 | [1]      | CRAN (R 3.5.0) |
##  Rcpp          1.0.0   2018-11-07 [1] CRAN (R 3.5.0)
##  readr       * 1.3.0   2018-12-11 [1] CRAN (R 3.5.0)
##  readxl        1.1.0   2018-04-20 [1] CRAN (R 3.5.0)
##  reshape2      1.4.3   2017-12-11 [1] CRAN (R 3.5.0)
##  rgdal         1.3-6   2018-10-16 [1] CRAN (R 3.5.0)
##  rlang         0.3.0.1 2018-10-25 [1] CRAN (R 3.5.0)
##  rmarkdown     1.10    2018-06-11 [1] CRAN (R 3.5.0)
##  rprojroot     1.3-2   2018-01-03 [1] CRAN (R 3.5.0)
##  rstudioapi    0.8     2018-10-02 [1] CRAN (R 3.5.0)
##  rvest         0.3.2   2016-06-17 [1] CRAN (R 3.5.0)
##  scales        1.0.0   2015-08-09 [1] CRAN (R 3.5.0)
##  selectr       0.4-1   2018-04-06 [1] CRAN (R 3.5.0)
##  sessioninfo * 1.1.1   2018-11-05 [1] CRAN (R 3.5.1)
##  sf            0.7     2018-10-24 [1] CRAN (R 3.5.0)
##  sp            1.3-1   2018-06-05 [1] CRAN (R 3.5.0)
##  spData        0.2.9.6 2018-12-03 [1] CRAN (R 3.5.0)
##  stringi       1.2.4   2018-07-20 [1] CRAN (R 3.5.0)
##  stringr       * 1.3.1   2018-05-10 [1] CRAN (R 3.5.0)
##  tibble        * 1.4.2   2018-01-22 [1] CRAN (R 3.5.0)
##  tidycensus    * 0.8.3   2018-12-14 [1] Github (jlegewie/tidycensus@8edf9fd)
##  tidyverse     * 1.2.1   2017-11-14 [1] CRAN (R 3.5.0)
##  tigris        0.7     2018-04-14 [1] CRAN (R 3.4.4)
##  units         0.6-2   2018-12-05 [1] CRAN (R 3.5.0)
##  utf8          1.1.4   2018-05-24 [1] CRAN (R 3.5.0)
##  uuid          0.1-2   2015-07-28 [1] CRAN (R 3.5.0)
##  withr         2.1.2   2018-03-15 [1] CRAN (R 3.5.0)
##  xml2          1.2.0   2018-01-24 [1] CRAN (R 3.5.0)
##  yaml          2.2.0   2018-07-25 [1] CRAN (R 3.5.0)

## [1] /Users/jlegewie/R-packages
## [2] /Library/Frameworks/R.framework/Versions/3.5/Resources/library