Mapping Police Misconduct Settlements by Neighborhood Classification

Becca Tramel

Introduction

My earlier map shows the incident locations for a stratified sample of police misconduct settlements by year (2008-2018). A friend noticed a similarity between the maps I made and maps of gentrification in New York City.

I found the Urban Displacement Project (https://www.urbandisplacement.org/maps/ny) while looking at this connection. They've created a data set classifying census tracts in New York City by level of gentrification using census data.

Since I'd already geocoded my sample using the Google API, I had latitude and longitude for almost all of my incidents. I then used Federal Communications Commission API (https://geo.fcc.gov/api/census/) to get census tract information for each incident (see my Jupyter notebook to see how.)

I now have a file with the coordinates of each incident along with a census tract. I'm going to join this with the classifications from the Urban Displacement Project and remake my map.

Adding a Neighborhood Classification to the Data

```
## Warning: package 'tidyverse' was built under R version 3.6.3
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.1
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tibble' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'readr' was built under R version 3.6.3
```

```
## Warning: package 'purrr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'stringr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Warning: package 'ggmap' was built under R version 3.6.3
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
# First I'll import the locations.
coords_with_tracts <- read_csv(url("https://raw.githubusercontent.com/becca-t/policemisconductny
c/main/coords_with_tracts"))
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   lon = col_double(),
##   lat = col_double(),
##   tract = col_character()
## )
```

```
# I'll now import the classifications (available at https://www.urbandisplacement.org/maps/ny)

classifications <- read_csv("udp_ny_final_typology_jan_2019.csv")
```

```
## Parsed with column specification:
## cols(
##   geoid = col_double(),
##   Type_1.19 = col_character()
## )
```

Now I'll combine the two data sets to add a classification column to the sample of incident locations.

```
# First we let R know that tract is numerical.
coords_with_tracts$tract <- as.numeric(coords_with_tracts$tract)

# Creating the joined table.
coords_with_class <- left_join(coords_with_tracts, classifications, by = c("tract" = "geoid"))
```

Studying the Neighborhood Classifications

Now that the classifications are added, I'll start by looking at some summary statistics.

```
prop.table(table(coords_with_class$Type_1.19))
```

```
##
##                     LI - At Risk of Gentrification
##                                         0.24041160
##              LI - Not Losing Low-Income Households
##                                         0.14031805
## LI - Ongoing Displacement of Low-Income Households
##                                         0.16744621
##                        LI - Ongoing Gentrification
##                                         0.20579981
##                      MHI - Advanced Gentrification
##                                         0.02712816
##                             MHI - Ongoing Exclusion
##                                         0.04770814
##                              MHI - Stable Exclusion
##                                         0.12722170
##                                       Missing Data
##                                         0.01496726
##            VHI - Super Gentrification or Exclusion
##                                         0.02899906
```

From this, we learn that

```
* 24.0% of cases in our sample are classified as "At Risk of Gentrification"
* 20.6% are classified as "Ongoing Gentrification"
* 16.7% are classified as "Ongoing Displacement of Low-Income Households.
```

According to the Urban Displacement Project, these are three categories that describe low-income neighborhoods.

- The neighborhoods described as "At Risk of Gentrification" are "Hot markets" from 2000 - 2016 (places where median rent changes are increasing more quickly than the regional average).
- Neighborhoods described as "Ongoing Gentrification" are "gentrifying" in 1990-2000 or 2000-2016 (changes in both median income and percentage of college educated residents are increasing greater than the regional average).
- "Ongoing Displacement of Low-Income Households" are neighborhoods described as losing low-income households in 2000-2016.

The other category of low-income neighborhoods, "Not Losing Low-Income Households", accounts for 14.0% of the data.

With this summary information in front of us, it might be helpful to compare with the percentages for neighborhoods in New York City overall.

```
# Some of the counties in the classifications data are outside of New York (in Connecticut and N
ew Jersey), or in nearby counties. First we'll filter down to the 5 counties of NYC.
nyc_classifications <- filter(classifications, str_detect(geoid, "36047[0-9]{6}") | str_detect(g
eoid, "36059[0-9]{6}") | str_detect(geoid, "36061[0-9]{6}") | str_detect(geoid, "36085[0-9]{6}")
| str_detect(geoid, "36005[0-9]{6}"))

prop.table(table(nyc_classifications$Type_1.19))
```

```
##
##                     LI - At Risk of Gentrification
##                                         0.09628378
##              LI - Not Losing Low-Income Households
##                                         0.15259009
## LI - Ongoing Displacement of Low-Income Households
##                                         0.11148649
##                        LI - Ongoing Gentrification
##                                         0.10754505
##                      MHI - Advanced Gentrification
##                                         0.06362613
##                           MHI - Ongoing Exclusion
##                                         0.10078829
##                            MHI - Stable Exclusion
##                                         0.29673423
##                                       Missing Data
##                                         0.02477477
##            VHI - Super Gentrification or Exclusion
##                                         0.04617117
```

From here, we can see that these four low-income neighborhood types make up a much smaller percentage of neighborhoods overall.

- At risk of gentrification: 9.6%
- Ongoing gentrification: 10.8%
- Ongoing displacement of low-income households: 11.1%
- Not losing low-income households: 15.3%

Some conclusions:

- Neighborhoods which are at risk for or currently undergoing gentrification are vastly overrepresented in the police misconduct settlement data.
- Specifically: neighborhoods “at risk of gentrification” show up about 2.5 times as much as expected, neighborhoods with “ongoing gentrification” show up about 1.9 times as much as expected, and neighborhoods with “ongoing displacement of low-income households” show up about 1.5 times as much as expected.
- Other low-income neighborhoods are not overrepresented in the data. (14% of the cases were in these neighborhoods, compared with 15.3% of neighborhoods overall fitting into these categories.)

Of course, these neighborhoods are not all equally sized! We’d need to account for population to better study this pattern.

Adding in Populations for Further Study

Populations for census tracts (as of 2010) are available from the Census website here (https://www.census.gov/geographies/reference-files/2010/geo/2010-centers-population.html). Here is the file for New York.

```
census_tract_pops <- read_csv(url("https://www2.census.gov/geo/docs/reference/cenpop2010/tract/CenPop2010_Mean_TR36.txt"))
```

```
## Parsed with column specification:
## cols(
##   STATEFP = col_double(),
##   COUNTYFP = col_character(),
##   TRACTCE = col_character(),
##   POPULATION = col_double(),
##   LATITUDE = col_double(),
##   LONGITUDE = col_double()
## )
```

To turn this into a compatible dataframe, we'll need to do a little bit of cleaning. In particular, we'll have to combine the first three columns to get the census tract.

```
census_tract_pops2 <- select(census_tract_pops, -LATITUDE, - LONGITUDE)

census_tract_pops2$TRACT <- str_c(as.character(census_tract_pops2$STATEFP), census_tract_pops2$COUNTYFP, census_tract_pops2$TRACTCE)

census_tract_pops2 <- select(census_tract_pops2, TRACT, POPULATION)
```

Now we can join this with our existing list of New York City census tracts to study the population of each classification.

```
census_tract_pops2$TRACT <- as.numeric(census_tract_pops2$TRACT)

ny_classifications_and_pop <- left_join(nyc_classifications, census_tract_pops2, by = c("geoid" = "TRACT"))
```

With this information, we can now figure out the percentage of the population of NYC living in each classification of neighborhood.

```
population_by_class <- ny_classifications_and_pop %>%
  group_by(Type_1.19) %>%
  summarize(population = sum(POPULATION))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
population_by_class$pop_pct <- population_by_class$population/sum(population_by_class$population)
```

```
write_csv(population_by_class, "population_by_class.csv")
```

With this information we can draw the following conclusions.

- 11.7% of New Yorkers live in “At Risk of Gentrification” neighborhoods, but 24% of incidents in our sample take place in such neighborhoods.
- 11.1% of New Yorkers live in neighborhoods with “Ongoing Displacement of Low-Income Households” but 16.7% of incidents in our sample take place in such neighborhoods.
- 12.0% of New Yorkers live in neighborhoods with “Ongoing Gentrification” but 20.6% of incidents in our sample take place in these neighborhoods.
- Interestingly, 13.8% of New Yorkers live in neighborhoods classified as Low-Income but “Not Losing Low-Income Households”, and 14% of incidents in our sample take place in such neighborhoods.

In other words, it doesn’t appear that being in a low-income neighborhood greatly increases the likelihood of reporting an incident of police misconduct leading to a settlement, but living in a gentrifying neighborhood does.

A Map to Display the Data

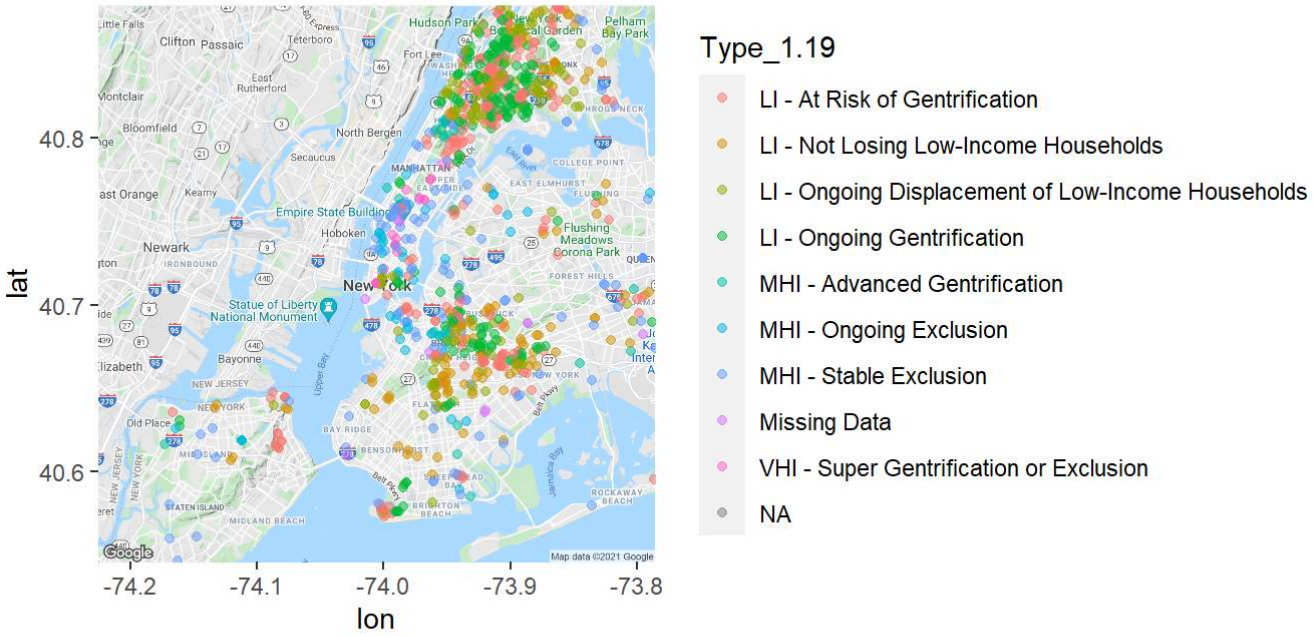
First I’ll make a map of New York City.

```
map_of_nyc <- get_googlemap(center = c(lon = -74.0060, lat = 40.7128), zoom = 11)
```

Now I’ll overlay the map with our cases, colored by neighborhood classification.

```
ggmap(map_of_nyc) +
  geom_point(data = coords_with_class, aes(x = lon, y = lat, color = Type_1.19), alpha = 0.5)
```

```
## Warning: Removed 93 rows containing missing values (geom_point).
```


lat
40.8
40.7
40.6
lon
-74.2
-74.1
-74.0
-73.9
-73.8
Type_1.19
LI - At Risk of Gentrification
LI - Not Losing Low-Income Households
LI - Ongoing Displacement of Low-Income Households
LI - Ongoing Gentrification
MHI - Advanced Gentrification
MHI - Ongoing Exclusion
MHI - Stable Exclusion
Missing Data
VHI - Super Gentrification or Exclusion
NA
Map data ©2021 Google