

36-315 Final Project Zip Code

Arthur Jakobsson, Alex Cheng, Liz Chu, Kevin Ren

November 18, 2022

Contents

Rat Sighting Dataset - Data Pre-analysis

1

Rat Sighting Dataset - Data Pre-analysis

```
# library imports
```

```
library(tigris)
library(dplyr)
library(leaflet)
library(tidyverse)
library(sp)
library(ggmap)
library(maptools)
library(broom)
library(httr)
library(rgdal)
library(gridExtra)
library(stringr)
library(geosphere)
library(gpclib)
library(broom)
library(geojsonio)
library(tidyverse)
library(plotly)
library(maps)
library(reshape2)
library(shiny)
```

```
zipnames$ZipCode <- as.character(zipnames$ZipCode)
zipnames = subset(zipnames, select = -c(X, X.1, X.2, X.3))
nycziips = nycziips %>% left_join(., zipnames, by = c("id" = "ZipCode"))
```

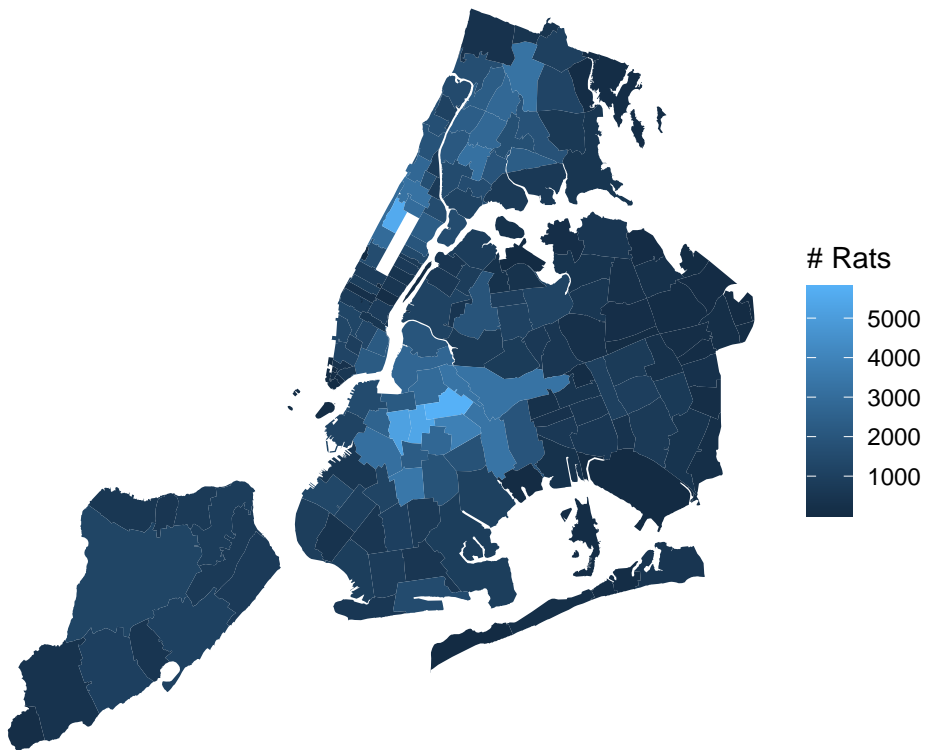
```
# SINGULAR VARIABLE ANALYSIS: # RATS, # RESTAURANTS, AVG RESTAURANT SCORE
ggplot() +
  geom_polygon(data = nycziips, aes(x = long, y = lat, group = group, fill = n_rats)) +
```

```

theme_void() +
# scale_fill_gradient2(low = "darkblue", mid = "purple", high = "pink", midpoint=3000) +
coord_map() + labs(
  title = "Rat Sightings by Zip Code",
  fill = "# Rats"
)

```

Rat Sightings by Zip Code

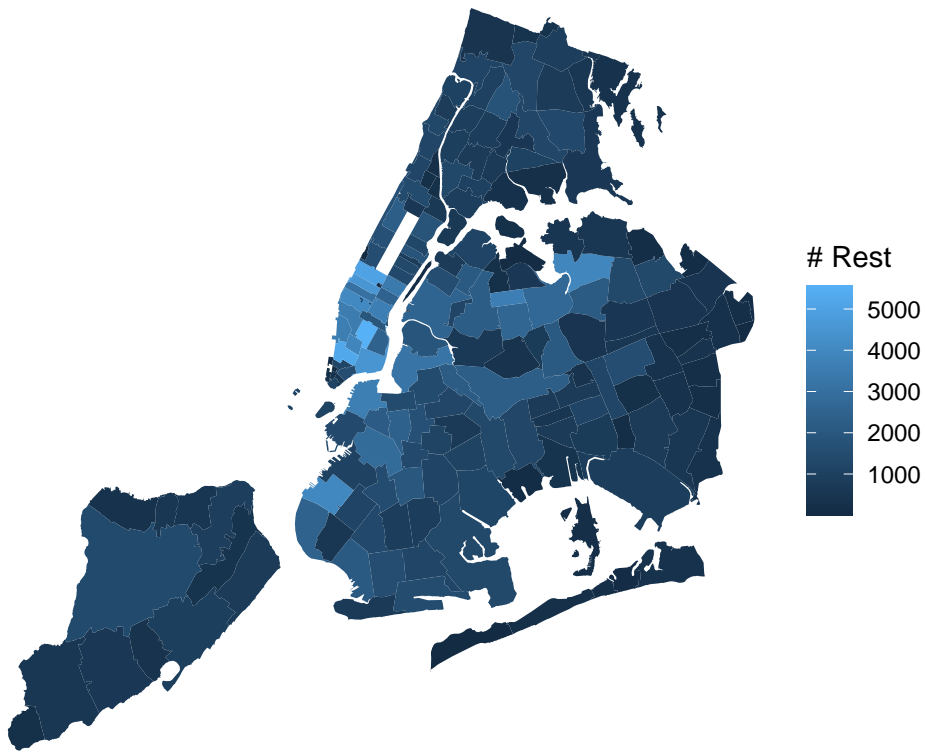


```

ggplot() +
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = n_rest)) +
  theme_void() +
  coord_map() + labs(
    title = "Number of Restaurants by Zip Code",
    fill = "# Rest"
  )

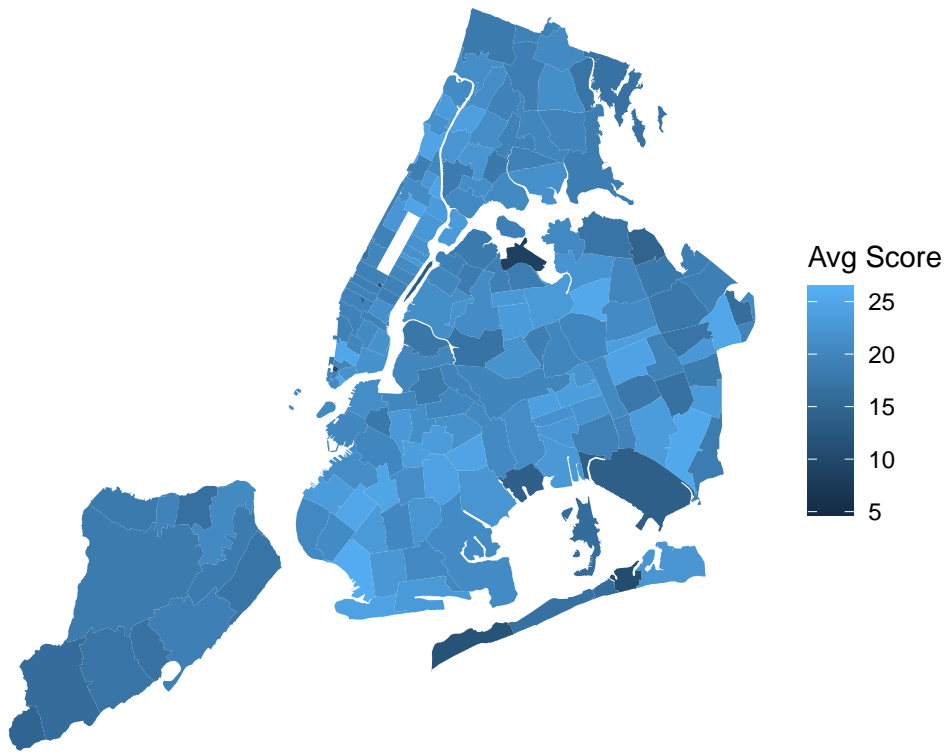
```

Number of Restaurants by Zip Code



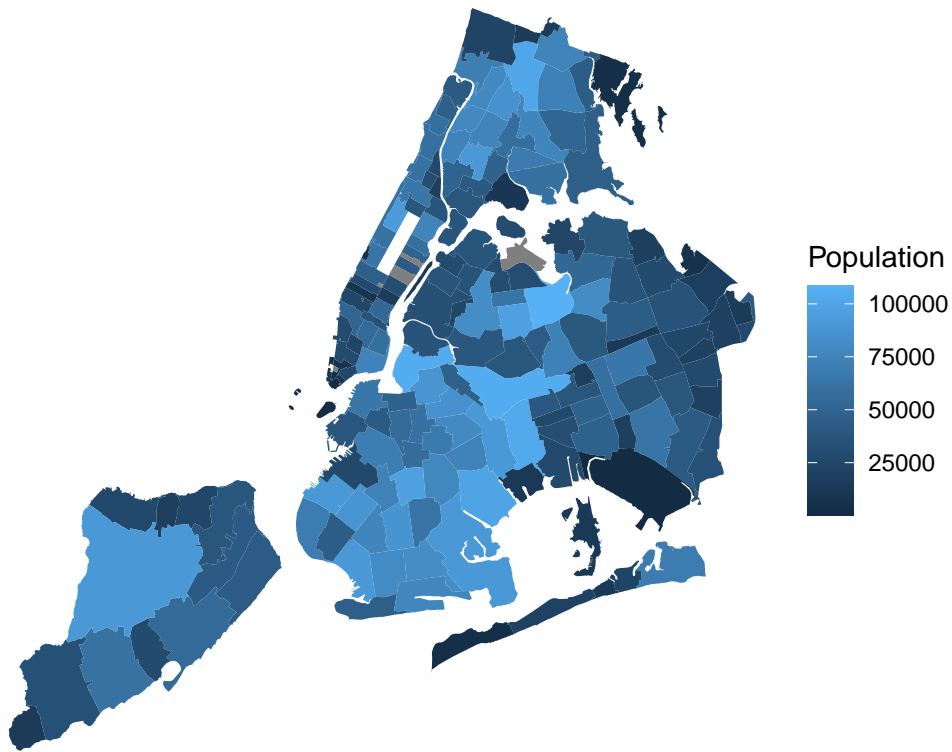
```
ggplot() +  
  geom_polygon(data = nyczips, aes(x = long, y = lat, group = group, fill = score_avg)) +  
  theme_void() +  
  coord_map() + labs(  
    title = "Restaurant Scores by Zip Code",  
    fill = "Avg Score"  
  )
```

Restaurant Scores by Zip Code



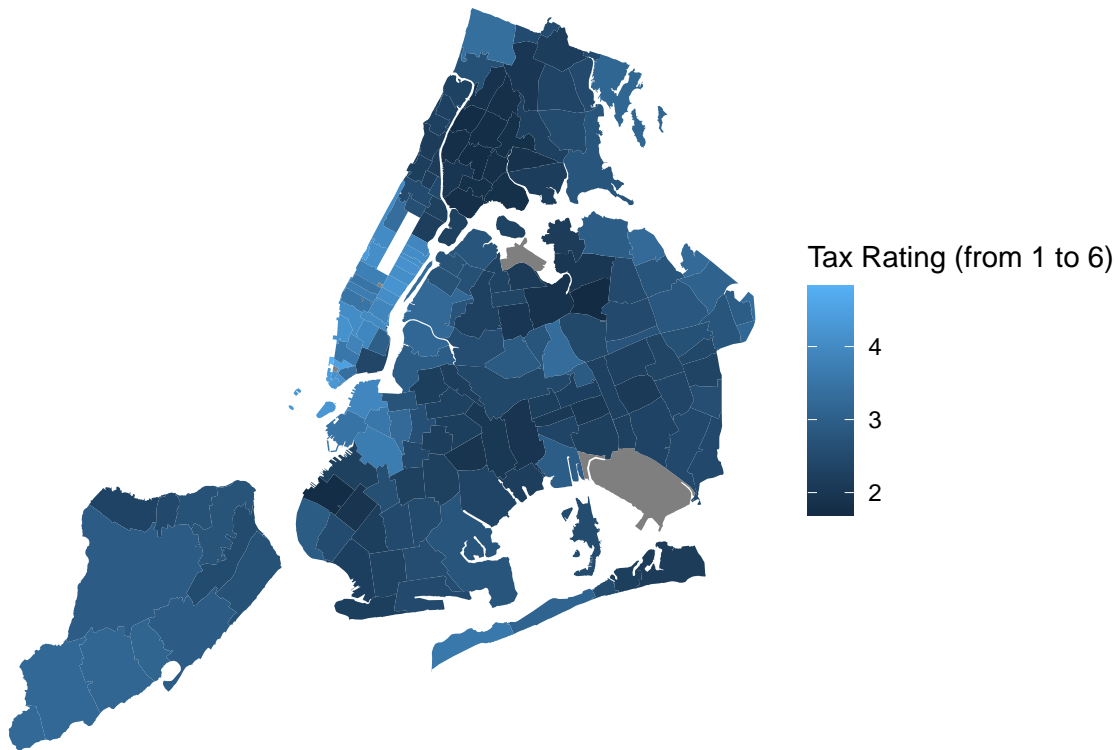
```
ggplot() +  
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = population)) +  
  theme_void() +  
  coord_map() + labs(  
    title = "Population by Zip Code",  
    fill = "Population"  
  )
```

Population by Zip Code



```
ggplot() +  
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = tax_rating)) +  
  theme_void() +  
  coord_map() + labs(  
    title = "Tax Rating by Zip Code",  
    fill = "Tax Rating (from 1 to 6)"  
  )
```

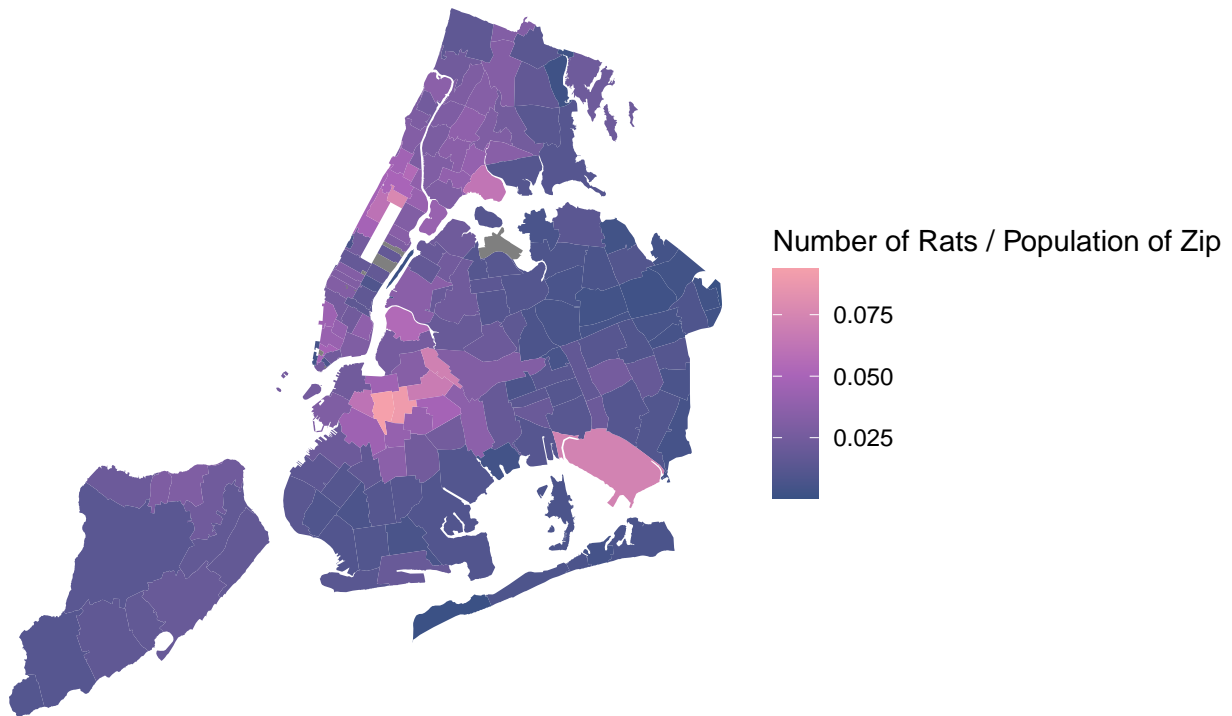
Tax Rating by Zip Code



```
# rats per capita

ggplot() +
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = rat_per_cap)) +
  theme_void() +
  scale_fill_gradient2(low = "#395184",
                       mid = "#A964B8",
                       high = "#FFA9A9", midpoint = 0.05) +
  coord_map() + labs(
    title = "Rats Per Capita",
    fill = "Number of Rats / Population of Zip"
  )
```

Rats Per Capita

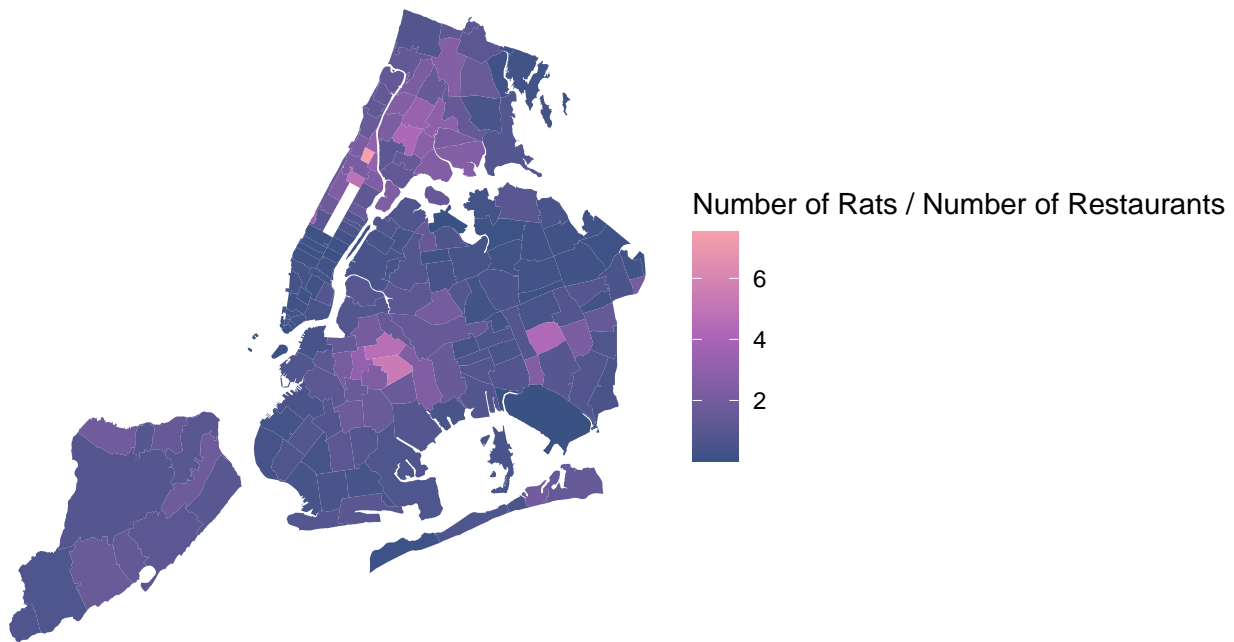


THROWN OUT: RESTAURANT TO RAT RATIOS

```
# ggplot() +  
#   geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = rest_to_rat)) +  
#   theme_void() +  
#   scale_fill_gradient2(low = "#395184",  
#                         mid = "#A964B8",  
#                         high = "#FFA9A9", midpoint = 200) +  
#   coord_map() + labs(  
#     title = "Restaurant to Rat Ratio by Zip Code",  
#     fill = "Number of Restaurants / Number of Rats"  
#   )
```

```
ggplot() +  
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = rat_to_rest)) +  
  theme_void() +  
  scale_fill_gradient2(low = "#395184",  
                      mid = "#A964B8",  
                      high = "#FFA9A9", midpoint = 4) +  
  coord_map() + labs(  
    title = "Rat to Restaurant Ratio by Zip Code",  
    fill = "Number of Rats / Number of Restaurants"  
  )
```

Rat to Restaurant Ratio by Zip Code



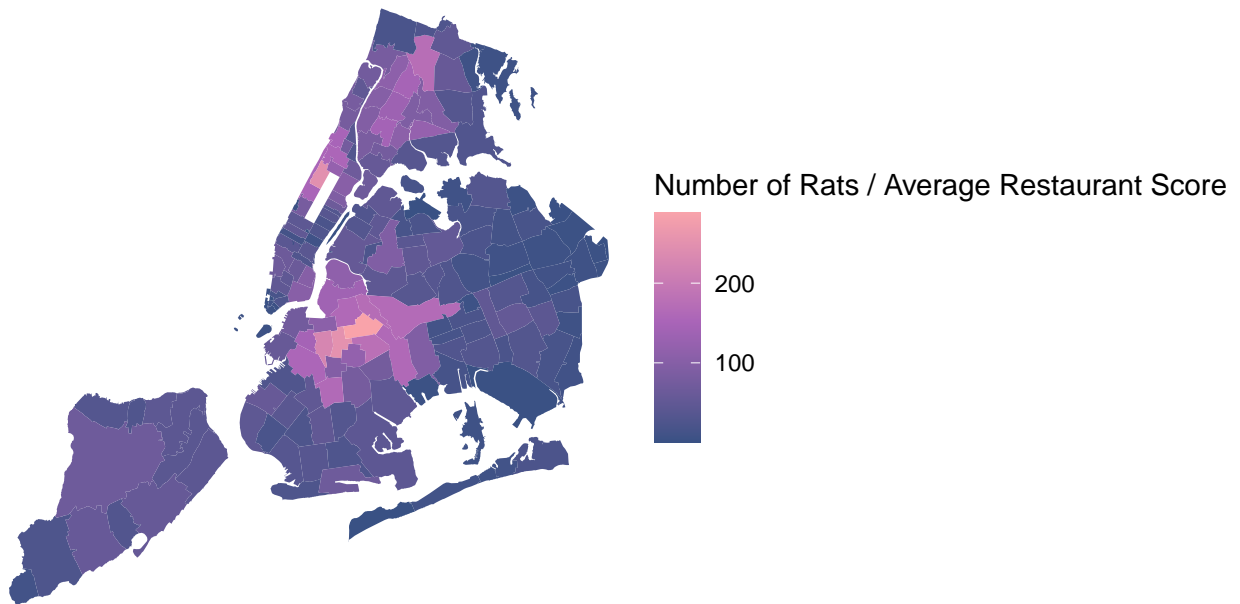
```
# THROWN OUT: SCORE TO RAT RATIOS

# ggplot() +
#   geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = score_to_rat)) +
#   theme_void() +
#   scale_fill_gradient2(low = "#395184",
#                         mid = "#A964B8",
#                         high = "#FFA9A9", midpoint = 10) +
#   coord_map() + labs(
#     title = "Restaurant Score to Number of Rats by Zip Code",
#     fill = "Average Restaurant Score / Number of Rats"
#   )

ggplot() +
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = rat_to_score)) +
  theme_void() +
  scale_fill_gradient2(low = "#395184",
                      mid = "#A964B8",
                      high = "#FFA9A9", midpoint = 150) +
  coord_map() + labs(
    title = "Rats to Restaurant Score Ratio by Zip Code",
    subtitle = "higher score = worse restaurant",
    fill = "Number of Rats / Average Restaurant Score"
  )
```


Rats to Restaurant Score Ratio by Zip Code

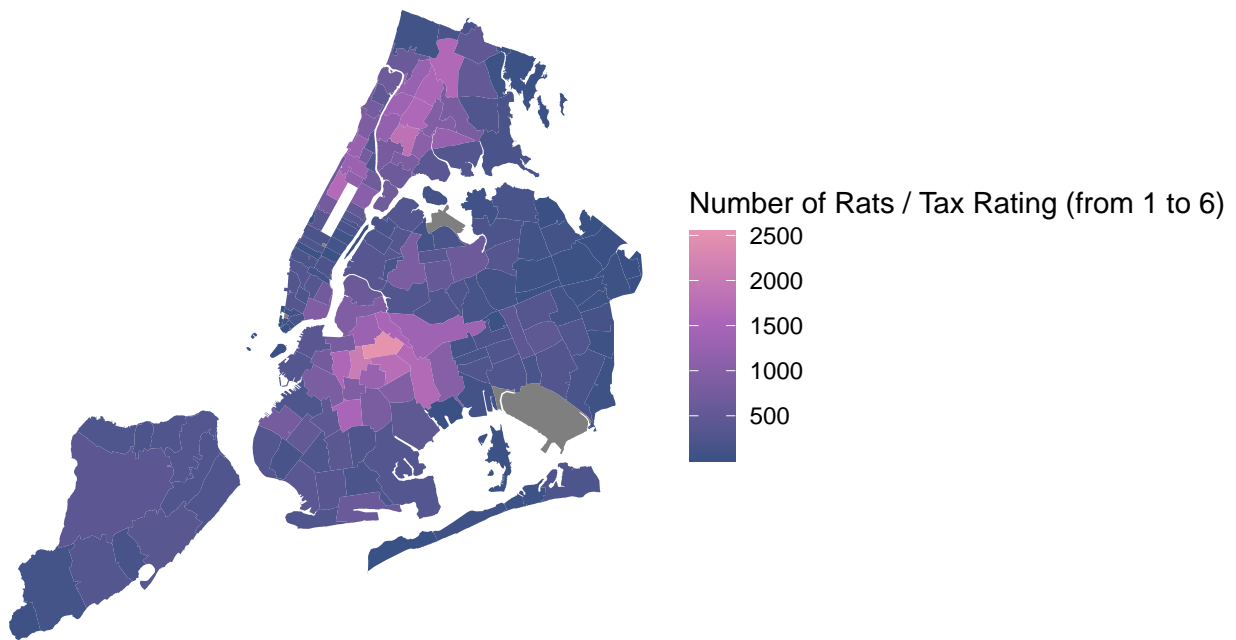
higher score = worse restaurant



```
# tax bracket stuff

ggplot() +
  geom_polygon(data = nyczip, aes(x = long, y = lat, group = group, fill = rat_to_tax)) +
  theme_void() +
  scale_fill_gradient2(low = "#395184",
                       mid = "#A964B8",
                       high = "#FFA9A9", midpoint = 1500) +
  coord_map() + labs(
    title = "Rat to Tax Rating Ratio by Zip Code",
    fill = "Number of Rats / Tax Rating (from 1 to 6)"
  )
```

Rat to Tax Rating Ratio by Zip Code



```
# ggplotly(p)

zipout = subset(nyczip, select = -c(long, lat, order, hole,
                                   piece, group))

colnames(zipout)[1] = "zipcode"
colnames(zipout)[3] = "borough"
colnames(zipout)[4] = "neighborhood"

zipout = zipout %>% distinct()

write.csv(zipout, file = "zipout.csv")
```