Google Movement Data Analysis

Albert Sun, Lily Zhu

```
library(tidyverse)
```

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

```
## v tibble  3.0.3     v purrr   0.3.4
## v tidyr   1.1.1     v dplyr   1.0.1
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x readr::guess_encoding()  masks rvest::guess_encoding()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x purrr::pluck()           masks rvest::pluck()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

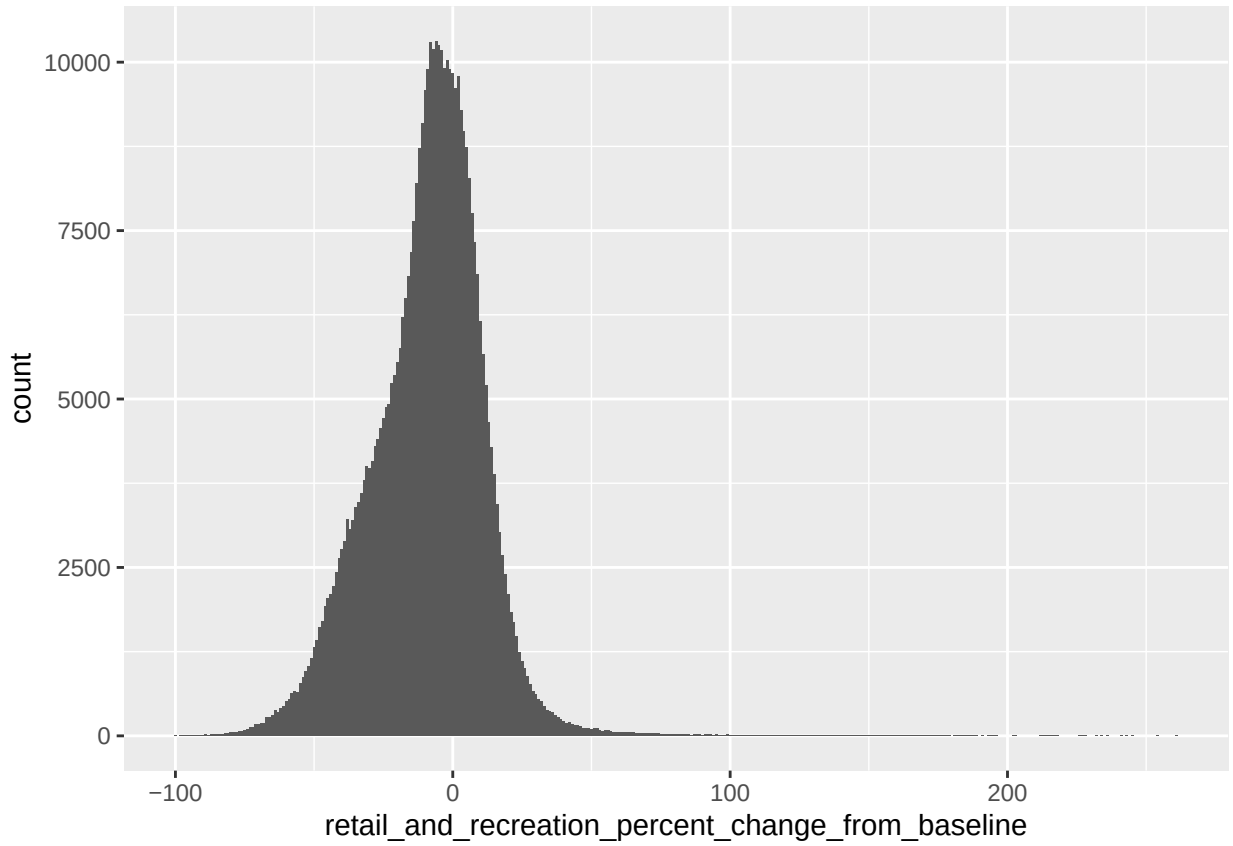
```
#library(hrbrthemes)
```

```
mobility <- read.csv("../data/US-Mobility-Report.csv")
```

```
mobility <- mobility %>%
  mutate(date_num = as.Date(date))
```

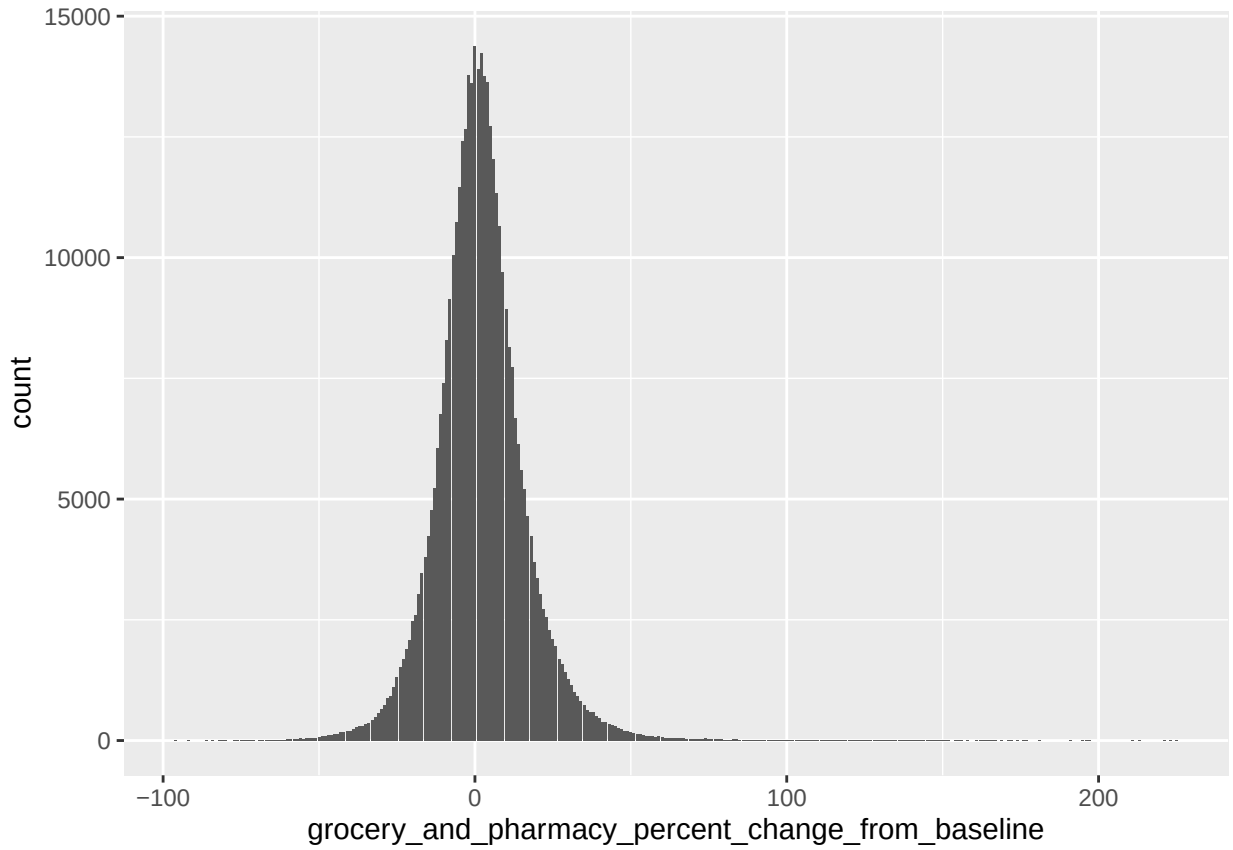
```
ggplot(data = mobility, mapping = aes(x = retail_and_recreation_percent_change_from_baseline)) +
  geom_bar()
```

```
## Warning: Removed 210137 rows containing non-finite values (stat_count).
```



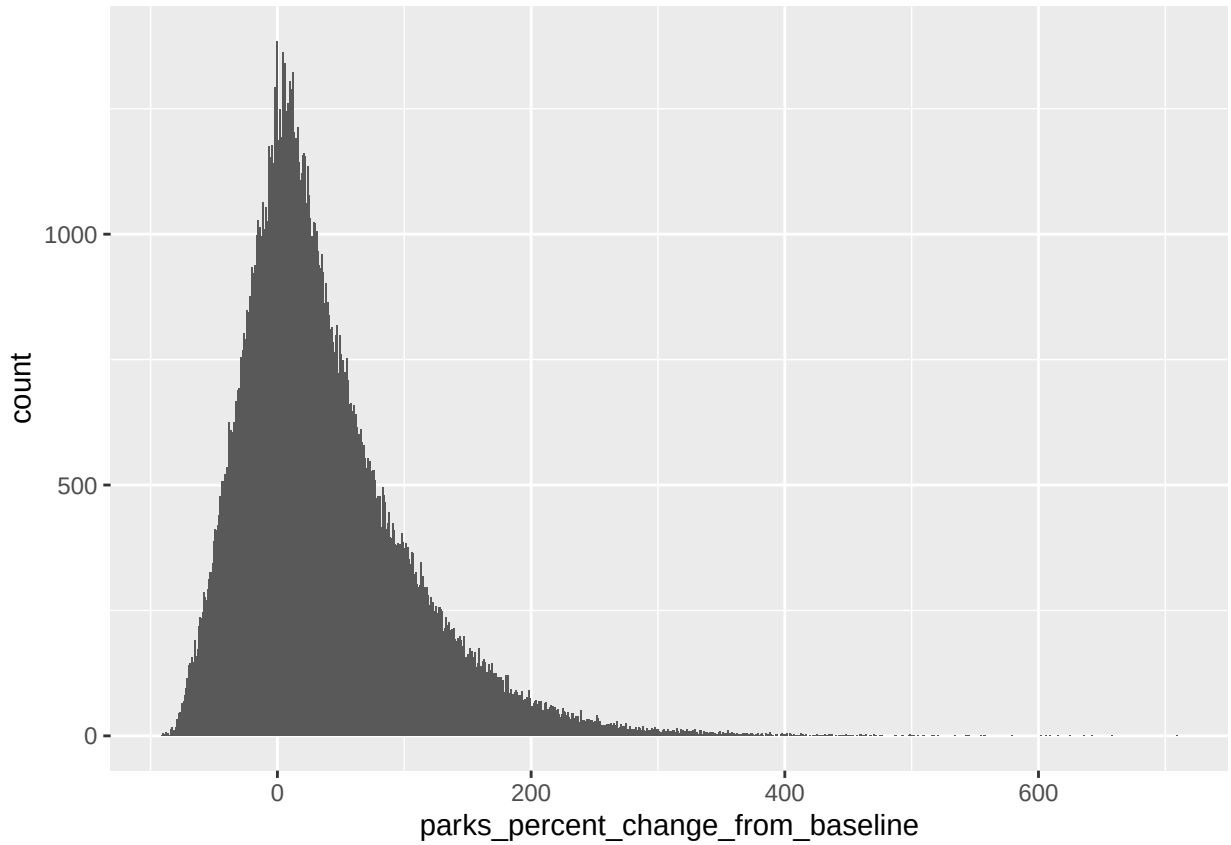

```
ggplot(data = mobility, mapping = aes(x = grocery_and_pharmacy_percent_change_from_baseline)) +
  geom_bar()
```

```
## Warning: Removed 247746 rows containing non-finite values (stat_count).
```



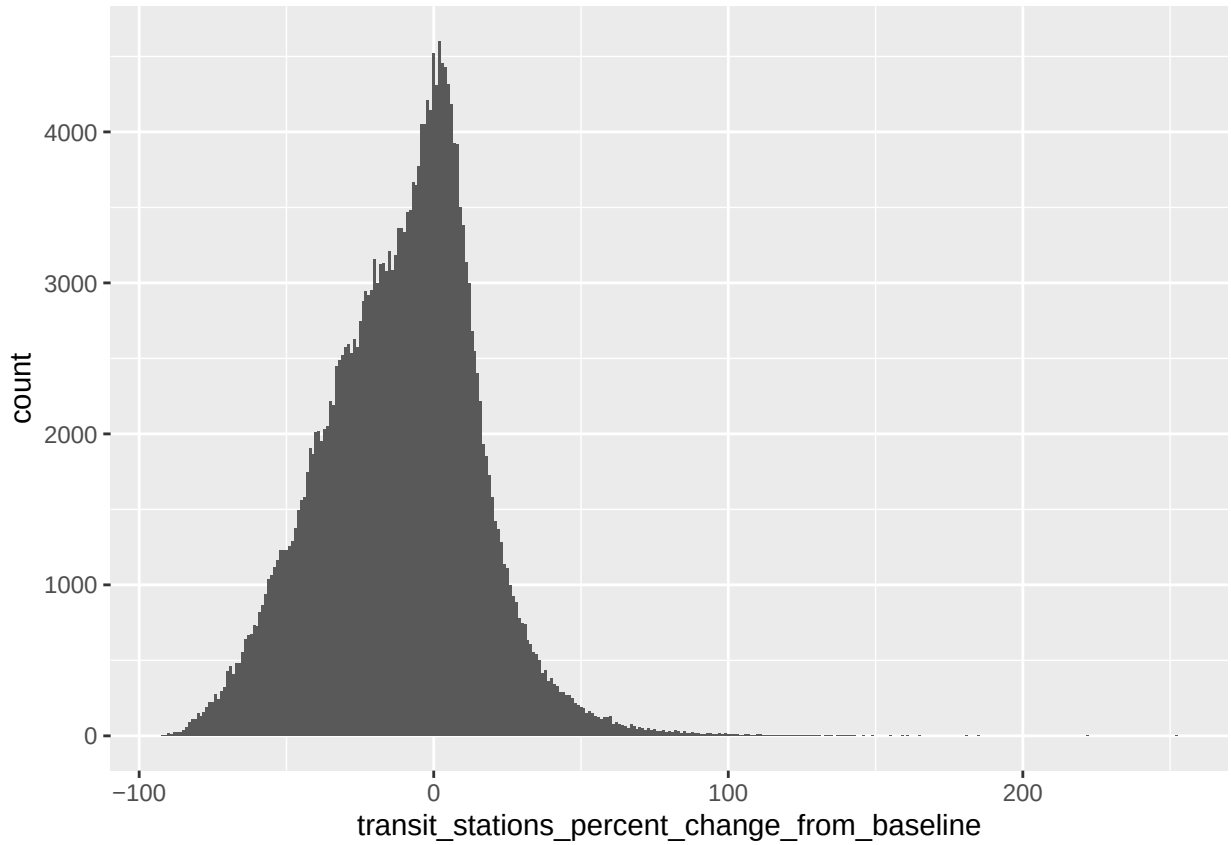

```
ggplot(data = mobility, mapping = aes(x = parks_percent_change_from_baseline)) +
  geom_bar()
```

```
## Warning: Removed 494324 rows containing non-finite values (stat_count).
```



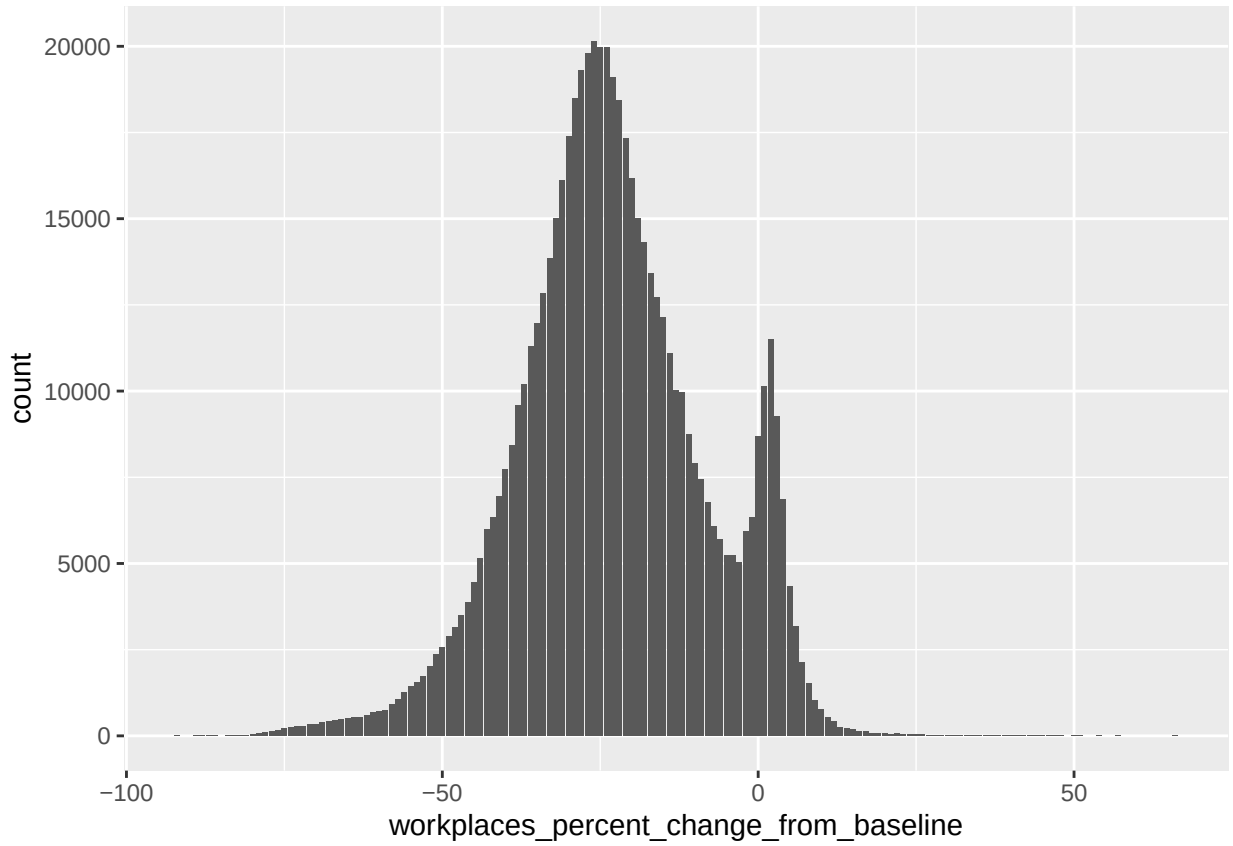

```
ggplot(data = mobility, mapping = aes(x = transit_stations_percent_change_from_baseline)) +
  geom_bar()
```

```
## Warning: Removed 399459 rows containing non-finite values (stat_count).
```



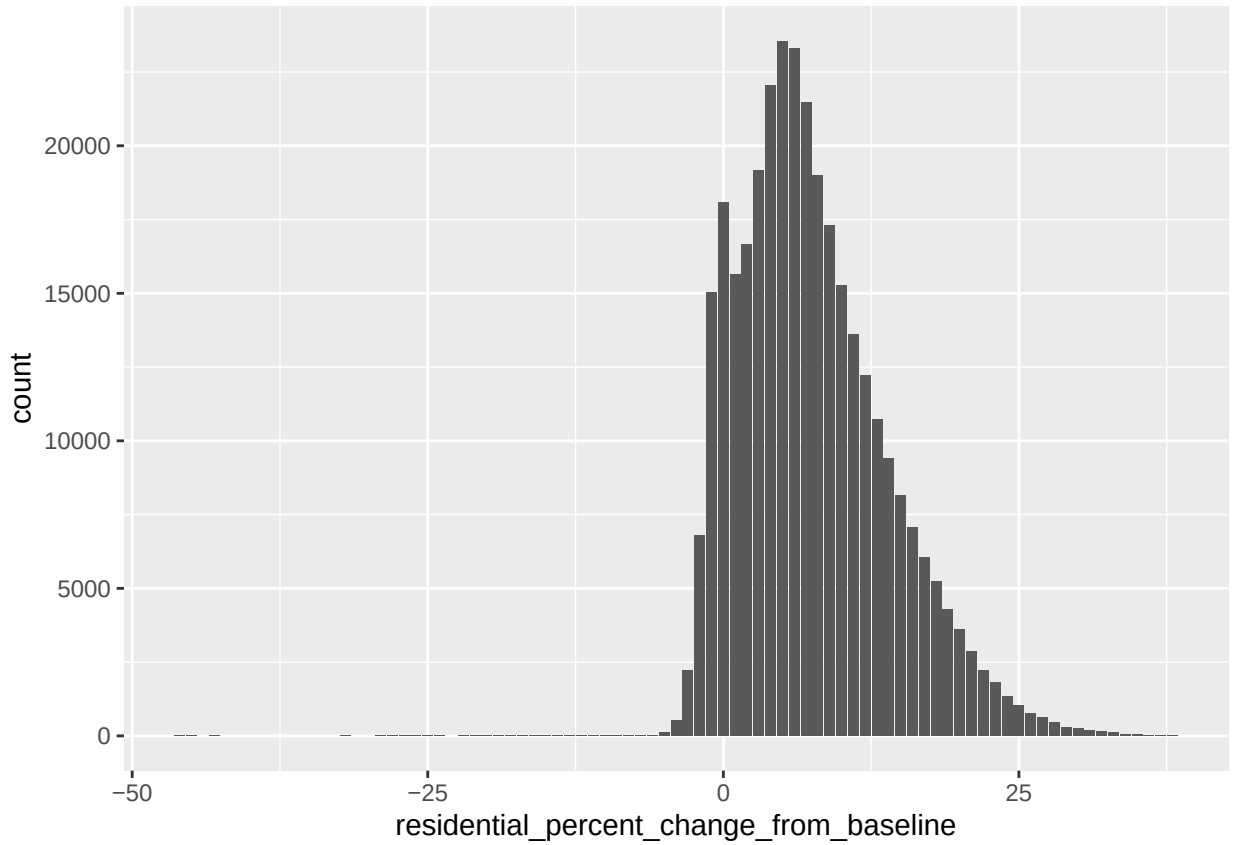

```
ggplot(data = mobility, mapping = aes(x = workplaces_percent_change_from_baseline)) +
  geom_bar()
```

```
## Warning: Removed 22327 rows containing non-finite values (stat_count).
```




```
ggplot(data = mobility, mapping = aes(x = residential_percent_change_from_baseline)) +
  geom_bar()
```

```
## Warning: Removed 312892 rows containing non-finite values (stat_count).
```




```
mobility %>%
  filter(!is.na(retail_and_recreation_percent_change_from_baseline)) %>%
  summarise(mean = mean(retail_and_recreation_percent_change_from_baseline),
            median = median(retail_and_recreation_percent_change_from_baseline),
            sd = sd(retail_and_recreation_percent_change_from_baseline),
            iqr = IQR(retail_and_recreation_percent_change_from_baseline),
            min = min(retail_and_recreation_percent_change_from_baseline),
            max = max(retail_and_recreation_percent_change_from_baseline))
```

```
##        mean median       sd iqr  min max
## 1 -9.100737     -7 20.51652  25 -100 261
```

```
mobility %>%
  filter(!is.na(grocery_and_pharmacy_percent_change_from_baseline)) %>%
  summarise(mean = mean(grocery_and_pharmacy_percent_change_from_baseline),
            median = median(grocery_and_pharmacy_percent_change_from_baseline),
            sd = sd(grocery_and_pharmacy_percent_change_from_baseline),
            iqr = IQR(grocery_and_pharmacy_percent_change_from_baseline),
            min = min(grocery_and_pharmacy_percent_change_from_baseline),
            max = max(grocery_and_pharmacy_percent_change_from_baseline))
```

```
##       mean median       sd iqr min max
## 1 2.047446      1 14.67781  15 -96 225
```

```
mobility %>%
  filter(!is.na(parks_percent_change_from_baseline)) %>%
  summarise(mean = mean(parks_percent_change_from_baseline),
            median = median(parks_percent_change_from_baseline),
```

```
          sd = sd(parks_percent_change_from_baseline),
          iqr = IQR(parks_percent_change_from_baseline),
          min = min(parks_percent_change_from_baseline),
          max = max(parks_percent_change_from_baseline))
```

```
##       mean median       sd iqr min max
## 1 37.33111     23 66.49339  76 -91 709
```

```
mobility %>%
  filter(!is.na(transit_stations_percent_change_from_baseline)) %>%
  summarise(mean = mean(transit_stations_percent_change_from_baseline),
          median = median(transit_stations_percent_change_from_baseline),
          sd = sd(transit_stations_percent_change_from_baseline),
          iqr = IQR(transit_stations_percent_change_from_baseline),
          min = min(transit_stations_percent_change_from_baseline),
          max = max(transit_stations_percent_change_from_baseline))
```

```
##        mean median       sd iqr min max
## 1 -10.91589     -9 26.18124  35 -92 252
```

```
mobility %>%
  filter(!is.na(workplaces_percent_change_from_baseline)) %>%
  summarise(mean = mean(workplaces_percent_change_from_baseline),
          median = median(workplaces_percent_change_from_baseline),
          sd = sd(workplaces_percent_change_from_baseline),
          iqr = IQR(workplaces_percent_change_from_baseline),
          min = min(workplaces_percent_change_from_baseline),
          max = max(workplaces_percent_change_from_baseline))
```

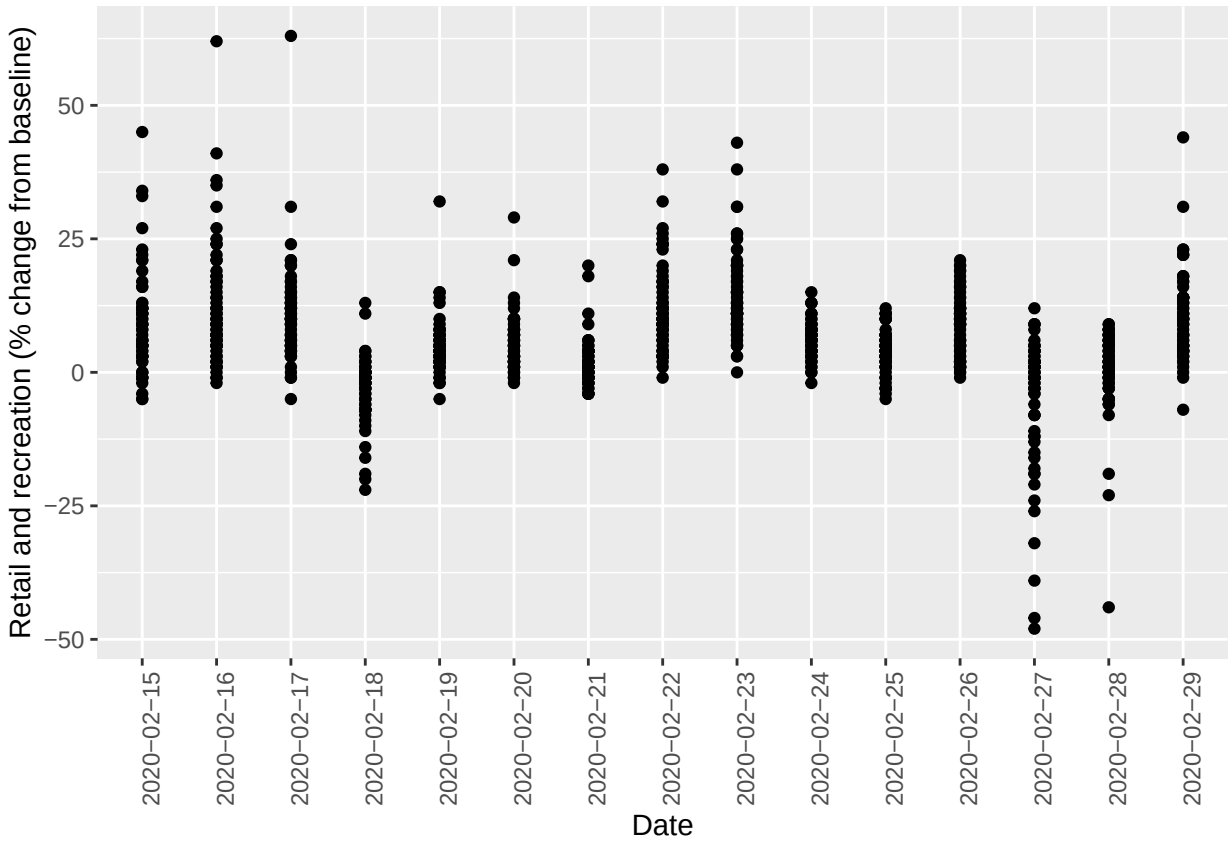
```
##       mean median       sd iqr min max
## 1 -23.0098    -24 14.96862  18 -92  66
```

```
mobility %>%
  filter(!is.na(residential_percent_change_from_baseline)) %>%
  summarise(mean = mean(residential_percent_change_from_baseline),
          median = median(residential_percent_change_from_baseline),
          sd = sd(residential_percent_change_from_baseline),
          iqr = IQR(residential_percent_change_from_baseline),
          min = min(residential_percent_change_from_baseline),
          max = max(residential_percent_change_from_baseline))
```

```
##       mean median       sd iqr min max
## 1 7.505763      7 6.300294   8 -46  38
```

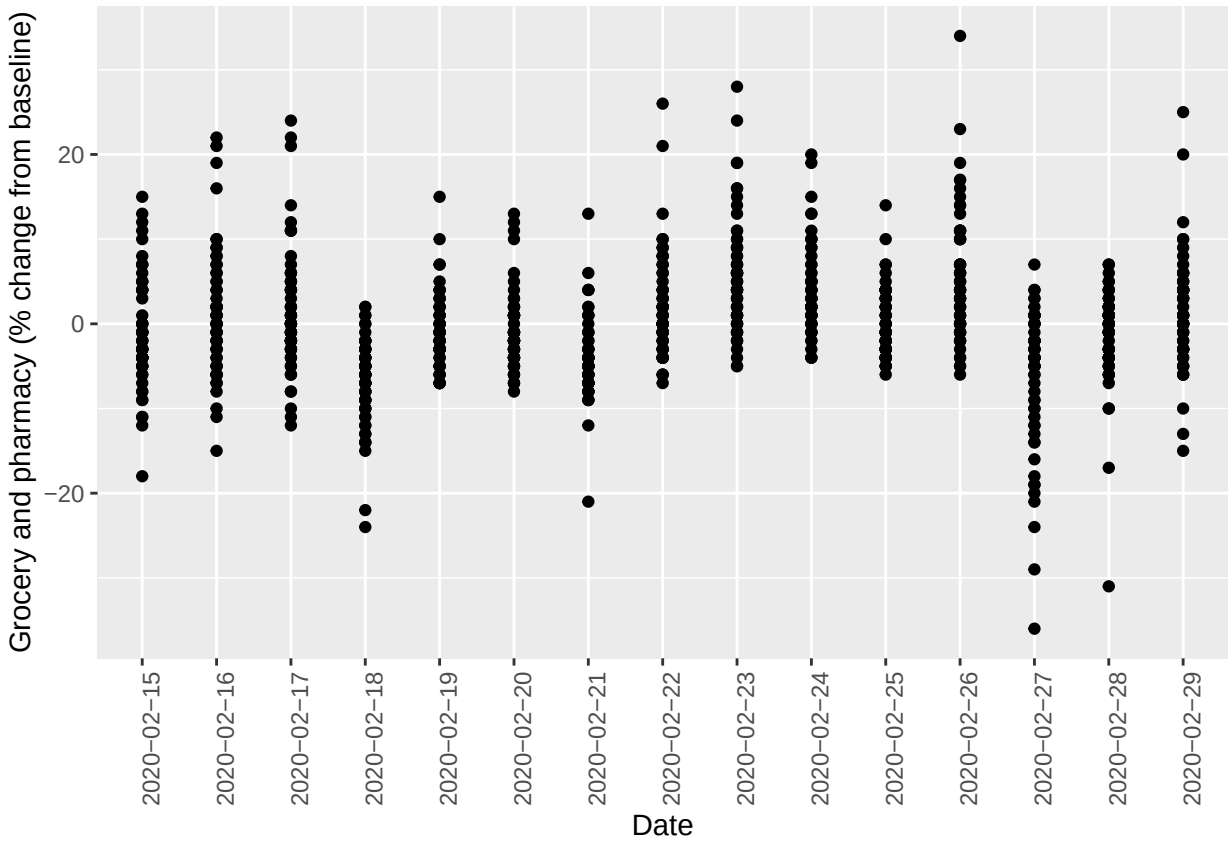
```
ny_feb <- mobility %>%
  filter(sub_region_1 == "New York", date_num >= as.Date("2020-02-15"), date_num < as.Date("2020-03-01"
```

```
ny_feb %>%
  ggplot(aes(x = date, y = retail_and_recreation_percent_change_from_baseline)) +
  geom_point() +
  labs(x = "Date", y = "Retail and recreation (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```



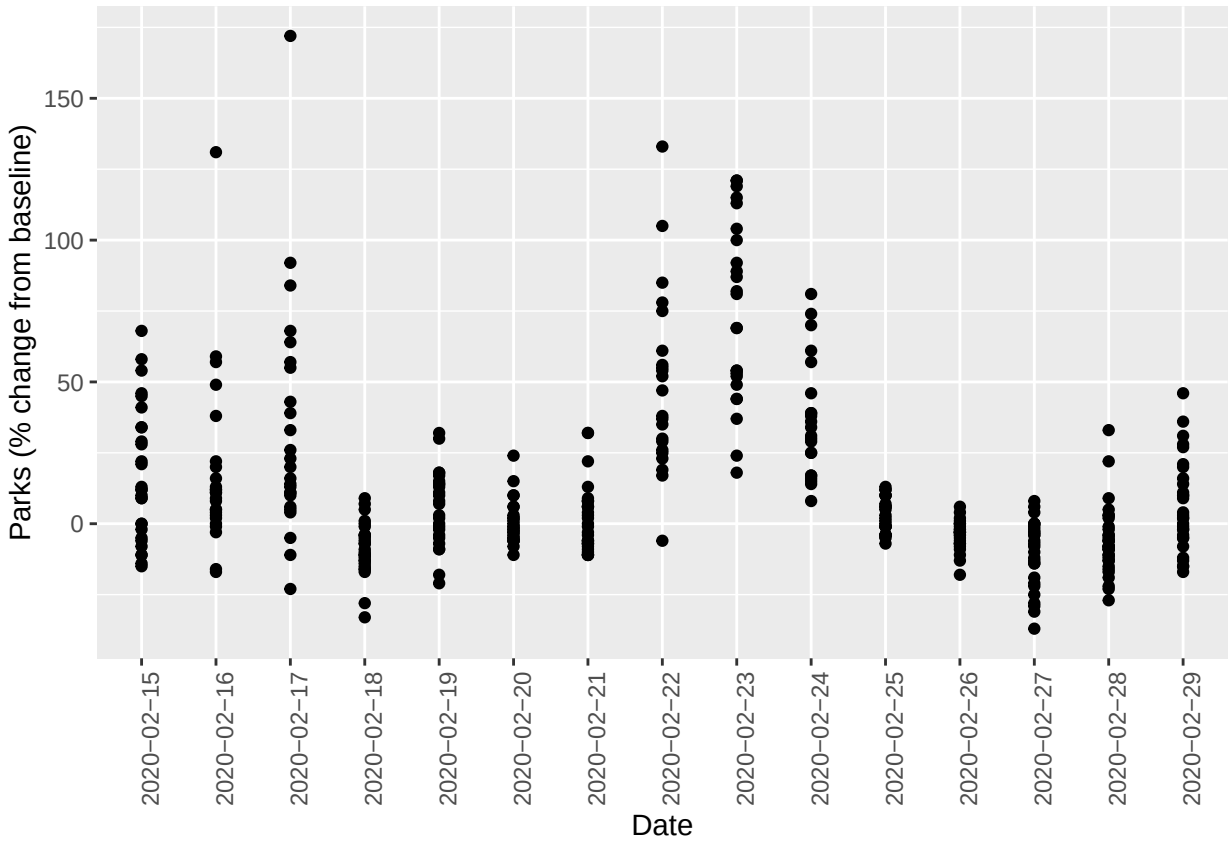

```
ny_feb %>%
  ggplot(aes(x = date, y = grocery_and_pharmacy_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Grocery and pharmacy (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



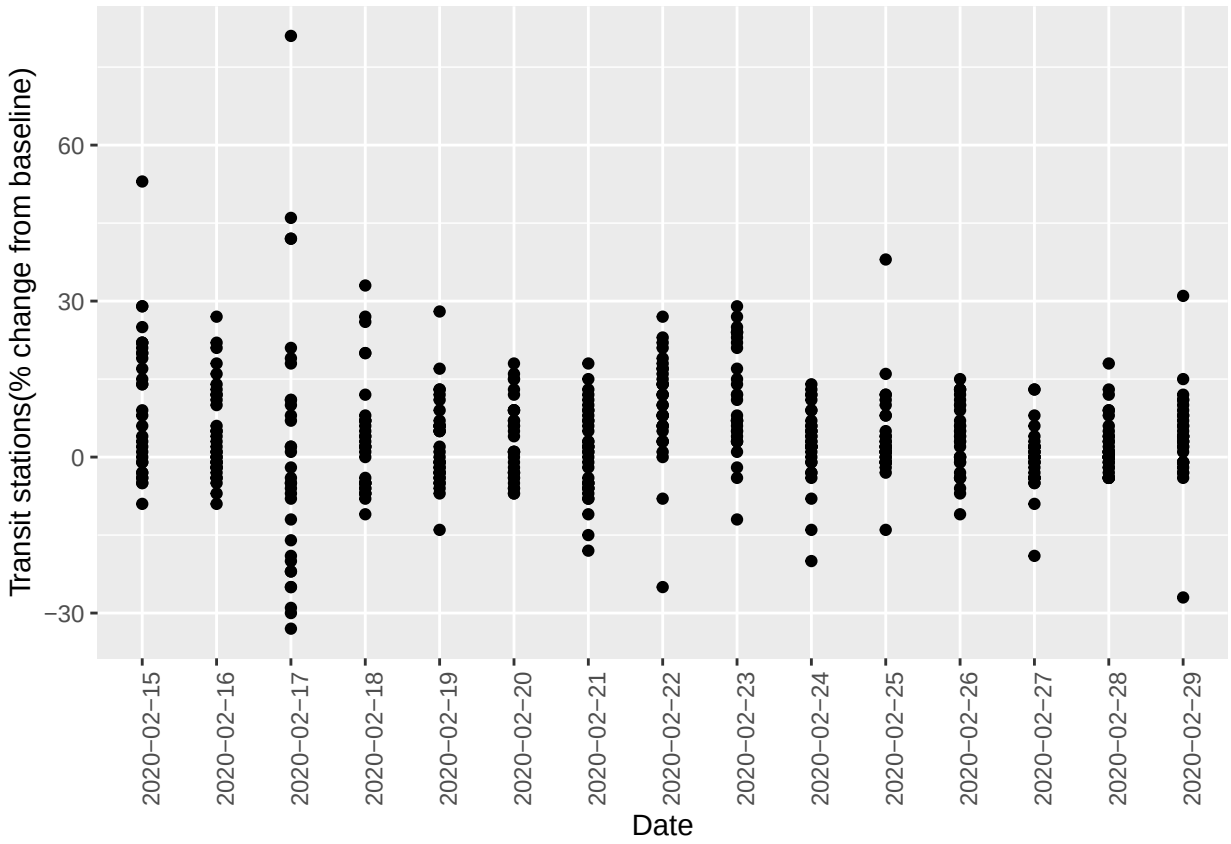

```
ny_feb %>%
  ggplot(aes(x = date, y = parks_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Parks (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```

```
## Warning: Removed 530 rows containing missing values (geom_point).
```

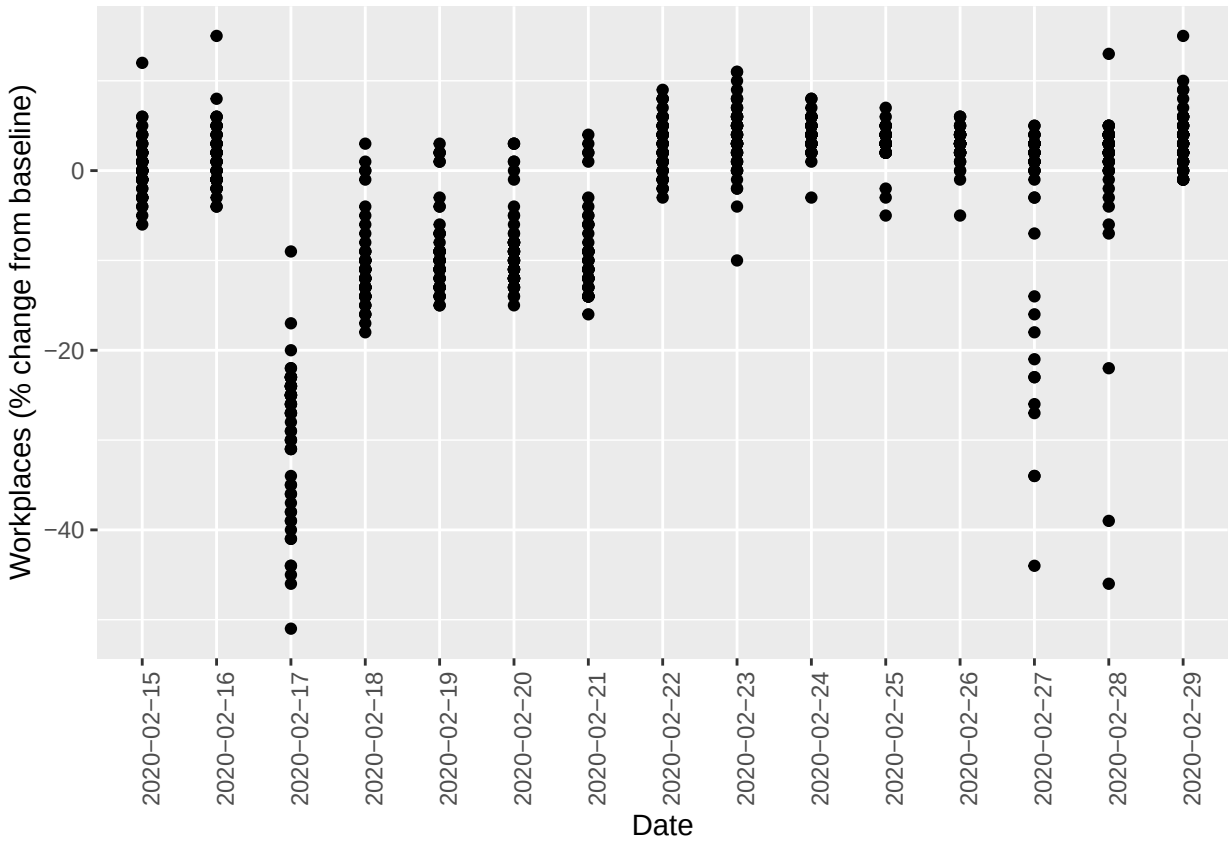



```
ny_feb %>%
  ggplot(aes(x = date, y = transit_stations_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Transit stations(% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```

```
## Warning: Removed 435 rows containing missing values (geom_point).
```

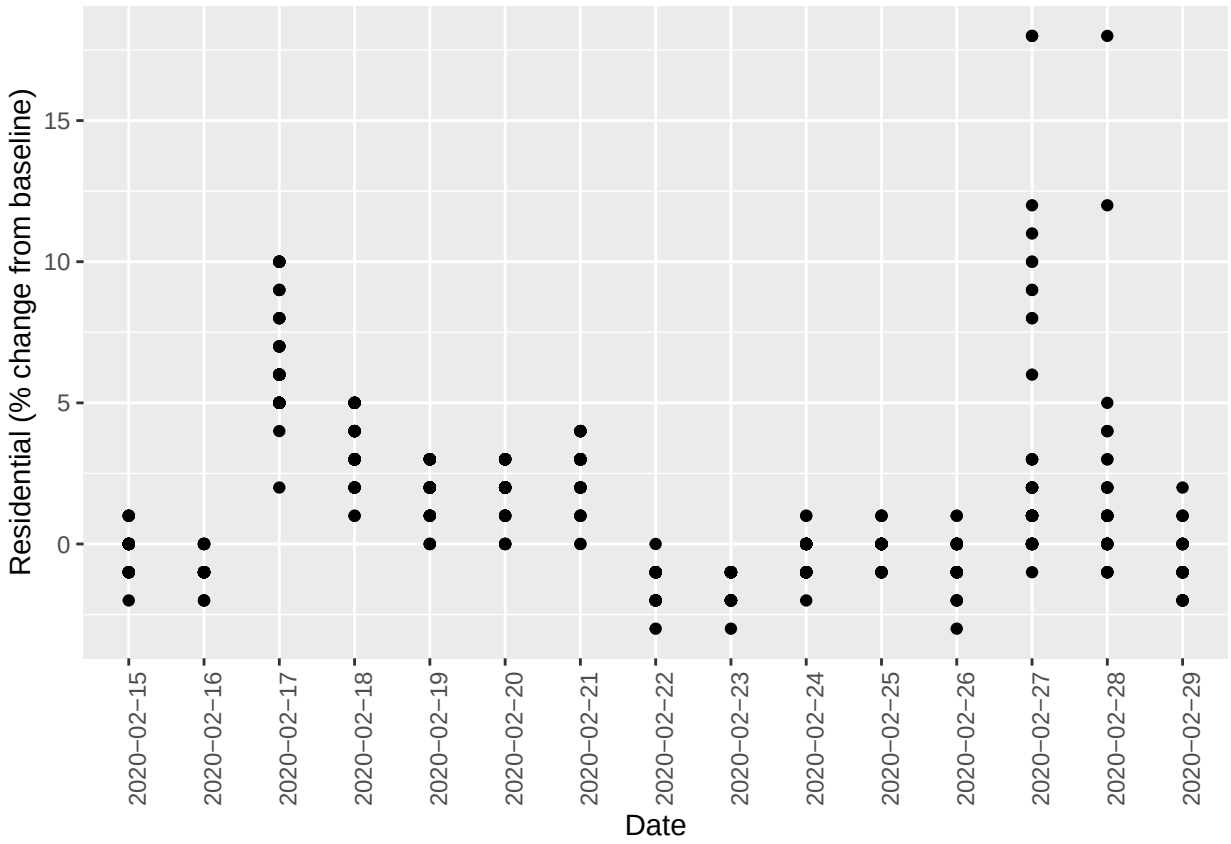



```
ny_feb %>%
  ggplot(aes(x = date, y = workplaces_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Workplaces (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```




```
ny_feb %>%
  ggplot(aes(x = date, y = residential_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Residential (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```

```
## Warning: Removed 159 rows containing missing values (geom_point).
```




```
durham <- mobility %>%
  filter(sub_region_2 == "Durham County")

durham %>%
  head(50) %>%
  ggplot(aes(x = date, y = workplaces_percent_change_from_baseline)) +
  geom_point() #+
  #theme_ipsum_rc()
```

```
nc_avg <- mobility %>%
  filter(sub_region_1 == "North Carolina") %>%
  group_by(sub_region_2, na.rm=TRUE) %>%
  summarise(retail_avg = mean(retail_and_recreation_percent_change_from_baseline, na.rm=TRUE),
            grocery_avg = mean(grocery_and_pharmacy_percent_change_from_baseline, na.rm=TRUE),
            parks_avg = mean(parks_percent_change_from_baseline, na.rm=TRUE),
            transit_avg = mean(transit_stations_percent_change_from_baseline, na.rm=TRUE),
            workplaces_avg = mean(workplaces_percent_change_from_baseline, na.rm=TRUE),
            residential_avg = mean(residential_percent_change_from_baseline, na.rm=TRUE))
```

```
## `summarise()` regrouping output by 'sub_region_2' (override with `.groups` argument)
```

```
nc_avg
```

```
## # A tibble: 99 x 8
## # Groups:   sub_region_2 [99]
##    sub_region_2 na.rm retail_avg grocery_avg parks_avg transit_avg
##    <chr>        <lgl>      <dbl>       <dbl>     <dbl>       <dbl>
##  1 ""           TRUE       -14.9       -1.01      45.4       -24.5
##  2 "Alamance C~ TRUE       -13.0        0.648     32.6        -2.04
```

```
##  3 "Alexander ~ TRUE       -7.44      11.7       NaN        NaN
##  4 "Alleghany ~ TRUE      -12.5       -9.75      NaN        NaN
##  5 "Anson Coun~ TRUE       -7.99      10.5       NaN        NaN
##  6 "Ashe Count~ TRUE        4.41       8.2       NaN        NaN
##  7 "Avery Coun~ TRUE      -12.6       13.1       NaN        NaN
##  8 "Beaufort C~ TRUE       -8.22       3.65       17.8      NaN
##  9 "Bertie Cou~ TRUE       18.8        7.02      NaN        NaN
## 10 "Bladen Cou~ TRUE       -3.61       5.86      -24.7      NaN
## # ... with 89 more rows, and 2 more variables: workplaces_avg <dbl>,
## #   residential_avg <dbl>
```

```
#test_plot <- ggplot(mtcars, aes(mpg, wt)) +
  #geom_point() +
  #labs(x="Fuel efficiency (mpg)", y="Weight (tons)",
       #title="Seminal ggplot2 scatterplot example",
       #subtitle="A plot that is only useful for demonstration purposes",
       #caption="Brought to you by the letter 'g'") +
  #theme_ipsum_rc()
```

```
#ggsave("../output/test-plot.png", test_plot, units = "in", width = 850/100, height = 600/100 )
```

EDA

Linear Regression Code