Google Movement Data Analysis

Albert Sun, Lily Zhu

```r
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()
```

```r
#library(hrbrthemes)
```

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

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

```r
ny_feb <- mobility %>%
  filter(sub_region_1 == "New York", date_num >= as.Date("2020-02-15"), date_num < as.Date("2020-03-01")
```
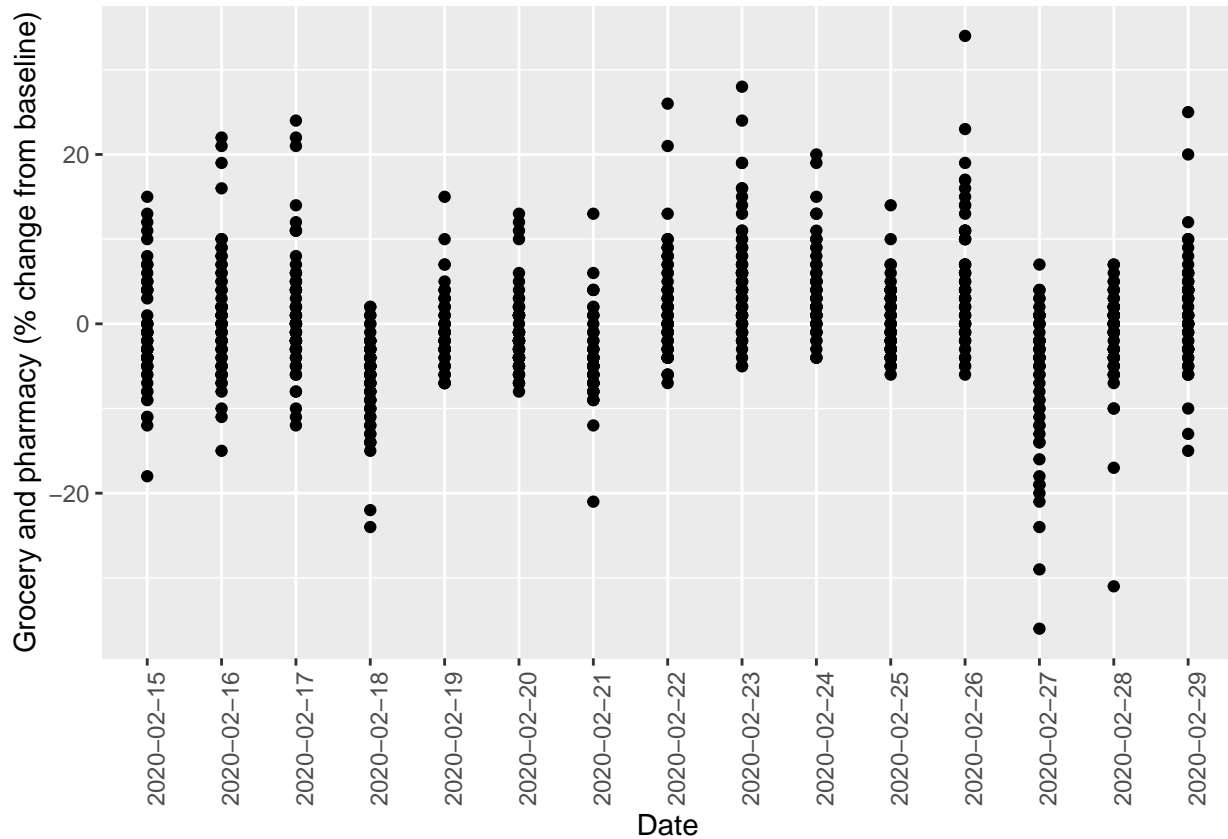
```r
ny_feb %>%
  ggplot(aes(x = date, y = retail_and_recreation_percent_change_from_baseline)) +
  geom_point() +
  labs(x = "Date", y = "Retail and recreation (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```
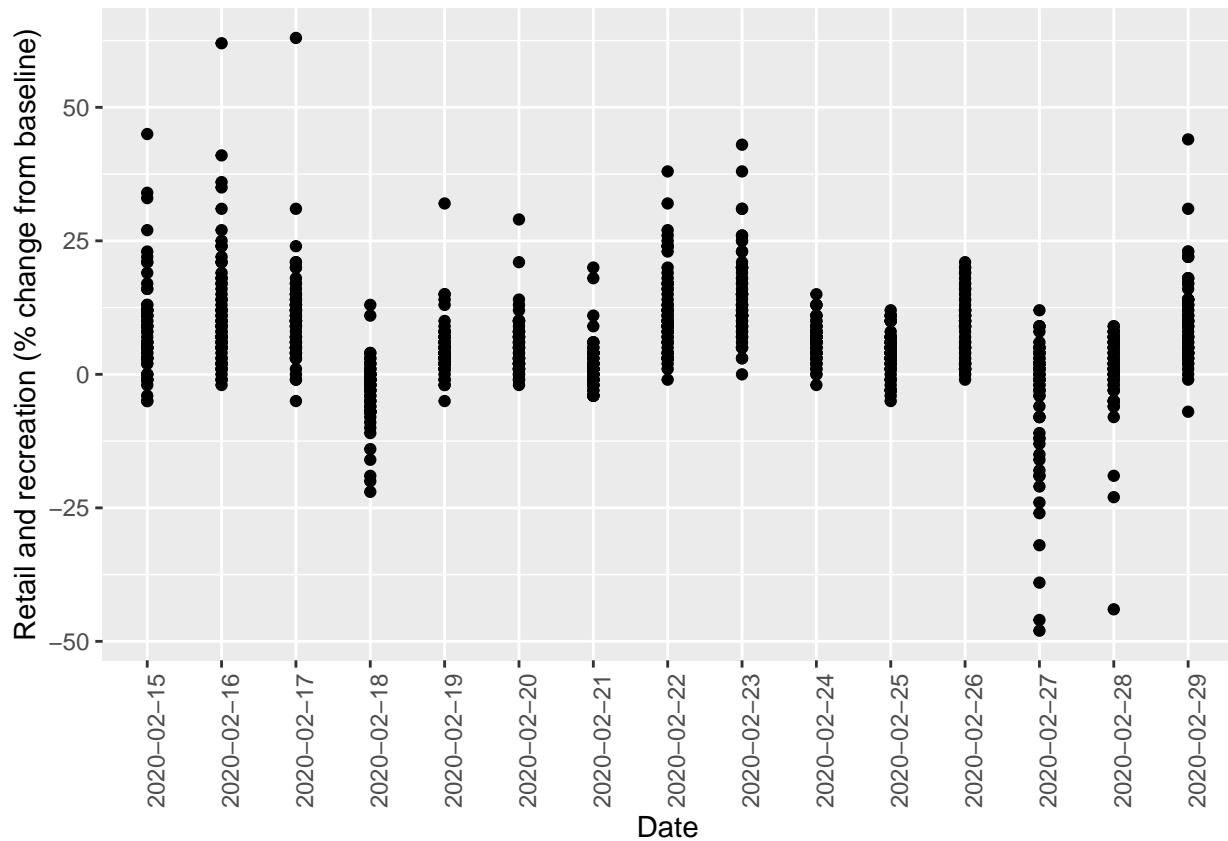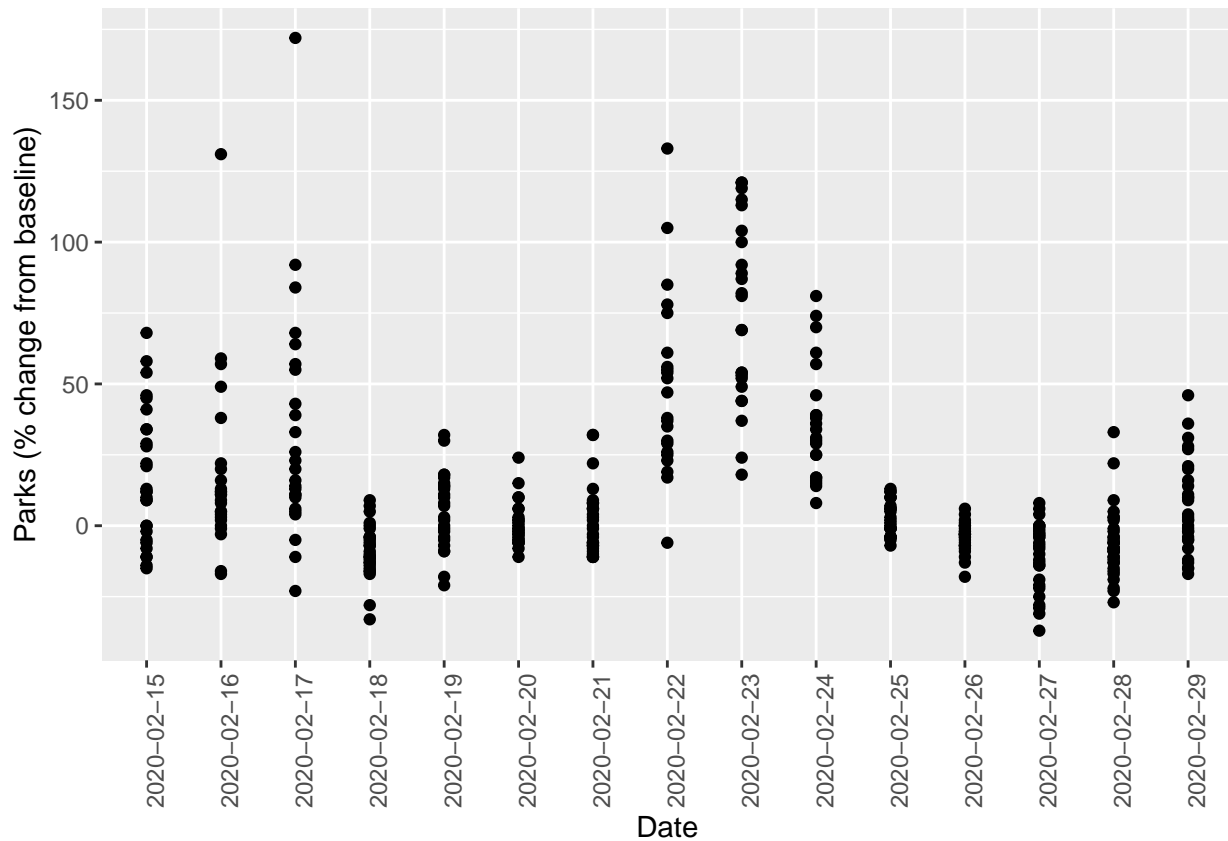
```
ny_feb %>%
  ggplot(aes(x = date, y = grocery_and_pharmacy_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Grocery and pharmacy (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```

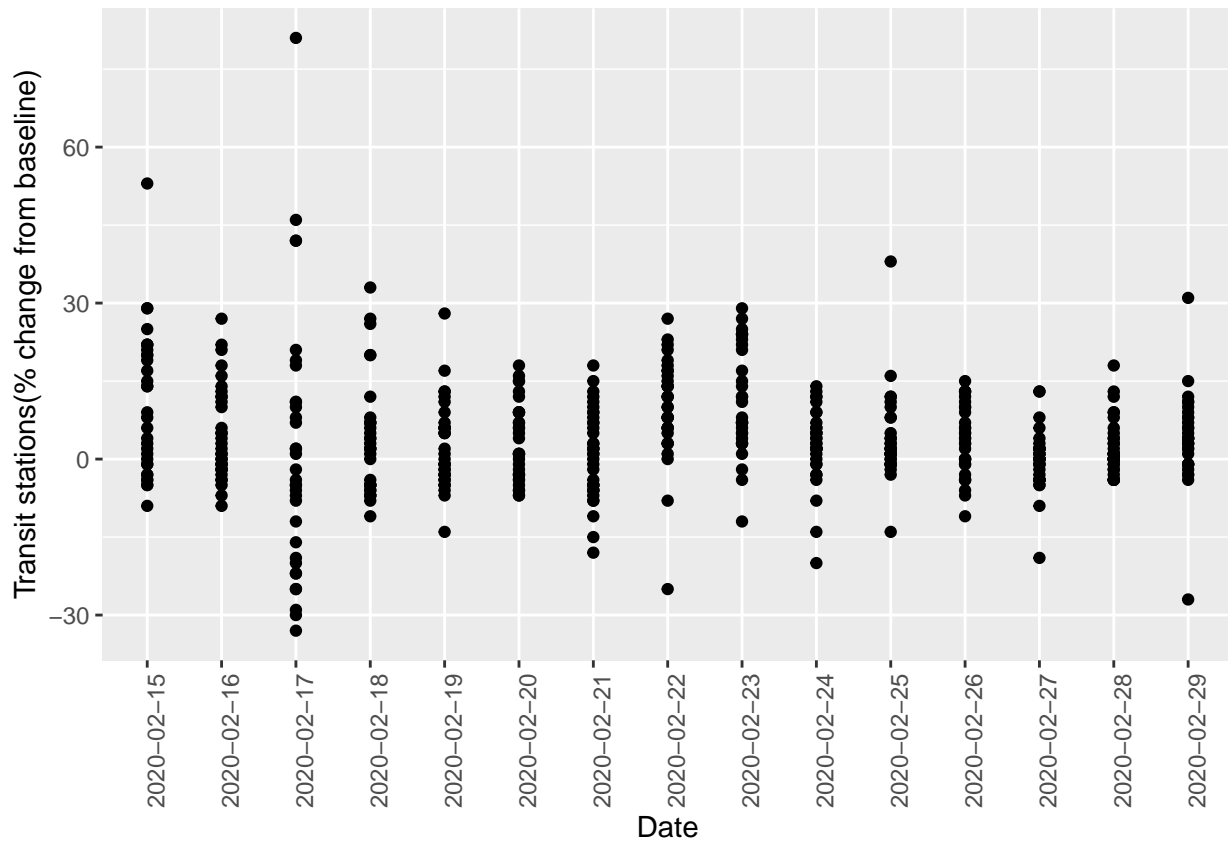## Warning: Removed 1 rows containing missing values (geom_point).

```
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))
```

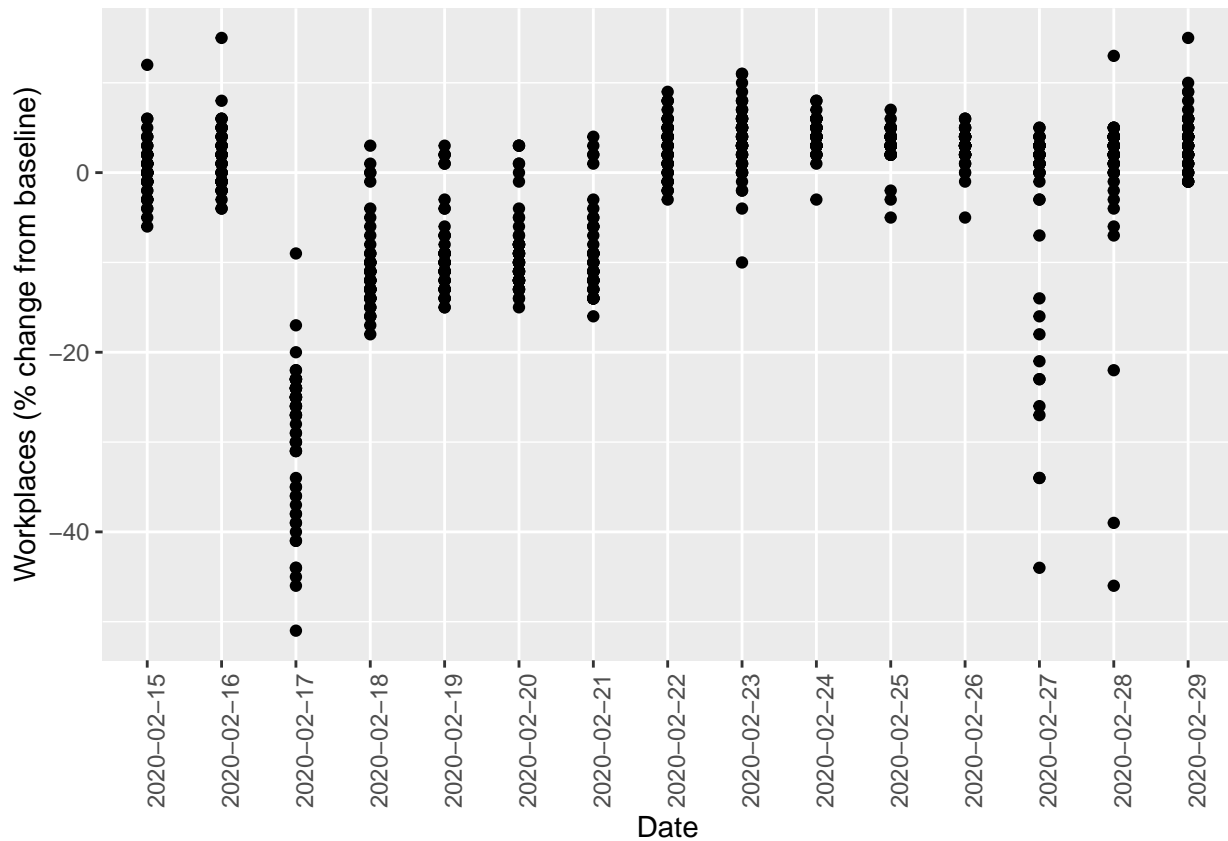## Warning: Removed 530 rows containing missing values (geom_point).

```
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))
```

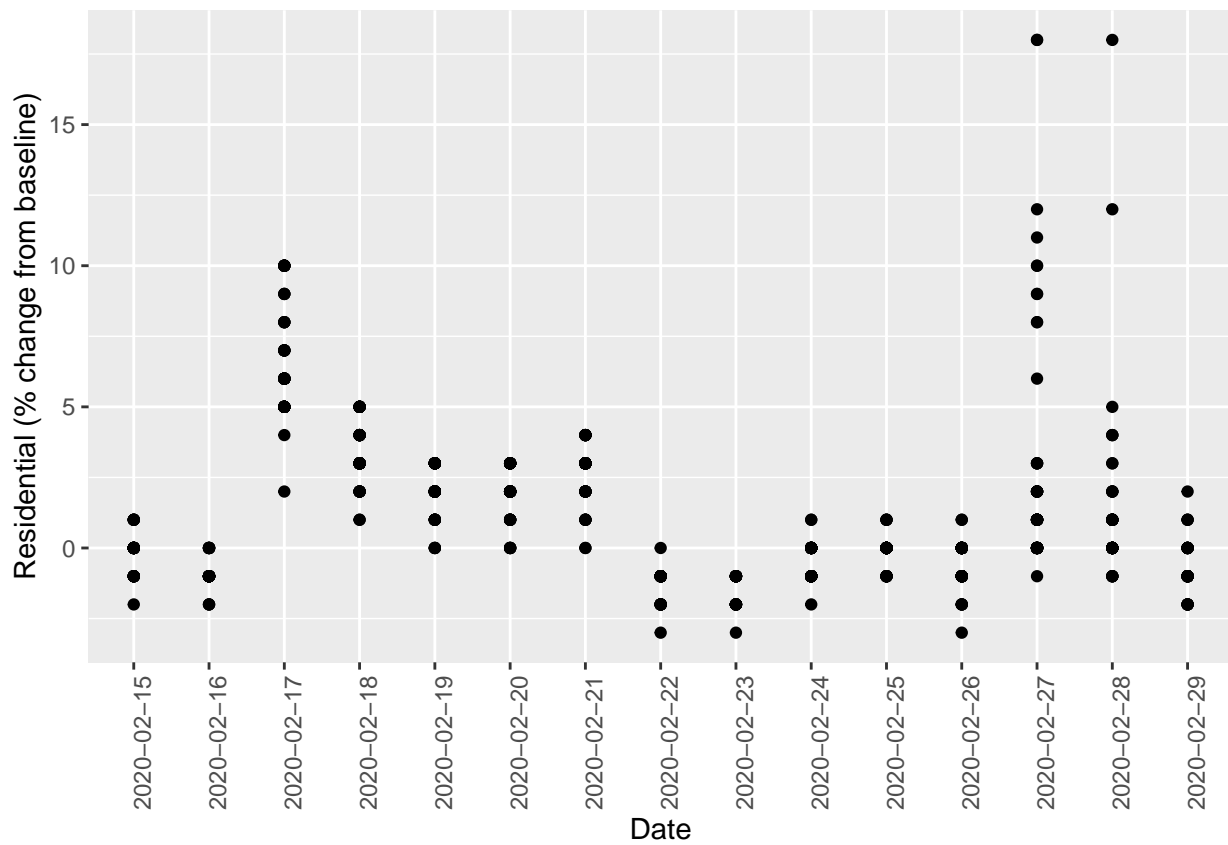## Warning: Removed 435 rows containing missing values (geom_point).

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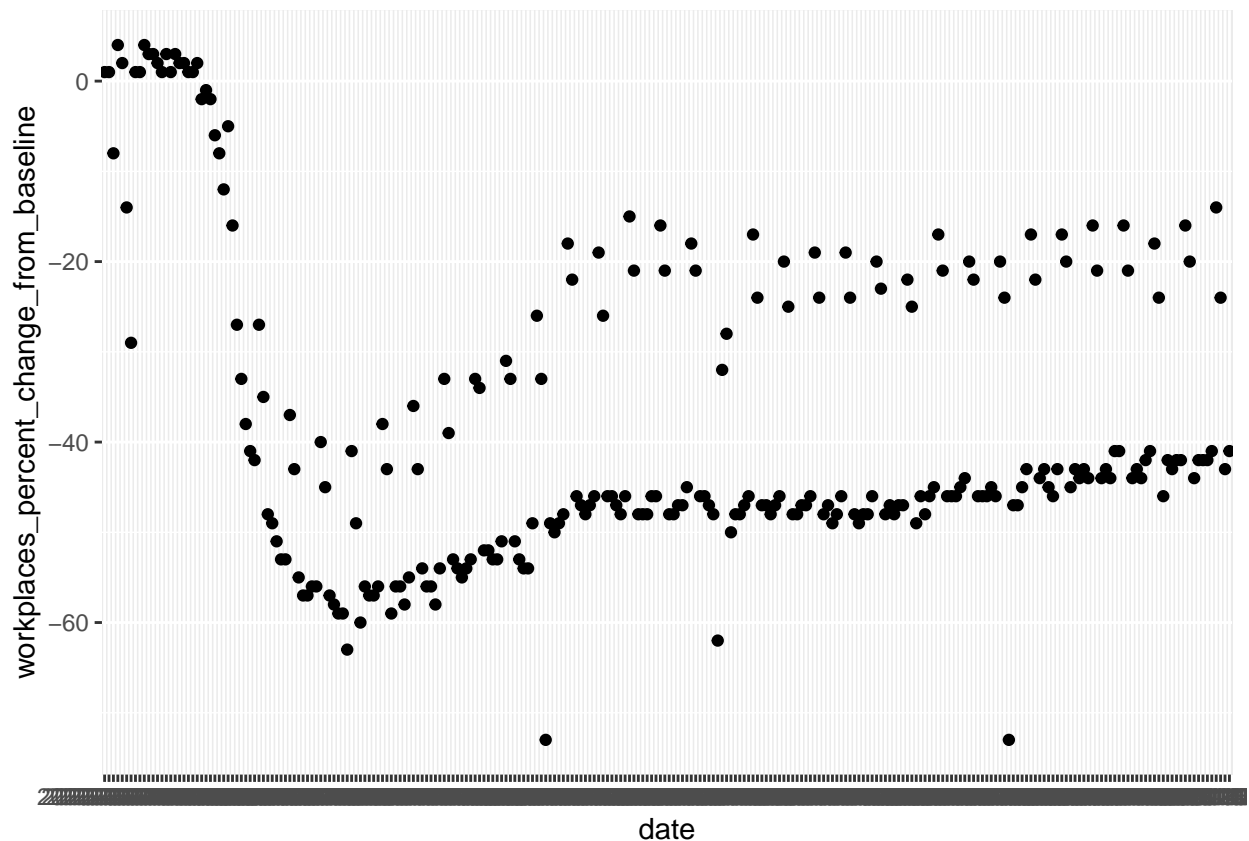
```
ny_feb %>%
  ggplot(aes(x = date, y = residential_percent_change_from_baseline)) +
  geom_point() +
   labs(x = "Date", y = "Residential (% change from baseline)") +
  theme(axis.text.x = element_text(angle = 90))
```

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

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

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

```
durham %>%
  filter(date == "2020-09-11") %>%
  select(workplaces_percent_change_from_baseline)
```

```
##   workplaces_percent_change_from_baseline
## 1                                     -43
```

```
#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