DataFest 2020: COVID-19 Analysis

Alyssa Shi and Drew Flanagan

2020-04-22

```
library(tidyverse)
library(broom)
library(knitr)
library(dplyr)
library(lubridate)
```

```
mobility <- read_csv("Data/Global_Mobility_Report.csv")
covid19 <- read_csv("Data/us_states_covid19_daily.csv")
covidus <- read_csv("Data/us_covid19_daily.csv")
states <- read_csv("Data/StatesFIPSCodes.csv")
```

Muating states to be able to join other datasets:

```
states <- states %>%
  select(STUSAB, STATE_NAME) %>%
  mutate(state = STUSAB,
         sub_region_1 = STATE_NAME) %>%
  select(state, sub_region_1)
```

Joining:

```
covid19 <- left_join(covid19, states)
```

Joining USA data with state data:

```
covidus <- covidus %>%
  mutate(state = "USA") %>%
  select(date, state, positive)

COVID <- full_join(covid19, covidus) %>%
  mutate(date = ymd(date),
         sub_region_1 = if_else(is.na(sub_region_1), "USA", sub_region_1)) %>%
  select(date, state, positive, negative, sub_region_1)
```

Altering mobility, imputing 0 for missing values:

```
mobility <- mobility %>%
  filter(country_region_code == "US") %>%
  mutate(date = ymd(date),
         sub_region_1 = if_else(is.na(sub_region_1), "USA", sub_region_1),
         retail_rec = if_else(is.na(retail_and_recreation_percent_change_from_baseline), 0, retail_and_
         groc_pharm = if_else(is.na(grocery_and_pharmacy_percent_change_from_baseline), 0, grocery_and_
         parks = if_else(is.na(parks_percent_change_from_baseline), 0, parks_percent_change_from_baseli
         transit = if_else(is.na(transit_stations_percent_change_from_baseline), 0, transit_stations_pe
         workplace = if_else(is.na(workplaces_percent_change_from_baseline), 0, workplaces_percent_chan
         residential = if_else(is.na(residential_percent_change_from_baseline), 0, residential_percent_
  select(sub_region_1, date, retail_rec, groc_pharm, parks, transit, workplace, residential)
```

Aggregating mobility by region and date to find average mobility:

```
mobility2 <- aggregate(mobility, by = list(mobility$sub_region_1, mobility$date), FUN = mean)

mobility2 <- mobility2 %>%
  mutate(date = Group.2,
         sub_region_1 = Group.1) %>%
  select(sub_region_1, date, retail_rec, groc_pharm, parks, transit, workplace, residential)
```

Joining datasets:

```
covid19_mob <- inner_join(COVID, mobility2)
```

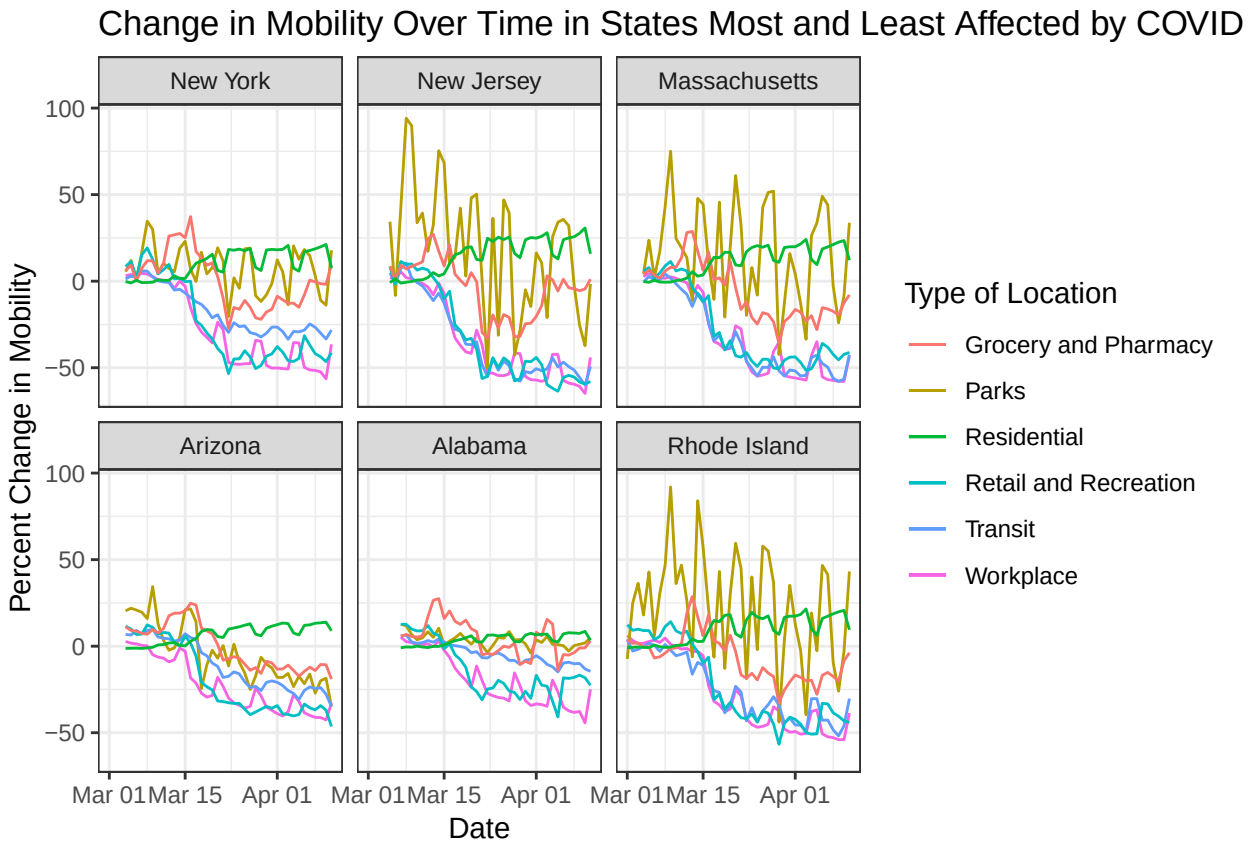
```
glimpse(covid19_mob)
```

```
## Observations: 1,969
## Variables: 11
## $ date         <date> 2020-04-11, 2020-04-11, 2020-04-11, 2020-04-11, ...
## $ state        <chr> "AK", "AL", "AR", "AZ", "CA", "CO", "CT", "DC", "...
## $ positive     <dbl> 257, 3191, 1226, 3393, 19472, 6510, 11510, 1778, ...
## $ negative     <dbl> 7475, 18058, 17352, 37137, 152604, 26143, 28321, ...
## $ sub_region_1 <chr> "Alaska", "Alabama", "Arkansas", "Arizona", "Cali...
## $ retail_rec   <dbl> -31.33333, -22.61404, -15.50000, -46.40000, -40.7...
## $ groc_pharm   <dbl> -4.6666667, 3.1578947, 0.0800000, -18.9333333, -9...
## $ parks        <dbl> 10.833333, 4.894737, -1.920000, -34.933333, -30.9...
## $ transit      <dbl> -39.00000, -14.54386, -12.38000, -34.80000, -33.5...
## $ workplace    <dbl> -32.16667, -24.91228, -22.28000, -32.86667, -36.6...
## $ residential  <dbl> 5.333333, 3.350877, 1.940000, 8.933333, 10.836364...
```

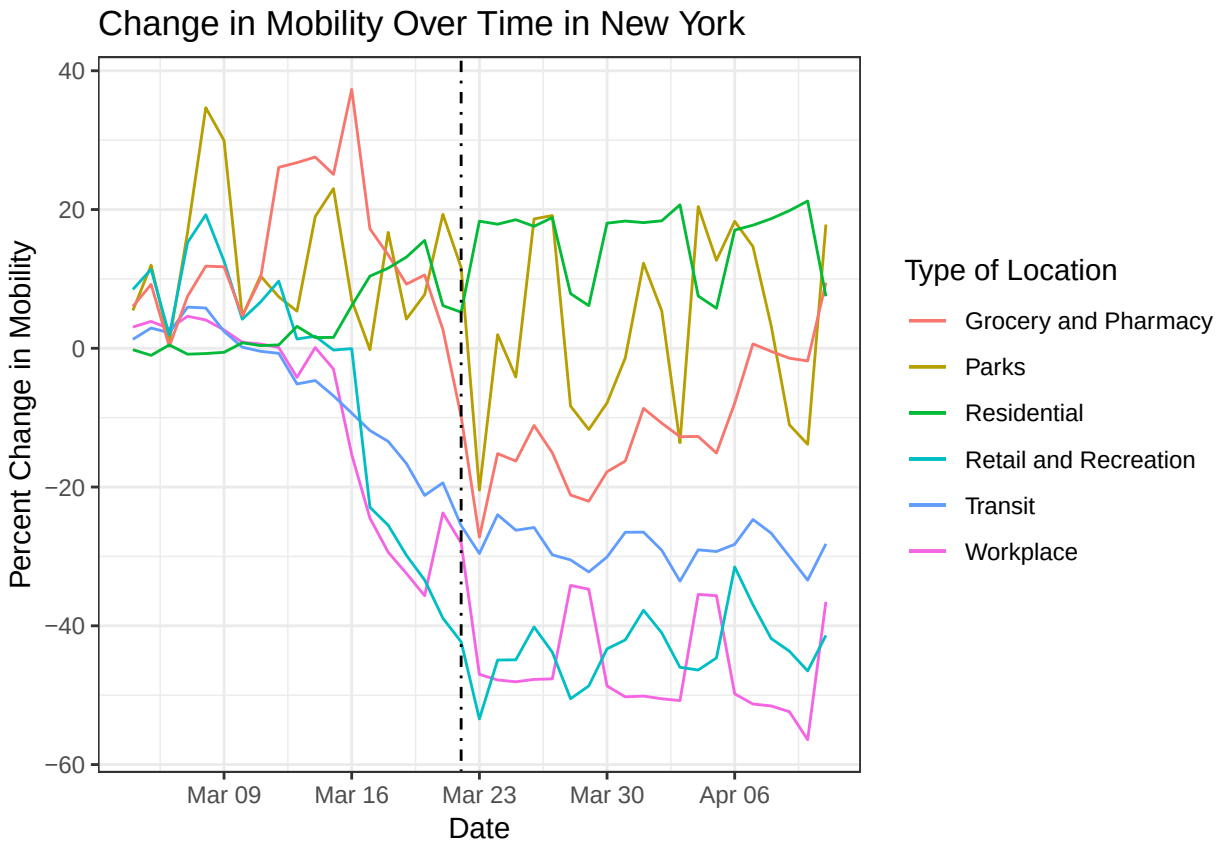
Introduction

```
covid19_mob <- covid19_mob %>%
  filter(state == "NY" | state == "NJ" | state == "MA" | state == "AZ" | state == "AL"| state == "RI") %
  mutate(sub_region_1 = factor(sub_region_1, levels = c("New York", "New Jersey", "Massachusetts", "Ari
```

```
covid19_mob %>%
  ggplot(mapping = aes(x = date)) +
  geom_line(aes(y = workplace, color = "Workplace")) +
  geom_line(aes(y = parks, color = "Parks")) +
  geom_line(aes(y = transit, color = "Transit")) +
  geom_line(aes(y = retail_rec, color = "Retail and Recreation")) +
  geom_line(aes(y = groc_pharm, color = "Grocery and Pharmacy")) +
  geom_line(aes(y = residential, color = "Residential")) +
  facet_wrap(vars(sub_region_1)) +
  labs(title = "Change in Mobility Over Time in States Most and Least Affected by COVID-19",
       x = "Date",
       y = "Percent Change in Mobility",
       color = "Type of Location") +
  theme_bw()
```




```
covid19_mob %>%
  filter(state == "NY") %>%
  ggplot(mapping = aes(x = date)) +
  geom_line(aes(y = workplace, color = "Workplace")) +
  geom_line(aes(y = parks, color = "Parks")) +
  geom_line(aes(y = transit, color = "Transit")) +
  geom_line(aes(y = retail_rec, color = "Retail and Recreation")) +
  geom_line(aes(y = groc_pharm, color = "Grocery and Pharmacy")) +
  geom_line(aes(y = residential, color = "Residential")) +
  labs(title = "Change in Mobility Over Time in New York",
       x = "Date",
       y = "Percent Change in Mobility",
       color = "Type of Location") +
  geom_vline(xintercept = as.numeric(ymd("2020-03-22")), linetype = 4) +
  theme_bw()
```



parks not decreasing same as others? discuss whether people are still getting infected going to parks despite stay at home order