DataFest 2020: COVID-19 Analysis

Alyssa Shi and Drew Flanagan

2020-04-21

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

```r
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:

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

Joining:

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

Joining USA data with state data:

```r
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:

```r
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_chang
         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)
```

Introduction