Group Project Check-in

submitted by Radhika Kaul, Odiche Nwabuikwu, & Ruochen Wang

Question of Interest

We will develop different models to predict the number of unemployment claims filed in each US state in the second week of May (when the project is due). Our primary data source for the outcome variable is the Unemployment Insurance Weekly Claims data from the U.S. Department of Labor (DOL)'s Employment & Training Administration.

Project Progress

We have decided on the following variables for our models:

Outcome Variables: either New Initial Claims filed per week OR percentage of the labor force filing for unemployment claims per week. New Initial Claims is one of the most sensitive and most frequently used official statistics in analyzing unemployment in the labor market (see Aaron Soujourner and Paul Goldsmith Pinkham).

Predictors *(all variables are at the state level)*

Structural predictors include:

- Region
- Population
- Working Population
- Per Capita Personal Income
- Percent Population w/ a High School Degree
- Top Industry in terms of GDP Contribution
- Democratic or Republican Governor

We are considering standardizing/normalizing/logging the populations variables due to concerns of high correlations between these variables and the outcome variable.

Real-Time Predictors include

- Number of COVID Cases
- Number of Days Since the issurance of Stay-at-Home Order
- Number of Days Before the Stay-at-Home Order ends
- Peak of Claims Predicted (we haven't found a specific data source but will keep looking)
- Whether there has been a protest/protests in the state

Algorithms: CART/RANDOM FOREST. We will have three predictive models based off of data from when the first case was reported in the US and when the first death was reported in the US.

Error Metric: RMSE (we'll have to explain why?)

Data Cleaning

```r
library(tidyverse)
```

```r
topindustry_bystate <-
  read.csv("data_raw/gdp_byindustry_bystate_2019.csv")

topindustry_bystate <- topindustry_bystate %>%
  filter(str_detect(string = Description,
                    pattern = "    ")) %>%
  filter(!str_detect(string = Description,
                     pattern = "      "))
```