St. Louis Crime Report

Ed Gonzalez

I wanted to demonstrate my capabilities in R by working on a project that is based on crime data in St. Louis from 2020. The following code chunks show the process of loading in the CSV files from a directory on my computer.

```
January2020 <- read_csv("Downloads/January2020.CSV")
February2020 <- read_csv("Downloads/February2020.CSV")
March2020 <- read_csv("Downloads/March2020.CSV")
April2020 <- read_csv("Downloads/April2020.CSV")
May2020 <- read_csv("Downloads/May2020.CSV")
June2020 <- read_csv("Downloads/June2020.CSV")
July2020 <- read_csv("Downloads/July2020.CSV")
August2020 <- read_csv("Downloads/August2020.CSV")
September2020 <- read_csv("Downloads/September2020.CSV")
October2020 <- read_csv("Downloads/October2020.CSV")
November2020 <- read_csv("Downloads/November2020.CSV")
December2020 <- read_csv("Downloads/December2020.CSV")
```

I ran into some errors trying to combine the datasets into a single data frame, so I did some troubleshooting and found that some of the columns were classified as numeric when the others were numeric. I fixed that with the following code chunk. Although the addresses aren't required for plotting purposes, it was causing a hang-up and they're a good reference for further analyzing.

```
November2020$ILEADSAddress <- as.numeric(November2020$ILEADSAddress)
November2020$CADAddress <- as.numeric(November2020$CADAddress)

December2020$CADAddress <- as.numeric(December2020$CADAddress)

January2020$x <- January2020$XCoord
January2020$y <- January2020$YCoord
```

Now I merged them into a single CSV file

```
st.lou.2020 <- bind_rows(January2020, February2020, March2020,
                         April2020, May2020, June2020,
                         July2020, August2020, September2020,
                         October2020, November2020, December2020)
```

At this point, I reached out to my surveyor friend who helped me convert the units in XCoord and YCoord from meters to Lat and Lon. dfjan2020 is the new file that I use to test my modeling.

```
dfjan2020 <- read.csv("/Users/ed/Downloads/EG - Coordinates.csv")
dfjan2020 <- dfjan2020 %>% filter(row_number() %% 2 == 0)
```

```
library(ggplot2)
library(maps)
library(mapdata)

df_sub <- dfjan2020 %>% select(Longitude.DD., Latitude.DD., Crime)
df_sub <- na.omit(df_sub)
```

```
mo_map <- map_data("state", region = "missouri")

ggplot() +
  geom_polygon(data = mo_map, aes(x = long, y = lat, group = group),
               fill = "gray90", color = "gray40") +
  geom_point(data = df_sub, aes(x = Longitude.DD.,
                                y = Latitude.DD.,
                                color = Crime),
             alpha = 0.5, size = 0.5) +
  scale_color_gradient(low = "green", high = "red", name = "Crime") +
  labs(title = "Crime Data in St. Louis",
       x = "Longitude", y = "Latitude") +
  coord_cartesian(xlim = c(-90.321, -90.18), ylim = c(38.54, 38.745))
```