1. R script (or R Markdown file) with some of your data visualizations (at least two plots)

See the data-wrangling.R file to see the visualization code (at the end after the data has been wrangled) in the attached git project for more information.

#make a variable for difference so we have single category to map in charts

For2020Visualizations <-prediction\_colorized\_2020 %>%

mutate(difference = paste(PreviousWinner,"->",NewWinner ))

#look at average for all the different types

For2020Visualizations %>%

group\_by(difference, NewWinner) %>%

summarize(mean\_migration=mean(totalMigration)) %>%

ggplot(aes(x=difference, y=mean\_migration, fill=NewWinner)) +

scale\_fill\_manual(values = c("Blue","Red")) +

geom\_col(stat = "identity")

Chart

Description automatically generated

#look at boxplot

For2020Visualizations %>%

ggplot(aes(x=difference,y=totalMigration)) +

geom\_boxplot()

Graphical user interface

Description automatically generated with low confidence

#Look at where the numbers are for each difference

For2020Visualizations %>%

ggplot(aes(x=difference, y=totalMigration, color=NewWinner)) +

geom\_count() +

scale\_color\_manual(values = c("Blue","Red")) +

scale\_y\_binned()

Graphical user interface, application, email

Description automatically generated

2. An explanation of what correlations or relationships your plot is displaying and how you are ensuring that the users of your visualization do not confuse the correlation displayed with causation (you can submit this as a separate text document or include this explanation in your markdown file together with your code)

The plots show a semi-regular relationship between states turning blue and total migration. But it is easy to not think it is casual because there are two big outliers (See plot2 -the boxplot). On average, higher migration from blue states leads to blue states but there is obviously another factor here not being shown. I think I need to bring in population of voters but I haven’t quite figure out how I want to handle that whether it is going to be total population or voter population because I can see advantages to both.