PDA Week 2

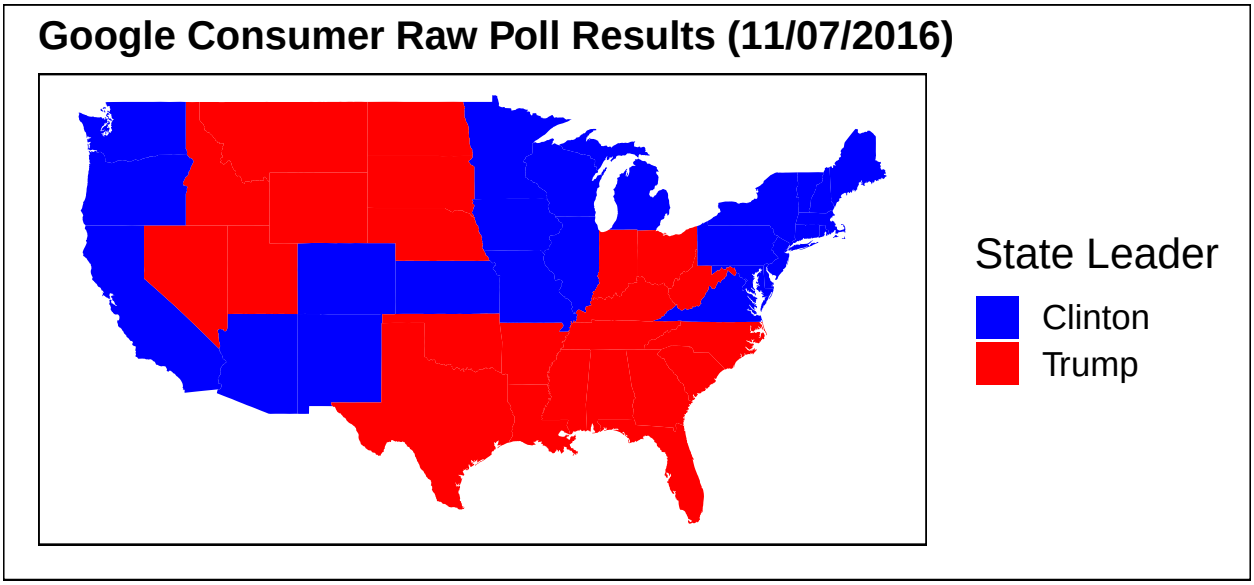
David Carbonello

9/15/2020

First, take a look at the different variables recorded in this dataset. Then, create a compelling visual based on a thought or question you have about the data. In a paragraph, explain how you used the principles emphasized this week to create an effective graphic.

You should submit a pdf file for this assignment containing your visual, the code behind it, and your paragraph of explanation.

The visual I am creating will show which candidate is leading each state using the raw polls from the latest Google Consumer Survey. It will also show how the candidate is doing across the entire US according to raw polls. The Google Consumer Survey was chosen because it included both information on raw polls from all 50 states in addition to total raw polls accross US.



Candidate	US Raw Poll Percentage
Clinton	38.03
Trump	35.69

** Johnson(5.46%) and Mcmullin (NA) not included*

The visual I created was designed to satisfy the characteristiscs of an effective visual we had discussed in class. I aimed at making the visual as simple as possible so it can be interpreted with minimal explanation needed. Additionally, I decided to focus on just Clinton and Trump because no other candidates were leading any of the states. This is why Johnson and Mcmullin are not included in the "state leader" key. Also, I removed the axis labels that contained latitude and longitude since I felt they did not offer any revelvant

additional information. I believe this helped to limit any potential confusion and improved the data to ink ratio. The use of color I believe is helpful because it clearly identifies who is leading in each state, and also red and blue were chosen to correspond to the Candidate's political party. Since the state graph alone does not indicate which Candidate is leading in the election, I included the table with US raw poll information. The table was the simplest way to include this relevant information, and is why I decided not to use a bar or pie chart. Although the polls do not add to 100%, I made sure to include Johnson and Mcmullin in the footnote to suggest why this is not the case. Google Survey did not record raw poll information on Mcmullin.

Code Appendix

```
knitr::opts_chunk$set(echo = FALSE,message = FALSE,warning = FALSE)
#install/Library packages
library(dslabs)
library(dplyr)
library(ggplot2)
library(tidyr)
library(ggthemes)
library(ggplot2)
library(dplyr)
library(maps)
library(gridExtra)
require(viridis)
library(ggpubr)
# read in data
data(package="dslabs")
data(polls_us_election_2016)
# https://www.datanovia.com/en/blog/how-to-create-a-map-using-ggplot2/  (reference for maps)
State_polls_us_election_2016=polls_us_election_2016%>%
  filter(state != 'U.S.')%>%
  mutate(state=tolower(state))%>%  # convertion state to lowercase (for join)
  rename(region=state)             # Renaming state as region (for join)

#looking at who won each state on last polling date according to Google Consumer
#max(State_polls_us_election_2016$enddate)

StatePolls_End<-State_polls_us_election_2016%>%
  filter(enddate=="2016-11-07" & pollster=='Google Consumer Surveys')%>%
  select(region,enddate,pollster,rawpoll_clinton,rawpoll_trump)%>%
  mutate(Winner=ifelse(rawpoll_clinton>rawpoll_trump,"Clinton","Trump"))


states_map <- map_data("state")
States_polls_map <- left_join(states_map, StatePolls_End, by='region')

p1<-ggplot(States_polls_map, aes(long, lat, group = group))+
  geom_polygon(aes(fill = Winner))+ theme_base()+scale_fill_manual( values = c("blue", "red"))+ggtitle(
        axis.ticks.y=element_blank())



# Create Table
GoogleConsumer<-polls_us_election_2016%>%
```

```
  filter(enddate=="2016-11-07"& pollster=='Google Consumer Surveys' & state=='U.S.')
Candidate<-c("Clinton","Trump")

US_Percentage<-c(GoogleConsumer$rawpoll_clinton,GoogleConsumer$rawpoll_trump)
df<-data.frame(Candidate, US_Percentage)
df <- setNames(df, c("Candidate","US Raw Poll Percentage"))
tbl<-ggtexttable(df, rows = NULL,theme = ttheme("classic"))
tbl<-tbl %>%
 table_cell_bg(row = 2,column = 1, fill = "blue",color = "White") %>%
 table_cell_bg(row = 3, column = 1, fill = "red",color = "White")%>%
 tab_add_footnote(text = "* Johnson(5.46%) and Mcmullin (NA) not included", size = 8, face = "italic")

# add tabe to plot together
grid.arrange(p1, tbl,
             nrow=2,
             as.table=TRUE,
             heights=c(2,1))
```