Applied Statistical Programming - ggplot

2/28/2022

Write the R code to answer the following questions. Write the code, and then show what the computer returns when that code is run. Thoroughly comment your solutions.

You have until the beginning of class 3/2 at 10:00am to complete the assignment below. You may use R, but not any online R documentation. Submit the Rmarkdown and the knitted PDF to Canvas. Have one group member submit the activity with all group members listed at the top.

Figuring out the Competition

You've been hired by a campaign to do some data analysis during the primary stage of an election. The campaign wants to understand competitiveness of certain candidates under different general election scenarios. You will plot some summary features of the provided `primaryPolls` data using `ggplot()`.

The data is associated with 2020 Democratic primary elections. Polling results for 38 states are provided. You will create a visualization of the state of the race using this data. For three states of your choosing, generate a summary figure that visualizes the support for each candidate in that state. Each plot must include:

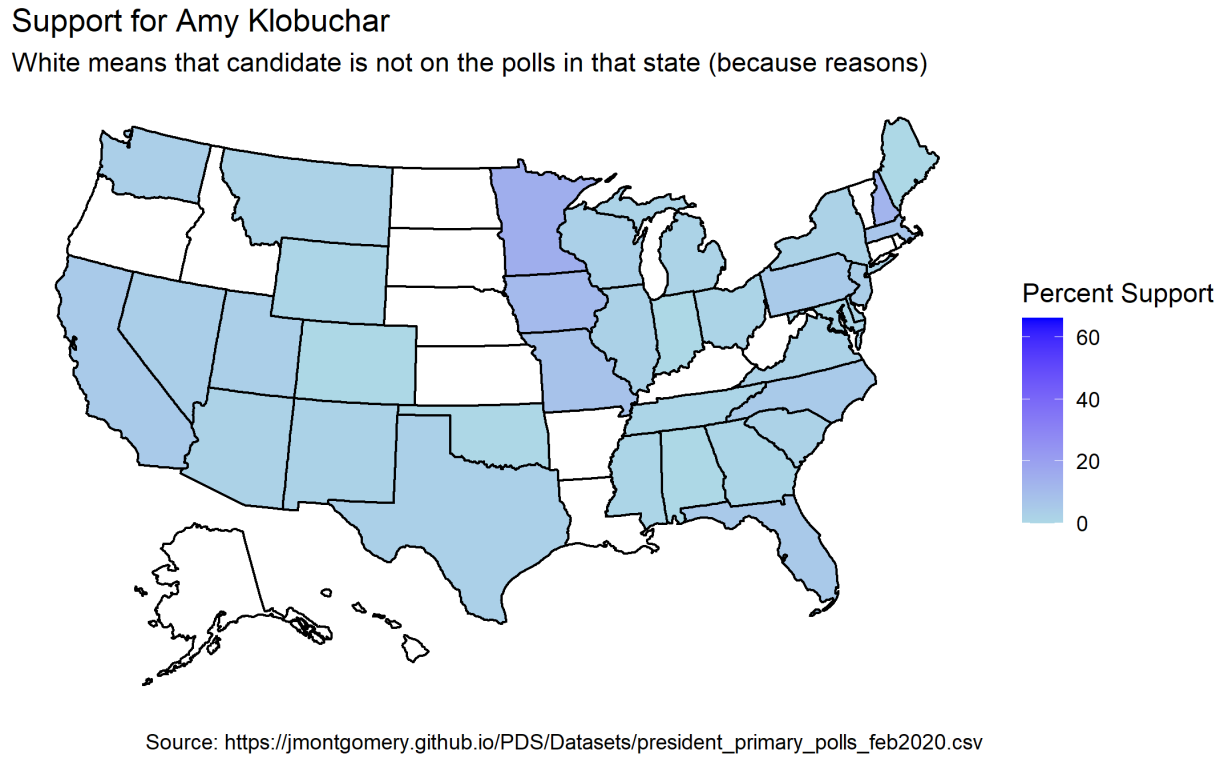
- a title,
- a subtitle,
- labeled axes,
- a legend for the candidates, and
- a source attribution to the GitHub URL for the data.

```
# Remove eval=FALSE to have this code block run.

# Load library dependencies
library(dplyr)
library(tidyr)
library(readr)

# Define path to the data
# dataURL <- 'https://jmontgomery.github.io/PDS/Datasets/president_primary_polls_feb2020.csv'
# Load the data
# Just got it manualy because of connection size issues
primaryPolls <- read.csv("president_primary_polls_feb2020.csv")

# Format the date
primaryPolls$start_date<-as.Date(primaryPolls$start_date, "%m/%d/%y")

# Your commented solution goes here.
# Here's our libraries, which includes my preferred mapping library
library(ggplot2)
library(usmap)
```

```r
# Setting up the mapping
primaryPolls$full <- primaryPolls$state

statemaps <- us_map(regions = "states")
statemaps$state <- statemaps$full

maps_state <- us_map(regions = "states")
maps_state$fips <- as.numeric(maps_state$fips)
maps_state$fips <- as.factor(maps_state$fips)

dat <- inner_join(maps_state, primaryPolls, by = "full")

# Our list of candidates
candidates <- unique(dat$candidate_name)
candidate_ids <- unique(dat$candidate_id)

candidate_list <- cbind("Candidate" = candidates,
                        "ID" = candidate_ids[1:60])

# A for-loop (this might be inefficient) that
# creates a map of the U.S., where the fill color
# is the `pct` variable. It's not clear which specific
# date is associated with each fill (and not every state
# has data on every date), but the prompt doesn't say
# I have to be a particularly competent analyst

# Note: not every candidate is on every poll in every state
mapmaker <- function(candidate_list){

  for (i in candidate_list[,2]) {
  subdat <- dat[dat$candidate_id == i,]
  print(subdat$candidate_name[1])

  map <<- ggplot(data = subdat)+
  geom_polygon(aes(x = x, y = y, group = group,
                   fill = pct),
               color = "grey")+
  labs(fill = "Percent Support") +
  scale_fill_continuous(low = "light blue", high = "blue", limits = c(0,66)) +
  geom_polygon(data = maps_state,
               aes(x = x, y = y, group = group),
               fill = NA,
               col = "black") +
  labs(title = paste("Support for", subdat$candidate_name),
       caption = paste("Source:", 'https://jmontgomery.github.io/PDS/Datasets/president_primary_polls_f
       subtitle = "White means that candidate is not on the polls in that state (because reasons)",
       x = NA) +
  theme_minimal() +
  theme(axis.title = element_blank(),
        axis.text = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

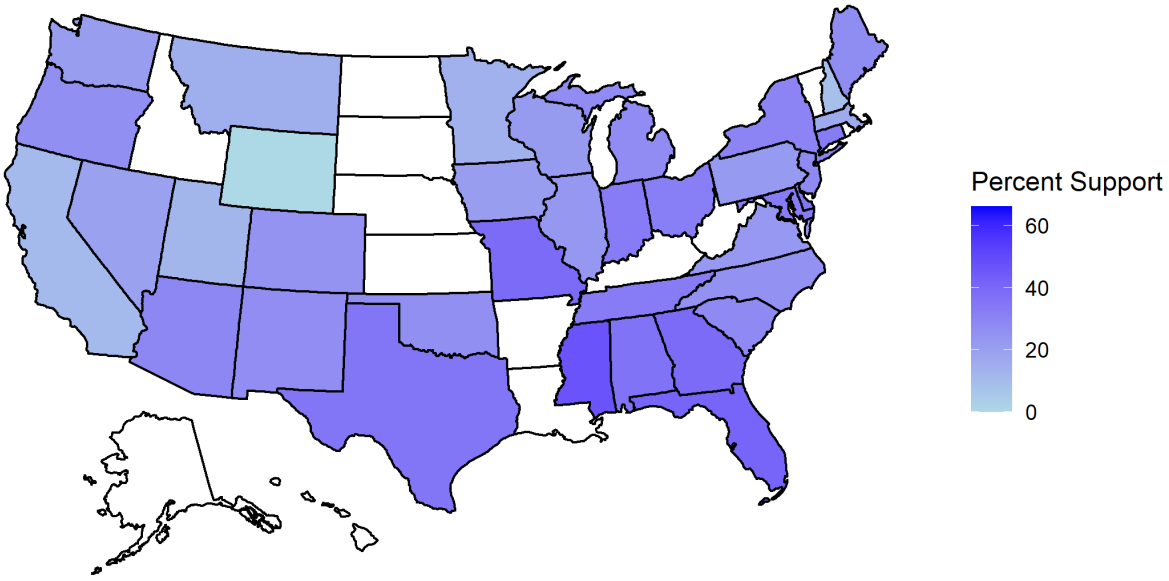
```
  print(map)
  ggsave(filename = paste0("./Maps/supportplot_", subdat$candidate_name[1], ".png"),
         plot = map)

  }

}

mapmaker(candidate_list)
```

Here's a few of the more populated ones:



Support for Joseph R. Biden Jr.

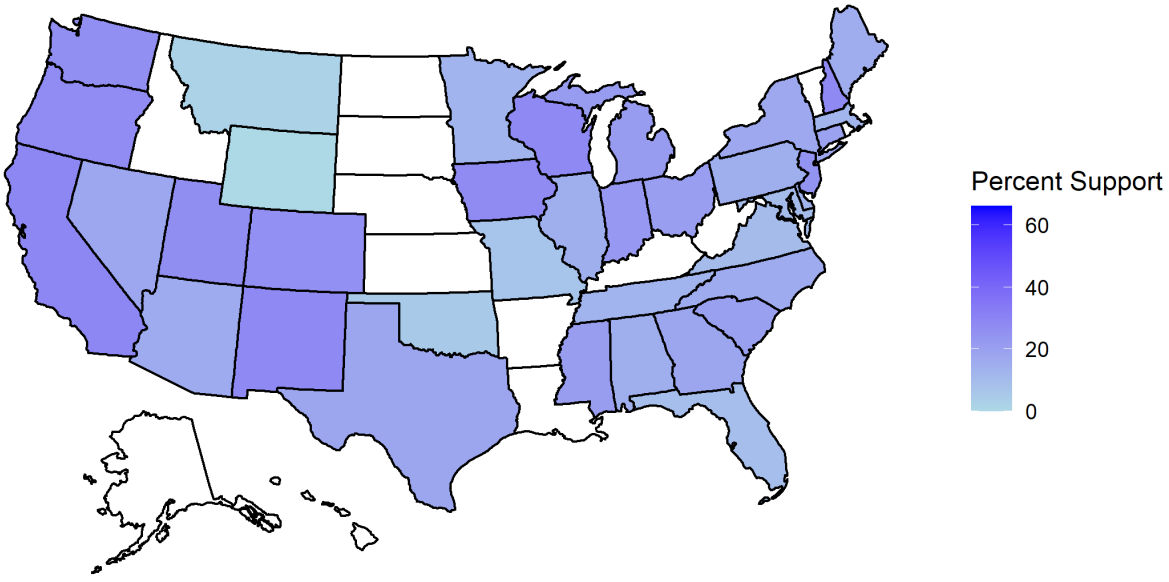
White means that candidate is not on the polls in that state (because reasons)



Source: https://jmontgomery.github.io/PDS/Datasets/president_primary_polls_feb2020.csv

Support for Bernard Sanders

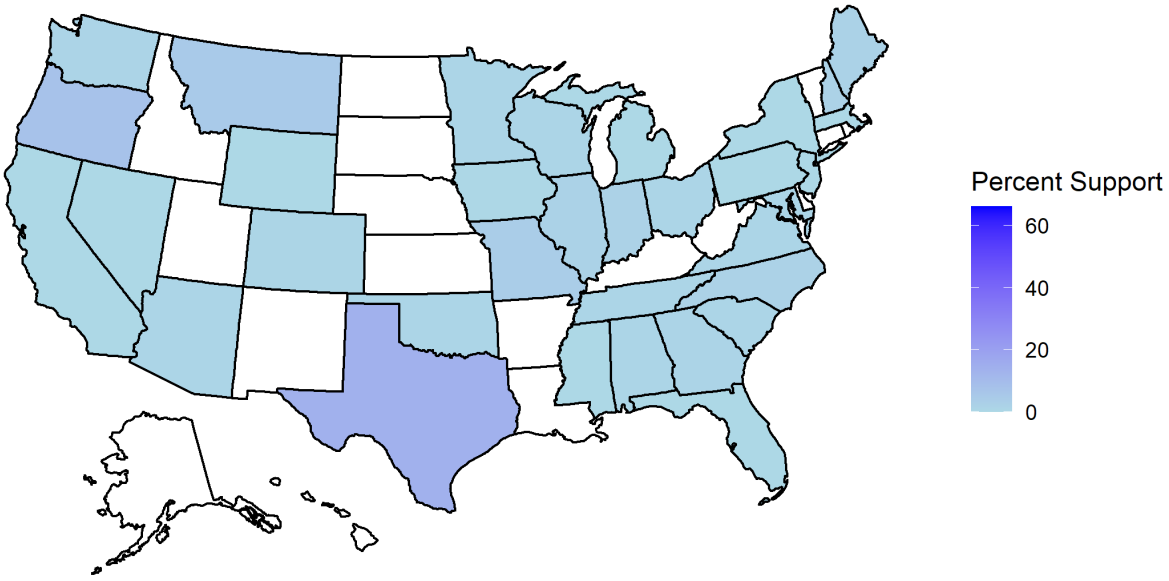
White means that candidate is not on the polls in that state (because reasons)



Source: https://jmontgomery.github.io/PDS/Datasets/president_primary_polls_feb2020.csv

Support for Beto O'Rourke

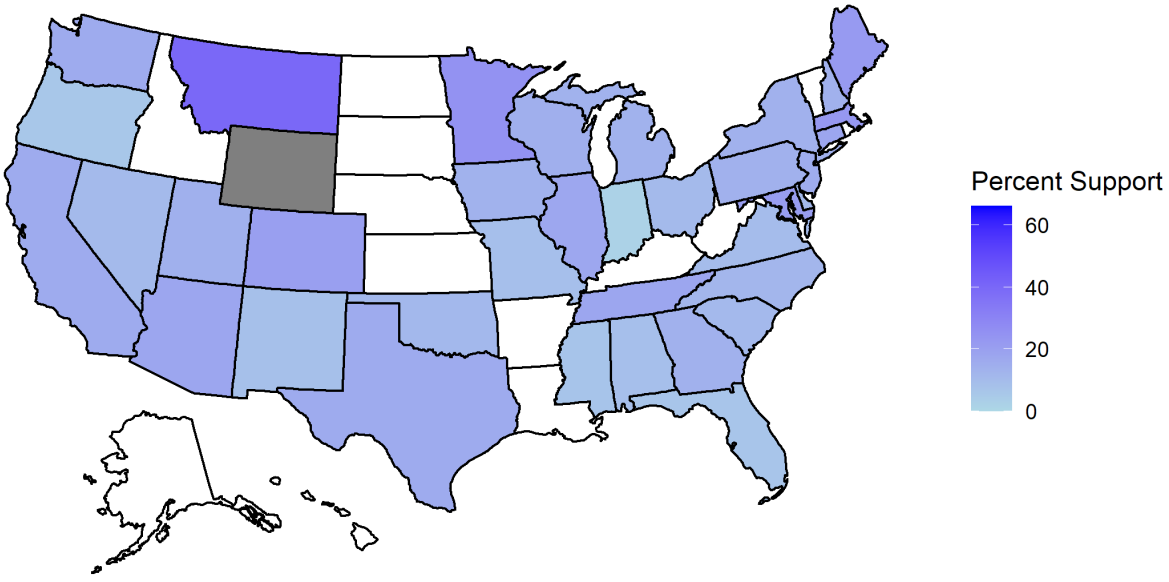
White means that candidate is not on the polls in that state (because reasons)



Source: https://jmontgomery.github.io/PDS/Datasets/president_primary_polls_feb2020.csv

Support for Elizabeth Warren

White means that candidate is not on the polls in that state (because reasons)



Source: https://jmontgomery.github.io/PDS/Datasets/president_primary_polls_feb2020.csv