

Applied Statistical Programming - dplyr and tidyr

3/2/2022

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 3/7 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.

dplyr & tidyr

You've been hired by a campaign to do some data analysis during the primary stage of an election. Before you can start, you need to organize their data that they've provided. Your regressions need to be oriented towards candidate-state pairs for an eventual analysis of expected vote shares under different general election scenarios.

You used this same data for the previous in class activity. Your task is to re-organize the `primaryPolls` data so that there is only one row for each candidate-state dyad. Limit the data down to only the relevant candidates. Once complete, compare the size of this dataset to the original dataset using the `object_size` command.

```
# 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
# NOTE: doing this locally
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.
dat <- primaryPolls %>%
  group_by(candidate_name, state, party) %>% # grouping by candidate, state, and party
  filter(state != "") %>% # dropping observations without a state
  summarise(mean_pct = mean(pct)) %>% # outputting the mean pct for each candidate, by state
  filter(mean_pct > 0) # getting rid of observations with percents = 0

head(dat)
```