dplyr

20 March, 2023

Write the R code to answer the following questions. You have until the beginning of next class to answer all of the questions below and commit to GitHub. **It's okay if you want to do this in a .R script. Because the data is so large, the code might run slowly, and you might not want to knit.**

Overview

We will continue using the polls data from class containing presidential primary polls for the 2020 election. As a reminder, these are data shared with me. Please do not use beyond class without inquiring with me further, and do not post publicly.

Question 1

- Filter the data so it includes only polls taken in Jan or Feb of 2019
- Select down to only the start_date, pct, state, and candidate
- Create a new variable that is the proportion of respondents in favor
- Find the median level of support by state limited only to candidate/state combinations with at least 5 polls

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(readr)
primaryPolls <- read_csv('president_primary_polls_feb2020.csv')
```

```
## Rows: 16661 Columns: 33
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (21): state, pollster, sponsors, display_name, pollster_rating_name, fte...
## dbl  (8): question_id, poll_id, cycle, pollster_id, pollster_rating_id, samp...
## num  (1): sponsor_ids
## lgl  (3): internal, tracking, nationwide_batch
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
primaryPolls$start_date <- as.Date(primaryPolls$start_date, "%m/%d/%y")

# code here
```