Project 1 process: Solved/confusing problems

September 20, 2023
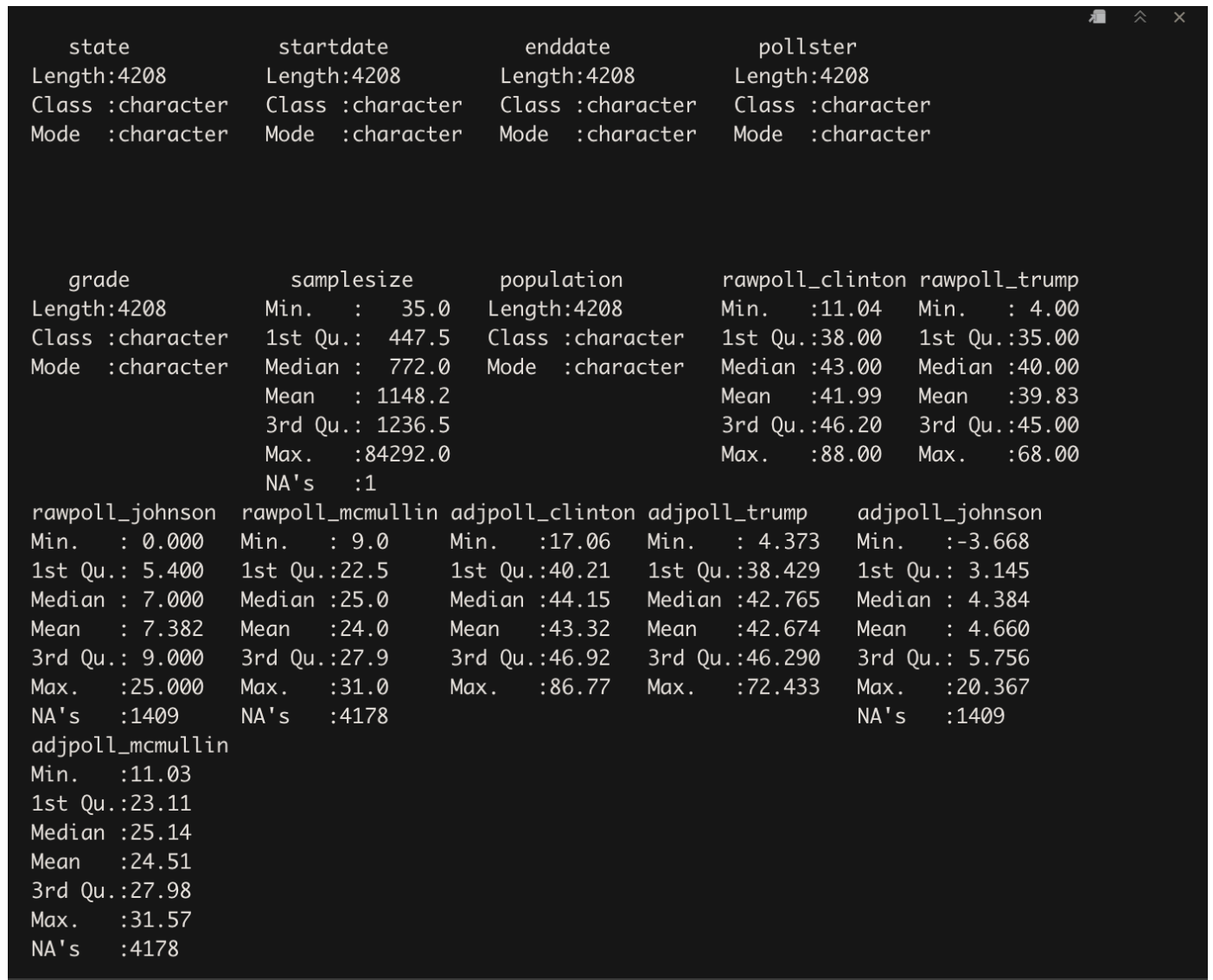
Contents

1 Done so far

Read the data polls_us_election_2016.csv
Summary data


```
    state             startdate           enddate            pollster
 Length:4208        Length:4208        Length:4208        Length:4208
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character




    grade             samplesize         population        rawpoll_clinton rawpoll_trump
 Length:4208        Min.   :    35.0   Length:4208        Min.   :11.04   Min.   : 4.00
 Class :character   1st Qu.:   447.5   Class :character   1st Qu.:38.00   1st Qu.:35.00
 Mode  :character   Median :   772.0   Mode  :character   Median :43.00   Median :40.00
                    Mean   :  1148.2                      Mean   :41.99   Mean   :39.83
                    3rd Qu.:  1236.5                      3rd Qu.:46.20   3rd Qu.:45.00
                    Max.   :84292.0                       Max.   :88.00   Max.   :68.00
                    NA's   :1
 rawpoll_johnson    rawpoll_mcmullin  adjpoll_clinton   adjpoll_trump     adjpoll_johnson
 Min.   : 0.000    Min.   : 9.0      Min.   :17.06     Min.   : 4.373    Min.   :-3.668
 1st Qu.: 5.400    1st Qu.:22.5      1st Qu.:40.21     1st Qu.:38.429    1st Qu.: 3.145
 Median : 7.000    Median :25.0      Median :44.15     Median :42.765    Median : 4.384
 Mean   : 7.382    Mean   :24.0      Mean   :43.32     Mean   :42.674    Mean   : 4.660
 3rd Qu.: 9.000    3rd Qu.:27.9      3rd Qu.:46.92     3rd Qu.:46.290    3rd Qu.: 5.756
 Max.   :25.000    Max.   :31.0      Max.   :86.77     Max.   :72.433    Max.   :20.367
 NA's   :1409      NA's   :4178                                          NA's   :1409
 adjpoll_mcmullin
 Min.   :11.03
 1st Qu.:23.11
 Median :25.14
 Mean   :24.51
 3rd Qu.:27.98
 Max.   :31.57
 NA's   :4178
```

Figure 1: Summary of polls_us_election_2016.csv

1.1 Explore variables

Variables	Size	Numbers of unique	Number of missings	Comment
state	4208	55	0	"U.S." means national polls
startdate	4208	352	0	E: 2015.11.06 L: 2016.11.06
enddate	4208	345	0	E: 2015.11.08 L: 2016.11.07
pollster	4208	196	0	just pollster's name
grade	4208	11 (including NA)	429	see pdf files for exact numbers
population	4208	4	0	exact numbers for each on pdf
rawpoll_clinton	4208	1312	0	Percentage
rawpoll_trump	4208	1385	0	Percentage
rawpoll_johnson	4208	585	1409	Percentage
rawpoll_mcmullin	4208	17	4178	Percentage
adjpoll_clinton	4208	4200	0	Percentage adjusted by 538
adjpoll_trump	4208	4204	0	Percentage adjusted by 538
adjpoll_johnson	4208	2211	1409	Percentage adjusted by 538
adjpoll_mcmullin	4208	31	4178	Percentage adjusted by 538

1.2 data dictionary

1.2.1 state

The name of the state where the election is held, "U.S." means national polls.

1.2.2 startdate

Start data of poll.

1.2.3 enddate

End data of poll.

1.2.4 pollster

Organization name that conducts or analyzes opinion polls.

1.2.5 grade

Grade assigned by Fivethirtyeight (sometimes as 538, an American website that focuses on opinion poll) to pollster.
"A letter grade from A+ to F that reflects the accuracy of a polling organization's polls and its predictive plus-minus score (a projection of how well we think a pollster's polls will do in the future). Pollsters with a small number of polls get a provisional rating (for example, A/B) rather than a precise letter grade" (Cite from `https://projects.fivethirtyeight.com/pollster-ratings/`).

1.2.6 samplesize

Sample size of polls for each pollster.

1.2.7 population

Type of population being polled.

- LV(likely Voters):
  population whose most likely to participate in 2016 election

- RV (Registered Voters):
  RV includes all eligible and registered voters, whether or not they end up voting.

- A (Adults):
  Refers to all persons of legal voting age.

- V (Voters):
  Persons who has voted in an election.

1.2.8 rawpoll_clinton

Percentage for Hillary Clinton.

1.2.9 rawpoll_trump

Percentage for Donald Trump.

1.2.10 rawpoll_johnson

Percentage for Gary Johnson.

1.2.11 rawpoll_mcmullin

Percentage for Evan Mcmullin.

1.2.12 adjpoll_clinton

Fivethirtyeight adjusted percentage for Hillary Clinton.

1.2.13 adjpoll_trump

Fivethirtyeight adjusted percentage for Donald Trump.

1.2.14 adjpoll_johnson

Fivethirtyeight adjusted percentage for Gary Johnson.

1.2.15 adjpoll_mcmullin

Fivethirtyeight adjusted percentage for Evan Mcmullin.

2 Problems

3 Plans