Problem Set 6

Daniel Shapiro

10/20/2022

Question 1 Setup:

Download the votes.csv dataset from the course website. These data describe the number of votes obtained by the Democratic and Republican candidates in each presidential election from 1932 to 2008.

```r
votes <- read.csv("votes.csv")
```

1a) Delete observations for Alaska, DC, and Hawaii (since residents of these states did not have the vote throughout this entire time period). Convert the data from wide to long format, with four variables (state, party, election, and vote).

```r
# I'm going to do this in kind of an annoying way.

votes <- votes %>%
  filter(state != "Hawaii") %>%
  filter(state != "Alaska") %>%
  filter(state != "District of Columbia") %>%
  pivot_longer(!state, names_to = "election", values_to = "vote") %>%
  separate(election, c("party", "election"))
```

1b) Transform the data to a state-election year dataset, with four variables (election, state, pcDem, turnout).

```r
# Had to suppress this really annoying message

options(dplyr.summarise.inform = FALSE)

newvotes <- votes %>%
  group_by(election, state) %>%
  summarize(turnout = sum(vote))

demvotes <- votes %>%
  filter(party == "d")

votemerge <- merge(newvotes, demvotes, by = c("election", "state")) %>%
  select(-party) %>%
  mutate(pcDem = vote/turnout) %>%
  select(-vote)
```