

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.

```
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).

```
# 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).

```
# 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*100) %>%
```

```
select(-vote)

votemerge <- votemerge[,c(1, 2, 4, 3)]
```

1c) Plot democratic vote share over time for Pennsylvania, South Carolina, and West Virginia using a `ggplot()` line graph with clear and appropriately labeled legends and axes.

```
# West Virginia is misspelled...edit out with recode() below

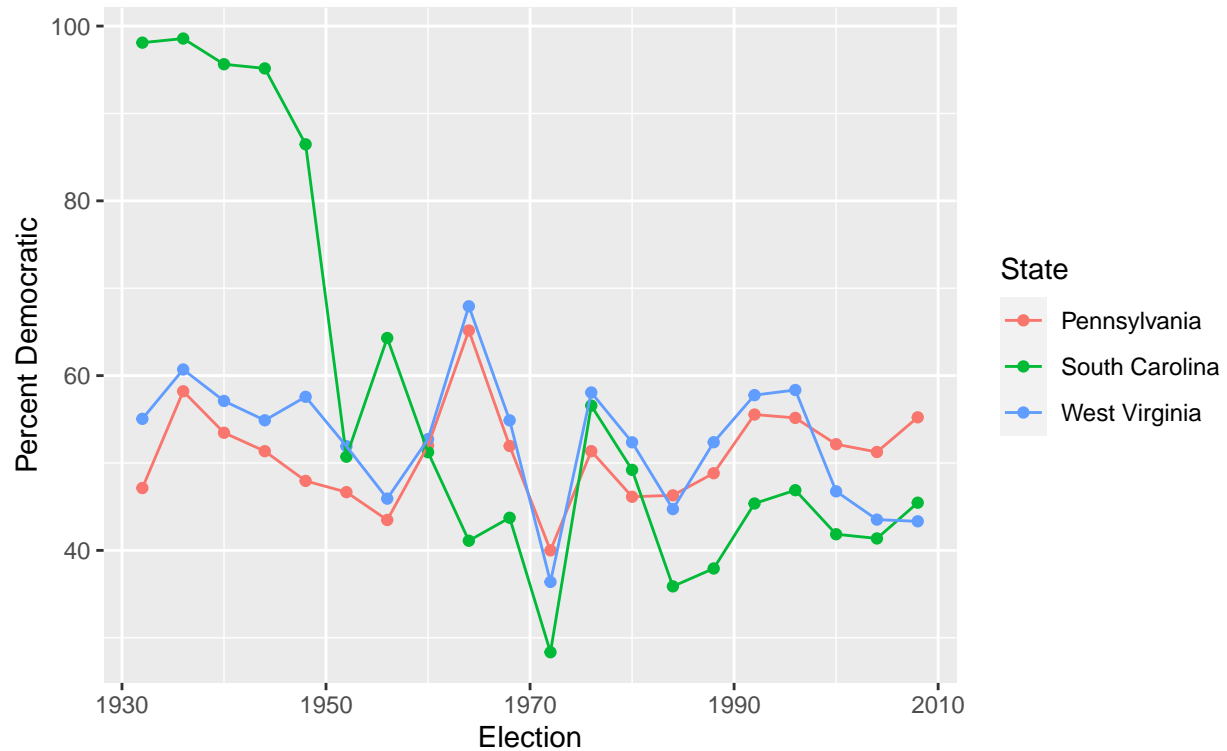
cdata <- votemerge %>%
  filter(state %in% c("Pennsylvania", "South Carolina", "West Vrginia"))

cdata$state <- recode(cdata$state, `West Vrginia` = "West Virginia")

cdata$election <- as.numeric(cdata$election)

ggplot(cdata, aes(x = election, y = pcDem)) +
  geom_point() +
  geom_line() +
  aes(color = state) +
  labs(x = "Election",
       y = "Percent Democratic",
       color = "State",
       title = "Democratic Vote Pct., 1932-2008",
       subtitle = "Data from Pennsylvania, South Carolina, West Virginia")
```

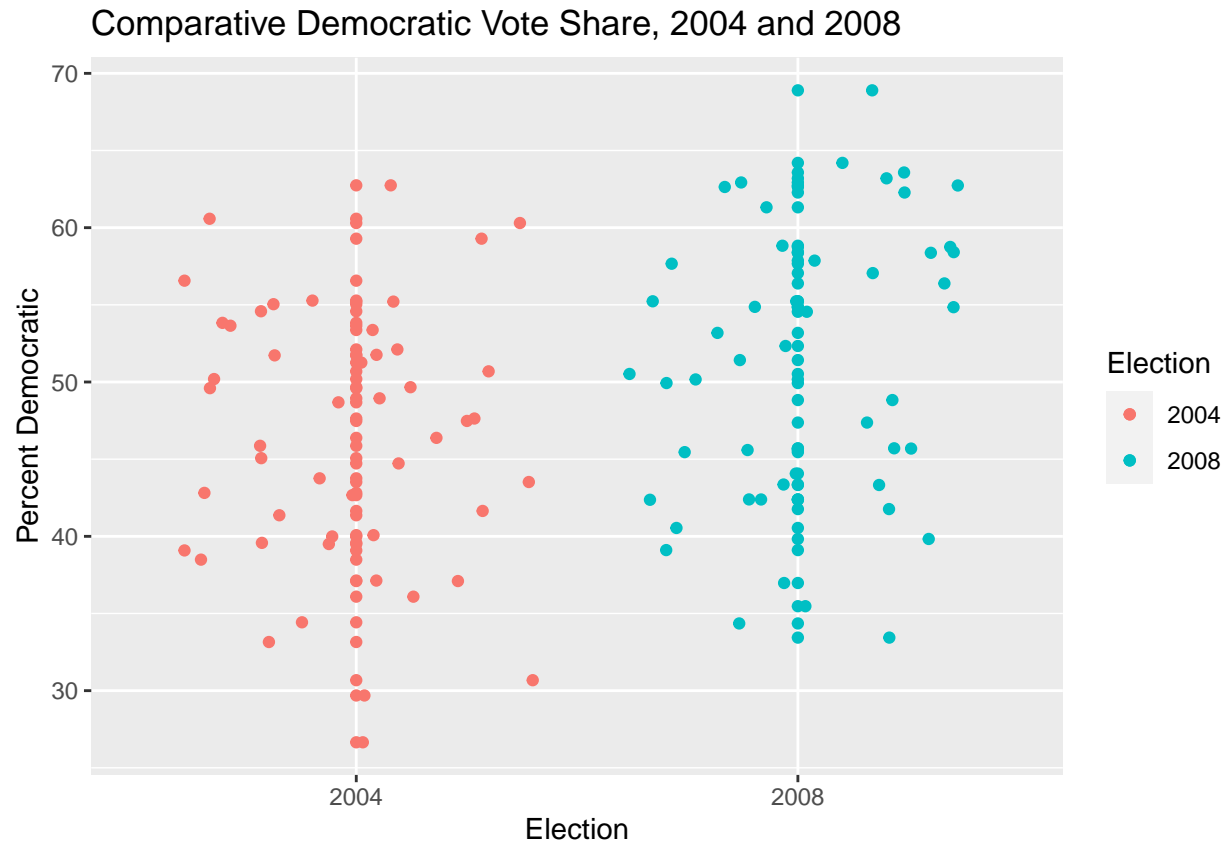
Democratic Vote Pct., 1932–2008
Data from Pennsylvania, South Carolina, West Virginia



1d) Use `ggplot()` to make a scatterplot showing the relationship between 2004 and 2008 democratic vote share.

```
ddata <- votemerge %>%
  filter(election %in% c(2004, 2008))

ggplot(ddata, aes(x = election, y = pcDem)) +
  geom_point(aes(color = election)) +
  geom_jitter(aes(color = election)) +
  labs(x = "Election",
       y = "Percent Democratic",
       color = "Election",
       title = "Comparative Democratic Vote Share, 2004 and 2008")
```



1e) Use a paired t-test to test whether the difference between the 2004 and 2008 democratic vote share was significantly different across states. Interpret your results.

```
# First, I split my data into two -- 2004 and 2008 and select pcDem.

e2004 <- ddata %>%
  filter(election == 2004) %>%
  select(pcDem)

e2008 <- ddata %>%
  filter(election == 2008) %>%
  select(pcDem)

t.test(e2004[,1], e2008[,1], paired = TRUE, alternative = "two.sided")

##
## Paired t-test
##
## data: e2004[, 1] and e2008[, 1]
## t = -10.415, df = 47, p-value = 8.493e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -5.861352 -3.963520
## sample estimates:
```

```
## mean of the differences
## -4.912436
```

The p-value of the test is 8.493^{-14} , which is lower than the significance level $\alpha = 0.05$. Thus, we can reject the null hypothesis and say that the average Democratic vote share in 2008 is significantly different than the average Democratic vote share in 2004, with a p-value at 8.493^{-14} .