

# Problem set 2

## Solutions

Due 10/11/2021 at 5pm

### Question 1: US presidential election results (again)

Download the file “tidy\_county\_pres\_results.csv.zip” from the repository (<https://github.com/UChicago-pol-methods/IntroQSS-F21/tree/main/data>), unzip it, and put the CSV file in the same directory as your Rmd file.

Then load the data:

```
library(tidyverse)
df <- read_csv("tidy_county_pres_results.csv")
```

For each US county (uniquely identified by FIPS and labeled with county and state) in each presidential election year, we have the total number of votes cast (total\_vote), number of votes for the Democratic candidate (dem\_vote), and number of votes for the Republican candidate (rep\_vote).

(1a) Add a variable called other\_vote\_share, which is the proportion of votes cast for candidates other than the Democratic and the Republican.

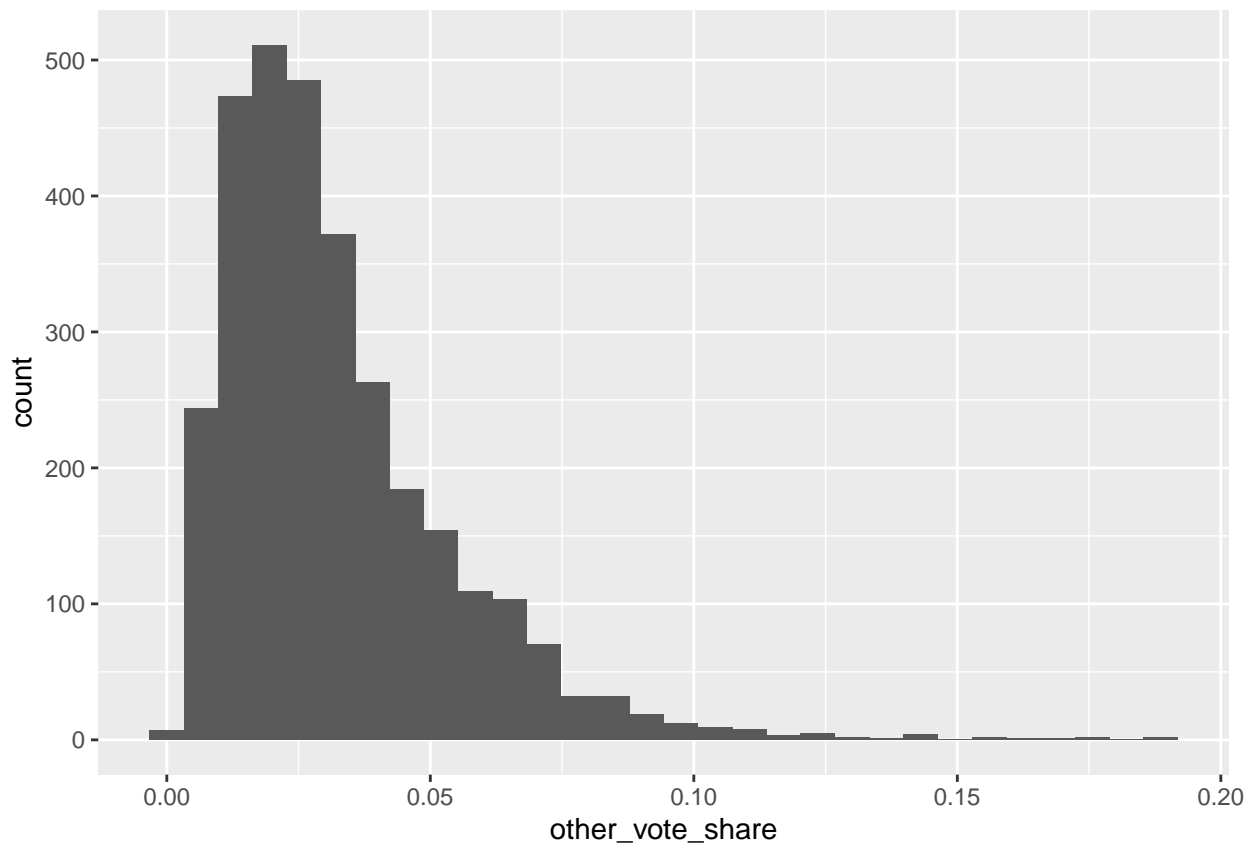
```
df %>%
  mutate(other_vote_share = 1 - (dem_vote + rep_vote)/total_vote) -> df2
```

(1b) Show a histogram of other\_vote\_share in 2000.

```
df2 %>%
  filter(year == 2000) %>%
  ggplot(aes(x = other_vote_share)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



(1c) Identify the counties with the highest `other_vote_share` in 2000. Output a table showing the county name, state, and `other_vote_share` for the six counties with the highest `other_vote_share` in 2000. (Don't worry about making the table look nice; just produce the raw R output.)

```
df2 %>%
  filter(year == 2000) %>%
  arrange(desc(other_vote_share)) %>%
  select(county, state, other_vote_share) %>%
  head()
```

```
## # A tibble: 6 x 3
##   county      state other_vote_share
##   <chr>      <chr>          <dbl>
## 1 Jefferson  IA              0.190
## 2 San Miguel CO              0.189
## 3 San Juan   CO              0.177
## 4 Grand      UT              0.175
## 5 Missoula   MT              0.169
## 6 Mendocino  CA              0.160
```

(1d) Using `group_by()` and `summarize()`, produce and store a new tibble showing the two-party vote share for the Democrat in each election year. (“Two-party vote share for the Democrat” is the votes for the Democrat divided by the votes for either the Democrat or the Republican.) Use it to make a plot showing the Democrats’ two-party vote share (vertical axis) across years (horizontal axis).

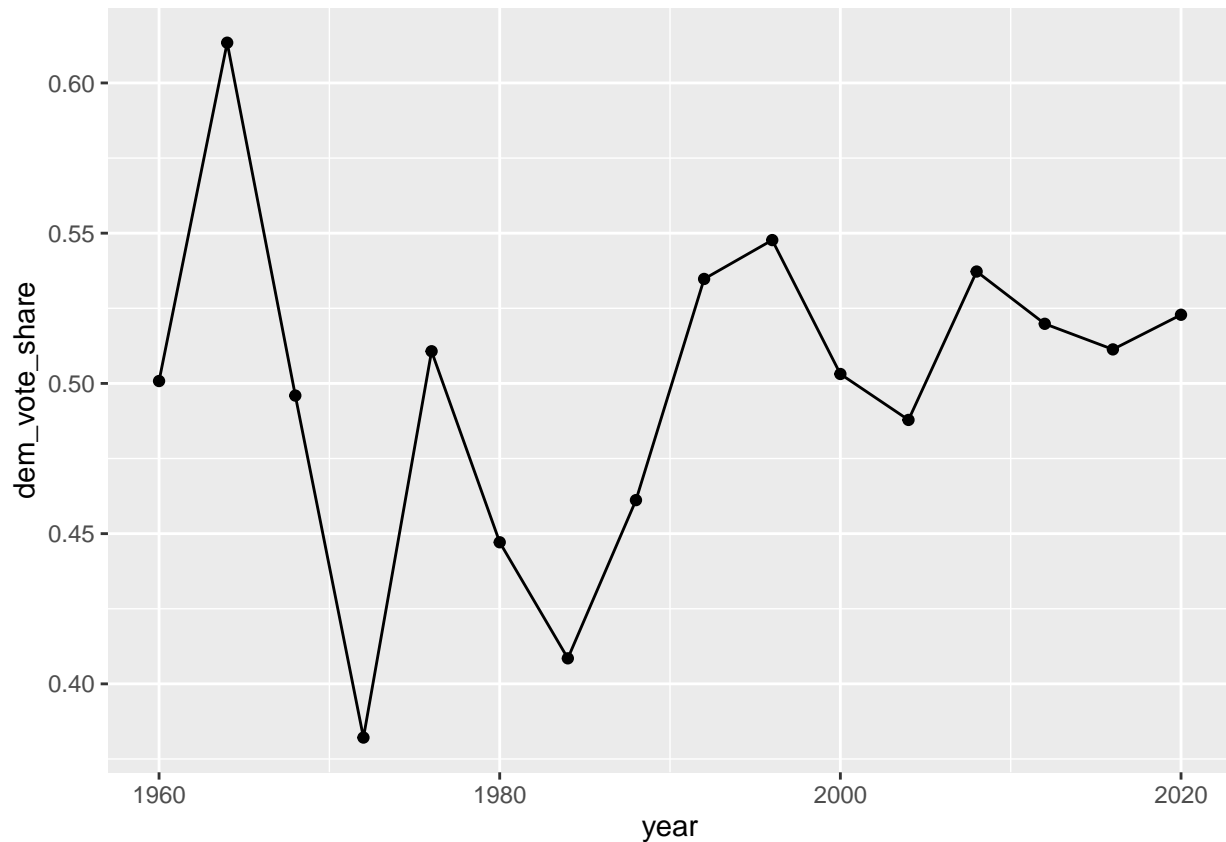
```
df %>%
  group_by(year) %>%
```

```

summarize(dem_sum = sum(dem_vote, na.rm = T),
          demrep_sum = sum(dem_vote + rep_vote, na.rm = T)) %>%
mutate(dem_vote_share = dem_sum/demrep_sum) -> df_vote_share

df_vote_share %>%
  ggplot(aes(x = year, y = dem_vote_share)) +
    geom_line() +
    geom_point()

```



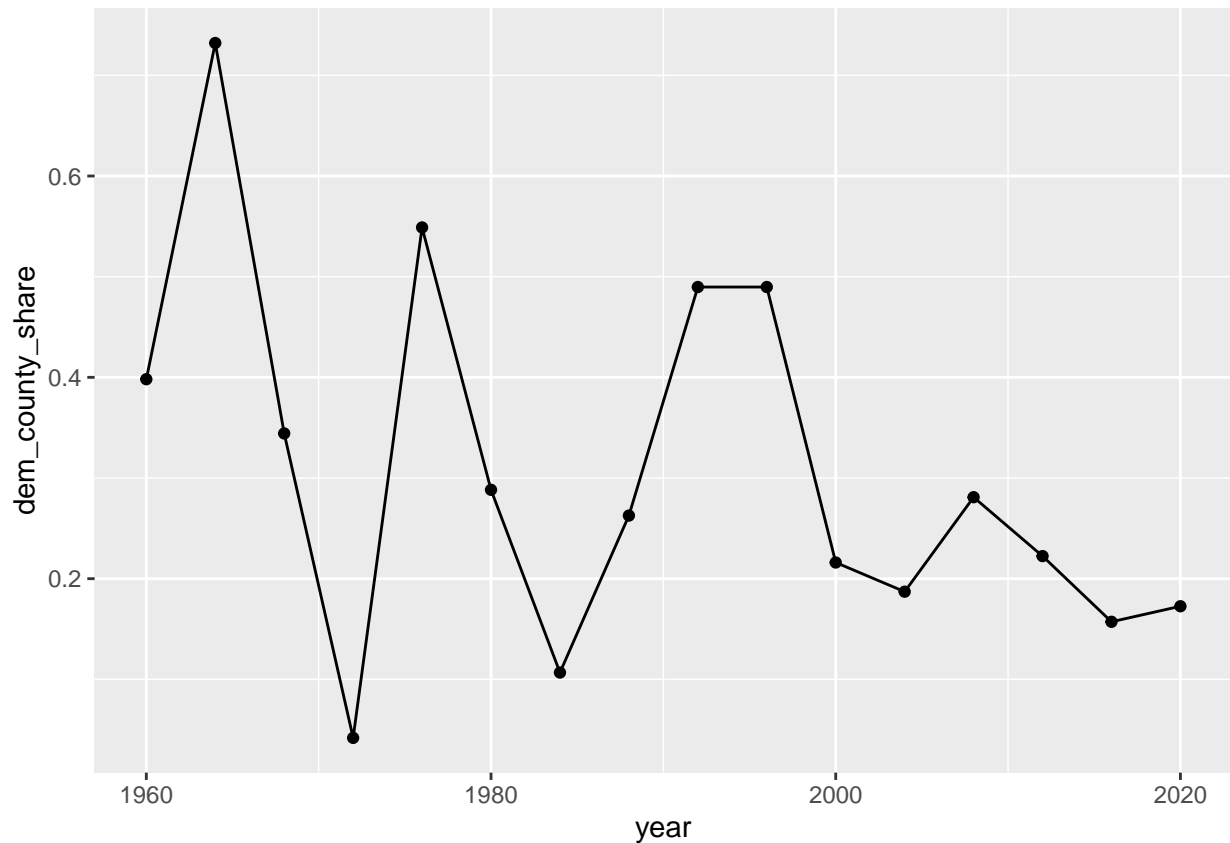
(1e) Using `group_by()` and `summarize()`, produce and store a new tibble showing the proportion of counties in which the Democrat got more votes than the Republican in each election year. Use it to make plot showing the share of counties won by the Democrat (vertical axis) across years (horizontal axis).

```

df %>%
  mutate(dem_beats_rep = dem_vote > rep_vote) %>%
  group_by(year) %>%
  summarize(dem_county_share = mean(dem_beats_rep, na.rm = T)) -> df_county_share

df_county_share %>%
  ggplot(aes(x = year, y = dem_county_share)) +
    geom_line() +
    geom_point()

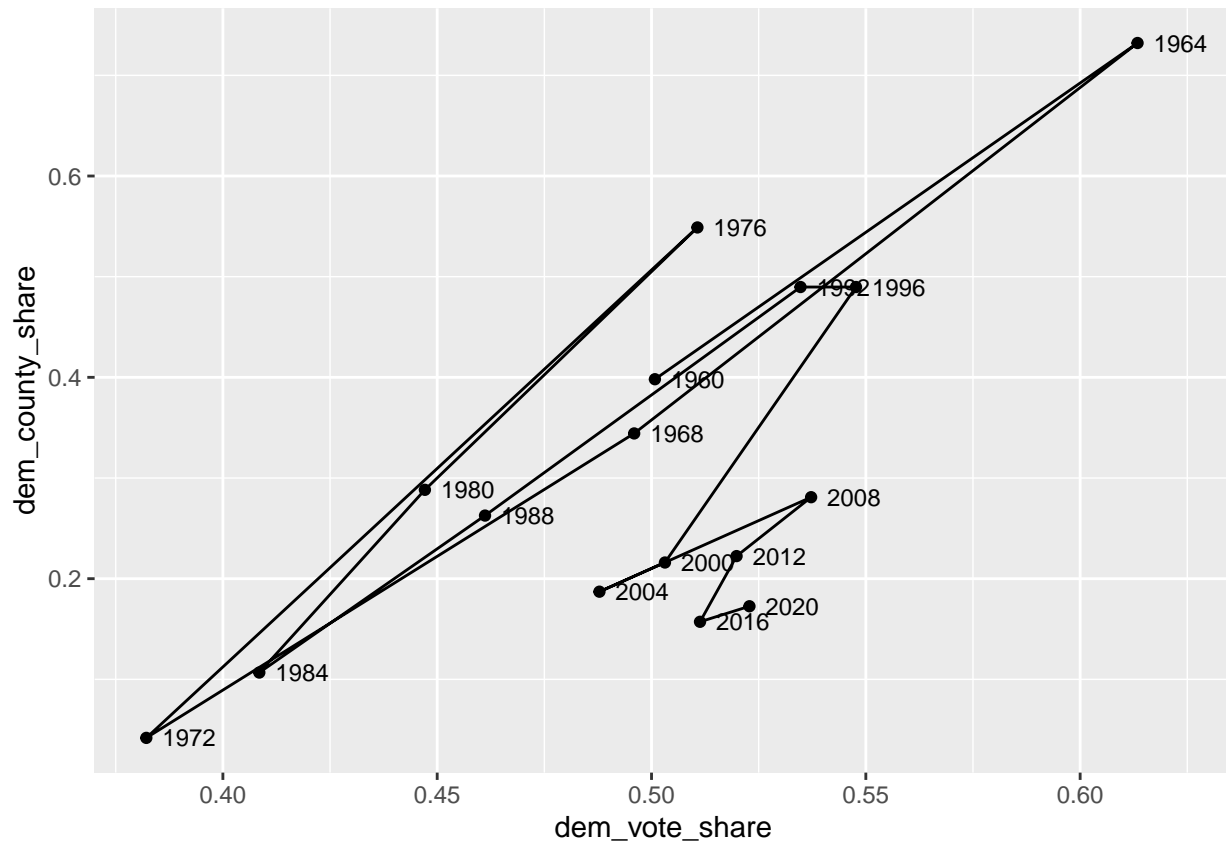
```



(1f) Use `left_join()` to merge the two tibbles (one with county share, the other with vote share) and store the result. Use this new tibble to plot the Democratic county share (vertical axis) against the Democratic vote share (horizontal axis) over time, as in the last problem set.

```
df_county_share %>%
  left_join(df_vote_share, by = "year") -> df_joined

df_joined %>%
  ggplot(aes(x = dem_vote_share, y = dem_county_share, label = year)) +
  geom_path() +
  geom_point() +
  geom_text(nudge_x = .01, size = 3)
```



(1g) Use `pivot_longer()` to convert the tibble created in the last question to a format appropriate for plotting both the Democratic vote share and the Democratic county share (vertical axis) against the year (horizontal axis) as on the last problem set. Make that plot.

```
df_joined %>%
  pivot_longer(cols = c(dem_county_share, dem_vote_share)) %>%
  ggplot(aes(x = year, y = value, col = name)) +
  geom_point() +
  geom_line()
```

