Project Part 3

Change of topic

I realized very quickly that there is very little data on wine that exists. As a result, I decided to change my topic to urbanization and politics, into which I would love to dive more deeply. I am particularly intrigued by how the percentage of urban population affects the way states vote.

Data Description

Research Question

Does the percent of a state population who lives in urban environments affect that state's election results?

General Context and Sources

The urbanization data set comes from Iowa State University's Iowa Community Indicators Program (1). These data are a population that includes all 50 US states and the District of Columbia. The data were collected through the US Decennial Census. Each row represents a state (or DC) and each column represents a year of the Decennial Census (1900, 1910, 1920, etc.). One potential issue is that because there are less than 15 observations, the sample size will be too small. These data are appropriate for answering the research question because they are the defined predictor.

The Senate election results data set comes from MIT's Election Data and Science Lab (2). These data are a population that includes all 50 US states. It is not clear on the MIT Election Lab website how the data were collected, but it is publicly available data. Each row represents a Senate candidate, and the columns include the year and state in which they were running, their name, their party, the number of votes they got, and the total number of votes. One potential issue is that the data will require a lot of quality checking, because there are a few results that seem to be coded incorrectly. For example, in Louisiana in 1978, where only one candidate was recorded, and they were only recorded to have recieved one vote.

Graphical Summary

```
scatter <- ggplot(data, aes(x=urban_prop, y=dem_prop, color=year))
scatter + geom_point() + facet_wrap(~year) +
  labs(
    title="Voting Behaviors and Urban Population",
    x="Proportion of Population Living in Urban Centers",
    y="Proportion of Major Party Votes for Democrats",
    color="Census Year"
  ) +
  scale_x_continuous(n.breaks=6, limits=c(0,1)) +
  scale_y_continuous(n.breaks=6, limits=c(0,1)) +
  theme(
    strip.background = element_rect(colour="gray50", fill="gray50"),
    strip.text=element_text(color="white", face="bold"),
  )
```
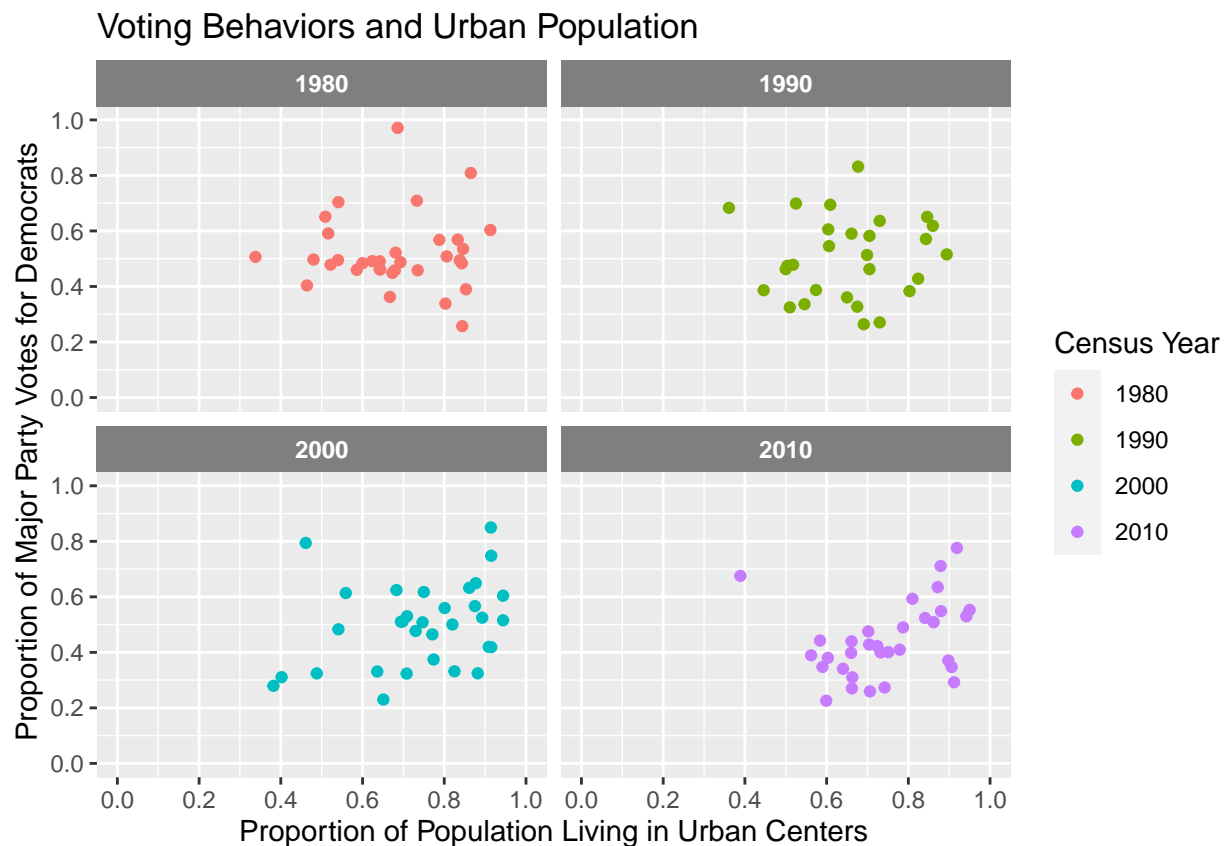

From the graphs, it looks like there is not much of a relationship between the two variables, except maybe in the years 2000 and 2010, where there is a small positive correlation.

Numerical Summary

```
data %>% group_by(year) %>%
  summarize(
    mean_urban=mean(urban_prop),
    sd_urban=sd(urban_prop),
    mean_dem=mean(dem_prop),
    sd_dem=sd(dem_prop),
    urban_dem_cor=cor(urban_prop, dem_prop))
```

```
## # A tibble: 4 x 6
##   year  mean_urban sd_urban mean_dem sd_dem urban_dem_cor
##   <fct>      <dbl>    <dbl>    <dbl>  <dbl>         <dbl>
## 1 1980       0.680    0.142    0.520  0.135      0.000568
## 2 1990       0.653    0.136    0.503  0.146      0.0480
## 3 2000       0.742    0.161    0.499  0.151      0.314
## 4 2010       0.747    0.135    0.443  0.136      0.311
```

The numerical summary tells us that the population has become more urban over time, but the spread has stayed essentially the same. The population as a whole has also, however, voted less Democratic and more Republican in Senate elections over time, with the spread staying more or less constant. The correlation values parallel the conclusions drawn from the graphical summary, which are that in the two most recent Decennial Censuses, the proportion of the population that lives in urban areas has a slightly positive correlation with the proportion of major party votes for Democrats.

References

1. Iowa State University Iowa Community Indicators Program https://www.icip.iastate.edu/tables/population/urban-pct-states
2. MIT Election Data and Science Lab https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/PEJ5QU