Khaykin-Proposal.Rmd

Alex Khaykin

2022-11-10

Data Preparation

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
# load data
pres_elect <- read_rds("../pres_elect.rds")
crime <- read_rds("../crime.rds")
votes_crime <- read_rds("../votes_crime_1979-2020_data.rds")
```

Research question

How does violent crime in the United Staes affect the popular vote trends across the states?

Cases

What are the cases, and how many are there?

The cases are total violent crime data by state from 1979 through 2020, merged with presidential election vote tallies by political party for all states for the same time period.

```
glimpse(votes_crime)
```

```
## Rows: 2,142
## Columns: 8
## $ state             <I<chr>> ALABAMA, ALASKA, ARIZONA, ARKANSAS, CALIFORNIA, C~
## $ year              <int> 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1979~
## $ tot_violent_crime <int> 15578, 1994, 14528, 7984, 184087, 14472, 12902, 3127~
## $ total_votes       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ DEMOCRAT          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ LIBERTARIAN       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ OTHER             <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ REPUBLICAN        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

Data collection

Describe the method of data collection.

The data for presidential elections by states was downloaded from MIT Election Data and Science Lab website. The link to the data is below:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX#

The violent crime rate data by state was downloaded from Federal Bureau of Investigation Crime Data Explorer website. The link to the data is below:

https://usafacts.org/data/topics/security-safety/crime-and-justice/crime-and-police/violent-crimes/?utm_source=google&utm_medium=cpc&utm_campaign=ND-StatsData

Type of study

What type of study is this (observational/experiment)?

This is an observation study of historical data.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

Please see links to the data and websites utilized:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX#

https://usafacts.org/data/topics/security-safety/crime-and-justice/crime-and-police/violent-crimes/?utm_source=google&utm_medium=cpc&utm_campaign=ND-StatsData

Dependent Variable

What is the response variable? Is it quantitative or qualitative?

The dependant response variable, is the party that von the popular vote in each state. This is a qualitative variable.

Independent Variable(s)

The independant variable is the violent crime totals in each year by state. This is a quantitative variable.

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

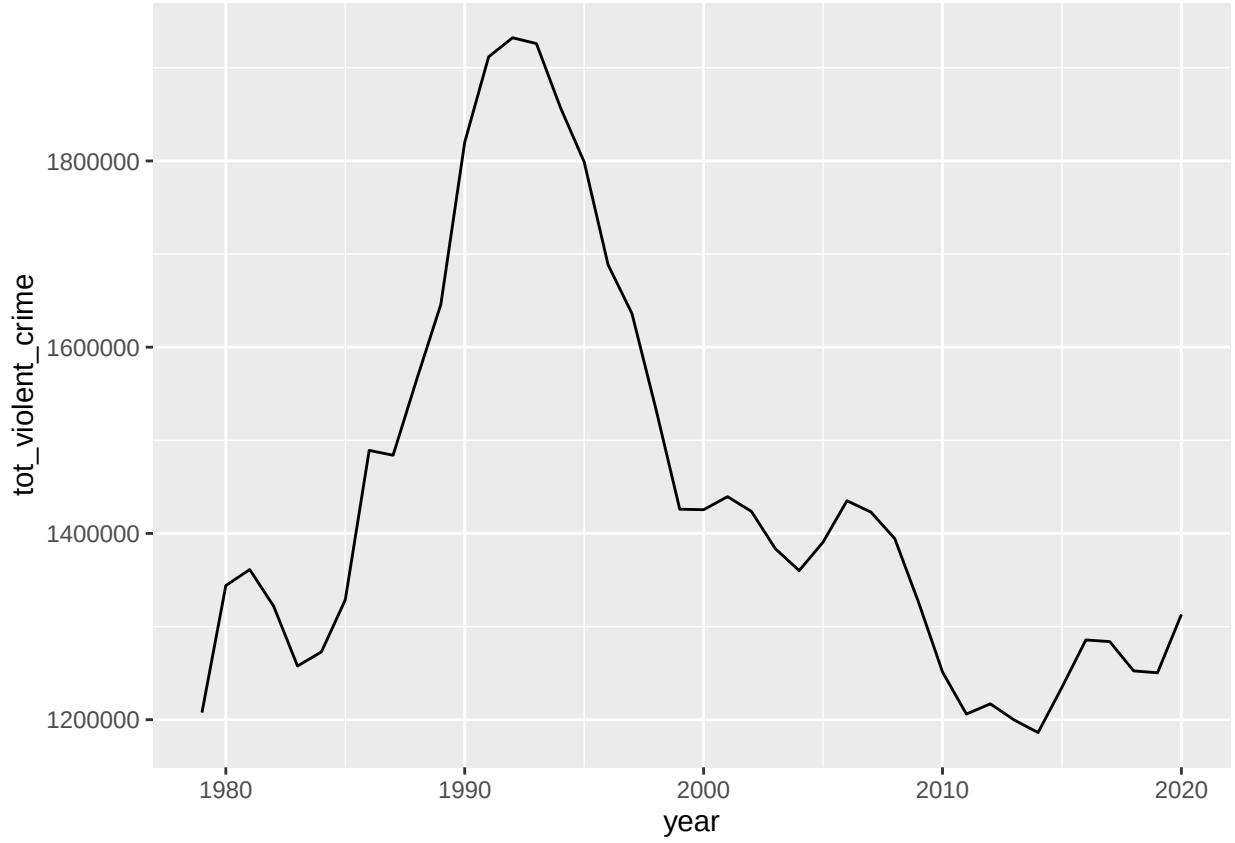
The summary statistics are total popular votes as a percentage and crime data for the entire country. Please see plots below.

```
vote_summ <- pres_elect %>%
  select(year, DEMOCRAT:REPUBLICAN) %>%
  pivot_longer(-year, names_to = "party", values_to = "votes") %>%
  group_by(year, party) %>%
  summarise(
    tot_votes = sum(votes, na.rm = TRUE),
    mean_votes = mean(votes, na.rm = TRUE),
    .groups = "drop_last"
  ) %>%
```

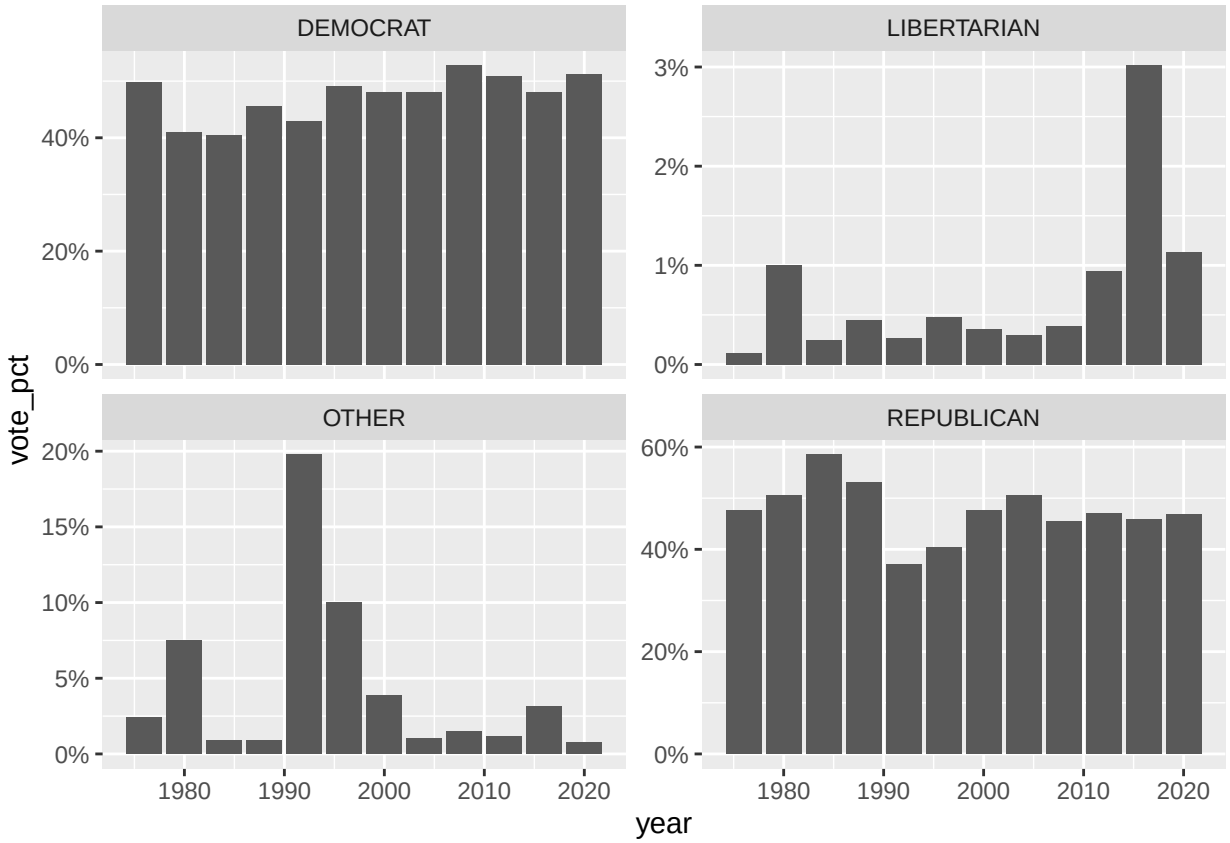
```
  mutate(vote_pct = tot_votes / sum(tot_votes))

crime_summ <- crime %>%
  group_by(year) %>%
  summarise(
    tot_violent_crime = sum(tot_violent_crime, na.rm = TRUE),
    mean_violent_crime = mean(tot_violent_crime, na.rm = TRUE),
    .groups = "drop"
  )

ggplot() +
  geom_line(
    data = crime_summ,
    aes(year, tot_violent_crime)
  )
```




```
ggplot() +
  geom_col(
    data = vote_summ,
    aes(year, vote_pct)
  ) +
  scale_y_continuous(labels = scales::percent) +
  facet_wrap(~party, scales = "free_y")
```




```
vote_summ
```

```
## # A tibble: 48 x 5
## # Groups:   year [12]
##     year party       tot_votes mean_votes vote_pct
##    <int> <chr>           <int>      <dbl>    <dbl>
##  1  1976 DEMOCRAT     40680446    797656.  0.499
##  2  1976 LIBERTARIAN     95626      3542.  0.00117
##  3  1976 OTHER         1954379     41583.  0.0240
##  4  1976 REPUBLICAN   38870893    762174.  0.476
##  5  1980 DEMOCRAT     35480948    695705.  0.410
##  6  1980 LIBERTARIAN    867401     17348.  0.0100
##  7  1980 OTHER         6505863    127566.  0.0752
##  8  1980 REPUBLICAN   43642639    855738.  0.505
##  9  1984 DEMOCRAT     37449813    734310.  0.404
## 10  1984 LIBERTARIAN    227204      5826.  0.00245
## # ... with 38 more rows
```

```
crime_summ
```

```
## # A tibble: 42 x 3
##     year tot_violent_crime mean_violent_crime
##    <int>             <int>              <dbl>
##  1  1979           1207653            1207653
##  2  1980           1344053            1344053
```

```
##  3  1981           1361239           1361239
##  4  1982           1321896           1321896
##  5  1983           1257651           1257651
##  6  1984           1272794           1272794
##  7  1985           1328757           1328757
##  8  1986           1489169           1489169
##  9  1987           1483999           1483999
## 10  1988           1566221           1566221
## # ... with 32 more rows
```