Exploratory Data Analysis

```
library(tidyverse)
```

```
Warning: package 'ggplot2' was built under R version 4.3.3

Warning: package 'purrr' was built under R version 4.3.3

-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.4
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
votes <- read_csv("data/cleaned_g24_sov_by_svprec.csv") |>
mutate(
total_votes = demvote + repvote
)
```

```
Rows: 38657 Columns: 77
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (49): fips, svprec, svprec_key, election, geo_type, assaip01, assdem01, ...
dbl (28): county, addist, cddist, sddist, bedist, totreg, demreg, repreg, ai...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Question 1

Which precincts had the closest races between Democratic and Republican votes?

Answer 1

```
closest_races <- votes |>
mutate(margin = abs(demvote - repvote)) |>
arrange(margin) |>
select(svprec, demvote, repvote, margin) |>
slice_head(n = 10)

closest_races
```

```
# A tibble: 10 x 4
   svprec demvote repvote margin
   <chr>    <dbl>   <dbl>  <dbl>
 1 453900      17      17      0
 2 542300      12      12      0
 3 835830      67      67      0
 4 SOVTOT       0       0      0
 5 SOVTOT       0       0      0
 6 CP38A       37      37      0
 7 SOVTOT       0       0      0
 8 4302        21      21      0
 9 4807         6       6      0
10 SOVTOT       0       0      0
```

Question 2

Which precincts had the highest total voter turnout?

Answer 2

```
highest_turnout <- votes |>
arrange(desc(total_votes)) |>
select(svprec, total_votes, demvote, repvote) |>
slice_head(n = 10)
highest_turnout
```

```
# A tibble: 10 x 4
   svprec   total_votes demvote repvote
   <chr>          <dbl>   <dbl>   <dbl>
 1 CNTYTOT      3324096 2273160 1050936
 2 BE03_TOT     3324096 2273160 1050936
 3 CNTYTOT      1433017  838994  594023
 4 BE04_TOT     1433017  838994  594023
 5 CNTYTOT      1357499  686186  671313
 6 BE04_TOT     1357076  686026  671050
 7 CNTYTOT       916947  468094  448853
 8 BE04_TOT      916947  468094  448853
 9 CNTYTOT       731693  367662  364031
10 CNTYTOT       627170  379025  248145
```

Question 3

Which party won the most precincts?

Answer 3

```
precinct_winners <- votes |>
mutate(
winner = case_when(
demvote > repvote ~ "Democrat",
repvote > demvote ~ "Republican",
TRUE ~ "Tie"
)
) |>
count(winner)
precinct_winners
```

```
# A tibble: 3 x 2
  winner         n
  <chr>      <int>
1 Democrat   19626
2 Republican 13924
3 Tie         5107
```