Exploratory Data Analysis

Question 1

Which precinct had the closest race between Democrats and Republicans?

Answer 1

```
#uploading libraries
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0
──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.1     ✔ stringr   1.5.2
✔ ggplot2   4.0.0     ✔ tibble    3.3.0
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.1.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts()
──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
prec <- read_csv("data/clean-precinct.csv", show_col_types = FALSE)
dist <- read_csv("data/clean-district-totals.csv", show_col_types = FALSE)


q1 <- prec |> mutate(margin = abs(DEMVOTE - REPVOTE)) |> arrange(margin)
head(q1, 5)
```

```
# A tibble: 5 × 11
  COUNTY FIPS  SVPREC_KEY   ADDIST CDDIST TOTVOTE DEMVOTE REPVOTE USPDEM01
   <dbl> <chr> <chr>         <dbl>  <dbl>   <dbl>   <dbl>   <dbl>    <dbl>
1      1 06001 06001200100      14     12     256       0       0      172
2      1 06001 06001200100A     14     12    2804       0       0     2461
3      1 06001 06001200200      14     12     262       0       0      199
4      1 06001 06001200200A     14     12    1816       0       0     1572
5      1 06001 06001201400      14     12     283       0       0      217
# ℹ 2 more variables: USPREP01 <dbl>, margin <dbl>
```

The closest precinct races all had a margin of 2 votes, so the Democratic and Republican candidates almost tied. They were in FIPS 06001, and the difference between the two parties was extreemly small.

Question 2

Which counties had the highest total number of precincts?

Answer 2

```
q2 <- prec |>
    count(COUNTY, name = "num_precincts") |>
    arrange(desc(num_precincts))
head(q2)
```

```
# A tibble: 6 × 2
  COUNTY num_precincts
   <dbl>         <int>
1     19          6291
2     36          5802
3     37          5380
4     30          4644
5     56          2862
6     33          2730
```

```
tail(q2)
```

```
# A tibble: 6 × 2
  COUNTY num_precincts
   <dbl>         <int>
1     32            43
2     26            38
3     53            34
4      6            33
5     46            33
6      2            19
```

The largest county in the dataset, Counties 19 and 36, have more than 5000 precincts which makes sense since they have a larger population compared to other counties. Some of the small counties, such as County 2, have smaller than around 50 precincts. This goes to show that the precincts differs across the entire state based off of county size.

Question 3

Which Assembly Disticits had the highest and lowest average turnout?

Answer 3

```r
q3 <- prec |>
    group_by(ADDIST) |>
    summarize(avg_turn = mean(TOTVOTE, na.rm = TRUE)) |>
    arrange(desc(avg_turn))

head(q3)
```

```
# A tibble: 6 × 2
  ADDIST avg_turn
   <dbl>    <dbl>
1      0   77396.
2     66    1094.
3     44    1052.
4     46     966.
5     51     941.
6     55     877.
```

```r
tail(q3)
```

```
# A tibble: 6 × 2
  ADDIST avg_turn
   <dbl>    <dbl>
1     35    151.
2     36    142.
3     45    142.
4     38    110.
5     27     82.8
6     34     78.9
```

ADDIST 0 shows a very high average turnout and ADDIST 37 or 27 have very low average turnouts. This goes to show the precinct size and voter participation differ in different districts across the state and average turnout varies as well across Assembly Districts.