Exploratory Data Analysis

Question 1

How many districts are D vs R, D vs D, and R vs R?

Answer 1

```
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
```

```
library(dplyr)
g24_house <- read_csv("data/g24_house_clean.csv",
show_col_types = FALSE)
district_sums <- g24_house %>%
  group_by(cddist) %>%
  summarise(
    dem_votes = sum(house_dem, na.rm = TRUE),
    rep_votes = sum(house_rep, na.rm = TRUE)
  ) %>%
  mutate(
    race_type = case_when(
      dem_votes > 0 & rep_votes > 0 ~ "D vs R",
      dem_votes > 0 & rep_votes == 0 ~ "D vs D",
      dem_votes == 0 & rep_votes > 0 ~ "R vs R",
      TRUE ~ "No major party votes"
    )
  )

district_sums %>% count(race_type)
```

```
# A tibble: 3 × 2
  race_type     n
  <chr>     <int>
1 D vs D        4
2 D vs R       48
3 R vs R        1
```

Most congressional districts were competitive (D vs R), but several had same-party matchups (D vs D or R vs R). D vs D = 4, D vs R = 48, R vs R = 1

Question 2

Which districts were the closest races?

Answer 2

```
library(tidyverse)
library(dplyr)
closest <- district_sums |>
  mutate(
    total = dem_votes + rep_votes,
    dem_share = dem_votes / total,
    margin = abs(dem_share - 0.5)
  ) |>
  filter(total > 0) |>
  arrange(margin)

head(closest, 5)
```

```
# A tibble: 5 × 7
  cddist dem_votes rep_votes race_type  total dem_share  margin
   <dbl>     <dbl>     <dbl> <chr>      <dbl>     <dbl>   <dbl>
1     45    158216    157522 D vs R    315738     0.501 0.00110
2     27    153942    145826 D vs R    299768     0.514 0.0135
3     47    181662    171393 D vs R    353055     0.515 0.0145
4     13     90617     85181 D vs R    175798     0.515 0.0155
5     41    171093    182893 D vs R    353986     0.483 0.0167
```

The closest district was district 45, the margin was 0.00110, dem_votes = 158216 while rep_votes = 15722. This showcases a close race

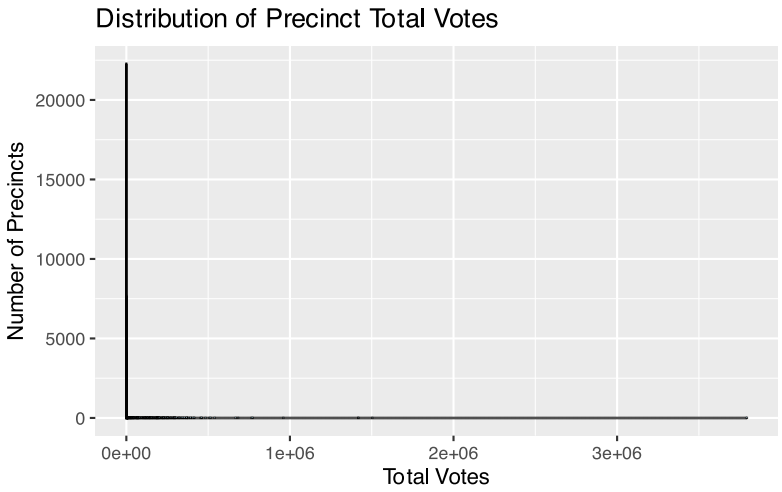
Question 3

What do precinct sizes look like? ### Answer 3

```
library(tidyverse)
library(dplyr)
summary(g24_house$totvote)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       1      91    2202     446 3793980
```

```
ggplot(g24_house, aes(totvote)) +
  geom_histogram(binwidth = 100, color = "black", fill = "lightblue") +
  labs(
    title = "Distribution of Precinct Total Votes",
    x = "Total Votes",
    y = "Number of Precincts"
  )
```



Most precincts cast between 0 and 3793980 votes. The mean for each district was 2202 votes cast.