Exploratory Data Analysis

```
library(tidyverse)

# Load cleaned data (or raw if not cleaned yet)
sov <- read_csv("data/g24_sov_by_g24_svprec.csv", show_col_types = FALSE)

# Ensure numeric vote columns
sov <- sov |>
  mutate(
    CNGDEM01 = as.numeric(CNGDEM01),
    CNGREP01 = as.numeric(CNGREP01),
    total_votes = CNGDEM01 + CNGREP01,
    margin = abs(CNGDEM01 - CNGREP01)
  )
```

Question 1

Which precincts had the closest congressional races in 2024?

Answer 1

Below are the top 5 closest precincts (margin = 0):

```
closest_precincts <- sov |>
  filter(total_votes > 0) |>
  arrange(margin) |>
  slice(1:5)

closest_precincts
```

```
# A tibble: 5 × 78
  COUNTY FIPS  SVPREC ADDIST SVPREC_KEY  ELECTION GEO_TYPE CDDIST SDDIST
BEDIST
   <dbl> <chr> <chr>   <dbl> <chr>       <chr>    <chr>     <dbl>  <dbl>
<dbl>
1      1 06001 453900     20 06001453900 g24      svprec       14      9
2
2      1 06001 542300     20 06001542300 g24      svprec       14      5
2
3      1 06001 835830     24 06001835830 g24      svprec       14     10
2
4      3 06005 CP38A       1 06005CP38A  g24      svprec        5      4
1
```

```
5      4 06007 4302        3 060074302   g24      svprec        1      1
1
# ℹ 68 more variables: TOTREG <dbl>, DEMREG <dbl>, REPREG <dbl>, AIPREG <dbl>,
#   GRNREG <dbl>, LIBREG <dbl>, NLPREG <dbl>, REFREG <dbl>, DCLREG <dbl>,
#   MSCREG <dbl>, TOTVOTE <dbl>, DEMVOTE <dbl>, REPVOTE <dbl>, AIPVOTE <dbl>,
#   GRNVOTE <dbl>, LIBVOTE <dbl>, NLPVOTE <dbl>, REFVOTE <dbl>, DCLVOTE <dbl>,
#   MSCVOTE <dbl>, PRCVOTE <dbl>, ABSVOTE <dbl>, ASSAIP01 <chr>,
#   ASSDEM01 <chr>, ASSDEM02 <chr>, ASSREP01 <chr>, ASSREP02 <chr>,
#   CNGDEM01 <dbl>, CNGDEM02 <chr>, CNGIND01 <chr>, CNGREP01 <dbl>, …
```

Question 2

Which congressional districts had the largest winning margins?

Answer 2

```
district_margins <- sov |>
  group_by(CDDIST) |>
  summarise(
    d_votes = sum(CNGDEM01, na.rm = TRUE),
    r_votes = sum(CNGREP01, na.rm = TRUE),
    margin = abs(d_votes - r_votes)
  ) |>
  filter(CDDIST != 0) |>
  arrange(desc(margin))

district_margins |> slice(1:5)
```

```
# A tibble: 5 × 4
  CDDIST d_votes r_votes margin
   <dbl>   <dbl>   <dbl>  <dbl>
1     11  274784   64312 210472
2     12  185144       0 185144
3     16  179463       0 179463
4      2  272384  106407 165977
5     37  160361       0 160361
```

Question 3

What is the distribution of precinct sizes (by total registration)?

Answer 3

```
summary(sov$TOTREG)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       0       0    3081     500 5745214
```