Exploratory Data Analysis

```r
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
```

```r
votes_precinct <- read_csv("data/g24_sov_by_g24_svprec.csv")
```

```
Rows: 51123 Columns: 76
── Column specification
────────────────────────────────────────────────────────
Delimiter: ","
chr (49): FIPS, SVPREC, SVPREC_KEY, ELECTION, GEO_TYPE, ASSAIP01,
ASSDEM01, ...
dbl (27): 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

What is the range and distribution of total votes cast across all precincts?

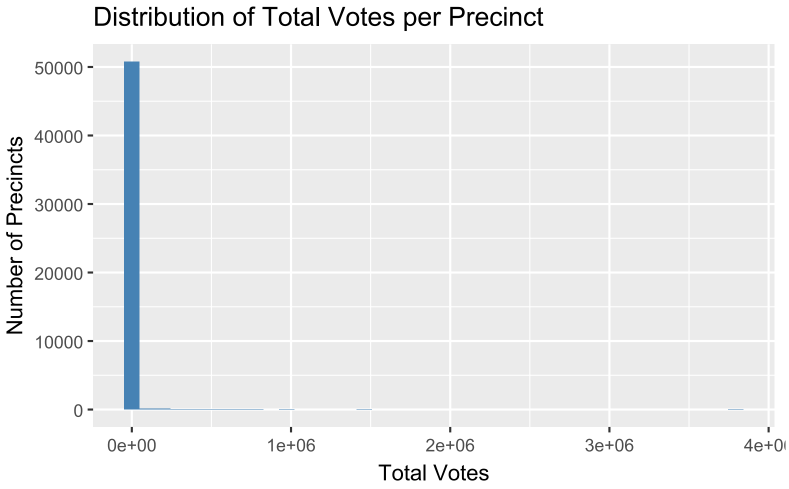
Answer 1

The total votes per precinct ranged from 0 to about 3.8 million, with a median of 91 and a highly skewed distribution where most precincts had relatively small totals.

```
summary(votes_precinct$TOTVOTE)
```

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

```
ggplot(votes_precinct, aes(x = TOTVOTE)) +
    geom_histogram(bins = 40, fill = "steelblue") +
    labs(title = "Distribution of Total Votes per Precinct", x = "Total
Votes", y = "Number of Precincts")
```



Question 2

Which counties or districts had the highest and lowest average turnout per precinct?

Answer 2

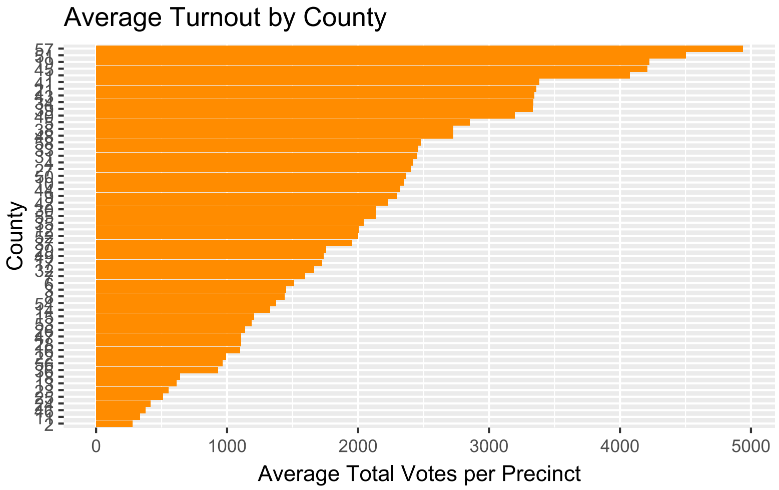
Across all counties, County 57 recorded the highest average turnout (around 4,937 votes per precinct), while other counties showed lower averages, reflecting variation in precinct size and voter participation.

```
votes_precinct |>
    group_by(COUNTY) |>
    summarize(mean_turnout = mean(TOTVOTE, na.rm = TRUE)) |>
    arrange(desc(mean_turnout))|>
    slice(1)
```

```
# A tibble: 1 × 2
  COUNTY mean_turnout
```

```
   <dbl>        <dbl>
1     57        4937.
```

```
votes_precinct |>
    group_by(COUNTY) |>
    summarize(mean_turnout = mean(TOTVOTE, na.rm = TRUE)) |>
    ggplot(aes(x = reorder(as.factor(COUNTY), mean_turnout), y =
mean_turnout)) +
    geom_col(fill = "darkorange") +
    coord_flip() +
    labs(title = "Average Turnout by County", x = "County", y = "Average Total
Votes per Precinct")
```



Question 3

Which Proposition had the closest statewide race between “Yes” and “No” Votes?

Answer 3

Among all statewide propositions, Proposition 32 had the closest race, with only about 1.4 million votes separating “Yes” and “No” totals, suggesting it was one of the most evenly contested measures.

```
prop_totals <- votes_precinct |>
    summarize(across(starts_with("PR_"), ~ sum(as.numeric(.), na.rm = TRUE)))
```

```
Warning: There were 20 warnings in `summarize()`.
The first warning was:
ℹ In argument: `across(starts_with("PR_"), ~sum(as.numeric(.), na.rm = TRUE))`.
Caused by warning:
```

```
! NAs introduced by coercion
ℹ Run `dplyr::last_dplyr_warnings()` to see the 19 remaining warnings.
```

```
prop_long <- prop_totals |>
    pivot_longer(cols = everything(),
                 names_to = c("Proposition", "Vote"),
                 names_pattern = "PR_(\\d+)_(.)",
                 values_to = "Votes") |>
    pivot_wider(names_from = Vote, values_from = Votes) |> mutate(Difference =
abs(Y - N))

arrange(prop_long, Difference)
```

```
# A tibble: 10 × 4
   Proposition        N        Y Difference
   <chr>          <dbl>    <dbl>      <dbl>
 1 32          53555144 52149025    1406119
 2 34          49659524 51461833    1802309
 3 6           54907683 48153959    6753724
 4 5           57421445 47039330   10382115
 5 2           43253760 61558433   18304673
 6 4           42405173 63198373   20793200
 7 33          62572749 41730224   20842525
 8 3           39388203 66176418   26788215
 9 35          33326868 70643366   37316498
10 36          33208544 71854516   38645972
```