Gerrymandering Metrics

Introduction

In this document I compute two standard gerrymandering metrics for the 2024 U.S. House elections in California:

- the **mean–median score**, and
- the **efficiency gap**.

I do this twice:

1. Using the **actual 2024 congressional district map**, and
2. Using the **proposed AB 604 congressional district map**, based on the area-weighted reallocation of votes created in `data-cleaning.qmd` (saved as `data/newmap-district-votes.csv`).

All calculations are based on Democratic vs. Republican vote totals in the congressional race, using the `cngdem01` and `cngrep01` columns.

```
library(tidyverse)
library(janitor)
```

Helper Functions

To keep things readable, I define a couple of small helper functions for the two gerrymandering metrics.

```
# Given vectors of Democratic and Republican votes by district,
# return the mean–median score for Democratic vote share.
mean_median_score <- function(d_votes, r_votes) {
  total <- d_votes + r_votes
  share <- d_votes / total
  mean(share, na.rm = TRUE) - median(share, na.rm = TRUE)
}

# Given vectors of Democratic and Republican votes by district,
# return the efficiency gap (Democratic perspective).
efficiency_gap <- function(d_votes, r_votes) {
  total <- d_votes + r_votes
  threshold <- floor(total / 2) + 1      # votes needed to win
  d_wins <- d_votes > r_votes

  wasted_d <- ifelse(d_wins, d_votes - threshold, d_votes)
  wasted_r <- ifelse(d_wins, r_votes, r_votes - threshold)
```

```
  sum(wasted_d - wasted_r, na.rm = TRUE) / sum(total, na.rm = TRUE)
}
```

2024 Election Results and the 2024 District Map

District-Level Votes Under the 2024 Map

First I load the cleaned SV-precinct data (clean-sov.csv) and aggregate Democratic and Republican congressional votes by 2024 congressional district (cddist).

```
sov <- read_csv("data/clean-sov.csv", show_col_types = FALSE) |>
  clean_names()

district_2024 <- sov |>
  group_by(cddist) |>
  summarise(
    d_votes = sum(cngdem01, na.rm = TRUE),
    r_votes = sum(cngrep01, na.rm = TRUE),
    .groups = "drop"
  ) |>
  # Drop districts with no two-party votes (just in case)
  filter(d_votes + r_votes > 0)

district_2024
```

```
# A tibble: 53 × 3
   cddist  d_votes  r_votes
    <dbl>    <dbl>    <dbl>
 1      0 52843210 34247705
 2      1   110472   208150
 3      2   272384   106407
 4      3   187960   233895
 5      4   227321   114644
 6      5   134467   214223
 7      6   165386   121625
 8      7   197361    98273
 9      8   201756    70932
10      9   130093   121006
# i 43 more rows
```

Seats Won Under the 2024 Map

```
seats_dem_2024 <- sum(district_2024$d_votes > district_2024$r_votes, na.rm =
TRUE)
seats_rep_2024 <- sum(district_2024$r_votes > district_2024$d_votes, na.rm =
TRUE)
```

```
tibble(
  plan    = "2024 Map",
  dem_seats = seats_dem_2024,
  rep_seats = seats_rep_2024
)
```

```
# A tibble: 1 × 3
  plan      dem_seats rep_seats
  <chr>         <int>     <int>
1 2024 Map         44         9
```

Mean–Median Score for the 2024 Map

The **mean–median score** is:

$$[= \{s\} - (s)]$$

where (s) is the Democratic vote share in each district. Positive values indicate that the mean share is higher than the median share, which can be consistent with a map that "packs" Democrats into fewer districts.

```
mm_2024 <- with(district_2024,
                mean_median_score(d_votes, r_votes))

mm_2024
```

```
[1] -0.01163046
```

Efficiency Gap for the 2024 Map

The **efficiency gap** compares the total wasted votes for each party:

- In districts Democrats win: Democratic wasted votes are (- (T + 1)) and Republicans waste all their votes.
- In districts Republicans win: the reverse is true.

The efficiency gap (from the Democratic perspective) is:

$$[=]$$

```
eg_2024 <- with(district_2024,
                efficiency_gap(d_votes, r_votes))

eg_2024
```

```
[1] -0.2596726
```

2024 Election Results and the Proposed 2025 District Map (AB 604)

For the proposed AB 604 map, I use the area-weighted results computed in `data-cleaning.qmd` and saved as `data/newmap-district-votes.csv`. This file contains one row per new district with the columns:

- `new_cdd` – new congressional district ID
- `d_votes` – estimated Democratic votes
- `r_votes` – estimated Republican votes

```
newmap_district_votes <- read_csv("data/newmap-district-votes.csv",
                                  show_col_types = FALSE)

newmap_district_votes
```

```
# A tibble: 52 × 3
   new_cdd d_votes r_votes
   <chr>     <dbl>   <dbl>
 1 01      189840. 160473.
 2 02      252190. 171693.
 3 03      197348. 174330.
 4 04      190249. 144307.
 5 05      117927. 205615.
 6 06      170484. 151681.
 7 07      194068. 152204.
 8 08      221721. 112912.
 9 09      169805. 119444.
10 10      248627. 127085.
# i 42 more rows
```

Seats Won Under the Proposed AB 604 Map

```
seats_dem_ab604 <- sum(newmap_district_votes$d_votes >
newmap_district_votes$r_votes,
                       na.rm = TRUE)
seats_rep_ab604 <- sum(newmap_district_votes$r_votes >
newmap_district_votes$d_votes,
                       na.rm = TRUE)

tibble(
  plan    = "AB 604 Map",
  dem_seats = seats_dem_ab604,
```

```
    rep_seats = seats_rep_ab604
)
```

```
# A tibble: 1 × 3
  plan       dem_seats rep_seats
  <chr>          <int>     <int>
1 AB 604 Map        47         5
```

Mean–Median Score for the AB 604 Map

```
mm_ab604 <- with(newmap_district_votes,
                 mean_median_score(d_votes, r_votes))

mm_ab604
```

```
[1] 0.03103683
```

Efficiency Gap for the AB 604 Map

```
eg_ab604 <- with(newmap_district_votes,
                 efficiency_gap(d_votes, r_votes))

eg_ab604
```

```
[1] -0.2057942
```

Comparison of 2024 Map vs. AB 604 Map

Finally, I summarize the results for both the current (2024) map and the proposed AB 604 map side by side.

```
comparison <- tibble(
  plan           = c("2024 Map", "AB 604 Map"),
  dem_seats      = c(seats_dem_2024, seats_dem_ab604),
  rep_seats      = c(seats_rep_2024, seats_rep_ab604),
  mean_median    = c(mm_2024, mm_ab604),
  efficiency_gap = c(eg_2024, eg_ab604)
)

comparison
```

```
# A tibble: 2 × 5
  plan       dem_seats rep_seats mean_median efficiency_gap
```

	<chr>	<int>	<int>	<dbl>	<dbl>
1	2024 Map	44	9	-0.0116	-0.260
2	AB 604 Map	47	5	0.0310	-0.206