Gerrymandering Metrics

2024 Election Results and the 2024 District Map

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.2     ✔ 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(scales)
```

```
Attaching package: 'scales'

The following object is masked from 'package:purrr':

    discard

The following object is masked from 'package:readr':

    col_factor
```

```
precinct_long <- read_csv("data/precinct_clean_long.csv")
```

```
Rows: 511230 Columns: 76
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (46): fips, svprec, svprec_key, election, geo_type, prsaip01, prsdem01, ...
```

```
dbl (30): 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.
```

```
congress <- precinct_long |>
  filter(race_type == "Congress") |>
  group_by(district, candidate_party) |>
  summarise(total_votes = sum(votes, na.rm = TRUE), .groups = "
  pivot_wider(names_from = candidate_party, values_from = total_
  mutate(
    total = DEM + REP,
    dem_share = DEM / total
  ) |>
  filter(total > 0)  # exclude empty districts

mean_share <- mean(congress$dem_share, na.rm = TRUE)
median_share <- median(congress$dem_share, na.rm = TRUE)
mm_diff <- mean_share - median_share

cat("Mean Democratic vote share:", round(mean_share, 3), "\n")
```

```
Mean Democratic vote share: 0.632
```

```
cat("Median Democratic vote share:", round(median_share, 3), "\n
```

```
Median Democratic vote share: 0.637
```

```
cat("Mean–Median Difference:", round(mm_diff, 3), "\n")
```

```
Mean–Median Difference: -0.005
```

```
congress <- congress |>
  mutate(
    dem_wasted = if_else(DEM > REP, DEM - (total / 2), DEM),
    rep_wasted = if_else(REP > DEM, REP - (total / 2), REP)
  )

total_dem_wasted <- sum(congress$dem_wasted, na.rm = TRUE)
total_rep_wasted <- sum(congress$rep_wasted, na.rm = TRUE)
```

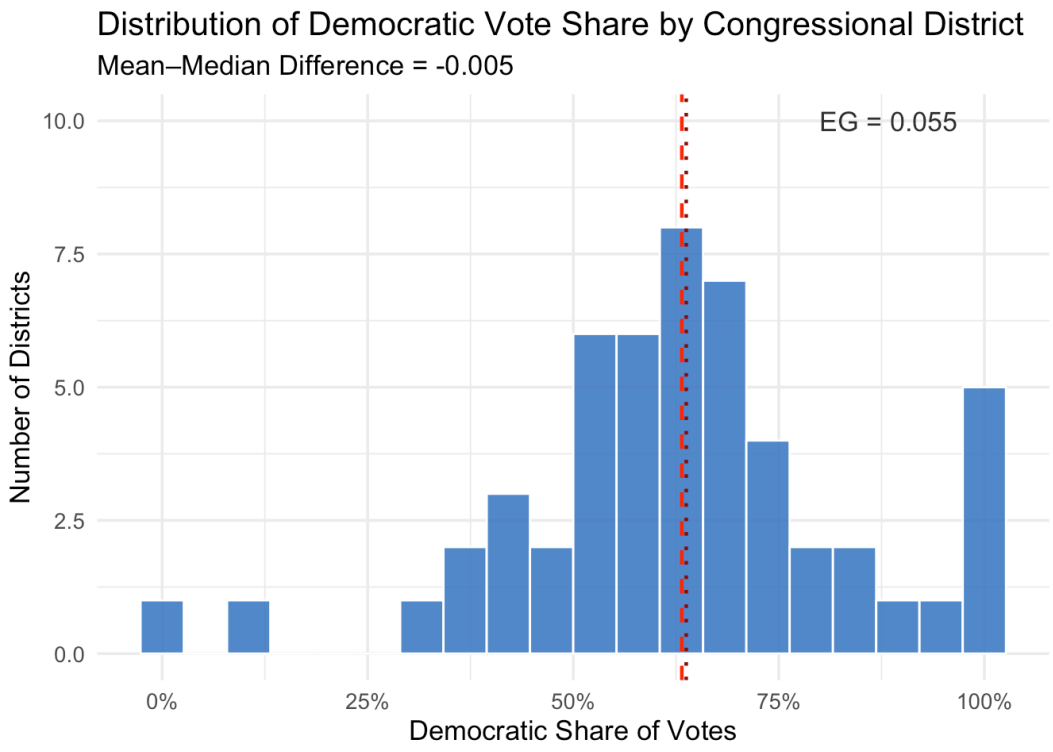
```
total_votes <- sum(congress$total, na.rm = TRUE)

efficiency_gap <- (total_rep_wasted - total_dem_wasted) / total_

cat("Efficiency Gap:", round(efficiency_gap, 3), "\n")
```

```
Efficiency Gap: 0.055
```

```
ggplot(congress, aes(x = dem_share)) +
  geom_histogram(bins = 20, fill = "#2E74C0", color = "white", a
  geom_vline(xintercept = mean_share, linetype = "dashed", colo
  geom_vline(xintercept = median_share, linetype = "dotted", co
  annotate("text", x = 0.8, y = 10,
           label = paste0("EG = ", round(efficiency_gap, 3)),
           hjust = 0, size = 4.5, color = "gray20") +
  scale_x_continuous(labels = scales::percent_format(accuracy =
  labs(
    title = "Distribution of Democratic Vote Share by Congressio
    subtitle = paste0("Mean–Median Difference = ", round(mm_dif
    x = "Democratic Share of Votes",
    y = "Number of Districts"
  ) +
  theme_minimal(base_size = 13)
```



2024 Election Results and the proposed 2025 District Map

```
library(sf)
```

```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is
TRUE
```

```
sr_geo <- st_read("data/shapefiles/srprecinct_cleaned.shp")
```

```
Reading layer `srprecinct_cleaned' from data source

`/Users/timroth/Documents/UC_Berkeley/Cal_Senior/Senior_Fall/st
at133/gerrymandering-
timrothtr/data/shapefiles/srprecinct_cleaned.shp'
  using driver `ESRI Shapefile'

Warning in CPL_read_ogr(dsn, layer, query,
as.character(options), quiet, : GDAL
Message 1:
/Users/timroth/Documents/UC_Berkeley/Cal_Senior/Senior_Fall/sta
t133/gerrymandering-
timrothtr/data/shapefiles/srprecinct_cleaned.shp
contains polygon(s) with rings with invalid winding order.
Autocorrecting them,
but that shapefile should be corrected using ogr2ogr for
example.

Simple feature collection with 26993 features and 81 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -380101 ymin: -605327 xmax: 540038 ymax:
450447
Projected CRS: NAD83 / California Albers
```

```
ab604_shp <- st_read("data/shapefiles/ab604_cleaned.shp")
```

```
Reading layer `ab604_cleaned' from data source

`/Users/timroth/Documents/UC_Berkeley/Cal_Senior/Senior_Fall/st
at133/gerrymandering-
timrothtr/data/shapefiles/ab604_cleaned.shp'
```

```
  using driver `ESRI Shapefile'
Simple feature collection with 52 features and 15 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -380102.2 ymin: -605326.6 xmax: 540036.5
ymax: 450447.3
Projected CRS: NAD83 / California Albers
```

```
votes_sr <- read_csv("data/srprecinct_votes_cleaned.csv")
```

```
Rows: 25167 Columns: 76

── Column specification
────────────────────────────────────────────────────────
Delimiter: ","
chr (48): srprec, election, srprec_key, geo_type, assaip01,
assdem01, assdem...
dbl (28): county, fips, addist, cddist, sddist, bedist, totreg,
demreg, repr...

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.
```

```
votes_sr <- votes_sr |>
  mutate(across(matches("dem|rep", ignore.case = TRUE), ~ suppre
  mutate(
    DEM = rowSums(across(matches("dem", ignore.case = TRUE)), na
    REP = rowSums(across(matches("rep", ignore.case = TRUE)), na
    total = DEM + REP
  )

sr_geo <- st_make_valid(sr_geo) |> st_transform(3310)
ab604_shp <- st_make_valid(ab604_shp) |> st_transform(3310)

sr_geo <- sr_geo |>
  left_join(votes_sr, by = "srprec")   # check the column name
```

```
Warning in sf_column %in% names(g): Detected an unexpected
many-to-many relationship between `x` and `y`.
ℹ Row 11 of `x` matches multiple rows in `y`.
ℹ Row 174 of `y` matches multiple rows in `x`.
```

```
ℹ If a many-to-many relationship is expected, set `relationship
=
  "many-to-many"` to silence this warning.
```

```
votes_ab604 <- st_join(sr_geo, ab604_shp, join = st_intersects,
```

```
district_results_ab604 <- votes_ab604 |>
  st_drop_geometry() |>
  group_by(DISTRICT) |>
  summarise(
    DEM = sum(DEM, na.rm = TRUE),
    REP = sum(REP, na.rm = TRUE),
    total = sum(total, na.rm = TRUE),
    .groups = "drop"
  ) |>
  mutate(
    dem_share = DEM / (DEM + REP)
  )


mm_diff_ab604 <- mean(district_results_ab604$dem_share, na.rm =
                 median(district_results_ab604$dem_share, na.rm

district_results_ab604 <- district_results_ab604 |>
  mutate(
    dem_wasted = if_else(DEM > REP, DEM - (total / 2), DEM),
    rep_wasted = if_else(REP > DEM, REP - (total / 2), REP)
  )

eff_gap_ab604 <- (sum(district_results_ab604$dem_wasted) -
                  sum(district_results_ab604$rep_wasted)) /
                  sum(district_results_ab604$total)

cat("Mean–Median Difference (AB604): ", round(mm_diff_ab604, 3),
```

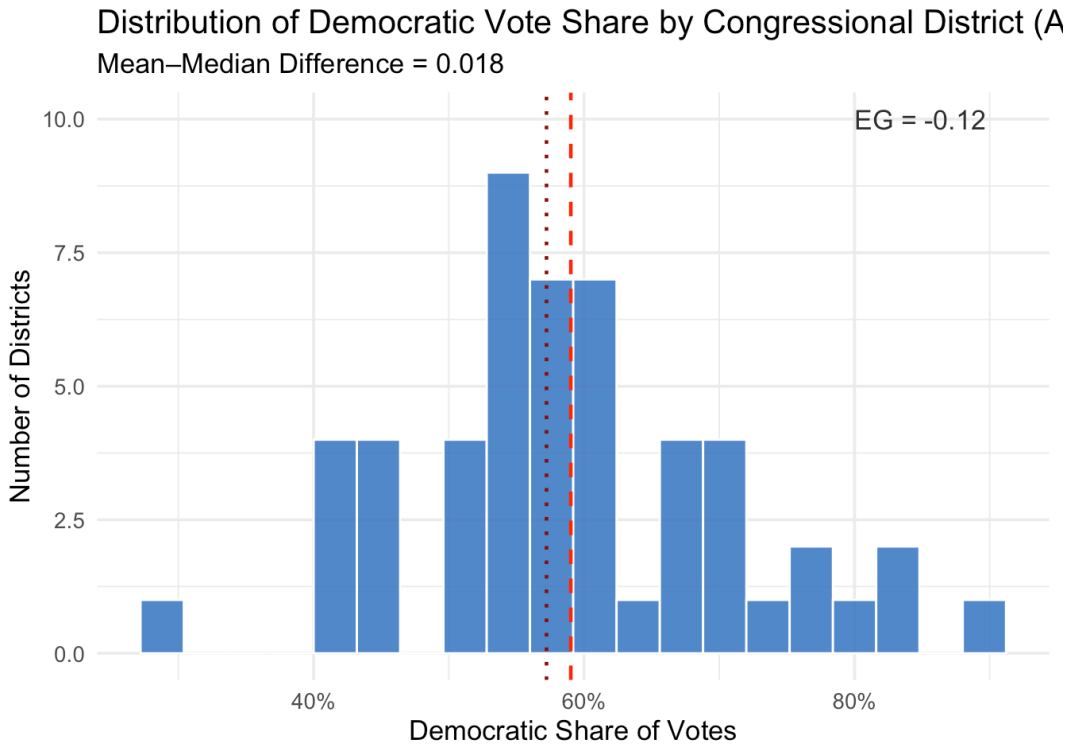
```
Mean–Median Difference (AB604):  0.018
```

```
cat("Efficiency Gap (AB604): ", round(eff_gap_ab604, 3), "\n")
```

```
Efficiency Gap (AB604):  -0.12
```

```
ggplot(district_results_ab604, aes(x = dem_share)) +
  geom_histogram(bins = 20, fill = "#2E74C0", color = "white", a
```

```
  geom_vline(xintercept = mean(district_results_ab604$dem_share,
             linetype = "dashed", color = "red", linewidth = 0.8
  geom_vline(xintercept = median(district_results_ab604$dem_sha
             linetype = "dotted", color = "darkred", linewidth =
  annotate("text", x = 0.8, y = 10,
           label = paste("EG =", round(eff_gap_ab604, 3)),
           hjust = 0, size = 4.5, color = "gray20") +
  scale_x_continuous(labels = scales::percent_format(accuracy =
  labs(
    title = "Distribution of Democratic Vote Share by Congressio
    subtitle = paste("Mean–Median Difference =", round(mm_diff_a
    x = "Democratic Share of Votes",
    y = "Number of Districts"
  ) +
  theme_minimal(base_size = 13)
```




```
write_csv(congress, "data/district_results_2024.csv")
write_csv(district_results_ab604, "data/district_results_ab604.c

metrics_summary <- tibble(
  map_version = c("2024 Actual", "AB604 Proposed"),
  mean_median_diff = c(mm_diff, mm_diff_ab604),
  efficiency_gap = c(efficiency_gap, eff_gap_ab604),
```

```
  prop_dem_wins = c(
    mean(congress$dem_share > 0.5, na.rm = TRUE),
    mean(district_results_ab604$dem_share > 0.5, na.rm = TRUE)
  )
)
write_csv(metrics_summary, "data/metrics_summary.csv")
```