Gerrymandering Metrics

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

udunits database from /Library/Frameworks/R.framework/Versions/4.5-arm64/
Resources/library/units/share/udunits/udunits2.xml

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.
```

```
Warning: There were 44 warnings in `mutate()`.
The first warning was:
ℹ In argument: `across(...)`.
Caused by warning:
! NAs introduced by coercion
ℹ Run `dplyr::last_dplyr_warnings()` to see the 43 remaining warnings.
```

Mean-Median Score and Efficiency Gap

```r
# Aggregate precinct votes to congressional districts
district_votes <- precinct_election |>
  group_by(CDDIST) |>
  summarize(
    dem_votes = sum(CNGDEM01 + CNGDEM02, na.rm = TRUE),
    rep_votes = sum(CNGREP01 + CNGREP02, na.rm = TRUE)) |>
  ungroup() |>
  mutate(
    total_votes = dem_votes + rep_votes,
    dem_share = dem_votes / total_votes)

# Mean-Median score
mean_median_score <- mean(district_votes$dem_share) -
median(district_votes$dem_share)

# Function to calculate wasted votes for efficiency gap
calc_wasted <- function(dem, rep) {
  votes_needed <- floor((dem + rep) / 2) + 1
  dem_wasted <- ifelse(dem > rep, dem - votes_needed, dem)
  rep_wasted <- ifelse(rep > dem, rep - votes_needed, rep)
  return(tibble(dem_wasted = dem_wasted, rep_wasted = rep_wasted))}

# Compute wasted votes per district
wasted_votes <- district_votes |>
  rowwise() |>
  mutate(
    dem_wasted = calc_wasted(dem_votes, rep_votes)$dem_wasted,
    rep_wasted = calc_wasted(dem_votes, rep_votes)$rep_wasted) |>
  ungroup()

# Efficiency gap
total_dem_wasted <- sum(wasted_votes$dem_wasted, na.rm = TRUE)
total_rep_wasted <- sum(wasted_votes$rep_wasted, na.rm = TRUE)
total_votes <- sum(district_votes$total_votes, na.rm = TRUE)

efficiency_gap <- (total_rep_wasted - total_dem_wasted) / total_votes

# Print results
mean_median_score
```

```
[1] -0.01169593
```

```r
efficiency_gap
```

```
[1] 0.1934897
```

The −0.01 mean-median score shows a very slight advantage for Democrats, as their votes are slightly more packed into districts and thus less spread out to try to win more districts. However a 1% mean-median score is so miniscule that there is definitely no gerrymandering going on and very little potential for advantage for Democrats.

The efficiency gap of 0.19 shows a very large mismatch in wasted votes for Republicans and Democrats. Republicans wasted 19% more votes than did Democrats, suggesting Republicans could have been packed into a few districts that they won by a lot or that they barely lost a bunch of races, wasting a lot of votes. My guess is the second option. Given that California is a heavy Democrat-majority state, I can see how Republicans might have lost a lot of congressional races (many being close), causing them to "waste" a lot of votes. However, 19% is so large that this could hint at potential gerrymandering.

2024 Election Results and the 2024 District Map

Part 5: Re-running the 2024 Election (Approach A)

```
setwd("~/Documents/stat133/gerrymandering-carsonkuehnert")

# Load + clean/fix data
# sr precinct vote data
sr_votes <- read.csv("data/state_g24_sov_data_by_g24_srprec.csv")

vote_cols <- c(
  "ASSAIP01", "ASSDEM01", "ASSDEM02", "ASSREP01", "ASSREP02",
  "CNGDEM01", "CNGDEM02", "CNGIND01", "CNGREP01", "CNGREP02",
  "PRSAIP01", "PRSDEM01", "PRSGRN01", "PRSLIB01", "PRSPAF01", "PRSREP01",
  "PR_2_N", "PR_2_Y", "PR_32_N", "PR_32_Y", "PR_33_N", "PR_33_Y",
  "PR_34_N", "PR_34_Y", "PR_35_N", "PR_35_Y", "PR_36_N", "PR_36_Y",
  "PR_3_N", "PR_3_Y", "PR_4_N", "PR_4_Y", "PR_5_N", "PR_5_Y", "PR_6_N",
"PR_6_Y",
  "SENDEM01", "SENDEM02", "SENREP01", "SENREP02",
  "USPDEM01", "USPREP01", "USSDEM01", "USSREP01")

sr_votes <- sr_votes |>
  mutate(across(all_of(vote_cols), ~ as.numeric(gsub("\\*", "", .x))))

# sr shapefiles
sr_shp <- st_read("data/shapefiles/srprec_state_g24_v01_shp/
srprec_state_g24_v01_shp.shp")
```

```
Reading layer `srprec_state_g24_v01_shp' from data source
  `/Users/carsonkuehnert/Documents/stat133/gerrymandering-carsonkuehnert/data/
shapefiles/srprec_state_g24_v01_shp/srprec_state_g24_v01_shp.shp'
  using driver `ESRI Shapefile'
```

```
Warning in CPL_read_ogr(dsn, layer, query, as.character(options), quiet, :
GDAL
Message 1:
/Users/carsonkuehnert/Documents/stat133/gerrymandering-carsonkuehnert/data/
shapefiles/srprec_state_g24_v01_shp/srprec_state_g24_v01_shp.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 24224 features and 6 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -124.482 ymin: 32.52883 xmax: -114.1312 ymax: 42.0095
Geodetic CRS:  NAD83
```

```
sr_shp <- sr_shp |>
    st_transform(3310) |>
    st_set_precision(1) |>
    st_make_valid()

# congressional map shapefiles
ab604_shp <- st_read("data/shapefiles/AB604/AB604.shp")
```

```
Reading layer `AB604' from data source
  `/Users/carsonkuehnert/Documents/stat133/gerrymandering-carsonkuehnert/data/
shapefiles/AB604/AB604.shp'
  using driver `ESRI Shapefile'
Simple feature collection with 52 features and 15 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -13857270 ymin: 3832931 xmax: -12705030 ymax: 5162404
Projected CRS: WGS 84 / Pseudo-Mercator
```

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

```
# Intersect precincts with new districts
sr_ab604 <- st_intersection(
  sr_shp |> left_join(sr_votes, by = "SRPREC"),
  ab604_shp)
```

```
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 344 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.
```

```
Warning: attribute variables are assumed to be spatially constant throughout
all geometries
```

```r
# Calculate area fractions per SR precinct
sr_ab604 <- sr_ab604 |>
  group_by(SRPREC) |>
  mutate(
    total_precinct_area = sum(st_area(geometry)),
    area_fraction = as.numeric(st_area(geometry) / total_precinct_area)) |>
  ungroup()
```

```r
# Allocate votes proportionally for all columns
vote_cols <- c(
  "ASSAIP01", "ASSDEM01", "ASSDEM02", "ASSREP01", "ASSREP02",
  "CNGDEM01", "CNGDEM02", "CNGIND01", "CNGREP01", "CNGREP02",
  "PRSAIP01", "PRSDEM01", "PRSGRN01", "PRSLIB01", "PRSPAF01", "PRSREP01",
  "PR_2_N", "PR_2_Y", "PR_32_N", "PR_32_Y", "PR_33_N", "PR_33_Y",
  "PR_34_N", "PR_34_Y", "PR_35_N", "PR_35_Y", "PR_36_N", "PR_36_Y",
  "PR_3_N", "PR_3_Y", "PR_4_N", "PR_4_Y", "PR_5_N", "PR_5_Y", "PR_6_N",
"PR_6_Y",
  "SENDEM01", "SENDEM02", "SENREP01", "SENREP02",
  "USPDEM01", "USPREP01", "USSDEM01", "USSREP01")

sr_ab604 <- sr_ab604 |>
  mutate(across(all_of(vote_cols), ~ .x * area_fraction, .names = "{.col}
_alloc"))
```

Part 6: Calculating Gerrymandering Again

```r
# Sum allocated votes by new district
district_votes <- sr_ab604 |>
  st_drop_geometry() |>
  group_by(DISTRICT) |>
  summarize(across(ends_with("_alloc"), sum, na.rm = TRUE))
```

```
Warning: There was 1 warning in `summarize()`.
ℹ In argument: `across(ends_with("_alloc"), sum, na.rm = TRUE)`.
```

```
ℹ In group 1: `DISTRICT = "01"`.
Caused by warning:
! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
Supply arguments directly to `.fns` through an anonymous function instead.

  # Previously
  across(a:b, mean, na.rm = TRUE)

  # Now
  across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
# Calculate votes + party shares
district_votes <- district_votes |>
  mutate(
    total_pres = PRSDEM01_alloc + PRSREP01_alloc + PRSGRN01_alloc +
PRSLIB01_alloc + PRSPAF01_alloc + PRSAIP01_alloc,
    dem_share = PRSDEM01_alloc / total_pres,
    rep_share = PRSREP01_alloc / total_pres)

# Mean-median score
mean_median_score_new <- mean(district_votes$dem_share) -
median(district_votes$dem_share)

# Function to calculate wasted votes for efficiency gap
calc_wasted <- function(dem, rep) {
  votes_needed <- floor((dem + rep) / 2) + 1
  dem_wasted <- ifelse(dem > rep, dem - votes_needed, dem)
  rep_wasted <- ifelse(rep > dem, rep - votes_needed, rep)
  return(tibble(dem_wasted = dem_wasted, rep_wasted = rep_wasted))}

# Compute wasted votes per district
wasted_votes_new <- district_votes |>
  rowwise() |>
  mutate(
    dem_wasted = calc_wasted(PRSDEM01_alloc, PRSREP01_alloc)$dem_wasted,
    rep_wasted = calc_wasted(PRSDEM01_alloc, PRSREP01_alloc)$rep_wasted) |>
  ungroup()

# Efficiency gap
total_dem_wasted_new <- sum(wasted_votes_new$dem_wasted, na.rm = TRUE)
total_rep_wasted_new <- sum(wasted_votes_new$rep_wasted, na.rm = TRUE)
total_votes_new <- sum(district_votes$total_pres, na.rm = TRUE)

efficiency_gap_new <- (total_rep_wasted_new - total_dem_wasted_new) /
total_votes_new
```

```
# Print results
mean_median_score_new
```

```
[1] 0.02170814
```

```
efficiency_gap_new
```

```
[1] 0.1880199
```

The mean-median score is 0.02, actually showing a slight favoring to Republicans, as the mean votes of democrats is slightly higher than their median, suggesting democrats are maybe a tiny bit packed into districts. However, 0.02 is so small that this suggests very little advantage to Republicans. However, an efficiency gap of 0.19 suggests that Republicans are quite packed into a few districts or cracked into many districts, wasting many votes. ### 2024 Election Results and the proposed 2025 District Map

```
# Aggregate 2024 results by district
district_votes_2024 <- precinct_election |>
group_by(CDDIST) |>
summarize(
dem_votes = sum(CNGDEM01 + CNGDEM02, na.rm = TRUE),
rep_votes = sum(CNGREP01 + CNGREP02, na.rm = TRUE)
) |>
ungroup() |>
mutate(
total_votes = dem_votes + rep_votes,
dem_share = dem_votes / total_votes,
rep_share = rep_votes / total_votes,
winner = ifelse(dem_votes > rep_votes, "Democrat", "Republican"))

# AB 604 results by district
district_votes_ab604 <- district_votes |>
mutate(
winner = ifelse(PRSDEM01_alloc > PRSREP01_alloc, "Democrat", "Republican"))

#Summary Statistics table
summary_table <- tibble(
Map = c("2024 Districts", "AB 604 Proposed"),
Seats_Dem = c(sum(district_votes_2024$winner == "Democrat"),
sum(district_votes_ab604$winner == "Democrat")),
Seats_Rep = c(sum(district_votes_2024$winner == "Republican"),
sum(district_votes_ab604$winner == "Republican")),
Mean_Median = c(mean(district_votes_2024$dem_share) -
median(district_votes_2024$dem_share),
```

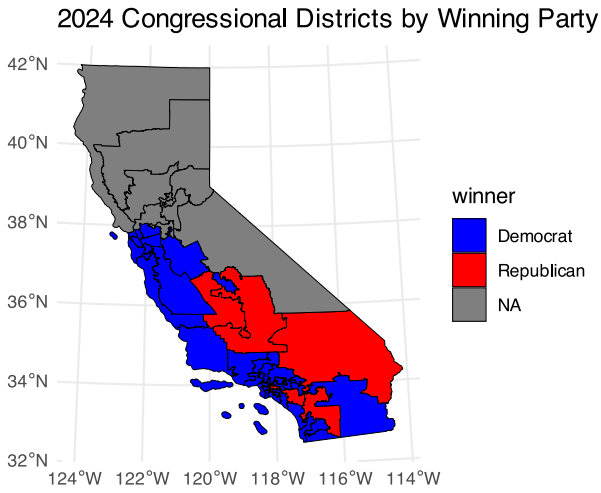
```
mean(district_votes_ab604$dem_share) -
median(district_votes_ab604$dem_share)),
Efficiency_Gap = c(efficiency_gap, efficiency_gap_new))

summary_table
```

```
# A tibble: 2 × 5
  Map             Seats_Dem Seats_Rep Mean_Median Efficiency_Gap
  <chr>               <int>     <int>       <dbl>          <dbl>
1 2024 Districts         44         9     -0.0117          0.193
2 AB 604 Proposed        47         5      0.0217          0.188
```

```
# 2024 Election results

district_votes_2024 <- district_votes_2024 |>
  mutate(CDDIST = as.character(CDDIST))
# Join the votes to the shapefile — start from the sf object
district_map_2024 <- ab604_shp |>
  st_transform(3310) |>
  left_join(district_votes_2024, by = c("DISTRICT" = "CDDIST")) |>
  mutate(
    winner = ifelse(dem_votes > rep_votes, "Democrat", "Republican"))
# Plot map
ggplot(district_map_2024) +
  geom_sf(aes(fill = winner), color = "black") +
  scale_fill_manual(values = c("Democrat" = "blue", "Republican" = "red")) +
  labs(title = "2024 Congressional Districts by Winning Party") +
  theme_minimal()
```



```r
# Proposed Map Hypothetical Election Results

district_votes_ab604 <- district_votes_ab604 |>
  mutate(DISTRICT = as.character(DISTRICT))
# Join allocated votes to the AB 604 shapefile
district_map_ab604 <- ab604_shp |>
  st_transform(3310) |>
  left_join(district_votes_ab604, by = c("DISTRICT" = "DISTRICT")) |>
  mutate(
    winner = ifelse(PRSDEM01_alloc > PRSREP01_alloc, "Democrat",
"Republican"))
# Plot map
ggplot(district_map_ab604) +
  geom_sf(aes(fill = winner), color = "black") +
  scale_fill_manual(values = c("Democrat" = "blue", "Republican" = "red")) +
  labs(title = "2024 Election Results by Proposed AB 604 Districts") +
  theme_minimal()
```