Gerrymandering Metrics – 2024 Map

Part 4 – Mean–Median score and Efficiency Gap for 2024 map

library(tidyverse)

1. Load cleaned precinct-level data

precincts24 <- read_csv("data/g24_precinct_clean.csv")

glimpse(precincts24)

2. Detect and standardize the congressional district column

Look for a column name that contains "cng" (case-insensitive)

but is NOT one of the vote columns like CNGDEMxx or CNGREPxx.

names_prec <- names(precincts24)

district_candidates <- names_prec[str_detect(names_prec, regex("cng", ignore_case = TRUE)) & ! str_detect(names_prec, regex("cngdem|cngrep", ignore_case = TRUE))]

district_candidates

if (length(district_candidates) == 0) { stop("Could not detect a congressional district column in precincts24. Please set it manually.") }

district_col <- district_candidates[1]

Rename that column to DIST so we can use it consistently

precincts24 <- precincts24 |> rename(DIST = all_of(district_col))

glimpse(precincts24)

3. Build district-level two-party results (Congress)

Assumes:

- CNGDEM01 = Democratic votes for Congress

- CNGREP01 = Republican votes for Congress

If your file uses slightly different names, adjust here.

district_results <- precincts24 |> group_by(DIST) |> summarise(dem_votes = sum(CNGDEM01, na.rm = TRUE), rep_votes = sum(CNGREP01, na.rm = TRUE)) |> mutate(total_two_party = dem_votes + rep_votes) |> # Drop any weird districts with no D/R votes at all filter(total_two_party > 0) |> mutate(dem_share = dem_votes / total_two_party)

district_results

4. Mean–Median score

MM = mean(dem_share) - median(dem_share)

mean_dem_share <- mean(district_results$dem_share, na.rm = TRUE) median_dem_share <- median(district_results$dem_share, na.rm = TRUE) mm_score <- mean_dem_share - median_dem_share

mean_dem_share median_dem_share mm_score

cat("–median score (mean - median Dem share):", round(mm_score, 4), "")

5. Efficiency Gap

For each district:

- If Democrats win:

Dem wasted = dem_votes - (total_two_party / 2)

Rep wasted = rep_votes

- If Republicans win:

Rep wasted = rep_votes - (total_two_party / 2)

Dem wasted = dem_votes

Then:

EG = (total_wasted_D - total_wasted_R) / total_two_party_votes

Here, EG > 0 means more wasted Dem votes → Republican advantage.

district_wasted <- district_results |> mutate(dem_wasted = if_else(dem_votes > rep_votes, dem_votes - total_two_party / 2, dem_votes), rep_wasted = if_else(rep_votes > dem_votes, rep_votes - total_two_party / 2, rep_votes))

district_wasted

total_wasted_dem <- sum(district_wasted$dem_wasted, na.rm = TRUE) total_wasted_rep <- sum(district_wasted$rep_wasted, na.rm = TRUE) total_two_party_votes <- sum(district_wasted$total_two_party, na.rm = TRUE)

efficiency_gap <- (total_wasted_dem - total_wasted_rep) / total_two_party_votes

total_wasted_dem total_wasted_rep total_two_party_votes efficiency_gap

cat("Efficiency gap ((D wasted - R wasted) / total two-party votes):", round(efficiency_gap, 4), "") cat("Interpretation: positive values imply more wasted Democratic votes (Republican advantage).")