

Data Cleaning

```
list.files("data", recursive = TRUE)
```

```
[1] "g24 Sov by g24_svpref.csv"      "g24-candidates-by-district.csv"  
[3] "g24-results-by-district.xlsx"
```

```
library(tidyverse)
```

```
— Attaching core tidyverse packages ————— tidyverse 2.0.0  
—  
✓ dplyr     1.1.4    ✓ readr     2.1.5  
✓ forcats   1.0.0    ✓ stringr   1.5.2  
✓ 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
```

```
library(readr)  
library(sf)
```

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

```
# -----  
# load data  
# -----  
votes_file <- "data/g24 Sov by g24_svpref.csv"  
df <- read_csv(votes_file)
```

```
Rows: 51123 Columns: 76  
— Column specification
```

```
-----  
Delimiter: ","
```

```

chr (49): FIPS, SVPREC, SVPREC_KEY, ELECTION, GEO_TYPE, ASSAIP01,
ASSDEMO1, ...
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.

```

```

# clean column names
clean_names <- function(x) {
  x %>%
    str_to_lower() %>%
    str_replace_all("[^a-z0-9]+", "_") %>%
    str_replace_all("_+", "_") %>%
    str_replace("^_|_$", "")
}

colnames(df) <- clean_names(colnames(df))

# detect precinct column automatically
prec_col <- colnames(df)[str_detect(colnames(df), "svprec")][1]

if(is.na(prec_col)) stop("✖ No precinct column found matching pattern
'svprec'.")

# convert vote-related columns to numeric
vote_cols <- df %>% select(contains("votes"), contains("total")) %>% names()

df <- df %>%
  mutate(across(all_of(vote_cols), ~
  suppressWarnings(parse_number(as.character(.)))))

# arrange and save cleaned file
cleaned_df <- df %>%
  arrange(across(all_of(prec_col))) %>%
  relocate(all_of(prec_col), .before = everything())

write_csv(cleaned_df, "cleaned_g24_precinct_votes.csv")

cat("✓ Cleaned precinct vote file saved as: cleaned_g24_precinct_votes.csv\n")

```

✓ Cleaned precinct vote file saved as: cleaned_g24_precinct_votes.csv

```
# -----  
# spatial processing setup (safe mode)  
# -----  
  
# we do NOT have shapefiles yet, so this is a placeholder  
cat("⚠ Spatial step skipped: required shapefiles not found in /data folder.  
\\n")
```

```
⚠ Spatial step skipped: required shapefiles not found in /data folder.
```

```
list(  
  file_cleaned = "cleaned_g24_precinct_votes.csv",  
  precinct_column_detected = prec_col,  
  vote_columns = vote_cols  
)
```

```
$file_cleaned  
[1] "cleaned_g24_precinct_votes.csv"  
  
$precinct_column_detected  
[1] "svprec"  
  
$vote_columns  
character(0)
```