

Data Cleaning

```
#data/g24 Sov by g24_svpref.csv: vote totals by precinct and contest
#data/g24-results-by-district.xlsx: aggregates votes to the congressional
district level
#g24-candidates-by-district.csv: lists candidates running in each district

library(tidyverse)
library(readxl)
library(readr)
library(stringr)

#load the data
#cleaning it by fixing the weird values
prec <- read_csv('data/g24 Sov by g24_svpref.csv', na = c("", "NA", "***",
"*****"), show_col_types = FALSE)

#only want a couple of columns from this dataset
#we want precinct key, county FIPS, assembly / congressional district IDs, and
the house results

#now lets clean the dataset
prec_clean <- prec |>
  select(COUNTY, FIPS, SVPREC_KEY, ADDIST, CDDIST, TOTVOTE, DEMVOTE,
REPVOTE, USPDEM01, USPREP01)

#save the cleaned dataset
write_csv(prec_clean, "data/clean-precinct.csv")

#district level data
dist <- read_excel("data/g24-results-by-district.xlsx", sheet = 1, col_names =
FALSE)

#give names to col
names(dist)[1:3] <- c("label", "dem", "rep")

dist_clean <- dist |>
  mutate(label = str_trim(label)) |>
  filter(label == "District Totals") |>
  transmute(
    district = row_number(),
    d_votes = as.numeric(dem),
    r_votes = as.numeric(rep))
```

```
write_csv(dist_clean, "data/clean-district-totals.csv")
```