

# Exploratory Data Analysis

```
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
```

## data

```
df <- read_csv("cleaned_g24_precinct_votes.csv")
```

```
Rows: 51123 Columns: 76
```

```
— Column specification
```

```
Delimiter: ","
```

```
chr (49): svprec, fips, 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.
```

```
glimpse(df)
```

```
Rows: 51,123
```

```
Columns: 76
```

```
$ svprec      <chr> "00000001", "00000001", "00000001A", "00000001A", "00000002",
```

```

"0...
$ county      <dbl> 10, 51, 10, 51, 10, 51, 10, 51, 10, 51, 10, 51, 10, 51,
10,...
$ fips        <chr> "06019", "06101", "06019", "06101", "06019", "06101",
"0601...
$ addist      <dbl> 27, 3, 27, 3, 27, 3, 27, 3, 27, 3, 27, 3, 27, 3,
27,...
$ svprec_key  <chr> "0601900000001", "0610100000001", "0601900000001A",
"061010000...
$ election    <chr> "g24", "g24", "g24", "g24", "g24", "g24", "g24", "g24",
"g2...
$ geo_type    <chr> "svprec", "svprec", "svprec", "svprec", "svprec",
"svprec",...
$ cddist      <dbl> 13, 1, 13, 1, 13, 1, 13, 1, 13, 1, 13, 1, 13, 1,
13,...
$ sddist      <dbl> 14, 1, 14, 1, 14, 1, 14, 1, 14, 1, 14, 1, 14, 1,
14,...
$ bedist      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,...
$ totreg      <dbl> 398, 1597, 0, 0, 25, 4529, 0, 0, 149, 3955, 0, 0, 1710,
201...
$ demreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ repreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ aipreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ grnreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ libreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ nlpreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ refreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ dclreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ mscreg      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ totvote     <dbl> 52, 68, 252, 1190, 3, 413, 11, 2880, 15, 290, 83, 2487,
218...
$ demvote     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ repvote     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ aipvote     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ grnvote     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...

```

```

0,...
$ libvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ nlpvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ refvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ dclvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ mscvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ prcvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ absvote <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,...
$ assaip01 <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0...
$ assdem01 <chr> "25", "9", "86", "188", "****", "109", "8", "1004", "8",
"81...
$ assdem02 <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0...
$ assrep01 <chr> "25", "55", "162", "956", "****", "269", "3", "1693", "6",
"...
$ assrep02 <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0...
$ cngdem01 <chr> "23", "8", "79", "213", "****", "138", "8", "1116", "6",
"97...
$ cngdem02 <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0...
$ cngind01 <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0...
$ cngrep01 <chr> "29", "57", "171", "932", "****", "241", "3", "1608", "9",
"...
$ cngrep02 <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0...
$ prsaip01 <chr> "1", "1", "3", "6", "****", "7", "0", "45", "0", "4", "0",
"...
$ prsdem01 <chr> "23", "13", "82", "267", "****", "126", "8", "1170", "8",
"9...
$ prsgrn01 <chr> "0", "1", "0", "6", "****", "0", "0", "20", "0", "1", "0",
"...
$ prslib01 <chr> "0", "0", "0", "3", "****", "1", "0", "10", "0", "4", "0",
"...
$ prspaf01 <chr> "1", "0", "1", "1", "****", "0", "0", "10", "0", "1", "1",
"...
$ prsrep01 <chr> "27", "53", "163", "886", "****", "274", "3", "1585", "7",
"...
$ pr_2_n <chr> "14", "34", "138", "778", "****", "152", "5", "1191", "2",

```

```

"..."
$ pr_2_y      <chr> "37", "31", "110", "359", "****", "229", "6", "1487", "11",
"..."
$ pr_32_n     <chr> "21", "48", "173", "957", "****", "202", "9", "1578", "6",
"..."
$ pr_32_y     <chr> "29", "17", "75", "188", "****", "176", "2", "1097", "8",
"9..."
$ pr_33_n     <chr> "26", "51", "174", "955", "****", "202", "4", "1730", "6",
"..."
$ pr_33_y     <chr> "23", "14", "73", "170", "****", "173", "7", "938", "7",
"10..."
$ pr_34_n     <chr> "16", "31", "111", "494", "****", "160", "5", "1092", "4",
"..."
$ pr_34_y     <chr> "33", "33", "130", "625", "****", "210", "6", "1533", "9",
"..."
$ pr_35_n     <chr> "13", "35", "100", "579", "****", "132", "3", "882", "3",
"8..."
$ pr_35_y     <chr> "36", "31", "139", "559", "****", "240", "8", "1785", "11",
"..."
$ pr_36_n     <chr> "20", "16", "47", "144", "****", "114", "2", "655", "2",
"71..."
$ pr_36_y     <chr> "29", "50", "198", "999", "****", "269", "9", "2026", "12",
"..."
$ pr_3_n      <chr> "19", "48", "147", "731", "****", "229", "4", "1482", "9",
"..."
$ pr_3_y      <chr> "29", "18", "95", "405", "****", "155", "6", "1191", "5",
"1..."
$ pr_4_n      <chr> "13", "35", "146", "815", "****", "158", "6", "1253", "6",
"..."
$ pr_4_y      <chr> "37", "29", "103", "326", "****", "223", "5", "1419", "8",
"..."
$ pr_5_n      <chr> "16", "47", "172", "948", "****", "205", "5", "1681", "6",
"..."
$ pr_5_y      <chr> "34", "18", "74", "185", "****", "171", "6", "966", "7",
"10..."
$ pr_6_n      <chr> "34", "50", "173", "914", "****", "234", "5", "1612", "11",
"..."
$ pr_6_y      <chr> "17", "15", "67", "205", "****", "135", "6", "1035", "3",
"8..."
$ sendem01    <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0..."
$ sendem02    <chr> "0", "0", "0", "0", "****", "0", "0", "0", "0", "0", "0",
"0..."
$ senrep01    <chr> "0", "43", "0", "764", "****", "230", "0", "1709", "0",
"150..."
$ senrep02    <chr> "0", "18", "0", "257", "****", "119", "0", "699", "0",
"93",...
$ uspdem01    <chr> "19", "12", "79", "244", "****", "110", "7", "1145", "8",

```

```
"9...
$ usprep01 <chr> "29", "50", "158", "898", "****", "247", "3", "1511", "5",
"
$ ussdem01 <chr> "22", "13", "78", "253", "****", "129", "8", "1181", "7",
"9...
$ ussrep01 <chr> "26", "52", "165", "897", "****", "253", "3", "1545", "7",
"
"
```

## Question 1

What is the distribution of precinct sizes?

## Answer 1

```
numeric_cols <- df %>% select(where(is.numeric))

# choose the numeric column with highest total sum (likely total votes)
total_col <- numeric_cols %>%
  summarise(across(everything(), sum, na.rm = TRUE)) %>%
  pivot_longer(cols = everything(), names_to = "col", values_to = "total") %>%
  arrange(desc(total)) %>%
  pull(col) %>%
  first()
```

```
Warning: There was 1 warning in `summarise()`.
i In argument: `across(everything(), sum, na.rm = TRUE)`.
```

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))
```

```
total_col
```

```
[1] "totreg"
```

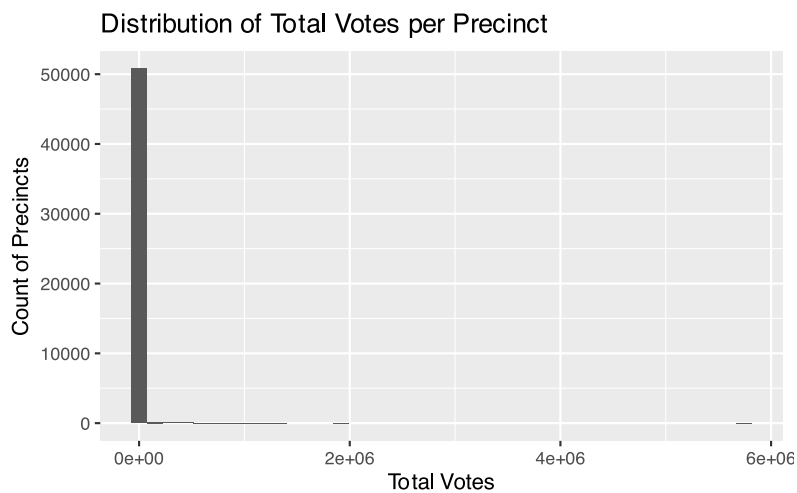
```
summary_stats <- df %>%
  summarise(
    min_votes = min(.data[[total_col]], na.rm = TRUE),
    median_votes = median(.data[[total_col]], na.rm = TRUE),
    mean_votes = mean(.data[[total_col]], na.rm = TRUE),
```

```
max_votes = max(.data[[total_col]], na.rm = TRUE)
)

summary_stats
```

```
# A tibble: 1 × 4
  min_votes median_votes mean_votes max_votes
    <dbl>      <dbl>      <dbl>      <dbl>
1         0          0      3081.    5745214
```

```
df %>%
  ggplot(aes(x = .data[[total_col]])) +
  geom_histogram(bins = 40) +
  labs(
    title = "Distribution of Total Votes per Precinct",
    x = "Total Votes",
    y = "Count of Precincts"
  )
```



## Question 2

Which precincts had the closest races?

## Answer 2

```
vote_cols <- df %>%
  select(contains("votes")) %>%
  names()

closest <- df %>%
```

```

rowwise() %>%
mutate(
  sorted_votes = sort(c_across(all_of(vote_cols)), decreasing = TRUE),
  margin = sorted_votes[1] - sorted_votes[2]
) %>%
ungroup() %>%
arrange(margin)

```

```

Error in `mutate()` :
i In argument: `margin = sorted_votes[1] - sorted_votes[2]`.
i In row 1.
Caused by error:
! object 'sorted_votes' not found

```

```
head(closest, 10)
```

```
Error: object 'closest' not found
```

### Question 3

Compare precinct-level totals with district totals

### Answer 3

```

library(readxl)
library(dplyr)

df <- read.csv("cleaned_g24_precinct_votes.csv")
district_results <- read_excel("data/g24-results-by-district.xlsx")

```

New names:

- `` -> `...2`
- `` -> `...3`

```

out <- list(
  precinct_preview = head(df),
  district_preview = head(district_results),
  precinct_rows = nrow(df),
  district_rows = nrow(district_results)
)

out

```

```

$precinct_preview
  svprec county fips addist svprec_key election geo_type cddist sddist
1 0000001 10 6019 27 060190000001 g24 svprec 13 14
2 0000001 51 6101 3 061010000001 g24 svprec 1 1
3 0000001A 10 6019 27 060190000001A g24 svprec 13 14
4 0000001A 51 6101 3 061010000001A g24 svprec 1 1
5 0000002 10 6019 27 060190000002 g24 svprec 13 14
6 0000002 51 6101 3 061010000002 g24 svprec 1 1
  bedist totreg demreg repreg aipreg grnreg libreg nlpreg refreg dclreg mscreg
1 1 398 0 0 0 0 0 0 0 0 0
2 1 1597 0 0 0 0 0 0 0 0 0
3 1 0 0 0 0 0 0 0 0 0 0
4 1 0 0 0 0 0 0 0 0 0 0
5 1 25 0 0 0 0 0 0 0 0 0
6 1 4529 0 0 0 0 0 0 0 0 0
  totvote demvote repvote aipvote grnvote libvote nlpvote refvote dclvote
1 52 0 0 0 0 0 0 0 0
2 68 0 0 0 0 0 0 0 0
3 252 0 0 0 0 0 0 0 0
4 1190 0 0 0 0 0 0 0 0
5 3 0 0 0 0 0 0 0 0
6 413 0 0 0 0 0 0 0 0
  mscvote prcvote absvote assaip01 assdem01 assdem02 assrep01 assrep02
cngdem01
1 0 0 0 0 25 0 25 0
23
2 0 0 0 0 9 0 55 0
8
3 0 0 0 0 86 0 162 0
79
4 0 0 0 0 188 0 956 0
213
5 0 0 0 *** *** *** *** ***
***
6 0 0 0 0 109 0 269 0
138
  cngdem02 cngind01 cngrep01 cngrep02 prsaip01 prsdem01 prsgrn01 prslib01
1 0 0 29 0 1 23 0 0
2 0 0 57 0 1 13 1 0
3 0 0 171 0 3 82 0 0
4 0 0 932 0 6 267 6 3
5 *** *** *** *** *** *** *** ***
6 0 0 241 0 7 126 0 1
  prspaf01 prsrep01 pr_2_n pr_2_y pr_32_n pr_32_y pr_33_n pr_33_y pr_34_n
1 1 27 14 37 21 29 26 23 16
2 0 53 34 31 48 17 51 14 31
3 1 163 138 110 173 75 174 73 111
4 1 886 778 359 957 188 955 170 494

```



```

5      ***      ***      ***      ***      ***      ***      ***      ***      ***
6      0      274      152      229      202      176      202      173      160
  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
1      33      13      36      20      29      19      29      13      37      16
2      33      35      31      16      50      48      18      35      29      47
3      130     100     139      47     198     147      95     146     103     172
4      625     579     559     144     999     731     405     815     326     948
5      ***      ***      ***      ***      ***      ***      ***      ***      ***
6      210     132     240     114     269     229     155     158     223     205
  pr_5_y pr_6_n pr_6_y sendem01 sendem02 senrep01 senrep02 uspdem01 usprep01
1      34      34      17          0          0          0          0          19      29
2      18      50      15          0          0          43         18          12      50
3      74      173     67          0          0          0          0          79     158
4      185     914     205          0          0         764        257        244     898
5      ***      ***      ***      ***      ***      ***      ***      ***      ***
6      171     234     135          0          0         230        119        110     247
  ussdem01 ussrep01
1          22         26
2          13         52
3          78        165
4         253        897
5          ***        ***
6         129        253

$district_preview
# A tibble: 6 × 3
  `1st Congressional District` ...2      ...3
  <chr>                        <chr>      <chr>
1 <NA>                        <NA>      <NA>
2 <NA>                        "Rose\r\nPenelope\r\nYee" "Doug\r\nLaMalfa*"
3 <NA>                        "DEM"      "REP"
4 Butte                       "41729"    "50979"
5 Colusa                      "2095"    "4528"
6 Glenn                       "2936"    "7197"

$precinct_rows
[1] 51123

$district_rows
[1] 543

```