

# Exploratory Data Analysis

```
— Attaching core tidyverse packages —
tidyverse 2.0.0 —
✓ dplyr     1.1.4      ✓ readr     2.1.5
✓forcats    1.0.0      ✓ stringr   1.5.1
✓ ggplot2   3.5.2      ✓ 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

Attaching package: 'janitor'
```

The following objects are masked from 'package:stats':

chisq.test, fisher.test

Rows: 511230 Columns: 76

— Column specification

---

Delimiter: ","
chr (46): fips, svprec, svprec\_key, election, geo\_type,
prsaip01, prsdem01, ...
dbl (30): 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.

Rows: 51123 Columns: 77

— Column specification

---

Delimiter: ","
chr (49): fips, svprec, svprec\_key, election, geo\_type,
assaip01, assdem01, ...

```
dbl (28): 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.

## Question 1

Which Races were the closest and which ones were blowouts?

```
Warning: Returning more (or less) than 1 row per `summarise()`  
group was deprecated in  
dplyr 1.1.0.
```

**i** Please use `reframe()` instead.

**i** When switching from `summarise()` to `reframe()`, remember that `reframe()`

always returns an ungrouped data frame and adjust accordingly.

#	A tibble: 10 × 8	race_type	district	winner	winner_party	loser	loser_party
		<chr>	<dbl>	<chr>	<chr>	<chr>	<chr>
1	Assembly	58	Leticia Castillo	REP	Clar...	DEM	
0.00362							
2	Congress	35	Mike Cargile	REP	Norm...	DEM	
0.00609							
3	Assembly	74	Laurie Davies*	REP	Chri...	DEM	
0.0167							
4	Assembly	47	Greg Wallis*	REP	Chri...	DEM	
0.0229							
5	Congress	7	Doris Matsui*	DEM	Tom ...	REP	
0.0290							
6	Assembly	36	Jeff Gonzalez	REP	Jose...	DEM	
0.0351							
7	Congress	27	George Whitesid...	DEM	Mike...	REP	
0.0448							
8	Assembly	40	Pilar Schiavo*	DEM	Patr...	REP	
0.0563							
9	Congress	40	Joe Kerr	DEM	Youn...	REP	
0.0617							
10	Congress	47	Dave Min	DEM	Scot...	REP	

```
0.0663  
# i 1 more variable: total_votes <dbl>
```

## Answer 1

```
Attaching package: 'scales'
```

```
The following object is masked from 'package:purrr':
```

```
discard
```

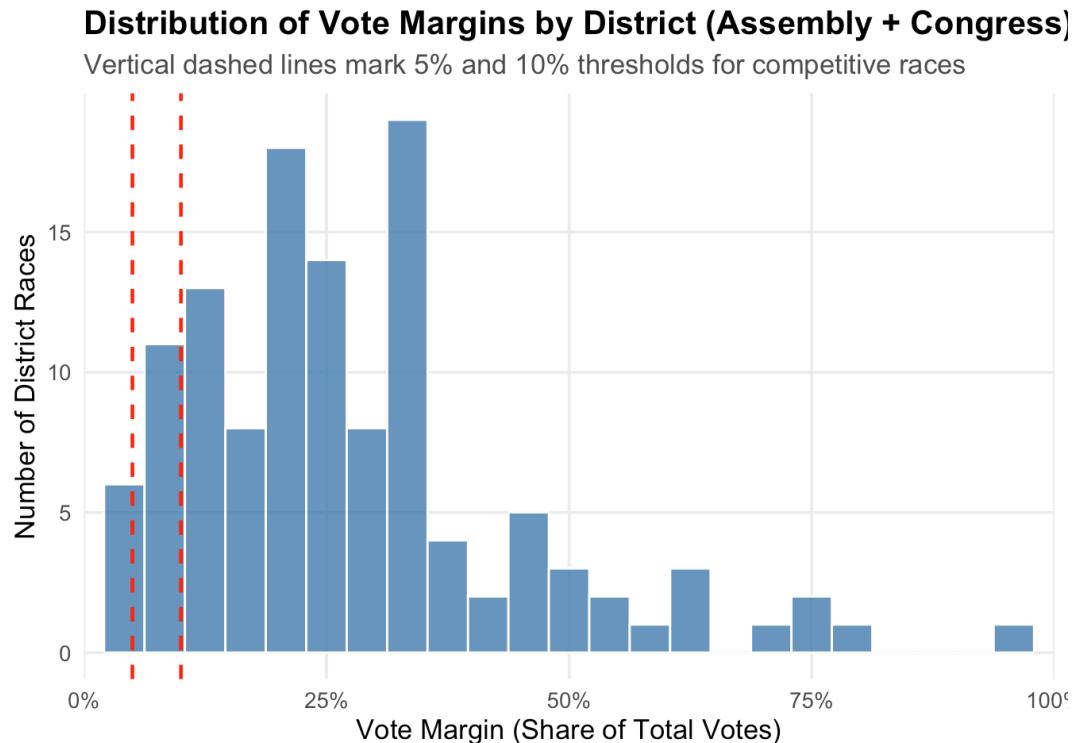
```
The following object is masked from 'package:readr':
```

```
col_factor
```

```
# A tibble: 10 × 8  
  race_type district winner      winner_party loser  
  <chr>     <dbl> <chr>       <chr>        <chr> <chr>  
  loser_party pct_margin  
  <chr>       <dbl> <chr>       <chr>        <chr> <chr>  
  <dbl>  
  1 Assembly    58 Leticia Castillo REP      Clar... DEM  
0.00362  
  2 Congress    35 Mike Cargile   REP      Norm... DEM  
0.00609  
  3 Assembly    74 Laurie Davies* REP      Chri... DEM  
0.0167  
  4 Assembly    47 Greg Wallis*  REP      Chri... DEM  
0.0229  
  5 Congress     7 Doris Matsui* DEM      Tom ... REP  
0.0290  
  6 Assembly    36 Jeff Gonzalez REP      Jose... DEM  
0.0351  
  7 Congress    27 George Whitesid... DEM      Mike... REP  
0.0448  
  8 Assembly    40 Pilar Schiavo* DEM      Patr... REP  
0.0563  
  9 Congress     40 Joe Kerr      DEM      Youn... REP  
0.0617  
 10 Congress    47 Dave Min      DEM      Scot... REP  
0.0663  
# i 1 more variable: total_votes <dbl>
```

```
Warning: Removed 2 rows containing missing values or values  
outside the scale range
```

```
(`geom_bar()`).
```

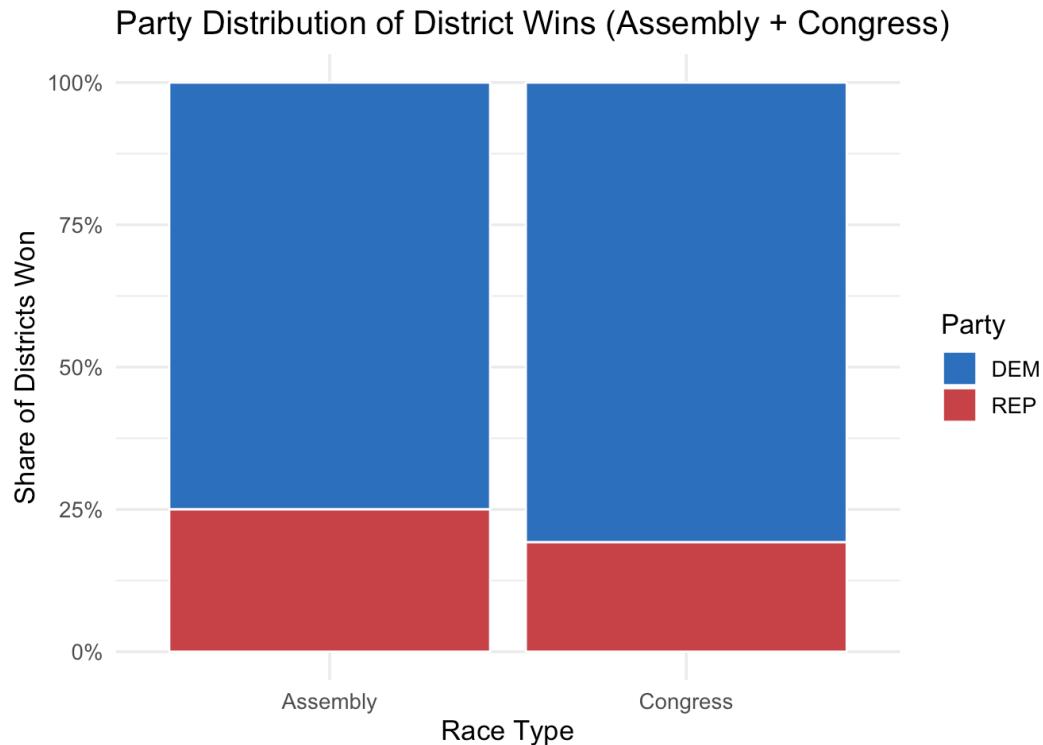


Leticia Castillo (REP) vs. Clarissa Cervantes (DEM), won by Castillo by just 0.003 was the closest race for Assembly, while Mike Cargile (REP) vs. Norma J. Torres (DEM), won by Cargile by ~0.007 was the closest race for Congress

## Question 2

Which party dominated across districts and were their wins evenly distributed?

## Answer 2



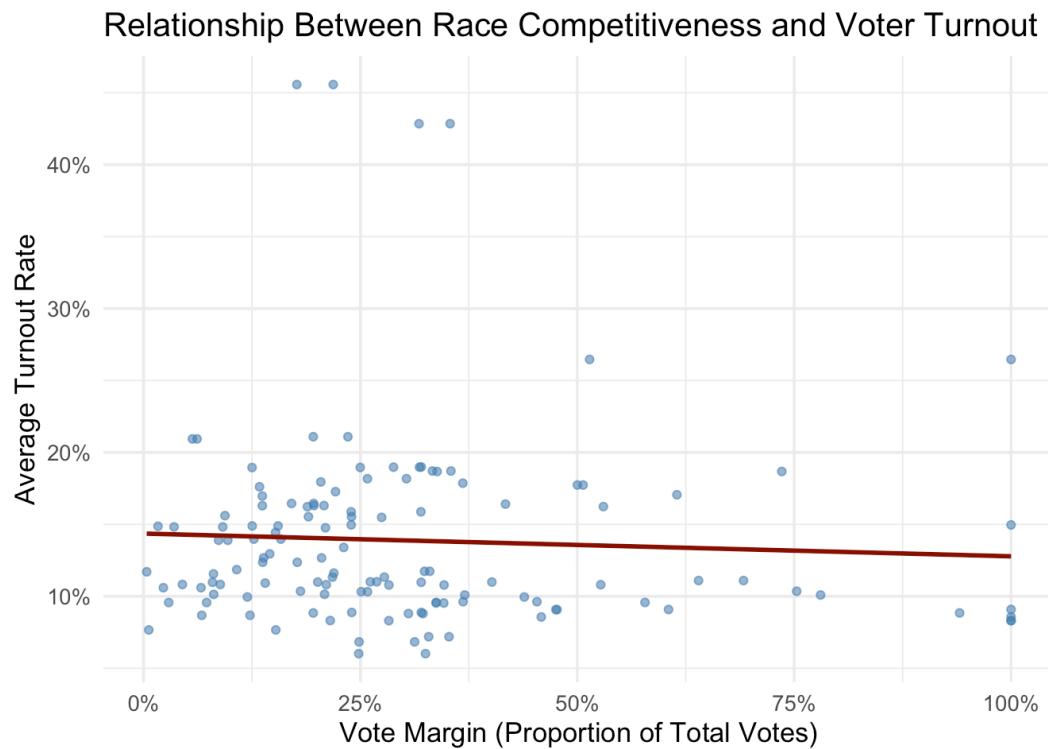
The Democratic party clearly dominated in both Congress and Assembly, although Republicans had a slightly higher share of Districts in the Assembly race.

## Question 3

Are turnout rates correlated with competitiveness?

## Answer 3

```
`geom_smooth()` using formula = 'y ~ x'
```



We see that Vote margins are lower when turnout is higher, suggesting closer races when people are participating in voting.