

Maps Mini Project: Gerrymandering

```
#install and load the necessary packages
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(sf)
```

Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE

```
library(mapproj)
```

Loading required package: maps

Attaching package: 'maps'

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

map

```
library(maps)
library(ggspatial)
library(prettymapr)
devtools::install_github("baumer-lab/fec12")
```

Skipping install of 'fec12' from a github remote, the SHA1 (aa02bfc1) has not changed since 1. Use `force = TRUE` to force installation

```
library(fec12)
library(leaflet)
library(mdsr)
```

Introduction

We are interested in seeing the difference in Democratic and Republican votes. The use of geospatial data will be useful in helping us understand. We are emulating our analysis of potential gerrymandering in Wisconsin. Our analysis focuses on the Federal Election Commission, specifically for the state of Wisconsin in the 2012 Presidential Election. This dataset, **results_house**, comes from the Federal Election Commission library. **Results_house** includes information on the the number of districts a particular state has, whether a candidate from that district is from the Republican or Democratic Party, the general number of votes in the election. We begin with determining how many congressional candidates there are for each district in the United States.

```
print(results_house, width = Inf)
```

A tibble: 2,343 x 13

	state	district_id	cand_id	incumbent	party	primary_votes	primary_percent
	<chr>	<chr>	<chr>	<lgl>	<chr>	<dbl>	<dbl>
1	AL	01	H2AL01077	TRUE	R	48702	0.555
2	AL	01	H2AL01176	FALSE	R	21308	0.243
3	AL	01	H2AL01184	FALSE	R	13809	0.158
4	AL	01	H0AL01030	FALSE	R	3854	0.0440
5	AL	02	H0AL02087	TRUE	R	NA	NA
6	AL	02	H2AL02141	FALSE	D	NA	NA
7	AL	03	H2AL03032	TRUE	R	NA	NA
8	AL	03	H2AL03099	FALSE	D	NA	NA
9	AL	04	H6AL04098	TRUE	R	NA	NA
10	AL	04	H2AL04055	FALSE	D	10971	0.514

```

runoff_votes runoff_percent general_votes general_percent won footnotes
      <dbl>      <dbl>      <dbl>      <dbl> <lgl> <chr>
1         NA         NA      196374      0.979 TRUE <NA>
2         NA         NA         NA         NA  FALSE <NA>
3         NA         NA         NA         NA  FALSE <NA>
4         NA         NA         NA         NA  FALSE <NA>
5         NA         NA      180591      0.636 TRUE  <NA>
6         NA         NA      103092      0.363 FALSE <NA>
7         NA         NA      175306      0.640 TRUE  <NA>
8         NA         NA       98141      0.358 FALSE <NA>
9         NA         NA      199071      0.740 TRUE  <NA>
10        NA         NA       69706      0.259 FALSE <NA>
# i 2,333 more rows

```

```

results_house |>
  group_by(state, district_id) |>
  summarize(N = n())

```

`summarise()` has grouped output by 'state'. You can override using the ``.groups` argument.

```

# A tibble: 445 x 3
# Groups:   state [56]
  state district_id     N
  <chr> <chr>      <int>
1 AK    00          10
2 AL    01           4
3 AL    02           2
4 AL    03           2
5 AL    04           3
6 AL    05           3
7 AL    06           6
8 AL    07           3
9 AR    01           6
10 AR   02           4
# i 435 more rows

```

Note that there are 435 Representatives in the US House, but there are 445 state and district combinations in our data because the `fec12` dataset includes the US Territories of: Guam, Puerto Rico, American Samoa, Northern Mariana Islands, District of Columbia, the Virgin Island.

However, we are only interested in the eight congressional districts in Wisconsin.

Analyzing Votes by District in Wisconsin

```
# summary of the 8 congressional WI districts and the 2012 voting results
district_elections <- results_house |>
  mutate(district = parse_number(district_id)) |>
  group_by(state, district) |>
  summarize(
    N = n(),
    total_votes = sum(general_votes, na.rm = TRUE),
    d_votes = sum(ifelse(party == "D", general_votes, 0), na.rm = TRUE),
    r_votes = sum(ifelse(party == "R", general_votes, 0), na.rm = TRUE),
    .groups = "drop"
  ) |>
  mutate(
    other_votes = total_votes - d_votes - r_votes,
    r_prop = r_votes / total_votes,
    winner = ifelse(r_votes > d_votes, "Republican", "Democrat")
  )
wi_results <- district_elections |>
  filter(state == "WI")
wi_results |>
  select(-state)
```

A tibble: 8 x 8

	district	N	total_votes	d_votes	r_votes	other_votes	r_prop	winner
	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
1	1	3	364891	158414	200423	6054	0.549	Republican
2	2	6	390111	265422	124683	6	0.320	Democrat
3	3	2	339425	217712	121713	0	0.359	Democrat
4	4	3	325321	235257	80787	9277	0.248	Democrat
5	5	2	368813	118478	250335	0	0.679	Republican
6	6	3	359381	135921	223460	0	0.622	Republican
7	7	3	359264	157524	201720	20	0.561	Republican
8	8	2	355161	156287	198874	0	0.560	Republican

Here, we are summing up all the `general_votes` for the Democratic and Republican Party for the state of Wisconsin. Notice that we are not super interested in any Third Party candidates, but we are interested in knowing the proportion of Democratic and Republican votes.

Proportion of Republican and Democratic Votes

```
# distribution of total number of votes is narrow by design
wi_results |>
  skim(total_votes) |>
  select(-na)
```

Variable type: numeric

var	n	mean	sd	p0	p25	p50	p75	p100
total_votes	8	357795.9	19345.37	325321	351227	359322.5	365871.5	390111

```
# compare total Dem and Rep votes across WI in 2012
wi_results |>
  summarize(
    N = n(),
    state_votes = sum(total_votes),
    state_d = sum(d_votes),
    state_r = sum(r_votes)
  ) |>
  mutate(
    d_prop = state_d / state_votes,
    r_prop = state_r / state_votes
  )
```

```
# A tibble: 1 x 6
      N state_votes state_d state_r d_prop r_prop
<int>   <dbl>   <dbl>   <dbl> <dbl> <dbl>
1     8   2862367 1445015 1401995  0.505  0.490
```

First, we are interested in the mean of the total votes across all 8 congressional districts. Approximately 357796 individuals from each congressional district voted, with a standard deviation of 19345. With that information, we are able to calculate the total amount of Republican and Democratic votes, and the proportion of Democratic and Republican votes for the entire state of Wisconsin.

Winner of Each District

```
# Proportion of Rep votes by district
wi_results |>
  select(district, r_prop, winner) |>
  arrange(desc(r_prop))

# A tibble: 8 x 3
  district r_prop winner
    <dbl>   <dbl> <chr>
1         5 0.679 Republican
2         6 0.622 Republican
3         7 0.561 Republican
4         8 0.560 Republican
5         1 0.549 Republican
6         3 0.359 Democrat
7         2 0.320 Democrat
8         4 0.248 Democrat
```

Now, we are able to analyze the proportion of Republican voters, in addition, to the party that won in that congressional district. Notice, a bigger difference from 0.50 the better. For the state of Wisconsin, the Republican Party seems to have won in a majority of the congressional districts. However, notice that the Republican Party wins by a slight margin, whereas the Democratic Party won by a much larger margin. In other words, there is a more significant difference between the Democratic and Republican votes. Now that we have the numbers, let us visualize the data.

Visualization of Potential Gerrymandering in Wisconsin

```
#first, let us download the fec12 dataset from UCLA

src <- "http://cdmaps.polisci.ucla.edu/shp/districts113.zip"
lcl_zip <- fs::path(tempdir(), "districts113.zip")
download.file(src, destfile = lcl_zip)
lcl_districts <- fs::path(tempdir(), "districts113")
unzip(lcl_zip, exdir = lcl_districts)
dsn_districts <- fs::path(lcl_districts, "districtShapes")

# read shapefiles into R as an sf object
st_layers(dsn_districts)
```

Driver: ESRI Shapefile

Available layers:

	layer_name	geometry_type	features	fields	crs_name
1	districts113	Polygon	436	15	NAD83

```
# be able to read as a data frame as well
districts <- st_read(dsn_districts, layer = "districts113") |>
  mutate(DISTRICT = parse_number(as.character(DISTRICT)))
```

Reading layer `districts113' from data source

```
`/private/var/folders/38/j8rsrqzs2b1_d3b742694ty80000gp/T/Rtmpkrl4Jh/districts113/districts113.shp'
using driver `ESRI Shapefile'
```

Simple feature collection with 436 features and 15 fields (with 1 geometry empty)

Geometry type: MULTIPOLYGON

Dimension: XY

Bounding box: xmin: -179.1473 ymin: 18.91383 xmax: 179.7785 ymax: 71.35256

Geodetic CRS: NAD83

```
head(districts, width = Inf)
```

Simple feature collection with 6 features and 15 fields

Geometry type: MULTIPOLYGON

Dimension: XY

Bounding box: xmin: -91.82307 ymin: 29.41135 xmax: -66.94983 ymax: 47.45969

Geodetic CRS: NAD83

	STATENAME	ID	DISTRICT	STARTCONG	ENDCONG	DISTRICTSI	COUNTY	PAGE	LAW
1	Louisiana	022113114006	6	113	114	<NA>	<NA>	<NA>	<NA>
2	Maine	023113114001	1	113	114	<NA>	<NA>	<NA>	<NA>
3	Maine	023113114002	2	113	114	<NA>	<NA>	<NA>	<NA>
4	Maryland	024113114001	1	113	114	<NA>	<NA>	<NA>	<NA>
5	Maryland	024113114002	2	113	114	<NA>	<NA>	<NA>	<NA>
6	Maryland	024113114003	3	113	114	<NA>	<NA>	<NA>	<NA>
	NOTE	BESTDEC	FINALNOTE	RNOTE	LASTCHANGE				
1	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626				
2	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626				
3	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626				
4	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626				
5	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626				
6	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626				
	FROMCOUNTRY		geometry						

```

1      F MULTIPOLYGON (((-91.82288 3...
2      F MULTIPOLYGON (((-70.98905 4...
3      F MULTIPOLYGON (((-71.08216 4...
4      F MULTIPOLYGON (((-77.31156 3...
5      F MULTIPOLYGON (((-76.8763 39...
6      F MULTIPOLYGON (((-77.15622 3...

```

```
class(districts)
```

```
[1] "sf"          "data.frame"
```

```

#Append election results to geospatial data
wi_merged <- districts |>
  filter(STATENAME == "Wisconsin") |>
  st_transform(4326) |>
  inner_join(wi_results, by = c("DISTRICT" = "district"))
head(wi_merged, width = Inf)

```

Simple feature collection with 6 features and 23 fields

Geometry type: MULTIPOLYGON

Dimension: XY

Bounding box: xmin: -92.808 ymin: 42.49198 xmax: -87.50719 ymax: 45.20957

Geodetic CRS: WGS 84

	STATENAME	ID	DISTRICT	STARTCONG	ENDCONG	DISTRICTSI	COUNTY	PAGE	LAW
1	Wisconsin	055113114001	1	113	114	<NA>	<NA>	<NA>	<NA>
2	Wisconsin	055113114002	2	113	114	<NA>	<NA>	<NA>	<NA>
3	Wisconsin	055113114003	3	113	114	<NA>	<NA>	<NA>	<NA>
4	Wisconsin	055113114004	4	113	114	<NA>	<NA>	<NA>	<NA>
5	Wisconsin	055113114005	5	113	114	<NA>	<NA>	<NA>	<NA>
6	Wisconsin	055113114006	6	113	114	<NA>	<NA>	<NA>	<NA>

	NOTE	BESTDEC	FINALNOTE	RNOTE	LASTCHANGE
1	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626
2	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626
3	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626
4	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626
5	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626
6	<NA>	<NA>	{"From US Census website"}	<NA>	2016-05-29 16:44:10.857626

	FROMCOUNTY	state	N	total_votes	d_votes	r_votes	other_votes	r_prop
1	F	WI	3	364891	158414	200423	6054	0.5492681
2	F	WI	6	390111	265422	124683	6	0.3196090

3	F	WI 2	339425	217712	121713	0 0.3585858
4	F	WI 3	325321	235257	80787	9277 0.2483301
5	F	WI 2	368813	118478	250335	0 0.6787586
6	F	WI 3	359381	135921	223460	0 0.6217914

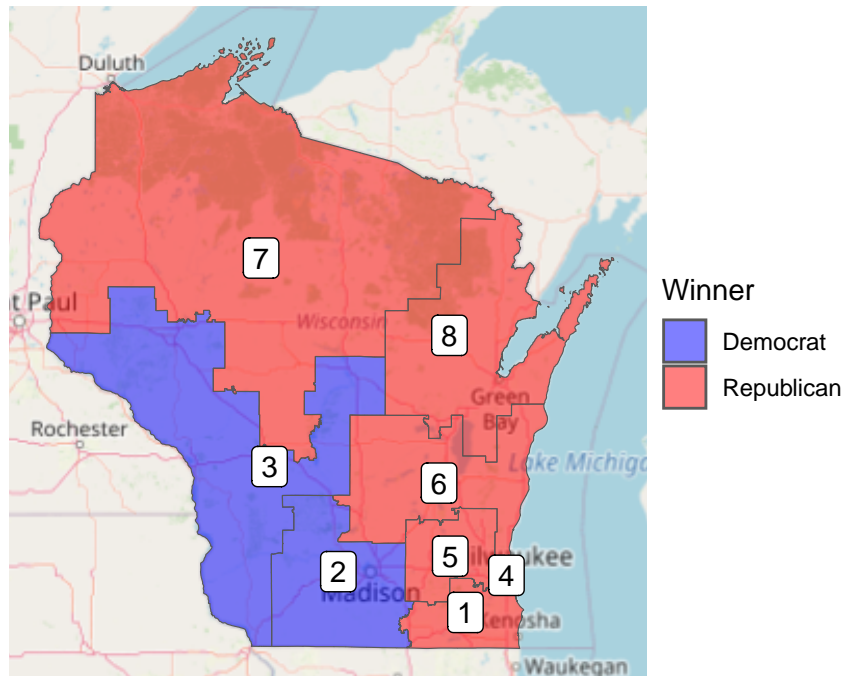
	winner	geometry
1	Republican	MULTIPOLYGON (((-89.08072 4...
2	Democrat	MULTIPOLYGON (((-90.43 43.1...
3	Democrat	MULTIPOLYGON (((-91.3984 44...
4	Democrat	MULTIPOLYGON (((-88.06601 4...
5	Republican	MULTIPOLYGON (((-89.01359 4...
6	Republican	MULTIPOLYGON (((-89.78555 4...

```
# Color based on winning party; shows the 8 congressional districts
```

```
wi <- ggplot(data = wi_merged, aes(fill = winner)) +
  annotation_map_tile(zoom = 6, type = "osm", progress = "none") +
  geom_sf(alpha = 0.5) +
  scale_fill_manual("Winner", values = c("blue", "red")) +
  geom_sf_label(aes(label = DISTRICT), fill = "white") +
  theme_void()
wi
```

Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not give correct results for longitude/latitude data

Loading required namespace: raster



This map shows the potential gerrymandering of the eight total congressional districts in the state of Wisconsin. Note that the map is color coded such that it matches the results we found above. There are 3 congressional districts where the Democratic Party won, and 5 congressional districts where the Republican Party won. From this map, I wanted to create a choropleth map which allows us to visualize the proportion of Republican and Democratic votes in each district.

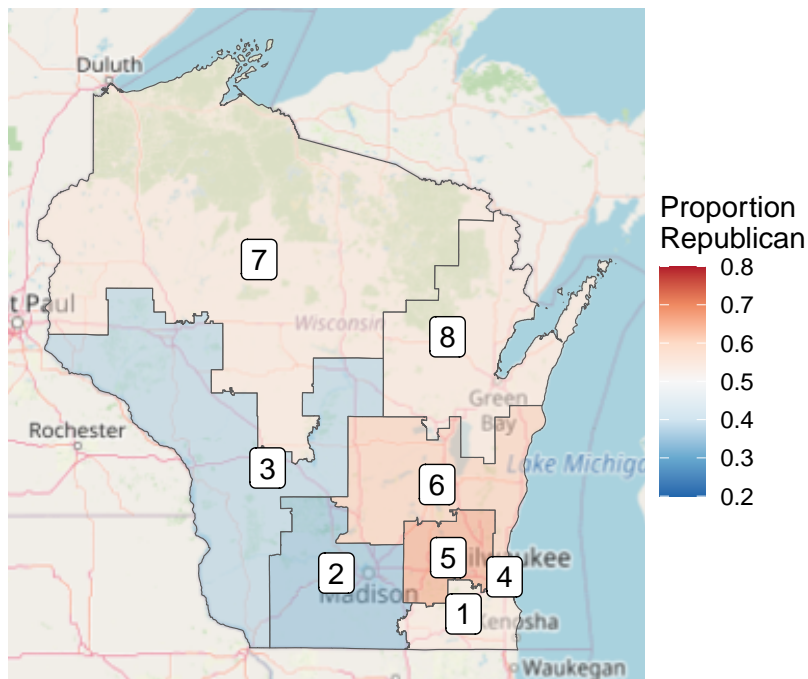
Another visualization of our potential gerrymandering

```
# Color based on proportion Rep.
wi +
  aes(fill = r_prop) +
  scale_fill_distiller(
    "Proportion\nRepublican",
    palette = "RdBu",
    limits = c(0.2, 0.8)
  )
```

Scale for fill is already present.

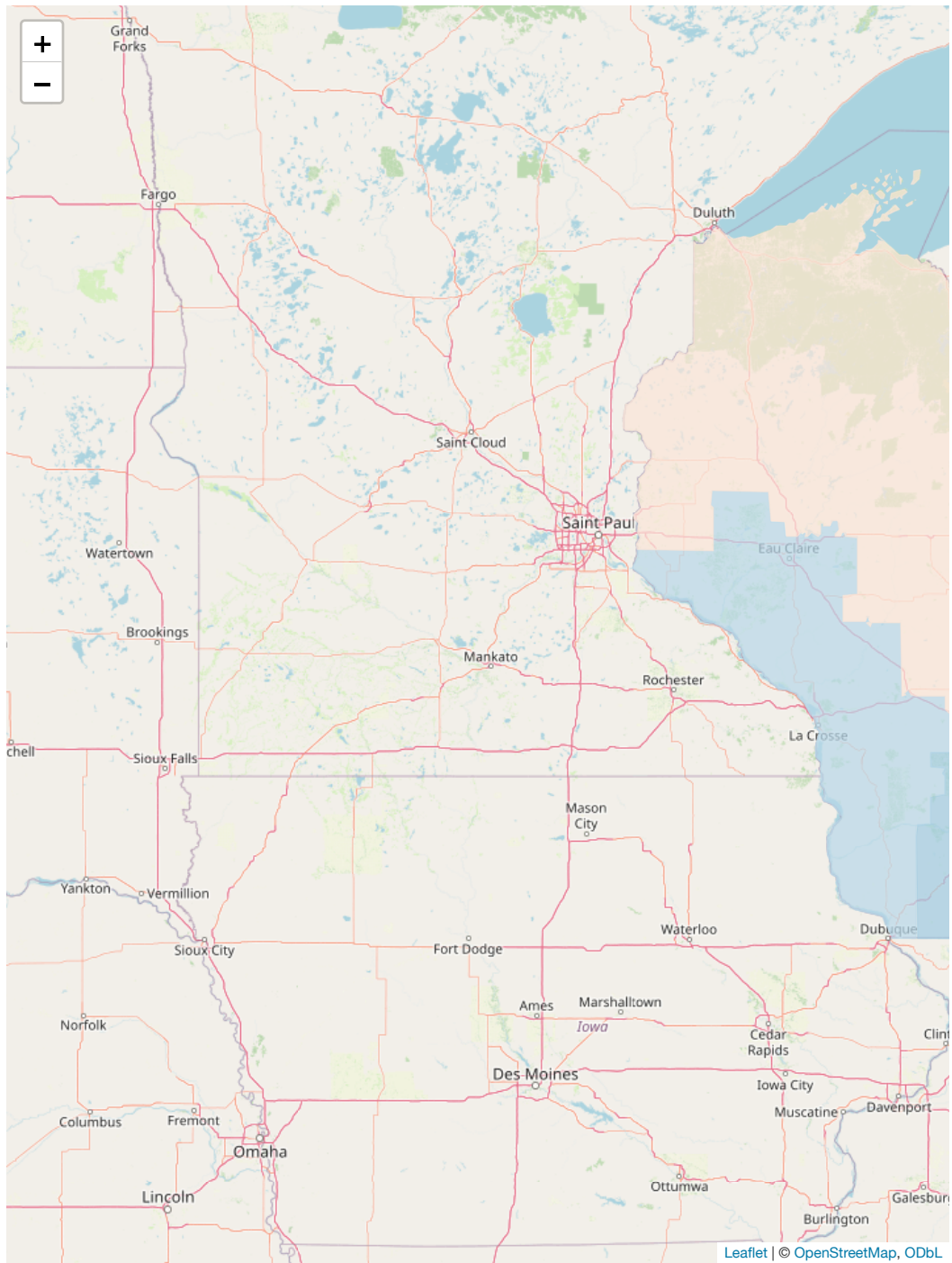
Adding another scale for fill, which will replace the existing scale.

Warning in `st_point_on_surface.sfc(sf::st_zm(x))`: `st_point_on_surface` may not give correct results for longitude/latitude data



```
# A leaflet map can allow us to zoom in and see where major cities fit, etc.
library(leaflet)
pal <- colorNumeric(palette = "RdBu", domain = c(0, 1))

leaflet_wi <- leaflet(wi_merged) |>
  addTiles() |>
  addPolygons(
    weight = 1, fillOpacity = 0.7,
    color = ~pal(1 - r_prop), # so red association with Reps
    popup = ~paste("District", DISTRICT, "<br>", round(r_prop, 4))
  ) |> # popups show prop Republican
  setView(lng = -89, lat = 44, zoom = 7)
leaflet_wi
```



Note, this is a choropleth map, where meaningful shading relates to the proportion of Democratic and Republican votes by congressional districts. Similar to the map above, the color blue represents the Democratic Party, while the color red represents the Republican Party. Notice, this map illustrates the same information as the map above. However, in this choropleth map, we can how much each party won, by congressional districts. We can tell the proportion of Democratic and Republican votes by how opaque the colors are on the map. If we hover our mouse over a particular place on the map, it tells us what district it is, and the proportion of votes they received.

Conclusion

It was quite interesting to explore the congressional districts data of Wisconsin. Knowing the results from the Presidential Election, I learned a lot about how the majority of the winning congressional districts does not translate to which party won in that particular state due to the electoral college. By analyzing potential gerrymandering data, we can learn more about each congressional district, and how the proportion of votes affect which party wins.

Citations

[Federal Election Commission:](#)