Congressional Reapportionment

Forecast

BELLEVUE UNIVERSITY

bellevue.edu

Matt Burns
DSC 680
Bellevue University
11/20/2020

Abstract

This project will investigate using readily available data to estimate the upcoming congressional apportionment. Each state will get a unique population forecast via linear regressions in RStudio. These populations will drive the apportionment.

Background) When the constitution was written over 200 years ago, it required that the people living in the United States be enumerated every 10 years. This count is used to reapportion the number of congressional seats per state.

Problem Statement) Censuses are expensive and slow.
I will be investigating an alternate methodology of reapportionment for the congressional seats. I will be comparing electricity usage, personal income, and to estimate the states' population which will be used to estimate the upcoming reapportionment.

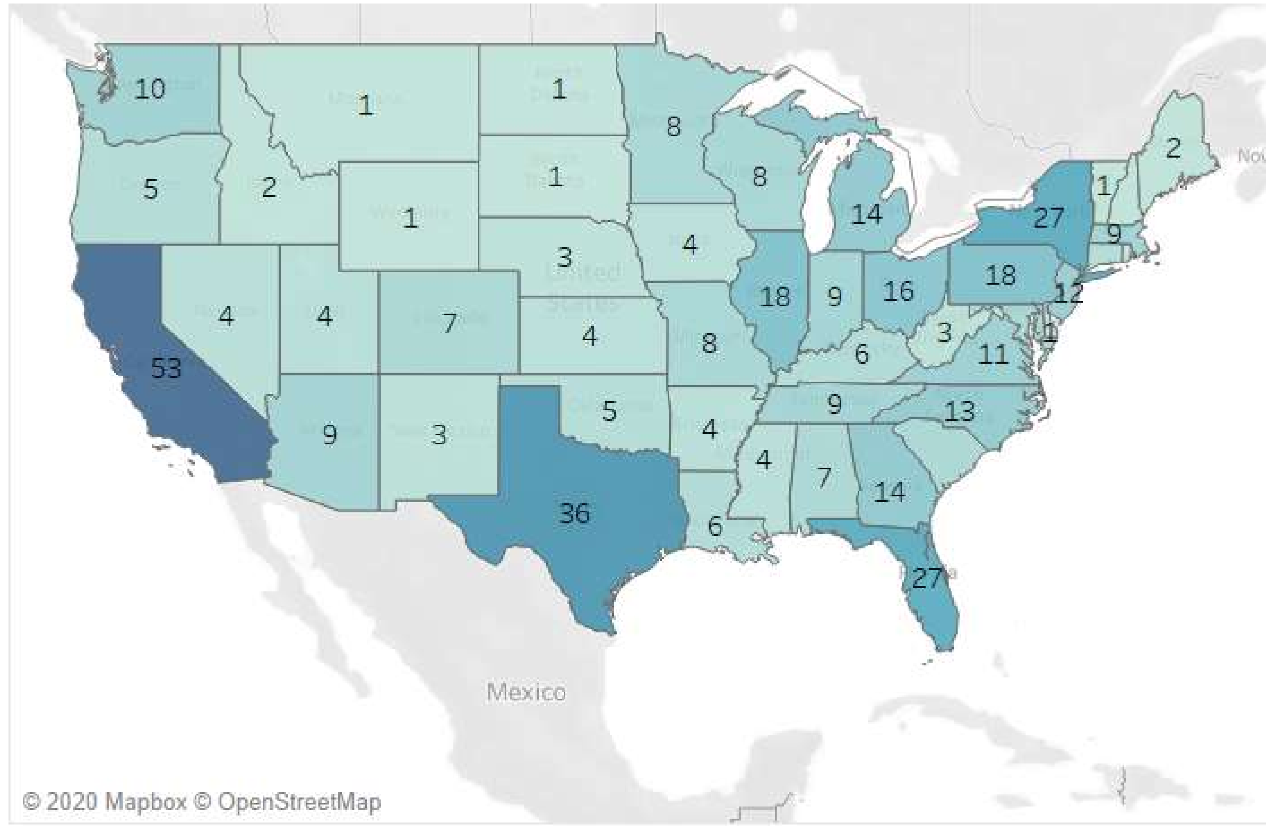
Scope) This analysis will forecast the population for each state using a unique linear regression. The forecasted population will then be used to allocate congressional seat. Additionally, apportionment using the raw data alone (e.g. electricity usage) with a regression will be used as a reference.

Congressional Seats

Reminders

- Congressional District: In this presentation I will refer to congressional districts as members.
- Senate Seats: Each state get two senators. They are largely left out of this analysis.
- Electoral Votes: This the apportionment is for the House of Representatives. Electoral college votes are derived by adding 2 to the number of members of the House of Representatives. This paper will only touch on this at the end.

Current House of Representative Members



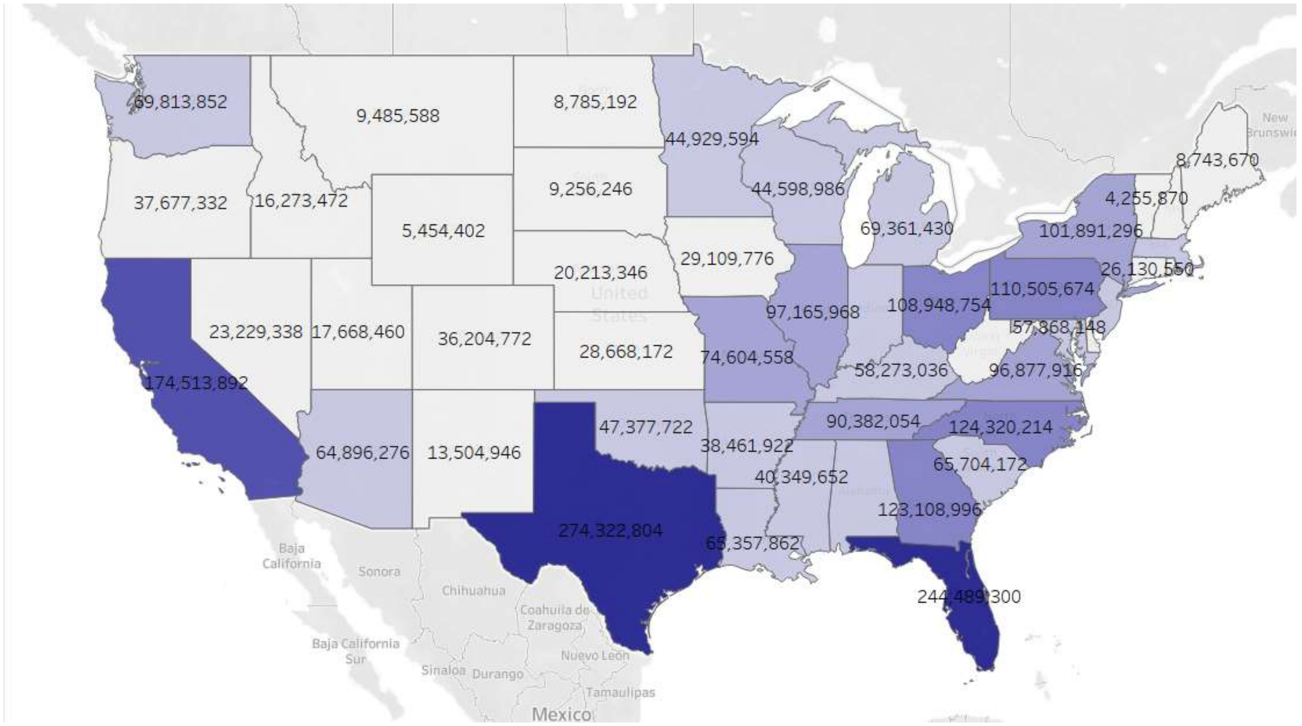
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Current Members. The marks are labeled by sum of Current Members. Details are shown for State. The view is filtered on State, which has multiple members selected.

Candidate Predictors

3 Candidates Macro Features

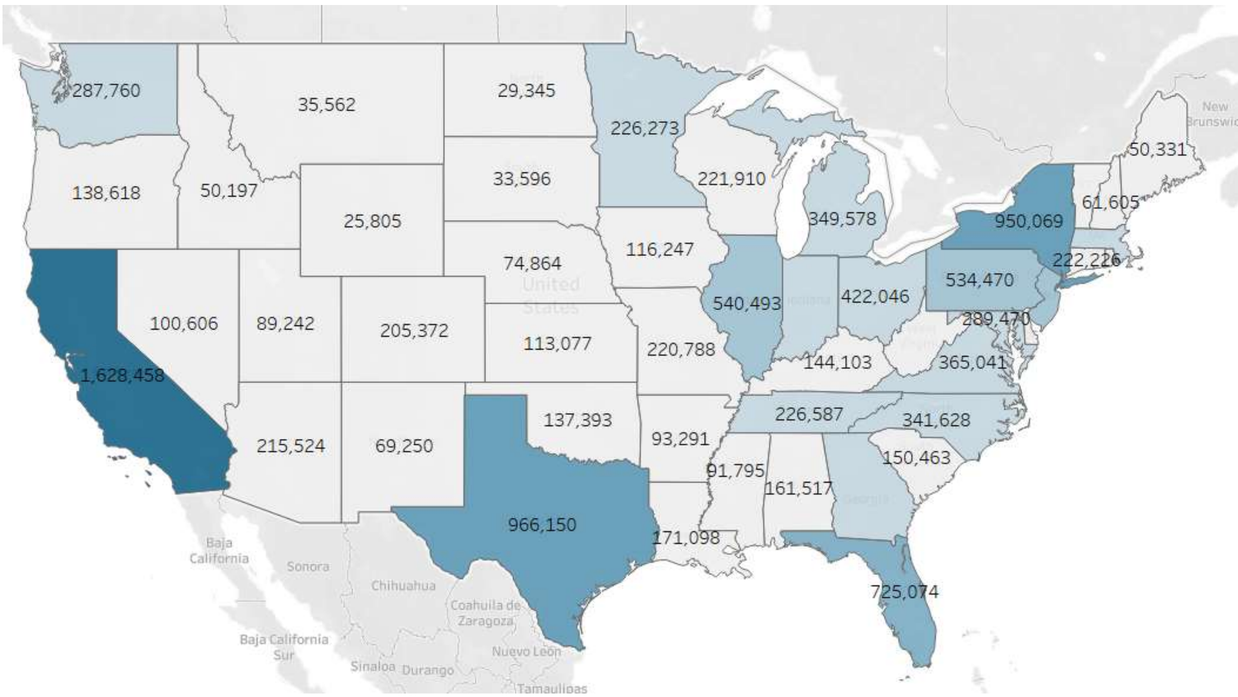
- Residential Electricity Usage: The Energy Information Administration providers annual residential consumption at the state level.
- Personal Federal Taxes: The Bureau of Economic Analysis provides the Personal Federal Income Taxes for each state.
- Personal Income: The Bureau of Economic Analysis provides Personal Income. It represents the gross income of individuals that is subject to taxation.

Residential Electricity Usage (MWh)

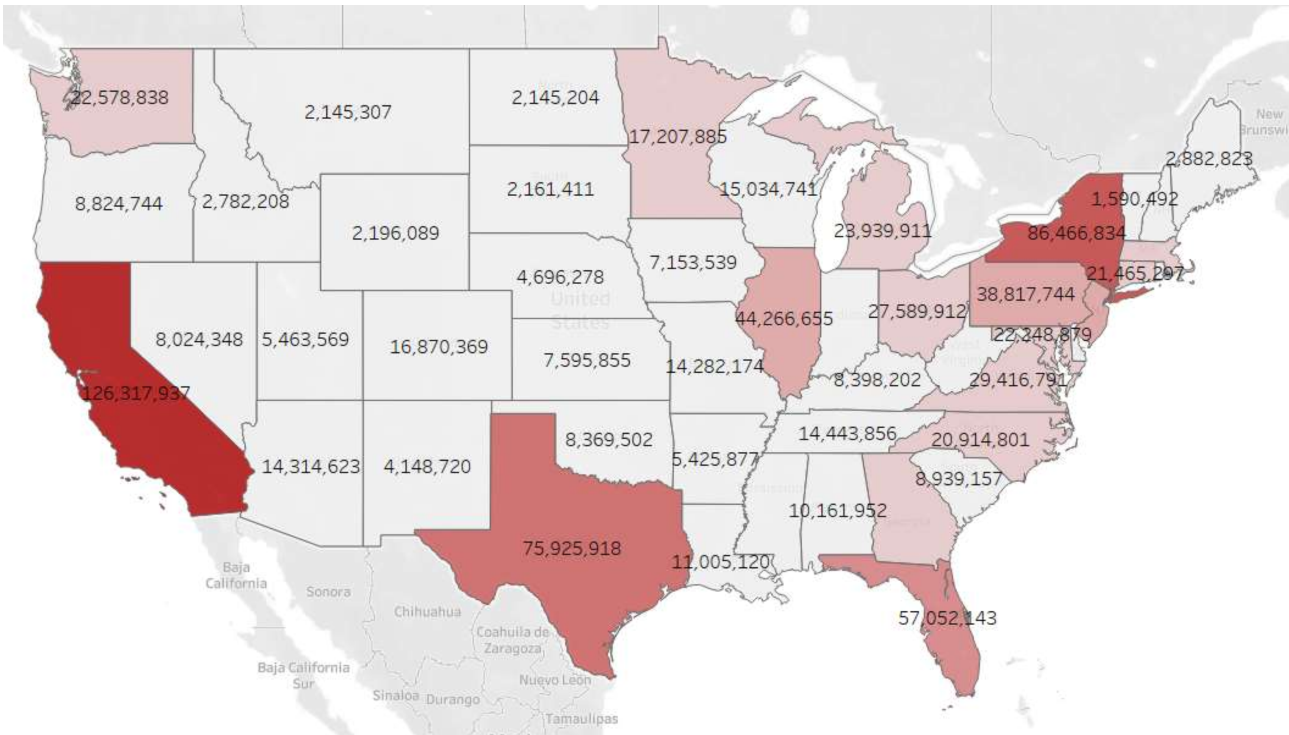


Candidate Predictors

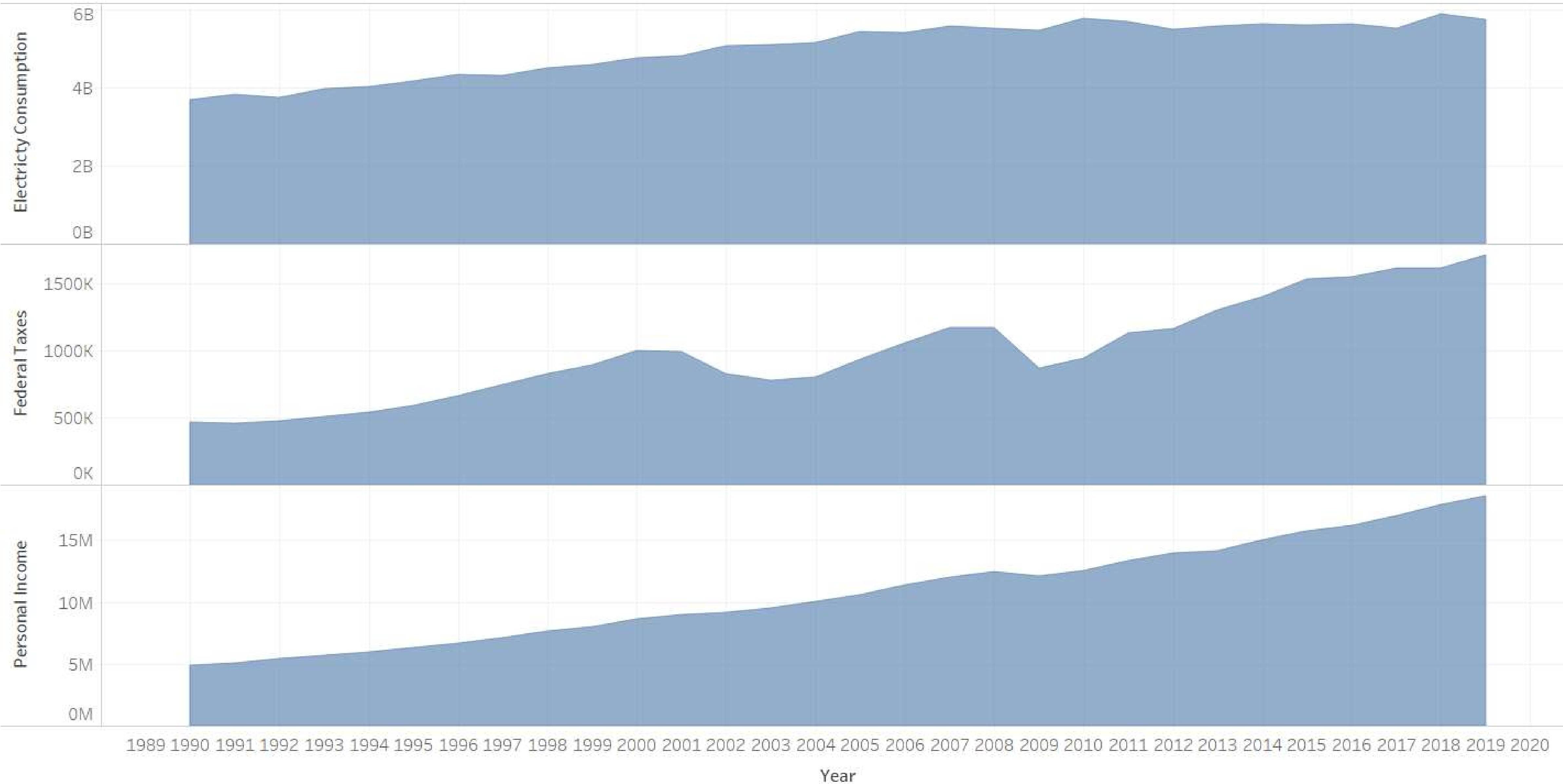
Personal Income ($, Millions)



Federal Taxes Paid ($, Thousands)



Macro Variables Time Series



The plots of sum of Electricty Consumption, sum of Federal Taxes and sum of Personal Income for Year.

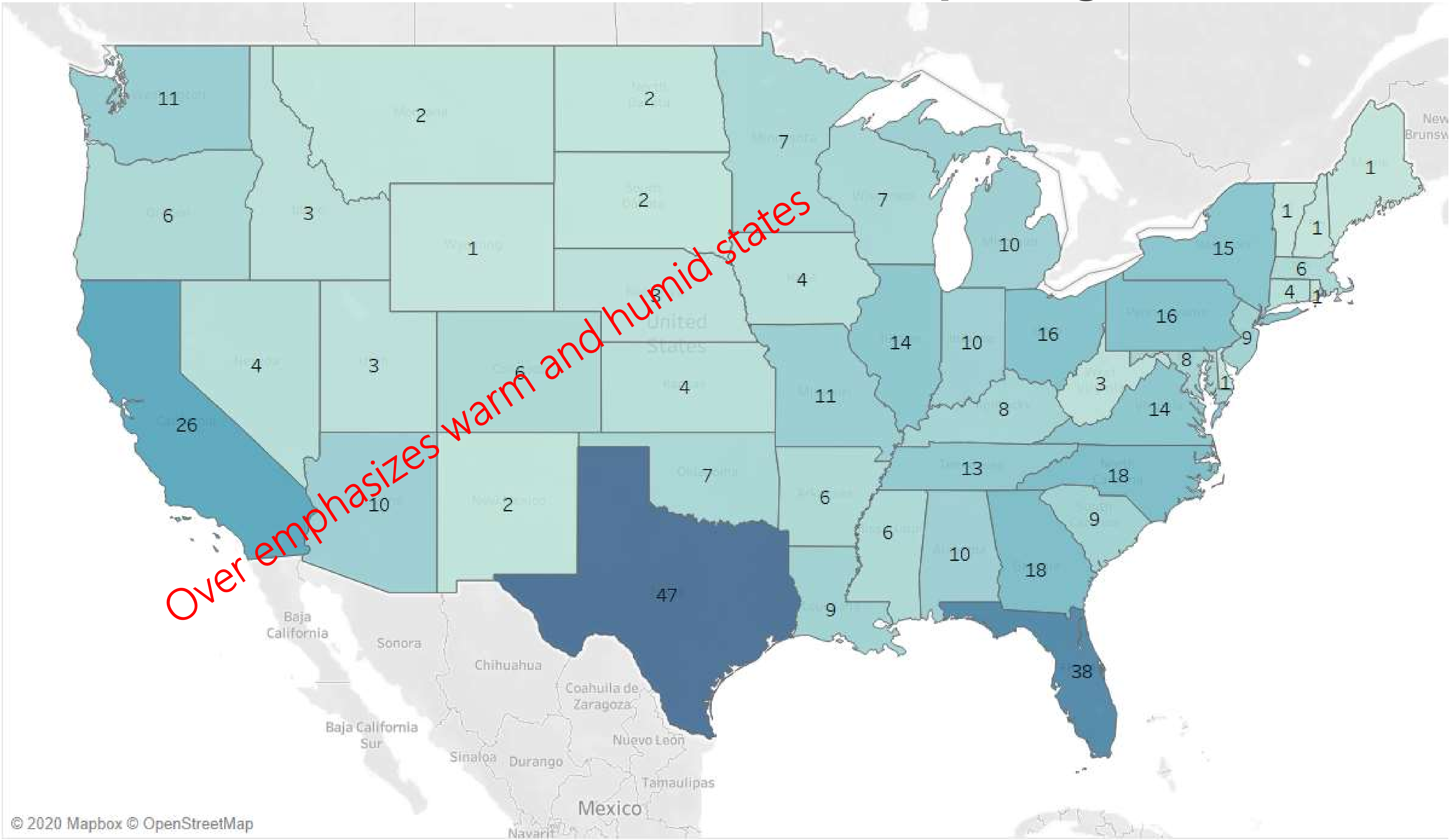
Apportionment by Raw 2019 Data

What happens when we take an oversimplified approach?

Let's explore apportionment based on the values of our drivers for 2019.

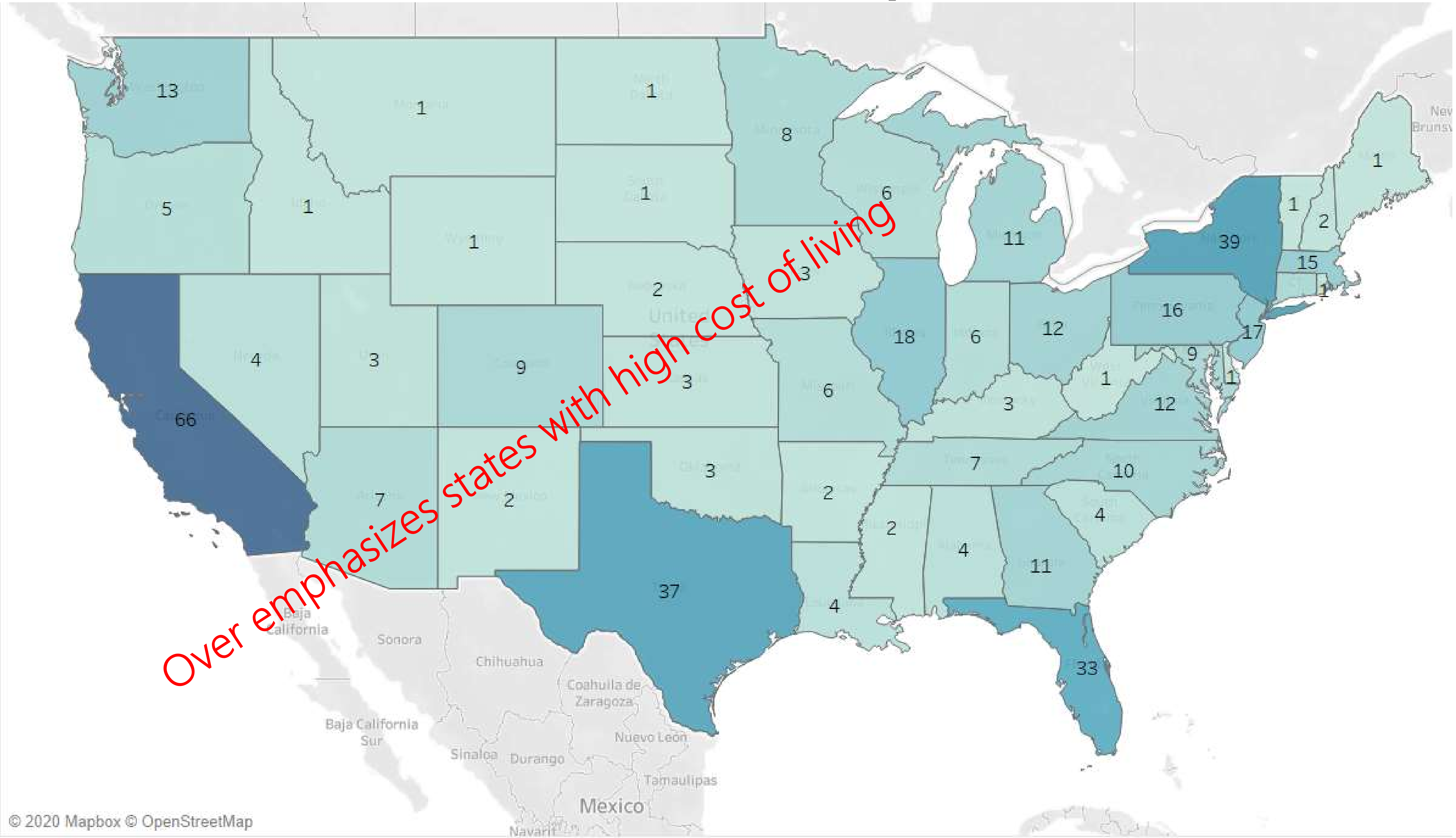
The shortcomings can be interesting.

2019 Residential Electricity Usage



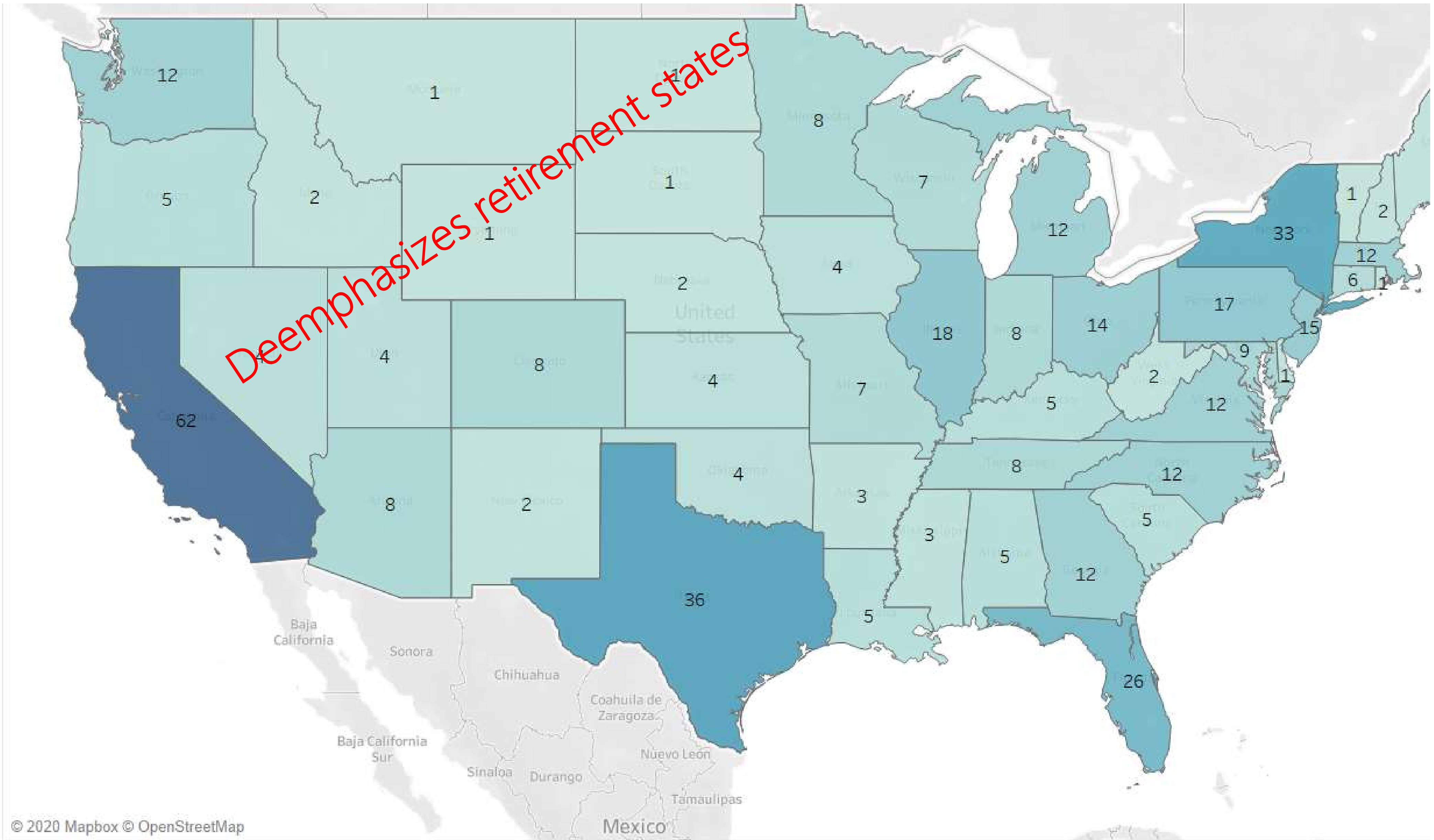
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Electrical. The marks are labeled by sum of Electrical. Details are shown for State. The view is filtered on State, which excludes AK and HI.

2019 Federal Tax Payments



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Taxes. The marks are labeled by sum of Taxes. Details are shown for State. The view i filtered on State, which excludes AK and HI.

2019 Residential Personal Income



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Personal Income. The marks are labeled by sum of Personal Income. Details are shown for State. The view is filtered on State, which excludes AK and HI.

Methodology

Since using 2019 values to drive apportionment is obviously not feasible, a more thoughtful method is required.

Using 30 years of population data along with predictors will yield smarter population estimations.

After a sensible model type is identified, a unique model will be fitted for each state. These models will be used to estimate the current population to drive apportionment.

3 Candidate Models

Option #1 Personal Income, Residential Electricity Usage & Federal Taxes

```
› Call:
› lm(formula = AZ_pop ~ AZ_pie + AZ_elct + AZ_tax, data = df)

› Residuals:
›     Min      1Q  Median      3Q     Max
› -284248  -88295   35961   89675  158123

› Coefficients:
               Estimate   Std. Error  t value  Pr(>|t|)
› (Intercept)  2.340e+06  1.408e+05   16.617   2.29e-15 ***
› AZ_pie       8.184e+00  1.510e+00   5.421    1.11e-05 ***
› AZ_elct      3.475e-02  4.900e-03   7.092    1.57e-07 ***
› AZ_tax       -4.474e-03 1.422e-02   -0.315   0.756
› ---
› Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

› Residual standard error: 123000 on 26 degrees of freedom
› Multiple R-squared:  0.9888,  Adjusted R-squared:  0.9875
› F-statistic: 765.9 on 3 and 26 DF,  p-value: < 2.2e-16
```

Option #2 Personal Income, Residential Electricity Usage & Adjusted Federal Taxes

```
› Call:
› lm(formula = AZ_pop ~ AZ_pie + AZ_elct + log(AZ_tax), data
  = df)

› Residuals:
›     Min      1Q  Median      3Q     Max
› -293504  -72892   17054   79210  184806

› Coefficients:
›              Estimate   Std. Error  t value Pr(>|t|)
› (Intercept)  -1.387e+06 2.240e+06   -0.619    0.541
› AZ_pie       7.004e+00  9.158e-01   7.648 4.07e-08 ***
› AZ_elct      3.139e-02  5.083e-03   6.174 1.57e-06 ***
› log(AZ_tax)  2.466e+05  1.485e+05   1.661    0.109
› ---
› Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

› Residual standard error: 117200 on 26 degrees of freedom
› Multiple R-squared:  0.9898,  Adjusted R-squared:  0.9887
› F-statistic: 844.8 on 3 and 26 DF,  p-value: < 2.2e-16
```

Option #3 Personal Income & Residential Electricity Usage

```
› Call:
› lm(formula = AZ_pop ~ AZ_pie + AZ_elct, data = df)

› Residuals:
›     Min      1Q  Median      3Q     Max
› -291773  -90137   34472   90230  148100

› Coefficients:
›              Estimate   Std. Error  t value Pr(>|t|)
› (Intercept)  2.327e+06  1.323e+05   17.590 2.57e-16 ***
› AZ_pie       7.786e+00  8.105e-01   9.606 3.34e-10 ***
› AZ_elct      3.505e-02  4.725e-03   7.418 5.58e-08 ***
› ---
› Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

› Residual standard error: 120900 on 27 degrees of freedom
› Multiple R-squared:  0.9888,  Adjusted R-squared:  0.9879
› F-statistic:  1188 on 2 and 27 DF,  p-value: < 2.2e-16
```

The first two options will be eliminated from contention.

3 Candidate Models

Option #4 Personal Income, Residential Electricity Usage, Adj. Federal Taxes & Growth

```
Call:
lm(formula = AZ_pop ~ AZ_pie + AZ_elct + year + log(AZ_tax)
- 1, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-293811  -76202   16842   81127  176575

Coefficients:
              Estimate   Std. Error   t value Pr(>|t|)
AZ_pie        7.173e+00  8.672e-01    8.272 9.40e-09 ***
AZ_elct       3.194e-02  5.068e-03    6.302 1.14e-06 ***
year         -4.032e+02  1.137e+03   -0.355    0.726
log(AZ_tax)   2.077e+05  1.495e+05    1.390    0.176
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 117800 on 26 degrees of freedom
Multiple R-squared:  0.9996,  Adjusted R-squared:  0.9996
F-statistic: 1.78e+04 on 4 and 26 DF,  p-value: < 2.2e-16
```

Option #5 Personal Income, Residential Electricity Usage, Adj. Federal Taxes & Growth

```
Call:
lm(formula = AZ_pop ~ AZ_pie + AZ_elct + year, data = df)

Residuals:
   Min     1Q Median     3Q    Max
-76534 -34973 -10239  28628  95078

Coefficients:
              Estimate   Std. Error   t value Pr(>|t|)
(Intercept)  -1.959e+08  1.808e+07  -10.834 3.90e-11 ***
AZ_pie       -1.493e+00  9.153e-01   -1.631    0.115
AZ_elct       2.543e-02  2.212e-03   11.493 1.08e-11 ***
year          9.997e+04  9.119e+03   10.963 3.02e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51980 on 26 degrees of freedom
Multiple R-squared:  0.998,   Adjusted R-squared:  0.9978
F-statistic:  4330 on 3 and 26 DF,  p-value: < 2.2e-16
```

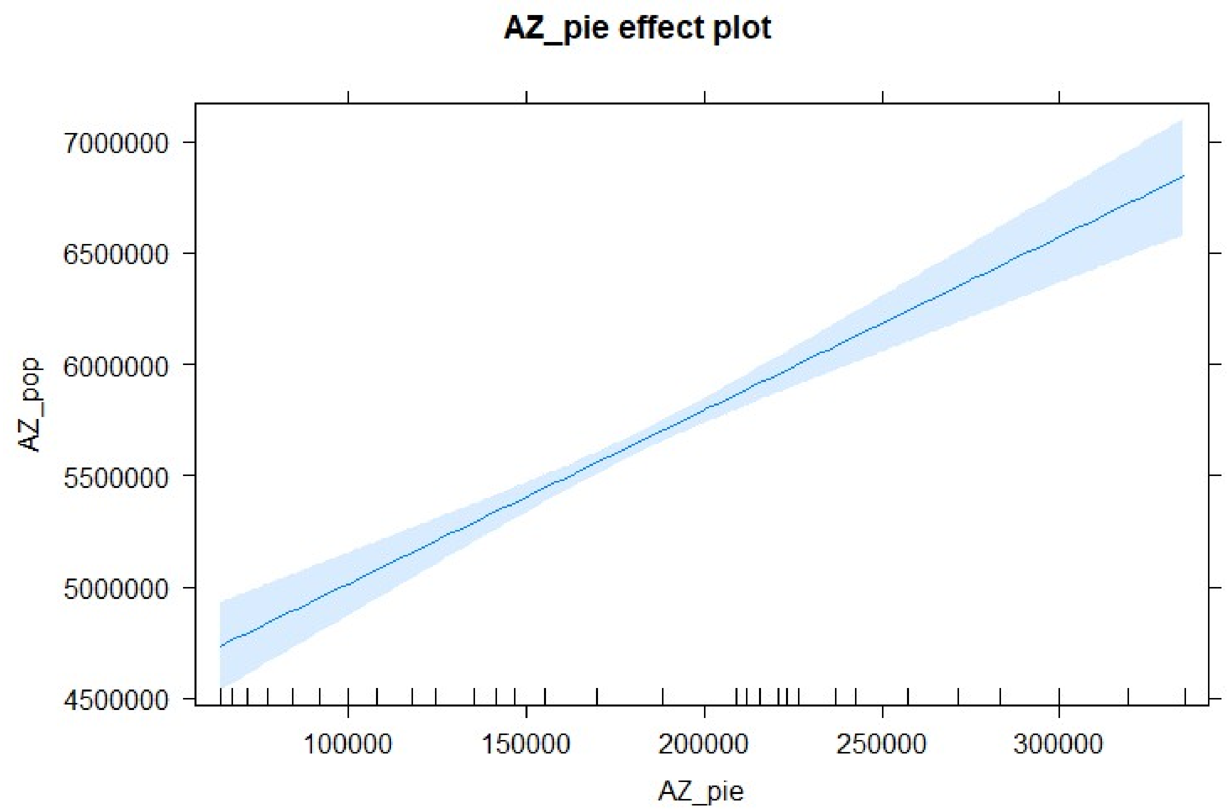
Option #3 Verification

```
Call:
lm(formula = IL_pop ~ IL_pie + IL_elct, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-296471  -51027   26994   92376  188510

Coefficients:
              Estimate   Std. Error   t value Pr(>|t|)
(Intercept)   9.624e+06  2.886e+05   33.346  < 2e-16 ***
IL_pie        9.977e-01  3.010e-01    3.315  0.00262 **
IL_elct       2.776e-02  4.661e-03    5.956 2.38e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 135400 on 27 degrees of freedom
Multiple R-squared:  0.9062,  Adjusted R-squared:  0.8992
F-statistic: 130.4 on 2 and 27 DF,  p-value: 1.34e-14
```

The Option #3 will be selected.

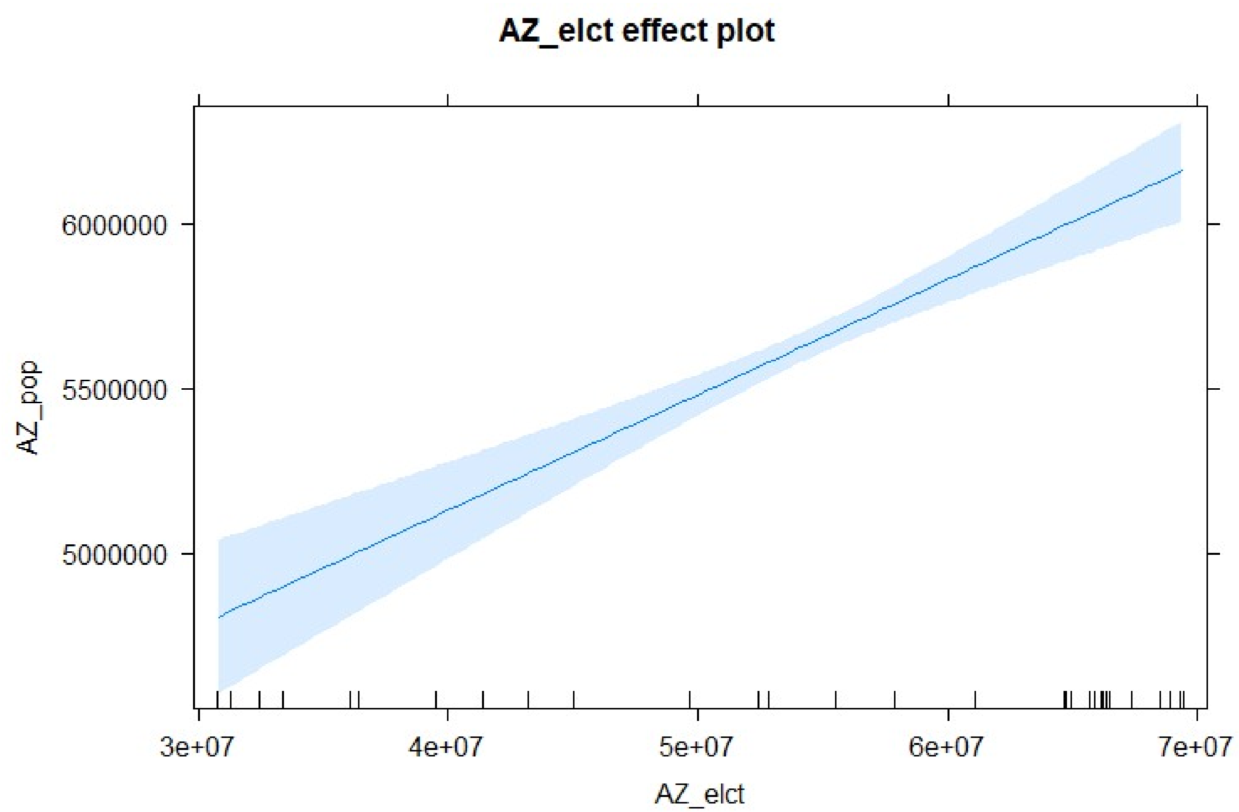
Personal Income's Influence

- Personal Income has a large influence on predicting population. For Arizona, Personal Income result in 4.7MM on the low end and 6.8MM on the high end.
- Also the T-Value is 9.6 which tells us that Personal Income's impact is large in comparison to its standard error.
- Additionally the slope of the relationship makes sense. The electrical usage goes up, so does the population.



Residential Electrical Consumption's Influence

- Residential electricity consumption also has a large influence on predicting population. For Arizona, residential electricity consumption result in 4.6MM on the low end and 6.3MM on the high end.
- Also the T-Value is 7.4 which tells us that residential electricity consumption's impact is decent in comparison to its standard error.
- Additionally the slope of the relationship makes sense. The residential electricity consumption goes up, so does the population.



50 State Population Forecasts

- Option 3 was fitted for each of the 50 states.
- The fitted results were dumped into a data frame
- After the data frame was exported, only the 2019 population estimates were used for apportionment.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Apportionment.R × | fitted ×

Source on Save

```
multi.HI = lm(HI_pop~HI_pie+HI_elct, data=df)
fitted$HI <- fitted(multi.HI)

multi.IA = lm(IA_pop~IA_pie+IA_elct, data=df)
fitted$IA <- fitted(multi.IA)

multi.ID = lm(ID_pop~ID_pie+ID_elct, data=df)
fitted$ID <- fitted(multi.ID)

multi.IL = lm(IL_pop~IL_pie+IL_elct, data=df)
fitted$IL <- fitted(multi.IL)

multi.IN = lm(IN_pop~IN_pie+IN_elct, data=df)
fitted$IN <- fitted(multi.IN)

multi.KS = lm(KS_pop~KS_pie+KS_elct, data=df)
fitted$KS <- fitted(multi.KS)

multi.KY = lm(KY_pop~KY_pie+KY_elct, data=df)
fitted$KY <- fitted(multi.KY)

multi.LA = lm(LA_pop~LA_pie+LA_elct, data=df)
fitted$LA <- fitted(multi.LA)

multi.MA = lm(MA_pop~MA_pie+MA_elct, data=df)
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

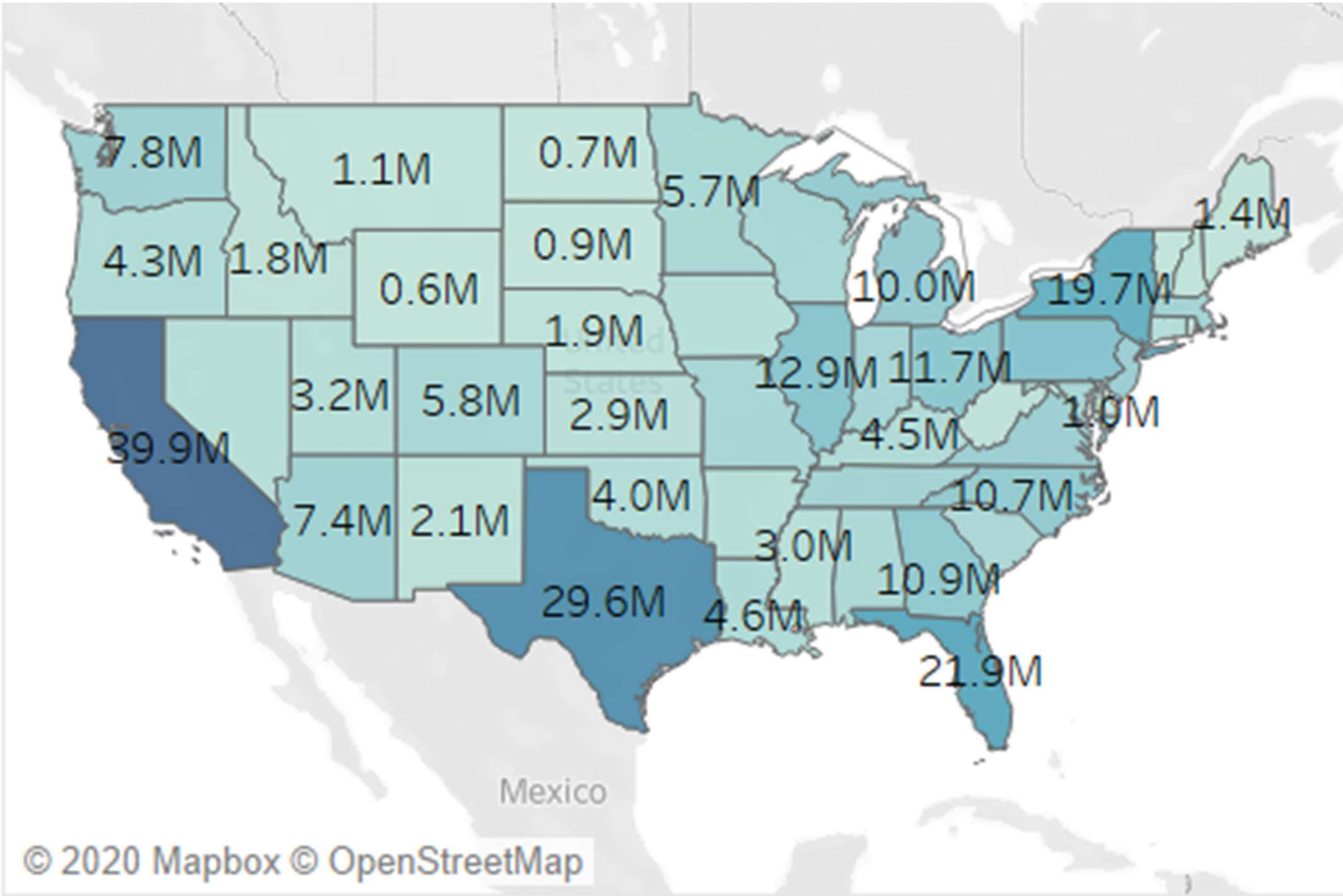
Go to file/function Addins

Apportionment.R × | fitted ×

Filter

	Year	AK	AL	AR	AZ	CA	CO	CT	DC	DE	FL	GA
6	1995	599065.8	4274213	2522188	4307174	31545151	3761047	3328109	568930.8	722674.8	14579731	7
7	1996	604630.0	4314564	2552270	4487219	32066203	3857390	3341433	570614.2	733757.3	14847640	7
8	1997	605699.9	4320177	2562358	4617591	32469283	3936900	3347242	569691.3	738485.7	14979762	7
9	1998	612212.6	4389695	2627190	4760804	33008199	4034933	3361745	573819.6	753968.0	15537219	7
10	1999	621291.4	4399091	2622439	4875723	33228173	4133141	3396638	577763.7	770722.1	15598338	7
11	2000	626675.5	4450834	2665342	5122399	34117726	4299909	3413367	581138.4	786697.7	16076239	8
12	2001	634215.7	4449081	2684956	5263860	33877881	4380670	3436096	586008.0	808380.7	16349254	8
13	2002	641365.9	4502474	2706826	5318936	34013394	4485145	3455217	590296.4	824635.9	16758529	8
14	2003	649412.3	4512189	2722074	5479606	34993368	4530607	3485009	589330.1	835002.2	17138213	8
15	2004	659283.1	4560325	2737220	5674889	35305768	4537686	3496943	597122.6	846617.2	17456915	8
16	2005	666320.9	4613986	2811058	5932701	35864922	4687469	3530944	604699.8	862654.4	18006518	9
17	2006	678011.3	4668398	2824434	6224001	36798928	4809754	3516702	603129.4	861594.7	18457588	9
18	2007	686824.3	4707391	2855891	6459454	36906514	4941541	3550111	614111.2	875591.2	18695522	9
19	2008	700882.1	4717309	2865526	6395627	37276695	4983156	3541511	613623.7	872056.4	18526444	9
20	2009	700350.2	4695313	2844090	6279785	36911442	4903242	3531692	613563.9	870290.9	18279055	9
21	2010	709159.3	4795998	2944534	6279461	36783447	5008329	3560970	628207.9	888536.9	18869864	9
22	2011	723000.1	4776252	2946621	6408110	37284216	5110865	3563722	631715.3	901704.9	18895818	9
23	2012	730595.7	4752875	2934759	6478773	37880916	5170723	3563061	632245.1	900586.9	18918290	9
24	2013	725430.4	4773280	2948102	6538231	37863743	5261088	3571412	635008.3	905898.7	18982729	9
25	2014	731146.3	4826864	2977157	6598448	38298197	5318044	3570865	642406.4	919851.8	19552918	9
26	2015	737267.1	4842367	2983527	6763836	38772974	5413098	3583023	667644.7	940712.5	20237701	10
27	2016	731389.3	4859892	2973562	6893520	38933269	5488889	3580930	672337.8	944151.0	20508464	10
28	2017	738139.9	4857489	2956582	7063683	39529059	5566125	3577265	670849.1	953496.9	20903988	10
29	2018	740690.1	4947104	3063916	7242005	39792033	5747426	3616932	682991.8	982970.7	21546170	10
	2019	744388.6	4966841	3056278	7370905	39934823	5840146	3605498	684530.1	991798.6	21924207	10

50 State Population Forecasts

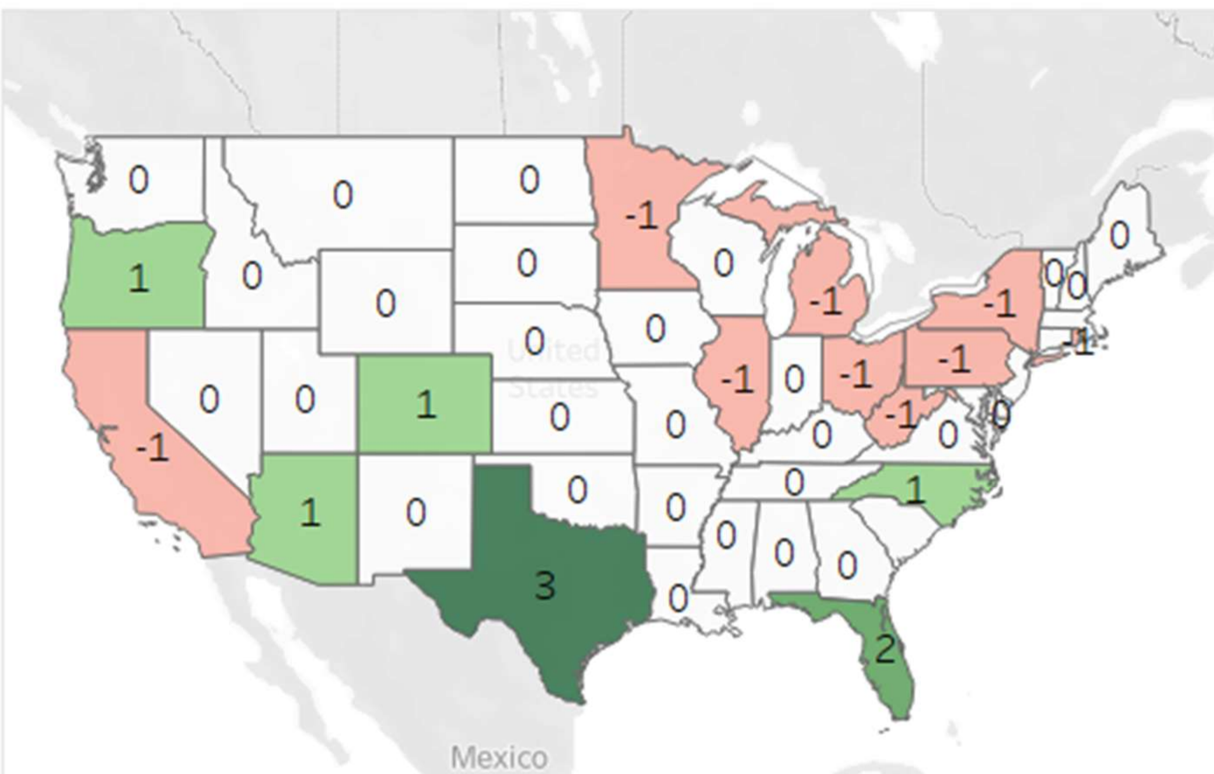


Apportionment

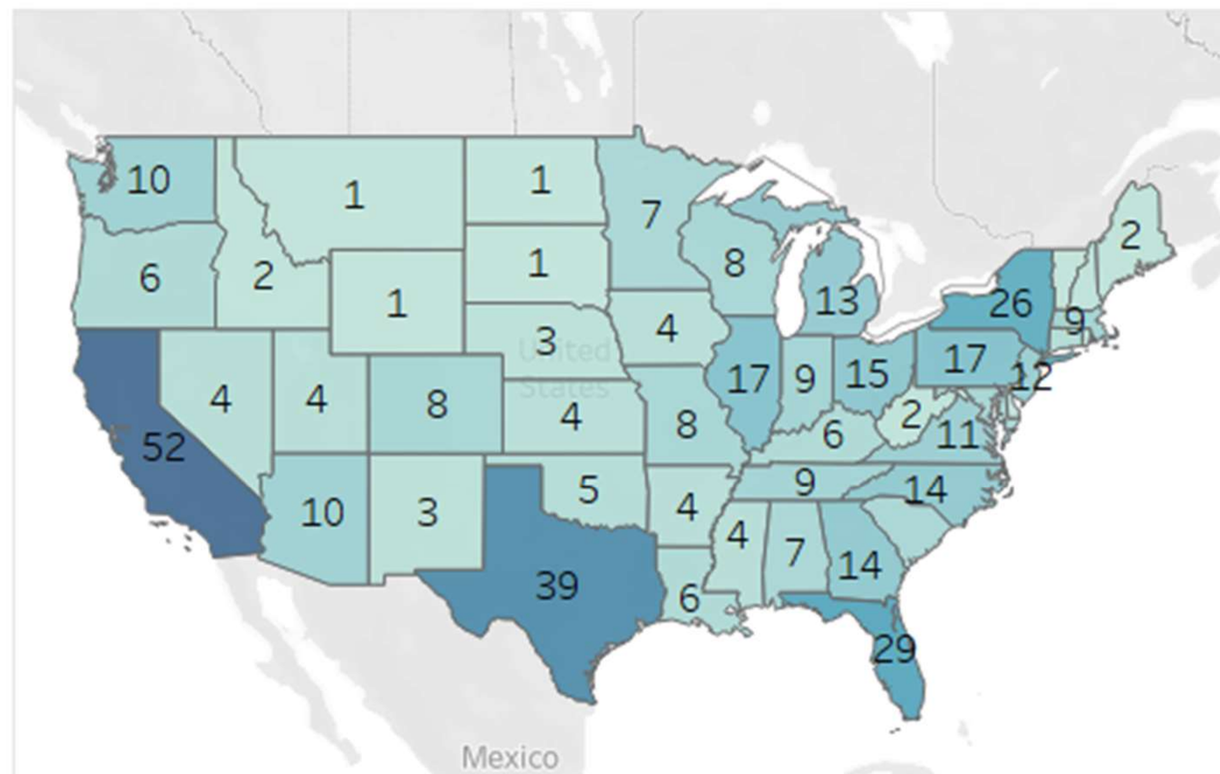
- Divide the 2019 fitted population values by 761k for each state.
- Round to nearest integer.

- This is very simple!

Winners & Losers

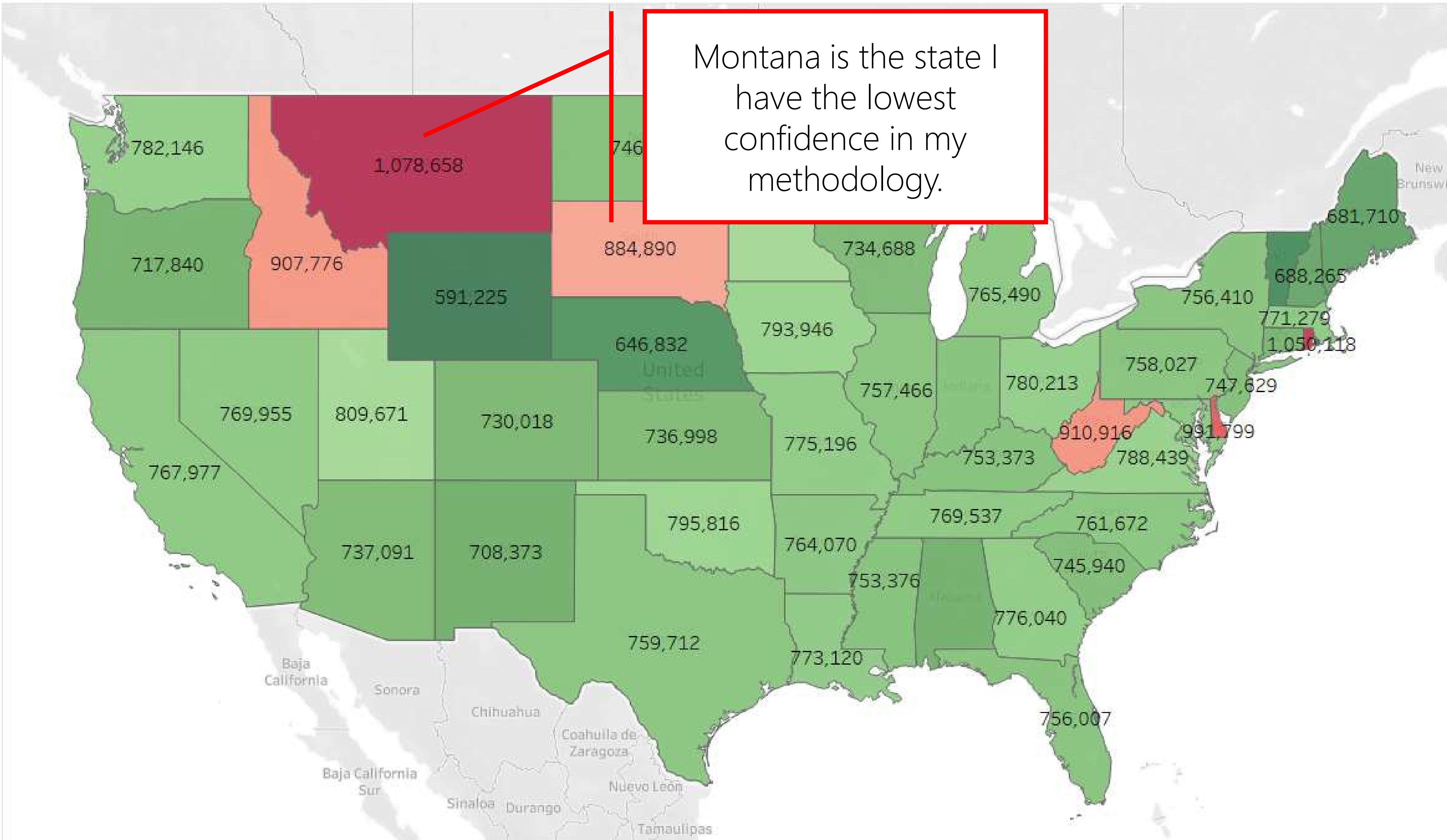


Projected Members



Apportionment Paradox Check

Residents per Seat



Potential Improvements

1. Think though population growth and consider potential ways to include it into the regressions.
2. Find an additional macro variables to compliment Personal Income and Electricity Usage.

Just for Fun

Residents per Elector