Congressional Reapportionment

Matt Burns

Fall 2020

https://github.com/mattburns963/680.git

Problem and Hypothesis

When the constitution was written over 200 years ago, it required that the people living in the United States be enumerated.

> "Article I, Section 2 of the U.S. Constitution mandates that an apportionment of representatives among the states must be carried out every 10 years. Therefore, apportionment is the original legal purpose of the decennial census, as intended by our Nation's Founders. Apportionment is the process of dividing the 435 memberships, or seats, in the U.S. House of Representatives among the 50 states, based on the state population counts that result from each decennial census. The apportionment results will be the first data published from the 2020 Census, and those results will determine the amount of political representation each state will have in Congress for the next 10 years." (Census)

Individually counting residents is still done to this day. Grant it we can fill out an online form for this enumeration, but it is an old fashioned and laborious way to understand the states' population.

For this paper I will be investigating an alternate methodology of reapportionment for the congressional seats. I will be comparing electricity usage, Personal Consumption Expenditures and federal taxes to estimate the states' population which will be used to estimate the upcoming reapportionment.

Data Retrieval Methodology

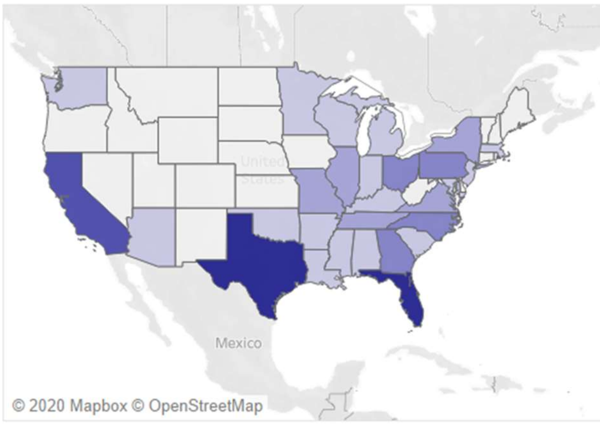
The data included 30 years of electrical usage data, federal tax data, personal income at the state level. These will be the independent variable. Population data (cut the same way) will be the dependent variable. This data is easily obtained from the following websites.

- https://www.eia.gov/electricity/data/state/
- https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1
- https://www2.census.gov/programs-surveys/susb/tables/time-series/us_state_totals_2007-2017.xlsx
- https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=30&isuri=1&major_area=0&area=xx&year=2019,2018,2017,2016,2015,2014,2013,2012,2011,2010,2009,2008,2007,2006,2005,2004,2003,2002,2001,2000,1999,1998,1997,1996,1995,1994,1993,1992,1991,1990&tableid=9&category=49&area_type=0&year_end=-1&classification=non-industry&state=0&statistic=70&yearbegin=-1&unit_of_measure=levels

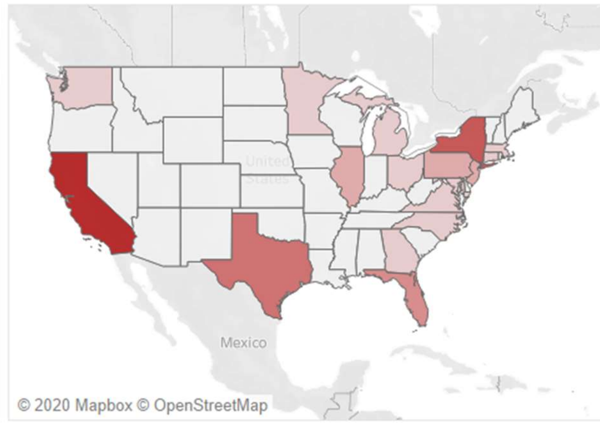
The U.S. Energy Information Administration and the Census Bureau are independent governmental bodies. The Bureau of Economic Analysis is a part of the Department of Commerce.

The data was based on 30 years, but the illustration before shows where the metrics stood for the last apportionment.

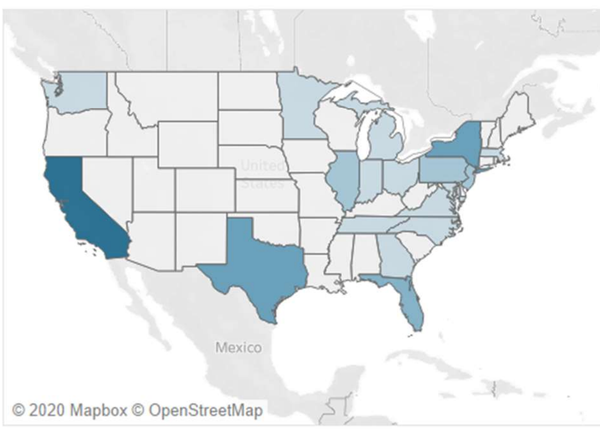
2010 Total Residential Electricity Usage



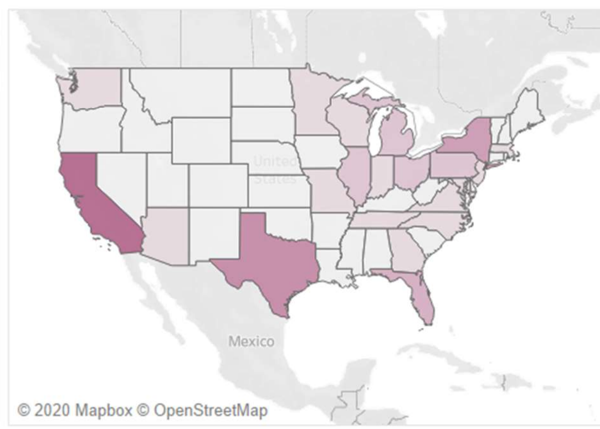
2010 Total Federal Taxes Paid



2010 Total Personal Income



2010 Total Population



Exploratory Oversimplified Approach

As a way to explore the data more and investigate simple apportionment methodologies, I investigated simple allocations of congressional seats based on the following the 2019 values of the following three metrics.

- Residential Electricity Usage
- Federal Taxes Paid
- Personal Income

As one would guess these methodologies had serious shortcomings. The way they interact with residents (i.e. per capita data) varies too greatly both by state and over time. Residential electricity favors warm and humid states. Using Federal taxes is unfavorable to states with large retiree populations. Finally, personal income favors expensive states.

Proper Analysis Methodology

Once the data was organized and cleaned, I explored 5 different candidate regression equations. They included combinations of our three main variables, a proxy for growth and adjusted value for taxes.

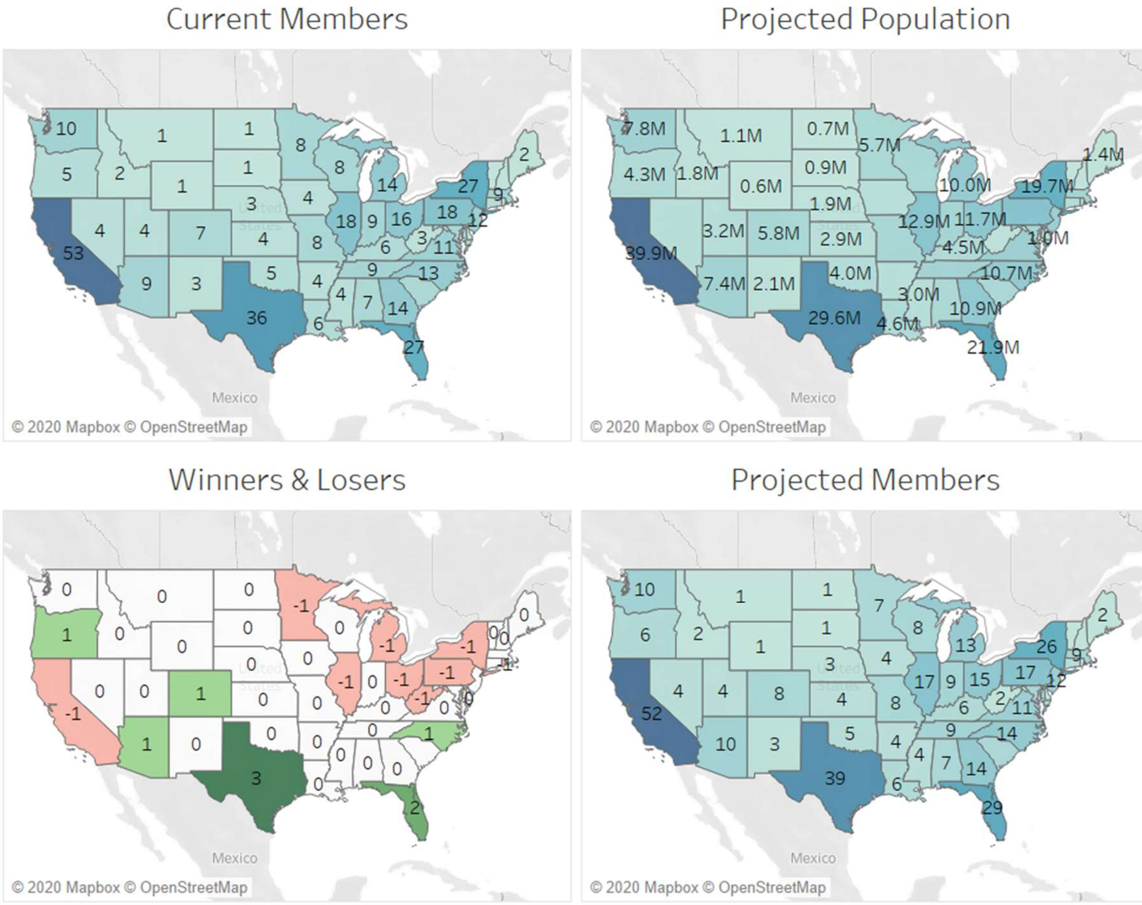
As is often the case, a simple model was the winner. The other candidates had coefficients that were incorrectly inversely related population. As an example, as electricity usage goes up, population shouldn't go down.

The winner was a simple linear model that used electricity consumption and personal income.

```
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
```

Then I had R Studio run though all 50 states and dump the forecasted state populations into a dataframe to export to Tableau. The same simple regression was used for each state.

Finally, I took the forecasted population for 2019 and apportioned the congressional seats. This done in two quick steps: Dividing each state's population by 761,000 and then rounding to the nearest integer. The results are illustrated on the next page.



These results largely align with the forecasts in the media with one exception, Montana. Many of the current estimates for Montana have it gaining a seat and having two districts. If this happens, it is likely to come at the expense of Nebraska.

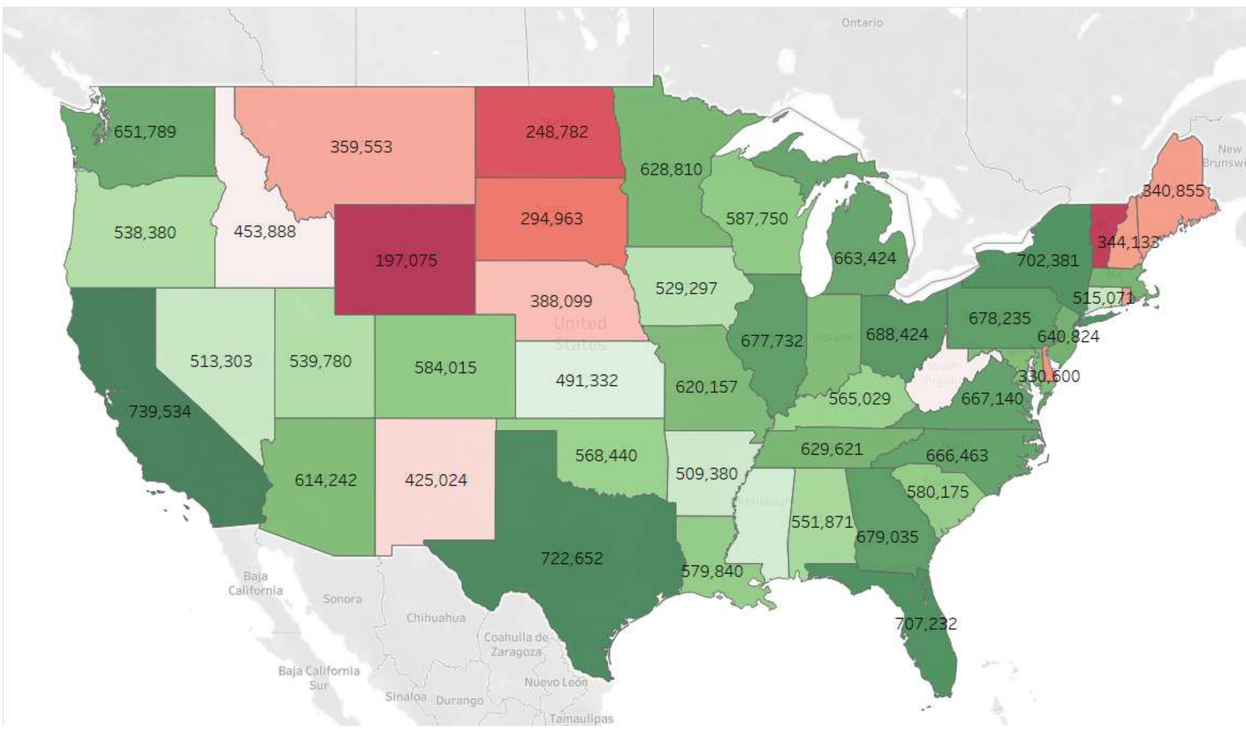
Conclusions

1. The following states will gain congressional seats:
   - Oregon
   - Arizona
   - Colorado
   - North Carolina
   - Florida (2)
   - Texas (3)
2. The following states will lose congressional seats:
   - California
   - Minnesota
   - Michigan
   - Ohio
   - Pennsylvania
   - New York
   - Rhode Island
   - West Virginia
   - Illinois
3. A simple (yet strong) regression model can be built for each state using electrical consumption and personal income
4. Tax payments will not work for this type of estimation. They vary too greatly during recessions.
5. The apportionment is fairly equitable with an average of 773k people per seat with a standard deviation of 87k. The apportionment paradox has little impact.

Appendix: Electoral College Influence

Out of curiosity I added 2 to the number of congressional districts per state to forecast each state's electoral college votes. Then I divided the population by the electoral votes to identify which states had grater influence that their population would suggest.

Residents per Elector



The Great Plains state as well as a few New England states are the biggest winner in this system.

APA references

Articles relating to the congressional reapportionment are below. There are also articles that electricity usage, tax payments and personal income. There are also articles to help inform my reapportionment methodology. They helped reinforce the idea that congressional seats from the house are distributed, not electoral votes. It is a simple but important clarification.

Chen, James. (Updated Sep 29, 2020) Nonfarm Payroll. Retrieved from
https://www.investopedia.com/terms/n/nonfarmpayroll.asp
This provides background regarding employment numbers. I will use it to smarten the usage of payroll data into regression that are used to understand the relationship between payrolls population.

U.S. Bureau of Labor Statistics. (October 30, 2020) Employment Cost Index Summary. Retrieved from
https://www.bls.gov/news.release/eci.nr0.htm
This news release provides near current payroll information. It provides context on how I may want to adjust the 2020 payroll number to calculate the population estimation.

Congressional Research Service. (August 4, 2010) The U.S. House of Representatives Apportionment Formula in Theory and Practice.
https://www.everycrsreport.com/reports/R41357.html
This article describes how the apportionment process works. It will be used to at the very end of my project to extrapolate my population estimations to actual congressional members.

House.gov. (downloaded Nov. 5, 2020) Congressional Apportionment, retrieved from
https://history.house.gov/Institution/Apportionment/Apportionment/
This article describes how the apportionment process works. It will be used to at the very end of my project to extrapolate my population estimations to actual congressional members.

United States Census Bureau. (Downloaded Nov. 5, 2020), About Congressional Apportionment Retrieved from
https://www.census.gov/topics/public-sector/congressional-apportionment/about.html
This page explains the constitutional and legal requirements to perform a census and how it is to be applied in apportionment. It is reference only.

Talk Business & Politics. (June 29, 2020) Global electricity use per capita increasing quicker than population. Retrieved from
https://talkbusiness.net/2020/06/global-electricity-use-per-capita-increasing-quicker-than-population/
This article compares population growth and electricity usage. I will use it to better understand the relationship between energy consumption and population.

Zabel, Graham. (April 20, 2009) Peak People: The Interrelationship Between Population Growth And Energy Resources, Retrieved from

https://www.resilience.org/stories/2009-04-20/peak-people-interrelationship-between-population-growth-and-energy-resources/
"This paper investigates the link between population growth, energy resources and carrying capacity at a global level" (Graham). I will use high level information to better understand the relationship between energy consumption and population.

United States Census Bureau. (Downloaded Nov. 5, 2020) Census in the Constitution, Retrieved from
https://www.census.gov/programs-surveys/decennial-census/about/census-constitution.html1
This page explains the constitutional and legal requirements to perform a census and how it is to be applied in apportionment. It is reference only.

Trading Economics. (Downloaded Nov. 5, 2020) United States - Personal consumption expenditures per capita, Retrieved from
https://tradingeconomics.com/united-states/personal-consumption-expenditures-per-capita-fed-data.html
This data summary highlights the interplay between energy consumption and population. This will be background information to help me design better population forecasts using energy.

NBC News. (Downloaded Nov. 5, 2020) New electoral map comes into focus ahead of 2020 census, Retrieved from
https://www.nbcnews.com/politics/meet-the-press/new-electoral-map-comes-focus-ahead-2020-census-n1110546
This article discusses analytical projections of the congressional apportionment based on the 2020 census. It can be used as a sanity check for my projection.

Appendix: R Code

```
library(readxl)
library(caret)
library(Rmisc)
library(ggplot2)
library(forecastML)
library(effects)
library(DataCombine)
require("forecast")
require("expsmooth") # required for the data
library(smooth)

# pull in monthly economic data from Jan. 2000 thru Dec. 2019
df <- read_excel("C:/Users/burns/OneDrive/Desktop/Matt/Grad School/DSC 680/Project 3/State Data.xlsx")
names(df)[1] <- "year"

# clean dataframe
df[!complete.cases(df),]

# Test candidate models
multi.fit1 = lm(AZ_pop~AZ_pie+AZ_elct+AZ_tax, data=df)
summary(multi.fit1)

multi.fit2 = lm(AZ_pop~AZ_pie+AZ_elct+log(AZ_tax), data=df)
summary(multi.fit2)

multi.fit3 = lm(AZ_pop~AZ_pie+AZ_elct, data=df)
summary(multi.fit3)

multi.fit4 = lm(AZ_pop~AZ_pie+AZ_elct+year+log(AZ_tax)-1, data=df)
summary(multi.fit4)

multi.fit5 = lm(AZ_pop~AZ_pie+AZ_elct+year, data=df)
summary(multi.fit5)

multi.fit6 = lm(IL_pop~IL_pie+IL_elct, data=df)
summary(multi.fit6)


#explore impacts from two variables
eff.fit3 <- allEffects(multi.fit3, xlevels=50)
for(i in 1:2) {plot(eff.fit3[i])}


# reference only
coefficients(multi.fit5) # model coefficients
confint(multi.fit5, level=0.95) # CIs for model parameters
fitted(multi.fit5) # predicted values
residuals(multi.fit5) # residuals
anova(multi.fit5) # anova table
vcov(multi.fit5) # covariance matrix for model parameters
influence(multi.fit5) # regression diagnostics

# make new data frame
fitted <- data.frame("Year" = df$year)


multi.AK = lm(AK_pop~AK_pie+AK_elct, data=df)
fitted$AK <- fitted(multi.AK)
```

```
multi.AL = lm(AL_pop~AL_pie+AL_elct, data=df)
fitted$AL <- fitted(multi.AL)

multi.AR = lm(AR_pop~AR_pie+AR_elct, data=df)
fitted$AR <- fitted(multi.AR)

multi.AZ = lm(AZ_pop~AZ_pie+AZ_elct, data=df)
fitted$AZ <- fitted(multi.AZ)

multi.CA = lm(CA_pop~CA_pie+CA_elct, data=df)
fitted$CA <- fitted(multi.CA)

multi.CO = lm(CO_pop~CO_pie+CO_elct, data=df)
fitted$CO <- fitted(multi.CO)

multi.CT = lm(CT_pop~CT_pie+CT_elct, data=df)
fitted$CT <- fitted(multi.CT)

multi.DC = lm(DC_pop~DC_pie+DC_elct, data=df)
fitted$DC <- fitted(multi.DC)

multi.DE = lm(DE_pop~DE_pie+DE_elct, data=df)
fitted$DE <- fitted(multi.DE)

multi.FL = lm(FL_pop~FL_pie+FL_elct, data=df)
fitted$FL <- fitted(multi.FL)

multi.GA = lm(GA_pop~GA_pie+GA_elct, data=df)
fitted$GA <- fitted(multi.GA)

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)
fitted$MA <- fitted(multi.MA)

multi.MD = lm(MD_pop~MD_pie+MD_elct, data=df)
fitted$MD <- fitted(multi.MD)

multi.ME = lm(ME_pop~ME_pie+ME_elct, data=df)
```

```
fitted$ME <- fitted(multi.ME)

multi.MI = lm(MI_pop~MI_pie+MI_elct, data=df)
fitted$MI <- fitted(multi.MI)

multi.MN = lm(MN_pop~MN_pie+MN_elct, data=df)
fitted$MN <- fitted(multi.MN)

multi.MO = lm(MO_pop~MO_pie+MO_elct, data=df)
fitted$MO <- fitted(multi.MO)

multi.MS = lm(MS_pop~MS_pie+MS_elct, data=df)
fitted$MS <- fitted(multi.MS)

multi.MT = lm(MT_pop~MT_pie+MT_elct, data=df)
fitted$MT <- fitted(multi.MT)

multi.NC = lm(NC_pop~NC_pie+NC_elct, data=df)
fitted$NC <- fitted(multi.NC)

multi.ND = lm(ND_pop~ND_pie+ND_elct, data=df)
fitted$ND <- fitted(multi.ND)

multi.NE = lm(NE_pop~NE_pie+NE_elct, data=df)
fitted$NE <- fitted(multi.NE)

multi.NH = lm(NH_pop~NH_pie+NH_elct, data=df)
fitted$NH <- fitted(multi.NH)

multi.NJ = lm(NJ_pop~NJ_pie+NJ_elct, data=df)
fitted$NJ <- fitted(multi.NJ)

multi.NM = lm(NM_pop~NM_pie+NM_elct, data=df)
fitted$NM <- fitted(multi.NM)

multi.NV = lm(NV_pop~NV_pie+NV_elct, data=df)
fitted$NV <- fitted(multi.NV)

multi.NY = lm(NY_pop~NY_pie+NY_elct, data=df)
fitted$NY <- fitted(multi.NY)

multi.OH = lm(OH_pop~OH_pie+OH_elct, data=df)
fitted$OH <- fitted(multi.OH)

multi.OK = lm(OK_pop~OK_pie+OK_elct, data=df)
fitted$OK <- fitted(multi.OK)

multi.OR = lm(OR_pop~OR_pie+OR_elct, data=df)
fitted$OR <- fitted(multi.OR)

multi.PA = lm(PA_pop~PA_pie+PA_elct, data=df)
fitted$PA <- fitted(multi.PA)

multi.RI = lm(RI_pop~RI_pie+RI_elct, data=df)
fitted$RI <- fitted(multi.RI)

multi.SC = lm(SC_pop~SC_pie+SC_elct, data=df)
fitted$SC <- fitted(multi.SC)

multi.SD = lm(SD_pop~SD_pie+SD_elct, data=df)
fitted$SD <- fitted(multi.SD)
```

```
multi.TN = lm(TN_pop~TN_pie+TN_elct, data=df)
fitted$TN <- fitted(multi.TN)

multi.TX = lm(TX_pop~TX_pie+TX_elct, data=df)
fitted$TX <- fitted(multi.TX)

multi.UT = lm(UT_pop~UT_pie+UT_elct, data=df)
fitted$UT <- fitted(multi.UT)

multi.VA = lm(VA_pop~VA_pie+VA_elct, data=df)
fitted$VA <- fitted(multi.VA)

multi.VT = lm(VT_pop~VT_pie+VT_elct, data=df)
fitted$VT <- fitted(multi.VT)

multi.WA = lm(WA_pop~WA_pie+WA_elct, data=df)
fitted$WA <- fitted(multi.WA)

multi.WI = lm(WI_pop~WI_pie+WI_elct, data=df)
fitted$WI <- fitted(multi.WI)

multi.WV = lm(WV_pop~WV_pie+WV_elct, data=df)
fitted$WV <- fitted(multi.WV)

multi.WY = lm(WY_pop~WY_pie+WY_elct, data=df)
fitted$WY <- fitted(multi.WY)

#Export fitted population
write.csv(fitted, "C:/Users/burns/OneDrive/Desktop/Matt/Grad School/DSC 680/Project 3//blurg.csv", row.names = FALSE)
```