

It's not fair: gerrymandering by whom?

Bruce Mallory

12/2/2020

ABSTRACT

INTRODUCTION

Every ten years, after the census is completed, the 435 House Representatives are reapportioned among the states based on the current populations. Then each state redraws its electoral districts, so the districts within a state have approximately equal populations. Aside from mandating that the districts have equal populations, the Constitution is silent about the process and leaves the redrawing of districts to the states. Redistricting laws in the different states vary as does the responsibility for redistricting. In some states redistricting is done by the courts, in other states there are independent commissions. And in the majority of states, redistricting is done by the state legislator and the governor. Of these states, in 2011, twenty-three had a state legislature and a governor who belonged to the same political party.

In assessing the fairness of the redrawn electoral districts, there are four overlapping and often conflicting, perspectives.

First is the hope that districts will be politically proportional, yet that doesn't always happen. In North Carolina, in 2018, the Democratic candidates (as a group) received 48.3% of the votes cast. That year, only 3 of the 13 representatives (23%) were Democratic. And in Massachusetts in 2018, Republican candidates received 36% of the votes cast, yet none of the 9 congressional representatives were Republican.

A second concern in redistricting is the competitiveness of the districts. Clearly there is a relationship between the number of competitive districts and proportional-political-representation. In the appendix I discuss my observations of the number of uncontested and competitive districts that I observed in the years that I've analyzed 2002 to 2018).

The third perspective about fairly drawn congressional districts is stipulated by the Voting Rights Act of 1965 and focuses on a minority groups' ability to elect representatives of their choice. The law addresses instances where this ability is diminished during redistricting and focuses on assuring that there are "majority-minority districts."

And the fourth perspective on drawing "fair" electoral districts is most clearly articulated in California's 2010 voter initiative that established a redistricting commission tasked with creating "communities of interest." The goal was to create districts that allow like-minded communities to have representation in congress.

The analysis that I am doing in this paper will only focus on proportional-political-representation and will not address minority representation or community representation. All the while understanding that the Voting Rights Act impacts the drawing of districts in a manner that can skew proportional-political-representation (ironically, assuring majority-minority districts is a form of packing and can dilute Democratic representation). And the goal of keeping communities of interest together can also impact proportional-political-representation.

In my current analysis, I will focus on the proportion of seats to votes. Specifically, I define the proportional-political-representation coefficient (PPR) to be the (% of seats) divided by the (% of votes). From a proportional-political-representation perspective, if an election is "fair," then the proportion of seat to votes

should be 1. And if you look at the seats and votes from one party’s perspective, if the proportion is greater than 1 then that party was advantaged. As an example, in Massachusetts the representation proportion in 2018 was $(100\% \text{ seats}) / (66\% \text{ votes}) = 1.52$. (In the appendix, I introduce another measure of political fairness called the “efficiency gap.” As I continue to work on this paper, I will look at how the “efficiency gap” compares with the PPR outcome that I’m currently using.)

In my analysis, I am looking for differences in redistricting between those states that were controlled by Republicans, by Democrats, and by Others.

Who draws the districts can and often does have an impact on the political fairness of the districts. There is no need for statistical analysis to convince you of this statement, nor do I need to argue the fact that gerrymandering is an aspect of politics. But there are other factors aside from who drew the lines, that can impact the “fairness” of the resulting districts.

Gerrymandering writ-large, is accomplished by either “packing” (drawing district lines to pack one group into a small number of districts, which gives the other group an advantage in the remaining districts), or by “cracking” (drawing district lines to dilute one group’s votes by distributing those voters among numerous districts).

Though we know that those who are drawing political boundaries can pack a district, voters also pack districts when they choose where they will live. And in a political environment where Democratic correlates with urban and Republican correlates with rural, this natural clumping can have an impact on proportional-political-representation. In the appendix I discuss some measures of clumping that I analyzed, and others that I will look at further. For this version of my analysis though, I have not included clumping as a predictor in my model fitting, and fully realize that any model that doesn’t include clumping is incomplete. I will include clumping as I continue to refine my model.

The other predictor variable that I did include in my model fitting, aside from who did the redistricting, is the number of districts in the state. Clearly if there are only 2 congressional seats, proportional-political-representation is unlikely. And at the other end of the spectrum as an example are the 61 seats the Democratically controlled legislature in California had to work with when they drew district boundaries in 2001 – a number that gives numerous opportunities for cracking and packing.

METHODS

In building the data frames for my analysis I’ve worked from: 1) election results from the Harvard Dataverse website, 2) An analysis done by the Brennan Center for Justice of who was responsible for redistricting in each state, and 3) population and land area data for 2000 and 2010, downloaded from Census.gov.

Aside from some data cleaning and data frame joining, the major data wrangling issue that I dealt with was imputing vote numbers for candidates who ran unopposed. The vote tallies for unopposed candidates are not reflective of the number of Democratic and Republican voters in a district. In it’s extreme, there are Florida’s vote records where the winning candidate is noted as having 1 vote. Since my outcome measure was comparing % of seats (an accurate number) I needed a method get a more accurate total of votes within a state.

For each unopposed election, I calculated the maximum number of votes received by candidates of that party in the other districts in that state and then assigned that maximum vote total to the unopposed candidate. And for the party that didn’t field a candidate, I created a candidate named “Imputed” and assigned them the minimum vote total from the other candidates of that party. (When I do my analysis using the “efficiency gap” and wasted votes, I will revisit this method of imputing uncontested elections.)

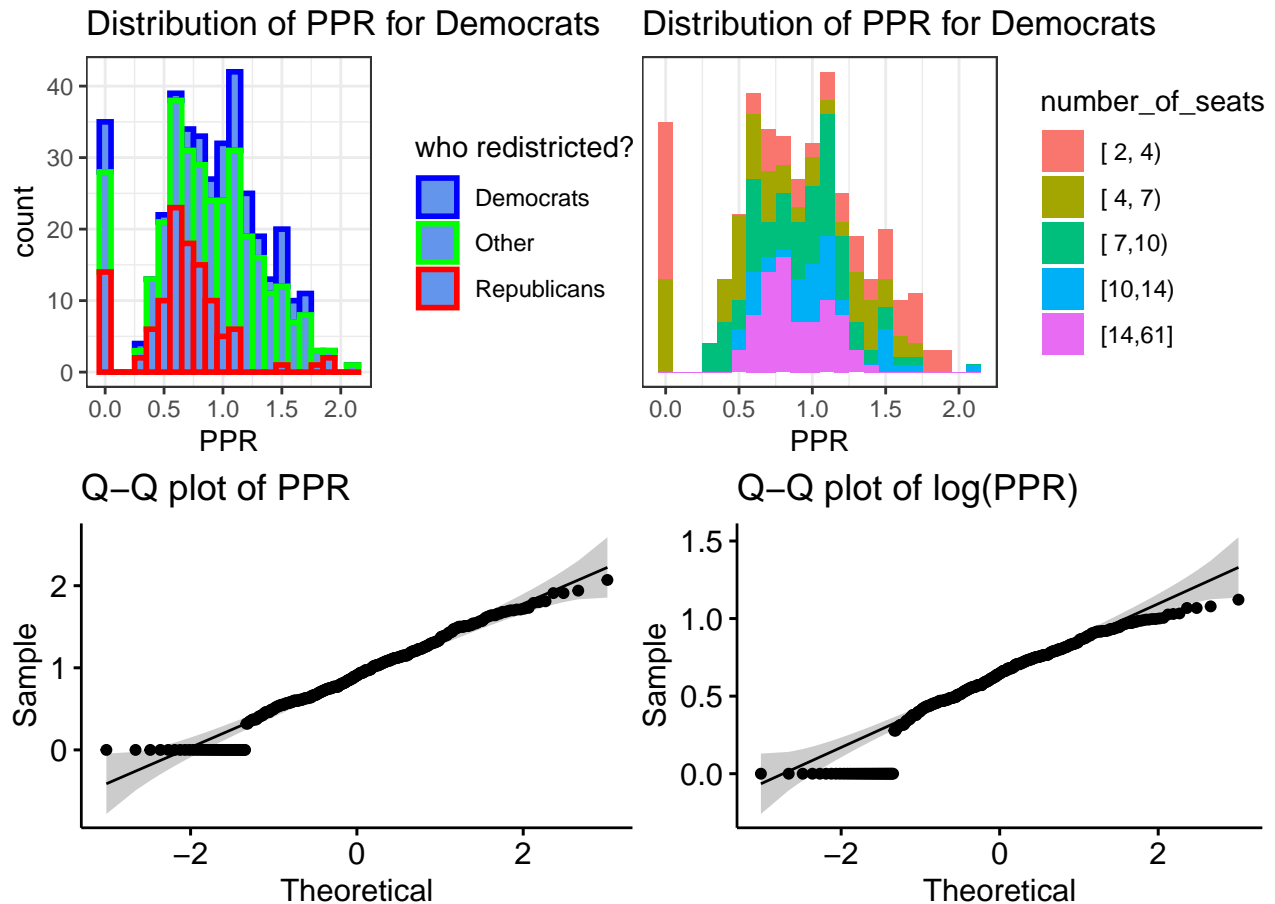
Finally, for each county within each state I calculated the population density for each year. Those values were used in my attempt to build a “clumping” metric (see the appendix).

When done, I had a data.frame for both Democratic results and Republican results. And I did my data analysis from the Democra’s perspective and using the Democratic Results data frame.

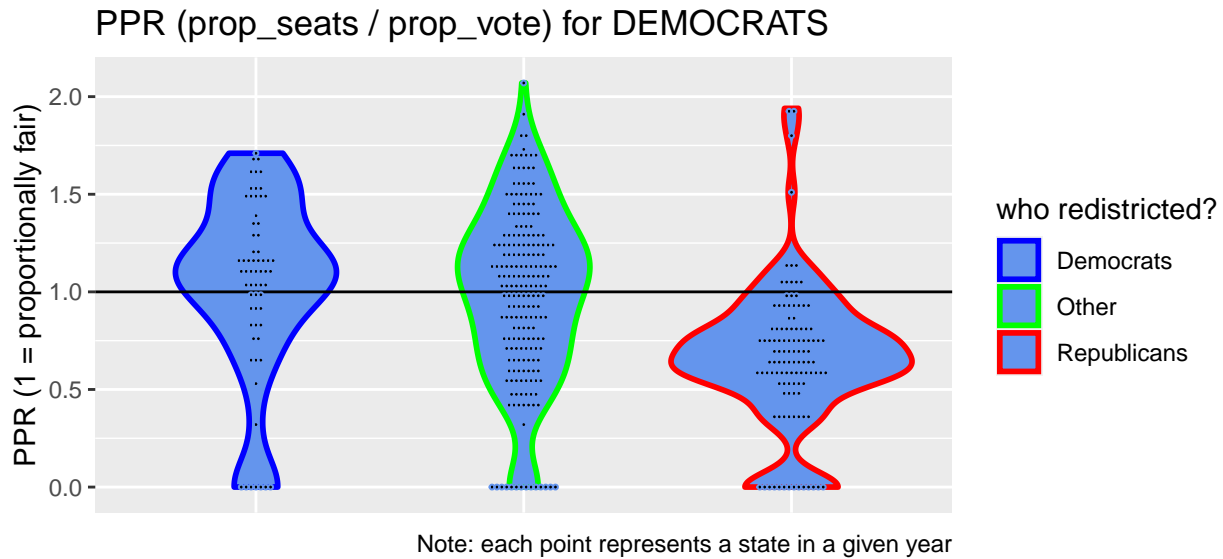
year	state	party	seats	who	PPR
2002	AL	democrat	7	D	0.63
2002	AR	democrat	4	O	1.31
2002	AZ	democrat	8	O	0.61
2002	CA	democrat	53	D	1.16
2002	CO	democrat	7	O	0.85

EDA

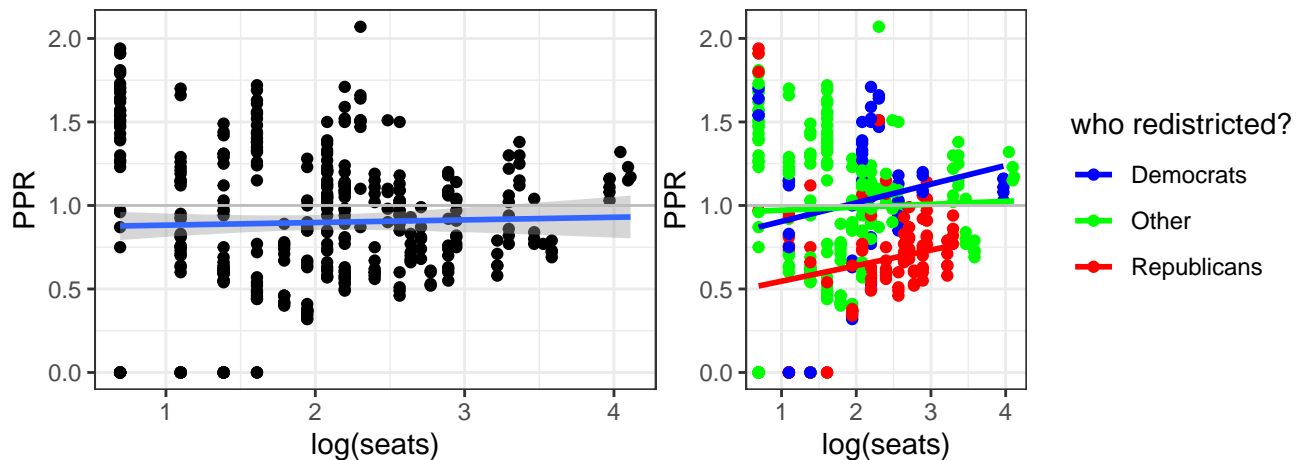
What's the distribution of PPR look like?: Visually the distribution of PPR looks normal, which the exception of the 35 elections where the Democrats won zero of the congressional seats in that state (regardless of their vote percentage). And when subset the PPR's by quantile, it's clear that with more congressional seats, the results are less varied and move toward a mean PPR of 1 (a politically proportional result). Finally a Q-Qplot of PPR (and $\log(\text{PPR})$) and a Shapiro-Wilk normality test ($p=.00001$) show that PPR is normally distributed (noting the exception of the 35 PPR results). Given this, I am assuming normality of my outcome variable for my regression fitting, and am using PPR instead of $\log(\text{PPR})$.



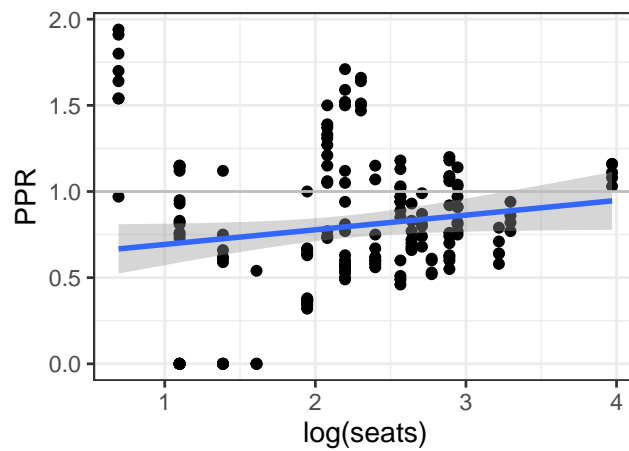
Who vs. PPR (seat/vote)?: Just looking at the comparison between the three groups who redistricted, it's visually clear that the democrats were advantaged when Democratic controlled states redistricted, and Democrats were disadvantaged when Republican controlled states redistricted. But we already knew this was true.



Size of district vs. political proportionality (seat/vote): In looking at the relationship between the size of the district and PPR, there is no visual pattern in the aggregate. And the grouped pattern is as expected when the redistricting was not done by either Democrats or Republicans: the size of the seat doesn't appear to influence the "fairness." But when looking at Democratly controlled redistricting it appears that more seats there are to redistrict the more likely the results advantage the Democrats. Whereas with Republican controlled redistricting, the opposite appears true.

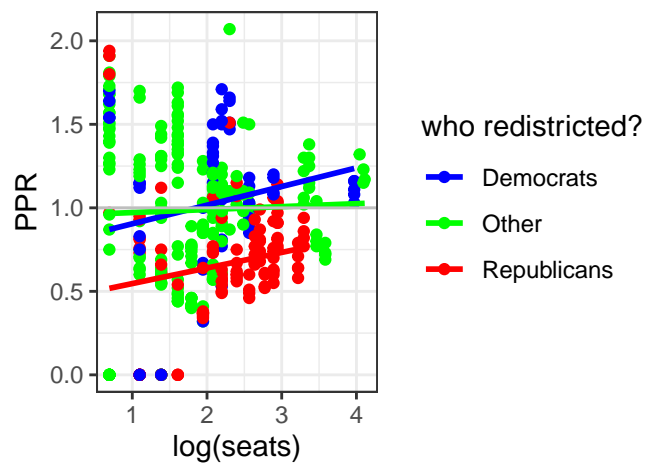


This is what it looks like with just the politically redistricted states over the years 2002 to 2018. I've removed all states where some other entity aside from Democratic or Republican controlled legislatures, did the redistricting.



```
##
## Call:
## lm(formula = PPR ~ log(seats), data = DemYearlyResults[DemYearlyResults$who !=
##      "D", ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74476 -0.24166 -0.04831  0.22869  1.27317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.60787    0.10056   6.045 8.8e-09 ***
## log(seats)   0.08506    0.04367   1.948  0.0531 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4392 on 175 degrees of freedom
## Multiple R-squared:  0.02122,    Adjusted R-squared:  0.01562
## F-statistic: 3.793 on 1 and 175 DF,  p-value: 0.05306
```

Model fitting



RESULTS

DISCUSSION

BIBLIOGRAPHY and APPENDIX

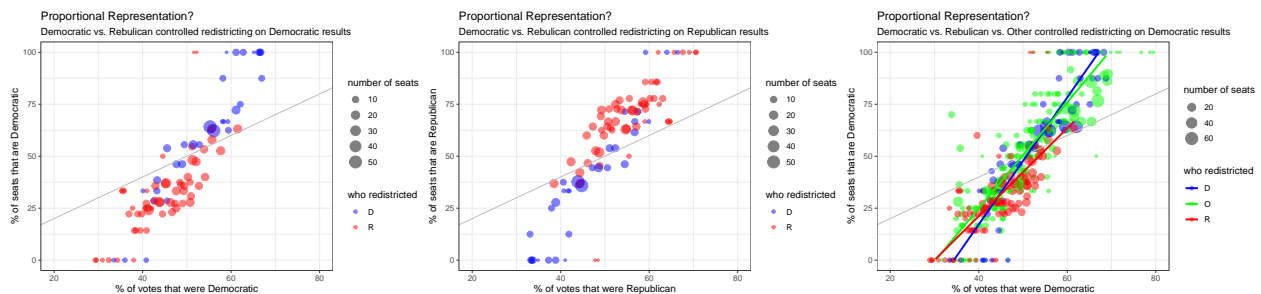
Metrics

(I) Who to include?

In looking at vote results I had to decide which parties to include. As such combined Minnesota's "democratic-farmer-labor" party with the democrat party. I looked at the entirety of listed parties, to see what parties aside from the democrats and the republicans were significant either in the number of candidates or the votes. The only party that was worthy of consideration aside from the two major parties was the Libertarian party. The Libertarians ran candidates in 2,018 elections between 1990 and 2018. But never got higher than 31% of the vote and their mean vote percentage was 3.4% (with a distribution of results that was heavily skewed right with "skewness" = 3.4). As such I have not included the "libertarian" party as a major party, and have confined my analysis to Democrats and Republicans.

(II) Is it politically fair?

(IIa) vote_percentage vs seat_percentage Originally, I looked at representing "politically fair" where the outcome variable was the proportion of house seats won and the predictor was the proportion of votes won. If an election was fair then these two proportions would be equal. State-years where the vote_proportion and the seat_proportion fell on the $y=x$ line (grey line in the graphs), were proportionally equal/fair. Additionally I used color to note who did the redistricting and how many seats were decided (with the notion that for a state with a small number of seats, it would not be surprising to have a big difference in vote and seat proportions, e.g. NH-2016 democrats had 46% of the vote but 0% of the two congressional seats). Ultimately I decided that this representation was harder to visualize. And the clumping variable was not a predictor variable for the proportion of seats won, but rather a possible predictor variable for political proportionality. As such I decided to center my analysis around a measure of political fairness (seat_proportion / vote_proportion).

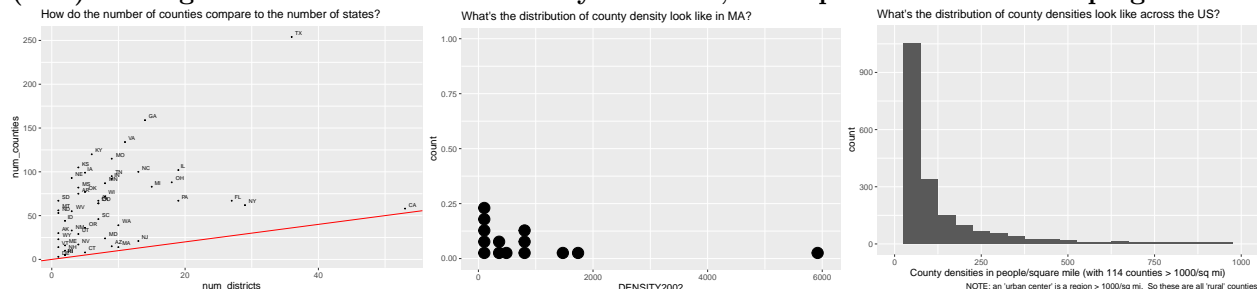


(IIb) The efficiency gap & wasted votes When "cracking" or "packing," voters are moved into or out of districts to dilute the power of their vote. The efficiency gap is a mathematical calculation that "counts the number of votes each party wastes in an election to determine whether either party enjoyed a systematic advantage in turning votes into seats." ("How the Efficiency Gap Works," Petry, 2015, http://bettergov.nc.lwvnet.org/files/how_the_efficiency_gap_standard_works.pdf). I have not yet taken the time to write the code to calculate the efficiency gap and then explore how illustrative it is as a measure of political fairness.

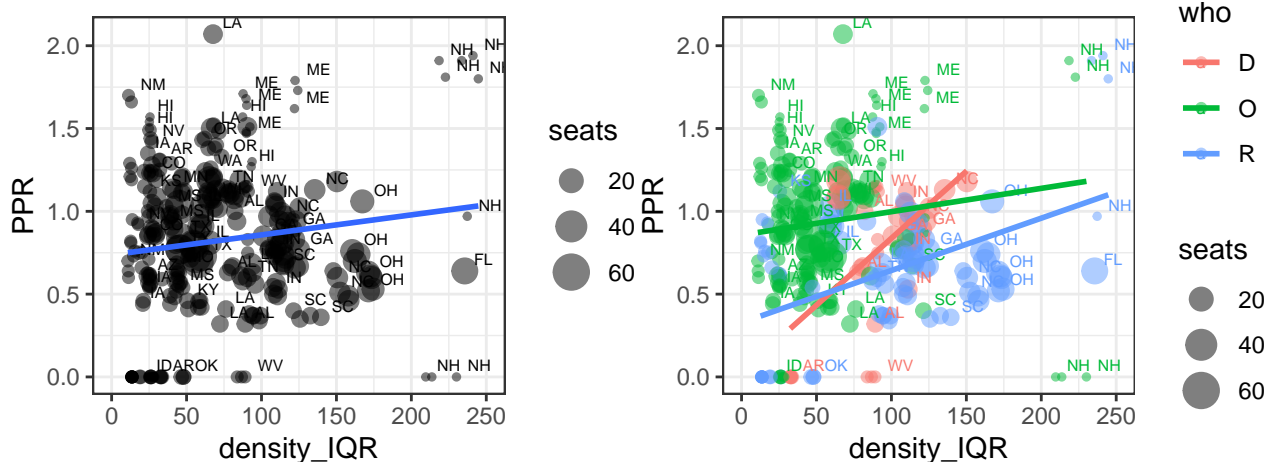
(III) How to measure clumping.

I'm still trying to wrap my head around how to appropriately measure "clumping" so that the measurement will have meaning in the context of packing and cracking.

(IIIa) looking at the distribution of county densities, to help structure a "clumping" metric

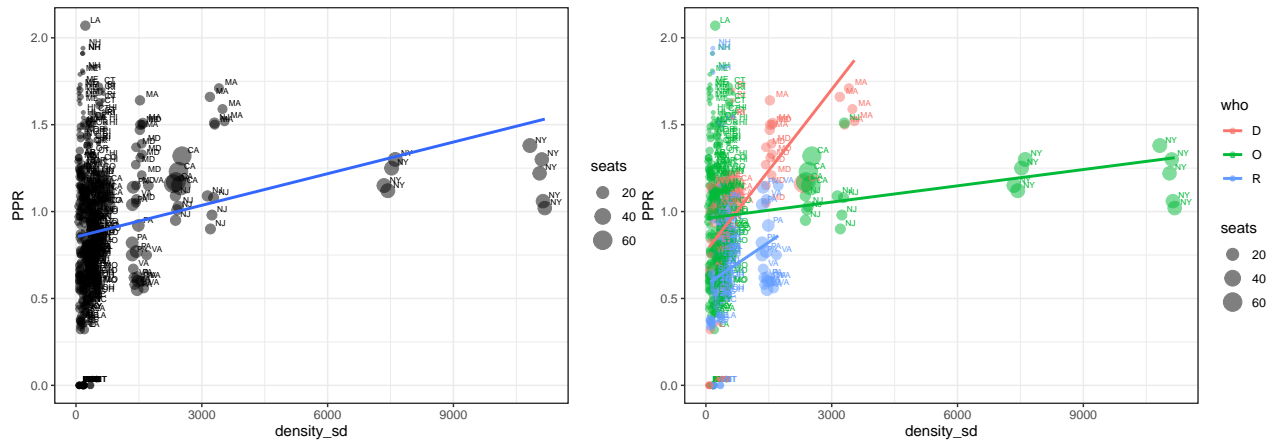


(IIIa) IQR of county level population density For each state, for each year, I've calculated the spread of county population densities



```
##
## Call:
## lm(formula = PPR ~ density_IQR, data = DemYearlyResults)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89770 -0.27926  0.01272  0.30302  1.22094
##
## Coefficients:
##      (Intercept)  8.288e-01  2.567e-02  32.289  < 2e-16 ***
##      density_IQR  2.993e-04  5.608e-05   5.338  1.61e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4368 on 384 degrees of freedom
## Multiple R-squared:  0.06908,    Adjusted R-squared:  0.06665
## F-statistic: 28.49 on 1 and 384 DF,  p-value: 1.612e-07
```

(IIIb) standard deviation of county level population density For each state, for each year, I've calculated the spread of county population densities



(IIIc) comparing state level population density, with county level population density, with Census defined “Urban centers” This will be my next clumping variable attempt.

(IV) who is making the redistricting decisions

These are the sources I used to create the group variable “who” for each election decade (2002-2010 & 2012-2018). The “who” variable attributes the responsibility for the redistricting to either D, R, or I (Democratically controlled state legislatures, Republican controlled state legislatures, or Other). I’ve defined “Other” as not Democratically controlled nor Republican controlled. “Other” includes court ordered redistricting, split-party redistricting, and non-political commissions.

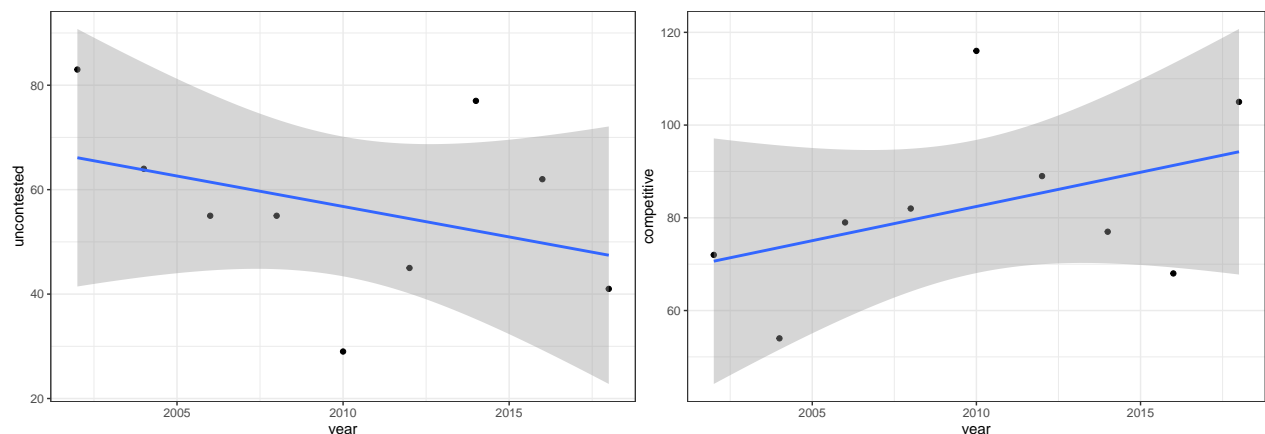
https://www.ncsl.org/documents/statevote/2010_Legis_and_State_post.pdf

<https://www.brennancenter.org/our-work/research-reports/redistricting-and-congressional-control-first-look>

https://www.brennancenter.org/sites/default/files/2019-08/Report_CGR-2010-edition.pdf

(V) looking for a pattern in the number of uncontested and competitive congressional races by year

Though there is a great deal that’s been written about how political gerrymandering is increasing the number of uncontested elections, and decreasing the number of competitive elections, my analysis doesn’t bear this out. Though my quick look is looking at all elections and is not looking at the grouping variable of who is doing the redistricting.




```
##
## Call:
## lm(formula = competitive ~ year, data = YearlySummary)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.294 -11.344   2.456   3.606  33.556
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2882.306   2364.110   -1.219   0.262
## year         1.475     1.176     1.254   0.250
##
## Residual standard error: 18.22 on 7 degrees of freedom
## Multiple R-squared:  0.1835, Adjusted R-squared:  0.0668
## F-statistic: 1.573 on 1 and 7 DF,  p-value: 0.2501

##
## Call:
## lm(formula = uncontested ~ year, data = YearlySummary)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.778  -6.444  -4.111  12.222  24.889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2401.778   2201.245    1.091   0.311
## year        -1.167     1.095   -1.065   0.322
##
## Residual standard error: 16.97 on 7 degrees of freedom
## Multiple R-squared:  0.1395, Adjusted R-squared:  0.01658
## F-statistic: 1.135 on 1 and 7 DF,  p-value: 0.3221
```

Citations

(need to figure out how to use LaTeX or ?? to display this citation) @incollection{DVN/IG0UN2/ELBYL3_2017, author = {MIT Election Data and Science Lab}, publisher = {Harvard Dataverse}, title = {1976-2018-house2.tab}, booktitle = {U.S. House 1976–2018}, UNF = {UNF:6:8iuXTceVO5a7EpOwUD5UPw==}, year = {2017}, version = {V7}, doi = {10.7910/DVN/IG0UN2/ELBYL3}, url = {https://doi.org/10.7910/DVN/IG0UN2/ELBYL3} }

<https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html#POP>

to get .xls for population and .xls for area:

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjHkPmp4bftAhV5FlkFHbDHBvurl=https%3A%2F%2Fwww2.census.gov%2Flibrary%2Fpublications%2F2011%2Fcompendia%2Fusa-counties%2Fexcel%2FMastdata.xls&usg=AOvVaw1ahgb3GWupqOb1UHYbZIMw>

Copyright Copyright 2020 Bruce C. Mallory

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER

CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.