Capstone Project 1: Milestone Report

Problem statement

Why is your capstone question import to answer and for whom?

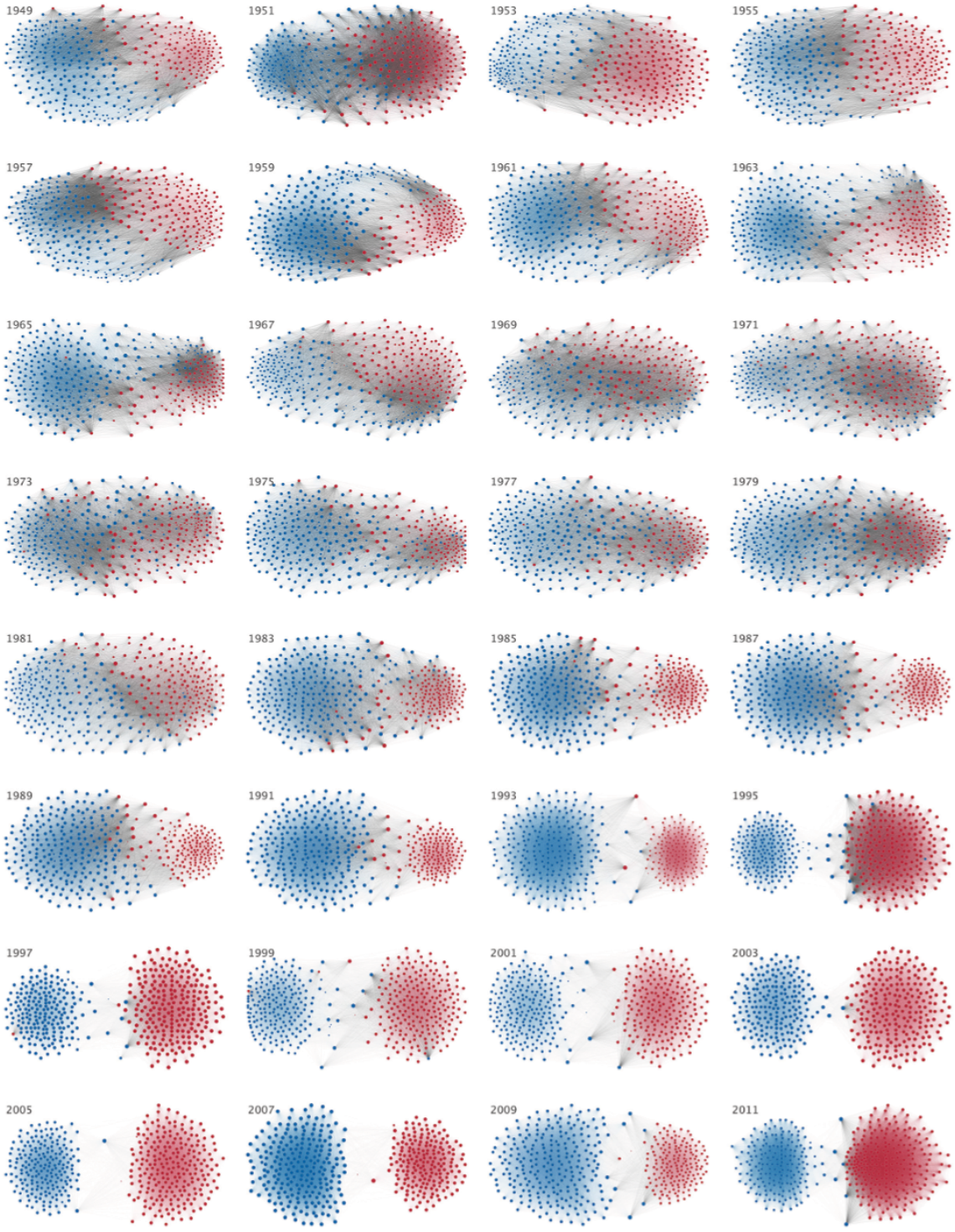
The question I am trying to answer with this project is whether I can predict the electoral outcome (i.e. Democrat or Republican) of a district, given that state's demographic and income data.

This is an important question to answer because over the past 50 years demographic and income trends have changed the United States drastically – from a largely ethnically homogenous and definitively middle-class nation, to a more diverse and more economically stratified society.

This has driven commensurate political change, as American politics has become more polarized (see visual on the next page[1]). These changes have upended the traditional political coalitions and has made political predictions more circumspect.

[1] Andris C, Lee D, Hamilton MJ, Martino M, Gunning CE, Selden JA (2015) The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives. PLoS ONE 10(4): e0123507. https://doi.org/10.1371/journal.pone.0123507

Division of Democrat and Republican Party members over time[2]



[2] Each member of the U.S. House of Representatives from 1949–2012 is drawn as a single node. Republican (R) representatives are in red and Democrat (D) representatives are in blue, party affiliation changes are not reflected.

My model strives to formulate a bottoms-up approach to the prediction problem – by analyzing the underlying demographic and income changes in order to make political predictions, specifically, which party is likely to carry a given district.

This question is of particular importance to political campaigns themselves. Each political party has a fixed amount of capital (i.e. monetary resources, volunteers, local party infrastructure, ect.) and so must make strategic decisions about how to allocate that capital. My model strives to determine, before a single vote is cast, the likely outcome of a given district race – entirely predicated on that district's demographic and income fundamentals.

Using this data, parties can make predictions about how likely a given district will vote Democrat or Republican and make strategic decisions about party resource allocation – avoiding wasting resources on districts that are more likely to vote for the other party or their own, allocating resources instead toward swing districts.

Data Description

What kind of cleaning steps did you perform?

The first step in cleaning the data was removing unnecessary columns from the dataset. This will help to prevent overcomplication and potential inaccuracies and will also improve the overall processing speed of the model.

I removed the following columns from the dataset:

- **abroadPrcnt** – percent of the district that lived abroad in the census used for this term. The field was removed because Censuses are updated yearly after the 109th Congress (2005-06). Prior to the 109th Congress, censuses were generally only updated once per decade, meaning that there are a significant number of missing values.
- **prcntExAliens** – percent of the foreign-born population that became a citizen. This is intended to capture a measure of assimilation. Unfortunately, this field also has a significant number of missing values.
- **totalHouseholds** – total number of households in the district at the time of the most recent census. This feature does not add anything to the dataset that totalPopBirthPlace doesn't already add.
- **meanIncome** – mean income of a district's households. This would be a great feature to include, however, there are a significant number of missing values, plus any significance contained in the income distribution is captured in the detailed income bracket information included in the model (i.e. variables **under10k**, **over10k**, **over15k**, etc.)

- **medianIncome** – median income of a district's households. This would be another income feature that would great to include. Unfortunately, there a significant number of missing values. Additionally, no inflation adjustments were used on the data. The values were estimated for the 1980 census using the percent of the population in each income bracket. Thus, for the 1980s there is a heavy rounding error involved in the estimation.
- **totalEmploymentPop** - total population used for the employment variables. Only includes those over the age of around 16, depending on the cutoff for that census. The more relevant feature here is **prcntUnemp**, which tracks the percent of the district's population that is unemployed but still in the labor force.
- **prcentNotEmploy** – percent of the district's population that is not in the labor force. This is a less meaningful variables as it excludes those who are unemployed but in the labor force – including young children, stay-at-home parents, the permanently disabled, the retired, etc.
- **totalPopRaceFile** – total population for the race variables. The numbers used here are irrelevant to determining the likelihood of a district voting Democrat or Republican. However, the race variables themselves are.
- **prcntBlackNotHisp** – percent of the district that is black but not Hispanic. Unfortunately, this variable has a significant number of missing values.
- **prcntMulti** – Percent of the district that is multiracial. Unfortunately, this variable has a significant number of missing values.
- **prcntWhite** – Percent of the district that is white but not hispanic. Unfortunately, this variable has a significant number of missing values.
- **prcntNotHisp** – Percent of the district that is not hispanic. Unfortunately, this variable has a significant number of missing values.
- **prcntOld** – Percent of the district that is over 62 or 60, depending on the age buckets used by the census for that decade. Unfortunately, this variable has a significant number of missing values.
- **medianAge** – median age in the district. Unfortunately, this variable has a significant number of missing values.
- **sponID** – govtrack id number for each member of Congress. All entries have a govtrack ID number.
- **Icpsr** – icpsr number for each member of Congress. This variable is irrelevant to our analysis, plus not all entries have an ICPSR number. Additionally, there are some discrepancies in how sources use ICPSR numbers, as some sources use 9's as the first digit to indicate party switches or other irregularities
- **state** – state name, two letter abbreviation (ie AK, CA). This is a repeat of information from **stateDist**.
- **district** – congressional district number. This is a repeat of **stateDist**.

- **lastName** – Some nontraditional characters (hyphens, accents) are garbled in translation. All names are recognizable by the human eye, but id numbers should be used for matching individuals.
- **firstName** – see **lastName**
- **middleName** – see **lastName**
- **age** – age at time of being sworn into Congress for that session. We are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **gender** – gender of MC. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **numberTerms** – number of terms served in the House. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **daysServed** – number of days of the full session served. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **fracServed** – fraction of the congressional session that the MC served in. The vast majority of MCs served in the full session (2 years). Those that didn't filled vacant seats or left mid-term. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **cd** – congressional district number. This is a repeat of **stateDist**.
- **statenm** – state name. Again, a repeat of information from **stateDist**, but this is the state name used by nominate tables, all in caps, and often abbreviated (ie MINNESO)
- **dwnom1** – first dimension nominate score. This variable's significance is unknown, so it is excluded.
- **dwnom2** – second dimension nominate score. This variable's significance is unknown, so it is excluded.
- **comPower** – how powerful were the committees the MC served in. Example provided is of the 109th Congress. Committees vary by congress, so points assigned vary by Congressional session. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **chair** – Binary variable. Takes the value of 1 if the MC was the chair of a committee in that congressional session, 0 if not. No MC chaired more than one committee. Committee information is currently available for the 109th-113th Congresses. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **rank** – Binary variable. Takes the value of 1 if the MC was the ranking member of a committee in that congressional session, 0 if not. No MC was ranking member of more than one committee. Committee information is currently available for the 109th-113th Congresses. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.

- **comName** – names of the committees served on. Committee information is currently available only for the 109th through 113th Congresses. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **numCom** – number of committees served on. Committee information is currently available for the 109th-113th Congresses. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **rankChair** – rankChair = 0.5rank + chair. This creates a single variable for power in a committee, where being a ranking member is half as valuable as being the chair of a committee. Committee information is currently available for the 109th-113th Congresses. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **black** – binary variable for being a member of the Congressional Black Caucus. 1 if a member of CBC, 0 if not. To my knowledge, all self-identifying black members of Congress are members of the CBC. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **Hispanic** – a binary variable for ebing a member of the Congressional Hispanic Caucus. 1 if a member of CHC, 0 if not. To my knowledge, all self-identifying Hispanic members of Congress are members of the CHC. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **numSpon** – number of bills an MC sponsored in a term. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **numCosp** – number of bills an MC cosponsored in a term. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **numPassH** – number of bills an MC sponsored and were approved by a full House vote in a term. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **numEnact** – number of bills an MC sponsored and were signed into law by a full House vote in a term. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **passPrcnt** – percent of bills that an MC sponsored and were enacted into law. Again, we are interested in the fundamental characteristics of the districts themselves, not on who runs for Congress in them.
- **partyControl** – 'D' if Democrats controlled Congress, 'R' if Republicans controlled Congress. We are interested in district specific predictions, not party control of Congress as a whole.
- **demSeats** – number of seats held by Democrats at the beginning of the Congressional session. Again, are interested in district specific predictions, not party control of Congress as a whole.
- **repSeats** – number of seats held by Republicans at the beginning of the Congressional session. Again, we are interested in district specific predictions, not party control of Congress as a whole.

- **otherSeats** – number of seats held by third parties (independents, green party, etc.). Note that this dataset does not provide information about which third party MCs belonged to. Again, we are interested in district specific predictions, not party control of Congress as a whole.
- **ses** – socioeconomic status measure: estimated from the income and education of a district. This was estimated by finding the weight of the common factor between education and income. The method by which this variable was calculated is unknown and so is excluded from the data.
- **sesNorm** – socioeconomic status measure, normalized: estimated from the income and education of a district. It is normalized to range between 0 and 100. This allows the magnitude of this coefficient to be directly compared to the coefficients for income and education, both of which are given in percentages. The method by which this variable was calculated is unknown and so is excluded from the data.

How did you deal with missing values, if any?

The next step in cleaning the data was dealing with missing values. Once we removed the above fields there were very few rows that had any missing values. However, those rows that did have missing data were simply removed from the data. As a result, the number of rows declined from 9,312 to 9,298 – an insignificant difference.

Were there outliers, and how did you handle them?

Due to the nature of the data, there weren't any obvious outliers. However, the last step in the data cleaning process was to apply one-hot encoding to our dependent variable – **party**. The goal of our analysis is to predict, given the fundamental features of the district in question, whether that district will vote Democrat or Republican.

Therefore, our dependent variable should either be Democrat or Republican – a binary choice – which can be represented by either a 0 (for Republican) or 1 (for Democrat). We can do this easily by applying one-hot encoding, which converts text-based features into numerical, binary values, represented as "1" or "0" – "True" or "False". This will be helpful going forward as many algorithms and also scatterplots are not compatible with non-numerical data.

When I applied one-hot encoding to the party's variable, it ended up splitting into four additional features – **party_Democrat, party_Republican, party_Independent**, and **party_Republican-Conservative**. The variable **party_Indpendent** was such a small value that rows containing this value could be easily removed.

Rows containing **party_Republican** and **party_Republican-Conservative** could easily be combined, until all that was left was **party_Democrat** and **party_Republican**.

Finally, since **party_Democrat** contained information – in binary – as to the outcome of the election in each district, we could remove **party_Republican**, leaving one, final, dependent variable – **party_Democrat** – whereby a 0 signified a Republican win and a 1 signified a Democratic win.

Initial Findings

Can you count something interesting? Can you find trends (e.g. high, low, increasing, decreasing, anomalies)? Can you make a bar plot or a histogram? Can you compare two related quantities? Can you make a scatterplot? Can you make a time-series plot?

The data is organized by Congress (specifically, Congress 93 through 113) so it made sense to visualize the data as a time-series plot, with Congressional Number as the x-axis and the various independent variables as the y-axis, in order to see the differences between Democratic and Republican district figures, plotted over time (see results on page 9).

Those variables that have a persistent difference between Democratic and Republican figures, over time, are likely to be the most significant variables in explaining the variation in party outcomes.

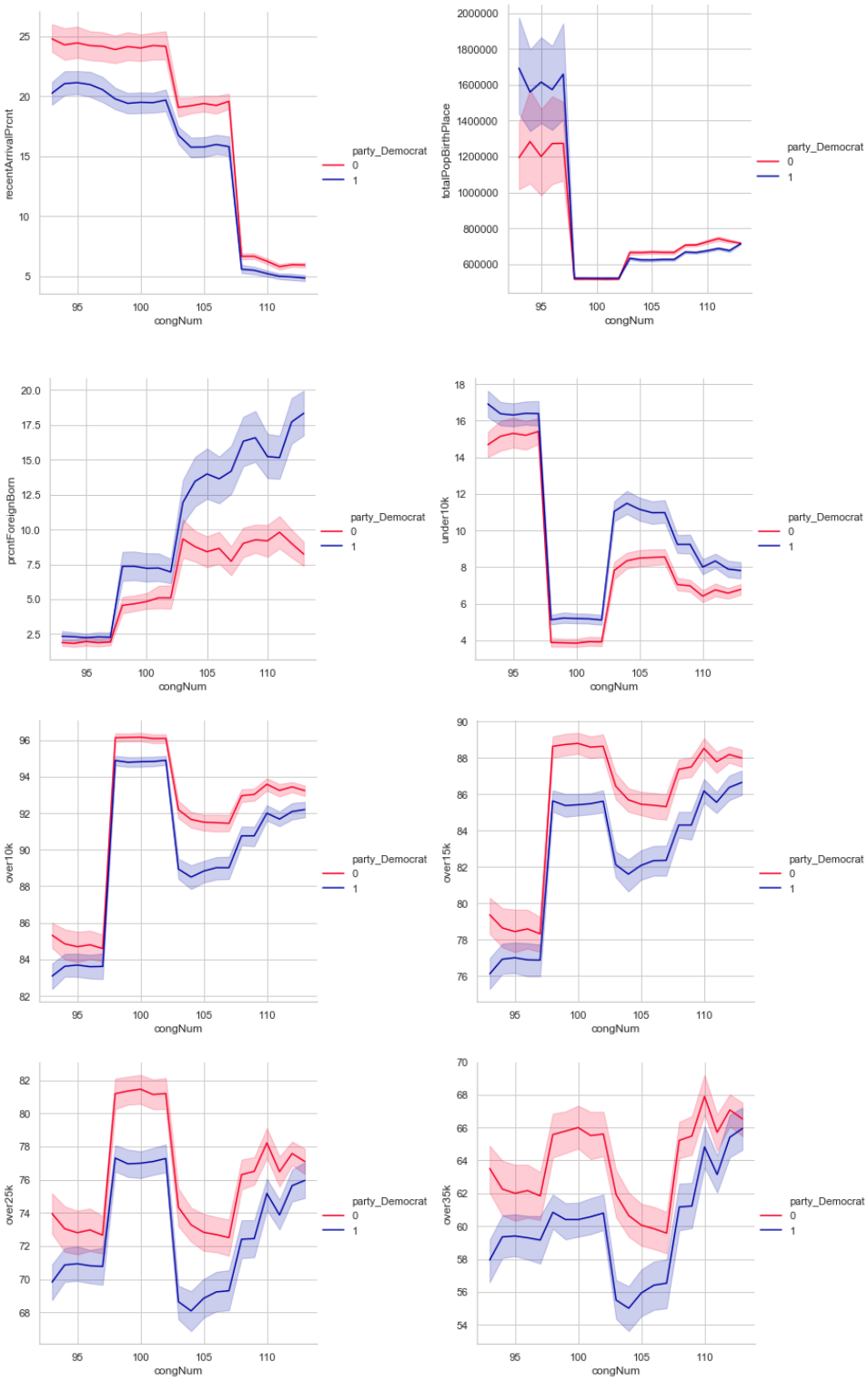
Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you'd like to investigate further? What other questions do the insights lead you to ask?

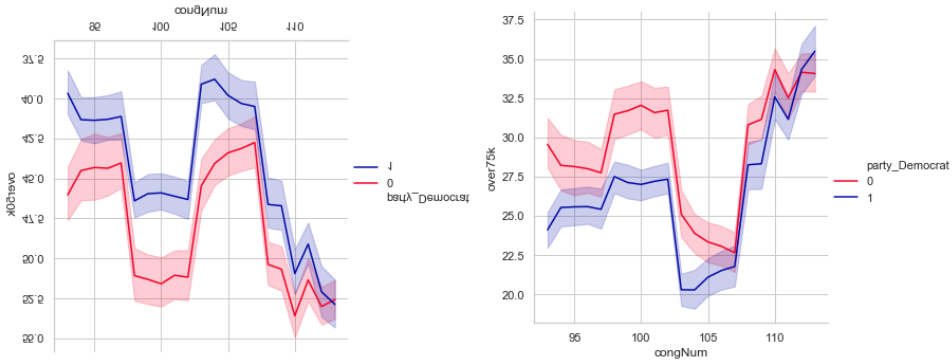
In general, it appears that there are large, national changes in certain variables that affect Democratic and Republican districts in the same way. For example, according to the graph plotting **recentArrivalPrcnt** there has been a distinct reduction in the mobility of Americans.

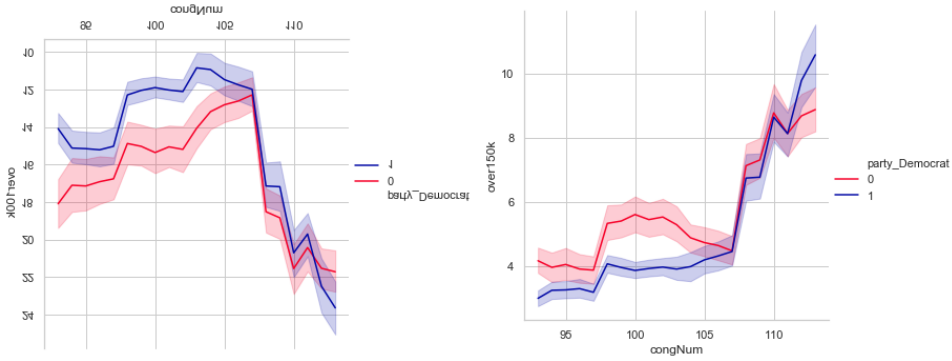
The **recentArrivalPrcnt** variable tracks the percent of the district that recently moved into the district from another county or state. It definitely appears that between the 100th and 110th Congresses, there was a distinct reduction in the percentage of the district populations that recently moved from another county or state, suggesting that fewer people are moving and that people are more locked into their districts than they were historically.

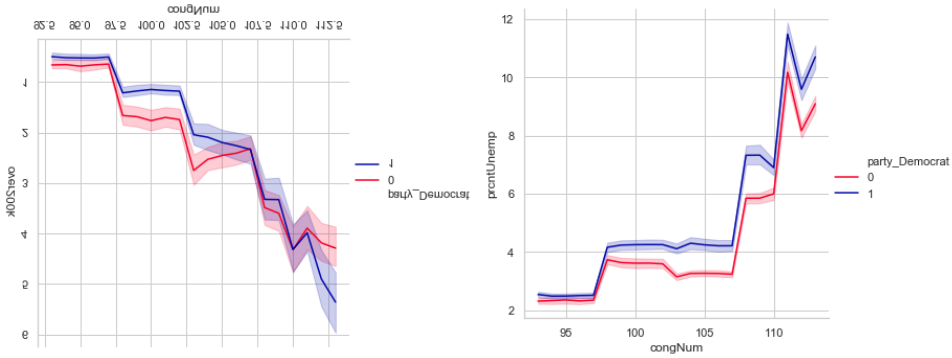
In addition, it appears that the percent of the districts that was born in a foreign country has also increased in both Democratic and Republican districts - though this trend is significantly more pronounced in Democratic districts.

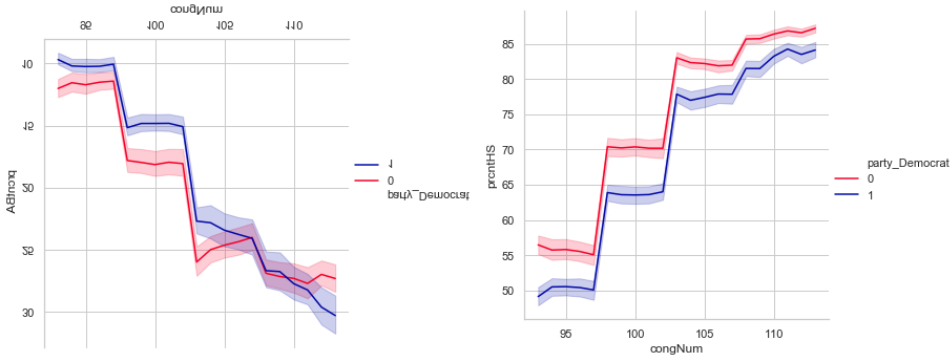
Additionally, unemployment rates have increased in both districts while High School and College graduation rates have increased in both as well.

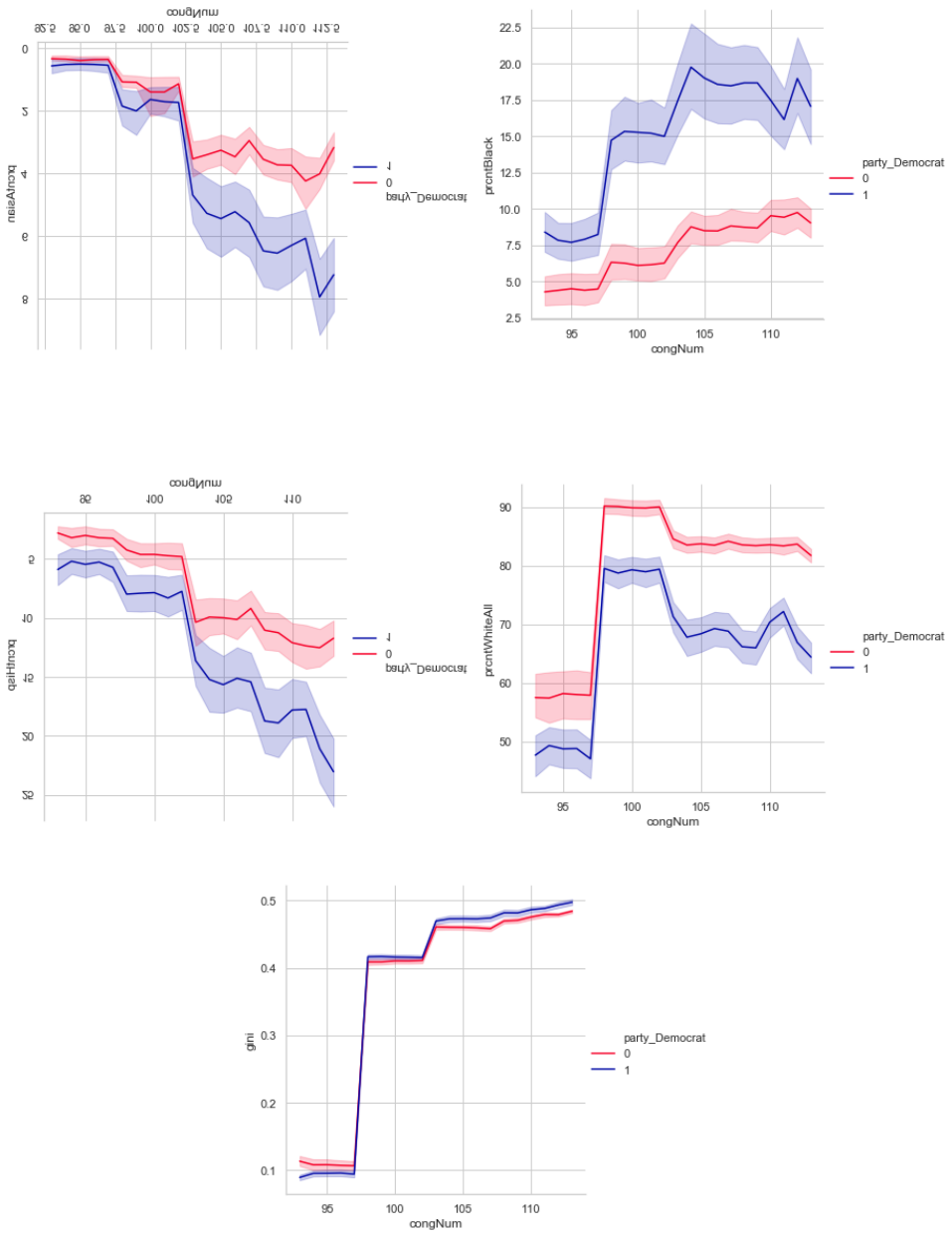

recentArrivalPrcnt
totalPopBirthPlace
prcntForeignBorn
under10k
over10k
over15k
over25k
over35k
congNum
party_Democrat
0
1


over75k
congNum
party_Democrat
0
1


over150k
congNum
party_Democrat
0
1


prcntUnemp
congNum
party_Democrat
0
1


prcntHS
congNum
party_Democrat
0
1



Finally, it appears that inequality has increased (as measured by the gini coefficient - a smaller gini value signifying a more equal district) in both Democratic and Republican districts.

Now that you've asked questions, hopefully you've found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are

some other trends/relationships you think will make the story more complete?

Of course, we are more interested in those things that differentiate Democratic and Republican districts. Though many of these national trends are represented in both types of district, we also see many of these trends are being more pronounced in one type of district versus another.

For example, though mobility has declined is has declined less in Republican districts; and though the percent of the district that was born in a foreign country has also increased in both Democratic and Republican districts, this trend is dramatically more pronounced in Democratic districts, and has actually declined in recent years amongst Republican districts.

There also appears to be a distinct gap between Republican and Democratic districts when it comes to the percent of households earning more than 35,000 dollars. There is a large and persistent inequality, with a greater percentage of Republican district households earning more than 35,000 dollars.

Though graduation rates have increased for both High School and College in both Democratic and Republican districts Republican districts have a consistently higher rate of High School graduation, while Democratic districts have only recently outpaced Republican districts in terms of College graduation - a reversal of a long-established trend.

The greatest divergences appear when we compare Democratic and Republican districts by racial composition. Democratic districts have consistently higher concentrations of Asian, Black, and Hispanic Americans than their Republican counterparts, and both Democratic and Republican districts are becoming less white overall.

In sum, the trends point toward a more static, highly educated, and diverse nation - but with these trends being far more pronounced in Democratic districts than in Republican ones. Therefore, it is my contention that as we look to predict a district's party affiliation, we will see that more static, highly educated, and diverse districts will vote Democratic, while more mobile, less educated, and more homogenous districts will vote Republican.

Are there variables that are particularly significant in terms of explaining the answer to your project question?

Of the original 21 variables tested, only 16 variables were determined to be statistically significant (determined by a p-value less than 0.05):

- **recentArrivalPrcnt**

- **totalPopBirthPlace**
- **prcntForeignBorn**
- **under10k**
- **over10k**
- **over15k**
- **over35k**
- **over50k**
- **over100k**
- **prcntBA**
- **prcntHS**
- **prcntAsian**
- **prcntBlack**
- **prcntHisp**
- **prcntWhiteAll**
- **gini**

Of these final 16 significant variables those of particular significance only **totalPopBirthPlace**, **prcntAsian**, and **prcntBlack** were even perceptibly close to the threshold of 0.05.

Are there strong correlations between pairs of independent variables or between an independent and a dependent variable?

When thinking about correlations between the independent and dependent variable we have to keep in mind that correlation is simply normalized covariation, and covariation measures how two random variables co-variate, that is, how change in one variable is related to change in another one.

Strictly speaking, correlation (like Pearson correlation, for example) cannot deal with categorical variables (mostly because categorical variables don't have a notion of mean, which Pearson is based on).

In terms of the correlations amongst the various independent variables we do see some strong correlations. On the previous page there is Pearson Correlation Cross-Table (see page 14), analyzing the correlations amongst all of the independent variables (here is a link to the original excel file, which is much easier to read).

Firstly, there is a high correlation between **prcntForeignBorn** and **prcntHisp**, which is not surprising, given that much of the U.S. Hispanic population are relatively recent arrivals.

Additionally, there is a great deal of correlation amongst the various income variables (i.e. under10k, over10k, over15k, etc.). It makes intuitive sense that, for example, under10k and over10K would be highly correlated (nearly perfectly inversely correlated in fact).

What are the most appropriate tests to use to analyze these relationships?

In terms of determining significance, the Logit function is the appropriate test to use. The Logit function goes hand in hand with the regular Logistic and produces the outputs on pages 14 and 16.

The p-values on both tables are outlined in red. The initial logit results suggested that over75k, over150k, over200k, and **prcntUnemp** were all statistically insignificant in explaining the variation in the dependent variable (i.e. **party_Democrat**).

After removing those variables from the data, we arrived at the second table containing the newly adjusted logit results. Fortunately, after removing the four insignificant variables outlined above, the rest of the variables remained significant.

Initial Logit Results (i.e. before removal of insignificant variables)

Results: Logit

Model:	Logit	Pseudo R-squared:	inf
Dependent Variable:	party_Democrat	AIC:	inf
Date:	2019-02-01 16:01	BIC:	inf
No. Observations:	9298	Log-Likelihood:	-inf
Df Model:	20	LL-Null:	0.0000
Df Residuals:	9277	LLR p-value:	1.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
recentArrivalPrcnt	-0.0370	0.0041	-8.9709	0.0000	-0.0450	-0.0289
totalPopBirthPlace	-0.0000	0.0000	-2.2300	0.0257	-0.0000	-0.0000
prcntForeignBorn	0.0624	0.0069	9.0382	0.0000	0.0489	0.0760
under10k	0.0381	0.0133	2.8674	0.0041	0.0121	0.0642
over10k	0.3353	0.0303	11.0500	0.0000	0.2758	0.3948
over15k	-0.2029	0.0441	-4.5979	0.0000	-0.2894	-0.1164
over25k	-0.1208	0.0396	-3.0490	0.0023	-0.1984	-0.0431
over35k	-0.1400	0.0475	-2.9453	0.0032	-0.2332	-0.0468
over50k	0.2718	0.0514	5.2902	0.0000	0.1711	0.3725
over75k	-0.0174	0.0439	-0.3952	0.6927	-0.1034	0.0687
over100k	-0.1259	0.0315	-3.9896	0.0001	-0.1877	-0.0640
over150k	-0.0227	0.0545	-0.4167	0.6769	-0.1296	0.0842
over200k	-0.0719	0.0567	-1.2683	0.2047	-0.1829	0.0392
prcntUnemp	0.0262	0.0137	1.9088	0.0563	-0.0007	0.0531
prcntBA	0.0816	0.0079	10.3425	0.0000	0.0661	0.0971
prcntHS	-0.0519	0.0056	-9.2018	0.0000	-0.0629	-0.0408
prcntAsian	0.0233	0.0081	2.8748	0.0040	0.0074	0.0391
prcntBlack	0.0166	0.0039	4.2046	0.0000	0.0088	0.0243
prcntHisp	-0.0152	0.0036	-4.2844	0.0000	-0.0222	-0.0083
prcntWhiteAll	-0.0254	0.0025	-10.0288	0.0000	-0.0303	-0.0204
gini	-4.5230	0.9051	-4.9974	0.0000	-6.2970	-2.7491

Pearson Correlation Cross-Table

	recentArrivalPrcnt	totalPopBirthPlace	prcntForeignBorn	under10k	over10k	over15k	over35k	over50k	over100k	prcntBA	prcntHS	prcntAsian	prcntBlack	prcntHisp	prcntWhiteAll	gini
recentArrivalPrcnt		0.123292934	-0.377221594	0.125285758	-0.125004124	-0.149867288	-0.070379322	-0.120455333	-0.17575878	-0.282593374	-0.317287594	-0.25093511	-0.213440854	-0.281981174	-0.062840384	-0.429403743
totalPopBirthPlace	0.123292934		-0.128723106	0.221588027	-0.221433775	-0.183049915	-0.047278259	-0.080564521	0.016933484	-0.145474206	-0.208267705	-0.082363107	-0.142302642	0.177434412	-0.531163605	-0.343076662
prcntForeignBorn	-0.377221594	-0.128723106		-0.207529043	0.206979829	0.238151988	0.176150914	0.236725254	0.316565864	0.443792338	0.249547585	0.62235129	0.080036106	0.722642598	-0.115930489	0.429804931
under10k	0.125285758	0.221588027	-0.207529043		-0.999943005	-0.941068958	-0.681806236	-0.710599533	-0.43462796	-0.471089071	-0.59030564	-0.215172501	0.231523629	-0.021948443	-0.469205214	-0.605156018
over10k	-0.125004124	-0.221433775	0.206979829	-0.999943005		0.941145808	0.682069724	0.710780364	0.434734239	0.470921932	0.590180817	0.214991568	-0.231713168	0.021512314	0.469243791	0.60463861
over15k	-0.149867288	-0.183049915	0.238151988	-0.941068958	0.941145808		0.851800267	0.857301303	0.625269876	0.58620255	0.658318901	0.282452188	-0.293001757	0.013467519	0.413555736	0.509166042
over35k	-0.070379322	-0.047278259	0.176150914	-0.681806236	0.682069724	0.851800267		0.983482708	0.862351689	0.862351689	0.434318607	0.282405068	-0.349605942	-0.080282239	0.174418852	0.112347248
over50k	-0.120455333	-0.080564521	0.236725254	-0.710599533	0.710780364	0.857301303	0.983482708		0.895902401	0.592843449	0.482846547	0.323750161	-0.283142782	-0.040520471	0.178395819	0.219074567
over100k	-0.17575878	0.016933484	0.316565864	-0.43462796	0.434734239	0.625269876	0.862351689	0.895902401		0.657197937	0.406969235	0.397564222	-0.203662446	0.04025702	-0.038142686	0.11341368
prcntBA	-0.282593374	-0.145474206	0.443792338	-0.471089071	0.470921932	0.58620255	0.862351689	0.592843449	0.657197937		0.799906707	0.491191366	-0.014703784	0.11209894	0.202565451	0.615448618
prcntHS	-0.317287594	-0.208267705	0.249547585	-0.59030564	0.590180817	0.658318901	0.434318607	0.482846547	0.406969235	0.799906707		0.319042631	-0.058686032	0.027725555	0.389957877	0.754319053
prcntAsian	-0.25093511	-0.082363107	0.62235129	-0.215172501	0.214991568	0.282452188	0.282405068	0.323750161	0.397564222	0.491191366	0.319042631		-0.029357377	0.30256067	-0.1231136	0.31245474
prcntBlack	-0.213440854	-0.142302642	0.080036106	0.231523629	-0.231713168	-0.293001757	-0.349605942	-0.283142782	-0.203662446	-0.014703784	-0.058686032	-0.029357377		-0.026378109	-0.416646362	0.222388265
prcntHisp	-0.281981174	0.177434412	0.722642598	-0.021948443	0.021512314	0.013467519	-0.080282239	-0.040520471	0.04025702	0.11209894	0.027725555	0.30256067	-0.026378109		-0.23252041	0.259090583
prcntWhiteAll	-0.062840384	-0.531163605	-0.115930489	-0.469205214	0.469243791	0.413555736	0.174418852	0.178395819	-0.038142686	0.202565451	0.389957877	-0.1231136	-0.416646362	-0.23252041		0.390290278
gini	-0.429403743	-0.343076662	0.429804931	-0.605156018	0.60463861	0.509166042	0.112347248	0.219074567	0.11341368	0.615448618	0.754319053	0.31245474	0.222388265	0.259090583	0.390290278	

Adjusted Logit Results (i.e. after removal of insignificant variables)

Results: Logit

Model:	Logit	Pseudo R-squared:	inf
Dependent Variable:	party_Democrat	AIC:	1102998.2642
Date:	2019-02-01 16:01	BIC:	1103112.4650
No. Observations:	9298	Log-Likelihood:	-5.5148e+05
Df Model:	15	LL-Null:	0.0000
Df Residuals:	9282	LLR p-value:	1.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
recentArrivalPrcnt	-0.0449	0.0034	-13.2248	0.0000	-0.0516	-0.0382
totalPopBirthPlace	-0.0000	0.0000	-3.0272	0.0025	-0.0000	-0.0000
prcntForeignBorn	0.0613	0.0068	8.9552	0.0000	0.0479	0.0747
under10k	0.0675	0.0108	6.2351	0.0000	0.0463	0.0888
over10k	0.2873	0.0259	11.0820	0.0000	0.2364	0.3381
over15k	-0.2311	0.0417	-5.5438	0.0000	-0.3128	-0.1494
over35k	-0.1802	0.0368	-4.8960	0.0000	-0.2523	-0.1080
over50k	0.2541	0.0270	9.3986	0.0000	0.2011	0.3071
over100k	-0.1593	0.0120	-13.3093	0.0000	-0.1828	-0.1359
prcntBA	0.0578	0.0065	8.8770	0.0000	0.0450	0.0705
prcntHS	-0.0349	0.0045	-7.7020	0.0000	-0.0438	-0.0260
prcntAsian	0.0164	0.0079	2.0800	0.0375	0.0009	0.0319
prcntBlack	0.0132	0.0037	3.5606	0.0004	0.0059	0.0205
prcntHisp	-0.0141	0.0034	-4.0823	0.0000	-0.0208	-0.0073
prcntWhiteAll	-0.0297	0.0024	-12.4645	0.0000	-0.0344	-0.0250
gini	-3.3956	0.7332	-4.6312	0.0000	-4.8326	-1.9585