Milestone #7

Debi Gonzalez

4/15/2020

Contents

1 Introduction

Research Paper: Local demographic changes and US presidential voting, 2012 to 2016

Authors: Seth J. Hilla, Daniel J. Hopkins, and Gregory A. Huberc

1.0.1 Abstract

Hopkins and Colleagues (2019) argue that demographic changes at low levels of aggregation are not associated with increased vote shifts toward an anti-immigration presidential candidate between 2012 and 2016. More specifically, the authors' estimates find little evidence that influxes of Hispanics or noncitizen immigrants benefited Trump relative to past Republicans, instead they indicate that such demographic changes were associated with shifts to Clinton. This is a replication and extension project of the "Local demographic changes and US presidential voting, 2012 to 2016" study by Seth J. Hilla, Daniel J. Hopkins, and Gregory A. Huberc, which explored how demographic changes from 2000 to 2016 affected the 2016 general election outcome. The topics of immigration and the increase of racial and ethnic minorities have been fairly prevalent in media, social media, and political rhetoric over the last few years, but do these phenomena actually influence voting behavior? This study analyzes voting patterns in areas that received an influx of immigrants prior to the 2016 general election. This document presents the working replication exercise of the aforementioned academic paper through a textual overview of the authors' findings and methods, a high-level visual presentation of the data, and a close-replication of the original analysis. It also contains an extension proposal, which will become the core of this project, as a form of critique, development, and/or expansion of the original findings. Currently, 3 out of the 6 proposed extension analysis have been addressed.

Dan Hopkins offers a link to the data for this research paper on his website. The full paper can be found here. For more information about this project please visit my project's Github repo[1].

[1] All analysis for this paper is available here.

1.0.2 Overview of Original Paper & Literature Review

As Dan Hopkins and colleagues explain in their paper, "immigration and demographic change have become highly salient in American politics, partly because of the 2016 campaign of Donald Trump." President Trump's victory seemed to solidify the previously proposed notion that demographic changes lead to "threatened reponse" in native voters realigning voting patterns on "the basis of ethnicity, nativity, nationalism, and education." The theory implies that as demographic changes occur, native voters shift towards more populist and outspoken anti-immigration candidates. Research of anti-immigrant and Brexit, U.K. independent party, support in the United Kingdom found that these attitudes are higher in "localities that have low immigrant shares but recent demographic changes." This trend is also true of other countries in continental Europe. In the United States, previous research with county-level data have shown that pro-GOP shifts are only associated with increases in low-skilled immigrants while increases in high-skill immigrants actually increases support for the Democratic parties. [insert source ref 27] Other studies have found an associationg between the percentage change in Hispanic population and shifts to the Republican Party from 2012-2016 at county-level. [insert ref13] Despite these results, the theory remains inconclusive as another study [insert ref 28] fails to find such a strong relationship between proportion change of Hispanics and Republican support using survey data and [insert ref 29] "show that the relationship between local demographic change and Trump favorability among Republicans was time-dependent."

Notably, using precinct-level dataset of election results and demographic measures for almost 32,000 precincts in the states of Florida, Georgia, Michigan, Nevada, Ohio, Pennsylvania, and Washington, the authors of our paper of interest find that the influxes in Hispanic and foreign born groups did not influence voting behavior in favor of Trump, but rather (in a slight way) benefited his opponent in those particular areas. In other words, local demographic changes are not, on their own, increasing support for anti-immigration candidates. The authors argue that this means the precincts in question are not engaging in the often suggested "threatened response" associated with voting behavior in light of immigration influxes. The authors acknowledge that "despite its disparate local impacts, immigration may be a symbolic, nationalized issue whose effects, do not depend on local experiences." However, the actual connection has yet to be proven empirically and it is tough to do so given several factors influencing voting patterns. Some of these factors include, but are not limited to, greater exposure to international trade and declining economic prospects for the less educated. The authors explain that 'positive intergroup contact' could be responsible for the seemingly 'supportive' voting patterns at low levels of aggregation (locally). This idea is partially supported by Enos & colleagues, whose study shows how individuals with long(er) exposures to minority groups, tend to have less exclusionary attitudes than those who are exposed to them only briefly or those who have never been exposed at all.

The authors employ regression analyses with varying explanatory variable(s) specifications and measures of 'demographic change' resulting in 8 main regression models as well as loess lines that visually and mathematically demonstrate the relevant results. They study change in proportion of Hispanics and proportion change of Hispanic proportions concluding that in places with demographic shifts, Trump did not benefit, but, in fact, Clinton did. These findings are further validated by a similar model utilizing foreign-born, noncitizens proportions to discard the posibility of increased Clinton support due to changing electoral composition (given that noncitizens cannot vote and would not influence the electoral outcome in such way) (Hilla, Hopkins, and Huberc 2019).

2 Proposed Extension

1) Examine the results at higher level of aggregation: construct same models with subset of data by state

- I am interested in exploring this topic further by analyzing the variables used and figuring out whether this conclusion stands across geographic area, when the data is analyzed at the state-level. I hope to be able to do this by building the same models utilizing only a subset of the data - i.e. by state. Perhaps I will do this for only some of the states. Another way I can approach this is by grouping by county, and inputting the average values of each variable by county into the model. This would produce values relevant to the association between the explanatory variables and the outcome (gop_vote_share) at the county level.

2) How economic indicators affect Republican vote share: construct a model isolating economic indicators

- I would also like to evaluate the association between economic indicators and changes in Republican voters. I have considered exploring these association in two ways: 1) Construct a model that explains shifts in GOP share of voters in terms of changes in hispanic proportion, changes in unemployment, and proportion change of poor to consider the economic influences on Republican voters share shifts. 2) Construct a model using only changes in unemployment and proportion change of poor as explanatory variables to GOP share.

3) Examine whether there has been shifting from majority rep/dem with changes in hispanic population.

- To examine whether there has been majority shifts towards Republican or Democrats with changes in hispanic population, I can construct a binary variable that assigns '1' who went from less than 50% republican to more than 50% votes and another variable that does the same for Democrats. These variables can the be evaluated in a regression model with hispanic proportion changes as one of the explanatory variables.

4) Robustness test: Bayesian perspective

- After testing their assumptions, I can try using using Bayesian stan_glm instead of lm() to test for the model's robustness. I can also run a loo() to see if the models are valid or if perhaps they are over-controlled.

5) Missing data

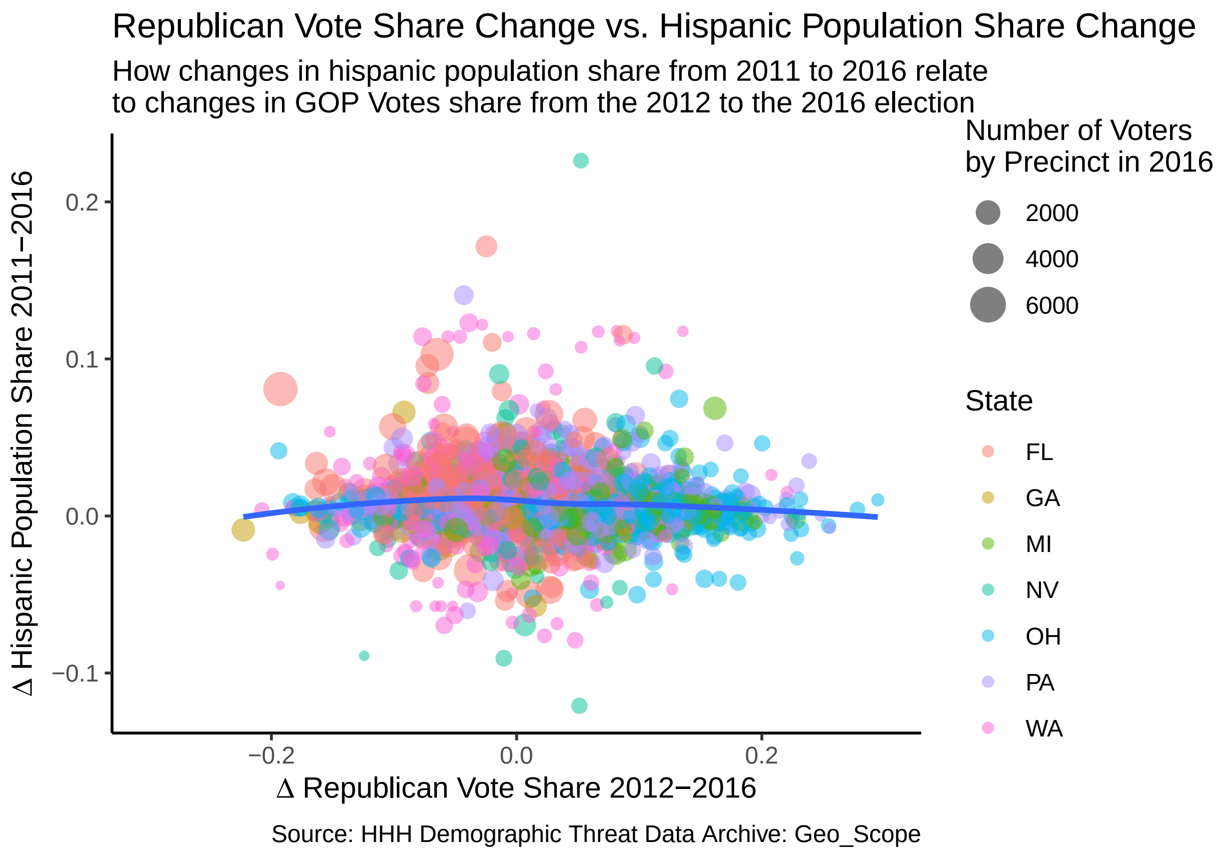
- I've been consistently receiving warning messages about ignored or dropped NA values. I'll try running the mice function and re-running some of the models to see how they compare.

6) Fixed Effects

- Heterogeneity in fixed effects models means different means among categories such as states and year. Unobserved heterogeneity is simply variation/differences among cases which are not measured. When the data can be grouped by such categories, and there are also some evidences indicating heterogeneity, the OLS is not sufficient to control the effects of these unobservable factors. However, fixed effects models can control and estimate these effects. I will construct mirror models using plm(y ~ x, data = mydata, within = "state") to see how a fixed effects model impacts original conclusions.

This paper evokes an interest in understanding how Trump's nationalist rhetoric and media's portrayal of immigration influxes may have attracted different demographic groups in cities that did not experience such influx (fearing that they might) i.e. the Deep South. At this point, the data used in the original paper cannot be used to explore this point as the 32,000 precincts in the sample are from the areas that have received an influx of immigrants (or experienced a demographic shift). However, this may be an interesting point to keep in mind for future research.

3 A Beautiful Graphic



My graphic shows the association between changes of Republican vote share and Hispanic population proportion changes over the same period of time. I am using the sample function to get a random sample of 2000 observations from my geo_scope dataset. I used the geoscope dataset because I wanted to color the observations by state, which the authors do not do, but I thought would be interesting and prettier. I was able to create the loess line using geom_smooth and it does not appear disimilar to the original paper's model(Consulting 2020). I also added the aesthetic enhancement of size as number of voters in 2016 in each precinct(Wickham et al. 2019).

To create this graphic I utilized several online guides cited in the References section of this paper(Xie, Allaire, and Grolemund 2019).

3.0.1 Extension #1: FL Subset

This extension takes the dataset and runs the same models on a smaller subset of the data. In this case, the precincts in the state of Florida. Ideally, comparing these coefficients and their significance to those of the original data models will further the central argument. At the very least, it should provide insight into the consequences of slicing the data in different ways, which could question some of their assumptions.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Apr 18, 2020 - 1:08:20 AM

Table 1: Change in Republican vote share 2012 to 2016 and change in Hispanic population, various time intervals in FL

	Dependent variable:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in Prop. Hispanic, 2011 to 2016	-0.039^{**}	-0.035^{***}	-0.068^{***}					
	(0.017)	(0.011)	(0.010)					
Prop. Change in Prop. Hispanic, 2011 to 2016				-0.003^{***}				
				(0.001)				
Prop. Hispanic 2011		-0.098^{***}	-0.121^{***}	-0.121^{***}				
		(0.003)	(0.003)	(0.003)				
Prop. Change in Prop. Hispanic, 2000 to 2016								0.002
								(0.001)
Change in Prop. Hispanic, 2000 to 2016					-0.017	0.025^{***}	-0.046^{***}	
					(0.011)	(0.008)	(0.008)	
Prop. Hispanic 2000						-0.105^{***}	-0.113^{***}	-0.114^{***}
						(0.004)	(0.004)	(0.004)
Constant	-0.007^{***}	0.004	0.110^{***}	0.108^{***}	-0.006^{***}	0.042^{***}	0.106^{***}	0.098^{***}
	(0.001)	(0.005)	(0.006)	(0.006)	(0.001)	(0.003)	(0.004)	(0.004)
Observations	4,938	4,938	4,938	4,923	4,938	4,938	4,938	4,938
R^2	0.001	0.552	0.619	0.617	0.0005	0.524	0.583	0.581
Adjusted R^2	0.001	0.551	0.618	0.616	0.0003	0.522	0.582	0.579
Note:						*p<0.1; **p<0.05; ***p<0.01		

3.0.2 Extension #2: Economic Indicators

This model isolates economic indicators to explore their effect on republican vote share as well as its association to hispanic proportion changes.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Apr 18, 2020 - 1:08:21 AM

Table 2: Change in Republican vote share 2012 to 2016 and change in Economic Indicators and Hispanic population, various time intervals

	Dependent variable:			
	(1)	(2)	(3)	(4)
Change in Prop. Hispanic, 2011 to 2016	-0.043^{***}			
	(0.010)			
Change in Prop. Poor, 2011 to 2016	0.095^{***}	0.093^{***}		
	(0.008)	(0.008)		
Change in Prop. Unemployed, 2011 to 2016	-0.296^{***}	-0.297^{***}		
	(0.016)	(0.016)		
Change in Prop. Hispanic, 2000 to 2016			-0.116^{***}	
			(0.007)	
Change in Prop. Poor, 2000 to 2016			0.247^{***}	0.235^{***}
			(0.007)	(0.007)
Change in Prop. Unemployed, 2000 to 2016			0.062^{***}	0.032^{*}
			(0.017)	(0.017)
Constant	-0.002^{***}	-0.002^{***}	-0.004^{***}	-0.008^{***}
	(0.0004)	(0.0004)	(0.001)	(0.001)
Observations	32,925	32,925	32,911	32,911
R^2	0.014	0.014	0.043	0.034
Adjusted R^2	0.014	0.013	0.042	0.034
Note:		*p<0.1; **p<0.05; ***p<0.01		

3.0.3 Extension #6: Fixed Effects by State

In statistics, a model that has fixed parameters or non-random quantities is called fixed effects model. In general, based on some observed factors, data can be divided into groups. The group means could be assumed as constant or non-constant across groups. And in a fixed effects model, just as its name implies, each group mean is a specifically fixed quantity. Furthermore, the assumption of fixed effect is that the group-specific effects are correlated with the independent variables. Heterogeneity in fixed effects models means different means among categories such as states. Unobserved heterogeneity is simply variation/differences among cases which are not measured. When the data can be grouped by such categories, and there are also some evidences indicating heterogeneity, the OLS is not sufficient to control the effects of these unobservable factors. However, fixed effects models can control and estimate these effects. Moreover, if these unobservable factors are time-invariant, then omitted variable bias can be eliminated by fixed effects regression.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Apr 18, 2020 - 1:08:32 AM

Table 3: Change in Republican vote share 2012 to 2016 and change in Hispanic population, various time intervals, fixed effects by State

	Dependent variable:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in Prop. Hispanic, 2011 to 2016	-0.007	-0.050^{***}	-0.072^{***}					
	(0.010)	(0.006)	(0.006)					
Prop. Change in Prop. Hispanic, 2011 to 2016				-0.004^{***}				
				(0.0004)				
Prop. Hispanic 2011		-0.101^{***}	-0.128^{***}	-0.128^{***}				
		(0.002)	(0.002)	(0.002)				
Prop. Change in Prop. Hispanic, 2000 to 2016								-0.006^{***}
								(0.0004)
Change in Prop. Hispanic, 2000 to 2016					0.003	-0.043^{***}	-0.105^{***}	
					(0.007)	(0.005)	(0.005)	
Prop. Hispanic 2000						-0.110^{***}	-0.123^{***}	-0.140^{***}
						(0.003)	(0.002)	(0.002)
Observations	32,929	32,311	32,311	31,368	32,913	32,911	32,911	32,909
R^2	0.00001	0.502	0.560	0.558	0.00001	0.504	0.553	0.550
Adjusted R^2	-0.0002	0.502	0.560	0.557	-0.0002	0.504	0.553	0.550
Note:						*p<0.1; **p<0.05; ***p<0.01		

This table shows a lack of statistical significance for Models #1 and #5 compared to the original and replicated models, which show statistical significance across all 8 models. I will further interpret this in my writeup.

4 Appendix

4.0.0.1 Replication of Table 1: Table 1 presents multiple least-squares regression estimates of how changes in Hispanic populations correlate with increases in Republican precinct-level vote share between 2012 and 2016.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Apr 18, 2020 - 1:08:36 AM

Table 4: Table 1. Change in Republican vote share 2012 to 2016 and change in Hispanic population, various time intervals

	Dependent variable:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in Prop. Hispanic, 2011 to 2016	-0.031^{***}	-0.045^{***}	-0.065^{***}					
	(0.010)	(0.007)	(0.006)					
Prop. Change in Prop. Hispanic, 2011 to 2016				-0.004^{***}				
				(0.0005)				
Prop. Hispanic 2011		-0.110^{***}	-0.132^{***}	-0.133^{***}				
		(0.002)	(0.002)	(0.002)				
Prop. Change in Prop. Hispanic, 2000 to 2016								-0.005^{***}
								(0.001)
Change in Prop. Hispanic, 2000 to 2016					-0.056^{***}	-0.044^{***}	-0.089^{***}	
					(0.007)	(0.005)	(0.005)	
Prop. Hispanic 2000						-0.122^{***}	-0.132^{***}	-0.147^{***}
						(0.003)	(0.002)	(0.002)
Constant	-0.00002	0.055^{***}	0.156^{***}	0.156^{***}	0.002^{***}	0.065^{***}	0.132^{***}	0.129^{***}
	(0.0004)	(0.002)	(0.003)	(0.003)	(0.0005)	(0.001)	(0.002)	(0.002)
Observations	28,934	28,934	28,934	28,934	28,934	28,934	28,934	28,934
R^2	0.0003	0.587	0.636	0.635	0.002	0.571	0.614	0.612
Adjusted R^2	0.0003	0.586	0.636	0.635	0.002	0.571	0.614	0.611
Note:						*p<0.1; **p<0.05; ***p<0.01		

I need to work on getting rid of the dependent variable row and adding the note. I also need to reorder the varibale output to best match the original image.

4.0.1 Original Table

Table 1. Change in Republican vote share 2012 to 2016 and change in Hispanic population, various time intervals

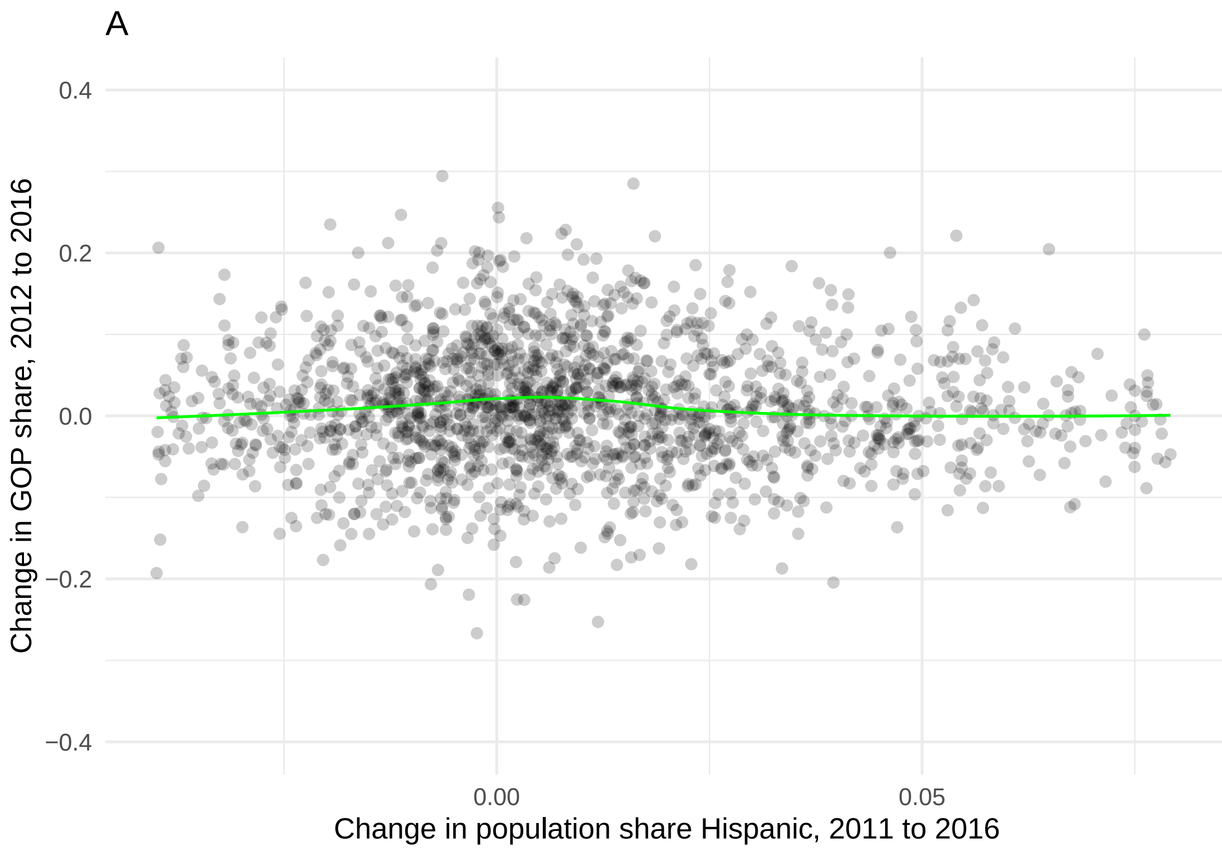
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in prop. Hispanic, 2011 to 2016	−0.040**	−0.071**	−0.077**					
	(0.01)	(0.01)	(0.01)					
Prop. Hispanic 2011		−0.13**	−0.15**	−0.15**				
		(0.00)	(0.00)	(0.00)				
Prop. change in prop. Hispanic, 2011 to 2016				−0.0041**				
				(0.00)				
Change in prop. Hispanic, 2000 to 2016					−0.077**	−0.047**	−0.085**	
					(0.01)	(0.01)	(0.01)	
Prop. Hispanic 2000						−0.13**	−0.14**	−0.15**
						(0.00)	(0.00)	(0.00)
Prop. change in prop. Hispanic, 2000 to 2016								−0.0055**
								(0.00)
Observations	31,949	31,352	31,352	31,352	31,949	31,949	31,949	31,949
R-squared	0.001	0.658	0.704	0.704	0.004	0.649	0.689	0.687
Control for levels	No	Yes	Yes	Yes	No	Yes	Yes	Yes
County fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Additional Census controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Republican share 2012	No	No	Yes	Yes	No	No	Yes	Yes

Robust standard errors are in parentheses. *$P < 0.05$; **$P < 0.01$. Precinct-level analysis; weighted to number of votes 2012; proportional changes top and bottom coded at 1 and −1. Note: Dependent variable is change in GOP vote share, 2012 to 2016. Prop., proportion.

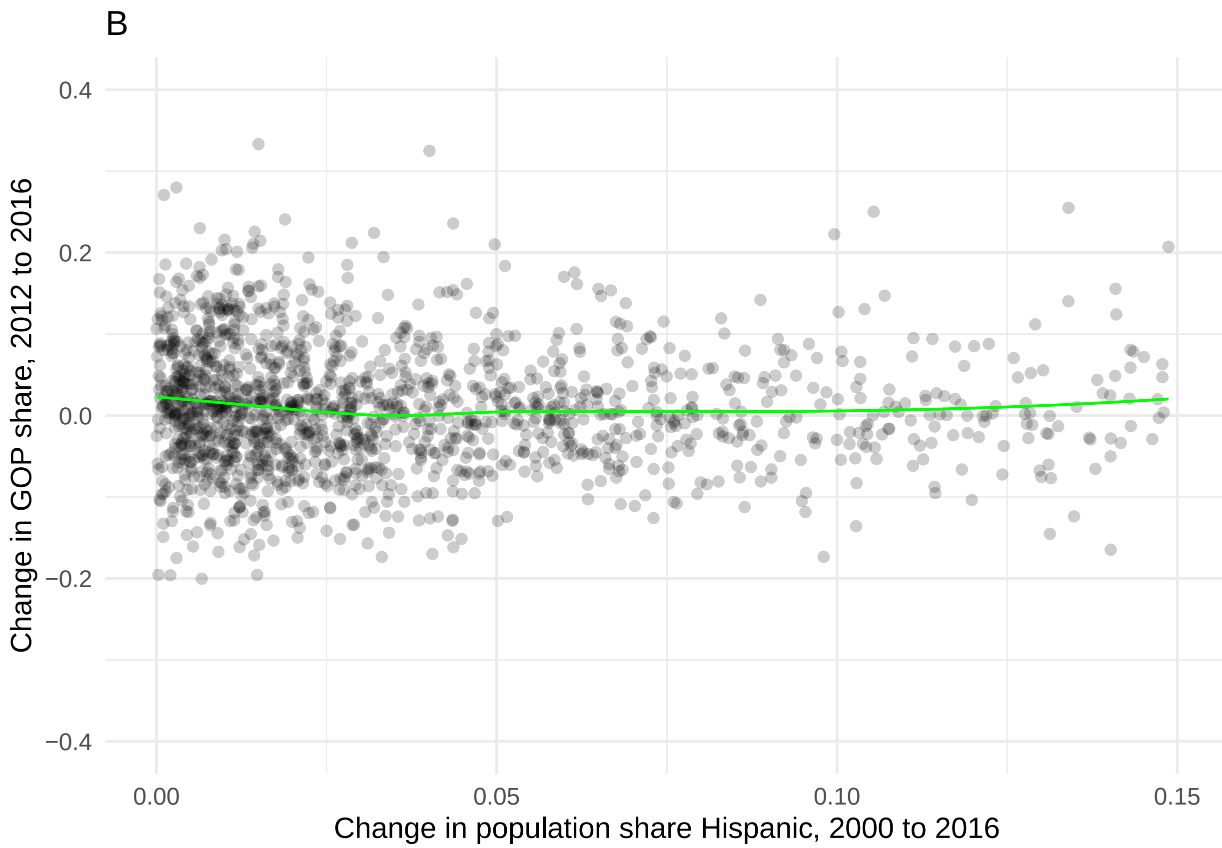
Figure 1: This table offers a summary of the 8 main regression models for the original paper while omitting several control variables in the output.

4.0.2 Figure 1

Figure 1 (A, B, C, & D) examine(s) how changes in Hispanic populations correlate with increases in Republican precinct-level vote share between 2012 and 2016. The figure shows change in the Republican share of the 2-party vote from 2012 to 2016 (positive values indicate pro-Republican shifts) against 4 different measures of change in the Hispanic population on the x axis. The 1st frame (A) measures changing population as the change in the Hispanic proportion of the overall population from 2011 to 2016, the 2nd (B) as the same change from 2000 to 2016, and the 3rd (C) and 4th (D) as proportional changes in the Hispanic population for each period.

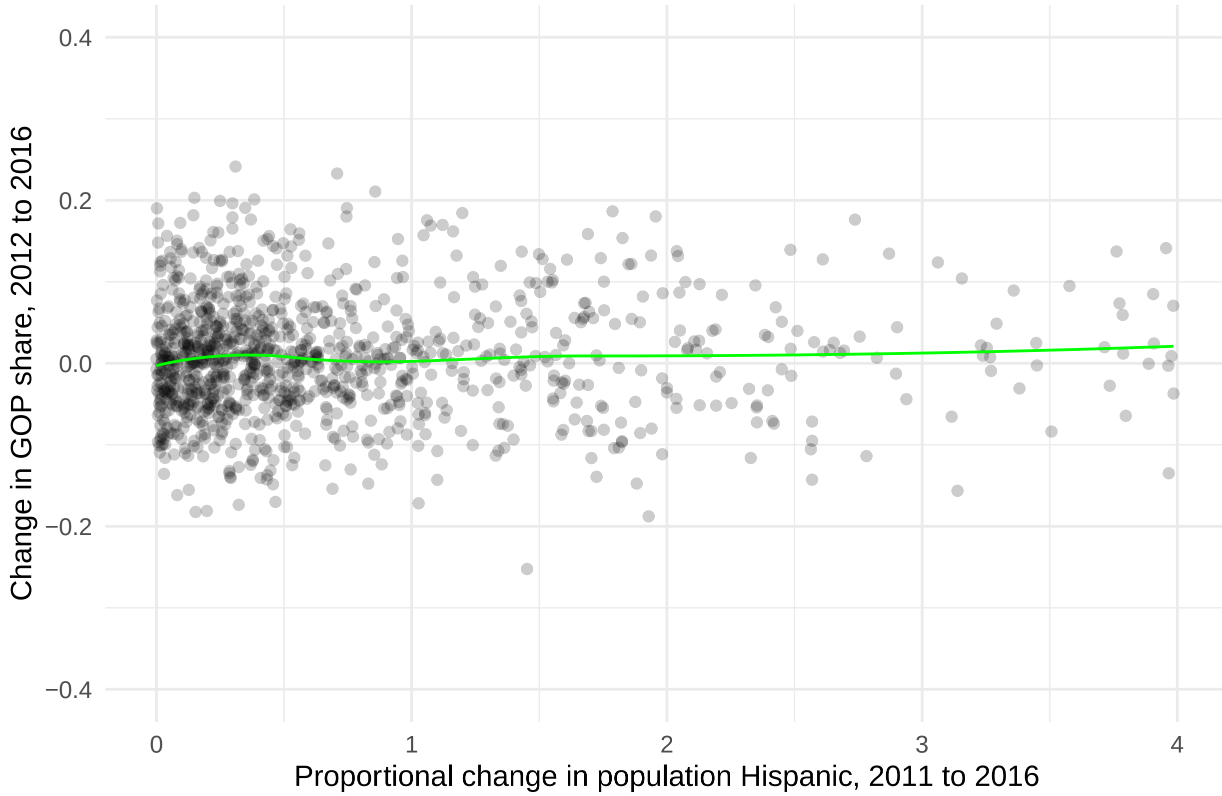



```
## Warning: Removed 420 rows containing non-finite values (stat_smooth).

## Warning: Removed 420 rows containing missing values (geom_point).
```



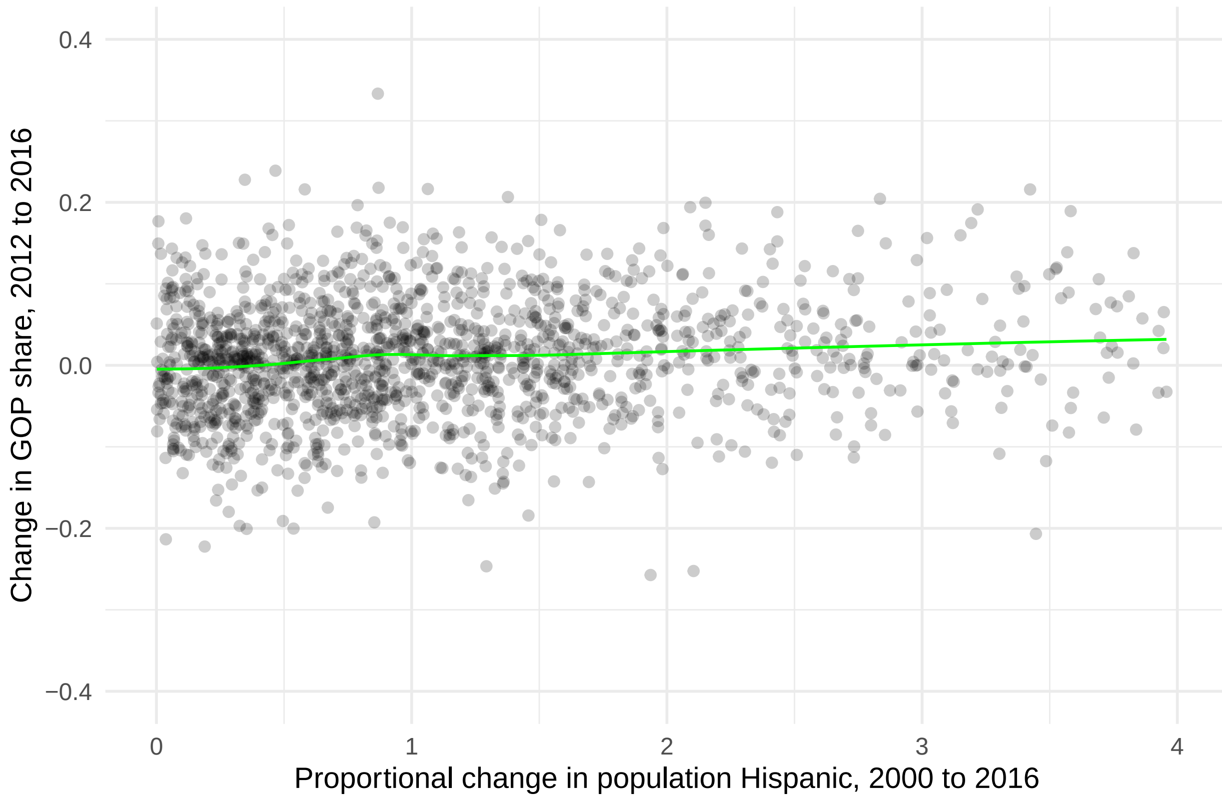

```
## Warning: Removed 887 rows containing non-finite values (stat_smooth).

## Warning: Removed 887 rows containing missing values (geom_point).
```

C




```
## Warning: Removed 432 rows containing non-finite values (stat_smooth).

## Warning: Removed 432 rows containing missing values (geom_point).
```

D


0.4
0.2
0.0
−0.2
−0.4
Change in GOP share, 2012 to 2016
0
1
2
3
4
Proportional change in population Hispanic, 2000 to 2016

4.0.3 Figure 2

I am currently working to replicate figure 2, but I'm at a bit of a crossroads with the original code. Hope to update soon. It appears that Figure 2 is putting 18 regression model coefficients and their standard deviations in a plot, some of these are not explored in the paper. When I replicate Fig.2, it will only include the main 8 regressions fully discussed in the paper. Per Alice's suggestion, I am moving on to extension as it is not necessary to include this figure.

References

Consulting, UCLA Statistical. 2020. "HOW Can I Explore Different Smooths in Ggplot2?" Available at https://stats.idre.ucla.edu/r/faq/how-can-i-explore-different-smooths-in-ggplot2/.

Hilla, Seth J., Daniel J. Hopkins, and Gregory A. Huberc. 2019. "Local Demographic Changes and Us Presidential Voting, 2012 to 2016." PNAS.org Political Science.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2019. "Smoothed Conditional Means." Available at https://ggplot2.tidyverse.org/reference/geom_smooth.html.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2019. *R Markdown: The Definitive Guide.*