PLSC 503 Final Paper

Mike Burnham
mlb6496@psu.edu

May 7, 2020

Abstract

In this research I examine partisan reactions to the COVID-19 pandemic in the United States. A reasonable theory is that partisan reactions have emerged in large part because rural and urban America have experienced a very different pandemic due to circumstance. I examine this theory to see if a partisan effect remains even after controlling for circumstance. I find that this partisan effect does in fact remain, and that while it is true that different circumstances create different experiences, we also construct these differences of our own volition. However, whether or not these partisan reactions result in different outcomes with regards to the damage done by the virus remains inconclusive.

1 Introduction

In the midst of the COVID-19 pandemic, a narrative emerged that the threat and experience of the virus in the United States is a partisan affair [1]. This is not altogether surprising. Political ideology has a strong correlation with both how we perceive threats [2], and who we blame when threats come to fruition [3]. Cursory statistics and polling data support this notion. Republicans are less likely to see the virus as a threat, and more likely to see restrictions on travel and socializing as an infringement of rights [4]. For their part, democrats are more likely to say they have been impacted by the virus and favor waiting longer before reopening the economy [5] [4]. This raises a few questions: First, if Republicans and Democrats are experiencing different pandemics, is this merely a function of different circumstances or does ideology shape our response and perception even when controlling for circumstantial differences? Second, if ideology is influencing the perception and reaction to the pandemic, to what extent has it influenced the outcome?

Understanding partisan divides has become foundational to understanding American politics. Data from the COVID crisis may help us further understand to what extent these divides are a result of geographic sorting, and to what extent they are of our own making. Do we have different experiences merely because we have chosen to self sort? Or has our partisanship also become a self-fulfilling prophecy in which our ideology governs our reactions, and thus we create different lived experiences that further reinforce this divide? This interplay between the world that circumstance or nature creates for us and that which we construct for ourselves is, of course, not a new topic to social science. Perhaps the most notable example is the push and pull between constructivism and other schools of thought in international relations. Yet, the opportunity to quantitatively measure this balance is rare. In this research I leverage data from the COVID pandemic to take a quantitative look at how partisanship constructs different experiences independent of circumstance.

2 Hypotheses

To address this question, I control for the circumstances in which a geographic area was affected by the pandemic, and then test to see if a correlation between partisanship and how politicians reacted remains. If it does, this provides evidence of a partisan effect, and thus social construction, independent of circumstances. Further, I test to see if these differences affected the outcome of the pandemic.

As the above polling data suggested, republicans are less likely to see the virus as a threat and more likely to question the validity of travel restrictions and social distancing measures. Accordingly I propose three hypotheses:

H1: All else being equal, areas that favor Republican candidates are less inclined to issue stay at home orders.

H2: When stay at home guidelines are issued, areas that favor Republican candidates will be relatively slower to issue them.

H3: If the above hypotheses are true, areas that favor Republican candidates will experience disproportionately more deaths after controlling for circumstance.

3 Data

To test the above hypotheses I compiled a data set of county level data that includes how the virus has affected the area, how the area responded to the virus, basic demographic information, as well as vote share from the 2016 election. Data on COVID-19 was collected from the Corona Data Scraper project, an open source effort that automates scraping data directly from all publicly available government sources across the world. Electoral data comes from the New York Times, and demographic data comes from a range of sources including the Census Bureau, RAND, and others. In total, the data consists of 3,142 counties from all 50 statues plus the District of Columbia. The data is publicly available online on a GitHub repository associated with this project.

This data offers fairly complete and perhaps the most granular level data publicly available on the COVID-19 pandemic. This fortunately allows me to sidestep many of the multicolinearity issues inherent to modeling ideology and demographic variables. While correlations between the variables still exist (see table 6 in the appendix for a correlation table of the primary variables used in the analysis), I will state here rather than calling specific attention to it for each model: multicolinearity by and large did not compromise the reliability of model estimates. Notable exceptions will be called attention to, but by and large the variance inflation factor for estimates rarely exceeded 2.

4 Analysis

4.1 Hypothesis 1: Areas that favor Republican candidates are less inclined to issue stay at home orders.

To test my first hypothesis I used logistic regression to model whether or not a county would issue stay at home order. As the independent variable I use the percentage of votes for the Republican candidate in the 2016 election. I use two models to test this variable. The first model uses population adjusted metrics (e.g. population density) and the second model uses the log of non-population adjusted metrics (e.g. the log of the total population). This allows me to test the robustness of the estimates as well as preform basic diagnostic tests to evaluate the models.

In addition, because I am trying to measure the ideological influence of a county's response I introduce a number of controls for other factors decision makers must weigh. Specifically, I aim to control for the vulnerability of a population and the visible impact of the virus. I reason that populations which are more vulnerable and have seen a more direct impact of the virus are more likely to implement travel restrictions. To control for the vulnerability of the population I use the size of the population (population density in model 1 and the log of the population in model 2) and the size of the population over 65 years of age (percentage over 65 in model 1 and the log of the total population over 65 in model 2). While there are other variables that contribute to the vulnerability of the population, population size and age are largely believed to be the most significant factors and I argue it is unlikely decision makers are considering every demographic attribute when making a decision. Accordingly, I restrict my model to those I believe they are most likely to consider.

The primary variable I use to control for the visible impact of the virus is the number of deaths. I chose this over the number of cases because of the general unreliability and inconsistency of testing for the virus within this time period. As an additional robustness check I run iterations of the models with cases included in addition to deaths. In model 1 these variables are represented by the percent of the total population killed or infected by the virus, and in model 2 I use the log of the total number of deaths or cases. Models that include controls for cases are labeled models 1A and 2A. The results of the models are shown in table 1. I will first briefly discuss diagnostics and then discuss the results.

To evaluate the fit of the models I plot the ROC curves in figures 4.1 and 4.1. The area under the curve for both models is slightly over 72%, indicating a relatively good fit for the model. The confusion matrices for models 1 and 2 are shown in in tables 2 and 3. Both models uniformly predicted stay at home orders for each county. While these models would not suffice for predictive purposes, they fit the data reasonably well for inferential purposes. From a model fit standpoint, both models seem to fit the data equally well.

While the model fits are nearly identical, model 2 does suffer from some multicolinearity. In both models 1 and 2 the variance inflation factor of all variables fell between 1 and 2 with the exception of the log(Pop) and log(Over65) in the second model. Both variables had a variance inflation factor of over 30. This, of course, make sense. When not using population adjusted metrics the total number of elderly citizens will strongly correlate with the size of the population.

Table 1: H1 Logistic Regression

	Dependent variable:			
	Stayhome Order Issued			
	Model 1	Model 1A	Model2	Model2A
	(1)	(2)	(3)	(4)
R Vote Perc.	-0.016***	-0.016***	-0.008**	-0.009**
	(0.004)	(0.004)	(0.004)	(0.004)
Perc. Killed	220.399***	244.439***		
	(68.390)	(74.727)		
Perc. Infected		-1.887		
		(2.270)		
Perc. Over 65	2.784**	2.687**		
	(1.171)	(1.177)		
Pop. Density	0.005***	0.005***		
	(0.001)	(0.001)		
log(Deaths)			0.026***	0.027***
			(0.008)	(0.008)
log(Cases)				-0.002
				(0.009)
log(Over 65)			1.276***	1.272***
			(0.238)	(0.239)
log(Pop)			-0.740***	-0.729***
			(0.225)	(0.229)
Constant	1.646***	1.716***	-0.823	-0.886
	(0.355)	(0.365)	(0.672)	(0.716)
Observations	3,062	3,062	3,068	3,068
Log Likelihood	-1,295.236	-1,294.914	-1,252.914	-1,252.881
Akaike Inf. Crit.	2,600.471	2,601.829	2,515.827	2,517.762
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 2: Model 1 Confusion Matrix

		FALSE	TRUE
1	FALSE	0	0
2	TRUE	514	2,548

Table 3: Model 2 Confusion Matrix

		FALSE	TRUE
1	FALSE	16	19
2	TRUE	498	2,535

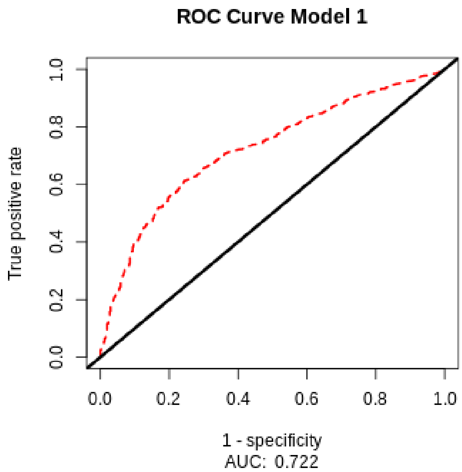

Figure 1: ROC Curve Model 1

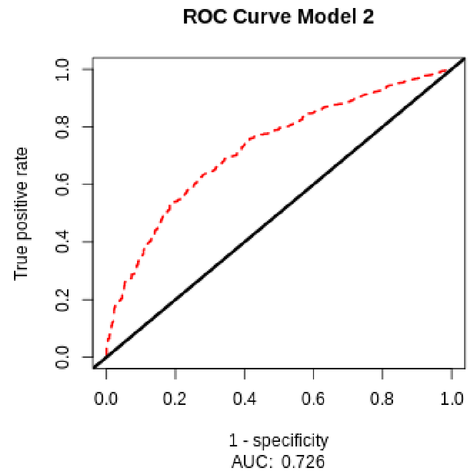

Figure 2: ROC Curve Model 2

Accordingly, the corresponding variables in model 1 provide more robust estimates for the effect of population and elderly population.

Consistent with my hypothesis, across all models the Republican vote share is negatively correlated with issuing stay at home orders. The signs on the coefficients of all variables are largely in the expected direction: areas that are more vulnerable and have had a higher visible impact from the virus are more likely to implement stay at home orders. The exception to this is the aforementioned population metrics that suffer from multicolinearity.

A reasonable theory to explain partisan reactions to the virus is that rural counties have not been hit as hard by the virus as densely populated ones, and that this rural/urban divide also correlates with ideological divides. While this may be true, these results provide evidence that partisan reactions extend beyond just circumstance.

4.2 Hypothesis 2: When stay at home guidelines are issued, areas that favor Republican candidates will be relatively slower to issue issue them.

For my second hypothesis I examine the speed at which communities reacted to the virus. To do this, I use a metric called exposure days as my dependent variable. Exposure days counts the number of days since a county's first case in which a stay at home order was not in place. For example, a county that issued stay at home orders two weeks after its first case would have 14 exposure days. Similar to the previous models, I use the Republican vote share in the 2016 presidential election as my independent variable and population adjusted statistics for the population vulnerability as my control variables. In addition, I add a control variable for the number of days since since a county had its first case as of the end of this data set (April 26, 2020). This variable is intended to control for the fact that the virus spread to some counties much later and thus those counties had a fewer number of possible exposure days. I ran two iterations of this model, one in which all counties are left in the data set and one in which only counties that issued a stay at home order are kept. I do this because the data was collected in the midst of the pandemic roughly three months after the first case in the United States. The assumption that the virus has not year fully penetrated the United States and that some counties may yet issue stay at home orders, and the counter assumption that all counties that will issue orders already have, are both reasonable. The model with all data is labeled model 3 and the model that subsets the data is model 3A.

The results are shown in table 4. Models 3 and 3A show inconclusive results. When including all data there is a positive correlation between the Republican vote share and exposure days. However, when you subset the data to instances where stay at home orders were issued, the result disappears. This indicates that the result is largely driven by counties that have not issued orders. When we look at the residual plots in figure 4.2 we see something else: residuals are not normally distributed and are skewed towards a cluster ranging from -10 to 10.

I hypothesize that the heteroskedasticity in the residuals is the result of a factor which the model is not accounting for: public awareness. In both models, the number of days since a county's first case has a strong positive correlation with with the number of exposure days. This indicates that the earlier in the year your county had a case, the slower it was to react. This makes some intuitive sense. Early in the pandemics spread it was unclear how severe or contagious

Table 4: H2 Linear Regression

	Dependent variable:			
	Exposure Days		Adjusted Exposure Days	
	Model 3	Model 3A	Model 4	Model 4A
	(1)	(2)	(3)	(4)
R Vote Perc.	0.047***	−0.011	0.095***	0.043***
	(0.014)	(0.008)	(0.016)	(0.009)
Perc. Over 65	−14.162***	−3.858	−19.132***	−8.341***
	(4.490)	(2.514)	(5.093)	(2.713)
Pop. Density	−0.00003	0.0001	−0.0002	−0.0001**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Days Since 1st Case	0.544***	0.571***		
	(0.020)	(0.012)		
Adj. Days Since 1st Cast			0.132***	0.015
			(0.023)	(0.013)
Constant	−7.656***	−10.438***	13.180***	11.164***
	(1.593)	(0.909)	(1.276)	(0.683)
Observations	2,748	2,333	2,748	2,333
R^2	0.254	0.588	0.045	0.023
Adjusted R^2	0.253	0.587	0.043	0.022
Residual Std. Error	9.934 (df = 2743)	5.214 (df = 2328)	11.269 (df = 2743)	5.628 (df = 2328)
Note:				*p<0.1; **p<0.05; ***p<0.01

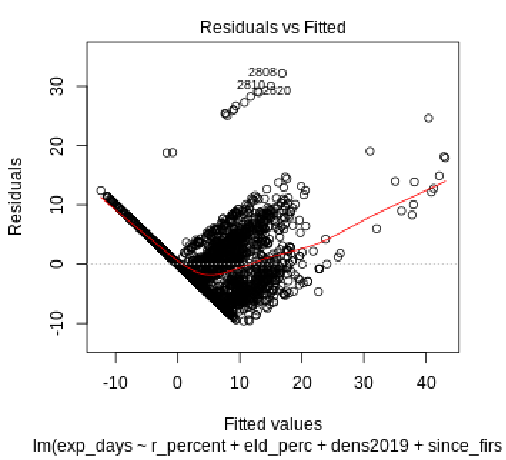

Figure 3: Model 3A Residuals

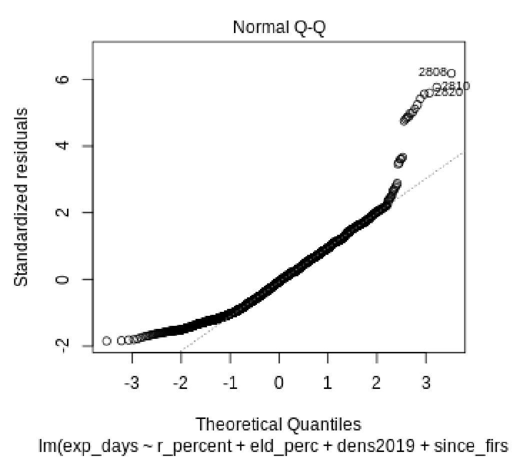

Figure 4: Model 3A QQ Plot

the virus was and public apprehensions were much lower. Accordingly, public officials were likely more measured in their responses. When we look at the distribution of the spread of the virus in figure 5 we see that democratic counties, on average, contracted the virus earlier than Republican ones. On average, Democratic counties contracted the virus 10 days earlier than Republican ones (95% CI:[8.99, 11.42]). In the context of exponential growth, this is a significant difference. With this in mind, the negative coefficient on Republican vote share in model 1A makes sense, despite its lack of significance. The average Republican county contracted the virus later in the pandemic and thus had the advantage of better public awareness when the threat of the virus was imminent.

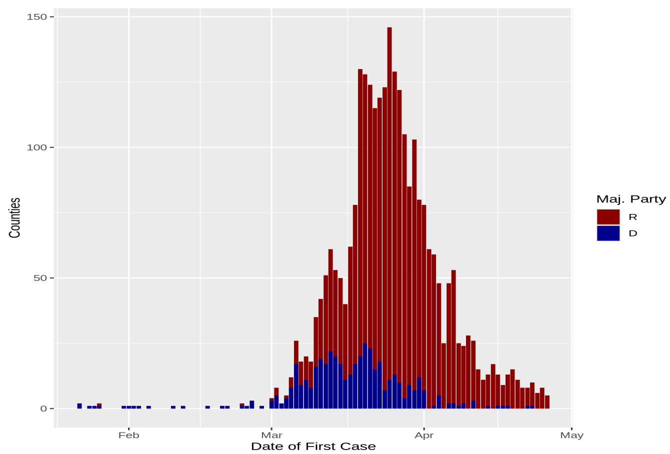

Figure 5: Distribution of when a county experienced its first case by majority party vote share in the 2016 election.

To accurately model the speed at which counties responded to the pandemic I thus need to control for the level of public awareness. To do so, I created two new variables. First, as the dependent variable I use adjusted exposure days. Adjusted exposure days centers the number of exposure days on a point in time at which public awareness of the pandemics reality had become appreciable. For this date I used March 16, 2020. This is the day the federal government issued social distancing guidelines to avoid gatherings of 10 or more as well as instructions to cancel all unnecessary travel. Additionally, rather than using the total number of days since the virus was first contracted in the county, I used the difference between the first case and March 16th.

Using these new metrics I reran models 3 and 3A and results can be seen in table 4 under models 4 and 4A. These transformed variables provide more consistent estimates of the coefficients and significantly alleviate the issue of heteroskedasticity as seen in figure 4.2. In both iterations of model 4 the Republican vote share is positively correlated the number of exposure days. This supports my second hypothesis and indicates that when adjusting for public

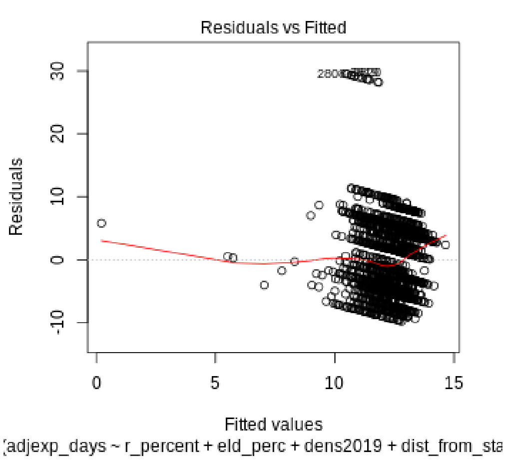

Figure 6: Model 4A Residuals

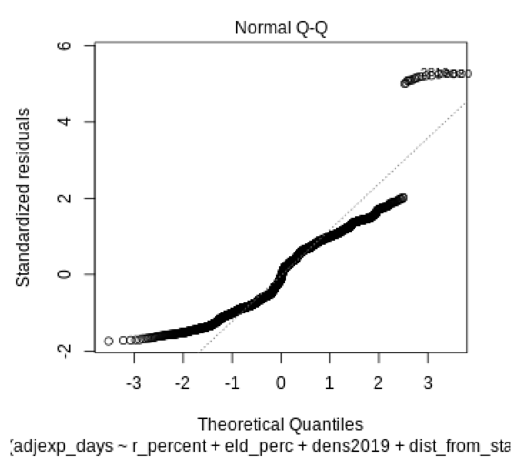

Figure 7: Model 4A QQ Plot

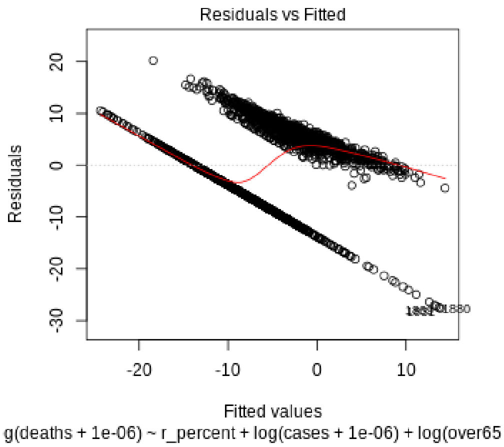

Figure 8: Model 6 Residuals

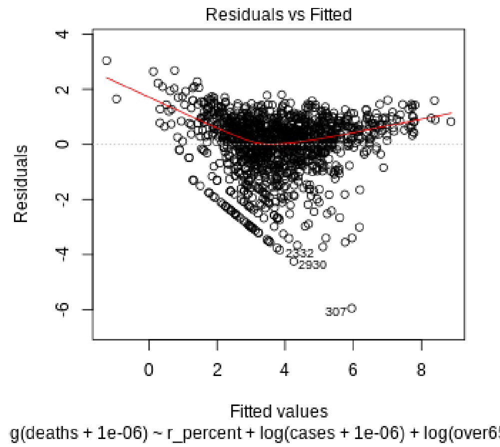

Figure 9: Model 6A Residuals

awareness and population vulnerability, counties that voted for Donald Trump in 2016 were slower to react and issue stay at home orders.

4.3 Hypothesis 3: Areas that favor Republican candidates will experience disproportionately more deaths

For the final hypothesis I test if ideological reactions to the virus have lead to measurably different outcomes. As with the first hypothesis I use two models, one with population adjusted metrics (model 5) and the other without (model 6). The first model uses the percentage of the population that died of the disease as its dependent variable and the second uses the log of the total number of deaths in the population. In addition to the previously used controls for the vulnerability of the population and exposure to the virus, I include an additional variable for obesity. I include this here because it is a known risk factor, but excluded it from previous models because I was modeling elite decision making and I argue that decision makers were unlikely to consider proportion of obese residents in their decision making. Additionally, I run two versions of each model. The first uses all available data and the second subsets the data to counties that have experienced at least one COVID-19 related death. This again has to do with varying assumptions about the spread of the virus. It is unclear if an absence of deaths is simply because the pandemic is still in its early stages, or if it is a function of appropriate countermeasures being taken.

Results for the models can be seen in table 5. Support for my third hypothesis is far less robust than hypothesis one and two. Both the sign and significance on the coefficient for the Republican vote share variable is inconsistent. In terms of model fit, models that included all counties again suffered from heteroskedasticity due to instances with zero deaths. Examining the log of the data did not alleviate that issue. Models using square roots as well as weighted least squares using 1/(fitted values)^2 as a weight were also attempted, but had little impact. This was partially alleviated by subsetting the data to those with at least one death as shown in the residual plots for models 6 and 6A in figures 4.3 and 4.3.

Given the persistence of heteroskedasticity and the nature of the data, the inconclusiveness of the models here is possibly due to incomplete data. For one, it may simply be too early in the pandemic's life cycle to draw definitive conclusions about outcomes. For another, significant variables may be missing or the data may simply not be granular enough. For example, adherence to social distancing guidelines is likely to play a significant role in outcomes but there is no such metric in this data set. Likely, it is probably a function of both missing variables and the premature nature of the data.

5 Discussion

Notably absent from this analysis is a lack of economic data as well as adherence to social distancing guidelines and granular public opinion data. Future research on the political implications of this pandemic could greatly enhance this analysis by controlling for these variables.

Table 5: H3 Linear Regression

	Dependent variable:			
	Deaths (Perc. of Pop.)		log(Total Deaths)	
	Model 5	Model 5A	Model 6	Model 6A
	(1)	(2)	(3)	(4)
R Vote Perc.	-0.00000^{**}	0.00000	0.016	0.004^{*}
	(0.00000)	(0.00000)	(0.010)	(0.002)
Perc. Infected	0.026^{***}	0.045^{***}		
	(0.001)	(0.001)		
Perc. Over 65	0.001^{*}	0.004^{***}		
	(0.001)	(0.001)		
Pop. Density	-0.00000^{***}	-0.00000^{***}		
	(0.00000)	(0.00000)		
log(Cases)			2.822^{***}	0.845^{***}
			(0.134)	(0.034)
log(Over 65)			1.027	0.694^{***}
			(0.631)	(0.152)
log(Pop)			-0.649	-0.693^{***}
			(0.618)	(0.148)
Obesity	0.006^{***}	0.005^{***}	32.480^{***}	1.365^{*}
	(0.001)	(0.001)	(3.221)	(0.822)
Exposure Days	-0.00000	-0.00001	0.054^{**}	-0.010^{*}
	(0.00001)	(0.00001)	(0.026)	(0.006)
Stayhome Orders	0.0001	-0.0001	1.968^{***}	0.041
	(0.0002)	(0.0003)	(0.722)	(0.189)
Days Since First Case	0.00001^{***}	0.00001	-0.084^{***}	0.017^{**}
	(0.00000)	(0.00001)	(0.026)	(0.007)
Constant	-0.002^{***}	-0.003^{***}	-35.237^{***}	-2.327^{***}
	(0.0004)	(0.001)	(2.092)	(0.556)
Observations	2,748	1,244	2,754	1,249
R^2	0.371	0.624	0.416	0.637
Adjusted R^2	0.369	0.622	0.414	0.635
Residual Std. Error	0.002 (df = 2739)	0.002 (df = 1235)	6.775 (df = 2745)	1.076 (df = 1240)
Note:				*p<0.1; **p<0.05; ***p<0.01

Despite these shortcomings, this analysis found support for the hypotheses that partisan responses to the COVID-19 crisis are not merely a function of our environment. When controlling for circumstance, Republican leaning counties were less likely to implement travel restrictions and slower to do so when they did. While the influence of circumstance in differentiating experiences is not to be understated, this is evidence that we are in part constructing these different experiences of our own volition. Whether or not this will lead to tangible differences in the damage caused by the virus remains to be seen. However, if we assume that our experiences shape our political beliefs, the different experiences red and blue America have chosen to construct for themselves can only exacerbate the partisan divide.

Appendix

Table 6:

	R vote%	Pop.	Elderly Pop.	Pop. Dens.	Cases	Deaths	obesity	Exp Days	Death %	Infected %	Stayhome Issued
R vote%	1	-0.356	-0.360	-0.239	-0.245	-0.244	0.236	-0.176	-0.163	-0.268	-0.130
Pop.	-0.356	1	0.984	0.333	0.511	0.541	-0.247	0.271	0.085	0.209	0.090
Elderly Pop.	-0.360	0.984	1	0.339	0.530	0.555	-0.269	0.269	0.094	0.224	0.103
Pop. Dens.	-0.239	0.333	0.339	1	0.640	0.172	-0.160	0.090	0.056	0.353	0.047
Cases	-0.245	0.511	0.530	0.640	1	0.426	-0.172	0.112	0.140	0.618	0.053
Deaths	-0.244	0.541	0.555	0.172	0.426	1	-0.125	0.135	0.387	0.320	0.061
obesity	0.236	-0.247	-0.269	-0.160	-0.172	-0.125	1	-0.097	0.064	-0.097	-0.059
Exp Days	-0.176	0.271	0.269	0.090	0.112	0.135	-0.097	1	0.048	0.119	-0.662
Death %	-0.163	0.085	0.094	0.056	0.140	0.387	0.064	0.048	1	0.572	0.078
Infected %	-0.268	0.209	0.224	0.353	0.618	0.320	-0.097	0.119	0.572	1	0.061
Stayhome Issued	-0.130	0.090	0.103	0.047	0.053	0.061	-0.059	-0.662	0.078	0.061	1

References

[1] Ronald Brownstein. Red and blue america aren't experiencing the same pandemic, March 2020. [Online; posted 22-March-2020].

[2] Becky L Choma, Yaniv Hanoch, Michaela Gummerum, and Gordon Hodson. Relations between risk perceptions and socio-political ideology are domain-and ideology-dependent. *Personality and Individual Differences*, 54(1):29–34, 2013.

[3] Cherie D Maestas, Lonna Rae Atkeson, Thomas Croom, and Lisa A Bryant. Shifting the blame: Federalism, media, and public assignment of blame following hurricane katrina. *Publius: The Journal of Federalism*, 38(4):609–632, 2008.

[4] The Economist/YouGov. April 19 - 21, 2020 - 1500 US Adult citizens, 2020.

[5] Pew Research Center. 2020 american Trends Panel Wave 64 March 2020, 2020.