

1 Abstract

- I'm interested in whether the effect of political bias on social distancing behavior is visible in the data, or is either the fringe behavior of a relative few people or just for show.
- I'm doing a descriptive regression analysis at the county level to correlate social distancing behavior with Trump voting, while controlling for a broad range of factors.
- I'm using a clustering strategy to enhance the comparability of observations underlying the regression estimates, to reduce the demands upon the linear model.
- I find that there is a significant degree of observed Trump effect - above and beyond the Republican effect.
- I use a machine learning model to show that the degree of Trump effect is higher in places that are more politically heterogenous/contentious.

2 The Set Up

- People are reacting to COVID in different ways.
- Given the options available, stopping the spread of the virus hinged upon one central policy proposition: implementing strong social distancing by issuing stay-at-home orders which amount to shutting down significant portions of the economy in order to reduce transmission through in person contact.
- There is a political divide evident among the ways both elected officials and citizens.
- Conservatives are and have been publicly prioritizing the economy over public health measures.
- Conservative politicians have also resisted or been slower to implement public health measures, which necessarily cause a restriction on economic activity.
- For instance, Mississippi, Texas...Florida.
- Conservative leaning citizens are also implementing protests against social distancing orders as of 4.15.
- Meanwhile, liberals have been publicly prioritizing the importance of public health measures, even at the expense of the economy.

- Liberal politicians have also called for public spending programs to smooth consumption and provide insurance for people who are negatively affected by public health interventions. Conservatives have resisted these measures.
- Political affiliation has analogs in media consumption, with conservative and liberal media outlets presenting different information, or the same information differently, with different points of emphasis, different levels of concern, and different takeaways regarding preparation and public health.
- Therefore, there is reason to suspect that liberal and conservative citizens may exhibit different behaviors during the COVID epidemic, due to different mental models primed by the information presented to them about the nature of the disease and the appropriate safety measures to take in response.
- However, these associations might be more in the nature of casual association, the ready caricatures of a stereotyping mind, than a broad look at the data might suggest. The appearance of protesters at state capitals armed not only with guns and signs, but also with prophylactic masks, suggests a level of caution even among those who are chafing at that the constraints imposed by the demands of public health authorities.
- Moreover, while appearances suggest the protesters come largely from the right, apparent relationships don't always bear out in the data, and the small group of people motivated to protest is probably less important from a public health standpoint than the overall behavior of the bulk of the population.
- So, has media messaging had an observable effect? Do conservatives demonstrate less social distancing response (and therefore higher spread) than liberals?

3 Social Distancing Metrics and the Early Spread of Coronavirus Map

- You've probably seen a zillion and one COVID-19 maps in the past 12 weeks, and, yes, I definitely have more for you.
- This first map shows a few variables which describe the spread of Coronavirus as well as social distancing metrics over time, and I include it because it's nice to have handy, and because it allows one to begin to correlate in their mind the severity of viral spread and the social distancing actions taken in response to that spread with the political map.

- We can see that the spread of the virus wasn't particularly time-varying. Yes, it arrived to the coasts, and Colorado, first, but there wasn't a significant amount of time before it had made itself known as a threat in most parts of the country.

4 Let's Make a Model

- I'm going to use a regression model to look at the effect of media consumption on social distancing behavior. I'll produce estimates of the changes in mobility corresponding to variation in likely media consumption.
- I say likely media consumption, because I don't have a broad, geographically precise and sufficiently disaggregated measure of the types of media being consumed.
- This data might exist, and an interesting add-on to this analysis would be to use something like Google Trends to correlate searches for misinformation keywords with social distancing behaviors.
- I'll be using the percentage of citizens in each county who voted for Trump in 2016 as a proxy for media consumption. This has advantages in that it is a fairly broad survey of opinion, and has a strong correlation to media consumption. While it isn't a direct measure, and suffers from degradation over time as a measure of current political opinion, its availability in terms of geographic detail makes it an appealing stand-in.
- If you're uncomfortable making the leap to media messaging, the use of Trump voting answers a slightly different question more directly: how much of the variation in spread of COVID is driven by political affiliation?

5 Controlling Confounders - also show Demographics, Economics, and Health Data

- Social distancing behavior in total is the result of many different factors, forces, and circumstances.
- For instance, the more educated are in general more healthy, more wealthy, and may be more receptive to the guidance of scientifically-oriented public health authorities.
- They are also less likely to vote for and listen to Trump.
- Wealthier people, by virtue of the ability to store goods, hire delivery people, hole up in larger homes, and work remotely, are much more able to undertake social distancing.

- All of which illustrates the analytical task - how do we isolate out the relationship between media messaging from other factors related to both media consumption and social distancing patterns?
- This map shows the geographic distribution of a large set of variables, which you can peruse through, that show distinct regional patterns and which are likely to be relevant to either health behaviors, political behaviors, or both.

6 Regression Strategy

- A natural experiment would be great, but I don't know of any broadly based shock to political affiliation or media consumption that could be pressed into service here.
- So, we won't be able to produce a causally strong estimate.
- Instead, we'll aim for the strongest descriptive evidence we can muster. With that in mind, the aim of our regression analysis is to form comparisons between very similar places - our unit of analysis will be US counties - that differ only in the variable of interest: the percentage of the presidential vote that went to Trump in the county in 2016.
- The idea we're pursuing is to replicate as nearly as possible the ideal, but impossible, experimental scenario of altering the political preferences of a county and observing any resulting change in social distancing behavior. If two counties show the same observable characteristics for a sufficiently comprehensive set of variables, then it is likely they'll show similar social distancing behaviors. If all characteristics are the same *except* political affiliation - measured by Trump voting, then this strongly suggests the difference in social distancing behaviors is related to Trump voting.

7 Identification

- Even though we don't have a strong causal methodology, it's worth discussing identification so that we understand where there may be biases leaking into the estimates I'll produce.
 - Political affiliation measured through voting patterns positively selects media messaging.
 - Political affiliation does not have an independent effect on social distancing behavior in the absence of regulations and of media consumption.
 - * Regulation timing and severity is correlated with political affiliation through choice of leadership. Therefore, we need to control

for the regulatory regime if we want to understand the effect of media consumption on political choices.

- There aren't any factors not directly included as controls or well-proxied-for by the included controls, which cause both the changes in movement by which we are measuring social distancing, and voting behavior by which we are measuring politics/media consumption.
 - * Job distribution might be incompletely proxied by education, and different jobs held by conservatives/rural citizens might not be as conducive to distancing. Which would mean that we see less distancing in the model, and this might or might not translate to infection rates. If the jobs themselves are not able to be done while distancing, and if they are not covered by stay-at-home orders, then you might see conservative areas with higher spread rates due to job composition.
 - * Fortunately, I am able to control for job composition.
- Timing issues. If the conservative areas of the country are able to benefit from the advance knowledge generated by the earlier spread of coronavirus to coastal and more liberal areas, then they might have been able to develop other strategies to control for and prevent the spread of COVID that free them of the need to rely on social distancing.
- The model's validity relies on the extent to which the above assumptions are justified.
- It's unlikely that we can attribute all the effect that we find to political affiliation, even with an extremely robust set of covariates.

8 Describing the Outcome Variable - Math Details

- I measure the amount of social distancing using data from Descartes Labs, which has provided time series data on the average (?) movement at a county level beginning on February XXth.
- Descartes Labs generates its estimates using (???)
- The Descartes Labs data is available at **Github Link**
- I process these time series into a single estimate of the amount of social distancing undertaken by a county.
- I generate this estimate by first calculating a date at which social distancing began in earnest in the county (I use the date when movements first dropped below 50% of their pre-COVID average). This accounts for time-based variation in the geographic spread of the virus, and in government

shut-down orders based on that spread. This helps us to obtain a uniform sample in the face of time variation in the spread of the virus.

- To calculate this date, I first take each time series, and generate a smoothed version by applying a flexible spline function. This removes day-to-day noise in the social distancing metric.
- Finally, I take the average movement after the initiation date through the end of April as the metric for the degree of social distancing behavior in each county.

9 Describing the Control Variables - Logic Model

- We need to control for as many factors which explain social distancing as possible.
- I control for: Income, Occupation, Health and Disease characteristics, Demographics, Other Stuff.
- Education: Which variables? Where does the data come from?
- Income: Which variables? Where does the data come from?
- Occupation: Which variables? Where does the data come from?
- Demographics: Which variables? Where does the data come from?
- Health factors: Which variables? Where does the data come from?
- Political factors: Which variables? Where does the data come from?
- Discuss factor extraction for vars repeated over time.

10 OLS - Building Up

- What's correlated with Social Distancing? Is it the things we expect?
- What's correlated with voting patterns?
- What's correlated with Social Distancing after we add voting patterns?
- What happens when we add Trump voting on top of the political factor?
- What happens when we use a Tikhonov Reg to control multicollinearity?

11 For Nerds Only - Covariance Structures and PCA

- The point of this section is to make it clear that the data does break down into a cohesive set of principal components, even though there are a large number of variables.
- This means that the clustering algorithm is going to be able to divide the data along meaningful dimensions - separating the data into groups that vary by their position along the main covariance channels in the data.
- In fact, PCA and k-means clustering are inherently related, with error-minimizing clusters equal to a transformation of the principal components.¹ In fact, the block-diagonal structure in the covariance matrix is indicative of clear clusters in the data.

12 Describing the Clusters - Maps

- Walkthrough the results - what are the takeaways of these maps?
- Notice how the adding clusters further subdivides, but does not shuffle cluster membership. This means there are stable clusters being formed.
- The metro areas and coasts separate themselves out fairly quickly.
- The south comes in several striations.
- The northern and western agricultural areas distinguish themselves as a stable homogenous region.
- Only with higher numbers of clusters do we see heavily hispanic areas of the country distinguish themselves.
- Health and Econ/Demo vars produce similar clusters, so it makes sense to lump them together for making the clusters for the regression analysis.

13 Basic OLS results - OLS Regression Map

- We see the Trump effect pretty much everywhere.
- Only when the clusters get small does the Rio Grande Valley emerge as showing a positive relationship between Trump voting and social distancing.

¹As noted here: <https://dl.acm.org/doi/pdf/10.1145/1015330.1015408>

- The effect is larger in metro/coastal areas, places where Trump voting is lower overall, and where standing out as a Trump-voting enclave means placing yourself in opposition to the opinions of your relatively near neighbors. Moreover, given the clustering strategy, counties in this cluster that voted for Trump were less likely to share average lifestyle characteristics that aligned with the bulk of Trump-voting counties in the country. Instead, they are more likely to share lifestyle characteristics with urban democratic voters. The fact that these counties were less likely to participate in social distancing may reflect the need these voters feel to establish their identity by acting on Trump’s messaging, so as to disassociate themselves from their immediate peers.
- The effect is also larger in the deep south, a band of counties that has a checkerboard pattern of Trump voting.

14 Individual County Effects Using Machine Learning Method

-

15 Explaining the Patterns: Higher Trump Effects in Contested Localities

Tableau Dashboards in Order

1. Intro Paragraph

On April 15/16th, small groups began to assemble outside the Michigan capitol building, simultaneously armed with guns, prophylactic masks, and a misunderstanding of the gravity of the public health crisis enfolding them.

Shortly after, on April 17th, President Trump tweeted out his support for the aims of these groups - issuing the rallying cry to liberate Michigan, Minnesota, and XXX from the tyranny of public health measures imposed by their liberal state governors.

Meanwhile, Gov. Cuomo of New York was gaining significant public plaudits from another slice of the populace for his series forthright press events that were clearly motivated by input from the scientific medical community and public health experts. Far from aiming at the emotions of an aggrieved set of voters who are clearly receptive to anti-elite sentiments, Cuomo's intent was clearly to provide understanding and context for the various restrictions his government was imposing.

Even much earlier than this, it was clear that there was a divide apparent along political lines between elected officials in how to understand and administrate the approaching pandemic. However, the onset of protests began to crystallize the notion that this divide might manifest itself in the actions of citizens of different political stripes.

Divides in political affiliation have analogs in media consumption - those digital bubbles within which we've understood we're trapped for at least the past four years - with conservative and liberal media outlets presenting different information, or the same information differently, with different points of emphasis, different levels of concern, and different takeaways regarding, in this case, preparation and public health.

Therefore, there is reason to suspect that liberal and conservative citizens may exhibit different behaviors during the COVID epidemic, due to different mental models primed by the information presented to them about the nature of the disease and the appropriate safety measures to take in response.

However visible, these political-behavioral associations might be more in the nature of casual associations, the ready caricatures of a stereotyping mind, than a broad look at the data might suggest. The appearance of armed protesters with protective masks and often confined to their cars, suggests a level of caution even among those who are chafing at that the constraints imposed by the demands of public health authorities.

Moreover, while appearances suggest the protesters come largely from the right, apparent relationships don't always bear out in the data, and the small group of people motivated to protest is probably less important (de-

spite the possibility of super-spreader events) from a public health standpoint than the overall behavior of the bulk of the population.

So, has media messaging had an observable effect? Do conservatives demonstrate less social distancing response than liberals? To answer this question, I'll run through some descriptive statistics, a regression analysis, and some machine learning techniques, to show us how areas that supported Trump may have differed in their movement levels relative to those that don't support Trump.

2. Social Distancing Metrics and Early Spread of Coronavirus

You've probably seen a zillion and one coronavirus maps in the past 12 weeks, and, yes, I definitely have more for you. To start, let's look at the measured spread of COVID and the adoption of social distancing behaviors over time.

*** MAP GOES HERE ***

This first map shows a few variables which describe the spread of Coronavirus as well as social distancing metrics over time, and it allows one to begin to correlate in their mind the severity of viral spread and the social distancing actions taken in response to that spread with the political map. Coronavirus infection data is from the NYTimes Github repository, and social distancing metrics are sourced from Google and Descartes Labs.

We can see that the spread of the virus wasn't particularly time-varying. Yes, it arrived to the coasts, and Colorado, first, but there wasn't a significant amount of time before it had made itself known as a threat in most parts of the country. It does not appear that there was a great time difference between its widespread infection in any one place and its initial appearance in another, and no region of the country has been spared of coronavirus infection.

What this means is that there was not a great deal of latitude for one region of the country to learn from the experiences of another part of the country, and thereby devise ways of slowing the virus without requiring distancing. So, to the extent that people took coronavirus seriously as a public health threat worthy of avoidance, distancing was likely the only avenue.

3. Let's Make a Model

To begin building up the regression model, let's first discuss the policy variable - media consumption. Unfortunately, I don't have a broad, geographically precise and sufficiently disaggregated measure of the types of media being consumed by a broad swath of individuals.

This data might exist, and an interesting add-on to this analysis would be to use something like Google Trends to correlate searches for misinformation keywords with social distancing behaviors.

Instead, I'll be using the percentage of citizens in each county who voted for Trump in 2016 as a proxy for media consumption. This has advantages in that it is a fairly broad survey of opinion, and has a strong correlation to media consumption. While it isn't a direct measure, and suffers from degradation over time as a measure of current political opinion, its availability in terms of geographic detail makes it an appealing stand-in.

If you're uncomfortable making the leap to media messaging, the use of Trump voting answers a slightly different question more directly: how much of the variation in social distancing is driven by political affiliation?

4. Controlling Confounders - also show Demographics, Economics, and Health Data

Social distancing behavior in total is the result of many different factors, forces, and circumstances. The situations people are embedded in do a lot to constrain their ability and willingness to forego public activity regardless of the messaging from their political officials and media outlets.

For instance, the more educated are in general more healthy, more wealthy, and may be more receptive to the guidance of scientifically-oriented public health authorities. They are also less likely to vote for and listen to Trump. Wealthier people, by virtue of the ability to store goods, hire delivery people, hole up in larger homes, and work remotely, are much more able to undertake social distancing.

All of which illustrates the analytical task - how do we isolate out the relationship between media messaging from other factors related to both media consumption and social distancing patterns?

*** MAP GOES HERE ***

This map shows the geographic distribution of a large set of variables, which you can peruse through, that show distinct regional patterns and which are likely to be relevant to either health behaviors, political behaviors, or both. These are the data that we'll use as controls in the regression analysis, and the patterns they show demonstrate their utility - for many of the many of the variables, the geographic distributions strongly resemble the political map. Adding these variables as controls also accounts for some variation from geographically varying factors. Using these variables as controls will automatically mean that we are forming comparisons among regions that are close in geography as well as accounting for constraints on the ability to restrict movement - or in underlying associations between political choices and social distancing behaviors that have nothing to do with media messaging.

Public health institutions may influence social distancing behaviors, and may be correlated with political choices through a common underlying factor or a mechanism such as funding priorities. So, including broad indicators of public health such as chronic disease incidence allows us to control for the state of these public health institutions.

Likewise, for economic and demographic data - not all jobs are equally able to social distance, and the composition of jobs varies by region in ways that may be connected to voting and media consumption. Common factors may underlie this association and are important to control.

It's also just cool to look at the distribution of the variables.

5. Regression

The outcome variable we're interested in predicting is the level of social distancing behavior. I measure the amount of distancing using data publicly provided by the movement tracking firm Descartes Labs, which has made available time series data on the average (?) movement at a county level beginning on February XXth. Descartes Labs generates its estimates using (???), and is available at **Github Link**

I process these time series into a single estimate of the amount of social distancing undertaken by a county. Collapsing the history of social distancing into a single metric allows me to control for timing in the arrival of the virus and government shut-down orders, as well as to smooth day-to-day noise in the metric.

I generate this estimate by first calculating a date at which social distancing began in earnest in the county (I use the date when movements first dropped below 40% of their pre-COVID average). To calculate this date, I first take each time series, and generate a smoothed version by applying a flexible spline function. Finally, I take the average movement after the initiation date through the end of April as the metric for the degree of social distancing behavior in each county.

*** Regression chart goes here ***

This chart plots the regression coefficients (on the vertical) against their precision (the inverse of their standard error, on the horizontal), and colors each observation by its p-value. The dropdown allows you to select different sets of included variables, to see how the inclusion of alternative information effects the analysis. The Trump coefficient changes little throughout various specifications.

What really stands out, is that the Trump effect is large and significant even when including a Republican voting factor which describes the average Republican vote over the *ELECTIONS*. This means that the relationship between political partisanship and social distancing is not simply related to a deeper conservative ideology, but appears to have a durable and powerful relationship to a specific ardor for Trump.

Given the large set of control variables and the lack of a causal methodology, we might well be concerned that the high Trump effect is the result of multicollinearity - the model may spit out high coefficient values for Trump as an offset to high coefficient values on other variables in a manner similar to an opponent process. Using a Tikhonov Regularization to

discipline the model's selection of high (absolute) value coefficients, we find that Trump's influence stands out persistently.

6. Clusters of Counties by Recent Public Health History

With our control variables, inherently we're asking the regression to form comparisons between places that are similar in as many dimensions as possible - other than Trump voting. To aid that, we'll follow a strategy similar in spirit to matched pairs or propensity scoring. Except instead of propensity to receive a treatment, because there is no "treatment" here - no way of changing political beliefs or media consumption, we'll use k-means clustering to group observations into similarity families.

This way, we're reducing the burden on our linear regression model to adjust dissimilar counties through control variables. If the regression model is trying to compare sets of observations that are very dissimilar, then multicollinearity issues are more likely to arise - its harder to pick a single reason to explain the differences in social distancing when there are many differences among counties that could give rise to the difference.

Observations that are similar on observable characteristics are more likely to be similar on unobservable characteristics, if the unobservable characteristics are correlated with the observable ones. Moreover, if they are correlated, then the observable characteristics will be good controls.

*** MAP GOES HERE ***

This map shows the results of clustering counties by health characteristics related to chronic diseases, using k-means clustering. As you increase the number of clusters, you'll notice identifiable geographic regions stand out - the north vs the south initially, then the deep south vs the bible belt and the upper midwest/mountain west, and eventually finer gradations of the south. Metro areas distinguish themselves after 5 clusters.

Another related way to visualize the strength of trends in these behaviorally-mediated health variables is using PCA. Factorization of the health data suggests there are a small number of underlying constructs that explain the many health variables used in this analysis, which in turn suggests that clusters formed on these data will separate reliable and meaningful groups, and that a small number of clusters is supported. In fact, you can pretty much read the structure of variation directly from the covariance matrix.

*** Health Covariance Matrix Goes Here ***

7. Clusters of Counties by Demographic Vars

The process of clustering only on the demographic data produces similar results to the health data, which is not entirely surprising, but it is good to find that it aligns with expectations. Interpreting the PCA factors formed on the demographic variables is not quite as visually direct, however, the components clearly identify relationships between wealth, race,

ethnicity, employment composition, population density, and so on that broadly conform to political characterizations.

*** Link to Map in (clustering only on the demographic data) *** ***
Link to Cov/PCA in (components clearly identify relationships) ***

8. OLS Regression Map

*** OLS REGRESSION MAP HERE ***

This shows the coefficient on Trump voting as we tighten the analysis in on narrower and smaller clusters of counties. In nearly all cases, Trump voting is associated with greater movement - that is, less social distancing.

Only when our number of groups is high enough to produce relatively low-observation clusters do we see that some of them produce Trump effects that go in the opposite direction.

On inspection of their geography, we can see that these opposite direction clusters are those that cover the Rio Grande valley, a heavily hispanic area. Here, with a relatively small number of counties and a high hispanic contingent, we see that the model indicates that more Trump voting is associated to higher social distancing. Given the consistency in Trump voting effects in other regions, it is possible that the model is having difficulty with multicollinearity - heavily hispanic districts might be associated with low social distancing in these regions for many reasons other than their media consumption, especially compared to the whiter counties in this cluster, who have a greater degree of social distancing and a higher Trump voting percentage. It may be that these whiter counties are wealthier and therefore more able to socially distance, but that the model is unable to distinguish between wealth, ethnicity, and voting patterns with a paucity of observations.

- How do the hispanic and wealth coefficients look, in comparison to the Trump coefficient?
- What happens if I do a ridge regression?

The regions that show the greatest effect of Trump voting on their social distancing behavior are those that are in or near densely populated metropolitan areas, or are on the coasts. It appears that standing out as a Trump voting county among these blue-enclaves is associated with a greater degree of action related to messaging about the lower severity of the virus coming from Trump.

Another region which has greater association of Trump voting and lax social distancing is a band that runs through the deep south, which is mottled with counties that narrowly voted for or against Trump in 2016.

9. Individual County Trump Effect Estimates

The OLS regression coefficients produce a single estimate per cluster, which is the average partial effect of Trump voting on social distancing

behavior across all counties in the cluster. To get individual estimate of the effect of pro-Trump sentiment on each county, we can use a non-linear machine learning model.

Since I have no *a priori* hypothesis as to the functional form of the non-linearity that should be applied, I employ a flexible machine learning method to build a mapping from the input variables to the social distancing values. Using this mapping, I can then produce an estimate of the Trump effect via the gradient of the model with respect to the Trump voting percentage.

Using an ensemble of ANN's as my estimator, the estimates for individual counties fall well within the ranges that we saw for the OLS estimation, suggesting that the ML estimator has good consistency with the appropriately sub-grouped OLS estimator.

**** Individual Effect Map Goes Here ****

While this method of estimation does not allow classical inferential tests of statistical significance, we can get a sense of the robustness of each county's Trump effect in a Monte Carlo fashion. If we make perturbations in the Trump voting percentage for each county, and estimate the gradient resulting from each perturbation, we will produce a range of coefficient estimates. If these estimates are stable across the range of perturbations, then we know that the method is not producing knife-edge results based on an overfit model.

With neural net methods, one concern is that the high number of parameters allows the model to identify observations on numerical means that have no bearing on the counterfactual question - what would have happened if the Trump vote were different, and people were less likely to be tuning in to low-risk messaging? When the gradient of the model is consistent across a range of counterfactual scenarios, then we have greater confidence that the model is producing meaningful information about counterfactual scenarios.

Estimating the counterfactual scenarios in this Monte Carlo fashion for each neural network in the ensemble also allows us to produce an estimate of the variation in the counterfactuals across models. When this variation across models is large, then our predictions are based on a set of models that are each overfitting the data in a different way - and while the average of models across the ensemble might still produce a good fit, we have less confidence that our model is valid out of sample, where our counterfactual scenarios inevitably lay. However, where the variation across models is small, the data is suggesting strongly that each model takes the same opinion with respect to the effect of Trump voting on social distancing behavior. That each component model would do so even though there is significant freedom of parameters to fit the data however it might choose suggests a reliable relationship between Trump voting and social distancing behavior.

In the above map, hovering over a county shows the counterfactual change in the percentage of Trump voters in 2016. The bar heights show the magnitude of the Trump effect on social distancing behavior (averaged across the ANN's in the ensemble) while the bar colors show the precision of the Beta estimates.

The map shows counties containing downtown cores to have negligible effects of Trump sentiment on their social distancing behavior, while suburban and surrounding counties have much larger effects. This indicates the model is isolating outliers, as high density urban cores show low support for Trump and have no effective analog as a basis for behavioral comparison, while suburban counties have a range of political opinions to compare between. Effect sizes for these suburban counties are somewhat smaller than found for the OLS cluster regressions, suggesting that the OLS results were in part driven by the comparison of urban and suburban counties, but that the suburban effect persists when the model is able to set urban counties aside.

10. Explaining the Patterns: Higher Trump Effects in Contested Localities

Visual inspection of the patterns of Trump effects from the ML model and the electoral map seem to suggest that the effect is strongest in areas of the map that are speckled with red and blue - areas in which ideas and preferences are likely to come into conflict, and in which the need to hold fast to one's beliefs, or to immerse oneself in a friendly media bubble is greater.

- Is there any research to draw on about how people hold tighter to their beliefs when they are challenged, rather than when they are around people who agree with them?
- Is there any research to suggest that people are more likely to demonstrate their identities through action when they are challenged than when they are simply allowed to have them without standing out?

It appears that the greater the degree of heterogeneity of opinion one was/is likely to face, the greater the effect of Trump's messaging became. To test this, I devised two metrics to capture the relationship between local heterogeneity of opinion and the social distancing effect.

*** SD Scatter Plots Here ***

The first is the average of the absolute value of the difference between a county and its immediate neighbors - adjacent counties. This captures the raw level of difference in opinion in a regional manner. However, the degree of difference between a county and its neighbors may not reflect the real interpersonal interactions that would be driving the need to more strongly identify with one's beliefs in the face of opposition. Thus, I created a second metric, which is the average absolute difference of a county with its neighbors multiplied by how close the county itself was to

a 50% vote in 2016 (this second metric, as an equation: Avg. Abs Diff * min[Trump% '16, 1-Trump% '16]). For both metrics, the Trump effect on social distancing behavior is stronger when political disagreement is greater.

As a final demonstration, I plot the two inputs to the second metric on independent axes and divide the resulting area into cells, with each cell showing the average social distancing undertaking by counties in that cell.

*** Insert Heat Map Here ***

Not only is there an observable effect of political partisanship on social distancing behavior, its evident that it is widespread both across the country and within broad population segments - not merely the active and vocal but small groups of protestors showing up at state capitols. Moreover, this effect seems to be the greatest where partisanship is the most contentious and where people are most likely to come into conflict with or have to negotiate with those who hold differing opinions. And while the effect of media selection cannot be conclusively, causally demonstrated here, it is strongly suggested by the fact that it is affection for Trump, not simply a place on the historical political axis, that influences risky action in the face of a global pandemic. Others have suggested that non-partisan messaging is critical to get around this divide, but it remains unclear what kind of public messaging is resistant to the influence of partisanship in 2020.

FUTURE

1. Difference in Differences using Before/After State Lockdown, county FE's, Time FE's, and interactions of before/after with Trump Vote and Republican Vote.
2. Produce the following $SD(t) = B1(t)*Trump\ Vote(t) + B2(t)*Repub.\ Vote(t) + B3(t)*Trump\ Vote(t)*my\ Polarization\ Metric + County\ FE + Time\ FE$ Accomplish the Diff-in-Diff feature by shifting the time indicator to be relative to an event.
- 3.
4. Add my Political Polarization metric to the OLS regression, and interact it with both Trump and Republican factors.
5. Do the Diff-in-Diff but replace Trump Vote with Google Trends searches for polarized terms.
6. Get a better polarization metric (later)

$$SD_t = \sum_{c \in C} 1[County = c] + \sum_{t \in T} 1[T = t] + \{\mathbf{Variables}\} * \sum_{t \in T} 1[T = t]$$

$$\{\mathbf{Variables}\} = \{Trump\ Vote, Health, Economics, Demographics\}$$

$$T = \{0 \text{ if Date } \leq 3/16/2020, 1 \text{ Otherwise}\}$$