Spring 5-1-2021

The Impact of Information Shocks and Partisanship on the Evolution of COVID-19 in Connecticut

Joslin Valiyaveettil
joslin.valiyaveettil@gmail.com

The Impact of Information Shocks and Partisanship on the Evolution of COVID-19 in Connecticut

Joslin Valiyaveettil

Thesis Advisor: Dr. David Simon

Honors Advisor: Dr. Richard Langlois

University of Connecticut

Honors Program

Department of Economics

Abstract

The COVID-19 pandemic has affected all aspects of life within the United States since early 2020. How people decided to behave during this time heavily influenced the trends that followed, triggering both health and behavioral economic concerns. Those trends seemed to vary based on the area and the beliefs of those constituents. This paper explores how partisan beliefs had an impact on the changes in case rates that occurred within the top 30 most populated towns in the state of Connecticut. In July 2020, former President Donald Trump sent out a tweet publicly endorsing face masks for the first time. This paper studies the effects and trends on case rates a month after that information shock across those top 30 towns. The initial assumption was that in democratic areas, an information shock coming from a republican source would either have a negative or no significant impact. However, the results show that the more democratic an area was, the tweet had a larger effect on decreasing rates. Although there are many other factors that can help explain this result, the research here suggests that information shocks can be very impactful in areas of high partisanship.

INTRODUCTION

The COVID-19 pandemic has been at the forefront of society for over a year now. In March 2020, state governors across the United States slowly began to implement lockdown procedures. Schools, companies, small businesses, churches, restaurants, and other institutions began to close as the country went into quarantine. COVID-19 cases and fatalities began to rise as time went on, and research ensued to identify the factors that could possibly be influencing them.

COVID-19 data and research identify political attitudes and partisanship as important elements of the discussion. Partisanship was already growing strong before the rise of COVID-19, but as cases started to spike, the divide grew and manifested itself into how citizens behaved during the pandemic. This is an important health and behavioral economics issue, where the actions of groups of citizens are affecting the wellbeing of society as a whole. The risks taken, particularly during a global pandemic, can have implications that go beyond those predicted by previous, pre-COVID 19 economic models. Furthermore, looking at how people perceive certain risks can help economists understand the reasons behind certain behavioral trends. Subjective beliefs, stemmed from strong partisan views, do tie in as one of those reasons. Recent data also shows how misinformation was a prominent issue during the pandemic, influencing the behaviors of its consumers. Especially in the early stages of the pandemic, faced with mainly uncertainty about all things COVID-19 related, media and news outlets tended to give skewed perspectives of the events occurring, and thus downplaying the severity of the growing COVID-19 virus. But it was not just the media, but political figures as well around the world. Those in roles of leadership are prominent influences on the behavior of their constituents. During the time of a global pandemic, the direction that leadership decides to take will determine the outcome within that area, and even outside it.

This paper investigates whether there was an impact in COVID-19 rates within the towns of Connecticut specifically from certain information shocks. Through the vast information available about the pandemic, it is clear that politics, political figures, and partisan media all had an influence on the COVID-19 pandemic. This paper decided to see if certain political media information, in this case those seen to be a major headline news at a certain point in time, could have been a factor in any changes in COVID-19 case rates. And furthermore, what if that political media shock was put out by the head of the nation himself, the President?

The information shock that is tested in this paper is a tweet issued by Former President of the United States Donald Trump. This tweet was posted on the social media network Twitter on July 20, 2020, regarding his endorsement of face masks. Posted on July 20, 2020 at 3:43pm, former President Trump tweeted the following: “We are United in our effort to defeat the Invisible China Virus, and many people say that it is Patriotic to wear a face mask when you can’t socially distance. There is nobody more Patriotic than me, your favorite President! https://t.co/iQOd1whktN” (Brown, 2020)

In relation to background prior to this event, President Trump, though never explicitly stating that masks were bad, did state that they were voluntary. He expressed how he would most likely not be wearing one and was not seen wearing one publicly until July 2020. (Mason, 2020) After this tweet was posted, it was picked up by almost all news networks and public information sites, sparking a nationwide discussion about this event. This tweet was the strongest endorsement the President had yet provided for publicly using face masks. (William, 2020)

There were many articles being published that day and week discussing the former President finally endorsing masks verbally.

Literature Review

Before conducting the data portion of the research process, background research was necessary to recognize what correlations were already being made about the COVID-19 pandemic. It must be recognized that the earliest studies, articles, and papers that were released about the pandemic were only able to utilize the information available at the time. The ones released towards the end of 2020 were able to include more of the research that came out in the later stages of the pandemic, some even disproving certain ideas released earlier.

The purpose of this paper overall is to explore any possible correlations between political ideologies of the public and the pandemic itself. Many questions were raised: Did political identities impact the changes seen in a town's COVID-19 rates? How much influence did a state’s governor have on COVID-19 rates? Do different views on social distancing make a difference? These preliminary questions helped to sort through the research available at the time (economic research released prior to January 2021) and narrow in on a more specific question.

Initial curiosity began in May 2020 with seeing how political the pandemic was becoming. Many papers were also being released which saw, through polling, how the COVID-19 pandemic was becoming a partisan issue. One of those was the paper “Partisan Differences in Physical Distancing are Linked to Health Outcomes during the COVID-19 Pandemic,” published in the Nature Human Behavior journal.

The published article gives an overview on how tensions between both political sides had been running high long before the pandemic. There has been a growing distrust between both the Democratic and Republican parties. But furthermore, politics do have a significant impact on society and the behaviors of its citizens. The source of one's media outlet is also important because different stations can be communicating very different outlooks about health issues, in this case the global pandemic. As taught in behavioral economics, people are not always rational, and their behaviors are heavily influenced by the environment around them. When that idea is used to analyze this published paper, we can see how crucial a person's information sources are in terms of how they will behave in the midst of a pandemic.

The article found that "US counties that voted for Donald Trump (Republican) over Hillary Clinton (Democrat) in the 2016 presidential election exhibited 14% less physical distancing between March and May 2020." (Gollwitzer, Martel, Brady, 2020) The timeline used in this study was between March 2020 and May 2020, the very early stages of the pandemic and country-wide lockdown regulations. The study calculated its findings through measuring the reduction of movement of the population in general, and to non-essential services.

All movement drastically decreased in early March, but that reduction becomes less and less as the timeline reaches May. The study linked partisanship to physical distancing, physical distancing to COVID-19 case rates and deaths, and both partisanship and physical distancing to case rates and deaths, all in order to get a better understanding of what was happening. The study looked into numerous other factors as well, including the areas COVID-19 case rates, median income, and racial and age demographics, and yet partisanship was still the most strongly linked with physical distancing. (Gollwitzer, Martel, Brady, 2020) The voting gap from the 2016 election was used to determine the partisanship among different counties, and those counties that voted Republican (for former President Donald Trump specifically) displayed less physical distancing from March to May 2020 than the counties that voted for Hillary Clinton. This use of voting data from the 2016 election in the published study is also a reason why it is used for this paper and its research as well, since it allows us to look at people's political views before they may have changed as the pandemic went on.

The study in the published article first hypothesized that the differences between the counties would decrease as the pandemic got worse in time and even those counties that were reluctant at first would cooperate with physical distancing measures. However, the opposite was seen, and the partisan gap actually increased. Reasoning for such differing attitudes were also attributed to different states' stay at home orders, and partisan media intake. The overall conclusion from the article's data suggested that the growing partisan outlooks influenced American citizens' attitudes towards the pandemic. Since partisanship had such an influence in relation to physical distancing, it begs the question of whether other partisan tools may have impacted COVID-19 rates as well, which is what we will look into in this paper.

Another important political and economic issue that arose during the pandemic was the spread of misinformation. It is known that media outlets present information from sharp perspectives. More right and/or left leaning sources in the early stages of the pandemic faced the growing uncertainty by informing their audiences with skewed takes. One working paper that studied this assessed whether conservative media had an effect on the behavior of its viewers. The paper and its study saw the two most watched cable news shows in all of the U.S, airing on the same Fox News network, resulting in calculable differences in the health of its own consumers. *Tucker Carlson Tonight*, in comparison to the other show *Hannity*, "warned viewers about the threat posed by COVID-19 from early February, while Hannity originally dismissed the associated risks before gradually adjusting his position starting late February" (Bursztyn, 2020) The *Hannity* show was described to be much more dismissive in its overall coverage of COVID-19 in the early month, with its content often downplaying the events happening nationwide and worldwide. Though the terminology and tone used by both networks became more similar by April 2020, the effects of the prior contrast in content could be seen in the difference in behaviors of its viewers. Specifically, the study's survey data revealed how Carlson's viewers changed their behavior in response to the COVID-19 virus days earlier than Hannity's viewers. With further research into COVID-19 case rates and mortality rates, the paper concluded that misinformation is a driving factor in COVID-19 effects. These results strengthen our assumptions that media does have a significant impact on behavior, and that partisan media influences individual actions that can collectively control the progression of the pandemic.

Risk behavior is another important factor, impacting development economics as well. When analyzing the behavior of a population, one must also look into the different risks that they are facing, and how external factors are influencing that public's risk perception. The hypothesis of this paper looks into whether a political leader's information shock impacted a group, and that is similar to the research conducted in Brazil in the working paper "More than Words: Leaders' Speech and Risky Behavior During a Pandemic." After seeing that the Brazilian president was dismissive of COVID-19 and measures associated with it, the study found that areas with less political support for the President had much stronger coronavirus safety measures, with larger impacts to the containment of cases. Evidence suggested that the impact was also influenced by whether the area was more media receptive, including news outlets and social media. While preventive measures were decreased in some areas (those that supported the Brazilian president), risky behavior increased within them. During the peak months from February to April 2020, travel in those areas were much higher, going to show how seriously supporters of the president were taking his statements dismissing the virus and its severity. (Ajzenman, Cavalcanti, Da Mata, 2020) This analysis show how severely harmful it is to that society, now much more susceptible to the growing fatalities approaching their area. Those in powers of leadership matter and have a dominating impact on their constituents and their resulting behaviors. Correlating this information to the US, it is concerning to see how a group's perspective on the severity of a potential risk can be dramatically changed from the words of a higher figure that they support. What is generally accepted now as risky behavior can be prevented with the right leadership viewing it that way early on, otherwise the actions of the leadership in charge can have devastating effects on its people.

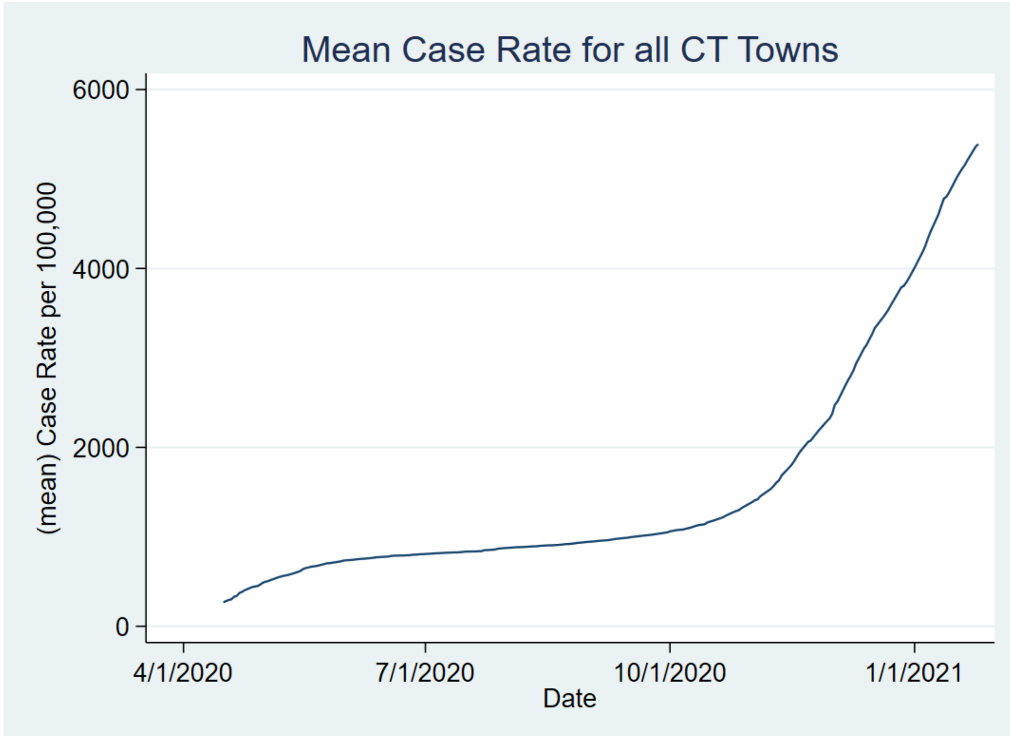
Another paper, exploring the political implications on the pandemic, went a step further to determine associations between governor political affiliation and COVID-19 cases and deaths. This article conducted an analysis investigating COVID-19 cases and fatalities for all US states and compared them with the governor's political affiliation. In the first literature review paper, we saw how more republican counties displayed less physical distancing as time went on (between March to May 2020). However, the interesting results of this article (published in November 2020) is that during the timeframe of March to June 2020, "Republican-led states had fewer per capita COVID-19 cases and deaths early in the pandemic, but these trends reversed in early June (for cases) and in July (for deaths). These early trends could be explained by high COVID-19 rates among Democratic-led states that are home to initial ports of entry for the virus in early 2020.However, the subsequent reversal in trends to Republican-led states may reflect policy differences that could have facilitated the spread of the virus" (Neelon, 2020)

It is crucial to recognize that there are many other factors that are impacting these COVID-19 statistics and trends, such as the democratic cities being places with higher populations, making their COVID-19 rates much higher than others. However, Republican governors were seen to be slower to adopt stay at home orders and issues mask mandates. This same outlook will be applied to this research paper as well, how external actions conducted by political figures could influence case rates. It is important to recognize the limits of the published study, including the fact that this analysis was done at the population level rather than the individual level, and since it is an observational study, it cannot infer causality. Another limit of the article, and one of this paper as well, is that governors and presidents are not the only authoritative actors within the state and country. Rather, there are many other legislative entities passing law-enforcing mandates concerning the pandemic, and thus also influencing the COVID-19 case and death rates.

Data Utilized

All covid-related data in Connecticut that was used in this research was retrieved from the Connecticut Open Data website. This archive led us to Connecticut's daily COVID-19 cases, hospitalizations, deaths, and tests from March 2020 to as recent as January 2021. Regarding this research, the CT Data sheet chosen listed COVID-19 Tests, Cases and Deaths by town. (Department of Health, 2020) This displayed the daily case rates (per 100,000 people) for all 169 towns in Connecticut.

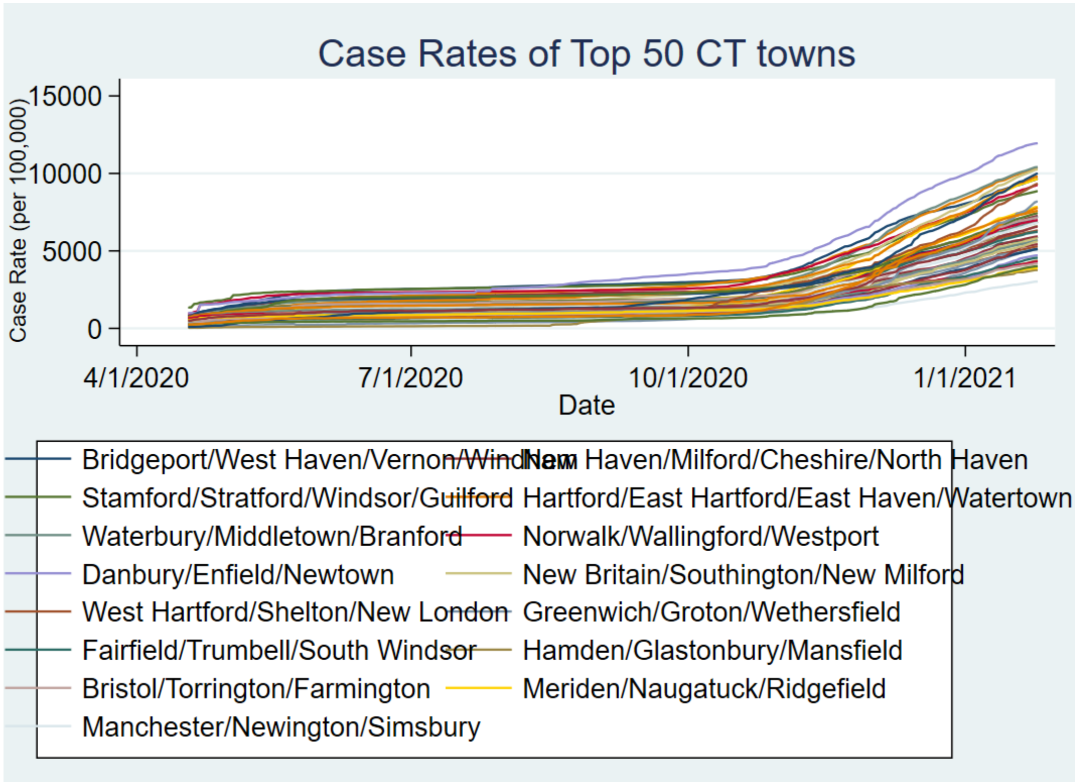
After downloading the data and organizing it, a data set was created that listed the daily case rate (per 100,000 people) for each town, and, for the purpose of this research, starting after April 1, 2020. This was done in order to create a graph that simply showed the average rate of cases (per 100,000 people) for all 169 towns in Connecticut. This gave a good starting point and an overview of what the case rate looked like on average throughout the entire state.



Simply looking at this graph, showing the average 169 towns, does not give much insight into what dips and peaks occurred, let alone into the specifics of what went on at the town level.

After looking at this graph of mean case rates, it was decided that the data set should be decreased in order to only look at the top 50 most populated cities in Connecticut. The Connecticut Demographics Report used population information gathered from the 2019 United States Census Bureau and ranked the data of all 169 towns by population. (Connecticut Demographics, 2020)

Using the top 50 towns in Connecticut, another graph was created to help see the case rates and their fluctuations better on an individual town level.

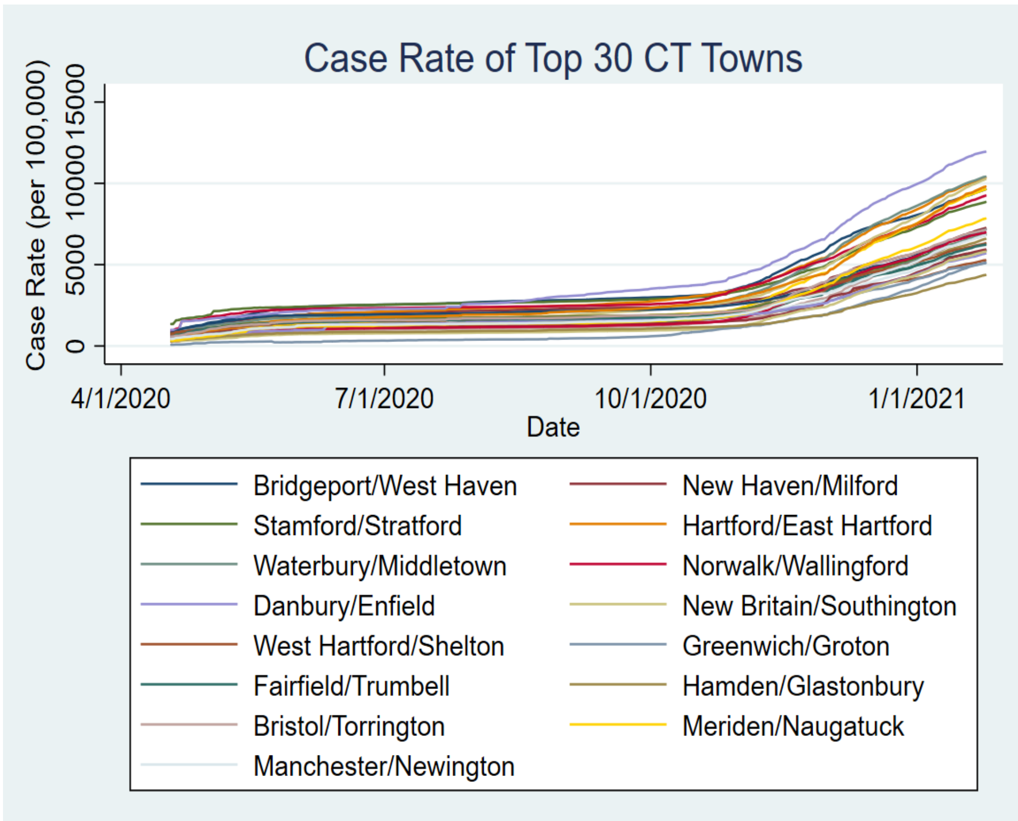


We can see more variation between the towns with this graph, especially as we get closer to the more recent dates. We are also able to determine how many towns shared such similar data that they could be graphed together.

However, the data is still too crowded for this research. The town size was decided to be narrowed down even further to the top 30 most populated towns in Connecticut. The towns were once again retrieved from the 2019 Census report and included the following towns, listed in the order of most to least populated:

Bridgeport, New Haven, Stamford, Hartford, Waterbury, Norwalk, Danbury, New Britain, West Hartford, Greenwich, Fairfield, Hamden, Bristol, Meriden, Manchester, West Haven, Milford, Stratford, East Hartford, Middletown, Wallingford, Enfield, Southington, Shelton, Groton, Trumbull, Glastonbury, Torrington, Naugatuck, Newington.

The case rates for the top 30 towns in Connecticut, between March 2020 and January 2021, can be seen here.



Our hypothesis looks into how and whether information shocks could be seen to influence these case rates. Specifically, we investigate former President Donald Trump's tweet on July 20, 2020, verbally endorsing masks in public for the first time. (Brown , 2020)

In order to analyze the political effects, election results from the 2016 presidential election were used. The official statewide statistics per town is archived on the Connecticut Secretary of the State's government database within Election Night Reporting. (Connecticut Secretary of State, 2020) For the top 30 towns that we narrowed in on, the number of votes from each town casted for candidates Donald Trump and Hillary Clinton are listed. The sum of both is then recorded to use as the total votes casted. For the purpose of this research, we decided to look at the democratic vote share in Connecticut within those top 30 towns. To do that, we found the percentage of the total voters from each of the top 30 towns that voted for the Democratic candidate Hillary Clinton. The data can be seen here, listed in order of lowest to highest democratic vote share:

Vote by town	Clinton	Trump	Total Votes	Dem vote share
Torrington	5,713	8,673	14,386	40%
Shelton	8,001	12,051	20,052	40%
Naugatuck	5,219	7,310	12,529	42%
Southington	9,890	12,383	22,273	44%
Enfield	8,646	9,233	17,879	48%
Trumbull	9,299	9,753	19,052	49%
Wallingford	10,651	10,940	21,591	49%
Bristol	12,499	12,752	25,251	49%
Milford	13,598	13,383	26,981	50%
Newington	8,425	6,557	14,982	56%
Stratford	13,729	10,534	24,263	57%
Danbury	16,084	11,626	27,710	58%
Groton	8,453	5,936	14,389	59%
Greenwich	17,630	12,215	29,845	59%
Glastonbury	11,074	7,533	18,607	60%
Meriden	12,788	8,660	21,448	60%
Fairfield	18,041	12,112	30,153	60%
Waterbury	19,870	12,837	32,707	61%
West Haven	12,477	7,774	20,251	62%
Manchester	15,109	8,358	23,467	64%
Middletown	12,959	7,126	20,085	65%
Norwalk	24,414	12,324	36,738	66%
Stamford	34,148	16,222	50,370	68%
Hamden	18,962	7,790	26,752	71%
East Hartford	13,180	5,213	18,393	72%
New Britain	15,468	6,055	21,523	72%
West Hartford	23,781	8,055	31,836	75%
Bridgeport	32,035	6,596	38,631	83%
New Haven	35,933	4,540	40,473	89%
Hartford	30,375	2,531	32,906	92%

Empirical Section

We have now been able to properly organize all the different data that we have available to use for our research. From here, we were able to narrow in on a specific topic, and more importantly a specific question to try to answer. The literature review all described an impact of partisan views and politics on the pandemic as a whole, on its case rates, fatalities, social distancing percentages and citizen behavior. Those previous conclusions suggested that republican areas were more likely to react poorly to pandemic control measures. Republican towns saw lower rates of social distancing and increasing COVID-19 case rates as time went on. (Gollwitzer, Martel, Brady, 2020) In that research, physical distancing was the variable being tested. Our initial doubts wondered what other factors could also be tested in relation to partisanship.

With this research, the hypothesis was that such an information shock coming from the President himself, a Republican, would have either a negative or no significant effect in democratic areas. Specifically, we decided to look to find any changes in COVID-19 case rates in the top 30 most populated towns in CT during the time of the tweet and one month after.

This research was a multi-step process to reach the final regression needed to see any possible effects of the tweet.

First, the democratic vote share was found for every one of the 169 towns in Connecticut. This percent showed the proportion of citizens who voted for Hillary Clinton, the Democratic candidate, in the 2016 presidential election against Donald Trump. After condensing the data to only include the top 30 towns, we first ran a simple regression looking at the case rate for each town and the democratic vote share percentage. The regression equation that was tested is the following, where “i” indexes the town and “t” indexes the date, and ε is the error term.

$$caseratePredicted_{i,t} = \alpha + \beta_1 voteshare_{i,2016} + \varepsilon_{i,t}$$

After analyzing that regression, we added in the dummy variable i.date for our date variable, allowing us to condense the large daily data set and rather look at the average across the entire timeline of April 2020 to January 2021. The regression equation that was tested is the following:

$$caseratePredicted_{i,t} = \alpha + \beta_1 voteshare_{i,2016} + \delta_t + \varepsilon_{i,t}$$

However, this still does not tell us enough about our data set and the connection between democratic vote share in the area and case rates specifically during the time of the tweet. We then created the new variable "tweet" to isolate the case numbers to only the time frame of the tweet, July 20, 2020 to August 20, 2020. After we created the new variable tweet_share that showed the interaction between tweet and vote_share during that time frame, the following regression was tested:

$$caseratePredicted_{i,t} = \alpha + \beta_1 voteshare_{i,2016} + \beta_2(voteshare_{i,2016} * tweet_{July\ 2020}) + \beta_3 tweet_{July\ 2020} + \gamma_i + \delta_t + \varepsilon_{i,t}$$

For the final regression, we created the last dummy variable i.townnumber, that allowed us to examine any effects of the tweet at the town level and during the time frame. The final and most important regression tested was the same as the previous, but now with the town data condensed with it:

$$caseratePredicted_{i,t} = \alpha + \beta_1 voteshare_{i,2016} + \beta_2 tweetshare_{July\ 2020} + \beta_3 tweet_{July\ 2020} + \gamma_i + \delta_t + \varepsilon_{i,t}$$

Methodology

With all the data sets that were compiled, they were combined to create one large set. This data set listed the case rates for the top 30 towns, in order of date from April 18, 2020, to January 25, 2021. The democratic vote share for each town was listed as well. Here, we see a brief portion of what that data set looks like:

Date	Town num	Town	Case rate	Dem Vote share
04182020	15	Bridgeport	827	0.83
04182020	17	Bristol	396	0.49
04182020	34	Danbury	993	0.58
04182020	43	East Hartfc	444	0.72
04182020	49	Enfield	414	0.48
04182020	51	Fairfield	415	0.6
04182020	54	Glastonbu	284	0.6
04182020	57	Greenwich	593	0.59
04182020	59	Groton	75	0.59
04182020	62	Hamden	625	0.71
04182020	64	Hartford	482	0.92
04182020	77	Mancheste	371	0.64

The above set goes on for about 6,600 rows to list the data for every date between that time frame of April 18, 2020 to January 25, 2021.

Using the data set of the top 30 towns in CT, two graphs were first compiled to look at an overview of what the average case rate between different percentage of Democratic vote share. The table on page 13 shows the towns listed from the lowest Democratic voteshare at 40%, thus indicating the town with the highest Republican voteshare, to the highest of 70%. The 30 towns were then split into 3 groups 40%-50% Democratic voteshare, 50%-70%, and 70%+. The average/mean case rate of all the towns within that percentage range was computed between the time frame of April 18, 2020 to January 25, 2021. The mean for each group was then plotted on a line graph, as well as a line showing the difference between the average of the lowest group (40%-50%) and the highest group (70%+). The second graph is the same graph as the first, it also includes a line for the average case rate across all 30 towns.

In the first regression ran, it shows the relationship between caserate and democratic voteshare. Our dependent variable is caserate, and the predictor variable here is only voteshare.

“Date” is a categorical variable in our data set, where there are variables for each date. In the data sets that were created, the data was arranged in a way to show the case rate and vote share of each town on each day between April 18, 2020 to the most recent data available at the time January 25, 2021. But that data itself reveals over 6000 rows of data since it shows the case rates for 30 towns on each day within that time frame. In order to comprehend that massive data set and to run a cleaner regression, we utilize the categorical/factor variable of “date” and create the dummy variable “i.date”. Date is the base category that all of the other variables, the case rate and vote share, are compared back to. Creating the i.date variable makes sure the date variable is denoted as a categorical/factor variable.

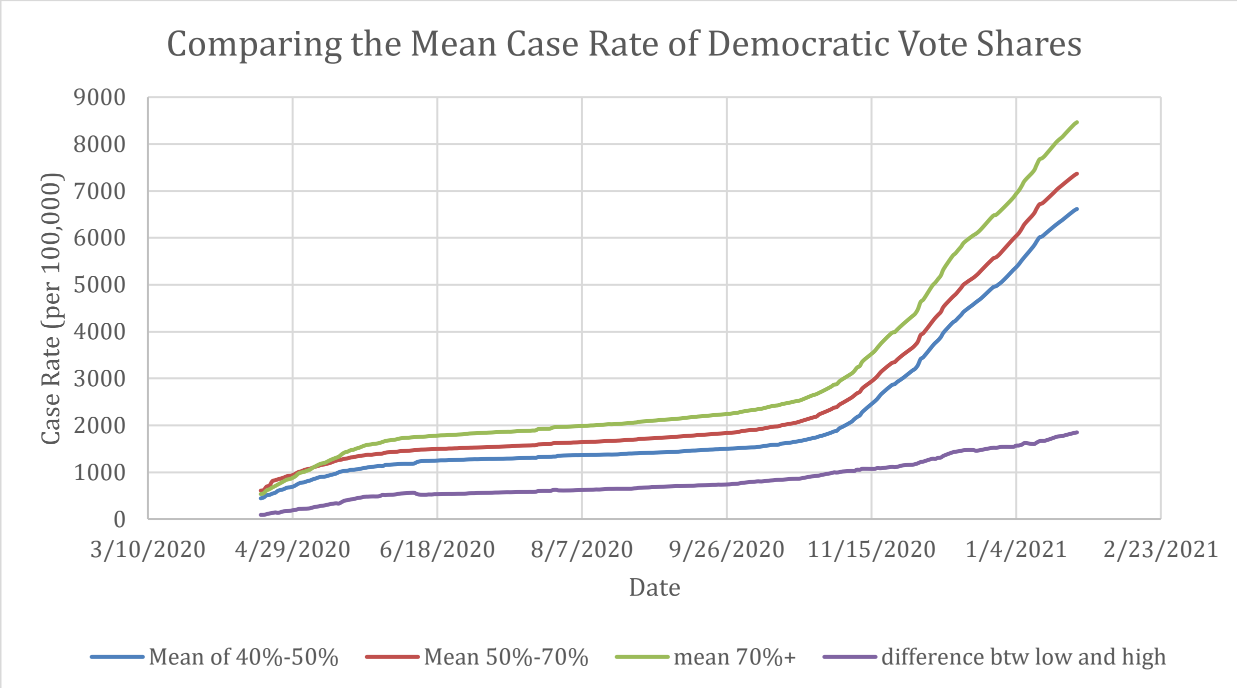
The second regression still shows the relationship between caserate and democratic voteshare, however it now takes into consideration the effects on the averages among the dates provided (between April 18,2020 to January 25, 2021). The regression uses the “i.date” variable to hold the fixed effects for our variable “date”. The dependent variable is caserate, and the predictor variable here is still voteshare.

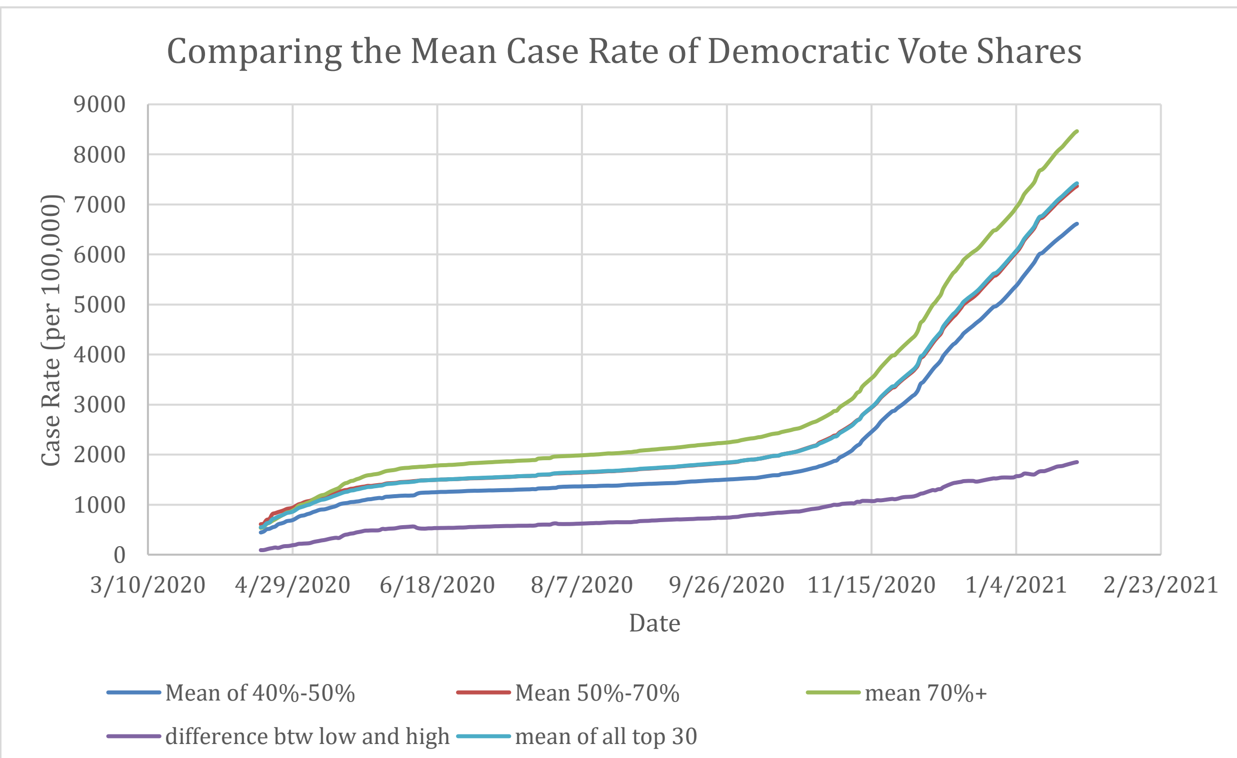
After analyzing the previous regression, the variable “tweet” was created to look into our original hypothesis, the implications of certain information shocks (in this case former President Trump’s tweet endorsing masks). After using the “tweet” variable to isolate the data within the timeframe of the tweet and its effects (July 20,2020 to August 20,2020-one month after the tweet), we created the variable “tweet_share” in order to create the interaction between the data of the tweet and vote share. Using the variable names, tweet_share = tweet (a variable either 1 if it was between July 20-August 20, or 0 if it was not) * voteshare (the democratic vote shares previously calculated from the votes for Clinton divided by the total votes).

The third regression that is ran used the new variables created to look at the dependent variable caserate against the variables tweet_share, voteshare and tweet.

For the fourth and final regression, the new variable “i.townnumber” was created. As done for the variable “date,” “i.townnumber” made the variable “town” a numerical value for each town. We can now further examine the implications of the tweet at the town level and analyze any trends that can be seen. The new “i.date” and “i.townnumber” variables will essentially put in a dummy variable for every town and absorb the average.

Results


Comparing the Mean Case Rate of Democratic Vote Shares
Case Rate (per 100,000)
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
3/10/2020
4/29/2020
6/18/2020
8/7/2020
9/26/2020
11/15/2020
1/4/2021
2/23/2021
Date
Mean of 40%-50%
Mean 50%-70%
mean 70%+
difference btw low and high


Comparing the Mean Case Rate of Democratic Vote Shares
Case Rate (per 100,000)
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
3/10/2020
4/29/2020
6/18/2020
8/7/2020
9/26/2020
11/15/2020
1/4/2021
2/23/2021
Date
Mean of 40%-50%
Mean 50%-70%
mean 70%+
difference btw low and high
mean of all top 30

As mentioned in the methodology, the first step was to create a graph to look at the average case rates overall between towns grouped together by Democratic voteshare. The results show a slow increase at all levels from the early months of May to September, but a drastic spike after then. The line showing the difference between the low and high vote share groups steadily increases without a significant change until the end of the 2020 year going into 2021. The second graph helps to show how the average case rate across all of the top 30 towns matches up with the line for the average cases between the towns with 50%-70% Democratic voteshare.

The data set, listing the top 30 towns and its caserate from April 2020 to January 2021, was imputed into the program STATA to be further analyzed. The following table shows the results from each regression that was ran:

caserate	**Regression 1**	**Regression 2**	**Regression 3**	**Regression 4**
voteshare	2507.816	2507.816	2560.25	4655.227
(s.e.)	(174.8224)	(76.08752)	(80.59325)	(228.776)
tweet_share			-480.648	-480.648
(s.e.)			(244.0082)	(151.0308)
tweet			-3897.474	-3897.474
(s.e.)			(256.6428)	(158.8511)
town fixed effects	NO	NO	NO	YES
date fixed effects	NO	YES	YES	YES
R-Squared	0.0302	0.8224	0.8225	0.9323

First, we looked at the correlation between caserate and democratic vote share. This is what the first regression yielded: caseratePredicted= 875.9114 + 2507.816*voteshare

The data shows that for every 1 percent increase in voteshare, since the unit of voteshare here is the percentage of democratic voters within the total votes of the towns, a 2507 increase in the rate of cases can be predicted when holding the other variables constant. The constant here of 875.9114 cases shows how even without voteshare, the predicted value of cases is still very high.

This must be taken into consideration since it reminds us how high COVID-19 cases are without even taking into consideration voting patterns.

Using the i.date variable to hold for date fixed effects, this second regression was ran with case rate, vote share and i.date. This is what the regression yielded: caseratePredicted= 4349.092 + 2507.816*voteshare

The coefficient for voteshare remains the same at 2507.816. However, the constant has now changed from 875.9114 to 4349.092 cases. This reveals how taking into consideration the effects of another variable such as "date" actually increased the number of cases there would be if all the other variables were held constant. As we get closer to looking into our hypothesis, this regression proves the need for more variables to really understand what is happening in terms of trends within this data.

Using the tweet_share variable, the third regression with case rate, tweet_share, vote share, tweet and i.date was ran. This is what the regression yielded:
caseratePredicted= 4317.282 - 480.648tweet_share + 2560.25*voteshare - 3897.474tweet

The coefficient for voteshare remains about the same, but the interesting find is the coefficients for the other two variables. The only purpose of the variables tweet and tweet_share are to look into the effects during the time of the tweet (July 20,2020 to August 20,2020). We see a negative coefficient of -480.648 for tweet_share, indicating that there was some sort of effect between the time of the tweet and voteshare (since that is what the variable tweet_share is). But the negative coefficient implies that there was a slight decrease in cases during the time.

However, it is still unclear if this is true for the democratic cities or rather just in general. Our final and most important regression will look more into this situation that raises many questions.

The fourth and last regression used the same variables as the previous one: case rate, tweet_share, vote share, tweet and i.date. But this one also held the fixed effects of the variable “town” as well, by using the variable “i.townnumber.” The fourth regression yielded the following:

caseratePredicted= 3253.066 - 480.648tweet_share + 4655.227*voteshare - 3897.474tweet

Analysis of Main Regression

This last regression is the main focus of our research, and what we had been working towards the entire time. This regression shows the relationship between caserate, tweet_share voteshare and tweet (defined in the previous regression), just like in the previous regression before this one. However, we use the dummy variables i.date and now i.townnumber. The coefficients for tweet_share and tweet remain the same. The coefficient for voteshare has increased from 2560 to 4655, indicating that a unit increase in voteshare yields a much higher increase in cases than previously seen. Examining this portion further, it is saying that for every 1 percent increase in democratic voteshare, the increase in the rate of cases is now more than 4000, much higher than we first assumed. Contrary to our initial assumptions, the data is now slowly revealing that as time went on, at least within our overall timeframe of May 2020 to January 2021, the more democratic a city was, the higher the likelihood of more cases.

With the i.date variable, the overall time pass is taken out. The regression is demeaning anything that is trending overtime. What is essential about the i.date portion of the data table are the coefficients that it lists for each day. Our initial hypothesis looks to see whether there were any changes in case rates after the posting of former President Trump's tweet. The timeline that we look at specifically is one-month post-tweet (July 20, 2020 to August 20, 2020).

The date coefficient tells us how many more cases there were that day than average (during that time frame) and the differences in the case rates. The town's fixed effects (the date and town variables) absorbed the average level of cases in each town over the entire period, giving us a more precise estimate. What it revealed was more in line with the results seen from the increase in the coefficient of the voteshare mentioned previously. In general, we see the cases increasing as time went on, but that is not exactly what we see within the July 20- August 20 time frame. When looking at the time directly after the tweet, you see a decline. Specifically, during this month the date coefficient is negative, and therefore less cases than average. This is extremely interesting because this seems to suggest there is an impact from the tweet, since cases are going down. We see a negative effect of the tweet, where the tweet itself is associated in that time with a large decrease in total cases. It does go with our assumption that the tweet seemed to have an effect, but the unexpected part is how much more democratic cities were affected by it than was first assumed.

It is important to note that this timeframe is only looking at a month after the tweet, and many other events can have influenced these numbers as well. This data must be looked at objectively for that reason: it being a small time frame after the tweet and the many other factors that were not considered and thus not held constant. The data does show the coefficient to stay negative, and cases decreasing from average, for a bit longer after the month, but it is still unclear by how much. The time of this tweet was the peak of the summer, where temperatures were very high, as well as the heated political climate of the upcoming presidential election in

November. Dummy variables are difficult to interpret in general, so that must also be taken into consideration when looking at a variable like i.date.

Overall, democratic cities have higher COVID-19 case rates in Connecticut. This is most likely from a majority of those cities being near the western shoreline and getting hit by the high numbers coming in from New York. The democratic cities make up 22 out of the 30 most populated cities in Connecticut, and cities are known to spread COVID-19 more. But, with the last regression, seeing the interaction between the tweet and the voteshare as being more negative is a new finding. It means that the more democratic an area is, the tweet had a bigger effect on decreasing rates. This was not obvious. From our personal analysis of the data and the literary research conducted before, there are a few reasons that this could be. If someone is partisan or independent in a more democratic area, where there is already more pressure to wear a mask or where cases are already higher, the tweet meant more. The citizens may have taken the tweet seriously either way, but this data suggests it was more so in more democratic areas.

Conclusion

Our initial assumption was that an information shock coming from a republican source, more importantly the President himself, would either have a negative effect or even no significant effect in democratic areas. However, the data suggests that in the more democratic cities and counties, the area did in fact take it more seriously than we thought they would have. This can be explained in different ways. It is difficult to imagine that the more democratic citizens are taking the strongly-right leaning former president extremely seriously, though it cannot be ruled out that they may be. But the more likely reasoning involves remembering that this predominantly democratic, highly populated cities also have a good number of republican citizens as well. It can be assumed that those republicans and independents in those areas, in combination to being in a place that is getting hit harder and the information shock of President Trump’s tweet, is what this data and regression is picking up on. It is harder when the area is getting hit harder, and the area is more democratic. Mask mandates are more likely in such areas, and it is more likely to be a part of the culture. This is simply a complementary between the information and the overall environment. And after putting in the fixed effects from our regression and the standard error goes down, the idea does hold up.

Bibliography

Gollwitzer, A., Martel, C., Brady, W.J. *et al.* Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nat Hum Behavior* **4,** 1186–1197 (2020). https://doi.org/10.1038/s41562-020-00977-7

Bursztyn, Leonardo, et al. "Misinformation During a Pandemic." *NBER*, 22 June 2020, www.nber.org/papers/w27417.

Ajzenman, N, Cavalcanti, T and Da Mata, D. 2020. 'More than Words: Leaders' Speech and Risky Behavior During a Pandemic'. London, Centre for Economic Policy Research. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14707

Neelon, Brian, *et al.* "Associations between Governor Political Affiliation and COVID-19 Cases, Deaths, and Testing in the United States." Nov. 2020, doi:10.1101/2020.10.08.20209619.

William, Kevin. "Trump Says Face Masks Are 'Patriotic' after Months of Largely Resisting Wearing One." *CNBC*, CNBC, 20 July 2020, www.cnbc.com/2020/07/20/trump-says-coronavirus-masks-are-patriotic-after-months-of-largely-resisting-wearing-one.html.

Mason, Melanie. "The Presidential Cover-up, or How Trump Went from Shunning to Wearing a Mask." *Los Angeles Times*, Los Angeles Times, 14 July 2020, www.latimes.com/politics/story/2020-07-14/trump-coronavirus-mask-evolution.

COVID-19 Tests, Cases, and Deaths (By Town), *Department of Health*, *CTData.gov* 2021 https://data.ct.gov/Health-and-Human-Services/COVID-19-Tests-Cases-and-Deaths-By-Town-/28fr-iqnx

"Connecticut Cities by Population." *Connecticut Demographics*, May 2020, www.connecticut-demographics.com/cities_by_population.

Brown, Brendan. *Trump Twitter Archive*, 2020 www.thetrumparchive.com/?searchbox=%22mask%22.

"Election Night Reporting." *Connecticut Secretary of the State*, 2020, ctemspublic.pcctg.net/#/selectTown.