

David Lattimer

DSC 680

6/27/2021

Vaccine Hesitancy

Overview:

Over the last 15+ months, the United States and the rest of the world have been hit hard by the COVID-19 virus and figuring out how to keep people safe. In the US, there have been a myriad of issues that have come about from how to handle such a deadly and contagious virus, and what the best way is to return back to some sense of normalcy. These issues include shutting down public events and businesses, when to open back up these public offerings, how useful masks are, if people should be forced to wear the masks, the overall safety and efficacy of the vaccines that have been used to combat the virus and more. During this time, there have been several factors in what people believe is the best way to handle the virus, one of which is political alignment, others being how affected the community has been by the virus. As people are closer to people with COVID, the more likely they are to take it seriously and take the necessary precautions, as well as get the vaccine to protect themselves. By looking through the data of estimated vaccine hesitancy in each US county, we are going to look into some approaches for how to help the most people and change the most minds so we can open back up and return to normal. With this, I will be looking from a couple different lenses, such as the national strategy that needs to be employed, the state strategy and try to locate the places that have relatively low vaccine rates while also are most likely to change their mind and get the vaccine if given the chance. By gauging the hesitancy to get the vaccine, we can predict who we will need to change the mind of to get to trust that this will protect both themselves and others. This is especially hard when there is so much misinformation and so many narratives that are putting doubt in people's minds, and we will

need to find a way to show them the truth and offer some sort of incentive to get the vaccine, especially if it means keeping more people safe, especially as more variants start sweeping the nation.

Defining the Problem:

The main issue with getting through the COVID-19 pandemic, especially as more variants are arriving and the regulations around the country are being lifted, it is more important than ever to get vaccinated and keep yourself and others safe. More and more people are getting vaccinated to protect themselves, but as a country, it is important to find the areas where people are more hesitant to get the vaccine and create a plan of action to tackle the problems, whether it's a state by state or a county by county approach. By creating incentives for people to get vaccines, as well as offering scientific approaches and being able to answer the questions people might have and alleviate any of the worries they have, we can find the places that are most important to focus on.

One of the aspects of defining the problem in this case also includes defining the solution. We need to look into approaches that some of the more successful counties have used to increase the number of people getting vaccinated and teach people the importance of the vaccination in terms of being able to return back to normal. Lots of companies and counties have offered incentives, such as Krispie Kreme and various lotteries across many of the states. In the end, as long as we find a way to change people's minds, the more people are going to end up being safe and we can fully open without fear.

The goal of this project will be to identify both the states and the counties that are underperforming. If we are looking at a federal view of the problem, we can see which states are worth putting resources into in terms of convincing people to get the vaccine and getting vaccines into the areas that need it. From a state by state view, we can find counties that are underperforming where we would expect them to be based on the indicators that lead to

people being hesitant about the vaccine and try to figure out why that may be and see if we can convince these people to get vaccinated and may change their mind if given the right incentives.

Data Understanding:

The datasets that were used for this project were data taken from data.gov as well as voting data from the most recent election. The first dataset had information about the vaccine hesitancy including an estimation of both the people that are considered hesitant and strongly hesitant, along with each county. The dataset also had information about the Social Vulnerability Index, which is defined as “It refers to the potential negative effects on communities caused by external stresses on human health. Such stresses include natural or human-caused disasters, or disease outbreaks. Reducing social vulnerability can decrease both human suffering and economic loss. It ranges from 0 to 1. Lower the SVI, the better it is.” This gives us a county by county look at places where COVID-19 might be taken more seriously because of previous disease outbreaks, and others that may not be as strongly affected by previous outbreaks. This also shows us places that are negatively affected by disasters and are more susceptible to disaster and live more troubled lives.

This dataset also had a variable called the CVAC level of concern, which measures how hard it is to roll the vaccine out to the county based on a myriad of historical reasons, such as “historic undervaccination, sociodemographic barriers, resource-constrained healthcare system, healthcare accessibility barriers, and irregular care-seeking behaviors.” By measuring how difficult it is for the vaccine to be administered to the county, even if there is interest in the vaccine is important to measure so that we can focus on overcoming these obstacles if we can get the vaccination numbers up and get more people the vaccine. Other variables in the dataset included the county and state, the percent of adults that were already fully vaccinated, and the demographic statistics of the county.

This dataset was combined with a voting dataset from the most recent election, mostly because a lot of the recent hesitancy or straight anti-vaccination views seem to be closely tied to political views. This dataset simply included all the political parties and the voter counts of each county in the United States. Although there are apparently some counties that have 24 political parties to vote for, the percentage of the county voting Republican and voting Democrat were what we found most important for trying to determine who might be hesitant about the vaccines. This dataset also had the population of each county, which is helpful for some of the smaller population counties that although may have a large percent of hesitant people, they may be such a small population that we could find a different county to impact more lives.

Using the combination of these datasets allows us to try to explain the vaccine hesitancy and try to find any counties that are far off where we would expect them to be. If we can build a model that shows a county is underperforming in terms of vaccines, but all of the factors in the county show that they should be performing better, we can target the county and know which counties are ready to be convinced and will be receptive to incentives we have for them.

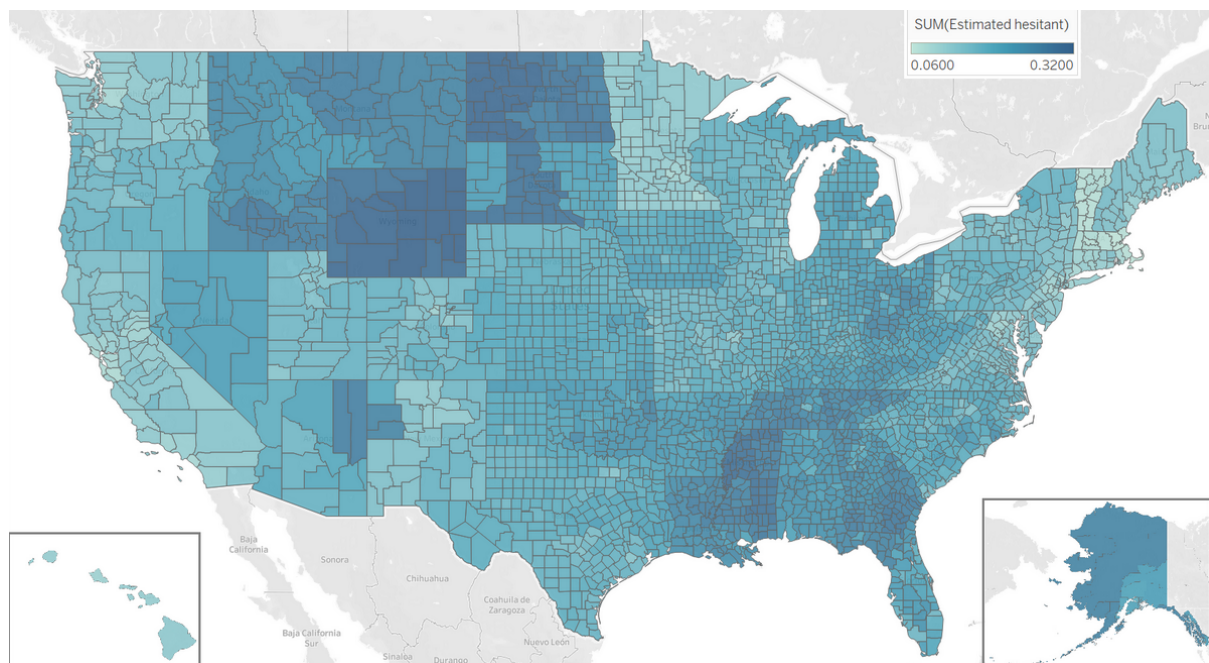
Data Preparation:

Data preparation was fairly straight forward to handle and throw into a dataset ready to be modelled. I did away with some of the redundant sections (the boundaries of the states/counties, the category of the SVI and CVAC sections so we could just use the number values, and the demographic data because it wasn't very helpful and wasn't the cause of any uncertainty), as well as created a few categories to help with the model. After pivoting and combining the voter data, to the original dataset, I then switched the total voting numbers to a percentage of voters in the county and got rid of all the minor political parties that were voted on. Although green or libertarian candidates might have some votes randomly dispersed, the

numbers were all fairly low and can still be seen when looking at the missing percentages in a county ($100\% - \text{Percent Republican} - \text{Percent Democrat}$). I also filled the missing values in the percent of fully vaccinated adults with the median, and the one missing value for SVI with the median as well, and now everything was ready to go.

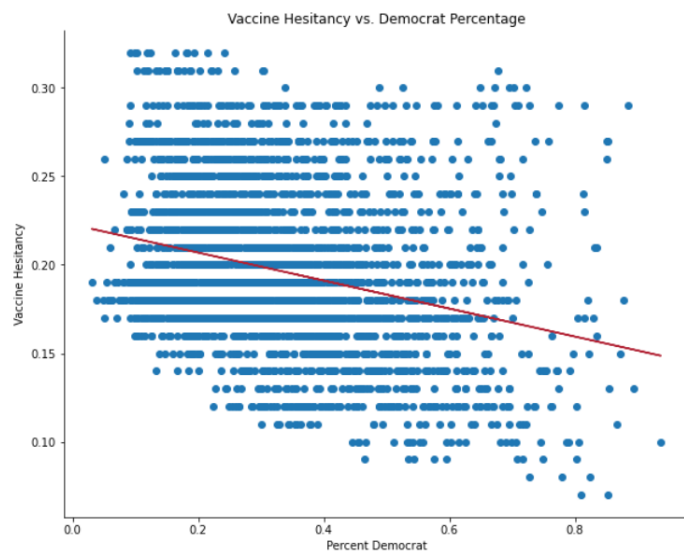
Modelling:

Modelling was done in several parts. Although it did not require modelling for the first aspect, the federal view, could be done fairly simply by looking for the problem areas and states, like in the graph below:



Clearly, there are some states that really stand out in terms of vaccine hesitancy, and the fact that the hesitancy is clearly defined along those state lines means that the state government isn't doing enough to keep their citizens safe and should be focused on to change the public perception of the vaccine. From a federal stance, focusing on states such as Wyoming and North Dakota would be a good place to start. Although these states are much more sparsely populated than many others, it still would be worth investing some resources to advertise the vaccines, answer any questions that they may have and incentivize getting the vaccine in these areas.

The more difficult part of this is identifying the counties that are underperforming based on the rest of their statistics that we measured. If we can maybe find the top 10 counties that are underperforming, we can target them with incentive programs, increased vaccine rollouts and information campaigns to stop the spread of the virus. Ideally these counties will be moderately to highly populated so we can get a large number of people (and we don't really want to focus on small populations), to be fairly democratic in voting (as seen



on the graph to the left), and not suffer from logistics problems which would make it difficult for us to roll vaccines out to those who need it.

Essentially, with this model, we will be using a simple linear regression model using all of the counties we

have data from. Instead of the usual creation of a model using training data and then testing the prediction model on test data, we will use all the counties, and then compare it to the actual hesitancy of those counties. We could easily create a model to predict hesitancy, but there is no need with this data because we already have the hesitancy rates and this doesn't help us make any decisions. Instead, we can compare the actual hesitancy with the predicted hesitancy, and any place that had a much higher actual hesitancy would be a good candidate for being able to change minds and get people vaccinated. We will need to filter out a few of the lower population counties that find their way into our top ten, but should be able to find some good areas to target and offer whatever is needed to get people vaccinated.

When looking into counties that are underperforming, but we would expect to be swayed toward getting the vaccines if they are targeted, we come across these counties that

are the most likely to be convinced. These 10 counties are underperforming the most, and

	County Name	State	Estimated hesitant	Predicted hesitant	Expected Growth
1943	Cibola County, New Mexico	New Mexico	0.26	0.175467	0.084533
1918	McKinley County, New Mexico	New Mexico	0.26	0.192939	0.067061
2512	Hughes County, South Dakota	South Dakota	0.29	0.228268	0.061732
2106	Onslow County, North Carolina	North Carolina	0.24	0.186974	0.053026
261	Navajo County, Arizona	Arizona	0.27	0.219235	0.050765
680	Blaine County, Idaho	Idaho	0.28	0.230232	0.049768
255	Apache County, Arizona	Arizona	0.27	0.225323	0.044677
2831	Norfolk city, Virginia	Virginia	0.20	0.160258	0.039742
1870	Lewis County, New York	New York	0.19	0.152114	0.037886
532	Bryan County, Georgia	Georgia	0.27	0.232662	0.037338

Would be the most likely to improve if pushed that way. This list also excludes any counties with a population in the lower 25 percentile of populations to ensure we are focusing on counties that have a big enough population to focus on. Although it seems that we can cover two counties in New Mexico, we have a spread of places country wide that we can figure out how to ease the hesitancy. Interesting enough for us, we have a cluster of counties here that all can be worked on at the same time. Cibola County, McKinley County, Navajo County and Apache County are all located adjacent to one another and could be focused on as a big source of alleviating the hesitancy all as one.

Conclusion:

Essentially what was done in this project was the creation of a linear regression model to take in a bunch of variables in order to predict the hesitancy we would expect to see from each county, along with the actual hesitancy that we got. By entering all the county information in, we could compare the expected and actual hesitancies, and then we could find the counties that are most likely to improve if we put some focus into them. By locating these counties, we learned there are areas where several counties are connected, even across state lines, that are having the same problem. We have 4 counties in the top 10 that are

underperforming, and they are all adjacent to each other from New Mexico to Arizona, and being able to focus on this much bigger plot of land would simplify the strategy we use to convince people.

On a larger scale, just mapping the hesitancy could show us insights into states that are having problems, which we learned we are seeing especially in Wyoming and North Dakota. We may need to take a different approach on how to handle state-wide hesitancy because none of our top 10 counties are located in these states where we are seeing the most hesitancy. This means either the populations of the counties are pretty low (which would not be surprising in North Dakota or Wyoming), or the HVAC, SVI and political landscape don't suggest that they should be doing much better. This could mean tackling the problem with possibly phone lines to answer questions from individuals, websites to clear up common misconceptions in the area, and possibly a lottery system like many states employed. This could also extend to working with local/state businesses in the area that can put together a way for people to get a free meal/discounts for fully vaccinated people. These seem like begging and hoping for the best, but in the end, especially with the Delta variant of the virus sweeping the nation, any reason for people to get vaccinated is a success.