James Yang

DATA 512
11/8/22

Does Temperature impact COVID-19?

Introduction

In America, there is a theory that the cold can make you sick. Whenever you get a runny nose, begin coughing, or feel week, the first instinct is to get warm. There are many studies that suggest that we don't get sick from being cold in general, but rather become more susceptible to getting sick. However, these studies are mostly in regards to the cold or flu. The problem in which I am trying to expand upon is whether COVID-19 is impacted by the weather, and how the Bergen county in New Jersey population fairs with this theory that the cold can affect the rate at which you become sick. Does temperature impact human susceptibility to COVID-19 like it does for the cold/flu?

Background/Related Work

In this project, we hope to extend upon the existing research that weather has increased susceptibility on the cold/flu and whether that concept applies to COVID-19. Dr. Fecher says "You can't get sick from being cold in general, whether you are outside or inside. Can you get sick from being cold? Yes, but not in terms of a cold or the flu" [1]. This begs the question, does COVID-19 follow the same trend as the cold or flu? How are we going to determine whether it follows a trendline? What is the true baseline?

In a research article written by Aly Kassem, he discusses the effects of temperature and whether it affects COVID-19 [1.1], he deduces that temperature "reports that the relationship between COVID-19 transmission and temperature is marginally and statistically confirmed ($p < 0.1$) in just two observations out of twelve." Indicating that he believes that the temperature does influence getting COVID. We will also try to deduce the validity of this article as well.

Data Used

For this problem, I first pulled from the Rutgers climate database that contained the max, min, and average temperature for each county in New Jersey [2]. I cross checked this data with different weather websites for archived temperatures and determined that they were accurate through some averaging of sampled weeks. The site allows for open use of its data and requests for a reference as its constraint of usage. This data will allow for further analysis of weather impacts within the Bergen county, obtaining one of our most important features. This feature being the average temperature recorded in each month. This dataset expands on the common analysis by introducing a new factor of weather; this feature is independent of COVID-19, as the weather will not depend on COVID-19 cases.

Another dataset I used is the New Jersey Outpatient Influenza surveillance dataset that contains a set of flu confirmed cases in the past 20 years [5]. Although it isn't the cold, it is still within the realm of the flu/cold.

Methodology

We start by doing data cleaning with regards to 3 datasets; the first dataset is a Kaggle repository of John Hopkins University COVID-19 data, the second being a CDC dataset of masking mandates by county, and the third being a New York Times mask compliance dataset [4]. In order to standardize and clean the three datasets, we needed to link the raw US deaths with the state associated to the New Jersey state and the columns associated with the Bergen county. Another thing we needed to clean is the dates associated with each state and the FIPS value that links with the mask use by county. Lastly, the number of deaths and the date associated with each date needs to be standardized across each of the datasets so they can be interpretable. These are the inner workings of the figure 1.1 produced below.

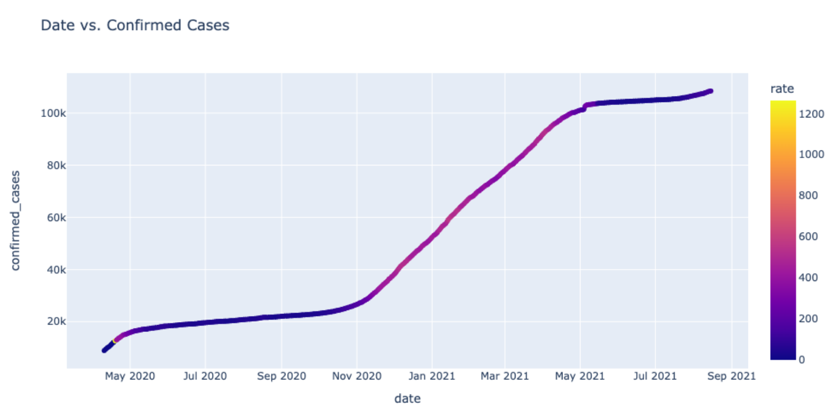

Figure 1.1: Depicts the visualization.

The visualization shows the dates vs the confirmed amount of cases with a corresponding rate on a color coded map of New Jersey in the county Bergen. The visualization itself contains all moments of masking policies in effect because New Jersey instantiated the mask policy from the beginning of COVID.

The next step is producing a rate progression of confirmed cases over a set amount of days. I have chosen 10 to average 3 rate changes per month. The rates increase a bit through mid November 2020 to May 2021 before flattening out again, creating a modal distribution in figure 1.2.

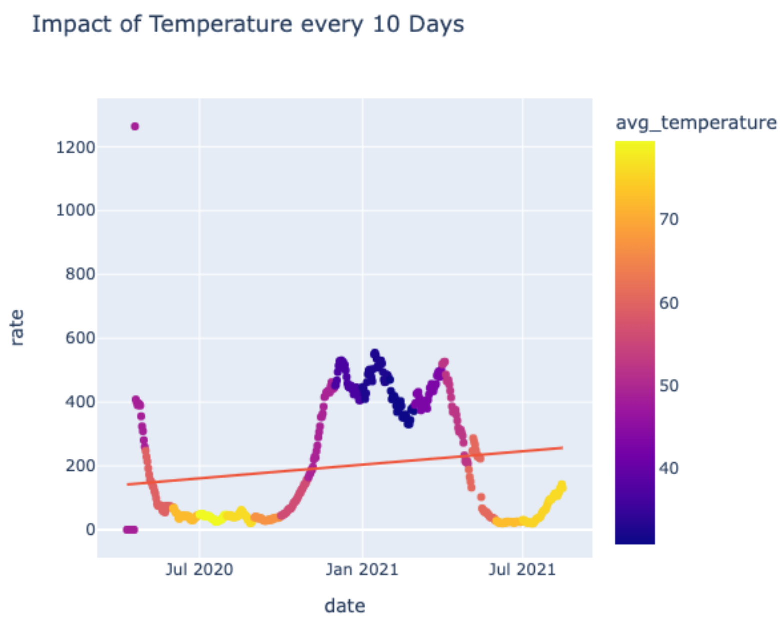

Figure 1.2: Depicts the rate progression of confirmed cases every 10 days in New Jersey, Bergen county.

When importing the temperature and transforming it to align with the dates associated with the cases, we notice an interesting trend with regards to the average temperature. In figure 1.3, we can see that the average temperature dramatically drops as the rate increases. We can see that the OLS trendline has the rate trending upwards with the scale of the same dates as the flu figure 1.6.

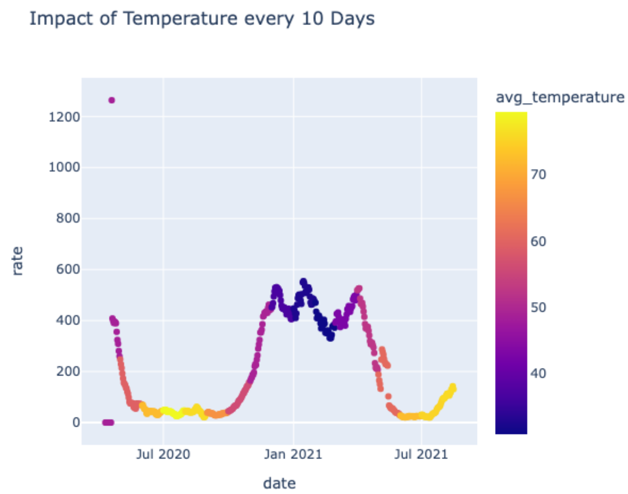

Figure 1.3: Depicts the impact of temperature every 10 days for New Jersey, Bergen.

When we compare with the confirmed cases of the flu in New Jersey, we will eventually see that there is indeed an impact on temperature with regards to rates of both COVID and the flu. This will be detailed in the findings section below.

Findings

The two cases show an instinctive difference in scalable average. As seen in figure 1.6, we can see that the trend decreases as the year goes on. There are a lot more cases of the FLU then there are with COVID-19. I chose to use 2019-2020 instead of 2020-2021 because the data from 2020-2021 could be skewed with false FLU and false COVID cases. By using the case the prior year, the number of labels for false COVID-19 could be eliminated.

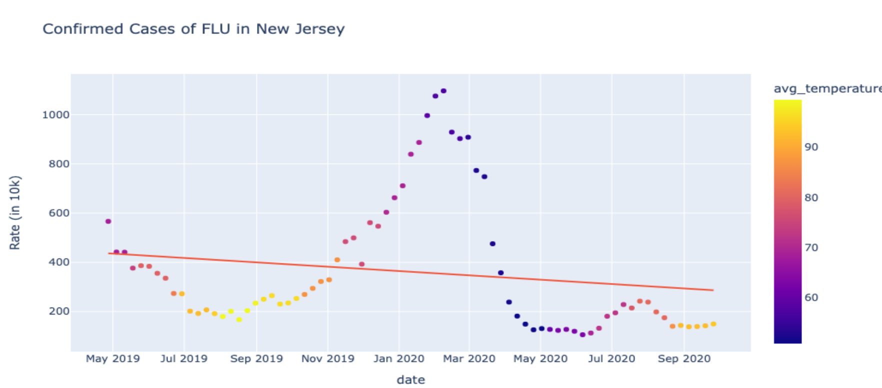

Figure 1.6: Depicts the average rate for the FLU in New Jersey.

When exploring further, we can also see the correlation of average temperature and confirmed cases being close to 0 which means they aren't necessarily correlated. However, this does not necessarily mean there isn't any causation.

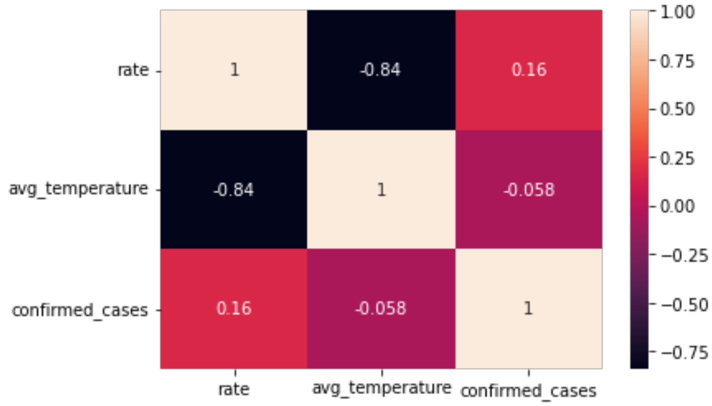

Figure 1.7: Depicts the correlation matrix between the three variables rate, average temperature, and confirmed cases.

For the hypothesis test, we need to statistically determine for an area of 0.05 ($\alpha = 0.05$), the z-score will approximately be 1.645, the outcome of our test. The null hypothesis that we will set is the claim that the temperature does not affect the rate for both diseases (($\mu1 = \mu2$) given (T1 = T2)). The alternative is that the temperature does affect the rate for both diseases (($\mu1$!= $\mu2$) given (T1 = T2)). We calculated that the average rate for COVID-19 is 220.3 and the average rate for the flu was 199.5. The standard deviation for COVID-19 is 188 and the number of samples is 450.

$$Z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

Figure 1.8: Depicts the equation to calculate the Z-score for the given rates.

Plugging in the values, we get that the Z value is equal to 2.31 which is greater than 1.645. This means that we reject the null hypothesis claim that temperature doesn't affect the rate for both diseases. We accept the alternative statement that temperature affects the rates for both diseases.

Discussion/Implications

Through the methodology, we found that we reject the null hypothesis claim and accept that temperature affects the rates for both diseases. This statement aligns with the statement written by Aly Kassem in the article discussed about regarding his test on impacts of temperature for

COVID. There is a distinct impact of rates correlating with the decrease in temperature in his study and it can be shown at a high level in this study as well.

The implications of this study revolve around human centered behavior as people will believe that temperature has an impact on most transmittable diseases such as the cold, flu, and COVID-19. This study does not make the claim that the affects are the exact same or replicate in behavior. The statement is simply claiming that cold temperature does impact rates of both the cold/flu and COVID-19.

Limitations / Unknowns

When recording average temperature, there is a difference in the temperature displayed versus the temperature felt. There is a difference between air temperature and "feels-like" temperature; air temperature is the actual temperature outside and "feels-like" temperature is the wind and humidity combined with the air temperature [3]. This discrepancy can cause a little bit of unknown as the dataset doesn't explicitly say which type of temperature it is referencing. There are also a lot of external factors that could blend in with temperature impact. For example, the rate could simply increase out of coincidence every year. A deeper analysis of whether COVID-19 impacts in years prior to 2019 would need to occur but that unfortunately does not exist since it started in 2019.

Conclusion

To conclude, we knew from research that the cold weather itself could not get a person sick. However, we also wanted to explore the idea of whether the cold weather had impacts on the cold/flu. Through this research, we can conclude that with the rates increasing through the winter times when temperature dips, that the cold weather does influence the rate of both COVID-19 and cold/flu cases through a standard hypothesis test and sampling. This is simply through the case of Bergen county in New Jersey. There could be cases in other counties where this could conflict, which would generate less conclusive results. However, in the simplified scenario of this research we can deduce that the cold does impact the rates of these diseases.

References/ Data Sources

[1]. "Do You Really Get Sick from Being Cold?" *UnityPoint Health*, https://www.unitypoint.org/livewell/article.aspx?id=9161c3a0-54dc-46d7-a633-e1d15a5227e2.

[1.1] Kassem, Aly Zein Elabdeen. "Does Temperature Affect COVID-19 Transmission?" *Frontiers*, Frontiers, 24 Nov. 2020, https://www.frontiersin.org/articles/10.3389/fpubh.2020.554964/full#:~:text=COVID%2D19%20can%20be%20seasonal,cases%20by%200.86%20(20).

[2]. *ONJSC :: Historical Monthly Summary Tables*, https://climate.rutgers.edu/stateclim_v1/nclimdiv/.

[3]. ThermoPro. "Temperature vs. Feels like Temperature." *Thermopro*, 15 Dec. 2021, https://buythermopro.com/knowledge/temperature-vs-feels-like-temperature/#:~:text=The%20air%20temperature%20is%20the,and%20how%20we%20should%20dress.

[4]. https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university
https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i
https://github.com/nytimes/covid-19-data/tree/master/mask-use

[5]. "Influenza Surveillance - Outpatient Influenza-like Illness Surveillance Data by Region and Influenza Season." *California Open Data*, https://data.ca.gov/dataset/influenza-surveillance/resource/78afd269-ea5b-41aa-9ca8-37af6bb35f5a.