Yujie Li
PSTAT 120C
Final Project
March 12 2021

Temperature relationship with Covid-19

Abstract

The pandemic of Covid-19 is currently a problem that attracts the attention of all the researchers all around the world. Many researchers in the world did many meaningful pieces of research to show that there is a relationship between local temperature and the Covid-19. Many pieces of research prove that the spread speed of Covid-19 has a negative relationship with the local temperature. In this paper, I did some research to test that whether the higher temperature really causes the local area to have fewer confirmed cases. However, the research shows that there is actually the not obvious relationship between the higher temperature and the final situation of Covid-19 in the local area. Many of the other factors will also play a significant role in the final situation of the Covid-19.

1. Introduction

In 2020, a series virus appears and greatly changes the life of all the people around the world. The pandemic of the Covid-19 has become a global problem, which is worth paying attention to for all the people around the world. Studying and trying to understand the virus and the spread of the virus should be really important for us to overcome the challenge and save the lives of people all around the world. Many people have done researches on this topic.

The researcher group of Alessio Notari is one of those researcher groups. According to the article "Temperature dependence of COVID-19 transmission", Alessio Notari uses the data and careful research process to show the audience that the virus of Covid-19 spread slower in the areas with lower local temperature. Alessio Notari uses data of three groups of areas all over the world to do some numerical analysis and produce many plots to show that the parameter of alpha, which is a parameter of the number of cases as a function of time, is negatively related to temperature. That means the higher temperature will cause a lower spreading speed of the Covid-19 Virus in that area.

Some other researches such as "Impact of Temperature and Relative Humidity on the Transmission of COVID-19: A Modeling Study in China and the United States" written by Jingyuan Yang et al and "Impact of temperature on the dynamics of the COVID-19 outbreak in China, Science of The Total Environment" written by Shi Peng et al also proves that the higher temperature can reduce the transmission of Covid-19 virus in some ways.

In this paper, I will also focus on a similar topic as those researchers. I decided to try to explore more on the topic of the relationship between Covid-19 and the local temperature of that area. In section 2, I will talk about the purpose and the goal of this research; in section 3, I will discuss the datasets and the methods I used to approach the final results of the research; in section 4, I will talk about the process of analyzing the data and plots of the research; in section 5, I will discuss the results of the research and some possible reason causing that; in section 6 and 7, I will make a conclusion of everything discussed before and provide a reference.

2. Research Goal

Based on the professional sources I discussed above, we can easily see that temperature plays an important role in the spread of the Covid-19 virus in the early stages. In this final project, I decided to do simple research to study the relationship between local temperature and the cumulative confirmed rate of the pandemic of Covid-19. I want to extend the situation to a broader context. That means I want to test that, besides the early stages of the spread, whether it is the same as that the higher temperature has a negative result on the Covid-19 virus as well.

3. Method

In order to achieve this purpose and solve the question, I choose to narrow the region of the research. I decide to choose just one country in the world as the target of research. In this way, I can make the research a little bit easier, and, at the same time, I can also control some variables, such as the medical condition, the population density, or the development of technology. In order to make sure that the country covers different temperature zones and enough population, I choose the United States as the target of my research.

I decide to use the dataset of US confirmed cases from the Data Analysis Project 2 from our PSTAT 120C project as my primary dataset. I choose this dataset because it is the dataset we used before, so I am more familiar with this dataset than other ones. Also, this dataset contains the basic information of the number of the confirmed cases through different areas all over the country that I need to fulfill my purpose for the research.

For the data I need for the temperature data, I decided to find a dataset of the average temperature for different areas all over the United States. I searched for datasets on Google, and I appropriate data from the website of USA.com. This website provides the average temperature of each state in the United States and the population of that states. These two are just what we want for the research. However, the website did not provide the data in a CSV file as we want it to be, all the information just shows on the web page. Therefore, I decide to use the technique of web-scraping to download those

data from the website. Since I do not know how to do this in R, so I used Python to write a simple code to obtain the information we need on that website and save it as a CSV file as we want.

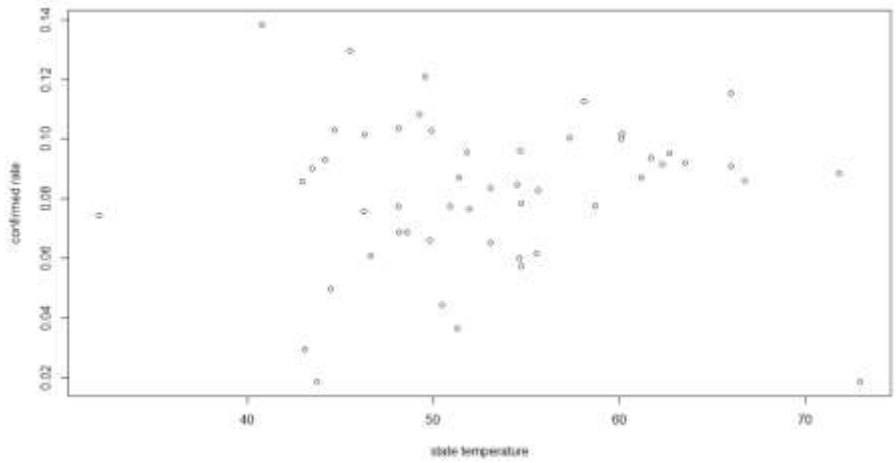
I selected data for each state of the United States and try to find the relationship between them and the average temperature of that state. I made some plots and try to see that is there any visible trends that can reveal the relationship between the spread of the virus and the local temperature.

4. Numerical analysis

In order to study the relationship between the Covid-19 and the local temperature, I decide to use the rate of the cumulative number of confirmed cases in a specific period over the total population of the state in the research process. Therefore, I first choose several different periods of time during the Covid-19 and then use R to calculate the cumulative number of confirmed cases for each state. After that, I divide it by the total population of each of the states from the temperature CSV file. Finally, I try to use R to produce several plots of the data and try to fit some model to it.

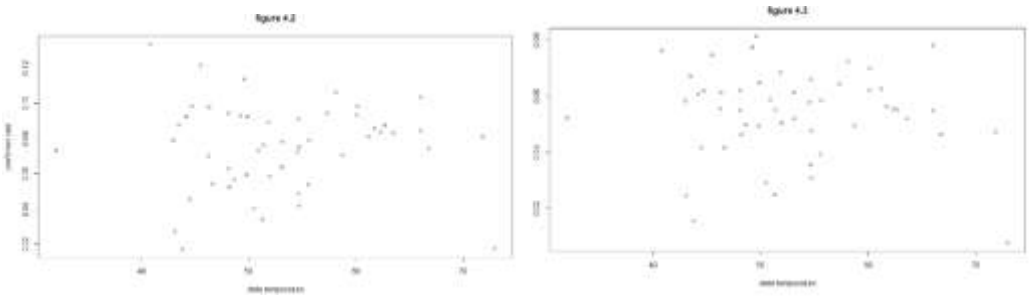
I first tried to fit the data from March 31 2020 to January 31 2021 and make a scatterplot of the confirmed rate against the state average temperature of all the states. Figure 4.1 below is the plot I get. Each dot in the plot represents a state; the x-axis of the dot is the average temperature of the state, and the y-axis of the dot is the confirmed rate of the state. Based on figure 4.1, actually, there is not a really obvious relationship between the state average temperature and the confirmed rate of the state. Then I tried to fit a linear regression model to the data. However, the summary report from R shows that the p-values for F-test and the parameters are really high and the R-square value is really low. This means that the model is invalid, and there is no linear relationship between these two variables.

figure 4.1

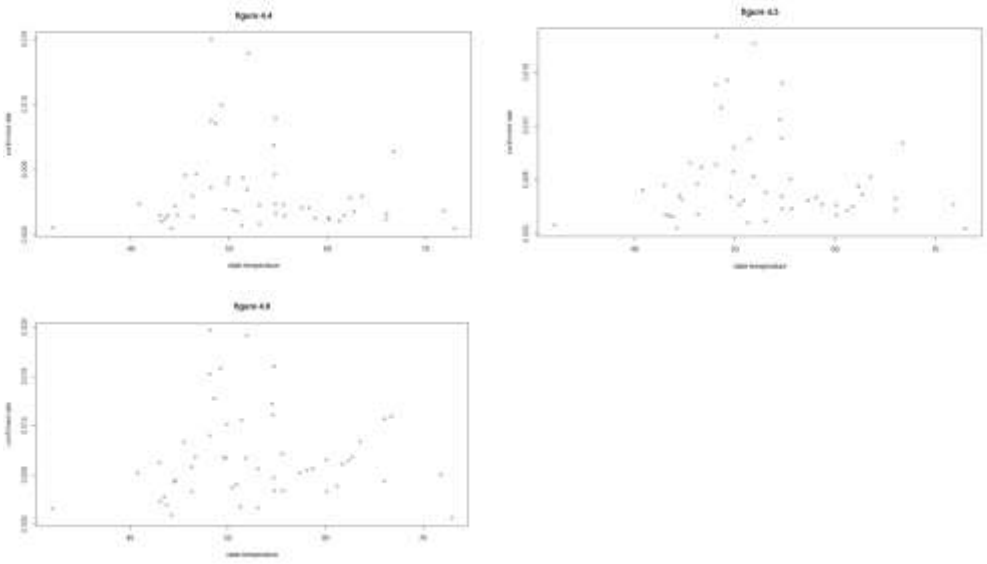


I think that one possible reason that there is no obvious relationship

between these two variables might be that I used the cumulative confirmed cases for different states. However, each state might live a different start date for the Covid-19 pandemic, and many of the states do not have Covid-19 at March 31, 2020, so the usage of cumulative confirmed cases since March 31 2020 may cause problems. Therefore, I tried to change the duration of the research period. I test data from June 30, 2020, to January 31 2021and October 31, 2020, to January 31, 2021, as well. Figure 4.2 and Figure 4.3 below show these two situations. Based on the plots and the summary report from R, there is still no linear relationship between the variables.



Another possible reason for this might be that the total number of days we use in the cumulative confirmed cases for each state is different. Thus, I try to control the number of days from the first day that each state has its more than 30 confirmed cases as Alessio Notari did. I changed the number of days after the first 30 confirmed cases and plot several figures as shown below. Based on figure 4.4 to figure 4.6, there is still no obvious relationship between the state average temperature and the confirmed rate of the state. The summary report from R shows that the p-values for F-test and the parameters are still high and the R-square value is still low. This means that the model still is invalid, and there is no linear relationship between these two variables.



Considering all the numerical analyses above, I found that there is actually no relationship between the temperature of the state and the local rate of cumulative confirmed cases.

5. Discussion

Based on the previous part of the Numerical Analysis, I tried several different ways of manipulating the data I have and make some plots to see that is there also a relationship between the cumulative confirmed rate and the local average temperature. The result of the analysis shows that there is actually no obvious relationship between them. However, based on the article from Alessio Notari, there is a relationship between initial spread speed and the local temperature. There are several possible reasons why there is a difference.

Generally speaking, there are many other factors that can contribute to the cumulative confirmed rate of an area. First of all, the population density of the state can play a significant role. No matter how fast the spread speed is, the spreading of the virus needs the contact of people. However, if the population density of a state is really low, which means that the chance for people there contacting with others can below as well. Therefore, although there spreading speed in that area is high, because of the lack of contact with each other, the total number of confirmed cases can still below, which causes that the cumulative confirmation rate is still low. For example, some of the states have a really large area with a relatively small population, so that the contacting rate between individuals can be really low. This can cause the cumulative confirmation rate of that state low as well.

Transportation of the state can also be an important factor in some ways. In the research, we just considered the situation of each state in the United States. However, the situation of other areas can also influence the situation in the states because people can travel around different states even different countries carrying the virus, which can cause more confirmed cases in the local area as well. For some large cities or developed areas with large international airports, there might be a greater chance to have imported cases causes spread locally. In this way, the local rate of cumulated confirmed cases can still be high in some of the areas of low spreading speed.

Another possible reason can be the policies of the governor. The governor of each state may have different opinions on the pandemic initially. In some of the states, the governor can be more serious about the virus that they may come up with some policies to enforce the social distance restrict traveling earlier than some other states. Thus, the chances for an individual to contact with each other and spread the virus can be relatively low, which can cause the final proportion of the confirmed cases in that area low as well.

Finally, some specific events may also have a great influence on the total confirmed cases. Some of the large events or social movements can let people

gather together and increase the probability of contact with each other to increase the possibility of spreading the virus. For example, last year there were some large parades about the social movement of "Black Life Matters" in many different areas of the state. In this way, the total number of the confirmed cases and the rate of cumulative confirmed cases can be larger in those areas.

6. Conclusion

According to all the discussion above, although, based on the research paper from Alessio Notari, the virus of Covid-19 spread relatively slowly in areas with higher temperature, we can easily come out with the conclusion that the rate of cumulative confirmed cases of states in the United States does not have a relationship with the average temperature of the state because of some possible reasons.

Above all, Covid-19 is a series of pandemic, which causes a huge loss for people all over the world. Therefore, we need to pay enough attention to this pandemic and the virus to overcome the difficulty together.

7. Reference

Notari, Alessio, "Temperature dependence of COVID-19 transmission", Science of The Total Environment, Volume 763, 2021, 144390, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.144390. (https://www.sciencedirect.com/science/article/pii/S0048969720379213)

Shi Peng, Yinqiao Dong, Huanchang Yan, Chenkai Zhao, Xiaoyang Li, Wei Liu, Miao He, Shixing Tang, Shuhua Xi, "Impact of temperature on the dynamics of the COVID-19 outbreak in China, Science of The Total Environment", Volume 728, 2020, 138890, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.138890. (https://www.sciencedirect.com/science/article/pii/S0048969720324074)

USA.com, "U.S. Average Temperature State Rank", accessed March 10, 2021http://www.usa.com/rank/us--average-temperature--state-rank.htm#

Wang, Jingyuan and Tang, Ke and Feng, Kai and Lin, Xin and Lv, Weifeng and Chen, Kun and Wang, Fei, "Impact of Temperature and Relative Humidity on the Transmission of COVID-19: A Modeling Study in China and the United States (March 9, 2020)". BMJ Open, 11(2), e043863, Available at SSRN: https://ssrn.com/abstract=3551767 or http://dx.doi.org/10.2139/ssrn.3551767