PM566 Final Report

Ruowen Wang

Introduction:

- Research question:
  I am trying to find the association between the social distance index and the incidence rate of COVID-19. To check whether keeping social distance is an effective way to slow down the spread of COVID-19.
- Background:
  COVID-19 is a worldwide contagious disease. COVID-19 seems to spread more easily than flu and causes more serious illnesses in some people. It can also take longer before people show symptoms and people can be contagious for longer. There is a vaccine to protect against flu. However, there is currently no vaccine to prevent COVID-19. The best way to prevent infection is to avoid being exposed to the virus. The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes.
- Explanation:
  The social distance index is an integer from 0~100 that represents the extent residents and visitors are practicing social distancing. "0" indicates no social distancing is observed in the community, while "100" indicates all residents are staying at home and no visitors are entering the county. Calculated by MTI.

Methods:

Data source:

- Social distance index:
  https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZAKKCE
- Total US confirmed cases:
  https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HIDLTK
- See the website for details of the social distance index's calculation:
  https://data.covid.umd.edu/methods/index.html

Preparation:

- Firstly, I check the dimensions, headers, footers, and the trend of the social distance. Both data sets have different number of columns and rows. For the social distance index, the date is updated from Jan/01/2020 to Sep/28/2020. For the total US confirmed cases, the date is updated from Jan/21/2020 to Oct/06/2020.
- Then I delete the extra columns and rows, focus on the same time period in both data sets, which is from Jan/21/2020 to Sep/28/2020.
- At last, I check the missing values in both data sets. It shows that there is no any

missing values in both data sets. The two data sets are ready to use in the following procedures.

Results:

statistic information	Alabama	Alaska	Arizona	Arkansas	California	Colorado
mean	26.83	20.82	33.46	25.37	38.9	31.32

Table 1: Statistical information of the social distance index in the first 6 states

This table is getting the statistical information of the social distance index (Jan/21/2020 ~ Sep/28/2020). The Kable format table contains the mean information of the first 6 states: “Alabama", "Alaska", "Arizona", "Arkansas", "California", and "Colorado". We can see the range of social distance index doesn't fluctuate much in these states.

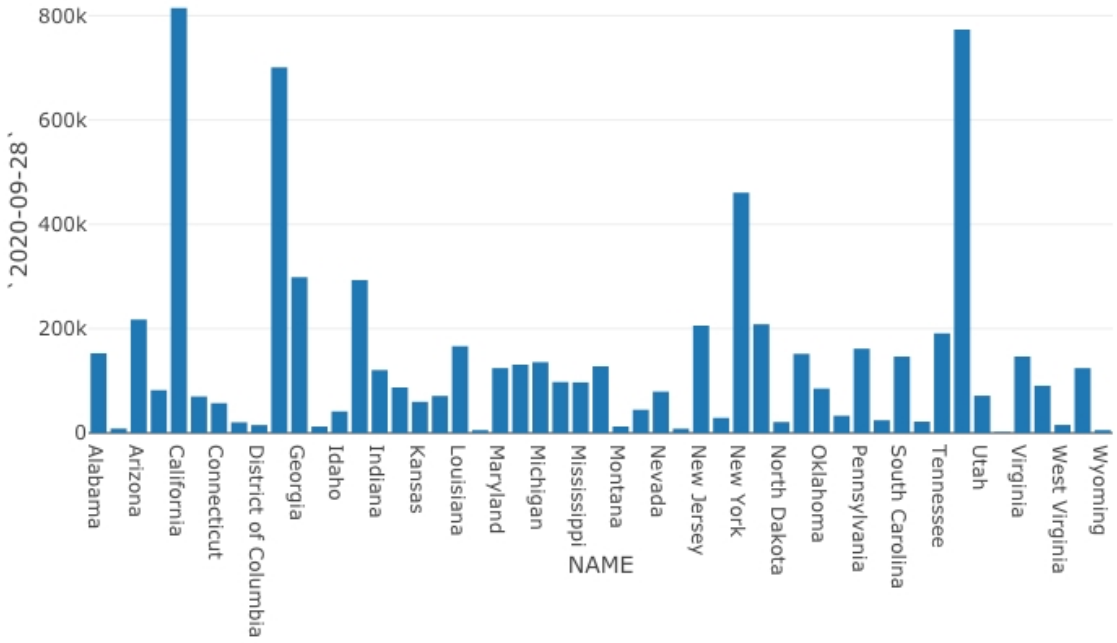

Figure 1: Interactive bar chart of total US confirmed cases on Sep/28/2020

It is clearly to see that California, Texas, Florida, and New York have the most confirmed cases on Sep/28/2020.

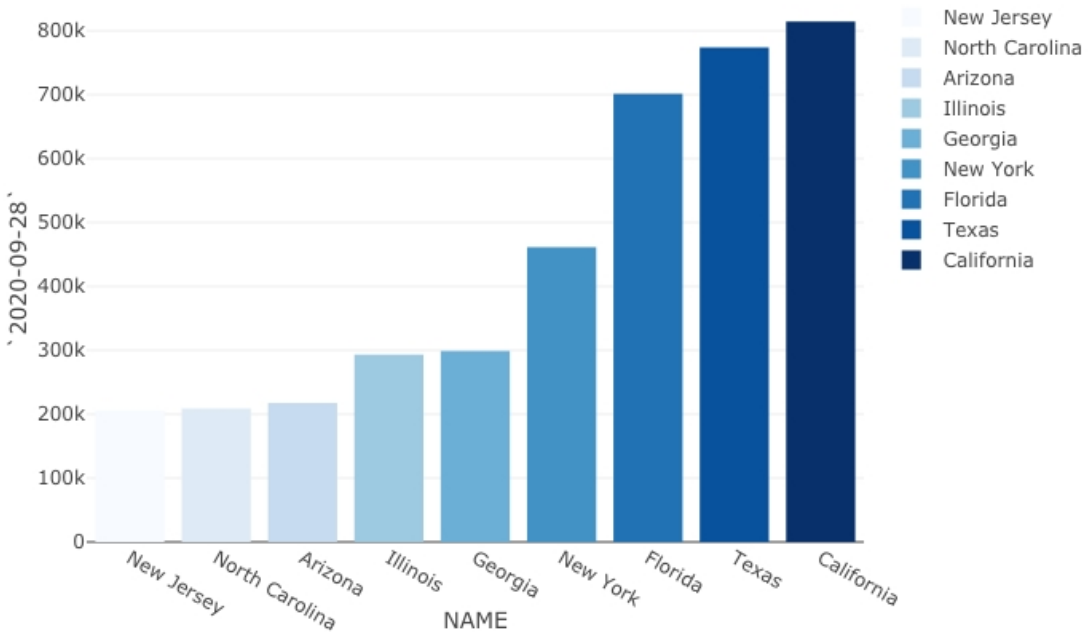

Figure 2: Bar chart of total confirmed cases on Sep/28/2020 which are over 200,000

The purpose of this interactive bar chart is to find which states have the total confirmed cases on Sep/28/2020 which are over 200,000 cases. I use the data set contains the total confirmed cases on Sep/28/2020 to draw this bar chart. The 9 states shown in ascending order on Figure 2 have the total confirmed cases over 200,000. Among these 9 states, California state has the most total confirmed cases, and New Jersey state has the least total confirmed cases.

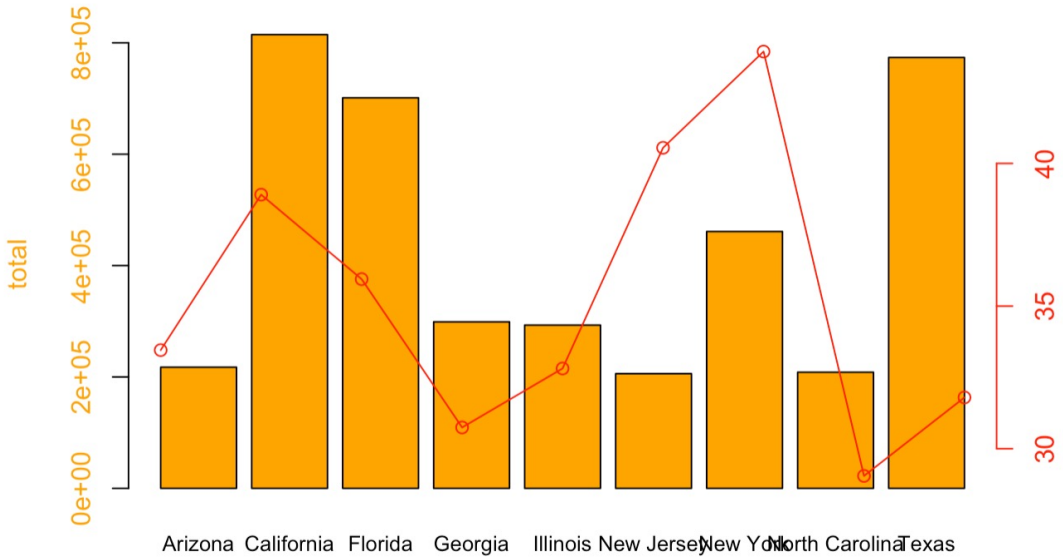

Figure 3: Biaxial chart of the SDI with total confirmed cases in 9 states

This biaxial chart is to compare the social distance index with the total confirmed cases in the 9 states which are over 200,000. From the figure 3, in the 9 states with more than 200,000 confirmed infected people. If we set the social distance index lager than 35 as the long social distance. Here are 4 states follow patterns are shown: those with fewer infected people have a longer social distance, such as New York and New Jersey; the number of infected people is higher in states with shorter social distance, such as California, Florida and Texas.

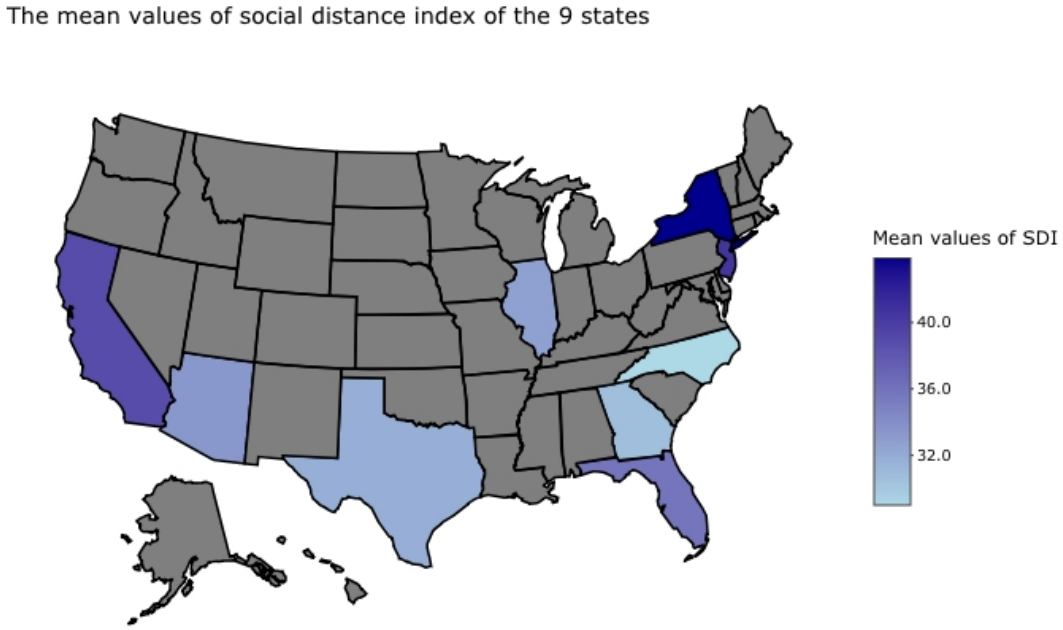

Figure 4: Plot the mean values of social distance index of the 9 states on the map

The states in grey color are those who have total confirmed cases lower than 200,000, which I did not calculate the mean values. The states in blue color are those who have total confirmed cases over than 200,000, which I have calculated the mean values before. The dark blue indicates the higher mean of social distance index, the light blue indicates the lower mean of social distance index.

```
        Pearson's product-moment correlation

data:  meandisz and totalz
t = -0.68561, df = 49, p-value = 0.4962
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3633023  0.1830221
sample estimates:
        cor
-0.09747817
```

Figure 5: Pearson's correlation coefficient

Using COR.TEST() to find whether the variable of total confirmed cases and the variable of mean social distance are related. The correlation coefficient is -0.097 which is very close to 0, the p-value is 0.4962 which means it is not statistically significant, so we can find that these two variables are not related, the association between these two variables is not significant either.

Conclusions:

1. Social distance is not an absolute influence on the number of people diagnosed.
2. The range of social distance doesn't fluctuate much in these states.
3. The total confirmed cases and the mean social distance are not related. The association between these two variables is not statistically significant. The reason may be due to the accuracy of the measurement of the social distance index. Although they use many different data sources and the calculation after multi-level weighting, they still can't guarantee the accuracy of the estimated index.
4. Although this conclusion is surprising me, I think we still need to focus on keeping social distance, because New York and New Jersey are two states have high population density, but with quite low confirmed cases numbers, we can also see that the social distance indexes of these two states are the highest two numbers in the plot. Maybe keeping a social distance is more efficient in big cities than in rural areas. Therefore, as the states with the same high population density, California and Texas should keep social distance and see if that can slow down the spread of Covid-19.
5. From the website of the World Health Organization, the best way to prevent and slow down transmission is to be well informed about the COVID-19 virus, the disease it causes and how it spreads. Protect yourself and others from infection by washing your hands or using an alcohol-based rub frequently and not touching your face. The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it’s important that you also practice respiratory etiquette (for example, by coughing into a flexed elbow). It is highly suggested to wear a mask when you are going outside.