Coursera Capstone Project

The Battle of the Neighborhoods

2020 coronavirus pandemic in the US

By Minh Cong Tran

1. **Introduction**
   1. **Problem**

The ongoing pandemic of coronavirus disease 2019 (COVID-19), a novel infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), spread to the United States in January 2020. The widespread disease transmission may force large numbers of people to seek healthcare, which could overload healthcare systems and lead to otherwise preventable deaths. The government needs to understand the situation at each state to have appropriate support. Moreover, all people having symptoms of COVID-19 are advised to visit the nearest hospitals to avoid getting the community infected.

* 1. **Target Audience**

The target audiences will be the government and organizations controlling the pandemic of coronavirus disease 2019 (COVID-19), the COVID-19 patients who need to access healthcare as soon as possible when having symptoms to avoid spreading diseases.

1. **Data**

Our data include States & number of patients (cases and deaths) collected from the link: <https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_the_United_States>

We also collect data on hospitals' locations from Foursquare API and geojon file data of US states.

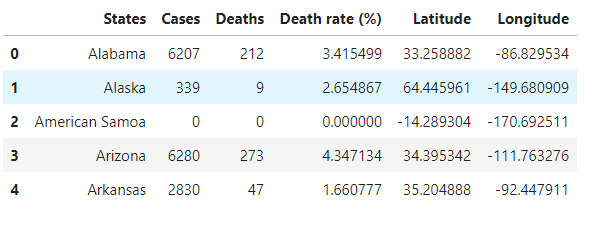
1. **Method**

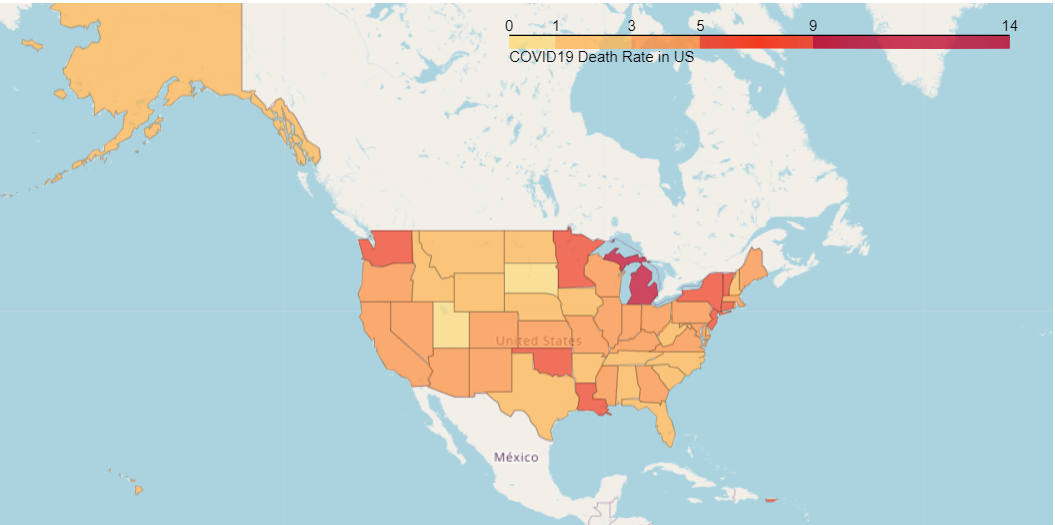
Using the data above, data is transferred to a data frame first to deal with data efficiently. Then data was modified based on the goal of this project. The death rate (deaths/cases) is used to rank the risk of each state. Visualized map of death rate is build to help understand the overall situation and identify the high-risk state. On the search of hospitals, the hospitals' location will be extracted from Foursquare API based on the patients’ location, then visualized by the map so that the patients have the choices to reach the nearest hospitals.

1. **Results and Discussion**

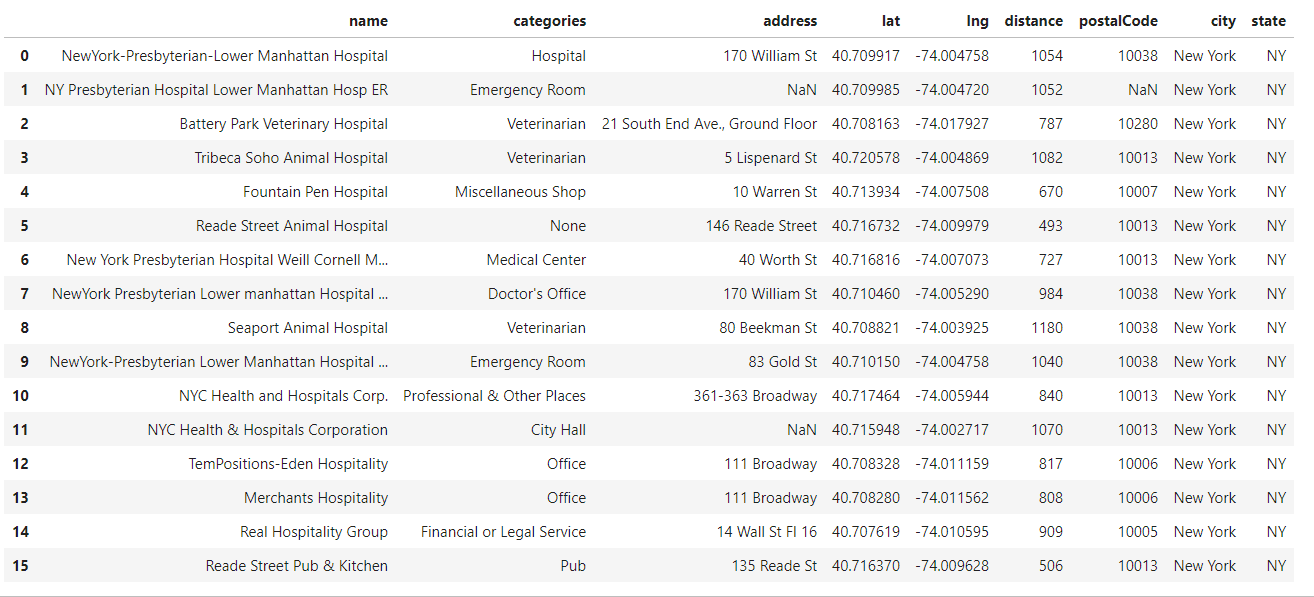
From the data analysis, we were able to identify which states are under high-risk, then more support (workforce, medical supply, finance..) should be prioritized to them. According to the death rate (deaths/cases) at each state, we divided states into 5 groups:

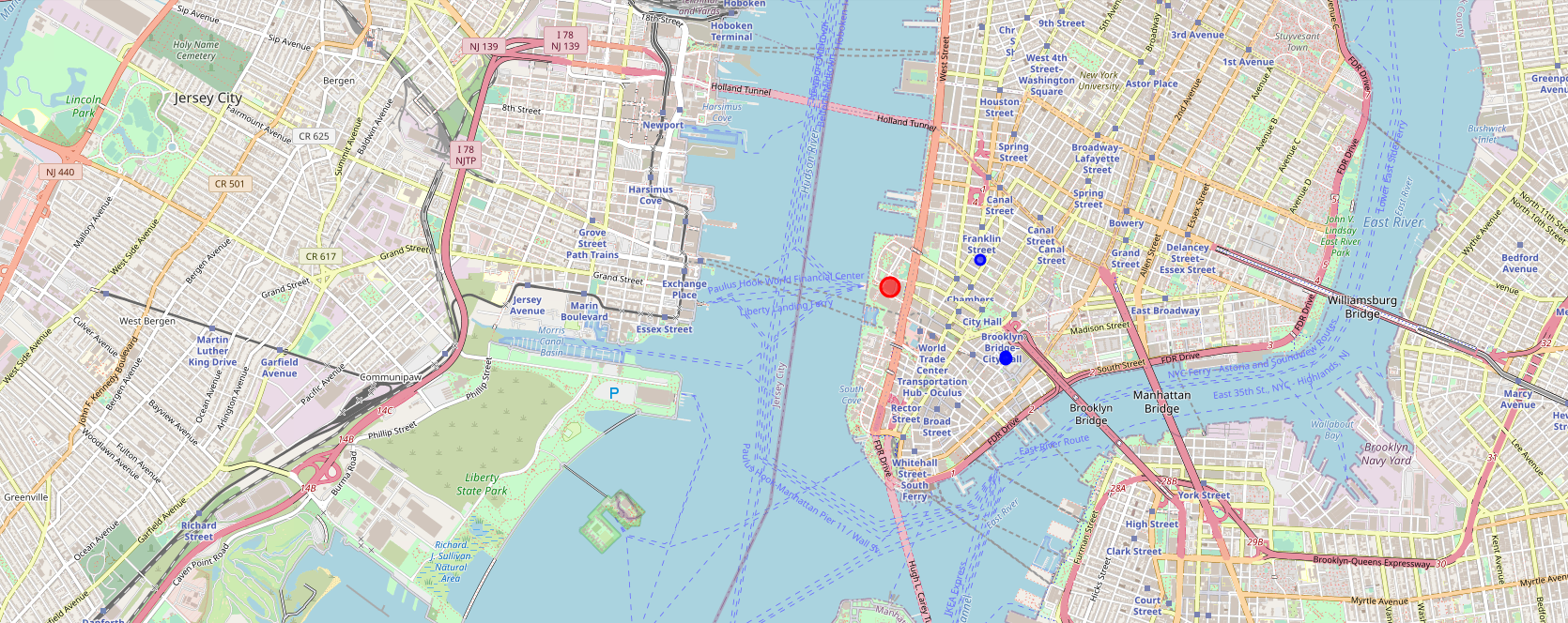
1. low risk: death rate < 1%
2. medium risk: 1% <= death rate < 3%
3. high risk: 3% <= death rate < 5%
4. very high risk: 5% <= death rate < 9%
5. extreme high risk: death rate >= 9%





Also, from the patient’s location, we can extract the hospitals’ locations nearby, then the patient can choose the nearest one to visit.





1. **Conclusion**

The analysis above will help the government understand the situation for making a proper supported plan over limited resources. For example, those states having high-risk status get prioritized in the distribution of medical supply (masks, medicine, protection kit...), workforce (specialist, nurses, doctors...), food,... And the patients can reach the nearest hospitals to have healthcare check-up as soon as possible before spreading the diseases.