Executive Summary: An Assessment of U.S. County Health System Resilience

Authored by: Charles Gallagher, David Wren, and Cameron Holt

References

A. Johns Hopkins University Coronavirus mapping, <https://systems.jhu.edu/research/public-health/ncov/>, 2020

B. Hatchett RJ, Mecher CE, and Lipsitch M, *Public health interventions and epidemic intensity during the 1918 influenza pandemic*, PNAS, 2007

Introduction

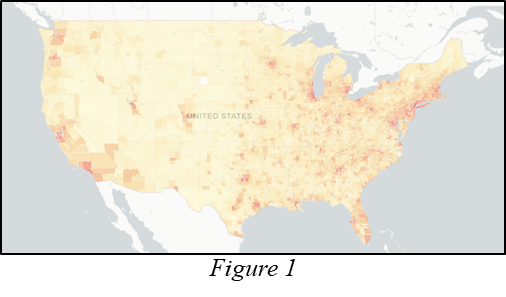
The team investigated the resiliency of the U.S. health system at the county level within the context of COVID-19. The study utilizes U.S. hospital bed1 and census data to establish health capacity at the county level and overlays this with known outbreaks (vide Ref A) and a COVID-19 risk factor to produce visualizations highlighting the net risk for counties throughout the country. The risk factor takes into account a range of factors including population density, proximity to confirmed COVID-19 cases, and other demographic data. A preliminary assessment of passenger flows through the US airport network was also conducted in order to identify connectivity of major population centers with highly infected countries.

Motivation and Learning Outcomes

This study was motivated by concern that there may be counties under-resourced within the United States which could be overwhelmed by a pandemic such as COVID-19. Historically, there are examples of how a health system can be overwhelmed through a lack of population control (e.g. during the 1918 Spanish flu vide Ref B2). The team assessed that data analysis afforded by Python would be a useful tool in assessing which counties were at the highest risk for the purpose of resource allocation by decision makers, as well as for the informative purposes of educating potential travelers. Key learning outcomes identified were centered on data preparation (primarily through the use of Pandas) and visualizations (through the use of Folium).

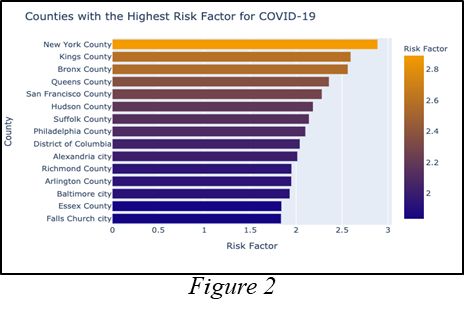
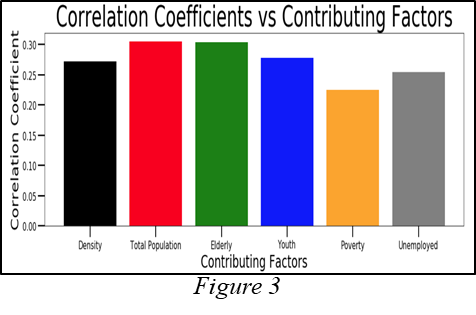
Workflow

The team collected data from the Census Bureau and Department of Homeland Security to create a data frame that allowed for statistical manipulation and analysis for each county in the United States. Through research and collaboration, the team utilized elderly and youth populations, number of hospital beds, unemployment percentage, and poverty as contributing factors to formulate a risk factor score for each county (See Figure 1). Network flow data from the Department of Transportation allowed the team to classify vulnerable entry points into the U.S. where additional population control measures can be implemented. Additionally, the team collected data daily from Johns Hopkins University for each Coronavirus case. This data evolved from first documenting cases in individual cities and then progressed to report larger and larger geographic regions, which demonstrated that adapting to data as it is changing is crucial to ensuring timely and accurate analysis. During the project, the team created nearly a thousand lines of code in several different Jupyter notebooks that created obstacles when the team members fused their work. As a result, the team learned that using descriptive and well annotated syntax is critically important for deconfliction and a shared understanding.

**

Analysis

Through the risk factor statistic, the team categorized those counties that are the least and most susceptible (See Figure 2) to an outbreak, which could enable timely decision-making and proper allocation of resources. By analyzing the cumulative count of Coronavirus cases, the team also identified that the number of cases in the U.S. is exponentially increasing. The team correlated Coronavirus cases per county against each of its contributing factors and determined that a statistical relationship is present (See Figure 3). The team modeled a potential scenario of county infections using its population and an understood transmittal rate against the county’s risk factor and the results proved promising. The results of this model combined with the correlation of each contributing factor highlighted potential areas for further exploration.

Conclusion

The team demonstrated that computational methods can lend insights into problems in a real-time environment. The evolution and treatment of Johns Hopkins’ data proved to be a valuable lesson that the team will carry with them in future projects. Combining code proved challenging at times and for future endeavors, leaning to use a Git framework would be beneficial especially if complexity were to increase. Python modules such as Pandas, Numpy, Plotly, and Folium, combined with Jupyter Notebooks allowed the team to analyze and present its findings to an audience in an informative and compelling manner. Interactive maps and plots allowed the audience to intuitively understand the team’s data and risk factor analysis. Overall, the team’s analysis is currently proving accurate. Densely populated counties faced with high populations of elderly and inadequate medical capacity face the greatest risks as this pandemic continues. Further consideration and evaluation of limited resources such as medical personal, expeditionary hospitals, and equipment will be vitally important in the coming days and weeks.