How Do Socioeconomic Factors Correlate to COVID-19 Cases and Deaths?

Anthony Rene Guzman, Jin Soung Yoo
Department of Computer Science
Purdue University Fort Wayne
Fort Wayne, IN, United States of America
Email: guzmar01@pfw.edu, yooj@pfw.edu

***Abstract*—The COVID-19 pandemic has spread around the world and had significantly affected every aspect of our day-to-day lives. Non-clinical socioeconomic factors may be important explanatory variables of COVID-19 cases and deaths. This work explores the correlations between various socioeconomic factors and the number of cases and deaths resulting from COVID-19. The study was conducted with county-level data from the U.S. Census Bureau and John Hopkins University, and examined the impact of ten different socioeconomic factors regarding population size, poverty, median household income, employment and education levels on COVID-19 prevalence across all counties in the United States. Correlation coefficients were computed between each of the socioeconomic factors and the total number of cumulative COVID-19 cases and deaths using various correlation analysis methods such as the Pearson, Kendall, and Spearman formulas. The results of the analyses echo the findings of similar research regarding COVID-19 and are visualized and discussed.**

Keywords- COVID-19; socioeconomic factors; correlation analysis.

I. INTRODUCTION

Coronavirus disease-19 (COVID-19) is a novel coronavirus that was first identified in Wuhan, China, in early December of 2019 [1]. Since its discovery, COVID-19 has affected world economies and lives across the globe. Due to the unknown and rapidly developing nature of the pandemic, governments, scientists and researchers have been working nonstop to develop a vaccine and understand more about this novel virus.

So far, old age and pre-existence of chronic conditions have been linked to more severe COVID-19 symptoms [2]. Other potential factors such as race/ethnicity and socioeconomic factors may also play an important role in the COVID-19 pandemic. The socioeconomic gradient in health is ubiquitous and has been described across pathologies, in life expectancy and mortality [3]. Low income might affect living conditions in many ways, such as residence in more deprived neighborhoods and housing conditions. A lower education level can be indirectly associated with several factors that may increase the risk of developing severe forms of COVID-19. However, the influence of socioeconomic factors on COVID-19 transmission, severity and outcomes is not yet known and is subject to scrutiny and investigation.

This study aims to identify any correlation between various socioeconomic factors and the number of cases and deaths resulting from COVID-19 across all counties in the United States. For this research, many socioeconomic data sets were considered and included various topics such as poverty, race, income, method of transportation, and more. The data for some factors, such as method of transportation to work and field of occupation, were not available on a granular level and were excluded from the analyses. Out of these many factors, ten were chosen to be considered for the correlation analysis. These socioeconomic factors were selected to correlate with COVID-19 cases and deaths due to their high data availability and novelty, Federal Information Processing Standards (FIPS) code-level data granularity, and prevalence in existing pandemic research. These factors are (1) total population size, (2) median household income, (3) population and percentage of population in poverty, (4) total employed population size, (5) total unemployed population size, (6) unemployment rate, (7) population and percentage of population without a high school diploma, (8) population and percentage of population with only a high school diploma, (9) population and percentage of population with some college education or an associate's degree, (10) population and percentage of population with a bachelor's degree or higher. This work explores each socioeconomic factor separately and analyzes the relationship with COVID-19 cases and deaths. The ten socioeconomic factors were grouped together based on the socioeconomic topic they fall under, which resulted in four groups: *population size*, *income and poverty*, *employment*, and *education*. Socioeconomic status might be one determinant that can tell us which regions are more vulnerable and high risk to the coronavirus disease. If we can find correlations between socioeconomic factors and COVID-19 outcomes, the findings will be useful in determining what regions may benefit most from a strategic allocation of health care resources.

This research used official county-level data sets gathered from the U.S. Census Bureau, U.S. Department of Agriculture (USDA) [4], and John Hopkins University [5]. The COVID-19 data sets provided by John Hopkins University contain time-series data for cumulative COVID-19 related cases and deaths, separated by county. Socioeconomic data sets provided by the U.S. Census Bureau and processed by the USDA contain data regarding education, population, poverty, and unemployment for each county of the United States in 2019.

The biggest challenge encountered when conducting this research was determining the best way to quantify the relationship between socioeconomic factors and COVID-19

 ISBN: 978-1-61208-891-4

cases and deaths. There are many different data mining tasks that could lead to interesting results, such as rule-based classification, time-series analysis, or clustering [6]. Another challenge of this research was determining how to correlate socioeconomic factors with COVID-19 cases and deaths. Quantitative values for socioeconomic factors are gathered infrequently and represent a value for a certain point in time (e.g., percentage of population in poverty in 2019). COVID-19 cases and deaths, on the other hand, take the form of a time-series with high variations in case and death frequency. These variations can be attributed to the time of the year, trends in social distancing, and changes in government ordinances. The last difficulty was that data on individual-level socioeconomic position are not being collected. The World Health Organization standard COVID-19 case report form only asks for each patient's age, sex/gender, place where the case was diagnosed, and usual place of residence. This impedes the sophisticated analysis of impact of socioeconomic factors.

This study uses Pearson, Kendall, and Spearman correlation coefficients in order to determine the relationships between socioeconomic factors and COVID-19 cases and deaths. By analyzing each socioeconomic factor separately, we investigated which factors most strongly correlate to COVID-19 cases and deaths. This research differentiates itself by using various correlation methods and is novel in the sense that the data being analyzed is more accurate due to its granularity. Similar research in this field has been done by using state-level data, but the data in this research is at the county level. Our analysis results offer some interesting findings. Visualization is also used to determine which geographical regions of the country are most vulnerable in the event of a pathogenic outbreak.

The following section describes the related literature. Section 3 describes the data and data preprocessing methods used. Section 4 describes correlation analysis and the methods used in detail. Section 5 describes the results of the research. Section 6 discusses the findings, and Section 7 concludes this paper.

II. Literature Review

Recently, there has been an abundance of research with respect to COVID-19 and data analysis. Applications for COVID-19 include COVID-19 detection and diagnosis, tracking and identification of the outbreak, infodemiology, biomedicine, and pharmacotherapy [7]. This section describes some related work in the field of statistically driven COVID-19 analysis.

Kurian et al. [8] evaluated the correlation between COVID-19 cases and Google Trends data on a state-by-state basis using the Pearson correlation method. Their findings suggest that certain keywords searched online were linked to high COVID-19 activity in the time leading up to spikes in cases, such as *face mask*, *Lysol*, and *COVID stimulus check*.

Shi et al. [9], using data on occupational position from 484 COVID-19 patients in Zhejiang Province of China, reported that severe cases were more likely to be agricultural workers and less likely to be self-employed than mild cases. Oh et al. [10] analyzed various socioeconomic factors of the South Korean population using a multivariable logistic regression model to determine if a lower socioeconomic standing increases the risk of contracting COVID-19.

Huang et al. [11] used k-means clustering ($k = 3$) to show how socioeconomically disadvantaged groups were disproportionately affected by stay-at-home orders. The results of their research showed that those who were more likely to stay at home were typically white, wealthy, well educated, and resided in regions with low unemployment and high median household income. Research conducted by Hawkins et al. [12] echoed these findings, where regions with low socioeconomic status are shown to have higher rates of COVID-19 cases and deaths.

Burton et al. [13] found that African Americans have been disproportionately affected by COVID-19. African American patients in the study were also found to be associated with older age, reside in a low-income area, and have a higher obesity rate. In turn, these socioeconomic factors contributed to their risk of contracting COVID-19. Gangemi et al. [14] also analyzed different socioeconomic factors to determine if significant correlations could be identified. They discovered that Gross Domestic Product (GDP) per capita and number of flights per capita were significantly correlated to the number of COVID-19 cases on a country-level basis.

Hatef et al. [15] developed an Area Deprivation Index (ADI) to measure a ZIP code's composite socioeconomic characteristics, using factors such as population size, age, gender, and race distribution. Populations in ZIP codes with higher ADI scores were found to be more at risk for COVID-19 related cases and deaths than those in ZIP codes with low ADI scores. Roser [16] found that the Human Development Index (HDI), a composite score that measures life expectancy, access to education, and standard of living, is significantly correlated to the number of COVID-19 cases in a country. The Distressed Communities Index (DCI) [17], a comprehensive estimate of a location's socioeconomic status, was used for analyzing COVID-19 case and fatality data in a Mann-Whitney U test.

III. Data and Data Preprocessing

This section describes the data and data preprocessing in detail.

A. Data

Two main data sets were gathered from several sources for use in the correlation analyses. The first main data set contains county-level COVID-19 cases and deaths data and was provided by the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at John Hopkins University [5]. The COVID-19 Data Repository is a publicly available data source that contains county-level time-series data for cumulative COVID-19 related cases and deaths that is updated daily. The data presents 3,340 counties and county-equivalents. The time series data begins on January 22, 2020 and is updated daily. For this research, we used the data entries up until March 27, 2021. This date was

 ISBN: 978-1-61208-891-4

selected to provide the most up-to-date correlation analyses. However, other dates such as thirty or ninety days after an outbreak has been identified may also be useful in determining how cases and deaths will continue to occur.

The second main data was provided by the U.S. Census Bureau and processed by the Economic Research Service of the U.S. Department of Agriculture [4]. There are 3,193 counties represented within this data set. We used the data of 2019. The data set contains county-level data for various socioeconomic factors, including poverty levels, median household income, population size, employment levels, and education levels. For this research, the socioeconomic factors analyzed are:

(1) total population size,
(2) median household income,
(3) population and percentage of population in poverty,
(4) total employed population size,
(5) total unemployed population size,
(6) unemployment rate,
(7) population and percentage of population without a high school diploma,
(8) population and percentage of population with only a high school diploma,
(9) population and percentage of population with some college education or an associate's degree,
(10) population and percentage of population with a bachelor's degree or higher.

B. Data Preprocessing

The data sets were first pre-processed for easier formatting. In order to eliminate confusion that may stem from how certain county names could be stored (e.g., "St. Mary's County" versus "Saint Marys County"), the data sources were merged based on FIPS codes [18], such that each FIPS code had a one-to-one relationship with each socioeconomic factor. FIPS codes were developed by the National Institute of Standards and Technology and are used to identify unique geographic regions such as states, counties, and county-equivalents (e.g., parishes, boroughs, and independent cities). Once the data sets were pre-processed and merged based on FIPS, the data sets were formatted and prepared for conducting data analysis.

IV. Analysis Methods

The primary goal of this study is to determine to what degree relationships exist between various socioeconomic factors and the number of COVID-19 cases and deaths. Correlation coefficients were used in this research to determine the degree to which these relationships exist.

Correlation is a value that describes the degree to which two variables are related [19]. A correlation coefficient falls within the range of -1 and 1. Correlation values closer to 1 describe a positive correlation, which means that as one variable increases, the other variable also increases. Correlation values closer to -1 describe a negative correlation, which means that as one variable increases, the other variable decreases. Correlation values close to 0 have little to no correlation, with a very weak or nonexistent pattern to the two variables [20]. The three correlation analysis methods used in this research were the Pearson, Kendall, and Spearman methods.

The Pearson correlation method is a parametric measure that produces a correlation coefficient *r*, which measures the direction and degree of relationships between pairs of variables. The Pearson correlation formula is shown below, where *r* is the correlation coefficient, x_i are the values of the first variable, $\bar{x}$ is the mean of the first variable, y_i are the values of the second variable, and $\bar{y}$ is the mean of the second variable [21]:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

The Kendall rank correlation method is a non-parametric measure that assesses the correspondence between the rankings of pairs of variables. The Kendall correlation formula is shown below, where *r* is the correlation coefficient, *c* is the number of concordant pairs, *d* is the number of discordant pairs, and *n* is the total number of points. For a pair of points (x_i, y_i) and (x_j, y_j), the pair is said to be concordant if $x_i > x_j$ and $y_i > y_j$, or if $x_i < x_j$ and $y_i < y_j$. If neither condition is true, that pair is said to be discordant [22]:

$$r = \frac{(c - d)}{\binom{n}{2}}$$

The Spearman rank correlation formula is another non-parametric measure that determines the correlation between the ranks of pairs of variables. The Spearman correlation formula is shown below, where *r* is the correlation coefficient, d_i is the difference between the two ranks of each point, and *n* is the number of points [23]:

$$r = \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Fahrudin et al. [24] suggest that the Kendall and Spearman correlation analysis methods are more appropriate for this type of research because they are non-parametric methods. These types of methods do not make any assumptions about the model of the data and can operate on incomplete data, making them better suited for uncertain and rapidly evolving events like pandemics.

V. Results

Correlation coefficients were computed between each of the socioeconomic factors and the total number of cumulative COVID-19 cases and deaths. Because correlation formulas determine the degree of association between two variables, the coefficients were calculated using one socioeconomic factor at a time. For each analysis, the Pearson, Kendall, and Spearman correlation methods were used. When discussing the results of the correlation analyses, we used the average of three correlation coefficients. Table 1

 ISBN: 978-1-61208-891-4

is used to describe the strength of the correlation coefficients [25]:

TABLE I. CATEGORIZATION OF CORRELATION COEFFICIENTS

Correlation Coefficient (*r*)	Degree of Association
[0.8, 1.0]	Very Strong (+)
[0.6, 0.8)	Strong (+)
[0.4, 0.6)	Medium (+)
(0.4, 0)	Weak (+)
(0, -0.4)	Weak (-)
[-0.4, -0.6)	Medium (-)
[-0.6, -0.8)	Strong (-)
[-0.8, 1.0]	Very Strong (-)

A. *Relationship with Population Size*

First, we examined the correlation with the socioeconomic factor of (1) total population size. Total population size with regards to COVID-19 cases has correlation coefficients of 0.970086 (Pearson), 0.842793 (Kendall), and 0.953558 (Spearman) with an average of 0.922146. With regards to COVID-19 deaths, the correlation coefficients are 0.930995 (Pearson), 0.727279 (Kendall), and 0.888893 (Spearman) with an average of 0.849055. The relationship between population size and COVID-19 cases and deaths has a very strong positive correlation.

B. *Relationship with Income and Poverty*

Second, we examined the correlation with the socioeconomic factor of (2) median household income with COVID-19 cases and deaths. Median household income with regards to COVID-19 cases has correlation coefficients of 0.220982 (Pearson), 0.220541 (Kendall), and 0.324824 (Spearman) with an average of 0.255449. With regards to COVID-19 deaths, the correlation coefficients are 0.188385 (Pearson), 0.144611 (Kendall), and 0.214694 (Spearman) with an average of 0.182564. Both show weak positive correlations. Figure 1 shows the relationship between COVID-19 cases and median house income. The socioeconomic factor is mapped in gray, with higher values being more gray and lower values being less gray. The number of COVID-19 cases by county are mapped in green, with counties having many cases being greener than counties with fewer cases. As shown in Figure 1, median household income is not a great indicator of the probability of contracting COVID-19. There are some regions with higher average household income that are greener, but not by a significant margin.

Third, we examined the correlation with the socioeconomic factor of (3) population and percentage of population in poverty. The population size in poverty with regards to COVID-19 cases has correlation coefficients of 0.956651 (Pearson), 0.776574 (Kendall), and 0.925292 (Spearman), with an average of 0.886172. With regards to COVID-19 deaths, the correlation coefficients are 0.93817 (Pearson), 0.731511 (Kendall), and 0.893441 (Spearman), with an average of 0.854374. Both show very strong positive correlations. In contrast, the percentage of a population in poverty with regards to COVID-19 cases has average correlation coefficients of -0.104471 and -0.044366, respectively, which show weak negative correlations.

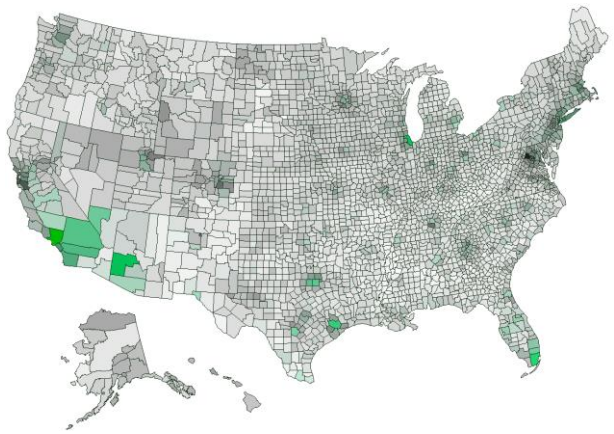

Figure 1. Relationship of Median House Income and COVID-19 Cases

C. *Relationship with Employed and Unemployed*

Fourth, we then examined the correlation with the socioeconomic factor of (4) total employed population size. Employed population size with regards to COVID-19 cases has correlation coefficients of 0.964435 (Pearson), 0.824678 (Kendall), and 0.945477 (Spearman) with an average of 0.911530. With regards to COVID-19 deaths, the correlation coefficients are 0.921836 (Pearson), 0.704132 (Kendall), and 0.872068 (Spearman) with an average of 0.832679. Both show very strong positive correlations.

Fifth, we examined the correlation with the socioeconomic factor of (5) total unemployed population size. Unemployed population size with regards to COVID-19 cases has correlation coefficients of 0.967212 (Pearson), 0.795141 (Kendall), and 0.933167 (Spearman) with an average of 0.898507. With regards to COVID-19 deaths, the correlation coefficients are 0.942800 (Pearson), 0.707763 (Kendall), and 0.875330 (Spearman) with an average of 0.841964. Both also show a very strong positive correlation.

Sixth, we examined the correlation with the socioeconomic factor of (6) unemployment rate. A region's unemployment rate with regards to COVID-19 cases has correlation coefficients of -0.043564 (Pearson), -0.013928 (Kendall), and -0.012187 (Spearman) with an average of -0.023227. With regards to COVID-19 deaths, the correlation coefficients are -0.018840 (Pearson), -0.025321 (Kendall), and 0.042350 (Spearman) with an average of 0.016277. They show a weak negative and weak positive correlation, respectively.

D. *Relationship with Education*

Seventh, we examined the correlation with the socioeconomic factor of (7) population and percentage of population with less than a high school diploma. The population size without a high school diploma with regards to COVID-19 cases has correlation coefficients of 0.965225 (Pearson), 0.764172 (Kendall), and 0.918743 (Spearman), with an average of 0.882713. With regards to COVID-19

 ISBN: 978-1-61208-891-4

deaths, the correlation coefficients are 0.945272 (Pearson), 0.734748 (Kendall), and 0.897835 (Spearman), with an average of 0.859285. Both show a very strong positive correlation. In contrast, the percentage of a population without a high school diploma with regards to COVID-19 cases and deaths has average correlation coefficients of -0.041244 and 0.017281, which show a weak negative and weak positive correlation, respectively.

Eighth, we examined the correlation with the socioeconomic factor of (8) population and percentage of population with only a high school diploma. The population size with only a high school diploma with regards to COVID-19 cases has correlation coefficients of 0.959932 (Pearson), 0.812512 (Kendall), and 0.943194 (Spearman), with an average of 0.905213. With regards to COVID-19 deaths, the correlation coefficients are 0.938317 (Pearson), 0.735719 (Kendall), and 0.89646 (Spearman), with an average of 0.856832. Both show a very strong positive correlation. In contrast, the percentage of a population with only a high school diploma with regards to COVID-19 cases and deaths has average correlation coefficients of -0.262308 and -0.208336, which show a weak negative and weak positive correlation, respectively.

Ninth, we examined the correlation with the socioeconomic factor of (9) population and percentage of population with some college education or an associate's degree. The population size with some college education or an associate's degree with regards to COVID-19 cases has correlation coefficients of 0.961137 (Pearson), 0.809469 (Kendall), and 0.938058 (Spearman), with an average of 0.902888. With regards to COVID-19 deaths, the correlation coefficients are 0.909433 (Pearson), 0.704526 (Kendall), and 0.871939 (Spearman), with an average of 0.828633. Both show a very strong positive correlation. In contrast, the percentage of a population with some college education or an associate's degree with regards to COVID-19 cases and deaths has average correlation coefficients of -0.113840 and -0.141439, respectively, which show weak negative correlations.

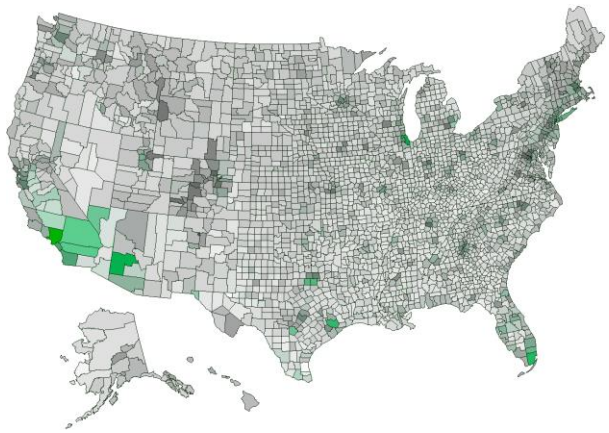

Figure 2. Relationship of percentage of population with a bachelor's degree or higher and COVID-19 Cases

Lastly, we examined the correlation with the socioeconomic factor of (10) population and percentage of population with a bachelor's degree or higher. The population size with a bachelor's degree or higher with regards to COVID-19 cases has correlation coefficients of 0.910476 (Pearson), 0.767488 (Kendall), and 0.912052 (Spearman), with an average of 0.863339. With regards to COVID-19 deaths, the correlation coefficients are 0.882014 (Pearson), 0.658304 (Kendall), and 0.831540 (Spearman), with an average of 0.790619. They show a very strong positive and strong positive correlation, respectively. In contrast, the percentage of a population with a bachelor's degree or higher with regards to COVID-19 cases and deaths has average correlation coefficients of 0.290129 and 0.225420, respectively, which show weak positive correlations. As shown in Figure 2, there is no significant relationship between the percentage of a population with a bachelor's degree or higher and COVID-19 cases and deaths.

VI. DISCUSSION

By analyzing each socioeconomic factor separately, we could determine which factors most strongly correlate to COVID-19 cases and deaths. The strongest correlations among the socioeconomic factors used in the data analysis are *total population size*, *population with only a high school diploma*, and *total employed population size*. The weakest correlations among the socioeconomic factors used in the data analysis are *unemployment rate*, *percentage of population with less than a high school diploma*, and *percentage of population in poverty*. No strong negative correlations were found from the analyses.

While the results of the correlation analyses are mostly expected, they did offer some surprising findings. The correlation coefficients are very similar between total employed population size and total unemployed population size. This could suggest that those that who are employed have roughly the same likelihood of contracting COVID-19 as those who are unemployed. The industry in which those employed work also plays a significant role. For example, someone who uses public transportation and travels to the workplace has a much higher chance of contracting COVID-19 than someone who works from home due to a constant proximity to people outside of their residence. This idea is further supported by the fact that there is no correlation between a region's unemployment rate and the number of COVID-19 cases and deaths.

From the results of the analyses, it is evident that population density is a very strong indicator for the severity of impact caused by COVID-19 in each region. This can be explained by the fact that a region with a higher population density is more prone to airborne transmission of COVID-19. Regions of the country with a higher population density will have more people contributing to economic activity (e.g., people working, shopping, eating out, running chores).

The results of the analyses show that the level of education a person has obtained has little correlation with COVID-19 cases and deaths. This suggests that a person without a high school education has roughly the same

 ISBN: 978-1-61208-891-4

likelihood of contracting COVID-19 as someone with a bachelor's degree or higher. Although jobs that require a bachelor's degree or other form of higher education are more likely to offer work-from-home solutions, these solutions have little bearing on the likelihood of contracting COVID-19.

Lastly, the results show that population-based socioeconomic factor values provide much stronger correlation coefficients than percentage of population values. This can be attributed to the fact that cumulative COVID-19 cases and deaths are represented by total counts, while percentages are orders of magnitudes smaller and represent a fraction of the population, rather than the individualistic counts. Further research in this field would benefit most from matching data types. In other words, total COVID-19 cases and deaths should be analyzed with the total number of people with a certain socioeconomic factor. Conversely, socioeconomic factors represented as percentages of a population should be analyzed with percentages of a population affected by COVID-19.

VII. CONCLUSION

The effects of COVID-19 have been far-reaching and ubiquitous; no public health event has taken such a toll on the global community for over a century. This work investigated correlations between various socioeconomic factors and the number of cases and deaths resulting from COVID-19 in the United States. Furthermore, this research showed how statistical computing and visualization can help determine which geographical regions of the country are most vulnerable in the event of a pathogenic outbreak. The information gained from this study will be useful in determining the proper distribution of medical resources when the next pandemic inevitably strikes.

It is imperative that data analysis is further conducted on COVID-19. For future works, this research can be continued by performing rule-based association analysis, which can determine which subsets of variables are most strongly associated with COVID-19 related cases and deaths.

REFERENCES

[1] G. Spiteri et al., "First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020", *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* vol. 25,9: 2000178. doi:10.2807/1560-7917.ES.2020.25.9.2000178, pp. 2-7, 2020.

[2] R. H. Shmerling, "COVID-19: If You're Older and Have Chronic Health Problems, Read This." *Harvard Health Publishing*, The President and Fellows of Harvard College. URL : www.health.harvard.edu/blog/covid-19-if-youre-older-and-have-chronic-health-problems-read-this-2020040119396. 2020. [retrieved : September, 2021]

[3] R. G. Wilkinson and K. E. Pickett, "Income inequality and socioeconomic gradients in mortality." *American journal of public health* vol. 98,4, doi:10.2105/AJPH.2007.109637, pp. 699-704, 2008.

[4] J. Pender and E. A. Dobis, Economic Research Service County-level Data Sets. *U.S. Department of Agriculture*. 2021. [retrieved: September, 2021]

[5] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time". Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1, pp. 544-534, 2020.

[6] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques*"*. Morgan Kaufmann Publishers. 2001.

[7] C. Hertzman, "Putting the concept of biological embedding in historical perspective". Proc Natl Acad Sci U S A. 109 Suppl 2, pp. 17160-17167, 2012.

[8] S. J. Kurian, "Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis". *Mayo Clinic Proceedings, 95*(11), pp. 2370-2381, 2020.

[9] Y. Shi et al., "Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan". Critical Care, 24:108. 2020.

[10] T. K. Oh, J. Choi, and I. Song, "Socioeconomic disparity and the risk of contracting COVID-19 in South Korea: an NHIS-COVID-19 database cohort study," *BMC Public Health*, vol. 21, pp. 1-12, 2021.

[11] X. Huang et al., "Time-Series Clustering for Home Dwell Time during COVID-19: What Can We Learn from It?". *ISPRS International Journal of Geo-information, 9*(11), 675, 2020.

[12] R. B. Hawkins, E. J. Charles, and J. H. Mehaffey, "Socio-economic status and COVID-19–related cases and fatalities". *Public Health (London), 189*, pp. 129-134, 2020.

[13] J. Burton, D. Fort, and S. Leonardo, "Hospitalization and Mortality among Black Patients and White Patients with Covid-19," *N. Engl. J. Med.*, vol. 382, (26), pp. 2534-2543, 2020.

[14] S. Gangemi, L. Billeci, and A. Tonacci, "Rich at Risk: Socio-Economic Drivers of COVID-19 Pandemic Spread." *Clinical and Molecular Allergy CMA*, vol. 18, no. 1, 2020, pp. 1–12.

[15] E. Hatef, H. Chang, C. Kitchen, J. P. Weiner, and H. Kharrazi, "Assessing the Impact of Neighborhood Socioeconomic Characteristics on COVID-19 Prevalence Across Seven States in the United States". *Frontiers in Public Health, 8*, 571808, 2020.

[16] M. Roser, "Human Development Index (HDI)." *Our World in Data*, Global Change Data Lab, 25 July 2014, https://ourworldindata.org/human-development-index. [retrieved: September, 2021]

[17] *Economic Innovation Group Distressed Communities Index*. URL: http://eig.org/dci. [retrieved: September, 2021]

[18] J. Holland, "ANSI (FIPS) Codes for Metropolitan and Micropolitan Statistical Areas." *Data.gov Data Catalog*, US Census Bureau, Department of Commerce, 11 Mar. 2021, catalog.data.gov/dataset/ansi-fips-codes-for-metropolitan-and-micropolitan-statistical-areas.

[19] W. M. K. Trochim, "Correlation." *Conjoint.ly*, Analytics Simplified Pty Ltd. URL: conjointly.com/kb/correlation-statistic/. [retrieved: September, 2021]

[20] W. M. LaMorte, "The Correlation Coefficient (r)." *Evaluating Association Between Two Continuous Variables*, Boston University School of Public Health, 21 Apr. 2021, sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module9-Correlation-Regression/PH717-Module9-Correlation-Regression4.html.

[21] "Pearson's product moment correlation", *Laerd Statistics*, Lund Research Ltd. URL: https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php. 2020. [retrieved: September, 2021]

[22] “Kendall tau metric”. *Encyclopedia of Mathematics*. URL: http://encyclopediaofmath.org/index.php?title=Kendall_tau_metric&oldid=51572. 2021. [retrieved: September, 2021]

[23] S. Glen, "Spearman Rank Correlation (Spearman’s Rho): Definition and How to Calculate it", *StatisticsHowTo.com: Elementary Statistics for the rest of us!*. URL: https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/spearman-rank-correlation-definition-calculate/. 2021. [retrieved: September, 2021]

[24] T. Fahrudin, D. R. Wijaya, and A. A. Agung, “COVID-19 Confirmed Case Correlation Analysis Based on Spearman and Kendall Correlation”. *2020 International Conference on Data Science and Its Applications (ICoDSA),* 1-4, 2020.

[25] Kent State University Libraries. (2021, Mar 22). *SPSS tutorials: Pearson Correlation.* URL: http://libguides.library.kent.edu/SPSS/PearsonCorr [retrieved: September, 2021]

 ISBN: 978-1-61208-891-4