6-27-2022

DEVELOPMENT OF A COVID-19 VULNERABILITY INDEX (CVI) FOR THE COUNTIES AND RESIDENTS OF NEW JERSEY

Remo Victor DiSalvatore
Rowan University

DEVELOPMENT OF A COVID-19 VULNERABILITY INDEX (CVI) FOR THE COUNTIES AND RESIDENTS OF NEW JERSEY

By

Remo Victor DiSalvatore

A Thesis

Submitted to the
Department of Civil Engineering
College of Engineering
In partial fulfillment of the requirement
For the degree of
Master of Science in Civil Engineering
at
Rowan University
May 3, 2022

Thesis Chair: Sarah K. Bauer, Ph.D., Assistant Professor, Department of Civil and Environmental Engineering

Committee Members:
Jeong Eun Ahn, Ph.D., Assistant Professor, Department of Civil and Environmental Engineering
Jahan Kauser, Ph.D., P.E., Professor and Head, Department of Civil and Environmental Engineering

Dedications

I would like to dedicate this thesis to my thesis advisors, Dr. Bauer, Dr. Ahn and Dr. Jahan for helping me tremendously throughout this process. They have provided their expertise and guidance through the course of this research. Without them I would not have had the opportunity to pursue my professional career and continue academics and for that I am extremely grateful.

I would also like to thank my family and friends for their emotional and physical support through these trying times. Without them I would not be where I am today and for that I am eternally grateful.

Abstract

Remo Victor DiSalvatore
DEVELOPMENT OF A COVID-19 VULNERABILITY INDEX (CVI) FOR THE COUNTIES AND RESIDENTS OF NEW JERSEY
2021-2022
Sarah K. Bauer, Ph.D.
Master of Science in Civil Engineering

The COVID-19 pandemic has impacted countless aspects of everyday life since it was declared a global pandemic by the World Health Organization in March of 2020. From societal to economic impacts, COVID-19 and its variants will leave a lasting impact on our society and the world. Approximately $9 trillion has been spent on fighting the pandemic around the world. During the pandemic, it became increasingly evident that indices, such as the Center for Disease Control (CDC) Social Vulnerability Index (SVI), were extremely important for predicting vulnerabilities in a community. The CDC's SVI provides important estimates on which communities will be more susceptible to 'hazard events' by compiling a variety of data from the U.S. census, as well as data from the American Community Survey. The SVI does not necessarily consider the susceptibility of a community to a Global Pandemic such as COVID-19. Thus, the objectives of this research were to develop a COVID-19 Vulnerability Index (CVI) to evaluate the community's susceptibility to future pandemics. The CVI was validated by comparing to real world COVID-19 data from New Jersey's 21 counties. The results of this study indicate that Essex County had the highest CVI, and Hunterdon County had the lowest CVI. This is due to factors such as disparity in wealth, population density, minority status, housing conditions and several other factors that were used to compose the CVI.

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

List of Figures

List of Tables

Chapter 1

Introduction

With the world facing uncertainty in the midst of a global pandemic, many people have turned to statistics in order to interpret data and understand the severity of the COVID-19 pandemic. Facets of everyday life have seen a change in response to the pandemic; some of these being seen from societal and economic perspectives. With a $9 trillion-dollar global expense and millions of lives lost, everyday life has changed drastically since COVID-19 had been declared a global pandemic by the World Health Organization in March of 2020 [1]. With a bleak future on the horizon, it became increasingly evident that there is a need to determine which communities are most affected and in turn vulnerable to COVID-19. The utilization of precedents and historical data has never been more important. By turning to the resources available such as the CDC's Social Vulnerability Index (SVI) [2] it is possible to create a COVID-19 Vulnerability Index (CVI).

The need for a CVI became evident due to the precedent set by COVID-19 as mentioned above. With millions of people infected in New Jersey alone, it became imperative that each county of New Jersey be assessed and given a statistically determined vulnerability evaluation in relation to COVID-19. As mentioned previously, the CDC's SVI is a great asset for the people of the U.S., though, there needs to be a distinction between the communities susceptible to natural disasters and to global pandemics. Even though the SVI may not provide the most accurate data in terms of pandemic vulnerability, portions of it can still be utilized as an asset for a CVI. The SVI has been validated in the literature; therefore, it is the best baseline for establishing the framework for a comprehensive CVI.

The Center for Disease Control (CDC) established the SVI in 2011, with the purpose of determining which communities of the U.S. are more susceptible to public health emergencies, and therefore, will need more resources allocated from the federal and state level [2]. Traditionally, public health emergencies involve consideration of natural disasters and how long a municipality and its economy will take to fully recover from a natural disaster. The SVI ranks every municipality in the U.S. that data has been gathered from and ranks them on a scale from 0 to 1 (i.e., least to most vulnerable). The SVI consists of four primary themes and multiple subcomponents related to each theme which contribute to the calculated SVI score. Themes are related to Socioeconomic Status, Household Composition and Disability, Minority Status and Language, and Housing Type and Transportation where each subcomponent is a more specific value related to the general theme. The SVI is updated every two years and utilizes data that is gathered from the U.S. Census Bureau as well as the American Community Survey (ACS) [3].

The SVI was developed in response to climate change and a need to determine which areas will be most afflicted in the occurrence of a 'disaster event'. Factors, such as emergency personnel, food, water, medical supplies, and other forms of relief, can be determined for each municipality in the U.S. based on the SVI values determined by the CDC. The SVI helps save lives in preparation for events, such as hurricanes, floods, and other life-threatening occurrences. By having an established index in place, there is a better general understanding of which communities will need the most aid. Having an index will not only help a community rebuild after a disaster but can also save lives. An index can be seen as a tool that communities possess, such as FEMA flood maps or Emergency Action Plans, which can be quickly referred to in order to determine the number of people affected and the courses of action that need to be taken. The establishment of the CVI is paramount in consideration of public health and welfare.

The CVI and SVI share many similarities and differences. One similarity being that they are both preventative measures/tools for communities to utilize in the event of a disaster (e.g., a global pandemic). By understanding which communities will be most at risk it is possible to enable preventative measures and awareness to those residents, as well as prepare preliminary relief.

The goal of this research was to create an index representing the level of vulnerability to future pandemics. New Jersey has the highest population density in the United States making it an extremely unique candidate case study [3] . New Jersey also happens to share borders with New York City, and Philadelphia, thus increasing the amount of traffic and populus commuting across state lines into densely populated cityscapes; thus, New Jersey was used as a representative state for a case study in this research. In the CVI, the level of vulnerability for each county will be represented by a number scaling from 0 to 1, least vulnerable to most vulnerable, respectively. The higher the SVI value, the more relief (policy making for emergency responder allocation) that can be anticipated for that county and thus the faster and more efficient the state and federal government can respond to aiding those areas. With an average population density of over 1,200 people per square mile, New Jersey has 21 counties ranging from 187 people per square mile (Salem County) to 14,568 (Hudson County) [3]. This makes it extremely critical for determining which New Jersey counties are at the highest risk for the spread of COVID-19.

Due to the nature of how the virus is spread, a higher population density directly relates to how many people may be in contact with one another, and thus, increase the likelihood of spreading the virus [18]. Though practices such as social distancing, contact tracing, and mask mandates have mitigated the spread of the virus in extremely populated urban areas. Therefore, it is important to gather a wide variety of variables that can contribute to how vulnerable a population or county in this case is to the transmission and mortality rates associated with COVID-19. However, the U.S. was much less restrictive with travel and social interaction compared to various other countries around the world [8]. Therefore, population density is going to be a rather important factor in determining the vulnerability of a county to COVID-19. Being a part of the 'Tristate area', New Jersey is considered part of New York's metropolitan area. The northern counties, share borders with New York City, NY and population-rich metropolitan areas; many people from these areas live in New Jersey for affordable housing and commuting purposes thus increasing the infection rate and transmissibility of COVID-19. The same can be applied to Western counties that border Philadelphia, PA. New Jersey is considered as a case study for these reasons, along with some of the Environmental hazards that are comprised within it such as having the highest number of Superfund sites in the country [28].

Research Objectives

The overall objective of this research was to develop a COVID-19 Vulnerability Index (CVI) and use New Jersey counties to validate the index. This research has two aims:

1) To develop a COVID-19 Vulnerability Index (CVI) to aid in preparedness for future global pandemics, and
2) To incorporate New Jersey county data into the CVI in order to validate the index.

By achieving these research objective, it research builds on the established field of indices and provides a better understanding of determining which communities are more susceptible to the COVID-19 pandemic, and future global pandemics. This study focused on developing a COVID Vulnerability Index for the counties of New Jersey by predominantly utilizing the existing framework and methodology from that the SVI has been developed. Previous studies [2,16,38,39], as well as publicly available data sets, were utilized in the development of the CVI.

Chapter 2

Literature Review

Concept of Established Indices

Indices have been around for centuries aiding in the research and development of cities and healthy populations [12]. The data composes an index depends on the need being filled at the time and the available data that is to be utilized in index construction. The need for an index may change over time depending on the technology available and the issues that are relevant to a population. Some examples of indices being stock, health and air quality indices. An index determines values by establishing a relative relationship between values in a dataset. These 'values' determine the ranking of the index. Data is normalized several times to make direct comparisons. By determining the index values in relation to one another, one can rank values on a scale in respect from least to greatest value. One example of this is the index set in place by the Center for Disease Control (CDC). The CDC's Social Vulnerability Index (SVI) was created to determine communities that are more vulnerable to disaster events such as hurricanes and floods [3]. This is done by ordering values on a scale from 0 to 1 (least to most vulnerable). By establishing a COVID Vulnerability Index (CVI), it is possible to measure vulnerability of a community to COVID-19. The applications of a CVI are similar to that of the SVI except instead of relief for a disaster event, like a hurricane, it would be in relation to the COVID-19 pandemic and future pandemics.

The foundation for building indices that incorporate socioeconomic, health, and other important values predominantly refers to the SVI. The SVI incorporates four themes that all attribute to a single score, which represents a population's susceptibility to a natural disaster occurring [2]. The reconstruction phase after a natural disaster can be an arduous process. The period that it takes a community to recover from an event depends on numerous factors. Though some of the most important factors are those that relate to socioeconomic status [10]. Communities that have a higher socioeconomic status can fund repairs and ensure that their community is operational as soon as possible. Where communities that may not have the extra fluid capital are reliant on government-funded campaigns for financial support. These campaigns are not time conscious and in some cases it may take years for money to transpire. Furthermore, impoverished communities have fewer facilities at their disposal than their more affluent counterparts. This is usually reflected in access to transportation, health care and numerous facets of general infrastructure.

The common stance that most of the current methods follow is that of developing a similar system to the CDC’s SVI [2]. Following this template is a very common practice in this field of study, particularly to the close relationship between the study of populations regarding specific themes and/or characteristics. Many studies relate SVI values and CVI values due to many of the factors that are used to compose the SVI [16,19]. The template that is followed is most often that of contributing various themes in generating a CVI. When these themes are added together, they can tabulate a specific value that indicates the susceptibility of a population to an outside occurrence, such as a natural disaster or global pandemic, such as the COVID-19 pandemic. This method has been proven to be effective in determining a specific population’s susceptibility and/or vulnerability to natural disasters as seen in current literature [2,16,38,39]. By using the framework and adjusting specific attributes such as relevant themes and factors that compose those themes; and thus, the CVI has a healthy framework that has been proven effective. Otherwise, the themes and their respective values can be determined by the area of research that is of question or interest of research. In other words, the composition of the themes and factors is entirely subjective, though should go through a series of steps to ensure validity [11]. Some want to analyze the country or global trends of the virus and its susceptibility [16] whereas others seek to identify the impacts on smaller entities such as municipalities and counties as seen by [14,15].

Different methodologies developing the CVI can be determined depending on the area of study and/or the objective of the research an example of this is Machine Learning utilizing the capacity of Artificial Intelligence to automatically compile real-time data in the autonomous computation of real time CVI values [13]. Machine learning has a wide range of applications and its application to this area of study is remarkably resourceful.

Concept of a CVI

There are few studies that pertain to the development of a comprehensive CVI. Most studies are relatively unproven considering the recentness of the pandemic. With the relevance and need for the development of a CVI, there is a scramble to determine which populations are most vulnerable to said pandemic and to proceed accordingly [2,4,9,17, 29-33]. The current studies range in focus from county and/or country level to specific municipalities. Comparisons need to be made from different communities and populations to hopefully draw conclusions as to who is most susceptible to COVID-19 and therefore further actions can be taken to ensure the fair distribution of aid amongst the afflicted population(s).

Many different approaches have been taken in hopes of developing the most accurate and reputable CVI [13-16,22,38,39]. In the creation of an index related to the susceptibility of a community to outside factors, many researchers have turned to the core principles established by the CDC's SVI [2]. These core principles have aided in establishing the fundamental foundation that contributed to the current understanding and development of a CVI. The SVI established an important baseline for taking what others may see as random data and organizing it into a quantitative index. Over the years, the SVI has been more refined and can accurately portray municipalities and counties all around the U.S. Though the SVI is a useful tool for determining the susceptible communities to natural disasters, it is not aimed at determining which communities will be at greater risk to pandemics. There has been a need in the current field of research to determine which communities will be impacted most by COVID-19 and a need for an analysis of what resources can be expected to help alleviate issues in these communities.

Relationships have been established between a community's socioeconomic status and a number of other factors that can attribute to the susceptibility of a community to a natural disaster. Through the CDC's SVI, based on a wide range of factors, a determination can be made as to how long it will take for a community to recover from a disaster event. At first, it was assumed that local topography and geography was the only concern that should be considered in such events. Over time, it has been proven that the composition of a community in relation to the factors established by the SVI will determine the resilience of the said community in the event of a disaster. With that being said, a CVI may take many different shapes and analyze different properties over time before a more refined version is created.

The scope of the current literature is rather wide and there has not been a determination as to which methodology is the most accurate in determining a susceptible population to COVID-19. Numerous attempts have been made at the creation of an unbiased CVI utilizing different areas of study and resources. The primary concepts that available literature highlights are the different methods that are being implemented at different scales around the world. Amongst the varying methods of creating a CVI, there are several methods that seem the most relevant and pioneering. Methods include AHP (Analytic Hierarchy Process), generalized propensity scoring being utilized in conjunction with time-varying data and machine learning, etc. [14, 15, 16] have been applied to CVI generation. Through the highly diverse areas of methodology, the one objective that they all share is to establish a better understanding of the COVID-19 pandemic and the susceptibility of different populations to it regardless of the methodology [13-16,22,38,39].

As previously mentioned, there are several steps that need to be taken to develop a comprehensive CVI. The CDC's SVI, as mentioned previously, has paved the way in terms of research and development for the generation of what an Index of this magnitude should look like. By this, they have established a fundamental framework by which this research can follow. This includes the use of publicly available datasets for the counties of New Jersey as well as theme development. The motivation behind this study is to hopefully generate public awareness for the counties of New Jersey that may have a more vulnerable population as well as generate an unbiased index for the state of New Jersey for the reasons mentioned previously. The inspiration for this study originally came from the University of Maryland's publication on COVID-19 and the relationship between racial inequalities [9]. There was mention of the creation of an index that mapped out contributing factors to

COVID-19 Vulnerability. The research involved in said study involved the supplemental inclusion of a community's demographics into the consideration of a CVI. Several articles have already proven that COVID-19 and other disasters affect minority groups disproportionately [10,29,30,31].

The first primary reference to CVI construction is the Surgo Foundation's development of a Community COVID Vulnerability Index (CCVI) [16]. Hereinafter, this study will be referred to as Surgo's CCVI. Surgo's CCVI aimed to utilize the SVI's proven methodology of determining susceptibility to natural disasters and rearrange the themes into those that suit the needs of a CVI as opposed to an SVI. Surgo's CCVI compares the counties of the U.S. to one another similarly to the SVI and rates them using the same methodology. Surgo's CCVI has not yet published their complete results on validation as of 21, April 2022. However, this methodology has been followed in this report predominantly because the CDC references it in their COVID vulnerability section. Though Surgo has established a comprehensive CVI, their validation results have not yet been published. This is where the most trouble comes from within this new territory of index creation. Where the SVI has had the better part of a decade to determine whether its claims are accurate in relation to disaster events and how long communities took to recover, the CVI has only had a year and several months to validate the effectiveness of its determinations. The Surgo Foundation was one of the first to publish a complete CVI or in their case, a CCVI (Community COVID-19 Vulnerability Index). The study used a similar methodology from the SVI for it has been proven over time and adapted it to relate to the pandemic as opposed to natural disaster susceptibility. The Surgo Foundation's work has given tremendous insight as to how the themes and the different variables regarding a CVI should be organized and what methodology should be employed when dealing with such an undertaking.

The methods of weighting as well as the current level of established validation methods are methods that have been developed through various assets of current literature. As seen in the CDC's SVI, Surgo's CCVI and other studies [2,16,38,39], even weighing factors are applied for numerous reasons. Some of them being that there need to be Subject Matter Experts (SME's) involved in determining weight values. In terms of validation, some researchers compare their findings with the data that has been collected through public health records such as the mortality and infection data represented in each area of study and/or with R-value deviation from established indices such as the SVI. As seen in some composite index development documents [21], the development and validation of an index is completely subjective and up to the researcher or direction of research. Though currently, there is no singularly established method of validation, one suggestion would be to analyze the individual areas of studies on a smaller scale such as counties per state and observe the infection and mortality data per population and analyze whether the CVI values relate to higher percentages of mortality or infections per population. This would be the most comprehensive way to compare hypothesized values to real-world data. Though, the validity of the comparison would heavily depend on the accuracy of the data collected from each respective county, state, or municipality. The different themes that were applied to the index help cover the broad spectrum of factors that all contribute to Covid Vulnerability. These factors were predominantly seen in the SVI and adjusted for the creation of a CVI. Many studies followed the same methodology in the creation of a CVI such as the Surgo Foundation's CCVI.

Many articles described how to analyze the percent of the population that may be infected through methods such as susceptible infected (SI) and susceptible exposed infected removed (SEIR). Others seek verification by comparing CVI and SVI values to one another due to articles published on comparisons [19,20]. (Though with an index that is thorough enough, there may be less reliance on analyzing real-time data and rather, a scramble to allocate funds and resources to communities that have already been determined to be at a higher risk.) The key difference between said studies and what is being presented is that an index based on a virus is effectively a snapshot of the future if done correctly.

A composite index considers multiple variables, normalizes the data, and relates values to one another. To create a composite index, a guide that has been established by OECD can be followed [11]. This guide maps out the different points of interest that need to be followed when establishing a composite index. This guide maps out ten steps that should be followed when establishing said index. These steps are as follows: Theoretical Framework, Data Selection, Imputation of Missing Data, Multivariate Analysis, Normalization, Weighing and Aggregation, Robustness, and Sensitivity, Back to the Real Data, Links to Other Variables, and, Presentation and Visualization [11]. These steps aid in establishing the missing framework from that of the CCVI and SVI.

Generally, there are many different avenues to follow when developing an index. Depending on the type of index being developed and the ultimate purpose that it fulfills, will determine the methodology that is used and developed in the research. In this case as stated above, OECD, SVI and CCVI were all key inspirations in developing the index found in this research. This includes weighting, aggregation, and several other methodological choices.

Chapter 3

Development of a COVID-19 Vulnerability Index

Introduction

Through this study, a CVI in the form of a composite index was developed to determine the vulnerability of a community to the COVID-19 pandemic. Validated steps from literature [2,11] were used to develop the index, including themes and factors that contribute to the vulnerability of a community to a global pandemic [2,16], as was as appropriate validation methods for the index [2,21]. This chapter will highlight the development of vulnerability index from other established works and explain how those works are related to this research as well as describe the mathematical methods utilized in composing the CVI. The methodology comprising this research primarily stems its inspiration from the Surgo Foundation's CCVI research as well as the CDC's SVI as mentioned previously. Publicly available datasets were the source of the data in which this CVI was developed.

The first objective of this study aimed to develop a COVID-19 Vulnerability Index (CVI) to aid in preparedness for future global pandemics. This includes developing different themes that generate the index as well as selecting the correct methodology for index creation and development. As stated previously, there is a large amount of interest in this field of study. This means that there are numerous different methods for index creation, development, and validation. Though, when considering a composite index, many articles stem their inspiration by one of the most well-known indices, the SVI [2].

Methodology

As can be found through various assets of current literature and research, the shape that many COVID Vulnerability studies and indices seem to conform to is the methodology established by the CDC's SVI [14,15,16,19,20]. As mentioned previously, the SVI has created a fundamental foundation for establishing the basis of an index related to public health. This is established by configuring themes and factors that are generalized and related to the index that is proposed. The key indices that have been established by current literature that are most applicable to this study are the CDC's SVI and Surgo Foundations CCVI. Where the SVI focuses more on the vulnerability aspect in relation to natural disasters, the CCVI focuses more on vulnerability to COVID-19. The themes and factors comprising these two indices differ to represent the populations vulnerability to natural disasters and COVID-19, respectively.

An example of this is the current CDC's SVI which establishes four themes (Socioeconomic Status, Household Composition and Disability, Minority Status and Language, and Housing Type and Transportation) whereas Surgo's CCVI utilizes seven themes (Socioeconomic Status, Minority Status and Language, Housing Type, Transportation, Household Composition and Disability, Epidemiological Factors, Healthcare System Factors, High Risk Work Environments, and Population Density) [2,16]. One of the first steps in creating a CVI is identifying the areas of research that need to be addressed. This would involve the determination of factors and the creation of themes and the respective categories of each theme. These themes and their counterparts will all contribute to creating a CVI. For this research, eight themes were chosen, and relevant factors were chosen to compile them as seen in Table 1. Those themes are: Socioeconomic Status, Minority Status and Language, Housing Type and Transportation, Epidemiological

Factors and Disability, Health System Capacity, High Risk Work Environments, Population Density and Environmental Factors. These themes and their composition may differ from other CVI's such as Surgo Foundation's CCVI since this index was specifically composed of only New Jersey counties and their respective data. Depending on the amount of data that is relevant as well as publicly accessible, the number of themes is relatively subjective, though one can determine whether the areas of the study comprise all the relevant information through these themes. As stated previously, by conducting research on the area of study and finding relevant literature, the shape of a CVI in relation to the themes and factors that comprise it are manifested.

As can be seen below in Figure 1, the eight themes that have been selected to represent the CVI are as follows: Socioeconomic Status, Minority Status and Language, Housing Composition and Transportation, Epidemiological and Disability, Healthcare Capacity, High Risk Work Environments, Population Density and Environmental Factors. These themes are comprised of numerous different factors which all contribute to the generation of a CVI. These themes and their respective factors were selected and generated based on current literature in the development of a CVI as well as indicators that have been determined to be directly related to COVID-19 susceptibility.

Figure 1

A Representation of the Different Themes that Contribute to the CVI

Table 1 below represents the themes and their respective factors which all contribute to the generation of a CVI. Further elaboration on the themes and their respective factors will be described in Chapter 4. This table predominantly stems its inspiration from the methodology established by [11,2,16].

Table 1

Composition of Themes and their Respective Factors

Theme	Theme Title	Factors
Theme 1	Socioeconomic Status	Below Poverty
		Unemployed
		Income
		No Highschool Diploma
		Uninsured
Theme 2	Minority Status and Language	Minority
		Age 5+ speak English less than well
Theme 3	Housing Type, Transporation, Houshold Composition & Disability	Multi Unit Structures
		Crowding
		Group Quarters
		Mobile Homes
		Aged 17 or Younger
		Single Parent Household
		No Vehicle
		Disability
Theme 4	Epedemiological & Disability Factors	High colosterol
		Stroke
		Heart disease
		COPD
		Cigarettes
		Asthma
		Cancer per 100k
		HIV per 100k
		BMI >30
		Diabetes
		Population 65+
Theme 5	Healthcare Capacity	ICU beds per 100k
		Hospital beds per 100k
		PQI Prevention Quality Indicator per pop
		Cost of Medical care
		Parmacies and Drug stores per 100k
		Primary Care physician per 1000
Theme 6	High Risk Work Environments	Long term care residents per 100k
		Prison population per 100k
		Population employed in high risk industry per 100k
Theme 7	Population Density	People per square mile
Theme 8	Environmental Factors	Average daily PPM
		Number of superfund sites per county
		Vehicle volume

Once general themes which are relevant to the research are established, factors of these themes can then populate. Selection of these factors is determined by whether the data is publicly accessible, published/verified and representative of the index construction (i.e., relate factors to themes and overall vulnerability). These themes and their respective factors can be applied to any area of study. This can be done by modifying the weighted factors that can be found in Equation 3.6. By determining what factors or themes are more important in different geographical locations, this study has a much broader scope. It is in this area of development that there is more freedom for the researcher to specify types of data and weighting factors they would like to incorporate into the themes and index. Factors were also chosen in relation to the available research connecting said factors to known vulnerability characteristics of COVID-19 [1]. Depending on the goal of the research, there may be limited data at the disposal of the researcher. Themes 1-7 were all inspired by SVI and CCVI construction, respectively [2,16]. Though as mentioned theme structure was re-worked to fit New Jersey, as well as validation through published articles as previously mentioned. Theme 8 was generated due to the number of unique variables surrounding New Jersey. The reasoning and methodology behind factor selection for Themes 1-8 will be further discussed in Chapter 4.

As far as the technical methodology in the creation of an index is concerned, the PERCENTRANK function on excel is used in the creation of the SVI and the Surgo Foundation follows the same format. The RANK.EQ function was chosen to standardize the data for this research. This was due to the sample size of the data set and the applications of this research. The PERCENTRANK function normalizes data from 0 to 1 regardless of the number of equal values. The RANK.EQ function does not normalize this data and thus if two or more values are equal, the function represents them as equal. The proceeding value will then continue the trend from 0 to 1. Figure 2 below represents the order in which equations are applied and when. Considering the size of this data set it was determined that the RANK.EQ function would benefit the research more due to the lack of modification to the data that PERCENTRANK imposes. The equations used to formulate the data related to the CVI can be seen below [2,11]. Sample calculations representing the methodology found in Figure 2 can be seen in Appendix A. The concurrent themes generated by the current level of research related to a CVI vary. Depending on the scope of the research and the intended research area, methodology, and intended use, the themes and factors that compose an index are generally subject to change as seen fit by the researcher. The novelty of this research is to compose a CVI that takes other established research into account to build an unbiased framework and introduces unique variables related to the case study area. As mentioned above, there are a number of ways that an index can take shape in regard to the methodology used. For the purposes of this study, and the creation of an index the methodology established by OCED [11] will be followed in the creation of a composite index. From this resource, there are ten steps that need to be addressed for the proper creation of a composite index. They are as follows: Theoretical Framework, Data Selection, Imputation of Missing Data, Multivariate Analysis, Normalization, Weighting and Aggregation, Robustness, and Sensitivity, Back to the Real Data, Links to Other Variables, and finally, Presentation and Visualization [11].

For this study to capture the full scope of index development, there is a combination of methodologies that are gathered from a multitude of sources namely, the SVI, Surgo

Foundation, OECD, [11,16,2] as stated previously. The SVI established the framework behind generating an index based on public datasets. Surgo's CCVI established one of the first comprehensive CVIs based on the SVI. OECD has established the methodology to follow when creating a composite index. With that said, there will be no use of imputation of missing data because the data collected has been methodically considered and the imputation of an average value in place of missing data has been seen to skew results as per the CDC's SVI analysis with Native American reserves and municipalities [4]. After careful deliberation, another factor from OECD's methodology that is not being directly applied is weighting. Weighting is rather subjective depending on the methodology used and thus will not be directly applied to this research. There are various forms of weighting and the method that was chosen falls outside of the realm of traditional methods. It is to be stated that there is an indirect method of weighting that is applied to the impact of individual factors on the total value of the theme.

As mentioned above, weighting evenly applied due to the current level of research applying even weight to their themes and/or factors [2,16,38,39]. Otherwise, weighting involves the use of Subject Matter Experts (SME's) and/or an extremely thorough analysis which is outside of this scope of research. The weighting method that was applied to this research was the construction of the themes and the number of factors that influence the final CVI value for that theme. Even distribution was used, and this study validated this methodology by using New Jersey as a case study. By increasing the number of factors in a theme, each individual factor will contribute less to the overall value of the theme. Depending on the number of factors in each theme, the less impact each factor has on the overall theme value or CVI value. An example of this can be seen in Table 1, Theme 1

(Socioeconomic Status) which is composed of 5 separate factors compared to Theme 7 (Population Density) is composed of one factor. By having 5 factors in Theme 1, each factor can only contribute a maximum of 20% to the final score of a theme. In Theme 7, one factor directly relates to the total score or CVI value. This system was applied with consideration of the SVI and Surgo Foundation as well as other published articles involving indices [2,16,38,39]. Aggregation of factors was also done with respect to the aforementioned articles and available resources. Though aggregation tends to be at the discretion of the researcher and the intended areas of research.

Continuing with the process that OCED's composite index development document outlines, the remaining eight steps have all been applied to this research in their own unique and respective ways. The theoretical framework has taken inspiration from the SVI, Surgo Foundation, areas of published research [2,13-16,22,38,39]. This aided in determining what themes, categories, and variables should be included in the framework of the index. By using a multitude of relevant variables, it was possible to ultimately generate a composite SVI score for the counties of New Jersey. This involved selecting the appropriate basis for themes and factors into generating a meaningful composite index [11].

Data selection follows the same methods as described in the Theoretical Framework. By using the resources and understanding of current, available literature and research, the data that was selected was determined to be suitable for the index and the factors/ themes that the data represents. Data was carefully collected and considered when developing this index. Real world COVID-19 data was selected and compared with CVI values determined by this study utilizing a regression analysis. The real-world data being COVID infection and death rates amongst the counties of New Jersey. The time frame that was selected was pre-vaccine (March-December 2020) due to the high amount of variance that could occur in data from uneven vaccine distribution or willingness to participate in vaccination. Selection of data from this time frame allowed the research to be focused more on susceptibility to a pandemic before a vaccine is widely available.

The multivariate analysis was considered when conducting the formation of the index in the later stages. This also happens to tie into the validation of the index which will be described in more detail later in the results section of the chapter. The aggregation portion from this step was also inspired by the SVI and Surgo Foundation's CCVI. This step in the development of a composite index is to ensure that there is no biased data that is going to skew the results in the favor of whomever is creating the index. It is the proper addition of relevant data to categories and themes that when aggregated represents a coherent larger picture. This analysis was conducted by experimenting with different factors and even combining themes to analyze whether it would represent the data in a more comprehensive way.

Normalization/standardization of the data was conducted by applying relevant populations as mentioned above. This was the first step in being able to compare the data from different populations to one another. The second step of normalizing the data involved similar methodologies to that of the CVI established by the Surgo Foundation which was modeled after the SVI. As mentioned previously, this methodology involves the use of Microsoft Excel's RANK.EQ function which normalizes the data on a scale depending on the size of the data set. Once this was applied, the COUNT function was then applied to the values to provide a 0.047 to 1 scale. These represent 1/21 to 21/21 scale which are normalized the data for each county of New Jersey; New Jersey consists of 21 counties (i.e., Atlantic, Bergen, Burlington, Camden, Cape May, Cumberland, Essex Gloucester, Hudson, Hunterdon, Mercer, Middlesex, Monmouth, Morris, Ocean, Passaic, Salem, Somerset, Sussex, Union, Warren). Normalizing the data using these methods is necessary when creating a composite index and allowed for a much more accurate depiction and representation of the data collected.

In terms of uncertainty and sensitivity, there have been many measures taken to represent the data collected in an unbiased manner and as close as possible to the current literature as possible. This involved the collection of long-term data from reputable and trusted sources. Unfortunately, much of the data collected in the census and/or public health data is oftentimes political and can be manipulated before publishing [40]. This also relates to the short-term COVID-19 data that will be compared in the validation portion of the paper. Values for the most part have remained constant from the sources. These values will be taken at face value for the purpose of this study but may change in the future as mentioned previously. Otherwise, sensitivity and robustness were tested by combining Themes 6 & 8 together as well as including/excluding specific factors. Validation through linear regression was the closest method compared to established sensitivity methods as published by OCED [11]. This methodology also coincides with Back to the Real Data step as described by OCED [11]. Simple linear regression was applied to acknowledge the relationships between multiple variables and outputs. By changing variables that were considered as well as combination of themes, linear regression ultimately determined whether the changes were beneficial for the study or not. The results of these regressions has provided a statistical analysis that can prove linear relationships between Real-World

COVID-19 data and calculated CVI values; this will be discussed and further elaborated on in.

Links to other indicators have been made as the whole analysis and composition of this index was to determine one thing: how susceptible a county is to COVID-19? This was done by including factors that were included from other analyses as well as at the discretion of the researcher. With composite indices, the general idea behind the creation of an index and the different factors or variables that comprise the index are predominantly subjective as stated numerous times in OCED [3]. This means that determining the links and other variables that may contribute or detriment the index need to be thoroughly considered. This was done through analysis of current research conducted as well as drawing connections and establishing relationships between COVID vulnerability as per the CDC [1] and research validating the factors and themes chosen [1,2,8,9,27-36].

In terms of visualization of the index, there are maps and graphs that were generated through ArcGIS and Excel which can be seen in the Results and Discussion section. The maps represent every county on a color-coded basis on a scale from 0 to 1, similar to the SVI. This scale is broken down into 4 sections. Low (0 -.25), Moderate (0.251-0.50), High (0.51-0.75) and Very High (0.751-1.0). With this representation of the data, it is rather easy to determine which counties are at a higher risk than their counterparts. This was possible by conducting an additional normalization through RANK.EQ for visualization purposes. The graphs and bar charts that display the data that has not gone through the additional normalization and represent the actual calculated CVI values compared to their averages as well as the real-world COVID infection and death data.

To specifically derive the index, the methods that were used as mentioned previously were the function(s) RANK.EQ, COUNT and SUM. More specifically, the order that they were used to normalize data and generate an SVI will be described below. By using the RANK.EQ function as seen in previous studies developed Surgo's CCVI and CDC's SVI, it is possible to rank each county or area of study based on statistical values. There were many different methodologies that were considered and researched in determining the most ideal for this research as described previously. Ideally, the most accurate and least intensive method would be chosen. For these reasons, the methodology of index creation that was chosen was the use of the RANK.EQ function in conjunction with other excel functions which were used from the composite CVI.

The first step in the generation of the CVI was first to gather all of the relevant information for each county and arrange the values in a way that benefited the integrity of the index (unemployed population = population > 18 years old without a job). The second step was to apply the RANK.EQ function to these values and thus a scale was created from least to greatest. The third step was to divide by the count function to generate all of the values on a 1/21 to 1 scale as opposed to a 1 to 21. The fourth step is to sum the values from each theme and divide them by the total number of categories that comprise each theme. The fifth step is to sum all of the themes and divide them by the total number of themes [8]. This then results in a singular value on a 0 to 1 scale that determines the susceptibility of each county to COVID-19.

One of the most important steps in index creation was organizing data for the CVI by standardizing applicable data if possible and necessary. Standardization is a crucial process that brings data into a common format so that further processes and analyses can be conducted. This involves first, finding valid and relevant data, then dividing it by the affected population or another variable that will standardize it. An example of this being the percentage of the population that is unemployed. Depending on the source, most raw data will provide values that are not in percentages, ratios or factors which means that it is up to the interpreter of the data to standardize it. An example of this being data provided from studies conducted by [3,4]. More specifically, the standardization of unemployed persons per county. By dividing the number of unemployed persons by the total population that can be considered unemployed (the total population above the age of 16) it is possible to standardize the data and compare unemployment rates from one community to another.

This process can be seen in Equations 3.1- 3.4. Equation 3.1 represents the data that needs to be oriented in the same respect that the scaling of the CVI follows; least to most vulnerable (0 to 1). An example of this being Per Capita Income which is inversely related with the scaling and thus needs Equation 3.1 to be concurrent with scaling. Equation 3.2. is used to standardize data by applicable populations where 'X_i' is the data that needs to be standardized and 'X_j' is the applicable population. The resulting standardized data being 'Y_n' by dividing these two values, it is possible to determine the percentage of the population that is afflicted by a factor. Equation 3.2 represents the data that has been standardized by 3.1 but is not yet ready to be computed into Equation 3.3. Only values that need to be inverted for the RANK.EQ function to work are going to go through Equation 3.2. An example of this being number of Primary Care Physicians. Considering our scaling is 0-1 (0 being least vulnerable and 1 being most vulnerable) values such as number of Primary Care physicians need to be inverted to conform to this scaling. This methodology has been applied to all applicable data. Equation 3.3. is the other form of standardization and utilizes the RANK.EQ function as well as the COUNT function. As can be seen below in Equation 3.3, the RANK.EQ function standardizes the data on a scale from 1 to 21 where the COUNT function divides by the number of values in the dataset (for this study: 21).

$1/ X_i$ Eq. 3.1

$Y_n = X_i / X_j$ Eq. 3.2

Z_n = RANK.EQ (Y_n, \$$Y_i$: \$$Y_j$,1) / COUNT (\$$Y_i$: \$$Y_j$) Eq. 3.3

SUM (\$$Z_n$: \$$Z_i$) / COUNT (\$$Z_n$: \$$Z_i$) Eq. 3.4

Where:

X_i = Raw Data

X_j = Representative Population

Y_n = Standardized Population

$Y_i : Y_j$ = The range of standardized data

Z_n = Factor values

Z_j = The range of factors

RANK.EQ = Normalization Function

COUNT = Count Function

SUM = Summation Function

Equation 3.3 is applied only to the factors of each theme after the standardization of data by applicable population as mentioned above. Once Equation 3.3 is applied, counties are then ranked on a scale from 0.04 (1/21) to 1(21/21) due to the size of the data set. After the values are standardized by Equation(s) 3.1- 3.3, Equation 3.4 is then applied to the newly standardized factors to calculate the theme value for each of the 8 themes. Equation 4 is then used again, to sum the themes and determine the overall CVI value.

In terms of validation methodology, linear regression was chosen based on available research relevant to this study [20,22] the linear regression function was utilized through Microsoft Excel. This involved the selection of dependent and independent variables as well as the confidence interval. Since the regression being conducted is comparing two variables at a time, the dependent and independent variable selection(s) are arbitrary; though it is to be noted that the dependent variables selected were the real-world values and the independent variables were the CVI values. A confidence interval of .95 or 95% was selected. Excel automatically applied Equation 3.5 (as seen below) which is the standard linear regression analysis for the selected dependent and independent variables [21]. The independent variable being the real-world data and the dependent variable being the CVI data. A further analysis of the application of this equation and the results from the regression will be discussed in Chapter 4.

$$Y_i = f(X_{i,} \beta) + e_i \qquad \text{Eq. 3.5}$$

Where:

Y_i = Dependent Variable (Real-world COVID-19 data)

f = Function

X_i = Independent Variable (CVI values)

β = Unknown Parameters

e_i = Error Items

If a weight was to be directly applied to the individual factors of the CVI as opposed to an even distribution of weight per factor as conducted in this study; the equation that could be used for including additional weights can be seen below in Equation 3.6. This equation represents the CVI theme summation once the individual factors are calculated. The respective summation of factors and averaging of each theme can be seen as well as the summation of themes and averaging by the total number of themes to generate a CVI value can also be seen in Equation 3.6.

$$\mathrm{CVI} = \sum_{i=1}^{m}\left[\frac{\sum_{i=1}^{n}(x_i * w_i)}{\sum_{i=1}^{n}(n)}\right]/\mathrm{m} \qquad \text{Eq. 3.6}$$

Where:

X= Factor Value

W= Weight

N = Number of factors

m = Number of Themes

Conclusion

Again, the purpose of the first objective was to develop a COVID-19 Vulnerability Index (CVI) to aid in preparedness for future global pandemics. By analyzing current literature, it was possible to determine the scope and number of themes that will contribute to the CVI as well as the factors that will comprise each respective theme. By conducting research on the statistical development of a composite index, the mathematical methodology was developed and derived from numerous sources of current research and literature. By analyzing the validation methods of said research, it was determined that the best course of validation for this index is a linear regression analysis between the CVI values and real-world COVID data. This sets the stage for the case study of New Jersey and the further development of a COVID Vulnerability Index for the counties of New Jersey.

Chapter 4

Development of a CVI: A New Jersey County Case Study

Introduction

New Jersey counties were used as a case study to validate the CVI index developed through this study. When developing an index in this area of study, it is extremely important to collect as much relevant information as possible. This is a monumental task due to the extent of the data that needs to be collected to form a coherent and well-rounded index. Therefore, it was in the best interest of the research to look to the established literature yet again for inspiration of how to conduct such a task. By acknowledging the CDC's SVI and Surgo Foundation's Community COVID Vulnerability Index (CCVI), it was possible to gain a greater understanding of what information and data would need to be gathered before the development of said CVI. By utilizing U.S. census data, Themes 1, 2, and 3 were easily completed. As for the other themes, there needed to be extensive research in filling the gaps in data that were not so easily accessible. By utilizing New Jersey's DOH portal, Theme 4 was able to be expanded and completed. As for Theme 5, multiple sources of data and information needed to be cross-referenced and checked for transparency and accuracy in relation to the source(s). Theme 6 includes high-risk work environments which were located on the BLS (Bureau of Labor Statistics) [41] website as well as sourced through research articles. Theme 7 is simply the given population density of each county; this value can be either manually calculated or located on the Census bureau's resources page [4]. Theme 8 comprises the environmental factors that make New Jersey a relatively unique case study. This involves three factors such as Air Quality data, Superfund sites, and Vehicle traffic data. By utilizing publicly available datasets from numerous sources, the CVI covers a wide range of factors and potential hazard indicators that can contribute to a county's COVID-19 vulnerability. This chapter will also discuss the methodology and results behind composing and validating the index as established by Surgo Foundation's CCVI and other published works. Validation involves linear regression modeling and comparison of the generated CVI values to real world data.

The development of a CVI for New Jersey will help determine which community is at risk and which community will need help before a pandemic happens which is considered as a preventative measure. Technically, pandemic vulnerability should have been considered before the event of a global pandemic which is not all too far from fiction as seen. The development of a CVI can only benefit the people of New Jersey by anticipating which populations may need more economic and/or medical assistance than others. No two counties in the entire U.S. are created equally, and New Jersey is no exception. Therefore, it is paramount to consider the people that are being afflicted more so than others based on the themes that were provided. Furthermore, a CVI is not biased. CVI's specifically run-on facts and data that are gathered from reputable sources such as the US census and local health data. With an equal determination of which counties are to receive more support than others, it is only ethical to ensure that the CVI is an accurate approximation and/or prediction of the counties/ communities that are represented.

Some critiques about the generation of a CVI is that there seems to be no formal singular method of validation established. This is to be expected considering different methodologies for index development arises different methods for validation and verification of the index. Comparing real-world data to compiled observations seems to be one of the only methods to verify whether the index created is applicable to real-world scenarios. Though there are so many different variables that real-world data brings, and it is nearly impossible to account for all variables without any discrepancies or sources of error. Otherwise, real-world data comparison has its own problems with the accuracy of the data being collected and compared to. The only true data verification may have to be years from now after all data has been verified and can be truly compared. The Surgo Foundation has yet to provide a full publishing of their results and methodologies established from validation. Though their processes involve the use of SVI to CVI comparison(s) via linear regression (R-value) as well as future work validating CVI values through real-world data [16]. Though that method of verification may be somewhat sound, it is expected that the CVI will deviate from the Social Vulnerability Index (SVI) due to the nature of the index and the values that either of them are comprised of.

The second objective of this study was to incorporate New Jersey county data as a case study into the CVI in order to validate the index. This involves the selection and justification of the factors that comprise each theme. Table 1 below shows the Themes, their respective factors, and the sources of data. This expands on the table that was developed in Chapter 3 (Table 1) and further elaboration of each theme and its respective factors will be highlighted in this portion of the research as stated above. The themes in the table represent the different generalized areas of research that were selected for the index. The categories column helps delineate some of the more specific areas of research that were covered into an umbrella category; this was strictly for convenience purposes and to make the table easier to understand. The factors column represents the specific values that were taken from data sources and input into an Excel document which was used to calculate the respective theme and CVI values for each county. The source column is linked to the sources by which each factor was found (1-12). The purpose of the table was to help display the factors of each theme, what categories were considered during development and the sources of the data that was collected.

Table 2

Themes, Categories, Factors, and Sources that Compose the CVI

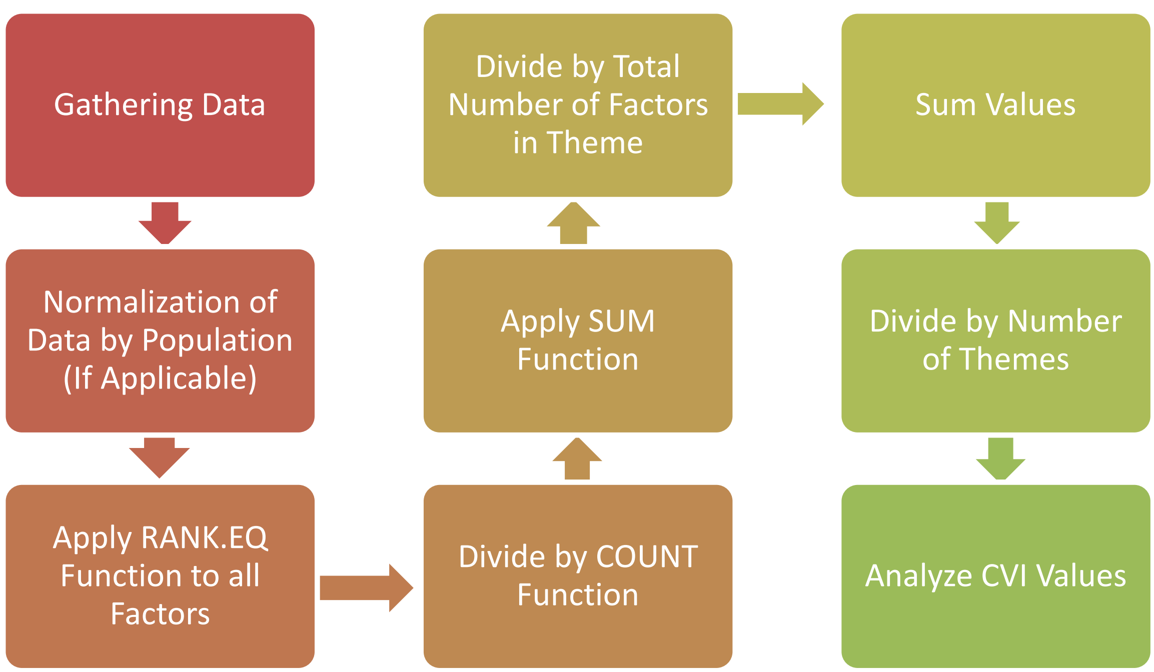
Theme	Theme Title	Categories	Factors	Source
			Below Poverty	
			Unemployed	
	Socioeconomic Status	Socioeconomic Status	Income	
			No Highschool Diploma	
Theme 1			Uninsured	
	Minority Status and Language	Minority Status and Language	Minority	
Theme 2			Age 5+ speak English less than well	
			Multi-Unit Structures	2
		Housing Type	Crowding	
			Group Quarters	
	Housing Type, Transportation, Household Composition & Disability		Mobile Homes	
		Household Composition	Aged 17 or Younger	
			Single Parent Household	
		Transportation	No Vehicle	
Theme 3		Disability	Disability	
			High cholesterol	
		Cardiovascular conditions	Stroke	
			Heart Disease	
			COPD	5
		Respiratory Conditions	Cigarettes	
	Epidemiological Factors		Asthma	
		Immune Compromised	Cancer per 100k	
			HIV per 100k	23
		Obesity	BMI >30	5
		Diabetes	Diabetes	
Theme 4		Elderly	65+	2
		Health system Capacity	ICU beds per 100k	6
			Hospital beds per 100k	7
	Healthcare System Factors	Health System Strength	PQI Prevention Quality Indicator	24
			Cost of Medical care	
		Healthcare Accessibility	Pharmacies and Drug stores per 100k	23
Theme 5			Primary Care Physician per 1000	
			Long term care Residents per 100k	25
	High Risk work environments	Percent of pop working/ living in high infection risk environment	Prison population per 100k	26
Theme 6			Population employed in high-risk Industry per 100k	16
Theme 7	Pop density	Pop density	People per Square Mile	2
		Average Daily PPM '16	Average Daily PPM '17	27
Theme 8	Environmental Factors	Number of superfund sites per county	Number of Superfund Sites Per County	28
		Vehicle Volume	Vehicle Volume	27

An example of this is the standardization of population data for each of the counties. There are 21 counties and by utilizing the percent rank (RANK.EQ) function, it is possible to create a scale in increments of 1 to 21. Generally, 21 is the highest risk and 1 is the lowest risk. After generating the standardized values, they are then divided by the COUNT function. This averages the data on percentile in increments of 1/21. This purpose is to help standardize the values and generate a scale from 0 to 1 similar to what the SVI had done. This process must be conducted for each and every factor inside of each theme. After each factor is standardized using the process mentioned above, they are then summed and divided again by the COUNT function which yields a theme value. After all of the theme values are calculated, the CVI value can then be created using the same process. By summing each theme and dividing by the COUNT function, an SVI value on the scale from 0 to 1 emerges. Considering there are 21 counties, there are 21 individual and representative SVI values for each county of New Jersey. As seen above in Table 2, the eight themes and their respective categories and factors are displayed. The factors that went into the creation of Table 2 and its respective themes are represented by the current state of knowledge regarding the current research being conducted as well as the current state of literature being published.

All themes (Themes 1-8) were considered using the methodology that has been described in Chapter 3. This process involved the utilization of many functions in Microsoft Excel in a specific order to achieve a composite index value for each county. This order can be seen in Figure 2 below. As mentioned previously, the statistical creation of a composite index was generated from the CDC's SVI. The use of the percent rank (RANK.EQ) function was mentioned in their derivation of the index. This function is also used in several other articles that are generating an index and need to relate values to one another on a coherent scale. Generally, the data gathered for each of the themes and their respective categories and factors are based on the percentage of the population which are affected and/or falls under the respective categories listed unless otherwise stated in the information listed below. Otherwise, the only difference in applying this process is the number of factors that comprises each theme. This is inherently a weighing system that is predominantly subjective but has been inspired by current literature [2,16,38,39] to uphold consistency and reduce bias.

Figure 2

Flowchart Representing the Different Steps in CVI Creation



Theme 1: Socioeconomic Status

As mentioned previously, Theme 1 is composed of various categories which populate the main theme of Socioeconomic Status. These categories and their respective factors are as follows: Percentage of the population that falls below poverty, Percentage of the population that is unemployed, Average household income, Percentage of the population without a high school diploma, and Percentage of the population that is uninsured. These factors were predominantly derived from the CDC's SVI [2]. This methodology is also followed by Surgo Foundation [16] in their analysis and derivation of a CVI. Socioeconomic status is relevant when making a CVI since lower-income areas generally have less access to healthcare and/or assets to subsist off of in the event of a pandemic. As can be seen in many facets of current literature, there is a tremendous impact on COVID-19 vulnerability from socioeconomic factors and social vulnerability. [29,30,31].

Theme 2: Minority Status & Language

The second theme of Minority Status and Language comprises the following factors: The percentage of the population that is considered a Minority and the percentage of the population that is above the age of 5 and speaks English less than well.

Minority status and language are relevant in the formation of a CVI due to the fact that public health messages and different regulations may not translate very effectively for the percentage of the population that cannot speak or understand English very well. Otherwise, Minority status (all persons except white, non-Hispanics) is also considered when forming the CVI due to the lack of facilities that minorities tend to have access to. As can be seen in [29,30,31] populations composed of minorities have been disproportionately affected by the pandemic. Therefore, Minority status and Language are important categories to consider in the creation of a CVI.

Theme 3: Housing Composition/Type, Transportation and Disability

The third theme is composed of numerous categories which relate to housing type and composition as well as population statistics. More specifically, Crowding, Multi unit structure, and Group Quarters data from the SVI has been compiled and used which is related to population density and directly relates to contact exposure. Other values that were considered for theme 3 were the percentage(s) of the population that lived in mobile homes, were aged 17 or younger, and/or did not have access to a personal vehicle. These factors were important in determining portions of the population living in less-than-ideal circumstances as well as finances. Otherwise, this theme indicates what percentage of the community will be using some form of public transportation. These factors are all extremely important when considering the creation of a CVI. This is due to the fact that though population density may be a generalized average for a large area, these factors are much more specific indirectly determining what percentage of the population is living in less-than-ideal conditions in relation to the pandemic. All of which contribute to the possibility of exposure to the virus and potential health risks associated with said living conditions/ choices [24,32,33].

Theme 4: Epidemiological Factors

This theme is titled Epidemiological factors and consists of statistical data gathered from the New Jersey Department of Health portal [5]. The factors that comprise Theme 4 are related to pre-existing health conditions, age, and lifestyle choices. The pre-existing categories that compose this theme are conditions such as cardiovascular health, respiratory health, immunocompromised factors, and age. For Cardiovascular health, the following values were considered. Heart disease, Chronic Obstructive Pulmonary Disease (COPD), diabetes, high cholesterol, stroke history, and obesity. These health factors increase the chance of mortality if exposed to COVID-19 as well as the likelihood of frequenting a hospital setting for treatment related to these conditions.

Respiratory health is an issue of large concern when it comes to COVID-19. This is due to the fact that the virus heavily affects the respiratory system and can cause pneumonia among other respiratory illnesses. For those with pre-existing respiratory conditions, this effect is catalyzed and can lead to high mortality rates. The factors that contributed to these factor areas mentioned previously are the percentage of the population that smokes cigarettes and those who have COPD.

The immunocompromised factors consist of two factors. Per capita that is afflicted with HIV and percentage of the population that is afflicted with cancer. These were both measured per 100,000 people. By having immune system deficiencies, the virus will most likely have a stronger effect and lead to higher mortality rates in those with said conditions. Another factor to consider is that New Jersey has the highest number of superfund sites (151) most of which are active and contaminated with known human carcinogens or toxins and therefore this value is rather important with respect to the New Jersey case study. As can be seen from various assets of current literature [20], there is a large role that pre-existing health factors play in the mortality rate of COVID-19. As mentioned, the more severe the pre-existing conditions, the higher the chance of mortality and/or severe health complications if one contracts the COVID-19 virus.

Theme 5: Health Care System Factors

This theme is titled Healthcare System Factors. This comprises numerous categories that represent the Healthcare systems related to each county of New Jersey. The categories that comprise this theme are as follows: Capacity, Strength, and Accessibility of Healthcare systems. Beyond these categories, there are a number of factors such as ICU beds per 100k, Hospital beds per 100k, Cost of Medical care, and the number of Primary physicians per 1000. The 100k and 1000 for these factors being a measure of population.

As can be seen in [23] the number of Emergency Department visits and total hospital visits surged from March 2020 to January 2022. This made it critical that those needing emergency care for treatment of COVID-19, or life-threatening illnesses have access to such facilities. By incorporating these categories in the CVI, it provides a picture of the various factors that come into play when assessing the effectiveness of the healthcare system(s) in each respective county. In any pandemic or disaster event, health system preparedness and strength are always an extremely important factor for obvious reasons. In relation to COVID-19, certain counties may not have as much capacity and/or resilience to serve their respective communities. Unfortunately, this was the case for many patients afflicted with the virus. By understanding the healthcare infrastructure that is established in each county it is possible to determine which counties may need state and/or federal assistance on the healthcare front. It is also important to note that counties that exceed capacity may have residents seeking medical attention in neighboring counties and even states.

Theme 6: High Risk Work Environments

This theme is titled High-Risk Work Environments. This comprises three categories that represent the percentage of the population living and/or working in high-risk environments with respect to COVID-19 [16]. This theme also comprises two factors which are prison population per 100,000 residents and long-term care residents per 100,000 residents of each county. These work and living environments are extremely high risk which means that there is a high chance of COVID-19 transmission in these areas. By determining the percentage of the population living and/or working in these environments, there is a better understanding of what percentage of the population risks COVID-19 exposure on a daily basis simply from their work and/or living environment. Examples of this high-risk industry include dental hygienists, healthcare workers, social workers etc. By having a high percentage of the population working in these professions, it is much more likely that a virus can spread and affect multiple families.

Theme 7: Population Density

This theme is one of the most important which is why it is in itself a stand-alone value from county to county. Population Density is one of the most important indicators of COVID-19 transmission as well as an indicator of numerous other factors [17]. This is in part from the nature of the virus and that it is predominantly spread in three ways which relate to breathing and bodily fluids. These routes of exposure are droplets which can be spread through various means such as breathing, touching the face from contaminated hands, and having droplets land on the face from a sneeze and/or cough [34] . Theme 3 is related to population density due to factors such as Multi-Unit Housing, Crowding, and Group

Quarters. Though these are in themselves independent, population density can provide a wider average population per-area basis. As mentioned earlier, population density is a more generalized value for how densely packed a county may be whereas Theme 3 is a little more specific.

Theme 8: Environmental Factors

This theme is titled Environmental factors. This seemed to be an appropriate addition to the index due to the nature of the study and the respective case study. The factors that comprise this theme are air quality data that included the average daily PM 2.5, location of superfund sites and vehicle traffic data. Average Daily PM 2.5 relates to the amount of particulate matter in the air which is a direct indicator of air quality. This is measured in µg/m3 and can have adverse health effects on entire communities if overexposed and thus higher COVID-19 infection/mortality rates [35,36]. Superfund sites were related to the amount of superfund sites per country and then divided by the total area of said county. This then determines the number of superfund sites per square mile. Considering New Jersey has the highest number of Superfund Sites in the United States, this was more than worth considering in the development of an index. Superfund sites are still being discovered and remediated to this day as more and more manufacturing processes are discovered. Vehicle traffic relates to the amount of traffic that passes through a county's major roads. It is well known that vehicle emissions include BTEX compounds (Benzene, Toluene, Ethylbenzene and Xylene) which are known carcinogens and toxins. Vehicle travel also relates to the amount of traffic a county might experience on a regular basis. Considering the avenues by which COVID-19 spreads, high vehicle traffic could relate to higher transmission rates.

Validation Methodology

After establishing the themes and relevant factors based on current levels of research as can be seen above, the next step is to calculate the CVI as well as analyze it in reference to data from the area of study. As stated in Chapter 3, the validation of this index is done through a linear regression analysis between the CVI values and real world COVID-19 data collected from the counties of New Jersey [26]. The real-world data being the number of COVID-19 infections and deaths both normalized by the total population of each county. These values were then compared to the CVI through a linear regression analysis as previously stated.

The methods behind the validation of the CVI were focused on validation through New Jersey case study data. As mentioned, this involved infection and death data collected from each county [38]. The time frame that was used for this comparison was in the beginning stages of the pandemic more specifically, March 2020- December 2020. This time frame was used due to the vaccine being widely distributed after December of 2020. Vaccine distribution entails a large list of complex variables such as willingness to be vaccinated, efficacy of vaccine brands, access to distribution points etc. These variables would need to be accounted for if there was to be a fair comparison of COVID-19 data past December of 2020.

Unfortunately, there is a lacking amount of methodology related to composite CVI development. There are other resources that highlight more complex and intricate ways of validation through programs like Python (AUC-ROC Modeling) [13] or other methods outside of this scope of research. Though there was a similar study conducted that applied a linear regression analysis and utilized data from COVID-19 death and infection rates in different counties [37]. Surgo Foundation's CVI was working on establishing validation methods for their CCVI; this involved comparison Pearson's correlation coefficient (R) between the CCVI and SVI. Surgo was also working on a similar methodology to that of [22] by conducting a linear regression analysis of their CVI with real world data. For these reasons, the most logical solution for validation is the linear regression analysis of the CVI through different sources of data. As mentioned, death and infection data that has been normalized by population was selected for analysis as well as a regression analysis of the SVI.

The generalized Linear Regression function (also known as least squares) can be seen in Chapter 3, Equation 3.5. As mentioned in said Chapter 3, the Excel function automatically conducts a linear regression analysis resulting in various outputs. Though it is important to understand what equations are used during this automation. Therefore, to fully appreciate and understand the mathematics behind the linear regression analysis, it is important to analyze each of the equations that is being used for validation. This includes R^2 (Coefficient of determination), R (Pearson's correlation coefficient) and Significance F (p-value). The Coefficient of determination (R^2) measures the amount of variance between the independent and dependent variables. Pearson's correlation coefficient measures the strength of the linear relationship between both the independent and dependent variables. The Significance F (p-value) needs to have a null and alternative hypothesis to function. The null hypothesis being that if the p-value exceeded .05 then there is no linear relationship between the dependent and independent variables. The alternative hypothesis being that if the p-value was under 0.05 then there is a linear relationship between the independent and dependent variable(s). As mentioned in Chapter 3, the dependent variable is the CVI and independent is the real-world data; infections/ population, deaths/population and SVI values. For the purpose of this study, independent linear regression testing is being conducted (single independent variable analysis) which means that the dependent and independent variables can be arbitrarily changed. Though it was noted that the CVI values should be dependent of the real-world data.

The coefficient of determination can be found by calculating the Pearson's correlation coefficient and then squaring that value. The Significance F/ p-value can then be calculated from finding both previous values (R and R^2). The process that is involved in calculating these values can be seen in Equations 4.1-4.3 and are as follows:

$$R = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \, \Sigma(y_i - \bar{y})^2}} \qquad \text{Eq. 4.1}$$

$$R^2 = R \times R \qquad \text{Eq. 4.2}$$

$$F = \frac{\frac{R^2}{k}}{\frac{1-R^2}{N-k-1}} \qquad \text{Eq. 4.3}$$

Where:

R = Pearson Correlation coefficient

R^2 = Coefficient of Determination

F = Significance F value

x_i = Independent variable values

$\bar{x}$ = Mean of the independent variable

y_i = Dependent variable values

$\bar{y}$ = Mean of the dependent variable

K = Number of independent variables

N = Sample size

Results

After establishing what index should be created, focus is then turned to how to properly construct a composite index and the methods that should be followed in the generation of this specific index. The result being that it is possible to produce a CVI and validate the procedure followed through a linear regression analysis. This includes the COVID-19 Vulnerability Index (CVI) for the counties of New Jersey as represented in Table 2. Maps have also been generated to visualize the CVI values that have been calculated as can be seen in Figures (3-6). Graphs representing the real-world data, CVI and SVI values can be seen in Figures (7-10).

Figures 3-6 represent the (3) CVI, (4) SVI, (5) Infections/Population and (6) Deaths/Population when normalized by the RANK.EQ function as stated previously. By conducting this process, comparisons can be made from one data set to another very easily. With each of them ranked with a color-coded scale spanning from: Low (0.0-0.25), Moderate (0.251-0.50), High (0.51-0.75), and Very High (0.751-1). Again, these maps are strictly meant for visualization and comparison purposes and do not represent the actual calculated data as represented in Table 5 in the column labeled 'CVI'. With that being said, there are many similarities between the different maps. As can be seen with the North-Eastern counties, there are 4 counties that are all considered Very High on all the maps.

This includes Essex, Passaic, Hudson, and Union. It is rather interesting that although the data is derived from completely different areas of research that there are trends emerging. This aids in validating that the themes and factors that were instilled in the CVI are matching real world comparisons and the index of the established SVI.

Figures 7-10 represent the data that has not been put through the RANK.EQ function for normalization as a final step. In other words, the values shown are not scaled from 0 to 1. These are the raw calculated values of the CVI, SVI, COVID-19 Infections/Population and COVID-19 Deaths/Population (per county respectively). Figure 7 shows the CVI values that have been taken from Table 5 and the average CVI value for the 21 counties of New Jersey. This bar chart helps delineate what counties may have had higher or lower values than depicted on the maps in Figures 3-6. By analyzing data in this way, it is beneficial in displaying the unmolested CVI values in a visually appealing graph. As can be seen, Essex is the highest followed by Hudson and 8 other counties that are above the average CVI value line. Figure 8 represents the CVI in comparison to SVI values. The reason for this graph is due to Surgo Foundation establishing a relationship and analysis between their CCVI and the SVI values. As can be seen in Figure 8, the SVI values tend to fluctuate depending on the county. A further analysis of the data will be discussed with the validation graphics (Figures 11-13). Figure 9 and Figure 10 represent the number of COVID-19 infections divided by the total population and number of COVID-19 deaths divided by total population per county, respectively. These graphics are important because they will help determine which counties should experience higher CVI values as well as form a validation argument for the generation of the CVI.

Figures 11-13 represent the linear regression analysis results for the CVI. This was described previously and therefore the results are as seen above. The calculated CVI values are shown as the blue points where the linear trend line is seen as the 'Predicted CVI' demarked as the orange points. Though examination of values in Table 5 (in the column labeled CVI vs. SVI) is necessary in order to ensure there is a linear relationship between the two variables. In Table 6, there are several parameters that were measured when conducting the linear regression analysis. The most important being the R value (Pearson's correlation coefficient), R^2 value (Coefficient of Determination) and Significance F (p-value). The R value generally relates to the strength between two variables in a linear trend. The higher the R value, the stronger the linear trend and subsequently the higher the R^2 value. The R^2 value is a measure of variance that can be seen through analysis of independent and dependent variables. This effectively represents how much the variables relate to one another numerically. The higher the R^2 the less variance there is between independent and dependent variable data. The Significance F value represents the p-value for the overall linear regression graph. This value should not exceed 0.05 if there is to be a case made for the alternative hypothesis (linear relationships) to be stated as true.

In the example of Figure 11, the CVI was plotted with the SVI through linear regression. The R^2 value was determined to be 0.5084. This means that there was a slight amount of variance between the independent and dependent variables when comparing CVI and SVI data and establishing a linear relationship. Though as mentioned above, Table 6 displays the R and Significance F value(s) to fully determine whether a linear relationship can be related to the two variables. The R and Significance F values need to be analyzed as mentioned previously. The R value in this case being labeled as 'Multiple R' and having a value of 0.7130. This indicates a strong linear relationship between the two variables. The Significance F value being 0.0002 which is well below 0.05 and thus the alternative hypothesis is accepted; the two variables share a strong linear relationship. With such a strong linear relationship and considering previous established methodologies by that of [2,16] it was determined that the best way to represent the data is unbiased by weighting factors. With the proper arrangement of factors and themes, the data can speak for itself in the New Jersey case study. As stated previously, the weighting factors that were applied to the calculation of the CVI for this study were equal respective to each theme. This can be seen in Equation 3.6 by applying 1 where the weight variables are considered.

Figure 12 and Figure 13 represent the real-world data that can be connected to the CVI values that have been calculated. More specifically, Figure 12 has the CVI value set as dependent and COVID-19 Infections/ Population as independent where Figure 12 has the CVI values set as dependent and COVID-19 Deaths/Population as independent. By performing the same methodology as previously described, there is a clear linear trend in the data. The R values being 0.7950 and 0.8138, respectively. This coincides with a very strong linear relationship between the dependent and independent variables. The R^2 values being 0.6320 and 0.6623, respectively. This coincides with a smaller amount of variance between the two data sets and the trend line. The Significance F values falling well below 0.05 are 1.66E-05 and 7.2E-06, respectively. This indicates that the relationships established in each of the datasets share an extremely strong linear relationship. This means that the CVI and the real-world COVID-19 data that has been collected are comparable to one another and thus the indication of a high CVI value relates linearly to a high chance of COVID-19 infection or death.

Figure 14 and Figure 15 represent the SVI plotted against the infection/ population and death/ population data respectively. By conducting a linear regression analysis with the same data that the CVI was plotted against, it is obvious that the CVI provides a greater linear relationship than that of the SVI. The values that are relevant from the SVI and real-world analysis can be seen in Table 7. The R value, R^2, and Significance F values are the most relevant in this analysis as mentioned previously. The R value for the SVI graphs are 0.6397 and 0.6111 respectively. These values are substantially lower than the CVI analysis that was conducted above. This results in the R^2 values also being much lower than the CVI analysis. The R^2 values being 0.4091 and 0.3734 respectively. This relates to a higher amount of variance from the linear trend line. Most importantly, the Significance F values for the SVI vs Infections/ Population and SVI vs Deaths/Population are 0.0018 and 0.0033 respectively. These values are substantially closer to the maximum significance F value allowed to relate a linear trend to two variables of 0.05. By analyzing the SVI values with a linear regression analysis utilizing real-world COVID-19 data, it was determined that the CVI shares a much stronger linear relationship with COVID-19 data than the SVI. Thus the CVI is more accurate in determining a population's vulnerability to COVID-19.

Table 3 below presents the values of each theme incorporated into the CVI for all 21 counties of New Jersey.

Table 3

Calculated Theme Values for Themes 1-8

County	Theme 1	Theme 2	Theme3	Theme 4	Theme 5	Theme 6	Theme 7	Theme 8
Atlantic	0.8095	0.6429	0.7347	0.6151	0.6190	0.5397	0.3333	0.3810
Bergen	0.3429	0.6429	0.4150	0.3532	0.4830	0.8571	0.8571	0.8571
Burlington	0.3048	0.3810	0.4490	0.6746	0.5102	0.6190	0.3810	0.5873
Camden	0.6857	0.5952	0.6395	0.6508	0.6463	0.7460	0.7143	0.8413
Cape May	0.5429	0.1667	0.5646	0.7857	0.4898	0.2698	0.2857	0.2063
Cumberland	0.8095	0.7857	0.7687	0.7262	0.6939	0.3333	0.2381	0.2857
Essex	0.8571	0.9048	0.7755	0.5159	0.6190	0.9206	0.9524	0.8413
Gloucester	0.4286	0.2143	0.4218	0.5317	0.5782	0.4127	0.4286	0.4603
Hudson	0.7619	1.0000	0.6122	0.4722	0.5850	0.6032	1.0000	0.9206
Hunterdon	0.0952	0.0952	0.2585	0.3532	0.2925	0.0794	0.1429	0.3333
Mercer	0.2571	0.6905	0.2449	0.3690	0.4966	0.8095	0.7619	0.3968
Middlesex	0.6857	0.7857	0.7347	0.4722	0.5782	0.7619	0.6190	0.6667
Monmouth	0.4476	0.3810	0.5238	0.3373	0.4762	0.5079	0.5238	0.4286
Morris	0.0857	0.4762	0.3197	0.4484	0.4422	0.5238	0.4762	0.3651
Ocean	0.7524	0.2143	0.6939	0.5992	0.5510	0.6508	0.6667	0.5556
Passaic	0.8286	0.9048	0.7007	0.4921	0.5578	0.5079	0.8095	0.4444
Salem	0.6952	0.3571	0.5646	0.7976	0.6735	0.2381	0.0476	0.3333
Somerset	0.2095	0.5000	0.3265	0.4286	0.3537	0.5238	0.5714	0.6190
Sussex	0.2571	0.0476	0.2313	0.4762	0.4422	0.1270	0.0952	0.2222
Union	0.6476	0.9048	0.5986	0.3968	0.4218	0.5397	0.9048	0.8095
Warren	0.4952	0.3095	0.3946	0.3849	0.3946	0.1111	0.1905	0.3333

Table 4 below represents the number of flagged factors that indicate which themes and factors contribute more to each county's CVI value. Calculated factors were determined using Equations 3.1-3.3 The objective of this figure is to represent the number of factors that are key contributors to the final theme and CVI value(s).

Table 4

Flagged Factors (Calculated Factor Value>0.7)

FLAGGED FACTORS > 0.7								
Counties	Theme 1: Socioeconomic Status	Theme 2: Minority Status and Language	Theme 3: Housing Type and Transportation	Theme 4: Epidemiological Factors and Disability	Theme 5: Health system Capacity	Theme 6: High-Risk Work Environments	Theme 7 : Population Density	Theme 8: Environmental Factors
Atlantic	3/5	1/2	4/7	5/12	1/7	2/3	0/1	0/3
Bergen	0/5	1/2	1/7	1/12	3/7	2/3	1/1	3/3
Burlington	3/5	0/2	1/7	5/12	1/7	1/3	0/1	1/3
Camden	2/5	0/2	3/7	6/12	3/7	2/3	1/1	3/3
Cape May	1/5	0/2	3/7	7/12	3/7	1/3	0/1	0/3
Cumberland	4/5	2/2	5/7	9/12	3/7	0/3	0/1	0/3
Essex	5/5	2/2	6/7	3/12	1/7	2/3	1/1	2/3
Gloucester	0/5	0/2	1/7	4/12	5/7	0/3	0/1	1/3
Hudson	3/5	2/2	4/7	4/12	1/7	1/3	1/1	3/3
Hunterdon	0/5	0/2	1/7	2/12	4/7	0/3	0/1	0/3
Mercer	0/5	0/2	0/7	2/12	1/7	2/3	1/1	1/3
Middlesex	3/5	2/2	3/7	2/12	2/7	1/3	0/1	1/3
Monmouth	0/5	0/2	1/7	3/12	1/7	1/3	0/1	0/3
Morris	0/5	0/2	0/7	3/12	2/7	0/3	0/1	0/3
Ocean	4/5	0/2	4/7	5/12	1/7	2/3	0/1	0/3
Passaic	4/5	2/2	4/7	5/12	2/7	1/3	1/1	1/3
Salem	3/5	0/2	2/7	9/12	3/7	1/3	0/1	1/3
Somerset	1/5	0/2	1/7	2/12	6/7	2/3	0/1	0/3
Sussex	0/5	0/2	1/7	2/12	3/7	0/3	0/1	0/3
Union	2/5	2/2	4/7	2/12	2/7	0/3	1/1	2/3
Warren	0/5	0/2	0/7	1/12	2/7	0/3	0/1	0/3

Table 5 below represents the CVI values that have been calculated for each county as well as a column representing the RANK.EQ values that were used to easily map the CVI results. As seen, it ranks counties from highest to lowest CVI. Essex has the highest CVI and Hunterdon the lowest.

Table 5

Final CVI Values and RANK.EQ CVI Values

County	CVI	RANK.EQ CVI
Essex	0.7983	1.0000
Hudson	0.7444	0.9524
Camden	0.6899	0.9048
Middlesex	0.6630	0.8571
Passaic	0.6557	0.8095
Union	0.6529	0.7619
Bergen	0.6010	0.7143
Ocean	0.5855	0.6667
Atlantic	0.5844	0.6190
Cumberland	0.5801	0.5714
Mercer	0.5033	0.5238
Burlington	0.4884	0.4762
Salem	0.4634	0.4286
Monmouth	0.4533	0.3810
Somerset	0.4416	0.3333
Gloucester	0.4345	0.2857
Cape May	0.4139	0.2381
Morris	0.3922	0.1905
Warren	0.3267	0.1429
Sussex	0.2374	0.0952
Hunterdon	0.2063	0.0476

Figures 3- 6 represent the CVI mapped onto the counties in the state of New Jersey. The ranking is ranged from Low to Very High in respect to the representative colors. The data is represented by RANK.EQ (standardized) and the following are represented: (3) CVI, (4) SVI, (5) Infections/Population and (6) Deaths/Population.

Figure 3

RANK.EQ CVI Values

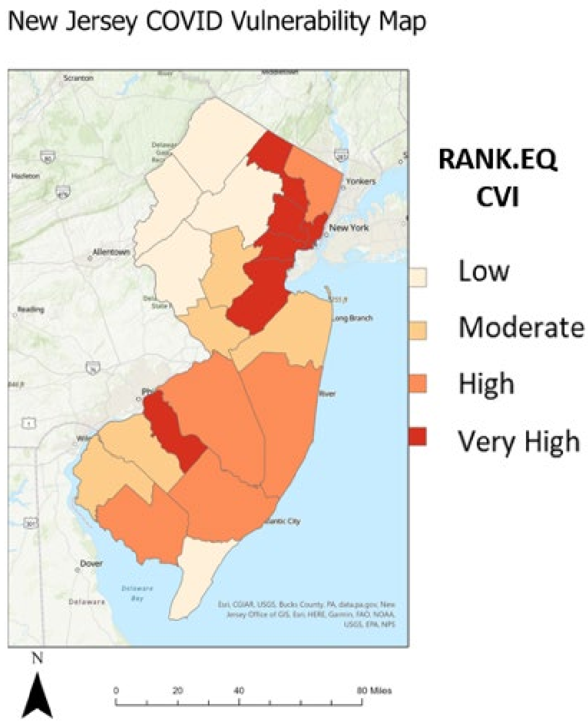


Figure 4

RANK.EQ SVI Values

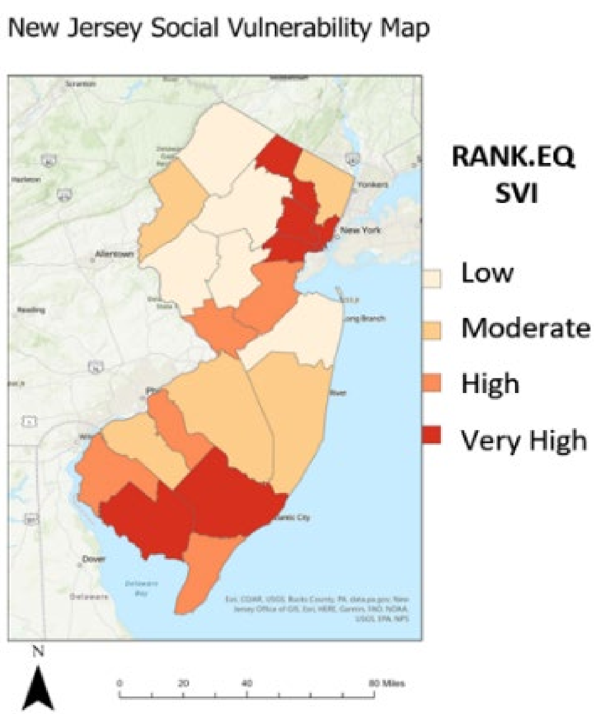


Figure 5

RANK.EQ Deaths/Population Values

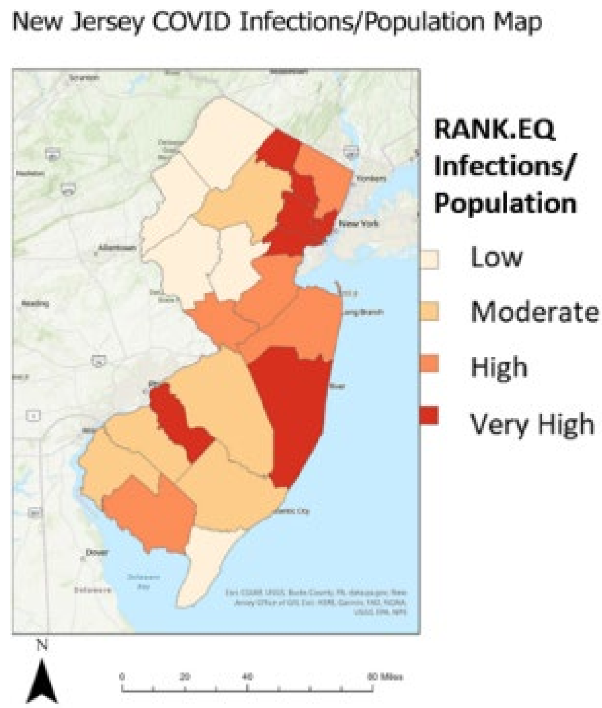


Figure 6

RANK.EQ Infections/Population Values

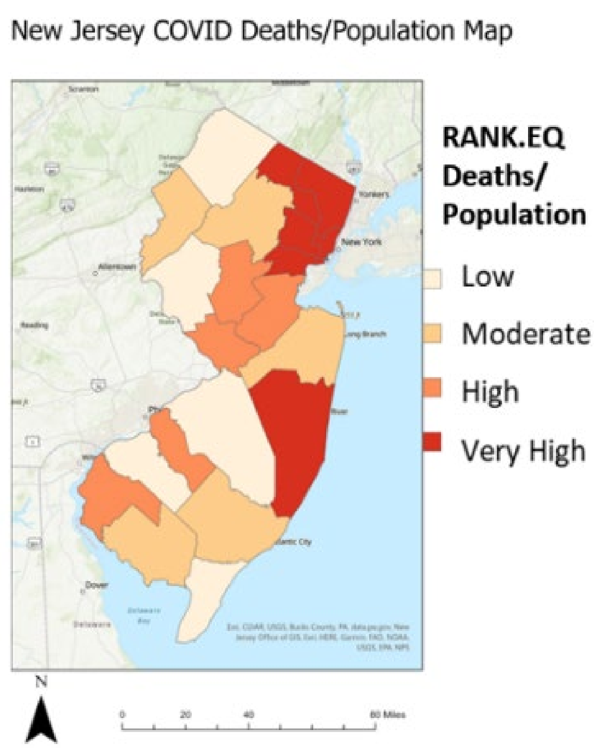


Figure 7 represents the calculated CVI values (not standardized by the RANK.EQ function at the end).

Figure 7

Bar Graph of CVI Values and Average

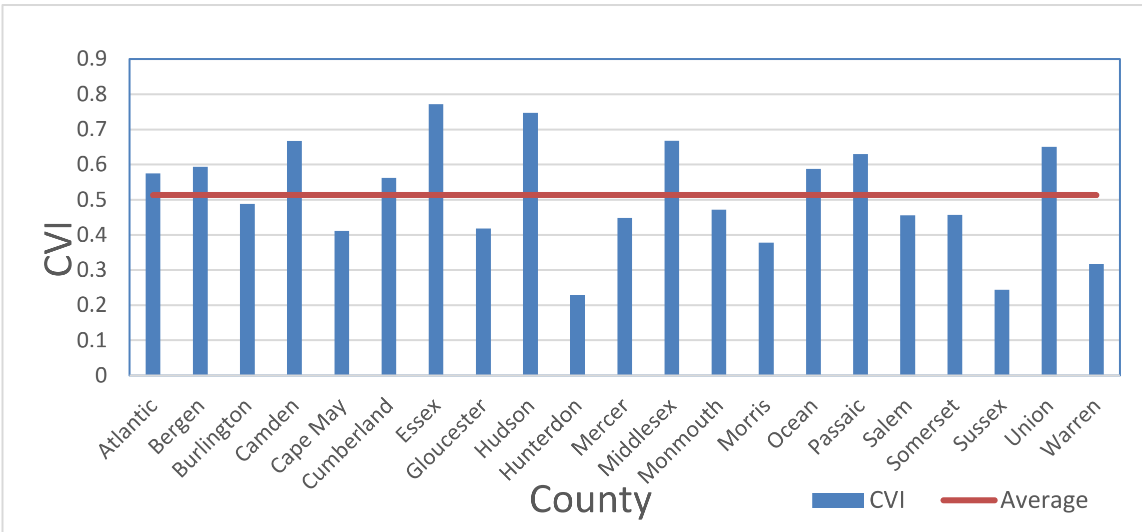


Figure 8 represents the CVI values and SVI values being compared to one another.

Figure 8

Bar Graph Comparing CVI and SVI Values

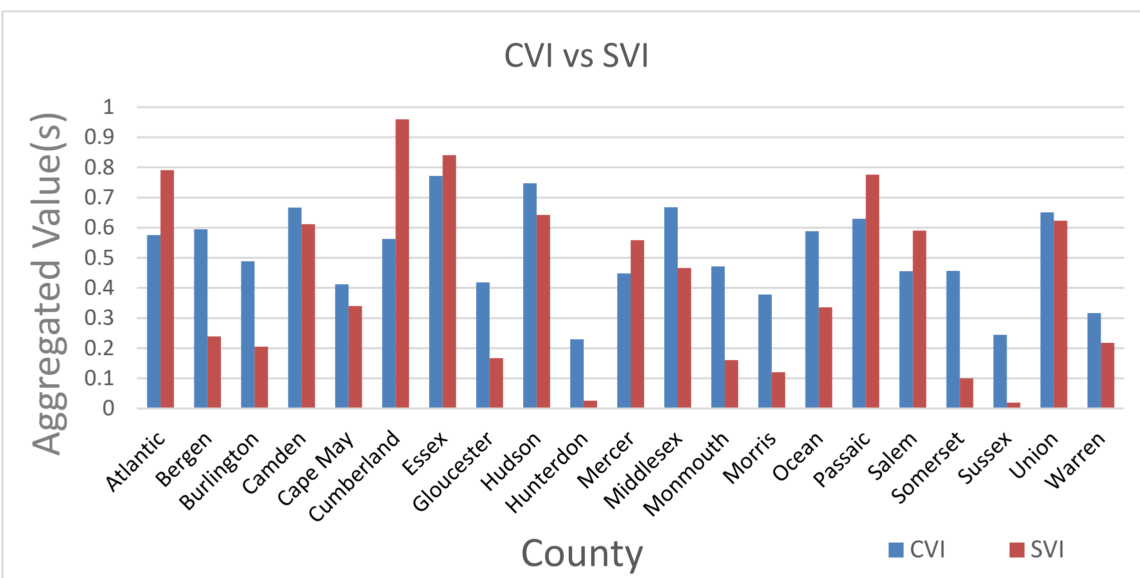


Figure 9 represents the number of COVID-19 Infections/ Population per county.

Figure 9

Bar Graph of the Amount of Infections/ Population per County

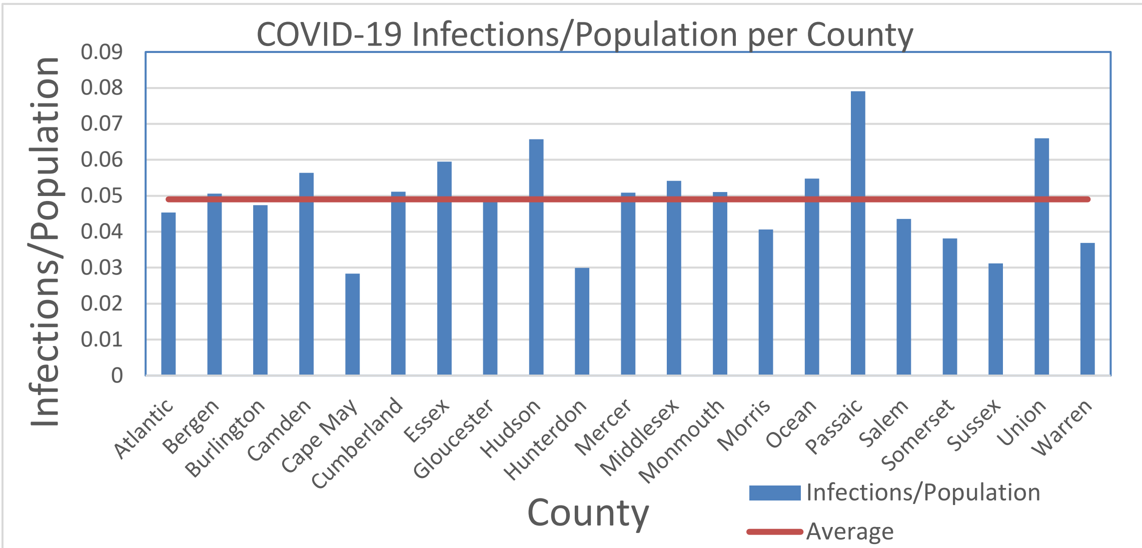


Figure 10 represents the number of COVID-19 Deaths/ Population per county.

Figure 10

Bar Graph of the Amount of Deaths/ Population per County

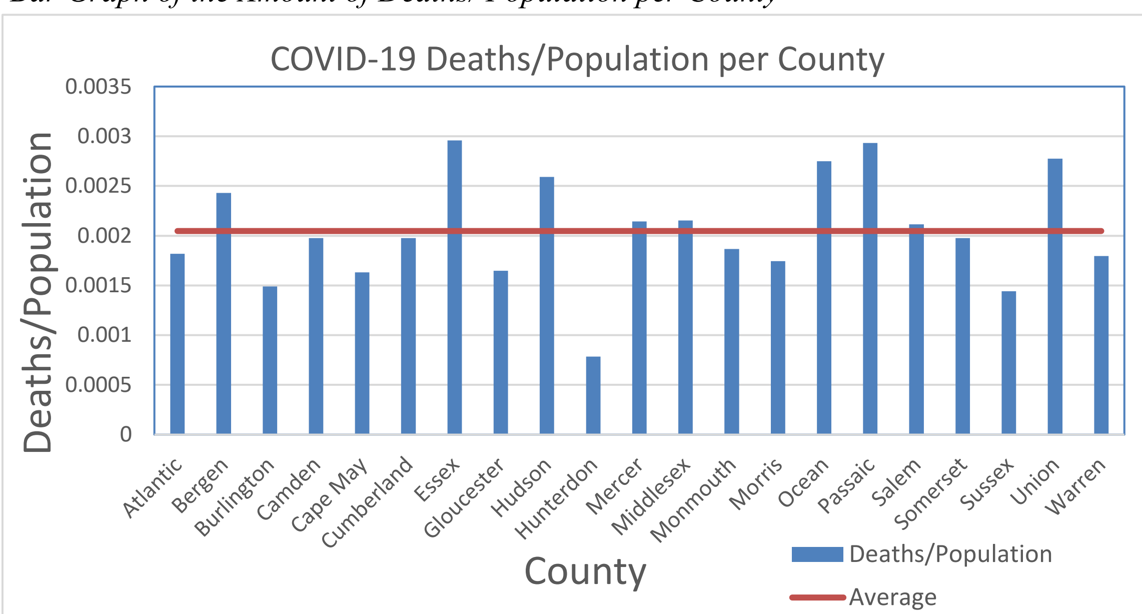


Figure 11 represents the Linear Regression Analysis that was conducted between the CVI and SVI values. As can be seen, the linear relationships between the two are not strong.

Figure 11

Linear Regression Plot of CVI vs. SVI

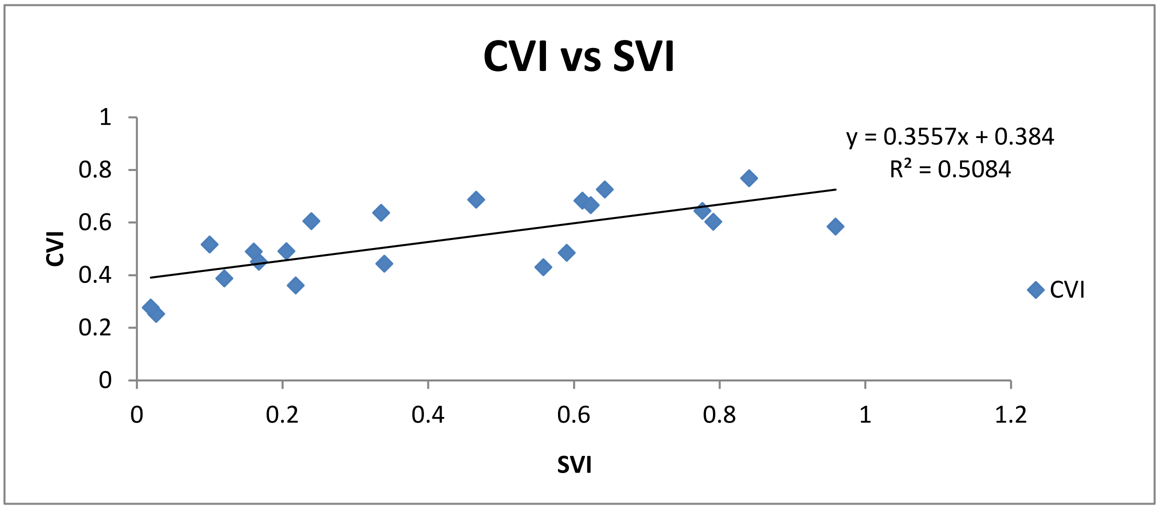


Figure 12 represents the Linear Regression Analysis that was conducted between the CVI values and COVID-19 Infections/ Population data.

Figure 12

Linear Regression Plot of CVI vs. Infections/ Population

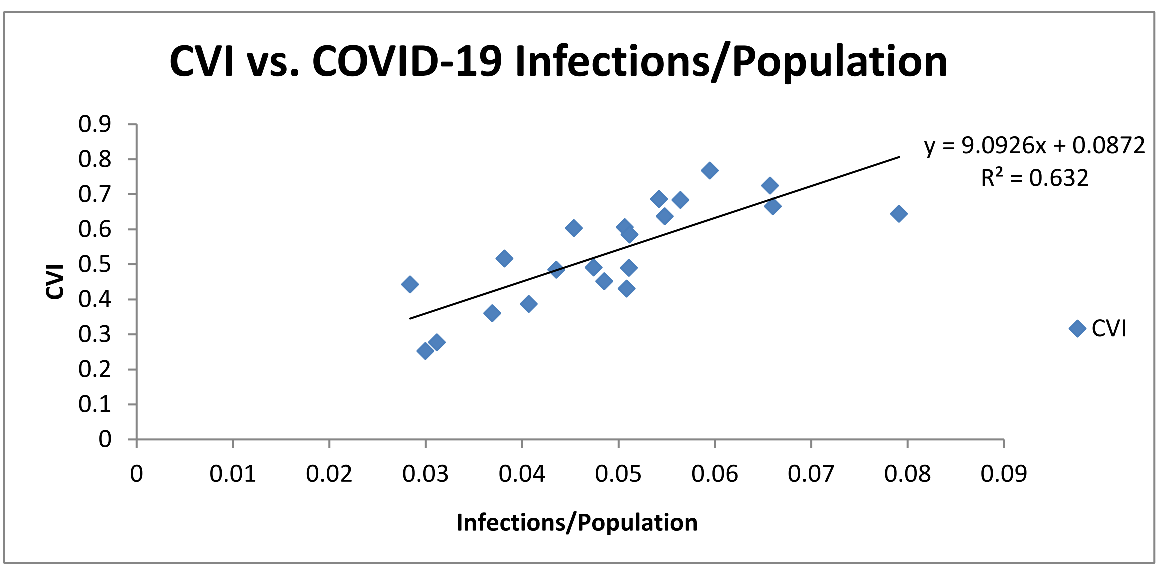


Figure 13 represents the Linear Regression Analysis that was conducted between the CVI values and COVID-19 Deaths/ Population data.

Figure 13

Linear Regression Plot of CVI vs. Death/ Population

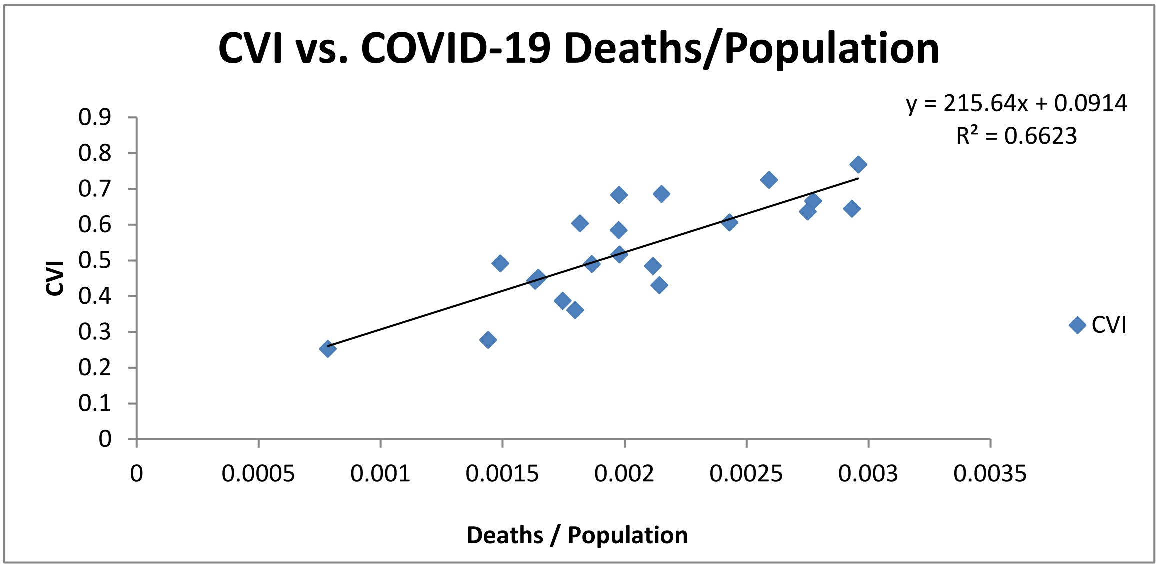


Table 6 represents the Linear Regression Analysis results that were conducted between the CVI, SVI, COVID-19 Infections/ Population, and COVID-19 Deaths/ Population.

Table 6

CVI Linear Regression Analysis

Linear Regression Analysis			
Regression Statistics	CVI vs. SVI	CVI vs. Infections/Population	CVI vs. Death/Population
Multiple R	0.712998	0.794955	0.813826
R Square	0.508366	0.631953	0.662312
Adjusted R Square	0.482491	0.612583	0.644539
Standard Error	0.10408	0.090053	0.086259
Observations	21	21	21
ANOVA: Significance F	0.000286	1.66E-05	7.2E-06

Figure 14 represents the Linear Regression Analysis that was conducted between the SVI values and COVID-19 Infections/ Population data. This graph highlights a weak linear relationship between the SVI and real world COVID data.

Figure 14

Linear Regression Plot of SVI vs COVID Infections/ Population Figure

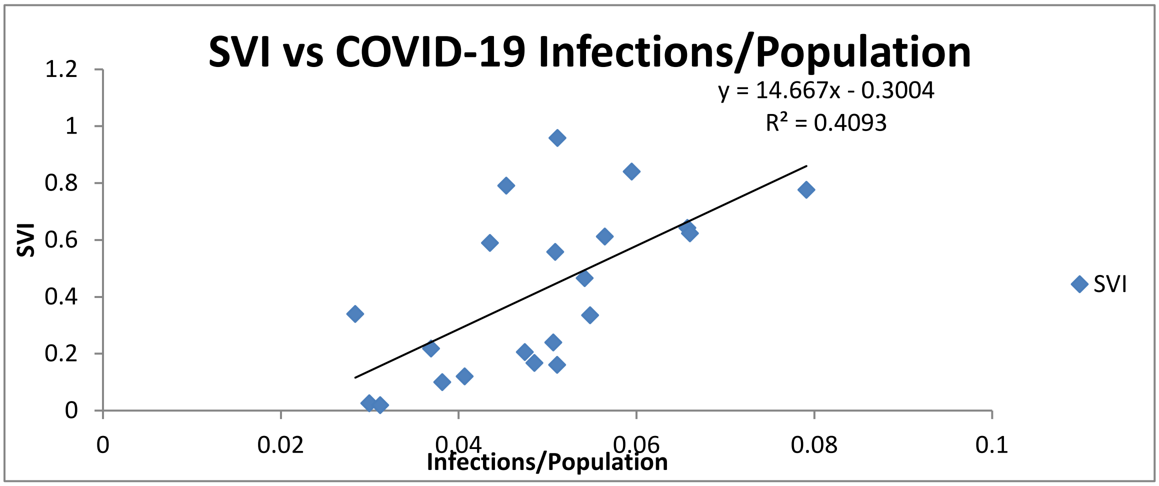


Figure 15 represents the Linear Regression Analysis that was conducted between the SVI values and COVID-19 Deaths/ Population data. This graph highlights a weak linear relationship between the SVI and real world COVID data.

Figure 15

Linear Regression Plot of SVI vs COVID Infections/ Population

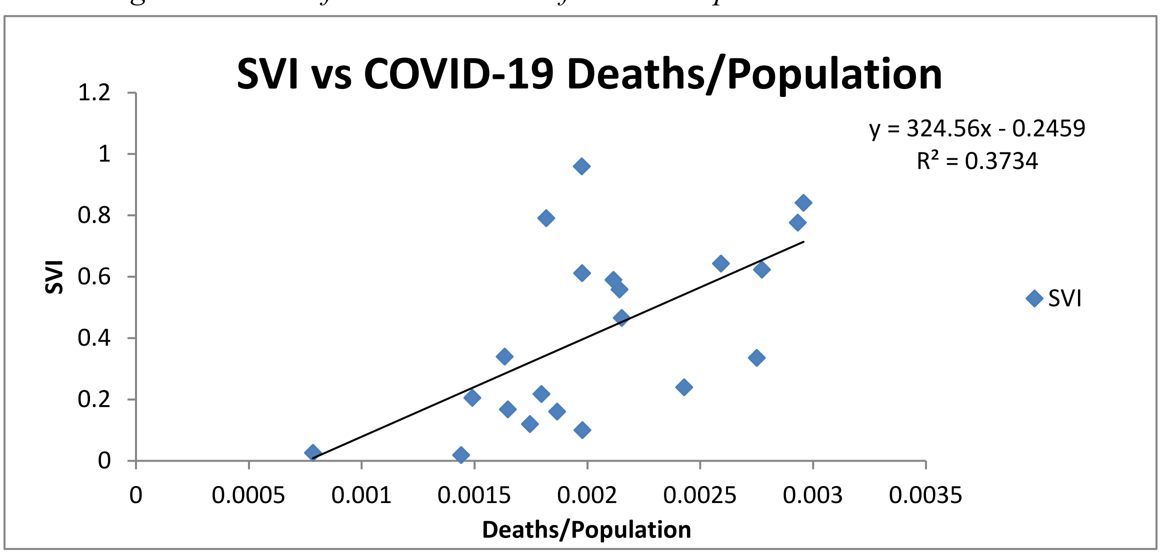


Table 7 represents the Linear Regression Analysis results that were conducted between the SVI, COVID-19 Infections/ Population, and COVID-19 Deaths/ Population.

***Table** 7*

SVI Linear Regression Analysis

Regression Statistics	*SVI vs Infection/Population*	SVI vs Deaths/ Population
Multiple R	0.639764403	0.611096913
R Square	0.409298492	0.373439437
Adjusted R Square	0.378208939	0.340462565
Standard Error	0.228674177	0.235512852
Observations	21	21
ANOVA: Significance F	0.001789075	0.003250658

Discussion

As shown, Table 3 represents the CVI theme values before they are summed and averaged into a final CVI value which can be then seen in Table 5. The values of each theme are in relation to the factors and their subsequent normalized value as illustrated in Figure 1 and described above. By acknowledging the different individual theme values, one can make delineations as to what counties may be more susceptible in certain areas/ themes as opposed to others. This was described in the SVI methodology as flagging specific values that may be higher than the average and taking note along the way. This was in part due to the number of themes and variables that contribute to the index and thus some indicators that are very high may not be represented in the final index value due to other lower values. That is why it is important to analyze the steps of index creation along the way and not whether the index is representative of the counties and factors that are being represented as a singular number.

Table 4 represents the number of 'Flagged' Factor values that are comprised in each theme. This is a methodology commonly applied to indices in order to determine what factors are the highest contributors to the overall index value. Table 4 aids in determining what themes and factors make the counites and their respective populations more susceptible to COVID-19. The values which are flagged are those above the 0.7 threshold on the scale from 0 to 1. The SVI utilizes the same methodology but caps the threshold at 0.9. Due to the size of the dataset and the highest resulting CVI values, it was determined that 0.7 would be more representative of the dataset and which values are to be of concern. This is due to the sample size of 21 counties as well as the maximum CVI value being 0.79. An example of this being Hunterdon County which has the lowest CVI score at 0.20. Even though the CVI score is low from all attributing themes and factors, there are 4 flagged factors under theme five, 'Health system capacity'. This means that in the case of Hunterdon County, accessibility to healthcare may be an issue in future pandemics even though the overall vulnerability score does not represent that specific information.

Table 5 as mentioned previously represents the final CVI values calculated for each county. This is the sole indicator from which all data and development has contributed to. The RANK.EQ column represents a normalized scaling factor: 1/21-21/21 as stated in Chapter 3. This is to help represent the data graphically as can be seen in Figures 3-6. This also helps rank which counties are most susceptible and least susceptible in relation to one another. Though this process hides the true calculated values as established in the column labeled CVI. That is the only reason why the RANK.EQ function is applied at the end; to get a quick determination of which county is highest, which is lowest and to represent that determination visually.

By conducting a linear regression plot of the CVI, SVI in comparison to real-world COVID data involving infection and mortality rates, it was determined that the calculated CVI values represent stronger linear relationships that that of the SVI. As can be seen in Figures 12, 13 and Table 6, the linear relationship between CVI values and COVID infection and mortality data is much higher in comparison to Figures 13, 14 and Table 7.

Conclusions

Again, Objective 2 of this study aimed to incorporate New Jersey county data into the CVI in order to validate the index. Through this study, a methodology for data collection and composition of a CVI was established, following validated steps when creating a composite index, generating a composite CVI, visually representing the data and validating the data through statistical analysis. By following current established literature, it was possible to conceptualize and eventually actuate a comprehensive vulnerability index for the counties of New Jersey. Not only that but establish strong linear relationships between established indices and real world COVID-19 data in order to validate the factors and methods comprised when creating the CVI. By establishing stated linear relationships, it is evident that the SVI is not a one size fits all in respect to pandemic vulnerability. This means that there is a need for a separate index for pandemic vulnerabilities due to the results gathered from the linear regression analyses. Granted, the SVI was never meant to be used in this application and thus by incorporating a different number of themes and relevant factors as described through methodology, it was possible to generate CVI values that represent COVID-19 vulnerability more favorably than that of the SVI.

Chapter 5

Conclusions and Future Work

The objective of this research was to establish a COVID-19 Vulnerability Index (CVI) and apply the CVI for the counties of New Jersey. This research is fundamental in establishing preventative measures in the event of future pandemics. By determining which counties are more likely to be vulnerable to COVID-19, it is possible to gain a better understanding of how pandemics affect different populations and how to account a for those populations through index development. By compiling a composite index through various avenues of publicly available data and researching previous index development works, a COVID-19 Vulnerability Index (CVI) for the counties of New Jersey was created and validated. This study is significant for future pandemics and index creation. This was possible by achieving the previously stated research objectives:

1) To develop a COVID-19 Vulnerability Index (CVI) to aid in preparedness for future global pandemics, and
2) To incorporate New Jersey county data into the CVI in order to validate the index.

Index Development

By analyzing the results from methodologies established in Chapter 3, the successful implementation of 8 themes and 39 different factors was accomplished. There are still opportunities for future studies to expand upon the research conducted in this study. As mentioned, some of the limitations of Chapter 3 are that there was effectively no direct method of weighting applied to the factors. For this reason, there were limited options regarding sensitivity and uncertainty analyses. In the future, it may be advisable to include a methodology of weighting for different case studies and areas that this research may be applied to. For the purposes of this study, it was seen as unnecessary to weight values considering the results as discussed. For other areas of study that may see prevalence in different factors a weighted scale will help determine COVID-19 vulnerability more accurately.

As seen in current literature there is a magnitude of angles that research can follow when establishing an index in this field. With that being said, the index could include more values that directly correlate to COVID-19 susceptibilities such as vaccine distribution and several other factors that may be relevant. Otherwise, further research should be focused on developing a time-varying index. With death, infection, and vaccination rates always changing through the event of a pandemic and thus the vulnerability of certain populations may change in the short or long term. This is a much more complex index to develop, but it would be ideal for long term impacts of a pandemic.

New Jersey Case Study

The second objective of this study was motivated by the state of New Jersey and the unique factors that surround it. By generating an index with factors specifically related to New Jersey and limiting the study area, the creation of a CVI for the counties of New Jersey was possible. Even though a CVI for each county of New Jersey is a wonderful and useful tool to have, it would be even more impressive if this research was broadened and included each of the municipalities of New Jersey and/or all of the counties of the United States. This would aid in understanding patterns related to the virus and provide an index for more people to use in the event of future outbreaks. This study can be expanded upon in effectively infinite dimensions. Though, the most practical direction for further research to follow would be to expand the study area. Otherwise, it would be nice to see if the CDC or future studies decide to incorporate unique factors specifically related to states, counties, or municipalities.

Conclusions

With the completion of this study, the following was accomplished and determined:

1) The development of a successful COVID-19 Vulnerability Index that used real-world COVID-19 data for validation.
2) An average CVI value of 0.5329 +/- 0.145 was determined for the 21 counties of New Jersey.
3) 36% of New Jersey's total population was determined to be in the High - Very High COVID-19 Vulnerability range.
4) Which counties that were most susceptible to COVID-19.

It seems that even in the 21st century, humans are not out of reach from global turmoil. The world was rather fortunate that the lethality of the virus was generally not reflected in younger generations. If COVID-19 had a higher mortality rate in lower age brackets (such as the bubonic plague) then it would have been a very rude awakening to for the scientific communities around the world. The past 2 years should be treated as a learning experience; when there is another pandemic with greater consequences than what has been experienced with COVID-19, the development of an index may be able to save thousands of lives. The development of technology, and the mutation of viruses that have been around for hundreds, if not thousands of years, will only contribute to this probability of a future pandemic. The development of an index to determine pandemic vulnerability is going to be more and more crucial over time. Therefore, it is imperative that humanity perseveres and takes preventative measures to preserve itself. As stated previously, the future work of this research is effectively limitless as time passes. More and more diseases and viruses will be developed or mutate and thus become the areas of study and concern for future indices and generations. Therefore, it is paramount that this research and future work is aimed at determining which populations are most susceptible to future pandemics.

References

1) Centers for Disease Control and Prevention. (2022, January 5). *CDC Museum Covid-19 Timeline*. Centers for Disease Control and Prevention. Retrieved April 17, 2022, from https://www.cdc.gov/museum/timeline/covid19.html

2) Centers for Disease Control and Prevention. (2021, August 27). CDC/ATSDR SVI data and documentation download. Centers for Disease Control and Prevention. Retrieved April 17, 2022, from: https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

3) Bureau, U. S. C. (2022, March 28). *American Community Survey (ACS)*. Census.gov. Retrieved April 15, 2022, from https://www.census.gov/programs-surveys/acs

4) Bureau, U. S. C. (2022, April 14). *Census.gov homepage*. Census.gov. Retrieved April 15, 2022, from https://www.census.gov/en.html

5) Health, D. of. (n.d.). New Jersey State Health Assessment Data New Jersey's Public Health Data Resource. NJSHAD -. Retrieved April 7, 2022, from https://www-doh.state.nj.us/doh-shad/

6) Blake Farmer, N. P. R., Pattani, A., Houghton, K., Bluth, R., Appleby, J., Peters, M., Graham, J., Wolfson, B. J., Rovner, J., & Yee, - S. (n.d.). Kaiser Health News. Retrieved April 7, 2022, from https://khn.org/

7) American Hospital Directory - information about hospitals from public and private data sources including MedPAR, OPPS, hospital cost reports, and other CMS files. (n.d.). Retrieved April 7, 2022, from https://www.ahd.com/NJ. (n.d.). Retrieved April 15, 2022, from: https://www.nj.gov/health/healthcarequality/documents/qi/pqi2018.pdf

8) *Mortality analyses*. Johns Hopkins Coronavirus Resource Center. (n.d.). Retrieved April 17, 2022, from https://coronavirus.jhu.edu/data/mortality

9) *Working together to show Covid's impact on communities of color*. Working Together to Show COVID's Impact on Communities of Color | Department of Civil & Environmental Engineering. (n.d.). Retrieved April 17, 2022, from https://cee.umd.edu/news/story/working-together-to-show-covids-impact-on-communities-of-color

10) R. C. Bolin and P. A. Bolton, “Race, religion, and ethnicity in disaster recovery,” *FMHI Publications,* vol. 88, 1986. [Online]. Available: USF Scholar Commons, http://scholarcommons.usf.edu/fmhi_pub/88. [Accessed April 2022]

11) OECD/European Union/EC-JRC (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, Paris, https://doi.org/10.1787/9789264043466-en

12) Centers for Disease Control and Prevention. (2012, May 18). *Principles of Epidemiology*. Centers for Disease Control and Prevention. Retrieved April 17, 2022, from https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section2.html

13) Anuj Tiwari, Arya V. Dadhania, Vijay Avin Balaji Ragunathrao, Edson R.A. Oliveira, using machine learning to develop a novel COVID-19 Vulnerability Index (C19VI), Science of The Total Environment, Volume 773,2021,145650, ISSN 0048-9697,https://doi.org/10.1016/j.scitotenv.2021.145650.(https://www.sciencedirect.com/science/article/pii/S004896972100718X)

14) Ofer Amram, Solmaz Amiri, Robert B. Lutz, Bhardwaj Rajan, Pablo Monsivais, Development of a vulnerability index for diagnosis with the novel coronavirus, COVID-19, in Washington State, USA, Health & Place, Volume 64,2020,102377, ISSN 1353-8292,https://doi.org/10.1016/j.healthplace.2020.102377.(https://www.sciencedirect.com/science/article/pii/S1353829220306298)

15) Morgan E. Gorris, Courtney D. Shelley, Sara Y. Del Valle, Carrie A. Manore,A time-varying vulnerability index for COVID-19 in New Mexico, USA using generalized propensity scores, Health Policy OPEN, Volume 2,2021 ,100052, ISSN25902296, https://doi.org/10.1016/j.hpopen.2021.100052. (https://www.sciencedirect.com/science/article/pii/S259022962100023X)

16) Surgo Ventures (2021) Vulnerable Communities and COVID-19: The Damage Done, and the Way Forward. https://precisionforcovid.org/ccvi.

17) Martins-Filho P. R. (2021). Relationship between population density and COVID-19 incidence and mortality estimates: A county-level analysis. *Journal of infection and public health*, *14*(8), 1087–1088. https://doi.org/10.1016/j.jiph.2021.06.018

18) *Code of ethics*. ASCE American Society of Civil Engineers. (n.d.). Retrieved April 17, 2022, from https://www.asce.org/career-growth/ethics/code-of-ethics

19) Zheng Li, Brian Lewis, Kevin Berney, Elaine Hallisey, Austin M Williams, Ari Whiteman, Luis O Rivera-González, Kristie E N Clarke, Heather Clayton, Terry Tincher, Jean D Opsomer, Michael P Busch, Adi Gundlapalli, Jefferson M Jones, vulnerability and rurality associated with higher SARS-CoV-2 infection-induced seroprevalence: a nationwide blood donor study, United States, July 2020 – June 2021, Clinical Infectious Diseases, 2022;, ciac105, https://doi.org/10.1093/cid/ciac105

20) Ibraheem M. Karaye, Jennifer A. Horney,The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships, American Journal of Preventive Medicine, Volume 59, Issue 3,2020,Pages 317-325,ISSN 0749-3797, https://doi.org/10.1016/j.amepre.2020.06.006. (https://www.sciencedirect.com/science/article/pii/S0749379720302592

21) Linear regression. (n.d.). Retrieved April 15, 2022, from http://www.stat.yale.edu/Courses/199798/101/linreg.htm#:~:text=A%20linear%20regression%20line%20has,y%20when%20x%20%3D%200).

22) Sameer Shadeed, Sandy Alawna, GIS-based COVID-19 vulnerability mapping in the West Bank, Palestine, International Journal of Disaster Risk Reduction, Volume 64,2021,102483, ISSN 2212-4209, https://doi.org/10.1016/j.ijdrr.2021.102483. (https://www.sciencedirect.com/science/article/pii/S2212420921004441)

23) *Healthy New Jersey*. Department of Health for the State of New Jersey Homepage. (n.d.). Retrieved April 7, 2022, from https://www.nj.gov/health/

24) Over 200,000 residents and staff in long-term care facilities have died from covid-19. KFF. (2022, February 3). Retrieved April 16, 2022, from https://www.kff.org/policy-watch/over-200000-residents-and-staff-in-long-term-care-facilities-have-died-from-covid-19/

25) Nursing homes. Homeland Infrastructure Foundation-Level Data (HIFLD). (n.d.). Retrieved April 7, 2022, from https://hifld-geoplatform.opendata.arcgis.com/datasets/geoplatform::nursing-homes/about

26) Vera-Institute. (n.d.). Vera-Institute/Incarceration-Trends: Incarceration trends dataset and Documentation. GitHub. Retrieved April 7, 2022, from https://github.com/vera-institute/incarceration-trends

27) County Health Rankings & Roadmaps. (n.d.). Retrieved April 7, 2022, from https://www.countyhealthrankings.org/app/new-jersey/2021/measure/factors/125/data

28) Environmental Protection Agency. (n.d.). EPA. Retrieved April 7, 2022, from https://www.epa.gov/superfund/search-superfund-sites-where-you-live

29) Monita Karmakar, P. D. (2021, January 29). *Association of sociodemographic factors with covid-19 incidence and death rates in the US*. JAMA Network Open. Retrieved April 16, 2022, from https://jamanetwork.com/journals/jamanetworkopen/article-abstract/2775732

30) Justin M. Feldman, S. D. (2021, November 23). *Covid-19 mortality in the US by race and ethnicity and educational attainment*. JAMA Network Open. Retrieved April 16, 2022, from https://jamanetwork.com/journals/jamanetworkopen/article-abstract/2786466

31) *September Public Meeting Minutes 2021 - Aspe.hhs.gov*. https://aspe.hhs.gov/sites/default/files/2021-09/low-income-covid-19-impacts.pdf

32) *LTC facilities outbreaks list - state*. (n.d.). Retrieved April 16, 2022, from https://www.state.nj.us/health/healthfacilities/documents/LTC_Facilities_Outbreaks_List.pdf

33) Ahmad K, Erqou S, Shah N, Nazir U, Morrison AR, Choudhary G, et al. (2020) Association of poor housing conditions with COVID-19 incidence and mortality across US counties. PLoS ONE 15(11): e0241327. https://doi.org/10.1371/journal.pone.0241327

34) Centers for Disease Control and Prevention. (n.d.). Coronavirus (covid-19) frequently asked questions. Centers for Disease Control and Prevention. Retrieved April 16, 2022, from https://www.cdc.gov/coronavirus/2019-ncov/faq.html

35) Yongjian Zhu, Jingui Xie, Fengming Huang, Liqing Cao, Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China, Science of The Total Environment, Volume727,020,138704, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.138704.(https://www.sciencedirect.com/science/article/pii/S004896972032221X)

36) X. Zhou, K. Josey, L. Kamareddine, M. C. Caine, T. Liu, L. J. Mickley, M. Cooper, F. Dominici, Excess of COVID-19 cases and deaths due to fine particulate matter exposure during the 2020 wildfires in the United States. Sci. Adv. 7, eabi8789 (2021).

37) Dpowell. (n.d.). Communicable disease service. Department of Health | Communicable Disease Service | New Jersey COVID-19 Dashboard. Retrieved September 8, 2021, from: https://www.nj.gov/health/cd/topics/covid2019_dashboard.shtml

38) Macharia, Peter & Joseph, Noel & Okiro, Emelda. (2020). A vulnerability index for COVID-19: spatial analysis at the subnational level in Kenya. 5. e003014. 10.1136/bmjgh-2020-003014

39) Claire E. Welsh, David R. Sinclair, Fiona E. Matthews, Static Socio-Ecological COVID-19 Vulnerability Index and Vaccine Hesitancy Index for England, The Lancet Regional Health Europe, Volume 14 ,2022, 100296, ISSN26667762, https://doi.org/10.1016/j.lanepe.2021.100296. (https://www.sciencedirect.com/science/article/pii/S2666776221002829)

40) *Federal system for tracking hospital beds and COVID-19 patients provides questionable data*. Science. (n.d.). Retrieved June 6, 2022, from https://www.science.org/content/article/federal-system-tracking-hospital-beds-and-covid-19-patients-provides-questionable-data

41) U.S. Bureau of Labor Statistics. (2022, June 3). U.S. Bureau of Labor Statistics. Retrieved June 6, 2022, from https://www.bls.gov/

Appendix

Sample Calculations and Data

Sample Calculations:

Equation 1: Theme 1 CVI calculation

$$
\begin{aligned}
&\Sigma\left[\frac{[a1 \ast s1 + a2 \ast s2 + a3 \ast s3 + a4 \ast s4 + a5 \ast s5]}{N(s)} + \frac{[b1 \ast t1 + b2 \ast t2]}{N(t)}\right. \\
&+ \frac{[c1 \ast u1 + c2 \ast u2 + c3 \ast u3 \ast c4 \ast u4 \ast c5 \ast u5 + c6 \ast u6 + c7 \ast u7]}{N(u)} \\
&+ \frac{[d1 \ast v1 + d2 \ast v2 + d3 \ast v3 + d4 \ast v4 + d5 \ast v5 + d6 \ast v6 + d7 \ast v7 + d8 \ast v8 + d9 \ast v9 + d10 \ast v10 + d11 \ast v11 + d12 \ast v12]}{N(v)} \\
&+ \frac{[e1 \ast w1 + e2 \ast w2 + e3 \ast w3 + e4 \ast w4 + e5 \ast w5 + e6 \ast w6 + e7 \ast w7]}{N(w)} + \frac{[f1 \ast x1 + f2 \ast x2 + f3 \ast x3]}{N(x)} \\
&\left. + \frac{[g1 \ast y1]}{N(y)} + \frac{[h1 \ast z1 + h2 \ast z2 + h3 \ast z3]}{N(z)}\right] / T
\end{aligned}
$$

Where:

a, b, c, d, e, f, g, h (x) = Weighting Factor(s)

s, t, u, v, w, x, y, z (x) = Factor Value

N (s, t, u, v, w, x, y, z) = Total number of Factors

T = Total Number of Themes

Raw Data:

	Theme 1: Socioeconomic Status				
County	**Poverty**	**Unemployed**	**No Highschool diploma**	**Uninsured**	**Per capita Income**
Atlantic	37471	13480	24304	24669	31366

Bergen	64674	22404	50921	76580	48426
Burlington	28151	14969	20101	19897	41517
Camden	62871	18876	40216	37569	34280
Cape May	10140	3490	6156	5510	38496
Cumberland	24830	4928	22944	14821	23946
Essex	127250	37731	75835	96506	37141
Gloucester	21275	9912	14261	13456	37888
Hudson	107718	22703	75223	93759	38147
Hunterdon	5828	2734	4335	3542	54984
Mercer	40026	13094	29717	29869	42155
Middlesex	67432	24243	57848	61923	38140
Monmouth	45873	17324	29723	39319	48959
Morris	22607	12939	19588	24961	55826
Ocean	62837	16052	33410	39050	34784
Passaic	82823	12613	55296	62230	30800
Salem	8067	2324	5464	3958	32526
Somerset	15319	8127	12158	17223	54393
Sussex	7191	4598	5291	6773	42639
Union	53602	18843	51661	64841	40201
Warren	8147	3700	6661	6589	38132

Atlantic	0.13954	0.06850	0.11559	0.09186	0.00003188
Bergen	0.06954	0.03162	0.06966	0.08234	0.00002065
Burlington	0.06307	0.04451	0.05722	0.04458	0.00002409
Camden	0.12392	0.05077	0.10294	0.07405	0.00002917
Cape May	0.10821	0.04740	0.07982	0.05880	0.00002598
Cumberland	0.16186	0.04401	0.19629	0.09662	0.00004176
Essex	0.16035	0.06655	0.12541	0.12161	0.00002692
Gloucester	0.07315	0.04593	0.06319	0.04626	0.00002639
Hudson	0.16110	0.04467	0.14169	0.14023	0.00002621
Hunterdon	0.04660	0.02817	0.04344	0.02832	0.00001819
Mercer	0.10854	0.02059	0.04579	0.03613	0.00002372
Middlesex	0.08157	0.05240	0.11882	0.09933	0.00002622
Monmouth	0.07359	0.04685	0.07678	0.07953	0.00002043
Morris	0.04573	0.02872	0.04226	0.04217	0.00001791

Ocean	0.10615	0.06050	0.11874	0.10589	0.00002875
Passaic	0.16432	0.03405	0.14435	0.12346	0.00003247
Salem	0.12737	0.04923	0.11032	0.06249	0.00003074
Somerset	0.04640	0.03285	0.04757	0.05216	0.00001838
Sussex	0.05053	0.04237	0.04677	0.04760	0.00002345
Union	0.09692	0.04668	0.12226	0.11724	0.00002488
Warren	0.07665	0.04573	0.07873	0.06199	0.00002622

Normalize Data by Population or Divide 1 by Value for Scaling:

Apply RANK.EQ Function

Atlantic	0.8095	1.0000	0.6667	0.6667	0.9048
Bergen	0.2857	0.1905	0.3810	0.6190	0.2381
Burlington	0.2381	0.4286	0.2857	0.1905	0.3810
Camden	0.7143	0.8095	0.5714	0.5238	0.8095
Cape May	0.6190	0.7143	0.5238	0.3810	0.4762
Cumberland	0.9524	0.3810	1.0000	0.7143	1.0000
Essex	0.8571	0.9524	0.8571	0.9048	0.7143
Gloucester	0.3333	0.5714	0.3333	0.2381	0.6667
Hudson	0.9048	0.4762	0.9048	1.0000	0.5238
Hunterdon	0.1429	0.0952	0.0952	0.0476	0.0952
Mercer	0.6667	0.0476	0.1429	0.0952	0.3333
Middlesex	0.4762	0.8571	0.7619	0.7619	0.5714
Monmouth	0.3810	0.6667	0.4286	0.5714	0.1905
Morris	0.0476	0.1429	0.0476	0.1429	0.0476
Ocean	0.5714	0.9048	0.7143	0.8095	0.7619
Passaic	1.0000	0.2857	0.9524	0.9524	0.9524
Salem	0.7619	0.7619	0.6190	0.4762	0.8571
Somerset	0.0952	0.2381	0.2381	0.3333	0.1429
Sussex	0.1905	0.3333	0.1905	0.2857	0.2857
Union	0.5238	0.6190	0.8095	0.8571	0.4286
Warren	0.4286	0.5238	0.4762	0.4286	0.6190

Sum Factor Values and Divide by Number of Factors

County	Theme 1
Atlantic	0.8095
Bergen	0.3429
Burlington	0.3048
Camden	0.6857
Cape May	0.5429
Cumberland	0.8095
Essex	0.8571
Gloucester	0.4286
Hudson	0.7619
Hunterdon	0.0952
Mercer	0.2571
Middlesex	0.6857
Monmouth	0.4476
Morris	0.0857
Ocean	0.7524
Passaic	0.8286
Salem	0.6952
Somerset	0.2095
Sussex	0.2571
Union	0.6476
Warren	0.4952

Sum Theme Values and Divide by Number of Themes