UNIVERSITY OF WATERLOO

Faculty of Environment

The trade-off between the volume of flight passengers and COVID-19: A case study in the United States in 2021

Ding Li (20866008), Lu Zhao (20922479), Wanxin Li (20922480),

Kioh Lin (20730918)

GEOG 481 GIS Project Report

Dr. Su-Yin Tan

April 4, 2022

Abstract

During the Novel Coronavirus disease 2019 (COVID-19) pandemic, flights services are declined significantly to control the infection rate. However, large amount of cancellation of flight services leads to much socio-economic loss and unsatisfied flight needs. Therefore, there is a trade-off between flight service capacity recovery and pandemic control. In this project, an index that can evaluate the performance of administrative units under different trade-off criteria is proposed with both theoretical proof and practical verification. By utilizing the index and a Moran's I analysis with queen-based spatial weight matrix, a case study in the United States (US) during January and October in 2021 will be conducted to research on the spatial pattern of how each state performs under some typical trade-off criteria. Afterwards, a modified Moran's I analysis method is designed based on a self-defined spatial weight matrix that emphasizes the flight service capacity between administrative units. By employing the modified the Moran's I analysis, the pattern of how flights and COVID-19 interact can be found. As a result, two states with higher volume of exchanged flight passengers tend to have similar pandemic control performance. The case applies especially on the southeastern part of the US.

GIS PROJECT

A CASE STUDY IN THE UNITED STATES IN 2021

Table of Contents

1. Introduction 3

1.1 Background 3

1.2 Mission Statement 7

1.3 Study Area 7

1.4 Literature Review 11

1.4.1 Influencing factors of the spread of COVID-19 11

1.4.2 Methods of monitoring population movement 12

1.4.3 Complex interactions between flights and COVID-19 need to be balanced 14

1.4.4 GIS applied to COVID-19 related research 16

1.4.5 Cartographic techniques used in the COVID-19 related research 18

1.5 Research Questions & Objectives 20

1.5.1 Research Questions 20

1.5.2 Research Objectives 20

1.6 Data Sources 21

2. Methodology 24

2.1 Mathematical symbol definition 24

2.2 Estimate of the severity of COVID-19 24

2.3 Index for evaluating the performance of a spatial unit under different criteria 27

2.4 Moran's I Analysis 32

2.5 Self-defined spatial weight matrix 35

2.6 General workflow 38

3. Results 41

3.1 Score 41

3.1.1 Distribution of Score 1 42

3.1.2 Distribution of Score 2 43

3.1.3 Distribution of Score 3 44

3.1.4 Comparing the distribution of Score 2 and Increased Infection Rate 45

3.2 Spatial Analysis and Mapping 47

3.2.1 Spatial Analysis for Score 1, Score 2 and Score 3 47

3.2.2 Spatial Analysis for increased rate of confirmed cases by modified Moran's I 51

3.2.3 Mapping for Input Passengers by Flight in Georgia 52

4. Discussion 55

4.1 Interpretation 55

4.2 Limitations of Data Sources, and Calculation of IR & OPE 58

5. Conclusion 60

5.1 Summary of Project Findings 60

5.2 Future Directions of Presented Work 61

5.3 Recommendation 61

References 62

Appendix I 75

Appendix II 79

1. Introduction

1.1 Background

Coronavirus disease 2019 (COVID-19) is one of the largest challenges for the world and scientists in any discipline, which has impacted everyone's life deeply. Starting from 2020, COVID-19 spreads globally from Wuhan, China and has caused 5,664,461 deaths among all 374,753,289 cases all over the world so far at the end of January 2022. According to the statistics provided by Johns Hopkins University (2022), the United States has more confirmed COVID-19 cases and more deaths than any other country so far. Therefore, the statistics support that the whole world has been affected by COVID-19 remarkably for both developed areas and developing areas.

Fortunately, some main factors that contribute to the spread of COVID-19 has already been proposed, based on which some regulations that can help control the pandemic can be clarified. According to the project by Chen et al. (2020), the transmission of COVID-19 can be influenced by population migration. Human mobility is also seen as one of the most crucial factors which is responsible for the continuous increase in the number of COVID-19 cases in the United States another research (Watts et al., 2020). Meanwhile, it is found that reducing population movements can effectively control the spread of the COVID-19 (Huang et al., 2020). Moreover, the COVID-19 outbreak in China has been brought under control in a relatively short period of time by the implementation of strict population movement control measures (Kraemer et al., 2020). Therefore, to control the pandemics, reducing population movements is necessary. Moreover, as shown in Figure 1.1, for distance larger than 1500 miles, about 82% people would like to travel by flight (U.S. Department of Transportation, 2011). Considering that there exists a lot of pairs of states have distance larger than 500 miles according to the data DataFromTo, it can be concluded that flight is a one of the most popular transportation methods for inter-state population movements (2022). Flights are likely to promote the propagation of COVID-19 as a main transportation for state-to-state population movements. There is also a study showing that the closer a region to an airport in the United States, the worse the COVID-19 (Desmet and Wacziarg, 2022). Therefore, declining flights is a necessary measure that must be taken to control the pandemic.

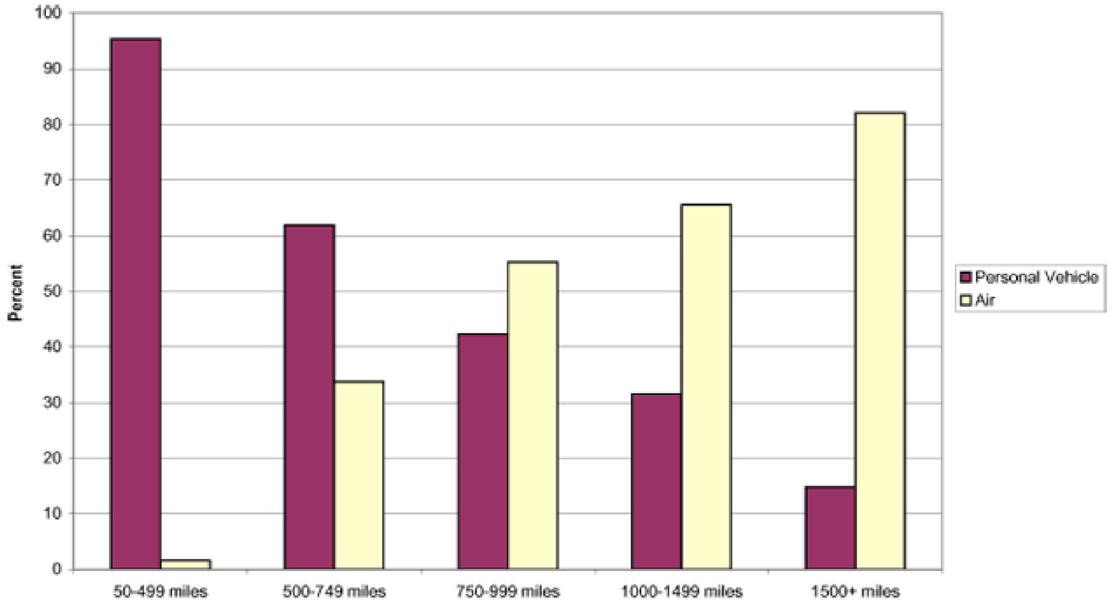


Figure 1.1 Trip length vs. transportation method selection from National Household Travel Survey (NHTS) in the U.S. in 2001 and 2002. Data source is from a report published by the U.S. Department of Transportation in 2011

However, based on the statistics of the Bureau of Transportation Statistics in the U. S. (2020), the United States has many domestic flights each year. The scheduled service airline passengers in 2021 have declined 27% to that in 2019 (the pre-pandemic period), which is demonstrated in the following Figure 2. A significant reduction in passengers and flights might cause much inconvenience and enforced negative changes in personal lives, which lowered residents' life satisfaction (Kessel et al., 2022). Furthermore, according to Figure 1.3 from a report from ICAO, over 40% of flight seats are declined comparing to pre-

COVID time. Passengers flown dropped by over 2000 million in 2021. The airline revenue loss is over 324 billion UDS. The case is even worse in 2020, which means people are trying to recover the flight service capacity, but there is still a long way to go (ICAO, 2022). The flight service capacity along with the whole aviation industry has been hit due to the cancellation of flights. As both flights and COVID-19 have a big impact on daily life, and they are contradicting each other, it is important to strike a trade-off between recovering flight service capacity and controlling COVID-19. Therefore, quantifying the performance of some administrative units under different trade-off criteria and seeing the spatial patterns worth a discussion, based on which some useful practical suggestions about how to balance flights and COVID-19 for some typical administrative units can be given.

Figure 1.2. U.S. Scheduled Service Airline Passengers from 2017 to 2021, from the Bureau of Transportation Statistics

Figure 1.3 Aviation industry hit by COVID-19. The figure is created by ICAO.

1.2 Mission Statement

This study is to verify that the flights and the spread of COVID-19 should be balanced based on the different background of states, create an index 'Score' to represent and evaluate the trade-off quantitatively, and apply that index to the case study in the United States. Meanwhile, the research will also focus on finding out the spatial distribution characteristics of trade-off flights and COVID-19 in each state in the United States. Finally, based on the obtained result of 'Score' and the spatial pattern several suggestions related to flight and COVID-19 policies will be provided.

1.3 Study Area

As a superpower, the United States is the center of global economic and cultural exchanges and one of the countries most affected by the COVID-19 pandemic, which has had the most significant number of infections and deaths due to COVID-19 in the world. Since the arrival of a vaccine to prevent COVID-19 in 2021, life has gradually returned to pre-outbreak levels as vaccination rates have improved across the United States. We note that domestic flights in the United States have recovered at least 60 percent more than in 2020 (Figure 3), and some states have recovered even more. The states in the United States have seen uneven changes in the pandemic situation, with some states showing a significant improvement in 2021 compared to 2020, while others are experiencing a more severe shortage (Figure 4). As the changes in flights and COVID-19 in the United States in 2021 compared with that in 2020 are represented globally, it is helpful to study the control of the pandemic in flights. Therefore, the United States is one of the most suitable places to explore the trade-off between flights and the spread of COVID-19.

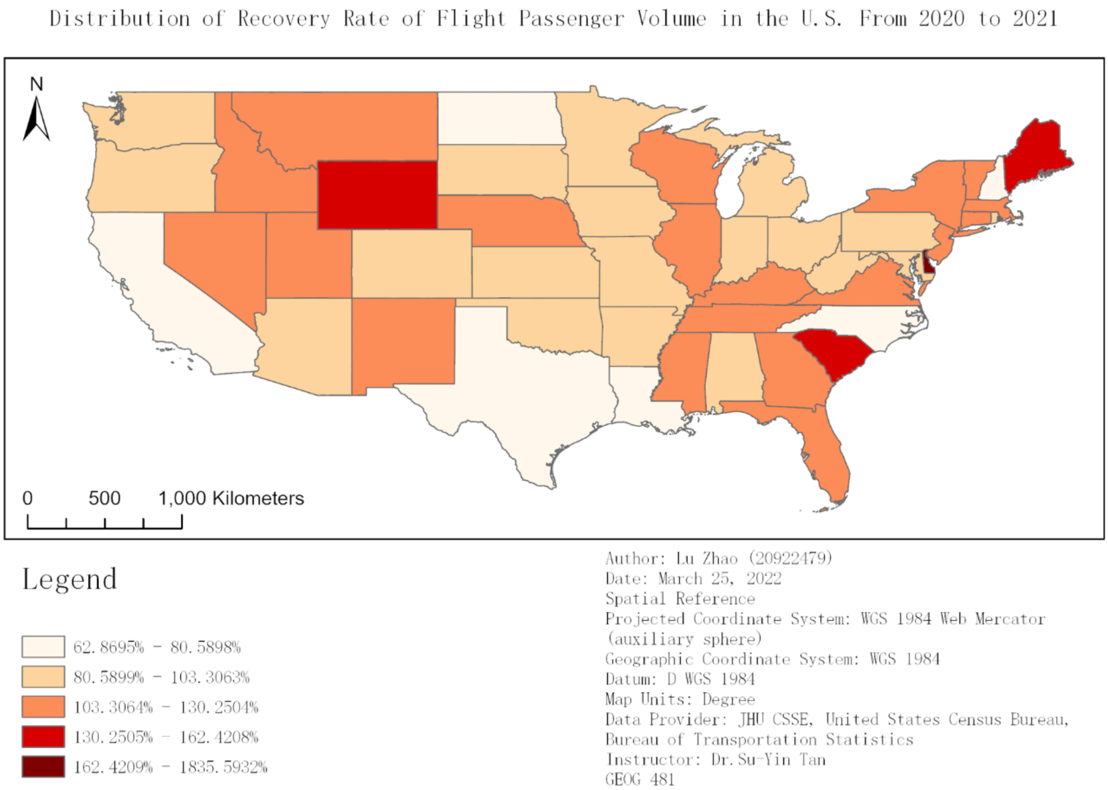


Figure 3. Distribution of recovery rate of flight passenger volume in the U.S. from 2020 to 2021

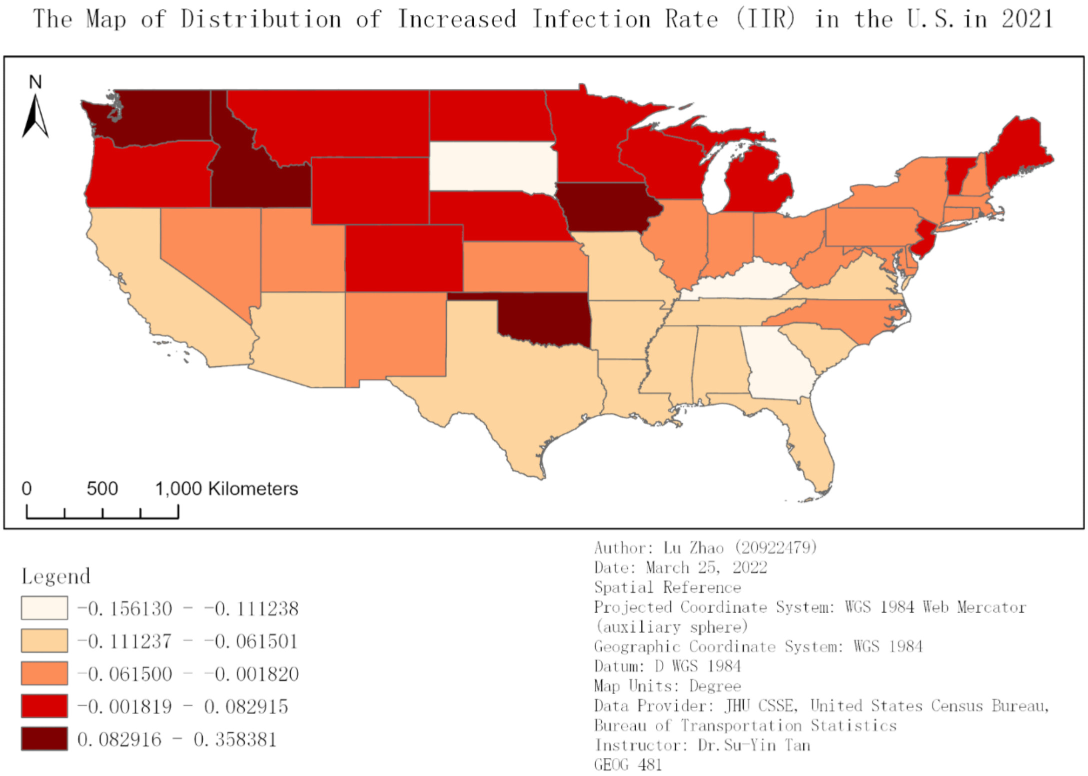


Figure 4. Distribution of increased rate of confirmed cases in the U.S. from January to October 2021

The study area is the continental United States. Considering the distances of Alaska, Hawaii, and other dependent islands from the continental states of the United States, maybe affect the results of conducting subsequent spatial autocorrelation analyses, this study is based on the 48 states of the continental United States (Figure 6). For the spatial analysis, the spatial unit states and the period unit were from January to October in 2021.

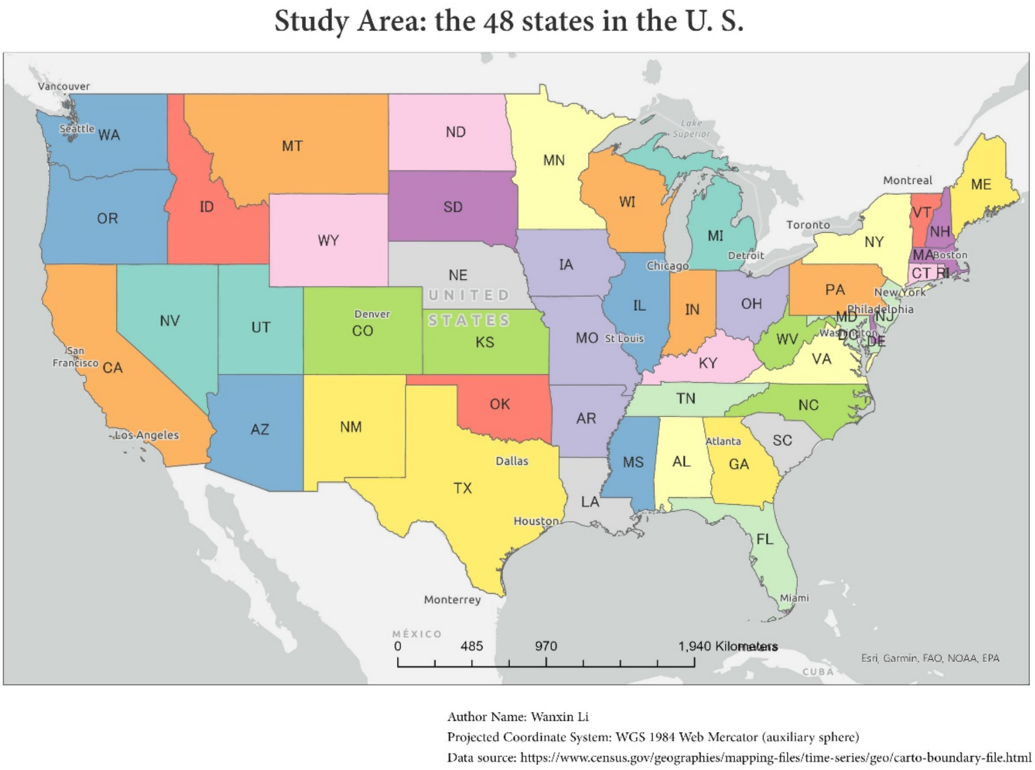


Figure 5. Study Area: The 48 States in the U.S.

1.4 Literature Review

COVID-19 has been considered as a hot issue in the recent two years, and research on factors affecting the spread of COVID-19 spread has been diversified worldwide, for instance, influencing factors.

1.4.1 Influencing factors of the spread of COVID-19

Previous research has studied various factors influencing the spread of COVID-19. Cordes and Castro (2020) support that race, education and income have significant effects on the infection rate in New York City, and there are spatial clusters in areas with severe epidemics. Based on a model Hu et al. (2020)

developed to predict the spread of outbreaks, and they found that an increase in the number of imported people and a lack of quarantine policies could increase the cumulative number of confirmed cases and even the possibility of renewed outbreaks in China. Hu et al. (2020) also believed that migration does contribute to the spread of the disease. Besides that, Yang et al. (2020) convinced that if increased population mobility, then the sprawl of COVID-19 pandemic would become more effortless and rapidly. Similarly, Kephart et al. (2021) analyzed the effect of population mobility on the COVID-19 spread based on the 314 Latin American cities and observed that when the weekly movement rate was reduced by 10%, the incidence of COVID-19 decreased by 8.6% in the following week. Chinazzi et al. (2020) used the global epidemic and mobility model (GLEAM) to analyze the impact of air travel restrictions on COVID-19, particularly in Wuhan, China, and they concluded that frequent air travel may highly affect the spread of COVID-19. Numerous studies on COVID-19 show the impact of the disease and the importance of doing research to find effective ways to mitigate the spread of the disease as soon as possible.

1.4.2 Methods of monitoring population movement

In previous studies on the impact of transportation or population movements on the spread of COVID-19, methods for monitoring or modeling population movements are also diversified. According to the data collected on the number of infected people in different provinces in China, Gross et al. (2020) found that the population of the province was correlated with the number of people migrating from Hubei and the distance to Hubei, and established a function to represent the corresponding relationship, which was used to analyze the impact of the population flowing out of Hubei on other regions. Gao et al. (2020) developed an interactive map platform based on web GIS to conduct real-time monitoring of population flow in the United States through quantitative data. It is a platform with high accuracy to simulate population flow, but it will pose a certain threat to users' privacy security (Gao et al., 2020). Gouissem et al. (2021) designed an Internet of Things sensor network to collect the movements of people to calculate exposure times between infected users and predict the probability of people becoming infected. They use different types of user protocols to protect user privacy and ensure the accuracy of the results (Gouissem et al., 2021). Mobile phone data can also be considered as an approach to quantify model population migration based on a small scale is more suitable and accurate than a national scale (Gabrielli et al., 2019). Twitter, Instagram, and other social media platforms can be used to monitor population mobility (Huang et al., 2020). Huang et al. (2020) analyzed human mobility at state scale in the United States based on more than 580 million tweets. However, this kind of method makes people who value their privacy feel uncomfortable (Huang et al., 2020; Kang et al., 2020). While these methods offer new insights into population mobility, the global penetration rate of smartphone users will only be about 38.5 percent by 2020., which is difficult to produce accurate population simulations with such low prevalence

(Gabrielli et al., 2019). Those are the reasons why there is very little research on quantifying population mobility directly and this method will not be used to calculate population flow in this study.

Beyond the above approaches to estimate the population movement, air travel is an important part of global population mobility, which can be used to represent the long-distance migration to some extent (Nizetic, 2020). Statistics collected by the authors show that 81% of government spending to stimulate transportation because of the COVID-19 pandemic went to airlines, air transport and airports, suggesting that flights are one of the essential modes of transportation in the United States (USAFacts, n.d.). Meanwhile, the European Commission (2021) supported that in the case of large-scale studies, it is not appropriate to establish and represent the situation and trend of population movement through road length statistics or vehicle data from toll stations Therefore, air passenger flow is used to be an indicator to simulate large-scale population movement in this study.

1.4.3 Complex interactions between flights and COVID-19 need to be balanced

Flights have been seriously affected in recent years, which suggests that there may indeed be a link between flights and the outbreak. Comparing and analyzing relevant statistics from 2019 to 2021, Duggal and Haddad (2021) supported that there were significantly fewer passengers and flights in 2020 and

2021 compared to the 2019 statistics. While daily life is gradually returning to normal in many countries and regions around the world in 2021, the number of flights and passengers is still nearly half of what it was before the pandemic (Duggal and Haddad, 2021). Moreover, Kommenda (2020) found that in many of the worst-affected countries, domestic flights in the United States declined sharply in the early days of the outbreak.

In addition to the direct impact of COVID-19 on flights, it will also have an indirect impact on financial, economics, tourism industries, energy market and many other aspects by affecting flights (Norouzi, 2021; Zhang and Zhang, 2021). Duggal and Haddad (2021) found that COVID-19 could be considered as a factor indirectly affecting global finance, particularly airline financing. Airlines have lost about $201 billion financially from 2020 to the present due to COVID-19 (Duggal and Haddad, 2021). Similarly, India has suffered from a similar trouble (Debata et al., 2020). Debata et al. (2020) found that the air transport ministry has lost more than 82 billion rupees since the mass cancellations and delays of international and domestic flights due to the outbreak in India. Machine learning and deep learning methods can also be used to discover the impact of the pandemic on flights and support the possibility that the transmission of novel Coronavirus may be one of the causes of flight cancellations (Mohammed et al., 2021). Meanwhile, Mohammed et al. (2021) and Nizetic (2020) argued that cancellations will negatively affect the reputation, operating costs, and future development of airlines, and increase the pressure on passengers to travel. Moreover, as the number of flights around the world has decreased by about 60%, oil markets, tourism and other industries have also been hit hard (Norouzi, 2021; Zhang and Zhang, 2021).

However, flights play a very important role which makes it impossible to cancel all flights due to control of the rapid spread of COVID-19 (Gheorghe and Sebea, 2010). Air Transport Action Group (2004) supported the presence of flights as creating jobs, contributing to GDP and tourism development, as well as providing humanitarian and medical assistance. Analogously, Gheorghe and Sebea (2010) provided the same conclusion about the significant benefits of air transportation. Therefore, the relationship between flights and the spread of COVID-19 is complex. So, in this study, we want to define an indicator to assess the balance between flights and the epidemic situation, which can be used by decision makers to timely adjust the policies related to flights and the epidemic situation and ensure people's normal life under the condition of maximizing benefits.

1.4.4 GIS applied to COVID-19 related research

In the research of the factors influencing the spread of COVID-19, there are several research methods related to the spread of COVID-19, among which GIS spatial analysis is commonly used (Franch-Pardo et al., 2020). The study of Guan et al. (2020) used GIS spatial analysis to analyze the spatio-temporal characteristics of the spread of COVID-19 in China, which can effectively help relevant personnel to determine the travel associated with confirmed patients and the scope of high-risk groups that may be infected. Padula and Davidson (2020) deem that GIS mapping and multi-level linear regression are good means to support the strong negative correlation between the registered nurse concentration and global COVID-19 mortality. Besides that, Oztig and Askin (2020) analyzed the relationship between human movement and the spread of COVID-19 using data related to air travel based on a negative binomial regression model and found a positive correlation between domestic travel and the incidence of COVID-19. This is another way to analyze factors influencing the spread of the epidemic in addition to spatial autocorrelation, and according to their results, flights have a negative correlation with the spread of the epidemic (Oztig and Askin, 2020).

Among the plenty of relevant studies based on GIS technologies, diversified spatial autocorrelation analysis is the most common to study factors that influence or mitigate the spread of the epidemic. Xiong et al. (2020) used spatial statistical model and GIS spatial analysis method to carry out spatial autocorrelation analysis of the outbreak of the COVID-19 pandemic in Hubei Province and found that this method can be used to find factors to alleviate the spread of the epidemic. Liu et al. (2021) used Moran's Index, Getis-Ord Gi, and Kulldorff's spatio-temporal scans as spatial analysis tools to find spatial clustering of confirmed COVID-19 cases. The authors also used a Geographically Weighted Regression (GWR) model to look for the impact of hypothetical factors on COVID-19 cases in mainland China (Liu et al., 2021). Sun et al. (2020) constructed a spatial weight matrix based on the first-order Queen rule to do a spatial correlation analysis to analyze the relationship between epidemic severity and geographic location in each state of the United States during the COVID-19 pandemic. However, using existing spatial relations to establish spatial weight matrix may not be suitable for all problems. Chen and Schumann (2013) convinced that it is very important to establish the algorithm of spatial weight matrix and to select the spatial weight matrix when calculating the Moran index by spatial autocorrelation analysis. Liu et al. (2012) developed a spatial clustering algorithm, called Density-Based Spatial Clustering Algorithm (DBSC), which is suitable for cluster analysis of flight status and dire epidemic status of high input populations. Also, Bourdin et al. (2021) analyzed the spread and concentration of COVID-19 in Italy based on a customized spatial weight matrix. Finally, a customized spatial weight matrix was constructed in this study.

1.4.5 Cartographic techniques used in the COVID-19 related research

In addition to research methods, mapping is also very important in the analysis of the relationship between population flow and COVID-19 pandemic. Mapping statics maps is a way to show the distribution and spread of COVID-19. Beecham et al. (2021) used custom coded ‘Glyphmaps’ to illustrate trends in confirmed cases, regional distribution patterns, and spatio-temporal patterns in the number of confirmed cases. Fang et al. (2021) tested 71 themes and different color schemes and data repositioning schemes, and found that maps with a mixture of blue, yellow, and red colors scored highest on contour maps, better representing the different levels of risk posed by the spread of COVID-19. Juergens (2020) considered that it is necessary to standardize the total number of confirmed cases in the production of thematic maps related to the epidemic, and to use ratios with corresponding meanings to represent confirmed figures. Besides that, the author also deem that contour maps can provide a better visual effect in space, thus more intuitively displaying COVID-19 prevalence rates and other data and information in each state (Juergens, 2020; Li, 2021). However, Kent (2020) supported the use of icons or sorting tables to represent acquired disease mortality more intuitively than contour maps, and the use of highly saturated red as a legend can be more appealing to the reader.

Dynamic maps are also a popular way to analyze the spread of the COVID-19 pandemic, which is timelier and time-sensitive than static maps (Rinner, 2021). Pase et al. (2021) believe that maps can be used not only to static represent quantitative data related to diagnosis, but also to create dynamic maps that reflect changes in the evolution of the epidemic and how it will change in the future.

1.5 Research Questions & Objectives

1.5.1 Research Questions

In this study, the research questions are in the following:

1) What is the relationship between the number of flight passengers and the spread of COVID-19 from January to October in 2021?

2) How to quantify and compare the trade-off between different states from January to October in 2021?

3) What is the spatial distribution pattern of the trade-off degree between the volume of the flight passengers and COVID-19 pandemic in the U.S. from January to October in 2021?

1.5.2 Research Objectives

This study aims to analyze the different patterns of trade-off in each state in the U.S. from January to October in 2021 and apply the self-defined index 'Score' to quantify and compare different trade-offs between different states in the United States from January to October in 2021. The specific objectives are in the following:

1) Create a ‘Score’ index to quantify the trade-off between flight passengers and the spread of COVID-19 pandemic in each state in the target research period, with ‘Score 1’’Score 2’’Score 3’ by R language and python.

2) Based on the spatial autocorrelation analysis to identify whether ‘Score 1’’Score 2’’Score 3’, and increased rate of confirmed cases are spatially correlated by using Global Moran’s I and Local Moran’s I in GeoDa and ArcGIS Pro.

3) Analyze the relationship between flight passengers and the increase of COVID-19 in the U.S. from January to October in 2021.

4) Analyze the factors that cause the differences in the trade-off pattern between states in the United States.

1.6 Data Sources

According to research questions and objectives, Flight Data, Boundary Vector Data and COVID-19 Datasets are be chosen. The summary of data source is demonstrated in Table 1.

Num#	Dataset name	Provider	Data type	Resolution/ Scale	Temporal Resolution	Spatial Resolution
1	T-100 Domestic Market (U.S. Carriers) – Flight Data	Office of Airline Information in the United States	Multivariate	Not applicable	Month	City
2	cb_2020_us_state_500k – U.S. Boundary Vector Data	United States Census Bureau	Vector data	1:500,000	Year of 2020	State
3	Johns Hopkins University CSSE COVID-19 Dataset – COVID-19 Datasets	JHU CSSE	Multivariate	Not applicable	Day	County

Table 1 Data Summary Source

The flight data, T-100 Domestic Market datasets contains domestic market data reported by U.S. air carriers, including carrier, origin, destination, and service class for enplaned passengers, freight, and mail when both origin and destination airports are located within the boundaries of the United States and its territories. It can be used to collect the interstate flight data from January to October in the U.S.

The cb_2020_us_state_500k – U.S. dataset includes the boundary vector data of the United States and states released by the United States Census Bureau contains attribute data (Name of 48 states and affiliated island) as well as spatial data (geographic location). It can be used to create a map of the movement of COVID-19 cases to display and analyze the direction and mode of transmission of the pandemic.

The Johns Hopkins University CSSE COVID-19 Dataset underlying the map, provided on the CSSE GitHub (url: https://github.com/CSSEGISandData/COVID-19), is licensed under the Creative Commons Attribution 4.0 International by the Johns Hopkins University on behalf of its Center for Systems Science in Engineering. Attribute the data to the "COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University" or "JHU CSSE COVID-19 Data" for short. It can be used to estimate the infection rate and increased infection rate of COVID-19.

2. Methodology

2.1 Mathematical symbol definition

For the convenience of expression and the purpose of this study, the mathematical expressions and symbols that will appear below will be defined here. First, let S denote the set of all spatial units in the study area. In this case study, S contains all the states in the US. Then, since the study focuses on domestic state-to-state travels by flight, two flight passenger transition matrices T^{2021} and T^{2020} are defined, where for any two states i, j in the study area, $T_{i,j}^{n}$ represents the overall number of passengers travel from state i to state j during January and October in year n. Those matrices are called flight passenger transition matrices. Based on the definition of flight passenger transition matrices, the overall flight passengers exchanged by flight (OPE) in state i during January and October in year n can be calculated by:

$$OPE_i^n = \sum_{j \in S} \left(T_{i,j}^n + T_{j,i}^n \right) \qquad (2.1)$$

. For the epidemic data, let c_i^n denote the number of confirmed COVID-19 cases in state i from January to October in year n. Let n_i^n denote the number of tested people for COVID-19 in state i from January to October in year n.

2.2 Estimate of the severity of COVID-19

To estimate the severity of the pandemics in a state, the infection rate (IR) of the state will be used. In the dataset provided by JHU, several potential measures of the severity of COVID-19 are included, which are the mortality rate, IR, confirmed cases, and dead cases. However, according to the metadata, the number of test cases vary in different states. Also, considering that different states themselves have different population attributes (e.g., the number of residents), only rate attributes are appropriate to compare the severity of COVID-19 in different states. Therefore, only mortality rate and IR are more appropriate. Nonetheless, since many mistakes are usually made when reporting death cases, which introduce too much inaccuracy to the mortality rate attribute and make the attribute unreliable according to the metadata. Furthermore, according to an experiment done by researchers from UK, the infection fatality ratio is just about 0.66% (Mahase, 2020), which is an extremely small number. That means, COVID-19 affects human's world mainly based on the infection itself, not death. Therefore, our study will only consider the IR of each state as the only appropriate measure of the severity of the pandemics in that state. Unfortunately, having every resident tested once a month is impossible, so the true value of IR of any state can't be obtained in real-life. Hence, sets of sample proportions $\hat{\sigma}_i^{2021}$ and $\hat{\sigma}_i^{2020}$ are introduced to estimate the IR in state i between January and October in year n, which is calculated by:

$$\hat{\sigma}_i^n = \frac{c_i^n}{n_i^n} \qquad (2.2)$$

Note that since $\hat{\sigma}_i^n$ is just a sample proportion, the capability for it to estimate the proportion in the whole population is uncertain. Thus, a statistical inference for the capability and variability must be conducted. In statistics, Bernoulli trials refer to the random trials that can only have two outcomes (e.g., success or failure), and the probability of each outcome remains the same each time (Papoulis, 1984). Hence, in the context of this study, each COVID-19 test for any person is either positive or negative, which has only two outcomes. Also, by selecting people randomly, the true IR σ_i^n can be well represented by the sample proportion $\hat{\sigma}_i^n$, which is also called the maximum likelihood estimate of the true proportion of infected population (Donnelly, 2016; LE CAM, 1990). In particular, since $\hat{\sigma}_i^n$ is expected to be very small in light of the data provided by JHU, but the sample size is extremely large, so a Normal model is the most common way to approximate the Binomial distribution with such a large sample size n_i^n, rather than a Poisson model (2021; Sheu, 1984; Chen, 1975; Feller, 1945; Hall, 1982). Specifically, the Normal model would be $N\left(n_i^n\hat{\sigma}_i^n, \sqrt{n_i^n p(1-\hat{\sigma}_i^n)}\right)$ in the context of this study. Therefore, the radius of the confidence interval (CI) that quantifies how certain that $\hat{\sigma}_i^n$can estimate $\hat{\sigma}_i^n$ in state i during January and October in year n by confidence level p is obtained by the formula:

$$rad_i^{(n)} = \alpha\sqrt{\frac{\hat{\sigma}_i^n(1-\hat{\sigma}_i^n)}{n_i^n}} \qquad (2.3)$$

, where α is defined such that $P(Z \leq \alpha) = \frac{1+p}{2}$ and Z follows a standard Normal distribution $N(0, 1)$. The analysis result will give very valuable information about the accuracy of the whole study. In general, the most plausible way of estimating IR of a state during January and October in year n is to divide the confirmed cases by the total test cases of the state during that time, but the error is unavoidable.

2.3 Index for evaluating the performance of a spatial unit under different criteria

In this section, an index that is a function of trade-off criteria is proposed for evaluating the performance of a spatial unit has during a period under different criteria. As mentioned before, too high volume of flight passengers will lead to worse pandemics, whereas too low volume will cause much socio-economic loss. In this scenario, Trade-off Theory proposed by Campbell and Kelly in 1994 can be well applied. In the theory, the community can hardly satisfy every requirement simultaneously when there exists inconsistency of social-choice axioms of Arrow. That is, decisions will be made based the preference of individuals (Campbell et al., 1994). Applied to our study, recovering flights will avoid the huge socio-economic loss, whereas keep limiting flights is expected to damage the pandemic control, which is exactly the contradiction and inconsistency proposed by Campbell and Kelly. Therefore, according to Trade-off theory, different people will have different preferences between flight passenger volume and pandemic control, which means there is no unified or optimal standards to measure how well the trade-off between flights and COVID-19 a unit does, which means the flexibility must be given when designing such an index. Therefore, criteria vector $\boldsymbol{v}$ is defined as $\boldsymbol{v} = (\alpha, \beta)$, where α represents by how much the individual cares about flight passenger volume, and β represents by how much the individual cares about the COVID-19 control. By adjusting $\boldsymbol{v}$, different criteria based on different preferences can be represented. Individuals that care about flight service recovery will assign higher scores to the states where the number of flight passengers recovers intensively comparing to the previous year. Hereby, the definition of the recovery rate of flight passenger volume (RRFP) in state i in 2021 can evaluate by how much the state recovers its flight service capacity, which is calculated by:

$$RRFP_i = \frac{OPE_i^{2021} - OPE_i^{2020}}{OPE_i^{2020}} \qquad (2.4)$$

. For the individuals that hope to control COVID-19, higher scores are expected to be assigned to the states with lower infection rates comparing to the previous year. Hence, the increase rate of confirmed cases (IRCC) in state i from 2020 to 2021 is proposed to quantify by how much the pandemic in the state is worsened. Based on Theorem 1 in the Appendix section, the attribute is estimated by:

$$IRCC_i = \frac{\hat{\sigma}_i^{2021} - \hat{\sigma}_i^{2020}}{\hat{\sigma}_i^{2020}} \qquad (2.5)$$

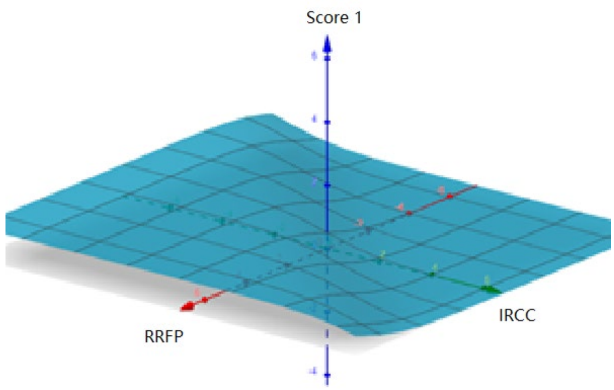
. However, there might be a case where there is a huge difference between the base unit of IRCC and RRFP. For example, if IRCC values are naturally larger than RRFP values, then setting $\boldsymbol{v} = (\alpha, \beta) = (1,1)$ does not make IRCC and RRFP equivalently weighted. Therefore, normalization of RRFP and IRCC can help minimize the problem to some extent (Choo et al., 1999). The normalization function g is defined in the Appendix, directly from the article finished by Patro et al. (2015). Afterward, to represent the score, a function that is decreasing and monotone with respect to IRCC, meanwhile increasing and monotone with respect to RRFP must be found. With the inspiration of sigmoid functions that are widely used in machine learning and statistics, the score is designed to inherit the excellent properties of the functions. Specifically, the definition of the score of state i that quantifies the performance of the state during

January and October in 2021 under criteria vector $\boldsymbol{v} = (\alpha, \beta)$ is:

$$Score_i(\boldsymbol{v}, RRFP_i, IRCC_i) = \frac{1}{1 + e^{-\alpha \cdot g(RRFP_i) + \beta \cdot g(IRCC_i)}} \qquad (2.6)$$

. By inheriting excellent properties from sigmoid functions, the Score function is always continuous everywhere, lying within the range $(0,1)$ that is convenient to convert to percentage as a usual score. Also, about the differentiability, any order derivative of the function is still differentiable, which is important for verifying the variation of the function with respect to any independent variables (Kyurkchiev et al., 2019; Ezeafulukwe et al., 2018). Hereby, people who prefer to recover flight service will set the criteria vector with α value significantly larger than β value. Then, under this criteria, higher scores will be assigned to the states with higher capacity of recovered flight service. People who prefer to see better pandemic control should set the criteria vector with α value significantly smaller than β value, so that higher scores will be given to the states with lower IRCC. Furthermore, people who think that flight service recovery and pandemic control are equally important, then the α value and the β value of the criteria vector should be equivalent. The proof of the above statements is provided in the Appendix. In this research, three typical scores will be implemented as a case study. The first score $Score1_i = Score_i\big((1,1), RRFP_i, IRCC_i\big)$ is derived from the most balanced trade-off criteria, where the RRFP and IRCC of the state are equally weighted. The function image and contour plot of the function is shown in Figure 2.1; The second score $Score2_i = Score_i\big((0,1), RRFP_i, IRCC_i\big)$ is from extreme trade-off criteria, where flight service recovery is completely ignored and all the concern is on pandemic control. The function image and contour plot of the function is shown in

Figure 2.2; The third score $Score3_i = Score_i\big((1,0), RRFP_i, IRCC_i\big)$ is from another extreme trade-off criteria, where flight service recovery is regarded as the most crucial task, whereas the evaluation for the pandemic control will be completely abandoned. The function image and contour plot of the function is shown in Figure 2.3. For each state in the study area, three scores will be assigned based on the definition of Score 1, Score 2, and Score 3. For each of Score 1, Score 2, and Score 3, a map of the study area that shows the distribution of the performance of each state will be output. By visualizing and doing some spatial analysis on the score distribution, a spatial pattern of the performances of the states under different trade-off criteria can be revealed. Additionally, besides the completeness of the Score function in mathematics, the hypothesis that the function is indeed a realistic index for estimating the performance of spatial units under different trade-off criteria will be further verified using the score maps of the study area in the result and interpretation section.



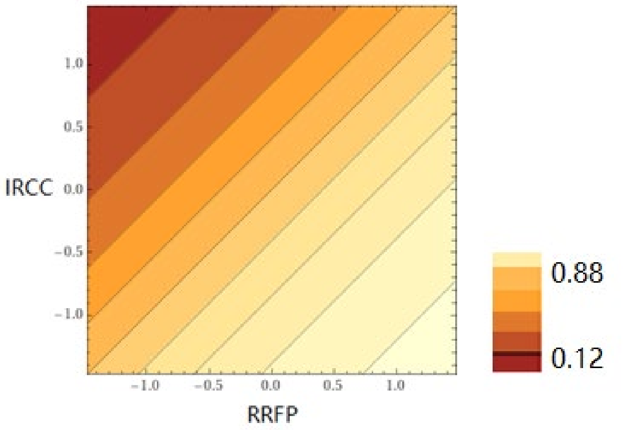


Figure 2.1 Function plot and contour plot of Score 1

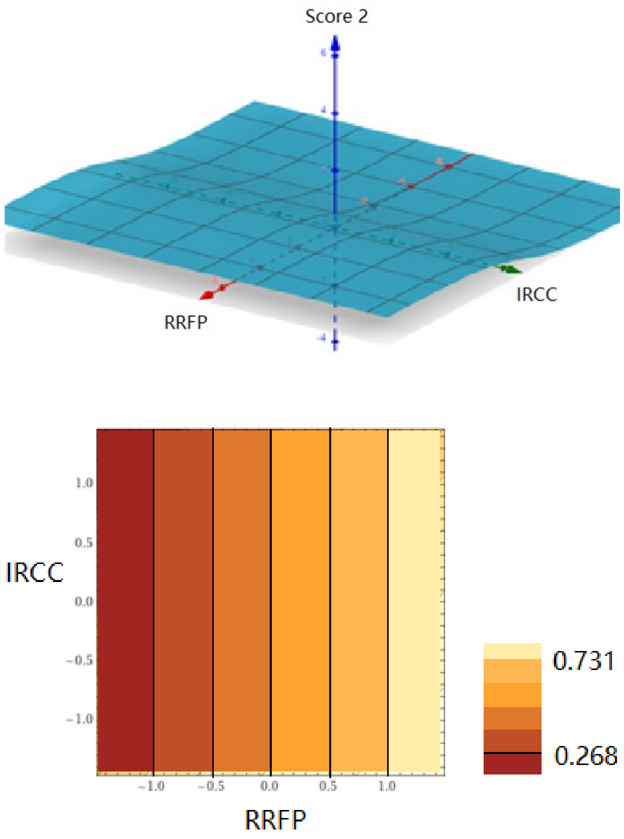


Figure 2.2 Function plot and contour plot of Score 2

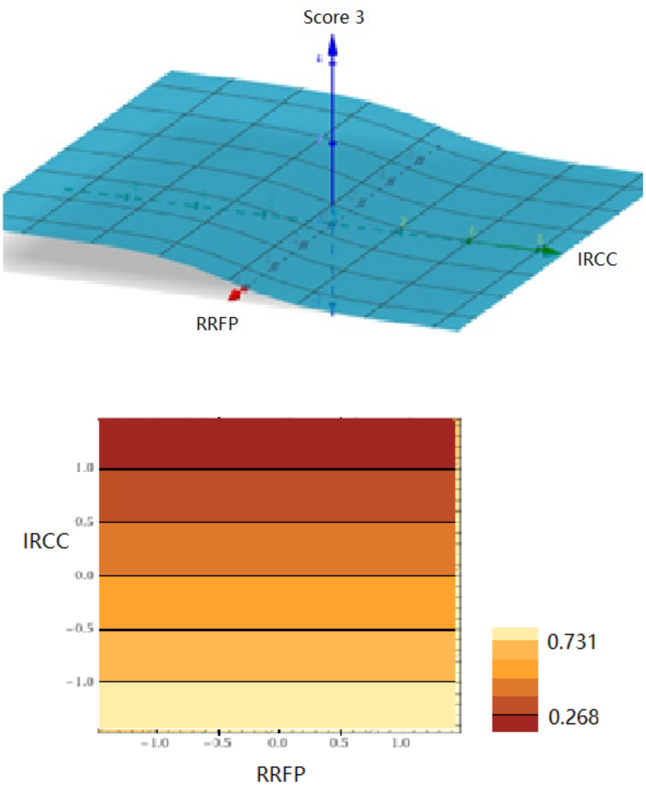


Figure 2.3 Function plot and contour plot of Score 3

2.4 Moran's I Analysis

In this project, Global Moran's I and Local Moran's I are the most critical spatial analyses. According to an article finished by Jared, spatial clustering analysis is one of the most powerful tools for epidemiologic and criminological studies (Jared, 2010). As one of the most popular spatial clustering analyses, Global Moran's I can determine the existence of spatial autocorrelation of an attribute of the spatial units in a feature class based on a spatial weight matrix (Dormann et al., 2007). For any two spatial units in the feature class, there will be a spatial relationship between them (Cheng et al., 2014). Traditionally, the spatial relationship between any pair of spatial units is negatively related to the spatial distance between the units. The rationale is from the famous

Toblers first law of geography, which indicates that closer spatial units have stronger spatial relationships (Miller, 2004). Also, contiguity based spatial relationships are also widely used, which indicate that there is relationship between neighboring spatial units. And there is no relationship if the units are not neighbors topologically (Getis, 2009). In Moran's I, a spatial weight matrix can be defined based on any types of spatial relationships. Then, based on a specific type of spatial relationship, if some attribute has a strong positive spatial autocorrelation, then spatial units with stronger spatial relationships will have similar attribute values (Smelser et al., 2001). In general, Global Moran's I have two outputs. The first output is a pseudo p-value, where lower pseudo-p-value indicates that the attribute does not have a strong autocorrelation. In other words, the spatial pattern of the attribute is random. Otherwise, there is likely a strong spatial autocorrelation (Li et al., 2007). In practice, a pseudo p-value smaller than 0.05 is strong enough to conclude that the attribute has a significant spatial autocorrelation. The second output of Global Moran's I is the Global Moran's I index, which is either positive or negative. For positive index with pseudo p-value that is small enough, the attribute is said to have a significant positive spatial autocorrelation. On the other hand, negative index with pseudo p-value that is small enough indicates that the attribute has a strong negative spatial autocorrelation, which means the difference between the attributes of the spatial units with stronger spatial relationships are even large (Smelser et al., 2001). However, Global Moran's I is sometimes too general and dependent on the scale of the study area, which can hardly identify which zonal areas are more significantly clustered. Therefore, to check the spatial units whose attributes are especially clustered based on the spatial correlation, Anselin proposed Local Moran's I (Anselin, 1995). In Local Moran's I, every spatial unit will have a pseudo p-value and a local Moran statistic (i.e., Local Moran's I index). For the spatial units with positive statistics and pseudo p-values, the attribute of those units has a strong positive spatial autocorrelation locally around the spatial units. Also, by utilizing cluster and outlier analysis, clusters of spatial units that have significant clustering or discrete spatial patterns will be revealed on map, which is called a LISA map (Anselin, 1995). In summary, for Global Moran's I, positive index with a pseudo p-value smaller than 0.05 indicates that in this feature class, spatial units with stronger spatial correlations will have similar attribute values in general. For Local Moran's I, spatial units with positive index and pseudo p-value smaller than 0.05 implies that other units that have strong spatial correlation with the unit will have similar attribute value with the unit. In this study, Global and Local Moran's I will be conducted based on the queen-based contiguity spatial weight matrix, which is described above. The reason is that for policy-oriented and socio-economic-oriented attributes, queen-based spatial relationship between spatial units is the most relevant according to some research (Crandall et al., 2004; Flores et al., 2019; Murray et al., 2014). Therefore, as the performance of flight service recovery and pandemic control are considered highly socio-economic and policy related according to the introduction section, queen-based contiguity spatial weight matrix is selected for the global and local Moran's I analysis. In conclusion, Global and Local Moran's I based on queen-based contiguity spatial weight matrix will be conducted to see whether neighboring states tend to share similar performance under different trade-off criteria. Next, a Moran's I method with a self-defined spatial weight matrix derived from OPE will be introduced.

2.5 Self-defined spatial weight matrix

In the project, a new spatial matrix defined based on the population exchanged by flight is introduced for modified Global and Local Moran's I analysis. As mentioned in the literature review section, some previous work involves defining new spatial weight matrix other than using the traditional ones. Also, according to the article finished by Chen, Moran's Index is just a characteristic parameter of spatial weight matrices, which reveals the significant relationship between Moran's Index and the weight matrix people use (2013). Hence, for unusual spatial analysis work, traditional matrix can hardly be suitable. Back to our study, flight is one of the main inter-state transportation methods for population movements, which plays a role in exchanging pathogen as discussed in the introduction section. Also, since in this project, other transportation methods (e.g., muting, driving, etc.) are not considered, and the likelihood of people selecting flight as the transportation method even becomes larger as the distance of the trip increases as shown in the introduction section, traditional distance-based or topology-based spatial weight matrix can't perform appropriately for the analysis of the relationship between flight and COVID-19 in this project. Therefore, a new definition of spatial weight matrix derived from population exchanged by flight is needed to satisfy the need of this project. Considering that high volume of flight passenger exchange (PE) between two states is expected to boost the relationships between the two states, the novel spatial weight matrix for year n is defined as:

$$w_{i,j}^{n} = T_{i,j}^{n} + T_{j,i}^{n} \qquad (2.7)$$

. To compare the novel spatial weight matrix with a traditional spatial weight matrix (in this case, inverse distance based), an example is provided. As demonstrated in Figure 2.4, the distance between Arizona and Illinois is about 1313 miles, and the distance between Arizona and New Mexico is about 244 miles. In the traditional spatial weight matrix, the spatial relationship between Arizona and Illinois is defined as $\frac{1}{1313}$. And the spatial relationship between Arizona and New Mexico is $\frac{1}{244}$. Since $\frac{1}{1313}$ is smaller than $\frac{1}{244}$, New Mexico has a much stronger spatial relationship with Arizona in the tradition spatial weight matrix. The rationale behind is that closer spatial units have stronger relationships. On the contrary, since during January and October in 2021, the total number of passengers exchanged between Arizona and Illinois by flight is 2133602, and there are only 443174 passengers exchanged between Arizona and New Mexico during that time. Since larger PE between two spatial units is considered to indicate stronger spatial relationship between them, Illinois should have a stronger spatial relationship with Arizona in the self-defined spatial weight matrix. Indeed, under the new definition of spatial weight matrix, the spatial relationship between Arizona and Illinois is 2133602 as the PE between them, which is significantly larger than the relationship between Arizona and New Mexico 443174. Therefore, by using the self-defined spatial weight matrix, the spatial relationships between spatial units are just the PE between them. Note that in this project, the modified global and local Moran's I based on the self-defined spatial weight matrix follow the formula and rationale in

Moran and Anselin’s papers, which was discussed under the previous section (Moran 1950; Anselin, 1995). And the outputs are generated from the R scrips we personally implemented. In this case, a significant positive autocorrelation means that the spatial units with higher PE between each other will have similar attribute values.

Figure 2.4 Traditional definition of spatial relation between Arizona-Illinois and Arizona-New- Mexico. The distance data is from travelmath

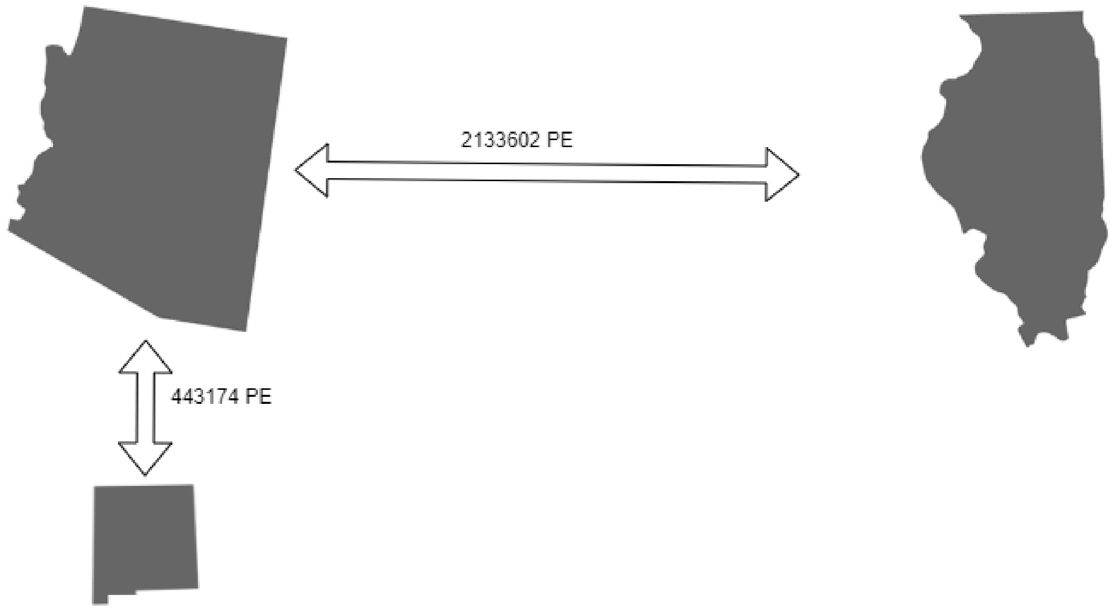


Figure 2.5 Novel definition of spatial relation between Arizona-Illinois and Arizona-New- Mexico based on PE between each pair of states. The PE data is from the T-100 domestic market dataset from Office of airline information in the US

2.6 General workflow

The workflow of this project is as demonstrated in Figure 2.6, which implements all the methods discussed above. First, the US boundary vector data is clipped according to the study area by using select by attribute tool in ArcGIS. Then, by developing and running some Python scripts, flight passenger data and COVID-19 data from January to October in 2020 and 2021 are filtered out based on the study area that includes 48 mainland states in the US. Then, by using database management tools in SQL and ArcGIS, the datasets are dissolved into a state basis with temporal resolution of 10 months. Additionally, for the flight passenger dataset, only the original state, destination state, number of passengers flew from the original state to the destination state are kept as the only attributes, based on which the flight passenger transition matrix T^{2020} and T^{2021} are generated. Afterward, RRFP and the self-defined matrix are calculated from T^{2020} and T^{2021} based on equation 2.1, 2.4, and 2.7. For the COVID-19 dataset, only the state name, number of total test cases, and number of confirmed cases are kept as the only attributes, based on which the IR and IRCC of each state are computed by equation 2.2 and 2.5. By using attributes RRFP and IRCC, Score 1, Score 2, and Score 3 are computed based on their definition and equation 2.6. Then, based on the US boundary vector data, the maps of performance of each state under different trade-off criteria are made. Furthermore, queen-based Global and Local Moran's I will be conducted to find the spatial patterns of the performance of the states under different trade-off criteria. After that, the modified Global and Local Moran's I algorithms are implemented in R based on the self-defined matrix and the formula provided from Moran and Anselin's papers. Then, after doing the modified Moran's I analysis, the pseudo p-values and the Moran's I indices will be returned. From the global p-value and the Global Moran's I index, the spatial patterns of the performance and the interaction between flight and COVID-19 will be explained. Finally, some practical suggestions for improving the flight recovery and pandemic control performance of some typical states will be given based on the results.

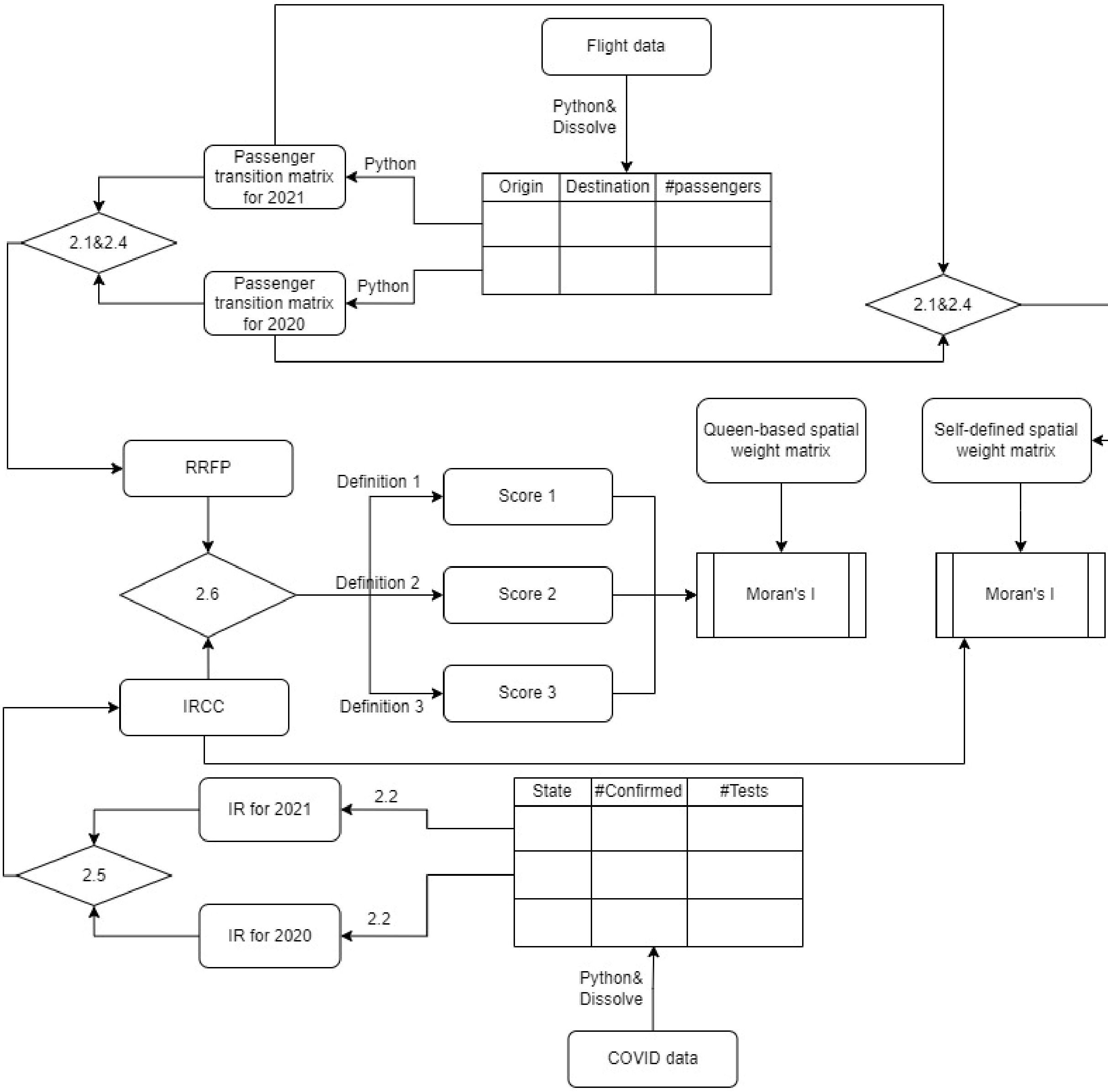


Figure 2.6 Workflow diagram

3. Results

In this section, the score maps of the US will be output. The results of traditional local and global Moran's I analysis for the score distributions based on queen-based spatial weight matrix will be shown to see if there is a positive spatial autocorrelation. Finally, the modified global and local Moran's I analysis for the IRCC will be demonstrated to see if the IRCC of different states cluster based on the volume of flight passengers exchanged between the states.

3.1 Score

The results in Table 1 review the selection of criterion vectors α, and β that are related to Score1, Score 2 and Score 3 defined in the methodology section.

	α	β	$Score(\alpha, \beta, RRFP, IRCC)$
Score 1	1	1	$\frac{1}{1+e^{-g(RRFP_i)+g(IRCC_i)}}$
Score 2	0	1	$\frac{1}{1+e^{g(IRCC_i)}}$
Score 3	1	0	$\frac{1}{1+e^{-g(RRFP_i)}}$

Table 2. Selection of criterion vectors for Score 1, Score 2 and Score 3

In this study case, Score 1 represents a (1:1) weighting of flight cancellation and increased overall infection rate, Score 2 (1:0) represents the extreme case where only IRCC is considered and flight cancellation is ignored, and Score 3 (0:1) represents the extreme case where only flight cancellation is considered and increased overall infection rate is ignored.

The results in Figure 4, Figure 5 and Figure 6 demonstrate the distribution of Score 1, Score 2 and Score 3 in the U.S. from January to October 2021. The Scores have been divided into five ranges from 0 to 1. The distribution of the range of scores for each of the 48 states in the study area can be obtained from the figures.

3.1.1 Distribution of Score 1

Since Score 1 represents a (1:1) weighting of flight cancellation and increased overall infection rate, a higher Score 1 indicates a lower flight cancellation and a lower increased overall infection rate simultaneously. According to Figure 4, we can find that Score 1 for most states is under 0.378086.   There is the only state that has a Score 1 higher than 0.378086, the state of Delaware located on the east part. In addition, in Figure 3, it can be found that score 1 is generally lower in the northern region than in the southern region. And Wyoming, North Dakota and Oklahoma, Washington have the lowest Score 1.

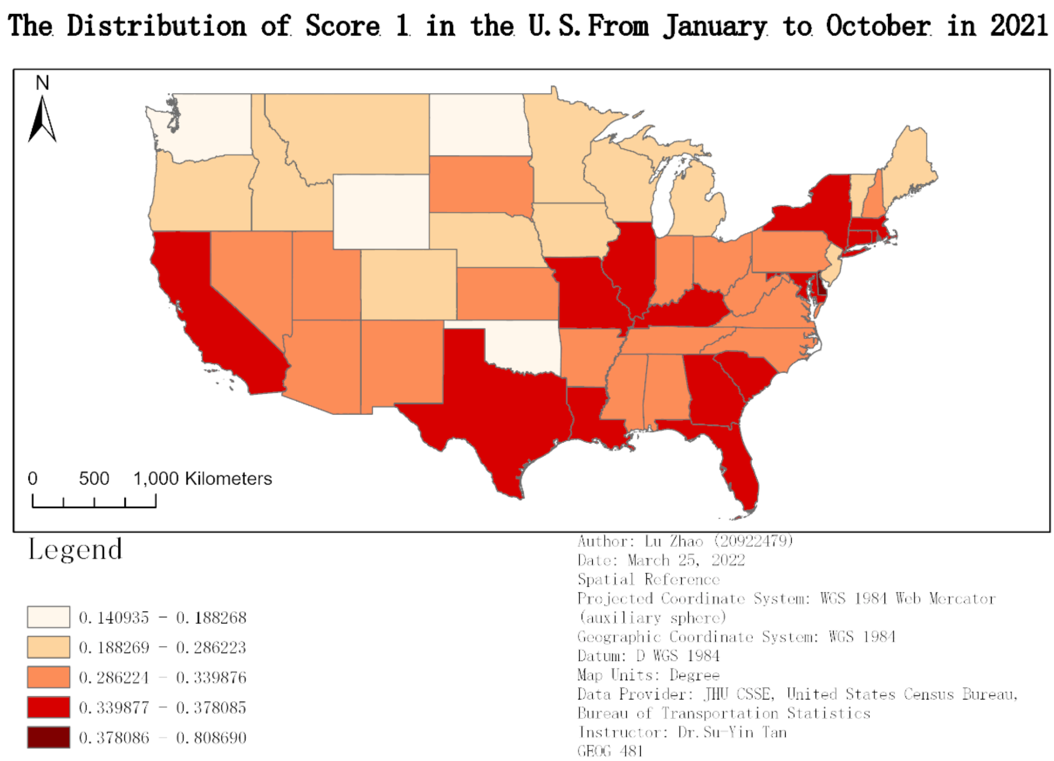


Figure 4. The distribution of Score 1 in the U.S. from January to October 2021

3.1.2 Distribution of Score 2

Score 2 represents the extreme case where only an increased overall infection rate is considered and flight cancellation is ignored, therefore a higher Score 2 indicates a lower IRCC. Based on Figure 5, it can be found that score 2 of the southern region is higher than that of the northern region. Take California, Florida, and Georgia as examples, these states in the south have both a higher Score1 and Score 2. Moreover, Wyoming, North Dakota, Oklahoma, and Washington, the four states with the lowest scores in score 1, still have the lowest scores in Score 2.

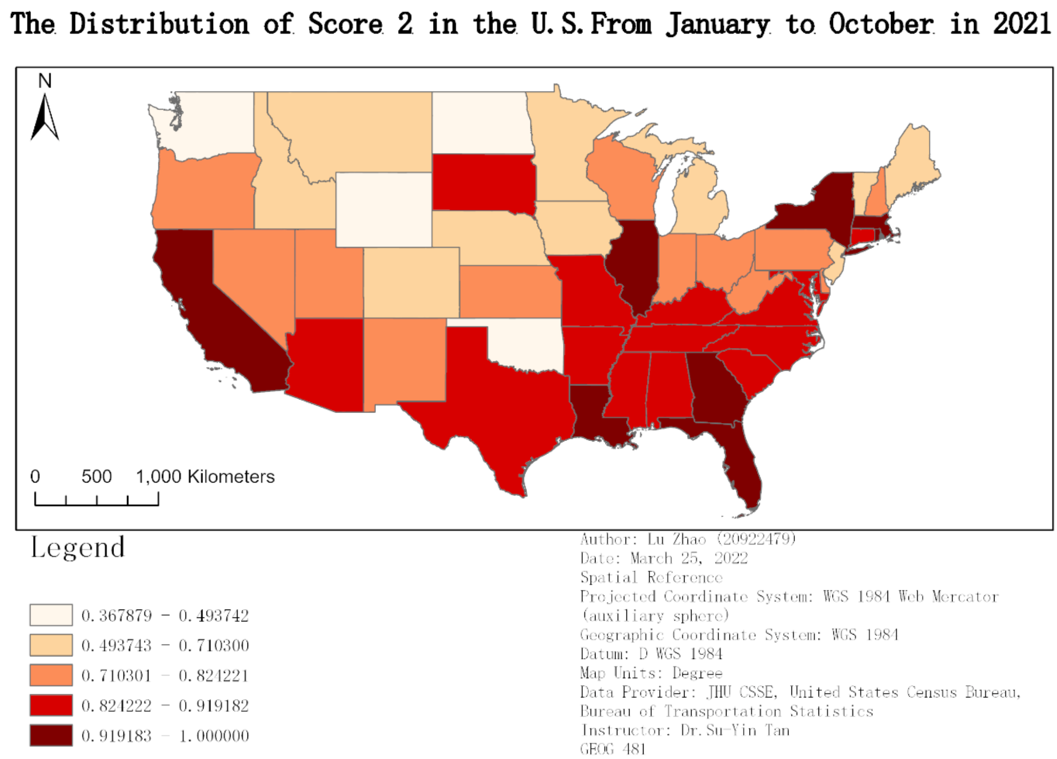


Figure 5. The distribution of Score 2 in the U.S. from January to October 2021

3.1.3 Distribution of Score 3

Score 3 (0:1) represents the extreme case where only flight cancellation is considered and increased overall infection rate is ignored, therefore a higher Score 3 indicates a lower flight cancellation. From Figure 6, we can find that most states have a score 3 below 0.39615. The values of Score 3 are mainly around 0.38. The range of Score 3 is from 0.378085 to 0.396150. Among all the states, Wyoming, Maine, and South Carolina have significantly higher scores. The score 3 of California, Texas, Louisiana, North Carolina, and North Dakota are sub-par.

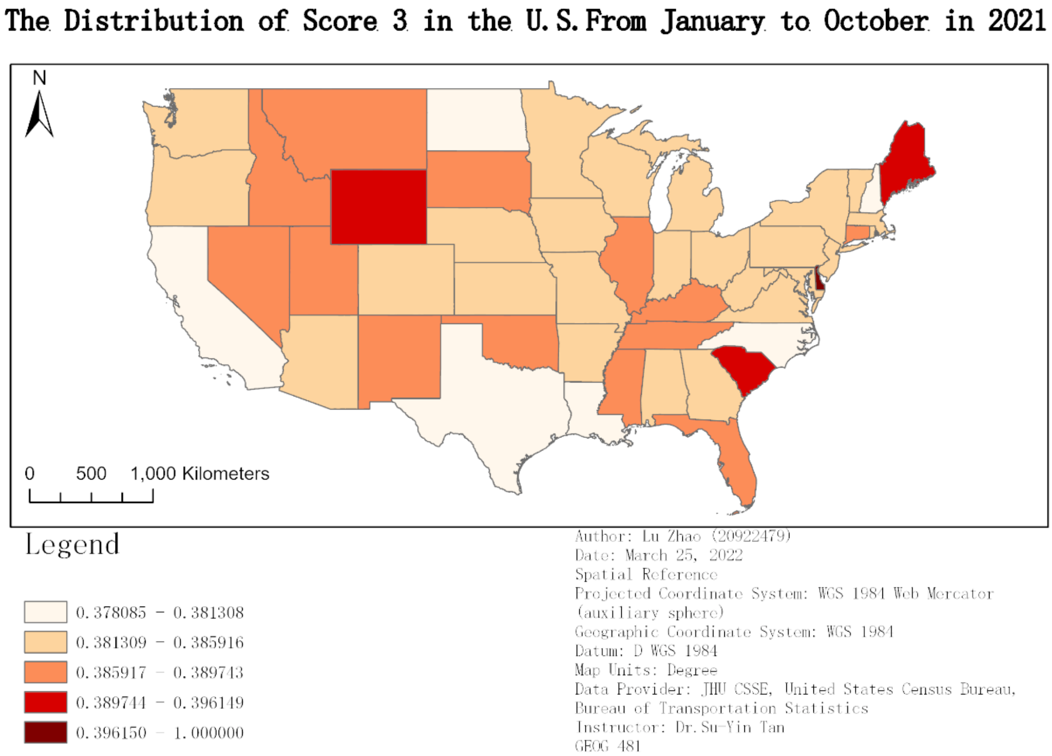


Figure 6. The distribution of Score 3 in the U.S. from January to October 2021

3.1.4 Comparing the distribution of Score 2 and Increased Infection Rate

As we discussed before, a higher Score 2 indicates a lower increased overall infection rate. In the map of the distribution of Increased Infection Rate, a darker color means a higher Increased Infection Rate, while in the map of the distribution of Score 2, a darker color means a lower increased overall infection rate. Therefore, the two maps are revered. For instance, the color of South Dakota in the map of Score 2 is red and it transforms to white in the map of Increased Infection Rate.

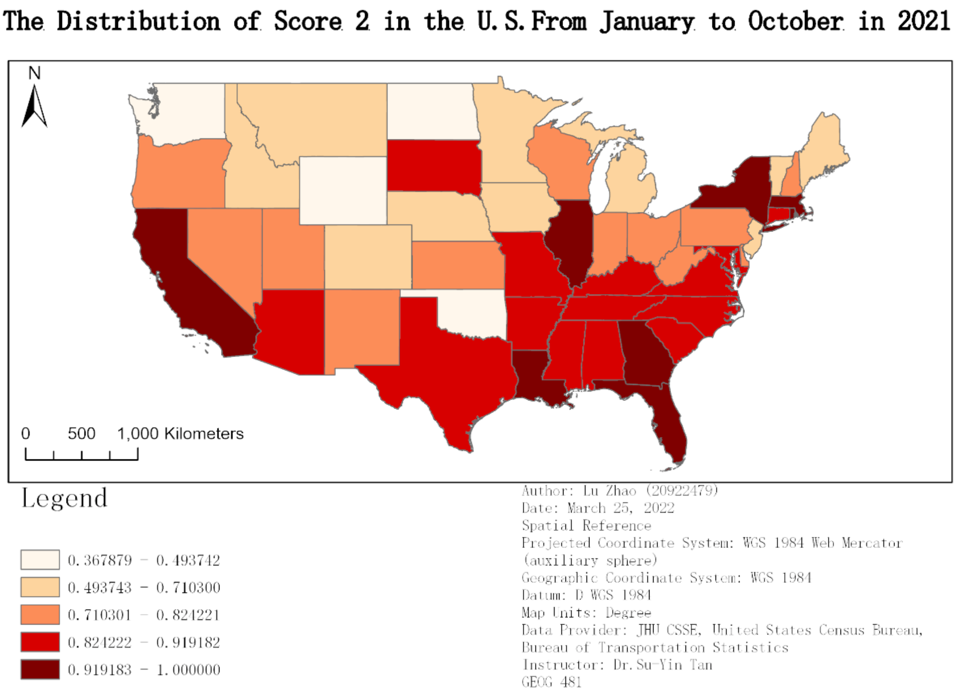


Figure 7. The distribution of Score 2 in the U.S. from January to October 2021 (2)

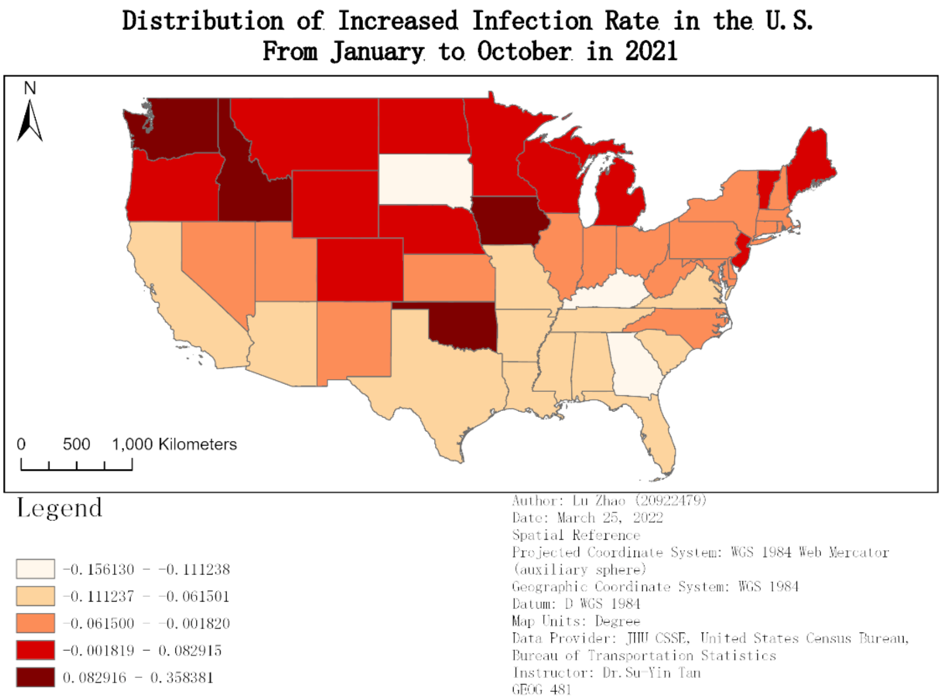


Figure 8. Distribution of increased rate of confirmed cases in the U.S. from January to October 2021

3.2 Spatial Analysis and Mapping

3.2.1 Spatial Analysis for Score 1, Score 2 and Score 3

The results of global Moran's I and Local Moran's I for Score 1 are shown in Figure 9. The global Moran's I index is 0.114, which is positive. The p-value of the global Moran's I is 0.037, which is lower than 0.05. Thus, score 1 of the 48 states in the US have a significant positive spatial autocorrelation, which indicates a clustering spatial pattern. According to the LISA map in Figure 9, the blue label states Montana, Idaho and Colorado show a low-low legend, which means the states are in a low-low cluster. That means, the local spatial autocorrelations of those states are significant and positive, but their score 1 values are sub-par. Maryland is a high-low outlier, which means the scores of the states around are much lower than its. Pennsylvania is a low-high outlier, which means the score 1 of the state is low, but the surrounding states have much higher score 1. The realistic interpretations of those phenomena and the results will be interpreted in the interpretation section.

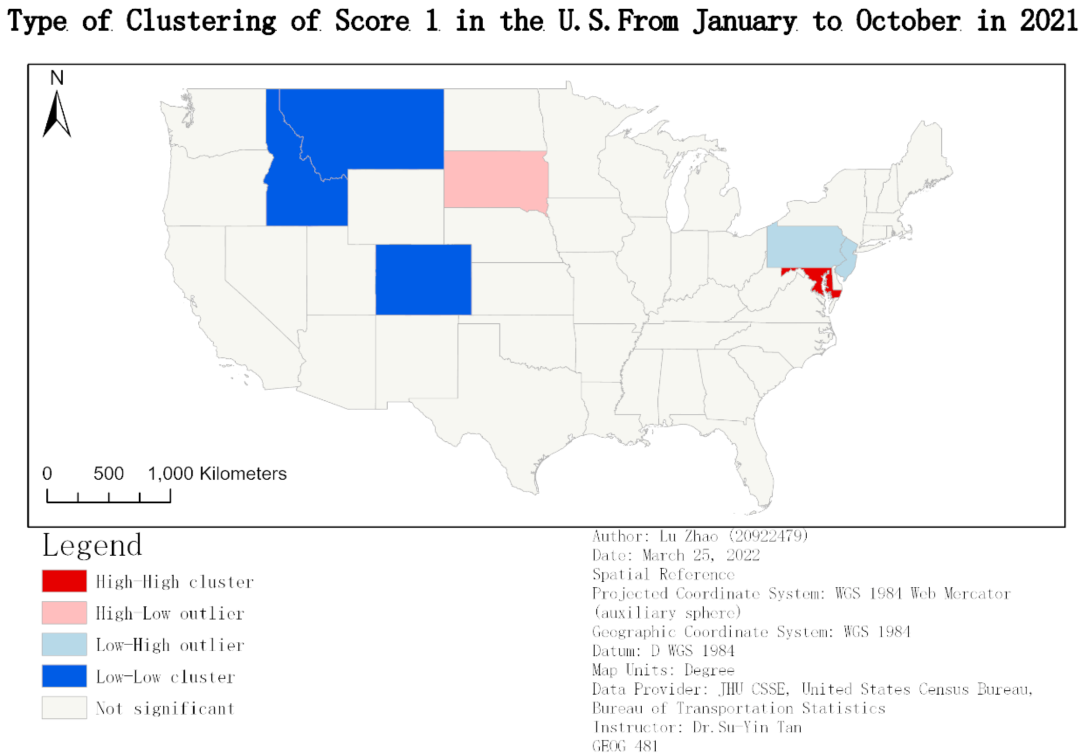


Figure 9. Type of clustering of Score 1 in the U.S. from January to October 2021

Moran's I:0.114, p-value:0.037

The results of global Moran's I and Local Moran's I for Score 2 are shown in Figure 10. The global Moran's I index is 0.039, which is positive. The p-value of the global Moran's I is 0.002, which is much lower than 0.05. Thus, score 2 of the 48 states in the US have a significant positive spatial autocorrelation, which indicates a clustering spatial pattern. According to the LISA map of score 2 in Figure 10, Idaho, Montana, and Colorado are in the low-low cluster, which means score 2 has an extremely significant positive spatial autocorrelation locally around the states. And the states themselves have a sub-par score 2. Tennessee and Alabama are two large states in the high-high cluster, which means score 2 has a strong local spatial autocorrelation around those states. And the states have relatively high score 2. Pennsylvania is a low-high outlier, which means the score 1 of the state is low, but the surrounding states have much higher score 2. The realistic interpretations of those phenomena and the results will be interpreted in the interpretation section.

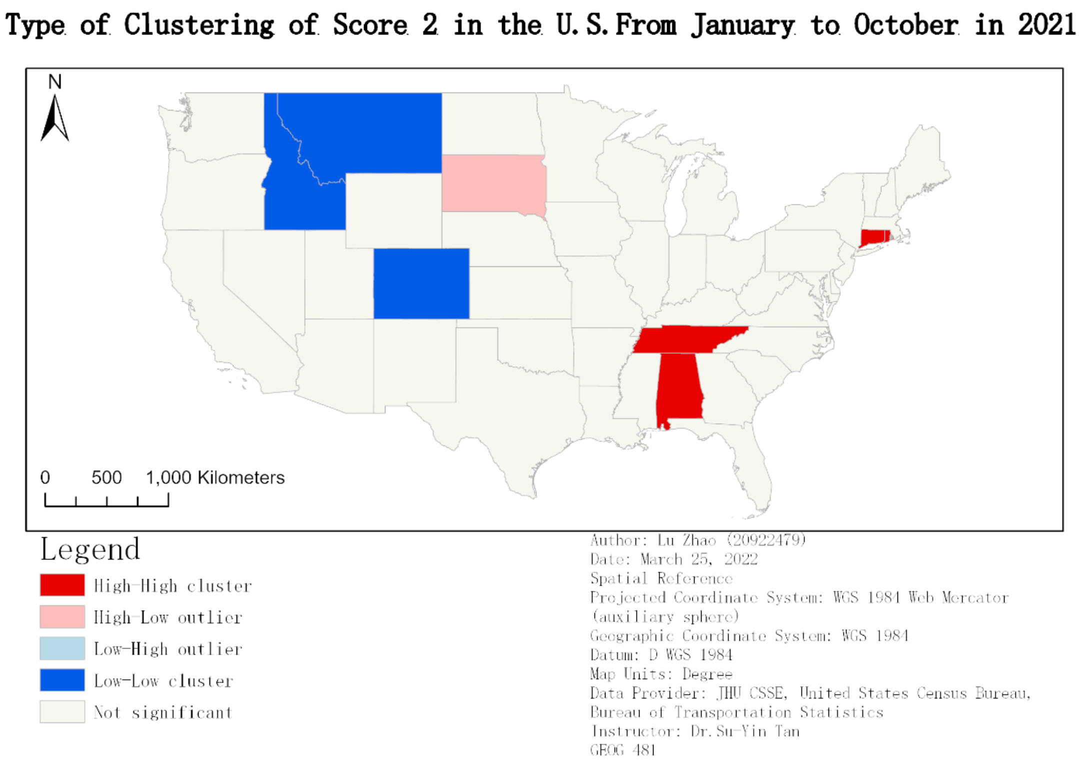


Figure 10. Type of clustering of Score 2 in the U.S. from January to October 2021

Moran's I: 0.309, p-value: 0.002

The results of global Moran's I and Local Moran's I for Score 3 are shown in Figure 11. The global Moran's I index is -0.018, which is negative. The p-value of the global Moran's I is 0.334, which is much bigger than 0.05. Thus, score 3 of the 48 states in the US does not have a significant global spatial autocorrelation, which indicates a random spatial pattern. However, according to the LISA map of score 3 in Figure 11, New Jersey is a low-high outlier. That means, the score 3 of the states around is very high, butt the score 2 of New Jersey itself is low. Still, the realistic interpretations of those phenomena and the results will be interpreted in the interpretation section.

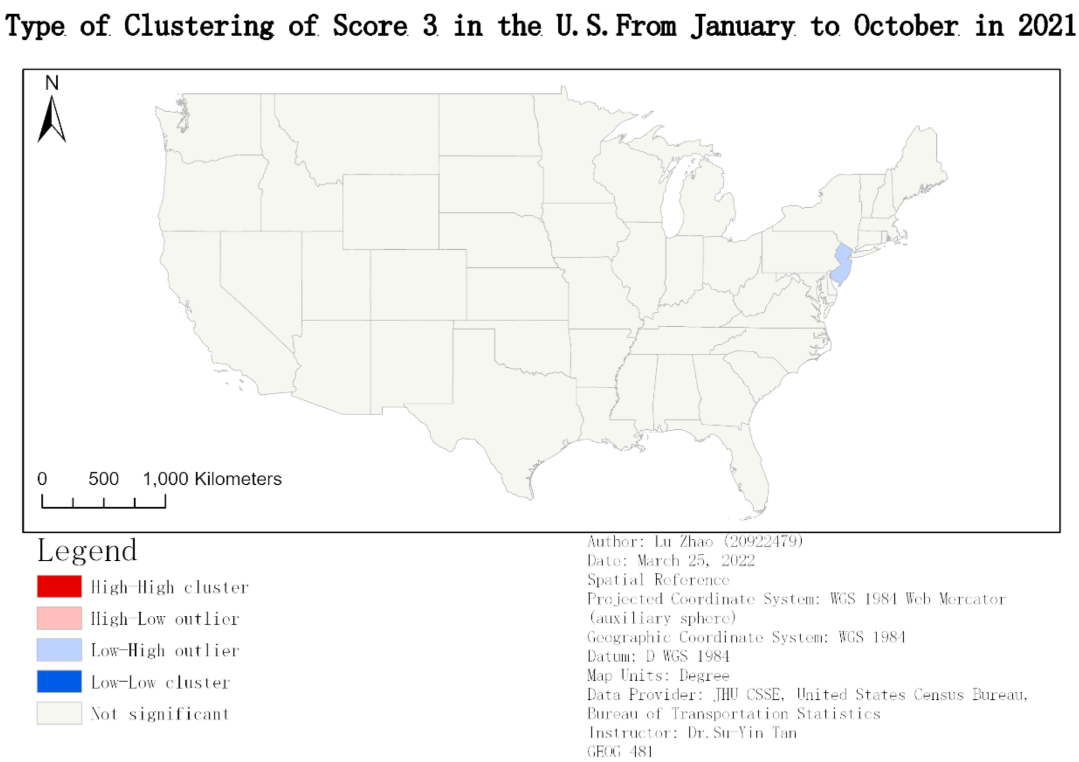


Figure 11. Type of clustering of Score 3 in the U.S. from January to October 2021

Moran's I: -0.018, p-value: 0.334

3.2.2 Spatial Analysis for increased rate of confirmed cases by modified Moran's I

The result of Moran's I and Local Moran's I for increased rate of confirmed cases are shown below. The global Moran's I index is 0.13, which is positive. The p-value of the global Moran's I is 0.009, which is much smaller than 0.05. Therefore, IRCC has a significant positive spatial autocorrelation under the self-defied spatial weight matrix using PE between states. And the spatial pattern is clustering. In the southeastern part of the map a large part of the blue aggregation area, represented by the states of Georgia, Florida, and Alabama, represents the low-low spatial cluster. It means that the positive spatial autocorrelation of IRCC is significant in Georgia, Florida, and Alabama. The IRCC values are low in these states and the states with strong spatial relationships with them. For Colorado and Kansas in the high-high cluster, the positive spatial autocorrelation of IRCC is significant in the states. The IRCC values are low in the states and the states with strong spatial relationships with them. Alabama is a low-high cluster, which means the IRCC value is low in Alabama, but the states with strong spatial relationships with Alabama has a high IRCC value. Since Georgia is one of the states in the low-low cluster, which is expected to have a low IIR. And the states with strong spatial relationships with Georgia are expected to have low IRCC values as well. In the next subsection, the expectation will be verified, so that some intuition of the meaning of the modified Moran's I analysis can be given.

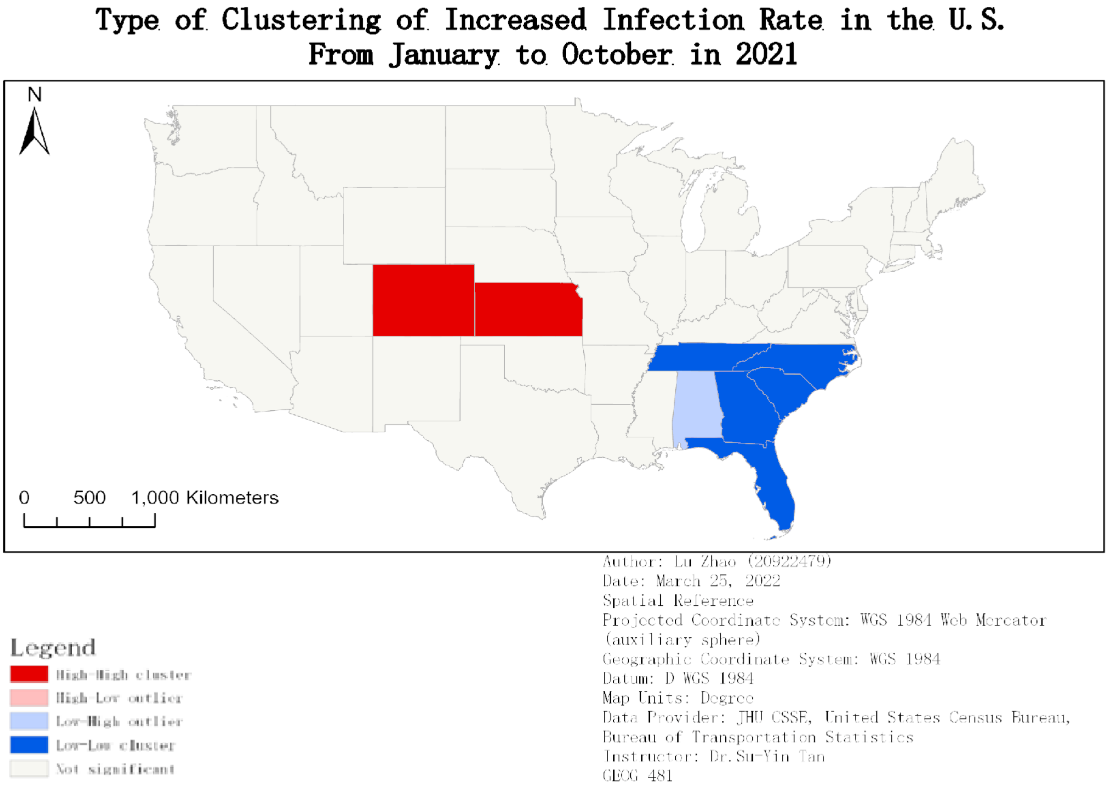


Figure 12. Type of clustering of increased rate of confirmed cases in the U.S. from January to October 2021

Moran's I: 0.13, p-value: 0.009

3.2.3 Mapping for Input Passengers by Flight in Georgia

Based on the modified local and global Moran's I analysis for IRCC using self-defined spatial weight matrix derived from PE between states, we found a significant positive spatial autocorrelation of IIR. To verify the claim made in the previous subsection that Georgia along with the states with high PE with Georgia have low IIR, we select the top 15 states with the highest PE with Georgia and output a map of input passengers flows by flight from the states to Georgia (i.e., PE between the states and Georgia) overlaid on the IRCC map.

According to Figure 13, Florida, Texas, and California rank in the top three in terms of total ridership input by flight to Georgia. The legends for all three states are in green, while the legend for Georgia is in yellow, indicating that their IRCC are similarly low. The top 15 states in terms of total passenger inputs are not from states with much higher IRCC than Georgia (i.e., states colored in dark blue in the map). In general, the darker the arrow that represents the volume of input passengers into Georgia by flight is, the more similar the colors that represent the IRCC of Georgia and the original state will be, which validates the expectation that significant positive spatial autocorrelation of IRCC from the modified Moran's I is indeed equivalent to saying that the states with higher PE between each other tend to have similar IRCC values.

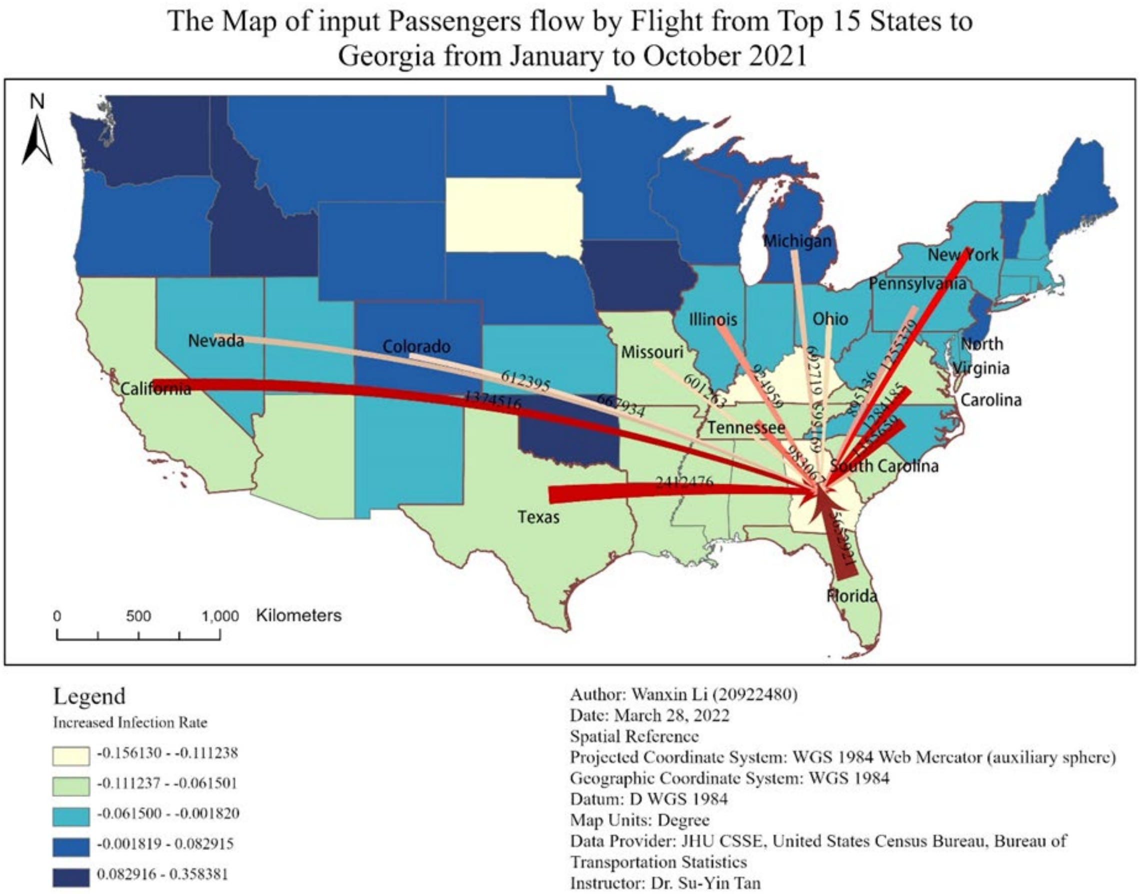


Figure 13. The map of input passengers flow by flight from Top 15 states to Georgia from January to October 2021

4. Discussion

4.1 Interpretation

According to the results, Delaware has a significantly higher Score 1 than other states, which means Delaware performs very well in balancing the recovery of flight service and the control of local pandemics during January to October in 2021. Also, northern region in the US has lower Score 1 than southern region, which means the north part of the US performs better in balancing the recovery of flight service and the control of local pandemics than the south part of the US during January to October in 2021. The result that Wyoming, North Dakota and Oklahoma, Washington have the lowest Score 1 indicates that Wyoming, North Dakota and Oklahoma, Washington performs terribly in either the recovery of flight service or the control of local pandemics during January to October in 2021 comparing to other states. From the score 2 figures, we found that the southern region has higher Score 2 than the northern region, which indicates that the south part of the US performs better than the north part of the US in controlling COVID-19 during January to October in 2021. Particularly, California, Florida, and Georgia in the south part of the US have both a higher Score1 and Score 2, which means in these states, they controlled COVID-19 well without sacrificing too much fight service capacity during January to October in 2021. For Wyoming, North Dakota, Oklahoma, and Washington that have the lowest Score 1, still have the lowest Score 2. That mean, Wyoming, North Dakota, Oklahoma, and Washington are still facing extreme difficulties in controlling

COVID-19 during January to October in 2021, which was not the correct time for recovering flight service. From the results of Score 3, Wyoming, Maine, and South Carolina have significantly higher scores, which means those states have already recovered a decent amount of flight service. On the other hand, the low Score 3 for California, Texas, Louisiana, North Carolina, and North Dakota indicates that the states were still not ready for recovering too much flight service capacity during January to October in 2021.

From the results of Moran's I for Score 1, Score 1 of the 48 states in the US have a significant positive spatial autocorrelation, which indicates a clustering spatial pattern. Thus, for the 48 states in the US, neighboring states tend to have similar performance of balancing flights and COVID-19. Among the states, Montana, Idaho, and Colorado are in a low-low cluster, which means Montana, Idaho and Colorado, and their neighboring states all perform terribly in balancing flights and COVID-19. Since Maryland is a high-low outlier, which means the scores of the states around are much lower than its.

From the results of the Moran's I and Local Moran's I for Score 2, the Score 2 of the 48 states in the US have a significant positive spatial autocorrelation, which indicates a clustering spatial pattern. Hence, for the states, neighboring states tend to have similar level of performance in controlling COVID-19. Among the states, Idaho, Montana, and Colorado are in the low-low cluster, which means the Idaho, Montana, Colorado, and the neighboring states all have a similarly low level of performance in controlling COVID-19. For Tennessee and Alabama in the high-high cluster, the states and the neighboring states all did a good job in controlling pandemics. Pennsylvania is a low-high outlier, which means Pennsylvania has a significantly worse performance in controlling pandemics in that area.

From the results of the Moran's I and Local Moran's I for Score 3, the score 3 of the 48 states in the US does not have a significant global spatial autocorrelation, which indicates a random spatial pattern. Thus, the performances of the states in recovering flights are likely to be independent with the neighboring states. However, since New Jersey is a low-high outlier, the performance of New Jersey is much below the average level in that area.

From the results of the modified the Moran's I analysis using self-defined spatial weight matrix derived from PE between states, IRCC has a significant positive spatial autocorrelation. Hence, for the 48 states in the US, the pair of states with higher PE are likely to have similar IRCC. Specifically, since the southeastern part of the map a large part of the blue aggregation area, represented by the states of Georgia, Florida, and Alabama, represents the low-low spatial cluster, the area has much lower IRCC than the average IRCC in the US. Also, the states with higher PE with the states in this area tend to have similarly low IIR. For Colorado and Kansas in the high-high cluster, Colorado and Kansas have high IRCC values than the average IRCC in the US. And the states with higher PE with the states in this area tend to have similarly high IIR. Alabama is a low-high cluster, which means the IRCC value is low in Alabama, but the states that have a large PE with Alabama tend to have high IRCC.

From the input population flow map of Georgia, states with higher volume of input passengers into Georgia tend to have similar IRCC values with Georgia. Hence, the interpretation of our modified Moran's I with a self-defined spatial weight matrix is valid.

4.2 Limitations of Data Sources, and Calculation of IR & OPE

The T-100 Domestic Market datasets we used have a limitation on the track of individual data of flight passengers, which may lead to potential extra counts of transition flight passengers. For instance, the overall population exchanged by flight between state A and state B would be counted twice if state C is the true destination for the passenger as state B is a transition flight point.

The IR of each state during January and October in 2020 and 2021 in the US estimated based on the JHU dataset is not necessarily identical to the true value of the IR. More discussion has been included in the methodology section, along with the rationale behind the estimate method. We have also calculated the radius of 95% confidence interval of the estimated IR for each state in 2021, which has been attached to the Appendix. Most of the radius of 95% confidence interval for the estimated IR for each state in 2021 are small. Even for the state whose estimated IR has the largest 95%, Iowa, the radius is still just 0.1297 %, which is very small comparing to the estimated IR value. Hence, if other assumptions of Bernoulli trial are satisfied in the COVID test in each state, the estimated infection rate tends to have a high enough accuracy. However, randomness of sample selection might still be violated. For example, some states may set more test centers in less populous areas, or near the refugee habitats, which is regarded as a bias in selection. Unfortunately, until governments publish where the test cases were located, the randomness of selection can never be certain.

5. Conclusion

Since the outbreak of COVID-19, there has been a huge negative impact on the world, including the United States, which is a superpower with the worlds top economy and medical security. The impact of COVID-19 on the daily lives of hundreds of millions of local people, but also on health and other aspects of life. Nowadays, when more and more human beings are gradually pursuing the coexistence of Coronavirus, there is a great significance to balance the spread of the pandemic and air flights and to explore the suitable approach to quantify this balance.

5.1 Summary of Project Findings

According to the results of the spatial autocorrelation analysis of increased rate of confirmed cases was carried out, we conclude that there is a strong positive correlation between increased rate of confirmed cases and overall population exchange. Beyond that, from the map of input passenger flow by flights and the distribution of increased rate of confirmed cases in the United States from January to October in 2021, it illustrates that Georgia is a good example.

Besides that, the index ‘Score’ is feasible to quantify and compare the trade-off between flights and the COVID-19 spread between different states in the U.S. based on the target period. We conclude that Delaware, the eastern United States, had a better trade-off between flights and COVID-19 during the research period, which has the highest Score 1. Meanwhile, California, Louisiana, and Florida, three states located in the southeast part of the U.S., had a better result of controlling the spread of COVID-19, which have the highest Score 2.

5.2 Future Directions of Presented Work

The future research will focus on improving the method of calculating the COVID-19 infection rate based on the existing dataset with limited specific information. Meanwhile, testing the rationality of the ‘Score’ index in a more rigorous way is the second future direction. Finally, it is planned to apply the ‘Score’ index in more cases to analyze the level of balance between flight passengers and the sprawl of the COVID-19 in other different types of countries and regions, such as China, the United Kingdom, Australia, and Kenya. It is also a method to test the application scope of the 'Score'.

5.3 Recommendation

We suggest that the spread of the COVID-19 should be controlled and a good balance between flights and the pandemic should be maintained to ensure people's normal travel as much as possible, especially the flights originating or transiting from the states which with low ‘Score’s.

References

Aldstadt, J. (2010). Spatial clustering. In *Handbook of applied spatial analysis* (pp. 279-300). Springer, Berlin, Heidelberg.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis, 27*(2), 93-115.

Beecham, R., Dykes, J., Hama, L., & Lomax, N. (2021). On the use of "glyphmaps" for analysing the scale and temporal spread of COVID-19 reported cases. ISPRS International Journal of Geo-Information, 10(4), 213–. https://doi.org/10.3390/ijgi10040213

Bourdin, S., Jeanne, L., Nadou, F., & Noiret, G. (2021). Does lockdown work? A spatial analysis of the spread and concentration of Covid-19 in Italy. Regional Studies, 55(7), 1182–1193. https://doi.org/10.1080/00343404.2021.1887471

Campbell, D. E., & Kelly, J. S. (1994). Trade-off theory. *The American Economic Review, 84*(2), 422-426.

Chen, Wu, D., Chen, H., Yan, W., Yang, D., Chen, G., Ma, K., Xu, D., Yu, H., Wang, H., Wang, T., Guo, W., Chen, J., Ding, C., Zhang, X., Huang, J., Han, M., Li, S., Luo, X., … Ning, Q. (2020). Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ (Online), 368. https://doi.org/10.1136/bmj.m1091

Chen, Y. & Schumann, G. (2013). New Approaches for Calculating Moran's Index of Spatial Autocorrelation. PloS One, 8(7), e68336–e68336._ https://doi.org/10.1371/journal.pone.0068336

Chen, Z., Zhang, Q., Lu, Y., Guo, Z., Zhang, X., & Zhang, W. et al. (2020). Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China. Chinese Medical Journal, 133(9), 1044-1050._ https://doi.org/10.1097/cm9.0000000000000782

Chen. (2013). New approaches for calculating Moran's index of spatial autocorrelation. *PloS One, 8*(7), e68336–e68336. https://doi.org/10.1371/journal.pone.0068336.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *Geographical Analysis, 46*(1), 75-97.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore Y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Elizabeth Halloran, M., Longini, I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science (American Association for the Advancement of Science), 368(6489), 395–400. https://doi.org/10.1126/science.aba9757

Choo, E. U., Schoner, B., & Wedley, W. C. (1999). Interpretation of criteria weights in multicriteria decision making. *Computers & Industrial Engineering, 37*(3), 527-541.

Cordes, J. & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spatial and Spatio-Temporal Epidemiology, 34, 100355–. https://doi.org/10.1016/j.sste.2020.100355

Crandall, M. S., & Weber, B. A. (2004). Local social and economic conditions, spatial concentrations of poverty, and poverty dynamics. *American Journal of Agricultural Economics, 86*(5), 1276-1281.

Debata, B., Patnaik, P., & Mishra, A. (2020). COVID-19 pandemic! It's impact on people, economy, and environment. Journal of Public Affairs, 20(4). https://doi.org/10.1002/pa.2372

Desmet, K. & Wacziarg, R. (2022). JUE Insight: Understanding spatial variation in COVID-19 across the United States. Journal of Urban Economics, 127, 103332–103332. https://doi.org/10.1016/j.jue.2021.103332

Duggal, H. & Haddad, M. (December 9, 2021). Visualising the Global Air Travel Industry. ALJZEERA. Retrieved from https://www.aljazeera.com/economy/2021/12/9/visualising-the-global-air-travel-industry-interactive

Ezeafulukwe, U. A., Darus, M., & Fadipe-Joseph, O. (2018). On analytic properties of a sigmoid function. *Int. Journal of Mathematics and Computer Science, 13*(2), 171-178.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography, 30*(5), 609-628.

Fang, H., Xin, S., Pang, H., Xu, F., Gui, Y., Sun, Y. & Yang, N. (2021). Evaluating the effectiveness and efficiency of risk communication for maps depicting the hazard of COVID-19. Transactions in GIS. https://doi.org/10.1111/tgis.12814

Flores-Chamba, J., López-Sánchez, M., Ponce, P., Guerrero-Riofrío, P., & Álvarez-García, J. (2019). Economic and spatial determinants of energy consumption in the European Union. *Energies, 12*(21), 4118.

Franch-Pardo, I., Napoletano, B., Rosete-Verges, F. & Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. The Science of the Total Environment, 739, 140033–140033. https://doi.org/10.1016/j.scitotenv.2020.140033

Gabrielli, L., Deutschmann, E., Natale, F., Recchi, E. & Vespe, M. (2019). Dissecting global air traffic data to discern different types and trends of transnational human mobility. EPJ Data Science, 8(1), 1–24. https://doi.org/10.1140/epjds/s13688-019-0204-x

Gao, S., Rao, J., Kang, Y., Liang, Y., & Kruse, J. (2020). Mapping county-level mobility pattern changes in the United States in response to COVID-19. SIGSpatial Special, 12(1), 16-26 https://doi.org/10.1145/3404820.3404824

Getis, A. (2009). Spatial weights matrices. *Geographical Analysis, 41*(4), 404-410.

Gheorghe, C. & Sebea, M. (2010). "The Economic and Social Benefits of Air Transport," Ovidius University Annals, Economic Sciences Series, Ovidius University of Constantza, Faculty of Economic Sciences, vol. 0(1), pages 60-66, May.

Gouissem, Abualsaud, K., Yaacoub, E., Khattab, T., & Guizani, M. (2021). A Novel Pandemic Tracking Map: From Theory to Implementation. IEEE Access, 9, 51106–51120. https://doi.org/10.1109/ACCESS.2021.3067824

Gross, B. Zheng, Z., Liu, S., Chen, X., Sela, A., Li, J., Li, D. & Havlin, S. (2020). Spatio-temporal propagation of COVID-19 pandemics. Europhysics Letters, 131(5). https://doi.org/10.1209/0295-5075/131/58003

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., & He, J. et al. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal Of Medicine, 382(18), 1708 1720. https://doi.org/10.1056/nejmoa2002032

Hu, Z., Cui, Q., Han, J., Wang, X., Sha, W. & Teng, Z. (2020). Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in Guangdong province, China. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases, 95, 231–240._ https://doi.org/10.1016/j.ijid.2020.04.010

Huang, H., Wang, Y., Wang, Z., Liang, Z., Qu, S., & Ma, S. et al. (2020). Epidemic Features and Control of 2019 Novel Coronavirus Pneumonia in Wenzhou, China. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3550007

Huang, X., Li, Z., Jiang, Y., Li, X. & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. PloS One, 15(11), e0241957–. https://doi.org/10.1371/journal.pone.0241957

Juergens, C. (2020). Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps. KN - Journal of Cartography and Geographic Information, 70(4), 155–161. https://doi.org/10.1007/s42489-020-00057-w

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J. & Kruse, J. (2020). Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Scientific Data, 7(1), 390–390. https://doi.org/10.1038/s41597-020-00734-5

Kent, A. (2020). Mapping and Counter-Mapping COVID-19: From Crisis to Cartocracy. Cartographic Journal, 57(3), 187–195._ https://doi.org/10.1080/00087041.2020.1855001

Kephart, J. L., Delclòs-Alió, X., Rodríguez, D. A., Sarmiento, O. L., Barrientos-Gutiérrez, T., Ramirez-Zea, M., Quistberg, D. A., Bilal, U., & Diez Roux, A. V. (2021). The effect of population mobility on COVID-19 incidence in 314 Latin American cities: a longitudinal ecological study with mobile phone location data. The Lancet. Digital health, 3(11), e716–e722. https://doi.org/10.1016/S2589-7500(21)00174-6

Kessel, P. van, Baronavski, C., Scheller, A., & Smith, A. (2022, January 18). How the COVID-19 pandemic has changed Americans' personal lives. Pew Research Center. Retrieved from https://www.pewresearch.org/2021/03/05/in-their-own-words-americans-describe-the-struggles-and-silver-linings-of-the-covid-19-pandemic/

Kommenda, N. (April 3, 2020). How Is the Coronavirus Affecting Global Air Traffic? The Guardian. Retrieved from https://www.theguardian.com/world/ng-interactive/2020/apr/03/how-is-the-coronavirus-affecting-global-air-traffic

Kraemer, M., Yang, C., Gutierrez, B., Wu, C., Klein, B., Pigott, D., du Plessis, L., Faria, N., Li, R., Hanage, W., Brownstein, J., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. & Scarpino, S. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. Science (American Association for the Advancement of Science), 368(6490), 493–497. https://doi.org/10.1126/science.abb4218

Kyurkchiev, N., Iliev, A., & Rahnev, A. (2019). *Some Families of Sigmoid Functions: Applications to Growth Theory*. Saarbrucken, Germany: Lap Lambert Academic Publishing.

Li, H., Calder, C. A., & Cressie, N. (2007). Beyond Morans I: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis, 39*(4), 357-375.

Li, R. (2021). Visualizing COVID-19 information for public: Designs, effectiveness, and preference of thematic maps. Human Behavior and Emerging Technologies, 3(1), 97–106. https://doi.org/10.1002/hbe2.248

Liu, M., Liu, M., Li, Z., Zhu, Y., Liu, Y., Wang, X., Tao, L. & Guo, X. (2021). The spatial clustering analysis of COVID-19 and its associated factors in mainland China at the prefecture level. The Science of the Total Environment, 777, 145992–. https://doi.org/10.1016/j.scitotenv.2021.145992

Liu, Q., Deng, M., Shi, Y. & Wang, J. (2012). A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity. Computers & Geosciences, 46, 296–309. https://doi.org/10.1016/j.cageo.2011.12.017

Liu, S. Qin, Y., Xie, Z. & Zhang, J. (2020). The Spatio-Temporal Characteristics and Influencing Factors of Covid-19 Spread in Shenzhen, China-An Analysis Based on 417 Cases. International Journal of Environmental Research and Public Health, 17(20), 7450–. https://doi.org/10.3390/ijerph17207450

Mahase, E. (2020). Covid-19: death rate is 0.66% and increases with age, study estimates. *BMJ: British Medical Journal (Online), 369.*

Mohammed, Asghar, M., & Kanwal, N. (2021). Analyzing the impact of COVID-19 on flight cancellation using machine learning and deep learning algorithms for a highly unbalanced dataset. 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), 1–6. https://doi.org/10.1109/ICECET52533.2021.9698693

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1/2), 17-23.

Murray, A. T., Grubesic, T. H., & Wei, R. (2014). Spatially significant cluster detection. *Spatial Statistics, 10*, 103-116.

Nizetic, S. (2020). Impact of coronavirus (COVID-19) pandemic on air transport mobility, energy, and environment: A case study. International Journal of Energy Research, 44(13), 10953–10961. https://doi.org/10.1002/er.5706

Norouzi. (2021). Post-COVID-19 and globalization of oil and natural gas trade: Challenges, opportunities, lessons, regulations, and strategies. *International Journal of Energy Research*, *45*(10), 14338–14356. https://doi.org/10.1002/er.6762

Miller, H. J. (2004). Toblers first law and spatial analysis. *Annals of the association of American geographers, 94*(2), 284-289.

Molloy, R., Smith, C. & Wozniak, A. (2011). Internal Migration in the United States. Journal of Economic Perspectives, 25 (3): 173-96. DOI: 10.1257/jep.25.3.173

Oztig, L. & Askin, O. (2020). Human mobility and coronavirus disease 2019 (COVID-19): a negative binomial regression analysis. Public Health (London), 185, 364–367. https://doi.org/10.1016/j.puhe.2020.07.002

Padula, W. & Davidson, P. (2020). Countries with High Registered Nurse (RN) Concentrations Observe Reduced Mortality Rates of Coronavirus Disease 2019 (COVID-19). SSRN, 3566190. https://doi.org/10.2139/ssrn.3566190

Patro, S., & Sahu, K. K. (2015). Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462.*

Pase, A., Presti, L., Rossetto, T. & Peterle, G. (2021). Pandemic cartographies: a conversation on mappings, imaginings and emotions. Mobilities, 16(1), 134–153. https://doi.org/10.1080/17450101.2020.1866319

Rinner, C. (2021). Mapping COVID-19 in Context: Promoting a Proportionate Perspective on the Pandemic. Cartographica, 56(1), 14–26._ https://doi.org/10.3138/cart-2020-0020

Smelser, N. J., & Baltes, P. B. (Eds.). (2001). *International encyclopedia of the social & behavioral sciences* (Vol. 11). Amsterdam: Elsevier.

Sun, Matthews, S. A., Yang, T.-C., & Hu, M.-H. (2020). A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? Annals of Epidemiology, 52, 54–59.e1._ https://doi.org/10.1016/j.annepidem.2020.07.014

Tan, L. F., & Seetharaman, S. (2020, April 2). Preventing the Spread of COVID -19 to Nursing Homes: Experience from a Singapore Geriatric Centre. Retrieved April 4, 2022, from https://covid-19.conacyt.mx/jspui/bitstream/1000/4710/1/1107144.pdf

"The economic & social benefits of air transport". (2004). Air Transport Action Group. Retrieved from_ https://www.icao.int/Meetings/wrdss2011/Documents/JointWorkshop2005

"Transnational Human Mobility Indicators Using Global Air Traffic Data." (August 5, 2021). *European Commission*. Retrieved from_ https://knowledge4policy.ec.europa.eu/dataset/ds00163_en

Travelmath. (n.d.). Retrieved April 5, 2022, from https://www.travelmath.com/drive-distance

U.S. Department of Transportation. (2011, December 22). Long Distance Transportation Patterns: Mode choice. Bureau of Transportation Statistics. Retrieved from_ https://www.bts.gov/archive/publications/america_on_the_go/long_distance_transportation_patterns/entire

Watts, A., Au, N., Thomas-Bachli, A., Forsyth, J., Mayah, O., Popescu, S. & Bogoch, I. (2020). Potential for inter-state spread of Covid-19 from Arizona, USA: analysis of mobile device location and commercial flight data. Journal of Travel Medicine, 27(8). https://doi.org/10.1093/jtm/taaa136

"What Does America Spend on Transportation and Infrastructure? Is Transportation Infrastructure Improving?." (n.d.). USAFacts. Retrieved from_ https://usafacts.org/state-of-the-union/transportation-infrastructure/

Xiong, Y., Wang, Y., Chen, F., & Zhu, M. (2020). Spatial Statistics and Influencing Factors of the COVID-19 Epidemic at Both Prefecture and County Levels in Hubei Province, China. International Journal Of Environmental Research And Public Health, 17(11), 3903. https://doi.org/10.3390/ijerph17113903

Yang, H., Chen, D., Jiang, Q., & Yuan, Z. (2020). High intensities of population movement were associated with high incidence of COVID-19 during the pandemic. Epidemiology and infection, 148, e177. https://doi.org/10.1017/S0950268820001703

Zhang, & Zhang, J. (2021). Long-term pathways to deep decarbonization of the transport sector in the post-COVID world. Transport Policy, 110, 28–36. https://doi.org/10.1016/j.tranpol.2021.05.018

Distance between United States cities. Distance Between Cities. (n.d.). Retrieved April 5, 2022, from https://www.distancefromto.net/distance-between-united-states-cities

ICAO. (2022). Economic Impacts of COVID-19 on Civil Aviation. Retrieved April 5, 2022, from https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx

Appendix I

Theorem 1: Assuming the IR of state i between January and October in year 2021 is σ_i^{2021}, the IR of state i between January and October in year 2020 is σ_i^{2020}, and the population of state i is P_i, which remains the same. Then, the increase rate of infected cases in state i from 2020 to 2021 is $IRCC_i = \frac{\sigma_i^{2021} - \sigma_i^{2020}}{\sigma_i^{2021}}$.

Proof:

According to the definition above, the number of infected people in 2021 during January and October is $\sigma_i^{2021} \cdot P_i$.

Similarly, the number of infected people in 2020 during January and October is $\sigma_i^{2020} \cdot P_i$.

Therefore, the increase rate of infected cases in state i from 2020 to 2021 is

$$\frac{\sigma_i^{2021} \cdot P_i - \sigma_i^{2020} \cdot P_i}{\sigma_i^{2021} \cdot P_i} = \frac{(\sigma_i^{2021} - \sigma_i^{2020}) \cdot P_i}{\sigma_i^{2021} \cdot P_i} = \frac{\sigma_i^{2021} - \sigma_i^{2020}}{\sigma_i^{2021}}$$

Definition 1: Normalization function $g: \{IRCC_i | i \in S\} \cup \{RRFP_i | i \in S\} \to R$ in this paper is defined by:

$$g(x) = \begin{cases} \dfrac{x - \min\{IRCC_i | i \in S\}}{\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}}, x \in \{IRCC_i | i \in S\} \\ \dfrac{x - \min\{RRFP_i | i \in S\}}{\max\{RRFP_i | i \in S\} - \min\{RRFP_i | i \in S\}}, y \in \{RRFP_i | i \in S\} \end{cases}$$

Theorem 2: For any value $x \in \{IRCC_i | i \in S\} \cup \{RRFP_i | i \in S\}$ normalized by g, $g(x)$ satisfies $g(x) \in [0,1]$.

Proof:

For any $x \in \{IRCC_i | i \in S\} \cup \{RRFP_i | i \in S\}$, x is in $\{IRCC_i | i \in S\}$ or x is in $\{RRFP_i | i \in S\}$.

If x is in $\{IRCC_i | i \in S\}$, then

$$\mathrm{x} \geq \min\{IRCC_i | i \in S\}$$

$$\mathrm{x} - \min\{IRCC_i | i \in S\} \geq 0$$

Also, since $\max\{IRCC_i | i \in S\} \geq \min\{IRCC_i | i \in S\}$, we have

$$\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\} \geq 0$$

Therefore,

$$g(x) = \frac{x - \min\{IRCC_i | i \in S\}}{\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}} \geq 0 \ (*)$$

Then, since $\mathrm{x} \leq \max\{IRCC_i | i \in S\}$, we have

$$x - \min\{IRCC_i | i \in S\} \leq \max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}$$

Also, since $\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\} \geq 0$, we have

$$\frac{x - \min\{IRCC_i | i \in S\}}{\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}} \leq \frac{\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}}{\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}}$$
$$= 1$$

$$\frac{x - \min\{IRCC_i | i \in S\}}{\max\{IRCC_i | i \in S\} - \min\{IRCC_i | i \in S\}} \leq 1 \ (**)$$

Therefore, by $(*)$ and $(**)$, we proved that $g(x) \in [0,1]$.

Similarly, proving that $g(x) \in [0,1]$ holds when x is in $\{RRFP_i | i \in S\}$ is trivial.

Corollary 1: Function $g: \{IRCC_i | i \in S\} \cup \{RRFP_i | i \in S\} \rightarrow [0,1]$ is increasing and monotone. Or equivalently, $g(x) < g(y)$ if and only if x < y.

Theorem 2: The rate by which the score increases with respect to an increment of x will be larger if α is larger; The rate by which the score decreases with respect to an increment of y will be larger if β is larger.

Proof:

Let $z = f(\alpha, \beta, x, y) = \frac{1}{1+e^{-\alpha \cdot x + \beta \cdot y}}$ where $x, y \in [0,1]$ denotes $Score_i(\boldsymbol{v}, RRFP_i, IRCC_i) = \frac{1}{1+e^{-\alpha \cdot g(RRFP_i) + \beta \cdot g(IRCC_i)}}$ for convenience.

Then, we have

$$\frac{\partial z}{\partial x} = \alpha \cdot \frac{e^{-\alpha \cdot x + \beta \cdot y}}{(1 + e^{-\alpha \cdot x + \beta \cdot y})^2} = \alpha \cdot z \cdot (1 - z) \ (*)$$

$$\frac{\partial z}{\partial y} = -\beta \cdot \frac{e^{-\alpha \cdot x + \beta \cdot y}}{(1 + e^{-\alpha \cdot x + \beta \cdot y})^2} = -\beta \cdot z \cdot (1 - z) \ (**)$$

From $(*)$, we know that the score increases by $\alpha \cdot z \cdot (1 - z)$ with an increment of x.

From $(*)$, we know that the score decreases by $\beta \cdot z \cdot (1 - z)$ with an increment of x.

Also, since $z = f(\alpha, \beta, x, y)$ has already been proved to be within $(0,1)$, we know that $z \cdot (1 - z) > 0$.

Therefore, the rate by which the score increases with respect to an increment of x will be larger if α is larger; The rate by which the score decreases with respect to an increment of y will be larger if β is larger.

Appendix II

States	Radius
Alabama	0.000449
Alaska	0.000319
Arizona	0.000179
Arkansas	0.000482
California	4.47E-05
Colorado	0.000161
Connecticut	0.000146
Delaware	0.000401
Florida	0.000118
Georgia	0.00026
Hawaii	0.000284
Idaho	0.001345
Illinois	7.64E-05
Indiana	0.000157
Iowa	0.001297
Kansas	0.000952
Kentucky	0.000309
Louisiana	0.000225
Maine	0.000313

Maryland	0.000124
Massachusetts	6.75E-05
Michigan	0.000175
Minnesota	0.000176
Mississippi	0.000783
Missouri	0.000318
Montana	0.000467
Nebraska	0.000308
Nevada	0.000356
New Hampshire	0.000361
New Jersey	0.00019
New Mexico	0.000242
New York	4.89E-05
North Carolina	0.000186
North Dakota	0.000758
Ohio	0.000184
Oklahoma	0.000555
Oregon	0.000197
Pennsylvania	0.000166
Rhode Island	0.000161
South Carolina	0.000229

South Dakota	0.000709
Tennessee	0.000319
Texas	0.000138
Utah	0.000436
Vermont	0.000223
Virginia	0.000253
Washington	0.000217
West Virginia	0.000296
Wisconsin	0.000188
Wyoming	0.000808

Radius of 95% confidence interval of IR for each state in 2021