

Article

Spatiotemporal Characteristics and Risk Factors of the COVID-19 Pandemic in New York State: Implication of Future Policies

Anran Zheng [†], Tao Wang ^{*} and Xiaojuan Li

College of Geospatial Information Science and Technology, Capital Normal University, North Road 105, Haidian District, Beijing 100048, China; 1173603035@cnu.edu.cn or anranz@design.upenn.edu (A.Z.); lixiaojuan@cnu.edu.cn (X.L.)

^{*} Correspondence: wangt@cnu.edu.cn; Tel.: +86-10-68903472

[†] Current address: Stuart Weitzman School of Design, University of Pennsylvania, 210 South 34th Street, Philadelphia, PA 19104, USA.

Abstract: The Coronavirus disease 2019 (COVID-19) has been spreading in New York State since March 2020, posing health and socioeconomic threats to many areas. Statistics of daily confirmed cases and deaths in New York State have been growing and declining amid changing policies and environmental factors. Based on the county-level COVID-19 cases and environmental factors in the state from March to December 2020, this study investigates spatiotemporal clustering patterns using spatial autocorrelation and space-time scan analysis. Environmental factors influencing the COVID-19 spread were analyzed based on the Geodetector model. Infection clusters first appeared in southern New York State and then moved to the central western parts as the epidemic developed. The statistical results of space-time scan analysis are consistent with those of spatial autocorrelation analysis. The analysis results of Geodetector showed that both temperature and population density were strong indications of the monthly incidence of COVID-19, especially in March and April 2020. There is a trend of increasing interactions between various risk factors. This study explores the spatiotemporal pattern of COVID-19 in New York State over ten months and explains the relationship between the disease transmission and influencing factors.

Keywords: COVID-19; Geodetector; New York State; spatial autocorrelation; space-time scan statistics



Citation: Zheng, A.; Wang, T.; Li, X. Spatiotemporal Characteristics and Risk Factors of the COVID-19 Pandemic in New York State: Implication of Future Policies. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 627. <https://doi.org/10.3390/ijgi10090627>

Academic Editor: Wolfgang Kainz

Received: 24 July 2021

Accepted: 16 September 2021

Published: 18 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes Coronavirus disease 2019 (COVID-19), has been spreading globally and seriously threatening public health worldwide [1]. On 30 January 2020, the World Health Organization declared COVID-19 a pandemic [2]. Since the outbreak of COVID-19 in the United States, New York State has been severely affected by COVID-19. As of 4 January 2021, there were 515,815 total confirmed cases and 25,868 mortality cases of COVID-19 in New York State [3]. In addition, COVID-19 has seriously affected the economy of the state. During the second quarter of 2020, the gross domestic product (GDP) of New York decreased by 36%, which was 5% lower than the national figure [4]. To improve public health and the economic development of the state as well as of the United States, how to control and prevent the spread of COVID-19 has become a matter of paramount importance for the New York State government and infectious disease prevention centers.

To identify the spread patterns and influencing factors of COVID-19, researchers have conducted spatial and temporal analyses of geographic clustering of COVID-19. Using the Bayesian space-time model and COVID-19 cases as of 30 January 2020, Chen et al. (2020) found strong correlations between the early incidence cases and emigration waves from Wuhan [5]. Kang et al. (2020) explored how COVID-19 spread to various types of neighborhoods in mainland China [6]. It was found that COVID-19 had significant spatial

autocorrelation in the early stage of regional transmission in Hubei, China [7]. Thakar (2020) conducted space and time analysis using visual analysis of clusters, kernel density estimation, and standard deviation ellipse with proxy data over Washington State, where the first community spread occurred in the United States during the first stage of the COVID-19 outbreak [8]. The research suggested that social distancing measures can help prevent the spread of COVID-19 on a large scale. However, it was not easy to acquire accurate information of infection cases in the early stage of COVID-19. Based on the emergency calls and ambulance dispatches of emergency medical services, Gianquintieri et al. (2020) applied a signal processing method and reconstructed the early spatiotemporal evolution of COVID-19 in the Lombardy region of Italy [9]. Desjardins et al. (2020) carried out the first study that utilized space-time scan analysis to monitor COVID-19 clustering in the United States, where New York City was identified as a high-risk area, and suggestions for the local health department were provided [10]. Hohl et al. (2020) investigated COVID-19 cluster patterns of the contiguous United States using time periodic surveillance with the prospective space-time scan based on daily data of the first half of 2020. A series of innovative figures and graphs along with a web application tracking the spatiotemporal distribution of significant clusters was developed to support decision making [11]. To further explore the role and functions of geospatial information, Müller et al. (2021) emphasized the requirements of a supranational approach of reliable and aligned statistical data in managing COVID-19 spread. The Nomenclature of Territorial Units for Statistics (NUTS) and the Infrastructure for Spatial Information in Europe (INSPIRE) together provide a framework for geo-information integration and interoperability in COVID-19 governing systems [12].

COVID-19 community transmission is influenced by various environmental factors, which have been investigated from various perspectives. Sun et al. (2020) found that spatial models, including ordinary least squares (OLS), spatial lag model (SLM), and spatial error model (SEM) models, can help partially explain the spatial disparities in COVID-19 period prevalence in the United States [13]. Liu et al. (2021) combined several models, including Moran's I, K-means clustering, SIR (susceptible-infected-removed), SLM, and OLS, to study the inter-county spatiotemporal interaction of COVID-19 transmission of 49 states in the United States. The result suggested the significant association of spatial heterogeneity with the socioeconomic factors of each state [14]. Mollalo et al. (2020) used geographically weighted regression (GWR) and multi-scale GWR (MGWR) models to examine the relationship between the COVID-19 incidence and environmental, demographic, and socioeconomic variables in the United States [15]. With data from 231 countries and regions, Meng et al. (2021) investigated annual and daily relationships between air pollution and early-stage COVID-19 incidence and observed significant spatially and temporally nonstationary variations between them, in which the United States showed significant linear patterns of O₃-confirmed cases [16]. Using Geodetector, Xie et al. (2020) analyzed influencing factors of COVID-19 in mainland China, which include population distribution, population inflow, traffic accessibility, economic connection, and average temperature. The interactions of factors were analyzed, and authors found out population inflow from Wuhan and economic connection intensity were the main factors affecting the COVID-19 spread rate and other factors had influence on the COVID-19 spread to different degrees [17].

Geospatial data and geographic information systems (GIS) can help researchers and public health experts analyze spatiotemporal trends and regional distribution and detect the influencing factors of COVID-19, thus providing support for governmental and public decision making [18]. Current studies have helped public health researchers better understand epidemic characteristics and transmission patterns of COVID-19 [13]. Environmental and socioeconomic factors increasing COVID-19 spread risks have been changing since the first occurrence of COVID-19.

To develop knowledge about COVID-19 dynamic clustering patterns in different seasons and understand relationships between influencing factors of the COVID-19 spread

over time, this study first collected a 10 month county-level COVID-19 dataset from 1 March to 31 December 2020 together with data of weather and population in New York State. Secondly, spatial autocorrelation analysis and space-time scan analysis were conducted to reveal spatiotemporal clustering patterns of COVID-19 in New York State. Finally, basic factors influencing the transmission of COVID-19 in New York State and their relationships are identified with the Geodetector model.

2. Materials and Methods

2.1. Study Area

The study area covers 62 counties of New York State, which is 141,000 km², accounting for 1% of land area of the United States [19]. New York State can be divided into seven geographic regions, including Albany–Schenectady, Buffalo–Cheektowaga, Elmira–Corning, Ithaca–Cortland, Rochester–Batavia–Seneca Falls, Syracuse–Auburn, and New York City [20]. Approximately 20.2 million inhabitants live in this area, accounting for 6.1% of the total U.S. population [19]. Population density varies greatly in different regions. Kings county has the highest population density of 2380.21/mi, while Hamilton has the lowest, at 4.13/mi [21].

New York State is one of most developed states in the United States and contributes 8% of the total GDP of the U.S [22]. New York State is also the leader in American manufacturing and international trade.

Figure 1 shows the COVID-19 epidemic development of New York State. From March to December 2020, the daily COVID-19 confirmed cases went through three stages: occurring and quickly developing (March–May 2020), stabilizing (June–September 2020), and quickly increasing (October–December 2020). Figure 1b exhibits seasonal variation concerning the COVID-19 incidence rate.

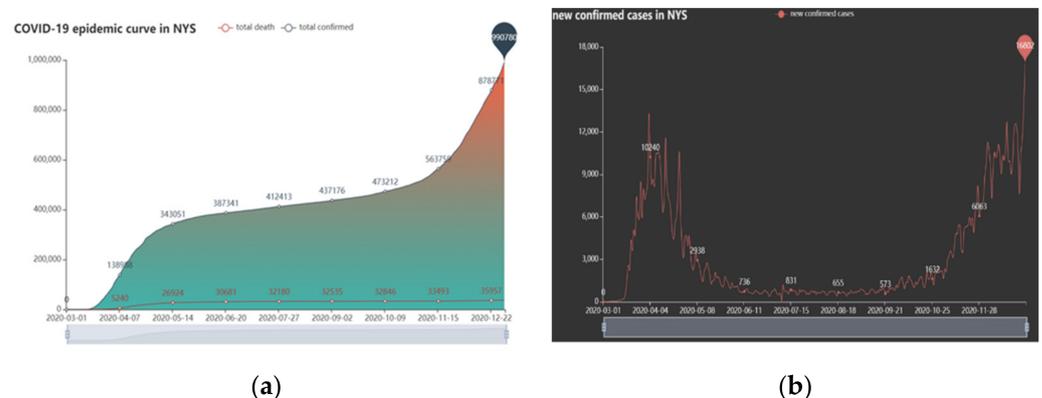


Figure 1. New York State epidemic development curve: (a) cumulative confirmed and deaths, (b) new confirmed COVID-19 cases.

2.2. Data

Daily COVID-19 data used in this work, including confirmed and death cases, was derived from data courtesy of USA Facts [23] which covers ten months, from 1 March to 31 December 2020. The boundaries of 62 counties of New York State were extracted from GIS.NY.GOV [24]. Six influencing factors were selected, and corresponding data were collected from various sources of spatiotemporal clustering analysis of COVID-19. Basic metadata of the six influencing factors are listed in Table 1.

Based on the monthly cumulative confirmed cases (CCC) and the population size (PS) from USA Facts [17], monthly incidence rate (IR, per 100 thousand people) can be derived:

$$IR = CCC/PS * 100,000 \quad (1)$$

The New York State Air Monitoring website [26] provides the PM 2.5 concentration data of 17 air stations in New York State. Inverse distance weighting (IDW) interpolation

was employed to produce a pollutant concentration surface of the entire region. To produce a reasonable interpolation along the border of New York State, this study collected the PM 2.5 concentration data of 13 air stations located in adjacent states, including New Jersey, Pennsylvania, Vermont, Massachusetts, and Connecticut, from US EPA [29].

By collecting the GPS data from mobiles, Unacast [27] grades social distancing behavior across the U.S. in different grades (A, B, C, D, and F). In this study, the grades of daily population mobility were converted to numerical scores (1, 2, 3, 4, and 5, respectively). Then, monthly average data was produced for Geodetector analysis.

Table 1. Influencing factors used in this study.

Category	Factors	Time Range	Source
Environmental	Temperature		National Centers for Environmental Information [25]
	Precipitation	March 2020–December 2020	National Centers for Environmental Information [25]
Demographic	PM 2.5 concentration		New York State Air Monitoring Website [26]
	Mobility		Unacast social distancing scoreboard [27]
Economic	Population density	2020	USA facts [23]
	Unemployment rate	March 2020–December 2020	FRED Economics Data [28]

2.3. Methods

In this section, the procedure to determine an optimal power parameter for interpolation was first implemented. Then, spatial autocorrelation analysis was conducted using global Moran's I to identify spatial dependence of COVID-19 cases in NYC. Getis–Ord G_i^* was used to reveal local correlations among neighboring counties. The prospective space-time scan analysis based on the Poisson model was used further on to investigate the emerging COVID-19 clusters at different stages. In order to explore the contribution of driving factors to the changing process of COVID-19 clusters, influencing factors were analyzed using Geodetector.

2.3.1. IDW Interpolation

Inverse distance weighting (IDW) was used to generate a pollutant's concentration surface based on PM 2.5 concentration data from air stations [14]. IDW interpolation determines cell values using a weighted average of surrounding points using the following equation:

$$Z_o = \frac{\sum_{i=1}^s z_i \frac{1}{d_i^k}}{\sum_{i=1}^s \frac{1}{d_i^k}} \quad (2)$$

where Z_o is the estimated value at point O; Z_i is the known attribute value of point i surrounding point O; d_i is the distance between point i and point O; S is the total number of known surrounding points; and k is the power parameter of distance. The value of the power parameter is how much a surrounding value contributes to the interpolated value based on the distance between the two. Root mean squared error was used to determine an optimal power value.

2.3.2. Spatial Autocorrelation Analysis

Both global and local Moran's I values were employed to investigate spatial autocorrelation of COVID-19 geographic distribution patterns in New York State. Global Moran's I statistics was applied to assess whether the cumulative number of confirmed COVID-19 cases in the entire New York State were spatially relevant [30]. It was calculated with the formula below.

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

where:

- $n = 62$ indicates 62 counties in New York;
- i and j indicate different counties;
- x_i and x_j , respectively, indicate the monthly incidence of COVID-19 in counties i and j ;
- \bar{x} indicates the monthly incidence of COVID-19 in counties in New York State;
- ω_{ij} represents the spatial weight matrix.

Global Moran's I ignores the instability of local spatial processes and cannot indicate the specific aggregation area [17]. Local spatial autocorrelation Getis-Ord G_i^* can reflect the distribution and spatial correlation characteristics of the local aggregation areas. Local spatial autocorrelation analysis was used to reflect the degree of correlation between a county and its neighboring counties in New York State, turning statistically significant clustering areas into "hot spot" and "cold spot" areas, which changed over time.

Spatial autocorrelation analysis was conducted with ArcGIS 10.5 Spatial Statistics.

2.3.3. Space-Time Scan Analysis

Aside from spatial dependence of the occurrence and development of COVID-19, space-time scan analysis was applied to detect the time, scope, and risk intensity of COVID-19 clustering in New York State.

The space-time scan analysis employs a moving cylindrical window that scans the research area to detect potential spatiotemporal clusters of COVID-19, with a radius changing continuously from zero to a specified maximum [30]. The radius of the cylinder's base represents the regional scope, and the height of the cylinder reflects time [10].

Under the Poisson model, the expected cases were calculated for each window using a proportional formula of $C * p/P$ where C is the total number of cases, P the total population and p the population inside the cylinder. According to the observed incidence cases, the logarithm of the likelihood ratio (LLR) was generated, which could indicate how the incidence cases within the scanning window differed from the cases outside the window. The calculation formula of LLR is as follows:

$$LLR = \frac{L_A}{L_0} = \frac{\frac{n_A}{u(A)}^{n_A} \left(\frac{n_G - n_A}{n_G - u(A)} \right)^{n_G - n_A}}{\left(\frac{n_G}{u_G} \right)^{n_G}} \quad (4)$$

where:

- L_0 and L_A are the likelihood function value under the hypothesis and the likelihood function value of window A ;
- n_A represents the actual number of cases in window A , n_G represents all cases, $u(A)$ is the expected cases in cylinder A , and $u_G = \sum u(A)$.

Then, the relative risk (RR) of each county was calculated using the formula below to measure the changes of the risk associated with that scanning window, and it was the ratio of observed cases to expected cases [31].

$$RR = \frac{\frac{c}{e}}{\frac{C-c}{C-e}} \quad (5)$$

where:

- c is the total confirmed cases of COVID-19 at the county level;
- e is the expected cases at the county level;
- C is the total confirmed cases of New York State.

This study used SaTScan 9.4 to perform space-time scan analysis. The upper spatial and temporal limits of the scanning radius were set to 30%, and the minimum duration was defined as 7 days. Based on the Poisson model, this study used Monte Carlo replications and considered a 0.05 level of significance.

2.3.4. Geodetector

To explore the main influencing factors of COVID-19 in the research period, this study analyzed the relationship between the monthly COVID-19 incidence rate and potential risk factors using Geodetector. Geodetector is designed and implemented to assess the association between the disease and relevant risk factors by analyzing their spatial differentiation [32]. To evaluate risk factors and reveal relationships among these factors, this study employed factor detection and interaction detection offered in Geodetector.

Factor detection is expressed by q -statistic:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST} \quad SSW = \sum_{h=1}^L N_h \sigma_h^2, \quad SST = N \sigma^2 \quad (6)$$

where q -statistic is to measure the spatial stratified heterogeneity, and it is also the explanatory power of factor X on the spatial heterogeneity of factor Y . The value of the q -statistic is between $[0, 1]$, and it increases as the stratified heterogeneity increases [33]. The study area is composed of N units and is stratified into $h = 1, \dots, L$ stratum. The h^{th} stratum is composed of N_h units; σ_h^2 and σ^2 are the variance of Y value for the h^{th} stratum and the whole study area. SSW and SST are the within sum of squares of a layer and the total sum of squares of New York State.

Interaction detection can identify the interaction relationship between two different factors. Following Equation (3), the q -statistics of X_1 and X_2 , $q(X_1)$ and $q(X_2)$, were calculated. Then, the q values of the interaction between X_1 and X_2 , $q(X_1 \cap X_2)$, were calculated. The interaction type between X_1 and X_2 was determined by comparing the values of $q(X_1)$, $q(X_2)$, and $q(X_1 \cap X_2)$.

In the Geodetector model, continuous values of risk factors had to be transformed into discrete values before their relationship with the COVID-19 incidence rate was analyzed. Discretization can be implemented using classification methods. There are four most frequently used classification methods including natural breaks, equal interval, geometrical interval, and quantile. None of these is versatile in all circumstances. To reveal the interaction variation of influencing factors in each specific month, q -statistics and p -values are used to determine an appropriate one for each month [34].

3. Results

3.1. IDW Interpolation Result

As indicated above, the power parameter of distance, k , influences the result of the interpolation process and the suggested value ranges from 0.5 to 3. The mostly suggested value is 2, which may not be the most suitable one to produce a useful interpolation. To determine a suitable value of k , this study calculated root mean squared error (RMSE) of IDW interpolation for comparison. According to Table 2, this study finally set the k value to 1 and then applied IDW to calculate the monthly average PM 2.5 concentration of each county.

3.2. Spatial Distribution Characteristics

Figure 2 shows the distribution of cumulative and average monthly incidence rate of COVID-19 in New York State. As seen on the maps, counties with the highest incidence rate are located around NYC, including Rockland, Richmond, Westchester, Suffolk, and Nassau. Counties with the lowest incidence rate are Clinton, Washington, Franklin, Essex, and Delaware, which are mainly located in the north of New York State. There is a strong regional heterogeneity of the COVID-19 incidence rate in New York State.

The counties with the highest incidence rate in the early stage of the epidemic (March–May) are concentrated in and around NYC, including Suffolk, Queens, Richmond, and Westchester. During the mid-epidemic period of this research (June–September), as shown in Figure S1, the incidence rate of COVID-19 in these areas decreased, and that in the central and northern areas increased over time. At the beginning of October, the overall incidence

rate of New York State increased. In November, the range of high incidence rate areas in New York State significantly expanded. In addition to NYC, the incidence rate in western Erie, Allegany, and Genesee also increased significantly. In December, the incidence rate of New York counties continued to rise, with decreasing spatial heterogeneity.

Table 2. RMSE of IDW interpolation with different k values.

k Value	1	1.5	2	2.5	3
March	0.85	0.88	0.92	0.96	1.09
April	0.98	1.02	1.07	1.12	1.17
May	1.24	1.28	1.33	1.37	1.42
June	1.47	1.48	1.51	1.55	1.59
July	1.34	1.37	1.42	1.48	1.54
August	1.21	1.27	1.33	1.4	1.46
September	1.1	1.16	1.22	1.28	1.33
October	1.1	1.15	1.2	1.25	1.3
November	0.78	0.82	0.96	0.96	0.96
December	0.86	0.86	0.88	0.91	0.94
Cumulative error	10.93	11.29	11.84	12.28	12.8

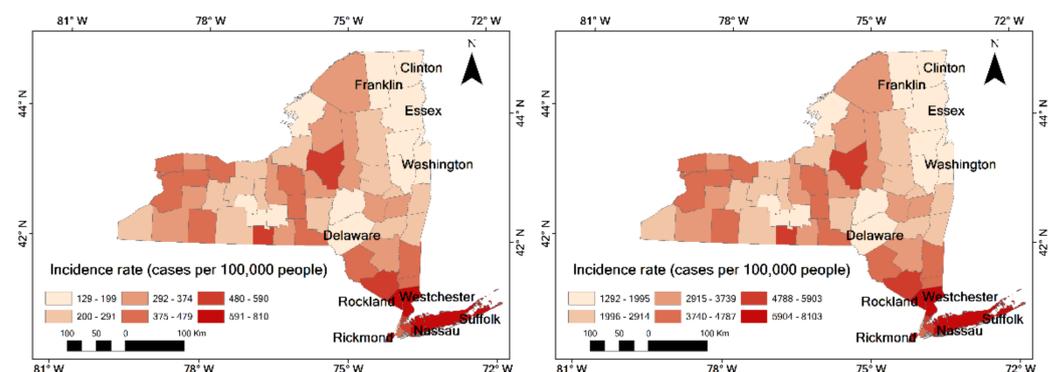


Figure 2. Distribution of monthly average and cumulative incidence rate of COVID-19 in New York State (March–December 2020).

3.3. Spatial Autocorrelation Analysis

The results of spatial autocorrelation analysis are shown in Table 3. As the table indicates, the global Moran's I was high at the beginning of the COVID-19 outbreak in NYC, then decreased sharply and reached the lowest point in September when the COVID-19 was spreading over the New York State. In general, the infection cases of COVID-19 in New York State were spatially clustered. Based on local spatial autocorrelation analysis, the COVID-19 hot spots of New York State were identified, and the results are shown in Figure S2. As the maps indicate, from March to May 2020, the High–High areas were the southern counties of New York State, and the Low–Low areas were the northern counties. Starting in June, High–High areas shrank to Westchester and Kings. Then, they moved from NYC to the central region of New York State. Starting in October, the range of High–High areas expanded to the west, while the northeastern Low–Low areas decreased.

Figure S3 shows that the hot spots from March to May were mainly located in NYC. Since May, cold spots appeared in western and northern New York State. From June to August, the range decreased, and hot spots appeared in the west. From September, hot spots were located in the Albany–Schenectady area in western New York. From June, cold spots could be found in the northern and western counties. The northern cold spots then greatly reduced after August. In November and December, a wide range of cold spots appeared in the northeast.

The frequency of monthly hot-spot occurrence of each county was produced by overlaying the results of hot-spot analysis, as shown in Figure 3. From March to December, Putnam and Westchester in southern New York State were hot spots for seven months, while counties in NYC were hot spots for four months. These were the key areas for prevention and control of COVID-19 spread. The cold spots were mainly concentrated in northeastern New York State, where the corresponding incidence rate of COVID-19 was low.

Table 3. Global Moran’s I analysis of COVID-19 incidence rate at the county level in New York State.

	Global Moran’s I	z-Score	p-Value	Spatial Pattern
March	0.7776	9.2978	0	cluster
April	0.7761	9.0308	0	cluster
May	0.5902	6.8092	0	cluster
June	0.3959	4.6962	0	cluster
July	0.2197	2.6244	0.0087	cluster
August	0.1373	1.7196	0.0855	cluster
September	0.0713	1.0222	0.3067	random
October	0.2443	3.1435	0.0017	cluster
November	0.2619	3.1161	0.0018	cluster
December	0.3545	4.1456	0.0000	cluster

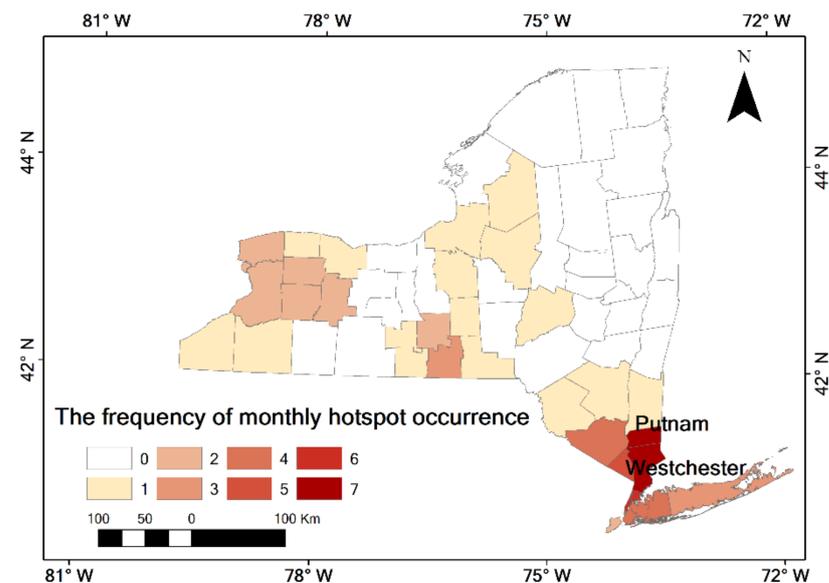


Figure 3. Occurrence of COVID-19 hot spots in New York State from March to December 2020.

3.4. Space-Time Scan Analysis

The space-time scan analysis produced duration, scope, and risk intensity of COVID-19 clustering in New York State as shown in Table S1. The cluster classes are ranked according to the value of likelihood ratio (LLR) in each month. As the epidemic developed, the number of clusters gradually increased from two in March to seven in September and then dropped to three in December. The aggregation region with the largest LLR (264053) was located in Westchester in December, and the aggregation region with the highest RR (8.31) was located in Otsego in September. A series of maps was produced to represent the COVID-19 spatiotemporal clustering area of each month in New York State, as shown in Figure S4, where the clusters first appeared in NYC in March, and then Nassau became the cluster center with a larger radius of 25.88 km in March and April. After June, the cluster area appeared in the western areas and gradually expanded into surrounding areas. In September, there were seven clusters in New York State. Among them, Chemung and Chautauqua were cluster centers with a radius of 83.86 km and 78.71 km, respectively.

Starting in November, the clustering area shrank. The radius of the cluster centered in Ontario reached 155.19 km. In December, the cluster in central New York State expanded again, and the clustering center in Tompkins appeared with a radius of 213.86 km.

To build an overview and evaluate the effectiveness of New York State's epidemic prevention and control policies, a timeline was produced, as shown in Figure 4, based on data collected from the state policy database of Boston University School of Public Health [35]. The results of space-time scan analysis and the daily confirmed cases were overlaid [36].

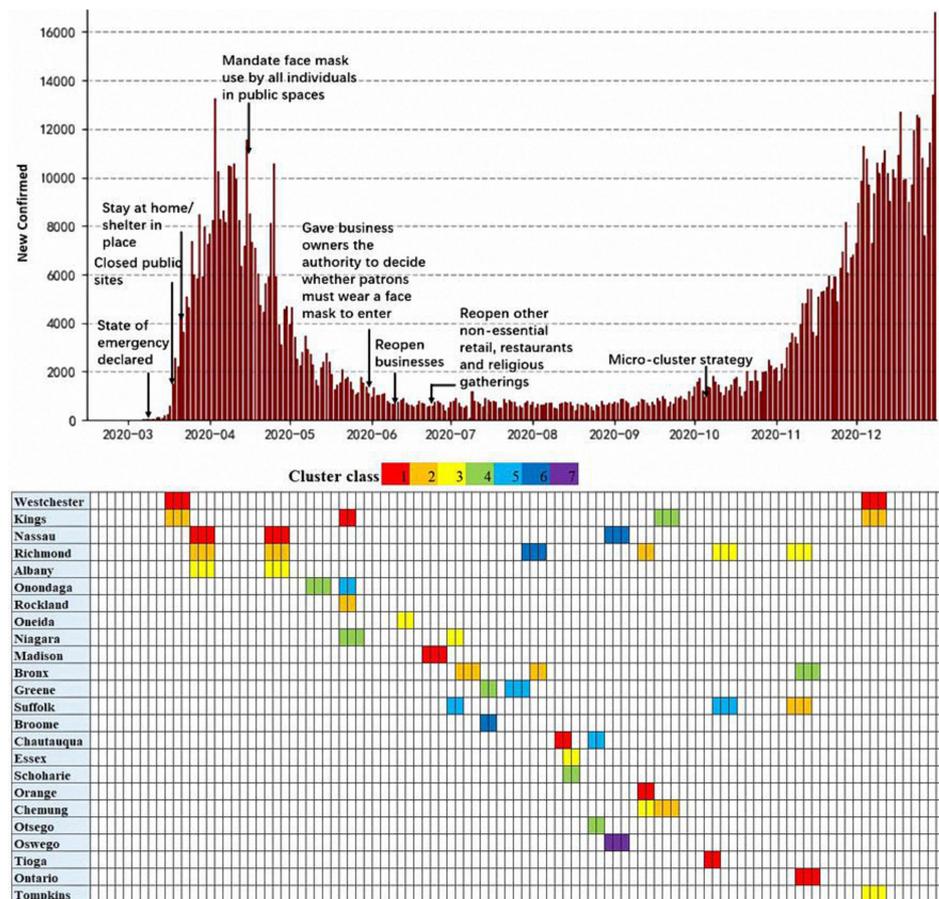


Figure 4. Timeline of newly confirmed cases, clusters, and related epidemic prevention policies of COVID-19 in New York State (the upper part shows the time series of the daily confirmed cases with the public health interventions, and the lower part shows the timeline and cluster time for each cluster).

When the first case was confirmed in New York State, a state of emergency was declared by the government. As the COVID-19 pandemic developed, the government took a series of preventive measures, including canceling inessential parties, closing public sites, etc. [35]. Daily confirmed New York State cases reached a peak in April and then decreased along with deterioration of the New York State economy. In June, as daily confirmed cases decreased sharply, New York State reopened the economy and relaxed the prevention policies. However, the fact that the COVID-19 coronavirus has an incubation period was not taken seriously. Subsequently, in November, there was a second peak of daily confirmed cases, which exceeded the first peak in April. When the situation became slightly optimistic, New York State began to reopen the economy, and personal prevention measures were not cautious, which caused the resurgence of the epidemic after October.

NYC became a hot spot of COVID-19 outbreaks in early 2020, partly due to its high population density. However, when New York State restarted the economy in June, the

epidemic clusters occurred in New York counties with relatively low population density, including Chemung (203.15 people/km²), Chautauqua (84.6 people/km²), and other counties in the Midwest. These places are not neighboring NYC. While the southern part was no longer a hot spot, the infection cases in the Midwest may have been imported from other regions when New York State's epidemic prevention policies were relaxed, and the population was highly mobile. In November, the New York epidemic deteriorated rapidly again. Large-scale clusters broke out in the west, while counties in northeastern New York State remained cold spots.

3.5. Geodetector

The COVID-19 infection cases exhibit variations on spatiotemporal heterogeneity over the ten-month research period. Geodetector was used to better analyze the driving factors in this process and their interaction. To prepare input data to Geodetector, it is necessary to select the best discretization method for each month. Therefore, this study utilized four different methods of discretization to classify the data of influencing factors. The q-statistics, which are the explanatory power of results, were used to determine the most suitable method corresponding to the highest q-statistics. Tables S2–S5 show the computation results. The natural breaks method was used for discretizing the data of March, April, November, and December. For the months in between, quantile was chosen in discretization.

This study set a condition of P value smaller than 0.05 to identify significant factors. According to Table 4, population density is the strongest factor concerning explanatory power, followed by temperature, unemployment rate, PM 2.5 concentration, population mobility, and precipitation, in a descending order. In the early epidemic stage, the explanatory power of these factors, especially population and temperature, were higher than 0.7. From May, all factors' explanatory power were decreasing. In October, the P values of all factors were greater than 0.05, meaning that all factors could not explain the COVID-19 incidence rate significantly. In November, the explanatory power of factors, including population density, unemployment rate, temperature, PM 2.5 concentration, and population mobility rose.

Table 4. Explanation of the power of influencing factors.

	Population Density	Unemployment Rate	Temperature	Precipitation	PM 2.5 Concentration	Population Mobility
March	0.8167	0.8167	0.7645	0.3746	0.3746	0.3154
April	0.7012	0.0322	0.8260	0.6195	0.3764	0.4343
May	0.3889	0.2465	0.4551	0.0842	0.5281	0.1939
June	0.4171	0.3013	0.3240	0.0429	0.3018	0.1663
July	0.4734	0.2266	0.2485	0.1203	0.0942	0.1453
August	0.3111	0.2597	0.1746	0.0237	0.1373	0.1347
September	0.2193	0.2168	0.1320	0.1059	0.0522	0.1302
October	0.0200	0.0426	0.1000	0.0940	0.0858	0.0683
November	0.1681	0.0977	0.2887	0.0698	0.1182	0.1565
December	0.2008	0.1156	0.1767	0.2009	0.1682	0.1532

■ represents $p < 0.005$; ■ represents $p < 0.01$; ■ represents $p < 0.05$.

As Table S6 shows, the q-statistic of each pair of interacting factors is greater than the q-statistics of a single factor, indicating that those interacting factors have had more significant impacts on the COVID-19 incidence rate than single factors during the research period. In March, population coupled with other factors generated reinforcing effects, as q values were greater than 0.8. In addition, interactions between temperature and other factors also produced reinforcing effects. From May, since the q-statistics of all single factors decreased, the interaction of any pair of factors decreased. In June, the interaction between PM 2.5 concentration and population density increased with q-statistics greater than 0.7, but then decreased after September.

4. Discussion

Much of the previous research on COVID-19 has explored the spatiotemporal patterns of COVID-19 and the key influencing factors of its transmission. However, research periods in most studies shorter than six months could not expose seasonal variations of spatiotemporal characteristics of COVID-19 and relationships among influencing factors. This paper investigated the spatiotemporal patterns using spatial autocorrelation and space-time scan analysis and analyses of the interaction of each pair of influencing factors of the COVID-19 pandemic in New York State between March and December 2020.

There are two important strongpoints in this study. First, this study revealed the spatiotemporal autocorrelation of COVID-19 spread in New York State in a period of ten months, and thereby identified changing patterns, covering three seasons, in more detail than previous works had done on relatively shorter research periods. It correlates with laboratories' findings that SARS-CoV-2 survives longer in low temperature conditions [37,38] which may further pose higher risks of infection in cold weather. Second, Geodetector analysis was employed in this study for analysis of influencing factors. The performance of four different discretization methods were evaluated based on q -statistic and the p -values of the results. The most appropriate method of discretization was determined for each month, which can adapt to the spatial heterogeneity of influencing factors in different months.

This study used both spatial autocorrelation and space-time scan analysis to explore the spatiotemporal clustering patterns more comprehensively. The clustering characteristics detected from both methods matched each other, in that high-risk clusters first appeared in NYC and then moved to the central and western parts of New York State as the epidemic developed. The spatiotemporal changes of these clusters were closely associated with the changes of the government's response to COVID-19. Since the daily confirmed cases decreased in May, New York State relaxed the prevention policies without effective restrictions and measures on population mobility across counties, which may have caused the epidemic spread from NYC to other parts of New York State, as reflected by spatial autocorrelation analysis. As was reflected in this analysis, economic activity and interpersonal contact between NYC and surrounding states may have also contributed to the spread of COVID-19 inside NYC.

Identifying the key factors influencing the spread of COVID-19 during different periods can help public health experts and policy makers adjust prevention measures more promptly. The monthly influencing factors and their relationships were analyzed through the Geodetector model. The results indicated that population density and temperature could correlate with the risks of COVID-19 spread. Of course, the COVID-19 infection and death rates are related to multiple factors, such as medical resources, economic conditions, and prevention strategies. These could indicate the decrease in all the factors' explanatory power since June, which we need to further explore with more data. The time resolution of analysis in this work was one month, and the daily data were aggregated. Detailed variations of relationships on a daily or weekly scale may not be determined accurately, especially since human mobility patterns can vary greatly from weekdays to weekends.

To prevent COVID-19 spread, COVID-19 vaccination has become an essential and effective solution. At present, since more than half the New York State population has been fully vaccinated, the incidence rate has gradually decreased [23], and New York State is now reopening its economy [39]. However, the present epidemic stage (July 2021) is similar to the stabilizing stage of the last year (June–September, 2020) of this research. The difference is that SARS-CoV-2 variants, including the Delta variant, have been spreading in the United States, and some of the variants might not be prevented effectively by current vaccines especially after nearly half a year since the earliest vaccination [40]. It is still too early to know about the duration of the antibodies generated from the available vaccines [41]. There is a risk that the epidemic might be resurgent in the future. If the effectiveness of COVID-19 vaccines could last for six months, people should be vaccinated periodically, and the government needs to guarantee sufficient supplies of vaccines. Considering these potential challenges, reopening policies shall be implemented gradually and cautiously.

Prevention measures, including social distancing, wearing facemasks, and hand hygiene should be suggested to members of the public.

5. Conclusions, Limitations, and Future Work

Based on the county-level COVID-19 dataset in New York State from March to December 2020, this study conducted both spatial autocorrelation and space-time scan statistical analysis to explore the spatiotemporal clustering patterns of COVID-19. Geodetector was applied to investigate the relationship between six environmental, demographic, and economic factors and the spread of COVID-19 across New York State. The results showed that the COVID-19 incidence rate has strong spatiotemporal heterogeneity and regional correlation. Both population density and temperature have been contributing to the spread of COVID-19 in different periods. To control and prevent the spread of COVID-19 and to ensure a conservative epidemic prevention strategy, New York State public health policies, based on careful consideration of social measures and supply of effective vaccines, need to be made for COVID-19 variants. Future research can divide New York State into some sub-regions based on the spatiotemporal patterns detected in this study and explore more underlying risk factors of COVID-19, such as medical resources, urbanization, and socioeconomic conditions in these regions. Surrounding regions shall also be included into analysis. The monthly aggregated data used in this work can be replaced with finer time scale data, for example weekly, daily, or even work days and weekends, to reveal cyclic temporal patterns due to different social mobility behaviors. A focused investigation can be conducted on determining suitable spatial and temporal radii in space-time scan analysis which may detect different clustering results and exhibit geographic scale effect.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijgi10090627/s1>, Figure S1. Distribution of Monthly Incidence of COVID-19 in New York State (March–December 2020); Figure S2. Cluster distribution of confirmed cases of COVID-19 in New York (March–December 2020); Figure S3. LISA cluster distribution map of the incidence of COVID-19 in New York State (March–December 2020); Figure S4. the COVID-19 spatiotemporal gathering area in 62 counties of New York State (March to December 2020); Table S1. Spatiotemporal scan statistics on the monthly cases; Table S2. Detection results of Geometrical interval; Table S3. Detection results of Equal interval; Table S4. Interactive detection results.

Author Contributions: Conceptualization, Anran Zheng and Tao Wang; methodology, Anran Zheng and Tao Wang; software, Anran Zheng; validation, Anran Zheng and Tao Wang; formal analysis, Anran Zheng; investigation, Anran Zheng; resources, Anran Zheng and Tao Wang; writing—original draft preparation, Anran Zheng; writing—review and editing, Anran Zheng, Xiaojuan Li and Tao Wang; visualization, Anran Zheng and Tao Wang; supervision, Tao Wang; funding acquisition, Tao Wang. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a NSFC grant (No. 41671403) and a NKRD program (No. 2017YFB050360103).

Acknowledgments: The authors acknowledge the comments and suggestions by the anonymous reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. COVID-19 Public Health Emergency of International Concern (PHEIC) Global Research and Innovation Forum. Available online: [https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-\(pheic\)-global-research-and-innovation-forum](https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum) (accessed on 13 March 2021).
2. COVID-19–PAHO/WHO Response, Report 42 (25 January 2021)–PAHO/WHO | Pan American Health Organization. Available online: <https://www.paho.org/en/documents/covid-19-pahowho-response-report-42-25-january-2021> (accessed on 13 March 2021).
3. CDC COVID Data Tracker. Available online: <https://covid.cdc.gov/covid-data-tracker/#county-view> (accessed on 4 January 2021).

4. New York's Economy and Finances in the COVID-19 Era. Available online: <https://www.osc.state.ny.us/reports/covid-19-october-14-2020> (accessed on 23 February 2021).
5. Chen, Z.L.; Zhang, Q.; Lu, Y.; Guo, Z.M.; Zhang, X.; Zhang, W.J.; Guo, C.; Liao, C.H.; Li, Q.L.; Han, X.H.; et al. Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China. *Chin. Med. J.* **2020**, *133*, 1044–1050. [[CrossRef](#)]
6. Kang, D.; Choi, H.; Kim, J.H.; Choi, J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int. J. Infect. Dis.* **2020**, *94*, 96–102. [[CrossRef](#)]
7. Liu, Y.; He, Z.; Xia, Z. Space-Time Variation and Spatial Differentiation of COVID-19 Confirmed Cases in Hubei Province Based on Extended GWR. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 536. [[CrossRef](#)]
8. Thakar, V. Unfolding Events in Space and Time: Geospatial Insights into COVID-19 Diffusion in Washington State during the Initial Stage of the Outbreak. *Int. J. Geo-Inf.* **2020**, *9*, 382. [[CrossRef](#)]
9. Gianquintieri, L.; Brovelli, M.A.; Pagliosa, A.; Dassi, G.; Brambilla, P.M.; Bonora, R.; Sechi, G.M.; Caiani, E.G. Mapping Spatiotemporal Diffusion of COVID-19 in Lombardy (Italy) on the Base of Emergency Medical Services Activities. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 639. [[CrossRef](#)]
10. Desjardins, M.R.; Hohl, A.; Delmelle, E.M. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Appl. Geogr.* **2020**, *118*, 102202. [[CrossRef](#)] [[PubMed](#)]
11. Hohl, A.; Delmelle, E.M.; Desjardins, M.R.; Lan, Y. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spat. Spatio-Temporal Epidemiol.* **2020**, *34*, 100354. [[CrossRef](#)] [[PubMed](#)]
12. Müller, H.; Louwsma, M. The Role of Spatio-Temporal Information to Govern the COVID-19 Pandemic: A European Perspective. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 166. [[CrossRef](#)]
13. Sun, F.; Matthews, S.A.; Yang, T.C.; Hu, M.H. A spatial analysis of COVID-19 period prevalence in US counties through June 28, 2020: Where geography matters? *Ann. Epidemiol.* **2020**, *52*, 54–59. [[CrossRef](#)] [[PubMed](#)]
14. Liu, L.; Hu, T.; Bao, S.; Wu, H.; Peng, Z.; Wang, R. The Spatiotemporal Interaction Effect of COVID-19 Transmission in the United States. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 387. [[CrossRef](#)]
15. Mollalo, A.; Vahedi, B.; Rivera, K.M. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci. Total Environ.* **2020**, *728*, 138884. [[CrossRef](#)]
16. Meng, Y.; Wong, M.S.; Xing, H.; Kwan, M.; Zhu, R. Yearly and Daily Relationship Assessment between Air Pollution and Early-Stage COVID-19 Incidence: Evidence from 231 Countries and Regions. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 401. [[CrossRef](#)]
17. Xie, Z.; Qin, Y.; Li, Y.; Shen, W.; Zheng, Z.; Liu, S. Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. *Sci. Total Environ.* **2020**, *744*, 140929. [[CrossRef](#)] [[PubMed](#)]
18. Zhou, C.; Su, F.; Pei, T.; Zhang, A.; Du, Y.; Luo, B.; Cao, Z.; Wang, J.; Yuan, W.; Zhu, Y.; et al. COVID-19: Challenges to GIS with Big Data. *Geogr. Sustain.* **2020**, *1*, 77–87. [[CrossRef](#)]
19. 2020 Census Apportionment Results. United States Census Bureau. Available online: <https://www.census.gov/data/tables/20/dec/2020-apportionment-data.html> (accessed on 7 March 2021).
20. Squizzato, S.; Masiol, M.; Rich, D.Q.; Hopke, P.K. PM_{2.5} and gaseous pollutants in New York State during 2005–2016: Spatial variability, temporal trends, and economic influences. *Atmos. Environ.* **2018**, *183*, 209–224. [[CrossRef](#)]
21. World Population Review. Available online: <https://worldpopulationreview.com/us-counties/states/ny> (accessed on 13 March 2021).
22. Market Insider. Available online: <https://markets.businessinsider.com/news/stocks/11-mind-blowing-facts-about-new-yorks-economy-2019-4-1028134328> (accessed on 25 March 2021).
23. US COVID-19 Cases and Deaths by State. USA Facts. Available online: <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/> (accessed on 4 July 2021).
24. GIS.NY.GOV. Available online: <http://gis.ny.gov/gisdata/inventories/details.cfm?DSID=927> (accessed on 15 December 2020).
25. National Centers for Environmental Information. Available online: <https://www.ncdc.noaa.gov/cag/county/mapping/30/tavg/202004/1/anomaly> (accessed on 14 January 2021).
26. Air Monitoring Website. Available online: <http://www.nyaqinow.net/> (accessed on 15 January 2021).
27. Unacast Social Distancing Scoreboard. Available online: <https://www.unacast.com/covid19/social-distancing-scoreboard?view=county&fips=36023> (accessed on 15 January 2021).
28. FRED Economics Data. Available online: <https://fred.stlouisfed.org/tags/series?t=new+york+county%2C+ny> (accessed on 8 January 2021).
29. US EPA. Available online: <https://www.epa.gov/outdoor-air-quality-data/download-daily-data> (accessed on 17 January 2021).
30. Wang, Q.; Dong, W.; Yang, K.; Ren, Z.; Huang, D.; Zhang, P.; Wang, J. Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors. *Int. J. Infect. Dis.* **2021**, *105*, 675–685. [[CrossRef](#)] [[PubMed](#)]
31. Boscoe, F.P.; McLaughlin, C.; Schymura, M.J.; Kielbaso, C.L. Visualization of the spatial scan statistic using nested circles. *Health Place* **2003**, *9*, 273–277. [[CrossRef](#)]
32. Wang, J.; Xu, C. Geodetector: Principle and prospective. *Acta Geograph. Sin.* **2017**, *72*, 116–134.
33. Wang, J.; Zhang, T.; Fu, B. A measure of spatial stratified heterogeneity. *Ecol. Indic.* **2016**, *67*, 250–256. [[CrossRef](#)]
34. Cao, F.; Ge, Y.; Wang, J. Optimal discretization for geographical detectors-based risk assessment. *GIScience Remote Sens.* **2013**, *50*, 78–92. [[CrossRef](#)]

35. Raifman, J.; Nocka, K.; Jones, D.; Bor, J.; Lipson, S.; Jay, J.; Cole, M.; Krawczyk, N.; Benfer, E.A.; Chan, P.; et al. *COVID-19 US State Policy Database*; Inter-University Consortium for Political and Social Research [Distributor]: Ann Arbor, MI, USA, 2021. [[CrossRef](#)]
36. Kim, S.; Castro, M.C. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *Int. J. Infect. Dis.* **2020**, *98*, 328–333. [[CrossRef](#)] [[PubMed](#)]
37. Matson, M.; Yinda, C.; Seifert, S.N.; Bushmaker, T.; Fischer, R.J.; van Doremalen, N.; Lloyd-Smith, J.O.; Munster, V.J. Effect of Environmental Conditions on SARS-CoV-2 Stability in Human Nasal Mucus and Sputum. *Emerg. Infect. Dis.* **2020**, *26*, 2276–2278. [[CrossRef](#)] [[PubMed](#)]
38. Chi, Y.; Wang, Q.; Chen, G.; Zheng, S. The Long-Term Presence of SARS-CoV-2 on Cold-Chain Food Packaging Surfaces Indicates a New COVID-19 Winter Outbreak: A Mini Review. *Front. Public Health* **2021**, *9*, 650493. [[CrossRef](#)] [[PubMed](#)]
39. Mathieu, E.; Ritchie, H.; Ortiz-Ospina, E.; Roser, M.; Hasell, J.; Appel, C.; Giattino, C.; Rodes-Guirao, L. A global database of COVID-19 vaccinations. *Nat. Hum. Behav.* **2021**. [[CrossRef](#)]
40. The New York Times. Available online: <https://www.nytimes.com/2021/05/19/nyregion/nyc-reopening-restrictions-masks-vaccines.html> (accessed on 30 June 2021).
41. CDC. Available online: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/keythingstoknow.html?s_cid=10536:%2Bcovid%20%2B19%20%2Bvaccine:sem.b:p:RG:GM:gen:PTN:FY21 (accessed on 30 June 2021).