

COVID-19 Dynamics: Unraveling Disparities in Metropolitan and Micropolitan Statistical Areas

Colin Qu (zq2271), Ines Figueira (ip2347), Ligao Ruan(lr2925), Qingyang Zhao(qz2567)

May 8, 2024

1 Introduction

The COVID-19 pandemic has underscored the critical need for nuanced epidemiological analyses that account for the spatial and urban dimensions of disease spread and control. This study investigates the dynamics of COVID-19 pandemic from 2020 to early 2023 using urban scaling and spatial autocorrelation techniques. To differentiate between smaller and larger cities, we rely on the definitions of metropolitan statistical areas (MSAs) and micropolitan statistical areas (MicroSAs) provided by the United States Office of Management and Budget (OMB). According to OMB guidelines, an MSA must encompass at least one urban area with a population of 50,000 or more, while each MicroSA must include at least one urban area with a population of at least 10,000 but fewer than 50,000 people [8, 15]. In a broader context, MSAs and MicroSAs represent core areas with major population hubs and surrounding communities characterized by a strong level of economic and social integration with its core. This distinction allows for a comprehensive examination of COVID-19 dynamics across different urban scales, providing insights into the interplay between urban structure and disease spread.

Our research employs urban scaling to quantify how COVID-19 metrics evolve with city size, a critical factor in understanding disease transmission and management efficacy in densely populated areas. Additionally, we use spatial autocorrelation to identify patterns of spread that are not random but rather influenced by geographic and social proximities. By integrating these methodologies, the study aims to reveal deeper insights into how urban structure and spatial dependencies influence pandemic outcomes.

Specifically, we compare the Scale-Adjusted Metropolitan Indicator (SAMI) and the Spatial and Scale-Adjusted Metropolitan Indicator (SSAMI) from the OLS and ESF models respectively, to assess disparities in COVID-19 cases, deaths, and vaccination rates. These indicators have been adjusted to reflect the unique urban configurations and interconnectivities that characterize MSAs and MicroSAs, providing a refined lens through which the pandemic's impacts can be evaluated.

Through a detailed comparative analysis of SAMI and SSAMI, this report highlights how spatial adjustments can alter our understanding of COVID-19 impacts, enhancing the accuracy of public health assessments and guiding more targeted interventions. Furthermore, the study incorporates correlation analysis to examine relationships between key COVID-19 metrics across MSAs and MicroSAs, followed by a cluster classification approach to categorize cities based on their risk profiles. These methodologies enhance our understanding of spatial dynamics and intervention effectiveness during the pandemic.

2 Data Sources

To assess the evolution and the impact of the COVID-19 pandemic, we collected data from the following resources:

- **COVID-19 Data Repository :** Daily time series summary tables that include the number of confirmed cases and deaths starting January 22th 2020 to March 9th 2023. This dataset is managed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [9].
- **COVID-19 Vaccinations in the United States:** Overall U.S. COVID-19 Vaccine administration and vaccine equity data at county level. Data represents includes the number of people with at least one administered dose by state of residence and was last updated May 12, 2023 [12].
- **Personal Income:** The Personal Income dataset is maintained by the Bureau of Economic Analysis (BEA), which is a part of the U.S. Department of Commerce. This dataset provides information regarding personal income at the county, metropolitan, and state level [14].
- **Primary Care Physicians:** Dataset containing the number of Primary Care Physicians in the United States by county.
- **Population Data:** Annual estimates of the resident population for Incorporated Places and Minor Civil Divisions in the United States provided by the Census Bureau [7].
- **CBSA (MSA and MicroSA):** Codes for metropolitan and micropolitan statistical areas used to delineate these [8].
- **MSA shapefile:** Cartographic boundary files of selected geographic areas from the Census Bureau's used to get compute geographical coordinates of the centroid of each metropolitan and micropolitan Statistical Area [6].

3 Methods

3.1 Urban Scaling

To analyze the COVID-19 pandemic's impact on different sized urban systems, we employ a scaling analysis methodology pioneered by [5]. Specifically, we examine how certain parameters such as the number of deaths, cases by COVID-19 and vaccination vary with the size of the urban system, typically represented by its population (N). This relationship can often be often expressed as a scaling law:

$$Y = Y_0 N^\beta e^\epsilon \quad (1)$$

Here, Y is the quantity of the city's property at a certain time, β is the exponent that indicates if Y follows a sublinear ($\beta < 1$), linear ($\beta = 1$), or superlinear ($\beta > 1$) relation.

This scaling law is often written in logarithmic form:

$$\ln Y_i = \ln Y_0 + \beta \ln N_i + \xi_i \quad (2)$$

Typically, $\ln Y_0$ and β are determined by fitting a linear model using ordinary least squares regression (OLS) on the logarithmically transformed data. Here, ξ_i are the scaling residuals that represent the Specific Attribute Metric Index (SAMI) for each city i , capturing its deviation from the expected scaling. SAMIs are often used for evaluating and comparing different sized cities [1]. It is important to note that the estimation of β directly influences the SAMIs, as these are functions of the estimated Y_i . Consequently, misestimating the scaling exponent can lead to biased predictions of a city's development. In fact, [17] shows that accounting for spatial autocorrelation on SAMIs allows us to generate better fitted results when estimating scaling relations. Following a similar methodology to the one described in that study we compare OLS model and ESF model performance.

3.2 Spatial Autocorrelation and Spatial Models

Spatial autocorrelation refers to the relationship between nearby values in a dataset. It is a key concept in spatial statistics and geographic information science that is also relevant in the modeling and analyses of spatial data. For example, when using linear models like urban scaling, the presence of non-zero spatial autocorrelation violates the assumption of independently and identically distributed errors, thus we need to account for spacial autocorrelation [11].

Moran's I is a common metric used to evaluate the spatial autocorrelation between variables in different locations within a given spatial domain. Therefore, Moran's I can be used to identify spatial clustering patterns, spatial heterogeneity, and spatial dependency.

$$Moran's I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2} \quad (3)$$

where z_i is the deviation of an attribute for feature i from its mean ($x_i - \bar{x}$), w_{ij} is the spatial weight between feature i and j , n is equal to the total number of features, and S_0 is the aggregate of all spatial weights:

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (4)$$

Note that if we row-standardized weights (i.e. $S_0 = n$), we can write Moran's I as shown in eq.3 [2].

Jong et al. [10] showed that given a spatial structure described by a spatial weights matrix, the largest value that Moran's I can take is the scaled value of the largest positive eigenvalue of the spatial weight matrix, and the most negative value that Moran's I can take is the scaled value of the most negative eigenvalue of this matrix. Furthermore, a positive Moran's I indicates spatial clustering (positive spatial autocorrelation), a negative value indicates spatial dispersion (negative spatial autocorrelation), and a value close to zero indicates spatial randomness (no spatial autocorrelation). In section 3.3 we will use this metric for the eigenvectors selection (selection of spatial predictors).

Various models have been proposed to tackle the issue of spatial autocorrelation in regression analyses. These models aim to filter spatial autocorrelation from regression results, thereby improving their reliability. Some of these include: (1) spatial error models, (2) spatial lag models, also known as spatial autoregression models, and (3) eigenvector spatial filtering (ESF) models [3]. Spatial lag models incorporate a lagged dependent variable term into the regression equation. It assumes that the value of the dependent variable in each location is influenced not only by the independent variables but also by the values of the dependent variable in neighboring locations [13]. Unlike the spatial lag model, the spatial error model assumes that the spatial autocorrelation is present in the error term rather than in the dependent variable. It introduces a spatially auto-correlated error term into the regression equation, capturing the residual spatial dependence not explained by the independent variables. The spatial error model is particularly useful when the spatial autocorrelation is a result of unobserved spatially correlated factors. The ESF model family utilizes the concept that the spatial data pattern can be converted into explanatory variables within the regression model to account for spatial effects. In this approach, eigenvectors of spatial matrices are included as spatial predictors in regression analyses to capture the spatial impact effectively [17].

3.3 Eigenvector Spatial Filtering (ESF) Model

To construct the ESF model we start by constructing the distance matrix \mathbf{W} , as described in [17]:

$$\mathbf{W} = [\omega_{ij}] = \begin{cases} 0, & \text{if } i = j \\ 0, & \text{if } d_{ij} > d_0 \\ 1 - d_{ij}/d, & \text{otherwise} \end{cases} \quad (5)$$

In the equation, w_{ij} represents the elements of the matrix \mathbf{W} . Here, d_{ij} denotes the distance between two locations, indexed as i and j . Additionally, d_0 is the threshold distance for spatial connectivity between cities and is assumed to be $d_0 = 989.5$ Km, consistent with the assumptions made in the original paper.

The spatial weight matrix is given by

$$\boldsymbol{\Omega} = (\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)^\top \quad (6)$$

Next, we compute the eigenvectors using the computed spatial weight matrix:

$$\boldsymbol{\Omega}\mathbf{V} = \mathbf{V}\lambda \quad (7)$$

where λ is the vector of eigenvalues and \mathbf{V} is the matrix of eigenvectors. To minimize the amount of spatial autocorrelation in the residuals we select suitable eigenvectors as spatial predictors are represented as $\mathbf{V}_{se} = (V_1, V_2, V_3, \dots, V_k)$.

The selection of the best subset of spatial eigenvectors (MEM, Moran's Eigenvector Maps) is done using the R package [16], more specifically the function *mem.select*. This function computes the spatial eigenvectors (MEM) of the spatial weigh matrix (Ω) provided and optimizes the selection of a subset of MEM variables relative to the response variable. The optimization is then done by minimizing the residual spatial autocorrelation using the *MIR* method (see details in [4]). For the *MIR* method, the global test consists in computing the Moran's I of the residuals of the model of the response variable against environmental variables) and tests it by permutation. If the Moran's I is significant, the function performs a selection procedure that searches among the set of generated spatial predictors the one that best minimizes the value of the Moran's I. A model is built, and the significance of the Moran's I of the model residuals is tested again. The procedure goes on until that Moran's I is not significant anymore, hence the name of Minimization of Moran's I in the Residuals (MIR).

Using Eq.2 and the selected spatial predictors we obtain the ESF model described by:

$$\ln Y_i = \ln Y_0 + \beta N_i + \gamma_1 V_1 + \gamma_2 V_2 + \dots + \gamma_k V_k + \xi_i \quad (8)$$

To assess the cities performance without the influence of spatial autocorrelation, we will use ξ_i , the residual of the ESF model, that is a spatial and scale-adjusted metropolitan indicator (SSAMIs).

4 Results

4.1 Temporal Evolution of the COVID-19 Epidemic

The daily newly added cases/deaths/vaccinations in the US are shown in Fig.1 Combining temporal trends in cases/deaths/vaccinations, this study divided the US COVID-19 epidemic from Jan 22, 2020 to May 9, 2023 into five phases. Phase 1 ended on April 22, 2020. Phase 2 continued until January 22, 2021. Compared to the other phases, the number of new cases in Phase 3 (April 15, 2021 to December 10, 2021) was relatively flat. However, new cases and the increase rate of the COVID-19 cases in Phase 4 (December 22, 2021 to March 1, 2022) were significantly higher than the other phases, while deaths did not change obviously. This wave (Phase 4) was mainly caused by the highly infectious Omicron variant [18]. Finally in Phase 5 (April, 2022 to March 2023), all three daily numbers are relatively stable.

When Covid-19 hit, people did not properly protect and treat the virus, resulting in many deaths in the first peak period in April, 2020. For the second peak period in December, 2020, as winter approaches, Covid-19 is showing an outbreak trend, and the daily number of confirmed cases and deaths remains high. Some areas have begun to encourage people to get vaccinated. With the large coverage of vaccination and the reduction of virus activity caused by warmer weather, the number of daily cases and deaths in the third period have been effectively controlled. During the fourth peak period (December, 2021 to January, 2022), the number of confirmed cases increased exponentially. In addition to the Omicron variant, it is speculated that the reasons may include: people's reduced vigilance against the virus, winter climate factors, etc. In phase five, daily cases and deaths stabilize, largely due to high coverage with a series of vaccines and widespread immunity.

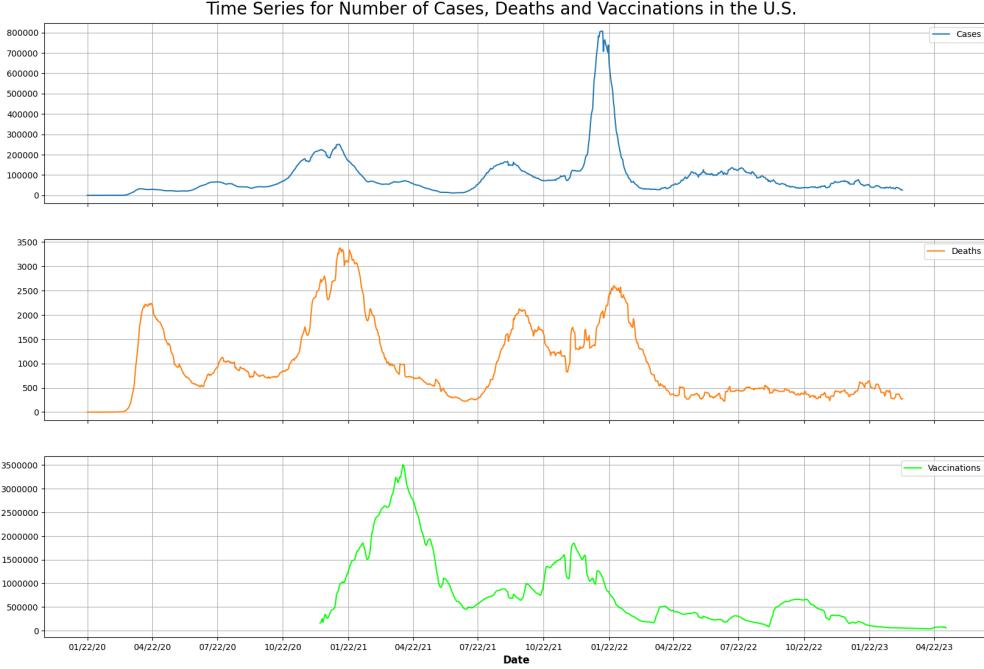


Figure 1: Temporal variations of COVID-19 cases, deaths and first administered doses in the United States. (a) Newly added COVID-19 cases per day. (b) Newly added deaths per day. (c) Newly administered vaccines per day.

4.2 Model Comparisons

In the present project all the scaling analyses were conducted by year (starting January 22th of 2022 to March 9th 2023) to allow for a better understanding of the temporal dynamics and trends of the pandemic that might have been overlooked if the data was aggregated. The results are shown in Table 1.

We can see the scaling coefficient β estimated with the non-spatial ordinary least square model (OLS) and the eigenvector spatial filtering model (ESF) for COVID-19 cases, deaths and vaccines. The adjusted R^2 of the ESF model is always bigger than the OLS model, indicating that using the ESF model could derive better-fitted results. A Welch's t-test was performed to see whether the estimated scaling exponent is significant different using both models and no significant differences were found (all $p - values > 0.05$). Nevertheless, our analyses shows that in some instances, like the annual cases in 2020, using OLS or ESF models will result in different conclusions regarding superlinear or sublinear scaling (Fig. 2).

Table 1: Estimation results for COVID-19 cases, deaths and vaccination between ordinary least square (OLS) and eigenvector spatial filtering (ESF) model.

Year	OLS		ESF		
	β	Adj- R^2	β	Adj- R^2	
Cases	2020	0.939 [0.913;0.965]	0.845	0.988 [0.970;1.007]	0.933
	2021	0.986 [0.972;0.999]	0.959	0.987 [0.976;0.998]	0.974
	2022	1.032 [1.018;1.047]	0.956	1.029 [1.016;1.042]	0.964
	2023	1.019 [0.996;1.041]	0.894	1.017 [0.998;1.036]	0.929
Deaths	2020	0.926 [0.889;0.963]	0.727	0.957 [0.924;0.989]	0.802
	2021	0.883 [0.855;0.911]	0.806	0.868 [0.844;0.892]	0.871
	2022	0.929 [0.902;0.955]	0.840	0.916 [0.894;0.938]	0.894
	2023	0.963 [0.928;0.998]	0.774	0.953 [0.919;0.987]	0.799
Vaccines	2021	1.073 [1.059;1.087]	0.962	1.060 [1.047;1.073]	0.971
	2022	1.070 [1.058;1.083]	0.968	1.059 [1.047;1.071]	0.975
	2023	1.070 [1.057;1.082]	0.968	1.061 [1.049;1.072]	0.975

Table 2 compares the results for the Moran's I test results for the scaling estimations using both models. It shows that there is some positive spatial autocorrelation in the residuals of each OLS model, indicating that the linear model using OLS might be misspecified due to the existence of spatial autocorrelation. In contrast, the Moran's I test results for the ESF model don't show significant spatial autocorrelation, indicating the estimated results are unbiased in terms of the influence of spatial autocorrelation. This can be seen if we look at Moran's I p-values. For the OLS model they are always < 0.01 , while for the ESF model is $p - value > 0.01$ indicating there is no statistically significant spatial autocorrelation in the dataset at the chosen significance level.

Table 2: Moran's I results for COVID-19 cases, deaths and vaccination using ordinary least square (OLS) and eigenvector spatial filtering (ESF) model.

Year	OLS		ESF		
	Moran's I	P-value	Moran's I	P-value	
Cases	2020	4.650e-1	< 0.01	1.082e-2	0.017
	2021	3.602e-1	< 0.01	2.748e-3	0.436
	2022	1.179e-1	< 0.01	3.737e-3	0.329
	2023	2.659e-1	< 0.01	3.984e-3	0.311
Deaths	2020	2.530e-1	< 0.01	5.177e-3	0.159
	2021	2.331e-1	< 0.01	5.584e-3	0.165
	2022	2.273e-1	< 0.01	3.936e-3	0.294
	2023	7.512e-2	< 0.01	5.338e-3	0.161
Vaccines	2021	1.439e-1	< 0.01	6.358e-3	0.094
	2022	1.356e-1	< 0.01	4.854e-3	0.177
	2023	1.280e-1	< 0.01	5.235e-3	0.152

Figure 2 tracks the scaling relation evolution for COVID-19 cases, deaths, and vaccination over four years. The scaling relationship between cases and population goes from being slightly sublinear to being superlinear. Not surprisingly, the scaling relation for deaths/vaccination and urban population size exhibits superlinear behaviour for the time considered, suggesting improved healthcare delivery and vaccination campaigns with urban growth.

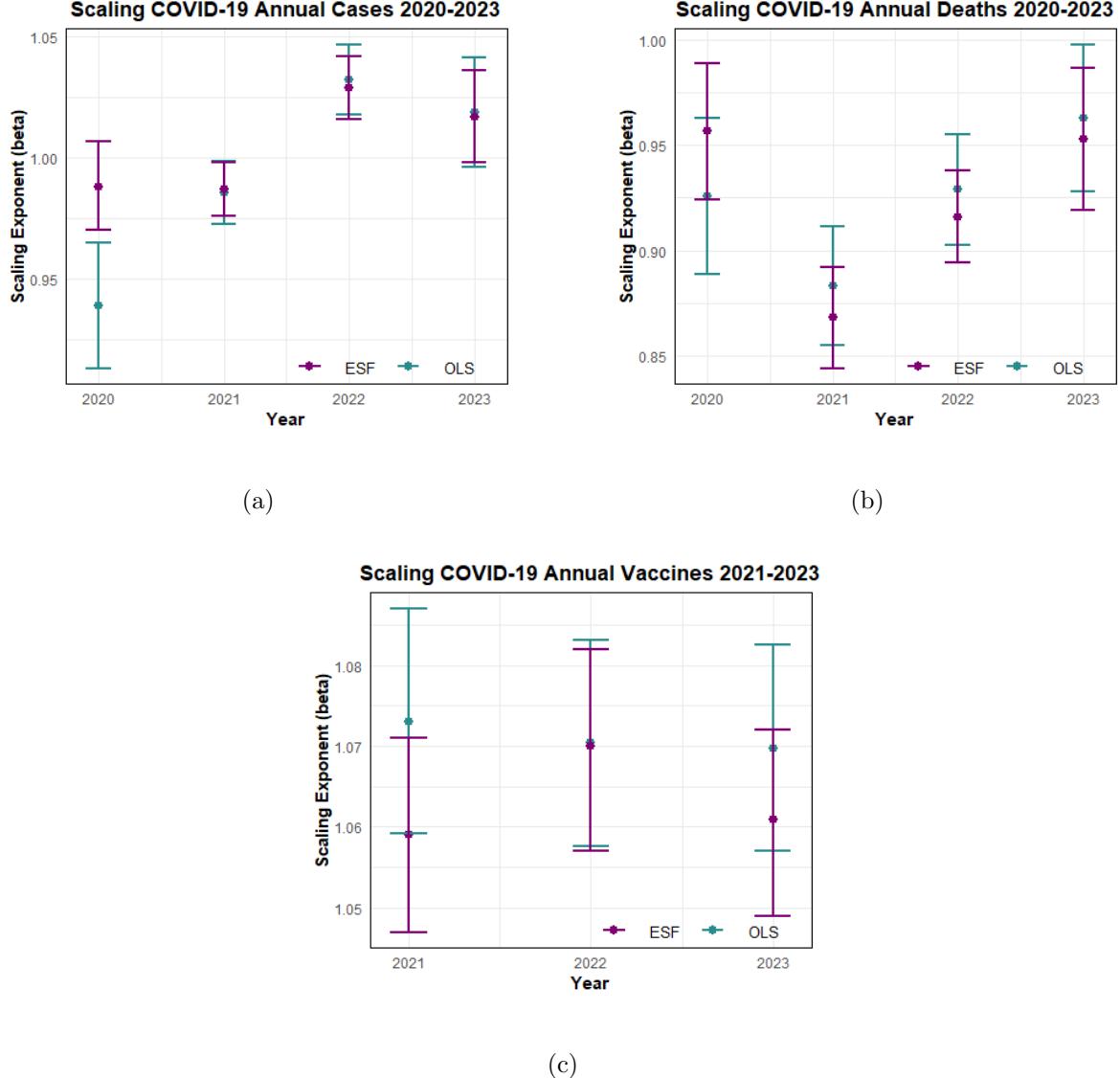


Figure 2: Scaling exponent (β) of COVID-19 cases, deaths and vaccination over time using ordinary least square (OLS) and eigenvector spatial filtering (ESF) model. The exponent β is reported along with its 95% confidence interval.

4.3 Comparison between SAMIs and SSAMIs

In order to show the difference between the SAMIs and SSAMIs, we compare the ranks of each city (R_{SAM} and R_{SSAM}) obtained from SAMIs and SSAMIs. In this comparison, cities are ranked from 1 to the total number of cities based on the descending order of their respective indicator values. Fig. 3 illustrates the correlation between these two sets of ranks. For all of the variables the Spearman correlation test shows a high correlation between the two indicators (0.783 for cases, 0.936 for deaths, and 0.813 for vaccination).

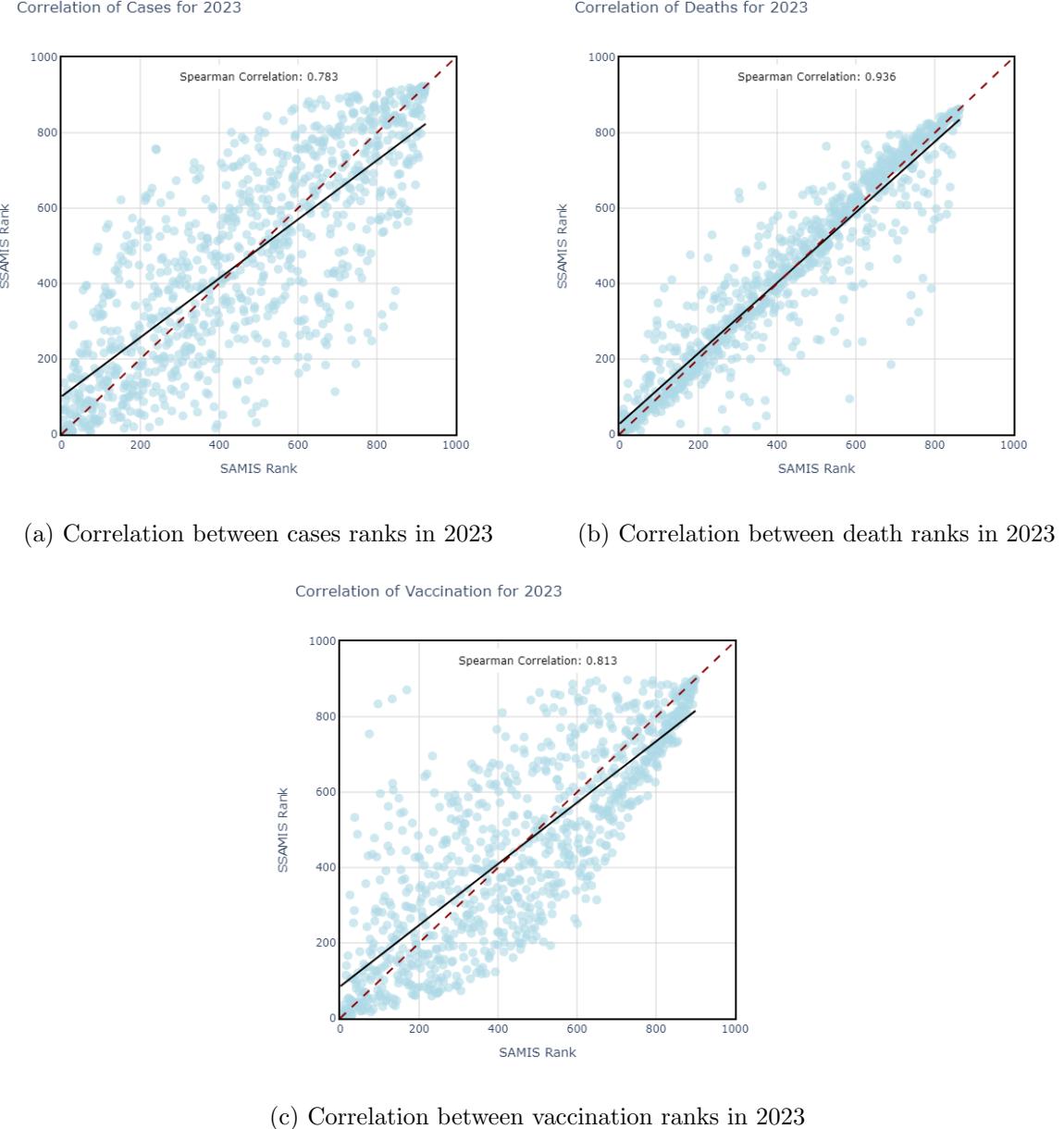


Figure 3: Scatterplot of SAMIs and SSAMIs Ranks in 2023.

Figure 4 illustrates the distribution of city ranks. We find that the distribution of R_{SAMI} and R_{SSAMI} follows a similar pattern for the variables considered. Nevertheless, the distribution of R_{SAMI} shows a more clustered pattern while the distribution of R_{SSAMI} is more even across space.

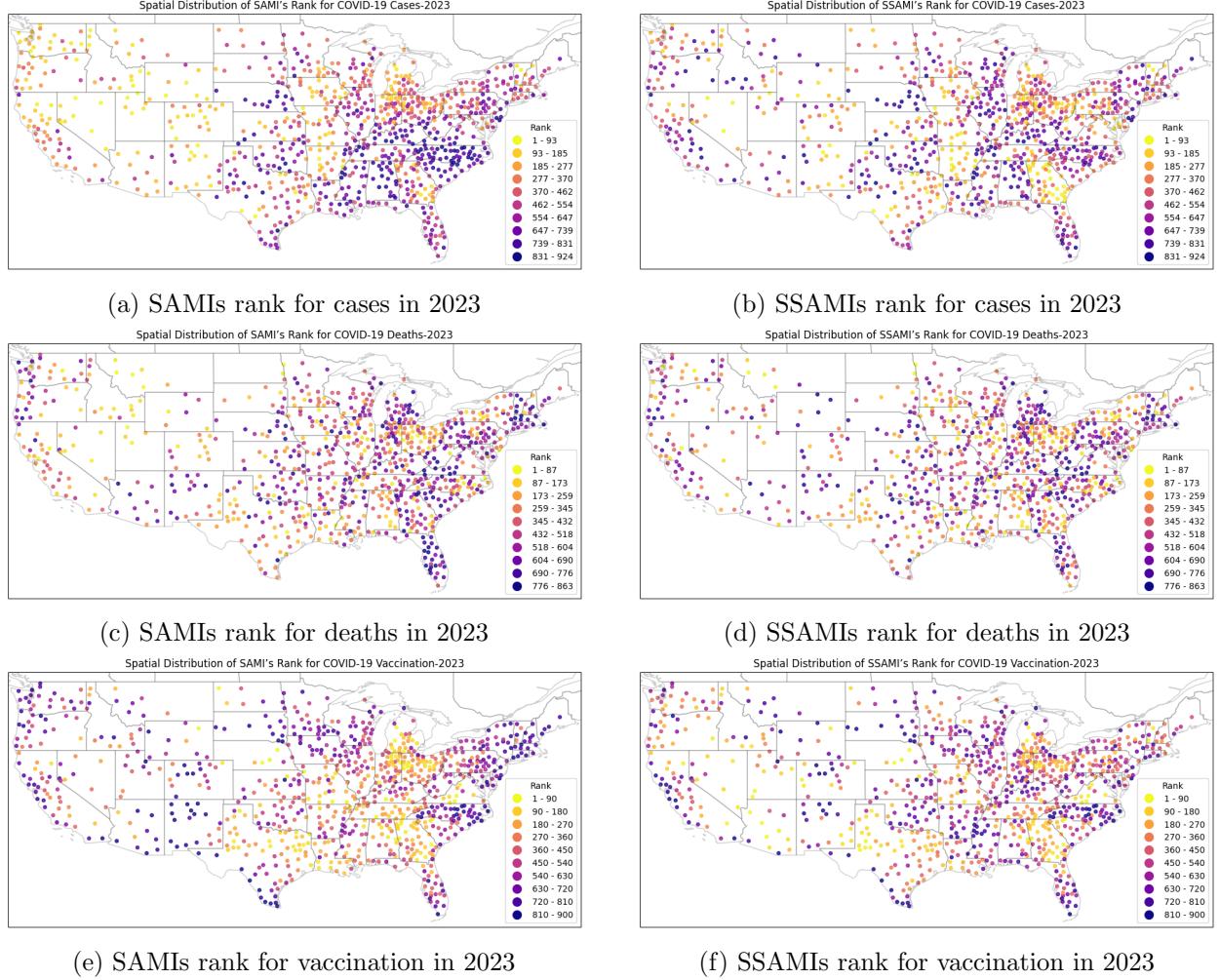


Figure 4: Spatial distribution of SAMIs and SSAMIs ranks in 2023.

To quantitatively analyze the disparity between R_{SSAMI} and R_{SAMI} , we visualize their differences ($R_{SSAMI} - R_{SAMI}$) spatially, as depicted in Fig. 6.

For deaths and cases, a positive difference indicates that SAMIS overestimates severity, as it assigns a more severe ranking (a lower numerical rank) than SSAMIS. This suggests that SAMIS perceives the situation as worse than it is evaluated by SSAMIS, potentially leading to an over-allocation of resources or overly cautious policy measures. Conversely, a negative difference suggests SAMIS underestimates the severity, possibly leading to insufficient responses or under-preparation.

In the context of vaccinations, the interpretation inverses due to the positive connotation of high vaccination rates. A positive difference indicates that SAMIS overestimates the effectiveness of vaccinations, attributing a better performance rank (lower numerical rank) than SSAMIS.

For example, in comparing Fig. 5c with Fig. 5d, it is noted that the New York-Newark-Jersey City, NY-NJ-PA MSA has a death rank of 324 under SAMI, which is numerically lower than its rank under SSAMI. This discrepancy suggests that SAMI overestimates the severity of conditions

compared to SSAMI. Conversely, regarding vaccinations, as shown in Fig. 5e and Fig. 5f, the SAMI rank is 344, significantly higher than the SSAMI rank of 612. This indicates that SAMI overestimates the effectiveness of vaccination coverage, as a lower rank denotes better performance. Such disparities underscore the importance of considering the context-specific implications of rankings when using SAMI and SSAMI for evaluating public health metrics.

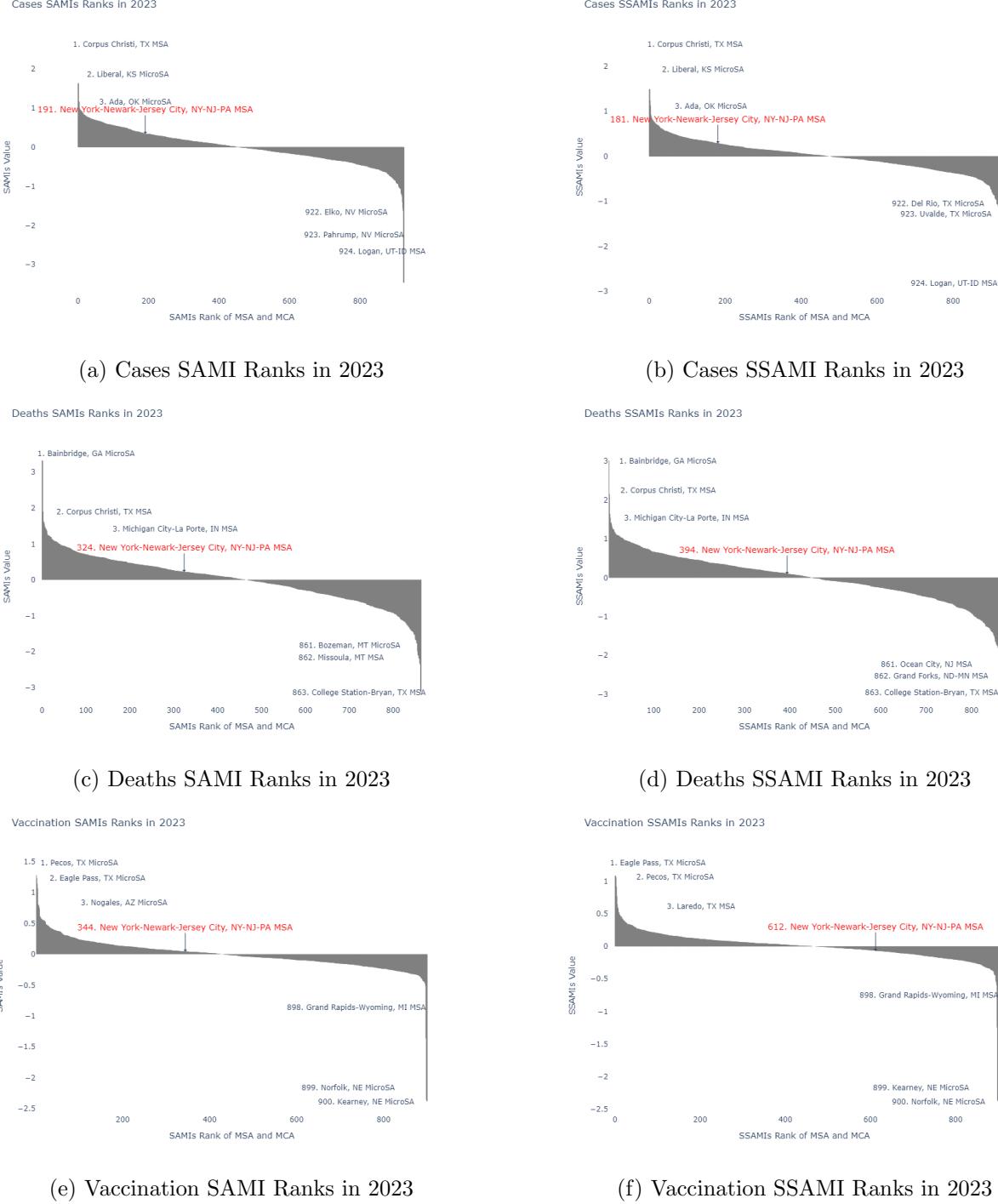
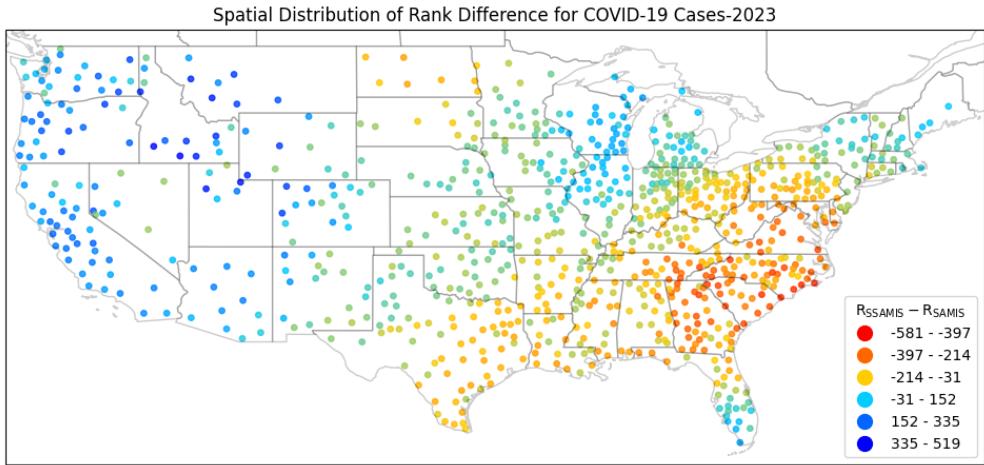
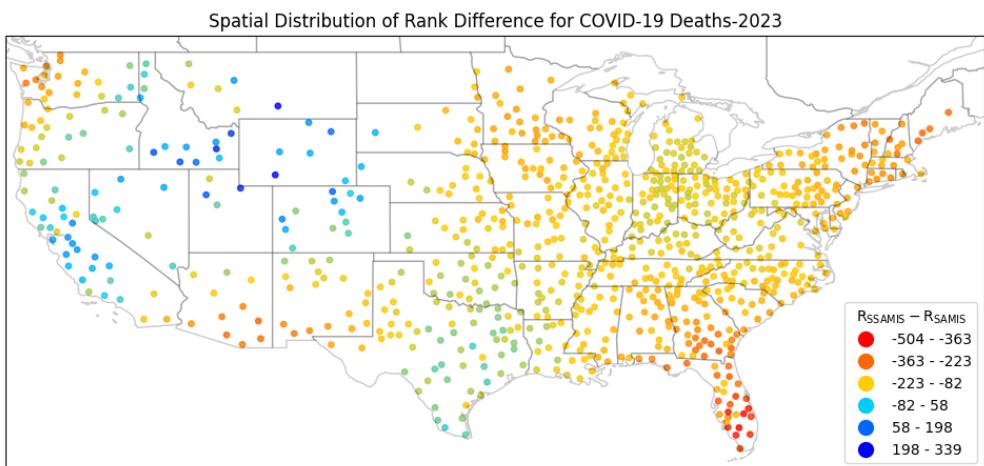


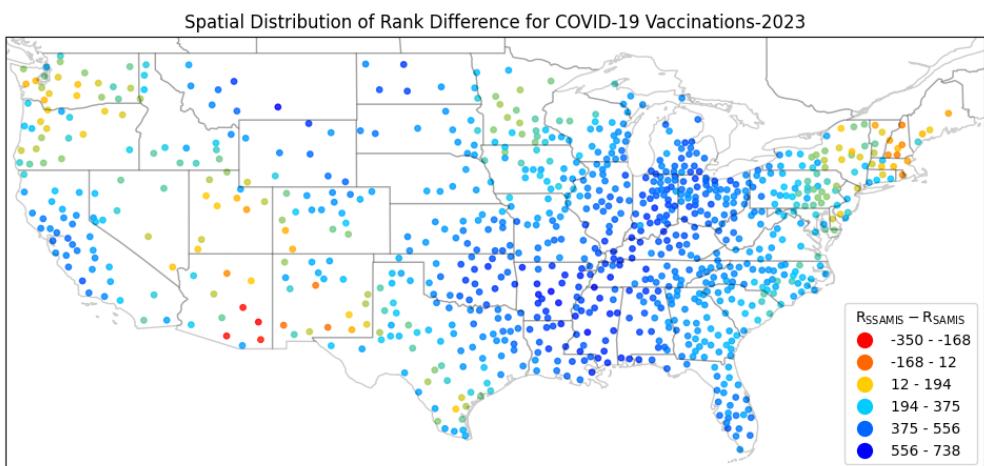
Figure 5: Ranking of MSAs and MicroSAs in 2023 by SAMIs (left) and SSAMIs (right).



(a) Cases rank difference in 2023



(b) Deaths rank difference in 2023



(c) Vaccinations rank difference in 2023

Figure 6: Difference of city ranks between two indicators ($R_{SSAMI} - R_{SAMIS}$) in 2023.

4.4 Correlation between SSAMIs

In this segment of our analysis, we explore the interdependencies among various spatially and scale-adjusted metropolitan indicators (SSAMIs) concerning COVID-19 cases, deaths, vaccinations, regional GDP, and the number of primary care physicians across MSAs. By focusing on these particular metrics from 2020 to 2023, this analysis aims to elucidate how different public health and economic indicators interact and potentially correlate with each other across geographic and temporal scales.

The method employed involves selecting relevant columns from a comprehensive dataset that includes SSAMIs for cases, deaths, vaccinations, GDP, and doctor counts across 2020 to 2023. Missing data within this dataset are imputed using median values to maintain consistency and integrity in the analysis. A correlation matrix is then calculated to quantify the relationships among these variables. The heatmap visualization provides a clear, visual representation of these correlations, offering insights into which factors are more closely interconnected. This approach is intended to provide a foundational understanding of the dynamics between health outcomes and economic factors, helping policymakers to tailor interventions more effectively.

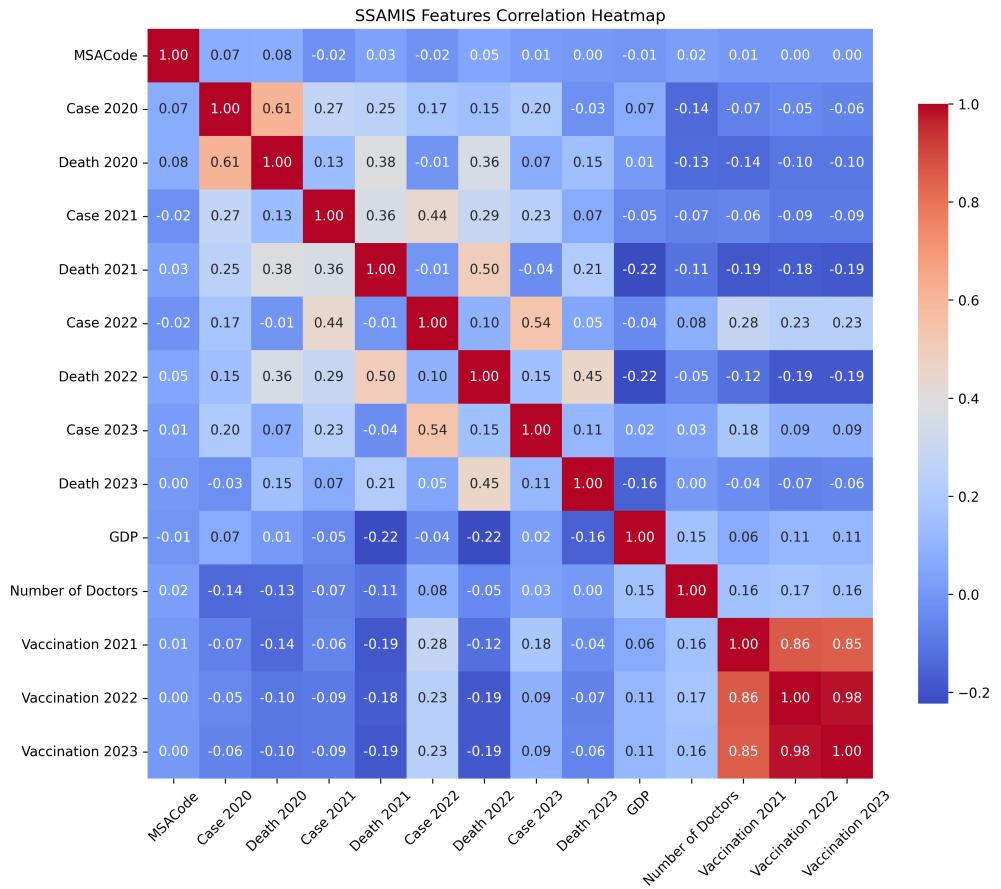


Figure 7: Correlation between SSAMIs Features.

Figure 7 represents the correlation coefficients between various SSAMIs across several years for cases, deaths, vaccinations, GDP, and the number of doctors. These coefficients range from -1 (perfect negative correlation) to +1 (perfect positive correlation), with 0 indicating no correlation.

Key observations from the heatmap include:

1. High Correlation Between Consecutive Years: There is a strong positive correlation between the same indicators across consecutive years (e.g., "Case 2020" with "Case 2021"). This consistency suggests that regions maintaining certain levels of COVID-19 cases, deaths, or vaccination rates tend to continue exhibiting similar levels in the following year, indicating persistent regional characteristics of effective continuity in public health response strategies.

2. Correlation Between Cases and Deaths: Moderate to high positive correlations exist between cases and deaths within the same year (e.g., "Case 2022" and "Death 2022"). This correlation underscores the direct impact of case surges on mortality rates, reflecting the immediate health outcomes of the pandemic's spread.

3. Vaccination Correlations: The vaccination rates for consecutive years show very high correlations, especially from 2021 to 2023 (e.g., "Vaccination 2021" and "Vaccination 2022"). High correlations in vaccination rates over the years may reflect ongoing vaccination campaigns and booster implementations which have continuity and sustained uptake among the population.

4. Low Correlation with GDP and Number of Doctors: Indicators such as GDP and the number of doctors show lower correlations with pandemic-specific indicators like cases and deaths, suggesting that while economic power and healthcare capacity are critical, they do not directly predict the immediate pandemic outcomes measured in cases and deaths within the same time frame.

This heatmap facilitates a deeper understanding of the interconnectedness of various health and economic factors in managing and responding to the pandemic. It highlights the importance of sustained public health measures and provides insights for policymakers on areas requiring continued focus or adjustment in strategy.

4.5 Cluster Classification

We apply the K-Means clustering technique to categorize MSAs and MicroSAs into three distinct risk levels based on SSAMIs for COVID-19 cases and deaths in the years 2020-2023. We then initiate the K-Means algorithm with three clusters, ensuring a consistent and unbiased seed for reproducibility through the 'random state' parameter set to zero. This parameter ensures that the clusters generated are the same every time the code is executed, which is crucial for reproducibility in scientific research.

Each MSA code, represented by its SSAMIs for cases and deaths, is assigned to one of three clusters — low, moderate, or high risk. This classification is inferred by the proximity of each point to the centroids of these clusters, calculated using the Euclidean distance, the default method in K-Means clustering. The centroids represent the average coordinates of all points in a cluster, effectively capturing the central trend of each risk group.

Post clustering, we map each cluster to a risk category based on the sorted order of centroids. This step translates the numerical labels into more understandable categories: low, moderate, and high risk. These categories are instrumental in visual analytics, where they are plotted to provide a visual representation of the distribution and intensity of risk across different regions. Such visualizations support intuitive understanding and are useful for stakeholders making public health decisions.

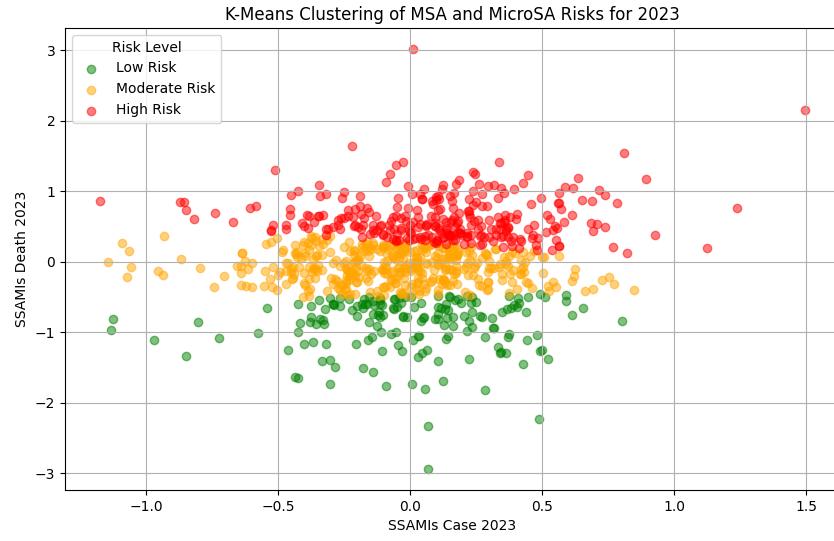


Figure 8: K-means clustering into Risk Levels for 2023.

The series of K-Means clustering results from 2020 to 2023 provides a visual evolution of the risk levels across Metropolitan and Micropolitan Statistical Areas. In 2020, the clustering reveals a high density of areas in the high-risk category, which can be associated with the initial and widespread impact of COVID-19. Over the subsequent years, the progression towards a larger distribution in the low-risk category by 2023 reflects the effectiveness of interventions and adaptive responses over time. This trend aligns with the documented improvements in public health strategies and the increased deployment of healthcare resources.

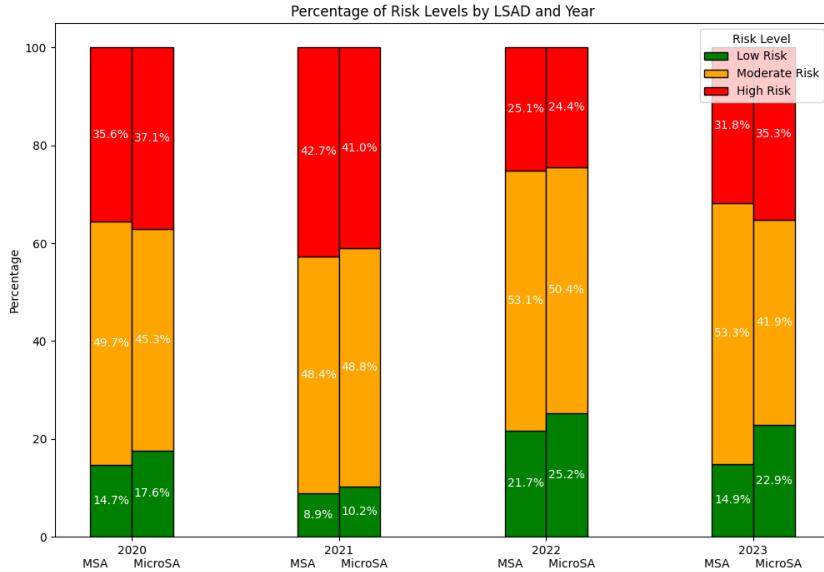


Figure 9: Risk Level Percentage for MSAs and MicroSAs from 2020 to 2023.

Figure 9 quantifies the transition of MSAs and MicroSAs through different risk categories over the four years. A notable shift from higher to lower risk categories is observable, especially in the increasing proportions of low-risk classifications from 2020 onwards. The marked reduction in high-risk categories by 2023 corroborates with the visual data provided by the K-Means clustering, underscoring the broad-scale improvement in managing health outcomes across various regions.

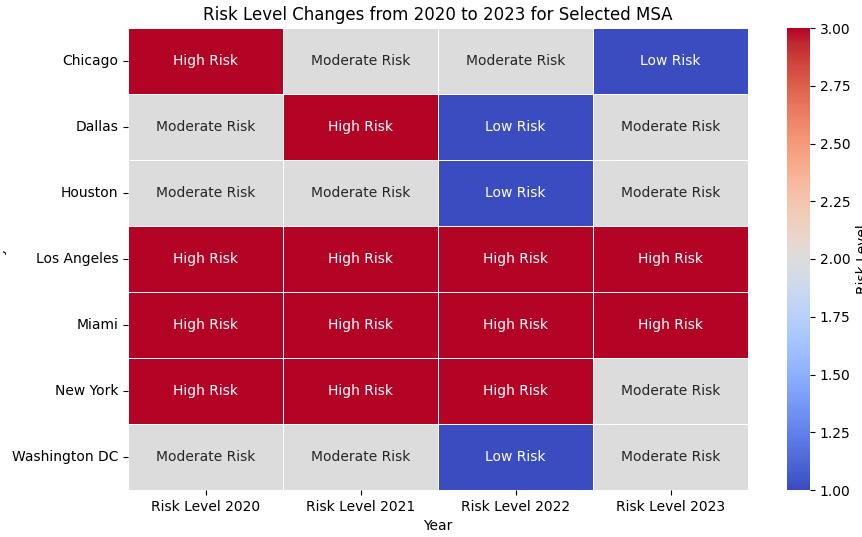


Figure 10: Change of Risk Level in Major MSA areas from 2020 to 2023.

Figure 10 offers a compelling snapshot of the changing risk landscapes across major MSA areas over four years. This visual representation highlights the dynamic nature of public health risks, revealing how cities like New York and Miami consistently sustained high risk levels, possibly due to their high population densities and international connectivity. In contrast, cities like Dallas and Washington D.C. show more varied transitions through risk levels, illustrating the impact of localized policy changes, health infrastructure improvements, or community compliance with health guidelines. The shifts in risk levels underscore the responsiveness of these regions to evolving epidemiological trends, highlighting the effectiveness of targeted public health interventions.

In summary, the cluster classifications over four years illustrate not only the dynamic nature of the pandemic's impact but also the adaptive responses of healthcare systems and policymakers. These visual and quantitative analyses offer a comprehensive overview of the changing landscape of health risks, providing valuable insights for future public health preparedness and response strategies.

5 Conclusion

This study investigates how the COVID-19 epidemic varied across different-sized cities in the United States from an urban scaling perspective. While the scaling law is commonly used for evaluating cities, many studies overlook spatial autocorrelation between them. Ignoring spatial autocorrelation can lead to misspecification of scaling exponents (β) and biased evaluations of city performance, as seen in section 4.3. Our results indicate that incorporating spatial models, such as the ESF model, improves the adjusted R^2 of the fitted linear scaling model estimations and slightly alters the scaling exponent compared to OLS models.

The Moran's I test on SAMIs reveals local clustering patterns, violating independence assumptions and potentially biasing city performance evaluations. Using ESF model residuals (SSAMIs) mitigates spatial autocorrelation effects, as demonstrated by Moran's I test. By analyzing the ranks from SAMI and SSAMI across various metropolitan and micropolitan statistical areas, we observe inconsistencies in how these indices evaluate performance. For example, SAMI sometimes overestimates or underestimates outcomes compared to SSAMI, demonstrating the need for careful interpretation of these metrics when assessing the efficacy of public health interventions in different regions. In essence, the SAMI of a city reflects not only its performance but also spatial trends, emphasizing the importance of accounting for spatial autocorrelation in urban evaluations.

Additionally, this study reveals the importance of understanding the relationship between COVID-19 dynamics and city size. By uncovering the non-linear scaling patterns, we can offer insights into the varying impacts of the pandemic across urban areas of different sizes. More specifically, higher infection rates are observed in larger cities as these exhibit superlinear scaling behavior, reflecting the challenges of controlling transmission in densely populated areas. Conversely, larger urban areas exhibit greater resilience in vaccination efforts (superlinear scaling) and lower mortality rates (sublinear scaling), highlighting the potential benefits of scale in healthcare infrastructure and resource allocation.

Lastly, the correlation and clustering analyses of SSAMIs can provide an understanding of urban health dynamics. Specifically, the clustering distinguishes cities by their evolving risk levels from 2020 to 2023, highlighting impactful shifts in cities like Dallas and Washington D.C., which reflect variations in public health responses. These analyses demonstrate the critical role of targeted interventions in urban areas, suggesting that effective public health strategies must consider both immediate correlations and longitudinal risk patterns to optimize responses to health crises. Overall, our findings emphasize the multifaceted nature of urban dynamics and the need for tailored approaches to address the complex challenges posed by infectious diseases like COVID-19. By integrating spatial models and accounting for spatial autocorrelation, we can enhance our understanding of urban epidemiology and inform more effective strategies for disease control and urban planning in the future.

References

- [1] Sahasranaman Anand and Bettencourt Luís M. A. Urban geography and scaling of contemporary indian cities. *Journal of the Royal Society Interface*, 16.
- [2] Luc Anselin. Local indicators of spatial association—lisa. *Geographical Analysis*, 27(2):93–115.
- [3] Luc Anselin. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267, 2002. Spatial Analysis for Agricultural Economists: Concepts, Topics, Tools and Examples.

- [4] D Bauman, T Drouet, M J Fortin, and S Dray. Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. *Ecology*, 99(10):2159–2166, 2018.
- [5] Luís MA Bettencourt. *Introduction to urban science: evidence and theory of cities as complex systems*. MIT Press, 2021.
- [6] U.S. Census. Zip code tabulation areas, 2020. Retrieved from <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html>.
- [7] U.S. Census. City and town population totals: 2020-2022, 2023. Retrieved from <https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html>.
- [8] U.S. Census. Delineating metropolitan and micropolitan statistical areas, 2023. Retrieved from <https://www.census.gov/programs-surveys/metro-micro/about.html>.
- [9] Johns Hopkins Coronavirus Resource Center. Covid-19 data repository, 2020 - 2023.
- [10] P. de Jong, C. Sprenger, and F. van Veen. On extreme values of moran's i and geary's c. *Geographical Analysis*, 16(1):17–24, 1984.
- [11] José Alexandre Felizola Diniz-Filho and Luis Mauricio Bini. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, 14(2):177–185, 2005.
- [12] Centers for Disease Control and Prevention. Covid-19 vaccinations in the united states, county, 2023.
- [13] Arthur Getis and Daniel A. Griffith. Comparative spatial filtering in regression analysis. *Geographical Analysis*, 34(2):130–140, 2002.
- [14] Bureau of Economic Analysis (BEA). Personal income by county, metro, and other areas, 2022.
- [15] OMB. 2020 standards for delineating core based statistical areas. *Federal Register*, 86(134), 2021.
- [16] Guillaume Blanchet Daniel Borcard Sylvie Clappe Guillaume Guenard Thibaut Jombart Guillaume Larocque Pierre Legendre Naima Madi Hélène H Wagner Aurélie Siberchicot Stéphane Dray ORCID iD, David Bauman. *adespatial*. R package version 0.3-23.
- [17] Yixiong Xiao and Peng Gong. Removing spatial autocorrelation in urban scaling analysis. *Cities*, 124:103600, 05 2022.
- [18] Gang Xu, Siyuan Zhang, Edwin McCulley, Ran Wu, Xinhua Li, and Limin Jiao. Evolving urban allometric scaling law of the covid-19 epidemic in the united kingdom. *Journal of Urban Management*, 03 2024.

Appendix A Scatterplot of SAMIs and SSAMIs Ranks in 2020, 2021 and 2022

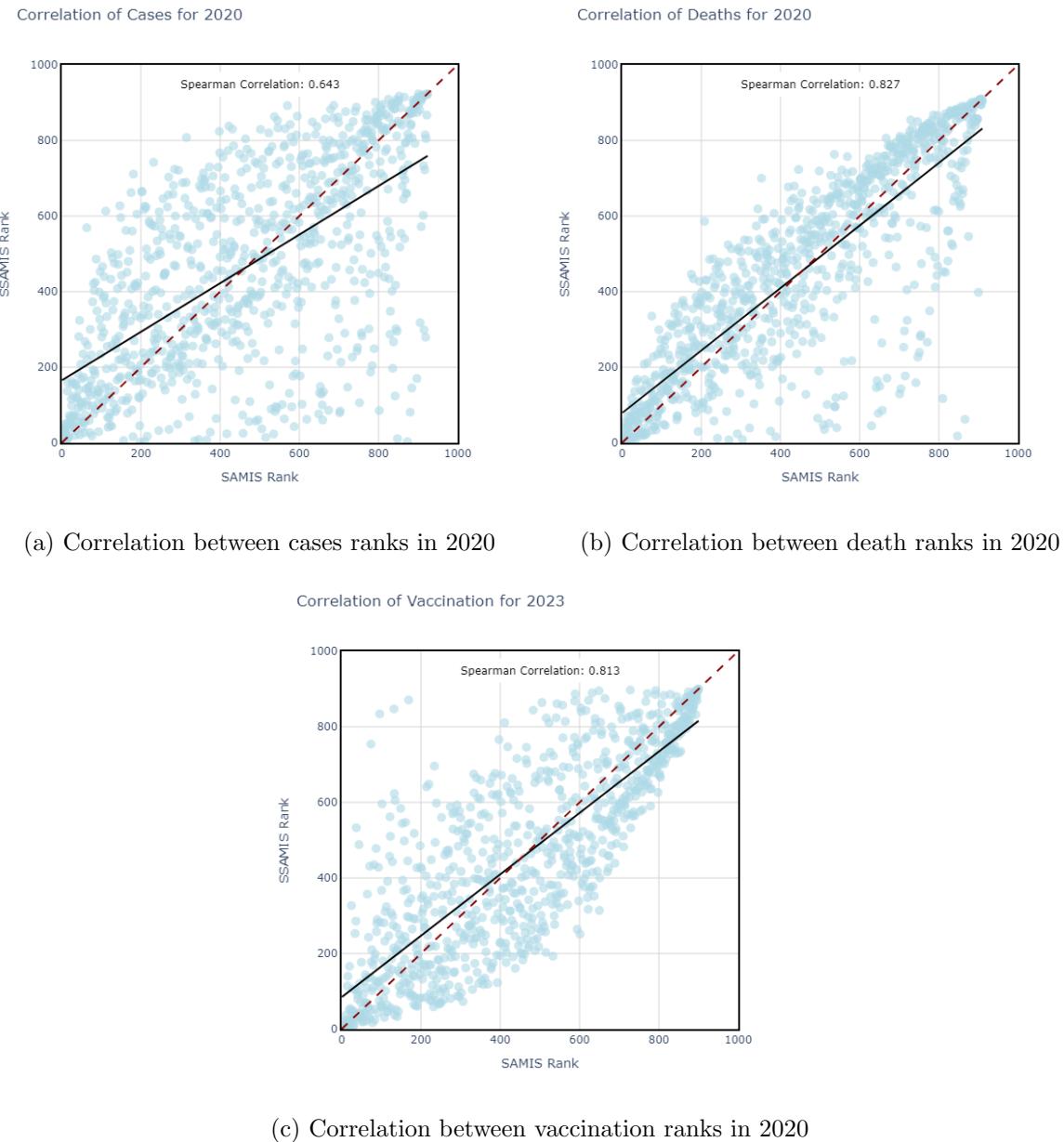
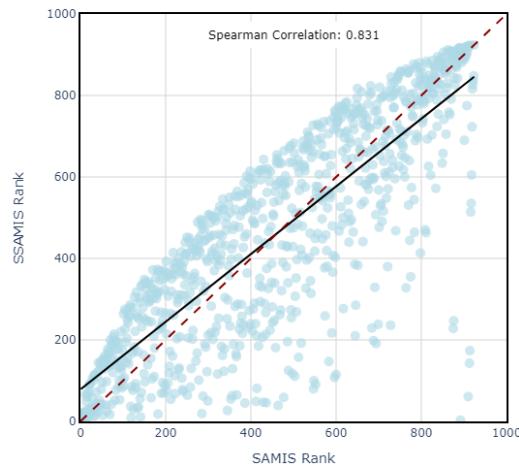


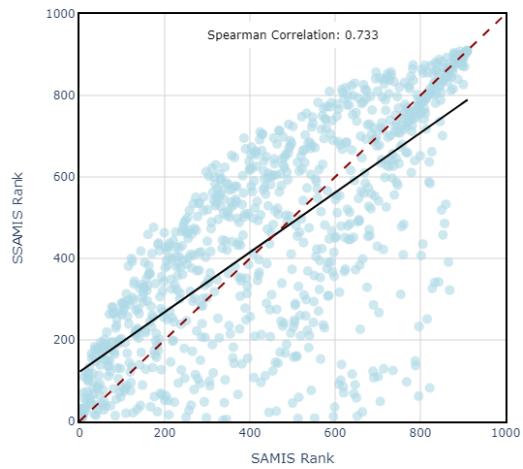
Figure 11: Scatterplot of SAMIs and SSAMIs Ranks in 2020.

Correlation of Cases for 2021



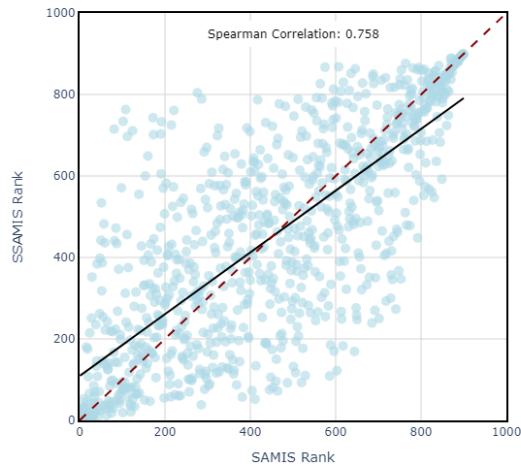
(a) Correlation between cases ranks in 2021

Correlation of Deaths for 2021



(b) Correlation between death ranks in 2021

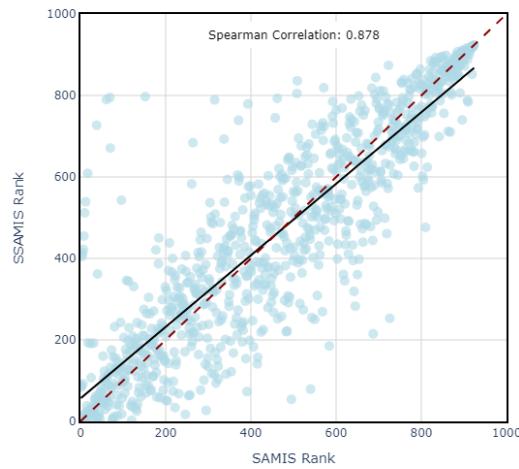
Correlation of Vaccination for 2021



(c) Correlation between vaccination ranks in 2021

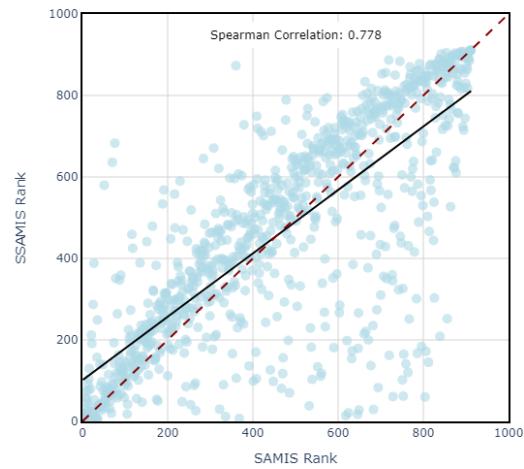
Figure 12: Scatterplot of SAMIs and SSAMIs Ranks in 2021.

Correlation of Cases for 2022



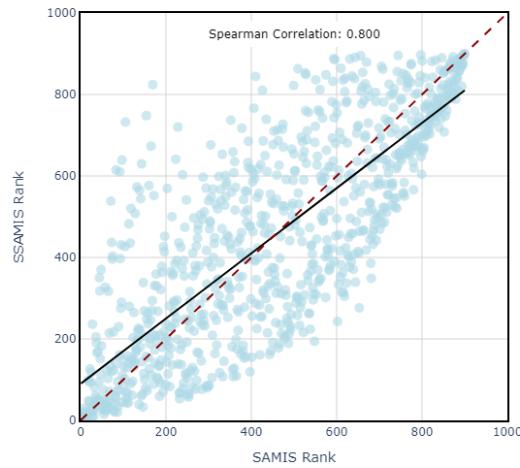
(a) Correlation between cases ranks in 2022

Correlation of Deaths for 2022



(b) Correlation between death ranks in 2022

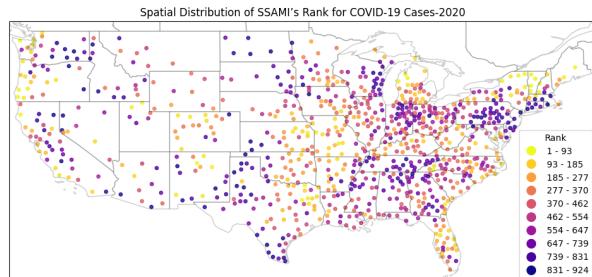
Correlation of Vaccination for 2022



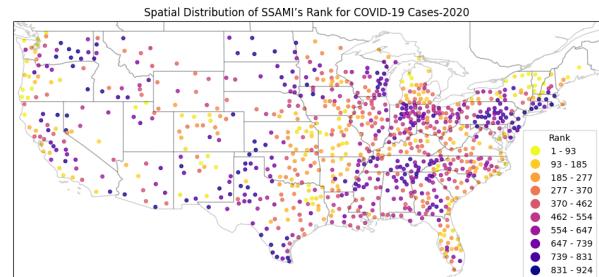
(c) Correlation between vaccination ranks in 2022

Figure 13: Scatterplot of SAMIs and SSAMIs Ranks in 2022.

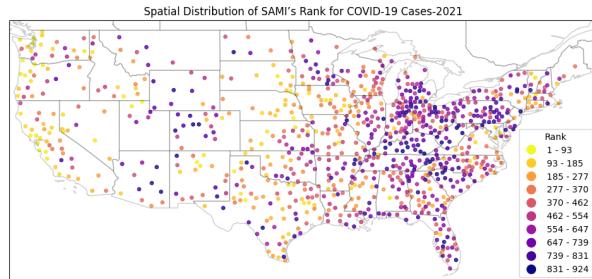
Appendix B Spatial distribution of SAMIs and SSAMIs ranks in 2020, 2021 and 2022.



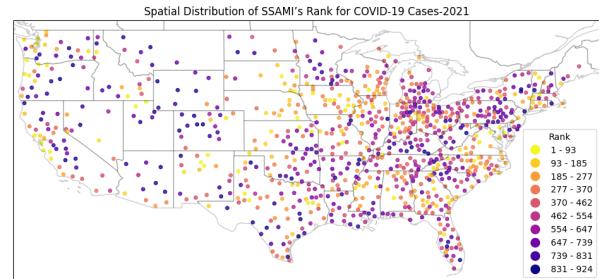
(a) SAMIs rank for cases in 2020



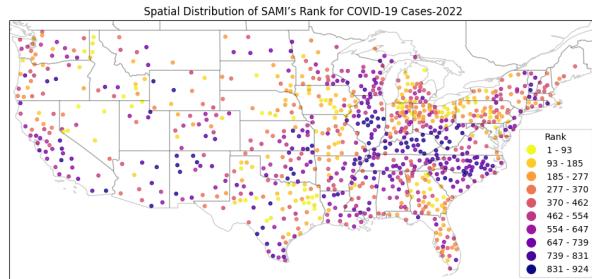
(b) SSAMIs rank for cases in 2020



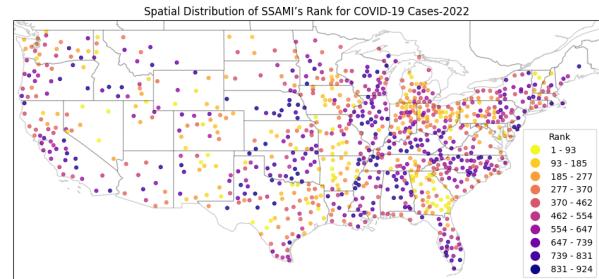
(c) SAMIs rank for cases in 2021



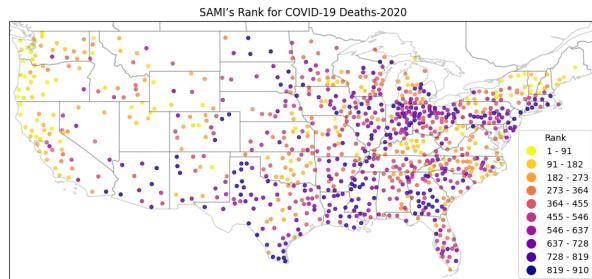
(d) SSAMIs rank for cases in 2021



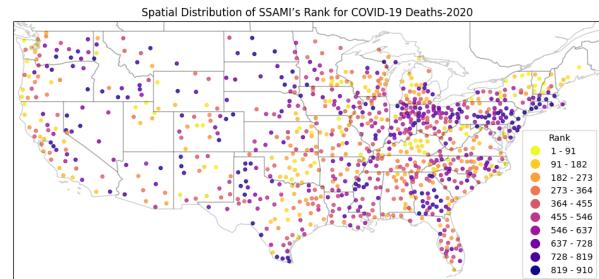
(e) SAMIs rank for cases in 2022



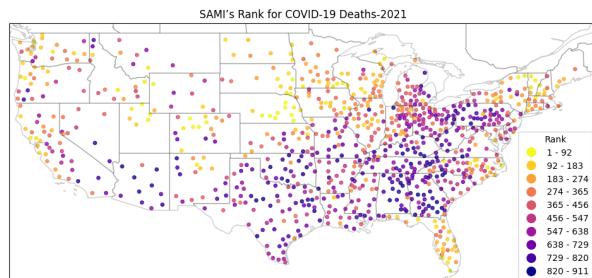
(f) SSAMIs rank for cases in 2022



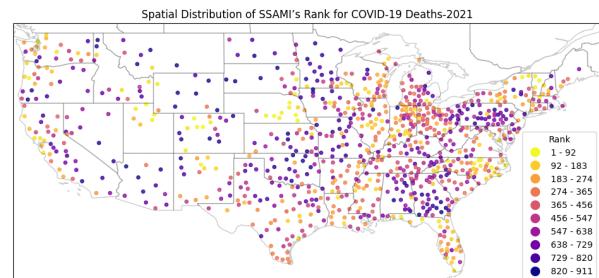
(g) SAMIs rank for deaths in 2020



(h) SSAMIs rank for deaths in 2020



(i) SAMIs rank for deaths in 2021



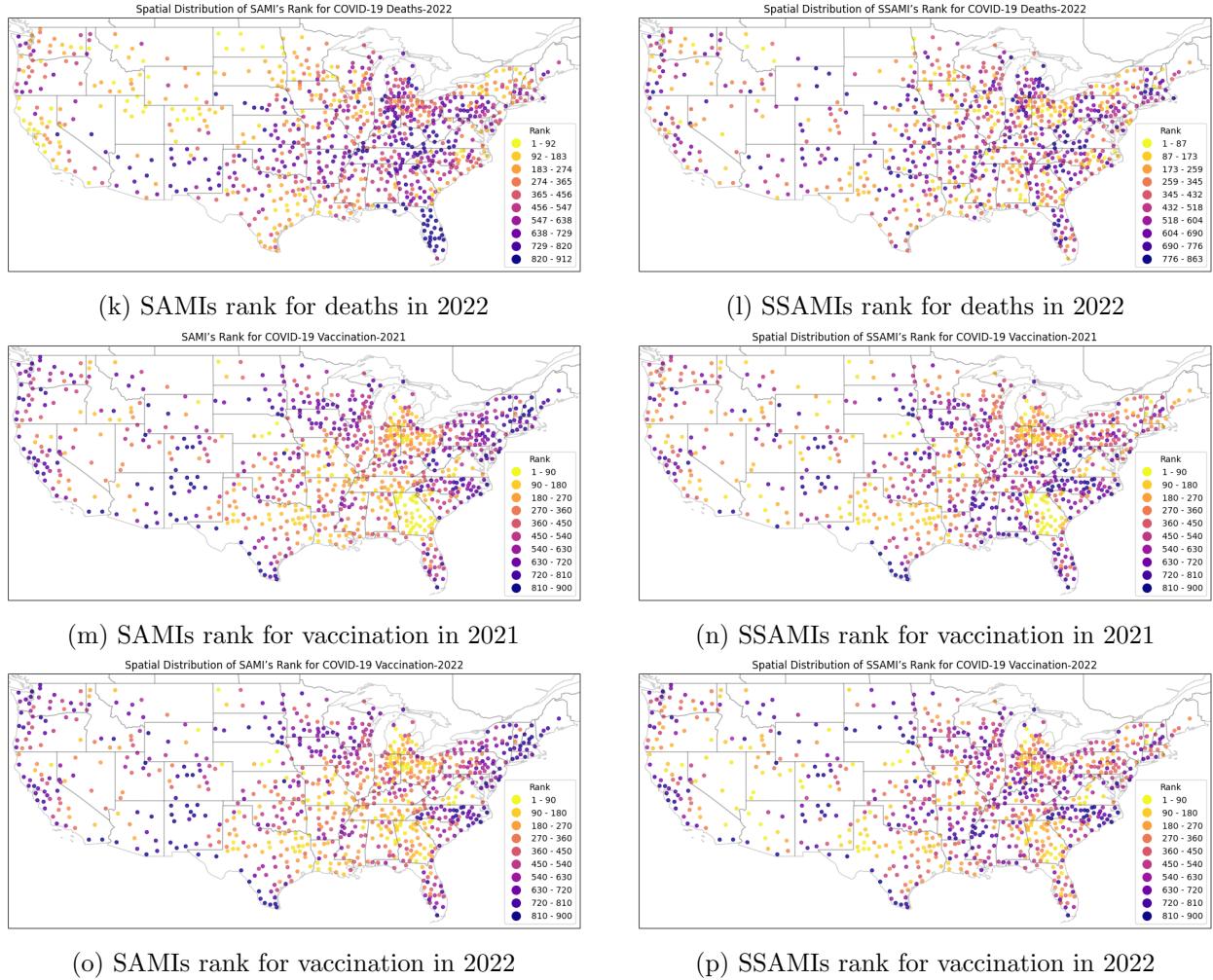
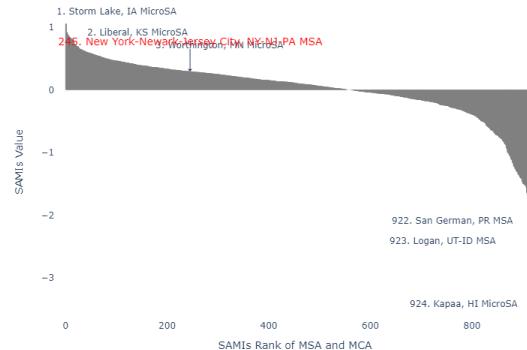


Figure 14: Spatial distribution of SAMIs and SSAMIs ranks.

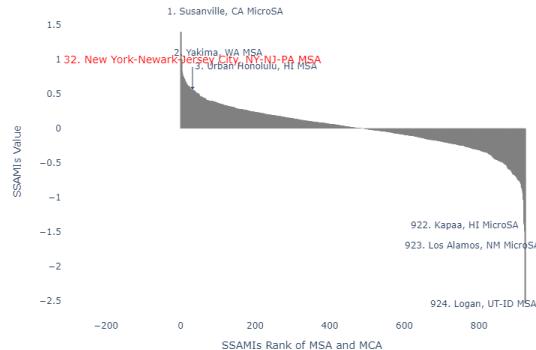
Appendix C Ranking of MSAs and MicroSAs by SAMIs and SSAMIs Figures

Cases SAMIs Ranks in 2020



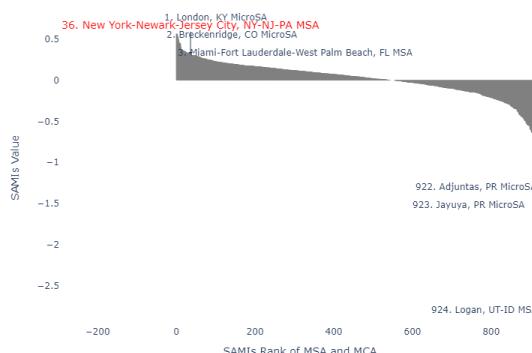
(a) Cases SAMI Ranks in 2020

Cases SSAMIs Ranks in 2020



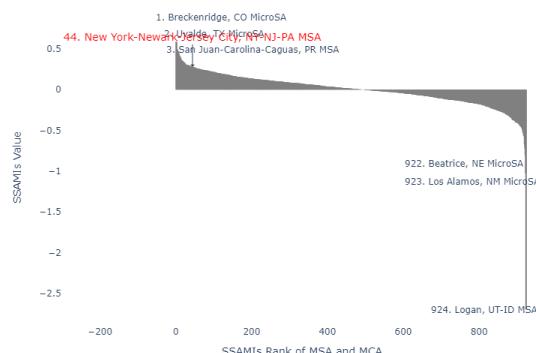
(b) Cases SSAMI Ranks in 2020

Cases SAMIs Ranks in 2021



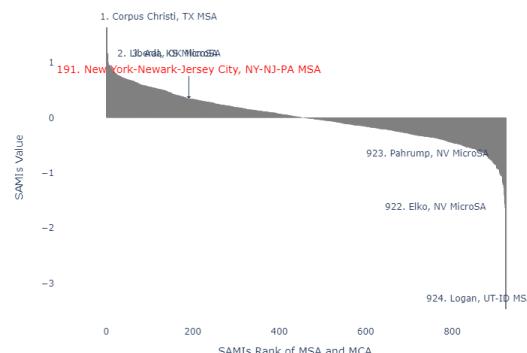
(c) Cases SAMI Ranks in 2021

Cases SSAMIs Ranks in 2021



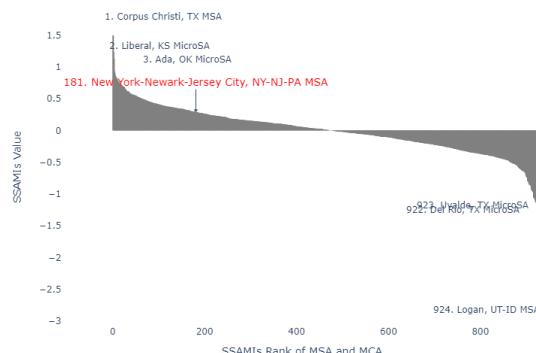
(d) Cases SSAMI Ranks in 2021

Cases SAMIs Ranks in 2022



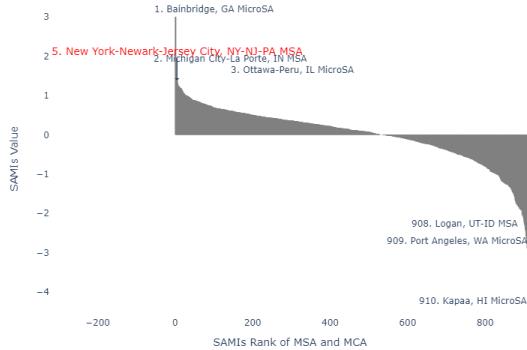
(e) Cases SAMI Ranks in 2022

Cases SSAMIs Ranks in 2022



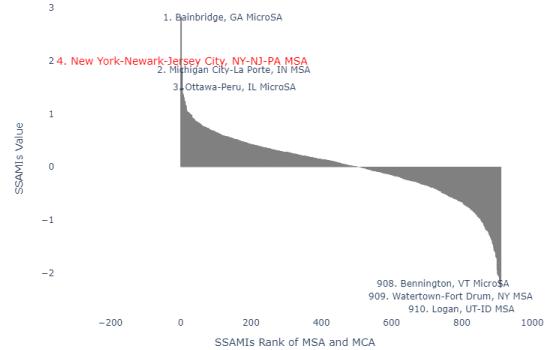
(f) Cases SSAMI Ranks in 2022

Deaths SAMIs Ranks in 2020



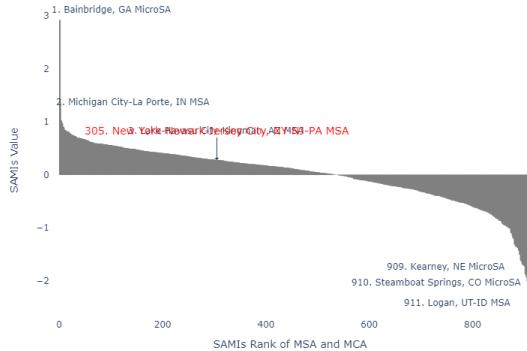
(g) Deaths SAMI Ranks in 2020

Deaths SSAMIs Ranks in 2020



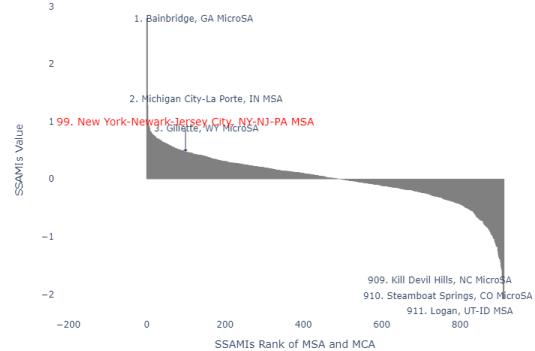
(h) Deaths SSAMI Ranks in 2020

Deaths SAMIs Ranks in 2021



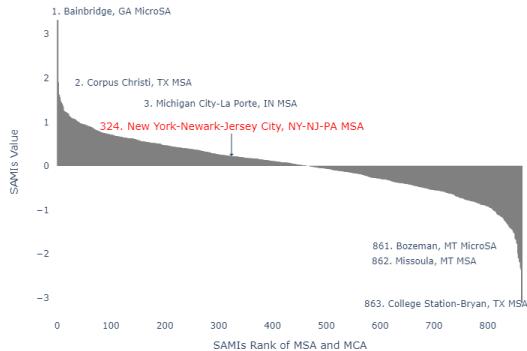
(i) Deaths SAMI Ranks in 2021

Deaths SSAMIs Ranks in 2021



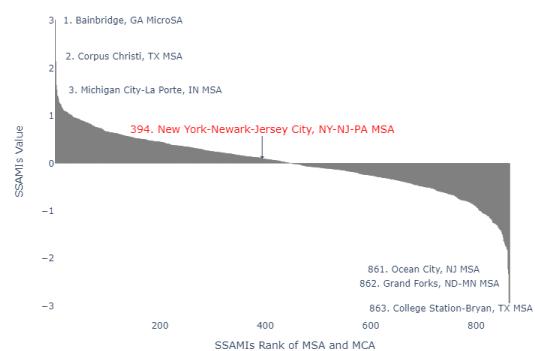
(j) Deaths SSAMI Ranks in 2021

Deaths SAMIs Ranks in 2022



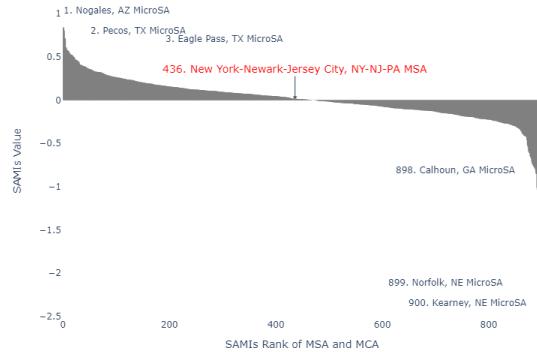
(k) Deaths SAMI Ranks in 2022

Deaths SSAMIs Ranks in 2022



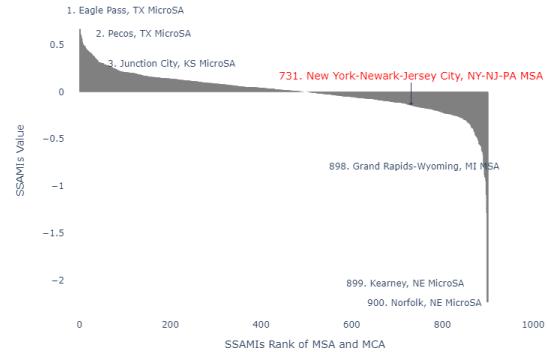
(l) Deaths SSAMI Ranks in 2022

Vaccination SAMIs Ranks in 2021



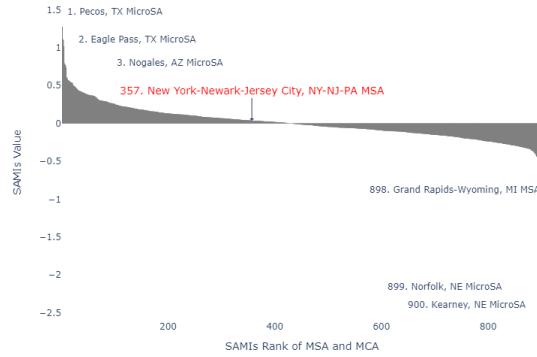
(m) Vaccinations SAMI Ranks in 2021

Vaccination SSAMIs Ranks in 2021



(n) Vaccinations SSAMI Ranks in 2021

Vaccination SAMIs Ranks in 2022



(o) Vaccinations SAMI Ranks in 2022

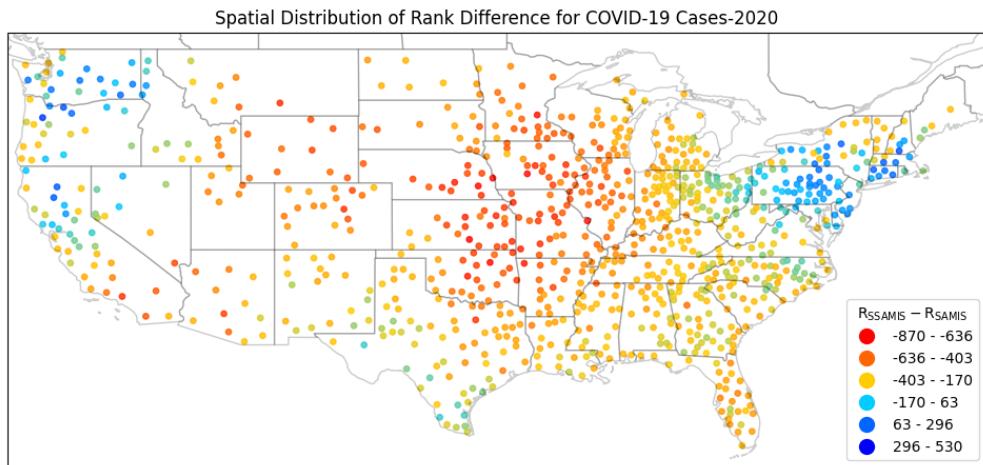
Vaccination SSAMIs Ranks in 2022



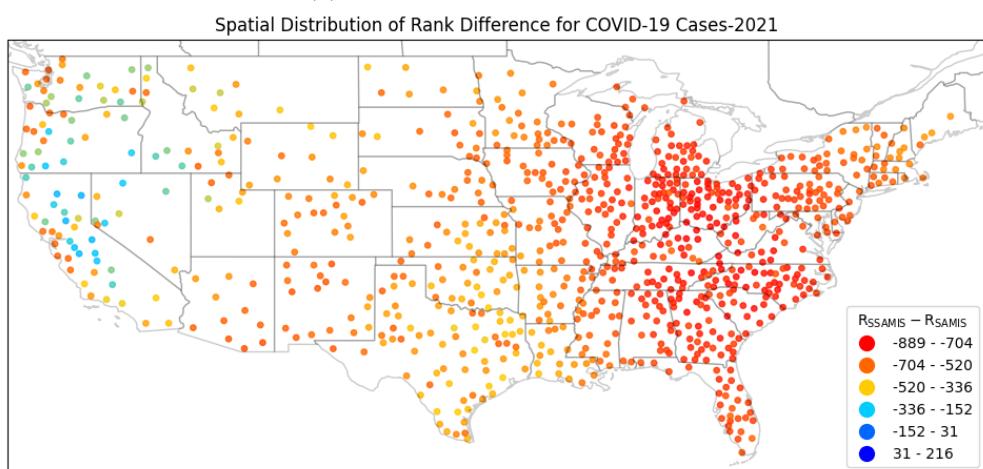
(p) Vaccination SSAMI Ranks in 2022

Figure 15: Ranking of MSAs and MicroSAs by SAMIs (left) and SSAMIs (right).

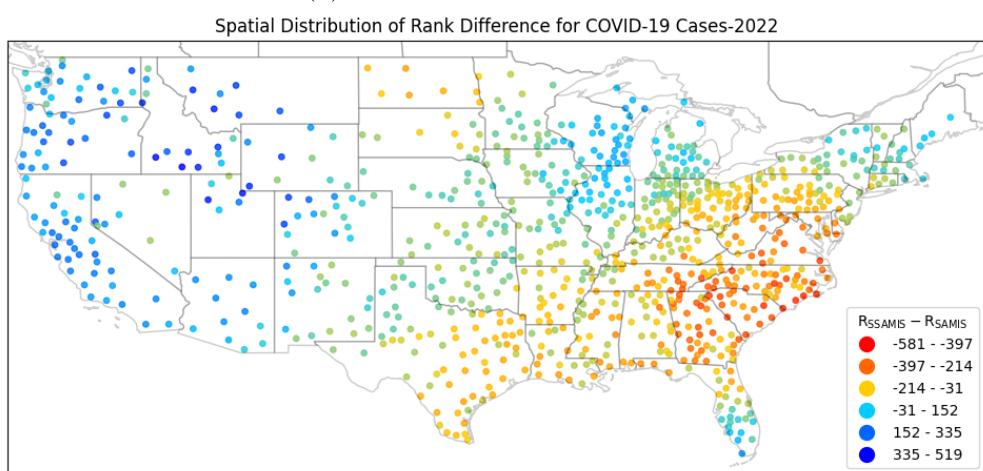
Appendix D Difference of City Ranks ($R_{SSAMI} - R_{SAMI}$)



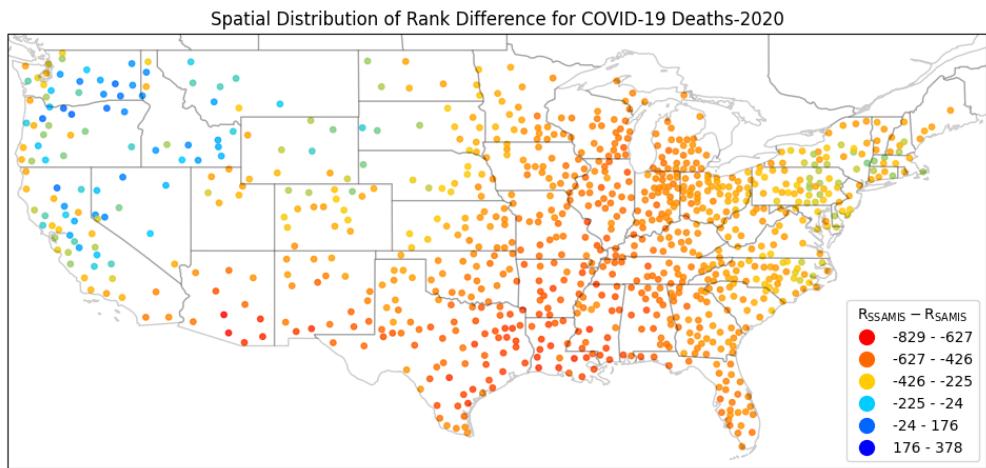
(a) Cases rank difference in 2020



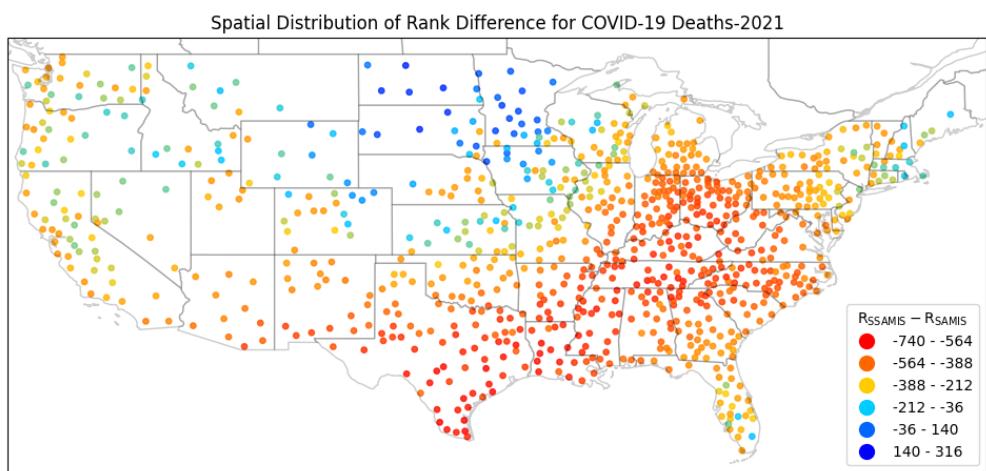
(b) Cases rank difference in 2021



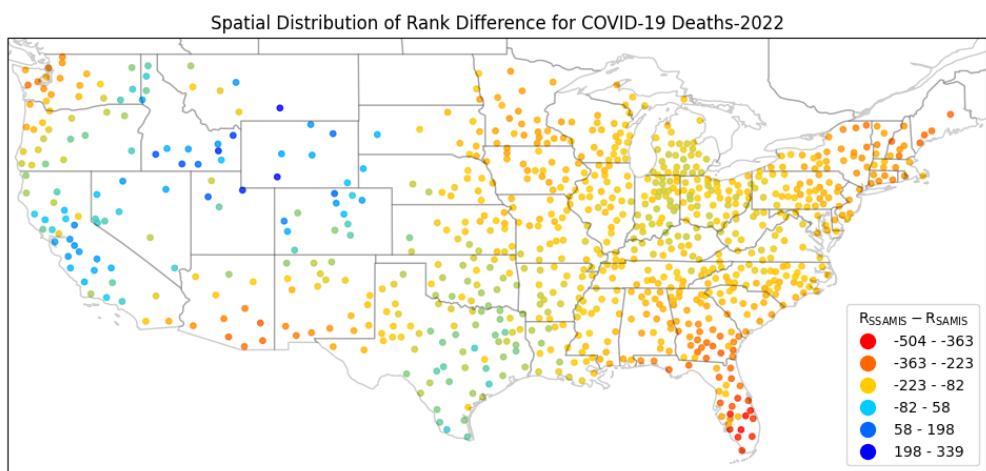
(c) Cases rank difference in 2022



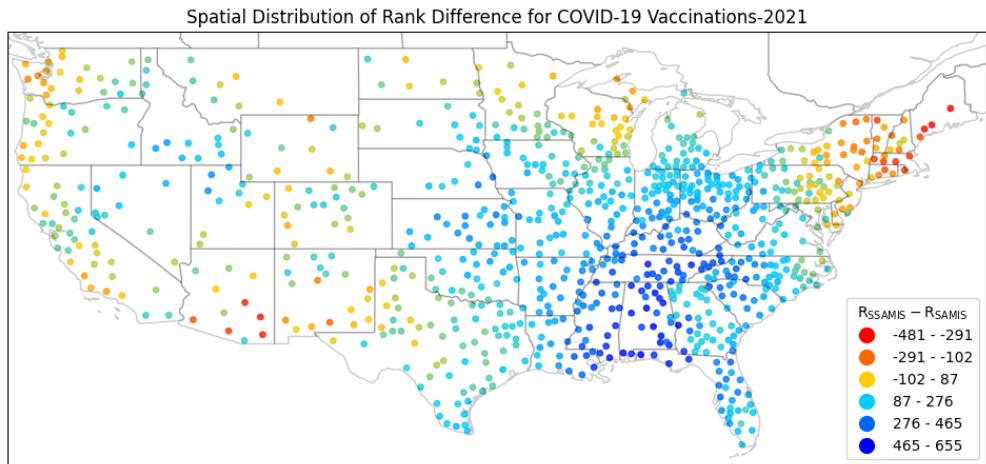
(d) Deaths rank difference in 2020



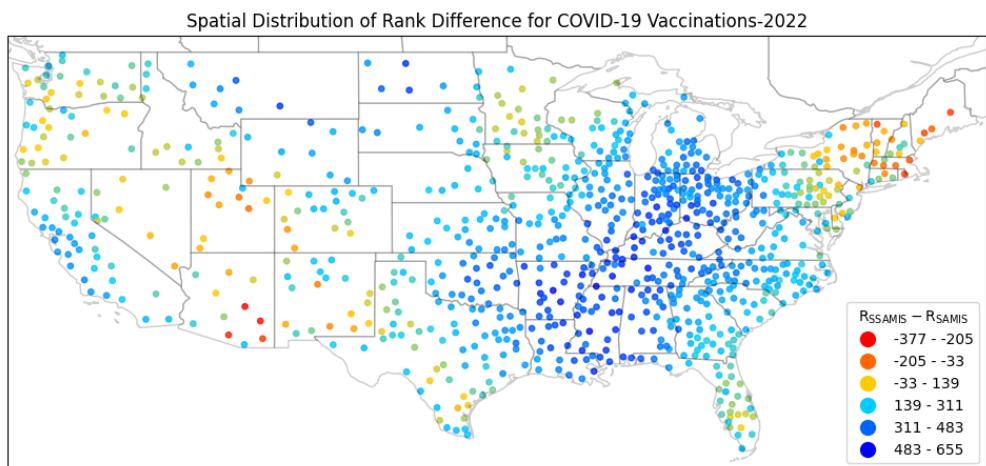
(e) Deaths rank difference in 2021



(f) Deaths rank difference in 2022



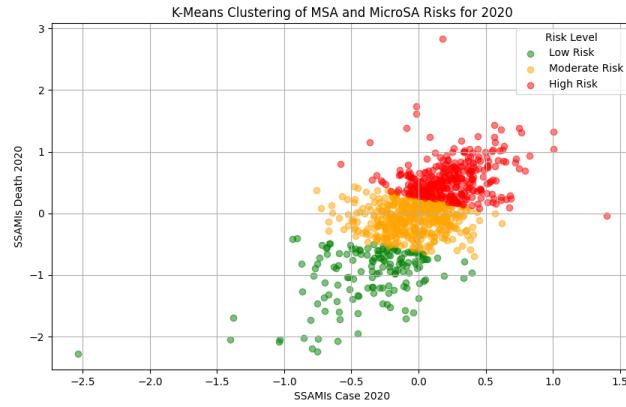
(g) Vaccinations rank difference in 2021



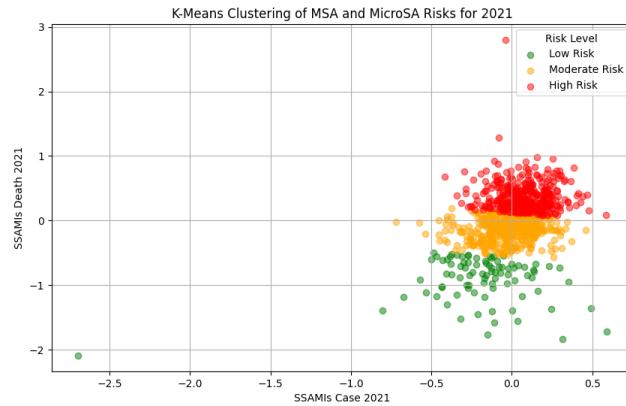
(h) Vaccinations rank difference in 2022

Figure 16: Difference of city ranks between two indicators ($R_{SSAMI} - R_{SAMIS}$).

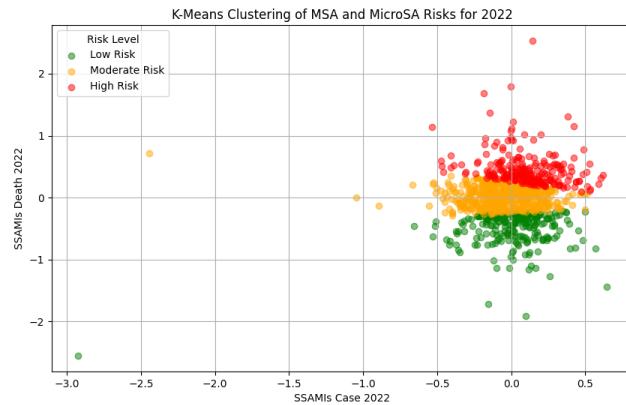
Appendix E K-Means Clustering



(a) K-Means Clustering of MSA and MicroSA Risks in 2020



(b) K-Means Clustering of MSA and MicroSA Risks in 2021



(c) K-Means Clustering of MSA and MicroSA Risks in 2022.

Figure 17: K-Means Clustering of MSA and MicroSA Risks.