

### **Introduction:**

A statistical analysis was performed to elucidate spatial patterns coinciding between the rate of positive covid cases reported and the diversity index in the New York City five borough region using the 2016 census tract data. This report aims to describe how the small multiples principle helps to highlight the relationship between diversity and COVID-19.

### **Methods:**

The principle of small multiples was coined by Edward Tufte as a solution for ease of readability of numerous related graphics or variables in order to make comparisons. Rather than depicting multivariate data in a single map, or displaying a series of maps on separate pages, Tufte's design of multiple, thumb-nail sized maps allow the reader view all the data at a glance. Though smaller maps require graphic simplification, differences in spatial patterns like the shape and spread of clustered data is more obvious when looking at small multiples.

A categorical bivariate map was created to compare positive COVID-19 cases and diversity (IDX) at a census tract level. The data is standardized by setting each respective median as the midpoint for the classification scheme. The data was categorized into four groups where each color corresponds to a 'High' or 'Low' occurrence of diversity and 'High' or 'Low' occurrence of positive COVID cases. The diversity index is set as the independent variable (x) and COVID rate is set as the dependent (y). Assuming there is a spatial correlation between the diversity and covid rates in census tracts, 'HH' and 'LL' occurrences are considered normal patterns, whereas data falling in the 'HL' and 'LH' categories should be considered anomalies and statistical testing should be performed to better understand the casualty and association between the variables at these geographic locations.

Once the data was categorized, a simple map was created for each category to highlight the distribution of data, then formatted into quadrants so they could be viewed together. Small multiples are appropriate for categorical mapping because the data is comparing more than one variable at a small scale where multiple comparisons need to be made between each category.

The unique scheme of contrasting colors used for each partition of data becomes confusing when viewed all on the same map. Using small multiples allows the reader to better discern spatial patterns in the data. Patterns of spatial heterogeneity and outliers in the data are especially hard to discern on a single bivariate categorical map where multiple colors inhibit data from being highlighted.

### **Discussion:**

Small clusters low in diversity and COVID cases (bright green) can be seen in Brooklyn, Manhattan and western queens, while large clusters high in diversity and COVID cases (bright orange) are in north Staten Island, the Bronx and form a corridor through the middle of Queens. Large clusters low in diversity with high covid rates (dull green) take up most of Staten Island, east Brooklyn, and east Queens, with smaller clusters in west Queens and the Bronx. Areas high in diversity with low covid cases (dull orange) are dappled throughout west Brooklyn, upper Manhattan, Queens and the Bronx.

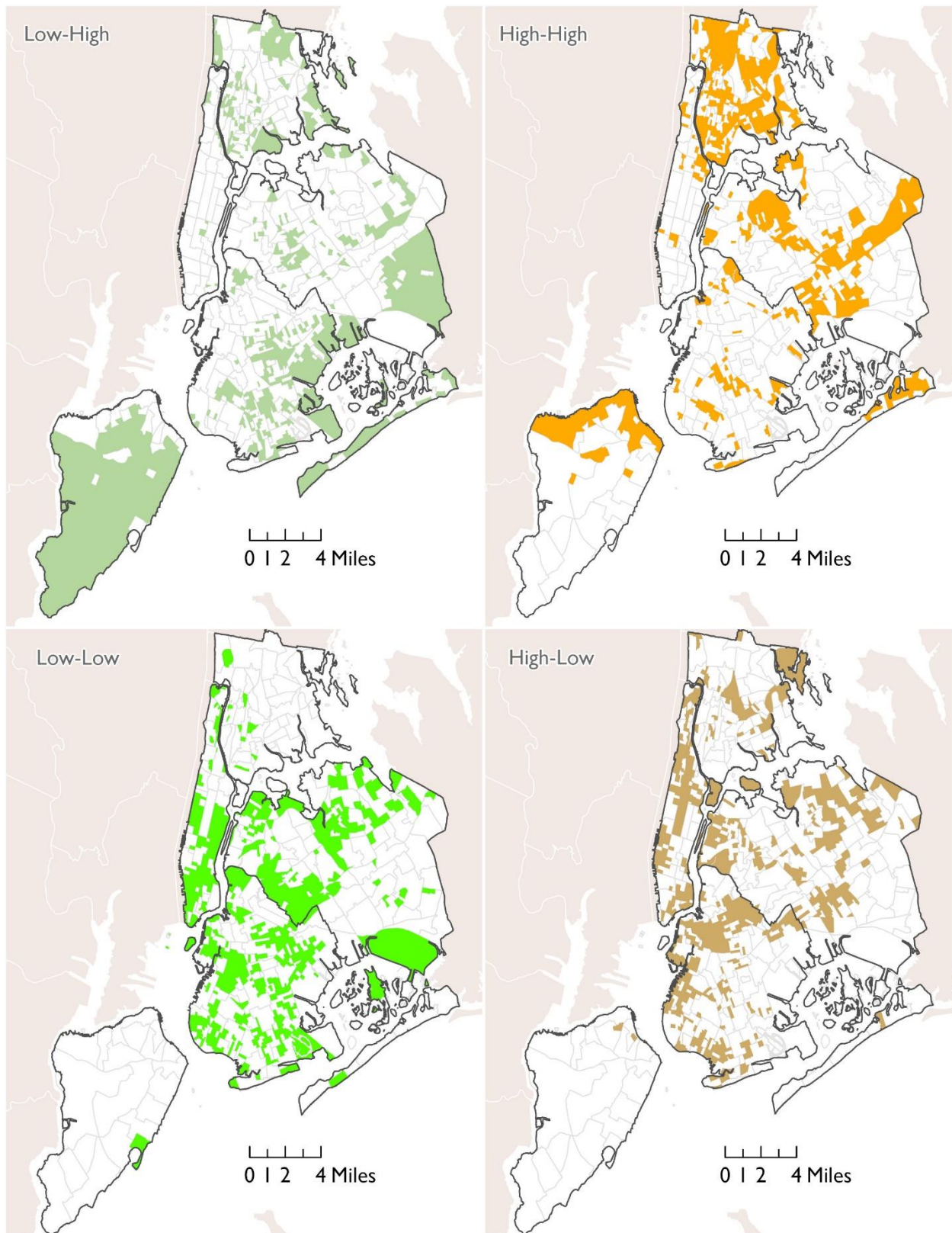
Staten Island appears to be the most homogenous in terms of categorization, where the data points to overall high covid cases and low diversity. Brooklyn neighborhoods are overall lower in diversity with some areas, but covid rates are relatively spatially heterogenous. East Brooklyn shows a fairly homogenous pattern of low integration and high covid cases. The Bronx is predominantly highly diversity with high positive covid cases. Manhattan has overall low covid cases and with higher diversity

uptown. And finally, Queens is the most spatially heterogenous of all the boroughs, containing clusters of all four categories, though many are high diversity clusters.

The limitation of using bivariate categories is due to the broad classification of the data, where data can only fall into four categories. This classification results in the loss of information about the range and distribution of values of both variables, which in effect disappears the gradient changes between the cores and the edges of the clusters.

### **Conclusion:**

This report describes the principle of small multiples and how it can be useful in portraying multivariate categorical data analysis. Small multiples is a helpful visualization design for categorical bivariate data because it allows for side-by-side comparisons, however there are drawbacks to looking at the data at a small scale.



**COVID and Diversity in NYC Small Multiples** – This series of bivariate maps depicts positive COVID-19 cases reported in the NYC-5 borough region as it relates to the diversity index per census tract. Data was obtained from the U.S. Census Bureau. Each quadrant shows a bivariate category where there are high or low occurrences of diversity and positive COVID cases. For example, the ‘High-Low’ quadrant represents tracts where there is a high occurrence of diversity and a low occurrence of covid cases.

