Examining the Correlation Between Social Variables and COVID-19 Morbidity in Minnesota at the Zip Code Tabulation Areas (ZCTA)s Level

Ulisses Jacobo
Department of Geography and Planning, School of Public Affairs, St. Cloud State University

Introduction

The COVID-19 pandemic has devastated economic and medical infrastructures, created political divisions, and has left many asking where do we go from here? As of April 6,, 2021 the World Health Organization has reported 131 million confirmed cases and 2.8 million deaths worldwide (WHO COVID-19 Dashboard). It has become evident-that COVID-19 is a socially driven disease, and, notwithstanding the rapid development and subsequent current rollout of vaccines, the best interventions for primary prevention of transmission during most of 2020 remained social and behavioral (De Ver Dye et al. 2020). In the US, state governments were charged with establishing preventative measures to reduce the spread of the virus. It is therefore instructive to explore data at the state level to examine geographical patterns of COVID-19 morbidity and explore possible connections to sociodemographic factors. Using data from the State of Minnesota, this project draws on work in medical geography, epidemiology, demography, and Geographic Information Systems (GIS)-to answer the following questions:

- Where are COVID-19 morbidity case rates concentrated in Minnesota at the zip code level?
- Where are COVID-19 hot/cold clusters in Minnesota at the zip code level?
- What socio-demographic characteristics have a statistically significant association with COVID-19 rates?

Methodology

Using available 2019 US Census Zip Code Tabulation Areas (ZCTAs) and monthly morbidity data at the Zip Code level between April and December 2020 from the Minnesota Health Department, I calculated morbidity case rates per 10,000 people in zip codes that had a population greater than 200. I utilized ESRI ArcMap Program tools Spatial Auto Correlation (Morans I), and Hotspot Analysis (Getis Ord Gi*) to identify and cartographically depict overall and monthly case rates (Figure 1 & 5) and identify Hot/Cold clusters (Figure 2) at the zip code level. Once statistical clusters and rates were determined, the next step was to take social demographic data collected from the US Census to fit a linear model comparing overall morbidity rates and certain demographic data. I utilized JMP software to identify any socio-demographic data associations to COVID-19 morbidity rates.

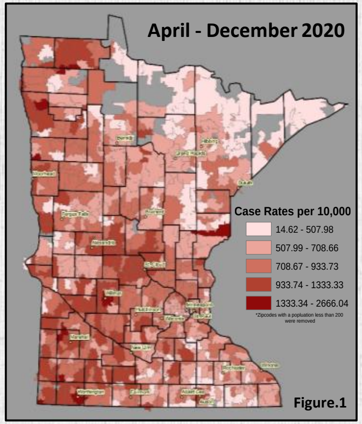


Figure.1

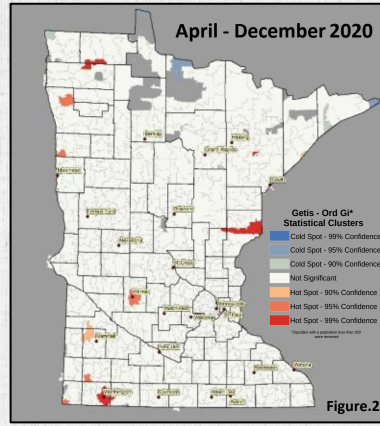


Figure.2

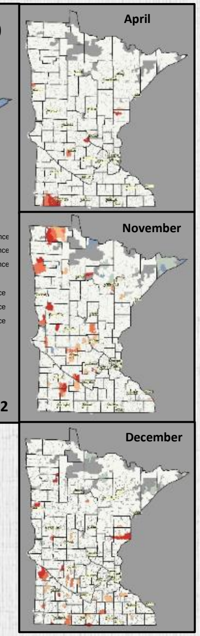
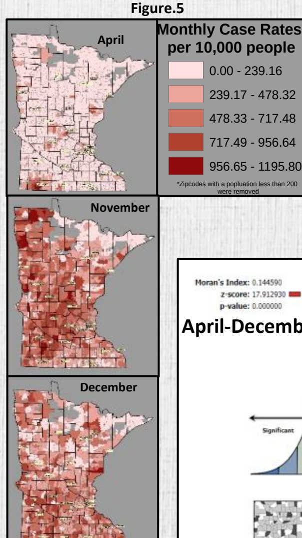


Figure.6

Figure.5



Parameter Estimates Figure.7

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	738.66453	15.77857	46.81	<.0001*
Higher Ed yrs 25+	-2266.29	4432.441	-5.11	<.0001*
Log[Percent_H_L]	60.39196	10.54056	5.73	<.0001*
Log[Black_Percent]	23.243327	8.452218	2.75	0.0062*

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.726582	0.951569	-1.81	0.0700
longitude	-0.087916	0.010103	-8.70	<.0001*

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.8211326	0.474263	18.60	<.0001*
latitude	-0.049774	0.010404	-4.78	<.0001*

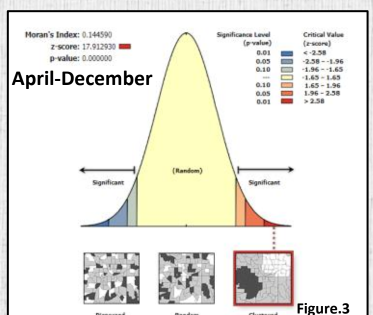


Figure.3

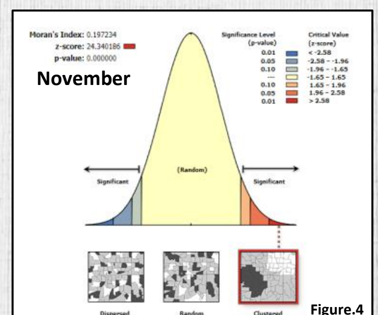


Figure.4

Results and Analysis

Data shows overall COVID-19 morbidity case rates to be higher in western and southwestern Minnesota with some outliers present in the eastern area. which were identified by utilizing latitude and longitude values with case rates in a linear model that provided strong p-values(Figure 1 & 7). Monthly morbidity case rates for April – December depict a generally increasing trend, with November presenting the highest case rates throughout the data timeframe. Interestingly, while April experienced higher case rates in more populated areas predominately in the south western region and lower-case rates in rural areas, by December case rates appear to have reduced slightly in the central, northwestern, and southwestern regions of the state with moderate to higher case rates in more rural areas (Figure 5).

There are few sustained hot/cold clusters for the overall study timeframe (Figure 2). April's COVID-19 hot/cold cluster map depicts multiple zip codes as hot spots in the state but with a more dominant presence in the south western area. With November's notably high morbidity rates, it interestingly presented the most polarization of hot/cold spatial clusters compared to the rest of the study timeframe. For the month of December hot spot spatial clusters have a more dominant scattered presence in the southern zip code areas of the state with some outliers in the northern areas (Figure 6)

Spatial Autocorrelation (Globan Moran's I) was conducted for overall and November case rates to comprehend if the values are dispersed, or clustered. (Figures 3 and 4) Initially there were six different social demographic variables inputted into a linear fit model. As the model presented variables with high p-values they were removed. Three linear models were developed with the third model presenting the following demographic data presenting statistically significant p-values Higher Education which is a calculated percentage of Bachelors and Masters or Professional degrees of a population with an age of 25+, Hispanic/Latino Population Percentage (of any race), and Black or African American Population Percentage (alone). (Figure 7)

Conclusions

Overall COVID-19 case rates are concentrated in the Southwestern, Central, and Northwestern areas with a general increase in rates from North to South and a decrease in case rates from West to East. There are also few statistically significant hot/cold spot clusters present within the study area for the overall timeframe which is due to COVID-19 being so widespread. Monthly case rates and hot/cold spot polarization are at a high peak during the month of November along with identifiable momentary hot/cold spots that revert to having no statistical significance for monthly data. Having conducted spatial auto correlation there is a less than 1% likelihood that the clustered patterns could be the result of random chance for overall case rates and the month of November. There is a negative correlation regarding Higher education and a Positive correlation regarding Hispanic/Latino and Black or African American Population percentages. This study is an initial exploration of the relationship between COVID-19 case rates, hot/cold spots, and statistical associations and provides a baseline for future study of patterns COVID-19 and the various factors that influenced its spread.