# Sociodemographic Factors, Immunization Sites, and COVID-19 Cases: A Spatial Analysis of Toronto's Neighborhoods

Ahmed Irshad[1], Kelvin Cho[1], Temoor Tayyab[1]

[1] Dalla Lana School of Public Health, University of Toronto

## Introduction

The COVID-19 pandemic has had a profound global impact, resulting in substantial excess mortality and placing immense strain on health systems worldwide.[1] As of April 2024, cumulative reports have documented over 700 million COVID-19 cases globally and more than 7 million deaths.[2] In Canada, cumulative reports have documented nearly 5 million cases and nearly 60,000 deaths.[3] Within Toronto, the country's largest, most diverse city and economic capital, the effects of COVID have been particularly pronounced.[4] As a densely populated urban center, Toronto has experienced 400,000 cases, and faces its own unique challenges in curbing transmission and ensuring equitable access to resources such as immunization clinics, testing sites, and healthcare services.[5]

Emerging research has suggested that COVID-19 disproportionately impacted certain socio-demographic groups in the city, reflecting broader health inequities.[4] As such, there is an interest in better understanding where these disparities exist and which groups might be more at risk of transmission. However, existing research often overlooks the spatial dimension of these disparities, particularly how they manifest at the neighborhood level within cities like Toronto. This gap in understanding calls for a closer examination of how sociodemographic factors and resource availability—such as the number and accessibility of immunization sites—interact with COVID-19 case rates across different areas. Addressing this gap could offer crucial insights into patterns of vulnerability and resilience during public health crises.

This study leverages Toronto's robust datasets, including nearly 400,000 COVID-19 cases and detailed census data, to conduct a granular spatial analysis at the neighborhood level. By exploring the spatial distribution of COVID-19 cases between January 2020 and February 2023, this research aims to identify potential hotspots and disparities tied to sociodemographic factors and resource allocation. The findings from this research will contribute to a deeper understanding of the social determinants of health during the pandemic, potentially guiding future studies and informing public health strategies for more equitable resource allocation and improved urban infrastructure for future public health challenges. Ultimately, this work lays the foundation for ongoing efforts to address health disparities and improve outcomes in diverse urban populations.

## Methods

All data was obtained from the City of Toronto's Open Data Portal.[5,6,7,8,9] Five total datasets were downloaded and imported into R for data cleaning. These included a boundary shapefile of Toronto by its 140 neighbourhoods, a shapefile of COVID-19 immunization sites as point data, a shapefile of COVID-19 testing sites as point data, a tabular dataset of 2016 Census data by Toronto's 140 neighbourhoods, and a tabular dataset of all COVID-19 cases since January 2020 by Toronto's 140 neighbourhoods.

The COVID-19 case data was filtered to include cases from January 2020 to February 15, 2023. This cutoff point was established since data on active immunization and testing sites was last updated on this date. The total number of immunization sites was summed and aggregated to get the total number of active immunization sites by each of Toronto's 140 neighbourhoods. Similarly, the total number of COVID-19 cases within the filtered time period were summed and aggregated to get total cases for each of Toronto's 140 neighbourhoods. There were 392,204 total cases. 12534 of these cases were missing information on which neighbourhood the cases occurred in. This made up ~3% of the data and was considered to be negligible given the large sample size and, as such, these cases were excluded from the analysis. Thus, the final sample size of included cases was 379,670. This data on total cases and total immunization sites was then merged with the census dataset, which had previously been merged with the Toronto neighbourhood boundary shapefile. This resulted in a final shapefile with all census data, the total number of COVID-19 cases, and the total number of COVID-19 immunization sites by each Toronto neighbourhood.

Using an a priori approach, we identified variables from the final dataset that were considered to be theoretically or logically relevant to include in descriptive analyses and model building. This selection was not based on statistical significance but rather on variables we believed could be interesting or potentially connected to the outcome of total COVID-19 cases. These variables included: the percentage of individuals with low income, the unemployment rate (%), population density per square kilometre, the percentage of individuals with a Bachelor's degree or higher, the percentage of individuals belonging to a visible minority group, the percentage of individuals living alone, the percentage of males, the percentage of non-English or non-French speakers, and the total number of immunization sites. We cleaned and created these variables as part of the data preparation process.

To address potential multicollinearity issues, a Variance Inflation Factor (VIF) analysis was done with a threshold value of 5. Any variables exceeding this threshold were excluded from further analysis. From this refined dataset, we created choropleth maps to visualize the distribution of the outcome variable—total COVID-19 cases—as well as each of the identified exposure variables. Additional maps were generated to explore areas where low immunization test site density intersected with COVID-19 case counts and other variables of interest related to population. Buffers techniques were also used with immunization and test sites to map the radius accordingly on choropleth maps. Spatial analysis techniques were also used to investigate patterns in the data. Global Moran's I was used to assess spatial autocorrelation, while the Getis-Ord statistic and permutation testing helped identify hotspot and coldspot areas for total COVID-19 cases by neighbourhood. Finally, geographically weighted regression was conducted to quantify the relationship between covariates and total COVID-19 cases, allowing us to examine how these relationships varied spatially across neighbourhoods.

# Results

## Descriptive Analysis

Choropleth Map of Total COVID-19 Cases from Jan 2020 to Feb 2023 by Toronto Neighbourhood

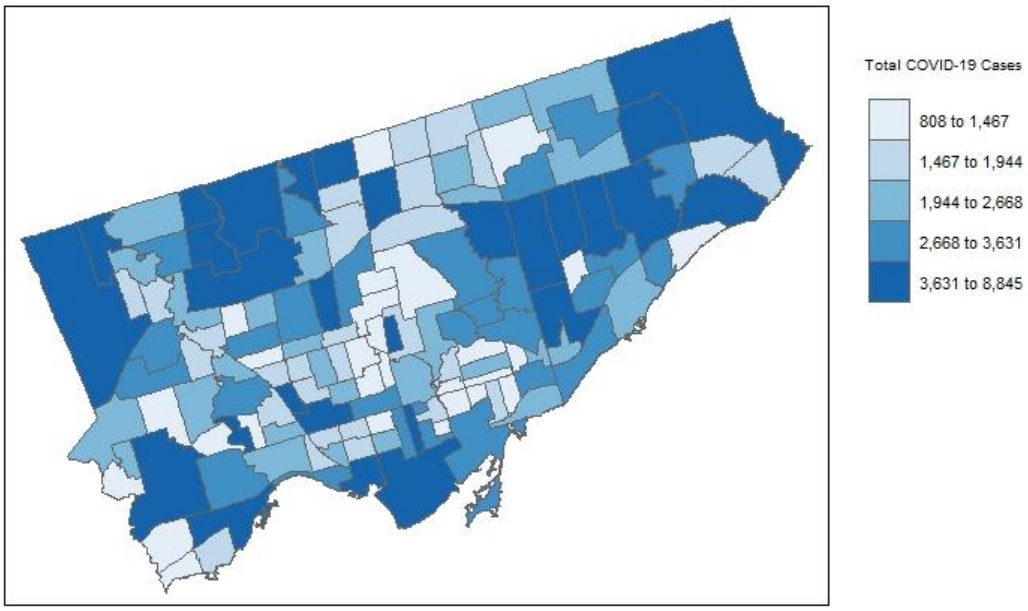
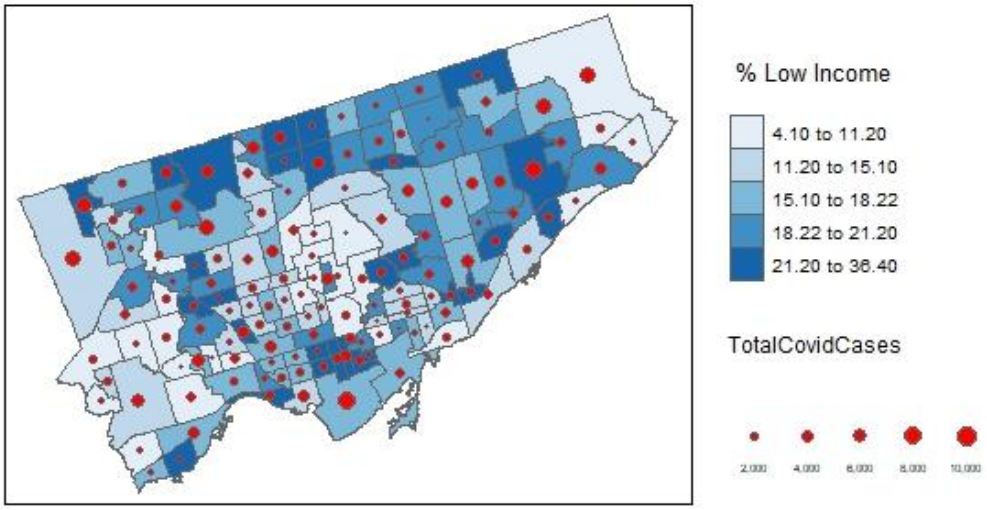
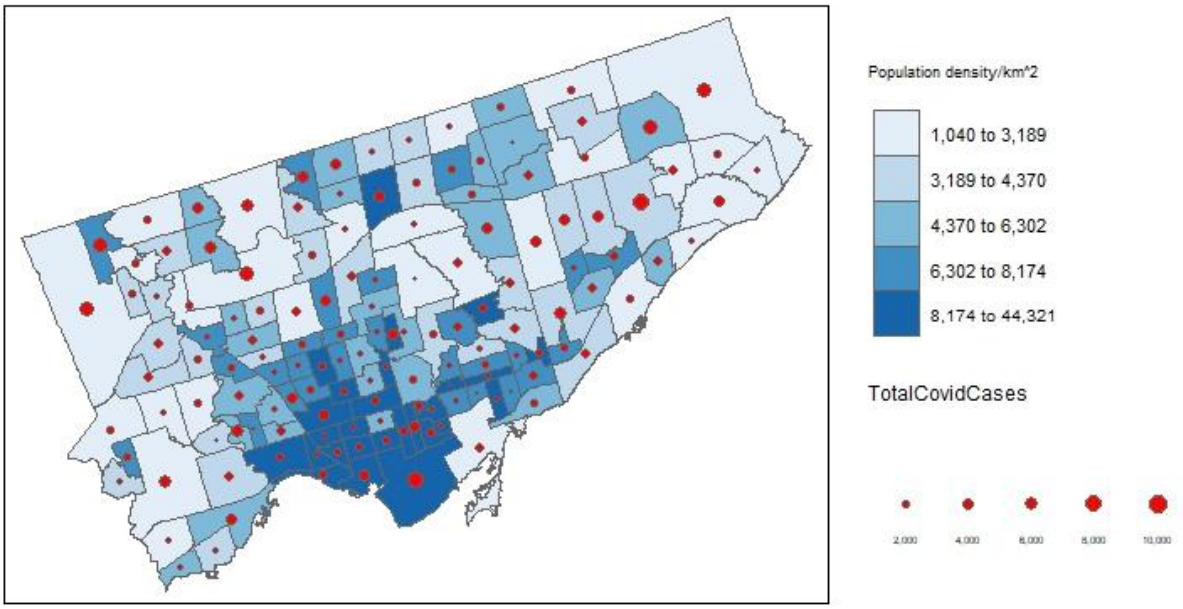



Figure 2. Map of % Low Income and Total COVID-19 Cases, by Toronto Neighbourhood



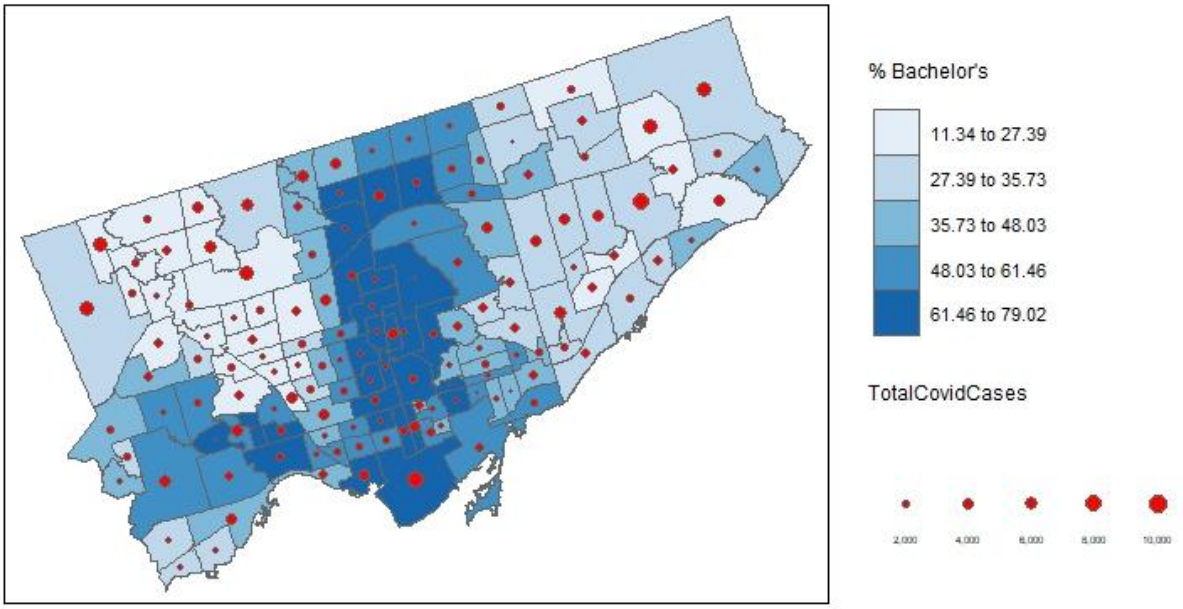

The % of low income individuals and total COVID-19 cases does not appear to have a clear association by neighourhood **(Figure 2).** As displayed, neighborhoods with higher proportions of low-income individuals are observed to have both high and low case counts, and neighborhoods with lower proportions also exhibit a range of case counts.

Figure 3. Map of Population Density and Total Covid-19 Cases, by Toronto Neighbourhood




Population density and total COVID-19 cases do not appear to have a clear association by neighourhood **(Figure 3).** As displayed, neighborhoods with higher proportions of population density are observed to have both high and low case counts, and neighborhoods with lower proportions of population density also exhibit a range of case counts.

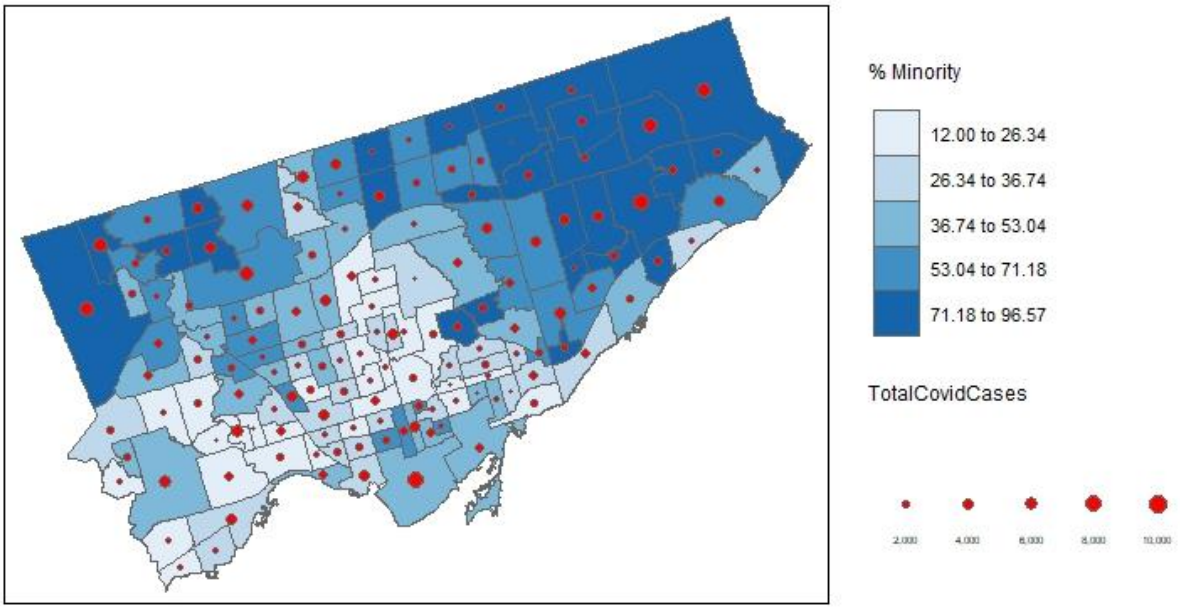
Figure 4. Map of % with Bachelor's degree and Total COVID-19 Cases, by Toronto Neighbourhood




Proportion of those with a bachelor's degree and total COVID-19 cases do not appear to have a clear association by neighourhood **(Figure 4)**. As displayed, neighborhoods with higher

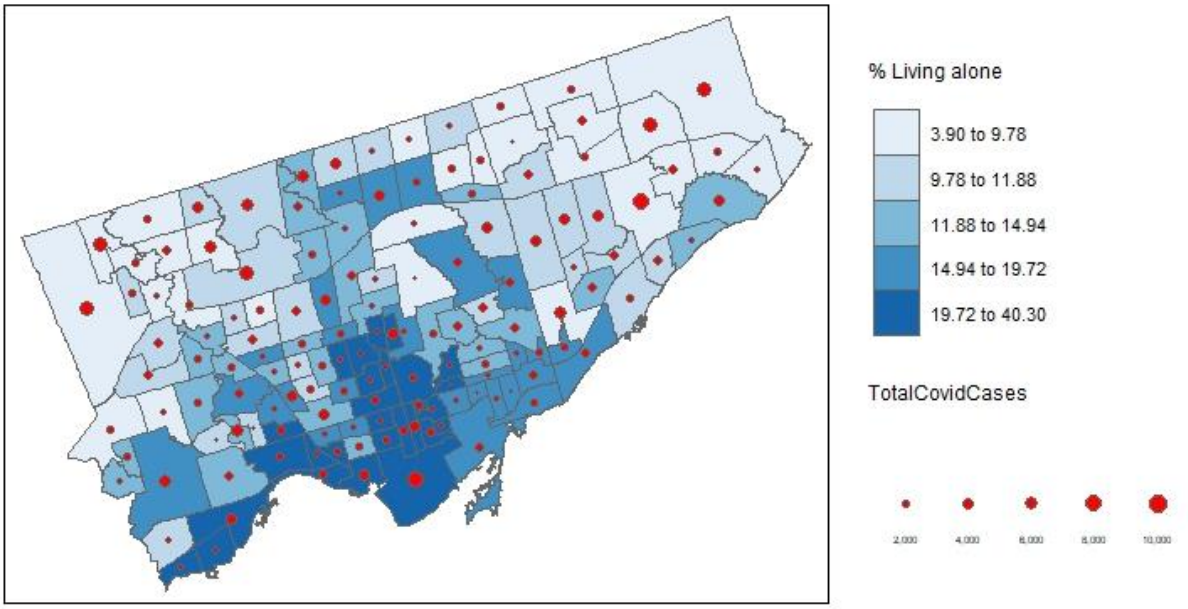
proportions of those with a bachelor's degree are observed to have both high and low case counts, and neighborhoods with lower proportions of those with a bachelor's degree also exhibit a range of case counts.

Figure 5. Map of % Visible Minority and Total COVID-19 Cases, by Toronto Neighbourhood



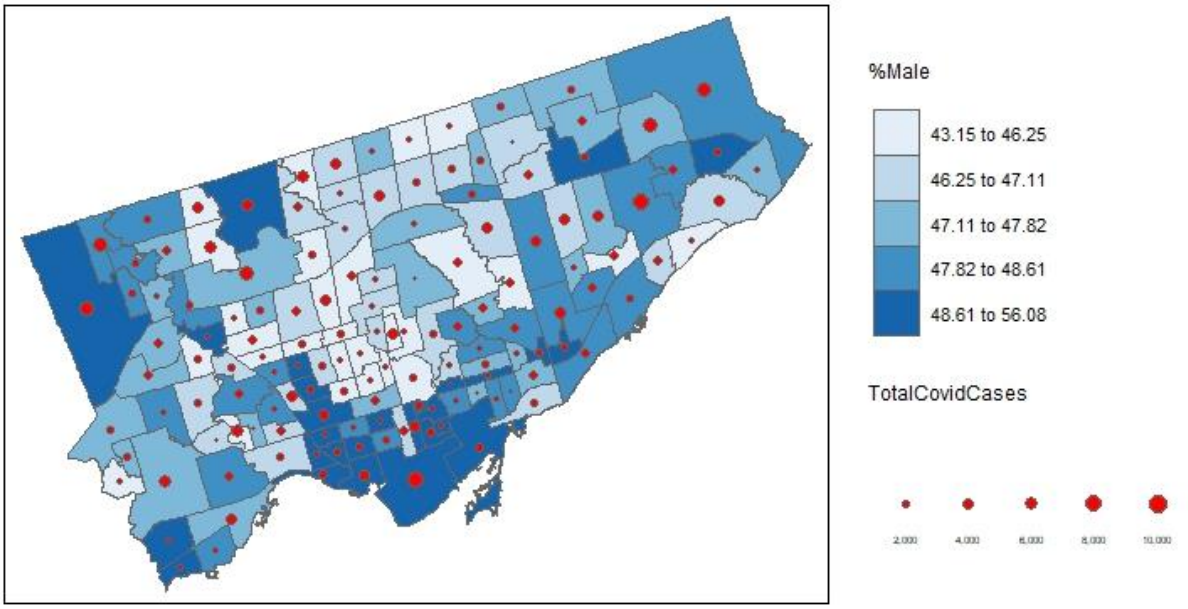

Percentage of visible minority and total COVID-19 cases do appear to have a more clear association by neighourhoods in Toronto **(Figure 5)** As displayed, neighborhoods with higher proportions of visible minority are observed to have high case counts, and neighborhoods with lower proportions of visible minorities exhibit a low case counts.

Figure 6. Map of % living alone and Total COVID-19 Cases, by Toronto Neighbourhood



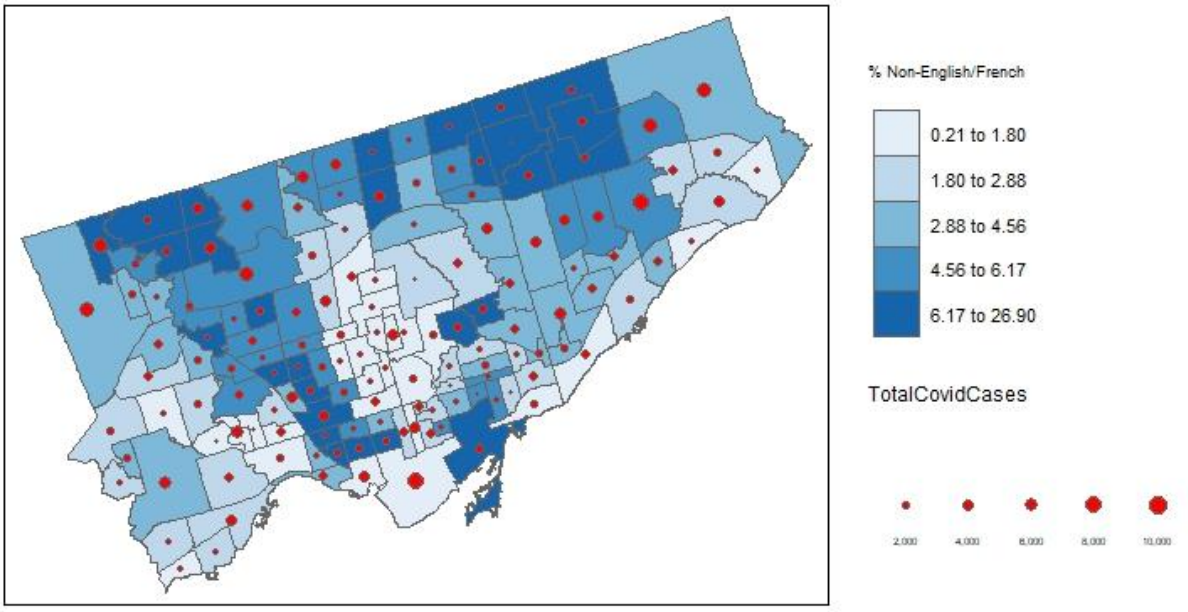

Percentage of those living alone and total COVID-19 cases do not appear to have a clear association by neighourhoods in Toronto **(Figure 6)**. As displayed, neighborhoods with higher proportions of those living alone are observed to have both low and high case counts, and neighborhoods with lower proportions of visible minorities exhibit a wide range of case counts as well.

Figure 7. Map of % Male and Total COVID-19 Cases, by Toronto Neighbourhood



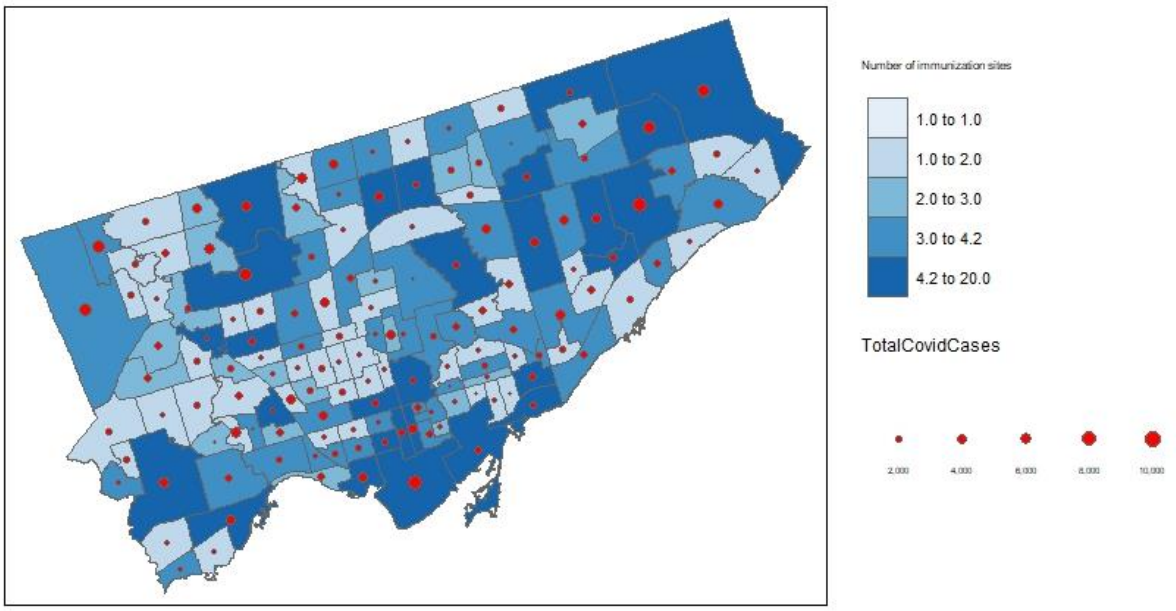
Percentage of males and total COVID-19 cases do not appear to have a clear association by neighourhood **(Figure 7)**. As displayed, neighborhoods with higher proportions of males are observed to have both high and low case counts, and neighborhoods with lower proportions of males exhibit a range of case counts.

Figure 8. Map of % Non-English/French and Total COVID-19 Cases, by Toronto Neighbourhood




Percentage of non-English/French individuals and total COVID-19 cases do not appear to have a clear association by neighourhood **(Figure 8)**. As displayed, neighborhoods with higher proportions of non-English/French individuals are observed to have both high and low case counts, and neighborhoods with lower proportions exhibit a range of case counts.

Figure 9. Map of Number of Immunization sites and Total COVID-19 Cases, by Toronto Neighbourhood




Percentage of those immunized and total COVID-19 cases appears to have a clear visual correlation by neighourhood which is unexpected **(Figure 9)**. As displayed, neighborhoods with higher numbers of immunization sites have high case counts, and neighborhoods with lower numbers exhibit low case counts.

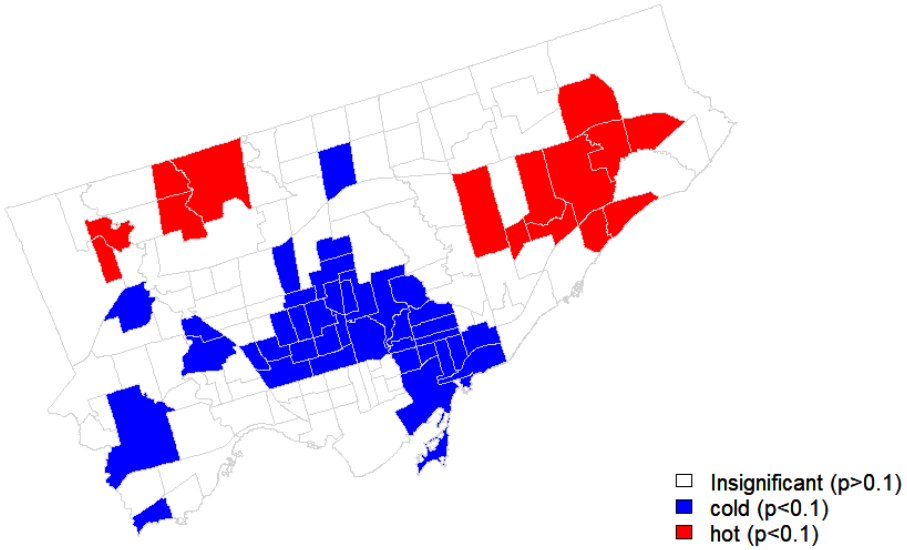
## Spatial Analysis

**Table 1. Global Moran's I Test for Spatial Autocorrelation**

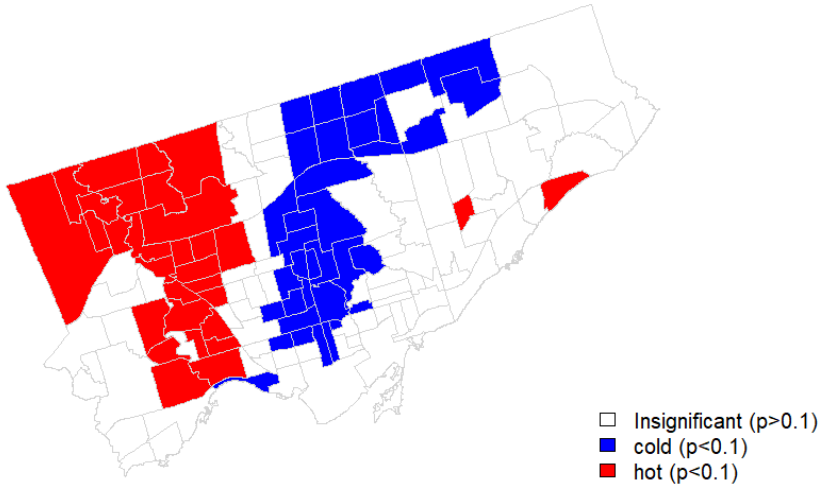
| Statistic | Value |
|---|---|
| Moran's I | 0.145 |
| Expected Moran's I | -0.007 |
| Variance | 0.002 |
| P-value | <0.001 |

**Figure 10.**

**Hot and Cold Spots of Total COVID-19 Cases by Toronto's Neighbourhoods Using the Getis-Ord Statistic**



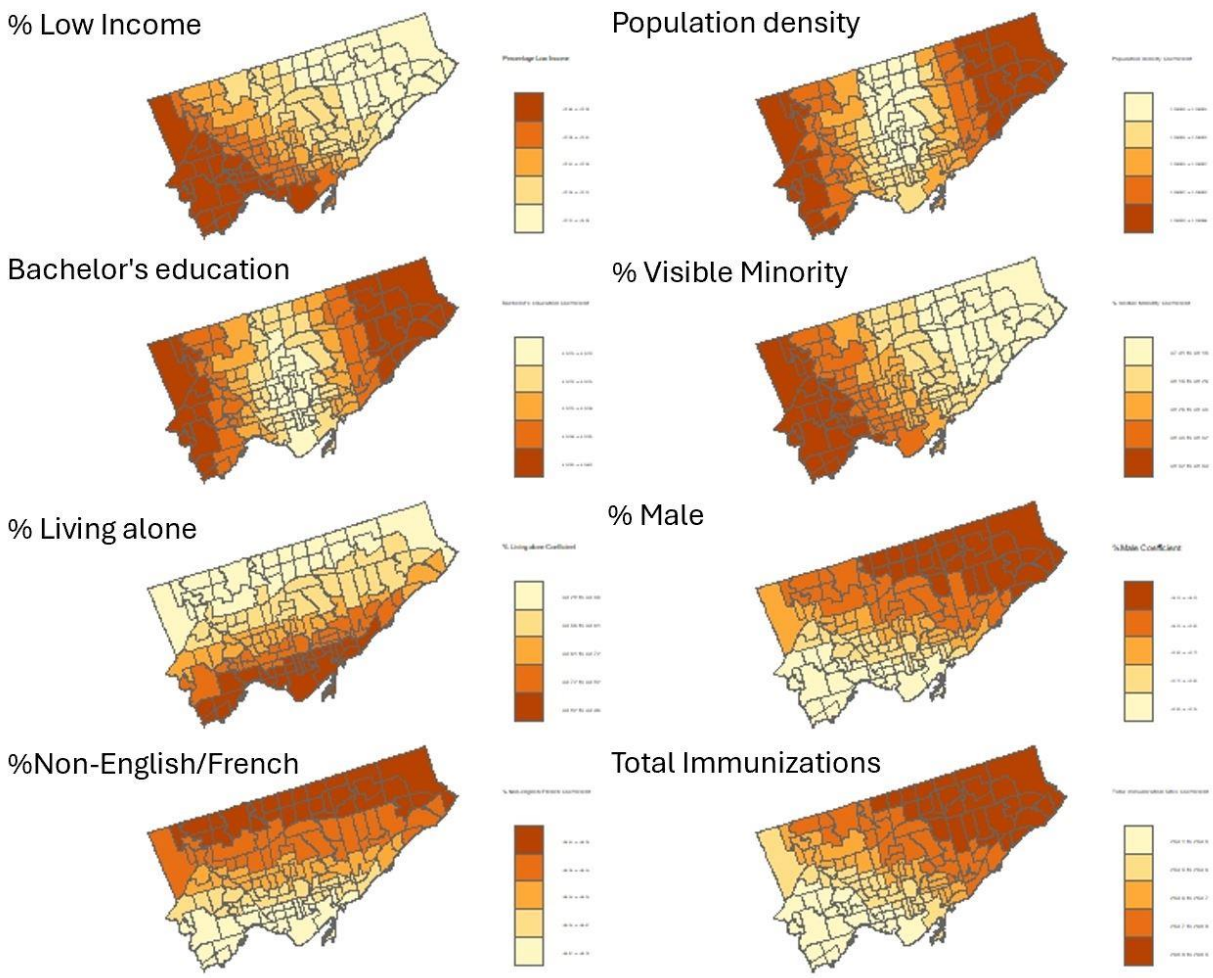

**Hot and Cold Spots of COVID-19 Cases per 1000 by Toronto's Neighbourhoods Using the Getis-Ord Statistic**




Mapping of total COVID-19 case clusters by neighbourhood indicates hotspots in the North Etobicoke, North York, and Scarborough regions of the city, and coldspots in the Midtown, East End, and Central Etobicoke regions **(Figure 10)**. Downtown Toronto shows insignificant findings of clustering. Looking at clustering for cases per 1000, hotspots are no longer prominent in the Scarborough region, but are even more pronounced in North York, North Etobicoke, and also extend into the West End. Midtown remains a cold spot, and the downtown region again shows primarily insignificant clustering.

**Geographically-Weighted Regression**

**Figure 11. Results of Geographically-Weighted Regression**

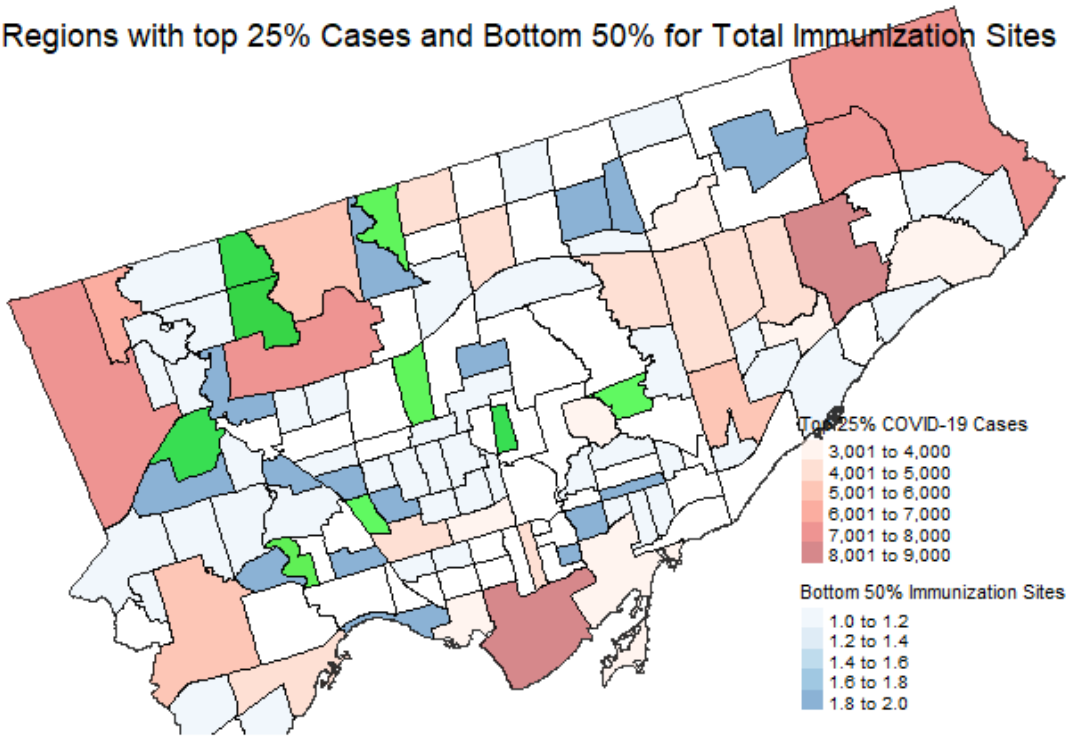



Although not displayed in the figures, the R2 values for the geographically-weighted regression were similar across neighbourhoods, indicating that a similar level of variance in the outcome is explained by the model across Toronto. The global R2 value was 0.455, indicating that the full model explained about 45.5% of the variability in the outcome. As you move towards western Toronto, coefficients between low-income and COVID-19 cases become more negative **(Figure 11)**. Population density and % of those with bachelor's education display a similar spatial relationship with total COVID-19 cases. Specifically, as you move away from central Toronto, coefficients become more positive. For % visible minority, coefficients are more positive as you move west. For % male, coefficients are more negative as you move west. For % living alone, coefficients are more positive as you move South. For % Non-English/French population, coefficients are more negative as you move North. Lastly, for total immunizations, coefficients are more positive as you move East. Overall, it is difficult to discern a general pattern among covariates.

**Immunization and Test Site Analysis**

This part of the analysis involved exploring the intersections of neighberhoods with the highest amount of immunization sites and neighbourhoods with the highest number of COVID-19 cases. The neighbourhoods that had the top 25% of COVID-19 cases and bottom 50% of total immunization sites is represented in green **(Figure 12).** These regions are mainly areas in West North York(Black creek, Glenfield-Jane Heights, Westminster-Branson) and Midtown.

**Figure 12.**




Neighbourhoods with the highest population and population density were also investigated. The neighbourhoods that had the top 25% of population and bottom 50% of total immunization sites is represented in green **(Figure 13).** These regions are mainly areas in West North York(Glenfield-Jane Heights, Westminster-Branson) and North Scarborough(Steeles, Agincourt North and Don Valle Village). The neighbourhoods that had the top 25% population density and bottom 50% of total immunization sites is represented in green **(Figure 14).** These regions are mainly areas in midtown. To expand the scope of population density outside the known density areas of downtown and midtown. We plotted the neighbourhoods that had the top 50% population density and bottom 50% of total immunization sites is represented in green **(Figure 15).** In addition to the neighbourhoods highlighted in Figure 14, West North York(Black Creek, Glenfield-Jane Heights, Westminster-Branson) and North Scarborough(Steeles, and Don Valley Village) are highlighted.

**Figure 13.**

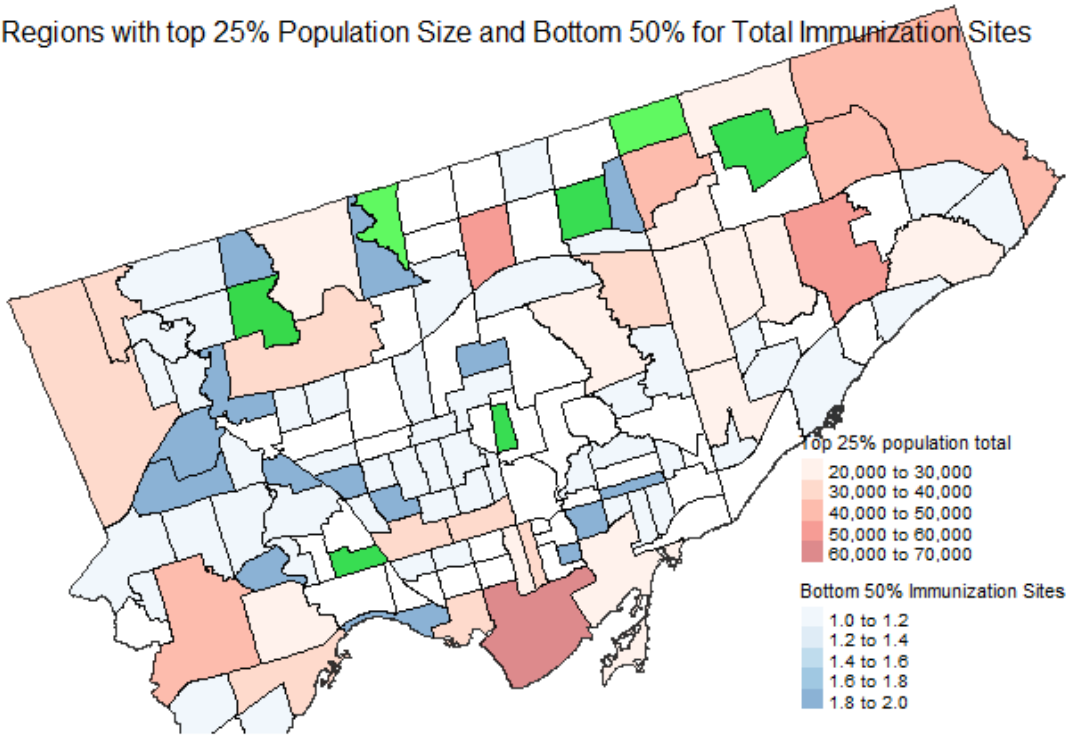



**Figure 14**.

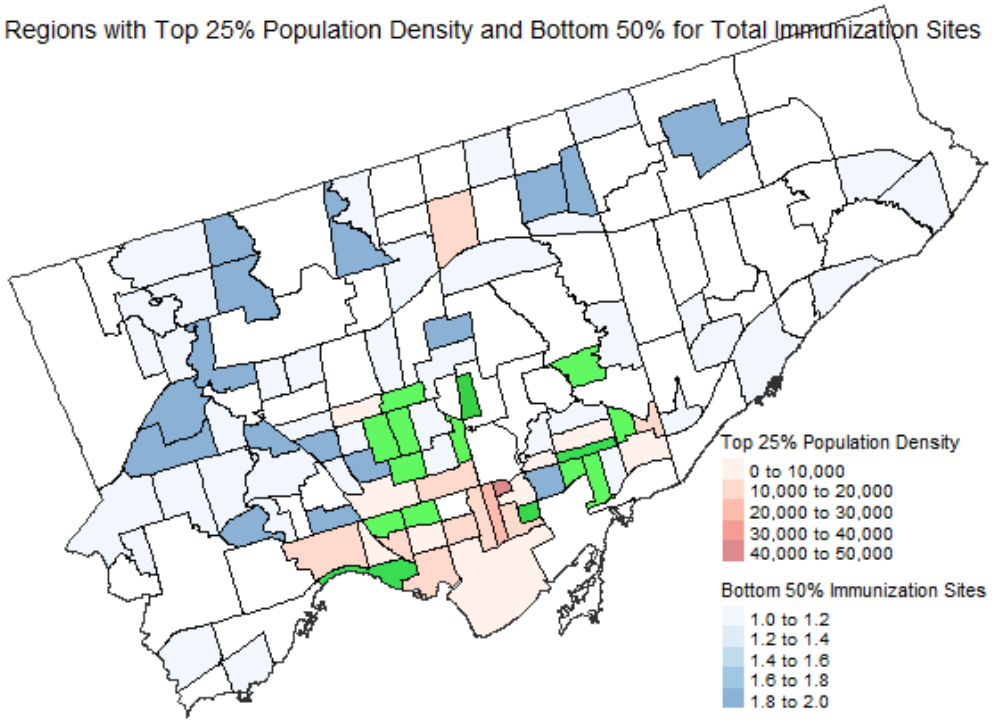
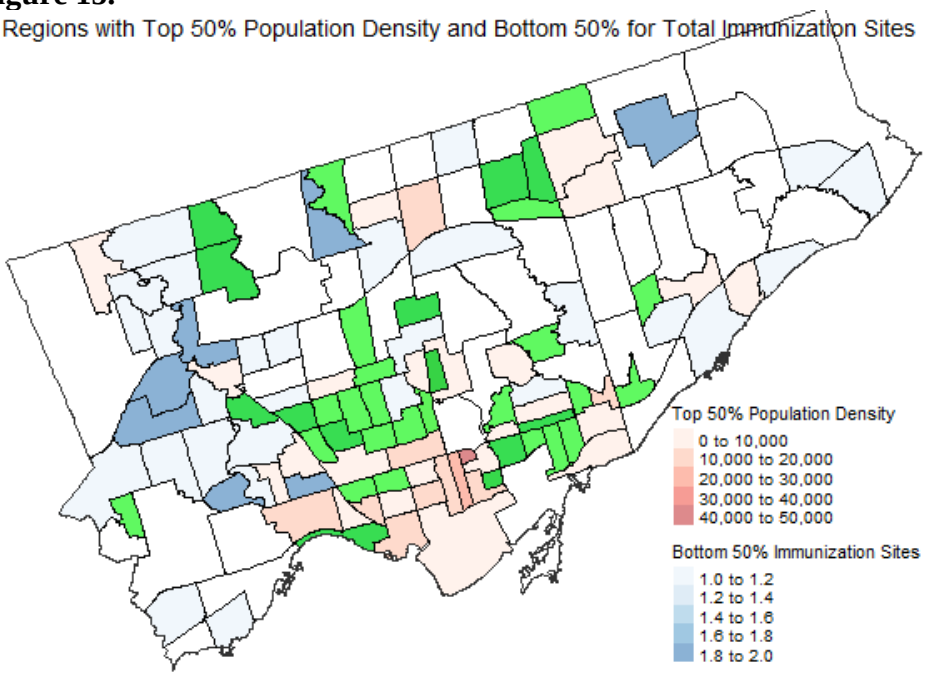


**Figure 15.**




Buffers showing 800m of radius around each immunization sites were then plotted on the same maps above. **Figure 16** shows each immunization site in Toronto with an 800m buffer radius mapped onto a map of Toronto showing the neighbourhoods that had the top 25% of COVID-19 cases and bottom 50% of total immunization sites is represented in green**.** The regions that showed lack of coverage within an 800m radius of an immuniation site are mainly areas in West North York(Glenfield-Jane Heights, Westminster-Branson) and some of north Etobicoke and North Scarborough.

**Figure 16.**

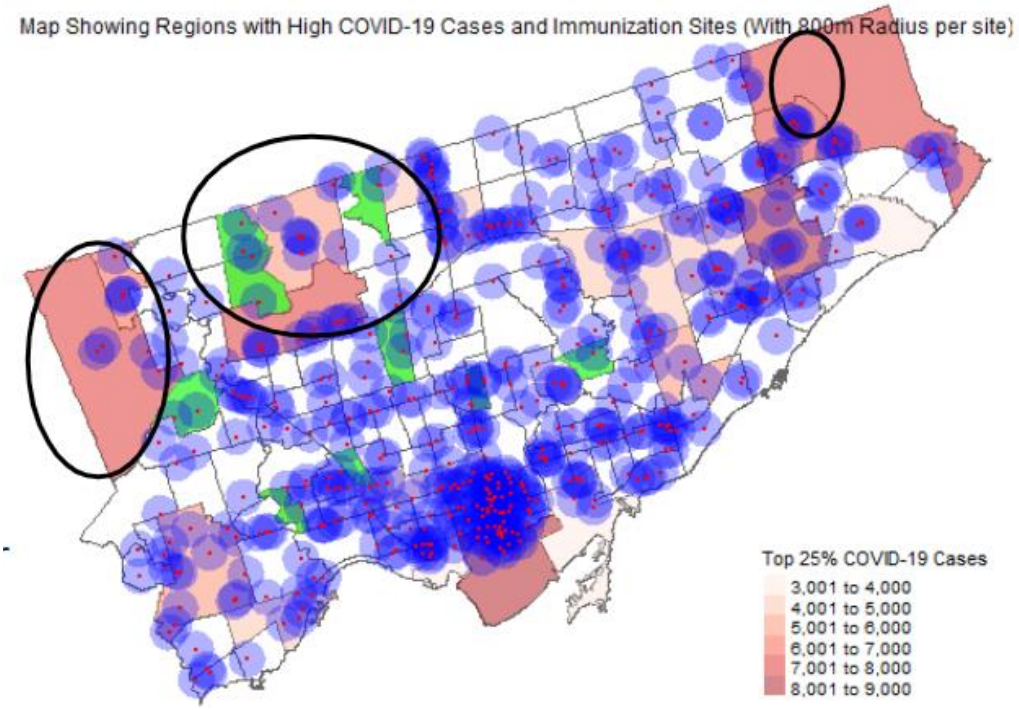



**Figure 17** shows each immunization site in Toronto with an 800m buffer radius mapped onto a map of Toronto showing the neighbourhoods that had the top 25% population and bottom 50% of total immunization sites is represented in green**.** The regions that showed lack of coverage within an 800m radius of an immuniation site are mainly areas in West North York( Westminster-Branson) and North Scarborough (Steeles and Agincourt North). **Figure 18** shows each immunization site in Toronto with an 800m buffer radius mapped onto a map of Toronto showing the neighbourhoods that had the top 50% population density and bottom 50% of total immunization sites is represented in green**.** The regions that showed lack of coverage within an 800m radius of an immuniation site are mainly areas in West North York(Glenfield-Jane Heights and Westminster-Branson) and North Scarborough(Steeles and L'Amoreaux).

**Figure 17.**

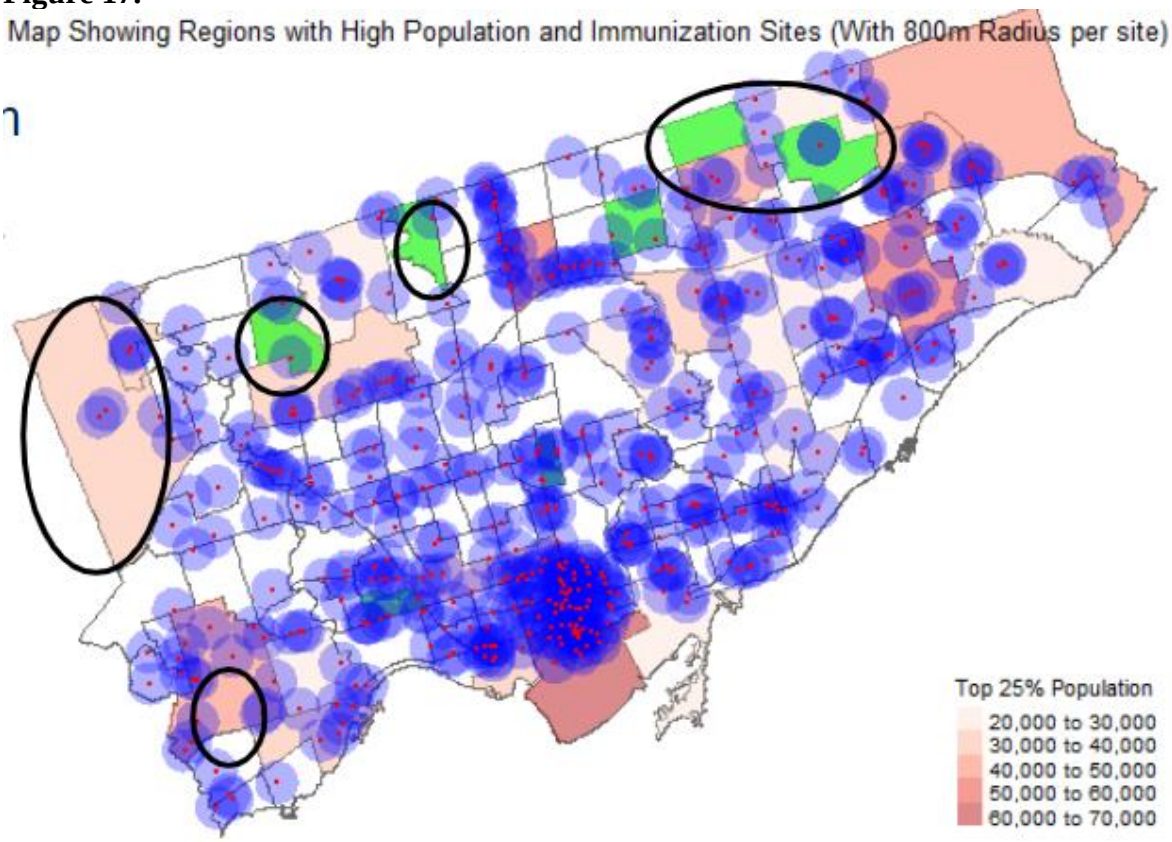
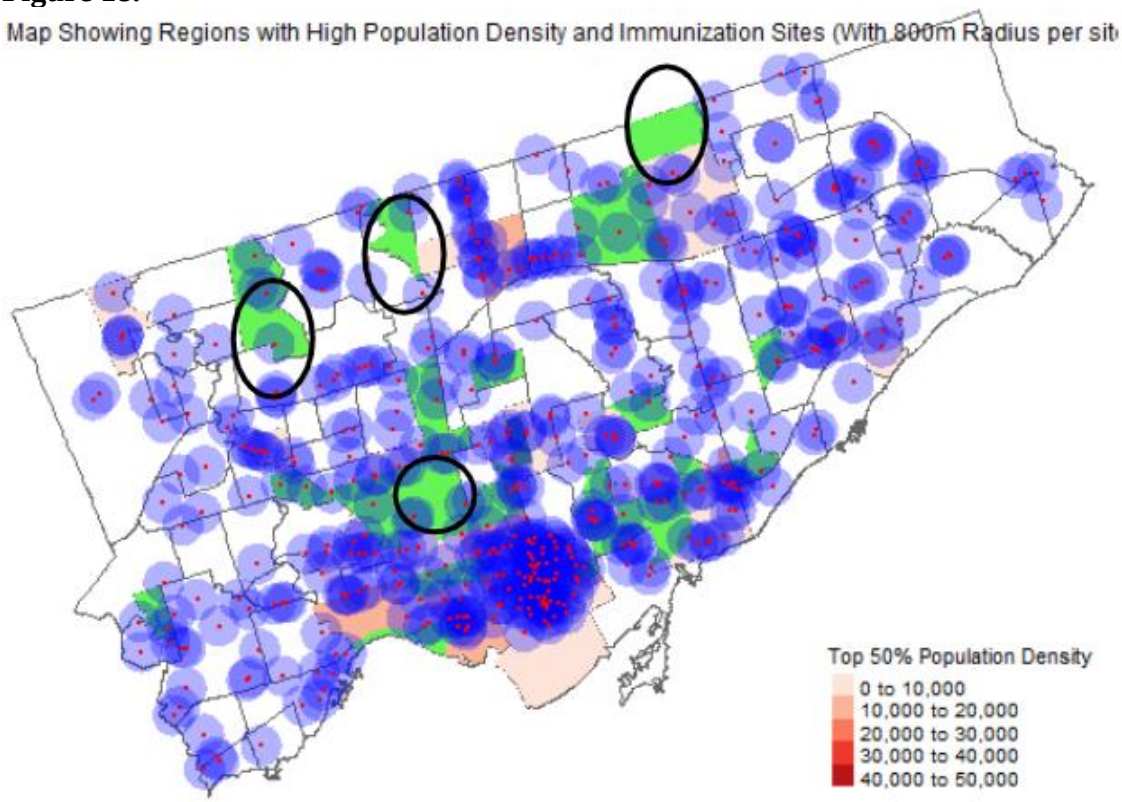

Map Showing Regions with High Population and Immunization Sites (With 800m Radius per site)
Top 25% Population
20,000 to 30,000
30,000 to 40,000
40,000 to 50,000
50,000 to 60,000
60,000 to 70,000

**Figure 18.**




Buffers showing 3000m of radius around each testing sites were then plotted on the same maps above. **Figure 19** shows each testing site in Toronto with an 3000m buffer radius mapped onto a map of Toronto showing the neighbourhoods that had the top 25% of COVID-19 cases and bottom 50% of total testing sites is represented in green**.** The regions that showed lack of coverage within an 3000m radius of an testing site are shown largely in Scarborough as well as Etobicoke and West North York.

**Figure 19.**

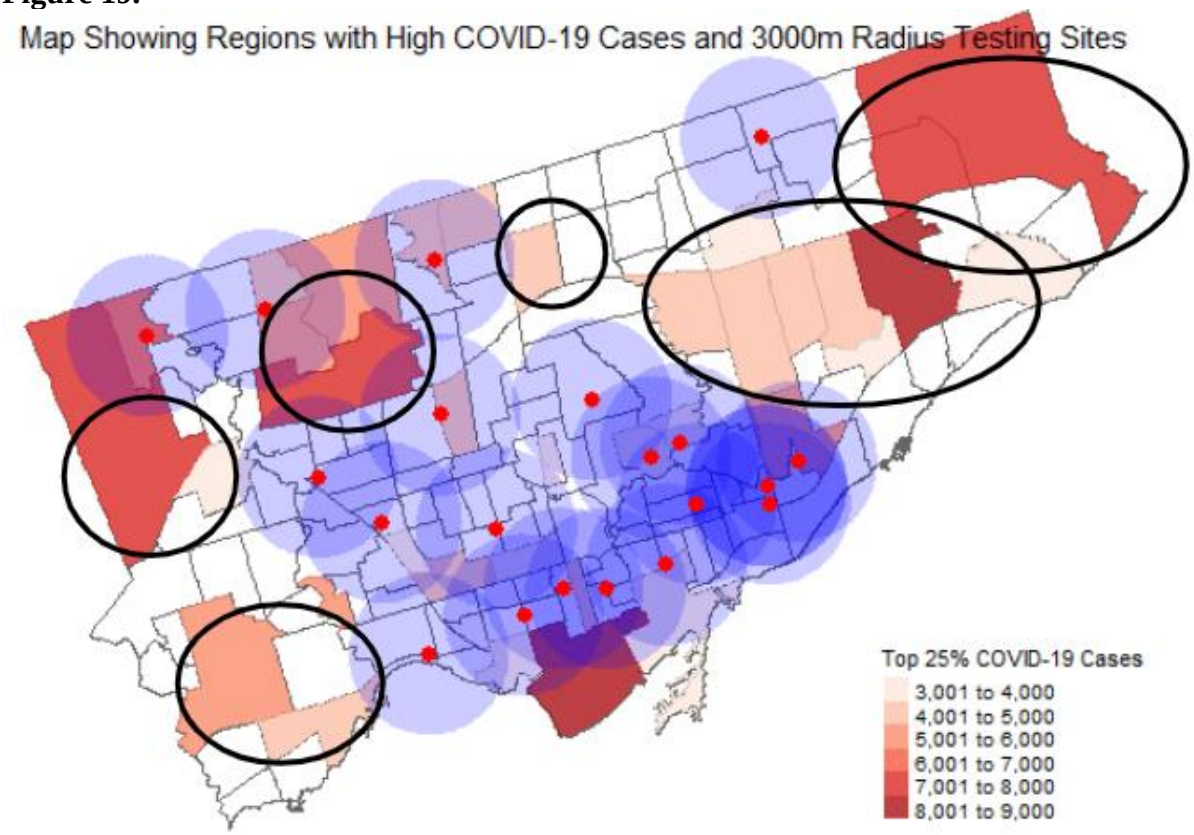



**Figure 20** shows each testing site in Toronto with an 3000m buffer radius mapped onto a map of Toronto showing the neighbourhoods that had the top 25% of population and bottom 50% of total testing sites is represented in green**.** The regions that showed lack of coverage within an 3000m radius of an testing site are similar and shown largely in Scarborough as well as Etobicoke and West North York. **Figure 21** shows each testing site in Toronto with an 3000m buffer radius mapped onto a map of Toronto showing the neighbourhoods that had the top 50% population density and bottom 50% of total testing sites is represented in green**.** The regions that showed lack of coverage within an 3000m radius of an testing site are shown in Scarborough as well as Willowdale East in North York.

**Figure 20.**

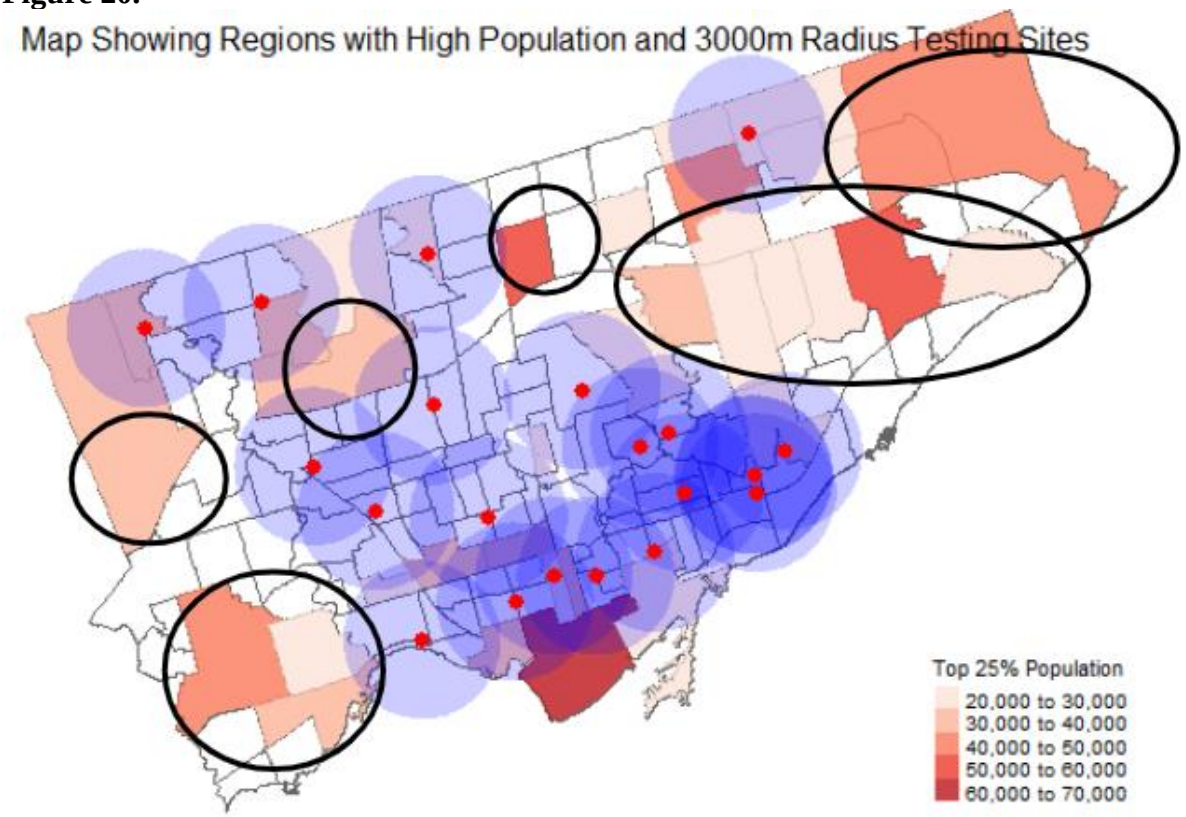
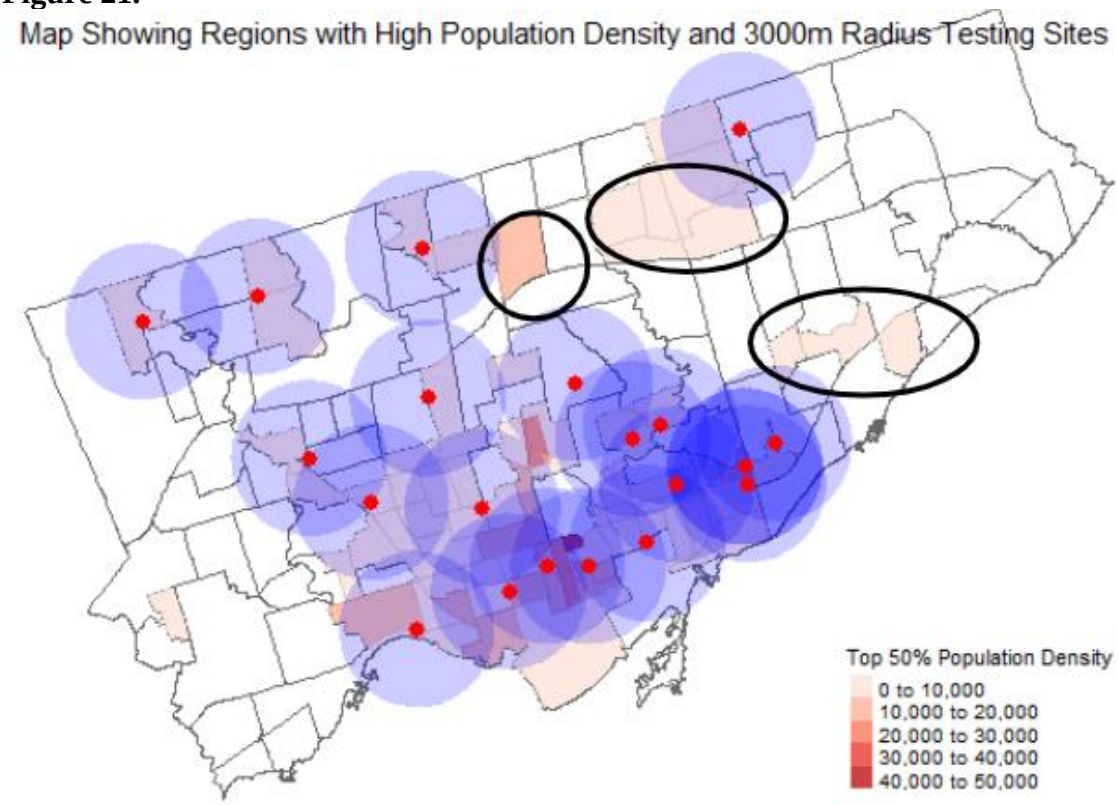

Map Showing Regions with High Population and 3000m Radius Testing Sites
Top 25% Population
20,000 to 30,000
30,000 to 40,000
40,000 to 50,000
50,000 to 60,000
60,000 to 70,000

**Figure 21.**




## Discussion

The Global Moran's I value (0.145) and the statistically significant p-value (<0.001) at the 5% level, give strong evidence to reject the null hypothesis of no spatial autocorrelation, suggesting that there is clustering of total COVID-19 cases by Toronto neighbourhood. This finding aligns with prior research suggesting that sociodemographic and geographic characteristics are potentially associated with disease burden.[4] Furthermore, the Getis-Ord hotspot analysis (**Figure 10)** revealed distinct geographic patterns, with various hotspots and coldspots of COVID-19 case density. These findings were expected, and may be attributed to socioeconomic disparities, as the identified hotspots—North Etobicoke, North York and Scarborough—are also neighbourhoods characterized by a higher percentage of low-income populations **(Figure 2)** and a greater percentage of residents belonging to visible minority groups **(Figure 5)**. Similarly, the identified coldspots—Central Etobicoke and Midtown—are also neighbourhoods characterized by a lower percentage of low-income populations **(Figure 2)** and a lower percentage of residents belonging to visible minority groups **(Figure 5)**. This suggests that regions that are more affluent may be less likely to be hotspots of case density. In addition, downtown Toronto was found to be an insignificant region, which was quite unexpected. However, due to the high population density in this region, it may have been a focus for targeted public health interventions which could have reduced the COVID-19 cases in the area. Additionally, many individuals commute from non-central Toronto areas to downtown Toronto, and with COVID-19 quarantine restrictions in place,

it would have reduced much of the foot traffic in the region, resulting in less cases among those who reside in the downtown neighbourhoods.

For geographically-weighted regression, higher education levels correlated with higher COVID-19 cases to a greater degree in non-central Toronto areas. For a possible explanation, central Toronto may be a hub for better jobs and individuals from outer Toronto areas could be commuting through trains and buses, which happen to be more densely packed and can increase the spread of COVID-19. Conversely, those living in Toronto with higher education may be able to walk/bike and avoid the commute through trains. Population density displays a similar distribution in which there is a better correlation with high COVID-19 cases in non-central Toronto areas. Although the population density in central Toronto is expected to be greater, thus should be more correlated with cases, it is also possible that this region had better public health interventions and measures to prevent or reduce the transmission of COVID-19. The spatial correlation with percentage of visible minority was also unexpected. There are more visible minorities in Eastern Toronto (i.e. Scarborough area), and we expected to observe a greater association with COVID-19 in those regions. However, the spatial maps for the geographically-weighted regression demonstrate the opposite.

As a potential explanation, public health interventions could have focused on the visible minority populations in Eastern Toronto as these regions may consist of a greater proportion of visible minorities, leading to less targeted interventions in Western regions for minorities. % living alone and COVID-19 cases were more strongly correlated in Southern parts of Toronto. This should be expected as these regions, which include downtown Toronto, could be linked to other factors, such as a denser population, reliance on shared facilities, etc. which may explain this relationship. Overall, spatial maps for geographically-weighted regressions showed complex patterns which were difficult to discern as many covariates had dissimilar patterns for their coefficients with total COVID-19 cases across neighourhoods in Toronto.

In general, the number of immunization sites falls short of meeting community demand in certain neighborhoods such as West North York (Glenfield-Jane Heights, Westminster-Branson), some areas of North Scarborough (Steeles, and Don Valley Village) and Midtown. These same regions in West North York and North Scarborough also have lack of immunization site coverage within 800m. This means that people living in these neighborhoods with high COVID and population have to travel further for an immunization site. This may further deter people from getting the vaccination. These neighbourhoods are also more likely to be dealing with socioeconomic challenges and barriers to healthcare access. Altogether , this shows the availability of immunization services falls short of meeting community demand or needs in these neighbourhoods.

Similarly with testing sites, there is a large lack of coverage observed across many areas of Scarborough. Other regions such as Etobicoke and West North York also show some lack of coverage when it comes to testing sites. These lack of coverage areas were observed even after expanding our coverage radius from 800M to 3000M. Ultimately, this shows the increased barriers to access for COVID-19 testing for residents of these neighborhoods in addition to the socioeconomic challenges such as low income, visible minority and immigrant status that they already face.

Limitations of the study can possibly include utilizing the Getis-Ord statistic rather than the Local Moran's I test. Getis-Ord test was utilized for purposes of simplicity to allow for a simpler interpretation. However, implementing the Local Moran's test would have allowed for the identification of outlier regions in which hot spots neighbourhoods may be found among cold spots, and vice versa, which could highlight the need to understand the needs of these regions. Additionally, although the geographically-weighted regression made it possible to observe spatial trends of covariates and their correlation with COVID-19 cases, it would have been ideal to calculate p-values through other tests to determine if the spatial relationships are significant. In addition, as the global R2 value only explained 45.5% of the variability in our outcome (i.e. COVID-19 case counts), additional sociodemographic indicators could be implemented (i.e. ethnicity, age, employment status, etc.) as long as this information is available and overfitting is not a concern.

As the sociodemographic data utilized in this study is from 2016, this limits the generalizability of our study, considering that this dataset is being linked with COVID-19 cases which occur from 2020-2023. The population demographics and socio-demographics could have changed during this period, and it would have been ideal to utilize more recent sociodemographic data. As an example, immigration could have increased substantially during this period, potentially increasing the number of visible minorities and low-income individuals in Toronto. For reasons such as these, our results may not generalize as well to the Toronto neighbourhoods during the pandemic. Future studies should repeat the methodology outlined in this project, and utilize more recent sociodemographic data, as well as conduct significance tests for geographically-weighted regression, to yield results which are more generalizable. Future studies should also include analysis of intervention outcome data such as percent immunization.

## Conclusion

Based on the data available and the results obtained, certain regions in Toronto (i.e. Scarborough, North Etobicoke, North York) consisted of high COVID-19 cases which may be due to a high proportion of low-income individuals and visible minorities who reside in these regions. These regions also lack coverage of immunization and testing sites. This highlights a need to implement public health interventions (e.g. such as key health messages displayed in multiple languages, programs for low-income populations such as food assistance, housing support, mobile clinics, etc.) in these areas in preparation for future outbreaks and to help reduce transmission in these vulnerable/disadvantaged communities.

## Acknowledgments

We would like to thank Dr. Jude Kong for teaching us the skills necessary for this project and guiding us in the right direction. We also acknowledge the City of Toronto Open Data Portal for providing the data essential to this project.

.

**References**

1. Ashmore P, Sherwood E. An overview of COVID-19 global epidemiology and discussion of potential drivers of variable global pandemic impacts. J Antimicrob Chemother. 2023 Nov 23;78(Suppl 2):ii2–11.
2. Worldometer. COVID - Coronavirus Statistics - Worldometer [Internet]. [cited 2024 Dec 4]. Available from: https://www.worldometers.info/coronavirus/
3. Worldometer. Canada COVID - Coronavirus Statistics - Worldometer [Internet]. [cited 2024 Dec 4]. Available from: https://www.worldometers.info/coronavirus/country/canada/
4. Mishra S, Ma H, Moloney G, Yiu KCY, Darvin D, Landsman D, et al. Increasing concentration of COVID-19 by socioeconomic determinants and geography in Toronto, Canada: an observational study. Ann Epidemiol. 2022 Jan;65:84–92.
5. City of Toronto. COVID-19 Cases in Toronto [Internet]. [cited 2024 Dec 4]. Available from: https://open.toronto.ca/dataset/covid-19-cases-in-toronto/
6. City of Toronto. Neighbourhoods [Internet]. [cited 2024 Dec 4]. Available from: https://open.toronto.ca/dataset/neighbourhoods/
7. City of Toronto. COVID-19 Immunization Clinics [Internet]. [cited 2024 Dec 4]. Available from: https://open.toronto.ca/dataset/covid-19-immunization-clinics/
8. City of Toronto. Neighbourhood Profiles [Internet]. [cited 2024 Dec 4]. Available from: https://open.toronto.ca/dataset/neighbourhood-profiles/
9. City of Toronto. COVID-19 Testing Sites [Internet]. [cited 2024 Dec 4]. Available from: https://open.toronto.ca/dataset/covid-19-testing-sites/