Understanding COVID-19 in Toronto Neighbourhoods – A Data-driven Approach For Vaccination Prioritization

Chris Blackwood

Introduction and Business Problem

COVID-19 has been a major health and economic issue worldwide since early 2020. For over a year we have watched daily updates on new cases detected, rising death counts, public health restrictions, and social and economic fallout. Dedicated research has led to the rapid development of safe and effective vaccines, and rollout has begun worldwide. The complexity of producing, distributing and administering these vaccines means that some will receive vaccines quickly, but many people will have to wait many months. Data scientists can play a valuable role in understanding the geographic and demographic distributions of COVID-19 infections at a detailed level. If patterns of infection can be understood, then areas for rapid vaccination can be prioritized.

In this study I examine COVID-19 infections in the city of Toronto, the capital of the Canadian province of Ontario. I am interested in the rates of COVID-19 infection by geographic area to see if certain neighbourhoods have more infections than others, or if the distribution is random. I will then attempt to understand the different regions of Toronto in terms of demographic characteristics. I will compare COVID-19 infection rates to a variety of census data, including population statistics, household income, and education level. For regions with particularly high or low COVID-19 cases, I will attempt to identify any demographic factors of significance in these areas. I will use an assortment of statistical and mapping analysis techniques to determine relevance of demographic factors.

The target audience for the study is government officials responsible for public health measures designed to limit the spread of COVID-19 and to coordinate vaccination distribution. My hope is that trends will become apparent, such as particular geographic regions or demographic factors correlate with high COVID-19 case rates. If I am successful

in identifying factors that correlate with high COVID-19 infection rates, then these correlations could be used to help prioritize neighbourhoods for further heath measures and vaccine distributions.

Data

The city of Toronto maintains a strong online database of city-specific information. I will reference three key databases from the city of Toronto:

- COVID-19 data for the city of Toronto – daily updates of infections sorted by neighbourhood, retrieved January 15, 2021. [1]
- City of Toronto neighbourhood profiles – a diverse collection of census data, most recently published 2016. [2]
- Toronto neighbourhood spatial information – Geojson file of 140 neighbourhoods. [3]

Methodology

The most fundamental question to answer is if certain neighbourhoods have higher rates of COVID-19 infections than others. The key metric is 'Covid19Rate', the number of infections per 100,000 residents of each Toronto neighbourhood. Figure1 shows Pandas dataframes with the neighbourhoods with the least (left) and most (right) infections per 100,000. There is significant variation, as the neighbourhood with the highest infection rate (Thistletown-Beaumond Heights) has a rate of more than ten times the neighbourhood with the lowest infection rate (The Beaches).

Neighbourhood	Covid19Rate
Bridle Path-Sunnybrook-York Mills	938.916469
Newtonbrook East	938.062993
Lawrence Park South	935.502998
Willowdale East	854.582226
Mount Pleasant East	846.497764
Rosedale-Moore Park	831.620704
Woodbine Corridor	813.332270
Forest Hill South	810.659709
Runnymede-Bloor West Village	645.481629
The Beaches	625.956322

Neighbourhood	Covid19Rate
Thistletown-Beaumond Heights	6650.579151
Mount Olive-Silverstone-Jamestown	6463.555259
Maple Leaf	6201.167046
Black Creek	6095.597369
Humbermede	5969.765198
West Humber-Clairville	5586.575408
Humber Summit	5565.399485
Glenfield-Jane Heights	5559.017415
Weston	5491.329480
Downsview-Roding-CFB	5232.226406

Figure 1. Neighbourhoods with lowest and highest rates of COVID-19 infection.

The neighbourhood profiles of demographic data were examined for factors correlating with COVID-19 infection rates. A subset of the full demographic suite was chosen for investigation, specifically data related to visible minority status, household income, employment, education, and commute to work.

For each demographic category, a similar workflow was employed. A correlation matrix was generated and plotted in tables and as heatmaps. The critical information is the correlation between Cvod19Rate and the demographic feature of interest. Regression plots were also examined to visualize correlations that warrant further attention. For brevity, not all figures are captured here.

Correlation matrices for visible minority data are shown in Figures 2 and 3. The column label abbreviations are explained in the figure captions. The visible minorities with the highest correlation to COVID-19 infections per 100,000 are Black (0.65), and Latin American (0.53). Japanese has the highest negative correlation at -0.45. Regression plots for visible minority black and visible minority Japanese are shown in Figure 4.

	NeighbourhoodNumber	Covid19Rate	Covid19CaseCount	VMAboriginal	VMLatinAmerica	VMBlack	VMArab	VMNotSpecified	VMMultiple	VMNot
NeighbourhoodNumber	1.000000	-0.160640	-0.027007	0.260068	-0.221717	0.022347	-0.070637	0.083444	0.183580	-0.073978
Covid19Rate	-0.160640	1.000000	0.730923	-0.096396	0.533566	0.654493	0.166010	0.493298	0.311711	-0.334786
Covid19CaseCount	-0.027007	0.730923	1.000000	0.223191	0.602199	0.851278	0.462577	0.778155	0.776954	0.040880
VMAboriginal	0.260068	-0.096396	0.223191	1.000000	0.173971	0.322458	0.206579	0.381450	0.435464	0.530763
VMLatinAmerica	-0.221717	0.533566	0.602199	0.173971	1.000000	0.649863	0.198034	0.431688	0.488615	0.192868
VMBlack	0.022347	0.654493	0.851278	0.322458	0.649863	1.000000	0.395387	0.806073	0.736919	-0.062478
VMArab	-0.070637	0.166010	0.462577	0.206579	0.198034	0.395387	1.000000	0.364912	0.524611	0.165470
VMNotSpecified	0.083444	0.493298	0.778155	0.381450	0.431688	0.806073	0.364912	1.000000	0.725973	-0.009871
VMMultiple	0.183580	0.311711	0.776954	0.435464	0.488615	0.736919	0.524611	0.725973	1.000000	0.229005
VMNot	-0.073978	-0.334786	0.040880	0.530763	0.192868	-0.062478	0.165470	-0.009871	0.229005	1.000000

Figure 2. Correlation matrix for a portion of visible minority demographic data. The categories shown are visible minority Aboriginal, Latin American, Black, Arab, unspecified, multiple visible minorities, or not a visible minority.

	NeighbourhoodNumber	Covid19Rate	Covid19CaseCount	VMSouthAsian	VMChinese	VMFilipino	VMSoutheastAsian	VMWestAsian	VMKorean	VMJapanese
NeighbourhoodNumber	1.000000	-0.160640	-0.027007	0.238937	0.218038	0.172835	-0.171758	-0.120980	-0.179417	0.087104
Covid19Rate	-0.160640	1.000000	0.730923	0.377711	-0.194829	0.410632	0.458677	-0.013907	-0.216567	-0.423823
Covid19CaseCount	-0.027007	0.730923	1.000000	0.740276	0.083278	0.666958	0.519743	0.214637	-0.015151	-0.060593
VMSouthAsian	0.238937	0.377711	0.740276	1.000000	0.182574	0.537870	0.059744	0.278101	-0.021373	0.025877
VMChinese	0.218038	-0.194829	0.083278	0.182574	1.000000	0.083815	-0.014296	0.413517	0.408836	0.410311
VMFilipino	0.172835	0.410632	0.666958	0.537870	0.083815	1.000000	0.159245	0.234661	0.070573	-0.028514
VMSoutheastAsian	-0.171758	0.458677	0.519743	0.059744	-0.014296	0.159245	1.000000	0.024845	-0.035139	-0.044588
VMWestAsian	-0.120980	-0.013907	0.214637	0.278101	0.413517	0.234661	0.024845	1.000000	0.834930	0.494523
VMKorean	-0.179417	-0.216567	-0.015151	-0.021373	0.408836	0.070573	-0.035139	0.834930	1.000000	0.651388
VMJapanese	0.087104	-0.423823	-0.060593	0.025877	0.410311	-0.028514	-0.044588	0.494523	0.651388	1.000000

Figure 3. Correlation matrix for a second portion of visible minority demographic data. The categories shown are visible minority South Asian, Chinese, Filipino, Southeast Asian, West Asian, Korean, and Japanese.

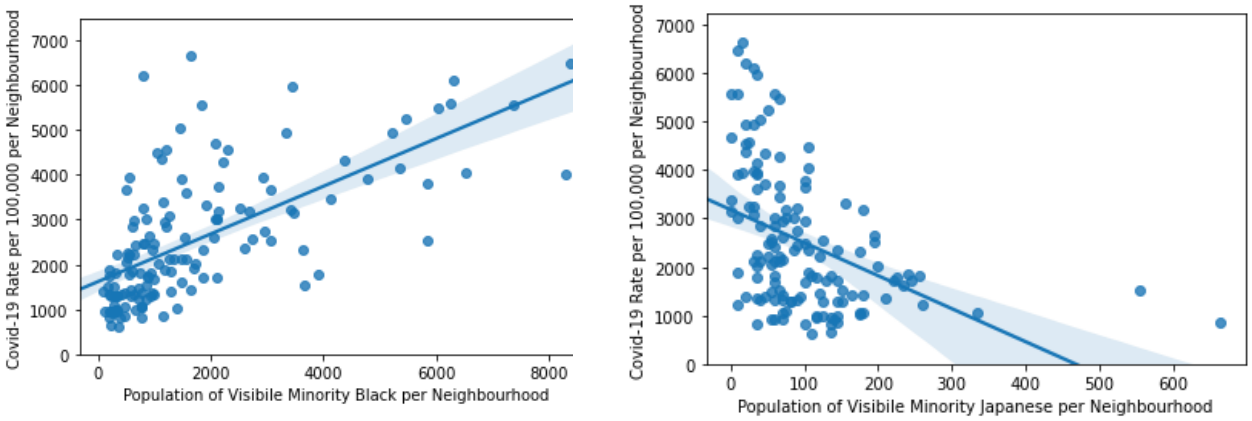

Figure 4. Left - Regression plot of COVID-19 cases per 100,000 people for visible minority Black populations showing a strong positive correlation. Right - Regression plot of COVID-19 cases per 100,000 people for visible minority Japanese populations showing a negative correlation.

Total household income was separated into bins. The heat map of the correlation matrix for total household income and COVID-19 infections is shown in Figure 5. The bins chosen for household income are below $25,000, $25,000-$40,000, $40,000-60,000, $60,000-$80,000, and greater than $80,000. The most important row to consider is Covid19Rate. Neither of the income bins has a very high correlation with Covid19Rate, with the max value 0.20 for $25,000-$40,000. But the highest income bin ($80,000 plus) has the largest negative correlation at -0.26.

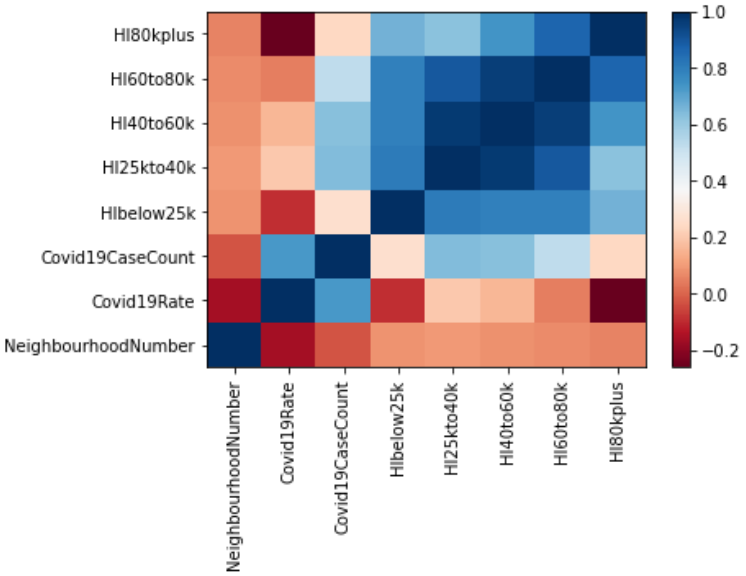

Figure 5. Heatmap of correlation matrix for total household income data.

An assortment of data describing commuting to work was analyzed, including duration of commute and method of transportation. There are high positive correlations between COVID-19 infection rates and both long commutes to work and commuting as a passenger

in a private vehicle. The strongest negative correlation is with those who cycle to work. There is no significant correlation between taking public transit and COVID-19 infections. Figure 6 shows two examples of commute data in regression plots.

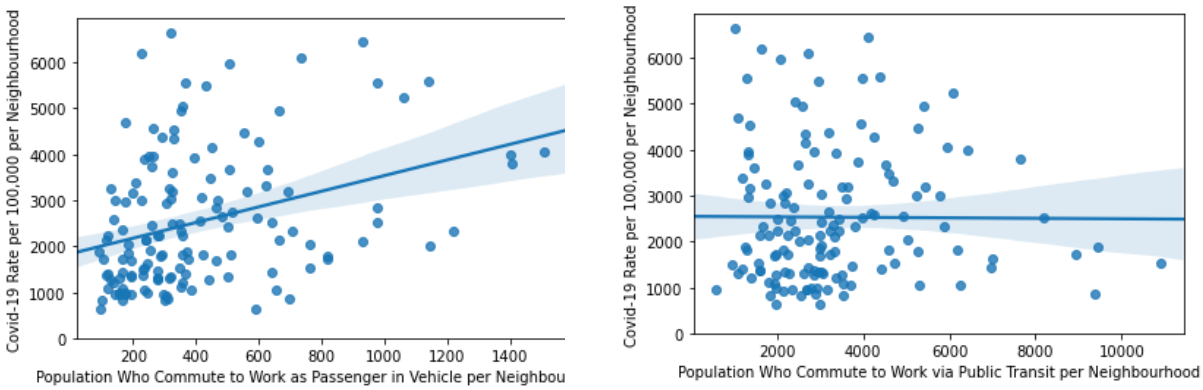

Figure 6. Left - regression plot showing positive correlation between commuting to work as a passenger in a vehicle and COVID-19 infections. Right - regression plot showing no correlation between commuting to work via public transit and COVID-19 infection.

Employment data were also examined, and the correlation matrix is shown in table form in Figure 7. COVID-19 infection rate has a strong positive correlation with unemployment, and a strong negative correlation with employment. The trend for unemployment is shown in Figure 8.

	NeighbourhoodNumber	Covid19Rate	Covid19CaseCount	EmploymentRate	UnemploymentRate
NeighbourhoodNumber	1.000000	-0.160640	-0.027007	0.044850	0.091673
Covid19Rate	-0.160640	1.000000	0.730923	-0.518157	0.605126
Covid19CaseCount	-0.027007	0.730923	1.000000	-0.376483	0.486920
EmploymentRate	0.044850	-0.518157	-0.376483	1.000000	-0.765290
UnemploymentRate	0.091673	0.605126	0.486920	-0.765290	1.000000

Figure 7. Correlation matrix for employment and unemployment rates.

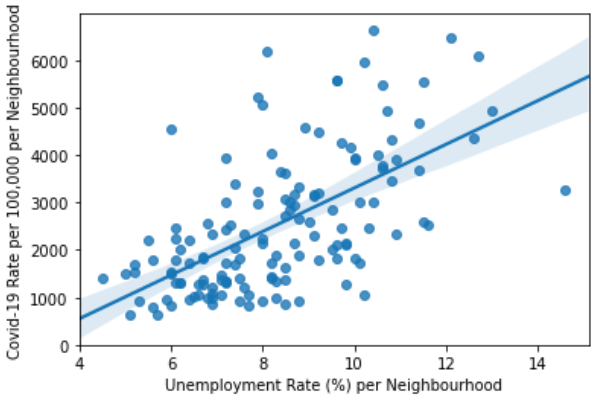

Figure 8. Regression plot of COVID-19 cases per 100,000 people vs. unemployment rate, showing a strong positive correlation.

The final demographic category considered was education level. Education was broken down into several bins, considering only adults aged 25-64. The categories and abbreviations are: no high school diploma (A25to64NoCert), high school diploma (A25to64HSdip), trade certificate or diploma (A25to64Trade), diploma from a non-university college or CEGEP (A25to64non-uniDip), university certificate or diploma below bachelor level (A25to64UniDipbelowBach), and university diploma of bachelor level or greater (A25to64UniBachorhigher). The heatmap of the correlation matrix is shown in Figure 9. Considering the row of Covid19rate, the highest positive correlations are associated with education less than high school and certified trades. University education bachelor or higher, has a negative correlation.

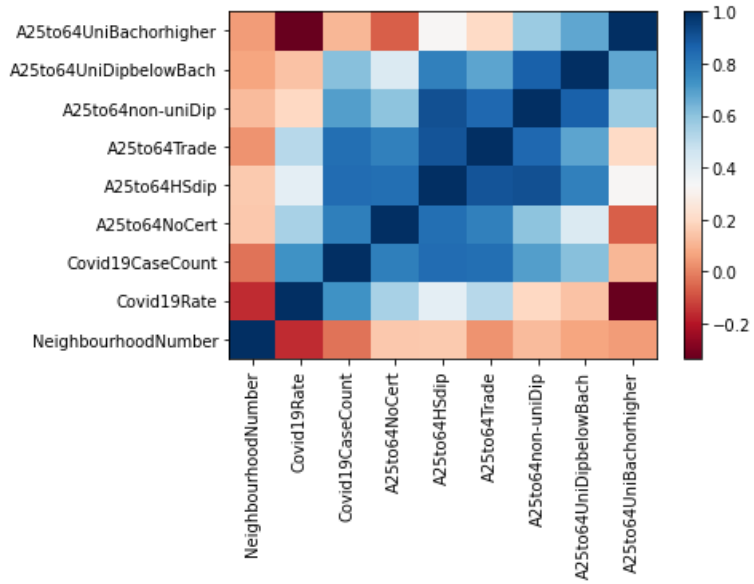

Figure 9. Heatmap of correlation matrix for education levels

Results

Neighbourhoods in Toronto with the highest and lowest number of COVID-19 infections per 100,000 people were identified in Figure 1. This is better visualized in the choropleth map shown in Figure 10. The highest rates are observed in neighbourhoods to the northwest, at the city limits. Case rates are also high in eastern neighbourhoods, and lowest in the centre and in the lakeside downtown core. Geography likely has no impact on this variation of COVID-19 infection, rather the demographic characteristics of neighbourhoods is relevant. Figures 11-15 are a series of choropleth maps constructed from the Toronto demographic data and should be compared with the infection rates by neighbourhood shown in Figure 10. The chosen figures help to describe the demographic characteristics associated (or not associated) with high rates of COVID-19 infections.

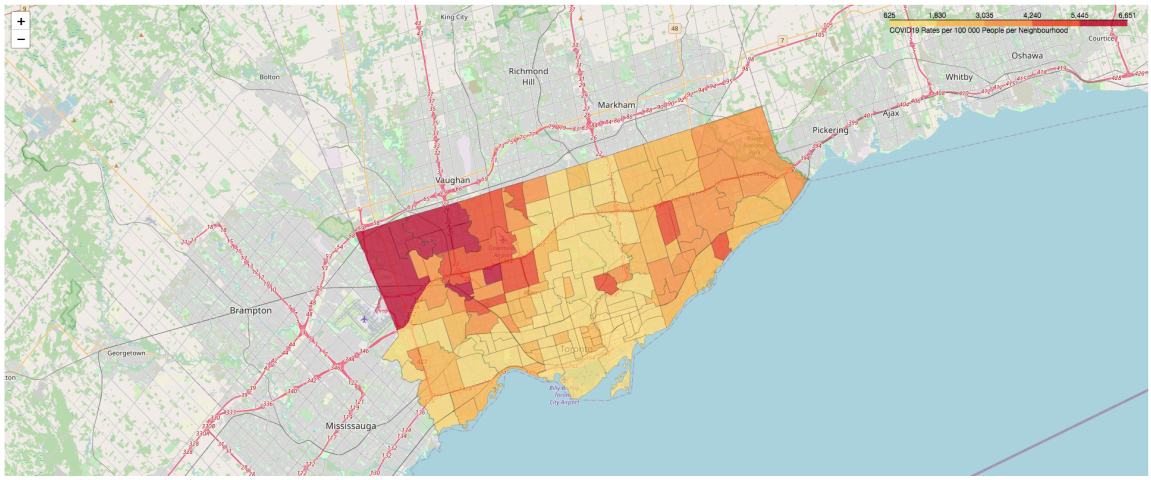

Figure 10. Choropleth map of COVID-19 infections per 100,000 residents of each Toronto neighbourhood.

Figure 11 shows the population of people who identify as visible minority Black for each Toronto neighbourhood. As was suggested in the statistical analysis, neighbourhoods with high populations of people who identify as visible minority Black generally have high rates of COVID-19 infection per 100,000 residents.

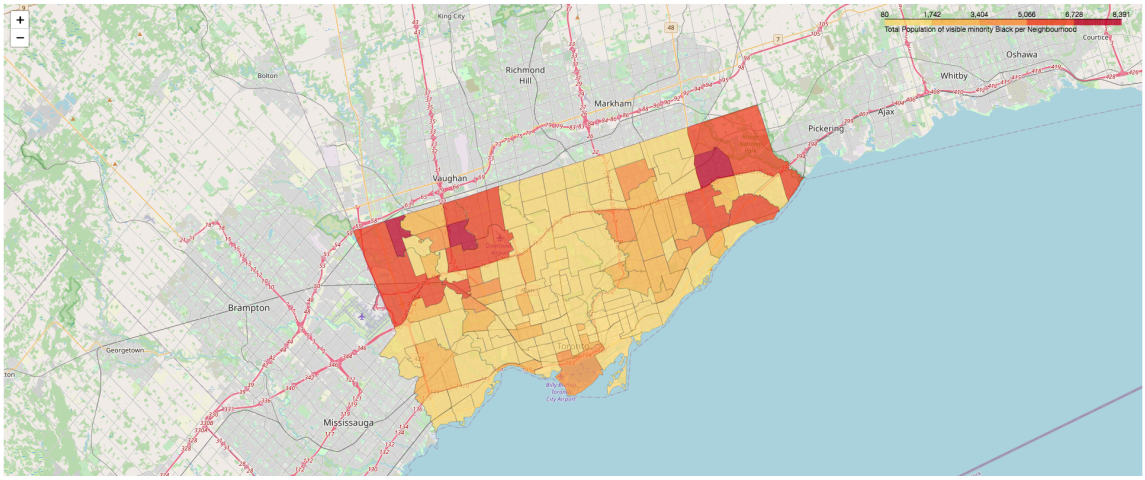

Figure 11. Choropleth map of population of people who identify as visible minority Black for each Toronto neighbourhood.

Figure 12 is a choropleth map for the population with total household income of $80,000 or greater for each Toronto neighbourhood. Neighbourhoods with high populations of wealthy people generally do not have high rates of COVID-19 infection. The number of people per neighbourhood who commute to work as a vehicle passenger (not public transit) is presented in Figure 13. This map resembles the COVID-19 infection rate map in Figure 20. Figure 14 shows the unemployment rate by neighbourhood. Many areas with high COVID-19 infection rates (Figure 10) are also regions of high unemployment.

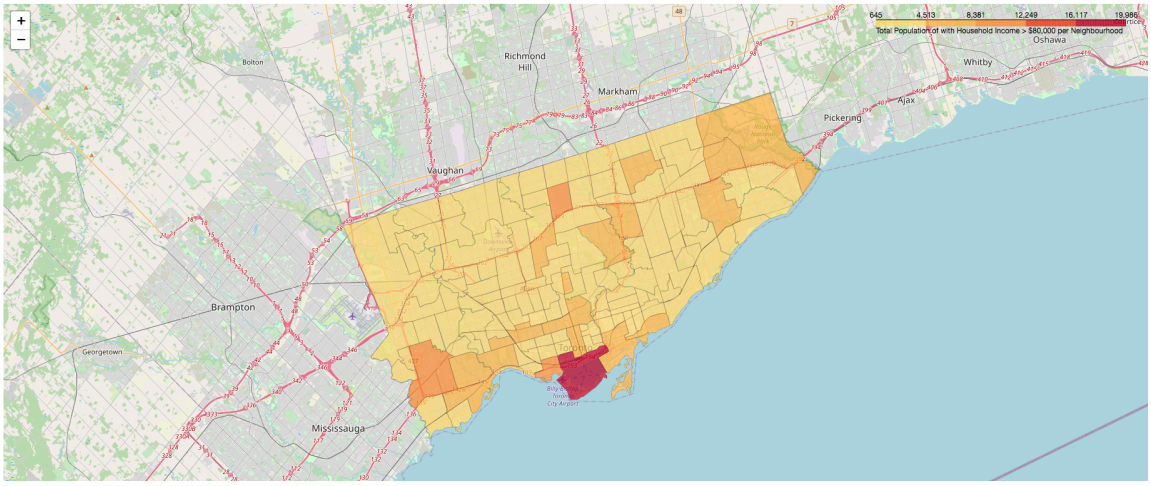

Figure 12. Choropleth map of population of people with total household income $80,000 and greater for each Toronto neighbourhood.

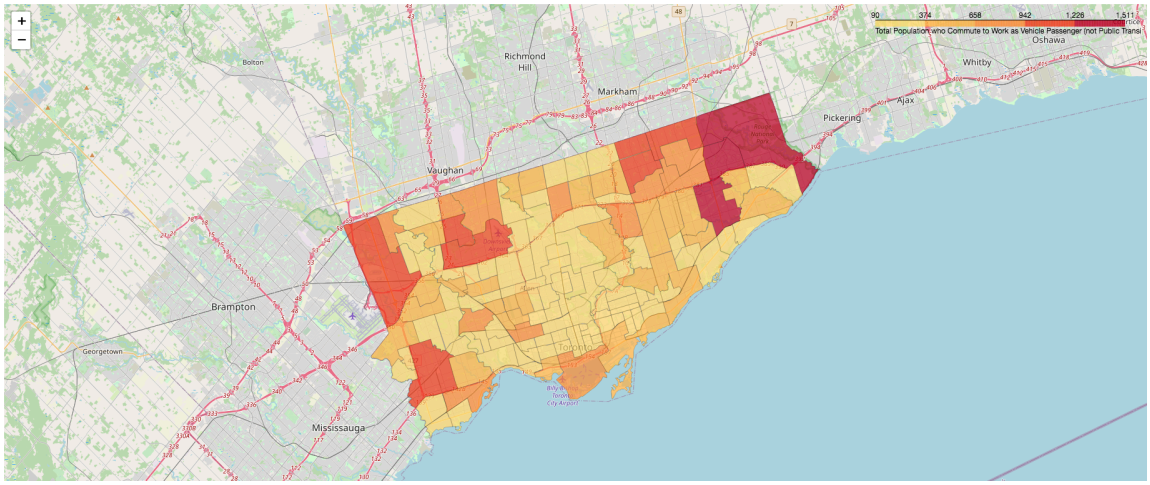

Figure 13. Choropleth map of population of people who commute to work in a private vehicle (not public transit) as a passenger for each Toronto neighbourhood.

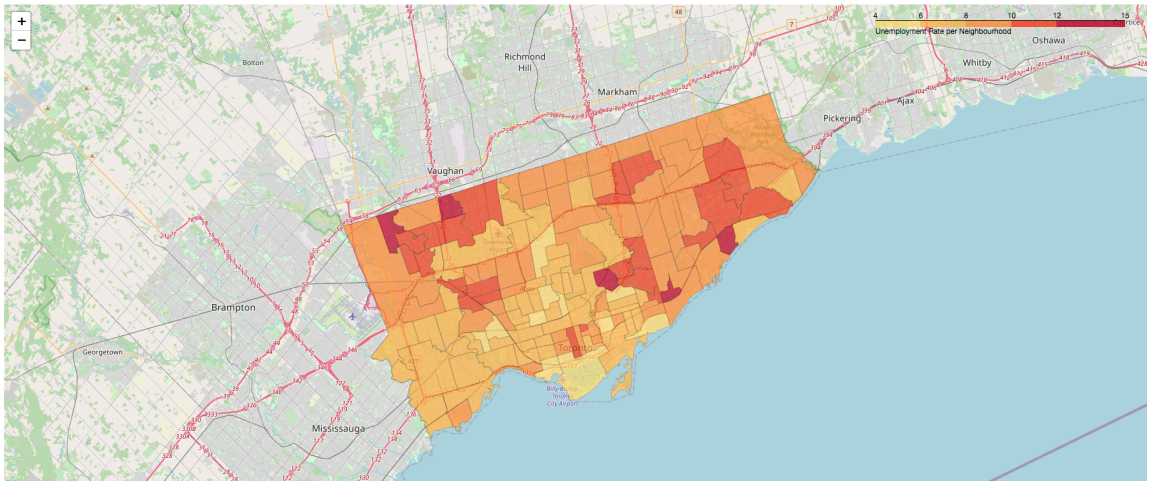

Figure 14. Choropleth map of unemployment rate for each Toronto neighbourhood.

Lastly, the number of people per neighbourhood with education level of Bachelor's Degree or higher is shown in Figure 15. The region of the most highly educated people corresponds with neighbourhoods of low rates of COVID-19 infection.

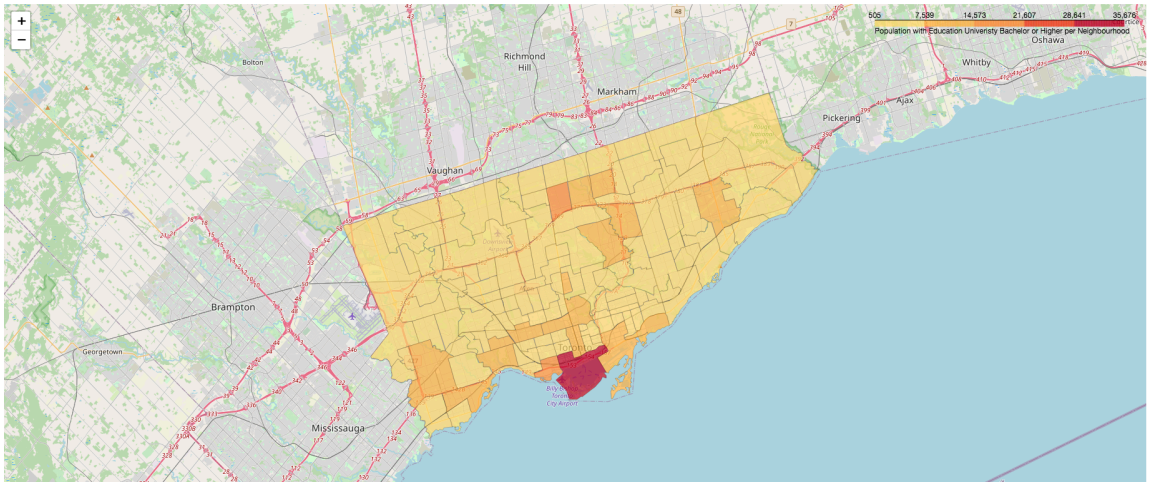

Figure 15. Choropleth map of number of people with university degrees at bachelor level or higher for each Toronto neighbourhood.

Discussion

A rich database was analyzed to understand the distribution of COVID-19 cases in Toronto. It was relatively easy to determine which neighbourhoods had the highest number of COVID-19 infections. The top five neighbourhoods in terms of infection rates are Thistletown-Beaumond Heights, Mount Olive-Silverstone-Jamestown, Maple Leaf, Black Creek, and Humbermede. These have been presented in table and map form.

Demographic data elucidated why infection rates were highest in these regions. Factors that strongly positively correlate with COVID-19 infections are common in neighbourhoods where infection rates are high. These factors include, but are not limited to, identifying as a visible minority black, commuting to work as a passenger in a private vehicle, and being unemployed. Conversely, factors that have strong negative correlations with infection rates were common in neighbourhoods where infection rates are low. These factors include high household income and high education.

Conclusions

The goal of this study was to help inform policy makers responsible for public heath measures. Regions with high infection rates of COVID-19 should be prioritized for restrictions aimed at reducing infections and for prioritized vaccinations. Based on this study, it is reasonable to conclude that neighbourhoods with the highest infection rates (Thistletown-Beaumond Heights, Mount Olive-Silverstone-Jamestown, Maple Leaf, Black Creek, and Humbermede) should be subjected to new health measures and prioritized for vaccinations.

Fortunately, the richness of the Toronto dataset allows for a more nuanced vaccination prioritization. This study suggests demographic groups that should be prioritized for

vaccination and those who should wait. *Vaccination priority groups* include those who identify as visible minority Black or Latin American, occupants of households with low to moderate incomes, those who commute to work as a passenger or for a long duration, the unemployed, and those with low education or who work in trades. Conversely, groups that should be *low vaccination priority* include visible minority Japanese or those who are not a visible minority, occupants of households with high income, and individuals who are highly educated.

References

[1] City of Toronto Daily COVID-19 updates

[2] Toronto neighbourhood profiles - census data

[3] Toronto Geojson