

A Spatial Understanding of Well-Being in the City of Toronto

By

Alexander Shatrov

Bachelor of Arts, Ryerson University, 2018

A Major Research Paper (MRP)

presented to Ryerson University

in partial fulfillment of the requirements

for the degree of Master of Spatial Analysis

in the program of Spatial Analysis

Toronto, Ontario, Canada, 2019

© Alexander Shatrov 2019

Author's Declaration

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP to other institutions for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

Acknowledgements

I would like to thank my supervisor, Dr. Brian Ceh. For his guidance and aid throughout the writing of this major research paper. I would also like to thank United Way Greater Toronto for graciously allowing me the usage of their data for the purposes of this MRP.

A Spatial Understanding of Well-Being in the City of Toronto, Alexander Shatrov, Master of Spatial Analysis, 2019, Ryerson University

Abstract

Social Capital is an emergent field of study that has the potential of applications in a wide variety of fields, from public health to economics, but most widely for human well-being. This study is a spatial and statistical analysis of the relation between commonly accepted indicators of well-being, commonly accepted indicators of social capital, and socio-economic factors within Toronto, as well as their respective spatial patterns. This research discovered that within Toronto there does not exist a clear link between wellbeing and social capital indicators, both in terms of statistical regression analysis and spatial pattern comparison. The results suggest that, at least within the context of Toronto, much of the research surrounding the effects of high social capital on wellbeing are not applicable, at least using the methods demonstrated in this study.

Table of Contents

Abstract	iv
List of Figures	vi
List of Tables	vii
1. Introduction	1
1.1 Research Objectives	1
2. Research Context	1
2.1 Social Capital Definition	1
2.2 Social Capital and Health	3
2.3 Social Capital and Civic Engagement	5
2.4 Social Capital and Economic Success	6
3. Data/Methodology Overview	8
3.1 Data Origin and Limitations.....	9
3.2 Data Cleaning	10
3.3 Analysis Process	11
4. Findings	12
4.1 Spatial Autocorrelation of Life Satisfaction.....	12
4.2 Spatial Autocorrelation of Self-Reported Health	15
4.3 Spatial Autocorrelation of Income	18
4.4 Other Relevant Variables	27
4.5 Regression Analysis.....	29
5. Conclusion	33
References	35

List of Figures

Figure 1: Average Respondent Number Histogram	10
Figure 2: Q35 Average	12
Figure 3: Q35 Moran's I	13
Figure 4: Q35 Cluster Map	14
Figure 5: Q36 Moran's I	15
Figure 6: Q36 Cluster Map	16
Figure 7: Q36 Average	17
Figure 8: Q58 Moran's I	18
Figure 9: Q59 Moran's I	19
Figure 10: Household Income Moran's I	20
Figure 11: Q58 Cluster Map	21
Figure 12: Q59 Cluster Map	22
Figure 13: Average Household Income Cluster Map	22
Figure 14: Census Tract Average Household Income	24
Figure 15: Q58 Average	25
Figure 16: Q59 Average	26
Figure 17: Q6 Average	27
Figure 18: Q25 Average	28

List of Tables

Table 1: Variable Descriptions	20
Table 2: Euclidian Distance Spatial Autocorrelation Results	21
Table 3: Variable Descriptions	29
Table 4: Model Summary	30
Table 5: Coefficients	30
Table 6: Model Summary	31
Table 7: Coefficients	31

1. Introduction:

Social Capital, defined as the network of relationships between people in a society that allows that society to function effectively, has an effect on everything from health, to life satisfaction to civic participation. Its subjective nature often makes it difficult to measure, which in turn makes research difficult, despite its importance. When research does happen, it rarely looks at the spatial context (in fact the study used for this paper is the first social capital survey ever conducted in Toronto). Toronto, as the largest city in Canada and one of the most multi-cultural in the world, benefits from social capital as much as anyone, perhaps even more so, as having a network of support amongst members of a community has the potential to be a critical element of an immigrant's success at building their life in Canada. However, little effort has been taken to seriously study social capital in the context of Toronto's environment, and even less to look at it from a spatial perspective, until the 2018 social capital survey.

1.1. Research Objectives

Using the data gathered during this survey, this research paper will have 2 research objectives.

1) To determine the spatial pattern of low or high social capital indicators across Toronto and how they compare to a spatial distribution of life satisfaction scores.

2) To determine if selected questions from the survey that are thought to be effective at predicting life satisfaction can do so with the data found in this survey. The questions chosen from the survey that will be selected with the help of a review of the existing literature.

3) To determine if census tract data measuring the same or similar variables is equally effective at predicting life satisfaction.

The hypothesis is that there will be a strong correlation between life satisfaction and at least some of the other survey questions and census data that will be looked at.

2. Research Context:

2.1. Social Capital Definition

Social Capital is a difficult subject to study, due to its subjective nature, with many researchers not even entirely agreeing on what it is. One of the more well-known works on the

subject, Robert Putnam's "Bowling Alone: America's declining social capital" defines it as a "features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit", while Francis Fukuyama (2000) describes it as an "informal norm that promotes cooperation between two or more individuals". Alejandro Portes (2000) argues that there are two separate definitions of social capital, one being a "feature of individuals" and the other being a "feature of communities". Social capital as a feature of individuals refers to social capital that takes the form of, for example, social connections that allow people to find jobs more easily, or social pressure that allows for reduced crime and therefore less danger for an individual walking the street at night. In short, social capital as a feature of individuals refers to social capital that benefits that benefits an individual's community. Social capital as a feature of communities on the other hand sees social capital a resource that can be harnessed by a government or community for its own betterment, with those communities that have higher "stocks" of social capital reaping benefits such as better governance and a better educated population. This is a view shared by Nan Lin (2017) who defines social capital as "investment in social relations with expected returns", but puts forward two opposing theories as to whether this investment and return is gained by individuals or by groups. They synthesize the two by describing social capital as something that is separate from the exact nature of their usage. Instead, it is described as containing "three elements intersecting structure and action: the structural (embeddedness), opportunity (accessibility) and action-oriented (use) aspects". This conceptualizes social capital as a resource to be obtained and used, with the exact nature of the use and users being not important to the overall definition. Putnam's instrument theory also identifies density as a key aspect of social capital accumulation. High density cities where a large number of voluntary organizations are created and organized are considered the most "productive", or getting the most use out of their social capital (Paldam, 2000). There is some criticism of Putnam's definitions of social capital, with some researchers claiming that interchangeably using common and individual good when talking about the positive aspects of social capital is incorrect. The same issue is found with the tendency of some social capital researchers to look at individual indicators of social capital and then scaling them up and expecting them to act the same. In particular, the idea that a community can "possess" social capital is criticized, as while social capital can be held by individuals and by institutions/other formal groups, the idea of a community is too abstract to assign it properties

like having social capital (DeFilippis, 2001). Even the narrowest definitions still leave a wide range of things that can be considered “social capital”, and many researchers use social capital interchangeably with social cohesion, or see the former as a subset of the latter but all of them seem to agree on social capital being largely defined by its usage and benefits as opposed to some inherent quality of itself (even the name social capital demonstrates its relationship with more traditional economic activity). However, this inability to agree on what exactly social capital actually is hinders its usage in further research, as different studies that are looking at the relation between social capital and another variable could arrive at vastly different conclusions not because of differences in methodology or study area, but simply because they are using two different definitions of social capital. This heavily impacts the ability of social capital researchers to build off of the work of others as different studies can be difficult to compare (Macinko & Starfield, 2001).

2.2. Social Capital and Health

One of these benefits, and perhaps the most easily measured is the effect on health, with high social capital being tied to better health of the people in question (it should be noted that while many social capital studies focus on the health benefits of interpersonal relationships, some, including the Toronto study being used for this paper consider direct access to healthcare an aspect of social capital and even explicitly asked people if they have access to medical care during their survey). One of the more famous examples of this is the Roseto Effect. Roseto is a town in Pennsylvania that, at the time the pattern was noticed, was overwhelmingly Italian-American, and was noted to have a much lower rate of heart attacks than the adjacent town of Bangor and other nearby communities, despite being serviced by the same hospital and physicians as those towns (Egolf, Lasker, Wolf, & Potvin, 1992). In addition, the unusual health of this town could not be explained by economics, as many of the residents worked in dangerous and toxic quarries, and were not noted to be particularly health-conscious, with smoking and consumption of high-cholesterol foods being common. Despite this, a comparison of death records showed that Roseto had a significantly lower rate of deaths as a result of heart attacks over an observed period of about 30 years, until 1965 when the rates began to grow more similar to other towns in the region, theorized to be a result of cultural changes that began to occur in Roseto around that time that caused social capital to decrease as the population adopted a more

“standard” American culture. Considering that Toronto is different from a small homogenous town, it may seem like the findings of the Roseto study may not be applicable, but numerous other studies have found that these kinds of health benefits can appear in a wide variety of environments. For example, a study performed by Kawachi & Berkman (2000) looked at mortality rates across 39 US states and performed a survey that asked participants about their membership in various organizations such as sports clubs, hobby groups and religious organizations, with a high participation rate found to strongly correlate with a lower rates of heart disease, tumors and even infant mortality. The participants were also asked questions about the level of trust they had in people in their community as well as strangers, with a regression analysis finding that variations in the levels of trust explained 58% of the variance in mortality rates. Self-rated health was also lower in states with lower social capital, even if other indicators of public health like obesity rates, smoking, etc. were identical. It should be noted that the questions asked in the survey that was compared to mortality rates are identical to some of the questions asked in the 2019 Toronto social capital survey that will be used as the basis for the analysis in this report. While that study was performed in a different and more varied environment, there may still be similarities. A similar study was performed in Saskatchewan, with frequency of socialization with co-workers and religious attendance being highly correlated with better health. This was followed closely by attendance of clubs and other forms of association, with a particularly pronounced effect on the health of the elderly population (Veenstra, 2000). There are different views on how exactly social capital relates to health. Some studies put forward the idea that the two are directly related in the form of higher social connection which leads to less stress and anxiety which in turn results in a lower frequency of health issues like heart attacks. Others put forward the idea that social capital affects health in a less direct way, such as higher amounts of social connections giving individuals something to do thus causing them to engage in fewer self-destructive behaviors, as well as giving people connections that are willing to help them out when health problems arise (Szreter & Woolcock, 2004). The opposite may be true if those connections result in social pressure to engage in unhealthy behavior however. In particular, one case study used to illustrate the latter point was the 1995 heat wave in Chicago which led to 739 deaths. One group that was hit particularly hard by the death toll was the Males over 65 group, and while a part of this was due to economic factors (many of the victims were poor and could not afford air conditioning) and age (the

elderly are generally considered to be more vulnerable to extreme temperatures), investigators noted that a lack of social trust may have contributed to the death toll, as many of the victims refused to leave windows open for fear of crime, and many of them were living alone, which meant that there was less of a chance for them to be found and transported to a hospital in time for treatment (ibid). The health benefits are not limited to physical health, but mental health as well. For example, loss of social ties has been found to be a leading cause of depression among the elderly, while higher social engagement has been found to correlate with lower rates of dementia and other forms of cognitive decline in both men and women over 65 (Kawachi & Berkman, 2001). This is both due to the positive effects of the social connectivity itself, and because individuals with high social capital are likely to be referred to a mental health professional or other form of intervention due to their issues being noticed sooner. However, unlike with physical health, there is also a negative side to high social capital with mental health. A study performed in the Outer Hebrides islands among heavily tight-knit communities found that while the residents, particularly the women did indeed suffer from lower rates of depression, they also suffered from higher rates of anxiety. This has been theorized to be because the kind of social capital that is provided by small tight-knit communities carries with it a pressure to conform, which has been tied to increased anxiety within the populations subjected to such pressures (ibid). It should be noted that the benefits of social capital, or at least elements of social capital, on health, have been researched widely enough that many organizations will openly reference them. For example, the government of Canada's own website has a page entitled "Social determinants of health" where they describe social status and social support, both key aspects of social capital theory as key determinants of health ("Social determinants of health", 2019). The fact that even governments are beginning to take this research seriously should show how much evidence there is behind these indicators.

2.3. Social Capital and Civic Engagement

Social capital has also been found to have a large effect on civic engagement and the health of democracy. In a 20 year study in Italy, Robert Putman (2000) found that there was a strong correlation between the social capital of a region, particularly the number of voluntary groups a person participates in, and the effectiveness of the local government. Areas with high social capital had governments that were efficient and creative in their solutions to various

problems, while areas with low social capital had governments that were more prone to corruption and low effectiveness. Putnam explained this by theorizing that high social capital in a population allows that population to more easily engage in collective action for their mutual benefit, making it easier to put pressure on poorly performing governments than in areas where the population was more insular and mistrustful of each other. In addition, both Putnam (2000) and others have noted that high social capital is correlated with higher voter turnout and political participation. Francis Fukuyama (2000) believes that social capital even has an effect on economy and science (beyond the direct economic effects of good health and effective government). According to him, an integral part of scientific advancement is informal information sharing, something that does not happen as often in an environment of mistrust (low social capital). In addition, an environment of high social capital supposedly results in a greater degree of fluid self-organization amongst workers which is better able to meet local challenges than a company relying exclusively on top down formal management. It has also been found that on average, the size of a person's social network is directly correlated with the spread of political information, and therefore political awareness of some kind in the individuals involved in the social network (La Due Lake & Huckfeldt, 1998). However, this study also found that this increase in political participation was much more noticeable in those whose social networks involved a large and diffuse group, as opposed to the tight and cohesive groups that have been identified as the most "optimal", which presents a unique challenge as it is clear that there is no one universal form of social capital formula that everyone can follow for the most optimal results. In a more abstract sense, social capital may also be beneficial to a democratic society because individuals or groups with higher levels of social capital may extend this trust not just to others within their immediate surroundings, but implicitly extend this even to people they have never met. This helps in promoting democratic participation and trust-based politics not just on a local but on a national scale (Newton, 1997).

2.4. Social Capital and Economic Success

Another study performed in several parts of the US looked at the relationship between social capital and economic success of a community in the form of poverty rates (Rupasingha & Goetz, 2007). This study found that most regions with higher amounts of social capital tended to have lower family poverty rates, but that this only applied to those regions that were not "metro

areas”, meaning counties with more than a million population, with the effect of social capital on poverty rates within metro areas seemingly being close to zero, and that there is no conclusive evidence that lower social capital results in lower poverty rates as opposed to the other way around. This is important to note as income is one of the demographic variables that was gathered by the Toronto study, and as such, according to the findings of this study, there should be little correlation between the social capital score and reported income in various parts of Toronto. However if this turns out to not be the case, it will be difficult to determine if the nature of the relationship between poverty and social capital in metro areas changed since the study was performed, as it is also possible that Toronto is an outlier. This can be mitigated by performing similar social capital surveys in other cities, but is beyond the scope of this paper. There is reason to believe that it may not in fact be the case, as another study concluded that economic well-being of a region actually improves social capital, with higher rates of employment and home ownership being correlated with higher social trust (Iyer, Kitson, & Toh, 2005). It has also been found that ensuring a high amount of social capital may be a vital component in the economic development of undeveloped/developing nations, particularly of their rural areas that often go underfunded even when the country as a whole is economically growing. This is because social capital provides a framework for economic development, as it allows for a more rapid and widespread exchange of information, improve commerce in areas that lack legal infrastructure, and promote responsible behavior over short-term profiteering (Woolcock & Narayan, 2000). But as with health, there may be a dark side to the economic benefits of high social capital. One of the examples often cited as a benefit of high social capital is the economic success that high social capital is considered to bring to communities that can rely on it to aid in job placement and such. However, some researchers claim that the kind of high social connectivity that is crucial for social capital to develop may actually hinder the economic success of newly arrived groups, as the existing social networks may close themselves off to protect the job security of existing members, hindering the economic success of the newly arrived groups (DeFilippis, 2001). There may also be an environmental component to the political benefits of social capital, or more specifically benefits to a community’s ability to adapt to and withstand environmental calamities. A study published in the *Global Environmental Change* journal claims that the social ties that can be found in communities and populations with high social capital contribute to climate change adaptation, both because individuals in such societies are more

likely to be aware of the problem and more politically active in a way that will allow them to effectively pressure the government or other organizations to tackle the issue, and because they will be more likely fight climate change themselves through a collective effort, such as through the organization of non-profit organizations and projects (Pelling & High, 2005). The study also touches on how high social capital environments result in greater knowledge sharing and managing of individual skills, which would benefit already existing organizations dedicated to fighting climate change.

The data that will be used in this research paper has already been used to write an overall report describing the state of social capital in Toronto, focusing on the relationship between rates of different responses and socio-demographic groups (for example, rates of high or low trust in police amongst different racial groups). This report also includes the creation of spatial indexes, with neighborhood groups consisting of multiple Toronto neighborhoods amalgamated into larger units. These units are assigned index values made up of the answers of the respondents who live in those neighborhoods, with the indexes being divided into Social Trust, Social Networks, Civic Connection and Neighborhood support (Toronto Foundation, 2018).

3. Data/Methodology Overview:

The data that will be used for this research paper is the 2018 Toronto Social Capital survey performed by the Environics Institute for Survey Research. This survey contains 70 questions asking participants about their demographic characteristics such as age and race, before asking them to describe the number and quality of social connections, trust in various institutions, overall life satisfaction and other questions. A research report describing the overall demographic patterns across the entirety of Toronto has been published, but as of yet the data has not been used for any kind of spatial analysis (or none that is publically available at least). Not every question will be looked at in this research paper, rather a few select ones that a review of existing literature suggests are the most important will be looked at. The spatial distribution of social capital will be examined using several methods. First, a hotspot analysis was run to determine which census tracts within Toronto contain the highest and lowest values of social capital indicators, followed by a standard deviation analysis to see if findings differ.

3.1. Data Origin and Limitations

The survey data was gathered through telephone calls and online questionnaires, with the phone calls having location-based quotas to ensure an even spread of respondents. These quotas were based on neighborhood profiles and were compiled with the intention of deliberately targeting marginalized population groups. After the initial round of surveys, secondary surveys were sometimes conducted to complete unfinished quotas. There are several limitations in this study. First, the survey only managed to capture just over 3200 respondents. When those who did not give their postal code or gave one that was outside the boundaries of Toronto and were therefore unusable for the analysis were removed, that number goes down to 3086. Secondly, there is a very large bias towards older populations, with people who are 65 years or older representing a quarter of the respondents, while those aged 18-24 make up just under 5%. Of the 3069 respondents who gave their age, the average age was 51.6 years old. There is also the limitation of all of the answers being self-reported as well as somewhat subjective and based on individual interpretation of a person's life, but that is difficult to avoid when dealing with something like social capital. After the data was obtained and the unusable postal codes were removed, each respondent was geocoded into a spatial point in accordance with their postal code. However, the kind of spatial analysis that was needed could not be done with point data, and as such, the respondent points were "dissolved" into the census tracts by calculating the average response for each census tract from the points geocoded inside of it. It should be noted that not all census tracts had multiple respondents, and there were even a few that had no respondents and were excluded from any subsequent analysis. In addition, despite the best efforts of the survey, there is an imbalance in the number of respondents from each CT, as can be seen below.

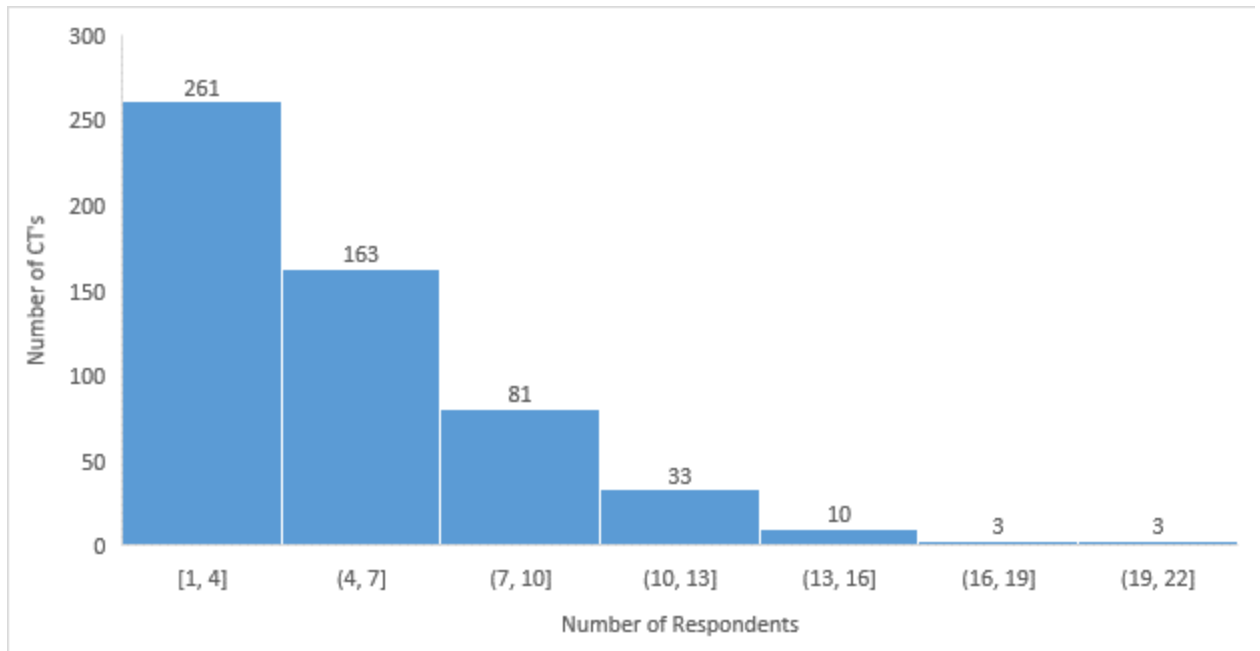


Figure 1: Average Respondent Number Histogram

The majority of CT's have 4 or less respondents, which puts some doubt on their ability to accurately represent their respective census tract.

3.2. Data Cleaning

Before that could happen, several issues needed to be solved. While most of the questions followed the format of “higher is better” eg answers that are considered better are coded as higher numbers while worse answers are coded as lower numbers, a few questions had this pattern in reverse. All such questions were recoded to consider low numbers as worse (eg if a question had options from 1 to 5 where 1 was considered the best and 5 the worst, it was recoded so that the best would be 5 and 1 the worst). In addition, another problem arose. The survey questions rarely had more than 5 answers, with a few open ended ones going as far as 10. However, if a person answered that they did not know or were not sure about an answer, the survey coded their answer as a 99. This presented an obvious issue as such a high number would heavily throw off census averages when they are calculated. As such, all responses were recoded to be multiplied by 10 (so if a person gave an answer corresponding to a 2, their value for this question was now 20) while all 99's were recoded as 1's. While this does bring a small degree of uncertainty into the average calculations, this was done in order to minimize the influence these answers would have on the outcome. For this change an assumption was made that a person who

answered cannot say on a question would likely give an answer that is considered poor or negative. For example, it was reasoned that someone who has a large number of close friends would very likely indicate as such on the report, while a person who does not have many close friends is more likely to answer cannot say as they attempt to remember if any of their acquaintances count as close friends, or simply because they did not want to disclose such a thing to the person running the survey. The rest of the numbers being recoded as multiplied by 10 was done both to ensure that the cannot say answers that were recoded as 1 do not get mixed in with those answers that were coded as 1 in the questionnaire. This also ensures that the higher values would allow for patterns in the data to come to the surface more easily. From there, the census tract averages were calculated for each question, and joined with census data from Statistics Canada. They were then run through a regression analysis, hotspot analysis and spatial auto correlation.

3.3. Analysis Process

Once this was accomplished, the next step was to perform Spatial Autocorrelation analysis to determine if there are spatial patterns to the survey responses. This was performed in the GeoDa software, where the spatial unit weight was created using the queen contiguity method (The rook contiguity method was considered, but it was found that there were no significant differences between results from the two methods). From there, a spatial autocorrelation was performed on the census tract averages of the measured variables, as well as census data, such as data by average household income per census tract in order to compare the spatial distribution of self-reported income and actual census income data (household income was used instead of individual income so as to be more easily compared to the survey questions which also asked specifically for household income). It should be noted that while the survey did ask people to provide their actual income range, it also asked a question where they described their income status using somewhat subjective terms like “Struggling” or “Good enough”. ArcGIS was used to create hotspot maps of the same variables, with the weight method being Euclidian distance as opposed to queen contiguity.

4. Findings:

4.1. Spatial Autocorrelation of Life Satisfaction

The first aspect of analysis, spatial auto correlation, proved to have some results, although nothing particularly concrete. Before the actual spatial autocorrelation was performed, maps of the census averages for each of the questions were created for context and potential comparison. For context, the question being looked at here (Q35) asks the respondents to rate their life satisfaction from 1 to 10.

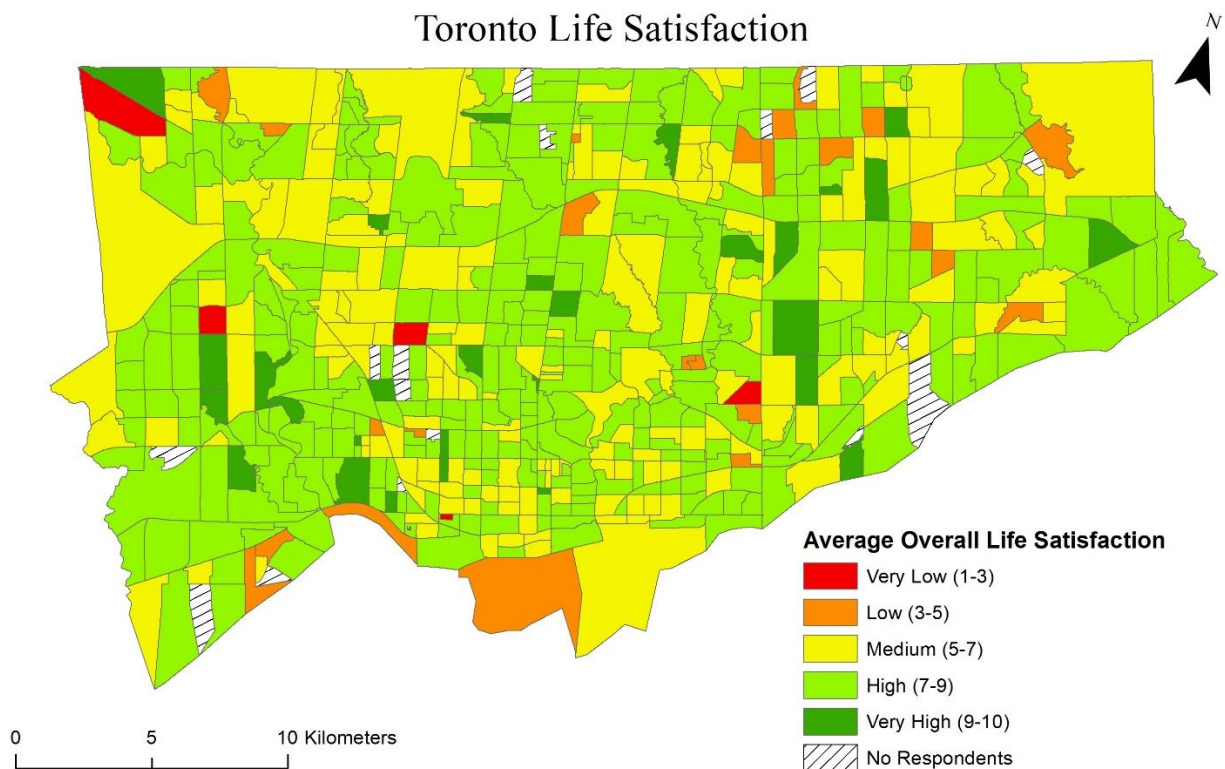


Figure 2: Q35 Average

When looking at the average life satisfaction across the census tracts, it becomes clear that at least among the respondents, the average life satisfaction is actually quite high. There are exceptions of course, but on average the census tracts with the lowest life satisfaction also have the lowest number of respondents, making it safe to assume that had these census tracts had more respondents their own averages would have gone up as well.

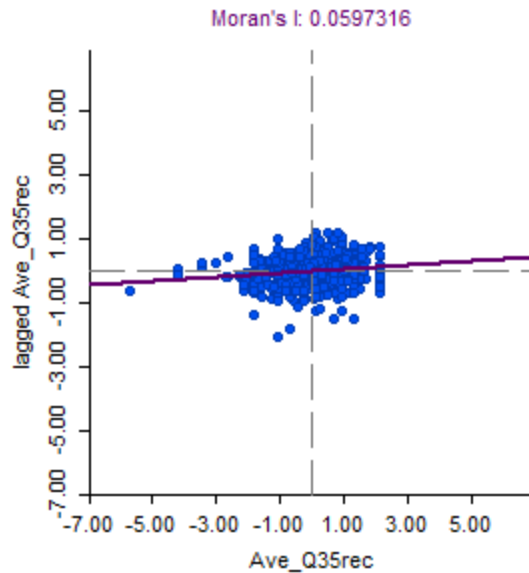


Figure 3: Q35 Moran's I

As can be seen, there is very little spatial autocorrelation in the average life satisfaction of Toronto, with the exception of a few particularly low outliers. It is likely that these outliers are the same ones that have been classified as Very Low in figure 1. This does seem to indicate that there is no noticeable spatial pattern to low life satisfaction in Toronto, and the outliers seen above can be explained not by those census tracts having a very low average life satisfaction, but by the fact that the survey was only able to get one responded from that census tract who also happened to be unsatisfied with their life.

The Z score for this spatial analysis was 2.4363 and the P value was 0.013. Despite the almost random distribution of values, the spatial autocorrelation process nonetheless indicates that there is a cluster pattern found in these variables, and a less than 5% chance that it is the result of random chance. This may be an indication that, although the overall sample is randomly distributed, there are internal clusters that are difficult to see through the overall scatterplot. To see if this is the case, a hotspot map can be created.

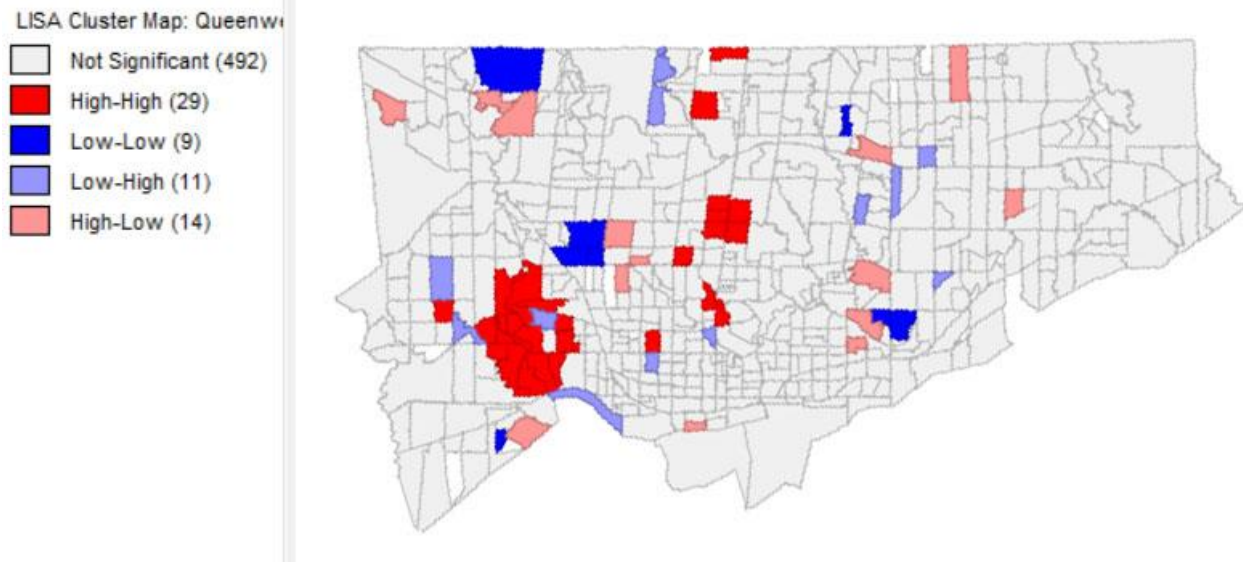


Figure 4: Q35 Cluster Map

As can be seen, there is a large cluster of high life satisfaction census tracts located approximately in South-Eastern Etobicoke, which may explain the relatively high Z score seen earlier. Other hot and cold spots are scattered around Toronto, but they do not appear to have any significant patterns, with most of these clusters being either individual census tracts, or small collections of 2 or 3. Most interestingly, none of the census tracts that were identified as having the lowest average life satisfaction are marked as not significant on the map, although given their infrequency this is likely the result of them simply not being common enough to have a significant effect on the calculations.

4.2. Spatial Autocorrelation of Self-Reported Health

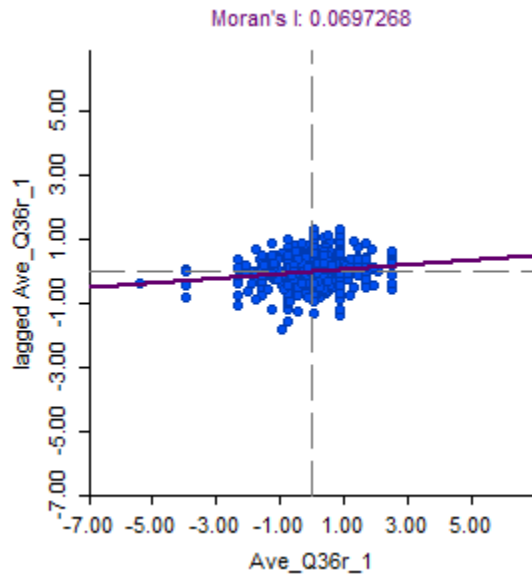


Figure 5: Q36 Moran's I

The question being looked at here (Q36) asks the respondents to describe their health, with their 5 options being poor, fair, good, very good and excellent. As with life satisfaction, self-described health seems to have an almost completely random spatial distribution. The few outliers are likely the result of those census tracts that had particularly low average health scores. However, when looking at these census tracts for more information, they are universally those census tracts that only had 1 responder

The Z score for Q36 was 3.1478, and the P value was 0.003, which is even higher and lower respectively than it was for life satisfaction, and there is a less than 1% chance that any clustered pattern that is found within it is the result of random chance.

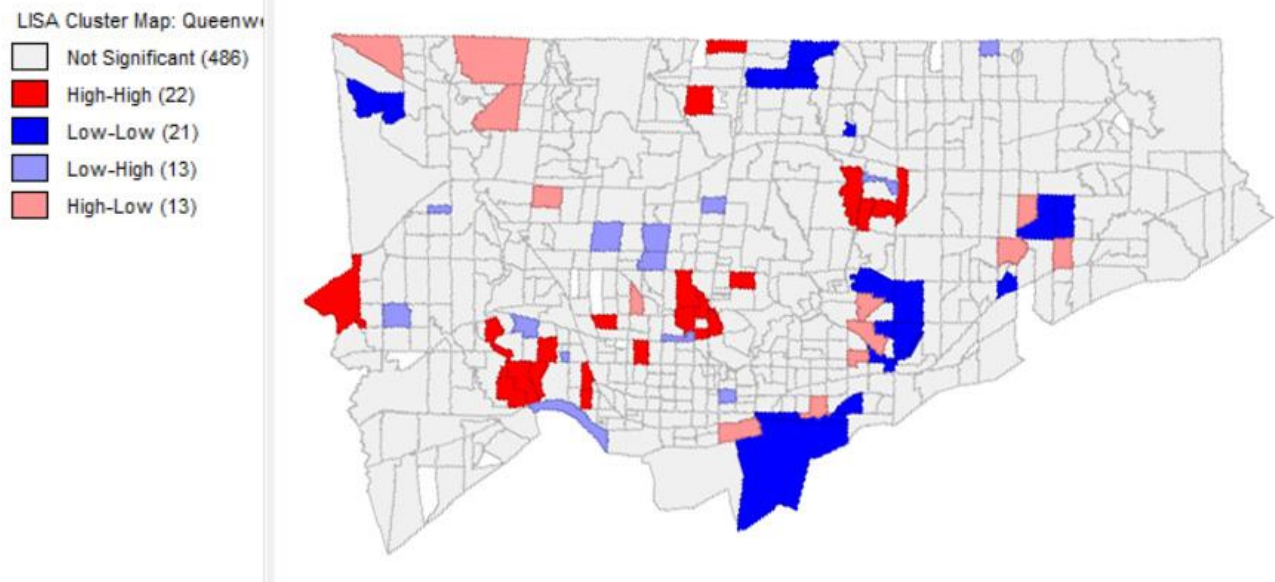


Figure 6: Q36 Cluster Map

Health status differs from the pattern found in figure 4, as here rather than one fairly large high cluster surrounded by seemingly random positioning of individual high and low census tracts, this cluster map shows the presence of several smaller clusters, both high and low. One of the high clusters is located roughly in the same region as the large high cluster from figure 4, with several more in central Toronto and roughly near on valley village, while a low cluster can be found the north of East York.

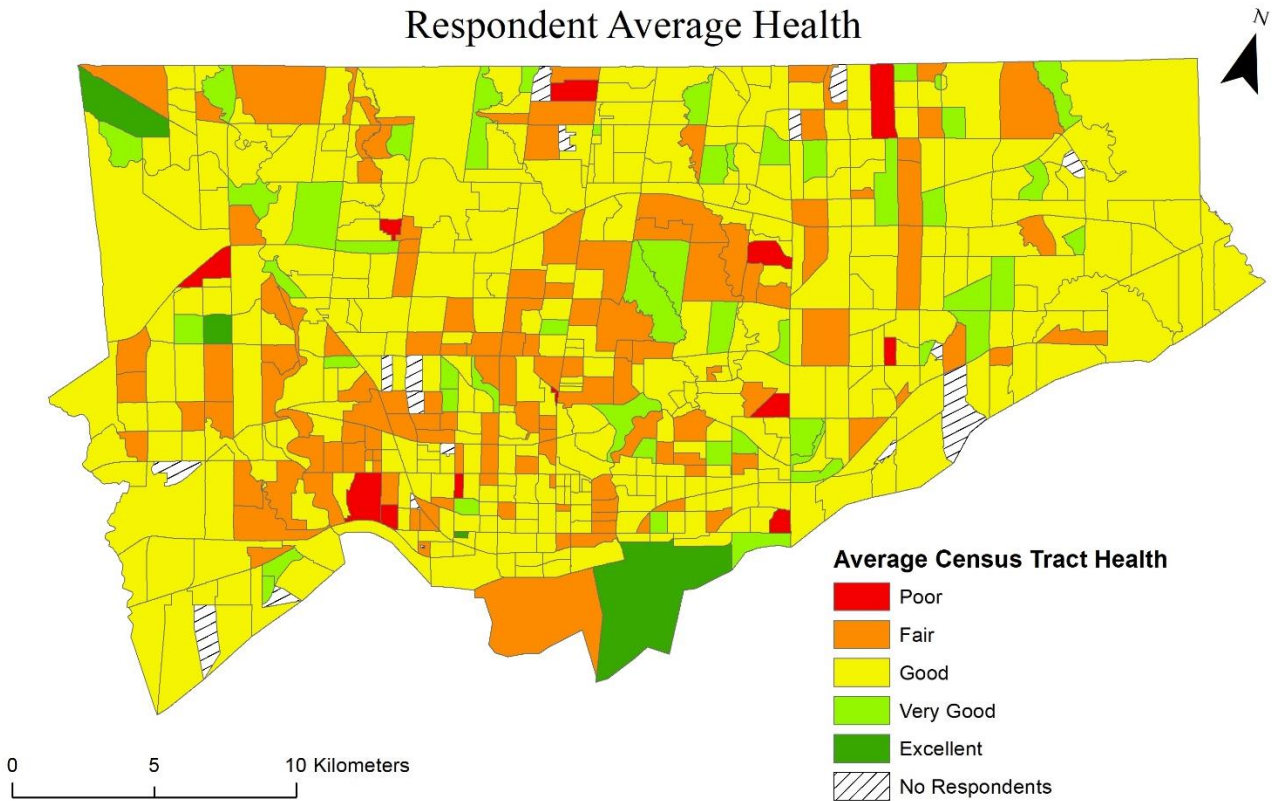


Figure 7: Q36 Average

Looking at this map it is apparent that a majority of the respondents consider themselves in good or better than good health, and what's more, the spatial pattern of those who described their health as poor or fair is not particularly consistent with the distribution of those who responded that they are unsatisfied with their life as seen in figure 1.

4.3. Spatial Autocorrelation of Income

The survey questions looked at for this section are the household income of the respondent, separated into 6 categories (Q58) and their self-described income status, with the 4 options being not enough and you are having a hard time, not enough and you are stretched, just enough and enough that you can save (Q59). The average household income for each census tract from Statistics Canada will also be used.

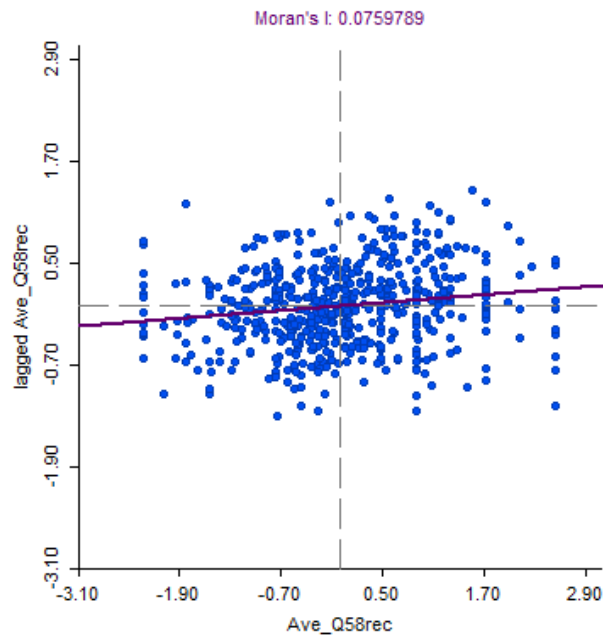


Figure 8: Q58 Moran's I

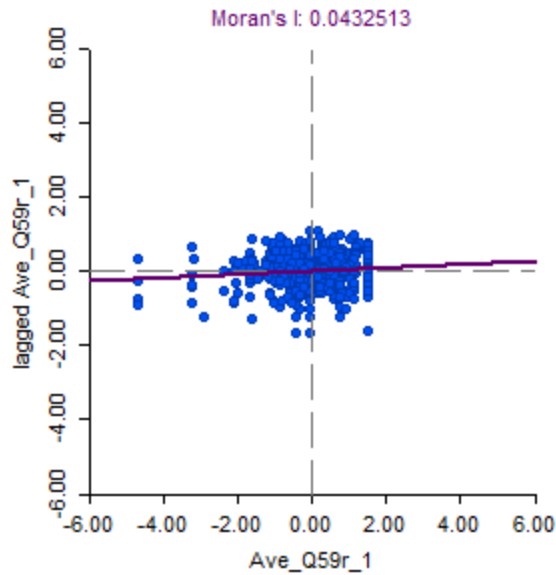


Figure 9: Q59 Moran's I

The Moran's I value continues to be extremely close to zero with every survey question on which spatial autocorrelation has been performed, suggesting a universal pattern of the responses having no discernable spatial pattern. What should be noted however, is that the scatter plot for Q58, appears a lot more spread out then the scatterplots for the other variables. This is not the result of the answers actually being more spread out, rather it is the result of this question having noticeably fewer outliers then the others. However, when compared to the scatter plot for Q59 which does have a number of outliers, it becomes clear that different respondents consider the same income category to be of different quality. This is logical, as the living costs of individual households differ significantly even when located within the same income category. For example, someone who lives in a relatively cheap part of the city or lives without dependents may find their income category to be enough to live on. At the same time, someone who lives in a more expensive part of the city, have a larger household or some kind of ongoing medical issue may struggle to meet their needs with the exact same income.

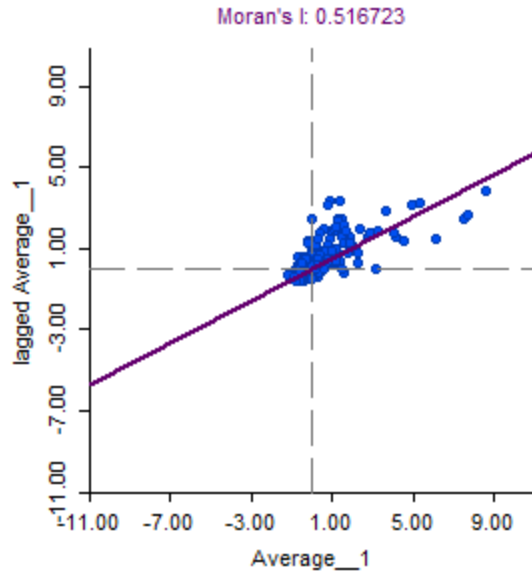


Figure 10: Household Income Moran's I

Household income as a variable for 2016 has a higher Moran's I value, which suggests that household income is more clustered than the survey found. This lack of clustering in the survey responses when compared to actual household income statistics is likely a result of the low number of survey respondents. As mentioned earlier, there were several census tracts that only had one respondent, while others may have only had a relatively low number like 3 or 5. What this means is that the survey simply does not have an accurate representation of average income on a smaller geographic level. In addition, the survey question did not ask people for their exact household income. Instead, the respondents were given a number of income ranges (IE \$60,000 to \$80,000) to choose from. This makes it difficult to assess the accuracy of even those census tracts where the survey was able to get a large number of respondents.

Table 1: Variable Descriptions

Variable	P Value	Z Score
Q58	0.003	3.1906
Q59	0.045	1.7901
Average Household Income	0.001	21.5107

The process of spatial autocorrelation was performed again using the Euclidian distance method as opposed to the previously used Queen's contiguity method to determine if there are

any noticeable differences. For the variables that were drawn from the survey, there was no noticeable difference. The numbers were slightly different, as can be expected from using a different weighing method, but the overall results remained the same, with a Moran's I value very close to zero. However, this did not happen with those variables that were taken from the census.

Table 2: Euclidian Distance Spatial Autocorrelation Results

Moran's I	P Value	Z Score
0.316147	0.00000	19.709279

For example, the Moran's I value of average household income grew significantly smaller, indicating that using this method is unable to find as much clustering in average household income. Considering the very heavy clustering seen in the actual data for this variable (see figure 19 below), this demonstrates that the Euclidian distance method is likely inferior to the queen contiguity method for the purposes of this analysis.

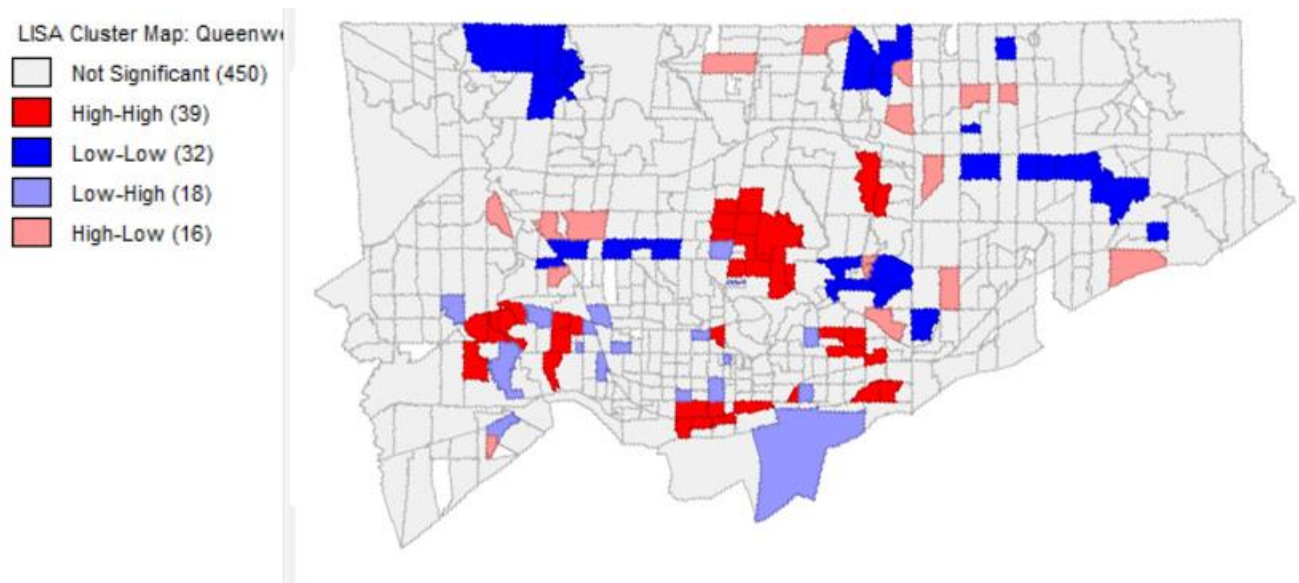


Figure 1: Q58 Cluster Map

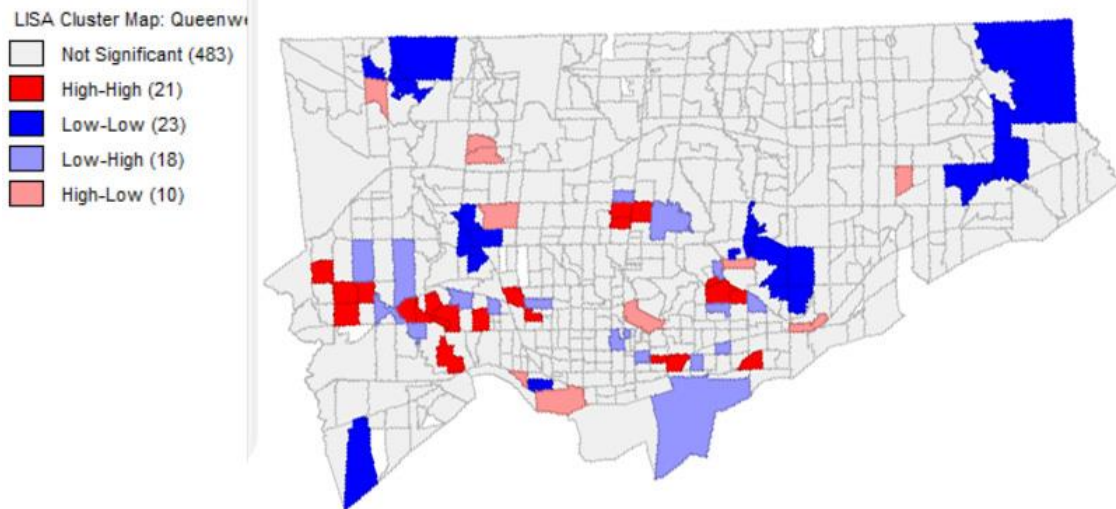


Figure 2: Q59 Cluster Map

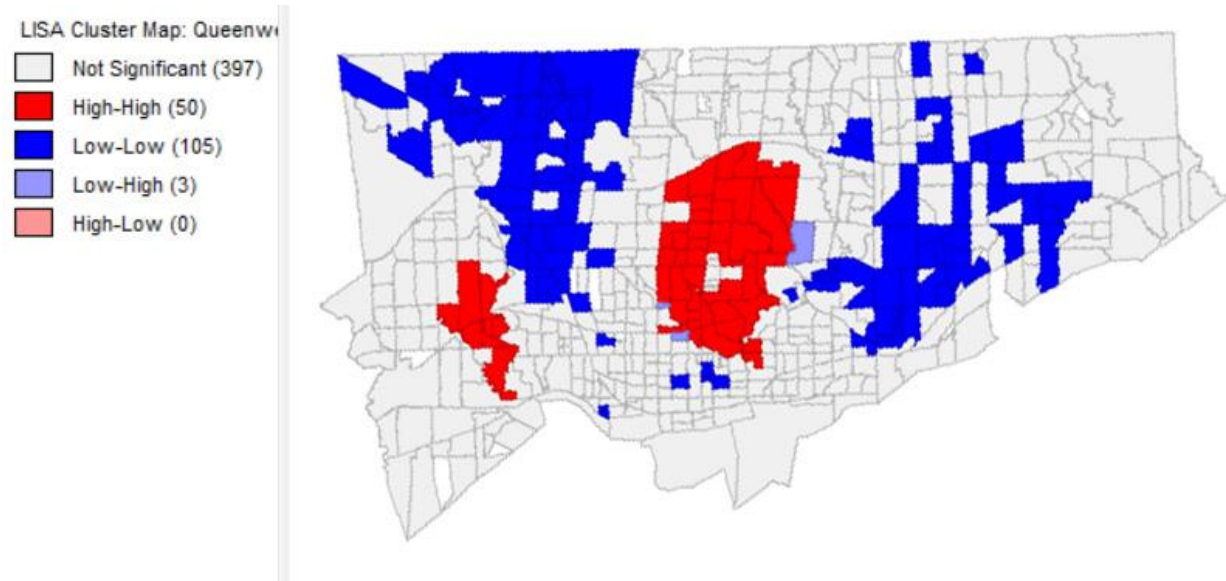


Figure 3: Average Household Income Cluster Map

Comparing cluster maps of average household income and security between the survey and census shows noticeable differences. Where there are a few similarities between the maps has to do with the center part of the city. In the census map (figure 18) there are clear and large hotspot clusters of high average household income in the center of Toronto and in the Islington region, while cold spots of low income are concentrated in the suburbs in the eastern and north-western portions of Toronto. Census tracts classified as low-high or high-low are either extremely scarce or non-existent, indicating a gradual shift from high to low average income between regions. However, the map of survey respondent incomes (figure 12) holds no such clusters. There is a

small cluster of high income census tracts in the center of Toronto corresponding to the primary cluster of high household income from the census, as well as a cluster of high income CT's in the Islington area like the census map, but other than that most hot or cold spots either correspond to the census map on an individual basis and not indicative of any kind of overall pattern, or actually go against it, with high-income respondent hotspots found in areas that are considered cold in the census map, and vice versa. The map of self-described income status follows even less of a pattern, although considering the subjective nature of the question this is perhaps to be expected, and is generally consistent with Q59's lower Z score. It should be noted that the Islington region seems to be the most reliable for high income hotspots, although individual census tracts within central Toronto have the same effect. This may be related to the fact that this region of Toronto also had the largest (and only) cluster of high life satisfaction, which indicates that the two variables may not be as unrelated as previously thought, despite their complete lack of relationship in the regression analysis. This pattern was observed in every other tested variable, that pattern being an almost completely random clustering complete with very high Z values, and a lack of any visible pattern in the clusters. A comparison of census tract average for the two income questions is also quite interesting.

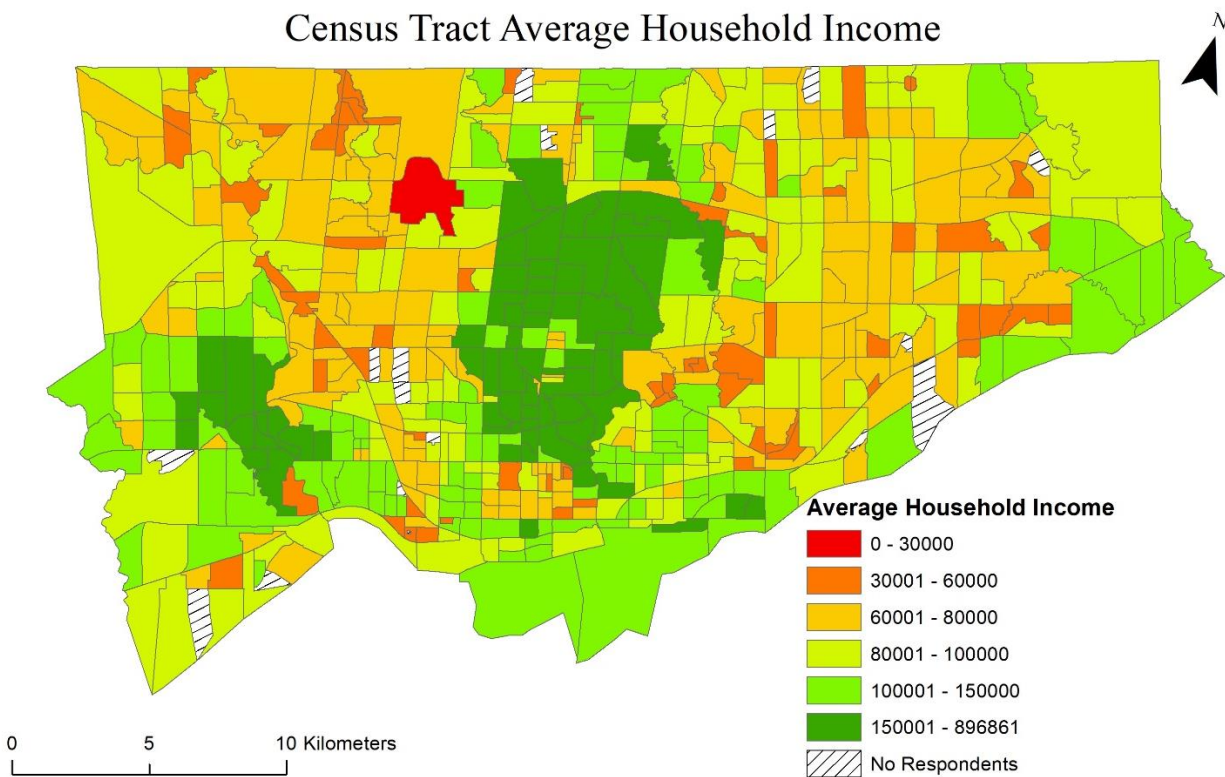


Figure 4: Census Tract Average Household Income

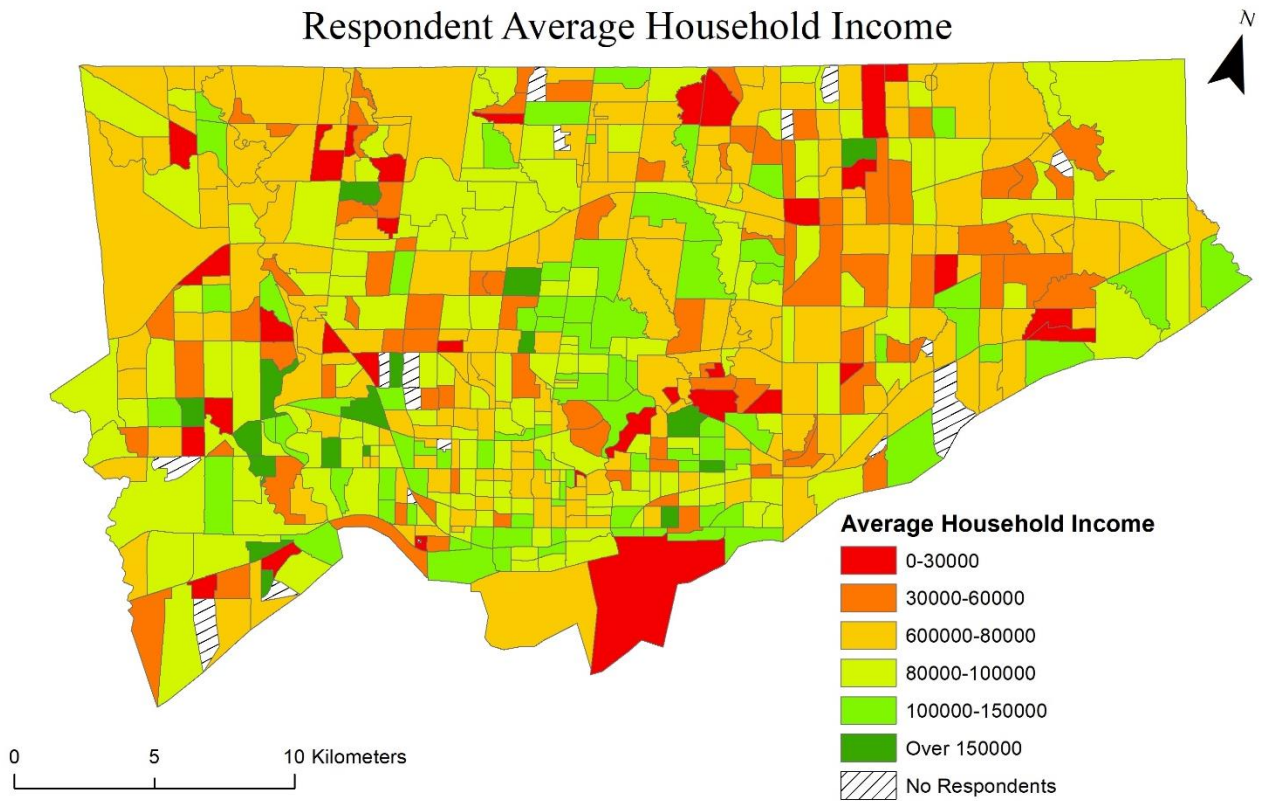


Figure 5: Q58 Average

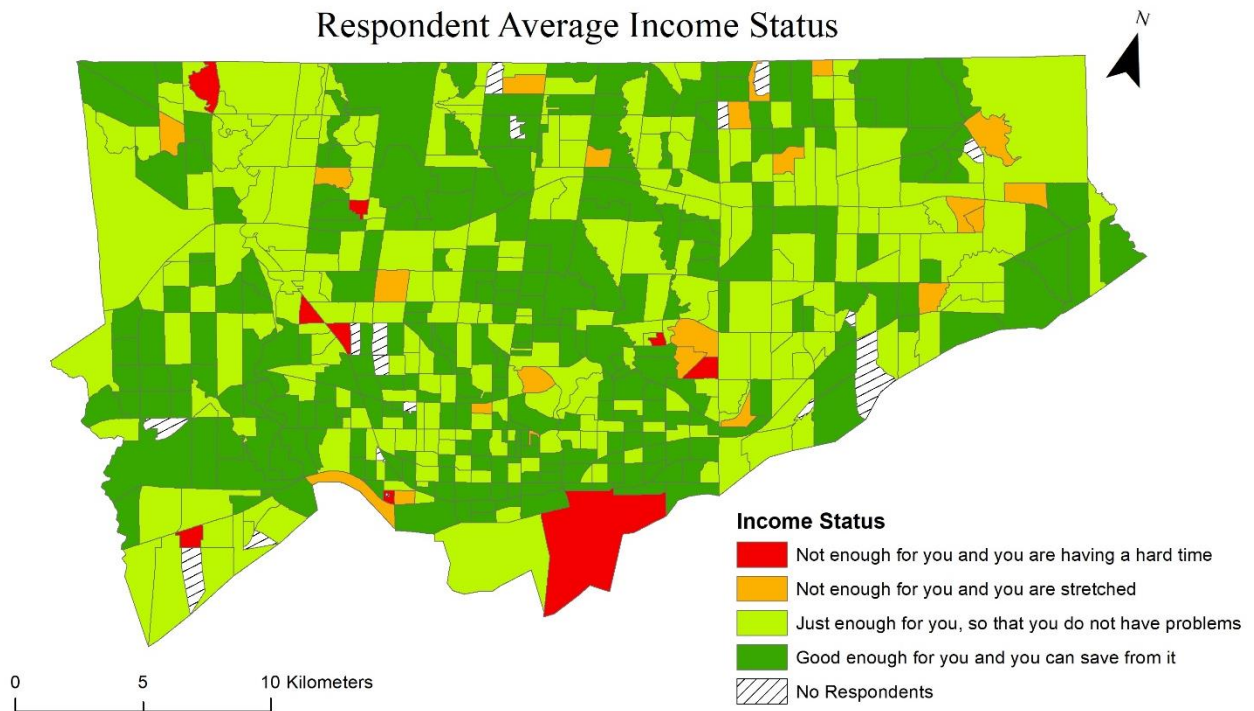


Figure 6: Q59 Average

As can be seen, the maps of the average census tract household income (figure 17) and the average survey household income (figure 16) have very little in common, outside of a slight tendency towards high income in the central Toronto area seen in the survey question map (consistent with both of their respective SAC cluster maps), with the main takeaway being that the survey has largely failed at accurately representing the income status of Toronto's census tracts. The difference between the survey income map and survey financial status map (figure 15) on the other hand paints a much more optimistic image of Toronto's economic situation. A part of this is likely due to the way the question is asked. The negative/low options may be difficult to distinguish for a respondent, leading to the possibility of them picking one at random. At the same time, the positive options may be too broad, as those respondents who describe themselves as having a good enough income they can save from may greatly vary from each other in their actual income. Nonetheless, it can be seen that the vast majority of the census tracts are, on average, not struggling with their income, with the exception of a few where the number of respondents was relatively low. This may be a result of bias in the results, as given the length

of the survey, it is much less likely that a person who is struggling with their financial situation will take the time to respond.

4.4. Other Relevant Variables

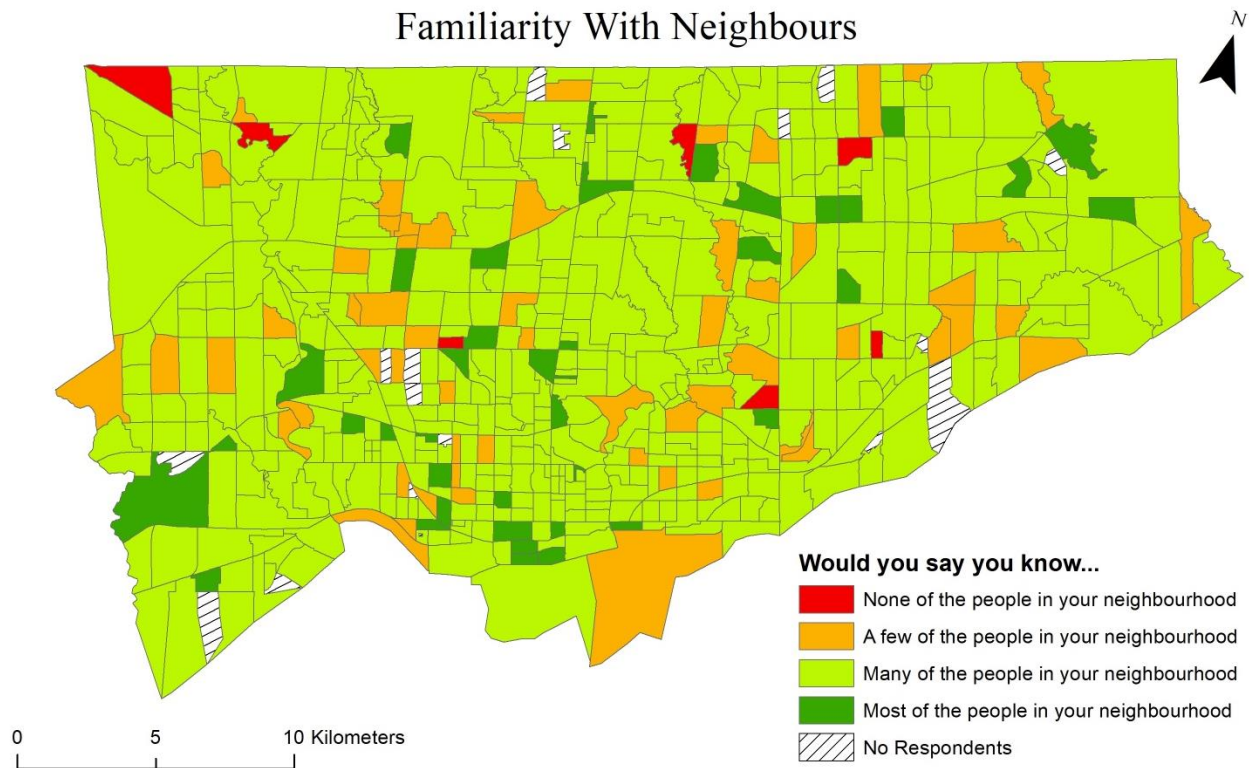


Figure 17: Q6 Average

As can be seen, most of the respondents know at least a few people in their neighborhoods, with a majority knowing many or even most. There is no visible spatial pattern or concentration of particularly high or low knowledge census tracts. In addition, there appears to be little correlation between the amount of people in a neighborhood known by the respondents and any of the other variables looked at so far. In this case, this is likely because the population density of Toronto and the relatively small size of the geographic units being looked at means that even people with healthy and thriving social networks may not know a lot of the people in their neighborhood. This is because it is much easier for a person living in Toronto to maintain relationships with people not living in their immediate vicinity than it would be for someone living in a small town, making this variable one of lesser importance.

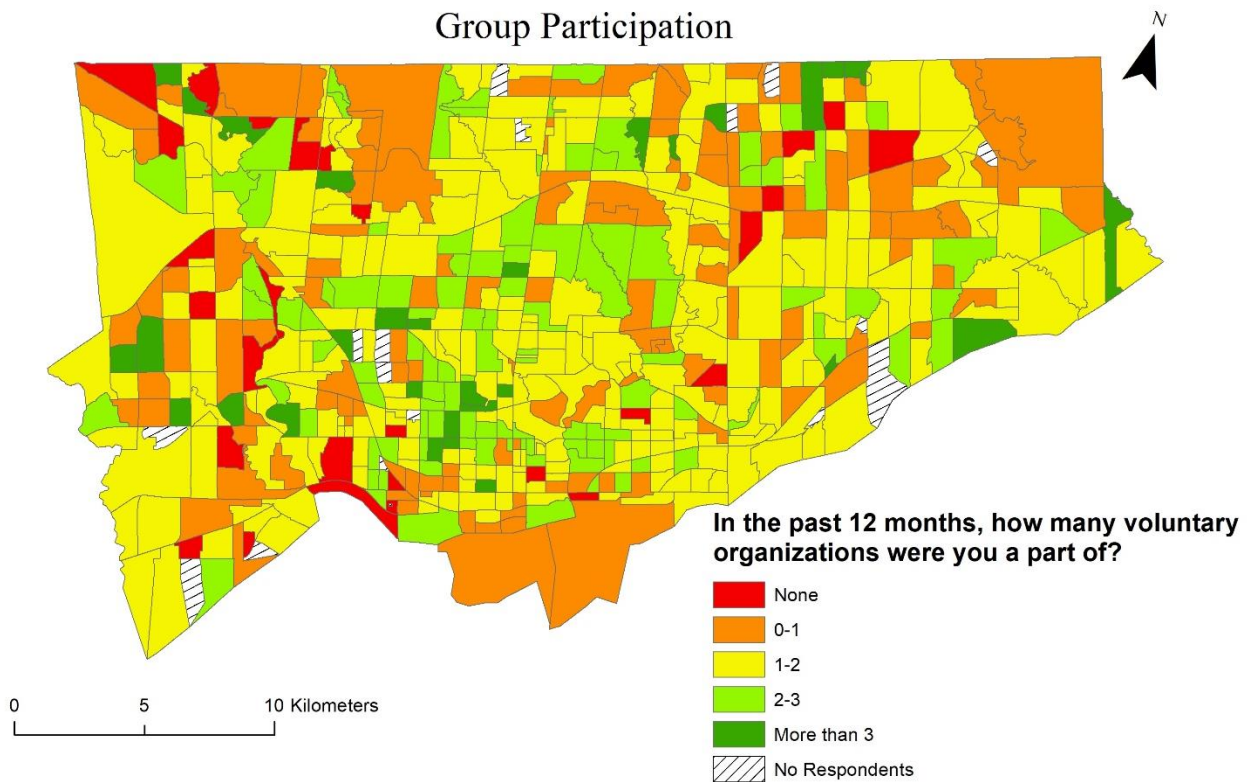


Figure 18: Q25 Average

This map is more interesting, as there is a small but nonetheless somewhat detectable spatial pattern. There appears to be a general cluster of higher group participation in the south-western corner of central Toronto, while low group participation seems to be more prevalent in the outer/suburban parts of Toronto. A part of this is likely due to population density, as the people in central Toronto will likely have more access to a potential organization of interest, both due to a higher number of those organizations and lower travel time. However, if compared to the census map of average household income (figure 15), it seems that low group participation is somewhat correlated with low household income, although the opposite does not appear to be true. The most likely reason for this is that people in lower income households simply cannot afford to participate in as many groups due to membership fees. There does not appear to be any correlation between life satisfaction, self-reported health and group participation however. Nonetheless, this does demonstrate that Toronto's socio-economic variables do in fact have a role in determining aspects of social capital.

4.5. Regression Analysis

The questions looked at in this section are as follows

Table 3: Variable Descriptions

Question	Description
Q1	How long have you lived in the Greater Toronto Area?
Q5	How long have you lived in your current neighborhood?
Q6	How many people within your neighbourhood do you know well?
Q14	How many close family members do you have that you regularly interact with?
Q16	How many close friends do you have that you regularly interact with?
Q35	How satisfied are you with your life right now?
Q36	How would you describe your health status?

The expected results were that there is a strong correlation between a respondent's life satisfaction and the other questions of the survey such as how many close friends a person has, or how connected they feel to their community. The regression analysis proved inconclusive, with most combinations of variables having extremely low R squared values (less than 0.1), while the relatively correlated ones still almost always had very low R squared values (less than 0.3). Every method taken to improve the correlation, from logging the variables to selectively going through every possible combination of variables to find the most correlated ones failed to overcome this. However there were some outliers that were somewhat less inconclusive than others, such as a higher than average correlation between self-described satisfaction with quality of life and the number of years that a person has lived in their current neighborhood and to a lesser extent in Toronto. For example, a regression analysis where the dependent variable measured satisfaction with life of the respondent, the results were as follows

Table 4: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	.433 ^a	.187	.186	11.39947	.187	127.581	1	553	.000	
2	.470 ^b	.221	.218	11.17269	.033	23.677	1	552	.000	
3	.483 ^c	.233	.229	11.09233	.013	9.027	1	551	.003	1.867

a. Predictors: (Constant), Q14

b. Predictors: (Constant), Q14, Q6

c. Predictors: (Constant), Q14, Q6, Q16

d. Dependent Variable: Q35

Table 5: Coefficients^a

Model		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		Beta			Tolerance	VIF
1	(Constant)		2.504	.013		
	Q14	.433	11.295	.000	1.000	1.000
2	(Constant)		-1.437	.151		
	Q14	.369	9.277	.000	.891	1.122
	Q6	.194	4.866	.000	.891	1.122
3	(Constant)		-2.101	.036		
	Q14	.281	5.724	.000	.576	1.736
	Q6	.169	4.203	.000	.856	1.168
	Q16	.149	3.005	.003	.564	1.772

a. Dependent Variable: Q35

Despite only being able to explain approximately 23.6% of the variation within the dependent variable, with slight positive autocorrelation, this is actually one of the best R squared values that could be gleaned from the various combinations of dependent and independent variables that were tried as part of the analysis. There is one notable exception to this trend, and that is the relationship between a person's self-reported satisfaction with their quality of life and their self-reported health quality. When put into a regression by itself, a person's self-reported health quality explains 32.4% of the variation in a person's life satisfaction, and when self-reported health quality is put into the regression analysis seen in table 1 above, the results are as follows.

Table 6: Model Summary^e

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.570 ^a	.324	.323	10.39504	.324	265.458	1	553	.000	
2	.613 ^b	.376	.373	10.00081	.051	45.457	1	552	.000	
3	.618 ^c	.382	.379	9.95709	.007	5.858	1	551	.016	
4	.624 ^d	.390	.385	9.90725	.007	6.558	1	550	.011	1.961

a. Predictors: (Constant), Q36

b. Predictors: (Constant), Q36, Q14

c. Predictors: (Constant), Q36, Q14, Q1

d. Predictors: (Constant), Q36, Q14, Q1, Q6

e. Dependent Variable: Q35

Table 7: Coefficients^a

Model		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		Beta			Tolerance	VIF
1	(Constant)		-4.562	.000		
	Q36	.570	16.293	.000	1.000	1.000
2	(Constant)		-6.952	.000		
	Q36	.472	12.903	.000	.844	1.184
	Q14	.247	6.742	.000	.844	1.184
3	(Constant)		-3.969	.000		
	Q36	.514	12.753	.000	.691	1.448
	Q14	.277	7.191	.000	.755	1.325
	Q1	-.101	-2.420	.016	.643	1.556
4	(Constant)		-4.197	.000		
	Q36	.499	12.317	.000	.676	1.478
	Q14	.272	7.098	.000	.753	1.328
	Q1	-.155	-3.327	.001	.512	1.955
	Q6	.108	2.561	.011	.630	1.588

a. Dependent Variable: Q35

The addition of that one variable increased the amount of variation explained by the variables from 23.6% to 41.2%. In addition, the Durbin-Watson value became a lot closer to 2 once Q36 (how would you describe your health right now) was added as a variable, indicating not only does it correlated better with life satisfaction then the other variables, but that it is highly independent of the other variables. It should be noted that this relationship does not work in the opposite direction, as life satisfaction does not correlate well with Q36 as the dependent variable,

only explaining 3.7 percent of the model. In addition, Q36 does not have this relationship with any other variables such as trust in neighbors or connection with neighborhood. It is logical that how healthy a person feels would heavily affect how satisfied they are with their life at any given moment, but otherwise the regression analysis results do not support the findings made by others studying social capital as discussed in the literature review section. It should be noted that most of the specific regression analysis discussed so far has had Q35 as the dependent variable. This is because Q35 has proven to be the dependent variable that consistently gets the highest R squared values, and as such is the most likely to expose relationships between variables in the study. Other variables that the research context identified as likely to contribute to high social capital had very poor R squared value results. The self-described income status of a respondent as well as group participation explained a very small amount of the variations in a person's life satisfaction (5.5 and 4.8% respectively), while the correlation between these variables and self-reported health were almost nonexistent (less than 1%), despite the fact that group participation has been directly tied to changes in health patterns by other studies. The difference here of course is that while those studies used medical records such as heart attacks and infant mortality rates, the survey asked for an immediate self-diagnosis on the part of the respondent, meaning there is a large possibility that how the respondent rates their health and their actual medical situation may be different, but at least some correlation was expected as opposed to the .000 R squared value that was obtained.

When census tract level census data was brought into the analysis, the results proved even worse with no regression analysis being able to break an R squared value of 0.1. This is likely due to a discrepancy of the datasets, as the census data looks at the entirety of the population while the social capital survey only looks at the information of those who answered the survey, and as such even if there is a strong connection between socio-economic variables and social capital it is unlikely that this connection would be identified in the survey. Nonetheless, this demonstrates that the survey data is likely not thorough enough to be effectively compared to census data, and subsequent research on the subject will need to run the survey in a manner that produces more respondents for more accurate results. More immediately, these results mean that a wellbeing index using these variables cannot be statistically justified, as it is almost guaranteed to be highly inaccurate.

5. Conclusion:

While the starting hypothesis assumed that findings on life satisfaction and other variables in Toronto would correlate with a number of socio-economic indicators, this was not the case. With the exception of self-reported health and life satisfaction, none of the variables had any notable correlation with one another. A part of the issue is likely the subjective nature of many of the questions. There is no way to verify or correct someone's answer if they make a mistake or deliberately lie, and two respondents may describe the same thing in two different ways. Even questions that deal with quantifiable variables like "how many close friends do you have" would vary significantly based on what each individual considers as the criteria for someone to be a "close friend" for example. Another part of the issue may also be the difference between the socio-economic variables of Toronto and the variables of the locations discussed in the various papers in the research context section. Had this survey been done in, for example, Vancouver, or across the rural parts of Southern Ontario, it is entirely possible that the results would have been different and a correlation would have been found.

While the regression analysis was not very useful, the maps that were created from the survey data were still quite useful, particularly the overall census average maps. This is not because they demonstrated a clear spatial pattern, but quite the opposite. They demonstrated that, at least within the limitations of the survey data, there are no significant or even noticeable spatial patterns. This indicates that low or high social capital and well-being are not confined to specific geographic regions, but rather something that varies from person to person all across Toronto. In addition, the lack of pattern similarity between the socio-economic data like income and well-being factors like life satisfaction demonstrates that using socio-economic data to determine where the population likely has the lowest social capital may not be the best method. If a program was created to tackle issues like low social trust and connectedness, and was using census data to determine where to focus their efforts, it would run the risk of letting people who are located in seemingly well-off areas fall through the cracks. This lack of connectedness between socio-economic and wellbeing variables may also be a unique feature of Toronto's environment, and a comparison of Toronto and a region where the established theories on social capital hold true would be a useful area of further study. One issue with these maps that nonetheless casts some doubt on their usefulness is the low response rate for the survey. As

mentioned in the methodology, there were just over 3000 usable respondents. However, simply getting more respondents may not be enough to get more accurate results, as a part of the issue was the disparity in respondents between different census tracts with some not even being a part of the analysis due to not having any respondents. A follow-up study would not just need to obtain more respondents, but also make sure that their respondents are evenly distributed across the geographic unit they are looking at the data by, whether it is census tracts or neighborhoods.

References

- DeFilippis, J. (2001). The myth of social capital in community development. *Housing policy debate*, 12(4), 781-806.
- Egolf, B., Lasker, J., Wolf, S., & Potvin, L. (1992). The Roseto effect: a 50-year comparison of mortality rates. *American journal of public health*, 82(8), 1089-1092.
- Fukuyama, M. F. (2000). Social capital and civil society. International Monetary Fund.
- Iyer, S., Kitson, M., & Toh, B. (2005). Social capital, economic growth and regional development. *Regional studies*, 39(8), 1015-1040.
- Kawachi, I., & Berkman, L. (2000). Social cohesion, social capital, and health. *Social epidemiology*, 174, 190.
- Kawachi, I., & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban health*, 78(3), 458-467.
- La Due Lake, R., & Huckfeldt, R. (1998). Social capital, social networks, and political participation. *Political Psychology*, 19(3), 567-584.
- Lin, N. (2017). Building a network theory of social capital. In *Social capital* (pp. 3-28). Routledge.
- Macinko, J., & Starfield, B. (2001). The utility of social capital in research on health determinants. *The Milbank Quarterly*, 79(3), 387-427.
- Newton, K. (1997). Social capital and democracy. *American behavioral scientist*, 40(5), 575-586.
- Ooka, E., & Wellman, B. (2006). Does social capital pay off more within or between ethnic groups? Analyzing job searchers in five Toronto ethnic groups. *Inside the mosaic*, 199-226.
- Paldam, M. (2000). Social capital: one or many? Definition and measurement. *Journal of economic surveys*, 14(5), 629-653.
- Pelling, M., & High, C. (2005). Understanding adaptation: what can social capital offer assessments of adaptive capacity?. *Global environmental change*, 15(4), 308-319.

Portes, A. (2000, March). The two meanings of social capital. In *Sociological forum* (Vol. 15, No. 1, pp. 1-12). Kluwer Academic Publishers-Plenum Publishers.

Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223-234). Palgrave Macmillan, New York.

Rupasingha, A., & Goetz, S. J. (2007). Social and political forces as determinants of poverty: A spatial analysis. *The Journal of Socio-Economics*, 36(4), 650-671.

Social determinants of health and health inequalities. (2019). Government of Canada. Retrieved from <https://www.canada.ca/en/public-health/services/health-promotion/population-health/what-determines-health.html>.

Szreter, S., & Woolcock, M. (2004). Health by association? Social capital, social theory, and the political economy of public health. *International journal of epidemiology*, 33(4), 650-667.

Toronto Foundation. (2018). Toronto Social Capital Study 2018.

Veenstra, G. (2000). Social capital, SES and health: an individual-level analysis. *Social science & medicine*, 50(5), 619-629.

Woolcock, M., & Narayan, D. (2000). Social capital: Implications for development theory, research, and policy. *The world bank research observer*, 15(2), 225-249.