

2016

# Analysis of the Geosocial Landscape in the City of Toronto

Courtney J. Jones

Wilfrid Laurier University, [jone6880@mylaurier.ca](mailto:jone6880@mylaurier.ca)

Follow this and additional works at: <http://scholars.wlu.ca/etd>

 Part of the [Geographic Information Sciences Commons](#), [Human Geography Commons](#), [Nature and Society Relations Commons](#), and the [Spatial Science Commons](#)

---

## Recommended Citation

Jones, Courtney J., "Analysis of the Geosocial Landscape in the City of Toronto" (2016). *Theses and Dissertations (Comprehensive)*. 1897.

<http://scholars.wlu.ca/etd/1897>

This Thesis is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact [scholarscommons@wlu.ca](mailto:scholarscommons@wlu.ca).

ANALYSIS OF THE GEOSOCIAL LANDSCAPE IN THE CITY OF TORONTO

by

Courtney Jones

BA, Wilfrid Laurier University, 2013

THESIS

Submitted to the Department of Geography and Environmental Studies

in partial fulfillment of the requirements for

MASTER OF SCIENCE

Wilfrid Laurier University

© Courtney Jones 2016

Wilfrid Laurier University

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

ANALYSIS OF THE GEOSOCIAL LANDSCAPE IN THE CITY OF TORONTO

by

Courtney Jones

BA, Wilfrid Laurier University, 2013

Supervisory Committee

---

Dr. Colin Robertson, Supervisor  
(Department of Geography and Environmental Studies, Wilfrid Laurier University)

---

Dr. Rob Feick, Member  
(School of Planning, University of Waterloo)

---

Dr. Ketan Shankardass, Member  
(Department of Health Sciences and Psychology, Wilfrid Laurier University)

## Abstract

Microblogging on geosocial platforms is a popular form of online communication where users post information about their daily lives and challenges. Since the launch of Twitter in 2006, information sharing through social media has become a largely unused data repository. Tweets often convey content about the users sentiment as it is happening. As such, Tweets can be viewed as a proxy of public mood. In this thesis, I performed a sentiment analysis of all public geo-located Tweets posted by a variety of Twitter users between September 2013 and October 2014. Each Tweet was processed through a custom algorithm to extract 8 different emotions: Anger, Confusion, Disgust, Fear, Happiness, Sadness, Shame, and Surprise. I then created an emotional landscape to display variance in emotion across the city of Toronto. The emotional landscape presented interesting emotional polarity change between the core and the periphery of the city. Neighbourhood profiles were then used to compare the emotional differences resource access could individual's ability to cope and mediate stress. I found that individuals living within close proximity to greenspace expressed increased levels of positivity though they have decreased access to built resources. I also found that individuals within Neighbourhood Improvement Areas experienced an increased risk of negativity. I believe large-scale analyses of public sentiment can provide valuable information for further analysis of resource use in an effort to reduce negative health effects long term.

---

Dr. Colin Robertson, Supervisor  
(Department of Geography and Environmental Studies, Wilfrid Laurier University)

---

Dr. Rob Feick, Member  
(School of Planning, University of Waterloo)

---

Dr. Ketan Shankardass, Member  
(Department of Health Sciences and Psychology, Wilfrid Laurier University)

## Acknowledgements

To begin, I would like to thank my supervisor, Dr. Colin Robertson - the writing process has been long but he was continually supportive and pushed me to the end. Dr. Robertson has taught me more through the completion of 2 theses than I could have ever imagined.

Second, I would like to thank Dr. Ketan Shankardass - without his primary research on stress and the relationship to the environment I would not have even considered an undergraduate thesis, let alone a Master's thesis as well. I would also like to thank both Dr. Rob Feick who was both reassuring and complimentary to the ideas and changes this research has endured along the way. I could not have completed this research without all of their guidance and support both mentally and academically. I have to thank the Spatial lab at Wilfrid Laurier University. Through the many years I spent there I met a lot of great academics and friends who were always there to draw out problems on the whiteboard and celebrate victories at Wilfs. Lastly, I would like to thank my friends and family who have always had the utmost confidence in me when I didn't think I could make it to through end.

## Contents

Acknowledgements.....	iv
List of Tables .....	viii
List of Figures.....	ix
Chapter One .....	1
1.0 Introduction .....	1
1.1 Background.....	1
1.2 Chronic stress and acute stress.....	5
1.3 Neighbourhood effects on health .....	6
1.4 Geosocial media data.....	7
1.5 Creating Emotional Landscapes .....	8
1.7 Objectives .....	9
References.....	10
Chapter Two .....	13
Abstract .....	13
2.1 Introduction .....	14
2.2 Methods.....	17
2.2.1 Study Area .....	17
2.2.2 Twitter .....	17
2.2.3 Neighbourhood Improvement Areas .....	19
2.2.4 Sentiment Analysis and Natural Language Processes .....	20
2.2.5 Kernel Density Estimates .....	22
2.2.6 Risk Rates.....	23
2.2.7 Boxplots.....	23

2.3 Results.....	24
2.3.1 Preliminary Statistics .....	24
2.3.2 Kernel Density Estimates .....	24
2.3.2 Risk Rates.....	25
2.3.3 Boxplots.....	26
2.4 Discussion .....	26
2.5 Conclusions and Future work .....	29
2.6 References.....	31
2.7 Figures.....	35
Chapter Three .....	44
Abstract .....	44
3.1 Introduction .....	45
3.2 Methods.....	47
3.2.1 Data .....	47
3.2.2 Proximity to Resources.....	50
3.2.3 Neighbourhood Profiles.....	51
3.2.4 Text Mining .....	54
3.2.5 Hierarchical Text Clusters .....	55
3.2.6 Sentiment Pattern.....	55
3.3 Results.....	56
3.3.1 Neighbourhood Profiles .....	56
3.3.2 Proximity.....	56
3.3.3 Text mining.....	57
3.3.4 Hierarchical Text Clusters .....	58

3.4 Discussion and Conclusion.....	59
3.5 References.....	63
3.6 Figures.....	66
Chapter Four.....	71
4.0 Conclusion .....	71
4.1 Discussions and Conclusions.....	71
4.2 Research Contributions .....	77



## List of Tables

Table 1 Personal resources .....	50
Table 2 Neighbourhood resource classification.....	54

## List of Figures

Figure 1.1	Stress as a Dynamic Process	5
Figure 2.1	Example of EMOTIVE classification	23
Figure 2.2	Spatial distribution of Tweets	37
Figure 2.3	Percentage of spatial distribution of Tweets	38
Figure 2.4	Monthly Variation in Tweets	39
Figure 2.5	Kernel Density Estimate (KDE) of positive Tweets	40
Figure 2.6	Kernel Density Estimate (KDE) of negative Tweets	41
Figure 2.7	Kernel Density Estimate (KDE) ratio of positive/negative	42
Figure 2.8	Risk rate	43
Figure 2.9	Risk rate	44
Figure 2.10	Risk rate of emotional polarity to NIA	45
Figure 3.1	Spatial location of neighbourhood profiles	69
Figure 3.2	Emotional frequencies by neighbourhood profile	70
Figure 3.3	Frequent terms by neighbourhood profile	71
Figure 3.4	Hierarchical text cluster of built neighbourhoods	72
Figure 3.5	Hierarchical text cluster of NIA	72
Figure 3.6	Hierarchical text cluster of social neighbourhoods	72
Figure 3.7	Hierarchical text cluster of green space neighbourhoods	73

## Chapter One

### 1.0 Introduction

### 1.1 Background

Urban environments often function as centres of culture, social life, and commerce. With the majority of the world's population living in urban centres, it is vital to understand how people interact with these spaces and how these spaces shape their well-being. Geographical analyses of health and neighbourhood over the last few decades have demonstrated many interacting processes which enact contextual influences on individual health (Diez Roux, 2001; Pearce, 2006). Stress in particular is a growing concern among health researchers focusing on both mental health (Langner and Michael 1963) and physical health (Rabkin and Struening 1976). Although there are theoretical linkages between the environment and health, the evidence is much less concrete. There are many barriers when studying stress, stemming from the very definition of it. Many researchers often disagree on a definitive answer as to what stress is. The overall consensus in research is that there are psychological effects from emotional stresses while at the same time those psychological effects can trigger a physical response. (DeLongis, 1988; Thoits, 1995; Turner et al., 1995; Matheson et al., 2006).

Stress is highly individualized, and according to Shankardass (2012), the way in which people cope with stress influences the impact of stress on their health. When people have proper access to healthier resources the negative effects of stress can be minimized, thus minimizing the future effects of stress such as stress-related illness. Stress related illness includes digestive health, cardiovascular distress, and possibly diabetes. (Blascovich

et al.,2007). Individual-level mediators of stress include early life events that affect cognitive and socio emotional development, prior experiences, personality characteristics, and personal capabilities, which include self-regulation and executive function (Shankardass, 2012).

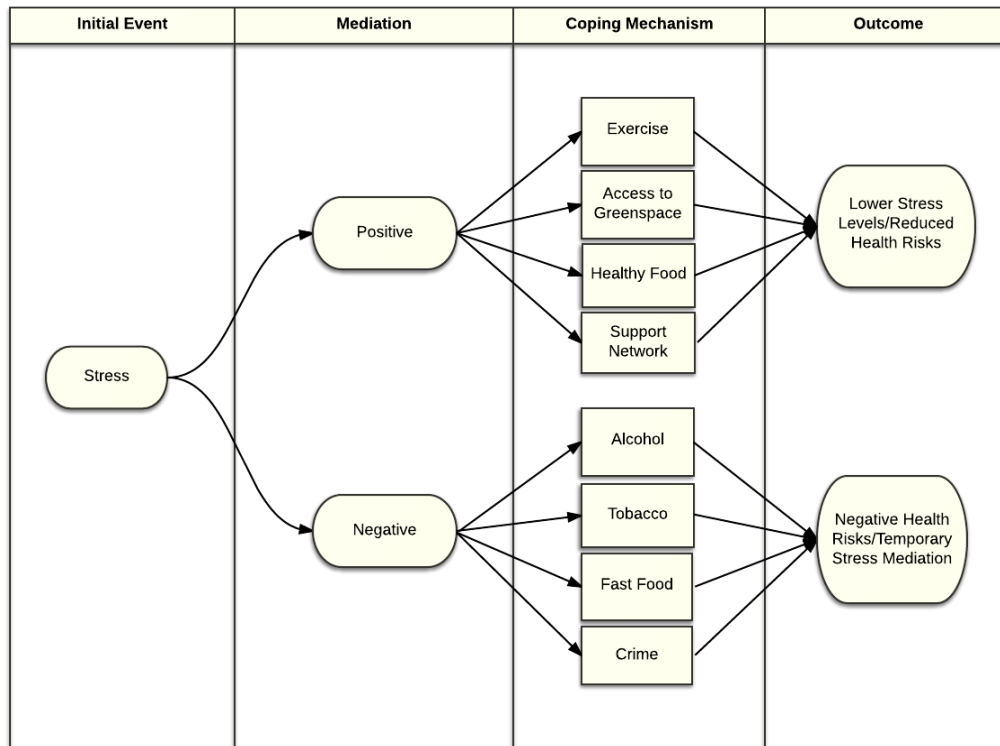
The urban environment is a collection of stressors and resources, as seen in figure 1.1, some resources within the urban environment are thought to help alleviate stress (Parkes, 1986). Therefore, the spatial accessibility of personal resources and the spatial structure of the urban environment may impact stress levels (Stigsdotter, 2011). Researchers suggest that there are 3 divisions of resources: built, natural, and social (Pearce et al.,2006; Shankardass,2012). Each of the 3 divisions contains both positive and negative resources. As stress can manifest in different ways, the way in which an individual interacts with a resource can vary (figure 1.1). Built resources are defined as a resource that is not naturally occurring within an environment such as grocery stores, fast food locations, or convenience stores (Gordon-Larson,2006; Pearce et al.,2006; Shankardass,2012). Natural resources are defined as resources that are mostly naturally occurring within the environment such as public parks or greenspace (MacKerron and Mourato, 2010;2013 ;Hartig and Hansmann,2010; Howell and Passmore, 2012). Social resources are defined as resources that affect the social interaction within the environment such as educational institutions, homeless shelters, or community health centres (Campbell et al., 1986; Holahan, 1991; Shankardass, 2012;). Most neighbourhoods contain multiple resources from within each resource category. As such, there may be utility in exploring how this varies across space.

A particular difficulty in the study of environment-stress relationships is access to large scale, granular information about the daily interactions of individuals. Within an urban setting, individuals are highly mobile which provides a challenge when identifying particular areas of high stress from traditional administrative survey data, as home environments may not be the main source of stress. In order to gain insight into how the environment interacts with and modifies individuals' experience of stress, it is important to know how and where one is encountering stressors, how stress manifests in their behaviour, and how and where one is coping with stress. Geosocial media streams combined with sentiment analysis may provide a more detailed individual measure of stress, which will allow extrapolation to the impact of features in the environment for different sub-populations.

Berke et al. (2007) found a positive association between obesity in the older population as a result of limited mobility and access to resources. Activity levels of adults aged 65-97 was largely based on accessibility of resources. As accessibility to resources decreased, physical activity also decreased, which lead to an increase in obesity (Berke et al., 2007). Walkability levels were assessed based on environmental variables that are known to affect walking along road networks, as well as distance to participant's homes. It was determined that the closer resources were considered more accessible and were therefore used more frequently regardless of health impact (Berke et al., 2007). Exposure to natural areas and greenspace can be highly restorative for some individuals during high stress times (Ulrich et al., 1991).

Through analysis of emotional content within social media expressions, an emotional landscape can be created to determine locations with high incidence of stress as well as areas with high incidence of negative sentiments. Increasing awareness of the emotional landscape of an area, and a better understanding of how features of the urban environment impacts individual stress, can aid in the development of an empirical definition of what a healthy urban environment looks like.

In this research, I hypothesize that patterns of emotional stress, as measured via the emotional mapping and analysis of the content, and location embedded within geosocial media streams (e.g., geo-tagged Tweets), may shed light onto the complex relationship between individual emotional state, stress, and level of interaction with their environment. In an effort to further the notion that coping and mediation within a positive environment can help reduce negative health outcomes.



**Figure 1.1 Stress as a dynamic process**

### 1.2 Chronic stress and acute stress

The 2 main types of stress are commonly referred to as chronic stress and acute stress. Both types of stress occur when an individual is exposed to a stressor (Selye, 1956; Holahan and Moos, 1991; Nasar, 1997; Shankardass, 2012). One notable difference between the manifestations of acute versus chronic stress is consistent exposure to a stressor. More specifically, daily habits can often include a stressor such as passing by a certain neighbourhood to get to work every day thus creating chronic stress. Whereas, passing by that neighbourhood one time may cause a moment of stress, defined as acute stress (Caspi et al., 1987; DeLongis et al., 1988; Affleck et al., 1994). The different classifications of stress could also indicate the different uses of resources. Individuals experiencing chronic stress may not be coping or mediating stress adequately, leading to

prolonged stress and increasing the risk of negative health outcomes. When an individual is experiencing acute stress, there is an opportunity to mediate that stress prior to the transition into chronic stress. The mediating resource could be positive such as exercise or social interaction or adversely the mediation choice could be creating negative health impacts such as alcohol or tobacco.

When experience of stress is measured, it is often through a single self-reported question in a survey (CCHS, n.d.). However, there can be important subtle differences in the emotional affect and sentiment that contributes to stress levels that are not captured through that one question. The difference in granular detail collected through traditional survey data and sentiment collected on social media prompted the need for further research into the cause of difference (Shih and Fan,2008). One notable difference between social media sentiment information and traditional survey data is that traditional data is completed based on home address at a larger scale of population, whereas social media is highly individualized. Highly individualized information can provide more information about the individual as well as the level of environmental influence when geo-located.

### 1.3 Neighbourhood effects on health

What defines a sense of place has been long explored within the discipline of geography. Davidson and Milligan (2004) suggest that emotions always take place within multiple levels of geography at multiple scales. That is to say, interaction with certain levels of geography can largely intensify the creation of a sense of place. An individual's experience within a neighbourhood is therefore influenced by multiple different elements of a neighbourhood. To further explore the neighbourhood effects on health there should be a focus on the health-influencing elements of a neighbourhood such as, resources.



There are variable amounts of resources within each neighbourhood which can lead to potential issue of access. However, resource access does not lead directly to the use of a resource as part of positive coping behaviour. For example, greenspace is typically considered a positive resource (MacKerron and Mourato, 2010;2013 ;Hartig and Hansmann,2010; Howell and Passmore, 2012); however, parks and greenspace can also be considered a negative space because of late-night recreational activities, such as crime, as well as the presence of a different demographic (Riger et al., 1982; Nassar and Jones,1997;Jorgensen et al., 2013). Perceived negative use of an overall more positive resource can then change the way an individual will use that resource.

There are also Socio-Economic factors within neighbourhoods that have been linked to more negative health outcomes. Neighbourhoods that are considered low income have also been associated with fewer resources, especially lower amounts of positive resources (City of Toronto,n.d.). In terms of health, neighbourhoods with fewer resources have been linked to diabetes, decreased mental health, as well as cardiovascular conditions (City of Toronto,n.d.).

#### 1.4 Geosocial media data

Sentiment analysis and opinion mining is the use of natural language processing as a means of extracting 'sentiment', or some aspect of emotional content from sources such as Twitter (Nasukawa, 2003; Bollen et al.,2011; Taheri et al., 2012). As Twitter gains popularity amongst users, the number of research studies exploring social media for a variety of purposes continues to grow (Kwak et al.2010; Hao et al., 2011). Twitter users are

often candid in their thoughts and opinions through Tweets, providing copious amounts of data on a variety of open ended topics. (Zhao & Rosson, 2009, Pak & Paroubek, 2010).

### 1.5 Creating Emotional Landscapes

Creation of emotional landscapes is an interesting sub-topic within geosocial sentiment analysis. Emotional landscapes are results that are produced from sentiment analysis after they have been related to the spatial landscape of the target location. Mapping sentiments may provide insight into how individuals interact in the space and even what contributes to the development of a sense of place.

“The Geography of Happiness” (Mitchell et al., 2013) is a sentiment study that used Twitter as a main source of data. Mapping happiness can be difficult when working with informal messages that are limited to 140 characters on Twitter however, the “Mappiness” app may change that (MacKerron and Mourato, 2010). Both studies were focused on the emotions of individuals across space. The creation of an emotional landscape in both studies displayed how users were interacting with their space and how that influenced their moods. An interesting finding within the Mitchell et al. (2013) study is that as city-size decreased, so did happiness levels within Tweets. Overall happiness scores were largely driven by the presence or absence of keywords in both applications of sentiment analysis. While the Mitchell et al. (2013) study is more of a target-based approach, researchers were able to find patterns related to the underlying environment after creating the emotional landscape.

Incorporating stress as the target in a target dependent application can further enhance creation of emotional landscapes. When looking for emotional patterns in geosocial media, a general approach to creating an emotional landscape could be to create a target independent approach. In this approach, all incoming Tweets are analyzed for the dominant emotions as defined before the classification, based on a pre-defined conceptual model of human emotions.

### 1.7 Objectives

The main objective of this thesis was to establish a connection between the built urban environment and individuals' emotional well-being. This approach encompassed a variety of techniques including, representation of the overall emotional landscape within the city as well as identifying key areas of emotion. Identification of potential emotional clusters was particularly useful for focusing the resource analysis portion of the study. The emotional clusters or patterns then gave focus on what type of resource within those particular areas may have been most influential between built, natural, and social as defined by Shankardass (2012).

## References

- Blascovich, J. J., & Katkin, E. S. (2007.). Psychological stress testing for coronary heart disease. *Cardiovascular Reactivity to Psychological Stress & Disease*, 27-45.  
doi:10.1037/10125-002
- CCHS (Canadian Community Health Survey) | Canadian Research Data Centre Network. (n.d.). Retrieved from <http://www.rdc-cdr.ca/datasets/cchs-canadian-community-health-survey>
- DeLongis, A. (n.d.). The Impact of Daily Stress on Health and Mood: Psychological and Social Resources as Mediators. *Journal of Personality and Social Psychology*, 54(3), 486–495.
- Diez Roux, A. V. (2001). Investigating Neighbourhood and Area Effects on Health. *American Journal of Public Health*, 91(11), 1783–1789. <http://doi.org/10.2105/AJPH.91.11.1783>
- Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case for emotional intensity. *Journal of Personality and Social Psychology*, 61(3), 427–434. <http://doi.org/10.1037/0022-3514.61.3.427>
- Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity. *Pediatrics*, 117(2), 417–424. <http://doi.org/10.1542/peds.2005-0058>
- Holahan, C. J., & Moos, R. H. (1991). Life stressors, personal and social resources, and depression: A 4-year structural model. *Journal of Abnormal Psychology*, 100(1), 31–38. <http://doi.org/10.1037/0021-843X.100.1.31>
- Howell, A. J., & Passmore, H.-A. (2013). The Nature of Happiness: Nature Affiliation and Mental Well-Being. In C. L. M. Keyes (Ed.), *Mental Well-Being* (pp. 231–257). Springer Netherlands. Retrieved from [http://link.springer.com/chapter/10.1007/978-94-007-5195-8\\_11](http://link.springer.com/chapter/10.1007/978-94-007-5195-8_11)
- Jorgensen, L. J., Ellis, G. D., & Ruddell, E. (2013). Fear Perceptions in Public Parks Interactions of Environmental Concealment, the Presence of People Recreating, and Gender. *Environment and Behavior*, 45(7), 803–820. <http://doi.org/10.1177/0013916512446334>
- Langner, T. S., & Michael, S. T. (1963). *Life stress and mental health: (by) thomas S. Langner and stanley T. Michael*. New York: the Free Press of Glencoe.

- MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, 23(5), 992–1000.  
<http://doi.org/10.1016/j.gloenvcha.2013.03.010>
- Matheson, F. I., Moineddin, R., Dunn, J. R., Creatore, M. I., Gozdyra, P., & Glazier, R. H. (2006). Urban neighbourhoods, chronic stress, gender and depression. *Social Science & Medicine*, 63(10), 2604–2616 <http://doi.org/10.1016/j.socscimed.2006.07.001>
- Nasar, J. L., & Jones, K. M. (1997). Landscapes of Fear and Stress. *Environment and Behavior*, 29(3), 291–323. <http://doi.org/10.1177/001391659702900301>
- NIA Profiles - Demographics - Your City | City of Toronto. (n.d.). Retrieved June 14, 2016, from <http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=e0bc186e20ee0410VgnVCM1000071d60f89RCRD>
- Parkes, K. R. (1986). Coping in stressful episodes: The role of individual differences, environmental factors, and situational characteristics. *Journal of Personality and Social Psychology*, 51(6), 1277–1292. <http://doi.org/10.1037/0022-3514.51.6.1277>
- Pearce, J., Witten, K., & Bartie, P. (2006). Neighbourhoods and health: a GIS approach to measuring community resource accessibility. *Journal of Epidemiology and Community Health*, 60(5), 389–395. <http://doi.org/10.1136/jech.2005.043281>
- Pile, S. (2010). Emotions and affect in recent human geography. *Transactions of the Institute of British Geographers*, 35(1), 5–20. <http://doi.org/10.1111/j.1475-5661.2009.00368.x>
- Rabkin, J. G., & Struening, E. L. (1976). Live events, stress, and illness. *Science*, 194(4269), 1013–1020. <http://doi.org/10.1126/science.790570>
- Riger, S., Gordon, M. T., & LeBailly, R. K. (1982). Coping with urban crime: Women's use of precautionary behaviors. *American Journal of Community Psychology*, 10(4), 369–386. <http://doi.org/10.1007/BF00893977>
- Selye, H. (1956). *The stress of life* (Vol. xvi). New York, NY, US: McGraw-Hill.
- Shankardass, K. (2012). Place-Based Stress and Chronic Disease: A Systems View of Environmental Determinants. In P. O'Campo & J. R. Dunn (Eds.), *Rethinking Social Epidemiology* (pp. 113–136). Springer Netherlands.
- Shih, T., & Fan, X. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, 20(3), 249–271. doi:10.1177/1525822x08317085

- Stigsdotter, U. K., & Grahn, P. (2011). Stressed individuals' preferences for activities and environmental characteristics in green spaces. *Urban Forestry & Urban Greening*, *10*(4), 295–304. <http://doi.org/10.1016/j.ufug.2011.07.001>
- Stress: The different kinds of stress. (n.d.). Retrieved April 01, 2016, from <http://www.apa.org/helpcenter/stress-kinds.aspx>
- Thoits, P. A. (1995). Stress, Coping, and Social Support Processes: Where Are We? What Next? *Journal of Health and Social Behavior*, *35*, 53–79. <http://doi.org/10.2307/2626957>
- Turner, R. J., Wheaton, B., & Lloyd, D. A. (1995). The Epidemiology of Social Stress. *American Sociological Review*, *60*(1), 104–125. <http://doi.org/10.2307/2096348>
- Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, *11*(3), 201–230. [http://doi.org/10.1016/S0272-4944\(05\)80184-7](http://doi.org/10.1016/S0272-4944(05)80184-7)

## Chapter Two

### Abstract

Stress is one of the most common illnesses society faces today (Almeida, 2005;Sharnkardass,2012). With such a large portion of the population affected by such an illness there is a lack of understanding what causes stress and the unique combination of emotions that cause stress. There are currently very few standardized surveys that collect information about how and where stress manifests. Geosocial sentiment analysis has recently become a popular form of data collection within the research community. Geosocial platforms such as Twitter are often open outlets for public ranting or discussion. Therefore, Tweets provide an excellent model for public behaviour as well as sentiment. Within the city of Toronto there are 2.8 million residents and an unknown amount of daily visitors. The Tweets collected have geographic coordinates, which support the use of Tweets in the creation of an emotional landscape created using public sentiment. With each Tweet broken down by emotion there are emotional patterns that form a broader expanse of emotional combinations linked to stress development and mediation. A variety of methods were used including Kernel Density Estimates (KDE) as well as a ratio of KDE to create a realistic view of emotions, I found that individuals who were beyond the core centre of the city were more likely to experience negativity. It was also determined that individuals within Neighbourhood Improvement Areas (NIA) expressed increased negativity. This research will provide a foundation for emotional compositions of stress as well as potential links to the neighbourhood resources.

## 2.1 Introduction

An urban centre such as Toronto provides an interesting array of individuals in terms of age, sex, and income that are likely all on social media in some form. Vast amounts of information are created daily through social platforms such as Twitter. Social information collected through Twitter is further amplified by the spatial information associated with some Tweets. The type of information collected through Twitter and other geosocial platforms may not be considered in a traditional style of information collection. When Twitter is used in combination with another data source such as census or neighbourhood profiles, the data can deliver a more detailed perspective on what is happening within an area.

The diversity of Twitter users within one dataset provides a much more intuitive sense of what is happening in the city. Twitter users “present” events through tweets as they are happening, they also likely to be tweeting where events are happening. Individuals on Twitter are not limited to questions asked post-event, they are open to posting exactly as things are happening while they are occurring. This can provide immediate insight into what happened from their perspective as well as how they are feeling in that moment.

Traditional survey methods have the advantage of household information such as income, age, and employment status. In some traditional surveys personal questions are not available or they are not open-ended. There can also a time delay on surveys as they are often conducted annually. Twitter can provide information that is temporally relevant about the everyday interactions of an individual. With the technology evolution, social media platforms such as Twitter have accelerated online communication. Within everyday



life an individual can communicate with multiple people simultaneously about a variety of topics and feelings. This study will focus on how spatial and social variation of geosocial sentiment can enhance our understanding of emotional affect across an urban landscape.

There are a variety of resources within a neighbourhood that can influence the health of an individual (Stallis and glanz, 2006; Mitchell et al., 2013). However, it is how each individual interacts with the neighbourhood and the resources within a neighbourhood that influence the mediating effects. For example, a resource such as an open greenspace may be considered positive mediation space to one individual, whereas that same greenspace could have a negative emotional impact on others (Maas et al.,2006; Mitchell and Popham, 2008). There are different classifications of resources: built, natural, and social. Built resources are any useful structure not naturally occurring in nature, for example a grocery store (Sloane et al.,2006). Natural resources include various living and non-living features that are occurring within a space without human intervention (Leader-Williams et al.,1990). Social resources include locations that foster relationships with other people such as, workplace, school, recreation locations (Macintyre et al.,2002).

The study of space and health has become closely intertwined with the study of place and health. The main difference between space and place is how an individual behaves within a space, shapes their sense of place (Tuan,1975;Nasar and Jones,1997). It is within that perception of place that certain behavioural patterns begin to develop. How an individual perceives a place is largely influenced by their emotional connection and level of interaction with that space. When an individual has a positive experience within a space, it

will likely form a positive sense of place. Adversely, when an individual has a negative experience in a space, it will likely form a negative sense of place. When an individual is exposed to negative places consistently, it can contribute to increased risk of stress (DeLongis, 1988; Steptoe and Feldman, 2001).

In order to develop an understanding of individual-level and community perceptions of stressors and resources, large-scale and granular data on mood and emotion are needed. It is important to first develop a general overview of the spatial distribution of emotions and resources. Establishing a baseline of emotional distribution and frequency will aid in the further research into the connection of the role of space and place in health geography. Emotional landscapes can provide a broad perspective on the emotional distribution across an urban landscape. Using the sentiment derived from tweets will provide the basis for the emotional landscape of the City of Toronto. The creation of an emotional landscape will also contribute to a further understanding of how resources or lack of resources contribute to the emotional outcome of an individual. Having the emotional landscape as a background, the resources can be overlaid to contribute to the understanding of the level of interaction.

There is a particular importance to the spatial relationship to disease because it can represent disparities in resources. Areas that are underserved in terms of resources may have increased demand for said resources due to population or demographic changes. Neighbourhood Improvement Areas (NIA) are one example within an urban environment where there is a need for change that is slow to arrive. NIA are neighbourhoods that match a set of criteria outlined by the city of Toronto. The list of criteria includes, socio-economic

factors, physical and mental health, as well as age of the neighbourhood infrastructure (City of Toronto,n.d.). Recent surveys have found that individuals living with NIA experience increased risk of various chronic diseases including mental health issues (City of Toronto,n.d.).

## 2.2 Methods

### 2.2.1 Study Area

The Study area for this research is the City of Toronto. Toronto is a city of approximately 2.8 million residents within 525 Census tracts and 140 neighbourhoods, organized into a roughly core-periphery development pattern around the central business district. Listed as one of the top five most liveable cities (Economist Intelligence Unit, 2012), Toronto is a hub of culture, education, and a bustling financial centre (fDi Intelligence, 2013). The city is Canada's largest commercial capital boasting the Toronto Stock Exchange as well the top five banks of Canada. The city's economy includes communications, education, transportation, media, and arts. Famed for well-defined neighbourhoods such as "Chinatown", "Little Italy", and the Greek "Danforth". There are many local attractions which draw the local population as well as first time visitors. Toronto provides an ideal context within which to study stress-environment relationships, as it encompasses a mix of age, ethnic, and socioeconomic stratifications.

### 2.2.2 Twitter

In-situ data collection in the form of geo-located Tweets has become more well-known and more widely adopted within the research community. Some of the first researchers to use geo-located Tweets began with studies of natural disaster and aid tracking (Tornens,

2010;Sakaki et al., 2010). Sakaki et al., created an algorithm that could better detect large scale events and create faster warning for evacuation and resource planning. Researchers were able to predict earthquakes with 96% accuracy through kalman filtering and reporting of earthquakes on geosocial media (Sakaki et al., 2010). Alternate research focused on post-natural disaster events and how people were coping with the major disasters (MacEachren et al., 2011). Near real-time data collection has proven to be useful in disaster avoidance and prevention (Rogstadius et al.,2011).

There has been a shift in the way individuals communicate with one another, progressively more interaction and communication is done online through platforms such as Twitter, facebook, flickr (Huberman et al., 2008). With the increase in daily online interaction there are increasingly more methods of data collection from geosocial data. Twitter provides different options for both location and frequency of collection. There are also options to harvest Tweets from a limited selection of individuals for targeted analysis. Targeted analysis can be particularly helpful when establishing individual behavioral patterns (Huberman et al., 2008; Zhao and Rosson, 2009 ). Targeted analysis can also be helpful when looking for trending topics or keyword searches. Alternatively, there is a post-event collection method where twitter streams are provided after the selected event has occurred (Twitter,n.d.).

Geo-Located Tweets for this research were collected within the python Twitter package "Tweepy". Tweepy is a programming library for python that supports access to

live streaming Tweets for analysis. Using the open Twitter Application Programming Interface (API), Tweets within the City of Toronto were collected. A bounding box of Toronto was defined as a geographic search criteria to ensure collected Tweets was limited to the study area. Tweets were collected from September 2013 through October 2014. In total there were approximately 9.2 million Tweets collected.

### 2.2.3 Neighbourhood Improvement Areas

Neighbourhood Improvement Area (NIA) boundaries were collected from the city of Toronto. As defined by the city of Toronto, NIA are neighbourhoods that have low neighbourhood equity scores that may require additional resources and services (City of Toronto, n.d.). Neighbourhood level indicators are divided into 6 main categories including economic opportunities, social development, participation and decision-making, physical surroundings, and health. Currently, 31 out of 140 neighbourhoods (22%) are included on the list for improvement, which has escalated from the 2005-2013 list of 13 identified neighbourhoods. Improvement plans are currently supported by the Toronto Strong Neighbourhood Strategy, which aims to invest in a variety of programs, services and facilities to strengthen conditions within the identified NIA. It is hypothesized that individuals within a NIA will experience increased risk of negative emotions, compared to individuals outside of the NIA. The hypothesized increased risk of negative emotions is estimated to be caused by a linkage between the amount of resources available within the NIA as well as the poor health outcomes. Resources can provide an outlet for individuals who are experiencing a negative emotion to aid in the mediation of stress, however, if there

are little to no resources available than there will be limited opportunities for coping (figure 1.1)(Holahan, 1991;).

#### 2.2.4 Sentiment Analysis and Natural Language Processes

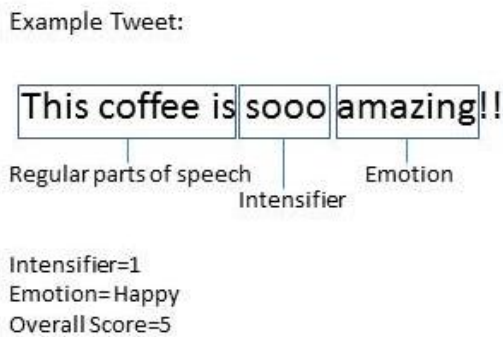
Sentiment analysis is defined as the computational process of identifying emotional terms within a text segment through Natural Language Processing (NLP) (Nasukawa and Yi,2003;Liu,2012). Sentiment analysis encompasses many methods that seek to determine patterns of emotion towards particular topics or to provide general insight within a geographic location.

The extraction method of sentiment from text is equally as unique as the applications (Pang and Lee,2008; Taboada et al.,2011; Davidov et al.,2011). The advancement of machine learning algorithms has also raised intrigue in extracting more fine-grained emotions rather than the polarity of positive or negative. (Agarwall et al., 2011;Pak and paroubeck,2011; Terrana et al., 2014).

Within this study, Tweet sentiments were extracted using an algorithm program called "EMOTIVE". EMOTIVE is an automated sentiment analysis program that is used for social media and informal terse messages (Sykora, et al., 2013). Based on Plutchik's wheel of emotions, eight high-level emotions were included in the EMOTIVE classification (Plutchik, 1980). There are six perceived negative emotions: anger, confusion, disgust, fear, sadness, shame, as well as two perceived positive emotions: happiness and surprise. Within each Tweet there were different classifications of words such as intensifiers. Intensifiers are associated with one emotional word. There can be multiple intensifiers that would

contribute to the overall score of that emotion. Tweets can pose more than one emotion, which is when the intensifier is largely useful. It should also be noted that not every Tweet will have emotional content. It is estimated that less and 10% of Tweets contain one of the eight high-level emotions within the EMOTIVE program.

Collected Tweets were first analyzed through a custom Natural Language Processor (NLP) to tokenize each Tweet. After each Tweet was tokenized into smaller part-of-speech elements, each token was matched with a selected ontology to determine the emotional content of the Tweet. The select ontology is unique in this NLP because there is no set dictionary of terms. The dynamic ontology allows for a greater range of classifications and better precision in analysis. When the tokens of the Tweets pass through the dynamic ontology, certain tokens are recognized as regular parts of speech while others are identified as emotional terms which are then classified and scored (Figure 2.1). The emotional score of the Tweet is dependent on the number of intensifiers, number of different emotions within the same Tweet, and negators. Using a combination of the positive emotions within the EMOTIVE classification, Tweets were then separated into positive and negative. Emotions such as “happy” and “surprise” were classified as overall positive. The negative emotions included: “anger”, “confusion”, “disgust”, “fear”, “sadness”, and “shame”.



**Figure 2.1 Example of EMOTIVE classified Tweet (Sykora et al.,2013)**

### 2.2.5 Kernel Density Estimates

Kernel Density Estimates (KDE) were created to visualize the density of point features at all locations in the study area. KDE are an ideal way to conduct exploratory analysis, as they can be used to visualize the distribution of continuous variables as a smoothed surface. Each KDE is unique in the sense that it is a non-parametric estimate of the Probability Density Function (PDF) of continuous data (Silverman, 1986). When creating a KDE there are no assumptions or bias in the distribution for input variable. The Twitter data that was collected included the geolocated coordinates, which were then used to create a spatial point data frame that facilitated analysis. A raster grid of the Tweet spatial point data frame was created with a standard cell size of 100. The Tweet raster grid was then used to create the KDE. Theoretically, this created a smoothly curved surface, which is fitted over each individual point of the input layer. Cell size contributed to the creation of the smoothing factor. When the cell size is larger, there is an “over smoothing” effect wherein the KDE is no longer accurately representative of the point features true location. Alternatively, there is “under smoothing” effect, which is too similar to the original



point feature location (Silverman, 1986; Jones, 2012). Secondary to the emotional polarity KDE, a ratio comparing the emotionally positive Tweets KDE to the emotionally negative Tweets KDE was created. The creation of a ration aided in the visualization in the difference in spatial distribution between positive and negative Tweets.

### 2.2.6 Risk Rates

Potential occurrences of each of the 8 emotion were calculated using a standardized mortality ratio. The traditional structure of standardized mortality (SMR) was used wherein observed cases of disease is divided by expected number of cases to create a ratio for potential Risk Rates (RR). For the purpose of this study, Tweets that were classified as one of the eight emotions within a neighbourhood were considered an observed case. The total number of Tweets in each emotion category was divided by the total amount of Tweets within the complete Tweet population database to create an expected percentage value. The expected percentage value was then applied to the total Tweet amounts in each neighbourhood to create the expected amount of emotional Tweets within that neighbourhood.

$$SMR = \frac{\text{Observed number of Tweets with } (x) \text{ emotion}}{\text{Expected number of Tweets with } (x) \text{ emotion}}$$

When the SMR variable is closer to 1 it is perceived to be closer to a “normal” or expected risk rate. When the SMR variable is higher than 1 that represents a higher than expected risk rate. SMR values lower than 1 indicate a lower than expected risk rate.

### 2.2.7 Boxplots

Boxplots provide statistical comparison between groups of data such as neighbourhoods. Boxplots were created to explore relationships between neighbourhoods

(NIA vs non-NIA) and the RR of emotional polarity.

### 2.2.8 Statistical Testing

Significance testing in the form of a t-test was performed to examine statistical significance of differences in mean emotional polarity among NIA and non-NIA neighbourhoods. A t-test is a form of hypothesis testing that compares the means between two datasets to determine the probability that differences are due to chance. The null hypothesis in this research was that there is no difference in mean negative tweet polarity between NIA and non-NIA neighbourhoods.

## 2.3 Results

### 2.3.1 Preliminary Statistics

After sentiment analysis it was determined that there were 300,806 Tweets that contained emotional content. The number of Tweets per neighbourhood varied largely, with the majority of the Tweets focused in the downtown core area of the City (Figure 2.2 and Figure 2.3).

### 2.3.2 Kernel Density Estimates

The KDE created for positive Tweets revealed a concentration of Tweets in the downtown core of the city with slight distribution upward through the centre of the city (Figure 2.5). The KDE created for negative Tweets showed a similar concentration in the downtown core of the city, though smaller than the positive KDE. The negative KDE also shows more dispersion throughout the city (Figure 2.6).

When comparing the positive KDE and the negative KDE, the patterns of Tweets are very similar making a direct comparison difficult. Creating a KDE ratio of positive Tweets to negative Tweets emphasized the differences between concentrations of Tweets (Figure

2.7). In the KDE ratio, there was a concentration of more positive Tweets in the downtown core that progresses up through the centre of the city. Though the northern portions of the city appear to be dominated primarily by more negative Tweets, there were also small isolated pockets of more positive Tweets throughout the city.

### 2.3.2 Risk Rates

Neighbourhoods that had a risk rate  $>1.2$  were considered to have increased rates of that emotion. Based on the observed number of Tweets and expected number of Tweets, neighbourhoods in the centreline of the city had increased occurrences of happiness (Figure 2.8). There was a similar pattern in the core neighbourhoods with an increased risk rate of surprise (Figure 2.9). Following an inverse pattern of happiness, there was an increase in the rate of sadness in the periphery neighbourhoods of the city (Figure 2.9). Risk rate for fear was increased mostly in the periphery neighbourhoods with smaller clusters of neighbourhoods in the core area (Figure 2.8). Increased risk for disgust occurred primarily along the edges of the city with fewer instances in the centre (Figure 2.7). There was a fairly distinct pattern for risk rate of anger. The neighbourhoods in the northern periphery experienced increased risk, while only one neighbourhood in the core had elevated risk rates (Figure 2.8 and Figure 2.9). Most of the emotions presented with a clustered pattern of elevated risk rates, though not all emotions had such distinct patterns. Shame and confusion were randomly distributed throughout the city, though shame also has neighbourhoods that presented the highest risk rates of any of the emotions (Figure 2.8 and Figure 2.9).

### 2.3.3 Boxplots

The output results of the risk rate maps showed the neighbourhoods with increased risk rates for negative Tweets also shared a potential spatial relationship with the NIA. The boxplot comparing risk rate of positive Tweets and the NIA shows that there were increased numbers of positive Tweets outside of the NIA (Figure 2.10). The boxplot comparing risk rate of negative Tweets and the NIA shows evidence there were more negative Tweets occurring within NIA (Figure 2.10).

Through a t-test for difference between means of the relationship between risk rate of positive Tweets and NIA, the t-statistic was 2.71, and the p-value was 0.009219, showing the difference was not random. The result of the t-test for significance between risk rate of negative Tweets and NIA was -3.11 with a p-value of 0.00292, also showing that there is a statistically significant relationship between emotional polarity on Twitter and NIA status in the City of Toronto, thus rejecting the null hypothesis.

## 2.4 Discussion

Knowing where particular emotions are occurring is important when establishing a connection with an individual's surrounding environment. When focusing on risk rates of emotions, the patterns that were formed were similar to the patterns that were presented through the KDE and KDE ratio, further contributing to definitive relationships between neighbourhoods and emotions. When individuals are consistently displaying one emotion in a neighbourhood it contributes to the emotional profile of that neighbourhood (Frohlich, 2013).

The KDE ratio in figure 2.7 validated an interesting scenario where the percentage of Tweets within the core location and they were more positive (figure 2.3). When also looking at the difference in socio-economic status (SES) between the core of the city and the periphery neighbourhoods there is also a noticeable shift in income (Census,2011). There is also a fairly significant difference between the number of residents within the core and the periphery as the core of the city tends to be more densely populated. There is also a shift in demographics between the core of the city and the more northern neighbourhoods. The core residents of the city are most often part of the younger demographic, while the north neighbourhoods are mostly older couples or families. All of these factors combined could help breakdown why there is a shift in amount of Tweets (figure 2.2) and the difference in emotions.

Neighbourhood Improvement Areas (NIA) present a unique set of conditions in which the neighbourhood as a whole does not meet living standards established by the city of Toronto. Various social and economic factors are considered when adding a neighbourhood to the list of NIA (City of Toronto, n.d.). When comparing the locations of NIA to the risk rates for emotions, it is interesting to note the difference between positive and negative areas in relation to the NIA. Although the content of the Tweet may not be directly related to the NIA it should be acknowledged that individuals within the NIA are overall more negative while individuals outside of the NIA tend to be more positive on social media.

The relationship between Tweet emotions and NIA has provided insight into how decreased resources may lead to negative emotions. There is a long understood connection

between low SES and increased stress levels (Baum et al.,1999; Steptoe and Feldman, 2001; DeHollander and Staatsen, 2003). However, the relationship between mediating resources and high stress areas is seldom discussed or empirically examined. With the addition of NIA, there was some evidence that a lack of resources such as healthy food, decreased walkability, and small greenspace may have contributed to negative emotions and acute stress.

A person's lifestyle has been attributed to nearly 25-30% of avoidable disease burden, which is a significant decrease from 70-80% in the early 20<sup>th</sup> Century (Murray and Lopez, 1997). With such a significant decrease in avoidable diseases, there should be a continued effort to reduce exposure to illness. There are also disparities between neighbourhoods and avoidable diseases, neighbourhoods with decreased access to resources experience delayed neighbourhood developments. When comparing NIA with percent population change and the negativity in the Tweets, there could be further emphasis on change and improvement. Some of the neighbourhoods have been part of the NIA for over 10 years with little to no improvement, and have experienced a decrease in population since being added to the list (City of Toronto, 2013).

Rapid population growth and a changing demographic within the city have contributed to the addition of resources in "newer" neighbourhoods. There is a primary focus on development and attraction of the younger working crowd, which places less focus on the improvement of pre-existing areas. When comparing population change (2006-2011) in Toronto neighbourhoods, the large majority of NIA are experiencing negative population changes while surrounding areas are experiencing increases between

7-66.7% within 5 years (Canadian Census, 2011). Adding and updating resources are changes that are simple enough that could intervene with neighbourhoods being added to the NIA list. As well as Integration of social programs to encourage physical activity and community involvement can help improve social cohesion and a sense of security within the neighbourhood (Thoits, 1995).

## 2.5 Conclusions and Future work

There has always been a gap in traditional data collection methods in regards to daily interactions and in-situ reactions to events. Typically when researchers are looking for individual data there is a certain level of aggregation required because data at the individual level is not collected. Alternatively, data collected at the individual level has to be aggregated to a higher level of geography to maintain anonymity. Twitter is a fairly new method of individual data collection, but it has already been useful within research. Timely information can be delivered about individual's encounters. In addition to the temporal advantage, sentiment and emotional content can also be collected. Including sentiment analysis with Twitter NLP is a logical partnership that should be leveraged within the research community. Although mental health and the relationship to physical health is a large area of study, Twitter remains a largely unused source of information about an individual's emotional status. Neighbourhood profiles remain a large source of health information that is enhanced with the addition of sentiment analysis.

Through the creation of an emotional landscape it became evident that there is a spatial pattern of emotions. Visually displaying the difference between neighbourhood emotions indicated the significant differences between emotional statuses of urban neighbourhoods. The general pattern of emotions shows an interesting link between the newer areas of the city versus the older more established areas of the city. The comparison between NIA and non-NIA neighbourhoods in terms of emotional risk rate have provided insight on areas that require further research and a more in-depth analysis of what may be causing the shift in emotional behaviours.

While acute stress may be temporary, there are more harmful health effects that can occur when consistently exposed to acute stressors. When there is constant exposure to stressors there is a shift in behaviour as well as a negative shift in health outcomes (Steptoe & Feldman, 2001). Temporary negativity may not pose health risks.

Some of the NIA experience increased levels of chronic illnesses such as cardiovascular conditions and diabetes (Eisenhauer, 2001). As mental health becomes a more prominent health issue, it is vital to consider the sentiment of residents in areas that may be consistently negative over a long duration. Addressing in-situ stress could minimize the development of chronic stress long term (Baum et al., 1999). Evidence of a relationship between NIA and emotional polarity presented in this paper has provided perspective on the relationship between the accessibility, or lack of accessibility of resources, which may contribute to an individual's mental and physical health.



## 2.6 References

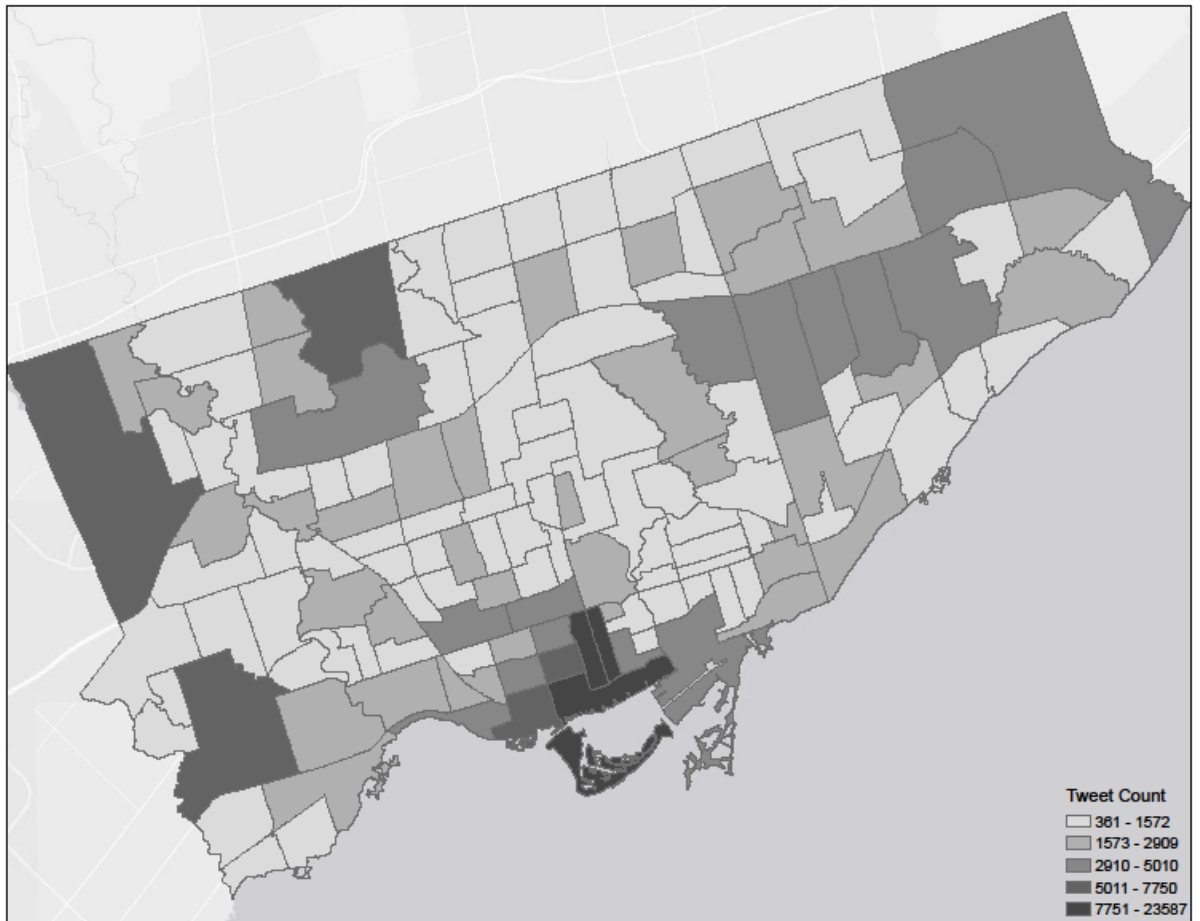
- Baum, A., Garofalo, J. P., & Yali, A. M. (1999). Socioeconomic Status and Chronic Stress: Does Stress Account for SES Effects on Health? *Annals of the New York Academy of Sciences*, 896(1), 131–144. <http://doi.org/10.1111/j.1749-6632.1999.tb08111.x>
- Boardman, J. D. (2004). Stress and physical health: the role of neighbourhoods as mediating and moderating mechanisms. *Social Science & Medicine*, 58(12), 2473–2483. <http://doi.org/10.1016/j.socscimed.2003.09.029>
- De Hollander, A., & Staatsen, B. (2003). Health, environment and quality of life: an epidemiological perspective on urban development. *Landscape and Urban Planning*, 65(1-2), 53–62. [http://doi.org/10.1016/S0169-2046\(02\)00237-2](http://doi.org/10.1016/S0169-2046(02)00237-2)
- DeLongis, A. (n.d.). The Impact of Daily Stress on Health and Mood: Psychological and Social Resources as Mediators. *Journal of Personality and Social Psychology*, 54(3), 486–495.
- Eisenhauer, E. (2001). In poor health: Supermarket redlining and urban nutrition. *GeoJournal*, 53(2), 125–133. <http://doi.org/10.1023/A:1015772503007>
- Ferrill, P. (2011). Python Scripting Utilities. In Pro Android Python with SL4A (pp. 165–194). Apress. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4302-3570-5\\_7](http://link.springer.com/chapter/10.1007/978-1-4302-3570-5_7)
- Freeman, L., & Braconi, F. (2004). Gentrification and Displacement New York City in the 1990s. *Journal of the American Planning Association*, 70(1), 39–52. <http://doi.org/10.1080/01944360408976337>
- Frohlich, K. L. (2013). Area Effects on Behaviour and Lifestyle: The Spatiality of Injustice. In C. Stock & A. Ellaway (Eds.), *Neighbourhood Structure and Health Promotion* (pp. 39–59). Springer US. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4614-6672-7\\_3](http://link.springer.com/chapter/10.1007/978-1-4614-6672-7_3)
- Holahan, C. J., & Moos, R. H. (1991). Life stressors, personal and social resources, and depression: A 4-year structural model. *Journal of Abnormal Psychology*, 100(1), 31–38. <http://doi.org/10.1037/0021-843X.100.1.31>
- Huberman, B. A., Romero, D. M., & Wu, F. (2008). *Social Networks that Matter: Twitter Under the Microscope* (SSRN Scholarly Paper No. ID 1313405). Rochester, NY: Social Science Research Network. Retrieved from <http://papers.ssrn.com/abstract=1313405>
- Leader-Williams, N., Harrison, J., & Green, M. J. B. (1990). Designing protected areas to conserve natural resources. *Science Progress (1933- )*, 74(2 (294)), 189–204.

- Lees, L. (2008). Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance? *Urban Studies*, 45(12), 2449–2470. <http://doi.org/10.1177/0042098008097099>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. <http://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Maas, J., Verheij, R. A., Groenewegen, P. P., Vries, S. de, & Spreeuwenberg, P. (2006). Green space, urbanity, and health: how strong is the relation? *Journal of Epidemiology and Community Health*, 60(7), 587–592. <http://doi.org/10.1136/jech.2005.043125>
- Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Social science & medicine (1982)*, 55(1), 125–139.
- McEwen, B. S. (1998). Stress, adaptation, and disease. Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840, 33–44.
- Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE*, 8(5), e64417. <http://doi.org/10.1371/journal.pone.0064417>
- Mitchell, R., & Popham, F. (2008). Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*, 372(9650), 1655–1660. [http://doi.org/10.1016/S0140-6736\(08\)61689-X](http://doi.org/10.1016/S0140-6736(08)61689-X)
- Murray, C. J., & Lopez, A. D. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet (London, England)*, 349(9063), 1436–1442. [http://doi.org/10.1016/S0140-6736\(96\)07495-8](http://doi.org/10.1016/S0140-6736(96)07495-8)
- Nasar, J. L., & Jones, K. M. (1997). Landscapes of Fear and Stress. *Environment and Behavior*, 29(3), 291–323. <http://doi.org/10.1177/001391659702900301>
- Nasukawa, T., & Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In *Proceedings of the 2Nd International Conference on Knowledge Capture* (pp. 70–77). New York, NY, USA: ACM. <http://doi.org/10.1145/945645.945658>
- NIA Profiles - Demographics - Your City | City of Toronto. (n.d.). Retrieved November 23, 2015, from <http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=e0bc186e20ee0410VgnVCM10000071d60f89RCRD>

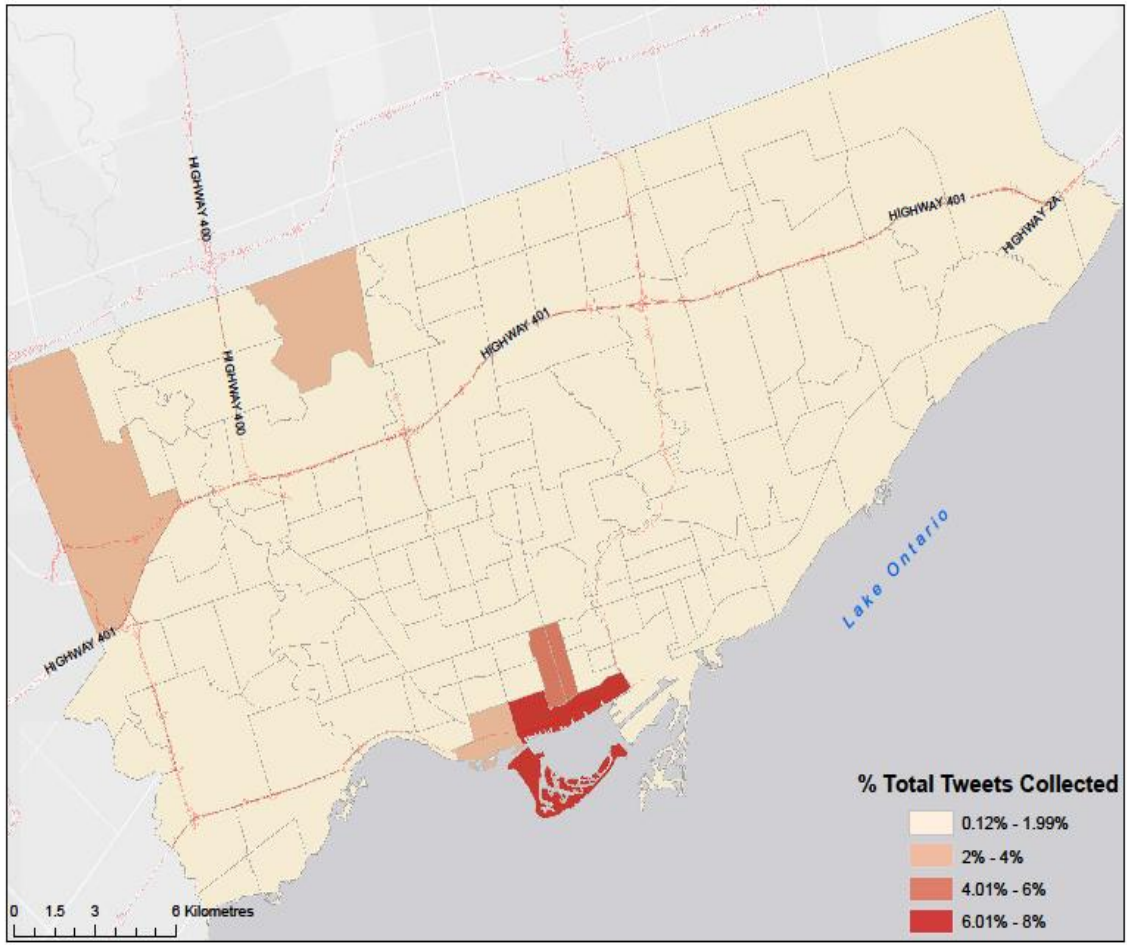
- Plutchik R., (1980). *Emotion: A Psychoevolutionary Synthesis*. Longman Higher Education.
- Public streams. (n.d.). Retrieved June 14, 2016, from <https://dev.twitter.com/streaming/public>
- Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. *ICWSM*, 11, 17-21.
- Sallis, J. F., & Glanz, K. (2006). The Role of Built Environments in Physical Activity, Eating, and Obesity in Childhood. *The Future of Children*, 16(1), 89–108. <http://doi.org/10.2307/3556552>
- Shankardass, K. (2012). Place-Based Stress and Chronic Disease: A Systems View of Environmental Determinants. In P. O'Campo & J. R. Dunn (Eds.), *Rethinking Social Epidemiology* (pp. 113–136). Springer Netherlands. Retrieved from [http://link.springer.com/chapter/10.1007/978-94-007-2138-8\\_6](http://link.springer.com/chapter/10.1007/978-94-007-2138-8_6)
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. CRC Press.
- Sloane, D., Nascimento, L., Flynn, G., Lewis, L., Guinyard, J. J., Galloway-Gilliam, L., ... Yancey, A. K. (2006). Assessing Resource Environments to Target Prevention Interventions in Community Chronic Disease Control. *Journal of Health Care for the Poor and Underserved*, 17(2S), 146–158. <http://doi.org/10.1353/hpu.2006.0094>
- Statistics Canada. (2006). *2006 Percent Population Change* . Retrieved from E-STAT <http://estat.statcan.ca>
- Step toe, A. & Feldman, J. (2001). Neighbourhood problems as sources of chronic stress: Development of a measure of neighbourhood problems, and associations with socioeconomic status and health. *Annals of Behavioral Medicine*, 23(3), 177–185. [http://doi.org/10.1207/S15324796ABM2303\\_5](http://doi.org/10.1207/S15324796ABM2303_5)
- Sykora, M. D., Jackson, T. W., O'Brien, A., & Elayan, S. (2013). Emotive ontology: Extracting fine-grained emotions from terse, informal messages. *Computer Science and Information Systems Journal*.
- Thoits, P. A. (1995). Stress, Coping, and Social Support Processes: Where Are We? What Next? *Journal of Health and Social Behavior*, 35, 53–79. <http://doi.org/10.2307/2626957>
- Torrens, P. M. (2010). Geography and computational social science. *GeoJournal*, 75(2), 133–148. <http://doi.org/10.1007/s10708-010-9361-y>

- Tuan, Y.-F. (1975). Place: An Experiential Perspective. *Geographical Review*, 65(2), 151–165.  
<http://doi.org/10.2307/213970>
- Zhao, D., & Rosson, M. B. (2009). How and Why People Twitter: The Role That Micro-blogging Plays in Informal Communication at Work. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (pp. 243–252). New York, NY, USA: ACM.  
<http://doi.org/10.1145/1531674.1531710>
- Ziersch, A. M., Baum, F. E., MacDougall, C., & Putland, C. (2005). Neighbourhood life and social capital: the implications for health. *Social Science & Medicine*, 60(1), 71–86.  
<http://doi.org/10.1016/j.socscimed.2004.04.02>

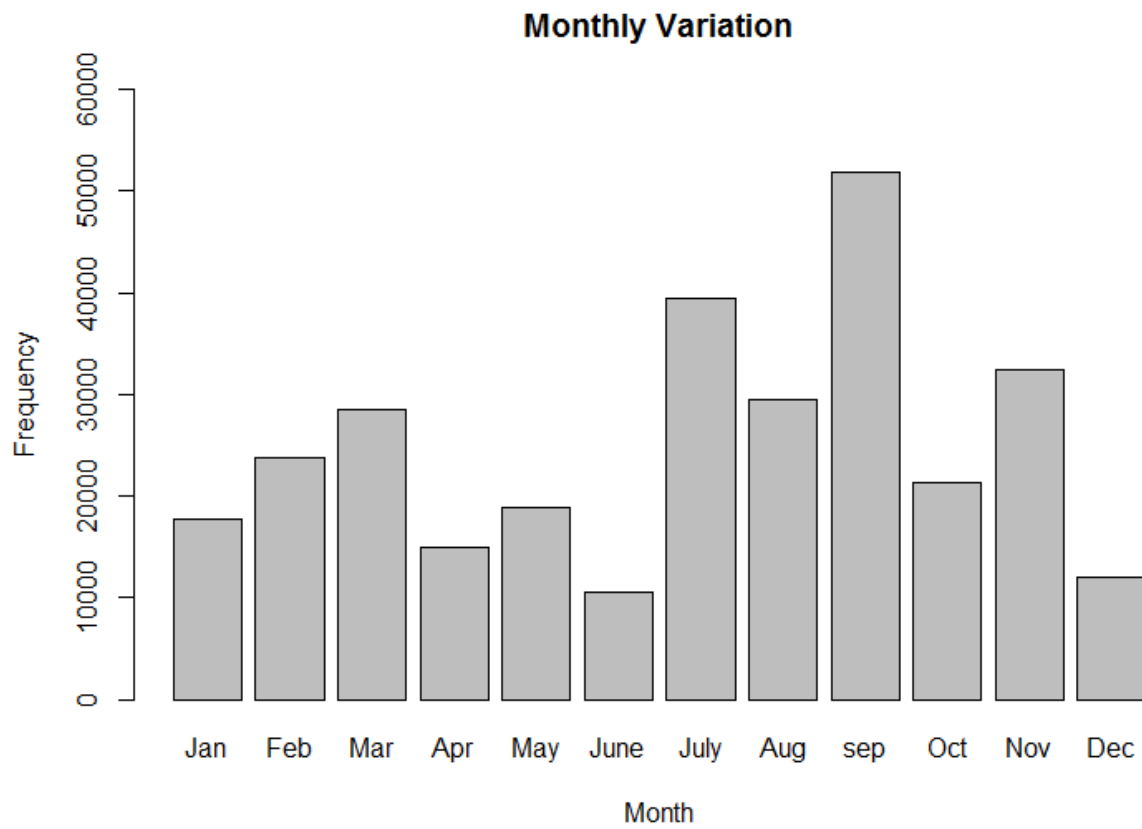
## 2.7 Figures



**Figure 2.2 Tweet count within each neighbourhood**



**Figure 2.3 Percentage of Tweets received from total Tweet collection**



**Figure 2.4 Monthly variation in Tweets**

## Positive Tweets

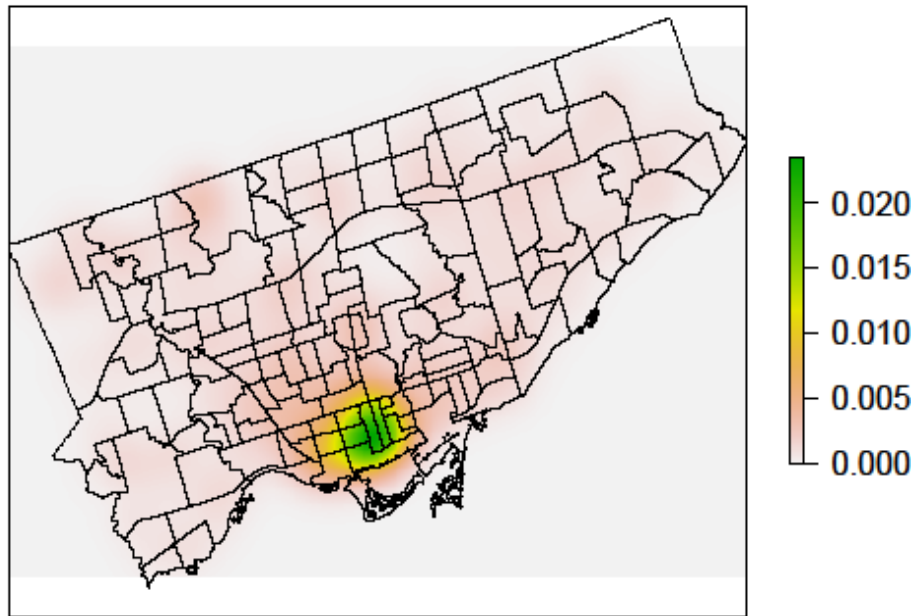


Figure 2.5 Kernel Density Estimate (KDE) of positive Tweets



## Negative Tweets

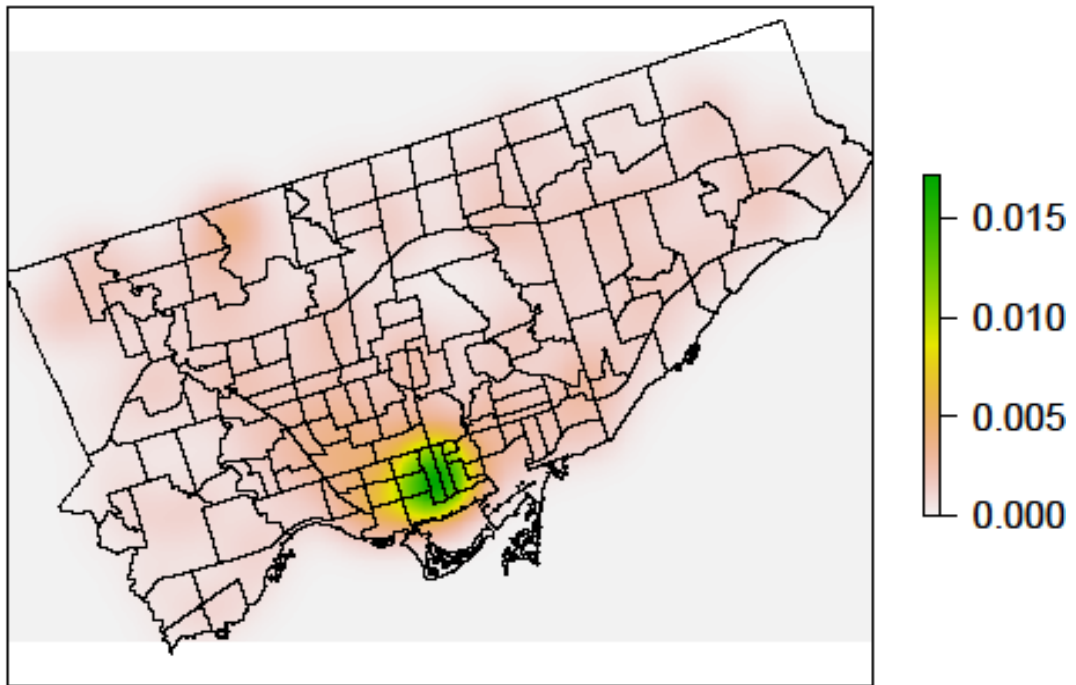


Figure 2.6 Kernel Density Estimate (KDE) of negative emotions

# Positive Tweets/Negative Tweets

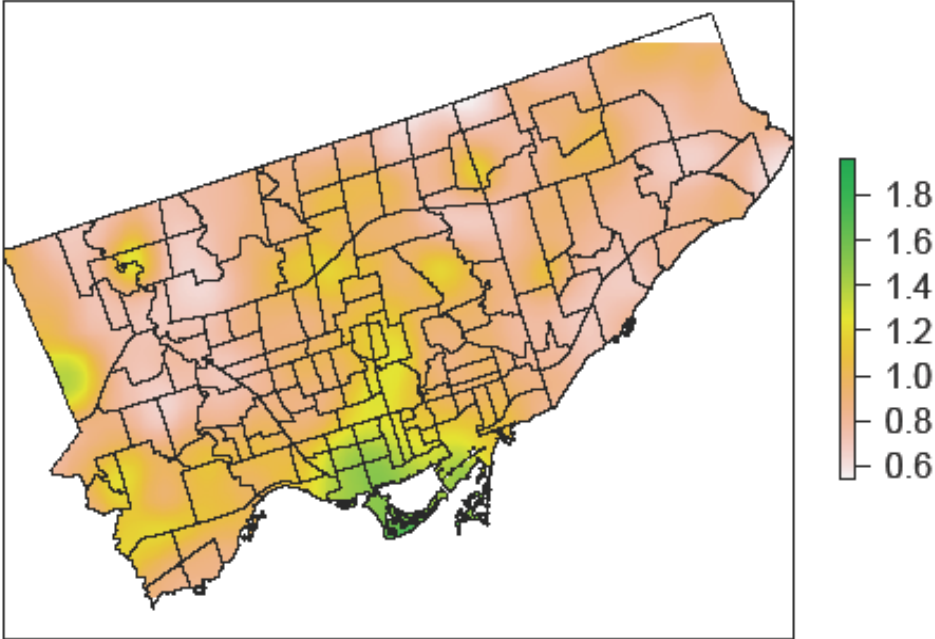
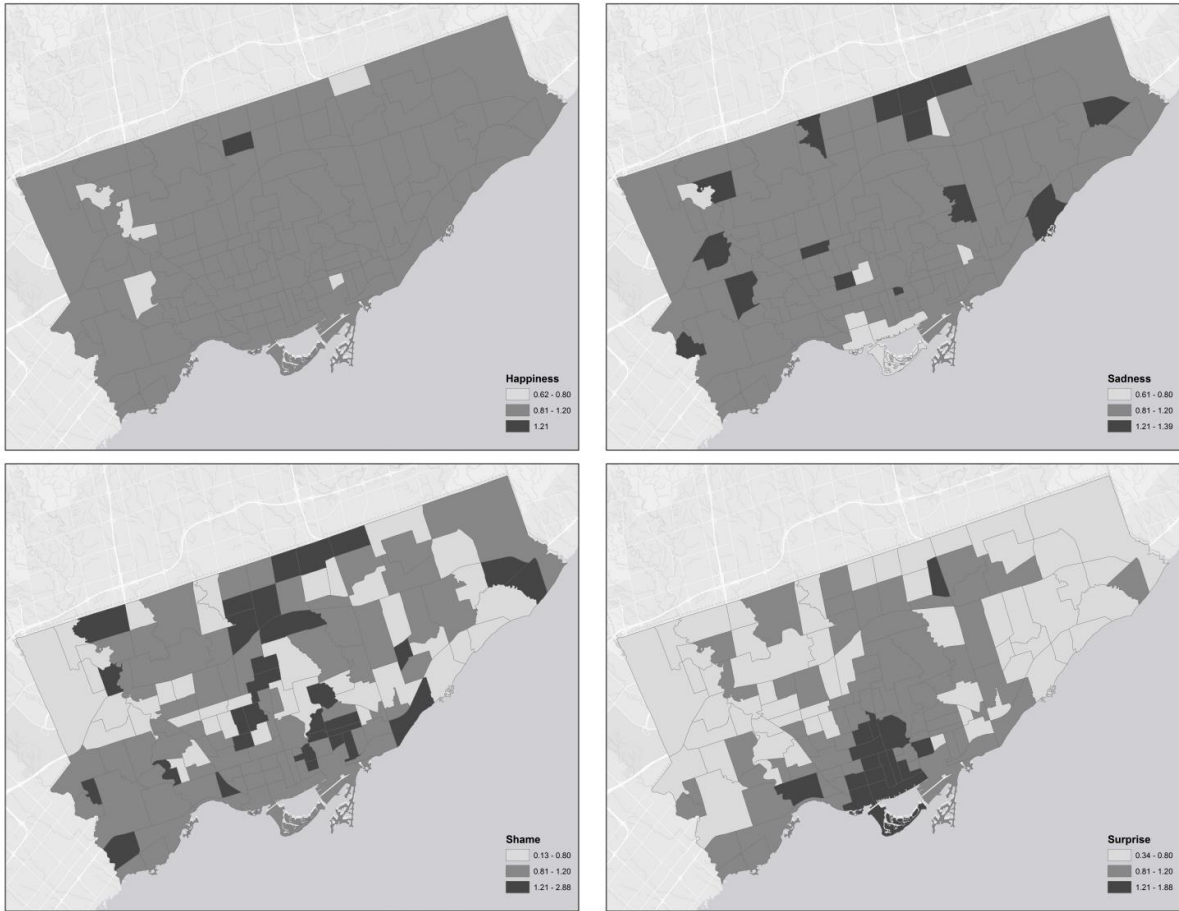


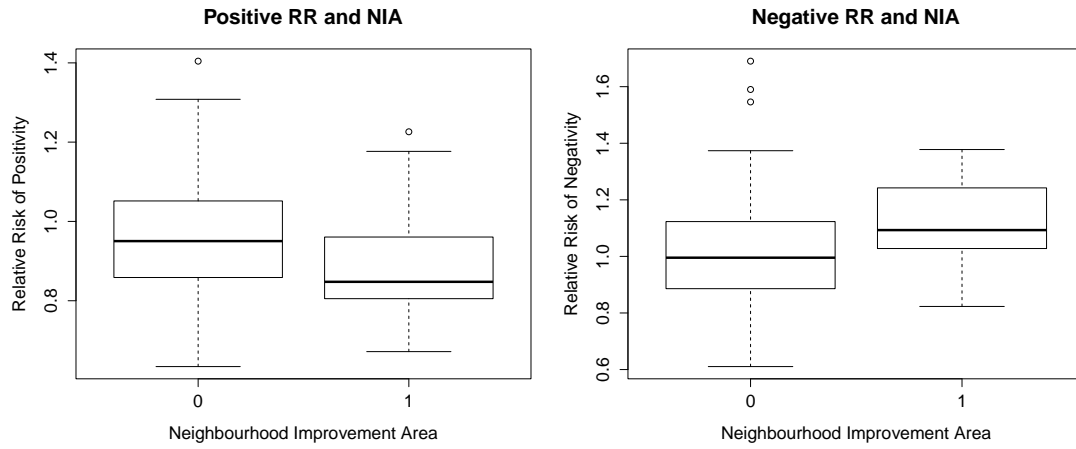
Figure 2.7 Kernel Density Estimate (KDE) ratio of positive/negative emotion



**Figure 2.8 Risk Rates: 1)Anger 2)Confusion 3)Disgust 4)Fear**



**Figure 2.9 Risk Rates: 5)Happy 6)Sadness 7)Shame 8)Surprise**



**Figure 2.10 Risk Rates of Emotion Compared to Neighbourhood Improvement Areas (NIA), where 0 are neighbourhoods outside of NIA and 1 are within NIA**

## Chapter Three

### Abstract

With the content and location data provided in geosocial data sources, there are potentially new opportunities to integrate quantitative and qualitative methods and provide a deeper contextualization of place as it relates to health. In this analysis, neighbourhood profiles were based upon the three main types of neighbourhood resources: built, natural, social. An additional fourth profile included Neighbourhood Improvement Areas (NIA). Using Natural Language Processes (NLP), Tweets were analysed for Tweet sentiment as well as frequency of specific terms. The inclusion of term frequency allowed for more context to better understand the spatial variations in detected emotion. I found that individuals who were within a greenspace experienced increased levels of positivity related to their location while individuals in neighbourhoods with mainly built resources did not experience the same levels of positivity. Overall, this research provided a foundational analysis of emotional and term frequencies within an urban setting.

### 3.1 Introduction

Empirical analyses of the relationships between environment and stress has received more attention in the past few years however, there is still a lot of details required to bridge the knowledge gap. One large gap is what type of environment is most influential to emotional well-being. According to Shankardass (2012), there are three categories of resources that may influence mental health; built, natural, and social. That is to say, environments can influence an individual's well-being based on their perception of that environment, and how each individual interprets their environment. For example, green space-like environments are believed to have a restorative psychological effect on individuals (Maas et al.,2006;Hartig et al.,2011). These positive effects are often increased when coupled with physical activity (Matheson, 2008), however, each individual may perceive green space differently. A recent study of students in Boston found that individuals of certain ethnic descents had a negative association with parks and green space, thus counteracting the expected restorative effects of the environment (Mitchell, 2008). Experiences of place-based stress are a dynamic and individualized process, which can be associated with various negative and positive impacts on human health (Davidson & Milligan, 2004; Shankardass, 2012). The traditional methods of spatial epidemiology, using large scale survey data, aggregated over polygonal units, obtained after an event, are insufficient to capture these sometimes acute processes.

What defines a sense of place has been long explored within the discipline of geography. Emotion is largely subjective to an individual's connection and experiences. Place-based stress is largely connected to an individual's sense of place, when an individual

is in a new environment there is often no sense of place causing place-based stress. Adversely, when an individual encounters the same locations and there is a negative sense of place this can contribute to chronic stress, as it is encountered regularly (Riger et al., 1982; Bowling, 2006). When an individual has a positive sense of place it can help the individual mediate and thus cope with the stress both acute and chronic. For many individuals a positive sense of place is associated with exercise and/or greenspace. Davidson and Milligan (2004) suggest that emotions always take place within multiple levels of geography at multiple scales. They argue that “thinking is different from feeling” implying that there is too much thought about how geography relates to our emotions. Interaction with certain levels of geography largely intensifies the creation of a sense of place.

Daily interactions within an urban center are constantly changing and as a result it can be difficult to classify features of the environment as either positive or negative resources. Although there are visual spatial relationships of negative emotions within Neighbourhood Improvement Areas (NIA), it can be difficult to quantify those relationships. Lack of resources is thought to be the cause of some of the negative health outcomes within the NIA (City of Toronto, n.d.), as closer resources are considered more accessible and are therefore used more frequently regardless of health impact (Berke et al., 2007). Exposure to natural areas and greenspace can be highly restorative for some individuals during high stress times (Ulrich et al., 1991).

With the highly variable nature of stress and the highly variable nature of the population within an urban area, it can be challenging to quantify and develop patterns



simply using a static survey. Through term frequency analysis of key neighbourhood profiles, more unique and identifiable patterns of emotion well-being will be demonstrated. This will help to identify how individuals are coping and mediating stress within the city. Berke et al. (2007) found a positive association between obesity in the older population as a result of limited mobility and access to resources. As accessibility to resources decreased, physical activity also decreased which lead to an increase in obesity (Berke et al., 2007; Foster and Giles-Corti, 2008). The type of resources an individual has access to will therefore influence their health (Ulrich et al., 1991; Evans, 2003).

Sense of place can be vital to an individual experiencing stress, both acute and chronic. The comfort of familiar resources or daily patterns may begin to create a sense of place in addition to a home environment. The familiarity of a location can also help define what resources are used in the coping process. This paper will determine if place-based information be used to enhance and contextualize geo social sentiment mapping of stress. Through text frequency and Term Frequency-Inverse Document Frequency (TF-IDF) analysis of geo-located Tweets within built, natural, and social neighbourhoods, a hierarchal k-means cluster of terms will be created to compare with sentiment patterns in the area. Comparison of resource proximity to Tweets in combination with term frequencies will also provide data on the connection between access and use of resources.

## 3.2 Methods

### 3.2.1 Data

The majority of personal resource data was obtained through a partnership with the Toronto Community Health Profile Partnership (TCHPP) team. The TCHPP initiative began

in the early 1990's to increase the availability of neighbourhood based health and well-being data, to demonstrate inequality in Toronto, and to combine the two in order to create partnerships and collaborations to improve the overall health of neighbourhoods. The data that was obtained from TCHPP included food security with variables such as fast food locations, grocery stores and convenience stores (Table 2). Crime data was also obtained which included non-violent crimes, violent crimes, theft, and vandalism. Point location and type of public recreation sites, as well as immigrant services have also been included in analysis as social resources.

Additional resource data was collected through the City of Toronto open data website, included variables such as neighbourhood boundaries, location of homeless shelters in the form of point data, the location of open park space as point data, location of police facilities in the form of point data, and the location, type, and size of all park space owned and maintained by the city of Toronto.

As an indicator of poor eating habits, fast food establishments were included in the analysis (Sallis & Glanz, 2006). Fast food data includes the location of all fast food establishments within the city of Toronto. Convenience stores were used in analysis to determine the accessibility to quick food options that are not necessarily fast food. It was hypothesized that convenience stores would also indicate unhealthy eating habits (Table 1). Convenience store data includes the location of all convenience stores. Grocery store locations were added as a personal resource to determine how accessible healthy food options were to neighbourhood residents (Crawford, 2013). The grocery store data

includes location as well as type of grocery store (supermarket or local). Recreation centres were included in the analysis as an indicator of healthy stress coping. Proximity to recreation activities is considered a healthy coping mechanism that can help reduce levels of stress and potentially reduce the risk of chronic diseases (Shankardass, 2012). Recreation centre data includes the location of all types of recreation centres both indoor and outdoor.

Lifestyle and neighbourhood safety data was provided by the Neighbourhood Environmental Health and Wellbeing (NEHW) project that was conducted in the City of Toronto. The NEHW study began in 2009 and focused on the individual neighbourhood factors that affect people’s health and well-being, both at the individual and the contextual level. Information such as neighbourhood security and perceived access to resources were also considered.

Theme	Factor	Rationale
Built	Neighbourhood divisions	These will provide the geographic regions that will be analyzed
	Police facilities	May be related to the perceived neighbourhood safety and location of police facilities
	Homeless shelters	May be related to the perceived level of neighbourhood safety
	Supermarkets	Accessibility of 'healthy' affordable food
	Recreation centers	Location of activities (includes softball fields, hockey rinks, swimming pools, soccer fields, and indoor gymnasiums),to measure emotional outcome of physical activity and social

		interaction
	Liquor stores	Determine the accessibility of alcohol
	Fast food restaurants	Determine the accessibility of fast food
	Convenience stores	Considered a food resource that will help to determine the accessibility of 'unhealthy' resources
Natural	Green Space	Will be included in the accessibility analysis as a method for coping
Social	Average household income	Used as a socioeconomic value to assess stress levels
	Unemployment rate	Used as a socioeconomic value to assess stress levels
	Services for immigrants	Used to determine accessibility of services based on location
	Canadian Community health survey (CCHS)	Comparing and contrasting stress clusters found within Twitter
	Tweets	Content will be used for accessibility measures as well as comparison of sentiment
	Crime	Used to determine neighbourhood trust and cohesion and can also contribute to stress

**Table 1 Categorical organization of data into “theme”**

### 3.2.2 Proximity to Resources

Resource access is not always directly related to the use of that resource. Research has shown that simply walking within close proximity to green space may help improve emotion and positivity (Nisbet et al.,2010; Ambrey and Fleming,2013;MacKerron and

Mourato, 2013). Similar to the effect of greenspace, is a negative influence from more negative resources. Increasing the frequency of resources should reflect more positive Tweets because coping and mediation is potentially more accessible to individuals.

Green space research has shown that simply walking within close proximity to green space may help improve emotional polarity (Nisbet et al.,2010; Ambrey and Fleming,2013;MacKerron and Mourato, 2013). While an individual who may be walking by a fast food resource while on a diet may be negatively impacted emotionally.

As a baseline for further analysis, a basic proximity analysis was performed. It was estimated that if a Tweet/individual was within a 10 minute walking distance (500 m) of a resource point source, that individual was more likely to use that resource. The core location of Tweets does correlate to the neighbourhoods with the most number of resources. It is hypothesized that as the number of resources decreases, the emotional pattern would shift from positive to mostly negative. The sidewalk network of Toronto was buffered to 10m to include the sidewalk on each side of a street. Tweets that were within the sidewalk network and were within 500m of a resource were considered within walking distance to a resource.

### 3.2.3 Neighbourhood Profiles

Within an urban area there are typically a few niche neighbourhoods that can develop their own set of distinct characteristics and reputations. In the city of Toronto, some of the more well-known groupings of neighbourhoods that form smaller niches include “Little Italy”, “the Danforth”, or “financial district”. There are also neighbourhoods

that are known for their more positive resources such as income or age of the neighbourhood. Typically, neighbourhoods that are older in age with higher income have increased neighbourhood cohesion and perceived safety, when compared to neighbourhoods that experience high levels of resident changes.

Based on research conducted by Shankardass in 2012, there are three main types of and resources within a neighbourhood: built, natural, social. Centered on these focal resource categories, neighbourhoods were classified into 4 profiles for this analysis: built, natural, and social resources, as well as NIA (Table 2). Neighbourhoods that were classified as built neighbourhoods were those that contained the most resources such as grocery stores, fast food locations, and convenience stores. The point in polygon sum function was used to determine which neighbourhoods contained the largest amount of resources within the boundaries.

Natural neighbourhoods were classified as neighbourhoods where the majority of the area within the boundary was what the City of Toronto outlined as environmentally significant areas. Based on the definition of a social neighbourhood proposed by (Campbell et al., 1986), neighbourhoods that contained or had access to a university such as York University, the University of Toronto, or Ryerson University were classified as social neighbourhoods. Neighbourhoods that contained the most recreation locations such as soccer fields, baseball diamonds, or hockey arenas were also classified as social neighbourhoods. NIA were areas considered to be low in resources and were generally predicted to have lower or more negative health outcomes including mental health. The four neighbourhood profiles will form the basis of this analysis.

Some neighbourhoods had a large number of built resources as well as social resources and therefore were included as part of both profiles. The neighbourhood profile classifications are not exclusive, as in some neighbourhoods may have increased levels of all resources as well as a large amount of Tweets. For the purpose of this study neighbourhoods that had the highest amount of a resource category were included in a profile.

CLASSIFICATION	DESCRIPTION
Built	<ul style="list-style-type: none"> <li>-Neighbourhoods with a high number of resources that are not naturally occurring within the landscape</li> <li>-There are positive and negative built resources</li> <li>i.e. Grocery stores, fast food restaurants, and convenience stores</li> </ul>
Natural/Greenspace	<ul style="list-style-type: none"> <li>-Neighbourhoods that have resources that are found naturally occurring within the city landscape</li> <li>-Greenspace is theoretically used as a healthy coping mechanism</li> <li>i.e. Parks, greenspace, and protected greenspace</li> </ul>
Social	<ul style="list-style-type: none"> <li>-Locations where factors such as socio-economic status are largely influential</li> <li>-There are both positive and negative social resources</li> <li>-Neighbourhoods that contain the most social resources were included for analysis</li> </ul>

	i.e. Universities and recreation locations
Neighbourhood Improvement Area (NIA)	-Neighbourhoods that statistically have lower income, lower health outcomes, and increased risk of mental illness

**Table 2 Neighbourhood resource classification**

**3.2.4 Text Mining**

Tweets from within the four neighbourhood profiles were used as part of the analysis of term frequency. The purpose of this portion of the study was to determine the most frequent terms used within each of the four neighbourhood profiles that were created. This was based on research that suggests that individuals who have increased access to resources are more positive compared to those with decreased access (Stigsdotter and Grahn, 2011). Term frequencies were completed in the programming language R using the text mining package. The text mining package eliminated all stop words, converted all letters into lower case, and removed all punctuation. After Tweet preprocessing was complete, all of the Tweets for that area were put into a term document frequency matrix as well as a document term frequency matrix where each document was a Tweet. Term document frequency matrix is comprised of the number of occurrences of a term within each document. A document term frequency matrix is the number of documents that contains the same term. After creating a term document matrix, the TF-IDF was calculated to minimize the amount of “common” word occurrences. Using the TF-IDF created a better overview of terms rather than the raw frequencies that may have included regular parts of speech (Zhu and Xiao,2011).



### 3.2.5 Hierarchical Text Clusters

Hierarchical text clusters were created to show the connection between words that are used most frequently together. This method also helped to determine the context in which most words were used. Knowing the terms that are used together helps build the emotional profile within neighbourhoods. The hierarchical text clusters also sift through potentially sarcastic or satirical Tweets that are not interpreted through basic text mining. Hierarchical text clusters were created from the text mined through the Tweets from within each neighbourhood profile. The text mining package in R was used to remove all sparse terms from the DTM created for Tweet text mining. A distance matrix was then computed and then the dendrogram for the “cluster tree” was plotted. The number of clusters was then arbitrarily chosen to represent the most commonly occurring combinations within the Tweets. The red outline of common word clusters was then added to further demonstrate patterns and context.

### 3.2.6 Sentiment Pattern

An individual’s emotional outcome can be affected by how they interact with their environment. It was estimated that increased access to resources such as recreation in greenspace, will have a more positive influence emotionally on a Tweet (Nisbet et al.,2010; Ambrey and Fleming,2013; MacKerron and Mourato, 2013). Conversely, it was hypothesized that decreased access to resources will have a more negative influence on the Tweet. Tweets located within the NIA were expected to have a more negative emotional score.

Sentiment analysis was performed on the Tweets using an algorithm called EMOTIVE where Tweets are parsed similar to the way that they are in the text mining

process, however, each word in the Tweet is assigned an emotion as well as a score based on the strength of that emotion (described in detail in Chapter 2). Some Tweets may have multiple emotions but there are words classified as intensifiers that will add to the score of that emotion. This means that whichever emotion had more intensifier words associated with it, would receive a higher score. The emotion that had the highest score within a Tweet would then be considered the dominant emotion. To enhance a neighbourhood profile with sentiment analysis, the most dominant emotion within each Tweet will be assigned to the neighbourhoods within each neighbourhood profile.

### 3.3 Results

#### 3.3.1 Neighbourhood Profiles

Most of the neighbourhood profiles included a broader spatial distribution throughout the city. The social neighbourhood profile was the only profile that was mostly located in the downtown core of the city. The creation of neighbourhood profiled enabled a broader perspective on resource use outside of the downtown core (Figure 3.1).

#### 3.3.2 Proximity

It was determined there was a total of 82365 tweets within the sidewalk network. Of the Tweets from individuals on the sidewalk network, 47269 (57%) were positive. There was a total of 77640 (92.4%) within a 500m walking distance to a resource. Additionally, there were 44526 positive Tweets within walking distance of a resource within all of the main neighbourhood profiles. The Tweets that were closest to built resources were primarily located in the downtown core. Many resources in the periphery of the city were not close to an individual's tweet location. Tweets that were located within the greenspace neighbourhood profile had the least number of Tweets within a close

proximity to a built resource. The built neighbourhood profile contained the highest number of tweets within a walking distance.

### 3.3.3 Text mining

It was determined that although the emotional frequency of the NIAs appears to be happy, that is not the most evident when looking at the most frequently used terms within those areas. The top three most frequent terms within a neighbourhood improvement area were “amazing”, “bad”, and “can” (Figure 3.2). It was determined that there were a lot of mixed emotions within the neighbourhood improvement areas and that the majority of the Tweet’s content were related to the future and less about the present.

When the minimum threshold for term frequency was set to 100, the only term was recognized was “happy”. When the 10 most frequently used terms within a green space were reviewed, it was determined that they were not necessarily directly related to that green space (Figure 3.3). For example, the top three most frequently used terms within a green space neighbourhood profile were “always”, “amazing” and “back”.

The built neighbourhood profile provided an interesting array of most frequently used terms, however, the top three most frequently used terms were “amazing”, “back” and “bad”(Figure 3.3). Within the built neighbourhood profile the most frequently used terms were primarily related to current events but also incorporated a temporal aspect to the content.

The majority of the social neighbourhood profile Tweets were derived from Tweets that were collected within the university area. Within the social neighbourhood profile, the

majority of the most frequently used terms were positive and were based more in the present tense. The top three most frequent terms in the social neighbourhoods were “amazing”, “bad”, and “can”.

#### 3.3.4 Hierarchical Text Clusters

Individually, the top 10 terms used within each neighbourhood profile are very similar. However, when they are combined with the most frequently used multiple term expressions, they are quite different. Figure 3.4 is the hierarchical text cluster for the built neighbourhood profile. The majority of the terms that are clustered together are positive and make reference to the present using the term “today”. There is also a slight reference to the future through the “looking, forward” cluster. There is also some confusion as evident in the cluster “don’t get”. Overall, the clustered terms within the built neighbourhoods imply a more positive context to the individual terms that were mined.

Within the NIA (Figure 3.5), there are some slightly different patterns. A few of the clusters include statements such as “get, mad”, and “today, sucks”. There are some positive connections such as “make, proud” and “enjoy, hope”.

The social neighbourhood profile followed the same general phrases as built and NIA profiles (Figure 3.6). There is one cluster of “really sad” and “get, mad” however, there are also positive phrases for instance “will, enjoy” and “excited, good”.

Within the Greenspace neighbourhood profile there was a mention of the “park” however, it was clustered with the word “sad” (Figure 3.7). There was also a mention of “looking forward” and “excited”. In reference to other people there is a grouping that

contains “really, great, people”. There is also mention of “today, enjoy” keeping the context of the Tweets within the present day.

### 3.4 Discussion and Conclusion

The terms that were most frequently used phrases in the NIAs that are related to the present could be an indication that there is a positive change occurring within the neighbourhood. As evident in the hierarchical text clusters people are focusing their Tweets more on the present or future, rather than on the past. The total number of NIAs within the City of Toronto has more than doubled in the last 10 years. The increased use of future tense in Tweets from these areas may indicate that there is a desire for change, including gentrification and improvement within those areas.

Providing better representation of emotional stressors and negators can lead to policy implications and more positive change. As part of positive gentrification and neighbourhood improvement, some cities have adopted the urban environment policy. The urban environment policy is primarily business focused with urban residents being a secondary focus (Freeman, 2004;Lees,2008). In recent years there has been a shift from suburban living to a more urbanized style of living, however the urban city has been slow to adapt to such changes. Along with the lifestyle change there is a change in demographic to a more mobile and young demographic, which requires adjustment to both living accommodations as well as recreation spaces. In the past, an indoor gym facility or soccer field would have sufficed as a method of mediating stress (Boardman, 2004;DeLongis, n.d.). Presently there is a need for more open space within the confines of the city limits. As stressors and demographics change, the urban city should be changing.

The data gathered from the social neighbourhood profile indicated that the most frequently used terms were based primarily in the present tense. This may be an indication that the individuals, who are mostly students in the university neighbourhoods, are focused on their current pursuits. Since recreation locations were also included in the social neighbourhood profile, the focus on the present tense could also be associated with the recreational activities that they were participating in. For example, when individuals live-Tweeted about a sporting event that they were participating in at one of the recreation locations.

As previously mentioned, Tweets that occurred within the green space neighbourhood profile provided interesting insight into how green space can affect one's thoughts. Some of the more frequently used terms that were Tweeted in the green space profile showed feelings of gratitude and positivity such as love and hope. The positivity that occurs within the natural neighbourhood profiles could be indicative of how green space often contributes to an individual's mental coping process (Ellaway et al. ,2013;Diez Roux,2001). However, the negative phrases used such as "park, sad" could be an indicator that green space does not always have a positive mediating effect. Adversely, the negative sentiment could be an individual who has gone to a park to help cope with sadness.

Term frequency analysis completed in each of the different neighbourhood profiles determined that the same terms were appearing in multiple profiles. Three words appeared in each of the most frequently used terms: "amazing", "bad", and "can". The word "amazing", which is a positive expression, was consistently one of the top three words used in Tweets, and appeared in each of the neighbourhood profiles. The term "bad" was the

second most frequently used term across all neighbourhoods and was found to be one of the top three most frequently used terms in both the built neighbourhood, as well as in the social neighbourhood. Both of these words have strong emotional connotations and further contributes to the idea that individuals are both highly positive and highly negative on social media (Markwick and Boyd,2012). Some keywords or phrases such as “happy birthday” may have skewed the data within the text mining process. Most instances of “Happy Birthday” were removed when spelled out in exactly that way however, some iterations were spelled in a different way therefore, not all instances of “Happy Birthday” were removed.

The number of resources in the city of Toronto may be considered somewhat insufficient based on the demand and population size. The built neighbourhood profile incorporated the neighbourhoods that had the highest number of built resources within the neighbourhood boundaries although, there were only 362 Tweets out of more than 75,000 within the built neighbourhood profile that were located within close proximity to a built resource. Access to resources, both healthy and unhealthy, provides an outlet for individuals to cope and process stress and emotions. When there is a lack of resources in an area, individuals are unable to experience proper coping techniques which may lead to potentially negative health outcomes.

With the establishment of niche neighbourhoods, there is often an associated reputation that goes along with that neighbourhood (Bowling, 2006). It has become evident that the initial reputation of an area is not necessarily accurate. For example, there are resources within NIA, however, when focusing on the density and location of Tweets, it

was determined that those resources did not necessarily improve neighbourhood sentiment. There is also a slight misconception that neighbourhoods that may be lacking resources are more negative when that is not the case.

There is contradicting research on how individuals interact with the resources around them, and whether they value proximity or satisfaction. A possible consideration for future research could be to consider that access to a resource does not necessarily translate into use of the resource. For example, the number of resources found within a neighbourhood does not necessarily correlate to the use of those resources. When an individual has a negative experience while using a resource in their neighbourhood, that individual may not return to that resource even though it may be the closest resource to their house. Although Twitter does provide information about the patterns of individual's preference for resources spatially, it does not necessarily mean that the individuals are using those resources or that they are enjoying those resources.



### 3.5 References

- Ambrey, C., & Fleming, C. (2013). Public Greenspace and Life Satisfaction in Urban Australia. *Urban Studies*, 0042098013494417. <http://doi.org/10.1177/0042098013494417>
- Berke, E. M., Koepsell, T. D., Moudon, A. V., Hoskins, R. E., & Larson, E. B. (2007). Association of the Built Environment With Physical Activity and Obesity in Older Persons. *American Journal of Public Health*, 97(3), 486–492. <http://doi.org/10.2105/AJPH.2006.085837>
- Bowling, A. (2006). Do perceptions of neighbourhood environment influence health? Baseline findings from a British survey of aging. *Journal of Epidemiology & Community Health*, 60(6), 476–483. <http://doi.org/10.1136/jech.2005.039032>
- Campbell, K. E., Marsden, P. V., & Hurlbert, J. S. (1986). Social resources and socioeconomic status. *Social Networks*, 8(1), 97–117. [http://doi.org/10.1016/S0378-8733\(86\)80017-X](http://doi.org/10.1016/S0378-8733(86)80017-X)
- Charreire, H., Weber, C., Chaix, B., Salze, P., Casey, R., Banos, A., ... Oppert, J.-M. (2012). Identifying built environmental patterns using cluster analysis and GIS: Relationships with walking, cycling and body mass index in French adults. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1), 59. <http://doi.org/10.1186/1479-5868-9-59>
- Crawford, F. (2013). Neighbourhoods and Healthy Nutrition. In C. Stock & A. Ellaway (Eds.), *Neighbourhood Structure and Health Promotion* (pp. 305–316). Springer US. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4614-6672-7\\_16](http://link.springer.com/chapter/10.1007/978-1-4614-6672-7_16)
- Davidson, J., & Milligan, C. (2004). Embodying emotion sensing space: introducing emotional geographies. *Social & Cultural Geography*, 5(4), 523–532. <http://doi.org/10.1080/1464936042000317677>
- Diez Roux, A. V. (2001). Investigating Neighbourhood and Area Effects on Health. *American Journal of Public Health*, 91(11), 1783–1789.
- Ellaway, A., Ferguson, N., Lamb, K., & Ogilvie, D. (2013). Access to Health-Promoting Facilities and Amenities. In C. Stock & A. Ellaway (Eds.), *Neighbourhood Structure and Health Promotion* (pp. 117–126). Springer US. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4614-6672-7\\_7](http://link.springer.com/chapter/10.1007/978-1-4614-6672-7_7)

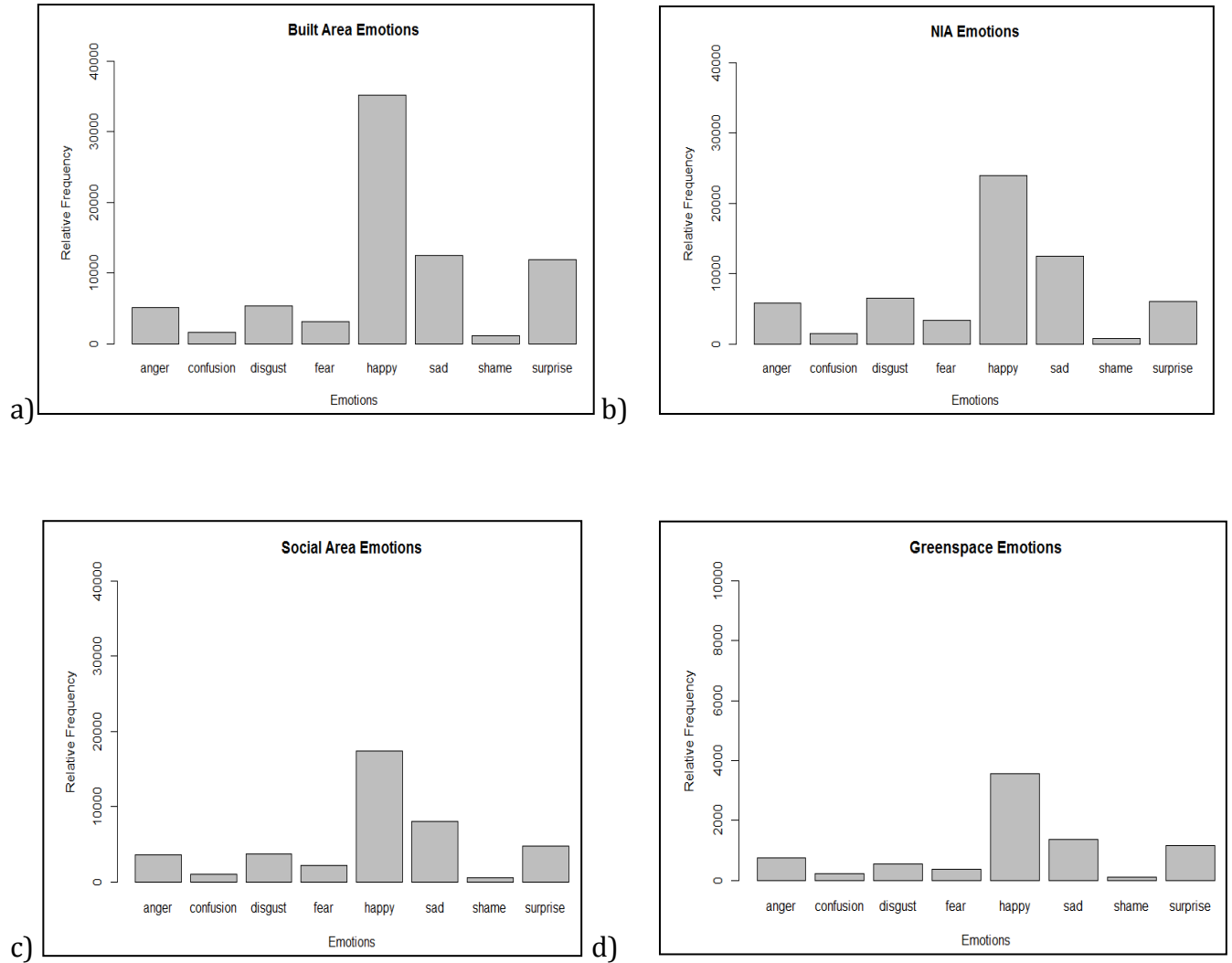
- Evans, D. G. W. (2003). The built environment and mental health. *Journal of Urban Health*, 80(4), 536–555. <http://doi.org/10.1093/jurban/jtg063>
- Foster, S., & Giles-Corti, B. (2008). The built environment, neighborhood crime and constrained physical activity: An exploration of inconsistent findings. *Preventive Medicine*, 47(3), 241–251. <http://doi.org/10.1016/j.ypmed.2008.03.017>
- Hartig, T., Berg, A. E. van den, Hagerhall, C. M., Tomalak, M., Bauer, N., Hansmann, R., ... Waaseth, G. (2011). Health Benefits of Nature Experience: Psychological, Social and Cultural Processes. In K. Nilsson, M. Sangster, C. Gallis, T. Hartig, S. de Vries, K. Seeland, & J. Schipperijn (Eds.), *Forests, Trees and Human Health* (pp. 127–168). Springer Netherlands. Retrieved from [http://link.springer.com/chapter/10.1007/978-90-481-9806-1\\_5](http://link.springer.com/chapter/10.1007/978-90-481-9806-1_5)
- Howell, A. J., & Passmore, H.-A. (2013). The Nature of Happiness: Nature Affiliation and Mental Well-Being. In C. L. M. Keyes (Ed.), *Mental Well-Being* (pp. 231–257). Springer Netherlands. Retrieved from [http://link.springer.com/chapter/10.1007/978-94-007-5195-8\\_11](http://link.springer.com/chapter/10.1007/978-94-007-5195-8_11)
- Maas, J., Verheij, R. A., Groenewegen, P. P., Vries, S. de, & Spreeuwenberg, P. (2006). Green space, urbanity, and health: how strong is the relation? *Journal of Epidemiology and Community Health*, 60(7), 587–592. <http://doi.org/10.1136/jech.2005.043125>
- MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, 23(5), 992–1000. <http://doi.org/10.1016/j.gloenvcha.2013.03.010>
- Marwick, A. E., & Boyd, D. (2011). I Tweet honestly, I Tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. <http://doi.org/10.1177/1461444810365313>
- Matheson, F. I., Moineddin, R., Dunn, J. R., Creatore, M. I., Gozdyra, P., & Glazier, R. H. (2006). Urban neighbourhoods, chronic stress, gender and depression. *Social Science & Medicine*, 63(10), 2604–2616. <http://doi.org/10.1016/j.socscimed.2006.07.001>
- Mitchell, R. (2008). Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*, 372(9650), 1655–1660.

- Mostafa, M. M. (2013). An emotional polarity analysis of consumers' airline service Tweets. *Social Network Analysis and Mining*, 3(3), 635–649. <http://doi.org/10.1007/s13278-013-0111-2>
- NIA Profiles - Demographics - Your City | City of Toronto. (n.d.). Retrieved June 14, 2016, from <http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=e0bc186e20ee0410VgnVCM10000071d60f89RCRD>
- Nisbet, E. K., Zelenski, J. M., & Murphy, S. A. (2010). Happiness is in our Nature: Exploring Nature Relatedness as a Contributor to Subjective Well-Being. *Journal of Happiness Studies*, 12(2), 303–322. <http://doi.org/10.1007/s10902-010-9197-7>
- Pak, A., & Paroubek, P. (2011). Twitter for Sentiment Analysis: When Language Resources are Not Available. In *2011 22nd International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 111–115). <http://doi.org/10.1109/DEXA.2011.86>
- Riger, S., Gordon, M. T., & LeBailly, R. K. (1982). Coping with urban crime: Women's use of precautionary behaviors. *American Journal of Community Psychology*, 10(4), 369–386. <http://doi.org/10.1007/BF00893977>
- Sallis, J. F., & Glanz, K. (2006). The Role of Built Environments in Physical Activity, Eating, and Obesity in Childhood. *The Future of Children*, 16(1), 89–108. <http://doi.org/10.1353/foc.2006.0009>
- Shankardass, K., & Dunn, J. R. (2012). How Goes the Neighbourhood? Rethinking Neighbourhoods and Health Research in Social Epidemiology. In P. O'Campo & J. R. Dunn (Eds.), *Rethinking Social Epidemiology* (pp. 137–156). Springer Netherlands. Retrieved from [http://link.springer.com/chapter/10.1007/978-94-007-2138-8\\_7](http://link.springer.com/chapter/10.1007/978-94-007-2138-8_7)
- Stigsdotter, U. K., & Grahn, P. (2011). Stressed individuals' preferences for activities and environmental characteristics in green spaces. *Urban Forestry & Urban Greening*, 10(4), 295–304. <http://doi.org/10.1016/j.ufug.2011.07.001>
- Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, 11(3), 201–230. [http://doi.org/10.1016/S0272-4944\(05\)80184-7](http://doi.org/10.1016/S0272-4944(05)80184-7)
- Zhu, D., & Xiao, J. (2011). R-tfidf, a Variety of tf-idf Term Weighting Strategy in Document Categorization. *2011 Seventh International Conference on Semantics, Knowledge and Grids*. doi:10.1109/skg.2011.44

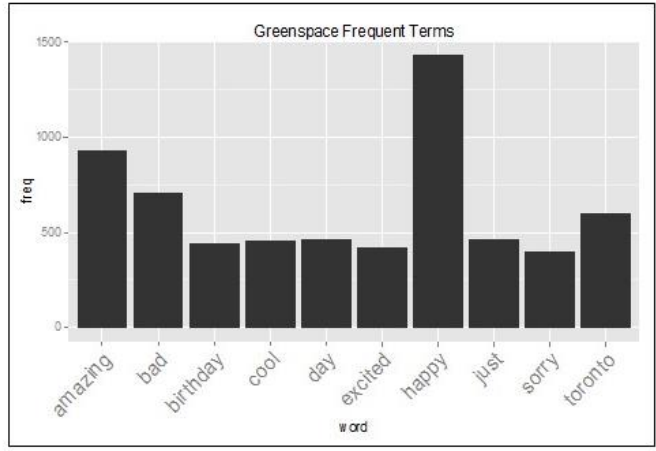
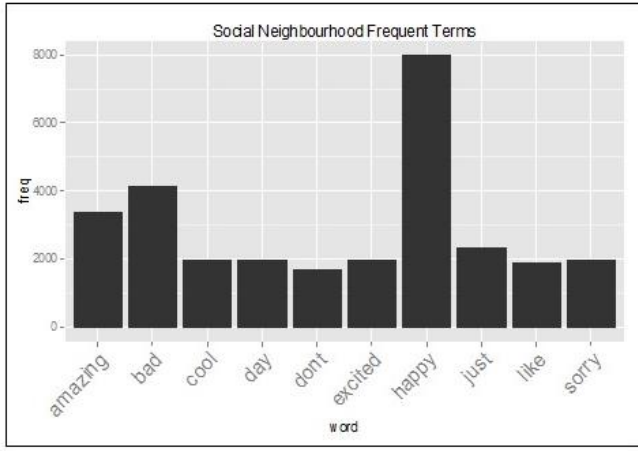
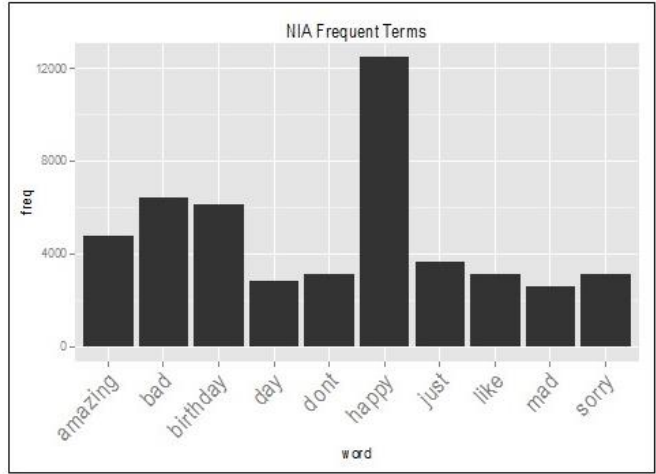
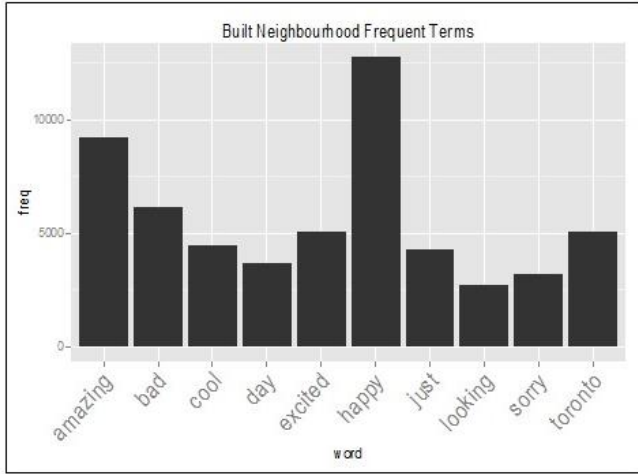
### 3.6 Figures



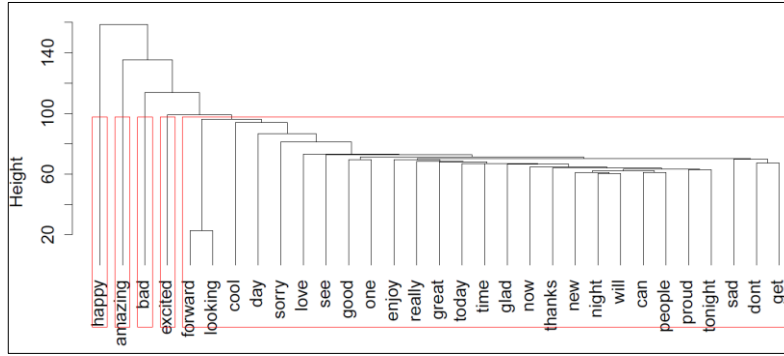
**Figure 3.1. a) Built Neighbourhoods , b) Neighbourhood Improvement Areas (NIA) , c) Social Neighbourhoods, and d) Natural Areas**



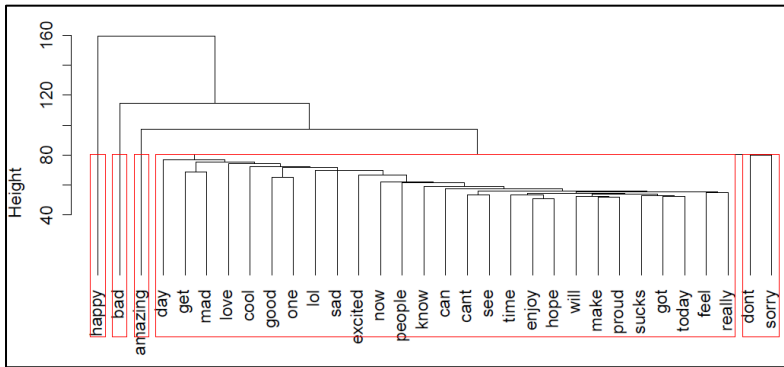
**Figure 3.2 Histogram of emotional frequencies within the 4 neighbourhood profiles. a)Built , b)NIA, c)Social, d)Greenspace**



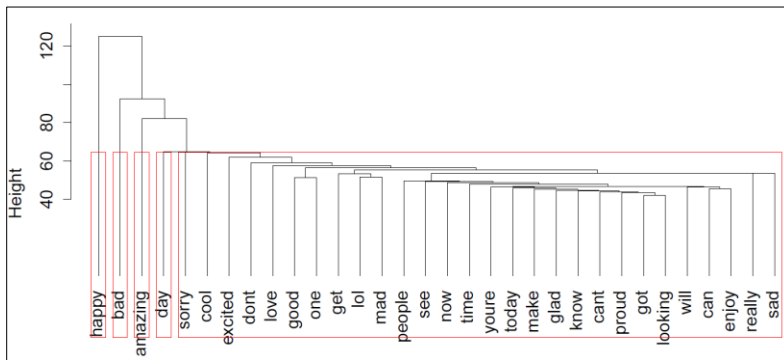
**Figure 3.3 Histogram of the most frequently used terms in the neighbourhood. 1.Built Neighbourhoods ,2.NIA, 3.Social Neighbourhoods, 4.Green Areas**



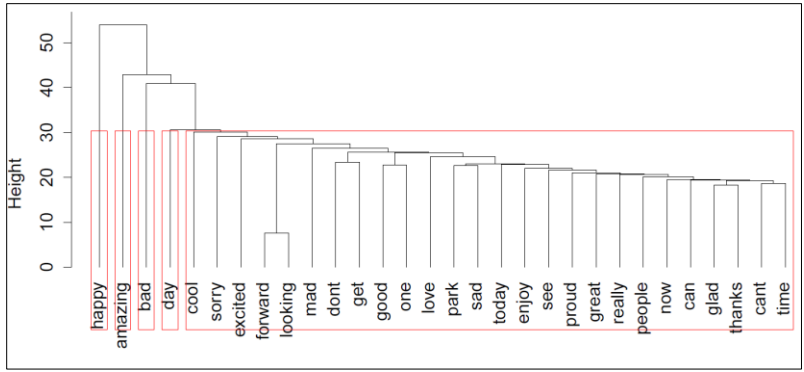
**Figure 3.4 Built neighbourhood word clusters**



**Figure 3.5 NIA word clusters**



**Figure 3.6 Social Neighbourhood word clusters**



**Figure 3.7 Neighbourhoods with Greenspace text clusters**



## Chapter Four

### 4.0 Conclusion

#### 4.1 Discussions and Conclusions

As natural language processes become more diverse and more dynamic, there is intrinsic value and incorporating them into quantitative analysis. Sentiment analysis has been the most recent addition to NLP, and it is becoming more widely incorporated into research. Researchers should be cautious when using geo social data as a form of sentiment analysis because due to the demographic of the users, geo social platforms are more widely used by a younger demographic. This thesis focused on a few aspects of geo social sentiment analysis. The main objective of this research was to establish a link between the built environments and individuals. Through a variety of different methods including kernel density estimates, term frequencies, and risk rates, there were clear patterns established within the research.

Chapter two found that there was a quantitative connection between emotional health and NIA. This finding is important because poor mental health as well as physical health is included in the criteria for establishing an NIA. By creating emotional landscapes, it was determined that there are other neighbourhoods that are consistently negative and are low in resources compared to other neighbourhoods. This information could be used to reevaluate the neighbourhoods that may need to be included as part of a NIA. The box plots in Figure 2.4 demonstrate that there is increased positivity in the areas around the neighbourhood improvement areas. One possible reason for this is that there is increased

access to resources outside of the NIA. For some neighbourhood improvement areas it is simply the age of the neighbourhood infrastructure that requires improvement.

An interesting finding was that the downtown core experience more overall positivity than the periphery neighbourhoods. When comparing the risk rate maps in Figure 2.3, the negative emotions are inversely proportional to the positive emotions within the city. This is of particular interest because as a result of the rise in condo culture, there is a mix of business and living in the downtown area now. In previous years, the downtown core would have been predominantly comprised of businesses, while residences were primarily located in the periphery of the city. With the increase of living space in the downtown core, there has been a shift towards the development and construction of resources in these neighbourhoods to meet the growing demand of the population. Conversely, this has resulted in less focus on the development of resources in the peripheries of the city. This theory does coincide with the location of neighbourhood improvement areas.

Chapter 3 focused on the interaction between resources and individuals. Through the use of term frequencies, it was estimated that if an individual is interacting with the resource around them, it will be included as a term in the Tweet. As Figure 3.3 shows, there are not many frequently occurring terms that are related to any resources, aside from the larger city itself. One interesting thing to note is that although one of the most frequently used terms is happy in each of the four neighbourhood profiles, that is not necessarily the predominant emotion associated with that area. The same is true for the emotional

frequencies when comparing emotional frequencies to the risk rate map. Based on the number of observed happiness Tweets were less than expected.

Overall, the objectives of this thesis were completed and the analysis proved successful. An emotional landscape was created to display the diverse range of emotions across the city landscape. Through a comparison of term frequency and emotional frequency within neighbourhood profiles, it was possible to determine which words were triggering or were associated with each of the eight perceived emotions as determined by the EMOTIVE classification system (Plutchik, 1980). In addition, it was determined that there is a positive link between individuals in NIA and increased risk of negativity. This study will be valuable and considered a contribution to the NIA neighbourhood profiles. It is possible that with a greater understanding of the emotional affect within each neighbourhood, there will be positive changes and integration within the neighbourhood.

The completion of this analysis further exemplifies the difference between acute stress and the development of chronic stress. Through the creation of an acute emotional landscape, it is proven there is a large amount of variability within the city in terms of emotional polarity. While there are variations of polar emotions, the negative emotions appear to form distinct patterns and clustered results. Negative emotional clusters indicate that there are common links between the neighbourhood's affected. There is a need for further research into the combination of negative emotions that most commonly develop into chronic stress. The varying nature of stress within individuals can create challenges when creating specific "acute stress clusters". Varying negative emotional patterns have

proven to be more insightful when creating a more comprehensive analysis of not just where stress is occurring but also in what way stress occurring.

There is no specific formula for the emotional groupings that cause stress. When an individual experiences certain groupings of emotions, it can often be perceived as stress. There are not only combinations of emotions within stress, but also various stressors that generate different emotional combinations. The way in which an individual responds to particular stressors may be different. Some stressors may create more sadness and cause the individual to become upset, while other stressors may manifest itself as anger and frustration. There can be distinguishable patterns within an individual's emotional state when they are inside their home environment versus their work environment.

Traditional data collection focuses on an individual's home profile, whereas geosocial data is focused on in-situ individual level data. When traditional surveys data is collected it is often after an emotional event has occurred, which is not reflective of the individual's emotional reaction to the event. In-situ data collection is crucial when establishing locations that produce acute stress and negative sentiment. Gaining perspective of a location through immediate emotional and sentiment reaction can provide more detail as well as finer data granularity. It is interesting to compare the results of a traditional survey such as the CCHS against the geosocial emotional landscape. The differences are reflective of chronic stress and acute stress as well as in-situ compared to household data collection.

There is a need to modify data collection techniques. Society is changing and with increasing interconnectivity and less physical human interaction, there is a need to adapt data collection to suit the current population. To establish more detailed behavioural data, the collection method as well as methods of analysis should be inclusive of new data types. Although Twitter may not be the solution or a replacement for traditional survey data collection, there is value in re-assessing the need to include in-situ data collection. There is also a need to collect more data at the individual level rather than constantly aggregating up to a standardized level.

Twitter was a sufficient entrance to geosocial sentiment analysis of a city. There are however some limitations of Twitter. The methods of Twitter stream collection are somewhat limiting, there is the option to collect by specific term or event, and there is the option to have the firehose which is meant to be the most open stream. The disadvantage of the firehose method of collection is that there is a limited number of users whose Tweets are geo-located. With a limited number of users with geo-located Tweets, the majority of Tweets come from the same few thousand individuals. There is also no age or gender information received through a Twitter stream. There is the option to search each users profile but that becomes a privacy issue. Twitter users also fall within a particular demographic at the moment. The microblog platform is increasing in popularity however, as commercial Tweets are filtered out there is not an exponential growth in individual users.

This thesis was limited in several aspects. The first was that the sentiment algorithm was developed in the United Kingdom, using a British ontology. The British ontology does

not necessarily match with North American slang and therefore, not all of the North American words were identified. A second aspect that could be improved with the sentiment algorithm is that there were approximately 9 million Tweets collected throughout the entire year of harvest. Of the nine million harvested Tweets, only 7%, or approximately 630,000 Tweets were returned with an emotional score. Although not all Tweets contain emotions, it was estimated that there would be approximately 10% of Tweets returned with emotional scores. Based on these predictions, there were nearly 270,000 Tweets that were not analyzed or not recognized as containing emotional content.

A suggestion for future research would be to include other urban centers to compare and contrast the emotional landscape as well as the differences in neighbourhood profiles. It would be interesting to compare a large Canadian city such as Toronto with a large American city such as New York to see the difference in emotional variation as well as resource interaction.

## 4.2 Research Contributions

The value of new data collection methods was explored through this research. When traditional survey-style data collection is compared with geosocial data, there were spatial differences as well as geographic levels of difference. The geographic levels of difference were attributed to the fact that traditional data is collected at the household level while newer methods of data collection were at the individual level. This research has shown that there is an increasing demand for individual level data, especially health and wellness information. This research has shown there is value in adding individual level information to existing methods of data collection.

There is significance in having increased accessibility to resources however, access does not directly translate into use. Being within close proximity to resources may help decrease negative health outcomes; however, it is the use of resources in positive mediation techniques that will ultimately aid in reducing occurrences of emotional stress in the future. Further research is required to identify more clearly what negative stressors impact individuals the most, as well as where they are located throughout the city. Identifying neighbourhoods with increased access to resources is one part of analysis that should be followed by and compared to the identification of specific stressors.

Currently, there is an expanding realm of user generated content and volunteered geographic information (VGI). Social media such as Flickr and Twitter are large contributors of user generated content that have not yet been fully explored. This thesis has contributed to the expanding field of sentiment analysis within user generated content and geosocial messages. Polarity of sentiment has been previously explored, however, the

use of a sentiment analysis algorithm is fairly new. The benefit of creating a sentiment analysis algorithm is that there can be more emotions extracted from large datasets at a much faster rate. Having a greater range of emotions for analysis creates more dynamic uses of the data.

This study has begun to indicate areas where stress may manifest due to the overlap of multiple negative emotions. There is so much variability within stress and within different individuals, and as a result, it is difficult to pinpoint which events will trigger stress within an individual. It is also difficult to monitor how an individual is using resources to cope or mediate their stress. This study did however, create a good foundation for the continued analysis of certain areas of Toronto where there were documented increased levels of negativity. It will provide a good indication of which resources foster positivity rather than negativity. While the evidence does not necessarily support a concrete relationship between resources and an individual's emotional state, it does support the idea that neighbourhoods can influence an individual's emotional outcome. Follow-up studies should include a greater number of individuals as well as more harvested Tweets.