Exploring the impact of the pedestrian environment on public transportation: A case study of Waterloo Region

by

Stephen Oliver

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Masters of Arts
in
Planning

Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Stephen Oliver

Abstract

In planning communities with balanced transportation options that were once defined by the personal vehicle a comprehensive understanding multi-modal relationships in transportation is required. Public transportation provides a mechanism to move many people through the same space effectively increasing distance and access of a resident. The journey of a public transportation user begins the moment they leave the door on route to the transit stop and only concludes after they disembark the transit vehicle and traveled to their destination. Understanding the influence of the environment between that door and that stop is the objective of this research.

This research is approached through quantitative analysis of the built environment and public transit ridership in the Region of Waterloo, Ontario, Canada. This is achieved through a bus stop level of analysis and linking the built environment within a standard 400 meter radius circular buffer to that stop. The response variable is provided in two forms, average boarding and alighting by stop through a one hour peak time or all day travel.

A literature review informed the selection of intervening and predictor variables. Intervening variables were selected to inform characteristics known to influence public transit use. These variables were *Population density, Employment density, Transit level of service and Transfer location.* Predictor variables were selected to measure different characteristics of the pedestrian environment. These variables included: Land use *Entropy, Sidewalk length, Intersection density, Traffic speed, Traffic signal density* and a *Ratio* of sidewalk length to road length.

Linear regression was used initially to correlate the relationship between public transit ridership and these variables. It was found that several of these variables showed no, or little statistical significance or impact. The best model was the variable combination *Population density, Employment density, Transit level of service, Transfer location, Entropy* and *Ratio*; with the response variable

measuring *All day average boarding and alighting* correlated to an adjusted R^2 of 0.436. It is found that the predictor variables have little impact on the adjusted R^2; however, they are statistically significant in their relationship to ridership.

Spatial regression is then used to further examine this relationship. This is conducted using the built environment variables identified as most influential using linear regression. Here it is found that the intervening variables correlate higher with ridership when a spatial lag model is used. The predictor variables however fail to achieve significance. It is concluded that the pedestrian environment has a low impact on overall public transportation ridership patterns. The pedestrian environment is however significant in informing analysis of the built environment around public transit stops.

This research informs academics quantifying the built environment for both public transit and pedestrian use. The conclusions suggest that sparse pedestrian infrastructure will not define ridership but an increase in the pedestrian environment supports a public transit system. Several variables examined here can inform planners and academics in their methods for conducting similar research supporting multi-modal travel.

Acknowledgements

Dr Clarence Woudsma, my supervisor, whose recommendations throughout all stages of my research was instrumental in both completing this thesis and broadening my academic experience.

My committee members, Dr Jeff Casello and Dr Geoff Lewis, who supported my research instrumentally providing guidance and were always willing to sit down with me when I randomly appeared in their office door.

The school of planning, including all academic and administrative staff, without which my experience would not have been as successful.

My colleagues and friends, those who came before me, those who joined with me and the years who follow, you kept me motivated and entertained as I required. Your willingness to sit down and deconstruct the many challenges proved a strong academic atmosphere can create the ideas that will one day change the world for the better.

My father, mother and siblings who have supported me throughout my life. Their enthusiasm towards professional success and support in my pursuits across all fields of life has and will always be a source of pride.

Finally, I will be forever grateful for the love and support of my soon to be wife, Meghan, who is an endless source of both inspiration and comfort.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Research Question

For a time the urban form was defined by the opportunities presented through cheap energy, a growth in personal wealth and technological innovation. In short, the urban form was defined by the car. Now we know that these communities were built for imperfect transportation solutions. Urban planners, and their colleagues, are daily refitting cities to be adaptable, energy efficient and human in scale. Establishing strong built environment characteristics which support the pedestrian and public transit networks is crucial to this phase of this urban evolution. Understanding what characteristics influence these journeys inform future development so limited public resources can be invested into capital projects which will garner the most mobility options. In this challenge, like many others, planners stand in the face of climate change, obesity and social equity while they move to remold a world once defined by only one mode of transportation.

Since the 1980s there has been a growing interest in how the built environment affects both mode choice (Pushkarev & Zupan, 1977) and health of residents (Villanueva, Giles-Corti, & McCormack, 2008). Current evidence from this research suggests that the built environment has an impact on local level walking behavior. The pedestrian is the affected by route choice, time of day, traffic and other interactions that will vary throughout the journey (Papadimitriou, Yannis, & Golias, 2009). The linkages between sidewalks and route choice, as well as the variables relating to density and land use mix, have been tied to pedestrians activities by previous researchers (Lee & Vernez Moudon, 2004).

Understanding the pedestrian environment in isolation however is limited in its help for users of modern cities nor the guide the construction of new developments that are sustainable and livable. We can all individually observe that with the exception of a few neighbourhoods, it is not possible to work, play and shop within walking distance of home. This speaks to the importance of combined pedestrian environment and public transportation as a mechanism for moving people between these key destinations within our urban centres (L. D. Frank & McKay, 2010). From his research Guo (2010) concludes that with a better understanding of walking behaviour to support public transportation, the built environment can be redeveloped to increase the distance people are willing to walk to take transit. Embracing the multi-modality of public transportation centers around the requirement to accommodate both modes of transportation.

This research presents the opportunity to examine the pedestrian portion of the public transportation journey. A public space designed for pedestrians is fundamentally different from one designed for other modes of transportation. In other modes of transportation the land use patterns only matter at the origin and the destination as the journey is only briefly experienced locally, this is not the case for pedestrians (Guo & Ferreira Jr, 2008). Currently in considering how to develop and intensify in the North American context there is a bias towards that destination oriented design (Wey & Chiu, 2013). This emphasis on the built environment at the point of origin, home or work, and at the public transit stop, leaves the pedestrians to fend for themselves in between.

Studies have shown that the pedestrian environment between the origin and destination is correlated with walks to and from a transit stop (Rodríguez & Joo, 2004). The understanding of this effect is however limited. Previous research has been force to omit the sidewalks, traffic signals, due to lack of information (L. D. Frank, Greenwald, Winkelman, Chapman, & Kavage, 2010; Wasfi, Ross, & El-Geneidy, 2013); furthermore, researchers have specifically identified that these variables need to be included in addressing these questions (Saelens, Sallis, & Frank, 2003).

The objective of this research is to explore the following question

- How does the pedestrian environment / walkability affect public transit ridership?

And in that exploration inform these following three sub-questions

- What is the most appropriate way to measure pedestrian infrastructure, as it relates to walkability, and what is its correlation with transit ridership?
- What walkability / built environment characteristics correlate best with transit ridership?
- In what way is answering this question informed through the use of linear regression and spatial regression models?

For the purposes of this research the built environment reflects all aspects of the urban form including buildings, roads, traffic and natural features. The pedestrian environment is those components shown to influence the pedestrian experience, pedestrian infrastructure being features which are built supporting that use, like sidewalks and trails. Walkable and walkability are used to express components of the built environment which relate to the pedestrian environment, this term is used with few exceptions, interchangeable with pedestrian environment. The public transportation being examined is exclusively run for local service by municipal or regional government and the term ridership is used to express the people who board and alight, get on and off, of the system. The roots of these definitions are founded in previous academic research further examined in Chapter 2.

This study provides an opportunity to explore the multi-modal element of public transit ridership as it explores the pedestrian and transit relationship within the built environment. This research aims to capitalize on quality of information available and clarify the relationship through pedestrian infrastructure. The intent of answering these questions is to develop further understanding which can inform future research for academics and professional planners alike.

1.2 Overview

Following this introduction, Chapter 2 - Literature Review, will establish the academic foundations for this research. The will be done through an examination of three relevant planning paradigms: New

Urbanism, Transit Oriented Development and Active Transportation. Subsequently an exploration of previous research surrounding the built environment's effect on the pedestrian environment and transportation ridership will follow. This chapter will inform discussion of potential variables and previous research conclusions.

Chapter 3 - Methods, provides the overview of the research design. This explains the study area, the Region of Waterloo, along with the details of the public transportation services, Grand River Transit, being studied. Each variable examined for this study is then identified and the collection/collation of that data explained. Some of the variables require statistical transformation and this will be articulated here. The use of both linear regression and spatial regression will be explained and methods explored.

The first regression results are presented in Chapter 4 – Results and discussion: Linear regression. Here the Base Model between the response variable and intervening variables, variables which are known to have a correlation, is established. Once the Base Model is established individual variables will be explored to assess their correlation and impact on transit ridership. The objective is to identify variables which indicate walkability and determine the best series of variables which will then be used in the spatial regression modeling.

The Base Model and walkability variables which reveal high linear correlation and statistical significance will be further explored in Chapter 5 - Results and Discussion: Spatial Regression. This section will aim to explore spatial auto correlation and its effect on this research question. The results will provide the foundation for a better understanding of both the variables and different methods to be used when examining the built environments impact on public transportation ridership.

Chapter 6 – Recommendations and Conclusion will provide a summary discussion about the research question and concluding thoughts on this research topic. This will include a discussion about limitations which influence the variability and directions for future research. Finally that chapter will re-visit the findings of this research and expand on the planning and the links within the topic.

Chapter 2
Literature Review

2.1 Introduction

The objective of this research is to answer the following question and sub-questions:

- How does the pedestrian environment / walkability correlate with public transit ridership?
    - What is the most appropriate way to measure pedestrian design, as it relates to walkability, and what is its correlation with transit ridership?
    - What walkability / built environment characteristics correlate best with transit ridership?
    - In what way is answering this question informed through the use of linear regression and spatial regression models?

The importance of understanding the relationship between the built environment and public transit ridership is a part of a growing field of literature. This chapter explores some of that scholarly material to create a foundational understanding of previous research and inform this study. This review will be broken into four sections: Planning theory, the built environment and public transit ridership, pedestrian infrastructure, and transit level-of-service.

2.2 Pedestrian and public transit in planning theory

Planning as a practice has developed many theories supporting and advising urban form with the intent of creating better spaces for people to live. This is highly evident in two of the planning theories presented in this section; New Urbanism, which pursues a high quality of life through creation of connected walkable communities (Congress of the New Urbanism, 2000), and Transit

Oriented Development (TOD), a term developed by the New Urbanist Peter Calthorpe (1993). This section will also explore the theories of active transportation, as both pedestrian activity and public transportation have a role to play in creating healthier communities.

2.2.1 New Urbanism

The principles of New Urbanism, along with the developments and designs influenced by them, state that the increase of public transit ridership and walkability is a fundamental element of creating a better community (CMHC, 2013; Gallagher, 2012; Handy, Boarnet, Ewing, & Killingsworth, 2002). A central tenant of New Urbanism is access to destinations through combinations of walking and a general reduction of car reliance as a priority in development (Congress of the New Urbanism, 2000; Newman & Kenworthy, 2006). Consequently the design contributions are inspired by the priorities of access and accessibility which are fundamentally shaped by the desire to reduce auto dependence.

Through New Urbanism, the objective goals of greater quality of life through urban design take form in the urban landscape. These goals have been integrated into the standard of planning for many metropolitan areas, shaping the development of new communities and renewal projects (Krizek, 2006). The promotion of walking in New Urbanist communities has its measurable benefits to resident behaviour. For example, a CMHC study (2013), showed that 51% of residents in the New Urbanist designed neighbourhood walk for local goods. This rate of walking is supported by pedestrian infrastructure, 24.7% more sidewalk coverage in New Urbanist communities versus Conventional suburbs. Favourable numbers in walking however are not extended to public transit; in both the conventional and New Urbanism community ridership levels were exactly the same at 9% (CMHC, 2013).

New Urbanist communities have been able to increase transit activity at the local scales, yet they are not inherently transit friendly, a point compounded by the fact that they are often not geographically situated to influence broader transportation decisions of the residents (L. D. Frank et al., 2010; Gallagher, 2012). Therfore urban design influenced by New Urbanism is defined in part by its new approach to accessibility and especially walkability. Public transit requires broader design changes than can be offered on the neighbourhood level, notably into regional and system design. The importance of New Urbanism in understanding this research is related specifically to pedestrian behaviour especially the prioritized presence of pedestrian infrastructure in community design.

2.2.2 Transit Oriented Development

TOD considers the role of the pedestrian as the keystone of a community where the role of transportation reflects the re-investment in the built form away from car oriented design (Calthorpe, 1993; Hess, 2011; Hester, 2010). A central principle of TOD is the design of communities to efficiently and effectively support public transportation, asserting this can increase ridership and subsequently the quality of life of residence (Chow, Zhao, Liu, Li, & Ubaka, 2000). The importance of integrating the pedestrian network and public transit is widely accepted in considering the goals of TOD. Through travel mode interconnections the objective of a TOD design is a decrease in overall auto vehicle travel and an increase in public transit, bicycling and walking (Cervero & Kockelman, 1997; L. D. Frank et al., 2010; Wey & Chiu, 2013).

TOD has been a staple in the redevelopment of downtowns and other intensified neighbourhood areas which support a variety of activities and amenities with naturally attractive and safe pedestrian environments (G. Thompson, Brown, & Bhattacharya, 2012). The interconnected aspects of the TOD design increases the ability of the user to access the area around the transit developments, sometimes referred to as permeability (Ratner & Goetz, 2013). As TOD design principles are imported to less intensified neighbourhoods, the challenge becomes understanding the interconnected dynamics of the pedestrian and public transit networks and using the appropriate tools to accurately support this permeability.

TOD as a concept has been challenged by its ability to quantitatively measure the impacts of the pedestrian environment and walkability factors (Ha, Joo, & Jun, 2011). That study explores this research area through quantifying experiential pedestrian qualities around subway stations. Ha et al. (2011) integrate their research with only limited built environment characteristics, as they focus on the qualitative pedestrian experience throughout the journey, between origin and destination.

The role of the pedestrian in developing a TOD urban form requires an understanding of the variables which affect that travel mode. While the principles of these designs have been extended to support bus activity, rail transit is often deemed more influential with higher ridership levels and customer satisfaction (G. Thompson et al., 2012). The core challenge is understanding no matter the type of public transit being used, the rider must be able to comfortably move as a pedestrian to and from the transit stop (Clarke, 2003).

For TOD to be successful in redeveloping a car oriented urban form they need to be designed to include high density, vibrant land use mix and urban design which facilitates pedestrian access (Calthorpe, 1993). Understanding quantitatively which of these variables has an impact on public transit ridership prioritizes investment in current and future developments. If it is shown the pedestrian infrastructure has little bearing on public transit ridership levels, the network interconnections required to facilitate TOD can reprioritise the use of resources towards elements of the built environment with higher impact.

2.2.3 Active Transportation

Like New Urbanism and TOD the study of Active Transportation is a planning paradigm of growing significance both in research and policy. Active Transportation is generally considered any mode of transportation which promotes physical activity, most commonly cycling or walking. The connection to the research question being asked is drawn from the understanding that every public transit user is by default a pedestrian. The riders have travelled from their point of origin to the transit stop and from their last stop to their destination (Hess, 2011; Mees, 2010).

Active Transportation is related to increasing public health in the face of unhealthy diets, sedentary life styles and climate change. Low levels of activity as a result of commuting behaviours have been directly linked with chronic disease and a mounting price tag in the social and health costs (L. D. Frank et al., 2010; Sallis, Frank, Saelens, & Kraft, 2004; Wasfi et al., 2013). Combining the cost to public health with the understanding from a study by Maibach, Steg & Anable (2009) that nearly half of trips by car are well within active transportation distances, less than 8 km, it can be observed that this public health epidemic is primed to be addressed. It has been shown that time spent walking for transit exceeds time spent on walking only trips (Agrawal & Schimek, 2007).

Combining public health and transportation research has moved the objectives for both fields of policy and study closer to harmony and subsequently success. One such example of this harmony is presented by Lee and Vernez Moudon (2004),where they found combining utilitarian activity and recreation activity, reduces the limitation of available time for each activity, increasing overall physical activity. Their research shows that promoting commuting, as the utilitarian activity, with walking, as a recreational activity has direct linkages with health benefits.

Studies have shown that public transit users get over 20% of their recommended daily activity through their journey to and from the transit stop (Morency, Trépanier, & Demers, 2011). That same study additionally found that the pedestrian environment promotes healthy living and reduces rates of obesity. Another study by Wey and Chiu (2013) showed that appropriate environmental design can increase pedestrian activity both in duration and frequency.

These findings reflects the importance of developing an environment that promotes pedestrian activity in the population. The users’ selection of travel mode presents an opportunity to change the transportation and public health dynamic (Clarke, 2003; Sallis et al., 2004). Provincial policy documents identify the role and implementation of pedestrian networks to be a crucial part of active transportation networks (Ontario Ministry of Transportation, 2012). As has been shown here there is a strong benefit to linking the active transportation network with public transportation. The studies of Active Transportation, New Urbanism and TOD all benefit from understanding the effect of pedestrian environment on public transit ridership as it effects pedestrian traffic and supports a healthier population.

2.3 The effect of the built environment on public transit ridership

The term the built environment encapsulates the non-natural physical features within an area of interest. Previous studies have indicated a relationship between the built environment and public transit ridership; in particular, these have studied higher densities and mixes of land use (Cervero, 2002; Ewing & Cervero, 2010; Guo, 2010; Pushkarev & Zupan, 1977). These and other similar studies support the design principles of New Urbanism and Transit Oriented Developments and Active Transportation as discussed in the previous section.

Some researchers have questioned the importance of changing the built environment as a method of changing attitudes and actions. Krizek (2006) found that only small portion of populations change their lifestyle based on built environmental factors. This, and other challenges, are heavily hinged on a part of an academic discussion commonly referred to as self-selection. Guo (2010, p. 4) explains one of the key complexities as this concept as: “a neighborhood that is more favorable to pedestrian activities might be more likely to request an improvement, and thus more likely to get it.” Approaching the residence self-selection towards their built environment and transit mode choice is an expanding academic research area (Lachapelle & Noland, 2012; Owen et al., 2007). For further discussion on self-selection the built environment and pedestrian activity see Cao, Handy, & Mokhtarian (2006).

The built environment has been used to examine its effect both on pedestrian choice behaviour and transit ridership. For example, path-choice modelling has previously explored characteristics of the pedestrian environment such as land-use, sidewalk convenience, sidewalk continuity, open space and topography, to relate to transit activity (Guo & Ferreira Jr, 2008). Their particular study precedes the conclusions from another study which suggests that the pedestrian environment does shape the utility of walking, and can be linked to the distance that people are willing to walk to a transit stop (Guo, 2010).

Further studies exploring the built environment and pedestrian behaviour have used elasticities, which indicate how much the response variable will shift when there is a small shift in an explanatory variable (G. Thompson et al., 2012). In a study by Ewing and Cervero (2010) between transit ridership and the built environment using: land use mix, population density, intersection density, destination accessibility and distance to transit, the relationship what found to be inelastic., This conclusion challenges policy and academic conventions and establishes an argument for the independence of each neighbourhood in affecting conclusions on this subject.

Measuring the pedestrian environment alone does not necessarily increase public transportation ridership. The argument has been made that high quality public transit actually competes with walking, and mediocre public transit will promote walking (Mees, 2010). Saelens et al. (2003) found a point where the proximity of the destination is supported by the tightly packed environment resulted in increased walking and cycling trips over public transportation. This concepts underlines the importance in understanding the relationship between public transit travel and pedestrian travel.

A categorical structure has been developed for addressing the highly variable and complex issues of transportation and built environment planning, Cervero and Kockelman (1997) defined the 3D's 3D categories: Density, Diversity and Design for use when analysing the effect of the built environment on travel behaviour. These three categories permit an academic study to examine specific comparable characteristics without becoming overwhelmed by the complexity of these built environment problems. To this end, the 3Ds appear prominently in this literature review and other chapters as a tool to answer the posed research question.

It is noted here that previous studies have shown a positive and statistically significant relationship between the combined 3Ds variables and public transit ridership, although this has been relatively marginal in scale (G. Thompson et al., 2012). The research question being asked in this research does not use the 3Ds exclusively to study the relationship, rather to categorise and better understand the variables of previous research. Additionally it is understood that the 3Ds are most valuable when examining the built environment at higher spatial levels, such as census tracts (Werner, Brown, & Gallimore, 2010). To better answer the pedestrian environment and public transit ridership we aim to examine individual variables of each of the 3Ds.

2.3.1 Density

The role of density is broadly understood as vital in a highly effective transit system. Newman and Kenworthy (2006) showed that over 90% of public transit ridership in the Los Angeles region is explained by this variable. Density is understood as: "sufficient human population to support a vibrant and economically viable community and is often indexed by census measures of population per unit of area" (Werner et al., 2010, p. 207). Density is also expanded beyond a population count and additionally reflects the level or intensity of activity, both employment and recreation within an area (Cervero, 2002).

Density is one measure used to establish the available population to access public transit systems. When the buffer area, explained in Section 3.2.3, is packed as tightly as possible with potential users, this increases those capable of accessing the service within close proximity. In order to effectively support a public transit system, density thresholds are linked with and increase walking and cycling trips and concurrently a reduction in automotive trips (Newman & Kenworthy, 2006). Guerra and Cervero (2010) noted, in their study of public transit cost, that fixed density benchmarks for transit programs were unreliable across different jurisdictions as each project maintains individual characteristic. The benchmarks were however all correlated.

In examining the built environment, the density around public transit location has a thoroughly examined effect on the ridership within that transit system. Table 2.1 provides a list of several different studies which have employed a measurement of density. Of note, this table clearly shows that that most studies use two variables to examine density, both employment density and population density. While this is not a comprehensive list it can be observed which types of density are frequently relied upon to operationalize this element of the built environment.

Table 2.1 Density variables used in previous built environment research

Variable	Explanation	Articles Used
Population Density	Number of people in a given area	Agrawal & Schimek, 2007; Besser & Dannenberg, 2010; Cervero & Kockelman, 1997; Chow et al., 2000; Delmelle, Haslauer, & Prinz, 2013; Duong & Casello, 2010; Edwards, 2008; Ewing & Cervero, 2010; Forsyth, Michael Oakes, Lee, & Schmitz, 2009; L. D. Frank et al., 2010; Guerra & Cervero, 2010; Handy et al., 2002; Hess, 2011; Hirsch, Moore, Evenson, Rodriguez, & Diez Roux, 2013; Krizek, 2006; McDonald &

		Trowbridge, 2009; Newman & Kenworthy, 2006; Quintero, Sayed, & Wahba, 2013; Rodríguez, Khattak, & Evenson, 2006; Samimi, Mohammadian, & Madanizadeh, 2009; Su, 2011; G. Thompson et al., 2012; Wasfi et al., 2013
Employment Density	Number of jobs within a given area	Cervero & Kockelman, 1997; Chow et al., 2000; Ewing & Cervero, 2010; Guerra & Cervero, 2010; Hirsch et al., 2013; Quintero et al., 2013; Saelens et al., 2003; Thompson et al., 2012a
Retail Density	Number of Retail establishments within a given area - This is a method of calculating potential destinations	Handy et al., 2002; Wasfi et al., 2013
Residential Density	Dwelling units within a given area, or population per area of residential land	CMHC, 2013; Lawrence D Frank, 2004; Hirsch et al., 2013; Oliver, Schuurman, & Hall, 2007; Owen, Humpel, Leslie, Bauman, & Sallis, 2004; Ryan & Frank, 2009; Saelens et al., 2003; Yang, Diez Roux, Auchincloss, Rodriguez, & Brown, 2011
Street/Lane/Path Density	Measure of Streets within a given area (ie. Miles/acre) - This variable is often measured differently and listed as a design variable discussed later	CMHC, 2013; Rodríguez et al., 2006; Samimi et al., 2009
Pedestrian Density	Road capacity evaluation using pedestrian environment indexes	Ha et al., 2011
Total Activity Density	Total population and employment divided by area	Cervero, 2002

The complexity of analysing the density variable was directly exposed in Dellmelle et al. (2013), which determined that social satisfaction was reduced in high density housing, apartments versus single family dwellings. The social satisfaction increased however in higher density neighbourhoods, postulating a link to chance social contact. That study was unable to provide a complete explanation about the nature of this relationship, showing that variables such as commuting travel times and affluence were obscured variables within the study.

Ewing and Cervero (2010) established that employment density had a much lower predictive capability on public transit ridership than population or residential density. This may be connected to the conclusions of Thompson et al. (2012) that the most high density employment areas generally have jobs based in legal, finance or other office employment, whose workers are not as transit-dependent. This conclusion may also represent the specifics of that study as other high density employment areas may not observe the same relationship.

The effect of density on transportation behaviour presents several complex conclusions that can confound transit analysis. Krizek (2006) found that high density populations show greater use of alternative forms of transportation. This conclusion was not necessarily synonymous with reduced auto dependence, as much as with an increase in available discretionary time to engage in those activities. Agrawal and Schimek (2007) revealed that the effect of low and high density built environment on recreational walking trips only showed up in the most extreme neighbourhood designs. This same study determined that walking for non-recreational, or utility purposes increased within each density category, and was strongest within the highest density category. Levinson (1998) articulated that transit trips are shorter than automotive trips in higher density areas, because transit can have a higher level of service, discussed in section 2.5. That concludes that where there are more transit riders, there is potential for higher levels or service while automotive trips must negotiate congestion where there are more vehicles.

The link between density and both walking behaviour and transit behaviour continues to be a contextual and complex analysis. Vital however to this thesis is that a relationship is consistently observed, especially in population and transit ridership. The inclusion of a measurement for this variable in this research is essential to establishing valid understanding of the built environment.

2.3.2 Diversity

Much like density, diversity is a variable of the built environment which is equally challenging to understand. Diversity refers to a measurement of land uses, amenities and opportunities which permits the broadest range of destination types within an area (Werner et al., 2010). There are more destination types in an area, therefore, more people are brought into the area for different reasons. Diversity combines two elements supporting public transportation: first it creates good destinations for trips, suggesting there is more to do at the end. Second, which most supports walkability, more destinations within proximity create neighbourhoods which are more likely to have amenities at various distances. Higher diversity serves the population as a whole more comprehensively and requiring less transportation to traverse.

Several studies by Cervero have provided different ways to operationalize the variable of diversity (Cervero & Kockelman, 1997; Cervero, 2002; Ewing & Cervero, 2010). In Cervero and Kockelman (1997) diversity was operationalized through the development of a dissimilarity index. This dissimilarity index calculates the proportion of dissimilar land uses as neighbouring land use parcels. Those studies found a strong relationship between travel behaviour and the built environment, especially walkability factors.

The importance of isolating the effect of design on walking as a form of transportation is a highlight of some academic research. For example the job-housing balance is a stronger measure of walking behaviour than land use mixture (Ewing & Cervero, 2010). This thesis measures pedestrian infrastructure specifically, which was not included in their examination of public transit ridership. The walking studies link with the earlier conclusion by Mees (2010) that a highly walkable environment will draw away users from public transit. As users’ origin and destination are both within walking range it removes the requirement for regular transit use.

Entropy is another measure of diversity used to examine the influence of the built environment. Entropy, a term which denotes a state of disorganisation, measures mixtures of activity across an activity type. High entropy indicats a more chaotic or random state and low entropy indicats a more uniform state. Previous research has measured mixtures of family types, with children and without (G. Thompson et al., 2012). In the case of the built environment however entropy is generally measured through land-use types (Cervero & Kockelman, 1997; Cervero, 2002; Ewing & Cervero, 2010; Ryan & Frank, 2009; G. Thompson et al., 2012). This measurement generally establishes a score between 0 and 1, with a score closer to 0 indicating a higher level of sameness. Previous studies have used the variable largely as a piece of an index, as in Ryan and Frank (2009), or in testing elasticity meant to inform future studies (Cervero, 2002). In the case of this research this will be an predictor variable, the equation and methods are detailed in section 3.3.2.5.

Overall, research has shown that diversity is a key measure of built form influence for both walkability and public transportation analysis. Walkability indices and pedestrian environment measurements frequently single out its importance. Saelens et al. (2003) concluded that the diversity of land use appeared related to commuting through walking and cycling, as it creates more residential destinations within proximity. The objective of this research is to determine the importance of these walkability variables in affecting public transportation ridership.

2.3.3 Design

The third D, design, is more difficult to operationalize than then either density or diversity. In this research the characteristic design refers to specifically examining the influence of design on the pedestrian activity, where the design of the environment is experienced at the individual level (Werner et al., 2010). The operationalization of design has resulted in mixed conclusions. This may be related to the fact that it is more subjective than the other two factors. Cervero (2002) showed that design has the lesser influence of the 3D's a point contrasted by Werner et al. (2010) which found twice as much walking in high walkable neighbourhoods.

A method of operationalization often used for design characteristics is intersection density, a variable of community design. Guo and Ferreira Jr (2008) explain that while intersections may discourage street crossings, due to traffic interaction and traffic lights, a dense street intersection design at the neighbourhood scale is indicative of denser development and associated with a more accessible pedestrian environment. Expanding on the range of operationalized variables Ewing and Cervero offer the following:

> "Design includes street network characteristics within an area. Street networks vary from dense urban grids of highly interconnected, straight streets to sparse suburban net- works of curving streets forming loops and lollipops. Measures include average block size, proportion of four- way intersections, and number of intersections per square mile. Design is also occasionally measured as sidewalk coverage (share of block faces with sidewalks); average building setbacks; average street widths; or numbers of pedestrian crossings, street trees, or other physical variables that differentiate pedestrian-oriented environments from auto-oriented ones." (Ewing & Cervero, 2010, p. 267)

Intersection density is an easily operationalized variable, used by Walk Score and indirectly by Transit Score, which measures distance to a transit stop from an address (Hirsch et al., 2013). The variable reflects the concepts of permeability and connectivity for pedestrian access. These terms, both reflect similar concepts: the ability of the individual to access more of the built environment in an area through pedestrian infrastructure. Many of the specific design elements are examined in the walkability section of this chapter.

"Pedestrian Friendly Parcels", used by Guo and Ferreira Jr (2008), is another example of a design variable which is also associated with the diversity characteristics. This variable measures land use types which are more favourable to pedestrian access, such as retail, mixed use and commercial versus residential, industrial and office. Their supposition is related to the fact that the more a business relies on pedestrian traffic, the more pedestrian friendly the building is. The idea however may be challenged based on the nature of that business or the business model, as their research was conducted in a downtown with a lack of big box stores or other business types.

The use of the 3Ds concept, while not exclusive in its approach to operationalizing the built environment to analyze urban transport issues, does break the complexity into manageable elements. Through establishing and examining how other academics have operationalized the built environment using density, diversity and design, this research ensures that each element of the built environment is accounted for in exploring the research question.

2.3.4 Walkability

The term "Walkability" is used to articulate the entirety of the built environment as it impacts the pedestrian experience - areas with higher walkability having more characteristics which support pedestrian activity. Walkability is typically expressed through combinations of variables and indicators measured in study of the pedestrian experience (Millward, Spinney, & Scott, 2013), this combination of variables is generally referred to as a walkability index. This section aims to identify different components of a walkability index from academic studies which are positively associated with high walkability.

A walkability index can be a combination of built form variables already individually considered within the previous discussion of the 3D's characteristics. The base for the walkability index used by Owen et al. (2007), in establishing the impact of walkable neighbourhoods on Australian adults, was measures of density operationalized through: street connectivity, land-use mix and retail area. These core variables were isolated into the 3D's and each D shown as an individual impact on pedestrian behaviour.

These can be broad in design including characteristics like aesthetics which can be highly subjective, often approached in qualitative studies which examine walking behaviour. Examinations of characteristics relating to litter, abandoned buildings and construction were included in the study by Ha et al. (2011), which compared walkability at subway stations. The perception of aesthetic quality, and perceived convenience of environmental facilities both revealed a strong positive association with walking behaviour (Ball, Bauman, Leslie, & Owen, 2001). These subjective elements are also more apt to influence differently neighbourhood to neighbourhood. Agrawal and Schimek (2007) describe the presence of these variables is more important for suburban recreational walkers than urban utilitarian walkers.

Table 2.2: Types of variables used in built environment and transportation research from Maghelal & Capp, 2011

Variable Type	Definition	Method of measurement	Examples
Objective	Variables that can be quantified, standardised and repeated in other studies	GIS or Audit	Intersection, Land-use mix
Subjective	Variables that can be quantified and standardised, but may	Survey	Perception, Architecture

	or may not be replicated		
Distinctive	Variables that can be quantified using a method of measurement that may not be replicated	Observation	Cautious driving

A 2011 audit of walkability indices by Maghelal and Capp (2011) divided the utilized variables into three categories defined in Table 2.2. This audit suggests walkability is operationalised differently by various authors. This non-standardised approach means that walkability analysis and its conclusions are not always transferable to other study locations. This is well cited within the academic literature, as data sources and scope outline limitations to the provided indices.

Two examples of the operationalization of walkability and its consequences in studying multi-modal transportation are now discussed. Thompson et al. (G. Thompson et al., 2012) concluded that walkability at the place of origin had no statistical effect on public transit ridership, however, the same was not true for walkability at destination. Their index used only a measurement of sidewalk length and crosswalk presence to establish a walkability score. Other variables like land-use mix, population density and employment density, were separated from the walkability index and analysed independently which has not been done in other studies. Similarly the walkability established by Hess (2011), using only sidewalk presence and intersection density, was statistically significant to public transit ridership in the case of riders over the age of 60. These differences illustrate a key challenge of the “walkability index”, which is non-standardised and therefore affects the ability to accurately compare academic conclusions.

Beyond academic research, walkability is generally presented in the popular internet tool Walk Score. Walk Score uses an algorithm which weighs proximity to key amenities, and adjusts the score based on the street network and built environment characteristics such as density and block length

(Hirsch et al., 2013). This product is considered a viable and relevant predictor of walking behaviour according to Hirsh et al. (2013) which relies on open source information to establish its metrics. The creators of Walk Score have also developed a Transit Score metric, and have effectively created consumer scale metrics to understand and apply the walkability concept.

Ryan and Frank (2009) developed a walkability index to measure directly against public transit ridership in San Diego. Walkability in this case was measured through land use mix, residential density, retail floor area ratio and intersection density. This walkability index explained approximately 0.5 percent of the variation in transit ridership across neighbourhoods. It was suggested in their conclusions that the use of land use data measured in acres versus a measurement in square feet had depressed this significance. As square footage more accurately reflects the height of the built environment.

The lack of a standard measurement relating to walkability affects the ability to use this metric in answering this research question. The issues of comparability and variability further justify the position that operationalization of the pedestrian environment through the 3D's and separating walkability variables for more detailed study present as a clear direction forward. While the term walkability still applies, it does not denote a specific variable set rather than an overall environmental concept.

2.4 Pedestrian infrastructure and pedestrian safety

Infrastructure has been developed which aims to increase the convenience of the pedestrian and increase safety. This section aims to provide an overview of some research related to this area. It needs to be acknowledged that this is a large area of research which has been expanding since the 1970's (Fruin, 1971). The information presented here is synopsis of some of the many topics and materials that relates to pedestrian safety and pedestrian infrastructure. Specifically this section will examine sidewalks and traffic calming infrastructure.

2.4.1 Sidewalks

When considering the pedestrian environment, especially as a user, one should start by asking if there is a stronger symbol of pedestrian priority than a sidewalk. In the same way as a road is clearly designed to move cars, a sidewalk is meant to move pedestrians. The presence of a sidewalk is a stated influence on walking behaviour which crosses income and gender lines (Brownson, Baker, Housemann, Brennan, & Bacak, 2001). Subsequently, when examining pedestrians and walking behaviour as it relates to public transportation ridership, the inclusion of a measurement for sidewalk must be considered.

The importance of a sidewalk in encouraging pedestrian activity is more than just psychological. Collisions between pedestrians and vehicles are more than twice as likely in areas without sidewalks (Campbell, Zegeer, Huang, & Cynecki, 2004; Retting, Ferguson, & McCartt, 2003). The role of a sidewalk in creating a safe trip for the pedestrians to travel without negative interaction with vehicles provides a basis for understanding the difference between examining the pedestrian infrastructure using the appropriate measurement, sidewalks not roads.

Owen et al. (2004) audited studies examining the built environments' effect on pedestrian activity. They found that the majority of studies, 37 versus 25, showed a statistically significant correlation between walking behaviour and presence of a sidewalk. High quality pedestrian facilities, which include sidewalks, have been shown to increase pedestrian activity even when land use and other built environment characteristics remain constant (Saelens et al., 2003). Further research by Cervero and Kockelman (1997) concluded that a sidewalk, among other pedestrian built form elements, positively related to promoting trips which did not rely on personal vehicles.

Often studies examine only the presence of a sidewalk and its influence on behaviour, not the satisfaction or perceived quality of that sidewalk (Wang, Li, Wang, & Namgung, 2012; Werner et al., 2010). This omission is related to the subjectivity of walkability, discussed in section 2.3.4, as perception develops a more subjective measurement of the utility. Using qualitative assessments and establishing level-of-service metrics, previous studies have been able to establish correlation between utility and pedestrian perception, a relationship which has not be established for crosswalks (Papadimitriou et al., 2009). This supports to the utilitarian purpose of the sidewalk primarily and can be used to justify level-of-service analysis which eliminates subjective elements.

Previous studies have also attempted to identify how sidewalks affect public transit. This pursuit has been challenging because it is difficult to get precise data for pedestrian activity. This difficulty is primarily due to the aforementioned complexity in measuring the wide range of influences involved in the pedestrian environment (Clarke, 2003; Guo & Ferreira Jr, 2008). Further complicating the situation, data collection of pedestrian activity frequently omits trips less than one kilometer (Mees, 2010). Despite these limitations Rodríguez and Joo (2004) determined that sidewalk continuity influenced mode choice both for accessing transit and in full journey. The inclusion of a sidewalk measurement specifically benefits this research by providing an objective variable within pedestrian access areas to bus stops.

Suburban planners have in the past omitted the pedestrian aspect of bus transportation. This is not a new discovery as Gassaway (1992) explores in his suburban study, where an intersection with over 300 bus riders each day and heavy vehicle traffic was built with no sidewalks to support their transportation. The objective to create environments that support modes of transportation other than cars, such as public transit, reveals the need for an integrated approach to infrastructure. Consider the significance putting a sidewalk in may have in reinforcing users of public transit.

2.4.2 Traffic Calming

The pedestrian experience, while linked to the built environment characteristics explored thus far, is also fundamentally linked to the interaction between the pedestrian and the car. A traffic calming feature is defined by Ha et al. as "features that reduce the negative impact of motor vehicles, therefore enhancing walking and bicycling conditions by slowing the hazards or providing pedestrian sanctuary" (2011, p. 141). Just as sidewalks have been established as a fundamental part of pedestrian infrastructure, traffic calming measures address another piece of the pedestrian experience, the shared corridor where sidewalks and roads are adjacent

In a study by Werner et al. (2010) the perceived safety from traffic during ingress and egress was an important element for regular riders of public transportation. Papadimitriou et al. (2009) argues that current pedestrian modelling using mostly crowd modeling methods, does not reflect pedestrian choice. Instead these models are traffic-oriented and pedestrians respond to traffic conditions by changing routes, times and crossing locations. Traffic calming measures typically reduce speed or volume of vehicles and shorten the road crossing distances thus benefiting the pedestrian (Campbell et al., 2004). It is worth noting that traffic lights are not consistently included in the studies of traffic calming, in the case of this research they are.

2.4.2.1 Traffic Lights

While intersection density measures the number of access points for pedestrians, traffic lights provide a method for the pedestrian to cross safely from vehicle traffic. A review of studies relating to traffic measures to reduce pedestrian-motor vehicle collisions conducted by Retting et al. (2003) found that every type of traffic light had the effect of reducing collisions. While the relationship with the pedestrian is not absolute, in many cases the results are considered promising and the study concluded more evaluation is needed.

In contrast with the benefit to permeability and access, discussed earlier, high intersection density in an area may not create a favourable environment for pedestrians. Each road crossing creates a potential conflict area between cars and pedestrians, the use of traffic lights regulates the activity of that traffic (Guo & Ferreira Jr, 2008). The study by Werner et al. (2010) included traffic signals in pedestrian safety survey conducted supporting ingress and egress to public transportation. In that study a marginally significant result was found on pedestrian perceptions. Signal lights have previously been presented in studies as a barrier for the pedestrian as the signalization can slow a pedestrians route and increase overall walking time (Hess, 2011). This research reveals the importance of this built environment characteristic in walkability which fundamentally links with larger research question being asked as to its influence on public transit ridership.

2.4.2.2 Traffic Speed

The reduction of traffic speed gives a driver more time to react to pedestrian activity. This principle also has the added benefit of making an environment more pleasant for the pedestrian. Wey and Chiu (2013) draw a linkage between enhancing pedestrian access to transit through reducing both automobile use and automobile traffic speed. The requirement for traffic control and traffic safety can be perceptually linked between the pedestrian and the speed of traffic Werner et al. (2010), this suggests that the need for traffic calming is less in lower speed environments. A conclusion which indicates that where speed is reduced the risk to the pedestrian is reduced. This safety is true in the case of both posted speeds and actual speed, which can be different depending on congestion and driver behaviour.

2.5 Transit level-of-service

Research about public transit would be remiss not to identify the huge amount of academic material and planning that already exists in examining variables that affect ridership. Factors such as socioeconomic characteristics of the community, unemployment levels and the cost of gas have strong system wide effects on transit ridership and transit studies (Tang & Thakuriah, 2012). Car ownership historically has been a determining factor in mode choice, specifically in public transit use (Levinson, 1998; Santoso, Yajima, Sakamoto, & Kubota, 2012). Transit Score considers two main variables in assessing a location: frequency and type of route (Hirsch et al., 2013). While not each known variable will be included in answering this research question, their impact is acknowledged as prominent in determining public transportation ridership.

This is especially true when considering the role that public transportation has in promoting ridership through level-of-service. Level-of-service, assessed through the mean waiting time at a stop (Delmelle et al., 2013), and transfer times have both shown significant relationships in explaining transit ridership (G. Thompson et al., 2012). Wasfi et al. (2013) showed that the longer the wait time for a bus, referred to as headway or frequency, the distance willing to walk to catch that bus shortens. Overall the frequency, routes and transfers are all decided in the planning stage of developing the transit systemFor a detailed review of material on this topic reference Guihaire and Hao (2008).

The research question being considered here is not focused on the transit system broadly, but rather on the built environment characteristics that influence pedestrian activity and related access to transit. In this respect transit level of service variables are included as intervening variables, which represent a viable predictor for latent demand at a bus stop. These variables are explained further in Section 3.3.3.2 Intervening variables: Transit. Understanding how these known transit variables are supported by the pedestrian environment is a primary objective in answering these thesis questions.

Chapter 3
Methods

3.1 Introduction

As outlined in the introduction the purpose of this research it to answer the following questions:

- How does the pedestrian environment / walkability correlate with public transit ridership?
    - What is the most appropriate way to measure pedestrian design, as it relates to walkability, and what is its correlation with transit ridership?
    - What walkability / built environment characteristics correlate best with transit ridership?
    - In what way is answering this question informed through the use of linear regression and spatial regression models?

This chapter defines: the study location, observation point, predictor, response and intervening variables, preliminary statistical transformations and spatial regression characteristics. All linear and spatial regression analysis and discussion is presented in Chapters 4 and 5.

The relationship between public transit, walkability and pedestrian behaviour has been explored from different directions, as established in Chapter 2.The resulting research design has been informed by previous academic material to best identify the variables which are to be analysed with respect to the pedestrian environment and public transit ridership. This research question will be answered through operationalising key variables and testing to establish the most effective linear regression model. After the most effective variables in linear regression models have been determined spatial regression will be used to further understand the observed relationship.

3.2 Research design

To answer the research question this research will employ a quantitative analysis of the impact of the predictor variables, walkability indicators, on the response variable, public transit ridership, while considering the effect of some intervening variables. A GIS framework is used to identify and create variables of interest in examining the built environment. Once all variables are standardised, linear regression is used to establish and examine the relationship. The best models are identified, as measured through adjusted R^2, these models and variables will be carried into spatial regression to further examine the relationship of the built environment.

This study employs the following tools:

- ARC GIS v10.1 to manipulate and collate variables,
- Microsoft Office Excel 2010 to manage data,
- SPSS v22.0 for descriptive and linear statistical analysis,
- Geoda v1.4.6 for spatial regression analysis

The research approach will employ several models to answer the research question. The research approach involves dividing variables into three categories: response, intervening and predictor. This variable language is similar to the use of the title "dependent and independent" from previous research design (Creswell, 2009). To thoroughly address this question this the statistical models will test two response variables, using linear regression models. The role of the intervening variable is to present a variable based on previous research which is known to affect public transportation ridership. These variables will assist in understanding the impact of the predictor variables and aim to operationalize the pedestrian environment. Intervening variables have been selected based on previous research and will be introduced in a base model to aid in analysis at all stages. The predictor variables, those operationalizing the pedestrian environment, will be introduced in subsequent models to observe impacts on public transit ridership. The order predictor variable introduction is based on previous research and anticipated result. Once the most explanatory response and predictor variables have been selected these will be combined into a model using spatial regression. This research design is heavily informed by Cardozo, Gardia-Palomares & Gutiérrez (2012).

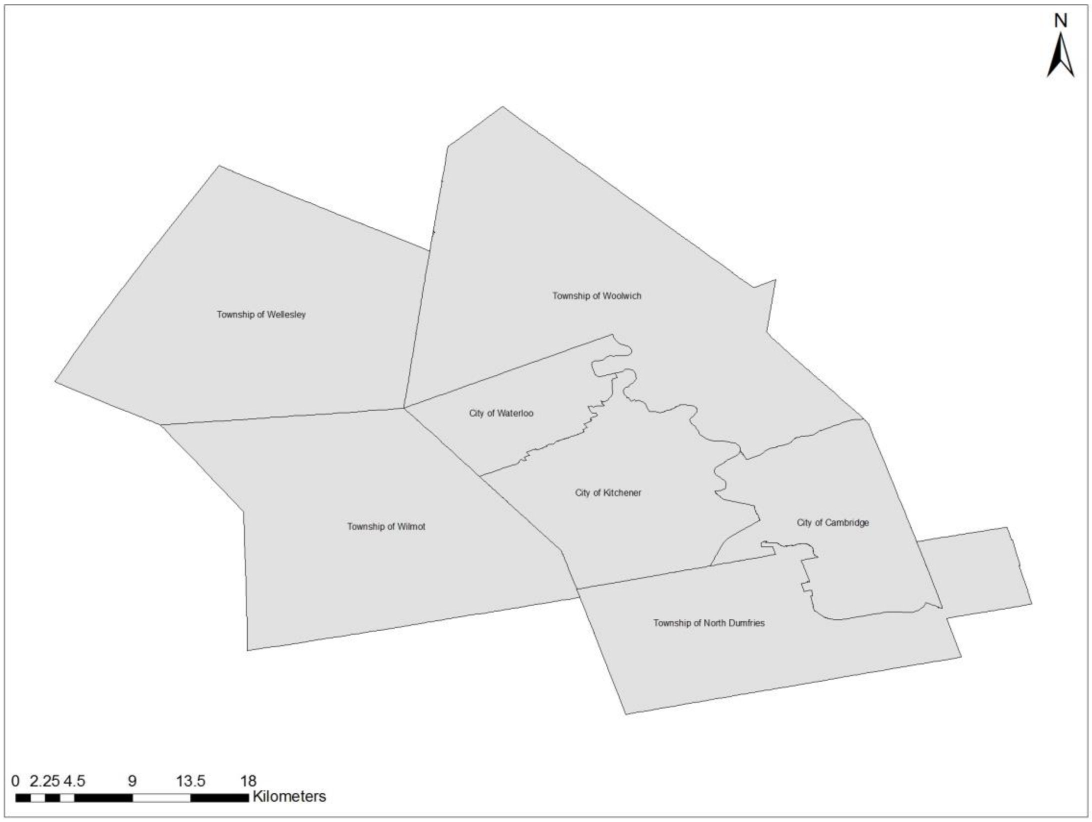
3.2.1 Study location

The study location is the Region of Waterloo, Ontario, Canada. The Region shown in figure 3.1, is comprised of three cities; Waterloo, Kitchener and Cambridge, and four Townships: Woolwich, Wellesley, Wilmot and North Dumfries. The combined population of Waterloo Region is over 550,000 (Region of Waterloo, n.d.). The 2011 Canadian Census reported that the region is comprised of 202,121 private dwellings which accounted for over 80% of total dwellings in the region (Region of Waterloo, 2011).

Public transit in the area is a service provided at the regional level by Grand River Transit (GRT), which is the sole public transit provider. The GRT has 66 regular routes, at the time of data collection, over 240 vehicles of which over 85% are accessible, which indicates the ability to lower and have low floors for access by mobility devices (Region of Waterloo, 2013). Each bus stop within this study is identified by the street name or intersection as well as a four digit stop identification code, referred to hereafter as stop ID.

This area has been selected based on location for study and availability of data.

Figure 3.1 Study location Region of Waterloo, Ontario, Canada



3.2.2 Public transit data

The GRT uses two methods to collect count ridership numbers. Primarily the fleet uses Mobile Statistics, which automatically measures boarding and alighting for busses equipped with Automatic Passenger Counters (APC). APCs use a combination of infrared signals to count numbers of people as they board or alight. The fleet is not entirely outfitted with the APCs and those busses not equipped are exclusively on six routes: 72, 73, 76, 9967, 9968, and 9977. These routes rely on the drivers to report manually at the end of the trip. Due to the inconsistency of the two methods of collection the data from the manual method of collection is removed from this research.

The data were collected using two time frames, one hour peak and all day ridership patterns. Provided data identify: date, time, route number, stop (by name and stop ID), and boarding and alighting (Grand River Transit, 2013). One hour peak travel time used the boarding and alighting data from 2773 bus stops during scheduled departure times between 15:00 and 16:00 over a 59 day period from January 2, 2013 to March 31, 2013. This results in 384,564 individual data points, where bus stop and route meet throughout that hour. During the 59 day period a route/stop/time combination reported between 1-59 times. The boarding and alighting are averaged at the route/stop/time combination to generate an average ridership over the hour or the day. For one hour peak data collection over 63% of the stops reported over half of the 59 days during the collection period.

All day data were collected between February 1, 2013 and March 31, 2013 based on scheduled departure times. The complete boarding and alighting numbers for the day was added and an average was taken for this time period. A total of 2492 stops reported all day averages.Those stops which did not report were omitted from this research. A map indicating all stops reporting *all day average boarding and alighting* is included in Appendix A.

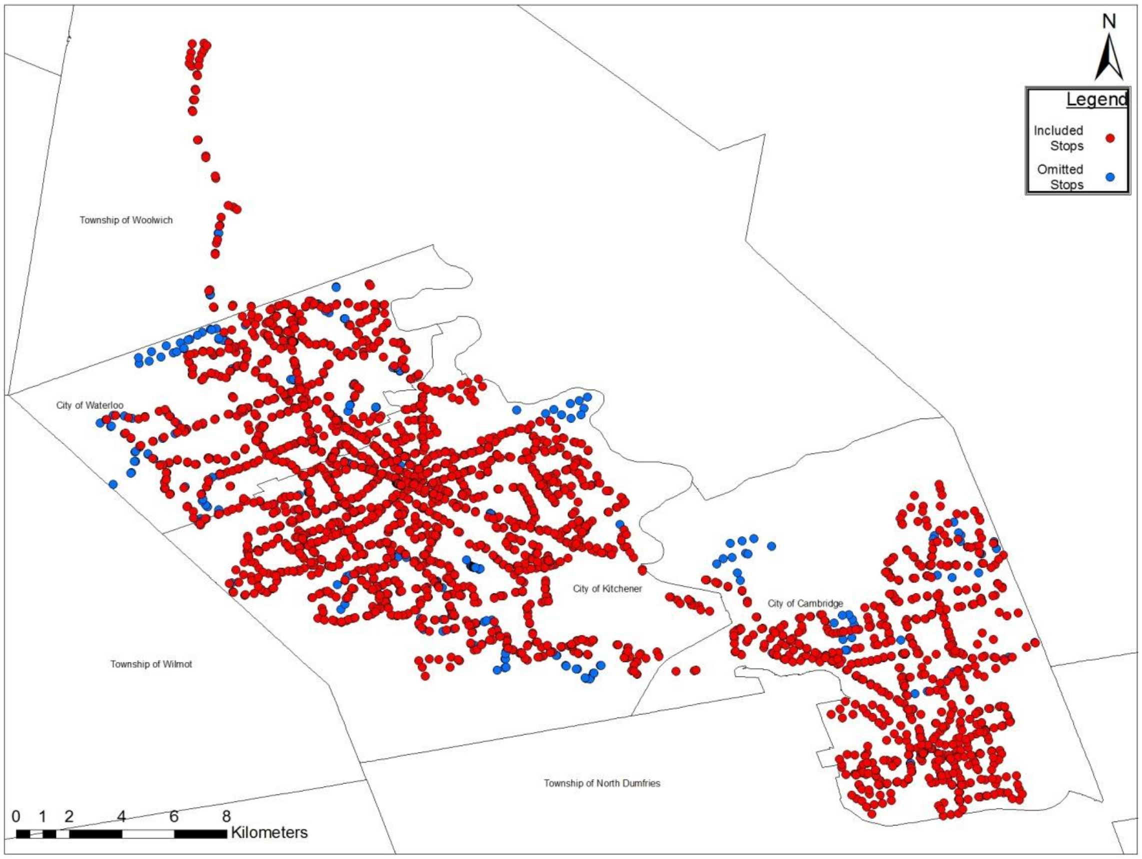
3.2.3 Observation point definition

This research uses individual bus stops as observation points for examining the pedestrian and transit relationship. In order to ensure that the built environment is standardized throughout the region, this study uses a 400m radius circular buffer around each bus stop to define the standard walking distance for each observation point. The 400m buffer size has been defined by academia and policies as the generally accepted distance a person is willing to walk to a bus stop (Hess, 2011; Mavoa, Witten, McCreanor, & O'Sullivan, 2012; Millward et al., 2013; Oliver et al., 2007; Ontario Ministry of Transportation, 2012; Wasfi et al., 2013). Due to the proximity of the observation points and the density of the bus stops, in many cases the stops have similar and overlapping areas. In this research however all bus stops and their areas are calculated independently, based on the argument that the predictor variables will affect them differently.

Due to proximity and built environment characteristics a circular buffer is used. A difference in results can develop through the use of different buffers shapes, such as a network buffer has been identified as a possible challenge in developing a consistent picture of the environment at each observation point (Oliver et al., 2007). Creating a network buffer would bias land-use variables towards the transportation mode elected to provide that network, either road or sidewalk.

Section 3.2.2 identifies that ridership data was collected for 2773 bus stops using the APC method. Section 3.3 will identify the predictor and intervening variables. Through this process it is recognized that 2488 have measurement for each of these variables, the remaining stops are omitted entirely from this study, figure 3.2 graphically presents the included and omitted data points as a result of data being complete.

Figure 3.2 Location of bus stops with indication of exclusion from research for incomplete data



3.3 Definition of variables

3.3.1 Response variable: boarding and alighting

Each data point collected by the GRT has associated number of passengers who got "on" and "off", referred throughout this research as boarding and alighting. While it is not contested that destination, origin and network each impact the transit journey, it is beyond the scope of this study to model those aspects of transit rider behavior. Rather as the study by Ryan and Frank (2009) articulates that boarding and alighting numbers are used as overall indication of transit demand at a stop. Their study showed that these numbers can be linked to the built environment. Linear regression has been used in academic research in the case of both these response variables. Su (2011) used one hour peak ridership and other studies have used average daily ridership (Ewing & Cervero, 2010; Ryan & Frank, 2009).

As described above all reporting ridership has been collated into averages, these averages have been linked by stop ID to determine overall traffic at every stop in the system. The data is collected separately counting each boarding and alighting. As the objective of this research is to measure pedestrian traffic at the stop both boarding and alighting variables are combined into one. Figure 3.3 and figure 3.4 show the number of stops where the average boarding and alighting for the one hour peak data and all day data respectively. This shows that the majority of the stops have a low traffic, which can result in skewed data. As normalcy is understood to be important in both linear and spatial regression models normalizing this data through transformation is addressed in section 3.4. There are other options in dealing with the data in this case, such as removing the cases with low average ridership, for example less than 1. The descriptive for this segmentation are presented in table 3.1. That table shows that even through this segmentation the need for transformation is no eliminated. Additionally, as the objective of this research is to examine the built environment the inclusion of these stops will help determine common land use characteristics around these low ridership stops. In this manner it is assessed that keeping the low ridership data will inform public transportation and pedestrian environment research equally.

Figure 3.3 Number of stops with average total boarding and alighting during one hour peak travel

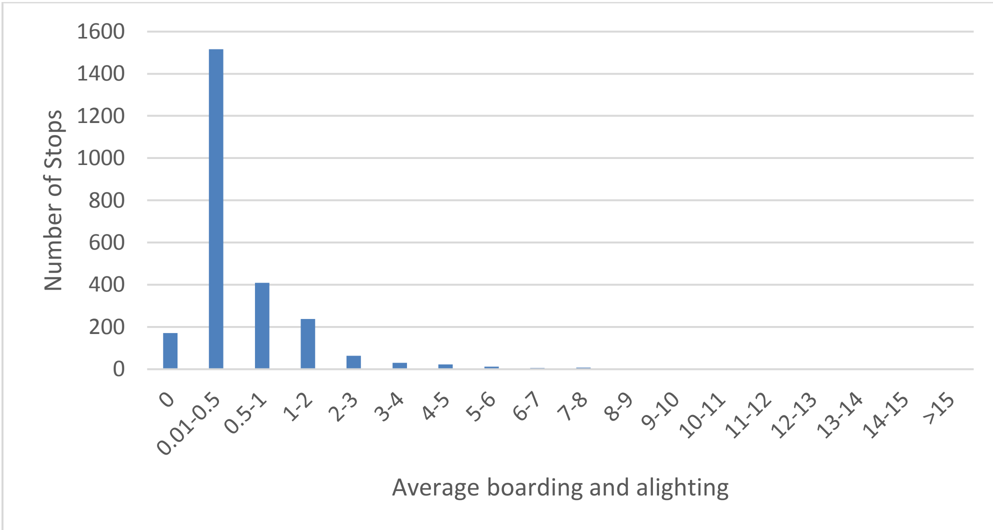


Figure 3.4 Number of stops with average total boarding and alighting during all day travel

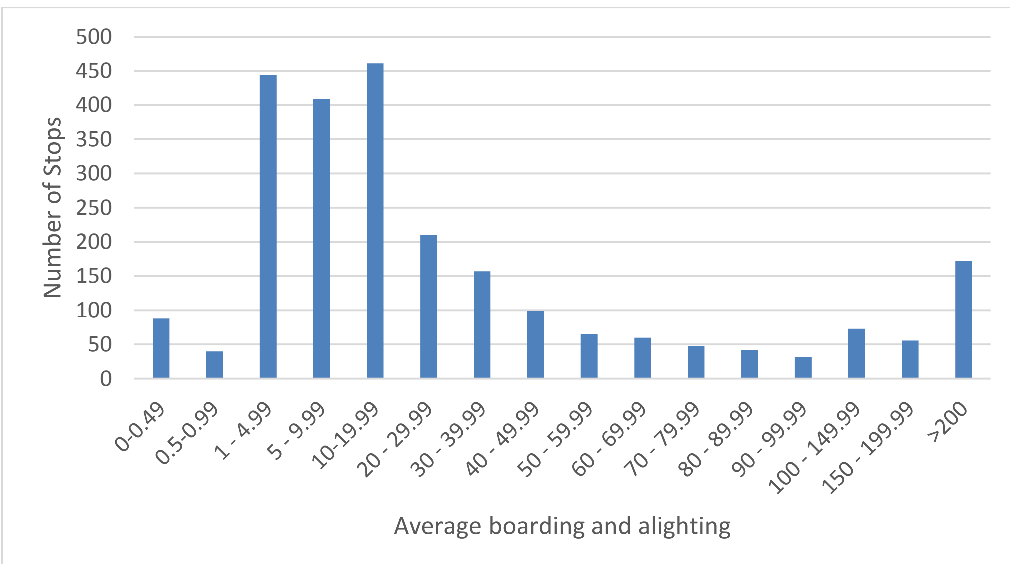


Table 3.1 Descriptive statistics for all day average boarding and alighting segmenting stops with low ridership

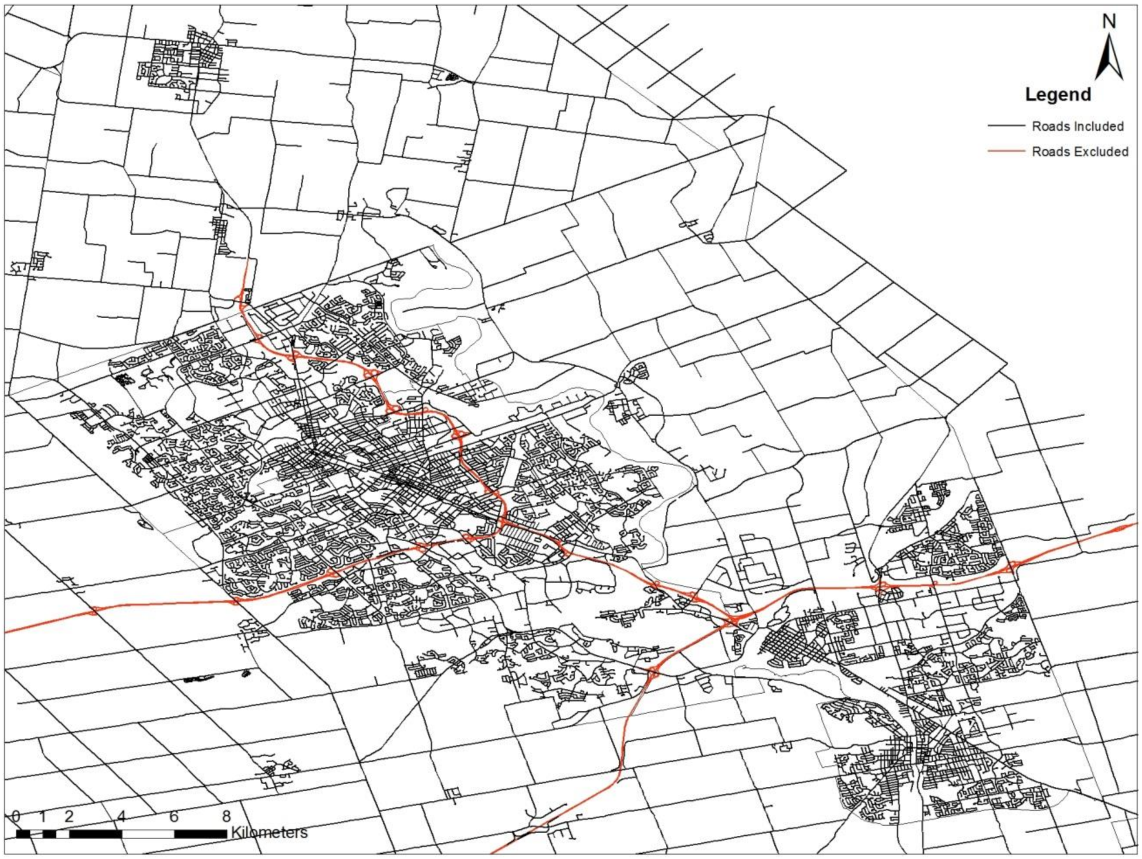
	Standard deviation	Skewness		Kurtosis		n
		Statistic	Std. Error	Statistic	Std. Error	
All stops	239.246	8.734	0.49	1114.359	0.98	2456
Stops with ridership <1	0.295	0.362	0.49	-0.238	0.98	128
Stops with ridership >=1	245.11	8.521	0.49	108.83	0.98	2328

3.3.2 Predictor variable: Pedestrian infrastructure

As explored throughout Chapter 2 pedestrian activity and walkability is shaped by elements of the built environment. Previous studies have identified the variables operationalized in this research which are all objective variables, as defined by Maghelal and Capp (2011), which can be quantified and replicated in other studies. This study aims to examine pedestrian infrastructure through various measurements: intersection density, sidewalk/road length, average designated traffic speed, and traffic lights density. These elements contribute heavily to both the perception and actual safety of pedestrians who are on route to bus stops as part of their travel (Campbell et al., 2004).

Sidewalk data is received as a shape file from the Region of Waterloo, this data is part of a corporate dataset and self-reported by the municipalities within the region. This data is considered current to August 2013 (Information Technology Services, 2013). Road data reflects all roads within the Region of Waterloo and is considered current to 2012 (Region of Waterloo, 2012). For relevancy; in this research expressways, including the associated ramps, are removed from road data as they do not serve a pedestrian function nor do they have public transit stops. This revisions of the data are represented in figure 3.5.

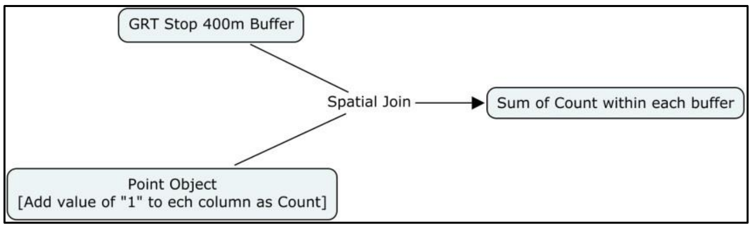
Figure 3.5 Region of Waterloo road network with indication of excluded roads from research



3.3.2.1 Intersections density

As illustrated in section 2.3 previous studies assessing walkability examine intersection density as a method for measuring pedestrian design, permeability and block design (Cervero, 2002; Ryan & Frank, 2009; Samimi et al., 2009). As this method has been used previously, it has been included as a measure of permeability and assesses previous methods and compare it with other variables used in this research. The variable intersection density is calculated as a point object. Figure 3.6 graphically explains how this data was collated. These figures are developed using the layout of GIS model builder to graphically depict the process of collating the data into the variable form. These processes are based on the different layout of variables to be merged. Using figure 3.6 the 400m buffer is the circle centred by the bus stop location, point objects are then joined as they overlap spatially. This process joins the point object with every buffer that encircles it and a total is summed. The final product is a buffer, identified by stop ID, which includes the characteristic of the sum of all point objects within the buffer.

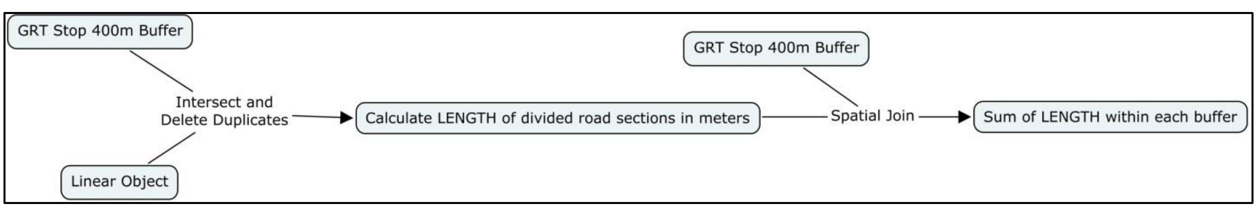
Figure 3.6 Methods point object



3.3.2.2 Sidewalk and road length

This research is examining specifically the role of pedestrian infrastructure, in order to assess its difference between the pedestrian infrastructure and previous methods using road analysis. The sidewalk data however, is not in a format which lends itself to analysis of intersection density. Instead the length of both roads and sidewalks within the observation points are calculated independently and the data collected based on the purpose. Figure 3.7 graphically explains how linear data was collated.

Figure 3.7 Methods linear object



The resulting data is used to create two variables:

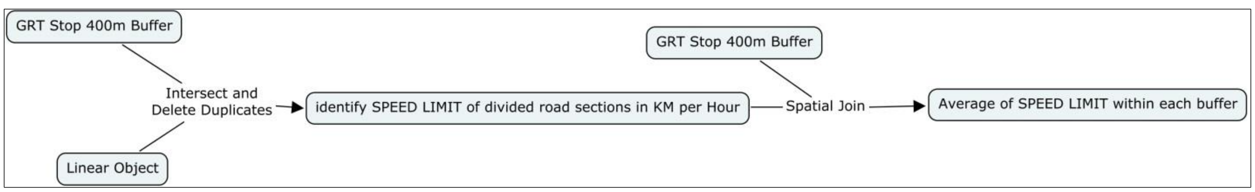
- Sum of sidewalk length within each buffer, this will establish an objective measurement of pedestrian infrastructure within the bus stop area.

- A ratio of sidewalk length to road length, which will be a direct comparison of how walk-car friendly the area is. A perfectly equal environment would have a Length Sidewalk: Length Road ratio of 2:1. While this does not consider width of either surface, this ratio provides a numerical representation of which mode of travel has priority to the public space. This method has been established previously by Cervero (2002).

3.3.2.3 Average designated traffic speed

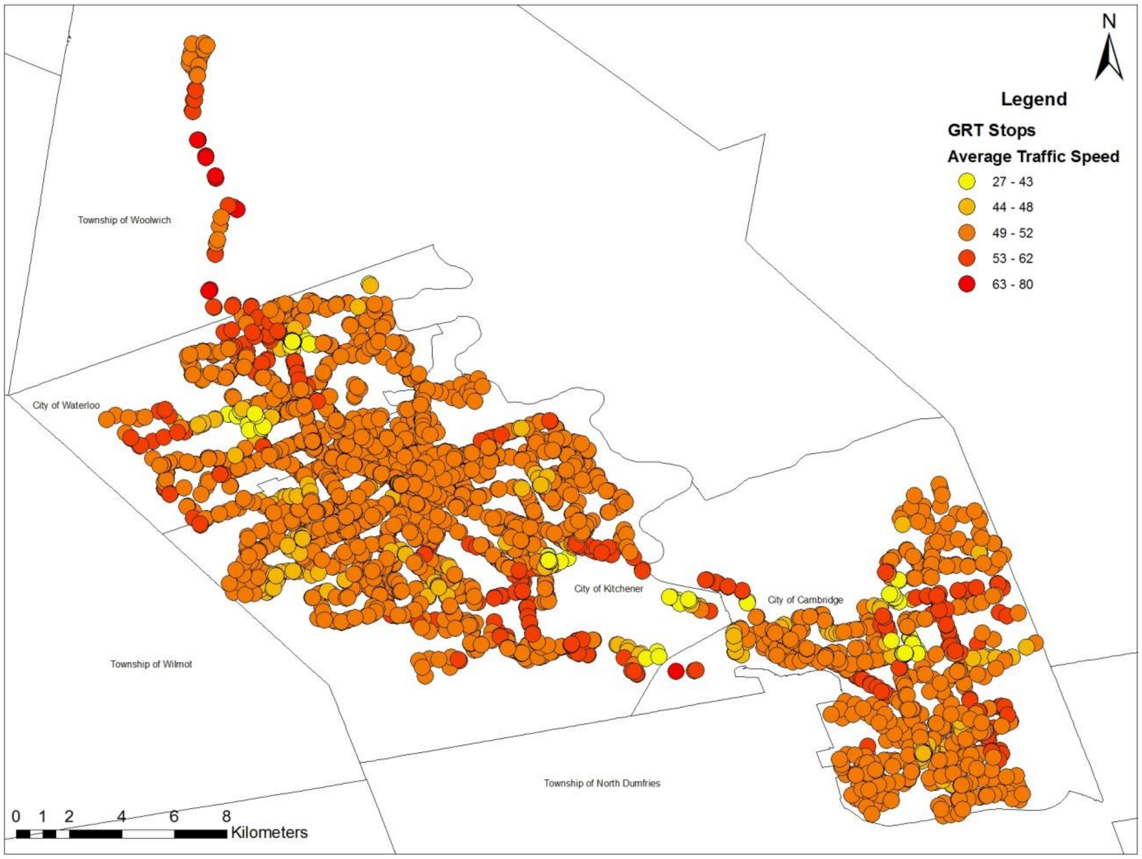
A reduction in traffic speed creates an environment which is safer for other modes of travel, in the case of this study specifically the pedestrian environment (Maibach et al., 2009; Werner et al., 2010; Wey & Chiu, 2013). In order to assess which bus stops are more pedestrian friendly the average speed limit within each buffer was collated, graphically represented in figure 3.8. Speeds were determined as part of the Region of Waterloo dataset considered current to 2012 (Region of Waterloo, 2012).

Figure 3.8 Methods traffic speed



The average traffic speed by bus stop buffer, using the *one hour peak* stops, is presented in figure 3.9, which indicates higher speeds with darker colours.

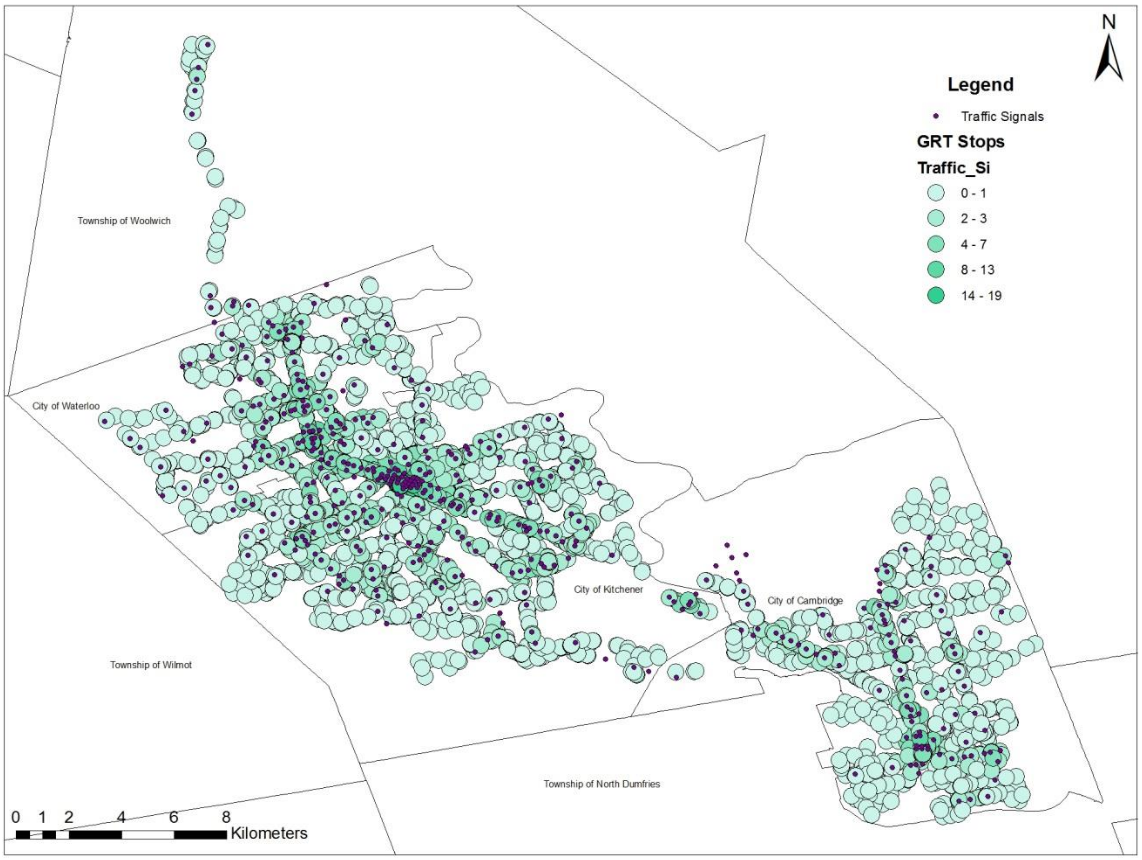
Figure 3.9 Map of average traffic speed per 400 meter bus stop buffer



3.3.2.4 Traffic signals

Traffic Signals are important measures which reduce traffic speeds and give pedestrians a safe crossing point (Gassaway, 1992; Papadimitriou et al., 2009). The Region of Waterloo has the GIS location of traffic lights throughout the Kitchener-Waterloo-Cambridge which is considered current to 2009 (Region of Waterloo, 2009). Each traffic light has been joined with the associated bus stops, this has been done as a point object. Figure 3.6 graphically explains how this data was collated. Figure 3.10 identifies all traffic lights in the region and associate traffic light density by bus stop, using the *one hour peak* stops, where darker colours have higher density of traffic lights. It can be seen that the highest densities are located in the core of the City of Kitchener and along core roads.

Figure 3.10 Map of traffic signal density within 400 meter bus stop buffer



3.3.2.5 Land use diversity

As is examined in section 2.3.3 mixed used communities have been established as favourable to the pedestrian experience. In order to create an index for mixed use previous studies have created methods of calculating entropy of land use (Cervero, 2002; Maghelal & Capp, 2011; Ryan & Frank, 2009). For this research the equation provided by Ryan and Frank (2009), shown as equation 1, will be used. This equation establishes a scale from 0-1 where areas with higher diversity in land use have values closer to 1. The manner in which land uses were identified and total area was determined is graphically represented as figure 3.11. The data used in creating this index reflects land use within the Region of Waterloo from 2006 and 2007 (Planning Housing and Community Services, 2007). For the purposes of this study land use codes roads and rail were removed from the data, the remaining land use codes are: agriculture, residential, commercial, industrial, extraction and open Space.

$$\boldsymbol{Equation}\ 1.$$

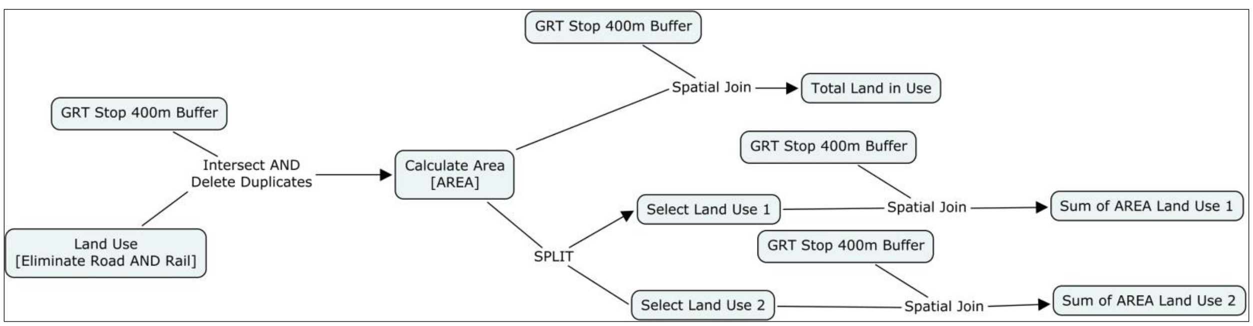
$$\boldsymbol{Entropy} = \frac{-\sum[\boldsymbol{P_n} * \ln(\boldsymbol{P_n})]}{\ln(\boldsymbol{N})}$$

Where:

P_n = proportion of area of the nth land use within buffer

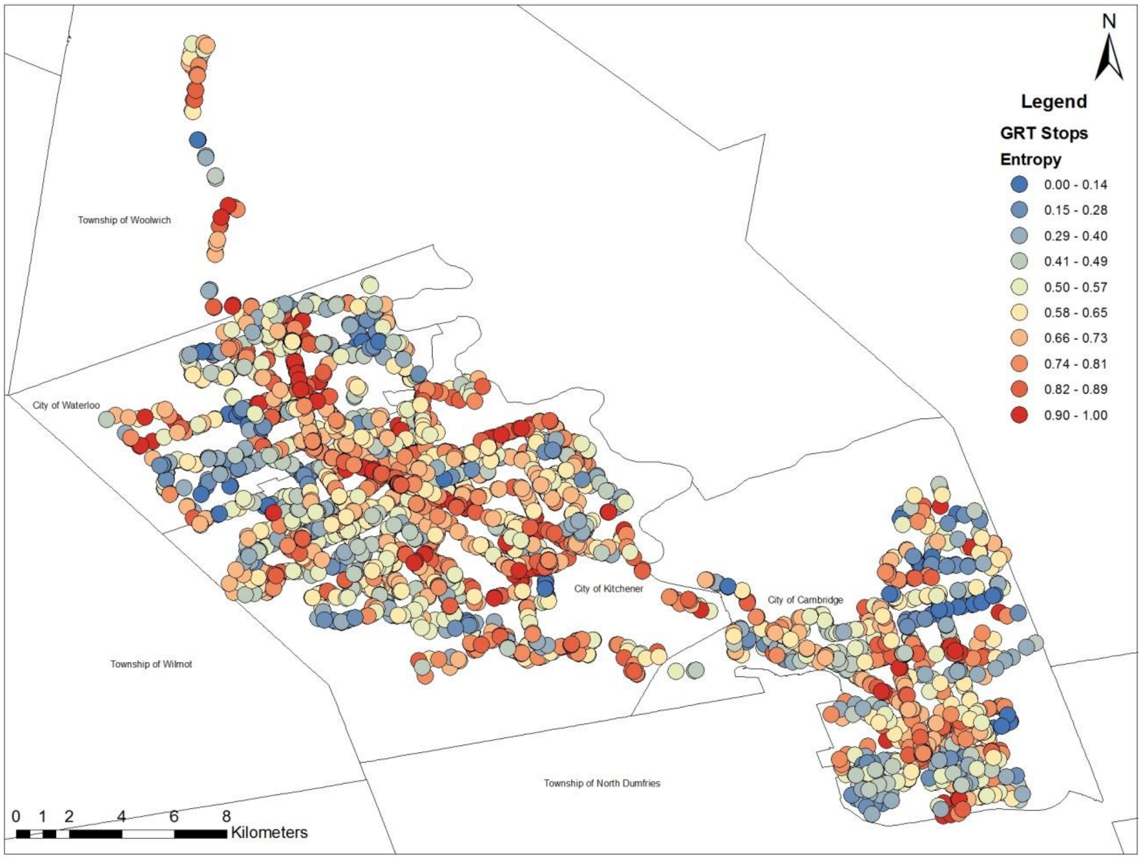
N = the number of different land uses within that buffer area

Figure 3.11 Methods land use entropy



The results of this equation are presented in figure 3.12, using the one hour ridership stops, where blue has less diversity and red has more diversity in land use. Here the concentrations of mixed use environments can be seen along the core of both the cities of Kitchener and Waterloo. There are several pockets of highly diverse land uses located throughout the bus network along with several very homogeneous neighbourhoods.

Figure 3.12 Map of land use entropy by 400 meter bus stop buffer



3.3.3 Intervening variables

Intervening variables are those which are known to affect or meditate relationship between the predictor variables and the response variable (Creswell, 2009). As was discussed in section 2.5 Transit level of service, there are many variables known from other research to affect public transportation ridership. As part of this research two types of intervening variables have been identified: built environment and transit system.

3.3.3.1 Intervening Variables: Built Environment

3.3.3.1.1 Population and Employment Density

Population and Employment are both crucial in predicting transportation ridership, as they indicate consumer access and availability (Cervero, 2002; Horner & Murray, 2002; D. Thompson, 2011). It is supported within previous research that built environment characteristics which affect walkability also affect public transportation ridership, as discussed in Chapter 2 – Literature Review. Since they are identified as impacting both mode choices, pedestrian and public transportation they are identified as intervening variables. Population data for the Region of Waterloo has been collected from the Statistics Canada 2011 Canadian census collated by Census Distribution Area (Statistics Canada, 2012). Employment data has been assembled from the Statistics Canada 2006 Canadian Census which established a place of work index throughout the region (Statistics Canada, 2008).

The data for these variables is presented as an example of the Modifiable Areal Unit Problem (MAUP) which has been extensively studied by academics. The data has been divided and coded digitally along borders which are spatially different from the study area; therefore, to align the data within the borders, an aggregate weight of the population is taken across each observation point area (Dark & Bram, 2007; Horner & Murray, 2002). This data has been collated as an MAUP object, figure 3.13 and figure 3.14 graphically depict how this was achieved.

Figure 3.13 Methods redefining borders to solve MAUP

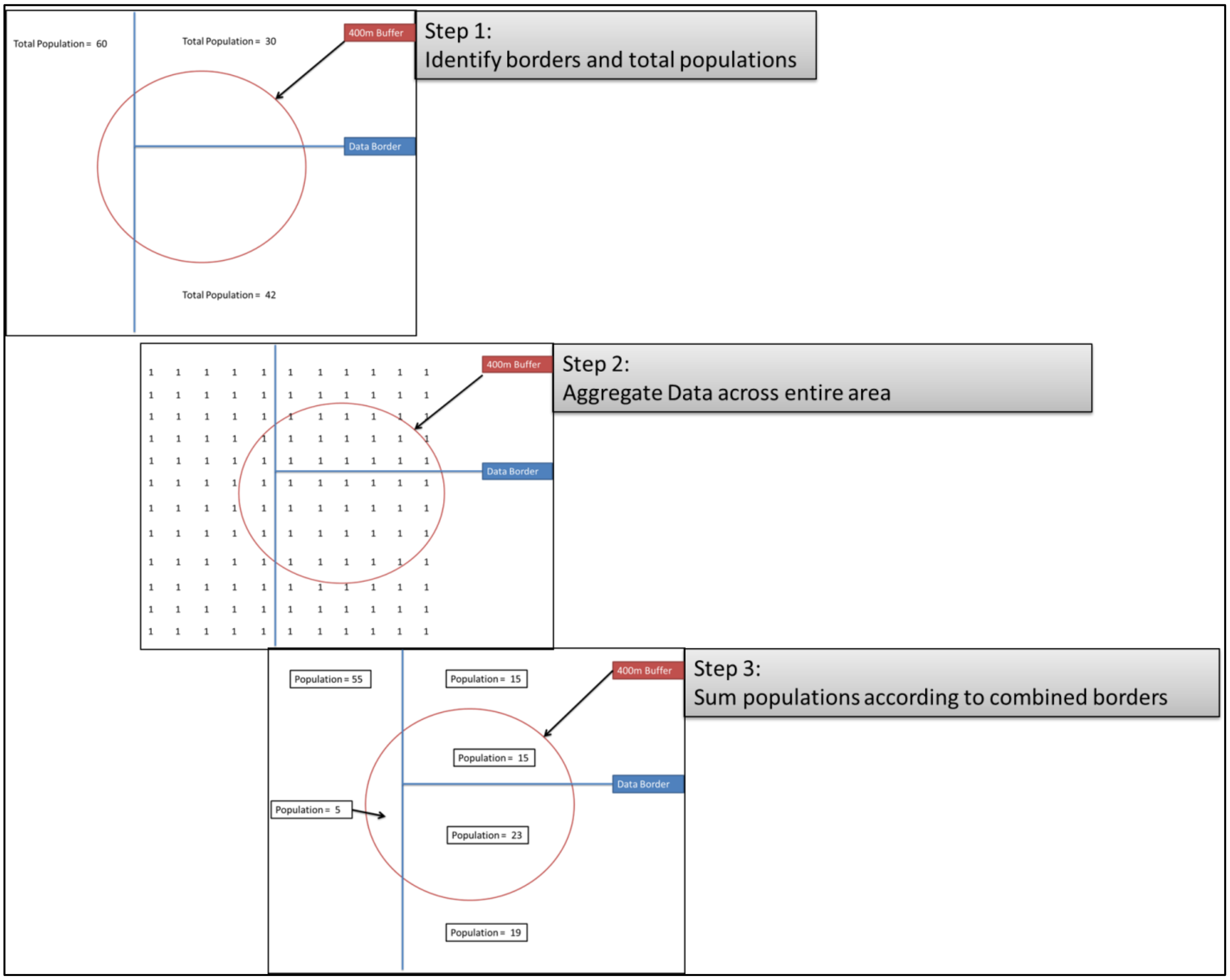
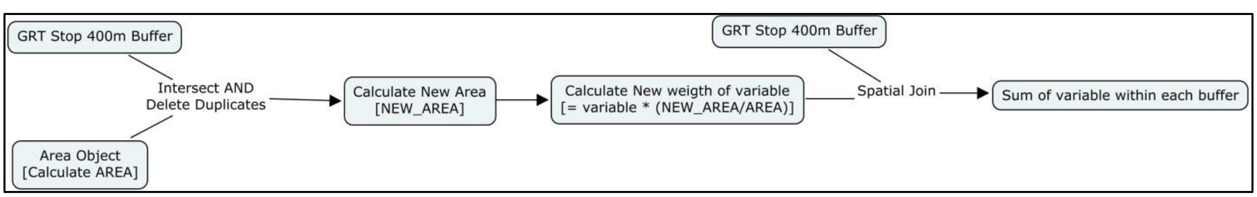


Figure 3.14 Methods collating data to solve MAUP



Population and employment density is presented, using the one hour average ridership, in figures 3.15 and 3.16 respectively, where darker colours indicate higher density. It can be seen by comparing these two maps that the Region of Waterloo has population concentrated in the outer areas, and employment concentrated in the core.

Figure 3.15 Map of population density within 400 meter buffer of bus stop

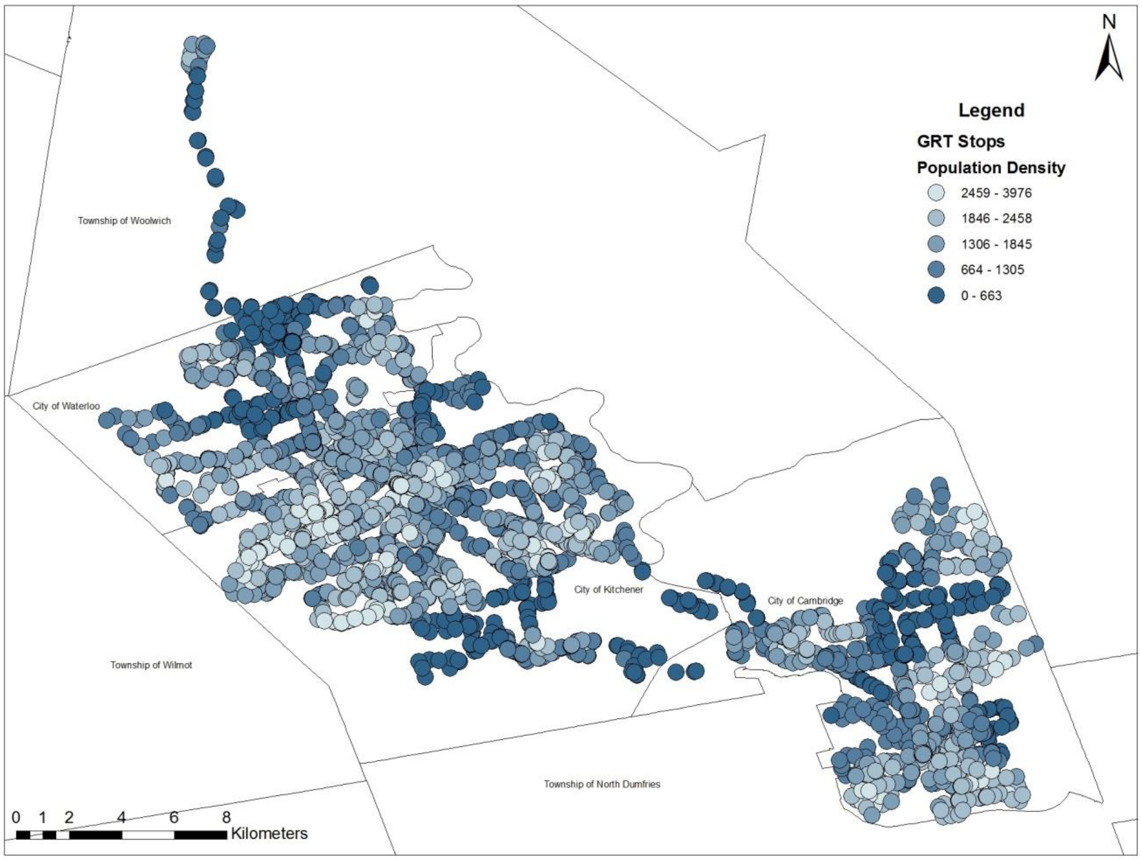
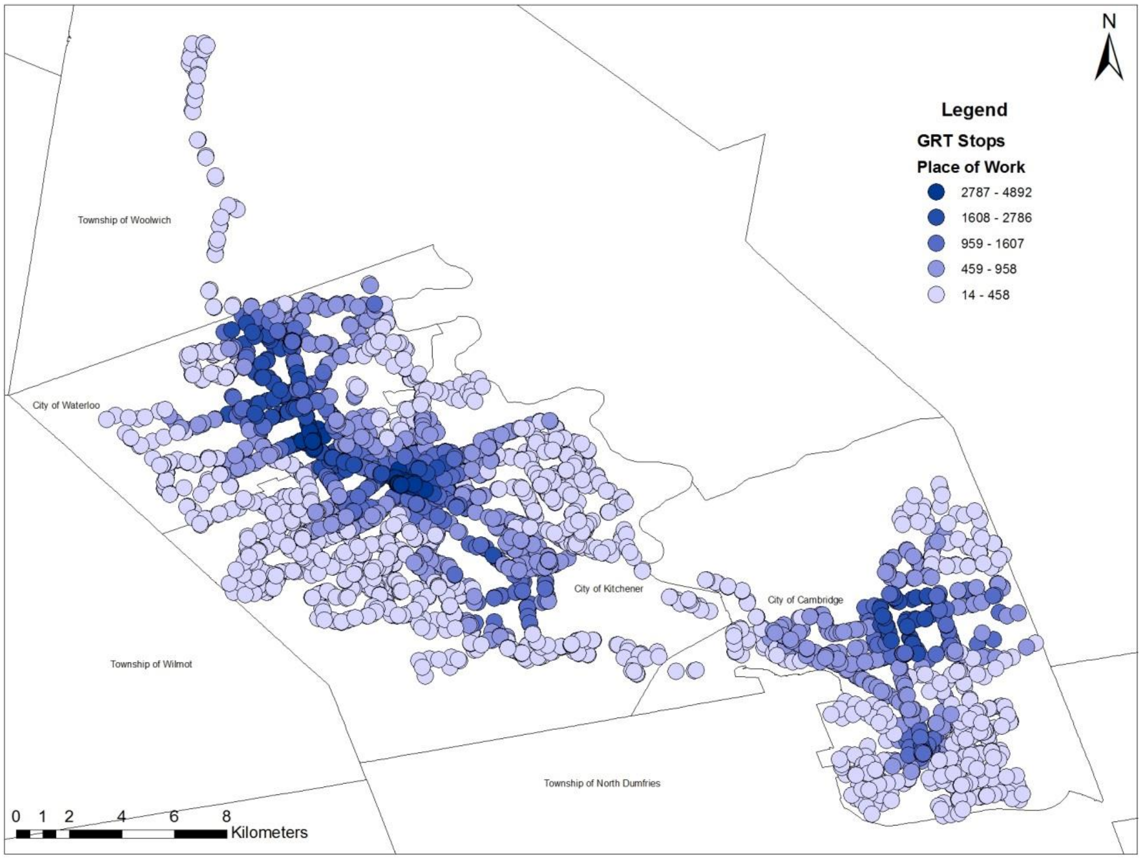


Figure 3.16 Map of employment density within 400 meter buffer of bus stop



3.3.3.2 Intervening variables: Transit system

3.3.3.2.1 Transit transfer location

Transfers create an opportunity for pedestrian travel still associated with public transit ridership (Guo & Ferreira Jr, 2008). The GRT has identified terminals, satellite terminals and major transfer points, listed in table 3.2: terminals having a central building and customer service staff, while satellite terminals and major transfer points have none (Region of Waterloo, 2011). Beyond the identified locations, it is important to recognize high number of corresponding transit stops as favourable transit locations which will generate pedestrian activity.

Table 3.2 GRT identified terminals, satellite terminals and major transfer points

Terminal Name	Terminal Type	Associate Bus Stop IDs
Charles Street Terminal	Terminal	2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2708, 2709, 2710, 2711
Ainslie Street Terminal	Terminal	1511, 1512, 1513, 1514, 1516, 1517, 1518, 1519, 1520, 1521, 1522
Conestoga Mall Terminal	Satellite Terminal	1125, 1126, 1127, 1128, 1129, 1130, 3798, 3799, 3800
Fairview Park Mall Terminal	Satellite Terminal	1046, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 3228
Cambridge Centre Mall Terminal	Satellite Terminal	1385, 1386, 1387, 1388, 1389, 1390, 1391
Highland Hills Mall Terminal	Satellite Terminal	2974, 2975, 2976, 2977, 2979, 2980, 3140
Forest Glen Plaza Terminal	Satellite Terminal	1765, 1766, 1767, 1768, 1769, 1770, 1771, 1772
Conestoga College	Major Transfer Point	3801, 3802, 3888, 1733, 1732, 1728, 3641, 1731, 1729, 3803, 3804
Hespeler	Major Transfer Point [note: this road is very long, very few of the associated bus stops are accessible from one another]	1476, 1427, 1478, 1480, 1459, 1321, 3806, 1392, 1454, 1325, 3527, 1475, 1428, 1477, 3537, 3538, 1458, 1481, 3539, 1462, 1426, 1479, 1384, 1460, 1461
University of Waterloo	Major Transfer Point	1122, 2519, 3700, 2517, 1123, 3699, 1124, 2516, 2515, 2518
Wilfrid Laurier University	Major Transfer Point	1167, 3619, 3620
Stanley Park Mall	Major Transfer Point	1017, 1070, 1667, 1668

3.3.3.2.2 Transit Level of Service/Frequency

Section 2.5 explains that frequency of service by public transit has been associated with the quality of the transit service. The higher the level of service the more access a user will have to the public transit as waiting times are reduced, this variable is known to have a positive effect on ridership levels (Delmelle et al., 2013; Lai & Chen, 2011; Ryan & Frank, 2009; Tang & Thakuriah, 2012; Wasfi et al., 2013). This variable is a indication of potential demand which will quantify a crucial role the GRT plays in influencing ridership.

To determine headway the number of bus routes and their appropriate frequency, was divided from 60 minutes. This is calculated from the scheduled arrival times during a peak hour for every stop. The results are shown in table 3.3. This indicates that 65 stops were at 60 minute headway, while during the same time period 84 stop were at 5 minutes or less. This method of calculating headway has the effect of standardizing all route types at a given stop. For the purposes of analysis this variable was collated into four levels, informed by Ontario Ministry of Transportation (2012):

- level A, 0-10 minute wait time
- level B, 10-20 minute wait time
- level C, 20-30 minute wait time
- level D, >30 minute wait time

Table 3.3 Frequency of scheduled bus arivals

Minutes Between Busses	Number of Stops
0-5	82
5-10	289
10-15	524
15-20	483
20-25	258
25-30	0
30-35	791
35-40	0

40-45	0
45-50	0
50-60	0
60+	61

3.4 Statistical Methods

Ewing and Cervero (2010) showed the academic common trend in the use of linear regression to examine relationships between travel and the built environment. In keeping with that tradition, this research will use linear regression as its foundation. Once the linear relationship is examined and predictor variables which show influence the response variable are identified, spatial regression will be used for further examination. This method of using linear regression to maximize the explanatory ability of spatial regression in built-environment and public transit academia has been previous established by Cardozo et al. (2012). As a requirement for linear regression a normal distribution of data is assumed. The next section will identify statistical transformations on an individual variable basis. Section 3.4.2 explains the spatial regression model, detailed analysis and measurements are conducted in Chapter 5.

3.4.1 Statistical transformations

Regression assumes normal distribution of data, several variables selected for this research show non normal distributions. Explained in Verma (2013) a skewness or kurtosis statistic of > +2 or < -2 is considered not normal, any variable with this condition is considered for transformation. Using the transformations, the Ladder of Powers, presented in De Veaux et al. (2012), the skewness and kurtosis presented in several variables will attempt to be addressed. A natural logarithm is a common transformation to address these issues, the solution used in presented as Equation 2.

Equation 2.

$$Transformed\ variable = \log(x + y)$$

Where:

x = the variable to be transformed

y = a small number added to ensure no answer of 0, often 1.

3.4.1.1 Average boarding and alightings

The average boarding and alighting variables are defined in section 3.3.1, it can be observed in figures 3.3 and 3.4 that this data is positively skewed resulting from the significant number of results between 0-1. Descriptive analysis of the *One hour peak average* variable reveal a skewness statistics of +5.491 (sig. 0.049) and kurtosis of 39.25 (sig. 0.098). Descriptive analysis of the *All day average* variable reveal a skewness statistics of 89.73 (sig. 0.049) and kurtosis of 114.359 (sig. 0.098).

The data for both these variables is leptokurtic, resulting from a concentration of data at the centre of the distribution. The natural logarithm was calculated using Equation 2 where y = 0.1. The *Log One hour average boarding and alighting* returned a skewness of 0.648 (sig. 0.049) and kurtosis of 0.371 (sig. 0.098). The *Log All day average boarding and alighting* returned a skewness of 0.097 (sig. 0.049) and kurtosis of 0.548 (sig. 0.098). The transformation has made these variables viable for this research.

3.4.1.2 Traffic Signal Density

Descriptive analysis of the variable *Traffic signal density* revealed a skewness statistics of 3.381 (sig. 0.049) and kurtosis statistic of 14.802 (sig. 0.098). This data is both positively skewed and leptokurtic. This is a consistent error, as seen in figure 3.10, most areas do not have traffic light and those areas with one traffic signal tend to have several. As above a natural logarithm is a common transformation to address this, achieved using Equation 2 were y = 1. This transformation results in a skewness of 0.519 (sig. 0.049) and kurtosis of 0.22 (sig. 0.098), this variable in this state is now viable for this research.

3.4.1.3 Average Traffic Speed

Descriptive analysis of this variable revealed a kurtosis statistic of 15.697 (sig. 0.098). The distribution is leptokurtic, resulting from a concentration of data in the centre of the distribution. This distribution error is anticipated as 50km/h is considered the normal posted speed within urban centres, therefore creating an average concentrated around that value. The transformation of this variable was not successful in creating a normal distribution, therefore this variable is not viable for regression analysis. As explored in section 2.4.2 *traffic speed* this is an important factor in pedestrian safety. In order to effectively create this variable more variance would be required in the data. A method of creating that variance is monitoring traffic, unfortunately monitored data was not available for analysis for this research.

3.4.1.4 Employment Density

Descriptive analysis of *Employment density* revealed a skewness statistic of 2.33 (sig. 0.049) and kurtosis statistic of 6.68 (sig. 0.099). This distribution is both positively skewed and leptokurtic. As seen in figure 3.16 employment in the Region of Waterloo is highly concentrated which has created a skewed result. Several transformations were attempted to create normalcy in the distribution. The best result was achieved using the natural logarithm of this variable, reference Equation 2 where y = 1. This transformation results in a skewness of -0.16 (sig. 0.049) and kurtosis of 0.25 (sig. 0.098), this variable is therefore viable for this research.

3.4.2 Spatial Regression

Spatial Regression considers the role of spatial location and incorporates that into the regression equation. In linear regression predicting the value of the spatial location is irrelevant, where spatial regression uses the coordinates of each data set to locate similarities across space (Cardozo et al., 2012). Spatial regression has been commonly used in planning and exploration of the built environment but with few exceptions has not been regularly used in examining public transit and built environment characteristics (Cardozo et al., 2012; Chow et al., 2000).

Prior to spatial analysis a test to demonstrate a spatial relationship is done, in this research Moran`s I. Moran's I is used to determine spatial autocorrelation, for a detailed examination refer to Anselin (1995). When spatial autocorrelation is shown to be present the use of spatial regression is a recommended course of analysis (Cardozo et al., 2012; Ward & Gleditsch, 2008). Further discussion about Spatial regression and the Moran's I statistic will be presented in Chapter 5.

3.5 Methods Conclusion

In this chapter the variables have been identified and the methods of regression are explained. The methods of data collation and transformation have been outlined for these variables. Table 3.4 presents a variable summary table with predicted sign and impact of the predictor and intervening variables upon the response variable of average boarding and alighting. The principal hypothesis is that public transit traffic is higher in areas that have a built environment that is more amenable to the pedestrian experience.

The list of pedestrian environment variables, presented as predictor variables, employs several different methods for measuring the pedestrian environments impact on public transit ridership. The uses of both linear regression and spatial regression been explained and an indication of a special auto-correlation has be shown with the predictor variables. Subsequent chapters will explore the results of the regression models and discuss the tools with which to examine how public transportation ridership is impacted by the pedestrian environment.

Table 3.4 Variable summary table

Variable Name	Variable Type	Measuring	Role	Sign/Impact
Log one hour average boarding and alighting	Scale	Natural logarithm of average boardings and alightings from 1 hour of peak travel time	Response variable	N/A
Log all day average boarding and alighting	Scale	Natural logarithm of average boardings and alightings from all day travel	Response variable	N/A
Intersection count	Scale	The number of intersections within 400 meter buffer of bus stop	Predictor variable	+
Sidewalk length	Scale	Length of sidewalk within 400 meter buffer of bus stop	Predictor variable	+
Ratio	Scale	Ratio of sidewalk and road within 400 meter bus stop buffer	Predictor variable	+
Traffic Speed	Scale	The average signed speed limit within 400 meter buffer bus stop	Predictor variable	omitted
Log traffic signal	Scale	Natural Logarithm of count of traffic signals within 400 meter buffer of bus stop	Predictor variable	+
Entropy	Scale	Mixed of land uses within 400 meter bus stop buffer	Predictor variable	+
Population density	Scale	Count of people within 400 meter bus stop buffer as MAUP from census tract	Intervening variable	+
Log employment	Scale	Natural logarithm of the count of place of work as MAUP	Intervening	+

density		problem from census tract	variable	
Employment density	Scale	Count of place of work as MAUP problem from census tract	Intervening variable	+
Transfer location	Ordinal	GRT designation of terminals	Intervening variable	+
Level of service	Ordinal	Frequence of all buses stopping at bus stop	Intervening variable	+

Chapter 4
Results and Discussion: Linear Regression

4.1 Introduction

Previous chapters have established the theoretical framework and methods for answering the following questions:

- How does the pedestrian environment / walkability correlate with public transit ridership?
    - What is the most appropriate way to measure pedestrian design, as it relates to walkability, and what is its correlation with transit ridership?
    - What walkability / built environment characteristics correlate best with transit ridership?
    - In what way is answering this question informed through the use of linear regression and spatial regression models?

The objective of this chapter is to establish a linear regression model between the variables presented in table 4.1. Subsequently the various predictor variables will be included over several linear models to examine the pedestrian environment. This chapter aims to identify which is more explanatory in the case of answering this research question. As part of preliminary analysis and establishing the viability of the various models, stepwise regression has been employed.This has not been included in analysis as conclusions are informed by the base model and subsequent predictor variable introductions presented here.

Analysis of regression models is done exclusively through the use of ordinary least squares regression, referred to as linear regression. This decision is founded on previous academic use of linear regression in examining travel behaviour and the built environment, as revealed in the exploration of academic material by Ewing and Cervero (2010). This thesis will employ the most explanatory set of variables, as assessed by adjusted R^2 and statistical significance, to be used in spatial regression models in Chapter 5.

This chapter approaches this analysis through first developing a Base Model, this is the model between intervening variables and response variables. The intent of this approach is to determine which variables known to affect public transportation ridership will correlate with the Grand River Transit (GRT) system. The Base Model will permit the identification of outlying cases. These outlying cases will be explored to determine if they show explanatory qualities in answering this research question. This Base Model will also serve to inform the relationship between the predictor variables to ensure their correlation remains consistent with previous research.

Once the Base Model is identified, predictor variables will be introduced testing specific models; the order of introduction is presented in table 4.1. Where cells are merged it indicates that these variables will be assessed against one another to determine their impact and significance. At the end of this chapter some conclusions will be examined and a model will be presented which will be used for spatial regression analysis in Chapter 5.

Table 4.1 Variable definition and point of interdiction for linear regression

Name	Measuring	Point of introduction
Log one hour average boarding and alighting	Natural logarithm of average boarding and alighting during one hour peak period	N/A
Log all day average boarding and alighting	Natural logarithm of average boarding and alighting from all day travel	
Population density	Count of people within bus stop buffer as MAUP from census	Base Model

	tract	
Transfer location	GRT designation of terminals	Base Model
Level of Service	headway of all buses stopping at bus stop	Base Model
Log employment density	Natural logarithm of the count of place of work as MAUP problem from census tract	Base Model
Employment density	count of place of work as MAUP problem from census tract	
Entropy	Mixed of land uses within bus stop buffer	Model 1
Intersection count	The number of intersections within buffer of bus stop	Model 2
Sidewalk length	Length of sidewalk within buffer of bus stop	
Ratio	Ratio of sidewalk and road length within bus stop buffer	
Log traffic signal	Natural logarithm of count of traffic signals within bus stop buffer	Model 3

4.2 Base Model

As stated in the introduction of this chapter, the Base Model is being used to measure variables which are known to affect public transit ridership and assess them in the context of the GRT. This model is the natural logarithm of the average boarding and alighting and the intervening variables. This regression analysis will be conducted independently for each response variable. The intervening variables are:

- Population density
- Employment density

- Transfer location
- Transit level of service

As previously explained these are not the only potential intervening variables known to affect public transportation ridership. These variables are four variables for which data could be collected and collated which are consistently shown to influence public transport ridership. This omission serves the necessity of minimising inaccuracy in data through further expanding the years of the data set, As well as ensuring the variables related to the pedestrian environment do not become obscured by known predictive characteristics. Understanding that not all variables known to correlate with ridership are included, the adjusted R^2 is not anticipated to be highly predictive.

The density variables serve a dual purpose, as established in section 2.3.1 density correlates as a base measurement for both walkability and public transportation. These variables provide explanatory abilities both in establishing the Base Model through revealing impact public transportation and in supporting the research question. The use of density as an intervening variable serves to limit conclusions about their effect on walkability when the impact on public transportation is already known.

The Base Model serves the purpose of examining the regression analysis through the use of both response variables: the natural logarithm of average boarding and alighting between one hour peak time and all day travel time. Preliminary analysis of the Base Model indicated the addition of another variable, the presence of a secondary school, may be required and the reasons for this are examined. Similarly preliminary analysis established that the variables *Employment density* and *Logarithm employment density* did not correlate as expected and these results are discussed. Finally this section will be used to explore outlying cases which will be examined to determine if these stops have higher or lower pedestrian environment characteristics than other stops.

4.2.1 Intervening Variables

This section explores two key variables which were considered while conducting preliminary testing using the Base Model. Firstly, the use of secondary schools: this variable was examined initially during exploration of the data and presented an interesting relationship particular to the *Log one hour peak average boarding and alighting*. The second variable explored here is the variable *Employment density*, which as discussed in section 3.4 presented in a non-normal distribution and was transformed using the natural logarithm. In this section it is shown that this transformation was unnecessary and regression was better served using a non-transformed variable.

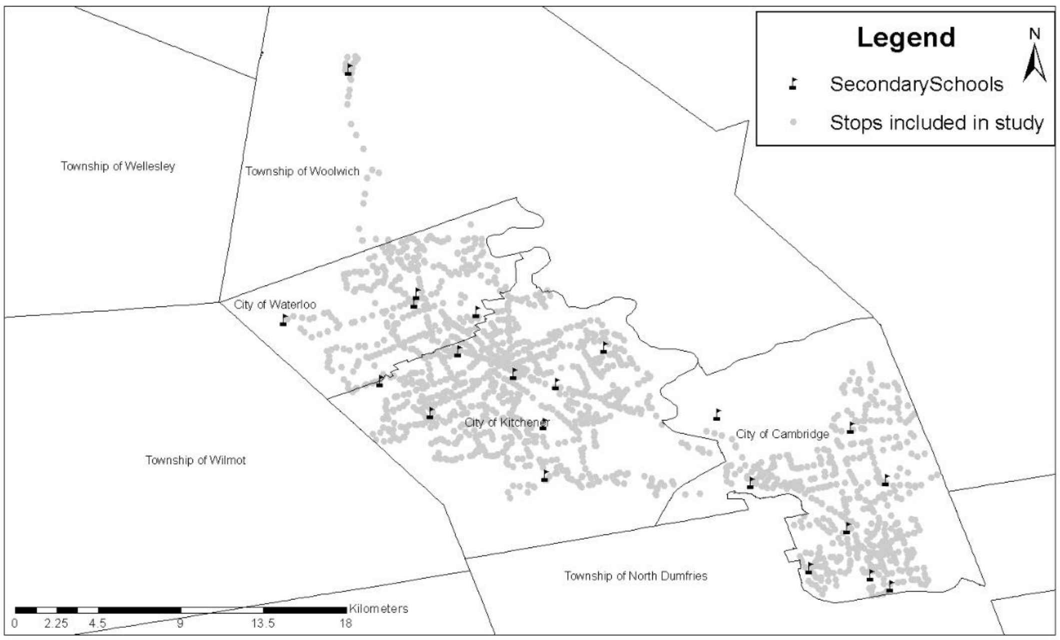
4.2.1.1 Secondary Schools

The role of secondary schools in impacting transit ridership is similar to that of employment and not to be undervalued or dismissed without exploration. This relationship is consistent with other studies which have shown the effect of schools on ridership patterns (Tang & Thakuriah, 2012). For this reason this section is used to inform this exploration of the research despite the fact this variable is not used further.

During preliminary regression analysis using the *One hour peak average* data cases were observed which had a school present within the bus stop 400m buffer and did not align with expected results. The effect of secondary schools on peak ridership is not unexpected as the time 3:00pm-4:00pm may increase end of day school day traffic. In order to determine if there was a correlation between ridership and secondary schools a variable was created.

Using GIS and regional data, the location for all secondary schools was identified and the variable *Secondary school* was created to identify the presence of a school within the buffer (Region of Waterloo, n.d.). This is a point object, discussed in section 3.3, where the variable is coded “1” for with a school and “0” without a school. The locations of secondary schools are identified geographically on the map presented as figure 4.1.

Figure 4.1 Map of secondary schools within the Region of Waterloo



During this analysis it was determined that, with the exception of six stops, the presence of a secondary school does not consistently increase public transportation ridership. The identified inconsistencies may be a consequence of the time of day, as classes are released or other confounding characteristics not examined by this research. For these reasons it was deemed that these cases as are to be removed from this study. This decisions results in a data set of 2478 observation points, and the variable *Secondary schools* is understood to be unnecessary. Appendix B presents regression analysis for three models: one without the *Secondary school* variable and all cases included, one with *Secondary school* variable, and the final one without *Secondary schools* and the identified outlying cases eliminated.

4.2.1.2 Employment Density

As examined in section 3.4.1 the variable *Employment density* showed statistical characteristics which suggested a distribution that was not normal. The requirement of normalcy is understood as important when doing linear regression so a standard transformation, the natural logarithm of the data, was used to normalize the curve. While conducting preliminary analysis it was noticed that this variable in the transformed state was not statistically significant. For these reasons the purpose of the transformation was revisited to ensure the best variable was being used.

The Base Model with the transformed variable *Log employment density* is presented in table 4.2 and that same model, run with the non-transformed variable *Employment density,* is presented in table 4.3. In both models this is done against the *Log one hour peak average boarding and alighting*. These tables show that despite the transformation the adjusted R^2 is static and significance of measuring employment density is reduced. The descriptive statistics of the studentized residual from the non-transformed regression model, table 4.4, shows that the non-transformed variable does not distort the result.

The use of a non-transformed variable is preferred as it decreases the complexity of equation while simplifying the conclusions. For this reason the Base Model will be examined using the non-transformed variable. This solution is reinforced later as the non-transformed variable increases significance when using *Log all day average boarding and alighting* as the dependant variable.

Table 4.2 Regression results: *One hour average* and Base Model with *Log employment density*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2478	*Log one hour average boarding and alighting*		0.277
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		

(Constant)	-1.779	0.138	-12.883	0.000
Level of service	0.239	0.023	10.211	0.000
Transfer location	1.071	0.043	24.855	0.000
Population density	0.000169	0.000024	7.100	0.000
Log Employment density	0.039	0.020	1.914	0.056

Table 4.3 Regression results: *One hour average* and Base Model with *Employment density*

Model type	**Number of observations**	**Response variable**		**Adjusted R^2**
Linear	2478	*Log one hour average boarding and alighting*		0.277
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.571	0.076	-20.722	0.000
Level of service	0.238	0.023	10.148	0.000
Transfer location	1.065	0.044	24.364	0.000
Population density	0.000168	0.000024	7.078	0.000
Employment density	5.240e-005	0.000027	1.974	0.049

Table 4.4 Descriptive statistics: Studentized residual from regression analysis using non-transformed *Employment density* variable

	Skewness		**Kurtosis**	
	Statistic	**Std. Error**	**Statistic**	**Std. Error**
Studentized Residual	0.362	0.49	-0.238	0.98

The role of *Employment Density* in affecting all mode travel has been established in Chapter 2, therefore the reasons for this variable's relative weakness when using *Log one hour average boarding and alighting* should be explained. As stated in section 3.3.3 the dataset used for this variable is oldest data used in this research originating from the 2006 Census. Additionally the geographic distribution of employment in the Region of Waterloo is highly concentrated around the downtowns cores, main routes, and employment lands, exactly the geographical pattern that created the need to the statistical transformation originally.

The above challenges acknowledged it is preferable to continue analysis using *Employment density*. Without employment being considered, the variable *Population density* becomes a dominant trip generator which may bias the results towards neighbourhood characteristics in residential areas. While this does increase the risk of distorting the analysis this will be mitigated with examination of the residuals for non-normal distribution. The role of the intervening variables is to expose inconsistencies in the effect of the predictor variables with the response variable in order to seek a balanced and complete statistical framework.

4.2.2 Base Model and response variable regression models

The response variable based on this research question came in two forms: the average of all day travel boarding and alighting, and the average of one hour peak travel time boarding and alighting at the bus stop level. In both cases this variable was transformed using the natural logarithm as detailed in section 3.4. Table 4.5 and table 4.6 shows the results of the Base Model using the *Log one hour peak average* and the *Log all day average* variables respectively. These regression models were conducted using the data points from the all day dataset to minimize the differences between the two models. It can be seen that the *One hour peak average* has an adjusted R^2 of .279 and the *All day average* an adjusted R^2 of .428 which reflects the foundation from which subsequent analysis will be added to answer this research question.

Table 4.5 Regression results: Base Model with *One hour peak average*

Model type	Number of	Response variable	Adjusted

	observations		R^2
Linear	2456	*Log one hour average boarding and alighting*	0.279

Predictor variables	Unstandardized coefficient		t	Sig.
	β	Standard Error		
(Constant)	-1.567	0.077	-20.539	0.000
Level of service	0.238	0.024	9.937	0.000
Transfer location	1.059	0.044	24.257	0.000
Population density	0.000147	0.000024	7.339	0.000
Employment density	6.067e-005	0.000027	2.244	0.025

Table 4.6 Regression results: Base Model with *All day average*

Model type	Number of observations	Response variable	Adjusted R^2
Linear	2456	*Log all day average boarding and alighting*	0.428

Predictor variables	Unstandardized coefficient		t	Sig.
	β	Standard Error		
(Constant)	-0.850	0.108	-7.884	0.000
Level of service	1.007	0.033	30.200	0.000
Transfer location	1.113	0.061	18.132	0.000
Population density	0.000261	0.000033	7.808	0.000
Employment density	0.00025	0.000038	6.583	0.000

The transit system variables *Level of service* and *Transfer location* indicate the draw of reliability and transfers. With both response variables these variables are significant at over 99%. This is consistent with previous research that system level of service provides a strong draw for users as established is section 2.5. These variables serve the function of identifying bus stops which have higher boarding and alighting not a result of the built environment characteristics but rather of system variables. If either of these system variables did not present as significant, the measurement of the response variable, or another study error, would have to be considered.

The density of the buffer reflects the number of people that can be drawn from a catchment area, this is explained in section 2.3.1. *Population density* is significant to 99% in the case of both response variables. *Employment density* shows a different result - in the case of *One hour peak average* the significance is 95% while in the case of *All day average* 99% significance is achieved. A lower correlation and significance is anticipated. For both these variables it can be associated with the MAUP method of calculating the variable. Specifically for *Employment density* the potential errors in data age are discussed in the previous section.

Despite these limitations, both these variables are significant and crucial to understanding and answering the research question. While other intervening variables could have been selected, these four variables indicate elements that are known to increase ridership without significant risk of correlation with other variables selected as walkability indicators.

4.2.3 Studentized Residual of Base Model

The studentized residual can inform the analysis of this question differently from the other regression characteristics. Examination of those stops which do not react to the intervening variables as expected may inform conclusions about their relationship to the built environment. Chow et al. (2000) state that a residual with an value greater than 3 is considered a significant outlier. Outlier stops are presented in figure 4.2, with individual maps in Appendix C. The *One hour peak average* model identifies six cases which meet this threshold, table 4.7. The *All day average* model identifies 18 stop which meet this threshold, table 4.8. There is only one stop, Stop ID 1278, which is an outlier in both models.

Exploring these 23 stops aims to inform what combination of built environment characteristics have affected ridership and therefore creat the outlier result. De Veaux et al. (2012) state that any significantly outlying case requires exploration. The Base Model includes: the ordinal *variable Level of service* and the nominal variable *Transfer location* and two scale variables *Population density* and *Employment density*. These variables are not a comprehensive list of all variables known to affect transit ridership and therefore this section will aim to understand other causes for the atypical results. To be considered, but not examined in this research, the route may have been designed for a specific role or purpose and abnormally high or low ridership is a result (Guihaire & Hao, 2008). The conclusions in this section are subjective in their interaction with walkability characteristics.

Figure 4.2 Map of outlier cases from Base Model regression analysis using both response variables

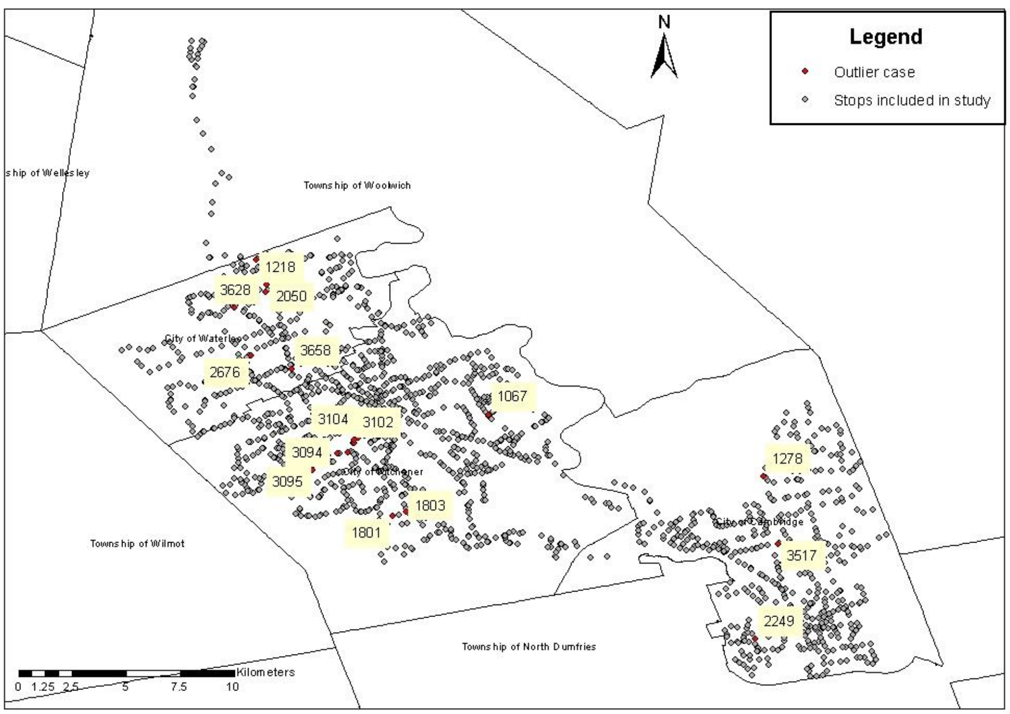


Table 4.7 Outlier cases from Base Model using *One hour peak average*

Model type		Response variable		
Linear		*Log one hour peak average boarding and alighting*		
Stop ID	**Standard Residual**	**Response variable**	**Predicted value**	**Residual**
3517	3.372	2.549	-0.514	3.063
3628	3.004	2.506	-0.222	2.728
3658	3.367	2.57	-0.478	3.058
2249	3.302	2.30	-0.697	3.000
1278	3.192	2.512	-1.041	3.554
1218	3.129	1.819	-1.022	2.84

Table 4.8 Outlier cases from Base Model using *All day average*

Model type		Response variable		
Linear		*Log all day average boarding and alighting*		
Stop ID	**Standard residual**	**Response variable**	**Predicted value**	**Residual**
2676	3.074	7.452	3.524	3.927
1278	3.016	5.205	1.352	3.853
2048	-3.116	-2.302	1.679	-3.981
2046	-3.076	-2.302	1.628	-3.930
2050	-3.091	-2.150	3.047	-5.198
3070	-3.188	-2.171	1.901	-4.073
3107	-3.714	-1.916	2.829	-4.746
1803	-3.198	-1.732	2.354	-4.086
1067	-3.181	-1.496	2.567	-4.064
3094	-3.459	-1.750	2.669	-4.419
3101	-3.694	-2.052	2.667	-4.719
3104	-3.181	-1.265	2.798	-4.064

3102	-3.227	-1.475	2.647	-4.123
3098	-3.435	-1.807	2.581	-4.389
3099	-3.581	-2.035	2.539	-4.575
1801	-3.598	-2.302	2.294	-4.597
3095	-3.471	-1.898	2.537	-4.435

Stop 1067 is an outlier when using the response variable *All day average boarding and alighting* in the regression model. This stop has low ridership levels with an average value of 0.12 boarding and alighting per day while the densities present just below the mean. The area around this stop is a suburban design and the low traffic may be attributed to the connectivity of the built environment for pedestrians, as these designs can increase walking distances. Less than 200 meters north - east this bus services two additional stops, stop ID 1219 and 1217, which are not outliers with an average ridership of 2.06 and 0.35 boarding and alighting respectively. It is possible that these ridership levels relate to the increased access and connectivity to those stops. While stop 1067 shows similar built environment characteristics the street design favours access to these two other stops which people may therefore favour.

Stop 1218 is an outlier when using the *One hour average boarding and alighting* response variable. This stop has high ridership while being located in an industrial area with low levels of both employment and population density. This stop highlights the difference between the *one hour average* ridership and *all day average* ridership. It is possible one of the employers ends a shift either during or just prior to the selected hour and as a result a disproportionate level of ridership is achieved. While other walkability characteristics may be considered, like sidewalk connectivity, this reason seems most probable.

Stop 1278 is an outlier in both the *All day average* and *One hour peak average* models, as they have high boarding and alighting in both. When boarding and alighting are examined separately this is attributed to a high number of boarding, 112 average daily boarding and 70 average daily alighting. Stop 1275 only 90 meters away also has high daily boarding and alighting. It is possible that these stops serve the same destination or are an unofficial transfer location. The walkability characteristics in this area seem overall low, suggested by the low street and sidewalk connectivity in the area. Due to the method of measurement, the 400 meter buffer, stop 1278 measures more of the land use from the multilane Hespler Road to the west and the on ramps to highway 401 than stop 1275. This results in lower population and employment density due to the large amount of undeveloped public land within the 400 meter buffer. Access to these major roadways may also promote multi-modal access, for example through carpooling, which is beyond the scope of this research.

Stops 1801 and 1803 are located just beyond the 400 meter buffer of one another, these stops are outliers in the *All day average* model with a ridership value of 0. It is observed that these stops have almost no connectivity beyond the lands they abut. The population density around these stops is measured higher as a result of the single family residential developments located to the north but those users have no access to these stops. Stop 1774, located in the north-east has a trail connecting to the interior of the suburban community, and has significant *All day average boarding and alighting* at 57.51. Stop 1802 directly between stops 1803 and 1801 has only marginally higher than 0 as ridership at 0.79 average all day boarding and alighting. Clearly the limited users benefit from the connectivity presented by 1802 and not 1803 or 1801.

Stops 2046, 2048 and 2050 are all outliers in the *All day average boarding and alighting* model. These stops each have 0 boarding or alighting these stops have similar land use and connectivity characteristics to the earlier discussed 1801 and 1803. Despite having more employment density, denoting higher potential, the location of the separated highway to the east likely draws more drivers or car commuters from further distances. This car connectivity would apply in several areas throughout the Region of Waterloo. The location here has a compounded issue of high level of service characteristics within walking distance on the flanking roads to the north and east which may draw transit riders to other routes. This stop exemplifies a challenge in using *all day averages*, where the *one hour peak* had many stops with no ridership, stops of this nature did not present as stark a contrast and are investigated with other low ridership stops. As is the case with *All day average* stops, which generate zero boarding and alighting and therefore stand out further.

Stop 2249 is an outlier of *One hour peak average boarding and alighting*. This stop has very high average traffic at 9.902, it is not considered an outlier in the case of *All day averages* thus suggesting a disproportionate number of trips generated during the observation hour. This stop likely generates riders from the residential community to the south where the residents may have created an impromptu access where the road or public network does not provide connectivity. This case recognises the ability of users to bypass the barriers of reduced pedestrian access.

Stop 2676 is located at the south access to University of Waterloo and is an outlier of All day average boarding and alighting. This stop generates large amounts of student traffic servicing the university, while the south half of the buffer is almost exclusively parking lot. The GRT system identifies Davis Centre stops 1123 and 3699 as a key transfer locations, generating 8557.49 all day average boarding and alighting, while this stop generates 1724.60. Therefore stop 2676 does inform the question of walkability as it speaks to the access and convenience to the south end of campus, which has good pedestrian connections on site.

Stops 3070, 3107 and 3108 are all outliers in the *All day average* model, located with access to a residential community. The street has very limited pedestrian access the east which is interfered with by a trail and water reservoir. These significant built environment barriers means the potential population that could access these three stops becomes limited to one row of single family dwellings. This highlights the challenge of access to bus stops where the built environment is defined by restricted corridors and topography. These stops highlight the importance of integrating pedestrian access to the population areas and the transit stop locations.

Stops 3094, 3095, 3098, 3099, 3101, 3102 and 3104 are outliers on the *All day average boarding and alighting* model. Each of these stops are serviced by one bus route through a low density residential neighbourhood. The urban form here reflects the suburban form identified in the article by Gassaway (1992). These urban forms have been known to create barriers to public transportation use as they do not provide pedestrian connections and sometimes lack pedestrian infrastructure. As was examined in Chapter 2 the urban design has been moving away from the conventional suburbs for just these reasons. The patterns identified in these stops and in this urban area exemplify the issues created by a low density suburban design.

Stops 3628, 3658 and 3517 are all outliers of the *One hour peak average boarding and alighting* model. These stops each have very high ridership and despite the high built environment indicators in both level of service and density, the results are beyond the predicted value of the regression model. These stops are not outliers when using the *All day average* as the response variable. This result could be a symptom of nearby destinations or route connections. For example stop 3658 is in a downtown with high mixed use and walkability characteristics. In those environments bus users would be able to walk to the buses once completed work and board directly, prospectively making this peak time earlier than other stops in the system which may rely on a transfer.

This examination of outliers have revealed how walkability variables like connectivity and mixed use can be used to understand several of the stops which are outliers in the Base Model. The conclusions here cannot be conclusive as this analysis does not account for all potential variables affecting ridership at each stop. The analysis does provide strong anecdotal support for the impact of pedestrian characteristics which may affect trip ridership. These conclusions are especially informative in cases where ridership below regression model expectations level, such as stops 2249 or 3094. The examination of these outliers helped inform the use of *Log all day average boarding and alighting* versus *Log one hour peak average boarding and alighting* as the response variable in future models. As the outlier cases have been used to inform the connectivity and pedestrian environment characteristics they will not be removed from the dataset.

4.3 Predictor Variables: Models 1, 2 and 3

The purpose of this research is to determine the relationship between public transportation ridership and pedestrian infrastructure. The previously established Base Model has been established in order to ensure the statistical relationship is consistent with conclusions from prior academic research. This section will take care to ensure that variables which may have multi-collinearity are not included in the same model, that can create difficulty in determining the relationship between variables when variables are found to be collinear (De Veaux et al., 2012). To inform this, a Pearson's R multi-collinearity matrix has been included as Appendix D. There it can be seen that this research has some variables which have a risk of multi-collinearity. This is especially true with the sidewalk variables and population density, which present a risk of multi-colliniarity.As not all three of these will be included in the same model, this will be acknowledged in conclusions and analysis. This will be achieved through running regression by testing variables that examine similar built environment characteristics separately. The strength of the correlation will be explored through significance of the variable and the adjusted R^2.

4.3.1 Model 1: Entropy

The variable *Entropy* was established using the method presented in Saelens et al. (2003) as explained in section 3.3.2. This variable was selected as it is the only variable to examine the land use characteristics around the bus stop and as a measurement of land-use diversity. The results of this model are presented in table 4.9, using *One hour peak average* as the response, and table 4.10, using *all day data average* as the response variables.

Table 4.9 Regression results: Model 1 using *One hour peak average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2478	*Log one hour average boarding and alighting*		0.278
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	B	**Standard Error**		
(Constant)	-1.651	0.090	-18.273	0.000
Level of service	0.234	0.024	9.931	0.000
Transfer location	1.059	0.044	24.151	0.000
Population density	0.000172	0.000024	7.224	0.000
Employment density	4.535e-005	0.000027	1.686	0.092
Entropy	0.151	0.093	1.624	0.104

Table 4.10 Regression results: Model 1 using *All day average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2456	*Log all day average boarding and alighting*		0.432
Predictor variables	**Unstandardized coefficient**		**t**	**Sig.**
	B	**Standard Error**		
(Constant)	-1.132	0.127	-8.892	0.000
Level of service	0.992	0.033	29.684	0.000

Transfer location	1.092	0.061	17.784	0.000
Population density	0.000277	0.000034	8.261	0.000
Employment density	0.000225	0.000038	5.864	0.000
Entropy	0.540	0.131	4.132	0.000

In the model using *One hour peak average boarding and alighting* it can be seen that the *Entropy* variable did not achieve 90% significance. The impact on the adjusted R^2 is insignificant with an increase of only 0.001 while this variable causes the significance of *Employment density* to drop to 90% from 95%. From these results it can be seen that the variable *Entropy* reduces the viability of the Base Model and is statistically insignificant in this model. In the model using the *All day travel average boarding and alighting* response variable, it can be seen that *Entropy* variable is significant to the 99% level. The variable also increases the adjusted R^2, from .428 to .432, with no reduction in the significance of the Base Model variables.

Some explanations can be developed in exploring this inconsistent result between response variables. *One hour peak* data may not equally reflect destinations and origins of journeys within the GRT system. The result that *Employment density* reduced significance when the *Entropy* variable was introduced suggests that these variables may be conflicting. During this one hour peak time employment trips may not be a significant trip origin or destination throughout the system. This challenges the use of *One hour peak* data in gathering overall activity. The role of *Employment density* as a known predictor of trip generation and the measurement of *Entropy* achieved here gives support to a decision to use the *All day average* as the response variable.

The variable *Entropy* is supported by other research which has linked this variable to travel behaviour choices using modes other than cars (Cervero, 2002). The objective of introducing the variable *Entropy* is as a measurement of diversity in the built environment around each bus stop. As explained previously, diversity is a strong indicator of a built environment which can both draw to a destination while simultaneously providing and internal population to support activity.

In addressing specifically the question the regression results here establish a relationship between a walkable environment and a public transportation ridership. Previous research has shown diversity has been shown to impact various mode choices, as discussed in section 2.3.2. The correlation presented in this model between public transportation ridership and the built environment is both significant and influential.

4.3.2 Model 2: Pedestrian design

Three variables have been included in this study which provide a measurement for pedestrian infrastructure: *Sidewalk length*, *Ratio* and *Intersection count*. The methods for these variables are presented in section 3.3.2. Model 2 aims to account for the design of the pedestrian infrastructure in the built environment that facilitate walkability within the bus. This analyses continue to use both response variables; *One hour peak average* and *All day average boarding and alighting*. The variable *Entropy*, discussed above, is included only in the case of the *All day average* response variable.

These pedestrian environment variables show significant a risk of multi-collinearity using Pearsons R correlation, table 4.11. As each of these variables are measuring the same fundamental infrastructure this relationship between variables is expected. The intent of this section is to determine which variable is the most significant in exploring the research question. While each of these variables measure a similar aspect of pedestrian infrastructure, the dissimilar methods of operationalization inform conclusions about this research differently.

Table 4.11 Pearsons R correlation for pedestrian design variables

Variables being tested	Pearsons R
Sidewalk length and *Ratio*	0.8597048

Sidewalk length and *Intersection density*	0.7127389
Ratio and *Intersection density*	0.402885

4.3.2.1 Sidewalk length

The variable *Sidewalk length* is a measurement of the length of sidewalk present within the 400 meter buffer of a bus stop. This measurement attempts to quantify the amount of pedestrian infrastructure within the 400 meter buffer. The hypothesis is that the higher the length of sidewalk within the buffer the more pedestrian friendly the community design. Table 4.12 and table 4.13 present the results of the regression model including the variable *Sidewalk length* using the *Log one hour peak average* and *Log all day average* as response variables respectively. It can be seen in both cases that this variable does not achieve significance and as a result is not explanatory.

Table 4.12 Regression results: Model 2 *Sidewalk length* using *One hour peak average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2478	*Log one hour average boarding and alighting*		0.277
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.570	0.076	-20.623	0.000
Level of service	0.238	0.023	10.142	0.000
Transfer location	1.064	0.044	24.253	0.000
Population density	0.000172	0.000035	4.972	0.000
Employment density	5.404e-005	0.000028	1.923	0.055
Sidewalk Length	-1.368e-006	0.000008	-0.178	0.859

Table 4.13 Regression results: Model 2 *Sidewalk Length* using *All day average*

Model type	Number of observations	Response variable	Adjusted R^2
Linear	2456	*Log all day average boarding and alighting*	0.432

Predictor variables	Unstandardized coefficient		T	Sig.
	β	Standard Error		
(Constant)	-1.142	0.128	-8.934	0.000
Level of service	0.990	0.033	29.582	0.000
Transfer location	1.097	0.062	17.793	0.000
Population density	0.000246	0.000049	5.048	0.000
Employment density	0.000213	0.000041	5.250	0.000
Entropy	0.545	0.131	4.164	0.000
Sidewalk length	9.378e-006	0.000011	0.872	0.383

4.3.2.2 Intersection density

The variable *Intersection density* is a common measurement found in academic material as a method to quantify pedestrian design (Ewing & Cervero, 2010; Guo & Ferreira Jr, 2008). This variable indicates the permeability or ability of a pedestrian to easily access different parcels of the area without walking long distances. While this variable does not measure pedestrian infrastructure directly, as is the case with the variable *Sidewalk length*, it does indicate pedestrian access. This may lead to a conclusions that pedestrians do not require infrastructure to access public transit as much as good connectivity within a the 400 meter buffer.

Table 4.14 and table 4.15 indicate the results of the regression model using *Log one hour peak average* and *Log all day average* as the response variables respectively. It can be seen that the variable *Intersection density* is significant to 95% in the *one hour peak* model and 90% in the *all day average* model. This significance is not surprising considering the use of this variable in previous research; however, in both models this variable has only a minor effect on the adjusted R^2.

This minor effect may relate to the conclusion by Guo and Ferreira Jr (2008) that *Intersection density* has a dual effect. The first effect is the above stated benefit of connectivity and permeability.

The second possible effect is that intersections create a barrier for pedestrians as they can increase the conflict points between pedestrian and cars. This barrier is a potential explanation for the results of this model that remains consistent with previous academic studies identified by Ewing and Cervero (2010). Research focusing on permeability could additionally be addressed through the use of network analysis as this research does not measure that aspect of connectivity.

Table 4.14 Regression results: Model 2 *Intersection density* using *One hour peak average*

<table>
<tr><th>Model type</th><th>Number of observations</th><th colspan="2">Response variable</th><th>Adjusted R^2</th></tr>
<tr><td>Linear</td><td>2478</td><td colspan="2">Log one hour average boarding and alighting</td><td>0.279</td></tr>
<tr><th rowspan="2">Predictor variables</th><th colspan="2">Unstandardized coefficient</th><th rowspan="2">T</th><th rowspan="2">Sig.</th></tr>
<tr><th>β</th><th>Standard Error</th></tr>
<tr><td>(Constant)</td><td>-1.529</td><td>0.078</td><td>-19.715</td><td>0.000</td></tr>
<tr><td>Level of service</td><td>0.239</td><td>0.023</td><td>10.203</td><td>0.000</td></tr>
<tr><td>Transfer location</td><td>1.083</td><td>0.044</td><td>24.463</td><td>0.000</td></tr>
<tr><td>Population density</td><td>0.000205</td><td>0.000028</td><td>7.327</td><td>0.000</td></tr>
<tr><td>Employment density</td><td>7.453E-005</td><td>0.000028</td><td>2.665</td><td>0.008</td></tr>
<tr><td>Intersection</td><td>-0.005</td><td>0.002</td><td>-2.493</td><td>0.013</td></tr>
</table>

density				

Table 4.15 Regression results: Model 2 *Intersection density* using *All day average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2456	*Log all day average boarding and alighting*		0.432
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.079	0.131	-8.264	0.000
Level of service	0.994	0.033	29.738	0.000
Transfer location	1.112	0.062	17.830	0.000
Population density	0.000314	0.000039	7.987	0.000
Employment density	0.000248	0.000040	6.134	0.000
Entropy	0.520	0.131	3.964	0.000
Intersection density	0.293	0.072	-1.798	0.072

4.3.2.3 Ratio

The variable *Ratio* is measurement of sidewalk length and road length within the 400 meter buffer of a bus stop. This variable quantifies the level of priority between pedestrian infrastructure and traffic. Table 4.16 and table 4.17 show the results of these models, using *Log one hour peak average* and *Log all day average* as the response variable respectively. This variable showed significance to 99% in both models and an increase in the adjusted R^2. This effect was most pronounced in the case of the response variable *Log one hour peak average boarding and alighting*.

In the *One hour peak average* model *Ratio* also has the effect of reducing the significance of Employment density to less than 90%. As a Base Model variable *Employment density* is included to indicate continuity with previous research and therefore this effect requires consideration. Discussed in section 4.2.1.2 the *Employment density* data has previously identified issues and is not as quantified as the other variables included in the Base Model. There is, however, no clear explanation for this variable's loss in significance.

In the model using *All day average*, the effect of *Ratio* is to increase the adjust R^2 only marginally by 0.003. This result has the effect of explaining very little but this does not reduce the significance of the variables from the Base Model. Unlike *Intersection density, Ratio* indicates the imbalance between travel modes and measures how much of the public space is prioritized for pedestrian use. The road is public space and the variable *Ratio* may indicate how much pedestrian traffic is invited in those neighborhoods by the community designers and planners. This variable while statistically significant does not inform conclusions about the effect of the built environment on public transportation at this stage.

Table 4.16 Regression results: Model 2 *Ratio* using *One hour peak average*

Model type	**Number of observations**	**Response variable**		**Adjusted R^2**
Linear	2478	*Log one hour average boarding and alighting*		0.281
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.662	0.079	-20.956	0.000
Level of service	0.227	0.024	9.628	0.000
Transfer location	1.091	0.044	24.722	0.000
Population density	9.260e-005	0.000031	3.001	0.003
Employment density	4.026e-005	0.000027	1.510	0.131
Ratio	0.192	0.051	3.792	0.000

Table 4.17 Regression results: Model 2 *Ratio* using *All day average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2456	*Log all day average boarding and alighting*		0.436
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.210	0.128	-9.428	0.000
Level of service	0.977	0.034	29.738	0.000
Transfer location	1.137	0.062	17.830	0.000
Population density	0.000158	0.000044	7.987	0.000
Employment density	0.000210	0.000038	6.134	0.000
Entropy	0.434	0.133	3.266	0.001
Ratio	0.293	0.072	4.049	0.000

4.3.2.4 Design Conclusion

The variables *Sidewalk length*, *Intersection density* and *Ratio* each provide a different measurement for the pedestrian infrastructure within the 400m buffer of the bus stop. These variables aim to quantify the pedestrian experience, thus informing a measurement of walkability. The objective of this section is to determine the most appropriate variable to use in further examination to answer the research questions presented. To determine the best design variable three alternative methods for quantifying pedestrian built environment and walkable design have been employed.

In answering this research question it is clear that the variable *Sidewalk length* is not significant or explanatory. The variable *Intersection density* is significant at 90% with an insignificant effect on the adjusted R^2. The variable *Ratio* is significant at 99% with a marginally greater result on the adjusted R^2. This suggests that the variable *Ratio* developed in Cervero (2002) is the most appropriate for further use in this study as it is the most informative with public transportation ridership. It is recognized that while the adjusted R^2 has not moved significantly, the β scores have changed which reflects the effect these variables are having on one another as suggested by the multi-collinearity results seen in Appendix D. The relationship between these design variables and the variable *Population density* in particular reflect the complexity in quantifying the built environment. This problem limits the ability to make conclusions based specifically on the design variables as they relate to public transportation ridership.

The effect of this conclusion is to suggest that the pedestrian environment is most appropriately measured as the public space balance between pedestrians and other modes of transportation, rather than the presence of sidewalks or overall connectivity of design. Considering a bus requires access to roads, therefore a pedestrian must have access to the road as well to access the bus. The variables *Sidewalk length* and *Intersection density* may show different results measuring public transportation by rail which does not have the same road dependence. This variable has characteristics of the pedestrian design and infrastructure, which will directly informs conclusions about the relationship between walkability and public transportation ridership.

4.3.3 Model 3: Log traffic signal

The variable *Log traffic signal* is a method of indicating pedestrian safety as traffic signals create a safe point of interaction between vehicles and the pedestrian. The collation of this variable is explained in section 3.3.2 and the required transformation explained in section 3.4.1. The results of this model are presented in table 4.18 and table 4.19 using *Log one hour peak average* and *Log all day average* as the response variables respectively. The multi-colliniarity results presented in Appendix D show that this variable presents a multi-collinearity risk with several of the variables across this study.

Seen in these results the variable *Log traffic signal* is significant at 99% in both models. In the case of *one hour peak average* it had the additional effect of reducing the significance of *Employment density* below 50%. While *Employment density* is also reduced in the *all day average* model it is still significant to 95%. The adjusted R^2 in both cases increased slightly, which may be a result of the large number of data points included in this study. Overall this variable does not indicate a strong effect on public transportation ridership and negatively impacts the Base Model variable. For these reasons this variable will be omitted in further models.

Table 4.18 Regression results: Model 3 using *One hour peak average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2478	*Log one hour average boarding and alighting*		0.286
Predictor variables	**Unstandardized coefficient**		**T**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.621	0.080	-20.168	0.000
Level of service	0.2045169	0.02434163	8.40194	0.000
Transfer location	1.056987	0.0446953	23.64872	0.000
Population density	9.023201e-005	3.100552e-005	2.910192	0.003
Employment density	-2.230152e-005	3.318688e-005	-0.6719979	0.5016519
Ratio	0.161	0.051	3.167	0.001
Log traffic signal	0.143	0.039	3.65	0.000

Table 4.19 Regression results: Model 3 using *All day average*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2456	*Log all day average boarding and alighting*		0.441
Predictor variables	**Unstandardized coefficient**		**t**	**Sig.**
	β	**Standard Error**		

(Constant)	-1.063	0.130	-8.123	0.000
Level of service	0.942	0.034	27.685	0.000
Transfer location	1.088	0.0626	17.370	0.000
Population density	0.000132	4.458e-005	2.951	0.003
Employment density	7.747e-005	4.639e-005	1.670	0.095
Entropy	0.286	0.135	2.116	0.034
Ratio	0.251	0.072	3.466	0.000
Log traffic signal	0.283	0.056	5.047	0.000

4.4 Conclusion

In this chapter several linear regression models were run in order to determine which variables were most appropriate in understanding the relationship between walkability, the built environment, and public transportation ridership. The variables presented in table 4.20 will be employed in spatial regression modelling in the next chapter. An examination of outlier results from linear regression using the Base Model also served to inform the conversation about walkability and the built environment.

In the spatial regression stage of this research only one response variable will be used. Based on the results from this chapter this research is best facilitated by using the response variable *Log all day average boarding and alighting*. This, in conclusion, is most significantly related to the correlation between the One hour average and the variable *Employment density* and *Entropy*.

Employment density is recognized as an important trip generator based on previous research (G. Thompson et al., 2012). In the case of the *One hour peak average* models *Employment density* was the first Base Model variable to lose statistical significance. This result suggests that this response variable was not capturing employment riders. The variable *Employment density* was significant when using the response variable *All day average boarding and alighting*.

Entropy was selected as a variable to test land use diversity based on availability of data. This variable was not significant when using the response variable *One hour peak average* and highly significant when using the response variable *All day average*. This again suggests that the *One hour peak average* variable is not capturing enough of the trip origins and destinations to achieve an accurate snapshot of the built environment as they relate to GRT riders.

This chapter served to answer some questions about the appropriate variable to measure the pedestrian environment or design. Models 2 and 3 concluded that the variable *Ratio* and *Log traffic signal* were statistically significant while their impact on the adjusted R^2 was minimal. It is acknowledged that the significance could result from the high number of cases included in the study and for that reason the significance of this conclusion is minimized. The variable *Ratio* is included as it best indicates design of those variables tested for that purpose. The variable *Log traffic signals* requires statistical transformation and serves less value in answering this research question and for that reason is omitted.

At this point, the results suggest that areas with pedestrian infrastructure are statistically significant in their correlation with public transit ridership. *Intersection density* was significant over 90% using both response variables, suggesting the importance of permeability and access to the public transportation stop. The variable *Ratio* reflects the need for the permeability around a bus stop to service the pedestrian equally as a mode of transportation. The variables selected here did not cause a strong change to the adjusted R^2; but they did show that depending on how the pedestrian environment is measured there is a statistically significant correlation with public transportation ridership at the bus stop level.

With the selection of the most appropriate pedestrian infrastructure variable this model includes a representative variable from each of the 3D's as defined by Cervero and Kockelman (1997). Density is represented throughout the study through the use of *Population density* and *Employment density* as Base Model variables. Diversity is represented through the use of the variable *Entropy* which was only significant in the case of *All day average boarding and alighting*. The variable *Ratio* has been selected to represent Design. These variables will serve to provide a statistical foundation for spatial regression going forward.

Table 4.20 Variable included in spatial regression as a result of linear regression models

Name	Measuring	Included in spatial regression model
Log one hour average boarding and alighting	Natural logarithm of average boarding and alighting during one hour peak period	Omitted
Log all day average boarding and alighting	Natural logarithm of average boarding and alighting from all day travel	Included
Population density	Count of people within bus stop buffer as MAUP from census tract	Included
Transfer location	GRT designation of terminals	Included
Level of Service	headway of all buses stopping at bus stop	Included
Log employment density	Natural logarithm of the count of place of work as MAUP problem from census tract	Omitted
Employment density	count of place of work as MAUP problem from census tract	Included
Entropy	Mixed of land uses within bus stop buffer	Included
Intersection count	The number of intersections within buffer of bus stop	Omitted
Sidewalk length	Length of sidewalk within buffer of bus stop	Omitted

Ratio	Ratio of sidewalk and road length within bus stop buffer	Included
Log traffic signal	Natural logarithm of count of traffic signals within bus stop buffer	Omitted

Chapter 5
Results and discussion: Spatial regression

5.1 Introduction

As shown in Chapter 4, the built environment has the ability to influence public transportation ridership. The intent of this research is to answer the following questions:

- How does the pedestrian environment / walkability correlate with public transit ridership?
  - What is the most appropriate way to measure pedestrian design, as it relates to walkability, and what is its correlation with transit ridership?
  - What walkability / built environment characteristics correlate best with transit ridership?
  - In what way is answering this question informed through the use of linear regression and spatial regression models?

To answer this question the variables listed in table 5.1 were selected from previous research, as detailed in Chapter 2, to test their correlation with public transportation ridership. Through testing the variables in several linear regression models, as detailed in Chapter 4, six variables and one response variable were selected that showed the strongest correlation. The variables included in this study were selected based on significance and impact on R^2. The method of variable selection for use in spatial regression models was informed by Cardozo et al. (2012).

This chapter will approach the spatial regression analysis by first establishing the methods framework required including identification of spatial regression model type and weight variables. Subsequently, this chapter will be exploring the walkability variables and their spatial interaction with ridership levels on the Grand River Transit (GRT) in the Region of Waterloo. This portion of the research is limited to use of spatial regression models, while these techniques present opportunities for further research such as clustering it is considered beyond the scope of this research.

Table 5.1 Variable definitions and role in spatial regression model

Name	Measuring	Inclusion in spatial regression model	Variable role
Log one hour average boarding and alighting	Natural logarithm of average boarding and alighting from one hour peak travel time	Omitted	
Log all day average boarding and alighting	Natural logarithm of average boarding and alighting from all day travel	Included	Response variable
Population density	Count of people within bus stop buffer as MAUP from census tract	Included	Base Model, Density indicator
Transfer location	GRT designation of terminals	Included	Base Model
Level of Service	headway of all buses stopping at bus stop	Included	Base Model
Log employment density	Natural logarithm of the count of place of work as MAUP problem from census tract	Omitted	
Employment density	count of place of work as MAUP problem from census tract	Included	Base Model, Density indicator
Entropy	Mixed of land uses within bus stop buffer	Included	Model 1, Diversity indicator
Intersection count	The number of intersections within buffer of bus stop	Omitted	
Sidewalk length	Length of sidewalk within buffer of bus stop	Omitted	
Ratio	Ratio of sidewalk and road within bus stop buffer	Included	Model 2, Design

			indicator
Log traffic signal	Natural logarithm of count of traffic signals within bus stop buffer	Omitted	

5.2 Methods framework

As discussed in Chapter 3, the objective of spatial regression is to take into consideration spatial influences and the impact of distance between data points. This method can be achieved through the use of two different measurement types: spatial lag and spatial error (Ward & Gleditsch, 2008). Spatial lag is a model which includes distance between two data points, in the case of this research the bus stop location, as a variable. This method assumes that the neighbouring locations are not independent of one another and have a similar relationship with the built environment.

Spatial error is used when the distance between observation points is considered a nuisance (Ward & Gleditsch, 2008), a variable is therefore added which accounts for the spatial effect on the predictor and response variables. Spatial error is recognized as generally less useful when conducting social science research as the indication of spatial error cannot necessarily lead to a conclusion about the origin of the error (Ward & Gleditsch, 2008). This potential error is amplified this research as the Base Model knowingly does not include variables which are known to affect public transportation ridership, for example auto-ownership and income level. For these reasons spatial lag is the only spatial regression model to be used in exploring this research.

The spatial weight is created using a threshold distance between bus stop centroids. This distance was determined through testing using the results of linear regression with the Base Model variables and the test for spatial autocorrelation Moran's I. The publications by Anselin (1995) and Ward and Gleditsch (2008) discuss the calculation of Moran's I and its use as an indicator of spatial autocorrelation in detail. Moran’s I provides a score between -1 and +1, where -1 indicates completely random distribution and +1 indicates a non-random distribution.

Table 5.2 shows the results of Moran’s I for several weight variable threshold distance. The distance 557 meters was specifically selected as it is the minimum distance to ensure all stops have a minimum of one neighbouring stop within the threshold distance, other distances were selected through testing. All the resulting Moran’s I scores indicate a non-random distribution and therefore it is understood that spatial autocorrelation exists. From these results it can be observed that the spatial impact increases as the threshold distance decreases. This trend plateaus once the threshold distance is reduced to 200 meters. The effect of these various weight files was further examined using various threshold distances in spatial lag regression models using only the Base Model variables and is included as Appendix E. This testing developed a viable weight file for use with the spatial lag regression model to include walkability indicators examined in the next section.

Table 5.2 Diagnostics of spatial dependence using Moran’s I

Threshold Distance in meters	Moran’s I	Value	Significance
100	0.563841	25.0808954	0.000
200	0.564252	28.7438459	0.000
400	0.439156	37.0730226	0.000
557	0.358688	41.3466099	0.000
800	0.269496	43.6768406	0.000

5.3 Spatial Lag Walkability model

The spatial regression model will include the response variable *Log all day average boarding and alighting*, intervening variables; *Level of service*, *Transfer location*, *Population density* and *Employment density*, and the variables which quantify the pedestrian environment; *Ratio* and *Entropy*. All these variables are defined in table 5.1 above. For reference purposes the results of a spatially lagged regression model using only the Base Model variables is presented as table 5.3.

Table 5.4 shows the results of the complete spatial regression model including the selected walkability variables, it can be seen that both these models have an R^2 of 0.63. Where the spatial lag regression model including only Base Model variables have *Employment density* significant at 90% it is entirely statistically insignificant in the walkability model. The variable *Entropy* also fails to be statistically significant in this model, while the variable *Ratio* is significant at the 95% level. The reduction in significance of *Population density* between the Base Model and walkability variables is also an indication of the unreliability of the latter.

These results lead to the conclusion that despite the clear indication of spatial autocorrelation, the walkability variables do not explain more of public transportation ridership when a spatial lag regression model is used. The resulting increase in R^2 between the linear regression model and spatial regression model shows that the use of spatial lag regression informs research on public transportation ridership. The comparison of these models, however, clearly indicate that when walkability is included a spatially lag regression model it is not more explanatory of the relationship.

Table 5.3 Spatial lag regression results: Base Model

Model type	**Number of observations**	**Response variable**	**Weight variable threshold distance in meters**	**R^2**
Spatial	2456	*Log all day average boarding and alighting*	200	0.632
Predictor variables	**Unstandardized coefficient**		**z-score**	**Sig.**
	β	**Standard Error**		
Weight Response variable	0.4335866	0.01290281	33.60404	0.000
(Constant)	-0.62946	0.08651896	-7.2754	0.000
Level of service	0.6672872	0.02835879	23.53018	0.000
Transfer location	0.7072966	0.05123063	13.80613	0.000
Population density	0.0001393892	2.708817e-005	5.14576	0.000

Employment density	5.78078e-005	3.078892e-005	1.877552	0.060

Table 5.4 Spatial lag regression results: Walkability variables

Model type	Number of observations	Response variable	Weight variable threshold distance in meters	R^2
Spatial	2456	*Log* all day average boarding and alighting	200	0.632

Predictor variables	Unstandardized coefficient		z-score	Sig.
	β	Standard Error		
Weight Response variable	0.4293349	0.01301996	32.97513	0.000
(Constant)	-0.764694	0.1039153	-7.35882	0.000
Level of service	0.6585565	0.02852699	23.08539	0.000
Transfer location	0.7261939	0.05226605	13.89418	0.000
Population density	8.777443e-005	3.594666e-005	2.441797	0.014
Employment density	4.371797e-005	3.122682e-005	0.1615094	1.400
Ratio	0.1433739	0.05853983	2.449168	0.014
Entropy	0.12982	0.1074232	1.208491	0.226

5.4 Conclusion

In exploring the sub question:

- In what way is answering this question informed through the use of linear and spatial regression models?

This chapter has shown that spatially lagged regression does inform public transportation research, especially in the case of known influence variables, included here in the Base Model. This effect reduces the importance of walkability and built environment characteristics as they measured against public transportation ridership. Public transportation ridership research which is exploring lower impact built environment characteristics should not rely on spatial regression as its effects can obscure results. This obscuring effect may be related to the level of analysis, at the individual bus stop, or other confounding features not identified within this research.

The conclusions from this chapter are consistent with previous research by Chow et al. (2000) into public transportation using linear and spatial regression modelling. That study concludes that spatial regression models:

> "…indicate that some variables are nonstationary. Their significance and influence vary by location, as indicated by the magnitude of their coefficients, which varies across space. An unexpected local sign of a variable may be an indication of multi-collinearity or insignificance or irrelevance of the variable at that location, which points to future research to explore possible different model structures within a geographic area as well as the need to develop better tools for model development." (Chow et al., 2000, p. 111)

That study examines specifically the role of home - work trips and established that spatial regression shows a strong improvement over linear regression in public transit ridership. In the conclusion of that study it is noted that there is a need to conduct this analysis at the bus stop level to determine that if the conclusions hold, the results presented in this research may inform that statement.

From these results it can be concluded that the spatial influence on proximal stops is more significant than the influence of the design or diversity characteristics selected for this study. The spatial influence may not have been accurate as the density variables were both negatively impacted by the spatial lag model. This research showed a strong spatial autocorrelation and the spatial lag model clearly revealed that an accurate threshold weight established a higher statistical correlation. In examining walkability or pedestrian variable however, the spatial significance may mask the influence of other variables.

Chapter 6
Recommendations and conclusion

6.1 Summary and recommendations

This thesis intends on informing research about how the built environment affect public transportation ridership through analysing the pedestrian environment at the individual bus stop. This was achieved through a system wide analysis of public transportation ridership in the Region of Waterloo. Using spatial data and statistical analysis this research aimed to explore several variables and methods in order to establish the most informative quantitative model. This chapter aims to summarize the conclusions of previous chapters, drawing themes which inform research about walkability and public transportation ridership.

Chapters 1 Introduction and Chapter 2 - Literature review, serve to define the scope and previous academic research in the fields of built environment effect on public transportation and pedestrian modal choice. Here, previous research has shown that the built environment influences transportation choice and can be studied in three segments known as the 3D's as defined by Cervero and Kockelman (1997). Previous research using qualitative and mixed methods into the built environments effect on public transportation showed that there was a stated preference towards walkable and pedestrian friendly designs. Research explored in those chapters established a strong academic precedence for the use of linear regression in conducting travel behavior and public transportation ridership research.

Chapter 3 - Methods, provides the overview of data collection, collation, and preliminary statistical transformations. The two potential response variables were defined as the average boarding and alighting at a bus stop during either a peak hour or all day travel time. The variables used to indicate public transportation ridership were established as intervening variable for use in the Base Model.

That model served to inform other models with consistency with previous research and establish a statistical foundation for examining the pedestrian built environment.

Finally, in that chapter, the variables to examine the pedestrian environment specifically were identified and collated. The variables *Entropy, Sidewalk length, Ratio, Intersection density* and *Traffic signal density* were each examined within the 400 meter buffer of a bus stop. These variables were selected to inform the research question differently and were statistically transformed to be employed in regression analysis.

Chapter 4 established which variables were most statistically significant and informative in exploring the relationship between the pedestrian environment and public transit ridership. The results of this inform research on built environments effect on public transit ridership at the bus stop level as it relates to the 400 meter buffer. Through this chapter it was shown that the Base Model variables had the highest correlation with public transportation ridership. Variables specifically selected to examine walkability characteristics showed that the measurements which also correlate to roads, such as *Intersection density* and the *Ratio* of sidewalks to roads, were of greatest significance to public transit ridership. Land use characteristics, as measured using the variable *Entropy*, were also informative when used with *all day average boarding and alighting* data.

That chapter established that the one hour peak average boarding and alighting data showed lower correlation with built environment characteristics, concluding that this was the result of incomplete journeys during the selected hour. The variables *Ratio*, a measurement of sidewalk length to road length within a bus stop buffer, and *Entropy*, a measurement of sameness in land uses within that buffer, were selected as most informative using linear regression. The use of linear regression showed that where key public transit variables, like level of service and transfer location, can heavily influence a correlation with public transportation ridership, walkability characteristics also correlate with ridership with a lower impact.

The resulting variables from linear regression modeling were used in spatial regression analysis presented in Chapter 5. The results showed that an analysis of public transportation ridership is informed by the use of spatial lag regression modeling; however, the analysis of lower impact variables, like those pertaining to walkability, becomes obscured. Where the results of spatial regression showed a marked increase in correlation with the Base Model, the pedestrian environment variables were not significant using those methods.

Previous research, as explored in Chapter 2, has shown a relationship between the built environment and public transportation and the built environment and pedestrian environment. The research questions here explores this through a quantitative methodological approach to understanding the relationships between the pedestrian environment and public transportation. The question asked aims to address the multi-modal interaction and two fields of research. Here, a brief summary of conclusions as they relate to the research sub questions and larger research question are presented.

The first sub question: "what is the most appropriate way to measure pedestrian infrastructure, as it relates to walkability, and what is its correlation with transit ridership?", explores the use of key indicators of pedestrian infrastructure. The variables *Traffic signal*, *Sidewalk length*, *Ratio*, and *Intersection density* each measure slightly different elements of the pedestrian environment. The common emphasis between these variable is the presence of sidewalks. Chapter 4 showed that the variables *Ratio* and *Intersection density* were both significant in correlation with public transportation ridership. These variables, developed in previous built environment research, provide two explanations of the pedestrian environment. The dual role of *Intersection density*, as presented by

Guo and Ferreira Jr (2008), is that each intersection presents a pedestrian barrier while simultaneously increasing pedestrian access. The role of that variable in measuring connectivity is a commonly used tool; however, the conclusions of this thesis suggest that examining the priority of that connectivity through the variable *Ratio* is more effective. This variable presentes the automotive space and pedestrian space as a ratio, while it does not include a measurement of width it provides a coarse measurement of priority in the built environment.

Like the above sub question, the second sub question: “what walkability / built environment characteristics correlate best with transit ridership?”, was informed through examining several different potential variables. In using Cervero & Kockelman (1997) 3D’s, the built environment was understood in three categories: Density, Diversity and Design. Density was measured through both employment density and population density and was included as an intervening variable as its key role is in promoting both public transit ridership and walkability. Diversity was operationalized through the variable *Entropy,* as presented by Ryan & Frank (2009). This variable provides a correlation with ridership especially when using *all day average boarding and alighting* data. The Design characteristics, explored partially above, showed that in the use of public transportation ridership measuring the balance between modes of transportation is crucial. While each of these variables are shown to correlate with public transportation ridership, the effect on the adjusted R^2 was minimal. Indicating that in order to change ridership levels through adjusting Density, Diversity, or Design characteristics would require a very radical change. This conclusion seems consistent with the role of the public transportation service, economic and automotive ownership characteristics in determining public transportation ridership.

The third sub question: “in what way is answering this question informed through the use of linear regression and spatial regression?”, is directly examined in Chapter 5. There it is observed that the effect of spatial characteristics on public transit ridership examined at the bus stop level obscures analysis of the pedestrian environment. Spatial regression as a tool can be used through other methods to inform this discussion, for example examining clustering or land use patterns, in the case of walkability however, the relationship is not significant when using specifically spatial regression. The impact of the spatial lag variable on the adjusted R^2 is significant and presents as an important tool for examining the built environment and public transportation more general in future research.

Each of these sub questions inform the conclusions of the main research question: “How does the pedestrian environment / walkability affect public transportation ridership?”. The results of this quantitative analysis show that the impact is marginal in scale but consistent across land use types. The high level of significance achieved could be the result of the high number of data points while equally indicating the consistent existence of the relationship. The examination of outlier cases in Chapter 4 and use of two response variables *One hour peak average boarding and alighting* and *All day average boarding and alighting* support the study from several different directions. In the end, the variables relating directly to walkability were found to be incidental in public transportation ridership. The role of mixed land use environment and pedestrian infrastructure to promote walkability are strong enough not to be dismissed. Variables which are known to affect walkability and public transportation ridership are shown to interact in this research. This conclusion strongly supports the multi-modal aspect of public transportation and linkages to active transportation and TOD planning paradigms. Those elements of the built environment which support walkability simultaneously support public transportation and while this relationship is not a keystone in creating a public transportation environment they cannot be dismissed in planning for a user friendly public transportation system.

6.1.1 Recommendations for future study

This research did not identify a golden bullet for increasing public transportation ridership at a bus stop. It did however, show that planners of the built environment and public transit systems alike disregard the pedestrian environment at their own peril. The significance of pedestrian connectivity shows that those who invest in the built environment should consider the integration and type of transportation they are designing to support.

In examining the pedestrian environment and public transit ridership several areas were beyond the scope of this study or shown to be of interest by the results presented here. The use of data, especially employment data, from a range of times was a limitation in the conclusions and may have negatively affected the results. To this end more current, finer detail population and employment data may show different results.

This study considered it beyond its scope to examine the quality of either pedestrian infrastructure or pedestrian amenities like lighting and snow removal. The multi-modal nature of public transportation, which is supported by either cycling or carpooling, was also not examined by this study, which may reveal clarity especially in the case of presented outliers. The impact of cycling infrastructure and its role in creating a better pedestrian environment would also inform the multi-modal element of this research. Research around all available modes and accessibility characteristics and the impact on public transportation ridership at the stop level would inform this topic thoroughly.

Within the methods of this research other approaches were considered. These took a sample of stops or segmented the data based on various characteristics that may reveal patterns not exposed here. The segmentation could have been achieved either from the transit perspective, for example only stops with higher than one average boarding and alighting, or from land use perspectives. These different methods, informed by the research here, may serve to further explore this area of academia.

The next step in quantifying the effect of the built environment on public transit ridership may be the use of a quasi-experimental longitudinal study. A time series study could hinge on the installation of various pieces of pedestrian infrastructure over time near stable public transit routes. Such research would serve to inform the effect of change and therefore identify the impact of pedestrian improvements on public transportation ridership. As these infrastructure projects are always challenging to identify awareness may create the opportunity for future study.

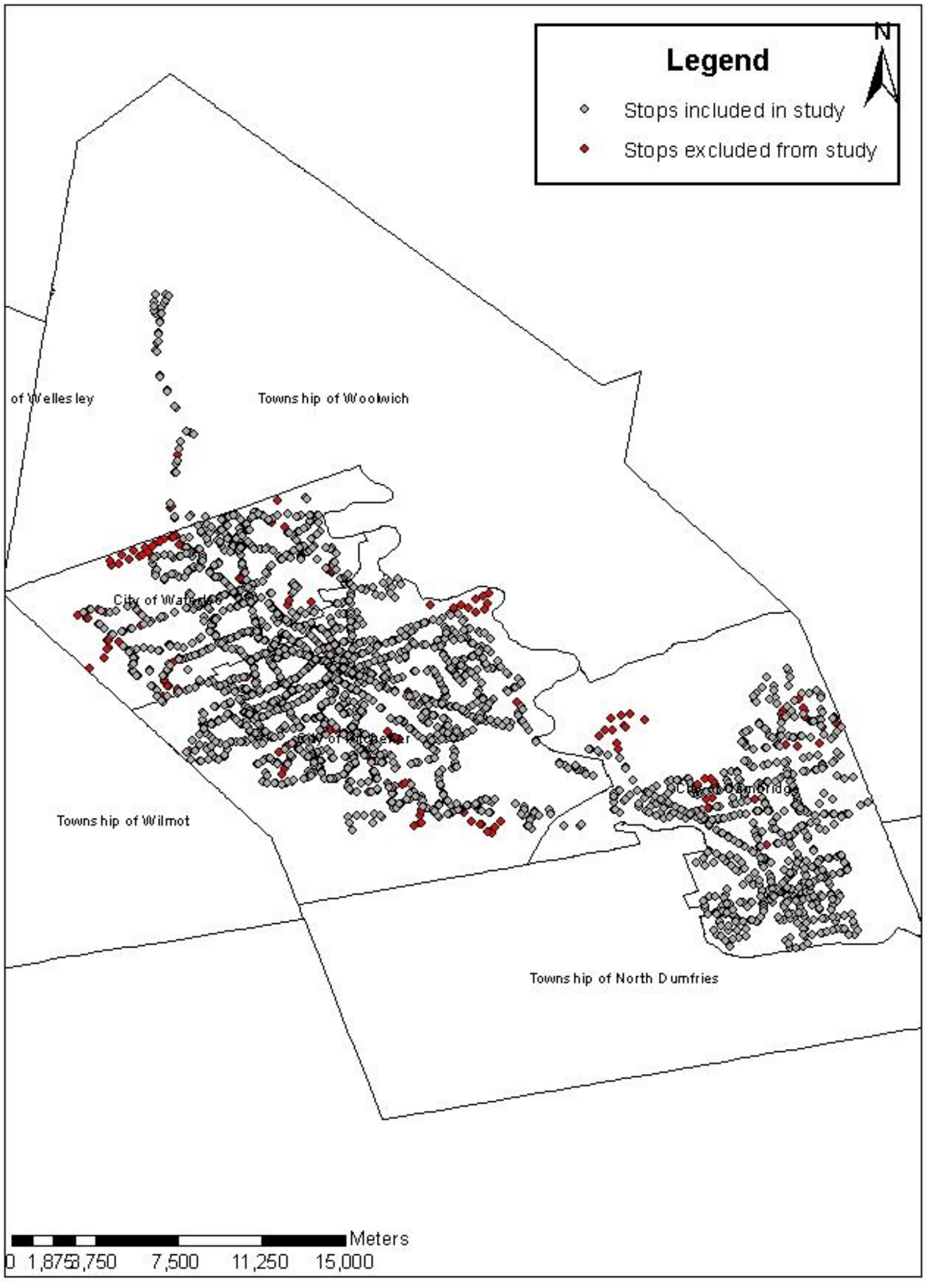
6.1.2 Conclusion

A sidewalk is perhaps the strongest indication of pedestrian priority within the built environment. While the public right of way is dominated by cars, pedestrian infrastructure can assist in the process of redefining how people move around the city. This research aims to inform how these pedestrian variables interacted with bus ridership. Through acknowledging the multi-modal nature of public transportation one is forced to acknowledge that the quality of the pedestrian infrastructure matters. The variables associated with walkability are often influenced by other elements of walking behaviour and transit ridership; this creates a complex issue which cannot be easily segmented.

Since every public transit journey being and ends as a pedestrian journey, the quality of the pedestrian environment should always be of consideration to the planners, architects, and other stakeholders involved. While this research showed that known transit characteristics dominate the statistical relationship, users will always benefit from a better pedestrian environment. The environment which reprioritizes public space away from the car and towards other modes of transportation serves all modes, as this facilitates ease of movement on a human scale. In redefining how people move, public transportation will be an ongoing requirement providing cross town and regional access and therefore reducing car use. A positive public transportation experience starts at the front door.

Appendix A

Maps of stops excluded from study



Appendix B

Regression models to define effect of secondary schools

All cases no *Secondary school* variable

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2486	*Log one hour average boarding and alighting*		0.251
Predictor variables	**Unstandardized coefficient**		**t**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.681	0.143	-11.732	0.000
Level of service	0.194	0.024	8.052	0.000
Transfer location	1.073	0.045	23.994	0.000
Population density	0.000146	0.000025	5.935	0.000
Log Employment density	0.052	0.021	2.458	0.014

All cases including variable *Secondary school*

Model type	Number of observations	Response variable		Adjusted R^2
Linear	2486	*Log one hour average boarding and alighting*		0.262
Predictor variables	**Unstandardized coefficient**		**t**	**Sig.**
	β	**Standard Error**		
(Constant)	-1.690	0.142	-11.886	0.000
Level of service	0.192	0.024	8.028	0.000
Transfer location	1.089	0.044	24.508	0.000
Population density	0.00015	0.000024	6.124	0.000
Log Employment density	0.049	0.021	2.311	0.021
Secondary school	0.524	0.082	6.401	0.000

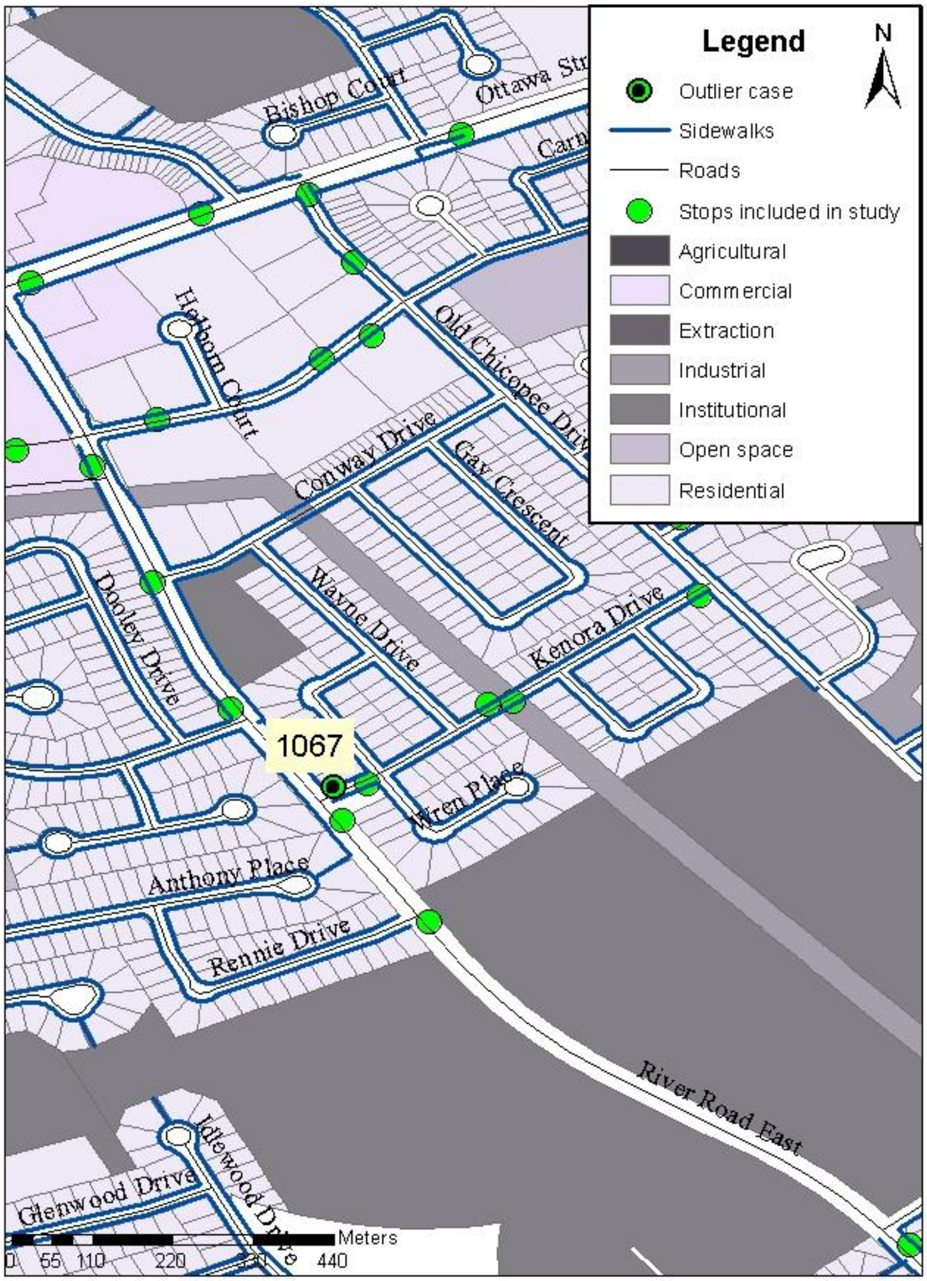
Outlying cases eliminated

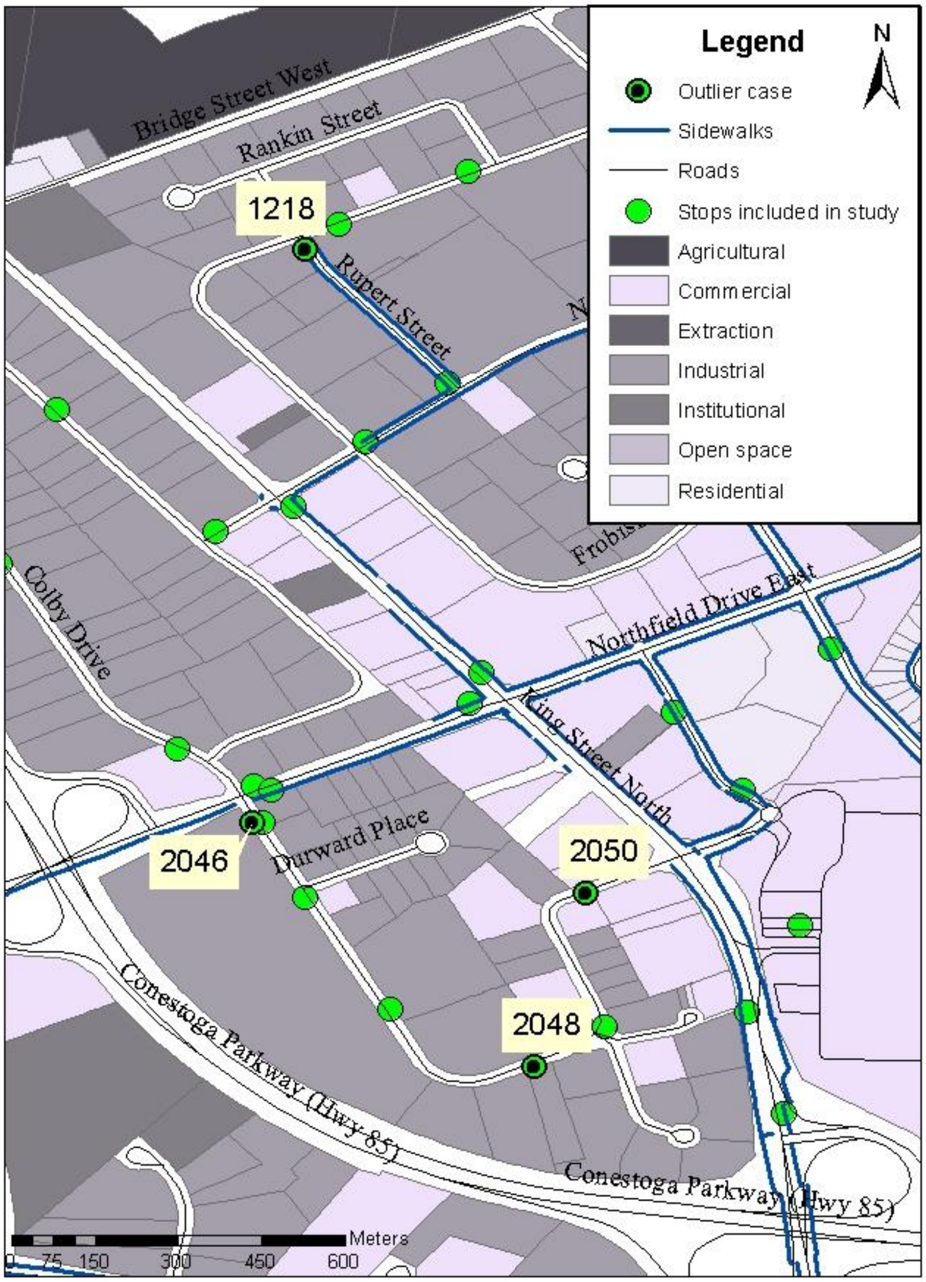
Model type	Number of observations	Response variable			Adjusted R^2
Linear	2478	*Log one hour average boarding and alighting*			0.277
Predictor variables	**Unstandardized coefficient**		**t**		**Sig.**
	B	**Standard Error**			
(Constant)	-1.779	0.138	-12.883		0.000
Level of service	0.239	0.023	10.211		0.000
Transfer location	1.071	0.043	24.855		0.000
Population density	0.000169	0.000024	7.100		0.000
Log Employment density	0.039	0.020	1.914		0.056

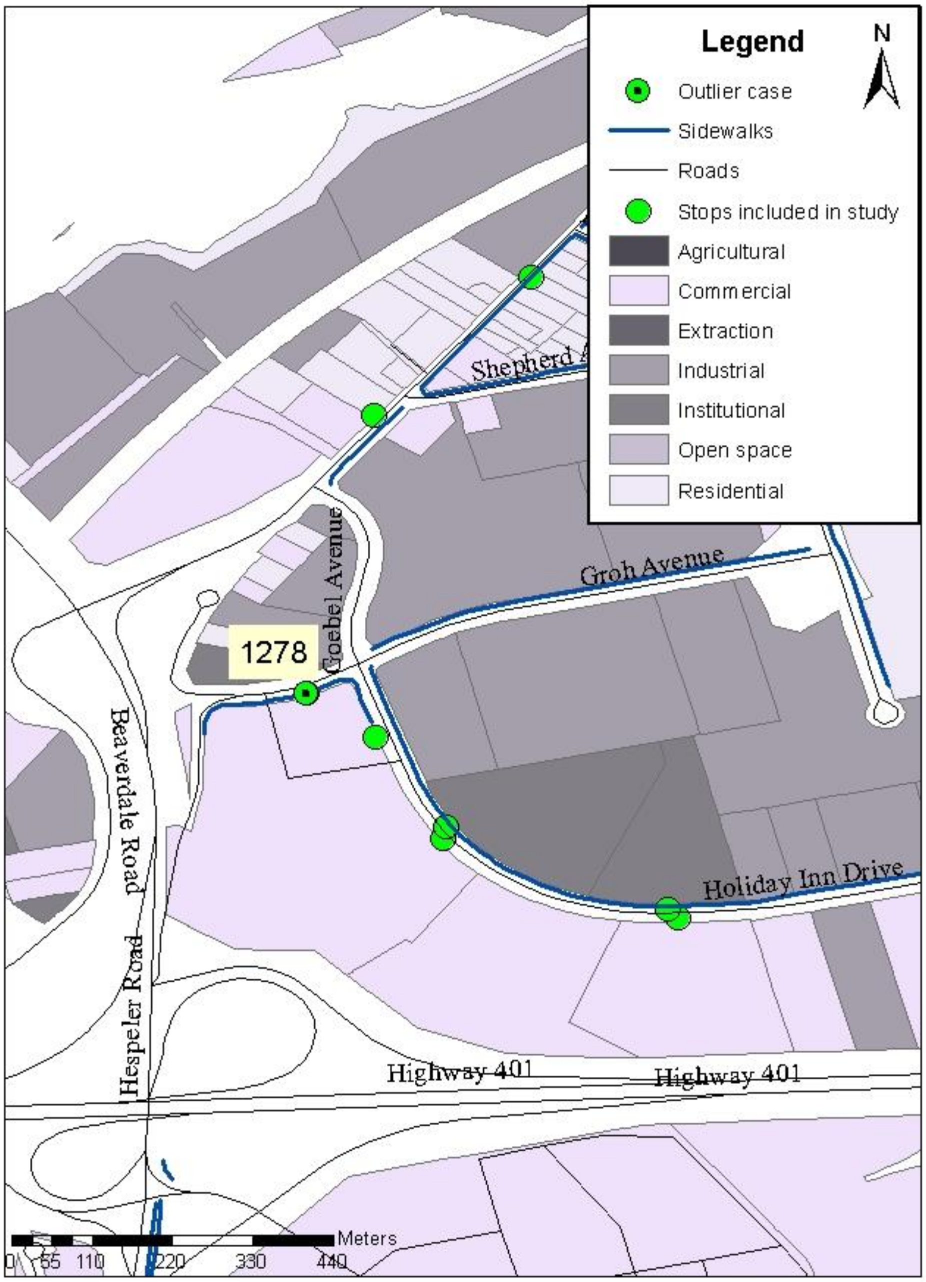
Appendix C

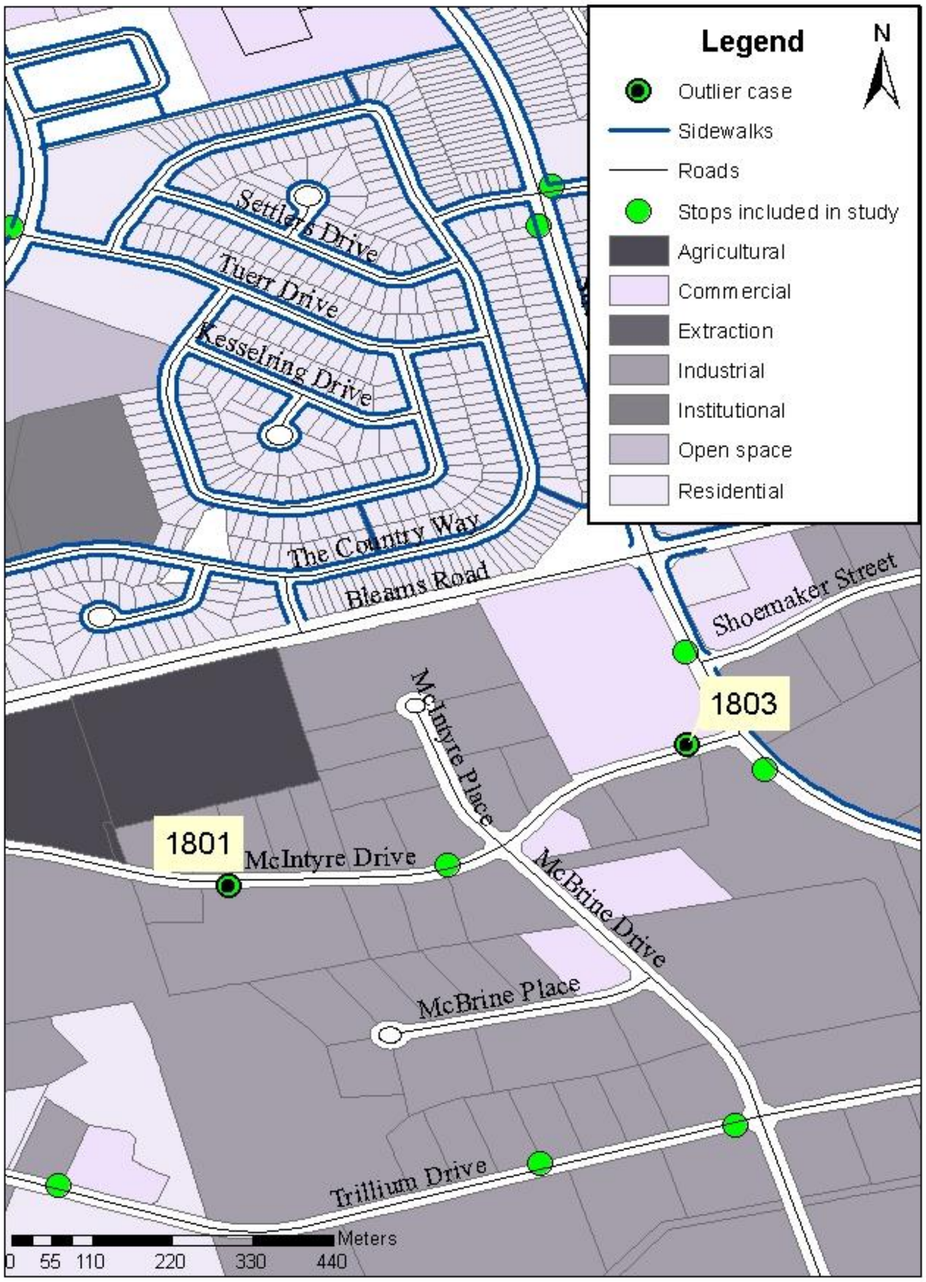
Maps of linear Base Model outlier stops

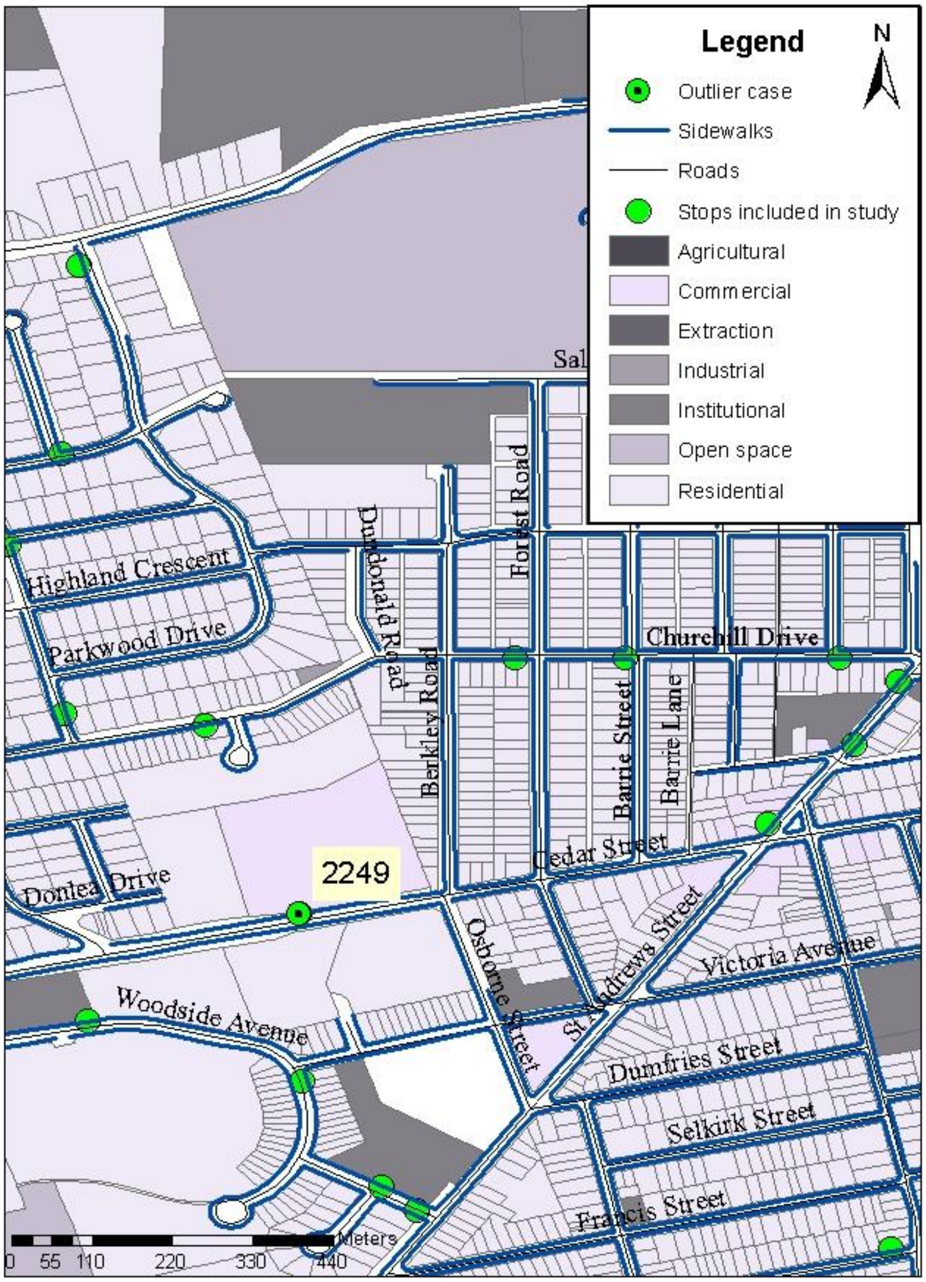
The following maps indicate land use, sidewalks, roads and bus stops around outlier cases from the Base Model.

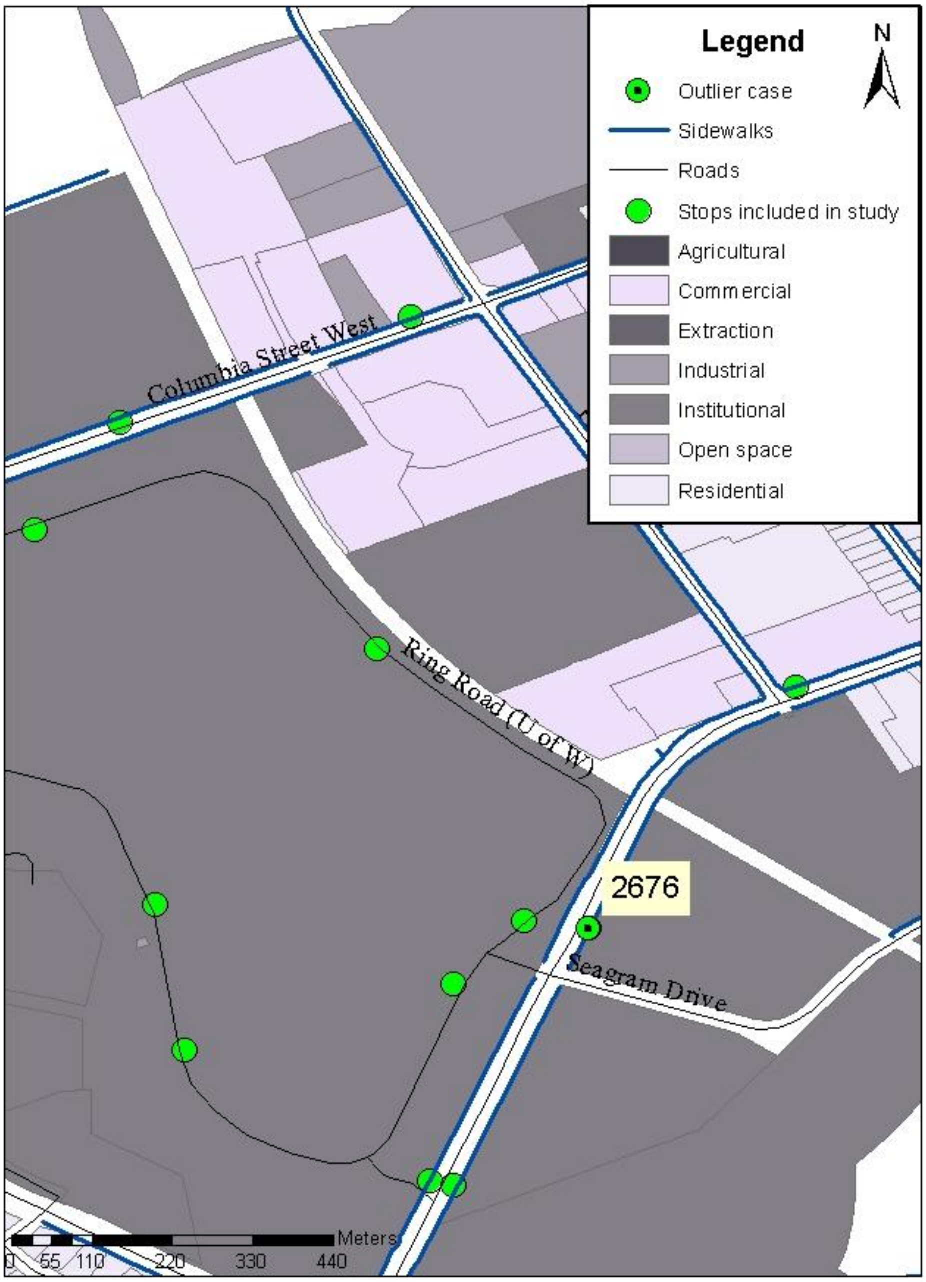


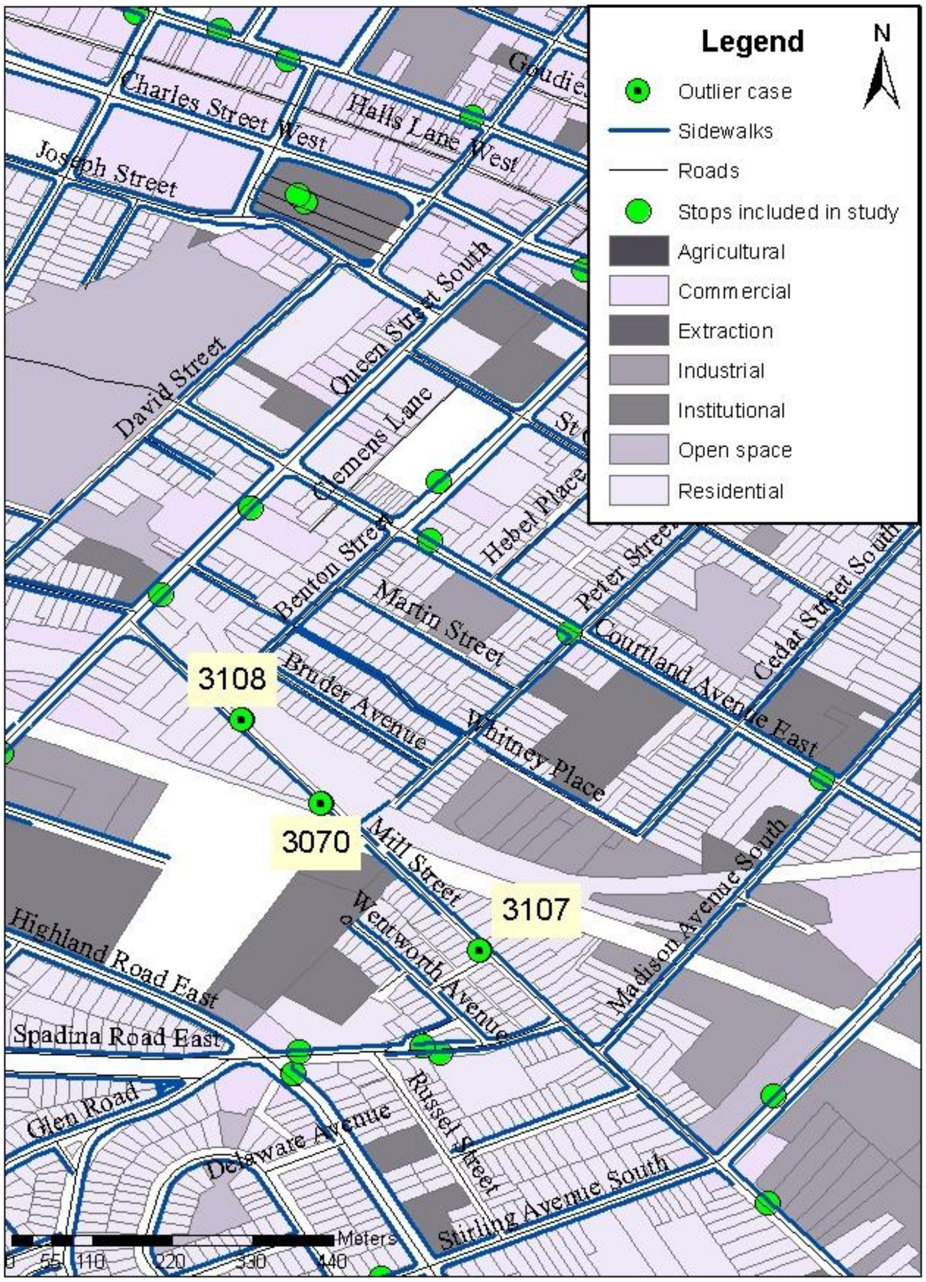

Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Bridge Street West
Rankin Street
Rupert Street
Colby Drive
Northfield Drive East
King Street North
Durward Place
Conestoga Parkway (Hwy 85)
Conestoga Parkway (Hwy 85)
1218
2046
2050
2048
Meters
0 75 150 300 450 600

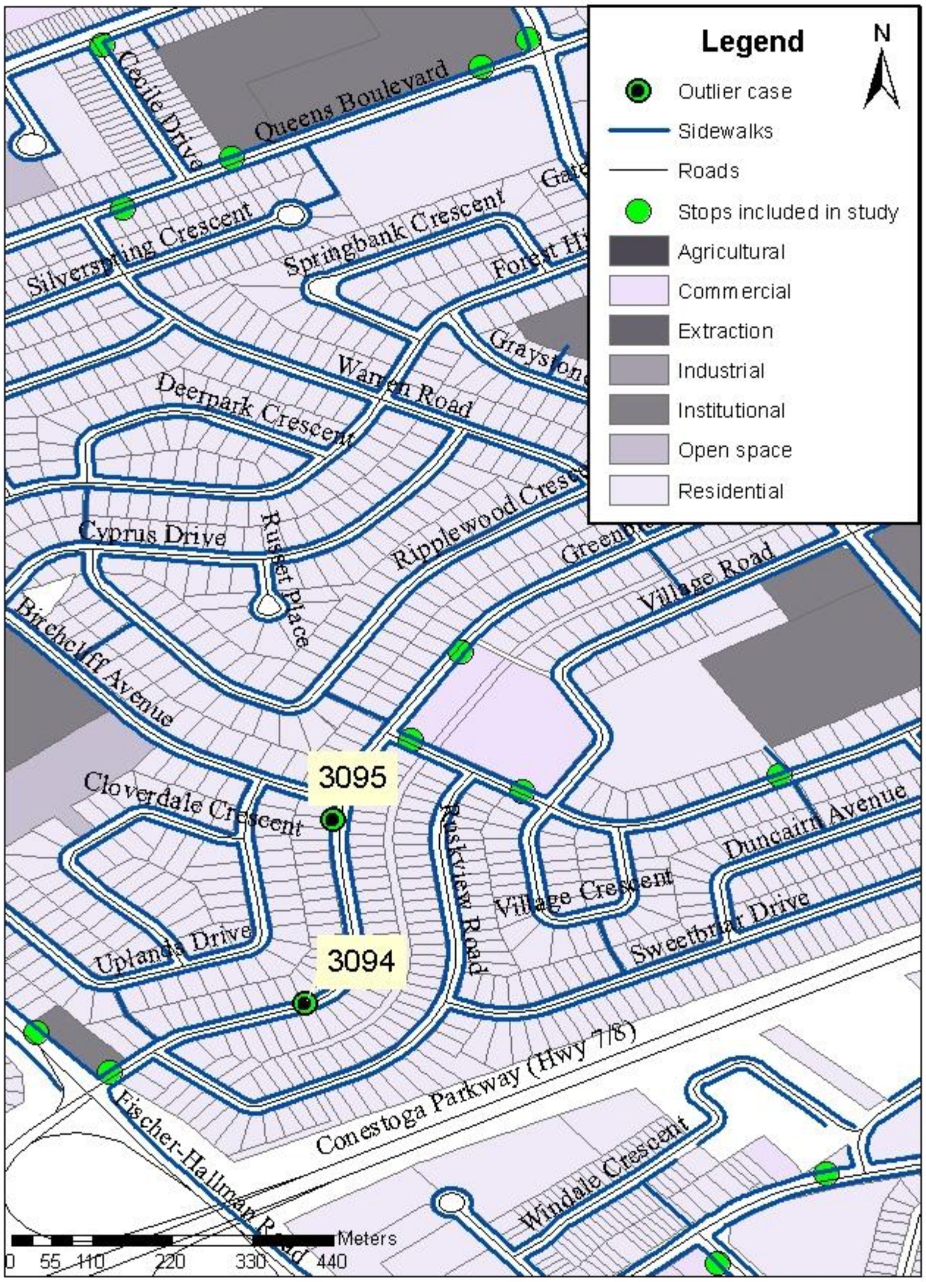

Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
1278
Goebel Avenue
Shepherd A
Groh Avenue
Holiday Inn Drive
Beaverdale Road
Hespeler Road
Highway 401
Highway 401
Meters
0 55 110 220 330 440


Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Settlers Drive
Tuerr Drive
Kesselring Drive
The Country Way
Bleams Road
Shoemaker Street
1803
McIntyre Place
1801
McIntyre Drive
McBrine Drive
McBrine Place
Trillium Drive
Meters
0 55 110 220 330 440

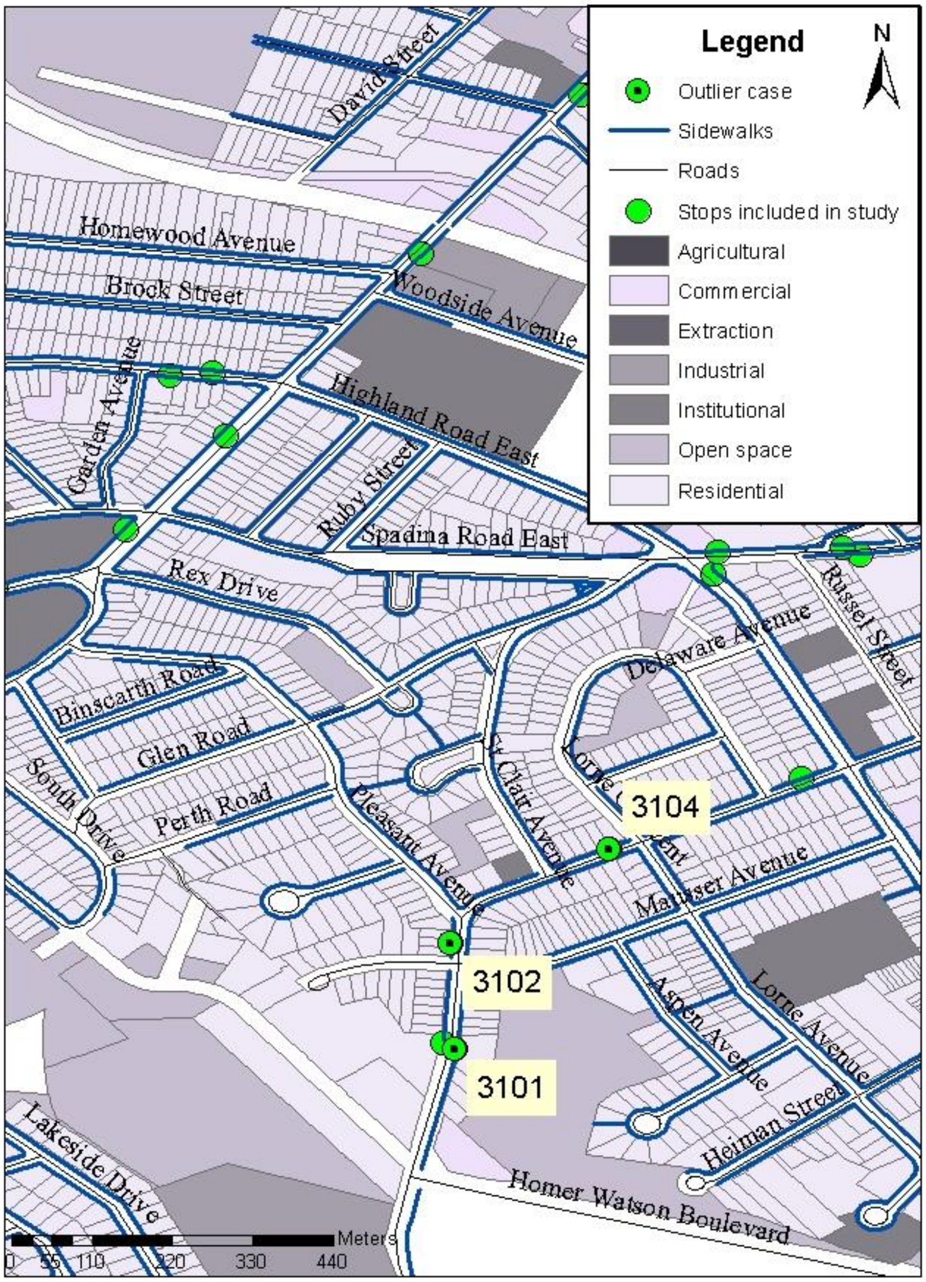

Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Highland Crescent
Parkwood Drive
Dundonald Road
Berkley Road
Forest Road
Churchill Drive
Barrie Street
Barrie Lane
Cedar Street
2249
Dontea Drive
Osborne Street
St Andrews Street
Victoria Avenue
Woodside Avenue
Dumfries Street
Selkirk Street
Francis Street
Meters
0 55 110 220 330 440

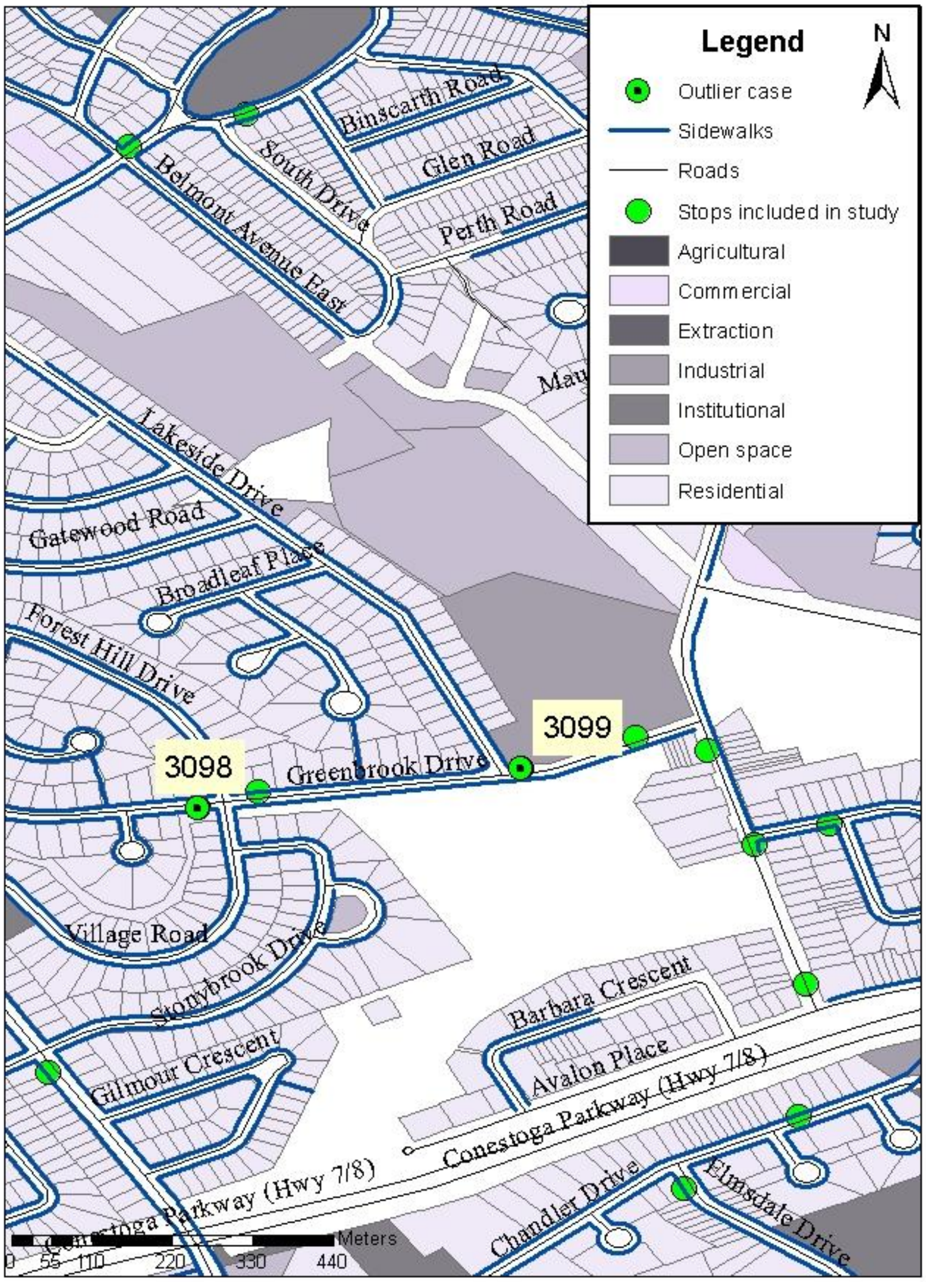

Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Columbia Street West
Ring Road (U of W)
Seagram Drive
2676
Meters
0 55 110 220 330 440

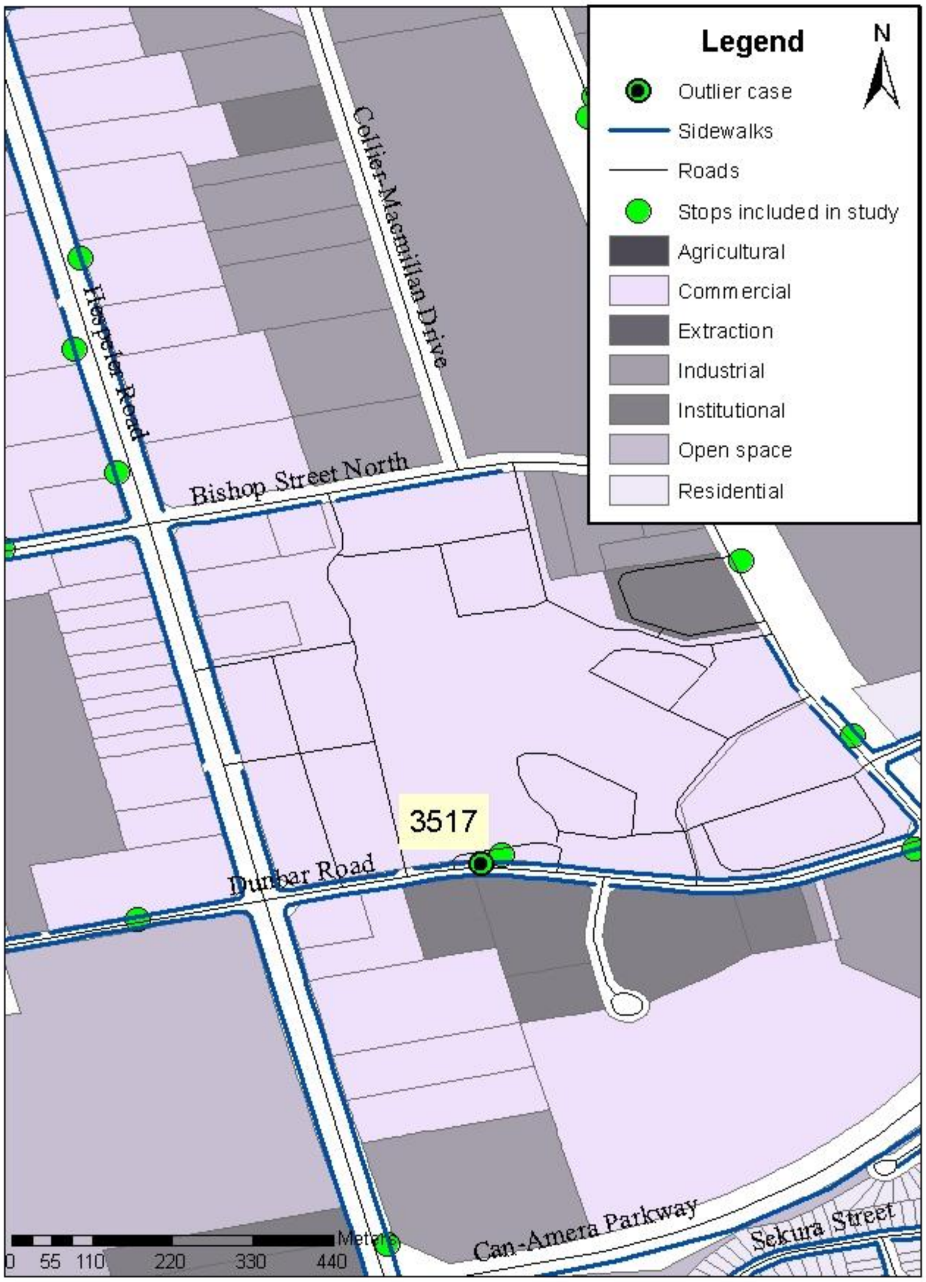

Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Charles Street West
Halls Lane West
Joseph Street
Queen Street South
David Street
Clemens Lane
Hebel Place
Benton Street
Martin Street
Peter Street
Courtland Avenue East
Cedar Street South
Bruder Avenue
Whitney Place
3108
3070
3107
Mill Street
Wentworth Avenue
Madison Avenue South
Highland Road East
Spadina Road East
Glen Road
Delaware Avenue
Russet Street
Stirling Avenue South
Meters
0 55 110 220 330 440

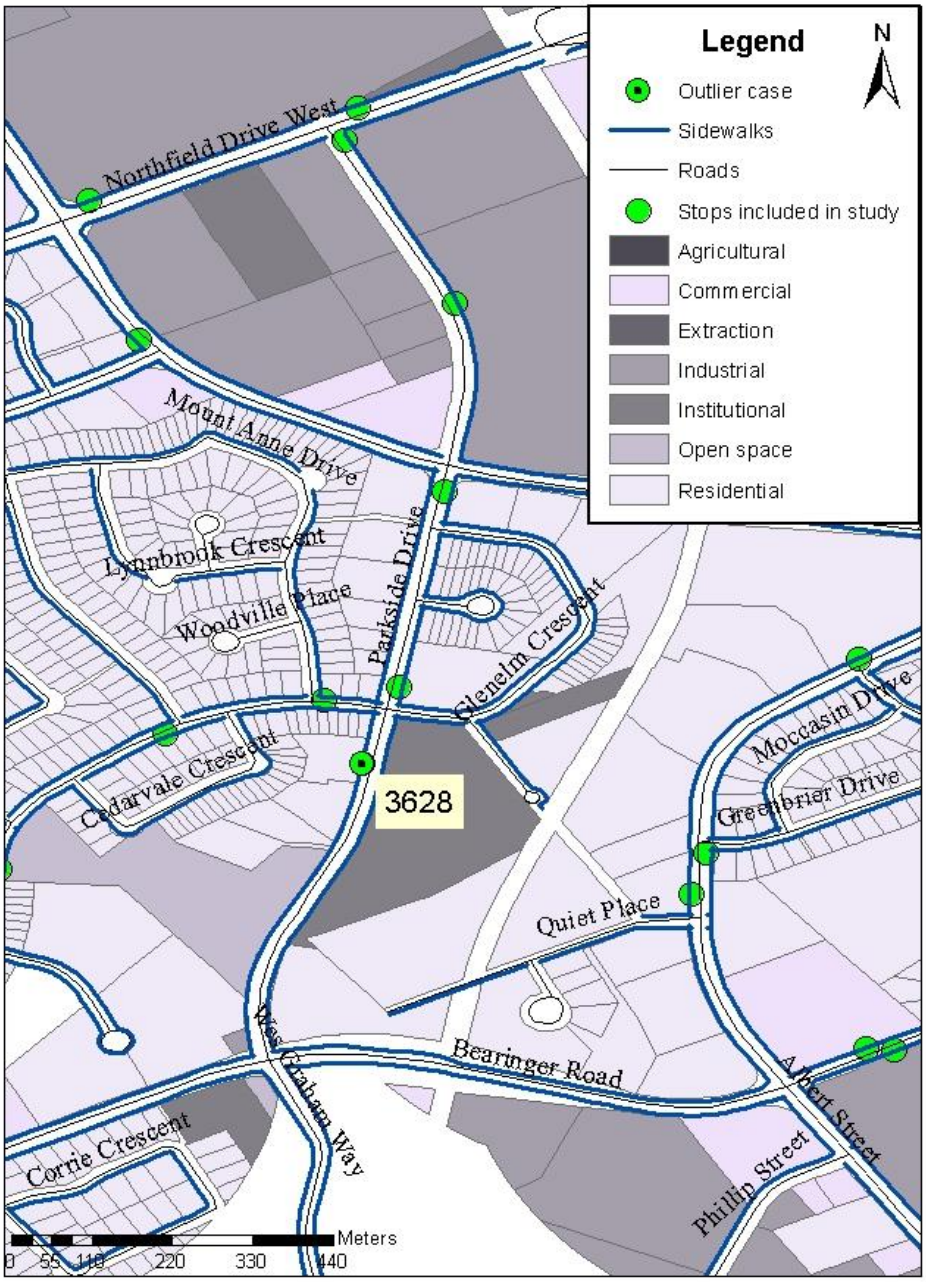


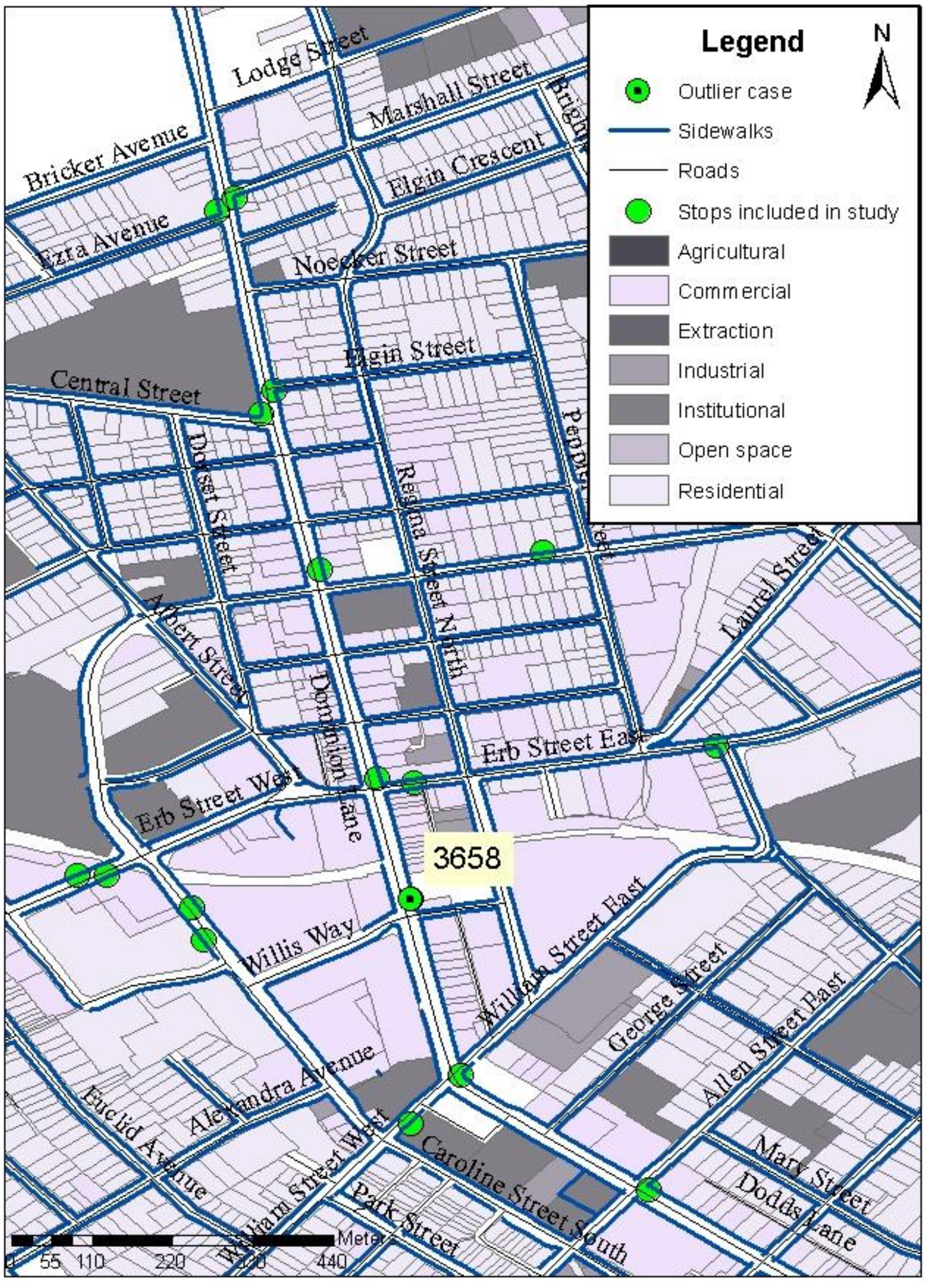
Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Queens Boulevard
Cecile Drive
Silverspring Crescent
Springbank Crescent
Warren Road
Deerpark Crescent
Cyprus Drive
Russet Place
Ripplewood Crescent
Village Road
Birchcliff Avenue
Cloverdale Crescent
Ruskview Road
Village Crescent
Duncairn Avenue
Sweetbriar Drive
Uplands Drive
Conestoga Parkway (Hwy 7/8)
Fischer-Hallman Road
Windale Crescent
3095
3094
Meters
0 55 110 220 330 440


Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
David Street
Homewood Avenue
Brock Street
Woodside Avenue
Garden Avenue
Highland Road East
Ruby Street
Spadina Road East
Rex Drive
Russel Street
Delaware Avenue
Binscarth Road
Glen Road
South Drive
Perth Road
Pleasant Avenue
St Clair Avenue
3104
Manser Avenue
3102
3101
Aspen Avenue
Lorne Avenue
Heiman Street
Lakeside Drive
Homer Watson Boulevard
Meters
0 55 110 220 330 440


Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Binscarth Road
Glen Road
South Drive
Perth Road
Belmont Avenue East
Lakeside Drive
Gatewood Road
Broadleaf Place
Forest Hill Drive
Greenbrook Drive
3098
3099
Village Road
Stonybrook Drive
Gilmour Crescent
Barbara Crescent
Avalon Place
Conestoga Parkway (Hwy 7/8)
Chandler Drive
Elmsdale Drive
Meters
0 55 110 220 330 440


Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Collier-Macmillan Drive
Hespeler Road
Bishop Street North
Dunbar Road
3517
Can-Amera Parkway
Sekura Street
0 55 110 220 330 440
Meters


Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Northfield Drive West
Mount Anne Drive
Lynnbrook Crescent
Woodville Place
Parkside Drive
Glenelm Crescent
Cedarvale Crescent
3628
Moccasin Drive
Greenbrier Drive
Quiet Place
Bearinger Road
Corrie Crescent
Albert Street
Phillip Street
Meters
0 55 110 220 330 440


Legend
N
Outlier case
Sidewalks
Roads
Stops included in study
Agricultural
Commercial
Extraction
Industrial
Institutional
Open space
Residential
Lodge Street
Marshall Street
Bricker Avenue
Elgin Crescent
Ezra Avenue
Noecker Street
Elgin Street
Central Street
Dorset Street
Regina Street North
Albert Street
Laurel Street
Dominion Lane
Erb Street East
Erb Street West
3658
Willis Way
William Street East
George Street
Allen Street East
Alexandra Avenue
Euclid Avenue
William Street West
Caroline Street South
Park Street
Mary Street
Dodds Lane
Meters
0 55 110 220 440

Appendix D

Multi-collinearity table including all intervening and predictor variables

Variable Name	*Transfer location*	*Level of service*	*Employment density*	*Population density*	*Entropy*	*Intersection count*	*Sidewalk length*	*Ratio*	*Log traffic signal*	*Secondary Schools*
Transfer location	1.000	0.158	0.273	0.035	0.142	0.240	0.037	-0.061	0.305	-0.050
Level of service	0.158	1.000	0.273	0.050	0.154	0.131	0.123	0.136	0.341	0.043
Employment density	0.273	0.273	1.000	-0.084	0.223	0.278	0.180	0.040	0.623	0.036
Population density	0.035	0.050	-0.084	1.000	-0.119	0.487	0.699	0.633	0.130	-0.007
Entropy	0.142	0.154	0.223	-0.119	1.000	-0.041	-0.059	0.087	0.318	0.111
Intersection count	0.240	0.131	0.278	0.487	-0.041	1.000	0.713	0.403	0.354	0.019
Sidewalk length	0.037	0.123	0.180	0.699	-0.059	0.713	1.000	0.713	0.339	0.045

Ratio	-0.061	0.136	0.040	0.633	0.087	0.403	0.713	1.000	0.229	0.084
Log traffic signal	0.305	0.341	0.623	0.130	0.318	0.354	0.339	0.229	1.000	0.094
Secondary schools	-0.050	0.043	0.036	-0.007	0.111	0.019	0.045	0.084	0.094	1.000

Appendix E

Spatial lag regression using Base Model to present impact of different threshold weights characteristics

Model type	Number of observations	Response variable	Weight variable threshold distance in meters	R^2
Spatial	2456	*Log all day average boarding and alighting*	100	0.604
Predictor variables	**Unstandardized coefficient**		**z-score**	**Sig.**
	β	**Standard Error**		
Weight Response variable	0.3357901	0.01138107	29.50426	0.0000000
(Constant)	-0.589861	0.0897937	-6.569069	0.0000000
Level of service	0.7304292	0.0288988	25.27541	0.0000000
Transfer location	0.7066167	0.05294631	13.34591	0.0000000
Population density	0.000211743	2.792247e-005	7.583249	0.0000000
Employment density	0.0001077121	3.18826e-005	3.378399	0.0007292

Model type	Number of observations	Response variable	Weight variable threshold distance in meters	R^2
Spatial	2456	*Log all day average boarding and alighting*	200	0.632
Predictor variables	**Unstandardized coefficient**		**z-score**	**Sig.**
	B	**Standard Error**		
Weight Response variable	0.4335866	0.01290281	33.60404	0.0000000
(Constant)	-0.62946	0.08651896	-7.2754	0.0000000
Level of service	0.6672872	0.02835879	23.53018	0.0000000
Transfer location	0.7072966	0.05123063	13.80613	0.0000000
Population density	0.0001393892	2.708817e-005	5.14576	0.0000003

Employment density	5.78078e-005	3.078892e-005	1.877552	0.0604424

Model type	**Number of observations**	**Response variable**	**Weight variable threshold distance in meters**	**R^2**
Spatial	2456	*Log all day average boarding and alighting*	400	0.613
Predictor variables	**Unstandardized coefficient**		**z-score**	**Sig.**
	β	**Standard Error**		
Weight Response variable	0.5610329	0.01687594	33.24454	0.0000000
(Constant)	-0.9579843	0.08864324	-10.80719	0.0000000
Level of service	0.6715501	0.03003288	22.3605	0.0000000
Transfer location	0.810443	0.05280345	15.34829	0.0000000
Population density	9.613956e-005	2.786116e-005	3.450665	0.0005593
Employment density	-8.994335e-006	3.179556e-005	-0.2828802	0.7772688

Model type	**Number of observations**	**Response variable**	**Weight variable threshold distance in meters**	**R^2**
Spatial	2456	*Log all day average boarding and alighting*	577	0.581
Predictor variables	**Unstandardized coefficient**		**z-score**	**Sig.**
	β	**Standard Error**		
Weight Response variable	0.5922852	0.01958521	30.24146	0.0000000
(Constant)	-1.092319	0.09218889	-11.8487	0.0000000
Level of service	0.6906175	0.0313798	22.00835	0.0000000
Transfer location	0.957049	0.05411976	17.68391	0.0000000
Population density	9.084406e-005	2.915831e-005	3.115546	0.0018362

Employment density	-3.70302e-005	3.34143e-005	-1.108214	0.2677693

Model type	**Number of observations**	**Response variable**	**Weight variable threshold distance in meters**	**R^2**
Spatial	2456	*Log all day average boarding and alighting*	800	0.537

Predictor variables	**Unstandardized coefficient**		**z-score**	**Sig.**
	B	**Standard Error**		
Weight Response variable	0.597565	0.02367049	25.24515	0.0000000
(Constant)	-1.316376	0.09709649	-13.5574	0.0000000
Level of service	0.7410248	0.03311676	22.37612	0.0000000
Transfer location	1.08272	0.05578141	19.41005	0.0000000
Population density	0.0001194671	3.084353e-005	3.873328	0.0001074
Employment density	-2.701869e-005	3.584223e-005	-0.7538228	0.4509555

Bibliography

Agrawal, A. W., & Schimek, P. (2007). Extent and correlates of walking in the USA. *Transportation Research Part D: Transport and Environment*, *12*(8), 548–563. doi:10.1016/j.trd.2007.07.005

Anselin, L. (1995). Local Indicators of Spatial Association-LISA, *27*(2).

Ball, K., Bauman, a, Leslie, E., & Owen, N. (2001). Perceived environmental aesthetics and convenience and company are associated with walking for exercise among Australian adults. *Preventive Medicine*, *33*(5), 434–40. doi:10.1006/pmed.2001.0912

Besser, L. M., & Dannenberg, A. L. (2010). Walking to Public Transit Steps to Help Meet Physical Activity Recommendations, *29*(4). doi:10.1016/j.ampre.2005.06.010

Brownson, R. C., Baker, E. a, Housemann, R. a, Brennan, L. K., & Bacak, S. J. (2001). Environmental and policy determinants of physical activity in the United States. *American Journal of Public Health*, *91*(12), 1995–2003. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1446921&tool=pmcentrez&rendertype=abstract

Calthorpe, P. (1993). *The next american metropolis: Ecology Community and the American Dream*. New York: Princeton Architectural Press.

Campbell, B., Zegeer, C., Huang, H., & Cynecki, M. (2004). *A review of pedestrian safety research in the United States and abroad*. McLean, VA. Retrieved from http://trid.trb.org/view.aspx?id=697038

Cao, X., Handy, S. L., & Mokhtarian, P. L. (2006). The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior: Evidence from Austin, TX. *Transportation*, *33*(1), 1–20. doi:10.1007/s11116-005-7027-2

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558. doi:10.1016/j.apgeog.2012.01.005

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment*, *7*(4), 265–284. doi:10.1016/S1361-9209(01)00024-4

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. Retrieved from http://www.sciencedirect.com/science/article/pii/S1361920997000096

Chow, L., Zhao, F., Liu, X., Li, M., & Ubaka, I. (2000). Transit Ridership Model Based on Geographically Weighted Regression, (1972), 105–114.

Clarke, A. (2003). Bikeways and pathways. *Public Roads*. Retrieved from http://trid.trb.org/view.aspx?id=664096

CMHC. (2013). *research highlight Comparing Canadian New Urbanist and Conventional Suburban Neighbourhoods (Updated)* (pp. 1–18).

Congress of the New Urbanism. (2000). *Charter of the New Urbanism*. New York: McGraw Hill.

Creswell, J. W. (2009). *Research Design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). London: SAGE Publications, Inc.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479. doi:10.1177/0309133307083294

De Veaux, R. D., Velleman, P. F., Bock, D. E., Vukov, A. M., & Wong, A. (2012). *Stats: Data and Models* (First Cana.). Pearson Education Canada.

Delmelle, E. C., Haslauer, E., & Prinz, T. (2013). Social satisfaction, commuting and neighborhoods. *Journal of Transport Geography*, *30*, 110–116. doi:10.1016/j.jtrangeo.2013.03.006

Duong, D., & Casello, J. (2010). An empirical comparison of public transportation systems' performance and function (Conference Paper). *Proceedings, Annual Conference - Canadian Society for Civil Engineering*, *2*(1482-1489).

Edwards, R. D. (2008). Public transit, obesity, and medical costs: assessing the magnitudes. *Preventive Medicine*, *46*(1), 14–21. doi:10.1016/j.ypmed.2007.10.004

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. doi:10.1080/01944361003766766

Forsyth, A., Michael Oakes, J., Lee, B., & Schmitz, K. H. (2009). The built environment, walking, and physical activity: Is the environment more important to some people than others? *Transportation Research Part D: Transport and Environment*, *14*(1), 42–49. doi:10.1016/j.trd.2008.10.003

Frank, L. D. (2004). Economic determinants of urban form: resulting trade-offs between active and sedentary forms of travel. *American Journal of Preventive Medicine*, *27*(3 Suppl), 146–53. doi:10.1016/j.amepre.2004.06.018

Frank, L. D., Greenwald, M. J., Winkelman, S., Chapman, J., & Kavage, S. (2010). Carbonless footprints: promoting health and climate stabilization through active transportation. *Preventive Medicine*, *50 Suppl 1*, S99–105. doi:10.1016/j.ypmed.2009.09.025

Frank, L. D., & McKay, H. a. (2010). Time to walk the talk: embracing the built environment to promote physical mobility. *British Journal of Sports Medicine*, *44*(9), 615–615. doi:10.1136/bjsm.2010.076265

Fruin, J. (1971). *Pedestrian planning and design*. New York: Metropolitan Association of Urban Designers and Environmental Planners.

Gallagher, P. (2012). Creating a Pedestrian Level-of-Service Index for Transit Stops: Evidence from Denver's Light Rail System. *Master's Theses*. Retrieved from http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1299&context=gs_theses

Gassaway, A. R. (1992). The adequacy of walkways for pedestrian movement along public roadways in the suburbs of an American city. *Transportation Research Part A: Policy and Practice*, *26*(5), 361–379. doi:10.1016/0965-8564(92)90001-N

Grand River Transit. (2013). *Boarding and Alighting Data*. Region of Waterloo: Grand River Transit.

Guerra, E., & Cervero, R. (2010). Cost of a Ride: The Effects of Densities on Fixed-Guideway Transit Ridership and Capital Costs. *Faculty Research, University of California Transport Center, UC Berkeley*.

Guihaire, V., & Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, *42*(10), 1251–1273. doi:10.1016/j.tra.2008.03.011

Guo, Z. (2010). Causality vs. Correlation: Rethinking Research Design in the Case of Pedestrian Environments and Walking. *NYU Wagner Research Paper*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1573754

Guo, Z., & Ferreira Jr, J. (2008). Pedestrian environments, transit path choice, and transfer penalties: understanding land-use impacts on transit travel. *Environment and Planning B: Planning and Design*, *35*(3), 461–479. doi:10.1068/b33074

Ha, E., Joo, Y., & Jun, C. (2011). An empirical study on sustainable walkability indices for transit-oriented development by using the analytic network process approach. *International Journal of Urban Sciences*, *15*(2), 137–146. doi:10.1080/12265934.2011.615977

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: views from urban planning. *American Journal of Preventive Medicine*, *23*(2 Suppl), 64–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12133739

Hess, D. B. (2011). Walking to the bus: perceived versus actual walking distance to bus stops for older adults. *Transportation*, *39*(2), 247–266. doi:10.1007/s11116-011-9341-1

Hester, R. T. (2010). *Design for Ecological Democracy*. Cambridge, Massachusetts: MIT Press.

Hirsch, J. a, Moore, K. a, Evenson, K. R., Rodriguez, D. a, & Diez Roux, A. V. (2013). Walk score(®) and transit score(®) and walking in the multi-ethnic study of atherosclerosis. *American Journal of Preventive Medicine*, *45*(2), 158–66. doi:10.1016/j.amepre.2013.03.018

Horner, M., & Murray, A. (2002). Excess commuting and the modifiable areal unit problem. *Urban Studies*, *39*(1), 131–139. doi:10.1080/0042098022009911

Information Technology Services. (2013). Sidewalk Data. Region of Waterloo: Unpublished Raw Data.

Krizek, K. (2006). Lifestyles, Residential Location Decisions, and Pedestrian and Transit Activity. *Transportation Research Record*, *1981*(1), 171–178. doi:10.3141/1981-25

Lachapelle, U., & Noland, R. B. (2012). Does the commute mode affect the frequency of walking behavior? The public transit link. *Transport Policy*, *21*, 26–36. doi:10.1016/j.tranpol.2012.01.008

Lai, W.-T., & Chen, C.-F. (2011). Behavioral intentions of public transit passengers—The roles of service quality, perceived value, satisfaction and involvement. *Transport Policy*, *18*(2), 318–325. doi:10.1016/j.tranpol.2010.09.003

Lee, C., & Vernez Moudon, A. (2004). Physical Activity and Environment Research in the Health Field: Implications for Urban and Transportation Planning Practice and Research. *Journal of Planning Literature*, *19*(2), 147–181. doi:10.1177/0885412204267680

Levinson, D. M. (1998). Accessibility and the journey to work. *Journal of Transport Geography*, *6*(1), 11–21. doi:10.1016/S0966-6923(97)00036-7

Maghelal, P., & Capp, C. (2011). Walkability: A Review of Existing Pedestrian Indices. *Journal of the Urban & Regional …*, *23*(2), 5–19. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Walkability:+A+Review+of+Existing+Pedestrian+Indices#0

Maibach, E., Steg, L., & Anable, J. (2009). Promoting physical activity and reducing climate change: opportunities to replace short car trips with active transportation. *Preventive Medicine*, *49*(4), 326–7. doi:10.1016/j.ypmed.2009.06.028

Mavoa, S., Witten, K., McCreanor, T., & O'Sullivan, D. (2012). GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of Transport Geography*, *20*(1), 15–22. doi:10.1016/j.jtrangeo.2011.10.001

McDonald, N., & Trowbridge, M. (2009). Does the built environment affect when American teens become drivers? Evidence from the 2001 National Household Travel Survey. *Journal of Safety Research*, *40*(3), 177–83. doi:10.1016/j.jsr.2009.03.001

Mees, P. (2010). Every Transit User is Also a Pedestrian. In *Transport for suburbia: beyond the automobile age* (pp. 183–193). London, UK: Earthscan.

Millward, H., Spinney, J., & Scott, D. (2013). Active-transport walking behavior: destinations, durations, distances. *Journal of Transport Geography*, *28*, 101–110. doi:10.1016/j.jtrangeo.2012.11.012

Morency, C., Trépanier, M., & Demers, M. (2011). Walking to transit: An unexpected source of physical activity. *Transport Policy*, *18*(6), 800–806. doi:10.1016/j.tranpol.2011.03.010

Newman, P., & Kenworthy, J. (2006). Urban Design to Reduce Automobile Dependence, *2*(1).

Oliver, L. N., Schuurman, N., & Hall, A. W. (2007). Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International Journal of Health Geographics*, *6*, 41. doi:10.1186/1476-072X-6-41

Ontario Ministry of Transportation. (2012). *Transit- Supportive Guidelines*. Queen's Printer for Ontario.

Owen, N., Cerin, E., Leslie, E., duToit, L., Coffee, N., Frank, L. D., … Sallis, J. F. (2007). Neighborhood walkability and the walking behavior of Australian adults. *American Journal of Preventive Medicine*, *33*(5), 387–95. doi:10.1016/j.amepre.2007.07.025

Owen, N., Humpel, N., Leslie, E., Bauman, A., & Sallis, J. F. (2004). Understanding environmental influences on walking; Review and research agenda. *American Journal of Preventive Medicine*, *27*(1), 67–76. doi:10.1016/j.amepre.2004.03.006

Papadimitriou, E., Yannis, G., & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 242–255. doi:10.1016/j.trf.2008.12.004

Planning Housing and Community Services. (2007). Land use parcel dataset. Region of Waterloo.

Pushkarev, B. S., & Zupan, J. M. (1977). *Public transportation and land use policy*. Bloomington: Indiana University Press.

Quintero, L., Sayed, T., & Wahba, M. M. (2013). Safety models incorporating graph theory based transit indicators. *Accident; Analysis and Prevention*, *50*, 635–44. doi:10.1016/j.aap.2012.06.012

Ratner, K. a., & Goetz, A. R. (2013). The reshaping of land use and urban form in Denver through transit-oriented development. *Cities*, *30*, 31–46. doi:10.1016/j.cities.2012.08.007

Region of Waterloo. (n.d.). Regional municipality of Waterloo schools shapefile. Region of Waterloo.

Region of Waterloo. (2009). Traffic signal shape file. Region of Waterloo.

Region of Waterloo. (2011). Terminals - Grand River Transit. Retrieved from http://www.grt.ca/en/travelwithus/terminalsstops.asp

Region of Waterloo. (2012). Road shapefile. Region of Waterloo.

Region of Waterloo. (2013). Fast Facts - Grand River Transit. Retrieved April 03, 2013, from http://www.grt.ca/en/aboutus/fastfacts.asp

Retting, R. a, Ferguson, S. a, & McCartt, A. T. (2003). A review of evidence-based traffic engineering measures designed to reduce pedestrian-motor vehicle crashes. *American Journal of Public Health*, *93*(9), 1456–63. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1447993&tool=pmcentrez&rendertype=abstract

Rodríguez, D. a., & Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D: Transport and Environment*, *9*(2), 151–173. doi:10.1016/j.trd.2003.11.001

Rodríguez, D. a., Khattak, A. J., & Evenson, K. R. (2006). Can New Urbanism Encourage Physical Activity?: Comparing a New Urbanist Neighborhood with Conventional Suburbs. *Journal of the American Planning Association*, *72*(1), 43–54. doi:10.1080/01944360608976723

Ryan, S., & Frank, L. (2009). Pedestrian environments and transit ridership. *Journal of Public Transportation*, 39–57. Retrieved from http://reconnectingamerica.org/assets/Uploads/JPT12-1Ryan.pdf

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*, *25*(2), 80–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12704009

Sallis, J. F., Frank, L. D., Saelens, B. E., & Kraft, M. K. (2004). Active transportation and physical activity: opportunities for collaboration on transportation and public health research. *Transportation Research Part A, 38,* 249–268.

Samimi, A., Mohammadian, A. (Kouros), & Madanizadeh, S. (2009). Effects of transportation and built environment on general health and obesity. *Transportation Research Part D: Transport and Environment*, *14*(1), 67–71. doi:10.1016/j.trd.2008.10.002

Santoso, D. S., Yajima, M., Sakamoto, K., & Kubota, H. (2012). Opportunities and strategies for increasing bus ridership in rural Japan: A case study of Hidaka City. *Transport Policy*, *24*, 320–329. doi:10.1016/j.tranpol.2012.09.005

Statistics Canada. (2008). Profile of Labour Market Activity, Industry, Occupation, Education, Language of Work, Place of Work and Mode of Transportation for Canada, Provinces, Territories, Census Divisions and Census Subdivisions, 2006 Census. Ottawa. doi:Statistics Canada catalogue no. 94-579-XCB2006001

Statistics Canada. (2012). Chapter 1 – Introduction. *2011 Census: Overview of the Census of Population*. Retrieved February 11, 2013, from http://www12.statcan.gc.ca/census-recensement/2011/ref/overview-apercu/pop1-eng.cfm

Su, Q. (2011). The effect of population density, road network density, and congestion on household gasoline consumption in U.S. urban areas. *Energy Economics*, *33*(3), 445–452. doi:10.1016/j.eneco.2010.11.005

Tang, L., & Thakuriah, P. (Vonu). (2012). Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transportation Research Part C: Emerging Technologies*, *22*, 146–161. doi:10.1016/j.trc.2012.01.001

Thompson, D. (2011). Invited Commentaries / Commentaires sollicités The Politics of the Census : Lessons from Abroad. *Canadian Public Policy*, *XXXVI*(3), 377–382.

Thompson, G., Brown, J., & Bhattacharya, T. (2012). What Really Matters for Increasing Transit Ridership: Understanding the Determinants of Transit Ridership Demand in Broward County, Florida. *Urban Studies*, *49*(15), 3327–3345. doi:10.1177/0042098012443864

Verma, J. P. (2013). *Data Analysis in Management with SPSS Software*. New York: Springer.

Villanueva, K., Giles-Corti, B., & McCormack, G. (2008). Achieving 10,000 steps: a comparison of public transport users and drivers in a university setting. *Preventive Medicine*, *47*(3), 338–41. doi:10.1016/j.ypmed.2008.03.005

Wang, W., Li, P., Wang, W., & Namgung, M. (2012). Exploring Determinants of Pedestrians' Satisfaction with Sidewalk Environments: Case Study in Korea. *Journal of Urban Planning and Development*, *138*(2), 166–172. doi:10.1061/(ASCE)UP.1943-5444.0000105

Ward, M. D., & Gleditsch, K. S. (Eds.). (2008). *Spatial Regression Models*. Thousand Oaks, CA: SAGE Publications, Inc. Retrieved from http://dx.doi.org.proxy.lib.uwaterloo.ca/10.4135/9781412985888

Wasfi, R. a, Ross, N. a, & El-Geneidy, A. M. (2013). Achieving recommended daily physical activity levels through commuting by public transportation: Unpacking individual and contextual influences. *Health & Place*, *23C*, 18–25. doi:10.1016/j.healthplace.2013.04.006

Werner, C. M., Brown, B. B., & Gallimore, J. (2010). Light rail use is more likely on "walkable" blocks: Further support for using micro-level environmental audit measures. *Journal of Environmental Psychology*, *30*(2), 206–214. doi:10.1016/j.jenvp.2009.11.003

Wey, W.-M., & Chiu, Y.-H. (2013). Assessing the walkability of pedestrian environment under the transit-oriented development. *Habitat International*, *38*, 106–118. doi:10.1016/j.habitatint.2012.05.004

Yang, Y., Diez Roux, A. V, Auchincloss, A. H., Rodriguez, D. a, & Brown, D. G. (2011). A spatial agent-based model for the simulation of adults' daily walking within a city. *American Journal of Preventive Medicine*, *40*(3), 353–61. doi:10.1016/j.amepre.2010.11.017