Determining the Value of Pedestrian Surfaces in Suburban DC

William Grove Ayers

Thesis submitted to the faculty of the

Virginia Polytechnic Institute and State University

In partial fulfillment of the requirements for the degree of

Master of Science
In
Civil Engineering

Randel L. Dymond, Chair
Kevin J. Boyle
Kathleen L. Hancock

February 20, 2014

Blacksburg, Virginia

Keywords: Walkability, Pedestrian, Hedonic Regression Analysis

Determining the Value of Pedestrian Surfaces in Suburban DC

William Grove Ayers

ABSTRACT

Recent demographic studies suggest a shift in consumer preference away from auto-centric suburban housing to more walkable suburban communities. In response to these changes, efforts have been made to model the walkability of a location and determine its effect on the market value of both residential and commercial real estate. Existing walkability models have considered the importance of amenities and potential pedestrian routes, but have neglected to identify the importance of pedestrian surfaces such as sidewalks and trails as a proportion of the route traveled, and have typically modeled pedestrian movement using exclusively street or trail centerline data. The following paper uses a new walkability model to provide insight on the effect pedestrian surfaces along these amenity routes have on the market value of single family detached and semi-detached homes in Fairfax County, VA. It was found that increases in pedestrian surfaces along amenity routes had little to no effect on home value, but that 3.3%, 1.2 %, and 0.7 % price premiums existed for single family homes that had amenity paths of less than 1 mile to public transportation, public spaces, and recreational facilities, respectively. Price reductions of 3.0 % were discovered for homes that had amenity paths within 1 mile of retail locations.

Acknowledgements

The number of people that I would like to thank for the help, support, and encouragement I received during my time at Virginia Tech is too much to list here. As a best attempt, I would to thank the following groups and individuals:

Fairfax County for providing the data necessary for me to complete my thesis.

The members of my committee, Dr. Kevin Boyle and Dr. Kathleen Hancock, I express sincere gratitude for your guidance and input during the thesis development, writing, and review process.

Dr. Dymond, I express my sincere thanks for the continuing guidance throughout thesis writing and graduate degree process.

My friends and fellow colleagues at Virginia Tech, I thank you for your continuing support.

Contents

Table of Figures

Table of Tables

1 Introduction

The concept of sustainability has developed multiple definitions over the years. What has become arguably the most commonly accepted definition was put forward by the United Nations General Assembly Brundtland Commission in 1987. The commission's report stated that "sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations General Assembly 1987) .This concept has been carried forward into multiple disciplines, including but not limited to public health, environmental preservation, and land development.

As the world's cities and surrounding suburbs continue to expand, the earth's ability to support an ever growing and more mobile population comes into question. It falls on urban planners, civil engineers, land developers, real estate professionals, and even property investors to make educated decisions that push society towards a more sustainable future. It would be a mistake, however, to see this responsibility as a strictly philanthropic venture. Due to changing demographics, lifestyle choices, and basic life values, the implementation of sustainable practices in the built environment provides a promise of financial benefits to well-educated investors. The difficulty comes with deciding which sustainable practices to pursue, and accurately estimating the financial value added by their implementation.

When one considers the many different sustainable land development improvements that can be made to an area, it becomes difficult to determine which features should be considered in this type of study and their relative importance on market value. There are two parameters that help dictate which features can and cannot be accurately studied: (1) the presence of sufficient feature data, and (2) the general acceptance by the public on whether or not a particular feature is sustainable. Fact (1) is simple to address: the researcher should only consider those features for which sufficient information exists, or seek to collect new information. Fact (2), however, represents a more complicated problem, as public opinion on what is healthy or sustainable for society has a tendency to go in and out of fashion. In the United States; however, one particular subject has both become of increasing interest to the research community and also increasingly relevant to the sustainability question: walkability.

Walkability, like sustainability, has numerous definitions that have changed as research on the subject has progressed. The study outlined in this thesis attempts to improve the current state of knowledge on both methods for accurately measuring the concept of walkability, and for determining the financial value of improvements such as sidewalks and trails that are assumed to encourage pedestrian activity. Fairfax County was chosen as the study area of interest since the Fairfax County GIS Department has been a historical collaborator with Virginia Tech's Civil Engineering Department, and is one of many suburban areas that suffers from a built environment that increasingly depends on automobiles for mobility.

2 Literature Review

2.1 Sustainability and Market Value

The importance of developing methods for determining the value of sustainable features is a well discussed topic in the literature. In an effort to improve the state of their discipline, Lorenz and Lutzkendorf identified several key arguments for why sustainable features need to be considered in real estate market valuation models. These arguments were: (1) analysts will be required to consider it through increased market demand for sustainable features, (2) it is the analyst's ethical prerogative to incorporate sustainable features for society's benefit, (3) a lack of consideration for sustainable features can lead to misallocation of capital by investors and undermine the property's true value, and (4) an increased incorporation of these methods will open up opportunities for younger, more socially "enlightened" investors (Lorenz and Lützkendorf 2008, 2011).

In her review of the current state of the literature on sustainability in real estate valuation, Warren-Meyers noted that there was a lack of actionable research that utilized current valuation methods to determine the value of sustainable features. Instead, the literature tended to favor theoretical methods using projections of potential price premiums and cost savings for homes and commercial buildings that implemented new "green" practices (Warren-Myers 2012). Austin as well as several other researchers have attempted to fill this void by providing a stepwise procedure that follows current valuation standards for evaluating a property that contains sustainable features (Austin 2012). Similarly, Levy has shown how adhering to general sustainability practices alone has little effect on the value of commercial real estate, while Garde and Son et al. identified price premiums associated with specific sustainable awards and certifications for buildings and communities such as the LEED rating system (Garde 2009; Levy 2013; Son et al. 2012). Peterson and Gammill goes into further detail, identifying how cost savings from increased energy and water efficiency in commercial buildings gives sustainably rated buildings a cost advantage over conventional buildings (Peterson and Gammill 2010).

Research has also been conducted on how to finance the implementation of sustainable features. Leinberger in his paper on financing progressive development points to the problem of a bias towards short term returns as a hurdle for sufficient investment in sustainable real estate (Leinberger 2001). In the same vein, Lang identifies two fixes that should be made in order to support financing of smart growth initiatives: (1) by focusing investments on the benefits of long term investment in sustainable properties due to their ability to hold market value, and (2) by changing the underlying urban form of the surrounding built environment so sustainable development can be supported (Lang et al. 2008).

The benefits of sustainable development for real estate portfolio investors has also been investigated. Pivo and Fisher researched the performance of portfolios that favored sustainable property investments, and found that investment properties with Energy Star ratings, located near public transit stations, and near urban regeneration areas either had the same or higher annual total returns than a portfolio of non-sustainable properties (Pivo and Fisher 2009). Muldavin identified ten principles that allow underwriters to accurately consider sustainable features when determining the risk associated with a mortgage, or when considering direct investment in sustainable properties (Muldavin 2010).

2.2 Trends towards Walkability

Several studies have shown recent shifts in American homeowner preferences away from auto-centric suburban housing towards more walkable communities. A web based survey of over 2,000 adults conducted by Belden, Russonello & Stewart showed that most homeowners prefer living in pedestrian oriented communities where amenities are within walking distance of their homes. Seventy seven percent of the adults surveyed in this study claimed that having available walking surfaces such as sidewalks and trails were one of the top community characteristics that dictate where they would buy a home (Belden 2011).

The decade review of housing in America put forward by the Urban Land Institute in 2010 showed trends towards the redevelopment of existing suburban areas into walkable, town centers. This is being attributed to an increasing senior baby-boomer population that places higher value on convenience to amenities than privacy and space. On the opposite side of the spectrum, younger generations are renting for longer periods of time due to fallout from the housing market crises of 2008 (McIlwain 2010).

2.3 Walkability Methods

Measuring and improving walkability in urban and suburban areas has been the topic of many journal articles. Southworth identifies six criteria for developing a successful pedestrian network in an urban area: (1) connectivity, (2) ability to access multimodal transit, (3) diverse land use patterns, (4) pedestrian safety, (5) path quality, and (6) path context (Southworth 2005). Ewing and Handy provided a more qualitative approach to measuring walkability with five different criteria including: (1) imageability, (2) enclosure, (3) human scale, (4) transparency, and (5) complexity. In these two studies alone there is both overlap and disagreement as to how walkability should be accurately measured (Ewing and Handy 2009).

Attempts have been made to identify the relative influence of the many walkability characteristics that could potentially be considered in these models. Adkins et al. found that the presence of green street features, parks, and increased connectivity improved the walkability of a particular location (Adkins et al. 2012). In 2012 Alfonzo identified a hierarchy of walking needs as seen by the pedestrian, with feasibility of the walk being the most important and pleasureability being the least (Alfonzo 2005).

Additional studies have developed quantitative metrics and indices to determine the walkability of a location. An example includes the Irvine-Minnesota Index (IMI), an audit tool that determines the degree of walkability for an area from a checklist of 162 built environment observations made by field research teams (Day et al. 2006). A more commonly used walkability model is Walk Score®, a web service that has developed a proprietary algorithm against Google Maps® to assign walkability values based on the distance to amenity groups (Carr et al. 2010). Walk Score® is particularly useful in real estate research since it generates a normalized value from 0 to 100 for a particular address or neighborhood, and can be calculated almost anywhere within the United States (Duncan et al. 2011).

The importance of accessibility to daily destinations in determining the walkability of a location has been researched by several groups. Saelens and Handy found positive correlations between walking and non-residential destinations and land use mixing (Saelens and Handy 2008). Hoehner et al. observed a positive correlation between pedestrian activity and the number of daily destinations accessible by foot (Hoehner et al. 2005). In a 2004 study by Owen et al., it was determined that the accessibility of sidewalks and trails along with the distance to daily destinations were among the most important indicators of walkability within a community (Owen et al. 2004).

2.4 Walkability and Housing Market Value

The economic effects of walkability on land and building market value has been studied over the past decade by several groups. In a 2011 paper by Rauterkus and Miller, the effects of walkability as measured by Walk Score on land value was investigated. The study identified a positive correlation between an increase in Walk Score values and the land value per square foot using hedonic regression analysis (Rauterkus and Miller 2011).

While land value is a strong determinant of market trends, it is important to consider the effects of walkability on a property as a whole. Cortright (2009) sought to determine the influence of walkability on the market value of residential properties in the D.C area. Using Walk Score as a walkability metric, he determined that home buyers find significant economic value in an increase in the walkability of their home.

Attempts have also been made to identify the combined benefits of increased walkability with environmentally conscious design. In 2010, a study was completed in Portland that found an increase in tree cover along pedestrian routes correlated with an increase in property value (Drake-McLaughlin and Netusil 2011). Huffman et al identified an 2%, 4%, and 5% price premiums on homes with access to trails, greenbelts, and both respectively (Asabere and Huffman 2007).

A recent study of interest that compared walkability and property value was conducted by Pivo and Fisher on commercial real estate. The study focused on the premium associated with an increase in the walkability of a commercial property, also using Walk Score as a metric. While controlling for other influential variables such as vacancy rates, inflation, and employment growth, Pivo and Fisher identified a positive correlation between Walk Score and the market value of commercial properties (excluding industrial locations), with little overall change in the monetary returns experienced by the owners. However, they identified several limitations in utilizing Walk Score as a walkability measure, particularly that Walk Score: (1) weights every amenity category equally, (2) accounts for proximity to amenities via all modes of transit, and (3) does not consider other factors such as walking surfaces and land use (Pivo and Fisher 2011).

Additional economic studies have been conducted that utilize methods of measuring walkability outside of Walk Score. Lienberger and Alfonzo studied the economic effects of walkability in regions surrounding the D.C. area using the Irvine-Minnesota Index (IMI). Similar to the other studies, hedonic regression analysis was utilized to determine the impact the IMI score had on various economic and social indicators. The result of this study determined that, based on the IMI values, higher walkability locations performed better economically with increased office, retail, and residential rents (Leinberger and Alfonzo 2012).

3 Determining the Value of Pedestrian Surfaces along Pedestrian Amenity Paths in Suburban DC

3.1 Abstract

Recent demographic studies suggest a shift in consumer preference away from auto-centric suburban housing to more walkable suburban communities. In response to these changes, efforts have been made to model the walkability of a location and determine its effect on the market value of both residential and commercial real estate. Existing walkability models have considered the importance of amenities and potential pedestrian routes, but have neglected to identify the importance of pedestrian surfaces such as sidewalks and trails as a proportion of the route traveled, and have typically modeled pedestrian movement exclusively using street or trail centerline data. This paper presents a new walkability model to provide insight on the effect pedestrian surfaces along these amenity routes have on the market value of single family detached and semi-detached homes in Fairfax County, VA. It was found that increases in pedestrian surfaces along amenity routes had little to no effect on home value, but that 3.3%, 1.2 %, and 0.7 % price premiums existed for single family homes that had pedestrian amenity paths of less than 1 mile to public transportation, public spaces, and recreational facilities respectively. Price reductions of 3.0 % were discovered for homes that had amenity paths to retail locations.

3.2 Introduction

Sustainability has been a recent topic of ongoing conversation within the real estate and land development industries (Lorenz and Lützkendorf 2008). While a significant collection of research has been published on the theoretical benefits of sustainable development practices, there is a notable lack of information on the relationship between sustainability and real estate market value (Warren-Myers 2012). This deficiency has been seen as a barrier to further investment in sustainable development (Sayce et al. 2010).

Lorenz and Lutzkendorf identified several key arguments for why sustainable features need to be considered in real estate market valuation models. These arguments were: (1) analysts will be required to consider it through increased market demand for sustainable features, (2) it is the analyst's ethical prerogative to incorporate sustainable features for society's benefit, (3) a lack of consideration for sustainable features can lead to misallocation of capital by investors and undermine the property's true value, and (4) an increased incorporation of these methods will open up opportunities for younger, more socially "enlightened" investors (Lorenz and Lützkendorf 2008, 2011).

Efforts to identify the potential financial benefits of sustainable development has taken many forms. Several studies have attempted to identify the potential price premium in homes and commercial buildings due to reduced maintenance costs (Austin 2012; Das et al. 2011), how green neighborhood characteristics provide price premiums on homes (Matthews and Turnbull 2007), and even attempts to identify the premiums associated with sustainable awards and certifications for buildings and communities (Garde 2009; Son et al. 2012). While most of these studies suffer from a lack of sufficient data, pedestrian oriented infrastructure is both in need of further research and has been historically implemented in American communities.

Several studies have shown recent shifts in American homeowner preferences away from auto-centric suburban housing towards more walkable communities. A web based survey of over 2,000 adults conducted by Belden, Russonello & Stewart showed that most homeowners prefer living in pedestrian oriented communities where amenities are within walking distance of their homes. Seventy seven percent of the adults surveyed in this study claimed that having available walking surfaces such as sidewalks and trails were one of the top community characteristics that dictate where they would buy a home (Belden 2011). The decade review of housing in America put forward by the Urban Land Institute in 2010 showed trends towards the redevelopment of existing suburban areas into walkable, town centers. This is being attributed to an increasing senior baby-boomer population that places higher value on convenience to amenities than privacy and space. On the opposite side of the spectrum, younger generations are renting for longer periods of time due to fallout from the housing market crises of 2008 (McIlwain 2010).

Recent studies have attempted to determine the relative impact of the built and natural environment on pedestrian behavior, and create indices that assign point locations or regions a walkability score (Adkins et al. 2012; Alfonzo 2005). An example includes the Irvine-Minnesota Index (IMI), an audit tool that determines the degree of walkability for an area from a checklist of 162 observations made by field research teams (Day et al. 2006). A commonly used walkability model is Walk Score®, a web service that has developed a proprietary algorithm against Google Maps® to assign walkability values based on the distance to amenity groups (Duncan et al. 2011). Walk Score® is particularly useful in real estate research since it generates a normalized value from 0 to 100 for a particular address or neighborhood, and can be calculated almost anywhere within the United States.

These studies point to a need for further research into both the concept of walkability in the built environment, and the financial value it and other sustainable development practices present to residential communities. In order to accurately study this we must explore existing methods available for measuring walkability, and identify an effective way of breaking out the contributory value of walkability with respect to property value.

3.3 Previous Research

The economic effects of walkability on land and building market value has been studied over the past decade by several groups. In a 2011 paper by Rauterkus and Miller, the effects of walkability as measured by Walk Score on land value was investigated. The study identified a positive correlation between an increase in Walk Score values and the land value per square foot using hedonic regression analysis (Rauterkus and Miller 2011).

While land value is a strong determinant of market trends, it is important to consider the effects of walkability on a property as a whole. Cortright (2009) sought to determine the influence of walkability on the market value of residential properties in the D.C area. Using Walk Score as a walkability metric, he determined that home buyers find significant economic value in an increase in the walkability of their home.

Attempts have also been made to identify the combined benefits of increased walkability with environmentally conscious design. In 2010, a study was completed in Portland that found an increase in tree cover along pedestrian routes correlated with an increase in property value (Drake-McLaughlin and Netusil 2011). Huffman et al identified an 2%, 4%, and 5% price premiums on homes with access to trails, greenbelts, and both respectively (Asabere and Huffman 2007).

A recent study of interest was conducted by Pivo and Fisher on commercial real estate market values. The study focused on the premium associated with an increase in the walkability of a commercial property, also using Walk Score as a metric. While controlling for other influential variables such as vacancy rates, inflation, and employment growth, Pivo and Fisher identified a positive correlation between Walk Score and the market value of commercial properties (excluding industrial locations), with little overall change in the monetary returns experienced by the owners. However, they identified several limitations in utilizing Walk Score as a walkability measure, particularly that Walk Score: (1) weights every amenity category equally, (2) accounts for proximity to amenities via all modes of transit, and (3) does not consider other factors such as walking surfaces and land use (Pivo and Fisher 2011).

Additional economic studies have been conducted that utilize methods of measuring walkability outside of Walk Score. Lienberger and Alfonzo studied the economic effects of walkability in regions surrounding the D.C. area using the Irvine-Minnesota Index (IMI). Similar to the other studies, hedonic regression analysis was utilized to determine the impact the IMI score had on various economic and social indicators. The result of this study determined that, based on the IMI values, higher walkability locations performed better economically with increased office, retail, and residential rents (Leinberger and Alfonzo 2012).

3.4 Data and Analysis

The study outlined in this paper seeks to identify the effect pedestrian surfaces such as sidewalks and trails have on the market value of single-family homes in suburban DC. This research question was addressed by splitting the analysis into 3 phases: (1) by identifying the preferred walking path for a resident from their home to a particular amenity type, (2) by taking this path and identifying what percentage of it is travelled along pedestrian surfaces, and (3) by identifying whether changes in this percentage, or changes in the availability of particular amenity types via said walking paths, influences the market value of a suburban home.

The following sections detail how this analysis was approached by explaining the data that was utilized, stepping through the walkability model that was developed, and describing the regression model that was implemented to break out the relative influence of these built environment characteristics on market value.

3.4.1 Fairfax County Data Set

Fairfax County was selected for this study due to its large size and the robustness of its GIS and real estate data set. The Fairfax County Mapping and GIS Services Department provided the residential property sales, dwelling characteristics, parcel characteristics and legal information, and surrounding infrastructure data used in this study. The sales data included over 1 million residential transactions over the past 50 years through August, 2013. Dwelling data included information regarding structural characteristics, parcel and legal data included information regarding lot characteristics, and the infrastructure data provided information regarding street and sidewalk networks (Table 3.1). All data tables included a parcel identification field (PARID) that allowed the sales data to be joined with its corresponding dwelling and parcel attributes. Once joined, the sales data was filtered down to a little over 200,000 homes in order to remove erroneous entries and increase the homogeneity of the data set. These filters included removing old sales data (sold before 2010), entries that had non-positive values for the number of total bedrooms or bathrooms, and sales that claimed to have square footages or lot acreages that were less than sufficient for a single family detached or semi-detached home.

Table 3.1. Available Information of interest from Fairfax County.

Characteristic of Interest	Data Element	Source
Sale price	PRICE	Sales Table
Date of sale	SALEDT	Sales Table
Type of sale	SALEVAL_DE	Sales Table
Year home was built	YRBLT (AGE)	Dwelling Table
Number of bedrooms	RMBED	Dwelling Table
Number of bathrooms	TOTBATH	Dwelling Table
Area of recreation room	RECROMAREA	Dwelling Table
Square footage of living area	SFLA	Dwelling Table
Type of residence	LUC_DESC	Parcel Table
Acreage of lot	ACRES	Legal Table

Amenity Categorization

Pedestrian paths are likely to provide value only when they serve a particular purpose. As was mentioned in the literature review, previous studies on walkability used amenity categorization to identify the utility of a particular destination by putting amenities that serve the same functional purpose to a resident in the same group. In addition, it is common practice to assume that residents will only place value on the most accessible (or closest) amenity in each category when purchasing a home (Hoehner et al. 2005; Owen et al. 2004; Saelens and Handy 2008). For the purposes of this study, amenity data was derived from the land use code descriptions provided with the Fairfax County data set. Amenities were then grouped into one of ten categories shown in Table 3.2 by assigning each of the 161 different land use codes to an amenity category, excluding residential uses. Once grouped together, these amenities were geospatially located by placing a point at the centroid of each parcel.

Table 3.2. Amenity categorization.

Amenity Category	Abbreviation	Category #
Groceries & Supermarkets	groc	1
Restaurants (sit down)	rest	2
Fast Food	ffood	3
Office	office	4
Services (medical, fire, police, laundry, etc.)	serv	5
Retail & hardware	retail	6
Recreational facilities & open space	rec	7
Educational facilities	edu	8
Public & religious places (non-recreational)	pub	9
Public Transit Stop	trans	10

3.4.2 Modeling Pedestrian Paths

Suburban pedestrian movement from a point of origin to a destination is difficult to model. Pedestrians, unlike cars, are not confined to specific surfaces and networks, and will frequently switch from defined pedestrian networks (i.e. sidewalks and trails), to non-pedestrian surfaces and networks (i.e. road shoulders and median strips) if necessary (Asabere and Huffman 2007). To accurately model this behavior, pedestrian movement was broken down conceptually into *pedestrian regions*, *pedestrian corridors*, and *pedestrian paths*.

Pedestrian Regions

To both reduce edge effects and provide boundaries to pedestrian activity, *pedestrian regions* were identified. The boundaries of these regions were delineated based on (1) the presence of high speed roadways (> 35 mph) which indicate dangers to pedestrian safety (Gårder 2004), and (2) the absence of pedestrian crossings across these roads. Additional consideration was given to the availability of sidewalk centerline data due to the incomplete nature of the Fairfax County sidewalk data set. Based on these parameters, four *pedestrian regions* were identified: (A) Herndon/Reston, (B) Pimmit, (C) Springfield/Burke, and (D) Alexandria (Figure 3.1). Only residences and amenities within these regions were considered in the final model. This resulted in a reduction of the filtered approximately 200,000 sales down to only 5,353 sales.

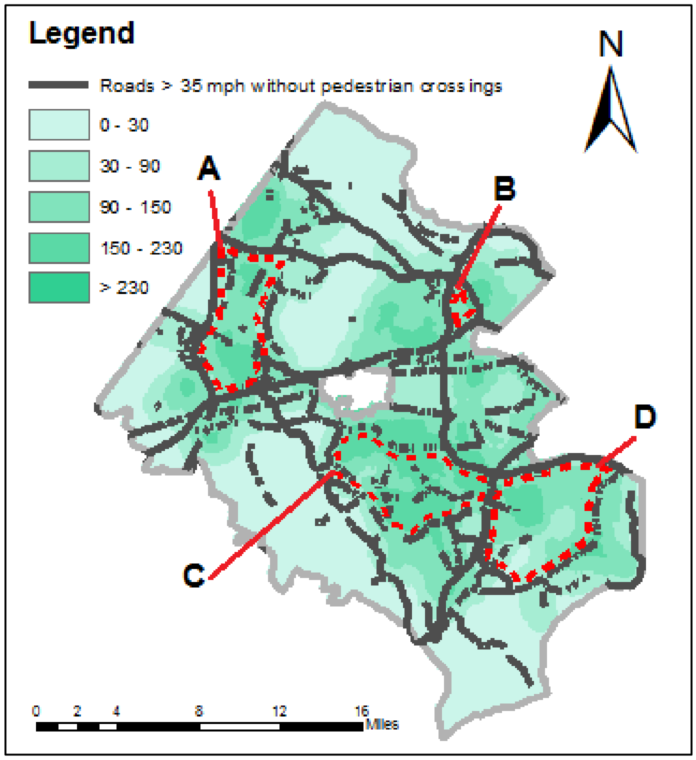

Figure 3.1**.** Pedestrian regions in Fairfax County based on sidewalk density and pedestrian barriers.

Pedestrian Corridors and Paths

The walkability model that was developed in this study attempts to identify the preferred walking path for a resident from their place of residence to an amenity that satisfies a particular need. For the purposes of this study, this has been defined as the *pedestrian path*, which lies within a *pedestrian corridor*, and leads to the closest amenity in a particular amenity category. In order to identify these paths, several key assumptions were made:

1. Residents will only walk along certain predefined surfaces. These surfaces include: (1) non-pedestrian paved areas with low speed limits (< 35 mph), (2) sidewalks, (3) maintained trails, and (4) unidentified areas that exist between surfaces that are within 50 feet of each other (hitherto referred to as "gap" areas). These gap areas were included to account for residents transitioning

between surfaces during their walk.

2. Residents will only utilize a *pedestrian path* that falls within the shortest *pedestrian corridor*. This corridor is defined as a 100 foot wide buffer surrounding the shortest route along these surfaces to the closest amenity in an *amenity category* (see Figure 3.2). This route is delineated assuming that the resident exhibits no preference for which surface type they travel along.

3. Given a *pedestrian corridor*, residents will always travel along pedestrian surfaces (sidewalks, trails) over non-pedestrian surfaces (roads, parking lots, gap areas) when they are available.

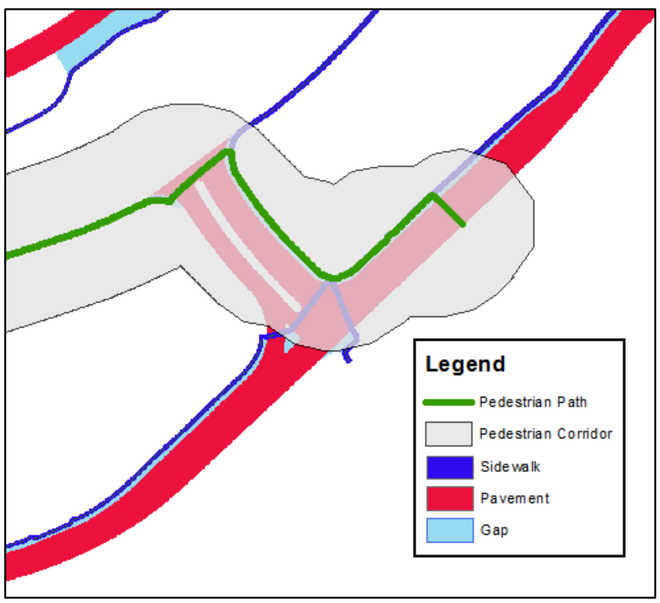

Figure 3.2. An example of a *pedestrian path* within its respective *pedestrian corridor*.

Figure 3.2 shows how a *pedestrian corridor* and a *pedestrian path* interact. The *pedestrian corridor* serves as a bounding container for the *pedestrian path* to ensure the shortest corridor is utilized, but that the resident is selecting sidewalk and trail surfaces whenever they are available.

Least Cost Path Analysis

Pedestrian corridors and *pedestrian paths* were determined using least cost path analysis (LCPA). LCPA is a technique that attempts to determine the most cost-effective route between an origin(s) and a destination(s) using special characteristics and relationships. For this analysis, LCPA takes as inputs a cost weighted distance raster, a direction raster, and a set of destination points. The cost weighted distance raster is calculated from a cost raster and a set of origin points. The cumulative cost of moving away from the origin is calculated for each cell, resulting in a cost weighted distance raster. This raster, along with its corresponding direction raster, allows least cost paths to be identified from the origin to the destination with the least possible cumulative cost en route.

LCPA and the additional modeling steps were completed using GRASS GIS, an open sourced spatial analysis toolset. Individual geoprocessing tools were linked together using python scripting to increase the speed of model iteration. Processing time required to calculate the cost distance raster proved to be a limiting factor in performing the analysis. Therefore, the final model reduced the geographic scope to bring processing times to a more reasonable level.

Two distinct types of least cost path analyses were conducted: one to identify the *pedestrian corridor*, and the other to delineate the *pedestrian path*. For each of these a unique cost raster was developed for each region. The *pedestrian corridor* cost raster gave equal weights to all surfaces to ensure the overall shortest route was identified from the home to a particular amenity (see Figure 3.4). The *pedestrian path* cost raster, however, had weights assigned based on the desirability of the surface for walking, with the sidewalk centerline being the most desirable (1), the remaining sidewalk surface being the second (5), the road being the third (50), and the gap areas the fourth (200) (see Figure 3.3). These weights were set arbitrarily, their significance being the magnitude of difference between them to ensure that, if a sidewalk were available in the direction of the amenity, it would be selected as the route of choice over a road or a gap surface.

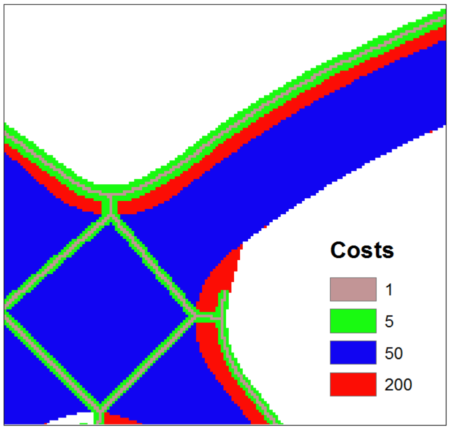

Figure 3.3. Pedestrian Cost Path

Figure 3.4. Pedestrian Corridor

Grass GIS Walkability Script

As mentioned previously, the walkability model made 3 key assumptions: (1) that residents will only walk along predefined surfaces or their gap areas, (2) that pedestrians will only find value in the shortest *pedestrian corridor* from their home to the closest amenity in a particular category, and (3) that, given the shortest *pedestrian corridor*, walkers will choose to travel along pedestrian surfaces such as trails and sidewalks if available. These assumptions are supported by previous walkability research (Duncan et al. 2011; Lo 2009; Moudon et al. 2006). Based on these assumptions, the walkability model iterated through the processing steps shown in Figure 3.5 for every amenity category in a *pedestrian region* to create the *pedestrian paths*. The basic steps shown in the flow chart and beyond are as follows:

1. Create a 1 mile buffer around the amenity. One mile was selected as the maximum distance based on previous walkability research (Duncan et al. 2011).

2. Clip the residential sales database to only include those that fall in this buffer region.

3. Conduct LCPA from the remaining sales to the amenity.

4. Once all the amenities in a category have been iterated through, keep only the shortest *pedestrian corridor* path for each sale.

5. Buffer out the *pedestrian corridor* path 100 feet on each side to create the corridor and clip the *pedestrian path* cost raster to this corridor.

6. Conduct LCPA using the clipped *pedestrian path* cost raster, and the sale and amenity that corresponds to the corridor.

7. Save the path as the *pedestrian path* from the sale in question to the closest amenity in the amenity category. Determine the distance traveled along sidewalks and trails, crosswalks, roads, and gap areas for each *pedestrian path* and output the results to a table.

8. Repeat for the next amenity category, and subsequently the next pedestrian region.

Once all the tables were generated for a particular region they were combined into a single master table that was subsequently used in the following regression analysis.

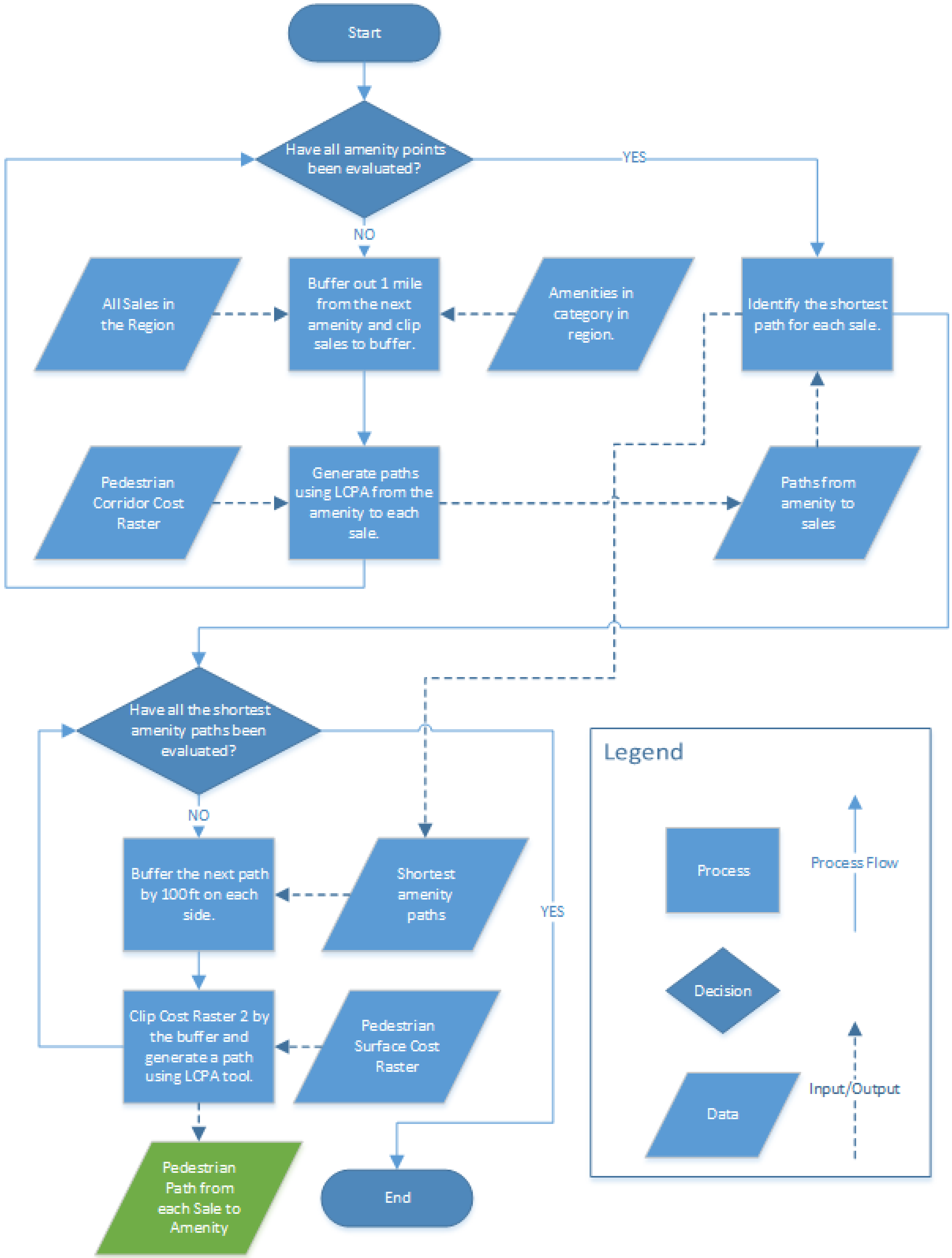

Figure 3.5. Flowchart describing the walkability model for all the amenity points in a particular region.

3.4.3 Hedonic Regression Analysis

Hedonic regression analysis attempts to identify the marginal contribution of different characteristics on the market value of a heterogeneous product. This type of analysis is based on the assumption that, for a particular product (such as a home), the buyer pays a price that represents the cumulative value of its characteristics (Rosen 2008). Therefore, hedonic analysis makes several key assumptions, primarily that (1) the purchaser is aware of and individually values certain components of the product, and (2) that the purchaser has the ability to consider and evaluate all potential combinations of these characteristics before settling on a product to purchase (Cohen et al. 2013). Based on these assumptions, it is possible to infer the marginal contribution each characteristic makes on the market value of a heterogeneous product. For further review of these concepts, please see (Lancaster 1966; Rosen 1974; Sheppard 1997; Sopranzetti 2010).

Conducting hedonic analysis on housing prices typically takes on the general form shown in (1), where S is a vector of structural characteristics, N is a vector of neighborhood characteristics, L is a vector of location based characteristics, and C is a vector of contract characteristics (Sopranzetti 2010). The data provided by Fairfax County included both structural and lot characteristics, census data was utilized to infer neighborhood and community characteristics, and contract characteristics are of little concern since this analysis focuses on single family detached and semidetached homes, not multifamily units. Lastly, the location category is defined by a vector of dummy variables identifying the existence and nature of *pedestrian paths* from a sale to a particular amenity type.

$$\text{Price} = f\{S, N, L, C\} \qquad (1)$$

Selection of predictor variables and the appropriate functional form in hedonic regression analysis depends on previous subject-matter research, basic statistical tests, and investigator intuition. The semi-logarithm functional form has been the prevailing methodology in hedonic amenity studies (Asabere and Huffman 2007) and was therefore used in this particular study. The predictor variables that were initially considered include the ones previously mentioned from the Fairfax County data set (Table 3.1), the median area income (MEDINC) of the census tract the property is located (2010 US Census), and dummy variables indicating the presence of a *pedestrian path* to an amenity category based on the walkability model. Evaluation of the variables and a review of the relevant literature resulted in the regression model shown in Equation 2.

$$\begin{aligned}\ln(Price) = \beta_0 &+ \beta_1 SFLA + \beta_2 RMBED + \beta_3 TOTBATH + \beta_4 AGE \\ &+ \beta_5 ACRES + \beta_6 EDU + \beta_7 FFOOD + \beta_8 GROC + \beta_9 OFFICE \\ &+ \beta_{10} PUB + \beta_{11} REC + \beta_{12} REST + \beta_{13} RETAIL + \beta_{14} SERV \\ &+ \beta_{15} TRANS + \epsilon\end{aligned} \tag{2}$$

Coding of Walkability Variables

The *pedestrian paths* were prepared for regression analysis by translating the percentage of path distances along pedestrian surfaces into binary (or dummy) regression variables. Four different models were developed:

a. Least Walkable Model: If an amenity path exists and is at least 0 % along pedestrian surfaces give a value of 1, else give a value of 0.
b. Fairly Walkable Model: If an amenity path is at least 25% along pedestrian surfaces, give a value of 1, else give a value of 0.
c. Moderately Walkable Model: If an amenity path is at least 50% along pedestrian surfaces, give a value of 1, else give a value of 0.
d. Most Walkable Model: If an amenity path is at least 75% along pedestrian surfaces, give a value of 1, else give a value of 0.

These models were then implemented to determine the relative contribution of pedestrian amenity routes on housing market value.

Spatial Regression

It is often necessary to consider spatial dependency when performing regression analysis on the sales prices of homes (Conway et al. 2010). To test for spatial dependency, the Moran's I statistic was calculated ($I = 45.55$, $p < 0.001$). The results of this test suggest that the null hypothesis of spatial independence cannot be rejected, and modifications should be made to the regression model. Two spatial dependency models are available to combat this problem: the spatial lag model and the spatial error model (Anselin 1988) which are both still in common use today (Cohen et al. 2013; Conway et al. 2010; Song and Knaap 2004).

The spatial lag model (Equation 3), and the spatial error model (Equation 4) each depend on a *n x n* spatial weights matrix W. W is a weights matrix representing the row-standardized Euclidean distances between neighbors, with an upper distance band set to the minimum threshold that allows for at least one neighbor per sale (~1,465 feet). The spatial lag model adds a spatially weighted dependent variable term and a spatial coefficient ρ to account for the effects the sales price has on the value of neighboring properties. Conversely, the spatial error model adds a spatially weighted error term and accounts for spatial correlation in regression residuals, with a spatial error coefficient λ. Calculation of the weights matrix and the spatial lag and error models were accomplished using the python spatial analysis library (pysal) toolset.

$$P = \rho WP + X_1\beta_1 + X_2\beta_2 + \epsilon \tag{3}$$
$$P = X_1\beta_1 + X_2\beta_2 + \lambda W\upsilon + \epsilon \tag{4}$$

In order to determine which spatial model is appropriate, robust Lagrange multiplier (LM) values were calculated for the base case of both the spatial lag (LM = 272.56, $p < 0.001$) and spatial error (LM = 3435.11, $p < 0.001$) models. These results suggest an application of both spatial models are needed. Since the results between the two models did not differ greatly, only the spatial lag results are shown in Table 3.3.

Table 3.3. Spatial Lag Regression Results for the four pedestrian surface models.

	PS >= 0%			PS >=25%			PS >=50%			PS >=75%		
Variable	**SpatLag**		**# Paths**	**SpatLag**		**# Paths**	**SpatLag**		**# Paths**	**SpatLag**		**# Paths**
CONSTANT	6.32E+00	***	NA	6.29E+00	***	NA	6.28E+00	***	NA	6.27E+00	***	NA
	1.36E-01			1.36E-01			1.37E-01			1.36E-01		
SFLA	1.19E-04	***	NA	1.18E-04	***	NA	1.18E-04	***	NA	1.18E-04	***	NA
	3.20E-06			3.20E-06			3.20E-06			3.20E-06		
RMBED	2.26E-02	***	NA	2.28E-02	***	NA	2.29E-02	***	NA	2.32E-02	***	NA
	2.28E-03			2.28E-03			2.29E-03			2.30E-03		
TOTBATH	5.46E-02	***	NA	5.46E-02	***	NA	5.46E-02	***	NA	5.46E-02	***	NA
	2.56E-03			2.57E-03			2.57E-03			2.58E-03		
AGE	-1.51E-03	***	NA	-1.49E-03	***	NA	-1.44E-03	***	NA	-1.42E-03	***	NA
	1.74E-04			1.73E-04			1.73E-04			1.73E-04		
ACRES	9.60E-02	***	NA	9.76E-02	***	NA	9.38E-02	***	NA	9.16E-02	***	NA
	1.14E-02			1.14E-02			1.15E-02			1.15E-02		
EDU	2.57E-03		3894	9.26E-04		3856	3.29E-03		3669	-8.50E-04		3191
	3.51E-03			3.48E-03			3.39E-03			3.34E-03		
FFOOD	2.07E-03		677	-8.22E-04		671	1.01E-03		635	2.62E-03		523
	5.40E-03			5.42E-03			5.56E-03			6.18E-03		
GROC	3.10E-03		849	3.07E-03		837	1.46E-03		790	5.42E-03		712
	4.88E-03			4.94E-03			4.94E-03			5.20E-03		
OFFICE	-1.70E-03		1559	-1.25E-03		1546	-7.35E-04		1519	1.43E-03		1413
	3.57E-03			3.59E-03			3.66E-03			3.78E-03		
PUB	1.15E-02	***	3803	1.23E-02	***	3762	6.21E-03	*	3581	9.72E-04		3047
	3.64E-03			3.59E-03			3.50E-03			3.41E-03		
REC	7.00E-03	*	4501	8.15E-03	**	4440	1.50E-03		4248	1.17E-03		3700
	4.17E-03			4.02E-03			3.73E-03			3.34E-03		
REST	-1.02E-02		525	-7.98E-03		522	-1.29E-02	*	483	-1.33E-02	*	445
	6.92E-03			6.91E-03			6.92E-03			7.10E-03		
RETAIL	-3.03E-02	***	898	-2.78E-02	***	891	-2.29E-02	***	835	-2.16E-02	***	739
	5.36E-03			5.34E-03			5.24E-03			5.70E-03		
SERV	-3.52E-04		2193	-1.14E-04		2157	2.85E-03		2022	1.21E-03		1726
	3.58E-03			3.60E-03			3.62E-03			3.78E-03		
TRANS	3.34E-02	***	582	3.15E-02	***	348	2.88E-02	***	312	2.76E-02	***	233
	5.66E-03			5.84E-03			6.15E-03			7.13E-03		
Lambda/Rho	0.48	***		0.48	***		0.48	***		0.49	***	
	0.01			0.01			0.01			0.01		
R^2	0.857			0.857			0.856			0.856		
# of Obs	5353.00			5353.00			5353.00			5353.00		

*** Denotes significance at the 1% level

** Denotes significance at the 5 % level

* Denotes significance at the 10% level

3.5 Results and Discussion

The four different models shown in Table 3.3 showed a minor decrease in R^2 as a higher percentage of pedestrian surfaces were considered. In addition, there was a general decrease in the significance and magnitude of walkability coefficients as the pedestrian surface percentage cutoff was increased, likely due to the consequently reduced number of paths being used to calculate these coefficients as paths were omitted. Additionally, it should be noted that the *pedestrian corridors* include roadways, and are also a measure of amenity access by other modes of transportation such as cars and bicycles. This effect was observed in several previous papers, and is further supported by the limited difference in these four pedestrian surface models (Cortright 2009; Leinberger and Alfonzo 2012; Pivo and Fisher 2011). Based on these observations, it is possible that suburban homebuyers more highly value accessibility to amenities via multiple modes of transportation (particularly vehicles), and not just specifically by walking.

Several amenity categories, such as access to fast food, grocery stores, office space, and services, showed little to no influence on the market value of the house across all the models. There are many possible reasons for this, only some of which are explored here. It is possible that the accessibility of office space to a home does not necessarily equate to proximity to one's job, which is more likely to affect value. Additionally, fast food locations and services industries (such as hospitals, doctors, painters, etc.) are not typically visited by pedestrians. Fast food restaurants are, by design, intended for an auto-oriented culture and are therefore frequently not seen as desirable destinations for pedestrians. Furthermore, most service industries are typically visited to fulfill a specific, infrequent need, or even possibly never visited at all. The reduced impact of grocery stores, however, is somewhat less expected. Conventional wisdom would dictate that grocery shopping is a frequent activity and that being able to access these stores by foot would provide a definitive advantage. There are many possible reasons why this relationship does not show up in this model, including the inability to strictly define grocery destination points in the data set due to combined retail and grocery stores.

Based on the spatial lag model for greater than or equal to zero percent sidewalk (model 1), access to public transportation via the walkability model described 3.3 % of the change in price, public facilities described 1.2 %, and recreational facilities described 0.7 %. Being in close proximity to mass transit is a logical benefit, especially for daily commuters in the D.C. area. The influence of public transportation access remained high across all four models suggesting that the majority of these routes likely have a significant amount of pedestrian surfaces along their path. The positive influence of having close access to public and recreational facilities can be expected as these are common daily use facilities.

Market value decreases of 3.0 % were identified with walkable access to retail amenities. While this is not expected, it can potentially be attributed to the conditions typically present around suburban retail sites. Many of the retail destinations in Fairfax County, outside of the Reston Town Center, can be considered auto centric and of the strip mall variety. Access to these locations by foot provides no clear advantage to residents since they are designed to be accessed by car, and therefore have a tendency to attract vehicular traffic and could cause a corresponding disamenity effects.

3.6 Conclusions

Market price premiums were identified for close access to transit (3.3 %), public facilities (1.2 %), and recreational facilities (0.7 %), while a price reduction was determined for close access to retail (3.0 %) as discussed above. These premiums and reduction should be taken into consideration when future development is proposed within this region. The results suggest potential financial benefits for a developer who builds public facilities within a new neighborhood to serve the community residents, or one who decides to design for metro bus or train access. It may, however, not be to the developer's benefit to retrofit existing corridors with pedestrian surfaces, as the model suggests no significant change in value due to their presence.

An additional facet that was touched upon but not fully explored in this research should be considered: the installation of new pedestrian corridors that shortcut access to public, recreational, or transit facilities. Since the model suggests value is placed on being close to these types of facilities, land developers should strongly consider walking paths that provide a more direct route. It is more likely value will be found in these types of pedestrian paths, as opposed to paths that follow existing pedestrian corridors.

Several potential improvements to the methods outlined in this study could be made. The four different regression models that were developed look at pedestrian surfaces as a percentage along the amenity path, and not the total distance walked along these surfaces. While this allows comparison across paths that may have varying lengths, one could envision situations where a pedestrian path is excessively short, but is considered to be 75 % along pedestrian surfaces due the presence of a short distance of sidewalk. This situation, from an analytical perspective, may not be comparable to a path that is close to 1 mile in length and has only 50 % of its path occupied by pedestrian surfaces.

Additional consideration should be given to other variables that could potentially have an effect on the relationship between the market value of a home and walkability but were not included in this study. These variables include features such as the number of cars per household, household income, the number workers in a home, and the number of trips taken per day by a household. In order to better understand the limitations of the model used in this paper and increase its robustness, future iterations of the models used in this paper should incorporate new parameters such as these.

One of the major contributions of this paper is in the development of a new, repeatable walkability model. Popular walkability models, such as the Walk Score® algorithm provides a universally available platform upon which these types of analyses can be conducted. Problems arise, however, when attempts are made to repeat and expound upon research that relies on one of these "black box" methods. Changes are constantly being made to these algorithms on a frequent basis, changing the resulting output. A transparent model such as the one in this paper can be replicated, reused, and improved upon by the academic community.

Modifications and improvements can be made to the walkability model developed in this study. The paths generated with the walkability script were generally jagged due to the nature of the least cost path tools and rasters that were utilized. Smoothing algorithms could be conservatively applied to better represent true pedestrian movement across these surfaces. In addition, usage of path length by the surface type instead of percentages may provide more robust results in the regression model, since the current model only considers the presence of an amenity within 1 mile along a path.

4 Conclusions & Engineering Significance

The results of the study provide insight on the value of pedestrian surfaces in suburban Northern Virginia. In the model that was developed, there was little to no increase in single family detached and semi-detached home value when a higher percentage of the *pedestrian path* included pedestrian surfaces. The study also identified price premiums for homes with pedestrian amenity paths within 1 mile of transit stations, public facilities, and recreational facilities. A disamenity effect was observed for proximity to retail.

Land developers are routinely faced with decisions regarding the construction of public infrastructure in a residential community development project. This study suggests that the value added by increasing the amount of pedestrian surfaces may not significantly increase home value. However, consideration should be given to the installation of new, shorter corridors to public, recreational, or transit facilities. While this is not to suggest the complete removal of sidewalks from residential neighborhoods, it does call into question the added value associated with including pedestrian surfaces along non-residential streets in suburban areas. Reduction in the amount of sidewalk placed along roads that are likely to see very little pedestrian activity may reduce construction costs for the developer, while having little to no effect on overall home value.

The results suggest that it may be to the developer's and resident's benefit to construct pedestrian surfaces that shortcut existing *pedestrian corridors* towards the desired amenities. These paths are shown to be valued by residents close to specific amenity types, likely due to the increased convenience associated with a shorter walking distance, and the obvious comparative advantage to other modes of transit. In other words, it is likely a resident sees no value in a pedestrian path located along a major road outside of their neighborhood where driving is the primary mode of transportation along the corridor. In this particular case, the resident would likely prefer to drive to the amenity in question. However, in cases where the pedestrian surfaces short-cut existing corridors (such as roads) providing a shorter path to the amenity, it is likely that the resident will find value in the addition of trails and sidewalks.

Developers should also consider the individual value found in the close proximity of public, recreational, and transit facilities. With this knowledge, a developer or planner may decide to increase accessibility to these types of features by including such facilities in a development plan, or by positioning the development closer to these types of amenities. The opposite is found for retail, with a price reduction of 3.0 % for homes with pedestrian amenity paths within 1 mile of stores. It is possible this effect is due to the overall nature of the retail industry within Fairfax County, which is dominated by big-box and strip mall businesses.

In the future, it would be of considerable interest to investigate the effects of the individual path lengths, as well as the surface sub lengths along the amenity paths. Several questions could be asked of this information: (1) how does the total distance of the path effect home value, (2) how does the frequency of road crossings in order to remain on pedestrian surfaces effect home value, (3) how do different neighborhoods and regions within Fairfax County compare to one another in any of these analyses, and (4) what is found in other municipalities where this model is applied?

Further testing and development of the walkability model that was developed in this study is also of interest. As was discussed previously, the model benefits from being transparent in an area where commonly used metrics, such as Walk Score® are neither transparent nor repeatable. The model in this study has several benefits and drawbacks that could be either fixed, or incorporated into another existing walkability model that does not have the same functionality or suffer from the same problems. The benefits of the model in this study include: (1) the consideration of sidewalks and trails in dictating pedestrian path distances, and (2) the ability to have pedestrian transitions across multiple surface types such as roads, medians, and sidewalks, while still being able to assign preferences and confining the pedestrian to the overall shortest corridor. Models exist that include more demographic information in their analyses (such as Walk Score®), or that provided a better overview of the surrounding environment and the conditions of the walk (such as the Irvine-Minnesota Index), could incorporate the methods used in this study's walkability model. Additional surfaces such as open lots and fields could be incorporated into the data set to account for pedestrian movement across open areas. Lastly, costs could also be assigned depending on the speed limit of the road, to determine whether walking along a sidewalk that follows a high speed road vs. one that follows a low speed road shows any difference in its effect on home value.

A final area of interest within the Fairfax County data set includes the regions around the silver line metro stations that are, as of this writing, opening within the next several weeks. Pedestrian paths and the corresponding home values that are within a specified distance of the metro stations could be evaluated during different points in time: before the announcement of the new metro stations, immediately following the announcement, immediately before construction finished, and immediately following the opening of the first train to the public. Observations on how home and property values change during these different time periods for both homes that have walkability access and those that don't could help clarify the effects new public transit stations have on home and property values in previously suburban areas.

References

Adkins, A., Dill, J., Luhr, G., and Neal, M. (2012). "Unpacking Walkability : Testing the Influence of Urban Design Features on Perceptions of Walking Environment Attractiveness." *Journal of Urban Design*, 17(4), 499–510.

Alfonzo, M. (2005). "To Walk or Not to Walk? The Hierarchy of Walking Needs." *Environment and Behavior*, 37(6), 808–836.

Anselin, L. (1988). "Spatial Econometrics: Methods and Models." Kluwer Academic Publishers, Boston, MA.

Asabere, P. K., and Huffman, F. E. (2007). "The Relative Impacts of Trails and Greenbelts on Home Price." *The Journal of Real Estate Finance and Economics*, 38(4), 408–419.

Austin, G. W. (2012). "Sustainability and Income- Producing Property Valuation: North American Status and Recommended Procedures." *Journal of Sustainable Real Estate*, 4(1), 78–122.

Belden, R. & S. (2011). *The 2011 Community Preference Survey: What Americans are looking for when deciding where to live*.

Carr, L. J., Dunsiger, S. I., and Marcus, B. H. (2010). "Walk score™ as a global estimate of neighborhood walkability." *American journal of preventive medicine*, Elsevier Inc., 39(5), 460–3.

Cohen, J., Blinn, C., Boyle, K. J., Holmes, T., and Moeltner, K. (2013). "Hedonic Valuation with Translating Commodities : Mountain Pine Beetles and Host Trees in the Colorado Front Range." *In Review*.

Conway, D., Li, C. Q., Wolch, J., Kahle, C., and Jerrett, M. (2010). "A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values." *The Journal of Real Estate Finance and Economics*, 41(2), 150–169.

Cortright, J. (2009). *Walking the Walk*.

Das, P., Tidwell, A., and Ziobrowski, A. (2011). "Dynamics of Green Rentals over Market Cycles: Evidence from Commercial Office Properties in San Francisco and Washington DC." *Journal of Sustainable Real Estate*, 3(1), 1–22.

Day, K., Boarnet, M., Alfonzo, M., and Forsyth, A. (2006). "The Irvine-Minnesota inventory to measure built environments: development." *American journal of preventive medicine*, 30(2), 144–52.

Drake-McLaughlin, N., and Netusil, N. (2011). "Valuing Walkability and Vegetation in Portland, Oregon."

Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., and Gortmaker, S. L. (2011). "Validation of walk score for estimating neighborhood walkability: an analysis of four US metropolitan areas." *International journal of environmental research and public health*, 8(11), 4160–79.

Ewing, R., and Handy, S. (2009). “Measuring the Unmeasurable: Urban Design Qualities Related to Walkability.” *Journal of Urban Design*, 14(1), 65–84.

Garde, A. (2009). “Sustainable by Design?: Insights From U.S. LEED-ND Pilot Projects.” *Journal of the American Planning Association*, 75(4), 424–440.

Gårder, P. E. (2004). “The impact of speed and other variables on pedestrian safety in Maine.” *Accident; analysis and prevention*, 36(4), 533–42.

Hoehner, C. M., Brennan Ramirez, L. K., Elliott, M. B., Handy, S. L., and Brownson, R. C. (2005). “Perceived and objective environmental measures and physical activity among urban adults.” *American journal of preventive medicine*, 28(2 Suppl 2), 105–16.

Lancaster, K. J. (1966). “A New Approach to Consumer Theory.” *Journal of Political Economy*, 74(2), 132–157.

Lang, R. E., Lefurgy, J., and Hornburg, S. (2008). *From Wall Street to Your Street : New Solutions for Smart Growth Finance*.

Leinberger, C. B. (2001). *Financing Progressive Development*.

Leinberger, C. B., and Alfonzo, M. (2012). *Walk this Way: The Economic Promise of Walkable Places in Metropolitan Washington, D.C.* 1–22.

Levy, D. (2013). “The effect of sustainability on commercial occupiers’ building choice.” *Journal of Property Investment & Finance*, 31(3), 267–284.

Lo, R. H. (2009). “Walkability: What is it?” *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 2(2), 145–166.

Lorenz, D., and Lützkendorf, T. (2008). “Sustainability in property valuation: theory and practice.” *Journal of Property Investment & Finance*, 26(6), 482–521.

Lorenz, D., and Lützkendorf, T. (2011). “Sustainability and property valuation: Systematisation of existing approaches and recommendations for future action.” *Journal of Property Investment & Finance*, 29(6), 644–676.

Matthews, J. W., and Turnbull, G. K. (2007). “Neighborhood Street Layout and Property Value: The Interaction of Accessibility and Land Use Mix.” *The Journal of Real Estate Finance and Economics*, 35(2), 111–141.

McIlwain, J. (2010). *Housing in America: The Next Decade*.

Moudon, A. V., Lee, C., Cheadle, A. D., Garvin, C., Johnson, D., Schmid, T. L., Weathers, R. D., and Lin, L. (2006). “Operational Definitions of Walkable Neighborhood : Theoretical and Empirical Insights.” *Journal of Physical Activity and Health*, 3(S1), 99–117.

Muldavin, S. (2010). *Value Beyond Cost Savings: How to Underwrite Sustainable Properties*.

Owen, N., Humpel, N., Leslie, E., Bauman, A., and Sallis, J. F. (2004). “Understanding environmental influences on walking; Review and research agenda.” *American journal of preventive medicine*, 27(1), 67–76.

Peterson, K., and Gammill, R. (2010). *The Economics of Sustainability in Commercial Real Estate*.

Pivo, G., and Fisher, J. D. (2009). “Investment Returns from Responsible Property Investments : Energy Efficient, Transit-oriented and Urban Regeneration Office Properties in the US from 1998-2008.”

Pivo, G., and Fisher, J. D. (2011). “The Walkability Premium in Commercial Real Estate Investments.” *Real Estate Economics*, 39(2), 185–219.

Rauterkus, S. Y., and Miller, N. G. (2011). “Residential Land Values and Walkability.” *Journal of Sustainable Real Estate*, 3(1), 23–43.

Rey, S. (2014). “Python Spatial Analysis Library.” <pysal.org>.

Rosen, S. (1974). “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competiton.” *Journal of Political Economy*, 82(1), 34–55.

Rosen, S. (2008). “Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition.” *Journal of Political Economy*, 82(1), 34–55.

Saelens, B. E., and Handy, S. L. (2008). “Built environment correlates of walking: a review.” *Medicine and science in sports and exercise*, 40(7 Suppl), S550–66.

Sayce, S., Sundberg, A., and Clements, B. (2010). *Is Sustainability Reflected in Commercial Property Prices: An analysis of the Evidence Base*.

Sheppard, S. (1997). “Hedonic Analysis of Housing Markets.” *Handbook of Regional and Urban Economics: Applied urban economics*.

Son, K., Asce, S. M., Choi, K., Asce, M., Woods, P., Park, Y. J., and Ph, D. (2012). “Urban Sustainability Predictive Model Using GIS : Appraised Land Value versus LEED Sustainable Site Credits.” *Journal of Construction Engineering and Management*, 138(September), 1107–1112.

Song, Y., and Knaap, G.-J. (2004). “Measuring the effects of mixed land uses on housing values.” *Regional Science and Urban Economics*, 34(6), 663–680.

Sopranzetti, B. J. (2010). “Hedonic Regression Analysis in Real Estate Markets: A Primer.” *Handbook of Quantitative Finance and Risk Management*, C.-F. Lee, A. C. Lee, and J. Lee, eds., Springer US, Boston, MA, 1201–1207.

Southworth, M. (2005). “Designing the Walkable City.” *Journal of Urban Planning and Development*, 131(4), 246–257.

Taylor, L. O., Champ, P., Boyle, K., and Brown, T. (2003). “The Hedonic Method.” *A Primer on Non-market Valuation*, Springer Netherlends, 331–393.

United Nations General Assembly. (1987). *Report of the World Commission on Environment and Development: Our Common Future*.

Warren-Myers, G. (2012). "The value of sustainability in real estate: a review from a valuation perspective." *Journal of Property Investment & Finance*, 30(2), 115–144.

A Analysis Details

The following appendix provides a detailed review of the analysis conducted in this thesis. It steps through each part of the analysis and provides references to tables, figures, and python scripts in appendices as applicable. It should be noted that GIS analysis was conducted using a combination of GRASS GIS, an open source GIS software tool, and ESRI's ArcMap software.

A.1 Filtering the Data

The initial data set provided by Fairfax County included over 30 GIS shape files of countywide features and 8 tables of parcel specific data. The raw data shape files and tables that were utilized for this analysis are shown in Table A.1 below. Each sale and amenity used in the analysis was identified using the parcel identification number (PARID). PARIDs are unique and provide a general location of the building by linking these IDs to the PARCELS.shp layer.

Table A.1. Descriptions of the raw data and their usage in the analysis.

Data Source	Description	Usage
PARCELS.shp	Polygons delineating parcel boundaries with the parcel identification number (PARID).	Used to locate amenities and sales using the PARID. Points were placed at the centroid of each parcel to signify building locations.
SIDEWALKS_CENTERLINE.shp	Lines delineating the centerline of the county sidewalks. Attributes included sidewalk width and whether it was a crosswalk or a sidewalk.	Used to create a sidewalk polygon layer.
MAJOR_TRANSPORTATION_ AREAS.shp	Polygons delineating the edge of pavement for major thoroughfares and highways in Fairfax County.	Combined with MINOR_TRANPORTATION_ AREAS.shp to create the edge of pavement polygon layer.
MINOR_TRANSPORTATION_ AREAS.shp	Polygons delineating the edge of pavement for minor thoroughfares and side streets in Fairfax County.	Combined with MINOR_TRANPORTATION_ AREAS.shp to create the edge of pavement polygon layer.
ROADWAYS.shp	Lines delineating the center of all major roads in Fairfax County. Attributes included speed limits.	Used to delineate *pedestrian regions*.
DWELLDATALL.dbf	A table containing physical characteristics for dwellings, identified by PARID.	Used for structural characteristics in regression analysis.
LEGDATALL.dbf	A table containing legal and tax information for parcels, identified by PARID.	Used for identifying parcel acreage.
PARDATALL.dbf	A table containing the land use code descriptions for each parcel.	Used to identify parcels with single family and semi-detached homes, as well as amenity types.
SALESALL.dbf	A table containing sales entries over the last 50 years for Fairfax County.	Used to identify a valid sale, price of the sale and date of the sale.

The raw sales data set (1,053,440 entries) was filtered based on the conditions shown in Table A.2. These conditions were set to either remove sales that contained data entry errors, such as homes that had negative values for the amount of living area, to reduce the effects of changing market conditions on the data set by taking only sales from 2010 to 2013, and to pull only sales that represented "arms-length" transactions (Taylor et al. 2003). The minimum sales price cutoff of $100,000 was determined by plotting the sales data set, and identifying where groupings of outliers began in the data.

Table A.2. Filters that were applied to the sales data in order to create a more representative sample of market conditions.

ID	Filter	Reasoning
LUC_DESC	Only include sales for single family detached and semi-detached homes.	The scope of this study only includes these types of residences.
SALEDT	Only include sales occurring after January 1st, 2010.	Adequate sample size while providing the most current analysis possible.
SALEVAL_DE	Only include sales that were recorded as "Valid and Verified"	These sales have been designated as being representative of
RMBED	Only include sales that had 1 or more bedroom.	Sales with values less than this were likely the result of data entry error.
SFLA	Only include sales that had at least 100 square feet of living area.[1]	Sales with values less than this were likely the result of data entry error.
TOTBATH	Only include sales that have at least one bathroom.	Sales with values less than this were likely the result of data entry error.
ACRES	Only include sales that have at least 0.1 acres of property.	Sales with values less than this were likely the result of data entry error.
YRBLT	Only include sales that have been built after 1700.	Sales with values less than this were likely the result of data entry error.
PRICE	Only include sales that sold for at least $100,000.	Sales with values less than this were likely a non-market sale.
PARID	For repeat sales of the same property, only include the most recent sale.	The regression model only considers the single sale of a property.

Pedestrian regions were identified in order to provide boundaries to pedestrian activity, reduce data edge effects, and to further filter the sales data into a more manageable subset. These regions were determined based on both the presence of sidewalk centerline data and the presence of un-walkable roads. The sidewalk centerline data set provided by Fairfax County, while sufficient for this study, is not entirely complete. Regions of Fairfax County remain to be digitized as can be easily seen in Figure 3.1 (see section 3.4.2). Additionally, it would be inaccurate to assume that pedestrians will attempt to cross high speed roadways where no pedestrian crossings are available. For this reason, high speed roads (>35 mph) along with sidewalk density were combined on a single map to allow for visual delineation of the four pedestrian regions utilized in this study. Only sales and amenities with these regions were included in the analysis.

A.2 The Walkability Model

The walkability model, which is contained in a python script "walk.py", relies on the following derived data sets:

1. Destination points for each filtered amenity, with a PARID identifying attribute.
2. Origin points for each filtered sale, with a PARID identifying attribute and the selected regression parameters for each property.
3. Two cost surfaces: one for the creation of the initial *pedestrian corridor*, and another to allow determination of the *pedestrian path*.

The following sections will attempt to explain how each of these data sets were created and utilized in the overall model, and subsequently describe the walkability model that was developed.

A.2.1 Generating the Amenity and Sales Data Points

Amenities were identified by grouping parcels based on their land use code description (see Table 3.2). The corresponding amenity category numbers of 1 through 10, as is shown in Table A.3, below, were used to determine pedestrian amenity paths in the walkability model. Each amenity feature's parcel was identified using GIS software, and a reference point was place at the parcel's centroid. Once amenities were grouped, these groups were place into their own separate shape files. These points were eventually projected to the nearest sidewalk polygon or edge-of-pavement (EOP) polygon to allow for least cost path analysis to be conducted. Each point was identified using the parcel identification number (PARID). The filtered sales data was also joined with the parcels layer, and in a similar fashion a point was identified and projected for each sale parcel.

Table A.3. Groupings based on land use codes.

Land Use Code Group	Abbreviation	Amenity ID #
Groceries & Supermarkets	groc	1
Restaurants (sit down)	rest	2
Fast Food	ffood	3
Office	office	4
Services (medical, fire, police, laundry, etc.)	serv	5
Retail & hardware	retail	6
Recreational facilities & open space	rec	7
Educational facilities	edu	8
Public & religious places (non-recreational)	pub	9
Public Transit Stop	trans	10
Hotel & Motel	hotel	11
Residential - (Single Family Detached)	ressfd	21
Residential - (Single Family Attached)	ressfa	22
Residential - (Multifamily)	resmf	23
Not Applicable	na	0
Unknown	unk	99

A.2.2 Generating the Cost Surfaces

Two cost surface rasters were generated for each *pedestrian region.* For the purposes of this analysis, these rasters have been identified as "cost surface 1" and "cost surface 2". Both cost surfaces were generated from the sidewalk polygons, the edge-of-pavement polygons, and the gap area polygons. The process for generating each of these rasters was as follows:

1. Buffer out the sidewalk centerlines ½ of the total width as defined by its width attribute. This created the sidewalk polygon layer.
2. Dissolve the major and minor transportation areas together into a single edge-of-pavement polygon layer.
3. Create a gap area polygon layer between the sidewalks and the edge of pavement using the "Create adjoining polygons" tool in ArcMap. The maximum distance for generating a polygon was set to 100 feet.
4. Convert the sidewalk, edge of pavement, and gap area polygon layers into rasters using ArcMap's "Polygon to Raster" tool. Two rasters were created for each surface type depending on which cost surface they were being added to.

5. Combine the rasters into a single cost surface raster using the "Mosaic" tool in ArcMap. Rasters were overlayed on top of each other with, the values of the sidewalk superseding the values of the edge-of-pavement raster, which in turn superseded the values of the gap areas raster.
6. Clip the countywide raster using the 4 *pedestrian region* boundaries and generate four, region specific cost surface rasters.

As a results of this process, 8 cost surface rasters were generated (2 for each region). An example of a "cost surface 2" is shown below in Figure A.1.

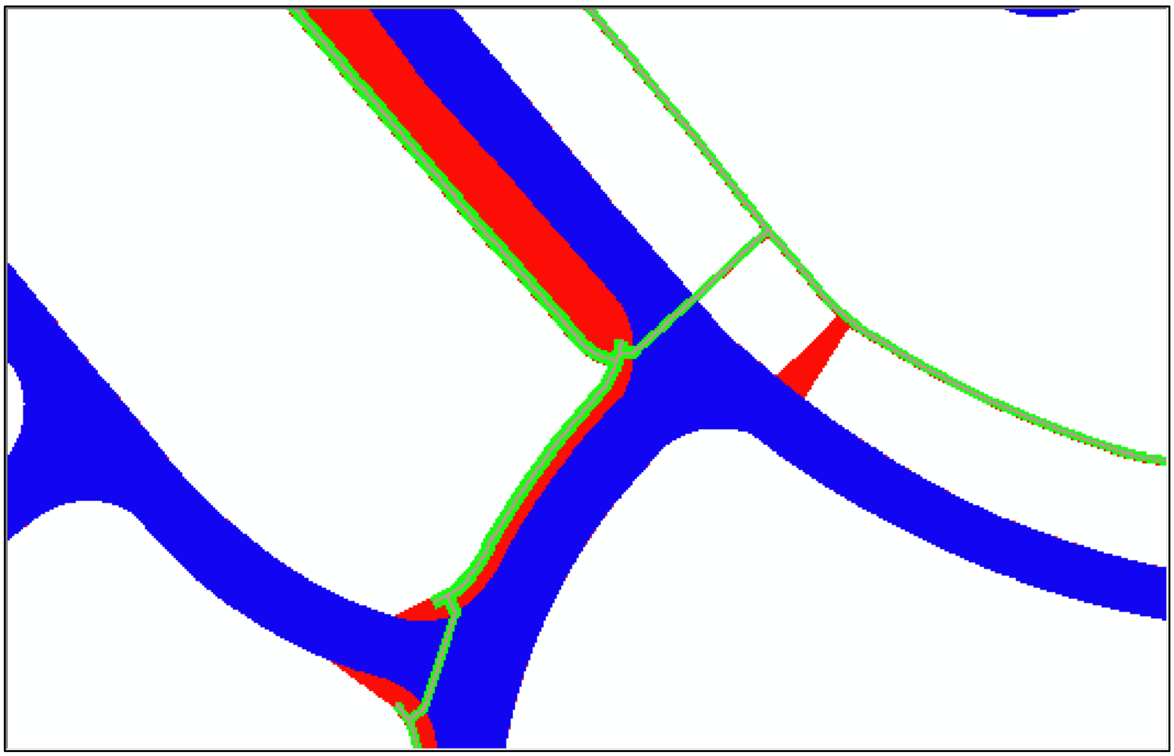

Figure A.1 An example of a cost surface 2 raster, where the sidewalk centerline has a value of 1 (brown lines), the sidewalk has a value of 5 (green), the edge-of-pavement has a value of 50 (blue), and the gap areas have a value of 200 (red).

A.2.3 The walk.py Script

The walkability model was developed by taking the data sources mentioned previously, and inputting them into the different functions sequentially in the walk.py python script. The walk.py script utilizes several dependencies to accomplish the goal of identifying pedestrian paths from a residence (sale) to an amenity in each category. The dependencies are:

1. The python data analysis (or pandas) library. This dependency allows for the use of data frames and allows for the manipulation of large data sets in tabular form.
2. The numerical python (or numpy) library. This dependency allows for matrix calculations to be performed within python.
3. The GeoDataSandbox (dataIO) library. This dependency allows for the direct import of .dbf files into pandas data frames. This is important since the attribute data of the different shape files lives in these .dbf files.
4. The GRASS GIS python scripting (grass) library. This dependency allows python to execute any of the geoprocessing tools that GRASS GIS has available.

Each function within the walk.py script serves an important purpose, and are performed in sequence for a particular amenity category in a specific region. The basic functionality of these different functions are as follows:

1. *setgrassenv()* – This function sets the location and mapset parameters for the GRASS geoprocessing environment. For the purposes of this analysis, the location is always Fairfax County, but the mapset changed depending on which region was being processed.
2. *getpaths()* – This function takes a sales and amenity point data set, as well as cost surface 1 for the region, and generates a line that will become the *pedestrian corridors* from each amenity within the amenity category to any sales within a 1 mile radius of the amenity using least cost path analysis.
3. *getstartend()* – This function identifies the starting and ending points for a particular corridor by taking the line that was generated by *getpaths()* and identifying the amenity and sale that correspond to the start and end points of the corridor. This is necessary since the LCPA tool does not identify the PARID of the source or destination when it conducts its analysis.
4. *getlines()* – This function identifies the shortest *pedestrian corridor* line from the *getpaths()* function for a particular sale, since it is possible that a single sale can exist within 1 mile of multiple amenities of the same category.
5. *getvalues()* – This function takes the shortest *pedestrian corridor* lines from the *getlines()* function and generates the corridor by buffering the line out 100 feet on each side. It then takes the second cost raster, constrains its analysis to this buffered region using a masking raster, and conducts LCPA to generate the *pedestrian path* from the amenity to the sale in question.

6. *getlengths()* – This function breaks apart the *pedestrian path* that is generated by *getvalues()* by surface type using the sidewalk, gap, and edge-off-pavement polygons. It then returns the length of each segment into a table to await further analysis.
7. *runpoi()* – This function just groups all the previous functions into a single command so that they can be easily run for a particular amenity group in a particular region.

At the completion of this script, all the *pedestrian paths* for a particular amenity category in a particular region will be generated, along with a table that provides the distances for each path along each type of walking surface.

A.3 Conducting Hedonic Regression Analysis

A.3.1 Selecting Predictor Variables

Selection of predictor variables in a hedonic regression equation is a difficult task, due to both the common lack of sufficient data and the need to prevent omitted variable bias in the results. The regressors that were selected were chosen because they were either: (1) consistently present in the literature where similar analyses were conducted, (2) found to be statistically significant in explaining the dependent variable based on an initial ANOVA analysis while not contributing to multicollinearity between predictor variables (Table A.4), or (3) one of the walkability variables of interest.

Table A.4 One-way ANOVA analysis of the non-walkability predictor variables for the final data set. All regressors are found to be significant to the 1 % level.

	Df	Sum of Sq	Mean of Sq	F Value	p
SFLA	1	269.75	269.75	17017.01	2.00E-16
AGE	1	18.25	18.25	1151.02	2.00E-16
ACRES	1	4.53	4.53	285.76	2.00E-16
TOTBATH	1	12.02	12.02	758.26	2.00E-16
RMBED	1	0.72	0.72	45.32	1.85E-11
MEDINC	1	4.3	4.3	271.19	2.00E-16
Residuals	5346	84.74	0.02		

The distances provided by the walkability script were translated into dummy (binary) variables based on the percentage of the total path that was along pedestrian surfaces (sidewalks and trails). Different ranges were set for the 4 different models based on at least 0 %, 25 %, 50 % and 75 % of the path existing along pedestrian surfaces (see section 3.4.3).

A.3.2 Conducting Spatial Regression Analysis

It is common for regression analyses involving real estate to exhibit spatial autocorrelation. In order to ensure that the results of the regression were not spatially biased, a Moran's I test was conducted using the Spatial Autocorrelation tool in ArcMap. The tool was run against the prices of all 5,353 properties in the final point shape file. The results showed that the calculated value for Moran's I ($I = 45.55$, $p < 0.001$) showed significant positive spatial autocorrelation in the dependent variable for the data set. Therefore, two spatial models that help compensate for this type of correlation were considered.

The spatial lag and spatial error models were computed using the python spatial analysis (pysal) library (Rey 2014). A weighted *nxn* matrix *W* was generated using pysal commands for the data set. *W* was calculated to be a 5353 x 5353 matrix for the 5353 sales entries that was first populated with the Euclidean distances between each pair of points that fell within a minimum threshold, which ensured each point had at least 1 neighbor. The matrix was then row-standardized by dividing 1 by each value across the matrix.

To determine which spatial model is appropriate, robust Lagrange multiplier (LM) values were calculated for the base case of both the spatial lag ($LM = 272.56$, $p < 0.001$) and spatial error ($LM = 3435.11$, $p < 0.001$) models. These results suggest an application of both spatial models are needed. The results of these spatial methods applied to the four different models are shown in Table 3.3.

B Additional Tables

Table B.1 Land use code descriptions with the assigned amenity category

Land Use Code Description	**Count**	**Amenity ID**
Single-family, Detached	189965	21
Townhouse in ownership development	81138	22
Garden Apartments condominium (=<4story)	29519	23
Multiplex in condominium development	10126	23
Vacant Land	9669	0
Private open space(planned development)	6938	0
High rise apartments condo(=>9 comm)	5990	23
High rise apartments condo(=>9 no comm)	4058	23
Condominium Office (< = 4 stories)	3156	4
Townhouse in condominium development	2904	22
Duplex, either vertical or horizontal	2263	22
Recreation Fac,Parks(govt) - outdoor	1493	7
Medium rise apartments condo(5to8 stry)	1224	23
Low Rise Office(< = 4 stories)	853	4
Whsle,wrhsing \& stg (not in IP/in condo)	827	99
Wholesale,warehousing \& stg (not in IP)	744	99
Garden Apartments rental (=<4 story)	689	23
Churches, Synagogues	454	9
Other public NEC	444	0
General med/hi rise off (= > 5 stories)	429	4
Two or more Single-family, detached	340	22
Community Center	273	9
Public Schools	245	8
Specialty Center	215	9
Other Retail NEC(not in shopping center)	194	6
Condominium Medical (< = 4 stories)	183	5
Condominium Boat Slips -private for sale	170	0
Medium rise apartments rental(5to8 stry)	170	23
Medical office (= > 5 stories)	159	5
Improved Land w dilapidated structure	158	0
Single-family, Semidetached,garden court	158	22
Condo office (= > 5 stories)	154	4
Gasoline and Service Station	150	0
Garage,barn,outhouse,shed adj prcl unit	146	0
Cluster Office (< = 4 stories)	141	4
Restaurant with alcohol	140	2

Neighborhood Center	132	9
Converted Residential office(exdwelling)	131	4
Single-family residences inf com/ind	127	21
Motor vehicle sales (new and used)	113	0
Carry-out with seating	103	3
Nursery Schools	101	8
Cemeteries	101	9
Other Industrial NEC	92	0
Water,pipeline ROW,plants,storage,etc.	90	0
Other automotive,marine,aircraft and NEC	89	0
Condo Center	87	23
Motor vehicle repair separately	85	5
Private Schools	83	8
Mini-Warehouses (not in IP)	79	99
Finance,insurance,real estate services	78	5
Convenience grocery	78	1
Townhouse in rental development	72	22
Electric,transmission ROW,plants,substat	70	0
Private open space(not planned develop)	68	0
Comm Use in Res Condo Dev	61	99
Medical/dental low rise (< = 4 stories)	56	5
Gasoline Sale Only	55	0
COMMUNITY SWIMMING POOL	48	7
Railroad,ROW,terminals,maintenance	48	10
Golf Courses (private)	45	7
Other residential NEC	45	0
Veterinary hospitals	43	5
Motel without restaurant \& other comm	43	11
Hotel with restaurant \& other comm	42	11
Auto parking	41	0
Civil,social,Fraternal, Prof \& Bus Assoc	41	99
Super Regional Center	41	99
Fire \& Rescue Stations	40	5
Research \& Testing(not in IP/not in off)	39	99
Promotional Center	38	99
Rail rapid transit,ROW,terminals,maint	37	10
Discount Store	36	6
Furniture, house furnishings	34	6
Gasoline Sales and Car Wash	33	0
Post Offices	33	5
Personal services (laundry,photo,beauty)	32	5

Hotel without restaurant \& other comm	32	11
Sewage,plants,etc	32	0
Other office NEC	32	4
Durable Manufacturing (not in Ind Park)	30	0
Telephone \& Telegraph	28	0
Department Store	28	6
Restaurant without alcohol	28	2
Town Center	28	99
Building Materials,Hardware, Farm Equip	27	6
Condo Retail (in office/Indust complex)	27	6
Government owned low rise(< = 4 stories)	27	4
Other consumer/business services NEC	26	99
Regional Center	25	99
Drug stores	25	6
Recreation Fac,Parks (public) - indoor	24	7
Gov owned med/hi rise(= > 5 stories)	24	4
Golf Courses (government-owned)	23	7
Single-family structure NEC	23	21
Golf Courses (commercial)	23	7
Nursing homes	22	5
Permanent Exhibition	22	99
Two-family NEC	22	22
Radio \& Television	21	0
Gas,pipeline ROW,plants,storage,etc.	21	0
Motel with restaurant \& other comm	21	11
Hospital \& Health Facilities	20	5
Mobile homes in park or court	20	0
Supermarket	19	1
Retirement homes \& orphanages	18	23
Sand \& Gravel Quarrying	18	0
Contract Construction (not in IP)	18	0
Street and highway ROW	18	0
Libaries	17	5
Combination of Structure types	17	99
Other group quarters NEC (not Military)	17	23
Apparel and accessories	15	6
Nondurable Manufacturing(not in IP)	15	0
Horticulture Activities \& services	15	5
Planned industrial park	15	0
Research \& Testing(not in IP/in condo)	14	99
Industrial conglomeration	11	0

Carry-out Kitchen	11	3
College,Universities	10	8
Other food NEC (include fruit,meat,fish)	10	1
High rise apartments rental(=>9 comm)	10	23
Other Educational Services NEC	9	8
Welfare \& Charitable services	9	5
Recreation Fac,Parks(public)-outdoor	9	7
Med/dental med/hi rise(= > 5 stories)	8	5
Religious quarters	8	9
Printing \& Publishing	7	5
Recreation Fac,Parks (govt) - indoor	7	7
High rise apartments rental(=>9 no comm)	7	23
Other repair services NEC	7	5
Permanent Conservation area,wildlife	7	0
Other communications, NEC	6	0
Public Assembly, Both Indoor \& Outdoor	6	9
Recreation Fac,Parks (private)-outdoor	6	7
Apartment, NEC including cooperatives	6	23
Other resources uses NEC	6	0
Service Station out of operation	4	0
Pipeline ROW and NEC (petroleum)	4	0
Government leased low rise(<= 4 stories)	4	4
Other utilities, NEC	4	0
Police Stations	4	5
Military Institutions	3	0
Conservation Easement	3	0
Swimming pools - outdoor	3	7
Recreation Fac, Parks (private)-outdoor	3	7
Supermarket plus general merchandise	3	1
Other cultural \& entertainment NEC	2	7
Office Park	2	4
Air,runways,terminals and maintenance	2	0
Agricultural Activities \& services	2	0
Rooming \& Boarding Houses	2	11
Special Training Schools	2	8
Tourist Home	2	0
Townhouse or Multiplex NEC	1	22
Marine terminals	1	0
Multiplex in rental development	1	23
Variety or junior department stores	1	6
Gov leased med/hi rise(= > 5 stories)	1	4

Table B.2 Basic statistics for the > 0 % data set, dummy variables

	count	mean	std	min	25%	50%	75%	max
edu	5353	0.727443	0.445317	0	0	1	1	1
ffood	5353	0.126471	0.332411	0	0	0	0	1
groc	5353	0.158603	0.365339	0	0	0	0	1
office	5353	0.291239	0.454376	0	0	0	1	1
pub	5353	0.710443	0.453599	0	0	1	1	1
rec	5353	0.840837	0.365862	0	1	1	1	1
rest	5353	0.098076	0.297445	0	0	0	0	1
retail	5353	0.167756	0.373685	0	0	0	0	1
serv	5353	0.409677	0.49182	0	0	0	1	1

Table B.3 Basic statistics for the > 25 % data set, dummy variables

	count	mean	std	min	25%	50%	75%	max
edu	5353	0.720344	0.448872	0	0	1	1	1
ffood	5353	0.12535	0.331147	0	0	0	0	1
groc	5353	0.156361	0.363231	0	0	0	0	1
office	5353	0.28881	0.453252	0	0	0	1	1
pub	5353	0.702783	0.457075	0	0	1	1	1
rec	5353	0.829441	0.376158	0	1	1	1	1
rest	5353	0.097515	0.296686	0	0	0	0	1
retail	5353	0.166449	0.372518	0	0	0	0	1
serv	5353	0.402952	0.490537	0	0	0	1	1

Table B.4 Basic statistics for > 50 % data set, dummy variables.

	count	mean	std	min	25%	50%	75%	max
edu	5353	0.68541	0.464396	0	0	1	1	1
ffood	5353	0.118625	0.323377	0	0	0	0	1
groc	5353	0.147581	0.354717	0	0	0	0	1
office	5353	0.283766	0.450867	0	0	0	1	1
pub	5353	0.668971	0.470628	0	0	1	1	1
rec	5353	0.793574	0.404778	0	1	1	1	1
rest	5353	0.09023	0.286537	0	0	0	0	1
retail	5353	0.155987	0.362877	0	0	0	0	1
serv	5353	0.377732	0.484865	0	0	0	1	1

Table B.5 Basic statistics for > 75 % data set, dummy variables.

	count	mean	std	min	25%	50%	75%	max
edu	5353	0.596114	0.490721	0	0	1	1	1
ffood	5353	0.097702	0.296939	0	0	0	0	1
groc	5353	0.13301	0.339617	0	0	0	0	1
office	5353	0.263964	0.440821	0	0	0	1	1
pub	5353	0.569214	0.495233	0	0	1	1	1
rec	5353	0.691201	0.462041	0	0	1	1	1
rest	5353	0.083131	0.276106	0	0	0	0	1
retail	5353	0.138053	0.344988	0	0	0	0	1
serv	5353	0.322436	0.467453	0	0	0	1	1

Table B.6 Basic statistics for non-walkability regressors

	acres	age	medinc	price	rmbed	totbath	sfla
count	5353.000000	5353.000000	5353.000000	5353.000000	5353.000000	5353.000000	5353.000000
mean	0.279651	34.984868	128155.087241	522133.426863	4.012330	3.258173	2044.379040
std	0.132288	14.696848	34006.377373	149756.839706	0.698759	0.923777	792.845033
min	0.058195	1.000000	36697.000000	204370.000000	2.000000	1.000000	1000.000000
25%	0.215100	26.000000	105642.000000	420000.000000	4.000000	3.000000	1409.000000
50%	0.251476	34.000000	122216.000000	500000.000000	4.000000	3.000000	1914.000000
75%	0.302991	45.000000	150350.000000	592500.000000	4.000000	4.000000	2430.000000
max	2.498771	140.000000	194241.000000	1610000.000000	8.000000	8.000000	6239.000000

C Python Scripts

C.1 walk.py

```
#!/usr/bin/env python

# Import necessary modules
import os
import dataIO as di
import pandas as pd
import grass.script as g
import grass.script.setup as gsetup
import numpy as np

# Function to setup GRASS environment
def setgrassenv(location, mapset):
   """Set the (location, mapset) for the GRASS Environment."""

   gisbase = os.environ['GISBASE']
   gisdb = "/Users/Liam/Desktop/thesis/analysis/grass"
   gsetup.init(gisbase, gisdb, location, mapset)
   g.run_command('g.message', message='GRASS environment set')

# Function to get least cost paths around point of interest
def getpaths(sales, poi, surface):
   """Set the (starting points, ending points, first cost surface)."""
   g.run_command('g.region', rast=surface)
   info = g.parse_command('v.info',flags='t', map=poi, quiet=True)
   n = int(info['points'])
   x = 0
   poiName = poi.split("_")[1]

   g.run_command('g.remove', vect=poiName + 'vFinalOut')
   g.run_command('v.edit',map=poiName + 'vFinalOut',tool='create')

   # Loop through end points and create paths
   for i in range(n):
      g.run_command('g.message', message='On feature ' + str(x+1) + ' of ' + str(n))
      g.run_command('g.region',rast=surface)
      g.run_command('v.extract',input=poi,
         list=i+1,output='currentPoi',type='point',overwrite=True)
      g.run_command('v.buffer',input='currentPoi',output='poiBuff',
         distance=5280,overwrite=True)
      g.run_command('g.region',vect='poiBuff')
      g.run_command('r.cost',input=surface,output='costDist1',outdir='costBacklink1',
         start_points='currentPoi',max_cost=5280,overwrite=True)
      g.run_command('r.drain',flags='d',input='costDist1',
         indir='costBacklink1',output='path1',
         voutput='tmpVOut',overwrite=True,vector_points=sales)
      g.run_command('v.category',input='tmpVOut',output='tmpVOutCat',option='sum',
         cat=x,overwrite=True)
      g.run_command('v.patch',input='tmpVOutCat',output=poiName + 'vFinalOut',
         flags='a',overwrite=True)
      x += 1

# Function to get start and end points for paths
def getstartend(sales, poi):
   """Set the (starting points, ending points)."""
   g.run_command('g.region', vect='boundary')
   info = g.parse_command('v.info',flags='t', map=poi, quiet=True)
   n = int(info['points'])
   poiName = poi.split("_")[1]
   g.run_command('g.region', vect='boundary')

   # Get start and end points
   g.run_command('v.category',input=poiName + 'vFinalOut', output='vFinalOutDel',
   option='del', overwrite=True)
   g.run_command('v.category',input='vFinalOutDel', output='vFinalOutCat',
```

```
    option='add', type='line', overwrite=True)
    g.run_command('v.db.addtable', map='vFinalOutCat', columns='cat integer')
    g.run_command('v.db.addcol', map='vFinalOutCat', columns='pathLength double precision')
    g.run_command('v.to.db', map='vFinalOutCat', option='cat', columns='cat')
    g.run_command('v.to.db', map='vFinalOutCat', option='length', columns='pathLength',
      units='me')
    g.run_command('v.extract', flags='r', input='vFinalOutCat',
      output=poiName + 'vFinalOutExtr', where='pathLength=0', overwrite=True, type='line')
    g.run_command('v.to.points', flags='n', input=poiName + 'vFinalOutExtr',
      output='vFinalOutPoints', overwrite=True)
    g.run_command('v.extract', input='vFinalOutPoints', output='vStartPoints',
      layer=2, where='along=0', overwrite=True, type='point')
    g.run_command('v.extract', flags='r', input='vFinalOutPoints', output='vEndPoints',
      layer=2, where='along=0', overwrite=True, type='point')
    g.run_command('v.db.addcol', map='vStartPoints', layer=2, columns='salesCat integer')
    g.run_command('v.db.addcol', map='vEndPoints',  layer=2, columns='poiCat integer')
    g.run_command('v.distance', _from='vStartPoints', to=sales, from_layer=2,
      upload='cat', column='salesCat', overwrite=True)
    g.run_command('v.distance', _from='vEndPoints', to=poi, from_layer=2, upload='cat',
      column='poiCat', overwrite=True)
    g.run_command('g.message', message='Get start end completed.')

# Function to get the lines to be used
def getlines(poi):
    """Set the ()."""
    g.run_command('g.region', vect='boundary')
    poiName = poi.split("_")[1]
    mapset = g.parse_command('g.gisenv',get='MAPSET')
    mapset = mapset.keys()[0]

    # Extract the shortest path lines
    dbfroot = "/Users/Liam/Desktop/thesis/analysis/grass/Fairfax/"+mapset+"/dbf/"
    lines = di.dbf2df(dbfroot + poiName + 'vFinalOutExtr.dbf')
    start = di.dbf2df(dbfroot + 'vStartPoints.dbf')
    sales = di.dbf2df(dbfroot + 'sales_since2010.dbf')
    lines.columns = ['lcat','pathLength']
    start = start.drop(['along','cat'],axis=1)
    start = start.set_index('lcat')
    lines = lines.set_index('lcat',drop=False)
    joined = lines.join(start)
    joined = joined.sort('pathLength')
    joinedGroups = joined.groupby('salesCat')
    linesUsed = joinedGroups.first()
    linesUsed = linesUsed.sort('lcat')
    sales = sales.set_index('cat')
    linesSales = linesUsed.join(sales)
    linesUsed['lcat'].to_csv("/Users/Liam/Desktop/thesis/analysis/tables/" + mapset + "_" + poiName + "_lines.txt", index = False)
    linesSales.to_csv("/Users/Liam/Desktop/thesis/analysis/tables/" + mapset + "_" + poiName + "_sales.csv")
    g.run_command('v.extract', input=poiName + 'vFinalOutExtr',
      output=poiName + 'vFinalOutUsed', file="/Users/Liam/Desktop/thesis/analysis/tables/" + mapset + "_" + poiName + "_lines.txt",
      overwrite=True)
    g.run_command('g.message', message='Get lines completed.')

# Function to get values
def getvalues(sales, poi, surface):
    """Set the (start points, end points, second cost surface)."""

    # Set region and variables
    mapset = g.parse_command('g.gisenv',get='MAPSET')
    mapset = mapset.keys()[0]
    poiName = poi.split("_")[1]
    g.run_command('g.region', rast=surface)
    infeat = poiName + 'vFinalOutUsed'
    outfeat = poiName + 'vFinalOut2'
    dbfroot = "/Users/Liam/Desktop/thesis/analysis/grass/Fairfax/"+mapset+"/dbf/"
    lcats = g.parse_command('v.db.select', flags='c', map=infeat, column='cat')
    lcats = lcats.keys()
    lcats = map(int, lcats)
    lcats.sort()
```

```
    # Set final output
    g.run_command('g.remove', vect=outfeat)
    g.run_command('v.edit',map=outfeat,tool='create')
    count = 0

    # Loop through output
    for i in lcats:
        count += 1
        g.run_command('g.message',
            message='On feature ' + str(count) + ' of ' + str(len(lcats)))
        g.run_command('g.region',rast=surface)
        g.run_command('v.extract',input=infeat,list=i,
            output='tmpline',type='line',overwrite=True)
        startPoint = g.parse_command('v.db.select', flags='c', map='vStartPoints',
            column='salesCat', where='lcat='+str(i),layer=2)
        startPoint = startPoint.keys()[0]
        endPoint = g.parse_command('v.db.select', flags='c', map='vEndPoints',
            column='poiCat', where='lcat='+str(i),layer=2)
        endPoint = endPoint.keys()[0]
        g.run_command('v.extract',input=sales, list=startPoint,
            output='tmpStart',type='point',overwrite=True)
        g.run_command('v.extract',input=poi, list=endPoint,
            output='tmpEnd',type='point',overwrite=True)
        g.run_command('v.buffer',input='tmpline',output='tmpbuff',
            distance=100,tolerance=10,overwrite=True)
        g.run_command('g.region',vect='tmpbuff')
        g.run_command('v.to.rast',input='tmpbuff',output='maskRast',use='val',
            overwrite=True)
        g.run_command('r.mask',input='maskRast',flags='o')
        g.run_command('r.cost',input=surface,output='costDist',outdir='costBacklink',
            start_points='tmpEnd',overwrite=True)
        g.run_command('r.drain',flags='d',input='costDist',
            indir='costBacklink',output='path',
            voutput='tmpVOut',overwrite=True,vector_points='tmpStart')
        g.run_command('r.mask',flags='r')
        g.run_command('v.category',input='tmpVOut',output='tmpVOutCat',option='add',
            cat=startPoint,overwrite=True)
        g.run_command('v.patch',input='tmpVOutCat',output=outfeat,flags='a',
            overwrite=True)

    # Categorize output
    g.run_command('v.category',input=outfeat, output='vFinalOutDel',
        option='del', overwrite=True)
    g.run_command('v.category',input='vFinalOutDel', output=outfeat,
        option='add', type='line', overwrite=True)
    g.run_command('v.db.addtable', map=outfeat, columns='cat integer')
    g.run_command('v.db.addcol', map=outfeat, columns='fpLength double precision')
    g.run_command('v.to.db', map=outfeat, option='cat', columns='cat')
    g.run_command('v.to.db', map=outfeat, option='length', columns='fpLength',
        units='me')

def getlengths(poi):
    g.run_command('g.region', vect='boundary')
    mapset = g.parse_command('g.gisenv',get='MAPSET')
    mapset = mapset.keys()[0]
    poiName = poi.split("_")[1]
    infeat = poiName + 'vFinalOut2'
    dbfroot = "/Users/Liam/Desktop/thesis/analysis/grass/Fairfax/"+mapset+"/dbf/"
    lines = di.dbf2df(dbfroot + infeat + ".dbf")
    lines = lines.set_index('cat')

    for area in ['sw_areas','eop_areas','cw_areas']:
        areaMap = poiName+area.split('_')[0]+"length"
        aLength = area.split('_')[0] + "Length"
        g.run_command('v.overlay',ainput=infeat,atype='line',binput=area,
            output=areaMap,operator='and',overwrite=True)
        g.run_command('v.db.addcol', map=areaMap, columns=aLength + ' double precision')
        g.run_command('v.to.db', map=areaMap, option='length', columns=aLength,
            units='me')
        areaDf = di.dbf2df(dbfroot + areaMap + ".dbf")
```

```
        areaGroup = areaDf.groupby('a_cat')
        areaSum = areaGroup[aLength].sum()
        areaSum.columns = [aLength]
        lines = lines.join(areaSum)

    sales = pd.read_csv("/Users/Liam/Desktop/thesis/analysis/tables/" + mapset + "_" + poiName + "_sales.csv")
    sales.index += 1
    lines = lines.fillna(0)
    lines.swLength = lines.swLength - lines.cwLength
    lines.eopLength = lines.eopLength - lines.cwLength
    lines['gapLength'] = lines.fpLength - lines.swLength - lines.eopLength - lines.cwLength
    lines = np.round(lines, 2)
    lines = lines.join(sales)
    lines.to_csv("/Users/Liam/Desktop/thesis/analysis/tables/" + mapset + "_" + poiName + "_lengths.csv")

def runpoi(location,mapset,sales,poi,surface1,surface2):
    setgrassenv(location,mapset)
    getpaths(sales,poi,surface1)
    getstartend(sales,poi)
    getlines(poi)
    getvalues(sales,poi,surface2)
    getlengths(poi)
    g.run_command('g.message', message=poi + " complete.")
```