Identifying and characterizing walkable environments in Vancouver, Washington

Tree Casiano, Seth Frazier, Johnathon Pantzlaff

May 4, 2017

Introduction

Before the advent of cars, humans traveled on foot as their primary mode of transportation, and neighborhoods were inherently walkable, with community amenities accessible to pedestrians by default. The car-dependent, sprawling suburbs that developed in the second half of the twentieth century have become the norm in the United States. Much of the infrastructure designed during this time did not take into account the needs of pedestrians, and it is not uncommon to find housing developments and newer neighborhoods devoid of sidewalks, street lights, or basic amenities like grocery stores or post offices within walking distance.

The 20-Minute Neighborhood concept characterizes walkable neighborhoods as those having "direct, obvious and safe routes with frequent connections to attractive destinations—places to which people need and want to go" (City of Portland Bureau of Planning and Sustainability, 2009, p. 3). Walkable neighborhoods cultivate community, sustainability, and physical fitness and are therefore of interest to city planners, public health professionals, and residents (City of Portland Bureau of Planning and Sustainability, 2009). Many people want to live somewhere that does not make them entirely car-dependent, as evidenced by popular real estate websites like RedFin and Zillow including a Walk Score index on their property listings. The Walk Score is a proprietary measure of walkability based on proximity to amenities such as restaurants, stores, and schools, access to transit, and pedestrian infrastructure (Vanderbilt, 2012).

Objectives

Our project was inspired by the Walk Score and is an attempt to create our own measure of walkability. We will look specifically at walkability in Vancouver, Washington, a city in the Pacific Northwest that is part of the rapidly growing Portland–Vancouver–Hillsboro Metropolitan Statistical Area (Christensen, 2015). We will use free and open data sources to create a PostgreSQL database that focuses on one aspect of walkability: proximity to amenities along the road network. Data for other factors that influence walkability, such as the quality or existence of sidewalks and street-lighting, could not be easily obtained and would result in a project of a scope larger than intended for this project.

Our datasets include OpenStreetMap (OSM) points of interest, TIGER/Line Shapefiles for public roads and census blocks, and median household income data from the American Community Survey for Clark County, Washington. We will generate service areas (polygons representing a 20-minute walk along roads at 4 km/h) for every relevant point of interest (POI) in OSM, using ArcGIS's Network Analyst in conjunction with public road data. Examples of the hundreds of different types of POI found in OSM data include bus stops, coffee shops, grocery stores, libraries, public parks, recreational facilities, restaurants, and retailers.

OSM's geographic data is crowdsourced and maintained by volunteer efforts much like open source software is maintained by community participants. In larger cities, coverage is better and the spatial data quality is generally higher because there are potentially more people available to keep it updated or to make corrections. Our assumption is that Vancouver data is well maintained because it is included in the area covered by an active OSM Local Chapter for Portland, Oregon and Southwest Washington (EliL, 2015).

Using PostGIS, we will perform intersect functions between these service areas and each census block to determine the number and importance of the POIs that are within the required walk time to any given block. Income data will be obtained for each block by joining the non-spatial American Community Survey median income by block group relation with the spatial census block relation. Analyses will then be conducted to correlate the number/importance of POIs within walking distance of blocks and the median income of such blocks.

Our database will be used to generate static maps and charts that answer the following questions:

1) *What areas in Vancouver have the greatest pedestrian accessibility to points of interest, including bus stops?*
2) *Which environments are most conducive to outdoor physical fitness, defined as having parks and recreational facilities within walking distance?*
3) *What is the correlation between the median neighborhood income and walkability?*

We suspect that walkability will be high in very few areas, with most of Vancouver and its adjacent communities exhibiting the low to occasionally-moderate walkability typical of modern suburban development in the United States. We also expect that our attempts to correlate median income and walkability will yield a positive correlation overall, i.e. the most walkable areas will be largely inhabited by those with relatively high incomes. It is also likely that the greatest pedestrian access to parks and other recreational facilities will occur in wealthier neighborhoods, with a positive correlation that could be even more significant than that between basic walkability and median income.

Database Design: Conceptual and Logical Modeling

We used the ERDPlus online modeling tool to create Entity-Relationship diagrams consisting of five entities, depicted in Figure 1: `poi`, `poi_weight`, `poi_service_areas`, `household_income`, and `block`.

poi

Each instance of the `poi` entity is an OSM point of interest with three attributes: a `gid` attribute that serves as a unique identifier and primary key, a `geom` attribute for the point geometry, and a `type` attribute based on the amenity types imported from OSM (“amenity”, “highway”, “leisure”, “shop”, and “tourism”). The `type` attribute serves as a foreign key to the `poi_weight` entity which holds the values of walkability scores assigned to each amenity type.

poi_weight

The `poi_weight` entity represents a table where each row is one of the OSM amenity types, identified with the primary key of `type`. Every amenity type in this table has two attributes: a score for outdoor recreation (`park_weight`) and one for general walkability (`general_weight`). These scores represent the amenity type's relative importance as a desirable recreational destination or as a useful destination for day-to-day living, respectively, based on an average of weights we assigned as individuals (see appendix).

poi_service_areas

`poi_service_areas` represent the point of interest service areas generated from each point of interest in our OSM dataset. The primary key in the `poi_service_areas` entity is the same number as the primary key in the `poi` entity, with which it has a one-to-one relationship. The spatial (polygon) attribute, `geom`, is the only other attribute.

household_income

The `household_income` entity is non-spatial, and has only two attributes: a primary key (`id`) based on the census block group with which it is associated and the `income` attribute for its median household income.

block

The block entity is a pared down version of a census block shapefile and has attributes for the geometry (`geom`), a primary key (`gid`) used by the Census to identify the unique blocks, and an attribute called `block_group` that can be used as a foreign key relating to the `id` attribute the `household_income` entity.

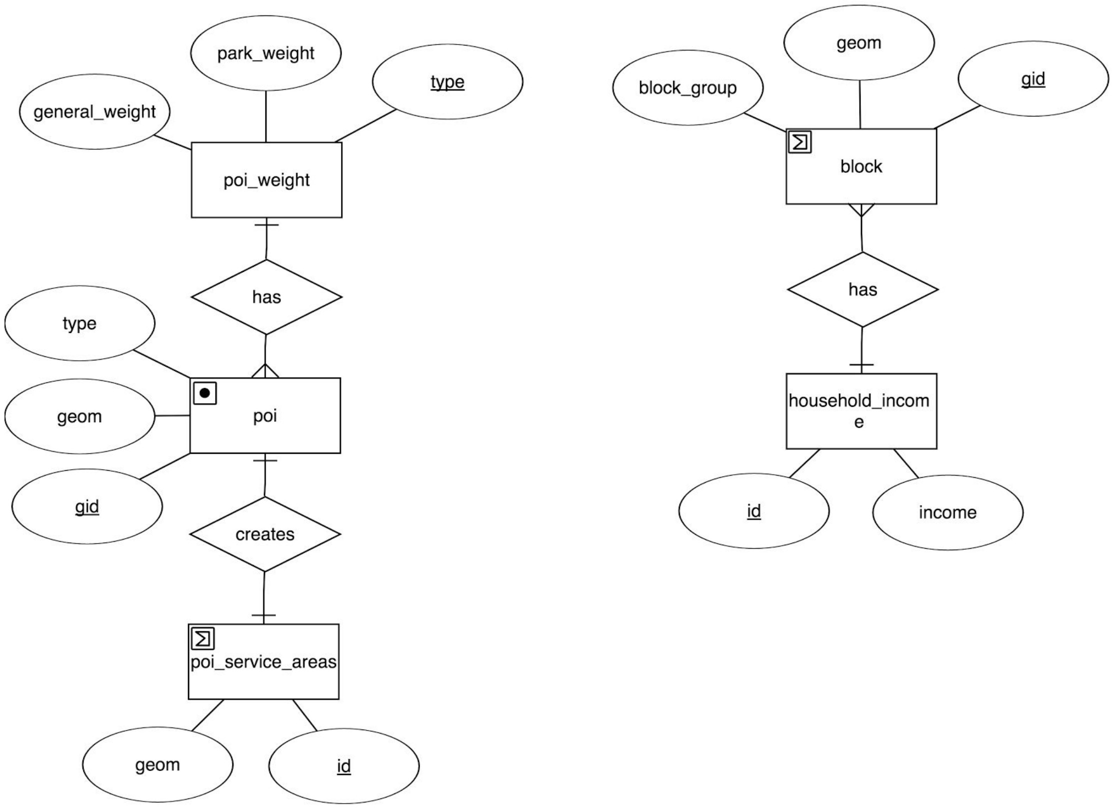

Figure 1. Conceptual diagram.

The foreign and primary key relationships and the data types for each attribute in our logical model are diagrammed in Figure 2.

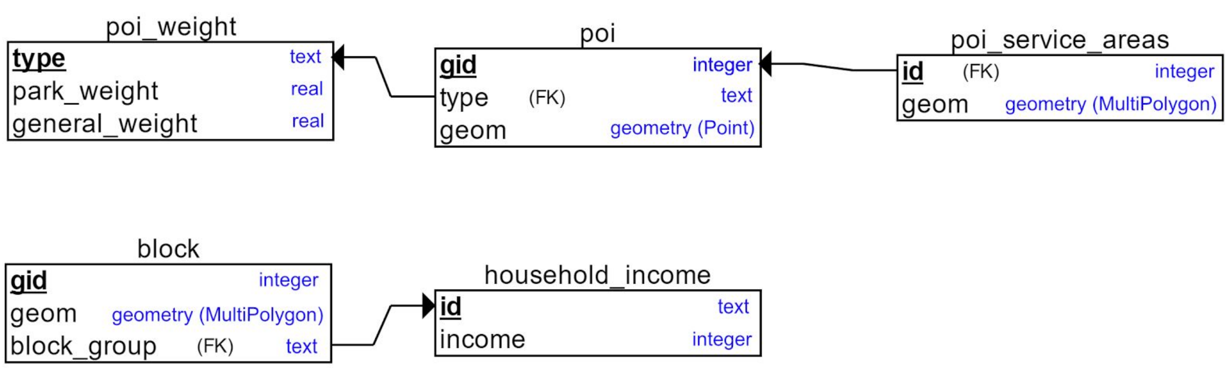

Figure 2. Logical model with PostgreSQL data types.

Data Collection, Data Cleaning, and Database Creation

Unless otherwise noted, all the tables with a spatial component (`block`, `poi`, `poi_service_area`) were first created in QGIS using the QGIS DB manager because it properly and easily sets up spatial columns and indexes. When necessary, pgAdmin was used to modify the table's columns and data types. QGIS was also used to copy the input shapefiles, reproject to EPSG:2856 as necessary, and import them into the target tables in our PostgreSQL database.

Table "block"

The data in the "block" table is the census block shapefile for all blocks in Clark County, Washington. It was imported using shp2pgsql. Irrelevant columns were removed using pgAdmin. The primary key is the US Census Bureau Geographic Identifier (geoid), a value assigned to every geographic entity about which the Census Bureau collects data. For census blocks, the geoid is a 15-character string which can also be used to identify the block group and census tract in which the block is found. The first 12 characters are the block group.

To create a "block_group" column that could serve as the foreign key to the table containing household income data per block group, we manipulated the geoid attribute of the census block shapefile with the SQL the statement `UPDATE block SET block_group = SUBSTRING(geoid10 from 0 for 13);`, which removed the last 3 characters of the original 15-character string.

Table "household_income"

The data for household income is from the American Community Survey, retrieved from the US Census Bureau's American FactFinder website in the form of a CSV file for the median income by census block groups for Clark County, Washington (United States Census Bureau, 2016). Block groups are the smallest geographic unit for which the data were available. The CSV file was opened with Excel and stripped of the irrelevant columns, leaving just the "GEO.id2" and "Estimate" columns. Values in the "GEO.id2" column had been rendered in scientific notation by Excel, so they were converted to standard notation before being exported as a CSV file and imported into a blank table (`household_income`) in our database with the following statement:

```
COPY household_income FROM '{path to CSV here}' DELIMITER ','
CSV;
```

Our logical model shows that `block_group` in the `block` table is a foreign key to the `id` attribute in `household_income`; however, we were unable to implement this foreign key relationship because there appears to be a small amount of income data missing, prohibiting the creation of the relationship.

Table "poi"

To collect the points of interest data, we used the QGIS OpenStreetMap import tool to pull in raw OSM data that covered the study area in general. This raw XML data was converted to a spatially enabled SQLite database in QGIS consisting of three separate tables as points, lines, and polygons, all of which were filtered by the presence of some value in the fields "amenity", "highway", "leisure", "shop", or "tourism" (i.e. if the values of all five of these columns were null, the feature was deleted). In order to conduct a walk time analysis on the points of interest, we needed all features to be, geometrically speaking, points. Any features that were not points were collapsed into points using the "Mean coordinates" processing tool for the lines and the "Polygon Centroids" processing tool for the polygons. These three separate sets of point data were then coalesced into a single set of data, and every column not named "amenity", "highway", "leisure", "shop", or "tourism" was deleted. The data were then clipped to the study area, defined as the area within the cities and Census-designated places of Baberton, Camas, Felida, Five Corners, Hazel Dell, Lake Shore, Minnehaha, Orchards, Salmon Creek, Vancouver, Walnut Grove, and Washougal, Washington (United States Census Bureau, 2016), and an additional 1333 meters (as this is how far one can walk in 20 minutes at 4 km/h). This point data set was then imported into the database, where rows that had values in more than one of the five abovementioned columns were found manually and values deleted to reduce the number of columns with a value down to one. The values in these five columns (four of which were then blank for any given feature) were then coalesced into a new "type" column with the statement

```
UPDATE poi SET type = CONCAT(amenity, highway, leisure, shop,
tourism);
```

The "amenity", "highway", "leisure", "shop", and "tourism" columns were then deleted.

Table "poi_weight"

This non-spatial table began with the query `SELECT DISTINCT type FROM poi;`. These query results were then pasted into a shared Google Sheets spreadsheet that was further manipulated to allow each project team member to cast a vote on how valuable he or she thought a feature of this type was to live by. Assigning the weights in an objective way posed a problem. Some types of amenities are more important to live near than others, and what those amenities are will largely depend on an individual's lifestyle. Short of doing a large-scale survey, which would be beyond the scope of this project, doing a survey among the three of us seemed like the best way to obtain weights with reasonable effort.

After voting was completed, the votes were averaged, resulting in a spreadsheet with three columns: `type`, `general_weight`, and `park_weight` (see appendix). This spreadsheet was exported as a CSV file and imported into the database with the statement

```
COPY poi_weight FROM '{path to CSV here}' DELIMITER ',' CSV;
```

Table "poi_service_areas"

Network Analyst tools in ArcGIS were used to generate the data for our point of interest service areas. First, we created a network in ArcGIS against which walk time analysis could be performed, the basis of which was a roads data set produced by the Census Bureau (United States Census Bureau, 2010). This dataset was initially imported into a file geodatabase feature class. Roads that are limited-access, such as freeways, were manually removed, as pedestrians are forbidden from using them.

A "minutes" column was added to the data, which represented the amount of time it would take, in minutes, to walk from one end of the road feature to the other. Considering that the length of the feature in meters was stored in the "Shape_Length" column and that a person walking at 4 km/h moves 66⅔ meters in a minute, this new column was populated by using the following statement in the Field Calculator: `Shape_Length / 66.667`. The “minutes” column served as the cost in the network. When calculating service areas, ArcGIS will look for the paths with the least cost. In this case cost was defined as time in minutes. The paths with the least cost are those that take the shortest amount of time to traverse.

The “poi” table was imported into another feature class in the file geodatabase, and 20-minute service areas were generated for each feature in the “poi” table using the Generate Service Areas tool in ArcGIS. The generated service areas were initially exported as another feature class in the file geodatabase and then were imported into our PostgreSQL database. Only the columns for the geometry and the primary key were kept.

Analyses

Four views were created to simplify the queries needed for our analysis. The views allow us to create shorter, more readable SQL statements that avoid executing multiple joins per statement.

View “poi_with_weight”

This view is a join between the “poi” and “poi_weight” tables, created for ease of later analysis.

```
CREATE VIEW poi_with_weight AS SELECT poi.id,  poi.geom, poi.type,
poi_weight.general_weight, poi_weight.park_weight FROM poi JOIN
poi_weight ON poi.type = poi_weight.type;
```

View “service_areas_with_poi”

This view is a join between the “poi_with_weight” view and “poi_service_area” table, created for ease of later analysis. It represents a complete amalgamation of all the related data in the “poi_service_areas”, “poi”, and “poi_weight” tables.

```
CREATE VIEW public.service_areas_with_poi AS SELECT
poi_service_areas.id, poi_service_areas.geom, poi_with_weight.type,
poi_with_weight.general_weight, poi_with_weight.park_weight FROM
poi_service_areas JOIN poi_with_weight ON poi_with_weight.id =
poi_service_areas.id;
```

View “blocks_with_walk_scores”

This view is where the intersection of features in the “service_areas_with_poi” view with the features in the “blocks” table is calculated. By calculating the sum of the “general_weight” of all of the POI service areas that intersect a given block, we get a general walkability score. Doing the same for “park_weight” gives us the outdoor recreation walkability score.

```
CREATE VIEW public.blocks_with_walk_scores AS SELECT a.gid,
a.block_group, a.geom, sum(b.general_weight) AS general_score,
sum(b.park_weight) AS park_score FROM block a, service_areas_with_poi
b WHERE st_intersects(a.geom, b.geom) GROUP BY a.gid;
```

View “complete_blocks”

This view is a join between the results in the “blocks_with_walk_scores” view and the income data in the “household_income” table. It represents the final product of all joins and spatial analyses.

```
CREATE VIEW public.complete_blocks AS SELECT a.gid, a.block_group,
a.geom, a.general_score, a.park_score, b.id, b.income FROM
blocks_with_walk_scores a JOIN household_income b ON a.block_group =
b.id;
```

Findings

The tables and views described above were used to create static maps and scatter plots that addressed our questions.

The maps displayed in Figures 3 - 6 identify areas in Vancouver that could be considered the most desirable for pedestrians. Each point represents the centroid of a census block. Depending on the map, the value symbolized is either the general walkability score or the outdoor recreation score. The continuum of colors starts at near-black for zero, then increases in intensity through purple, pink, orange, and yellow up to a value of 700 for general walkability or 210 for outdoor recreation, respectively. The size of a point is also determined by the same attribute. In the general walkability score maps, the width of a point in meters on the ground is the equal to the value of its score (i.e. a block with a general walkability score of 100 will have a point that is 100 meters wide). In the maps showing proximity to outdoor recreation points of interest, a similar rule applies, but the value of the score is multiplied by three (i.e. a block with a outdoor recreation score of 100 will have a point that is 300 meters wide). Where these points overlap, they affect each other's’ colors through addition. For example, in places where many yellow points overlap, such as the urban core of Vancouver, an intense white glow is exhibited.

With regard to our first two questions about walkability and outdoor physical fitness, the areas identified appear to be largely coterminous. The Vancouver neighborhoods of Esther Short, Hough, Arnada, Carter Park, Central Park, Hudsons Bay, and Edgewood Park have the highest concentration of walkable areas. These neighborhoods are older and surround the urban core of Vancouver, so this finding is not surprising.

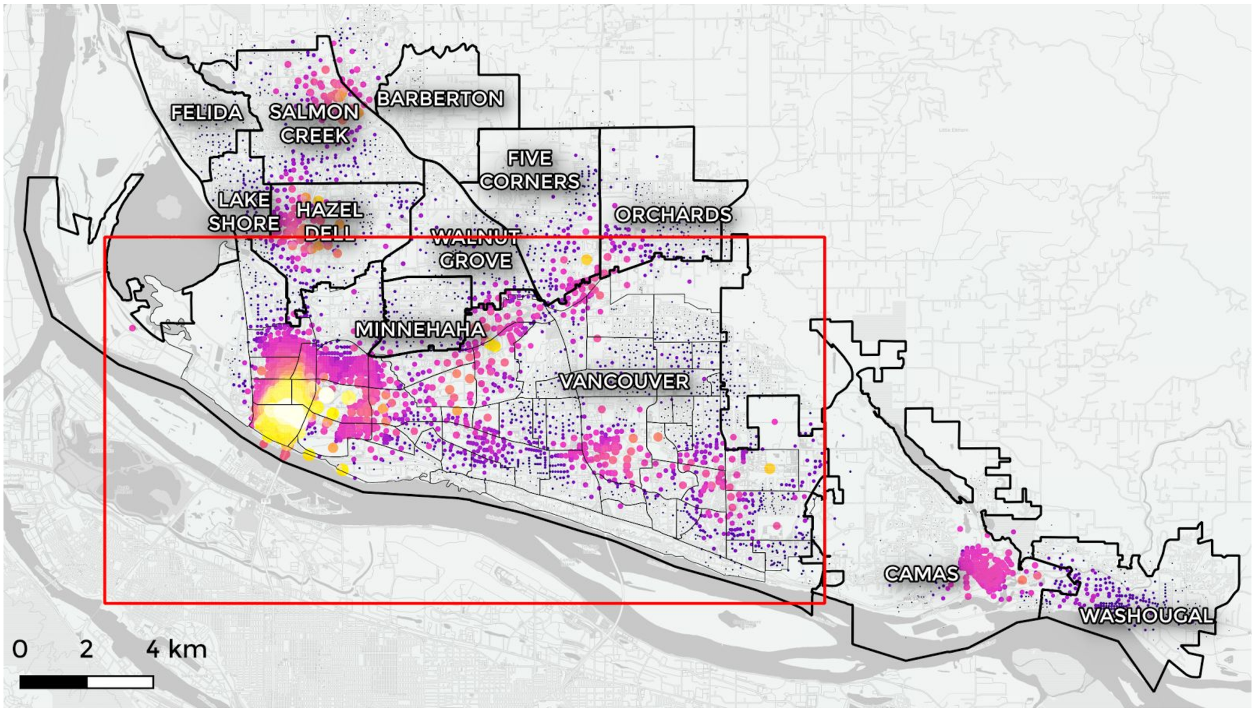

Figure 3. The most walkable areas in the entire study area. *Basemap © OpenStreetMap contributors*

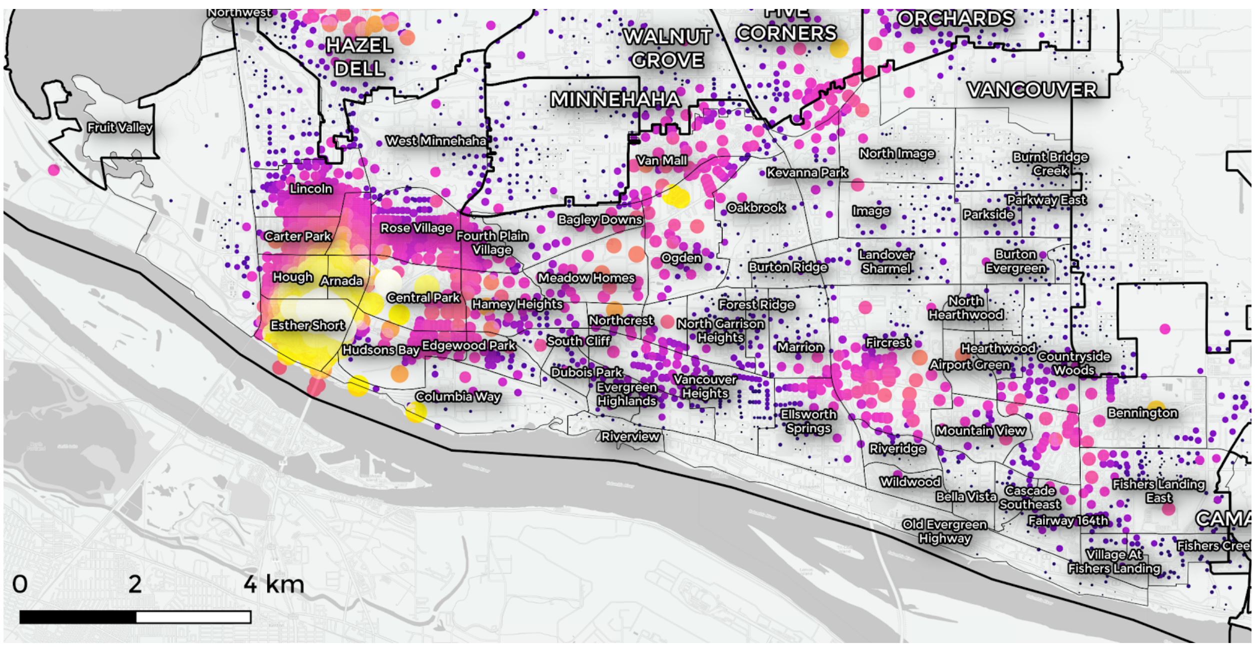

Figure 4. The most walkable areas in Vancouver neighborhoods. *Basemap © OpenStreetMap contributors*

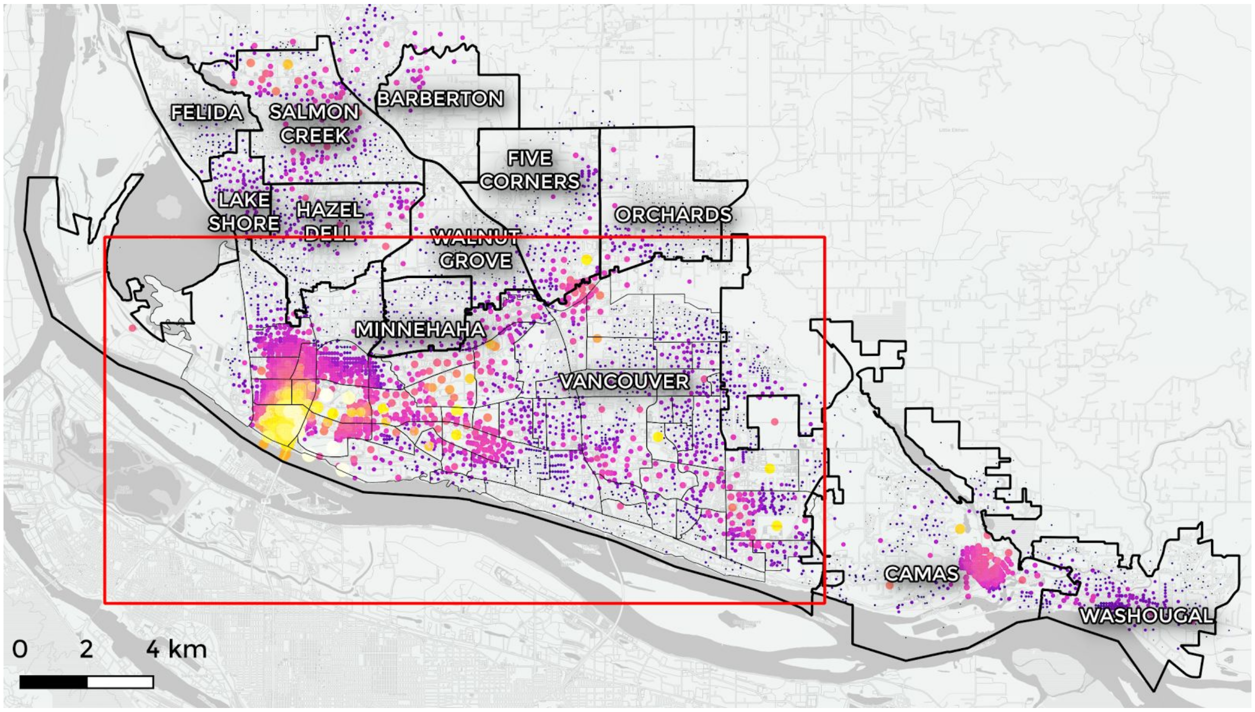

Figure 5. Areas most conducive to walking to outdoor recreational points of interest (entire study area). *Basemap © OpenStreetMap contributors*

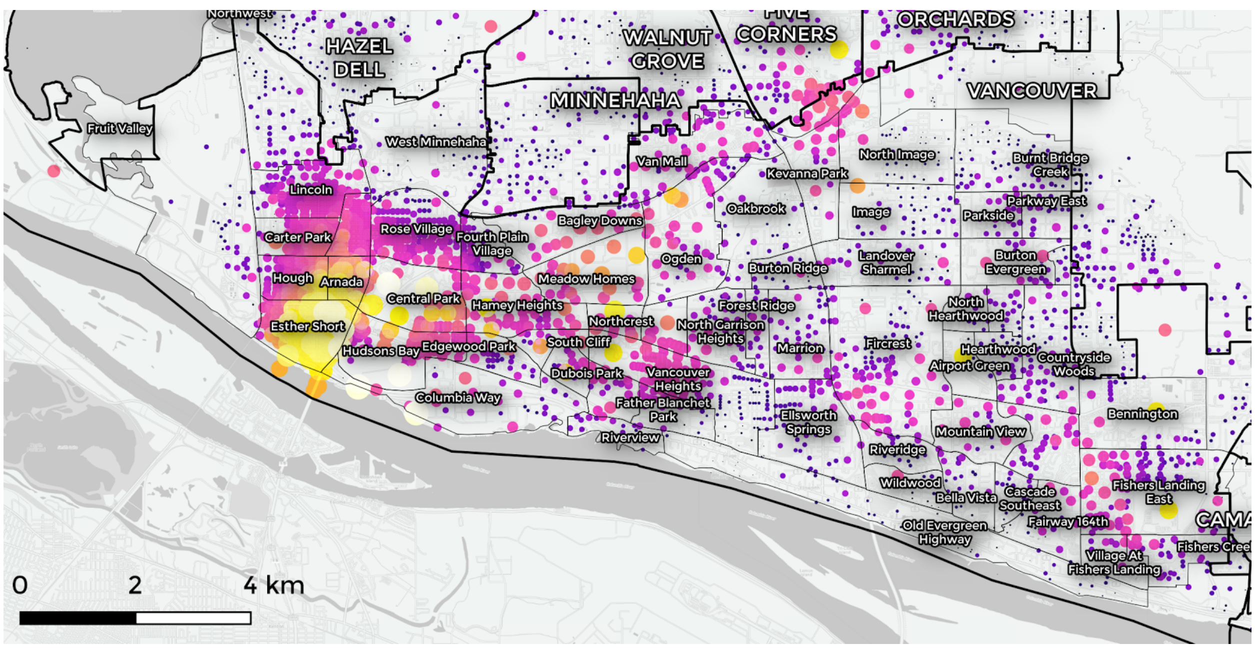

Figure 6. Areas most conducive to walking to outdoor recreational points of interest (Vancouver neighborhoods). *Basemap © OpenStreetMap contributors*

Our third and final question aimed to assess a relationship between median neighborhood income and walkability. For this assessment, we created two scatter plots (Figures 7 and 8). In our initial hypotheses, we expected a positive correlation between a census block's walkability and its median household income; however, our scatter plots show a correlation that, especially in the lower end of incomes, is strongly negative.

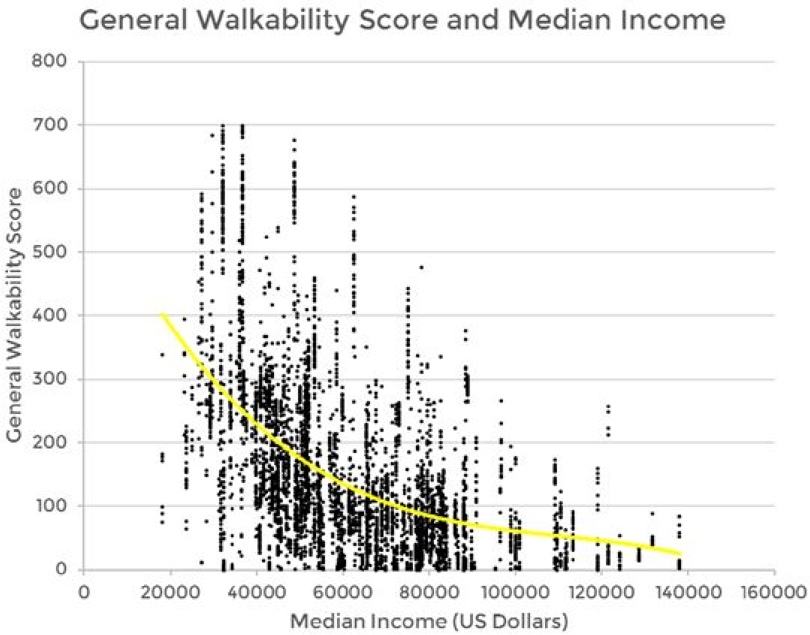

Figure 7. Correlation between general walkability score and median income.

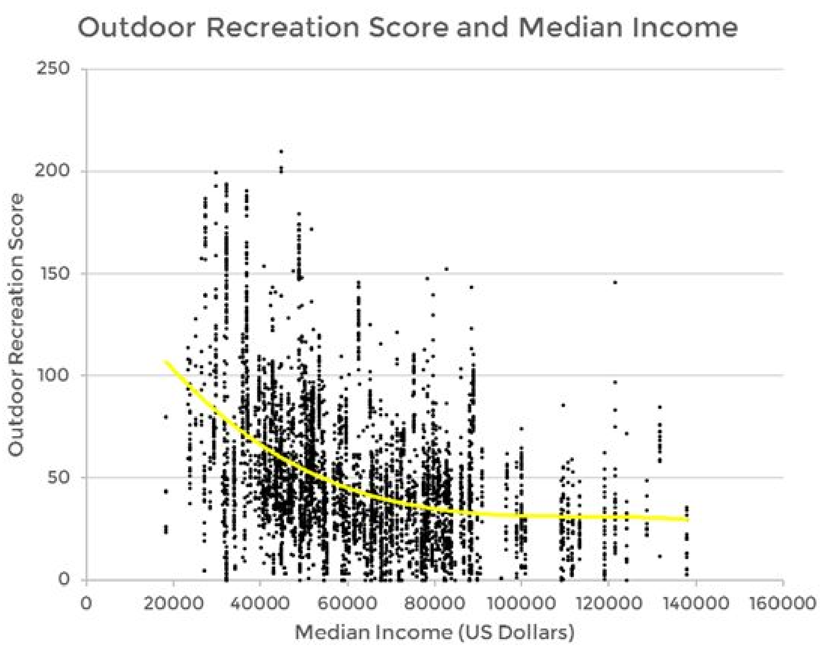

Figure 8. Correlation between outdoor recreation score and median income.

Summary of Conclusions

Our project demonstrated that Vancouver's most walkable environments are its central business district in the Esther Short neighborhood and the neighborhoods adjacent to it. These older neighborhoods have the greatest density of desirable places to walk and the street connectivity to encourage pedestrian travel. Our hypothesis that the correlation between household income and walkability would be positive was refuted; in Vancouver, Washington, the areas that are most walkable tend to be inhabited by individuals with relatively low incomes.

As discussed in the objectives of this study, other data could have been used to further influence the walkability score. Additionally, using a larger sample of people to determine the relevant importance of the various amenities to have within walking distance could have influenced the makeup of the walk score values. With walkability becoming more of a focus amongst city planners, this database and resulting maps could be valuable to those in city and county governments for exploring how to make areas more desirable for pedestrian use.

References

Christensen, N. (2015). *Portland region grows to 2.35 million residents, Census estimates, with newcomers leading the way*. Retrieved from http://www.oregonmetro.gov/news/portland-region-grows-235-million-residents-census-estimates-newcomers-leading-way

City of Portland Bureau of Planning and Sustainability (2009). *Status report: twenty-minute neighborhoods.* Retrieved from https://www.cityofmadison.com/sustainability/community/documents/20minNeigh.pdf

EliL (2015). *Portland, OR Chapter of OSGEO* Wiki. Retrieved from http://wiki.osgeo.org/wiki/PDX-OSGEO

United States Census Bureau (2016). *American FactFinder*. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.

United States Census Bureau (2010). *TIGER/Line® Shapefiles: Blocks - Washington*. Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2016&layergroup=Places.

United States Census Bureau (2016). *TIGER/Line® Shapefiles: Places - Washington*. Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2016&layergroup=Places.

United States Census Bureau and State of Washington Office of Financial Management (2010). *Census geographic files: Roads*. Retrieved from http://www.ofm.wa.gov/pop/geographic/tiger.asp.

Vanderbilt, T. (2012). *What's Your Walk Score?* Retrieved from http://www.slate.com/articles/life/walking/2012/04/walking_in_america_how_walk_score_puts_a_number_on_walkability_.html

Appendix: Weights assigned to OpenStreetMap amenities

amenity	general	outdoor
atm	3	0
bicycle	3	2.67
bus_stop	3	0.67
chemist	3	0
clinic	3	0
dentist	3	0
supermarket	3	0
bakery	2.67	0
bank	2.67	0
bus_station	2.67	0.67
department_store	2.67	0
library	2.67	0.33
pharmacy	2.67	0
deli	2.33	0
doctors	2.33	0
fast_food	2.33	0
food	2.33	0
food_court	2.33	0
greengrocer	2.33	0
hospital	2.33	0
post_office	2.33	0
pub	2.33	0
restaurant2.33	0	
social_facility	2.33	1
bar	2	0
beauty	2	0
bicycle_parking	2	2.67
butcher	2	0
cafe	2	0
cinema	2	0
clothes	2	0
community_centre	2	1
fitness_centre	2	1.33
garden	2	2.67
hairdresser	2	0
nightclub	2	0
playground	2	3
recreation_ground	2	3
school	2	0
sports_centre	2	3
theatre	2	0.33
convenience	1.67	0
houseware	1.67	0
laundry	1.67	0
mall	1.67	0
museum	1.67	0
optician	1.67	0
park	1.67	3
sports	1.67	3
swimming_pool	1.67	3
variety_store	1.67	0
baby_goods	1.33	0
childcare	1.33	0
confectionery	1.33	0
dry_cleaning	1.33	0
kindergarten	1.33	0
medical_supply	1.33	0
mobile_phone	1.33	0
nutrition_supplements		
	1.33	0
place_of_worship	1.33	0
trade	1.33	0

amenity	general	outdoor
university	1.33	0
viewpoint	1.33	1
alcohol	1	0
animal_boarding	1	0
art	1	0
artwork	1	0
attraction	1	1
books	1	0
boutique	1	0
bowling_alley	1	0
cannabis	1	0
car_rental1	0	
car_repair1	0	
charging_station	1	0
college	1	0
computer	1	0
courthouse	1	0
craft	1	0
dog_park	1	2.67
doityourself	1	0
drinking_water	1	1
fire_station	1	0
florist	1	0
fountain	1	1.33
garden_centre	1	0.67
gift	1	0
hardware	1	0
ice_cream	1	0.33
massage	1	0
miniature_golf	1	2.67
money_lender	1	0
musical_instrument	1	0
nature_reserve	1	3
outdoor	1	2.67
picnic_site	1	2
picnic_table	1	1.67
police	1	0
post_box	1	0
public_bookcase	1	0
recycling	1	0
second_hand	1	0
stadium	1	0.67
tool_rental	1	0
townhall	1	0
toys	1	0.33
track	1	2
video	1	0
video_games	1	0
water_park	1	2.33
antiques	0.67	0
conference_centre	0.67	0.33
cosmetics	0.67	0
furniture	0.67	0
golf_course	0.67	1.67
hotel	0.67	0
interior_decoration	0.67	0
jewelry	0.67	0
paint	0.67	0.67
pet	0.67	0.67
pitch	0.67	1.33
shoes	0.67	0
tobacco	0.67	0

amenity	general	outdoor
veterinary	0.67	0
waste_disposal	0.67	0
bench	0.33	1.33
car	0.33	0
clock	0.33	0
compressed_air	0.33	0
electronics	0.33	0
fabric	0.33	0
fuel	0.33	0
hearing_aids	0.33	0
Property Management Company		
	0.33	0
slipway	0.33	0.33
stationery	0.33	0
tabacco	0.33	0
tattoo	0.33	0
telephone	0.33	0
ticket	0.33	0
toilets	0.33	0.33
trampoline_park	0.33	2
travel_agency	0.33	0
tyres	0.33	0
vending_machine	0.33	0
waste_basket	0.33	0.33
weapons	0.33	0
appliance	0	0
bed	0	0
boat_repair	0	0
bureau_de_change	0	0
camp_site	0	1
car_parts	0	0
car_wash	0	0
caravan_site	0	1
carpet	0	0
Carpet Cleaning Service		
	0	0
chandler	0	0
charity	0	0
e-cigarette	0	0
fishing	0	2
frame	0	0
funeral_directors	0	0
glaziery	0	0
grave_yard	0	0
horse_riding	0	2
information	0	0
marina	0	0.67
motel	0	0
motorcycle	0	0
motorcycle_parking	0	0
nursing_home	0	0
parking	0	0
parking_space	0	0
party	0	0
pawnbroker	0	0
prison	0	0
scuba_diving	0	1
shelter	0	0
smoking_area	0	0
tanning_salon	0	0
vacuum_cleaner	0	0