Assessing Urban Quality at the Parcel Level

Carole Voulgaris and Elizabeth Christoforetti

Contents

Chapter 1

Introduction

Elizabeth will write 2,000 words or so about the motivations and vision for this project.

1.1 Report organization

The remainder of this report proceeds as follows. Chapter 2 discusses related work that we and others have done on the topic of evaluating urban quality and the challenges of highly-disaggregated spatial data. In Chapter 3, we describe a set of workshops we conducted with a diverse set of experts on urban development to identify values associated with urban quality. We go on to propose a method for evaluating urban quality at the parcel level using readily available data for Allegheny County, Pennsylvania in Chapter 4, and summarize the results of that analysis in Chapter 5. In Chapter 6, we discuss the alignment of the values suggested in our workshops with the outcomes of our quantitative analysis. Chapter 7 concludes the report with our key takeaways and potential directions for future work.

Chapter 2

Related work

2.1 Neighborhood classification

There is a large body of literature that seeks to apply quantitative methods to describe or classify urban environments. In 2011, Urban Geography released a special issue devoted to neighborhood classification approaches, including a review of neighborhood classification work that had been done to date [Reibel, 2011]; a study identifying five distinct neighborhood types in Cleveland and demonstrating how locations transition among types [Mikelbank, 2011]; a method for identifying ethnic neighborhoods [Logan et al., 2011]; and an analysis of New Urban developments that classifies them by how well they meet the ambitions of the New Urbanism movement [Trudeau and Malloy, 2011]. A common approach to classifying neigborhoods has been to employ factor analysis (principal component analysis) to reduce a large number of variables into a smaller set of factors (or principal components), followed by cluster analysis to group neighborhoods sharing similar characteristics [Voulgaris et al., 2016, Chow, 1998, Li and Chuang, 2009, Shay and Khattak, 2007, Song and Knaap, 2007, Song and Quercia, 2008, Vicino et al., 2011]. Although the purpose of neighborhood classification studies is to develop a set of categorical neighborhood types, the initial factor analysis step yields a set of indices that can be used as continuous variables describing various dimensions of neighborhood characteristics.

2.2 Quantification of sprawl

In general, neighborhood classification studies have differentiated between neighborhoods with a more urban character and those with a more suburban character. Indeed this is often the explicit purpose of such analyses. A related body of work has specifically sought to quanitify sprawl. Hamidi et al. [2015] offers a helpful review of the early work on this topic, noting that early studies emphasized density as a measure of sprawl and that some used satellite imagery to incorporate parameters like fragmentation and fractal dimension. Ewing et al. [2002] have developed a widely-cited measure of sprawl using principal component analysis to develop indices for four separate dimensions of sprawl (density, land-use diversity, centering, and street accessibility) and averaging them (with equal weights) to generate a single overall sprawl index. Hamidi et al. [2015] later repeated this method with updated data and have published a dataset of county-level and tract-level values for the resulting sprawl index.

2.3 Conceptualizing urban quality at the site level

Chapter 3

Expert Perspectives

Summarize the workshops.

Chapter 4

Quantitative Methods

How might results of quantitative approach to evaluating urban quality at the parcel level align with the values identified in the workshops described in Chapter 3? In this chapter, we propose one such approach, using readily-available parcel-level data for Allegheny County, Pennsylvania.

4.1 Data

We obtained data on property addresses, land uses, assessed values (for both land and buildings), and sale prices from Allegheny County Office of Property Assessments [2022], which includes information on 582,116 properties in Allegheny County.

We also obtained latitude and longitude coordinates for each property from a geocoder file provided by Western Pennsylvania Regional Data Center [2021]. Over 99.5 percent of properties included in the assessment dataset are included in the geocoder file. Properties without geocoded locations are excluded from our analysis.

Potential development sites were identified as those

1. classified as "residential" (residential properties with one to four housing units) or "commercial" (which includes mixed-use developments and residential properties with more than four housing units), and

2. with a land use description in one of 59 possible categories[1]. The most common of these are listed Table 4.1.[2].

Table 4.1: Most common land uses categorized as potential sites

USEDESC	Number of potential sites	Percent of potential sites	Cumulative percent of potential sites
SINGLE FAMILY	370,513	73.2	73.2
VACANT LAND	62,672	12.4	85.5
TWO FAMILY	17,293	3.4	89.0
TOWNHOUSE	14,670	2.9	91.8
ROWHOUSE	11,082	2.2	94.0
VACANT COMMERCIAL LAND	5,817	1.1	95.2
THREE FAMILY	3,968	0.8	96.0
RES AUX BUILDING (NO HOUSE)	3,601	0.7	96.7
RETL/APT'S OVER	3,354	0.7	97.3
COMM AUX BUILDING	2,825	0.6	97.9
APART: 5-19 UNITS	2,771	0.5	98.4
FOUR FAMILY	2,058	0.4	98.9
BUILDERS LOT	1,230	0.2	99.1
PARKING GARAGE/LOTS	891	0.2	99.3
OFFICE/APARTMENTS OVER	854	0.2	99.4
MOBILE HOME	666	0.1	99.6
APART:40+ UNITS	529	0.1	99.7
DWG USED AS OFFICE	440	0.1	99.8
APART:20-39 UNITS	400	0.1	99.8
CONDEMNED/BOARDED-UP	132	0.0	99.9

Potential building sites were further filtered to exclude those with missing data on the most recent sale (about one percent of all sites).[3]

The focus of this analysis is on potential development sites rather than on properties. Some properties in the assessor dataset are condominums where multiple properties share a single parcel of land. We aggregated these to the site level by identifying all properties with an assessed building value greater than zero, a land value of zero, and a land use description that did not indicate the land was vacant. If multiple such properties share an address, we classified all properties at that address as a condominium and aggregated them to the parcel level. This led to a final sample of 506,405 sites.

4.1.1 Tax assessment data

Three variables (total assessed fair market value, assessed fair market value of the building, and lot area) were taken directly from the county tax assessment data for use in our analysis. We also included the most recent listed sales price, adjusted for inflation.

[1] One site (3008 Phillip Dr in Clairton) is missing a land use description in the assessment data. We checked this address on Zillow to determine that this is a single-family home and classified it as such in our data.

[2] The land use descriptions that were classified as potential development sites but are not listed in Table 4.1, which combine to represent less than one percent of all sites are "RIGHTOF WAY - RESIDENTIAL", "CONDOMINIUM UNIT", "DWG USED AS OFFICE", "APART:20-39 UNITS", "CONDO GARAGE UNITS", "COMMON AREA", "CONDO DEVELOPMENTAL LAND", "CONDEMNED/BOARDED-UP", "CONDOMINIUM OFFICE BUILDING", "INDEPENDENT LIVING (SENIORS)", "DWG USED AS RETAIL", "OTHER COMMERCIAL", "MOBILE HOMES/TRAILER PKS", "RIGHT OF WAY - COMMERCIAL", "GROUP HOME", "TOTAL/MAJOR FIRE DAMAGE - COMM", "OTHER COMMERCIAL HOUSING", "TOTAL/MAJOR FIRE DAMAGE", "COMM APRTM CONDOS 5-19 UNITS", "MUNICIPAL URBAN RENEWAL", "COMMERCIAL LAND", "CAMPGROUNDS", "COMMON AREA OR GREENBELT", "CHARITABLE EXEMPTION/HOS/HOMES", "INCOME PRODUCING PARKING LOT", "DWG APT CONVERSION", ">10 ACRES VACANT", "MINOR FIRE DAMAGE", "COMM APRTM CONDOS 20-39 UNITS", "COMMERCIAL/UTILITY", "H.O.A RECREATIONS AREA", "COMM APRTM CONDOS 40+ UNITS", "MINOR FIRE DAMAGE - COMM", "OTHER", "OTHER RESIDENTIAL STRUCTURE", "OWNED BY METRO HOUSING AU", "RESIDENTIAL VACANT LAND", "HUD PROJ #221", and "VACANT LAND 0-9 ACRES"

[3] Four sites had sales prices listed that were unreasonably high. 3039 Liberty Avenue in Pittsburgh is listed as having sold for $511,945,000 on August 30, 2021. Zillow lists this property as having sold on that date for $511,945 (https://www.zillow.com/homedetails/3039-W-Liberty-Ave-Pittsburgh-PA-15216/2070262638_zpid/, accessed 5/4/2022), so the value was corrected for what appears to have been a typo. 220 Hyeholde Dr in Coraopolis is listed as having sold for $28,100,000 in 1967. This may also be a typo, and it also does not seem to be the most recent sale. Zillow lists this home as having sold for $350,000 in 2004 (https://www.zillow.com/homes/220-hyeholde-dr,-Coraopolis,-PA_rb/11552817_zpid/, accessed 5/4/2022), so the data was corrected to add that as the most recent sale. Two other sites were identified as having unreasonably high sales values: 1339 Arlington Avenue in Pittsburgh is a three-bedroom single-family home that is listed as having sold for $57,010,813 in 1976 and a 0.06-acre vacant lot with tax ID 0165G00270000000 is listed as having sold for $24,920,232 in 1936. The sales data for these sites were treated as missing.

To aggregate properties identified as condominiums to the site level, we summed the total values for lot area, assessed land value, assessed building value, and inflation-adjusted sale price. We log-transformed these four variables prior to including them in our analysis. Their distributions are shown in Figure 4.1.

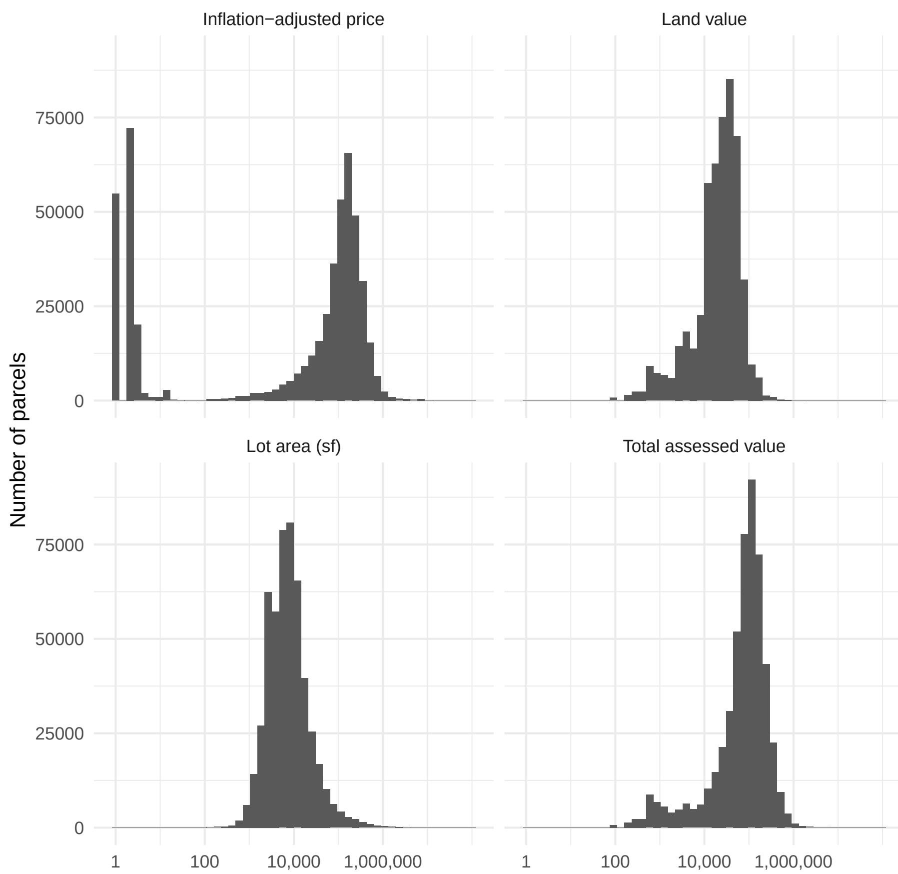

Figure 4.1: Distribution of variables from tax assessor database

4.1.2 Accessibilty data

Accessibilty was calculated from each of the 518,032 sites in our sample to each of several location types described below.

4.1.2.1 Destination parcels

We used land use codes from the county assessor parcel data to identify *destination parcels* that residents might value access to. The most common land use codes of identified destination parcels are listed in Table 4.2.

4.1.2.2 Job locations

We identified *job locations* based on data from a Longitudinal Employer-Household Dynamics (LEHD) dataset published by the United States Census Bureau [United States Census Bureau, 2021]. The LEHD dataset provides the total number of jobs in each census block in the United States, based on employment tax records. The location of each job was defined as the centroid of the block in which it was located. We downloaded job location data for Pennsylvania and filtered it to include locations in the Pittsburgh metropolitan area (Allegheny, Armstrong, Beaver, Butler, Fayette, Washington, and Westmoreland counties).

In addition to calculating the accessibility to jobs of all categories, we also calculated accessibility to several subsets of jobs. We disaggregated jobs by earnings, reasoning that the usefulness of a job might vary depending on how well it matches a workers skills or wage expectations. *High-paying job locations* are a subset of job locations where the worker earns more than $3333 per month. *Low-paying job locations* are those where the worker earns $1250 per month or less.

We also disaggregated jobs based on employment industry, based on the North American Industry Classification System (NAICS), reasoning that the presence of jobs particular industries might represent a shopping or recreation destination. *Retail job locations* are a subset of job locations in NAICS sector 44-45 (retail trade); *Entertainment job locations* are those in NAICS sector 71 (arts, entertainment, and recreation); and *Hospitality job locations* are those in NAICS sector 72 (accommodation and food services).

Finally, we identified three location types that correspond with common non-work trips: schools, grocery stores, and parks. *Grocery store locations* were identified as vendors participating in the Supplemental Nutrition Program for Women, Infants, and Children (WIC). WIC vendor locations and *school locations* were obtained from the Allegheny County GIS portal [Allegheny County Office of Information Technology, 2018, 2020]. *Park locations* were taken from the Pennsylvania Geospatial Data Clearinghouse [Pennsylvania Department of Conservation and Natural Resources, 2015]. Park locations were downloaded for Pennsylvania and filtered to Allegheny county.

We used the r5r package in the R programming language [Pereira et al., 2021] to calculate accessibility each destination type described above, for each of four transportation modes (walking, cycling, driving, and transit). The r5r package

Table 4.2: Land uses identified as potential destinations

USEDESC	Number of identified destinations	Percent of identified destinations	Cumulative percent of identified destinations
MUNICIPAL GOVERNMENT	10,376	29.88	29.88
CHURCHES, PUBLIC WORSHIP	1,946	5.60	35.49
COMMERCIAL GARAGE	1,735	5.00	40.48
OFFICE - 1-2 STORIES	1,649	4.75	45.23
SMALL DETACHED RET(UNDER 10000)	1,646	4.74	49.97
OFFICE/WAREHOUSE	1,386	3.99	53.96
COUNTY GOVERNMENT	1,287	3.71	57.67
WAREHOUSE	1,252	3.61	61.27
OWNED BY BOARD OF EDUCATION	1,086	3.13	64.40
TOWNSHIP GOVERNMENT	855	2.46	66.86
LIVESTOCK O/T D & P-CAUV	805	2.32	69.18
LIGHT MANUFACTURING	799	2.30	71.48
PUBLIC PARK	710	2.04	73.53
RESTAURANT, CAFET AND/OR BAR	697	2.01	75.54
GENERAL FARM	607	1.75	77.28
OWNED BY COLLEGE/UNIV/ACADEMY	458	1.32	78.60
MEDICAL CLINICS/OFFICES	445	1.28	79.88
RETL/OFF OVER	442	1.27	81.16
OFFICE-ELEVATOR -3 + STORIES	412	1.19	82.34
LODGE HALL/AMUSEMENT PARK	386	1.11	83.46

calculates accessibility as the weighted total number of destinations reachable by a given mode, where destinations are weighted according to a decay function, such that destinations that can be reached within less time are assigned greater weight. We used a logistic decay function, as illustrated in 4.2. For motorized modes, the decay function had a mean (inflection) of 40 minutes and a standard deviation of 10 minutes. For non-motorized modes, the decay function had a mean of 20 minutes and a standard deviation of 5 minutes.

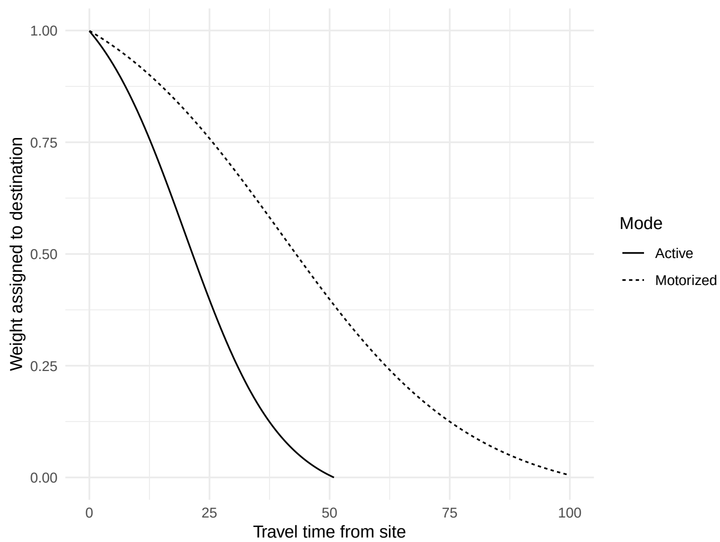

Figure 4.2: Decay functions for accessibility calculations

Calculating accessibility metrics for a combination of four transportation modes and ten destination types yields 40 different accessibility variables. Figure 4.2 illustrates the distributions of each of these variables.

4.1.3 Disamenity proximity

We categorized several land uses in the county assessor data as disamenities. The land use codes we used to identify disamenities are listed in 4.3[4].

We included a disamenity proximity index in our analysis that we calculated as the logarithm of the average distance from each site to the ten closest disamenity sites. The distribution of this index is shown in 4.4.

[4]289 properties related to coal mining (with land use descriptions of either "COAL RIGHTS, WORKING INTERESTS" or "COAL LAND, SURFACE RIGHTS") are co-located and are treated as a single site.

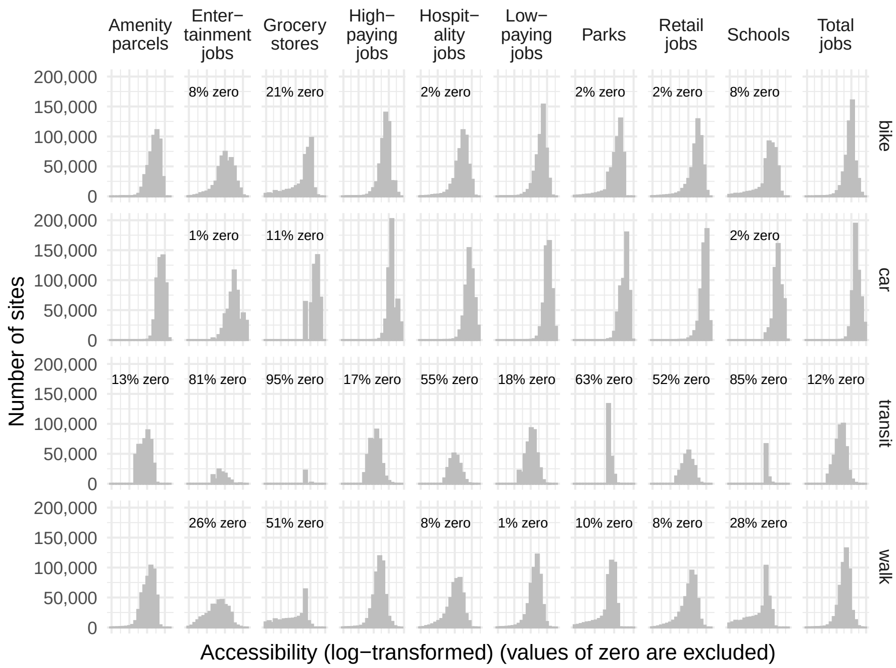

Figure 4.3: Distributions of accessibility variables

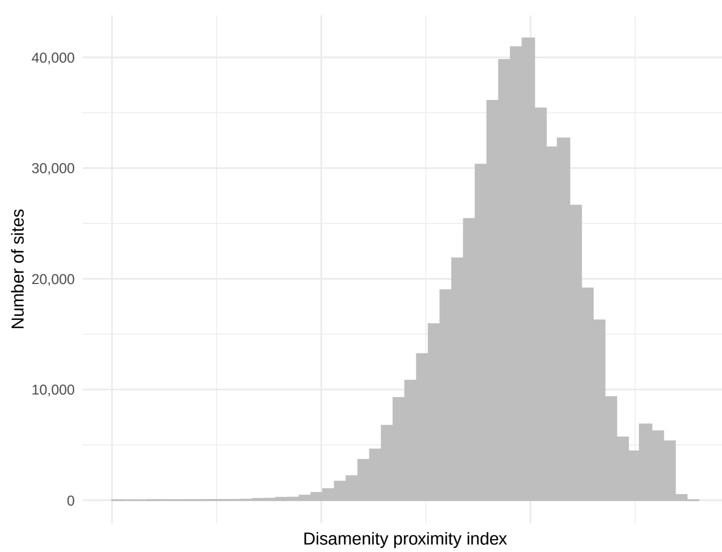

Figure 4.4: Distribution of average distance to nearest ten disamenity sites

Table 4.3: Land uses identified as disamenities

USEDESC	Number of identified locations	Percent of identified locations	Cumulative percent of identified locations
VACANT INDUSTRIAL LAND	614	68.68	68.68
HEAVY MANUFACTURING	89	9.96	78.64
FOOD & DRINK PROCESSING	40	4.47	83.11
INDUSTRIAL/UTILITY	39	4.36	87.47
COMMERCIAL TRUCK TERMINAL	26	2.91	90.38
RECYCLING/SCRAP YARDS	23	2.57	92.95
MINES AND QUARRIES	17	1.90	94.85
P.P. - P.U. - OTHER THAN R.R.	14	1.57	96.42
INDUSTRIAL TRUCK TERM	9	1.01	97.43
BULK TRANSFER TERMINAL	8	0.89	98.32
INDUSTRIAL LAND	3	0.34	98.66
POULTRY FARM	3	0.34	98.99
COAL LAND, SURFACE RIGHTS	2	0.22	99.22
COAL RIGHTS, WORKING INTERESTS	2	0.22	99.44
OIL & GAS RIGHTS WORKING INTEREST	2	0.22	99.66
COAL RIGHTS SEP. ROYALTY INTEREST	1	0.11	99.78
MINERAL LAND	1	0.11	99.89
OTHER MINERALS	1	0.11	100.00

4.1.4 Density

To represent the residential density around each site, we used the sf [Pebesma, 2018], nngeo [Dorman, 2022] and tidycensus [Walker and Herman, 2022] R packages to determine the smallest circular buffer around each site containing a population of at least two thousand people, based on the 2020 census. In denser places, a buffer with a smaller radius would encompass two thousand residents. In more sparsely-populated places, a buffer containing two thousand residents would be larger. The distribution of radii for two-thousand-person site buffers is shown in 4.5.

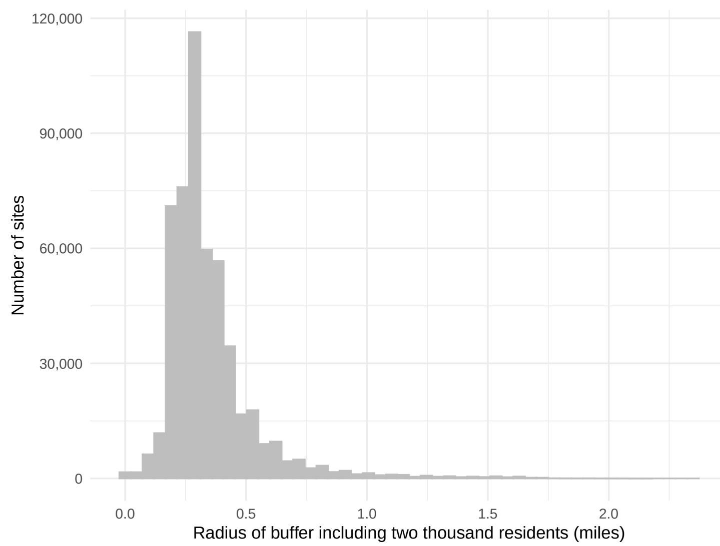

Figure 4.5: Histogram of radii of buffer containing 2000 residents

4.1.5 Population diversity

The two-thousand-resident buffers described above were also used as a basis to estimate the racial diversity of residents in the immediate vicinity. For each buffer, we calculated the percentage of residents that who identified in the 2020 census as non-Hispanic white, non-Hispanic Black, and Hispanic. The distributions of these variables are shown in 4.6.

4.1.6 Land use diversity

We also calculated the total number of different land uses within each two-thousand-resident buffer and used this as a measure of land-use diversity. 4.7.

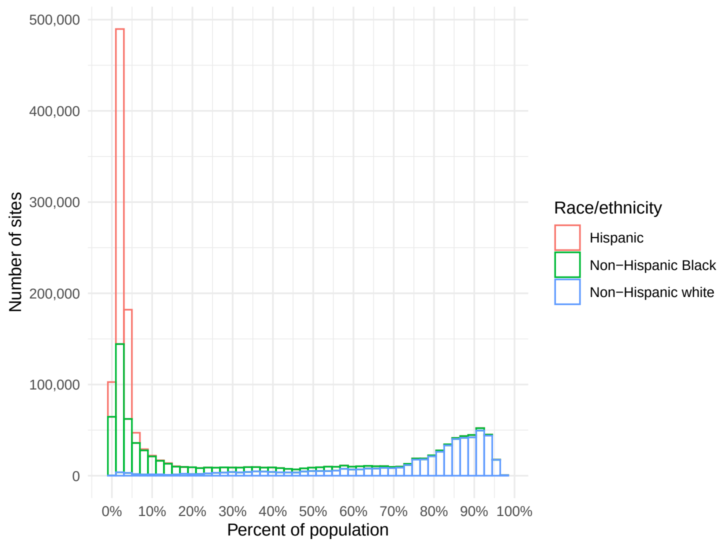

Figure 4.6: Distributions of population diversity variables

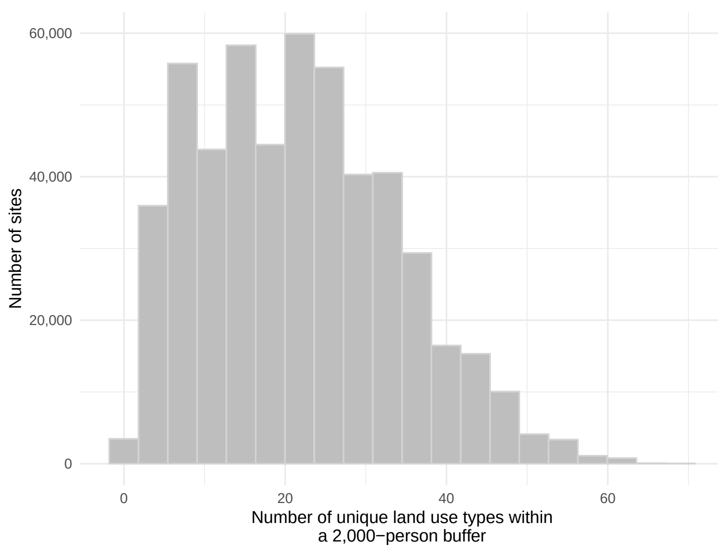

Figure 4.7: Histogram of land use diversity

4.2 Index development

The methods described above yielded a set of fifty parcel-level variables, forty of which are accessibility metrics, for each of 506,405 parcels. We used the EFAtools R package [Steiner and Grieder, 2020] to develop a set of parcel level indices from these variables using factor analysis. The Kaiser-Meyer-Olkin criterion for the dataset is 0.9, suggesting a “marvellous” case for factor analysis [Kaiser, 1974].

We determined the appropriate number of factors based on the Kaiser-Guttman criterion and the Hull method.The Hull method suggests an optimal number of factors that balances model fit and number of parameters, with a goal or keeping only major factors. Potential solutions with various number of factors are plotted on a graph of goodness-of-fit versus degrees of freedom, where the optimal solution will be on the boundary of a convex hull [Lorenzo-Seva et al., 2011]. The Kaiser-Guttmat criterion is a recommendation to retain as many factors as there are sample eigenvalues greater than one [Guttman, 1954].

We computed factor loadings using an oblimin rotation. We applied these loadings to calculate a set of index scores (one for each factor) for each potential development site.

4.3 Index validation

The indices we developed through factor analysis might represent dimensions of urban quality. If they are valid quality metrics, one might expect them to be predictive of an activity associated with desirable locations for development.

We hypothesize that more desirable locations for development might be those where plans have been made for new development activity and that building permits for new construction or demolition are indicative of such plans.

We estimated a logistic regression model using the indices developed through factor analysis as independent variables and predicting the likelihood that a site in the city of Pittsburgh was issued a a building permit for either construction or demolition over a one-year period (June 2021 - May 2022). Building permit data were obtained from Western Pennsylvania Regional Data Center [2022]. Out of 269,151 potential residential development sites in Pittsburgh, 139 (one twentieth of one percent) had building permits issued for new construction or demolition of a residential structure during the study period.

4.4 Combined index

If the indices we developed represent distinct dimensions of urban quality, the relative importance of each dimension (and its associated index) might vary depending on the values of the individual or institution seeking to assess urban quality. However, the results of the regression analysis might offer insight into the typical or average values of active housing developers and property owners in Pittsburgh.

We used the regression coefficients estimated to predict the likelihood of a recent building permit (as described above) to generate weights for each index, scaled such that the highest coefficient represented a weight of one hundred percent. We used these weights to calculate a combined index value for each site.

Chapter 5

Quantitative Results

This chapter summarizes the results of the factor analysis, index development, index validation, and the development of a combined index to describe site-level variation in urban quality.

5.1 Factor analysis

Both the Hull method and the Kaiser-Guttman criterion suggested a five-factor solution would be appropriate for out data set. Figure 5.1 illustrates the results of the Hull method with a plot of goodness-of-fit versus degrees of freedom for potential solutions with numbers of factors ranging from zero to fourteen. Figure 5.2 illustrates that there are five eigenvalues greater than one, suggesting a five-factor solution according to the Kaiser-Guttman Criterion.

We assigned names to each factor based on a visual inspection of the results. The *drivable* factor had the highest loadings for variables representing access by car to most destination types. The *walkable* factor has high loadings for variables representing access by walking and transit. The *diverse* index is characterized by diversity of people (high percentages of black residents and low percentages of white residents), diversity of land use (a greater number of distinct land uses in the immediate vicinity and a shorter average distance to disamenities), and lower assessed property values. The *dense* factor is characterized by lower values for the radius of the smallest buffer containing two thousand residents (i.e. higher population densities) and higher access to retail and grocery locations by non-motorized modes. The *amenities* factor is characterized by non-motorized and transit access to retail and grocery locations. Figure 5.3 illustrates the loadings of each individual variable onto each of the five factors.

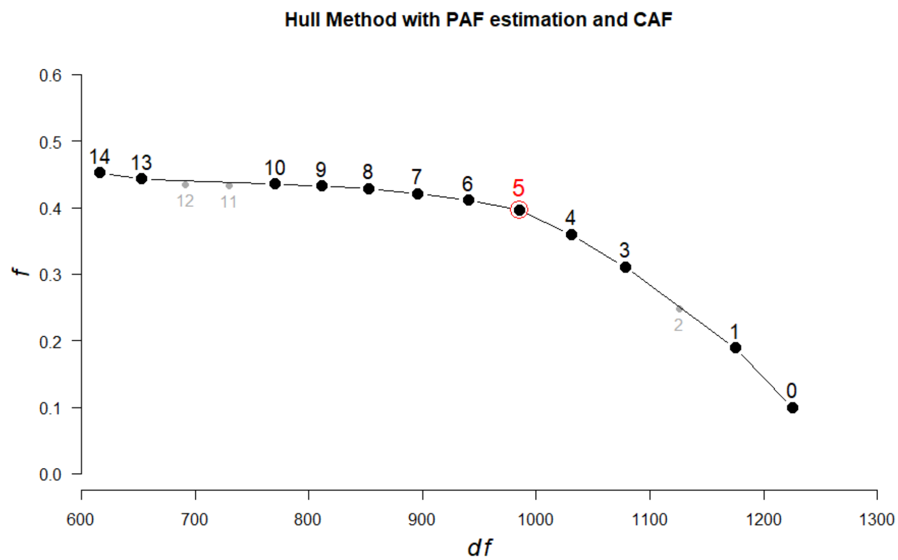

Figure 5.1: Results of Kaiser-Guttman criterion for determining the number of factors

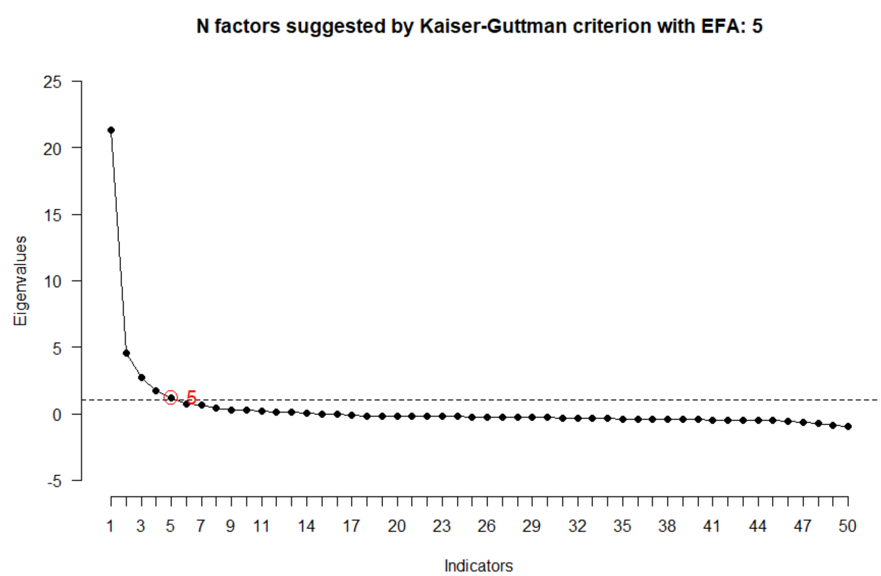

Figure 5.2: Results of Hull method for determining the number of factors

Figure 5.3: Factor loadings

5.2 Indices from factors

Figures 5.4 through 5.8 show the spatial variation in the drivability, walkability, density, diversity, and amenity-richness indices, respectively.

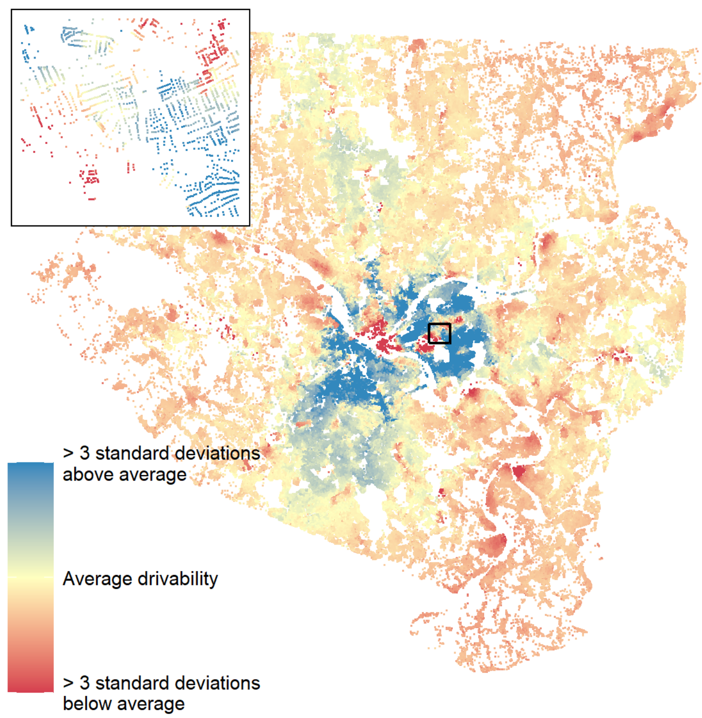

Figure 5.4: Spatial variation in drivability index

As illistrated in Figure 5.4, there is an area in the center of the region (downtown Pittsburgh) with particularly low drivablity (likely due to traffic congestion), surrounded by a ring around the center with particularly high drivability.

Figure 5.5 shows the opposite pattern for walkabilty, with particularly high walkability in the center of the region and a ring of surrounding the center with particularly low walkability.

Figure 5.6 shows pockets of low density in the center of the region, possibly because the density index is driven by residental density and these are commercial centers with relatively few residents. With the exception of those pockets, the density is generally highest in the center of the county and lowest closer to the boundaries.

Figure 5.7 suggests that the greatest diversity (with an index representing both

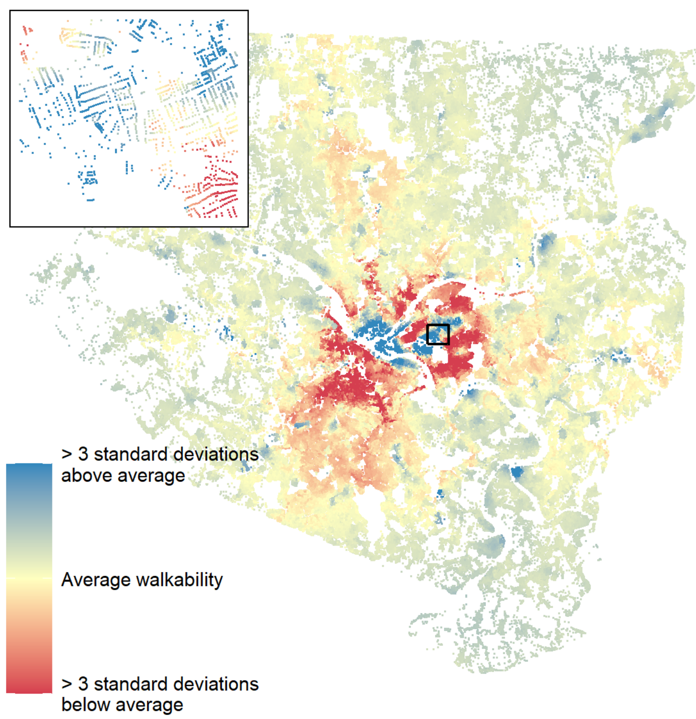

Figure 5.5: Spatial variation in walkability index

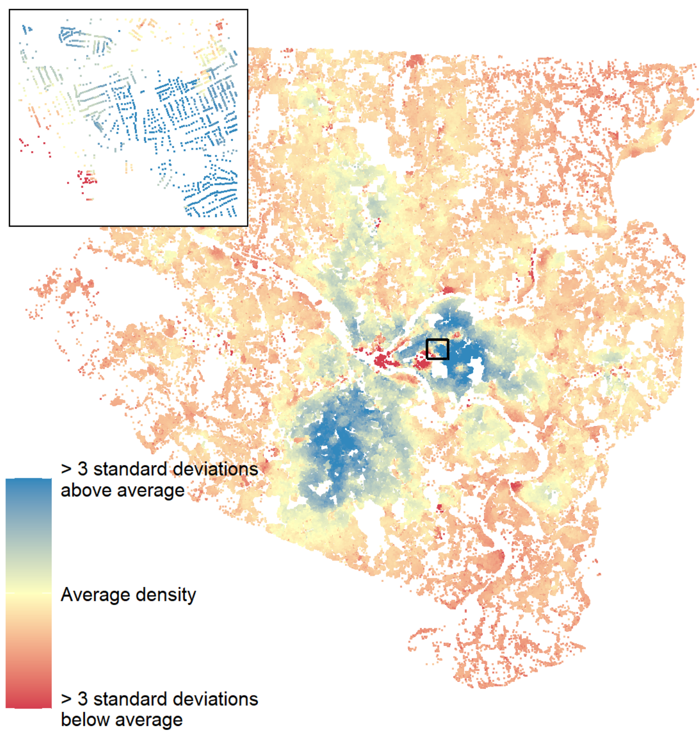

Figure 5.6: Spatial variation in density index

sociodemographic and land-use diversity) is found in areas adjacent to the center of the region.

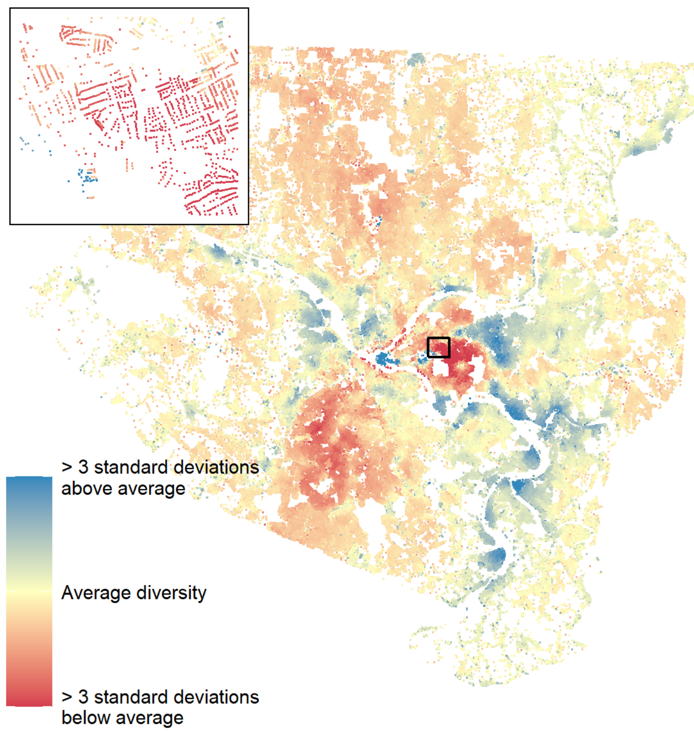

Figure 5.7: Spatial variation in diversity index

Finally, Figure 5.8 suggests an amenity-poor area at the center of the region surrounded by and amenity-rich ring, with remainder of the county having an amenity richness closer to the county average.

Figure 5.9 illustrates the distribution of each factor and the relationships among them. In general, there appears to be a trade-off between drivability and walkability and between drivability and diversity. There also appears to be a negative association between density and diversity. This may be because land-use diversity was measured as the number of unique land uses within the smallest buffer containing at least 2,000 residents. In very dense places, this buffer might be too small to include a large number of unique land uses.

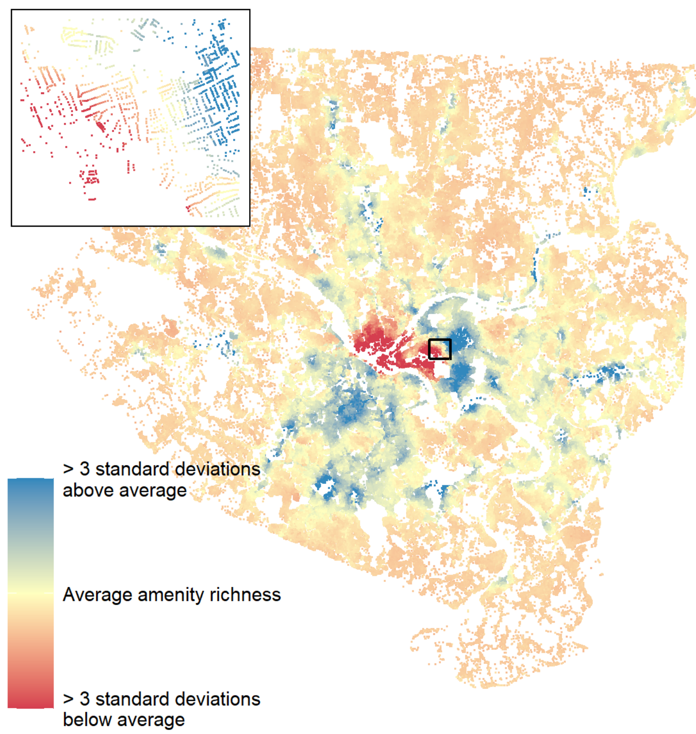

Figure 5.8: Spatial variation in amenity-richness index

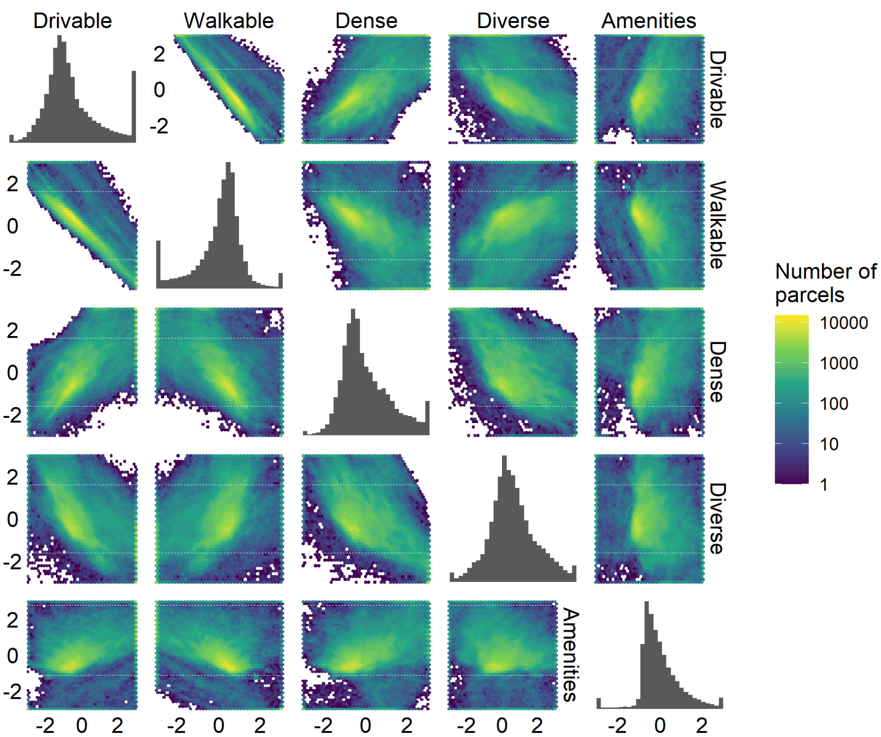

Figure 5.9: Spatial variation in amenity-richness index

5.3 Factor validation through regression

Table 5.1 shows the results of three alternative regression models. The first of these is a null model, which assumes that likelihood of the modeled outcome (in this case, a building permit for construction or demolition) is constant across all sites. The second model (labeled as the "full model") predicts that likelihood based on the variation in the three indices generated by the factor analysis. As shown, four of the five factors are statistically significant predictors of the likelihood of a building permit at a 99.9 percent confidence level, and this model fits the data better than the null model based on three measures of model fit: The Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the pseudo R^2. Based on the results of the full model, the amenity-richness index is not a significant predictor of the likelihood of a building permit, so a reduced model was estimated without that variable. In this reduced model, the remaining model coefficients were unchanged and the model fit was essentially unchanged.

Figure 5.10 shows how the predicted probability of development changes as each index varies from -5 to 5 (i.e. from five standard deviations below the average to five standard deviations above the average).

5.4 Combined index

Recent development activity has been consistent with a hypothesis that four of the five indices developed from the factor analysis results represent dimensions of urban quality that matter to developers and property owners. We can combine these four indices into a combined index by calculated a weighted average, where weights are derived from the coefficients of the regression model described above. The distribution of the resulting index is shown in Figure 5.11. Figure 5.12 shows the spatial distribution of this index across Pittsburgh, with the locations of the building permits used to estimate the model shown for reference.

Figure 5.13 shows the spatial distribution of the combined index across the entire study area. It is noteworthy that there is less variation within the inset area than there is for any of the four indices it comprises. This is because low scores on one index generally compensate for high scores on another.

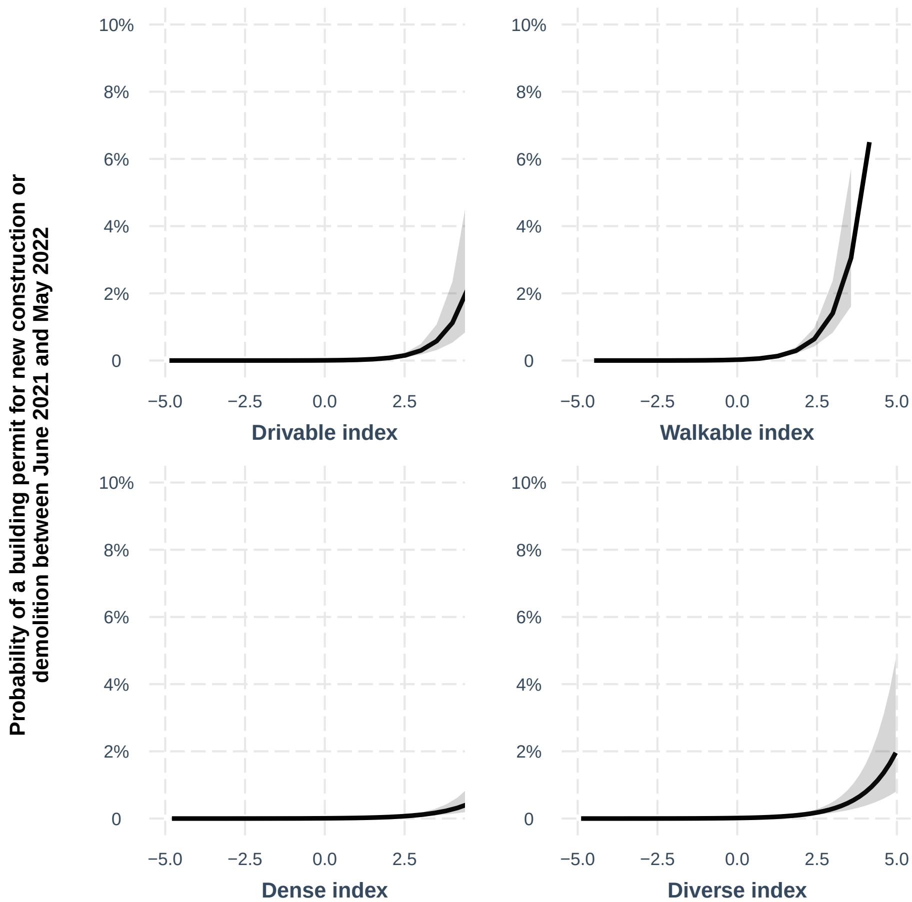

Figure 5.10: Effect of drivability, walkabilty, diversity, and density on development probability

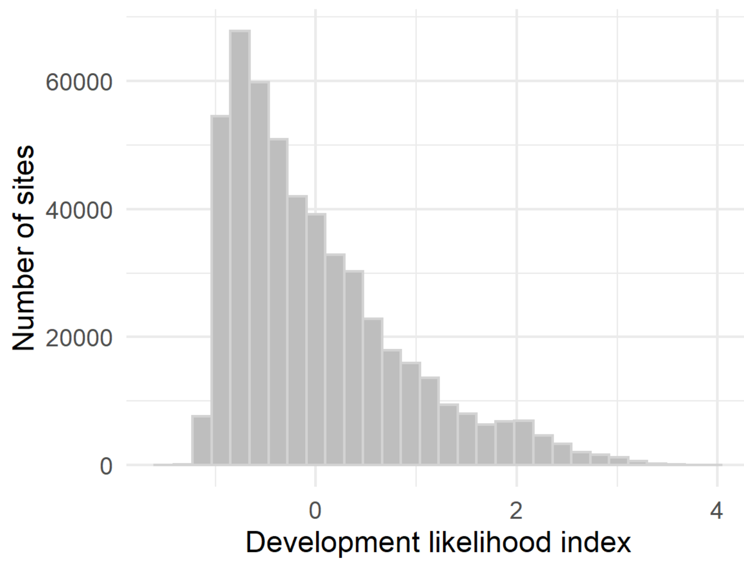

Figure 5.11: Distribution of combined index, weighted according to coefficients from regression predicting the likelihood of a building permit for construction or demolition

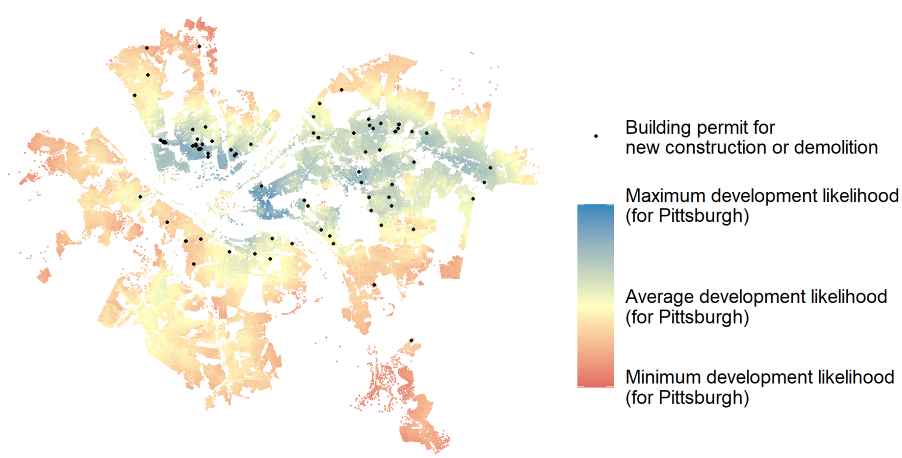

Figure 5.12: Locations of building permits for new construction and demolition and their estimated likelihood

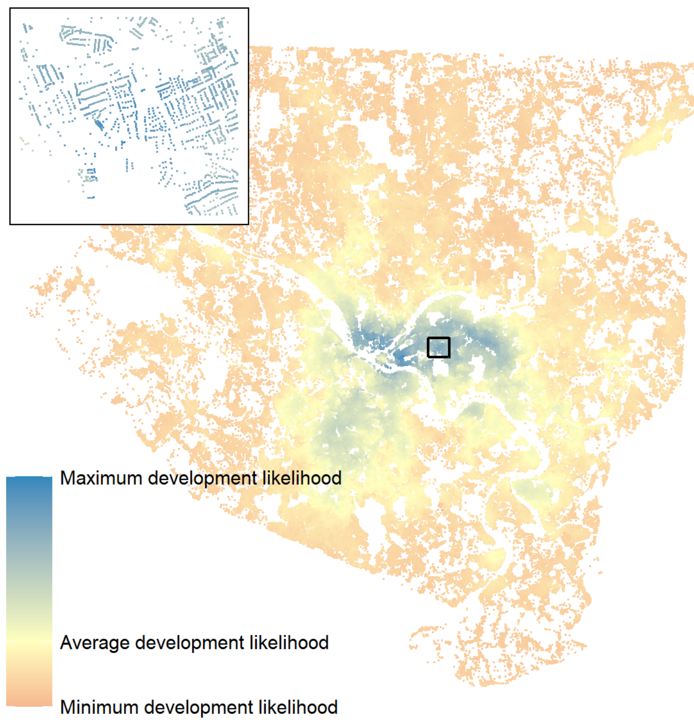

Figure 5.13: Spatial variation in combined index, weighted according to coefficients from regression predicting the likelihood of a building permit for construction or demolition

Table 5.1: Results of logistic regression predicting likelihood of a demolition or construction permit over a one-year period.

	Null model	Full model	Reduced model
(Intercept)	-8.04 ***	-9.57 ***	-9.57 ***
	(SE = 0.11)	(SE = 0.25)	(SE = 0.24)
Drivable		1.35 ***	1.35 ***
		(SE = 0.16)	(SE = 0.14)
Walkable		1.38 ***	1.38 ***
		(SE = 0.15)	(SE = 0.11)
Dense		0.90 ***	0.90 ***
		(SE = 0.12)	(SE = 0.12)
Diverse		0.98 ***	0.98 ***
		(SE = 0.12)	(SE = 0.11)
Amenities		0.00	
		(SE = 0.07)	
N	269151	269151	269151
AIC	1574.43	1414.33	1412.33
BIC	1584.93	1477.35	1464.84
Pseudo R2	0.00	0.11	0.11

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Chapter 6

Relating Indices to Values

This chapter will draw connections between the workshop results and the quantitative analysis results. Elizabeth will write this.

Chapter 7

Conclusions and Future Directions

Elizabeth to write this chapter.

Bibliography

Allegheny County Office of Information Technology. Allegheny County WIC Vendor Locations, April 2018. URL https://openac-alcogis.opendata.arcgis.com/datasets/ab9ec54e46d8403db31cff6bdc890aff_0/explore?location=40.458725%2C-79.972398%2C10.20. type: dataset.

Allegheny County Office of Information Technology. Allegheny County Public Schools / Local Education Agency (LEAs) Locations, December 2020. URL https://openac-alcogis.opendata.arcgis.com/datasets/AlCoGIS::allegheny-county-public-schools-local-education-agency-leas-locations/about. type: dataset.

Allegheny County Office of Property Assessments. Allegheny County Property Assessments, May 2022. URL https://data.wprdc.org/dataset/2b3df818-601e-4f06-b150-643557229491. type: dataset.

Julian Chow. Differentiating urban neighborhoods: A multivariate structural model analysis. *Social Work Research*, 22(3):131–142, September 1998. ISSN 1070-5309, 1545-6838. doi: 10.1093/swr/22.3.131. URL http://swr.oxfordjournals.org/content/22/3/131.

Michael Dorman. *nngeo: k-Nearest Neighbor Join for Spatial Data*, 2022. URL https://CRAN.R-project.org/package=nngeo. R package version 0.4.5.

Reid Ewing, Rolf Pendall, and D. Chen. Measuring Sprawl and Its Impacts. Technical report, Smart Grown America, Washington DC, 2002.

Louis Guttman. Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2):149–161, 1954.

Shima Hamidi, Reid Ewing, Ilana Preuss, and Alex Dodds. Measuring sprawl and its impacts: An update. *Journal of Planning Education and Research*, 35(1):35–50, 2015.

Henry F Kaiser. An index of factorial simplicity. *psychometrika*, 39(1):31–36, 1974.

Yu-Sheng Li and Ying-Chih Chuang. Neighborhood Effects on an Individual's Health Using Neighborhood Measurements Developed by Factor Analysis and Cluster Analysis. *Journal of Urban Health*, 86(1):5–18, January 2009. ISSN 1099-3460, 1468-2869. doi: 10.1007/s11524-008-9306-7. URL http://link.springer.com/article/10.1007/s11524-008-9306-7.

John R Logan, Seth Spielman, Hongwei Xu, and Philip N Klein. Identifying and bounding ethnic neighborhoods. *Urban Geography*, 32(3):334–359, 2011.

Urbano Lorenzo-Seva, Marieke E Timmerman, and Henk AL Kiers. The hull method for selecting the number of common factors. *Multivariate behavioral research*, 46(2):340–364, 2011.

Brian A Mikelbank. Neighborhood déjà vu: Classification in metropolitan cleveland, 1970-2000. *Urban Geography*, 32(3):317–333, 2011.

Edzer Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446, 2018. doi: 10.32614/RJ-2018-009. URL https://doi.org/10.32614/RJ-2018-009.

Pennsylvania Department of Conservation and Natural Resources. Pennsylvania Local Parks Access Points, November 2015. URL https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=308.

Rafael H. M. Pereira, Marcus Saraiva, Daniel Herszenhut, Carlos Kaue Vieira Braga, and Matthew Wigginton Conway. r5r: Rapid Realistic Routing on Multimodal Transport Networks with R^5 in R. *Findings*, page 21262, March 2021. doi: 10.32866/001c.21262. URL https://findingspress.org/article/21262-r5r-rapid-realistic-routing-on-multimodal-transport-networks-with-r-5-in-r. Publisher: Findings Press.

Michael Reibel. Classification approaches in neighborhood research: Introduction and review. *Urban Geography*, 32(3):305–316, 2011.

Elizabeth Shay and Asad Khattak. Automobiles, Trips, and Neighborhood Type: Comparing Environmental Measures. *Transportation Research Record: Journal of the Transportation Research Board*, 2010(-1):73–82, January 2007. doi: 10.3141/2010-09. URL http://dx.doi.org/10.3141/2010-09.

Yan Song and Gerrit-Jan Knaap. Quantitative Classification of Neighbourhoods: The Neighbourhoods of New Single-family Homes in the Portland Metropolitan Area. *Journal of Urban Design*, 12(1):1–24, 2007. ISSN 1357-4809. doi: 10.1080/13574800601072640. URL http://dx.doi.org/10.1080/13574800601072640.

Yan Song and Roberto G. Quercia. How are neighbourhood design features valued across different neighbourhood types? *Journal of Housing and the Built Environment*, 23(4):297–316, December 2008. ISSN 1566-4910, 1573-7772. doi: 10.1007/s10901-008-9122-0. URL http://link.springer.com/article/10.1007/s10901-008-9122-0.

Markus D. Steiner and Silvia Grieder. Efatools: An r package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5(53):2521, 2020. doi: 10.21105/joss.02521. URL https://doi.org/10.21105/joss.02521.

Dan Trudeau and Patrick Malloy. Suburbs in disguise? examining the geographies of the new urbanism. *Urban Geography*, 32(3):424–447, 2011.

United States Census Bureau. LEHD Origin-Destination Employment Statistics (LODES), October 2021. URL https://lehd.ces.census.gov/data/#lodes. Type: dataset.

Thomas J. Vicino, Bernadette Hanlon, and John Rennie Short. A Typology of Urban Immigrant Neighborhoods. *Urban Geography*, 32(3):383–405, 2011. ISSN 0272-3638. doi: 10.2747/0272-3638.32.3.383. URL http://www.tandfonline.com/doi/abs/10.2747/0272-3638.32.3.383.

Carole Turley Voulgaris, Brian D. Taylor, Evelyn Blumenberg, Anne Brown, and Kelcie Ralph. Synergistic neighborhood relationships with travel behavior: An analysis of travel in 30,000 US neighborhoods. *Journal of Transport and Land Use*, 10(1), August 2016. ISSN 1938-7849. doi: 10.5198/jtlu.2016.840. URL https://www.jtlu.org/index.php/jtlu/article/view/840.

Kyle Walker and Matt Herman. *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*, 2022. URL https://CRAN.R-project.org/package=tidycensus. R package version 1.2.

Western Pennsylvania Regional Data Center. Geocoders, February 2021. URL https://data.wprdc.org/dataset/6bb2a968-761d-48cf-ac5b-c1fc80b4fe6a.

Western Pennsylvania Regional Data Center. City of pittsburgh permit summary, 2022. URL https://data.wprdc.org/dataset/city-of-pittsburgh-building-permit-summary.