Masters Program in **Geospatial Technologies**

The Edges of Areal Units

A Case Study in the Heterogeneous Effects of Assessment District Edges

Thomas Daniel Buckley

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

UNIVERSIDADE NOVA
ISEGI

UNIVERSITAT JAUME·I

WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

The Edges of Areal Units

A Case Study in the Heterogeneous Effects of Assessment District Edges

Dissertation supervised by

Prof Dr Tiago Oliveira

Prof Dr Mark Padgham

Prof. Dr. Jorge Mateu

March 2013

ACKNOWLEDGMENTS

I would like to thank Dr Marco Painho for the patience, thoughtfulness, open-mindedness, and support that he imbued in every facet of my academic development.

I would like to thank Professors Tiago Oliveira, Marcus Padgham, and Jorge Mateu for their patience, guidance, and support in developing this work. In particular, I would like to thank Dr Padgham for his help in making a complicated, vaguely formulated idea into a simple and testable hypothesis. I would like to thank Dr Oliveira for his instrumental role in pushing me to understand the methods and modeling necessary to test this hypothesis. I must also thank Drs Werner Kuhn and Simon Scheider for the discussions at the MUSIL seminar group at IFGI, as well as Schuyler Erle and Melissa Santos. I would also like to thank every person that I have spoken with about this work: classmates, colleagues and friends, and most of all: my family.

Finally, I am grateful to the European Commission for the Erasmus Mundus program. ISEGI, IFGI, and UJI are unique and excellent place for higher education, and I hope that the Erasmus Mundus program will continue to support all of the excellence in academic work and cross-cultural exchange that makes it such a rich place in which to be a student.

The Edges of Areal Units

A Case Study in the Heterogeneous Effects of Assessment District Edges

ABSTRACT

Areal units are used in a broad range of demographic and physical description and analysis related to surveying, reporting, navigation, and modeling. In The Modifiable Areal Unit Problem, Openshaw (1983) described how the arbitrariness of an areal unit's boundaries means that any measurement aggregated to it is to some extent arbitrary as well. Therefore, those who survey, model, and report information based on these units must be aware of their shortcomings as models for describing phenomena that they aggregate.

Here we propose to test an aspect of the Modifiable Areal Unit Problem, namely that the boundaries used by an application of modeling with areal units are not homogeneous in their relationship to the phenomena that they model. That is, here we focus on the general problem that the boundaries of a set of areal units aren't entirely arbitrary. Boundaries for these areal units specifically—and many others generally—are along physical and social features of the environment, which may have an internal effect on the phenomena that they describe as homogenous in the aggregate.

In this thesis, we use real estate sales data and assessor's neighborhood boundaries to develop a method for describing differences in the effect of the boundaries of areal units. It is hoped that the methods developed here could be applied to the analysis of other urban phenomena that are restricted, afforded, described, and modeled by boundaries.

KEYWORDS

Modifiable Areal Unit Problem

Spatial Ontology

Cartographic Methods

Residual Analysis

Regression Modeling

TABLE OF CONTENTS

Page:

INDEX OF FIGURES

INDEX OF TABLES

1. INTRODUCTION

Theoretical Framework

Areal units are used in a broad range of demographic and physical description and analysis related to surveying, reporting, navigation, and modeling. For example, the United States Census Bureau surveys and publishes information according to a nested set of areal units (e.g. Tracts, Block Groups, and Blocks). Consumer location services direct users according to areal units assigned with colloquial neighborhood names. And real estate assessors and agents model and describe the value of properties by areal units. In each of these cases, the definition of an areal unit is based on expert knowledge with a particular use model.

Census units are drawn for efficient surveying, homogeneity of demographic characteristics, and spatio-temporal continuity. Neighborhoods used in consumer location services are built to describe colloquial words for describing Place. And real estate assessment neighborhoods are constructed for accuracy and equity in describing how a complex set of physical and social characteristics relevant to housing are imbued with economic value.

In *The Modifiable Areal Unit Problem*, Openshaw (1983) described how the arbitrariness of an areal unit's boundaries means that any measurement aggregated to it is to some extent arbitrary as well. Therefore, those who survey, model, and report information based on these units must be aware of their shortcomings as models for describing phenomena that they aggregate.

Here we propose to test a qualification to the Modifiable Areal Unit Problem, namely that the boundaries used by an application of modeling with areal units are not homogeneous in their relationship to the phenomena that they model. That is, here we focus on the general problem that the boundaries of a set of areal units aren't entirely arbitrary. Boundaries for these areal units specifically—and many others generally—are along physical and social features of the environment, which may have an internal effect on the phenomena that they describe as homogenous in the aggregate.

Background

Roads, parks, highways, and other physical features cut through urban areas, affording and prohibiting activity in a city. In addition, a physical feature—such as a particular street—can often serve as a named boundary or delineating feature between areas or neighborhoods.

Homogeneous Statistical Areas, such as Census Blocks often based on these boundaries. For example, the US Census explains about their areal units:

Census blocks, the smallest geographic area for which the Bureau of the Census collects and tabulates decennial census data, are formed by streets, roads, railroads, streams and other bodies of water, other visible physical and cultural features, and the legal boundaries shown on Census Bureau maps. (Census, 1994)

As indicated in the text, the smallest census units are sensitive to higher-level areal units such as administrative boundaries. Of course, there are strong reasons to believe that these administrative units are critical determinants of the things that happen within them. Laws and taxes vary across them. And at the areal units below, we can see that significant environmental features make up their boundaries.

Other HSA's are often based on census boundaries: real estate assessor's neighborhoods, urban planning neighborhood clusters. In the Annex we present a comparison of these units in our area of interest.

While much has been written about HSA's, including doing analysis based on demographic data based on them, the theory behind their construction (in urban planning or economics), the effect of scale and boundary delineation on qualitative analysis (the Modifiable Areal Unit Problem), the effect of the boundaries which define them on people psychologically, less attention has been given to how the boundaries of these HSA's may effect the processes that occur within them.

It is reasonable to expect that urban boundaries affect human behavior, and therefore may create homogeneous areas. However, it may also be reasonable to expect that boundaries have qualitatively different effects on human behavior and related urban processes. We might expect, for example, that two different kinds of boundaries have different effects.

Objective

In this thesis, we use real estate sales data and assessor's neighborhood boundaries to develop a method for describing differences in the effect of boundaries. It is hoped that the methods developed here could be applied to the analysis of other urban phenomena that are restricted and afforded by boundaries, such as crime and movement.

Hypothesis

The aim of this work was to test the following:

1. Residential home sales can be modeled with regression
2. The residuals of said regression models reveal patterns of spatial dependence and heterogeneity in the housing market.
3. The boundaries used to model and understand spatial dependence and heterogeneity are heterogeneous in their relationship to the housing market they are used to model.

General Methodology

The general methodology of the thesis is borrowed from regression analysis for housing in economics, although it differs in overall objective. In real estate economics, there is typically a focus on the identification of the effect of a particular parameter in a global model. In the case of this thesis, we do not hypothesize that boundaries in general have an effect, but that certain boundaries may have an affect on houses that they circumscribe. Ex-ante, we may be able to test for the effect of these particular boundaries using regression models, but our concern here was on developing methods for first identifying those boundaries among a set that we might want to look at more closely.

Dissertation Organization

This study is comprised of five chapters. The first chapter—the introduction—includes a brief overview of the background, objective, hypothesis, and general methodology. The second chapter—the literature review—reviews relevant literature on urban planning & design, real estate economics, and homogeneous statistical areas. The third chapter provides a description of the data and the methods used to test the hypothesis, the results of which are discussed in the fourth chapter. Finally, the fifth chapter details some conclusions based on the results, discussed their academic relevance, and suggests future directions for research.

2. Literature Review

Overview

Much of this literature review will describe methods for modeling house prices. While in theory the spatial processes that describe some important facets of housing may be modeled directly, in practice many economists and assessors use homogeneous geographic statistical areas. Little is understood about how the cartographic nature of these geographic units, in particular their boundaries, relate to the processes that they circumscribe. Our hypothesis is that these boundaries are heterogeneous in their relation to the housing sales process. Because they are often drawn along common physical features of the urban environment (roads, parks, etc), we borrow some of the language of urban design to describe them.

Housing (Hedonic) Regression

First, we review applied modeling for housing economics. In the analysis section of this thesis we will use a method that is traditionall known as "hedonic" regression to hold constant all other variables as we look at the relationship between boundaries and real estate sales transactions. Sheppard (1997) provides one of the simplest definitions of the theory of how and why hedonic regression might be used.

> *Imagine, for a moment, that you are a private investigator or market researcher studying the demand for food. You have a particular disadvantage, however, in that you have been banned from entering the local grocer. You have found a place outside where you can sit and photograph shoppers as they approach the checkout counter, and from these photographs you can pretty much tell what foods each customer has purchased (although some items may be obscured in the shopping basket) and the total cost of all items combined. By bribing a contact at the local bank, you are able to find out each shoppers income. That is all the information that you have. From this, can you infer the demand for eggs? Can you determine how much households would be willing to pay to remove sugar import quotas? (Sheppard, 1999)*

In the housing market, hedonic regression is used to determine the demand for bathrooms, for example, rather than for eggs. In our case, the model is simplified. We only have to use the information about sales and the components of the home to be able to accurately predict what the price of a house is. That is, were we doing food market research, we would only be interested in using the information available to us to accurately predict the price that a particular basket at the grocer would cost.

Two features of this literature motivate this approach to hedonic regression. The first feature is that the theoretical basis for the specification of the hedonic regression and the interpretation of the parameters is not well developed. As such, there is considerable variation in the variables and model specifications that are used, and little common agreement on which ones are appropriate. In turn, the success of these models is typically assessed in terms of accuracy, for house prices and for specific parameters (Malpezzi, 2003).

The second motivating feature of the review is that, in housing models, missing variables are a given, but there is a high degree of correlation among missing variables and those included in the regression. Because of this correlation, it can be difficult to interpret individual parameters. However, this high correlation among descriptive variables also means that models can be explanatory for price overall, even with missing variables. The objection that in modeling a missing variable may not be accounted for is legitimate, but in housing models, variables often proxy for one another in the overall goal of price prediction (Malpezzi, 2009). For our question in particular, the goal is in holding price, based on independent variables, consistent, so that we can inspect variation in value across space. To that extent, any set of variables that accurately holds price constant across space will be useful to us. On the other hand, location itself is a proxy for many variables, a fact we will discuss in the next section.

2.3 Spatial Aspects of a Housing Model

"Location" is a common parameter or vector of parameters in real estate modeling. However, the geographic structure of housing data is difficult to parse because it contains at least two related but distinct aspects. It can be difficult to identify the effect of one or the other in the model (Bourassa, 2003).

The first aspect is that the qualities and value of one's house depend on the qualities and value of one's neighbors house. Some studies target this dependency directly by specifying spatial interaction among the prices and parameters of houses (Anselin, 2002); others map variation in parameters estimates geographically (Brunsdon, Fotheringham, & Charlton, 1996); some inspect the spatial distribution of residuals (Dubin, 1992); and most just use aggregate statistical regions and neighborhoods (Bourassa, 2003).

2.4 Geographic Statistical Units

These aggregate regions and neighborhoods might be understood to capture both the first aspect of location (dependence) and the second: that many variables that are important to housing (e.g. school, work, noise, demographics) are spatially fixed (Bourassa, 2003). For example, because one of the requirements of the establishment of new census tracts is homogeneity of demographics among the households within them, we might expect that not only is a sum of demographics described by the tract's geography, but so too is the spatial distribution of a group of highly similar people.

Housing economists have long theorized that there are market segments in housing, and many assume that they are probably geographic (Tiebout, 1956). A market segment contains a set of goods that are substitutes, and therefore more or less similar in use value to the buyer. Assuming census tracts are homogenous, most economists have assumed that they are the best representations of market segments (Goodman & Thibodeau, 2003; Wachter & Wong, 2008).

We might thank that it better to model housing prices by areal units because demographic characteristics correspond to them. Bourassa (2003) tests this hypothesis by comparing the accuracy of two models: one, with market segments based on geographic units delineated by assessors and another, with market segments that are derived on similarity comparison across social and physical variables and then applied to houses irrespective of geographic continuity. They find that the assessor's geographically contiguous submarket definitions are more accurate.

Another approach is to create a model of neighborhoods using only the physical components of houses in a regression, and then inspecting residuals for dependence and heterogeneity. Dubin (1991) creates contour-plots of Baltimore based on this method. She also suggests that it may be possible to create appraisal boundaries based on these surfaces. Work on the effectiveness of neighborhoods created from residuals continues, with the authors finding that "spatial trend analysis and census tract variables do not perform nearly as well as neighboring residuals(Case, Bradford, & al., 2004).

One method for accounting for the spatial error in residuals is to introduce the error into the model via a spatial weights matrix. Anselin outlines a plethora of ways to do this (Anselin, 2002). However, there is evidence that the results of the regression model become highly dependent on the characteristics of the spatial weights matrix itself. Furthermore, the methods for applying spatial weights matrices do not account for the heterogeneity in distances among houses (Bell, 2000). Given the sensitivity of our question to the heterogeneous qualities of individual neighborhoods and their boundaries, a regression model with a spatial weights matrix may obfuscate heterogeneity in service of global accuracy.

In sum, there are numerous methods for account for both spatial dependence and local neighborhood qualities in a housing regression. In this thesis, we focus on the method most commonly used, that of areal units. We take the tack of many economists, modeling dependence in the form of demographic characteristics, assuming that characteristics such as age, income, race, and education as aggregated to census areal units are useful variables for describing spatial dependence as well as important qualities of a neighborhood. By using these variables, rather than a spatial regression, we elude the difficulty of having a secondary model of space, obscuring the focus on the spatial relationship between houses and boundaries.

The Psychology of Urban Boundaries

As discussed previously, the boundaries that circumscribe areal units, in particular assessment neighborhoods, are based on physical features of the urban environment. Therefore, in order to understand the biases that may be present in modeling human activity with areal units, it may be useful to have a general theory for how physical features of the urban environment affect people psychologically. Kevin Lynch offers us a general and widely cited general theory in *The Image of The City.* He uses a mix of personal interviews, fieldwork, and cartographic comparison to develop an ontology of urban design psychology based around: Paths, Edges, Districts, Nodes, and Landmarks. At the risk of oversimplifying: Paths allow movement, Edges inhibit it, Nodes and Landmarks may be important for ordinal sense of direction and place, and Districts are places that may be defined by all of the above, about which one can feel inside of or outside of.

His theory of Edges and Nodes was nuanced enough to capture and describe several kinds of relationships that might occur at a boundary. For example, here he is on Edges and Paths "An edge may be more than simply a dominant barrier if some visual or motion penetration is allowed through it—if it is, as it were, structured to some depth with the regions on either side. It then becomes a seam rather than a barrier, a line of exchange along which two areas are sewn together (Lynch, 1960, P 100)."

He compares the qualities of these ontological elements across three cities with very different urban designs: Boston, Jersey City, and Los Angeles, producing cartographic visualizations of their elements, such as the image of Boston in Figure 1.

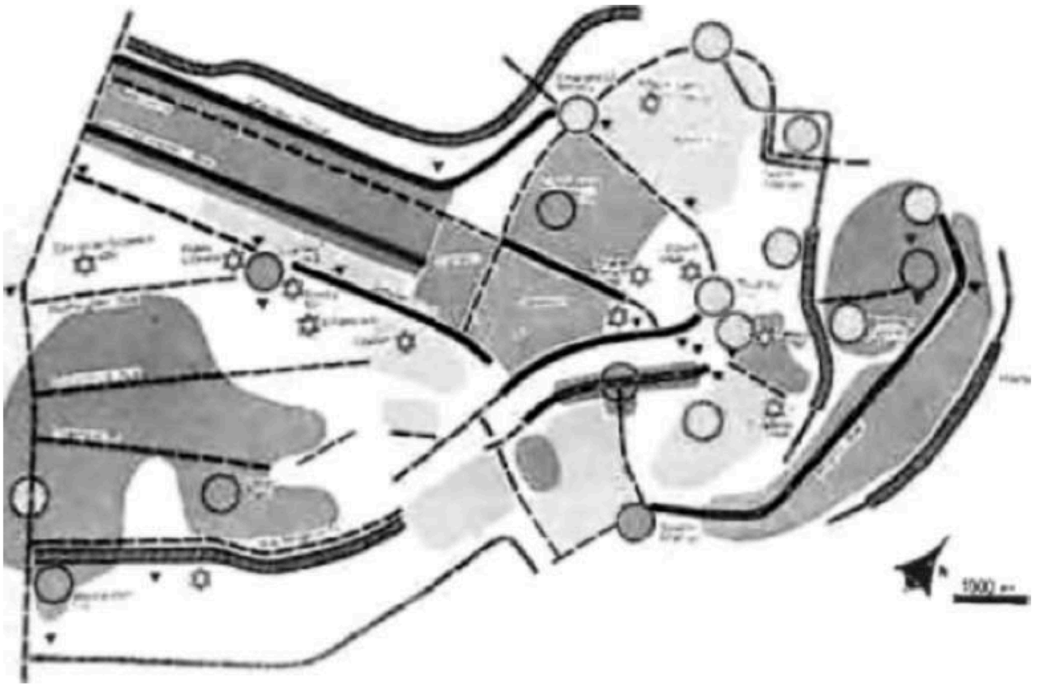


FIG. 3. ***The visual form of Boston as teen in the field***

Figure 1 – Edges, Nodes, and Districts in Boston (Lynch, 1960)

Lynch's *The visual form of Boston as seen in the field* provides a cartographic example for how we might think about the hypothesis tested in this thesis. Lynch used field interviews and asked individuals to draw maps of the city in order to sketch, for example, the Edges, Nodes and Paths displayed in the map. In our case, we would hope to derive Edges from transactions in the real estate market. In economic terms, we would hope to derive boundaries from revealed preferences, which are "tastes that rationalize an economic agent's observed actions (Beshears, 2008)." While there is much to be said for simply asking people about these boundaries, it is possible that a revealed preferences method based on real estate transactions could capture psychological boundaries that may operate on different temporal and contextual scales. To that extent, in the following sections, we describe how real estate investment was central to the history of neighborhood development in Washington, D.C., using Lynch's theory to briefly describe how Edges and Nodes interacted with real estate investment.

3. Data and Methods

This chapter introduces the study area, the data, and the methodology for analysis. It is divided into two sections. The first introduces some general facts about Washington, D.C., its urban plan and neighborhoods, and housing. The second reviews relevant considerations about the modes of analysis and the data used.

Study Area

Washington, D.C. is the capitol city of the United States. It is an independent District with a 2010 population of roughly 6 hundred thousand. It is located within the Washington Metropolitan Statistical Area, population 5.6 million, at the southernmost tip of the US Megalopolis, population 44 million (US Census, 2010). As the center of a metropolitan region, Washington has seen waves of population gain and loss that roughly follow those of other United States cities. However, Washington has recently experienced population growth, after decades of decline.

Population Growth, Housing, and Urban Planning (1790-1900)

While the city saw early occupation in the form of trading posts and ports by Native Americans and then European Settlers for many hundreds of years, it only achieved the pretense of what a current inhabitant might recognize as a city within the past two hundred years. In his first major urban design project Pierre L'Enfant was commission by George Washington, the first President of the United States to design the city (Rybczynski, 2010). L'Enfant designed a plan that is still relevant to the boundaries of the city today (Figure 2).

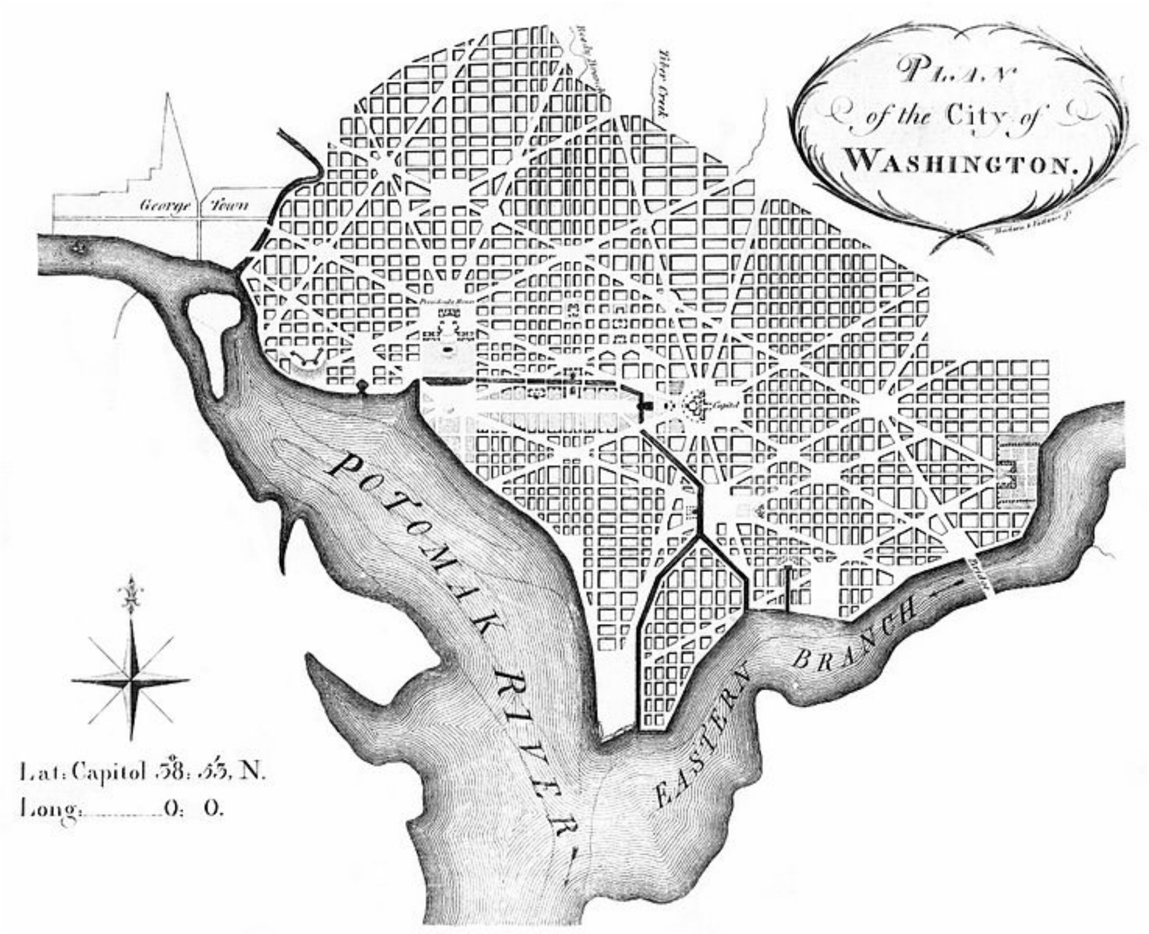
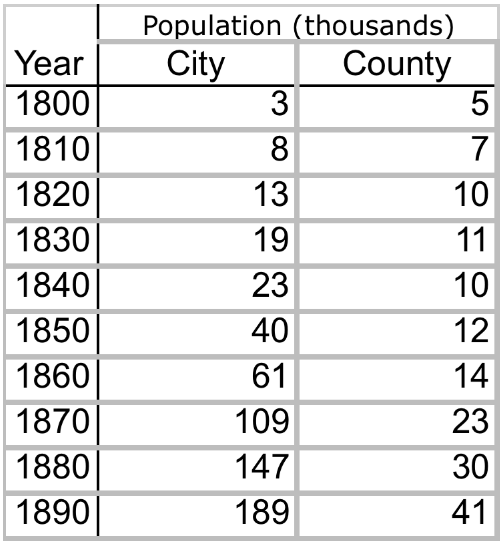


Figure 2 - English: Plan of the City of Washington, March 1792, Engraving on paper

Washington’s population grew steadily through the 1800s, reaching nearly half of its present-day population in 1890, with the majority of the population residing within the boundaries of L’Enfants plan, known then as “Washington City.”(See Figure 3) City and County were officially consolidated in 1871, although census records are available for both separately until 1890.

Table 1 – Population Growth Patterns in Washington, D.C. from 1800 to 1890 (Source: US Census)



	Population (thousands)	
Year	City	County
1800	3	5
1810	8	7
1820	13	10
1830	19	11
1840	23	10
1850	40	12
1860	61	14
1870	109	23
1880	147	30
1890	189	41

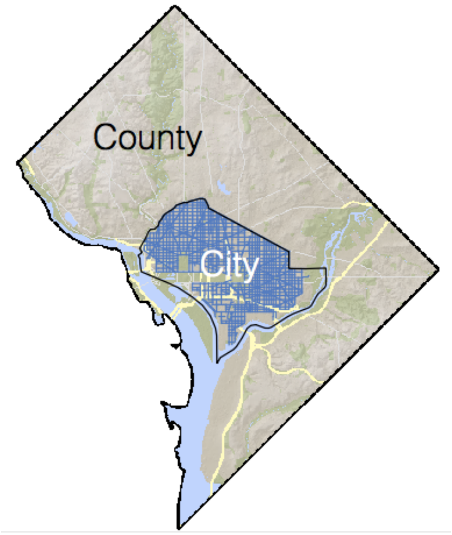


Figure 3 - Map of Present Day Washington D.C. with L'Enfants Historic Plan for "Washington City" and Washington County

While Historic "City" and "County" Washington kept pace through the beginning of the century, its clear that population grew much more quickly in the historic "City" starting around 1830 and continuing through 1890. Building construction data is somewhat inconsistent for these years, but population data provides us with a proxy for the history settlement in the center of Washington, D.C. mainly.

Population Growth, Housing, and Urban Planning (1900-2000)

As we see in Figure 3, almost all construction of new housing after 1920 takes place outside of the historic "City." By comparing Table 2 and Figure 3, we find that a post-World War II population boom roughly corresponds with waves of construction across the city. These waves in construction also result in clusters of residential developments of a consistent aesthetic quality. While we will account for building age specifically in our methods, it is important to note the historical spatial clustering of development as it corresponds to population growth. Building age has been shown in the literature review to be an important dimension along which housing markets may be segmented.

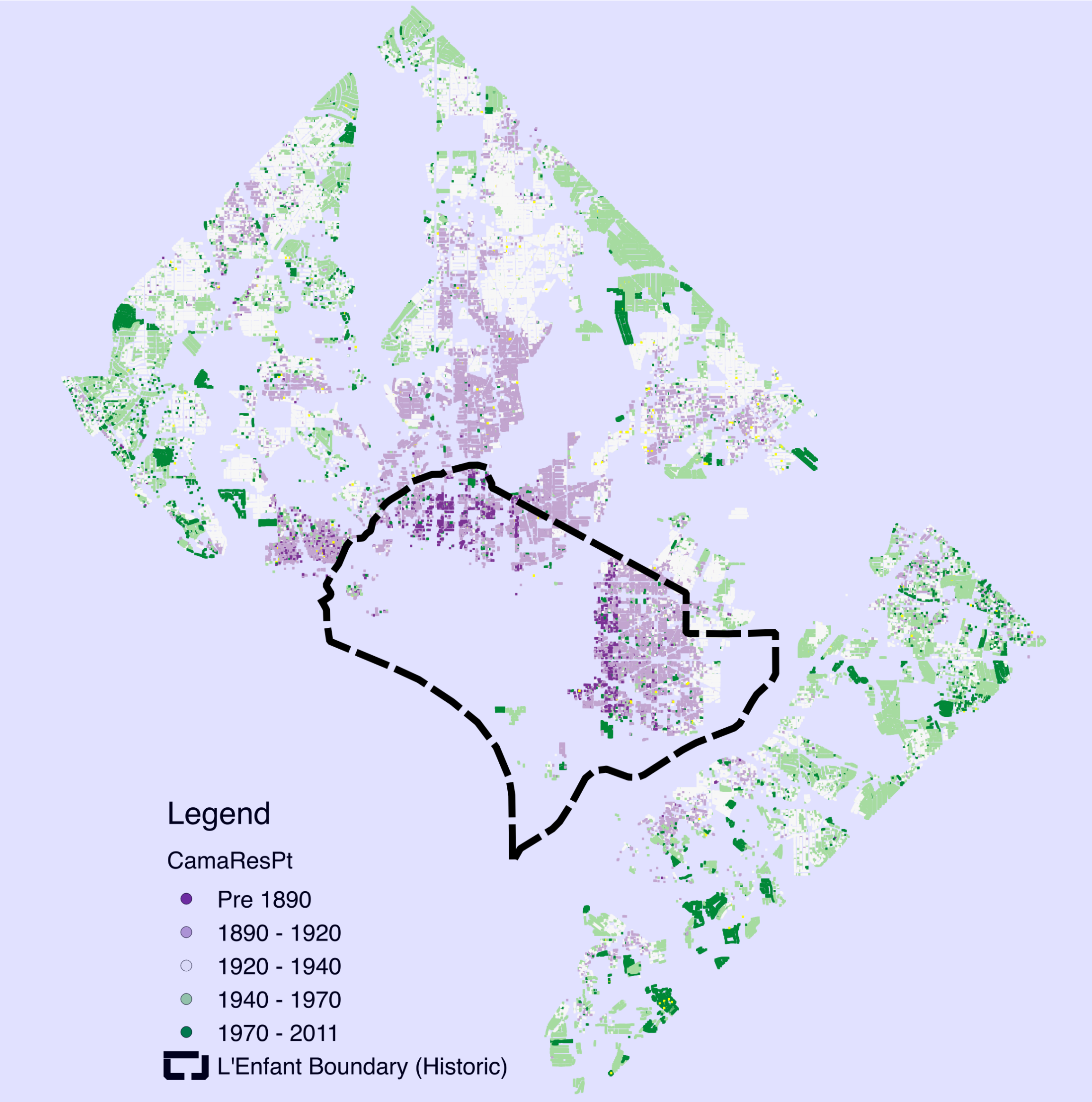


Figure 4 - Year of Construction of Residential Buildings in Washington DC (source: DCGIS)

Table 2 - 20th Century Population (in thousands, source: US Census)

Year	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
Popul.	279	331	438	487	663	802	764	757	638	607	572	602

History of Neighborhoods in Washington D.C.

Residential real estate investment is an important component of neighborhood definition in the history of the development of neighborhoods in Washington, DC. In the following section, we provide a brief historical background on some of the history and geography of neighborhoods in Washington, D.C. While the definition of the word neighborhood is not our focus, it is hoped that this introduction will provide the reader with more historical understanding of what a neighborhood is historically in Washington. This history is relevant to our purposes because assessment neighborhoods share some historical spatial similarity and continuity with colloquial and official neighborhood names. However, it is perhaps more important for our purposes to note that colloquial and official neighborhoods can also change dramatically in name and spatial extent over time.

Given that the majority of Washington's population in 1887 was within the L'Enfant Plan area, the neighborhood names available to us from that time, for that area, provide us with a useful starting point to understanding the dynamics of DC's neighborhoods historically. The persistence of colloquial neighborhoods might be thought of as the persistence of a name and of the spatial extent that it refers to (Maue, 2013).

In Washington, there is at least one example of a colloquial neighborhood that has persisted in name since 1887. The colloquial neighborhood "Foggy Bottom" in the map of Official and Colloquial Neighborhood Names in 1877 is still used today, and its geography is defined similarly, although it is also now an official neighborhood both according to the Office of Planning and the Assessor's Neighborhoods. Foggy Bottom has gone through numerous changes in use, from a mix of Military Encampments (Civil War) and Industrial Sites to Residential, University, and Performing Arts center.

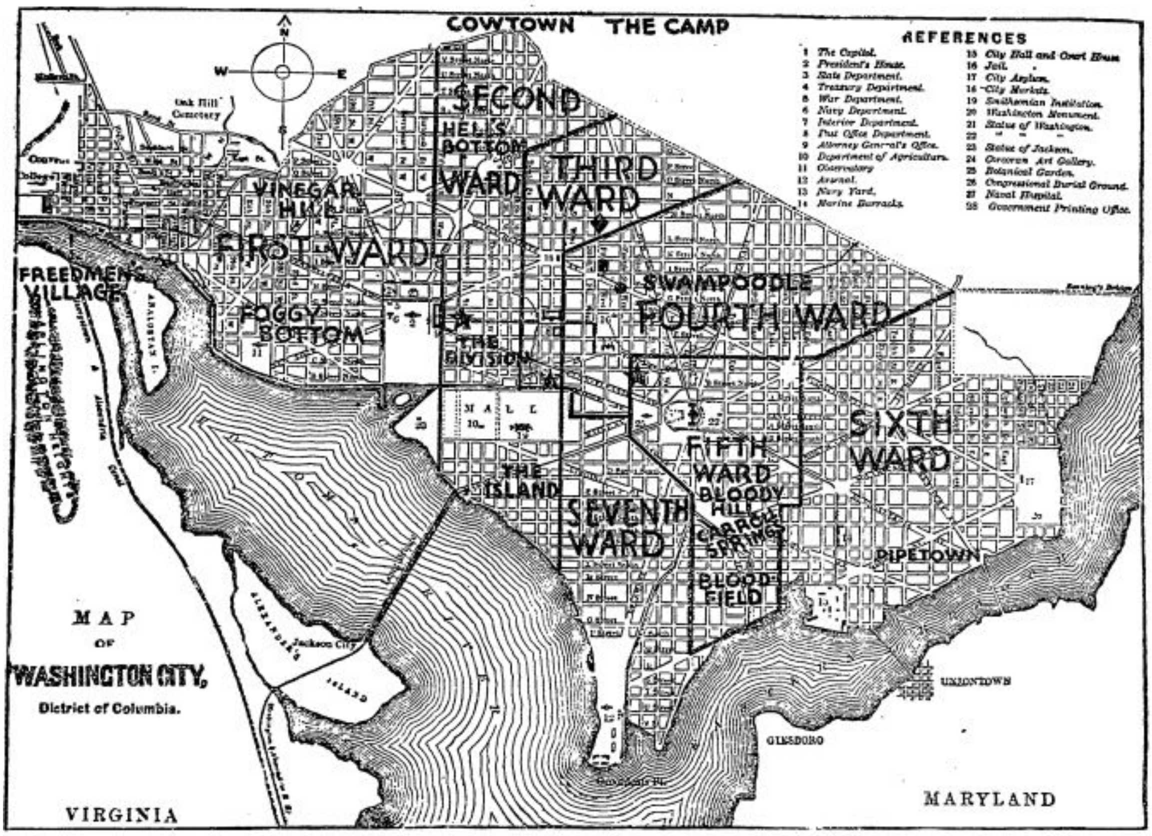


Figure 5 - Official and Colloquial Neighborhoods in Washington, D.C. in 1877 (Washington Post)

On the other hand, most of the other colloquial neighborhood names from this map are not presently used, with the exception of Swampoodle, which is used in the Gazetteer of Yahoo! Inc., but it is very rarely—if at all—used in present day colloquial language. The installation of a major trans-city train station in the middle of historical Swampoodle in 1907 (Amtrak History and Archives, 2013) may account for why what was once considered an identifiable geographic area lost its usefulness. We will see later that this train line forms an important present-day boundary for real estate in the city.

The historical neighborhoods Vinegar Hill and Hell's Bottom are now both commonly known by the traffic circle parks at their center, Dupont Circle and Logan Circle, respectively. Both circles were only formally developed as part of a major urban planning project several years before the map in Figure 4 was made. These projects were followed by significant residential construction projects, with the most exceptional architectural examples adjacent to the circles (Lanius & Park, 1995). To that extent, we can see how—to use Lynch's terminology—Nodes also define a sense of place in DC's history of residential neighborhoods. In these cases, the Nodes were actually a kind of improvement on the pre-existing historical plan that coincided with residential investment.

The historical plan for historical Washington is still important in the city's definition of neighborhoods today. When describing Washington's communities present-day Planners have described it as "Sixteen Cities Within the City. (Figure 20-Appendix)" Here was can see again that the boundaries for these communities are largely drawn along the historic boulevards of the city plan. Later, when we look at assessment neighborhoods and census tracts, these boulevards will predominate as lines along which borders are drawn.

Neighborhoods in Washington, D.C. are defined in large part according to an urban planning history, although they critically depend upon residential investment. In addition, their names are also defined according to the uses of those who refer to them. In Washington, colloquial names and planned neighborhood improvements often—though not always—change together, shifting into new neighborhoods according to the Edges and Nodes that prohibit and enable activity within the city. Our goal in this thesis is to determine whether residential housing purchases relate to the Edges defined by assessment neighborhoods in present day. As such, in the next section we turn to more recent developments in residential real estate investment.

Housing Sales (2000-2012)

In this section, we will broadly review the temporal and spatial dimensions of the housing sales transactions process that we considered in our analysis. Housing sales data, collected for tax purposes, represent a complex, multidimensional process.

The median sales price of homes in Washington has roughly doubled over the past 12 years across the entire city (See Figure 8). While the Northwest Quadrant represents a much larger total increase in price, all Quadrants have increased.

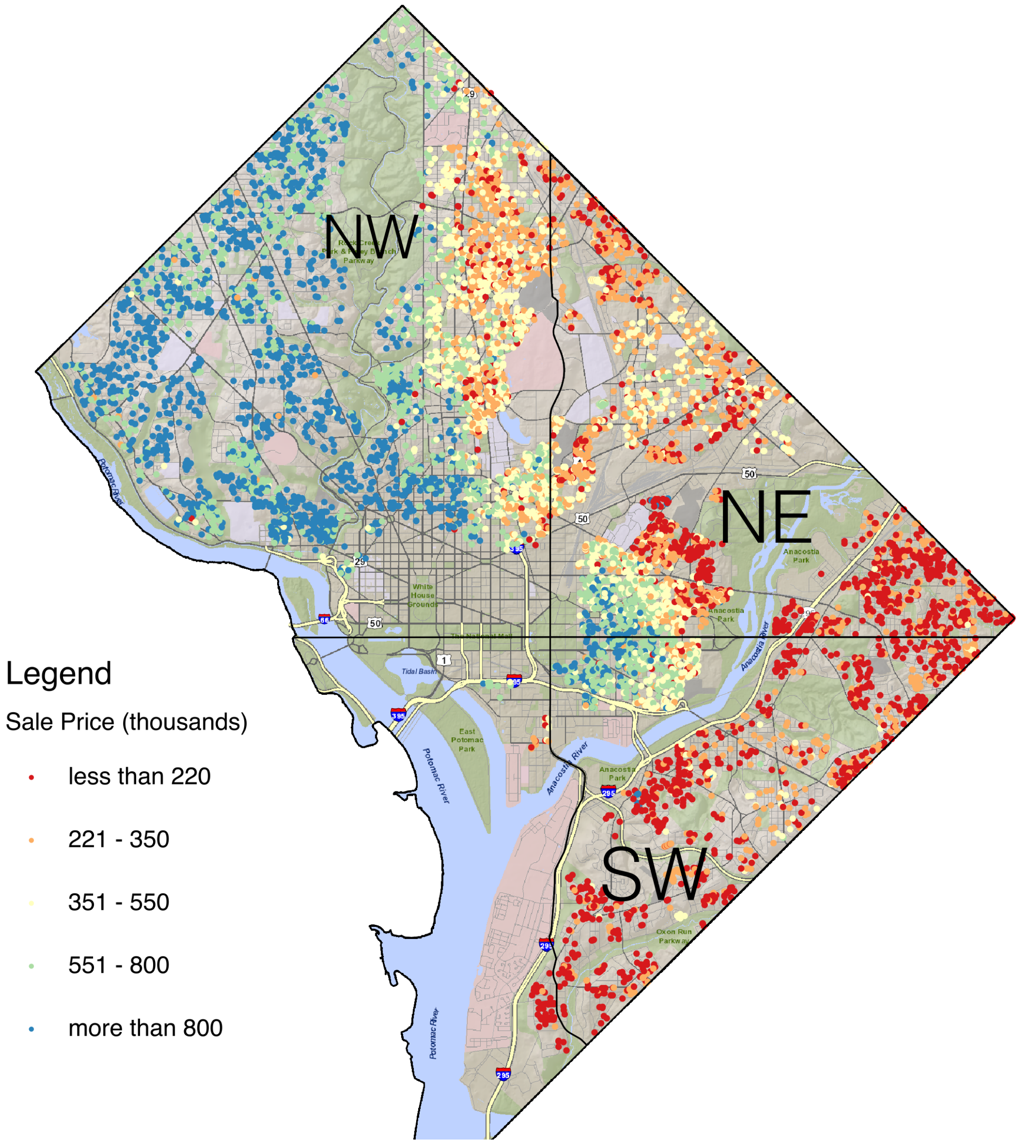


Figure 7 - Sales Prices of Houses from 2010-2012 with city Quadrants

The spatial distribution of the prices of the years we selected, the most recent available is shown in Figure 7. Perhaps unsurprisingly, there appears to be some local spatial autocorrelation in sales prices. We model this local autocorrelation according to census demographic characteristics, assuming based on literature reviewed that these census demographic characteristics correspond to social and geographic market segments.

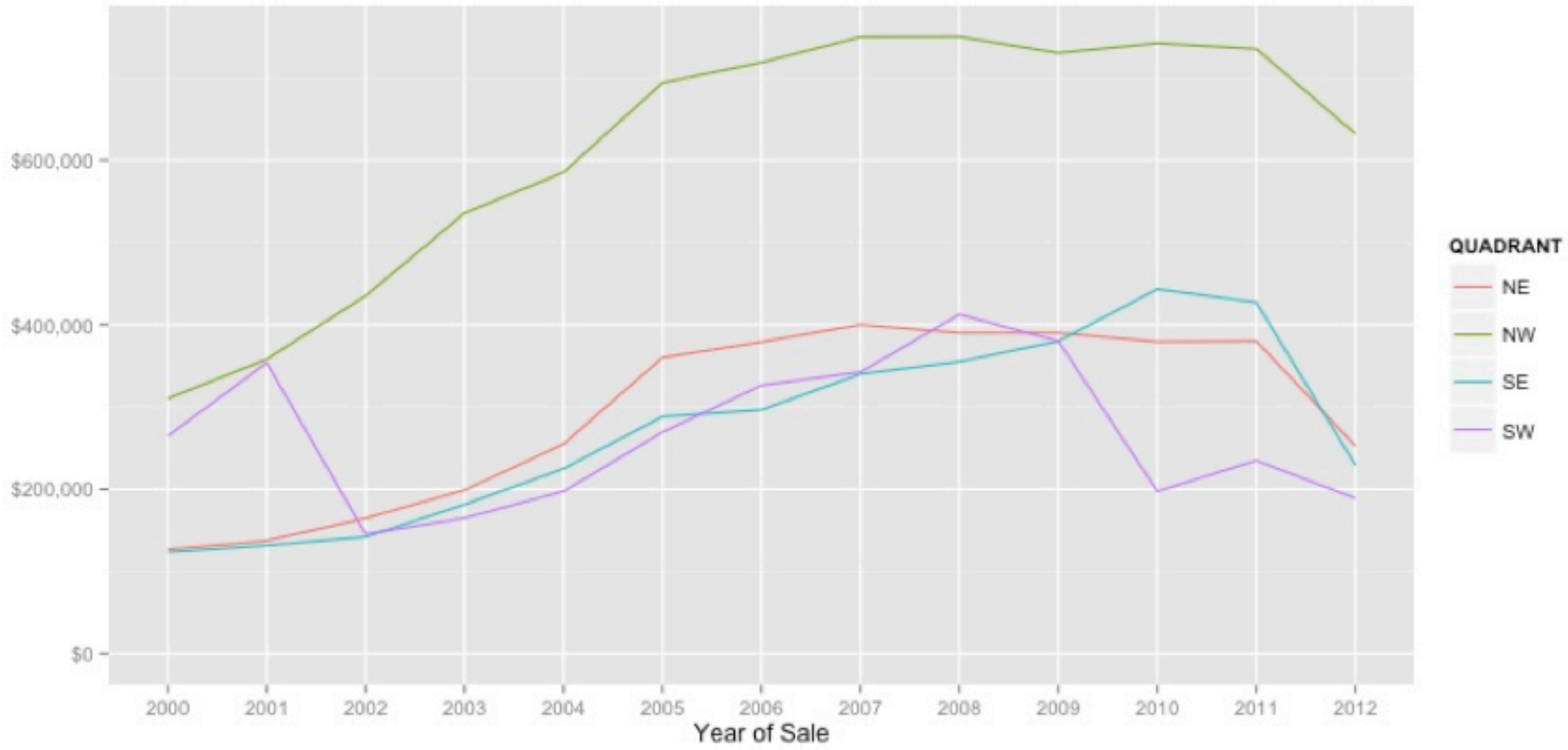
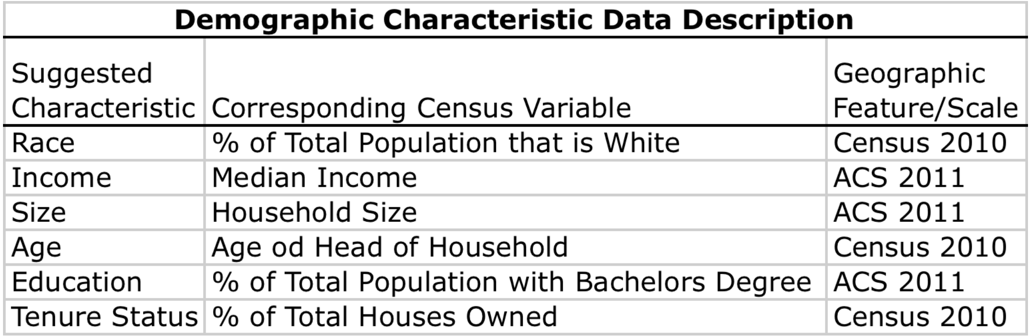


Figure 8 - Median Sales Price (2000-2012, Source: DCGIS, OTR)

Demographic Characteristics by Census Areal Units (2010-2011)

In our analysis, we assume that demographics are one of the key descriptors of how markets are spatially and socially segmented. Demographic characteristics were selected from the 2010 US Census and the American Community Survey 2011 3-year estimates. The latter were available at the census tract, and the former at census block group (see chart). We chose all available variables that were identified in the literature review as variables known to explain variation in house prices.

Table 3 - Demographic Characteristics Considered for Analysis



Demographic Characteristic Data Description		
Suggested Characteristic	Corresponding Census Variable	Geographic Feature/Scale
Race	% of Total Population that is White	Census 2010
Income	Median Income	ACS 2011
Size	Household Size	ACS 2011
Age	Age od Head of Household	Census 2010
Education	% of Total Population with Bachelors Degree	ACS 2011
Tenure Status	% of Total Houses Owned	Census 2010

Below, we map these demographic characteristics. Our goal in presenting maps of them is twofold: 1) to give the reader a better understanding of the geographies at which these important variables are available and 2) As discussed in the literature review, housing economists have found that demographic characteristics may be a direct or indirect (via signaling) explanation for local spatial autocorrelation. Therefore, we also present demographic data by census tract.

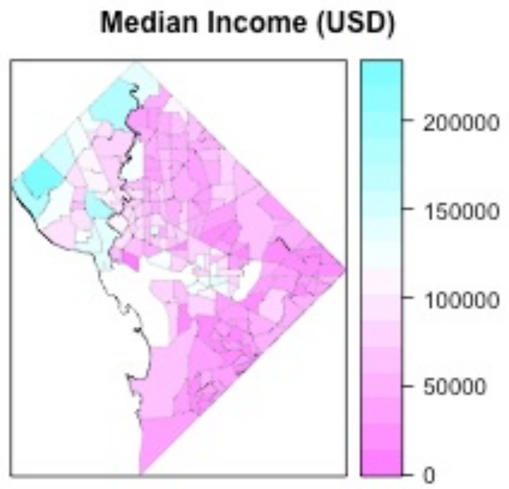


Figure 6 - Map of Median Income by Census Tract (2011 ACS)

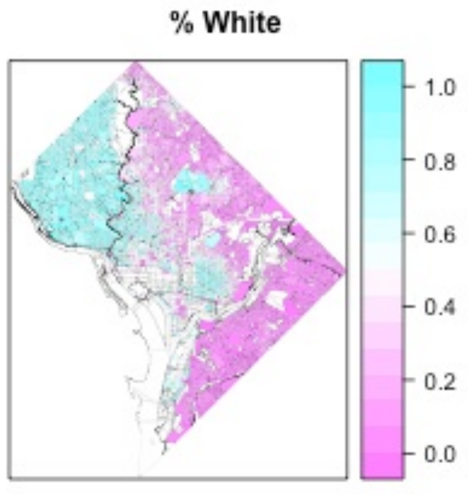


Figure 7 - Map of % of White Population by Census Block Group (2010)

Both median income and percent white population might seem to explain a significant amount of the spatial distribution in house prices. However, their spatial distribution is not identical.

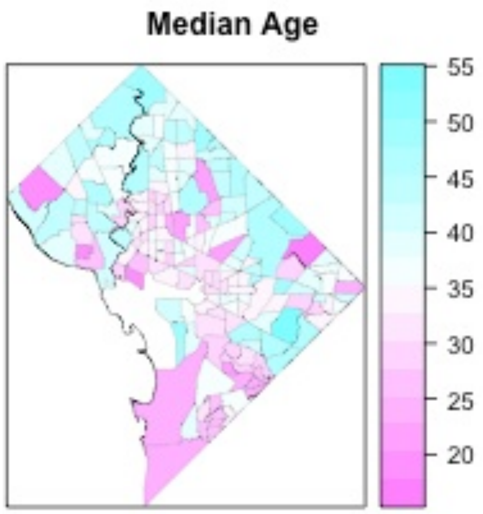


Figure 8 - Map of Median Age by Census Tract (2011 ACS)

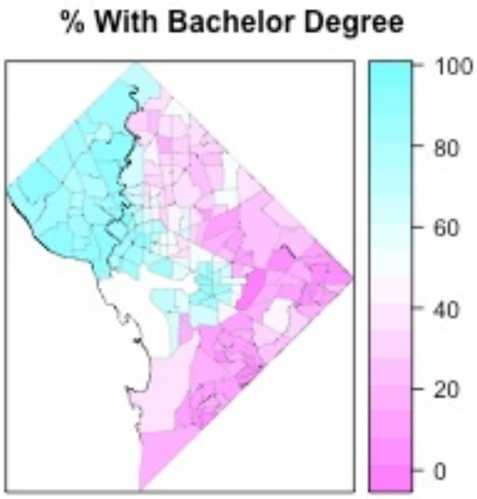
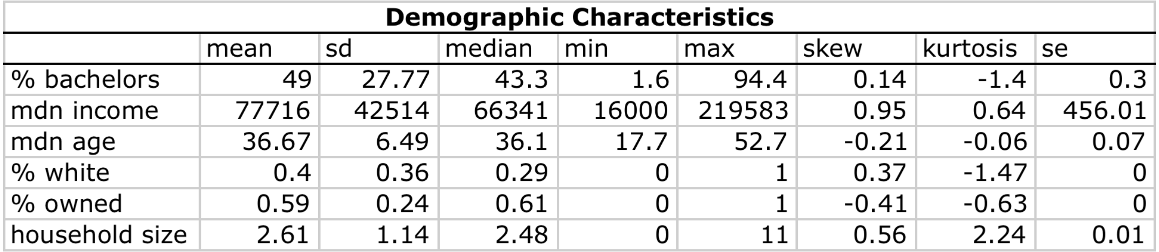


Figure 9 - Map of % of Population with Bachelors Degree by Census Tract (2011 ACS)

While it is unsurprising that Median Age of the head of the household does not seem to correspond spatially to sales prices, we include it in the later principal component analysis and then regression analysis because, as discussed in the literature review, there is evidence that housing markets are segmented by age. On the other hand, the percent of the population with Bachelor's degrees seems to be collinear with Median Income and Percent of Population that is White. We will speak in more depth about dealing with such colinearity in the methods section on Principal Component Analysis.

Since our dependent variable is the sale price of an individual house, we interpolate all demographic characteristics to the house. Therefore, summary statistics in Table 4 describe demographic characteristics of houses that were sold. Note that variables are defined in Table 3.

Table 4 - Demographic Characteristics Summary (2010-2012)



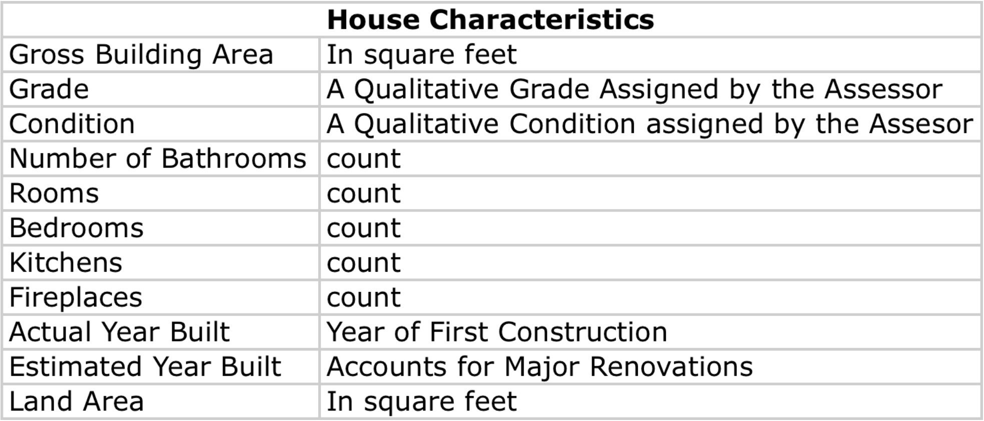
Demographic Characteristics

	mean	sd	median	min	max	skew	kurtosis	se
% bachelors	49	27.77	43.3	1.6	94.4	0.14	-1.4	0.3
mdn income	77716	42514	66341	16000	219583	0.95	0.64	456.01
mdn age	36.67	6.49	36.1	17.7	52.7	-0.21	-0.06	0.07
% white	0.4	0.36	0.29	0	1	0.37	-1.47	0
% owned	0.59	0.24	0.61	0	1	-0.41	-0.63	0
household size	2.61	1.14	2.48	0	11	0.56	2.24	0.01

House Characteristics (2010-2012)

We use variables on characteristics of houses to describe heterogeneity in home prices as it relates to amenities of the house itself. Records of home sales and the characteristics of the houses sold are available from the DC Office of Tax and Revenue. For our initial model, the selected houses sold between January 1, 2010 and Feb 28th, 2012. It later became obvious that we could also use 2 other samples of house sales from 2006 to 2007 and from 2008 to 2009 in order to further test both the model and our results. Summary data for these years is presented in the Annex.

Table 5 - House Characteristics Data Dictionary



	House Characteristics
Gross Building Area	In square feet
Grade	A Qualitative Grade Assigned by the Assessor
Condition	A Qualitative Condition assigned by the Assesor
Number of Bathrooms	count
Rooms	count
Bedrooms	count
Kitchens	count
Fireplaces	count
Actual Year Built	Year of First Construction
Estimated Year Built	Accounts for Major Renovations
Land Area	In square feet

Grade and Condition are both qualitatively assigned by assessors as part of the assessment process, which includes field inspections based on outlier detection over time. Grade refers to the overall quality of the house apart from the state of its maintenance. Condition refers to the maintenance condition. Estimated year built is based on a weighting system of the most recent year of major renovation and the first year that the house was built (Office of Tax and Revenue, 2013).

Table 6 - House Characteristics (2010-2012)

House Characteristics

	mean	sd	median	min	max	skew	kurtosis	se
GBA	3061.9	1573.1	2779.5	360	23401	2.54	14.38	16.87
GRADE	4.29	1.43	4	2	12	1.38	2.21	0.02
CNDTN	3.71	0.82	4	1	6	0.18	0.88	0.01
BATHRM	2.91	1.25	3	0	12	0.73	1.81	0.01
ROOMS	7.39	2.36	7	0	24	1.68	4.7	0.03
BEDRM	3.46	1.16	3	0	16	1.63	5.87	0.01
KITCHENS	1.26	0.66	1	0	4	2.95	8.68	0.01
FIREPLACES	0.65	0.94	0	0	13	2.39	10.94	0.01
AYB	1932	30	1927	1791	2011	0.86	1.46	0.33
EYB	1966	17	1961	1930	2012	1.22	1.23	0.18
LANDAREA	3051	3191	2115	0	155905	15.27	624.1	34.22

It is evident from the summary statistics that the median house was first built in 1927, underwent major renovations by 1961 or after, has 3 bathrooms, 7 rooms, 3 bedrooms, 1 kitchen, no fireplace, and has more square footage of built area than the land on which it sits. There are, of course, extreme houses with as many as 13 fireplaces, 12 bathrooms, and at least one house that was just a shell.

Many Variables, Multicollinearity, and Principal Component Analysis

Given the large number of variables, we had to consider ways of reducing the number of dimensions that describe housing. Furthermore, because many of these variables are collinear, and because our eventual mode of analysis is dependent upon linear regression, we attempted to use Principal Component Analysis(PCA). In the regression stage, the meaning of these principal components became unclear, in cases in which all regression independent variables were principal components, and in mixed models. However, we did find that some of the transformations of the data for the regression into their ordinal equivalents were still useful. As such, we describe them below.

One challenge in using Principal Component Analysis was identifying which of the many characteristics of a house could be considered ordinal. In particular, the architectural qualities of the building (e.g. a row middle, row-end, or detached house), the interior flooring type (e.g. cement, parquet, tile), the exterior wall type (e.g. stone, stucco, vinyl) do not have an immediate ordinal identification. However, we do have access to previous sales records, which indicate the relative value of these qualities. Since we are interested in explaining variation in price, we can use the observed previous value of these characteristics to construct an ordinal relationship among them. The table below shows summary statistics for the relative value of these variable types. One can interpret each number as a dollar value per square foot for the qualities that the house has. For example, the median house is a row-house (middle), which according to analysis of previous sales is assessed at $117 per square foot (Office of Tax and Revenue, 2013).

Table 7 - Qualitative Variable Description (2010 - 2012)

Qualitative Variables as Ordinal Equivalents

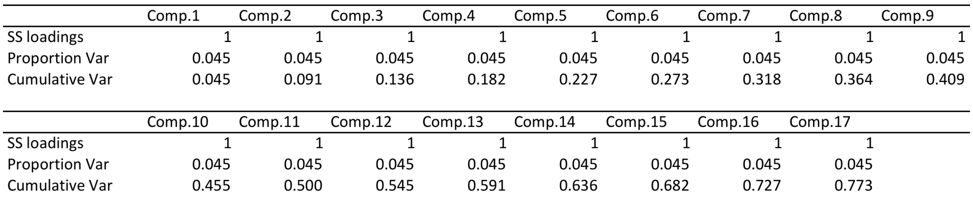
	mean	sd	median	min	max	skew	kurtosis	se
use code	123	12.31	117	90	141.5	-0.25	0.57	0.13
interior wall	6.52	1.24	7.17	0.75	8.53	-1.95	3.23	0.01
exterior wall	3.32	1.55	3.95	0	9.38	-0.76	2.3	0.02
roof type	0.41	0.81	0	-0.43	2.93	2.44	4.67	0.01

Principal Components Analysis

While the loading of the variables on the principal components did not leave us confident in our ability to interpret the results of a regression based on them, Principal Component Analysis did reveal some useful collinearities that we later use to better understand confusing signs on several regression coefficients.

The first three Principal Components explain 14% of the variance. Gross Building Area (House Size), Land Area, and Median Income are the only variables that load on these components (See PCA Table). Median income is the only variable loaded on the first component. It explains 4.5 percent of the cumulative variance. It is interesting to note that the orthogonal second and third components describe the complex relationship between land area(LA) and Gross Building Area(GBA), where GBA and LA can vary together, but also inversely.

In the DC Assessor's manual (2012), it is pointed out that Land Area has a complex relationship to Price, which they describe by the neighborhood that a house falls in. In some neighborhoods, as the assessor models price, the marginal value of each addition square foot of land is modeled by one of four possible elasticity curves. This model is borne out by the PCA.



	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
SS loadings	1	1	1	1	1	1	1	1	1
Proportion Var	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045
Cumulative Var	0.045	0.091	0.136	0.182	0.227	0.273	0.318	0.364	0.409

	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17
SS loadings	1	1	1	1	1	1	1	1
Proportion Var	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.045
Cumulative Var	0.455	0.500	0.545	0.591	0.636	0.682	0.727	0.773

Figure 10 - Proportional and Cumulative Variance of PCA of all Variables

Hedonic Regression

Our method for holding all housing characteristic constant as we examine the effect at boundaries is regression. As discussed in the literature review, regression applied in the sense that we use it is called “hedonic” in the economics literature.

In linear form, the hedonic regression is a regression of price onto characteristics. That is, for $(z_{1i},\ldots,z_{ki})$, a vector of housing characteristics, the regression is:

Equation 1 - Hedonic Regression

$$P_i = \beta_0 + \sum_{j=1}^{k} \beta_i z_{ji} + \xi_i,$$

, where β_i are coefficients and ξ_i are independent and identically distributed. In general, the literature suggests that using a log-linear form of regression is preferred for simple regression (Malpezzi, 2003). The reason that this form is preferred is that, theoretically, it allows coefficients to vary together. That is, by exponentiation of the observed housing variables, their effect is mutually informative for price, rather than simply being a linear relationship between price and an individual component.

Variable Selection for Regression Model

We evaluated several methods for modeling house prices using physical and social characteristics. A model based entirely around PCA and around mixed PC's and variables left us with less confidence in our model than the linear model based on variables. However, all three methods generally identified similar boundaries. Because the regression model based on variables is more informative to the reader, we present the results of that model in the results section and include the PCA-based regression models in the Appendix.

Principal Components Analysis was theoretically desirable given the large number of collinear variables. However, many of the variables load alone on a single component, and the interpretation of the others was unclear. In light of literature that suggests that PC's cannot be selected for regression by the size of their eigenvectors alone, and that all PC's must be considered in terms of their explanatory power for the desired variable (Jolliffe, 1982), we found that PC's did not provide us with a model about which we could be any more confident in explaining than the traditional linear model based on variables themselves.

This left us with the problem of variable selection. We used stepwise regression over 3 different pairs of years to identify a set of variables to use in the regression model. Stepwise regression uses a series of F-Tests to determine whether variables can be said to have an effect on the accuracy of the model that is significantly different from zero (Mickey, 1967). We attempt to overcome the shortcoming of stepwise regression, that models constructed by it may be less generalizable (Whittingham, 2006), by applying it over 3 sets of years and reviewing the results in the context of variables recommended in the literature review.

For each of the three pairs of years of houses sold, stepwise regression identified the same 19 variables. Only Crime was not identified as significant via an F-Test for the 2008-2009 and 2010-2011 home sales data. However, because crime was identified for the 2006-2007 data, we include it in the regression model applied to each pair of years.

A Heuristic for Examining the Spatial Variation in Residuals at Boundaries

As discussed in the literature review, the analysis of the spatial distribution of residuals has been shown to be a valid method for developing better modeling techniques for housing prices.

We developed a simple technique for inspecting the variation in residuals and their distance from the boundary. After modeling with regression, we inspect residuals for each house in each neighborhood with respect to their distance from every boundary for that neighborhood.

The first step in this analysis was identifying boundaries of interest. We chose to use only boundaries internal to the District of Columbia. This decision was driven by the initial hypothesis that house price residuals at boundaries might exhibit discontinuities. In the final analysis, this choice also removes the possibly confounding factor of boundaries of a different administrative order. However, it is also possible that future research could focus in particular at how state boundaries relate to house prices. Prices are certainly discontinuous across state boundaries, given differences in tax rates and services. However, since our focus was on the internal effect of boundaries, we focused on those boundaries within DC.

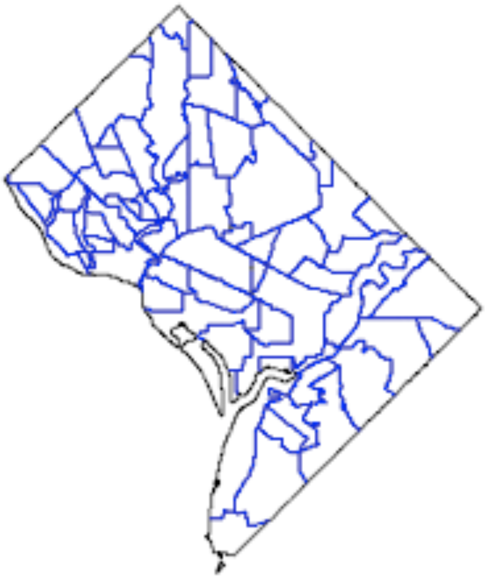


Figure 11 - Assessment Boundaries of Interest

We define a boundary as a continuous line between adjacent assessment neighborhoods. Above we plot the 235 boundaries within the District of Columbia that make up assessment neighborhoods. 40 of these boundaries were less than 100 meters and therefore removed from the analysis. That left 195 boundaries. Because each boundary has two sides, a total of 390 boundaries were analyzed with respect to the residuals for houses sold within their assessment neighborhoods.

4. Results

Introduction

In this chapter we present the results of the regression model, the relationship of residuals at neighborhood boundaries to those boundaries, and finally, maps and boxplots for residuals for houses for which we can be confident that there is a difference in residual value with respect to the boundary.

Regression Model Results

The results of the regression model are shown in the Linear Model Regression Results tables in the Appendix for all 3 pairs of years of home sales.

Of the variables that describe the physical house, all variables display the same sign and similar magnitudes for the sign, despite summary statistics that at times show very different underlying distributions for the data. For example, the distribution of Land Area in 2010-2011 is very positively skewed and with high kurtosis. However, the estimated coefficient is very similar in magnitude to 2006 and 2007 sales.

Estimated coefficients for social variables are more difficult to interpret. For example, in all 3 sets of years, median income (mdninc) is estimated to have a negative coefficient. One possible explanation for this result is that an expansion in demand resulted in some submarkets with lower median income overall to experience increases in price, and a larger volume of sales. In addition, the sign on median age flips from 2006-2007 to 2008-2009. In this case, the interpretation is not as worrisome. In the literature review, we found that median age might be seen simply as a way that people shop in the market, and while its bearing on price may not be direct, it is useful to us because it describes a dimension along which people purchase homes. Percent White (pwhite) is useful to the model in the same way. In all sets of years, educational attainment (pbchlr—Percentage of population with a Bachelors Degree) exhibits a positive sign. In this latter case, the direct relationship between the coefficient and the dependent variable may it may be more straightforwardly interpretable.

Dummy variables on Condition and Grade are generally easy to interpret and are consistent with expectations. For Use Code dummy variables (E.G. USECODE012), the base comparison is against row-houses. The interpretation here may be a bit confusing, because we might assume that a row house will always be less desirable than a detached or semi-detached house. The estimated coefficients generally do not indicate that this is the case. However, many of the row houses in Washington, D.C. are historic, and may be desirable for their historic character. Therefore, the implicit positive sign on row houses may be understood in light of the explicit negative sign on Actual Year Built (ayb) which we find in the regression for all years. Indeed, in the PCA, we see complex interaction between these variables represented by the loadings on components 6 and 7. Thankfully, we are not interested in the most accurate estimate for the coefficient on these variables, but only on their overall accuracy in predicting price. As explained in the literature review, any missing or latent correlation at play here should not hinder our overall goal of accuracy in price.

Residuals and Assessment Neighborhood Boundaries

Our method for relating residuals to boundaries was to use distance from a given boundary as a dimension along which residuals might vary.

By way of example, in the figure below, we plot a histogram for residuals for all houses within an individual assessor's neighborhood, binned by distance from a given boundary. Further examples can be found in the Annex.

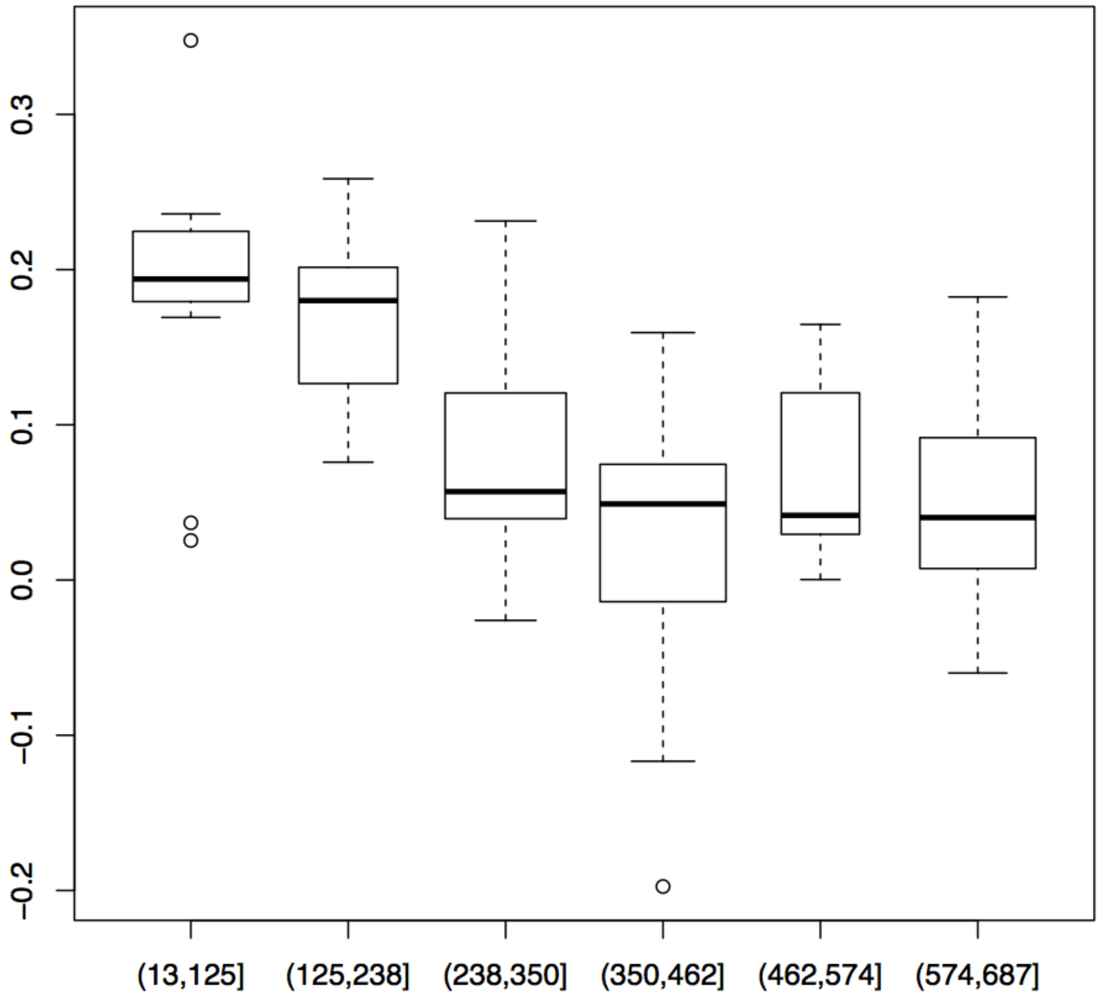


Figure 12 - Example Boxplot of Residuals (Y Axis) Binned by Distance from Example Boundary (X Axis)

In this case, we can see that the residuals for houses in the first 2 distance bins have actual sales prices that are consistently higher than their predicted price,. In the following, we plot the sale point of each house, the assessor's neighborhood that it falls in, and highlight the boundary of interest for this particular box plot.

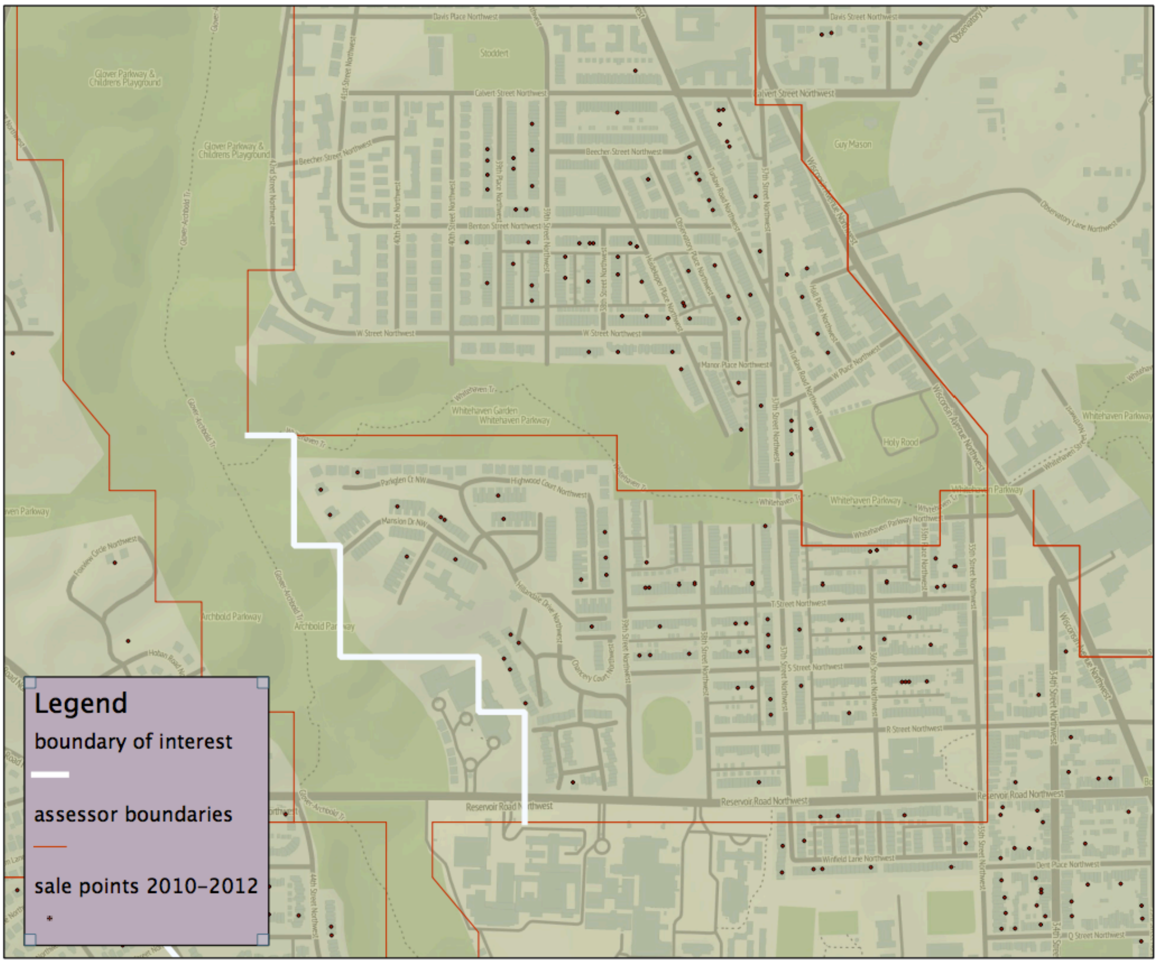


Figure 13 – Example Selected Neighborhood, Boundary, and House Sales Locations (base data c/o OSM)

This boundary adjoins a natural park, like many of the other boundaries at which nearby houses displayed a significantly different mean residual than the others houses within the same assessment neighborhood.

While boxplots provide a method for inspecting boundaries and residuals individually, we needed a method that would allow us to describe variation in the residuals with respect to boundaries for the city as a whole. After a review of hundreds of boxplot residuals, we settled on defining "nearby" houses as those within the first 15% of the furthest distance from the boundary. Then we compared the mean residual value of houses "nearby" to a boundary with the mean residual for the rest of the houses within the assessment neighborhood. Below we plot the boundaries at which this t-test resulted in a p value of less than 0.05, indicating that we could be 95% confident that the mean residual value for houses near the boundary was significantly different than for houses not near the boundary, but within the same assessment neighborhood.

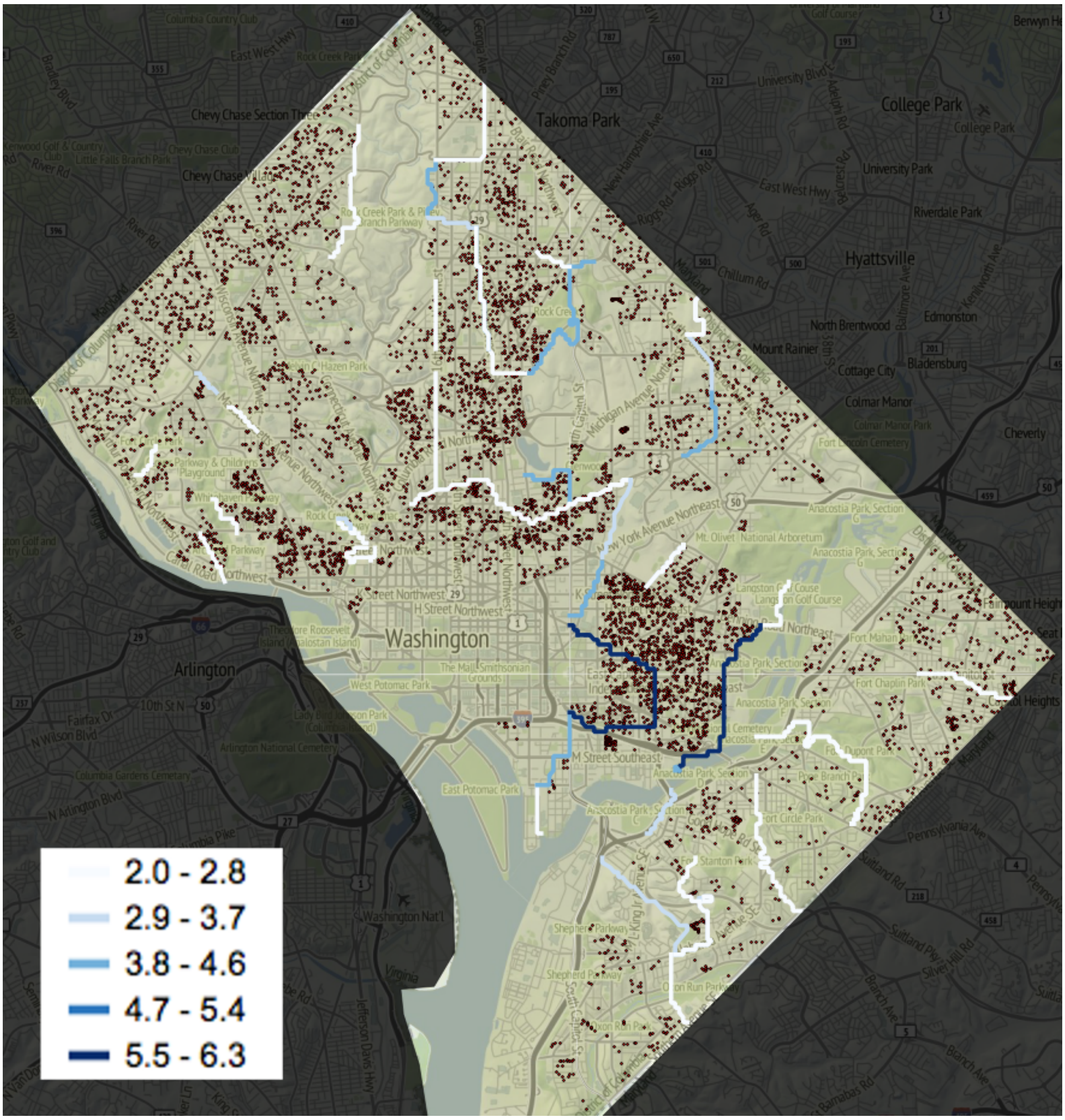


Figure 14 - Map of Boundaries at which Residuals are Significantly Different

(absolute value of t-statistic in legend)

That many of these boundaries outline natural features is a testament to the validity both of the regression model and the heuristic method for identifying boundaries of interest. Furthermore, not all of these boundary effects could be modeled, for example, by distance of the house from a natural feature. In some cases, the boundary was along a major road, such as Florida Ave, NW. Below we show an example of the context of Florida Avenue and then the boxplots for residuals nearby. In this case, the boxplot is a plot of residuals as they move from north to south.

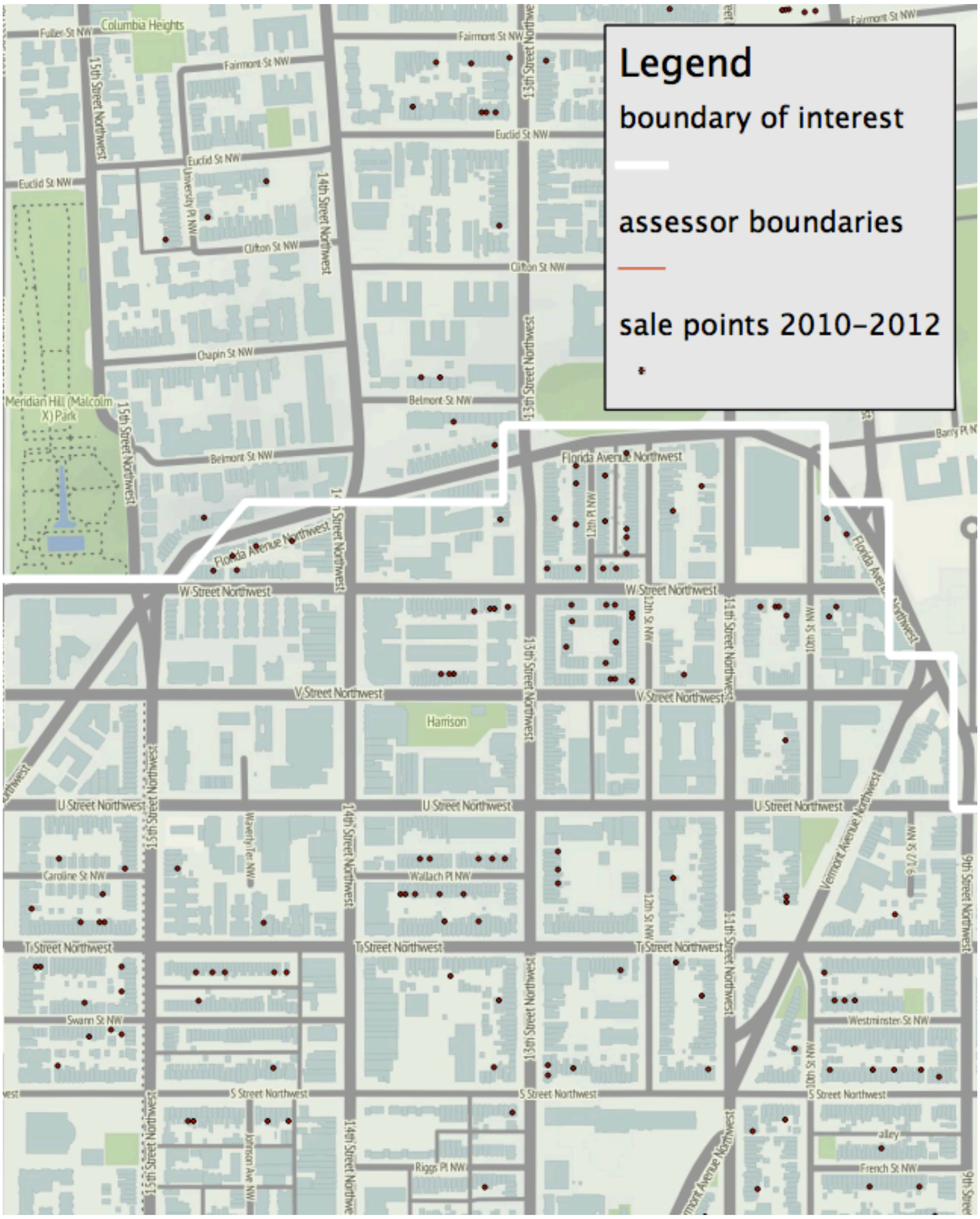


Figure 15 - A Significant Boundary Adjoining a Major Road

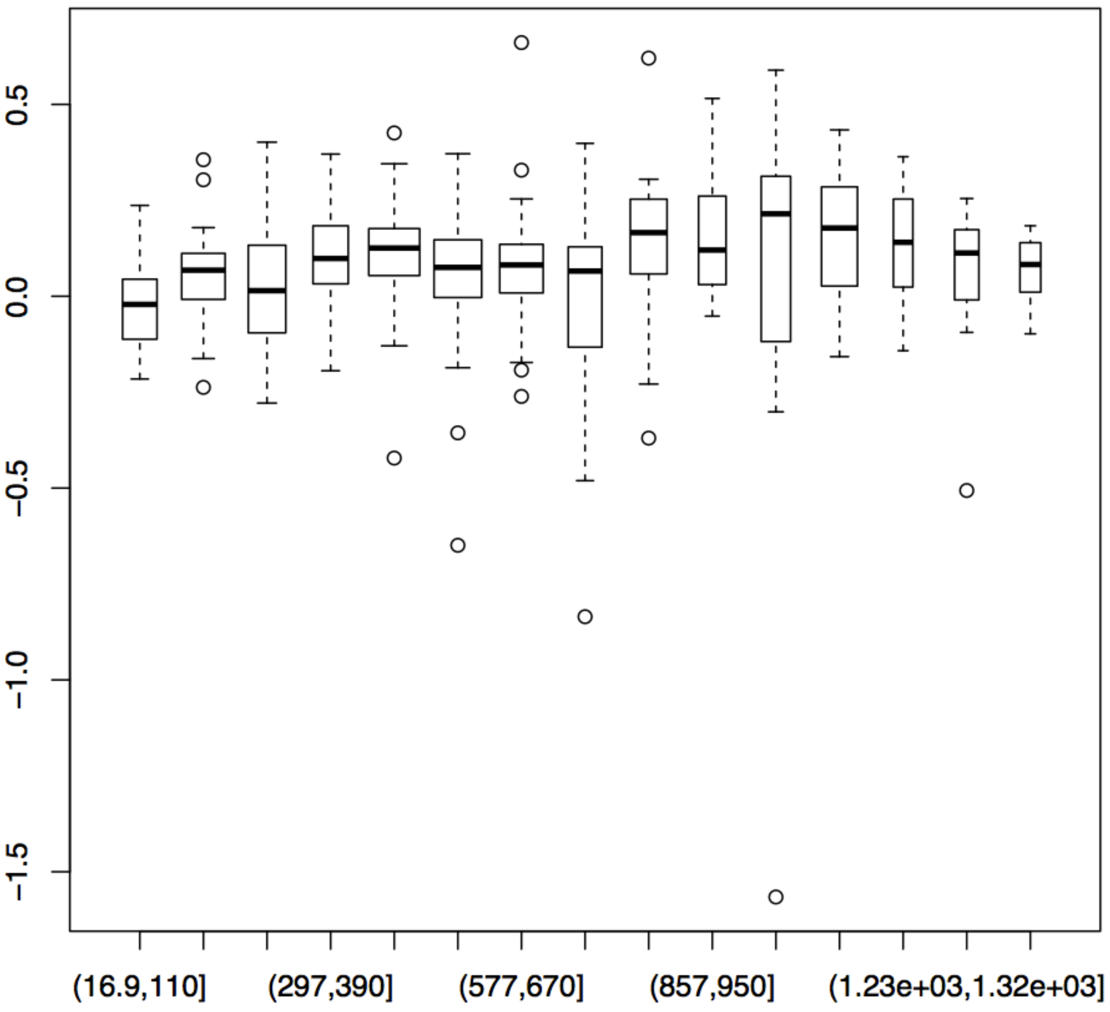


Figure 16 - Boxplot of Residuals (Y Axis) going south from Florida Avenue Binned by Distance from Edge (X Axis)

Its important to note that in this boxplot the y axis for residuals has a much wider range than in Figure 12. Here the trend for houses close to the assessment boundary along Florida Avenue is to sell for less closer to the boundary, as opposed to more, in the case of the boundary along the park in the previous assessment district. Other Avenues that make up significant boundaries include Georgia Avenue and 16th Streets Northwest. Both of these roads have a functional definition as a "Principal Arterial" by the District Department of Transportation. This functional definition means that these streets are deisgned to support a maximum speed of 40 mph, which is second only to the designed freeway speed of 55 mph (Department of Transportation, 2013). While further research would be required, we could hypothesize that the real estate market reacts to these functional definitions for roads. Furthermore, these roads become clear boundaries along which boundaries for neighborhoods are drawn. However, if the boundary itself is something that the real estate market responds to, it seems that there might be a latent variable, based on a physical feature of the environment, in the explicit geographic definition of the areal unit.

5. Conclusion

To review, this thesis was motivated by the main hypothesis: that the effect of neighborhood boundaries on their internal phenomena is heterogeneous, in particular for housing sales. That is, our hypothesis implied that the boundaries used to model and describe spatial dependence and heterogeneity are heterogeneous in their relationship to the housing market they are used to model.

We found generally that the hypothesis was valid. In particular, we found examples of where boundaries to areal units had a strong relationship with house prices, holding all other factors constant. Many of them were along obvious features of the environment: major arterial roads, railways, and natural parks. While these are perhaps obviously important features in making housing decisions, what this analysis reveals is that these physical features in particular, as boundaries for areal units, are heterogeneous features, and not simply the arbitrary geometric components of areal units. One implication for this result is that we might expect that areal units fall short as models of spatial processes in which there are hard physical Edges in the urban environment.

Further research could attempt to investigate how boundaries for other phenomena relate to the processes that they circumscribe. It seems intuitively clear that more concrete processes would be easier to understand. For example: how does the spatial pattern of crime relate to police districts? However, processes based on more abstract concepts, such as politics, could present interesting relationships between boundaries and individual choices. For example, how are political voting districts in the U.S. shaped by demographic changes?

Bibliography

Amtrak History and Archives. (2013). Retrieved 12 10, 2013, from Amtrak.com: http://history.amtrak.com/archives

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics* , 247-267.

Bell, K. P. (2000). Applying the generalized-moments estimation approach to spatial problems involving micro-level data.". *Review of Economics and Statistics , 82* (1), 72-82.

Beshears, J. (2008). How are preferences revealed? *Journal of Public Economics* , 1787-1794.

Bourassa, S. C. (2003). Do housing submarkets really matter? *Journal of Housing Economics* , 12-28.

Brunsdon, C. A., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis* , 281-298.

Case, Bradford, & al., e. (2004). Modeling spatial and temporal house price patterns: a comparison of four models. *The Journal of Real Estate Finance and Economics* , 167-191.

Census, U. (1994). *Geographic Areas Reference Manual.* Washington: US Census Bureau.

Department of Transportation, D. (2013). *Design and Engineering Manual.* Washington: D.C. Government.

Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional science and urban economics* , 433-452.

Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics* , 181-201.

Lanius, J. H., & Park, S. C. (1995). *Martha Wadsworth's Mansion: The Gilded Age Comes to Dupont Circle.* Washington, DC: Historical Society of Washington.

Malpezzi, S. (2003, 01 01). Hedonic pricing models: a selective and applied review. *Section in Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan* , pp. 01-60.

Maue, P. (2013). *Places in the Long Tail.* Gloucester, UK: Ios Pr Inc.

Office of Tax and Revenue, D. (2013). *Assessment Manual.* DC: DC Government.

Rybczynski, W. (2010). *Makeshift metropolis: Ideas about cities .* New York: SimonandSchuster.

Sheppard, S. (1999). In S. Sheppard, *Handbook of regional and urban economics 3* (pp. 1595-1635). : APA .

Tiebout, C. M. (1956). A pure theory of local expenditures. *The journal of political economy* , 416-424.

Wachter, S. M., & Wong, G. (2008). What Is a Tree Worth? Green-City Strategies, Signaling and Housing Prices. *Real Estate Economics* , 213-239.

Annex

Table 8 - 2006-2007 Data Summary

2006-2007 Data Summary (n: 4657)

	var	mean	sd	median	trimmed	mad	min	max	skew	kurtosis	se
BATHRM	2	2.76	1.24	3	2.69	1.48	0	12	0.87	1.77	0.02
BEDRM	4	3.32	1.09	3	3.19	1.48	0	10	1.46	3.82	0.02
ROOMS	9	7.29	2.24	7	7.02	1.48	0	26	1.52	4.94	0.03
INTWALLP	10	6.65	1.19	7.17	6.95	0	0.75	8.53	-2.5	5.8	0.02
KITCHENS	13	1.26	0.62	1	1.12	0	0	6	2.93	9.5	0.01
FIREPLACES	14	0.7	0.98	0	0.53	0	0	11	2.35	10.89	0.01
GBA	1	2,973.78	1,521.52	2,711.00	2,787.67	1,036.34	407.00	23,120.00	2.77	17.06	22.30
LANDAREA	8	2,837.96	2,594.33	2,024.00	2,364.12	1,058.58	294.00	64,205.00	5.62	80.60	38.02
EYB	11	1,964.64	16.03	1,961.00	1,962.31	10.38	1,927.00	2,011.00	1.31	1.47	0.23
AYB	12	1,933.03	29.14	1,930.00	1,930.32	25.20	1,754.00	2,011.00	0.53	1.74	0.43
pbchlr	3	50.68	27.95	48.2	50.82	39.59	1.6	94.4	0.05	-1.45	0.41
pwhite	5	0.41	0.36	0.31	0.4	0.44	0	1	0.25	-1.57	0.01
mdninc	6	78,918.12	43,091.49	72,340.00	74,584.60	43,499.48	13,672.00	219,583.00	0.91	0.57	631.45
mdnage	7	36.87	6.73	36	37.02	5.49	17.7	52.7	0.11	-0.1	0.1
crimepp	15	0.11	0.08	0.1	0.1	0.05	0	1.48	3.25	25.22	0

Grade Dummy	
Average	1561
Above Avg	1365
Good	982
Very Good	419
Excellent	159
Superior	121
(Other)	50

Condition Dummy	
Average	1941
Default	0
Excellent	39
Fair	46
Good	2251
Poor	6
Very Gd	374

Use Code Dummy	
11	2306
12	922
13	711
24	540
23	178
1	0
(Other)	0

Table 9 - 2008-2009 Data Summary

2008-2009 Data Summary (n: 3639)

	var	mean	sd	median	trimmed	mad	min	max	skew	kurtosis	se
BATHRM	2	2.98	1.24	3	2.93	1.48	0	10	0.74	1.5	0.02
BEDRM	4	3.39	1.1	3	3.28	1.48	0	16	1.69	7.7	0.02
ROOMS	9	7.36	2.22	7	7.11	1.48	0	25	1.6	5.34	0.04
INTWALLP	10	6.6	1.12	7.17	6.84	0	0.75	8.53	-2.01	3.64	0.02
KITCHENS	13	1.22	0.55	1	1.09	0	0	4	3.06	10.67	0.01
FIREPLACES	14	0.79	0.97	1	0.64	1.48	0	10	1.91	6.7	0.02
GBA	1	3092	1558	2795	2905	1075	252	16842	2	11	26
LANDAREA	8	2896	2572	2000	2420	1186	397	35124	3	20	43
EYB	11	1967	17	1963	1965	9	1900	2011	1	1	0
AYB	12	1933	32	1927	1929	25	1776	2011	1	1	1
pbchlr	3	57.22	26.6	62.1	58.59	34.54	1.6	94.4	-0.28	-1.32	0.44
pwhite	5	0.49	0.35	0.5	0.5	0.52	0	1	-0.1	-1.57	0.01
mdninc	6	87,813	42,709	85,484	84,642	47,049	13,672	219,583	1	0	708
mdnage	7	36.83	6.46	36	36.91	5.49	17.7	52.7	-0.1	0.02	0.11
crimepp	15	0.11	0.09	0.1	0.1	0.06	0	1.48	3.21	21.9	0

Grade Dummy	
Above Average	1081
Good Quality	924
Average	909
Very Good	438
Superior	109
Excellent	108
(Other)	70

Condition Dummy	
Average	1252
Default	0
Excellent	93
Fair	20
Good	1908
Poor	4
Very Good	362

Use Code Dummy	
11	1873
12	822
13	463
24	376
23	105
1	0
(Other)	0

Table 10 - 2010-2012 Data Summary

(n:4969)

	var	mean	sd	median	trimmed	mad	min	max	skew	kurtosis	se
BATHRM	2	3.07	1.21	3	3.01	1.48	0	12	0.84	2.85	0.02
BEDRM	4	3.45	1.11	3	3.35	1.48	0	16	1.67	7.3	0.02
ROOMS	9	7.34	2.16	7	7.1	1.48	0	20	1.57	4.66	0.03
INTWALLP	10	6.59	1.14	7.17	6.82	0	0.75	8.53	-1.96	3.37	0.02
KITCHENS	13	1.23	0.57	1	1.1	0	0	4	3.06	10.51	0.01
FIREPLACES	14	0.77	1	1	0.61	1.48	0	13	2.19	9.91	0.01
GBA	1	3139	1598	2868	2947	1085	360	23401	3	17	23
LANDAREA	8	2843	3523	1877	2292	1023	266	155905	19	732	50
EYB	11	1968	16	1963	1966	9	1937	2012	1	1	0
pbchlr	3	56.15	26.45	56	57.24	38.55	3.1	94.4	-0.19	-1.33	0.38
pwhite	5	0.48	0.35	0.47	0.48	0.53	0	1	-0.01	-1.57	0
mdninc	6	86,698	42,868	81,326	83,142	42,917	16,000	219,583	1	0	608
mdnage	7	36.34	6.27	35.7	36.42	5.49	17.7	52.7	-0.1	0.12	0.09
crimepp	15	0.11	0.09	0.1	0.1	0.06	0.01	1.48	3.18	20.06	0

Grade Dummy	
Above Average	1465
Average	1391
Good Quality	1090
Very Good	533
Excellent	236
Superior	176
(Other)	78

Condition Dummy	
Average	1427
Default	0
Excellent	134
Fair	60
Good	2744
Poor	11
Very Good	593

Use Code Dummy	
11	2724
12	1059
13	538
24	508
23	140
1	0
(Other)	0

Table 11 - Estimated Coefficients for House Sales 2006-2007

	Estimate	Std. Error	t value	Pr(>|t|)	
GBA	8.19E-05	3.93E-06	20.854	< 2e-16	***
BATHRM	4.75E-02	3.90E-03	12.176	< 2e-16	***
AYB	-2.11E-03	1.76E-04	-11.986	< 2e-16	***
FIREPLACES	3.26E-02	3.89E-03	8.384	< 2e-16	***
BEDRM	1.96E-02	3.96E-03	4.936	8.26E-07	***
INTWALLP	1.78E-02	2.49E-03	7.181	8.01E-13	***
ROOMS	9.33E-03	2.04E-03	4.583	4.70E-06	***
LANDAREA	8.25E-06	1.82E-06	4.529	6.07E-06	***
EYB	1.69E-03	3.43E-04	4.915	9.18E-07	***
KITCHENS	2.43E-02	8.84E-03	2.75	5.98E-03	**
pbchlr	6.68E-03	3.29E-04	20.308	< 2e-16	***
pwhite	3.85E-01	2.15E-02	17.888	< 2e-16	***
mdninc	-7.56E-07	1.37E-07	-5.536	3.27E-08	***
crimepp	-1.23E-01	3.81E-02	-3.221	1.29E-03	**
mdnage	1.07E-03	4.68E-04	2.289	0.02214	*
USECODE012	-1.44E-02	1.13E-02	-1.275	0.20223	
USECODE013	-5.96E-02	8.76E-03	-6.801	1.18E-11	***
USECODE023	-1.18E-01	2.44E-02	-4.816	1.51E-06	***
USECODE024	-2.86E-02	1.30E-02	-2.199	2.79E-02	*
GRADE: Average	-5.08E-02	8.03E-03	-6.326	2.76E-10	***
GRADE: Excellent	1.81E-01	1.88E-02	9.641	< 2e-16	***
GRADE: Exceptional-A	3.97E-01	3.88E-02	10.237	< 2e-16	***
GRADE: Exceptional-B	5.32E-01	5.90E-02	9.018	< 2e-16	***
GRADE: Exceptional-C	3.88E-01	9.02E-02	4.303	1.72E-05	***
GRADE: Exceptional-D	6.96E-01	1.43E-01	4.867	1.17E-06	***
GRADE: Good Quality	1.43E-02	9.38E-03	1.528	1.27E-01	
GRADE: Low Quality	-1.09E-01	2.10E-01	-0.519	0.60382	
GRADE: Superior	2.78E-01	2.13E-02	13.056	< 2e-16	***
GRADE: Very Good	8.94E-02	1.32E-02	6.795	1.22E-11	***
CONDITION: Excellent	1.70E-01	3.27E-02	5.194	2.15E-07	***
CONDITION: Fair	-1.29E-01	2.86E-02	-4.492	7.23E-06	***
CONDITION: Good	9.08E-02	6.58E-03	13.799	< 2e-16	***
CONDITION: Poor	3.10E-02	8.52E-02	0.364	7.16E-01	
CONDITION: Very Good	1.90E-01	1.30E-02	14.692	< 2e-16	***
Signif. codes	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Residual standard error	0.1897 on 4621 degrees of freedom
Multiple R-squared	0.8938
F-statistic	1111 on 35 and 4621 DF, p-value: < 2.2e-16

Table 12 - Estimated Coefficients - 2008-2009 Housing Sales

	Estimate	Std. Error	t value	Pr(>|t|)	
GBA	6.73E-05	3.75E-06	17.944	< 2e-16	***
BATHRM	5.87E-02	3.84E-03	15.29	< 2e-16	***
AYB	-1.66E-03	1.77E-04	-9.391	< 2e-16	***
FIREPLACES	4.00E-02	3.93E-03	10.174	< 2e-16	***
BEDRM	1.95E-02	3.88E-03	5.038	4.94E-07	***
INTWALLP	1.28E-02	2.70E-03	4.734	2.29E-06	***
ROOMS	5.33E-03	1.93E-03	2.756	0.00588	**
LANDAREA	1.42E-05	2.01E-06	7.088	1.63E-12	***
EYB	2.01E-03	3.70E-04	5.428	6.08E-08	***
KITCHENS	1.96E-02	9.36E-03	2.096	0.03615	*
pbchlr	6.82E-03	3.31E-04	20.638	< 2e-16	***
pwhite	4.43E-01	2.17E-02	20.465	< 2e-16	***
mdninc	-6.87E-07	1.29E-07	-5.329	1.05E-07	***
crimepp	-4.00E-02	3.45E-02	-1.159	0.24661	
mdnage	-1.33E-03	4.93E-04	-2.708	0.0068	**
USECODE012	-2.48E-02	1.14E-02	-2.169	3.02E-02	*
USECODE013	-3.04E-02	9.43E-03	-3.224	0.00128	**
USECODE023	-5.55E-02	2.62E-02	-2.12	3.41E-02	*
USECODE024	-3.99E-02	1.34E-02	-2.982	0.00289	**
GRADE: Average	-5.32E-02	8.71E-03	-6.108	1.11E-09	***
GRADE: Excellent	1.77E-01	1.95E-02	9.089	< 2e-16	***
GRADE: Exceptional-A	4.03E-01	3.12E-02	12.906	< 2e-16	***
GRADE: Exceptional-B	3.56E-01	4.10E-02	8.686	< 2e-16	***
GRADE: Exceptional-C	8.22E-01	1.23E-01	6.683	2.70E-11	***
GRADE: Exceptional-D	8.58E-01	1.25E-01	6.84	9.29E-12	***
GRADE: Good Quality	4.36E-02	8.83E-03	4.94	8.17E-07	***
GRADE: Low Quality	-9.78E-01	1.98E-01	-4.948	7.85E-07	***
GRADE: Superior	3.04E-01	2.03E-02	14.987	< 2e-16	***
GRADE: Very Good	1.06E-01	1.21E-02	8.769	< 2e-16	***
CONDITION: Excellent	1.79E-01	2.15E-02	8.314	< 2e-16	***
CONDITION: Fair	-1.59E-01	3.87E-02	-4.098	4.26E-05	***
CONDITION: Good	9.80E-02	6.93E-03	14.141	< 2e-16	***
CONDITION: Poor	-1.12E-01	9.91E-02	-1.132	2.58E-01	
CONDITION: Very Good	2.16E-01	1.19E-02	18.118	< 2e-16	***
Signif. codes	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Residual standard error	0.1704 on 3603 degrees of freedom
Multiple R-squared	0.9119
F-statistic	1065 on 35 and 3603 DF, p-value: < 2.2e-16

Table 13 - Estimated Coefficients 2010-2012 Housing Sales

2010-2011 Linear Model Regression Results

	Estimate	Std. Error	t value	Pr(>|t|)	
GBA	8.32E-05	3.52E-06	23.669	< 2e-16	***
BATHRM	6.56E-02	3.68E-03	17.817	< 2e-16	***
AYB	-3.08E-03	1.64E-04	-18.747	< 2e-16	***
FIREPLACES	2.48E-02	3.64E-03	6.823	1.00E-11	***
BEDRM	1.99E-02	3.76E-03	5.284	1.32E-07	***
INTWALLP	1.33E-02	2.56E-03	5.19	2.19E-07	***
ROOMS	2.54E-03	1.98E-03	1.286	0.198632	
LANDAREA	2.87E-06	1.12E-06	2.56	0.010493	*
EYB	3.49E-03	3.44E-04	10.139	< 2e-16	***
KITCHENS	4.38E-03	8.40E-03	0.521	0.60235	
pbchlr	8.64E-03	3.13E-04	27.641	< 2e-16	***
pwhite	3.92E-01	2.01E-02	19.498	< 2e-16	***
mdninc	-4.34E-07	1.26E-07	-3.453	0.00056	***
crimepp	1.44E-02	3.38E-02	0.427	6.70E-01	
mdnage	-8.50E-04	4.92E-04	-1.728	0.084009	.
USECODE012	7.85E-03	1.02E-02	0.768	0.442771	
USECODE013	-8.15E-02	9.51E-03	-8.569	< 2e-16	***
USECODE023	-8.26E-02	2.47E-02	-3.351	0.000811	***
USECODE024	-3.44E-02	1.25E-02	-2.739	0.00619	**
GRADE: Average	-9.43E-02	7.95E-03	-11.856	< 2e-16	***
GRADE: Excellent	1.34E-01	1.65E-02	8.165	4.03E-16	***
GRADE: Exceptional-A	3.09E-01	3.13E-02	9.856	< 2e-16	***
GRADE: Exceptional-B	4.62E-01	4.93E-02	9.355	< 2e-16	***
GRADE: Exceptional-C	3.44E-01	7.39E-02	4.647	3.45E-06	***
GRADE: Exceptional-D	3.73E-01	1.00E-01	3.719	0.000202	***
GRADE: Good Quality	8.64E-03	8.73E-03	0.989	0.322609	
GRADE: Low Quality	NA	NA	NA	NA	NA
GRADE: Superior	2.09E-01	1.80E-02	11.625	< 2e-16	***
GRADE: Very Good	4.45E-02	1.20E-02	3.701	0.000217	***
CONDITION: Excellent	2.98E-01	2.19E-02	13.589	< 2e-16	***
CONDITION: Fair	-9.14E-02	2.56E-02	-3.572	3.57E-04	***
CONDITION: Good	1.13E-01	6.82E-03	16.584	< 2e-16	***
CONDITION: Poor	-1.91E-01	5.82E-02	-3.283	1.04E-03	**
CONDITION: Very Good	2.28E-01	1.09E-02	20.823	< 2e-16	***

Signif. codes 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error	0.1905 on 4934 degrees of freedom
Multiple R-squared	0.9047
F-statistic	1378 on 34 and 4934 DF, p-value: < 2.2e-16

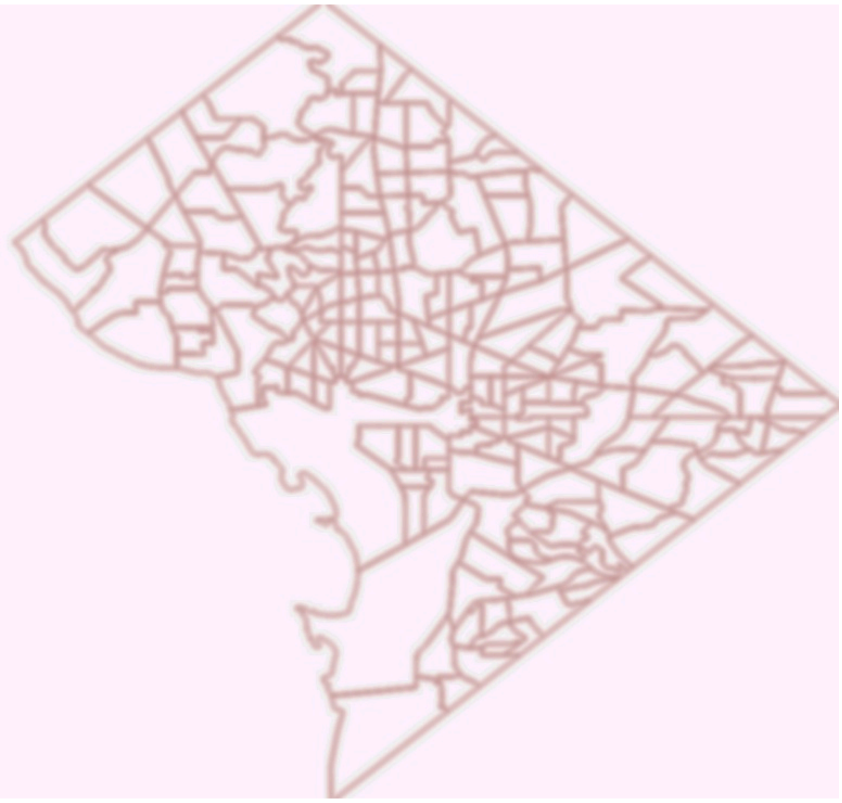


Figure 17 - **Washington, DC Census Tracts**

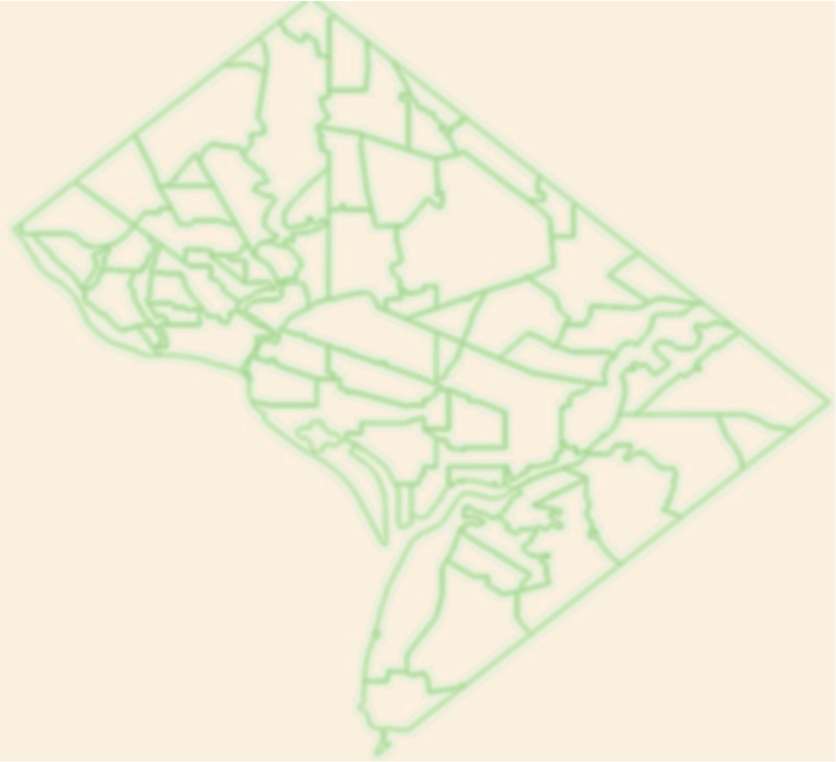


Figure 18 - **DC Assessment Neighborhoods**

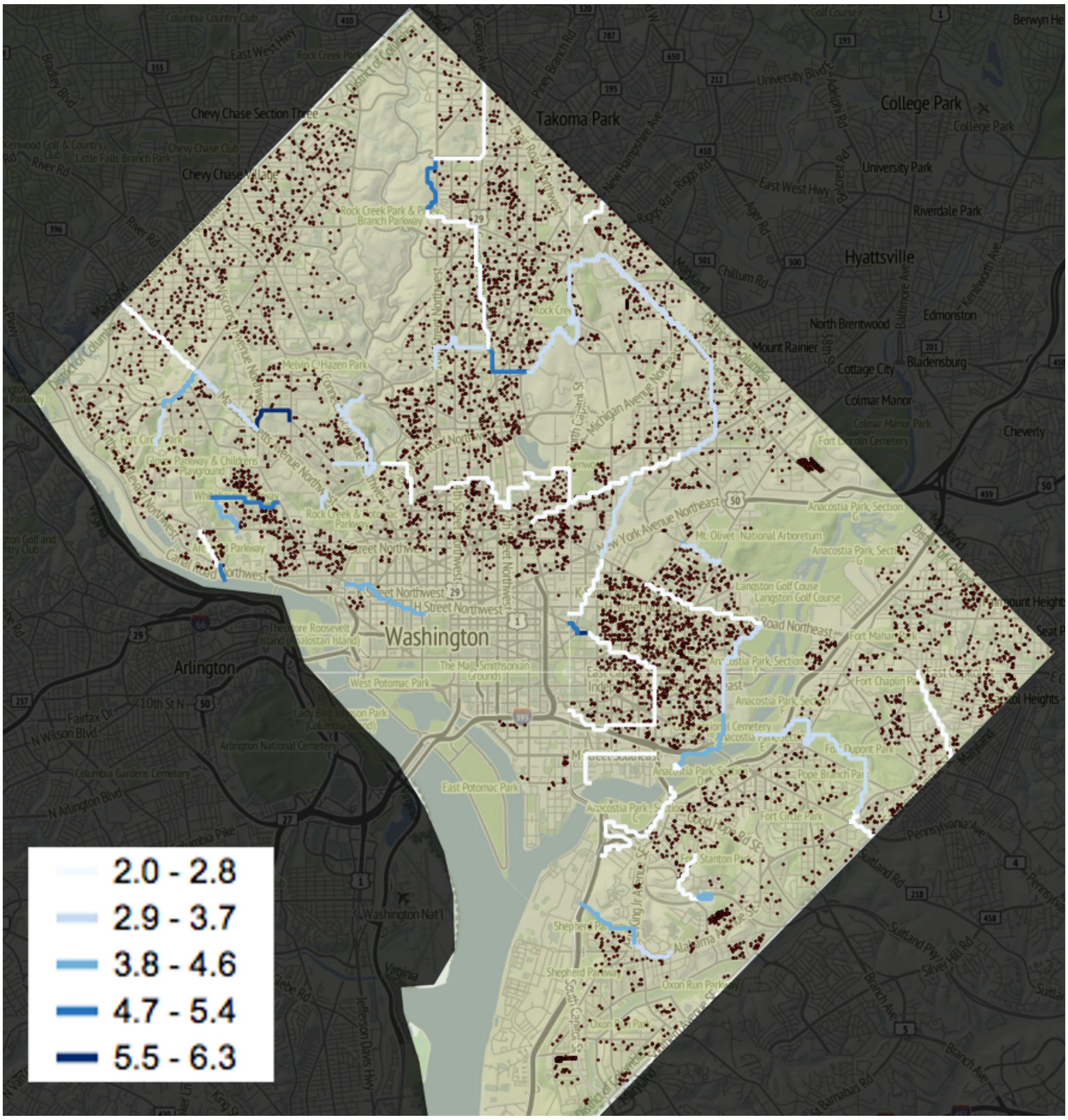


Figure 19 - Boundaries with Significantly Different Adjacent Home Sale Prices (2006-2007)

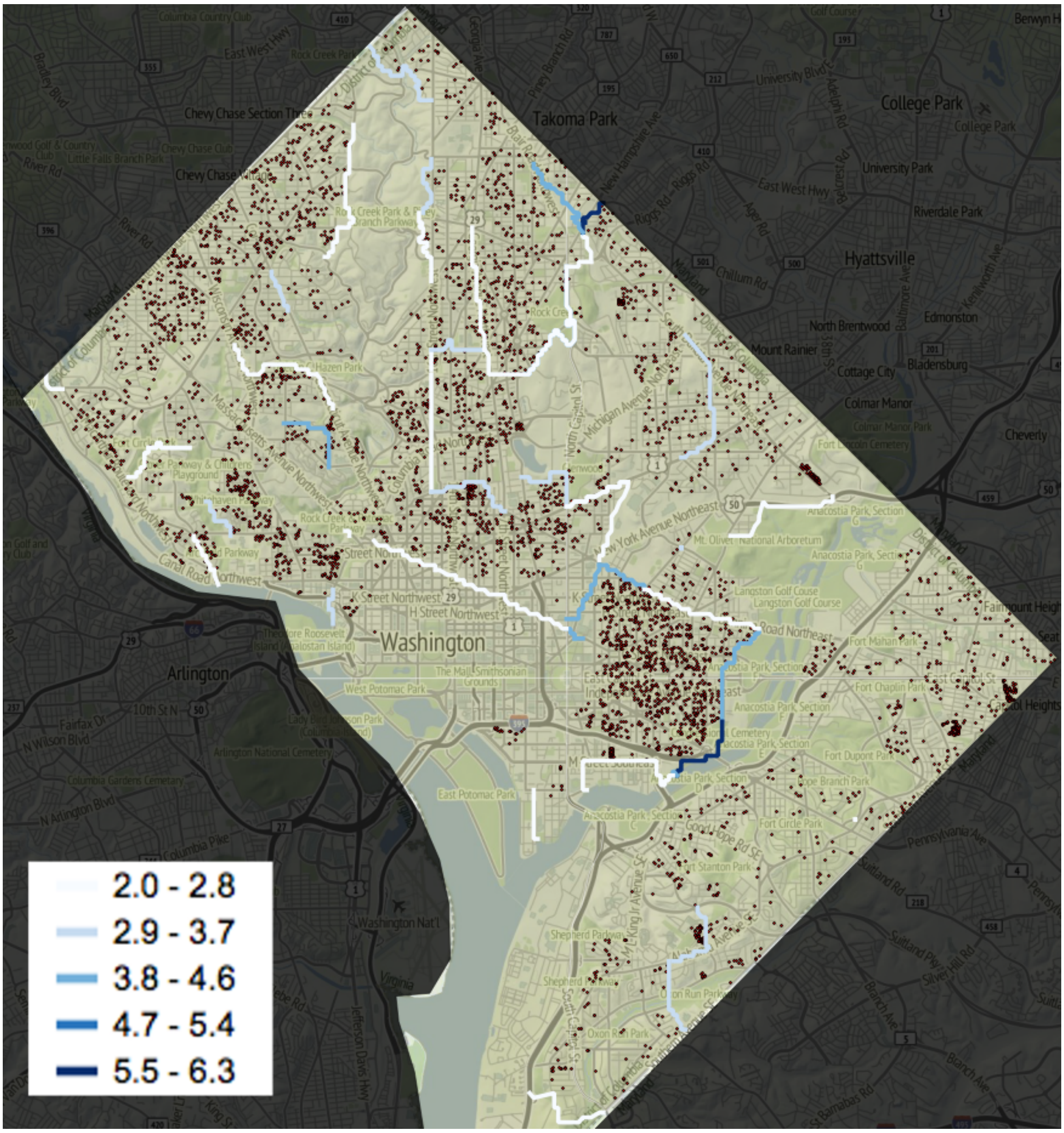
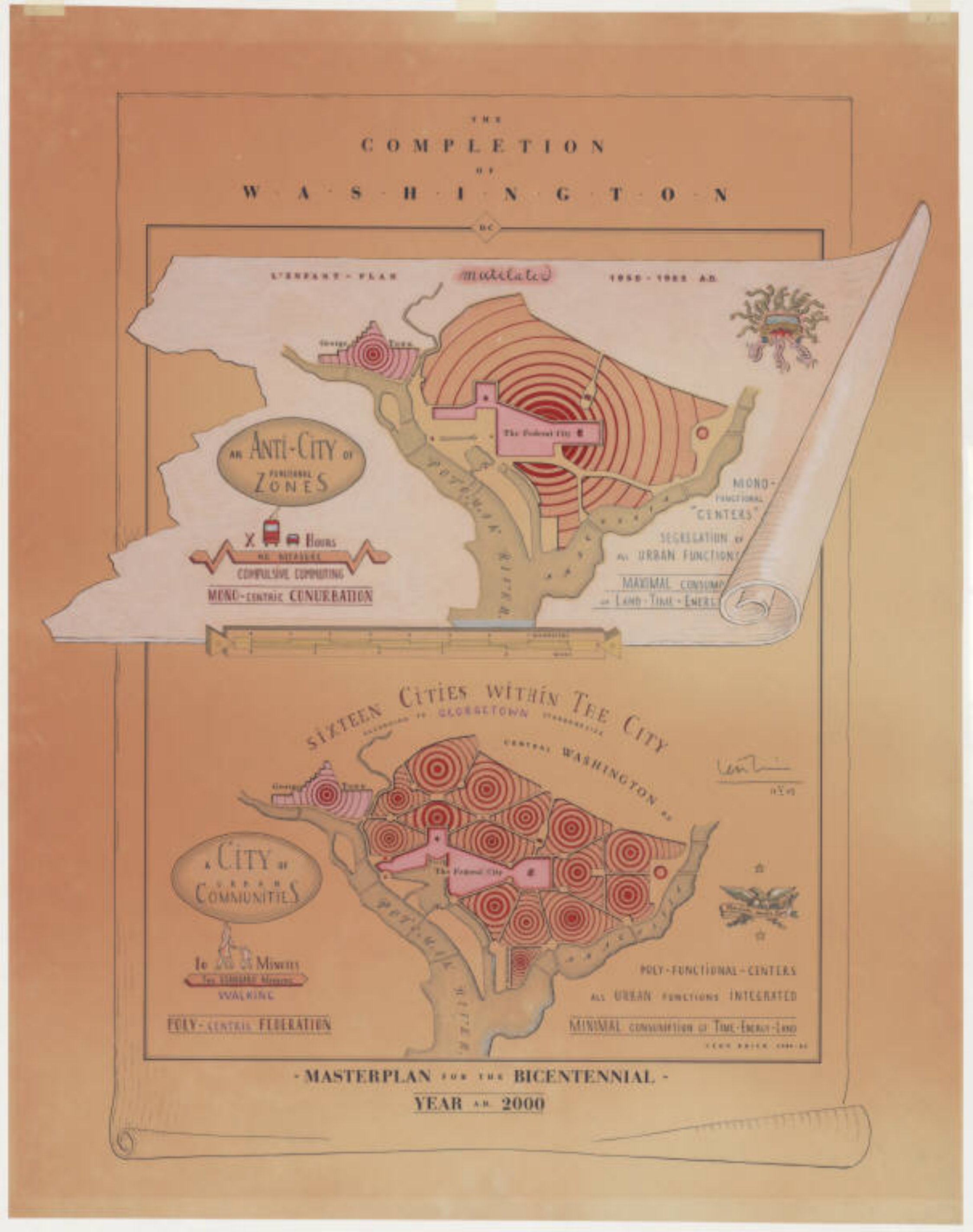


Figure 20 - Boundaries with Significantly Different Adjacent Home Sale Prices (2008-2009)



THE COMPLETION OF WASHINGTON
DC
L'ENFANT - PLAN mutilated 1950 - 1982 AD
George Town
The Federal City
POTOMAK RIVER
AN ANTI-CITY OF FUNCTIONAL ZONES
X Hours
COMPULSIVE COMMUTING
MONO-CENTRIC CONURBATION
MONO-FUNCTIONAL "CENTERS"
SEGREGATION OF ALL URBAN FUNCTIONS
MAXIMAL CONSUMP... OF LAND-TIME-ENERG...
SIXTEEN CITIES WITHIN THE CITY
GEORGETOWN
CENTRAL WASHINGTON
A CITY OF URBAN COMMUNITIES
10 Minutes
WALKING
POLY-CENTRIC FEDERATION
POLY-FUNCTIONAL-CENTERS
ALL URBAN FUNCTIONS INTEGRATED
MINIMAL CONSUMPTION OF TIME-ENERGY-LAND
- MASTERPLAN FOR THE BICENTENNIAL -
YEAR AD 2000

Figure 21 - Sixteen Cities Within the City (Krier, 2000)