

Disparities in Access to Amenities that Promote Health and Well-Being in Washington, DC Neighborhoods: A Preliminary Exploration

Peter A. Soyka

Introduction/Business Problem

Washington, DC is a racially and economically diverse city and the capitol of the United States. Historically, the population of Washington has been majority African-American, but the demographics of the city are changing rapidly. The local economy has grown markedly in recent years, propelled by large increases in U.S. federal government activity. The population of the city has expanded significantly during the past two decades, and a large percentage of new immigrants are younger, more upwardly mobile, and more racially diverse than many longer-term residents.

Despite these substantial changes, there remain major differences in the racial and socioeconomic composition of different areas of the city. For example, neighborhoods in the western end of the city generally house people who are Caucasian, have college degrees, are working professionals, and earn higher than average salaries. In contrast, neighborhoods on the eastern side, particularly east of the Anacostia River, typically contain residents of color who are less extensively educated and have lower incomes. For many years, residents of these latter areas (Wards 7 and 8) have expressed concerns that they receive fewer city services, have higher crime rates, and have access to fewer amenities (both public and private) than in wealthier areas of the District of Columbia. The lack of grocery stores selling fresh foods, drug stores, and health care facilities has been of particular concern. Studies have shown that this lack of or limited access to high quality food (i.e., “food deserts”) and other amenities that promote health and well-being is not uncommon in less affluent areas of the United States.

As a long-standing resident of the greater Washington, DC area and a person with substantial professional interest in human health and environmental quality, I wish to determine whether and to what extent there are significant spatial disparities in access to goods and services that influence the health and well-being of area citizens. Ensuring that all citizens are not only willing, but able, to pursue a healthy lifestyle is a key component of making our society more resilient, fair, and sustainable. Doing so requires, as an initial step, developing an understanding of where access to key goods and services by all residents of a particular community is adequate and where it is not.

The purpose of this study is to determine whether there are in fact substantial differences in convenient access to providers of goods and services that tend to promote good health and well-being across defined neighborhoods across Washington, DC as well as in the presence of entities that may have detrimental effects on residents’ health. To address these issues, I will examine location information on both public and private sector entities that can reasonably be considered to have either positive or negative impacts on the health and well-being of neighborhood residents, and then compare and contrast the patterns observed in wealthier (and whiter) areas of the city with those found in poorer (and less white) neighborhoods. Examples of entities that would be expected to promote good health and/or prosperity include supermarkets, parks and athletic fields, doctor’s offices, houses of worship,

and banks. Examples of entities that might limit or detract from the health and well-being of residents include bars, waste management facilities, and check cashing services.

Data

To perform the analysis described above, I will make use of the Foursquare location database supplemented with neighborhood-level location information obtained from the on-line real estate service Zillow. Foursquare contains not only precise location data (street address and latitude/longitude) but also descriptions of each entity's basic function/purpose, which is defined as one of several hundred "Categories" or sub-categories. Table 1 below provides examples of the categories that I will examine for this project. The presence of several types of entities within a neighborhood is likely to promote better health, convenience, satisfaction, and general well-being. At the same time, the presence of other types of entities within a neighborhood may work to the detriment of local residents, due to possible exposure to pollutants, unsafe conditions, excessive noise, or predatory economic activity. Finally, the presence of some entities may have positive effects in some cases, but negative effects in others or have impacts that are difficult to predict in isolation.

Table 1 Illustrative Foursquare Entity Categories of Interest: Presumed Impact on Human Health and Well-Being		
<i>Positive</i>	<i>Neutral or Ambiguous</i>	<i>Negative</i>
Supermarket	Convenience Store	Bar
Pharmacy/Drug Store	Military Base	Waste Facility
Bank	Warehouse	Check Cashing Facility
Spiritual Center (place of worship)	Gas Station	Prison
Athletic and Sports Facility	Recycling Facility	

One issue that must be resolved in conducting a project based on entity location is that of scale. Washington, DC is a medium-sized city bounded by somewhat arbitrarily drawn lines, which date from the city's designation as the U.S. national capitol, on three sides and the Potomac River on the remaining (south) side. Washington has approximately 20 zip codes, many of which are of very irregular size and shape. Accordingly, an analysis performed on the basis of zip code might not yield sufficient granularity and consistency to support an analysis of the type of interest here. Rather, an analysis conducted at the community level, according to widely recognized neighborhoods found throughout the city, would be more likely to yield valid results.

The identity of specific neighborhoods is not a field contained within Foursquare's data set. Fortunately, I have been able to obtain location data with this level of granularity from a different source. The real estate services firm Zillow has compiled data on neighborhoods in major U.S. cities and made them available for public use. I have located and downloaded this information, containing the latitude, longitude, and boundary locations of Washington, DC's neighborhoods. The data may be viewed and downloaded at the following link: <https://public.opendatasoft.com/explore/dataset/zillow-neighborhoods/export/?q=DC&refine.state=DC>. According to the designations used in Zillow's data set, the District of Columbia has 135 identified neighborhoods, yielding a level of granularity that should be

sufficient to meaningfully analyze information on this city of approximately 700,000 residents. It should be noted, however, that neighborhoods in the District of Columbia do not have officially designated boundaries, and different information sources may draw the boundaries between adjacent neighborhoods somewhat differently. Moreover, the lack of official neighborhood boundaries makes collecting demographic information on this level for the District of Columbia somewhat problematic.

Methods

The analysis described in this report was performed within a Jupyter notebook hosted on IBM's Skills Networks Labs web site. The analysis consisted of a series of commands written in Python code that moved through a sequence of activities to assemble, organize, display, and analyze data on amenities available to residents of specific neighborhoods in Washington, DC.

Creating the Frame of Reference

The initial step was to load the data described above provided by Zillow that names and delineates the boundaries and center point of each neighborhood by use of latitude/longitude coordinates. These data are resident in a .json file. After reading in these data, I created a pandas dataframe and then sequentially loaded in each neighborhood's name and the latitude and longitude of its center.

Mapping DC's Neighborhoods

To display the center and boundaries of DC's neighborhoods, I used Python's folium library to create a map showing this information. The result is shown in Figure 1 below, which also demonstrates that each neighborhood's name can be obtained by clicking on the blue circle found at its center.

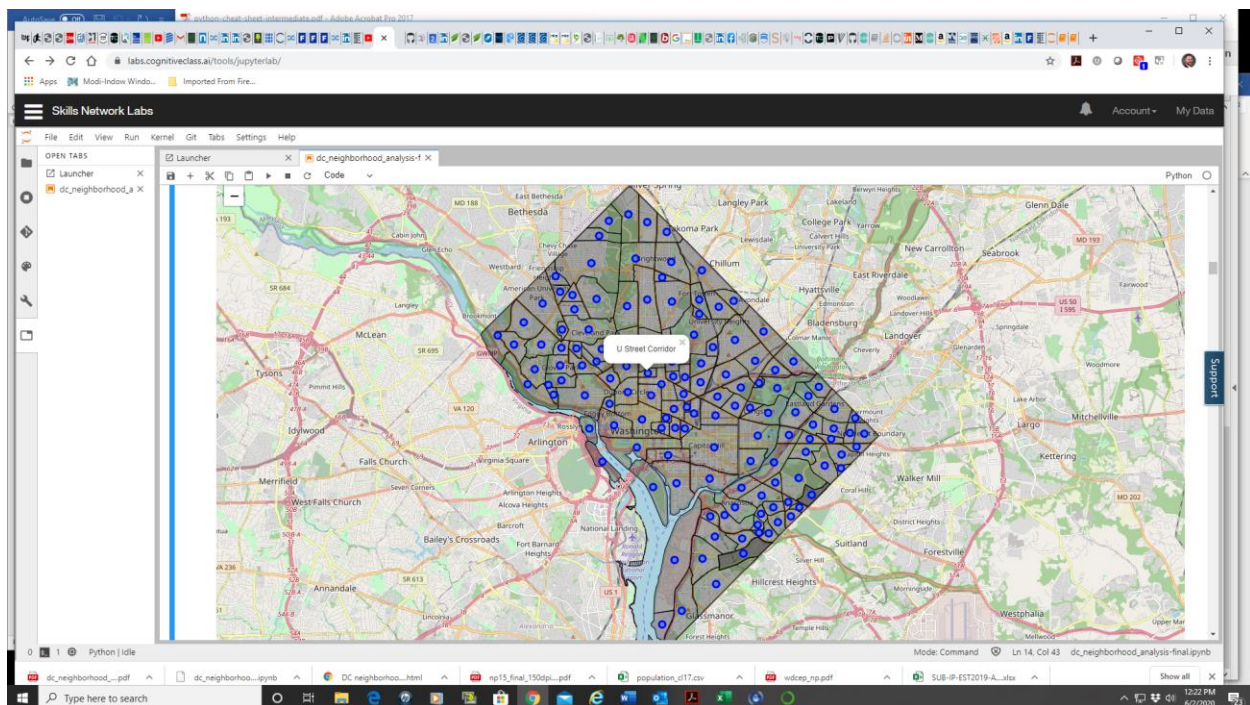


Figure 1 - Neighborhoods of Washington, DC

Searching for Venues by Neighborhood

To determine whether and to what extent residents of DC neighborhoods have convenient access to amenities that promote health and well-being, I performed a “venue search” of the Foursquare database. Foursquare maintains a very large set of detailed information on commercial and public entities that is searchable by location and category. For the present inquiry, I conducted a search around the center of each DC neighborhood using a radius of 1000 meters. Although this method is not entirely consistent when applied across more than 100 neighborhoods of widely varying sizes and shapes, it does produce a reasonable means by which we can begin to understand the number and types of amenities that are readily available to residents in different parts of the city.

This approach yielded more than 6,600 unique venues in more than 340 categories across DC’s neighborhoods. For a number of neighborhoods, Foursquare returned 100 venues, the maximum available using my queries, while others had far fewer.

Normalizing the Data for Analysis

To prepare the data for further analysis, I ran some Python code to assign dummy variables to the presence or absence of each venue type for each neighborhood, and then calculated the frequency with which each venue type is found in each neighborhood. An example of this information is presented in Figure 2.

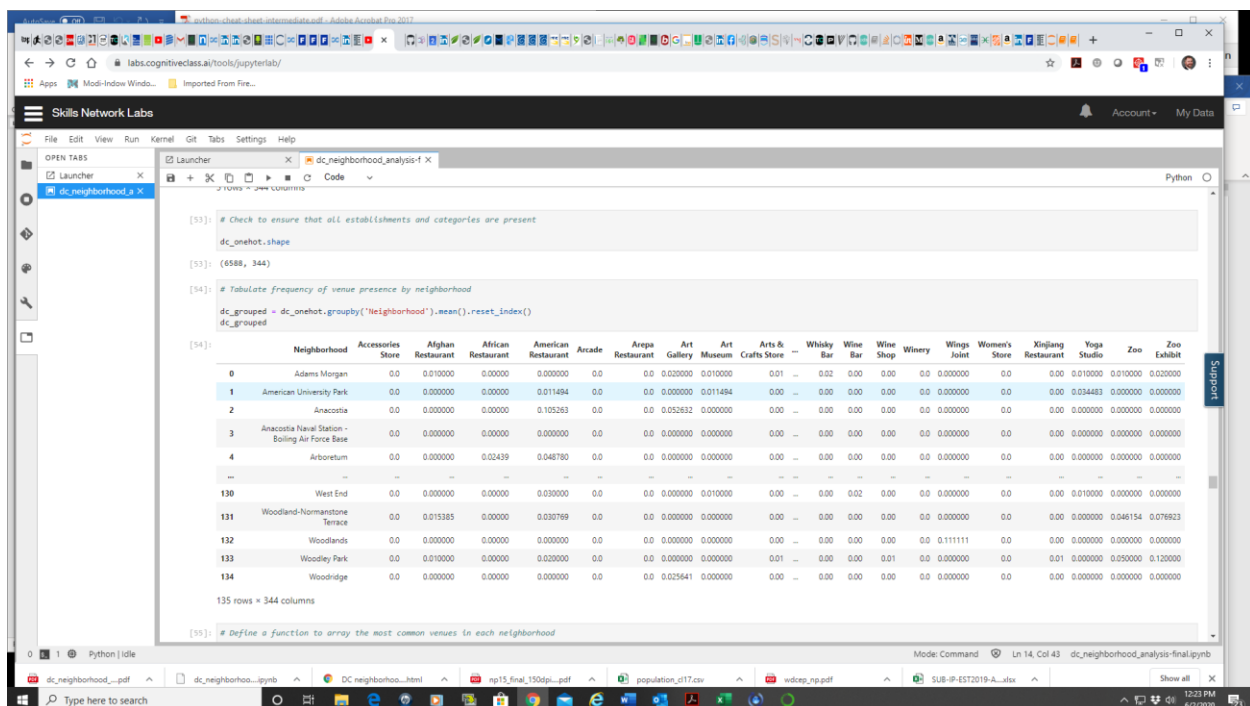


Figure 2 - Frequency of Venue Type Presence in DC Neighborhoods

Pattern Identification

Given the thousands of individual unrelated venues within the District of Columbia and the hundreds of distinct venue types found by the Foursquare database, I needed a means by which patterns across individual neighborhoods might be identified. One such tool is K-Means clustering, which partitions a set of data into non-overlapping clusters that have no internal structure. This technique is suitable in situations in which one wants to identify data points (neighborhoods in this case) that share certain similarities (form a “cluster”) and distinguish them from others that belong to another “cluster” that are in some way quite different. My working hypothesis was that neighborhoods in more affluent areas of Washington, DC would be served by different types of venues than neighborhoods in less affluent areas of the city.

The key issue that must be decided upon in using the K-Means technique is the number of clusters that will be defined. Washington is an unusual, if not unique, U.S. city because it serves as the nation’s capitol. As a consequence, there are many areas of the city that are dominated by U.S. federal government buildings, museums, monuments, and other infrastructure as well as parks and open space available for Americans and visitors to enjoy. For this reason, my conjecture was that Washington, DC might be well represented by three clusters—one representing the U.S. government presence, one representing affluent areas of the city, and one representing areas not well served by providers of amenities associated with good health and well-being.

Results

Results of applying the K-Means technique are shown in Figure 3 below. One can observe one very small cluster (three neighborhoods) at the northernmost point of Washington, DC, and two much larger clusters occupying most of the western and central portions of the city in one case, and the eastern and portions of the northeastern portion in the other. The map shown in Figure 3 very clearly shows a distinction between neighborhoods that lie primarily on the eastern side of the Anacostia River and those on the west side. The former roughly correspond to DC’s Wards 7 and 8, which as noted above historically have had higher African American populations, lower incomes, and higher crime rates than most other areas of the city.

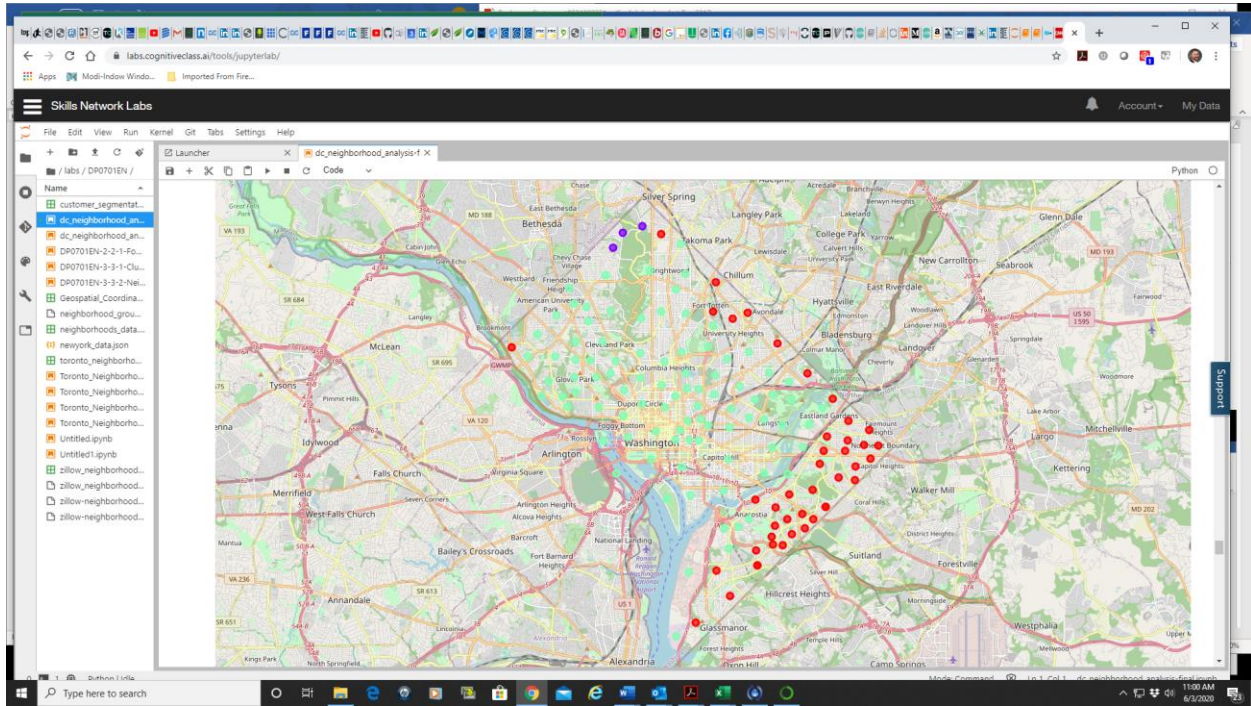


Figure 3 - Neighborhood Clusters in Washington, DC

To further examine what factors might lead to the three clusters being defined in the way assigned by K-Means, I examined each cluster individually, looking at the types of venues that appear to be most common in neighborhoods within that cluster. Some results of this inquiry are presented in Table 2.

Table 2 Most Common Venues in Neighborhoods Across Two Large Clusters						
Venue Category	Affluent Neighborhoods (n = 91)			Less Affluent Neighborhoods (n = 41)		
	Number of Neighborhoods			Number of Neighborhoods		
	Most Common	2 nd Most Common	Third Most Common	Most Common	2 nd Most Common	Third Most Common
Restaurant	6	16	14	3	2	6
Coffee Shop	16	11				
Park	8	7	9	4	5	
Bank						
Hotel	7	6	6			
Bar		6	5			
Bus Stop		4				
Intersection				3	3	5
Carry Out					2	
Convenience Store				8	15	6
Liquor Store	6			10	4	3

Although it appears that many neighborhoods in both of the larger clusters offer access to parks and other similar green space, the most common commercial venues available to residents appear to vary

substantially between the two clusters. The cluster representing the more affluent areas of the city houses many neighborhoods in which coffee shops and restaurants are very common, and in some (presumably, in or near the Downtown area), hotels and bars also are frequently available. In contrast, neighborhoods in the cluster on the eastern edge of the city more commonly house convenience stores and liquor stores as their primary commercial entities.

The cluster containing the three neighborhoods just south of Silver Spring, MD comprises areas that adjoin parkland, which appears to have caused K-Means to distinguish it from other areas of Washington, DC. The most common venue in all three of these neighborhoods is “trail,” i.e., access to a walking trail or path. Review of the other most common venues in these two neighborhoods reveals no other distinguishing characteristics.

Discussion

This document outlines an analysis of potential geographic disparities in access to amenities that promote or detract from the health and well-being of residents at the neighborhood level in Washington, DC. Results generated from this analysis suggest that there are identifiable differences in such access based upon where residents live. In particular, DC residents who live east of or near to the Anacostia River (Wards 7 and 8) have access to a different set of choices than those in most other parts of the city. Although this analysis is preliminary and was performed at an exploratory level, several findings may merit further study.

The large number of coffee shops (e.g., Starbucks) and restaurants in many neighborhoods may be obscuring the frequency with which other key venue types (e.g., grocery stores) are present and available to residents across the city of Washington. A more detailed or secondary analysis focused more tightly on these other categories may be warranted.

Somewhat surprisingly, neighborhoods in which I expected U.S. federal buildings and other properties to dominate were not selected by K-Means for their own cluster. This may reflect the extensive growth in commercial activity and residential construction that has pushed north and west from the downtown area during the past 15 years or so. Because of this growth, there are many new residents (and businesses to serve them) in areas that previously were rather quiet after normal business hours. Some further study, including possibly defining more clusters, might shed some light on where the main differences across areas of the city lie, and why.

Finally, integration of more extensive data including information on racial, economic, and educational characteristics of resident populations could bring some added sophistication and potential value to the preliminary analysis presented here. These data and appropriate further analysis could help determine how deep neighborhood-level differences run and the extent to which they reflect (or predict) quality of life for the people of Washington, DC.

Conclusions

Based upon the preliminary analysis described in this report, it appears that my conjecture that residents of less affluent areas of Washington, DC have access to a different set of amenities than their

neighbors in more affluent areas is confirmed. Further work will be needed to explore whether and to what extent these differences may have health and well-being implications for city residents.