



Data Science
Capstone Project

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

Peter A. Soyka

Outline

- Introduction/Business Problem
- Data
- Methods
- Results
- Discussion

Introduction/Business Problem

- Washington, DC - a racially and economically diverse city and the capitol of the United States
- Historically majority African-American, but demographics are changing rapidly
- Many new immigrants are younger, more upwardly mobile, and more racially diverse
- Geographic, racial, and economic divides:
 - Center and west areas – more caucasian, higher education levels and incomes, working professionals
 - Eastern and northeast areas – more african-American, less education, and lower incomes
- Residents east of the Anacostia River (Wards 7 and 8) concerned they receive fewer city services, have higher crime rates, and have access to fewer amenities
 - Limited access to fresh food, drug stores, and health care facilities of particular concern

Study Purpose

Determine whether there are detectable differences in access to providers of goods and services that promote good health and well-being (or detract from it) across Washington, DC neighborhoods

Illustrative Provider Categories of Interest: Presumed Impact on Human Health and Well-Being

Positive	Neutral or Ambiguous	Negative
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	

Data and Methods

Needs:

- Geographic areas of DC to be studied
 - Political boundaries (Wards) and zip codes are too large
 - **Neighborhoods** are not defined precisely but are widely used and understood
 - Need names, boundaries, geographic coordinates
- Categories of venues that might enhance or detract from health and well-being of neighborhood residents
 - Many available, some with ambiguous effects

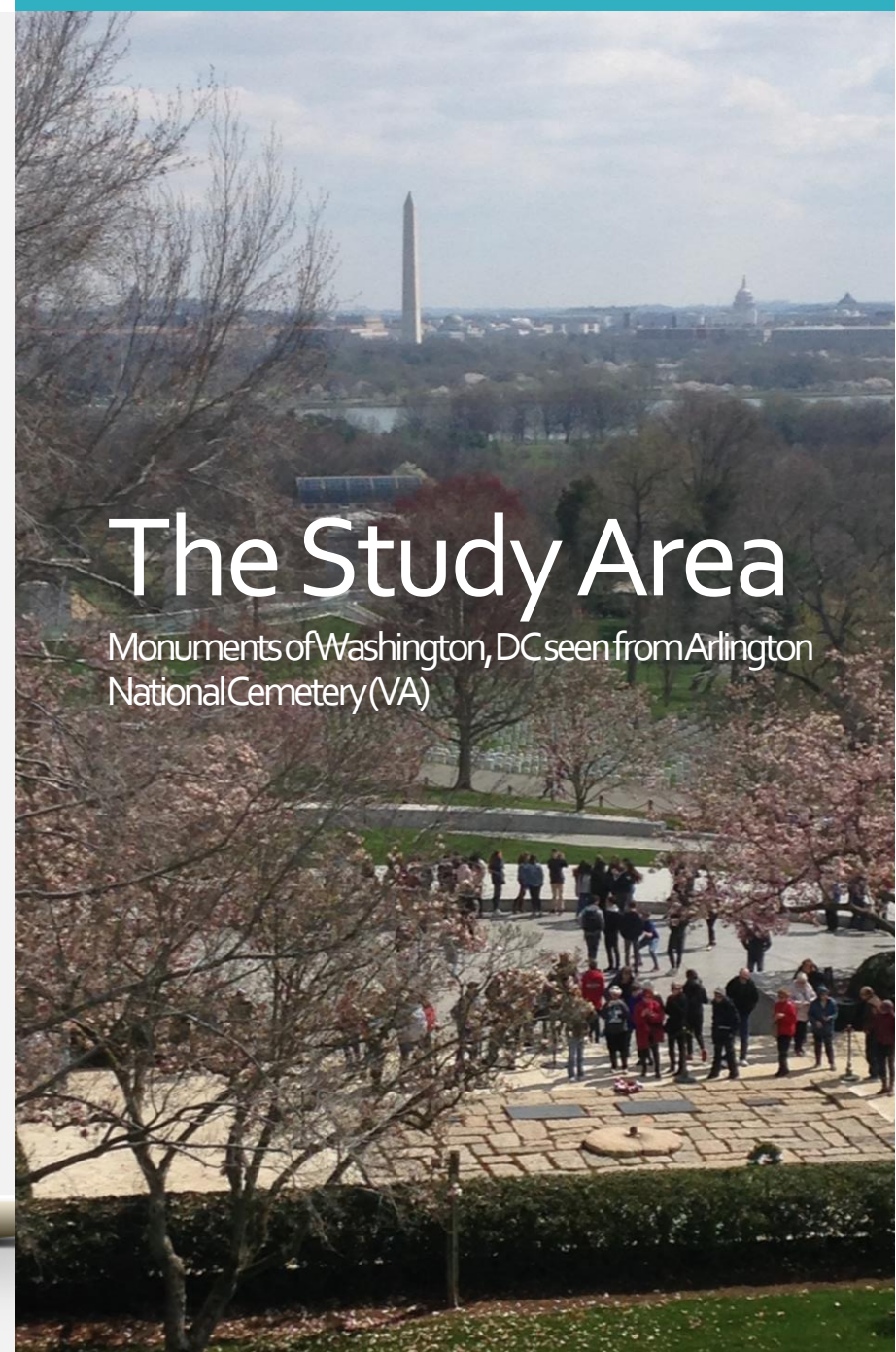
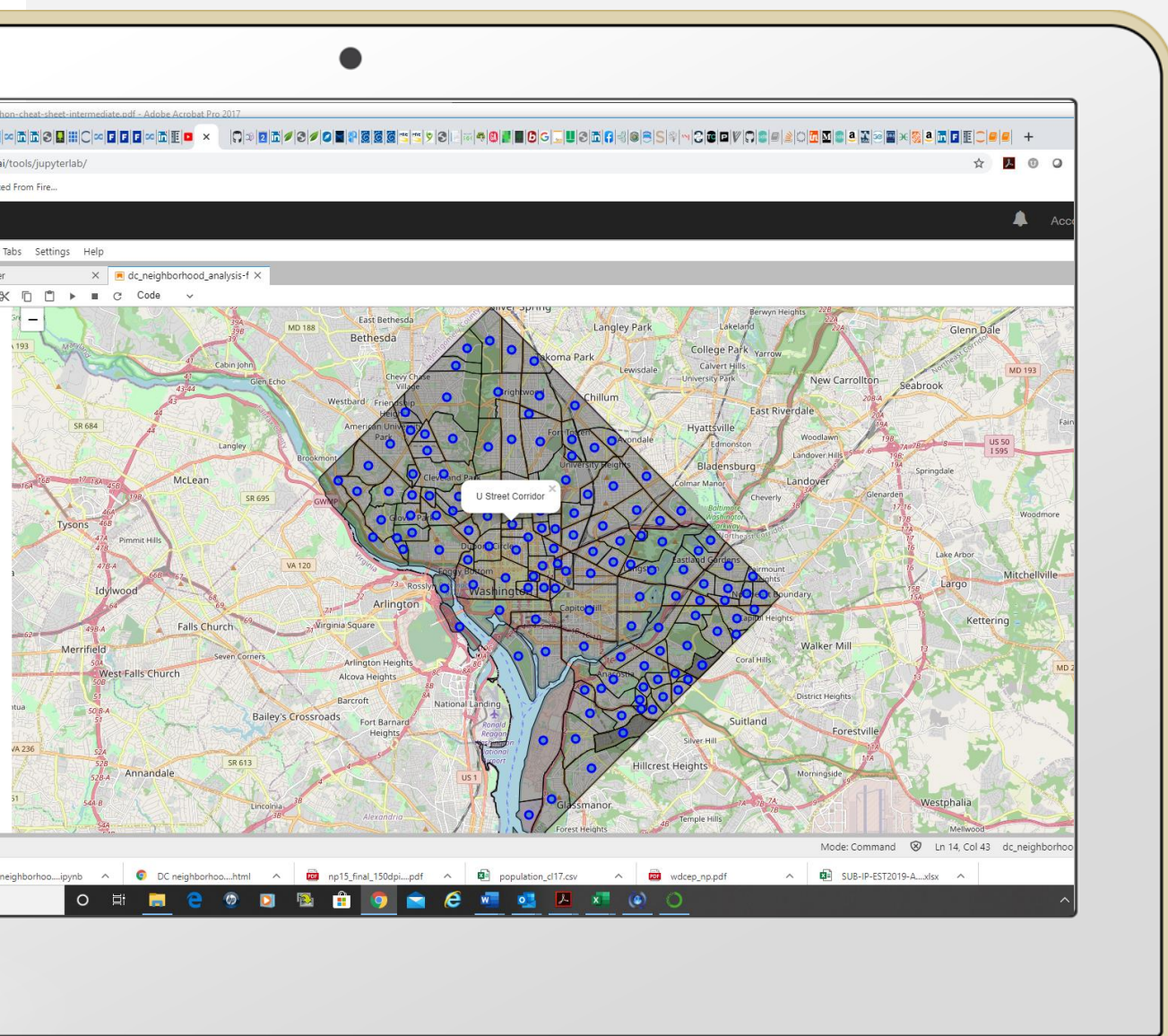
Sources:

Zillow – data on neighborhood locations and boundaries

Foursquare database – information on venues of many types, with specific geolocation

Methods:

Operate in a Jupyter notebook, use python code to assemble and manipulate data, and apply additional libraries to perform mapping, computations, and other tasks



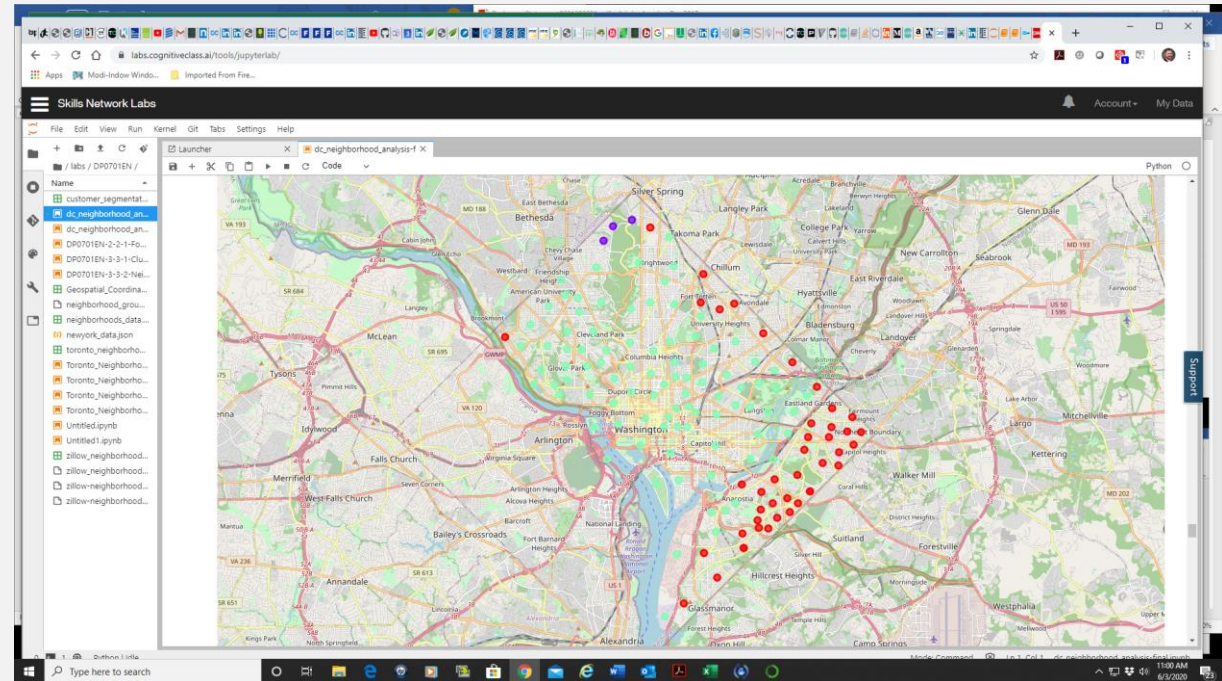
The Study Area

Monuments of Washington, DC seen from Arlington National Cemetery (VA)

Search Parameters and Results

Foursquare search using a 1,000 m radius around the center of each neighborhood

- 135 neighborhoods
 - More than 6,600 venues
 - More than 340 unique categories
- Pattern Identification – Use K-Means Clustering
 - 3 clusters defined
 - More affluent
 - Less affluent
 - Heavy federal government presence



Results – 3 Clusters

- 1 very small (3 neighborhoods), adjacent to extensive parkland. Most common venue – trail!
- 1 large (91 neighborhoods), comprising the central business district and most areas to the west. Largely affluent and well-educated population.
- 1 moderate in size (41 neighborhoods), consisting of all areas east of the Anacostia River and several more on the northeaste periphery of the city (and one outlier on the west side).
- *What distinguishes the neighborhoods in the latter two clusters?*

A Closer Look

Most Common Venues in DC 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

A Closer Look

- Many neighborhoods in both of the larger clusters offer access to parks and other similar green space
- More affluent neighborhoods often have...
 - Coffee shops
 - Restaurants
 - Hotels and bars
- In less affluent neighborhoods, the most common venues are often...
 - Convenience stores
 - Liquor stores

Discussion and Limitations

These preliminary results suggest differences in access to amenities that promote or detract from the health and well-being of Washington, DC residents based upon where they live

- The Anacostia River appears to be meaningful dividing line
- The expected “federal” cluster did not materialize. This may be due to large amounts of new residential construction and population growth downtown and in the areas north of most federal buildings.

Limitations

- The “Starbucks effect:” the large number of coffee shops and restaurants in many neighborhoods may be obscuring the presence of other key venue types (e.g., grocery stores). A more detailed analysis using more tightly defined venue categories may be warranted.
- This initial analysis was limited to a “first cut” at the presence of different types of amenities at the neighborhood level and was exploratory in nature. Making some refinements to this analysis and integrating racial, economic, and educational characteristics of resident populations could bring some added sophistication and potential value to the preliminary analysis presented here.



Thank You

Peter Soyka 

petersoyka2@gmail.com 