Development of Geospatial Methods for Assessing Food Environment

Michelle Poblete

Masters of Science in Geographic Information Science (MSGISci)

Department of Geography, California State University, Long Beach

Introduction

One in five children, ages 6-19, are obese and more than one third of U.S adults have obesity (CDC 2015). The CDC's Department of Health and Human Services found that lack of access to healthy food retailers, such as supermarkets, in communities has been associated with a poor-quality diet and increased risk of obesity. Similarly, studies suggest greater access to convenience stores and fast food restaurants have been associated with greater likelihood of obesity.

The purpose of this study was to compare food environment in different cities of Los Angeles county using GIS technology. I attempted to find an accurate measurement of food environment which is the ratio of healthy and unhealthy food retailers in a given area since this information is likely to be of relevance to studies looking at the factors influencing obesity.

The study areas were chosen so that it would be possible to compare cities that are affluent, poor, and a mix of both. As shown in figure 1, the cities selected for this project were Beverly Hills, Inglewood, Lancaster, and Maywood.

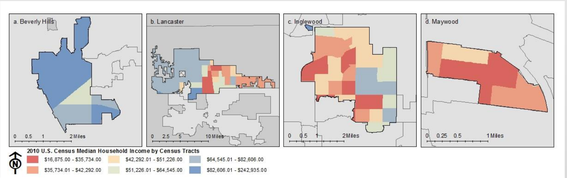


Figure 1. *Project study areas symbolized by census block's median house hold income*

Data and Data Sources

The 2010 income data for the county of Los Angeles was downloaded from the U.S Census Bureau to assist with choosing the four study areas with different ranges of median household income. To obtain food retailer information, CSV files were pulled from ReferenceUSA, which included addresses and NAICS codes. The LA County CAMS stress address locator was obtained from the LA County GIS data portal to geocode the food retailer data. Census block data was also obtained from the U.S Census Bureau and was used as a scale for measurement of food environment within the study areas.

Table 1. *List of data and data sources used in the project*

Dataset	Source
Food Retailer	ReferenceUSA
Census Blocks	U.S Census Bureau 2010
Median Household Income	U.S Census Bureau 2010
Address Locator	LA County GIS Data Portal

Timeline

Table 2. *Project timeline*

Proposed Date of Completion	Description	Actual Date Completed	Description
5/2/2017	Obtain business data for three additional study areas and clean data in CSV file. Find public transportation data.	5/8/2017	Collected all data needed for analysis
5/11/2017	Complete base map for each study area. Create script for buffer and network analyst closest facility.	5/11/2017	Completed base map for each study area
5/15/2017	Data processing: run buffer and network analyst for each study area and check for errors in script.	5/12 -5/30	Personal leave
5/18/2017	Finish buffer, network analysis. Create script for GWR analysis for available public transportation and food retailer coverage.	6/2/2017	Revised project – outline new analysis
6/2/2017	Finish GWR analysis and all static maps. Begin design and production of interactive web map	6/17/2017	Created Thiessen polygons and calculated mRFEI for one study area. Created multiple ring buffers around each food retailer and calculated food environment score
6/9/2017	Test web map and search for any errors.	6/27/2017	Created Model and ran analysis on the remaining study are
6/16/2017	Publish web map	7/12/2017	Completed Python script for food environment scores for each census block
		7/15/2017	Published web map

Methodology

Three methodologies were considered for assessing the food environment of each of the four selected cities. The first method that I developed relied on the Thiessen polygons which were created using ArcMap's Create Thiessen Polygons tool. The input dataset was the healthy food retailer point feature class and the extent was the city's boundary. The modified food environment index (mRFEI) score was then calculated for each polygon using the following equation:

$$\frac{1}{\#\ of\ unhealthy\ food\ retailers\ +1} * 100$$

For all methods, food retailers were classified by their North American Industry Classification System (NAICS) codes. The healthy food retailers included for this project were supermarkets (445110), fast food restaurants (722513) and convenience stores (445120) were categorized as unhealthy retailers.

The second method that I developed was designed to show the influence each store has in a given area. To do this I created multiple ring buffers around each healthy and unhealthy food retailer and assigned a score depending on its health classification. The scores ranged from -5 to 5, areas closest to healthy food retailers were assigned a score of 5 while areas closest to unhealthy food retailers were assigned a score of -5. The overlapped areas represent the various choices residents have in their immediate area. I took the score of each buffer in the overlapped area and calculated the sum, which was the new score for the given area.

In the third method, the scores from the multiple ring buffers were unioned with 2010 U.S. Census blocks to see how the food retailers impact the whole city and not just the area where they are clustered. The scores for each 2010 U.S. Census block were calculated by adding up the food retailer influence score for each polygon within its boundaries and these values for the block. Because this operation was rather complex, a Python script, shown in figure 3, was created to avoid manually calculating each score for every study area. Each method was repeated four times for the four study areas and to efficiently run the data processes a model shown in figure 2 was created.

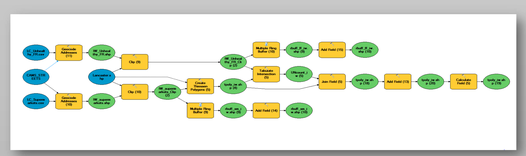


Figure 2. *Model of Thiessen polygon and multiple ring buffer methods*

```
import arcpy
workspace = arcpy.env.workspace = r"C:\Users\Michelle Poblete\Desktop\Thesis2\cb_scores.gdb"

city = "maywood"
fipslist = []
cbtotalscores = []
fipscursor = arcpy.da.SearchCursor(city, "FIPS")

#Create list with all FIPS (census block IDS)
for fips in fipscursor:
    if fips[0] not in fipslist and fips[0] != " ":
        fipslist.append(row[0])

#Create list to input census block scores - for all census blocks in the fipslist
for censusblocks in fipslist:
    cbtotalscores.append(0)

del cursor

#define new cursor to search through fields used to calculate polygon area score
polycursor = arcpy.da.SearchCursor(city, ["FIPS", "cb_area", "poly_area", "score"])

#define expression used to calculate polygon area score
exp = "!poly_area!/!cb_area! * !score!"

#Calculates the polygon area score for all attributes that have a census block ID (FIPS)
for row in polycursor:
    if row[0] in fipslist:
        arcpy.CalculateField_management(city, "poly_ar_score", exp, "PYTHON_9.3")
        #Excludes polygons who do not fall within the census block boundary
        if cb[1] > 1:

            polyarscore = (row[2]/row[1]) * row[3]
            fipspositioninlist = fipslist.index(row[0])
            cbtotalscores[fipspositioninlist] += polyarscore
            #cbtotalscores[fipspositioninlist] = cbtotalscores[fipspositioninlist] + polyarscore

#Adds all scores sharing the same FIPS together for the census block score
with arcpy.da.UpdateCursor(city,["FIPS","cb_score"]) as fipscursor:
    for fips in fipscursor:
        #Creates an order to the FIPS
        fipspositioninlist = fipslist.index(fips[0])
        print str(cbtotalscores[fipspositioninlist])
        fips[1] = cbtotalscores[fipspositioninlist]
        fipscursor.updateRow(fips)
```

Figure 3. *Python script calculating food environment scores for each census block*

Results

All three methodologies indicated that Maywood had the highest overall food environment score while Beverly Hills had the lowest. Also, within each city there are specific areas where food environment is poor and this pattern is consistent throughout all three methods which was expected. In figure 4, the symbology of each polygon depends on the mRFEI score, the higher the score the darker the color. Higher scores represent a smaller number of fast food restaurants and convenience stores in the area, and indicate that health food retailers face little competition from unhealthy food retailers. Greater levels of competition within a polygon indicate an increased chance that a resident will chose to purchase food from an unhealthy retailer.

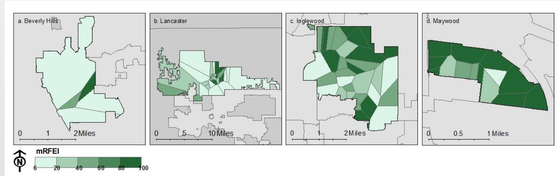


Figure 4. *Thiessen polygons around each supermarket in the study area symbolized by mRFEI score*

Figure 5 shows maps of the combined multiple ring buffers created around each healthy and unhealthy food retailer, with the symbology indicating the level of influence on consumers from health and unhealthy nearby food retailers. Scores of 0 were symbolized with the color yellow, as these areas are considered to have a neutral influence.

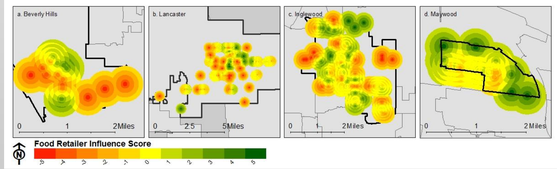


Figure 5. *Multiple ring buffers around each food retailer symbolized by their influence score*

Figure 6 show the results of the last methodology conducted for this project. The census block score was calculated using the scores generated from the ring buffers and normalized by the given area. The symbology for these maps are the same for figure 5, bold red represents low negative scores and the opposite for dark green. The areas where there are no healthy or unhealthy food retailers are symbolized with a crosshatch.

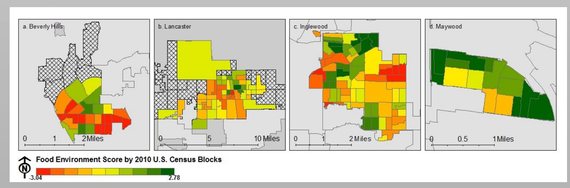
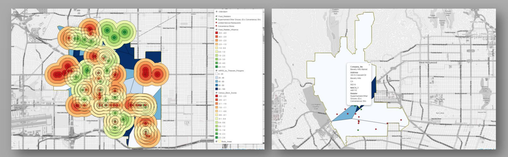


Figure 6. *Food environment score per census block in each study area*

Figure 7 and 8 are screenshots of the interactive web map created to the results of all three methodologies as well as the food retailer points for all four study areas. The web map allows users to control the visibility of each layer using an interactive legend. The web map also allows users to click on each feature which will then open a pop up that contains feature attributes including the mRFEI or food environment score. The food retailer points can also be selected, showing the address, NAICS code, and name of retailer.



Figures 7 & 8. *Screenshots of the interactive maps created to display the result of this project*

Discussion

There were no monetary costs for this project. All datasets were obtained through free data sources and a student software license for ArcMap, which was used to conduct the analysis, was provided to me through CSULB. QGIS, a free, open source geographic information system software program, was also used to create the interactive web map to display the results of this project.

Overall, the methods were efficient but had a few limitations. The majority of this study's limitations were caused by the chosen extent for each study area. This project focused on studying individual cities in Los Angeles and only analyzed food environment within the study areas' boundary. Therefore, food retailers and census blocks were clipped to fit within the city's borders, ignoring the extent of the influence they may have on the study area. In the future, I would advise keeping census blocks whole and including all those that touch within the study area. I would also include all food retailers outside of the city boundary whose ring buffers intersect the city. The distance extent of each ring buffer in this study was chosen based on personal experience and surveys should ideally be conducted to better understand how far an individual is willing to walk to nearby food retailers.

The results of each method would have more validity if the study area extent was not confined to each city's boundary. The accuracy of these results decreases in the outer regions of each study area. However, I believe the results are still significant as each map accurately shows the ratio and influence of healthy versus unhealthy food retailers within a city boundary. This project successfully compared food environment in different cities of Los Angeles. The interactive web map effectively allows users to visualize the food environment of each city and makes it it easy to compare one area to another. Local governments may utilize this tool to determine the need for healthier options such as farmers markets or local gardens within their respective city.

This study assumed that all fast food restaurants and convenience stores exert equally negative influences on their surrounding areas, though this is not likely to be the case. To avoid personal bias, I categorized each fast food restaurant and convenience store by its given NAICS code and defined it as unhealthy as had been done in several previous studies found during a review of relevant literature. I could further increase the accuracy in the future by surveying the inventory of each food retailer to determine the extent to which its food options are healthy or unhealthy.

Conclusion

While this study produced interesting results, both the quality and scope of the project can be expanded in the future. I suggest using the methods of this study to analyze the influence of food environment on various diet related diseases using statistical techniques such as a geographic weighted regression in addition to an ordinary least squares regression. I also recommend factoring in the demographic differences in each city, such as race and income to examine the relationship between these factors and food environment.

Public transportation plays a large role in food accessibility and could help determine a better ring buffer distance from each food retailer. I recommend using public transportation data and surveying residents on their transportation preference to produce more detailed and accurate results than I was able to obtain in this study. Another preference to consider is which food retailers residents have the tendency to visit and how this is affected by proximity. I suggest examining the sales volume of each store or surveying residents about stores they frequent to give an accurate influence weight to each food retailer.

The methods of this project can help influence policymakers to enact policies that can help improve the low ratios of healthy to unhealthy food retailers in a given area. Because I failed to include all food retailers that may influence each study area, the results are not currently accurate enough to support these types of efforts. However, the results were still able to identity areas where unhealthy food retailers exert a greater than normal level of influence. These findings also demonstrate how food environment can differ depending on the scale at which it is measured.

Submitted in partial fulfillment of the requirements of the Masters of Science in Geographic Information Science(MSGISci), August 12, 2017.

For additional information please contact: Michelle Poblete

Michelleapoblete@gmail.com