**A case study of food access data sources in Los Angeles, CA.**

by

Charles Maurice Hall

A Thesis Presented to the

Faculty of the USC Graduate School

University of Southern California

In Partial Fulfillment of the

Requirements for the Degree

Master of Science

(Geographic Information Science and Technology)

July 2016

Copyright ® 2016 by Charles Hall

To MK, this wouldn’t have happened without you.

Table of Contents

List of Figures v

List of Tables vi

Acknowledgements vii

List of Abbreviations viii

Abstract ix

Chapter 1 Introduction 1

1.1 Defining healthy food 3

1.2 Defining food access 3

1.3 Traditional data and AGI 3

1.4 Primary research question 4

1.5 Motivation 4

1.6 Thesis Organization 5

C 6

2.1 Traditional Food Access Studies 6

2.2 USDA Food Access Estimates 7

2.3 Classification and AGI 7

Chapter 3 Methods 8

3.1 Methods 8

3.2 US Census Data 8

3.3 Esri Business Analyst Data 13

3.4 Ambient Geographic Information 13

3.4.1 Computing environment 13

3.4.2 Google Places API 13

3.4.3 Yelp API 14

3.4.4 In field evaluation of facilities 14

REFERENCES 15

Appendix A – Code 16

Apendix B - Market Evaluation Worksheet 20

# List of Figures

Figure 1 Study Area 2

Figure 2 Site Selection Map 10

# List of Tables

Table 1 A short title 5

Table 2 Summary of previous research including examples of work completed by other   
scholars working around the world 5

Table 3 A title which has been shortened from the actual caption 6

# Acknowledgements

I am grateful for the support and encouragement offered by my partner MK, both in preparing for and completing this program. I am grateful to the faculty at the USC Spatial Sciences Institute for all of the support and encouragement offered throughout my journey. I would like to thank Jennifer Swift for encouraging me and supporting my presentation at the Esri Developer Summit. I would like to thank USC and Adnan Choudhary for giving me the time, space, and encouragement required to complete this project. I would like to thank Alex Sosa for his assistance wrangling data.

# List of Abbreviations

ACS American Community Survey

AGI Ambient Geographic Information

GIS Geographic information system

GISci Geographic information science

SSI Spatial Sciences Institute

USC University of Southern California

VGI Volunteered Geographic Information

# Abstract

This thesis performs a comparative analysis of traditional models of food access and a proposed model of food access that uses ambient geographic information (AGI). Moreover, food businesses are often manually classified, which limits the number of businesses used for a given study. This paper explores AGI as a potential improvement in the classification of food businesses. Field research is conducted in a subset of the selected facilities in order to determine the actual quality of the data retrieved from the experimental sources. The goal is to create a more nuanced and accurate representation of food access for a given person in a given place. Finally, data is compared for areas with different socio-economic conditions. Median income, car access, and percent minority from the 2010-2014 American Community Survey (ACS) 5-year estimates are used to define contrasting study areas. Two census tracts in Los Angeles have been selected for the study area using these criteria. An affluent area near La Canada, and a less affluent area in South Los Angeles were selected. This paper explores the quality and completeness of three data sets for census tracts with contrasting socio-economic conditions in order to identify whether or not problems exist with traditional methods and data. Furthermore, this paper compares the data from census tracts with contrasting socio-economic conditions in order to determine whether or not the data varies based on the community served.

# Chapter 1 Introduction

Food security and food access have become popular topics of discussion in and out of academia in the recent years. Books, articles, and films have been produced investigating these topics; however, they often rely on commercial data sources that describe the businesses in the study area. Though these data sets provide information about the businesses such as size, income, and number of employees, the data remains problematic because it lacks any measure of the variety and quality of goods offered. Consequently, the results of studies performed with this data can be difficult to interpret and often require field verification in order to meaningfully interpret the results. This thesis investigates the overall quality and consistency of commercial data, and the utility of augmenting commercial data with ambient geographic information (AGI) from Google and Yelp. In so doing this paper will evaluate commercial data and investigate whether or not AGI can reduce the need for field verification. The paper will use the most current data available and is not concerned with how access changes over time. Moreover, though this paper addresses access to food, it does not address health outcomes. Finally, although community gardens, farmer’s markets, and food trucks represent meaningful additions to the food environment, they are not discussed in this paper. The primary focus is testing, verifying, and augmenting the data that is often used in current models.

The initial study will compare data for two contrasting census tracts in the greater Los Angeles area. Data from the 2010 – 2014 American Community Survey (ACS) was used to classify census tracts with regard to their median income, vehicles per occupied dwelling, and percentage of the population who is white. The initial site selection resulted in two tracts: 460700 and 224020. Tract 460700 is located in La Canada, CA, an affluent area north of downtown Los Angeles. It was chosen because the population is predominantly white with high income and multiple cars per household. Tract 224020 is located in an area of South Los Angeles, CA, and was selected because it contains a large minority population with low income and limited access to cars. The study area is visualized in figure 1 on the following page.

## Defining healthy food

A definition of healthy food is necessary in order to address the question of food security and access. The absence of a definition would erroneously show individuals who only have access to low quality food to have good access. The US Department of Agriculture (USDA) publishes dietary guidelines that identify healthy food and diet choices. The USDA suggests a diet rich in fruits and vegetables that avoids processed sugars and other calorie rich, nutrient poor foods. Moreover, the guidelines suggest limiting saturated fats and sodium.

A worksheet was created in order to standardize the evaluation foods available in markets within the study area. The USDA has published a Food Store Survey Instrument designed to be used by a lay person in order to evaluate the quality of a market. The University of Pennsylvania has also published their Nutrition Environment Measures Survey (NEMS), which has also been designed to allow a lay person to evaluate the quality of a local market. These documents were reviewed and adapted in order to produce the final worksheet that was used in the evaluation of markets in the study area. The sheet includes both qualitative and quantitative measures that indicate the presence, quality, and cost of fresh food, dairy, and meat.

## Defining food access

This study considers vehicle access and income when assessing access. These variables will be considered for each of the two census tracts that are being evaluated. The density of facilities per local population will be considered in addition to the overall quality and variety of food available in each location. This is important because food access studies often use one quarter of a mile as an acceptable walkable distance for a person without a vehicle.

Moreover, the cost of staple items in each facility will be evaluated. A person in a given place needs to be able to afford healthy food in order to choose to consume it. Prices will be collected and analyzed for staple goods in each facility being studied. This will provide some insight into how process change from facility to facility. Furthermore, patterns could emerge that demonstrate variation from census tract to census tract. It might begin to address the question of how access to healthy food based on price and availability varies.

## 1.3 Traditional data and AGI

results obtained from the field survey.

## 1.4 Primary research question

How accurately does commercial data represent the reality of the businesses in the study area, and what benefit is gained through the inclusion of AGI? Businesses will be surveyed with a worksheet in order to evaluate the variety and quality of foods available. AGI from Yelp and Google places will be aggregated with the commercial data and investigated in order to determine whether or not they provide valuable additional data about the classification and quality of food facilities in the study area. Results will then be interrogated in order to determine whether or not the commercial data provides meaningful insight into the environment, and what benefit is reaped through the inclusion of AGI. Finally, field research results will be used to evaluate the overall quality of access in each of the study areas.

## 1.5 Motivation

Numerous studies of food access have been conducted with the aid of GIS technology. However, the data sources used in these studies are often problematic and generally require the author to visit the facilities or make assumptions about them. Traditional data sets contain limited details about the businesses that they represent, which complicates interpretation of the results. It is not sufficient to calculate the distance to the nearest market when determining whether or not a person has access to healthy food in a given place. Additional information about the market would greatly inform the results.

Traditional food access studies often use North American Industry Classification System (NAICS) codes in order to select and classify food related businesses. NAICS codes in Esri Business Analyst include a proprietary two-digit suffix created by Dunn and Bradstreet. However, there is no key available that defines the meaning of the suffixes. Esri was unaware of the existence of a key when contacted by telephone. Moreover, a Dunn and Bradstreet representative was unaware of the existence of the proprietary suffix. Finally, a business librarian was also unaware of the suffix and unable to provide insight into their meaning. Manual aggregation and investigation of the codes revealed that neither the six digit NAICS code, nor the six digit code with the two digit suffix, classified food facilities into more than very general categories.

NAICS codes aggregate businesses into general categories such as market or restaurant. Consequently, food access studies often struggle with the classification of food facilities because of the use of NAICS codes. For example, it can be difficult to distinguish different types of markets and restaurants based solely on NAICS codes. Julienne Gard chose to work from the assumption that medium and extra-large facilities represented supermarkets in urban and suburban neighborhoods respectively (Gard 2016). Others have resorted to manual classification, with the stipulation that the data set will only include recognizable national chains (Morganstern 2015). This paper addresses the classification problem through the use of volunteered geographic information (VGI).

AGI tends to be classified to a higher level because of the number of people who are able to contribute to the effort. Though the risk of misclassification exists, the increased quality of classification is a worthwhile trade. API’s like those provided by Yelp or Google Places allow for the selection of facilities by geographic location, and provide significant attribute data. Unlike NAICS codes, the classification is often textual, such as restaurants -> family -> burgers. A more robust classification system will result in more robust and nuanced results.

Finally, food access studies often indicate that urban dwellers have high access scores when compared to suburban and rural populations. This conclusion, however, could be misleading because it assumes that all food facilities are of equal quality. This study intends to avoid that assumption and interrogate the data in greater detail in order to avoid this and better understand what the results mean for a given person in a given place.

## 1.6 Thesis Organization

The remainder of this thesis includes a literature review, methodology, results, and conclusions. The literature review, offered in chapter 2, provides background for this study. It will examine data and methods used in previous food access studies and provide context for the results produced by this study. The methodology described in chapter 3, reviews the tools, data sources, and methods used to conduct this study. The Python code used to acquire data from Google and Yelp can be found in appendix 1. The results presented in chapter 4 describe the data produced by the study, and provides some descriptive statistics for the data. Chapter 5 interrogates the results and offers conclusions about how much value AGI could potentially add to a food access study. Moreover, the results will be evaluated with respect to each census tract in the study area in order to determine whether or not quality differences exist. Finally, the results will be evaluated with respect to the data gathered in the field in order to inform the discourse surrounding food access studies with a quality assessment of the food facilities indicated by the data.

# Chapter 2 Literature Review

As discussed in the previous chapter, the data used in food access studies continues to be problematic. This chapter will review methods used in food access studies that have been conducted in order to illustrate the challenges that the data presents. Various classification methods have been employed in food access studies; however, many of them rely on indirect indicators such as square footage of the facility or annual sales. Other methods employ local knowledge in order to manually classify facilities.

The intention of this chapter is to present a clear representation of recent food access studies. Particular attention is paid to the methods used to classify food facilities because classification is the main challenge presented by traditional data sets.

The remainder of this chapter is an interrogation of recent food access studies with some discussion of the methods used by each study. A brief discussion of USDA food access methods follows because the USDA is very involved in the discussion of food access and food deserts. The chapter concludes with a brief discussion of the need for improved classification methods and makes the case for testing AGI as a potential adjunct to traditional data.

## 2.1 Traditional Food Access Studies

Traditional food access studies use NAICS codes and commercial data in order to locate and interrogate food facilities in the study area. Helen Lee employs NETS data from Dunn and Bradstreet in one such study. (Lee 2012) She classifies facilities into five categories that are “primarily based on 6-digit NAICS codes, although in some cases the 8-digit SIC codes are used to refine the definition, as well as business name, trade name, employee size, and annual sales information.” (Lee 1196, 2012) Lee’s classifications include: supermarkets stores, corner stores, convenience stores, restaurants, and fast food restaurants. Though her method seeks to identify access to healthy food, it is difficult to discern quality from the classification. Consequently, her results focus on counts and densities for each of the categories.

An and Sturm employ similar methods in their study of food access in California. They source their data from InfoUSA; however, they too use NAICS codes when classifying facilities. An and Sturm sampled the data and used local knowledge to identify NAICS codes that represented different business classes:

Although there is no NAICS code for fast-food restaurants, 63 major fast-food franchises are identifıed with main menus containing items such as hotdogs, burgers, pizza, fried chicken, subs, or tacos under the NAICS codes 72221105-6. Convenience stores are identifıed as NAICS code 44512001, and small food stores (annual sales \_$1 million); midsize grocery stores (annual sales $1–$5 million); and large supermarkets (annual sales\_$5 million) are identifıed as NAICS codes 44511001-3. (An and Sturm, XXXX)

The NAICS codes that they use in their study include the two digit proprietary suffix as well as a third digit followed by a dash. Their methods do not explain where the additional data comes from, and a review of the InfoUSA FAQ did not yield any clues. It is possible that InfoUSA has done some work to extend the classification, and could be a potential source of improved facilities data.

Shier, An, and Sturm used similar methods in their paper “Is there a robust relationship between neighborhood food environment and childhood obesity in the USA?” Their classification method uses six digit NAICS codes and annual sales in order to classify facilities. (Shier et al., 2012) Much like Lee, their methods focus on counting facilities within their study area. In this case, Shier et al. calculate the percentage of census tracts that have at least on of each of the defined classes. Though annual sales can be used to estimate the size of a market, it is not necessarily a good indicator of the quality of goods that the market sells.

Zick et al. use Dun and Bradstreet data in their study of food environments and obesity. Their methods use SIC codes, the predecessor to NAICS codes, to classify facilities into four categories. There are corresponding NAICS codes for each SIC code, and Esri Business Analyst data includes both codes for each facility listed. Zick et al. then identify whether or not each block group has a single type of facility or some mix of facilities in order to define food access for a given place. (Zick et al., 2009) Like the previously discussed studies, it is difficult to assess the quality of food available given this classification method.

## 2.2 USDA Food Access Estimates

The US Department of Agriculture has published a report about healthy food access in the United States. The USDA combines a list of stores that accept SNAP with commercial data from Trade Dimensions TDLinx in order to identify retail food facilities that offer a wide range of products. (USDA, 2012) The USDA uses square footage and annual sales to select and classify outlets into three classes: super-center, super market, and grocery store. Though their method works around the NAICS selection problem, it admittedly excludes numerous smaller businesses that sell healthy food. Moreover, size and sales are not necessarily a good indicator of the quality of food being sold by a given facility.

## 2.3 Classification and AGI

The classification methods discussed in this chapter do not provide a direct representation of the quality of the food provided by a given facility. They instead classify facilities with proxies that attempt to identify the quality of food available. Other methods classify facilities by using the local knowledge of the author; however, this method does not scale well.

Food access studies would benefit greatly from an improved classification method that represents the quality of food available more directly. Google Places and Yelp both provide access to their data which includes classification, ratings, and reviews. The additional data could be used instead of, or in conjunction with, traditional commercial data in order to generate a more robust and nuanced measure of food access. The remainder of this thesis investigates data from Google and Yelp in order to take a step toward understanding whether or not AGI is a useful adjunct to commercial data.

# Chapter 3 Methods

## 3.1 Methods

This methods employed by this study required initial data aggregation, site selection, Python development, in filed data collection, and analysis. This allowed for the collection and comparison of data, ambient geographic information, and observed data in the study area. Government data from the USDA and US Census were first collected in order to identify the study area.

## 3.2 US Census Data

A census tract shape file and tabular demographic data were downloaded from the US Census website for use in ArcMap. The shape file included tracts for the entire state of California, and was first pared down for ease of processing. Select by attribute with COUTYFPS = ‘037’ was used to reduce the overall number of tracts that had to be processed in ArcGIS during the site selection process.

Fields that represent white population, total population, aggregate number of cars, inhabited homes, and median income were identified. The object id, geoid, B00XXX, and BXXXXXX were selected in the X01\_Sex\_and\_Age table, object id and BXX was selected in the X19\_Income table, and finally, object id, BXXX and BXX were selected from X25\_Household\_Characteristics. The selection was necessary in order to reduce the overall size of the data set, which reduces processing time.

The site selection method used is derived from site selection methods that use the raster calculator. It was necessary to retain the data in vector format because the fishnet created by raster pixels would not line up neatly with the irregular boundaries of the census tracts. Consequently, a method was adopted that produces an aggregate score for each tract, which is subsequently visualized with a choropleth map. The score is a summation of the percent white population, car access, and median income.

Some data manipulation and cleaning was necessary in order to facilitate the calculation of scores for the census tracts. Specifically, fields form several different US Census tables needed to be aggregated into a single table before the score could be calculated. Consequently, the selected data was exported from each table into a comma separated values file, which was subsequently imported into Microsoft SQL Server. The import resulted in three tables that were aggregated into a single view, joined on the geoid. The resulting view was opened in arc map and saved in the original file geodatabase for subsequent processing.

Fields were added for percent white population, car access, median income , log(median income), and overall score. All of the fields were of type double. Any row that had missing data was removed because it would not be useful for the purposes of this study. Finally, the field calculator was used to populate the calculated fields. Percent white population was calculated by dividing the white population by the total population. Car access was computed by dividing aggregate car by inhabited homes. The log of median income was computed in order to be included in the final score. The value of the final score is the summation of the three previously computed fields.

The resulting tabular data was joined with the census tracts shape file in ArcMap on the geoid field. The data for the final score was visualized by creating a choropleth map of the census tracts in Los Angeles County. The tracts were symbolized by quantity and classified by standard deviation. The resulting map, figure 2, was then employed to visually identify tracts that would be useful for this study.

Census tracts with large standard deviations from the mean were visually identified. Tracts with significant deviations were more thoroughly investigated for inclusion in the study. Facilities data was included form Esri Business Analyst in order to determine whether or not a tract was a good candidate for inclusion in the study. This was necessary because tracts without facilities would not be useful for the purposes of this study.

Businesses within 1500 meters of the mean center of each tract were selected from each of the three data sources and visualized. 1500 meters was chosen because it produced a sufficient number of businesses to be compared. The radius also helps to mitigate edge effects that would occur if the facilities were clipped to the tract boundary.

Census tract 224020, figure 3, was identified as a low income area with limited car access and high minority population within the Los Angeles study area, and will be used in this study. This tract is 33% white, has a median income of $22042, and has .72 cars per household.

Census tract 460700, figure 4, was selected because it has contrasting characteristics. The area is 71% white, has a median income of $177578, and 2.6 cars per household.

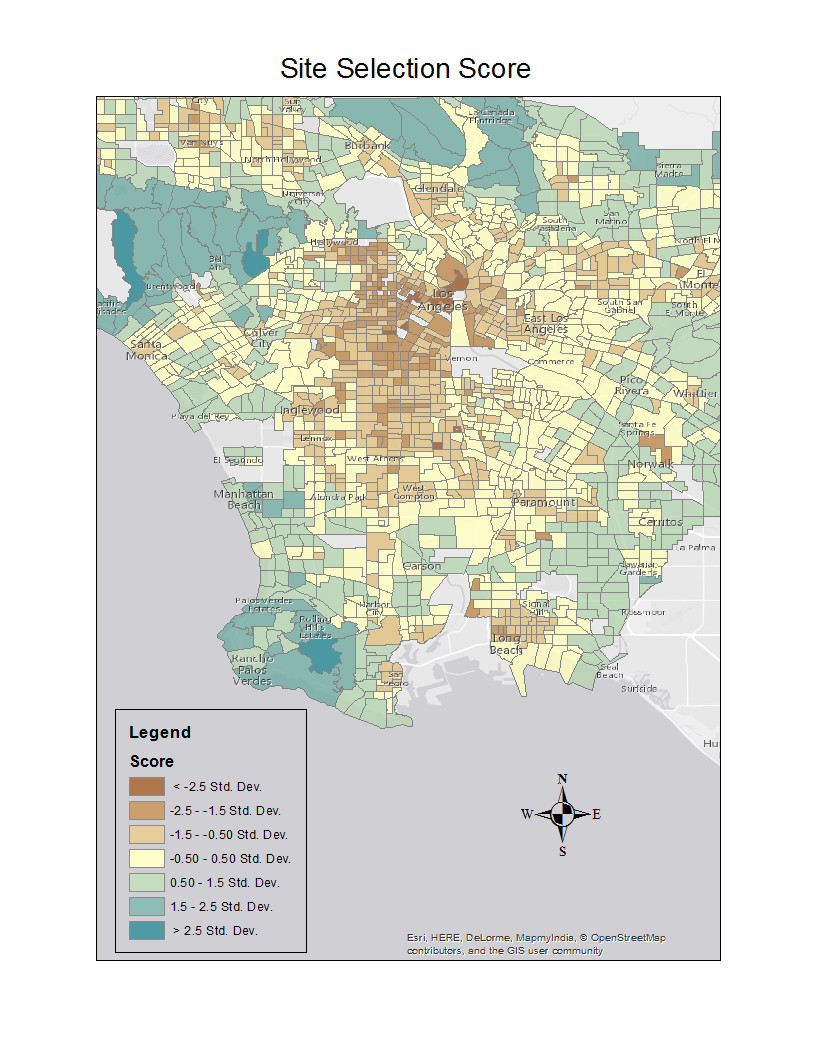


Figure 2

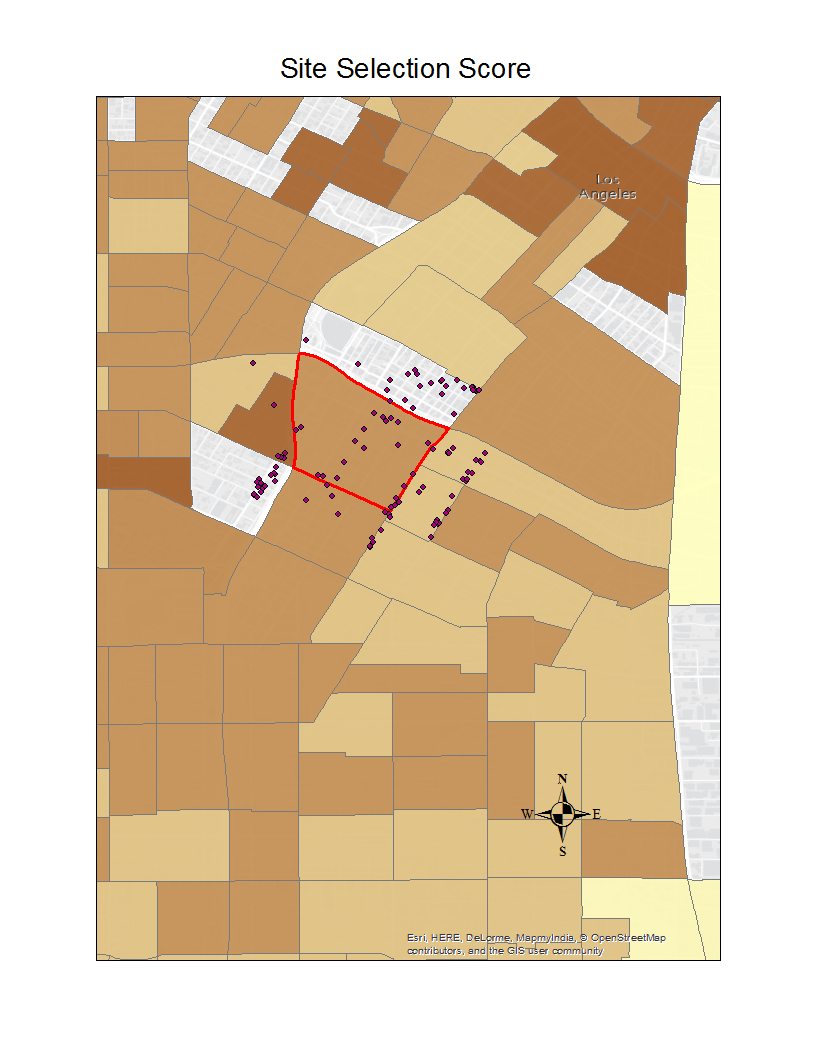


Figure 3

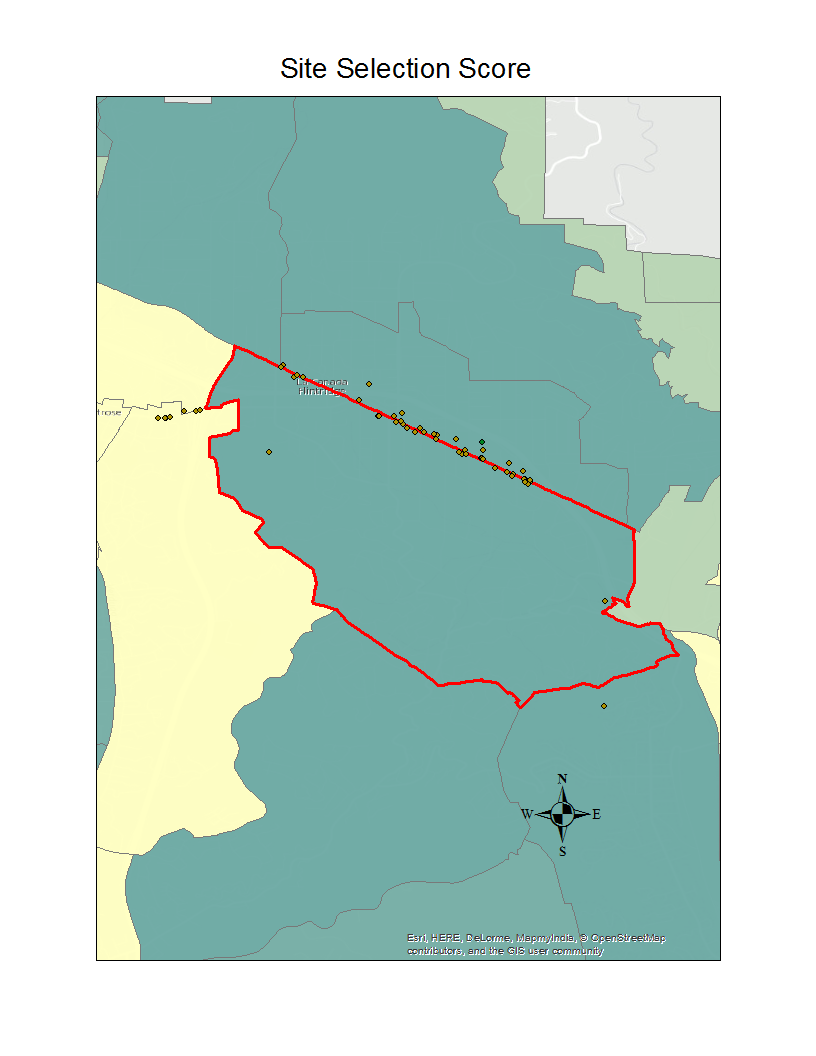


Figure 4

## 3.3 Esri Business Analyst Data

Esri Business Analyst was used as a source of commercial data for this project. Data was selected by NAICS codes for both markets (445%) and restaurants (7225%). Business Analyst produced a shapefile containing point data for the businesses within the greater Los Angeles area. Businesses within 1500 meters of the mean center of each tract were selected. The selection resulted in 33 grocery businesses in tract 460700 and 91 grocery businesses in tract 224020.

The join was performed in ArcMap with the select by location tool. Businesses that were within each tract were selected, and a layer was then created from the selection. The process resulted in two layers, one for each tract, that contained all of the Esri Business Analyst data for each tract.

## 3.4 Ambient Geographic Information

Ambient geographic information (AGI) was collected from both Google Places and Yelp for both of the census tracts in the study. Python code was used to access and record the data from both API’s. The data was subsequently imported into ArcMap and point data was derived from the latitude and longitude provided in the results.

### 3.4.1 Computing environment

The Anaconda package was used to install Python 2.7 onto a Windows computer; however, Anaconda was chosen because it is compatible with Windows, Mac, and Linux. Both API’s can be accessed with any programming language that can authenticate and make a secure http request. Python was used in this case because of the Arcpy integration with ArcGIS. Moreover, Yelp provides a Python client library that simplifies access to their API. The library is installed on the command line with pip with the following syntax: pip install yelp. The Yelp API requires OAUTH authentication, and the client library simplifies the process.

### 3.4.2 Google Places API

The Google Places API is accessed via secure http request and returns paginated data. The API requires a client key to be passed with each https request as part of the query string. Python’s native httplib was used to establish a connection with the Google API and collect results. A query string with all of the necessary components was assembled with concatenation operators in the Python code, and then passed to the API by httplib.

In addition to the key, the API request requires a location and a business category. The location is a latitude and longitude in an unprojected coordinate system (WGS84) . The ArcGIS find mean center tool is used to identify the center of each census tract, which is then used as a location input for the API call. The categories are provided in tabular format in the API documentation. The restaurant and market categories were used for the purposes of this study.

The Google Places API returns twenty items per page along with a next page token. The Python code written for this study wraps the API call inside of a function that can be recursively called with the next page token in order to aggregate the data from multiple pages into a single data set. The API query string is concatenated and sent to the Googe API server. The HTTP response code is checked, and if it is 200 OK, the response data is loaded into a data structure and parsed into JSON format. An iterator is then used to step through the JSON and output the latitude, longitude, business name, and classification are written in CSV format. The CSV can then be opened in ArcMap and the plot XY feature can be used to create point features that represent each business.

### 3.4.3 Yelp API

The Yelp API requires OAUTH1 authentication prior to being queried. Yelp provides a client library for Python that greatly simplifies the authentication process. Consequently the client library was employed in the code written for this study. The installation of the client library via pip is a prerequisite to executing any of the Yelp code provided in this paper.

The Yelp API returns twenty businesses per function call, and uses an offset integer in order to access additional pages of data. The response data includes the total number of records returned. Consequently, and iterator is used to call the function recursively, incrementing the offset by twenty each time, until the offset is greater then the total. The results are then aggregated into csv format so that they can be opened in ArcMap and point features can be created from the plotXY function.

### 3.4.4 In field evaluation of facilities

Food facilities identified in the previous steps will be visited and evaluated. A worksheet, appendix 2, will be employed in order to ascertain the quality and cost of food in each of the facilities within the study area.

The results of the field investigation will be used to assess the quality of the results obtained through traditional methods. <This needs more development>

# REFERENCES

An, Ruopeng, and Roland Sturm. 2012. "School and residential neighborhood food environment and diet among california youth." *American Journal of Preventive Medicine 42 (2):* 129-35.

Boone, Christopher G. 2008. "Improving resolution of census data in metropolitan areas using a dasymetric approach: applications for the Baltimore Ecosystem Study." *Cities and the 1 (1):* Article 3,25 pp.

Briggs, David J., John Gulliver, Daniela Fecht, and Danielle M. Vienneau. 2007. "Dasymetric modelling of small-area population distribution using land cover and light emissions data. ." *Remote Sensing of Environment 108 (4):* 451-66.

Cummins, S., E. Flint, and SA Matthews. 2014. "New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity." *Health Affairs 33 (2):* 283-91.

Hay, S. I., A. M. Noor, A. Nelson, and A. J. Tatem. 2005. "The accuracy of human population maps for public health application." *Tropical Medicine & International Health 10 (10):* 1073-86.

Holt, James B., C.P. Lo, Thomas W. Hodler. 2004. "Dasymetric Estimation of Population Density and Areal Interpolation of Census Data." *Cartography and Geographic Information Science 31 (2):* 103 - 21.

Langford, Mitchel, Gary Higgs, Jonathan Radcliffe, and Sean White. 2008. "Urban population distribution models and service accessibility estimation." *Computers, Environment and Urban Systems 32 (1):* 66-80.

Langford, Mitchell. 2006. "Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method." *Computers, Environment and Urban Systems 30 (2):* 161-80.

Langford, Mitchell. 2007. "Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. ." *Computers, Environment and Urban Systems 31 (1):* 19-32.

Lear, Scott A., Danijela Gasevic, and Nadine Schuurman. 2013. "Association of supermarket characteristics with the body mass index of their shoppers." *Nutrition Journal 12 (1):* 117-.

Lee, Helen. 2012. "The role of local food availability in explaining obesity risk among young school-aged children." *Social Science and Medicine 74 (8):* 1193-203.

Mennis, Jeremy, and Torrin Hultgren. 2006. "Intelligent dasymetric mapping and its application to areal interpolation." *Cartography and Geographic Information Science 33 (3):* 179-94.

Mooney, Peter, Padraig Corcoran, and Blazej Ciepluch. 2013;2012;. "The potential for using volunteered geographic information in pervasive health computing applications." *Journal of Ambient Intelligence and Humanized Computing 4 (6):* 731-45.

Shier, V., R. An, and R. Sturm. 2012. "Is there a robust relationship between neighbourhood food environment and childhood obesity in the USA?" *Public Health 126 (9):* 723-30.

Zick, Cathleen D., Ken R. Smith, Jessie X. Fan, Barbara B. Brown, Ikuho Yamada, and Lori Kowaleski-Jones. 2009. "Running to the store? the relationship between neighborhood environments and the risk of obesity." *Social Science & Medicine 69 (10):* 1493-500.

# Appendix A – Code

#-------------------------------------------------------------------------------  
# Name: Google Places API Nearby Request  
# Purpose: USC SSI Master's Thesis Project  
# Food access in Los Angeles  
# Author: CMH  
#  
# Created: 06/03/2016  
# Copyright: (c) CMH 2016  
# Licence: Attribution-NonCommercial-ShareAlike 4.0 International  
#-------------------------------------------------------------------------------  
  
# Import all of the libraries  
import httplib  
import urllib  
import json  
import pprint  
  
#Yelp Python Client Support  
from yelp.client import Client  
from yelp.oauth1\_authenticator import Oauth1Authenticator  
  
auth = Oauth1Authenticator(  
 consumer\_key='NLz9tZJ465VfGE1ZF4wSvw',  
 consumer\_secret='hPmTpTwFLNskvTNO-lu4HUNskMo',  
 token='HLBajJ7AO9-EVkz-eRiaD0DDfFz4kE1q',  
 token\_secret='RLEhQPqvobQ9umvCKiCN\_wJxFHc'  
)  
  
yelpClient = Client(auth)  
  
  
def yelpit( lat, lng, types, offset ):  
 #This funciton pulls business data from the Yelp API  
  
 params = {  
 'term': 'food',  
 'radius\_filter': '1500',  
 'offset': offset,  
 'sort': '0',  
 'category\_filter': types  
 }  
 response = yelpClient.search\_by\_coordinates(lat,lng,\*\*params)  
 #print response.total  
 for key in response.businesses:  
 print key.location.coordinate.latitude, ",",  
 print key.location.coordinate.longitude, ",",  
 print key.name, ",",  
 print key.categories[0][1], ",",  
 print key.rating, ",",  
 print key.review\_count  
  
 offset += 20  
 #print offset, "\n"  
 if offset < response.total:  
 yelpit(lat, lng, types, offset)  
  
def deets(bus\_id):  
 conn = httplib.HTTPSConnection("maps.googleapis.com")  
 APIkey = "AIzaSyB864Llir0-1NrFaQ1yr3TIzG9fB09IP7c"  
 reqstring = "/maps/api/place/details/json?placeid=" + bus\_id + "&key=" + APIkey  
 #print reqstring  
 conn.request("GET", reqstring)  
 response = conn.getresponse()  
 #print response.status, response.reason  
 if response.status == 200:  
  
 # Get and print the actual data  
 data = response.read()  
  
 # parse the json into a more useful data structure  
 parsed\_json = json.loads(data)  
 #pp = pprint.PrettyPrinter(indent=4)  
 #pp.pprint(parsed\_json)  
 if "rating" in data:  
 print parsed\_json['result']['rating'], ",",  
 else:  
 print "0,",  
 if "user\_ratings\_total" in data:  
 print parsed\_json['result']['user\_ratings\_total']  
 else:  
 print "0"  
  
 conn.close()  
  
def googit(lat, lng, types, next\_page\_token = None):  
 conn = httplib.HTTPSConnection("maps.googleapis.com")  
  
 #headers = {"":""}  
 #Lat Long for center of census tract  
 #lat = "34.0465960"  
 #lng = "-118.2515835"  
  
 radius = "1500"  
 #types = "restaurant"  
 #types = "grocery\_or\_supermarket"  
 APIkey = "AIzaSyB864Llir0-1NrFaQ1yr3TIzG9fB09IP7c"  
  
 if next\_page\_token is None:  
 conn.request("GET", "/maps/api/place/nearbysearch/json?location=" + lat + "," + lng + "&radius=" + radius + "&types=" + types + "&key=" + APIkey)  
 else:  
 conn.request("GET", "/maps/api/place/nearbysearch/json?location=" + lat + "," + lng + "&radius=" + radius + "&key=" + APIkey + "&pagetoken=" + next\_page\_token)  
  
 # Get the response and print the response information eg. 200 OK or 404 Not Found  
 response = conn.getresponse()  
 #print response.status, response.reason  
  
 if response.status == 200:  
  
 # Get and print the actual data  
 data = response.read()  
  
 # parse the json into a more useful data structure  
 parsed\_json = json.loads(data)  
  
 # Load the pretty printer so that we can better see the structure of the data  
 #pp = pprint.PrettyPrinter(indent=4)  
 #pp.pprint(parsed\_json)  
 #pp.pprint(parsed\_json['pagination'])  
 #pp.pprint(parsed\_json['meta'])  
 #pp.pprint(parsed\_json['results'])  
 #pp.pprint(data)  
 #json.dumps( parsed\_json, sort\_keys=True, indent=4, separators=(',', ': ') )  
  
 items = parsed\_json['results']  
 for item in items:  
 #theLine = ""  
 try:  
  
 #print item['place\_id']  
 #pp.pprint(item)  
 print item['geometry']['location']['lat'], ",",  
 #theLine = theLine + item['location']['latitude'] + ","  
 print item['geometry']['location']['lng'], ",",  
 #theLine = theLine + item['location']['longitude'] + ","  
 print item['name'], ",",  
 print item['types'][0], ",",  
 deets(item['place\_id'])  
 #print ""  
 #theLine = theLine + item['link'] + ","  
 #print item['images']['standard\_resolution']['url']  
 #theLine = item['location']['latitude']  
 #theLine = theLine + item['location']['latitude'] + "," #+ item['location']['longitude'] + "," + item['link'] + "," + item['images']['standard\_resolution']['url'] + "\n"  
 #print theLine  
 #print (",")  
 #print "\n"  
 #f.write(theLine)  
 except TypeError:  
 print ",type error"  
 pass  
 except KeyError:  
 print ",key error"  
 pass  
 # Close the connection  
 #f.close()  
 conn.close()  
  
 if "next\_page\_token" in data:  
 #print "recurse"  
 #print parsed\_json['next\_page\_token']  
 googit(lat,lng,types,parsed\_json['next\_page\_token'])  
  
print "lat, long, name, type, rating, review\_count"  
#googit("34.1929284","-118.1988009","grocery\_or\_supermarket");  
#yelpit("34.1929284","-118.1988009","grocery,convenience",0);  
  
googit("34.0304827","-118.2686569","grocery\_or\_supermarket");  
#yelpit("34.0304827","-118.2686569","grocery,convenience",0);

# Apendix B - Market Evaluation Worksheet