FOOD DESERTS BY ZIPCODE
DATA SOURCES

How Big is the Problem in Three Largest U.S. Cities?

Los Angeles

Chicago

New York

Data Sources for Food Deserts by Zip Code

- There are three sets of data needed for this analysis
  - Income by Zip Code
  - Zip Codes in target cities
  - Grocery Stores near a Zip Code

The next pages explain these data sources and their use.

Data Sources – Income by Zip Code

- For income by zip code, this analysis will be using the 2017 income tax file available at https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2017-zip-code-data-soi. The IRS data contains information by 6 major income groups for each zip code and indicates the individual line item information on U.S. Income Tax forms 1040, 1040A, 1040EZ. The information will be used to create an average income by zipcode for correlation and comparative purposes.

- The Data Dictionary for this file can be downloaded here. (This will open in a Word instance.) Some rows from the Data Dictionary are below:

VARIABLE NAME	DESCRIPTION	VALUE/LINE REFERENCE	Type
STATEFIPS	The State Federal Information Processing System (FIPS) code	01-56	Char
STATE	The State associated with the ZIP code	Two-digit State abbreviation code	Char
ZIPCODE	5-digit Zip code		Char
AGI_STUB	Size of adjusted gross income	1 = $1 under $25,000 2 = $25,000 under $50,000 3 = $50,000 under $75,000 4 = $75,000 under $100,000 5 = $100,000 under $200,000 6 = $200,000 or more	Num
N1	Number of returns		Num
MARS1	Number of single returns	Filing status is single	Num
MARS2	Number of joint returns	Filing status is married filing jointly	Num

Data Sources – Target City Zip Codes

- To retrieve zip codes for each city, this analysis will be using information from for www.zip-codes.com the three cities using the city names as keys for the data. The data will be scraped from the web site and refined as needed.

Screen shot of Los Angeles Zip Code data

Data Sources – Grocery Stores by Zip Code

- To identify grocery stores by Zip code, FourSquare data will be used.

- The following *categoryId* keys will be used to identify the grocery stores:

  - grocery_store = '4bf58dd8d48988d118951735'

  - organic_grocery = '52f2ab2ebcbc57f1066b8b45'

  - supermarket = '52f2ab2ebcbc57f1066b8b46'

- The range for the grocery stores will be 804 meters, approximately ½ mile. This is in line with the USDA's definition of reasonable grocery store access.

- Additional information about venue categories within FourSquare can be found here.

FOURSQUARE