# FOOD DESERTS BY ZIPCODE

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


Los Angeles

Chicago

New York

# FOOD DESERT ANLYSIS BY ZIP CODE FOR 3 MAJOR US CITIES

1. Introduction
2. Data Section
3. Methodology
4. Results
5. Discussion
6. Conclusion

# 1. INTRODUCTION

# 1.1 Introduction - Background

- Healthful diets have a significant positive effect on reducing the occurrence of heart disease, diabetes, and even cancer. Yet, healthful diets are dependent on access to healthy food, which is generally available in supermarkets and generally not available in alternative stores like convenient stores. If the grocery stores are too few and far between, something we call a food desert, the purchase of healthy diet food is confounded.
- The food desert concept was first discussed in the 1990s. The metaphor is built on a desert paradigm. The wandering person is in the desert and looking for water at an oasis. If the person cannot find the desert, the person's chance of survival are limited.
- In the food desert, the wandering person is in the neighborhood (the desert) looking for healthy food at a supermarket (the oasis). If the person cannot find the supermarket, his alternatives for healthy food purchases and therefore consumption are limited.
- Research has shown that when food deserts are found, either convenient stores or fast food chains will provide nutritional support, both of which marginalize nutrition consumptions.

# 1.2 Introduction - Problem

- Determine if Food Deserts are rampant based on Zip Code Analysis
- Use 3 large U.S. Cities
  - *Los Angeles* *(6,900 per mile$^2$)*
  - *New York* *(27,000 per mile$^2$)*
  - *Chicago* *(11,900 per mile$^2$)*

# 1.3 Introduction - Interest

- Agencies of Interest
  - *State Health Departments*
  - *Federal Health Departments*
  - *Local Health Departments (County, Municipality, City)*
  - *Independent Agencies – CDC, IFIC, NIH*
- Possible Actions
  - *Improve public transportations to stores*
  - *Implement Construction Incentives*
    - Tax Breaks
    - Subsidies
  - *Incent healthy food purchase*
    - Coupons
    - Rebates

# 2. DATA SECTION

# 2.1 Data Sources – Income by Zip Code

- For income by zip code, this analysis will be using the 2017 income tax file available at "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 zip code for correlation and comparative purposes.
- The Data Dictionary for this file can be downloaded "www.irs.gov/pub/irs-soi/17zpdoc.docx". 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<br>2 = $25,000 under $50,000<br>3 = $50,000 under $75,000<br>4 = $75,000 under $100,000<br>5 = $100,000 under $200,000<br>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 |

# 2.1 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

# 2.1 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‘
- Additional information about venue categories within FourSquare can be found at “developer.foursquare.com/docs/build-with-foursquare/categories/”.

FOURSQUARE

# 2.1 Data Sources – Population by Zip Code

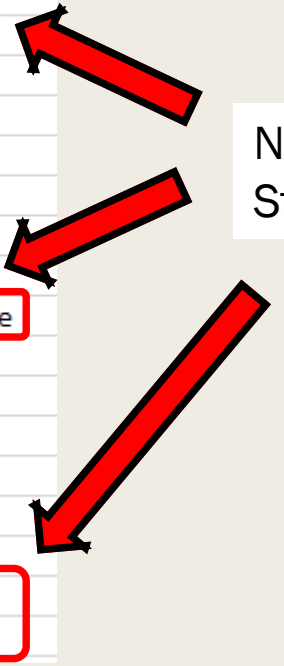
- To identify population used in calculating per capita income, a file developed by Splitwise will be used. It can be reviewed at https://blog.splitwise.com/2013/09/18/the-2010-us-census-population-by-zip-code-totally-free/.
- A sample of the data rows is below:

| Zip Code ZCTA | 2010 Census Population |
|---|---|
| 01001 | 16769 |
| 01002 | 29049 |
| 01003 | 10372 |
| 01005 | 5079 |
| 01007 | 14649 |
| 01008 | 1263 |
| 01009 | 741 |
| 01010 | 3609 |
| 01011 | 1370 |
| 01012 | 661 |
| 01013 | 23188 |
| 01020 | 29668 |
| 01022 | 2451 |
| 01026 | 946 |
| 01027 | 17660 |
| 01028 | 15720 |
| 01029 | 789 |
| 01030 | 11669 |

# 2.2 Data Cleaning – FourSquare Data

- Unexpected Results
  - *While the "category_Id" was grocery store, supermarket, and organic grocery were requested, the Venue categories returned included locations which were not desired.*

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Zip Code | Zipcode La | Zipcode Lo | Venue Name | Venue Latitude | Venue Longitude | Venue Category | |
| 135 | 90099 | 34.0522 | -118.244 | Best Market | 34.044483 | -118.244066 | Liquor Store | |
| 136 | 90189 | 34.0515 | -118.256 | Best Market | 34.044483 | -118.244066 | Liquor Store | |
| 137 | 90005 | 34.0585 | -118.301 | Beto's Market | 34.04338 | -118.30793 | Grocery Store | |
| 138 | 90006 | 34.0493 | -118.292 | Beto's Market | 34.04338 | -118.30793 | Grocery Store | |
| 139 | 90018 | 34.029 | -118.315 | Beto's Market | 34.04338 | -118.30793 | Grocery Store | |
| 140 | 90077 | 34.1112 | -118.45 | Bev Glen Market | 34.09963226 | -118.4438553 | Grocery Store | |
| 141 | 90033 | 34.0487 | -118.208 | Beverly Carniceria Market | 34.046595 | -118.205486 | Grocery Store | |
| 142 | 90063 | 34.0451 | -118.186 | Beverly Carniceria Market | 34.046595 | -118.205486 | Grocery Store | |
| 143 | 90077 | 34.1112 | -118.45 | Beverly Glen Marketplace | 34.12802195 | -118.4443044 | Grocery Store | |
| 144 | 90067 | 34.0551 | -118.41 | Beverly Hills Market & Deli | 34.06992522 | -118.3986681 | Sandwich Place | |
| 145 | 90031 | 34.0783 | -118.211 | Big Saver | 34.07342 | -118.210159 | Grocery Store | |
| 146 | 90031 | 34.0783 | -118.211 | Big Saver | 34.08474252 | -118.2222611 | Grocery Store | |
| 147 | 90032 | 34.0818 | -118.175 | Big Saver Foods | 34.08903086 | -118.1699286 | Grocery Store | |
| 148 | 90041 | 34.1339 | -118.208 | Bilo Market | 34.132717 | -118.202177 | Grocery Store | |
| 149 | 90033 | 34.0487 | -118.208 | Binky's Market | 34.052342 | -118.220622 | Grocery Store | |
| 150 | 90031 | 34.0783 | -118.211 | Bi-Rite Market | 34.074048 | -118.211104 | Grocery Store | |
| 151 | 90034 | 34.029 | -118.401 | Bob's Market & Liquor | 34.02886961 | -118.4032186 | Snack Place | |
| 152 | 90016 | 34.0298 | -118.353 | Bodega "R" Ranch Market | 34.03205589 | -118.3559542 | Deli / Bodega | |



Non-Grocery Store Entities

# 2.2 Data Cleaning – FourSquare Data

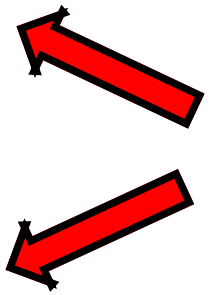
- Unexpected Results
  - ➢ *These categories were cleaned up by specificing selecting the desired Venue Categories*

```
In [25]: ## some of the stores have incorrect venue ids and need to be cleaned up.  Only retain the original 3 categories.

         NY_grocery_stores = clean_data(NY_grocery_stores_rough,"Y")              ## Change parameter to "Y" to print statistics

         ##NY_grocery_stores.head()                  ## DEBUGGING - uncomment line to see dataframe head
```

```
**********Original file venue names **********
Grocery Store                 6215
Supermarket                   2526
Deli / Bodega                  260
Gourmet Shop                   230
Convenience Store              160
Food & Drink Shop              138
Organic Grocery                131
Buffet                          48
Fruit & Vegetable Store         48
Farmers Market                  39
Bakery                          38
Butcher                         36
Health Food Store               34
Pharmacy                        33
Market                          19
Nutritionist                    14
Mexican Restaurant              12
Juice Bar                       12
```



Removed programmatically

# 2.2 Data Cleaning – Zip-codes.com

- Unexpected Results
  - *Some data retrieval grabbed incorrect states or cities. Shown below is an example of a search for Los Angeles which retrieved Puerto Rico and Texas data.*

# 2.2 Data Cleaning – Zip-codes.com

- Unexpected Results
  - *To find the data for New York city required using the individual boroughs which were listed as counties and selecting the data accordingly*


County used as a representation for Boroughs

# 2.3 Data Limitations

- The information for populations comes from the 2010 census. The Census Bureau has estimated population changes from 2010 to 2019 for New York, Los Angeles, and Chicago as 1.98%, 4.93%, and -0.06%. (See data tables found at https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html#tables.) The analysis for Los Angeles could be impacted by this increase. There may be growth in population in a zip code causing that area's average income to be overstated. This would cause an area to appear to be more affluent than it is actually is.
- The Census Bureau uses Zip Code Tabluation Areas (ZCTA) are slightly different in size from the true U.S. Postal Service Zip Codes. (A discussion of this can be found at http://gis.washington.edu/phurvitz/zip_or_zcta/index.html.) Because this analysis uses the center of a Zip Code area to determine distance, it is not believed the differences in boundaries will vary the information significantly.

# 2.3 Data Limitations – 2019 Census Calculations

| Geographic Area | April 1, 2010 | | Population Estimate (as of July 1) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Census | Estimates Base | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | |
| New York | 8,175,133 | 8,175,031 | 8,190,209 | 8,272,948 | 8,346,693 | 8,396,091 | 8,433,806 | 8,463,049 | 8,469,153 | 8,437,478 | 8,390,081 | 8,336,817 | 1.98% |
| Los Angeles | 3,792,621 | 3,793,139 | 3,795,512 | 3,820,876 | 3,851,202 | 3,881,622 | 3,909,901 | 3,938,568 | 3,963,226 | 3,975,788 | 3,977,596 | 3,979,576 | 4.93% |
| Chicago | 2,695,598 | 2,695,652 | 2,697,477 | 2,708,114 | 2,719,141 | 2,725,731 | 2,727,066 | 2,724,344 | 2,716,723 | 2,711,069 | 2,701,423 | 2,693,976 | -0.06% |

Based on Census Department projections and applied for 9 years to original 2010 data.

# 3. METHODOLOGY

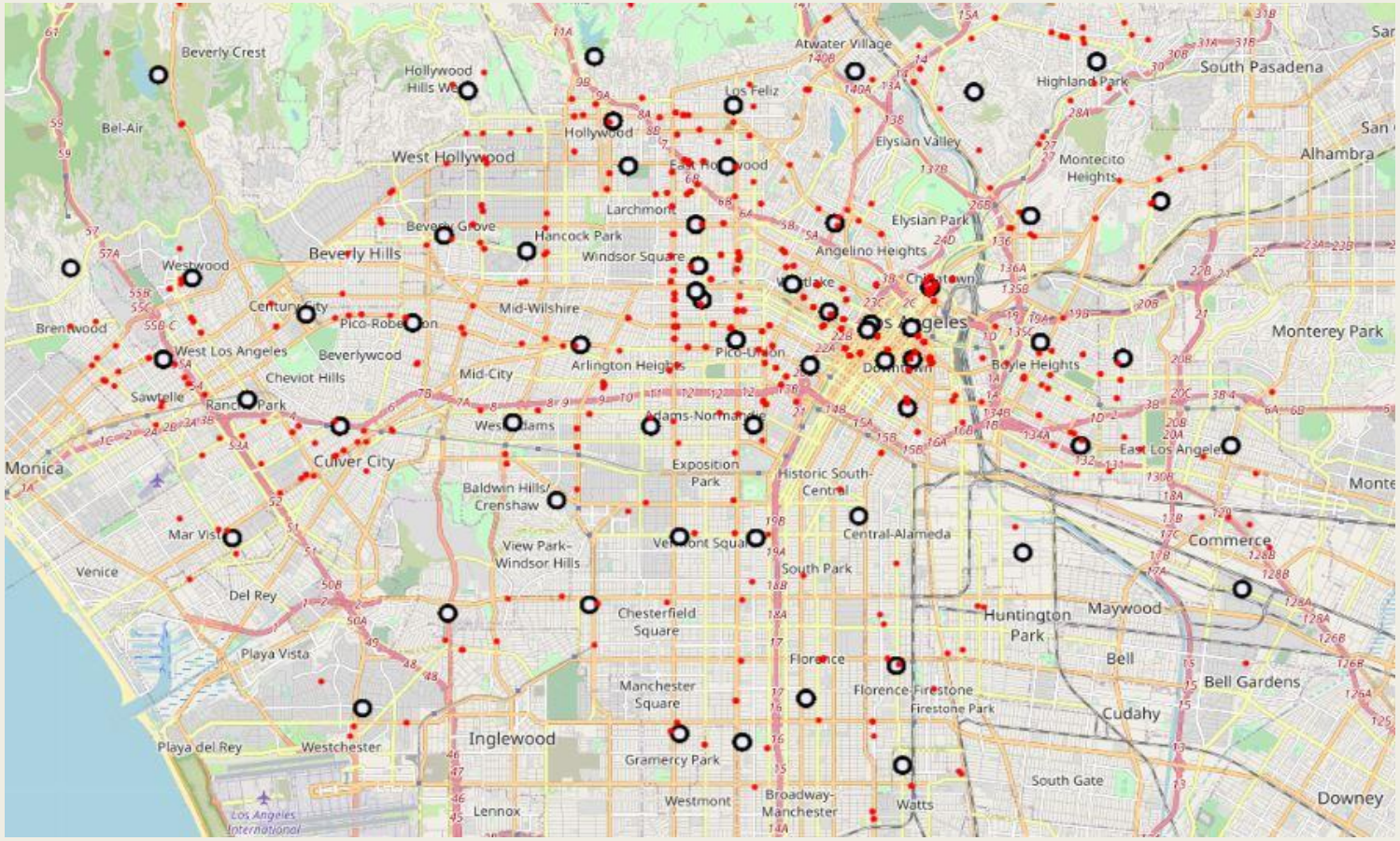
# 3. Methodology

- Income levels by zip code were determined by using the Average Gross Income line from the IRS tax data for each zip code. The census population was used to determine income by individuals in the area. This allowed family size to influence the calculation of the gross income as it accounts for additional expense associated with the support of dependents. In essence, a family of 2 making $35,000 is better off than a family of 5 making the same amount.
- Distance to grocery stores was considered as 1 mile. The distance measurements were from the longitude and latitude of the zip code center to the longitude and latitude of the grocery store.

# 4. RESULTS

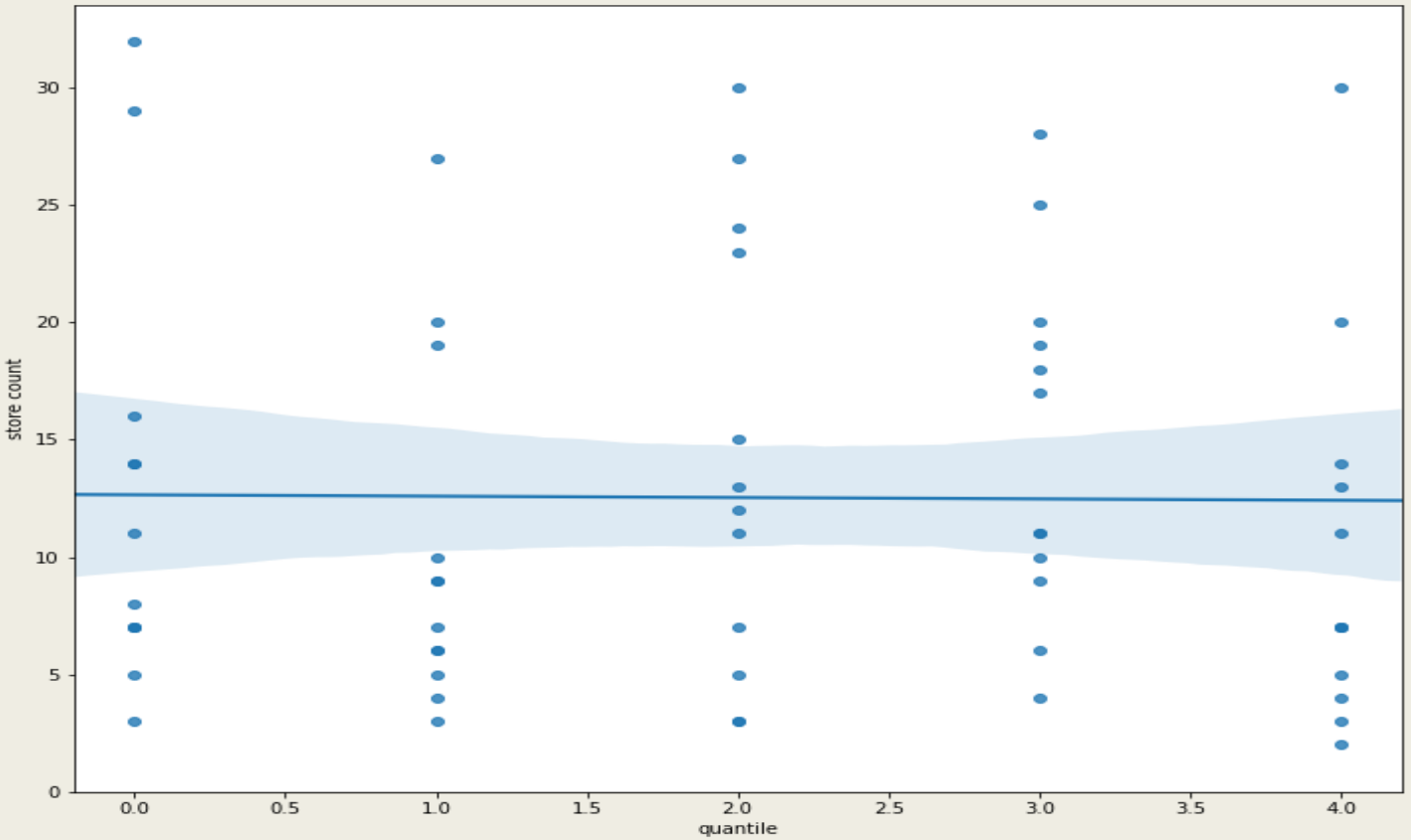
# 4.1 Conclusions - Los Angeles

Population Density is fairly low in LA at 6,900 per mile$^2$. This is revealed in the mapping of the grocery stores for each zip code area. Without significant concentration of population, a concentration of stores would be unsound economically.



# 4.1 Conclusions - Los Angeles

Not surprisingly then, and relative to the other cities, the concentration of grocery stores is low.  Even in the highest income level, access to stores is not particularly easy. Given the distribution across all income quantiles, the data struggle to characterize the lowest income quantiles as food deserts.



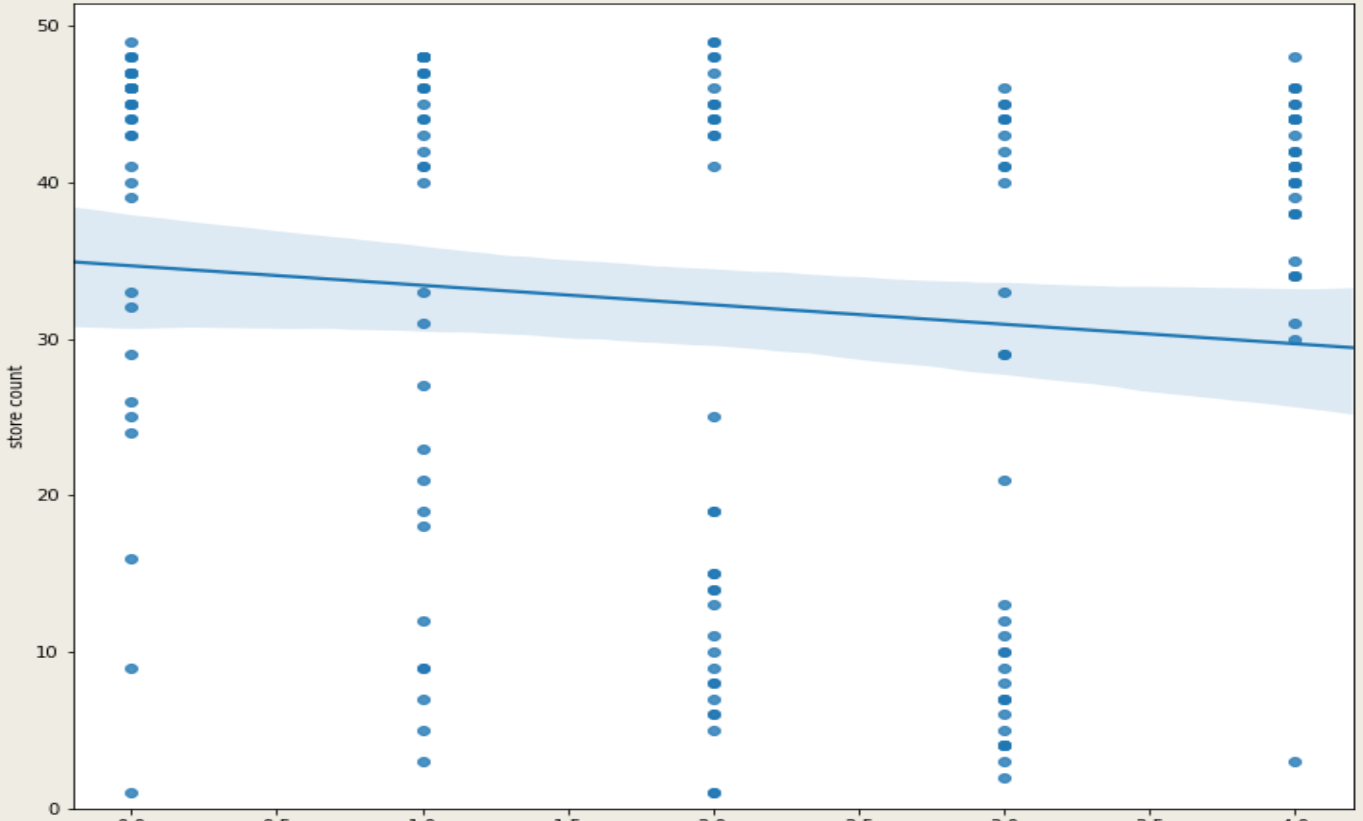
# 4.2 Conclusions – New York

Population Density is highest in NY at 27,000 per mile$^2$. The number of grocery stores shown in the map is plentiful in certain parts of the city such as Manhattan that is both significantly populated and affluent.

There are zip code areas with limited grocery stores such as on the peninsulas. With this level of density, the income divides become more muddled as both affluent and poor intersect side by side at times.

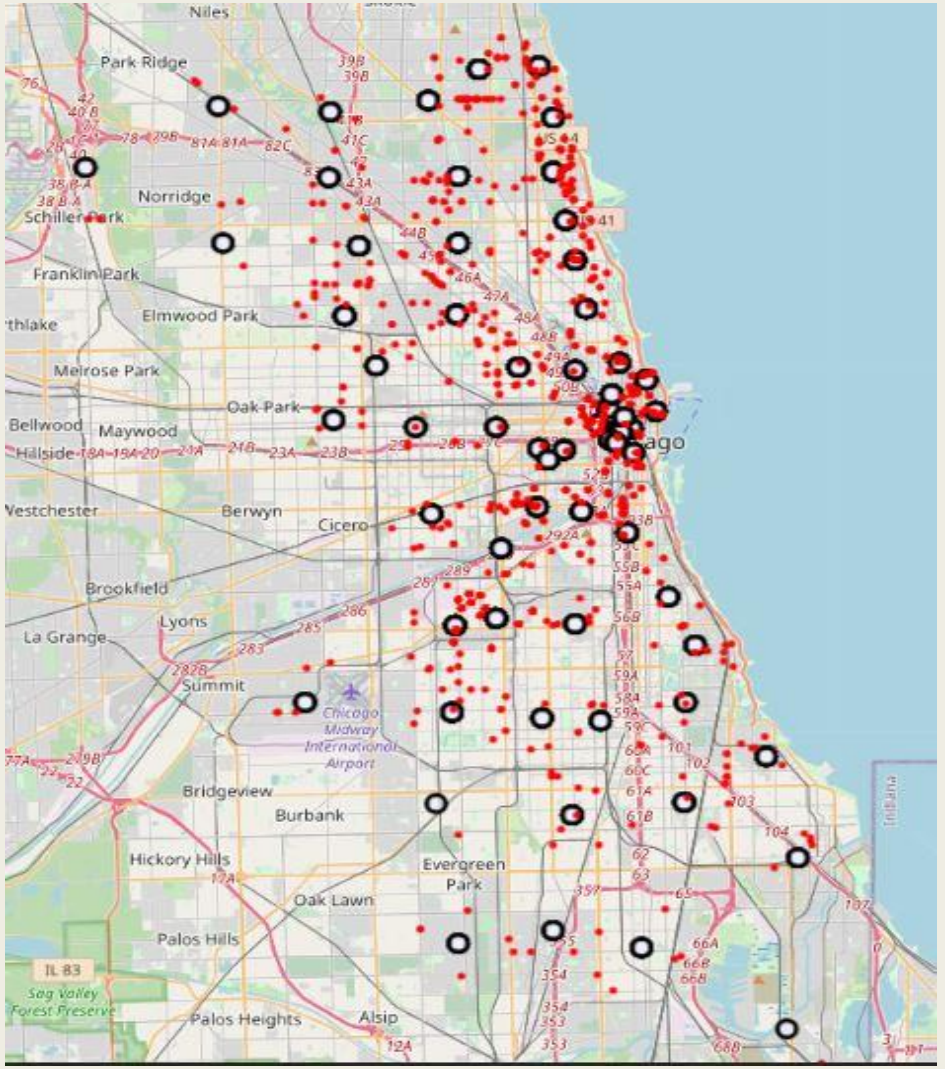
# 4.2 Conclusions – New York

Grocery store concentration shows only a few low income zip codes with accessibility problems.  While some upper income quantiles also show low cross over, these people are likely living in high rent and therefore the most dense areas.  Only a few grocery stores would be needed to support them.



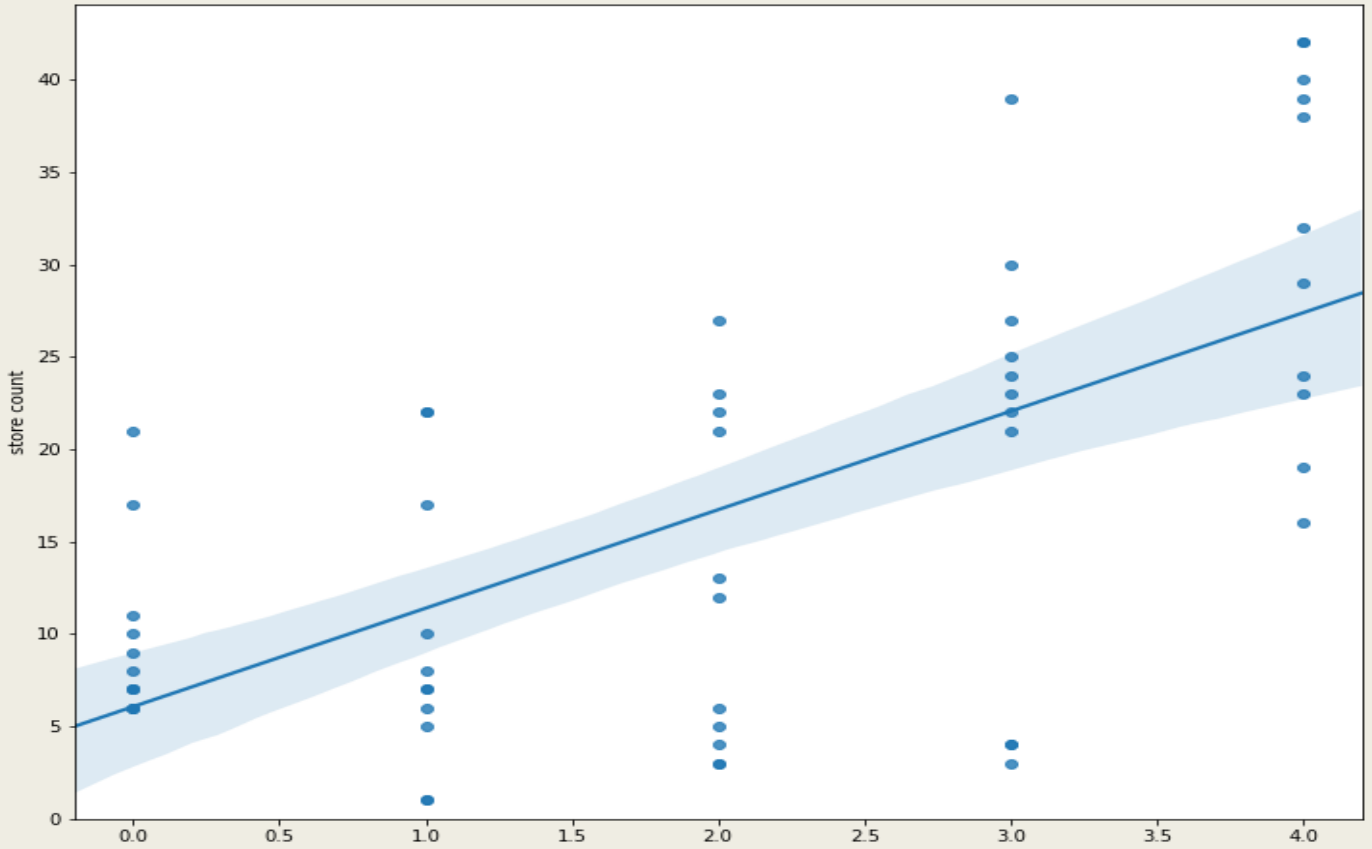
# 4.3 Conclusions - Chicago

Chicago's population density falls between LA and NY at 11,900 per mile$^2$. The sparsity of some of the grocer stores becomes apparent even in this depiction with the south and west sides of Chicago, noted for their lower income, showing much fewer grocery stores.



# 4.3 Conclusions - Chicago

Chicago shows the largest trend with food deserts appearing in the lowest income quantile and with significant access for all zip codes of the highest quantile.  Chicago then would appear to have the most problems with food deserts relative to the low income groups.



# 5. DISCUSSION

# 5. Discussion

- Food Desert research has self-constrained to only look at income levels and grocery store availability.  This research has clearly demonstrated that overall density may be an independent variable.  Results from at least two cities demonstrated lack of access equally for both high income and low income quantiles.  Additional research should be conducted to understand the general population density on the economics of grocery store construction which may operate independently of income levels.

# 6. CONCLUSIONS

# 6. Conclusions

- The largest cities in the U.S. vary in their food-desertness.
- New York has an inherent advantage to provide food access because of its significant population density, roughly 2 times that of Chicago, and 4 times that of Los Angeles. This density likely provides economic surety for the construction of new grocery stores.
- A city like Los Angeles, is so spread out as shown with the lowest population density shows all income quantiles having equal lack of access to food grocery stores. However, keep in mind more affluent areas can more easily support grocery stores, so this equivalency in access is unusual. Further research of the data however showed several grocery chains that were specific to heritage such as Vietnamese and Hispanic chains. Further research should determine if ethnic concentrations foster ethnic grocery store construction displacing what would be food deserts.
- Chicago has the highest correlation of grocery stores to income levels. There are clearly some food deserts and less affluent areas. Additionally, Chicago doesn't have the influx of ethnic groups driving a different economic. Further study should be conducted to determine what the impact is of ethnic change on the economics that support grocery store construction.