Nutrition in Chicago Neighborhoods

Andrew Leonard
Connor Binnig
Nic Carlson

Executive Summary

Utilizing food source (grocery store, urban farm, and farmer's market) and census, USDA, and Health Atlas datasets by community, determine areas in Chicago with poor or inadequate nutritional profiles and recommend solutions to those district's respective ward Alderman's to improve citizen nutrition.

ROADMAP:

- Clean data using OpenRefine
- Create database
- Normalize data
- Data exploration and categorization (e.g. for grocery stores, how will we categorize and define access)
- Carry out analysis of related neighborhood data, distilling possible reasons for food insecurity disparity
- Create visualizations
- Summarize findings

TASKS:

- Ranked list in SQL of neighborhoods that do not regularly consume fruits and vegetables
- Distill possible reasons: lack of access, lack of income,
- Tableau visualizations to support analysis of influences for food insecurity
- Actionable list of potential next steps

Business Case

CONTEXT:

- The USDA defines food insecurity as "a household-level economic and social condition of limited or uncertain access to adequate food.[1]"
- Chicago is a 1 in 7 Chicagoans will experience food insecurity in 2020[2]
- 'Food insecure households' are classified by three realities:[3]
    - They're worried food will run out before they get money for more.
    - The food they bought won't last, and they won't have money to get more.
    - They can't afford to eat balanced meals.

OBJECTIVE:

- Our project plans to measure the third bullet point by analyzing easy access to fruits and vegetables by Chicago neighborhoods
- Explore related data for potential reasons for lack of access, such as proximity to grocery stores and income levels

FINDINGS:

- Create visualizations that support potential conversations with Ward Alderman's/City of Chicago Department of Public Health
- Recommend viable solutions to Alderman districts with inadequate nutritional profiles and/or CPH

1. Link, 2. According to Feeding America's Map the Meal Gap Study, 3. Link

Data Tools

Open Refine
For data cleaning and preparation

MySQL
For building the database

Google Cloud SQL
For collaboration

Tableau
For analysis and visualization

Our Data Sources

Source	Dated	Description	Source link
Grocery Stores - Chicago Data Portal	2013	Contains a list of grocery stores in Chicago from 2013	Link
Chicago Health Atlas	2016-2018	Survey data includes obesity rates, access to fruits and vegetables, soda pop consumption by Chicago neighborhood.	Link
Per Capita Income - Chicago Data Portal	2007-2011	This dataset contains a selection of six socioeconomic indicators of public health significance and a “hardship index,” by Chicago community area, for the years 2007 – 2011.	Link
Urban Farm Data	2017	The location of select urban farms throughout Chicago.	Link
USDA food access/desert data by census tract (2017)	2010-2017	Provides food access data for populations within census tracts	**Link**
Farmers Market	2013	Location of Chicago’s farmers markets by neighborhood.	**Link**

Design Considerations

Model/method considerations

Sought out data to fit the business requirements. Main outputs were visualization of nutritional trends in Chicago and statistics linked to aldermen for community members to reference.

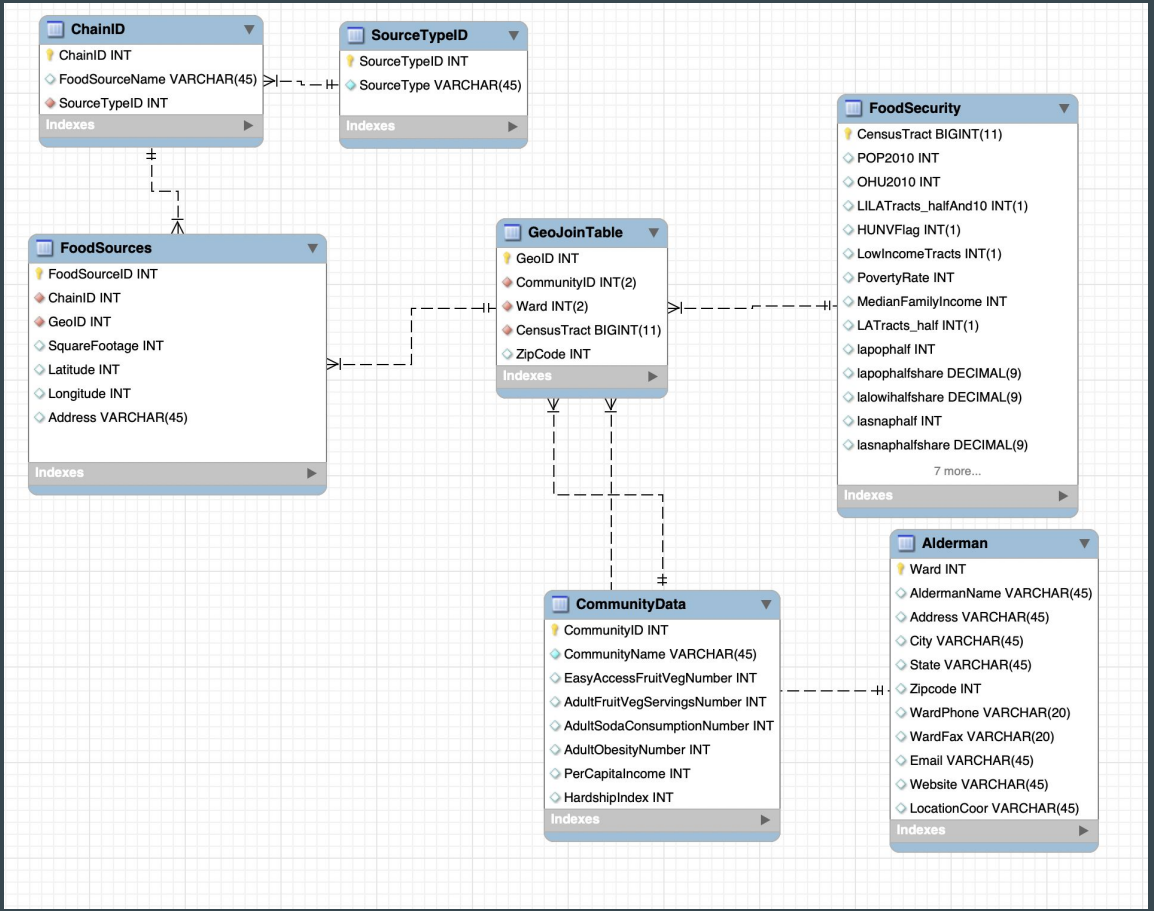
Model type: We ended up using a SQL relational model with a join table to solve the many to many relationship issues, and it also fit our needs since our data is highly normalized and entity based. We hosted our database on Google Cloud SQL for easier collaboration.

In our model, we used GeoID as the primary keys to connect most tables. Since most data is numbers, we used mostly INT as parameter.. Some of the attributes are restricted to number of digits, such as zip code (5) and CensusTract (11), and a few allow nulls, such as square footage of food sources due to lack of data

Granularity:

- Initially chose to look at data by neighborhood and ward, but issues with data forced us to look at a deeper level (census tracts).
- Decided to limit the scope of the project to the city of Chicago since most databases use it as the cutoff.
- For food security stats, we chose data that would be appropriate for city. For example, limiting definition of food security to food 'within ½ mile' as opposed to 1 mile (realistic walking distance for groceries)

Model Overview



Database Table Contents

Table Name	Cardinality	Table Content
CommunityData	1:M	Health survey data by Chicago neighborhoods.
FoodSecurity	1:M	Food access, median income, and race metrics for populations within census tracts.
Alderman	1:M	Contact information for each Alderman by Ward.
FoodSources	M:1	Coordinates, address, square footage of all Chicago grocery stores, urban farms, and farmers markets.
GeoJoin	M:1	Join table with GeoID as primary key.
ChainID	1:M	Foodsource name lookup table.
SourceTypeID	1:M	Source type lookup table.

Extraction, Transformation & Loading

ETL - Data Extraction

CSV extraction from the following sources:

FoodSecurity

Grocery Stores

Urban Farm Data

Farmers Market

FoodSources

USDA food access/desert data by census tract

CommunityData

Chicago Health Atlas

Hardship Index

Alderman

Ward Offices

ETL - Transformation

Data cleansing and manipulation

Food sources - wanted to put all food sources together in one table so that we can visualize easily

- Used OpenRefine clustered and renamed retailers
- Created food source ID and chain-ids for distinct retailers
- Added food source IDs to differentiate.
- Farmer Markets - fix addresses and named based on neighborhood
- Lookup census tracts using Census geocoder website
- Clean/union urban farm/farms markets tables with grocery stores
    - Removed records outside Chicago proper

GeoJoin table:

- Removed any geographies outside city of Chicago proper
- Created GeoID to give each distinct census tract / ward combination its own identifier
- Used census master tract file to pull in zip codes

Food security - removed unnecessary columns

Normalized tables, with central join table being geography

CommunityData - Joined together using CommunityID as primary key

AldermanData

- Added contact data where missing

ETL - Loading

MySQL Workbench to reverse engineer and generate DDL SQL scripts. Import wizard to write DML SQL scripts. Database uploaded to our Google Cloud SQL server.

MySQL Workbench
Run DDL & DML Scripts

Google Cloud SQL

ETL - Loading

Sample SQL DDL script:

```
126  DROP TABLE IF EXISTS `FoodSources`;
127  CREATE TABLE `FoodSources` (
128    `FoodSourceID` int(11) NOT NULL,
129    `ChainID` int(11) NOT NULL,
130    `GeoID` int(11) NOT NULL,
131    `SquareFootage` int(11) DEFAULT NULL,
132    `Latitude` int(11) DEFAULT NULL,
133    `Longitude` int(11) DEFAULT NULL,
134    `Address` varchar(45) DEFAULT NULL,
135    PRIMARY KEY (`FoodSourceID`),
136    KEY `fk_FoodSources_ChainID1_idx` (`ChainID`),
137    KEY `fk_FoodSources_FoodSources_has_TractDemographics1_idx` (`GeoID`),
138    CONSTRAINT `fk_FoodSources_ChainID1` FOREIGN KEY (`ChainID`) REFERENCES `chainid` (`ChainID`),
139    CONSTRAINT `fk_FoodSources_FoodSources_has_TractDemographics1` FOREIGN KEY (`GeoID`) REFERENCES `geojointable` (`GeoID`)
140  ) ENGINE=InnoDB DEFAULT CHARSET=utf8;
141
```

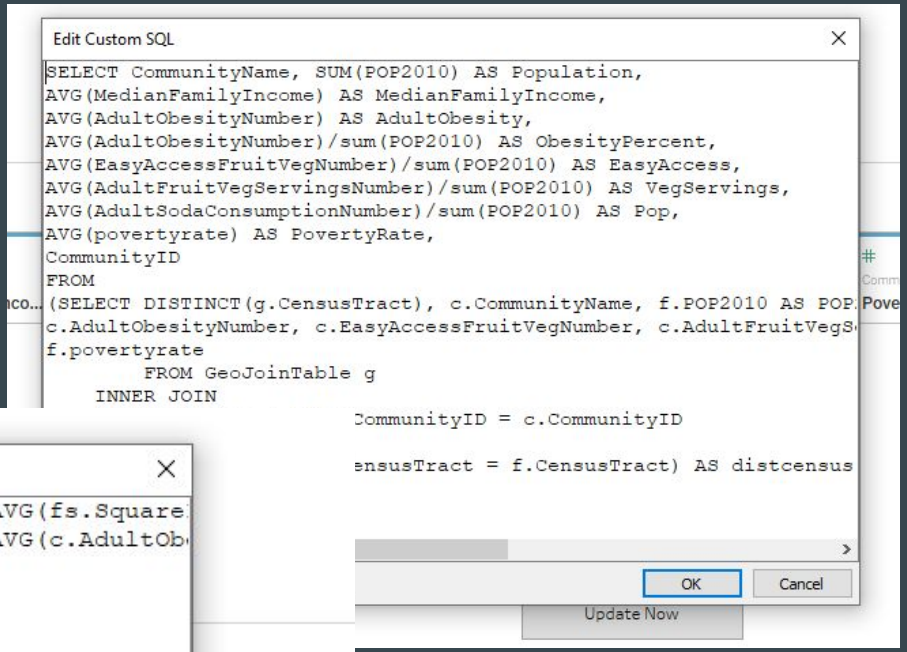
ETL - Loading

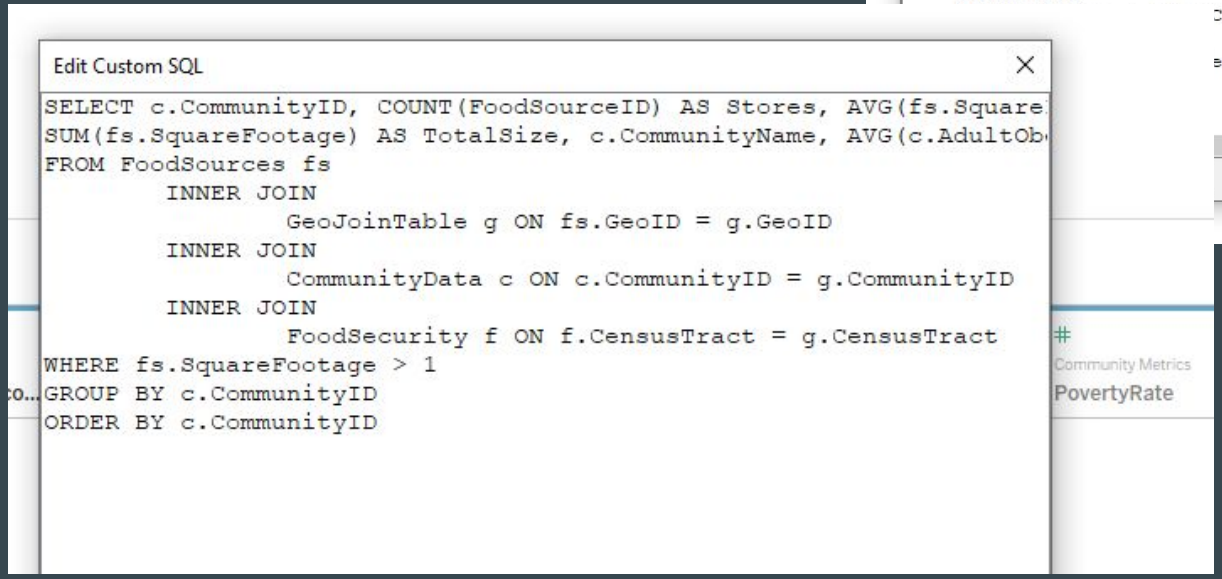
Sample SQL DML script.

```
INSERT INTO `FoodSources` VALUES (1,47,707,NULL,42,-88,'10900 S DOTY AVE W'),
(2,50,1129,NULL,42,-88,'600 W 35th St'),
(3,51,610,NULL,42,-88,'5301 S Hyde Park Blvd'),
(4,52,1095,NULL,42,-88,'300 N Central Park Ave'),
(5,53,1121,NULL,42,-88,'233 S WACKER DR'),
(6,54,1019,NULL,42,-88,'601 S Central Ave'),
(7,55,1017,NULL,42,-88,'1838 W 95th St'),
(8,56,155,NULL,42,-88,'705 W Armitage Ave'),
(9,57,464,NULL,42,-88,'5610 W Lake St'),
(10,58,163,NULL,42,-88,'39 W Division St'),
(11,59,648,NULL,42,-88,'7844S S Farragut Dr'),
```

Findings

(switch to Tableau)





Suggest researching 3 levers to achieve healthier Chicago

Improve access to existing healthy food sources

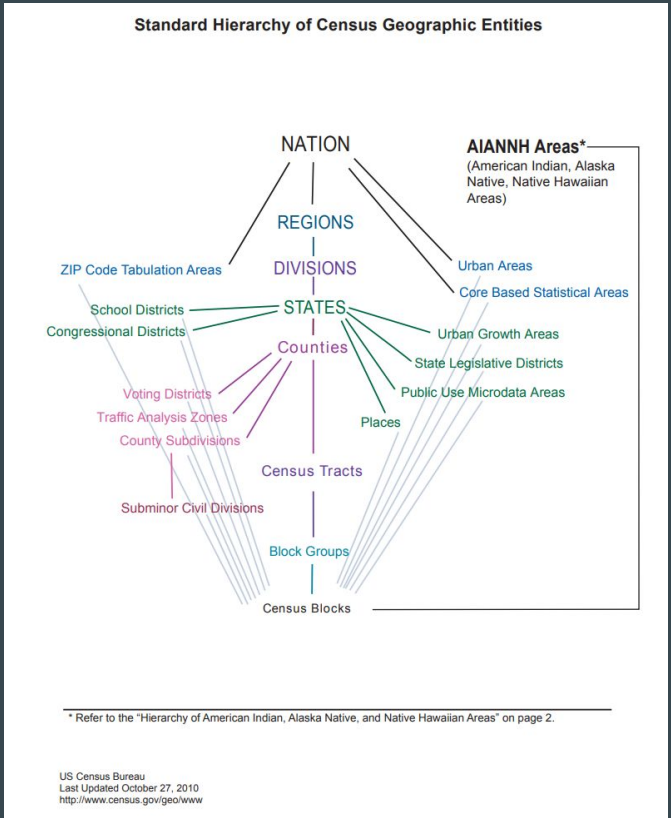
Develop new sources

Income

Mobility

Low cost food

Appendix (or for later use)



Measures of food insecurity

Chart 1. Derived and fundamental food insecurity measures.

Method	Principles	Advantages	Disadvantages
Derived measures			
FAO method	- Outcome: national calories available *per capita* per day - Inputs: food balance sheets, energy intake coefficient of variation, single cut-off point to estimate at-risk population	- Inexpensive - Applied worldwide on annual basis	- Does not identify at-risk households or individuals - Dietary quality not taken into account - High measurement error - Low standardization on data collection methods across countries - Evidence-base of cut-off point is questionable
Household expenditure surveys	- Outcomes: caloric intake *per capita* per household, dietary variety score - Inputs: money spent on food and other needs, foods consumed and market value, reference time period, food composition tables to convert food expenditures and consumption into energy intake	- Identifies vulnerable households - Can take into account dietary quality - Used for evaluating national anti-poverty and assistance programs	- Measures food available but not necessarily consumed during period of interest (periodicity bias) - Difficult to estimate foods consumed outside the household, fed to animals, exchanged as gifts or payment of work - Difficult to standardize methodology across countries - Expensive and logistically difficult - Data usually not available annually
Dietary intake	- Outcomes: Individual's food group intake counts, nutrient intake - Inputs: 24-hour recalls, FFQ, food records, food composition tables, known nutrient requirements, reference time period	- Measures actual food consumption - Can assess short medium and long term food intake - Deals with both dietary quality and quantity - Identifies at-risk households and individuals	- Memory "recall" bias - High intra-subject variability in food and nutrient intakes - Difficult to assess portion sizes - Food composition tables need to be of high quality and culturally appropriate - Uncertainty about human requirements for most nutrients - High cost especially for inclusion of 24-hour recalls in national surveys - National data unavailable annually
Anthropometry	- Outcome: % population malnourished - Inputs: weight height, other body dimensions	- Highly standardized - Evidence-based cut-off points - Inexpensive - Frequently applied in national surveys	- Nutritional status indicator - FI-obesity relationship difficult to interpret
Fundamental measure			
Experience-based food insecurity scales	- Outcome: Household level of food insecurityInputs: scale containing items representing the conceptual and multidimensional nature of FI, algorithm to convert scale scores into FI categories	- Fundamental measure of FI - HFSSM and adaptations valid across diverse socio-cultural settings - Captures the physical and psycho-emotional dimensions of FI - Low cost allowing for decentralization	- Does not capture food safety dimension - Different reference time periods and frequency response options needed in different settings - Difficult to standardize cut-off points across regions/countries -"Benefit" bias

FFQ: food frequency questionnaire; FI: food insecurity; HFSSM: US Household Food Security Survey Measure; FAO: Food and Agriculture Organization.