Food Banks Analytics - Improving Chronic Disease Control

Yi-Hsuan Hsieh
Georgia Institute of Technology
yhsieh35@gatech.edu

Luffina Huang
Georgia Institute of Technology
lhuang315@gatech.edu

Lei Jiang
Georgia Institute of Technology
ljiang87@gatech.edu

Mario Wijaya
Georgia Institute of Technology
mwijaya3@gatech.edu

Keith Woh
Georgia Institute of Technology
keith.woh@gatech.edu

1 INTRODUCTION

Food insecurity is the state of not having reliable access to adequate and affordable quantity of nutritious food for household members to lead an active and healthy lifestyle. Multiple studies [3][14][9][16] have shown that there is a strong correlation between food insecurity status and chronic health conditions, and suggested that food banks could be a promising avenue for health promotion by tackling both problems of food insecurity and chronic disease. Our goal is to develop analysis and visualization tools that help organizations and government agencies allocate resources more efficiently, and ultimately help reduce food insecurity and disease risk. For this purpose we

- Collected, cleaned, and scrapped several datasets: Demographic & population, food insecurity, disease prevalence, and food pantries' location,
- Applied analysis to study relationship between chronic disease prevalence and food insecurity,
- Applied clustering algorithms to predict potential food insecurity area,
- Created comprehensive visualization of the data and results of analysis, and provided a user-friendly interface.

2 PROBLEM DEFINITION

More specifically, our visualization tools and analyses help answer the following questions:

(1) What are the relations between food insecurity and the prevalence of obesity and diabetes?
(2) How does geographic and demographic characteristics of the population affect their susceptibility to the risk of chronic diseases induced by food insecurity?
(3) Can we predict areas with high risk of chronic diseases due to lack of food?
(4) How many people can a supermarket/food pantry impact if built at a specific location?

3 SURVEY

3.1 Concept

Food insecurity is a widely discussed topic in the literature. For example, Berry et. al.[4] offered a conceptual analysis of food security and related terms (e.g. accessibility). Their theoretical work helps ensure that we understand and use our key terms correctly. Buczynski et. al. [6] defined the concept of food desert, and this provides a guide on how we can visualize food insecurity. This only explores data in Baltimore, and our project can scale it up to a national level.

3.2 Visualization

There is an absence of comprehensive visualization integrating multiple categories of information. FeedingAmerica has a choropleth map (Figure 1) showing food insecurity at county-level.[1] Figure 2 shows an interactive choropleth map provided by USDA[2], where users can visualize various indexes, such as access to grocery stores and socioeconomic characteristics.

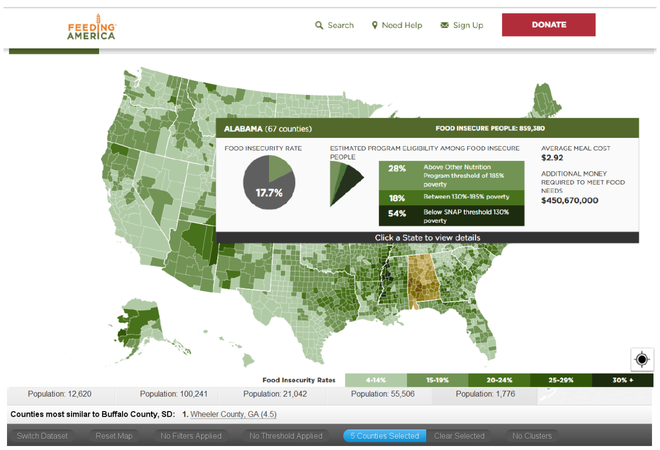

Figure 1: Feeding America Choropleth Map

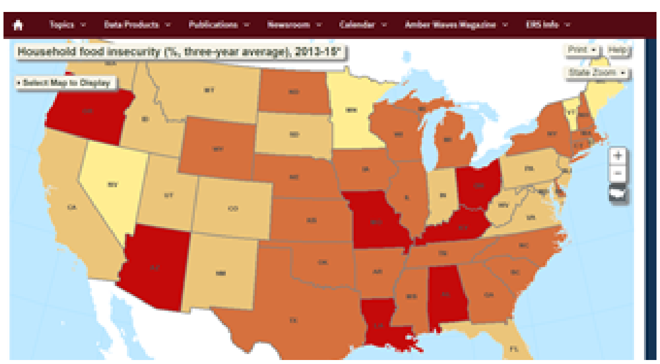

Figure 2: USDA Food Environment Atlas

However, these two visualizations, as well as other methodologies we surveyed, do not allow users to view information from multiple categories at the same time. Users are limited to visualizing one factor at a time, making it difficult to spot any relationships between factors visually.

Joshi et. al. [17] showed visualization using disease, economic, and educational factor. The visualization gave us idea on how to visualize our data with metrics that makes visualization meaningful. However, this visualization did not incorporate information regarding food insecurity.

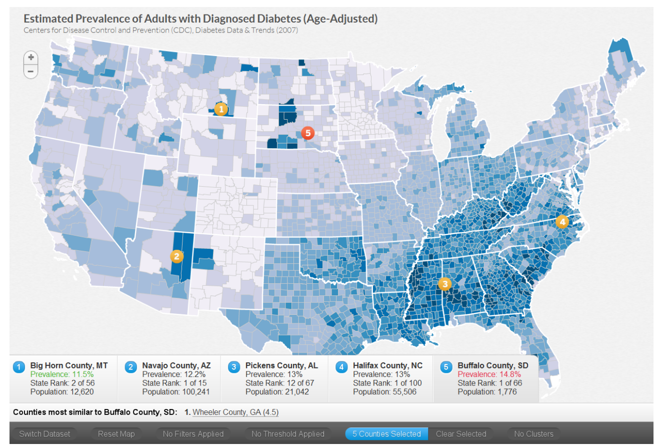

Figure 3: Visualization using disease, economic, and educational factor

We also surveyed works regarding theoretical principles in visualizing food-related and geographic information. Kronenfeld and Wong [13] stated that choropleth map might lead to a biased visual perceptions due to size alone, without regard for density, thus we can avoid this by utilizing cartograms. Goldsberry and Duvall [8] built a geospatial visualization of food accessibility. Their visualization is on a city-level. We can use this as a guide to scale it up to national level. Jung et. al. [12] introduces a methodology to increase the power of GIS and geovisual techniques to visualize the socio-spatial inequality by incorporating social dimensions. Though the data resource needed to implement the visualization is hard to access, we can adapt this method to implement our visualization.

[1]http://map.feedingamerica.org/
[2]https://www.ers.usda.gov/data-products/food-environment-atlas/go-to-the-atlas/

3.3 Analysis

For data analysis, Amarasinghe et. al. [2] applied ordinary least square to county-based data to analyze the relationship between food insecurity and various factors, such as water source and poverty. This serves as a reference to analyze the relationship between disease and food insecurity. Glenn Hyman et. al. [11] explored 6 cases of spatial analysis on poverty and food insecurity. The small area analysis and geographically weighted regression mentioned can be tentative models for our project which uses county-based datasets.

Seligman et. al. [15] provided evidence that food insecurity contributes to poor diabetes self-management. They suggested potential interventions that our project can improve on. Gundersen et. al. [10] inferred that food insecurity is related with chronic metabolic diseases and food-sensitive disease. This provides us with the disease types on investigating the effect of food insecurity and the potential intervention for each disease. The data analysis and visualization on this paper still requires nationwide data to make concrete conclusion.

Susan J. Algert et. al. [1] utilized clustering to find areas with severe food insecurity. However, the data granularity is at a individual level, and we might not able to access such detailed data. Brock et. al. [5] used neural network to estimate the amount of food available for collection at supermarket. The approach expands our modeling technique given limited information. The approach used in the paper is narrow focused and used single model to approach the problem.

4 PROPOSED METHOD

This section details the approach we took: Section 4.1 explains our data gathering process, and our list of innovations is described in sections 4.2 - 4.4.

4.1 Data Gathering and Pre-Processing

All datasets are gathered from public sources: disease prevalence datasets are obtained from CDC, geographic and demographic information are obtained from US Census Bureau, food-insecurity rates are obtained from FeedingAmerica, and locations of food pantries are scraped from public web-page of foodpantries.org. After collecting all datasets, we cleaned and merged all datasets to improve ease of use for algorithm analysis and visualization. Python libraries Pandas and Numpy are used for data preparation.

4.2 Multiple Layers in a Map

We want to create an interactive and comprehensive geo-spatial visualization with toggling capabilities using these layers:

(1) Food insecurity rates
(2) Demographics and disease prevalence
(3) Food pantries location
(4) Algorithm results (see section 4.4)

To provide a user-friendly interface, our map allow users to visualize different layers on the same map. As shown in Figure 4, user has selected 2 layers: base layer is a choropleth map of our clustering results (see section 4.4), red dots refer to diabetes prevalence, and the radius of circles reflect prevalence rates. Refer to Figures 5 - 8 for various map layers. User has full flexibility to show/hide the layers that they want to see by toggling the buttons on the left panel. We have a total of 6 layers, including algorithm results (such as clustering), and various information layers (such as disease rates).

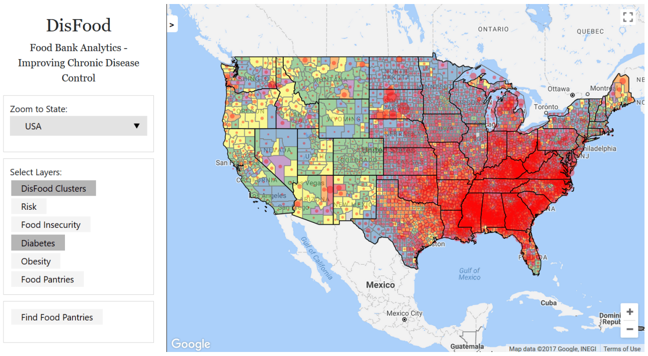

Figure 4: Integrated view

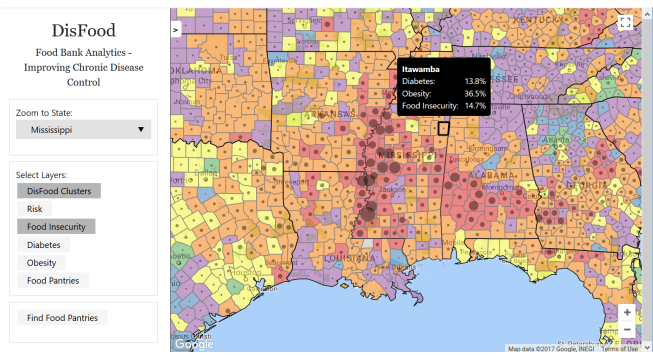

Figure 5: Zoomed-in view, with tooltip

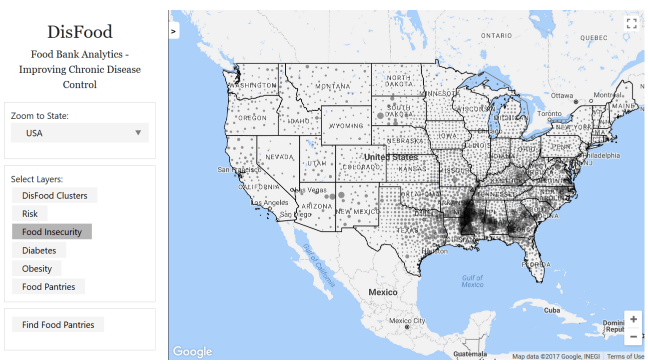

Figure 6: Food Insecurity Layer

4.3 Comparison Capability between Counties

To provide additional analysis, we added a sidebar to allow users to compare disease prevalence rates between two counties over time. Users are able to toggle this additional sidebar via a button on the top-left corner of the map. Figure 9 shows a screen-shot of our sidebar. Once the user selects a county, the d3 line chart is then updated with relevant information.

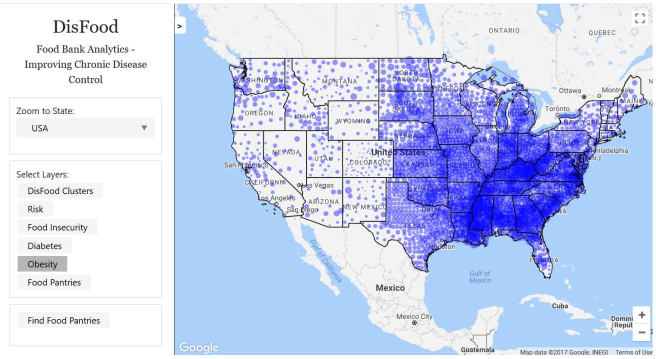

Figure 7: Obesity Layer

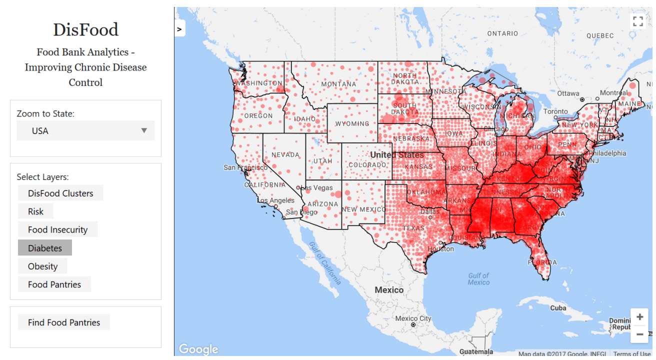

Figure 8: Diabetes Layer

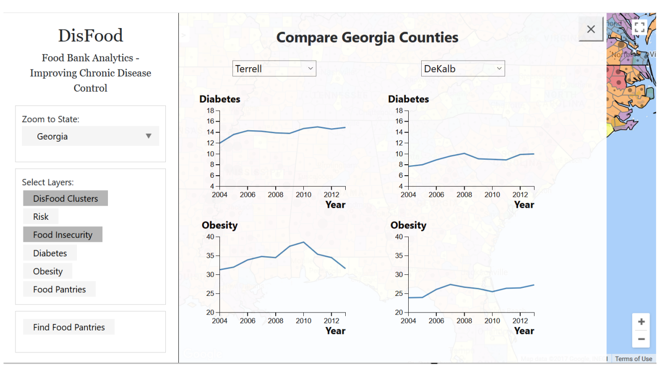

Figure 9: Comparison between Georgia Counties

4.4 Algorithms

4.4.1 ***Clustering.*** To decide the optimal number of clusters, we use Elbow method and find the optimal number of clusters where the change in sum of squared errors become less significant.

We used K-Means clustering to cluster counties into 6 groups based on food insecurity rate, obesity rate, and diabetes rate to better understand condition of food insecurity and chronic disease in U.S.

4.4.2 ***Find High-Risk Areas.*** A high-risk area in our project is defined as a county that has relatively high food-insecurity rate compared to other counties with similar living condition, such as: food access distance, income, disease rate, and poverty rate etc. The goal is to analyze how similar of the severity from a county to another. From [9][14], we also know that diabetes and obesity are associated with food insecurity. Therefore, the team decided to add more features: diabetes and obesity prevalence.

Radius based nearest neighbor algorithm is used to find counties that has similar living condition. Compared to the K-nearest neighbor algorithm, this approach enables us to prevent some counties that have all neighbors far from them by setting a threshold of distance, i.e. the radius, between counties and their neighbors to remove outliers. Food insecurity rate for counties is transformed into food insecurity level according to the threshold obtained from FeedingAmerica as the label for our dataset.

The neighbors for each county is identified by measuring the similarity of living condition between them. First, we compute the statistic distance of given living conditions between each pair of counties. Based on these distances, we consider 2 counties to be neighbors if the distance between them is in the first percentile of all the distances, which is the radius for our nearest neighbor algorithm.

We found that the features in the dataset have dependency to some degree. To take this into consideration, instead of using Euclidean Distance, we chose to use Mahalanobis Distance because it can effectively measure scale-invariant features and takes into account the correlations of the dataset.

Once all of the similar neighbors of a county is determined, new food insecurity level for each county is generated by neighbors voting. That is the expected food insecurity level with the given living condition. By comparing the expected food insecurity level against the existing food insecurity level, if the original food insecurity rate is higher, we can claim that this county has more severe living situation in regards to food access and food stability. This could act as an indicator for organizations to pay more attention and thus can adjust related policies.

Figure 10 shows our risk layer. Gray colored area indicates high-risk counties, whereas green indicates low-risk.

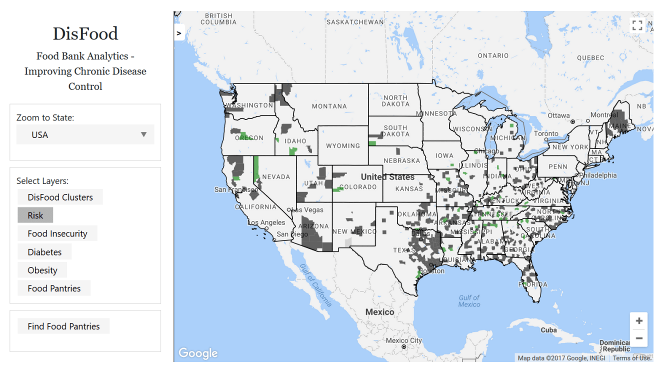

Figure 10: Risk Layer

5 EXPERIMENTS

5.1 Visualization Comparison

As mentioned in Section 3.2, all existing visualizations are either focused on food insecurity rate, as shown in Figure 1, or disease prevalence, as shown in Figures 2 and 3. The team experimented with these layers of visualization using d3.js and Google Map API:

- Risky area
- Food insecurity

- Diabetes prevalence
- Obesity prevalence
- Food pantries location
- Disease and food insecurity
- Supermarket location

Clearly, our tool is far more superior than any existing tool that exist out there. It allows users to look for things that they are interested in and can analyze disease prevalence and food insecurity at once. It serves as one-stop shop for those interested in food insecurity and disease prevalence.

5.2 Optimized Visualization Performance

Our map visualization used to have slight lagging issues. When users zoom and pan freely, the map has to be re-rendered and recolored for states and counties. The team experimented with plotting the map using Google Map API instead of native d3 based map. However, the difference is not noticeable hence we decide to use the native d3.js based visualization.

To reduce the lagging issue, the team created a drop down menu that contains all of the states in U.S. When a user select a state on the drop down list, the map will be zoomed to that particular location.

5.3 Food Insecurity Implication

By visualizing the result of the clusters obtained using chronic disease prevalence and food insecurity rate as specified in Section 4.4. One can answer the following questions:

- Which county/state has higher food insecurity rates and chronic disease prevalence?
- Which county/state has worse healthy food access and stability of healthy food resource compared to other county/state with similar living condition such as income, disease rate, ratio of car ownership, etc?

Based on the algorithm and visualization, we found that Mississippi appears to be the state with highest number of red clusters. Red clusters tend to have higher food insecurity rates and chronic disease prevalence, indicating severe condition. To further validate the result, we visualized the food insecurity layer as black dots visualization as shown in Figure 11 to check if our result match with our intuition. Clearly enough, the visualization showed exactly what we expected.

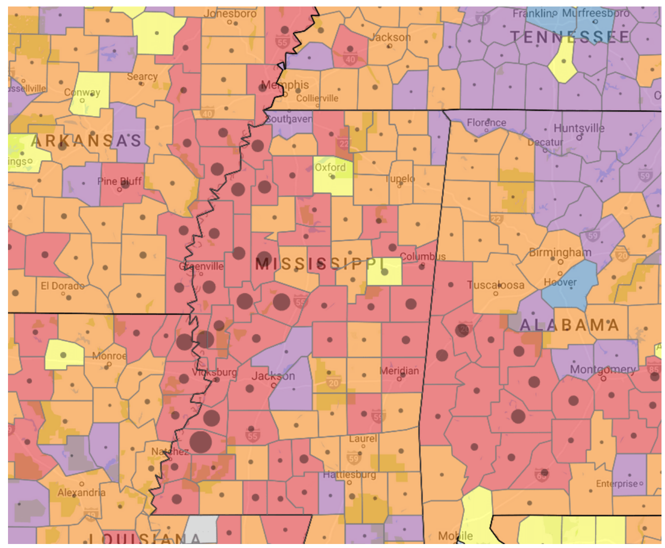

Figure 11: Result Validation

5.4 Relationship Between Food Pantries and Food Insecurity

The team is interested in understanding whether there is any connection between food pantries and food insecurity. The hypothesis that the team came up with initially was that the lesser food pantries available in the region, the riskier it is. To verify this, the team visualized food pantries location and clusters based on similar living condition. Gray cluster areas are those that have similar living condition such as: food assessment, income, disease rate, poverty rate, etc to their neighbors but tend to have relatively higher food insecurity rate. White cluster areas are those that have no difference in food insecurity rate, while the green cluster areas are those that have relatively lower food insecurity rate.

From Figure 12, gray cluster areas appears to have fewer food pantries, this supports our hypothesis, indicating the areas with less food pantries tend to be riskier than those that have more food pantries.

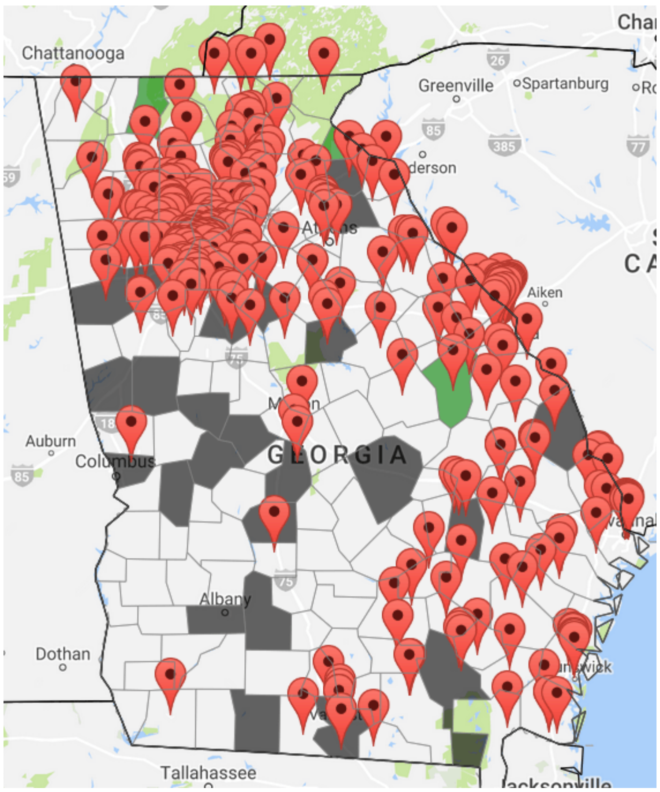

Figure 12: Georgia high-risk areas and food pantries location

5.5 Locating Closest Food Pantries

The team has added an additional interaction on the map by providing users with the closest food pantries given their approximate location. There are multiple food banks in a state, which raise the question on which location is closest to the user. Users can click on the "Find Food Pantries" button and the system will show 3 closest food pantries. Top 3 closest locations are determined by calculating the distance of where the user is located compared to the location of food pantries. Figure 13 shows an example, closest 3 food pantries are shown as markers on the map, and the food pantry's specific information is shown when user clicks on the markers.

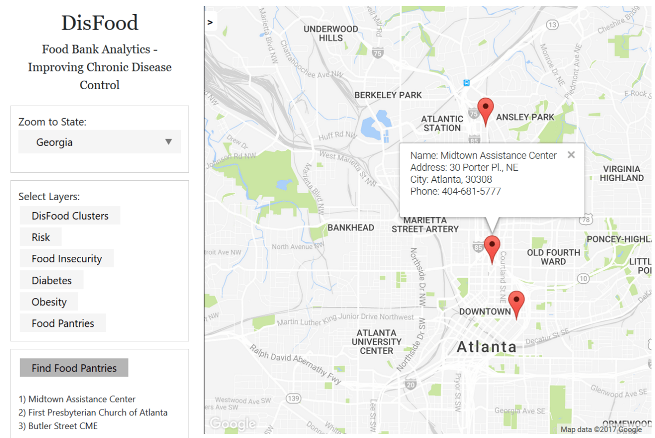

Figure 13: Locate user's closest 3 food pantries

5.6 Supermarket Locations

The team developed another visual analytic module to analyze accessibility to supermarkets in Georgia as shown in Figure 14. Accessibility is measured both in terms of distance and travel time to the nearest supermarket. We were unable to pinpoint the location of every individual or household in the population, but we utilized census-tract level data. For each tract, we use a representative location and query Google Maps API to get travel time and distance from that location to the nearest supermarket, and use those data to describe typical situation of the population in the tract. Then, we combine these data with demographic data and aggregate for each county. Users can query average distance or travel time to supermarkets, and apply multiple filters by demographic groups, income, and distance threshold to be considered as low accessibility. The results under certain conditions were compared against the USDA data and were shown to be consistent with the latter. When compared to USDA food insecurity data, this tool offers a more flexible and interactive way to explore and analyze food accessibility in the state.

Furthermore, the novel analysis that this module enabled is to recommend tracts to build a new supermarkets that can affect the most people in the group of interest, or result in the greatest reduction of travel time. For each census tract, we query Google's Distance Matrix API to compute which other nearby tracts will benefits from a new supermarket in the current tract. Based on data from about 20,000 API calls, the system can compute the recommended place to build a new supermarket given the population and distance filter the user provides. This provides the users with an interactive way to analyze the impact of interventions on food insecurity.

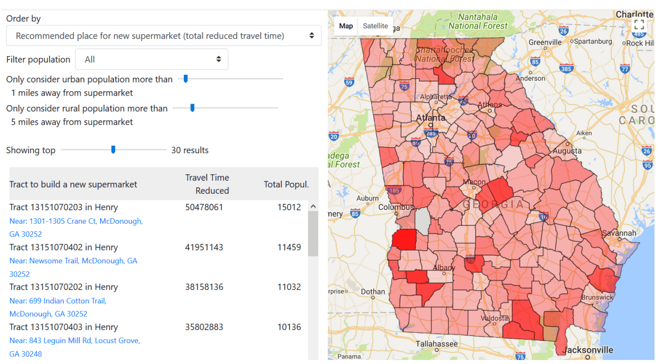

Figure 14: Visualizing Food Inaccessibility

5.7 Children Health Issue and Food Insecurity

Child food insecurity is a pressing issue, as children are particularly vulnerable to inadequate food intake. In 2015, nearly 13.1 million U.S. children, or 16.6 percent, lived in food-insecure households. Relatively little research has been conducted to find out what causes food insecurity among children [7].

Using state level data, our team's preliminary analysis showed that there is a positive correlation between child food insecurity rates and child obesity and death rates as shown in Figure 15. The linear regression model used children obesity and child death rate as covariates. The p-value of coefficients for child death rate and children obesity are 0.029 and 0.001 respectively, which are significant with type I error (α) of 0.05.

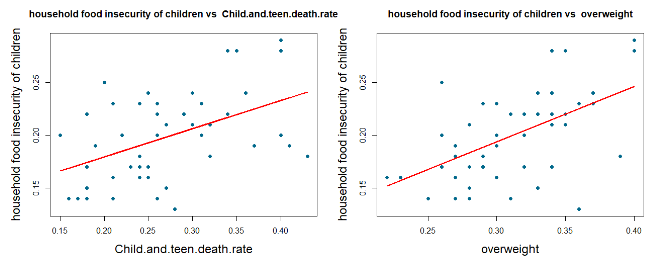

Figure 15: Child Food Insecurity

6 CONCLUSIONS AND DISCUSSION

6.1 Better Allocation of Resources

Our study results shed light on improving the food distribution efficiency and enhancing the health concurrently. Our analysis provides food pantries, with more detailed information on the distribution of chronic disease population. They can take advantage of these information to pinpoint the area with high disease prevalence and allocate the food resource more appropriately. For example, the food bank can allocate more low-sugar food to area with higher prevalence of diabetes, or more low-salt food to areas with higher prevalence of hypertension.

6.2 Future Work

6.2.1 County Level Data for Children Health Issue and Food Insecurity. Prevention of chronic disease as early as possible is better than cure. The team explored the data to find whether there is any relationship between children health condition and food insecurity rate. We found that children food insecurity correlated with children obesity and death rate of children & teenagers at state level. Since the detailed data for counties is not available, we are not able to conclude the relationship at the county level. If county level data is available in the future, the team would like to do comprehensive analysis to gain more insight.

7 DISTRIBUTION OF TEAM EFFORT

All members contribute similar amount of time and effort. Below shows concrete details of task performed by each member.

All members	Research, writing documents, brainstorming visualization.
Yi-Hsuan Hsieh	Problem formulation for algorithms surveyed. Validation of data to use for analysis.
Luffina Huang	Found dataset for the project, surveyed algorithms. Proposed visualization design.
Lei Jiang	Optimization of the code. Algorithm exploration and formulation.
Mario Wijaya	Merged and Preprocessed dataset. Visualization design and implementation.
Keith Woh	Proposed samples for map visualization. Created API for visualization in d3.js.

REFERENCES

[1] Susan J Algert, Aditya Agrawal, and Douglas S Lewis. 2006. Disparities in access to fresh produce in low-income neighborhoods in Los Angeles. *American journal of preventive medicine* 30, 5 (2006), 365–370.

[2] Upali Amarasinghe, Madar Samad, and Markandu Anputhas. 2005. Spatial clustering of rural poverty and food insecurity in Sri Lanka. *Food Policy* 30, 5 (2005), 493–509.

[3] Seth A Berkowitz, Theodore SZ Berkowitz, James B Meigs, and Deborah J Wexler. 2017. Trends in food insecurity for adults with cardiometabolic disease in the United States: 2005-2012. *PloS one* 12, 6 (2017), e0179172.

[4] Elliot M Berry, Sandro Dernini, Barbara Burlingame, Alexandre Meybeck, and Piero Conforti. 2015. Food security and sustainability: can one exist without the other? *Public health nutrition* 18, 13 (2015), 2293–2302.

[5] Luther G Brock and Lauren B Davis. 2015. Estimating available supermarket commodities for food bank collection in the absence of information. *Expert Systems with Applications* 42, 7 (2015), 3450–3461.

[6] A Buczynski, Holly Freishtat, and Sarah Buzogany. 2015. Mapping Baltimore City's food environment: 2015 report. *Baltimore Food Policy Initiative and Johns Hopkins Center for a Livable Future* (2015).

[7] Alisha Coleman-Jensen, Matthew Rabbitt, Christian Gregory, and Anita Singh. 2016. Household food security in the United States in 2015. (2016).

[8] Kirk Goldsberry and Chris Duvall. 2009. Visualizing nutritional terrain: an atlas of produce accessibility in Lansing, Michigan, USA. *American Journal of Preventative Medicine* 36 (2009), 485–499.

[9] Christian Gregory and Alisha Coleman-Jensen. 2017. Food Insecurity, Chronic Disease, and Health Among Working-Age Adults. (2017).

[10] Craig Gundersen and James P Ziliak. 2015. Food insecurity and health outcomes. *Health affairs* 34, 11 (2015), 1830–1839.

[11] Glenn Hyman, Carlos Larrea, and Andrew Farrow. 2005. Methods, results and policy implications of poverty and food security mapping assessments. *Food Policy* 30, 5 (2005), 453–460.

[12] Jin-Kyu Jung and Christian Anderson. 2017. Extending the conversation on socially engaged geographic visualization: representing spatial inequality in Buffalo, New York. *Urban Geography* 38, 6 (2017), 903–926.

[13] Barry J Kronenfeld and David WS Wong. 2017. Visualizing statistical significance of disease clusters using cartograms. *International journal of health geographics* 16, 1 (2017), 19.

[14] Barbara A Laraia. 2013. Food insecurity and chronic disease. *Advances in Nutrition: An International Review Journal* 4, 2 (2013), 203–212.

[15] Hilary K Seligman, Terry C Davis, Dean Schillinger, and Michael S Wolf. 2010. Food insecurity is associated with hypoglycemia and poor diabetes self-management in a low-income sample with diabetes. *Journal of health care for the poor and underserved* 21, 4 (2010), 1227.

[16] Hilary K Seligman, Courtney Lyles, Michelle B Marshall, Kimberly Prendergast, Morgan C Smith, Amy Headings, Georgiana Bradshaw, Sophie Rosenmoss, and Elaine Waxman. 2015. A pilot food bank intervention featuring diabetes-appropriate food improved

glycemic control among clients in three states. *Health affairs* 34, 11 (2015), 1956–1963.
[17] Jared Shenson and Alark Joshi. 2012. Visualizing disease incidence in the context of socioeconomic factors. In *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction*. ACM, 29–38.