

DATA VISUALIZATION FINAL PROJECT

OBESITY IN MASSACHUSETTS - VISUALIZING OBESITY RELATED FACTORS AND FEDERAL AID PROGRAMS

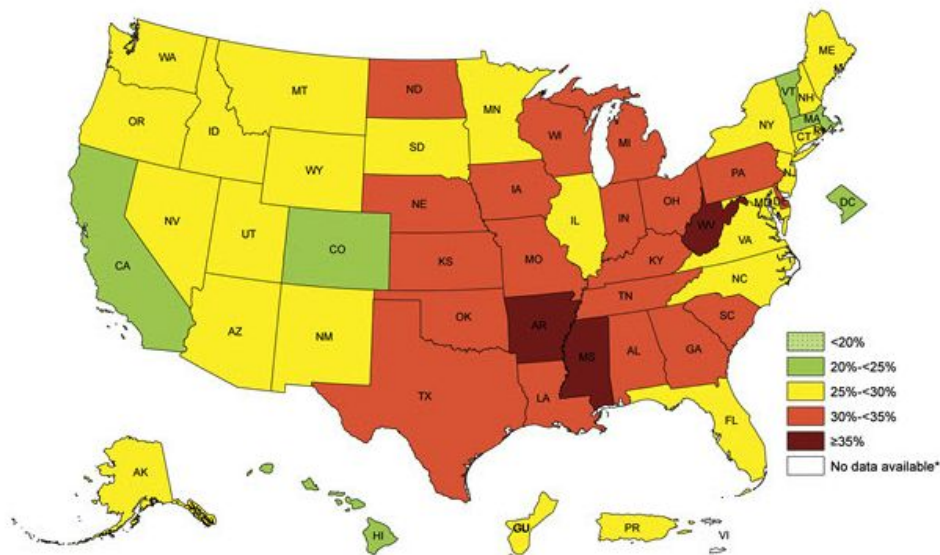
- *Chitra Pichaimuthu Kanickaraj*
- *Tabassum Kakar*
- *Suwodi Dutta Bordoloi*

Overview and Motivation:

Obesity is a severe public health issue in the US, where currently, one in three U.S. adults is obese. More than 35% of US Adults are obese, 22 states have rates above 30% and every state is above 20%. Obesity is known to cause heart disease, stroke, type 2 diabetes and certain types of cancer. Through this project we want to highlight the condition of obesity in Massachusetts, socio-economic data of the state, some of the key obesity related factors and data of federal programs. Our goal is to help obesity policy and program decision makers in planning and allocating federal budgets to various food assistance programs.

Related Work:

We were inspired by the State obesity prevalence maps[1] provided by the Center of Disease Control and Prevention (Figure 1) and wanted to dig further in order to understand and investigate the factors such as socioeconomic status, access to grocery stores and restaurants etc. effecting or causing obesity in these states. Moreover we want to build a tool that provides insights that ultimately helps agencies in decision making processes in an effective way through visualizations to tackle obesity.



Questions:

The objective of our project is to understand and explore which counties within Massachusetts are mostly affected by obesity and what are the factors such as socioeconomic status, income and physical activity etc. contributing to obesity. Initially we thought of focusing on exploring the

factors affecting obesity using visualizations and showing some statistics of the distribution of obesity across different counties and people's access to different facilities such as grocery stores etc. Through our evolving designs, our primary objective slightly changed to build visualizations that will finally help state government to understand the counties and demography most prone to obesity in Massachusetts, how successful have the past programs been and to help in decision making process for budget allocation.

Data:

The dataset for this study has been chosen from the United States Department of Agriculture (USDA) Economic Research Service portal. This data is collected from a variety of sources and cover varying years and geographic levels. It comprises of about 168 predictors categorized across 9 different categories.

01. *Access and Proximity to Grocery Store*: It provides an overview of community's access to grocery store.
02. *Store Availability*: It provides numbers of supermarkets, grocery stores, supercenters and club stores, and various other establishments that are primarily engaged in retailing a general line of food.
03. *Restaurant Availability and Expenditures*: It provides numbers of limited-service restaurants and full-service restaurants along with their average expenditures on food purchased.
04. *Food Assistance*: It gives us statistics as to how providing nutrition assistance has helped people make healthy food choices and the effect of economic incentives on making informed decisions about diet quality.
05. *Food Insecurity*: It provides percentage of prevalence of household-level food insecurity which affects the food intake of household members and disrupted eating patterns because of insufficient money and other resources for food.
06. *Food Prices and Taxes*: It provides statistics on price and sales tax of few food items like milk, soda, chips and pretzels to help compare their prices with each other and other general food items.
07. *Local Foods*: It provides information on locally sourced and locally available foods and the programs used to enlighten the nutrition education of these farms.
08. *Health and Physical Activity*: It provides information on recreation and fitness facilities along with the diabetes and obesity rate among different age groups categorized by income .
09. *Socioeconomic Characteristics*: It provides a census of population based on race, age group, poverty rate etc.

Data cleanup -

One of our main problems with the dataset is the large number of predictors and many of them are highly correlated because there are multiple variables from the same category. Variables represent the attributes in percentage, percent change between two years as well in absolute numbers. We had to perform attribute selection (both apply statistical techniques) and manually to drop redundant variables. To address missing value problems, we have replaced them with mean or median in many cases. Also there are some data inconsistencies and repeated records that needed to be removed.

Exploratory Data Analysis:

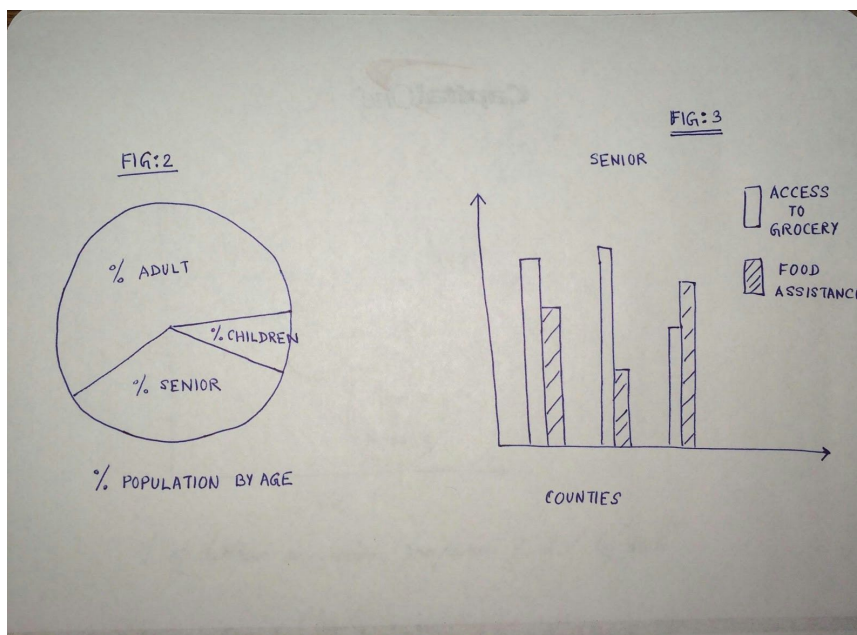
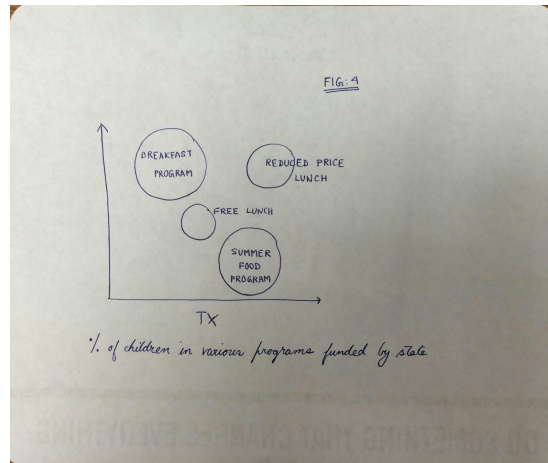
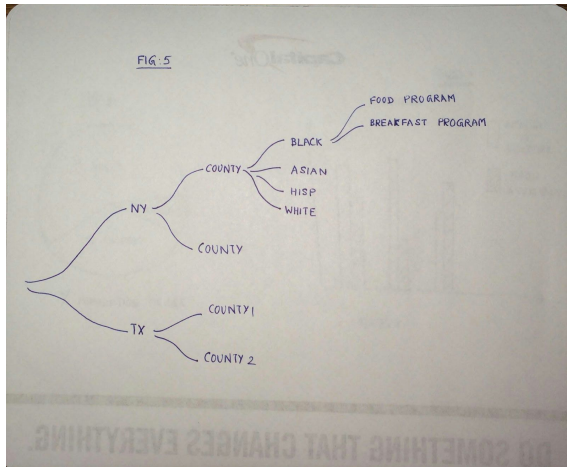
We used Weka and Orange (Python) to explore the data set and started with basic plots like pairwise scatter plots comparing each pair of predictors and explored the distribution of each of the attributes. It was very helpful to see the predictors which are highly correlated within themselves. We also found that some of the variables like accessibility to stores and local foods have a very skewed distribution.

From the scatter plots, we saw child poverty, number of participants in free lunch, number of participants in free breakfast programs, accessibility to groceries, etc are all correlated with obesity. Another insight was that poverty rate is positively correlated with Black and Hispanic population (while it isn't true for white, asian or other races) as well as the benefit programs (like free lunch, snap participants, etc). Convenience stores have a positive correlation with adult obesity. We also found that snap benefits per capita is correlated with convenience stores - which is not very evident. But we saw that lower income households do participate more in the snap or other food assistance programs.

The insights gained were very crucial since we selected the variables correlated with obesity as some of our key factors for visualization.

Design Evolution:

We considered various visualizations for displaying our data such as maps, bar-charts, pie-charts, bubble-charts and treemaps, but we selected only maps, bar charts and pie charts as they are well suited for the type of analysis we intend to do. Below are the initial sketches of our design.



We chose map to show the obesity density distribution across Massachusetts across counties. The color density is mapped to the obesity density, i.e. darker color implies more obesity whereas lighter color implies less obesity. User can select a county and view the obesity distribution across different races and age groups through bar charts.

At first we planned to implement a US state-level overview of the obesity distribution. But as per Professor's suggestion and our thought of local health officials being more accountable for controlling obesity, we settled to show obesity per state across each of its counties.

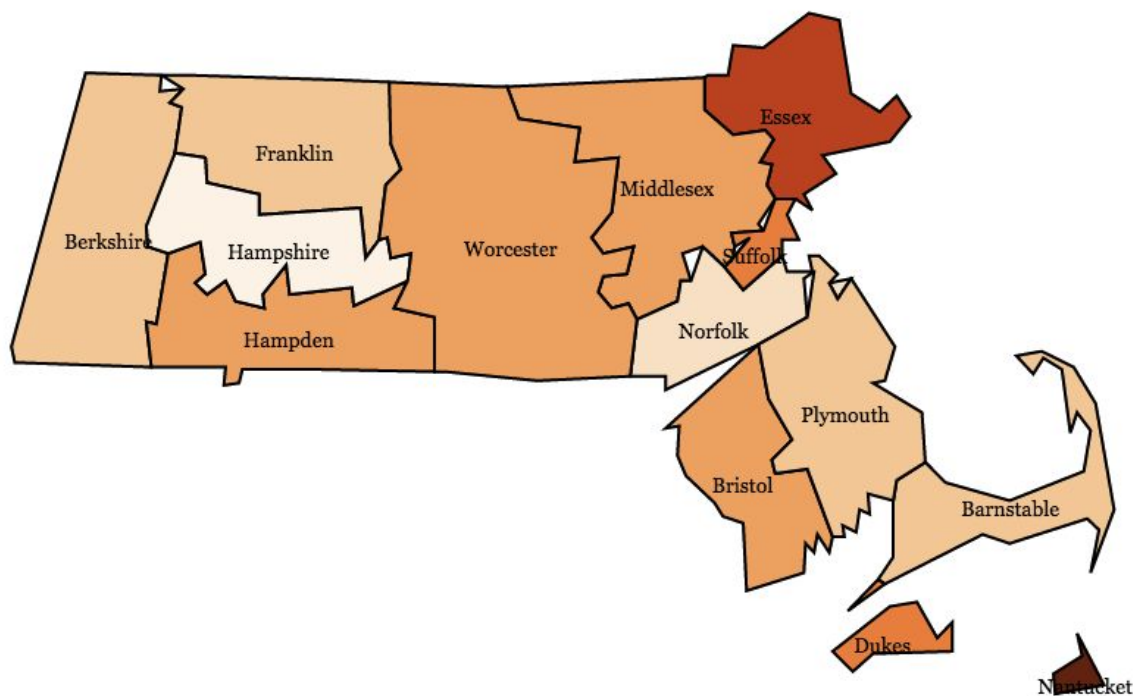
Implementation:

We have implemented the visualizations such that with each additional visualization, the user has an additional layer of information about a county that ultimately helps in learning the prevalence of obesity, how the factors affecting obesity are prevailing in the counties, and the government funded programs. At each level, we also provide different references to compare to in each of the graph (for example either the state average or the most/least obese county).

The visualizations are all interactive and any county updated on any chart updates all the charts for that particular county. This feature is extremely useful in exploring and investigating every county on a case by case basis as well compare with others.

Visualization 1: MA Interactive Obesity Map

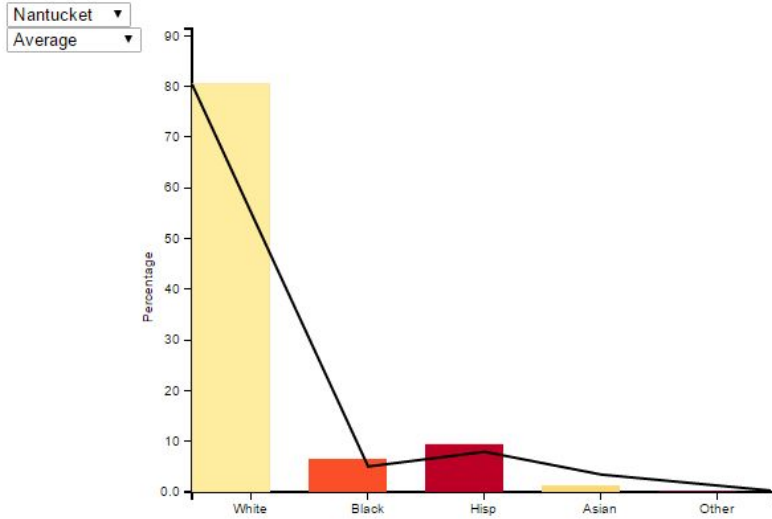
For the year of 2014, we show the child obesity map for MA counties with low to high obesity percentages represented by the color gradient. The user will be able to select any county, and the following visualizations show details for that county only.



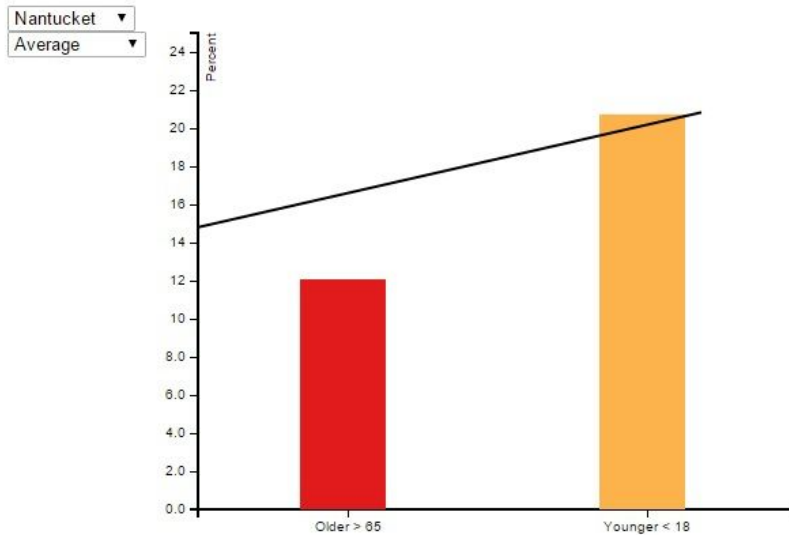
Visualization 2: Obesity Race Distribution

For the county selected by the user, we show the demographics of the county in the barcharts (Obesity Distribution by race and age) for that specific county. We will also show the average of

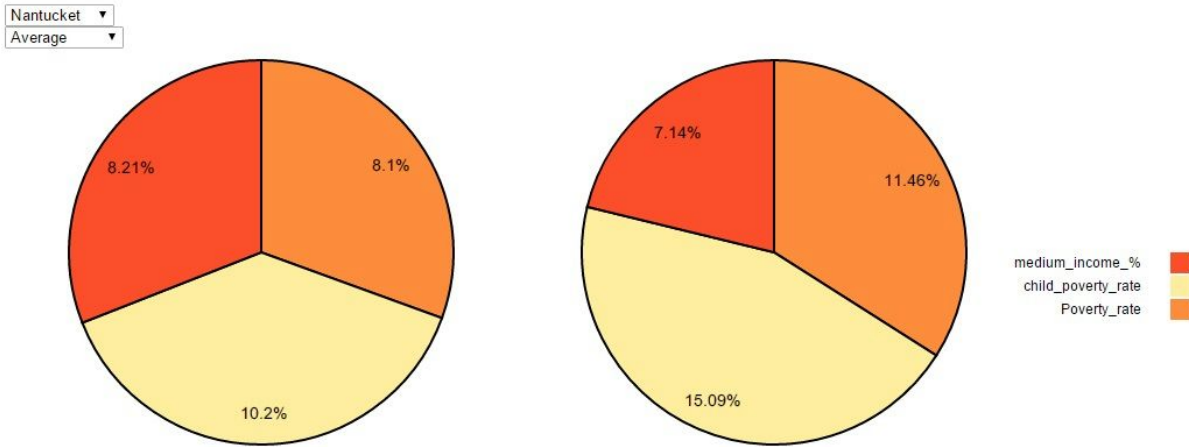
the obesity for race and age respectively for entire state of massachusetts for comparison. Our intent is that for each selected county, the user gets an overview of the demography and socio-economic conditions of the county and is able to compare with the state average.



Visualization 3: Obesity Age Distribution



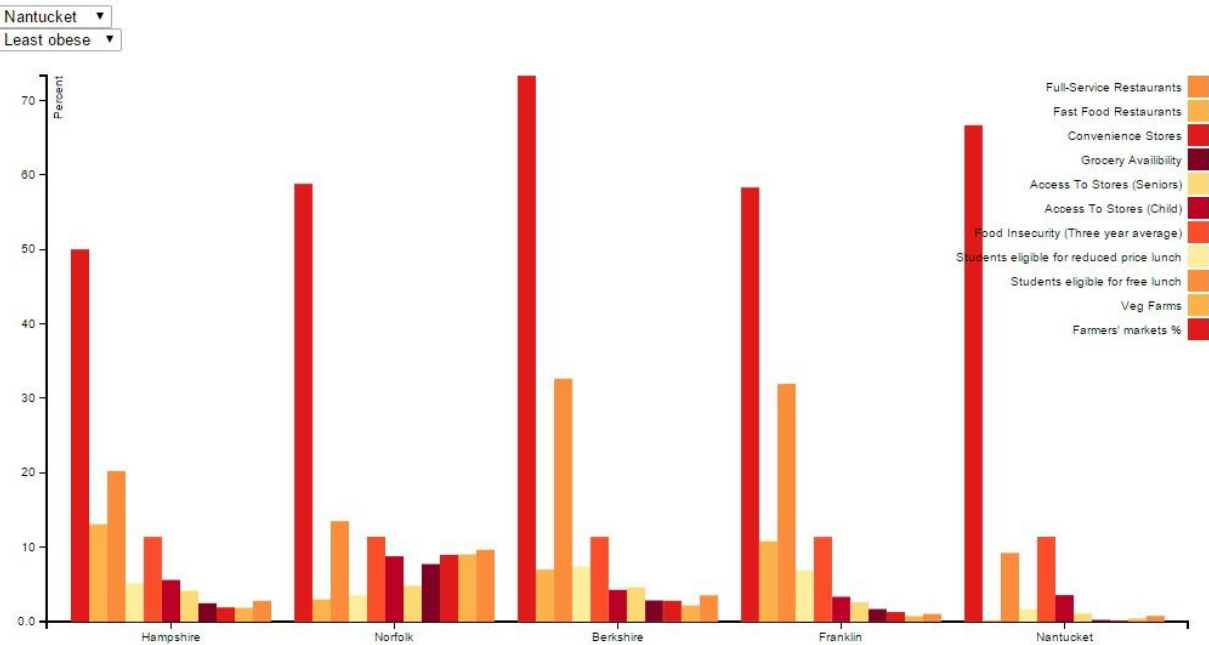
Visualization 4: Pie charts showing poverty rates, child poverty rate and median income percent of state average versus county



Visualization 5: Interactive grouped bar chart - displaying obesity related factors

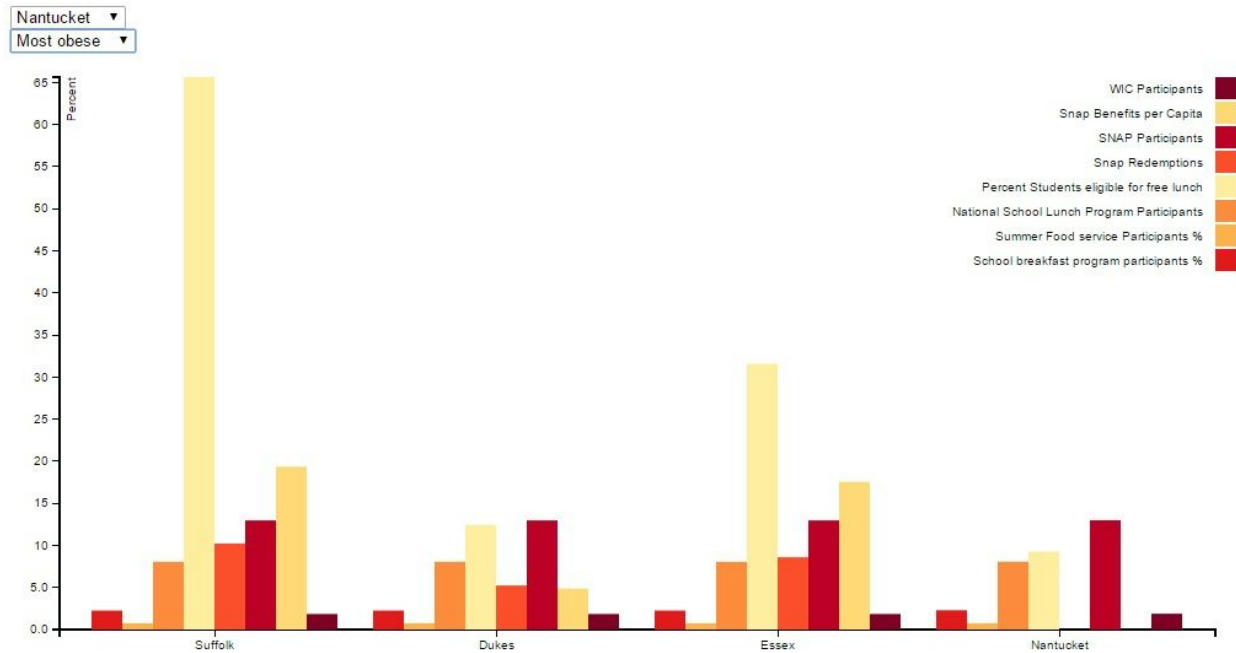
The grouped bar chart compares the factors of a specified county to others - it will compare the factors (strong predictors of obesity) such as food insecurity, access to grocery stores, convenience stores and available fast-food or full restaurants in the county.

A drop down menu is provided with each chart so that the user can explore the charts easily without scrolling back to the map. Also, the user will have an option to compare the selected county (from the map or menu) with either the state average or top four obese or bottom four obese counties.



Visualization 6: Interactive grouped bar chart showing food programs

This visualization is similar in implementation to visualization 5. It helps in understanding how the government funded programs are distributed across counties, participation in these programs, the eligibility of students to participate in such programs, etc. This will provide a definite view of how the programs are being funded in other counties as opposed to the particular county.



Evaluation:

Socio-economic evaluation : The majority of population in Massachusetts is white with a small percent of black, hispanics and asian population. Counties with higher obesity rate have a relatively higher percent of hispanic and black population than state average. Eg: Suffolk county (third highest obesity rate) has the highest black and hispanic population and is the third highest obese county in Massachusetts. On the contrary, Dukes has above 80% of white population and is second most obese county.

Except Barnstable, all counties have more young obese population than older obese population. Also obese counties tend to have an above average child poverty rate and lower median household income with an exception of Nantucket (which is the most obese county). Suffolk and Hampden have the lowest median income percents and high poverty rates.

Factors and food programs evaluation: Many of the more obese states have lower percent of students eligible for free lunch while the case is opposite for some of the lower obese counties. Suffolk is an exception with high child obesity rate and where the county has a high percent of student eligible for free lunch or reduced priced food. Also many of the obese counties have a lower percent of local vegetable farms. For most of the programs the percent of participants is equal across all counties in spite of a very highly disproportionate child poverty rate and student's eligibility percentage within the counties.

ANALYSIS:

The government aided food programs are responsible for the healthcare and nutrition of low income families, especially women and children. It is crucial that these programs are administered such that children have more access to a nutritioned diet as opposed to cheap fast food. We see that all counties have equal percent of student participants whereas there is a highly disproportionate poverty rate among the counties. We need to investigate why poorer counties with larger percent of students eligible for food assistance programs cannot attract more participants. Unfortunately some of these also have higher obesity rates. For example - Suffolk is the fourth highest obese county, has a high percent of eligible students for food programs as well as high hispanic and black population. An exception to this case is Berkshire, which on the higher end of poverty rates in the state, but has is the third least obese county. On the other hand, the most obese county is Nantucket - the island south of Cape Code is also the least populous county with higher income rates and an above state average hispanic population. The island has low accessibility to grocery stores, restaurants and no local farms. The participants in food programs is also least in Nantucket which can be explained by the low poverty rates. But because of the uniqueness of Nantucket, we need a customized plan to implement the programs to help fight obesity.

Future improvements: In the future, our design can be improved by including US map, so that user can select any US state, and drill down to the counties within that state and get insights about the obesity distribution for each county. In addition we can extend this work to multiple years and provide the user an option of selecting data from a particular year and comparing across different years.

References:

1. State Obesity Prevalance Maps: <http://www.cdc.gov/obesity/data/prevalence-maps.html>
2. <http://bl.ocks.org/mbostock/5872848>
3. <http://bl.ocks.org/mbostock/3887051>
4. <http://bl.ocks.org/Caged/6476579>

