

Proposal: Food Accessibility at the County-Level

DATA 450 Capstone

Tenzin Sherpa

2/5/23

1 Introduction

Maintaining a good diet and nutrition makes for our most basic and important needs. Many people, however, do not have easy access to fresh and healthy foods. This comes about due to a lack of grocery stores, food markets, and other food sources near the residence. This problem can most adversely affect vulnerable populations like seniors, children, and low-income groups. Although government-funded food assistance programs are available throughout the country, they may not be enough to tackle the problem.

This project looks at the problem of food accessibility in the U.S. using the dataset made available by the U.S. Department of Agriculture. Many variables such as low access population distribution, food source availability, food insecurity, assistance programs, and socioeconomic factors will be examined to find trends and changes in food accessibility between 2010 and 2015. The goal of this project is to analyze the effect of food environment factors on food accessibility.

2 Dataset

The project will use the Food Environment Atlas Data from 9/10/2020 found [here](#) (Economic Research Service, n.d.). The dataset has been compiled by the US Department of Agriculture (USDA) Economic Research Service (ERS) for purposes of studying factors that affect food choices and the accessibility of healthy foods in communities. The dataset contains county-level information on food environment factors such as access to grocery stores/supermarkets/restaurants, local food sales, food prices, food assistance programs like SNAP (Supplemental Nutrition Assistance Program), National School Lunch Program, etc., socioeconomic characteristics, and some health/physical activities. State-level information on

food taxes, and household food insecurity are also available. More extensive documentation on each variable can be found on the webpage’s documentation section.

The dataset has been compiled using various reports from the USDA, the Bureau of the Census, the U.S. Department of Commerce, and more. Some variable values, like those from the 2010 census, were measured directly while others were estimates calculated in the source reports. The data cover the years 2006 to 2019 as separate reports published over a decade were used to create this dataset. However, the focus of the dataset is on the years 2010 and 2015. Hence, this project will also work on analysis relevant to these years.

Some of the variables that are of interest to this project are listed below. They are listed under categories because of the large number of variables.

- FIPS: Code to uniquely identify counties
- Population with low access to stores: Living more than 1 mile from a supermarket or large grocery store if in an urban area, or more than 10 miles from a supermarket or large grocery store if in a rural area.
 - Population, low access (%) (2010, 2015)
 - Low income and low access to store (%)
 - Households, no car and low access to store (%) (2010, 2015)
 - SNAP households, low access to store (%)
 - Children, low access to store (%)
 - Seniors, low access to store (%)
 - Race group, low access (%) 2015
- Stores:
 - Grocery stores, supercenters, convenience stores, SNAP-authorized stores, specialized food stores (all in ‘/1000 population’ values for 2011 and 2016).
- Restaurants: (2011 and 2016)
 - Fast-food restaurants, Full-service restaurants (in ‘/1000 population’ values).
- Food Assistance: **(2012, 2017)**
 - SNAP participants (percent pop)
 - National School Lunch Program participants (percent children)
 - School Breakfast Program participants (percent children)
 - Summer Food Service Program participants (percent children)
 - Food Distribution Program on Indian Reservations (FDPIR) Sites (2012, 2015)
- State Food Insecurity (from 2012-14 and 2015-17)

- Household food insecurity (percent, 3-year average): Households unable to provide enough food during a year for one or more household members.
- Local Foods:
 - Farmer markets /1,000 pop (2013, 2018)
 - Farms with direct sales
 - Food hubs (2018)
- Socioeconomic characteristics
 - 2010 Percentage of the county resident population that is White, Black, Hispanic, Asian, Hawaiian or Pacific Islander, 65 years or older, under 18
 - Median household income, poverty rate, and child poverty rates in 2015
 - Metro/non-metro county (labels for whether the county is metropolitan or non-metropolitan)

3 Data Acquisition and Processing

Before analysis work, I will check if any variables of interest have missing values. Initial exploration of the dataset showed that some fields have a high number of missing values which makes such variables less useful. For example, data on the WIC (Women, Infants, and Children) program had excessive missing values across the dataset. Such variables will be discarded from any analysis. Any remaining variables should be checked to see if they have the correct data type.

Next, I will work on imputing other missing values using similar/average values for the surrounding counties. If this is not possible, I will use median values for imputation. I will also check for outliers for quantitative variables.

Additionally, the dataset holds data in multiple spreadsheets. An important task then would be to join the variables of interest based on the FIPS number which are the unique identifiers for counties. After attaining a final dataframe with cleaned data, I will also work on recoding variables. This recoded variables will also have a lookup/list of descriptions for the new names.

4 Research Questions and Methodology

- 1) Which counties have the lowest accessibility to stores? Are they concentrated in certain cities/regions/states? (6 hours)

I will investigate this question by creating a geographical heatmap of low accessibility populations in the counties. The variables that give the percentage of the population

considered to have low access will be useful for charting this. I will create separate heatmaps for the years 2010 and 2015. In the heatmaps, a darker shade of color would indicate a higher percentage of the population with low access to food, meaning these areas would have a greater problem with food accessibility.

Additionally, I will work on how to filter for counties with low access to food. This categorization may help scale down the problem and focus on areas that need the most help.

2) What race and age groups are affected the most by low food accessibility? (6 hours)

- Race: To understand differences in food accessibility for different race groups, I will create a Racial Disparity Score. I will use the low accessibility population values given as a percentage of the racial group population living in the county. The disparity score will be computed by subtracting these values for the groups that have the highest or the lowest values. To visualize the results, I will use a heatmap on a geographical map of the U.S.; the darker shades would represent higher disparities and vice versa. Additionally, I will find the average of the population % with low accessibility for each group. A bar plot can be used to visualize this national statistic.

I will also store which race groups had the lowest and highest accessibility for each county. I will create a grouped bar plot to show the number of times a race group appears to have the best or the worst accessibility. For each group, there will be two bars counting the number of times the group had the lowest and the highest accessibility.

Lastly, I will also use the above-mentioned strategies after aggregating data based on states or custom territories.

- Age: First, I will look at the average low accessibility population percentages for elders and children. I will also create geographical heatmaps for the two variables to check if some counties have low food accessibility. I will compare it to the graphs for Question 1 to identify if these groups are especially worse off in some counties. Furthermore, I will create two variables that measure the difference between the overall population % with low access and the population % values for elders and age. Using summary statistics and creating a boxplot/distribution plot for the difference scores may help with investigating if there are large differences which would suggest that seniors/children face greater challenges in some counties

3) How does food insecurity level compare with the availability of vehicles? (1 hour)

For this question, I will find the average % of households with vehicles separately for metropolitan and non-metropolitan counties. To compare with the low accessibility population, I will create a side-by-side bar plot of the average population % values.

4) How much has food insecurity changed in states? (1 hour)

I will select counties that had some of the lowest and highest population percentages with low access to food, and create a table that records the percentage change in food security from 2010 to 2015.

- 5) What is the relationship between income and food access? (2 hours)
I will use a scatterplot of the median household income and low access percent population for different counties to investigate any trends.
- 6) How is the type of store available related to accessibility? How are the demographic & socioeconomic characteristics of the population and food access in the county connected? (6 hours)
I will use linear regression that will try to explain the low accessibility population percentages in a county using food environment variables such as demographic & socioeconomic characteristics dominant in that county. I will also explore other machine-learning models and approaches for this problem.
Additionally, I will use these models to investigate the relationship between the type and number of stores available in the county, and the demographic & socioeconomic characteristics information for that county.
- 7) Do food assistance programs strongly benefit counties with low food access? (4 hours)
I will compare the low access percent population with the percent population receiving benefits using a scatterplot.

5 Work plan

Week 4 (2/6 - 2/12):

- Data tidying and transforming (4 hours)
- Question 1 (3 hours)

Week 5 (2/13 - 2/19):

- Question 1 (3 hours)
- Question 2 (6 hours)

Week 6 (2/20 - 2/26):

- Question 3 (1 hour)
- Question 4 (1 hour)
- Question 5 (2 hours)
- Question 6 (4) hours)

Week 7 (2/27 - 3/5):

- Question 6 (1 hour)

- Question 7 (4 hours)
- Presentation prep and practice (4 hours)

Week 8 (3/6 - 3/12): *Presentations in class on Thurs 3/9.*

- Presentation peer review (1.5 hours)

Week 9 (3/20 - 3/26):

- Poster prep (4 hours)

Week 10 (3/27 - 4/2): *Poster Draft 1 due Monday 3/27. Peer feedback due Thursday 3/30.*

- Peer feedback (2.5 hours)
- Poster revisions (2 hours)

Week 11 (4/3 - 4/9): *Poster Draft 2 due Monday 4/3. Final Poster due Saturday 4/8.*

- Poster revisions (1 hour).

Week 12 (4/10 - 4/16):

Week 13 (4/17 - 4/23): *Finish all project work*

- Draft blog post (5 hours).

Week 14 (4/24 - 4/30): *Blog post draft 1 due Monday 4/24. Peer feedback due Thursday 4/27. Blog post draft 2 due Sunday 4/30.*

- Peer feedback (2.5 hours)
- Blog post revisions (2 hours)

Week 15 (5/1 - 5/7): *Final blog post due Tuesday 5/2.*

- Final presentation prep and practice.

Final Exam Week (5/8): *Final Presentations during final exam slot, Monday May 9th 3:20-6:40pm.*

References

Economic Research Service, U. S. Department of Agriculture (USDA). n.d. "Food Environment Atlas."