Final Report

Team 39 - Coast 2 Coast

Sandra Azubuike

Paulina Cucalon

M. Alejandra Gomez

Christina Liu

Zaire Wade

Table of Contents

Introduction

Context

Food banks have been a staple of citizenship and generosity in America for decades. As we all know, the COVID-19 pandemic has caused a public health and economic crisis, and the effects of these are extensive. The repercussions will include an increase in hardship for those populations who were already vulnerable, including the number of people experiencing food insecurity. Food insecurity is defined as “the state of being without reliable access to a sufficient quantity of affordable, nutritious food.”

According to the United Nations, COVID-19 could double the global food insecurity rate! In 2019, the countrywide average U.S. food insecurity rate was 10.5%. At the state level, the projected rate of food insecurity among the overall population for 2020 ranges from 12.0% in North Dakota to 24% in Mississippi, and among all counties, it ranges from as low as 8.6% to a high of 34%.

As the pandemic continues, unemployment has ascended, and demand has spiked at food banks and food pantries across the United States. This will likely result in worse health outcomes for the general population who rely on food banks, and more so in times of crisis. If people cannot obtain enough nutritious food to eat, their health tends to decline. This is intensified by the current pandemic, which tends to target those with poorer health.

COVID-19 has highlighted the necessity for food banks in many communities in America. People have reported waiting for hours in line to reach their local food bank and find it to be out of resources, making them have to choose between trying again at the next closest location or going to bed hungry.

All of this sparked an interest in our team to see how well distributed the network of food banks is in the US. After extensive research and analysis, we decided to focus on the following question: **On the U.S. County Level, is Distance to Food Banks a Predictor for Food Insecurity?**

Project Definition

The team decided to look into what factors can predict food insecurity in a county in the U.S. Our main focus was the distances to food banks, however, we also looked at unemployment and poverty rates, as well as education levels in all counties. We believe this project to be important and applicable especially right now with the pandemic because of the rise in unemployment and poverty rates, and also the risk of more students dropping out of schools due to lockdowns and online learning protocols.

Our solution is a linear regression model showing the significance of all these factors as predictors of food insecurity for all counties in the United States. This will show how much importance should be put onto each of the factors to better recover from the pandemic's repercussions.

Data Analysis & Computation

Datasets + Data Wrangling & Cleaning

Once we defined our question of interest, we realized we had an empirical challenge on our hands. A single dataset would not suffice to answer our question and to obtain the county information needed for our analysis, our research encompassed numerous data sources.

Among which were the following:

	Year	Source	Link	Date Retrieved
Food Insecurity Rate	2018-2020	Feeding America – *The Impact of Coronavirus on Food Insecurity*, Map the Meal Gap (October, 2020).	FI Rates	December 8, 2020
FIPS Codes Dataset	2010	Data.world retrieved from the U.S. Government Census (2010).	FIPS Codes	January 20, 2021
Food Bank Locations	Current	Feeding America Organization	FB Locations	January 19, 2021
County Demographics	2019	U.S. Department of Agriculture – Economic Research Service (USDA)	Demographics	December 12, 2020
County Distance Dataset (100 miles)	2010	National Bureau of Economic Research (NBER)	Distance Data	January 18, 2020

The county data/information needed for our investigations was as follows:

- Food Insecurity Rates per county (Pre-covid and Covid Data).
- Food Bank Presence.
- Distance to the nearest county with a Food Bank.
- County Demographics.

We began our analysis by retrieving the food bank locations (via web-scraping) and food insecurity rate data from the years 2018 and projected 2020 (by online request) from the nonprofit organization, Feeding America. Feeding America's food bank network distributes and secures over 4.3 billion meals each year throughout the country (Hake, M., 2020).

Two data frames were created to store the datasets received from the nonprofit organization, in which one contained foodbank county location data and another food insecurity county rates.

Our foodbank location data frame only included zip codes, so we had to find the corresponding FIPS (unique county identifier) to each zip code to facilitate the merge with the food insecurity data frame. And for that, we used a “Zip Code to FIPS” dataset from the data.world website, and merged it to our Food Bank locations data frame.

Multiple counties can share the same zip code, and this was the case with several of our foodbank locations, which shared the same zip code but were in different counties. To determine the FIPS codes we had to phone the particular food bank locations and obtain the correct county information.

To determine the distance of a county with no foodbank to the nearest Food Bank location, we retrieved from the National Bureau of Economic Research (NBER) their distance between counties' dataset and merged it to our food insecurity data frame.

Lastly, our final data frame was created to support our geo-temporal analysis and to contain extra features such as unemployment rate, household income, poverty rate, and education level retrieved from the U.S. Department of Agriculture.

The figure below illustrates the final data frame creation process:

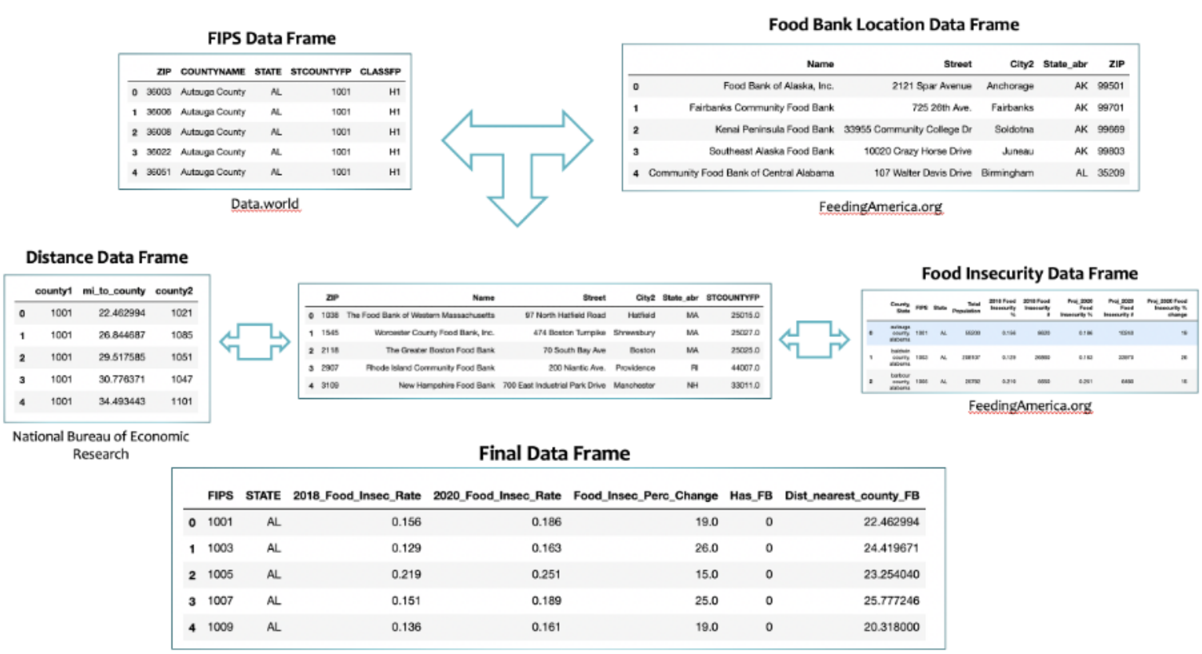

Figure 1. Data frame creation process

Our final data frame consisted of approximately 3k observations and included the following county features:

- Food Insecurity Percent Change (2018-2020)
- Has a Food Bank? Yes-1 / No-0
- Distance to the nearest county with Food Bank (within 100 miles)
- Demographics: Poverty Rate, Unemployment Rate, Education Level

Exploratory Data Analysis

To begin exploring if the distance to food banks is a predictor of food insecurity on the county level, it is essential to fundamentally investigate our datasets. First, an exploratory data analysis was conducted to summarize the main characteristic of food insecurity data. The key variables of our food in food insecurity dataset included: [County, State], FIPS, total population, and food insecurity rates over the years (2018-2020), annual unemployment change, and total child population.

When exploring if the distance to food banks is a predictor of food insecurity on the county level, it is essential to visualize the number of foodbanks compared to a county’s population. In our first visualization seen below in Figure 1, we look in the Greater Boston Area through a heat map lens.

The first layer of this map shows populations' variation, where large populations are represented by “warmer” colors.

The second layer of this map represents the number of food banks in which the “bubbles” show that number. Once a bubble is clicked, the pins will show the exact location of food banks. Figure 2 visualization shows areas with higher populations tend to have a higher number of food banks, seen as those areas with warmer colors have bubbles with a greater count.

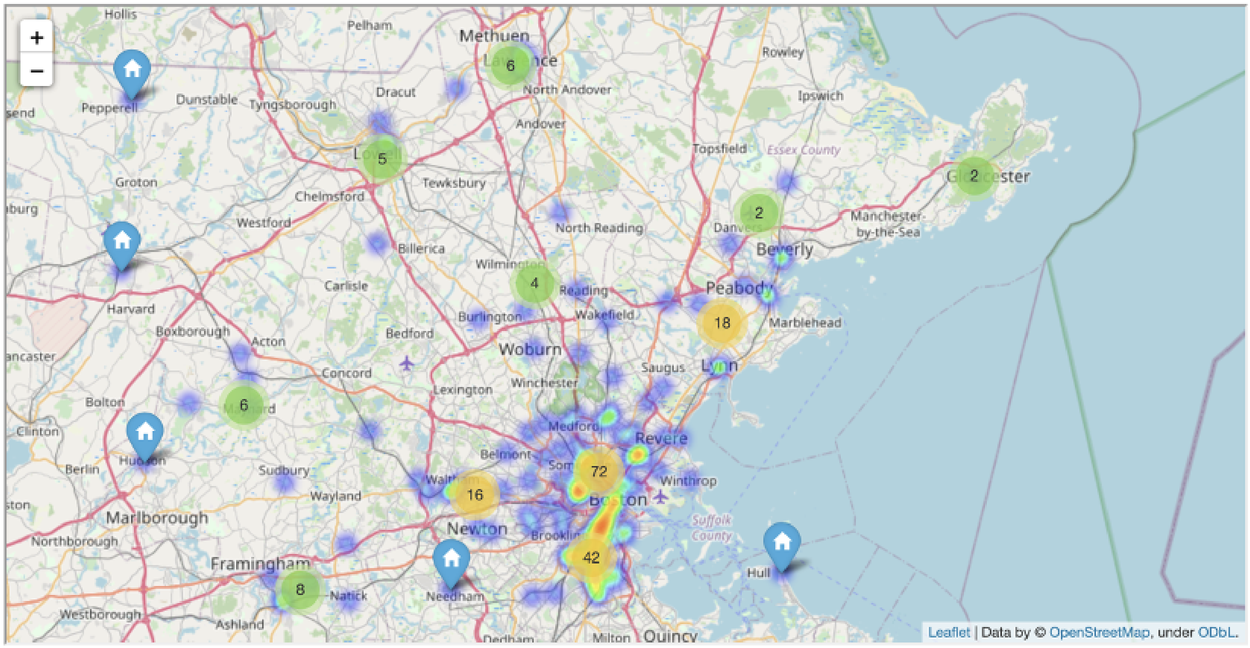

Figure 2. Boston Heat Map

Figure 2 visualization shows areas with higher populations tend to have a higher number of food banks, seen as those areas with warmer colors have bubbles with a greater count.

In our next visualization, Figure 3, we take a look at food insecurity in the US. In this [dashboard](), large bubbles show counties with food banks within 100 miles, while smaller ones represent counties with food banks within a 100-mile radius

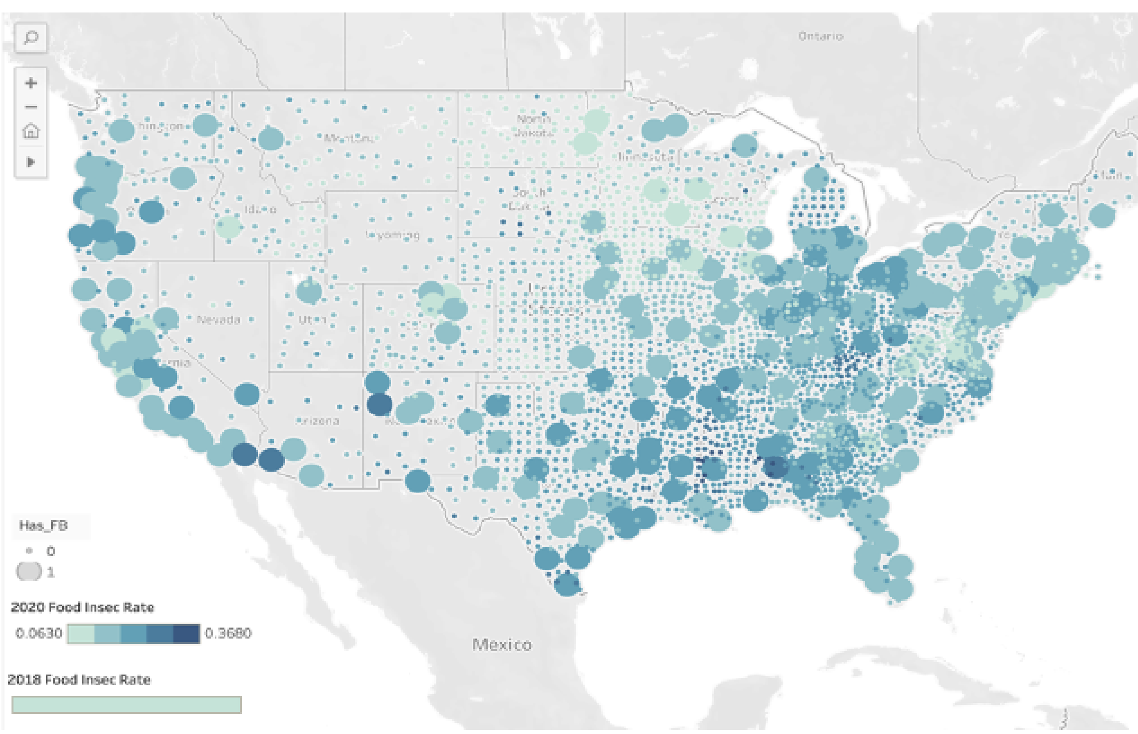

Figure 3. Screenshot of Tableau Dashboard

The range of colors represents fluctuating food insecurity rates. Figure 3 includes the point's FIPS code, whether they have a foodbank within 100 miles or not (0 or 1), and their particular food insecurity rate. This dashboard can help us see that the east coast has more food banks, while the Mountain states seem to have the least food banks. Overall, the southeast has a lower food insecurity rate, as most points are darker colors, while the Northeast has the bluest points showing higher food insecurity rates.

Statistical Analysis & Predictive Modeling

Linear Regression Model

Linear regression is used to model the relationship between a scalar variable and one or more explanatory variables by fitting a linear equation. To answer our question, we are using a simple linear regression to model the food insecurity rate change from 2018 to 2020 as a function of distance to the nearest food bank in miles in United States counties. Our scalar variable is distance and our explanatory variable is food insecurity. If a food bank exists in a county, the distance would be 0. Each point in our scatter plot represents a county. In this simple regression problem, we are looking at a model:

$$y = B_0 + B_1*x$$

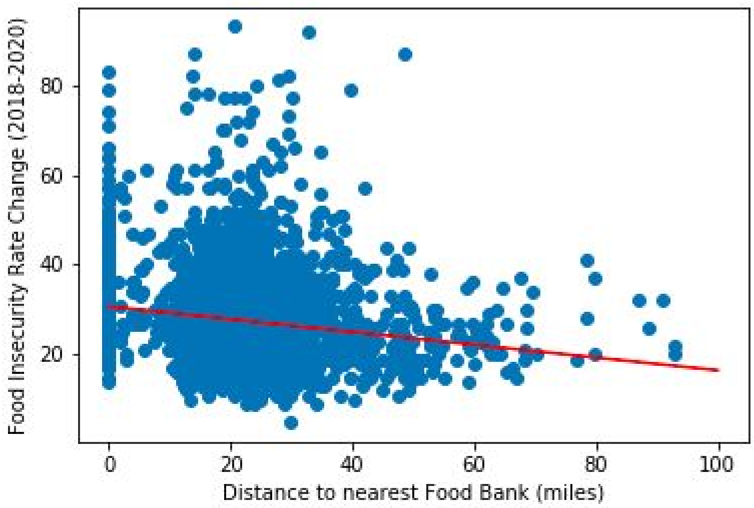

Figure 4. Linear regression model

The slightly downward slope of the line indicates there is a negative correlation between food insecurity rate change and distance to the nearest food bank. So, we would expect to see less food insecurity in areas further away from a food bank. The coefficients we extracted from this model are:

B_0 = 30.73083653626483

B_1 = -0.14210314

As B_1 is close to 0, we cannot conclude a strict correlation from linear regression, which leads us to look at different models. Using linear regression assumes the relationship between food insecurity rate change from 2018 to 2020 and distance to food banks is linear, however, through this analysis we realized it is not linear.

Linear Probability Models

Linear probability models are a special case of linear regression models, as they include a binary dependent variable. We were interested in estimating the outcome of a variable that can take on two possible values. In this case, our dependent variable was if a county contains a food bank. We decided to use this model to interpret the county demographic data we had obtained: poverty rate, unemployment rate, and education. So we wanted to see how the demographic of a county affects the probability that a county will contain a food bank.

We first modeled the probability of having a food bank with poverty and unemployment rate in 2019.

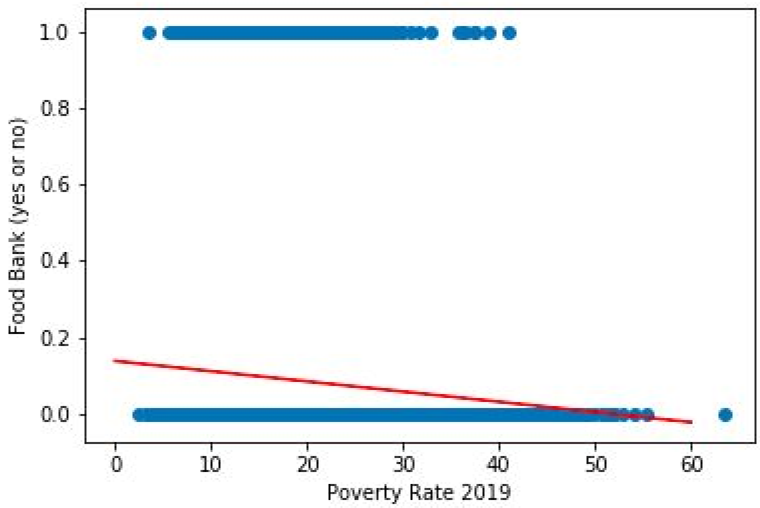

Figure 5. Linear probability model for poverty rate

The coefficients obtained for the poverty model are:

B_0: 0.13786498304023248

B_1: -0.00267764

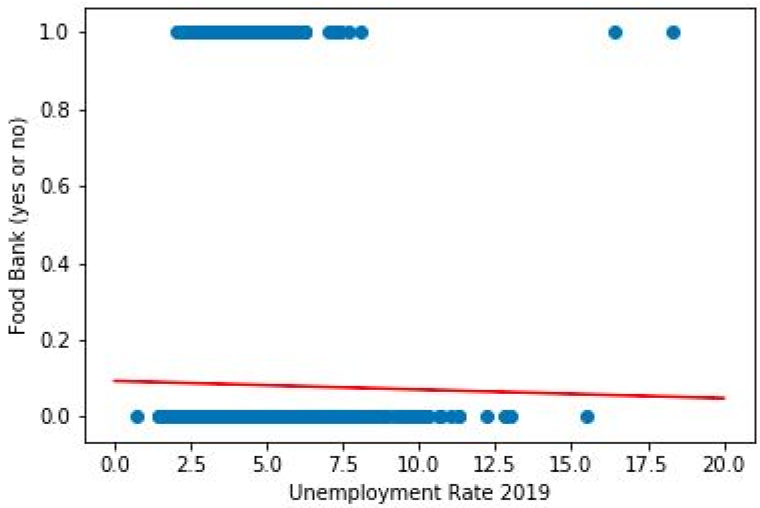

Figure 6. Linear probability model for unemployment rate

The coefficients obtained for the unemployment model are:

B_0: 0.09326605513139503

B_1: -0.00224568

The downward slope of the line shows there is a negative correlation, which suggests that the further away a county is from a food bank, the higher the poverty/unemployment rate, but the coefficients are so low that there is not a strong correlation.

We then modeled the probability of having a food bank with the variables for education levels: looking at the percentage of adults with less than a high school diploma, percentage of adults with a high school diploma only, percentage of adults with an associate's or some college degree, and percentage of adults with a bachelor's degree or higher.

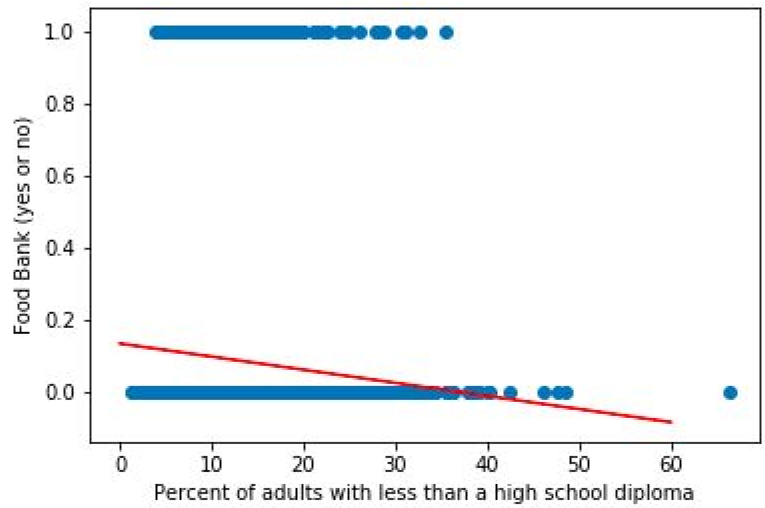

Figure 7. Linear probability model for percentage of adults with less than a high school diploma

The coefficients obtained for the percentage of adults with less than a high school diploma model are:

B_0: 0.1332009633830952

B_1: -0.00363783

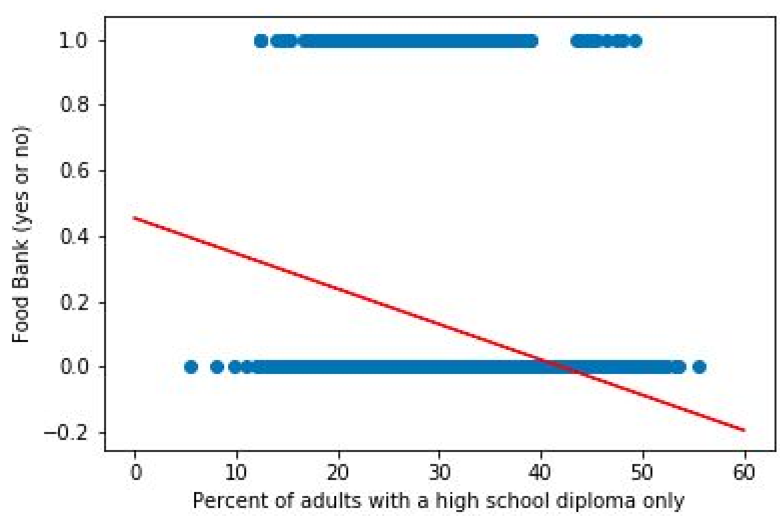

Figure 8. Linear probability model for percentage of adults with a high school diploma only

The coefficients obtained for the percentage of adults with a high school diploma only model are:

B_0: 0.45453011022217615

B_1: -0.01079596

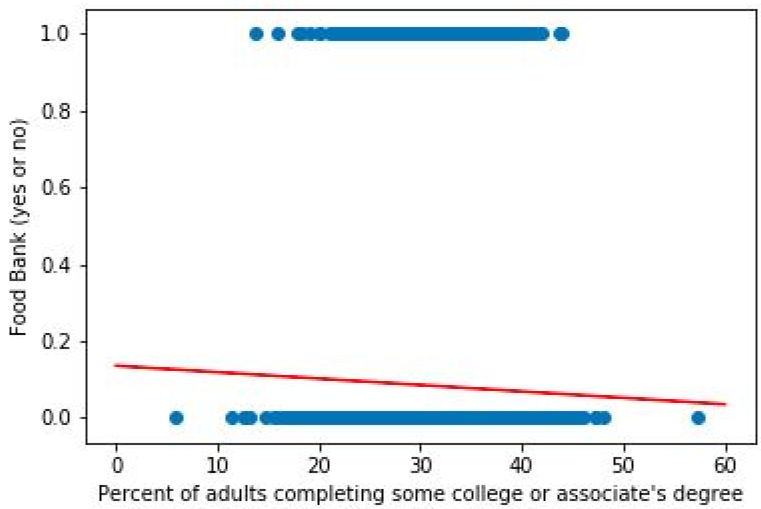

Figure 9. Linear probability model for percentage of adults with a college or associate's degree

The coefficients obtained for the percentage of adults with an associate's or some college degree model are:

B_0: 0.1355588320672214

B_1: -0.00166693

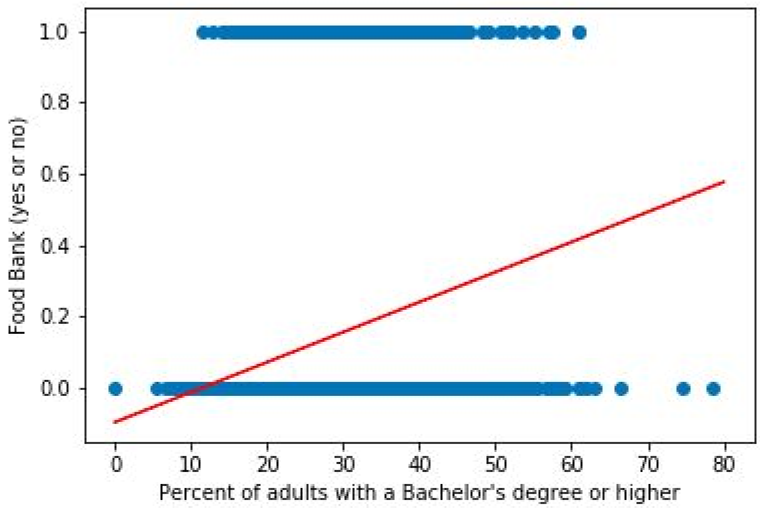

Figure 10. Linear probability model for percentage of adults with a Bachelor's degree of higher

The coefficients obtained for the model of the percentage of adults with a Bachelor's degree or higher are:

B_0: -0.09738444236960447

B_1: 0.00842924

The downward slope of the line shows there is a negative correlation for three of the education categories, but the coefficients are so low that there is not a strong correlation. The interesting point here is that counties with a higher percentage of people with bachelor's degree or higher have a higher probability of having a food bank within the county or nearby.

In terms of analysis, we would have wanted to explore more than the linear probability model. Linear probability models are a good first step and are easy to interpret, but to get a more comprehensive model, we would have wanted to use a logistic model for more statistically accurate results. However, linear probability models are a good first step in modeling because of the easy to interpret coefficients, since it is a form of linear regression. If we had more time, we would have investigated these relationships in the form of a logit or probit regression for more accuracy.

Linear Regression with Interaction Variable

After looking at the linear probability models, we decided to add another variable to our original linear regression model -- the percentage of adults with a Bachelor's degree or higher. We took the median of the percentage of people with a Bachelor's degree or higher and then assigned a binary variable to all the counties, one higher than the medium percentage of people with a Bachelor's degree or more and the other lower than the medium percentage of people with a Bachelor's degree or more. In this more complex linear regression model, we are looking to model a simple linear regression with an interaction term, which results in modeling an equation like this:

$$y = B_0 + B_1*x + B_2*y + B_3*x*y.$$

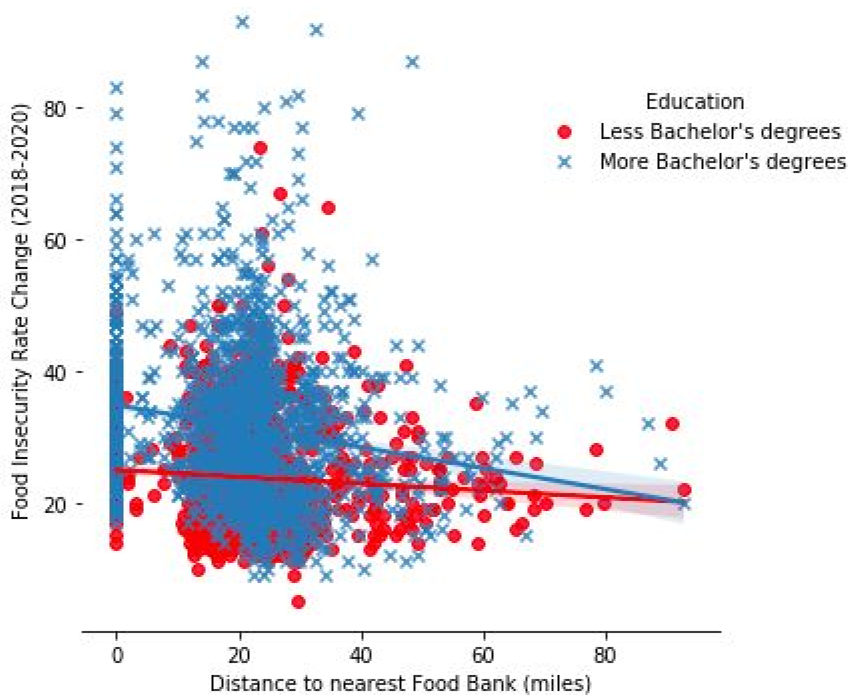

Figure 11. Linear regression model with interaction variable

Counties with more college graduates (people who received a Bachelor's degree or higher) tended to have a lower food insecurity change before and during COVID the further away they were from a food bank.

Counties with fewer college graduates (people who received an associates' degree, a high school diploma, or less than a high school diploma) tended to see less of a change in food insecurity from before and during COVID the further away they were from a food bank.

The difference in slope between counties with more college graduates and those with fewer college graduates suggests that food banks have less of an impact on the latter since the slope is less pronounced.

We used ordinary least squares regression, which is a type of linear least squares method for estimating the unknown parameters in a linear regression model. Basically, given a regression line through the data OLS calculates the distance from each data point to the regression line, square it, and sum all of the squared errors together. This is a very good regression to use for multiple variables. Our OLS regression results show how factors such as distance to the nearest food bank and college education do affect food insecurity.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.139
Model:                            OLS   Adj. R-squared:                  0.138
Method:                 Least Squares   F-statistic:                     167.4
Date:                Sat, 06 Feb 2021   Prob (F-statistic):          1.28e-100
Time:                        17:29:09   Log-Likelihood:                -11692.
No. Observations:                3118   AIC:                         2.339e+04
Df Residuals:                    3114   BIC:                         2.342e+04
Df Model:                           3
Covariance Type:            nonrobust
==========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                    24.9568      0.639     39.076      0.000      23.705      26.209
Dist_nearest_county_FB   -0.0521      0.026     -1.997      0.046      -0.103      -0.001
more_college              9.8422      0.808     12.176      0.000       8.257      11.427
dist*more_college        -0.1073      0.033     -3.240      0.001      -0.172      -0.042
==============================================================================
Omnibus:                      842.579   Durbin-Watson:                   0.970
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2725.504
Skew:                           1.353   Prob(JB):                         0.00
Kurtosis:                       6.696   Cond. No.                         148.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 12. OLS Regression Model Results

- The "const" coefficient (B_0) is the food insecurity of a county with a food bank.

- The “Dist_nearest_county_FB” coefficient (B_1) shows the rate of food insecurity change per mile for counties that do not have a food bank. The negative coefficient postulates that there is a slightly lower chance of having food insecurity the further you are from a food bank.
- The “more_college” coefficient (B_2) is the amount that food insecurity increases with communities with more college-educated people with a food bank. The positive coefficient suggests there is a strong chance of having a higher rate of food insecurity in counties that have food banks and a higher number of college-educated adults.
- The “dist*more_college” coefficient (B_3) shows the rate of food insecurity change per mile for counties that do not have a food bank but do have more college-educated adults. The negative coefficient postulates that there is a slightly lower chance of having food insecurity the further you are from a food bank even in counties with higher college graduates.

Our results indicate that while a college education is a significant factor in predicting food insecurity, the distance to the nearest food bank in counties without food banks is not a significant predictor of food insecurity.

Description of Application

The application of this project is presented in a webpage deployed on Herokuapp & GitHub (https://c2c-team39.herokuapp.com/). The webpage includes a home page, data visualization, analysis, takeaways, and about the team page. Beginning with the home page, a user will view a brief project overview statement, three quick links to other tabs, and an in-depth overview of the project’s methodology.

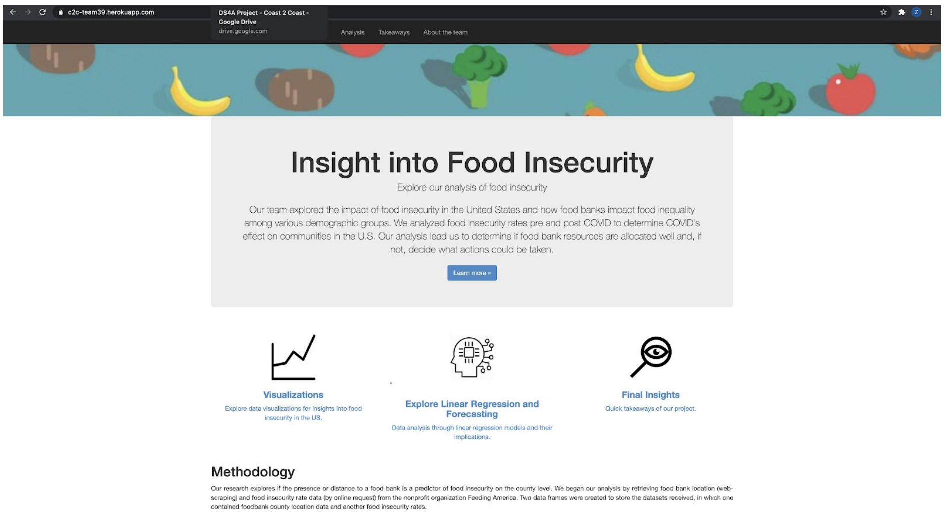

Figure 13. Home page of website

Next, the data visualizations tab provides users with the project's "Boston Heat Map" and an embedded Tableau interactive dashboard titled "Food Bank Locations & Food Insecurity Rate (2020)". Both visualizations provide additional analytical insights within the project.

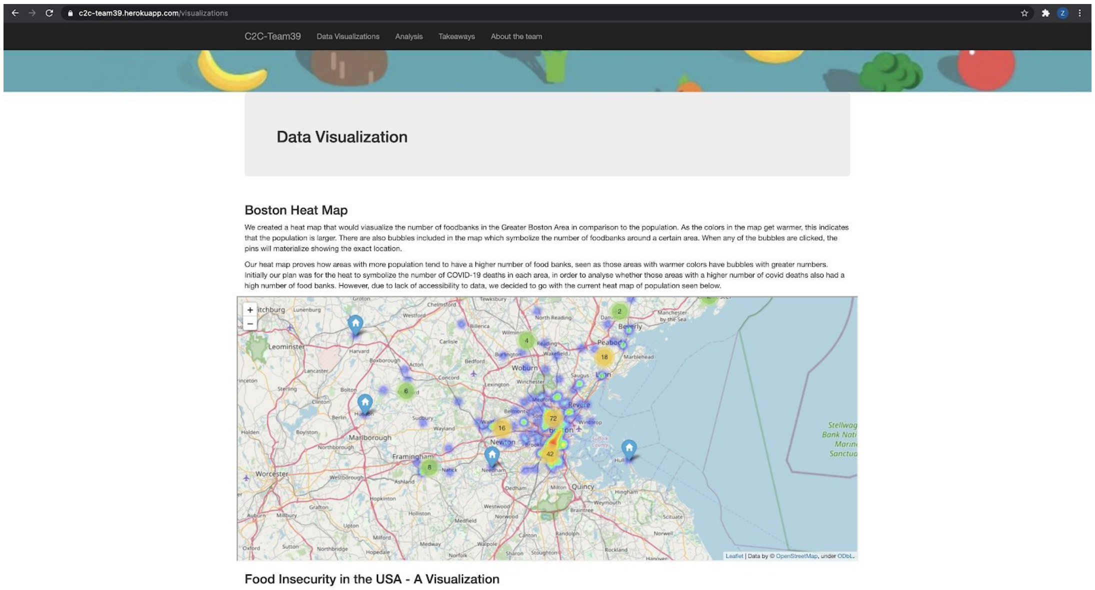

Figure 14. Data Visualization page of website

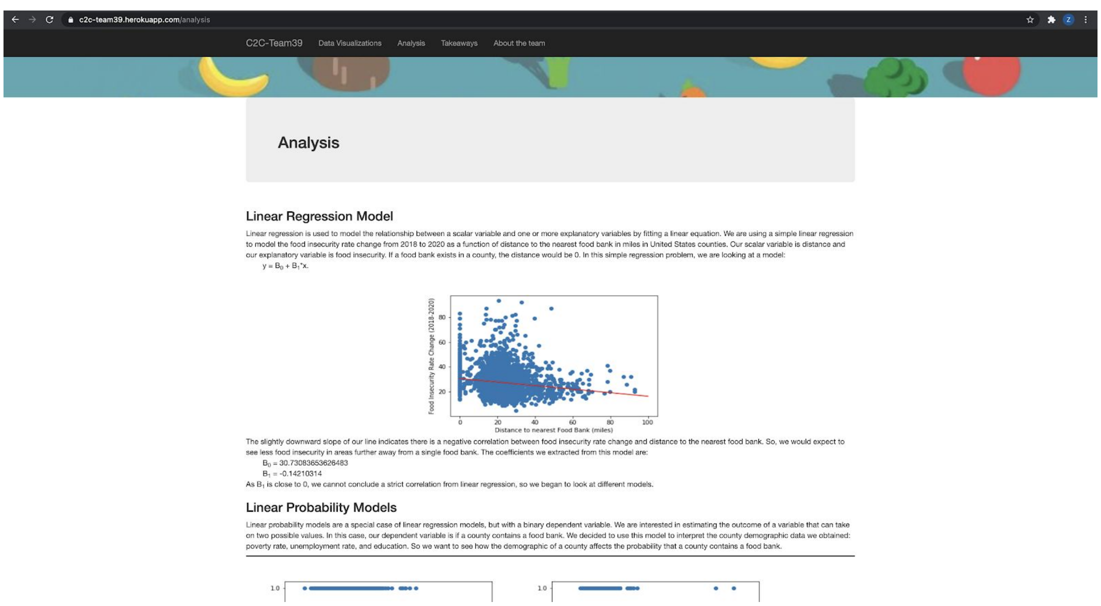

Figure 15. Analysis page of website

The analysis tab gives any users a description of the project's statistical models such as linear regression, linear probability, and interaction models. Our takeaway tab provides this project's purpose and details the various complications that came up during the project's research. In addition, the takeaways tab includes the implication of what we discovered and future steps. Lastly, we include an "About the Team" page to highlight all team members who took part in this project effort and links to their respective LinkedIns.

Takeaways

Why did we choose this problem?

Our team wanted to see if there was any impact on food insecurity due to COVID-19. We are aware that food insecurity is a significant issue in the US and around the world. However, the team also believes that there is not enough effort in working to improve this problem, so we decided to focus our project on trying to shine a light on food insecurity in the US and the different factors that affect it.

The Implications of What We Found

Our findings postulated that food insecurity decreased the further away a county was from the nearest food bank. The demographics we chose such as education, poverty level, and unemployment were based on the availability of data we could find related to food insecurity. Graphs with a steep slope and a strong correlation such as the percent of adults with a bachelor's degree vs food bank show that there is a greater chance of finding a food bank in counties with more college-educated adults. This was our strongest correlation out of all the demographic variables tested. Additionally, our interaction model shows that there is a higher likelihood of food insecurity in counties with a food bank and with more bachelor or higher degree holders. It may seem that the results show that food banks are not making an impact on food insecurity, however, if you consider that food banks tend to be located in cities this can explain the seemingly contradictory results of counties with food banks.

Larger cities and urban areas tend to have more college-educated individuals and more inequality as well, according to the Gini coefficient which measures wealth dispersion within a particular area. Large cities tend to have more degree holders and also a higher cost of living. This means that college-educated adults may have a higher probability of affording housing and having a higher income, but an individual with a lower income may not be able to afford the same housing they could have in a more rural, lower cost of living area and thus would have to choose where to allocate their resources. Money may go to paying rent instead of food leading to more food insecure individuals with access to a food bank even though many others in the population may be much more financially stable.

Future work

In the development of this case, we realized there are several more projects that can be performed to further investigate the issue of food insecurity, especially as it pertains to food banks. There are additional projects that would provide more insight regarding food inequality in counties that have food banks and more urban areas which are listed below:

- Two county comparison: One way we would hope to compare the effects of food banks is looking at two counties with similar socioeconomic indicators in different parts of the country and determine how each has fared with food insecurity from pre-COVID until now. One county would have a food bank and the other wouldn't so we can better study the effect of having the food bank with controls for other forms of inequality.
- In-depth Urban area: Another idea for a project to further study the impact of food banks would be to study how food insecurity changes the further you get away from

a food bank within a particularly low-income neighborhood in a large city. This could show a more relevant breakdown of food insecurity among areas where people need access to food most and how other factors such as transportation can affect access to food banks and food insecurity.

Final Thoughts

We started our project by exploring the relationship between food insecurity and food banks and how COVID has impacted food insecurity. After devising methods to explore this relationship we have realized a multitude of factors affect food insecurity and while food banks may help reduce food insecurity, there is still a lot of research that needs to be done on how best to utilize food banks and how to reduce the rate of food insecurity. Our research for this project highlighted the lack of information on food banks and many other factors that lead to food insecurity. Though our data at first glance suggests that food banks do not have a significant impact on food insecurity, when we look closer at the areas where food banks are located, we are likely to see other forms of inequality. We hope that there is further research into food banks and food inequality to tackle the serious impact of food insecurity so that more people have access to a basic human need.

References

Brownlee, J. (2020, August 14). Linear regression for machine learning. Retrieved from https://machinelearningmastery.com/linear-regression-for-machine-learning/

Cohen, S. (2020, December 07). Millions of hungry Americans turn to food banks for first time. Retrieved February 13, 2021, from https://www.pbs.org/newshour/economy/millions-of-hungry-americans-turn-to-food-banks-for-first-time

Hake, M., Dewey, A., Engelhard, E., Gallagher, A., Summerfelt, T., Malone-Smolla, C., . . . Gundersen, C. (2020, May 19). The Impact of the Coronavirus on Local Food Insecurity. Retrieved from https://www.feedingamerica.org/sites/default/files/2020-05/Brief_Local%20Impact_5.19.2020.pdf

Perraillon, M. C. (2019). Week 12: Linear Probability Models, Logistic and Probit. Retrieved from https://clas.ucdenver.edu/marcelo-perraillon/sites/default/files/attached-files/week_12_lpn_logit.pdf

Urbanization: Expanding opportunities, but deeper DIVIDES | UN DESA Department of economic and social affairs. (2020, February 21). Retrieved February 13, 2021, from https://www.un.org/development/desa/en/news/social/urbanization-expanding-opportunities-but-deeper-divides.html#:~:text=Economically%2C%20inequality%20is%20generally%20greater,more%20unequal%20than%20smaller%20cities.

Weiss, E., & García, E. (2020, September 10). COVID-19 and student performance, equity, and U.S. EDUCATION policy: Lessons From PRE-PANDEMIC research to inform RELIEF, recovery, and rebuilding. Retrieved from https://www.epi.org/publication/the-consequences-of-the-covid-19-pandemic-for-education-performance-and-equity-in-the-united-states-what-can-we-learn-from-pre-pandemic-research-to-inform-relief-recovery-and-rebuilding/