Alexandra Koken

Foundations of Data Science

610:560:90

Final Project: How Food Deserts Compare to Food Waste

**Introduction: The Problem of Food Waste and Food Insecurity**

Food insecurity and food waste are two problems in the United States that impact the most basic need for survival. Covid-19 has created more of both. Prices are continuing to rise, wages are not, and therefore this problem is likely to persist or worsen. The purpose of this proposal is to explore the idea that the United States does *not* actually have a food insecurity problem, but a policy and distribution problem. Also, that the problem can be fixed with common sense policies and regulations. The importance of this goes beyond meeting basic needs, but also the costly problems that come along with poor nutrition, and the mental health issues created and compounded by food insecurity.

**Background: Food Abundance, Food Insecurity & Food Deserts**

Food is abundant in the United States. Portion sizes are relatively large, shelves in stores are stocked, and restaurants can be found everywhere. One of the main ways we have been able to reflect this idea statistically is through food waste. Food waste is the amount of food that stores and consumers throw away. The factors that contribute to food waste include spoilage, past an expiration date, and unappealing appearance to consumers. There is more food in U.S. landfills than any other waste item[[1]](#footnote-1). For as much research as there is on waste, there is a similar amount about food insecurity.

Food insecurity is a term that covers everything from individuals having anxiety about having enough food or enough money for food to missing meals[[2]](#footnote-2). Food insecurity is a problem in the United States that disproportionately impacts children and marginalized groups. Coivd-19 exacerbated the food insecurity problem. Rising prices, stock shortages, and school closures all had a dramatic effect for those already impacted. In 2020, 38 million people lived in food insecure households, and over half a million children did not have enough food[[3]](#footnote-3). Some of the factors such as income and household size, or single parenthood are expected. These factors are also often attributed to food deserts, or areas where *access* to food is limited. Limited access happens when there is no local supermarket (Caporusico, J., 2020). Individuals are left with difficult choices for food, and often times these choices have negative consequences for their financial situation, their health, or both.

Food deserts are areas where *access* to food is limited. This occurs in both urban and rural environments. There are some communities that do not have local access to larger supermarkets and grocery stores and are forced to shop at smaller markets and corner stores. Pricing can be arbitrary in these shops and can be up to 30% higher than in a supermarket. This forces communities to live off of less food, or higher density, lower nutrition foods (Caporusico, J., 2020). Individuals are left with difficult choices for food, and often times these choices have negative consequences for their financial situation, their health, or both.

**Hypothesis, Questions, and Data**

Given the contrast between food abundance and food insecurity in the United States, the question becomes: is there really a problem with food insecurity, or are we ignoring the resources we have to correct it? To address this issue, I will explore data comparing food waste to food insecurity, comparing how much food waste there is to the number of people who are food insecure. I will also attempt to narrow down areas considered food deserts and find commonalities among them in order to understand what creates them, how these characteristics compare to individuals who do not live in food deserts yet are still food insecure. Additionally, I would like to explore the health impacts of food deserts and food insecurity to demonstrate any cost advantage to solving the problem.

Planned data to test the hypothesis included data from the Consumer Price Index and the Agricultural Commodity Price Index, to help demonstrate cost issues. Kaggle has several datasets that offer data on food waste and food insecurity. Those were to be used as a comparison point to determine whether or not there are enough resources already available to solve the problem. There is also a report from the World Food Program that has links to data demonstrating health issues relating to income and food. Links to all of the data sets that have already been located are provided in the next section.

Much of the data proved to be difficult to use. For example, some Kaggle datasets were missing too much data to filter out anything of relevance. There is also a disconnect between the way food waste data is reported and the way food insecurity data is reported. Food insecurity is reported by region or by population/number of people. It is either not specific enough or too specific. The Food and Agriculture Organization of the United Nations has excellent data on food loss and waste. However, the data is categorized by stage, and would therefore require a lot of steps to create a comparison that time constraints did not allow for.

The focus shifted to understanding the overlap between poverty and food deserts, to try to understand why some live in poverty but not food deserts and the other way around. To do this, I obtained a data set from the U.S. Census that offered information on how many people were in the Snap program (government food assistance) throughout the United States. The data set contained data for every county in each state, and was mapped by state. This decision was made to give an overhead view of states with the highest number of people with food insecurity. States are important because part of the hypothesis is that there may be a policy issue. If that is the case, either individual states could enact effective policy, or the aggregate of state data would be more useful at the federal level.

The second step was to plot the county data for all of the counties in the United States where people were food insecure. Data for this plot was obtained through the USDA. This data set was very large and returned an error in R that it was over the memory size. To remedy the problem, the dataset was loaded in Excel to reduce the columns down to just the three that would be required. It was then filtered for only counties where people experienced food insecurity at 1 mile away from a supermarket in urban areas, and 10 miles away in rural areas. This was selected as the most broadly applicable of categories, as it did not account for any other demographic factors. The data was then filtered for unique counties, and resulted in 2084 rows of data. It took a memory reset and several restarts of R to get R to accept the data and not return an error. One last try got all the code to work, however, the map did not load, and neither did the next line. Over one hour later, the map appeared. It is a bit distorted because after trying to resize it froze again.

**Plots**

Map

Description automatically generated

This map depicts the states with the highest percentage of the population receiving food assistance benefits by state. If poverty and low income are pointed at as common factors among those living in food deserts, then this data should line up well with a map of those who are food insecure.

Map

Description automatically generated

This plot demonstrates that there is little correlation between food deserts and receiving Snap benefits. California and Texas, for example, have lower numbers of food deserts, and yet some of the highest Snap benefits percentages in the county. Another qualitative choropleth map would have been clearer, here, however the limitations of the data available made this the next best choice.

**Code**

Code used to create the plot “Snap Benefits Received as a Percentage of Total Population by State”:

install.packages("dplyr")

Installing package into ‘C:/Users/alexa/OneDrive/Documents/R/win-library/4.1’

(as ‘lib’ is unspecified)

also installing the dependency ‘generics’

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/generics\_0.1.1.zip'

Content type 'application/zip' length 76274 bytes (74 KB)

downloaded 74 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/dplyr\_1.0.7.zip'

Content type 'application/zip' length 1345496 bytes (1.3 MB)

downloaded 1.3 MB

package ‘generics’ successfully unpacked and MD5 sums checked

package ‘dplyr’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\alexa\AppData\Local\Temp\Rtmpwl1HQd\downloaded\_packages

> #create data frame

> bdf<-data.frame(character(), integer(), integer(), stringsAsFactors = FALSE)

>

> total\_population<- nrow(Population)

Error in nrow(Population) : object 'Population' not found

> total\_pop<- nrow(total\_pop)

Error in nrow(total\_pop) : object 'total\_pop' not found

> total\_pop<-nrow(total\_pop)

Error in nrow(total\_pop) : object 'total\_pop' not found

> total\_pop<-nrow(Total\_Pop)

Error in nrow(Total\_Pop) : object 'Total\_Pop' not found

> PercentOn<-(Received/Total\_Pop)\*100

Error: object 'Received' not found

> PercentOn<-(Received/Total\_Pop)\*100

Error: object 'Received' not found

> b<-read.csv("csnap.csv")

> head(b)

Total.Pop Total.Pop.MOE Reveived.SNAP Received.SNAP.MOE No.Snap

1 1897576 10370 247490 8214 1650086

2 252199 3658 25988 2450 226211

3 2670441 12014 251413 7528 2419028

4 1163647 8530 127488 5875 1036159

5 13157873 23844 1107229 17691 12050644

6 2235103 10026 153588 7498 2081515

MOE.No.Snap id Geographic.Area.Name Percent.of.Pop

1 11159 0400000US01 Alabama 13.042429

2 4303 0400000US02 Alaska 10.304561

3 13441 0400000US04 Arizona 9.414662

4 10654 0400000US05 Arkansas 10.955900

5 27013 0400000US06 California 8.414954

6 10595 0400000US08 Colorado 6.871630

>

> bdf<-data.frame(character(), integer(), stringsAsFactors = FALSE)

> state\_base <-ggplot(data=states, mapping=aes(x=long, y=lat, group=group))+ coord\_fixed(1.3) + geom\_polygon(color="black", fill="gray")+ggtitle("SNAP as a Percentage of Population by State")

Error in ggplot(data = states, mapping = aes(x = long, y = lat, group = group)) :

could not find function "ggplot"

> library(ggplot)

Error in library(ggplot) : there is no package called ‘ggplot’

> library(ggplot2)

> state\_base <-ggplot(data=states, mapping=aes(x=long, y=lat, group=group))+ coord\_fixed(1.3) + geom\_polygon(color="black", fill="gray")+ggtitle("SNAP as a Percentage of Population by State")

> colnames(b)<-<-c("Total\_Pop", "Total\_Moe", "Received", "Received\_MOE", "No\_Snap", "No\_Snap\_MOE", "id", "State", "Percentage")

Error: unexpected assignment in "colnames(b)<-<-"

> colnames(b)<-c("Total\_Pop", "Total\_Moe", "Received", "Received\_MOE", "No\_Snap", "No\_Snap\_MOE", "id", "State", "Percentage")

>

> stco<-merge(states, b, by="State")

>

> head(stco)

[1] State long lat group order

[6] subregion Total\_Pop Total\_Moe Received Received\_MOE

[11] No\_Snap No\_Snap\_MOE id Percentage

<0 rows> (or 0-length row.names)

> b$State<-tolower(b$State)

> head(stco)

[1] State long lat group order

[6] subregion Total\_Pop Total\_Moe Received Received\_MOE

[11] No\_Snap No\_Snap\_MOE id Percentage

<0 rows> (or 0-length row.names)

> stco<-merge(states, b, by="State")

> head(stco)

State long lat group order subregion Total\_Pop Total\_Moe

1 alabama -87.46201 30.38968 1 1 <NA> 1897576 10370

2 alabama -87.48493 30.37249 1 2 <NA> 1897576 10370

3 alabama -87.52503 30.37249 1 3 <NA> 1897576 10370

4 alabama -87.53076 30.33239 1 4 <NA> 1897576 10370

5 alabama -87.57087 30.32665 1 5 <NA> 1897576 10370

6 alabama -87.58806 30.32665 1 6 <NA> 1897576 10370

Received Received\_MOE No\_Snap No\_Snap\_MOE id Percentage

1 247490 8214 1650086 11159 0400000US01 13.04243

2 247490 8214 1650086 11159 0400000US01 13.04243

3 247490 8214 1650086 11159 0400000US01 13.04243

4 247490 8214 1650086 11159 0400000US01 13.04243

5 247490 8214 1650086 11159 0400000US01 13.04243

6 247490 8214 1650086 11159 0400000US01 13.04243

>

> b1<-state\_base + geom+polygon(data=stco, aes(fill=Received), color="white") +geom\_polygon(color="black", fill=NA)+coord\_fixed(1.3)+theme\_bw()+ggtitle("Snap Benefits Received as a Percentage of Total Population by State")

Error: object 'geom' not found

> b1<-state\_base + geom\_polygon(data=stco, aes(fill=Received), color="white") +geom\_polygon(color="black", fill=NA)+coord\_fixed(1.3)+theme\_bw()+ggtitle("Snap Benefits Received as a Percentage of Total Population by State")

Coordinate system already present. Adding new coordinate system, which will replace the existing one.

> b2<-b1+scale\_fill\_gradient(low="lightred", high="darkred")

Error: Unknown colour name: lightred

> b2<-b1+scale\_fill\_gradient(low="lightblue", high="darkblue")

> b2

Code used to create the plot “Food Deserts by County:

> library(mapdata)

Loading required package: maps

> library(ggplot2)

> usa<-map\_data("usa")

> states<-map\_data("state")

> counties<-map\_data("county")

> state\_base <-ggplot(data=states, mapping=aes(x=long, y=lat, group=group))+ coord\_fixed(1.3) + geom\_polygon(color="black", fill="gray")+ggtitle("Food Deserts by County")

> state\_base

> des<-state\_base + geom\_polygon(data=counties, fill=NA, color="gray") + geom\_polygon(color="black", fill=NA)

> des

> at<-read.csv("catlas3.csv")

> colnames(at)<-c("State", "County", "OneandTen")

> head(at)

State County OneandTen

1 Alabama Autauga County 1

2 Alabama Baldwin County 1

3 Alabama Barbour County 1

4 Alabama Bullock County 1

5 Alabama Butler County 1

6 Alabama Calhoun County 1

> at$OneandTen<-as.numeric(as.character(at$OneandTen))

> at$State<-tolower(at$State)

> names(counties)[names(counties)=="region"]<-"State"

> countyat<-merge(counties, at, by="State")

> head(countyat)

State long lat group order subregion County

1 alabama -86.50517 32.3492 1 1 autauga Chambers County

2 alabama -86.50517 32.3492 1 1 autauga Chilton County

3 alabama -86.50517 32.3492 1 1 autauga Autauga County

4 alabama -86.50517 32.3492 1 1 autauga Baldwin County

5 alabama -86.50517 32.3492 1 1 autauga Coffee County

6 alabama -86.50517 32.3492 1 1 autauga Colbert County

OneandTen

1 1

2 1

3 1

4 1

5 1

6 1

> fdc<-subset(countyat, countyat$OneandTen>=1)

> fddf<-data.frame(character(), integer(), stringsAsFactors = FALSE)

> foodd<-des+geom\_point(inherit.aes = FALSE, aes(x=long, y=lat), color="blue", size=0.5, data=fdc)

> food

Error: object 'food' not found

> foodd

**Plot and Data Selection**

As described previously, it was difficult to obtain data for this project that fit with other data. Due to time constraints and software limitations very little insight was able to be gained. I felt maps were important because it suited the information best. Since food deserts are location-based a map of other location-based data should be the most revealing. The idea is that since food assistance programs and food deserts have the common factor of low income, then the map should show some correlation between the two. The first map had to be used as a percentage of the population because the number of people in each area would not have told us much since the United States is geographically varied, and the data set in comparison accounted for both urban and rural areas.

**Conclusions & Recommendation**

The plots that were created are a successful beginning to understanding the conditions of food deserts. Overall, however, it is a largely unsuccessful project. Data sets proved more challenging than initially anticipated, and unanticipated software issues coupled with time constraints made it difficult to attempt to prove the hypothesis. What the data that is presented here does manage to show is that poverty likely isn’t the only large contributing factor for food deserts, despite being one of the key factors pointed to by reputable sources such as the Department of Agriculture. While it is difficult to offer any recommendations based on these very early-stage findings, there is absolutely reason here to look further into food deserts, their causes, and how they can be remedied. A quick Google search provided the following information: the average American eats 1,996 lbs. of food per year, and the U.S. as a whole wastes an estimated 108 billion pounds of food per year, which means that 54,108,216 people can be fed per year by the amount of food that is wasted. There are approximately 38.3 million food insecure people in the United States. This is why the recommendation is that more research in this area is absolutely crucial. Comparisons between more common factors of food deserts and food insecurity may provide more insight. It could also be beneficial if there were data on geographic locations of supermarkets (there is some, but it is not public), to understand the relationship between supermarket locations and food deserts, where waste is produced and if it is feasible to turn food waste into food security from a geographic standpoint.

**Data Sets**

Used for this project:

U.S. Census Data

<https://data.census.gov/cedsci/table?q=food%20insecurity&g=0100000US%240400000,%240500000>

Atlas of Food Deserts

<https://www.ers.usda.gov/data-products/food-access-research-atlas/go-to-the-atlas/>

Data Sets Considered, but not used:

Data on food deserts: <https://foodispower.org/access-health/food-deserts/>

World Food Programme: <https://docs.wfp.org/api/documents/WFP-0000117811/download/?_ga=2.202424085.5276826.1640885801-1571597065.1640885801>

BLS Data Labs/CPI Data: [https://beta.bls.gov/dataQuery/find?fq=survey:[cu]&s=popularity:D](https://beta.bls.gov/dataQuery/find?fq=survey:%5bcu%5d&s=popularity:D)

Kaggle: <https://www.kaggle.com/search?q=food+insecurity>

USDA on food security: <https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx>

**References**

Caporusico, J., (2020, June 22). *What are food deserts, and how do they impact health?* Medical News Today. <https://www.medicalnewstoday.com/articles/what-are-food-deserts>

1. https://www.rts.com/resources/guides/food-waste-america/ [↑](#footnote-ref-1)
2. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security.aspx [↑](#footnote-ref-2)
3. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx#children [↑](#footnote-ref-3)