**IST 652: Scripting for Data Analysis**

****

**What Factors Influence Food Insecurity?**

**IST 652**

**Project Report**

Food Insecurity Analysis

**By:**

Avery Curcio

Emani Jones

**Our Data and its Source**

The topic we want to investigate is what type of factors will predict whether or not certain individuals and countries are food insecure. The datasets we used are from online resources from kaggle.com, as well as the World Bank data catalog. We used multiple datasets such as food insecurity by country as well as a dataset containing economic indicators and demographic data as well.

**Data Preprocessing**

The first step we took towards preprocessing our data was reading in all of our data. From there we selected only the columns we wanted to use in the data. These columns include Area, Item, Value and Year for our first dataset. We did this because we did not want irrelevant data in our analysis. Next, we used the pivot function to pivot the data. After using pivot, the next preprocessing step we took was resetting the index so that the variables Area and Year are columns rather than indexes. This is important because it is easier to work with the data when it is in columns. We then converted selected columns to a numeric data type. The final step we took to preprocess our data was to convert selected columns into a smaller subset, which contained the majority of the food insecurity data. We named this subset“threeyearaverage”. We then divided the values out of 100 so we could get a decimal to measure food insecurity rates. At the end of preprocessing, we joined the dataset containing food insecurity data and economic factors by joining the data frames by country and year. In the third dataset we read in, we had to remove spaces in the column names, as well as remove commas in numbers, so that we can convert certain columns into numeric data types.

**Method of Analysis**

Our analysis will be able to determine questions like: Which country is most severely food insecure? What economic factors make a country insecure? Which demographic is the most food insecure? Which region in the United States is the most food insecure? We answered our analysis questions by exploring the average food insecurity rate in each country over a three year time period. To answer our first question about which country is the most food insecure, we sorted countries by, “Prevalence of moderate or severe food insecurity in the total population”. This displayed to us the top 10 countries and their food insecurity rates. We then created a bar graph using matplotlib to help us visualize the data more efficiently in terms of which country has the most food insecurity. See the figure below for the top 10 countries and their percetages as well as the bar graph below.

| Area | Prevalence of moderate or severe food insecurity in the total population (percent) (3-year average) |
| --- | --- |
| Congo | 0.874250 |
| South Sudan | 0.849333 |
| Malawi | 0.818250 |
| Central African Republic | 0.813000 |
| Sierra Leone | 0.809000 |
| Liberia | 0.806000 |
| Somalia | 0.791000 |
| Guinea | 0.739250 |
| Angola | 0.718000 |
| Mozambique | 0.697500 |

*Figure 1*

To answer our second question of what economic factors make a country food insecure, we used a correlation matrix to determine which variables were most correlated to the variable “Prevalence of moderate or severe food insecurity in the total population” also seen in Figure 3.

We used the json command to read in our data set which explored the economic indicators of food insecurity and created different variables to illustrate the different years named, “avg20152017”, “avg20162018”, and “avg20182020”. We then used the concat function to combine all the data from these years together. Next we had to rename the variable Area that we used in our first dataset so that we could merge our country dataset and economic factors dataset together. After we merged the datasets we ran a correlation so we could see which economic factors affect food insecurity in different countries. After running our correlation, we extracted the head of our correlation which showed the top 10 factors. These factors were:

* Prevalence of moderate or severe food insecurity in the total population (percent)
* Prevalence of severe food insecurity in the total population (percent)
* Prevalence of undernourishment (percent)
* Share of dietary energy supply derived from cereals, roots and tubers (kcal/cap/day)
* Inflation: percent change in the Consumer Price Index
* Number of moderately or severely food insecure people (million)
* Value of food imports in total merchandise exports (percent)
* Cereal import dependency ratio (percent)
* Number of people undernourished (million)
* Unemployment rate

We then imported ipywidgets so we could create a countplot of average dietary energy supply adequacy in respect to prevalence of malnourishment over a 3 year average.

Lastly, we imported our third data set which included demographic information concerning food insecurity. With this dataset we plan to answer the questions, which demographic is the most food insecure and which region in the United States is the most food insecure? In our code, we ran a groupby clause to group by race and food insecurity percentage. We then did the same thing for categories: household income to poverty ratio, household composition, area of residence and census geographic region. We then made a bar graph to display these results with all categories combined together.

**Output of the Python program**

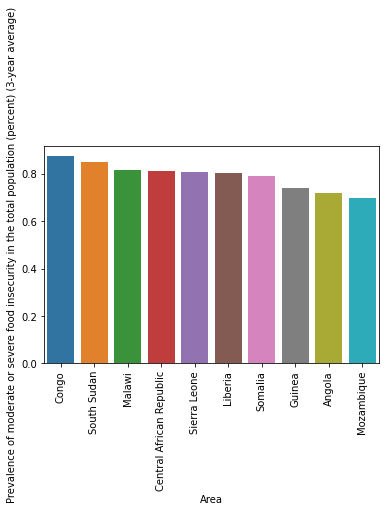
Our results of the first bar graph we created showed us that the country with the most food insecurity was the Congo with a 87% rate of food insecurity on average during a three year time. The 10th most food insecure country appeared to be Mozambique with a 69% rate of food insecurity. Our correlation matrix shows the most influential economic factors of food insecurity in various countries, with “Prevalence of severe food insecurity in the total population” being the number one reason for a country to be food insecure, seen also in Figure 3. Figure 4 shows average dietary energy supply adequacy in respect to prevalence of undernourishment over a 3 year average.

When looking at the demographic categories, we see that Black non Hispanics have the highest food insecurity percentage out of all the races. Families with single mothers have the highest food insecurity percentage out of all of the household composition subcategories. The South also has a higher food insecurity percentage than the Midwest, West and North East. Figure 5, shows that households with children have the highest food insecurity.

Figure 6 is very interesting because it shows food insecurity by demographic and year. We can see that in 2020, food insecurity went up for many categories with individuals with children older than 18 being the most likely individuals to be food insecure. For people with no children over 18 years old, the greatest spike for food insecurity was seen in 2016. The next category is elderly, which saw the greatest spike for food insecurity around 2019. For White non-Hispanics, there was a greatest spike in food insecurity during 2020, for Black non-Hispanics the highest for food insecurity was found in 2016. Hispanics saw the most food insecurity around 2017. While individuals under 1.30 saw a spike in 2020.

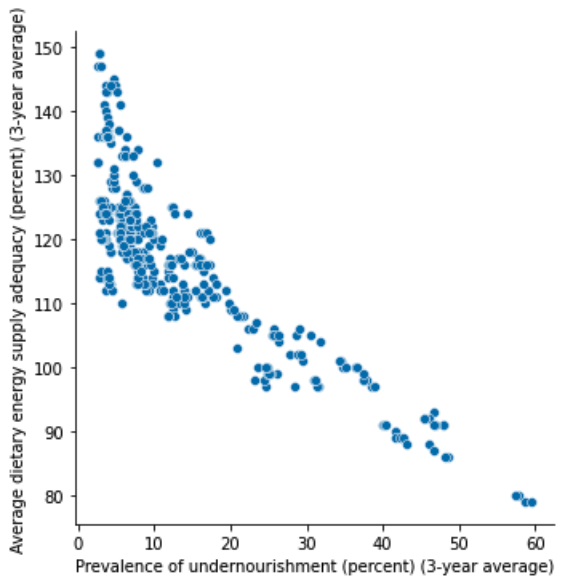
The income to poverty ratio was also measured in this graph. We get this ratio by dividing income by the poverty threshold. When it comes to individuals who fell below a 1.0 poverty ratio threshold, food insecurity was highest for these individuals during the year 2019. For people under a 1.85 scpre, they saw the greatest food insecurity during 2018. This is comparable to individuals who fell above 1.85, who saw high food insecurity during 2019.

People living inside a metropolitan area saw the greatest spike of food insecurity during 2020, whereas people living outside a metropolitan area saw the most food insecurity during 2017. This is interesting to compare with the individuals living in the Midwest who had the most food insecurity in 2019. The South had a surge of food insecurity during 2020, the Northeast saw a surge in 2018, and the West saw a spike between 2015-2017. It is interesting to note how different parts of the country have been affected by food insecurity at different times. The spikes in 2020 can most likely be related to COVID. While spikes in the other years could be due to recessions or other political factors.

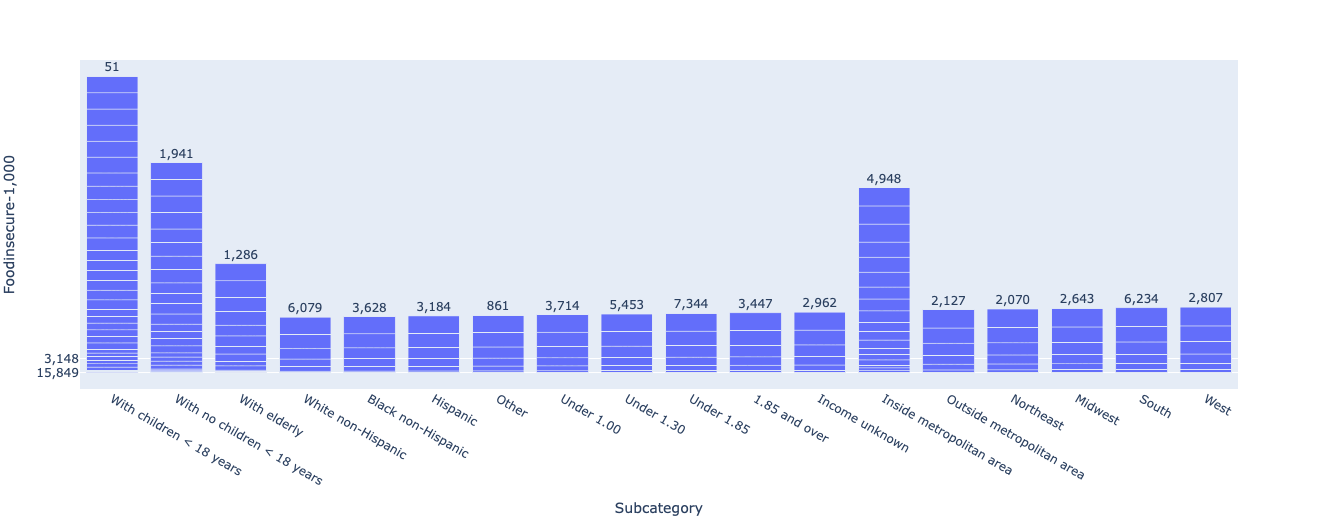


*Figure 2*

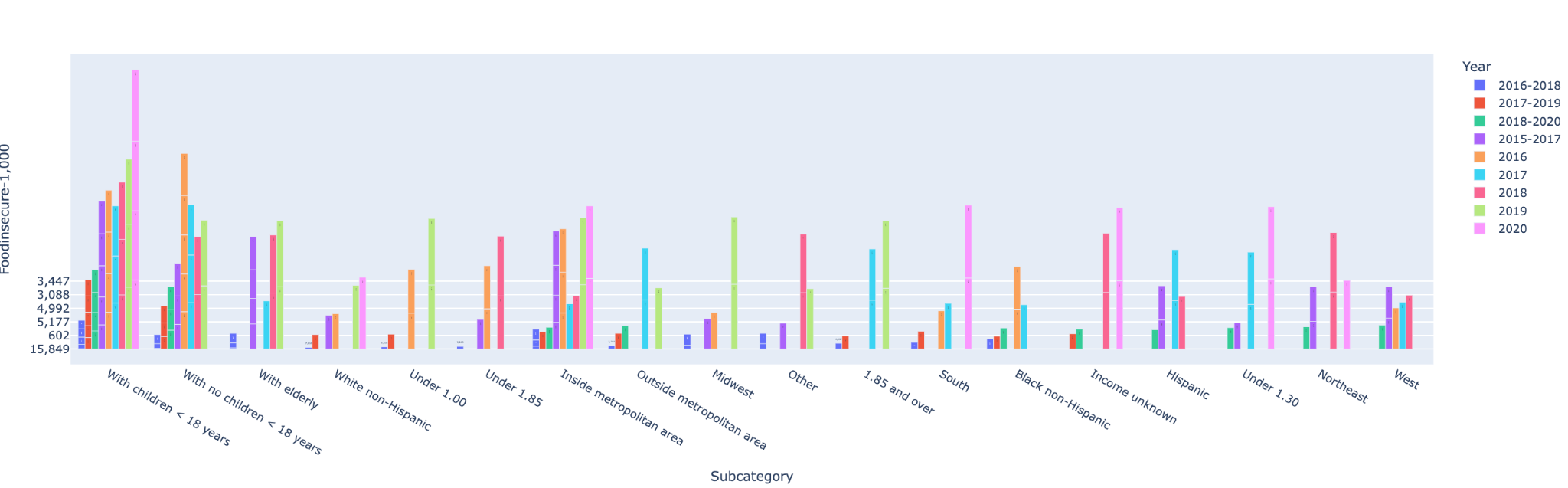
*Figure 3*



*Figure 4*

**

*Figure 5*

**

*Figure 6*

**Conclusion**

From our analysis, we discovered that the Congo is the most food insecure nation in the world. At the same time we saw that the economic factor that is correlated to food insecurity the most is “Prevalence of severe food insecurity in the total population”. As we compared average dietary energy supply adequacy and prevalence of undernourishment, we could see a correlation between these two things. We examined food insecurity by demographic, and our results showed us that 2020 showed the highest spike of food insecurity for individuals with children older than 18 years old. The demographic who saw the most food insecurity in 2017 was Hispanics. While the region with the most food insecurity is the South, which has slightly higher food insecurity rates than a metropolitan area. Individuals who fell below 1.30 on the poverty threshold were the most food insecure compared to others on the scale.

Overall, this food insecurity analysis is important because it can bring awareness to the issue of food insecurity to people who might not be aware. This analysis also targeted the specific economic factors that cause food insecurity, this type of analysis could hypothetically help policy makers make better economic solutions. Lastly, our analysis also illustrated which subcategories of demographic are the most food insecure.