Analyzing Food Security in United States

Kishore Vasan kishorev@uw.edu

Madeline Holmes mholmes2@uw.edu

Nirmalya Ghosh nirmalya@uw.edu

Simran Bhatia simran18@uw.edu

ABSTRACT

With increasing population and consumption growth, it is predicted that the global demand for food will increase for at least another 40 years. A multifaceted and linked national strategy is needed to ensure sustainable and equitable Food Security in the United States. This paper explores three different spheres that affect Food Security, that is, Food Affordability, Nutrition, and Food Accessibility. It analyzes each of the three spheres across different states and regions in the United States and presents results for the same. The preliminary results across all categories reveal Alaska as the most food-secure state in the country; while District of Columbia, West Virginia, Arkansas and Massachusetts were identified currently at the highest risk for food security. This study also aims to find clusters of geographic locations, based on accessibility, affordability, and nutrition. To do so, the study uses K-Means clustering as an unsupervised statistical model; the results show that with sufficient evidence for further exploration that there is a relationship between Food Security and geographic location of a state. Along with this report, the authors also developed a Shiny web application that can be used to interact and further explore the analysis.

KEYWORDS

Food Security; United States; Food Affordability; Food Accessibility; Nutrition; Population Health;

ACM Reference Format:

Kishore Vasan, Madeline Holmes, Nirmalya Ghosh, Simran Bhatia. 2018. "Analyzing Food Security in the United States". For the proceedings of Population Health Informatics (INFO 498-D) presentation.

1. INTRODUCTION

Food insecurity is estimated to affect 10% of the United States population.[1] In fact, one in every eight Americans suffers from food insecurity. Furthermore, previous studies have shown that food insecurity leads to a higher risk of adverse health outcomes as people who are food insecure tend to eat nutrient-poor foods to survive.[2] Therefore, limited access to food can pose a threat to the overall public health and is something that should be researched further.

One of the early definitions of Food Security came from the United Nations World Food Conference in 1974 with "availability at all times of adequate world supplies of basic food-stuffs...to sustain a steady expansion of food consumption... and to offset fluctuations in production and prices".[3] Since then, it has been impossible to speak about Food Security without including the aspects of the food supply, food access and food entitlement.[4]

Additionally, the American Dietetic Association [5] proposed a solution to the food problems- "To eliminate food insecurity, interventions are needed, including adequate funding for an increased utilization of food and nutrition assistance programs…[to] support individual and household economic self-sufficiency".

In this paper we consider Food Security as the following three categories, taking inspiration from the previously curated definitions of food access in *Juggling the Five Dimensions of Food Access: Perceptions of Rural Low Income Residents*,[6] namely:

1. Affordability - Food prices and people's perceptions of worth relative to food cost.
2. Nutrition - The nutritional index of the population based on factors of obesity/weight status and physical inactivity.
3. Accessibility - The geographic location of the food supply and ease of getting to that location.

Specifically, there has been significant research done analyzing nutrition and food security. For instance, the idea of a food desert-an area with limited access to food-was analyzed in John Coveney's research: *Effects of mobility and location on food access*. It is clear in this research that people all over the world are impacted by food insecurity and limited access to food, and much of this is caused by the limited transportation services available to their residents.[7]

Furthermore, a study conducted by Christina Mushi-Brunt,[8] *Fruit and Vegetable Intake and Obesity in Preadolescent Children: The Role of Neighborhood Poverty and Grocery Store Access*, determined that access to food can have detrimental impacts on a child's overall health. This research shows that children that have low access to foods typically have lower access to a healthy, nutritious meal which can lead to obesity.

The motivation to look at Food Security from the broader scope of geographic location comes from assessing the risk of Food Insecurity due to climate change.[9] In that case, the geographic location of a region is considered to be an

important factor in determining the food security of a region, which led to the initial insight for this exploratory study.

The purpose of this work is to understand the status of food security in the United States and thereby analyze the similarity between states. We use multiple identifiers from different sources to establish food security in a state. We employ visualizations as our primary tool to analyze food security through the lens of aforementioned subcategories and use K-Means clustering [10] to find a relationship between geographic location and food security.

2. METHODS

In the methods sections, we will first give a descriptive introduction to the datasets we used in our visualizations and analysis. Then we will present our visualizations on each of the subcategories of Food Security, Affordability, Nutrition, and Accessibility. After which we will include the methods used to determine clusters among states.

2.1 Data Overview

1. Personal Consumptions Expenditure [11]

Collected by: Bureau of Economic Analysis

The data used from this dataset focuses on the Total Personal Consumptions Expenditure as well as the Food and Beverages Expenditure subset, for each of the 51 states categorized based on the eight main regions from 2014 - 2016 were used. These 8 distinct regions are *Far West* (Alaska, California, Hawaii, Nevada, Oregon, Washington), *Great Lakes* (Indiana, Michigan, Ohio, Wisconsin), *Mideast* (Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania), *New England* (Connecticut, Maine, Massachusetts New Hampshire, Rhode Island, Vermont), *Plains* (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota), *Rocky Mountains* (Colorado, Idaho, Montana, Utah, Wyoming), *South East* (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia), *South West* (Arizona, New Mexico, Oklahoma, Texas).

2. Nutrition, Physical Activity, and Obesity - the Behavioral Risk Factor Surveillance System [12]

Collected by: Centers for Disease Control & Prevention

The data from this dataset focuses on the Obesity/Weight Status and Physical Inactivity percentages of the population for each of the 51 states, from 2011 - 2016. The data is part of a survey, and the questions used for the dataset were - “Percentage of Adults who were obese” and “Percentage of adults who have no-leisure time physical activity”. For further analysis, the variable of “StratificationCategory1” was used for analysis about Gender, Income, Education and Age (in years).

3. Food Environment Atlas [13]

Collected by: US Department of Agriculture - Economic Research Service

The dataset describes the accessibility starting at a county level for each state depending on a variety of variables. The variables used for the analysis were: States, Farmers Markets(2009, 2016), Grocery Stores(2009, 2014), Supercenters(2009, 2014), and Convenience Stores(2009, 2014) per 1000 population. Accessibility to grocery Stores for different population types - children, senior citizens, and population in 2010 and 2015 were also used. In conclusion, Food Insecurity was analyzed using the variable Household Food Insecurity percent change from 2010-2012.

2.2 Data Visualizations

For each of the discussed variables analyzing food security, we present multiple data visualizations created to show how each variable impacts food security. Through the data visualizations, it is clear that specific states have more difficulty ensuring their residents have proper food security. We employ scatter plots, bar graphs, heat maps, and predictive clustering, to further our argument that food security is impacted by geographic locations.

2.2.1 Food Affordability

U.S. Bureau of Economic Analysis (BEA) [11] groups from all 50 states and the District of Columbia into 8 distinct regions for purposes of presentation and analysis.

The two variables used to plot *Fig. 1*, Per Capita Total Expenditure and Food and spending per capita, but in USD. The size represents the percentage change in food and beverage spending from the previous year. Greater the size, more the change in spending.

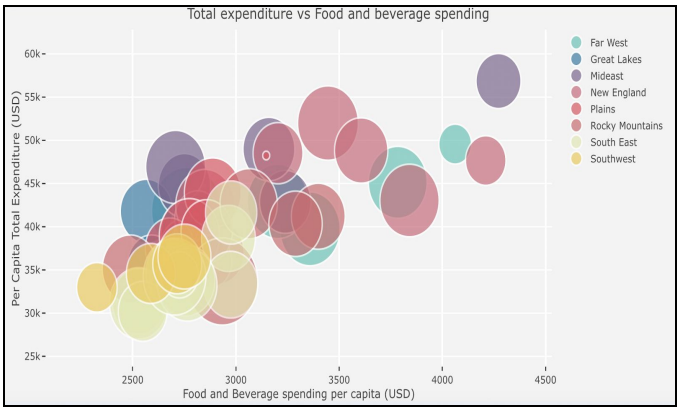

Fig. 1

2.2.2 Nutrition

The graph *Fig. 2* shows the trend of how nutritional index (in terms of Physical Inactivity and Obesity/Weight Status) has

changed over the past 5-years.

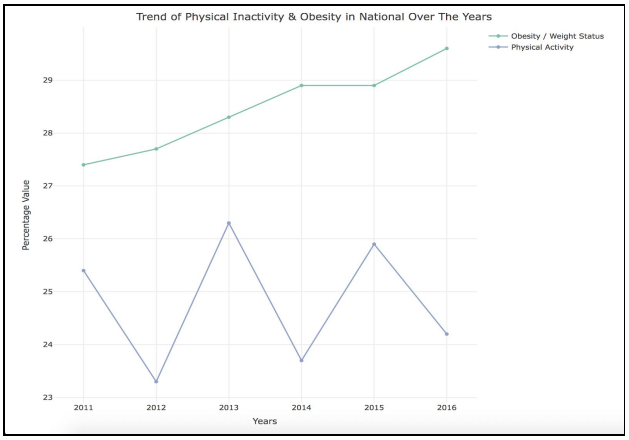

Fig. 2

Fig. 3 shows a state-wise comparison for the percentage of the population that is physically inactive and obese.

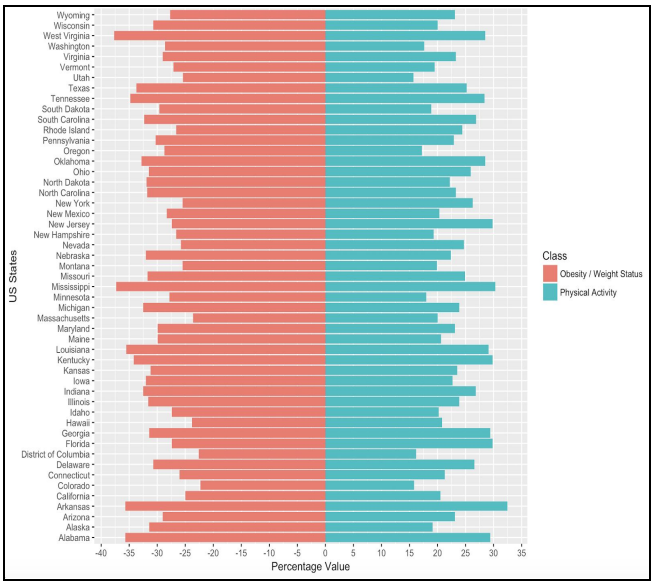

Fig. 3

2.2.3 Food Accessibility

Fig. 4 displays the average number of people within each state's population that have low access to grocery stores. It is for the overall population in 2015 to show the most current data. Data is also available for 2010, and for different age groups – children and senior citizens.

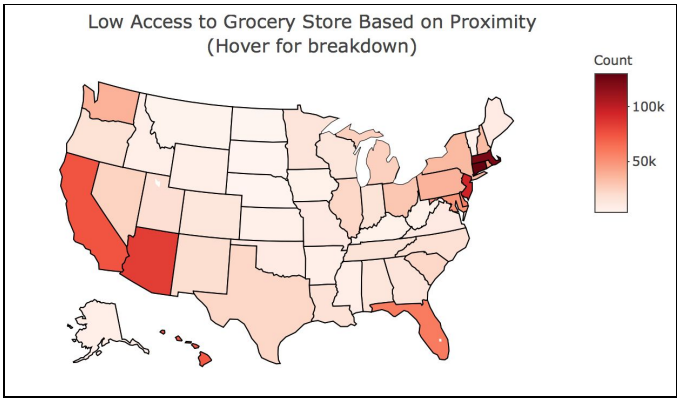

Fig. 4

An analysis was also done to display data for Connecticut, Massachusetts, and New Jersey-the states that had the highest number of people with low access to grocery stores in 2015. It displayed the past and recent data for each type of food store and the number of stores per 1000 people, and all the findings from the three graphs were very similar as each state had similar values for overall accessibility.

2.3 Predictive Clustering

It employ K-Means clustering as our tool to analyze the relationship between states based on their Food Security status. K-Means clustering is a type of unsupervised learning, which is used when dealing with unlabeled data (i.e., data without defined categories or groups). The algorithm works iteratively to assign each data point to one of the *k* groups based on the features that are provided. In this section, initial clustering among states is displayed, select features for the algorithm, find the optimum number of clusters, and visualize the clusters of states interactively.

For this analysis, a combination of the variables are presented in the above sections: percent population with low access, number of grocery stores per 1000 population, number of convenience stores per 1000 population, percentage of Adults who are obese, percentage of adults who have no-leisure time physical activity, number of farmers market per 1000 population, per capita food and beverages spending, number of supercenters per 1000 population.

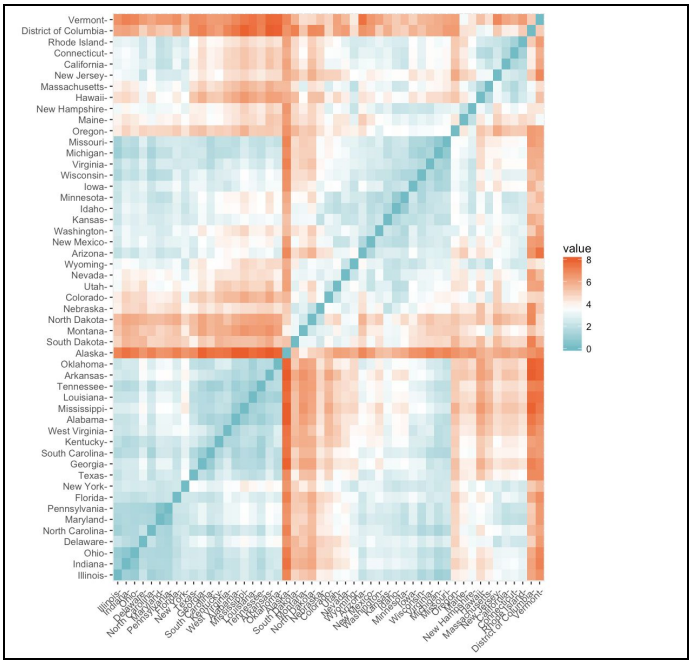

Fig. 5

In *Fig. 5,* the heatmap based on the euclidean distance between states based on aforementioned variables.

In a K-Means Analysis, it is important to select only the variables that uniquely identify a cluster and refrain from using variables that do not play a significant role in the clustering [15]. In *Fig. 6*, a dendrogram of the variables based on ascendant hierarchical clustering of a set of variables is shown. It is observed that obesity and physical inactivity measures are highly correlated with each other and separate from the rest of the variables. Hence, obesity is removed moving forward from our clustering analysis for more accurate results.

Fig. 6

Silhouette method is used to evaluate the performance of the K-Means algorithm. Silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. The measure of the average silhouette value for all the clusters is used to evaluate the algorithm. In *Fig. 7*, you can see the silhouette value for different values of *k.*

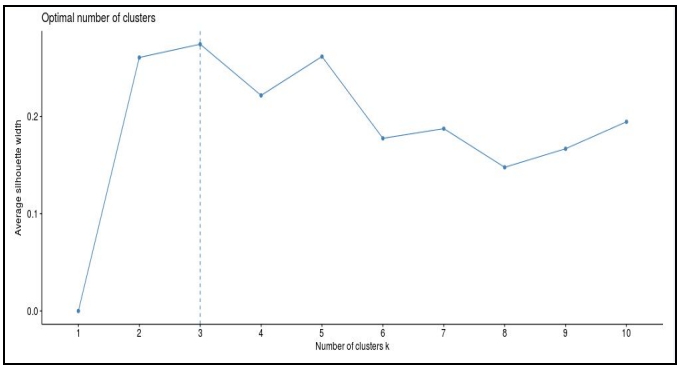

Fig. 7

For this project, *k=3* is set to be the optimum number of clusters for the algorithm. *Fig. 8* shows how the clusters are present in a 2D space and the states within each of the clusters.

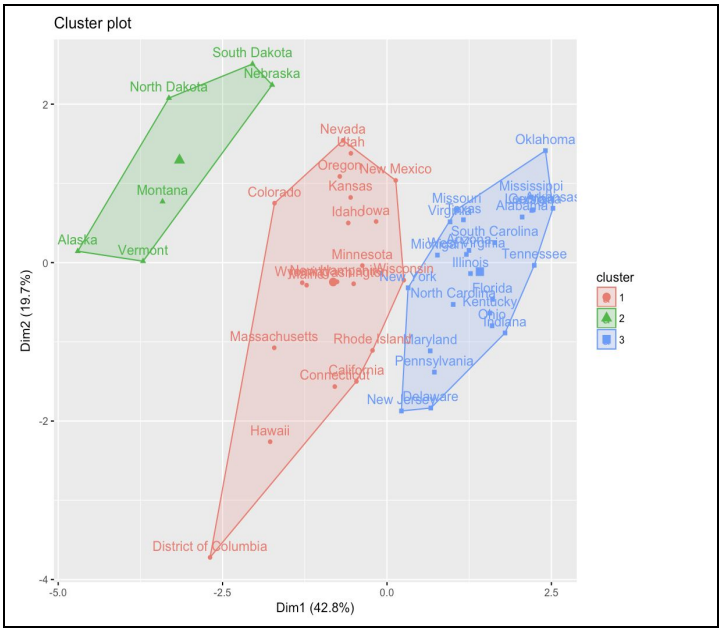

Fig. 8

3. RESULTS AND DISCUSSION

3.1 Food Affordability

1. *Far West*: Alaska is the highest in both variables. While most of the states in this region are close to the average per capita total expenditure and food and beverage spending per capita, Nevada is the lowest spender on food and beverage.

2. *Great Lakes*: Illinois has the highest per capita expenditure but the lowest expenditure on Food and Beverage, while within this region, Indiana has the lowest in both the variables. Highest Food and beverage spender: Wisconsin

3. *Mideast*: Out of all the states, District of Columbia stands out, as if it were an outlier, with the highest per capita total expenditure and food and beverage spending per capita. Most of the states are between the range of ~$40-49k for Per capita total expenditure and ~$2.7-3.2k for F&B spending.

4. *New England**:*** It is interesting to see that most of the states don't have similar values. Massachusetts have the highest per capita total expenditure and an average F&B spending for the region. While Vermont is the highest spender for F&B and Rhode Island being the lowest in both variables.

5. *Plains*: North Dakota stands out instantly as the change reported was negative but also remains the highest F&B spender and highest Per capita total expenditure in the region. Kansas is the lowest per capita total expenditure.

6. *Rocky Mountains*: Idaho has the highest percent change at 8.2% and yet remains the lowest per capita total expenditure state in the region and a bit below the average F&B spending.

7. *South East*: It is the region with the most number of states(12). Most of the states are concentrated between the $2600-$2800 range for F&B spending and Per capita total expenditure range of $32k-36k. It is interesting to find that, increase in per capita total expenditure doesn't mean increase in F&B spending for this subset of state. While the highest per capita spender is Virginia, it only spends ~$3000 on F&B, not significantly higher than most of the states concentrated in the average zone. Mississippi is the lowest in both variables for this region.

8. *South West*: We can see a linear plot of dots, there is an almost linear increase in total expenditure. There seems to be a more direct relationship between the per capita total expenditure and the F&B spending in these states. Texas is the highest in both variables while Oklahoma is the lowest.

3.2 Nutrition

The average value of obesity is 29.8% and the average value of physical inactivity is 23.4% nationally in 2016. The top 10 high-risk zone states (with Obesity > 34% or Physical Inactivity > 29%) are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, New Jersey, Tennessee, West Virginia (in no particular order).

These values show that 1 in 5 adults in the US is obese. This matches with the findings from the OECD Update about Obesity in 2017, which states that "Adult obesity rates are highest in the United States, Mexico, New Zealand and Hungary."

In addition, even though the US is not a high-risk country for physical inactivity as compared to other countries, it still lies in the second risk zone according to WHO's report on Prevalence of Insufficient Physical Activity.

The highest rate of obesity is found in West Virginia, wherein 37.7% of the population is obese in 2016. The analysis found that males are marginally more obese than women. Additionally, low-income households have higher obesity (with 16.2%) as compared to high-income households (with 12.5%). The age group at the highest risk for obesity in West Virginia is 45-54 years (with 20.7%) and the lowest risk for obesity is 18-24 years (with 11.6%).

For physical inactivity, the highest rate is in Arkansas, with a reported percentage of 32.5 percent in 2016. Even though Arkansas has been more physically inactive in the past years than in 2016, its own highest obesity rates are in 2016 as well. The analysis reported that men are physically more inactive in Arkansas than women, that lower the household income, higher is the rate of physical inactivity.

Overall, it is seen that the United States is at a relatively high-risk zone for both Obesity and Physical Inactivity, which makes it not only nutritionally insecure in current times but more so in the future.

3.3 Food Accessibility

In a recent study analyzing the relationship between access and the amount of healthy food available. The authors found that the children that had more access to grocery stores in low poverty areas tended to have more servings of fruits and vegetables.[9] Therefore, having a grocery store accessible to the population, no matter what their socioeconomic standing is, is important because it can impact the number of health food that person is consuming. Access to grocery stores is the first step to providing the population with healthy food and is why this variable was chosen to analyze how food secure residents are in different states.

Overall, the states with the most consistent high rates of low access to grocery stores based on the overall population are California, Massachusetts, New Jersey, Connecticut, and Arizona. Specifically, Connecticut and Massachusetts are consistently the top two states that have the highest population with low access to grocery stores in both 2010 and 2015. In fact, the Boston Globe recently discussed how many residents in Massachusetts struggle to have access to grocery stores.

Felice Freyer describes how the nearest grocery store for a resident is about two hours away, but also requires a two-hour bus ride with multiple transfers.[14] One of the biggest issues Massachusetts faces is the transportation to the grocery store.[14] If a resident does not have a car, they must rely on the subpar public transportation which causes problems for how accessible the grocery stores are for their residents. Massachusetts is consistently one of the states with the largest population that has low access to grocery stores, and this can lead to unhealthy diets.

In Massachusetts in 2015, the number of children that had low access to grocery stores was 28,452 and the number of Senior Citizens with low access to grocery stores was 17,844. Although there are much fewer Senior Citizens struggling with grocery store access than children, this number of Senior Citizens in Massachusetts was one of the highest values when compared to other states. It is interesting to see how the population of each state differs when looking at these age groups. Some encouraging data proves that there has been a decline in the number of Senior Citizens in Delaware, New Jersey, and Massachusetts that have low access to grocery stores from 2010 to 2015. Clearly, certain states are attempting to increase the accessibility to grocery stores, and certain states have been fairly successful.

In terms of Convenience Stores, the number of stores per 1000 people slightly decreased in Connecticut, remained the same in Massachusetts, and increased in New Jersey from 2009 to 2014. Each of these states had a different change in the number of Convenience Stores, but all continued to struggle with low access in recent years.

In Connecticut, Massachusetts, and New Jersey, the number of farmers markets increased from 2009 to 2016. This could provide more of the population with healthy, organic foods that are sold by farmers markets.

For each of the three states, the number of grocery stores per 1000 people is approximately the same value—0.25. This explains why these three states are the most concerning with access to grocery stores, and explain why their unweighted values are so similar. As discussed earlier, a limited supply of grocery stores can have a negative impact on the population's ability to adequately feed everyone.

Finally, the number of supercenters per 1000 people for all three states was almost insignificant and under 0.05. It was interesting to see such a small value for a type of food supply that can provide the public with an extremely large supply of food and other retail items.

Each of the three states had similar values for their number of food supply stores per 1000 people and is consistent with the understanding that these are the top three states that are similar in their service to providing the public with access to food. These states continue to have extremely large numbers of people that have low access to grocery stores, but have larger access to convenience stores. The concerning aspect of this access is how convenience stores typically do not sell as many fresh fruits and vegetables. Therefore, the public's access to food is limited by what is supplied by convenience stores for all three of the states analyzed.

3.4 Relationship between States

The goal of the paper was to explore the relationship between geographic location and Food Security. The authors used a couple of different methods to analyze the same: Euclidean distance between states based on their Food Security values and K-Means clustering.

In *Fig. 5,* a heat map shows clustering of states. For observations, Alaska acts as this central state that has very similar Food Security to most of the states and thus represents the average Food Security situation in the United States. Vermont and District of Columbia have very similar Food Security results, and they are also geographically close to each other.

The K-Means clustering algorithm was used in forming clusters between states in an unsupervised manner. The results of which were presented in *Fig. 8.* The green cluster, Nebraska, South Dakota, North Dakota and Montana have similar Food Security situation and are also neighboring states, while red cluster, observes that Nevada, Utah, Oregon, New Mexico, Kansas and Colorado have similar Food Security situation and are also neighboring states and the blue cluster observes that Illinois, Indiana, Ohio, Kentucky, Tennessee, South Carolina have similar Food Security situation and are also neighboring states.

While there are a lot of caveats to this interpretation, however, the presence of multiple neighboring states in each of the clusters show that there exists some relationship between geographic location and food security.

4. LIMITATIONS

There were multiple limitations in conducting this study; most of them were related to the datasets used in the study. The main issues were related to data timeline, data formatting, data collection, and causation.

The three datasets in the study had multiple time discrepancies, in terms of years in which the data was collected. This made it difficult to connect the analysis and visualisations together. Hence the study's major focus was upon the trend for all factors for food security over geographic location, and time was standardized as *Past* or *Recent* times

for all factors. The timeline was only considered for individual factor analysis.

The datasets were hard to directly load into R[16] and visualise, due to errors in the data formatting. The datasets had to be selected according to individual columns and creation of separate files for the data was required for each factor.

The data collection methods were not standardised, as each dataset was collected by a different agency using different methods. Different agencies use their own data collection methods, which makes it difficult to connect data from multiple sources.

Lastly, the study's last section focuses on finding the relationship between states, based on the correlation between factors. However, it is important to note that this correlation does not lead to causation and should be considered for all results and future analysis from this study.

4. FUTURE DIRECTIONS

After the analysis about the Food Security problem in the United States, it was identified that this analysis could be used for multiple policy-makers in their work as well as potential future research areas were identified.

For policy-making, it is important to note the risk zone for each factor. For affordability, District of Columbia; for obesity, West Virginia; for physical inactivity, Arkansas; for accessibility, Massachusetts, Connecticut and New Jersey. Each of these state legislatures should focus on the respective factors for consequent policies in the state.

For future research, there are multiple potential areas that could help in this study. Affordability, nutrition, and accessibility could be researched in depth, for example, more factors that affect nutrition such as Fruits/Vegetable intake could be analysed. Another factor that could be analysed would be the effect of season on each location. For finding causes of the correlation analysed in this study, there is scope to research the social determinant for each geographic location, especially the risk-zones identified in the study. Lastly, another important question to answer is how different geographic locations in the United States compare to other locations in other countries, based on similar factors.

5. CONCLUSION

In this project, Food Security is analysed in the United States through three categories - Food Affordability, Nutrition, and Food Accessibility. The preliminary analysis we presented in this report shows that Food Security is a pressing and important issue that needs to be addressed as a Population Health Issue. Additionally, our clustering analysis reveals that there may be an effect of geographic location on Food Security. We provided promising results for further exploring in this field.

REFERENCES

1. Pheley, Alfred M., et al. "Food Security and Perceptions of Health Status: A Preliminary Study in Rural Appalachia." *Freshwater Biology*, The Journal of Rural Health, 28 June 2008, onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-0361.2002.tb00909.x.
2. "Food Insecurity: A Public Health Issue." *Public Health Reports* 131.5 (2016): 655–657. *PMC*. Web. 2 June 2018.
3. World Food Conference Rome, Italy). Report of the World Food Conference, Rome, 5-16 November, 1974. New York: United Nations, 1975.
4. Parry, Martin, et al. "Climate Change and World Food Security: a New Assessment." *ScienceDirect*, Elsevier, 25 Oct. 1999, www.sciencedirect.com/science/article/pii/S0959378099000187.
5. Holben, David. "Position of the American Dietetic Association: Food Insecurity in the United States." *ScienceDirect*, Elsevier, 25 Aug. 2010, www.sciencedirect.com/science/article/pii/S0002822310011946?via=ihub.
6. Andress, Lauri, and Cindy Fitch. "Juggling the Five Dimensions of Food Access: Perceptions of Rural Low Income Residents." *ScienceDirect*, Elsevier, 18 May 2016, www.sciencedirect.com/science/article/pii/S0195666316301908.
7. Coveney, J, and L A O'Dwyer. "Effects of Mobility and Location on Food Access." *Advances in Pediatrics.*, U.S. National Library of Medicine, Mar. 2009, www.ncbi.nlm.nih.gov/pubmed/18396090.
8. Mushi-Brunt, Christina, et al. "Fruit and Vegetable Intake and Obesity in Preadolescent Children: The Role of Neighborhood Poverty and Grocery Store Access." *American Journal of Health Education* 38.5 (2007): 258-65. *ProQuest.* Web. 29 May 2018.
9. Ahmad, M., G. Mustafa, and M. Iqbal. "Impact of Farm Households' Adaptations to Climate Change on Food Security: Evidence from Different Agro-Ecologies of Pakistan." *Pakistan Development Review*, vol. 55, no. 4, 2016, pp. 561-588. *SCOPUS.*
10. Lloyd, Stuart P. "Least squares quantization in PCM." Information Theory, IEEE Transactions on 28.2 (1982): 129-137.
11. US Department of Commerce. "Bureau of Economic Analysis." U.S. Bureau of Economic Analysis (BEA), US Department of Commerce, 4 Oct. 2017, www.bea.gov/newsreleases/regional/pce/pce_newsrelease.htm.

12. CDC. “Nutrition, Physical Activity, and Obesity - Behavioral Risk Factor Surveillance System | Chronic Disease and Health Promotion Data & Indicators.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 22 Jan. 2018, chronicdata.cdc.gov/Nutrition-Physical-Activity-and-Obesity/Nutrition-Physical-Activity-and-Obesity-Behavioral/hn4x-zwk7.
13. USDA. “Data Access and Documentation Downloads.” *USDA ERS - Sharing the Economic Burden: Who Pays for WIC's Infant Formula*, United States Department of Agriculture, 27 Mar. 2018, www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/.
14. Freyer, Felice J. “Want Healthy Food? In Much of Mass., It's Hard to Get.” *BostonGlobe.com*, The Boston Globe, 3 May 2017, www.bostonglobe.com/metro/2017/05/02/want-healthy-food-much-mass-hard-get/6tdKRBgRmyjprBPRDvl31H/story.html.
15. Raftery, Adrian E, and Nema Dean. “Variable Selection for Model-Based Clustering.” *Taylor & Francis*, Journal of the American Statistical Association, 1 Jan. 2012, www.tandfonline.com/doi/abs/10.1198/016214506000000113.
16. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.