Proceedings of the IISE Annual Conference & Expo 2023
K. Babski-Reeves, B. Eksioglu, D. Hampton, eds.

Clustering Partner Agencies for a Local Food Bank

Henry Ivuawuogu, Steven Jiang and Lauren Davis
Department of Industrial & Systems Engineering
North Carolina Agricultural and Technical State University
Greensboro, NC 27411, USA

Abstract

Food insecurity is a serious problem in America and the pandemic makes the problem even worse. Feeding America has more than 200 food banks. that These food banks and their partner agencies are the key players in the battle against food insecurity. Partner agencies may vary in size and location depending on the service area and the variety of the partner agencies and the complexities of their operations make equitable food distribution very challenging. There is a need for a meaningful to group those partner agencies to assist food bank operations managers to make informed decisions. This study uses data from a local food bank and its partner agencies. Each agency is unique in terms of its behavior. Therefore, *k*-means clustering was used to categorize agencies into groups based on the number of persons served and the amount of food received. The results of the study will provide evidence-based information to assist the food bank in making informed decisions.

Keywords
Cluster, Partner agencies, Food bank, Hunger relief

1. Introduction

Food insecurity, an economic and social condition where households have limited access to nutritious food [1], is a serious problem in America. Feeding America is the nation's largest nonprofit domestic hunger relief organization. There are more than 200 food banks affiliated with Feeding America and these food banks serve all fifty states, the District of Columbia, and Puerto Rico. Food banks and their partner agencies are the key players in the battle against food insecurity. Food banks acquire donated food, safely store food, and distribute food to partner agencies such as food pantries and soup kitchens. These partner agencies will then provide safe and nutritious food to people who experience food insecurity. Partner agencies must meet certain requirements (i.e., non-profit) to be eligible to receive food from the food bank.

The distribution network of food banks has a lot of complexity with several configurations. These would consist of various charitable partner agencies that help collect and distribute donated food, decisions about food distribution need to ensure food equitably, efficient utilization, and cost-effectively distribution. Also, high uncertainty in supply from donors and demand from clients could create additional complexity in understanding existing supply and food needs. Hasnain et al. [2] highlighted the challenges faced by the food bank network, such as transportation disruptions and communication breakdowns, especially during disasters. The study found that the food bank network was able to maintain its critical services during the hurricane by implementing various strategies, such as pre-positioning supplies, collaborating with partners, and adjusting distribution schedules. Partner agencies typically visit a food bank branch to receive the food they offer to their clients, and then individual clients or families visit the partner agencies for food assistance. This involves a huge volume of unstructured and complex data coming from food donations, distribution events, and inventory management.

The Food Bank of Central and Eastern North Carolina (FBCENC) works with over 800 partner agencies including food pantries, shelters, soup kitchens, and group homes, that help serve 34 counties in central & eastern North Carolina. The food bank serves its partner agencies through six branches in Durham, Greenville, New Bern, Raleigh, Sandhills (Southern Pines), and Wilmington with Raleigh as the headquarter.

According to FBCENC's fact card, there are nearly 510,000 people facing hunger across its 34-county service area. In the fiscal year 2021-2022, the food bank distributed about 138 million pounds of food by partnering with 813 agencies in its network [3]. In meeting these needs, the food bank typically uses the amount of food disturbed and the number of people served by partner agencies in the previous year in determining how much the partner agencies would receive the next year. The food bank also has a fair share percentage it tries to meet among its different branches to ensure equal and equitable distribution. The existing clustering pattern of the food bank lacks the capabilities to interpret how agencies operate in the distribution of food and persons served correspondingly.

The complexity of the system makes it very challenging for the food bank to make an equitable distribution to the partner agencies. Currently, FBCENC does not have a very effective way to group its partner agencies and there is a strong need for that.
In this preliminary study, clustering analysis was used based on the weight of food distributed and the number of persons served to help decision-makers make strategic decisions about the amount of food each branch or county within the food bank network receives and tailor their support to each agency's specific needs.

2. Method

2.1 Data collection and data analysis

Data on the amount of food distribution and persons served by the active partner agencies was used from the most recent active agencies for the fiscal year of 2019-2020 and 2020-2021. Agencies that stopped activities before 2019, and those that started operations after 2020 were not considered in this preliminary study. The data report included the gross weight (in pounds) of food distributed by the partner agency per transaction aggregated by year and the number of persons served also aggregated by year. Other characteristics of each agency such as the exact address including county and zip code, the branch serving that agency, and the posting date were obtained from data provided by FBCENC. After aggregating the data, 643 partner agencies were active for both fiscal years, 66 of them had no address and were excluded. Overall, 577 partner agencies were used for the analysis.

2.2 K-means clustering analysis

Clustering is a commonly used unsupervised machine-learning technique that does not require a response variable. However, it helps group instances into similar clusters by identifying relations between the instances in a dataset. *k*-means clustering is a popular method that splits the data points into a pre-specified set *k* groups. The data points are formed into one of the *k*-clusters using the centroid of the cluster and each cluster is acknowledged by its centroid. The centroid is a location representing the center of the cluster. The aim is to minimize the sum of distances between the instances and the cluster centroid, to identify the correct group each instance should belong to [4]. The most common approaches for generating the distance matrix are Euclidean and Manhattan distances. Euclidean distance is generally considered to determine the distance between each data object and the cluster centers [5]. The Euclidean distance between one vector x = $(x_1, x_2, \ldots x_n)$ and another vector y= $(y_1, y_2, \ldots y_n)$, The Euclidean distance $d_{euc}(x_i, y_i)$ can be obtained as follow in Equation 1:

Euclidean distance:

$$d_{euc}(x_i, y_i) = [\sum_{i=1}^{n}(x_i - y_i)^2]^{\frac{1}{2}} \tag{1}$$

The Manhattan distance is the sum of absolute differences. The Manhattan distance $d_{man}(x_i, y_i)$ can be obtained as follow in Equation 2:

Manhattan distance:

$$d_{man}(x_i, y_i) = \sum_{i=1}^{n}|(x_i + y_i)| \tag{2}$$

2.3 Defining the Optimal Number of Clusters

Clustering results may mostly be subject to the number of clusters stated. It is essential to provide educated guidance for defining the number of clusters to attain suitable clustering results [6]. A few methods are known in the literature for selecting the optimal number of clusters: The elbow method, the Silhouette method, and the Gap statistic. Kaufman and Rousseeuw [7] introduced the Silhouette index which is constructed to show graphically how well each object is classified in each clustering output. The silhouette value measures how similar a point is with its cluster relative to other clusters. In 2001, Tibshirani et al. [8] proposed an approach to estimate the number of clusters in a dataset via the gap statistic. The Elbow method is probably the most well-known method for defining the optimal number of clusters. It calculates the Sum of Squared Errors (SSE) within clusters for different values of *k* and selects the *k* for which SSE initially starts to reduce. In the plot of Total within SSE-versus-*k*, it becomes visible as an elbow. For this study, the elbow method was used.
The use of *k*-clustering in different applications across literature has been well recorded. Kuhn and Culhane [9] applied cluster analysis to test a typology of homelessness by pattern of shelter utilization, and the cluster analysis was used to generate three groups using the number of shelter days and the number of shelter episodes. Black & Seto [10] used data from a food bank group in Vancouver, Canada to carry out a cluster analysis using the *k*-means algorithm to

classify patterns of food bank use into short-term or transitional users, medium or episodic users, and longer-term or chronic users. The clusters were built on the number of exclusive food bank sites visited, the time in between the first and last recorded visit, usage features including the total number of visits to the food bank, and the number of 90-day pauses between the visits.

The *k*-means clustering used on the partner agencies in this study was 3. The decision was based on how the local food banks desired to see clusters for easy operations. Agencies were grouped into - small, medium, and large – based on the gross weight (in pounds) of food they received, and then clustered again based on the number of persons served.

2.4 Development of Visualizations

Tableau was used to explore the data collected, the boxplots of the gross weight (in pounds) and the number of persons served for the respective years are shown in Figure 1 below.

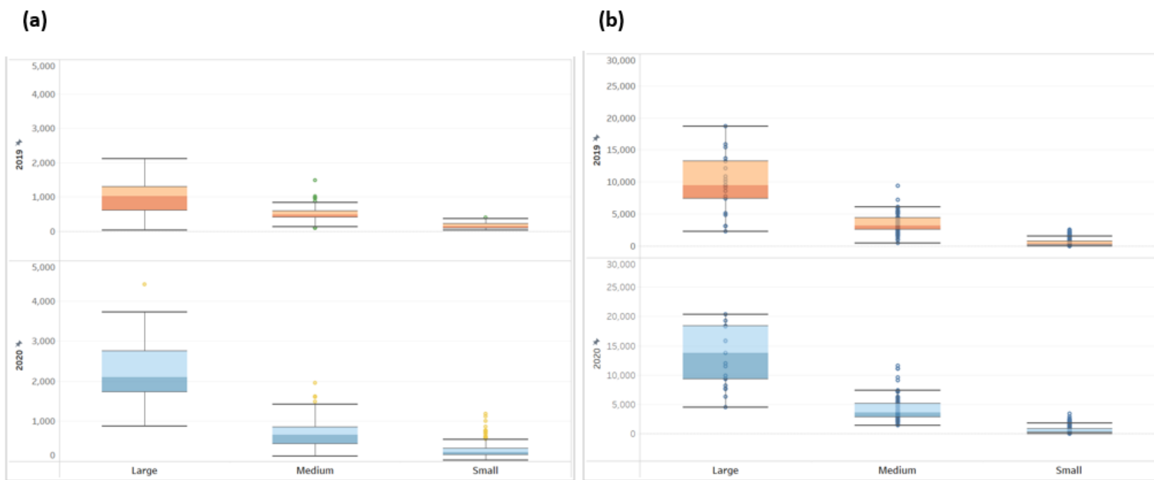

Figure 1: Box plot of data analysis in both years (**a**) the box plot of agency clusters for food distribution by gross weight, and **(b)** box plot of agency clusters for number of persons served.

3 Results

3.1 Cluster by gross weight

Using *k*-means clustering, the 577 active agencies in 2019 and 2020 were clustered into groups based on the gross weight of food demand by these agencies. This was done using in-built functions in the R programming language for the *k*-means clustering technique. *k*=2 was observed as the optimal number of clusters from the elbow chart. However, due to the large imbalance of the number of agencies between clusters when *k*=2, the preference from the food bank sizing *k*=3 was used. The Elbow method and cluster plot is shown in the figure below in Figure 2.

To better understand the variation between groups, the minimum, mean and maximum values of the observations within each cluster were calculated as shown in table 1 below.

Table 1: Descriptive Analysis of Clusters by Gross Weight

	Min		Mean		Max	
	2019	2020	2019	2020	2019	2020
Small	19	34	150	258	394	1184
Medium	90	125	504	698	1480	1954
Large	120	872	1225	2582	15000	10000

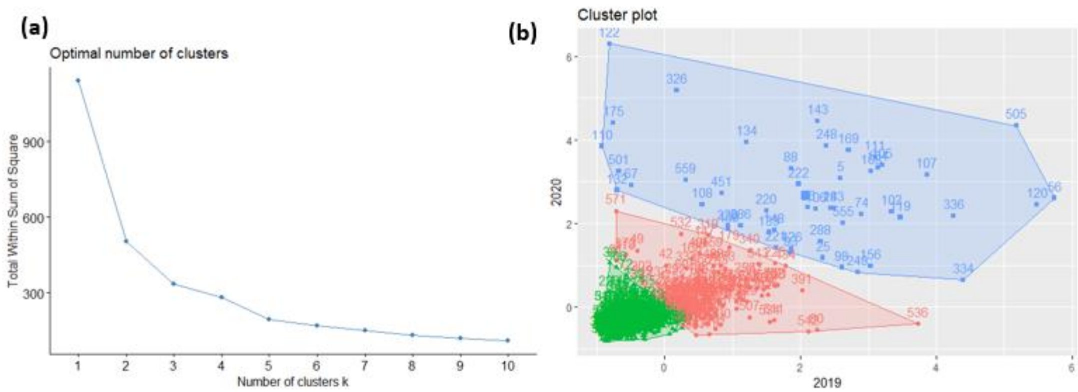

Figure 2: Cluster analysis for food distributed by gross weight **(a)** the elbow method demonstrating optimal number of clusters, and **(b)** cluster plot of agencies for food distributed by gross weight.

3.2 Cluster by number of people served

Similar to the cluster by gross weight, $k = 2$ was observed as the optimal number of clusters for the agencies based on the number of people served, going by the plot shown below. Again, due to the large imbalance of the number of agencies between clusters when k=2, the preference from the food bank sizing k=3 was used. The Elbow method and cluster plot is shown in the figure below in Figure 3.

The three clusters were color coded as follows: the red map showing the large cluster, the blue map showing the medium cluster, and the green showing the small cluster.

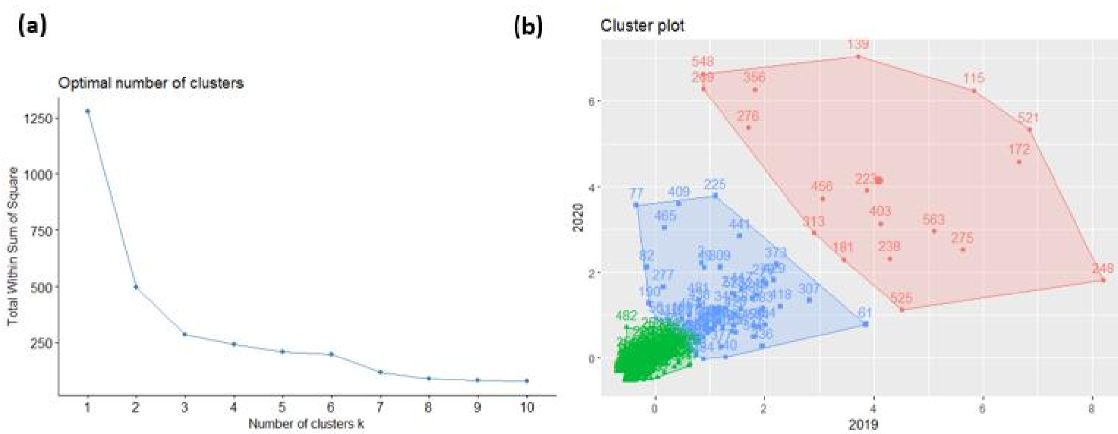

Figure 3: Cluster analysis for food distributed by the number of persons served **(a)** the elbow method demonstrating optimal number of clusters, and **(b)** cluster plot of agencies for food distributed by gross weight.

Table 2: Descriptive Analysis of Clusters by Number of Person Served

	Min		Mean		Max	
	2019	2020	2019	2020	2019	2020
Small	3	2	874	1120	15838	61261
Medium	15	10	2192	2648	36454	60793
Large	13	15	1478	3670	15419	78724

Table 2 below shows descriptive analysis within each cluster and in analysis of the data geographically, Figures 4 a and b show the filled map and the dot map of the various agencies within the network. Tableau was used to develop the filled map and the dot map of the various agencies. The dot map is a mapping method where each dot serves as a proxy location for each partner agency, and the filled map uses geospatial analysis to discover patterns correlating to geographic locations.

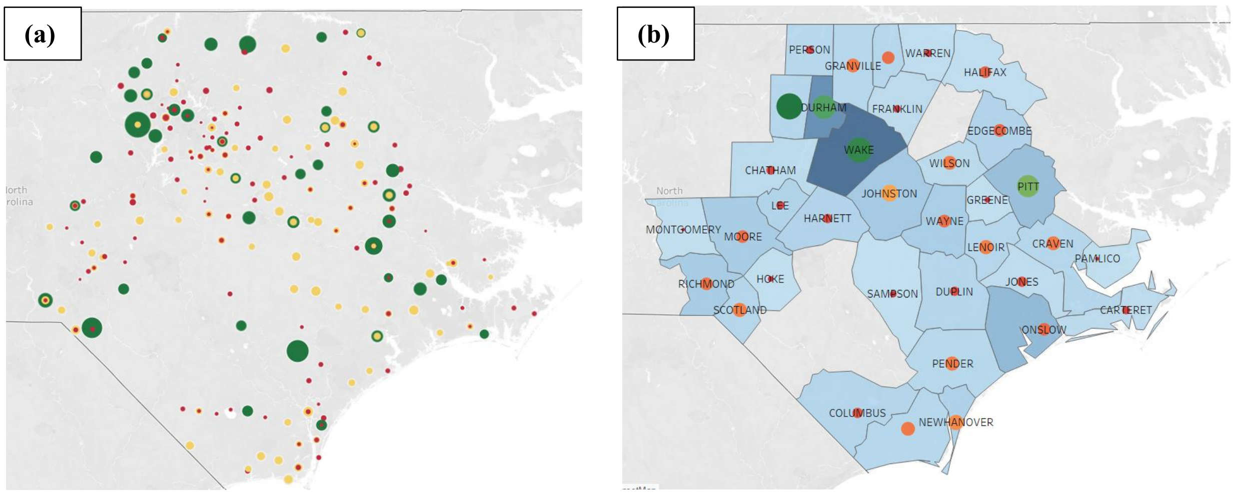

Figure 4: Visualizations of agencies **(a)** the dot map of agencies by gross weight, and **(b)** the filled and dot map of agencies in network.

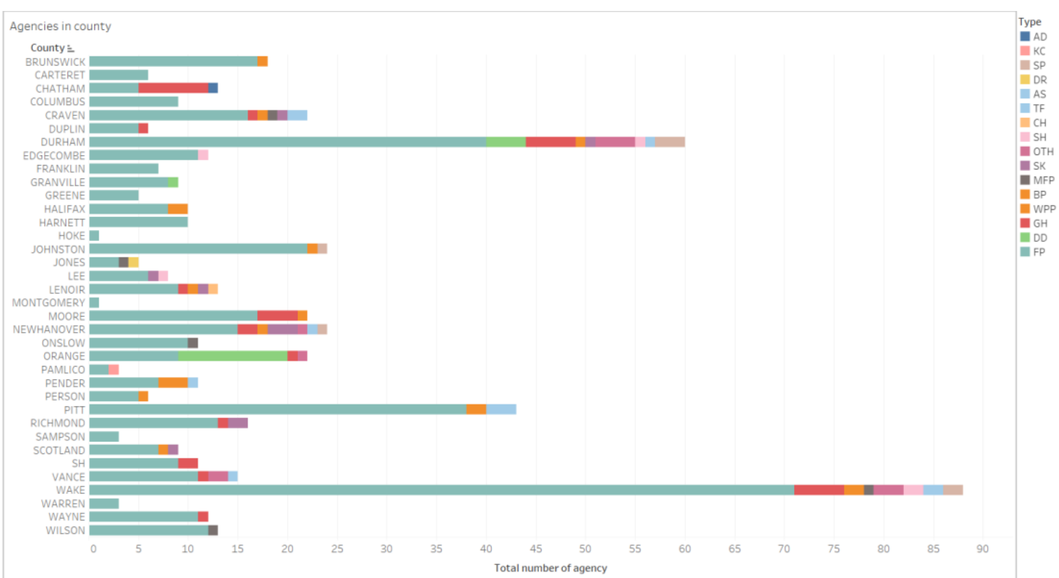

Figure 5: Visualization of agencies in counties by type

The various partner agencies in the counties depending on the activities carried out could be considered as an Adult daycare (AD), Kid's café (KC), School pantry (SP), Disaster relief (DR), After School (AS), Soup kitchen (SK), The Emergency Food Assistant Program (TEFAP), Mobile food pantry (MFP), Backpack (BP), Weekend power pack (WPP), Group home (GH), Direct distribution (DD), Food pantry (FP), serving children (CH), shelter (SH), or others (OTH). Figure 5 above shows the visualization of partner agencies by type in the different counties served.

4.0 Discussion and Conclusion

The intent of this study was to explore clustering analysis and visualization as a tool in assisting Food Bank decision-makers working with multiple partner agencies. From the cluster analysis conducted using the food distributed by gross weight, there are 54 agencies in cluster 1, or the large cluster, 153 agencies in cluster 2, or the medium cluster, and 370 agencies in cluster 3, or the smaller cluster. Therefore, agencies in cluster 1 distributed more food than those in cluster 2 and 3 for the years considered; 2017 and 2018. Cluster analysis for the person-served dataset when partitioned into 3 clusters of the 577 agencies considered, 21 were placed in cluster 3 or the large cluster, 75 were placed in cluster 2 or the medium cluster, and 481 were placed in cluster 1 or the small cluster. Accordingly, agencies in the large cluster served the most people compared to the other clusters.

The Dot map in Figure 4a gives a better perspective of the location of partner agencies and the size based on the amount distributed. They were color-coded based on their clusters and the sizes of the dots were based on the total amount of food by pound distributed in both years. Figure 4b also gives clarity on a county level. The colored filled demonstrates the number of persons served, with darker colors indicating a higher number. The dot placed on it demonstrates the amount of food distributed and the various sizes indicate the amount. As shown in Figure 5, the differences in the number and types of programs offered by each partner agency in a county, the number of people served, the amount of food distributed, and the frequency of visit-which affects the total number of food the agency has in-store make the behavior of each agency unique.

In summary, visualization of the partner agencies based of their clusters provides a quick and easy way for operations managers to take advantage of the analytical power in making critical decisions as to how food should be distributed among partner agencies. This has the potential to increase the effectiveness and the efficiency of the food bank operations. In the future, more complex and unstructured data will be utilized to give a better representation of the various agencies in the counties.

Acknowledgements

The authors would like to thank Charlie Hale and Ansen Gunawan of the Food Bank of Central and Eastern North Carolina (FBCENC) for their help and support. This project is partially supported by NSF award #2100855 and NSF Award #173528. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] USDA. (2021, May 29). Stores accepting snap online. Retrieved January 25th 2023, from https://www.fns.usda.gov/snap/online-purchasing-pilot

[2] T. Hasnain et al., “Resiliency within the Socio-Ecological System of a Large Food Bank Network: Preparing, mitigating, responding, and recovering from Hurricane Florence,” International Journal of Disaster Risk Reduction, vol. 88, p. 103580, Apr. 2023, doi: https://doi.org/10.1016/j.ijdrr.2023.103580.

[3] Food Bank of Central and Eastern North Carolina, 2022, “Hunger Studies, “Last retrieved January 20, 2022. http://www.foodbankcenc.org/site/PageServer?pagename=hunger_studies

[4] S. Mannor et al., “K-Means Clustering,” Encyclopedia of Machine Learning, pp. 563–564, 2011, doi: 10.1007/978-0-387-30164-8_425.

[5] S. Na, L. Xumin, and G. Yong, “Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm,” 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, Apr. 2010, doi: 10.1109/iitsi.2010.74.

[6] C. Patil and I. Baidari, “Estimating the Optimal Number of Clusters *k* in a Dataset Using Data Depth,” Data Science and Engineering, vol. 4, no. 2, pp. 132–140, Jun. 2019, doi: 10.1007/s41019-019-0091-y.

[7] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” Journal of Computational and Applied Mathematics, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

[8] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 2, pp. 411–423, May 2001, doi: 10.1111/1467-9868.00293.

[9] R. Kuhn and D. P. Culhane, “Applying Cluster Analysis to Test a Typology of Homelessness by Pattern of Shelter Utilization: Results from the Analysis of Administrative Data,” American Journal of Community Psychology, vol. 26, no. 2, pp. 207–232, Apr. 1998, doi: 10.1023/a:1022176402357.

[10] J. L. Black and D. Seto, “Examining Patterns of Food Bank Use Over Twenty-Five Years in Vancouver, Canada,” VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, vol. 31, no. 5, pp. 853–869, Oct. 2018, doi: 10.1007/s11266-018-0039-2.