Redefining Food Access in Rural Counties of Pennsylvania

Claudia Shrefler, Professor Abby Flynt
Clare Boothe Luce Research Program
Department of Mathematics, Bucknell University, Lewisburg, Pa.

Introduction

Food access is defined as the percent of individuals in urban census blocks living more than one mile from a major supermarket or grocery store, and the percent in rural census blocks living more than ten miles from a major supermarket or grocery store (USDA).

Through comprehensive literature review, groups of variables were chosen that may explain food access better than the current definition:

- **Variable group 1:** low food access %, meal cost, commute time, and high school diploma %
  - Identified in "Understanding Food Access in a Rural Community" by Rosalie M. Rodriguez and Kamini Maraj Grahame[3]
- **Variable group 2:** unemployment %, median household income, obesity %, diabetes %, and SES index
  - Identified in "Diet-related chronic disease in the northeastern United States: a model-based clustering approach" by Abby Flynt and Madeleine I. G. Daepp[4]

Goal:
Use publicly available data to better define food access, specifically for rural counties

Data

Multiple publicly available datasets were gathered and carefully cleaned from government websites, such as Food Environment Atlas, American Community Survey, US Census Bureau, and other listed below.

The cleaned data were then merged into one large dataset for analyzing, and the variables were annotated. There were approximately 260 variables gathered in total for exploration.

Shape files for the counties were also gathered which allowed the analyses to be visualized using maps.

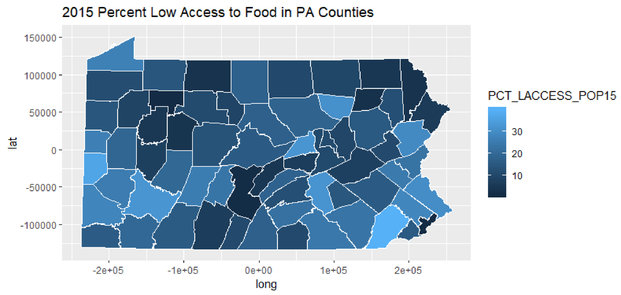
Data Sources:

1. American Community Survey
2. Data.gov
3. Feeding America
4. Food Environment Atlas
5. PA Legislature Data
6. PA County Data
7. US Census Bureau
8. US Department of Labor
9. US Department of Agriculture

Exploratory Data Analysis

Exploratory Data Analysis was conducted on different variables and visualizations were created by mapping these analyses onto the map of Pennsylvania using the ggmap package in R.

Below is a map of the low access to food percentage (determined by physical distance from grocery stores) for counties in 2015, where a lighter shade represents a larger percentage of the county with lower access to food.



Methods

Spatial Hierarchical Clustering

Spatial hierarchical clustering was used to partition observations into groups based on similarities in the variables to see patterns in the data.[1]

- Hierarchical clustering: looks at distances between observations and merges based on different criteria
- Spatial clustering: counties can only merge if they are neighbors, i.e. they share a border

The gap statistics was used to choose the number of clusters in the final solution.[2]

This is an example of a dendrogram which illustrates a hierarchical clustering solution. Each letter represents a county and they are sequentially merged based on their distance/proximity to each other. The dendrogram is horizontally cut at a particular height to create a final solution.

Results

Rodriguez Spatial Clustering Summary:

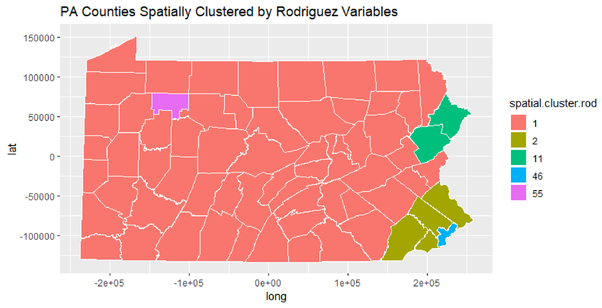
CLUSTER	LOW ACCESS (%)	MEAL COST INDEX	COMMUTE TIME (MIN)	HS DIPLOMA (%)
ELSE	3-36	2.60-3.50	18-32	83-93
C, D, M, B	39.7, 15.8, 28.4, 30.2	3.88, 3.62, 3.50, 3.41	27.8, 28.6, 28.1, 29.3	93, 92, 94, 93
M, P	29.4, 6.2	3.15, 3.44	38.6, 44	89, 90
PHILADELPHIA	1.2	3.23	32.7	82
FOREST	1.3	2.81	24.3	80

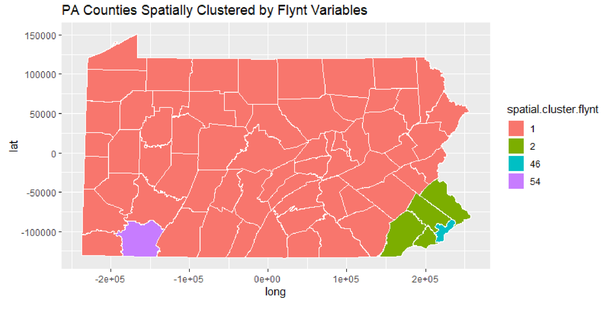
Flynt Spatial Clustering Summary:

CLUSTER	PCT DIABETES	PCT OBESE	UNEMPLOYMENT RATE	SES INDEX	MEDIAN HH INCOME
ELSE	7-15	26-38	3.5-6.3	-4 – 3.7	39-66
C, D, M, B	7.9, 9, 7.4, 9.6	23, 24, 24, 25	3.5, 4.5, 3.8, 4.1	-3.2, 1.7, -2.8, -3.7	92, 67, 84, 80
PHILADELPHIA	11.5	29.8	6.1	15.2	41
FAYETTE	15.9	38.5	6.3	3.18	42

Key:
C,D,M,B = Chester, Delaware, Montgomery, Bucks Counties
M,P = Monroe, Pike Counties

The numerical summaries above highlight the unique observation values that are responsible for the clustering of certain counties using variable group 1 and variable group 2.





Conclusion

At the conclusion of this research, it became clear that one variable representing someone's physical distance from a grocery is a poor way to determine their food access / food insecurity.

There were multiple relationships between variables that proved interesting throughout the research. For example, Philadelphia county had good access to food as seen in the EDA section, but has the highest rate of food insecurity in Pennsylvania (food insecurity rate measures the percentage of the population that experienced food insecurity at some point during the year)[5].

This demonstrated that the proximity to a grocery store is not a direct indication of one's food insecurity. In fact, there appears to be a negative relationship between the two.

Future research on this subject should
- Look at new / more subsets of variables
- Try different clustering parameters
- Extend to other states

One variable that was mentioned frequently in literature and previous research that was believed to have a high impact on food access is transportation. This not only includes how long it takes to get to a grocery store, but also the public transportation available in the area among other things.

This variable was unfortunately not present in any of the data that were gathered but can perhaps be constructed through existing data in future work.

References

1. Craig Anderson, Duncan Lee, Nema Dean, Identifying clusters in Bayesian disease mapping, *Biostatistics*, Volume 15, Issue 3, July 2014, Pages 457–469,
2. Tibshirani, R., Walther, G. and Hastie, T. (2001), Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63: 411-423.
3. Rosalie M. Rodriguez & Kamini Maraj Grahame (2016) "Understanding Food Access in a Rural Community. *Food, Culture & Society*. 19:1, 171-194.
4. Flynt, Abby, and Madeleine I.G. Daepp, "Diet-related Chronic Disease in the Northeastern United States: A Model-based Clustering Approach." *International Journal of Health Geographics* 14.1 (2015): n. pag. Print.
5. Gundersen, C., A. Dewey, M. Kato, A. Crumbaugh & M. Strayer. *Map the Meal Gap 2019: A Report on County and Congressional District Food Insecurity and County Food Cost in the United States in 2017.* Feeding America, 2019.

Acknowledgements

Clare Boothe Luce Research Scholarship

I'd like to thank the Clare Boothe Luce Research Program for funding this research, as well as for their work in supporting women in STEM fields in which they are underrepresented.