Food Deserts: a Meta Analysis of Determinant Factors

Aidan Flinton

Introduction

1. Food deserts are considered to be regions in which many individuals do not have cheap or convenient access to healthy food. This could occur due to an area's low income population, or a lack of nearby stores carrying unprocessed food, forcing citizens to instead subsist on low quality food simply because they lack options to acquire beneficial nutrients.

2. According to the Annie E. Casey Foundation (2021), nearly thirteen percent of the total population of the United States, totalling around forty million people, live in areas classified as food deserts due to their low income and low food accessibility status (p. 2). Therefore, the problem of food insecurity should be addressed as a national concern with attempts to improve the abysmal inaccessibility percentage. However, the avenue through which the goal of decreasing food-inaccessible areas remains a point of contention amongst scholars with some insisting that specific regional factors may be targetable in improving an area's food accessibility (Luan et. al. 2015 p. 9). Conversely, others have found that examining specific factors such as "populations, rates of abandoned or vacant homes, and residents levels of education, incomes, and unemployment" could be used to predict a region's food desert status (Dutko et al. 2020, p.1).

3. Therefore, the aim of this study is to determine whether individual regional factors could have a strong direct correlation with whether an area is considered to be a food desert. Additionally, this study focuses on whether corrective actions addressing each

---

Keywords: food deserts, food insecurity, meta analysis, quantitative research methodologies, regional factors

Citation (modified APA)
Flinton, Aidan. (2021, November 4). *Food Deserts: a Meta Analysis of Determinant Factors*. Hokies Write. http://hokieswrite.com.

---

determined and analyzed criterion are warranted. If this study finds that an individual factor holds a significant effect on a state's population living in food deserts, the factor's effect will be quantified to determine which should be corrected more immediately. This study holds the potential to improve the lives of those living in food insecurity by recommending methods to address factors directly contributing to their subsistent way of life.

Methodology

4. The primary data sets analyzed throughout this project include the 2017 Food Environment Atlas and the 2019 Food Access Research Atlas, FEA and FARA, respectively. These sources were chosen for analysis due to their established credibility and large collections of data. The US Department of Agriculture, USDA, published both and has established its credibility through its various other studies, datasets, and legislation. Additionally, both sources contain expansive avenues to research, with over 350 combined possible features related to each census tract's classification as a food desert.

5. The data from each collection was collected, aggregated, and analyzed using the Python programming language within Google Colaboratory. This programming language was used due to Colaboratory's (colab) ease of use and maximum user data limit--a limit that none of the data fittings were expected to exceed. Additionally, several Python libraries were used to improve the ease with which large Comma Separated Value (csv) data sets could be analyzed. These included NumPy, pandas, Matplotlib, and scikit-learn (the erratic capitalization of these libraries was intentional as that is a Python naming convention for some of them). These libraries were chosen because of their common usage across data science and math fields alike for manipulating and displaying large quantities of matrix data (Millman, 2011, p. 9).

6. Data from each of the data sets were first imported to colab using an integrated file reader method before being converted to a pandas data frame in order to make them more accessible. However, before the data was analyzed the data frame was converted into a NumPy array in order to make the rows and columns of the array more easily separable for regression lines to be fit onto the data and to be graphed with other libraries.

7. Then the NumPy arrays holding the data from each data set were separated into a column to be considered as the label and graphed along the y-axis and a set of columns to be considered as the features

used to determine each label. These collections were further divided, with 80% of each being used to train the machine learning models that will correlate many features to each label. Another 10% of each collection was used to validate the model's fit and make final corrections for accuracy and the last 10% was used to determine how accurate the model's predictions are in relation to the actual data.

8. Following the separation of the data into verifiable sets, the training set was entered into the multiple regression model function offered by the Sci kit Learn library. This determined the importance of each feature offered by the data set by determining whether that census tract corresponded to an area flagged as a food desert within the data set. After, this model was fitted to the data, its coefficients were output which indicated which features were most influential on a census tract's food desert classification based on the magnitude of the coefficient, with larger coefficients being more influential and negative coefficients decreasing the likelihood of an area to be considered a food desert.

9. Next, the data corresponding to the coefficients of the greatest magnitudes was deemed of the greatest importance and considered in greater depth when developing the results. However, at this point in the analysis, the narrow scope of the FARA was revealed, so subsequent methods were applied only across subsections of the FEA (Rhone 2021, Rhone 2019).

10. This dataset was grouped by the state to find the average population percentage of each state that was considered low access and low income. Similarly, average values per state of several other features, from subcategories of the FEA, including nearby stores, restaurants, and local food accessibility, are to be compared to the state access percentages. This comparison was intended to provide insight into which of the features from each subcategory held the greatest bearing on each state's low access rates.

11. The values of low access and each subcategory by state were converted into their z-scores (a unitless value that allows unlike units to be compared based on their variation from their mean) to invite a comparison amongst like elements. This was all done before having a linear regression model from Scikit learn fit across 80% of their data. The coefficients were output to determine which features were the most influential. Similarly, the mean squared deviation of the training and validation sections of the data was output to determine the model's accuracy.

12. Finally, the aspects of each subcategory that held the highest correlation coefficients with state access percentages were graphed against low access percentages with a line of best fit applied to them. This indicated how strongly each subcategory correlated with a state's access rates.

Results

13. Analysis of each dataset revealed information about separate factors on state food access, or lack thereof.

14. Unfortunately the results from the FARA were largely irrelevant to the purposes of this study, only revealing what was already generally understood from the research of Dwane Jones (2016)-- low access areas that may be considered food deserts are more likely to receive higher volumes of federal aid (p.239). The only other thing the FARA taught us was that there is a significant relationship between the financial statuses of racial groups and the area's designation as low access. However, this data was considered too niche to be applicable as general factors for a region's consideration as a food desert because it is too greatly influenced by racial stigmas that decrease an area's shared wealth rather than targetable features that may hold a similar or larger influence.

15. Luckily, analysis of the FEA dataset provided much more insightful results, in fact, too many results to analyze in the short time frame of this project, so factors were considered based on which factors from each analyzed subset had the greatest positive and negative effects on a state's average county food desert population percentage. Each of these factors' effects on a state's low access population percentage were graphed linearly, exponentially, and with power law relationships in order to demonstrate the type of effect individual county characteristics hold on a state's population percentage living within a food desert. Which of these graphs holds the closest to a linear pattern amongst the data points reveals which relationship that feature follows when indicating a county's consideration as a food desert. However, if none of the graphs for a data feature align with a linear pattern then it is likely that a more complex relationship is present or none at all, with a merely coincidental effect on data during the model fitting process, and further research would be required to make this distinction.

16. The effect that the presence of different types of stores in a county held on that state's low access population was narrowed down to the effects of total convenience stores and the change in grocery stores from the past 5 years. The Pearson correlation (r-value) was

graph in z-space in order to show which relationships have the strongest suggested correlations, however, Pearson's r was lower than expected for all examined relationships for both the graphs with positive and negative model weighting. The Pearson Correlation is a score between -1 and 1 suggesting the closeness of a graph to the line of best fit associated with it, with scores closer to + or - one being a stronger positive or negative correlation respectively, while scores closer to 0 indicate a weaker correlation. When examining the positive correlation between convenience stores and low access states r = -4.16 x 10-19, -4.72 x 10-18, and 2.43 x 10-18 for the linear, exponential, and power law relationships, respectively (Figure 7). These values were similarly low for the relationship between percent change in a state's grocery stores and its low access percentage with r = -1.08 x 10-17, -4.16 x 10-18, and -5.27 x 10-18 for each respective relationship (Figure 8).

17. The effect that different types of restaurants had on a state's low access population was concised to the effects of fast food restaurants and the effects of traditional dine in restaurants. When considering the Pearson correlations for restaurant accessibility versus low access population, these correlations were similarly low with the positively related full restaurants versus state low access percentage scoring r = 0, -8.47x10-18, and 2.78 x 10-18 for their respective mathematical relationships (Figure 9). Correlations for the negatively related fast food restaurants versus state low access percentage was nearly as low with r = -2.78 x 10-19, -3.82x10-18, and 9.33x10-18 for their respective mathematical relationships (Figure 10).

18. The indication that different levels of access to locally grown food held on a state's low access population was limited to the effects of farmer's markets per county and the effects of a five changes in the percentage of county area that was vegetable farm land. Similarly, the pearson correlations for the positively related farmers markets were r = 4.16x10-19, 1.67x10-18, and -3.12x10-18 while the correlations for the negatively related Vegetable farm percent change were r = -1.11x10-18, -5.55x10-19, and 0 (Figure 11, 12).

19. The importance of these low R values is indicated because the graphs with the highest Pearson correlation were graphed with a line of best fit that demonstrates how minute the relationships between specific data points are to a state's low access population percentage (Figure 1, 2, 3, 4, 5, 6). However, these graphs were primarily included to demonstrate how completely underwhelming and inconclusive the correlation between even the most important individual features and

and food-insecurity measurement were. Therefore, these graphs do not warrant further individual discussion since the primary conclusions that can be drawn from this data are that even the features with the greatest correlation to food scarcity have a near inconsequential individual correlation and should not be individually targeted while attempting to improve food security.

Analysis

20. The linear regression models from scikit learn were able to locate notably important features in determining a state's low access population percentage, with mean squared errors (MSE) consistent across the train, test, and validate data segments, meaning that they were likely fit correctly to the dataset. Despite this, the MSEs were quite high--which would indicate that the model remained somewhat inaccurate. However, since many data point values exceeded those of the MSE, it was deemed acceptable to allow for further analysis of the relationships revealed by this model.

21. This further analysis revealed the miniscule Pearson correlations for every mathematical relationship of every feature versus low access. It would also indicate that there is enough excess noise, or outlying data, in every category used to determine a region's food desert status to prevent the use of a single factor in the determination of this percentage (Benesty, 200, p. 758). Again, this lack of single factor accuracy is represented by the visually poor fitting lines graphed along the relationships with the highest Pearson correlations from each averaged data category.

22. While the overall models for each of these relationships fell within reasonably satisfactory ranges of variation, examination of the individual vectors that comprised them showed largely irrelevant correlation. This indicates that factors need to be considered in context to each other in order to predict a region's low access status. Therefore, a county or state's likelihood of increasing its percentage of low access population, or harboring more food deserts must be determined through a more holistic analysis than individual statistical examination. Additionally, it is possible that the statistics recorded by the USDA in the FEA do not hold a significant correlation with low access populations and that data based on a currently unknown factor must be collected in order to predict how a state's food accessibility percentage will change.

Discussion

23. Despite the shortcomings of known specific factor analysis in predicting a state's food inaccessible population, the models created

from composite factors returned accuracies that were far more acceptable, indicating that holistic review provides a more accurate determination than specific review. This aligns with the findings of researcher Francine Rodier (2017) that "geographical access to supermarkets is not the main factor fostering the purchase" of healthy foods because their conclusion that "a changing mediation process through… diversification seem to be more significant," identifies that a broad set of features need to be examined with context to one another to make the determination of what influences a food desert's formation (p. 2). Furthermore, the conclusion: individual factors do not have an impact great enough to be scaled to a national level, is validated by researcher Jerry Shannon(2016) who found that focus should be on drawing a "broader body of research on cities and mobility to develop an understanding of food provisioning that moves beyond just proximity to major food retailers" in order to combat the multifaceted formation of food deserts, rather than focusing on specific impact factors.

24. Despite the evidence gathered from dataset analysis and support from other researchers, there remain some authors that insist specific factors contribute clearly enough to food deserts' existence that a targeted approach may still be feasible. One such researcher explained that increased "support [for] retail food environments" would serve to decrease food deserts due to their finding of a correlation between a lack of local retailers and food deserts. However, the source used to validate this claim, Evaluating Food Environment Assessment Methodologies, seems to indicate the opposite of their claim, instead stating that food desert "perceptions were not highly correlated with objective FE [food environment] measures" (Minaker 2013 p.4). Therefore, arguments in favor of single factor food desert contributors are tangentially supported at best, with suggested solutions providing general relief for those living in food deserts rather than providing relief by decreasing a specific factor.

Conclusion

25. In conclusion, holistic review of a state and its counties is required in order to predict that state's food inaccessibility severity because analysis of individual recorded factors is insubstantial in making this judgement. However, this may be a result of limitations in the currently accessible information about factors surrounding a state's food accessibility. Therefore, an avenue for future researchers to pursue could involve an evaluation of not previously considered state and county socioeconomic elements that may hold less immediately

apparent effects on food accessibility to determine if any of these factors have a strong individual correlation to food deserts. Additionally, the analysis conducted in this study could be replicated on a smaller scale by examining how each of the factors from the FEA impact food desert status on a state level, because it is possible that the factors tested in this study do correlate to food deserts, but are overruled by other factors on a nationwide scale, decreasing their perceived correlation. This could potentially be conducted by examining county averages based on census tract data and comparing that to county food desert status rather than by examining state averages based on county data.

26. Limitations for this study included the limited time to process and analyze datasets and the inaccessibility to datasets that contained national records for parameters that held a high enough individual correlation to state low access percentages to offer significant insight into which factors could be targeted to alleviate food desert stresses on families. While other credible national datasets are published, such as the Food Expenditure Series, FoodAPS National Household Food Acquisition and Purchase Survey, or the Food Price Outlook that may have offered insights yielding more concrete conclusions on which factors to target to absolve some areas of food insecurity, the narrow time frame of this project prevented their further analysis. Additionally, this project's research was limited in that there were no feasible methods through which to obtain new data on national food statistics to develop conclusions outside of what was discernible from existing data.

27. While data collection on a national scale would not be feasible in most college level research projects, it does offer an avenue for further research through data collection of new regional aspects to be compared to food desert populations. Further research that could be proposed to follow this study would be to research how profit margins would be influenced by large grocers if they were to open more locations in areas considered food deserts. While large healthy grocers already almost certainly conduct internal studies that influence their decision to abstain from opening stores in locations whose food security would benefit from their presence, publicly releasing expected data on the profits or losses associated with opening stores in high need locations could influence government intervention to subsidize expansion in these areas to cover potential losses until a foothold can be established in the area and incentivize further development. These incentives would not be out of place alongside other farming subsidies designed to improve healthy food distribution across the country

such as the "Farm Bill" passed in 2018 to boost dairy, improve SNAP benefits, and improve farm loans (Agriculture Improvement Act of 2018).

References

Benesty, J., Chen, J., & Huang, Y. (2008). On the Importance of the Pearson Correlation Coefficient in Noise Reduction. IEEE Transactions on Audio, Speech, and Language Processing, 16(4), 757–765. https://doi.org/10.1109/TASL.2008.919072

Conaway, K. M. (2018, December 20). H.R.2 - 115th Congress (2017-2018): Agriculture Improvement Act of 2018 (2017/2018) [Legislation]. https://www.congress.gov/bill/115th-congress/house-bill/2

Dutko, P., Ver Ploeg, M., & Farrigan, T. (Eds.). (2012). Characteristics and Influential Factors of Food Deserts. https://doi.org/10.22004/ag.econ.262229

Foundation, T. A. E. C. (2021, February 14). Exploring America's Food Deserts. The Annie E. Casey Foundation. https://www.aecf.org/blog/exploring-americas-food-deserts

Gregg, C. T. (2017, October 18). Food Deserts and Food Swamps: A Primer | National Collaborating Centre for Environmental Health | NCCEH - CCSNE. National Collaborating Centre for Environmental Health. https://ncceh.ca/documents/evidence-review/food-deserts-and-food-swamps-primer

Jones, D. (2016). Addressing Urban Health and Food Policy through Resiliency Food Hubs: A Case Study from Washington, D.C. Innovations in Public Health. Saint Louis University Journal of Health Law & Policy, 10(2), 239–248.

Luan, H., Law, J., & Quick, M. (2015). Identifying food deserts and swamps based on relative healthy food access: A spatio-temporal Bayesian approach. International Journal of Health Geographics, 14(1), 37. https://doi.org/10.1186/s12942-015-0030-8

Millman, K. J., & Aivazis, M. (2011). Python for Scientists and Engineers. Computing in Science Engineering, 13(2), 9–12. https://doi.org/10.1109/MCSE.2011.36

Minaker, L. M. (2013, Spring). Evaluating food environment assessment methodologies: A multi-level examination of associations between food environments and individual outcomes. ERA. https://doi.org/10.7939/R3X39P

Rhone, A. (2018). 2017 Food Environment Atlas Data Documentation. [Data File]. Retrieved from https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/

Rhone, A. (2021). 2019 Food Access Research Atlas Data Documentation. [Data File]. Retrieved from https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/

Rodier, F., Durif, F., & Ertz, M. (2017). Food deserts: Is it only about a limited access? British Food Journal, 119(7), 1495–1510. https://doi.org/10.1108/BFJ-09-2016-0407

Shannon, J. (2016). Beyond the Supermarket Solution: Linking Food Deserts, Neighborhood Context, and Everyday Mobility. Annals of the American Association of Geographers, 106(1), 186–202. https://doi.org/10.1080/00045608.2015.1095059

Appendix

Figure 1.

Line of Best Fit attached to Relationship with greatest R from Figure 7

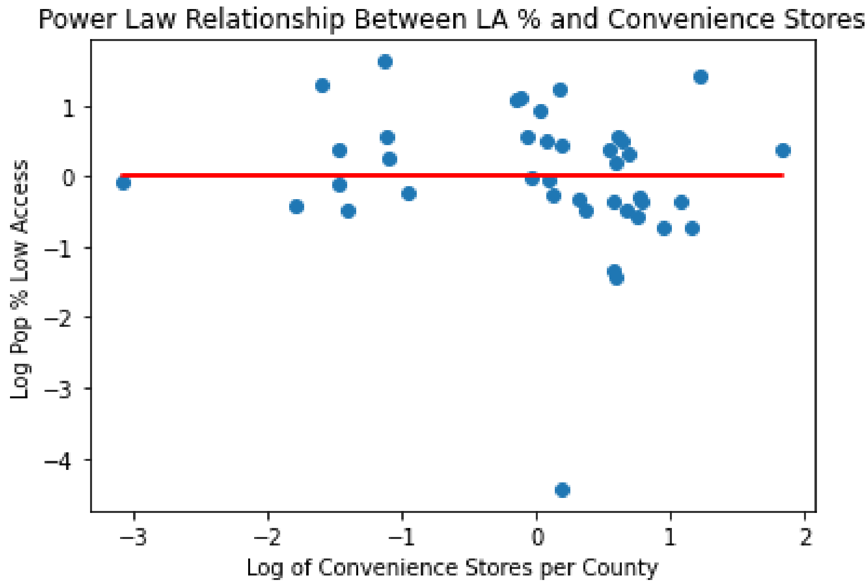

Figure 2.

Line of Best Fit attached to Relationship with greatest R from Figure 8

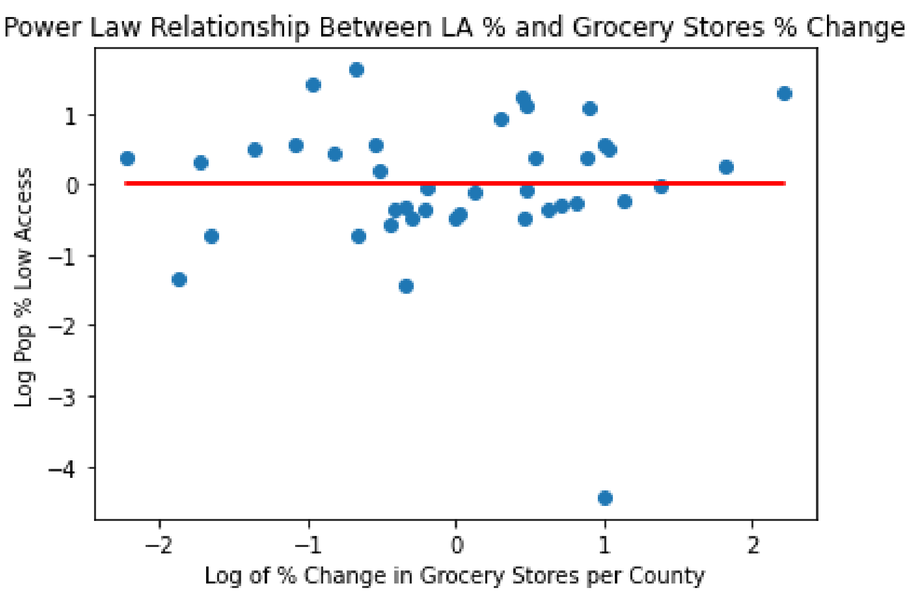

Figure 3.

Line of Best Fit attached to Relationship with greatest R from Figure 9

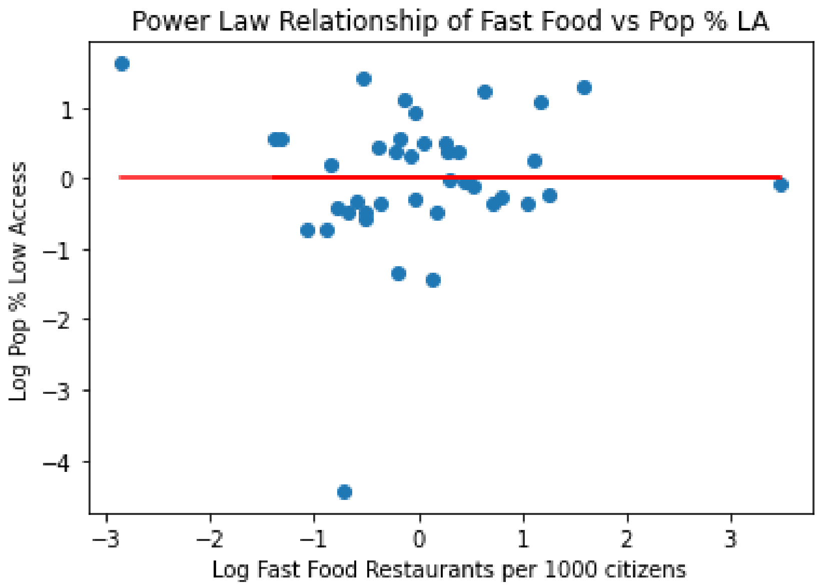

Figure 4.

Line of Best Fit attached to Relationship with greatest R from Figure 10

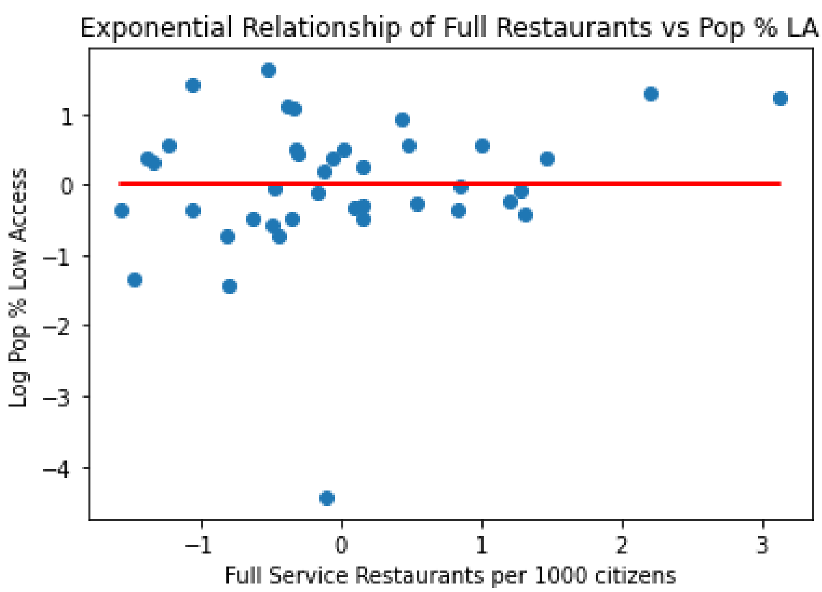

Figure 5.

Line of Best Fit attached to Relationship with greatest R from Figure 11

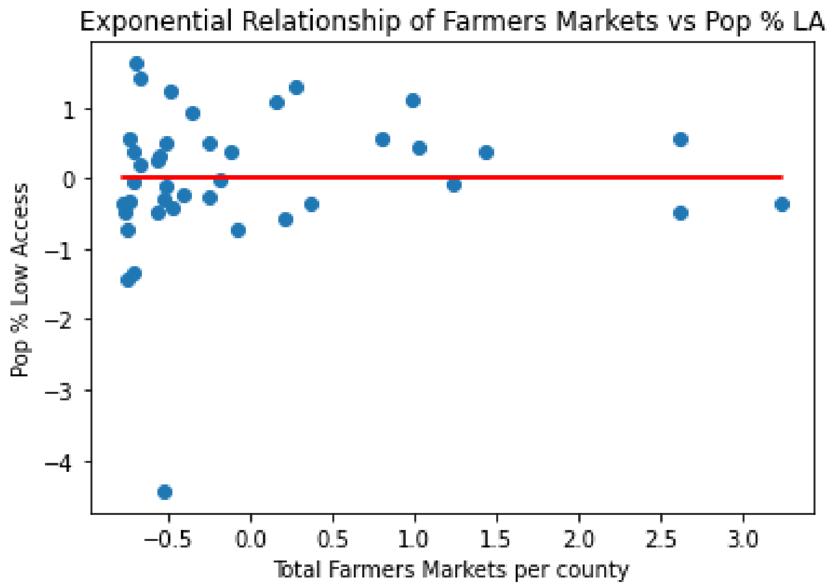

Figure 6.

Line of Best Fit attached to Relationship with greatest R from Figure 12

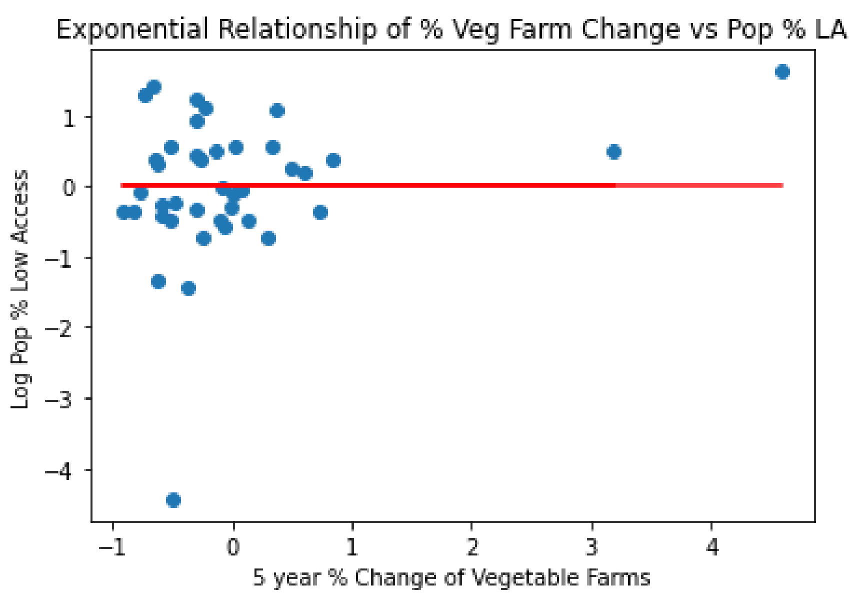

Figure 7.

Graphs of Each Mathematical Relationship for Convenience Stores vs Food Access

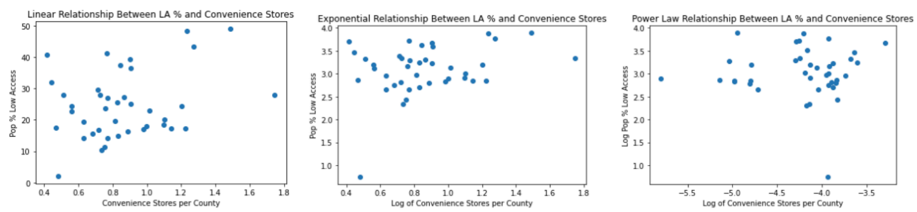

Figure 8.

Graphs of Each Mathematical Relationship for Change in Number of Grocery Stores in 5 Years vs Food Access

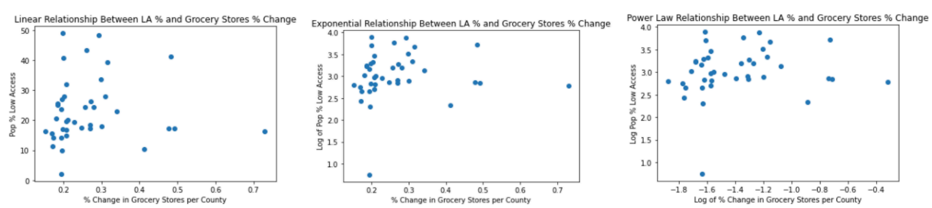

Figure 9.

Graphs of Each Mathematical Relationship for Fast Food vs Food Access

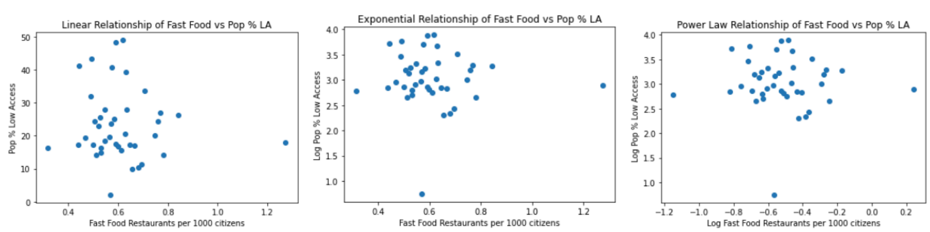

Figure 10.

Graphs of Each Mathematical Relationship for Restaurant vs Food Access

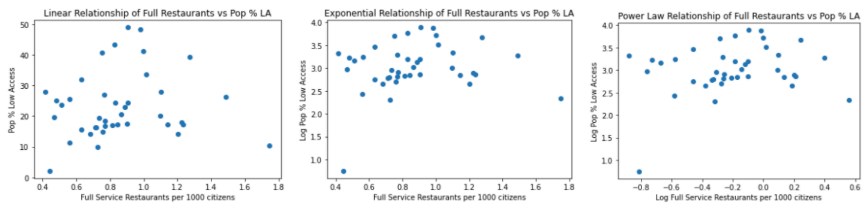

Figure 11.

Graphs of Each Mathematical Relationship for Farmers Markets vs Food Access

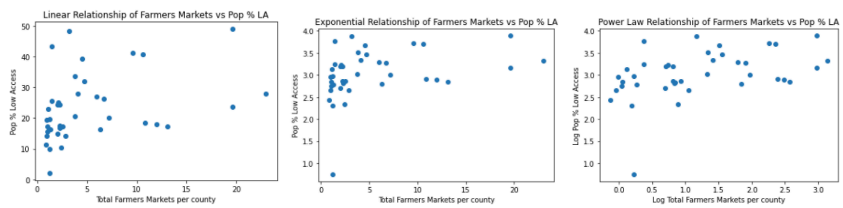

Figure 12.

Graphs of Each Mathematical Relationship for Change in Vegetable Farms in 5 Years vs Food Access

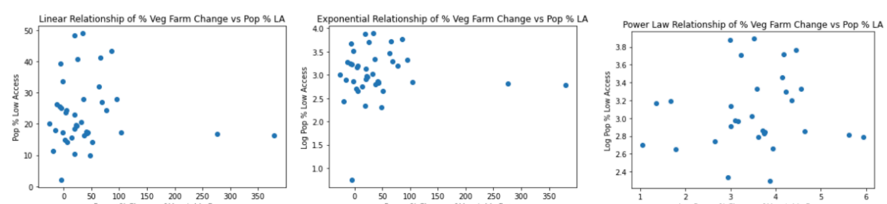

Figure 13.

A Link to My Code
https://colab.research.google.com/drive/1p6DAh9gCpVzFwF_SqRCuQIk-wCgVaFmNy?usp=sharing

Instructor: Sherri Craig (Assistant Professor)

Author Permissions:
I grant permission to include this project/paper in the Document Gallery on HokiesWrite.com.
I grant permission for this project/paper, if selected, to be published in the program's textbook.
I grant permission for this project/paper to be shared in-class.
I grant permission for this project to be used in professional workshops and for GTA education purposes.

Note: The citation information on p. 1 follows modified APA, incorporating the writer's full name as an acknowledgment of a more fully humanistic regard for authorship.