A State-wide Analysis of Food Access and Agriculture

Mohammed Abdul Karim
Computational and Data Sciences
George Mason University

***Abstract*—The aim of this paper is to see the correlation and relationship between the state and production of agriculture and a populations' access to food. The data was collected from the Food Environment Atlas provided by the Economic Research Service of the United States Department of Agriculture. The Atlas records these variables at a county-level, where each row represents a county in the United States. By using Python and the Pandas library, the dataset was cleaned and converted to a state-level. Using JMP Pro 15 to get the correlation and relationship between the variables, we can see that there is no significant relationship between food access and agriculture despite other similar studies' results. This begs the question of what other indicators, either in the original raw data or data provided by another party, have an effect on our accessibility to food, if any.**

***Index Terms*—USA, agriculture, food, access, security, insecurity, state, county, analysis, correlation, research**

I. Introduction

The number of farms in the United States continues to slowly decrease. However, productivity has been increasing over the years as a result of the development of technology to help with farming; farms have increased their outputs without having to increase their inputs (United States Department of Agriculture, 2022). According to the US Department of Agriculture, about 10.5% of houses in the United States were food insecure throughout 2020 (2022). Using the Food Environment Atlas provided on their website, my goal was to find the correlation between food insecurity and agriculture. According to the Australian International Food Security Research Centre, food insecurity is also becoming a global issue; agricultural research, development and productivity helps improve the livelihoods and is beneficial for those who do not have easy access to food. Food insecurity also comes in many forms, such as what is available in supply or how easy it is to physically access; it is a complex issue that needs to be addressed (2014). From reviewing somewhat similar studies, I hypothesized that there would be some correlation or relationship between agriculture and food access. However, to fully explore the issue, a data set involving two sets of indicators that best represent the state of our agriculture and accessibility to food needs an in-depth analysis using linear regression where the R^2 value is measured and obtaining the correlation coefficient between multiple sets of indicators.

My motivation behind researching this topic is because I believe the issue is going to continue to spread, even in areas where the economy may appear to be doing well. I believe continued research of this topic will lead closer to answers on how to address and possibly fix the issue of people not being able to afford food. Even if the results show that no significant relationship or correlation exists between the two variables, that presents the notion that the issue may lie elsewhere; whether it is recorded or not. Based on the lack of studies involving this specific topic, the primary objective of this research paper is to be used as a form of groundwork stemming other works, hopefully scoping towards regions where food insecurity is a bigger issue than in the United States.

II. Literature Review

Most studies among the topic of agriculture and food security involve the use of urban agriculture as a method of supporting a sustainable food system. One such work is from Kubi Ackerman and his colleagues from Columbia University on a multi-year study done in New York City. They concluded that urban agriculture has the potential to mitigate access to nutritional food; its just a matter of policies to help promote this form of agriculture (2014). As previously stated, food insecurity comes in many forms; a study was done in India to find the correlation between agricultural performance and malnutrition indicators. Similarly, the methods of analysis done were a correlation analysis and a simple linear regression model to study the relationship between agricultural performance and malnutrition. Results from the study show that there was a significant negative relationship between agricultural performance and 'undernutrition' (Gulati et al., 2012). Based on these works, which were similar enough to mine, I surmised the analysis of the relationship between my indicators to produce similar results. However, my work was being done on the entirety of the United States; so the results may not necessarily be the same.

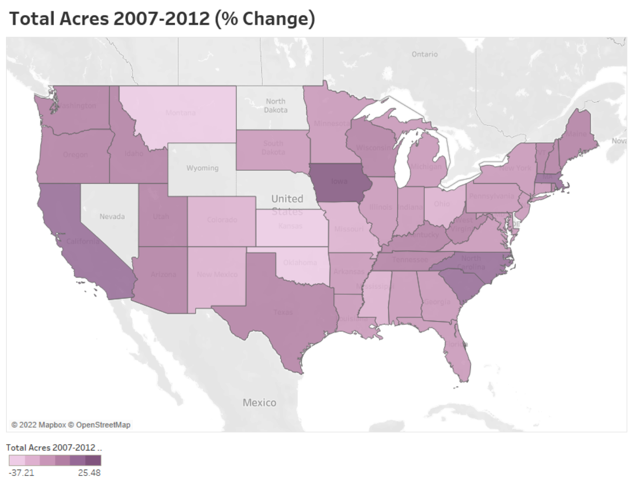

Fig. 1. Choropleth map of total farm acres in each state.

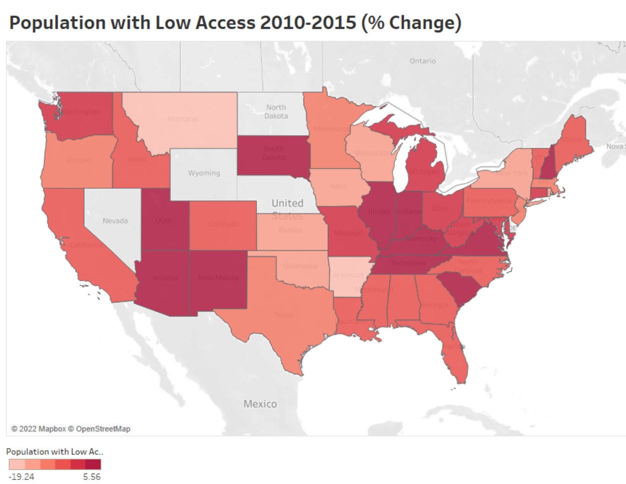

Fig. 2. Choropleth map of population with low access to stores in each state.

The figures above were produced in Tableau using a cleaned version of the data set explained in the following section. The darker shades reflect a higher percent change while the lighter shades reflect a lower percent change. Looking at some of the states in both maps, primarily across the Midwestern states, I notice some of them having a darker shade in one map and a lighter shade in the other. For example, we can see that Iowa has had an overall increase in total farm acres between 2007 and 2012; while having a lower amount of the population with access to stores between 2010 and 2015. Some states are also not colored in, this is due to the lack of data and outliers present in the final cleaned version of the data set.

III. Data

The primary data set I am using for research is from the Economic Research Service, an agency of the US Department of Agriculture. The data set contains indicators on food choices, farms and community characteristics measured at the county level unless otherwise stated. The explanatory variables I am interested in are the percent changes in vegetable acres harvested between 2007-2012, orchard acres between 2007-2012, berry acres between 2007-2012 and the total acres, which is a total of the previous three acre types, between 2007-2012. The response variables are the percent change in population with low access to stores between 2010-2015 (which means the percent change in the number of people living more than 1 mile from a supermarket or large grocery store if in an urban area, or more than 10 miles from a supermarket or large grocery store if in a rural area), and the percent changes in grocery stores and fast food restaurants between 2011-2016. All of these variables are listed in Table 1, along with their corresponding years and type of data. These indicators were what I believed to represent agriculture and food accessibility the most. Most of the research done using the Food Environment Atlas is at a preliminary level. More work needs to be done to identify correlations between interactions. The Atlas is constantly updated to provide more statistics to stimulate more research and to be used as a spatial overview of each indicator. This is inline with the main objective of this research; to find the correlation and relationship between indicators provided by the Atlas and to make it easier to visualize these indicators using mapping technology provided by Tableau. In terms of cleaning the data, despite having most of what I needed, there was some work needed to be done on the raw data. The variables needed to be measured at the state level and the percent change in total acres needed to be calculated. Using Python and the Pandas library, I created a data frame of each observation in the raw data, which was approximately equal to the total amount of counties in the United States, and filtering out the rows I did not need. I then created multiple lists adding all the aggregate data based on each county's state; these variables were measured in counts. I then calculated the percent change of these variables based on their year. These values are then added to a second data frame, where non-existing values are removed. I exported the data frame to a comma separated value file to be put in the software used for analyzing the relationship between the variables.

Explanatory Variables	Response Variables
Vegetable acres harvested (% change), 2007 - 12	Population, low access to store (% change), 2010 - 15
Orchard acres (% change), 2007 - 12	Grocery stores (% change), 2011-16
Berry acres (% change), 2007 - 12	Fast-food restaurants (% change), 2011-16
Total acres (% change), 2007 - 12	

TABLE I
TABLE OF EXPLANATORY AND RESPONSE VARIABLES COLLECTED AND/OR CALCULATED USING THE LATEST VERSION (UPDATED 9/10/2020) OF THE FOOD ENVIRONMENT ATLAS PROVIDED BY THE UNITED STATES DEPARTMENT OF AGRICULTURE

Summary Statistics		Summary Statistics		Summary Statistics	
Mean	-4.144594	Mean	2.2787173	Mean	8.3523001
Std Dev	6.4182117	Std Dev	8.5280753	Std Dev	5.4172082
Std Err Mean	0.9787681	Std Err Mean	1.3005193	Std Err Mean	0.8261165
Upper 95% Mean	-2.16936	Upper 95% Mean	4.9032715	Upper 95% Mean	10.019471
Lower 95% Mean	-6.119828	Lower 95% Mean	-0.345837	Lower 95% Mean	6.6851295
N	43	N	43	N	43

Fig. 3. Data retrieved and cleaned from Food Environment Atlas by the United States Department of Agriculture. Descriptive statistics on percent change in population with low access to a grocery store between 2010 and 2015, percent change in number of grocery stores between 2011 and 2016 and the percent change in number of fast food restaurants between 2011 and 2016, respectively.

Summary Statistics		Summary Statistics		Summary Statistics		Summary Statistics	
Mean	-7.5772	Mean	-8.743119	Mean	18.48797	Mean	-7.666202
Std Dev	18.473463	Std Dev	18.224818	Std Dev	36.210107	Std Dev	11.395415
Std Err Mean	2.8171767	Std Err Mean	2.7792586	Std Err Mean	5.5219895	Std Err Mean	1.7377845
Upper 95% Mean	-1.891907	Upper 95% Mean	-3.134348	Upper 95% Mean	29.631796	Upper 95% Mean	-4.159211
Lower 95% Mean	-13.26249	Lower 95% Mean	-14.35189	Lower 95% Mean	7.3441441	Lower 95% Mean	-11.17319
N	43	N	43	N	43	N	43

Fig. 4. Data retrieved and cleaned from Food Environment Atlas by the United States Department of Agriculture. Descriptive statistics on percent change in number of vegetable acres harvested, orchard acres, berry acres, and total acres between 2007 to 2012, respectively.

IV. Analysis

The analysis portion of the research was done through a JMP Pro 15. JMP Pro is an analytic software designed for engineers and scientists. Some features of the software are modeling and prediction. For this project, we are focused on regression analysis and correlation analysis; both of these tests can be done in JMP Pro just by opening the data set through the software. The built-in methods to perform these types of analysis are ”Fit Y by X” and ”Multivariate and Correlations” evaluations under the Analyze tab. For this research, to find the correlation and relationship between the two sets of indicators, I have set the α significance level to 0.05. For the regression portion of the analysis, I tested the linear fit between percent change in total acres between 2007 and 2012 and all of the response variables. The main values I am observing for each linear fit is the R^2 value and the p-value. The R^2 value, or the coefficient of determination, measures the proportion of the variance for a dependent or response variable that is explained by the independent variable, which in this case is the percent change in total acres between 2007 and 2012. It also represents the direction of the relationship depending on whether it is a positive or negative value. The p-value is used to determine the statistical significance of the observed difference, where a lower p-value indicates a statistically significant difference. In this case, a p-value lower than 0.05 indicates that the alternative hypothesis that there is a significant relationship between the percent change in total acres and the corresponding response variable. The correlation portion of the analysis mainly involves getting the correlation matrix, a heat map and a pairwise table of the Pearson correlation coefficient between each interaction. The p-value is also observed for this analysis.

V. Results

A. Regression Analysis

Linear Fit

Grocery Stores 2011-2016 (% Change) = 2.2519716 - 0.0034888*Total Acres 2007-2012 (% Change)

Summary of Fit

RSquare	2.173e-5
RSquare Adj	-0.02437
Root Mean Square Error	8.631356
Mean of Response	2.278717
Observations (or Sum Wgts)	43

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.0664	0.0664	0.0009
Error	41	3054.5125	74.5003	Prob > F
C. Total	42	3054.5789		0.9763

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.2519716	1.592284	1.41	0.1648
Total Acres 2007-2012 (% Change)	-0.003489	0.116876	-0.03	0.9763

Fig. 5. Regression analysis results displaying the relationship between percent change in grocery stores between 2011 and 2016 and the percent change in total acres between 2007 and 2012.

Linear Fit

Population with Low Access 2010-2015 (% Change) = -3.463278 + 0.0888727*Total Acres 2007-2012 (% Change)

Summary of Fit

RSquare	0.024898
RSquare Adj	0.001115
Root Mean Square Error	6.414632
Mean of Response	-4.14459
Observations (or Sum Wgts)	43

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	43.0771	43.0771	1.0469
Error	41	1687.0474	41.1475	Prob > F
C. Total	42	1730.1245		0.3122

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3.463278	1.18335	-2.93	0.0056*
Total Acres 2007-2012 (% Change)	0.0888727	0.086859	1.02	0.3122

Fig. 6. Regression analysis results displaying the relationship between percent change in population with low access to stores between 2010 and 2015 and the percent change in total acres between 2007 and 2012.

Linear Fit

Fast Food Restaurants 2011-2016 (% Change) = 8.2341986 - 0.0154055*Total Acres 2007-2012 (% Change)

Summary of Fit

RSquare	0.00105
RSquare Adj	-0.02331
Root Mean Square Error	5.479994
Mean of Response	8.3523
Observations (or Sum Wgts)	43

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.2944	1.2944	0.0431
Error	41	1231.2437	30.0303	Prob > F
C. Total	42	1232.5381		0.8366

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.2341986	1.010931	8.15	<.0001*
Total Acres 2007-2012 (% Change)	-0.015405	0.074204	-0.21	0.8366

Fig. 7. Regression analysis results displaying the relationship between percent change in fast food restaurants between 2011 and 2016 and the percent change in total acres between 2007 and 2012.

Figures 5 through 7 are the results of the linear regression portion of the analysis. Taking a look at the R^2 value for each of the runs, you can see that there is practically no relationship between the variables at all. The extremely high p-values of 0.9763, 0.3122 and 0.8366 further signifies that the relationship between the percent change in total acres and the response variables are not statistically significant at all. Based on the evidence collected, I fail to reject the null hypothesis that there is a relationship between total farm acres and indicators of food accessibility.

B. Correlation Analysis

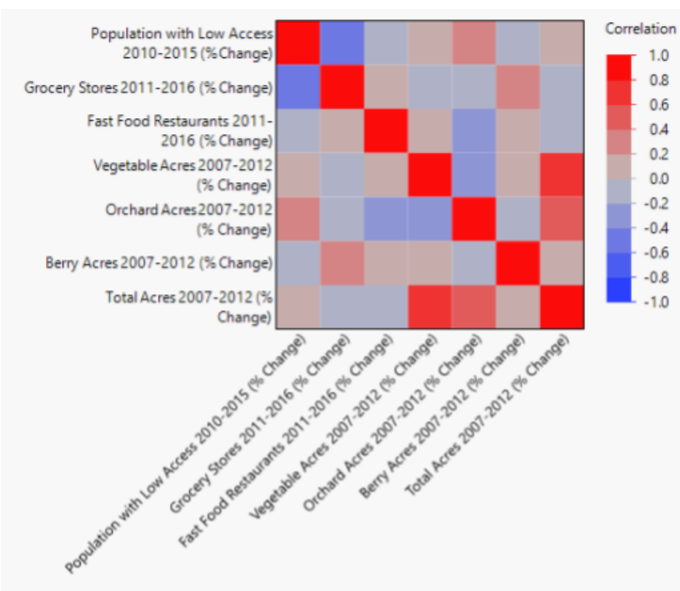

Fig. 8. Heat map of correlation between each variable. The redder the square, the higher the correlation between the two variables. The bluer the square, the lesser the correlation.

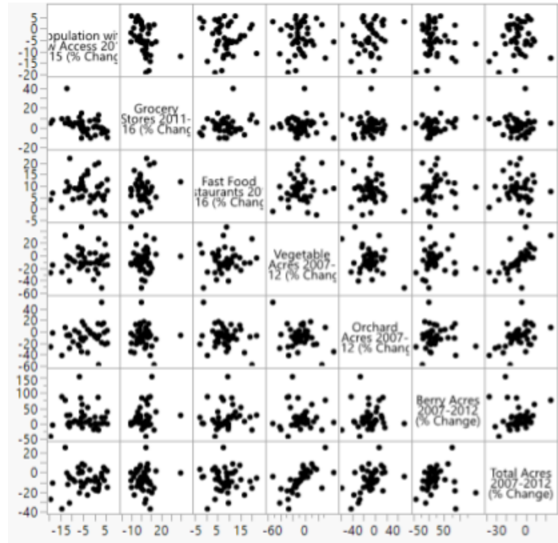

Fig. 9. Correlation matrix with multiple plots as interaction between variables. Each dot represents a state with the x-axis and y-axis being an indicator.

Pairwise Correlations

Variable	by Variable	Correlation	Signif Prob
Vegetable Acres 2007-2012 (% Change)	Population with Low Access 2010-2015 (% Change)	0.0236	0.8804
Vegetable Acres 2007-2012 (% Change)	Grocery Stores 2011-2016 (% Change)	-0.0074	0.9623
Vegetable Acres 2007-2012 (% Change)	Fast Food Restaurants 2011-2016 (% Change)	0.1309	0.4029
Orchard Acres 2007-2012 (% Change)	Population with Low Access 2010-2015 (% Change)	0.2140	0.1681
Orchard Acres 2007-2012 (% Change)	Grocery Stores 2011-2016 (% Change)	-0.0621	0.6922
Orchard Acres 2007-2012 (% Change)	Fast Food Restaurants 2011-2016 (% Change)	-0.2301	0.1377
Berry Acres 2007-2012 (% Change)	Population with Low Access 2010-2015 (% Change)	-0.0845	0.5902
Berry Acres 2007-2012 (% Change)	Grocery Stores 2011-2016 (% Change)	0.2226	0.1513
Berry Acres 2007-2012 (% Change)	Fast Food Restaurants 2011-2016 (% Change)	0.1084	0.4891
Total Acres 2007-2012 (% Change)	Population with Low Access 2010-2015 (% Change)	0.1578	0.3122
Total Acres 2007-2012 (% Change)	Grocery Stores 2011-2016 (% Change)	-0.0047	0.9763
Total Acres 2007-2012 (% Change)	Fast Food Restaurants 2011-2016 (% Change)	-0.0324	0.8366

Fig. 10. Pairwise table containing the Pearson correlation coefficient and p-value for each interaction between an indicator of agriculture and food access

Figures 8 through 10 are the results of the correlation analysis portion of the research. Looking at the heat map, there does not seem to be any strong correlation between indicators of agriculture and food accessibility. Based on the Pearson correlation coefficient provided by the Pairwise Correlations table in figure 10, which is quite low for each interaction, and the high p-value, there is no significant evidence to reject the null hypothesis that there is no correlation between indicators of agriculture and food accessibility.

VI. Conclusion & Discussion

Through my findings, I can conclude that there is no significant relationship or correlation between agriculture and access to food in the United States. After reviewing similar literature, getting the descriptive statistics of each indicator and visualizing the data on a map of the United States, I can say that these results were somewhat surprising to me.

This, however, gives a lot of room for future research, such as observing other indicators like state GDP or demographics. Other options to explore, given more data is provided, other elements of the agricultural sector can be researched such as a farm's income or types of animals owned. Broadening the scope of the project by observing another country's state of agriculture and food security could also be beneficial. My original idea was keep the data at a county-level, however, I believed that state-level data would be easier and more beneficial to interpret and analyze.

Even though the research shows that agriculture does not have an affect on the state of our food system, the growing issue of food insecurity persists. This also does not mean that agriculture definitely does not have an affect on the United States' food system; despite the results, the development of our agriculture will only benefit the quality of food provided to us. The goal of this paper was to see if the state of our agriculture affected people having access to their daily necessities. In this case, no significant relationship or correlation was found. However, there are many more indicators, in the Atlas and data provided by other parties, to be researched. Any work with significant breakthroughs should bring the issue on a political level, so that policies can be made to help people in need.

ACKNOWLEDGMENTS

The dates between each variable may not match exactly. For example, the percent change in population with low access to grocery stores is between 2010 and 2015, while percent change in vegetable acres harvested is between 2007 and 2012. This presents the possibility of data spikes not presented in the raw data set. The lack of being able to get data using the years I wanted was the greatest limitation for this research. However, in some sense, it may be beneficial to use these years as it may show the impact of the state of agriculture better.

Nebraska, North Dakota, Hawaii, Nevada, Wyoming, Alaska, Washington DC, Delaware were not included due to insufficient data or outliers. Also, there were no variables calculating vegetable acres, only vegetable acres harvested. So the total acres adds acres harvested with just acres. This means that the aggregate variable may not be its best real-life representation. Despite these shortcomings, I believe the research still has merit.

REFERENCES

[1] Farming and Farm Income. United States Department of Agriculture Economic Research Service. (2022, February 4). Retrieved April 14, 2022, from https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/farming-and-farm-income

[2] Key Statistics & Graphics. United States Department of Agriculture Economic Research Service. (2021, September 8). Retrieved April 14, 2022, from https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-u-s/key-statistics-graphics/

[3] Food security and why it matters. Australian International Food Security Centre. (2014, December 22). Retrieved March 3, 2022, from https://aifsc.aciar.gov.au/food-security-and-why-it-matters.html

[4] Ackerman, K., Conard, M.J., Culligan, P.J., Plunz, R., Sutto, M.P., Whittinghill, L.J. (2014). Sustainable Food Systems for Future Cities: The Potential of Urban Agriculture*. Economic and Social Review, 45, 189-206.

[5] Gulati, A., Kumar, A. G., Shreedhar, G., Nandakumar, T. (2012). Agriculture and Malnutrition in India. Food and Nutrition Bulletin, 33(1), 74–86. https://doi.org/10.1177/156482651203300108

[6] U.S. Department of Agriculture. (2020). Food Environment Atlas. Retrieved from https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/