Christopher De La Rosa
MARS 5470
Dr. Cheryl Harrison

State Population as an Indicator of Economic Health for Organic Agriculture

Abstract

Organic agriculture has seen a boom in popularity over the last few decades, with a global cultural shift towards a more sustainability and health-oriented society. Organic agriculture, however, is an expensive endeavour that necessitates great sales figures to economically viable. There have been many noted benefits of organic agriculture, and market trends and sales figures have shown the industry to be seeing great success. Analysis of per state sales metrics, within the United States, can help shed light on the health of organic markets within each individual state. Sales data was taken from the United States Department of Agriculture database for 2019 to analyze the relationships between Population estimates, Total Dollar Value in sales, and Total Dollar Value in sales per capita to establish if any represents a good predictor of health. It has been found that two relationships are statistically significant following regression testing: Population and Total Dollar Value in sales, and Total Dollar Value in sales and Total Dollar Value in sales per capita.

Introduction and Background

Organic farmers face substantial financial hurdles on the path towards and the maintenance of organic certification. It is a time-consuming and expensive endeavour, but one that has rich tangible and intangible rewards. Ostapenko et al. (2020) found that organic agriculture generally has higher per hectare output, for example. Organic agricultural practices are highly specific, and materials are generally more expensive for these practices, such as organic feed for livestock and organic fertilizers and/or compost can be seen as some examples. Further, it can be more labor-intensive due to the restrictions for such aspects such as weed management and soil rejuvenation. The highly restrictive nature of organic farming means sales are extremely important for this industry, and it is worth analyzing for aspiring organic farmers.

Over the past few decades, there has been a massive cultural shift in favor of organic produce. Sales figures across the country saw early amounts of tremendous year over year growth that has led to a healthy and economically sustainable industry (Lotter, 2003). The Research Institute of Organic Agriculture in 2019 provided a global picture of all facets of organic agriculture that has shown tremendous growth and sustainability since then, as well. Throughout the United States, however, it is worth noting that there is limited analysis of the per state health of organic markets. The United States Department of Agriculture has provided census data in certain years for organic agriculture, and it is through analyzing the sales data they have produced that a portrait of the health of organic sales in each state shall be derived.

Question and Hypotheses

With the data at hand, can population and/or total dollar sales be accurate predictors of total per capita dollar sales of organic agricultural products? Is there a statistically significant correlation between population and total dollar sales, population and total per capita dollar sales, and total dollar sales and total per capita dollar sales?

H0: The null hypothesis is simply that there is no statistical significance to the relationships, regardless of correlation.
H1: The first hypothesis is that at least one of the relationships analyzed will be found to have been statistically significant.
H2: The second hypothesis builds on the first that two of the three relationships analyzed will be statistically significant.
H3: The third and final hypothesis is that all three of the relationships analyzed will be statistically significant.

H2 is the favored and expected hypothesis.

Methodology

The initial dataset was derived from the 2019 Organic Survey that was derived from the 2017 Census of Agriculture Special Study conducted by the National Agricultural Statistics Service (NASS) under the United States Department of Agriculture. Specifically, the data was taken from Table 26: Marketing Practices on Certified Organic Farms: 2019. The data was downloaded in a zip file that contained csv files of each data table, and only the selected table was extracted. Two columns were added to the csv file, one that summed up the Total Value for each state based on the three dollar value columns, and another one that added the estimated population for each state. The population estimates were taken from the US Census QuickFacts section of their website, with the last date of estimation being July 1, 2019.

Once the csv file was cleaned up in excel to clear out any cells that were filled with anything non-numerical, the data was uploaded as a pandas DataFrame simply titled 'data' into JupyterLab. Afterwards, the data was modified so that all empty 'NaN' entries were filled with 0s, and the index was set to the 'Geographic area,' which was simply the names of each state. Next, the total dollar value per capita was calculated through using a for loop that ran through the indices and stored as a list. Per capita numbers for each of the individual dollar value columns were also calculated, but these quickly proved non-essential. Each of these lists were then added to the DataFrame.

A stacked bar chart with each of the three subsets of value labeled distinctively was created to highlight and visualize the immense disparities between certain states. This bar chart can be found in the Appendix as Figure A. Afterwards, each of the three relationships to be analyzed were then plotted out in scatterplots through the matplotlib package. A trendline for each was calculated as well through the numpy packages poly functions. These trendlines were then plotted into each respective graph. The graphs can be found in the Appendix as Figures B, C, and D, and are titled accordingly to each relationship.

Finally, the stats subpackage from the scipy package was used to run linear regression testing on each of the relationships. Through this function, the slope, intercept, r-value, p-value, and standard error were able to be derived for each relationship, and the trendline equation was then able to be derived.

Analysis of Findings

Initial viewing of the scatter plots and trend lines would indicate that none of the relationships have any statistical significance. The data points are exceptionally spaced out among each due to the extreme outliers that some of the more populous states represent, especially California. The trend lines, as well, are generally err closer to the flatter side, especially when analyzing the relationship between Population and Total Dollar Value per capita. Running the linear regression testing, however, helps to provide deeper insight into the relationships. For the following table, the relationships will be as follows: 1 - Population vs Total Dollar Value per capita, 2 - Population vs Total Value, and 3 - Total Value vs Total Dollar Value per capita. The values will be rounded to the nearest thousandth.

Relationship	Slope	Intercept	R-Value	P-Value	Std. Error
1	1.791e-7	6.534	0.069	0.636	3.756e-7
2	19.802	-69645256.36	0.630	9.574e-7	3.525
3	3.916e-8	5.355	0.472	0.001	1.055e-8

The statistical values derived here indicate at least two relationships that should be statistically significant. None of the relationships here have a strong correlation, with the strongest being relationship 2 and it still is only considered moderately strong. However, the p-values for both relationships 2 and 3 are <0.05, which indicates statistical significance. Through this analysis, it can be determined that H2 can be confirmed as at two of the three relationships indicate statistical significance.

Discussion

Organic produce has grown in popularity and there are more adopters of these practices than ever before. In the United States, however, this does not seem to be the case equally from state to state. It can be concluded that population can be a good indicator of an economically healthy organic market, however in a more transitive sense. Population can be a good predictor of total sales, which can then in turn be a decently good predictor of per capita total sales. This can be an important data point for aspiring organic farmers who hope to transition their current practices or ones who wish to enter the industry altogether.

References

Bialik, K., & Walker, K. (n.d.). Organic farming is on the rise in the U.S. Pew Research Center. Retrieved April 4, 2021, from https://www.pewresearch.org/fact-tank/2019/01/10/organic-farming-is-on-the-rise-in-the-u-s/

Donald W. Lotter. (2003). Organic Agriculture, Journal of Sustainable Agriculture, 21:4, 59-128, DOI: 10.1300/J064v21n04_06

National Agricultural Statistics Service. (2020). 2017 Census of Agriculture - 2019 Organic Survey. United States Department of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Organics/ORGANICS.pdf.

Ostapenko R, Herasymenko Y, Nitsenko V, Koliadenko S, Balezentis T, Streimikiene D. Analysis of Production and Sales of Organic Products in Ukrainian Agricultural Enterprises. *Sustainability*. 2020; 12(8):3416. https://doi.org/10.3390/su12083416

United States Census Bureau. (2019). QuickFacts Population Estimates, July 1, 2019 (V2019). https://www.census.gov/quickfacts/geo/chart/US/PST045219.

Willer, E. H., & Lernoud, J. (2019). The World of Organic Agriculture Statistics and Emerging Trends 2019. 351.

Appendix

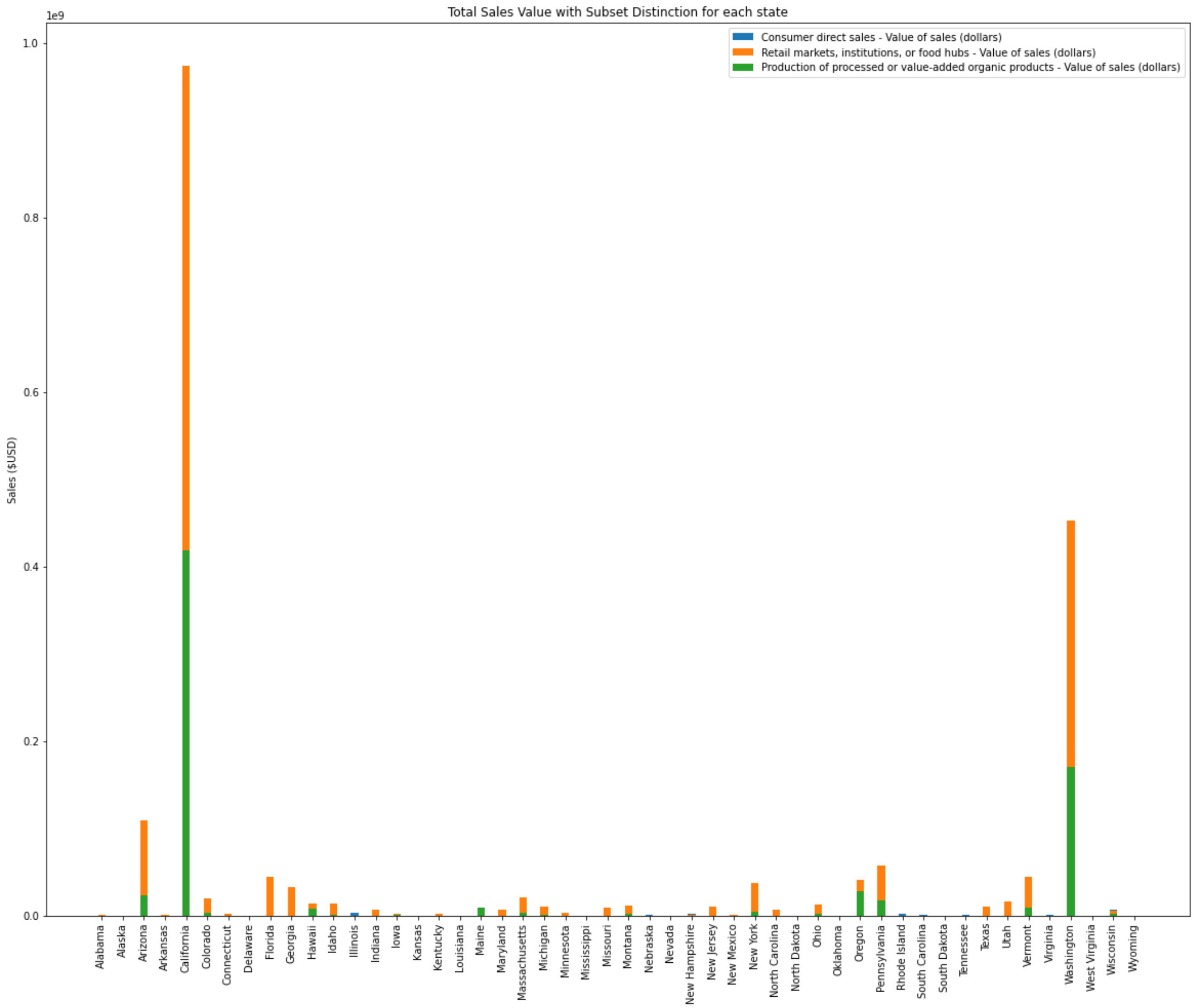


Figure A

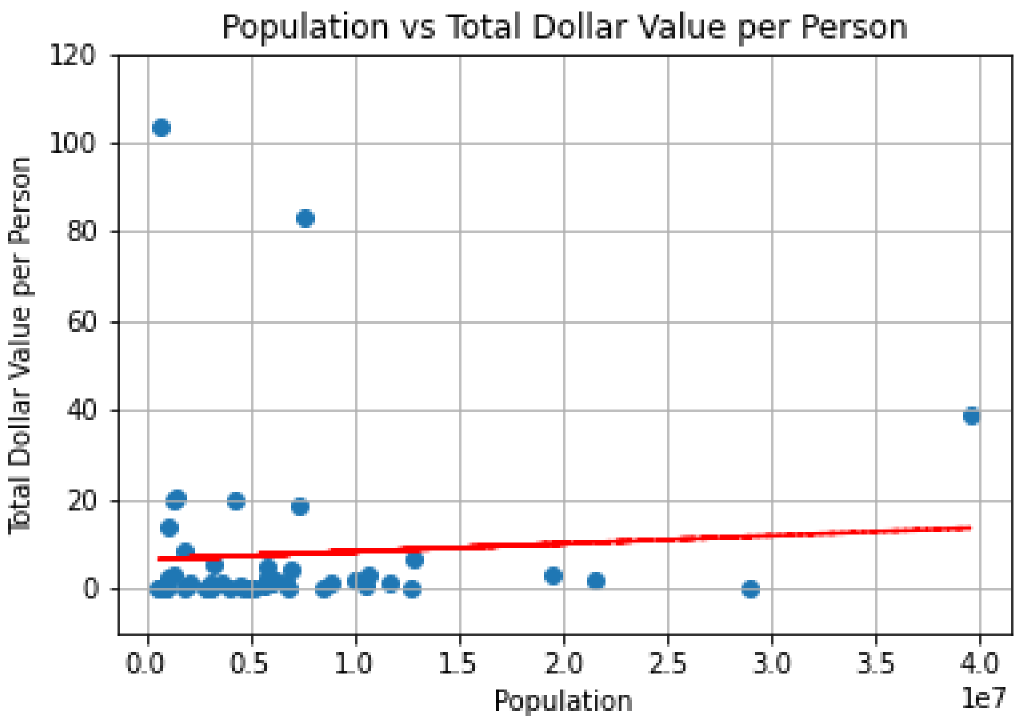


Figure B

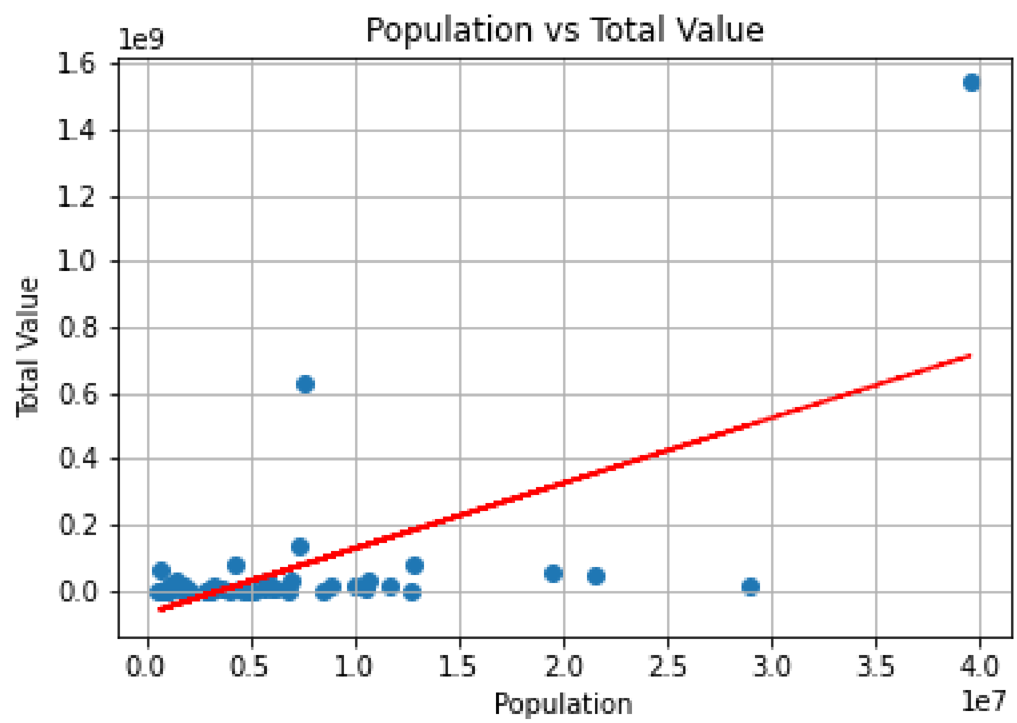
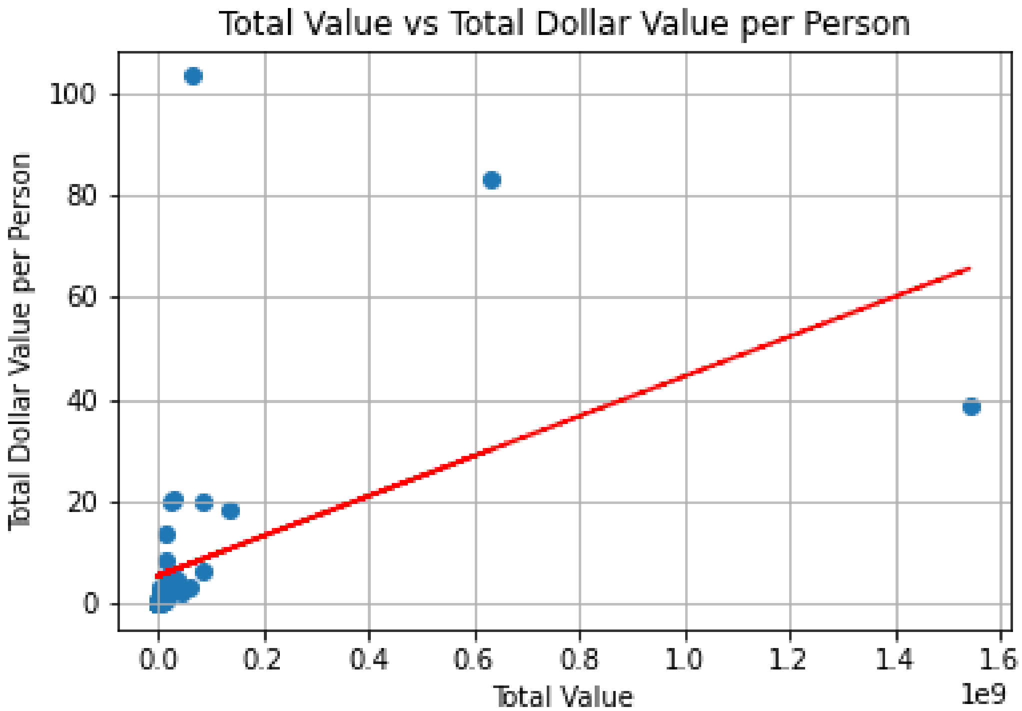


Figure C


Total Value vs Total Dollar Value per Person
Total Dollar Value per Person
Total Value
1e9
0
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Figure D

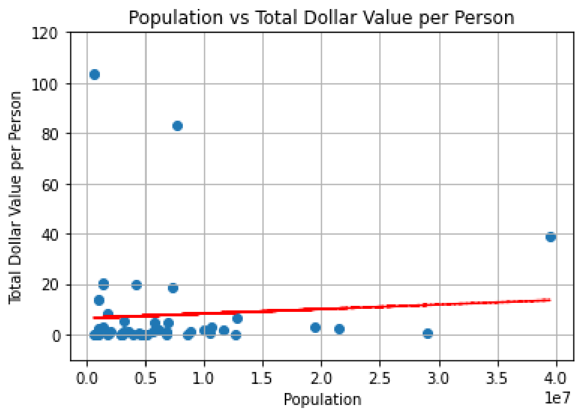
```
[19]: #next get percentages for each and then maybe do a linreg? main thing is visualization
```

```
[21]: z = np.polyfit(data.loc["Alabama":]["Population"], data.loc["Alabama":]["Total Dollar Value per Person"], 1)
p = np.poly1d(z)

fig,ax = plt.subplots()

plt.scatter(data.loc["Alabama":]["Population"], data.loc["Alabama":]["Total Dollar Value per Person"])
plt.xlabel("Population")
plt.ylabel("Total Dollar Value per Person")
plt.title("Population vs Total Dollar Value per Person")
plt.grid(True)
ax.set_ylim([-10,120])
plt.plot(data.loc["Alabama":]["Population"],p(data.loc["Alabama":]["Population"]),"r--")

plt.savefig("PopulationVSTotalDollarValueperPerson.png")
plt.show()
```



Sample Code 1

```
[26]: totalp_slope, totalp_intercept, totalp_r_value, totalp_p_value, totalp_std_err = stats.linregress(data.loc['Alabama':]["Population"],data.loc['Alabama':]["Total Dollar Value per Person"])

[27]: stats.linregress(data.loc['Alabama':]["Population"],data.loc['Alabama':]["Total Dollar Value per Person"])

[27]: LinregressResult(slope=1.7905883335147504e-07, intercept=6.534231104837948, rvalue=0.0686626721184530З, pvalue=0.6356438871936771, stderr=3.755158355942288e-07)

[28]: totalxtotal_slope, totalxtotal_intercept, totalxtotal_r_value, totalxtotal_p_value, totalxtotal_std_err = stats.linregress(data.loc["Alabama":]["Total Value"], data.loc["Alabama":]["Total Dollar Value per Person"])

[29]: stats.linregress(data.loc["Alabama":]["Total Value"], data.loc["Alabama":]["Total Dollar Value per Person"])

[29]: LinregressResult(slope=3.915898018429209e-08, intercept=5.35503936484205, rvalue=0.47213099707506306, pvalue=0.0005369884427943572, stderr=1.0553209234289893e-08)

[30]: total_slope, total_intercept, total_r_value, total_p_value, total_std_err = stats.linregress(data.loc["Alabama":]["Population"], data.loc["Alabama":]["Total Value"])

[31]: stats.linregress(data.loc["Alabama":]["Population"], data.loc["Alabama":]["Total Value"])

[31]: LinregressResult(slope=19.802136635385786, intercept=-69645256.35984263, rvalue=0.6298046482414749, pvalue=9.573773383410818e-07, stderr=3.525078796502857)
```

Sample Code 2