Data Story

Sara Maxwell

1. Introduction

Farmer's markets are a key factor in the health of a population due to greater proximity to quality food sources and greater trust from consumers because they can get to know their local farmers and trust that the food is produced to their standards.

With news stories of food recalls due to pathogens such as E. Coli O157:H7 in romaine lettuce (https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html) or foreign matter such as hard plastic in chicken breasts (https://www.fsis.usda.gov/wps/portal/fsis/topics/recalls-and-public-health-alerts/recall-case-archive/archive/2018/recall-026-2018-release), farmer's markets can be seen as a source of food produced closer to home, with less risk of contamination. The greater the variety of foods and payment methods available, the likelier the markets are to draw more clients and grow their presence in the locality.

This data is focusing on the numbers of markets per 1,000 people in 2009 and 2016, as well as some characteristics: * Types of foods sold (produce, animal, baked, other) * Methods of payment accepted (credit card, SNAP, WIC, WICCash, SFMNP)

2. The dataset

This dataset was accessed from the United States Department of Agriculture. [https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/]. These are the factors of interest:

- FMRKTPTH09 <- Farmer's markets per 1,000 people in each county in 2009
- FMRKTPTH16 <- Farmer's markets per 1,000 people in each county in 2016
- FMRKT_SNAP16 <- Farmer's markets in each county that accepted Supplemental Nutrition Assistance Program (SNAP) benefits in 2016.
- FMRKT_WIC16 <- Farmer's markets in each county that accepted Women, Infants, and Children (WIC) benefits in 2016.
- FMRKT_WICCASH16 <- Number of farmer's markets in each county that accepted WIC Cash Value Vouchers (WICCash) in 2016.

- FMRKT_SFMNP16 <- Number of farmer's markets in each county that accepted Senior Farmers' Market Nutrition Program (SFMNP) benefits in 2016.
- FMRKT_CREDIT16 <- Number of farmer's markets in each county that accepted credit cards in 2016.
- FMRKT_FRVEG16 <- Number of farmer's markets in each county that sold fruits and vegetables in 2016.
- FMRKT_ANMLPROD16 <- Number of farmer's markets in each county that sold animal products in 2016.
- FMRKT_BAKED16 <- Number of farmer's markets in each county that sold baked goods in 2016.
- FMRKT_OTHERFOOD16 <- Number of farmer's markets in each county that sold other food products in 2016.

3. Scope and limitation

This data set has some limitations. For example, the data does not show which types of markets increased or decreased between 2009 and 2016. Also, the data does not have variables for markets that sell nonfood items, have entertainment, or have food truck vendors. Finally, there is no data on dollar sales figures, so we cannot see how people spend at farmer's markets.

4. Data wrangling

- Columns of little value were removed. These columns included data that do not pertain to farmer's markets, columns showing the total number of farmer's markets in each county or county equivalent, the percent change from 2009 to 2016, and the percent of markets in each county of each type of market.
- In the process of formatting the data, the column names were converted to shorter, more descriptive names.
- Columns with decimals were made easier to read and distinguish by converting the numbers in percentage columns to numbers rounded to the nearest thousandth.
- To help with analysis, the following columns were added: *1.* Dummy variables for each kind of market. *2.* Columns showing numbers of each kind of market per 1,000 people. *3.* The difference between markets per 1,000 people in 2009 and 2016. *4.* Whether or not farmer's markets increased their presence or not. *5.* Normalized data for t-test analysis.

- In columns already present and in columns added, all blank cells were replaced with zeroes for consistency.

5. Exploration

When exploring the data, I decided I would compare the data from counties that saw increases in markets per 1,000 people with the overall data and see if the counties with increases offer a greater variety of food and payment options.

I calculated percentages of counties that sold all types of food and accepted all noncash methods of payment, both in counties that saw increases and in all counties. A greater proportion of the counties that saw an increase in farmer's markets had markets that sold all food types and accepted all noncash methods of payment.

To visualize the differences between the counties with increases and counties overall, I made histograms showing the numbers of counties with each type of market. I noticed that there were smaller numbers of counties with none of a certain type of market in counties that saw an increase in markets per 1,000 people.

After examining the histograms, I thought that having fewer counties with none of a certain type of market would bring up the mean of the type of market compared to the counties overall. I chose a two-sided t-test to compare the means, whether the mean of the counties that saw increases is identical to the mean of counties overall. The p-values for all types were over 0.05, so the mean number of each type of food and method payment was not significantly higher in counties that saw an increase in farmer's markets than in all counties.

Finally, linear model plots showed correlation of the different types of food and methods of payment between counties with increases and counties overall. Every food type and every method of payment showed a positive correlation with the number of markets per 1,000 people in 2016, though I did see weaker correlations with animal products and WICCash. Interestingly, there is a weaker correlation between markets that sell animal products and the overall number of markets per 1,000 people than there is between markets that sell baked goods and the overall number of markets per 1,000 people. Producing baked goods is easier than producing meat, dairy, or eggs.

The lines representing SNAP and WIC were close, which didn't surprise me as these are the most common government benefit programs used to purchase food. Interestingly, the senior benefit line showed a stronger correlation with markets per 1,000 people overall.