Analyzing Demographic Shifts and Their Economic Impact on U.S.Agriculture

HDSI Agri Datathon 2024
CodeCultivators
6th October 2024

Krishna Venkatesh
Northeastern University
360 Huntington Ave
Boston, MA 02115
hosahudyavenkatesh.s@northeastern.edu

Laasya Anantha Prasad
Northeastern University
360 Huntington Ave
Boston, MA 02115
ananthaprasad.l@northeastern.edu

Rakshak Kunchum
Northeastern University
360 Huntington Ave
Boston, MA 02115
kumchum.r@northeastern.edu

Abstract

This project investigates the relationship between farmer age demographics and agricultural production across the contiguous United States at the state level, using data from the National Agricultural Statistics Service (NASS) from 1997 to 2022. Given the challenges posed by an aging farming population, understanding the impact on agricultural productivity is essential. The study focuses on three core areas: Spatiotemporal analysis to examine changes in farmer age over time; regional economic impact to assess how aging trends affect larger, economically significant farms; and the application of inflation-adjusted crop sales data to uncover any persistent trends. By integrating these data sources and using statistical modeling, the findings will highlight regions most at risk of agricultural decline and provide insights into sustaining the agricultural economy.

1. Introduction

In the evolving landscape of U.S. agriculture, demographic shifts in the farming population present significant challenges and opportunities for productivity and economic sustainability. This research examines the relationship between farmer age demographics and agricultural production across various states, using extensive data from the National Agricultural Statistics Service (NASS) spanning 1997 to 2022. With the average farmer age now at 58.1 years—a trend that has persisted over recent decades—concerns about diminished innovation and productivity have grown, posing potential risks to national food security and the economic health of rural communities that depend on agriculture.

This project seeks to clarify the economic implications of an aging agricultural workforce by exploring the connection between farmer age and the scale of agricultural output, particularly focusing on large-scale, economically significant farming operations managed predominantly by older individuals. Additionally, it investigates the evolution of farmer age across states and its correlation with farm scale using SARIMA forecasting and visualizations. To deepen insights, this analysis includes interactive choropleth maps and Plotly plots, examining pre- and post-inflation production trends to reveal how real economic value relates to farmer age. These findings aim to highlight critical regions and commodities affected by demographic shifts, informing strategic interventions to sustain productivity in U.S. agriculture.

2. Data and Methods

The analysis utilized four datasets:
1. Sales Data: Included agricultural sales data across various products and counties, along with the number of operations. Missing or placeholder values were noted for certain regions and products.
2. Land Total Demos Data: Focused on land use, ownership, and age-related demographics of producers and

Figure 1: Average age of farmers SARIMA forecasting per state

operators across counties, with some missing values.
3. State-Level Data: Contained state-level demographic information of farm operators across sales categories.
4. Inflation GDP Price Index Data: Provided inflation-adjusted indices from 1990 to 2022, used for adjusting state-level sales data.

Sales Data Preprocessing: To handle missing or disclosed values in the sales data, specific strategies were employed. Values marked as (D), indicating disclosure limitations, were imputed with the respective state and year mean. This method preserved the insights at both state and year levels, thus preventing the loss of critical data that might otherwise have occurred if these values were excluded. This approach ensured that key trends were retained while minimizing the loss of information. Similarly, values marked as (Z), which represent negligible percentages, were replaced with 0. This technique ensured data completeness while avoiding the introduction of skewness by not imputing values with a mean when actual percentages were nearly zero. Empty strings were also replaced with the corresponding State FIPS code, and missing values were imputed using the state mean. This helped maintain the data integrity, ensuring consistency throughout. It is important to note that ”999” is not a county code but represents the total of all county data (including disclosure values) for a given year of the census. This ”999” state-level data is crucial to the analysis since it focuses on state-wise trends rather than county-specific insights.

Land Total County Demos Data Preprocessing: Feature selection was conducted by retaining age-related data while addressing inconsistencies between producers and operators. Age groups for both producers and operators were standardized by creating unified categories. Missing values were replaced with 0, ensuring no data points were lost during this step. To maintain consistency across age groups, similar age categories were merged, providing a coherent dataset for both producers and operators. For the 2017 and 2022 data, only the primary producer was considered as the operator, ensuring consistency across the ”Operator” category in all years. This approach addresses discrepancies in data classification and ensures reliable and comparable analysis. Redundant producer data was removed to streamline the dataset.

State-Level Data Preprocessing: Missing state names were addressed by imputing them using an ANSI-to-state mapping. This ensured that any missing state names in the GEO column were accurately replaced, maintaining data consistency. To further refine the dataset, special characters such as spaces and dashes were replaced, and missing values were imputed using K-means clustering. K-means was particularly useful in identifying underlying patterns in the data to predict missing values, as it operates in a multi-dimensional space, imputing missing data based on feature similarities.

Merging Datasets: An inner join was performed between the sales county data and the land totals data. This was followed by a second merge with the state-level data, ensuring a comprehensive and unified dataset for analysis.

Inflation Adjustment: The GDP price index was applied to adjust crop sales data for inflation, using 2022 as the base year. The formula used to calculate real GDP from nominal GDP was as follows:

$$\text{Real GDP} = \left(\frac{\text{Nominal GDP} \times 100}{\text{GDP Deflator}}\right)$$

Where nominal GDP represents the market value of all goods and services produced, and the GDP Deflator reflects price changes (inflation or deflation). The base year GDP price index was set at 117.996 for 2022, ensuring all economic analysis reflected inflation-adjusted values.

A statistical choropleth map provides insights into farmer demographics across the contiguous U.S. Green shades in Midwest states like Iowa and Nebraska indicate younger farmers, possibly due to education initiatives, while lighter in California and the Southwest shows older demographics, posing succession challenges. Darker shades in the Northeast, such as New York and Pennsylvania, suggest higher age averages, with barriers to younger entrants. Light beige in states like Florida and Texas further highlights the aging trend.

From the early 2000s to 2010, nominal and inflation-adjusted sales grew steadily with minimal inflation impact. Post-2010, nominal sales rose sharply, while inflation-adjusted sales showed moderate growth, illustrating inflation's growing impact on agricultural sales and the need for adjustment to assess long-term performance accurately.

The 1997 dataset for age-by-sales groups does not adhere to the same methodology as data from other years, resulting in inconsistencies and missing values. To mitigate this issue, adjustments were made to the 1997 values by applying the difference observed between the 2002 and 2007 data for corresponding columns. This approach provides continuity and maintains dataset consistency, ensuring that the 1997 data aligns more accurately with subsequent years and supports reliable comparative analysis across the dataset.

Method: This project began by analyzing spatiotemporal trends in the average age of farmers across the United States, covering counties, states, and the national level from 1997 to 2022. A heatmap was employed to visually represent the changes over time, highlighting regional variations in farmer demographics. The heatmap helped identify areas where the average farmer age has significantly increased over time. This tool utilized a color gradient, where lighter shades represented older farmers and darker shades indicated younger farmers. The methodology focused on identifying trends at the state level, comparing state averages with the national trend to understand how the aging of farmers differed across regions.

Following the heatmap analysis, regional variations and the economic impact of an aging farming population were examined. A Seasonal Autoregressive Integrated Moving Average (SARIMA) forecasting model was applied to predict future trends in the average farmer age by state. This model considered historical data and regional patterns to forecast changes in farmer age across different regions, with a focus on identifying how aging farmers may impact larger, economically significant farms compared to smaller farms operated by younger farmers. The analysis also compared age trends across states with differing farm sizes to investigate correlations between farmer age and economic scale.

The third phase of the analysis involved examining inflation-adjusted sales data. This was done to determine whether inflation-adjusted trends would reveal different relationships between farmer age and agricultural production. The GDP price index was used to adjust for inflation, focusing on how changes in real terms (inflation-adjusted sales) compared to nominal sales affected various commodities. By correlating farmer age with production and sales, the study aimed to uncover whether older farmers were less involved in high-production commodities and whether their operations scaled differently as they aged.

Table 1: States with a significant percentage of younger farmers

Year	**Rhode Island (44)**	**Alaska (2)**	**Maine (23)**
1997	52.3	54.2	51.0
2002	54.3	55.2	53.7
2007	56.3	56.2	56.4
2012	56.7	57.1	57.0
2017	56.9	55.2	56.5
2022	56.5	56.7	57.5

Table 2: States with significant percentage of older farmers

Year	**Mississippi (28)**	**Florida (12)**	**California (06)**
1997	55.8	55.6	55.2
2002	57.2	57	58.8
2007	58.6	58.4	58.4
2012	60.4	59.8	60.1
2017	58.9	58.9	59.2
2022	59.6	59.5	59.9

3. Results

The spatiotemporal analysis revealed a consistent nationwide trend of increasing farmer age. In 2002, Mississippi recorded the highest average farmer age at 57.2 years, while Minnesota had the lowest at 52.9 years. By 2022, Hawaii had the highest average age of 60.7 years, while Pennsylvania had the lowest at 55.8 years. An interactive plot comparing the average farmer age by state and at the national level

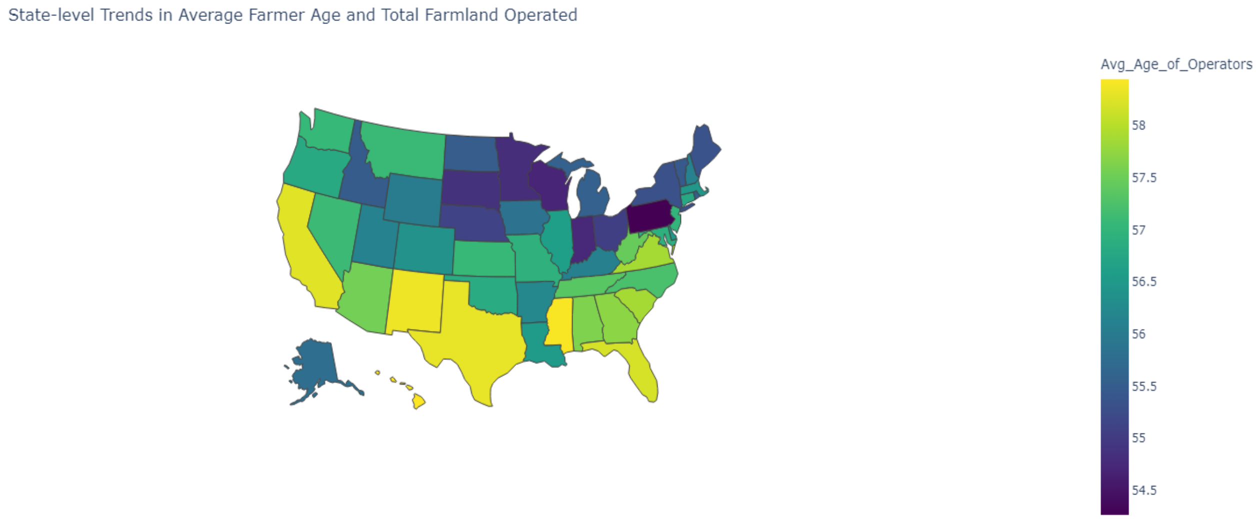

Figure 2: Choropleth map depicting the farmer demographics for each state across contiguous United States

confirmed a steady upward trajectory across most states, indicating a clear aging trend in the farming population.

The analysis of regional variations highlighted significant differences in farmer demographics across states. Rhode Island, Alaska, and Maine were identified as having a higher percentage of younger farmers. In contrast, states like Mississippi, Florida, and California maintained a consistently lower-than-average farmer age. Arizona showed a particularly high average farmer age in 2022, reaching 60.1 years, while Delaware experienced a marked increase in the average farmer age over time.

One notable finding was the peak in average age around 2012, where many states displayed the darkest shades on the heatmap, indicating a significant increase in age. After 2012, some states saw a slight drop or stabilization in the average age of farmers. This trend suggests that while the aging trend is ongoing, there may be regional efforts or factors contributing to a slowdown or stabilization in certain areas.

The analysis of the economic impact of aging farmers through the SARIMA forecasting model showed that larger farms are typically managed by older farmers, who have accumulated significant resources and land over time. These larger farms are often more economically significant, reinforcing the correlation between farm size and the operator's age. Conversely, younger farmers were more likely to operate smaller or niche farms due to limited capital and resources. This suggests a divide in the farming landscape, where younger generations may be entering the industry through smaller, less resource-intensive operations, while larger, more capital-intensive farms remain under the control of older farmers.

Finally, the analysis of inflation-adjusted sales revealed important trends in the relationship between farmer age and agricultural production. Several key commodities, such as poultry and wheat, showed negative correlations with farmer age after adjusting for inflation, suggesting that as farmers age, their involvement in high-production activities or the scalability of their operations declines. The correlation analysis between nominal and inflation-adjusted sales further indicated that commodities like rice, cotton, and corn also exhibited negative correlations with farmer age. This highlights that as the farming population ages, production or sales of these commodities tend to decrease, possibly due to a reduced focus on large-scale operations among older farmers.

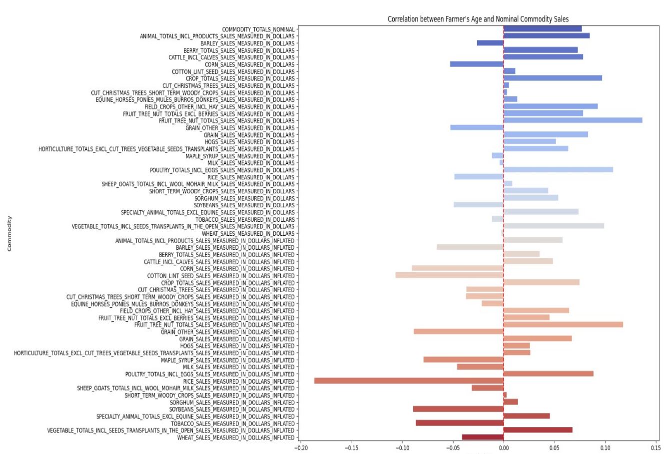

Figure 3: Correlation between farmers age and sales of commodities before and after inflation after inflation adjustment.

4. Conclusion

The project highlights the upward trend in the average age of farmers across the United States, reflecting a steady generational shift within the agricultural workforce. The data reveals notable regional variations, with northeastern states like Rhode Island, Alaska, and Maine having a higher proportion of younger farmers, while southern and southwestern states, particularly Arizona, Florida, and California, consistently show older farming populations. States such as Texas and Montana, which boast larger agricultural lands, exemplify the correlation between farm size and operator age, as larger farms are more often managed by older farmers who have accumulated resources and experience.

Our findings also underscore a significant peak in the average age of farmers around 2012, followed by a stabilization, possibly due to advancements in agricultural technology that help older farmers maintain productivity with reduced physical strain. This trend is particularly visible in large-scale farming regions where high-tech methods support sustained activity among older farmers. Economic scale further impacts age demographics, as larger farms are generally operated by older farmers, while younger farmers are more commonly involved in smaller or niche farms due to resource constraints. The steady rise in the average age of farmers reflects a lower rate of young entrants relative to retirements among older farmers. SARIMA forecasting confirms this upward trend, projecting further increases in 2025 and 2028. This supports the hypothesis that larger farms, indicated by total sales and average acres of harvested land, are more commonly managed by older operators.

The analysis of the correlation between farmer age and commodity sales, both nominal and inflation-adjusted, sheds light on the influence of aging on production. Inflation-adjusted correlations reveal that commodities like rice, corn, and cotton tend to show decreased production or sales as farmer age increases. This suggests that older farmers may be less engaged in high-output crops or face challenges in scaling production as effectively with age. These insights provide valuable guidance for policy and investment in agriculture, emphasizing the need for strategic support to address succession challenges and encourage younger generations to enter the profession.

Please see this link to our submission video.
Please see this link to our Github Repository.
Please see this link to our Google Colab Notebooks:

1. Data preprocessing and merging,
2. Exploratory Data Analysis,
3. Trend Analysis and modeling

References

[1] Phillip Warsaw, Steven Archambault, Arden He, Stacy Miller, *The Economic, Social, and Environmental Impacts of Farmers Markets: Recent Evidence from the US*, Sustainability, vol. 13, no. 6, pp. 3423, 2021, https://www.mdpi.com/2071-1050/13/6/3423

[2] Sylvia Szabo, Chidchanok Apipoonyanon, Malay Pramanik, Takuji W. Tsusaka, Kevin Leeson, *Agricultural Productivity, Aging Farming Workforce, Sustainable Agriculture, and Well-Being: Household Survey Data From Central Thailand*, Frontiers in Sustainable Food Systems, vol. 5, 2021, https://www.frontiersin.org/articles/10.3389/fsufs.2021.728120/full

[3] Áron József Borda, Balázs Sárvári, Jeremiás Máté Balogh, *Generation Change in Agriculture: A Systematic Review of the Literature*, Economies, vol. 11, no. 5, pp. 129, 2023, https://www.mdpi.com/2227-7099/11/5/129