Developing a User-Friendly Shiny Application for Visualizing and Comparing United States Crop Data Procured by NASS

December 9, 2024

Abstract:
Agricultural trends and insights are critical for stakeholders to make informed decisions, yet the complexity of raw data often limits its accessibility. The United States Department of Agriculture (USDA) conducts surveys and censuses to collect data on crops and other relevant sectors. The National Agricultural Statistics Service (NASS) compiles such data for public release. This project uses crop data from 2000 to 2024 to visualize agricultural information to benefit users without advanced statistical expertise. Data from NASS were processed and visualized using RStudio, Shiny App, and associated packages. The interactive application presents geospatial and temporal trends for four Midwest crops: corn, soybeans, wheat, and potatoes. The resulting tool enables users to analyze crop data at both state and county levels, offering insights into production, yield, and harvested acreage. NASS has commended the resulting app, highlighting its potential to influence agricultural planning and decision-making, including adjustments to acreage planted and the possibility of increased yields.

Introduction:

The United States Department of Agriculture is a federal department that oversees United States agriculture and related operations. The USDA is composed of twenty-nine various agencies, one being the National Agricultural Statistics Service (NASS). NASS is responsible for coordinating surveys annually and the Census of Agriculture once every five years. These provide agriculture info that can be utilized when analyzing agricultural trends. While some information is compiled into reports that are made available to the public, much of the data remains in an online database without pre-defined tools for analysis. Data obtained from both the surveys and census, excluding data that cannot be disclosed for security reasons, is available on the QuickStats database within the NASS website. While this data is easily accessible, it is not necessarily as easily understood by the general public. Extraction of data from the QuickStats database is complex. Only 50,000 observations can be queried at once, accounting for only a fraction of available data. Additionally, due to the organization of the filters, data about specific commodities may not be organized as expected; if one is not familiar with the database, it is possible that relevant data may not be successfully queried for a specific area of study in one download. Some visualizations are available on the NASS website, but few allow for customizable visualizations or easily depict trends over time. NASS provides relevant and important data regarding agriculture, but data users must sift through complex and convoluted data to understand and interpret agricultural trends.

In an effort to alleviate the data complexity issue, this project aimed to generate an output portal that would display various insightful visualizations and an interactive map for data users looking to analyze specific agricultural trends over a 24-year period. To provide customizable information, both county and state-level comparisons are available. Data representing four significant crops found in the Midwest, corn, soybeans, wheat, and potatoes, were displayed within the portal. Density of variables for each crop can be displayed geospatially via an interactive map, as well as compared to each other over time using line plot (timeline) visualizations. These tools allow for the transformation from elaborate tabular data to helpful visualizations of agricultural trends geographically and over time.

Methods:

The first steps in this project involved data collection and gathering ideas for useful visualizations. While the final product scope involved only crop data, initial experimentation involved data from the following sectors: animals and products, crops, demographics, economics, and environmental data. There are three ways to collect data from the QuickStats database provided by NASS. The first method is to download entire datasets in one large, compressed file. This is the way data was collected for the final product. The second method is to pull data in queries of up to 50,000 observations at once from the QuickStats user interface. The third option is to utilize API calls that download up to 50,000 observations in the data through a script, using an API key provided by NASS.

In the final rendition of the visualization portal, data is collected from QuickStats in a large download of all crop data provided. While experimenting with the data, it was determined which values and features were valuable to interpret. This involved variables such as production, yield, acres harvested, and acres planted. Additionally, crops other than corn, soybeans, wheat, and potatoes were not included in the project scope. Using RStudio, after many attempts to filter, concatenate, and calculate data values, final datasets were developed for county and state-level data. These datasets were stored in a GitHub repository for online access and the ability to load data into the app online without the need to do calculations and create as many datasets within the app. At the point that data creation was complete, visualization could be produced.

Following the creation of project data, the application was developed using Shiny, a framework for building interactive web applications in R. The app's user interface and server components were designed to dynamically process and display data based on user selections. Key features include:

1. Summary Tab: Provides plots summarizing crop data at a high level, including yearly trends.
2. Interactive Map Tab: Displays geospatial data for states and counties, allowing users to customize their view by selecting variables and filters.
3. Comparison Tab: Enables temporal and statistical comparisons between selected states or counties, offering insights into regional agricultural performance.

These features transform raw tabular data into visual insights, facilitating user engagement and understanding.

To further enhance user accessibility, the app includes a walkthrough YouTube video explaining its features and functionality. Additionally, simple explanations accompany certain visualizations to help users with limited data analysis experience understand the insights being presented. The app is also deployed on an AWS server using RStudio Connect, allowing users to access it directly through a web browser using the app's URL.

Results:

The application's outputs include:

1. Temporal Trends: Line graphs depicting annual changes in yield, production, and harvested acres, helping users identify key patterns and anomalies over time.
2. Geospatial Visualizations: Interactive maps at state and county levels showing variable densities, with selectable criteria for customization. For example, users can compare the density of harvested acres for corn across multiple states.
3. Comparison Graphs: Box plots and line graphs standardizing variables for comparative analysis across regions and timeframes.

Figure 1 displays a visual representation of the scaled values of average yield, total production, and total harvested acres. Trends in Figure 1 help determine which years saw an increase in yields and how this relates to harvested acres and production. Figures 2 and 3 display the interactive map tabs, with Figure 2 showing state data and Figure 3 showing county data. The selected filters for these maps are shown on the left-hand side, and the stats of the selected county and state are shown on the right-hand side. Next, Figures 4 and 5 display the comparison graphs produced after a second variable is selected. Figure 4 shows a z-score standardized box plot comparing these selected variables, while Figure 5 shows the bar graphs and a line graph showing the comparison across time. Figures 6, 7, and 8 are all crop condition and progress graphs. Figure 6 shows separate line plots across the weeks of the year showing the values of "Good" and "Excellent" crops in the selected area across the surrounding years. Figure 7 shows the cumulative percentages of crops in each crop condition category across the planting and harvesting timeline in the selected area and year. Finally, Figure 8 displays the progress of each variable listed in the legend as a percentage throughout the weeks of the crop year.

From further analysis using these visualizations, quantitative conclusions can be drawn about agricultural practices. For example, counties with consistently high yields over a decade may provide benchmarks for best practices. This will benefit agricultural industries and emphasize this project's significance. Future results will be drawn by NASS and the US Department of Agriculture after this app is deployed publicly and feedback is received from users across the region.

Discussions/Conclusions:

The project was concluded with the Shiny App being deployed for public use. The visualizations throughout the app are able to assist farmers in making strategic decisions for agriculture in their selected states or counties. Additionally, data users can observe the visualizations and realize trends. The initial goal of this project was to create a portal where helpful and easy-to-read visualizations could be made using data from NASS's survey and census programs. Allowing the general public to have access to the portal was important when working on the project. The portal is a success in this area due to a link being easily accessible to deploy the app, having a clear simple design for the app and its functions, providing a walkthrough YouTube video explaining the use of the app, and providing simple explanations of certain visualizations. Farmers will be able to use the app to observe if their farm is competitive when compared to nearby counties, and they could then make more informed and strategic decisions regarding their farm.

In the future, a weather data source could be integrated, and climate change implications could be observed with the comparison of the weather data and current data in the app. Although the app is impressive and a success when compared to the initial goals, there is room for improvement. One shortcoming would be that new data from NASS will not be integrated automatically into the app.

Overall, this tool offers significant value to the agricultural sector by enabling data-driven insights and strategic planning. Its deployment marks a step forward in making complex agricultural data accessible and actionable for a wide range of users.

References:

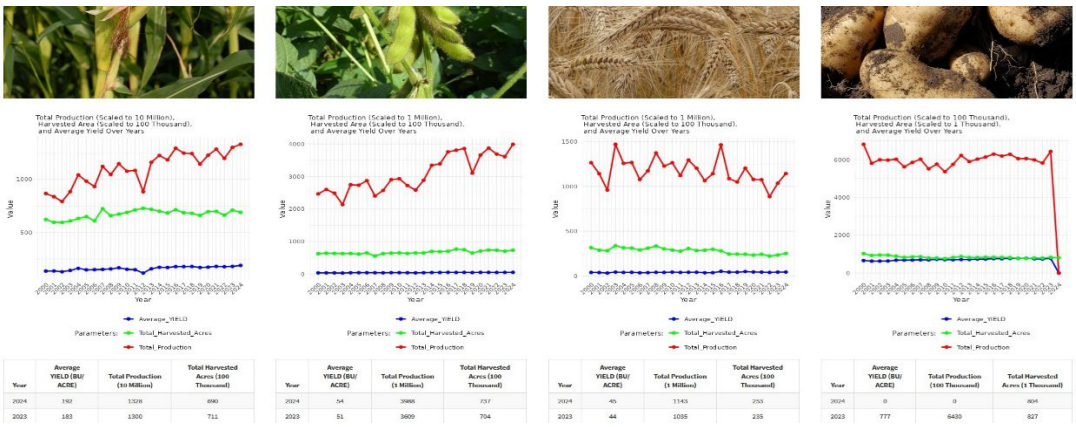
[1] “NASS - Understanding Agricultural Statistics - Education and Outreach - Understanding Agricultural Statistics.” Accessed: Dec. 10, 2024. [Online]. Available: https://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/index.php

[2] “USDA.” Accessed: Dec. 10, 2024. [Online]. Available: https://www.usda.gov/

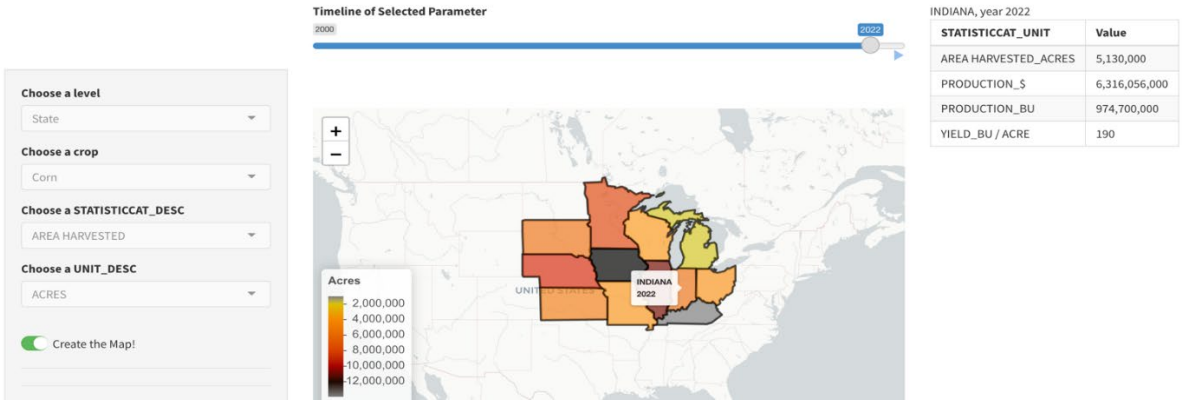
[3] “USDA – National Agricultural Statistics Service – National Crop Progress.” Accessed: Dec. 10, 2024. [Online]. Available: https://www.nass.usda.gov/Publications/National_Crop_Progress/

[4] “USDA/NASS QuickStats Ad-hoc Query Tool.” Accessed: Dec. 10, 2024. [Online]. Available: https://quickstats.nass.usda.gov/

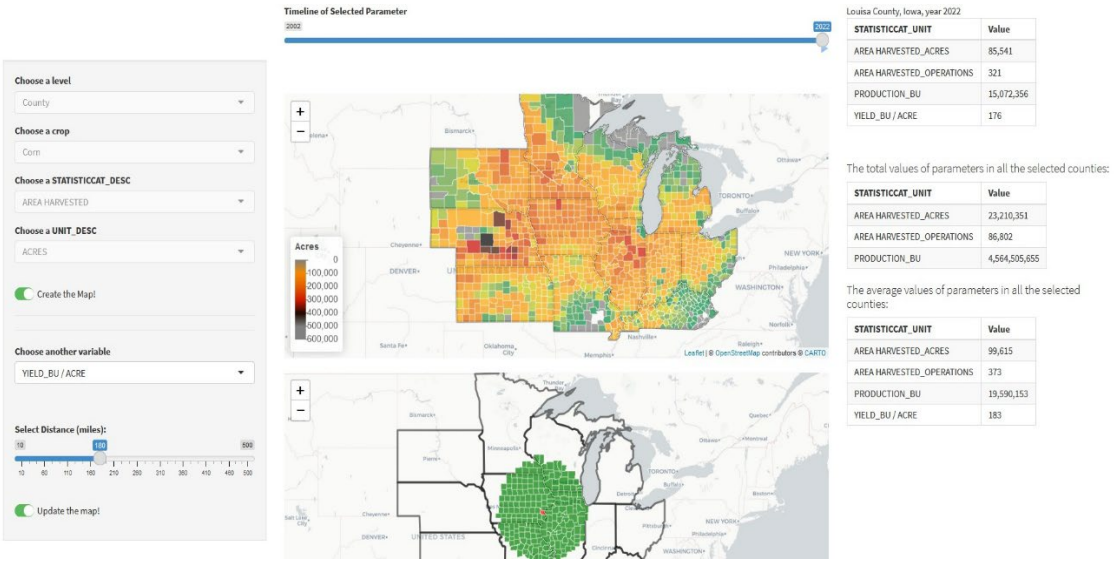
Appendix



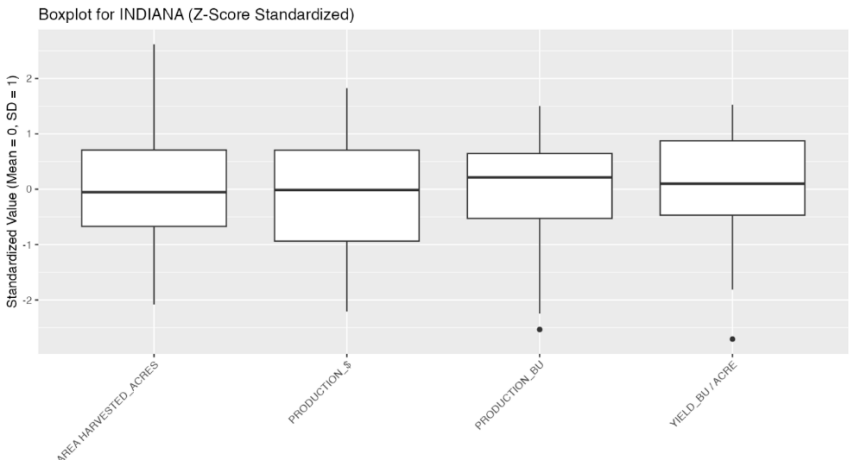
[Figure 1]



[Figure 2]

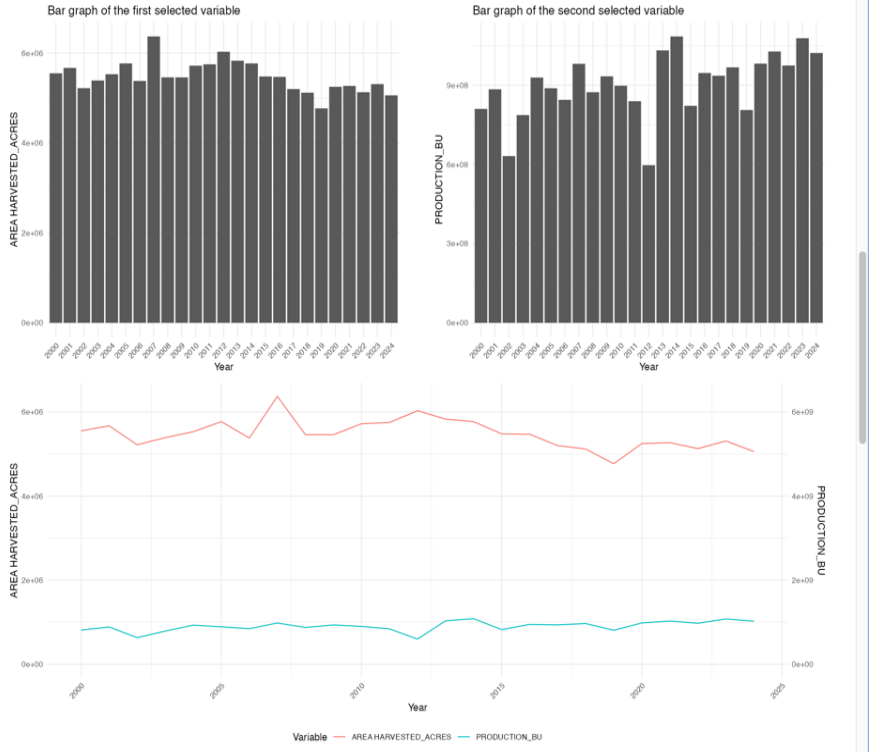


[Figure 3]

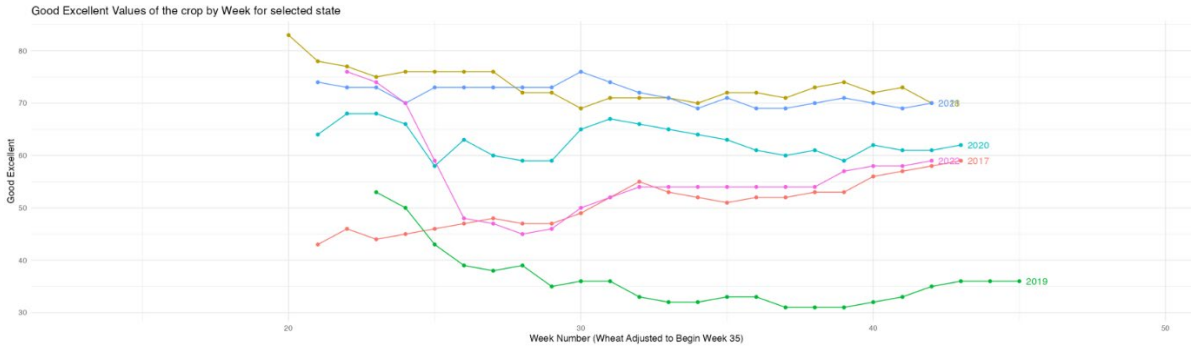


[Figure 4]

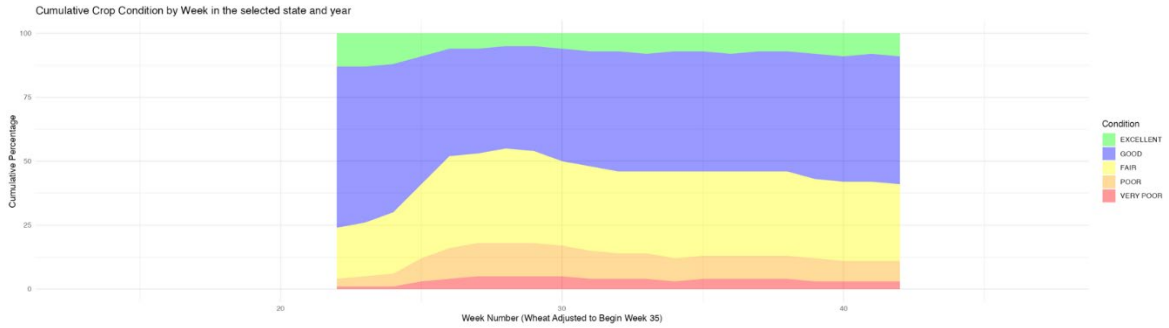
YEAR	AREA HARVESTED ACRES	PRODUCTION BU
2000	5550000.00	810300000.00
2001	5670000.00	884520000.00
2002	5220000.00	631620000.00
2003	5390000.00	786940000.00
2004	5530000.00	929040000.00
2005	5770000.00	888580000.00
2006	5380000.00	844660000.00
2007	6370000.00	980980000.00
2008	5460000.00	873600000.00
2009	5460000.00	933660000.00
2010	5720000.00	898040000.00
2011	5750000.00	839500000.00
2012	6030000.00	596970000.00
2013	5830000.00	1031910000.00
2014	5770000.00	1084760000.00
2015	5480000.00	822000000.00
2016	5470000.00	946310000.00
2017	5200000.00	936000000.00
2018	5120000.00	967680000.00
2019	4770000.00	806130000.00
2020	5250000.00	981750000.00
2021	5270000.00	1027650000.00
2022	5130000.00	974700000.00
2023	5310000.00	1077930000.00
2024	5060000.00	1022120000.00



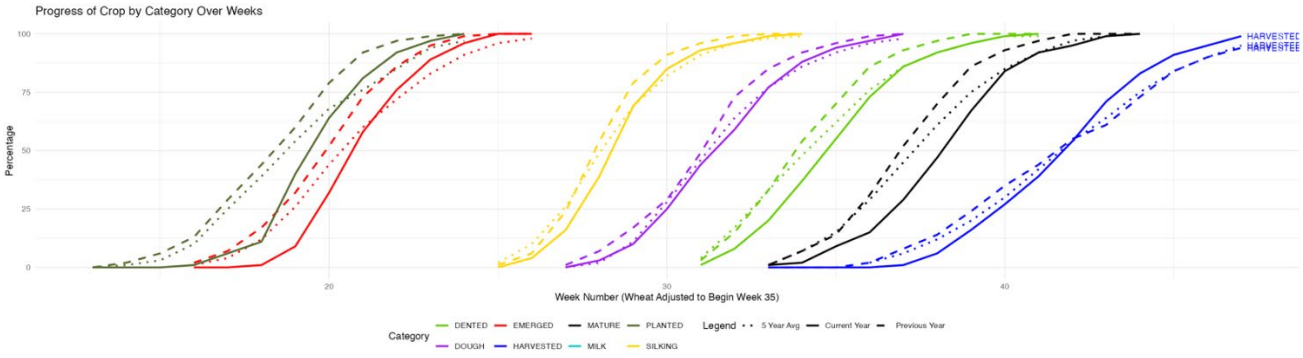
[Figure 5]



[Figure 6]



[Figure 7]



[Figure 8]