Exploring GDP Deflator in the Top Ten Agriculture Producing States in the U.S. from 2000 to 2017 using Data Science

By

Radhika Kataria

Submitted to the Board of Purchase College School of Natural Sciences In partial fulfillment of the requirements for the degree of Bachelor of Arts

Purchase College State University of New York

May 2019

Sponsor: Dr. Irina Shablinsky

Second Reader: Professor Knarik Tunyan

Abstract:

As technology is becoming more and more prevalent in everyday use, it is important to consider the benefits of its uses to help solve problems. The field of agriculture requires technology to measure resources and its impacts in different sectors. By using Applications of Data Science to study the GDP Deflator t in American Agriculture, we can measure the growth of GDP to see the effect of farming on the economy. In this paper I will study the GDP Deflator in American Agriculture using applications of Data Science, specifically focusing on 10 specific states that have been the largest producers of agriculture in 2017. Data Science can help analyze certain economic trends and can help identify the growth of GDP. This paper will examine the top ten states with the biggest agricultural producers in the United States, to find the impact of farming on the American economy and the overall growth of the GDP from 2000 to 2018. I'm writing this paper because I believe that agriculture is a very important source of income for a nation, and I think it has enormous potential to contribute towards the American economy. Agriculture contributes to job growth, diversity of resources, and is a large source of food production.

Table of Contents:

Chapter 1: Literature Review

Literature Review of *The 20th Century Transformation of U.S. Agriculture and Farm Policy* by Carolyn Dimitri, Anne Effland, and Neilson Conklin

---

This literature review is an analysis of the article *The 20th Century Transformation of U.S. Agriculture and Farm Policy* written by Carolyn Dimitri, Anne Effland, and Neilson Conklin. This article was published by the United States Department of Agriculture specifically by the Economic Research Service. This article examined the structural and gradual evolution of American Agriculture in the 20th century, specifically how certain 20th century political policies shaped American farming. From the Start of American History, farming has shaped the surrounding community through it's productivity and outputs. Many social and political policies have changed the Agriculture process in the United States. The authors discussed the intensity and rigor of agricultural development from the 20th century into the 21st century. They talked about variables involved in the difficulties of farming during the 1900's such as the lack of technological innovation as well as the small commodities such as production amount. Because of policy changes and innovation, agriculture has managed to grow and become a part of the global economy by growing the GDP (Gross Domestic Product). The authors talk about the growth of farming households compared to individual farms as well as the production of farm commodities. They also found that from 1930 all the way up to 2002, farm households had a large growth of off-farm income, which meant they received a part of their household income from non-farming occupations. Because farms have become more specialized from the 20th to the 21st century, the amount of commodities produced per farm has decreased as well (Dimitri et al.). Many of the graphs also display the Nonmetro farming income, which is defined as receiving income that was non farming based. In one specific graph, the number of farms were graphed against the years starting from 1900 to

2002. The results found that the average farm size per acre increased as time went on, while the total number of farms decreased. These changes were driven by specific factors. The push for both technological and economic development, helped to develop and diversify the farming sector in the United States. The authors analyzed the mechanics involved in maintaining the farming process. Technological development such as the invention and implementation of the tractor contributed to efficient farming. Tractors assisted with replacing work animals and improving work conditions, as technological innovation increased. The variety of machinery seemed to grow and replace work animals from the 20th into the 21st century. Other advancements in the creation of pesticides and fertilizer also contributed to farming efficiency. The concept of Consumer Influence also boosted the farming industry. Constant pressure from the consumer made it more demanding for farmers. In parallel, as consumerism increased so did environmental concern. Many people demanded and advocated for environmentally friendly farming practices. Global Markets also assisted in drawing the agriculture sector in the Global economy. At the start of 1910 agriculture exports reached their peak and did outstandingly well up to 1914. They remained constant until the 1970's, where they started to rise again. During the 1990's, the second wave of globalization helped boost the farming sector once again. Farming was also influenced by political and economic policies implemented during these times. Farm policy and income have always been the heart of American Agriculture policy. After World War I, the Agricultural Adjustment Act (AAA) was formed as a response to the weak economy. The program has evolved over the years to be more helpful and efficient. Because of the Great Depression farm household incomes were very low. Federal reform changed the AAA and made it more helpful and efficient. After World War II farming productivity increased because of mechanical and chemical innovation. In 1965 the Food and Agricultural Act was created to monitor elements of supply and manage control of supply but also merged new income support for payments for farmers. This Act assisted American farmers with exports. In 1996 the Federal Agriculture Improvement and Reform Act ended supply control for farmers and allowed more freedom in terms of what they could do with their commodities. Agriculture is often shaped by political, economic, and even social policies and beliefs. It is important to consider the evolution and growth of farming on a global scale from the 20th century into the 21st century. This article truly defines Agriculture growth as an adaptation to the factors in or around it. Political and Economic policy has had a large impact on farming as it affects the income, exports, commodities, and overall growth of farms. Innovation and technological advancements have also impacted American farming, helping it's growth and income. This article was well written and used excellent visual and graphs to depict the growth of American farms. It also appropriately analyzes and dissects the factors that contribute to efficient farming practices, on the economic scale as well as in a historical context, which is beneficial to my project as it helped me understand what affects the economics of agriculture. It is beneficial to study the evolution of farming, and the factors that contributed to that evolution. This article was advantageous to my research as it helped me understand what conditions lead to current day farming and how sensitive the agricultural economy can be. It was interesting to see the progression of American agriculture.

Chapter 2: Thesis and Purpose including Topics of Study

For my senior research I explored data science applications to study economic development of farming, specifically studying the growth of the GDP of the top ten agriculture producing states in the U.S. from the years 2000 to 2017. I believed that analyzing the GDP Deflator of the top ten states would help highlight economic growth of farming. I studied the increase in GDP Deflator from 2000 to 2017 to analyze the economics of agriculture. Finally I attempted to analyze my findings to see the effects of the GDP Deflator on the ten states.

1. Data Science

Data Science is an interdisciplinary field that combines programming, databases, and mathematics to organize and analyze data that can provide answers to future problems, as well as verify the efficiency of existing solutions. Data Science problem solutions can be implemented in various programming languages like Python, Excel, Java, SAS, etc. Certain software packages such as Tableau were created to specifically implement data science practices. Python has an advantage in data science, as it is a very versatile language and has a large number of built-in libraries. The general process includes gathering data (via API or key, and loading in datasets), cleaning data, feature engineering it, and implementing models and methods to help analyze and study specific patterns and points in the collected information. Data modeling is a very powerful aspect of data science. It involves creating graphs, histograms, charts, etc. to analyze and draw conclusions for problems. It’s an efficient way of visualizing large amounts of data. Built-in libraries such as Pandas, Matplotlib, and TensorFlow can also be very helpful in organizing data and studying patterns. Pandas is a library that creates data frames and charts that organize and sort data.

After the data is plotted, another way of measuring how well the data can be fitted to results or patterns is the implementation of algorithms. TensorFlow and Scikit-Learn are machine learning libraries that utilize certain predictive algorithms. They also assist in pre-processing and selecting data. Linear regression is a common regression method used to predict and estimate relationships between different variables. The beauty of linear regression is that it assists in helping find patterns and trends in data. It can also help outline the fit of the data, predicting how close the data is packed together or fitted, giving us an idea of how precisely the data points are plotted. Another algorithm that is useful in studying patterns in data is KNN also known as (K-nearest neighbors). KNN is used for testing predictions, as it generates the mean of a target based on a similar measure of data points. It requires one point to be set as the standard point of measure, and then scales the other data points around it. These algorithms are helpful in determining how efficiently and accurately the data was fitted while also pointing out trends.

2. Python

Python is a very diverse language, with many different features and vast built-in libraries. It is a highly adaptive programming language, that is able to accomplish whatever the user needs it to do. Many of its advanced built-in libraries, such as Pandas, Scikit-Learn, and Matplotlib can perform specific tasks and implement algorithms based on what needs to be done. The Pandas library works very efficiently with large data sets grouping and creating large data frames for data analysis. Scikit-Learn is beneficial for applying machine learning methods to a data set, such as clustering or classification algorithms. Linear regression is implemented to observe whether the plotted data has some sort of correlation or relationship that exists between variables. I used an open source software called Jupyter notebooks to implement and plot the data. Jupyter notebooks is an open source cloud platform, that allows data charts and graphs to be visualized, while also executing python based functions and calculations. In Jupyter notebooks, code can be written, executed, and presented visually through Matplotlib. For my project, I created graphs using code and functions to analyze economic development of agriculture by looking at the GDP Deflator.

Chapter 3: Subjects and Topics Involved in Study

1. Agriculture

The field of agriculture is highly specialized and diverse in production. Farming is a very important source of food production. Farms are also diverse in terms of what they produce. The United States Department of Agriculture (USDA) is the federal agency in charge of recording all agriculture related functions in the United States. The USDA predicted that there is a need to increase farming productivity to boost economic development in the United States. Multiple factors dictate how efficient the outcome of farming can be (inputs & outputs). Factors such as what item is being produced, in what state, how many units of that item are being produced, and the effect of what is produced are all factors to consider when measuring the economic development of farming. Studying and analyzing farming data, may point to trends in how efficient farming methods can be. In order to analyze how farming trends can possibly contribute to economic growth and patterns I looked at different data sources. In the *Statista* dataset *"Distribution of total farm production expenditures in the United States in 2017, by type"*, the average costs involved in the agriculture industry were organized by amount and type of expenditure. These two variables gave insight as to their impact on the economic development of agriculture.

In that statistic, the main farming expenditure were feed (15%), farm services (12.2%), and livestock poultry related expenses (11.7%) (*Total U.S. Farm Production Expenditures by Type 2017 | Statistic*). According to the USDA, in 2017 the top ten agricultural producing states in terms of cash receipts (compared to individual value of items) were in descending order: California, Iowa, Texas, Nebraska, Minnesota, Illinois, Kansas, North Carolina, Wisconsin, Indiana ("U.S. Farming"). In the interactive dashboard, *Get to know your State*, farm facts are tabulated by number of farms, acres of farmland, net farm income, government payments, federal insurance premiums, and federal insurance indemnities. These factors helped analyze economic development in American Agriculture based on their contribution towards efficient farming outcomes.

The leading commodities based on value in 2017 farm production were (in descending order): cattle/calves, soybeans, dairy products/milk, broilers, miscellaneous crops, hogs, wheat, chicken eggs, and cotton lint. According to the Economic Research Service (ERS) in 2017 approximately 88.2% of American households were food secure throughout the year, while 11.8% of Americans were food *insecure*, and finally 4.5% had very low food security. This potentially led to disruption of normal eating patterns, and reduction of food intake for members of those households. This statistic seemed to be a little too high for a developed nation like the United States. Emphasizing economic development of farming could help lower that percentage. In 2015 farming added a 4.5% increase to the nominal GDP (*USDA ERS - FAQs*). In the USDA dataset that I worked with titled, *Farm Income Wealth Statistics*, there were 11 variables, and approximately 402,000 entries. The data ranged from 1910 to 2018. The variables used to organize the data were year, state, artificial key, variable description total, variable description part 1, variable description part 2, amount, unit description, publication date, source, and GDP deflator. These variables listed the expenditures and farm income based on individual state and by year. The variables also listed what items were produced, and how many units of that item were produced by state in total. I worked with the top ten agriculture producing states in the year 2017, specifically within the range of years from 2000 to 2017. I studied and analyzed the growth of the GDP Deflator in the top ten agriculture producing states in 2017, by graphing the growth of the GDP deflator from 1910 to 2018 and from 2000 to 2018 using Python.

2. Economics

The GDP (gross domestic product) is a measurement of the contribution of resources and goods that a nation produces (exports - imports), and sums up the total as a value of income. The GDP Deflator is a measurement of the change of prices of all new domestically produced services and final goods within a country. The deflator exists as a ratio or fractional value, and is used as a statistic to measure the value of a data point from a set base year. GDP can fluctuate due to many factors (environmental, political, social, etc.), and farming does contribute to a nations GDP. In the book, *The Economics of Agriculture: Evolution and Global Development* author Steven C. Blank wrote about the impact of agriculture on the American economy and it’s effect on global development. He made a point of talking about how agriculture is declining as a profession in the United States, when comparing 1950 to the later 1990's. Blank also pointed out that a declining trend of an industry is an "inferior good" in the context of national economic development and investment. It is also considered a positive sign of economic advancement, as nations that are well developed tend to have more emphasis and choices of non-agriculture related occupations. This is derived from the idea that agriculture provides a better return on investments which are comparatively lower than returns that come from investing in labor and other industries in the economy. Interestingly economic development of agriculture also requires analysis of socio-political events of that nation.

Much of the evolution that arose from the 20th century into the 21st century in the United States came from the advancement in political, social, and economic policies as well as technological innovation (Dimitri et al.). Many political factors such as treaties, wars, and regulations impacted the growth and production of Agriculture. The introduction of technological innovation, such as the tractor helped to boost farming production while lowering the need for work animals. This advancement was helpful in creating more farm capital by boosting production and lowering costs compared to using work animals. Many economic changes were caused by political issues in that time. World War I and World War II had a big influence on Agriculture output and income, as trade was a part of a nations asset and wealth. Exports were also halted during the wars due to global dispute. This was not beneficial for the agriculture sector, as exports were a major source of income and created an economic presence for the United States. Exports boosted the economy as it helped with international business, and assisted with raising the GDP.

The macro effect of agriculture, helped to structure economic trends and political relationships. Because agriculture has a large effect on the production of consumer products, it is important to consider its significance in the United States. Agriculture is essential for sourcing food and different resources. The Micro effect of agriculture helped analyze the developments needed to have productive farm related outputs. The micro effect is sensitive to changes in environment and government. Studying item production and volume can also point out small economic changes. The macro and micro effects of agriculture help analyze economic development in the United States.

Chapter 4: Methodology and Materials

Methodology

In my project, I implemented multiple different methods to accomplish my task

1. State my Hypothesis

The applications and processes of data science can help highlight and study economic trends or *ideas* in American Farming. I believed that American Agriculture helped contribute to the growth of the American GDP, and studying the 2017 GDP Deflator of the top ten agriculture producing states in the U.S. would be helpful in understanding economic development. The purpose of this project was to explore the GDP Deflator of the top agriculture producing states of 2017, while learning about the process and functionality of Data Science using Python.

2. Investigate with the use of software

I used Jupyter notebooks (Anaconda Cloud platform) to load in the data, and visually depict it through graphs using Python. After loading in data, I had cleaned a part of the data, attempting to separate any erroneous values from true values. This helped me visualize the data more efficiently. Once the data was cleaned, I looked at variables that have potential economic relationship (ex: GDP

Deflator & Year, etc.). The I presented the data through graphs and plots using Pandas & Matplotlib libraries to create my visuals. Finally I measured and studied any trends or relationships in the plotted data. Jupyter Notebooks is an open source platform that runs Anaconda. It was a valuable resource for data visualization and analysis.

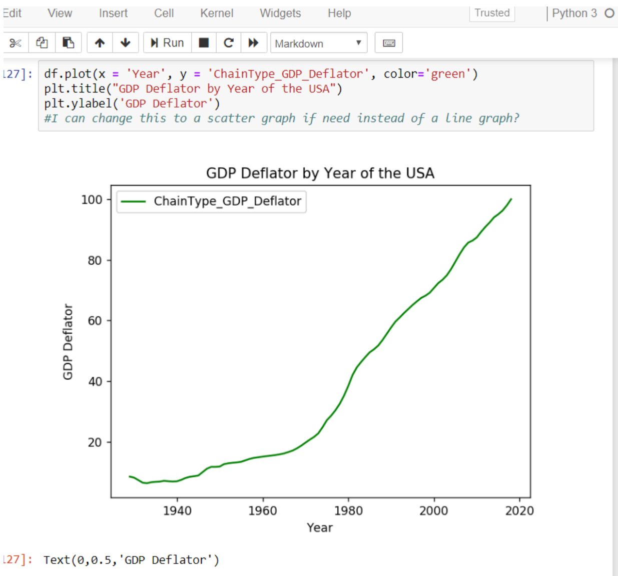
3. Working with the Data

The process of working with the data was interesting. I obtained a dataset from the United States Department of Agriculture (USDA). The original dataset was an Excel dataset with over 400,000 entries. There were eleven variables in the dataset. The variables included were *year, state, artificial key, variable description 1, variable description 2, total variable description, amount, unit description, publication, source, and GDP Deflator*. The variables I focused on were *Year, State*, and *GDP Deflator*. Looking at those three variables, would show the relationships between states and economic growth. The other variables such as *variable description 1* had specifics as to what individual products each state produced, and *variable description 2* contained what type of farming was done based off of products that were produced by that state. *Total variable description* combined *variable description 1* and *2* with all commodities. *Publication* and *Source* variables specified that the data had come from the USDA. The *Amount* variable had the specific number of units produced and *Unit Description* was a constant variable that had the cost of a single unit that was produced. The ten states that I looked at were *California, Iowa, Texas, Nebraska, Minnesota, Illinois, Kansas, North Carolina, Wisconsin,* and *Indian*a. According to the USDA, these states were the largest producers of agriculture in 2017 (*USDA ERS - FAQs*).

Using these three variables, I studied specific economic growth within a specific range of years, specifically the GDP Deflator among the ten states that were the largest agriculture producers of 2017. The first thing I did was extract the data, and read in the file into the software. I used an API key from the USDA to open and down load the file, after which the file was converted from a GZ file and then finally into a readable CSV file. I considered using an online file converter, but instead manually unzipped the file from a GZ into a ZIP file, and lastly into a proper CSV file.

The data cleaning process required me to iterate through the individual values in the data and find the null sets, which had to either be removed or separated. I considered cleaning the data using Excel but instead I cleaned the data in Python. It was both a new experience and learning process for me. I had some previous experience programming in Python, but I also wanted to learn more about Python, especially all the inbuilt libraries and tool-kits. I worked with my erroneous values by using the built in python functions *isNull*() and *dropna()* to attribute for my missing data. The main data that needed to be cleaned was data from 1910 to 1920. I had worked on cleaning my null values in the dataset, but all the erroneous values where in the range of years that I wasn't working with. After consideration, I decided that I didn't need to clean those data points, as I was working with data from 2000 to 2017. As I started creating my graphs, I had stumbled upon an issue when I decided to graph the GDP Deflator by state. I was unable to graph the data at the specific year intervals I needed. To fix this, I implemented a data frame as to selectively extract the data from the CSV file based on the specific years that I needed. A data frame is an built-in data structure in Python that is often used to cluster variables together. The data then can be easily retrieved compared to pulling the specific values from the original data file. I then graphed the data, to visualize the growth of the GDP Deflator.

I graphed the number of years against the GDP Deflator of the entire United States from 1910 to 2017 in figure 1, and from a specific interval of years from 2000 to 2017 in figure 2, but this method was not working for the ten specific states that I need to graph. Using the x and y variable assignment in Python was beneficial for graphing a single variable.



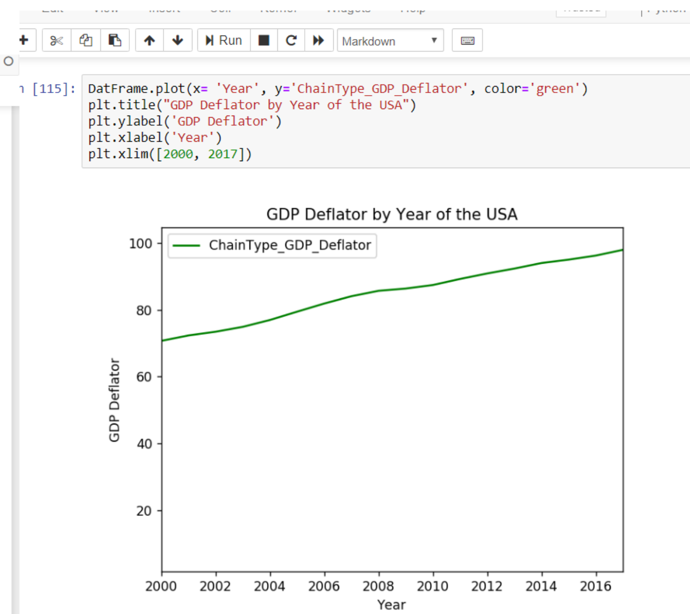
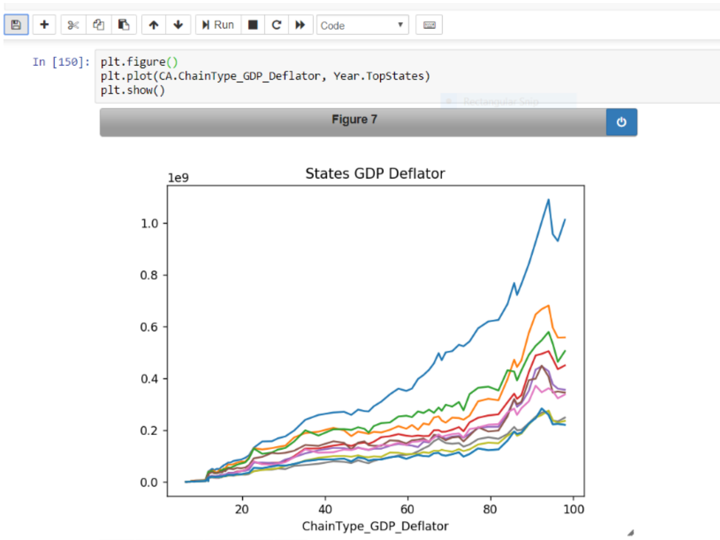


Figure 1

Figure 2

Afterwards, I graphed the GDP Deflator vs. the years 2000 – 2017 for the top ten agriculture producing states, but struggled with implementing the correct value for my y variable. When I created this graph, my y-axis was an unexpected logarithmic scale. At first, my graph had the GDP Deflator, instead of the number of years automatically set as the x-axis which can be seen in figure 3. I solved this issue by setting my group-by variable parameter as the Year, which then graphed the States by year vs. GDP Deflator as seen in figure 4. The group-by variable was helpful in clumping together different variables that I wanted to look at.



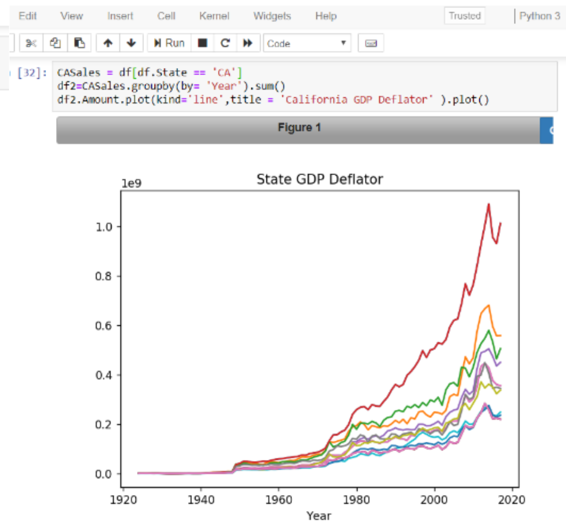


Figure 3

Figure 4

Although I had a graph for the top 10 states from the years 1910 – 2017, I only needed a snippet of that graph for my data analysis because I was focusing on the specific range of years from 2000 to 2017. Again, I implemented the *xlim*() function, which scaled the x-axis appropriately as shown in figure 4. I graphed each state using the group-by function, which pulled an individual state from the State variable from the original data set. I ended up creating the code for each individual state, of the ten states that I needed, in order to graph the Deflator of each state. The y variable was still displaying a logarithmic scale with unknown units as seen in figure 5.

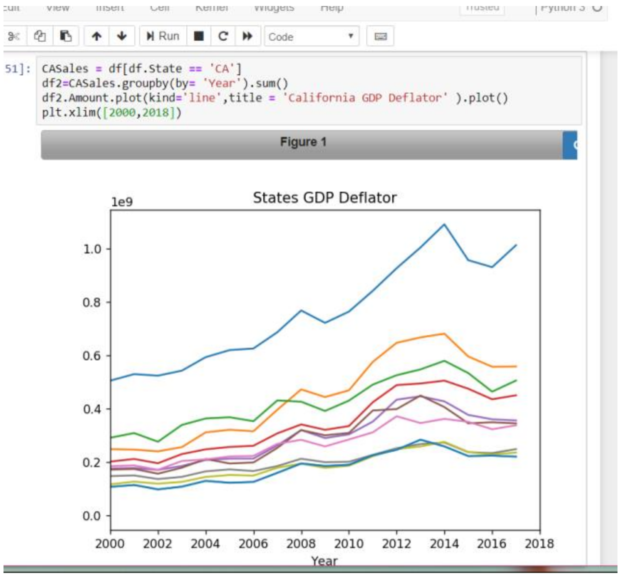


Figure 5

Because the graph in figure 5 was still somewhat abstract and hard to understand, I created another graph with the same variables. I implemented 2017 as a base year, for the GDP Deflator, so all the other data from the other years was to be compared to 2017. I had graphed the GDP Deflator by Year for each individual state of all fifty states using x and y variable assignment in Python. This was a simple yet effective way to graph two variables independent of one another. It was done by assigning x and y values to the variable, and then plotted as a figure. The graph displayed all fifty states, by GDP Deflator, thus becoming a more effective way to plot the Deflator against the number of years.

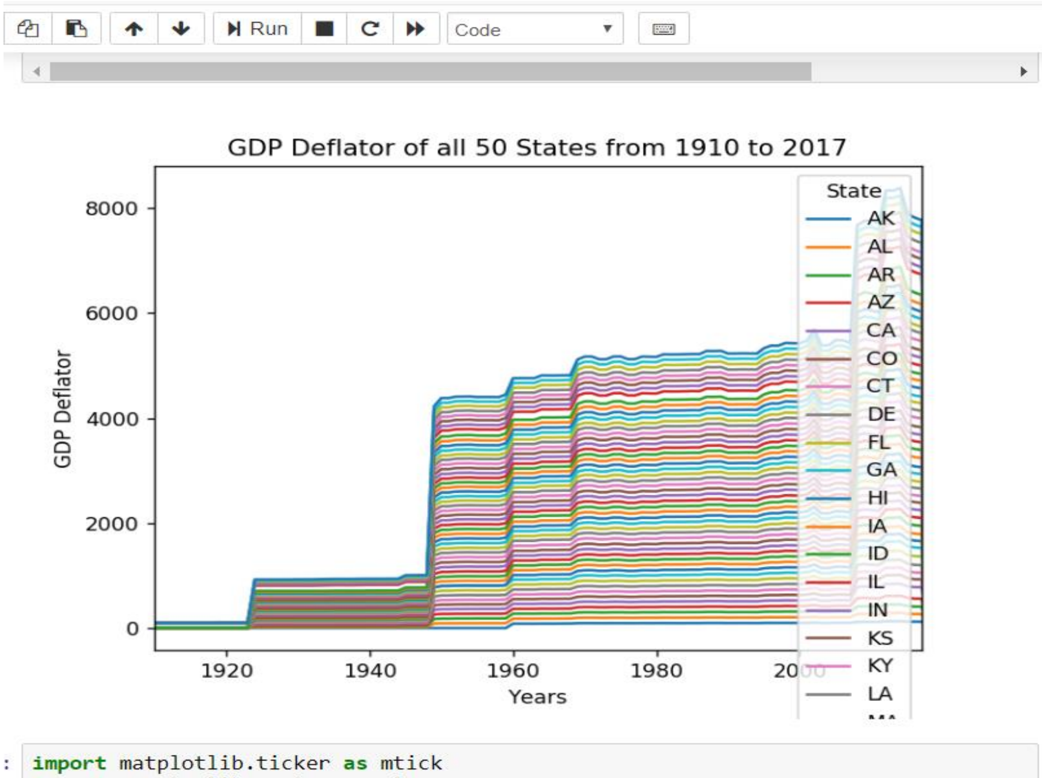


Figure 6

I next made a graph for the years 2000-2017, using a different group by function that created a better graph for the states GDP Deflator ratio, which ultimately became my graph seen in figure 6. I used the *xlim()* function that allowed me to give my x-axis the appropriate range. I implemented the lambda function, which created a function within a function based on what I needed by defining different parameters. I used the lambda function to create a function that would scale the GDP Deflator to a full 100%, and thus it allowed me to compare the graphed data to the set base year data. These graphs showed the evolution of this project, in terms of the diverse use of data visualization and the different types of functions that creates these graphs. For me, this was an evolving process, as each graph was better than the one that was created before it. I also found, that I learned more with the creation of each graph.

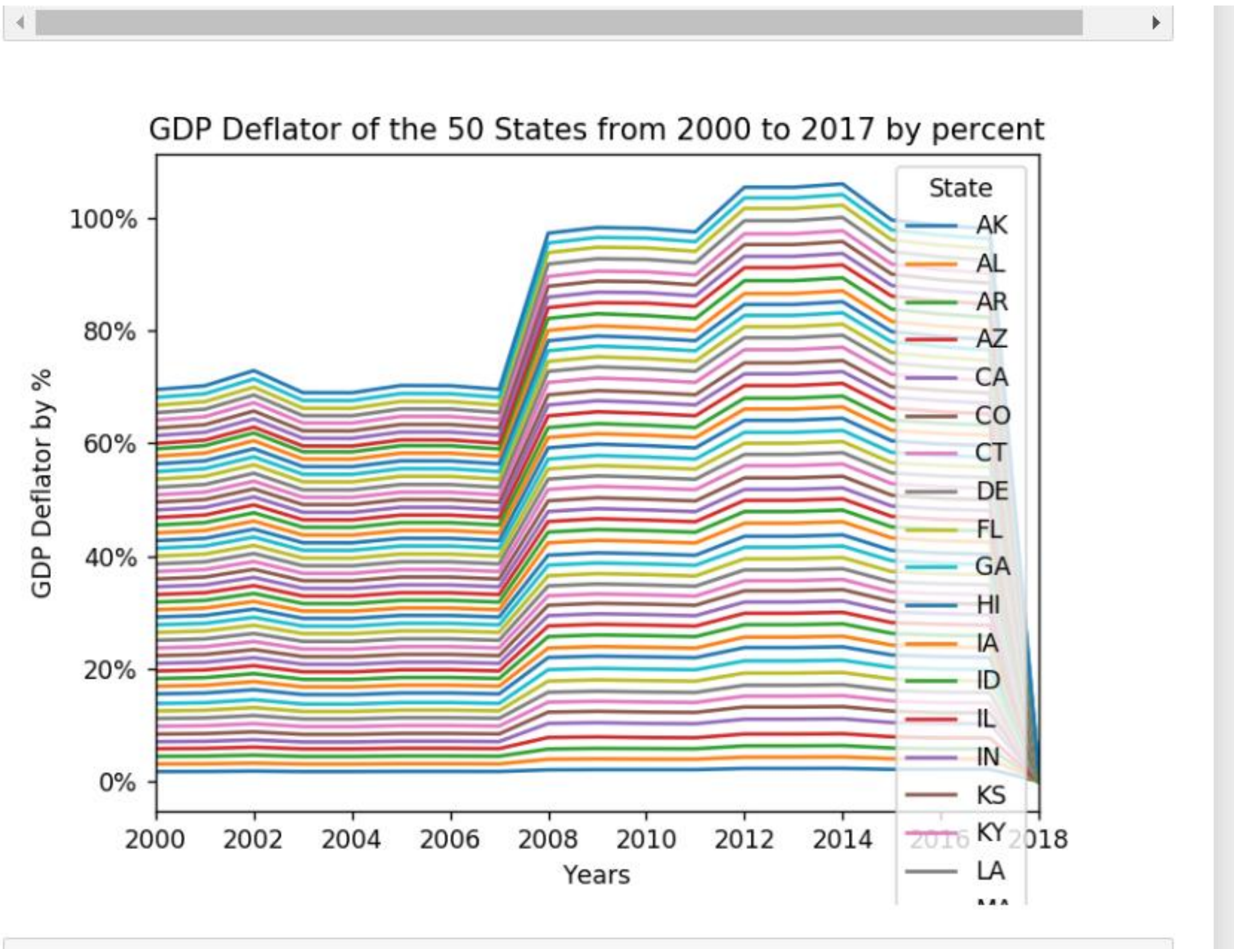


Figure 7

Another library that I used to create these graphs, other than Pandas was Matplotlib which implemented tick marks on axis. The Matplotlib library is very useful when graphing visual plots of data. It contained a good number of add on toolkits (pyPlot, Ticker, etc.) for data analysis and visualization. In figure 7, I implemented the lambda function and the add on toolkits to scale the y variable as a percentage of the GDP Deflator variable. Matplotlib was a helpful library for the graphs as it helped me implement specific axis and titles for the plot. Although I used Pandas to create some of my first graphs, Matplotlib was a very helpful inbuilt library that helped me visualize and categorize the data.

4. Discuss Results

After I plotted the data, I reviewed my graphs to see if the graphs were helpful and analogous to my thesis. Although I had a graph that displays the GDP Deflator against the number for all the fifty states, I think it would've been more beneficial to graph the top ten agriculture producing states in 2017. Because of this being a learning process, I wasn't fully able to graph the ten states with the correct x and y variables. For future development of my project, I would like to be able to graph the GDP Deflator against the specific interval of years for the ten states that are the biggest agriculture producers of 2017. This would help analyze economic growth and development of farming. Using Python I was able to group by function the GDP Deflator by state and year as a list. I found that all ten states had the same GDP Deflator from 2000 to 2017. The GDP deflator grew from 70.729 in 2000 all the way up to 97.958 in 2017. Because GDP Deflator measures changes in price level, having an increase in the deflator means that the overall aggregate level of pricing will also increase.

The deflator seemed static, but another interesting thing I found was that the states that had the highest amount of farming production per year in descending order were California, Iowa, Texas, Nebraska, Minnesota, Illinois, Kansas, North Carolina, Wisconsin, and Indiana. California had the largest rate of production in 2017 compared to the other 9 states. I think having a growth in production rate could have added to economic development and possibly even contribute to overall growth of farming. This finding also supports the USDA statistic of the ten states that had the highest farming production in 2017. I was happy that I found the deflator to be static and the amount of production to be the same as the states that had the highest farming production in 2017. GDP Deflator is a good indicator of price changes in an economy, but for my project it did not seem to tell much about economic development in American agriculture. A future expansion for this project can be narrowing down the ten states based on amount of production as well as looking into variable descriptions of what was produced in each state.

In my project, the use of data visualization through graphs was very advantageous to me in understanding growth of variables. Using the inbuilt libraries Pandas and Matplotlib, I was able to create appealing graphs that were helpful in data analysis. The GDP Deflator has gradually increased from the year 1910 to 2017, as well as from 2000 to 2017 with small bumps along the gradual progression of the data based on my graphs. Although my graphs are a work in progress, they still show the increase of the deflator against the number of years.

5. Conclusion

Using data science to study the growth of the deflator of the ten states that had the highest farming production in 2017 was a great experience for me. For this project, I don't believe that the GDP Deflator depicted significant changes on the economy of Agriculture of the ten specific states I was working with. Although the ten states had the same rate of change in deflator that increased gradually by year, they all had different rates of production. GDP Deflator measures fluctuations in price level, thus having growth in the deflator means that the total aggregate level of pricing will also increase. GDP is very sensitive to changes in price level. Studying the deflator gives insight into the development of the farming economy and its price levels. Because the deflator was the same for the ten states, and only increased consistently I don’t think that the deflator gave too much insight into economic development of agriculture. Studying other factors, such as production rate, and amount of units produced may also help us understand the economic development of farming in the U.S. Especially because each states have specific funds that are allocated for agriculture development, which can also impact the economics of farming. I think that using data science was very helpful in this project, as it helped me learn more about data analysis and visualization while learning about the economic development of Agriculture. Using Python to manipulate specific variables and graphing them for visuals was helpful for me to understand whether the GDP Deflator had changed or not. I felt that this project helped me expand my knowledge on the economics of the farming sector in the United States and helped me study the growth of the GDP Deflator using Python. Farming is an intricate occupation. It has many factors that contribute to its overall production. The economics of farming is dependent on technological innovation and political decisions that drive it’s efficiency. I’m thankful for this project as I learned so much, and it always seemed to evolve to become better each and every time that I worked on it.

Struggles/Frustrations/Challenges/Inhibitors

A struggle I faced when I started working with this project was loading in the dataset into the Jupyter notebooks software. I retrieved the data from the USDA website, using an API key. Once I got the data on my computer, it was packed in a GZ file and had to be converted to a ZIP file and then unpacked. I also considered using a unpacking software, called Winzip to unpack the dataset. After a consistent struggle, my files successfully downloaded as an Excel file. I then had issues reading in my excel file. I tried to convert the excel file into a CSV file, but that didn't work. After constant struggle I tried to think about what the issue could be.

I believed that the path of my file maybe the problem, because the physical file was present on my computer, but the software, Jupyter notebooks wasn't able to read it. I tried tracing the path of the file using the terminal – linux/unix commands. I also thought of using a specific function that reads in the Microsoft excel libraries to help me. I finally solved the problem over Winter break using a fairly simple solution. I renamed the file, test file, and converted it to a CSV file and copied the test file into the specific path of the jupyter notebooks software into the senior project folder. I then loaded in the file using a *pd.read_CSV()* file function and it worked! Another struggle that I faced was graphing the GDP from the year 2000 to 2017. It was a challenge getting the ten specific states that I needed graphed with the properly scaled y variable. The function that assisted me in helping me graph the growth of the GDP Deflator against the interval of years was the group-by function. It allowed me to group my variables together as a line graph, but in some instances it would still graph the GDP Deflator as an odd logarithmic value which did not make sense. Graphing all ten states with the correct y variable was a challenge for me, that required more attention to detail. I am very thankful to be able to be given the opportunity to learn so much about the economics of American Agriculture and Data Science in Jupyter Notebooks

Bibliography

Picazo-Tadeo, Andrés J., et al. “Assessing Farming Eco-Efficiency: A Data Envelopment Analysis Approach.” *Journal of Environmental Management*, vol. 92, Apr. 2011, pp. 1154–64. *EBSCOhost*, doi:10.1016/j.jenvman.2010.11.025.

The Future of Agriculture | The Economist. https://www.economist.com/technology-quarterly/2016-06-09/factory-fresh. Accessed 13 Oct. 2018.

McKinney, Wes. *Python for Data Analysis: Data Wrangling With Pandas, Numpy, And Ipython*. 2nd ed., October 2017. Accessed 21 Oct. 2018.

USDA ERS - Summary of Recent Findings. https://www.ers.usda.gov/data-products/agriculturalproductivity-in-the-us/summary-of-recent-findings/. Accessed 23 Oct. 2018.

Cowen, Tyler, and Alex Tabarrok. *Modern Principles: Macroeconomics*. 3rd ed., Worth Publishers. Accessed 23 Oct. 2018.

“U.S. Farming: Top 10 States Based on Number of Farms 2017 | Statistic.” *Statista*, https://www.statista.com/statistics/196114/top-10-us-states-by-number-of-farms/. Accessed 13 Feb. 2019.

USDA ERS - FAQs. https://www.ers.usda.gov/faqs/. Accessed 4 Mar. 2019.

USDA ERS - Data Visualizations. https://www.ers.usda.gov/data-products/data-visualizations/. Accessed 8 Mar. 2019.

Dimitri, Carolyn, et al. *The 20th Century Transformation of U.S. Agriculture and Farm Policy*. p. 17.

Pandas Dataframe: Plot Examples with Matplotlib and Pyplot. http://queirozf.com/entries/pandas-dataframe-plot-examples-with-matplotlib-pyplot. Accessed 30 Apr. 2019.

Total U.S. Farm Production Expenditures by Type 2017 | Statistic. https://www.statista.com/statistics/246516/distribution-of-total-us-farm-production-expenditures-by-type/. Accessed 11 May 2019.

Steven Blank. *The Economics Of American Agriculture: Evolution and Global Development*. Accessed 5 Mar. 2019.

Chen, James. “GDP Price Deflator.” *Investopedia*, https://www.investopedia.com/terms/g/gdppricedeflator.asp. Accessed 12 Mar. 2019.