

Production of Strawberries and Fertilizer Use in the United States

Researched by: George Acostalemus, Madeleine Livaudais, Yuan Wang Zhang

When you go the market and look in the produce section, you see all sorts of fruits and vegetables, but do you ever stop to think about where they are coming from or the farming process that got them there for you to purchase and consume?

Strawberries are one of the most widely cultivated fruit crops in the United States, with production heavily concentrated in states such as California and Florida. To sustain high yields and consistent quality, strawberry growers depend on fertilizers to supply essential plant nutrients. However, fertilizer use also raises important questions about environmental impact, soil health, and sustainability.

The United States Department of Agriculture (USDA) collects census and survey data regularly on all of the major crops grown in the United States. To analyze the strawberry crop, we will look at data collected by the USDA through census' and surveys from 1997 to 2025. We will focus on the six states that produce the majority of the strawberries consumed: California, Florida, Washington, Oregon, North Carolina and New York. The goal is to explore patterns, trends, and potential pros/cons of different fertilizer type including nitrogen, phosphate, potash and sulfur to understand how these inputs vary by geography and time.

Our Research Process

We begin by cleaning and structuring the data to ensure technical correctness, then conduct exploratory analyses that focus on:

- Usage distribution by fertilizer type
- Geographic difference in fertilizer use
- Trends in fertilizer use over time
- Relationships among fertilizer types

The approach follows a structured analytically strategy:

1. Clean → 2. Explore → 3. Interpret → 4. Report

The purpose of the exploratory data analysis is to not draw conclusion, but to hopefully have a better understanding on how fertilizer use reflects broader management practices in strawberry agriculture. The results can serve as a foundation for future work linking fertilizer management to yield, soil health and environmental sustainability.

What is in plant fertilizer?

Fertilizer use is a key input determining yield because, just like humans, plants need essential nutrients to thrive. The most important nutrients are carbon, hydrogen, and oxygen. Plants get these naturally though from the air and water. The other necessary nutrients are obtained from soil, but sometimes the soil isn't naturally nutrient-dense, and the plants need supplemental nutrients. The critical nutrients are the macro nutrients: nitrogen, phosphorus, potassium, calcium, sulfur, magnesium, and the micro nutrients iron, boron, chlorine, manganese, zinc, copper, molybdenum, and nickel.

Our data on the use of fertilizer on strawberry crops focuses on the four macro nutrients below. Here is the main purpose of each:

Chemical	Role	Typical Form	Organic Status
Nitrogen (N)	Leaf/vegetative growth	Urea, Ammonium Nitrate	Not organic unless it came from compost
Phosphate (P)	Root/fruit development	Super phosphate	Synthetic
Potash (K)	Disease resistance	Potassium chloride or Sulfate	Natural forms like potassium from mineral sources can be organic
Sulfur (S)	Disease control	Elemental sulfur	Allows in organic systems if naturally derived

For a bit more detail, Nitrogen promotes leafy growth and overall plant vigor, but can cause environmental runoff issues if over-applied. Phosphate supports root development and early growth, though excessive use can lead to water pollution. Potash improves fruit size, sweetness, and disease resistance, while sulfur helps regulate soil pH and supports nutrient uptake.

What constitutes organic produce?

For a crop to be considered organic, there are certain conditions that need to be followed by the farmers. This means farmers cannot use certain typical farming practices used to grow conventional products. Farmers cannot use inorganic fertilizers, genetic engineering, or radiation on their products.

For a product to have the USDA Organic seal, the product must have more than 95% certified organic ingredients. One of the biggest differences is that organic produce is exposed to fewer toxic pesticides and chemicals during the growing process.

Can fertilizer be used on conventional and organic strawberries?

Since all four of these nutrients are essential for crop growth, they are used by farmers growing conventional and organic strawberries. Organic strawberry farms may also use fertilizers, but they must come from natural sources (e.g., compost, fish emulsion, mined minerals), whereas conventional farms can use synthetic fertilizers. This explains why some fertilizer categories appear across both systems, although organic volumes tend to be lower.

Inorganic fertilizers are typically manufactured, water-soluble products and are not carbon-based. They generally tend to be cheaper due to their manufacturing process and have easy calculations to determine the application rate. Depending on the form of the fertilizer, a liquid fertilizer will have quicker absorption or a slow-release pellet will release nutrients over time. One of the negatives to inorganic fertilizers is that some of the synthetic compounds used can cause environmental harm through leaching or runoff.

Organic fertilizers, on the other hand, are typically derived from living organisms and are carbon-based. Some fertilizers might be derived from fish emulsion or blood meal as the base. Since they are derived from these living organisms, organic fertilizers are generally more expensive and considered to be slow-release. This is because the soil needs a couple of days to weeks to transform the organic nutrients into a form the plants can more readily absorb. Because of this, it can be difficult to calculate application rates with precision. One of the negatives of organic fertilizer is that it may contain weeds or pathogens, like Salmonella, since both can be found in the various living organisms that the fertilizer is derived from.

While the choice of fertilizer is important, it is still recommended to add other organic materials to help support the soil, specifically carbon-based materials like compost or grass clippings. Items like these or other ground cover can help plants receive the other macro and micro nutrients needed for healthy growth.

Why is fertilizer important?

Fertilizers play a vital role in strawberry farming by supplying essential nutrients, improving soil fertility, and enhancing fruit yield and quality. Nitrogen fertilizer, invented by Sam Eyde and Kristian Birkeland in 1905, revolutionized agricultural production and made today's high crop yields possible. As crop cultivation depletes soil nutrients over time, fertilizers restore soil vitality and ensure a strong foundation for future harvests, helping sustain the world's growing population.

The Data Wrangling Process

Cleaning the Conventional Strawberry Data Set

We begin by reading in and taking a look at our data from the USDA data base. Here is a glimpse of the raw data set:

```
Rows: 2710 Columns: 21
-- Column specification -----
Delimiter: ","
chr  (12): Program, Period, Geo Level, State, State ANSI, watershed_code, Co...
dbl  (1): Year
lgl  (7): Ag District, Ag District Code, County, County ANSI, Zip Code, Reg...
date (1): Week Ending

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Now that we've taken a look at the data, the first step is to remove any columns that are entirely one data value and thus will not be useful to our analysis. This removal of unnecessary columns also makes the rest of the data easier to manage. Next, we analyze what we have left. Upon looking at the data, we have a couple columns of numeric data that are currently character structure. In order to be able to work with the data and perform exploratory analysis, we must make sure the structure of all numeric columns are in the correct format for analysis and manipulation, so we made those changes.

Next, we investigated the `Data Item` column. This column gives us a lot of different information all in one cell. In order to make it more useful, we will remove the phrases common to all elements of the data set. For example, removing the "strawberries -" from the beginning of each line. Thus, it becomes a more informative column about the numeric data values.

In this data set, the data comes from surveys and census'. There are some differences in the data collection for each of those methods that will make it difficult to clean the data set while it is all together. So next, we will split the data into separate survey and census data frames for further cleaning. We also have a separate data set for the organic data.

Cleaning the Census Portion

Once we split the data, the census data did not need further cleaning. We also pivoted it to it's wider form to be able to see the data is a more tabular format. In the wide format, each row represents the year's output for a state and each data observation for that year is in a column.

Cleaning the Survey Portion

The survey data is a bit more complicated because each state has a bit of variability in how they report the information. We will need to further split up the survey data by state. Recall, the states we are looking at are California, Florida, New York, North Carolina, Oregon, and Washington.

Before making that split, we chose to remove the rows where the data was a forecast value. The rationale behind this was that the actual year value was also provided. After that, we split the survey data up into a data frame for each of the six states and look at each separately.

For most of the states, we will remove the **Week Ending** column and the **State ANSI** column, as they would be removed by the drop one value column function. We also wrote and used a function to drop duplicate rows. In some of the states, the survey recorded “marketing year” and “year” with the same value. We did not need to keep both rows as they represent the same value.

Cleaning the data sets for New York, Washington, and Oregon were the easiest. We dropped the unnecessary columns and ran the drop duplicate rows function and finished with a clean data set.

North Carolina was the only state that reported their data in weekly increments rather than yearly. To address that, we took the sum of all of the weeks in the year and now have a value for the year. This will allow us to compare the data with the other states.

California and Florida began the same way as the rest of the states. They both had extra data on monthly values in addition to yearly values. Since we are looking at yearly values to compare between states, we removed the monthly value rows for better comparisons.

Once each state has a clean data set, we can recombine the data into a single data frame for further exploration and also pivot the data wider for visualization in a more tabular format.

Cleaning the Organic Strawberry Data Set

First, we read in the organic strawberry data and we will take very similar steps to those taken in cleaning the conventional strawberry data set.

The data had lots of unnecessary columns, so we ran our drop one value columns functions. After that, we made sure to covert the numeric value columns into the correct structure for analysis.

The last step in getting the organic data clean is to remove the phrase common to the beginning of every row of the **Data Item** column. Once this is removed, the organic data is ready for analysis.

Cleaning the Fertilizer Data Set

This fertilizer data set from the USDA contains information about fertilizer application types, quantities, and related descriptors by US states and year.

We began the same as before: reading in the data, running the drop one value columns function to remove unnecessary columns, and changing the numeric value columns to the correct structure for analysis.

Next, we have a new column compared to the other data sets that tells us the type of fertilizer used. In this column, similar to the `Data Item` column previously, we need to remove some common words to every entry. We will do that to the `Fertilizer` column and again to the `Data Item` column so that it is clean what data value is being collected for each fertilizer type.

Now, all of our data sets are clean and ready for analysis.

Visualizations of the Data

Conventional vs. Organic Strawberry Production

To get a general sense of strawberry production, we first created two graphs that show the comparison between conventional and organic strawberry production. The graph shows production over time in years (x-axis), and the amount of strawberries produced per one hundred pounds (y-axis).

