Geospatial Technology

Seattle WA 2023

This project won 1st Place at the 2023 Washington Technology Student Association Conference.

Table of Contents

For access to the technical notebook, please click here.

Interpretation and explanation of the challenge

Precipitation has always been central to economic growth, as it's tied to agricultural output. With global warming increasing the temperature of the world, the ice caps are melting and causing rising sea levels. This is creating a polarization in the world where areas that generally don't get much precipitation are getting less and less, while areas in the world that get a lot of rain are getting even more. Thus, the ideal conditions for raising regional plants are being destroyed. As global warming is causing increasingly fluctuating and unreliable levels of precipitation, farmers are being increasingly affected. They are no longer able to produce reliable agricultural outcomes and thus their sources of income are greatly affected. As local farmers struggle to supply food, it will cause major ramifications for the consumer as they struggle to get access to food. Crop yields, as a whole, eventually declined by 18% in 2015, says Wolfram Schlenker, a professor at Columbia University (Nuwer). The impact of climate change raises many repercussions for agriculture that impacts everyone. Though solutions can be approached through a multitude of ways, many include making the transition to genetically modified plants that are more resistant to variations in precipitation, or other environmentally adaptive methods of agriculture. However, to truly understand the effects of global warming on agricultural farms, an analysis and prediction of agricultural output and precipitation will help illustrate the urgency for farmers to change their methods.

The purpose of this investigation is to determine the effect global warming has had and will have on precipitation. Consequentially, this investigation will then predict the agricultural output depending on the precipitation. To better observe the effect global warming has on these factors, the temperature will be used as a variable. Precipitation, not just average rainfall, is used to holistically measure the effects of global warming as water can manifest itself in diverse ways in a region. For example, it could appear as hydrological disasters (floods, tsunamis, etc.) that are not conducive to agriculture, and are a manifestation of global warming. Total Sales in thousands of USD is used because it captures the economic impact of global warming in a more accurate way than Total Sales in Bushels. Lastly, the years 1997, 2007, and 2017, are used to track how global warming changes these variables over time.

From the data collected and analyzed, a conclusion can be drawn regarding the economic makeup of the U.S. and actions that could be taken to change this makeup to sustain or reform the agricultural sector during times of environmental change.

Collected geospatial data & analysis

Data for this investigation was collected via various trusted sources after hours of research. The sources required a time range to allow for analysis over an interval of time.

This interval of time should be from a start date to an end date, with a midpoint date, all with an equal time gap between them. Six datasets in total have been collected from various sources, below. In this case, the dates were determined to be 1997, 2007, and 2017, due to the availability of data at these time periods and the quality and completeness of this data.

- National Centers for Environmental Information (NCEI), National Oceanic and Atmospheric Administration (NOAA)
  - Data from the NOAA measured annual precipitation values across various counties in the contiguous U.S.
  - For data cleaning, the "Rank" column was removed due to the arbitrary nature of the value compared to the much more quantitative and specific "Value" column, which stated annual rainfall up to four significant figures. Further columns, such as "1901-2000 Mean" and "Anomaly" were removed since our investigation does not cover this time period, only more recent periods past 1997.
- United States Department of Agriculture (USDA)
  - Data from the USDA measures total sales and production from agricultural operations, from the amount to monetary value in USD.
  - For data cleaning, most columns other than "State", "County", "County ANSI" and "Value" were removed due to information being extraneous (for example, "watershed_code" and "Domain", which are not relevant for the purposes of this investigation).

These datasets were then stitched together with county data with the GeoJSON format to allow for graphical representation on a contiguous US map.

The first step to predict was to understand the data through multiple visualizations. It is important to observe the effect of time on all 3 factors and also how they affect each other.

Univariate Precipitation Patterns

The data was first collected and cleaned and then the precipitation was mapped onto a histogram to see the changes in precipitation over time. The global mean sea level was added to observe the effect of global warming over time.

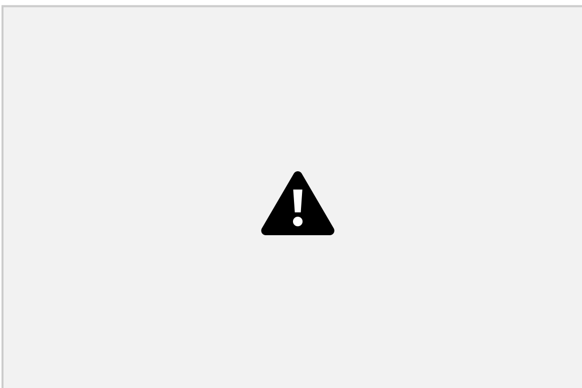

Figure 1*: Precipitation in 2017 Histogram*

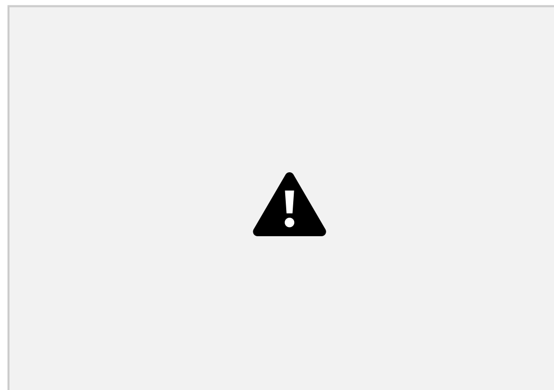

Figure 2: *Precipitation in 2007 Histogram*

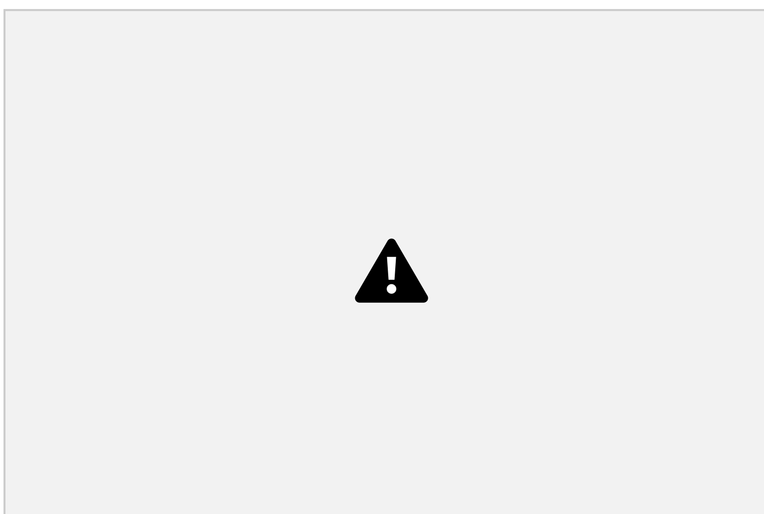

Figure 3: *Precipitation in 1997 Histogram*

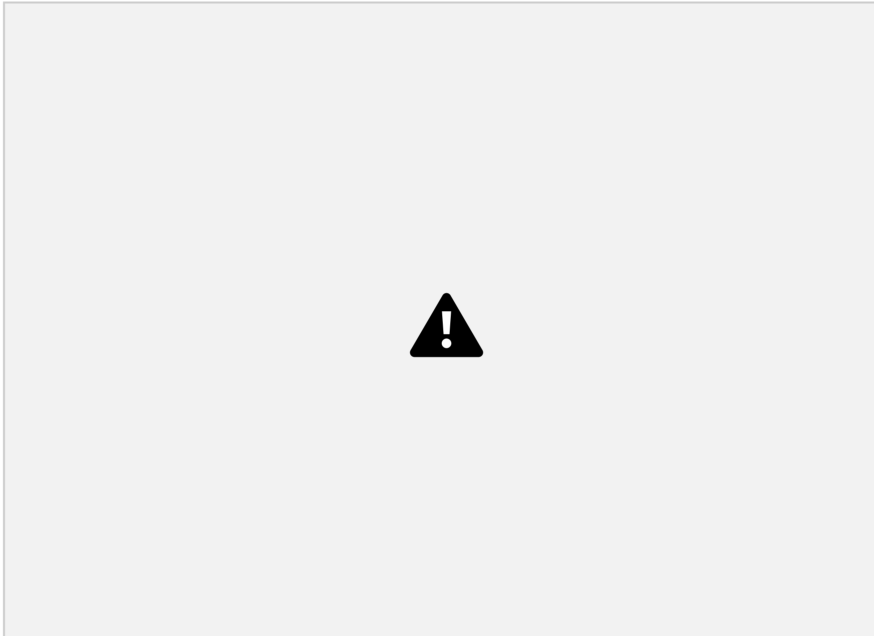

Figure 4: *Global mean sea level, time series data*

The data shows that the mode of the data is similar for 2007 and 2017 at around 42 inches per year. However, the mode for precipitation in 1997 is around 52 inches per year which means that global warming did affect the precipitation. Additionally, the outlier values of 1997 are greater, with the highest value being 150 inches per year, rather than 120 inches per year in 2007 and 2017. The global mean sea level supports this as there was a greater rate of change in sea level between the years 2007 and 2017, than between 1997 to 2007. Thus, the compounding effect of precipitation is seen by the decreasing precipitation.

Then, the data was mapped to see how the geographic location affected the precipitation.

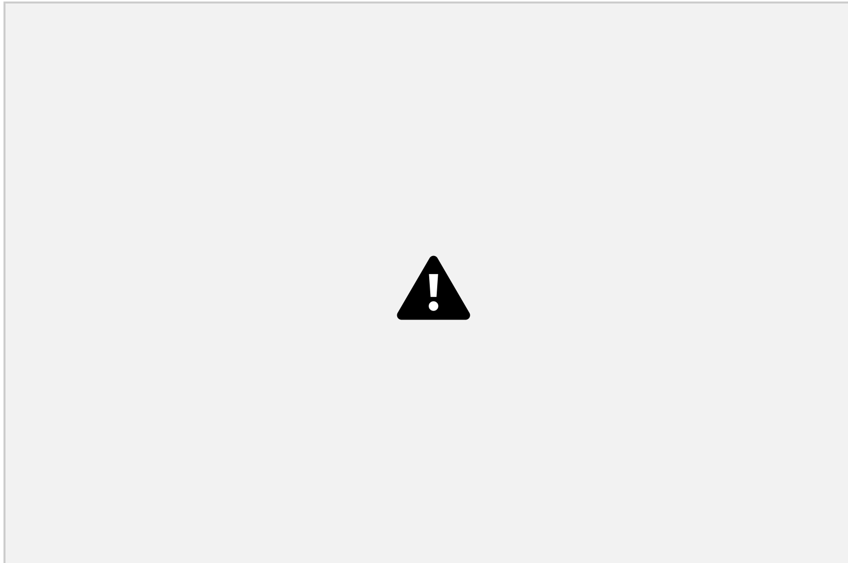

Figure 5*: Annual precipitation (inches) in 1997 around the contiguous United States*

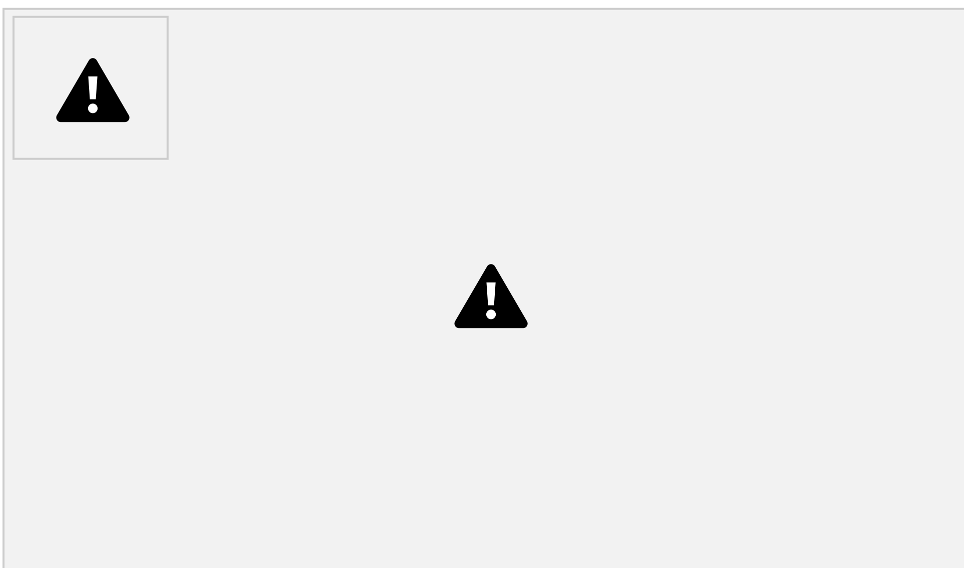

Figure 6: *Annual precipitation (inches) in 2007 around the contiguous United States*

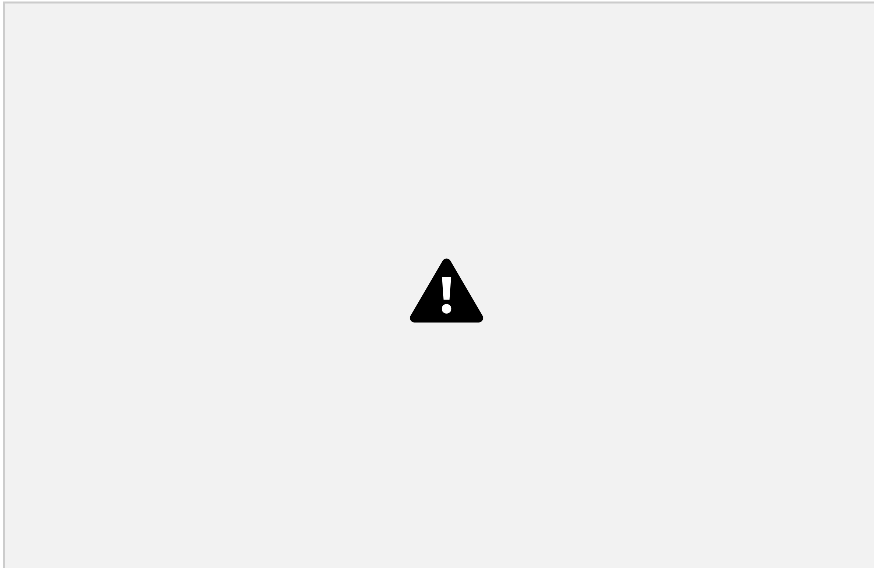

Figure 7*: Annual precipitation (inches) in 2017 around the contiguous United States*

The counties that were white in the maps did not have precipitation data available at a county level. Initially, it is important to acknowledge general precipitation trends across various regions in the U.S. In the coastal areas of the Pacific Northwest, it is notable that the amount of precipitation is much higher than in other regions in the United States. As the Pacific Northwest transitions into the West, precipitation suddenly becomes much more scarce, a pattern also shown in the Four Corners region and the Great Plains.

The Great Lakes, Northeast, and South regions all possess around median precipitation levels and also have much smaller county sizes. Notably, precipitation appears to be higher along the ridge of the Appalachian Mountains.

In 2007, there appears to be more rainfall than previously, with all regions having a higher blanket precipitation, likely due to the effects of climate change (warmer oceans lead to more evaporation, and thus more precipitation) and an increase in hydrological disasters (such as hurricanes, especially visible in the South).

However, in 2017, precipitation appears to decrease again, notably in the Pacific Northwest and the Midwest, as well as the Atlantic Coast, especially in North and South Carolina. In the Four Corners and Great Plains regions, precipitation appears to vary less between counties compared to 1997.

Univariate Agricultural Productivity Patterns

After exploring the precipitation data over time, the agricultural output over time was explored. The histogram is the total sales in thousands of US dollars.

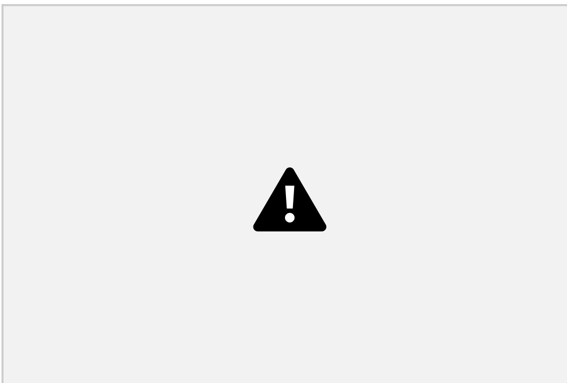

Figure 8: *Agricultural Output in 2017 Histogram*

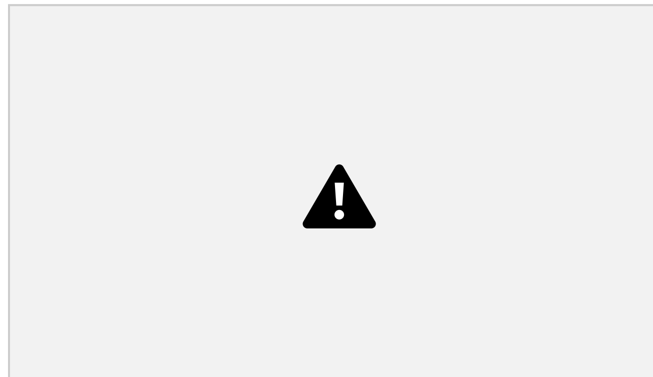

Figure 9: *Agricultural Output in 2007 Histogram*

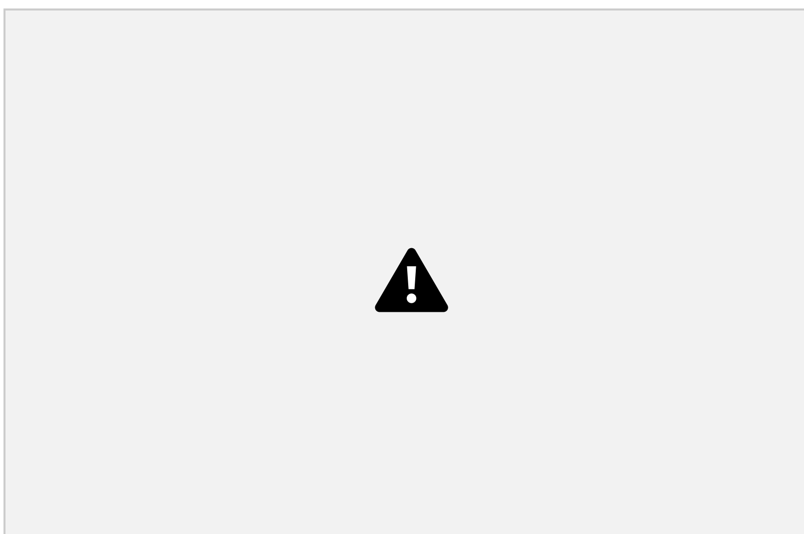

Figure 10: *Agricultural Output in 2017 Histogram*

Despite wildly varying precipitation changes over time, the agricultural output stays the same. The mean continues to be centered around 0, and the density is around 0.204. The outliers continue to be more than $1,000,000 per county per year.

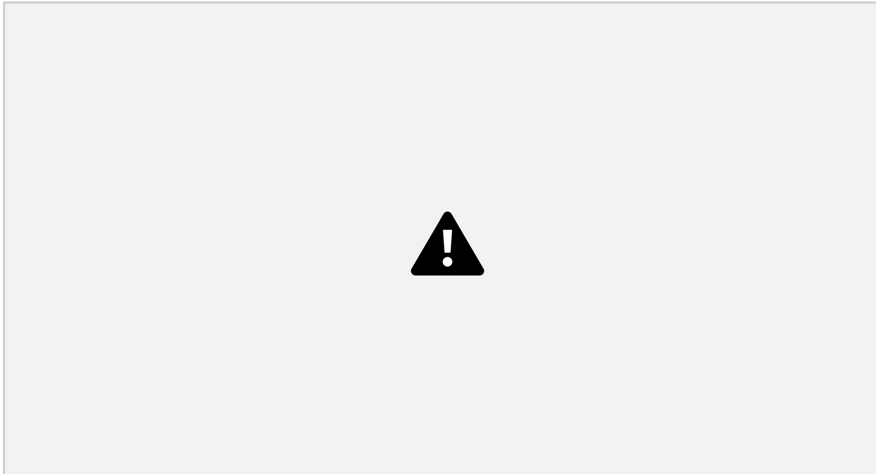

Figure 11: *Total sales from agricultural operations in 1997 in the contiguous United States*

The counties that were white in the maps did not have precipitation data available at a county level. Overall, across the U.S., agricultural sales appear to be maximized around the Great Plains, the Midwest, and some parts of the Pacific Northwest and West. For example, Fresno County in California, a consistent top performer in agriculture, can be seen with a high color density.

Out of the remaining regions, the Northeast and the South appear to be mediocre in terms of agricultural output, often having fewer counties with high agricultural sales. The Four Corners region performs the worst, having barely any counties with agricultural sales, the majority of which are in southern Arizona.

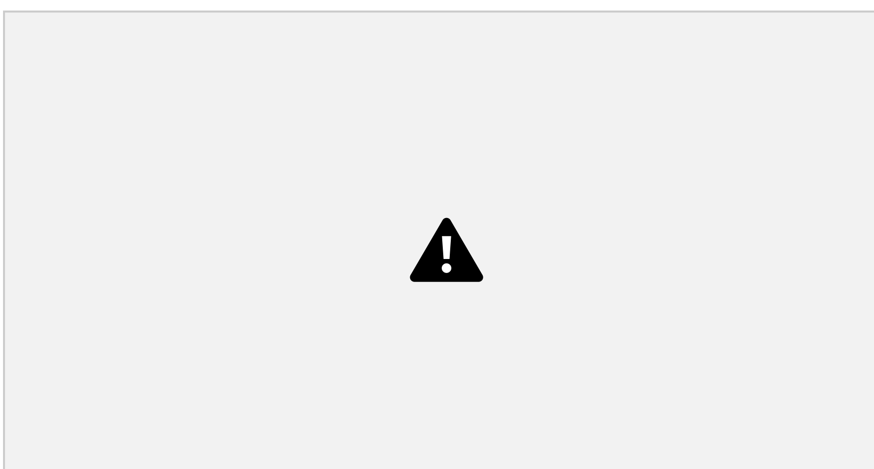

Figure 12: *Total sales from agricultural operations in 2007 in the contiguous United States*

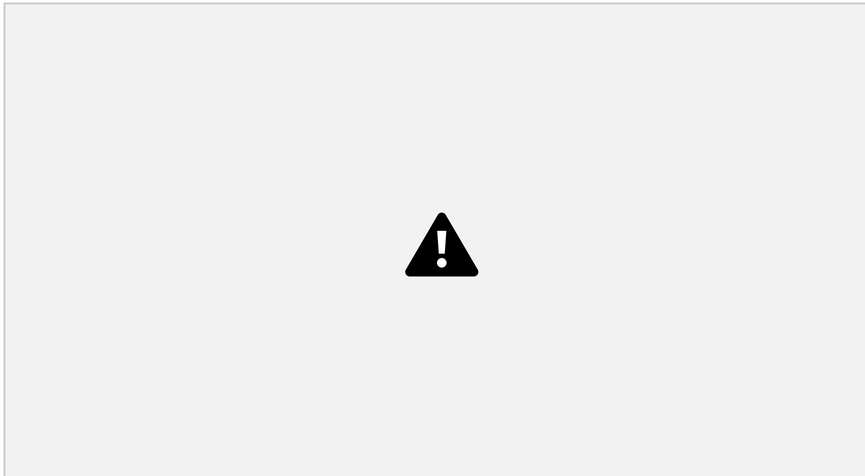

Figure 13: *Total sales from agricultural operations in 2017 in the contiguous United States*

Overall, across the U.S., agricultural sales appear to be maximized around the Great Plains, the Midwest, and some parts of the Pacific Northwest and West. For example, Fresno County in California, a consistent top performer in agriculture, can be seen.

Out of the remaining regions, the Northeast and the South appear to be mediocre in terms of agricultural output, often having fewer counties with high agricultural sales. The Four Corners region performs the worst, having barely any counties with agricultural sales, the majority of which are in southern Arizona.

In 2007, the trends seen from 1997 mainly continued, but agricultural production and sales appeared to slow down. Interestingly, the agricultural production of Navajo county (the strip of land in northern Arizona) appears to spike and retains this pattern through 2017.

In 2017, agricultural productivity in various areas such as northern Arizona and the Rocky Mountains increased. There are also increases in other areas of the United States.

Univariate Temperature Patterns

To have a tangible value, other than year, to measure the increase of global warming, the temperature was used. It was first plotted in histograms to understand how the temperature changed over the 3 years used in this study.

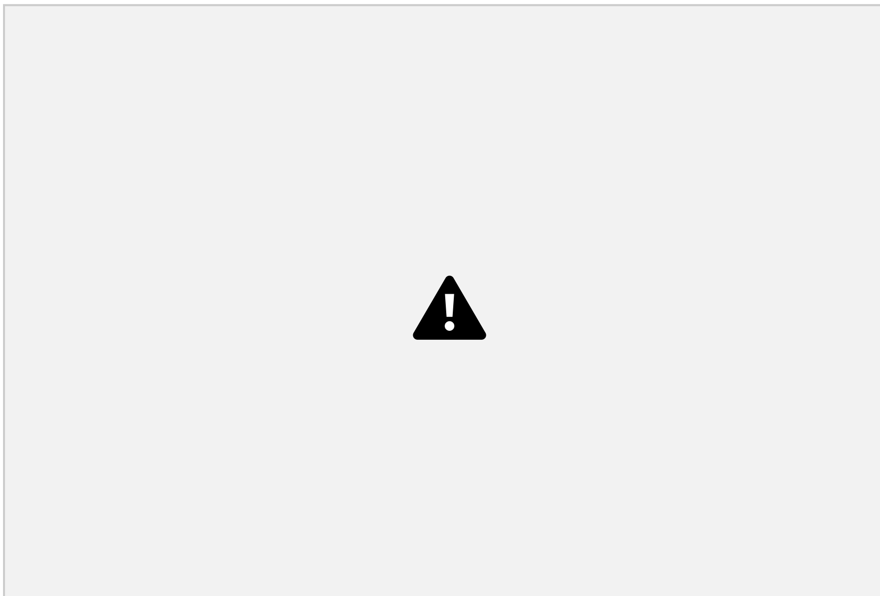

Figure 14*: Average Temperature in 1997 Histogram*

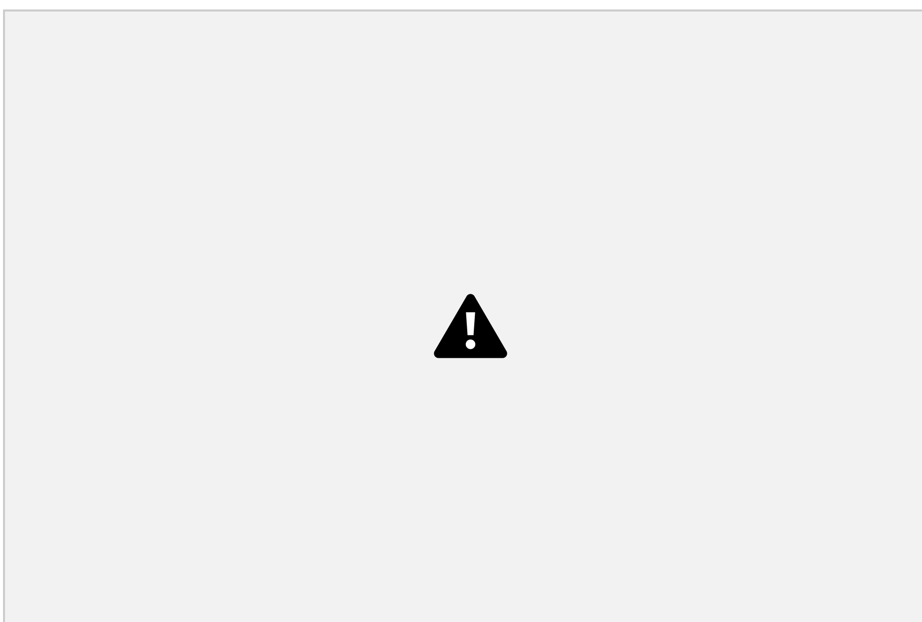

Figure 15*: Average Temperature in 2007 Histogram*

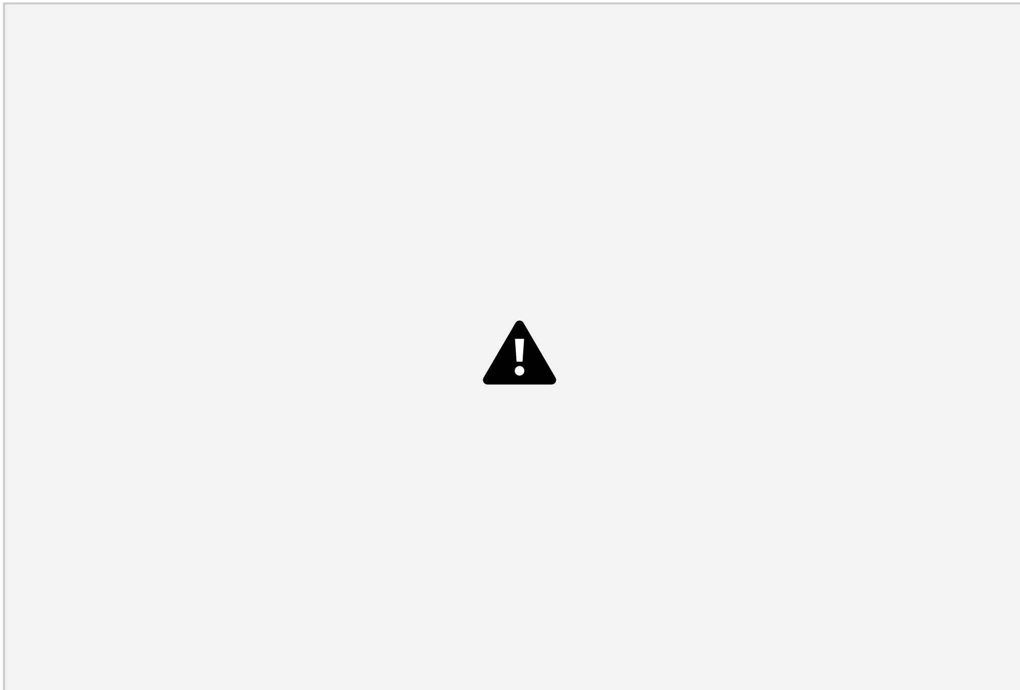

Figure 16: *Average Temperature in 2017 Histogram*

As expected, the mode of the histograms changed by about 2 degrees Fahrenheit as the 10 years passed. This follows the expectation of climatologists for the increase in temperature due to global warming and also accounts for the exponential growth of the temperature as the US continues to urbanize. (Lindsey). In 1997, 2007, and 2017, the temperature was about 51 degrees, 53 degrees, and 57 degrees Fahrenheit respectively.

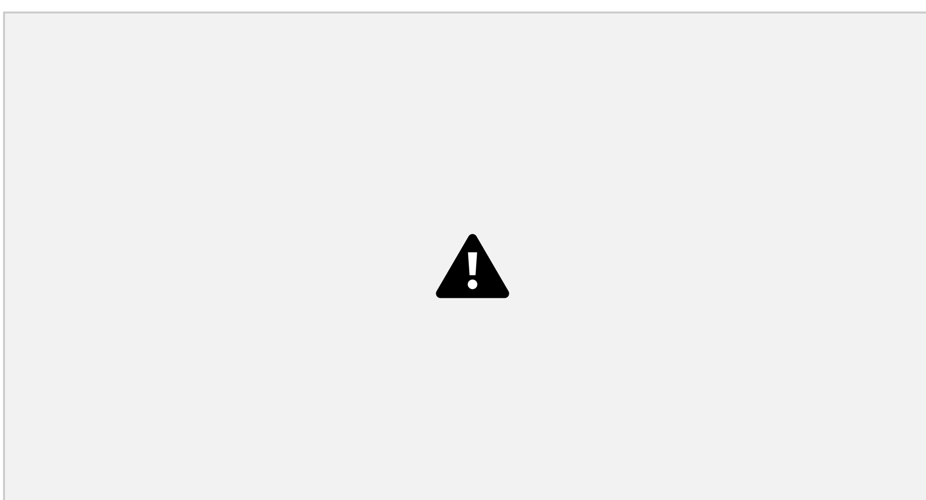

Figure 17: *Average temperature in each county around the U.S. in 1997*

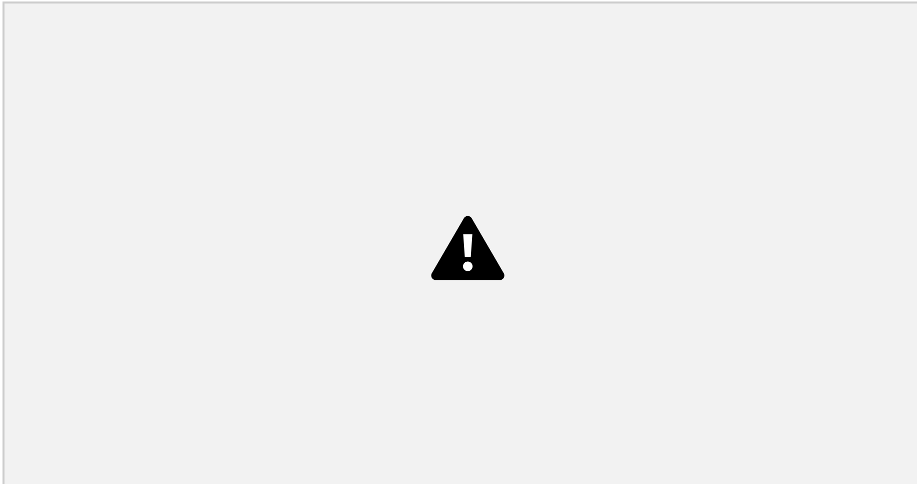

Figure 18: *Average temperature in each county around the U.S. in 2007*

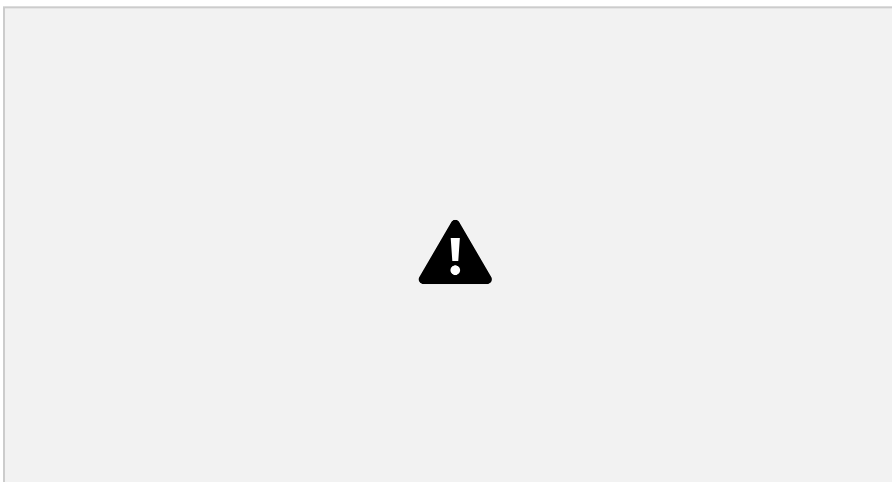

Figure 19: *Average temperature in each county around the U.S. in 2017*

The counties that were white in the maps did not have precipitation data available at a county level. These subtle shifts are also seen in the mapping graphs. From 1997 to 2017, there were small visible increases in temperature across the U.S. as a result of global warming. In the Southern part of the United States, the brown seems to spread cautiously upward as the decades go by. Counties in California and Oregon, which were initially counties that were around 50 degrees Fahrenheit in 1997, inches closer to 55 degrees in 2017.

Multivariate Precipitation, Temperature, and Agricultural Productivity Patterns

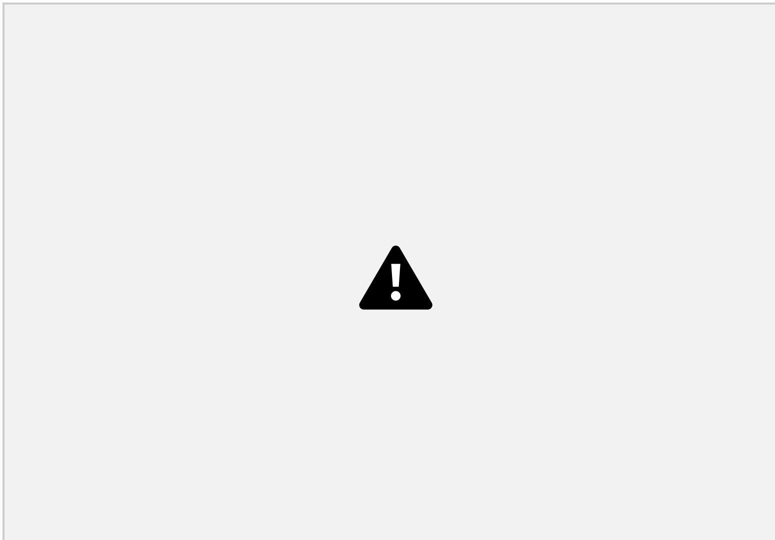

Figure 20: Correlation heatmap with the variables

"Precip" corresponds to total precipitation in inches. "STATE_NAME" corresponds to the postal number of the state, such as 1 for Alabama. "Total sales" corresponds to total sales in thousands of US Dollars of agricultural products. "Temp" corresponds to the average temperature in degrees Fahrenheit. "Year" corresponds to which year the data is representing, which can be either 2007, 1997, or 2017. Region corresponds to the following numbers with the following regions:

0. South
1. Rocky Mountains
2. Four Corners
3. Northeast
4. Pacific Northwest
5. Great Lakes
6. Great Plains

The heatmap shows the correlations between certain variables. Values close to 1 mean a high positive linear relationship, while close to -1 show a high negative relationship. Because this data is taken over such a large scale, spanning states, and years, the threshold for the value is lowered significantly. The temperature has a correlation of 0.32 with precipitation, indicating that temperature is an important variable to analyze for changes of precipitation. Since temperature is the de facto variable to determine global warming, global warming can be said to have a causal relationship with precipitation.

Temperature has a correlation of 0.19 with year, again signifying that these two variables provide a good substitute for global warming. Total sales have a correlation with temperature, at 0.036. Thus, in a general sense, when temperature increases, total sales increase. Since average temperature is used, areas that have extreme temperatures may still have low agriculture, causing the correlation to be low. Though total sales and precipitation have a weak correlation, the positive correlation suggests that as precipitation increases, total sales increase. The reason the relationship is not extremely strong is that the geographical location is not considered in this correlation. Thus, areas that have high precipitation but poor growing conditions, such as the Pacific Northwest.

After analyzing the univariate patterns in precipitation and agricultural productivity, the relationship and the impact between those will be analyzed here. A closer look at how geolocation plays a role in agriculture and total sales changed over time.

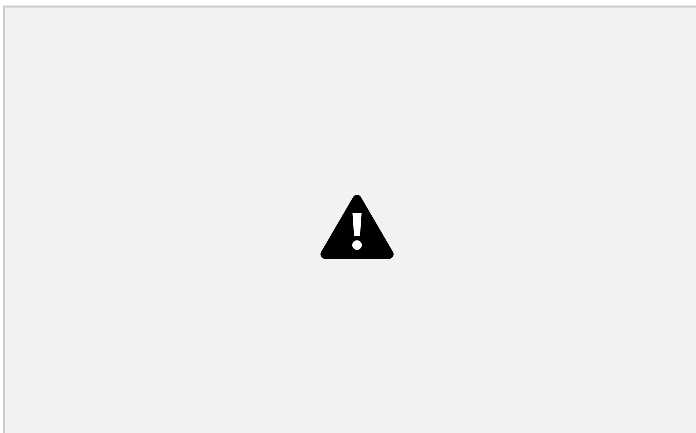

Figure 21: *South precipitation versus total sales scatterplot*

In the South, agricultural productivity appears to have decreased over time, with data points having lower total sales the later the year. This is a pattern measured in the other regions as well, due to the increasing industrialization of the U.S.

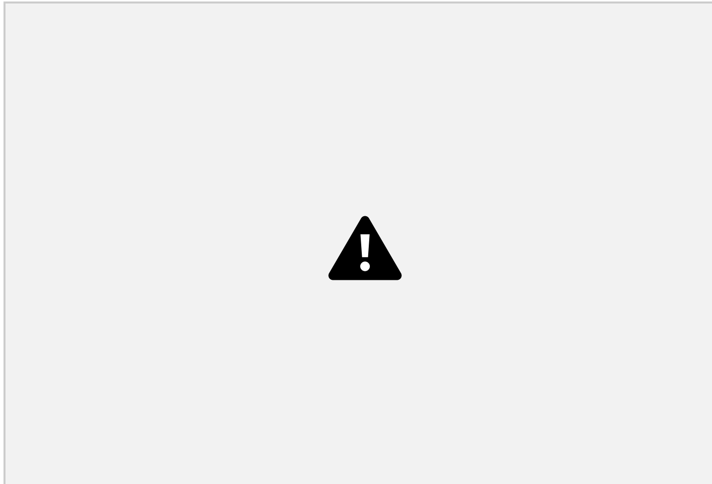

Figure 22: *Rocky Mountains precipitation versus total sales scatterplot*

In the Rocky Mountains, agricultural productivity was generally higher in 1997 due to the many outliers. However, it remained mostly steady, regardless of the precipitation.

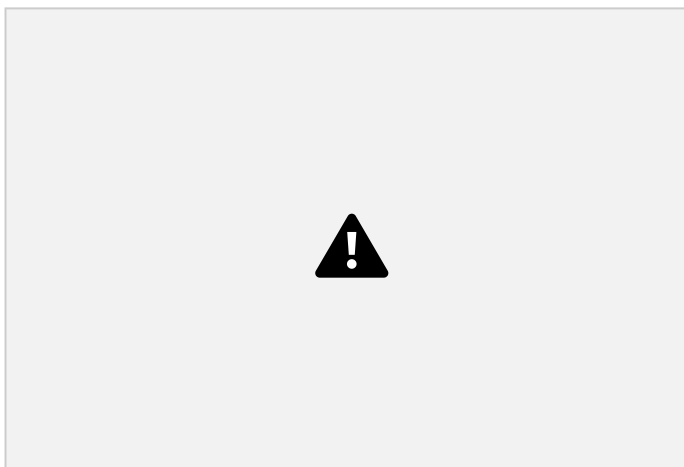

Figure 22*: Four Corners precipitation versus total sales scatterplot*

Interestingly, in the Four Corners region, agricultural productivity appears to remain relatively the same, likely since much of the region is desert. From the beginning, the Four Corners region lacked agricultural potential, so changes in industrialization did not impact it as heavily.

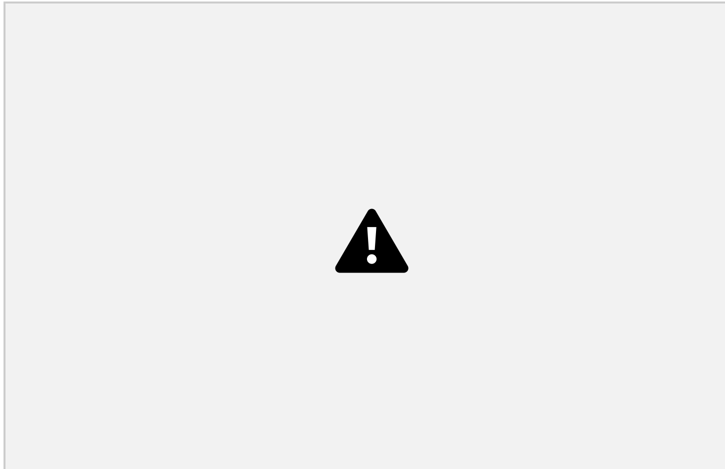

Figure 23: *Northeast precipitation versus total sales scatterplot*

The Northeast region does have a weak trend that as precipitation increases, total sales increase. There are also many outliers in the graph, which shows that this region isn't as heavily affected by precipitation as other regions.

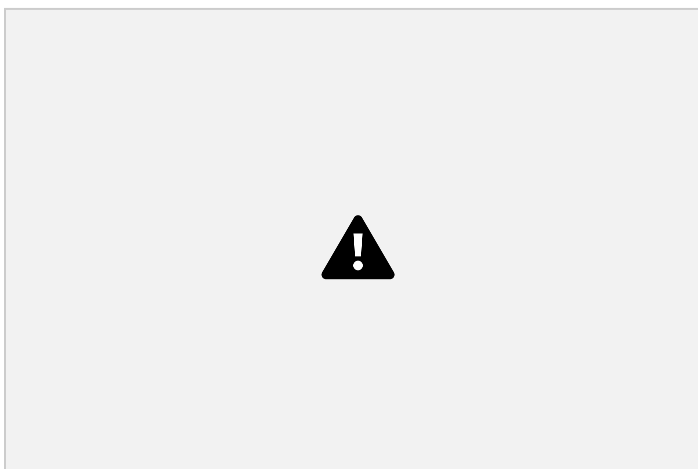

Figure 24: *Pacific Northwest precipitation versus total sales scatterplot*

The lack of data points in the Pacific Northwest region is notable since it reflects the circumstances in that area that prevent the extensive agricultural practice from occurring. For one, the terrain is either very mountainous, and thus inconvenient to farm on, or near the coastline, where it could be affected by floods or tides. In addition, much of the Pacific Northwest is covered in federal lands, such as national forests, excluding the possibility of agricultural activity. Thus, the scale of the total sales is only up to $800, rather than the $1000 in the other figures.

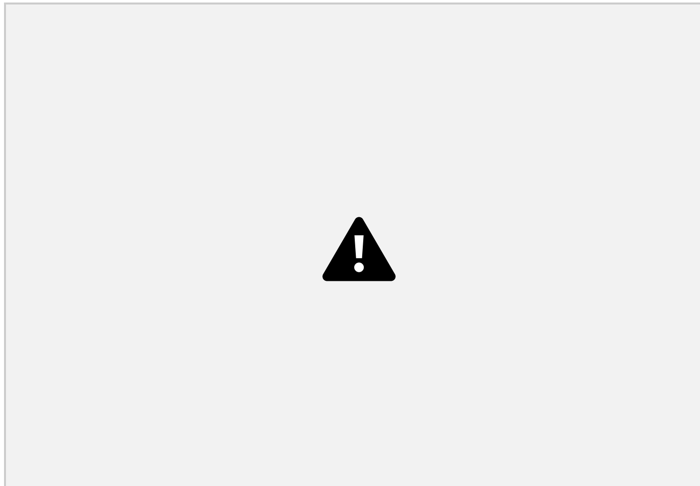

Figure 25: *West precipitation versus total sales scatterplot*

There is a weak trend with precipitation and total sales in the West. In 1997, higher sales were found in counties that didn't have much precipitation. However, it dwindled as the years went by and areas of more precipitation were able to produce more sales.

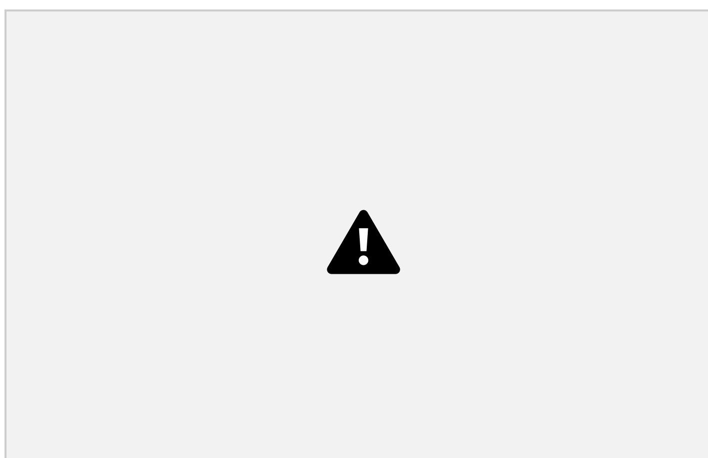

Figure 26: *Great Lakes precipitation versus total sales scatterplot*

The Great Lakes had very quantized precipitation, not straying from the integer values. This also causes no clear trend to be found. As precipitation increases, total sales remain more or less the same.

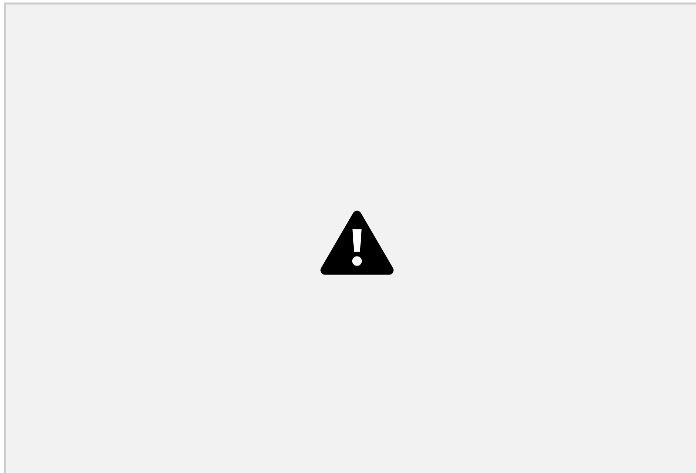

Figure 27: *Great Plains precipitation versus total sales scatterplot*

In the Great Plains, there is a clear trend that as precipitation increases, total sales increase. This is supported by the fact that this region experiences many droughts. Thus, precipitation is required for agricultural production.

After examining the regions at a smaller scale, general trends were examined.

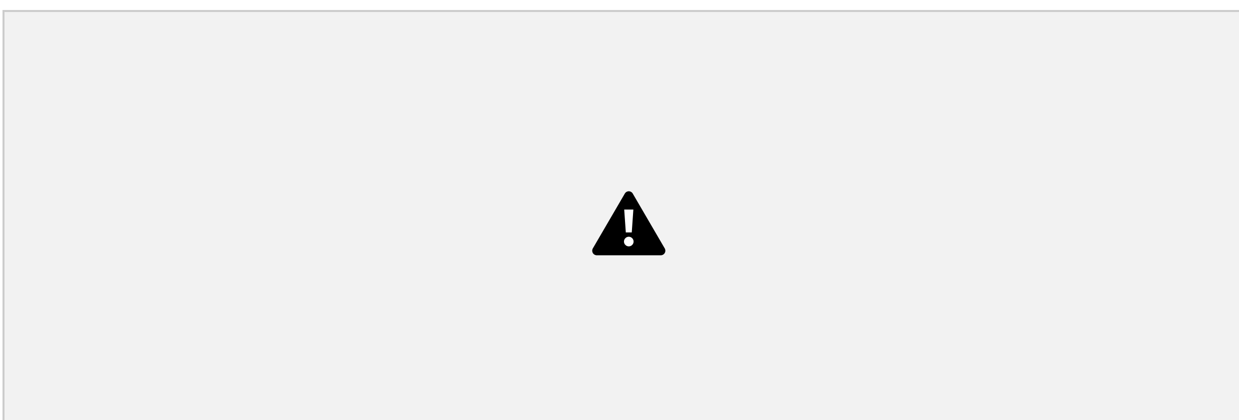

Figure 28: Catplot with temperature to total sales

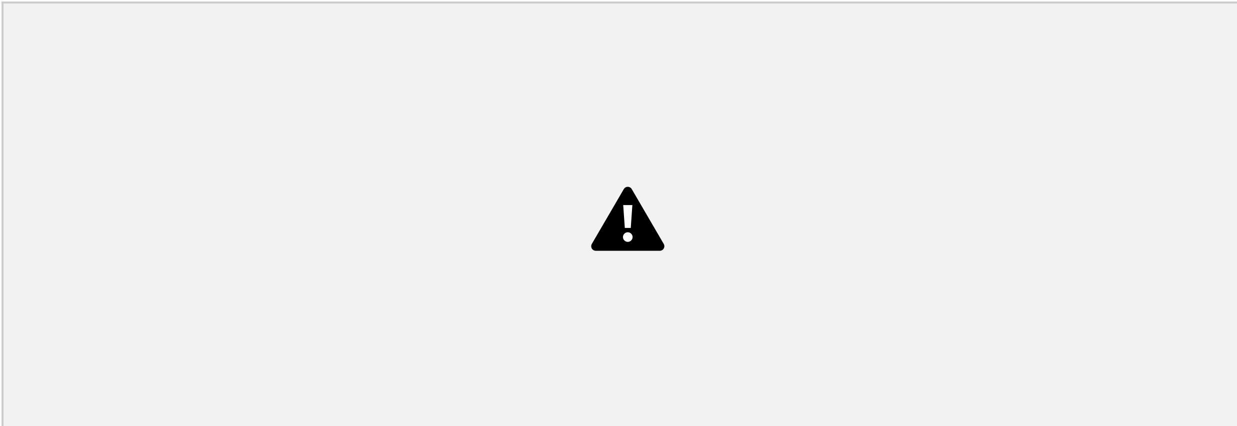

Figure 29: Catplot with precipitation to total sales

For the most part, annual precipitation is directly related to agricultural sales. This is especially visible in the comparison between the Four Corners regions and the Rocky Mountains. The Midwest also visibly has high agricultural productivity compared to the more arid regions around it such as the Great Plains. The Northeast and the South are at median precipitation and agricultural sales.

However, it appears that the counties with highest agricultural sales are those with a high, but not exceedingly high, precipitation. This is likely due to the fact that areas with high precipitation are either susceptible to floods or other hydrological disasters, or often receive snowfall, which is not conducive towards agricultural practices.

In 2017, the trend remained similar, although agricultural sales in areas such as the South and the Great Plains decreased despite precipitation remaining very similar. Some areas such as the Northeast and northern Arizona also experienced increased agricultural sales.

Prediction

After initial exploration, I decided which methods would be the best to use to predict both precipitation and agriculture. This was done in two parts. First, the precipitation was predicted, using 3 counties in each of the areas. The data used was a time series type of the average precipitation of three random counties per geographic region from 1987 to 2022.

An autoregressive integrated moving average model was used to predict the precipitation. The actual values (blue) are graphed under the predicted values (red) for each of the geographic regions.

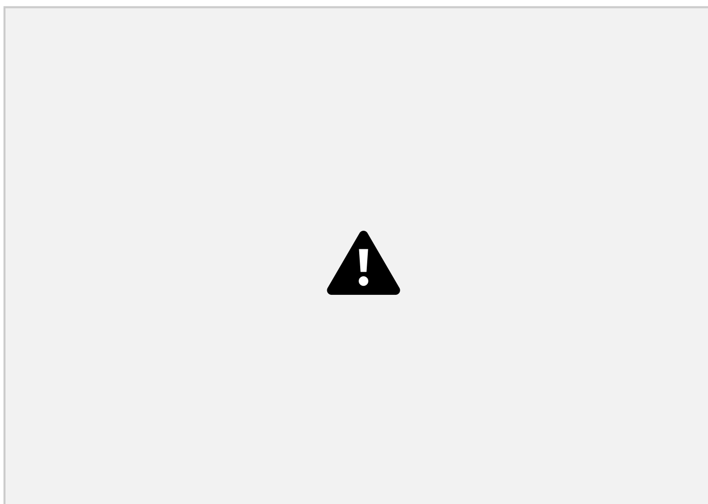

Figure 1: Four Corners Prediction

The precipitation values for the regions are as follows:

Region	**Rank (Precipitation prediction in inches per year)**
PNW	85.046282
Northeast	45.456955
Great Lakes	33.977177
South	33.138507
Great Plains	27.326068
Four Corners	26.514427

Rocky Mountains	25.440934
West	21.167662

Table 1: Average Precipitation Prediction Per Geographic Region In Inches Per Year

Then, Machine Learning models for agriculture output prediction were used. There were multiple regression models to choose from, but they were evaluated using Root Mean Squared Error (RMSE) because RMSE doesn't have errors of large magnitude. This is useful for this problem because there are many counties who are significantly above the average of agricultural output and the model shouldn't penalize it. After evaluating models, the average of XGboost and BayesianRidge predictions were used:

Geographic Region	**Agriculture value (in 1000 USD/ year)**
Four Corners	109.0095776
West	87.6034522
RM	78.29915819
GL	77.87831323
PNW	76.7444275
GP	64.65274827
South	63.85172074
Northeast	57.14519799

Table 2: Average Agricultural Output Prediction Per Geographic Region In $1000 Per Year

Reflection

Location Factors

There were some location factors involved in this investigation. For one, imperial units are used for many variables, notably temperature, which is measured in Fahrenheit, and depth/rain, which is measured in inches. Although this is the standard in the United States and some other countries, the data would need to be adjusted for countries that use the metric system.

Agricultural sales were also measured in USD, as opposed to other forms of currency in other countries. This is not a large problem, as the investigation focuses on solely the United States anyway.

Lastly, the agricultural sales data was relatively free from state and county-specific location factors, such as taxes. However, there may still be some variation in how counties and states measure agricultural production, which could affect the investigation.

Resources

"Agriculture and Climate | US EPA." EPA, 17 January 2023, https://www.epa.gov/agriculture/agriculture-and-climate. Accessed 15 March 2023.

"Agriculture's Contribution to Climate Change and Role in Mitigation Is Distinct From Predominantly Fossil CO2-Emitting Sectors." Frontiers, 14 December 2020, https://www.frontiersin.org/articles/10.3389/fsufs.2020.518039/full. Accessed 1 February 2023.

"Climate Change and Agriculture." Union of Concerned Scientists, 20 March 2019, https://www.ucsusa.org/resources/climate-change-and-agriculture. Accessed 3 February 2023.

"Climate Change." USDA ERS, 10 June 2022, https://www.ers.usda.gov/topics/natural-resources-environment/climate-change/. Accessed 18 February 2023.

"Climate Change Impacts on Agriculture and Food Supply | US EPA." EPA, 13 December 2022, https://www.epa.gov/climateimpacts/climate-change-impacts-agriculture-and-food-supply. Accessed 21 February 2023.

"Climate-Smart Agriculture." World Bank, 5 April 2021, https://www.worldbank.org/en/topic/climate-smart-agriculture. Accessed 19 February 2023.

"Eight warmest years on record witness upsurge in climate change impacts." *World Meteorological Organization* |, 6 November 2022, https://public.wmo.int/en/media/press-release/eight-warmest-years-record-witness-upsurge-climate-change-impacts. Accessed 14 February 2023.

Lindsey, Rebecca, and LuAnn Dahlman. "Climate Change: Global Temperature | NOAA Climate.gov." *Climate.gov*, 18 January 2023, https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature. Accessed 23 February 2023.

Nuwer, Rachel. "Here's how climate change will affect what you eat." *BBC*, 28 December 2015, https://www.bbc.com/future/article/20151228-heres-how-climate-change-will-affect-what-you-eat. Accessed 8 February 2023.

Smith, Issac. "3 Best metrics to evaluate Regression Model? | by Songhao Wu." Towards Data Science, 23 May 2020, https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b. Accessed 20 March 2023.

"USDA - National Agricultural Statistics Service - Quick Stats." *USDA National Agricultural Statistics Service*, 30 April 2019, https://www.nass.usda.gov/Quick_Stats/index.php. Accessed 19 February 2023.