Global Warming, Always a Hot Topic

An Analysis of Crop Yield in Relation to Temperatures

Plant Friends: Dante Rousseve, Sindura Sriram, Huiyuan Wu, Zihan Hu

Hypothesis

Given the unnatural rise of temperatures due to global warming, the climates of regions across the world are changing over time. Our hypothesis is that crop yield will be inversely correlated with this positive temperature change, because plants will not be able to flourish as the climates they are native to or cultivated in alter rapidly.

Data Collection

We collected data from two sources:

- The Global Dataset of Historical Yield from PANGEA
- The Global Land-Ocean Temperature Index from NASA

Both datasets were uniformly distributed across the globe in 0.5° increments for the years from 1981 to 2014. We used bounding box analyses to narrow our dataset to the contiguous U.S. and identify the corresponding state of each coordinate. The output of this process was 248 tables representing the temperature data and yield data for each year across the four crops. We performed joins across those tables to create our final dataset, which consists of 4 tables, one for each major crop in our analysis, each with ~4000 data points.

For some states, no data is present. Alaska and Hawaii were intentionally left out of the data to focus on the contiguous U.S. Vermont, West Virginia, Nevada, Wisconsin, and Rhode Island were subject to bounding box overlap, and their data was assigned to neighboring states. Since we go through the states alphabetically, and some states' bounding boxes overlap other states, larger states sometimes encompassed smaller states' data points.

Background

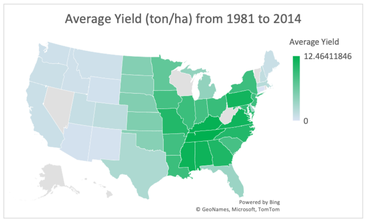

Figure II. Average Yield of Maize, the crop most widely grown in the U.S., from 1981 to 2014, state-by-state. The average change in yield from 1981 to 2014 was -0.021 ton/ha.

Our initial intuition was that as temperature increased, crop yield would decrease. We came to this conclusion based on our understanding of plant life's ability to thrive in changing climates.

However, either the temperature variations within the U.S. are not robust enough to display this relationship, or the relationship between yield and temperature is actually positive.

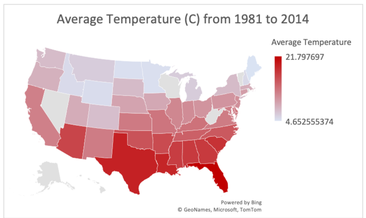

Figure III. Average Temperature from 1981 to 2014, state-by-state. The average change in temperature from 1981 to 2014 was 0.165° Celsius.

U.S. Analysis

Our initial analysis of the data across the entire U.S. provided little insight into any correlation between temperature and yield. This is the result of the vast temperature differences present across the states. Looking for a trend across all fifty states led to an inaccurate model.

Crop	Mean of Test Residuals	One Sample t-statistic	One Sample p-value
Wheat	7.3773	181.1982	0.0
Maize	4.5373	111.9525	0.0
Rice	9.2234	117.9709	0.0
Soybeans	9.5427	211.6966	0.0

Figure IV. T-test results on regression predicted values on the U.S. dataset.

We verified this inaccuracy by performing statistical tests on our regression's predictions on the test dataset.

Comparing the mean of the error values to zero, one-sample t-tests across all four crops resulted in p-values of zero, so we cannot reject the null hypothesis, meaning the average error is not statistically equal to zero.

Limitations

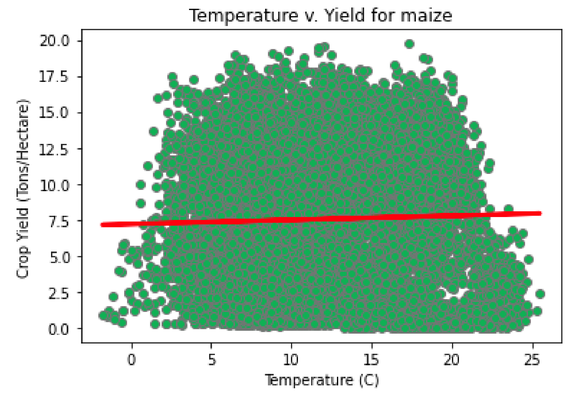

Figure VII. Country-Level Regression across the maize dataset.

The data across the U.S. was distributed widely across both temperature and crop yield. This led to key issues with the linear regression, and created similar issues and challenges when a quadratic or cubic polynomial regression was used to fit the data.

The wide distribution across both temperature and yield is the result of the U.S. having many different climates and geographies within its borders. This variation was somewhat controlled for when we shifted to a state-by-state analysis.

In the state-by-state data, the key limitation was that not every state grew all four crops we wanted to analyze. This led to some state datasets having little or no data to analyze, model, or predict with.

Methodology

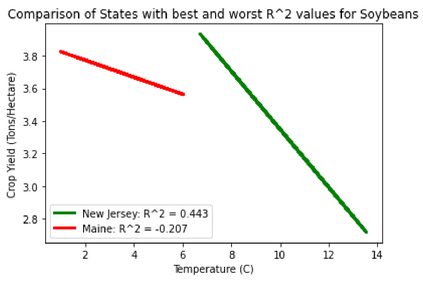

Figure I. Comparison of outlying R^2 values across Soybean dataset.

Our analysis was initially focused on the U.S. as a whole. We analyzed the relationship between temperature and crop yield using data points from across the entire U.S. The regression across this data was inaccurate, so we decided to analyze the data on a state-by-state basis.

Our second analysis focused on the state level, creating and fitting a regression model for each state. We clustered the data by state, then ran our regressions on those datasets. These regression models were more accurate, but still not accurate enough to reject the null hypothesis.

To further explore the models, we analyzed the R^2 values of all 43 models, across all 4 crops to understand where the model was lacking.

State Level Analysis

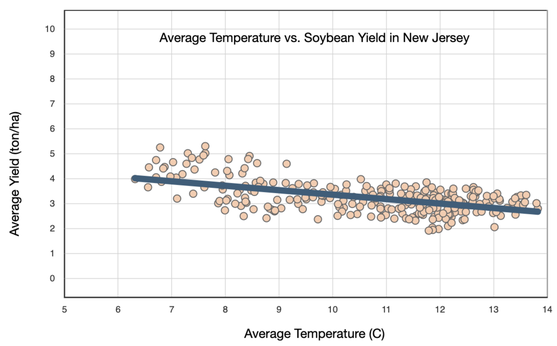

Figure V: State-Level Regression with the maximum R^2 value, 0.44.

Our average R^2 score for all of the states was 0.08 for maize, 0.09 for wheat, 0.04 for rice, and 0.06 for soybean, which we did not find to be statistically different from our R^2 scores of 0.0 for our analysis over all US data. We thought it might be productive to consider the state yielding the maximum R^2 score for each crop.

Doing so showed that our results were still not statistically significant for any individual state: our maximum R^2 scores are 0.39 for maize in the state of New Jersey, 0.36 for wheat in the state of Mississippi, 0.29 for rice in the state of Texas, and 0.44 for soybeans, again in the state of New Jersey.

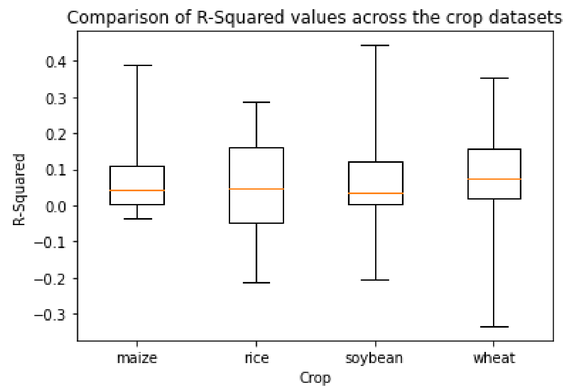

Figure VI. Box-and-Whiskers plot showing variance of R^2 values.

To further illustrate the inconsistencies in our R^2 values when modeling state by state, we have included the following plot of the range of our R^2 values for each crop. As can be seen, the mean R^2 values are near 0.0, and R^2 values range from the maximum values stated left to negative values of similar magnitude.

Our best indicator of correlation, the highest R^2 value of 0.44 for soybeans grown in New Jersey, suggests a negative correlation between temperature and crop yield. However, the highest R^2 value for a different crop, wheat in Mississippi, suggests a positive correlation between temperature and crop yield.

Conclusions

Our initial intuition was to perform our crop yield vs. temperature regression across all US data, and after seeing that the data was so disparate that there was no way of generalizing a trend from it, we wanted to show this result and compare it to the results we could achieve by doing this analysis state by state.

However, we still did not find any consistent results in our state-wise analysis. From the inconsistent and insignificant values from our statistical tests, we are not able to reject our null hypothesis and confirm that there is a relationship between crop yield and average temperature for any of the four crops.

As some of our results directly contradict our initial hypothesis (shown in our state-level analysis), we believe that this points to external factors, like improvements in technology, that we may not be considering in our model.

To consider these external factors, we would continue to develop the project by analyzing the data across shorter timescales (to account for the improvement in farming technology across the 33 years of data) and by analyzing more data from regions all around the world that have similar climates (to account for the large temperature disparities present within our current datasets).