# Multi-Scale Weather Patterns for Crop Yield Potential

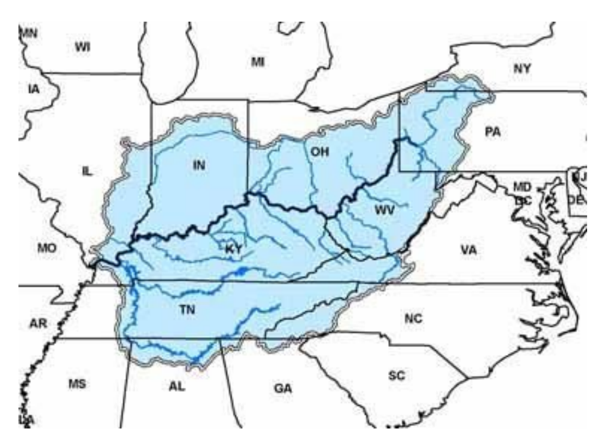

**Team Grain Gang**

Mahita Mudunuri, Krystian Misiewicz, Sri Bhamidipati, Mitchell Jones



## Abstract

- The purpose of this project was to benchmark crop yield potential for corn in the Central Ohio Valley region in the United States using various biophysical parameters and weather patterns
- Research Question 1: What is the variation in precipitation and temperature between states? Is there a pattern for temperature and precipitation?
- Research Question 2: What is the relationship between precipitation and temperature in the Ohio Valley region? What is the climate of the Ohio Valley region?
- Research Question 3: Do temperature and precipitation play a role in crop yield based on the questions above?

## Introduction

- Convective storm systems during the growing seasons increase difficulty in forecasting patterns.
- Precipitation is the driver for the biophysical system in Ohio Valley region's rainfed agriculture.
- Multi-scale weather patterns, which include biophysical parameters such as precipitation, surface temperature, humidity, soil moisture, and wind speed, affect crop yield tremendously
- Determining the extent of this influence and being able to predict future crop yields is an important aspect of data-driven agricultural company like John Deere.
- Corn consumes 3.5 million acres across 25,000 farms in the Ohio Valley region. The 2015 harvest was valued at $1.9 billion, making it a vital part of this region's economy

## Hypothesis

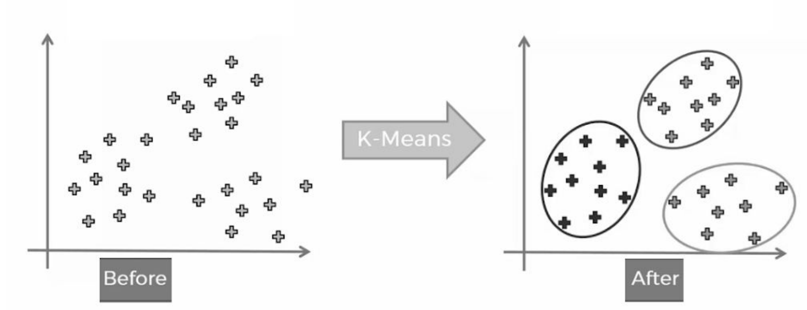
- We hypothesized that increasing changes in precipitation and air temperature over 2014-2018 have slowed the growth of the corn yield compared to 1990-1994 in Ohio Valley Central. K-means clustering was used for analyzing how the crop yields are affected by temperature and precipitation in different states.

## Methods

Clustering

**K-Means Cluster Analysis:**

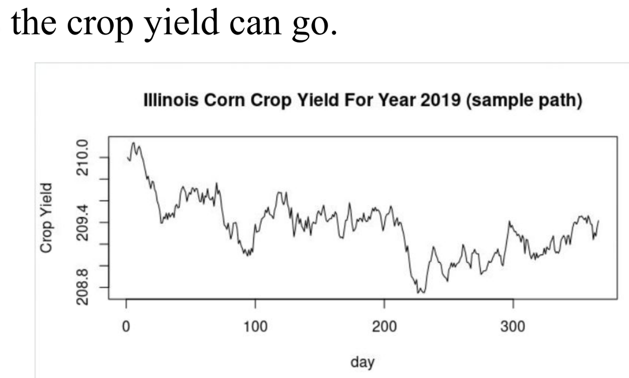
To find the subgroups of States that have similar climates (precipitation and temperature), we will use k-means cluster analysis in R to determine the number of subgroups and which States belong to which subgroup that minimizes variation in the precipitation and temperature.




Prediction

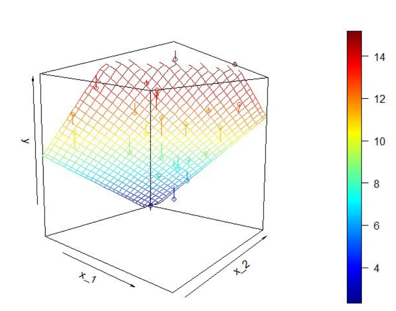
**Monte Carlo Simulation (Brownian Motion):**

To predict the upcoming year's crop yield for each state, we will use the Monte Carlo approach in R to create simulations for different runs. This Monte Carlo approach uses Brownian Motion (random stepping) to create a map of different paths the crop yield can go.




**Multiple Linear Regression:**

To predict the upcoming years crop yield for each state, we will use multiple linear regression in R to outline a linear relationship between precipitation and temperature, and crop yield. This approach uses least squares mean to minimize the squared residuals in the linear relationship



## Results

| Year | State | Crop Yield | Min Pre | Max Pre | Avg Pre | Min Temp | Max Temp | Avg Temp |
|---|---|---|---|---|---|---|---|---|
| 2014 | ILLINOIS | 200 | 0 | 55.87 | 41.18378981 | 53.2 | 66.5 | 48.65794393 |
| 2015 | ILLINOIS | 175 | 0 | 63.57 | 46.93318792 | 34.6 | 49.5 | 61.95238095 |
| 2016 | ILLINOIS | 197 | 22.7 | 60.46 | 39.6866129 | 37.1 | 51.3 | 63.5115384[illegible] |
| 2017 | ILLINOIS | 201 | 23.24 | 57.45 | 39.44387195 | 38.2 | 48.5 | 63.82952381 |
| 2018 | ILLINOIS | 210 | 27.95 | 66.2 | 45.89868902 | 35.8 | 49.6 | 61.39732143 |
| 2014 | INDIANA | 188 | 30.34 | 59.81 | 43.81726027 | 34.6 | 65.4 | 49.43561644 |
| 2015 | INDIANA | 150 | 31.08 | 67.44 | 47.09846512 | 37.7 | 68.2 | 52.07820513 |
| 2016 | INDIANA | 173 | 30.87 | 62.18 | 43.86247706 | 39.8 | 49.8 | 64.06666667 |
| 2017 | INDIANA | 180 | 32.01 | 65.495 | 46.49692346 | 39.1 | 50.1 | 63.06096492 |
| 2018 | INDIANA | 189 | 33.15 | 68.81 | 49.13136986 | 38.4 | 50.4 | 62.05526316 |

Figure 1: The Data table displays the calculated values of the minimum, maximum and average precipitation and temperature values in the states of ILLINOIS and INDIANA through the years 2014-2018.

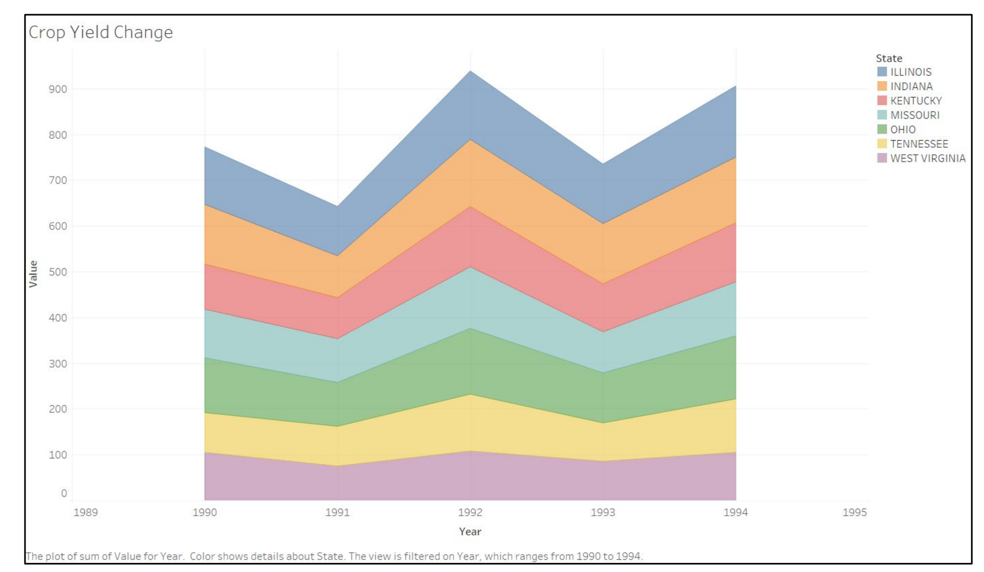
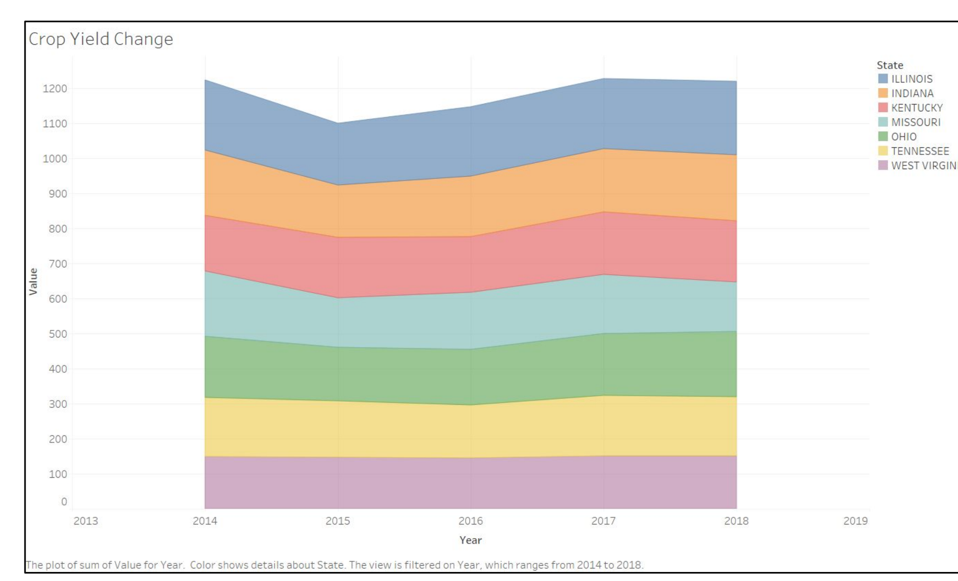



Figure 2: Change in the sum value of the corn crop yield over the years 1990 - 1994.



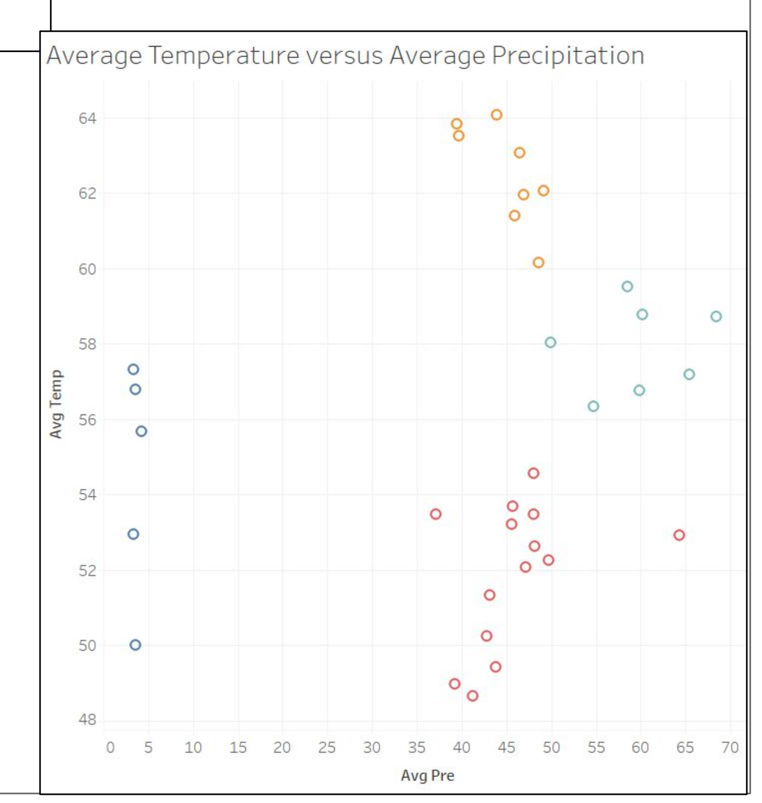

**Exploratory Data Analysis:**

- As seen in the results, there were drastic changes across the years, with the year 1992 having the maximum crop yield for all the states.
- Change in crop yield is mild in recent years ( 2014 - 2018) and can be attributed to similar precipitation and temperatures over the years. The overall sum value of crop yield is higher in recent years.

Figure 3: Change in the sum value of the corn crop yield over the years 2014 - 2018..

**K-Means Clustering:**

- To find the clusters of States that have similar climates, we will use k-means clustering in Tableau.
- Tableau sums up the the Euclidean distances between the data points and calculates the average silhouettes to find the optimal number of clusters.
- As a result, we were able to find these clusters that minimize the variation in precipitation and temperature: Illinois-Indiana, Kentucky-Tennessee, Ohio-West Virginia, and Missouri.




## Discussion & Analysis

- Based on our results from initial data exploration and the various methods used, we were able to create specific models pertaining to our Ohio Valley crop and weather data
- The K- Means method showed that there were certain clusters of states that had similar patterns of crop yield throughout the years: Illinois & Indiana, Kentucky & Tennessee, Ohio & West Virginia, and Missouri
- Based on the Monte Carlo Simulation, we were able to predict that Illinois would have the greatest crop yield.
- Though the models were not the most accurate, it gave a glimpse of how corn yield was being affected over the years along with weather changes, which ties into our hypothesis.

| State | Prediction | Actual | Percentage Error |
|---|---|---|---|
| Illinois | 213.4497 | 181 | 17.92801105 |
| Indiana | 190.9699 | 169 | 12.99994083 |
| Missouri | 132.2199 | 155 | -14.69683871 |
| Kentucky | 180.056 | 169 | 6.542011834 |
| Tennessee | 168.4651 | 177 | -4.821977401 |
| Ohio | 190.6555 | 164 | 16.25335366 |
| West Virginia | 152.8096 | 165 | -7.388121212 |

## Conclusions & Next Steps

- Our models and analyses showed that throughout the years of 2014-2018, the corn crop yield had constant growth and not as many drastic changes in comparison to 1990-1994 years.
- Evaluating the impact of this is advantageous to John Deere because if they know how much crop to expect, they can better make decisions about machinery, equipment, etc., which will aid in business operations.
- **Future Steps:**
- Use neural networks (machine learning) for predictive analysis
- Test these models on other regions and crops for comparison of crop yield.

Purdue University 2019-2020 Data Mine Program


## References & Acknowledgements

Thank you to Angela Bowman for guidance throughout this project, and John Deere for presenting us with this opportunity. Weather data is available at NOAA.gov, and crop yield data is available at USDA.gov.