Project #1: Crop Yields & Weather

Kerry Olden, Erin Nazetta, Nour Hamdan, Onuwa Nwosu

The Hypothesis Statement

Crop yields are driven by weather variables such as temperature and precipitation during summer months.

We looked at crop yields of soybeans, corn and wheat for five U.S. states--Missouri, Kansas, Illinois, Iowa and Texas--for a span of 20 years.

Compared the difference between the state's average temperature and its effect on yield and the average precipitation and its effect on the amount of yield.

We also looked at different counties,within a state, to see how different weather patterns across a state affects the crop's yield.

The Questions Asked

1. What temperature produces the highest yield for our crops?

2. At what precipitation will the crop yield the most?

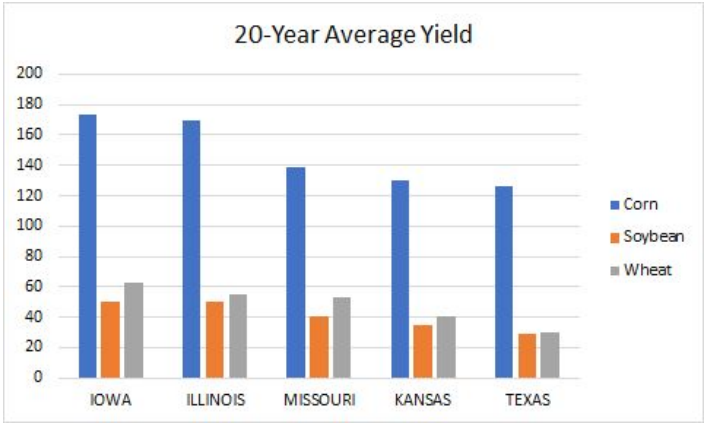
3. What states tend to have better temperature / precipitation for the best yield of crops?

4. What does the crop growth look like with higher/lower precipitation/weather?

The Data

- We gathered precipitation and temperature data for counties in five states (Missouri, Illinois, Kansas, Texas and Iowa) from National Centers for Environmental Information - National Oceanic and Atmospheric Administration: https://www.ncdc.noaa.gov/data-access

- We pulled annual crop yield data by county for three crops (corn, soybeans, wheat) from United States Department of Agriculture - National Agricultural Statistics Service - Quick Stats: https://quickstats.nass.usda.gov/



Data Exploration

Crop Yield Data:

https://github.com/nourhamdan/databootcamp_project1/blob/main/Project1_pt2.ipynb

- Remove unnecessary columns
- Merge county_state to make column to merge
- Split data to 3 data frames, one for each crop
- Merge all 3 crops back into one data frame
- Add columns in weather data for month and year

Weather Data:

https://github.com/nourhamdan/databootcamp_project1/blob/main/Project1_pt3.ipynb

- Extract weather data for May, June, July & August
- Separate weather into 5 data frames by state
- Sum precipitation
- Take average maximum temperature
- Merge weather data and yield data into 5 csvs

	COUNTY_STATE	Year	MONTH	Precipitation (Inches)	Maximum temperature (F)	Minimum temperature (F)	Unnamed: 0	State	County	Corn Yield	Wheat Yield	Soybean Yield
0	BARBER, KANSAS	2000	August	0.01	100.806452	70.225806	16322	KANSAS	BARBER	110.0	37.0	29.0
1	BARBER, KANSAS	2000	July	2.90	92.357143	67.928571	16322	KANSAS	BARBER	110.0	37.0	29.0
2	BARBER, KANSAS	2000	June	4.46	84.666667	63.100000	16322	KANSAS	BARBER	110.0	37.0	29.0
3	BARBER, KANSAS	2000	May	3.77	81.935484	55.967742	16322	KANSAS	BARBER	110.0	37.0	29.0
4	BARBER, KANSAS	2001	August	0.63	96.548387	68.903226	15024	KANSAS	BARBER	111.0	38.0	NaN

The Graphs

1. We made a variable for each Yield to filter out all the data for the specific yield we wanted to graph
2. We labelled the x and y axis for the Yield and the data we were comparing it to
3. We gave the graph a title and marked the x and y axis
4. We wanted to show the regression line and show how the data points compared to each other
5. Saved the chart into a png file

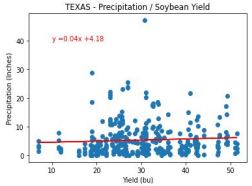
In [23]:
```
soybean_tx = tx_df[tx_df['Soybean Yield'].notnull()]
soybean_tx.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 344 entries, 84 to 1667
Data columns (total 12 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   COUNTY_STATE              344 non-null    object
 1   Year                      344 non-null    object
 2   MONTH                     344 non-null    object
 3   Precipitation (Inches)    344 non-null    float64
 4   Maximum temperature (F)   344 non-null    float64
 5   Minimum temperature (F)   344 non-null    float64
 6   Unnamed: 0                344 non-null    int64
 7   State                     344 non-null    object
 8   County                    344 non-null    object
 9   Corn Yield                284 non-null    float64
 10  Wheat Yield               132 non-null    float64
 11  Soybean Yield             344 non-null    float64
dtypes: float64(6), int64(1), object(5)
memory usage: 34.9+ KB
```

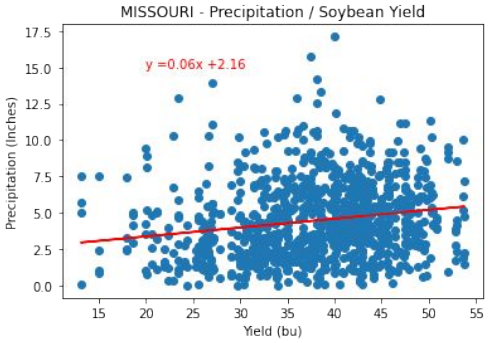
In [24]:
```
x_axis = soybean_tx["Soybean Yield"]
y_axis = soybean_tx['Precipitation (Inches)']
```

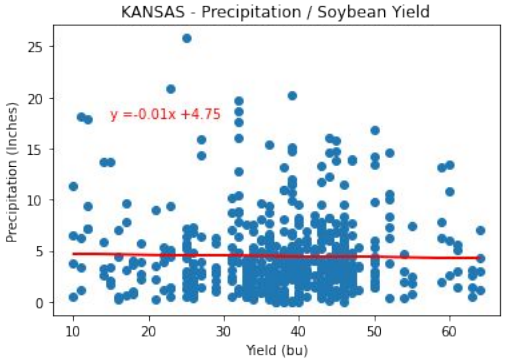
In [21]:
```
x_axis = soybean_tx["Soybean Yield"]
y_axis = soybean_tx['Precipitation (Inches)']
plt.scatter(x_axis, y_axis)
plt.title('TEXAS - Precipitation / Soybean Yield')
plt.xlabel("Yield (bu)")
plt.ylabel('Precipitation (Inches)')
(slope,intercept, rvalue, pvalue,stderr)  = st.linregress(x_axis, y_axis)
regress_values = x_axis * slope + intercept
line_eq = 'y =' +str(round(slope,2)) + "x +" + str(round(intercept,2))
plt.annotate(line_eq,(10,40), fontsize =10, color="red")
plt.plot(x_axis,regress_values, color = 'red')
plt.savefig("soybean_tx")
print(slope)
print(intercept)
print(rvalue)
print(pvalue)
print(stderr)
print(f"The r-squared is: {rvalue**2}")
```

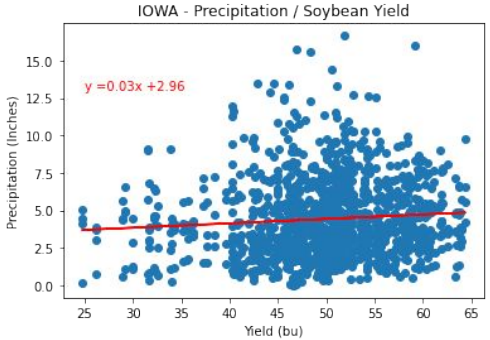
```
0.038022819955809355
4.180294707989933
0.0649099563244618
0.22983459045651344
0.031608443719259506
The r-squared is: 0.004213302430043538
```



Precipitation - Soybean Yield

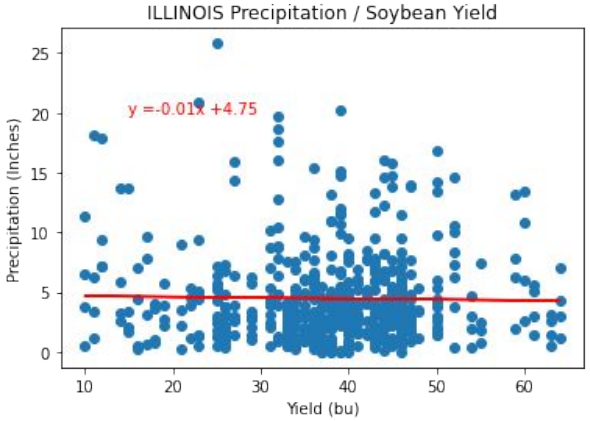


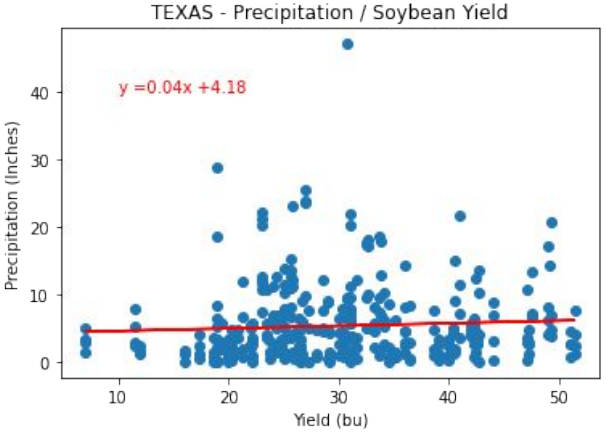




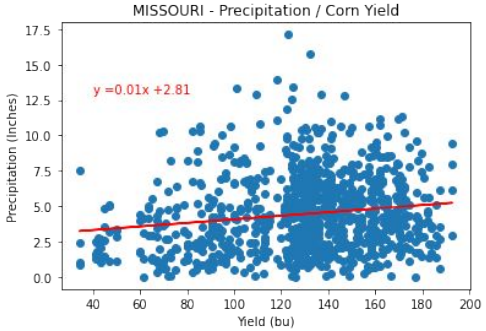
The r-squared is: 3%

Missouri had the most correlation compared to the other graphs and shows around the 2.5-5 inch range for precipitation to have the most yield grown.

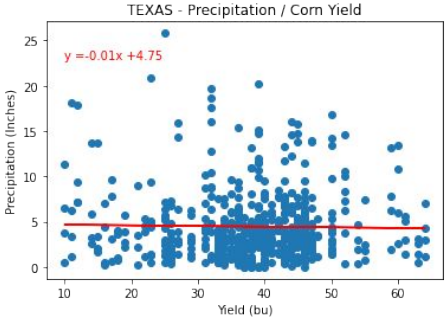


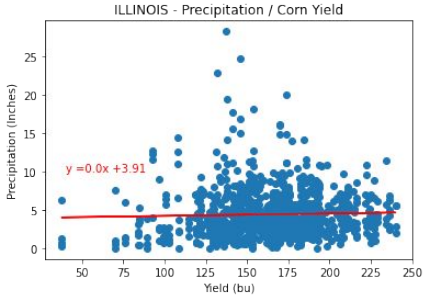


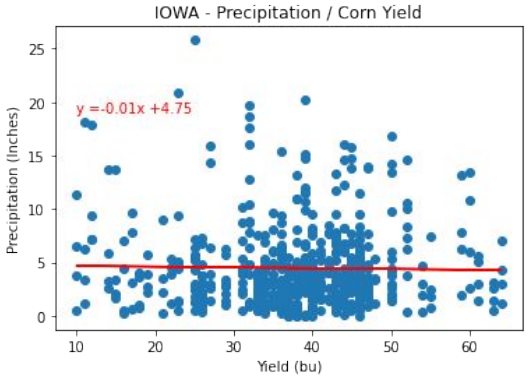
Precipitation - Corn Yield

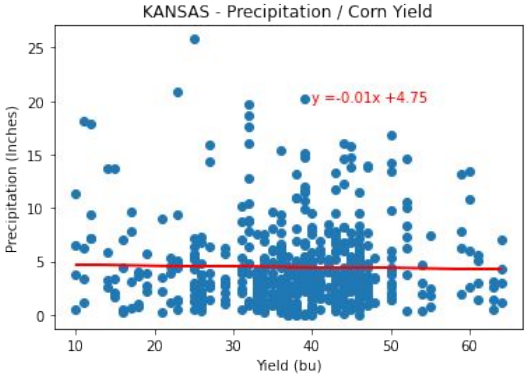


The r-squared is : 2%

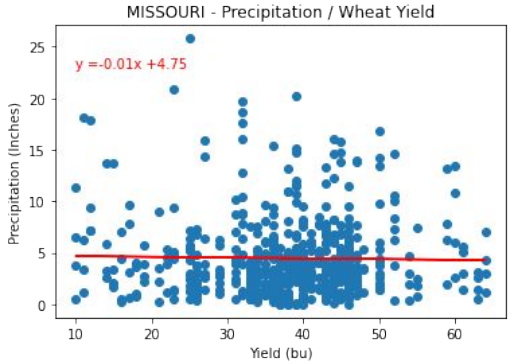


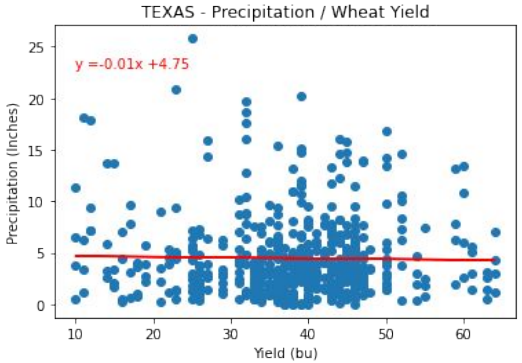


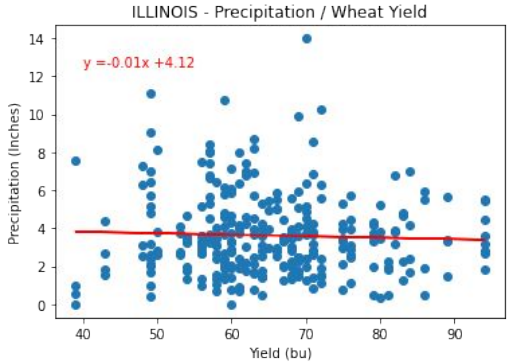


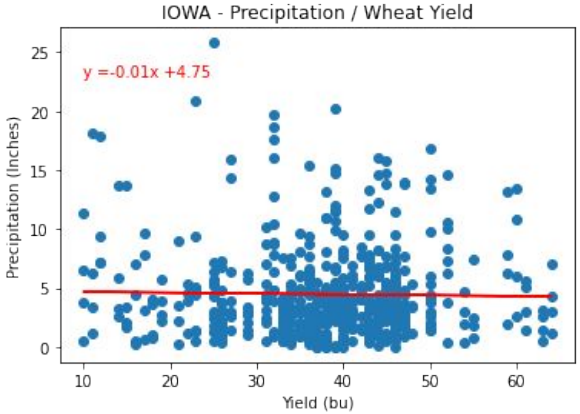


Precipitation - Wheat Yield

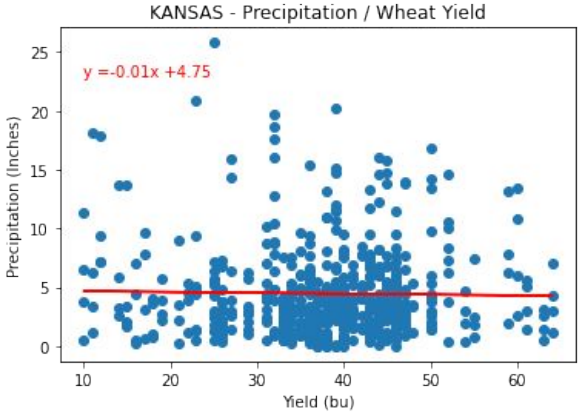




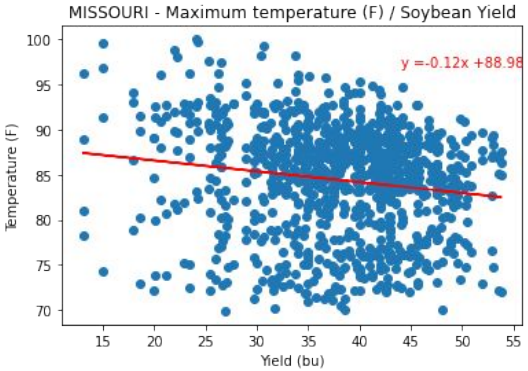


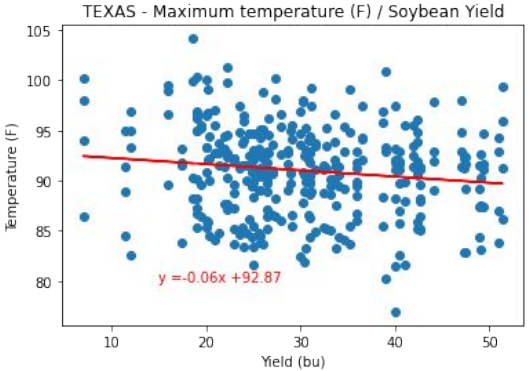


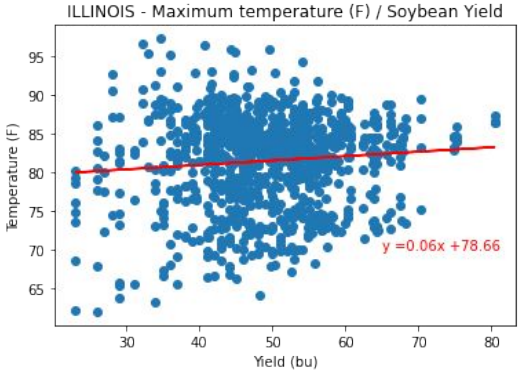
The r-squared is: 3%



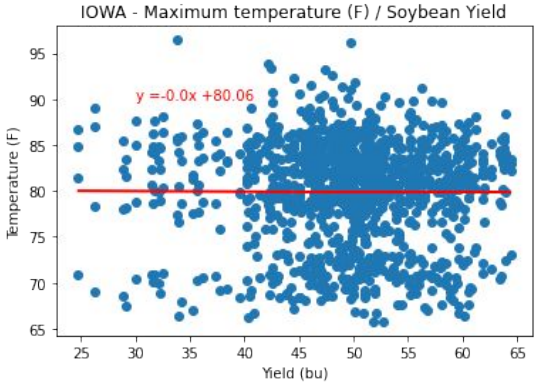
Temperature - Soybean Yield

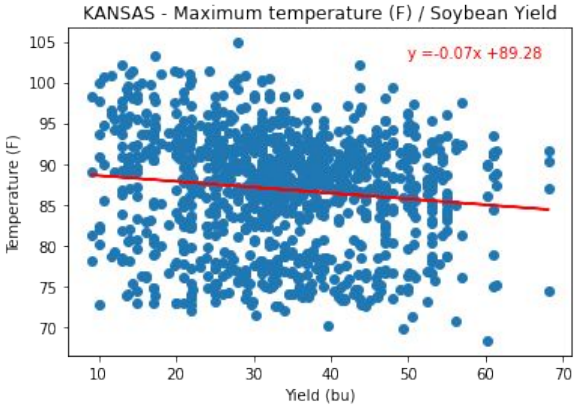




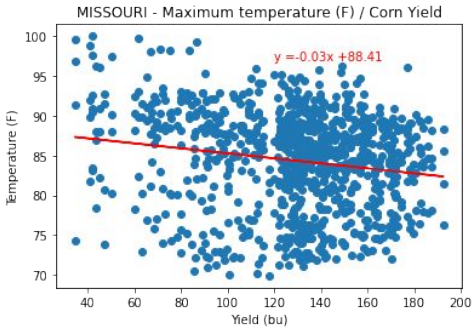


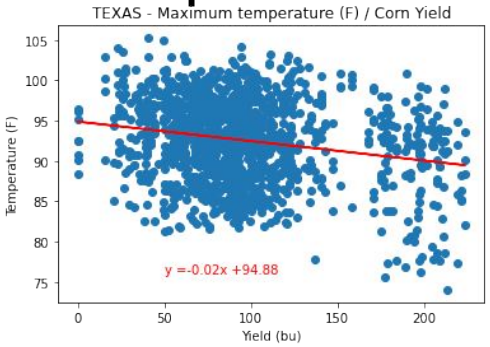
The r-squared is: 2%



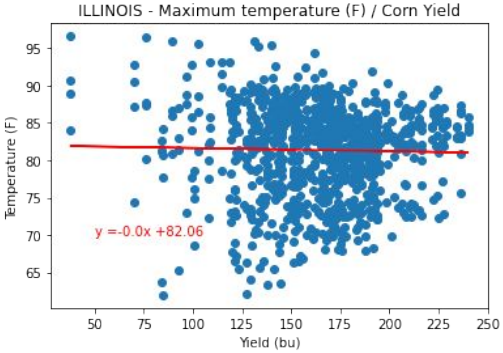


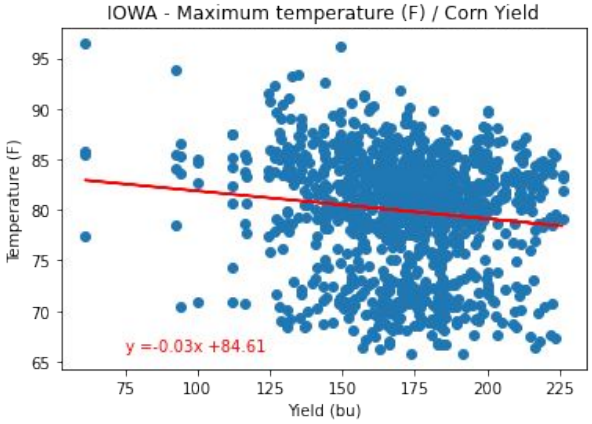
Temperature - Corn Yield

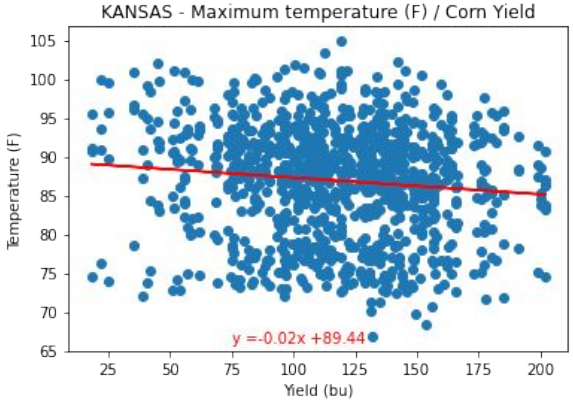




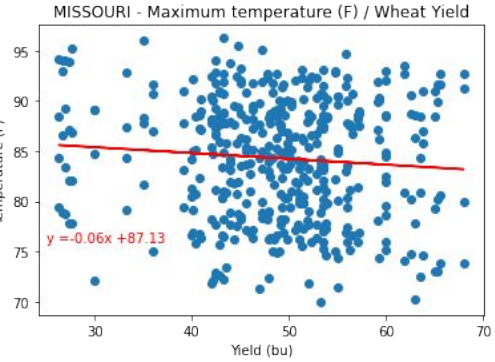
The r-squared is: 4%

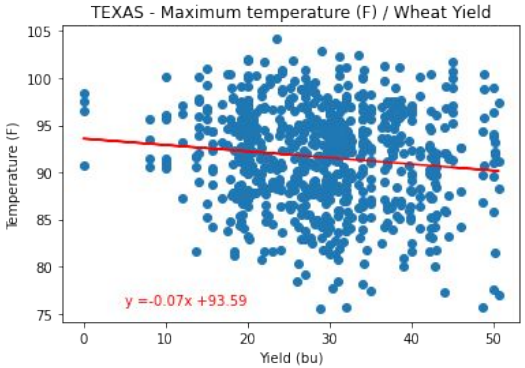




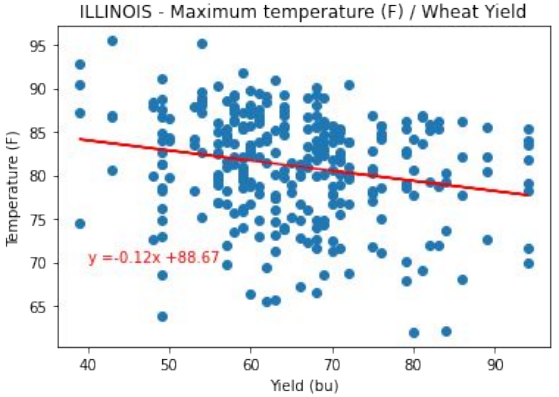


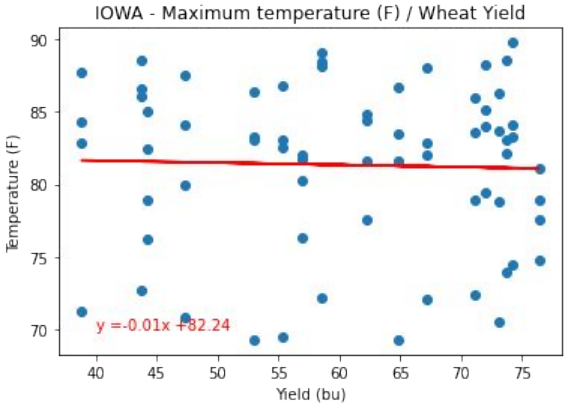
Temperature - Wheat Yield

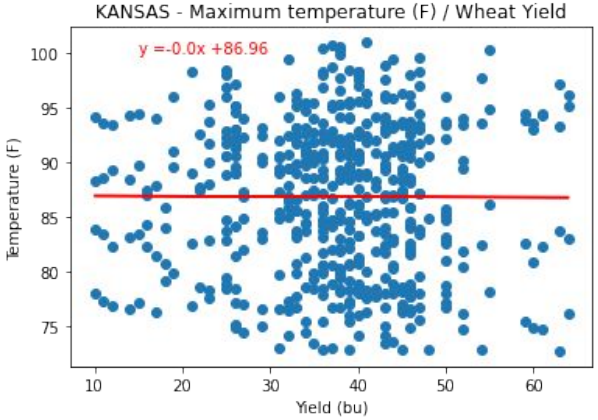




The r-squared is: 4%

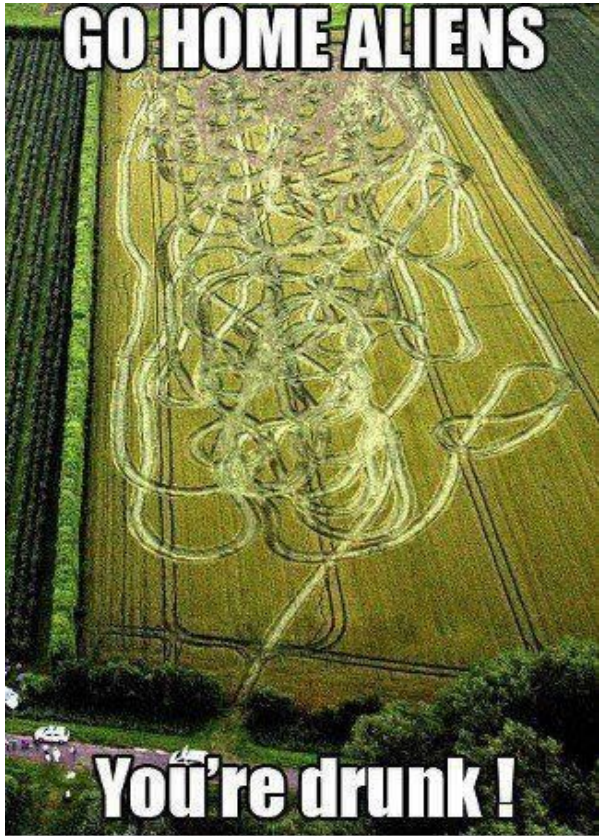






The Summary

1. What temperature produces the highest yield for our crops?
   - 85℉
2. At what precipitation will the crop yield the most?
   - Around the 5 inch mark is when crops yielded the most
3. What states tend to have better temperature / precipitation for the best yield of crops?
   - Missouri had the highest correlation of weather data and crop yields
4. What does the crop growth look like with higher/lower precipitation/weather?
   - Precipitation of 1 to 7.5 inches tends to correlate to higher crop growth; Precipitation rates over 8 inches, and under 1 inch, led to the least amount of growth
   - The average maximum temperature range between 75℉ and 90℉ tended to yield the majority amount of crops; While the average maximum temperatures of >90℉ and <75℉ yield the least amount of crops



The Conclusion

- There was little correlation between the weather data we analyzed and crop yields, as most of the regressions had an r-squared values ranging from 2%-4%.
- Some of the reasons for this could be, not well enough aligned weather data time period to crop development, using max temp vs average, inability to control for different soil types or farming techniques, use of different seed types or equipment.
- We do not have enough of a relationship to use this data to forecast crop yields based on weather and would need to do further analysis.