LA 558 Web Mapping and Spatial Data Visualization Spring 2023 Final Project Documentation

Zahra Khalilzadeh

Impacts of Weather Variables on Corn Yield in Iowa

Data Source:

- The data used in this project included the weather variables and corn yield for years of 1980-2018 corn_data_soilgrid250_modified_states_9.csv from this link: https://github.com/saeedkhaki92/CNN-RNN-Yield-Prediction

Dashboard 1: Show Iowa Counties' Corn Yield Percentage Change between 1980 and 2018

1-1-Data Preprocessing

- For the first part of the project I wanted to show the yearly corn yield in Iowa counties from 1980 to 2018 and have a wide format of data. In this format, the data will have the yield values for each county on each row and the corresponding year on the columns. Then, by creating a parameter in Tableau, users can select a year and view the corn yield for different counties of Iowa for the selected year.

Dashboard 1: Show Iowa Counties' Corn Yield Percentage Change between 1980 and 2018

1-1-Data Preprocessing

- So, I wrote a python code in CornDataForeachyearIowa.ipynb to make the data ready.
- 1-The data included the corn yield data along with soil and weather variables for all counties of all US states, and I just got the data for Iowa counties, and pivote the data to have the wide format:

```python
import pandas as pd

# Read the CSV file
df = pd.read_csv('corn_data_soilgrid250_modified_states_9.csv')
df2 =  pd.read_csv('IowaFiinal.csv')

# Define a list of loc_ID values to consider
loc_ids = df2['loc_ID']

# Filter the data to only include the specified loc_IDs
df = df[df['loc_ID'].isin(loc_ids)]

# Pivot the data on the 'year' column and fill the columns with values from the 'yield' column
df = df.pivot(index='loc_ID', columns='year', values='yield')

# Print the resulting dataframe
print(df)
df.to_csv('IowaLocIDsYielForallYears.csv')
```

Dashboard 1: Show Iowa Counties' Corn Yield Percentage Change between 1980 and 2018

1-1-Data Preprocessing

- Python code to filter the data for Iowa counties and pivote the data to have a wide format:

```python
import pandas as pd

# Read the CSV file
df = pd.read_csv('corn_data_soilgrid250_modified_states_9.csv')
df2 =  pd.read_csv('IowaFiinal.csv')

# Define a list of loc_ID values to consider
loc_ids = df2['loc_ID']

# Filter the data to only include the specified loc_IDs
df = df[df['loc_ID'].isin(loc_ids)]

# Pivot the data on the 'year' column and fill the columns with values from the 'yield' column
df = df.pivot(index='loc_ID', columns='year', values='yield')

# Print the resulting dataframe
print(df)
df.to_csv('IowaLocIDsYielForallYears.csv')
```

loc_ID	year	yield
0	2010	121.9
0	2011	135.8
0	2012	76
0	2013	153.3
0	2014	202.4
0	2015	158.1
0	2016	193.1
0	2017	192.2
0	2018	197.8
1	1980	49
1	1981	100
1	1982	105
1	1983	55
1	1984	89
1	1985	105
1	1986	96
1	1987	110
1	1988	80
1	1989	102
1	1990	112

Long format to wide format

loc_ID	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
194	89.7	121	102.2	77.3	97.4	127.9	128.3	128.4	86.1	131.4	115.1	126.4	140	76.8
195	78.4	119.6	95.3	60.1	70	126.1	119.6	125.8	73.5	110.1	119.4	133.7	135.4	79
196	114.5	127.2	116.3	109.6	112.7	114.6	138.7	136.2	66.1	122.2	132.7	132.6	129.4	83.2
198	90.6	115.4	112	89.6	101.6	130.6	126.1	133.5	91.7	132	122	132	147	80.6
199	105.7	126.1	125.1	90.8	121.2	113.1	134.3	138	57.4	107.3	120.5	93.2	140.9	84.8
200	117.8	132.5	128.2	111	106.5	130.3	144.9	143	74.1	92.9	125	111.2	156.3	73.6
201	118.8	121.8	128.7	111.4	120.8	125.2	144.9	142.9	86.8	130.5	118.8	123.4	163.8	84
202	123.1	135	124.5	113.6	115.2	137.7	142.5	137.6	70.5	99.3	133.6	130.9	155.8	79.7
203	126.5	133.3	130	100.4	118.4	117	145.6	141.4	75.3	113.6	133.3	123.5	151.1	76.6
204	121.8	129.9	124.9	81.4	98.3	132.4	131.6	131.4	113	132.4	130.4	120.5	153.7	80.3
205	124.3	127	123.9	104.5	113.7	132.5	134.2	131.6	94.2	94.3	129.5	125.6	152.3	73.9
206	123.4	135	127.8	101.2	104.4	127.1	134.6	143.2	123.4	134.6	125.4	116	167.6	83.3
207	98.1	124.6	114.2	87.8	109.3	129	138.5	135.8	115.9	129.4	120.7	122.1	153.5	70.5
208	88.4	116.2	110.4	91.3	107	131.9	129.8	132.9	91.5	131.4	123.9	131.4	148.8	72.1
209	124.6	141.4	138.1	90.9	134.7	134.7	144	120.6	61.3	131.8	136.5	108.6	147.7	91.3
210	130.2	134.1	126.9	109.3	111.9	125.8	144.7	133	100.2	97	136.8	120	143.7	64.8
211	111.9	122.6	123.5	89.8	116.4	126.4	134.8	129.3	120	132.7	130.7	135.8	150.7	87.5
212	122.3	124.2	111.2	97.2	103.4	115	136.3	133	69	98.8	130	130.1	140.6	71
213	83.7	114.4	79.6	27.1	61.2	130.4	100.5	100	39.3	68.2	104.3	102.3	121	43
214	126.3	134.7	113.5	84.4	104.2	134	136.9	136.5	100.3	130.5	133.7	112.2	147.4	66
215	116	126	123.8	106.6	116.4	120.8	146.6	139	66.9	134	141.4	133.7	134.7	89.9
216	123	134.5	131.2	87.9	113.6	116.1	135.1	109.3	53.3	122.2	137.3	105.8	142	100.7
217	67.8	94.7	110.4	80	110.6	116.9	126.9	127.1	103.7	122.6	125.5	122.8	144.8	76.7
218	96.9	134.8	126.5	89	114.7	113.6	132.1	140	88.2	139.5	127.5	124.3	153.6	95.4
220	83.3	113.7	89.5	28.1	84.2	120.4	117.4	108	39.3	84.8	97.4	105.7	124.6	48.8
221	124.6	134.5	131.9	89	118.2	114	150.4	139	64.9	122.5	138.9	125.4	141.5	77.3

Dashboard 1: Iowa Counties' Corn Yield Percentage Change between 1980 and 2018

1-1-Data Preprocessing

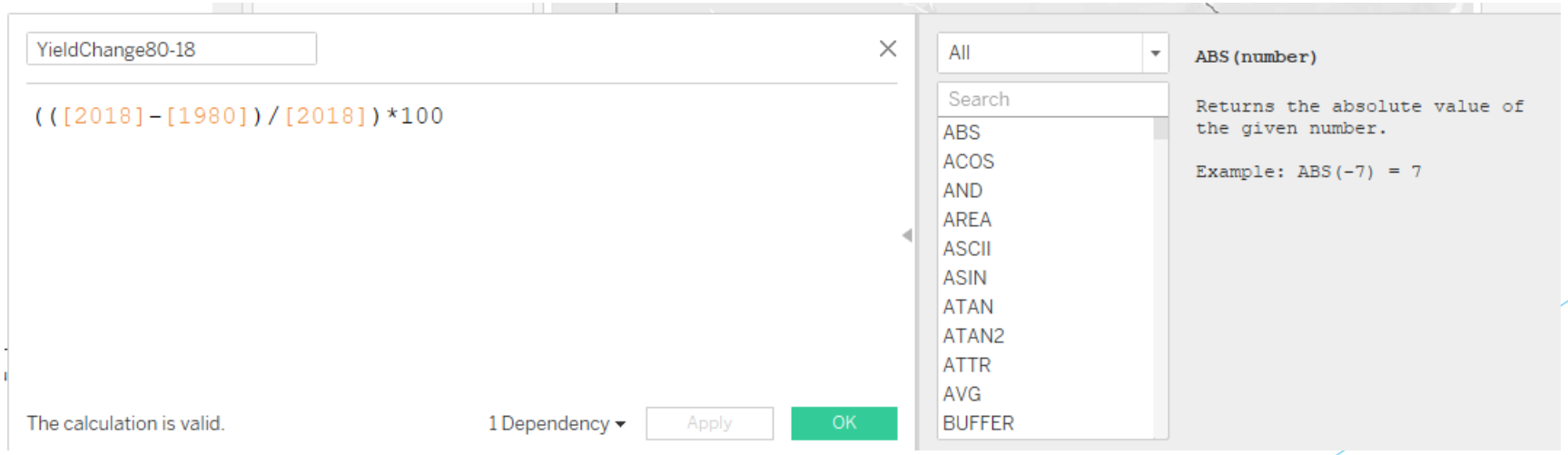
- Python code merge the wide data with loc_ID to the Iowa counties shapefiel details. So that for each county we have these column information as well: ['PERIMETER', 'DOMCountyI', 'FIPS', 'FIPS_INT', 'CountyName', 'StateAbbr', 'SHAPE_Leng', 'OBJECTID', 'CountyName_1', 'State', 'lat', 'long', 'DOMCountyI_1', 'Shape_Length', 'Shape_Area']

```
df = pd.read_csv('IowaLocIDsYielForallYears.csv').reset_index()
df2 =  pd.read_csv('IowaFiinal.csv').reset_index()

print('df', df['loc_ID'].dtype)
print('df2', df2['loc_ID'].dtype)

merged_df = pd.merge(df, df2.loc[:, ['loc_ID', 'PERIMETER', 'DOMCountyI', 'FIPS', 'FIPS_INT', 'CountyName', 'StateAbbr', 'SHAPE_Leng', 'OBJECTID', 'CountyName_1',

# Display merged dataframe
print(merged_df)
merged_df.to_csv('IowaLocIDsYielForallYearsfinal.csv')
```

AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE
PERIMETE	DOMCoun	FIPS	FIPS_INT	CountyNa	StateAbbr	SHAPE_Le	OBJECTID	CountyNa	State	lat	long	DOMCoun	Shape_Le	Shape_Ar
153765.3	1	19001	19001	Adair	IA	1.610861	1	adair	iowa	41.31278	-94.4953	1	204908.5	2.62E+09
134211.6	2	19003	19003	Adams	IA	1.43068	2	adams	iowa	41.01616	-94.685	2	177974.2	1.94E+09
172890.3	3	19005	19005	Allamake	IA	1.847862	3	allamakee	iowa	43.29617	-91.3962	3	237607	3.22E+09
144819.2	5	19009	19009	Audubon	IA	1.500155	4	audubon	iowa	41.67021	-94.9107	5	194034.2	2.06E+09
174069.6	6	19011	19011	Benton	IA	1.805887	5	benton	iowa	42.05997	-92.0776	6	234735.5	3.38E+09
157108.5	7	19013	19013	Black Haw	IA	1.669052	6	black haw	iowa	42.46075	-92.303	7	213122.3	2.73E+09
154073.6	8	19015	19015	Boone	IA	1.624895	7	boone	iowa	42.01718	-93.9254	8	207587	2.69E+09
136161.8	9	19017	19017	Bremer	IA	1.475098	8	bremer	iowa	42.76696	-92.31	9	185568.8	2.11E+09
156690.4	10	19019	19019	Buchanan	IA	1.663029	9	buchanan	iowa	42.46349	-91.8449	10	212545.7	2.73E+09
155119.5	11	19021	19021	Buena Vis	IA	1.645523	10	buena vis	iowa	42.74458	-95.1128	11	211253.6	2.79E+09
155223	12	19023	19023	Butler	IA	1.64582	11	butler	iowa	42.73365	-92.7819	12	211462.1	2.8E+09
163055.1	13	19025	19025	Calhoun	IA	1.736286	12	calhoun	iowa	42.38082	-94.6332	13	220859.1	2.72E+09
153707.8	14	19027	19027	Carroll	IA	1.61976	13	carroll	iowa	42.01936	-94.8603	14	207050.1	2.68E+09
153027.1	15	19029	19029	Cass	IA	1.601998	14	cass	iowa	41.31228	-94.9215	15	203902.1	2.6E+09
155209.7	16	19031	19031	Cedar	IA	1.632346	15	cedar	iowa	41.76086	-91.1263	16	208213.6	2.71E+09
154230.9	17	19033	19033	Cerro Gor	IA	1.641674	16	cerro gord	iowa	43.07424	-93.2549	17	211295.1	2.79E+09
154682	18	19035	19035	Cherokee	IA	1.640159	17	cherokee	iowa	42.7248	-95.625	18	210619.1	2.77E+09
144990.8	19	19037	19037	Chickasaw	IA	1.557933	18	chickasaw	iowa	43.05994	-92.3167	19	198539.2	2.46E+09
134994.3	20	19039	19039	Clarke	IA	1.438093	19	clarke	iowa	41.01521	-93.7843	20	179045.6	1.97E+09
153944.1	21	19041	19041	Clay	IA	1.638979	20	clay	iowa	43.08094	-95.1408	21	210828.3	2.78E+09
204158.7	22	19043	19043	Clayton	IA	2.202352	21	clayton	iowa	42.85477	-91.3667	22	278459.9	3.83E+09
208161.8	23	19045	19045	Clinton	IA	2.243617	22	clinton	iowa	41.84447	-90.1887	23	279584.2	3.32E+09
173298.9	24	19047	19047	Crawford	IA	1.855344	23	crawford	iowa	42.02058	-95.3812	24	233350.9	3.36E+09
160342.5	25	19049	19049	Dallas	IA	1.688815	24	dallas	iowa	41.67442	-94.0394	25	214843.6	2.75E+09

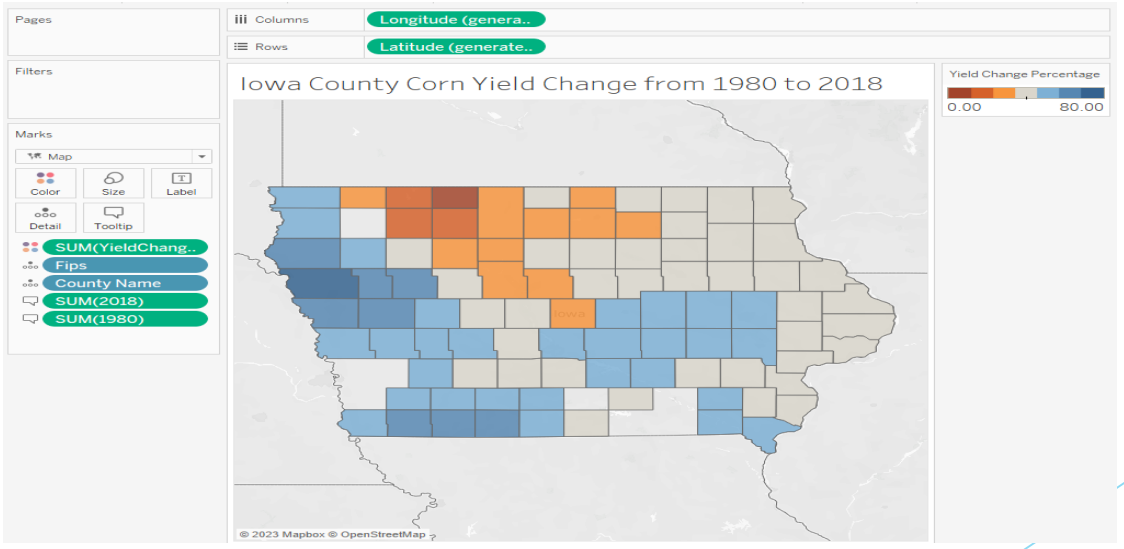
Dashboard 1: Iowa Counties' Corn Yield Percentage Change between 1980 and 2018 1-2-Create first Tableau dashboard

- After uploading the csv final csv file: IowaLocIDsYielForallYearsfinal.csv to Tableau, we can show the percentage of yearly corn change from 1980 to 2018. So, I created a calculated field and use it as a colur marks in the map. I also used the county names and yields of 1980 and 2018 to show as labels on each county when the user select it.
- Calculated field



Dashboard 1: Iowa Counties' Corn Yield Percentage Change between 1980 and 2018 1-2-Create first Tableau dashboard

- After uploading the csv final csv file: IowaLocIDsYielForallYearsfinal.csv to Tableau, we can show the percentage of yearly corn change from 1980 to 2018. So, I created a calculated field and use it as a colur marks in the map. I also used the county names and yields of 1980 and 2018 to show as labels on each county when the user select it.
- Results of the first dashboard:



Dashboard 2: Show the Yearly Corn Yield in Iowa Counties from 1980 to 2018
2-1-Data Preprocessing

- Here I wanted to keep the long format of data and just wanted to filter the data to have only Iowa Counties' data. So, I wrote a python code to filter the data for Iowa counties.

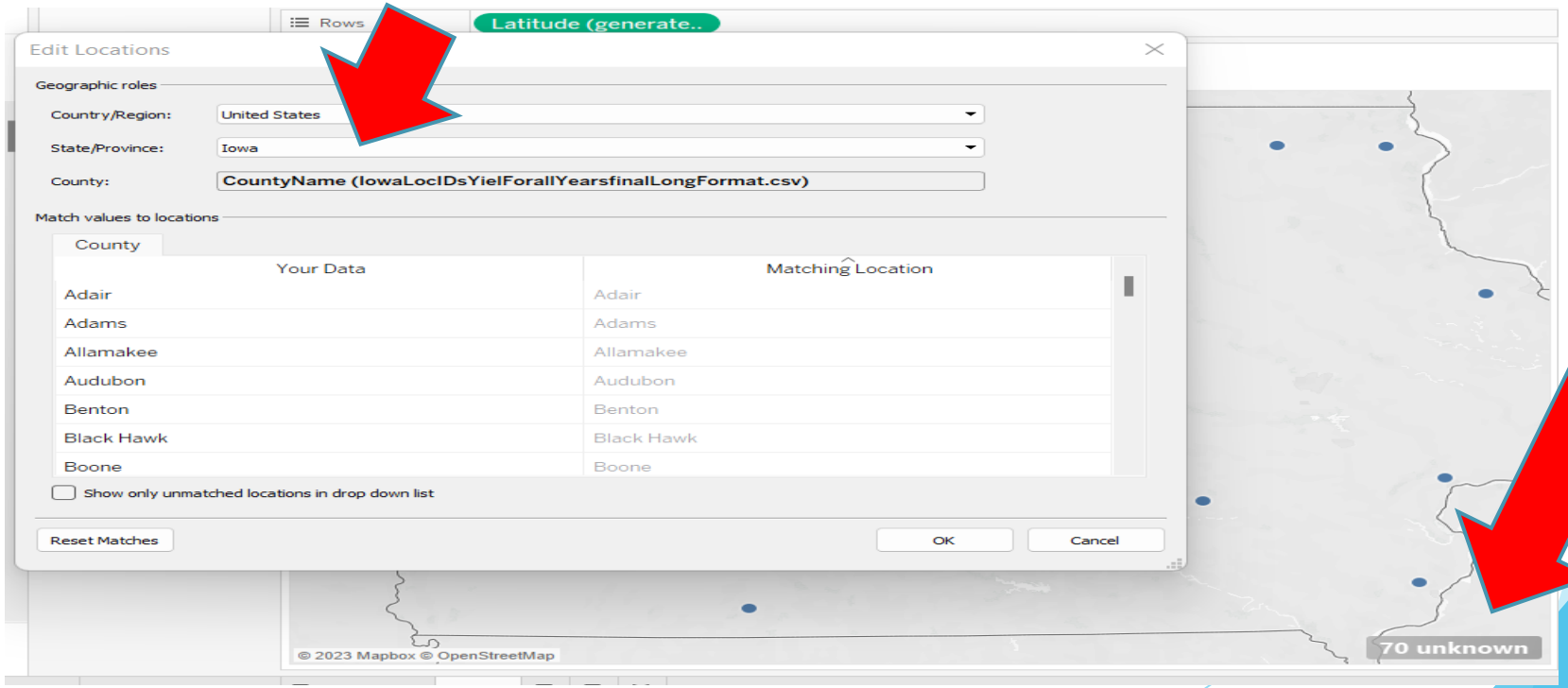
```
import pandas as pd

# Read the CSV file
df = pd.read_csv('corn_data_soilgrid250_modified_states_9.csv')
df2 =  pd.read_csv('IowaFiinal.csv')

# Define a list of loc_ID values to consider
loc_ids = df2['loc_ID']

# Filter the data to only include the specified loc_IDs
df = df[df['loc_ID'].isin(loc_ids)]


# Print the resulting dataframe
print(df)
df.to_csv('IowaLocIDsYielForallYearsLongFormat.csv')
```

- Then merge merge the wide data with loc_ID to the Iowa counties shapefiel details. So that for each county we have these column information as well: ['PERIMETER', 'DOMCountyI', 'FIPS', 'FIPS_INT', 'CountyName', 'StateAbbr', 'SHAPE_Leng', 'OBJECTID', 'CountyName_1', 'State', 'lat', 'long', 'DOMCountyI_1', 'Shape_Length', 'Shape_Area']

```
df = pd.read_csv('IowaLocIDsYielForallYearsLongFormat.csv').reset_index()
df2 =  pd.read_csv('IowaFiinal.csv').reset_index()

print('df', df['loc_ID'].dtype)
print('df2', df2['loc_ID'].dtype)

merged_df = pd.merge(df, df2.loc[:, ['loc_ID', 'PERIMETER', 'DOMCountyI', 'FIPS', 'FIPS_INT', 'CountyName', 'StateAbbr', 'SHAPE_Leng', 'OBJECTID', 'CountyName_1',

# Display merged dataframe
print(merged_df)
merged_df.to_csv('IowaLocIDsYielForallYearsfinalLongFormat.csv')
```

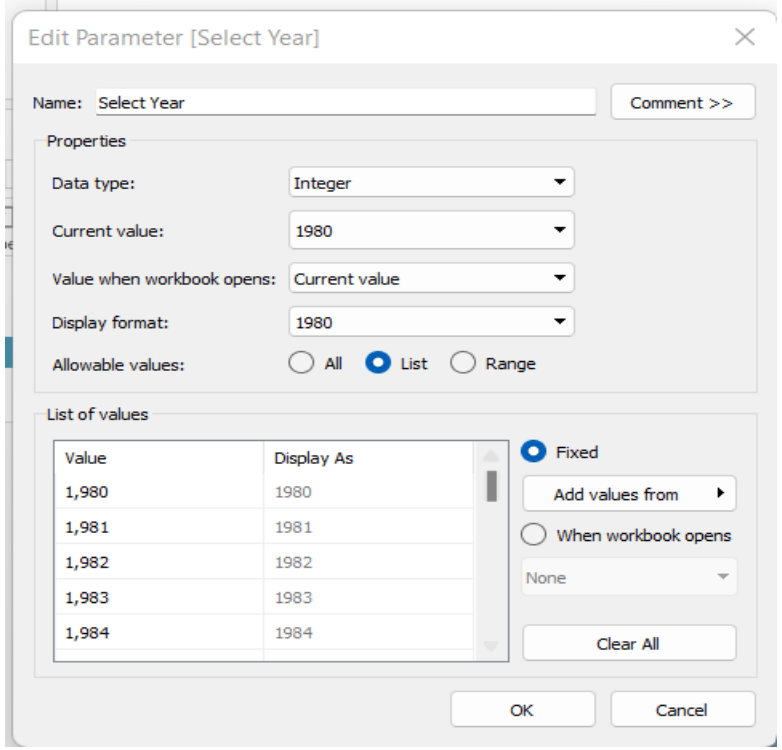
Dashboard 2: Show the Yearly Corn Yield in Iowa Counties from 1980 to 2018
2-2-Create Second Tableau dashboard

- After uploading the second csv final in the long format: IowaLocIDsYielForallYearsfinalLongFormat.csv, I faced the unknown problem. I was able to solve this problem by choosing state of Iowa in the following table:



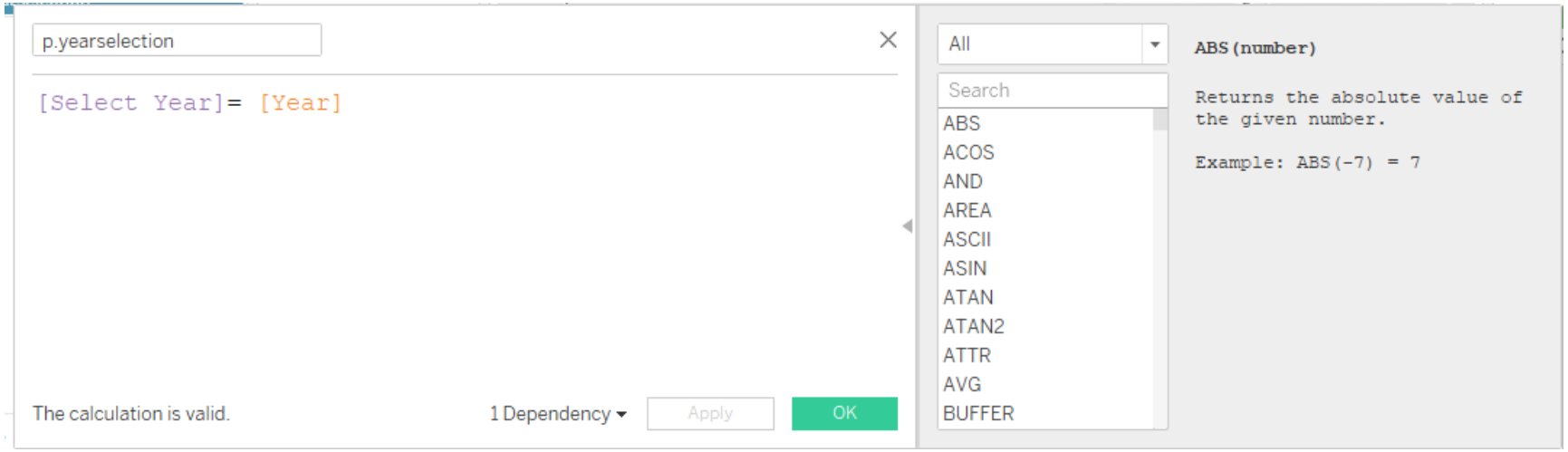
Dashboard 2: Show the Yearly Corn Yield in Iowa Counties from 1980 to 2018
2-2-Create Second Tableau dashboard

- After uploading the csv final csv file: IowaLocIDsYielForallYearsfinalLongFormat.csv to Tableau, we created a parameter and named it as select year, and select years as a list from Year column.



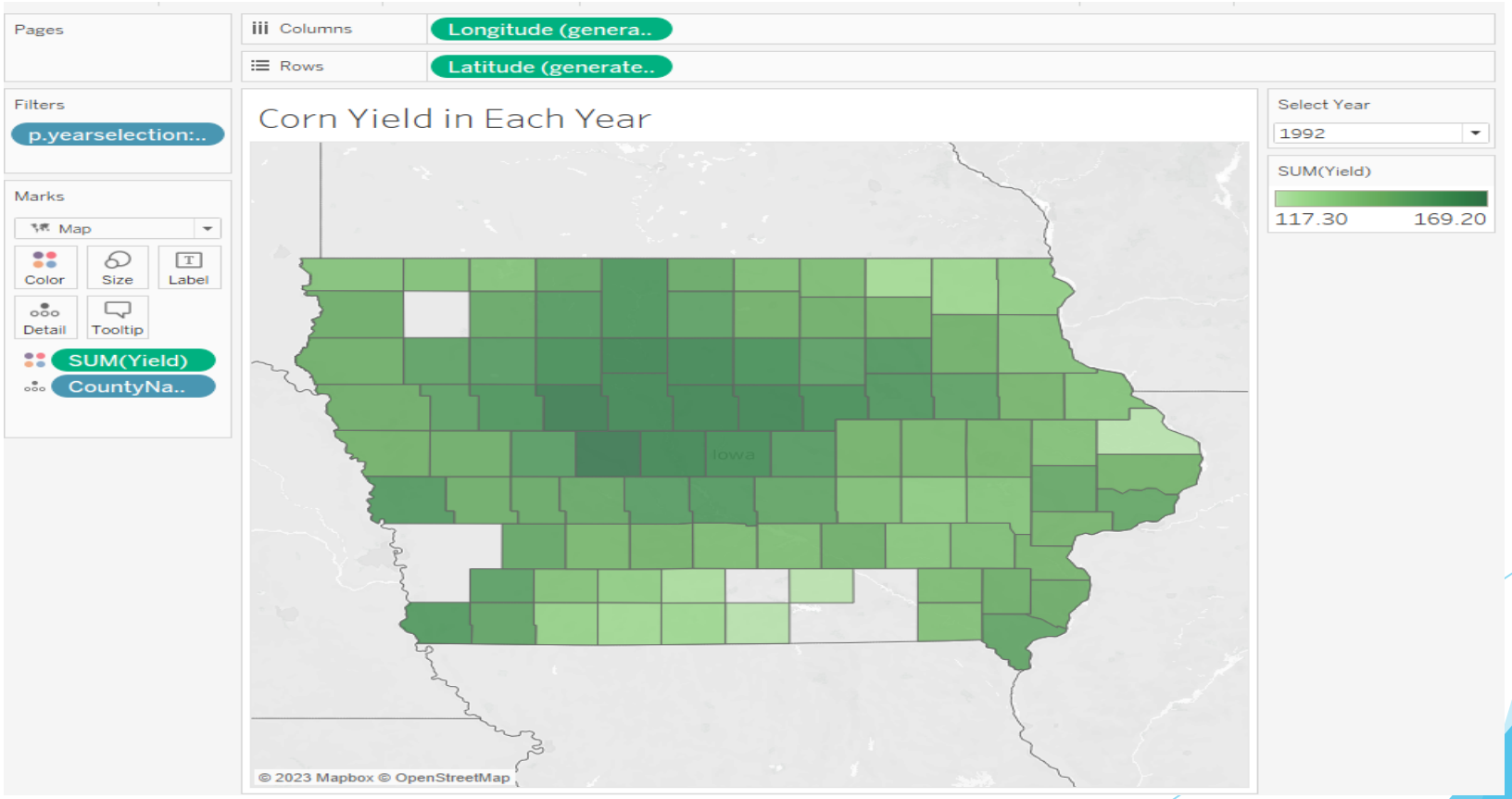
Dashboard 2: Show the Yearly Corn Yield in Iowa Counties from 1980 to 2018
2-2-Create Second Tableau dashboard

- Then we created a calculated field and used it as a filter so that it will show the yield of counties based on the selected year by the user.



Dashboard 2: Show the Yearly Corn Yield in Iowa Counties from 1980 to 2018
2-2-Create Second Tableau dashboard

- Results of the second dashboard:



Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-1-Data Preprocessing

- First, we put all the weather variables that we wanted to analyze including, precipitation, solar radiation, and average temperature from weeks of growing season which are week 10 to week 44 and took the average over the whole weeks for each county in excel.

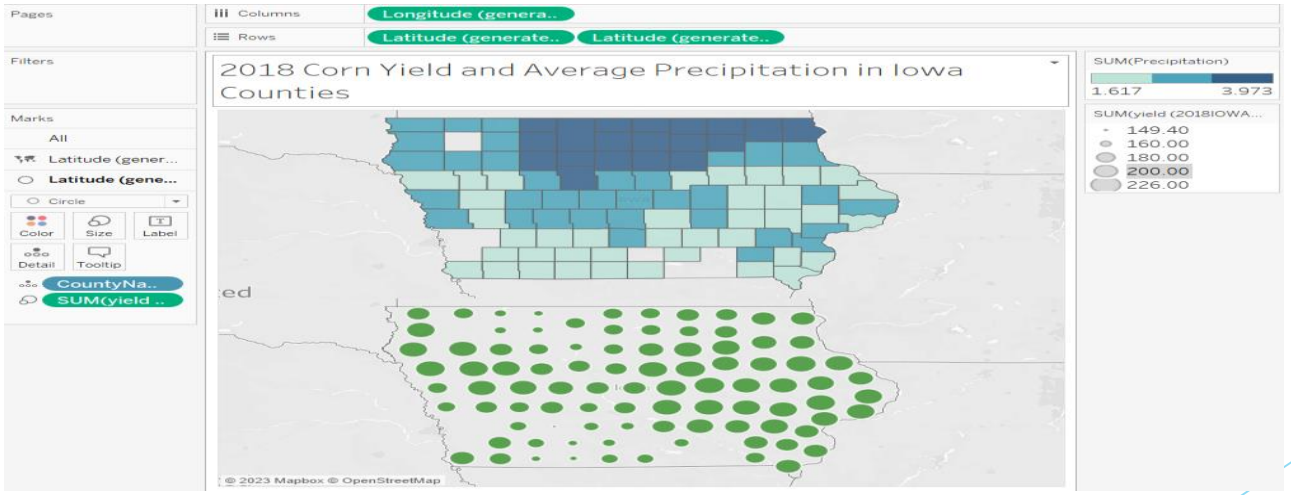
	A	B	C	D	E	F	G	H	I
1	CountyNa	loc_ID	year	yield	Precipitation	SolarRadiation	AverageMax	AverageMin	AvergaeTemperature
2	adair	194	2018	149.4	2.356060606	393.1913424	13.1525974	1.277056277	7.21482684
3	adams	195	2018	171.5	1.753246753	393.4857146	13.91477273	1.668831169	7.791801948
4	allamakee	196	2018	198.4	3.973026974	409.4337667	11.26723277	-1.135864136	5.065684315
5	audubon	198	2018	179.4	2.612554113	389.01472	12.44588745	0.701298701	6.573593074
6	benton	199	2018	214.5	2.280844156	399.5688311	12.39975649	0.144480519	6.272118506
7	black haw	200	2018	206.3	2.096320346	398.4935065	11.43777056	-0.252164503	5.59280303
8	boone	201	2018	193.7	2.73051948	394.5974031	12.5	1.112554113	6.806277057
9	bremer	202	2018	212.4	2.918831169	398.2233765	10.74188312	-0.517045455	5.112418831
10	buchanan	203	2018	207.1	2.206709957	403.4493504	11.46266234	-0.555735931	5.453463204
11	buena vis	204	2018	193.1	2.863636364	388.8519485	10.67748918	0.197510823	5.4375
12	butler	205	2018	211.3	3.215367965	398.7428569	10.74729437	-0.230519481	5.258387446
13	calhoun	206	2018	193.3	2.87012987	389.1906483	11.53181818	0.596103896	6.063961039
14	carroll	207	2018	208.5	2.655122655	387.4031754	11.9978355	0.665223666	6.331529582
15	cass	208	2018	169.7	1.942640693	388.9038965	13.39231602	1.231601731	7.311958874
16	cedar	209	2018	222.2	2.360750361	395.9457438	13.1962482	1.12049062	7.158369408
17	cerro gord	210	2018	184.3	3.848484849	414.5777785	9.726551226	-0.744588745	4.490981241
18	cherokee	211	2018	213.4	2.797979798	399.3443001	11.0966811	0.094516594	5.595598846
19	chickasaw	212	2018	193	3.847402598	411.5324675	10.29707792	-0.9375	4.679788961
20	clarke	213	2018	161.9	2.030844156	393.6415585	13.9301948	1.870941558	7.900568181
21	clay	214	2018	158.7	2.982495765	395.8947486	10.71795596	-0.343873518	5.18704122
22	clayton	215	2018	201.6	2.772136954	403.324675	11.540732	-0.574380165	5.483175916
23	clinton	216	2018	210.8	2.494434137	397.6578861	12.86363636	0.92903525	6.896335807

Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-1-Data Preprocessing

- To assess whether there is also an interaction effect between these three independent variables on a corn yield, we implemented a three-way ANOVA in R. For this matter, we used the following formula in excel to create 3 levels for weather variables that we wanted to analyze including, precipitation, solar radiation, and average temperature .
- For precipitation: =IF(E2<2.35,"P1",IF(E2<3.05,"P2",IF(E2<3.98,"P3")))
- For solar radiation: =IF(F2<394.6,"S1",IF(F2<405.88,"S2",IF(F2<423.46,"S3")))
- For average temperature : =IF(I2<5.49,"T1",IF(I2<6.96,"T2",IF(I2<8.81,"T3")))

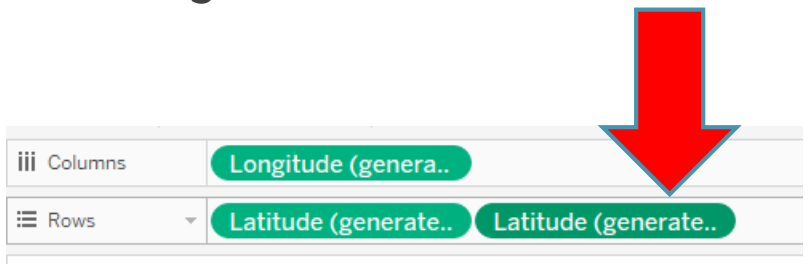
Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-2-Create 3rd dashboard including 3 maps

- We did the same process to generate 3 maps that show 2018 corn yield of Iowa counties along with average precipitation, average solar radiation, and average temperature through the growing season.
- To do it first we generated two maps first with the value of the weather variable with colors, and second with the values of yield in shapes and selected circle as the shape of the second map.

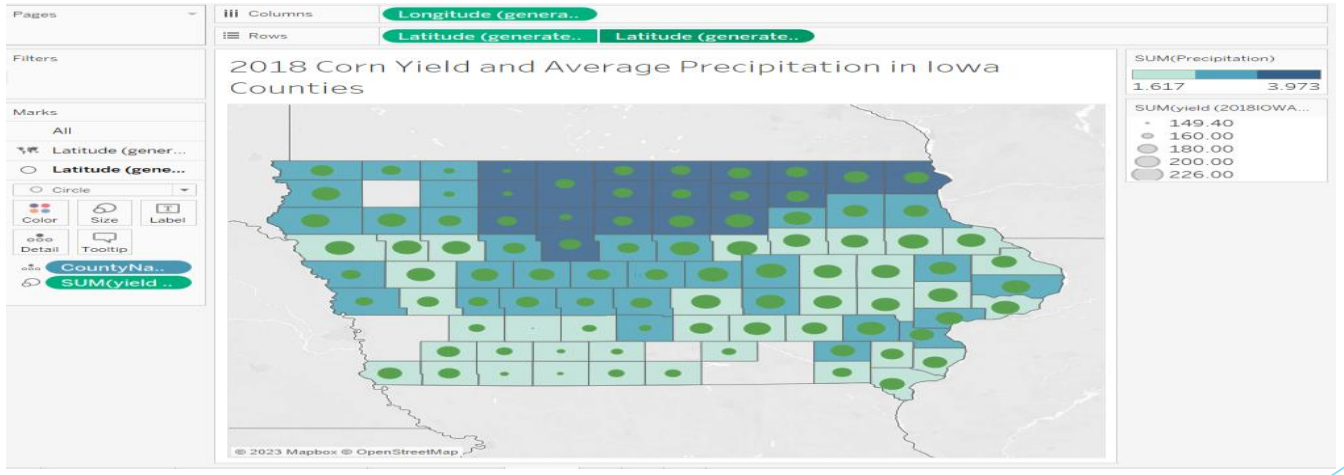


Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-2-Create 3rd dashboard including 3 maps

- The to merge them right click on the second Latitued and select Dual axis:

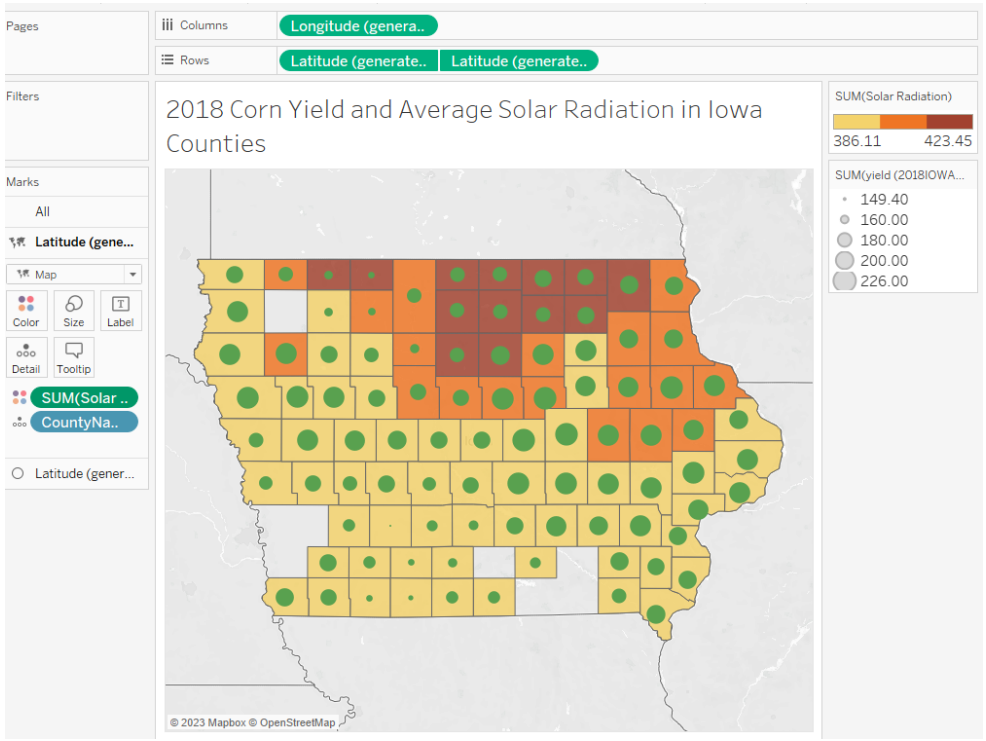


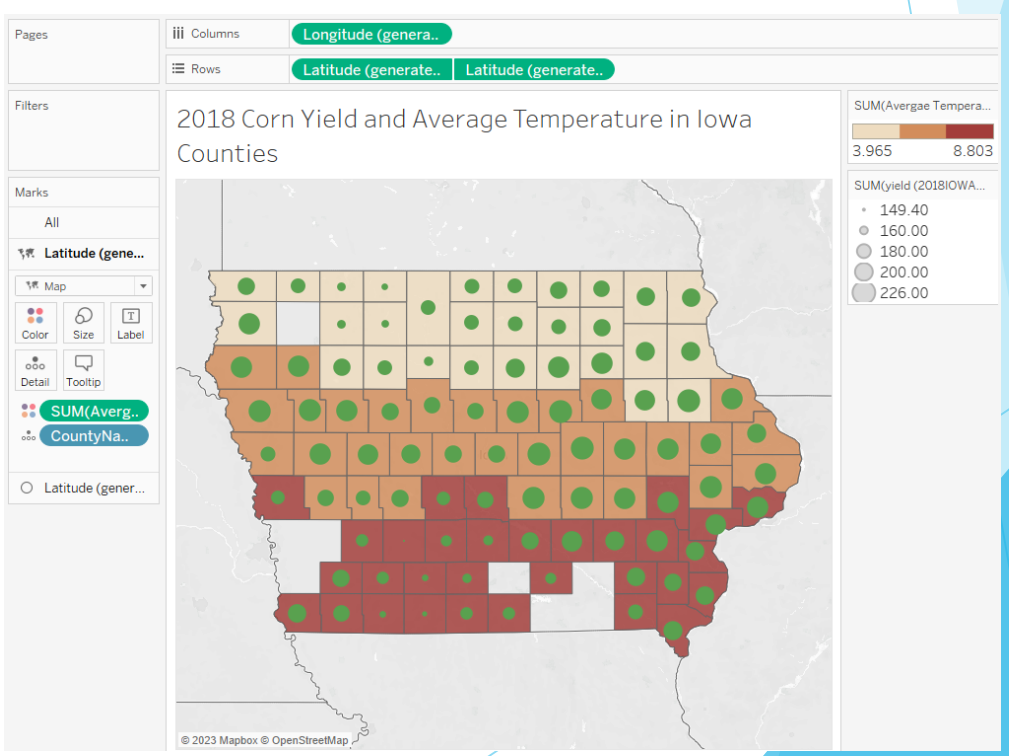
- Then we can see the merged map:



Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-2-Create 3rd dashboard including 3 maps

- We did the same process for average solar radiation and average temperature and got the following maps:





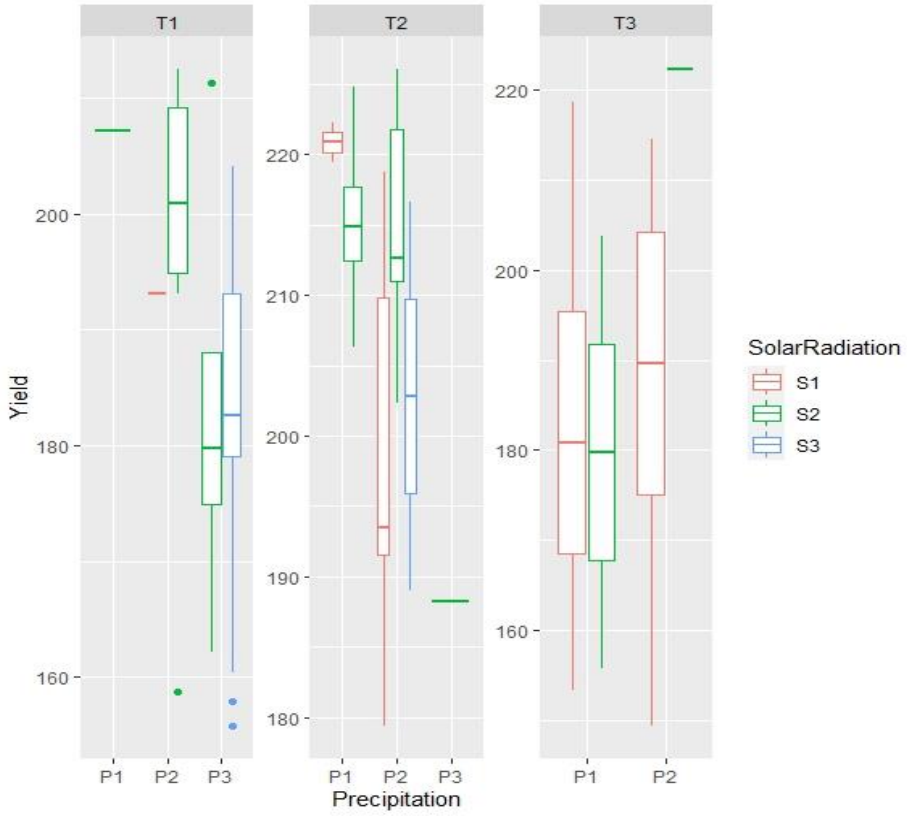
Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-4-A three-way ANOVA in R

- We used R to do the three-way ANOVA. The following is my code in R and the box plot from my code.

```
install.packages("ggplot2")
library(ggplot2)
data3Month=read.csv("C:/Users/zahra/Documents/Spring2023/CRP558-WebMapping/FinalProject
                    /2018IOWAMarchMidMay.csv")

data<-data.frame(Yield=data3Month$yield, Precipitation=data3Month$PrecipitationLevels,
                 SolarRadiation=data3Month$SolarRadiationLevels,
                 AverageTemperature=data3Month$AvergaeTemperatureLevels)
str(data)
data$Precipitation=as.factor(data$Precipitation)
data$SolarRadiation=as.factor(data$SolarRadiation)
data$AverageTemperature=as.factor(data$AverageTemperature)

Anova<-lm(Yield~SolarRadiation*Precipitation*AverageTemperature, data = data)
summary(Anova)
boxplot(Yield~SolarRadiation,data=data)
ggplot(data,
       aes(x = Precipitation, y = Yield,colour = SolarRadiation)) +
  geom_boxplot() + facet_wrap( ~ AverageTemperature, scales="free")
```

Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-4-A three-way ANOVA in R

- We used R to do the three-way ANOVA. The following is my code in R and the box plot from my code.



Dashboard 3: Impacts of Weather Variables on Corn Yield in Iowa
3-4-A three-way ANOVA in R

- Results from ANOVA: To assess whether there is also an interaction effect between these three independent variables on a corn yield, we implemented a three-way ANOVA in R. The results indicate that the solar radiation and precipitation are statistically significant at level 2 (between 394.15 and 405.87 and between 2.34 and 3.04 respectively) and average temperature is statistically significant at level 3 (between 6.95 and 8.8). The results also show that solar radiation and precipitation interaction is statistically significant at level 2 (between 394.15 and 405.87 and between 2.34 and 3.04 for solar radiation and precipitation respectively). Finally, the results indicate that that precipitation and average temperature interaction is statistically significant at level 3 (between 3.04 and 3.97 and between 6.95 and 8.8 for precipitation and average temperature respectively).

Link to my Tableau public

- https://public.tableau.com/views/FinalProject-LA558-ZahraKhalilzadeh/Resultsofthethree-wayANOVA?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link