# EDA Agriculture

Matt Oehler

February 12, 2018

1. Problem Summary

Irrigation helps farmers to avoid the effects of drought. In cases of water scarcity, actions neeeded to be taken to help optimize crop output per drop of water. Using measurements of crop water stress index (CWSI), farmers can know when they need to water their crops. However, in order to know how much water their crops need farmers use measurements of soil water content (SWC) which are expensive relative to CSWI. We hope to be able to determine if their is a relationship between SWC and CSWI. If we can understand this relationship then farmers can hopefully predict the amount of water that their crops need without paying for the expensive SWC measurements.

2. Data Summary

We have 78 measurements in our data set for SWC and CWSI. The summary statistics and a plot of CWSI vs SWC are shown below. We can see that their seems to be a relationship, but it doesn't seem to be linear.

```
      swc             cwsi
 Min.   :23.14   Min.   :0.0100
 1st Qu.:23.73   1st Qu.:0.2925
 Median :24.02   Median :0.6000
 Mean   :24.67   Mean   :0.5487
 3rd Qu.:25.32   3rd Qu.:0.7975
 Max.   :28.49   Max.   :1.0000
```

Crop Data
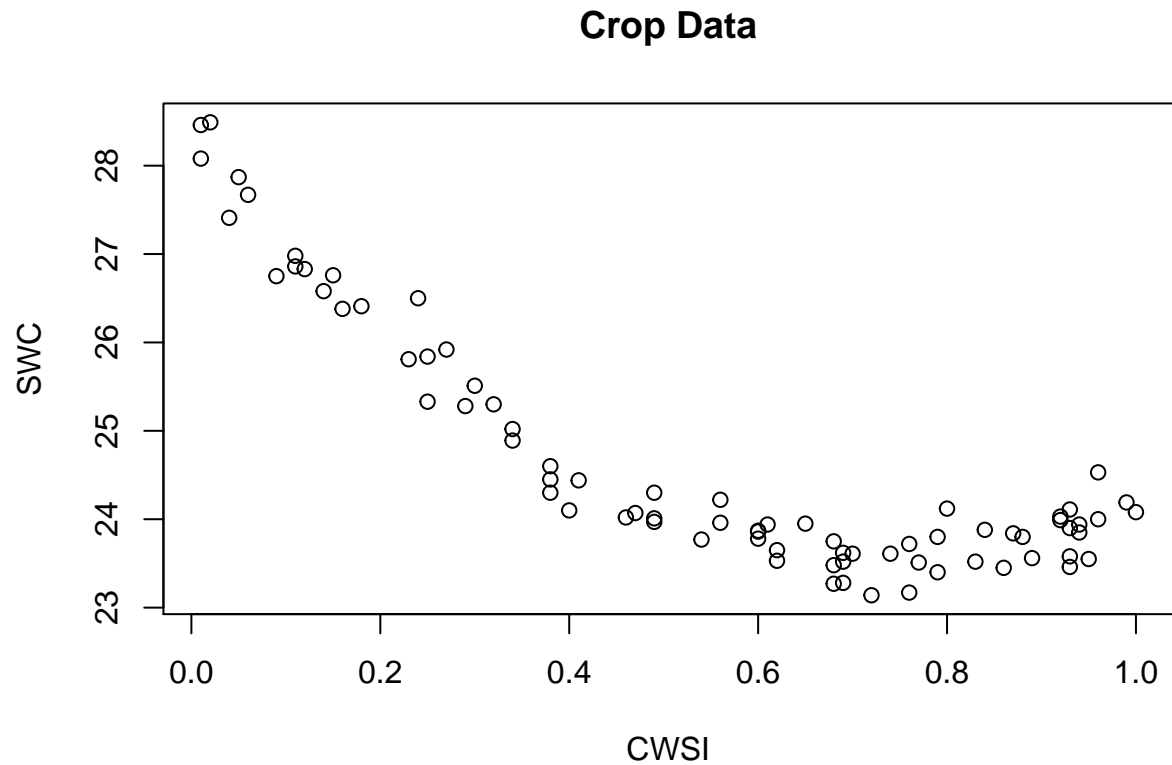
3. Statistical Method

I think that some kind of polynomial regression method would be best for this problem because the data aren't linearly related. Creating a non-linear model that fits this data would help us to understand the relationship between CWSI and SWC, and also help us to make predictions of SWC given CWSI, which is what we are most interested in.

4. Things I don't know

I don't know exactly how to make a non-linear model (beyond what was briefly mentioned in class). I'm also unsure how consistent a polynomial model would be in terms of prediction accuracy as compared to a linear model, or if prediction accuracy is even affected by the type of model as long as it fits the data well.