Project Earth

Toby Duncan, Katie Kessler, Karl Maier, Wes Merrick, Jen Mince, Katie Vervack

Data Set and Question

- Actuaries Climate Index data set
- ACI = mean(T90 – T10 + Rx5 + CDD + W + SL) -> ACI= mean(T90 - T10 + Rx5 + CDD) to see how SL changes
- How do the seasonal components of high temperatures, low temperatures, heavy rain, and drought affect the seasonal sea levels in North American regions and by how much?

Societal Need

- Why the research of rising sea levels is important:
  - Storms move closer inland (coastal flooding)
  - Infrastructures in danger (roads, bridges, power plants…)
  - Almost 40% of U.S. population lives in high-population-density coastal areas (NOAA)
  - U.S. economy threatened by impending flooding (coastal activities generate “58% of national gross domestic product” (EPA))

Results Already Available

- ACI used to make a risk index for various regions
  - Using linear regression along with disaster and mortality rates for each region a risk index was built to show the risk with living in each region
  - Each compared data set was put into a historical impact index from 1-10
- The goal of this is to help insurance companies with setting rates for the various regions
- The only other result available was an update to the ACI to include the UK and Europe.

Methodology and Justification

- The ACI had monthly data available but we chose to stick with the seasonal
  - Seasonal data is split by:
    - Winter- December, January, February
    - Spring-March, April, May
    - Summer- June, July, August
    - Fall-September, October, November
- Regions of Central Arctic and Midwest had no sea level data
- With monthly data, the changes over time are much smaller
  - Seasons give us a more visual change

Methodology

- Given the nature of our question, we did not look into using classifier ML algorithms such as SVM or KNN during our work on the project
- First we started by plotting data by variables versus sea level for each season and region
  - Sea level versus CDD, T10, T90, RX5, and AVG

Random Forest Regression

- Tried Random Forest Regression…
- Did not produce any good results that were better than other regression
- Too much noise
- Did not like the variance in the regression
  - One time a graph could be positive than another time would be negative

Linear Regression

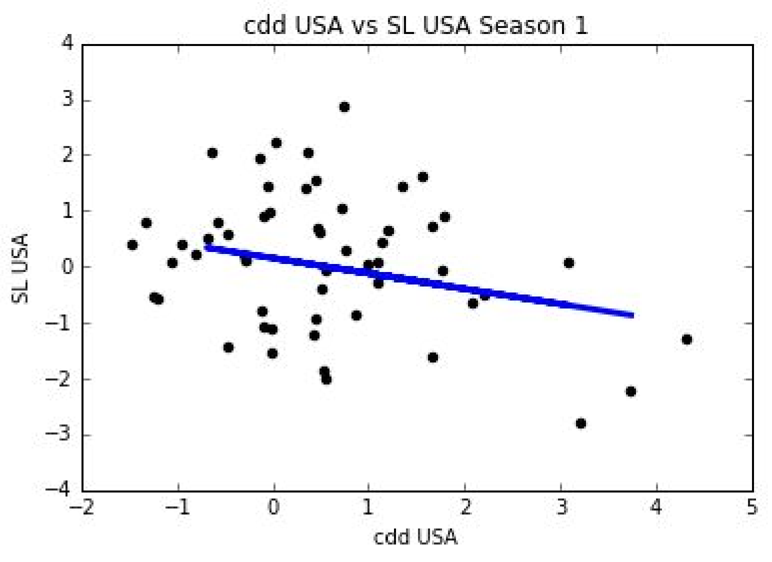
- Based on the graphs, we determined that there were linear trends in some data
- We implemented linear regression on the same set of data that we plotted earlier
  - Sea level versus the variables
  - Plotted the linear regression line on the graphs
- We examined the R^2 of the regression line

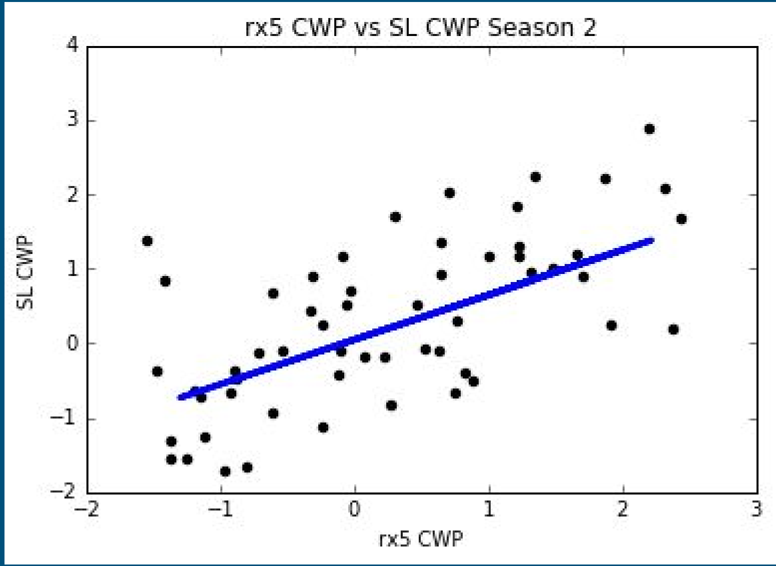
Results

Variable	Number of positive R^2 values
CDD	5
Rx5	12
T10	8
T90	3
AVG	4

- Most positive R^2 values found in Rx5
- Least positive R^2 values found in T90

Maximum R^2 Values

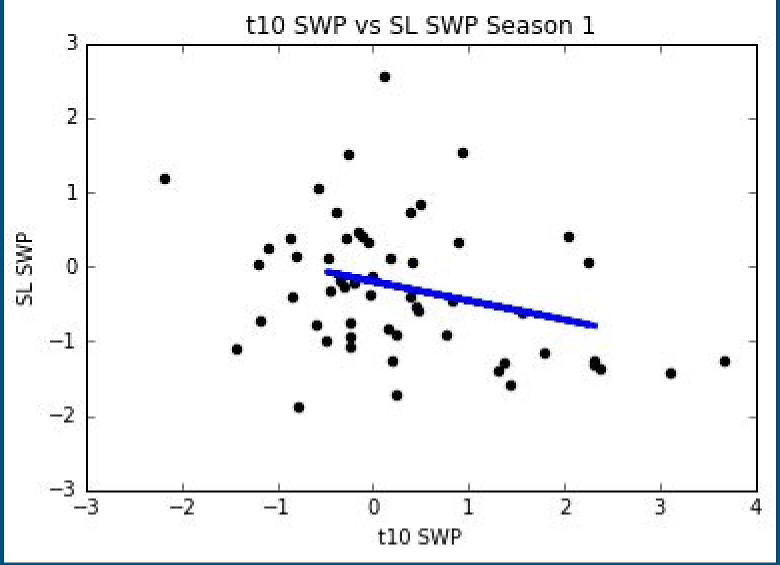


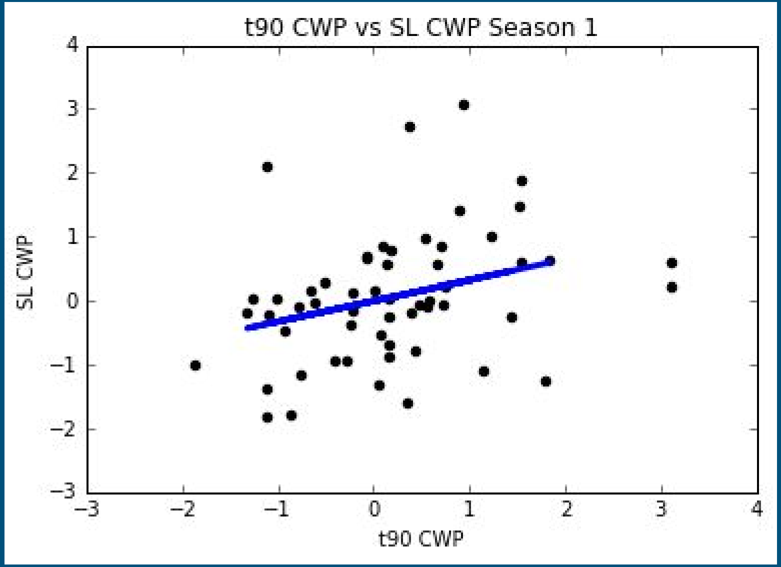


$R^2 = 0.085$

$R^2 = 0.602$

Maximum R^2 Values

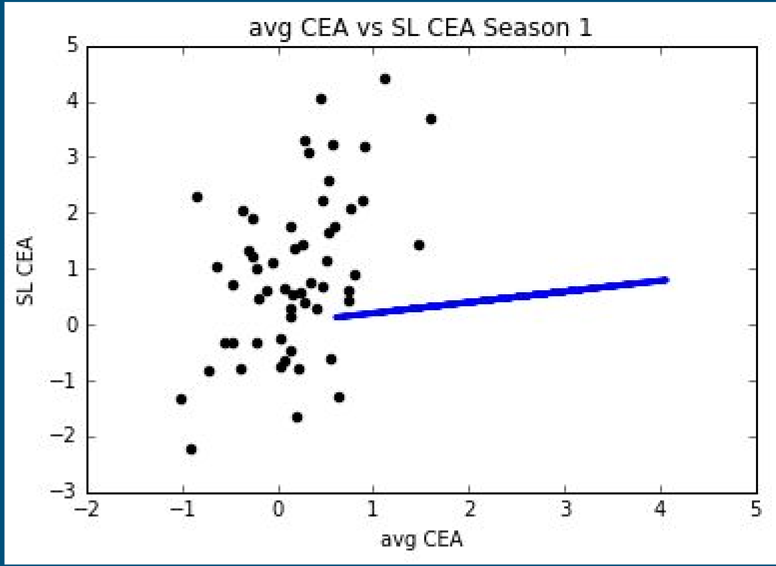




$R^2 = 0.19$

$R^2 = 0.081$

Maximum R^2 Values



$R^2 = 0.086$

Future Steps...

- Split testing and training set in a random fashion - our approach used the earliest 80% of data as the training set and the most recent 20% for testing
- Find ways to leverage classifier algorithms
  - This would require more specific questions to be asked
  - Could provide more powerful predictive models to answer these narrower questions
- Time-series manipulation
  - Particularly important to consider when using monthly data
  - So-called “sliding window data enrichment” could benefit classification attempts

Questions?