Project 2 Instructions

# Goal

To understand the complexities of gathering data to answer a question that might not necessarily have data readily available, designing and validating a regression model to predict defined outcomes and presenting model results effectively.

# Instructions

Come up with a question independent of data gathering, gather data, conduct EDA and develop a regression model that answers your question.

Using Rmarkdown develop a report of your project and publish to Rdocs so all students can see your work. Make sure your report answers at least the following questions:

1. How did you develop your question and what relevant research has already been completed on this topic?

- Interest in climate change data; looking at other analyses gave idea that Linear Regression to predict increase in GLOBAL (?) temperatures// CITIES is feasible  
- One slide Background of climate change

- Files will be GlobalLandTemperaturesByCity.csv and GlobalTemperatures.csv

1. How did you gather and prepare the data for analysis?

- Berkley data- look at how well populated data are for CONUS; make call on how far to go back.   
- Data - Impute

1. How did you select and determine the correct regression model to answer your question?  
   - Looking at plot, linear reg seems feasible
2. How reliable are your results?

- Analyzing model output AUC/ROC, R^2 etc

- Performance on test/train (accuracy)

1. What predictions can you make with your model? Examples

- Future global temperature

- Rates of rise in temp

1. What additional information or analysis might improve your model results or work to control limitations?

-TBD

Develop a presentation that provides an overview of your results, inclusive of the limitations and be prepared to demonstrate your knowledge in class. **10 slides max.**

# Deliverables

Please turn in your final copy of each of these items in Blackboard:

* Rmarkdown report answering the questions detailed organized in alignment with the Data Lifecycle.
* Any slides or resources used for the presentation

# Grading (See Rubric)

1. 33% Summary Report
2. 33% Model Development, Selection and Usage
3. 33% Presentation
4. Bonus points for use of github and google cloud

Slide Outline:

1. Question and why?
2. Existing Climate Change Research
3. Data Gathering, Impute/Prep methods
4. EDA
5. Maps
6. Regression Model
   1. Graph
   2. Summary Stats
7. Predictions
   1. On future years
   2. On test/train (reliability/accuracy)
8. Examples