Lucas Laughlin, Gregor Tzinov, Eugene Miller
Applied Regression

Regression Project Proposal

1. What are the data that you plan to work with?
US Accidents in the US (3 million records)
https://www.kaggle.com/sobhanmoosavi/us-accidents

2. Where did the data come from? Are they experimental or observational?
Observational data; recorded accidents on 49 different variables.

3. Why is this data interesting to you? What questions do you hope to answer about it?
This data is interesting as it relates to our everyday lives. We hope to answer what are some common characteristics of different types of accidents, from severe ones that create large traffic jams, to those that are specifically inflicting the driver side of the vehicle. We also want to see what factors most impact the severity of the accident.

4. What are the relationships between the variables? Does theory suggest that they are related in some way?
Severity and side struck are correlated, precipitation and humidity, humidity and severity are correlated, visibility and weather are correlated, visibility and severity are correlated.

5. What random components are there (e.g., measurement error)?
Whether the data was recorded correctly by first responders if there was an observer when the accident took place.

6. What prior research on your topic might be helpful to consider?
US Department of Transportation studies regarding collisions and accidents in America.

7. What methods might be useful in analyzing this data?
Categorical Predictors could prove very useful as there are many factor datatype variables in our data. We hope to show that we can predict whether there was rain based on severity and humidity. We'll also be using all of the methods learning in class to strengthen our predictions, and will also be open to researching additional methods that can be applied to our intended analysis as needed.