By: Branden Addicott

South Caroline Accident Severity

Data Set

- US Accident June 20
  - Kaggle: https://www.kaggle.com/sobhanmoosavi/us-accidents
  - 3.5 Million Observations
  - 42 variables: Data Dictionary

Data Cleaning

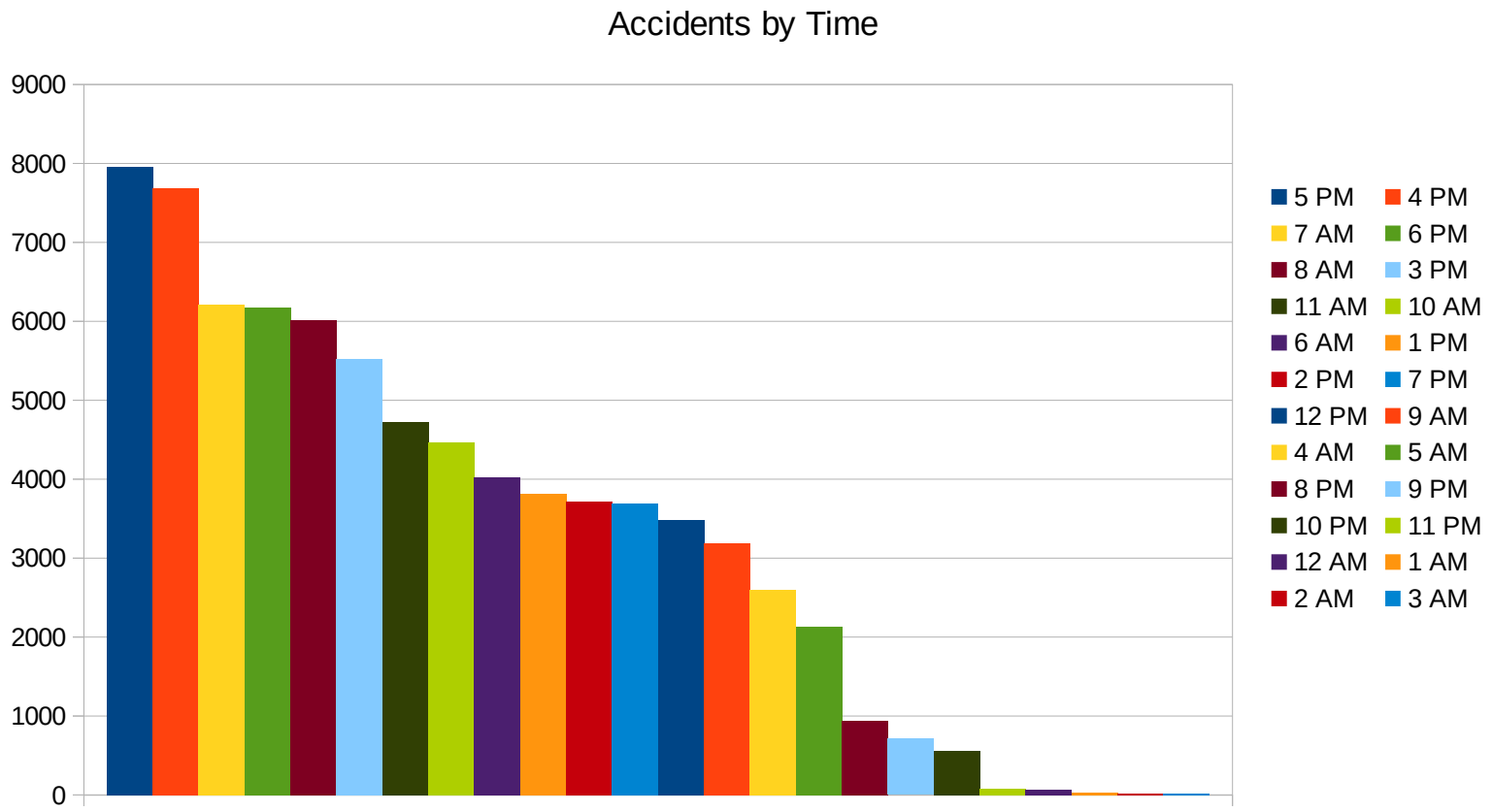
- Selected only observation for the state of South Carolina
- Retained 9 variables
  1) Severity
  2) Year
  3) Month
  4) Day of Week
  5) Hour
  6) Temperature
  7) Visibility
  8) Wind Speed
  9) Precipitation

Exploration

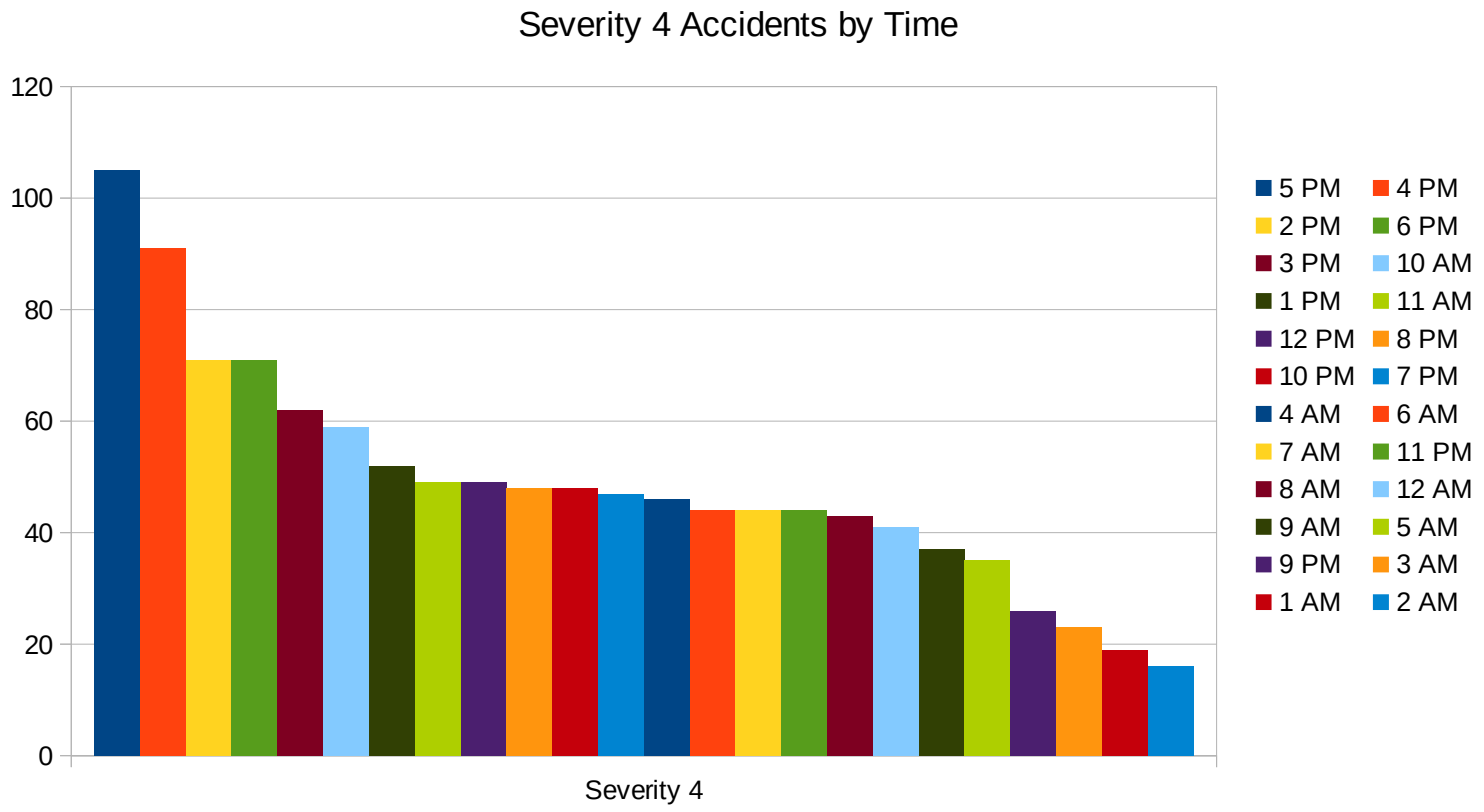
Record Counts and Percentages by Severity

Severity	Count	Percent of Total
1	116	0.07%
2	137,371	79.28%
3	34,620	19.98%
4	1170	0.68%

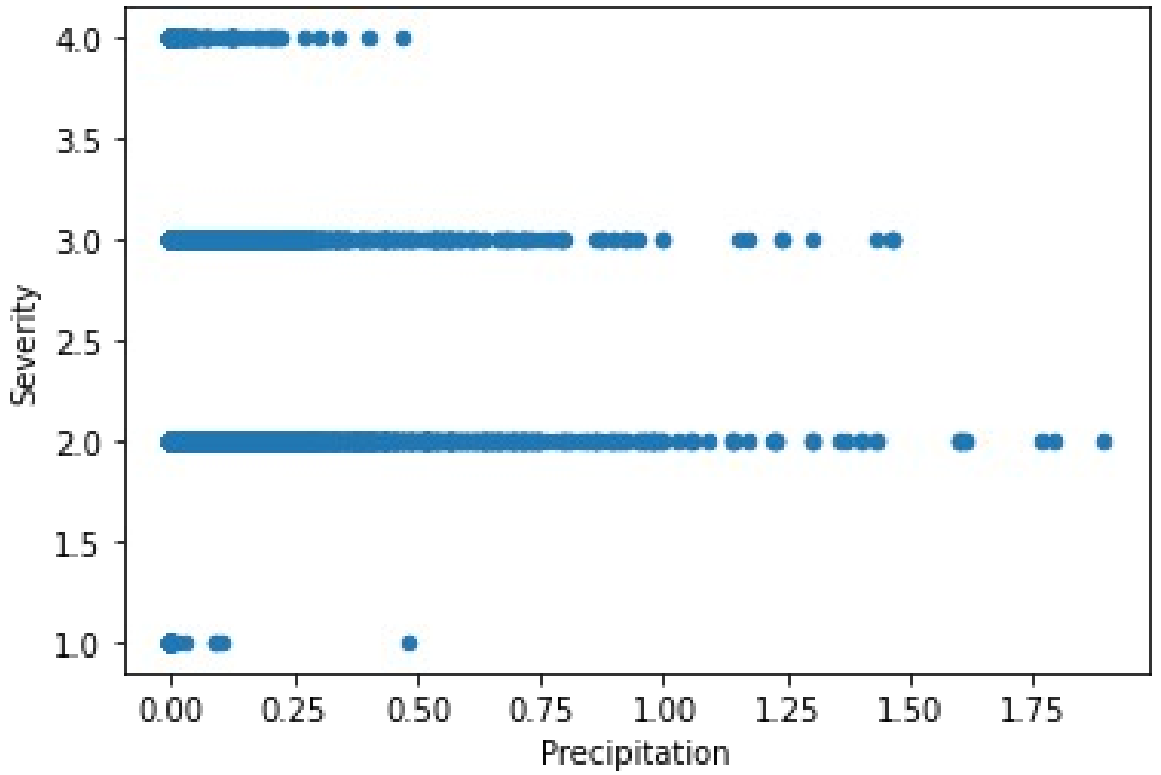
Accident Time



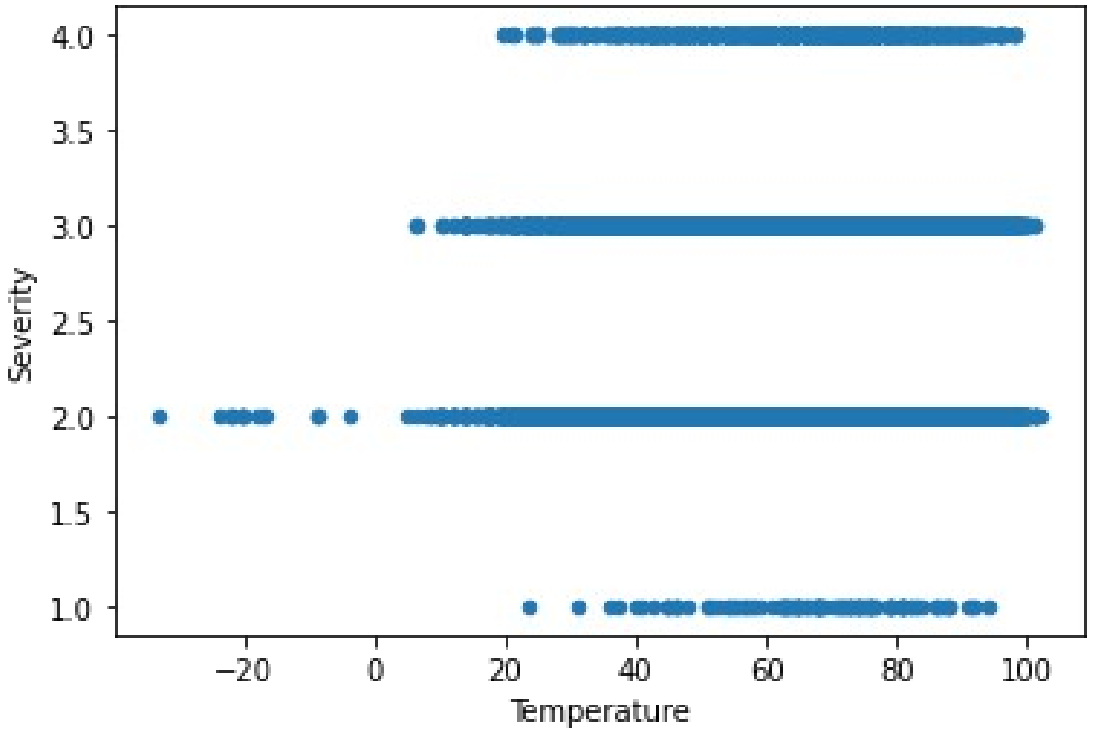
Sever Accidents by Time



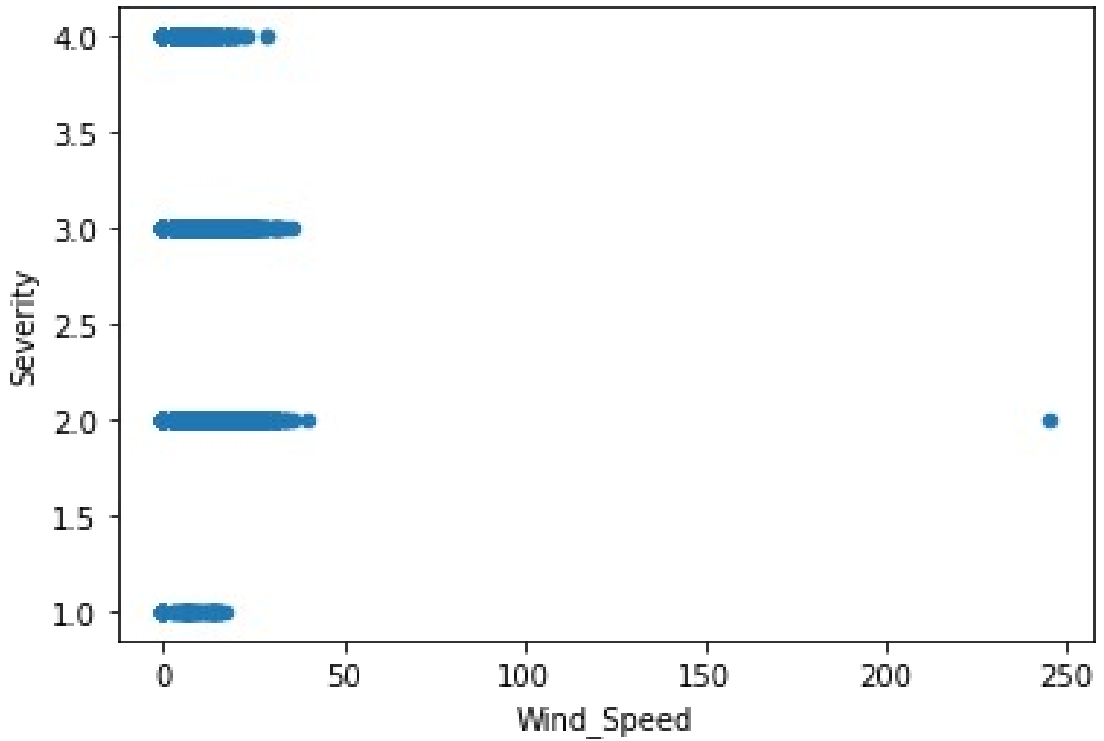
Precipitation Plot



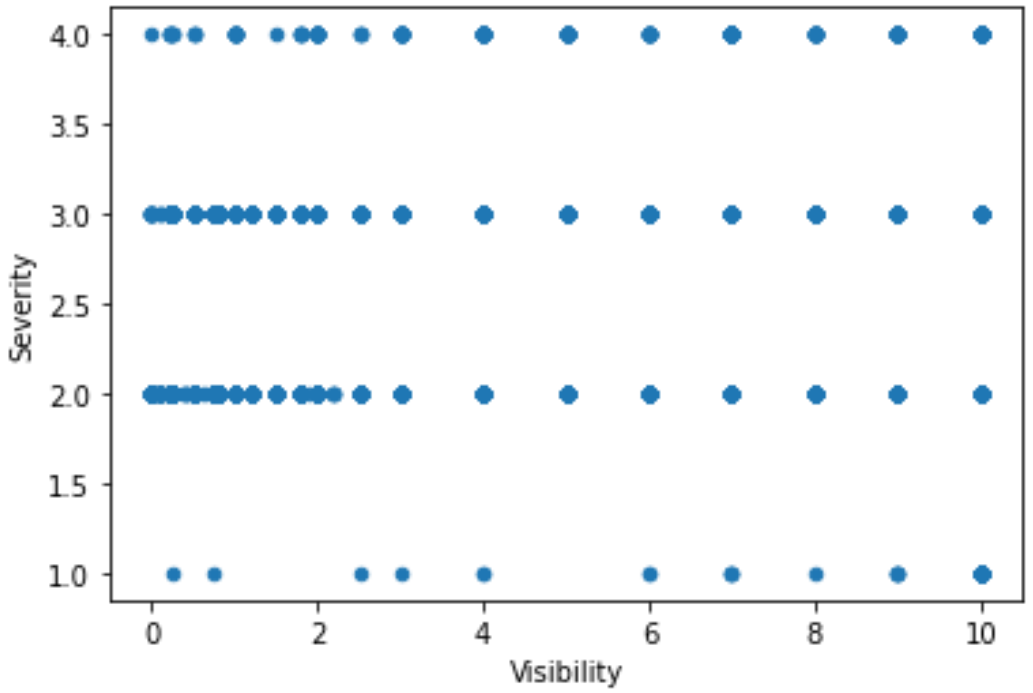
Temperature Plot



Wind Speed Plot



Visibility Plot



Linear Regression

- Not Significant
  - Dependent variable not continues
  - Lack of variation

Severity = 2.184 – 0.047 (Year) – 0.014 (Month) + 0.006 (Day) – 0.012 (Hour) + 0.013 (Temp) – 0.015 (Visibility) + 0.004 (Wind Speed) + 0.013 (Precipitation)

Decision Tree

- Split data set into train and test for tuning depth parameter
  - Optimal depth determined to be 9
- Re-fit will full data set
- Accuracy of 83%
- First Branch split on Year

Conclusion

- Most accidents occur during work commuting hours
- Sever accidents occur during the commuting hours also
- Accident severity has been decreasing with Time

Further Research

- Measurement and Accuracy of Severity
- Analyses Remaining States
- Identify factors of severity decrease over time