CONCLUSION

I have demonstrated a strategy for prediction of property losses due to tornado activity using NOAA tornado dataset. The modeling is done using a multiple linear regression model formulation where the loss sustainability is classified using classification model. The principal findings of the study include the following:

The logistic regression model proved useful for initial classification of the observations, producing a classification rate of about 76% percent on the validation data. This proved successful in the ability to predict positive values of property damage.

Similarly, the predictors were also able to estimate the dollar amount of property loss using multiple linear regression. Looking carefully we can see that the high RMSE is due to a few outliers producing high error residuals in the predictions of some of the calculated MSE's, where few number of error terms are severely dragging the average upward while randomly sampling.

After we have classified our test dataset we subset it by removing all the observations that we predicted had less than a 50% chance of having positive property damage. Using only the observations that we predict to have positive property damage, we use our linear regression model to predict the dollar amount. When predicting on the test dataset using the linear regression model we calculated a mean absolute error of 3.1 million, which was even less than our validation MAE. Since there are tornados included in the dataset that produced over a billion dollars in property loss, an average deviation of a few million dollars is reasonably adequate. Thus, the entire procedure seems to have significant predictive abilities for predicting property loss of future tornados.