INTRODUCTION

Traffic safety, which continues to remain a critical issue worldwide, has led to a myriad of modeling techniques to improve analytical capabilities with respect to crash modeling and prediction. The State and Metropolitan Transportation Planning processes must be consistent with Strategic Highway Safety Plans. This research aims to identify machine learning models and methods to improve the ability to capture variables that have the most significant impact on traffic safety through crash prediction with modeling parametric and nonparametric approaches. In machine learning models, we use different models for prediction and inference with the aim of minimizing the reducible error.

DATA PREPARATION

Crash data has been obtained from UMass Safe Data Warehouse and has been geocoded and provided with several attributes. This GIS layer represented the actual locations of crashes. Information including Roadway Characteristics, Inventory, traffic characteristics, demographic, and socioeconomic data have been collected from U.S. Census Bureau and the Office of Geographic Information (Mass GIS). The linked dataset allows for the creation of a model that directly relates level of service and level of safety. The final product creates a specific geometric elements can be manipulated to evaluate safety effects.

MODELS

Moving Beyond Linearity (GAM)

There are several methods that offer a lot of flexibility, without losing the ease and interpretability of linear models:
- Polynomial regression
- Step functions
- Regression and smoothing splines
- Local regression
- Generalized additive models (GAMs) which use
  - above models

In this model, we define separate \( f_j \) for each \( X_j \) in each TAZ, and then add together all of their contributions.

GAM are very nice and effective way to lead to interpretable Models: easily mix terms in GAMs, some linear and some Non Linear terms.

RESULTS

Total number of crashes in each TAZ in Springfield and Boston, MA.

Few variables were associated with both total and severe crashes (red color). Seven variables in particular were present in the top of the variable rankings in terms of increasing for both total and severe crashes. The effect of these variables should be considered while developing a strategy for improving the safety of a zone. For example, a TAZ with higher number of AADTs can be prioritized for allocating funds for safety treatment. If necessary, TAZs with higher lengths of roadway may be scrutinized carefully by transportation officials to reduce severe crashes as well as total crashes. Alternatives such as installing speed-calming devices or lowering the speed limit may be additionally taken into account to improve safety.

DISCUSSION

Output of this study can improve modeling of transportation safety planning at the macro level. Using machine learning models to improve prediction of traffic crashes and improve safety with considering the boundary analysis can make these predictions more accurate. Development of new models to analyze drivers’ contribution with developing policies regarding these analyses can decrease the crashes and using variables that have not traditionally been used in previous studies can improve the TSP.