Final Project Proposal
Due: 5/8/18

Jason Goode
Jordan Zandi

Street Safety in NYC:
Studying the Effectiveness of the VisionZero Program for Reducing Traffic-related Fatalities/Injuries

Introduction: In 2016, traffic-related accidents resulted in 186 fatalities and 41,466 injuries in New York City. Mayor Bill DeBlasio has made reducing the number of injuries and fatalities a priority of his administration. The main initiative to achieve this goal, "Vision Zero", offers a rich dataset to observe the relationship between various street design techniques and traffic outcomes. Our objective with this project is to study the relationship between said safety measures (e.g. speed bumps, bike lanes) in order better to understand the effectiveness of each initiative. Our hypothesis is that the NYCDOT's traffic safety initiatives (detailed below) have a statistically significant reductive impact on the odds of fatality and the incidence of injury.

Data: The majority of our data will come from the NYC DOT's *Vision Zero* data feed and visualization tool. It contains four CSV files on crash data along with 19 additional GeoJSON files indicating the locations of the street safety initiatives. We also plan to pull weather data from Kaggle, as we hypothesize that it will have an impact on crash outcomes.

Our number of samples for 2016 crash data is 41,652, with 41,466 injuries and 186 fatalities. The outcomes are further specified by injury type (motor vehicle, bicycle, pedestrian). A full list of our predictors is as follows: Arterial Slow Zones, Bike Priority Districts, Enhanced Crossings, Leading Pedestrian Interval, Left Turn Traffic Calming, Neighborhood Slow Zones, Safe Streets For Seniors, Speed Humps, Street Improvement Projects (Intersections), Street Improvement Projects (Corridors), VZ Priority Corridors, VZ Priority Intersections, VZ Priority Zones, 25mph Signal Timing, Speed Limit, Weather, and Month.

Methods: The main data-wrangling challenge of this project will be to cross-reference street attribute data (GeoJSON) with the crash data (CSV), and determine whether or not an accident location also has a given street attribute. We plan to use Python's GeoPandas module in order to render the GeoJSON file entries into dataframes and identify points of intersection. This dataframe will then be formatted in such a way that we can run a multiple logistic regression in R using the various predictors in order to generate our training model. We will use a 70/30 split of the data for training and testing. From this model we will be able to observe the effectiveness of implementing various street safety measures and their impact on the odds of fatality and injury.

Timeline:

- **Data Cleaning** (May 7 - May 18):
  - Parse GeoJSON Data into Dataframes: Jason
  - Clean Crash CSVs, Weather Data, and join with parsed GeoJSON: Jordan
- **Modeling/Interpretation**: May 19 - May 25
  - Prepare & Run Logistic Regressions: Jason
  - Perform Model Analysis/Interpretation: Jordan
- **Presentation Preparation**: May 26 - May 29
  - Create Outline: Jordan
  - Make Visualizations: Jason
  - Interpretation/Construction: Jason & Jordan