Take Home Data Analysis Problem

As a low-cost safety solution, rumble strips have been installed at many state-owned highways over the past decade. Rumble strips are engineered to alert the inattentive drivers through audible and vibratory warning as the vehicle strays across the center or edge line, and hence to improve the chance for a safe return to the current driveway.

The dataset "PennCrash.txt" (`https://stat.duke.edu/~fl35/data/PennCrash.txt`) provides number of crashes occurred on about 2000 road segments (sites) in Pennsylvania in the year 2004, 2008 and 2012. Among these sites, rumble strips were installed in 331 of them between year 2008 and 2012. Each row represents a site.

Below is the list of the variables for each site:

Variable	Definition
Treat	Rumble strip was installed between 2008 and 2012 (1:yes; 0:no)
totxx	Total number of crash in year xx (04, 08, 12)
adtxx	Annual average daily traffic (AADT) volume in year xx (unit is vehicles per day)
length	Roadway segment length in miles
width	Pavement width in feet as three categories: width ≤ 20; $20 <$ width < 24; $24 \leq$ width
speed	Posted speed limit: less than or equal to 45 mph; greater than 45 mph
shoulder	Average shoulder width in feet: width ≤ 3; $3 <$ width ≤ 6; $6 <$ width
driveways	Number of driveways is categorical: none ; $0 <$ # driveways ≤ 10; $10 <$ # driveways
intersections	Inclusion of intersections: no intersection; at least 1 intersection
curves	Existence of horizontal curves: no curves; at least 1 curve
curvature	average degree of curvature

Analyze the data to address the following question: Are rumble strips effective in reducing crashes?

Write a report (maximum 3 pages) based on your analysis describing your findings. Be thorough in your exploratory analyses and exploratory use of models; applied work that overly emphasizes complicated modeling to begin is often less valuable than careful, incisive evaluation of data through simpler, exploratory models– at least to begin.

Your report should discuss all relevant aspects of your analysis (exploratory and modeling) with graphical and numerical summaries that are important for communicating results.

Take-home Applied Exam

- Present your results in a three page (maximum) report addressing the primary questions posed. Keep your answers concise and to the point.

- You may include code and other plots in a supplemental appendix; BUT, you should not assume that graders will read beyond the main report; all relevant material should be within the three page limit.

- You may use all notes, books, software etc from courses and studies to date, and build on your cumulated experience in applied modeling and data analysis.

- You may freely use other resources– code, literature, etc– from whatever source you like, so long as you do not violate condition 4 below.

- **To Confirm**– you are also bound by this honor pledge and must sign below to confirm this:

    1. I confirm that this Take-home Exam submission is my work alone.
    2. I have not consulted at all with any other students, whether they are taking the exam or not.
    3. I have not copied nor adapted the work of others, nor provided help or advice to others on this exam.
    4. I have not sought out or used any external sources (past student projects, publications, web sites, etc) that explicitly address any aspects of the specific data set and applied problem here. In particular, I have not used web searches to find previous references to the data and earlier analyses of this specific data set and problem, of any kind.

- Sign below and hand this in with your solution before or at 12pm, May 10 (Thursday) 2018 to Lori Rauch at Room 214, Old Chem.

Name:

Signature:

Date: May 10, 2018