Trials and Tribulations of Projecting 2D Car Crash Data onto 1 Dimension

Paul Gaona
Scout Jarman
Kenneth Pomeyie
Tom Kerby
Department of Mathematics and Statistics
Utah State University

April 7, 2022

Outline

Problem Motivation

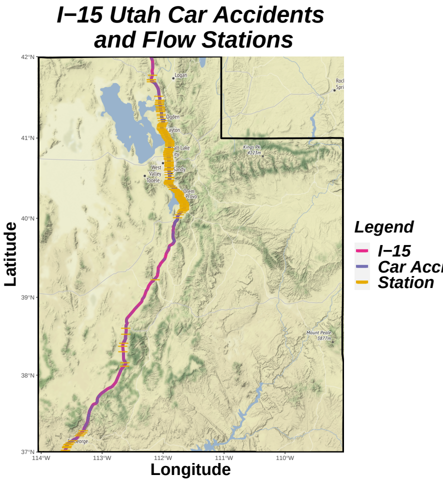
Can we determine if there are parts of I-15 that are more or less dangerous than others?

Data Sources:

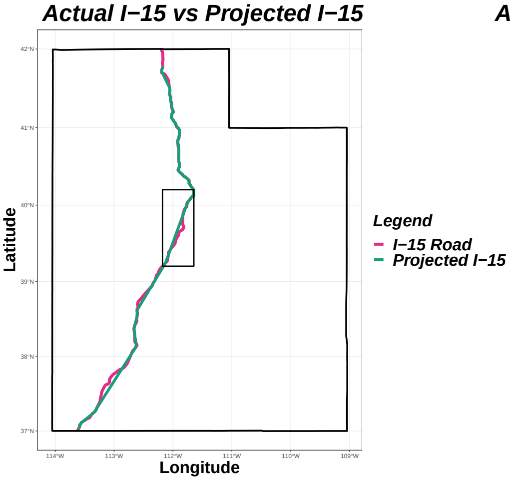
- Car Crash Data - Utah Department of Public Safety
  - Latitude Longitude of accident
- I-15 Data - Performance Measurement System (PeMS)
  - Flow (# Cars per hour)
  - Station Postmile
  - Conversion table from P.M. to lon/lat

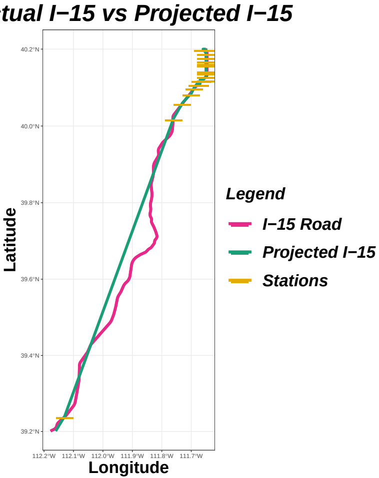
Data Cleaning

- Combine north bound and south bound data
- Aggregate station data when they are located in the same place
- Convert from postmiles to longitude and latitude coordinates

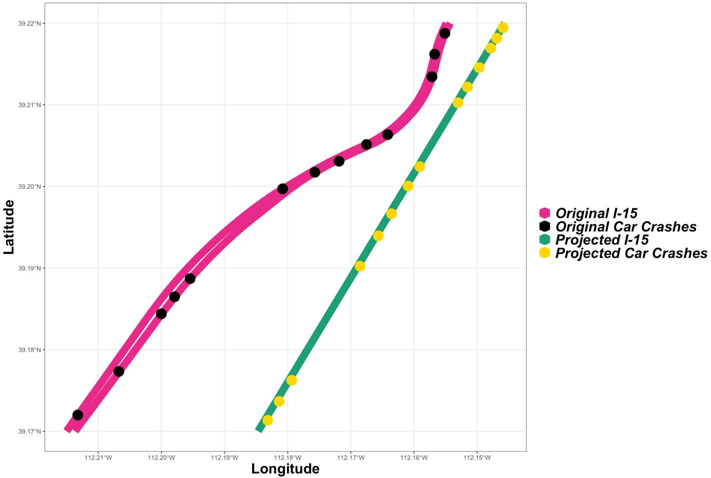


Create a Single Line string for I-15





Projected vs Unprojected Car Crashes

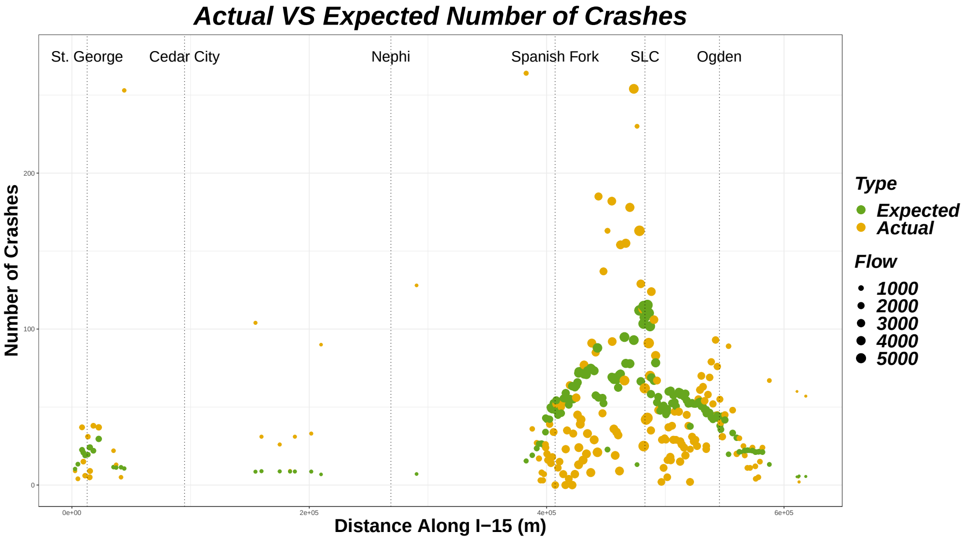


Null Hypothesis

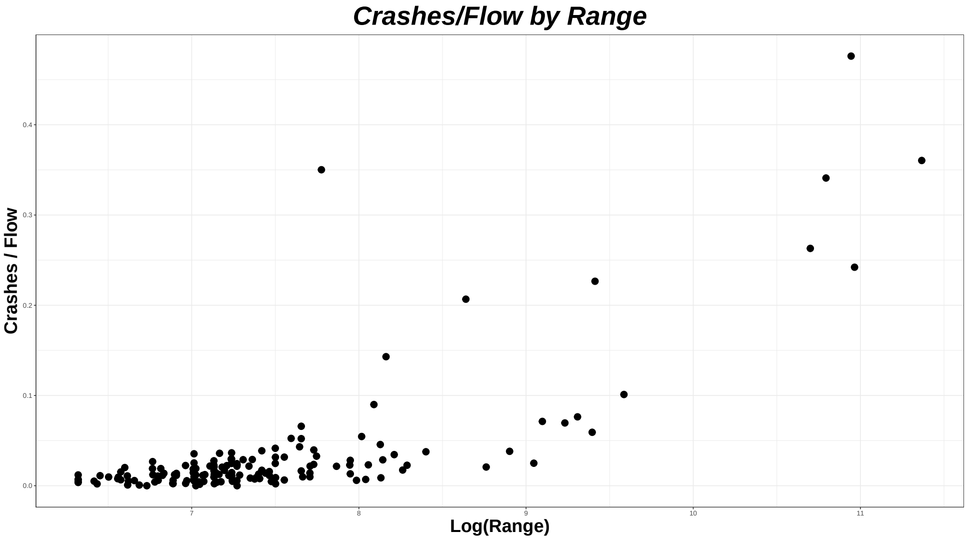
- H_0: Car crashes are uniformly distributed according to station flow.
- We will find the expected number of crashes per station, and conduct a χ^2 test to see if there is a difference.
- Let f_i be the flow for station i, and let n_i be the number of crashes at station i, and let N be the total number of stations.
- For a station s, the expected number of crashes will be:

$$E_s = \left(\sum_{i=1}^{N} n_i\right) \cdot \frac{f_s}{\sum_{i=1}^{N} f_i}$$

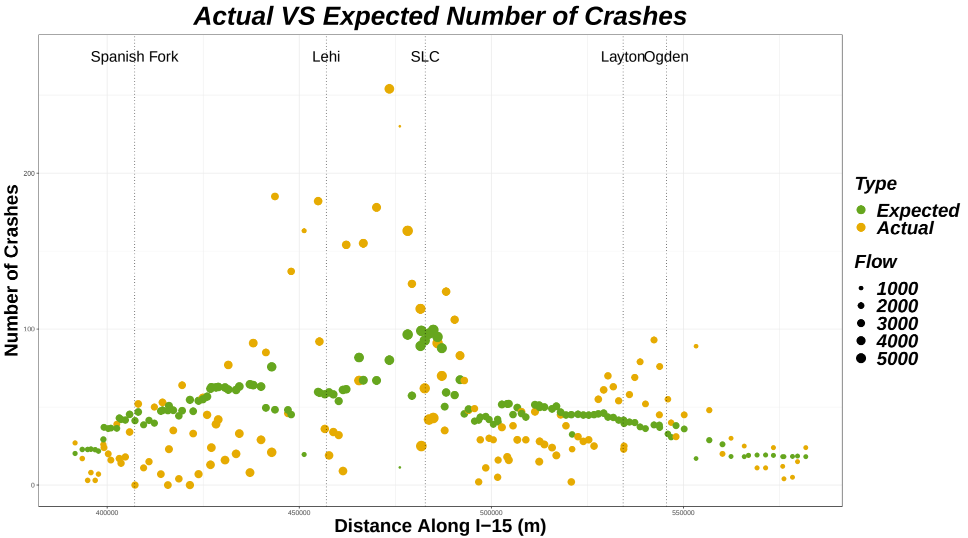
All Stations, $\chi^2 = 23323$, p-value ≈ 0



Effect of Station Range



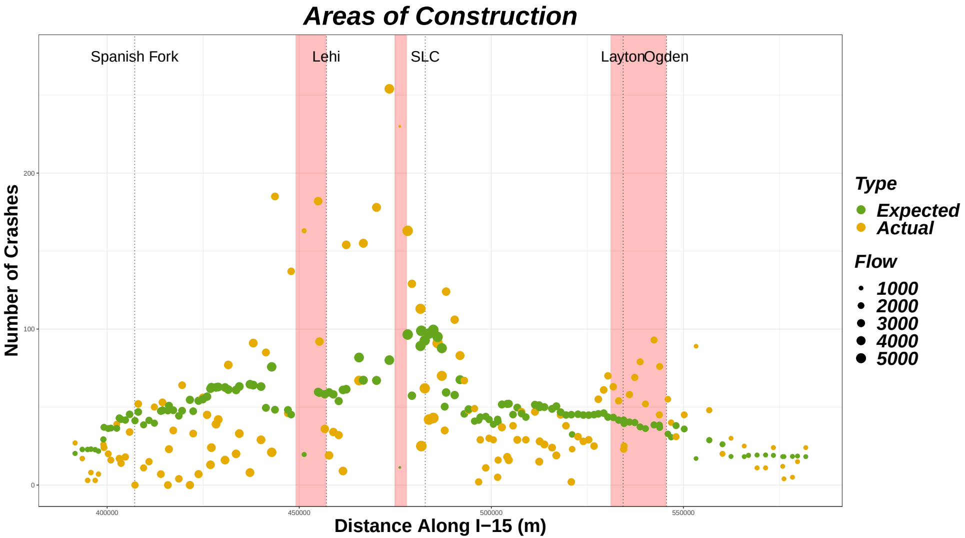
Wasatch Front: $\chi^2 = 9183$, p-value≈ 0



Results Summary

- Considering all stations, the test is very significant ($\chi^2 = 23323, p \approx 0$).
- The range of a station plays a role after accounting for flow, so we only consider the Wasatch Front.
- Consider the Wasatch Front, the test is still significant ($\chi^2 = 9183, p \approx 0$).
- We conclude there are more factors that influence the number of crashes on I-15 besides how busy the road is (flow).

Road Work


Areas of Construction
Spanish Fork
Lehi
SLC
LaytonOgden
Number of Crashes
0
100
200
400000
450000
500000
550000
Distance Along I−15 (m)
Type
Expected
Actual
Flow
1000
2000
3000
4000
5000

Future Work

- Improve data quality (find better data source), including a better representation of I-15, and having a smoother representation of flow.
- Fully investigate and model the dynamics between flow, range of station, and the number of crashes.
- Have a better way to find which stations are particularly off.
- What is the effect of direction of travel (Northbound/Southbound), what about time of travel?