# How R Helps to Evaluate the Safety Performance of a Metropolitan Highway Network

Peter Fortunato, Data Scientist | 2022 | OKI

## Introduction

The OKI Regional Council of Governments is the metropolitan planning organization (MPO) for Greater Cincinnati. Our primary function is to carry out our region's transportation planning process which involves allocating federal and state funds for transportation improvement projects. We accomplish this by using our award-winning project prioritization process to guide us in deciding which projects get funded.

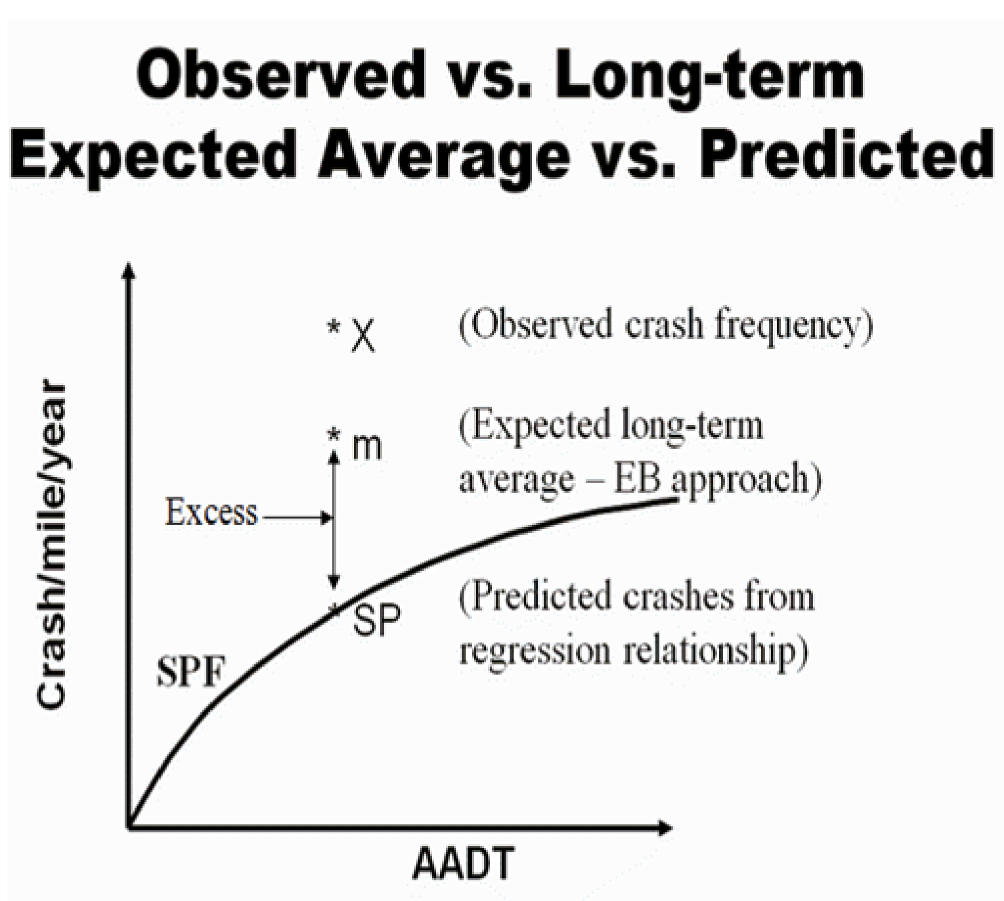
Roadway safety is a critically important component to improving a network, and here we demonstrate how to use R to develop **Safety Performance Functions (SPFs)** in order to output the safety metric **Excess Expected Crashes (EECs).**

### What are SPFs?

**Safety Performance Functions** are negative binomial regression models that output EEC's.

There are a variety of metrics used by transportation officials to evaluate roadway safety performance. The data the metrics use is gathered from the networks themselves and include physical attributes (number of lanes, urban or rural, stop control), roadway statistics (traffic volume, functional class), and crash data (both nominal and discrete). From this, crash rates are the most common evaluation metric and a proxy for roadway safety. However, they have two primary issues: there is no account for statistical biases or ability to differentiate roadway features and geography.

A more statistically robust safety performance metric can be developed using count regression models known as SPFs. These are negative binomial regression models that model crashes against traffic volume and length (for segments) plus any additional features that an agency has data for.

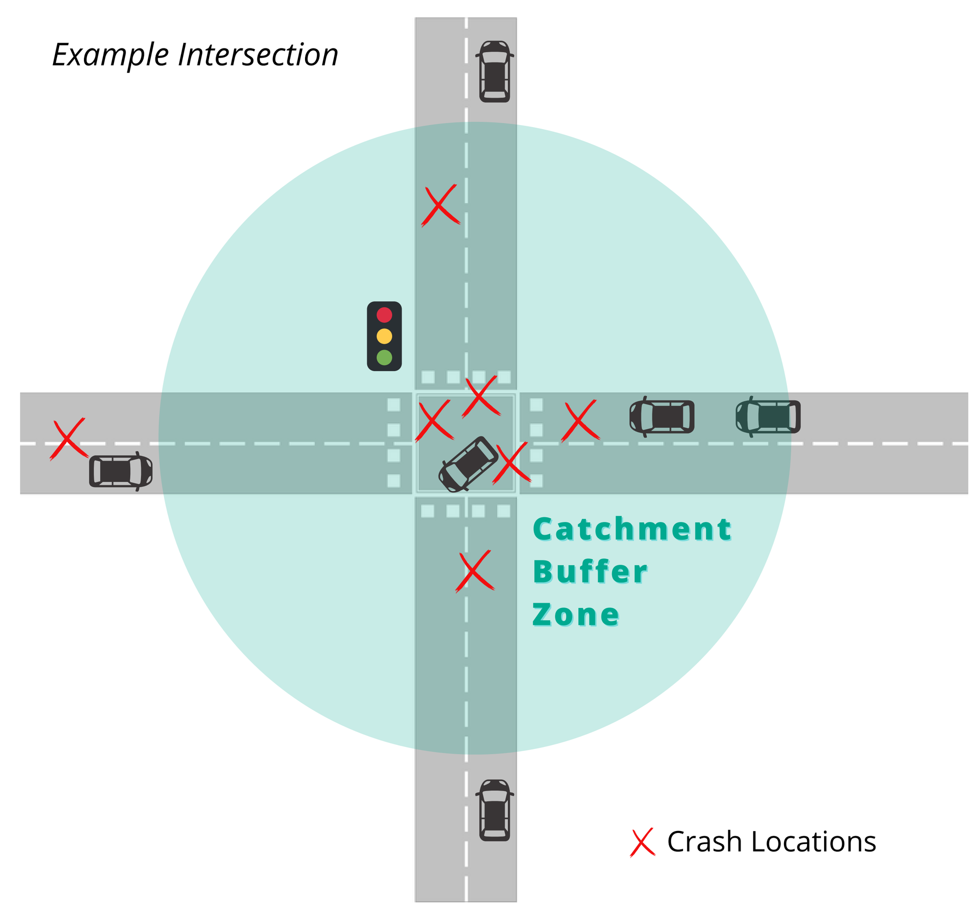


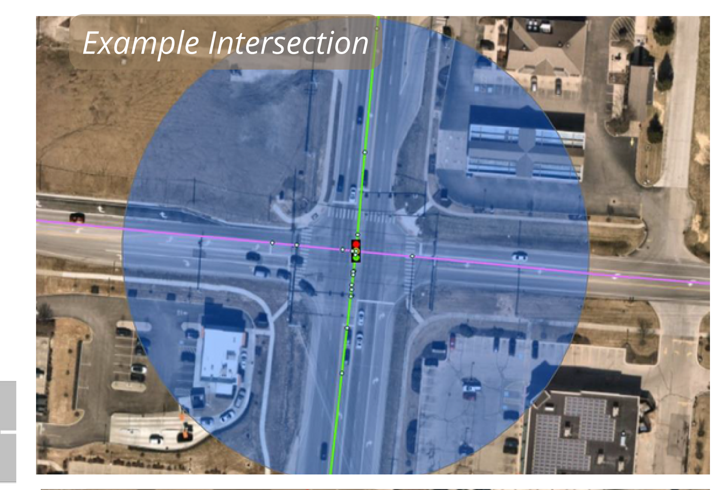

SPFs have the advantage of applying models to roadways with similar site types, i.e. 3-legged rural intersections with a minor stop control are modeled separately from 4-legged signalized urban sites. While the initial output of the model is predicted average crashes, the **Empirical Bayes** method can be applied to derive both expected average crashes as well as **Excess Expected Crashes (EEC)**, the difference between expected and predicted crashes.

Agencies can either calibrate pre-calculated SPFs from the Highway Safety Manual (HSM) or develop their own jurisdiction-specific SPFs. SPFs can screen a whole network for high-potential sites or to determine the safety impacts of different site improvements at the project level. For OKI's network screening, we developed our own SPFs using R & RStudio with ArcGIS Pro assisting with geospatial engineering and preprocessing.

The geospatial engineering component of this analysis began by geolocating crashes to the network segments and intersections. Next, traffic control features (such as one-way directions and two-way left turn lanes for segments and stop control for intersections) were taken into account. Afterwards, crashes were assigned to intersections using a catchment buffer around each intersection.

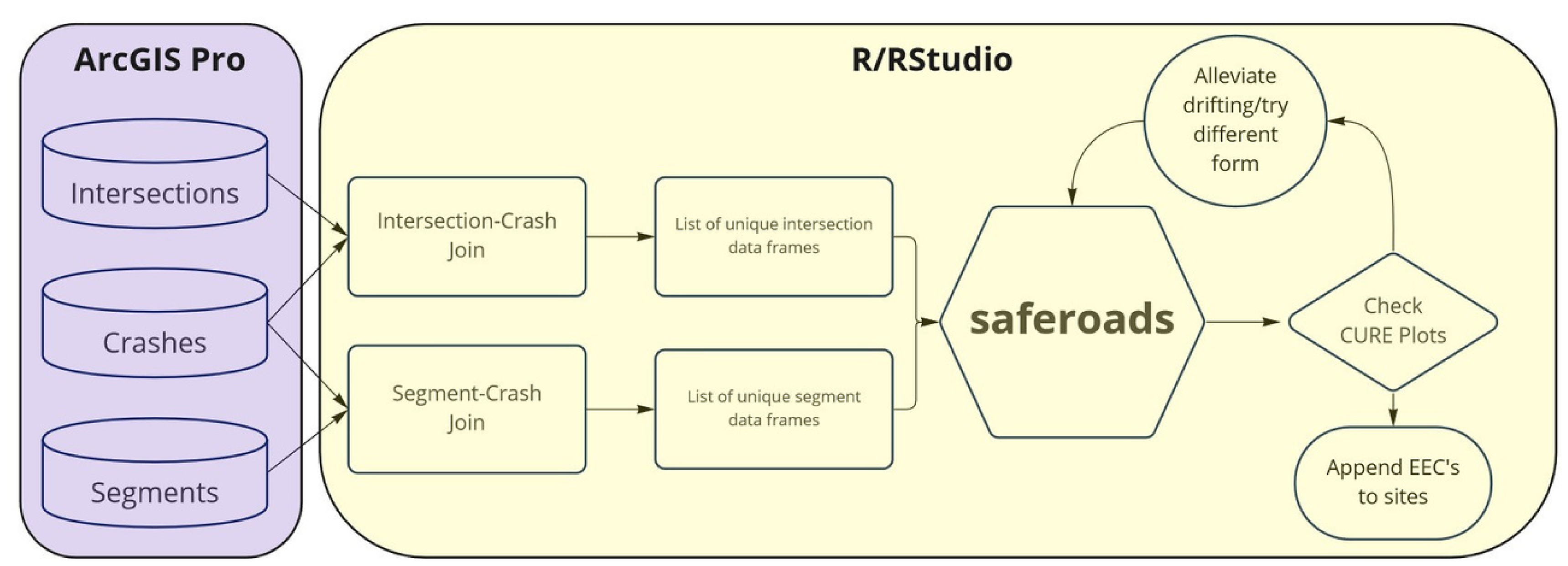
*Note: all intersection crashes were also assigned to their adjacent segment legs per the HSM guidelines stating a segment's length is measured between the centers of two intersections.*





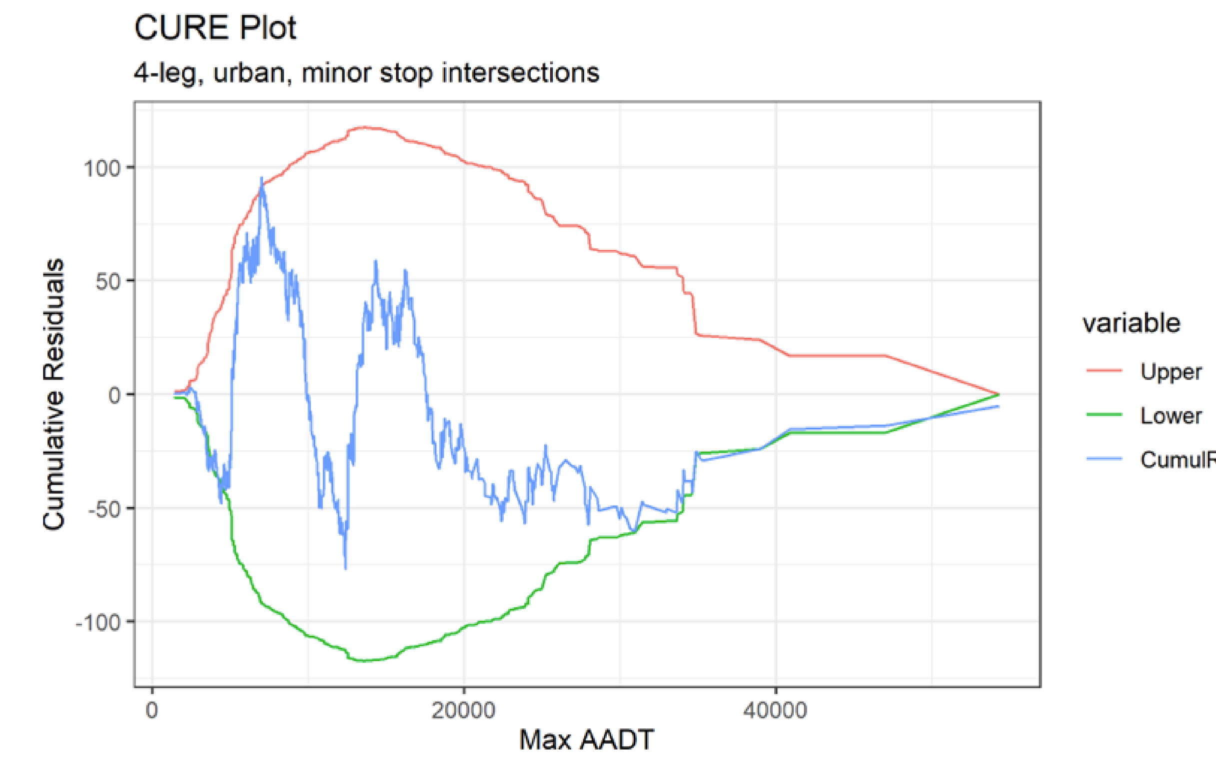
Once the segment, intersection and crash databases have been preprocessed, the remainder of the analysis is completed in R.




### What's a *CURE* plot?

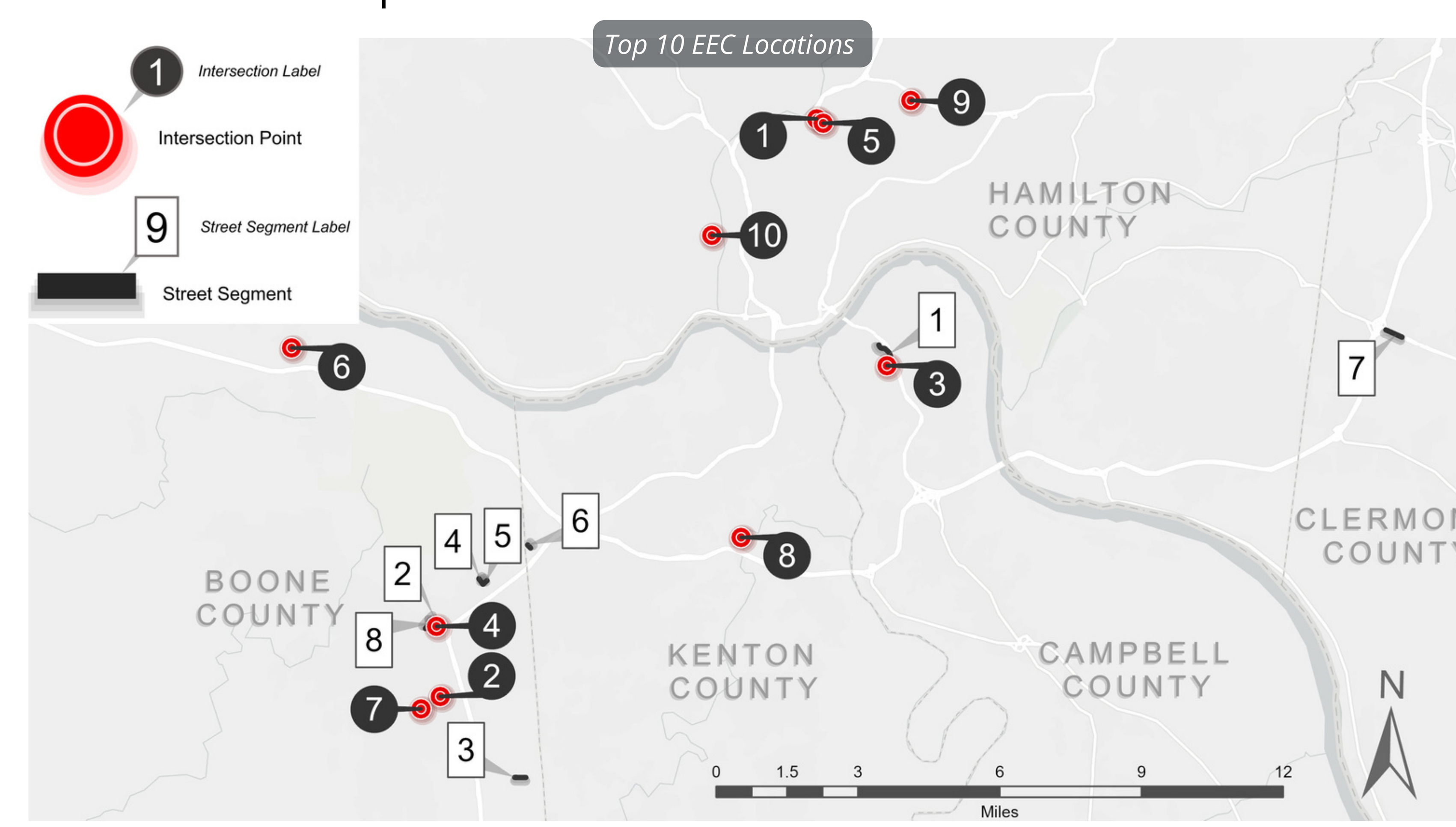
**Cu**mulative **Re**sidual Plots is the clearest evaluator of SPF performance.

## Results




The **dplyr** package aided in joining datasets together and creating nested data frames for each site-type combo using the `group_by()` and `nest()` functions (for ensuring SPFs are applied to unique site types). The **MASS** package provided the `glm.nb()` function to peform negative binomial regression on these datasets, while **purrr** helped with iterative mapping of the model to the nested data frames. And **ggplot2** helped with outputting CURE plots which easily identify how well an SPF is performing: the cumulative residuals should oscillate about y=0 while staying within the upper and lower confidence levels. The plot above indicates that the SPF for the particular site type is performing well.

Finally, using the predicted values from the SPFs in combination with the Empirical Bayes method, we are able to produce the EEC's which communicate the potential for a reduction in crashes.




EECs provide a more meaningful metric as it accounts for Regression-to-the-Mean bias and communicates a more actionable threshold for comparison. This analysis enables agencies to better understand why crashes happened where they do and how to best to best allocate safety improvement resources.