Research question

* What impacts the likelihood of a car accident involving the injury or fatality of a cyclist?
  + Location of accident
  + Which side of road bike is on (with traffic or against traffic)
  + Weather, road conditions(**rdconditio**)
* Is there an increase in bike accidents around the fall daylight savings time? (This is getting to the impact of daylight/time of day on frequency of bike accidents)

Rows and Columns

* Each row represents an accident involving a cyclist
* Each column represents different pieces of information about the accidents
* **Objected** = identification number for each individual accident

Aim of my visualizations

* Compare frequency of bike accidents to time of day, season, daylights savings time
* Frequency of bike accidents at type of roadway so whether their at intersection or just passing (crash type)
* Bike accidents by gender
* Frequency of bike accidents involving driver alcohol
* Frequency of bike accidents involving bike alcohol
* Determine the most dangerous intersection
* Determine most dangerous time of day for riding your bicycle
* Determine the city or region or place where the most bike accidents happen
* cyclist age v bike accident
* frequency of bike accidents or injury versus **bikedir**
* compare driver speed and speed limit and frequency of accident or bike injury

|  |  |
| --- | --- |
| **Categorical** | **Quantitative** |
| **Crashsevr**  **Bikedir**  **Traffcntrl**  **Crashtype**  **Crashgrp**  **Bikeinjury**  **County**  **Weather**  **Lightcond**  **Drvragegrp**  **bikeage** | **Crashtime**  **crashmonth** |

Notes

EDA

Structure: the “shape” of the data file  
Granularity: how fine/coarse is each datum

* Fine grained data

Scope: how (in)complete is the data  
Temporality: how is the data situated in time  
Faithfulness: how well does the data capture “reality”

Notes on Variables

Questions about columns

* + What is the difference between **locality** and **development**

Unknown columns

* + **Rteinvdcd**
  + **Numbikesci**
  + **Dsctncmifrm**
  + **Numbikesui**
  + **Numbikeski**
  + **Numbikesto**
  + **Numbikesai**

Changed variable types

* + Changed **crashtime** from military time and nominal data type to 12-hour clock and continuous data type
  + Changed age from nominal to continuous

Column Changes

* Created a combined column with **crashdate** and **crashtime**
* Recoded lightcond to lightcond 2 & did this to help with new column DaylightSavingsTime
* Created new columns for **crash day**  and **crash month** and then a **daylightSavingsTime**  with an if condition