Leeds Accident Data – 2015

Part II

Michael Sloan

Department of Math and Computer Science

Lewis University

Romeoville, Illinois

michaeldsloan@lewisu.edu

Abstract - This report contains visualizations of data regarding accidents reported in 2015 for the city of Leeds in the United Kingdom and the immediate surrounding area.

# Introduction

This paper serves two purposes. It provides insight into characteristics of accidents occurring in and about the area of Leeds in the United Kingdom during the year of 2015. Additionally, it provides a demonstration the author’s use of data visualization techniques acquired in the Lewis University Data Visualization course (CPSC-53000-002-FA16). The author completed this report for submission on December 18th, 2016.

## Domain of Interest

Auto accidents regardless of injury impact accident participants directly and indirectly. Beyond the direct economic impact on victims, every accident adds additional costs to non-participants as well. For example, increased costs increase insurance rates and delays in local traffic increase transportation costs.

Injury adds layers of costs beyond that of property damage. Given the open-ended costs of medical treatment, as well as the emotional burden born by accident victims, understanding the nature of injuries and the contributing variables helps individuals who manage risks and outcomes do so more efficiently.

The Leeds UK data provides a peek into injury outcomes and helps explain factors that coincide with those outcomes. The visualizations in this document assist in providing insights that help explain those factors.

## Approach

This report serves its first purpose by providing visualizations and interpretations for the reader. It serves the second purpose by showing both new and original visualizations with appropriate narration.

The Author used Processing 3 to extract data and produce all visualizations in this report except Fig 4. Injuries Per Month.

## Structure of This Document

This paper initially provides a description of the dataset used for investigative purposes. A sequence of visualizations follows this description in the section *Figures and Explanation.* This document presents each new visualization with appropriate narrative and then displays the first visualization with a narrative describing the motivation for the changes.

# Dataset

The Leeds City Council provides the Road Traffic Accidents dataset [1] under the Open Government License[2].

This dataset covers years 2009 through 2015. The data provider packaged as a single zip file with all years, or simply as CSV files of separate years. This analysis used only accidents occurring in 2015, and all data points experienced some degree of bodily injury. The data set used contains 2664 records representing injuries resulting from automobile accidents near Leeds, UK.

Table I provides the schema for the dataset and the purpose of the features within it and appears on the following page.

| **Field Name** | **Schema Metadata** | | |
| --- | --- | --- | --- |
| ***Type*** | ***Category*** | ***Purpose*** |
| Grid Ref: Easting | Numeric-Continuous | Feature | Geographic X Axis |
| Grid Ref: Northing | Numeric-Continuous | Feature | Geographic Y Axis |
| Number of Vehicles | Numeric-Discrete | Feature | Tche total vehicles involved in the accident. |
| Time (24 hr) | Numeric-Continuous | Feature | Time of the incident, |
| 1st Road Class | Nominal | Feature | Classification of Road Quality |
| Road Surface | Nominal | Feature | Road Surface at Time of Accident |
| Lighting Conditions | Nominal | Feature | Lighting Conditions at Time of Accident |
| Weather Conditions | Nominal | Feature | Weather Conditions at Time of Accident |
| Casualty Class | Nominal | Feature | The Role of Injured in Accident (Driver, Passenger, etcectera.) |
| Casualty Severity | Nominal | Feature | Degree of Injury: Slight, Serious, Fatal |
| Sex of Casualty | Nominal | Feature | Male of Female |
| Age of Casualty | Numeric - Continuous | Feature | Age of Casualty in Years |
| Type of Vehicle | Nominal | Feature | Vehicle Class |
| Reference Number | Discrete ID | Meta | ID of the Accident, 1 to Many relationships between Accident and injury observation. |
| Accident Date | Date | Meta | Trhe Date of the Accident |

Table

Leeds Accident Data Set

# Figures and Explanation

## Table of Visualizations

This document includes the following visualizations:

Table

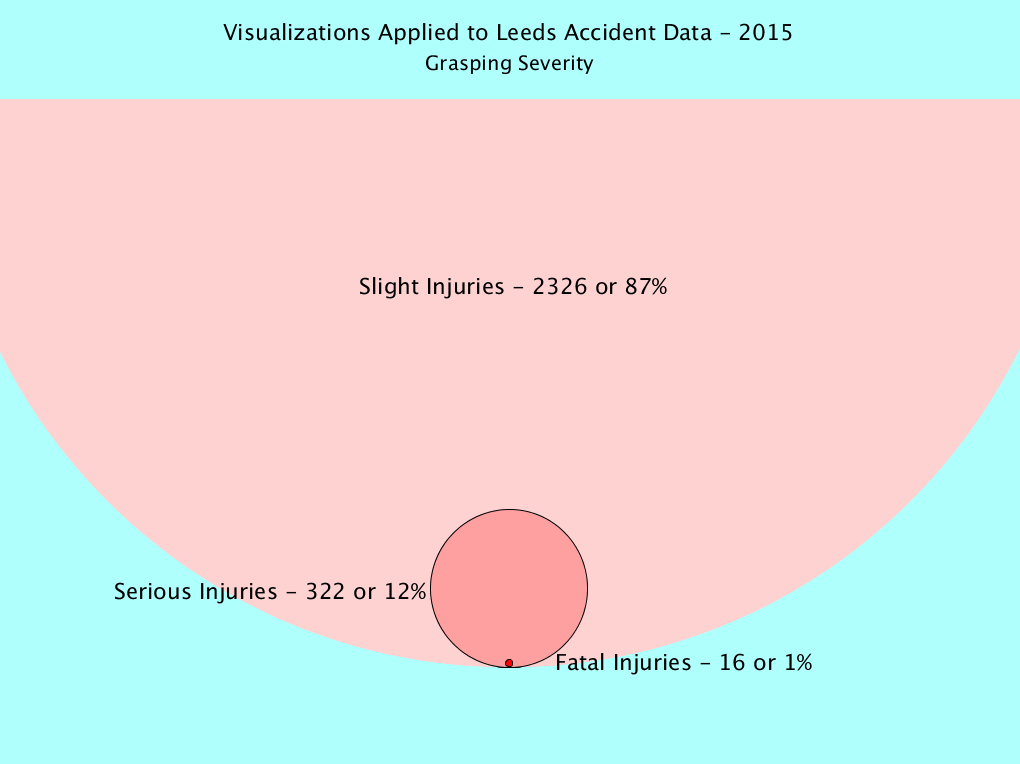
**List of Visualizations**

|  |  |  |
| --- | --- | --- |
| **Visualization** | **Content** | **Type** |
| Grasping Severity | Comparing severity of injuries by type. | Circle |
| Injuries Per Month | Time series view of monthly injury counts. | Line, Time series |
| Road Conditions | View of injuries by road conditions across severity. | Comparative Bar Chart |
| Casualty Location | Locations of injury occurrences. | Map, Scatterplot |
| Sex of Victim | The impact of sex classification on injury severity. | Comparative Bar Chart |
| Color Blindness Checks | A collection of images with filters applied to demonstrate the applicability of visuals for colorblind readers. | Transformed Images |

## Visualization 1 – Grasping Severity

Figure

**Grasping Severity – Understanding the Relative Frequency of Injuries by Severity**



### About this Visualization

Accidents may result in injuries. What does the degree of injury differ?

Fig. 1 demonstrates that eighty-seven percent of the injuries in 2015 in Leeds resulted in injuries categorized as slight. One percent resulted in fatalities. Sixteen percent resulted in serious injuries.

When analyzing the data, the author found that the relative difference in the population’s injury categorization made predicting a fatal injury relating to a set of features impossible. However, some characteristics may define populations that have a relatively reduced risk of participating in a fatal accident.

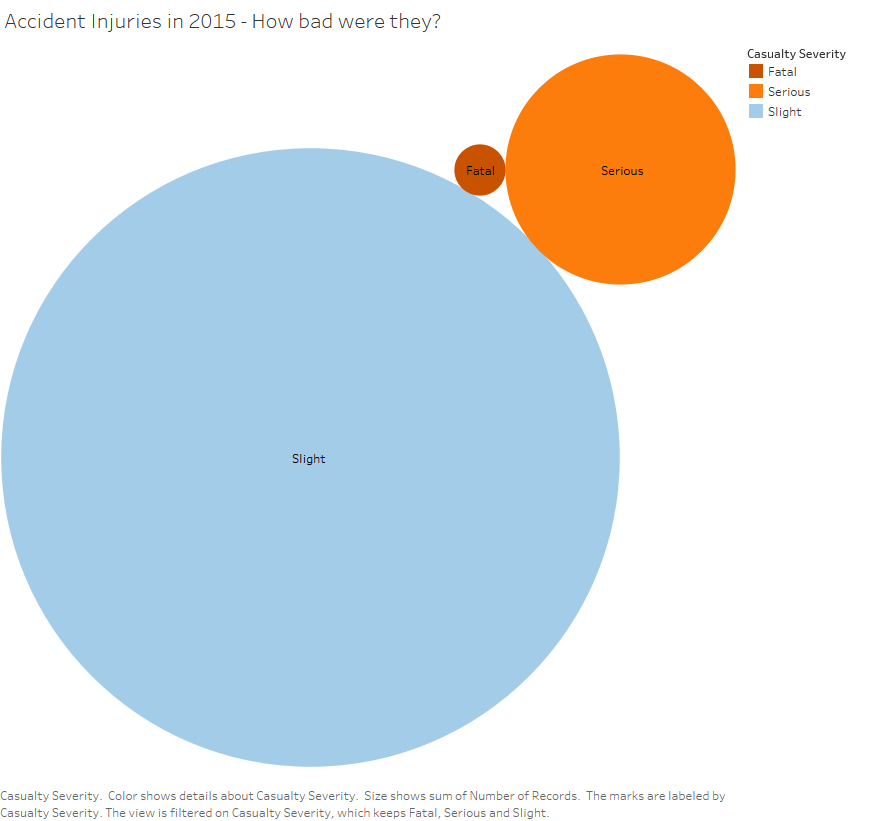
### Design Choices and Encodings

This graphic had the following design goals and encodings:

* Goal: Show the reader the relative frequency of Slight, Serious, and Fatal Injuries and variance between the categories.
* Size demonstrates frequency volume differences between the fatality types. Comparing the size of the Slight Injuries circle to the Serious Injuries circle drives home the frequency differences between the two.
* Variation in shade (brightness) adds order to the severity level.
* Labels provide additional detail and serve to replace a legend explaining colors and values.

Figure

**Original Image – Accident Injuries in 2015**



### Comparing the Original

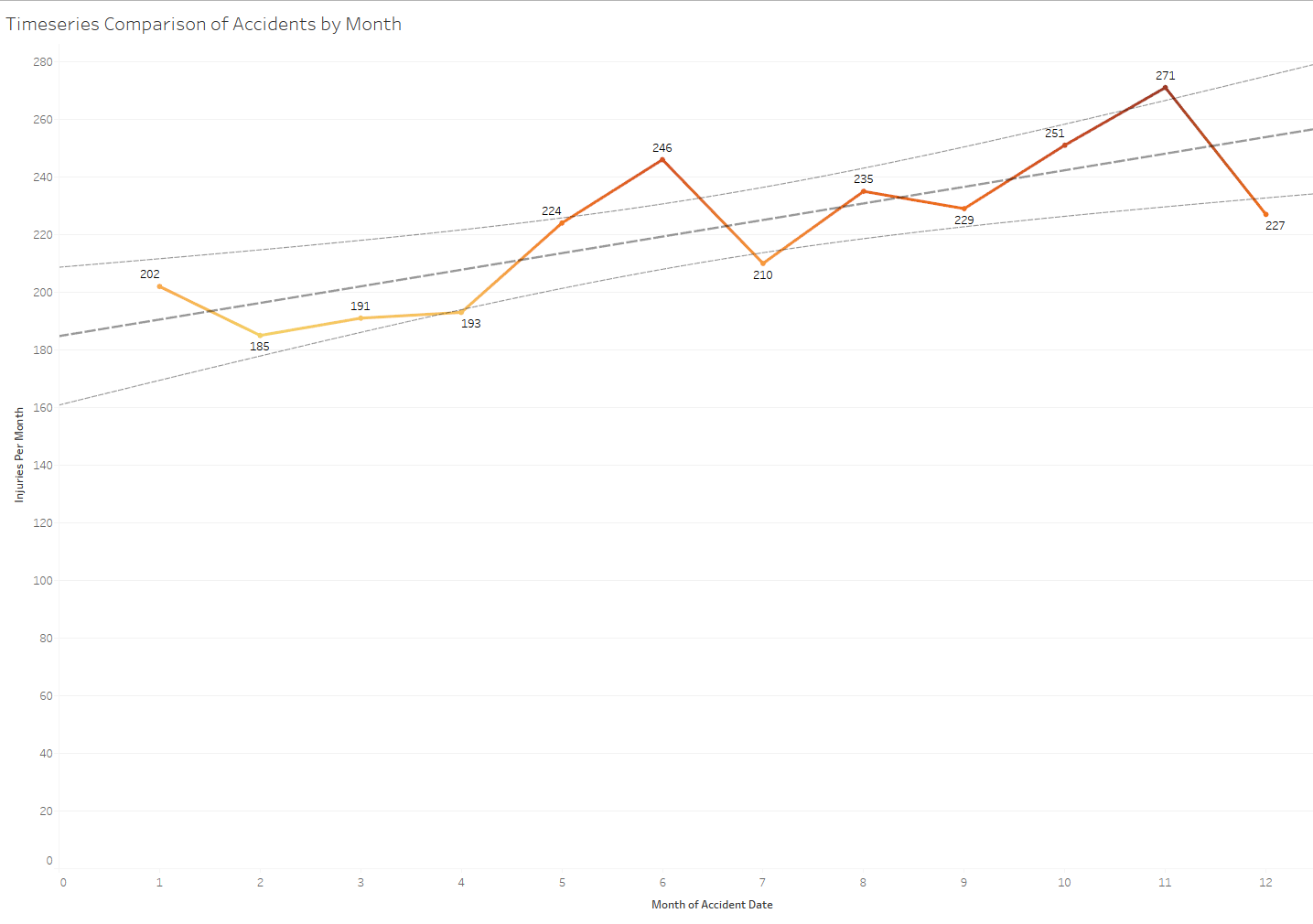
Fig. 2 used a circle metaphor to show the relative size difference. However, the new version addresses issues in the original version.

* Injury proportions did not accurately match the data. Superimposing proportional shapes gives a better representation of the relative frequency of injury by type.
* The new version maintains proportionality between the frequency circles, better conveying the relationship. Increasing the slight injury circle to exceed the screen allows the smaller serious injury and fatality circles to remain large enough for the reader to understand.
* Using saturation as a variable in the circle colors established an ordinality not present in the original drawing. This ordinality appears in later visualizations.
* The original image contained unnereded elements. By incorporating labels in proximity to the referenced elements, readers no longer require a legend. Including injury counts and percentages conveys more information in the visualization.

## Visualization II – Injuries Per Month

Figure

**Injuries per Month**



### About this Visualization

The gross quantity of injuries appears to increase over time. The number of injuries seems to increase in volatility from month to month.

Fig 3. demonstrates that as time continues, injuries increase. Factors related to time (weather, road development, travel volume increases) potentially impact the growth in accidents. What impacts the injury type?

### Design Choices and Encodings

This graphic had the following design goals and encodings:

* Goal: Show the reader the increase in the frequency of all injuries over time.
* Line charts relay time and frequency data well. Using this format allows the reader to grasp changes quickly.
* Variation in shade (brightness) reinforce the highs and lows of injury events by establishing ordinality beyond the frequency grid.
* Lines showing the linear regression of the points, along with variance boundaries strengthen the move and growth of frequency trends.

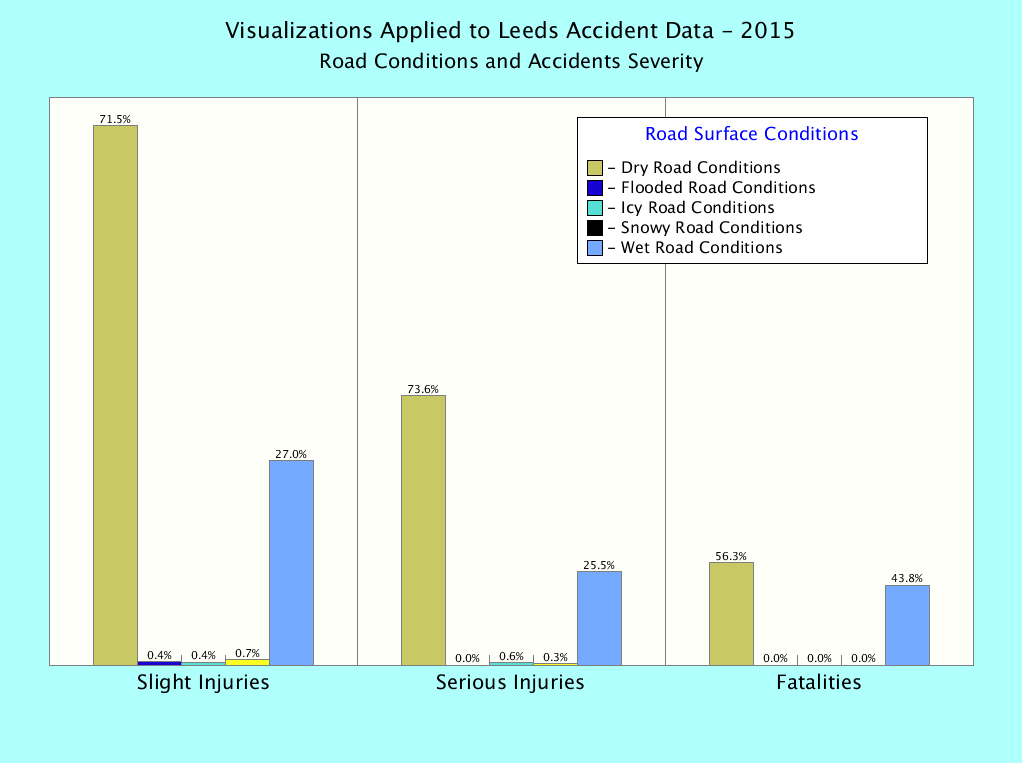
### Comparing the Original

Fig 3. differs from the original only in ordinal placement within the report, and a change in wording on the Y Axis to better describe the frequency data. The author considered the original relevant without major revision.

## Visualization III – Road Conditions

Figure

**Road Conditions**



### About this Visualization

Could the weather be a defining factor in the severity of a casualty? Analysis of the data indicates that weather plrays an increasing role in the degree of injury in an accident. As injuries increase in severity, wet weather coincides with more and more injuries. Interestingly, some extreme types of weather play less and less of a role. Flooded, snowy and icy road conditions rarely coincide with injuries and not at all with fatalities. While dry surface accidents are far more widespread for slight injuries, this proportionality changes substantially as injury severity increases.

### Design Choices and Encodings

This graphic had the following design goals and encodings:

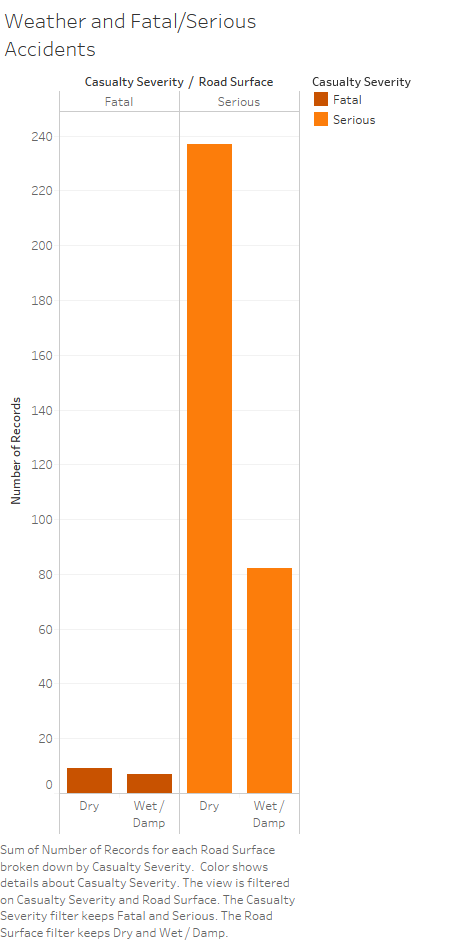
* Goal: Show the reader weather types that coincide with injuries.
* The visual uses the enclosure encoding to group conditions related to a specific injury severity (Slight, Serious, Fatal).
* Distinct colors separate different road conditions but link similar features in different groups.
* Labels provide additional detail andr link them to groups and individual bars. Use of a color-coded legend avoids screen noise given the number of conditions presented. Different visualizations use a different metaphor.

### Comparing the Original

Fig. 4 and 5 used a similar comparative bar metaphor however the new version has substantial changes.

Figure

**Original - Image**



* This report attempts to show differences in weather-related road conditions on accident severity. The shading scheme used in the original focused on injury severity rather road condition. Road condition serves as the primary focus of the visualization within the context of injury severity. The new version reinforces this difference and uses display order and positioning to relate category differences.
* Use of a color-coded legend with proximity placed labels avoids the need for a grid and on-screen text.
* Increasing the number of bars allows for the integration of additional road condition.
* The additional condition categories draw a distinction in the proportional changes in changes Wet and Dry road conditions as injury severity increases.
* The addition of the “Slight” injury category to the “Serious” and “Fatal” adds trending information not available in the original visualization.

## Visualization IV – Casualty Location

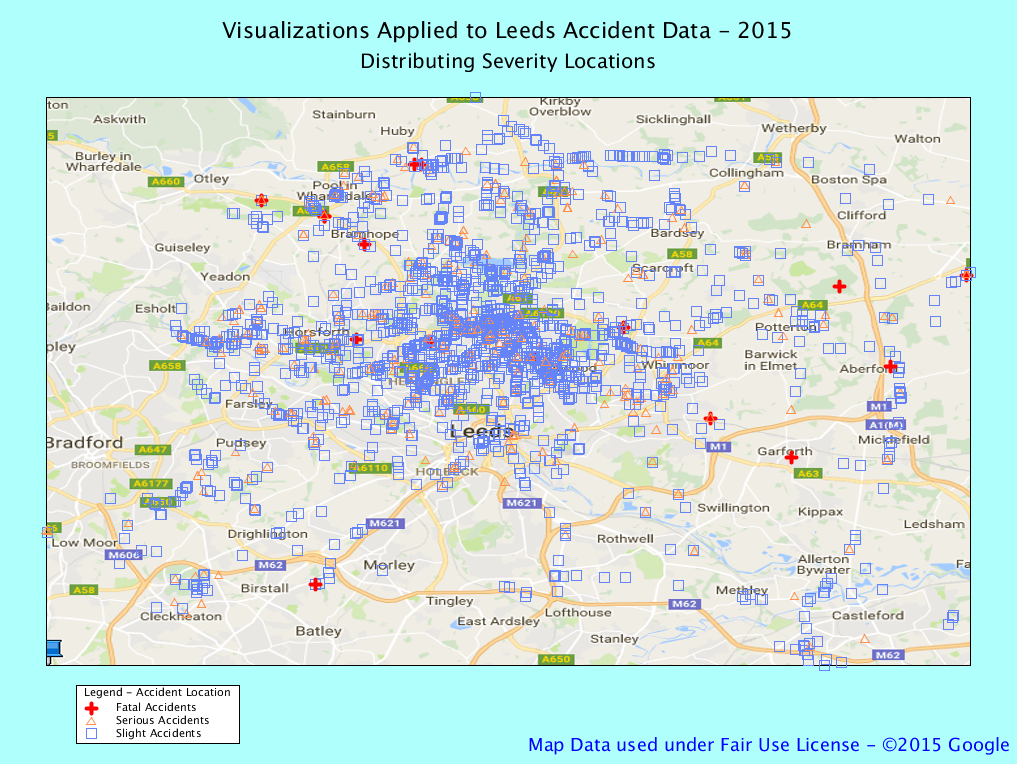
### About this Visualization

Data indicates an increasing role in wet weather conditions and the degree of injury casualty incurred in an accident. Road conditions only seem to play a part in severity.

Fig. 6 shows where accidents occur and interesting facts begin appearing. Numerous accidents with slight injuries concentrate near the city center. Likewise, feeder roads and congested urban areas show a high propensity for numerous, lower-impact accidents. Distribution moves outward for more serious accidents. Fatalities move outside the city distribution. The statistical review indicates the centroids for the city-center and “Slight” injury locations are only about 300 meters apart. The centroid for the Fatal accidents is over 2200 meters from the slight injury accidents.

The graph clearly shows fatal accidents occur primarily outside the urban center close to major thoroughfares.

Figure

**Casualty Location**

### Design Choices and Encodings

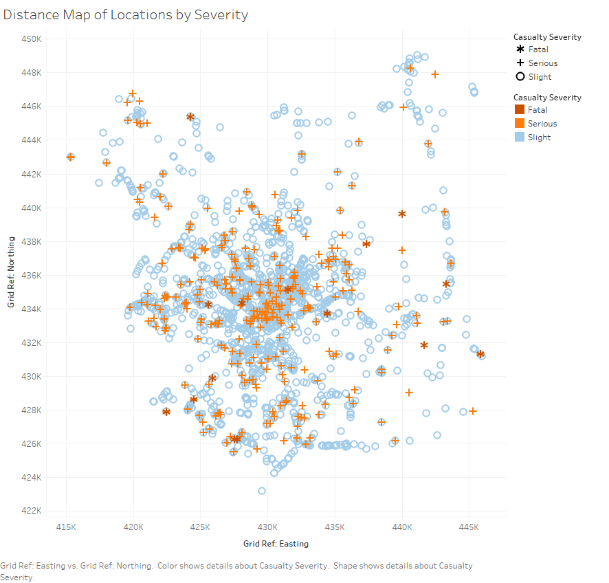
This graphic had the following design goals and encodings:

* Goal: Show the reader where accidents occur relating to Leeds.
* Goal: Show the reader how concentrations differ for Slight, Serious, and Fatal Accidents.
* The visual incorporates a scatterplot plotted a Google Streets Map image [4] using UK National Grid northing and easting coordinates.
* The primary category uses the same color scheme (red) that was used to indicate fatalities in prior visuals.
* “Serious” and “Slight” accidents use low saturation colors and higher transparency to ensure that the reader perceives fatalities as more important.

### Comparing the Original

The original visualization displayed in Fig 7 only displayed distance information and thus only provide abstract insight into the location data available. The changes in Fig. 8 now display a complete picture of accident locations and the relationship to physical terrain characteristics.

* The introduction of a scale map shows proximity to key roads and relative distances between accidents.
* Distinct color and pattern distances allow the reader to gain a picture of accident concentrations by severity.

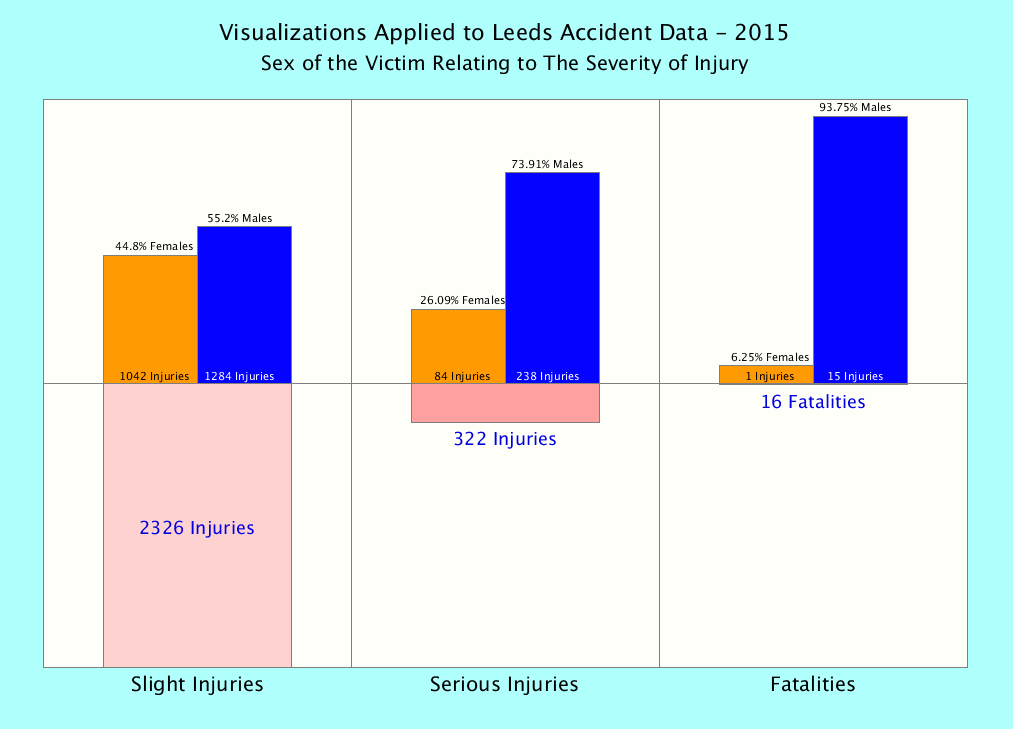


Figure

Distance Map of Locations by Severity

## Visualization V – Victim’s Sex

Figure

**Sex of the Victim Relating to The Severity of the Injury**

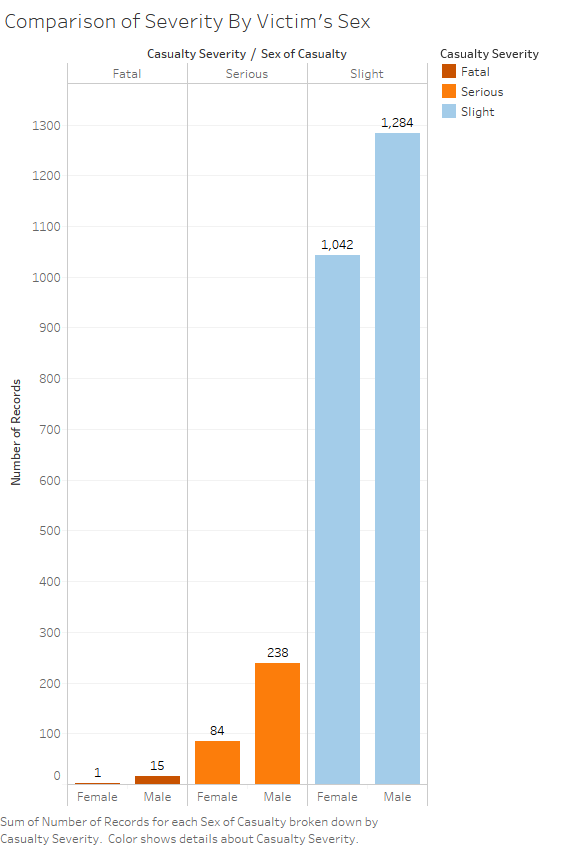
### About this Visualization

This visualization conveys the role of an accident victim’s sex and the increasing impact casualty severity. It also balances the number of injuries against type, reinforcing the relative quantity of impacts. Clearly, accident fatalities in Leeds are almost exclusively male.

As the graph in Fig. 8 demonstrates, when an injury is more critical, the likelihood that the victim is male increases. This distribution shift increases almost linearly. Conversely, females are much less likely to be in a sustain “Serious” or “Fatal” injuries than a male.

### Design Choices and Encodings

This graphic had the following design goals and encodings:

* Goal: Show the reader that men are more likely to be in severe accidents.
* Goal: Show the reader that Leeds accident fatalities are almost exclusively male.
* Goal: Remind the reader of relationships between severity frequency.
* Enclosing data features within boxes defined by injury severity helps group the related data elements for the message delivered. Fig. 4 uses this metaphor as well.
* Data summaries are grouped by proximity to the bars they represent, helping the reader quickly associate the correct quantitative data with the appropriate feature.
* Use of labels and consistent coloring eliminates the need for a formal legend.

Figure

**Comparison of Severity by Victim Sex**

### Comparing the Original

The original uses similar grouping encodings, however, there are several key differences.

* While both versions group by enclosing the casualty severity together, Fig. 8 uses different color selections to differentiate features. Fig. 9 emphasized different severities over sex differences.
* The version in Fig. 8 uses color to first separate male from female, then grouping to indicate a difference in severities.
* The version in Fig. 8 also displays count and percentage information, as well as frequency volume for each severity.
* Distinct labeling provides attribute information by proximity and eliminates the need for a specific legend.

## Comparisons of Colorblind Test

All colorblindness simulations were generated using Vischeck [5].

|  |  |  |
| --- | --- | --- |
| **Deuteranopia** | **Protanopia** | **Original** |
|  | C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Protanope_Viz1.jpg | C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\GraspingSeverity.png |
|  | C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Protanope_Viz2.jpg | C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Injury and Sex.png |
| C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Deuteranope_Viz3.jpg | C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\Protanope_Viz3.jpg | C:\Users\mdsloan2000\AppData\Local\Microsoft\Windows\INetCacheContent.Word\CasualtyLocations.png |

# cConclusion

Accidents tend to result in injuries that are “Slight.” However as injury severity increases, certain characteristics more clearly define the victim. Based on research conducted for this document, fatal accidents tend to occur in either Wet or Dry road conditions but do not seem to increase in Flooded, Icy, or Snowy conditions. The most severe injuries, “Fatalities,” tend to occur outside of the city center. Additionally, as injury severity increases, the sex of the victim tends to be male.

These findings lead to new questions that require additional data. Why are males prone to fatal accidents? Do vehicle speeds contribute injury severity? Can additional analysis predict accidents that result in fatalities? This analysis allows us to characterize the victim, but not predict conditions that uniquely lead to death.

##### References

1. Michael Sloan, [Unpublished Paper] Key Factors of Fatal Injuries in Leeds, UK, <https://www.dropbox.com/s/vux06qif0c5np6b/2015LeedsTraffic.pdf?dl=0>, October 22, 2016
2. Leeds City Council, “Road traffic accidents.” Internet: https://data.gov.uk/dataset/road-traffic-accidents, May 19, 2016.
3. The National Archives, “Open Government License for the Public Center.” Internet: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/, October 21, 2016.
4. Google Maps, “Map of Leeds” Internet: https://www.google.com/maps/place/Leeds,+UK/@53.8059209,- 1.6758131, Retrieved December 17, 2016.
5. Vischeck, “Vischeck.” Internet: <http://www.vischeck.com/vischeck/>, December 17, 2016.