Data Mining, Fall 2015

#### Analytical Project Assignment

Team D

Understanding patterns of road safety

## Associated data sets

DfTRoadSafety\_Accidents\_2012.csv

Road-Accident-Safety-Data-Guide.xls

# Data description

The associated data is a subset of the public collection of data of the circumstances of personal injury road accidents in Great Britain in 2012. The statistics relate only to personal injury accidents on public roads that are reported to the police, and subsequently recorded, using the STATS19 accident reporting form. Information on damage-only accidents, with no human casualties or accidents on private roads or car parks is not included in this data. Please note that some of the integer values are codes rather than actual values, please refer to the supplied code book (Road-Accident-Safety-Data-Guide.xls) to understand the actual meanings. This file may contain references to additional datasets that can be ignored.

Very few, if any, fatal accidents do not become known to the police although it is known that a considerable proportion of non-fatal injury accidents are not reported to the police. Figures for deaths refer to persons killed immediately or who died within 30 days of the accident.  This is the usual international definition, adopted by the Vienna Convention in 1968.

As well as giving details of date, time and location, the accident file gives a summary of all reported vehicles and pedestrians involved in road accidents and the total number of casualties, by severity.

### Key tasks

1. Identify a group of features that you think will be most useful. You may use a combination of intuition and variable selection techniques for this task. If you wish, you may derive new variables which you think could be useful in your analysis and discard variables which you think may not be. Search for interesting patterns in data by summarizing and visually inspecting it.
2. Based on the available data, what are the factors that best discriminate between different severities of accidents? Can the knowledge of these factors be helpful in the practice of reducing frequency of higher-severity accidents?
3. Can the given data be used to infer areas for improvement within different police jurisdictions? If so, provide the reasoning for your analysis with examples.