

MATH550/SCC461: Statistics in Practice

Lab 2

Assessment

Tom Palmer

- **Deadline** for coursework submission is **9am Monday 16th October 2017**.
- Include your 8 digit library card number in the author field when creating your knitted report.

Assessment

There are four parts to the assessment. Before you start, read the content below. Then complete the tasks.

Leeds Accident data

The internet site data.gov.uk/dataset is host to a wide range of officially collected statistics databases. Amongst these databases is road traffic accidents such as `accidents2014.csv` which contains road accident data for Leeds in 2014. The data set can be downloaded from the 'Data sets' folder on the Moodle page. This contains 16 variables on road accidents in Leeds and there is an accompanying `.csv` file `accidentguidance.csv` with details concerning the variables. Furthermore the variables in the data set are listed in Table 1.

Variable(s)	Description
Reference.Number	Reference number of accident.
Grid.Ref..Easting	East(-west) grid reference for accident.
Grid.Ref..Northing	North(-south) grid reference for accident.
Number.of.Vehicles	Number of vehicles involved in the accident.
Number.of.Casualties	Number of casualties as a result of the accident.
Accident.Date	Date of accident.
Time..24hr.	Time of accident on the 24 hour clock.
X1st.Road.Class	Class of road (details given in accidentguidance.csv).
Road.Surface	Road surface conditions (details given in accidentguidance.csv).
Lighting.Conditions	Lighting conditions (details given in accidentguidance.csv).
Weather.Conditions	Weather conditions (details given in accidentguidance.csv).
Casualty.Class	Class of casualty (details given in accidentguidance.csv).
Casualty.Severity	Severity of casualty (details given in accidentguidance.csv).
Sex.of.Casualty	Sex of casualty (1 - Male; 2 - Female).
Age.of.Casualty	Age of casualty (in years).
Type.of.Vehicle	Type of vehicle (details given in accidentguidance.csv).

Table 1: List of variable description in the Leeds road accident data set.

Tasks

These tasks require you to make use of the functions provided by the `dplyr`, and `ggplot2` packages.

1. Read the data into R, check the names of the variables match those in the table, and find out the number of rows.

Print the dimensions of this data frame.

[2 mark]

2. Use `select()` to modify the data frame, dropping the following variables; `Accident.Date`, `Time..24hr.`, `Road.Surface`, `Lighting.Conditions` and `Weather.Conditions` and keeping the others.

Use `filter()` to modify the data frame, so that the accidents we are studying involve: a private car (vehicle type 9) and are not on a motorway.

Print the dimensions of your newly modified data frame.

[2 marks]

3. The centre of Leeds has Easting-Northing coordinates, 429967, 434260. Add a variable to accident data frame using `mutate()` to give the distance of the accident from the centre of Leeds in metres.

Hint: Note that each change in an Easting/Northing coordinate corresponds to one metre.

Using the modified data from Q2, and `arrange()`, reorder the accidents in ascending order from the centre of Leeds and print out the final entries of the data frame.

[4 marks]

4. Using `ggplot()`, create a histogram of age of casualties. Modify the binwidth into groups of 10 years. Set the axis labels to be Casualty age and No. of casualties.

[2 marks]