

SCC.461: Coursework 1

Chris Jewell

2024

Before you start

- Read the instructions carefully. Failure to follow them will result in mark deduction.
- All coursework are to be submitted on the module Moodle page. Do not email your coursework to the lecturer or the teaching office.
- Each weekly coursework involves you submitting an HTML file (a webpage that shows both your code and your output) generated from an Rmarkdown (.Rmd) file. You **must** submit the HTML version. **Pdf files or Word files are not accepted.**
- You also need to submit the raw Rmarkdown file, of which the file extension **needs to be changed from .Rmd to .txt upon submission**, because Moodle only allows file types accepted by turnitin.
- The coursework is marked out of 10.
- Include your library card number at the top of your document as your **author** :.
- Deadline for coursework submission is **1700 Friday 18th October 2024**.

Coursework

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

Leeds Accident data

The internet site <https://data.gov.uk/search> 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 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.

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

Variable	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 <code>accidentguidance.csv</code>)
Road.Surface	Road surface conditions (details given in <code>accidentguidance.csv</code>)
Lighting.Conditions	Lighting conditions (details given in <code>accidentguidance.csv</code>)
Weather.Conditions	Weather conditions (details given in <code>accidentguidance.csv</code>)
Casualty.Class	Class of casualty (details given in <code>accidentguidance.csv</code>)
Casualty.Severity	Severity of casualty (details given in <code>accidentguidance.csv</code>)
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 <code>accidentguidance.csv</code>)

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 print the dimensions of the data frame. [2 marks]
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 category 9) and are not on a motorway (class of road category 1).

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 the accident data frame using `mutate()` to give the distance of the accident from the centre of Leeds in metres.

Hint: Easting-Northing coordinates are a rectangular geospatial coordinate system in units of metres. This enables you to use Pythagoras' Theorem to compute distances.

Using the modified data from question 2 and `arrange()` reorder the accidents in ascending order from the centre of Leeds and print out the bottom few rows from this data frame.

[4 marks]

4. Continuing with the data modified in questions 2 and 3, 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]