

# Coursework 1

36071280

15/10/2021

## Leeds Accident data

Q1. Read the data into R, check the names of the variables match those in the table, and print the dimensions of the data frame.

```
library(dplyr)

## 
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
## 
##     filter, lag
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
library(ggplot2)

#Reading data fromCSV file
accidentsData = read.csv("accidents2014.csv")

#data.frame(accidentsData)

#Check column names -no need to print
#colnames(accidentsData)

#Get dimensions of the dataframe
dim(accidentsData)
```

## [1] 2533 16

Q2. 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.

```
# Do it using column names and not strings
accidentsData_select = select(accidentsData, -c("Accident.Date", "Road.Surface",
                                              "Lighting.Conditions",
                                              "Weather.Conditions",
                                              "Time..24hr."))
#dim(accidentsData_select)
```

Q. 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.

```
accidentsData_filtered = filter(accidentsData_select,
                                accidentsData_select$Type.of.Vehicle == 9,
                                accidentsData_select$X1st.Road.Class != 1)

dim(accidentsData_filtered)

## [1] 1515   11
```

Q3. 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.

```
#Defining leeds centre coordinates
x1 = 429967
y1 = 434260

#Defining function to calculate the distance from city centre
calc_distance <- function (x1, x2, y1, y2) {
  sqrt((x2-x1)^2 + (y2-y1)^2)
}

#Getting accidents data with distance column
accidentsData_withDistance = mutate(accidentsData_filtered,
                                      Distance.From.Centre = calc_distance
                                      (x1, accidentsData_filtered$Grid.Ref..Easting,
                                       y1, accidentsData_filtered$Grid.Ref..Northing))
```

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.

```
accidentsData_orderedbyDistance = arrange(accidentsData_withDistance,
                                            accidentsData_withDistance$Distance.From.Centre)
tail(accidentsData_orderedbyDistance)

##      Reference.Number Grid.Ref..Easting Grid.Ref..Northing Number.of.Vehicles
## 1510          1BU1133        440547        448561                  3
## 1511          1BU1133        440547        448561                  3
## 1512          1BU1133        440547        448561                  3
## 1513          17V0436        439873        449526                  1
## 1514          13L0235        440411        449270                  1
## 1515          1AH0546        441101        449222                  2
##      Number.of.Casualties X1st.Road.Class Casualty.Class Casualty.Severity
## 1510                 3             4             1             3
## 1511                 3             4             1             3
## 1512                 3             4             2             3
## 1513                 1             6             3             3
## 1514                 1             4             3             3
## 1515                 1             2             1             3
##      Sex.of.Casualty Age.of.Casualty Type.of.Vehicle Distance.From.Centre
## 1510                 1              91               9       17789.18
## 1511                 1              65               9       17789.18
## 1512                 2              63               9       17789.18
## 1513                 1              42               9       18198.34
## 1514                 1              14               9       18285.98
## 1515                 1              56               9       18650.13
```

Q4. 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”.

```
# Creating histogram
ggplot(accidentsData_orderedbyDistance, aes(Age.of.Casualty)) +
  geom_histogram(binwidth = 10, color="yellow",
                 position = "identity", fill="red") +
  labs(title="Histogram for Age", x="Casualty age", y="No. of Casualties")
```

