# R Coursework 1

Name: Congye Wang Student ID: 35427962

Oct 7th, 2020

```
# Import Packages
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------ tidyverse 1.3.0 --
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## -- Conflicts --------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

## Question 1

```
# Load Data
df <- read.csv("./accidents2014.csv", header = TRUE)
t1 <- data.frame(

  Variable = c("Reference.Number", "Grid.Ref..Easting", "Grid.Ref..Northing",
"Number.of.Vehicles","Number.of.Casualties", "Accident.Date", "Time..24hr.",
"X1st.Road.Class", "Road.Surface", "Lighting.Conditions", "Weather.Conditions",
"Casualty.Class", "Casualty.Severity", "Sex.of.Casualty", "Age.of.Casualty",
"Type.of.Vehicle"),

  Description = c("Reference number of accident", "East(-west) grid reference for accident",
"North(-south) grid reference for accident", "Number of vehicles involved in the accident",
"Number of casualties as a result of the accident", "Date of accident",
"Time of accident on the 24 hour clock",
"Class of road (details given in accidentguidance.csv)",
"Road surface conditions (details given in accidentguidance.csv)",
"Lighting conditions (details given in accidentguidance.csv)",
"Weather conditions (details given in accidentguidance.csv)",
"Class of casualty (details given in accidentguidance.csv)",
"Severity of casualty (details given in accidentguidance.csv)",
"Sex of casualty (1 - Male; 2 - Female)", "Age of casualty (in years)",
"Type of vehicle (details given in accidentguidance.csv)")
)

# Check Headers
df_headers_names <- sort(names(df))
```

```r
std_headers_names <- sort(t1$Variable)

for (i in 1:length(df_headers_names)) {

  if (df_headers_names[i] != std_headers_names[i]) {

    print(paste("False", "No.", i, "\n"))

  }

  else {

    next

  }

}

# Print Dimensions
nrow(df)
```

```
## [1] 2533
```

```r
ncol(df)
```

```
## [1] 16
```

```r
dim(df)
```

```
## [1] 2533   16
```

As a result, the number of the data frame's columns is 16, and that of rows is 2533.

## Question 2

```r
# Screen Data
vars <- c("Accident.Date", "Time..24hr.", "Road.Surface", "Lighting.Conditions",
"Weather.Conditions")
df_2_1 <- select(df, -one_of(vars))
df_2_2 <- filter(df_2_1, (Type.of.Vehicle == 9) & (X1st.Road.Class != 1))

# Print Dimensions
nrow(df_2_2)
```

```
## [1] 1515
```

```r
ncol(df_2_2)
```

```
## [1] 11
```

```r
dim(df_2_2)
```

```
## [1] 1515   11
```

As a result, the number of the data frame's columns is 11, and that of rows is 1515.

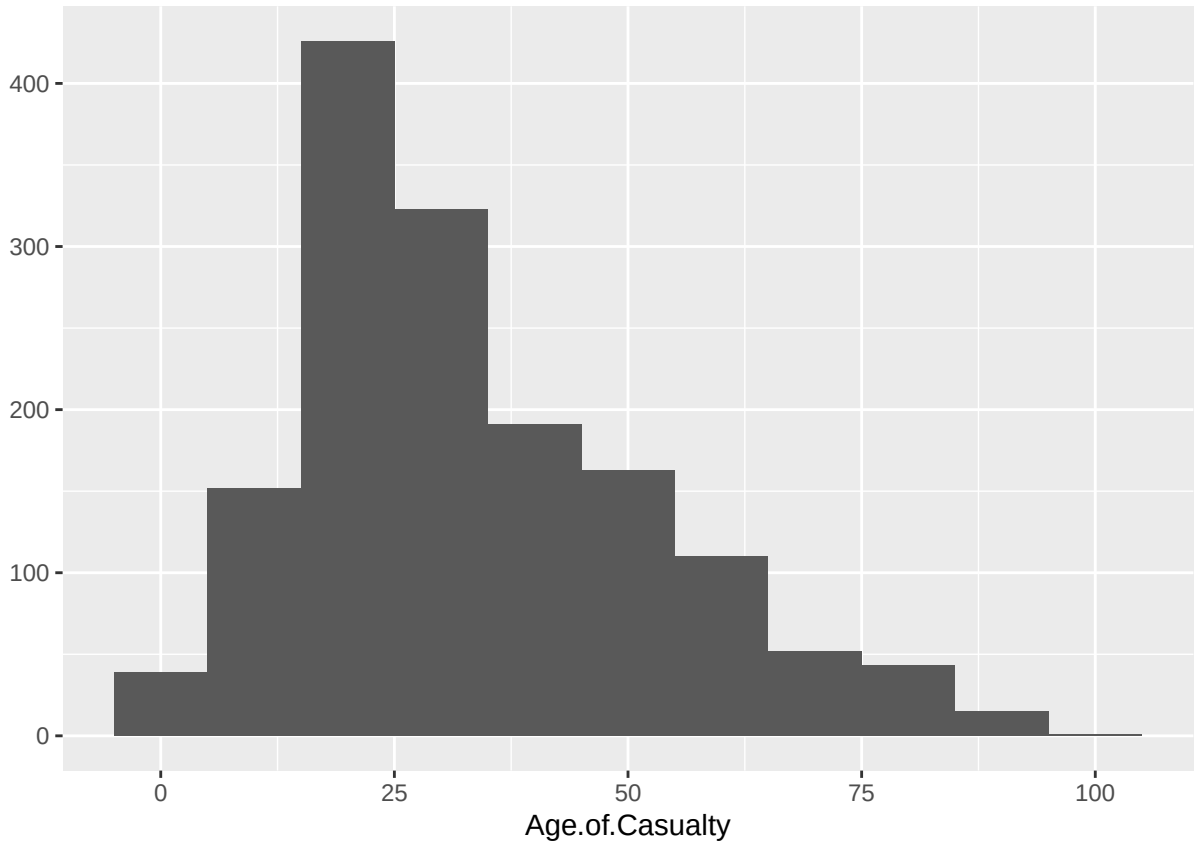
## Question 3

```
# Distance Calculation Function
distance_calculate <- function(a, b) {

  return(sqrt((a - 429967)^2 + (b - 434260)^2))

}

# Add the Distance of the Accident from the Centre of Leeds in Metres to the Accident Data Frame
df.distance.of.accident <- distance_calculate(df$Grid.Ref..Easting,
df$Grid.Ref..Northing)
df <- mutate(df, Distance.of.Accident = df.distance.of.accident)

# Using the Modified Data from Question 2
df_3 <- mutate(df_2_2, Distance.of.Accident = distance_calculate(
df_2_2$Grid.Ref..Easting,
df_2_2$Grid.Ref..Northing))
df_3 <- arrange(df_3, df_3$Distance.of.Accident)
tail(df_3)
```

```
##      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.of.Accident
## 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
```

## Question 4

```
# Plot Histogram
qplot(x=Age.of.Casualty, data=df_3, geom="histogram", binwidth = 10)
```


400
300
200
100
0
0
25
50
75
100
Age.of.Casualty