Lab 2 coursework - Leeds accident data

Tom Palmer

8th October 2017

```
#
# 1.
#
# Read in data.
#

setwd("H:\\all\\teaching\\Math550 Stats in Practice\\R\\MATH550\\R_course\\Lab2\\Coursework")

#install.packages("dplyr")
#install.packages("ggplot")
# or
#install.packages("tidyverse") # tidyverse: ggplot2, dplyr, tidyr, readr, purrr, tibble
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)

accidents <- read.csv("accidents2014.csv")
print(dim(accidents))
```

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

```
#
# 2.
#

# Select the required varibles
acc1 <- select(accidents, Reference.Number,
               Grid.Ref..Easting, Grid.Ref..Northing,
               Number.of.Vehicles, Number.of.Casualties,
               X1st.Road.Class, Casualty.Class,
               Casualty.Severity, Sex.of.Casualty,
               Age.of.Casualty, Type.of.Vehicle)

# Keep cars and non-motorway accidents.
acc2 <- filter(acc1, Type.of.Vehicle==9, X1st.Road.Class!=1)
print(dim(acc2))
```

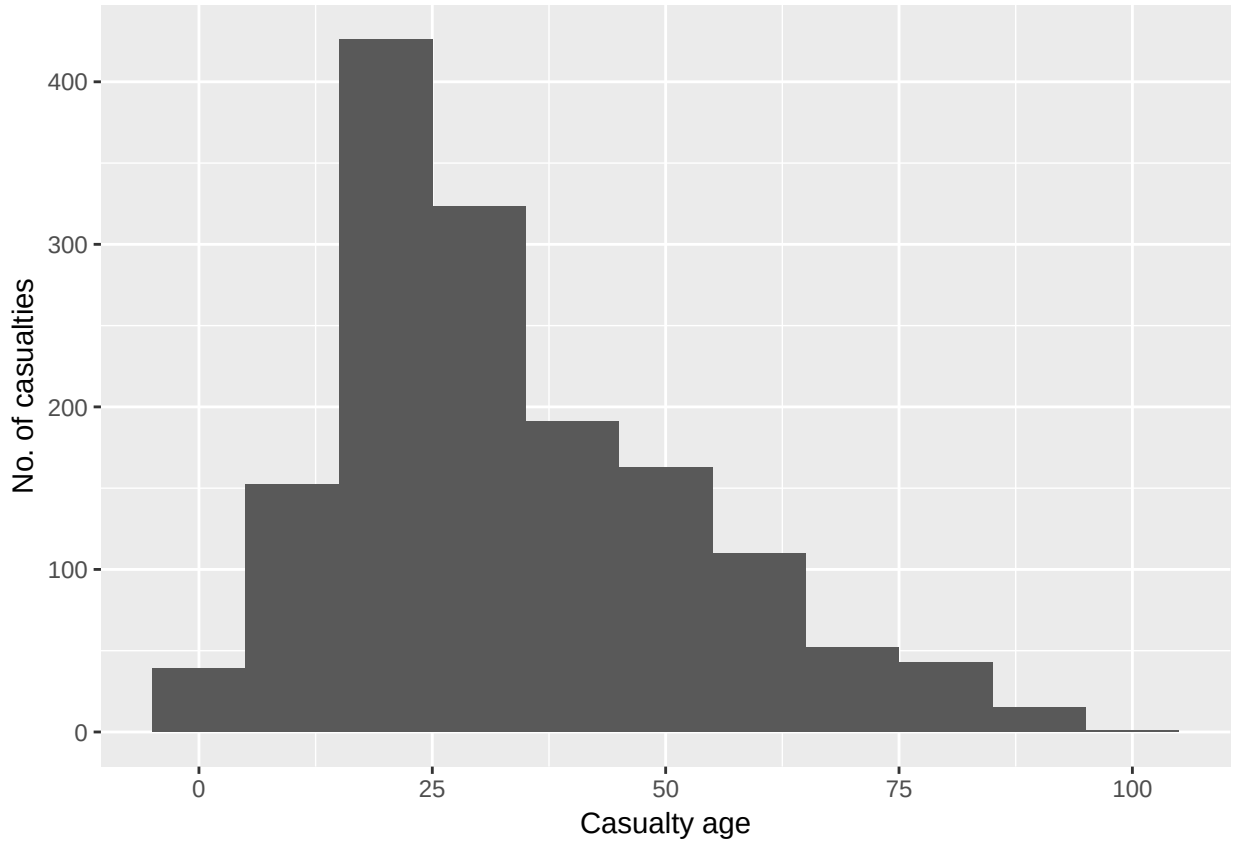
```
## [1] 1515   11
```

```
#
# 3.
#

#Leeds_Easting <- 429967
#Leeds_Northing <- 434260

#
# Function for computing the distance from the centre of Leeds.
#
Leeds_dist <- function(Easting, Northing){
  Leeds_Easting <- 429967
  Leeds_Northing <- 434260
  distance <- sqrt((Easting - Leeds_Easting)^2 +
                   (Northing - Leeds_Northing)^2)
  return(distance)
}

#
# Compute the distance from the centre of Leeds.
#
acc3 <- mutate(acc2, distance=Leeds_dist(Grid.Ref..Easting,
                                         Grid.Ref..Northing))

#
# Rearrange the accidents by distance.
#
acc4 <- arrange(acc3, distance)

# output
print(tail(acc4))
```

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

```
## 1510                9 17789.18
## 1511                9 17789.18
## 1512                9 17789.18
## 1513                9 18198.34
## 1514                9 18285.98
## 1515                9 18650.13
```

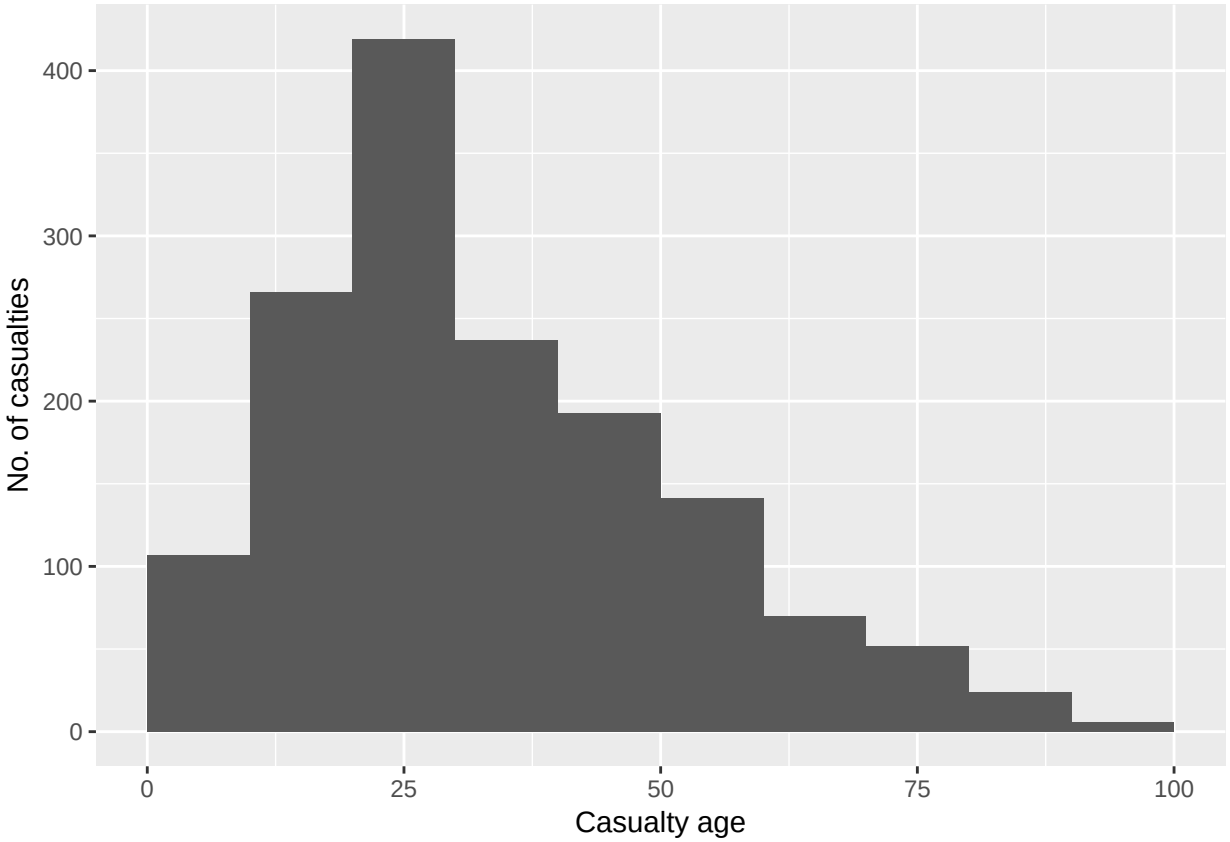
```
#
# 4.
#

# Producing the plot
Age <- ggplot(acc4) +
     geom_histogram(aes(x=Age.of.Casualty), binwidth=10) +
     labs(x="Casualty age", y="No. of casualties")
Age
```




```
# Saving it as a file.
ggsave(Age, file="Age.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
# Making sure the bins do not go below zero

Age2 <- ggplot(acc4) +
  geom_histogram(aes(x=Age.of.Casualty), breaks=seq(0, 100, 10)) +
  labs(x="Casualty age", y="No. of casualties")
Age2
```


400
300
200
100
0
No. of casualties
0
25
50
75
100
Casualty age