Analysis of the cars data set in R

by Osita Onyejekwe

Part 1

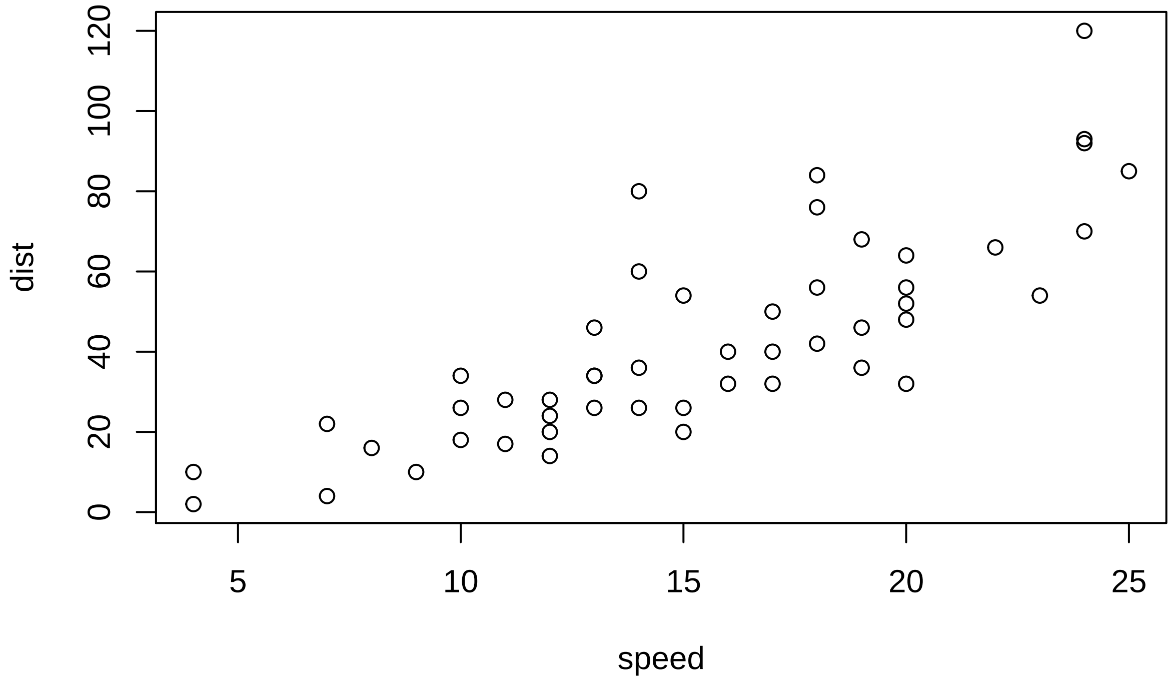
```
data(cars)
str(cars)
```

```
## 'data.frame':    50 obs. of  2 variables:
##  $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
##  $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

```
summary(cars)
```

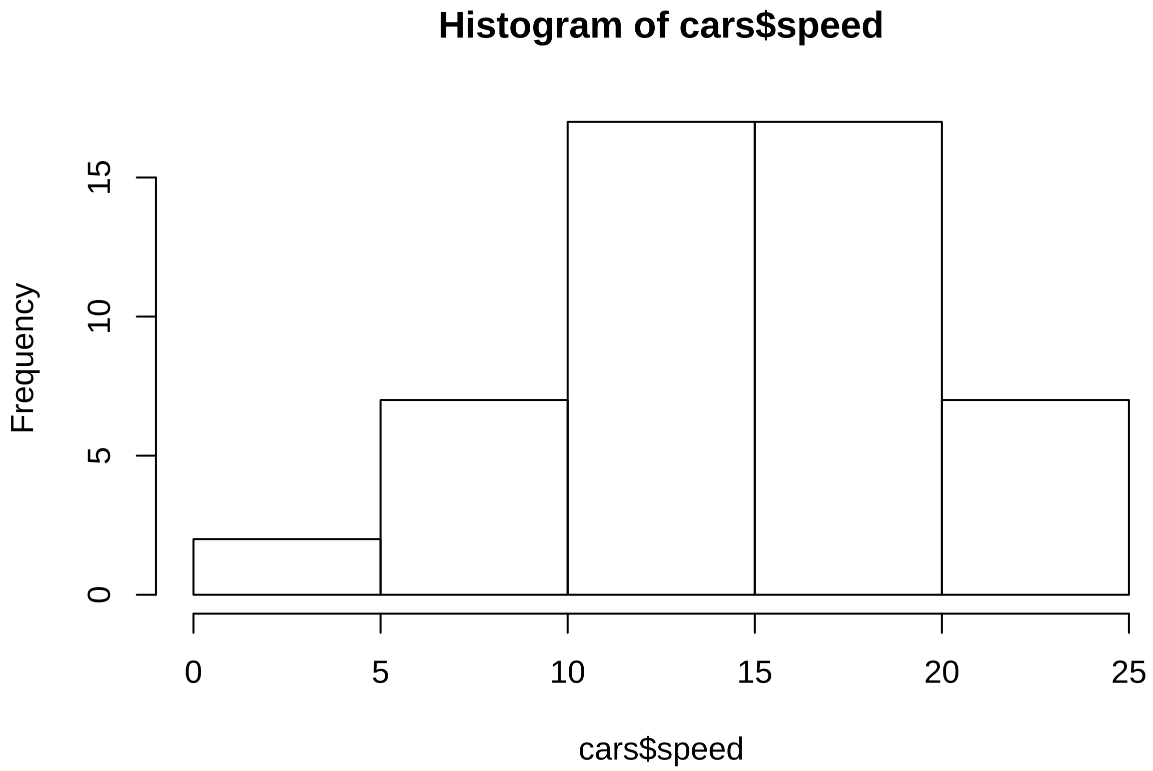
```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

```
plot(cars)
```

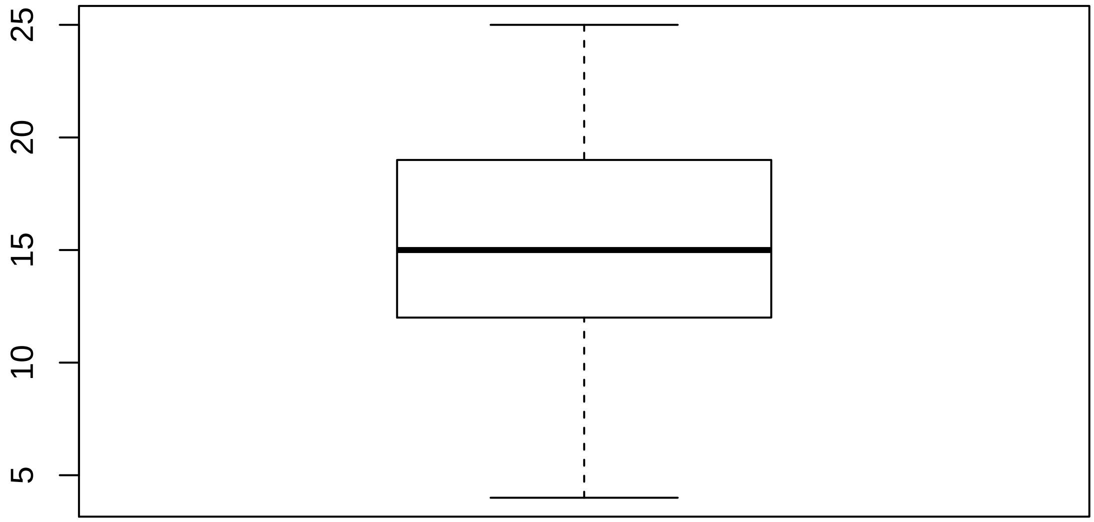


Part 2

```
hist(cars$speed)
```



```
boxplot(cars$speed)
```



The mean speed of the car is 15.4

```
hello i can't wait for the patriots to win the next superbowl
```

Part 3: Data Import

```r
survey<- read.csv('/Users/osita/OneDrive/Desktop/STAT 2600 SPRING 2021/Coding/survey_data2020.csv')
class(survey)
```

```
## [1] "data.frame"
```

```r
head(survey, 3)
```

```
##   Program                PriorExp      Rexperience OperatingSystem TVhours
## 1     PPM         Some experience       Never used         Windows    10.5
## 2   Other    Extensive experience Basic competence        Mac OS X     3.0
## 3    MISM Never programmed before Basic competence         Windows     0.0
##           Editor
## 1          Other
## 2 Microsoft Word
## 3 Microsoft Word
```

```r
survey<- read.csv('/Users/osita/OneDrive/Desktop/STAT 2600 SPRING 2021/Coding/survey_data2020.csv')
class(survey)
```

```
## [1] "data.frame"
```

```r
head(survey,3)
```

```
##   Program                PriorExp      Rexperience OperatingSystem TVhours
## 1     PPM         Some experience       Never used         Windows    10.5
## 2   Other    Extensive experience Basic competence        Mac OS X     3.0
## 3    MISM Never programmed before Basic competence         Windows     0.0
##           Editor
## 1          Other
## 2 Microsoft Word
## 3 Microsoft Word
```

Lecture 2 Day 2

STAT 3400

Part 1: Simple Summary

Use the **str()** function to get a simple summary of your data frame object

```
str(survey)
```

```
## 'data.frame':    57 obs. of  6 variables:
##  $ Program        : Factor w/ 3 levels "MISM","Other",..: 3 2 1 3 3 3 3 3 3 2 ...
##  $ PriorExp       : Factor w/ 3 levels "Extensive experience",..: 3 1 2 2 2 3 2 3 3 3 ...
##  $ Rexperience    : Factor w/ 4 levels "Basic competence",..: 4 1 1 4 4 1 4 3 1 1 ...
##  $ OperatingSystem: Factor w/ 3 levels "Linux/Unix","Mac OS X",..: 3 2 3 3 3 2 2 2 3 3 ...
##  $ TVhours        : num  10.5 3 0 10 4 0 2 20 4 0 ...
##  $ Editor         : Factor w/ 5 levels "Excel","LaTeX",..: 4 3 3 1 3 3 3 4 3 3 ...
```

Factor refers to categorical data wherease TVhours is a numerica variable

```
summary(survey)
```

```
##   Program                      PriorExp                Rexperience
##  MISM : 9   Extensive experience   : 8   Basic competence    :24
##  Other:10   Never programmed before: 8   Experienced         : 6
##  PPM  :38   Some experience        :41   Installed on machine: 7
##                                          Never used          :20
##
##
##    OperatingSystem    TVhours                 Editor
##  Linux/Unix: 2     Min.   : 0.000   Excel           : 1
##  Mac OS X  :19     1st Qu.: 3.000   LaTeX           : 5
##  Windows   :36     Median : 5.000   Microsoft Word:40
##                    Mean   : 6.763   Other           : 8
##                    3rd Qu.:10.000   R Markdown      : 3
##                    Max.   :21.000
```

Data Frame Basics

Lists, and data frames (and their “tidy” variants)—> next week but for now some basics Goal here is to observe what an R object is made up off, using **attributes()

```
attributes(survey)
```

```
## $names
## [1] "Program"         "PriorExp"        "Rexperience"     "OperatingSystem"
## [5] "TVhours"         "Editor"
##
## $class
## [1] "data.frame"
##
```

```
## $row.names
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
## [26] 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
## [51] 51 52 53 54 55 56 57
```

An R **data frame** is a list whose columns can refer to by name or index When you see $ symbol it tells you that it's a list of some kind

**Data Frame Dimensions

We use **nrow()** and **ncol** to determine the number of survey responses and the number of survey questions

```r
nrow(survey) # number of rows (responses)
```

```
## [1] 57
```

```r
ncol(survey) # number of columns (questions)
```

```
## [1] 6
```

57

We collected data on 6 survey questions from 57 respondents. Respondents represnted 3 CU programs. 38 of the respondents were from PPM

We collected data on 6 survey questions from 57 repondents. Respondents represented 3 CU programs. 38 of the respondents were from PPM

Mondays — Indexing of data frames