Project - Cardio Good Fitness

Chai Kok Soon

2/16/2020

Contents

[1. Project Objective 2](#_Toc33186519)

[2. Assumptions 2](#_Toc33186520)

[3 Exploratory Data Analysis – Step by step approach 3](#_Toc33186521)

[3.1 Environment Set up and Data Import 3](#_Toc33186522)

[3.1.1 Install necessary Packages and Invoke Libraries 3](#_Toc33186523)

[3.1.2 Set up working Directory 3](#_Toc33186524)

[3.1.3 Import and Read the Dataset 3](#_Toc33186525)

[3.2 Variable Identification 4](#_Toc33186526)

[3.2.1 Variable Identification – Inferences 5](#_Toc33186527)

[3.3 Univariate Analysis 6](#_Toc33186528)

[3.4 Bi-Variate Analysis 9](#_Toc33186529)

[3.5 Missing Value Identification 20](#_Toc33186530)

[3.6 Outlier Identification 20](#_Toc33186531)

[4 Conclusion 20](#_Toc33186532)

[5 Appendix A – Source Code 20](#_Toc33186533)

# 1. Project Objective

The objective of the report is to explore the cardio data set (“CardioGoodFitness”) in R and generate insights about the data set. This exploration report will consist of the following:  
- Importing the dataset in R  
- Understanding the structure of dataset  
- Graphical exploration  
- Descriptive statistics  
- Insights from the dataset

# 2. Assumptions

The analysis is done based on the data stored in the CardioGoodFitness.csv file. We assume that the data provided in the file is accurate without major errors.

# 3 Exploratory Data Analysis – Step by step approach

A typical data exploration activity consists of the following steps:  
     1.Environment Set up and Data Import  
     2.Variable Identification  
     3.Univariate Analysis  
     4.Bi-Variate Analysis  
     5.Missing Value Treatment (Not in scope for our project)  
     6.Outlier Treatment (Not in scope for our project)  
     7.Variable Transformation / Feature Creation  
     8.Feature Exploration  
We shall follow these steps in exploring the provided dataset.

Although Steps 5 and 6 are not in scope for this project, a brief about these steps (and other steps as well) is given, as these are important steps for Data Exploration journey.

## 3.1 Environment Set up and Data Import

### 3.1.1 Install necessary Packages and Invoke Libraries

##################################################  
# 1. Data Preparation  
##################################################  
  
########################  
## 1.1 Import Libraries  
library(readr) ## To import .csv  
library(ggplot2)  
library(esquisse)  
library(cowplot)  
library("ggpubr")

### 3.1.2 Set up working Directory

    Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project.  
    Please refer Appendix A for Source Code.

########################  
## 1.2. Set Up Working Directory  
setwd("C:/Users/koksoon.chai/OneDrive - Kaer Pte Ltd/Machine\_learning/GreatLearning/Week 3")

### 3.1.3 Import and Read the Dataset

     The given dataset is in .csv format. Hence, the command ‘CardioGoodFitness.csv’ is used for importing the file.     Please refer Appendix A for Source Code.

########################  
## 1.3. Importing the dataset of cardio fitness  
file\_name <- "CardioGoodFitness.csv"  
cardio\_fitness\_data = read\_csv(file\_name)  
summary(cardio\_fitness\_data) # List all the parameters and view their statistic information for further analysis

## Product Age Gender Education   
## Length:180 Min. :18.00 Length:180 Min. :12.00   
## Class :character 1st Qu.:24.00 Class :character 1st Qu.:14.00   
## Mode :character Median :26.00 Mode :character Median :16.00   
## Mean :28.79 Mean :15.57   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :21.00   
## MaritalStatus Usage Fitness Income   
## Length:180 Min. :2.000 Min. :1.000 Min. : 29562   
## Class :character 1st Qu.:3.000 1st Qu.:3.000 1st Qu.: 44059   
## Mode :character Median :3.000 Median :3.000 Median : 50597   
## Mean :3.456 Mean :3.311 Mean : 53720   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.: 58668   
## Max. :7.000 Max. :5.000 Max. :104581   
## Miles   
## Min. : 21.0   
## 1st Qu.: 66.0   
## Median : 94.0   
## Mean :103.2   
## 3rd Qu.:114.8   
## Max. :360.0

View(cardio\_fitness\_data) # View the data of the parameters

## 3.2 Variable Identification

The following source code is written to display the structure of dataset

##################################################  
# 2. Perform uni-variate analysis  
##################################################  
  
########################  
## 2.1 Understanding the structure of dataset  
colnames(cardio\_fitness\_data) # retrieve the column names of the cardio fitness data set

## [1] "Product" "Age" "Gender" "Education"   
## [5] "MaritalStatus" "Usage" "Fitness" "Income"   
## [9] "Miles"

colnames(cardio\_fitness\_data) = make.names(colnames(cardio\_fitness\_data)) # Make syntactically valid names out of character vectors  
cardio\_fitness\_data$Gender = as.factor(cardio\_fitness\_data$Gender) # set the factor levels for gender  
cardio\_fitness\_data$MaritalStatus = as.factor(cardio\_fitness\_data$MaritalStatus) # set the factor levels for marital status  
cardio\_fitness\_data$Product = as.factor(cardio\_fitness\_data$Product) # set the factor levels for product

### 3.2.1 Variable Identification – Inferences

knitr::opts\_chunk$set(echo = TRUE)  
dim(cardio\_fitness\_data) # returns the dimension (e.g. the number of columns and rows) of a matrix

## [1] 180 9

names(cardio\_fitness\_data) # to get or set the names of an object

## [1] "Product" "Age" "Gender" "Education"   
## [5] "MaritalStatus" "Usage" "Fitness" "Income"   
## [9] "Miles"

str(cardio\_fitness\_data) # display the structure of an R object

## Classes 'spec\_tbl\_df', 'tbl\_df', 'tbl' and 'data.frame': 180 obs. of 9 variables:  
## $ Product : Factor w/ 3 levels "TM195","TM498",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Age : num 18 19 19 19 20 20 21 21 21 21 ...  
## $ Gender : Factor w/ 2 levels "Female","Male": 2 2 1 2 2 1 1 2 2 1 ...  
## $ Education : num 14 15 14 12 13 14 14 13 15 15 ...  
## $ MaritalStatus: Factor w/ 2 levels "Partnered","Single": 2 2 1 2 1 1 1 2 2 1 ...  
## $ Usage : num 3 2 4 3 4 3 3 3 5 2 ...  
## $ Fitness : num 4 3 3 3 2 3 3 3 4 3 ...  
## $ Income : num 29562 31836 30699 32973 35247 ...  
## $ Miles : num 112 75 66 85 47 66 75 85 141 85 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Product = col\_character(),  
## .. Age = col\_double(),  
## .. Gender = col\_character(),  
## .. Education = col\_double(),  
## .. MaritalStatus = col\_character(),  
## .. Usage = col\_double(),  
## .. Fitness = col\_double(),  
## .. Income = col\_double(),  
## .. Miles = col\_double()  
## .. )

head(cardio\_fitness\_data, 10) # display the first 10 rows of the cardio fitness data

## # A tibble: 10 x 9  
## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## <fct> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 TM195 18 Male 14 Single 3 4 29562 112  
## 2 TM195 19 Male 15 Single 2 3 31836 75  
## 3 TM195 19 Female 14 Partnered 4 3 30699 66  
## 4 TM195 19 Male 12 Single 3 3 32973 85  
## 5 TM195 20 Male 13 Partnered 4 2 35247 47  
## 6 TM195 20 Female 14 Partnered 3 3 32973 66  
## 7 TM195 21 Female 14 Partnered 3 3 35247 75  
## 8 TM195 21 Male 13 Single 3 3 32973 85  
## 9 TM195 21 Male 15 Single 5 4 35247 141  
## 10 TM195 21 Female 15 Partnered 2 3 37521 85

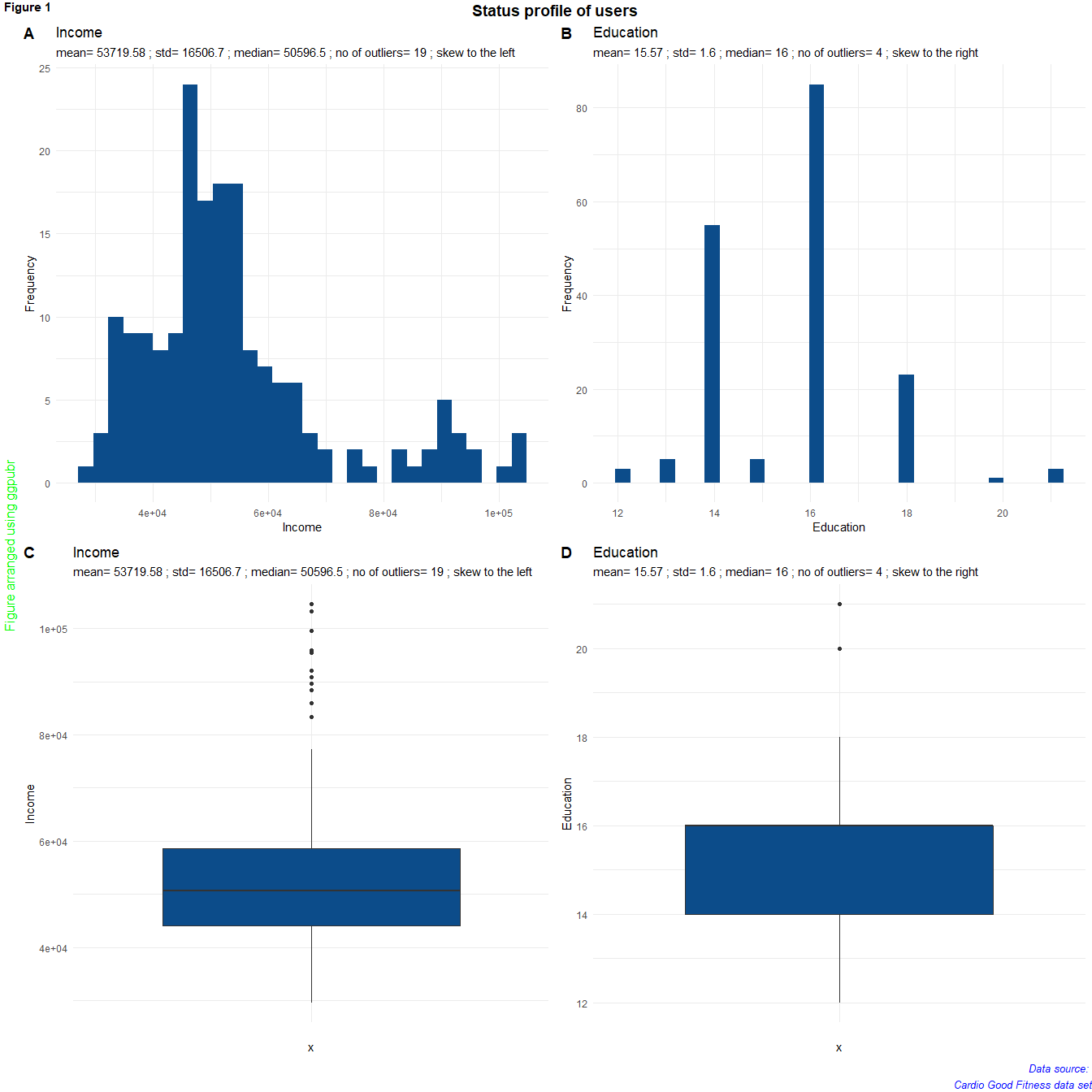
tail(cardio\_fitness\_data, -10) # display the last 10 rows of the cardio fitness data

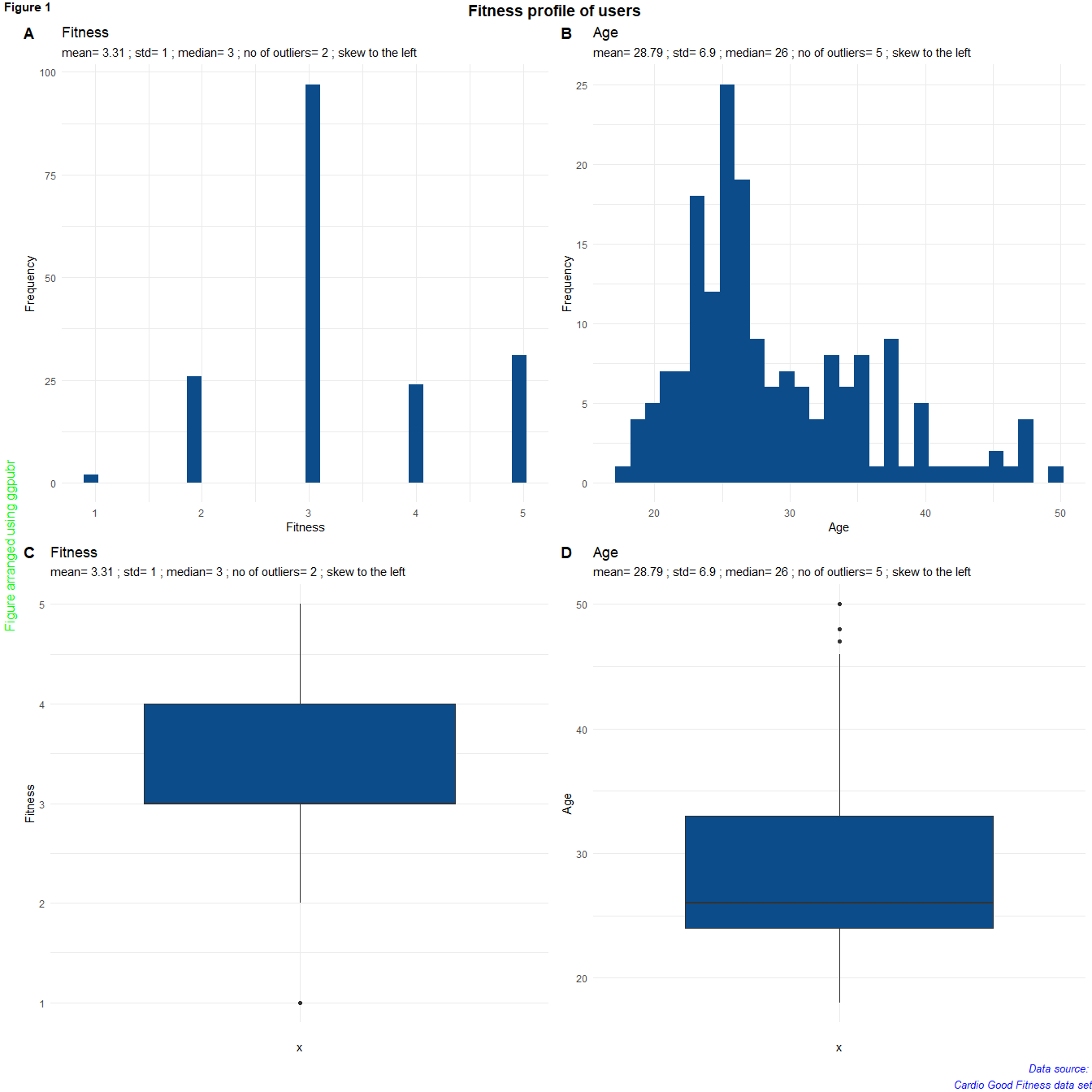
## # A tibble: 170 x 9  
## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## <fct> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 TM195 22 Male 14 Single 3 3 36384 85  
## 2 TM195 22 Female 14 Partnered 3 2 35247 66  
## 3 TM195 22 Female 16 Single 4 3 36384 75  
## 4 TM195 22 Female 14 Single 3 3 35247 75  
## 5 TM195 23 Male 16 Partnered 3 1 38658 47  
## 6 TM195 23 Male 16 Partnered 3 3 40932 75  
## 7 TM195 23 Female 14 Single 2 3 34110 103  
## 8 TM195 23 Male 16 Partnered 4 3 39795 94  
## 9 TM195 23 Female 16 Single 4 3 38658 113  
## 10 TM195 23 Female 15 Partnered 2 2 34110 38  
## # ... with 160 more rows

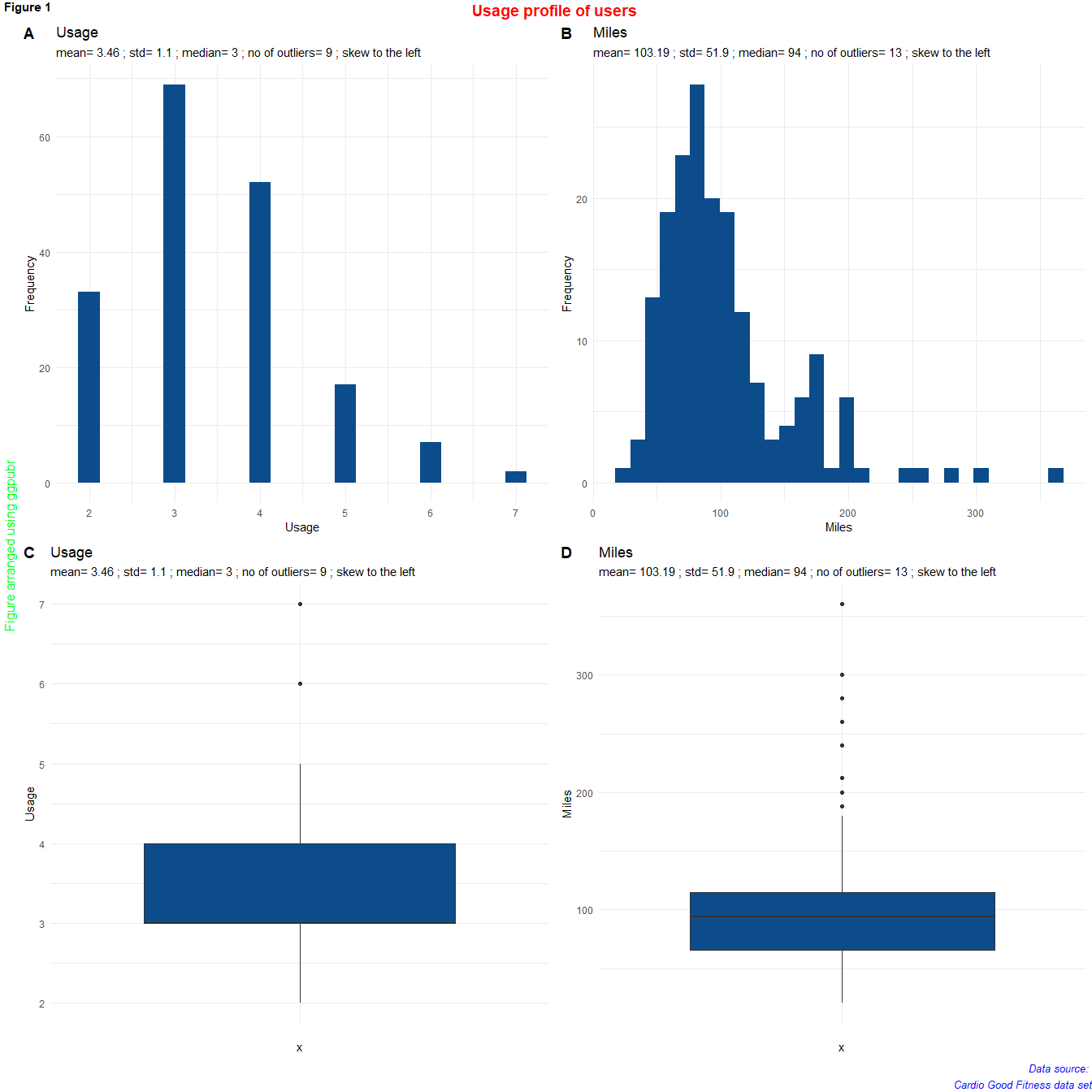
## 3.3 Univariate Analysis

The following source code is written to perform univariate analysis of the data set

knitr::opts\_chunk$set(echo = FALSE)  
# 2.4 Generate status profile of users  
status\_profile <- ggarrange(income\_bin, edu\_bin, income, education, nrow = 2, ncol= 2, labels = "AUTO")  
annotate\_figure(status\_profile,  
 top = text\_grob("Status profile of users", face = "bold", size = 14),  
 bottom = text\_grob("Data source: \n Cardio Good Fitness data set", color = "blue",  
 hjust = 1, x = 1, face = "italic", size = 10),  
 left = text\_grob("Figure arranged using ggpubr", color = "green", rot = 90),  
 fig.lab = "Figure 1", fig.lab.face = "bold"  
)

    
The statistic information for income variable is {mean= 3.46 ; std= 1.1 ; median= 3 ; no of outliers= 9 ; skew to the left} and the statistic information for education variable is {mean= 15.57 ; std= 1.6 ; median= 16 ; no of outliers= 4 ; skew to the right}.

    
The statistic information for variable fitness is {mean= 3.31 ; std= 1 ; median= 3 ; no of outliers= 2 ; skew to the left} and there statistic information for variable age is {mean= 28.79 ; std= 6.9 ; median= 26 ; no of outliers= 5 ; skew to the left}.

    
The statistic information for variable fitness is {mean= 3.31 ; std= 1 ; median= 3 ; no of outliers= 2 ; skew to the left} and there statistic information for variable miles is {mean= 103.19 ; std= 51.9 ; median= 94 ; no of outliers= 13 ; skew to the left}.

## 3.4 Bi-Variate Analysis

##   
## Female Male  
## TM195 40 40  
## TM498 29 31  
## TM798 7 33

##   
## Female Male  
## TM195 22.222222 22.222222  
## TM498 16.111111 17.222222  
## TM798 3.888889 18.333333

##   
## Partnered Single  
## TM195 48 32  
## TM498 36 24  
## TM798 23 17

##   
## Partnered Single  
## TM195 26.666667 17.777778  
## TM498 20.000000 13.333333  
## TM798 12.777778 9.444444

##   
## Young Age Middle Age Senior Age  
## TM195 55 19 6  
## TM498 35 23 2  
## TM798 30 6 4

##   
## Young Age Middle Age Senior Age  
## TM195 30.555556 10.555556 3.333333  
## TM498 19.444444 12.777778 1.111111  
## TM798 16.666667 3.333333 2.222222

##   
## 1 2 3 4 5  
## TM195 1 14 54 9 2  
## TM498 1 12 39 8 0  
## TM798 0 0 4 7 29

##   
## 1 2 3 4 5  
## TM195 0.5555556 7.7777778 30.0000000 5.0000000 1.1111111  
## TM498 0.5555556 6.6666667 21.6666667 4.4444444 0.0000000  
## TM798 0.0000000 0.0000000 2.2222222 3.8888889 16.1111111

##   
## Low Education Medium Education High Education  
## TM195 39 41 0  
## TM498 27 33 0  
## TM798 2 34 4

##   
## Low Education Medium Education High Education  
## TM195 21.666667 22.777778 0.000000  
## TM498 15.000000 18.333333 0.000000  
## TM798 1.111111 18.888889 2.222222

##   
## Low Income Medium Income High Income  
## TM195 62 18 0  
## TM498 47 13 0  
## TM798 8 13 19

##   
## Low Income Medium Income High Income  
## TM195 34.444444 10.000000 0.000000  
## TM498 26.111111 7.222222 0.000000  
## TM798 4.444444 7.222222 10.555556

##   
## Rare exercise Moderate exercise High exercise Very high exercise  
## TM195 12 50 18 0  
## TM498 5 39 15 1  
## TM798 0 8 27 4

##   
## Rare exercise Moderate exercise High exercise Very high exercise  
## TM195 6.7039106 27.9329609 10.0558659 0.0000000  
## TM498 2.7932961 21.7877095 8.3798883 0.5586592  
## TM798 0.0000000 4.4692737 15.0837989 2.2346369

##   
## 2 3 4 5 6 7  
## TM195 19 37 22 2 0 0  
## TM498 14 31 12 3 0 0  
## TM798 0 1 18 12 7 2

##   
## 2 3 4 5 6 7  
## TM195 10.5555556 20.5555556 12.2222222 1.1111111 0.0000000 0.0000000  
## TM498 7.7777778 17.2222222 6.6666667 1.6666667 0.0000000 0.0000000  
## TM798 0.0000000 0.5555556 10.0000000 6.6666667 3.8888889 1.1111111

## cardio\_fitness\_data$Product: TM195  
## Product Age Gender Education MaritalStatus  
## TM195:80 Min. :18.00 Female:40 Min. :12.00 Partnered:48   
## TM498: 0 1st Qu.:23.00 Male :40 1st Qu.:14.00 Single :32   
## TM798: 0 Median :26.00 Median :16.00   
## Mean :28.55 Mean :15.04   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :18.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :1.000 Min. :29562 Min. : 38.00   
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:38658 1st Qu.: 66.00   
## Median :3.000 Median :3.000 Median :46617 Median : 85.00   
## Mean :3.087 Mean :2.962 Mean :46418 Mean : 82.79   
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:53439 3rd Qu.: 94.00   
## Max. :5.000 Max. :5.000 Max. :68220 Max. :188.00   
## income\_bins education\_bins age\_bins   
## Low Income :62 Low Education :39 Young Age :55   
## Medium Income:18 Medium Education:41 Middle Age:19   
## High Income : 0 High Education : 0 Senior Age: 6   
##   
##   
##   
## miles\_bins  
## Rare exercise :12   
## Moderate exercise :50   
## High exercise :18   
## Very high exercise: 0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Product: TM498  
## Product Age Gender Education MaritalStatus  
## TM195: 0 Min. :19.00 Female:29 Min. :12.00 Partnered:36   
## TM498:60 1st Qu.:24.00 Male :31 1st Qu.:14.00 Single :24   
## TM798: 0 Median :26.00 Median :16.00   
## Mean :28.90 Mean :15.12   
## 3rd Qu.:33.25 3rd Qu.:16.00   
## Max. :48.00 Max. :18.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :1.0 Min. :31836 Min. : 21.00   
## 1st Qu.:3.000 1st Qu.:3.0 1st Qu.:44912 1st Qu.: 64.00   
## Median :3.000 Median :3.0 Median :49460 Median : 85.00   
## Mean :3.067 Mean :2.9 Mean :48974 Mean : 87.93   
## 3rd Qu.:3.250 3rd Qu.:3.0 3rd Qu.:53439 3rd Qu.:106.00   
## Max. :5.000 Max. :4.0 Max. :67083 Max. :212.00   
## income\_bins education\_bins age\_bins   
## Low Income :47 Low Education :27 Young Age :35   
## Medium Income:13 Medium Education:33 Middle Age:23   
## High Income : 0 High Education : 0 Senior Age: 2   
##   
##   
##   
## miles\_bins  
## Rare exercise : 5   
## Moderate exercise :39   
## High exercise :15   
## Very high exercise: 1   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Product: TM798  
## Product Age Gender Education MaritalStatus  
## TM195: 0 Min. :22.00 Female: 7 Min. :14.00 Partnered:23   
## TM498: 0 1st Qu.:24.75 Male :33 1st Qu.:16.00 Single :17   
## TM798:40 Median :27.00 Median :18.00   
## Mean :29.10 Mean :17.32   
## 3rd Qu.:30.25 3rd Qu.:18.00   
## Max. :48.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :3.000 Min. :3.000 Min. : 48556 Min. : 80.0   
## 1st Qu.:4.000 1st Qu.:4.000 1st Qu.: 58205 1st Qu.:120.0   
## Median :5.000 Median :5.000 Median : 76569 Median :160.0   
## Mean :4.775 Mean :4.625 Mean : 75442 Mean :166.9   
## 3rd Qu.:5.000 3rd Qu.:5.000 3rd Qu.: 90886 3rd Qu.:200.0   
## Max. :7.000 Max. :5.000 Max. :104581 Max. :360.0   
## income\_bins education\_bins age\_bins   
## Low Income : 8 Low Education : 2 Young Age :30   
## Medium Income:13 Medium Education:34 Middle Age: 6   
## High Income :19 High Education : 4 Senior Age: 4   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise : 8   
## High exercise :27   
## Very high exercise: 4   
## NA's : 1   
##

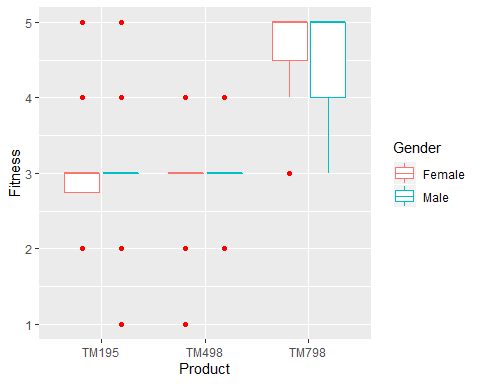
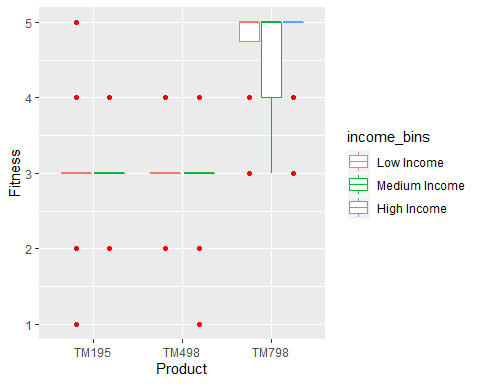
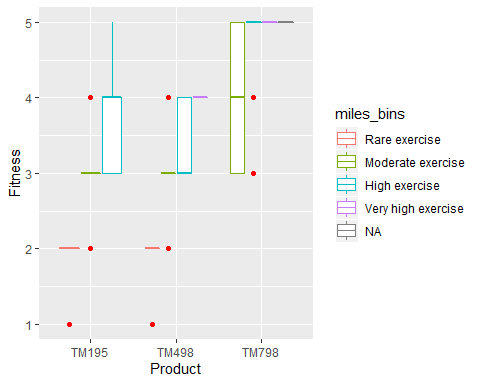
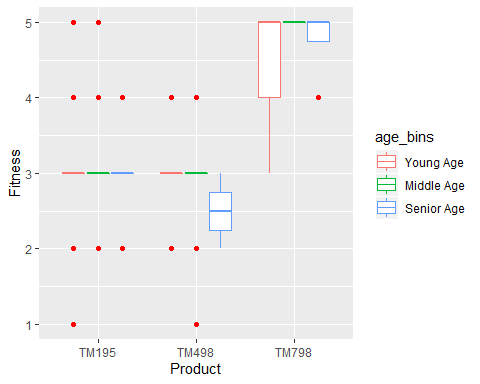
## cardio\_fitness\_data$Fitness: 1  
## Product Age Gender Education MaritalStatus  
## TM195:1 Min. :23 Female:1 Min. :16.0 Partnered:1   
## TM498:1 1st Qu.:25 Male :1 1st Qu.:16.5 Single :1   
## TM798:0 Median :27 Median :17.0   
## Mean :27 Mean :17.0   
## 3rd Qu.:29 3rd Qu.:17.5   
## Max. :31 Max. :18.0   
## Usage Fitness Income Miles income\_bins  
## Min. :2.00 Min. :1 Min. :38658 Min. :21.0 Low Income :1   
## 1st Qu.:2.25 1st Qu.:1 1st Qu.:45299 1st Qu.:27.5 Medium Income:1   
## Median :2.50 Median :1 Median :51939 Median :34.0 High Income :0   
## Mean :2.50 Mean :1 Mean :51939 Mean :34.0   
## 3rd Qu.:2.75 3rd Qu.:1 3rd Qu.:58580 3rd Qu.:40.5   
## Max. :3.00 Max. :1 Max. :65220 Max. :47.0   
## education\_bins age\_bins miles\_bins  
## Low Education :0 Young Age :1 Rare exercise :2   
## Medium Education:2 Middle Age:1 Moderate exercise :0   
## High Education :0 Senior Age:0 High exercise :0   
## Very high exercise:0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 2  
## Product Age Gender Education MaritalStatus  
## TM195:14 Min. :20.00 Female:16 Min. :12.00 Partnered:18   
## TM498:12 1st Qu.:23.25 Male :10 1st Qu.:14.00 Single : 8   
## TM798: 0 Median :25.50 Median :14.00   
## Mean :28.27 Mean :14.73   
## 3rd Qu.:32.50 3rd Qu.:16.00   
## Max. :46.00 Max. :16.00   
## Usage Fitness Income Miles income\_bins  
## Min. :2.000 Min. :2 Min. :32973 Min. :38.00 Low Income :22   
## 1st Qu.:2.000 1st Qu.:2 1st Qu.:40932 1st Qu.:43.25 Medium Income: 4   
## Median :2.000 Median :2 Median :45480 Median :47.00 High Income : 0   
## Mean :2.538 Mean :2 Mean :46180 Mean :51.69   
## 3rd Qu.:3.000 3rd Qu.:2 3rd Qu.:53155 3rd Qu.:53.00   
## Max. :4.000 Max. :2 Max. :60261 Max. :85.00   
## education\_bins age\_bins miles\_bins  
## Low Education :15 Young Age :17 Rare exercise :15   
## Medium Education:11 Middle Age: 7 Moderate exercise :11   
## High Education : 0 Senior Age: 2 High exercise : 0   
## Very high exercise: 0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 3  
## Product Age Gender Education MaritalStatus  
## TM195:54 Min. :19.00 Female:45 Min. :12.00 Partnered:57   
## TM498:39 1st Qu.:23.00 Male :52 1st Qu.:14.00 Single :40   
## TM798: 4 Median :26.00 Median :16.00   
## Mean :28.75 Mean :15.24   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :3 Min. :30699 Min. : 53.00   
## 1st Qu.:3.000 1st Qu.:3 1st Qu.:40932 1st Qu.: 75.00   
## Median :3.000 Median :3 Median :48891 Median : 85.00   
## Mean :3.165 Mean :3 Mean :48853 Mean : 87.19   
## 3rd Qu.:4.000 3rd Qu.:3 3rd Qu.:54576 3rd Qu.: 95.00   
## Max. :5.000 Max. :3 Max. :88396 Max. :170.00   
## income\_bins education\_bins age\_bins   
## Low Income :71 Low Education :42 Young Age :64   
## Medium Income:25 Medium Education:54 Middle Age:28   
## High Income : 1 High Education : 1 Senior Age: 5   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise :78   
## High exercise :19   
## Very high exercise: 0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 4  
## Product Age Gender Education MaritalStatus  
## TM195:9 Min. :18.00 Female: 8 Min. :13.00 Partnered:13   
## TM498:8 1st Qu.:25.00 Male :16 1st Qu.:14.00 Single :11   
## TM798:7 Median :26.00 Median :16.00   
## Mean :28.33 Mean :15.79   
## 3rd Qu.:32.25 3rd Qu.:16.50   
## Max. :44.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :3.000 Min. :4 Min. : 29562 Min. : 74.0   
## 1st Qu.:3.000 1st Qu.:4 1st Qu.: 44059 1st Qu.:106.0   
## Median :4.000 Median :4 Median : 49460 Median :127.0   
## Mean :3.917 Mean :4 Mean : 54995 Mean :131.6   
## 3rd Qu.:5.000 3rd Qu.:4 3rd Qu.: 59906 3rd Qu.:160.0   
## Max. :6.000 Max. :4 Max. :103336 Max. :212.0   
## income\_bins education\_bins age\_bins   
## Low Income :15 Low Education :10 Young Age :17   
## Medium Income: 6 Medium Education:13 Middle Age: 5   
## High Income : 3 High Education : 1 Senior Age: 2   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise : 5   
## High exercise :18   
## Very high exercise: 1   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 5  
## Product Age Gender Education MaritalStatus  
## TM195: 2 Min. :22.00 Female: 6 Min. :14.00 Partnered:18   
## TM498: 0 1st Qu.:24.00 Male :25 1st Qu.:16.00 Single :13   
## TM798:29 Median :28.00 Median :16.00   
## Mean :29.81 Mean :17.06   
## 3rd Qu.:33.50 3rd Qu.:18.00   
## Max. :48.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :3.000 Min. :5 Min. : 44343 Min. : 80.0   
## 1st Qu.:4.000 1st Qu.:5 1st Qu.: 53536 1st Qu.:150.0   
## Median :5.000 Median :5 Median : 77191 Median :170.0   
## Mean :4.839 Mean :5 Mean : 74397 Mean :178.9   
## 3rd Qu.:5.500 3rd Qu.:5 3rd Qu.: 91509 3rd Qu.:200.0   
## Max. :7.000 Max. :5 Max. :104581 Max. :360.0   
## income\_bins education\_bins age\_bins   
## Low Income : 8 Low Education : 1 Young Age :21   
## Medium Income: 8 Medium Education:28 Middle Age: 7   
## High Income :15 High Education : 2 Senior Age: 3   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise : 3   
## High exercise :23   
## Very high exercise: 4   
## NA's : 1   
##

Table below shows that young people use three different types of product so ages do not seem to be the prominent reason for the users of TM798 to achieve exceptional fitness compared to other users.

## , , = Young Age  
##   
##   
## 1 2 3 4 5  
## TM195 0.5555556 5.0000000 21.1111111 3.3333333 0.5555556  
## TM498 0.0000000 4.4444444 12.2222222 2.7777778 0.0000000  
## TM798 0.0000000 0.0000000 2.2222222 3.3333333 11.1111111  
##   
## , , = Middle Age  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 2.2222222 6.6666667 1.1111111 0.5555556  
## TM498 0.5555556 1.6666667 8.8888889 1.6666667 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.0000000 3.3333333  
##   
## , , = Senior Age  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 0.5555556 2.2222222 0.5555556 0.0000000  
## TM498 0.0000000 0.5555556 0.5555556 0.0000000 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.5555556 1.6666667

Table below shows that young people use three different types of product so ages do not seem to be the prominent reason for the users of TM798 to achieve exceptional fitness compared to other users.

## , , = Rare exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.5586592 6.1452514 0.0000000 0.0000000 0.0000000  
## TM498 0.5586592 2.2346369 0.0000000 0.0000000 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
##   
## , , = Moderate exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 1.6759777 25.6983240 0.5586592 0.0000000  
## TM498 0.0000000 4.4692737 16.2011173 1.1173184 0.0000000  
## TM798 0.0000000 0.0000000 1.6759777 1.1173184 1.6759777  
##   
## , , = High exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 0.0000000 4.4692737 4.4692737 1.1173184  
## TM498 0.0000000 0.0000000 5.5865922 2.7932961 0.0000000  
## TM798 0.0000000 0.0000000 0.5586592 2.7932961 11.7318436  
##   
## , , = Very high exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## TM498 0.0000000 0.0000000 0.0000000 0.5586592 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.0000000 2.2346369

The table that links product to fitness has showed that majority of the users of product TM798 achieved high fitness of 4 and above. The following pictures are generated trying to compare the high fitness of product TM798 with other products. 

The figures above show that i) users of TM798 achieve higher fitness levels compared to other products, ii) users of TM798 achieve higher fitness levels even for similar rates of exercise.

## 3.5 Missing Value Identification

Number of missing data is 1.

## 3.6 Outlier Identification

## [1] "The outliers are:"  
## [1] 47 50 48 47 48

## [1] 5

# 4 Conclusion

The objective of the report is to understand the relationship of customers, products and their fitness benefits after using the products.

The products are used by similar percentage of male and female users, except that TM798 is purchased by more male users. About 66% of the users of the products are young people. This indicates that the products are affordable and it is comfirmed that about 66% of the users have income that is in the lower income bracket.

Tables and figures are used to compare the products versus the fitness of the corresponding product users. They indicate that the majority of users of TM798 are able to achieve higher fitness of 4 and 5. In fact, about 16.1% of the total users achieve highest fitness of 5 when using TM798. However, only about 1.1% of the total users achieve highest fitness of 5 when using products TM195 and TM498.

It is recommended that more study is done to find out the cost and price of TM798 which has showed the benefits of allowing users achieve much higher fitness levels compared to other products. It is interesting to find out that about 50% of the users of TM798 are in the lower and middle income brackets. Sales and marketing department may leverage on the excellent fitness level result of TM798 for strategic marketing.

# 5 Appendix A – Source Code

## Product Age Gender Education   
## Length:180 Min. :18.00 Length:180 Min. :12.00   
## Class :character 1st Qu.:24.00 Class :character 1st Qu.:14.00   
## Mode :character Median :26.00 Mode :character Median :16.00   
## Mean :28.79 Mean :15.57   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :21.00   
## MaritalStatus Usage Fitness Income   
## Length:180 Min. :2.000 Min. :1.000 Min. : 29562   
## Class :character 1st Qu.:3.000 1st Qu.:3.000 1st Qu.: 44059   
## Mode :character Median :3.000 Median :3.000 Median : 50597   
## Mean :3.456 Mean :3.311 Mean : 53720   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.: 58668   
## Max. :7.000 Max. :5.000 Max. :104581   
## Miles   
## Min. : 21.0   
## 1st Qu.: 66.0   
## Median : 94.0   
## Mean :103.2   
## 3rd Qu.:114.8   
## Max. :360.0

## [1] "Product" "Age" "Gender" "Education"   
## [5] "MaritalStatus" "Usage" "Fitness" "Income"   
## [9] "Miles"

## [1] 180 9

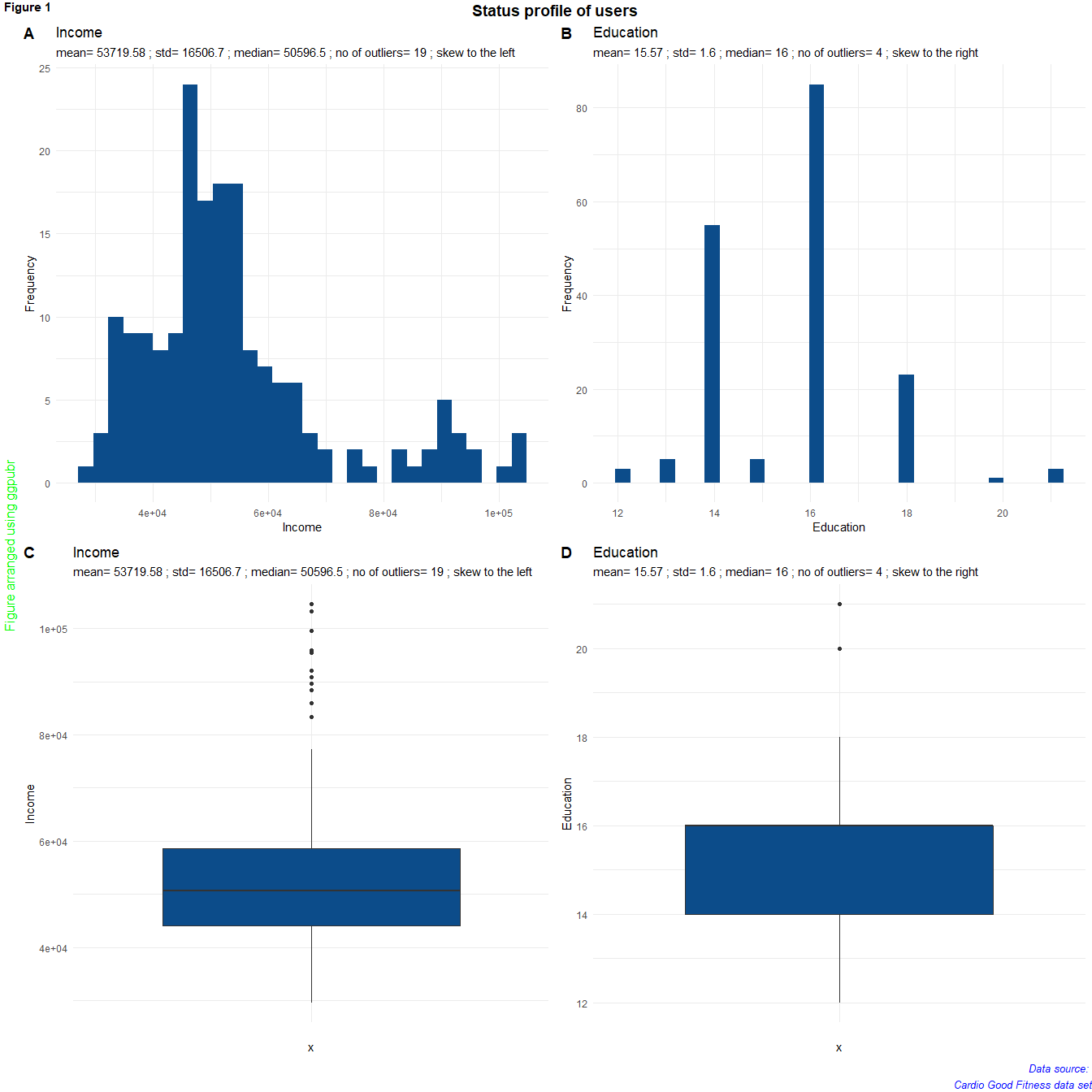
## [1] "Product" "Age" "Gender" "Education"   
## [5] "MaritalStatus" "Usage" "Fitness" "Income"   
## [9] "Miles"

## Classes 'spec\_tbl\_df', 'tbl\_df', 'tbl' and 'data.frame': 180 obs. of 9 variables:  
## $ Product : Factor w/ 3 levels "TM195","TM498",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Age : num 18 19 19 19 20 20 21 21 21 21 ...  
## $ Gender : Factor w/ 2 levels "Female","Male": 2 2 1 2 2 1 1 2 2 1 ...  
## $ Education : num 14 15 14 12 13 14 14 13 15 15 ...  
## $ MaritalStatus: Factor w/ 2 levels "Partnered","Single": 2 2 1 2 1 1 1 2 2 1 ...  
## $ Usage : num 3 2 4 3 4 3 3 3 5 2 ...  
## $ Fitness : num 4 3 3 3 2 3 3 3 4 3 ...  
## $ Income : num 29562 31836 30699 32973 35247 ...  
## $ Miles : num 112 75 66 85 47 66 75 85 141 85 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Product = col\_character(),  
## .. Age = col\_double(),  
## .. Gender = col\_character(),  
## .. Education = col\_double(),  
## .. MaritalStatus = col\_character(),  
## .. Usage = col\_double(),  
## .. Fitness = col\_double(),  
## .. Income = col\_double(),  
## .. Miles = col\_double()  
## .. )

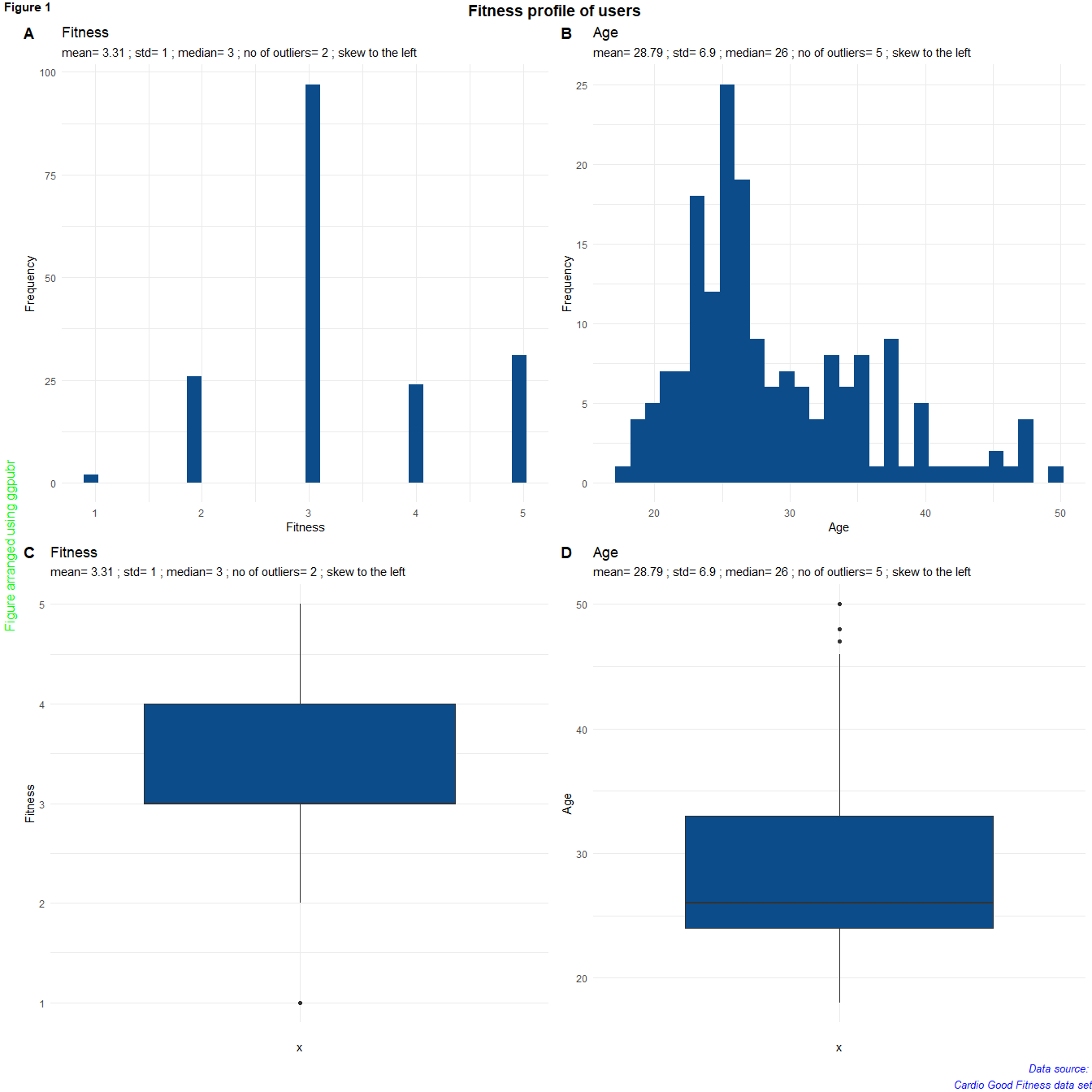
## # A tibble: 10 x 9  
## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## <fct> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 TM195 18 Male 14 Single 3 4 29562 112  
## 2 TM195 19 Male 15 Single 2 3 31836 75  
## 3 TM195 19 Female 14 Partnered 4 3 30699 66  
## 4 TM195 19 Male 12 Single 3 3 32973 85  
## 5 TM195 20 Male 13 Partnered 4 2 35247 47  
## 6 TM195 20 Female 14 Partnered 3 3 32973 66  
## 7 TM195 21 Female 14 Partnered 3 3 35247 75  
## 8 TM195 21 Male 13 Single 3 3 32973 85  
## 9 TM195 21 Male 15 Single 5 4 35247 141  
## 10 TM195 21 Female 15 Partnered 2 3 37521 85

## # A tibble: 170 x 9  
## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## <fct> <dbl> <fct> <dbl> <fct> <dbl> <dbl> <dbl> <dbl>  
## 1 TM195 22 Male 14 Single 3 3 36384 85  
## 2 TM195 22 Female 14 Partnered 3 2 35247 66  
## 3 TM195 22 Female 16 Single 4 3 36384 75  
## 4 TM195 22 Female 14 Single 3 3 35247 75  
## 5 TM195 23 Male 16 Partnered 3 1 38658 47  
## 6 TM195 23 Male 16 Partnered 3 3 40932 75  
## 7 TM195 23 Female 14 Single 2 3 34110 103  
## 8 TM195 23 Male 16 Partnered 4 3 39795 94  
## 9 TM195 23 Female 16 Single 4 3 38658 113  
## 10 TM195 23 Female 15 Partnered 2 2 34110 38  
## # ... with 160 more rows

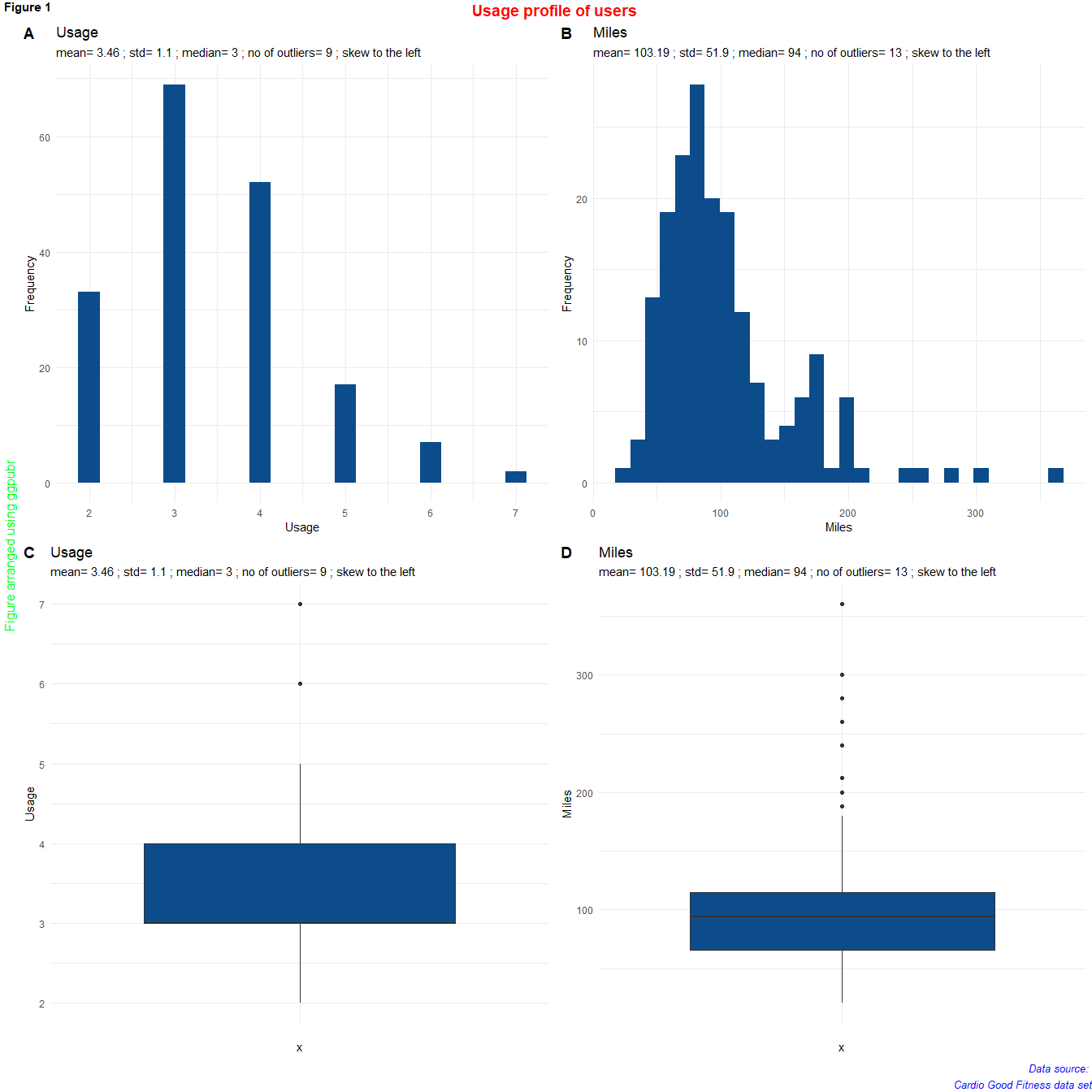
knitr::opts\_chunk$set(echo = TRUE)  
########################  
# 2.2. Explore the data  
count\_number\_of\_outliers <- function(x) {  
 Q3 = quantile(x,0.75)  
 Q3 = unname(Q3)  
 Q1 = unname(quantile(x,0.25))  
 age\_iqr = IQR(x)  
   
 Q3\_outlier <- Q3+1.5\*age\_iqr  
 Q1\_outlier <- Q1-1.5\*age\_iqr  
   
 # Select the outliers  
 outliers <- x[which( x < Q1\_outlier | x > Q3\_outlier )]  
   
 # Print the value of outliers  
 #print(outliers)  
   
 return(length(outliers))  
}  
  
explore\_variable <- function(x) {  
 mean\_x <- round(mean(x),2)  
   
 median\_x <- round(median(x),2)  
 range(x)  
  
 std\_x <- round(sd(x),1)  
   
 count\_number\_of\_outliers(x)  
 #print(x)  
   
 summary\_of\_variable <- paste("mean=", toString(mean\_x), "; std=", toString(std\_x), "; median=", toString(median\_x))  
 summary\_of\_variable <- paste(summary\_of\_variable, "; no of outliers=", toString(count\_number\_of\_outliers(x)))  
   
 # Check the centrality of data  
 if (median\_x < mean\_x)  
 summary\_of\_variable <- paste(summary\_of\_variable, "; skew to the left")  
 else  
 summary\_of\_variable <- paste(summary\_of\_variable, "; skew to the right")  
   
 # print(summary\_of\_variable)  
   
 return(summary\_of\_variable)  
}  
  
#explore\_variable(cardio\_fitness\_data$Age)  
#count\_number\_of\_outliers(cardio\_fitness\_data$Age)  
  
  
########################  
# 2.3 Univariate Analysis  
#esquisser()   
  
# Classification of various variables  
# 2.3.1 Plot the fequency of age bin  
age\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Age) +  
 geom\_histogram(bins = 30L, fill = "#0c4c8a") +  
 labs(x = "Age", y = "Frequency", title = "Age", subtitle = explore\_variable(cardio\_fitness\_data$Age)) +  
 theme\_minimal()  
  
age <- ggplot(cardio\_fitness\_data) +  
 aes(x = "", y = Age) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 labs(title = "Age", subtitle = explore\_variable(cardio\_fitness\_data$Age)) +  
 theme\_minimal()  
  
## 2.3.2 Plot the frequency of education bin  
edu\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Education) +  
 geom\_histogram(bins = 30L, fill = "#0c4c8a") +  
 labs(x = "Education", y = "Frequency", title = "Education", subtitle = explore\_variable(cardio\_fitness\_data$Education)) +  
 theme\_minimal()  
  
education <- ggplot(cardio\_fitness\_data) +  
 aes(x = "", y = Education) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 labs(title = "Education", subtitle = explore\_variable(cardio\_fitness\_data$Education)) +  
 theme\_minimal()  
  
## 2.3.3 Plot the frequency of gender bin  
gender\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Gender) +  
 geom\_bar(fill = "#0c4c8a") +  
 labs(x = "Gender", y = "Frequency", title = "Gender") +  
 theme\_minimal()  
  
# 2.3.4 Plot the frequency of gender bin  
marital\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = MaritalStatus) +  
 geom\_bar(fill = "#0c4c8a") +  
 labs(y = "Frequency", title = "Marital Status") +  
 theme\_minimal()  
  
# 2.3.5 Plot the frequency of usage bin  
usage\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Usage) +  
 geom\_histogram(bins = 21L, fill = "#0c4c8a") +  
 labs(y = "Frequency", title = "Usage", subtitle = explore\_variable(cardio\_fitness\_data$Usage)) +  
 theme\_minimal()  
  
usage <- ggplot(cardio\_fitness\_data) +  
 aes(x = "", y = Usage) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 labs(title = "Usage", subtitle = explore\_variable(cardio\_fitness\_data$Usage)) +  
 theme\_minimal()  
  
# 2.3.6 Plot the frequency of fitness bin  
fitness\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Fitness) +  
 geom\_histogram(bins = 30L, fill = "#0c4c8a") +  
 labs(y = "Frequency", title = "Fitness", subtitle = explore\_variable(cardio\_fitness\_data$Fitness)) +  
 theme\_minimal()  
  
fitness <- ggplot(cardio\_fitness\_data) +  
 aes(x = "", y = Fitness) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 labs(title = "Fitness", subtitle = explore\_variable(cardio\_fitness\_data$Fitness)) +  
 theme\_minimal()  
  
# 2.3.7 Plot the frequency of income bin  
income\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Income) +  
 geom\_histogram(bins = 30L, fill = "#0c4c8a") +  
 labs(y = "Frequency", title = "Income", subtitle = explore\_variable(cardio\_fitness\_data$Income)) +  
 theme\_minimal()  
  
income <- ggplot(cardio\_fitness\_data) +  
 aes(x = "", y = Income) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 labs(title = "Income", subtitle = explore\_variable(cardio\_fitness\_data$Income)) +  
 theme\_minimal()  
  
# 2.3.8 Plot the frequency of mile bin  
miles\_bin <- ggplot(cardio\_fitness\_data) +  
 aes(x = Miles) +  
 geom\_histogram(bins = 30L, fill = "#0c4c8a") +  
 labs(y = "Frequency", title = "Miles", subtitle = explore\_variable(cardio\_fitness\_data$Miles)) +  
 theme\_minimal()  
  
miles <- ggplot(cardio\_fitness\_data) +  
 aes(x = "", y = Miles) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 labs(title = "Miles", subtitle = explore\_variable(cardio\_fitness\_data$Miles)) +  
 theme\_minimal()  
  
  
# 2.4 Generate status profile of users  
status\_profile <- ggarrange(income\_bin, edu\_bin, income, education, nrow = 2, ncol= 2, labels = "AUTO")  
annotate\_figure(status\_profile,  
 top = text\_grob("Status profile of users", face = "bold", size = 14),  
 bottom = text\_grob("Data source: \n Cardio Good Fitness data set", color = "blue",  
 hjust = 1, x = 1, face = "italic", size = 10),  
 left = text\_grob("Figure arranged using ggpubr", color = "green", rot = 90),  
 fig.lab = "Figure 1", fig.lab.face = "bold"  
)



knitr::opts\_chunk$set(echo = TRUE)  
# 2.5 Generate fitness profile of users  
fitness\_profile <- ggarrange(fitness\_bin, age\_bin, fitness, age, nrow = 2, ncol= 2, labels = "AUTO")  
annotate\_figure(fitness\_profile,  
 top = text\_grob("Fitness profile of users", face = "bold", size = 14),  
 bottom = text\_grob("Data source: \n Cardio Good Fitness data set", color = "blue",  
 hjust = 1, x = 1, face = "italic", size = 10),  
 left = text\_grob("Figure arranged using ggpubr", color = "green", rot = 90),  
 fig.lab = "Figure 1", fig.lab.face = "bold"  
)



knitr::opts\_chunk$set(echo = TRUE)  
# 2.6 Generate usage profile of users  
usage\_profile <- ggarrange(usage\_bin, miles\_bin, usage, miles, nrow = 2, ncol= 2, labels = "AUTO")  
annotate\_figure(usage\_profile,  
 top = text\_grob("Usage profile of users", color = "red", face = "bold", size = 14),  
 bottom = text\_grob("Data source: \n Cardio Good Fitness data set", color = "blue",  
 hjust = 1, x = 1, face = "italic", size = 10),  
 left = text\_grob("Figure arranged using ggpubr", color = "green", rot = 90),  
 fig.lab = "Figure 1", fig.lab.face = "bold"  
)



knitr::opts\_chunk$set(echo = TRUE)  
##################################################  
# 3. Perform Bi-Variate Analysis  
##################################################  
  
cardio\_fitness\_data$income\_bins <- cut(cardio\_fitness\_data$Income, 3, labels=c("Low Income", "Medium Income", "High Income"))  
cardio\_fitness\_data$education\_bins <- cut(cardio\_fitness\_data$Education, 3, labels=c("Low Education", "Medium Education", "High Education"))  
cardio\_fitness\_data$age\_bins <- cut(cardio\_fitness\_data$Age, 3, breaks = c(0,30, 40, 100), labels=c("Young Age", "Middle Age", "Senior Age"))  
cardio\_fitness\_data$miles\_bins <- cut(cardio\_fitness\_data$Miles, 3, breaks = c(0, 50, 100, 200, 300), labels=c("Rare exercise", "Moderate exercise", "High exercise", "Very high exercise"))  
View(cardio\_fitness\_data)  
  
##--------------------------------------------------------------  
  
## 3.1 Products vs Gender  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Gender)

##   
## Female Male  
## TM195 40 40  
## TM498 29 31  
## TM798 7 33

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Gender))\*100

##   
## Female Male  
## TM195 22.222222 22.222222  
## TM498 16.111111 17.222222  
## TM798 3.888889 18.333333

##--------------------------------------------------------------  
   
## 3.2 Products vs marital status  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$MaritalStatus)

##   
## Partnered Single  
## TM195 48 32  
## TM498 36 24  
## TM798 23 17

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$MaritalStatus))\*100

##   
## Partnered Single  
## TM195 26.666667 17.777778  
## TM498 20.000000 13.333333  
## TM798 12.777778 9.444444

##--------------------------------------------------------------  
   
## 3.3 Products vs Age  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$age\_bins)

##   
## Young Age Middle Age Senior Age  
## TM195 55 19 6  
## TM498 35 23 2  
## TM798 30 6 4

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$age\_bins))\*100

##   
## Young Age Middle Age Senior Age  
## TM195 30.555556 10.555556 3.333333  
## TM498 19.444444 12.777778 1.111111  
## TM798 16.666667 3.333333 2.222222

##--------------------------------------------------------------  
   
## 3.4 Products vs fitness  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Fitness)

##   
## 1 2 3 4 5  
## TM195 1 14 54 9 2  
## TM498 1 12 39 8 0  
## TM798 0 0 4 7 29

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Fitness))\*100

##   
## 1 2 3 4 5  
## TM195 0.5555556 7.7777778 30.0000000 5.0000000 1.1111111  
## TM498 0.5555556 6.6666667 21.6666667 4.4444444 0.0000000  
## TM798 0.0000000 0.0000000 2.2222222 3.8888889 16.1111111

##--------------------------------------------------------------  
   
## 3.5 Products vs education  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$education\_bins)

##   
## Low Education Medium Education High Education  
## TM195 39 41 0  
## TM498 27 33 0  
## TM798 2 34 4

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$education\_bins))\*100

##   
## Low Education Medium Education High Education  
## TM195 21.666667 22.777778 0.000000  
## TM498 15.000000 18.333333 0.000000  
## TM798 1.111111 18.888889 2.222222

##--------------------------------------------------------------  
  
## 3.6 Products vs income  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$income\_bins)

##   
## Low Income Medium Income High Income  
## TM195 62 18 0  
## TM498 47 13 0  
## TM798 8 13 19

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$income\_bins))\*100

##   
## Low Income Medium Income High Income  
## TM195 34.444444 10.000000 0.000000  
## TM498 26.111111 7.222222 0.000000  
## TM798 4.444444 7.222222 10.555556

##--------------------------------------------------------------  
   
## 3.7 Product vs miles  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$miles\_bins)

##   
## Rare exercise Moderate exercise High exercise Very high exercise  
## TM195 12 50 18 0  
## TM498 5 39 15 1  
## TM798 0 8 27 4

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$miles\_bins))\*100

##   
## Rare exercise Moderate exercise High exercise Very high exercise  
## TM195 6.7039106 27.9329609 10.0558659 0.0000000  
## TM498 2.7932961 21.7877095 8.3798883 0.5586592  
## TM798 0.0000000 4.4692737 15.0837989 2.2346369

##--------------------------------------------------------------  
  
## 3.7 Product vs usage  
table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Usage)

##   
## 2 3 4 5 6 7  
## TM195 19 37 22 2 0 0  
## TM498 14 31 12 3 0 0  
## TM798 0 1 18 12 7 2

prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Usage))\*100

##   
## 2 3 4 5 6 7  
## TM195 10.5555556 20.5555556 12.2222222 1.1111111 0.0000000 0.0000000  
## TM498 7.7777778 17.2222222 6.6666667 1.6666667 0.0000000 0.0000000  
## TM798 0.0000000 0.5555556 10.0000000 6.6666667 3.8888889 1.1111111

##--------------------------------------------------------------  
  
by(cardio\_fitness\_data,INDICES = cardio\_fitness\_data$Product, FUN=summary)

## cardio\_fitness\_data$Product: TM195  
## Product Age Gender Education MaritalStatus  
## TM195:80 Min. :18.00 Female:40 Min. :12.00 Partnered:48   
## TM498: 0 1st Qu.:23.00 Male :40 1st Qu.:14.00 Single :32   
## TM798: 0 Median :26.00 Median :16.00   
## Mean :28.55 Mean :15.04   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :18.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :1.000 Min. :29562 Min. : 38.00   
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.:38658 1st Qu.: 66.00   
## Median :3.000 Median :3.000 Median :46617 Median : 85.00   
## Mean :3.087 Mean :2.962 Mean :46418 Mean : 82.79   
## 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:53439 3rd Qu.: 94.00   
## Max. :5.000 Max. :5.000 Max. :68220 Max. :188.00   
## income\_bins education\_bins age\_bins   
## Low Income :62 Low Education :39 Young Age :55   
## Medium Income:18 Medium Education:41 Middle Age:19   
## High Income : 0 High Education : 0 Senior Age: 6   
##   
##   
##   
## miles\_bins  
## Rare exercise :12   
## Moderate exercise :50   
## High exercise :18   
## Very high exercise: 0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Product: TM498  
## Product Age Gender Education MaritalStatus  
## TM195: 0 Min. :19.00 Female:29 Min. :12.00 Partnered:36   
## TM498:60 1st Qu.:24.00 Male :31 1st Qu.:14.00 Single :24   
## TM798: 0 Median :26.00 Median :16.00   
## Mean :28.90 Mean :15.12   
## 3rd Qu.:33.25 3rd Qu.:16.00   
## Max. :48.00 Max. :18.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :1.0 Min. :31836 Min. : 21.00   
## 1st Qu.:3.000 1st Qu.:3.0 1st Qu.:44912 1st Qu.: 64.00   
## Median :3.000 Median :3.0 Median :49460 Median : 85.00   
## Mean :3.067 Mean :2.9 Mean :48974 Mean : 87.93   
## 3rd Qu.:3.250 3rd Qu.:3.0 3rd Qu.:53439 3rd Qu.:106.00   
## Max. :5.000 Max. :4.0 Max. :67083 Max. :212.00   
## income\_bins education\_bins age\_bins   
## Low Income :47 Low Education :27 Young Age :35   
## Medium Income:13 Medium Education:33 Middle Age:23   
## High Income : 0 High Education : 0 Senior Age: 2   
##   
##   
##   
## miles\_bins  
## Rare exercise : 5   
## Moderate exercise :39   
## High exercise :15   
## Very high exercise: 1   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Product: TM798  
## Product Age Gender Education MaritalStatus  
## TM195: 0 Min. :22.00 Female: 7 Min. :14.00 Partnered:23   
## TM498: 0 1st Qu.:24.75 Male :33 1st Qu.:16.00 Single :17   
## TM798:40 Median :27.00 Median :18.00   
## Mean :29.10 Mean :17.32   
## 3rd Qu.:30.25 3rd Qu.:18.00   
## Max. :48.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :3.000 Min. :3.000 Min. : 48556 Min. : 80.0   
## 1st Qu.:4.000 1st Qu.:4.000 1st Qu.: 58205 1st Qu.:120.0   
## Median :5.000 Median :5.000 Median : 76569 Median :160.0   
## Mean :4.775 Mean :4.625 Mean : 75442 Mean :166.9   
## 3rd Qu.:5.000 3rd Qu.:5.000 3rd Qu.: 90886 3rd Qu.:200.0   
## Max. :7.000 Max. :5.000 Max. :104581 Max. :360.0   
## income\_bins education\_bins age\_bins   
## Low Income : 8 Low Education : 2 Young Age :30   
## Medium Income:13 Medium Education:34 Middle Age: 6   
## High Income :19 High Education : 4 Senior Age: 4   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise : 8   
## High exercise :27   
## Very high exercise: 4   
## NA's : 1   
##

by(cardio\_fitness\_data,INDICES = cardio\_fitness\_data$Fitness, FUN=summary)

## cardio\_fitness\_data$Fitness: 1  
## Product Age Gender Education MaritalStatus  
## TM195:1 Min. :23 Female:1 Min. :16.0 Partnered:1   
## TM498:1 1st Qu.:25 Male :1 1st Qu.:16.5 Single :1   
## TM798:0 Median :27 Median :17.0   
## Mean :27 Mean :17.0   
## 3rd Qu.:29 3rd Qu.:17.5   
## Max. :31 Max. :18.0   
## Usage Fitness Income Miles income\_bins  
## Min. :2.00 Min. :1 Min. :38658 Min. :21.0 Low Income :1   
## 1st Qu.:2.25 1st Qu.:1 1st Qu.:45299 1st Qu.:27.5 Medium Income:1   
## Median :2.50 Median :1 Median :51939 Median :34.0 High Income :0   
## Mean :2.50 Mean :1 Mean :51939 Mean :34.0   
## 3rd Qu.:2.75 3rd Qu.:1 3rd Qu.:58580 3rd Qu.:40.5   
## Max. :3.00 Max. :1 Max. :65220 Max. :47.0   
## education\_bins age\_bins miles\_bins  
## Low Education :0 Young Age :1 Rare exercise :2   
## Medium Education:2 Middle Age:1 Moderate exercise :0   
## High Education :0 Senior Age:0 High exercise :0   
## Very high exercise:0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 2  
## Product Age Gender Education MaritalStatus  
## TM195:14 Min. :20.00 Female:16 Min. :12.00 Partnered:18   
## TM498:12 1st Qu.:23.25 Male :10 1st Qu.:14.00 Single : 8   
## TM798: 0 Median :25.50 Median :14.00   
## Mean :28.27 Mean :14.73   
## 3rd Qu.:32.50 3rd Qu.:16.00   
## Max. :46.00 Max. :16.00   
## Usage Fitness Income Miles income\_bins  
## Min. :2.000 Min. :2 Min. :32973 Min. :38.00 Low Income :22   
## 1st Qu.:2.000 1st Qu.:2 1st Qu.:40932 1st Qu.:43.25 Medium Income: 4   
## Median :2.000 Median :2 Median :45480 Median :47.00 High Income : 0   
## Mean :2.538 Mean :2 Mean :46180 Mean :51.69   
## 3rd Qu.:3.000 3rd Qu.:2 3rd Qu.:53155 3rd Qu.:53.00   
## Max. :4.000 Max. :2 Max. :60261 Max. :85.00   
## education\_bins age\_bins miles\_bins  
## Low Education :15 Young Age :17 Rare exercise :15   
## Medium Education:11 Middle Age: 7 Moderate exercise :11   
## High Education : 0 Senior Age: 2 High exercise : 0   
## Very high exercise: 0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 3  
## Product Age Gender Education MaritalStatus  
## TM195:54 Min. :19.00 Female:45 Min. :12.00 Partnered:57   
## TM498:39 1st Qu.:23.00 Male :52 1st Qu.:14.00 Single :40   
## TM798: 4 Median :26.00 Median :16.00   
## Mean :28.75 Mean :15.24   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :3 Min. :30699 Min. : 53.00   
## 1st Qu.:3.000 1st Qu.:3 1st Qu.:40932 1st Qu.: 75.00   
## Median :3.000 Median :3 Median :48891 Median : 85.00   
## Mean :3.165 Mean :3 Mean :48853 Mean : 87.19   
## 3rd Qu.:4.000 3rd Qu.:3 3rd Qu.:54576 3rd Qu.: 95.00   
## Max. :5.000 Max. :3 Max. :88396 Max. :170.00   
## income\_bins education\_bins age\_bins   
## Low Income :71 Low Education :42 Young Age :64   
## Medium Income:25 Medium Education:54 Middle Age:28   
## High Income : 1 High Education : 1 Senior Age: 5   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise :78   
## High exercise :19   
## Very high exercise: 0   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 4  
## Product Age Gender Education MaritalStatus  
## TM195:9 Min. :18.00 Female: 8 Min. :13.00 Partnered:13   
## TM498:8 1st Qu.:25.00 Male :16 1st Qu.:14.00 Single :11   
## TM798:7 Median :26.00 Median :16.00   
## Mean :28.33 Mean :15.79   
## 3rd Qu.:32.25 3rd Qu.:16.50   
## Max. :44.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :3.000 Min. :4 Min. : 29562 Min. : 74.0   
## 1st Qu.:3.000 1st Qu.:4 1st Qu.: 44059 1st Qu.:106.0   
## Median :4.000 Median :4 Median : 49460 Median :127.0   
## Mean :3.917 Mean :4 Mean : 54995 Mean :131.6   
## 3rd Qu.:5.000 3rd Qu.:4 3rd Qu.: 59906 3rd Qu.:160.0   
## Max. :6.000 Max. :4 Max. :103336 Max. :212.0   
## income\_bins education\_bins age\_bins   
## Low Income :15 Low Education :10 Young Age :17   
## Medium Income: 6 Medium Education:13 Middle Age: 5   
## High Income : 3 High Education : 1 Senior Age: 2   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise : 5   
## High exercise :18   
## Very high exercise: 1   
##   
##   
## ------------------------------------------------------------   
## cardio\_fitness\_data$Fitness: 5  
## Product Age Gender Education MaritalStatus  
## TM195: 2 Min. :22.00 Female: 6 Min. :14.00 Partnered:18   
## TM498: 0 1st Qu.:24.00 Male :25 1st Qu.:16.00 Single :13   
## TM798:29 Median :28.00 Median :16.00   
## Mean :29.81 Mean :17.06   
## 3rd Qu.:33.50 3rd Qu.:18.00   
## Max. :48.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :3.000 Min. :5 Min. : 44343 Min. : 80.0   
## 1st Qu.:4.000 1st Qu.:5 1st Qu.: 53536 1st Qu.:150.0   
## Median :5.000 Median :5 Median : 77191 Median :170.0   
## Mean :4.839 Mean :5 Mean : 74397 Mean :178.9   
## 3rd Qu.:5.500 3rd Qu.:5 3rd Qu.: 91509 3rd Qu.:200.0   
## Max. :7.000 Max. :5 Max. :104581 Max. :360.0   
## income\_bins education\_bins age\_bins   
## Low Income : 8 Low Education : 1 Young Age :21   
## Medium Income: 8 Medium Education:28 Middle Age: 7   
## High Income :15 High Education : 2 Senior Age: 3   
##   
##   
##   
## miles\_bins  
## Rare exercise : 0   
## Moderate exercise : 3   
## High exercise :23   
## Very high exercise: 4   
## NA's : 1   
##

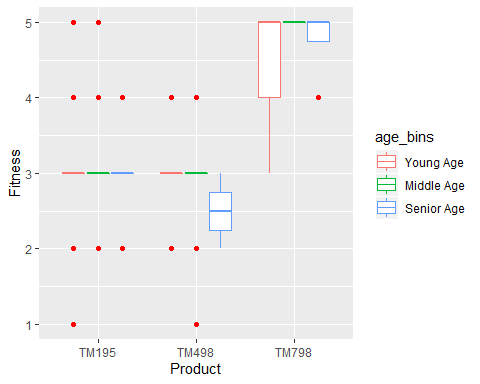
prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Fitness, cardio\_fitness\_data$age\_bins))\*100

## , , = Young Age  
##   
##   
## 1 2 3 4 5  
## TM195 0.5555556 5.0000000 21.1111111 3.3333333 0.5555556  
## TM498 0.0000000 4.4444444 12.2222222 2.7777778 0.0000000  
## TM798 0.0000000 0.0000000 2.2222222 3.3333333 11.1111111  
##   
## , , = Middle Age  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 2.2222222 6.6666667 1.1111111 0.5555556  
## TM498 0.5555556 1.6666667 8.8888889 1.6666667 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.0000000 3.3333333  
##   
## , , = Senior Age  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 0.5555556 2.2222222 0.5555556 0.0000000  
## TM498 0.0000000 0.5555556 0.5555556 0.0000000 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.5555556 1.6666667

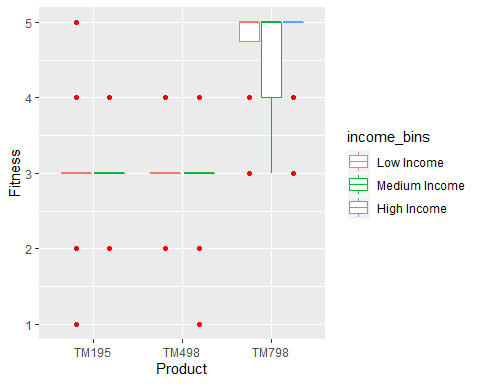
prop.table(table(cardio\_fitness\_data$Product, cardio\_fitness\_data$Fitness, cardio\_fitness\_data$miles\_bins))\*100

## , , = Rare exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.5586592 6.1452514 0.0000000 0.0000000 0.0000000  
## TM498 0.5586592 2.2346369 0.0000000 0.0000000 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
##   
## , , = Moderate exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 1.6759777 25.6983240 0.5586592 0.0000000  
## TM498 0.0000000 4.4692737 16.2011173 1.1173184 0.0000000  
## TM798 0.0000000 0.0000000 1.6759777 1.1173184 1.6759777  
##   
## , , = High exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 0.0000000 4.4692737 4.4692737 1.1173184  
## TM498 0.0000000 0.0000000 5.5865922 2.7932961 0.0000000  
## TM798 0.0000000 0.0000000 0.5586592 2.7932961 11.7318436  
##   
## , , = Very high exercise  
##   
##   
## 1 2 3 4 5  
## TM195 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000  
## TM498 0.0000000 0.0000000 0.0000000 0.5586592 0.0000000  
## TM798 0.0000000 0.0000000 0.0000000 0.0000000 2.2346369

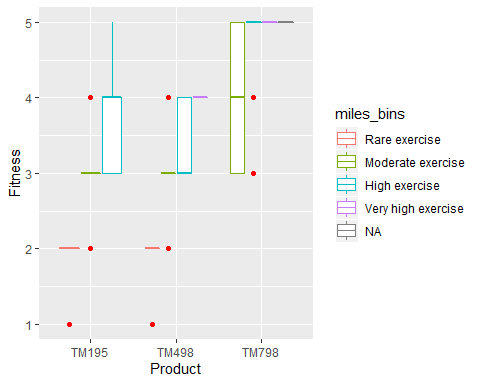
user\_with\_high\_fitness <- cardio\_fitness\_data[which(cardio\_fitness\_data$Fitness>=4),]  
  
#qplot(user\_with\_high\_fitness$Age~user\_with\_high\_fitness$Product, data = user\_with\_high\_fitness)  
  
#library(rpivotTable) # No need to explicitly load htmlwidgets: this is done automatically  
#rpivotTable(user\_with\_high\_fitness ,rows="Product", cols=c("Age","Fitness"),width="100%", height="100px")  
  
ggplot(data = cardio\_fitness\_data, mapping = aes(x = Product, y = Fitness)) + geom\_boxplot( aes(colour = age\_bins), outlier.colour = "red")



ggplot(data = cardio\_fitness\_data, mapping = aes(x = Product, y = Fitness)) + geom\_boxplot( aes(colour = income\_bins), outlier.colour = "red")



ggplot(data = cardio\_fitness\_data, mapping = aes(x = Product, y = Fitness)) + geom\_boxplot( aes(colour = miles\_bins), outlier.colour = "red")



count\_number\_of\_outliers <- function(x) {  
 Q3 = quantile(x,0.75)  
 Q3 = unname(Q3)  
 Q1 = unname(quantile(x,0.25))  
 age\_iqr = IQR(x)  
   
 Q3\_outlier <- Q3+1.5\*age\_iqr  
 Q1\_outlier <- Q1-1.5\*age\_iqr  
   
 # Select the outliers  
 outliers <- x[which( x < Q1\_outlier | x > Q3\_outlier )]  
   
 # Print the value of outliers  
 print("The outliers are:")  
 print(outliers)  
   
 return(length(outliers))  
}  
count\_number\_of\_outliers(cardio\_fitness\_data$Age)

## [1] "The outliers are:"  
## [1] 47 50 48 47 48

## [1] 5