**Mini Project - Cardio Good Fitness**

# 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 consists of the following:

* Importing the dataset in R
* Understanding the structure of dataset
* Graphical exploration
* Descriptive statistics
* Insights from the dataset

# Assumptions : NA

# 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. Variable Transformation / Feature Creation

6. Feature Exploration

## 3.1 Environment Set up and Data Import

## 3.1.1 Install necessary Packages and Invoke Libraries

Two packages has been used for this project.

Dplyr() : For data manipulation

ggplot2 () : For data visualization

## 3.1.2 Set up working Directory or Exporting data

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.

Command to get the current directory ‘getwd()’

To change the working directory ‘setwd()’

Or we can use file.choose function to choose the file from drive.

Cardio\_Goods=read.csv(file.choose())

Please refer Appendix A for Source Code.

### 3.1.3 Import and Read the Dataset

The given dataset is in .csv format. Hence, the command ‘read.csv’ is used for importing the file.

Please refer Appendix A for Source Code.

## **3**.2 Variable Identification

Here is list of all R commands used for variable identification

### 3.2.1 dim () - To identify rows and column in data set .

> dim(Cardio\_Goods)

[1] 180 9

### 3.2.2 str() - structure of data set

|  |
| --- |
| 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 : int 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 : int 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 : int 3 2 4 3 4 3 3 3 5 2 ...  $ Fitness : int 4 3 3 3 2 3 3 3 4 3 ...  $ Income : int 29562 31836 30699 32973 35247 32973 35247 32973 35247 37521 ...  $ Miles : int 112 75 66 85 47 66 75 85 141 85 ... |
|  |
| |  | | --- | | > | |

### 3.2.3 summary () - Summarise data set .

> summary(Cardio\_Goods)

Product Age Gender Education

TM195:80 Min. :18.00 Female: 76 Min. :12.00

TM498:60 1st Qu.:24.00 Male :104 1st Qu.:14.00

TM798:40 Median :26.00 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

Partnered:107 Min. :2.000 Min. :1.000

Single : 73 1st Qu.:3.000 1st Qu.:3.000

Median :3.000 Median :3.000

Mean :3.456 Mean :3.311

3rd Qu.:4.000 3rd Qu.:4.000

Max. :7.000 Max. :5.000

Income Miles

Min. : 29562 Min. : 21.0

1st Qu.: 44059 1st Qu.: 66.0

Median : 50597 Median : 94.0

Mean : 53720 Mean :103.2

3rd Qu.: 58668 3rd Qu.:114.8

Max. :104581 Max. :360.

### 3.2.4 class () - to find out the data type of object

> class(Cardio\_Goods)

[1] "data.frame"

### 3.2.5 library() -> confirm package listing .

> library(dplyr)

### 3.2.6 Head () – first six rows of data set .

|  |
| --- |
| head(Cardio\_Goods)  Product Age Gender Education MaritalStatus Usage Fitness Income Miles  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 |
|  |
| |  | | --- | | > | |

### 3.2.7 Tail() - last six rows of data set

|  |
| --- |
| > tail(Cardio\_Goods)  Product Age Gender Education MaritalStatus Usage Fitness Income Miles  175 TM798 38 Male 18 Partnered 5 5 104581 150  176 TM798 40 Male 21 Single 6 5 83416 200  177 TM798 42 Male 18 Single 5 4 89641 200  178 TM798 45 Male 16 Single 5 5 90886 160  179 TM798 47 Male 18 Partnered 4 5 104581 120  180 TM798 48 Male 18 Partnered 4 5 95508 180 |
|  |
| |  | | --- | |  | |

## 3.3 Univariate Analysis

This section talks about each feature of data set .

### 3.3.1 Product : Categorical variable

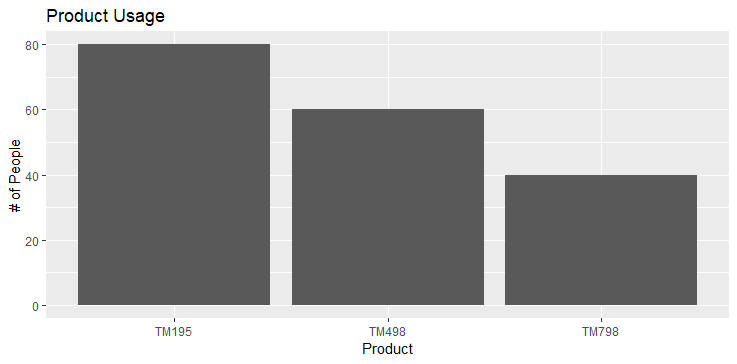
> summary(Cardio\_Goods$Product)

TM195 TM498 TM798

80 60 40

> ggplot(Cardio\_Goods\_tbl,aes(x=Product)) + geom\_bar() +

+ labs(y='# of People', title = 'Product Usage')



Inference : Total 3 products are being used in for Cardio in this data set .

TM195 is most used and TM798 is least used .

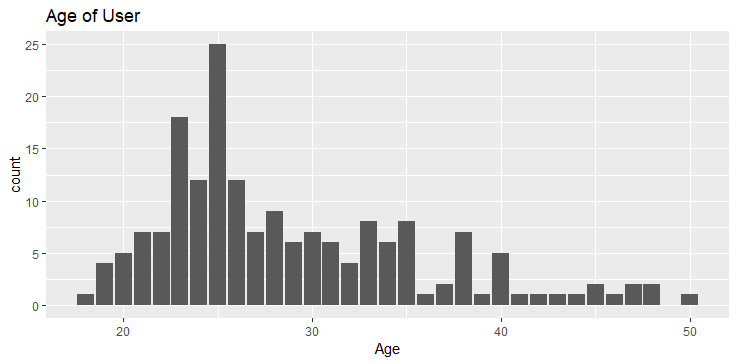
### **3.3.2 Age** : Continuous variable

> summary(Cardio\_Goods$Age)

Min. 1st Qu. Median Mean 3rd Qu. Max.

18.00 24.00 26.00 28.79 33.00 50.00

ggplot(Cardio\_Goods\_tbl,aes(x=Age)) + geom\_bar()



Inference : Summary shows that majority of population age in data set belongs to 24 to 33.

### **3.3.3. Gender** : Categorical variable

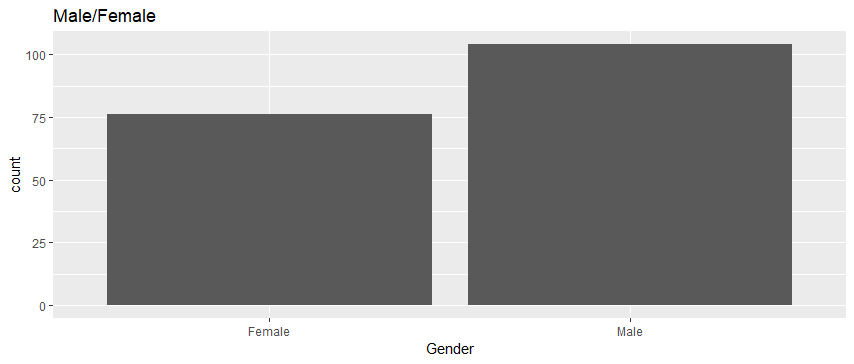
summary(Cardio\_Goods$Gender)

Female Male

76 104

> ggplot(Cardio\_Goods\_tbl,aes(x=Gender)) +

+ geom\_bar() + labs(y='count', title = 'Male/Female')



Inference: Summary shows that male population is more than female

### 3.3.4 Education : Continuous variable

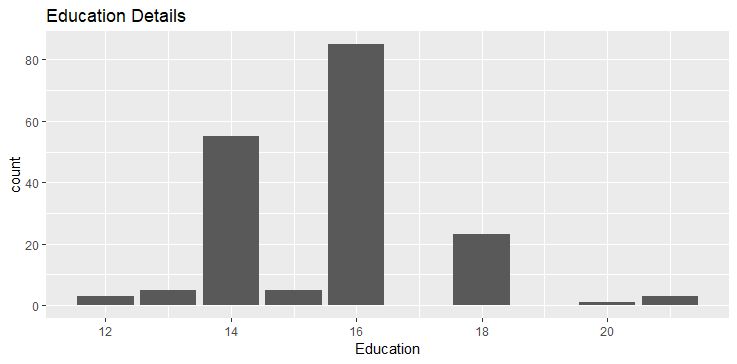
summary(Cardio\_Goods$Education)

Min. 1st Qu. Median Mean 3rd Qu. Max.

12.00 14.00 16.00 15.57 16.00 21.00

> ggplot(Cardio\_Goods\_tbl,aes(x=Education)) + geom\_bar() +

+ labs(y='count', title = 'Education Details')



Inference : Most of people in data set are in rage of 14 to 16 education .

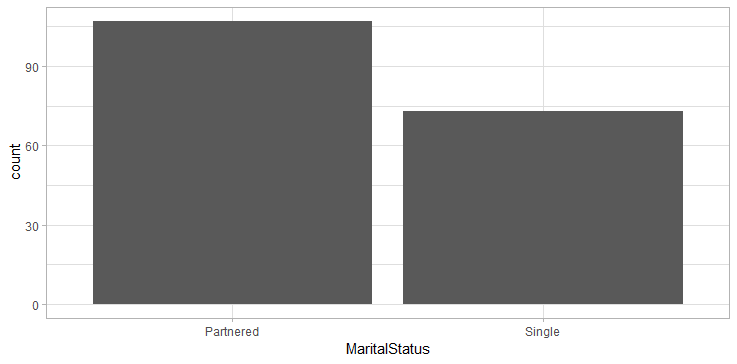
### **3.3.5 Marital Status**: Categorical variable

summary(Cardio\_Goods$MaritalStatus)

Partnered Single

107 73

> ggplot(Cardio\_Goods\_tbl,aes(x=MaritalStatus)) + geom\_bar() + theme\_light()



Inference : Summary shows that there are more people in data set who has partner compare to single .

### 3.3.6 Usage : continuous variable

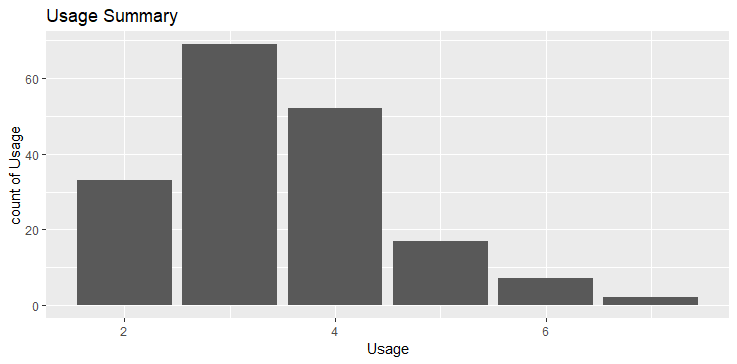
summary(Cardio\_Goods$Usage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.000 3.000 3.000 3.456 4.000 7.000

> ggplot(Cardio\_Goods\_tbl,aes(x=Usage)) + geom\_bar()+

+ labs(y='count of Usage', title = 'Usage Summary')



**Inference : Data shows that most of the users usage is between 3 and 4.**

### 3.3.7 Fitness: Continuous variable

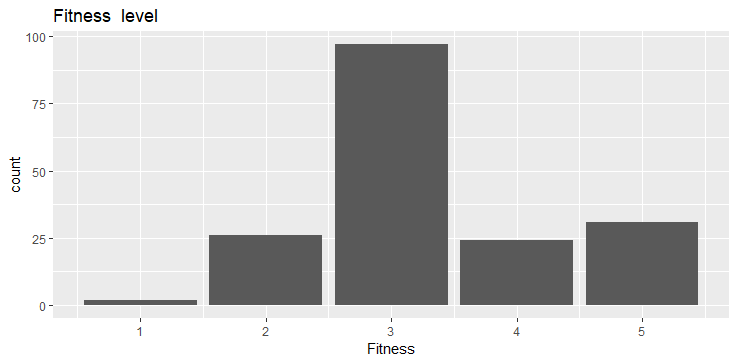
> summary(Cardio\_Goods$Fitness)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 3.000 3.000 3.311 4.000 5.000

> ggplot(Cardio\_Goods\_tbl,aes(x=Fitness)) + geom\_bar() +

+ labs(y='count', title = 'Fitness level')



**Inference : most of users fitness at level 3 .**

### 3.3.8 Income : Continuous variable

summary(Cardio\_Goods$Income)

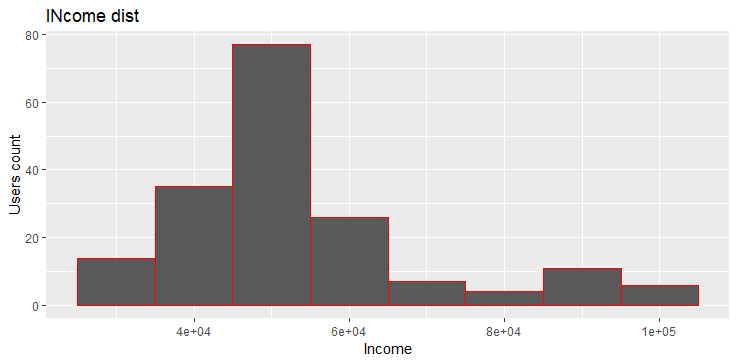
Min. 1st Qu. Median Mean 3rd Qu. Max.

29562 44059 50597 53720 58668 104581

ggplot(Cardio\_Goods\_tbl, aes(x=Income)) +

+ geom\_histogram(binwidth = 10000, colour='red' )+

+ labs(y='Users count', title = 'INcome dist')



Inference : Most of users income is between 48k to 58k

### **3.3.9 Miles :** continuous variable

summary(Cardio\_Goods$Miles)

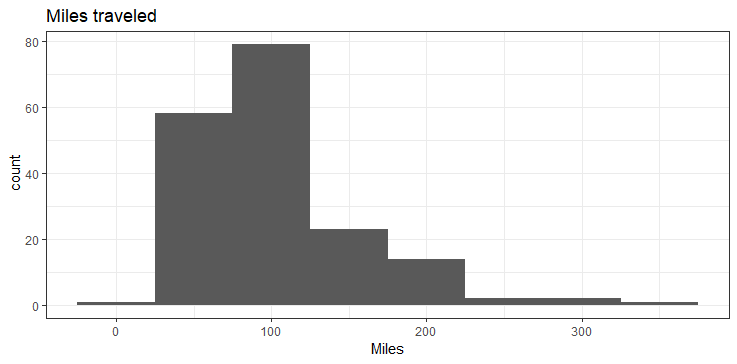
Min. 1st Qu. Median Mean 3rd Qu. Max.

21.0 66.0 94.0 103.2 114.8 360.0

ggplot(Cardio\_Goods\_tbl, aes(x=Miles)) +

+ geom\_histogram(binwidth = 50 )+ theme\_bw() +

+ labs(y=' count', title = 'Miles traveled')



Inference : Data shows that mean of miles is around 103.

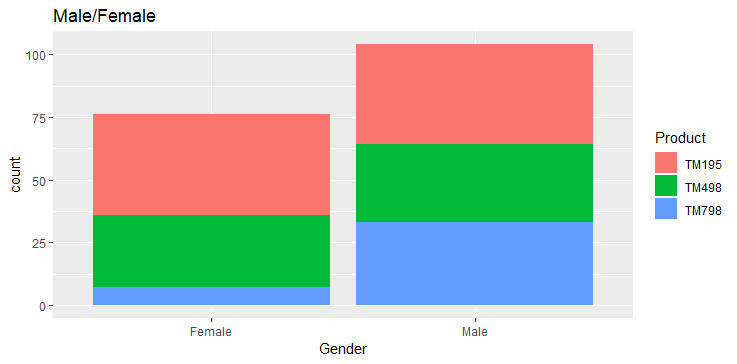
## 3.4 Bi-Variate Analysis

This section is about identifying relationship between two features.

### 3.4.1 Product and Gender

> ggplot(Cardio\_Goods\_tbl,aes(x=Gender, fill=Product)) +

+ geom\_bar() + labs(y='count', title = 'Male/Female')



**Inference :**

TM798 product is not much used by females as they are mostly using TM195.

### 3.4.2 Product and age

> Cardio\_Goods\_tbl %>%

+ group\_by(Product) %>%

+ summarise(Age\_mean = mean(Age))

# A tibble: 3 x 2

Product Age\_mean

*<fct>* *<dbl>*

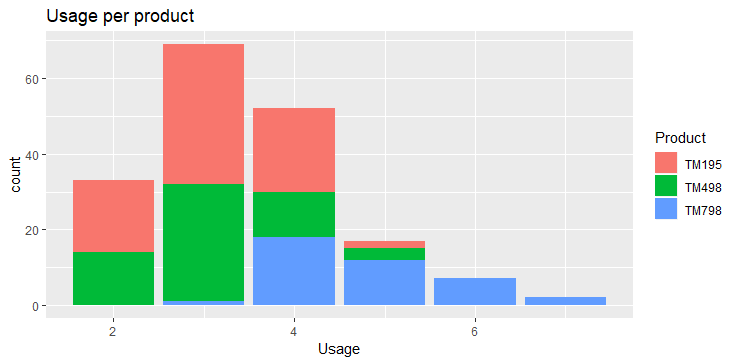
1 TM195 28.6

2 TM498 28.9

3 TM798 29.1

Mean age per product is nearly same .

### 3.4.3. Usage and Product



> Cardio\_Goods\_tbl %>%

+ group\_by(Product) %>%

+ summarise(Fit\_mean = mean(Usage))

# A tibble: 3 x 2

Product Fit\_mean

*<fct>* *<dbl>*

1 TM195 3.09

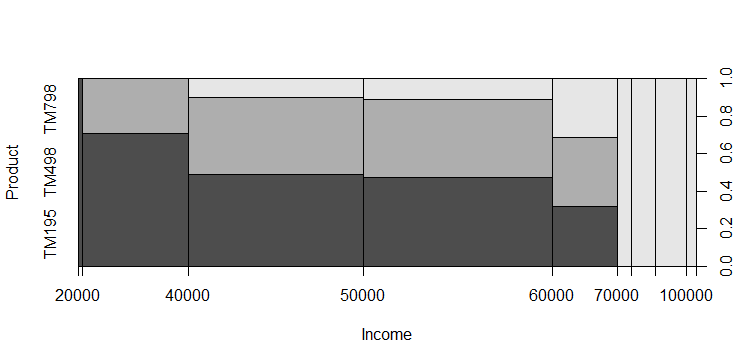
2 TM498 3.07

3 TM798 4.78

**Inference : TM798 product is mostly used 4 or more days.**

### 3.4.4 Product and income

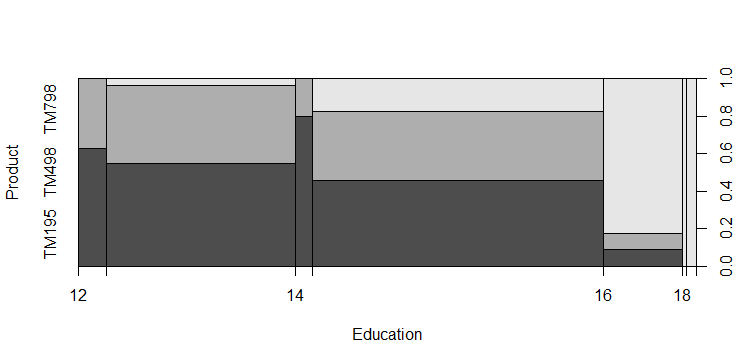
plot(Product ~ Income , Cardio\_Goods\_tbl)



**Inference : Product TM798 is being used by higher income user .**

**Product and Education :**

> plot(Product ~ Education , Cardio\_Goods\_tbl)



**Inference : high educated users are mostly using TM798 product .**

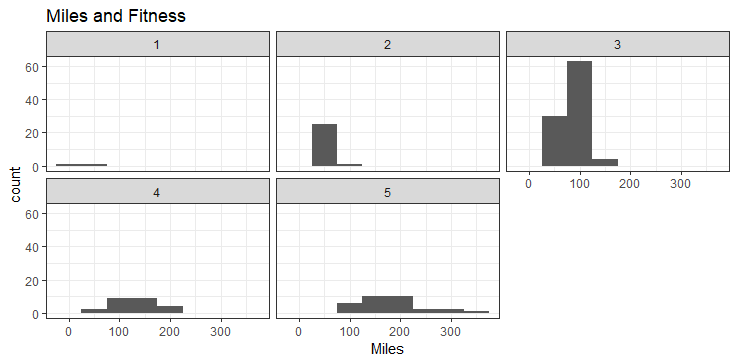
### 3.4.5 Miles and Fitness

ggplot(Cardio\_Goods\_tbl, aes(x=Miles)) +

+ facet\_wrap(~Fitness) +

+ geom\_histogram(binwidth = 50 )+ theme\_bw() +

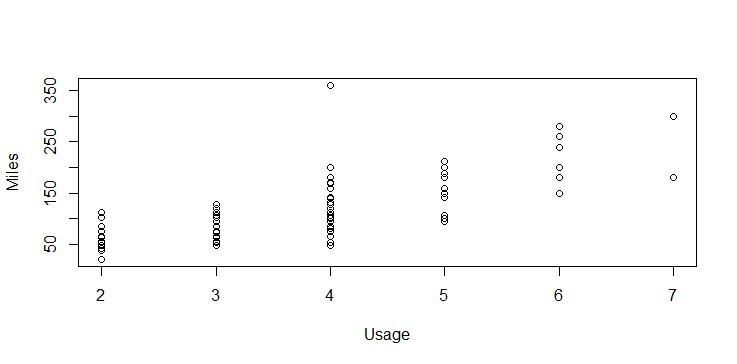
+ labs(y=' count', title = 'Miles and Fitness')



**Inference : Histogram shows that as the miles travelled increase Fitness level is also going up.**

### 3.4.6 Miles and usage

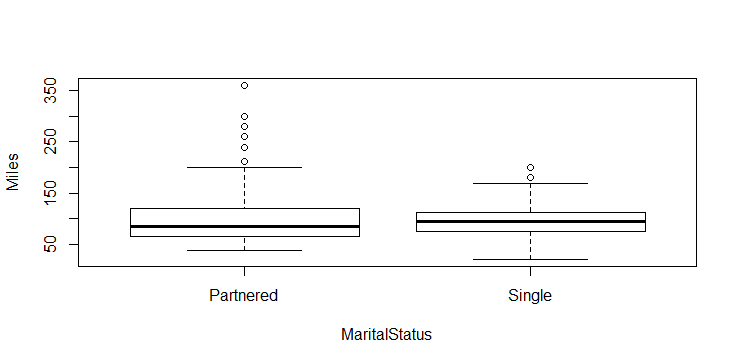
> plot(Miles ~ Usage , Cardio\_Goods\_tbl)



**Inference : as usage increase miles also increase**

### 3.4.7 Miles and Marital status

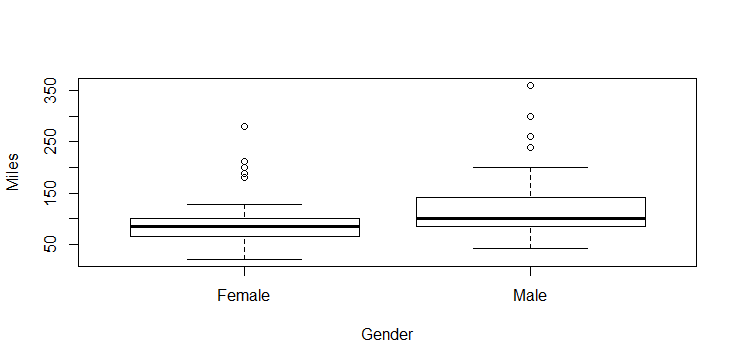
plot(Miles ~ MaritalStatus , Cardio\_Goods\_tbl)



**Inference : Single ‘s have lower miles compare to Partnered .**

### 3.4.8 Miles and Gender

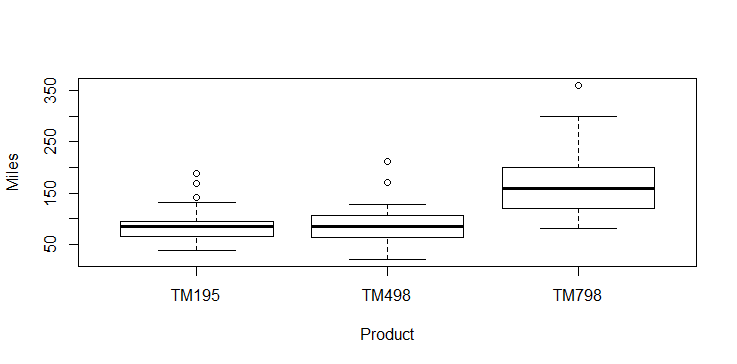
plot(Miles ~ Gender , Cardio\_Goods\_tbl)



**Inference : Males have more miles then female**

### 3.4.9 Product and Miles

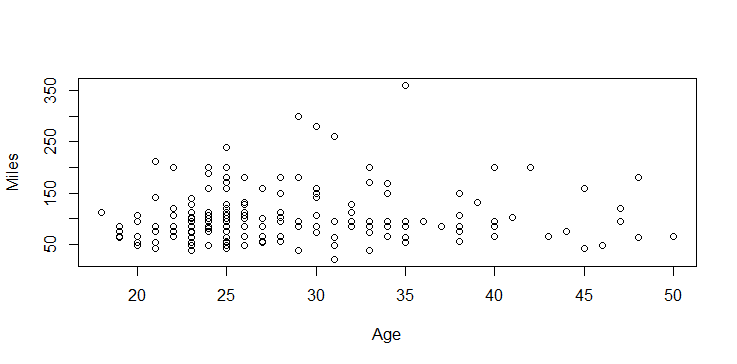
> plot(Miles ~ Product , Cardio\_Goods\_tbl)



**Inference : maximum miles recorded Is for TM798 .**

### 3.4.10 Miles and age

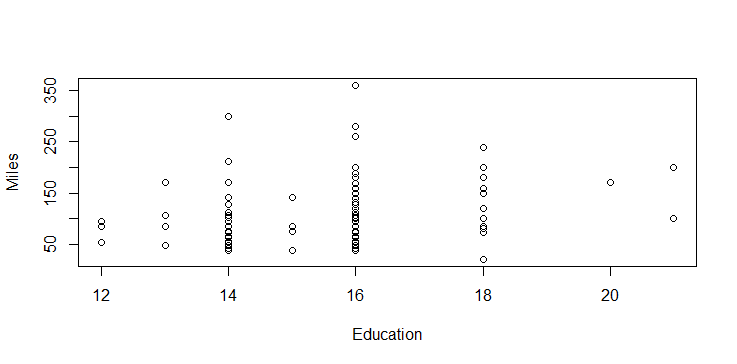
plot(Miles ~ Age , Cardio\_Goods\_tbl)



**Inference : Most of the miles are acquired by users in between the age of 29 and 35**

### 3.4.11 Miles and Education

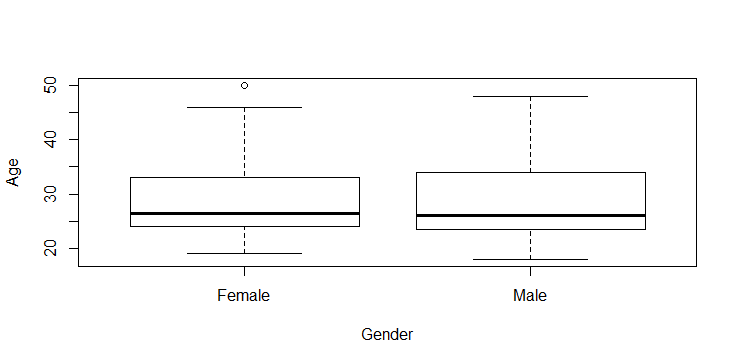
plot(Miles ~ Education , Cardio\_Goods\_tbl)



**Inference : Maximum miles are acquired by users who are at education level 16.**

### 3.4.12 Age and gender

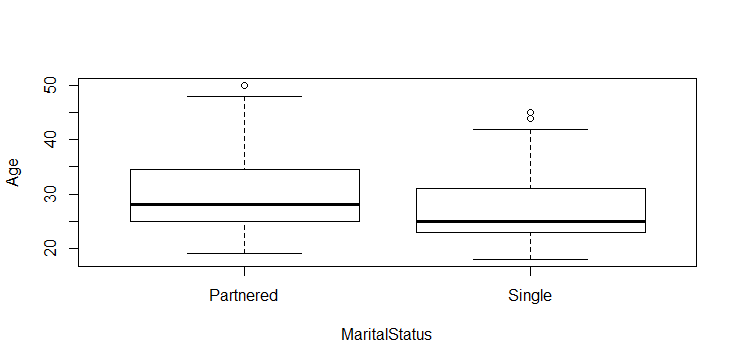
plot(Age ~ Gender , Cardio\_Goods\_tbl)



**Inference : Males and Females are uniformly distributed in age .**

### 3.4.13 Age and MaritalStatus

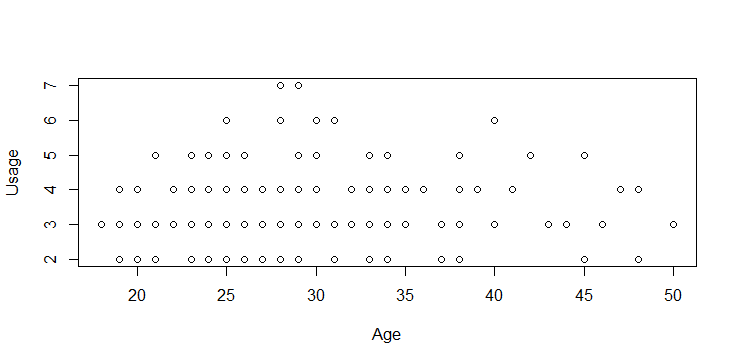
plot(Age ~ MaritalStatus , Cardio\_Goods\_tbl)



**Inference – Mean age of single is lesser then Partnered .**

### 3.4.14 Age and Usage

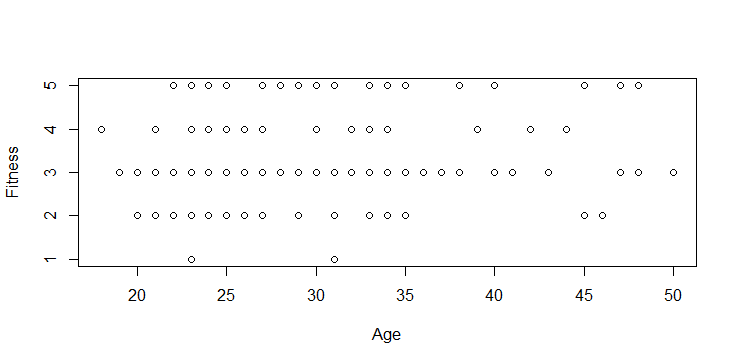
plot(Usage ~ Age , Cardio\_Goods\_tbl)



**Inference : Higher usage is between age group 25 to 40**

### 3.4.15 Age and Fitness

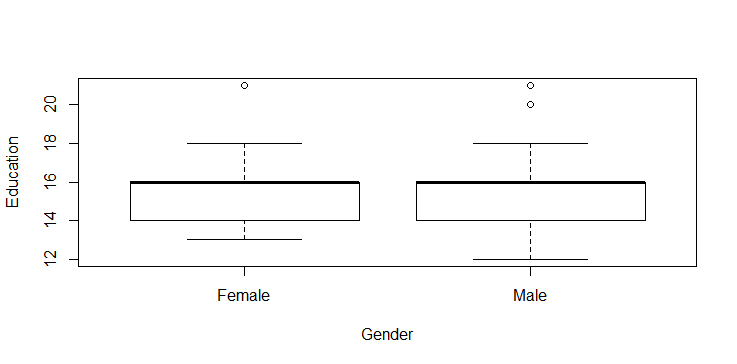
plot(Fitness ~ Age , Cardio\_Goods\_tbl)



**Inference : Fitness level is same across all age group.**

### 3.4.16 Gender and Education

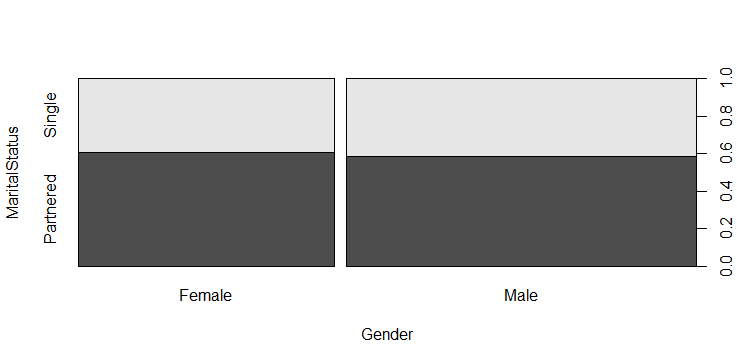
plot(Education ~ Gender , Cardio\_Goods\_tbl)



**Inference : Education level is same across male and female**

### 3.4.17 Gender and Marital Status

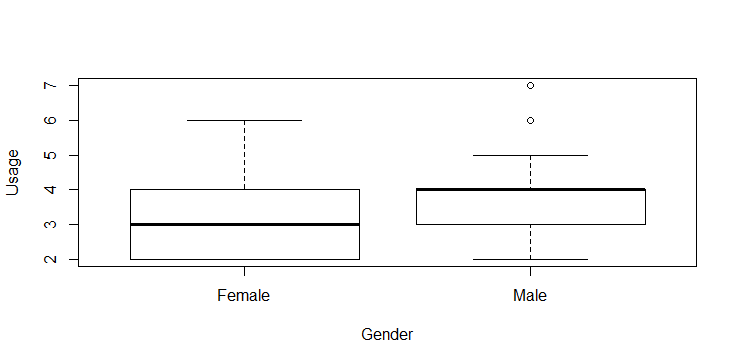
plot(MaritalStatus ~ Gender , Cardio\_Goods\_tbl)



**Inference : Marital status is same across male and female**

### 3.4.18 Gender and Usage

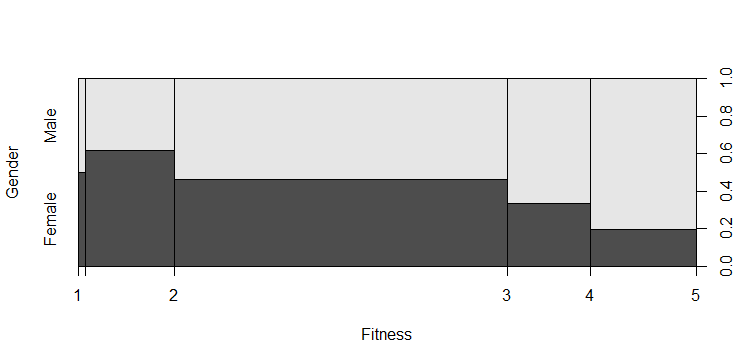
plot(Usage ~ Gender , Cardio\_Goods\_tbl)



**Inference : Male usage is more than Female**

### 3.4.19 Gender and Fitness

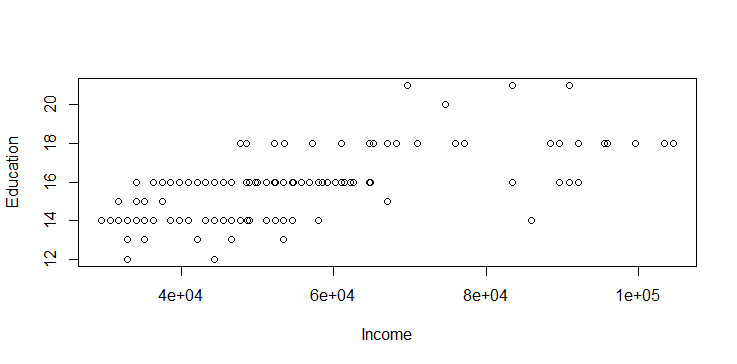
plot(Gender ~ Fitness , Cardio\_Goods\_tbl)



**Inference : Fitness level for male is more then female**

### 3.4.20 Income and Education

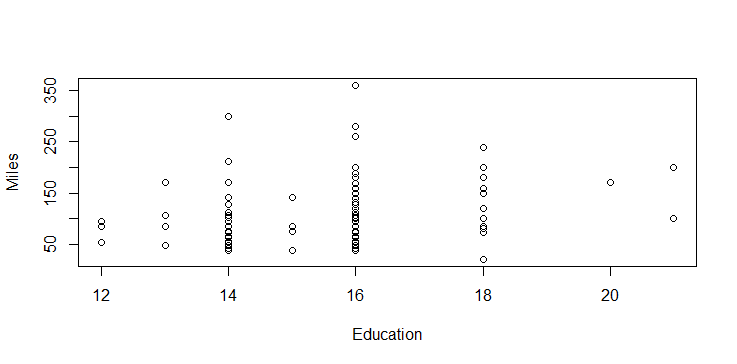
> plot(Education ~ Income , Cardio\_Goods\_tbl)



**Inference: Higher income earner are from education level 16 and above**

### 3.4.21 Education and Miles

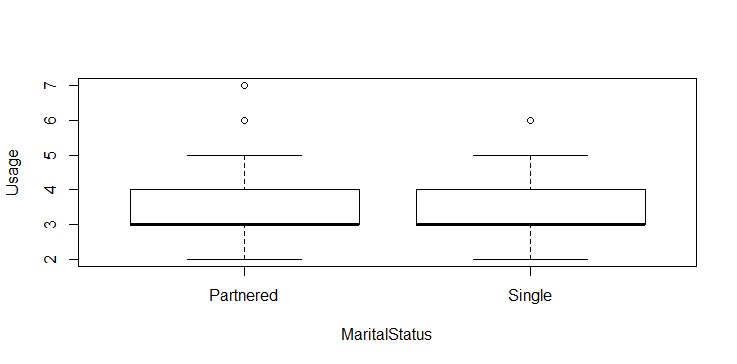
> plot(Education ~ Miles , Cardio\_Goods\_tbl)



**Inference : Highest miles earned are from education level 16.**

### 3.4.22 Marital Status and Usage

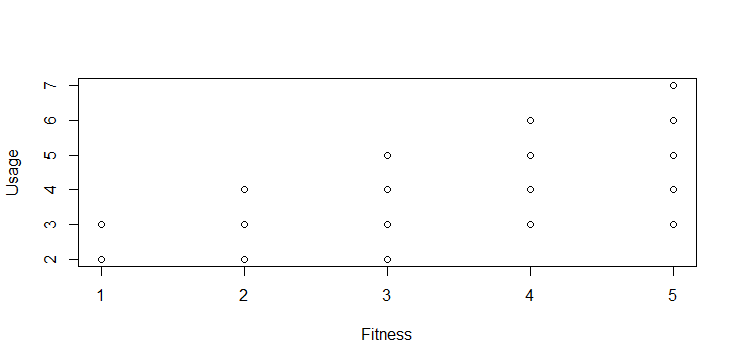
plot(Usage ~ MaritalStatus , Cardio\_Goods\_tbl)



**Inference : Marital status doesn’t effect the usage .**

### 3.4.23 Usage and Fitness

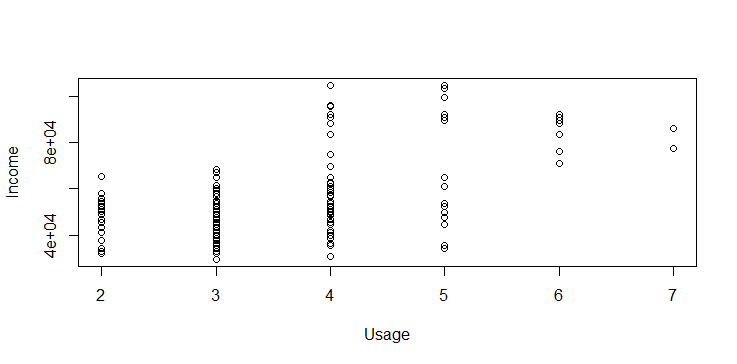
plot(Usage ~ Fitness , Cardio\_Goods\_tbl)



**Inference : Fitness increase with the usage .**

### 3.4.24 Usage and income

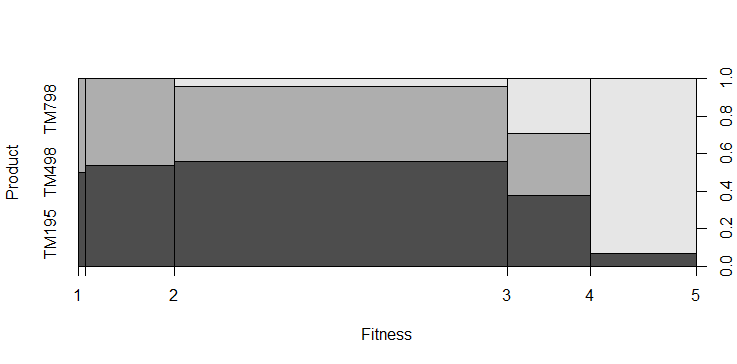
plot(Income ~ Usage , Cardio\_Goods\_tbl)



**Inference : Usage is more for high income users.**

### 3.4.25 Product and Fitness

plot(Product ~ Fitness , Cardio\_Goods\_tbl)



> Cardio\_Goods\_tbl %>%

+ group\_by(Product) %>%

+ summarise(Fitness\_mean = mean(Fitness))

# A tibble: 3 x 2

Product Fitness\_mean

*<fct>* *<dbl>*

1 TM195 2.96

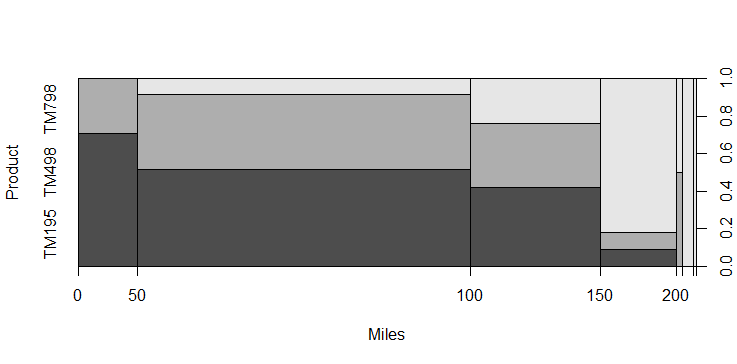
2 TM498 2.9

3 TM798 4.62

Inference : Fitness level is more for product TM798.

### 3.4.26 Product and Miles

plot(Product ~ Miles , Cardio\_Goods\_tbl)



> Cardio\_Goods\_tbl %>%

+ group\_by(Product) %>%

+ summarise(Miles\_mean = mean(Miles))

# A tibble: 3 x 2

Product Miles\_mean

*<fct>* *<dbl>*

1 TM195 82.8

2 TM498 87.9

3 TM798 167.

Inference : Miles mean is maximum for TM798

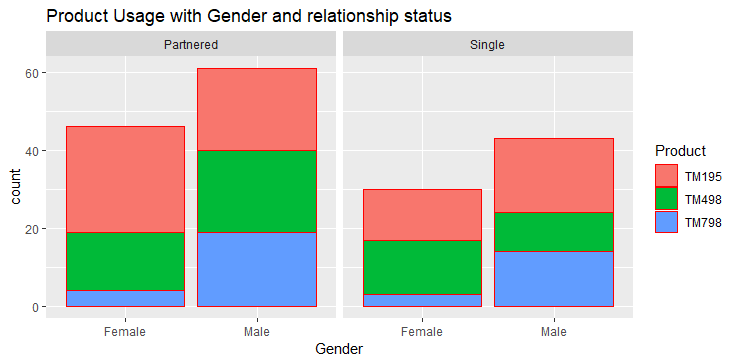
## 3.5 Multivariate analysis :

### 3.5.1 Gender , usage and Product

> ggplot(Cardio\_Goods\_tbl,aes(x=Gender, fill=Product)) +

+ facet\_wrap(~ MaritalStatus)+

+ geom\_bar(colour="red") + labs(y='count',title = 'Product Usage with Gender and relationship status')



**Inference : Users who are in relationship are more using products .**

**Usage for TM798 is very less for female singles .**

**Males who are partnered are using all 3 products in same ratio**

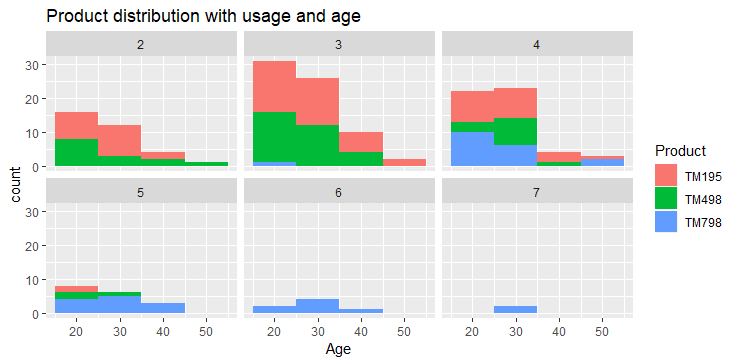
### 3.5.2 Product , age and usage

ggplot(Cardio\_Goods\_tbl, aes(x=Age, fill=Product)) +

+ facet\_wrap(~ Usage ) +

+ geom\_histogram(binwidth = 10 )+

+ labs(y=' count', title = 'Product distribution with usage and age')



**Inference :**

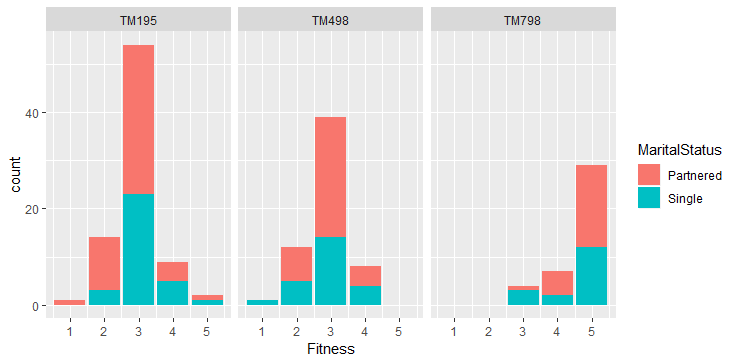
**For user who are using 6 and 7 days in a week , product used is only TM798**

### 3.5.3 Fitness , product and Marital status

ggplot(Cardio\_Goods\_tbl,aes(x=Fitness, fill=MaritalStatus)) +

+ facet\_wrap(~ Product) +

+ geom\_bar()



**Inference: Fitness level of users using TM798 is more then or equal to 3.**

## 3.5 Variable Transformation / Feature Creation : NA

# Conclusion

1. Product TM798 is most popular product in this data set as:

Its usage is maximum

User has maximum fitness level

Miles travelled is also more then double as compared to other products .

1. Mean Fitness level is in direct proportion to number of Usage days per week .

# Appendix A – Source Code

# First R project DEC 18 BACP

# check for vaiable

# ls()

# Reading file

# Cardio\_Goods=read.csv(file.choose())

getwd()

## [1] "C:/Users/Mudgal/Documents/BACP/Module-1IntrotoAnalytics/R Programing/Programming/CardioGoodsFitness"

Cardio\_Goods=read.csv("CardioGoodFitness.csv")  
  
Cardio\_Goods

## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## 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  
## 11 TM195 22 Male 14 Single 3 3 36384 85  
## 12 TM195 22 Female 14 Partnered 3 2 35247 66  
## 13 TM195 22 Female 16 Single 4 3 36384 75  
## 14 TM195 22 Female 14 Single 3 3 35247 75  
## 15 TM195 23 Male 16 Partnered 3 1 38658 47  
## 16 TM195 23 Male 16 Partnered 3 3 40932 75  
## 17 TM195 23 Female 14 Single 2 3 34110 103  
## 18 TM195 23 Male 16 Partnered 4 3 39795 94  
## 19 TM195 23 Female 16 Single 4 3 38658 113  
## 20 TM195 23 Female 15 Partnered 2 2 34110 38  
## 21 TM195 23 Male 14 Single 4 3 38658 113  
## 22 TM195 23 Male 16 Single 4 3 40932 94  
## 23 TM195 24 Female 16 Single 4 3 42069 94  
## 24 TM195 24 Female 16 Partnered 5 5 44343 188  
## 25 TM195 24 Male 14 Single 2 3 45480 113  
## 26 TM195 24 Male 13 Partnered 3 2 42069 47  
## 27 TM195 24 Female 16 Single 4 3 46617 75  
## 28 TM195 25 Female 14 Partnered 3 3 48891 75  
## 29 TM195 25 Male 14 Partnered 2 3 45480 56  
## 30 TM195 25 Female 14 Partnered 2 2 53439 47  
## 31 TM195 25 Female 14 Partnered 3 3 39795 85  
## 32 TM195 25 Male 16 Single 3 4 40932 113  
## 33 TM195 25 Female 16 Partnered 2 2 40932 47  
## 34 TM195 25 Male 16 Single 3 3 43206 85  
## 35 TM195 26 Female 14 Partnered 3 4 44343 113  
## 36 TM195 26 Female 16 Partnered 4 3 52302 113  
## 37 TM195 26 Male 16 Partnered 2 2 53439 47  
## 38 TM195 26 Male 16 Partnered 3 3 51165 85  
## 39 TM195 26 Female 16 Single 3 3 36384 66  
## 40 TM195 26 Male 16 Partnered 4 4 44343 132  
## 41 TM195 26 Male 16 Single 3 3 50028 85  
## 42 TM195 27 Female 14 Partnered 3 2 45480 66  
## 43 TM195 27 Male 16 Single 4 3 54576 85  
## 44 TM195 27 Female 14 Partnered 2 3 45480 56  
## 45 TM195 28 Female 14 Partnered 2 3 46617 56  
## 46 TM195 28 Female 16 Partnered 2 3 52302 66  
## 47 TM195 28 Male 14 Single 3 3 52302 103  
## 48 TM195 28 Female 14 Partnered 3 3 54576 94  
## 49 TM195 28 Male 14 Single 4 3 54576 113  
## 50 TM195 28 Female 16 Partnered 3 3 51165 56  
## 51 TM195 29 Male 18 Partnered 3 3 68220 85  
## 52 TM195 29 Female 14 Partnered 2 2 46617 38  
## 53 TM195 29 Female 16 Partnered 4 3 50028 94  
## 54 TM195 30 Male 14 Partnered 4 4 46617 141  
## 55 TM195 30 Male 14 Single 3 3 54576 85  
## 56 TM195 31 Male 14 Partnered 2 2 54576 47  
## 57 TM195 31 Female 14 Single 2 2 45480 47  
## 58 TM195 32 Female 14 Single 3 4 46617 113  
## 59 TM195 32 Male 14 Partnered 4 3 52302 85  
## 60 TM195 33 Female 16 Single 2 2 55713 38  
## 61 TM195 33 Female 16 Partnered 3 3 46617 85  
## 62 TM195 34 Male 16 Single 4 5 51165 169  
## 63 TM195 34 Female 16 Single 2 2 52302 66  
## 64 TM195 35 Male 16 Partnered 4 3 48891 85  
## 65 TM195 35 Female 16 Partnered 3 3 60261 94  
## 66 TM195 35 Female 18 Single 3 3 67083 85  
## 67 TM195 36 Male 12 Single 4 3 44343 94  
## 68 TM195 37 Female 16 Partnered 3 3 37521 85  
## 69 TM195 38 Male 16 Partnered 3 3 46617 75  
## 70 TM195 38 Female 14 Partnered 2 3 54576 56  
## 71 TM195 38 Male 14 Single 2 3 52302 56  
## 72 TM195 38 Male 16 Partnered 3 3 56850 75  
## 73 TM195 39 Male 16 Partnered 4 4 59124 132  
## 74 TM195 40 Male 16 Partnered 3 3 61398 66  
## 75 TM195 41 Male 16 Partnered 4 3 54576 103  
## 76 TM195 43 Male 16 Partnered 3 3 53439 66  
## 77 TM195 44 Female 16 Single 3 4 57987 75  
## 78 TM195 46 Female 16 Partnered 3 2 60261 47  
## 79 TM195 47 Male 16 Partnered 4 3 56850 94  
## 80 TM195 50 Female 16 Partnered 3 3 64809 66  
## 81 TM498 19 Male 14 Single 3 3 31836 64  
## 82 TM498 20 Male 14 Single 2 3 32973 53  
## 83 TM498 20 Female 14 Partnered 3 3 34110 106  
## 84 TM498 20 Male 14 Single 3 3 38658 95  
## 85 TM498 21 Female 14 Partnered 5 4 34110 212  
## 86 TM498 21 Male 16 Partnered 2 2 34110 42  
## 87 TM498 21 Male 12 Partnered 2 2 32973 53  
## 88 TM498 23 Male 14 Partnered 3 3 36384 95  
## 89 TM498 23 Male 14 Partnered 3 3 38658 85  
## 90 TM498 23 Female 16 Single 3 3 45480 95  
## 91 TM498 23 Male 16 Partnered 4 3 45480 127  
## 92 TM498 23 Female 16 Partnered 3 2 43206 74  
## 93 TM498 23 Female 14 Single 3 2 40932 53  
## 94 TM498 23 Male 16 Partnered 3 3 45480 64  
## 95 TM498 24 Female 14 Single 3 2 40932 85  
## 96 TM498 24 Male 14 Single 3 4 48891 106  
## 97 TM498 24 Female 16 Single 3 3 50028 106  
## 98 TM498 25 Female 14 Partnered 2 3 45480 85  
## 99 TM498 25 Female 14 Single 3 4 43206 127  
## 100 TM498 25 Male 16 Partnered 2 2 52302 42  
## 101 TM498 25 Female 14 Partnered 5 3 47754 106  
## 102 TM498 25 Male 14 Single 3 3 45480 95  
## 103 TM498 25 Female 14 Single 2 3 43206 64  
## 104 TM498 25 Male 14 Partnered 4 3 45480 170  
## 105 TM498 25 Male 14 Partnered 3 4 43206 106  
## 106 TM498 25 Male 16 Partnered 2 3 50028 53  
## 107 TM498 25 Female 14 Single 2 2 45480 42  
## 108 TM498 25 Male 14 Single 4 3 48891 127  
## 109 TM498 26 Female 16 Partnered 4 3 45480 85  
## 110 TM498 26 Female 16 Single 4 4 50028 127  
## 111 TM498 26 Male 16 Single 4 3 51165 106  
## 112 TM498 27 Male 14 Single 4 2 45480 53  
## 113 TM498 29 Female 14 Partnered 3 3 51165 95  
## 114 TM498 30 Female 14 Single 3 3 57987 74  
## 115 TM498 30 Female 13 Single 4 3 46617 106  
## 116 TM498 31 Male 16 Partnered 3 3 52302 95  
## 117 TM498 31 Female 16 Partnered 2 3 51165 64  
## 118 TM498 31 Female 18 Single 2 1 65220 21  
## 119 TM498 32 Male 16 Single 4 3 60261 127  
## 120 TM498 32 Male 16 Partnered 3 3 53439 95  
## 121 TM498 33 Male 13 Partnered 4 4 53439 170  
## 122 TM498 33 Female 16 Partnered 2 3 50028 85  
## 123 TM498 33 Male 16 Partnered 3 3 51165 95  
## 124 TM498 33 Female 16 Partnered 5 3 53439 95  
## 125 TM498 33 Female 18 Single 3 4 47754 74  
## 126 TM498 34 Female 16 Partnered 4 3 64809 95  
## 127 TM498 34 Male 16 Partnered 3 4 59124 85  
## 128 TM498 34 Male 15 Single 3 3 67083 85  
## 129 TM498 35 Female 14 Partnered 3 2 52302 53  
## 130 TM498 35 Male 16 Partnered 3 2 53439 53  
## 131 TM498 35 Female 16 Single 3 2 50028 64  
## 132 TM498 35 Male 16 Partnered 3 3 53439 95  
## 133 TM498 37 Female 16 Partnered 2 3 48891 85  
## 134 TM498 38 Female 16 Partnered 4 3 62535 85  
## 135 TM498 38 Male 16 Partnered 3 3 59124 106  
## 136 TM498 40 Female 16 Partnered 3 3 61398 85  
## 137 TM498 40 Female 16 Single 3 3 57987 85  
## 138 TM498 40 Male 16 Partnered 3 3 64809 95  
## 139 TM498 45 Male 16 Partnered 2 2 54576 42  
## 140 TM498 48 Male 16 Partnered 2 3 57987 64  
## 141 TM798 22 Male 14 Single 4 3 48658 106  
## 142 TM798 22 Male 16 Single 3 5 54781 120  
## 143 TM798 22 Male 18 Single 4 5 48556 200  
## 144 TM798 23 Male 16 Single 4 5 58516 140  
## 145 TM798 23 Female 18 Single 5 4 53536 100  
## 146 TM798 23 Male 16 Single 4 5 48556 100  
## 147 TM798 24 Male 16 Single 4 5 61006 100  
## 148 TM798 24 Male 18 Partnered 4 5 57271 80  
## 149 TM798 24 Female 16 Single 5 5 52291 200  
## 150 TM798 24 Male 16 Single 5 5 49801 160  
## 151 TM798 25 Male 16 Partnered 4 5 49801 120  
## 152 TM798 25 Male 16 Partnered 4 4 62251 160  
## 153 TM798 25 Female 18 Partnered 5 5 61006 200  
## 154 TM798 25 Male 18 Partnered 4 3 64741 100  
## 155 TM798 25 Male 18 Partnered 6 4 70966 180  
## 156 TM798 25 Male 18 Partnered 6 5 75946 240  
## 157 TM798 25 Male 20 Partnered 4 5 74701 170  
## 158 TM798 26 Female 21 Single 4 3 69721 100  
## 159 TM798 26 Male 16 Partnered 5 4 64741 180  
## 160 TM798 27 Male 16 Partnered 4 5 83416 160  
## 161 TM798 27 Male 18 Single 4 3 88396 100  
## 162 TM798 27 Male 21 Partnered 4 4 90886 100  
## 163 TM798 28 Female 18 Partnered 6 5 92131 180  
## 164 TM798 28 Male 18 Partnered 7 5 77191 180  
## 165 TM798 28 Male 18 Single 6 5 88396 150  
## 166 TM798 29 Male 18 Single 5 5 52290 180  
## 167 TM798 29 Male 14 Partnered 7 5 85906 300  
## 168 TM798 30 Female 16 Partnered 6 5 90886 280  
## 169 TM798 30 Male 18 Partnered 5 4 103336 160  
## 170 TM798 30 Male 18 Partnered 5 5 99601 150  
## 171 TM798 31 Male 16 Partnered 6 5 89641 260  
## 172 TM798 33 Female 18 Partnered 4 5 95866 200  
## 173 TM798 34 Male 16 Single 5 5 92131 150  
## 174 TM798 35 Male 16 Partnered 4 5 92131 360  
## 175 TM798 38 Male 18 Partnered 5 5 104581 150  
## 176 TM798 40 Male 21 Single 6 5 83416 200  
## 177 TM798 42 Male 18 Single 5 4 89641 200  
## 178 TM798 45 Male 16 Single 5 5 90886 160  
## 179 TM798 47 Male 18 Partnered 4 5 104581 120  
## 180 TM798 48 Male 18 Partnered 4 5 95508 180

#inoking Library package for data wrangling , data plotting and knit  
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(knitr)  
library(ggplot2)

## high level overview

# rows and col

dim(Cardio\_Goods)

## [1] 180 9

#structure  
str(Cardio\_Goods)

## '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 : int 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 : int 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 : int 3 2 4 3 4 3 3 3 5 2 ...  
## $ Fitness : int 4 3 3 3 2 3 3 3 4 3 ...  
## $ Income : int 29562 31836 30699 32973 35247 32973 35247 32973 35247 37521 ...  
## $ Miles : int 112 75 66 85 47 66 75 85 141 85 ...

#summary  
summary(Cardio\_Goods)

## Product Age Gender Education MaritalStatus  
## TM195:80 Min. :18.00 Female: 76 Min. :12.00 Partnered:107   
## TM498:60 1st Qu.:24.00 Male :104 1st Qu.:14.00 Single : 73   
## TM798:40 Median :26.00 Median :16.00   
## Mean :28.79 Mean :15.57   
## 3rd Qu.:33.00 3rd Qu.:16.00   
## Max. :50.00 Max. :21.00   
## Usage Fitness Income Miles   
## Min. :2.000 Min. :1.000 Min. : 29562 Min. : 21.0   
## 1st Qu.:3.000 1st Qu.:3.000 1st Qu.: 44059 1st Qu.: 66.0   
## Median :3.000 Median :3.000 Median : 50597 Median : 94.0   
## Mean :3.456 Mean :3.311 Mean : 53720 Mean :103.2   
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.: 58668 3rd Qu.:114.8   
## Max. :7.000 Max. :5.000 Max. :104581 Max. :360.0

class(Cardio\_Goods)

## [1] "data.frame"

head(Cardio\_Goods)

## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## 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

tail(Cardio\_Goods)

## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## 175 TM798 38 Male 18 Partnered 5 5 104581 150  
## 176 TM798 40 Male 21 Single 6 5 83416 200  
## 177 TM798 42 Male 18 Single 5 4 89641 200  
## 178 TM798 45 Male 16 Single 5 5 90886 160  
## 179 TM798 47 Male 18 Partnered 4 5 104581 120  
## 180 TM798 48 Male 18 Partnered 4 5 95508 180

# Univariate analysis

# summary(Cardio\_Goods,Product)

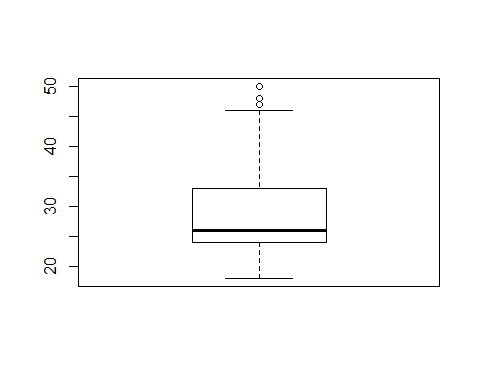
summary(Cardio\_Goods$Product)

## TM195 TM498 TM798   
## 80 60 40

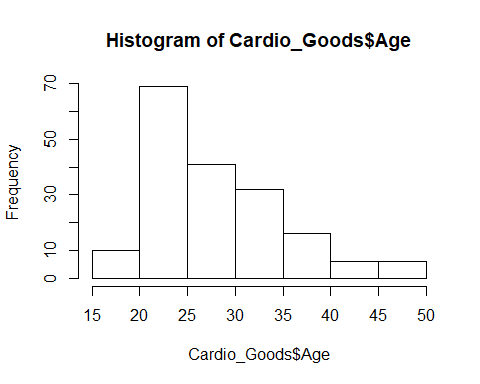
summary(Cardio\_Goods$Age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 18.00 24.00 26.00 28.79 33.00 50.00

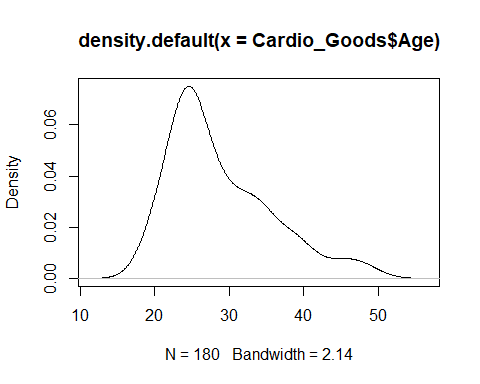
boxplot(Cardio\_Goods$Age)



hist(Cardio\_Goods$Age)



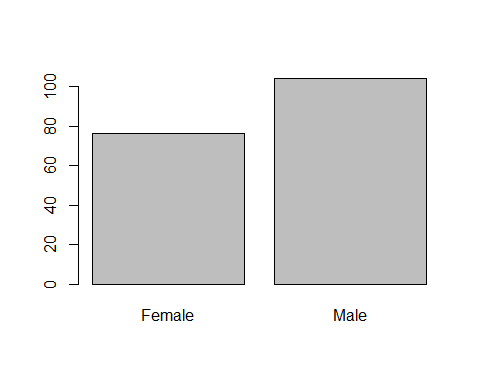
plot(density(Cardio\_Goods$Age)) # not sure how its working



summary(Cardio\_Goods$Gender)

## Female Male   
## 76 104

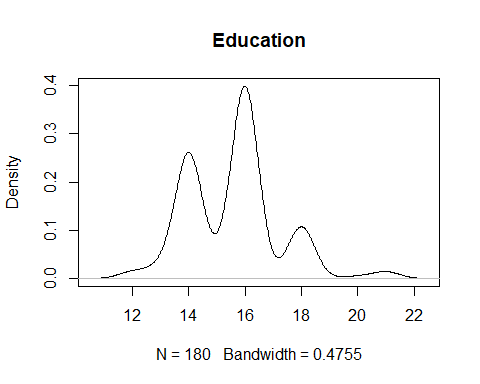
plot(Cardio\_Goods$Gender)



summary(Cardio\_Goods$Education)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.00 14.00 16.00 15.57 16.00 21.00

plot(density(Cardio\_Goods$Education),main='Education')



summary(Cardio\_Goods$MaritalStatus)

## Partnered Single   
## 107 73

summary(Cardio\_Goods$Usage)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 3.000 3.456 4.000 7.000

summary(Cardio\_Goods$Fitness)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 3.000 3.000 3.311 4.000 5.000

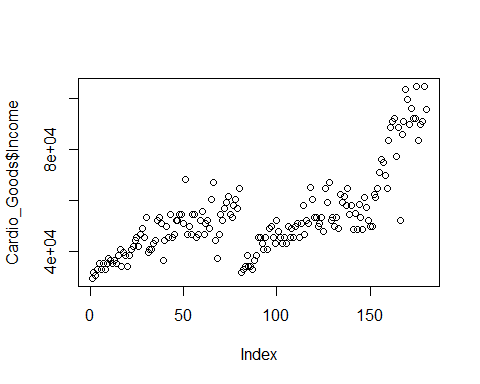
summary(Cardio\_Goods$Income)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 29562 44059 50597 53720 58668 104581

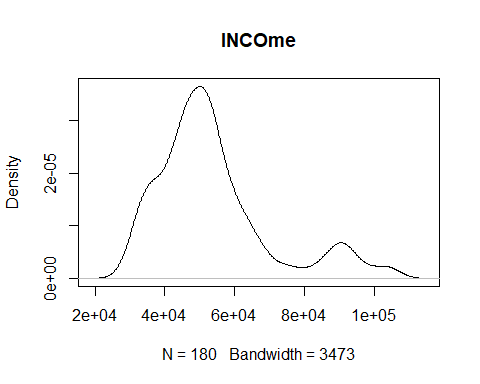
summary(Cardio\_Goods$Miles)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 21.0 66.0 94.0 103.2 114.8 360.0

plot(Cardio\_Goods$Income)



plot(density(Cardio\_Goods$Income),main='INCOme')



## using tblyr librabry

# convert to local data frame

Cardio\_Goods\_tbl = tbl\_df(Cardio\_Goods)  
Cardio\_Goods\_tbl

## # A tibble: 180 x 9  
## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## <fct> <int> <fct> <int> <fct> <int> <int> <int> <int>  
## 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  
## # ... with 170 more rows

#Bivariate analysis   
#filter on some condition   
filter(Cardio\_Goods\_tbl,Product=='TM195' & Gender=='Male')

## # A tibble: 40 x 9  
## Product Age Gender Education MaritalStatus Usage Fitness Income Miles  
## <fct> <int> <fct> <int> <fct> <int> <int> <int> <int>  
## 1 TM195 18 Male 14 Single 3 4 29562 112  
## 2 TM195 19 Male 15 Single 2 3 31836 75  
## 3 TM195 19 Male 12 Single 3 3 32973 85  
## 4 TM195 20 Male 13 Partnered 4 2 35247 47  
## 5 TM195 21 Male 13 Single 3 3 32973 85  
## 6 TM195 21 Male 15 Single 5 4 35247 141  
## 7 TM195 22 Male 14 Single 3 3 36384 85  
## 8 TM195 23 Male 16 Partnered 3 1 38658 47  
## 9 TM195 23 Male 16 Partnered 3 3 40932 75  
## 10 TM195 23 Male 16 Partnered 4 3 39795 94  
## # ... with 30 more rows

#select few columns   
select(Cardio\_Goods\_tbl, Age,Gender,Usage)

## # A tibble: 180 x 3  
## Age Gender Usage  
## <int> <fct> <int>  
## 1 18 Male 3  
## 2 19 Male 2  
## 3 19 Female 4  
## 4 19 Male 3  
## 5 20 Male 4  
## 6 20 Female 3  
## 7 21 Female 3  
## 8 21 Male 3  
## 9 21 Male 5  
## 10 21 Female 2  
## # ... with 170 more rows

#chaining   
Cardio\_Goods\_tbl %>%   
 select(Age,Gender) %>%   
 filter(Age > 40)

## # A tibble: 12 x 2  
## Age Gender  
## <int> <fct>   
## 1 41 Male   
## 2 43 Male   
## 3 44 Female  
## 4 46 Female  
## 5 47 Male   
## 6 50 Female  
## 7 45 Male   
## 8 48 Male   
## 9 42 Male   
## 10 45 Male   
## 11 47 Male   
## 12 48 Male

# Mean age for each product user   
Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Age\_mean = mean(Age))

## # A tibble: 3 x 2  
## Product Age\_mean  
## <fct> <dbl>  
## 1 TM195 28.6  
## 2 TM498 28.9  
## 3 TM798 29.1

# Mean Usage for each product -  
Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Usage\_mean = mean(Usage))

## # A tibble: 3 x 2  
## Product Usage\_mean  
## <fct> <dbl>  
## 1 TM195 3.09  
## 2 TM498 3.07  
## 3 TM798 4.78

# Mean Fitness level for each product   
Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Fitness\_mean = mean(Fitness))

## # A tibble: 3 x 2  
## Product Fitness\_mean  
## <fct> <dbl>  
## 1 TM195 2.96  
## 2 TM498 2.9   
## 3 TM798 4.62

# Mean INcome level for each product user  
Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Income\_mean = mean(Income))

## # A tibble: 3 x 2  
## Product Income\_mean  
## <fct> <dbl>  
## 1 TM195 46418.  
## 2 TM498 48974.  
## 3 TM798 75442.

# Mean Education level for each product user  
Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Edu\_mean = mean(Education))

## # A tibble: 3 x 2  
## Product Edu\_mean  
## <fct> <dbl>  
## 1 TM195 15.0  
## 2 TM498 15.1  
## 3 TM798 17.3

Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Miles\_mean = mean(Miles))

## # A tibble: 3 x 2  
## Product Miles\_mean  
## <fct> <dbl>  
## 1 TM195 82.8  
## 2 TM498 87.9  
## 3 TM798 167.

# Mean age level for each gender   
Cardio\_Goods\_tbl %>%   
 group\_by(Gender) %>%   
 summarise(Fitness\_mean = mean(Age))

## # A tibble: 2 x 2  
## Gender Fitness\_mean  
## <fct> <dbl>  
## 1 Female 28.6  
## 2 Male 29.0

# Mean usage level for each gender  
Cardio\_Goods\_tbl %>%   
 group\_by(Gender) %>%   
 summarise(Usage\_mean = mean(Usage))

## # A tibble: 2 x 2  
## Gender Usage\_mean  
## <fct> <dbl>  
## 1 Female 3.18  
## 2 Male 3.65

# Mean Income level for each gender  
Cardio\_Goods\_tbl %>%   
 group\_by(Gender) %>%   
 summarise(Inc\_mean = mean(Income))

## # A tibble: 2 x 2  
## Gender Inc\_mean  
## <fct> <dbl>  
## 1 Female 49829.  
## 2 Male 56563.

# Mean fitness level for each gender  
Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise(Fit\_mean = mean(Usage))

## # A tibble: 3 x 2  
## Product Fit\_mean  
## <fct> <dbl>  
## 1 TM195 3.09  
## 2 TM498 3.07  
## 3 TM798 4.78

Cardio\_Goods\_tbl %>%   
 group\_by(Gender) %>%   
 summarise(Fit\_mean = mean(Fitness))

## # A tibble: 2 x 2  
## Gender Fit\_mean  
## <fct> <dbl>  
## 1 Female 3.03  
## 2 Male 3.52

# Mean Miles level for each gender  
Cardio\_Goods\_tbl %>%   
 group\_by(Gender) %>%   
 summarise(Miles\_mean = mean(Miles))

## # A tibble: 2 x 2  
## Gender Miles\_mean  
## <fct> <dbl>  
## 1 Female 90.0  
## 2 Male 113.

#Fitness mean per marital status   
Cardio\_Goods\_tbl %>%   
 group\_by(MaritalStatus) %>%   
 summarise(Miles\_mean = mean(Fitness))

## # A tibble: 2 x 2  
## MaritalStatus Miles\_mean  
## <fct> <dbl>  
## 1 Partnered 3.27  
## 2 Single 3.37

# INcome mean per marital status

Cardio\_Goods\_tbl %>%   
 group\_by(MaritalStatus) %>%   
 summarise(Miles\_mean = mean(Income))

## # A tibble: 2 x 2  
## MaritalStatus Miles\_mean  
## <fct> <dbl>  
## 1 Partnered 55763   
## 2 Single 50724.

#Uage mean per marital status   
Cardio\_Goods\_tbl %>%   
 group\_by(MaritalStatus) %>%   
 summarise(Miles\_mean = mean(Usage))

## # A tibble: 2 x 2  
## MaritalStatus Miles\_mean  
## <fct> <dbl>  
## 1 Partnered 3.45  
## 2 Single 3.47

#Miles mean per marital status   
Cardio\_Goods\_tbl %>%   
 group\_by(MaritalStatus) %>%   
 summarise(Miles\_mean = mean(Miles))

## # A tibble: 2 x 2  
## MaritalStatus Miles\_mean  
## <fct> <dbl>  
## 1 Partnered 104.  
## 2 Single 102.

# Edu mean per marital status

# Cardio\_Goods\_tbl %>%

# group\_by(MaritalStatus) %>% #summarise\_(Miles\_mean = mean(Education))

#summarise each  
Cardio\_Goods\_tbl %>%   
 group\_by(MaritalStatus) %>%   
 summarise\_each(funs(mean), Income,Usage,Fitness,Education,Age , Miles )

## `summarise\_each()` is deprecated.  
## Use `summarise\_all()`, `summarise\_at()` or `summarise\_if()` instead.  
## To map `funs` over a selection of variables, use `summarise\_at()`

## # A tibble: 2 x 7  
## MaritalStatus Income Usage Fitness Education Age Miles  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Partnered 55763 3.45 3.27 15.7 29.9 104.  
## 2 Single 50724. 3.47 3.37 15.4 27.2 102.

Cardio\_Goods\_tbl %>%   
 group\_by(Product) %>%   
 summarise\_each(funs(mean), Income,Usage,Fitness,Education,Age, Miles )

## `summarise\_each()` is deprecated.  
## Use `summarise\_all()`, `summarise\_at()` or `summarise\_if()` instead.  
## To map `funs` over a selection of variables, use `summarise\_at()`

## # A tibble: 3 x 7  
## Product Income Usage Fitness Education Age Miles  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 TM195 46418. 3.09 2.96 15.0 28.6 82.8  
## 2 TM498 48974. 3.07 2.9 15.1 28.9 87.9  
## 3 TM798 75442. 4.78 4.62 17.3 29.1 167.

Cardio\_Goods\_tbl %>%   
 group\_by(Usage) %>%   
 summarise\_each(funs(mean), Income,Fitness,Education,Age )

## `summarise\_each()` is deprecated.  
## Use `summarise\_all()`, `summarise\_at()` or `summarise\_if()` instead.  
## To map `funs` over a selection of variables, use `summarise\_at()`

## # A tibble: 6 x 5  
## Usage Income Fitness Education Age  
## <int> <dbl> <dbl> <dbl> <dbl>  
## 1 2 46871. 2.52 14.9 28.5  
## 2 3 47922. 3 15.1 29.0  
## 3 4 56403. 3.60 15.9 28.5  
## 4 5 66396. 4.41 16.5 29.1  
## 5 6 84483. 4.86 17.9 29.6  
## 6 7 81548. 5 16 28.5

Cardio\_Goods\_tbl %>%   
 group\_by(Gender) %>%   
 summarise\_each(funs(mean), Income,Fitness,Education,Age,Usage )

## `summarise\_each()` is deprecated.  
## Use `summarise\_all()`, `summarise\_at()` or `summarise\_if()` instead.  
## To map `funs` over a selection of variables, use `summarise\_at()`

## # A tibble: 2 x 6  
## Gender Income Fitness Education Age Usage  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Female 49829. 3.03 15.4 28.6 3.18  
## 2 Male 56563. 3.52 15.7 29.0 3.65

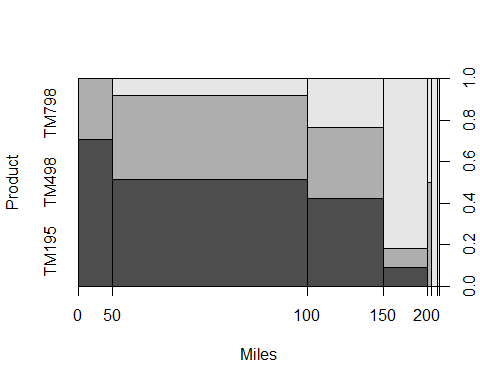
#group by more then 1 column   
Cardio\_Goods\_tbl %>%   
 group\_by(Product,Gender,MaritalStatus) %>%   
 summarise\_each(funs(mean), Income,Usage,Fitness,Education,Age, Miles)

## `summarise\_each()` is deprecated.  
## Use `summarise\_all()`, `summarise\_at()` or `summarise\_if()` instead.  
## To map `funs` over a selection of variables, use `summarise\_at()`

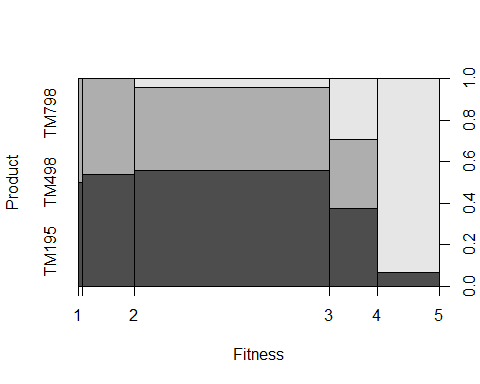
## # A tibble: 12 x 9  
## # Groups: Product, Gender [?]  
## Product Gender MaritalStatus Income Usage Fitness Education Age Miles  
## <fct> <fct> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 TM195 Female Partnered 46154. 2.85 2.85 14.9 28.3 74.9  
## 2 TM195 Female Single 45742. 3 2.92 15.5 28.7 78.8  
## 3 TM195 Male Partnered 50028 3.29 2.86 15.4 31.4 80.2  
## 4 TM195 Male Single 43266. 3.26 3.26 14.5 25.6 99.5  
## 5 TM498 Female Partnered 49725. 3.33 2.93 15.2 30 94   
## 6 TM498 Female Single 48920. 2.93 2.79 15.2 28.1 80.2  
## 7 TM498 Male Partnered 49378. 2.86 2.90 15.3 30.4 87.2  
## 8 TM498 Male Single 47072. 3.3 3 14.5 25.2 91.1  
## 9 TM798 Female Partnered 84972. 5.25 5 17.5 29 215   
## 10 TM798 Female Single 58516 4.67 4 18.3 24.3 133.   
## 11 TM798 Male Partnered 81431. 4.84 4.63 17.4 30 176.   
## 12 TM798 Male Single 68216. 4.57 4.64 16.9 28.9 148.

# Bivariate - between cont. feature

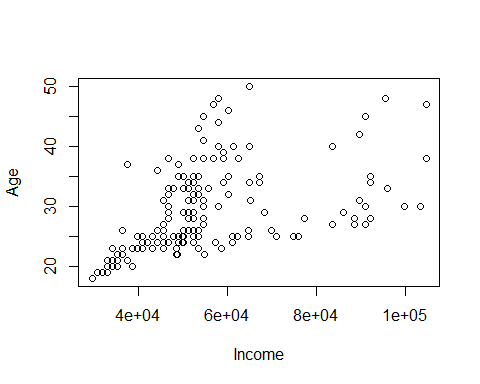
plot(Product ~ Miles , Cardio\_Goods\_tbl)



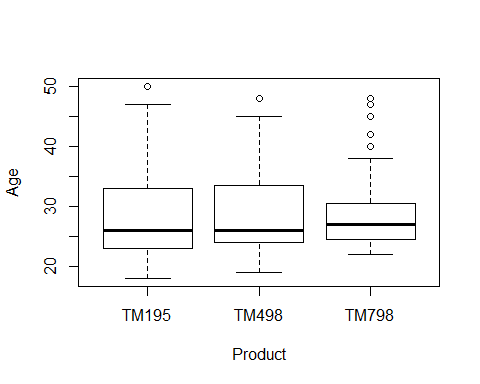
plot(Product ~ Fitness , Cardio\_Goods\_tbl)



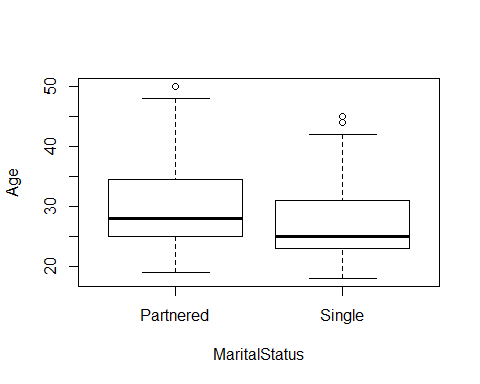
plot(Age ~ Income , Cardio\_Goods\_tbl)



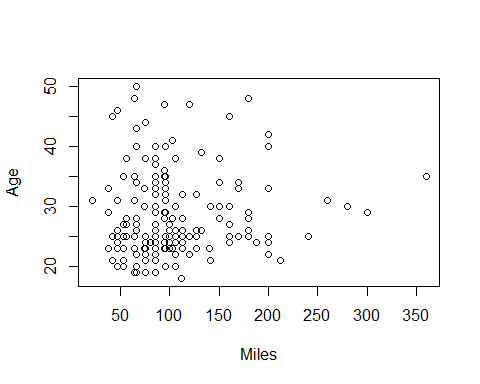
plot(Age ~ Product , Cardio\_Goods\_tbl)



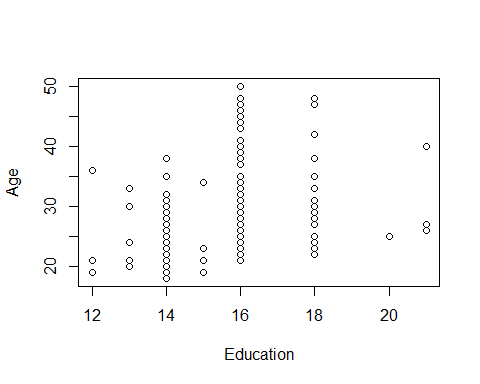
plot(Age ~ MaritalStatus , Cardio\_Goods\_tbl)



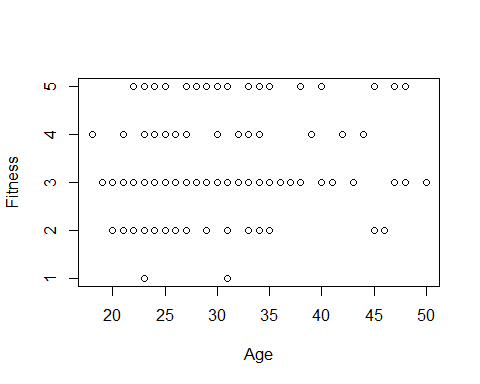
plot(Age ~ Miles , Cardio\_Goods\_tbl)



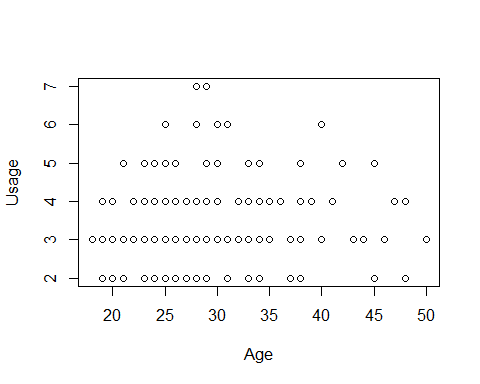
plot(Age ~ Education , Cardio\_Goods\_tbl)



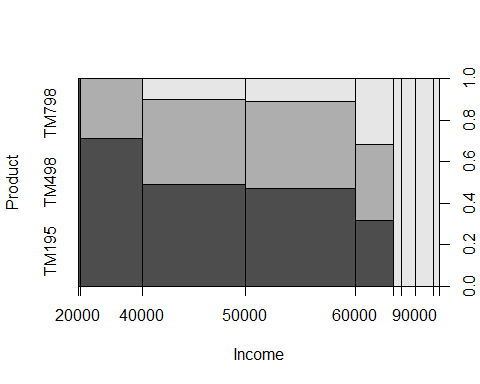
plot(Fitness ~ Age , Cardio\_Goods\_tbl)



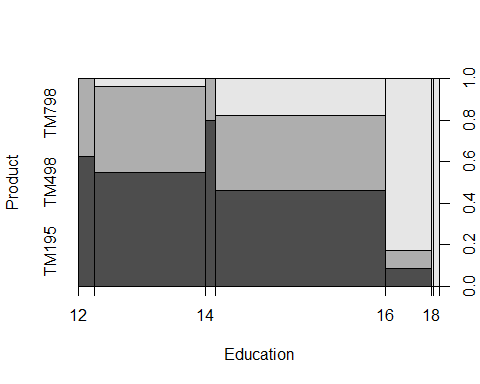
plot(Usage ~ Age , Cardio\_Goods\_tbl)



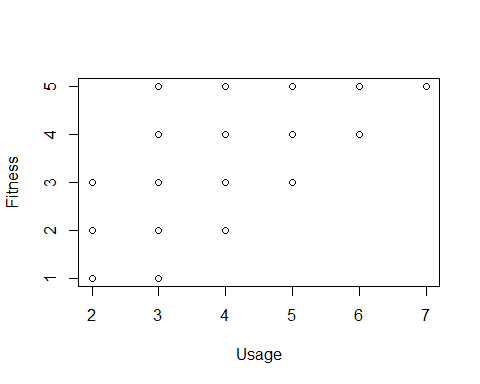
plot(Product ~ Income , Cardio\_Goods\_tbl)



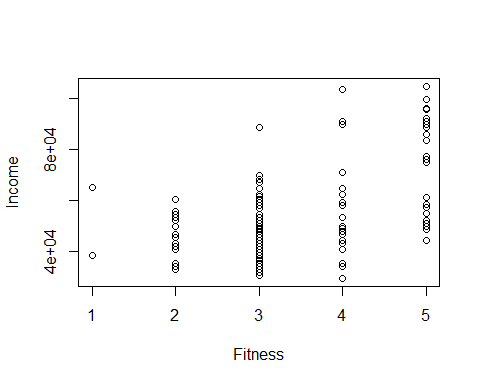
plot(Product ~ Education , Cardio\_Goods\_tbl)



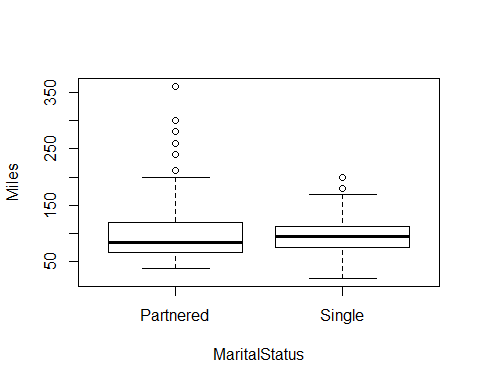
plot(Fitness ~ Usage , Cardio\_Goods\_tbl)



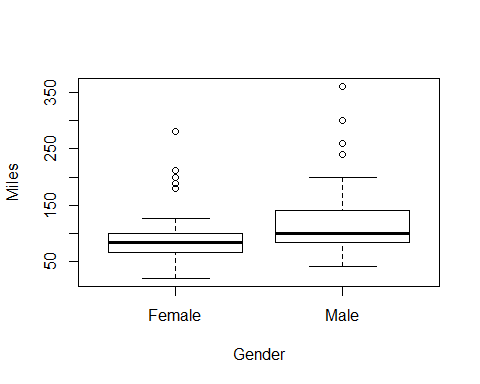
plot(Income ~ Fitness , Cardio\_Goods\_tbl)



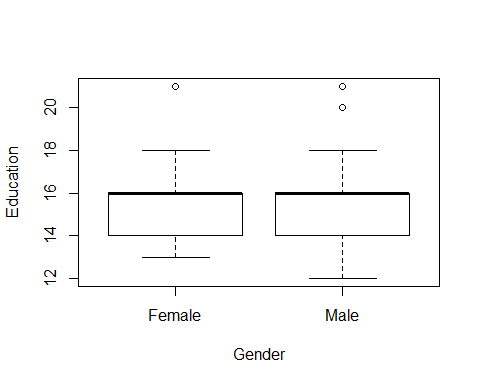
plot(Miles ~ MaritalStatus , Cardio\_Goods\_tbl)



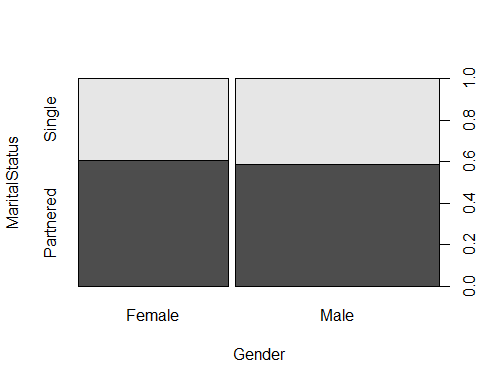
plot(Miles ~ Gender , Cardio\_Goods\_tbl)



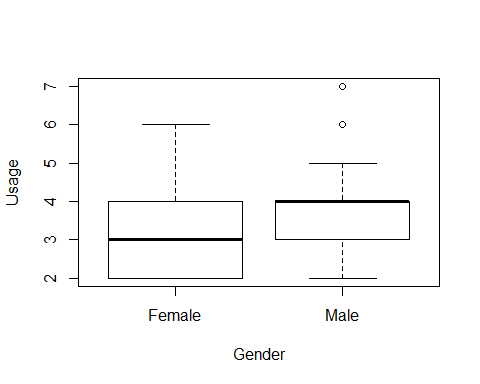
plot(Education ~ Gender , Cardio\_Goods\_tbl)



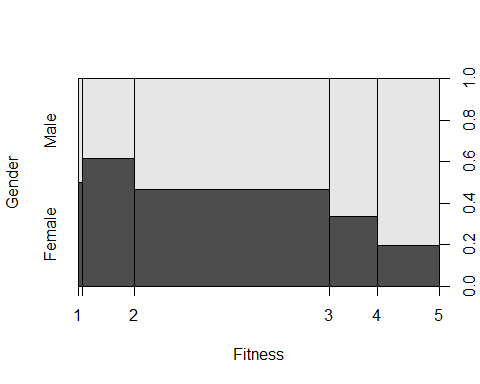
plot(MaritalStatus ~ Gender , Cardio\_Goods\_tbl)



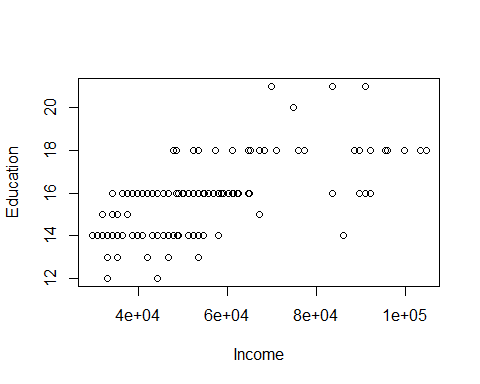
plot(Usage ~ Gender , Cardio\_Goods\_tbl)



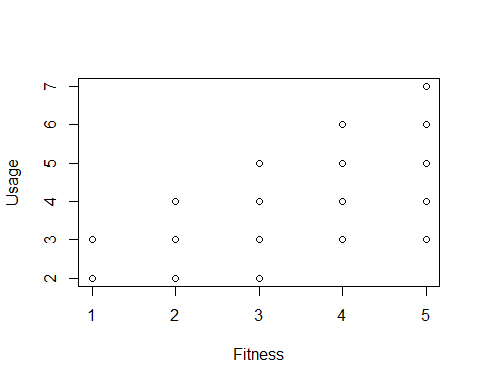
plot(Gender ~ Fitness , Cardio\_Goods\_tbl)



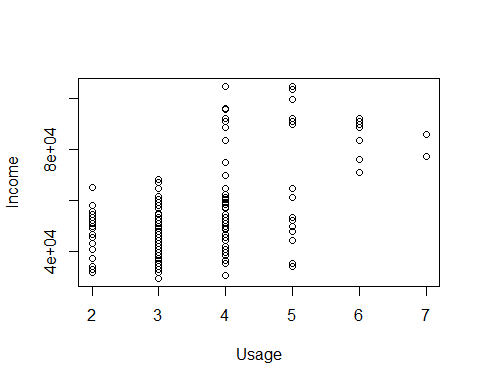
plot(Education ~ Income , Cardio\_Goods\_tbl)



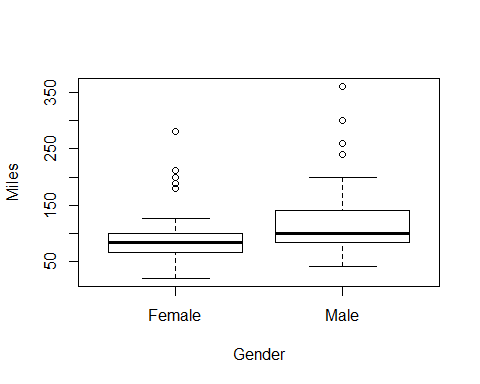
plot(Usage ~ Fitness , Cardio\_Goods\_tbl)



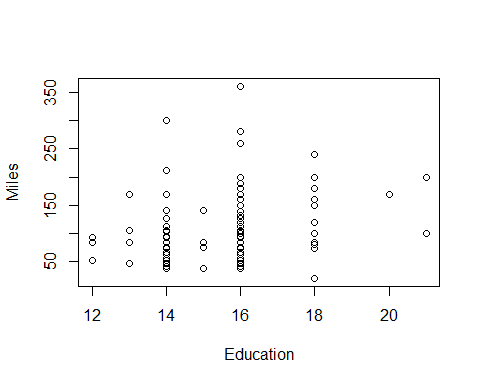
plot(Income ~ Usage , Cardio\_Goods\_tbl)



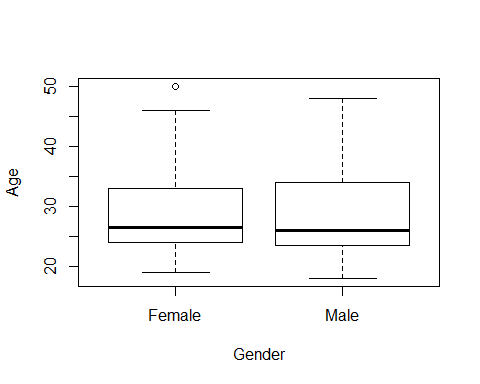
plot(Miles ~ Gender , Cardio\_Goods\_tbl)



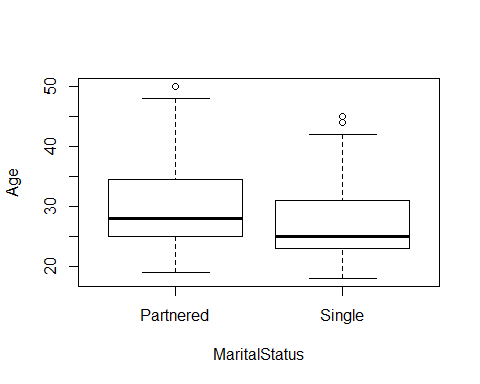
plot(Miles ~ Education , Cardio\_Goods\_tbl)  
  
plot(Miles ~ Education , Cardio\_Goods\_tbl)



plot(Age ~ Gender , Cardio\_Goods\_tbl)



plot(Age ~ MaritalStatus , Cardio\_Goods\_tbl)



# (Cardio\_GoodsUsage, Cardio\_Goods)

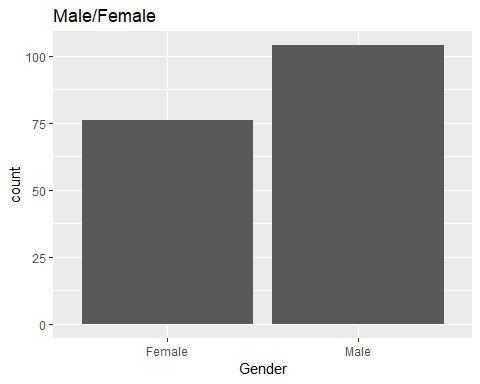
# vars(Cardio\_Goods$Age)

# ggplot2

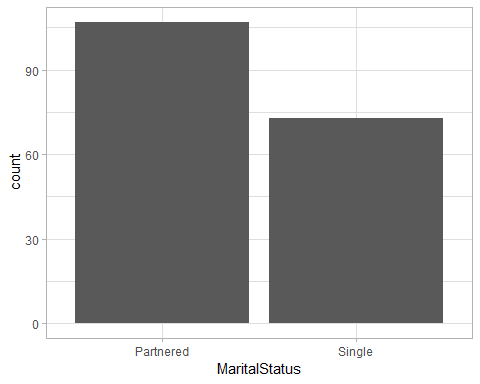
#install.packages("ggplot2")  
library(ggplot2)  
str(Cardio\_Goods\_tbl)

## Classes '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 : int 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 : int 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 : int 3 2 4 3 4 3 3 3 5 2 ...  
## $ Fitness : int 4 3 3 3 2 3 3 3 4 3 ...  
## $ Income : int 29562 31836 30699 32973 35247 32973 35247 32973 35247 37521 ...  
## $ Miles : int 112 75 66 85 47 66 75 85 141 85 ...

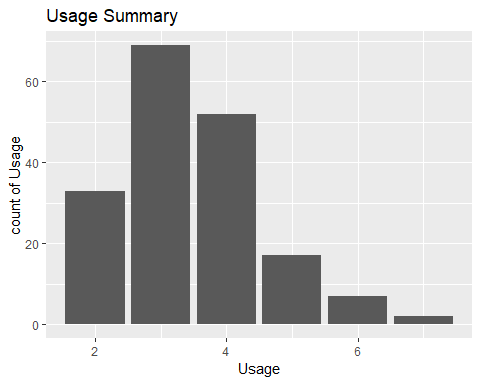
##Categorical variables   
ggplot(Cardio\_Goods\_tbl,aes(x=Gender)) +  
 geom\_bar() + labs(y='count', title = 'Male/Female')



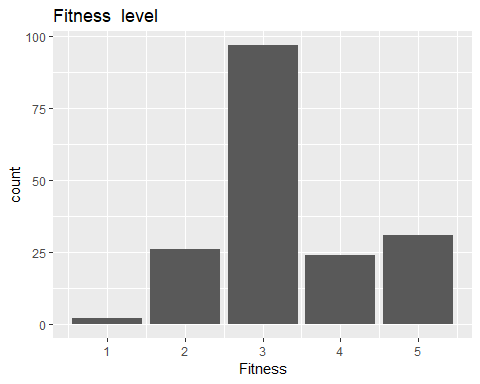
ggplot(Cardio\_Goods\_tbl,aes(x=MaritalStatus)) + geom\_bar() + theme\_light()



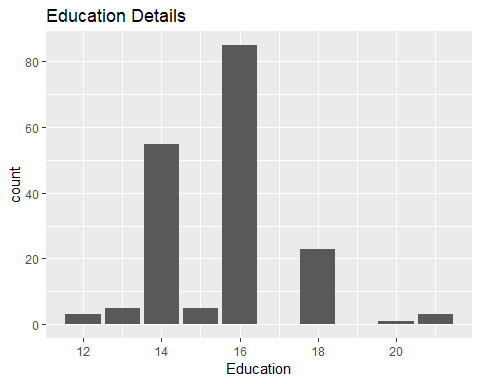
ggplot(Cardio\_Goods\_tbl,aes(x=Usage)) + geom\_bar()+  
 labs(y='count of Usage', title = 'Usage Summary')



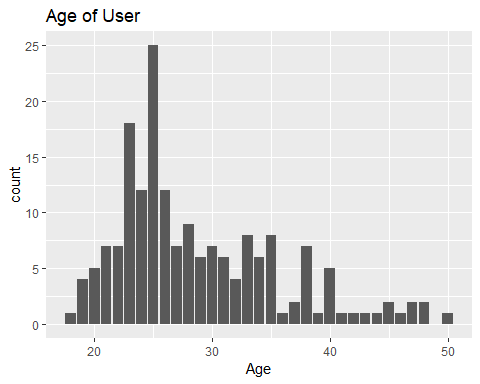
ggplot(Cardio\_Goods\_tbl,aes(x=Fitness)) + geom\_bar() +  
 labs(y='count', title = 'Fitness level')



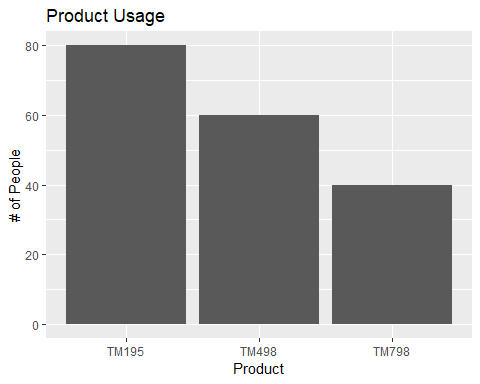
ggplot(Cardio\_Goods\_tbl,aes(x=Education)) + geom\_bar() +  
 labs(y='count', title = 'Education Details')



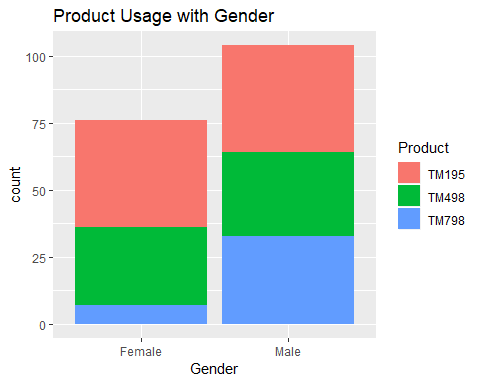
ggplot(Cardio\_Goods\_tbl,aes(x=Age)) + geom\_bar() +   
 labs(y='count', title = 'Age of User')



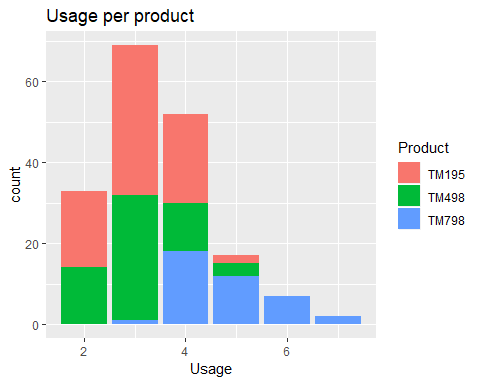
ggplot(Cardio\_Goods\_tbl,aes(x=Product)) + geom\_bar() +   
 labs(y='# of People', title = 'Product Usage')



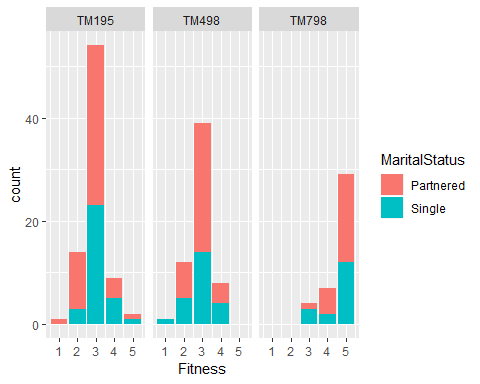
ggplot(Cardio\_Goods\_tbl,aes(x=Gender, fill=Product)) +  
 geom\_bar() + labs(y='count', title = 'Product Usage with Gender')



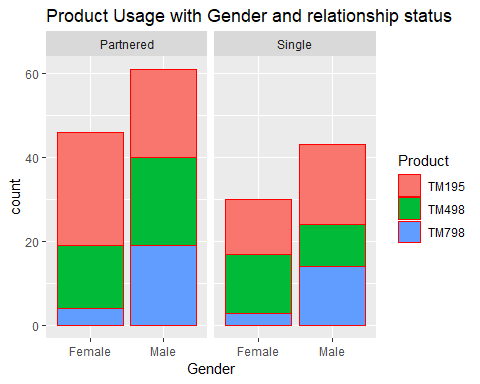
ggplot(Cardio\_Goods\_tbl,aes(x=Usage, fill=Product)) +  
 geom\_bar() + labs(y='count', title = 'Usage per product')



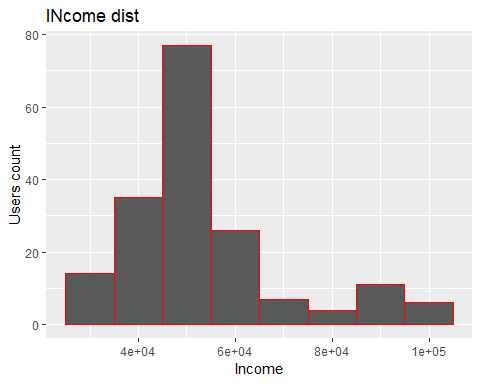
ggplot(Cardio\_Goods\_tbl,aes(x=Fitness, fill=MaritalStatus)) +  
 facet\_wrap(~ Product) +  
 geom\_bar()



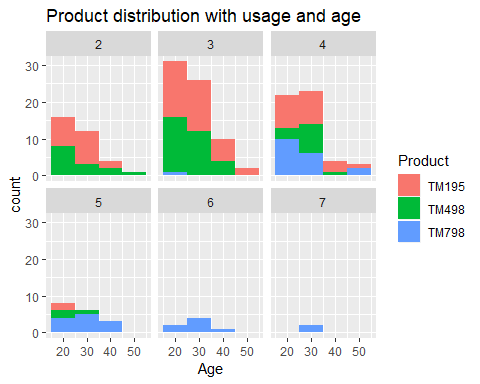
ggplot(Cardio\_Goods\_tbl,aes(x=Gender, fill=Product)) +  
 facet\_wrap(~ MaritalStatus)+  
 geom\_bar(colour="red") + labs(y='count',title = 'Product Usage with Gender and relationship status')



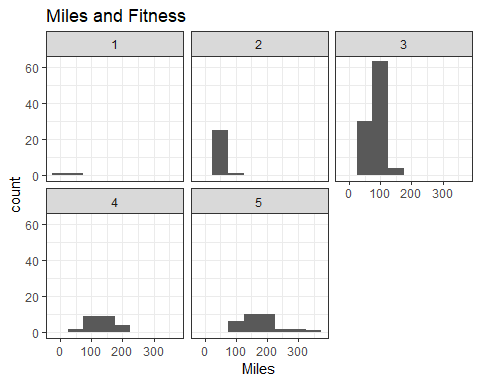
##Comtineous variables   
 ggplot(Cardio\_Goods\_tbl, aes(x=Income)) +  
 geom\_histogram(binwidth = 10000, colour='red' )+  
 labs(y='Users count', title = 'INcome dist')



ggplot(Cardio\_Goods\_tbl, aes(x=Age, fill=Product)) +  
 facet\_wrap(~ Usage ) +  
 geom\_histogram(binwidth = 10 )+  
 labs(y=' count', title = 'Product distribution with usage and age')



ggplot(Cardio\_Goods\_tbl, aes(x=Miles, fill=Usage)) +  
 facet\_wrap(~Fitness) +  
 geom\_histogram(binwidth = 50 )+ theme\_bw() +  
 labs(y=' count', title = 'Miles and Fitness')



ggplot(Cardio\_Goods\_tbl, aes(x=Miles)) +  
 facet\_wrap(~Income) +  
 geom\_histogram(binwidth = 50 )+ theme\_bw() +  
 labs(y=' count', title = 'Miles and Fitness')

