

# R\_File\_Young\_Survey

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

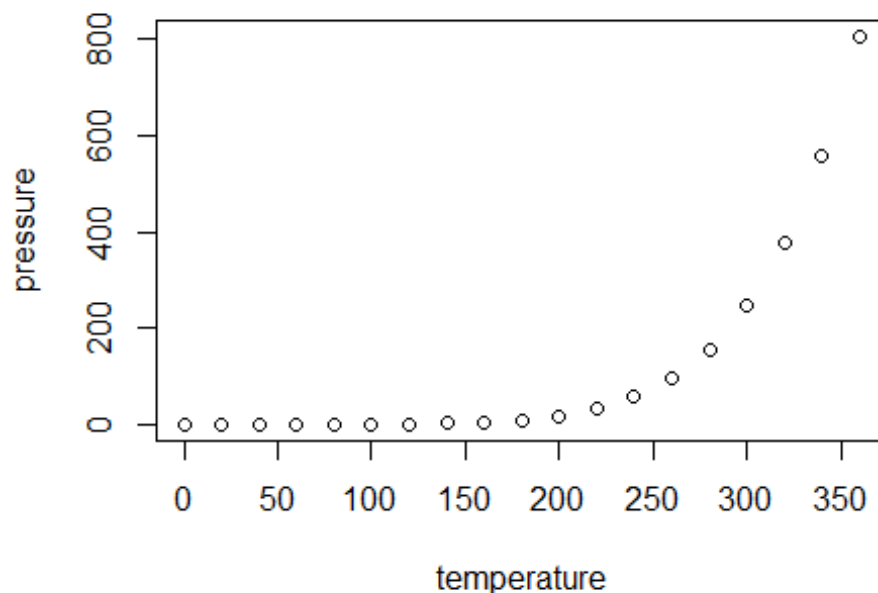
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

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

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
data <- read.csv("G:/Rutgers/MVA/Project/responses.csv")
View(data)

# The structure of the datasets
dim(data)

## [1] 1010 150

## Separate out numeric variables and categoric variables
data_cat <- data[,sapply(data, is.factor)]
data_num <- data[,!sapply(data, is.factor)]
dim(data_cat) # 11 features

## [1] 1010 11

dim(data_num) # 139 features

## [1] 1010 139

# Grouping of columns accoring to the preferences
music_data = data[,1:19]
movie_data = data[,20:31]
hobbies_data = data[,32:63]
phobia_data = data[,64:73]
health_data = data[,74:76]
```

```
traits_data = data[,77:133]
spend_data = data[,134:140]
demo_data = data[,141:150]
```

## MISSING VALUES ANALYSIS

```
sum(is.na(data))
```

```
## [1] 571
```

```
colSums(is.na(data)) #suming the na values as per the column level
```

```
##           Music           Slow.songs.or.fast.songs
##           3                2
##           Dance                Folk
##           4                5
##           Country           Classical.music
##           5                7
##           Musical                Pop
##           2                3
##           Rock           Metal.or.Hardrock
##           6                3
##           Punk           Hiphop..Rap
##           8                4
##           Reggae..Ska           Swing..Jazz
##           7                6
##           Rock.n.roll           Alternative
##           7                7
##           Latino           Techno..Trance
##           8                7
##           Opera                Movies
##           1                6
##           Horror           Thriller
##           2                1
##           Comedy           Romantic
##           3                3
##           Sci.fi                War
##           2                2
##           Fantasy.Fairy.tales           Animated
##           3                3
##           Documentary           Western
##           8                4
##           Action           History
##           2                2
##           Psychology           Politics
##           5                1
##           Mathematics           Physics
##           3                3
##           Internet                PC
##           4                6
##           Economy.Management           Biology
```

##	5	6
##	Chemistry	Reading
##	10	6
##	Geography	Foreign.languages
##	9	5
##	Medicine	Law
##	5	1
##	Cars	Art.exhibitions
##	4	6
##	Religion	Countryside..outdoors
##	3	7
##	Dancing	Musical.instruments
##	3	1
##	Writing	Passive.sport
##	6	15
##	Active.sport	Gardening
##	4	7
##	Celebrities	Shopping
##	2	2
##	Science.and.technology	Theatre
##	6	8
##	Fun.with.friends	Adrenaline.sports
##	4	3
##	Pets	Flying
##	4	3
##	Storm	Darkness
##	1	2
##	Heights	Spiders
##	3	5
##	Snakes	Rats
##	0	3
##	Ageing	Dangerous.dogs
##	1	1
##	Fear.of.public.speaking	Smoking
##	1	0
##	Alcohol	Healthy.eating
##	0	3
##	Daily.events	Prioritising.workload
##	7	5
##	Writing.notes	Workaholism
##	3	5
##	Thinking.ahead	Final.judgement
##	3	7
##	Reliability	Keeping.promises
##	4	1
##	Loss.of.interest	Friends.versus.money
##	4	6
##	Funniness	Fake
##	4	1
##	Criminal.damage	Decision.making

##	7	4
##	Elections	Self.criticism
##	3	5
##	Judgment.calls	Hypochondria
##	4	4
##	Empathy	Eating.to.survive
##	5	0
##	Giving	Compassion.to.animals
##	6	7
##	Borrowed.stuff	Loneliness
##	2	1
##	Cheating.in.school	Health
##	4	1
##	Changing.the.past	God
##	2	2
##	Dreams	Charity
##	0	3
##	Number.of.friends	Punctuality
##	0	0
##	Lying	Waiting
##	0	3
##	New.environment	Mood.swings
##	2	4
##	Appearance.and.gestures	Socializing
##	3	5
##	Achievements	Responding.to.a.serious.letter
##	2	6
##	Children	Assertiveness
##	4	2
##	Getting.angry	Knowing.the.right.people
##	4	2
##	Public.speaking	Unpopularity
##	2	3
##	Life.struggles	Happiness.in.life
##	3	4
##	Energy.levels	Small...big.dogs
##	5	4
##	Personality	Finding.lost.valuables
##	4	4
##	Getting.up	Interests.or.hobbies
##	5	3
##	Parents..advice	Questionnaires.or.polls
##	2	4
##	Internet.usage	Finances
##	0	3
##	Shopping.centres	Branded.clothing
##	2	2
##	Entertainment.spending	Spending.on.looks
##	3	3
##	Spending.on.gadgets	Spending.on.healthy.eating

```
##           0           2
##           Age           Height
##           7           20
##           Weight       Number.of.siblings
##           20           6
##           Gender       Left...right.handed
##           0           0
##           Education     Only.child
##           0           0
##           Village...town House...block.of.flats
##           0           0
```

*# Finding missing values with more than 1%*

*# Create a function*

```
pMiss <- function(x){sum(is.na(x))/length(x)*100}
```

```
perc_cat <- apply(data_cat, 2, pMiss)
```

```
perc_num <- apply(data_num, 2, pMiss)
```

*perc\_cat # this shows the percentage of missing value in the categorical data sat*

```
##           Smoking           Alcohol           Punctuality
##           0           0           0
##           Lying           Internet.usage           Gender
##           0           0           0
##           Left...right.handed           Education           Only.child
##           0           0           0
##           Village...town House...block.of.flats
##           0           0
```

*# this shows the percentage of missing value in the numerical data*

*perc\_num*

```
##           Music           Slow.songs.or.fast.songs
##           0.2970297           0.1980198
##           Dance           Folk
##           0.3960396           0.4950495
##           Country           Classical.music
##           0.4950495           0.6930693
##           Musical           Pop
##           0.1980198           0.2970297
##           Rock           Metal.or.Hardrock
##           0.5940594           0.2970297
##           Punk           Hiphop..Rap
##           0.7920792           0.3960396
##           Reggae..Ska           Swing..Jazz
##           0.6930693           0.5940594
##           Rock.n.roll           Alternative
##           0.6930693           0.6930693
##           Latino           Techno..Trance
##           0.7920792           0.6930693
```

##	Opera	Movies
##	0.0990099	0.5940594
##	Horror	Thriller
##	0.1980198	0.0990099
##	Comedy	Romantic
##	0.2970297	0.2970297
##	Sci.fi	War
##	0.1980198	0.1980198
##	Fantasy.Fairy.tales	Animated
##	0.2970297	0.2970297
##	Documentary	Western
##	0.7920792	0.3960396
##	Action	History
##	0.1980198	0.1980198
##	Psychology	Politics
##	0.4950495	0.0990099
##	Mathematics	Physics
##	0.2970297	0.2970297
##	Internet	PC
##	0.3960396	0.5940594
##	Economy.Management	Biology
##	0.4950495	0.5940594
##	Chemistry	Reading
##	0.9900990	0.5940594
##	Geography	Foreign.languages
##	0.8910891	0.4950495
##	Medicine	Law
##	0.4950495	0.0990099
##	Cars	Art.exhibitions
##	0.3960396	0.5940594
##	Religion	Countryside..outdoors
##	0.2970297	0.6930693
##	Dancing	Musical.instruments
##	0.2970297	0.0990099
##	Writing	Passive.sport
##	0.5940594	1.4851485
##	Active.sport	Gardening
##	0.3960396	0.6930693
##	Celebrities	Shopping
##	0.1980198	0.1980198
##	Science.and.technology	Theatre
##	0.5940594	0.7920792
##	Fun.with.friends	Adrenaline.sports
##	0.3960396	0.2970297
##	Pets	Flying
##	0.3960396	0.2970297
##	Storm	Darkness
##	0.0990099	0.1980198
##	Heights	Spiders
##	0.2970297	0.4950495

##	Snakes	Rats
##	0.0000000	0.2970297
##	Ageing	Dangerous.dogs
##	0.0990099	0.0990099
##	Fear.of.public.speaking	Healthy.eating
##	0.0990099	0.2970297
##	Daily.events	Prioritising.workload
##	0.6930693	0.4950495
##	Writing.notes	Workaholism
##	0.2970297	0.4950495
##	Thinking.ahead	Final.judgement
##	0.2970297	0.6930693
##	Reliability	Keeping.promises
##	0.3960396	0.0990099
##	Loss.of.interest	Friends.versus.money
##	0.3960396	0.5940594
##	Funniness	Fake
##	0.3960396	0.0990099
##	Criminal.damage	Decision.making
##	0.6930693	0.3960396
##	Elections	Self.criticism
##	0.2970297	0.4950495
##	Judgment.calls	Hypochondria
##	0.3960396	0.3960396
##	Empathy	Eating.to.survive
##	0.4950495	0.0000000
##	Giving	Compassion.to.animals
##	0.5940594	0.6930693
##	Borrowed.stuff	Loneliness
##	0.1980198	0.0990099
##	Cheating.in.school	Health
##	0.3960396	0.0990099
##	Changing.the.past	God
##	0.1980198	0.1980198
##	Dreams	Charity
##	0.0000000	0.2970297
##	Number.of.friends	Waiting
##	0.0000000	0.2970297
##	New.environment	Mood.swings
##	0.1980198	0.3960396
##	Appearance.and.gestures	Socializing
##	0.2970297	0.4950495
##	Achievements	Responding.to.a.serious.letter
##	0.1980198	0.5940594
##	Children	Assertiveness
##	0.3960396	0.1980198
##	Getting.angry	Knowing.the.right.people
##	0.3960396	0.1980198
##	Public.speaking	Unpopularity
##	0.1980198	0.2970297



```
##          Life.struggles          Happiness.in.life
##          0.2970297          0.3960396
##          Energy.levels          Small...big.dogs
##          0.4950495          0.3960396
##          Personality          Finding.lost.valuables
##          0.3960396          0.3960396
##          Getting.up          Interests.or.hobbies
##          0.4950495          0.2970297
##          Parents..advice          Questionnaires.or.polls
##          0.1980198          0.3960396
##          Finances          Shopping.centres
##          0.2970297          0.1980198
##          Branded.clothing          Entertainment.spending
##          0.1980198          0.2970297
##          Spending.on.looks          Spending.on.gadgets
##          0.2970297          0.0000000
##          Spending.on.healthy.eating          Age
##          0.1980198          0.6930693
##          Height          Weight
##          1.9801980          1.9801980
##          Number.of.siblings
##          0.5940594
```

IMPUTE MISSING VALUES : Imputation based on predictive method using features

```
# Numeric variable imputation
library(mice)

## Warning: package 'mice' was built under R version 3.6.2

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.6.2

##
## Attaching package: 'mice'

## The following objects are masked from 'package:base':
##
##      cbind, rbind

# methods(mice)

impu_num = mice(data_num, m=1, method = 'pmm', maxit = 1, seed = 200)

##
## iter imp variable
## 1 1 Music Slow.songs.or.fast.songs Dance Folk Country
Classical.music Musical Pop Rock Metal.or.Hardrock Punk Hiphop..Rap
Reggae..Ska Swing..Jazz Rock.n.roll Alternative Latino Techno..Trance
Opera Movies Horror Thriller Comedy Romantic Sci.fi War
Fantasy.Fairy.tales Animated Documentary Western Action History
```

Psychology Politics Mathematics Physics Internet PC Economy.Management  
 Biology Chemistry Reading Geography Foreign.languages Medicine Law  
 Cars Art.exhibitions Religion Countryside..outdoors Dancing  
 Musical.instruments Writing Passive.sport Active.sport Gardening  
 Celebrities Shopping Science.and.technology Theatre Fun.with.friends  
 Adrenaline.sports Pets Flying Storm Darkness Heights Spiders Rats  
 Ageing Dangerous.dogs Fear.of.public.speaking Healthy.eating Daily.events  
 Prioritising.workload Writing.notes Workaholism Thinking.ahead  
 Final.judgement Reliability Keeping.promises Loss.of.interest  
 Friends.versus.money Funniness Fake Criminal.damage Decision.making  
 Elections Self.criticism Judgment.calls Hypochondria Empathy Giving  
 Compassion.to.animals Borrowed.stuff Loneliness Cheating.in.school Health  
 Changing.the.past God Charity Waiting New.environment Mood.swings  
 Appearance.and.gestures Socializing Achievements  
 Responding.to.a.serious.letter Children Assertiveness Getting.angry  
 Knowing.the.right.people Public.speaking Unpopularity Life.struggles  
 Happiness.in.life Energy.levels Small...big.dogs Personality  
 Finding.lost.valuables Getting.up Interests.or.hobbies Parents..advice  
 Questionnaires.or.polls Finances Shopping.centres Branded.clothing  
 Entertainment.spending Spending.on.looks Spending.on.healthy.eating Age  
 Height Weight Number.of.siblings

```
# summary(impu_num)
```

```
impu_num$imp$Age # Imputed data at each iterations
```

```
##      1
## 138 20
## 143 22
## 463 19
## 550 22
## 736 25
## 903 18
## 961 20
```

```
impu_num$imp$Weight
```

```
##      1
## 138 74
## 143 55
## 165 92
## 210 89
## 277 65
## 406 63
## 454 64
## 496 51
## 510 57
## 552 70
## 559 58
## 647 68
## 704 76
## 713 64
```

```
## 791 57
## 843 50
## 876 88
## 890 60
## 903 58
## 961 60
```

Parameters: 'pmm' - predictive mean matching method m=5 - no.of multiple imputed datasets maxit = 10 - no.of iterations The computational time is dependent on the 'maxit' - for me it took more than 20 min

```
# Get the numeirc imputed data
impu_num_compl = complete(impu_num,action = 1)
```

Imputing categorical missing variables

```
impu_cat = mice(data_cat, m=5, maxit = 10, seed = 200, method = 'pmm')
```

```
##
##  iter imp variable
##  1   1
##  1   2
##  1   3
##  1   4
##  1   5
##  2   1
##  2   2
##  2   3
##  2   4
##  2   5
##  3   1
##  3   2
##  3   3
##  3   4
##  3   5
##  4   1
##  4   2
##  4   3
##  4   4
##  4   5
##  5   1
##  5   2
##  5   3
##  5   4
##  5   5
##  6   1
##  6   2
##  6   3
##  6   4
##  6   5
##  7   1
```

```
## 7 2
## 7 3
## 7 4
## 7 5
## 8 1
## 8 2
## 8 3
## 8 4
## 8 5
## 9 1
## 9 2
## 9 3
## 9 4
## 9 5
## 10 1
## 10 2
## 10 3
## 10 4
## 10 5

# Get categorical imputed data
impu_cat_compl = complete(impu_cat, 1)

sum(is.na(impu_num_compl)) # No missing values
## [1] 0

sum(is.na(impu_cat_compl)) # No missing values
## [1] 0
```

Outlier detection:

```
library(outliers)

outlier(data_num)
```

##	Music	Slow.songs.or.fast.songs
##	1	1
##	Dance	Folk
##	1	5
##	Country	Classical.music
##	5	5
##	Musical	Pop
##	5	1
##	Rock	Metal.or.Hardrock
##	1	5
##	Punk	Hiphop..Rap
##	5	5
##	Reggae..Ska	Swing..Jazz
##	5	5
##	Rock.n.roll	Alternative

##	1	5
##	Latino	Techno..Trance
##	5	5
##	Opera	Movies
##	5	1
##	Horror	Thriller
##	5	1
##	Comedy	Romantic
##	1	1
##	Sci.fi	War
##	1	1
##	Fantasy.Fairy.tales	Animated
##	1	1
##	Documentary	Western
##	1	5
##	Action	History
##	1	1
##	Psychology	Politics
##	1	5
##	Mathematics	Physics
##	5	5
##	Internet	PC
##	1	1
##	Economy.Management	Biology
##	5	5
##	Chemistry	Reading
##	5	1
##	Geography	Foreign.languages
##	1	1
##	Medicine	Law
##	5	5
##	Cars	Art.exhibitions
##	5	5
##	Religion	Countryside..outdoors
##	5	1
##	Dancing	Musical.instruments
##	5	5
##	Writing	Passive.sport
##	5	1
##	Active.sport	Gardening
##	1	5
##	Celebrities	Shopping
##	5	1
##	Science.and.technology	Theatre
##	1	1
##	Fun.with.friends	Adrenaline.sports
##	2	5
##	Pets	Flying
##	1	5
##	Storm	Darkness

##	5	5
##	Heights	Spiders
##	5	5
##	Snakes	Rats
##	1	5
##	Ageing	Dangerous.dogs
##	5	1
##	Fear.of.public.speaking	Healthy.eating
##	5	1
##	Daily.events	Prioritising.workload
##	1	5
##	Writing.notes	Workaholism
##	1	5
##	Thinking.ahead	Final.judgement
##	1	5
##	Reliability	Keeping.promises
##	1	1
##	Loss.of.interest	Friends.versus.money
##	5	1
##	Funniness	Fake
##	1	5
##	Criminal.damage	Decision.making
##	5	1
##	Elections	Self.criticism
##	1	1
##	Judgment.calls	Hypochondria
##	1	5
##	Empathy	Eating.to.survive
##	1	5
##	Giving	Compassion.to.animals
##	5	1
##	Borrowed.stuff	Loneliness
##	1	5
##	Cheating.in.school	Health
##	1	1
##	Changing.the.past	God
##	5	1
##	Dreams	Charity
##	1	5
##	Number.of.friends	Waiting
##	1	5
##	New.environment	Mood.swings
##	1	1
##	Appearance.and.gestures	Socializing
##	1	1
##	Achievements	Responding.to.a.serious.letter
##	5	1
##	Children	Assertiveness
##	1	1
##	Getting.angry	Knowing.the.right.people

```
##          1          1
##      Public.speaking      Unpopularity
##          1          1
##      Life.struggles      Happiness.in.life
##          1          1
##      Energy.levels      Small...big.dogs
##          1          5
##      Personality      Finding.lost.valuables
##          1          5
##      Getting.up      Interests.or.hobbies
##          1          1
##      Parents..advice      Questionnaires.or.polls
##          1          5
##      Finances      Shopping.centres
##          1          1
##      Branded.clothing      Entertainment.spending
##          1          1
##      Spending.on.looks      Spending.on.gadgets
##          1          5
##      Spending.on.healthy.eating      Age
##          1          30
##      Height      Weight
##          62          165
##      Number.of.siblings
##          10
```

Demographic category - Height, Weight, Age, No.of siblings have maximum no.of outliers

```
library(ggplot2)
```

The `boxplot.stats` function; is a ancillary function that produces statistics for drawing boxplots. It returns among other information a vector `stats` with five elements: the extreme of the lower whisker, the lower 'hinge', the median, the upper 'hinge' and the extreme of the upper whisker, the extreme of the whiskers are the adjacent values (last non-missing value, i.e. every value beyond is an outlier).

```
id1 = boxplot.stats(impu_num_compl$Weight)
```

```
id1$stats
```

```
## [1] 41 55 64 75 105
```

```
id1$stats[1] #The lower adjacent value
```

```
## [1] 41
```

```
id1$stats[5] # The upper adjacent value
```

```
## [1] 105
```

```
id2 = boxplot.stats(impu_num_compl$Height)

id2$stats[1] #The Lower adjacent value
## [1] 148

id2$stats[5] # The upper adjacent value
## [1] 197

id3 = boxplot.stats(impu_num_compl$Age)

id3$stats[1] #The Lower adjacent value
## [1] 15

id3$stats[5] # The upper adjacent value
## [1] 26

id4 = boxplot.stats(impu_num_compl$Number.of.siblings)

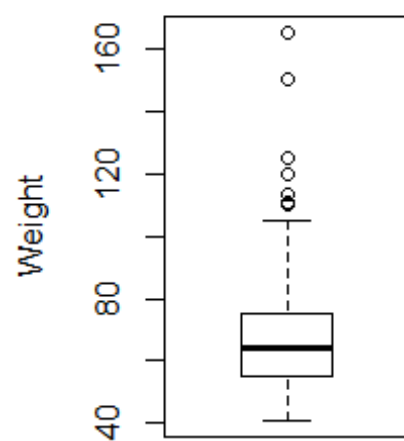
id4$stats[1] #The Lower adjacent value
## [1] 0

id4$stats[5] # The upper adjacent value
## [1] 3

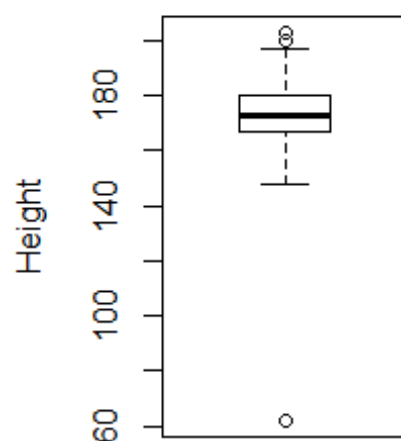
# Boxplot
par(mfrow=c(1,2))
boxplot(impu_num_compl$Weight, main = 'Outliers in Weight', ylab = 'Weight')
boxplot(impu_num_compl$Height, main = 'Otliers in Height', ylab = 'Height')
```



**Outliers in Weight**

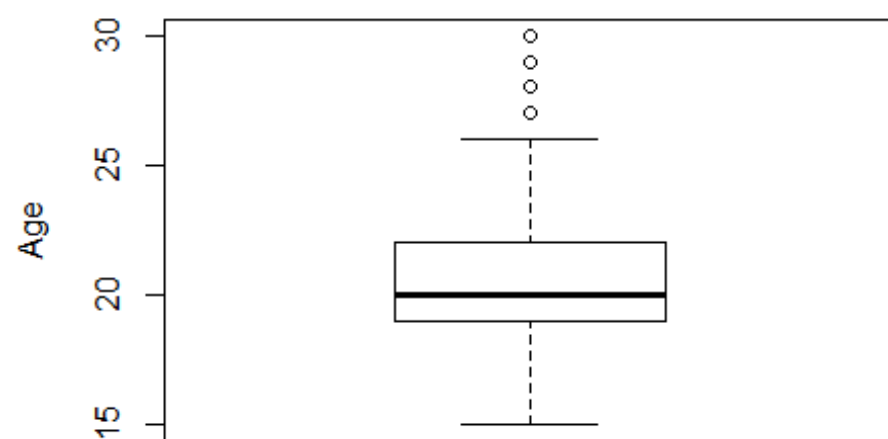


**Outliers in Height**

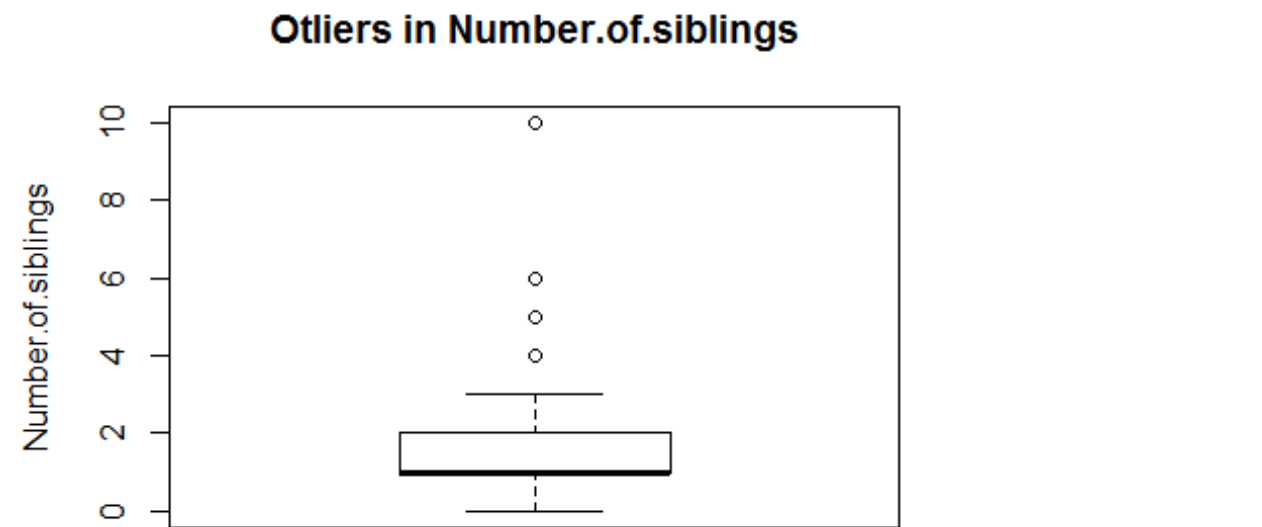


```
boxplot(impu_num_compl$Age, main = 'Outliers in Age', ylab = 'Age')
```

**Outliers in Age**



```
boxplot(impu_num_compl$Number.of.siblings , main = 'Outliers in
Number.of.siblings', ylab = 'Number.of.siblings')
```



```
# You can get the actual values of the outliers with this
```

```
boxplot(impu_num_compl$Weight, plot=FALSE)$out
```

```
## [1] 120 110 111 120 113 125 165 120 150
```

```
boxplot(impu_num_compl$Height, plot=FALSE)$out
```

```
## [1] 200 200 203 62 203 200
```

```
# no of rows having outliers
```

```
Outlier_height = boxplot(impu_num_compl$Height, plot=FALSE)$out
```

```
impu_num_compl[which(impu_num_compl$Height %in% Outlier_height),]
```

```
##      Music Slow.songs.or.fast.songs Dance Folk Country Classical.music
```

Musical

```
## 98      5      3      2      1      1      1
```

##	221	5	3	4	3	3	4
----	-----	---	---	---	---	---	---

```
## 548      5      4      1      4      2      5
```

## 677	5	4	2	2	1	2
--------	---	---	---	---	---	---

```
## 800      5      3      1      2      3      4
```

5

## 993	4			4	4	1		4		4
--------	---	--	--	---	---	---	--	---	--	---

1

##	Pop	Rock	Metal.or.	Hardrock	Punk	Hiphop..	Rap	Reggae..	Ska	Swing..	Jazz
## 98	2	1			1	1	5		2		2
## 221	1	3			1	2	5		4		4
## 548	1	4			5	2	2		2		1
## 677	1	2			2	1	2		1		2
## 800	2	4			5	4	1		1		2
## 993	3	4			4	3	2		3		2

## Rock.n.roll Alternative Latino Techno..Trance Opera Movies Horror Thriller

## 98		1		1	1		2	1	5	2
## 221		2		5	3		1	3	5	3
## 548		4		3	1		1	5	5	3
## 677		1		5	2		2	2	5	1
## 800		4		4	2		2	4	5	1
## 993		1		1	2		5	2	4	4

4

##	Comedy	Romantic	Sci.fi	War	Fantasy.	Fairy.tales	Animated	Documentary	Western
## 98	5		3	5	5		3	5	5
## 221	5		3	2	3		4	5	4
## 548	5		2	2	4		4	4	5
## 677	2		3	1	2		3	4	5
## 800	4		3	4	4		3	4	5
## 993	4		3	5	3		4	4	4

3

##	Action	History	Psychology	Politics	Mathematics	Physics	Internet	PC
## 98	5		3	2	2	4	1	5 5
## 221	4		3	3	1	1	1	3 4
## 548	4		5	3	5	1	1	3 3
## 677	3		3	1	1	1	1	5 1
## 800	3		4	2	3	4	3	5 5
## 993	4		5	2	3	5	5	4 3

## Economy.Management Biology Chemistry Reading Geography Foreign.languages

## 98		2	1	1	3	1
## 221		1	5	5	3	1

5									
## 548		1	2	1	5	5			
5									
## 677		1	1	1	1	3			
4									
## 800		2	3	2	5	3			
4									
## 993		1	4	4	2	4			
2									
##	Medicine	Law	Cars	Art.exhibitions	Religion	Countryside..outdoors			
Dancing									
## 98	1	2	2		1	4		4	
1									
## 221	3	2	4		3	2		4	
2									
## 548	2	2	1		4	4		3	
1									
## 677	1	5	2		3	1		4	
1									
## 800	3	2	3		4	2		4	
2									
## 993	1	1	2		1	1		1	
1									
##	Musical.instruments	Writing	Passive.sport	Active.sport	Gardening				
## 98		1	1	1	5	2			
## 221		5	1	5	5	3			
## 548		5	4	5	5	2			
## 677		1	1	5	5	2			
## 800		3	3	3	2	4			
## 993		1	5	4	1	1			
##	Celebrities	Shopping	Science.and.technology	Theatre	Fun.with.friends				
## 98	1	1		4	1	3			
## 221	1	3		4	2	5			
## 548	1	1		2	5	5			
## 677	4	3		3	5	5			
## 800	3	3		4	5	4			
## 993	1	1		5	2	5			
##	Adrenaline.sports	Pets	Flying Storm	Darkness	Heights	Spiders	Snakes		
Rats									
## 98		3	5	1	1	2	3	2	4
1									
## 221		3	5	1	3	2	2	2	3
2									
## 548		5	4	1	1	2	3	1	1
1									
## 677		2	3	1	2	2	2	3	3
4									
## 800		3	5	2	1	3	2	3	3
2									
## 993		1	1	3	1	1	5	5	3

```

2
##      Ageing Dangerous.dogs Fear.of.public.speaking Healthy.eating
Daily.events
## 98      1      1      2      3
1
## 221     1      4      3      4
2
## 548     1      4      1      2
5
## 677     3      3      3      4
2
## 800     2      2      4      3
3
## 993     2      3      4      1
3
##      Prioritising.workload Writing.notes Workaholism Thinking.ahead
## 98      1      1      3      2
## 221     2      2      1      2
## 548     2      1      2      2
## 677     1      2      1      2
## 800     3      2      4      3
## 993     1      1      1      3
##      Final.judgement Reliability Keeping.promises Loss.of.interest
## 98      1      4      5      5
## 221     2      3      5      2
## 548     1      4      4      1
## 677     3      3      3      4
## 800     1      4      4      2
## 993     3      3      5      1
##      Friends.versus.money Funniness Fake Criminal.damage Decision.making
## 98      3      4      4      5      1
## 221     4      4      2      5      4
## 548     5      5      1      1      3
## 677     4      2      2      5      2
## 800     4      3      3      2      3
## 993     1      2      1      1      3
##      Elections Self.criticism Judgment.calls Hypochondria Empathy
## 98      5      4      3      1      4
## 221     4      2      4      1      4
## 548     5      5      5      1      4
## 677     1      2      4      2      4
## 800     2      4      4      2      4
## 993     5      2      3      1      4
##      Eating.to.survive Giving Compassion.to.animals Borrowed.stuff
Loneliness
## 98      4      1      2      5
5
## 221     2      3      5      5
2
## 548     1      1      5      5

```

1						
## 677	1	2		1		2
2						
## 800	2	3		5		4
4						
## 993	1	4		2		4
3						
##	Cheating.in.school	Health	Changing.the.past	God	Dreams	Charity
## 98	5	1		5	5	4
## 221	5	3		1	2	3
## 548	5	2		2	1	3
## 677	5	3		4	4	3
## 800	3	3		2	2	4
## 993	1	3		4	5	3
##	Number.of.friends	Waiting	New.environment	Mood.swings		
## 98	2	4		5		5
## 221	4	2		4		3
## 548	3	5		5		1
## 677	3	3		3		5
## 800	2	4		2		2
## 993	3	2		2		4
##	Appearence.and.gestures	Socializing	Achievements			
## 98		2	3		4	
## 221		3	4		3	
## 548		1	4		2	
## 677		3	3		3	
## 800		3	2		3	
## 993		2	1		3	
##	Responding.to.a.serious.letter	Children	Assertiveness	Getting.angry		
## 98		3	5		5	2
## 221		4	3		3	2
## 548		3	4		2	1
## 677		4	4		1	2
## 800		4	3		3	2
## 993		3	3		5	3
##	Knowing.the.right.people	Public.speaking	Unpopularity	Life.struggles		
## 98		4	2		3	1
## 221		4	5		3	1
## 548		3	1		3	1
## 677		4	2		3	3
## 800		4	4		3	4
## 993		3	5		3	1
##	Happiness.in.life	Energy.levels	Small...big.dogs	Personality		
## 98		3	3		3	3
## 221		5	4		3	4
## 548		4	4		3	3
## 677		3	2		2	4
## 800		3	2		2	3
## 993		2	3		4	3
##	Finding.lost.valuables	Getting.up	Interests.or.hobbies	Parents..advice		

```

## 98          1          1          5          3
## 221         3          4          5          4
## 548         4          4          5          4
## 677         4          2          3          3
## 800         4          2          2          3
## 993         1          5          3          3
##      Questionnaires.or.polls  Finances  Shopping.centres  Branded.clothing
## 98          1          2          1          1
## 221         3          2          3          2
## 548         3          3          2          1
## 677         1          3          4          2
## 800         3          4          3          3
## 993         2          1          1          1
##      Entertainment.spending  Spending.on.looks  Spending.on.gadgets
## 98          5          1          5
## 221         5          4          2
## 548         3          1          1
## 677         2          4          2
## 800         2          3          4
## 993         5          1          1
##      Spending.on.healthy.eating  Age  Height  Weight  Number.of.siblings
## 98          2  19    200    75          1
## 221         3  18    200    90          2
## 548         4  21    203    80          2
## 677         4  20     62    55          2
## 800         4  18    203    89          2
## 993         4  30    200   150          1

outlier_siblings = boxplot(impu_num_compl$Number.of.siblings, plot=FALSE)$out
impu_num_compl[which(impu_num_compl$Number.of.siblings %in%
outlier_siblings),]

##      Music  Slow.songs.or.fast.songs  Dance  Folk  Country  Classical.music
Musical
## 13         5          3      1      2          1          4
3
## 33         5          5      3      1          3          2
3
## 35         5          4      3      2          1          3
4
## 54         5          3      4      4          3          4
5
## 92         5          5      1      5          1          2
5
## 125        4          4      1      4          5          5
5
## 139        5          3      2      3          3          4
5
## 150        2          3      1      3          1          3
4

```

## 230	4				3	4	2	4		1		
1												
## 260	2				3	3	3	3		2		
2												
## 366	5				3	2	2	3		4		
2												
## 372	5				3	5	5	5		5		
5												
## 443	5				3	4	2	4		3		
3												
## 476	5				3	3	2	5		4		
5												
## 503	5				4	4	3	2		4		
2												
## 525	5				3	2	1	2		4		
2												
## 571	5				3	3	1	1		2		
1												
## 620	5				1	3	3	1		5		
4												
## 683	4				4	4	2	3		3		
3												
## 695	5				3	3	3	4		5		
2												
## 743	5				3	1	3	4		5		
3												
## 751	5				3	5	5	2		5		
3												
## 804	5				5	2	1	1		2		
4												
## 847	5				3	5	2	2		3		
3												
## 849	5				2	2	1	1		3		
1												
## 854	5				3	4	4	2		3		
4												
## 931	5				3	3	2	1		2		
3												
## 938	5				3	4	1	1		5		
4												
## 941	5				3	2	1	1		3		
3												
## 958	5				4	2	1	1		1		
2												
## 1007	4				4	5	1	3		4		
1												
##	Pop	Rock	Metal	or.	Hardrock	Punk	Hiphop	..Rap	Reggae	..Ska	Swing	..Jazz
## 13	3	5			4	2		3		1		1
## 33	3	4			3	4		4		2		2
## 35	4	5			3	4		2		3		3



## 54	4	3	2	2	3	3	3
## 92	3	5	5	5	1	5	1
## 125	3	5	3	3	1	2	4
## 139	2	1	1	1	3	3	4
## 150	1	2	2	2	3	2	1
## 230	4	3	1	4	2	5	3
## 260	3	4	2	2	2	1	3
## 366	1	5	4	5	1	5	4
## 372	5	5	3	3	2	2	5
## 443	5	4	1	4	3	4	4
## 476	3	5	4	4	3	4	3
## 503	2	5	3	4	3	3	4
## 525	2	5	5	5	2	5	2
## 571	3	3	1	2	2	1	1
## 620	3	3	2	2	3	2	3
## 683	4	4	3	4	3	4	3
## 695	3	5	4	2	3	3	5
## 743	2	4	2	1	3	4	2
## 751	3	3	1	1	5	1	5
## 804	3	5	4	5	1	1	2
## 847	4	2	1	1	4	2	4
## 849	3	5	3	4	1	4	2
## 854	3	3	1	1	2	3	2
## 931	3	2	2	3	5	3	1
## 938	5	5	4	5	4	5	5
## 941	3	3	1	1	4	4	5
## 958	2	2	1	1	5	1	2
## 1007	4	1	1	4	1	1	2

## 366	5	5	1	2	4	5	1
5							
## 372	5	3	5	2	5	5	1
2							
## 443	4	2	2	2	1	5	5
4							
## 476	5	3	4	1	2	4	4
3							
## 503	5	3	1	3	3	5	3
5							
## 525	3	5	2	1	3	5	3
4							
## 571	3	2	2	1	1	5	4
2							
## 620	3	2	3	2	5	5	2
4							
## 683	4	3	2	5	2	5	2
4							
## 695	5	1	5	4	3	5	4
5							
## 743	3	4	5	5	2	5	2
3							
## 751	4	5	5	1	4	4	3
4							
## 804	3	2	1	1	2	5	3
4							
## 847	2	2	3	5	3	5	3
3							
## 849	3	5	1	2	1	3	1
2							
## 854	2	2	3	1	4	5	3
3							
## 931	1	1	3	3	2	5	3
2							
## 938	5	5	5	2	5	5	1
3							
## 941	2	2	1	5	1	4	4
2							
## 958	1	1	2	2	1	4	4
4							
## 1007	3	1	3	4	1	5	2
5							
##	Comedy	Romantic	Sci.fi	War	Fantasy.Fairy.tales	Animated	Documentary
## 13	4	3	1	4	5	3	3
## 33	5	5	2	3	5	5	3
## 35	5	4	4	2	2	5	3
## 54	5	5	1	5	5	3	3
## 92	5	5	3	1	5	5	1
## 125	5	4	5	5	5	5	5
## 139	5	5	1	3	5	4	4

## 150	4	2	5	5	4	4	4
## 230	4	1	2	2	3	2	3
## 260	5	1	1	4	3	1	5
## 366	2	2	1	5	2	2	5
## 372	5	5	3	2	5	5	3
## 443	5	4	3	3	5	5	4
## 476	5	5	2	3	5	5	3
## 503	4	2	5	5	5	5	3
## 525	3	3	3	3	4	5	4
## 571	5	3	1	5	3	3	3
## 620	4	2	2	5	3	3	5
## 683	4	2	4	5	5	4	4
## 695	5	3	4	5	4	4	5
## 743	5	4	3	5	5	5	4
## 751	5	4	4	4	5	5	5
## 804	5	4	1	2	3	2	2
## 847	5	5	5	2	2	4	4
## 849	3	4	4	2	1	2	1
## 854	5	5	1	3	5	3	4
## 931	5	5	2	3	5	5	3
## 938	5	5	4	2	4	5	3
## 941	5	3	5	5	4	3	5
## 958	5	4	4	4	3	2	3
## 1007	5	1	5	5	1	5	5

## Western Action History Psychology Politics Mathematics Physics  
Internet PC

## 13	1	1	4	4	4	1	1
3 2							
## 33	2	4	4	5	3	1	1
5 4							
## 35	1	5	4	5	4	3	1
5 4							
## 54	1	5	3	3	4	1	1
3 2							
## 92	1	3	1	1	1	1	1
5 5							
## 125	5	4	4	4	4	1	3
4 5							
## 139	3	2	2	3	2	1	1
3 1							
## 150	2	2	2	3	1	5	5
3 5							
## 230	3	5	2	1	1	4	2
5 5							
## 260	1	1	5	4	4	1	2
2 2							
## 366	3	4	5	5	4	1	1
3 2							
## 372	3	2	2	5	1	5	3
3 3							

## 443	3	3	2	2	3	3	2
3 4							
## 476	1	2	4	5	3	3	2
4 2							
## 503	5	5	2	1	4	3	1
5 5							
## 525	1	1	3	4	1	3	2
5 2							
## 571	1	5	4	3	2	1	1
5 3							
## 620	2	4	3	2	2	5	5
5 4							
## 683	3	5	3	3	2	4	2
4 4							
## 695	2	4	5	5	5	5	3
5 5							
## 743	5	5	3	5	3	4	3
5 4							
## 751	3	3	4	5	2	4	3
3 3							
## 804	1	2	3	3	3	1	1
5 3							
## 847	2	3	2	1	1	2	2
5 4							
## 849	1	3	1	5	2	1	2
5 3							
## 854	4	3	3	3	2	3	3
4 3							
## 931	2	3	3	2	2	1	2
5 4							
## 938	5	5	3	5	1	5	4
5 3							
## 941	4	5	4	3	2	3	3
5 5							
## 958	2	5	3	3	3	1	1
4 4							
## 1007	2	5	4	1	1	5	4
5 5							
##	Economy.Management Biology Chemistry Reading Geography						
Foreign.languages							
## 13		1	5	5	5	3	
5							
## 33		3	1	1	3	3	
3							
## 35		4	3	1	3	2	
5							
## 54		2	4	2	4	5	
5							
## 92		1	5	1	1	1	
5							

## 125 4	3	2	2	4	4
## 139 5	2	5	4	5	3
## 150 3	3	2	3	2	3
## 230 3	2	3	2	1	5
## 260 4	2	4	3	5	4
## 366 5	4	4	4	4	5
## 372 3	3	5	5	5	2
## 443 3	2	3	3	2	3
## 476 3	2	3	2	5	3
## 503 4	4	1	1	1	1
## 525 3	1	5	4	3	4
## 571 5	2	2	1	1	4
## 620 4	2	3	2	1	5
## 683 3	2	2	1	3	4
## 695 5	3	2	2	2	4
## 743 5	5	4	1	5	4
## 751 2	2	3	2	2	2
## 804 3	2	2	1	2	2
## 847 5	3	1	1	3	2
## 849 3	5	2	2	1	2
## 854 3	3	3	3	4	4
## 931 3	3	2	1	1	3
## 938 4	2	4	2	5	5
## 941 3	1	2	1	2	1
## 958 1	1	1	1	1	1

## 1007		2	2	1	1	2
2						
##	Medicine	Law	Cars	Art.exhibitions	Religion	Countryside..outdoors
Dancing						
## 13	5	2	3	1	1	5
3						
## 33	1	2	3	4	2	2
1						
## 35	3	2	5	3	5	5
2						
## 54	4	3	5	5	5	4
5						
## 92	5	1	1	1	1	5
5						
## 125	4	4	4	5	3	5
2						
## 139	5	1	1	1	5	5
2						
## 150	2	1	2	1	3	4
1						
## 230	2	2	2	2	2	5
2						
## 260	3	4	4	2	3	5
5						
## 366	4	4	5	4	4	5
1						
## 372	5	2	2	3	5	5
5						
## 443	2	2	2	4	3	3
2						
## 476	2	3	1	2	2	4
2						
## 503	1	2	5	1	1	2
1						
## 525	5	1	1	3	3	5
2						
## 571	1	1	5	1	1	5
1						
## 620	2	3	4	1	1	2
1						
## 683	2	2	4	2	4	4
1						
## 695	1	3	4	3	5	4
3						
## 743	4	4	2	3	5	5
4						
## 751	5	3	5	5	5	2
5						
## 804	1	1	3	3	2	4
1						

## 847	2	1	3	2	2	3
5						
## 849	1	2	1	2	1	2
2						
## 854	3	3	3	3	3	5
3						
## 931	3	2	4	3	3	4
2						
## 938	3	2	3	3	3	2
2						
## 941	1	1	1	2	4	5
1						
## 958	1	1	5	1	1	3
1						
## 1007	1	1	5	1	5	5
5						

##	Musical.instruments	Writing	Passive.sport	Active.sport	Gardening
## 13		4	1	5	3
## 33		2	1	2	2
## 35		3	2	5	3
## 54		4	1	5	5
## 92		1	1	5	3
## 125		5	3	4	5
## 139		5	3	2	4
## 150		1	1	2	3
## 230		4	1	5	3
## 260		2	2	5	2
## 366		3	2	4	3
## 372		5	3	2	5
## 443		5	2	3	4
## 476		5	3	3	3
## 503		3	1	2	3
## 525		1	1	5	1
## 571		2	1	5	5
## 620		1	1	2	1
## 683		1	1	4	1
## 695		5	1	4	5
## 743		3	4	3	5
## 751		1	3	2	5
## 804		4	1	4	5
## 847		2	1	2	2
## 849		2	1	5	1
## 854		3	4	5	3
## 931		4	1	3	1
## 938		5	1	5	2
## 941		5	1	5	1
## 958		1	1	2	1
## 1007		5	1	1	5
##	Celebrities	Shopping	Science.and.technology	Theatre	Fun.with.friends
## 13	3	2	3	2	4

## 33	4	5	3	3	5
## 35	1	3	5	3	5
## 54	1	3	3	5	5
## 92	3	3	1	3	5
## 125	1	1	5	4	4
## 139	2	3	2	2	4
## 150	1	2	4	2	3
## 230	1	2	4	2	4
## 260	1	1	3	5	5
## 366	1	3	2	3	4
## 372	1	3	3	3	5
## 443	4	3	4	4	5
## 476	4	2	2	3	5
## 503	1	4	5	1	5
## 525	5	3	1	4	5
## 571	3	5	4	3	4
## 620	1	1	5	1	5
## 683	2	3	4	3	4
## 695	2	2	5	2	5
## 743	1	3	4	4	5
## 751	2	5	5	5	5
## 804	3	3	1	3	5
## 847	1	5	2	2	5
## 849	1	4	3	1	5
## 854	3	4	3	5	5
## 931	3	5	3	2	5
## 938	1	3	3	4	5
## 941	1	3	2	2	5
## 958	1	2	4	2	4
## 1007	1	2	5	1	5



## 260	1	1	1	1	1	3	1	1
1								
## 366	5	1	1	1	1	1	1	1
1								
## 372	2	3	2	4	1	3	2	3
2								
## 443	2	5	2	1	2	2	2	4
3								
## 476	2	5	2	3	5	1	1	4
3								
## 503	3	1	2	1	1	2	3	3
1								
## 525	5	5	1	2	3	2	2	1
1								
## 571	4	3	1	1	1	4	1	5
4								
## 620	3	1	1	1	2	2	3	5
2								
## 683	3	5	3	1	1	2	2	5
3								
## 695	3	1	2	2	1	1	3	5
2								
## 743	4	1	1	1	2	4	2	2
3								
## 751	5	1	1	1	1	1	1	1
1								
## 804	5	3	1	1	1	2	5	5
4								
## 847	2	5	5	2	2	4	5	5
5								
## 849	4	2	1	1	1	1	1	1
1								
## 854	3	5	1	1	1	1	1	2
3								
## 931	2	1	3	3	3	4	5	5
5								
## 938	4	5	1	3	4	2	3	5
4								
## 941	3	4	1	3	4	5	3	5
1								
## 958	1	1	1	1	1	3	3	3
1								
## 1007	1	5	1	1	1	1	1	1
1								
##	Ageing	Dangerous.dogs	Fear.of.public.speaking	Healthy.eating				
Daily.events								
## 13	5	2			2		3	
4								
## 33	1	5			2		3	
3								

## 35	1	2	3	3
5				
## 54	1	3	5	3
3				
## 92	5	5	1	3
3				
## 125	1	2	1	4
4				
## 139	1	4	3	1
4				
## 150	2	3	4	2
2				
## 230	3	2	2	2
2				
## 260	1	5	1	4
4				
## 366	3	1	1	4
4				
## 372	1	5	1	1
2				
## 443	3	3	2	3
3				
## 476	1	2	3	2
3				
## 503	2	3	2	3
2				
## 525	3	2	5	2
3				
## 571	5	4	1	5
3				
## 620	4	3	1	1
3				
## 683	2	4	3	2
3				
## 695	1	2	1	4
3				
## 743	1	3	2	3
5				
## 751	1	2	1	4
5				
## 804	3	3	1	2
3				
## 847	4	5	3	3
3				
## 849	1	1	1	4
3				
## 854	3	4	1	3
3				
## 931	3	5	4	4
3				

## 938	4	3	4	3
3				
## 941	5	3	4	1
3				
## 958	1	2	3	1
1				
## 1007	2	5	5	3
1				
##	Prioritising.workload	Writing.notes	Workaholism	Thinking.ahead
## 13		3	5	4
## 33		1	1	1
## 35		3	1	5
## 54		3	5	2
## 92		3	5	5
## 125		2	2	3
## 139		2	2	1
## 150		3	2	3
## 230		1	2	1
## 260		5	3	4
## 366		3	1	3
## 372		3	3	4
## 443		3	4	3
## 476		1	2	2
## 503		4	2	5
## 525		3	2	3
## 571		3	2	3
## 620		3	5	5
## 683		1	1	2
## 695		3	4	3
## 743		2	5	2
## 751		3	4	4
## 804		1	4	1
## 847		3	4	3
## 849		3	3	4
## 854		5	5	3
## 931		2	3	2
## 938		2	4	1
## 941		2	1	1
## 958		2	1	1
## 1007		3	1	5
##	Final.judgement	Reliability	Keeping.promises	Loss.of.interest
## 13		1	5	5
## 33		1	1	1
## 35		5	4	4
## 54		3	5	5
## 92		5	3	5
## 125		2	4	4
## 139		2	5	5
## 150		3	2	2
## 230		3	3	5

## 260	4	3	3	1	
## 366	5	4	5	2	
## 372	3	4	5	1	
## 443	3	4	5	2	
## 476	3	2	4	2	
## 503	3	5	5	2	
## 525	3	4	5	1	
## 571	1	4	4	5	
## 620	1	4	3	1	
## 683	3	1	2	3	
## 695	5	5	4	2	
## 743	5	2	5	5	
## 751	2	4	4	2	
## 804	5	5	5	2	
## 847	5	5	3	3	
## 849	3	3	1	1	
## 854	3	4	4	2	
## 931	3	4	3	1	
## 938	3	4	5	2	
## 941	5	1	1	4	
## 958	4	3	3	2	
## 1007	5	5	4	1	
##	Friends.versus.money	Funniness	Fake	Criminal.damage	Decision.making
## 13	4	3	1	2	5
## 33	1	1	1	3	1
## 35	5	5	1	1	3
## 54	5	1	1	1	3
## 92	5	3	1	1	5
## 125	2	3	1	1	2
## 139	4	3	1	1	4
## 150	5	3	3	3	2
## 230	4	3	3	3	3
## 260	5	3	1	1	3
## 366	5	4	2	2	2
## 372	4	1	1	1	2
## 443	5	3	1	2	2
## 476	4	5	2	1	2
## 503	2	4	2	1	4
## 525	4	4	2	3	3
## 571	1	3	2	5	4
## 620	3	3	2	3	4
## 683	2	3	3	2	1
## 695	5	5	1	2	3
## 743	5	5	2	3	2
## 751	5	2	2	2	3
## 804	3	3	1	5	3
## 847	5	4	2	5	3
## 849	4	3	2	4	2
## 854	3	3	1	1	3
## 931	4	4	2	5	4

## 938	4	2	1	1	3
## 941	4	5	2	5	2
## 958	4	3	1	4	4
## 1007	2	1	1	3	5

##	Elections	Self.criticism	Judgment.calls	Hypochondria	Empathy
----	-----------	----------------	----------------	--------------	---------

## 13	5	5	2	5	5
## 33	1	1	5	1	5
## 35	5	4	3	1	2
## 54	5	3	5	1	5
## 92	1	3	3	3	5
## 125	3	5	3	1	3
## 139	5	4	5	1	5
## 150	3	1	2	1	2
## 230	4	4	4	1	5
## 260	4	5	5	5	5
## 366	5	2	4	1	2
## 372	5	1	5	1	4
## 443	1	4	4	1	4
## 476	5	4	4	2	5
## 503	5	2	5	2	4
## 525	5	5	4	2	4
## 571	3	2	5	1	3
## 620	1	3	4	1	3
## 683	1	1	1	5	4
## 695	1	5	5	1	5
## 743	2	2	5	2	5
## 751	5	3	4	2	5
## 804	3	3	2	3	5
## 847	2	2	2	3	3
## 849	3	5	4	1	2
## 854	3	4	3	1	3
## 931	1	3	5	3	5
## 938	4	2	5	4	4
## 941	1	2	4	2	4
## 958	5	4	3	2	5
## 1007	5	5	5	1	2

##	Eating.to.survive	Giving	Compassion.to.animals	Borrowed.stuff	Loneliness
----	-------------------	--------	-----------------------	----------------	------------

## 13	3	2	5	5
5				
## 33	1	1	5	5
1				
## 35	2	3	1	3
2				
## 54	5	5	3	5
3				
## 92	3	3	1	2
2				
## 125	2	4	2	4
2				

## 139	2	1	2	5
4				
## 150	3	2	1	2
3				
## 230	5	3	3	4
2				
## 260	1	1	5	5
3				
## 366	2	3	4	4
2				
## 372	3	3	4	4
1				
## 443	1	3	4	4
4				
## 476	2	2	5	4
3				
## 503	1	3	2	4
2				
## 525	1	1	5	3
3				
## 571	1	3	3	3
3				
## 620	2	3	5	4
2				
## 683	2	2	4	4
3				
## 695	4	1	3	5
5				
## 743	4	1	4	2
1				
## 751	3	4	3	4
2				
## 804	1	5	5	5
3				
## 847	4	3	5	5
4				
## 849	1	2	3	1
1				
## 854	3	5	5	5
3				
## 931	2	5	2	4
5				
## 938	2	1	4	5
5				
## 941	2	2	3	2
2				
## 958	3	1	5	4
3				
## 1007	1	4	3	4
1				

##	Cheating.in.school	Health	Changing.the.past	God	Dreams	Charity
## 13	2	5	3	1	3	2
## 33	5	1	1	3	3	1
## 35	2	2	4	5	4	4
## 54	5	3	5	5	3	1
## 92	5	3	3	5	3	2
## 125	2	2	2	1	4	4
## 139	4	3	5	5	3	3
## 150	2	2	1	4	3	2
## 230	5	2	2	5	3	1
## 260	3	5	5	5	4	3
## 366	3	1	3	5	4	3
## 372	2	3	1	5	4	3
## 443	5	3	2	5	4	2
## 476	5	3	2	5	3	2
## 503	5	2	2	4	4	1
## 525	4	3	4	5	3	1
## 571	5	4	4	3	4	3
## 620	3	5	2	1	3	1
## 683	3	2	4	5	3	2
## 695	4	3	3	5	2	2
## 743	5	1	2	5	4	3
## 751	3	3	2	5	3	5
## 804	5	5	5	5	3	3
## 847	4	4	5	5	3	2
## 849	5	3	1	5	4	1
## 854	3	5	3	5	4	3
## 931	4	4	3	4	3	2
## 938	4	4	1	3	4	2
## 941	5	4	5	4	2	1
## 958	5	3	3	4	3	1
## 1007	1	3	4	5	3	3
##	Number.of.friends	Waiting	New.environment	Mood.swings		
## 13	3	2	1		4	
## 33	5	2	5		2	
## 35	4	4	3		2	
## 54	5	1	4		2	
## 92	3	1	4		4	
## 125	4	4	4		2	
## 139	4	5	5		2	
## 150	2	2	4		3	
## 230	4	3	4		2	
## 260	4	1	5		2	
## 366	3	4	4		2	
## 372	5	3	4		2	
## 443	3	4	4		3	
## 476	5	3	2		3	
## 503	4	3	5		1	
## 525	2	1	3		3	
## 571	5	3	5		3	

## 620	4	3	3	3
## 683	3	3	5	4
## 695	4	2	5	4
## 743	5	2	5	4
## 751	3	3	4	2
## 804	3	2	5	2
## 847	2	2	4	4
## 849	5	5	4	2
## 854	5	4	4	4
## 931	2	3	4	5
## 938	1	2	4	4
## 941	4	2	4	4
## 958	2	3	3	3
## 1007	3	5	5	1
##	Appearance.and.gestures	Socializing	Achievements	
## 13		4	2	3
## 33		2	4	2
## 35		3	4	3
## 54		1	5	2
## 92		3	5	3
## 125		3	2	3
## 139		3	5	3
## 150		2	3	4
## 230		4	4	3
## 260		2	4	3
## 366		4	3	3
## 372		4	5	3
## 443		4	3	3
## 476		2	1	4
## 503		3	3	4
## 525		2	3	1
## 571		5	4	3
## 620		4	3	3
## 683		2	3	4
## 695		4	5	3
## 743		1	5	4
## 751		4	3	2
## 804		4	3	3
## 847		3	2	4
## 849		5	3	2
## 854		4	4	4
## 931		4	3	3
## 938		2	3	1
## 941		4	1	5
## 958		3	3	3
## 1007		3	3	3
##	Responding.to.a.serious.letter	Children	Assertiveness	Getting.angry
## 13		4	5	5
## 33		3	3	4
## 35		3	3	4



## 54	1	5	5	2
## 92	3	4	4	5
## 125	1	4	4	1
## 139	5	5	4	1
## 150	3	5	2	2
## 230	4	4	4	1
## 260	2	3	3	3
## 366	3	4	4	1
## 372	2	5	4	1
## 443	2	5	5	3
## 476	4	4	4	2
## 503	2	3	5	1
## 525	4	3	2	2
## 571	3	5	4	4
## 620	2	3	5	3
## 683	4	5	3	3
## 695	5	4	3	1
## 743	3	4	4	2
## 751	1	5	3	3
## 804	1	4	3	3
## 847	2	5	4	3
## 849	4	4	4	1
## 854	2	5	3	3
## 931	2	4	3	4
## 938	2	5	3	4
## 941	3	3	4	3
## 958	4	3	2	4
## 1007	5	5	4	1
##	Knowing.the.right.people	Public.speaking	Unpopularity	Life.struggles
## 13	3	3	2	5
## 33	3	4	3	3
## 35	4	4	4	1
## 54	2	2	5	2
## 92	5	5	5	2
## 125	3	2	4	1
## 139	2	5	5	5
## 150	3	5	4	2
## 230	3	4	3	2
## 260	3	3	3	2
## 366	1	1	3	2
## 372	3	2	3	2
## 443	4	4	4	4
## 476	3	3	3	4
## 503	5	2	2	1
## 525	4	4	4	3
## 571	3	5	5	3
## 620	5	1	3	1
## 683	3	3	4	1
## 695	4	1	3	1
## 743	4	2	2	1

## 751	5	1	3	2
## 804	3	3	3	3
## 847	3	5	5	4
## 849	5	2	5	2
## 854	4	4	3	5
## 931	3	4	4	5
## 938	4	3	5	5
## 941	2	2	4	4
## 958	2	5	4	3
## 1007	2	5	1	1

##	Happiness.in.life	Energy.levels	Small...big.dogs	Personality
----	-------------------	---------------	------------------	-------------

## 13	4	3	4	3
## 33	4	5	1	4
## 35	3	5	3	3
## 54	4	5	4	3
## 92	5	5	1	3
## 125	3	4	5	3
## 139	4	4	3	3
## 150	3	2	3	2
## 230	5	4	3	3
## 260	4	5	3	3
## 366	4	4	4	3
## 372	5	4	2	4
## 443	4	3	5	3
## 476	4	3	3	3
## 503	4	4	3	4
## 525	3	2	3	3
## 571	4	4	3	3
## 620	4	4	3	4
## 683	4	3	3	3
## 695	4	4	5	3
## 743	4	5	5	4
## 751	4	4	5	3
## 804	3	4	3	4
## 847	4	3	1	3
## 849	5	5	3	5
## 854	3	4	3	3
## 931	3	3	3	3
## 938	4	3	3	2
## 941	4	4	4	3
## 958	3	2	4	3
## 1007	4	4	3	5

##	Finding.lost.valuables	Getting.up	Interests.or.hobbies
Parents..advice			

## 13	1	3	2
4			
## 33	1	5	5
3			
## 35	5	2	3
2			

## 54	3	5	5
4			
## 92	3	5	4
4			
## 125	3	3	4
3			
## 139	4	3	3
5			
## 150	3	2	3
3			
## 230	1	3	4
4			
## 260	5	3	5
3			
## 366	2	2	4
3			
## 372	3	3	5
3			
## 443	3	2	4
3			
## 476	3	5	5
3			
## 503	1	2	4
4			
## 525	3	3	3
4			
## 571	1	3	3
3			
## 620	3	3	1
3			
## 683	3	5	2
3			
## 695	5	5	5
3			
## 743	3	5	4
1			
## 751	1	3	3
3			
## 804	2	5	4
4			
## 847	5	4	4
4			
## 849	1	4	5
2			
## 854	4	1	5
5			
## 931	4	3	4
3			
## 938	2	5	5
2			

## 941	2	5	3
4			
## 958	4	4	2
4			
## 1007	3	1	3
4			
##	Questionnaires.or.polls	Finances	Shopping.centres
## 13	2	4	1
## 33	3	1	3
## 35	2	3	3
## 54	4	3	5
## 92	3	5	5
## 125	2	3	3
## 139	5	4	1
## 150	3	3	2
## 230	2	1	2
## 260	5	4	1
## 366	1	3	3
## 372	3	3	3
## 443	1	2	4
## 476	2	1	2
## 503	3	2	5
## 525	5	4	2
## 571	1	3	5
## 620	1	2	2
## 683	3	1	4
## 695	3	3	2
## 743	3	1	4
## 751	3	4	3
## 804	1	3	3
## 847	5	4	4
## 849	2	3	4
## 854	3	4	2
## 931	3	3	5
## 938	4	5	2
## 941	3	2	5
## 958	3	3	2
## 1007	3	3	1
##	Entertainment.spending	Spending.on.looks	Spending.on.gadgets
## 13	2	3	2
## 33	4	5	1
## 35	3	3	1
## 54	3	1	2
## 92	5	3	2
## 125	3	4	3
## 139	1	1	1
## 150	2	1	2
## 230	4	1	2
## 260	4	3	1
## 366	3	4	2

## 372	3		3		4
## 443	4		5		3
## 476	5		2		1
## 503	4		4		5
## 525	3		2		2
## 571	3		5		5
## 620	3		2		4
## 683	4		4		4
## 695	3		2		4
## 743	5		4		4
## 751	3		3		3
## 804	3		3		3
## 847	4		4		4
## 849	5		5		5
## 854	1		3		2
## 931	2		3		2
## 938	4		2		3
## 941	4		4		4
## 958	3		3		3
## 1007	1		1		5
##	Spending.on.healthy.eating	Age	Height	Weight	Number.of.siblings
## 13	3	24	168	55	10
## 33	1	20	158	46	4
## 35	2	20	177	67	5
## 54	1	18	164	51	4
## 92	2	18	168	50	4
## 125	5	20	176	69	4
## 139	1	20	168	50	4
## 150	4	19	184	85	5
## 230	2	21	173	73	5
## 260	3	19	175	72	4
## 366	4	24	175	70	5
## 372	3	19	164	53	4
## 443	3	18	163	58	6
## 476	3	18	172	60	4
## 503	5	22	193	86	6
## 525	2	19	171	60	4
## 571	5	23	189	88	4
## 620	5	27	175	83	4
## 683	1	19	172	75	4
## 695	4	21	187	73	4
## 743	5	26	183	90	5
## 751	3	22	175	86	6
## 804	3	27	173	79	5
## 847	4	23	171	54	4
## 849	5	19	184	60	4
## 854	2	30	175	68	4
## 931	4	18	165	45	5
## 938	5	18	167	62	4
## 941	4	16	172	59	4

```
## 958          3  20    175    65          5
## 1007         3  27    183    80          5

# Function for outlier treatment
# Capping and Flooring function
treat_outlier <- function(x){
  qnt <- quantile(x, probs=c(.25, .75), na.rm = T)
  caps <- quantile(x, probs=c(.05, .95), na.rm = T)
  H <- 1.5 * IQR(x, na.rm = T)
  x[x < (qnt[1] - H)] <- caps[1]
  x[x > (qnt[2] + H)] <- caps[2]
  return(as.data.frame(x))
}

impu_num_compl$Age <- treat_outlier(impu_num_compl$Age)$x
impu_num_compl$Height <- treat_outlier(impu_num_compl$Height)$x
impu_num_compl$Weight <- treat_outlier(impu_num_compl$Weight)$x

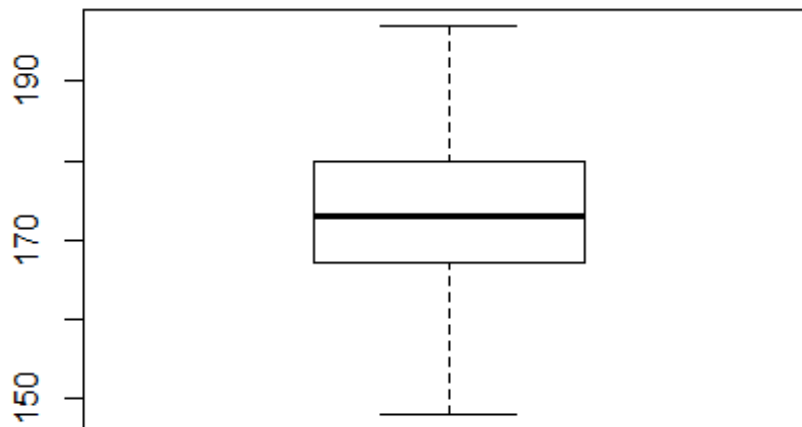
dim(impu_num_compl)

## [1] 1010  139

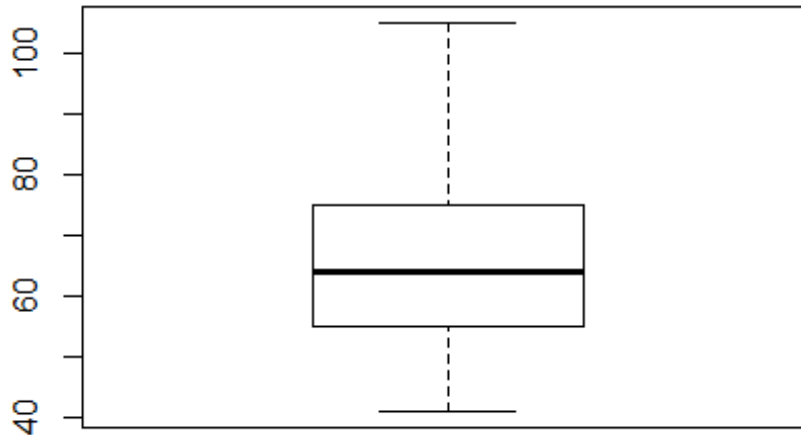
dim(impu_cat_compl)

## [1] 1010   11

boxplot(impu_num_compl$Height)
```



```
boxplot(impu_num_compl$Weight)
```



```
data_transformed = cbind(impu_num_compl, impu_cat_compl)
dim(data_transformed)
```

```
## [1] 1010 150
```

Corelation Analysis:

```
data_num_trans <- data_transformed[,!sapply(data_transformed, is.factor)]
dim(data_num_trans)
```

```
## [1] 1010 139
```

```
music_trans = data_transformed[,names(music_data)]
dim(music_trans)
```

```
## [1] 1010 19
```

```
library(corrgram)
```

```
## Warning: package 'corrgram' was built under R version 3.6.2
```

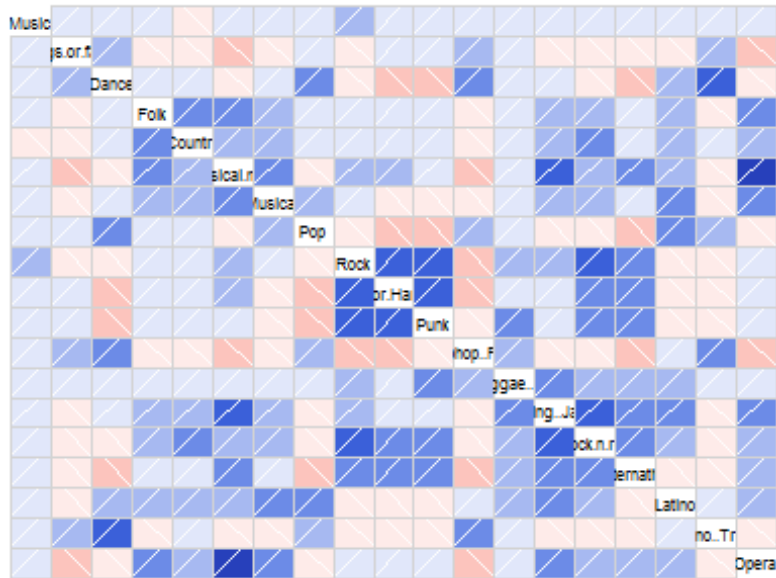
```
## Registered S3 method overwritten by 'seriation':
```

```
##   method      from
## reorder.hclust gclus
```

```
##
```

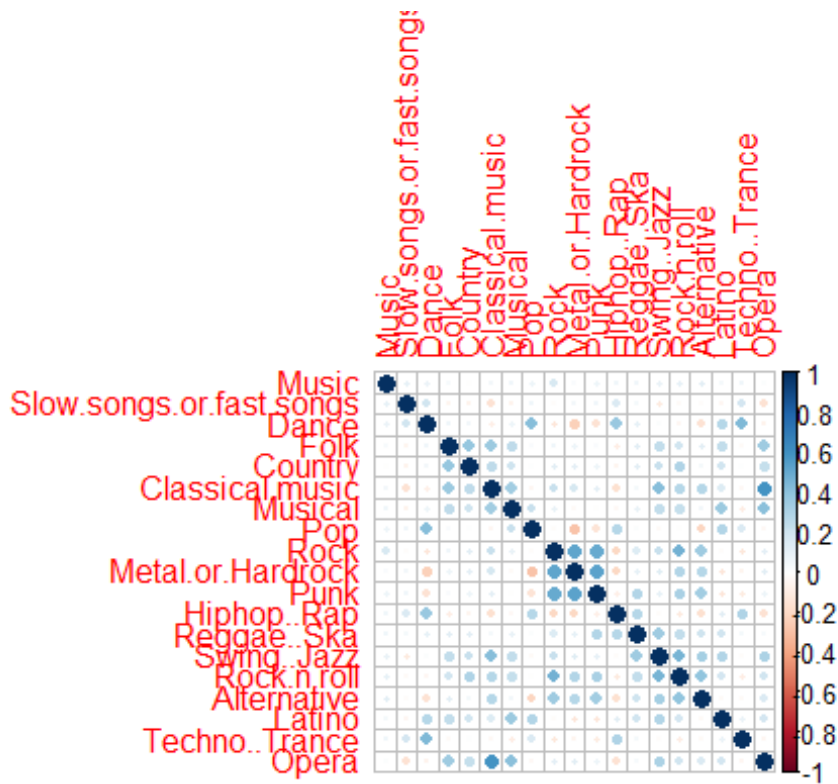
```
## Attaching package: 'corrgram'
```

```
## The following object is masked from 'package:lattice':
##
##      panel.fill
corrgram(music_trans)
```



```
library(corrplot)
## Warning: package 'corrplot' was built under R version 3.6.2
## corrplot 0.84 loaded
music_cor = cor(music_trans)
corrplot(music_cor)
```

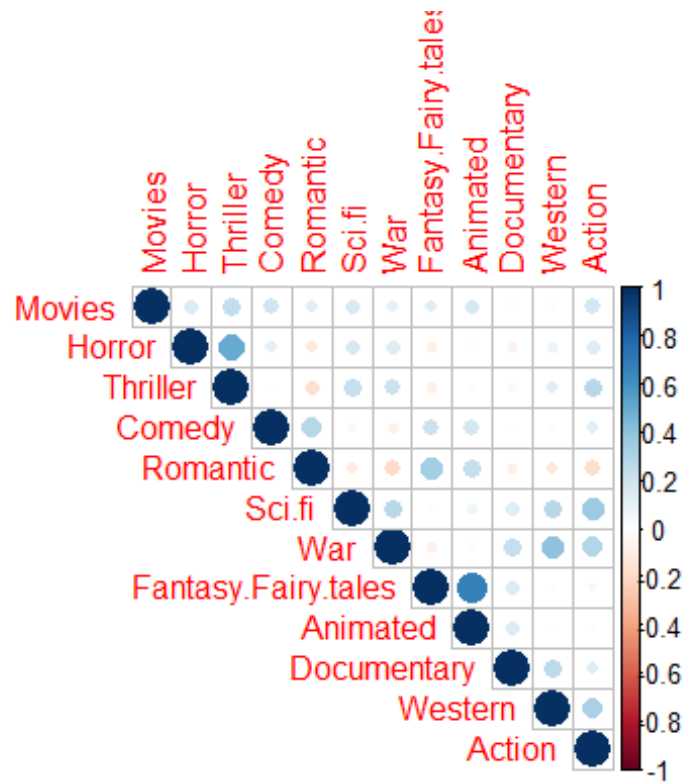




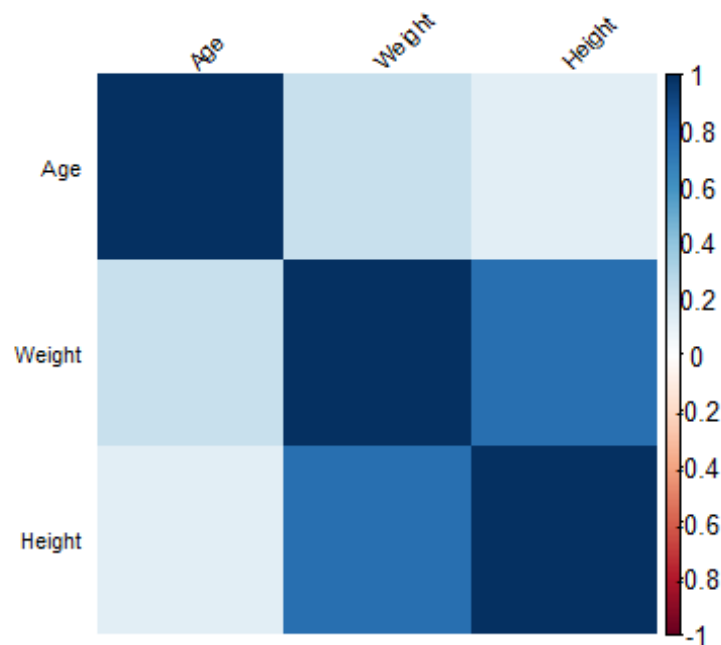
```
movie_trans = data_transformed[,names(movie_data)]
dim(movie_trans)

## [1] 1010 12

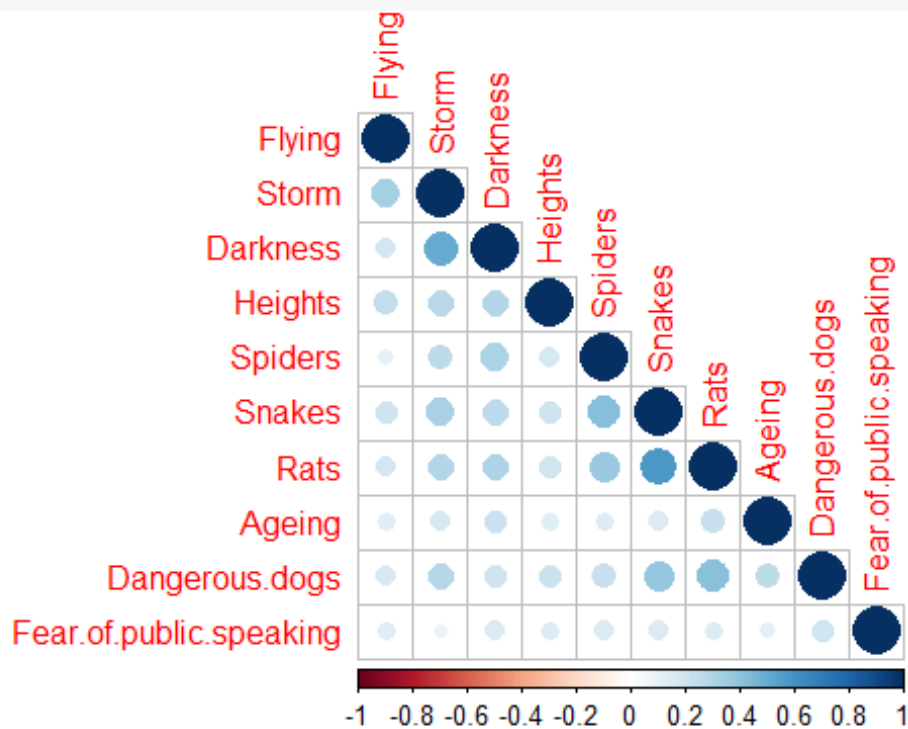
movie_cor = cor(movie_trans)
corrplot(movie_cor,type="upper")
```



```
demo_cor <- cor(data_transformed[,c('Age', 'Weight', 'Height')])
corrplot(demo_cor, method="shade", shade.col=NA, tl.col="black", tl.srt=45,
tl.cex =0.7)
```



```
phobia_cor = cor(data_transformed[,names(phobia_data)])
corrplot(phobia_cor,type = "lower")
```



```
spend_cor = cor(data_transformed[,names(spend_data)])
corrplot(spend_cor)
```



## Exploratory Data Analysis

data\_transformed\$Gender

```
## [1] female female female female female male female male female
female
## [11] female female female female female male female female male male
## [21] male male female female female female female female male
female
## [31] male female female female male female female male female
female
## [41] female male female female female male female female female
female
## [51] female female female female female female female female male male
## [61] male male male male female male male female female
female
## [71] male male female male male male male female female
female
## [81] female male female female female male male female female
female
## [91] female female female male male male male male male
female
## [101] female male female male male male male female female male
## [111] male female female female female female female female male
female
## [121] male male female female male female female female female
female
## [131] female male female female female male male female
female
## [141] female female male female male female male male male male male
## [151] female male male female male female male male male female male
## [161] male female male male male female female male female
female
## [171] male male male male male female male male female male
## [181] male male female male female female female female female
female
## [191] female female female male male male female female male male
## [201] male male female female female female male female female male
## [211] male female male female female male female female male male
## [221] male male female female male male female female male
## [231] female female female female female female female female female
female
## [241] male female female female female female male female female
female
## [251] female female female female female female female male female male
## [261] male female female female female female female female female
female
## [271] female male female female female female female male male
female
## [281] female male female male male female female female male male
```

```

## [291] female male    female female female female male    male    female
female
## [301] female male    female female female female female female female
female
## [311] male    female male    female male    male    female male    female
female
## [321] male    female female male    female male    female female male
female
## [331] female male    female female female male    male    male    female
female
## [341] female male    female female female female male    male    male
female
## [351] male    female male    female female male    male    male    male
female
## [361] female female female female male    male    male    male    female male
## [371] female female male    male    female female female female female male
## [381] female female female male    female male    female female female
female
## [391] male    female female male    male    male    female female female
female
## [401] male    female male    male    female female female female female
female
## [411] male    female female male    male    female female male    female
female
## [421] female female female female female female female male    female
female
## [431] female male    male    female male    female male    female female male
## [441] female female female female female female male    female female
female
## [451] male    male    female female female female male    male    female
female
## [461] male    male    female female female female male    female female male
## [471] female male    male    male    male    female female female male
female
## [481] female male    male    female female female female female male
female
## [491] female female female male    female female female female female male
## [501] female female male    female female male    male    female male
female
## [511] male    male    female male    male    male    male    female female male
## [521] female female female male    female female male    female male    male
## [531] female male    female male    female male    female male    female male
## [541] male    female    male    male    male    male    male    female
female
## [551] female female female female male    female female female female
female
## [561] female male    male    male    male    female female male    female
female
## [571] male    male    female male    female female male    female female male
## [581] male    female male    male    female male    male    female female

```

female

## [591] male female female female male male male male male male male

## [601] female female male female female male male female male male

## [611] male female male female female male male female female male

## [621] male female male female female female female female male

female

## [631] female female male male female female male male female male

## [641] female male male female female male male female female

female

## [651] male male male female male female female female female male male

## [661] female female female male female female female male male male

## [671] female female female female male female female female female

female

## [681] female female male female female female male male female male

## [691] male female male female male female female female female

female

## [701] female female male male female female male female male

female

## [711] female female female female female male male female female male

## [721] female female female female female male male male female male

## [731] female female female female female female female female male male

## [741] male female male male male male female female female male

## [751] female female female female male female male female male male

## [761] female male male male female male female male male

female

## [771] female male female female female female male male female

female

## [781] male female female male male male male male female male

## [791] female female male male male female male male female male

## [801] male female female female female female male male male male

## [811] male male female male female male male male male

female

## [821] female male female male female female female female male

female

## [831] female female female male male male male male female

female

## [841] female male female female male female female female male

female

## [851] female male male female male male female male male male

## [861] male female male male male female male female male

female

## [871] female male male male female female male male male male

## [881] female female female female female female male male female

female

## [891] male male male female female male male male female male

## [901] male female male male female male female female male

female

## [911] female female female female female female male male female

female

## [921] female male female female female female male female female

```

female
## [931] female male   female female           female female female female
female
## [941] male   female male   female female male   male   female female
female
## [951] male   male   male   female female female male   male   female
female
## [961] female male   female male   male   female female male   male   male
## [971] male   male   female female female female female male   female male
## [981] female female male   female female female female male   male
female
## [991] male   male   male   female female female male   female male   male
## [1001] female female male   female male   female male   female female male
## Levels:  female male

a = split(data_transformed$Gender, 3, drop = TRUE)

str(data_transformed$Gender)

## Factor w/ 3 levels "", "female", "male": 2 2 2 2 2 3 2 3 2 2 ...

library(ggplot2)
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.6.2

library(magrittr) # needs to be run every time you start R and want to use
%>%
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##   combine

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(reshape2)

## Warning: package 'reshape2' was built under R version 3.6.2

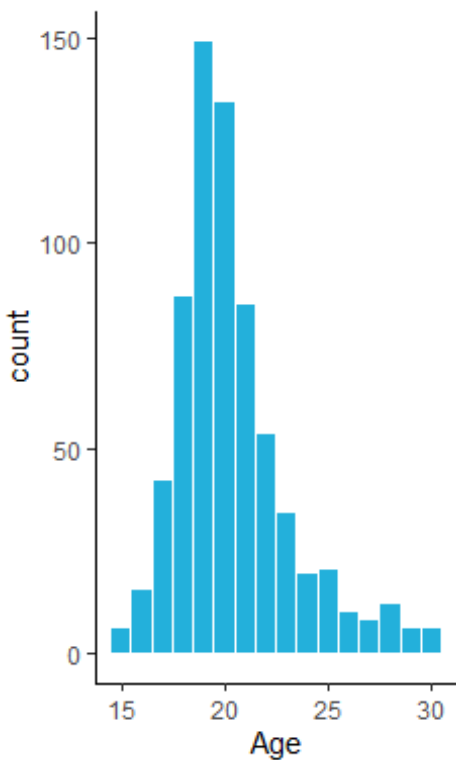
p1 = ggplot(na.omit(data), aes(x=Age)) + geom_bar(fill = "#23b0db") +
theme_bw() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),

```



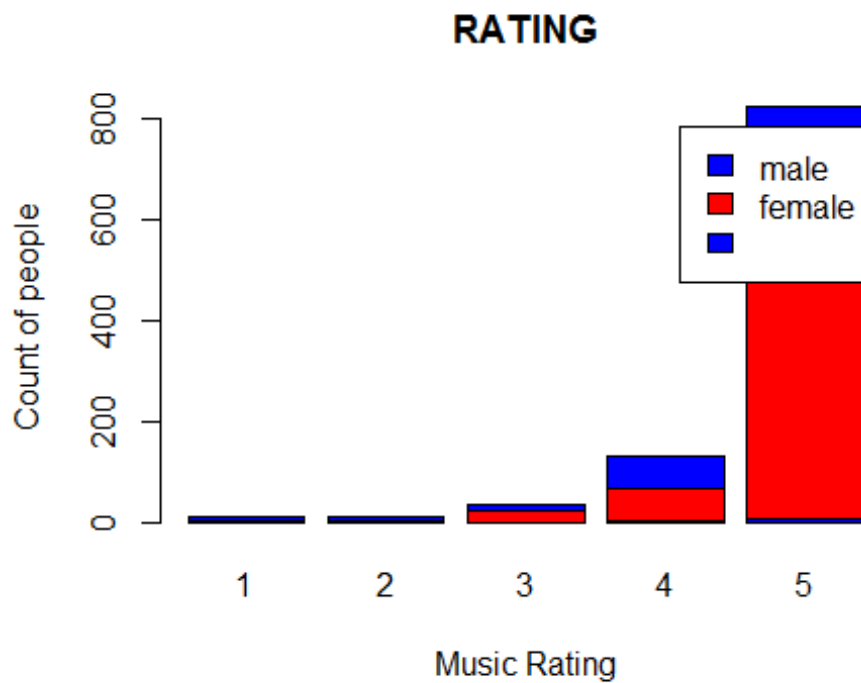
```
panel.grid.minor = element_blank(), axis.line = element_line(colour =
"black"))
```

```
age = data %>% select(Age) %>% group_by(Age) %>% summarize(count = n()) %>%
arrange(desc(count))
age = tableGrob(as.data.frame(age))
grid.arrange(p1, head(age, 7), ncol=2)
```

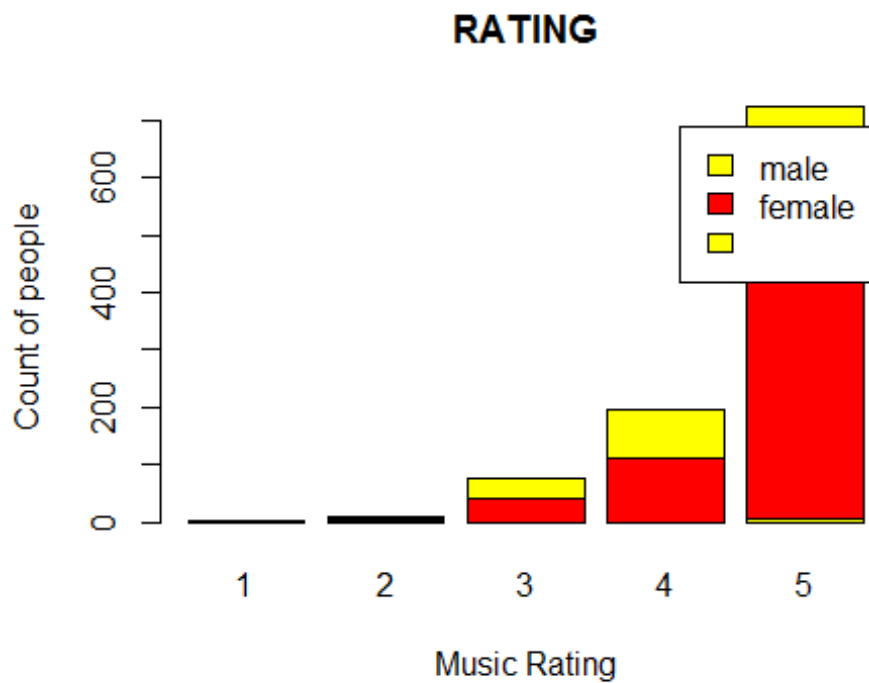


	Age	count
1	19	210
2	20	194
3	21	127
4	18	123
5	22	84
6	17	53

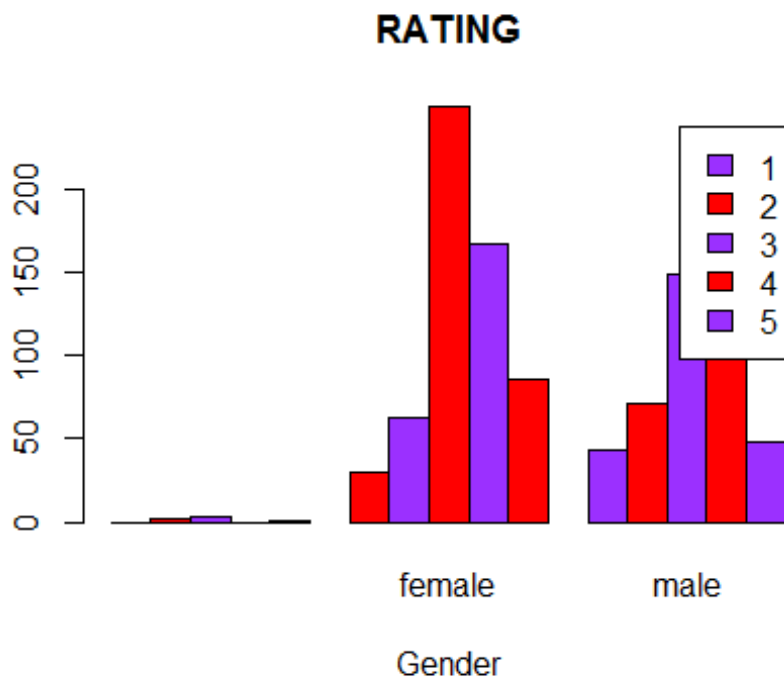
```
counts <- table(data_transformed$Gender, data_transformed$Music)
barplot(counts, main="RATING",
        xlab="Music Rating", ylab = "Count of people", col=c("blue","red"),
        legend = rownames(counts))
```



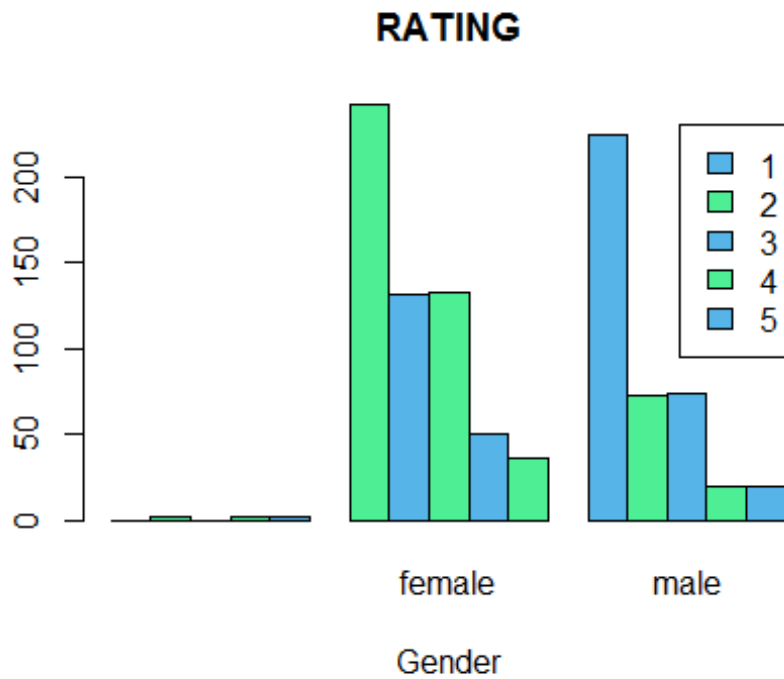
```
counts <- table(data_transformed$Gender, data_transformed$Movies)
barplot(counts, main="RATING",
        xlab="Music Rating", ylab = "Count of people", col=c("Yellow","red"),
        legend = rownames(counts))
```



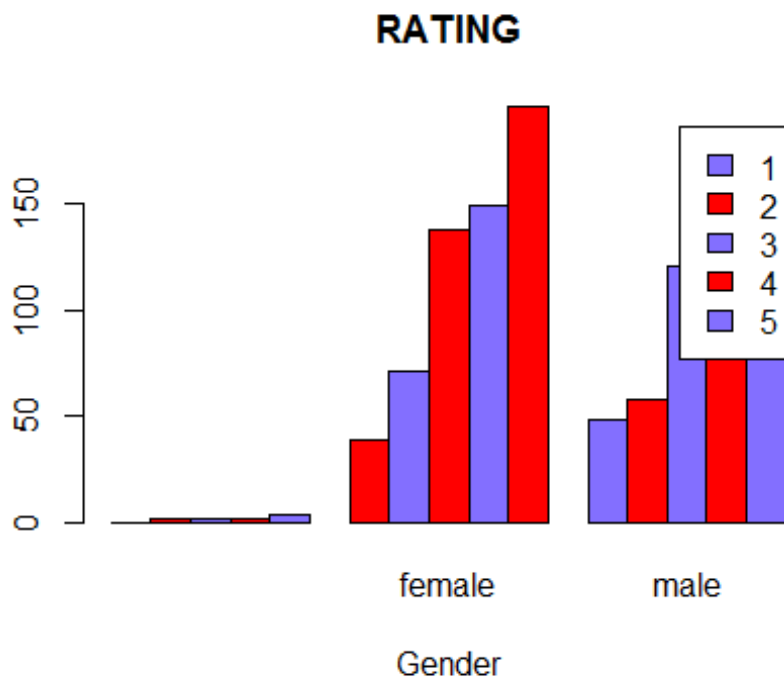
```
#Health
counts <- table(data_transformed$Health, data_transformed$Gender)
barplot(counts, main="RATING",
        xlab="Gender", col=c("purple1", "red"),
        legend = rownames(counts), beside=TRUE)
```



```
#Flying_----- Phobias
counts <- table(data_transformed$Flying, data_transformed$Gender)
barplot(counts, main="RATING",
        xlab="Gender", col=c("#56B4E9","seagreen2"),
        legend = rownames(counts), beside=TRUE)
```



```
#Public Speaking-----Phobias
counts <- table(data_transformed$Public.speaking, data_transformed$Gender)
barplot(counts, main="RATING",
        xlab="Gender", col=c("slateblue1","red"),
        legend = rownames(counts), beside=TRUE)
```



```
#Left/Right Handed
counts <- table(data_transformed$Left...right.handed,
data_transformed$Gender)
barplot(counts, main="RATING",
        xlab="Gender", col=c("deepskyblue1","red"),
        legend = rownames(counts), beside=TRUE)
```

## RATING

