

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:

```
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 numeric 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 categoric 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

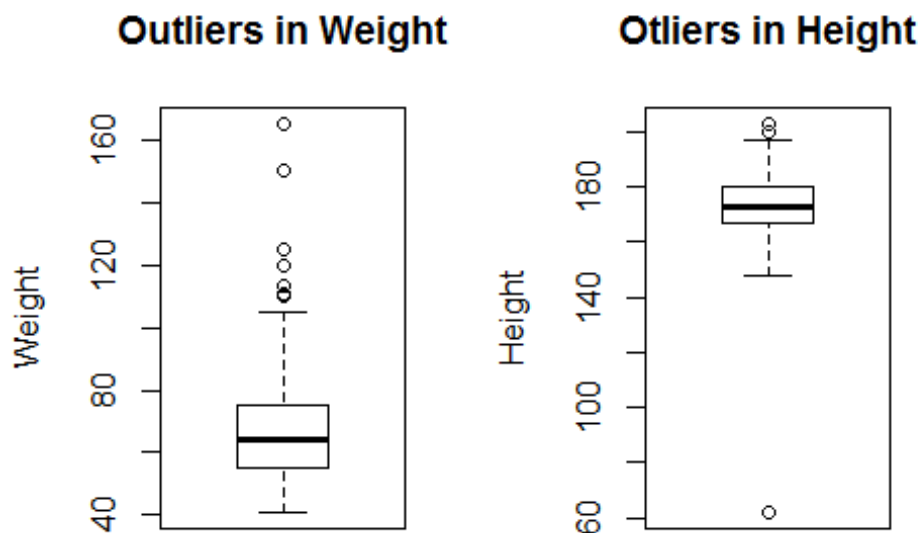
id4 = boxplot.stats(impu_num_compl$Number.of.siblings)

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')

```

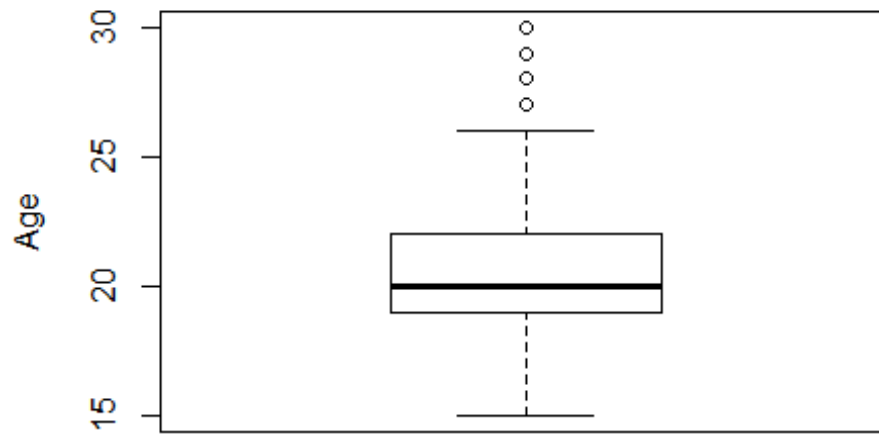


```

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

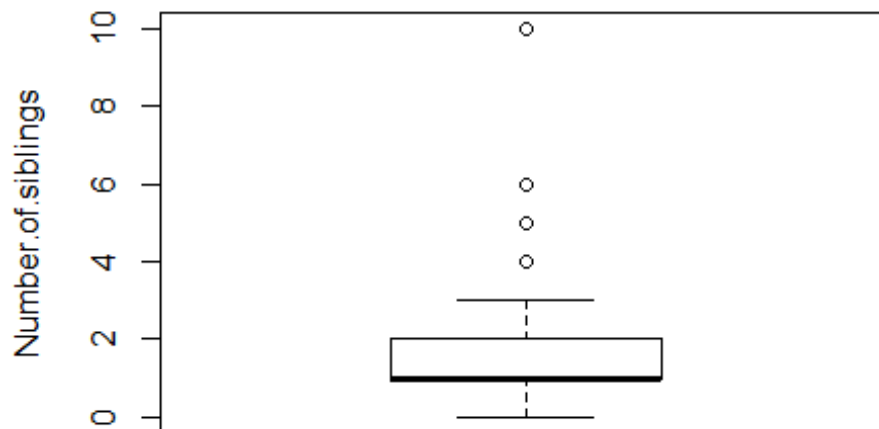
```

Outliers in Age



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

Otliers in 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
```

```
1
```

```
## 221      5              3      4      3      3              4
```

```
2
```

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

```
3
```

```
## 677      5              4      2      2      1              2
```

```
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
```

```
3
```

```
## 221              2              5      3              1      3      5      3
```

```
3
```

```
## 548              4              3      1              1      5      5      3
```

```
5
```

```
## 677              1              5      2              2      2      5      1
```

```
2
```

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

```
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
3								
## 221	5	3	2	3		4	5	4
3								
## 548	5	2	2	4		4	4	5
3								
## 677	2	3	1	2		3	4	5
1								
## 800	4	3	4	4		3	4	5
3								
## 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	
4								
## 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		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
1								4
## 221		3	5	1	3	2	2	2
2								3
## 548		5	4	1	1	2	3	1
1								1
## 677		2	3	1	2	2	2	3
4								3
## 800		3	5	2	1	3	2	3
2								3
## 993		1	1	3	1	1	5	5
2								3
##	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
## 221	4	4	2	5
## 548	5	5	1	1
## 677	4	2	2	5
## 800	4	3	3	2
## 993	1	2	1	1
##	Elections	Self.criticism	Judgment.calls	Hypochondria
## 98	5	4	3	1
## 221	4	2	4	1
## 548	5	5	5	1
## 677	1	2	4	2
## 800	2	4	4	2
## 993	5	2	3	1
##	Eating.to.survive	Giving	Compassion.to.animals	Borrowed.stuff
## 98	4	1	2	5
## 221	2	3	5	5
## 548	1	1	5	5
## 677	1	2	1	2
## 800	2	3	5	4
## 993	1	4	2	4
##	Cheating.in.school	Health	Changing.the.past	God Dreams
## 98	5	1	5	5
## 221	5	3	1	2
## 548	5	2	2	1
## 677	5	3	4	4
## 800	3	3	2	2
## 993	1	3	4	5
##	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
##	Appearance.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
## 98	2	19	200	75
## 221	3	18	200	90
## 548	4	21	203	80
## 677	4	20	62	55
				Number.of.siblings
## 98				1
## 221				2
## 548				2
## 677				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
##	Rock.n.roll Alternative Latino Techno..Trance Opera Movies Horror Thriller							
## 13		4		3	2		1	2
5								
## 33		4		4	2		3	1
5								
## 35		4		4	2		1	1
3								
## 54		3		3	5		2	5
3								
## 92		3		1	1		1	1
4								
## 125		5		2	3		1	3
5								
## 139		4		2	3		1	2
2								
## 150		1		1	1		1	2
2								
## 230		3		2	2		3	3
3								
## 260		5		2	5		1	5
1								
## 366		5		5	1		2	4
5								
## 372		5		3	5		2	5
2								
## 443		4		2	2		2	1
4								
## 476		5		3	4		1	2
3								
## 503		5		3	1		3	3
5								
## 525		3		5	2		1	3
4								
## 571		3		2	2		1	1
2								
## 620		3		2	3		2	5
4								
## 683		4		3	2		5	2
4								
## 695		5		1	5		4	3
5								
## 743		3		4	5		5	2
3								
## 751		4		5	5		1	4
4								
## 804		3		2	1		1	2

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		3	2	2	4	4		
4								
## 139		2	5	4	5	3		
5								
## 150		3	2	3	2	3		
3								
## 230		2	3	2	1	5		
3								
## 260		2	4	3	5	4		
4								
## 366		4	4	4	4	5		
5								
## 372		3	5	5	5	2		
3								
## 443		2	3	3	2	3		
3								
## 476		2	3	2	5	3		
3								
## 503		4	1	1	1	1		
4								
## 525		1	5	4	3	4		
3								
## 571		2	2	1	1	4		
5								
## 620		2	3	2	1	5		

4						
## 683		2	2	1	3	4
3						
## 695		3	2	2	2	4
5						
## 743		5	4	1	5	4
5						
## 751		2	3	2	2	2
2						
## 804		2	2	1	2	2
3						
## 847		3	1	1	3	2
5						
## 849		5	2	2	1	2
3						
## 854		3	3	3	4	4
3						
## 931		3	2	1	1	3
3						
## 938		2	4	2	5	5
4						
## 941		1	2	1	2	1
3						
## 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	4
## 33		2	1	2	2	1
## 35		3	2	5	3	1
## 54		4	1	5	5	3
## 92		1	1	5	3	1
## 125		5	3	4	5	1
## 139		5	3	2	4	1
## 150		1	1	2	3	1

## 230	4	1	5	3	2
## 260	2	2	5	2	1
## 366	3	2	4	3	1
## 372	5	3	2	5	3
## 443	5	2	3	4	2
## 476	5	3	3	3	1
## 503	3	1	2	3	1
## 525	1	1	5	1	1
## 571	2	1	5	5	1
## 620	1	1	2	1	1
## 683	1	1	4	1	1
## 695	5	1	4	5	1
## 743	3	4	3	5	1
## 751	1	3	2	5	4
## 804	4	1	4	5	3
## 847	2	1	2	2	2
## 849	2	1	5	1	2
## 854	3	4	5	3	4
## 931	4	1	3	1	1
## 938	5	1	5	2	1
## 941	5	1	5	1	1
## 958	1	1	2	1	1
## 1007	5	1	1	5	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
##	Adrenaline.sports	Pets	Flying	Storm	Darkness	Heights	Spiders	Snakes
Rats								
## 13		1	2	1	1	1	1	3
1								
## 33		1	5	3	5	5	5	5
5								
## 35		5	1	1	2	2	1	2
2								
## 54		3	1	3	1	1	5	1
3								
## 92		4	1	1	3	1	1	5
5								
## 125		5	4	1	1	1	2	1
1								
## 139		1	3	1	1	1	1	2
2								
## 150		3	1	1	1	2	3	2
1								
## 230		5	1	1	1	1	1	2
1								
## 260		1	1	1	1	1	3	1
1								
## 366		5	1	1	1	1	1	1
1								
## 372		2	3	2	4	1	3	3
2								
## 443		2	5	2	1	2	2	4
3								
## 476		2	5	2	3	5	1	4
3								
## 503		3	1	2	1	1	2	3
1								
## 525		5	5	1	2	3	2	1
1								
## 571		4	3	1	1	1	4	5
4								
## 620		3	1	1	1	2	2	5
2								
## 683		3	5	3	1	1	2	5
3								
## 695		3	1	2	2	1	1	5
2								
## 743		4	1	1	1	2	4	2
3								
## 751		5	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	3
## 620	3	5	5	4
## 683	1	1	2	2
## 695	3	4	3	5
## 743	2	5	2	2
## 751	3	4	4	3
## 804	1	4	1	4
## 847	3	4	3	3
## 849	3	3	4	2
## 854	5	5	3	3
## 931	2	3	2	2
## 938	2	4	1	2
## 941	2	1	1	3
## 958	2	1	1	3
## 1007	3	1	5	5

##	Final.judgement	Reliability	Keeping.promises	Loss.of.interest
----	-----------------	-------------	------------------	------------------

## 13	1	5	5	4
## 33	1	1	1	1
## 35	5	4	4	4
## 54	3	5	5	1
## 92	5	3	5	3
## 125	2	4	4	1
## 139	2	5	5	5
## 150	3	2	2	2
## 230	3	3	5	2
## 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	5
## 33	3	3	4	1
## 35	3	3	4	1
## 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	Branded.clothing
## 13	2	4	1	1
## 33	3	1	3	1
## 35	2	3	3	4
## 54	4	3	5	3
## 92	3	5	5	5
## 125	2	3	3	3
## 139	5	4	1	1
## 150	3	3	2	1
## 230	2	1	2	2
## 260	5	4	1	3
## 366	1	3	3	3
## 372	3	3	3	2
## 443	1	2	4	2
## 476	2	1	2	2
## 503	3	2	5	5
## 525	5	4	2	5
## 571	1	3	5	5
## 620	1	2	2	4

## 683	3	1	4	4
## 695	3	3	2	3
## 743	3	1	4	5
## 751	3	4	3	3
## 804	1	3	3	2
## 847	5	4	4	3
## 849	2	3	4	5
## 854	3	4	2	1
## 931	3	3	5	2
## 938	4	5	2	2
## 941	3	2	5	5
## 958	3	3	2	5
## 1007	3	3	1	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
```

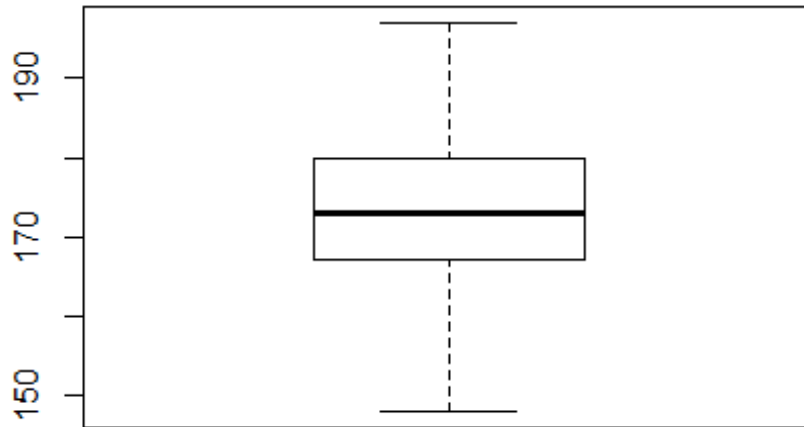
```
dim(impu_num_compl)
```

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

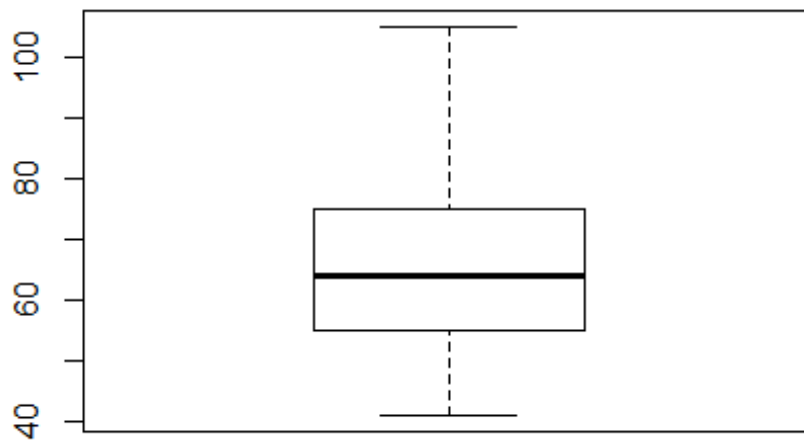
```
dim(impu_cat_compl)
```

```
## [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)

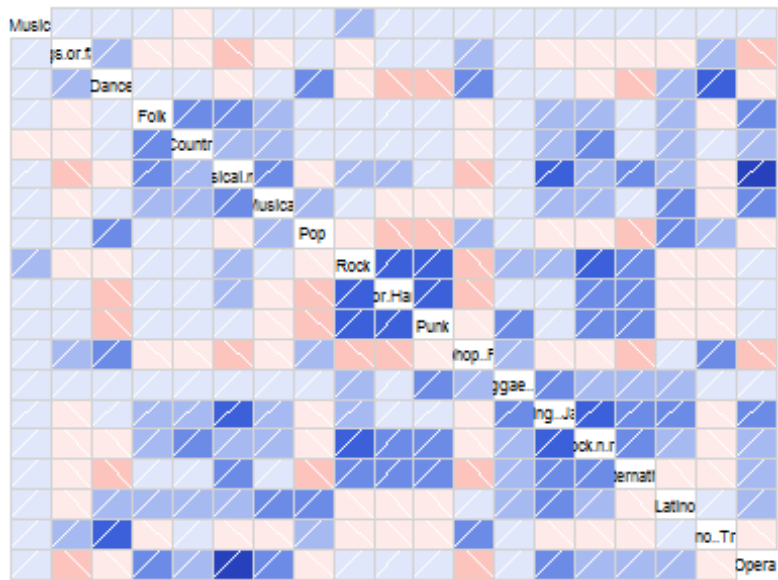
## Warning: package 'corrgram' was built under R version 3.6.2

## 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)
```

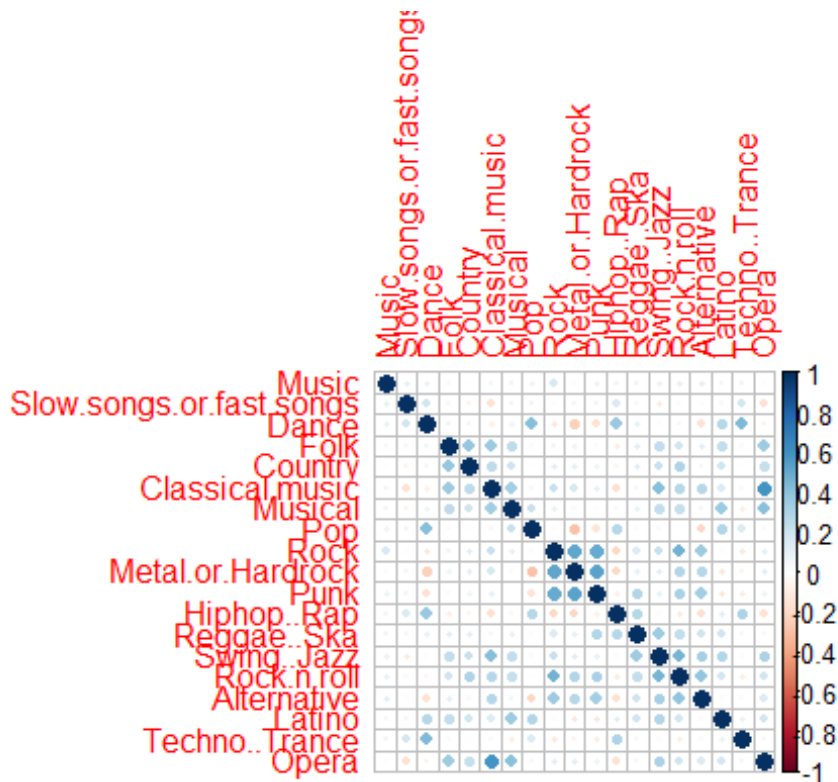


```
library(corrplot)

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

## corrplot 0.84 loaded

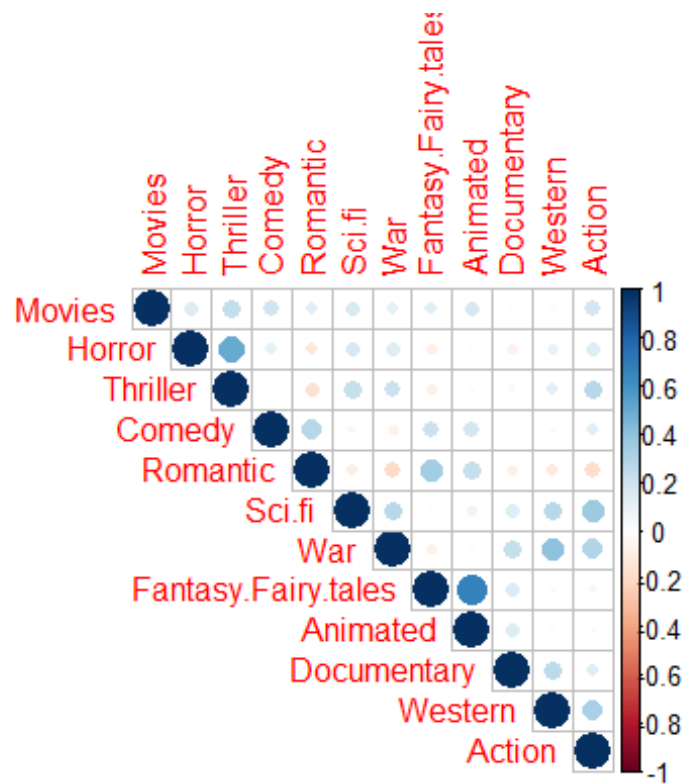
music_cor = cor(music_trans)
corrplot(music_cor)
```



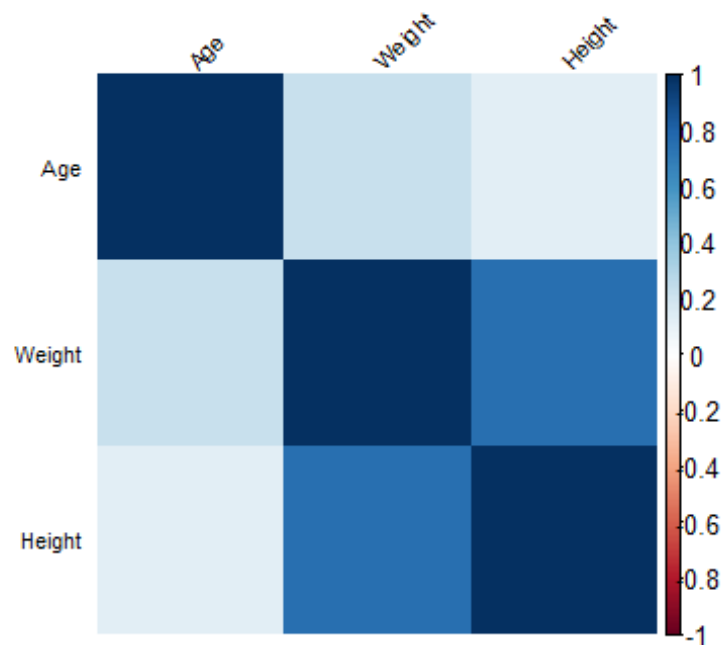
```
movie_trans = data_transformed[,names(movie_data)]
dim(movie_trans)

## [1] 1010 12

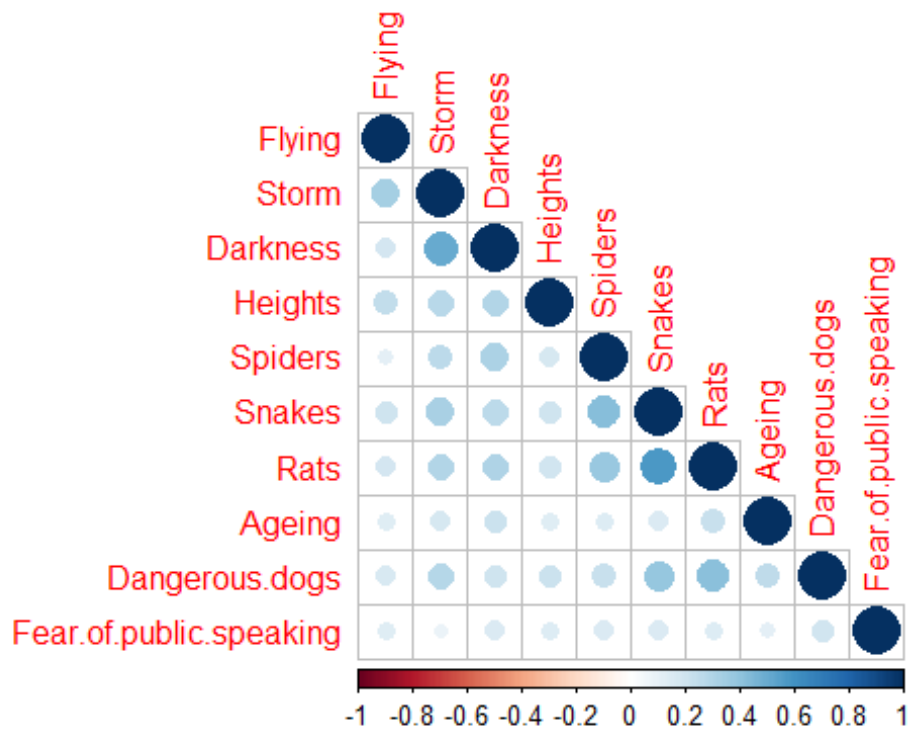
movie_cor = cor(movie_trans)
corrplot(movie_cor,type="upper")
```

```
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)

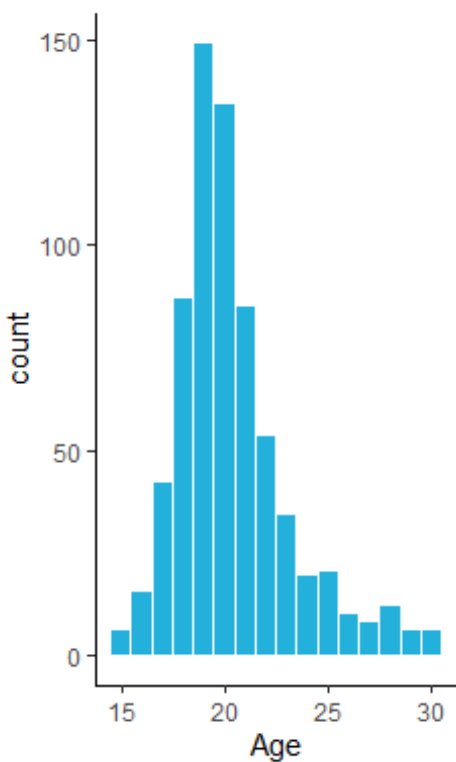
## Warning: package 'reshape2' was built under R version 3.6.2

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

```

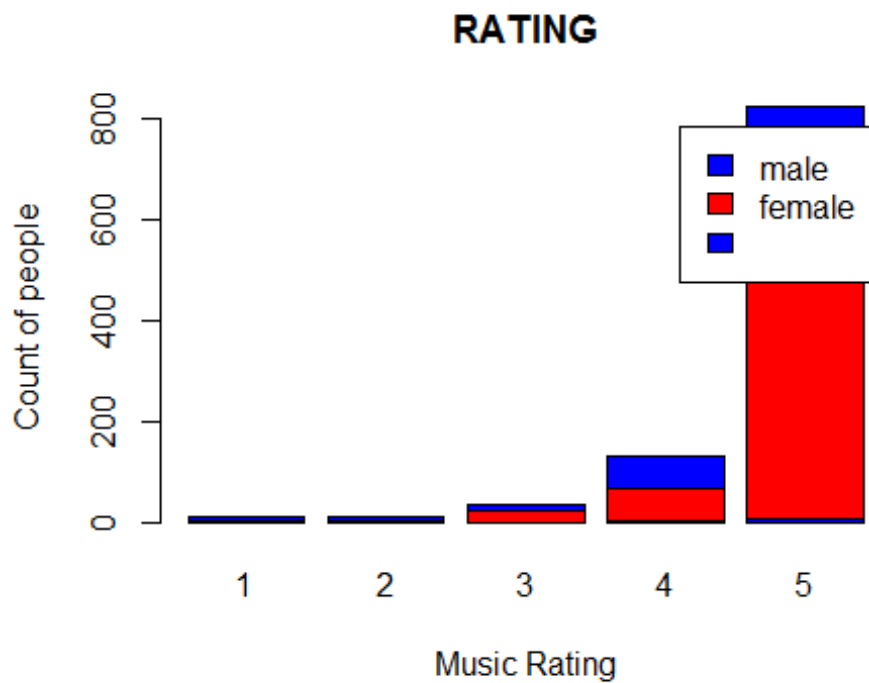
```
panel.grid.minor = element_blank(), axis.line = element_line(colour =
"black"))
```

```
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)
```

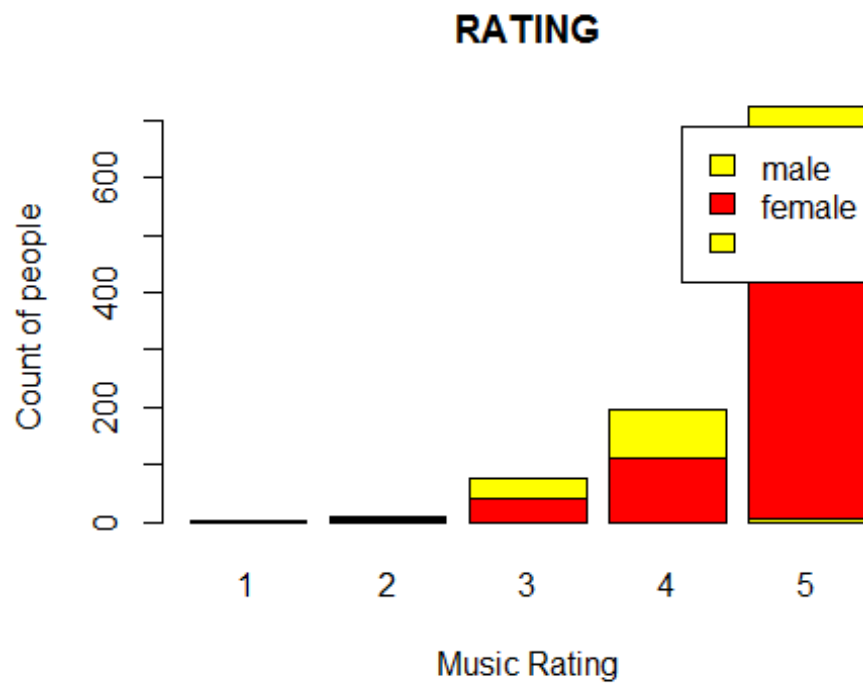


	Age	count
1	19	210
2	20	194
3	21	127
4	18	123
5	22	84
6	17	53

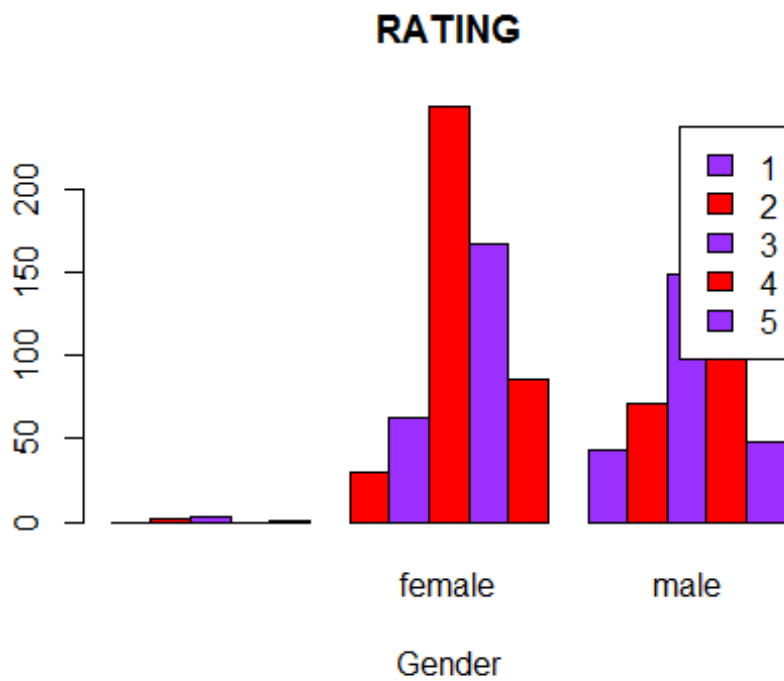
```
counts <- table(data_transformed$Gender, data_transformed$Music)
barplot(counts, main="RATING",
        xlab="Music Rating", ylab = "Count of people", col=c("blue","red"),
        legend = rownames(counts))
```

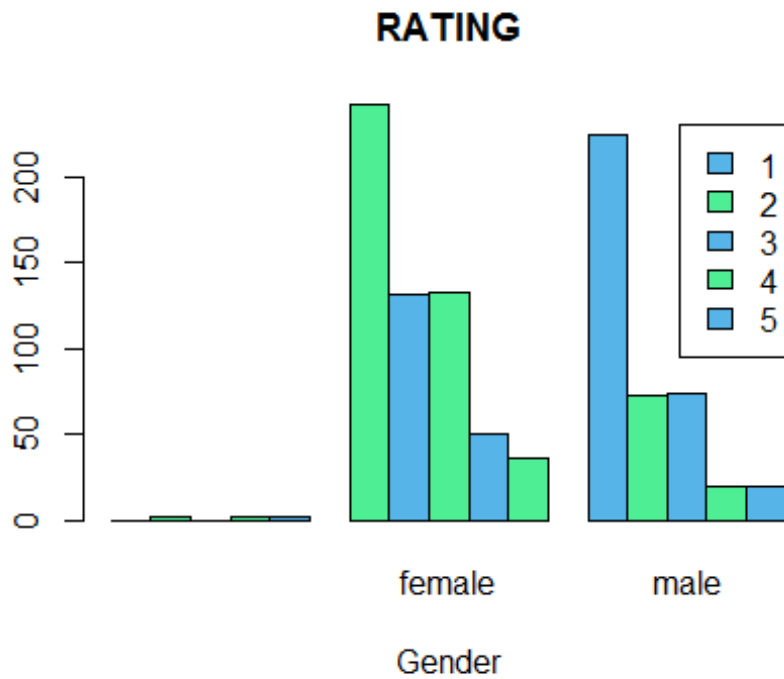
```
counts <- table(data_transformed$Gender, data_transformed$Movies)
barplot(counts, main="RATING",
        xlab="Music Rating", ylab = "Count of people", col=c("Yellow","red"),
        legend = rownames(counts))
```



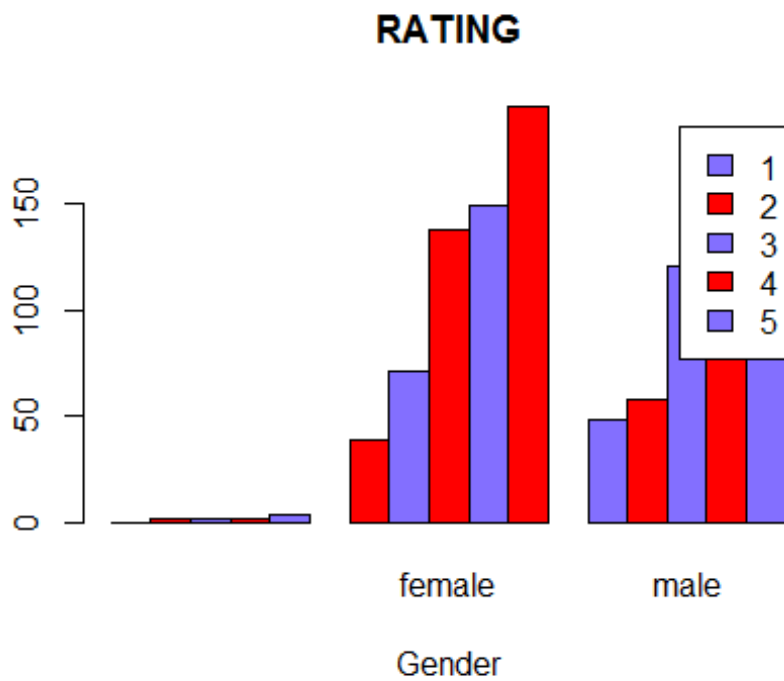
```
#Health
counts <- table(data_transformed$Health, data_transformed$Gender)
barplot(counts, main="RATING",
        xlab="Gender", col=c("purple1", "red"),
        legend = rownames(counts), beside=TRUE)
```



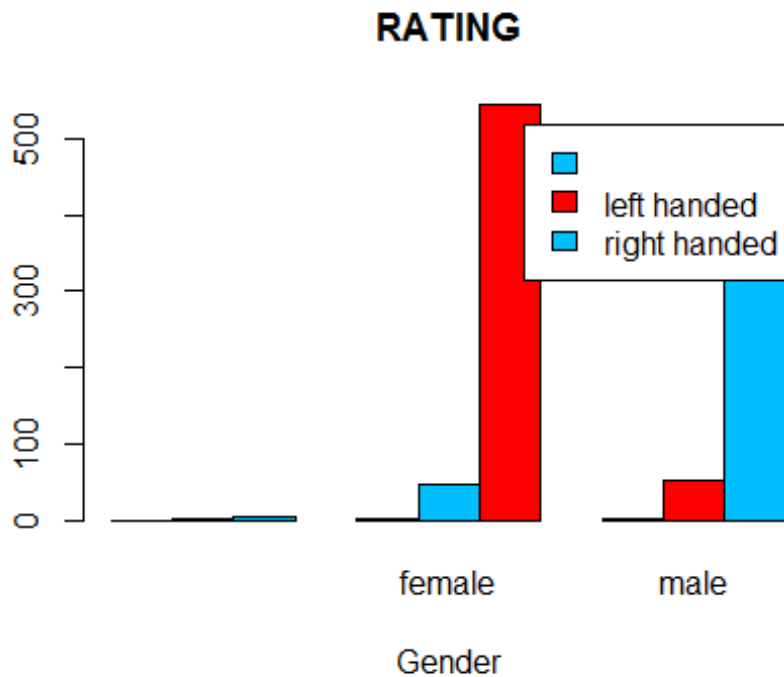
```
#Flying_----- Phobias
counts <- table(data_transformed$Flying, data_transformed$Gender)
barplot(counts, main="RATING",
        xlab="Gender", col=c("#56B4E9", "seagreen2"),
        legend = rownames(counts), beside=TRUE)
```



```
#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)
```



```
dim(hobbies_data)
## [1] 1010 32

hobbies_transformed <- data_transformed[,names(hobbies_data)]
hobbies_pca <- prcomp(hobbies_transformed, scale= T)
names(hobbies_pca)

## [1] "sdev" "rotation" "center" "scale" "x"

#rotation - represents the loadings of each variables on components
hobbies_pca$rotation

##          PC1          PC2          PC3          PC4
## History    0.188458129 -0.09972504  0.191659752 -0.22991541
## Psychology  0.230036558  0.01786097  0.099845621 -0.12092486
## Politics    0.123498901 -0.20221515  0.261041540 -0.21465574
## Mathematics 0.071298342 -0.28497563 -0.148794306 -0.11696119
## Physics     0.117537829 -0.32127215 -0.264318100 -0.13787033
## Internet   -0.047107481 -0.24945405  0.114615245  0.10689102
## PC         -0.040810044 -0.37916578 -0.026702872 -0.01947711
## Economy.Management -0.004247973 -0.19189231  0.297016846  0.01081594
## Biology     0.275152415  0.02013486 -0.351237734  0.16443892
## Chemistry   0.220249354 -0.02970157 -0.402713435  0.09655225
## Reading     0.273290822  0.20243199  0.075074553 -0.17176942
## Geography   0.165733285 -0.15102913  0.150868856 -0.07075296
## Foreign.languages 0.196795893  0.01943294  0.224850707 -0.01840982
## Medicine    0.283815963 -0.01061050 -0.296379491  0.11658962
```

## Law	0.139885598	-0.14966276	0.276167023	-0.02315421
## Cars	-0.054065830	-0.35287817	0.054454401	0.15197566
## Art.exhibitions	0.307286358	0.08363239	0.106223775	-0.05528705
## Religion	0.234554023	-0.02118824	-0.009334782	-0.16050540
## Countryside..outdoors	0.196673276	-0.02578241	-0.009371219	0.11783931
## Dancing	0.232312679	0.05907080	0.086621432	0.24672526
## Musical.instruments	0.216399667	-0.01219800	-0.012291430	-0.08626889
## Writing	0.231718611	0.07397364	0.067636420	-0.15644741
## Passive.sport	-0.023439504	-0.15300983	0.051616951	0.20587472
## Active.sport	0.069058573	-0.20429023	0.043485929	0.24414843
## Gardening	0.192890484	0.03464648	-0.080288637	0.18959498
## Celebrities	0.015776035	0.09462840	0.186510020	0.34123592
## Shopping	0.083917204	0.12707686	0.199581102	0.39774969
## Science.and.technology	0.097311047	-0.33700031	-0.117622329	-0.03484404
## Theatre	0.293866571	0.13860458	0.110875136	-0.03254494
## Fun.with.friends	0.062460243	-0.05187930	0.139364342	0.26018379
## Adrenaline.sports	0.074602609	-0.25199639	0.062705429	0.22378600
## Pets	0.087440530	0.03835047	-0.026758239	0.25545973
##	PC5	PC6	PC7	PC8
## History	0.092591265	-0.232015433	0.25239113	-0.06548575
## Psychology	0.148720516	0.014508725	-0.14185033	0.08552131
## Politics	0.247013176	-0.132333726	0.12069825	0.12790917
## Mathematics	-0.002370872	0.290108551	-0.24097120	0.17220703
## Physics	-0.005937990	0.138911115	-0.15417460	0.13472563
## Internet	-0.031563650	0.409201245	0.15277805	-0.33138111
## PC	-0.140059444	0.319267598	0.09215982	-0.08493418
## Economy.Management	0.219596773	0.149171977	-0.20031212	0.17069411
## Biology	0.237872115	-0.035475460	0.02578910	-0.11433832
## Chemistry	0.254623212	-0.044583322	-0.03483529	-0.07935483
## Reading	0.009166497	0.073519057	-0.01308066	-0.14064121
## Geography	-0.001296033	-0.207056165	0.27032491	-0.27703787
## Foreign.languages	0.062444420	0.094568340	-0.14992093	-0.41035039
## Medicine	0.274674151	-0.056459180	0.01393986	-0.12804866
## Law	0.358448817	-0.101036149	0.02537942	0.16761041
## Cars	-0.003879326	-0.088122453	0.11289934	0.04308159
## Art.exhibitions	-0.194064673	0.083454620	-0.08848909	0.01818948
## Religion	-0.066276692	0.064782579	0.19193962	0.06085112
## Countryside..outdoors	-0.352637463	-0.097031086	0.22136307	-0.14052420
## Dancing	-0.079618956	0.012731339	-0.16346012	0.18230175
## Musical.instruments	-0.366971152	0.082456456	-0.01926015	0.09180239
## Writing	-0.205552004	0.136106746	0.03084835	0.23993099
## Passive.sport	-0.146071143	-0.127491042	0.15782316	-0.17095880
## Active.sport	-0.141964721	-0.325163564	-0.12763567	0.26173364
## Gardening	-0.147538683	0.082339921	0.35510845	0.31300178
## Celebrities	0.162685242	0.307917512	0.18206845	0.13487536
## Shopping	0.157865883	0.241111976	0.02183875	0.02799432
## Science.and.technology	0.001477316	0.083055353	0.04019640	-0.08928700
## Theatre	-0.101931633	0.064560558	-0.21810241	-0.07438009
## Fun.with.friends	-0.109742421	-0.063689827	-0.39486085	-0.29098669
## Adrenaline.sports	-0.143782184	-0.329756935	-0.24642978	0.03323823

## Pets	0.051462435	-0.009519456	0.21944914	0.10229865
##	PC9	PC10	PC11	PC12
## History	0.2342319487	0.013213679	0.02656929	0.014686171
## Psychology	0.1634428977	-0.173316277	0.29922905	-0.217844043
## Politics	0.0003330098	0.028409473	0.11560478	0.089453472
## Mathematics	-0.1227967977	0.415744678	-0.03102544	0.146137501
## Physics	0.0605717983	0.187165070	-0.05964467	0.118388354
## Internet	0.0501169630	-0.135904736	0.13256657	-0.038345456
## PC	0.0836480239	-0.194955920	-0.08441881	-0.110075826
## Economy.Management	-0.2578599951	0.237036055	-0.05440312	-0.099314337
## Biology	-0.0656992949	-0.068240695	0.06665055	-0.004194024
## Chemistry	-0.0448282578	-0.012743686	-0.03919365	0.109878401
## Reading	0.1311171386	0.121373314	-0.16553035	0.121776808
## Geography	-0.2121255816	0.075893927	-0.40988342	0.173529969
## Foreign.languages	-0.2743702248	-0.085613319	-0.18114315	-0.019851505
## Medicine	-0.1044512509	-0.176308133	0.08116972	0.001417123
## Law	-0.0129844112	0.024797536	0.13392287	0.042392594
## Cars	0.1126257763	-0.100383522	0.05174534	-0.138464290
## Art.exhibitions	0.1480556405	0.028108790	-0.03140235	0.049578834
## Religion	-0.2035988869	-0.004848608	0.35849183	-0.393017524
## Countryside..outdoors	-0.1682045383	0.334944022	-0.04810741	-0.379477869
## Dancing	-0.2718126978	-0.035144419	-0.02929581	-0.169902419
## Musical.instruments	-0.1518894347	-0.228909445	0.26370229	0.136685469
## Writing	0.0859320390	-0.365865217	-0.11663001	0.303731172
## Passive.sport	-0.1388167380	0.193524389	0.49466560	0.554056046
## Active.sport	-0.0978546087	-0.191441181	-0.20297874	0.037383030
## Gardening	-0.0174921768	0.134950497	-0.07051273	0.075730596
## Celebrities	-0.0213124787	-0.112807899	-0.10859489	0.173889435
## Shopping	0.0241152752	-0.022607496	-0.02547038	-0.019461421
## Science.and.technology	0.2525523547	-0.146066026	-0.18943901	-0.072974790
## Theatre	0.2373060604	0.203906538	-0.02908785	0.045437909
## Fun.with.friends	0.2492036609	0.070038751	0.20242860	0.038986631
## Adrenaline.sports	0.1002526616	-0.115681528	-0.07578508	-0.052770482
## Pets	0.4875204274	0.294586372	-0.01534318	-0.099785036
##	PC13	PC14	PC15	PC16
## History	-0.11918292	-0.264841906	0.23562129	0.048499587
## Psychology	-0.53147080	-0.134003144	-0.25960270	-0.212537201
## Politics	0.14225150	0.020275698	0.17110829	-0.030416836
## Mathematics	-0.17389206	-0.088680894	0.04575537	0.190550851
## Physics	-0.02725997	-0.176143809	0.01568440	0.150602790
## Internet	-0.24231809	-0.094006707	0.14534745	-0.080953151
## PC	0.03656889	0.065041281	0.07195299	-0.073035218
## Economy.Management	-0.02685626	0.258389802	-0.09402371	-0.306934336
## Biology	-0.03674544	0.014119227	-0.03946773	-0.083243298
## Chemistry	0.07105813	0.018197628	0.10117714	-0.021308793
## Reading	0.05977048	-0.012847994	0.11893710	0.045485171
## Geography	-0.12965737	-0.141435945	0.01708068	0.037413428
## Foreign.languages	-0.11516304	0.336766162	-0.04416348	0.114506778
## Medicine	0.05568575	0.112221784	-0.05484492	-0.011739787
## Law	0.20052654	0.220965966	0.03155761	0.053131138

## Cars	0.35424863	0.088685964	-0.19272303	0.127627879
## Art.exhibitions	0.15766714	-0.095611890	-0.35873345	0.021981645
## Religion	0.14689966	-0.232714089	0.02768653	0.224113200
## Countryside..outdoors	-0.02248433	0.064114677	-0.13582649	-0.052047476
## Dancing	-0.07621054	-0.156530043	0.29739895	0.006507375
## Musical.instruments	0.08865918	0.372599208	0.20443150	0.279091075
## Writing	-0.13370538	0.143648944	-0.01997338	-0.120877714
## Passive.sport	-0.10655182	-0.033802622	-0.27309109	-0.053951301
## Active.sport	-0.16497639	-0.187370312	0.05512601	0.165136097
## Gardening	0.08029240	0.051113842	0.16972346	-0.577177451
## Celebrities	0.06850153	-0.234754041	-0.08292139	0.248466200
## Shopping	0.06411123	-0.170558272	-0.09922488	0.126536851
## Science.and.technology	0.17199686	-0.004297953	-0.15015236	-0.130311971
## Theatre	0.26338123	-0.078838818	-0.21457617	-0.028072652
## Fun.with.friends	0.17384884	-0.101511359	0.47844396	-0.185474306
## Adrenaline.sports	-0.10302385	0.065904906	-0.16953678	-0.015564040
## Pets	-0.33225220	0.440512655	0.11433088	0.322198191
##	PC17	PC18	PC19	PC20
## History	0.140568535	0.019058804	-0.063137882	-0.340299967
## Psychology	0.005707387	0.129875824	0.031426925	-0.214641018
## Politics	-0.121565756	0.217081565	-0.068302094	0.003366495
## Mathematics	-0.087735929	-0.110900070	-0.023429064	-0.143880906
## Physics	-0.073782415	0.068293026	0.076768503	-0.050267426
## Internet	0.217522404	-0.215441375	0.162339326	0.176880420
## PC	0.211440075	-0.111475474	-0.031609704	-0.047651596
## Economy.Management	-0.045930537	0.027840840	0.036412935	0.102489385
## Biology	0.003257511	-0.012333933	0.069147808	-0.038975437
## Chemistry	0.065622261	-0.045303826	0.024145774	-0.027908462
## Reading	0.118537094	-0.343128291	-0.273039466	0.051648044
## Geography	-0.176918197	0.137768750	0.384144671	0.184347919
## Foreign.languages	-0.095618470	-0.048243639	-0.400141090	-0.196940783
## Medicine	0.028008379	-0.062090399	0.091488004	0.042749446
## Law	0.232253891	-0.212837437	0.163565688	0.047394133
## Cars	0.101290718	-0.105382064	0.064788378	-0.077041582
## Art.exhibitions	0.153430645	-0.122627223	0.216359971	0.014236907
## Religion	-0.385963102	-0.232352629	-0.186134785	0.334236445
## Countryside..outdoors	0.041465898	0.063372624	0.104381153	-0.213741811
## Dancing	0.347373197	0.315367885	0.078103955	0.303012386
## Musical.instruments	0.023551049	0.261318628	0.136422927	-0.305837922
## Writing	-0.232484887	-0.091472367	0.092251598	0.294789399
## Passive.sport	0.080514277	0.071173847	-0.224338422	0.159425948
## Active.sport	0.244062647	-0.171008779	-0.379805344	0.038079046
## Gardening	-0.083220637	-0.152562035	-0.158094901	-0.204521380
## Celebrities	-0.162035703	0.019266520	0.116912722	-0.286153981
## Shopping	-0.169084120	0.142479980	-0.132551335	-0.061052617
## Science.and.technology	-0.123494310	0.487730771	-0.345660363	0.152263003
## Theatre	0.223948702	0.105474154	-0.003676628	0.108807117
## Fun.with.friends	-0.282540907	-0.008602802	0.091997421	-0.028442837
## Adrenaline.sports	-0.331133280	-0.278594683	0.153394827	-0.068272781
## Pets	-0.046438159	0.058743608	-0.021156648	0.241707897

##	PC21	PC22	PC23	PC24
## History	0.089378024	0.27249982	0.081685958	0.404197186
## Psychology	0.105591700	-0.00341174	-0.142315832	-0.212594526
## Politics	-0.169374235	-0.31925693	-0.086141769	-0.168492841
## Mathematics	0.077297990	0.03806932	-0.006096263	-0.035969797
## Physics	-0.013570922	0.11217012	0.162977863	-0.240395516
## Internet	-0.155409428	-0.09517320	0.174672673	0.013745440
## PC	-0.029353779	-0.12653046	-0.172011307	0.097874903
## Economy.Management	0.056489720	0.03141804	-0.161727689	0.522600837
## Biology	0.036599800	-0.09641642	-0.019273983	0.149770112
## Chemistry	-0.042457150	0.10832498	0.063636914	0.192233868
## Reading	-0.143087905	0.20705552	-0.348080118	-0.001703690
## Geography	0.213153785	-0.12339851	-0.024883673	-0.022706934
## Foreign.languages	0.213276835	-0.00187073	0.098434706	-0.251287644
## Medicine	-0.033030822	-0.05857201	-0.124398783	-0.031707678
## Law	-0.261291638	0.07208552	0.310180864	-0.283896857
## Cars	0.487331938	0.36616198	-0.200557843	-0.139096644
## Art.exhibitions	0.296753202	-0.33764367	0.195876942	0.058571883
## Religion	0.061461577	-0.07981032	-0.032296540	0.098511884
## Countryside..outdoors	-0.405737882	0.18979428	0.044101941	-0.119357905
## Dancing	0.100580228	0.23799480	-0.165722489	-0.146517269
## Musical.instruments	-0.022222980	-0.10777622	0.013830321	0.222434963
## Writing	-0.068715353	0.37683012	0.070487540	-0.025514325
## Passive.sport	-0.006673113	0.10716349	-0.094521343	0.014066381
## Active.sport	-0.035453278	-0.29817392	0.081032254	0.066455305
## Gardening	0.191596177	-0.13655688	0.078457744	-0.157125078
## Celebrities	-0.193970300	-0.08718504	-0.443689860	-0.071644312
## Shopping	0.011889605	0.18851704	0.504202132	0.182827722
## Science.and.technology	-0.167691163	0.02308271	0.015102689	-0.001437206
## Theatre	-0.137914691	-0.12133188	-0.093273415	0.082180350
## Fun.with.friends	0.107272447	-0.03169795	-0.033741293	-0.096182782
## Adrenaline.sports	-0.274955137	0.04690266	-0.068476535	0.117502955
## Pets	0.120812178	-0.12080933	-0.059398012	0.020557043
##	PC25	PC26	PC27	PC28
## History	-0.015325238	-0.053328696	-0.30396841	0.165928980
## Psychology	0.110054126	0.151113617	0.19481959	-0.132513914
## Politics	-0.347059475	-0.438008513	0.26221721	-0.074297731
## Mathematics	-0.022970305	0.072390127	0.01836953	0.102939921
## Physics	0.024431136	-0.066306695	-0.02225712	-0.033656426
## Internet	-0.155178320	-0.041469767	0.22223796	0.398051285
## PC	0.015283657	0.026053169	-0.21824761	-0.580466363
## Economy.Management	0.183985283	-0.095838917	0.04118931	0.052189789
## Biology	0.071131033	-0.088068770	-0.04602056	-0.060382767
## Chemistry	-0.131050003	-0.155411984	0.07054633	-0.055676060
## Reading	-0.059989639	0.116711124	0.34129790	-0.258231158
## Geography	0.118727757	0.292600948	0.15721901	-0.122039954
## Foreign.languages	-0.101072969	-0.085507934	-0.23360523	0.135039521
## Medicine	0.107828167	-0.005904270	0.02810581	0.138094731
## Law	0.226302203	0.291996307	-0.21384712	-0.089548162
## Cars	-0.012959607	-0.077885679	0.29806334	0.143823156

## Art.exhibitions	-0.088515158	-0.305624403	-0.19551497	-0.128551909
## Religion	0.051195327	0.108817872	-0.11200946	0.024296446
## Countryside..outdoors	0.207300073	-0.308268090	0.08157090	-0.043944572
## Dancing	-0.306851079	0.006402375	-0.21306071	-0.040119998
## Musical.instruments	0.011472790	0.237780789	0.20544425	0.020807957
## Writing	0.183600601	-0.295185892	-0.02768596	0.070409544
## Passive.sport	-0.013564230	-0.001914892	-0.11556287	-0.129317040
## Active.sport	0.328790684	-0.091217882	0.19676068	0.081321547
## Gardening	-0.124510434	0.217420577	-0.01340209	0.117973210
## Celebrities	0.132522166	-0.065358357	-0.21211129	0.135374613
## Shopping	-0.109395978	0.069785761	0.29544826	-0.299961683
## Science.and.technology	0.087363698	0.163581113	-0.10841696	0.113833901
## Theatre	-0.023644657	0.226538771	0.05989364	0.309597289
## Fun.with.friends	0.321433616	-0.102270132	-0.02456618	-0.068338982
## Adrenaline.sports	-0.498461041	0.180729640	-0.11356323	0.003381675
## Pets	0.006329268	-0.038696539	-0.06279122	-0.011001691
##	PC29	PC30	PC31	PC32
## History	0.0436252264	0.08885972	-0.103643613	0.024873786
## Psychology	0.0003107664	-0.11531229	0.115281418	0.089776263
## Politics	0.1014169997	-0.04752005	-0.112039360	-0.049791245
## Mathematics	-0.1355805402	-0.34968646	-0.460256858	0.006634840
## Physics	0.2005685773	0.54567187	0.397435033	-0.060552789
## Internet	-0.1195715221	0.03743514	0.085149185	-0.047635005
## PC	0.3364108291	-0.07193913	-0.109636364	0.059974932
## Economy.Management	-0.0416391116	0.18651618	0.161571740	0.028396348
## Biology	0.0034737992	-0.03368850	-0.076829898	-0.782877286
## Chemistry	0.0016929119	-0.46994875	0.452349152	0.368489468
## Reading	-0.3029578433	0.18862194	0.049639285	-0.076930129
## Geography	0.0434073886	-0.07451515	0.028324979	0.026165782
## Foreign.languages	0.1328496283	-0.01348350	0.107185899	-0.019764701
## Medicine	0.0670882078	0.40844571	-0.477668691	0.434044599
## Law	-0.1208068942	-0.11101063	0.041150511	-0.055407938
## Cars	0.0285055732	-0.07887428	0.005240120	-0.085151482
## Art.exhibitions	-0.3798787386	0.05354058	0.010294708	0.096928939
## Religion	0.0538250641	-0.05636915	0.116502577	0.030215774
## Countryside..outdoors	-0.0140779316	-0.05361508	-0.042472613	0.031224932
## Dancing	-0.0817104451	0.02524419	-0.036321042	-0.022423769
## Musical.instruments	-0.1095417829	0.05151530	0.088708506	-0.022241308
## Writing	0.1553783156	-0.13250198	-0.106361802	-0.046039769
## Passive.sport	0.0056073362	0.02706482	0.029672665	0.031963639
## Active.sport	0.0329459508	-0.01366870	0.033893795	0.005157876
## Gardening	0.0286857110	0.04824095	0.041577449	0.037166519
## Celebrities	-0.0600954317	-0.03197759	0.154158929	0.019391548
## Shopping	0.1108551014	0.07020307	-0.165240833	0.033732846
## Science.and.technology	-0.4034422108	-0.05396951	-0.022175035	0.003311783
## Theatre	0.5421808200	-0.14293657	-0.003959493	-0.046385582
## Fun.with.friends	-0.0469591626	-0.03661335	-0.013638638	0.037373544
## Adrenaline.sports	-0.0248298425	0.05120411	-0.004412017	-0.026971282
## Pets	0.0339082048	0.01427524	0.039473323	0.037922218

```
# Calculate the variance explained by each component
```

```
eigen_hobbies <- hobbies_pca$sdev^2
```

```
names(eigen_hobbies) <- paste("PC",1:32,sep="")
```

```
eigen_hobbies
```

```
##      PC1      PC2      PC3      PC4      PC5      PC6      PC7
PC8
## 4.2391294 3.2188434 2.5628009 2.1936649 1.5704999 1.3868499 1.1689198
1.1287278
##      PC9      PC10      PC11      PC12      PC13      PC14      PC15
PC16
## 1.0887872 0.9921651 0.9212331 0.8913134 0.8595703 0.7790779 0.7677160
0.7298606
##      PC17      PC18      PC19      PC20      PC21      PC22      PC23
PC24
## 0.6634885 0.6386832 0.6241286 0.5868250 0.5521857 0.5316185 0.5036362
0.4961543
##      PC25      PC26      PC27      PC28      PC29      PC30      PC31
PC32
## 0.4709276 0.4429519 0.4079236 0.3757610 0.3676952 0.3061206 0.2978055
0.2349349
```

```
sum_hobbies <- sum(eigen_hobbies)
```

```
sum_hobbies
```

```
## [1] 32
```

```
## Proportion of variance explained by each component
```

```
propvar_hobbies <- eigen_hobbies/sum_hobbies
```

```
propvar_hobbies
```

```
##      PC1      PC2      PC3      PC4      PC5      PC6
## 0.132472795 0.100588857 0.080087528 0.068552030 0.049078121 0.043339059
##      PC7      PC8      PC9      PC10      PC11      PC12
## 0.036528744 0.035272745 0.034024598 0.031005160 0.028788535 0.027853544
##      PC13      PC14      PC15      PC16      PC17      PC18
## 0.026861573 0.024346185 0.023991126 0.022808142 0.020734014 0.019958849
##      PC19      PC20      PC21      PC22      PC23      PC24
## 0.019504019 0.018338280 0.017255805 0.016613077 0.015738632 0.015504822
##      PC25      PC26      PC27      PC28      PC29      PC30
## 0.014716488 0.013842245 0.012747613 0.011742532 0.011490476 0.009566268
##      PC31      PC32
## 0.009306422 0.007341716
```

```
cumvar_hobbies <- cumsum(propvar_hobbies)
```

```
cumvar_hobbies
```

```
##      PC1      PC2      PC3      PC4      PC5      PC6      PC7
PC8
## 0.1324728 0.2330617 0.3131492 0.3817012 0.4307793 0.4741184 0.5106471
```

```

0.5459199
##          PC9          PC10          PC11          PC12          PC13          PC14          PC15
PC16
## 0.5799445 0.6109496 0.6397382 0.6675917 0.6944533 0.7187995 0.7427906
0.7655987
##          PC17          PC18          PC19          PC20          PC21          PC22          PC23
PC24
## 0.7863328 0.8062916 0.8257956 0.8441339 0.8613897 0.8780028 0.8937414
0.9092462
##          PC25          PC26          PC27          PC28          PC29          PC30          PC31
PC32
## 0.9239627 0.9378050 0.9505526 0.9622951 0.9737856 0.9833519 0.9926583
1.0000000

matlambdas <- rbind(eigen_hobbies,propvar_hobbies,cumvar_hobbies)
rownames(matlambdas) <- c("Eigenvalues","Prop. variance","Cum. prop.
variance")
round(matlambdas,4)

##          PC1    PC2    PC3    PC4    PC5    PC6    PC7
PC8
## Eigenvalues      4.2391 3.2188 2.5628 2.1937 1.5705 1.3868 1.1689
1.1287
## Prop. variance    0.1325 0.1006 0.0801 0.0686 0.0491 0.0433 0.0365
0.0353
## Cum. prop. variance 0.1325 0.2331 0.3131 0.3817 0.4308 0.4741 0.5106
0.5459
##          PC9    PC10    PC11    PC12    PC13    PC14    PC15
PC16
## Eigenvalues      1.0888 0.9922 0.9212 0.8913 0.8596 0.7791 0.7677
0.7299
## Prop. variance    0.0340 0.0310 0.0288 0.0279 0.0269 0.0243 0.0240
0.0228
## Cum. prop. variance 0.5799 0.6109 0.6397 0.6676 0.6945 0.7188 0.7428
0.7656
##          PC17    PC18    PC19    PC20    PC21    PC22    PC23
PC24
## Eigenvalues      0.6635 0.6387 0.6241 0.5868 0.5522 0.5316 0.5036
0.4962
## Prop. variance    0.0207 0.0200 0.0195 0.0183 0.0173 0.0166 0.0157
0.0155
## Cum. prop. variance 0.7863 0.8063 0.8258 0.8441 0.8614 0.8780 0.8937
0.9092
##          PC25    PC26    PC27    PC28    PC29    PC30    PC31
PC32
## Eigenvalues      0.4709 0.4430 0.4079 0.3758 0.3677 0.3061 0.2978
0.2349
## Prop. variance    0.0147 0.0138 0.0127 0.0117 0.0115 0.0096 0.0093
0.0073

```

```

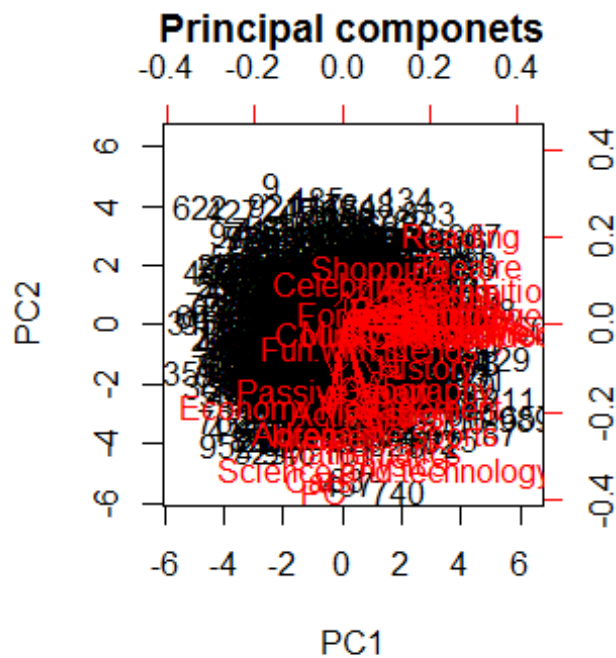
## Cum. prop. variance 0.9240 0.9378 0.9506 0.9623 0.9738 0.9834 0.9927
1.0000

summary(hobbies_pca)

## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6
PC7
## Standard deviation      2.0589 1.7941 1.60088 1.48110 1.25320 1.17765
1.08117
## Proportion of Variance 0.1325 0.1006 0.08009 0.06855 0.04908 0.04334
0.03653
## Cumulative Proportion 0.1325 0.2331 0.31315 0.38170 0.43078 0.47412
0.51065
##              PC8      PC9      PC10      PC11      PC12      PC13
PC14
## Standard deviation      1.06242 1.04345 0.99607 0.95981 0.94409 0.92713
0.88265
## Proportion of Variance 0.03527 0.03402 0.03101 0.02879 0.02785 0.02686
0.02435
## Cumulative Proportion 0.54592 0.57994 0.61095 0.63974 0.66759 0.69445
0.71880
##              PC15      PC16      PC17      PC18      PC19      PC20
PC21
## Standard deviation      0.87619 0.85432 0.81455 0.79918 0.7900 0.76605
0.74309
## Proportion of Variance 0.02399 0.02281 0.02073 0.01996 0.0195 0.01834
0.01726
## Cumulative Proportion 0.74279 0.76560 0.78633 0.80629 0.8258 0.84413
0.86139
##              PC22      PC23      PC24      PC25      PC26      PC27
PC28
## Standard deviation      0.72912 0.70967 0.7044 0.68624 0.66555 0.63869
0.61299
## Proportion of Variance 0.01661 0.01574 0.0155 0.01472 0.01384 0.01275
0.01174
## Cumulative Proportion 0.87800 0.89374 0.9093 0.92396 0.93780 0.95055
0.96230
##              PC29      PC30      PC31      PC32
## Standard deviation      0.60638 0.55328 0.54572 0.48470
## Proportion of Variance 0.01149 0.00957 0.00931 0.00734
## Cumulative Proportion 0.97379 0.98335 0.99266 1.00000

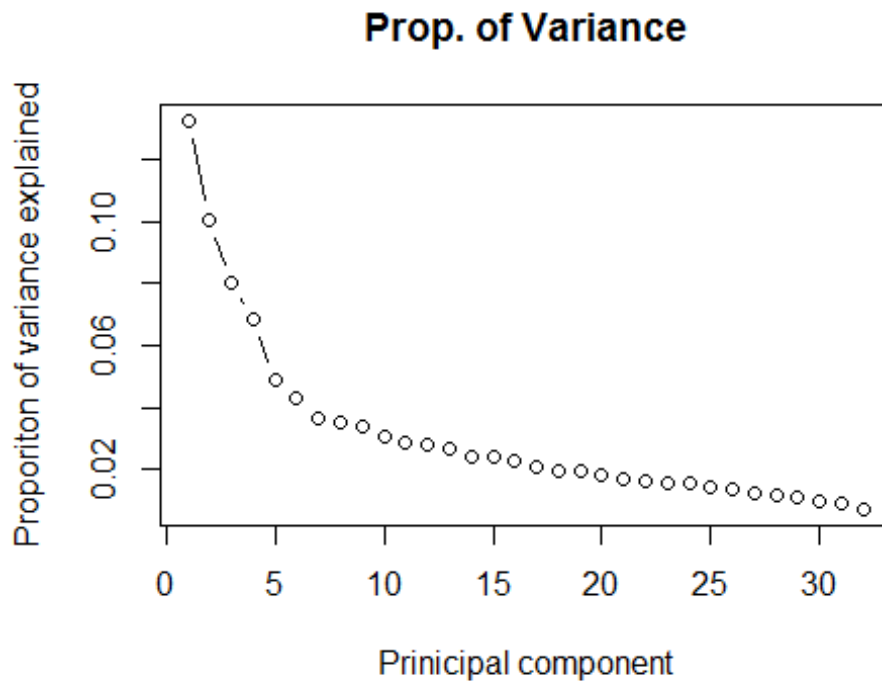
## Plot a biplot to view components on n-dimensional plane
biplot(hobbies_pca, scale = 0, main = 'Principal componets')

```



##scree plot - to identify the elbow point

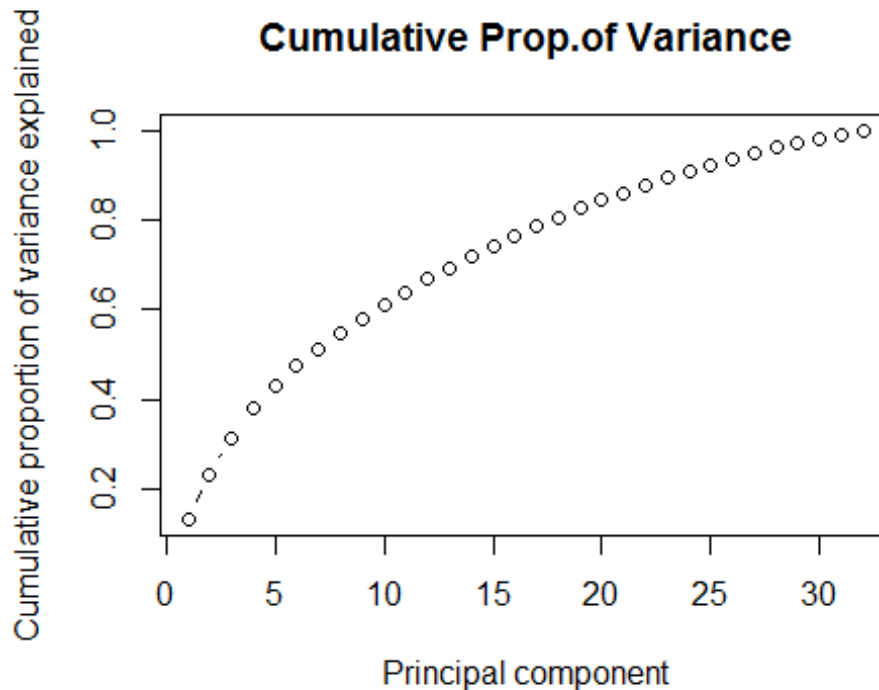
```
plot(propvar_hobbies,xlab = 'Prinicipal component',ylab = 'Proporiton of
variance explained',type = 'b', main = 'Prop. of Variance')
```



```
#The optimum number of components are ~ 8 i.e PC1 : PC8
```

```
# cumulative scree plot
```

```
plot(cumvar_hobbies,xlab = 'Principal component',ylab = 'Cumulative  
proportion of variance explained',type = 'b', main = 'Cumulative Prop.of  
Variance')
```



```
#Approx: ~ 75% of the variance is explained by 20 components i.e PC1 to PC20
```

```
dim(movie_data)
```

```
## [1] 1010 12
```

```
movie_transformed <- data_transformed[,names(movie_data)]
```

```
movie_pca <- prcomp(movie_transformed, scale= T)
```

```
names(movie_pca)
```

```
## [1] "sdev" "rotation" "center" "scale" "x"
```

```
#rotation - represents the Loadings of each variables on components
```

```
movie_pca$rotation
```

```
##          PC1          PC2          PC3          PC4
## Movies    -0.18067387  0.2943522148 -0.29434144  0.1753636709
## Horror    -0.29774492  0.0281326824 -0.48865014 -0.3468640042
## Thriller  -0.38637808  0.0264578000 -0.39885764 -0.3255496550
## Comedy     0.00957783  0.3601720121 -0.23777270  0.5045943972
## Romantic   0.25313650  0.3635153351 -0.09666751  0.2900854864
```



```

## Sci.fi          -0.37660398 0.1182255631 0.04518061 0.1849454105
## War            -0.40483304 0.0003113374 0.24428376 -0.0005316756
## Fantasy.Fairy.tales 0.13147403 0.5530319979 0.14626774 -0.2789775438
## Animated       0.05080348 0.5454401638 0.10708136 -0.3574114799
## Documentary    -0.18889650 0.1506598727 0.47688883 -0.1906702214
## Western        -0.36669650 0.0334098269 0.35553267 0.1424325175
## Action         -0.41171847 0.0941099146 0.03483277 0.3282782261
##               PC5          PC6          PC7          PC8
## Movies         -0.58171966 0.542758857 -0.063103472 -0.05424721
## Horror         0.40181487 -0.053550228 0.101364547 0.06897342
## Thriller       0.03984132 0.113480694 -0.031760285 0.07288430
## Comedy         0.50260892 -0.003243588 -0.318932106 -0.08324422
## Romantic       0.14940500 0.206681921 0.476449707 0.32906483
## Sci.fi         -0.26744131 -0.481445533 0.116396679 0.65269730
## War            0.02545531 0.268568571 0.383612801 -0.25097709
## Fantasy.Fairy.tales -0.03100975 -0.148531492 0.071361143 -0.14464158
## Animated       -0.10495198 -0.237208831 -0.005510402 -0.17859759
## Documentary    0.20900897 0.403789550 -0.512050412 0.41447721
## Western        0.28455333 0.055422480 0.377146568 -0.15168470
## Action         -0.12172688 -0.313963737 -0.292826826 -0.36763211
##               PC9          PC10         PC11         PC12
## Movies         -0.003660409 0.33290656 -0.098891004 0.052868440
## Horror         0.087255268 0.19915946 -0.553794897 0.134689408
## Thriller       -0.297400380 -0.33471692 0.595001950 -0.098215100
## Comedy         0.319128438 0.11521465 0.283954563 -0.044854364
## Romantic       -0.296782981 -0.40454959 -0.215875985 -0.107768793
## Sci.fi         0.212854244 0.10154056 0.087993301 0.041342734
## War            0.604553122 -0.35768442 -0.001701204 -0.009188477
## Fantasy.Fairy.tales -0.017579240 -0.03239820 0.148441139 0.710786335
## Animated       0.071388848 0.08856167 -0.065700856 -0.666511889
## Documentary    -0.061502779 -0.08527072 -0.161446239 0.002305173
## Western        -0.425260296 0.51196209 0.154642040 -0.039639315
## Action         -0.336418372 -0.37950074 -0.343795999 0.054460149

```

Calculate the variance explained by each component

```

eigen_movie <- movie_pca$sdev^2
names(eigen_movie) <- paste("PC",1:12,sep="")
eigen_movie

```

```

##      PC1      PC2      PC3      PC4      PC5      PC6      PC7
PC8
## 2.5328872 2.1154265 1.4605655 1.0037483 0.8077760 0.7887751 0.7483881
0.6730806
##      PC9      PC10      PC11      PC12
## 0.5946969 0.5502757 0.4200133 0.3043667

```

```

sum_movie <- sum(eigen_movie)
sum_movie

```

```
## [1] 12
```

Proportion of variance explained by each component

```
propvar_movie <- eigen_movie/sum_movie
propvar_movie

##          PC1          PC2          PC3          PC4          PC5          PC6
PC7
## 0.21107393 0.17628555 0.12171380 0.08364569 0.06731467 0.06573126
0.06236568
##          PC8          PC9          PC10          PC11          PC12
## 0.05609005 0.04955807 0.04585630 0.03500111 0.02536389

cumvar_movie <- cumsum(propvar_movie)
cumvar_movie

##          PC1          PC2          PC3          PC4          PC5          PC6          PC7
PC8
## 0.2110739 0.3873595 0.5090733 0.5927190 0.6600336 0.7257649 0.7881306
0.8442206
##          PC9          PC10          PC11          PC12
## 0.8937787 0.9396350 0.9746361 1.0000000

matlambdas_movie <- rbind(eigen_movie,propvar_movie,cumvar_movie)
rownames(matlambdas_movie) <- c("Eigenvalues","Prop. variance","Cum. prop.
variance")
round(matlambdas_movie,4)

##          PC1          PC2          PC3          PC4          PC5          PC6          PC7
PC8
## Eigenvalues          2.5329 2.1154 1.4606 1.0037 0.8078 0.7888 0.7484
0.6731
## Prop. variance          0.2111 0.1763 0.1217 0.0836 0.0673 0.0657 0.0624
0.0561
## Cum. prop. variance 0.2111 0.3874 0.5091 0.5927 0.6600 0.7258 0.7881
0.8442
##          PC9          PC10          PC11          PC12
## Eigenvalues          0.5947 0.5503 0.4200 0.3044
## Prop. variance          0.0496 0.0459 0.0350 0.0254
## Cum. prop. variance 0.8938 0.9396 0.9746 1.0000

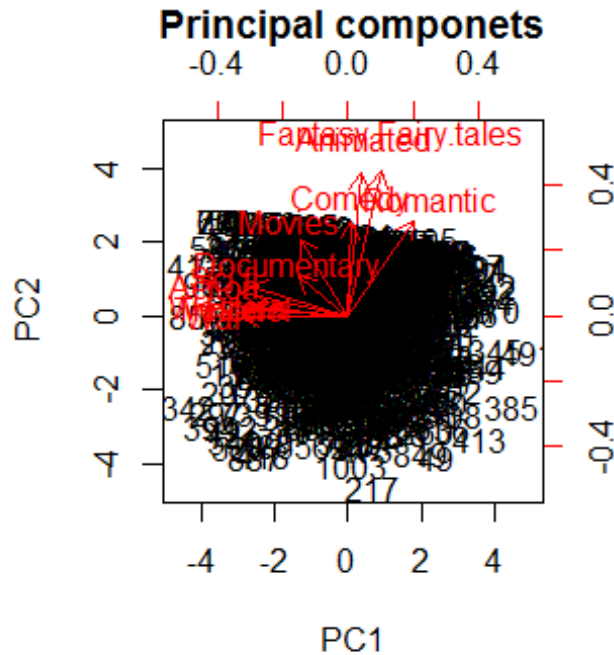
summary(movie_pca)

## Importance of components:
##          PC1          PC2          PC3          PC4          PC5          PC6
PC7
## Standard deviation      1.5915 1.4545 1.2085 1.00187 0.89876 0.88813
0.86509
## Proportion of Variance 0.2111 0.1763 0.1217 0.08365 0.06731 0.06573
0.06237
## Cumulative Proportion 0.2111 0.3874 0.5091 0.59272 0.66003 0.72576
0.78813
```

```
##
##          PC8      PC9      PC10     PC11     PC12
## Standard deviation 0.82041 0.77117 0.74181 0.6481 0.55169
## Proportion of Variance 0.05609 0.04956 0.04586 0.0350 0.02536
## Cumulative Proportion 0.84422 0.89378 0.93963 0.9746 1.00000
```

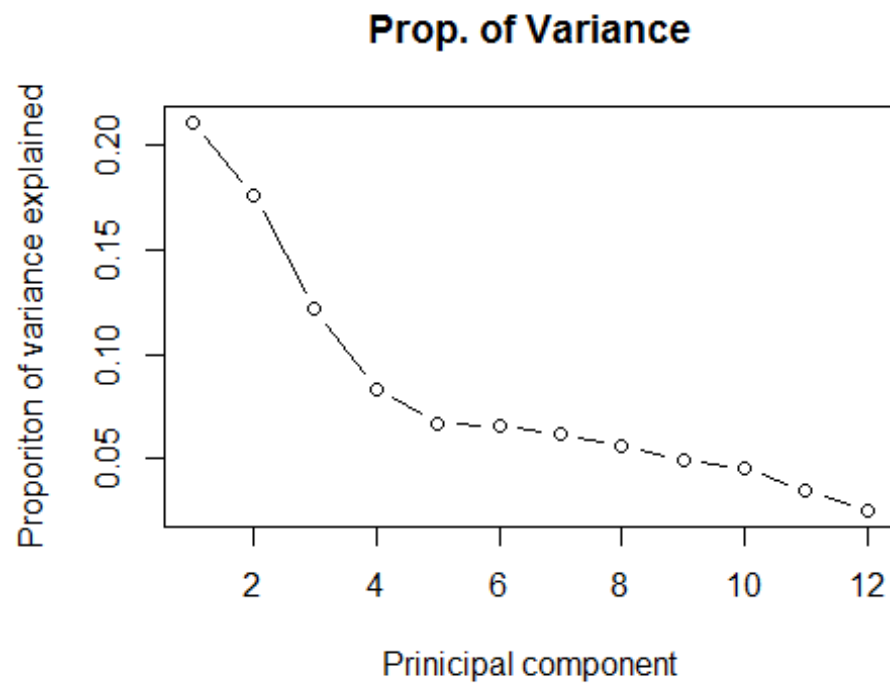
Plot a biplot to view components on n-dimensional plane

```
biplot(movie_pca, scale = 0, main = 'Principal componets')
```



##scree plot - to identify the elbow point

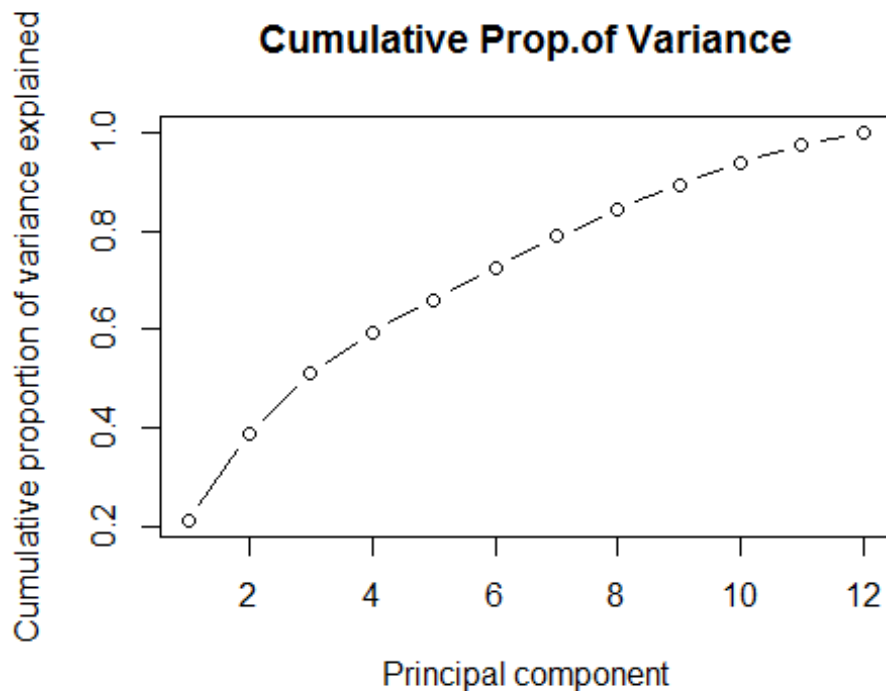
```
plot(propvar_movie,xlab = 'Prinicpal component',ylab = 'Proporiton of
variance explained',type = 'b', main = 'Prop. of Variance')
```



#The optimum number of components are ~ 5 i.e PC1 : PC5

cumulative scree plot

```
plot(cumvar_movie,xlab = 'Principal component',ylab = 'Cumulative proportion  
of variance explained',type = 'b', main = 'Cumulative Prop.of Variance')
```



#Approx: ~ 75% of the variance is explained by 8 components i.e PC1 to PC8

```
dim(music_data)
```

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

```
music_transformed <- data_transformed[,names(music_data)]
```

```
music_pca <- prcomp(music_transformed, scale= T)
```

```
names(music_pca)
```

```
## [1] "sdev" "rotation" "center" "scale" "x"
```

#rotation - represents the Loadings of each variables on components

```
music_pca$rotation
```

	PC1	PC2	PC3	PC4
## Music	-0.07348587	-0.06637258	0.20393520	-0.27285334
## Slow.songs.or.fast.songs	0.07909389	-0.03715789	0.33673187	-0.38692865
## Dance	0.10593604	-0.40686613	0.25266992	-0.18809943
## Folk	-0.23536427	-0.20395715	-0.17735902	-0.15261096
## Country	-0.22165587	-0.16378995	-0.08826069	-0.07278389
## Classical.music	-0.33427216	-0.11363745	-0.22530674	-0.19648989
## Musical	-0.21306339	-0.27761360	-0.19176926	-0.04813964
## Pop	0.07948200	-0.36864939	0.10610310	-0.08628967
## Rock	-0.31408320	0.16381242	0.24497619	-0.13760973
## Metal.or.Hardrock	-0.26788754	0.27493723	0.20888397	-0.20728971
## Punk	-0.25877936	0.22880832	0.32746991	-0.01776324
## Hiphop..Rap	0.14668504	-0.28399372	0.32226482	0.21487767

## Reggae..Ska	-0.17063495	-0.13437788	0.33404941	0.51149310
## Swing..Jazz	-0.32246691	-0.16834386	0.02878952	0.30067030
## Rock.n.roll	-0.35567777	-0.02784195	0.15705958	0.11136720
## Alternative	-0.29273952	0.12627082	0.16107910	0.04377580
## Latino	-0.11432176	-0.39837369	-0.03575959	0.17249177
## Techno..Trance	0.09648161	-0.23597716	0.27927306	-0.33307335
## Opera	-0.28734109	-0.14638664	-0.29342147	-0.21923779
##	PC5	PC6	PC7	PC8
## Music	0.32592753	-0.47761802	0.53270546	-0.19974554
## Slow.songs.or.fast.songs	-0.07423270	0.07177854	0.12369657	-0.47694134
## Dance	-0.05577652	0.06055365	-0.05942948	0.21887240
## Folk	-0.26544668	0.20827679	0.30003736	-0.17663204
## Country	-0.24305255	0.49366779	0.40798558	0.10106895
## Classical.music	-0.13812859	-0.25098533	-0.22358870	-0.01890419
## Musical	0.32990755	0.04105766	-0.22303756	-0.29312977
## Pop	0.46499371	0.15721120	-0.19059791	0.13285282
## Rock	0.31499048	0.14180077	-0.08393359	0.20370011
## Metal.or.Hardrock	-0.07920439	0.21455617	-0.19275758	-0.01276431
## Punk	0.03787992	0.19456396	-0.23696787	-0.16828955
## Hiphop..Rap	-0.14591724	-0.05330011	-0.16891849	-0.21654015
## Reggae..Ska	-0.18318806	0.01510496	-0.02608224	-0.33284595
## Swing..Jazz	-0.09855736	-0.27677329	0.06829684	0.13525914
## Rock.n.roll	0.15407054	0.08154945	0.17375175	0.27605014
## Alternative	-0.10328380	-0.37016042	0.05279782	0.17869982
## Latino	0.16866976	0.09636174	0.05687371	0.07392996
## Techno..Trance	-0.40272503	-0.16159061	-0.17221071	0.36838201
## Opera	-0.12224120	-0.15441720	-0.32365349	-0.21004774
##	PC9	PC10	PC11	
PC12				
## Music	0.407713408	-0.08118246	0.163265029	-
0.046192469				
## Slow.songs.or.fast.songs	-0.626930018	-0.15449805	-0.117591247	
0.142963357				
## Dance	-0.005060525	0.15836972	-0.001963950	
0.034498068				
## Folk	0.101978314	0.59768376	-0.044187886	
0.266968801				
## Country	0.187947460	-0.34757560	-0.176241400	-
0.296622461				
## Classical.music	0.074840699	-0.11353868	-0.197110022	
0.215496597				
## Musical	-0.127942176	-0.06550861	0.043660995	-
0.530976566				
## Pop	0.089375763	0.17576542	-0.360479495	
0.206155249				
## Rock	0.093377372	-0.01603115	0.012947358	
0.194189967				
## Metal.or.Hardrock	0.095803135	-0.09200573	0.378608727	
0.148009287				
## Punk	0.188934780	0.18211426	-0.009197961	-

0.136728601			
## Hip-hop..Rap	0.291127995	-0.19951768	-0.212893673
0.016816754			
## Reggae..Ska	0.109227439	0.09666073	0.055259444 -
0.033726863			
## Swing..Jazz	-0.180981008	-0.25859960	0.011764742
0.393901884			
## Rock.n.roll	-0.227212134	-0.21678701	-0.222790454 -
0.050344008			
## Alternative	-0.252502677	0.43193360	-0.239920191 -
0.392991837			
## Latino	-0.233292591	0.10051712	0.644329108
0.009843656			
## Techno..Trance	0.049583019	-0.05716672	0.202874089 -
0.232242161			
## Opera	0.098588439	-0.11619488	0.012943917
0.014313820			
##	PC13	PC14	PC15
PC16			
## Music	0.049792397	-0.02112800	-0.021711972 -
0.024489556			
## Slow.songs.or.fast.songs	0.003296287	-0.12450736	0.072331948 -
0.011696895			
## Dance	0.305288988	0.10278145	-0.636427969
0.312758934			
## Folk	-0.152396906	0.33786207	0.078929868 -
0.115326873			
## Country	0.033698701	-0.30836572	0.067818499
0.203982029			
## Classical.music	-0.081254762	-0.22963497	0.078587122
0.183762909			
## Musical	0.012225873	0.44293088	0.107137446
0.250791103			
## Pop	0.039315867	-0.21197082	0.351688766 -
0.023506412			
## Rock	-0.019546117	0.05416288	0.143052503
0.071888142			
## Metal.or.Hardrock	-0.239674138	0.16759903	0.028215133
0.274542443			
## Punk	0.158206848	-0.23587526	-0.223276307 -
0.151087803			
## Hip-hop..Rap	-0.650067679	0.07539266	-0.148338423 -
0.004739777			
## Reggae..Ska	0.409507673	-0.01913020	0.255708315 -
0.040735037			
## Swing..Jazz	0.161510400	0.13579529	0.038880226
0.271349720			
## Rock.n.roll	-0.090920432	0.29860848	-0.205885619 -
0.545870152			
## Alternative	-0.244048277	-0.23543333	-0.004240393

```

0.204720244
## Latino -0.277520061 -0.40294002 -0.040696024 -
0.117740140
## Techno..Trance 0.085649349 0.14642498 0.421244095 -
0.240536719
## Opera 0.142288289 -0.17563687 -0.235717510 -
0.398788272
## PC17 PC18 PC19
## Music -0.066875434 2.567921e-02 0.07345943
## Slow.songs.or.fast.songs 0.033053577 -1.410316e-02 -0.04974487
## Dance 0.140182891 -2.906194e-02 0.11125996
## Folk -0.051850387 -3.490200e-02 -0.10739444
## Country 0.006149017 9.814582e-02 -0.06497751
## Classical.music 0.028773192 -5.759827e-01 0.34337770
## Musical -0.066238082 -8.145992e-02 -0.07751286
## Pop -0.216265041 2.928257e-01 0.17206160
## Rock 0.583382189 -1.360268e-01 -0.43938942
## Metal.or.Hardrock -0.119696729 3.588220e-01 0.43444249
## Punk -0.523054151 -2.601956e-01 -0.25901292
## Hiphop..Rap 0.056186635 8.017653e-03 -0.15671264
## Reggae..Ska 0.302779309 9.179124e-05 0.28654917
## Swing..Jazz -0.369360329 1.762901e-01 -0.35365088
## Rock.n.roll -0.039227337 -8.664268e-02 0.31816965
## Alternative 0.084954381 2.479808e-01 0.01148615
## Latino 0.008442401 -1.256294e-01 -0.01498152
## Techno..Trance -0.095918082 -8.103699e-02 -0.10837567
## Opera 0.206245145 4.752428e-01 -0.12849210

# Calculate the variance explained by each component
eigen_music <- music_pca$sdev^2
names(eigen_music) <- paste("PC",1:19,sep="")
eigen_music

## PC1 PC2 PC3 PC4 PC5 PC6 PC7
PC8
## 3.8053009 2.6634739 1.9555901 1.1307190 1.1176877 1.0349666 0.8881561
0.8590806
## PC9 PC10 PC11 PC12 PC13 PC14 PC15
PC16
## 0.7991110 0.6495392 0.6424224 0.5720013 0.4996263 0.4616361 0.4427409
0.4127463
## PC17 PC18 PC19
## 0.3801299 0.3576233 0.3274485

sum_music <- sum(eigen_music)
sum_music

## [1] 19

## Proportion of variance explained by each component

```



```

propvar_music <- eigen_music/sum_music
propvar_music

##          PC1          PC2          PC3          PC4          PC5          PC6
PC7
## 0.20027900 0.14018284 0.10292580 0.05951152 0.05882567 0.05447193
0.04674506
##          PC8          PC9          PC10          PC11          PC12          PC13
PC14
## 0.04521477 0.04205847 0.03418627 0.03381170 0.03010533 0.02629612
0.02429663
##          PC15          PC16          PC17          PC18          PC19
## 0.02330215 0.02172349 0.02000684 0.01882228 0.01723413

cumvar_music <- cumsum(propvar_music)
cumvar_music

##          PC1          PC2          PC3          PC4          PC5          PC6          PC7
PC8
## 0.2002790 0.3404618 0.4433876 0.5028992 0.5617248 0.6161967 0.6629418
0.7081566
##          PC9          PC10          PC11          PC12          PC13          PC14          PC15
PC16
## 0.7502151 0.7844013 0.8182130 0.8483184 0.8746145 0.8989111 0.9222133
0.9439368
##          PC17          PC18          PC19
## 0.9639436 0.9827659 1.0000000

matlambdas_music <- rbind(eigen_music,propvar_music,cumvar_music)
rownames(matlambdas_music) <- c("Eigenvalues","Prop. variance","Cum. prop.
variance")
round(matlambdas_music,4)

##          PC1          PC2          PC3          PC4          PC5          PC6          PC7
PC8
## Eigenvalues          3.8053 2.6635 1.9556 1.1307 1.1177 1.0350 0.8882
0.8591
## Prop. variance          0.2003 0.1402 0.1029 0.0595 0.0588 0.0545 0.0467
0.0452
## Cum. prop. variance 0.2003 0.3405 0.4434 0.5029 0.5617 0.6162 0.6629
0.7082
##          PC9          PC10          PC11          PC12          PC13          PC14          PC15
PC16
## Eigenvalues          0.7991 0.6495 0.6424 0.5720 0.4996 0.4616 0.4427
0.4127
## Prop. variance          0.0421 0.0342 0.0338 0.0301 0.0263 0.0243 0.0233
0.0217
## Cum. prop. variance 0.7502 0.7844 0.8182 0.8483 0.8746 0.8989 0.9222
0.9439
##          PC17          PC18          PC19
## Eigenvalues          0.3801 0.3576 0.3274

```

```

## Prop. variance      0.0200 0.0188 0.0172
## Cum. prop. variance 0.9639 0.9828 1.0000

summary(music_pca)

## Importance of components:
##
##          PC1      PC2      PC3      PC4      PC5      PC6
PC7
## Standard deviation    1.9507 1.6320 1.3984 1.06335 1.05721 1.01733
0.94242
## Proportion of Variance 0.2003 0.1402 0.1029 0.05951 0.05883 0.05447
0.04675
## Cumulative Proportion 0.2003 0.3405 0.4434 0.50290 0.56172 0.61620
0.66294
##
##          PC8      PC9      PC10     PC11     PC12     PC13
PC14
## Standard deviation    0.92687 0.89393 0.80594 0.80151 0.75631 0.7068
0.6794
## Proportion of Variance 0.04521 0.04206 0.03419 0.03381 0.03011 0.0263
0.0243
## Cumulative Proportion 0.70816 0.75022 0.78440 0.81821 0.84832 0.8746
0.8989
##
##          PC15     PC16     PC17     PC18     PC19
## Standard deviation 0.6654 0.64245 0.61655 0.59802 0.57223
## Proportion of Variance 0.0233 0.02172 0.02001 0.01882 0.01723
## Cumulative Proportion 0.9222 0.94394 0.96394 0.98277 1.00000

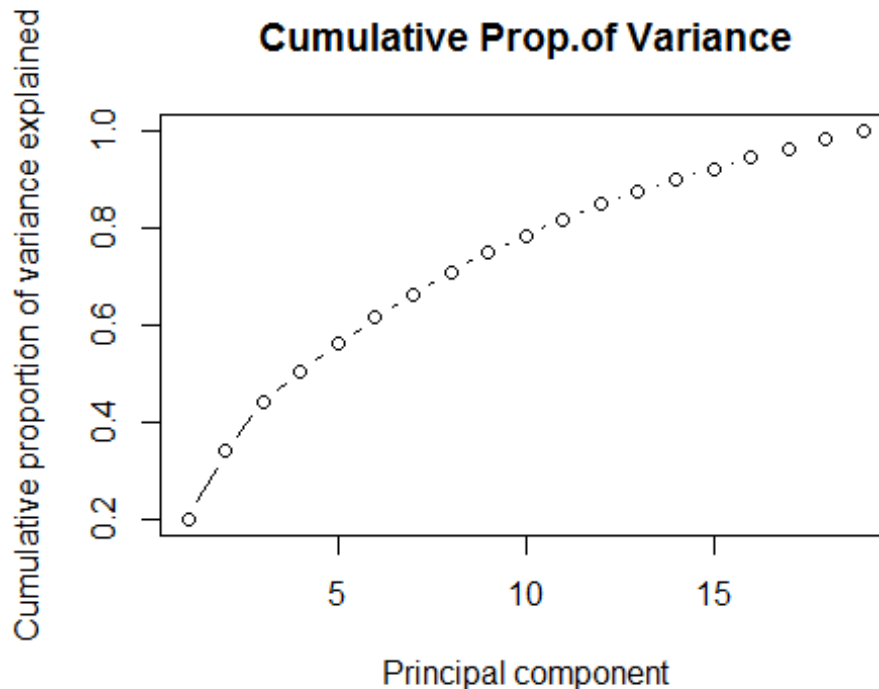
## Plot a biplot to view components on n-dimensional plane
biplot(music_pca, scale = 0, main = 'Principal componets')

```


#The optimum number of components are ~ 4 i.e PC1 : PC4

cumulative scree plot

```
plot(cumvar_music,xlab = 'Principal component',ylab = 'Cumulative proportion  
of variance explained',type = 'b', main = 'Cumulative Prop.of Variance')
```



#Approx: ~ 75% of the variance is explained by 8 components i.e PC1 to PC8

```
library(cluster)
```

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

Data Cluster

Movie Preferences Cluster (Hierarcical Clustering)

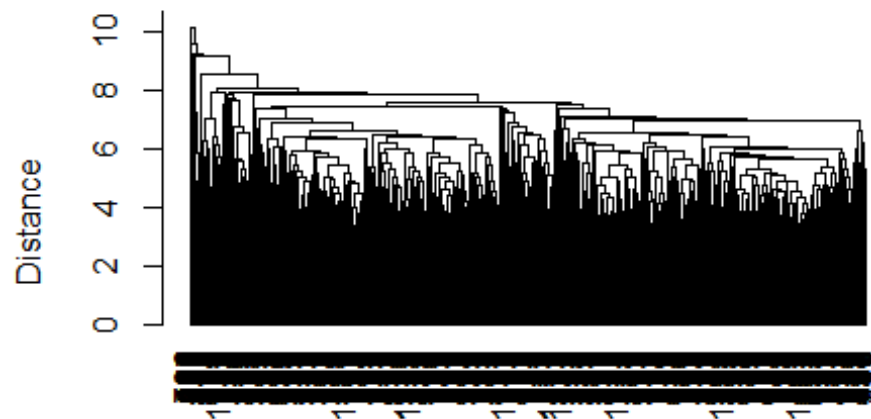
```
dim(music_trans)
```

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

Given to our variables value is the same dimension, We don't need to normalize the data and form the cluster fit directly.

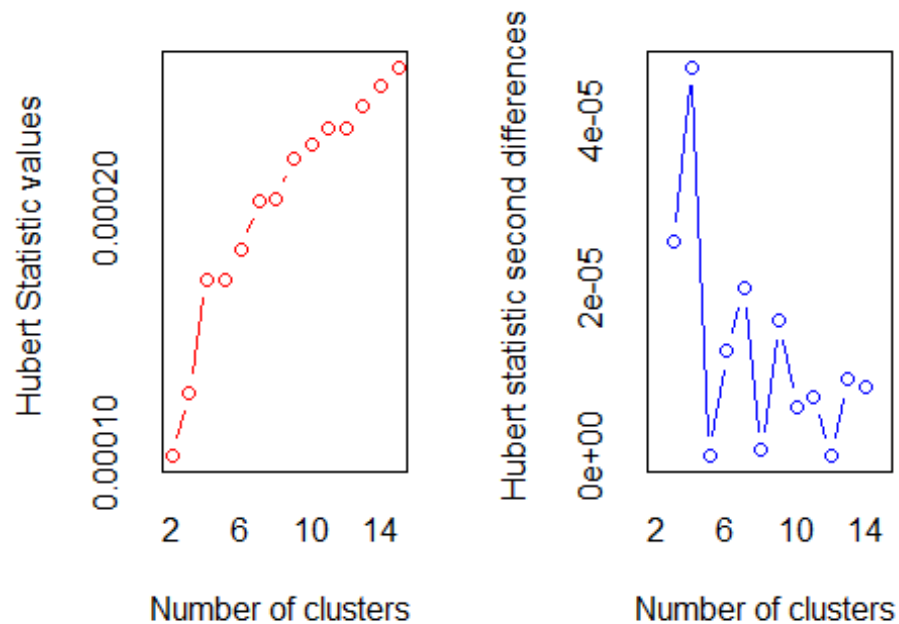
```
dist.mat5 = music_trans  
mat5.nn <- hclust(dist(dist.mat5, method = "euclidean"), method = "average")  
plot(mat5.nn, hang=-1,xlab="Object",ylab="Distance",  
main="Dendrogram. Nearest neighbor linkage")
```

Dendrogram. Nearest neighbor linkage

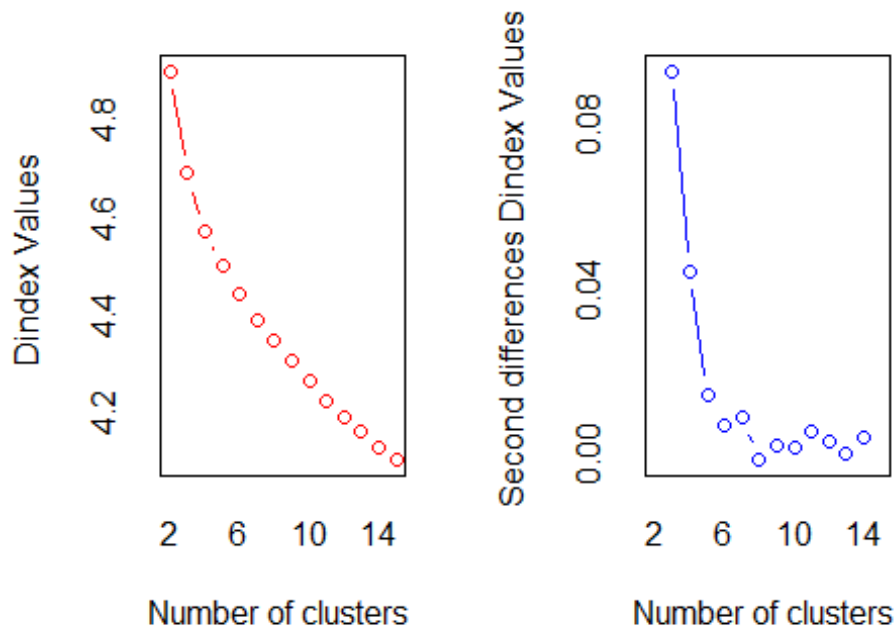


Object
hclust (*, "average")

```
#use agglomerative hierarchical clustering to cluster  
hc_mod = hclust(dist(music_trans, method = "euclidean"), method = "ward.D")  
plot(hc_mod, hang = -0.01, cex = 0.7)
```

```
## *** : The Hubert index is a graphical method of determining the number of
clusters.
##           In the plot of Hubert index, we seek a significant knee
that corresponds to a
##           significant increase of the value of the measure i.e the
significant peak in Hubert
##           index second differences plot.
##
```



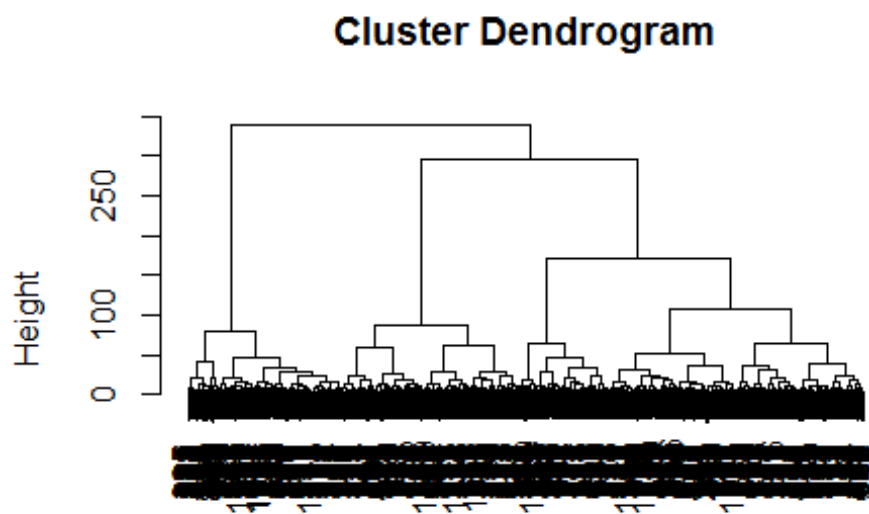
```
## *** : The D index is a graphical method of determining the number of
clusters.
##           In the plot of D index, we seek a significant knee (the
significant peak in Dindex
##           second differences plot) that corresponds to a significant
increase of the value of
##           the measure.
##
## *****
## * Among all indices:
## * 4 proposed 2 as the best number of clusters
## * 11 proposed 3 as the best number of clusters
## * 4 proposed 4 as the best number of clusters
## * 1 proposed 9 as the best number of clusters
## * 1 proposed 10 as the best number of clusters
## * 2 proposed 15 as the best number of clusters
##
##           ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is  3
##
## *****
```

Next step is cut the tree to into different clusters.

```
fit_model = cutree(hc_mod, k=3)
```



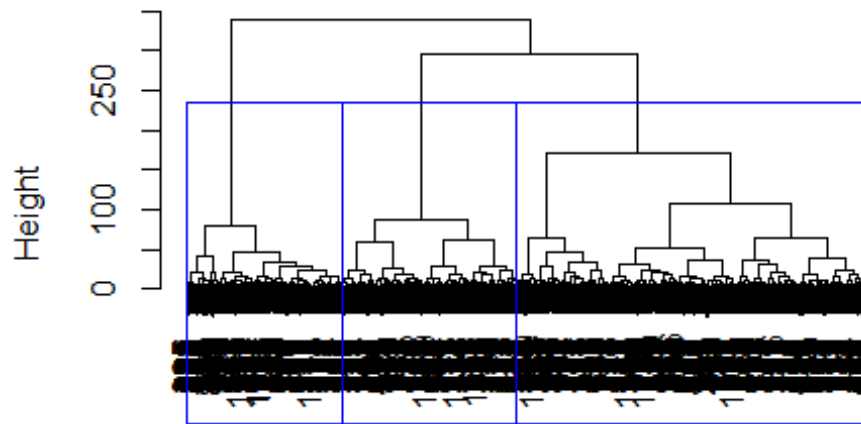
```
plot(hc_mod)
```



```
dist(music_trans, method = "euclidean")  
hclust (*, "ward.D")
```

```
plot(hc_mod)  
rect.hclust(hc_mod, k= 3, border = "blue")
```

Cluster Dendrogram



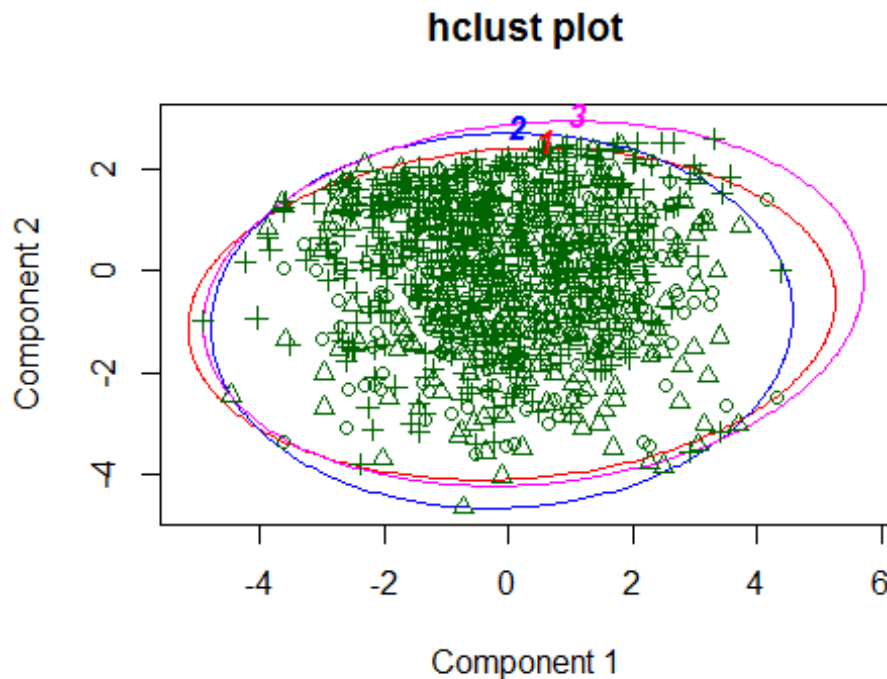
```
dist(movie_trans, method = "euclidean")
hclust(*, "ward.D")
```

#segment inspection

```
aggregate(movie_trans, by=list(cluster=fit_model), mean)
```

```
## cluster Movies Horror Thriller Comedy Romantic Sci.fi War
## 1 1 4.584416 3.034632 3.463203 4.597403 3.316017 3.017316 3.225108
## 2 2 4.616858 2.823755 3.517241 4.298851 3.042146 3.210728 3.272031
## 3 3 4.623552 2.667954 3.277992 4.550193 3.791506 3.098456 3.065637
## Fantasy.Fairy.tales Animated Documentary Western Action
## 1 3.376623 3.519481 3.406926 1.874459 3.679654
## 2 3.513410 3.689655 3.724138 2.130268 3.440613
## 3 4.034749 3.959459 3.706564 2.239382 3.515444
```

```
clusplot(movie_trans, fit_model, color=TRUE, labels = 4, lines=0, main
="hclust plot")
```

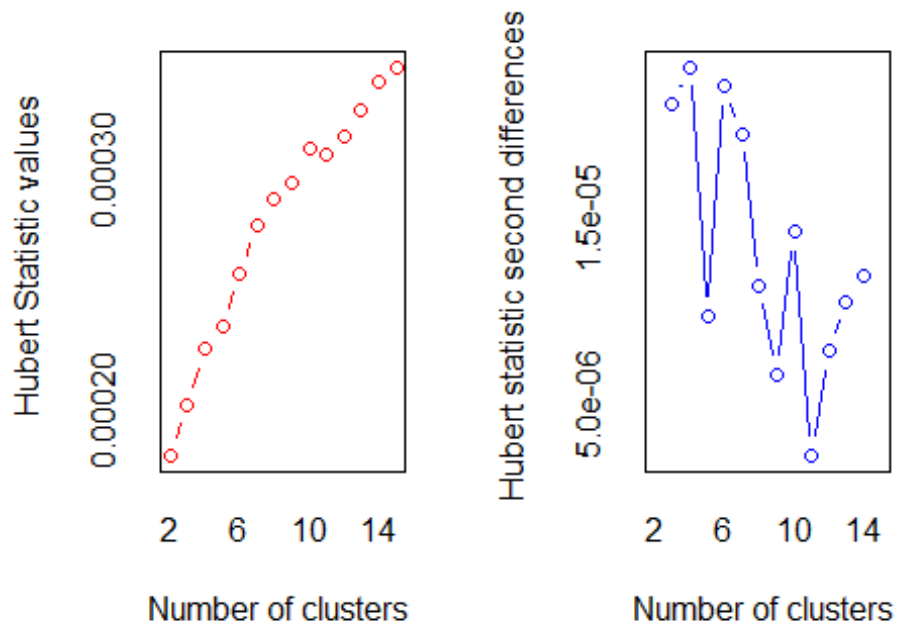


These two components explain 38.74 % of the point variab

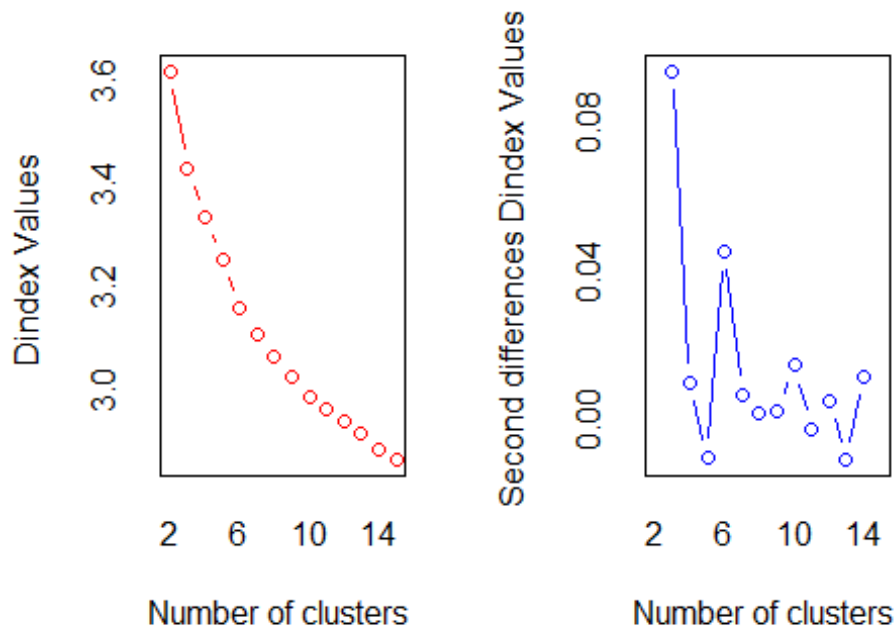
The cluster is cut clearly, but as the segment description shows that this cluster is kind of odd. Three groups are unbalance distribution, the first variable, which is general in this topic, has only a value.

K-means cluster

```
#Decide the number of cluster before starting the K-mean fit form.  
nb_clust2 = NbClust(movie_trans, min.nc=2, max.nc=15, method="kmeans")
```



```
## *** : The Hubert index is a graphical method of determining the number of
clusters.
##           In the plot of Hubert index, we seek a significant knee
that corresponds to a
##           significant increase of the value of the measure i.e the
significant peak in Hubert
##           index second differences plot.
##
```



```
## *** : The D index is a graphical method of determining the number of
clusters.
##           In the plot of D index, we seek a significant knee (the
significant peak in Dindex
##           second differences plot) that corresponds to a significant
increase of the value of
##           the measure.
##
## *****
## * Among all indices:
## * 7 proposed 2 as the best number of clusters
## * 11 proposed 3 as the best number of clusters
## * 1 proposed 6 as the best number of clusters
## * 1 proposed 9 as the best number of clusters
## * 1 proposed 10 as the best number of clusters
## * 2 proposed 15 as the best number of clusters
##
##           ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is  3
##
## *****
#table(nc2$Best.n[1,])
```

It's still 3 with Kmeans.

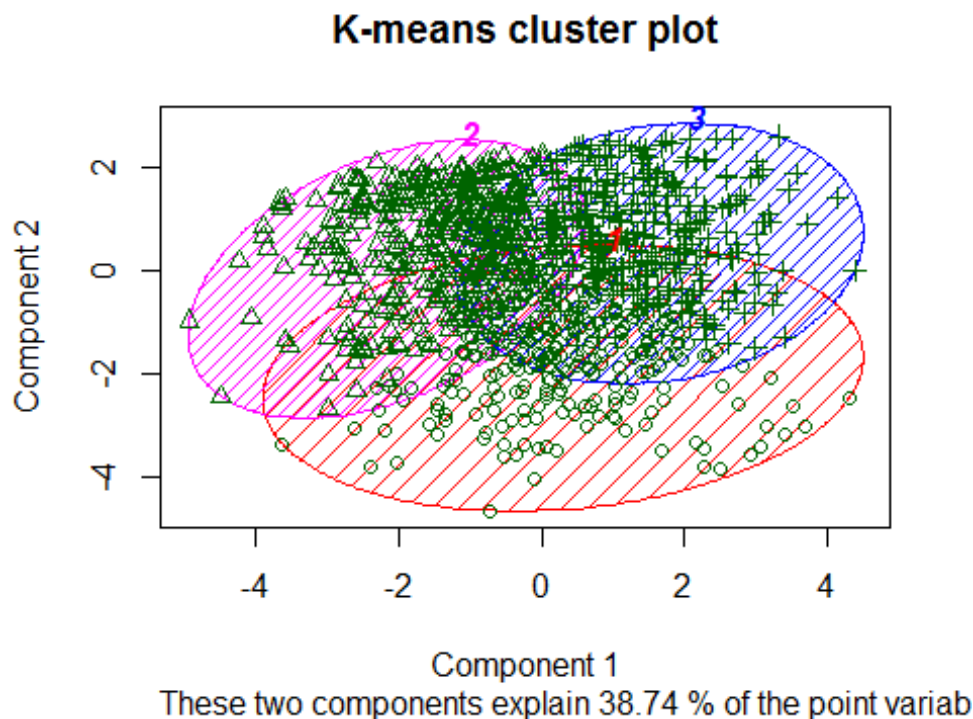
```

set.seed(101)
km_mo <- kmeans(movie_trans, 3)
aggregate(movie_trans, by=list(cluster=km_mo$cluster), mean)

##   cluster  Movies  Horror Thriller  Comedy Romantic  Sci.fi  War
## 1      1 4.449827 2.913495 3.494810 4.145329 2.837370 2.951557 3.173010
## 2      2 4.576316 2.176316 2.721053 4.607895 4.068421 2.378947 2.334211
## 3      3 4.791789 3.375367 4.023460 4.668622 3.395894 4.055718 4.055718
##   Fantasy.Fairy.tales Animated Documentary Western Action
## 1      2.484429 2.422145      3.346021 2.086505 3.477509
## 2      4.421053 4.350000      3.463158 1.600000 2.873684
## 3      4.073314 4.322581      4.093842 2.750733 4.316716

clusplot(movie_trans, km_mo$cluster, color=TRUE, shade=TRUE,
          labels=4, lines=0, main="K-means cluster plot")

```



According to the kmeans cluster result, I can barely name the three cluster as:

cluster 3: movie enthduasiasts, love all kind of movie except romantic and Fantasy/Fairy tales
 cluster 1: normal movie consumer, focus on romantic, Fantasy/Fairy tales, comedy, and no feeling on horror, thriller, war, and western
 cluster 2: specific kind of movie lover, focus on war, thriller, action

comparing the three clusters demographics characters:

```

demo_trans = data_transformed[,names(demo_data)]
aggregate(demo_trans[,1:4], by = list(km_mo$cluster), mean)

```

```
##   Group.1      Age   Height   Weight Number.of.siblings
## 1      1 20.67474 175.3253 69.12111      1.231834
## 2      2 20.12895 169.1082 60.02105      1.300000
## 3      3 20.37537 177.0254 70.24047      1.351906

round(prop.table(table(km_mo$cluster, demo_trans$Gender),1),2)

##
##           female male
## 1 0.01    0.45 0.54
## 2 0.01    0.90 0.10
## 3 0.01    0.35 0.64

round(prop.table(table(demo_trans$Education, km_mo$cluster),2),2)

##
##                1      2      3
##                0.00 0.00 0.00
## college/bachelor degree      0.22 0.22 0.20
## currently a primary school pupil 0.01 0.01 0.01
## doctorate degree            0.01 0.00 0.00
## masters degree              0.10 0.08 0.07
## primary school              0.06 0.09 0.09
## secondary school            0.60 0.62 0.63

round(prop.table(table(km_mo$cluster, demo_data$Village...town),1),2)

##
##           city village
## 1 0.01 0.72    0.27
## 2 0.00 0.69    0.31
## 3 0.00 0.70    0.30

round(prop.table(table(km_mo$cluster,
demo_trans$House...block.of.flats),1),2)

##
##           block of flats house/bungalow
## 1 0.01      0.60      0.39
## 2 0.01      0.58      0.42
## 3 0.00      0.59      0.41
```

As you can see, the most difference between them is gender:

cluster 3: movie enthduasiasts are mainly guys cluster 1: normal movie consumers are mostly girls cluster 2: specific kind of movie lover are more elder but quite equal on gender