**Solution Provider Name: Dipali Bharat Jadhav**

**College: V. P. Institute of Management Studies & Research, Sangli  
R Version: 4.0.3**

**R Studio Version: 1.1.463 Operation System: Windows 10 OS**

**Exercise No. 5**

# Install package ggplot and import it

* install.packages("ggplot2")

package ‘ggplot2’ successfully installed

* library(ggplot2)

**Data Insert**

Use c() function to insert data.

**Syntax**: vector\_name <- c(“value1”, “value2”,……,“value\_n”) Where, vector\_name is used to define name of the vector.

c() is used to create a function for passing values or arguments. The value may be an integer, string, float values, etc.

# Command

>**App** <- c ("Flow Free","Zombie Catchers","Gems or jewels ?","Candy Crush Jelly Saga","Jewels Star: OZ adventure","Hungry Shark Evolution","Pdf Book Download ","Free Book Reader","eBoox new","Golden Dictionary (EN-AR)","All Language Translator Free","Azpen eReader","ADP Mobile Solutions","OfficeSuite","Job Search by ZipRecruiter","Google Primer","SuperLivePro","TED","English Communication","Khan Academy","Learn English with Wlingua","Princess Coloring Book","English Grammar Test","Speed Reading","English for beginners")

>**Category** <-

c ("GAME","GAME","GAME","GAME","GAME","GAME","BOOKS","

BOOKS","BOOKS","BOOKS","BOOKS","BOOKS","BUSINESS","BU SINESS","BUSINESS","BUSINESS","BUSINESS","EDUCATION","E DUCATION","EDUCATION","EDUCATION","EDUCATION","EDU CATION","EDUCATION","EDUCATION")

>**Rating**

<-c(4.3,4.7,3.7,4.3,2.5,4.5,4.4,3.4,4.9,4.4,3.4,3.5,1.3,4.3,4.8,2.4,4.3,3.6,4.7, 4.6,2.7,4.5,1.8,4.6,2.6)

>**Reviews**

<-c(1295625,990796,171448,1300619,21892,6074627,1322,1680,2739,512 69,30105,156,85185,1002861,148945,62272,46353,181893,2544,85375,

314299,9770, 4075,10611,9321)

>**Size.mb**

<-c(11,75,13,78,14,100,3.5,4,21,6.1,2.1,42,29,35,25,18,21,18,18,21,3.3,39, 5.1,11,27)

**data. frame()**this function is used to storing data tables in tabular format.

AppData <- data.frame(App,Category,Rating,Reviews,Size.mb)

**View()**function is displays data of the data frame which is created.

**Syntax: View(x)**Where x is object or data frame name View(AppData)

**Write.csv()**function is used to create or store data in .csv file. write.csv(AppData,"NewAppData.csv")

**read.csv()**used to read the CSV file which is created.

NewAppData <- read.csv("NewAppData.csv")

**Print()**function is used to display data.

print(NewAppData)

# Draw pie chat for showing size require for different education application

**Command**

* subset.data.frame(AppData,Category=="EDUCATION")

|  |  |  |  |
| --- | --- | --- | --- |
| App Category | Rating | Reviews | Size.mb |
| 18 TED EDUCATION | 3.6 | 181893 | 18.0 |
| 19 English Communication EDUCATION | 4.7 | 2544 | 18.0 |
| 20 Khan Academy EDUCATION | 4.6 | 85375 | 21.0 |
| 21 Learn English with Wlingua EDUCATION | 2.7 | 314299 | 3.3 |
| 22 Princess Coloring Book EDUCATION | 4.5 | 9770 | 39.0 |
| 23 English Grammar Test EDUCATION | 1.8 | 4075 | 5.1 |
| 24 Speed Reading EDUCATION | 4.6 | 10611 | 11.0 |
| 25 English for beginners EDUCATION | 2.6 | 9321 | 27.0 |

Subset.data.frame() this function is used to select variables and observation

# Command

* x <- c (18,18,21,3.3,39,5.1,11,27)
* labels <- c("TED","English Communication","Khan Academy","Learn Englis h with Wlingua","Princess Coloring Book","English Grammer Test","Speed Re ading","English for beginners")
* pieper <- round(100 \* x/sum(x),1)
* pie(x,labels,main = "Education pie chart",col = rainbow(length(x)))

**Pie chart**: In R programming using the pie() function to create a pie chart. A pie chart is a graphical representation of slices of a circle with different colors.

# Syntax: pie(x, labels, radius, main, col, clockwise)

Where x is vector which is present in number data type

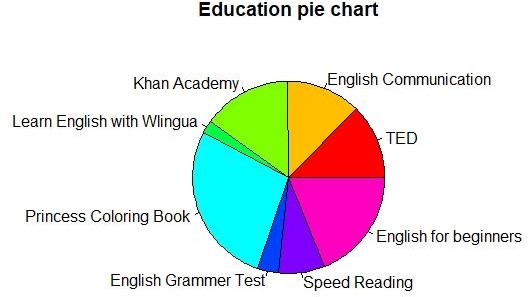
Labels indicate name (description) of slices in pie chart

Radius indicates radius of the circle

Main indicates title of pie chart

Col indicates colors of slices of circle Clockwise display direction of pie chart

# Pie Chart:



1. **Draw histogram for ratings of e-books apps with proper labeling.**

* subset.data.frame(AppData,Category=="BOOKS")

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| App | Category | Rating | Reviews | Size.mb |
| 7 Pdf Book Download | BOOKS | 4.4 | 1322 | 3.5 |
| 8 Free Book Reader | BOOKS | 3.4 | 1680 | 4.0 |
| 9 eBoox new | BOOKS | 4.9 | 2739 | 21.0 |
| 10 Golden Dictionary (EN-AR) | BOOKS | 4.4 | 51269 | 6.1 |
| 11 All Language TranslatorFree | BOOKS | 3.4 | 30105 | 2.1 |
| 12 Azpen eReader | BOOKS | 3.5 | 156 | 42.0 |

**Histogram**: Histogram represents the frequency of values in variables. The histogram is similar to a bar chart but it displays a continuous range.

In R programming create a histogram using the hist() function.

**Syntax**: hist(v, main, xlab, xlim, ylim, breaks, col, border) Where V is vector which contain numeric data or value

   Main indicates the title of the histogram.

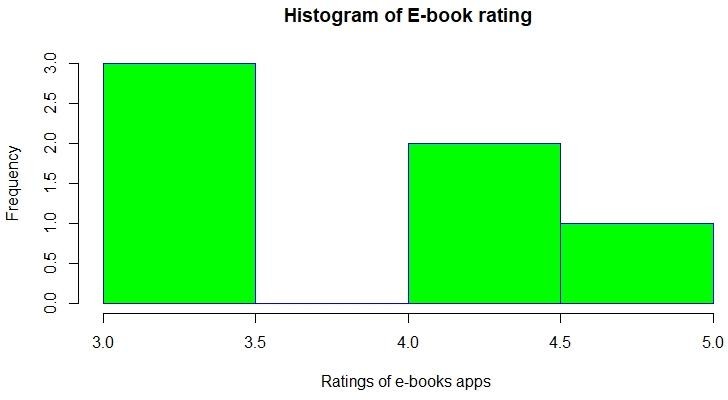
   Xlab indicates description /labels of x-axis Xlim it indicates x-axis limit

   Ylim is used to specific range of y-axis Col is used to display colors of histogram Border is used to set border color

# Command

* v <- c(4.4,3.4,4.9,4.4,3.4,3.5)
* hist(v, xlab = "Ratings of e-books apps ", ylab=" Frequency", col = "green", b order = "blue", main = "Histogram of E-book rating")

# Histogram:



1. **Draw bar chart for showing app-wise rating with proper labeling.**

**Bar plot:**Bar plot represents data in a rectangle bar with its length. In R programming barplot() function is used to create a bar chart.

**Syntax**: barplot(H,xlab,ylab,main, names.arg,col)

Where, H is value of bar chart which is present in number format Xlab is label of x-asix

ylab is label of y-axis

Main indicates title of bar cart

names. arg indicates vector name in each bar

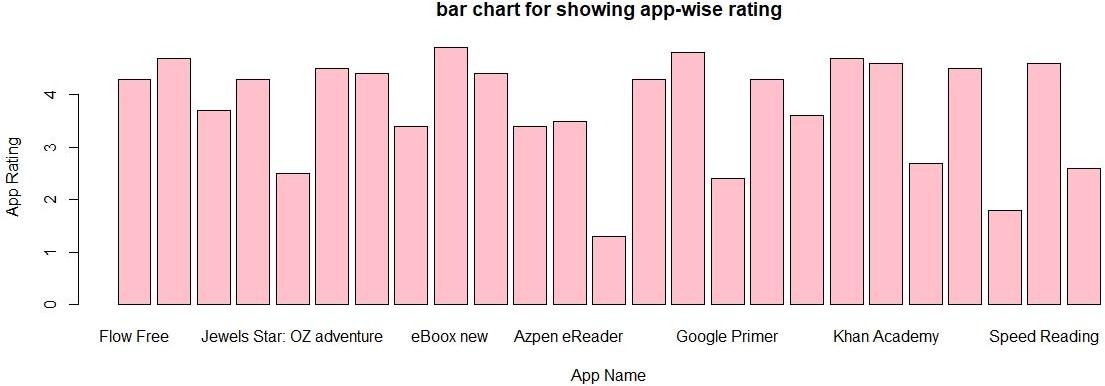
col is used to display different colors in barplot

# Command

* barplot(AppData$Rating,ylab = "App Rating",xlab = "App Name",main

= "bar chart for showing app-wise rating",col = "pink",names.arg = AppD ata$App)

# Bar Plot:



1. **Reveal the relationship between rating and reviews of business apps.**

lm() the function is used to create a relationship between two variables.

**Syntax**: lm(formula, data)

Command

* + Rating <- c (1.3,4.3,4.8,2.4,4.3)
  + Reviews <- c(85185,1002861,148945,62272,46353)
  + relation <- lm(Rating ~ Reviews)
  + print(relation)

Call:

lm(formula = Rating ~ Reviews) Coefficients:

(Intercept) Reviews 3.069e+00 1.303e-06

**Correlation:**Correlation is part of data visualization in R. Use the cor. test() function to find a correlation between two vectors.

**Relation Type:**

* + Rating <- c (1.3,4.3,4.8,2.4,4.3)
  + Reviews <- c(85185,1002861,148945,62272,46353)
  + cor.test(Rating, Reviews)

# Pearson's product-moment correlation

data: Rating and Reviews

t = 0.66435, df = 3, p-value = 0.554

Alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval:

-0.7662471 0.9425738

Sample estimates: cor

0.358122

# Draw bar char for showing category wise application count.

# Description:

table() function is used to find the frequency of a particular column.

In this example it display different category and frequency using table() function

**Bar plot:**Bar plot represents data in a rectangle bar with its length. In R programming barplot() function is used to create a bar chart.

**Syntax**: barplot(H,xlab,ylab,main, names.arg,col)

Where, H is value of bar chart which is present in number format Xlab is label of x-asix

ylab is label of y-axis

Main indicates title of bar cart

names. arg indicates vector name in each bar

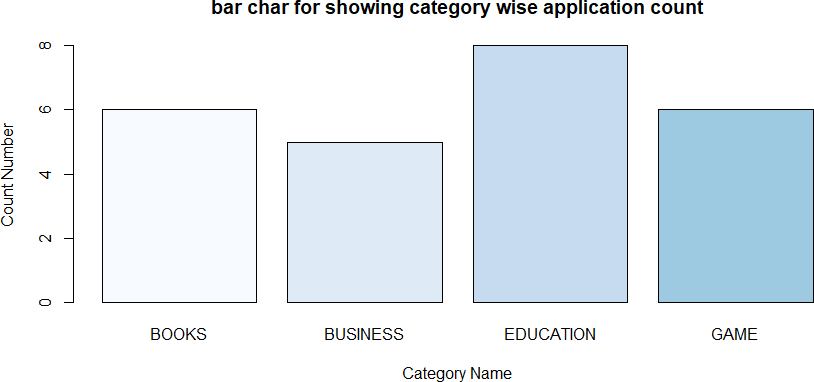
col is used to display different colors in barplot

# Command

* + table(Category) Category

BOOKS BUSINESS EDUCATION GAME 6 5 8 6

* + count=table(Category)
  + barplot(count,main = "bar char for showing category wise application coun t",col = blues9,ylab = "Count Number",xlab = "Category Name")



# Find the frequency distribution of app category.

# Description:

**Frequency Distribution:**Frequency distribution is the representation of data in tabular format and observation with a given interval.

**Using read.**table() function access data from the table and calculate frequency.

# ****Using transform****() function to transfer data or modified data frames. Syntax: transform(data, x = x\_transformed)

# Command:

* AppCategoryData <- read.table("NewAppData.csv",header = TRUE,sep = ",")
* table(AppCategoryScore)

AppCategoryScore

BOOKS BUSINESS EDUCATION GAME

6 5 8 6

* AppCategoryScore <- AppData$Category
* transform(table(AppCategoryScore))

AppCategoryScore Freq

1. BOOKS 6
2. BUSINESS 5
3. EDUCATION 8
4. GAME 6

# Draw bar graph for showing rating of education system.

# Description:

**Subset()**function is a simple way to select variables.

**Data.frame()**function is used to create a frame with collection of variables. These two functions combine and easily select variables and create the frame.

# Command:

>subset.data.frame(AppData,Category=="EDUCATION")

|  |  |  |  |
| --- | --- | --- | --- |
| App Category | Rating | Reviews | Size.mb |
| 18 TED EDUCATION | 3.6 | 181893 | 18.0 |
| 19 English Communication EDUCATION | 4.7 | 2544 | 18.0 |
| 20 Khan Academy EDUCATION | 4.6 | 85375 | 21.0 |
| 21 Learn English with Wlingua EDUCATION | 2.7 | 314299 | 3.3 |
| 22 Princess Coloring Book EDUCATION | 4.5 | 9770 | 39.0 |
| 23 English Grammar Test EDUCATION | 1.8 | 4075 | 5.1 |
| 24 Speed Reading EDUCATION | 4.6 | 10611 | 11.0 |
| 25 English for beginners EDUCATION | 2.6 | 9321 | 27.0 |

# Description:

**Bar plot:**Bar plot represents data in a rectangle bar with its length. In R programming barplot() function is used to create a bar chart.

**Syntax**: barplot(H,xlab,ylab,main, names.arg,col)

Where, H is value of bar chart which is present in number format Xlab is label of x-asix

ylab is label of y-axis

Main indicates title of bar cart

names. arg indicates vector name in each bar

col is used to display different colors in barplot

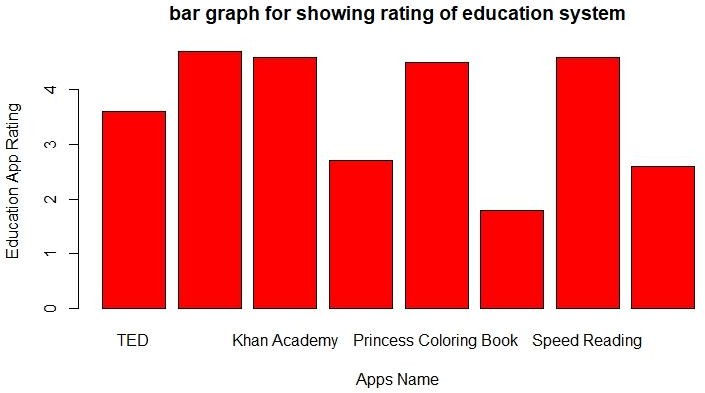
# Command:

>R <- c(3.6,4.7,4.6,2.7,4.5,1.8,4.6,2.6)

>labels <- c("TED"," English Communication","Khan Academy","Learn English with Wlingua","Princess Coloring Book","English Grammar Test","Speed Reading","English for beginners")

>barplot(R,names.arg = labels, ylab = "Education App Rating",xlab = "Apps Name", main = "bar graph for showing rating of education system",col = "red")

# Bar Plot:



1. **Find the relative frequency distribution of app categories.**

**Description:**

**Relative frequency distribution:** Relative frequency is the absolute frequency of that event divided by the total number of events. It represents the proportion of a particular data category present in the data vector.

# Syntax: table(x)/length(x)

# rf is a variable

**Command**

* rf <- table(AppData$Category)/length(AppData$Category)
* print(rf)

# Output

|  |  |  |  |
| --- | --- | --- | --- |
| BOOKS | BUSINESS | EDUCATION | GAME |
| 0.24 | 0.20 | 0.32 | 0.24 |

1. **Draw the scatter plot of reviews and rating given to education app. Does it revels the relationship between variables? If yes which type of relationship.**

**Description:**

**Scatter plot:**Scatterplots show many points plotted in the Cartesian plane. Each point represents the values of two variables. One variable is chosen in the horizontal axis and another in the vertical axis.

In R programming to create scatter plot using **plot()**function **Syntax: plot (x, y, main, xlab, ylab, xlim, ylim, axes)**Where, x is a data set which indicates horizontal coordinates

Y is a data set indicates vertical coordinates Main indicates title of scatter plot

Xlab indicates labels of horizontal coordinates ylab indicates labels of vertical coordinates

axes indicate whether both axes should be drawn on the plot.

# Command

* + View(AppData)
  + dim(AppData) [1] 25 5

# Description:

**Subset()**function is a simple way to select variables.

**Data.frame()**function is used to create a frame with collection of variables. These two functions combine and easily select variables and create a frame.

>subset.data.frame(AppData,Category=="EDUCATION")

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| App | Category | Rating | Reviews | Size.mb |
| 18 | TED EDUCATION | 3.6 | 181893 | 18.0 |
| 19 | English Communication EDUCATION | 4.7 | 2544 | 18.0 |
| 20 | Khan Academy EDUCATION | 4.6 | 85375 | 21.0 |
| 21 | Learn English with Wlingua EDUCATION | 2.7 | 314299 | 3.3 |
| 22 | Princess Coloring Book EDUCATION | 4.5 | 9770 | 39.0 |
| 23 | English Grammar Test EDUCATION | 1.8 | 4075 | 5.1 |
| 24 | Speed Reading EDUCATION | 4.6 | 10611 | 11.0 |
| 25 | English for beginners EDUCATION | 2.6 | 9321 | 27.0 |

# Command

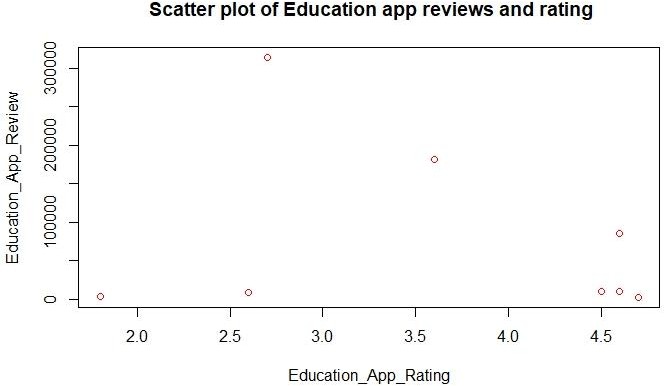
>Education\_App\_Rating <- c(3.6,4.7,4.6,2.7,4.5,1.8,4.6,2.6)

>Education\_App\_Review <- c(181893,2544,85375,314299,9770,4075,10611

,9321)

>plot(x=Education\_App\_Rating,y=Education\_App\_Review,main = " Scatter plot of Education app reviews and rating",col="red")

# Scatter\_Diagram



**Relationship between Variables:**

**Correlation:**Correlation is part of data visualization in R. Use the cor. test() function to find the correlation between two vectors.

**Pearson's product-moment correlation:**It is the statistical correlation that measures the linear correlation between two variables.

# Command:

* cor.test(AppData$Rating,AppData$Reviews)

**Pearson's product-moment correlation data**: AppData$Rating and AppData$Reviews t = 1.1179, df = 23, p-value = 0.2751

Alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval:

-0.1846795 0.5709346

sample estimates: cor0.2270175

# Find the cumulative frequency distribution of the rating in AppStore

# Description:

Using the **cumsum()**function to calculate cumulative frequency.

# Syntax: cumsum(x)

Where x is numeric object

# Command:

* + cumsum(AppData$Rating)

# Output:

[1] 4.3 9.0 12.7 17.0 19.5 24.0 28.4 31.8 36.7 41.1 44.5 48.0 49.3 53.6 5

8.4 60.8

[17] 65.1 68.7 73.4 78.0 80.7 85.2 87.0 91.6 94.2

1. **Determine total number of applications having rating below 3.5.**

**Description:**

**Subset()**function is a simple way to select variables. It is an indexing variable to accessing the object element.

* + subset(AppData,Rating < 3.5)

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No | App Category Rating | Reviews | Size.mb |
| 5 | Jewels Star: OZ adventure GAME 2.5 | 21892 | 14.0 |
| 8 | Free Book Reader BOOKS 3.4 | 1680 | 4.0 |
| 11 | All Language Translator Free BOOKS 3.4 | 30105 | 2.1 |
| 13 | ADP Mobile Solutions BUSINESS 1.3 | 85185 | 29.0 |
| 16 | Google Primer BUSINESS 2.4 | 62272 | 18.0 |
| 21 | Learn English with Wlingua EDUCATION 2.7 | 314299 | 3.3 |
| 23 | English Grammar Test EDUCATION 1.8 | 4075 | 5.1 |
| 25 | English for beginners EDUCATION 2.6 | 9321 | 27.0 |

# Description:

table() function is used to find the frequency of a particular column.

In this example it display different rating of apps and calculate its frequency using table() function.

Using count function, calculate rating below 3.5

* count=table(AppData$Rating < 3.5)
* print(count)

FALSE TRUE 17 8