

Session 6 – Visualization

and Plotting

Assignment - 1



Contents

[Introduction](#page3) [2](#page3)



[Objective](#page3) [2](#page3)



[Prerequisites](#page3) [2](#page3)



[Associated Data Files](#page3) [2](#page3)



[Problem Statement](#page3) [2](#page3)



[Expected Output](#page3) [2](#page3)



|  |  |
| --- | --- |
| Copyrights© 2017, AcadGild. All Rights Reserved | 1 |
|  |  |



**Introduction**



This assignment will help you to understand the key concepts learnt in this session.

**Objective**



This assignment will test your skills on Visualization and Plotting operations in R.

**Prerequisites**



Not Applicable

**Associated Data Files**



Not Applicable

**Problem Statement**



1. Import the Titanic Dataset from the link [Titanic Data Set.](file:///D:/R/new%20assignments/session1/Assignment1.docx)

TitanicData<-read.csv("E:\\Acadgil assignments\\TitanicData.txt",header=F,sep="," )

colnames(TitanicData)<-c("PassengerId",

"Survived",

"Pclass",

"Name",

"Sex",

"Age",

"SibSp",

"Parch",

"Ticket",

"Fare",

"Cabin",

"Embarked")

View(TitanicData)

Perform the following:

* 1. Preprocess the passenger names to come up with a list of titles that represent families and represent using appropriate visualization graph.

head(TitanicData)

tail(TitanicData)

str(TitanicData$Name) # check structure, as only charecter vectors can be split using strsplit function

TitanicData$Name<-as.character(TitanicData$Name)

str(TitanicData$Name)

#telling R to call rbind, on two charecters split by strsplit.

#in strsplit, as the data has many " ", and all breaks in many pieces

# hence, using sub() {and not gsub()}, which replaces only first pattern

# so, sub changes first space in ; and the strsplit splits along ; and then rbind binds along colums, which is called by do.call

namessplit<-do.call(rbind,strsplit(sub(" ",";",TitanicData$Name),";"))

head(namessplit)

#converting the charecters to data frame and naming the columns

namessplit<-data.frame(namessplit)

names(namessplit)<-c("family name", "first name")

head(namessplit)

str(namessplit)

#getting title separated from first name

Title<-(do.call(rbind,strsplit(TitanicData$Name)," "))[,2]

table(Title)

head(Title)

#merging the rownames in titanic survival data to form new data set

#similar to text to columns in excel

#tried merge function which didnt work as expected, but cbind is simpler and gives right data.

str(TitanicData)

TitanicData<-cbind(namessplit,TitanicData)

head(TitanicData)

View(TitanicData)

* 1. Represent the proportion of people survived from the family size using a graph.

SurvivedTitle<-table(TitanicData$Survived, TitanicData$Title)

#survived is 0, first row. we will take only that

p<-SurvivedTitle[1,]

#barplot of survived numbers per title

barplot(p,xlab = "Title", ylab = "survived",

main= "Survival as per title", col=rainbow(length(p)))

#pie chaart showing proportion of survival title wise

pie\_chart<-pie(p, main = "Pie-Chart of Titles survived", col = rainbow(length(p)) )

legend("topright", names(p), cex= 0.5, fill = rainbow(length(p)))

* 1. Impute the missing values in Age variable using Mice Library, create two different graphs showing Age distribution before and after imputation.

**Expected Output**



Not Applicable

|  |  |
| --- | --- |
| Copyrights© 2017, AcadGild. All Rights Reserved | 2 |
|  |  |