Prepared By: Sandeep Singh
Student ID: C0727422
Subject: 2018S-T2 BDM 1053 - Big Data Algorithms and Statistics 01
Title: Assignment 6
Week: 02 July - 08 July
Phone: +1 705-977-1868
Email (Internal): c0727422@mylambton.ca
Email (External): official.sandeep93@gmail.com
Submitted To: Mr. Pedram Habibi
Submission Date: 12 July, 2018

Contents

Explanation

Pre-requisites

```
titanic = read.csv("titanic.csv")

titanic_copy = titanic

View(titanic_copy)
```

We have first imported the dataset and created a copy of it in order to work upon it without amending the original dataset.

Task 1.1

Draw a box plot for class VS survived

```
titanic_copy$Class_copy = as.integer(titanic_copy$Class)
titanic_copy$Survived_copy =
as.integer(titanic_copy$Survived)
boxplot(Class_copy~Survived_copy, data = titanic_copy,
main = "Class v/s Survived" ,xlab="Class",
ylab="Survived")
```

To plot the box plot for the Class and Suvived,

1. We will convert the columns to integer type.
2. Use boxplot function to plot the graph.

Task 1.2

Draw a box plot for Age vs Sex

To plot the box plot for the Age and Sex:

1. We will convert the columns to integer type.
2. Use boxplot function to plot the graph.

Task 2

Create correlation matrix for mtcars and analyse which variables are linearly correlated and document the analysis.

```
library(corrgram)

corrgram(mtcars)
```

1. We will import the relevant library
2. Using corrgram function, we can plot the correlation between the columns

The following section describes more about the correlation analysis between different columns:

Column1	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	No Correlation	Highly Negative	Highly Negative	Highly Negative	Highly Positive	Highly Negative	Positive	Highly Positive	Highly Positive	Positive	Negative
cyl	Highly Negative	No Correlation	Highly Positive	Highly Positive	Highly Negative	Highly Positive	Highly Negative	Highly Negative	Highly Negative	Highly Negative	Positive
disp	Highly Negative	Highly Positive	No Correlation	Highly positive	Highly Negative	Highly Positive	Highly Negative	Highly Negative	Highly Negative	Highly Negative	Positive
hp	Highly Negative	Highly Positive	Highly Positive	No Correlation	Highly Negative	Highly Positive	Highly Negative	Highly Negative	Negative	Very Slightly Negative	Highly Positive
drat	Highly Positive	Highly Negative	Highly Negative	Highly Negative	No Correlation	Highly Negative	Slightly Positive	Highly Positive	Highly Positive	Highly Positive	Very Slightly Negative
wt	Highly Negative	Highly Positive	Highly Positive	Highly Positive	Highly Negative	No Correlation	Slightly Negative	Negative	Negative	Negative	Positive
qsec	Positive	Highly Negative	Highly Negative	Highly Negative	Slightly Positive	Slightly Negative	No Correlation	Highly Positive	Slightly Negative	Slightly Negative	Highly Negative
vs	Highly Positive	Highly Negative	Highly Negative	Highly Negative	Highly Positive	Negative	Highly Positive	No Correlation	Slightly Positive	Slightly Positive	Negative

am	Highly Positive	Highly Negative	Highly Negative	Negative	Highly Positive	Negative	Slightly Negative	Slightly Positive	No Correlation	Highly Positive	Very Slightly Positive
gear	Positive	Highly Negative	Highly Negative	Very Slightly Negative	Highly Positive	Negative	Slightly Negative	Slightly Positive	Highly Positive	No Correlation	Slightly Positive
carb	Negative	Positive	Positive	Highly Positive	Very Slightly Negative	Positive	Highly Negative	Negative	Very Slightly Positive	Slightly Positive	No Correlation

	Very Slightly Negative
	Slightly Negative
	Negative
	Highly Negative
	No Correlation
	Highly Positive
	Positive
	Slightly Positive
	very Slightly Positive

A **correlation** can range between -1 (perfect negative **relationship**) and +1 (perfect positive **relationship**), with 0 indicating no straight-line **relationship.** The colour scheme of the columns illustrate the correlation. Red colour denotes a negative correlation and blue colour denotes the positive correlation whereas white colour is used to denote no correlation.