1. Import the Titanic Dataset from the link Titanic Data Set.

Perform the following:

a. Is there any difference in fares by different class of tickets?

Note - Show a boxplot displaying the distribution of fares by class

Ans:

# Load packages

**library**('ggplot2')

**library**('ggthemes')

**library**('scales')

**library**('dplyr')

**library**('mice')

**library**('randomForest')

train <- **read.csv**('C:\Users\jai sriram\Desktop\Acadgild/train.csv', stringsAsFactors = F)

test <- **read.csv**('C:\Users\jai sriram\Desktop\Acadgild/test.csv', stringsAsFactors = F)

full <- **bind\_rows**(train, test) *# bind training & test data*

> titanic %>%

+ filter(!is.na(PCLASS)) %>%

+ ggplot() +

+ geom\_boxplot(aes(PCLASS, FARE))

b. Is there any association with Passenger class and gender?

Note – Show a stacked bar chart

rm(list=ls())

**library**(NLP)

**library**(tm)

**library**(R.utils)

**library**(stringi)

**library**(data.table)

**library**(ggplot2)

**library**(SnowballC)

**library**(caret) *# Confusion Matrix*

**library**(pscl) *# For R^2*

**library**(ROCR) *# for ROC and AUC*

setwd("'C:\Users\jai sriram\Desktop\Acadgild")

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

*#Columns available in the original Data-set*

colnames(train)

sum(is.na(train$Age))

max(train$Age)

min(train$Age)

counts <- table(train$Survived, train$Sex)

counts

female <- nrow(train[train$Sex=='female',])

male <- nrow(train[train$Sex=='male',])

slice <- c(female, male)

divi <- c("Female", "Male")

survival\_percent <- c(counts[2]/(counts[1]+counts[2]), counts[4]/(counts[3]+counts[4]))

survival\_percent

Ccounts <- table(train$Survived, train$Pclass)

Ccounts

Class\_survival\_percent <- c(Ccounts[2]/(Ccounts[1]+Ccounts[2]), Ccounts[4]/(Ccounts[3]+Ccounts[4]), Ccounts[6]/(Ccounts[5]+Ccounts[6]))

Class\_survival\_percent