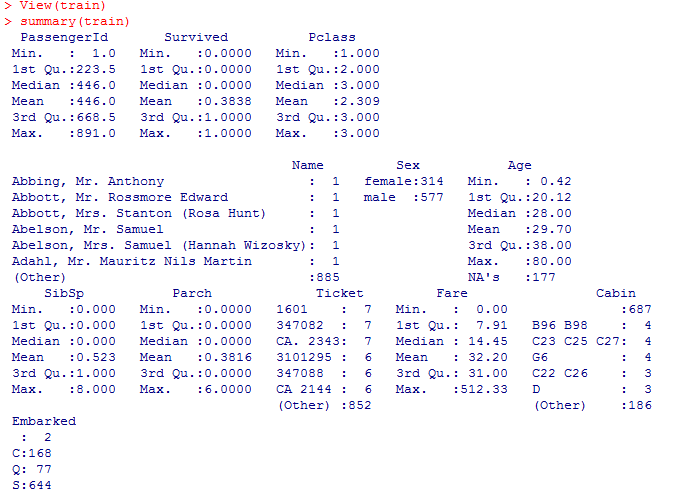
Summary:

The following is the summary of the data as is from the summary function of R.

As seen below there are 891 entries in the dataset; out of these there were 177 “Not Available” (NA) data; hence effectively 714 entries were used for further the analysis.

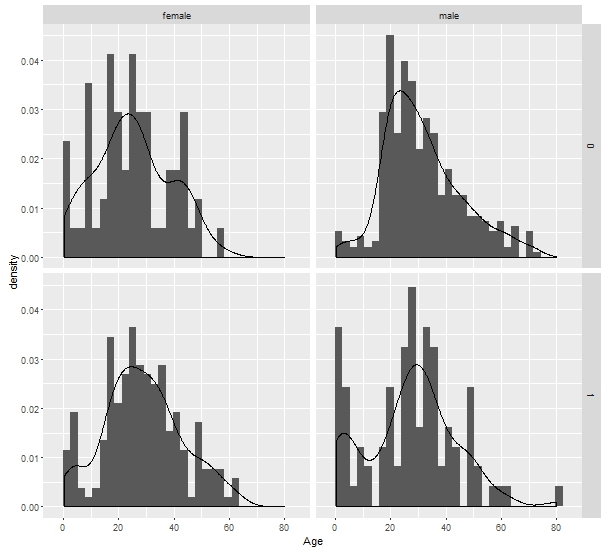
All the passengers id were in sequence starting from 1 through 891; there were 314 Females and 577 Males. The passengers age varied between from few months to 80 years with majority of the passengers were of the age between 29 and 30 and the average Age was 28 years. The fare of the tickets range was between 0 and little over 512 units; the most common fare was between 14 and 15 units with an average of 32 units. Majority of the passengers, 687, were assumed to be in general class. About one third of the passengers survived.



Facet Grid:

The below shows the plot using histogram with two facets, Factor Survived and Sex. Thereby the below histogram shows details about Male/Female, theirs age and survival.

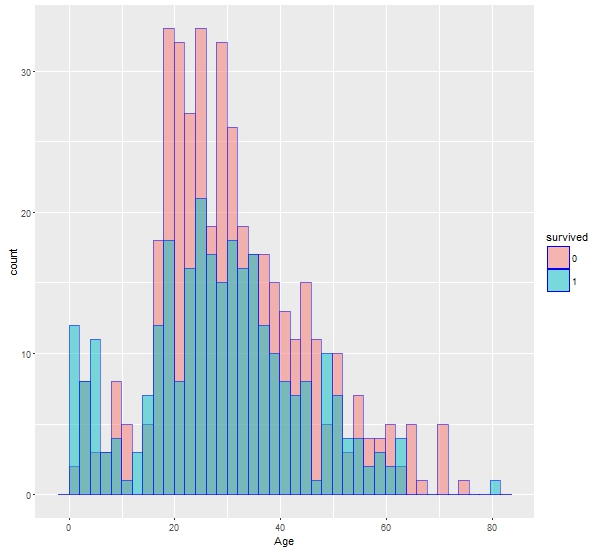
Among the survived – there were more females than males between the age between 20 and 40 ; more male infants survived than females; there were not that many males or females above age 60 survived only a few males with age 80 survived. In addition to that it seemed that predominant age of both males and females was around 28 or 29; and there were more males than females in this category.



Among those did not survive – there were more young males than females aged between 20 and 40. It seems that more males than females did not survive. There were more old males than females of the age group 60 to 80.

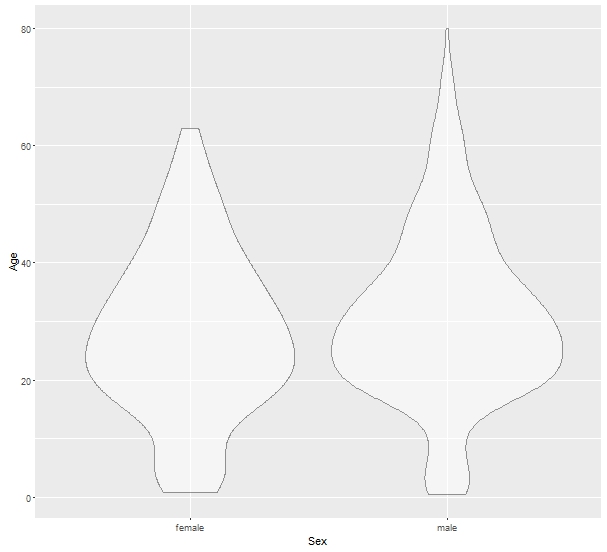
Histogram:

From the Histogram below it seems most of the passengers were of the age between 18 and 50 and of these majority of them did not survive. There were considerable number of kids and teenagers of age between 0 and 18 and maximum number of passengers of these survived. There seems to be one or two passengers of age 80, the oldest age group, and all that survived. Looks like the passengers of the age 22 and 46 have more causalities than survived.



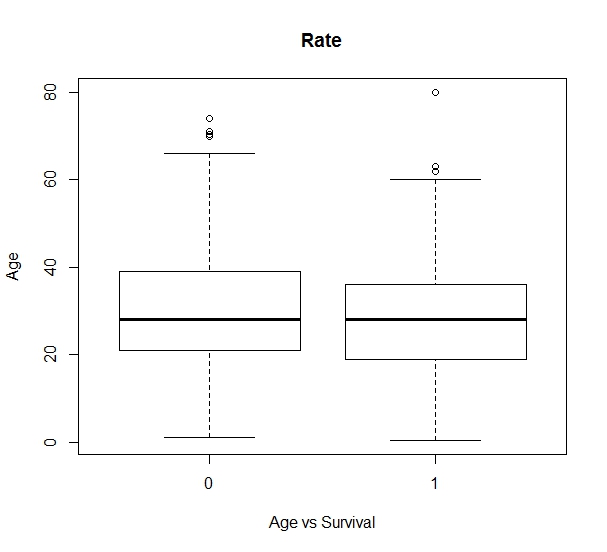
Violin plot:

Similar to the other plots it is evident that the oldest passengers were of age 80. There were more males than females of the age group of 60 and 80. There were more females than males between the age of 20 and 40. There were more female children than male of age between 0 and 10. The oldest female passenger was 62 years of age and that of male was 80.



Whisker Plot:

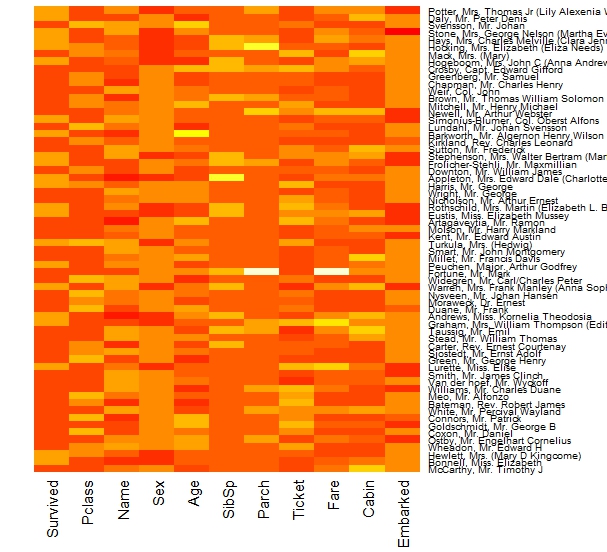
The Median age of those survived and for those did not survive was around 28 years as it is evident from box in the plot diagram below. This is consistent with the other plots summarized so far. The oldest survived happened to be of age 80. There were some passengers of age above 60 that did not survive. It seems an equal number of teenage passengers or young children in both survived and did not survive categories.



Heat Map:

The below Heat map provides more details on the all the parameters from the data. We have considered the data of passengers of age 50 and above to keep simple and not to clutter the plot.

There were few survived than died represented in light color box for “Survived” parameter. There were more females dies than males as represented by light color box for “Sex” parameter. There were more passengers in the lower “Pclass” died of age between 50 and 60 or 65; majority of these are females.



R – Programming Code:

The following is the R programming code copied below ; though submitted separately as R file.

#Setting Working Directory

setwd("C:\\Users\\Venkatesh\\Desktop\\Assignment4")

#Reading the csv into a Data Frame

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

#Opon verifying the data, found some rows with N/A values. Removing such rows.

train <- na.omit(train)

#Loading ggplot2 Library

library(ggplot2)

#There Only 2 levels for Survival. So, Converting them to factors.

survived <- as.factor(train$Survived)

#Plotting Histogram with Age on X-Axis and the Count on Y-Axis

ggplot(train, aes(Age, fill = survived)) + geom\_histogram(alpha = 0.5,colour="blue",position="identity",bins = 40)

#Whisker Plot showing information about survival and age

boxplot(Age~survived,data=train, main="Rate", xlab="Age vs Survival", ylab="Age")

#Facet Grid, Showing Age, Survival and Density in a single plot.

ggplot(train, aes(x=Age)) + geom\_histogram(aes(y = ..density..)) + geom\_density() +facet\_grid(Survived ~ Sex)

#Violin plot Age vs Sex

ggplot(train) + geom\_violin(alpha=0.5,aes(x = Sex, y = Age))

#Heatmap

# Had filter on the Age because, the plot was not clear considering all the passengers or putting a filter on any other Factor

#Selecting passengers with Age>50

train1 <- subset(train, Age>50)

#Removing Passenger Id Column. Will use the Passenger Names in the Plot. Will make more sense.

train1 <- train1[,2:12]

#Selecting Passenger Names.

row.names(train1) <- train1$Name

#Converting the Data into Matrix

train\_matrix <- data.matrix(train1)

#HeatMap Plot

heatmap(train\_matrix,Rowv=NA, Colv=NA, col = heat.colors(15), scale="column", margins=c(5,10))