

## task2\_titanic\_dataset

rehana\_bensha

2024-05-17

```
#import libraries
library(ggplot2)

library(dplyr)

library(reshape2)

#read the dataset
tita<-read.csv("C:/Users/abi00/Downloads/dataset/tested.csv")
#View First % element in the dataset
head(tita)

##   PassengerId Survived Pclass                                Name
## 1         892         0       3                                Kelly, Mr. James
## 2         893         1       3      Wilkes, Mrs. James (Ellen Needs)
## 3         894         0       2                Myles, Mr. Thomas Francis
## 4         895         0       3                Wirz, Mr. Albert
## 5         896         1       3 Hirvonen, Mrs. Alexander (Helga E Lindqvist)
## 6         897         0       3      Svensson, Mr. Johan Cervin
##      Sex  Age SibSp Parch  Ticket       Fare Cabin Embarked
## 1  male 34.5     0     0  330911   7.8292      Q
## 2 female 47.0     1     0  363272   7.0000      S
## 3  male 62.0     0     0  240276   9.6875      Q
## 4  male 27.0     0     0  315154   8.6625      S
## 5 female 22.0     1     1 3101298  12.2875      S
## 6  male 14.0     0     0    7538   9.2250      S

#rows
nrow(tita)

## [1] 418

#columns
ncol(tita)

## [1] 12

#summary
summary(tita)

##   PassengerId      Survived      Pclass      Name
##  Min.   : 892.0   Min.   :0.0000   Min.   :1.000   Length:418
##  1st Qu.: 996.2   1st Qu.:0.0000   1st Qu.:1.000   Class :character
##  Median :1100.5   Median :0.0000   Median :3.000   Mode  :character
```

```

## Mean :1100.5 Mean :0.3636 Mean :2.266
## 3rd Qu.:1204.8 3rd Qu.:1.0000 3rd Qu.:3.000
## Max. :1309.0 Max. :1.0000 Max. :3.000
##
## Sex Age SibSp Parch
## Length:418 Min. : 0.17 Min. :0.0000 Min. :0.0000
## Class :character 1st Qu.:21.00 1st Qu.:0.0000 1st Qu.:0.0000
## Mode :character Median :27.00 Median :0.0000 Median :0.0000
## Mean :30.27 Mean :0.4474 Mean :0.3923
## 3rd Qu.:39.00 3rd Qu.:1.0000 3rd Qu.:0.0000
## Max. :76.00 Max. :8.0000 Max. :9.0000
## NA's :86
## Ticket Fare Cabin Embarked
## Length:418 Min. : 0.000 Length:418 Length:418
## Class :character 1st Qu.: 7.896 Class :character Class :character
## Mode :character Median : 14.454 Mode :character Mode :character
## Mean : 35.627
## 3rd Qu.: 31.500
## Max. :512.329
## NA's :1

#structure
str(tita)

## 'data.frame': 418 obs. of 12 variables:
## $ PassengerId: int 892 893 894 895 896 897 898 899 900 901 ...
## $ Survived : int 0 1 0 0 1 0 1 0 1 0 ...
## $ Pclass : int 3 3 2 3 3 3 3 2 3 3 ...
## $ Name : chr "Kelly, Mr. James" "Wilkes, Mrs. James (Ellen Needs)"
"MYles, Mr. Thomas Francis" "Wirz, Mr. Albert" ...
## $ Sex : chr "male" "female" "male" "male" ...
## $ Age : num 34.5 47 62 27 22 14 30 26 18 21 ...
## $ SibSp : int 0 1 0 0 1 0 0 1 0 2 ...
## $ Parch : int 0 0 0 0 1 0 0 1 0 0 ...
## $ Ticket : chr "330911" "363272" "240276" "315154" ...
## $ Fare : num 7.83 7 9.69 8.66 12.29 ...
## $ Cabin : chr "" "" "" "" ...
## $ Embarked : chr "Q" "S" "Q" "S" ...

#check_missing_values
sum(is.na(tita))

## [1] 87

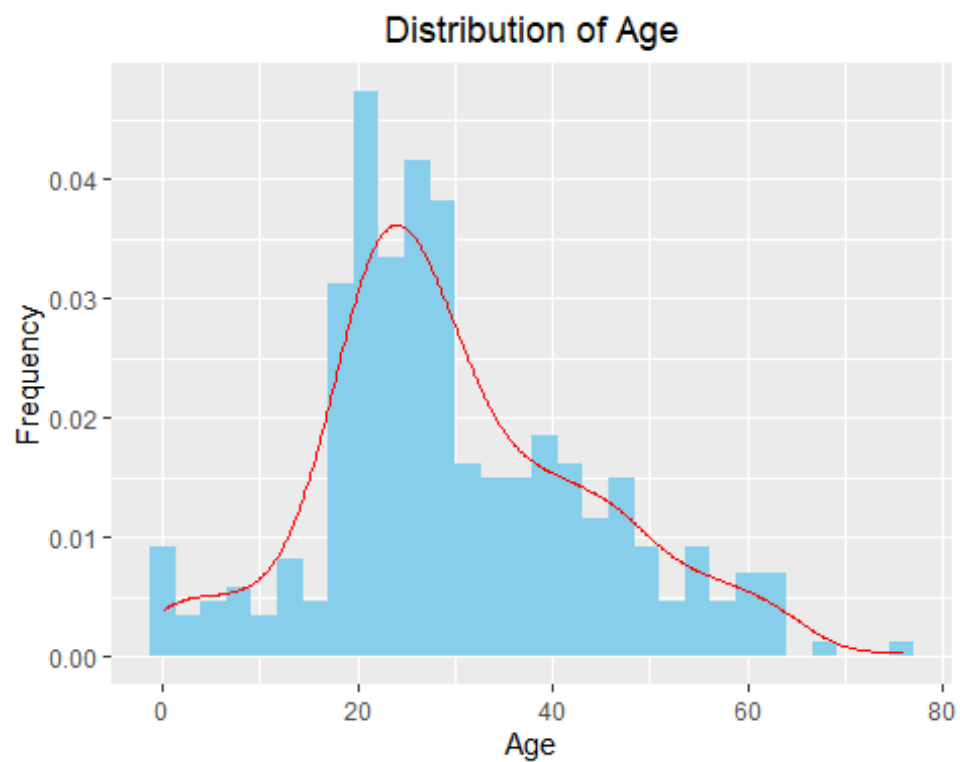
```

```
#clear_missing_values
titan<-na.omit(tita)
titan
```

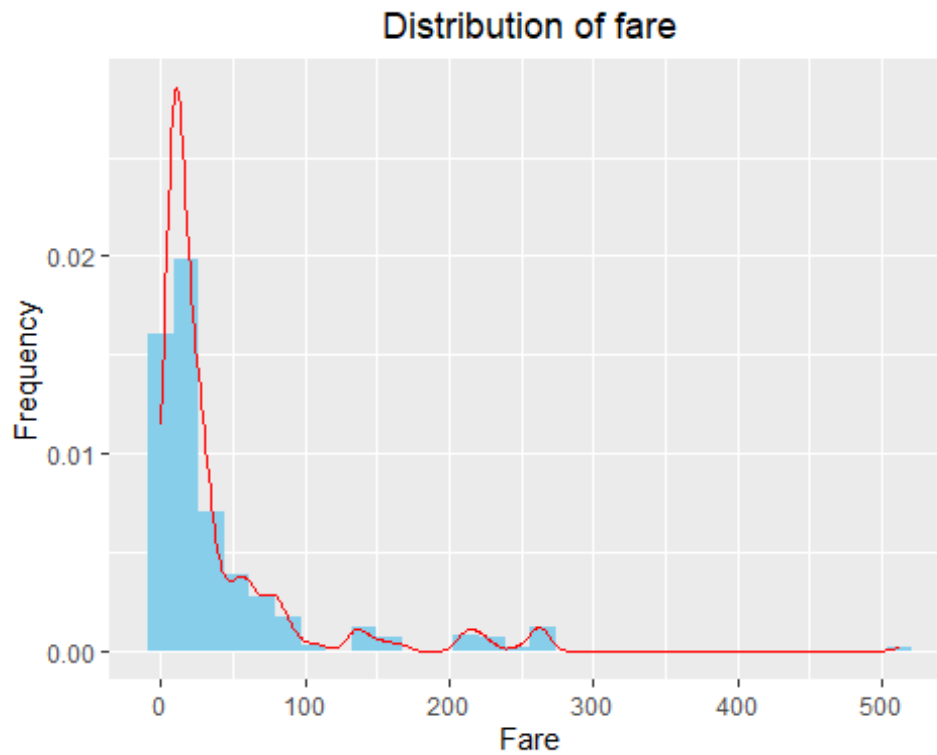
```
#Univariate
```

```
#Distribution of Age
```

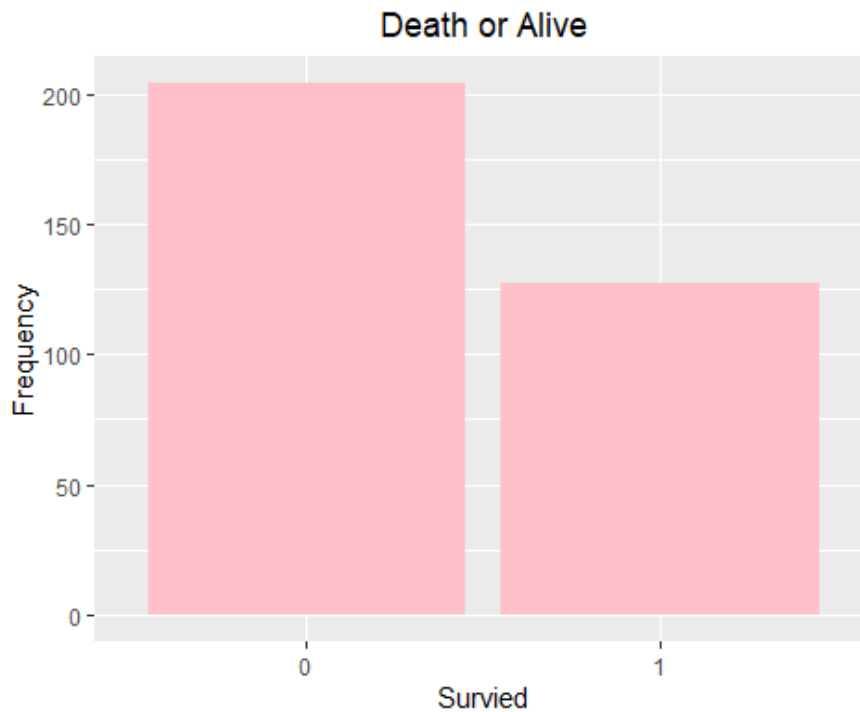
```
ggplot(titaa,aes(titaa$titan.Age))+geom_histogram(aes(y=..density..),fill='skyblue')+xlab('Age')+ylab('Frequency')+geom_density(col='red')+labs(title='Distribution of Age')+theme(plot.title = element_text(hjust=0.5))
```



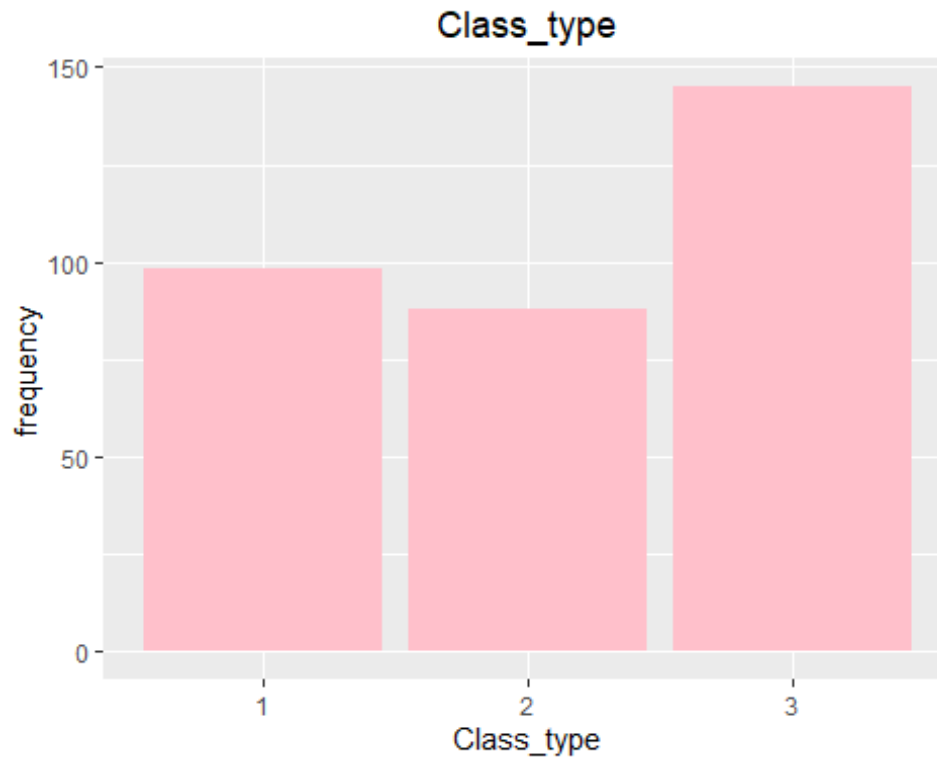
```
ggplot(titaa,aes(titaa$titan.Fare))+geom_histogram(aes(y=..density..),fill='skyblue')+xlab('Fare')+ylab('Frequency')+labs(title='Distribution of fare')+theme(plot.title = element_text(hjust=0.5)) +geom_density(col='red')
```



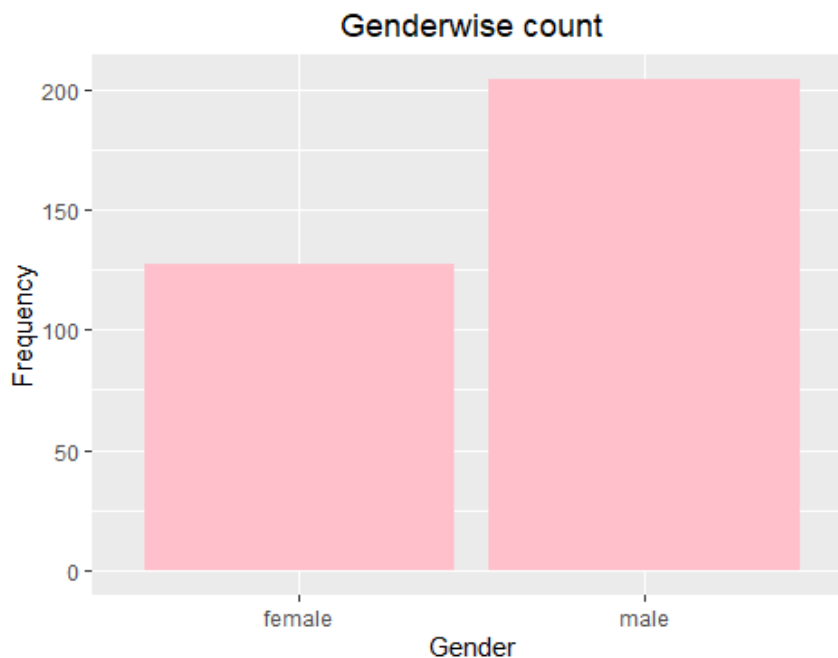
```
#Survied or not ?  
ggplot(titaa,aes(x=factor(titan.Survived)))+geom_bar(fill='pink')+xlab('Survived')+ylab('Frequency')+labs(title='Death or Alive')+theme(plot.title = element_text(hjust = 0.5))
```



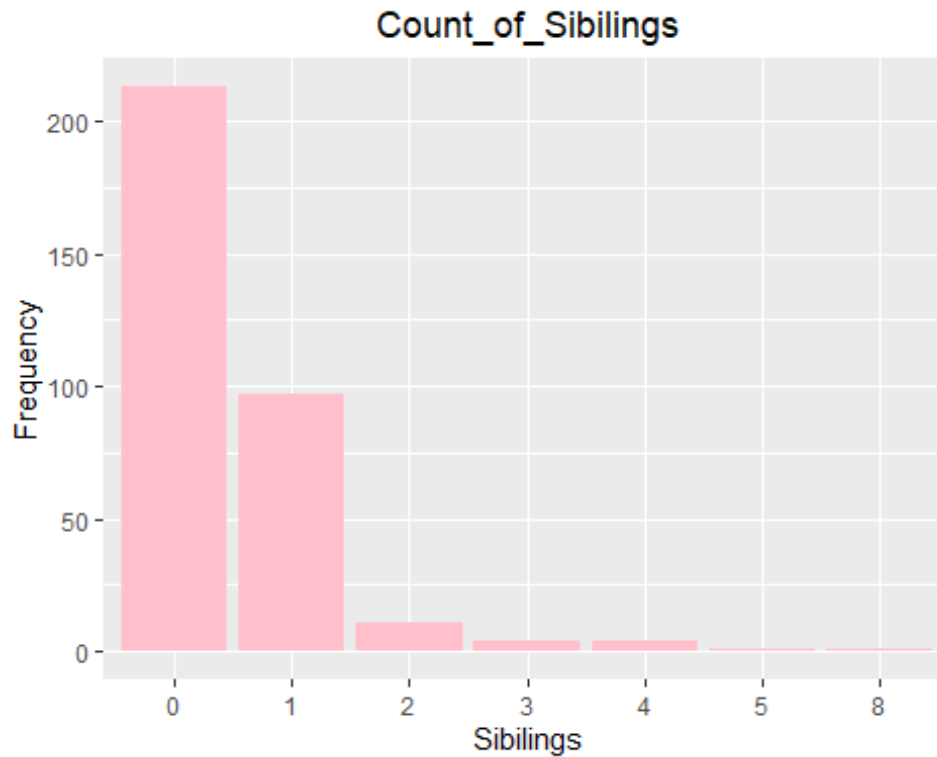
```
#Class
ggplot(titaa,aes(x=factor(titan.Pclass)))+geom_bar(fill='pink')+xlab('Class_type')
+ylab('frequency')+labs(title='Class_type')+theme(plot.title =
element_text(hjust=0.5))
```



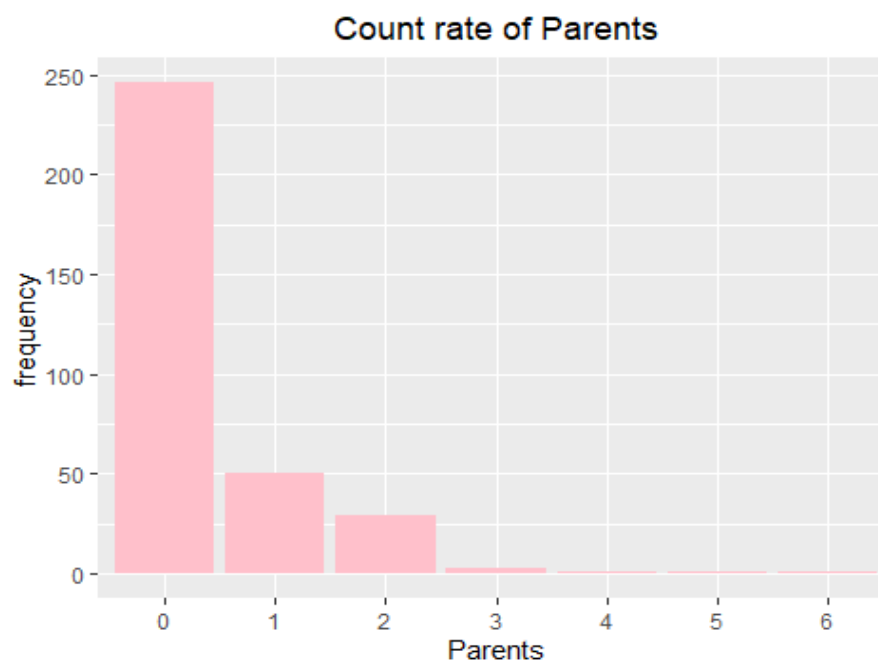
```
#genderwise frequency
ggplot(titaa,aes(titan.Sex))+geom_bar(fill='pink')+xlab('Gender')+ylab(
'Frequency')+labs(title = 'Genderwise count')+theme(plot.title =
element_text(hjust=0.5))
```



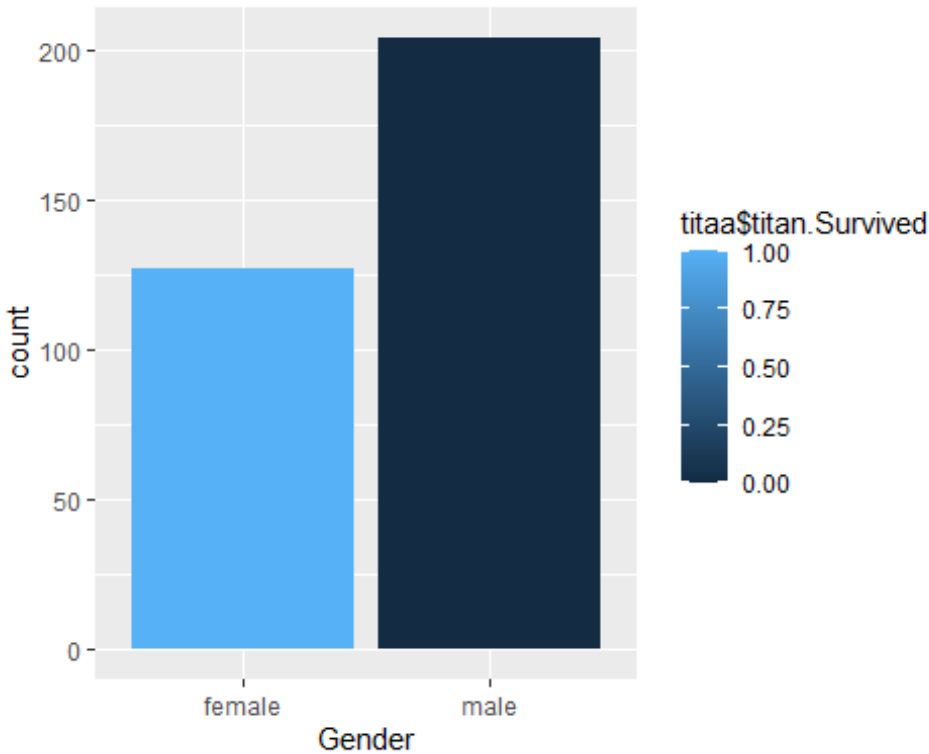
```
ggplot(titaa,aes(x=factor(titaa$titan.SibSp)))+geom_bar(fill="pink")+xlab('Siblings')+ylab('Frequency')+labs(title='Count_of_Siblings')+theme(plot.title = element_text(hjust=0.5))
```



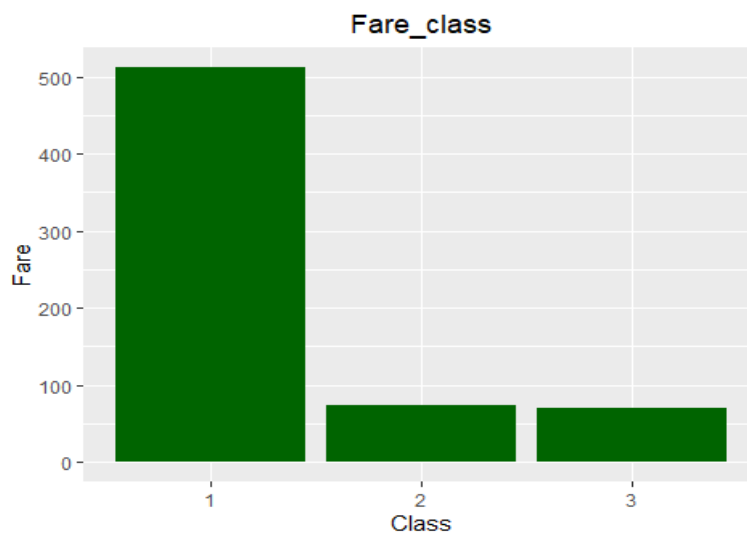
```
ggplot(titaa,aes(factor(titaa$titan.Parch)))+geom_bar(fill='pink')+xlab('Parents')+ylab('frequency')+labs(title='Count rate of Parents')+theme(plot.title = element_text(hjust=0.5))
```



```
#Bivariate
# Assumption 1: Compare the male and female death rate
ggplot(titaa, aes(x=factor(titaa$titan.Sex), fill=titaa$titan.Survived)) + geom_bar() + xlab('Gender')
```



```
#Assumption 2: Which class requires more fare ?
ggplot(titaa, aes(x=factor(titaa$titan.Pclass), y=titaa$titan.Fare, fill=titan.Survived)) + geom_bar(stat = 'identity', position='dodge', fill='darkgreen') + xlab('Class') + ylab('Fare') + labs(title = 'Fare_class') + theme(plot.title = element_text(hjust=0.5))
```

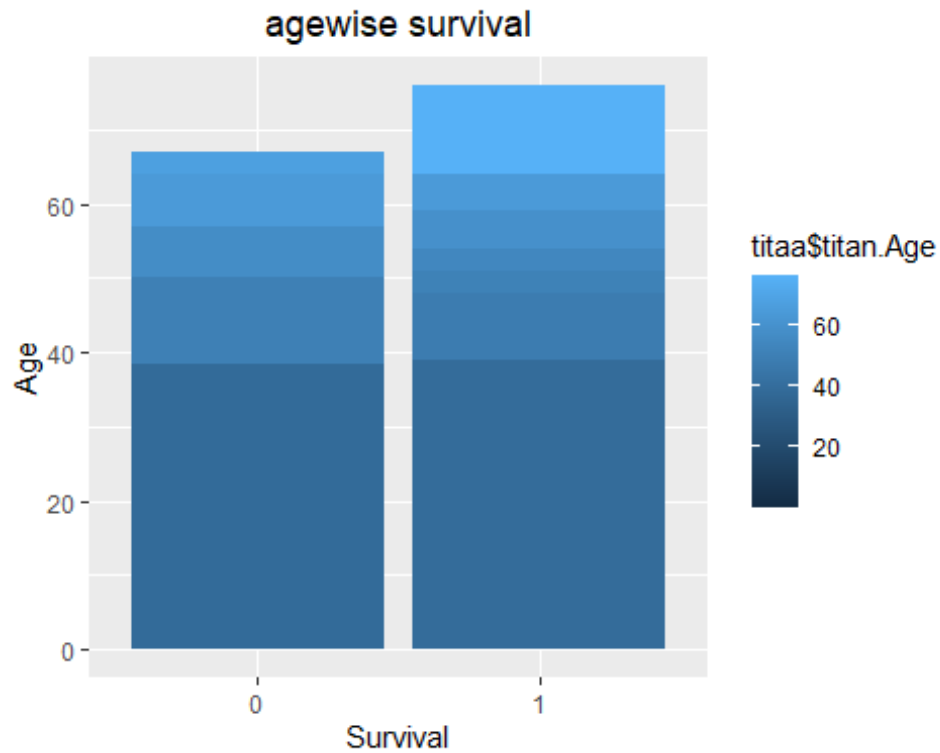


```
#Assumption 3: Find Which gender Spend more for fare ?
ggplot(titaa,aes(x=titaa$titan.Sex,y=titaa$titan.Fare))+geom_bar(stat='identity',position = 'dodge',fill='pink')+xlab("gender")+ylab("Fare")+labs(title="Fare depends upon gender")+theme(plot.title = element_text(hjust = 0.5))
```



```
#Assumption 4: Agewise survival
ggplot(titaa,aes(factor(titaa$titan.Survived),titaa$titan.Age,fill=titaa$titan.Age))+geom_bar(stat='identity',position='dodge')+xlab('Survival')+ylab('Age')+labs(title = 'agewise survival')+theme(plot.title = element_text(hjust = 0.5))
```



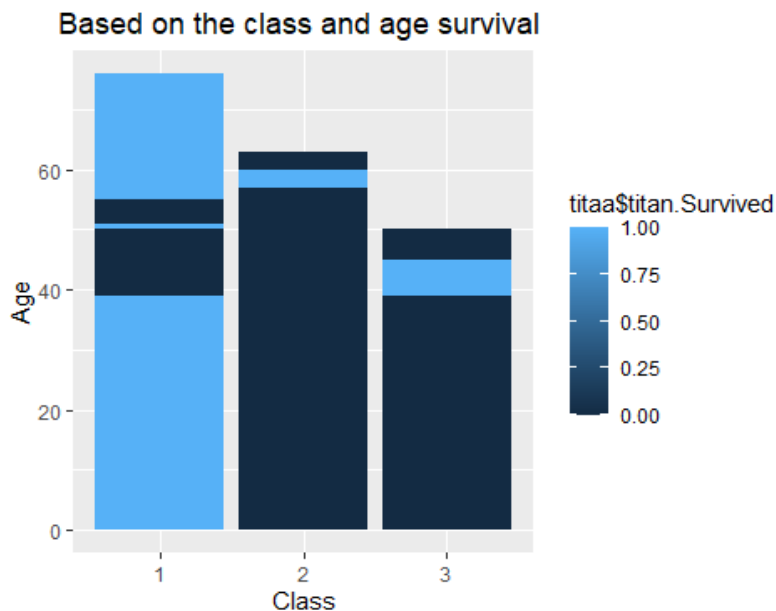


*#Assumption 5: Based on the class and age survival ?*

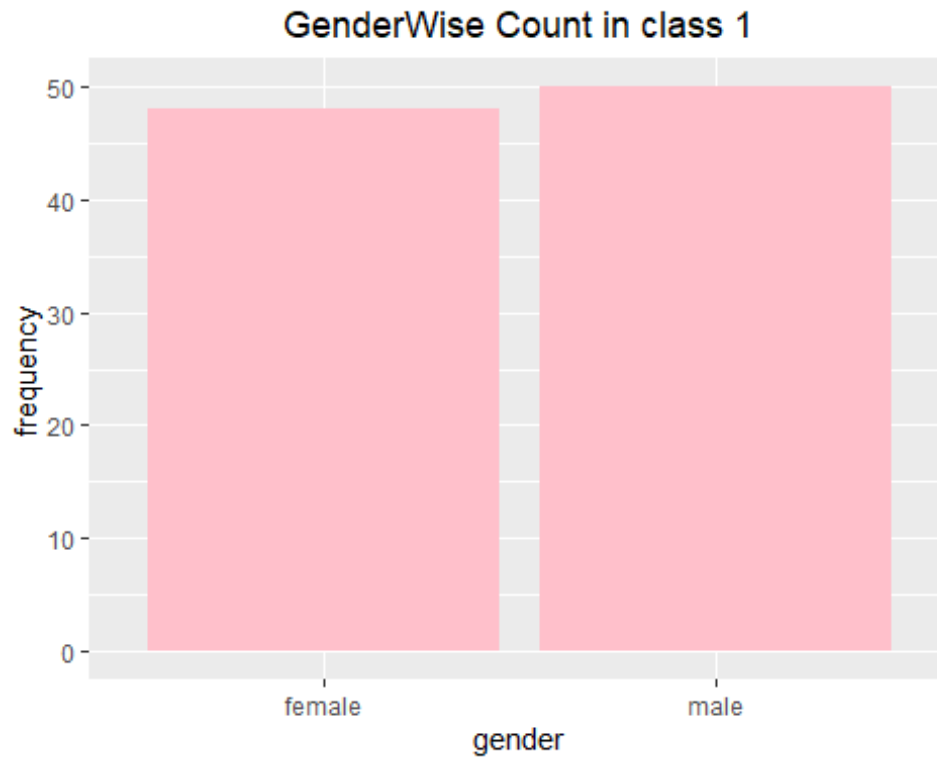
```
ggplot(titaa,aes(factor(titaa$titan.Pclass),titaa$titan.Age,fill=titaa$titan.Survived))+geom_bar(stat='identity',position='dodge')+xlab('Class')+ylab('Age')+labs(title = "Based on the class and age survival ")+theme(plot.title = element_text(hjust = 0.5))
```

*#Assumption : Which Gender more travel in class 1*

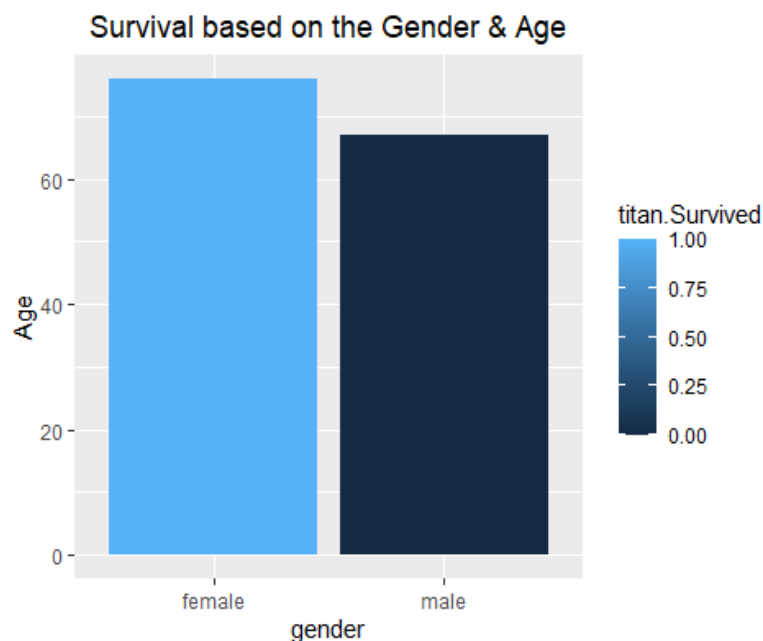
```
fil<-titaa %>% filter(titan.Pclass==1)
head(fil)
```



```
#GenderWise Count in class 1
ggplot(fil,aes(fil$titan.Sex))+geom_bar(fill='pink')+xlab('gender')+ylab('frequency')+labs(title = 'GenderWise Count in class 1')+theme(plot.title = element_text(hjust = 0.5))
```



```
ggplot(fil,aes(factor(fil$titan.Sex),fil$titan.Age,fill=titan.Survived))+geom_bar(stat='identity',position='dodge')+xlab('gender')+ylab('Age')+labs(title = 'Survival based on the Gender & Age')+theme(plot.title = element_text(hjust = 0.5))
```



```

#multivariate
corrr<-data.frame(titaa$titan.Fare,titaa$titan.Age)
cor_tita<-cor(corrr)
melt_tita<-melt(cor_tita)
ggplot(melt_tita,aes(Var1,Var2,fill=value))+geom_tile()+scale_fill_gradient(1
ow='skyblue',high='pink')+labs(title='correlation_between_titanic_dataset')+t
heme(plot.title = element_text(hjust=0.5))

```

