First name :Hadj ALI

Last name : WISSAL

Group :11

Project name: visualisation of Tanzania tourists dataset (Train)

Deliver date : 22/05/2022

Title

Code

install.packages("plot3D")

install.packages("plotly")

install.packages("rgl")

library(plot3D)

library(plotly)

library(rgl)

scatter3D(x=Train$night\_zanzibar,y=Train$night\_mainland,z=Train$total\_cost ,

ticktype="detailed",clab="cost",xlab="night zanzibar",ylab="night mainland",

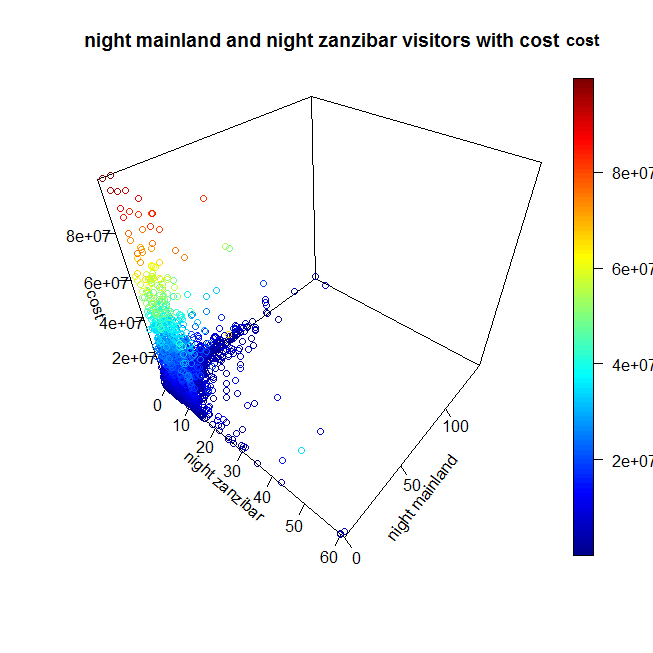
zlab="cost",

main=" night mainland and night zanzibar visitors with cost")

Comment

According to the chart below night zanzibar visitors pay much more than those of night mainland

Output:



1.Title : Analyse cost of tourists by international transport .

2.Code:

ggplot(Train, aes(package\_transport\_int,total\_cost,fill=package\_transport\_int )) +

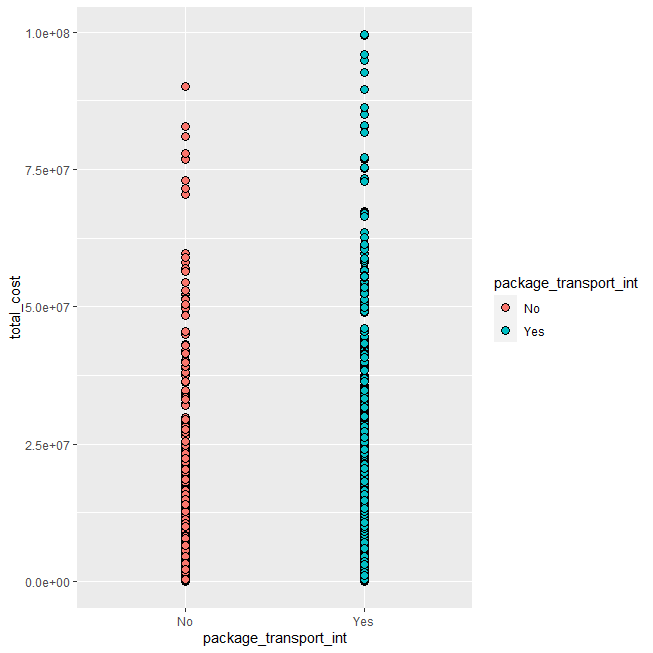
geom\_point(shape = 21,size = 3,

main="international transport and total\_cost")

3 comment:

The graph below shows that most Tanzanian tourists use international transport .

4.Output:



Title:

Analyse cost of tourists by international transport

Code:

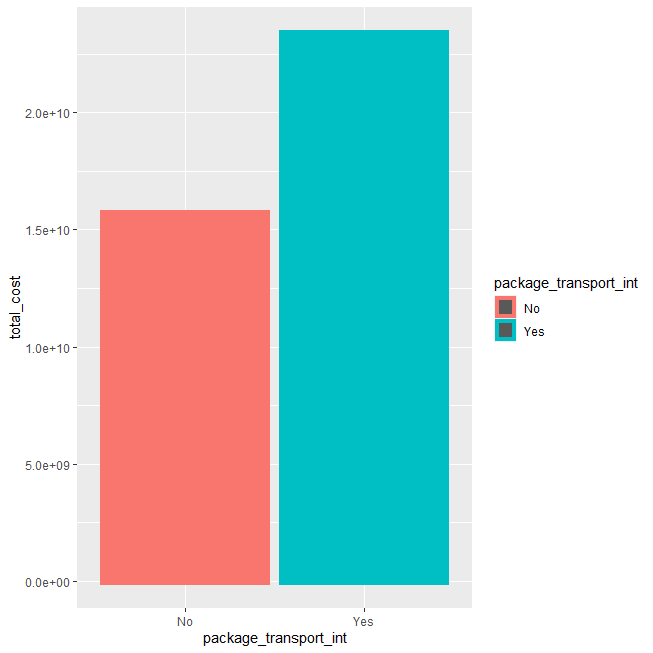
ggplot(Train,aes(package\_transport\_int,total\_cost,color=package\_transport\_int,

shape=package\_transport\_int)) +

geom\_col(size=3)

Comment:

3 comment:from the graph below it seems that most Tanzanian tourists use international transport .



Title: Analyse cost of tourists by tanzania transport

Code:

ggplot(Train,aes(package\_transport\_tz,total\_cost,color=package\_transport\_tz,

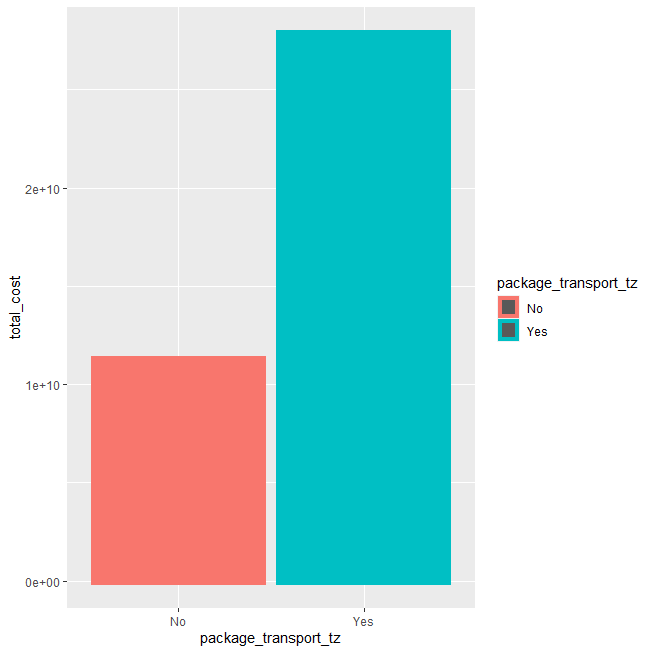
shape=package\_transport\_tz)) +

geom\_col(size=3)

Comment:

3 comment:from the graph below it seems that most Tanzanian tourists use tanzania transport .

Output



Train1 <- Train %>% select(package\_transport\_int,package\_transport\_tz,total\_cost) %>%

group\_by(total\_cost)%>%

summarise(cost = mean(total\_cost),nb= n())%>%

filter(nb > 150) %>%

ggplot(aes(country,nb,fill= c("green","blue"))) +

geom\_col()

Title:Night zanzibar guests

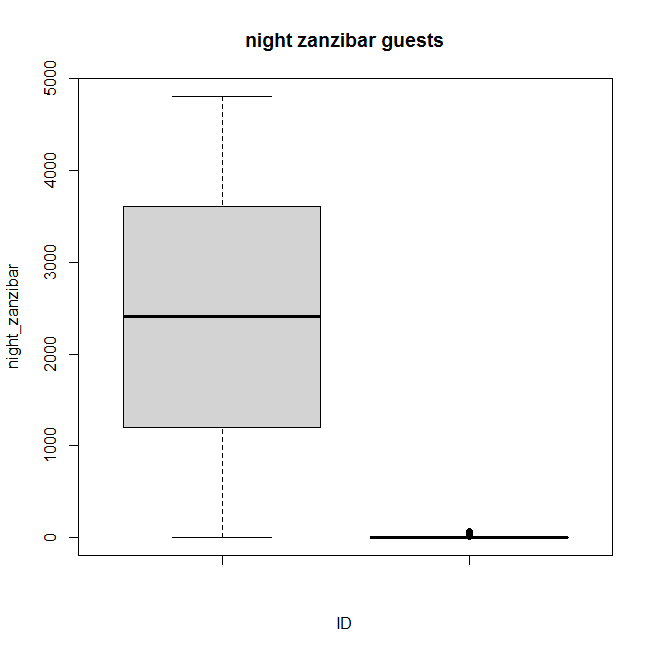
Code boxplot(Train$ID,Train$night\_zanzibar,xlab="ID",ylab="night\_zanzibar",

main="night zanzibar guests",

color="dark")

Comment: The boxplot below shows that night zanzibar has alot of tourists .

Output



Title: Tanzania tourists age categories

Code

summary(Train)

summary(catAge)

ctage = na.omit(Train$age\_group)

summary(ctage)

tbl\_ctage = table(ctage)

tbl\_ctage

pie(tbl\_ctage, clockwise = F,col = rainbow(4, s = 0.5),init.angle = 90,

main = "tourists age categories",

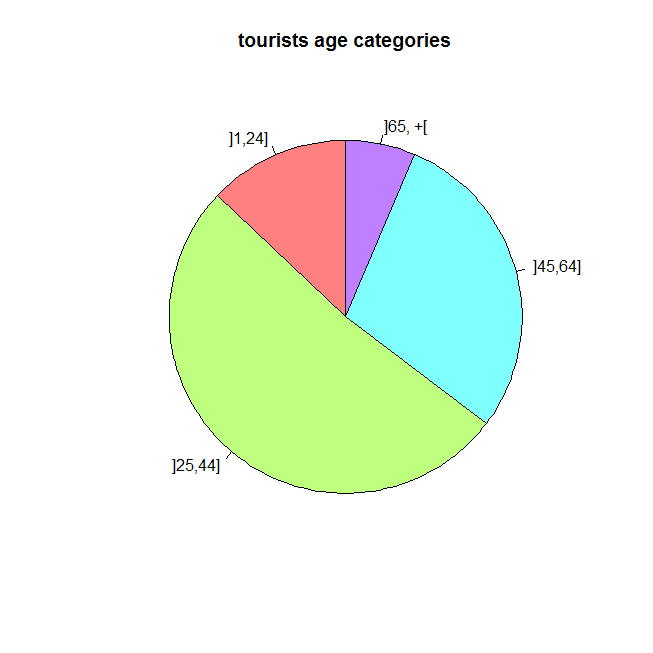
legend = T,

labels = c("]1,24]","]25,44]","]45,64]","]65, +["))

Comment

:From pie graph below most tanzania tourists ages are from two categories

[25,44] ,[45,64]



Title The main activity for tanzania tourists

Code

summary(Train$main\_activity)

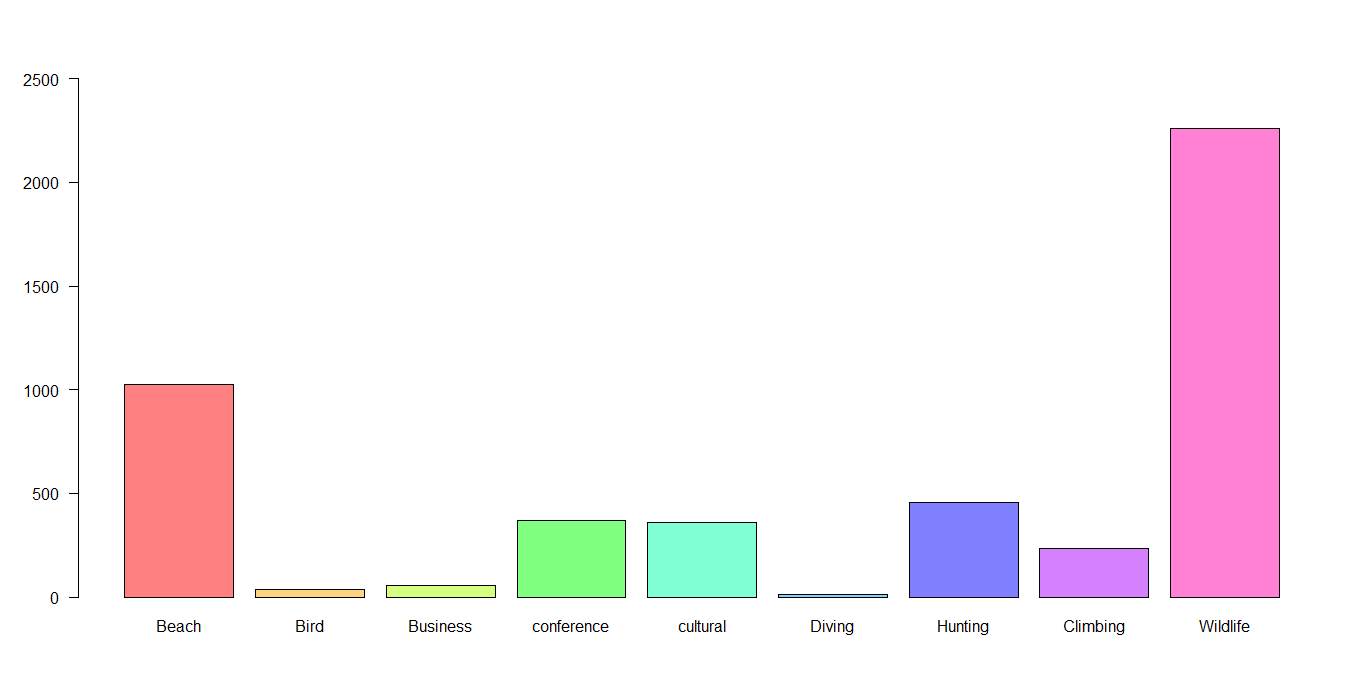
tbl = table(Train$main\_activity)

barplot(tbl,col = rainbow(9, s = 0.5),names.arg = c("Beach","Bird","Business","conference","cultural",

"Diving","Hunting","Climbing", "Wildlife"), ylim = c(0,2500),las = 1)

Comment: Most Tanzanian tourists' main activity is wildlife than swimming than hunting ,conference and finally cultural.

Output



Title Analysing many variables with boxplot .

Code :

pack1 = summary(Train$package\_transport\_int)

pack1

pack2 = summary(Train$package\_accomodation)

pack2

pack3 = summary(Train$package\_food)

pack3

pack4 = summary(Train$package\_transport\_tz)

pack4

pack5 = summary(Train$package\_sightseeing)

pack5

pack6 = summary(Train$package\_guided\_tour)

pack6

pack7 = summary(Train$package\_insurance)

pack7

par(mfrow=c(3,3))

barplot(pack1,col = rainbow(2,s=0.1),las = 1,xlab = "transport\_int")

barplot(pack2,col = rainbow(2,s=0.2),las = 1,xlab = "package\_accomodation")

barplot(pack3,col = rainbow(2,s=0.3),las = 1,xlab = "package\_food")

barplot(pack4,col = rainbow(2,s=0.4),las = 1,xlab = "package\_transport\_tz")

barplot(pack5,col = rainbow(2,s=0.5),las = 1,xlab = "sightseeing")

barplot(pack6,col = rainbow(2,s=0.6),las = 1,xlab = "guided\_tour")

barplot(pack7,col

Comments

***Comment 01:***According to the boxplot below the tourists who doesn't use international transport are more than non users

***comment02:*** According to the boxplot below tourists with package accommodation are less than those with no package accommodation

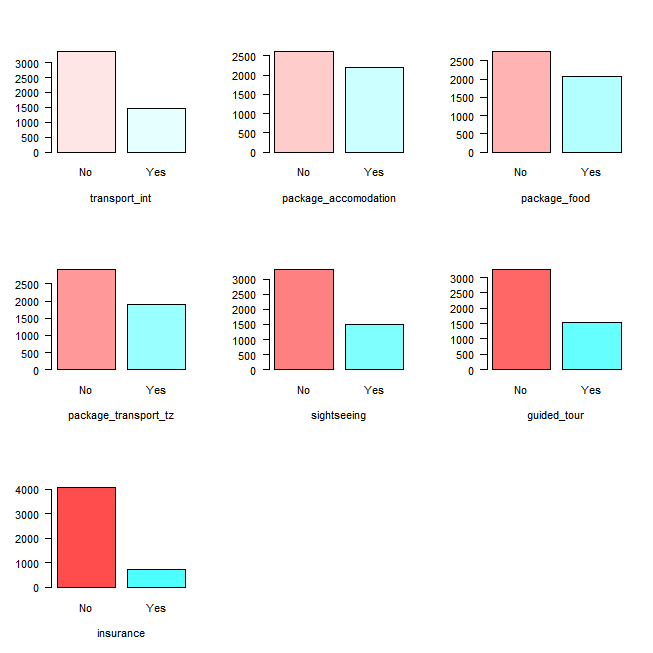
***Comment03:***According to the boxplot tourists with package food are less than those without package food

***Comment04:***According to the boxplot below tourists who use tanzania transport are less than those who doesn’t use it .

***Comment05:***According to the boxplot below tourists who use tanzania transport are less than those who doesn’t use it .

***Comment06:***According to the boxplot below tourists belongs to guided tour are less than those who doesn’t belong to guided tour

***Comment07:***According to the boxplot below tourists who have insurance are less than those who haven't it .

******

Title The food and total cost

Code

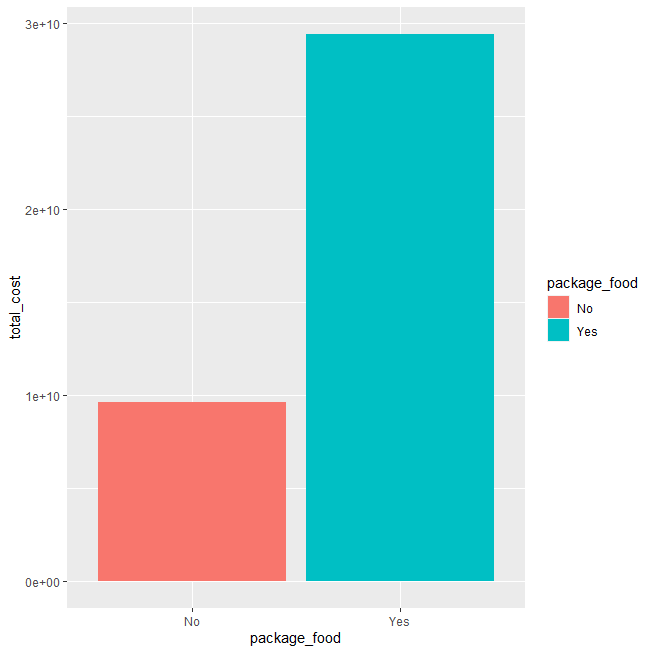
ggplot(Train,aes(package\_food,y=total\_cost,fill=package\_food))+

geom\_histogram(stat = "identity")

Comment

The histogram below shows that tourists that take food from the organised trip pay more .

Output



Title Night mainland visitors and total cost payed

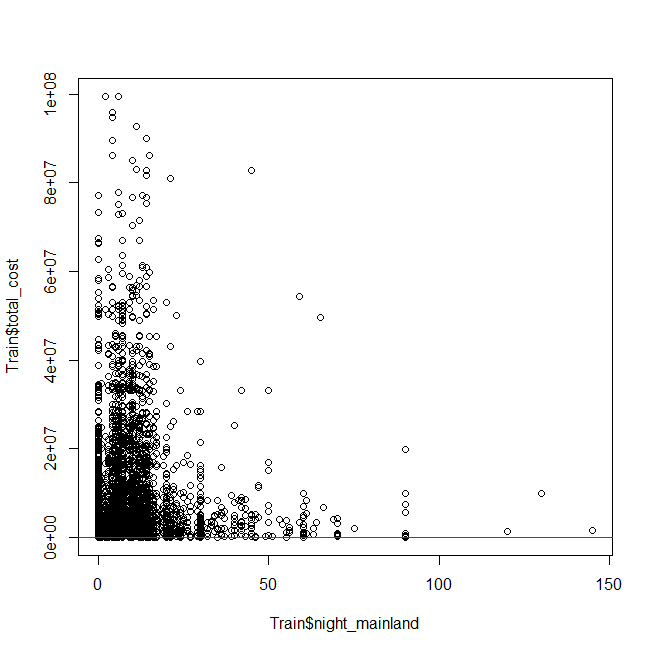
Code

plot(Train$night\_mainland,Train$total\_cost)

abline(lm(Train$night\_mainland ~ Train$total\_cost),col="red")

Comment: The result in figure show that the correlation between the two variables night\_mainland and total\_cost is very weak that means there’s no statistically significant relation

Output



Title :Age categories of night zanzibar visitors .

Code :

ggplot(Train,aes(country,night\_mainland,fill=country))+

geom\_col()+

coord\_flip()

Comment:

According to the graph below ,the most tourists of night mainland are from :

India , Germany , United Kingdom , Kenya , France , Netherland , Oman , Martdus , Uganda,

Bermuda , Pakistan and Belgium .

Output:



Title: Correlation between night mainland and night zanzibar .

Code:

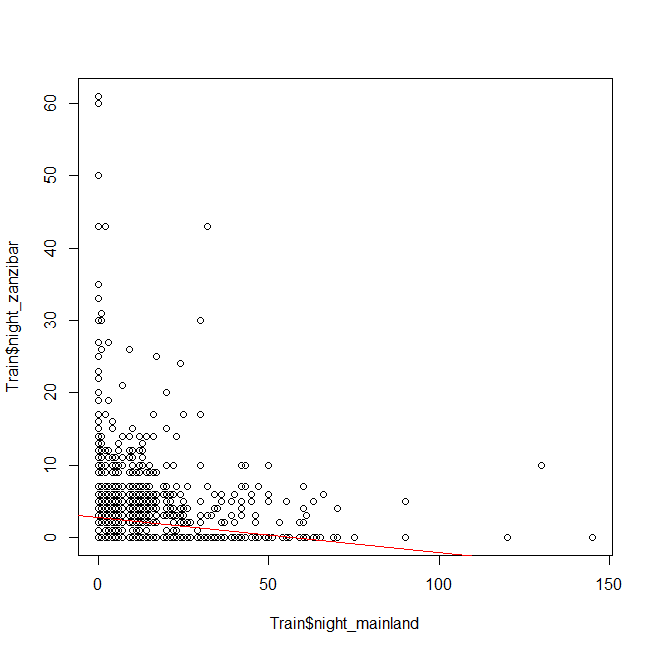
plot(Train$night\_mainland,Train$night\_zanzibar)

abline(lm(Train$night\_zanzibar ~ Train$night\_mainland),col="red")

Comment:

The result in the figure shows that the correlation between the two variables night\_mainland and night \_ zanzibar is very weak, which means there’s no statistically significant relation .

Output :



Title: :Age categories of night zanzibar visitors .

Code:

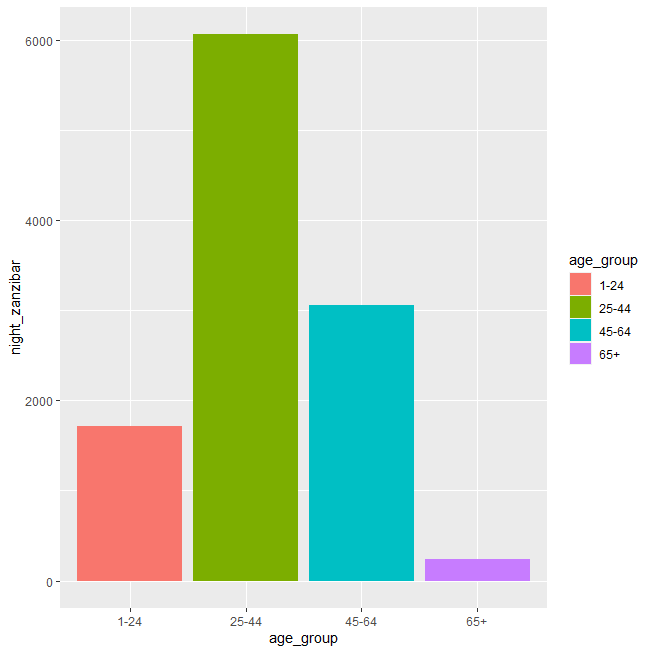
ggplot(Train,aes(age\_group,night\_zanzibar,fill=age\_group))+

geom\_col()

Comment:The most night zanzibar tourists ages categories are :

25-44 . and 45-64 .

Output



Title Tourists purpose and the total cost.

Code

plot(Train$purpose, Train$total\_cost, col =Train$purpose,pch = 16,

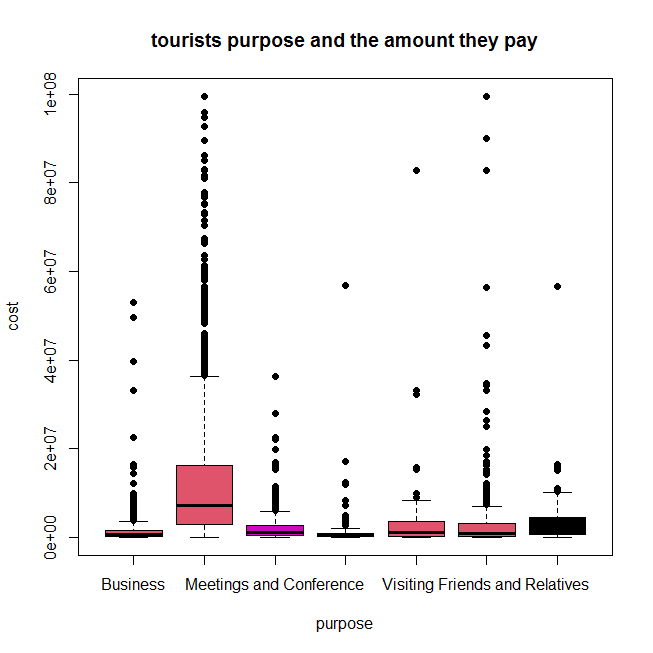
xlab="purpose",ylab="cost",

main="tourists purpose and the amount they pay")

Comment:

Tourists for Meeting ,Volunteers,scientific and Academic

Output



Title : Night mainland tourists and the tour arrangement .

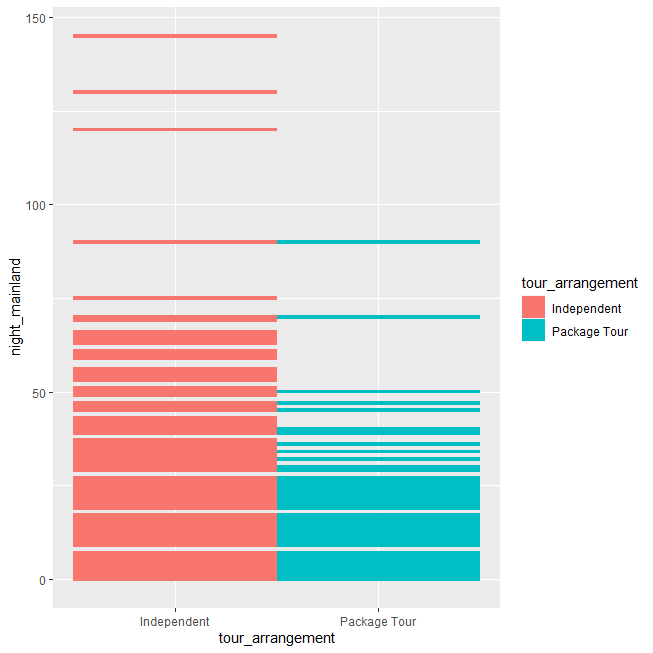
Code:

ggplot(data = Train) +

geom\_raster(aes(x =tour\_arrangement, y = night\_mainland, fill =tour\_arrangement))

Comment : Most night mainland tourists are independent .

Output



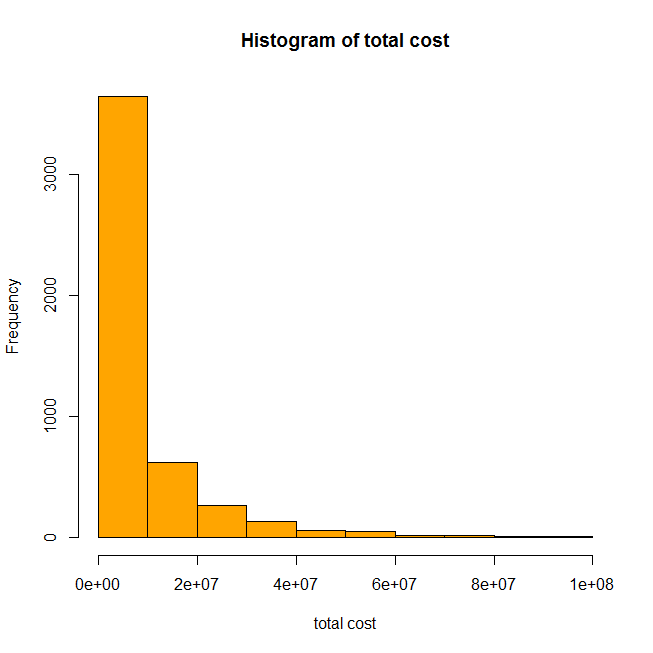
Title Total cost

Code:

hist(Train$total\_cost, breaks = 10, col = "orange", main = "Histogram of total cost", xlab = "total cost")

Comment Most tourists total total cost is 20000000 .

Output



Title

Correlation between total male and total female .

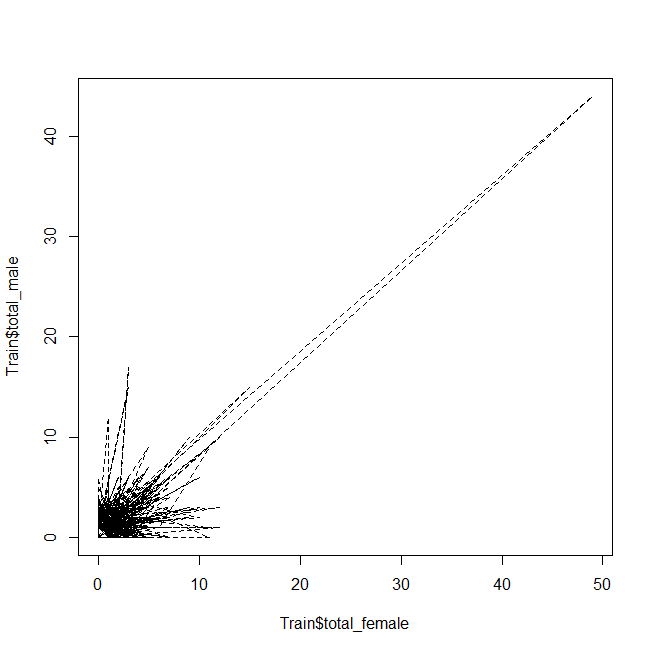
Code

plot(Train$total\_female,Train$total\_male, type = "l", lty = "dashed")

Comment

The line chart shows correlation between total female and male .

Output



Title

Correlation between total male and total female .

Code

plot(Train$total\_female,Train$total\_male)

abline(lm(Train$total\_female ~ Train$total\_male),col="red")

Comment

The line chart shows correlation between total female and male .

Output

