PIMPRI CHINCHWAD EDUCATION TRUST'S

PIMPRI CHINCHWAD COLLEGE OF ENGINEERING



Department of AS&H

Report of IE-2 Activity

Subject : Statistical Data Analysis using R

Academic Year : 2022-23

Semester **I**

Submitted by:

| Name of Student | Roll no | Branch |
| --- | --- | --- |
| Rishi Shinde | TYCOC192 | Computer Engineering |
| Ayush Talegaonkar | TYCOC198 | Computer Engineering |
| Amaan Tamboli | TYCOC201 | Computer Engineering |
| Abhinav Tekam | TYCOC204 | Computer Engineering |
| Bhushan Wakchaure | TYCOC212 | Computer Engineering |

| Submitted to : Dr. Ganesh Tarte  Date : /11/2022  Sign : |  |
| --- | --- |

**Project Title:**

**Data used: Mall customer dataset**

**Details of the data:** Data contains 200 observations and 5 variables as follows:

1. Customer id : ID of the customers.
2. Gender : Gender of the customers
3. Age : Age of the customers
4. Annual income : The total annual income of the customers (in k$)
5. Spending score : The score calculated by taking into account the gender, age and annual income of the customers.

Code:

customer\_data=read.csv("/home/dataflair/Mall\_Customers.csv")

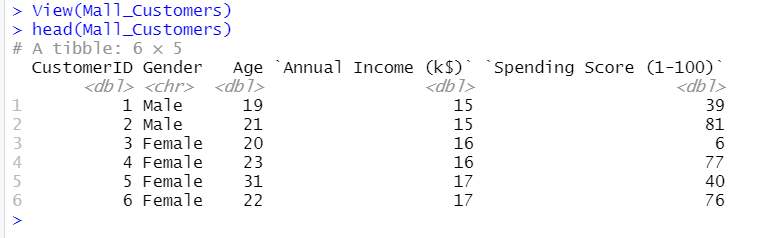
View(Mall\_Customers)

head(Mall\_Customers)

library(tidyr)

library(ggplot2)

Output:



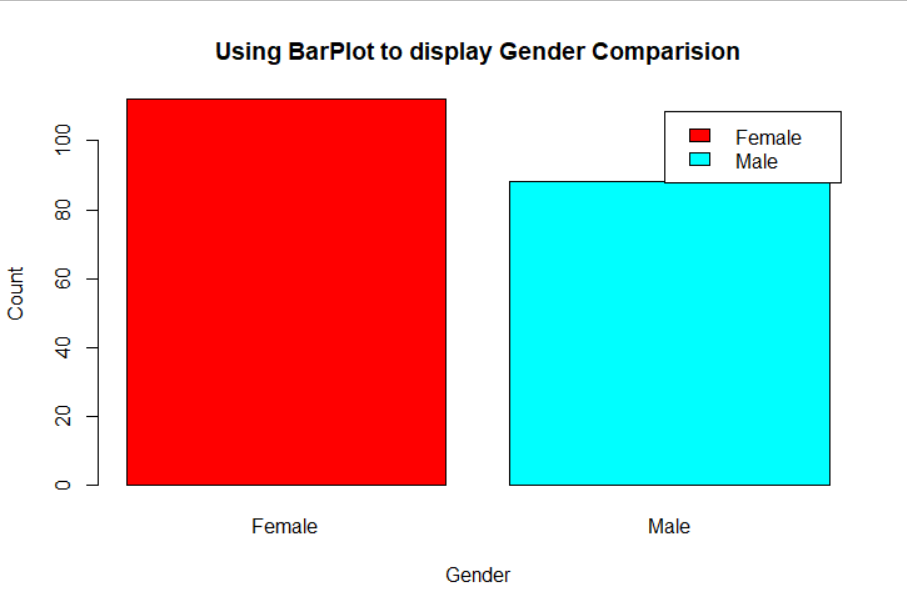
**Problem statement 1: What is the gender comparison among customers?**

Code:

a=table(Mall\_Customers$Gender)

barplot(a,main="Using BarPlot to display GenderComparision",ylab="Count",xlab="Gender",col=rainbow(2),legend=rownames(a))

output:

****

It can be seen clearly that the number of females is more than that of males. Hence it can be concluded that Females live to shop more than men.

# **Problem statement 2: What is the age class of the customers?**

Code:

hist(Mall\_Customers$Age,

col="blue",

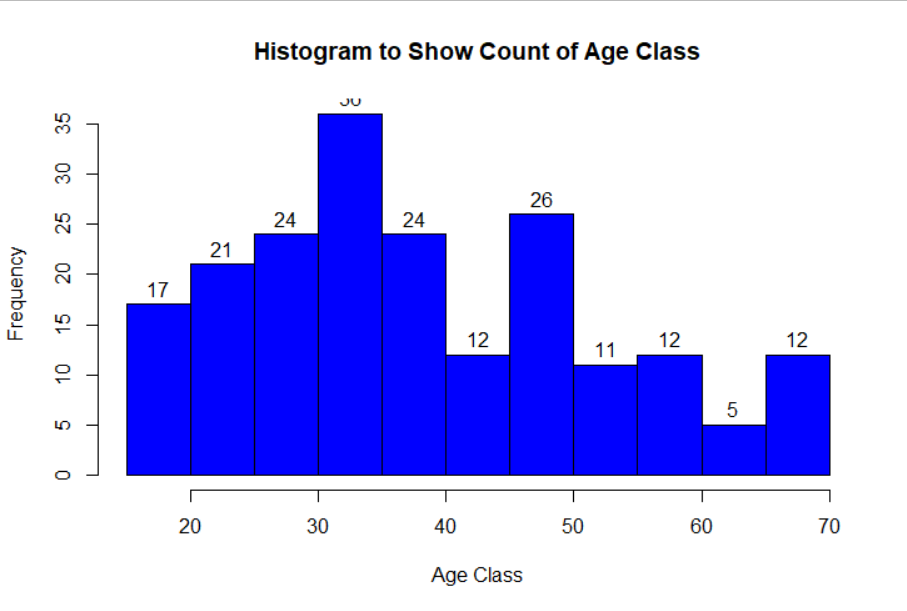
main="Histogram to Show Count of Age Class",

xlab="Age Class",

ylab="Frequency",

labels=TRUE)

Output:



From the histogram, it is evident that the majority of the customers are from the age group 25-40 i.e. the active working age group.

# 

# **Problem statement 3: What is the demographics of annual income of the customers?**

# **Code:**

plot(density(Mall\_Customers1$annual\_income),

col="yellow",

main="Density Plot for Annual Income",

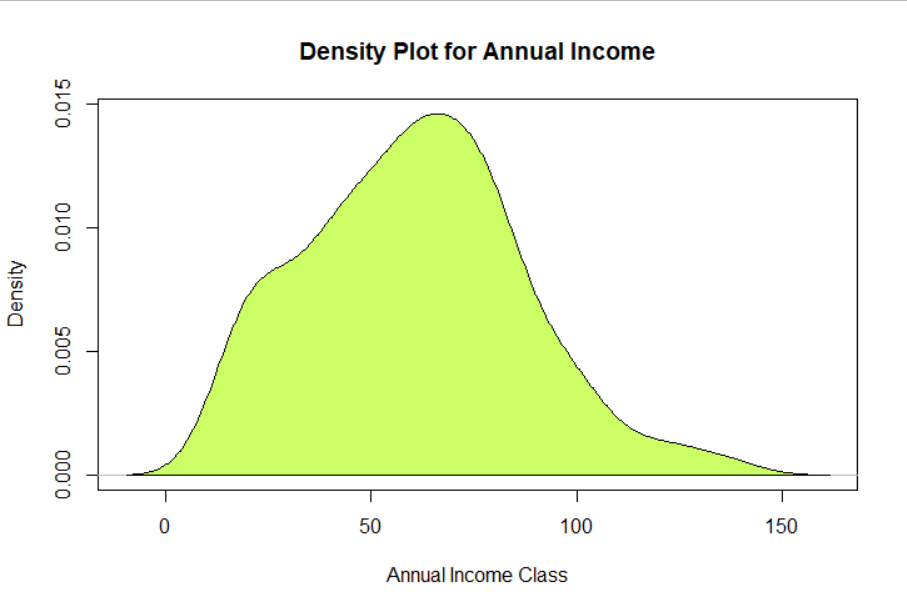
xlab="Annual Income Class",

ylab="Density")

polygon(density(Mall\_Customers1$annual\_income),

col="#ccff66")

Output:



From the density plot, we can conclude that the median annual income of the customers is around $60k

# 

# 

# 

# **Problem statement 4: What is the spending score demographic of the customers?**

Code:

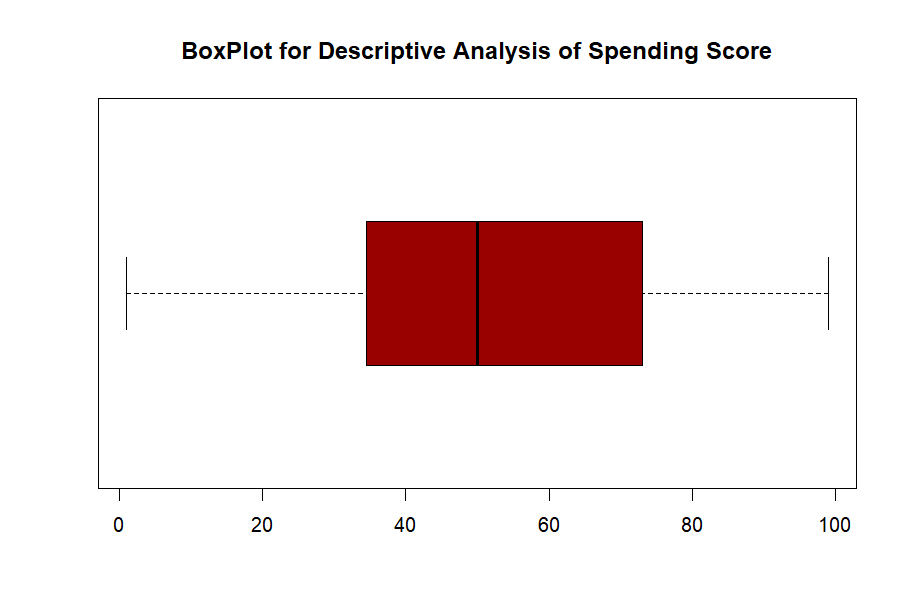
boxplot(Mall\_Customers1$spending\_score,

horizontal=TRUE,

col="#990000",

main="BoxPlot for Descriptive Analysis of Spending Score")

Output:



The boxplot above shows that the median spending score of the customers is around 48.

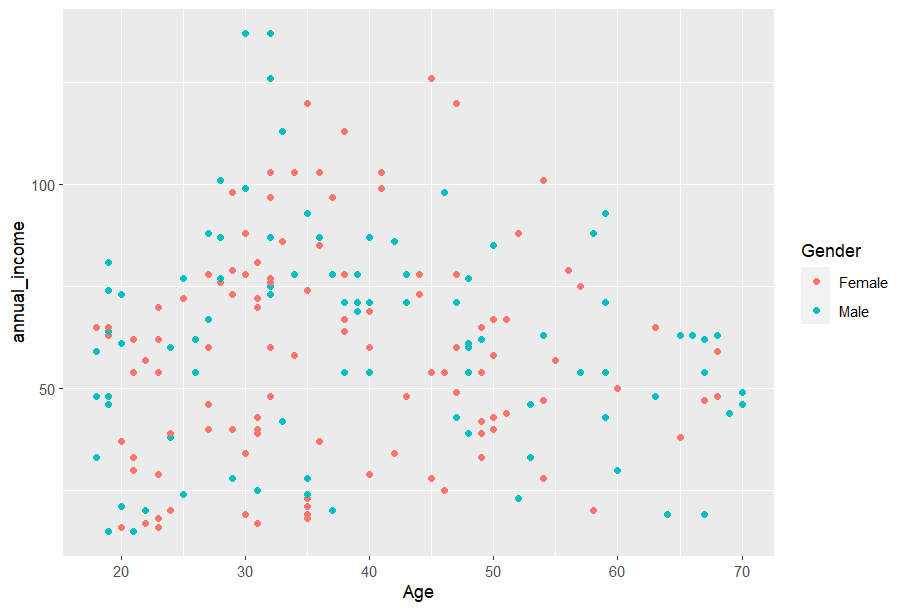
# 

# **Problem statement 5: The demographic of annual income vs the age of the customers?**

Code:

ggplot(Mall\_Customers1) +

geom\_point(aes(x = Age, y = annual\_income, col = Gender))



The cluster concludes that young customers are more likely to earn more than the old ones. Men earn more than females.

# 

# 

# **Problem statement 6: The demographic of Age vs the Spending score of the customers?**

Code:

ggplot(Mall\_Customers1) +

geom\_point(aes(x = Age, y = spending\_score, col = Gender))

Output:



Young customers spend more of their income than the old customers, as the needs decrease as we grow old. Young females and males seem to be on equal footing in case of spending

**Overall Conclusion:**

1. The number of Females that go to the mall is more than the number of males.
2. Active working population frequently visits malls.
3. The mall customers have a median income og $60k.
4. The spending score of the customers is about 48.
5. The customers in the age range of 30-40 earn more.
6. Young population is more likely to spend more at malls.