```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("stats")
library("ggplot2")
library("dplyr")
library("stats")
customer<- read.csv("/cloud/project/Mall_Customers.csv")
str(customer)
names(customer)
head(customer)
summary(customer$Age)
sd(customer$Age)
summary(customer$Annual.Income..k..)
sd(customer$Annual.Income..k..)
summary(customer$Age)
sd(customer$Spending.Score..1.100.)
x<- "Customer Gender Visualization"
x
a=table(customer$Gender)
barplot(a,main="Using BarPlot to display Gender Comparision",
        ylab="Count",
        xlab="Gender",
        col=rainbow(2),
        legend=rownames(a))
install.packages("plotrix")
library("plotrix")
pct=round(a/sum(a)*100)
lbs=paste(c("Female","Male")," ",pct,"%",sep=" ")
library(plotrix)
pie3D(a,labels=lbs,
      main="Pie Chart Depicting Ratio of Female and Male")
y<- "visualization of age"
y
summary(customer$Age)
hist(customer$Age,
     col="blue",
     main="Histogram to Show Count of Age Class",
     xlab="Age Class",
     ylab="Frequency",
     labels=TRUE)
boxplot(customer$Age,
        col="ff0066",
        main="Boxplot for Descriptive Analysis of Age")
summary(customer$Annual.Income..k..)
hist(customer$Annual.Income..k..,
     col="#660033",
     main="Histogram for Annual Income",
     xlab="Annual Income Class",
     ylab="Frequency",
     labels=TRUE)
plot(density(customer$Annual.Income..k..),
     col="yellow",
     main="Density Plot for Annual Income",
     xlab="Annual Income Class",
     ylab="Density")
polygon(density(customer$Annual.Income..k..),
        col="#ccff66")
summary(customer$Spending.Score..1.100.)
boxplot(customer$Spending.Score..1.100.,
        horizontal=TRUE,
        col="#990000",
        main="BoxPlot for Descriptive Analysis of Spending Score")
hist(customer$Spending.Score..1.100.,
     main="HistoGram for Spending Score",
```

```
      xlab="Spending Score Class",
      ylab="Frequency",
      col="#6600cc",
      labels=TRUE)

install.packes("purrr")
library(purrr)
set.seed(123)
iss <- function(k) {
  kmeans(customer[,3:5],k,iter.max=100,nstart=100,algorithm="Lloyd" )$tot.withinss
}
k.values <- 1:10

iss_values <- map_dbl(k.values, iss)

plot(k.values, iss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total intra-clusters sum of squares")


install.packages("cluster")
install.packages("gridExtra")
install.packages("grid")
library(cluster)
library(gridExtra)
library(grid)
k2<-kmeans(customer[,3:5],2,iter.max=100,nstart=50,algorithm="Lloyd")
s2<-plot(silhouette(k2$cluster,dist(customer[,3:5],"euclidean")))

install.packages("NbClust")
install.packages("factoextra")
library(NbClust)
library(factoextra)
fviz_nbclust(customer[,3:5], kmeans, method = "silhouette")
set.seed(125)
stat_gap <- clusGap(customer[,3:5], FUN = kmeans, nstart = 25,
                    K.max = 10, B = 50)
fviz_gap_stat(stat_gap)

k6<-kmeans(customer[,3:5],6,iter.max=100,nstart=50,algorithm="Lloyd")
k6
pcclust=prcomp(customer[,3:5],scale=FALSE) #principal component analysis
summary(pcclust)
pcclust$rotation[,1:2]
set.seed(1)
ggplot(customer, aes(x =Annual.Income..k.., y = Spending.Score..1.100.)) +
  geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
  scale_color_discrete(name=" ",
                       breaks=c("1", "2", "3", "4", "5","6"),
                       labels=c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4", "Cluster
5","Cluster 6")) +
  ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")

ggplot(customer, aes(x =Spending.Score..1.100., y =Age)) +
  geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
  scale_color_discrete(name=" ",
                       breaks=c("1", "2", "3", "4", "5","6"),
                       labels=c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4", "Cluster
5","Cluster 6")) +
  ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")

kCols=function(vec){cols=rainbow (length (unique (vec)))
return (cols[as.numeric(as.factor(vec))])}
```

```
digCluster<-k6$cluster; dignm<-as.character(digCluster); # K-means clusters

plot(pcclust$x[,1:2], col =kCols(digCluster),pch =19,xlab ="K-means",ylab="classes")
legend("bottomleft",unique(dignm),fill=unique(kCols(digCluster)))
```