Untitled1* | Untitled2* | RGLOB_A9.R | Untitled3* | Mall_Customers

Source on Save | Run | Source

```
1  customer_data=read.csv("C:\\Users\\manne\\Downloads\\Mall_Customers.csv")
2  str(customer_data)
3
4  names(customer_data)
5  head(customer_data)
6
7  summary(customer_data$Age)
8  sd(customer_data$Age)
9  summary(customer_data$Annual.Income..LPA.)
10 sd(customer_data$Annual.Income..LPA.)
11 summary(customer_data$Age)
12 sd(customer_data$Spending.Score..1.100.)
13
14
15 #Customer Gender Visualization
16
17 a=table(customer_data$Gender)
18 barplot(a,main="Using BarPlot to display Gender Comparision",
19
```

12:41 (Top Level) R Script

Console | Terminal | Background Jobs

R 4.3.2 · ~/

```
> sd(customer_data$Age)
[1] 13.95149
> summary(customer_data$Annual.Income..LPA.)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.00   41.50   61.50   60.56   78.00  137.00
> sd(customer_data$Annual.Income..LPA.)
[1] 26.23179
> summary(customer_data$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   28.75   36.00   38.85   49.00   70.00
> sd(customer_data$Spending.Score..1.100.)
[1] 25.79114
>
```

RStudio

e Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
13
14
15  #Customer Gender Visualization
16
17  a=table(customer_data$Gender)
18  barplot(a,main="Using BarPlot to display Gender Comparision",
19          ylab="Count",
20          xlab="Gender",
21          col=rainbow(2),
22          legend=rownames(a))
23
24
25  pct=round(a/sum(a)*100)
26  lbs=paste(c("Female","Male")," ",pct,"%",sep=" ")
27  library(plotrix)
28  pie3D(a,labels=lbs,
29        main="Pie Chart Depicting Ratio of Female and Male")
30
31
32  #Visualization of Age Distribution
33
```

26:50 (Top Level) R Script

Console | Terminal | Background Jobs

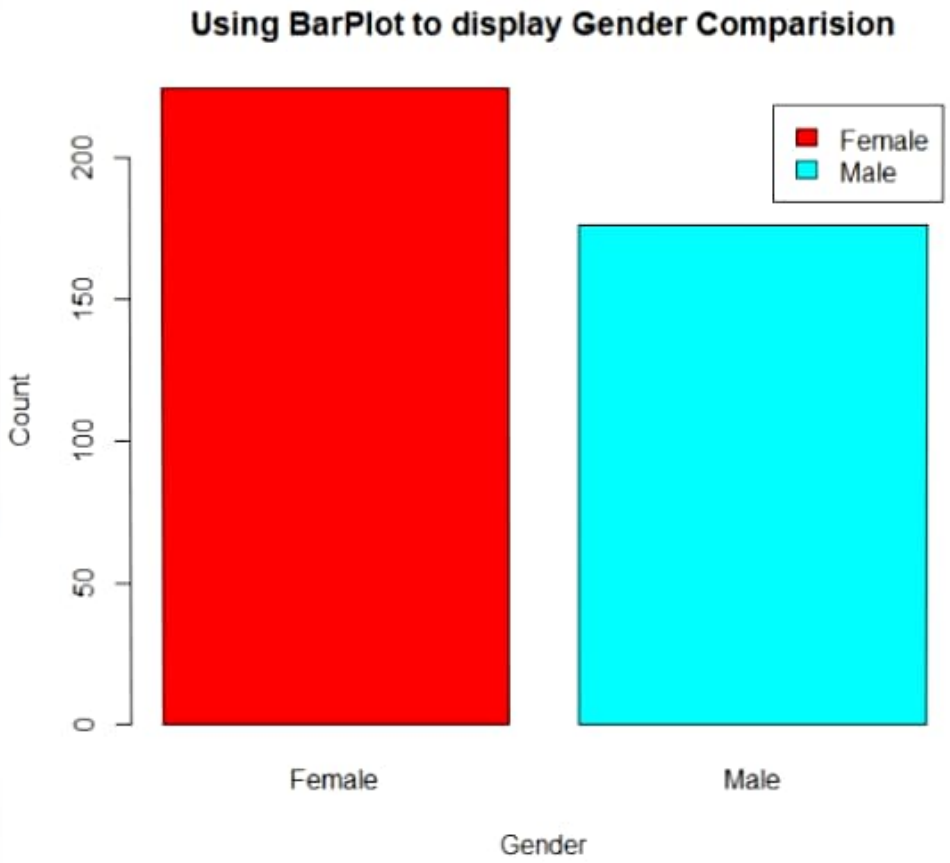
R 4.3.2 · ~/

```
> a=table(customer_data$Gender)
> barplot(a,main="Using BarPlot to display Gender Comparision",
+         ylab="Count",
+         xlab="Gender",
+         col=rainbow(2),
+         legend=rownames(a))
>
>
> pct=round(a/sum(a)*100)
> lbs=paste(c("Female","Male")," ",pct,"%",sep=" ")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search

ENG IN 20:09 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
20          xlab="Gender",
21          col=rainbow(2),
22          legend=rownames(a))
23
24
25  pct=round(a/sum(a)*100)
26  lbs=paste(c("Female","Male")," ",pct,"%",sep=" ")
27  library(plotrix)
28  pie3D(a,labels=lbs,
29        main="Pie Chart Depicting Ratio of Female and Male")
30
31
32  #Visualization of Age Distribution
33
34
35  summary(customer_data$Age)
36
37  hist(customer_data$Age,
38       col="blue",
39       main="Histogram to Show Count of Age Class",
40
```

29:59 (Top Level) R Script

Console | Terminal | Background Jobs

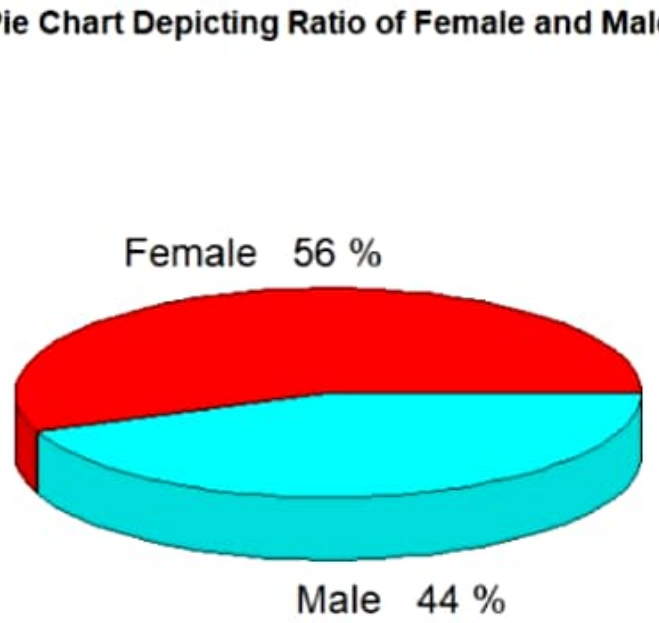
R 4.3.2 · ~/

```
+          xlab="Gender",
+          col=rainbow(2),
+          legend=rownames(a))
>
>
> 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")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:09 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
28 pie3D(a,labels=lbs,
29       main="Pie Chart Depicting Ratio of Female and Male")
30
31
32 #Visualization of Age Distribution
33
34
35 summary(customer_data$Age)
36
37 hist(customer_data$Age,
38      col="blue",
39      main="Histogram to Show Count of Age Class",
40      xlab="Age Class",
41      ylab="Frequency",
42      labels=TRUE)
43
44 boxplot(customer_data$Age,
45         col="#ff0066",
46         main="Boxplot for Descriptive Analysis of Age")
47
48
```

42:18 (Top Level) R Script

Console | Terminal | Background Jobs

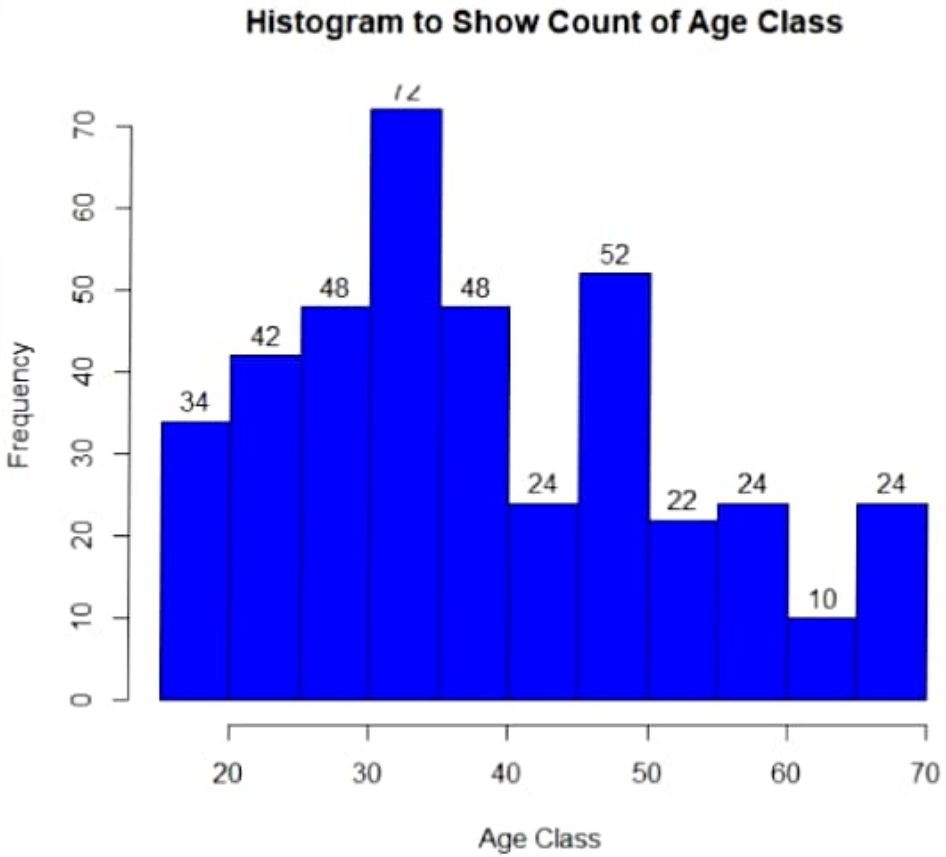
R 4.3.2 · ~/

```
> summary(customer_data$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   28.75   36.00   38.85   49.00   70.00
>
> hist(customer_data$Age,
+      col="blue",
+      main="Histogram to Show Count of Age Class",
+      xlab="Age Class",
+      ylab="Frequency",
+      labels=TRUE)
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



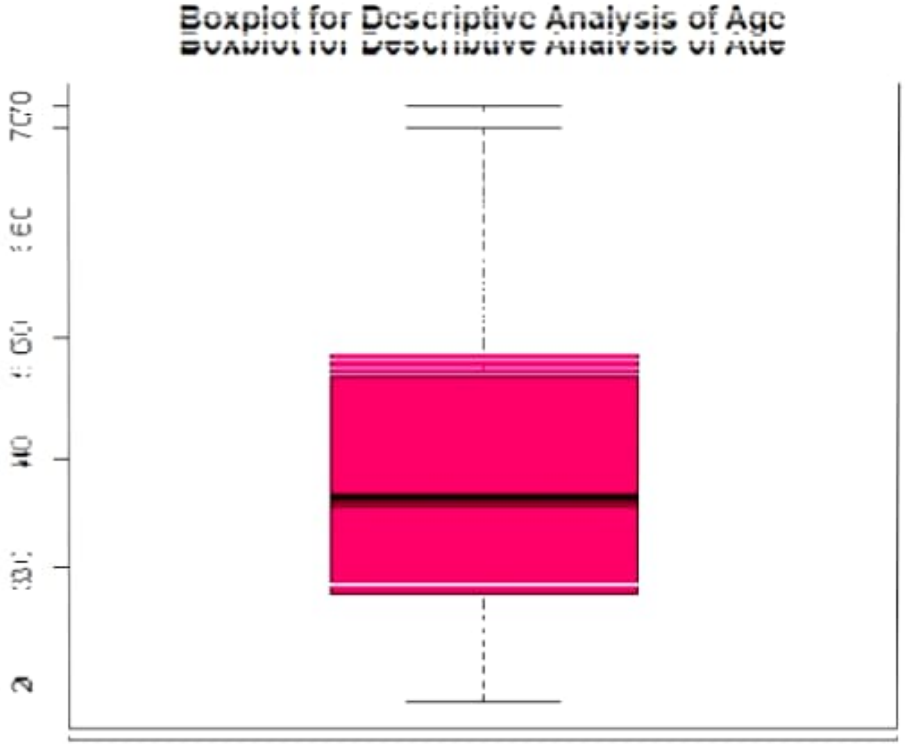
Search | ENG IN | 20:10 04-01-2024

RStudio

e Edit Code View Plots Session Build Debug Profile Tools Help

```
boxplot(customer_data$Age,
boxplot(customer_data$Age,
        col="#ff0000",
summary(customer_data$Annual.Income..LPA.)
hist(customer_data$Annual.Income..LPA.,
     col="#660033",
     main="Histogram for Annual Income",
     xlab="Annual Income Class",
     ylab="Frequency",
     labels=TRUE)
plot(density(customer_data$Annual.Income..LPA.),
```

```
> boxplot(customer_data$Age,
+         col="#ff00bb",
+         main="Boxplot for Descriptive Analysis of Age")
>
```



Search

ENG
IN

20:10
04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
46          main="Boxplot for Descriptive Analysis of Age")
47
48  #Analysis of the Annual Income of the Customers
49
50  summary(customer_data$Annual.Income..LPA.)
51  hist(customer_data$Annual.Income..LPA.,
52       col="#660033",
53       main="Histogram for Annual Income",
54       xlab="Annual Income Class",
55       ylab="Frequency",
56       labels=TRUE)
57
58  plot(density(customer_data$Annual.Income..LPA.),
59       col="blue",
60       main="Density Plot for Annual Income",
61       xlab="Annual Income Class",
62       ylab="Density")
63  polygon(density(customer_data$Annual.Income..LPA.),
64          col="#ccff66")
65
66
```

56:18 (Top Level) | R Script

Console | Terminal | Background Jobs

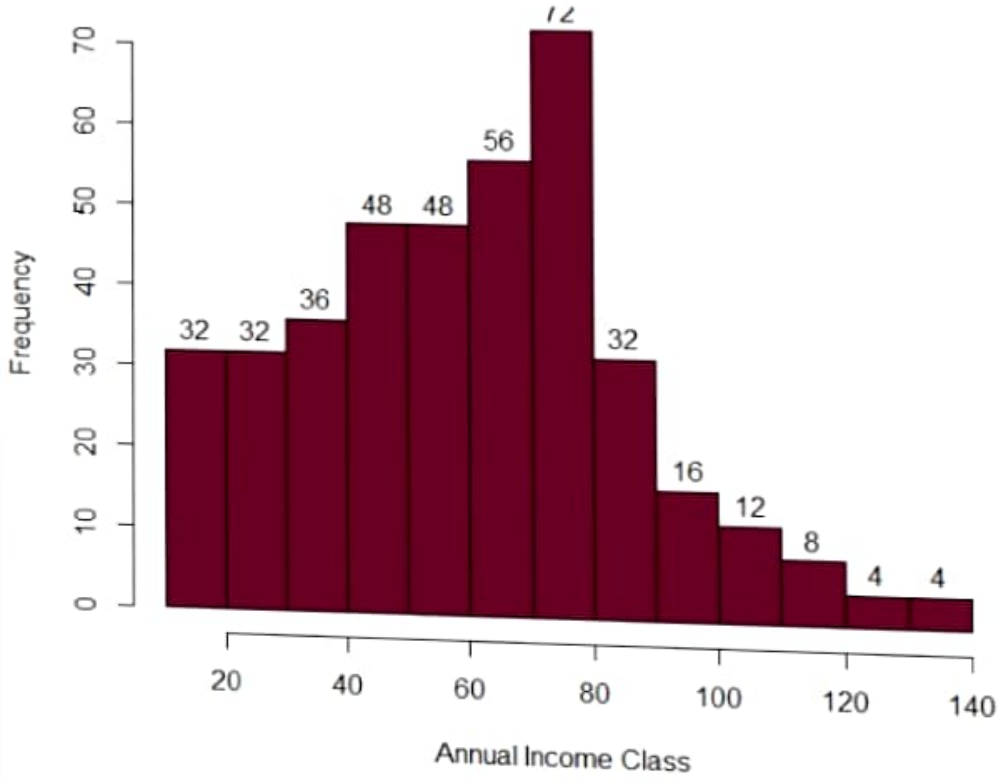
R 4.3.2 · ~/

```
> summary(customer_data$Annual.Income..LPA.)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.00   41.50   61.50   60.56   78.00  137.00
> hist(customer_data$Annual.Income..LPA.,
+      col="#660033",
+      main="Histogram for Annual Income",
+      xlab="Annual Income Class",
+      ylab="Frequency",
+      labels=TRUE)
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
56        labels=TRUE)
57
58  plot(density(customer_data$Annual.Income..LPA.),
59       col="blue",
60       main="Density Plot for Annual Income",
61       xlab="Annual Income Class",
62       ylab="Density")
63  polygon(density(customer_data$Annual.Income..LPA.),
64          col="#ccff66")
65
66  boxplot(customer_data$Spending.Score..1.100.,
67          horizontal=TRUE,
68          col="#990000",
69          main="BoxPlot for Descriptive Analysis of Spending Score")
70
71  hist(customer_data$Spending.Score..1.100.,
72       main="HistoGram for Spending Score",
73       xlab="Spending Score Class",
74       ylab="Frequency",
75       col="#6600cc",
76
```

62:21 (Top Level) R Script

Console | Terminal | Background Jobs

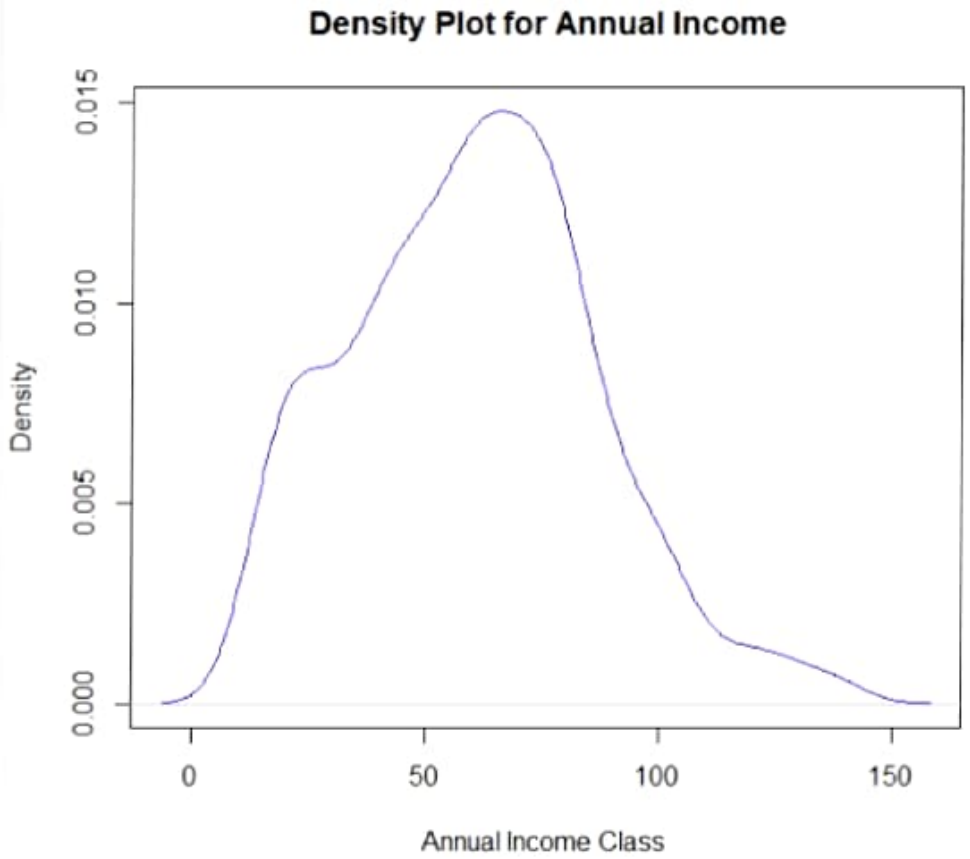
R 4.3.2 · ~/

```
> plot(density(customer_data$Annual.Income..LPA.),
+      col="blue",
+      main="Density Plot for Annual Income",
+      xlab="Annual Income Class",
+      ylab="Density")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search

ENG IN

20:11 04-01-2024

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
      col="blue",
      main="Density Plot for Annual Income",
      xlab="Annual Income Class",
      ylab="Density")
polygon(density(customer_data$Annual.Income..LPA.),
        col="#ccff66")

boxplot(customer_data$Spending.Score..1.100.,
        horizontal=TRUE,
        col="#990000",
        main="BoxPlot for Descriptive Analysis of Spending Score")

hist(customer_data$Spending.Score..1.100.,
     main="HistoGram for Spending Score",
     xlab="Spending Score Class",
     ylab="Frequency",
     col="#6600cc",
     labels=TRUE)

#K-means Algorithm
```

64:23 (Top Level) R Script

Console | Terminal | Background Jobs

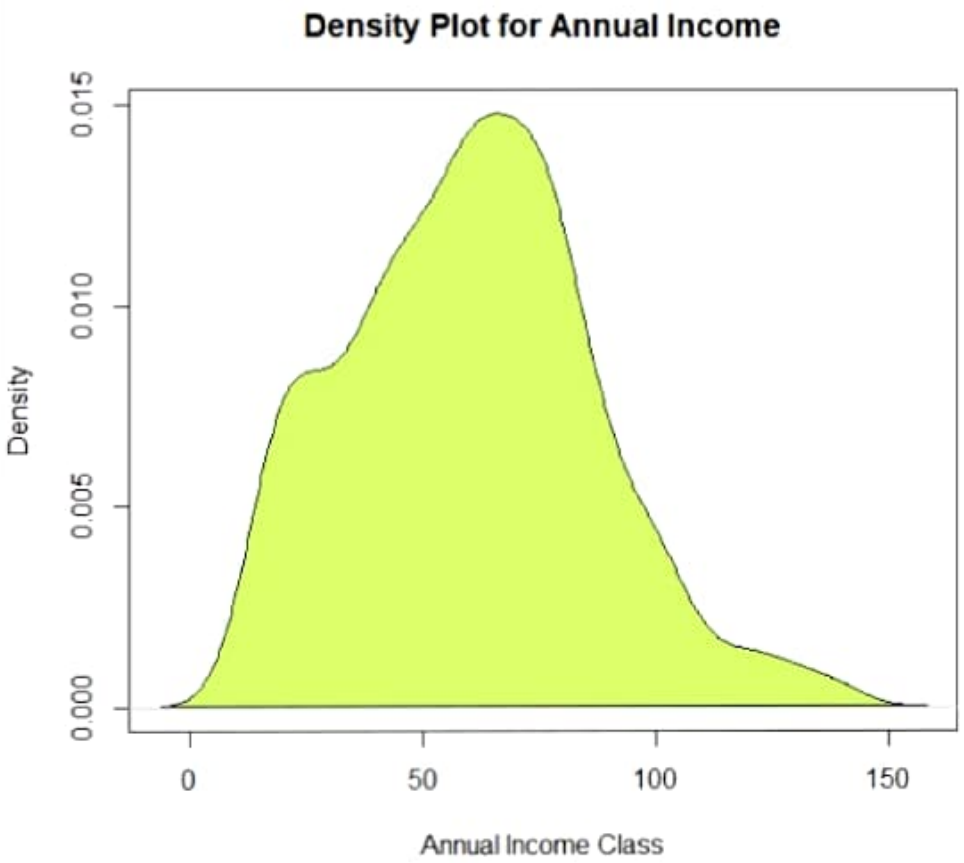
R 4.3.2 · ~/

```
> plot(density(customer_data$Annual.Income..LPA.),
+      col="blue",
+      main="Density Plot for Annual Income",
+      xlab="Annual Income Class",
+      ylab="Density")
> polygon(density(customer_data$Annual.Income..LPA.),
+         col="#ccff66")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:11 04-01-2024

RStudio

e Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
polygon(density(customer_data$Annual.Income..LPA.),
        col="#ccff66")

boxplot(customer_data$Spending.Score..1.100.,
        horizontal=TRUE,
        col="#990000",
        main="BoxPlot for Descriptive Analysis of Spending Score")

hist(customer_data$Spending.Score..1.100.,
     main="HistoGram for Spending Score",
     xlab="Spending Score Class",
     ylab="Frequency",
     col="#6600cc",
     labels=TRUE)

#K-means Algorithm

library(purrr)
set.seed(123)
# function to calculate total intra-cluster sum of square
```

69:67 (Top Level) R Script

Console | Terminal | Background Jobs

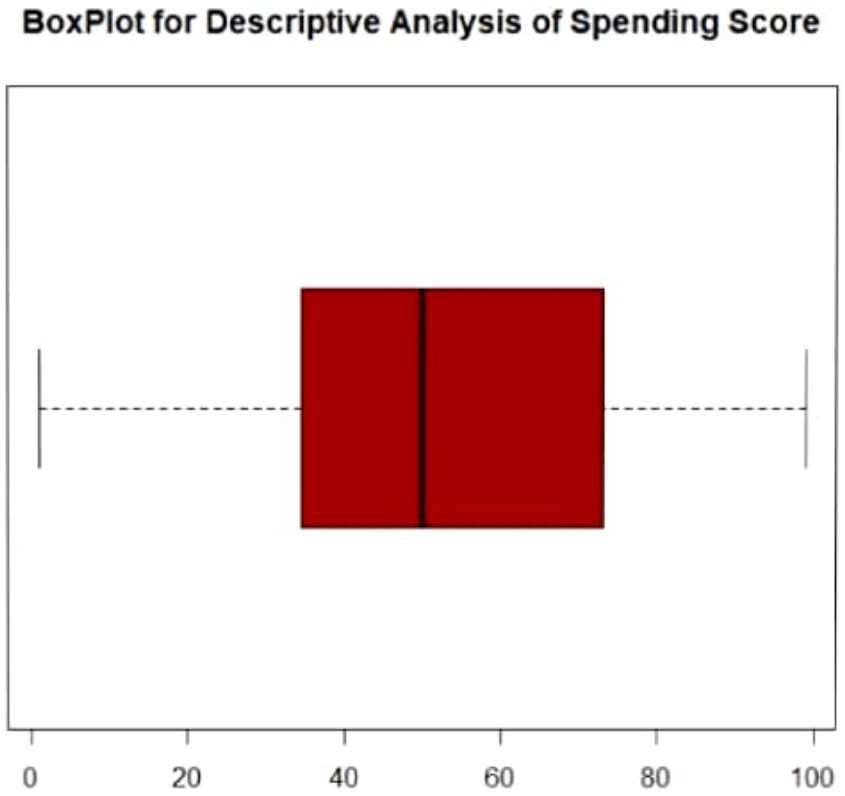
R 4.3.2 · ~/

```
> boxplot(customer_data$Spending.Score..1.100.,
+         horizontal=TRUE,
+         col="#990000",
+         main="BoxPlot for Descriptive Analysis of Spending Score")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search

ENG IN 20:12 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
68          col="#990000",
69          main="BoxPlot for Descriptive Analysis of Spending Score")
70
71  hist(customer_data$Spending.Score..1.100.,
72       main="HistoGram for Spending Score",
73       xlab="Spending Score Class",
74       ylab="Frequency",
75       col="#6600cc",
76       labels=TRUE)
77
78  #K-means Algorithm
79
80  library(purrr)
81  set.seed(123)
82  # function to calculate total intra-cluster sum of square
83  iss <- function(k) {
84    kmeans(customer_data[,3:5],k,iter.max=100,nstart=100,algorithm="Lloyd" )$tot.w
85  }
86
87  k.values <- 1:10
```

76:18 (Top Level) | R Script

Console | Terminal | Background Jobs

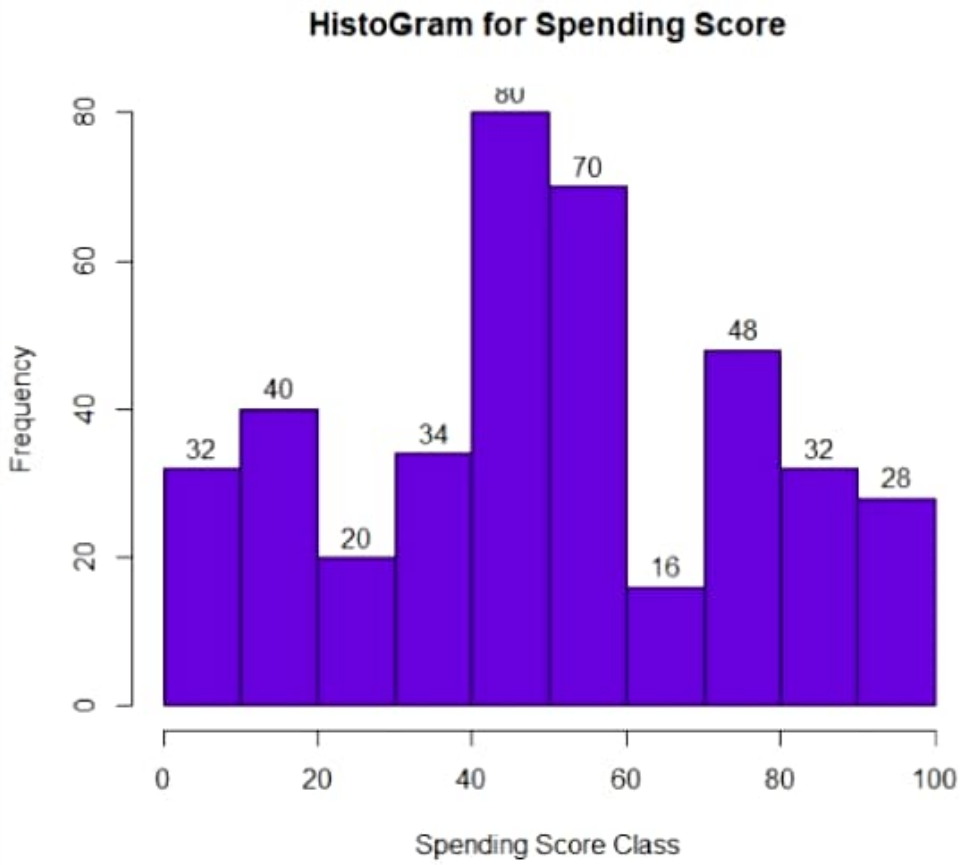
R 4.3.2 · ~/

```
> hist(customer_data$Spending.Score..1.100.,
+      main="HistoGram for Spending Score",
+      xlab="Spending Score Class",
+      ylab="Frequency",
+      col="#6600cc",
+      labels=TRUE)
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:12 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
82  # function to calculate total intra-cluster sum of square
83  iss <- function(k) {
84    kmeans(customer_data[,3:5],k,iter.max=100,nstart=100,algorithm="Lloyd" )$tot.w
85  }
86
87  k.values <- 1:10
88
89
90  iss_values <- map_dbl(k.values, iss)
91
92  plot(k.values, iss_values,
93       type="b", pch = 19, frame = FALSE,
94       xlab="Number of clusters K",
95       ylab="Total intra-clusters sum of squares")
96
97  #Average Silhouette Method
98
99  library(cluster)
100 library(gridExtra)
101 library(grid)
```

95:49 (Top Level) | R Script

Console | Terminal | Background Jobs

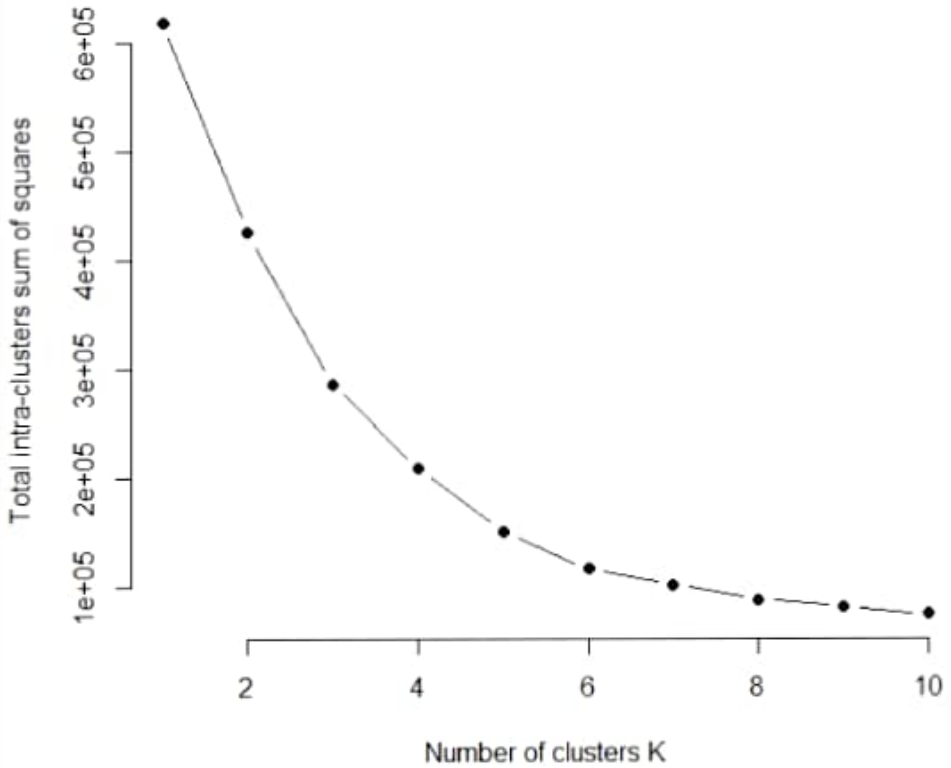
R 4.3.2 · ~/

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

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:14 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

customer seg rlab.R* | customer_data

Source on Save | Run | Source

```
101 library(factoextra)
102
103
104 k2<-kmeans(customer_data[,3:5],2,iter.max=100,nstart=50,algorithm="Lloyd")
105 s2<-plot(silhouette(k2$cluster,dist(customer_data[,3:5],"euclidean")))
106
107 k3<-kmeans(customer_data[,3:5],3,iter.max=100,nstart=50,algorithm="Lloyd")
108 s3<-plot(silhouette(k3$cluster,dist(customer_data[,3:5],"euclidean")))
109
110 k4<-kmeans(customer_data[,3:5],4,iter.max=100,nstart=50,algorithm="Lloyd")
111 s4<-plot(silhouette(k4$cluster,dist(customer_data[,3:5],"euclidean")))
112
113 k5<-kmeans(customer_data[,3:5],5,iter.max=100,nstart=50,algorithm="Lloyd")
114 s5<-plot(silhouette(k5$cluster,dist(customer_data[,3:5],"euclidean")))
115
116 k6<-kmeans(customer_data[,3:5],6,iter.max=100,nstart=50,algorithm="Lloyd")
117 s6<-plot(silhouette(k6$cluster,dist(customer_data[,3:5],"euclidean")))
118
119 k7<-kmeans(customer_data[,3:5],7,iter.max=100,nstart=50,algorithm="Lloyd")
120 s7<-plot(silhouette(k7$cluster,dist(customer_data[,3:5],"euclidean")))
121
122 k8<-kmeans(customer_data[,3:5],8,iter.max=100,nstart=50,algorithm="Lloyd")
123 s8<-plot(silhouette(k8$cluster,dist(customer_data[,3:5],"euclidean")))
124
125 k9<-kmeans(customer_data[,3:5],9,iter.max=100,nstart=50,algorithm="Lloyd")
126 s9<-plot(silhouette(k9$cluster,dist(customer_data[,3:5],"euclidean")))
127
128 k10<-kmeans(customer_data[,3:5],10,iter.max=100,nstart=50,algorithm="Lloyd")
129 s10<-plot(silhouette(k10$cluster,dist(customer_data[,3:5],"euclidean")))
130
131
132
```

104:1 (Top Level) | R Script

Console | Terminal | Background Jobs

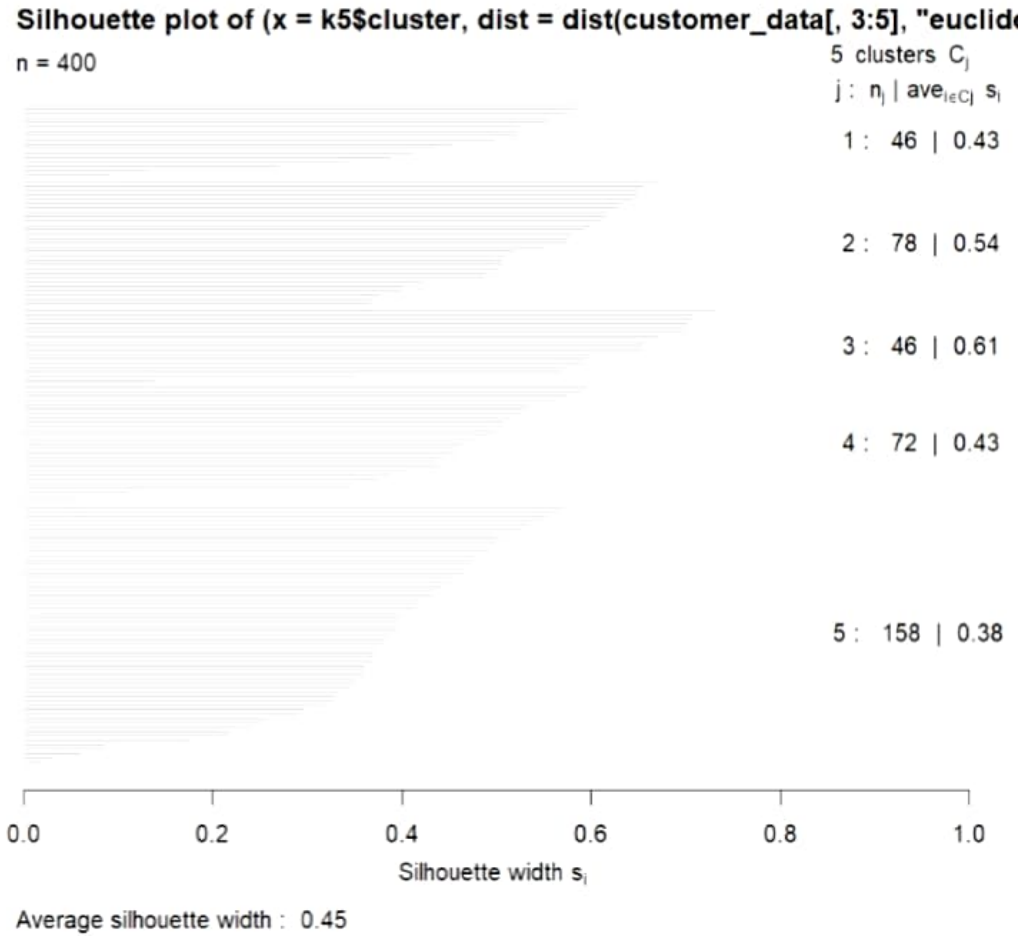
R 4.3.2 · ~/

```
> k9<-kmeans(customer_data[,3:5],9,iter.max=100,nstart=50,algorithm="Lloyd")
> s9<-plot(silhouette(k9$cluster,dist(customer_data[,3:5],"euclidean")))
>
> k10<-kmeans(customer_data[,3:5],10,iter.max=100,nstart=50,algorithm="Lloyd")
> s10<-plot(silhouette(k10$cluster,dist(customer_data[,3:5],"euclidean")))
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
129 s10<-plot(silhouette(k10$cluster,dist(customer_data[,3:5],"euclidean")))
130
131
132
133 library(NbClust)
134 library(factoextra)
135
136 fviz_nbclust(customer_data[,3:5], kmeans, method = "silhouette")
137
138 set.seed(125)
139 stat_gap <- clusGap(customer_data[,3:5], FUN = kmeans, nstart = 25,
140                     K.max = 10, B = 50)
141 fviz_gap_stat(stat_gap)
142
143 k6<-kmeans(customer_data[,3:5],6,iter.max=100,nstart=50,algorithm="Lloyd")
144 k6
145
146
147 #visualizing the Clustering Results using the First Two Principle Components
148
149
```

137:1 (Top Level) | R Script

Console | Terminal | Background Jobs

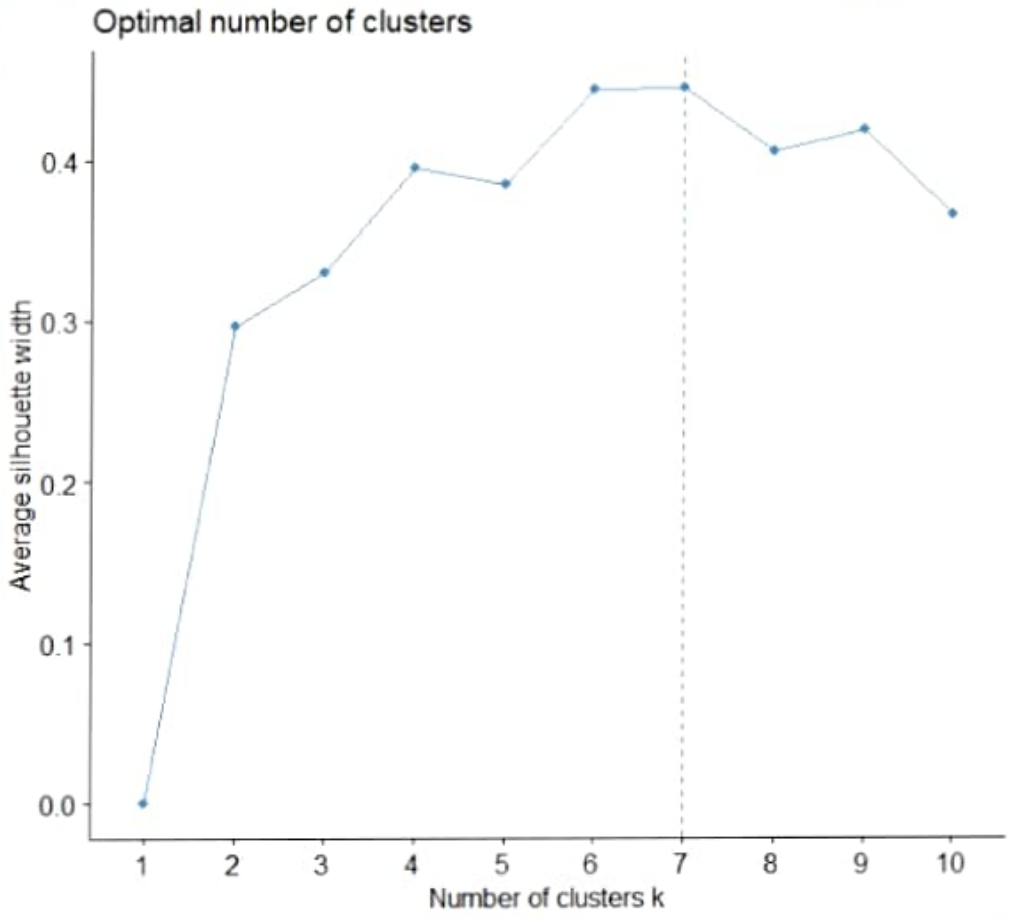
R 4.3.2 · ~/

```
> library(NbClust)
> library(factoextra)
>
> fviz_nbclust(customer_data[,3:5], kmeans, method = "silhouette")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search

ENG IN

20:15 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
129 s10<-plot(silhouette(k10$cluster,dist(customer_data[,3:5],"euclidean")))
130
131
132
133 library(NbClust)
134 library(factoextra)
135
136 fviz_nbclust(customer_data[,3:5], kmeans, method = "silhouette")
137
138 set.seed(125)
139 stat_gap <- clusGap(customer_data[,3:5], FUN = kmeans, nstart = 25,
140                     K.max = 10, B = 50)
141 fviz_gap_stat(stat_gap)
142
143 k6<-kmeans(customer_data[,3:5],6,iter.max=100,nstart=50,algorithm="Lloyd")
144 k6
145
146
147 #Visualizing the Clustering Results using the First Two Principle Components
148
149
```

144:3 (Top Level) R Script

Console | Terminal | Background Jobs

R 4.3.2 · ~/

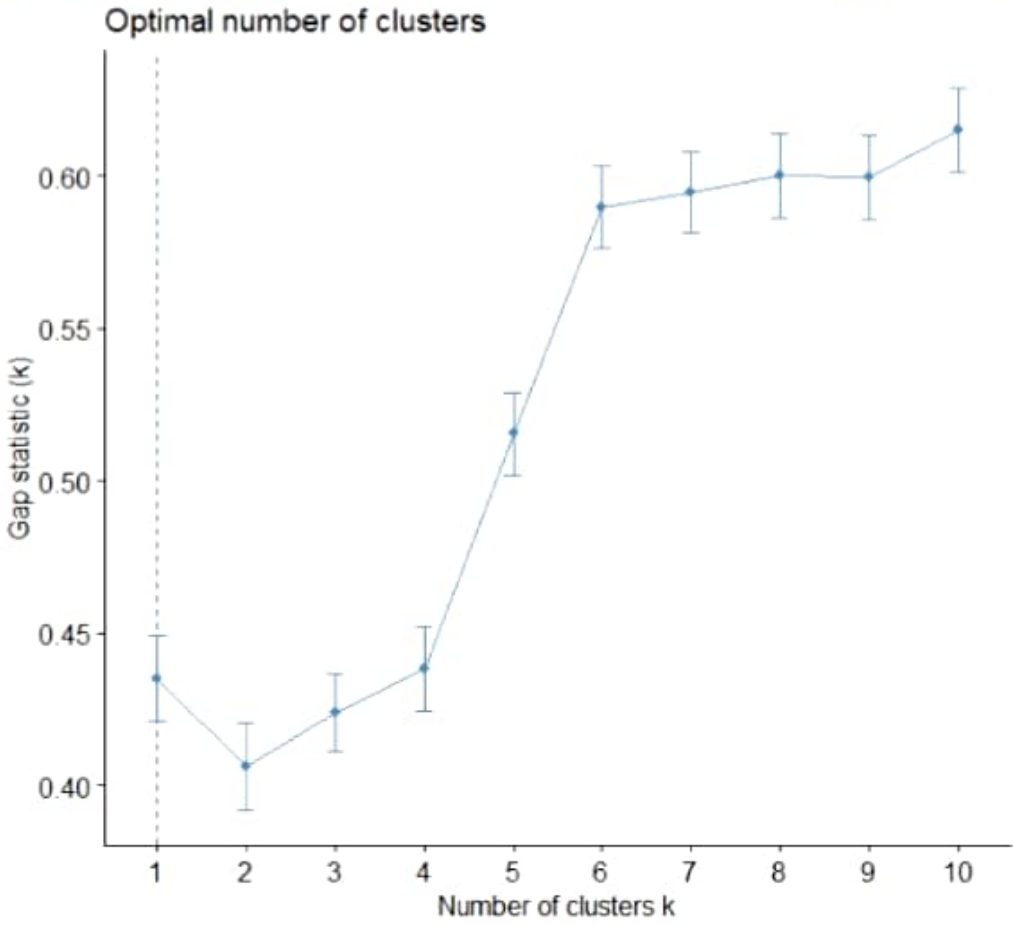
```
Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
> set.seed(125)
> stat_gap <- clusGap(customer_data[,3:5], FUN = kmeans, nstart = 25,
+                     K.max = 10, B = 50)
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 50)  [one "." per sample]:
.................................................
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
#Visualizing the Clustering Results using the First Two Principle Components

pcclust=prcomp(customer_data[,3:5],scale=FALSE) #principal component analysis
summary(pcclust)

pcclust$rotation[,1:2]



set.seed(1)
ggplot(customer_data, aes(x =Annual.Income..LPA., 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
  ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")



ggplot(customer_data, aes(x =Spending.Score..1.100., y =Age)) +
```

163:1 (Top Level) | R Script

Console | Terminal | Background Jobs

R 4.3.2 · ~/

```
Annual.Income..LPA.     -0.5886410 -0.8083757
Spending.Score..1.100. -0.7859965  0.5739136
> set.seed(1)
> ggplot(customer_data, aes(x =Annual.Income..LPA., 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")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish

Search | ENG IN | 20:17 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
157 ggplot(customer_data, aes(x =Annual.Income..LPA., y = Spending.Score..1.100.)) +
158   geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
159   scale_color_discrete(name=" ",
160                        breaks=c("1", "2", "3", "4", "5","6"),
161                        labels=c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4
162   ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")
163
164
165
166 ggplot(customer_data, aes(x =Spending.Score..1.100., y =Age)) +
167   geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
168   scale_color_discrete(name=" ",
169                        breaks=c("1", "2", "3", "4", "5","6"),
170                        labels=c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4
171   ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")
172
173
174
175 kCols=function(vec){cols=rainbow (length (unique (vec)))
176
```

172:1 (Top Level) | R Script

Console | Terminal | Background Jobs

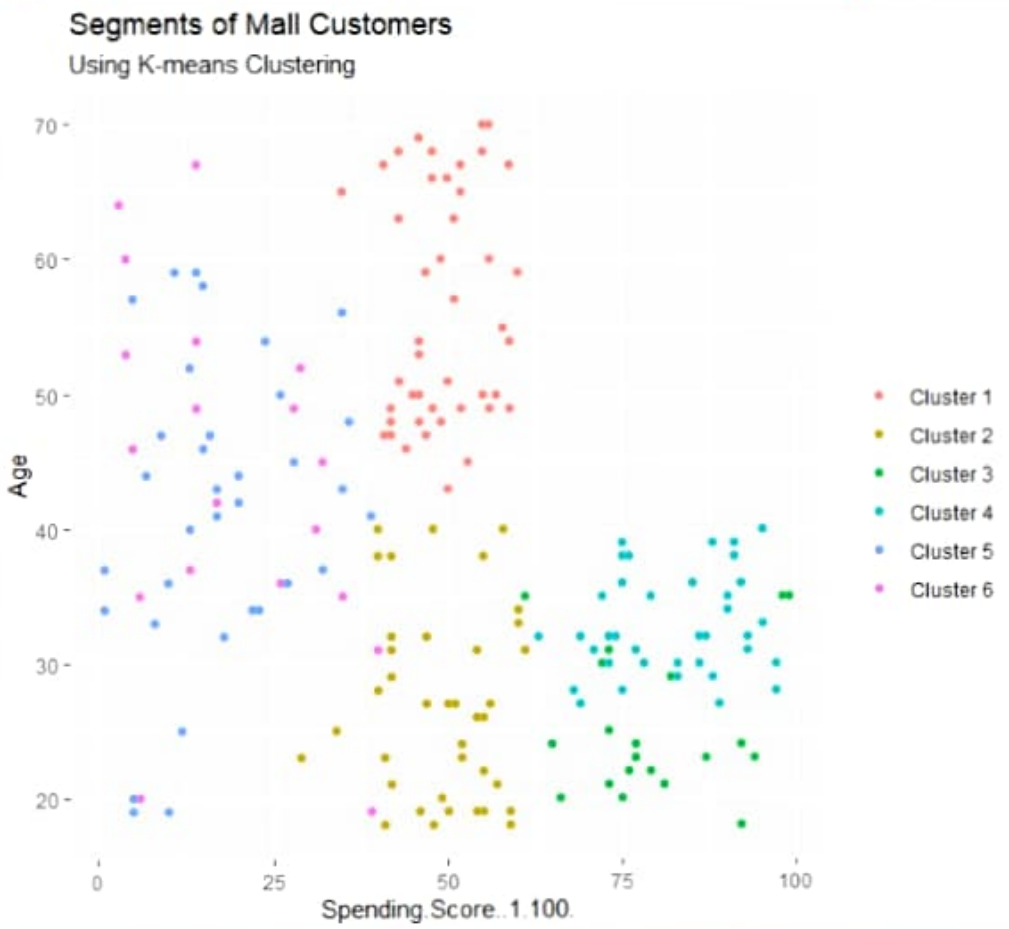
R 4.3.2 · ~/

```
+                        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_data, 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")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:17 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
172
173
174
175 kCols=function(vec){cols=rainbow (length (unique (vec)))
176 return (cols[as.numeric(as.factor(vec))])}
177
178 digCluster<-k6$cluster; dignm<-as.character(digCluster); # K-means clusters
179
180 plot(pcclust$x[,1:2], col =kCols(digCluster),pch =19,xlab ="K-means",ylab="classe
181 legend("bottomleft",unique(dignm),fill=unique(kCols(digCluster)))
182
183
184 cor_matrix <- cor(customer_data[,3:5])
185 heatmap(cor_matrix,
186         col = colorRampPalette(c("blue", "white", "red"))(50),
187         main = "Correlation Heatmap")
188
189 library(ggplot2)
190
191 ggplot(customer_data, aes(x = Age, fill = as.factor(k6$cluster))) +
192
```

182:1 (Top Level) | R Script

Console | Terminal | Background Jobs

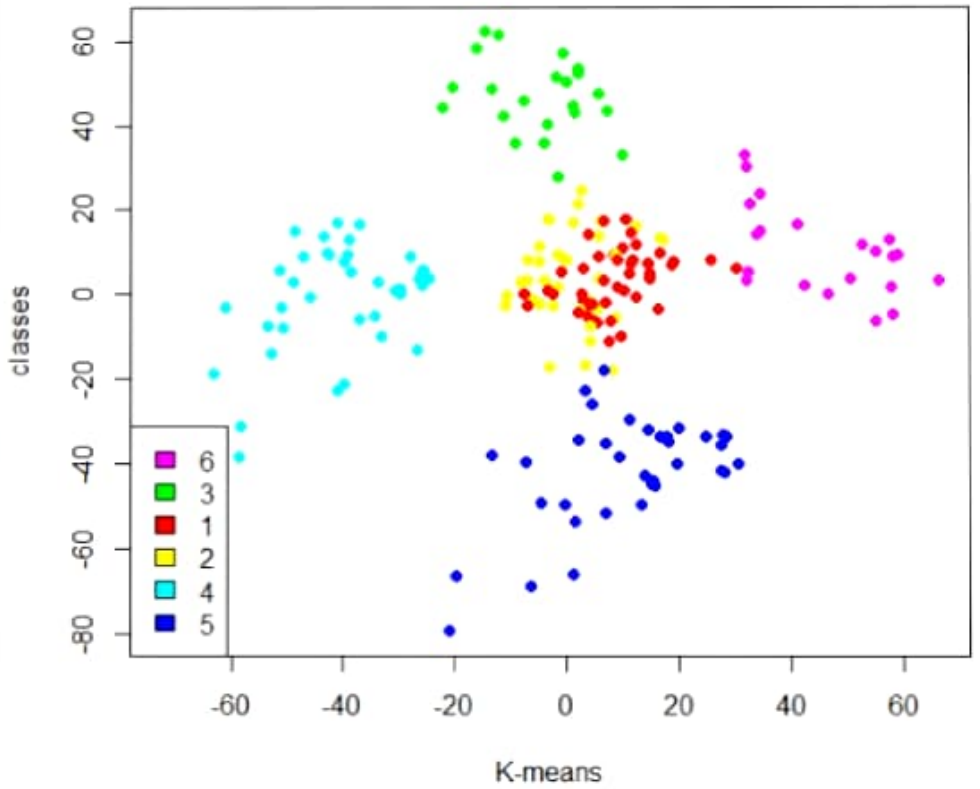
R 4.3.2 · ~/

```
+ return (cols[as.numeric(as.factor(vec))])}
>
> digCluster<-k6$cluster; dignm<-as.character(digCluster); # K-means clusters
> kCols=function(vec){cols=rainbow (length (unique (vec)))
+ return (cols[as.numeric(as.factor(vec))])}
>
> digCluster<-k6$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)))
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_AS.R | Mall_Customers

Source on Save | Run | Source

```
179
180 plot(pcclust$x[,1:2], col =kCols(digCluster),pch =19,xlab ="K-means",ylab="classe
181 legend("bottomleft",unique(dignm),fill=unique(kCols(digCluster)))
182
183
184 cor_matrix <- cor(customer_data[,3:5])
185 heatmap(cor_matrix,
186         col = colorRampPalette(c("blue", "white", "red"))(50),
187         main = "Correlation Heatmap")
188
189 library(ggplot2)
190
191 ggplot(customer_data, aes(x = Age, fill = as.factor(k6$cluster))) +
192   geom_density(alpha = 0.5) +
193   ggtitle("Density Plot of Age by Cluster")
194
195 # Repeat for Annual Income and Spending Score
196
197 cluster_means <- aggregate(. ~ k6$cluster, data = customer_data[,3:5], mean)
198
```

188:1 (Top Level) | R Script

Console | Terminal | Background Jobs

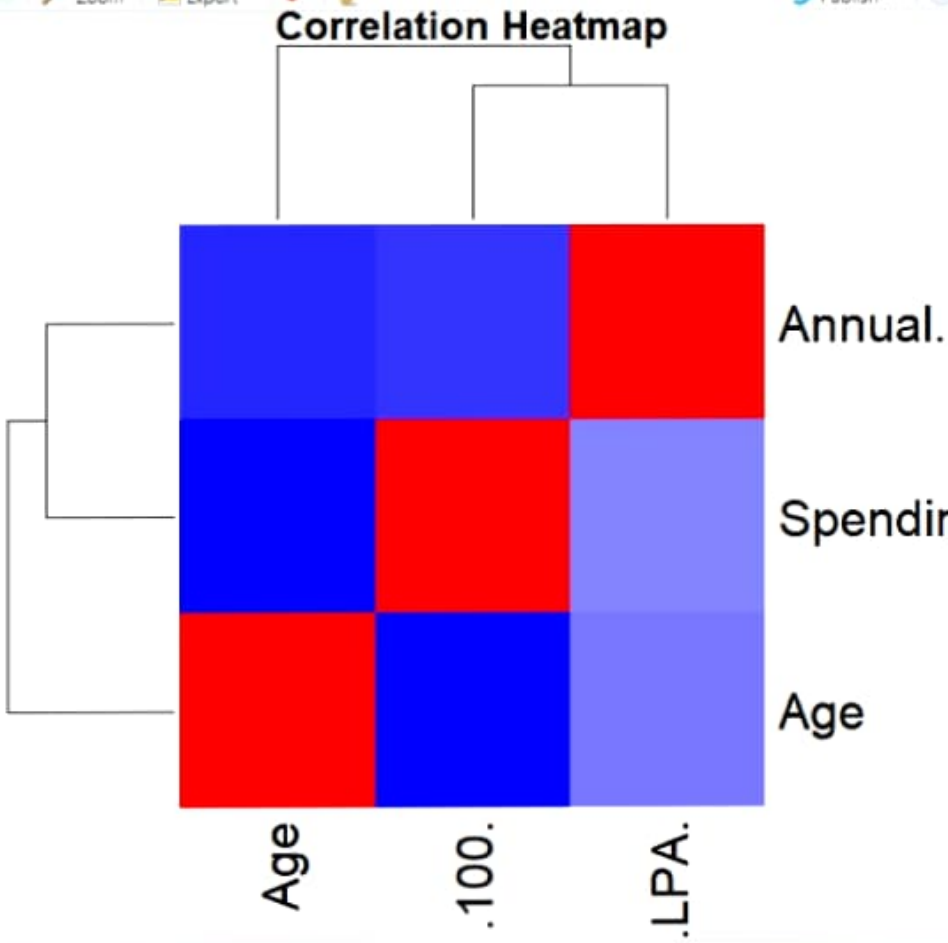
R 4.3.2 · ~/

```
+ 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)))
> cor_matrix <- cor(customer_data[,3:5])
> heatmap(cor_matrix,
+         col = colorRampPalette(c("blue", "white", "red"))(50),
+         main = "Correlation Heatmap")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:18 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
187         main = "Correlation Heatmap")
188
189 library(ggplot2)
190
191 ggplot(customer_data, aes(x = Age, fill = as.factor(k6$cluster))) +
192   geom_density(alpha = 0.5) +
193   ggtitle("Density Plot of Age by Cluster")
194
195 # Repeat for Annual Income and Spending Score
196
197 cluster_means <- aggregate(. ~ k6$cluster, data = customer_data[,3:5], mean)
198
199 library(scatterplot3d)
200
201 scatterplot3d(customer_data$Age, customer_data$Annual.Income..LPA.,
202               customer_data$Spending.Score..1.100.,
203               color = as.factor(k6$cluster),
204               main = "3D Scatter Plot by Cluster")
205
206
```

194:1 (Top Level) | R Script

Console | Terminal | Background Jobs

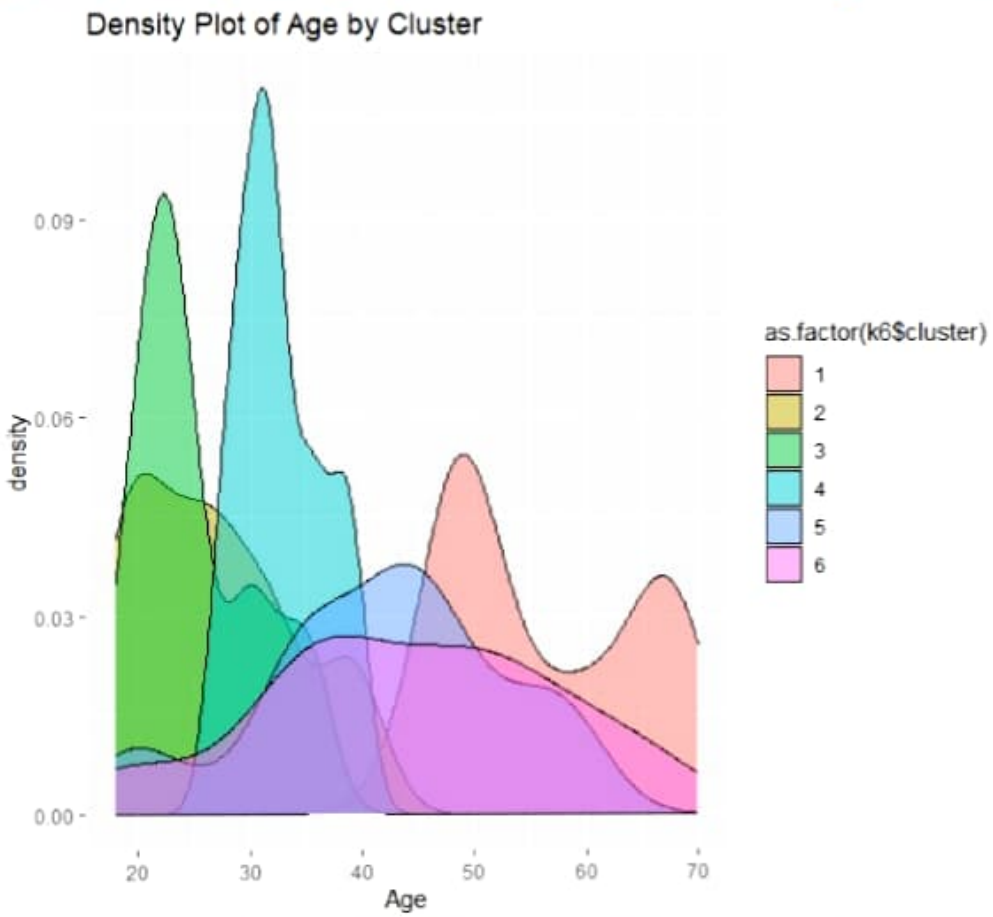
R 4.3.2 · ~/

```
> library(ggplot2)
>
> ggplot(customer_data, aes(x = Age, fill = as.factor(k6$cluster))) +
+   geom_density(alpha = 0.5) +
+   ggtitle("Density Plot of Age by Cluster")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search | ENG IN | 20:19 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
193   ggtitle("Density Plot of Age by Cluster")
194
195 # Repeat for Annual Income and Spending Score
196
197 cluster_means <- aggregate(. ~ k6$cluster, data = customer_data[,3:5], mean)
198
199 library(scatterplot3d)
200
201 scatterplot3d(customer_data$Age, customer_data$Annual.Income..LPA.,
202               customer_data$Spending.Score..1.100.,
203               color = as.factor(k6$cluster),
204               main = "3D Scatter Plot by Cluster")
205
206
207 library(wordcloud)
208
209 wordcloud(names(a), freq = a, scale=c(3,0.5), colors=brewer.pal(8, "Dark2"))
210 library(plotly)
211
212 plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1
213
```

204:51 (Top Level) | R Script

Console | Terminal | Background Jobs

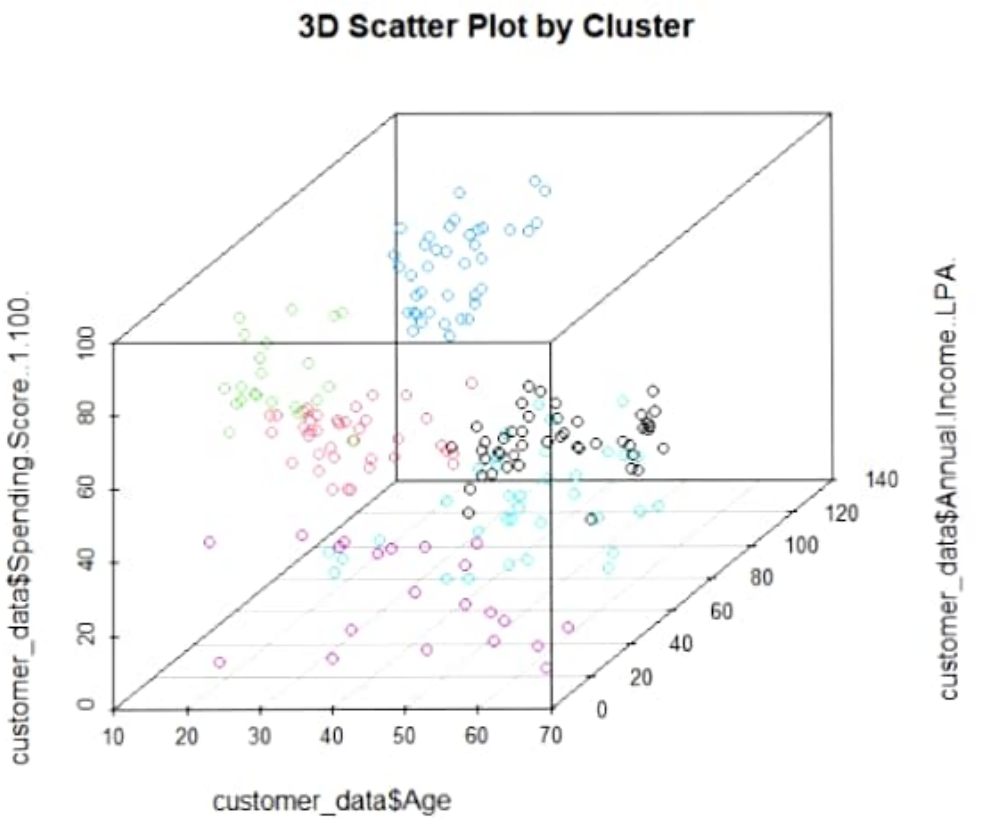
R 4.3.2 · ~/

```
> # Repeat for Annual Income and Spending Score
>
> cluster_means <- aggregate(. ~ k6$cluster, data = customer_data[,3:5], mean)
>
> library(scatterplot3d)
>
> scatterplot3d(customer_data$Age, customer_data$Annual.Income..LPA.,
+               customer_data$Spending.Score..1.100.,
+               color = as.factor(k6$cluster),
+               main = "3D Scatter Plot by Cluster")
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



Search

ENG IN

20:19 04-01-2024

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

RGLOB_A9.R | Mall_Customers

Source on Save | Run | Source

```
202                customer_data$Spending.Score..1.100.,
203                color = as.factor(k6$cluster),
204                main = "3D Scatter Plot by Cluster")
205
206
207  library(wordcloud)
208
209  wordcloud(names(a), freq = a, scale=c(3,0.5), colors=brewer.pal(8, "Dark2"))
210  library(plotly)
211
212  plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1
213          color = as.factor(k6$cluster), type = "scatter3d", mode = "markers")
214
215
216
217
218
219
220
221
```

209:77 (Top Level) R Script

Console | Terminal | Background Jobs

R 4.3.2 · ~/

```
> library(scatterplot3d)
>
> scatterplot3d(customer_data$Age, customer_data$Annual.Income..LPA.,
+               customer_data$Spending.Score..1.100.,
+               color = as.factor(k6$cluster),
+               main = "3D Scatter Plot by Cluster")
> library(wordcloud)
>
> wordcloud(names(a), freq = a, scale=c(3,0.5), colors=brewer.pal(8, "Dark2"))
>
```

Environment | History | Connections | Tutorial

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish

Search

ENG IN

20:20 04-01-2024

```
                color = as.factor(k6$cluster),
                main = "3D Scatter Plot by Cluster")

library(wordcloud)

wordcloud(names(a), freq = a, scale=c(3,0.5), colors=brewer.pal(8, "Dark2"))
library(plotly)

plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1
        color = as.factor(k6$cluster), type = "scatter3d", mode = "markers")
```

```
> library(plotly)
> plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1.100.,
```