

```
1 customer_data=read.csv("C:\\Users\\manne\\Downloads\\Mail_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 x Background Jobs x
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
>
```

RGL0B_A9.R Mail_Customers

Source on Save

Run

```
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

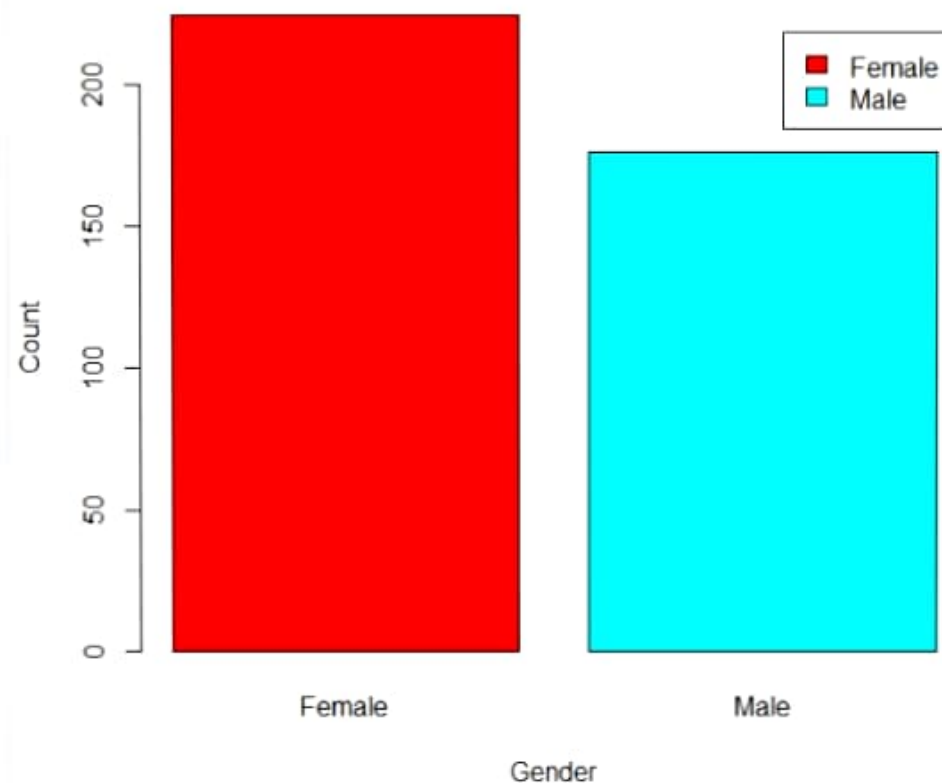
Console Terminal Background Jobs

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

Using BarPlot to display Gender Comparision



Search

ENG
IN

20:09

04-01-2024

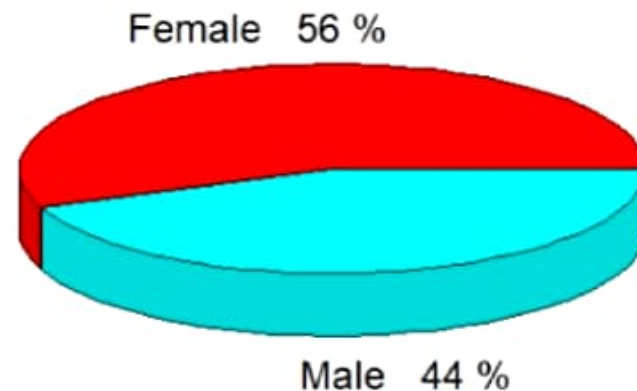
```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
RGLob_A9.R Mail_Customers
Source on Save Run Source
20 xlab="Gender",
21 col=rainbow(2),
22 legend=rownames(a))
23
24
25 pct=round(a/sum(a)*100)
26 lbs=paste(c("Female","Male")," ",pct,"%",sep=" ")
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

Pie Chart Depicting Ratio of Female and Male



```
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
```

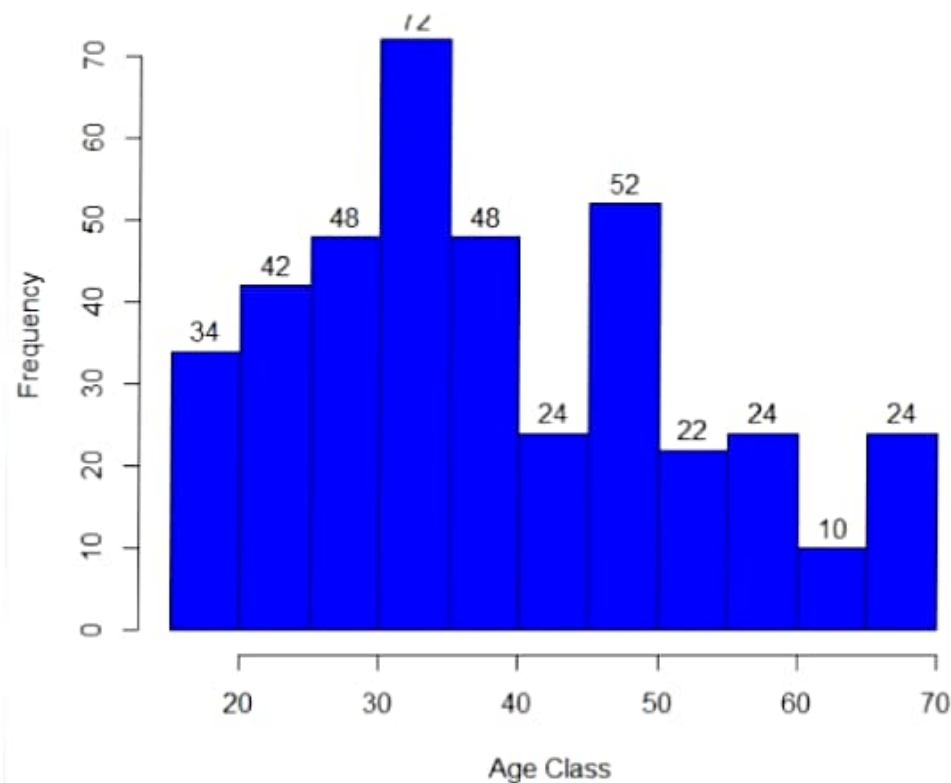
42:18 (Top Level)

R Script

```
> 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)
```

Histogram to Show Count of Age Class

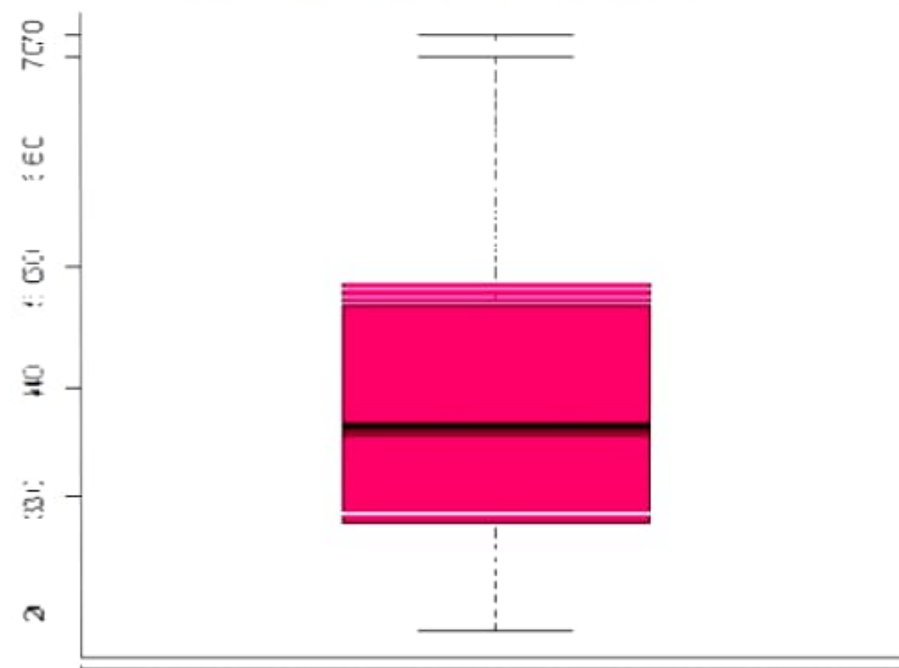


Search

ENG
IN20:10
04-01-2024

Boxplot for Descriptive Analysis of Age

Boxplot for Descriptive Analysis of Age



```
46 main="Boxplot for Descriptive Analysis of Age")
```

```
47 #Analysis of the Annual Income of the Customers
```

```
48 summary(customer_data$Annual.Income..LPA.)
```

```
49 hist(customer_data$Annual.Income..LPA.,
```

```
50 col="#660033",
```

```
51 main="Histogram for Annual Income",
```

```
52 xlab="Annual Income Class",
```

```
53 ylab="Frequency",
```

```
54 labels=TRUE)
```

```
55 plot(density(customer_data$Annual.Income..LPA.),
```

```
56 col="blue",
```

```
57 main="Density Plot for Annual Income",
```

```
58 xlab="Annual Income Class",
```

```
59 ylab="Density")
```

```
60 polygon(density(customer_data$Annual.Income..LPA.),
```

```
61 col="#ccff66")
```

```
62
```

```
63
```

```
64
```

```
65
```

```
66
```

```
67
```

```
68
```

```
69
```

```
70
```

```
71
```

```
72
```

```
73
```

```
74
```

```
75
```

```
76
```

```
77
```

```
78
```

```
79
```

```
80
```

```
81
```

```
82
```

```
83
```

```
84
```

```
85
```

```
86
```

```
87
```

```
88
```

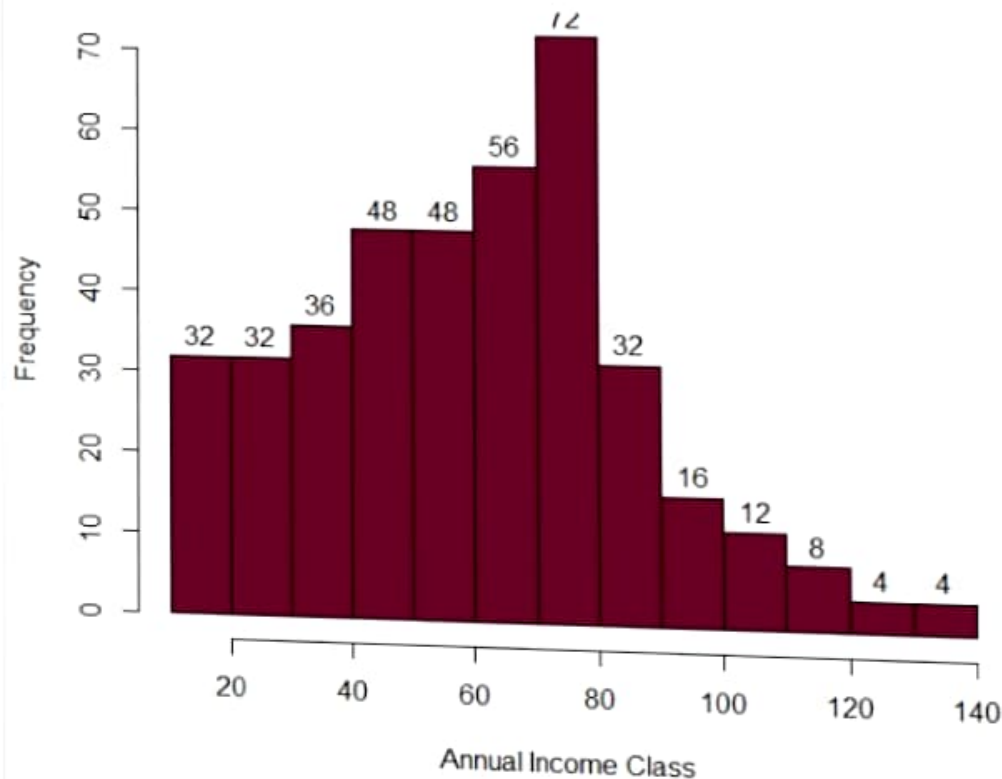
```
89
```

```
90
```

```
91
```

```
92
```

Histogram for Annual Income



```
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      )
```

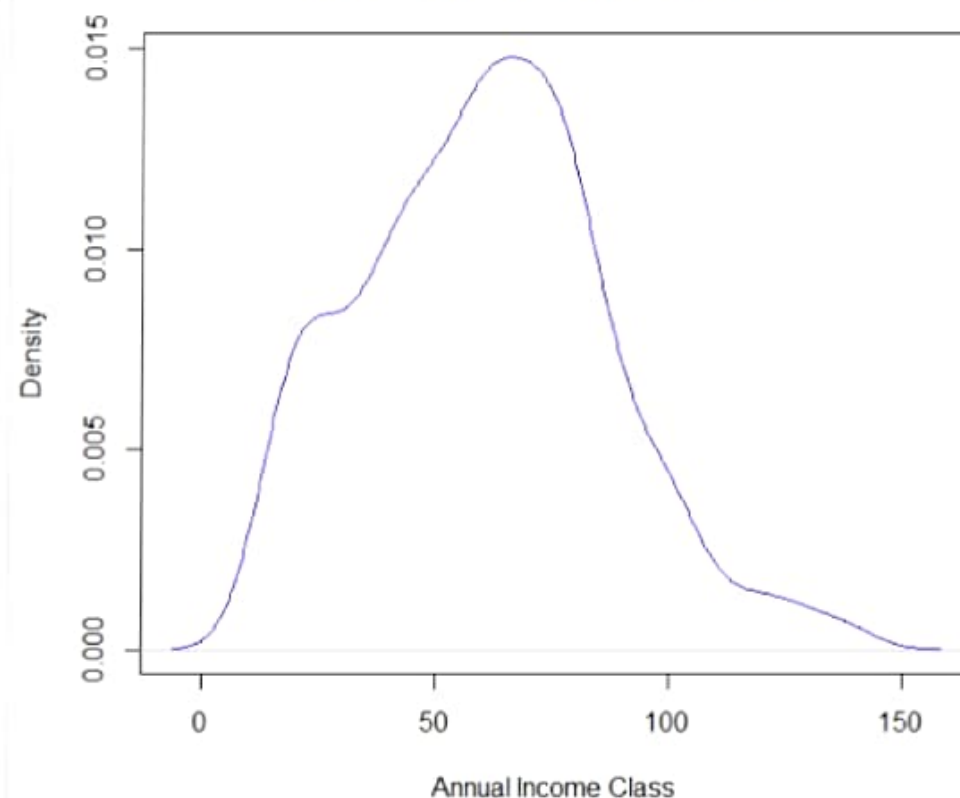
62:21 (Top Level)

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")
>
```

Density Plot for Annual Income



Search

ENG
IN20:11
04-01-2024

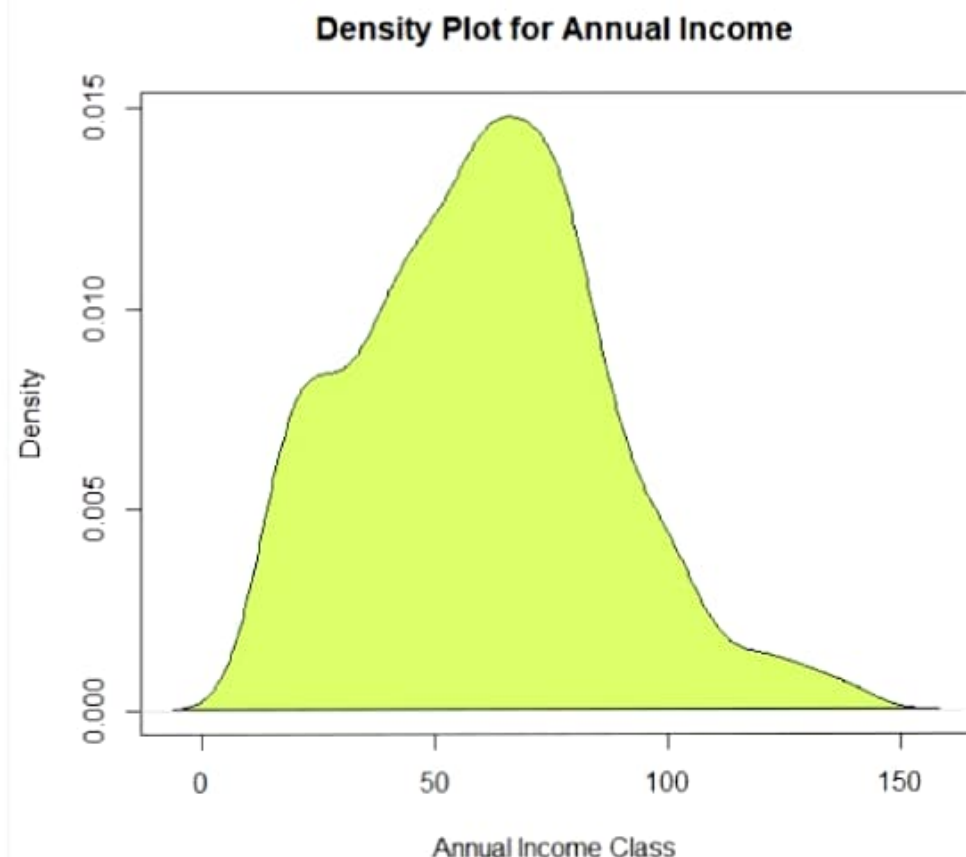
```
RGLOB_A9.R x Mail_Customers x
Source on Save | Run | Source
59 col="blue",
60 main="Density Plot for Annual Income",
61 xlab="Annual Income Class",
62 ylab="Density")
63 polygon(density(customer_data$Annual.Income..LPA.),
64 col="#ccff66")
65
66 boxplot(customer_data$Spending.Score..1.100.,
67 horizontal=TRUE,
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 labels=TRUE)
77
78 #k-means Algorithm
79
64:23 (Top Level) : R Script :
```

```
Console Terminal Background Jobs x
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




```
RGLOB_A9.R x Mail_Customers x  
Source on Save  
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       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
```

69:67 (Top Level)

R Script

Console Terminal Background Jobs

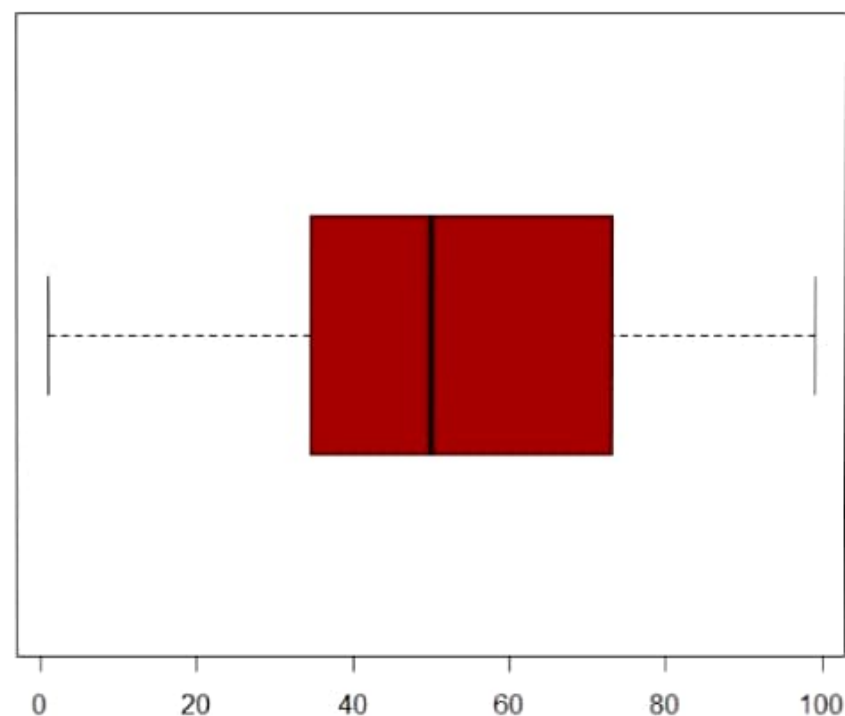
```
R 4.3.2 ~/  
> boxplot(customer_data$Spending.Score..1.100.,  
+         horizontal=TRUE,  
+         col="#990000",  
+         main="BoxPlot for Descriptive Analysis of spending score")  
>
```

Environment History Connections Tutorial

Files Plots Packages Help Viewer Presentation

Zoom Export Publish

BoxPlot for Descriptive Analysis of Spending Score



Search

ENG
IN20:12
04-01-2024

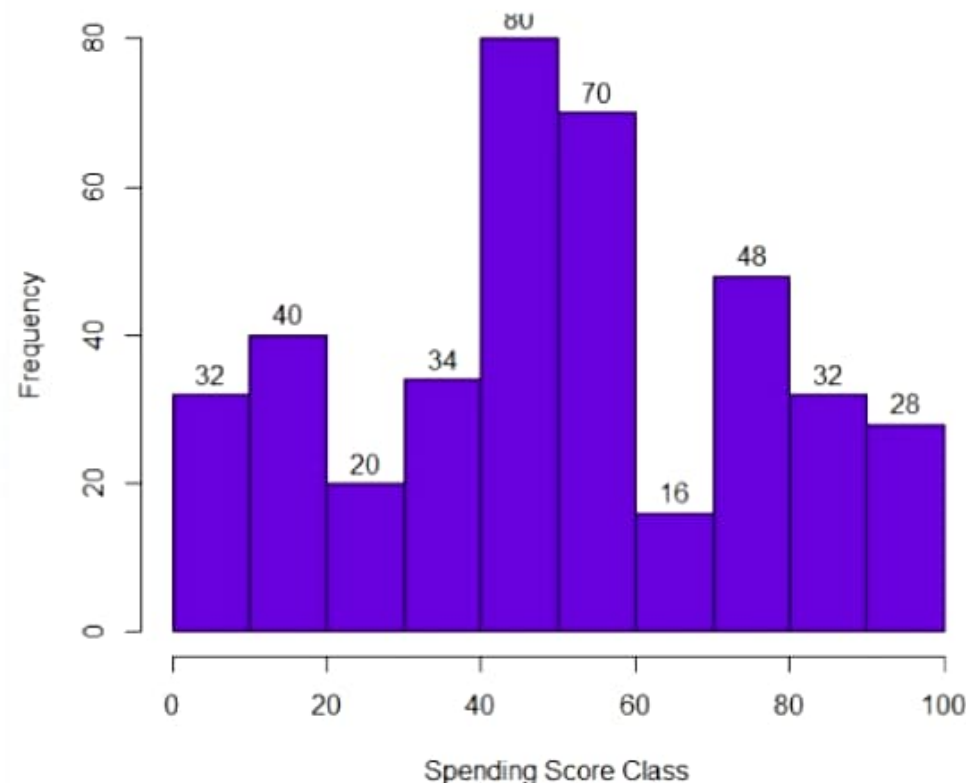
```
RGLOB_A9.R x Mail_Customers x  
Source on Save Run Source  
68 col="#990000",  
69 main="BoxPlot for Descriptive Analysis of Spending Score")  
70  
71 hist(customer_data$Spending.Score..1.100.,  
72       main="Histogram for Spending Score",  
73       xlab="Spending Score Class",  
74       ylab="Frequency",  
75       col="#6600cc",  
76       labels=TRUE)  
77  
78 #K-means Algorithm  
79  
80 library(purrr)  
81 set.seed(123)  
82 # function to calculate total intra-cluster sum of square  
83 iss <- function(k) {  
84   kmeans(customer_data[,3:5],k,iter.max=100,nstart=100,algorithm="Lloyd")$tot.w  
85 }  
86  
87 k_values <- 1:10  
88
```

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)  
>
```

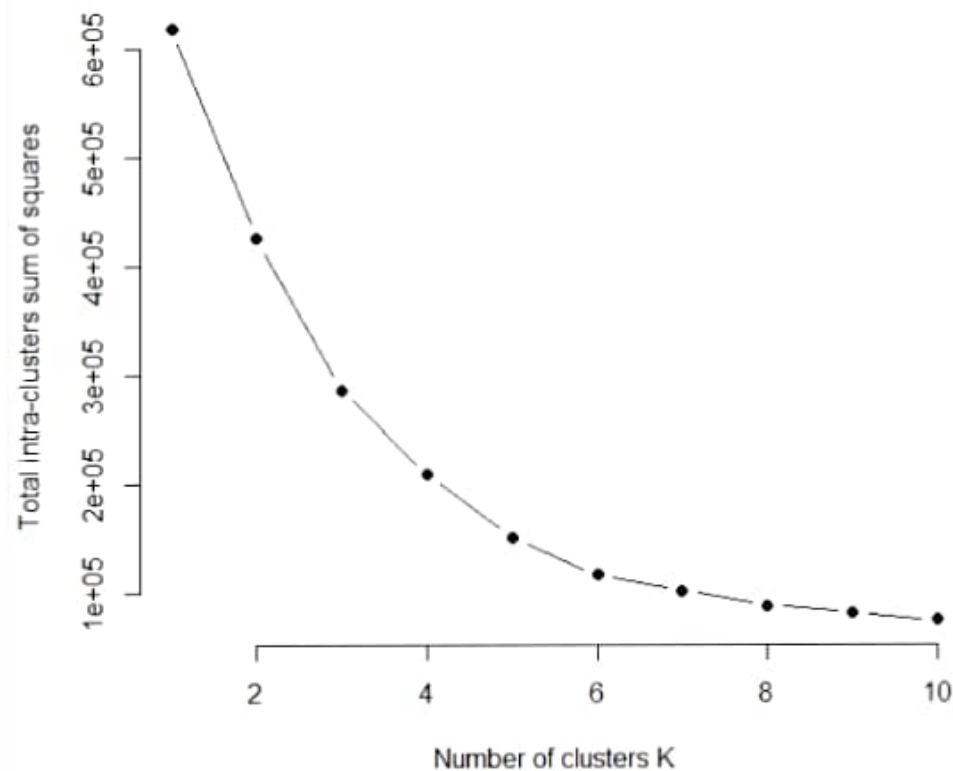
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Zoom Export Publish

HistoGram for Spending Score



```
RGLOB_A9.R x Mail_Customers x  
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(ggplot2)  
102  
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")  
> |
```



```
customer_seg_rlab.R* customer_data*
library(factoextra)
k2<-kmeans(customer_data[,3:5],2,iter.max=100,nstart=50,algorithm="Lloyd")
s2<-plot(silhouette(k2$cluster,dist(customer_data[,3:5],"euclidean")))
k3<-kmeans(customer_data[,3:5],3,iter.max=100,nstart=50,algorithm="Lloyd")
s3<-plot(silhouette(k3$cluster,dist(customer_data[,3:5],"euclidean")))
k4<-kmeans(customer_data[,3:5],4,iter.max=100,nstart=50,algorithm="Lloyd")
s4<-plot(silhouette(k4$cluster,dist(customer_data[,3:5],"euclidean")))
k5<-kmeans(customer_data[,3:5],5,iter.max=100,nstart=50,algorithm="Lloyd")
s5<-plot(silhouette(k5$cluster,dist(customer_data[,3:5],"euclidean")))
k6<-kmeans(customer_data[,3:5],6,iter.max=100,nstart=50,algorithm="Lloyd")
s6<-plot(silhouette(k6$cluster,dist(customer_data[,3:5],"euclidean")))
k7<-kmeans(customer_data[,3:5],7,iter.max=100,nstart=50,algorithm="Lloyd")
s7<-plot(silhouette(k7$cluster,dist(customer_data[,3:5],"euclidean")))
k8<-kmeans(customer_data[,3:5],8,iter.max=100,nstart=50,algorithm="Lloyd")
s8<-plot(silhouette(k8$cluster,dist(customer_data[,3:5],"euclidean")))
k9<-kmeans(customer_data[,3:5],9,iter.max=100,nstart=50,algorithm="Lloyd")
s9<-plot(silhouette(k9$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"))
```

```
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"))
>
```

Silhouette plot of (x = k5\$cluster, dist = dist(customer_data[, 3:5], "euclidean"))

n = 400

5 clusters C_j $j : n_j \mid \text{ave}_{i \in C_j} s_i$

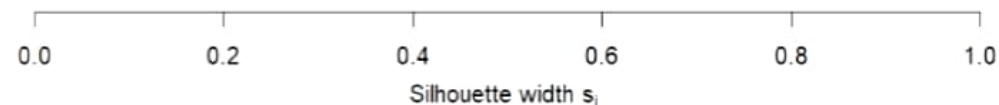
1 : 46 | 0.43

2 : 78 | 0.54

3 : 46 | 0.61

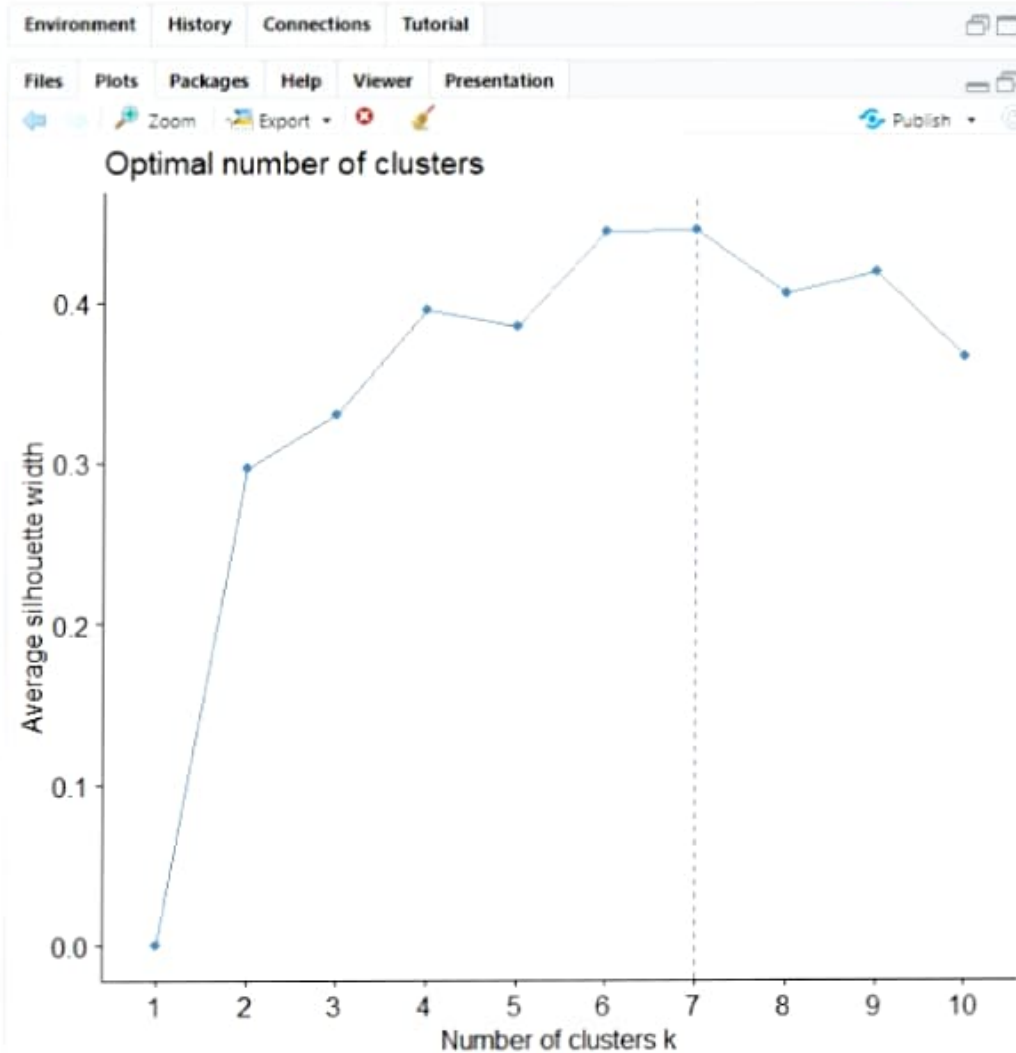
4 : 72 | 0.43

5 : 158 | 0.38

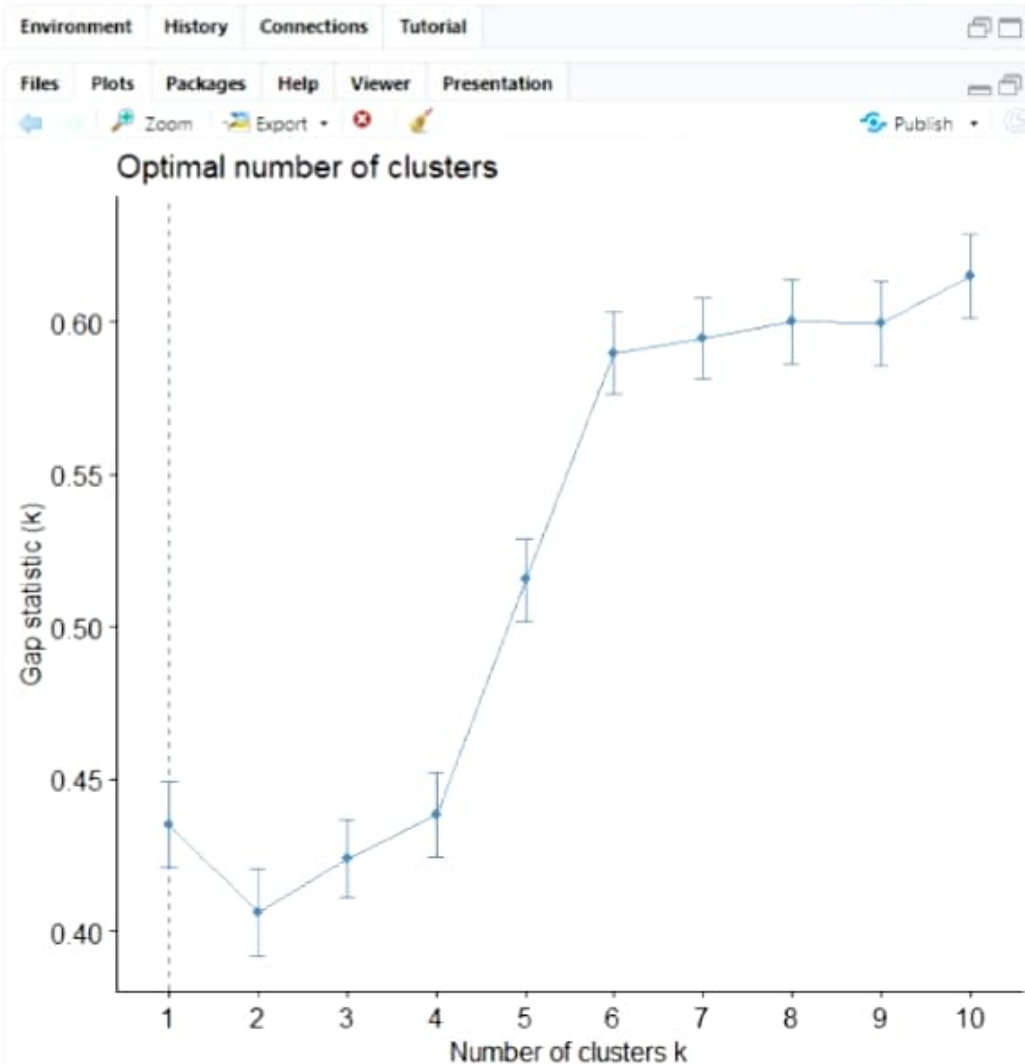


Average silhouette width : 0.45

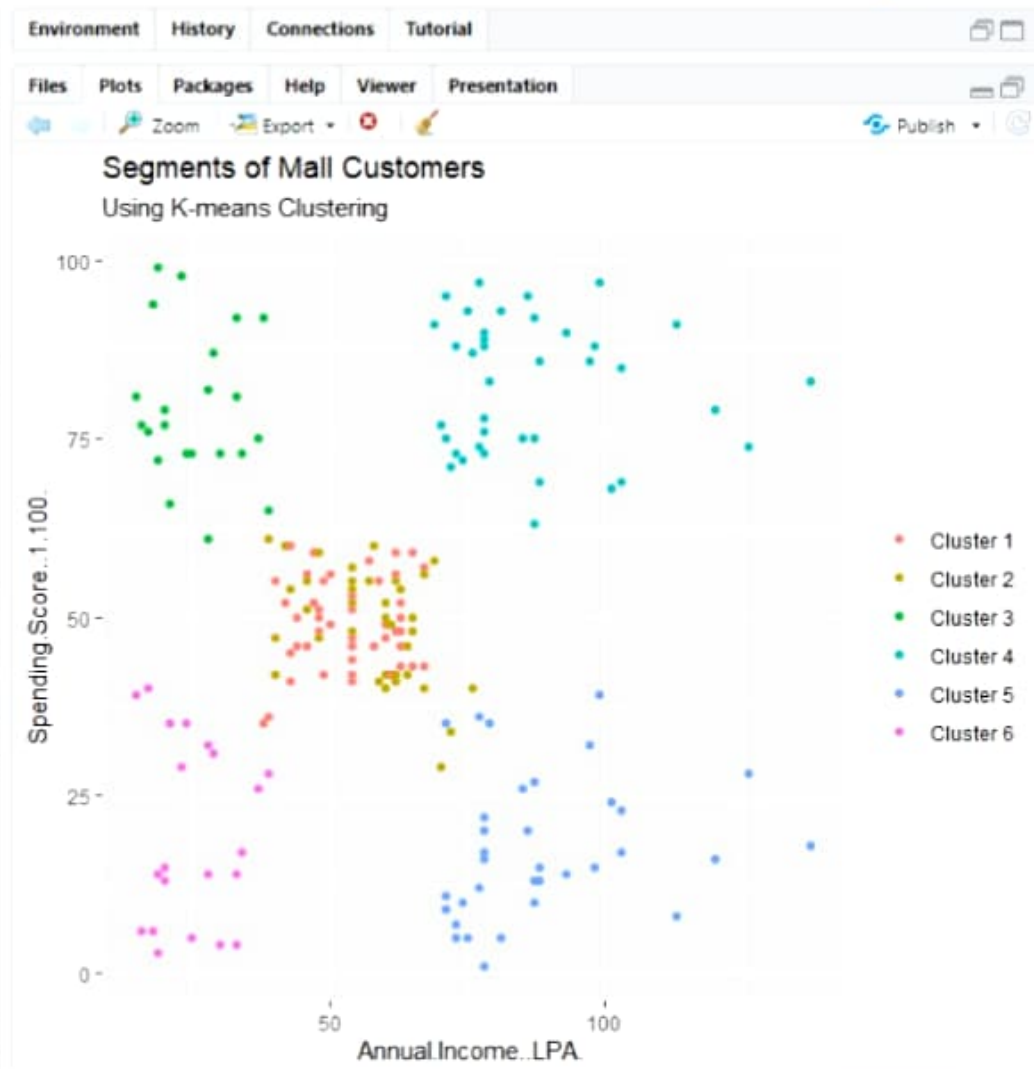

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
RGLOB_A9.R Mail_Customers
Source on Save Run Source
129 s10<-plot(silhouette(k10$cluster,dist(customer_data[,3:5],"euclidean")))
130
131
132
133 library(NbClust)
134 library(factoextra)
135
136 fviz_nbclust(customer_data[,3:5], 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")
>
```



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
• • • • • Go to file/function • Addins •
RGLob_A9.R * Mail_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]:
.....
```



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
RGLOB_A9.R Mail_Customers
Source on Save Run Source
147 #Visualizing the Clustering Results using the First Two Principle Components
148
149 pcclust=prcomp(customer_data[,3:5],scale=FALSE) #principal component analysis
150 summary(pcclust)
151
152 pcclust$rotation[,1:2]
153
154
155
156 set.seed(1)
157 ggplot(customer_data, aes(x =Annual.Income..LPA., y = Spending.Score..1.100.)) +
158   geom_point(stat = "identity", aes(color = as.factor(k6$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       "cluster 5", "cluster 6")) +
163   ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")
164
165
166 qqplot(customer_data, aes(x =Spending.Score..1.100., y =Aqe)) +
167   +
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```




```
157 ggplot(customer_data, aes(x = Annual.Income..LPA., y = Spending.Score..1.100.)) +
158   geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
159   scale_color_discrete(name = "",
160     breaks = c("1", "2", "3", "4", "5", "6"),
161     labels = c("cluster 1", "cluster 2", "cluster 3", "cluster 4",
162       "cluster 5", "cluster 6"))
162 ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")
163
164
165
166 ggplot(customer_data, aes(x = Spending.Score..1.100., y = Age)) +
167   geom_point(stat = "identity", aes(color = as.factor(k6$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       "cluster 5", "cluster 6"))
171 ggtitle("Segments of Mall Customers", subtitle = "Using K-means Clustering")
172
173
174
175 kcols = function(vec) { cols = rainbow(length(unique(vec)))
176   }
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
```

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")
>
```

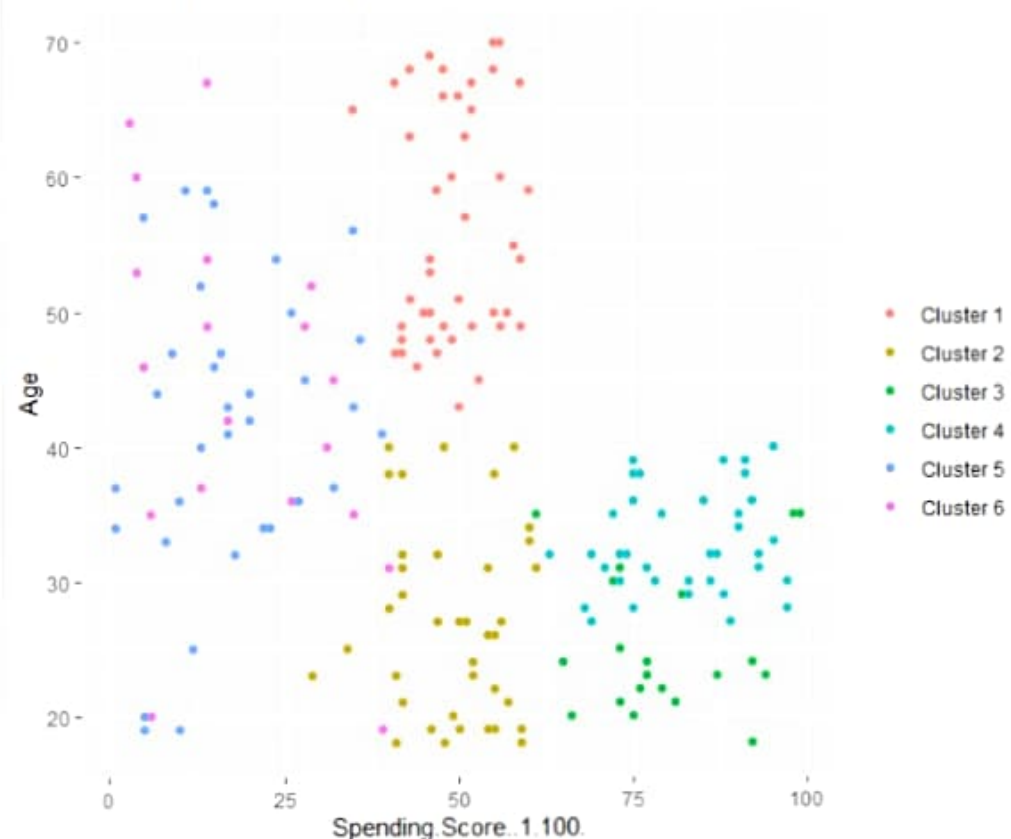
Environment History Connections Tutorial

Files Plots Packages Help Viewer Presentation

Zoom Export Publish

Segments of Mall Customers

Using K-means Clustering




```

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="classes")
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)))
>

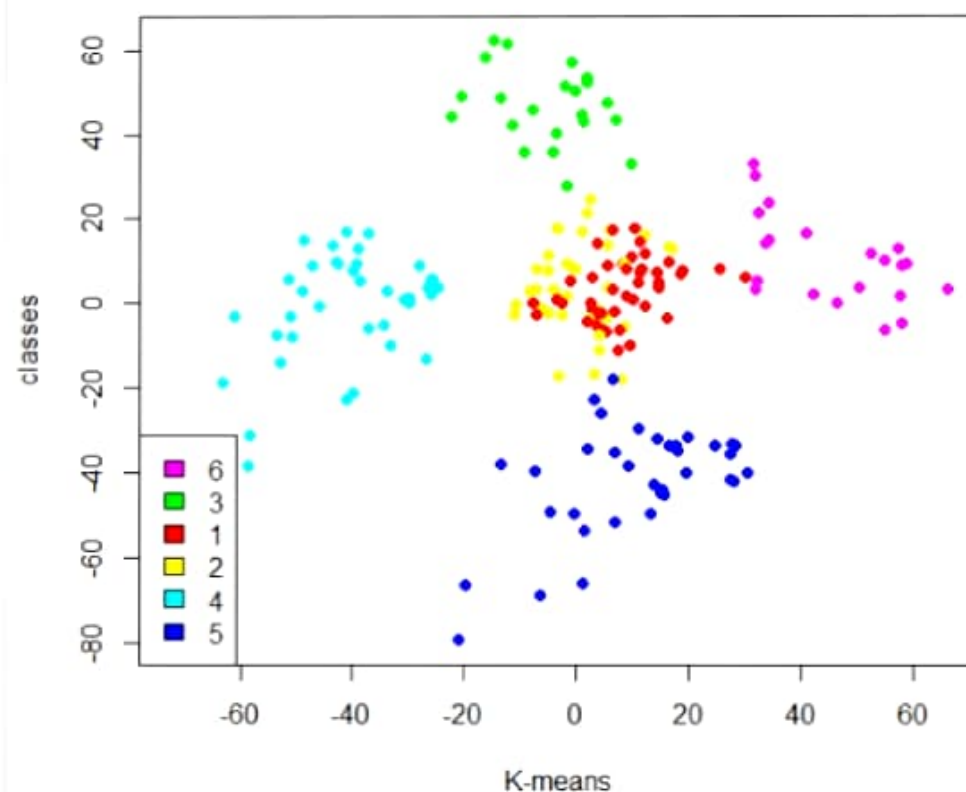
```

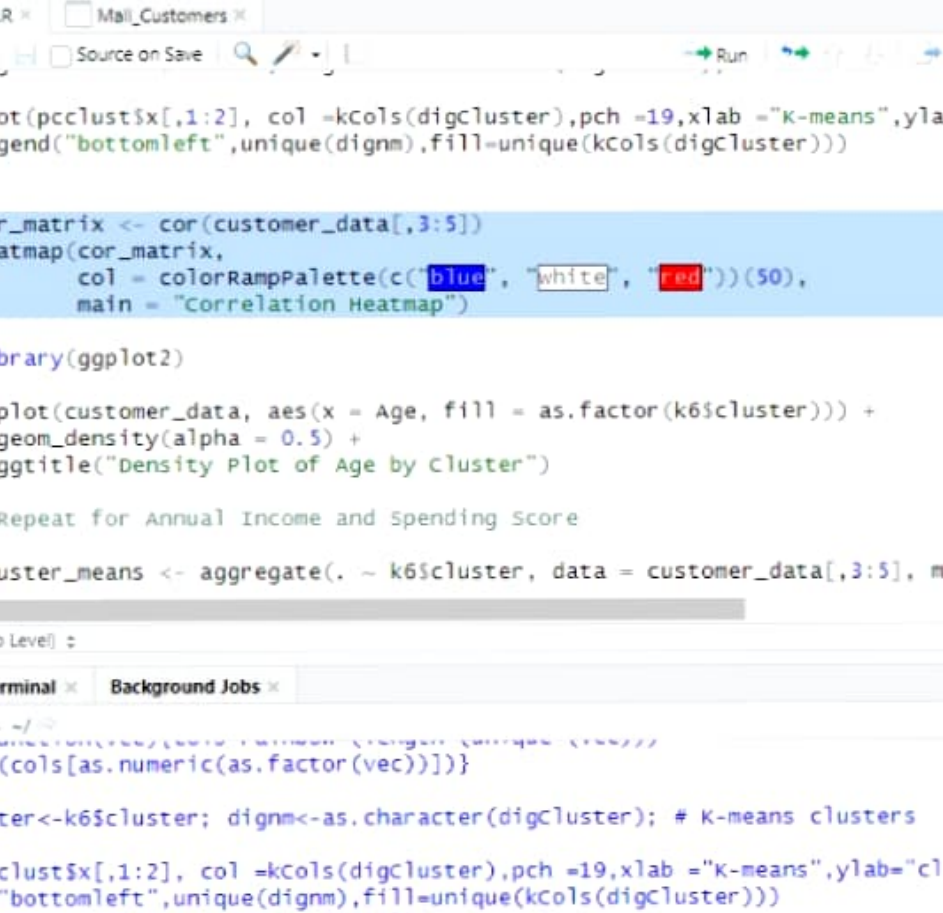
Environment History Connections Tutorial

Files Plots Packages Help Viewer Presentation

Zoom Export

Publish



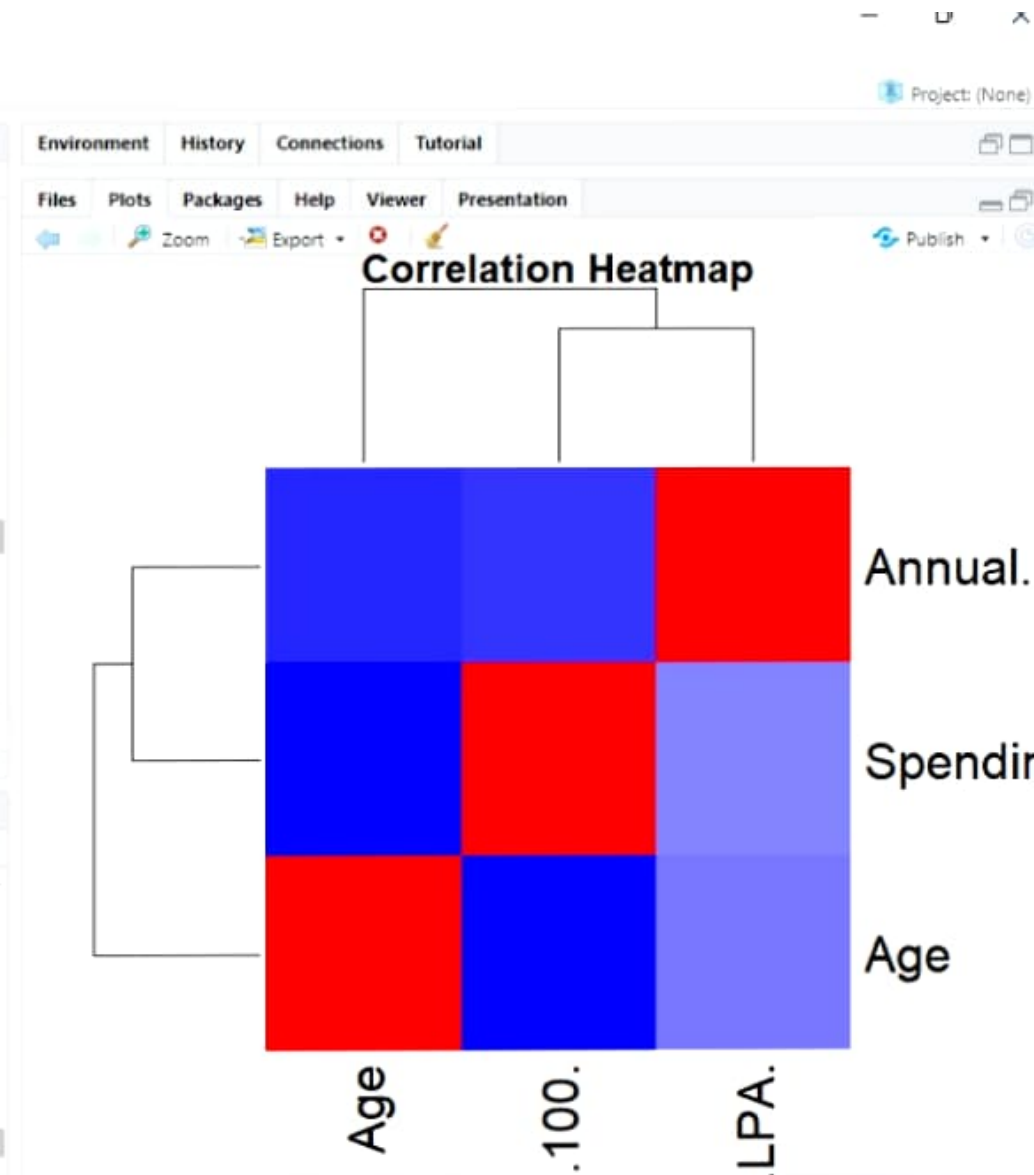


The screenshot shows the RStudio IDE with the following components:

- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations, a search bar with "Go to file/function", and buttons for "Run" and "Source".
- Script Editor:** Contains R code for K-means clustering and visualization. The code is as follows:


```
179
180 plot(pclust$x[,1:2], col = kcols(digcluster), pch = 19, xlab = "K-means", ylab = "classes")
181 legend("bottomleft", unique(dignm), fill = unique(kcols(digcluster)))
182
183
184 cor_matrix <- cor(customer_data[,3:5])
185 heatmap(cor_matrix,
186         col = colorRampPalette(c("blue", "white", "red"))(50),
187         main = "Correlation Heatmap")
188
189 library(ggplot2)
190
191 ggplot(customer_data, aes(x = Age, fill = as.factor(k6$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)
```
- Console:** Shows the execution of the code, including the output of the K-means clustering and the creation of the density plot. The console text is:


```
R 4.3.2 ~/>
+ return (cols[as.numeric(as.factor(vec))])
+
+ digcluster <- k6$cluster; dignm <- as.character(digcluster); # K-means clusters
+
+ plot(pclust$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")
+ 
```



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
RGLOB_A9.R x Mail_Customers x
Source on Save
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
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

```

```

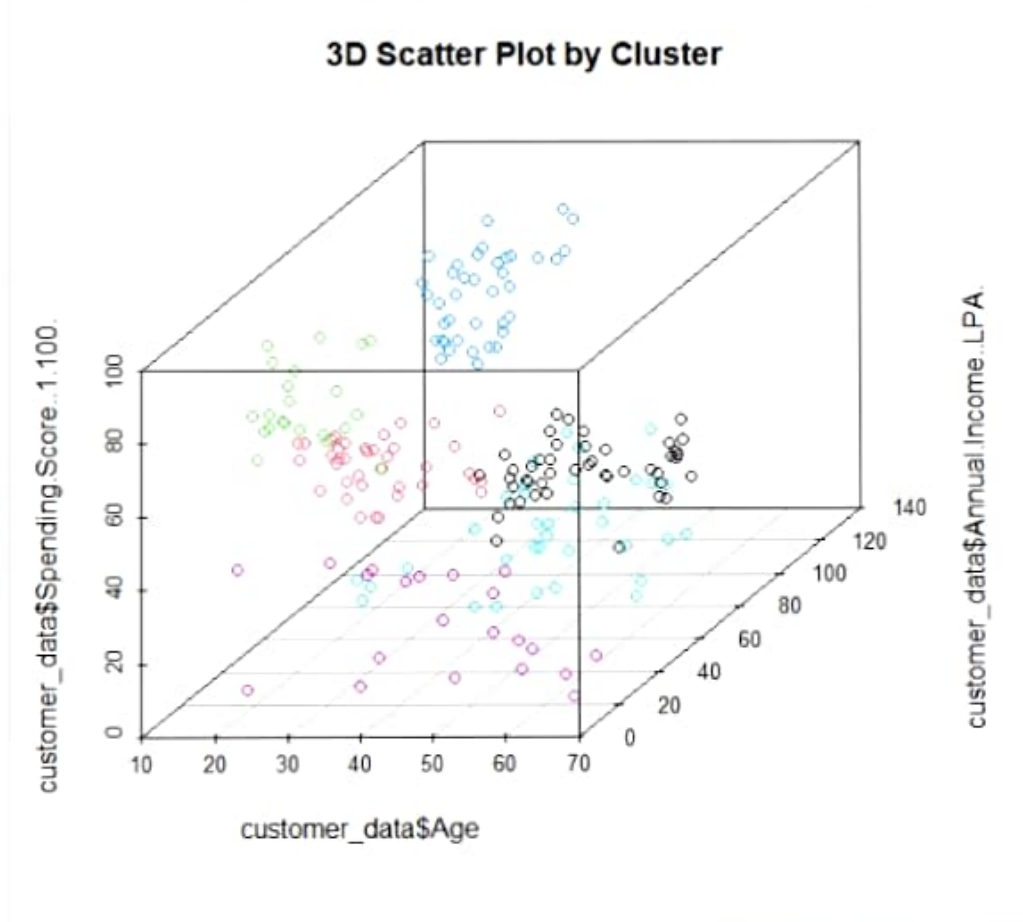
194:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.2 ~ /
> library(ggplot2)
>
> ggplot(customer_data, aes(x = Age, fill = as.factor(k6$cluster))) +
+   geom_density(alpha = 0.5) +
+   ggtitle("Density Plot of Age by Cluster")
>

```




```
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..1.100.)
```

```
R 4.3.2 ~\
> cluster_means <- aggregate(. ~ k6$cluster, data = customer_data[,3:5], mean)
> # Repeat for Annual Income and Spending Score
>
> cluster_means <- aggregate(. ~ k6$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")
>
```




```
202 scatterplot3d(customer_data$Age, customer_data$Annual.Income..LPA.,
203               customer_data$Spending.Score..1.100.,
204               color = as.factor(k6$cluster),
205               main = "3D Scatter Plot by cluster")
206
207 library(wordcloud)
208 wordcloud(names(a), freq = a, scale=c(3,0.5), colors=brewer.pal(8, "Dark2"))
209 library(plotly)
210
211 plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1
212         color = as.factor(k6$cluster), type = "scatter3d", mode = "markers")
213
214
215
216
217
218
219
220
221
```

209:77 (Top Level)

R Script

```
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"))  
>
```

Male
Female



Search

ENG
IN20:20
04-01-2024

```
File Edit View Plots Session Build Debug Runfile Tools Help
[Icons] [Go to file/function] [Addins]

RGL08_A9K Mail_Customers
[Icons] [Full] [Zoom] [Jupyter]

202 #-----#-----#-----#-----#-----#-----#-----#-----#-----#
203 color = as.factor(k6$cluster),
204 main = "3D scatter plot by cluster")
205
206
207 library(wordcloud)
208
209 wordcloud(names(a), freq = a, scale=c(3,0.5), colors=brewer.pal(8, "Dark2"))
210 library(plotly)
211
212 plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1.100.,
213         color = as.factor(k6$cluster), type = "scatter3d", mode = "markers")
214
215
216
217
218
219
220
221
214:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.3.2 ~ /
> library(plotly)
> plot_ly(customer_data, x = ~Age, y = ~Annual.Income..LPA., z = ~Spending.Score..1.100.,
:         color = as.factor(k6$cluster), type = "scatter3d", mode = "markers")
>
```

