Tugas Besar Data Mining

PRAKTIKUM

Disusun oleh:

Shaliha Putri Ninda	**3311901001**
Rokki Samuel T	**3311901015**
Bryant Axell Hang	**3311901020**

Disusun untuk memenuhi salah satu syarat kelulusan
matakuliah IF317 Data Mining

PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNIK INFORMATIKA

POLITEKNIK NEGERI BATAM

BATAM

2021

HALAMAN PENGESAHAN

Tugas Besar Data Mining

Disusun oleh:

Shaliha Putri Ninda	**3311901001**
Rokki Samuel T	**3311901015**
Bryant Axell Hang	**3311901020**

Batam, 14 Januari 2021

Disetujui dan disahkan oleh:

Dosen pengajar,

Muhammad Nashrullah, SST., M.Sc

HALAMAN PERNYATAAN

Dengan ini, kami:

1. NIM: 3311901001, Nama: Shaliha Putri Ninda
2. NIM: 3311901015, Nama: Rokki Samuel T
3. NIM: 3311901020, Nama: Bryant Axell Hang

adalah mahasiswa Program Studi Teknik Informatika Politeknik Negeri Batam menyatakan bahwa Tugas Besar disusun dengan:

1. Tidak melakukan plagiat terhadap naskah karya orang lain
2. Tidak melakukan pemalsuan data
3. Tidak menggunakan karya orang lain tanpa menyebut sumber asli atau tanpa ijin pemilik

Jika kemudian terbukti terjadi pelanggaran terhadap pernyataan di atas, maka kami bersedia menerima sanksi apapun termasuk pencabutan gelar akademik.

Lembar pernyataan ini juga memberikan hak kepada Politeknik Negeri Batam untuk mempergunakan, mendistribusikan ataupun memproduksi ulang seluruh hasil Tugas Praktikum ini.

Batam, 14 Januari 2021

Anggota 1,	**Anggota 2,**	**Anggota 3,**
Shaliha Putri Ninda	**Rokki Samuel T**	**Bryant Axell Hang**

DAFTAR ISI

KATA PENGANTAR

Puji dan syukur Tim Penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat-Nya yang telah dilimpahkan sehingga dapat menyelesaikan laporan praktikum yang merupakan salah satu Tugas Besar terstruktur Data Mining.

Dalam laporan praktikum ini kami mempraktikkan mengenai Dataset yang sudah diberikan intruksi menggunakan aplikasi R Studio.

Dalam menyelesaikan laporan ini, kami telah banyak mendapat bantuan dan masukan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini kami ingin menyampaikan terima kasih kepada :

1. Bapak Muhammad Nashrullah, SST., M.Sc selaku dosen Data Mining yang telah memberikan tugas mengenai laporan ini sehingga pengetahuan kami dalam penulisan laporan praktikum makin bertambah dan hal itu sangat bermanfaat bagi penyusunan laporan kami di kemudian hari.
2. Pihak-pihak yang tidak dapat kami sebutkan satu persatu yang telah turut membantu sehingga laporan ini dapat terselesaikan dengan baik dalam tepat waktu.

Kami menyadari bahwa penyusunan laporan ini masih jauh dari kesempurnaan, namun demikian telah memberikan manfaat bagi kami. Akhir kata kami berharap laporan ini dapat bermanfaat bagi kita semua. Kritik dan saran yang bersifat membangun akan kami terima dengan senang hati.

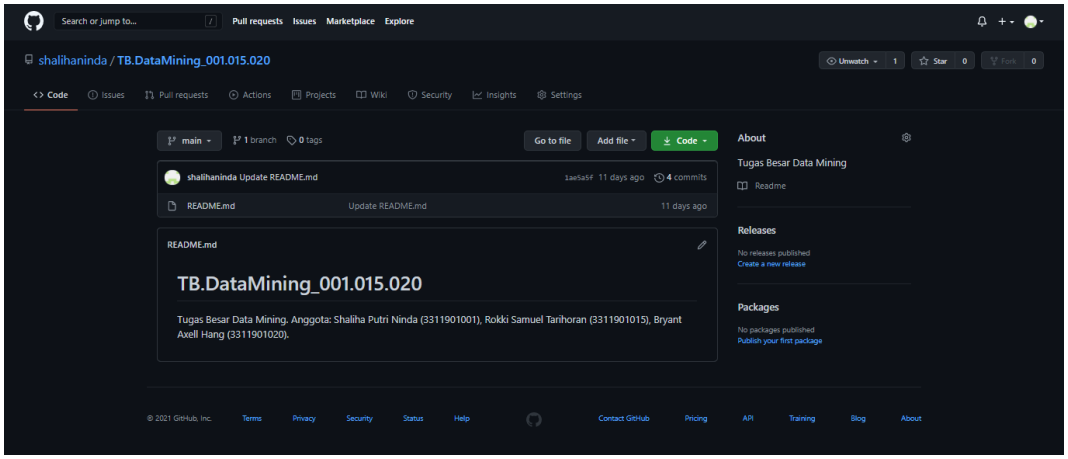
Batam, 14 Januari 2021

Tim Penulis

GITHUB

Repository

https://github.com/shalihaninda/TB.DataMining_001.015.020

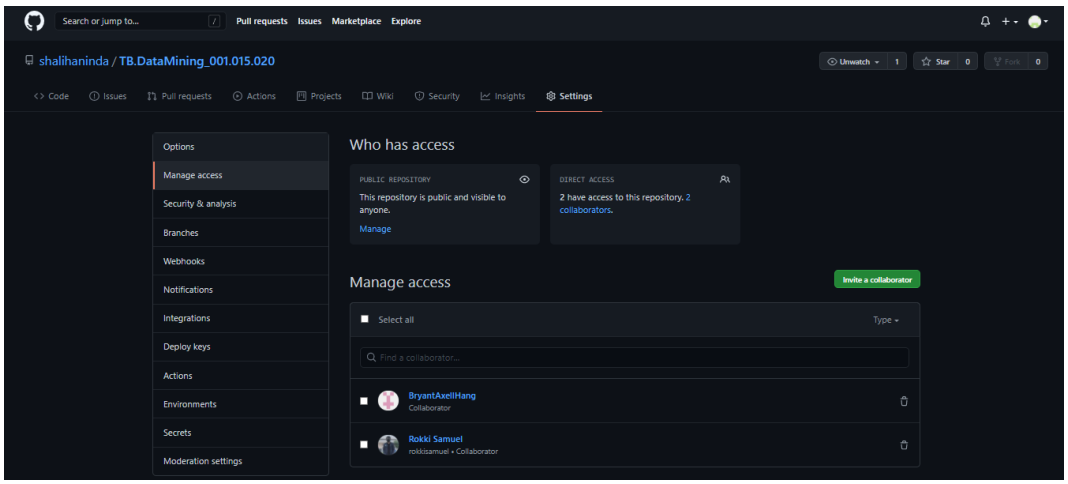


Anggota

https://github.com/shalihaninda

https://github.com/rokkisamuel

https://github.com/BryantAxellHang

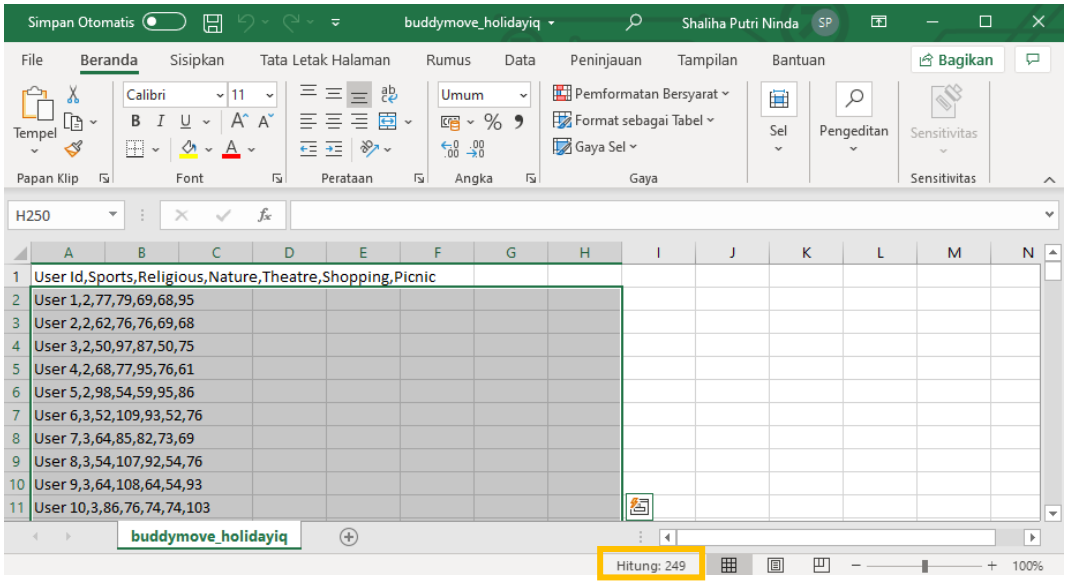


DOKUMEN

Dataset

Dataset dengan nama “BuddyMove Dataset”, berisi 249 data.

Link: https://archive.ics.uci.edu/ml/datasets/BuddyMove+Data+Set



Dataset ini memiliki 8 kolom, yaitu:

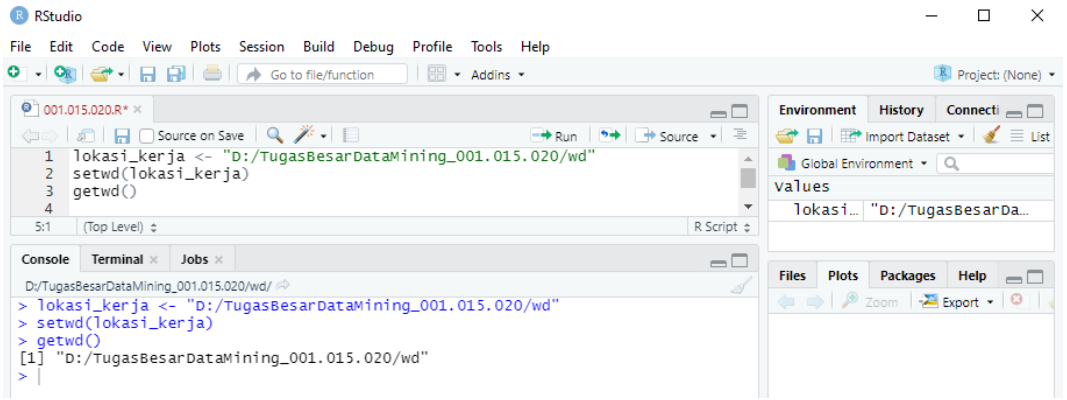
1. User ID : Id user / seseorang yang mendatangi suatu tempat / kunjungan liburan
2. Sports : Tempat olahraga
3. Religious : Tempat keagamaan
4. Nature : Lokasi pemandangan alam
5. Theatre : Tempat pertunjukan
6. Shopping : Tempat belanja
7. Picnic : Tempat piknik

Full Code

A. K-Means

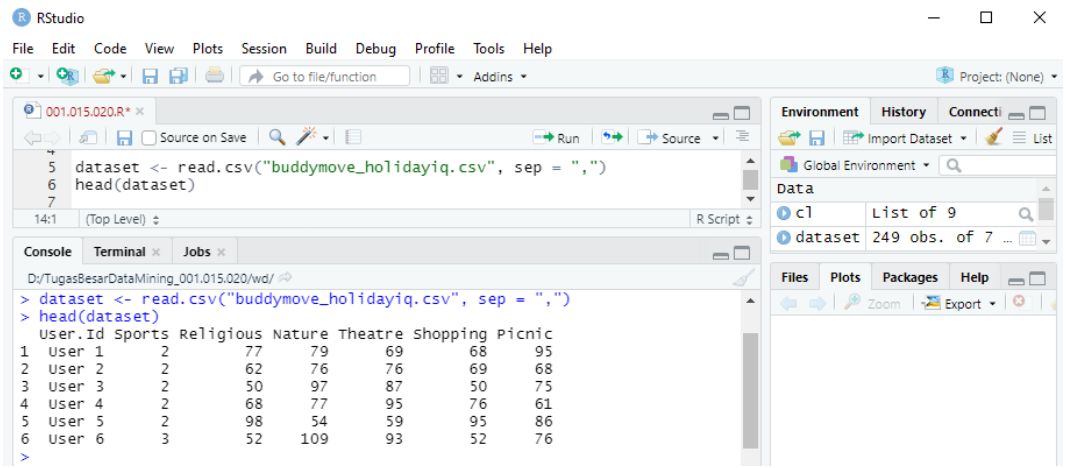
1. Pengaturan lokasi directory

```
lokasi_kerja <- "D:/SHALIHA.001/3A/DATA
MINING/Tugas Besar/wd"
```



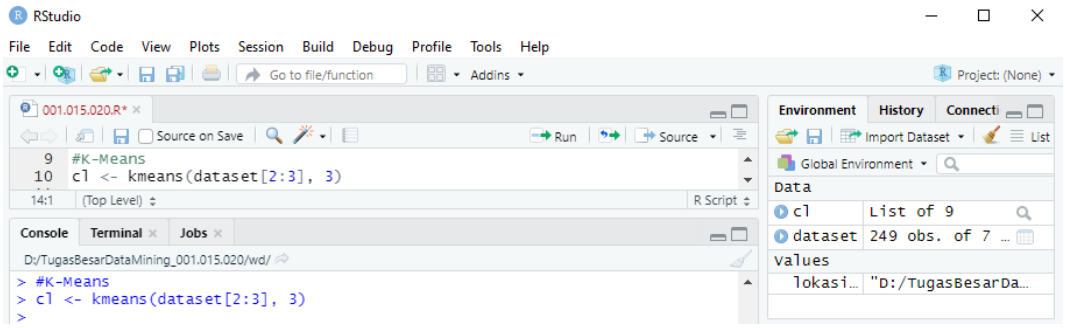
2. Membaca data dengan format csv dan dibedakan berdasarkan cell

```
dataset <- read.csv("data.csv", sep = ";")
```



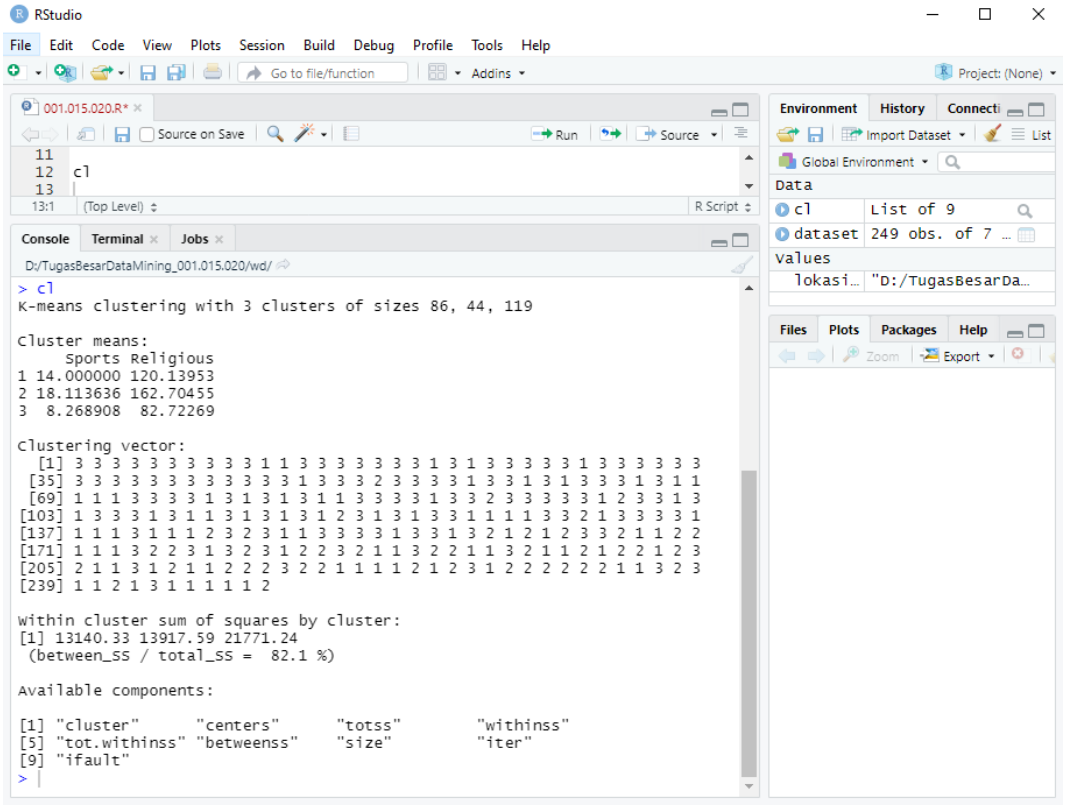
3. Clustering menggunakan kmeans dengan 3 buah titik cluster

```
cl <- kmeans(dataset, 3)
```



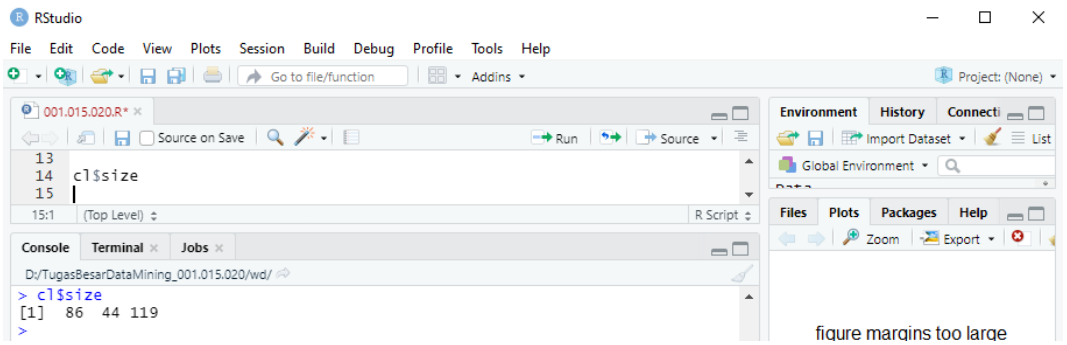
4. Jalankan perintah

```
Cl
```



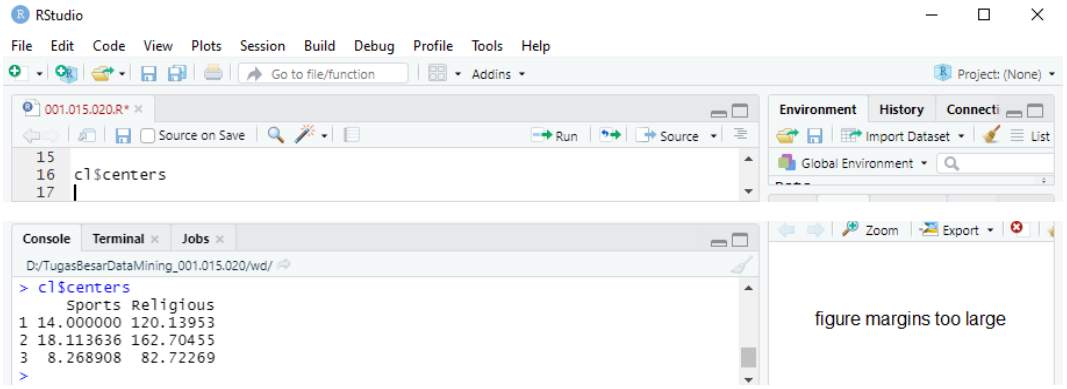
Untuk mengakses nilai pada model kmeans, dapat dijalankan perintah dengan contoh

```
cl$size
```



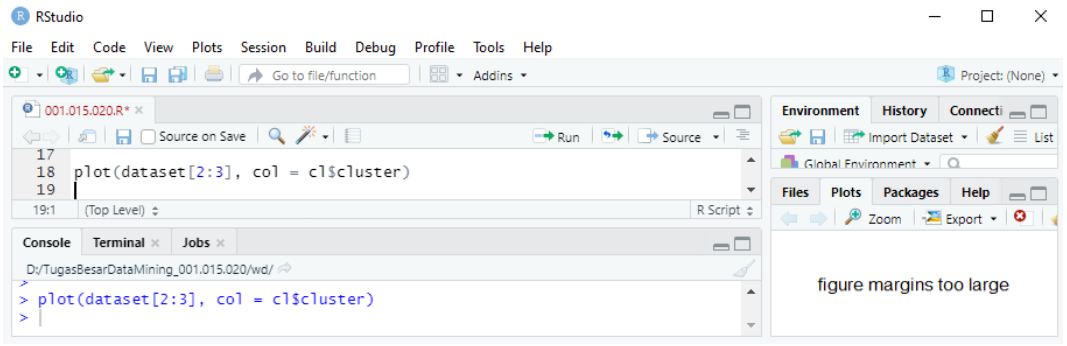
Menampilkan titik pusat cluster

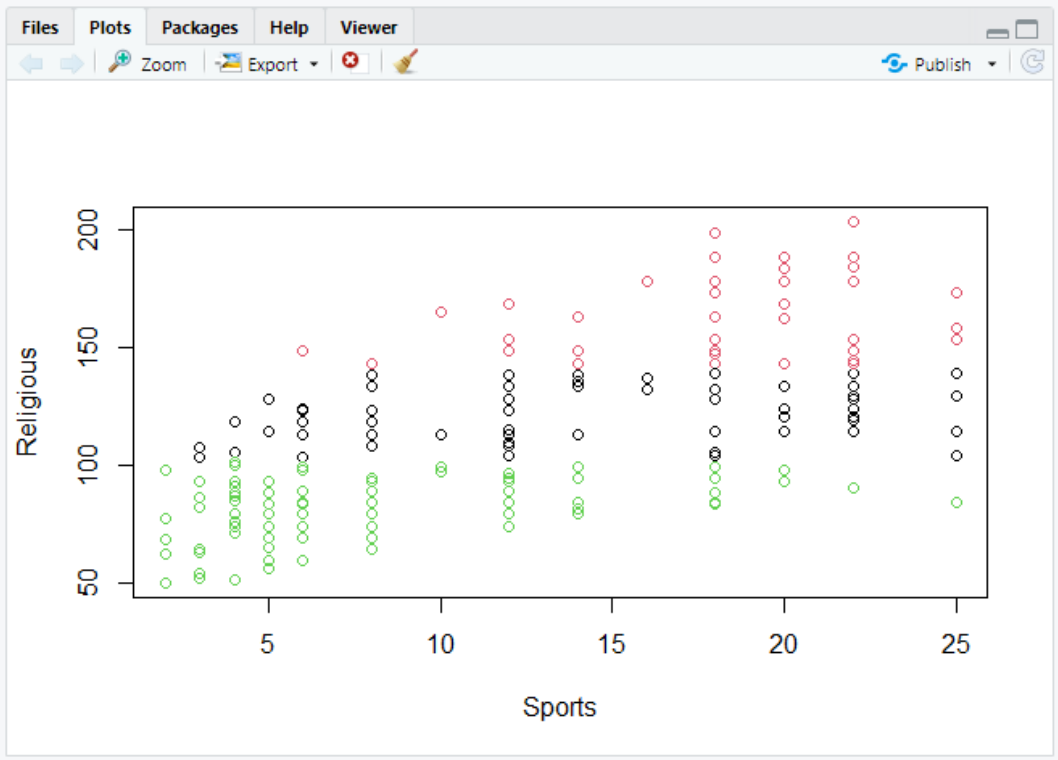
```
cl$centers
```



5. Menampilkan objek dalam bentuk gambar

```
plot(dataset, col = cl$cluster)
```



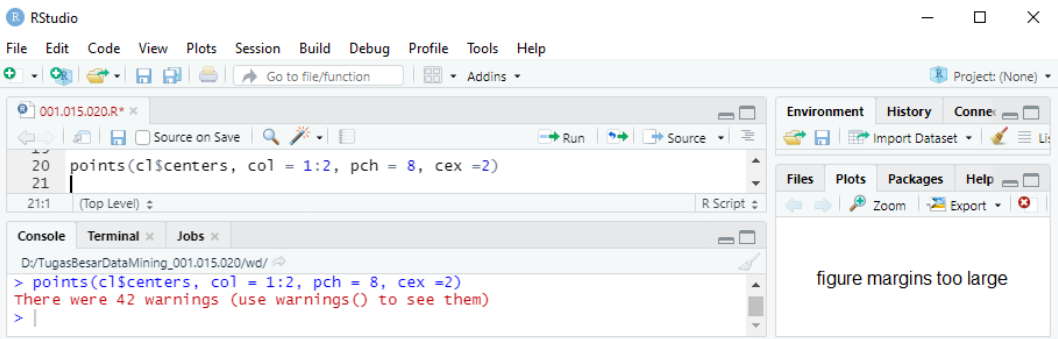


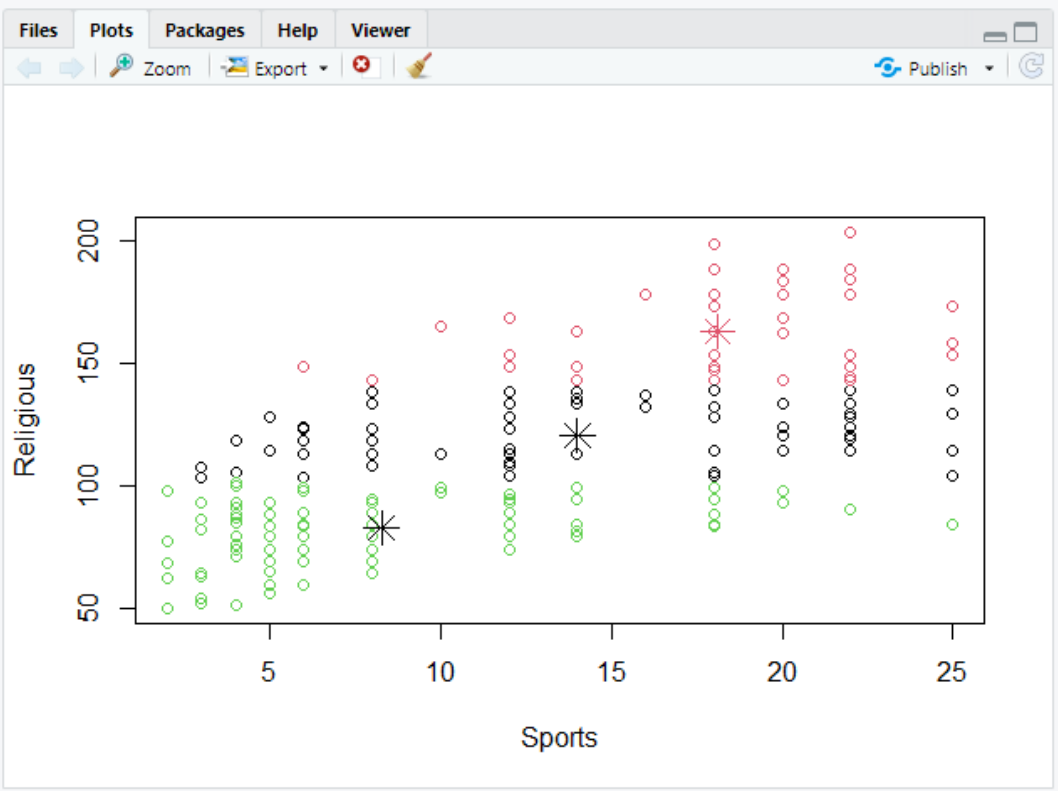
Data objek tersebut dibedakan dengan warna berdasarkan nilai cluster

`(col = cl$cluster)`

6. Pada gambar diatas, ditambahkan point baru

`points(cl$centers, col = 1:2, pch = 8, cex =2)`





Data yang digunakan pada gambar diatas adalah titik cluster

Atribut **pch** bernilai integer, menggambarkan symbol yang diinginkan, contoh

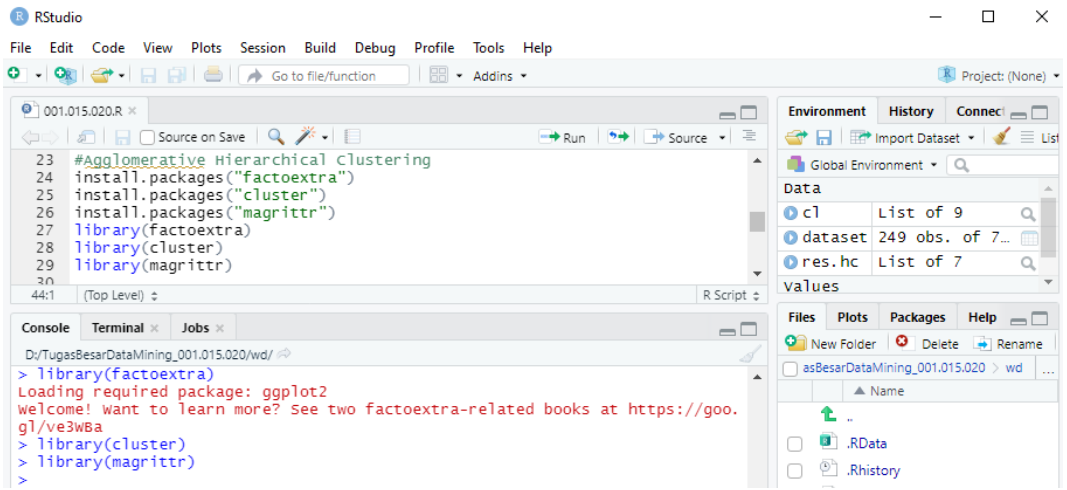
pch=8

Atribut **(cex = 2)** menentukan ukuran dari symbol tersebut

B. Agglomerative Hierarchical Clustering

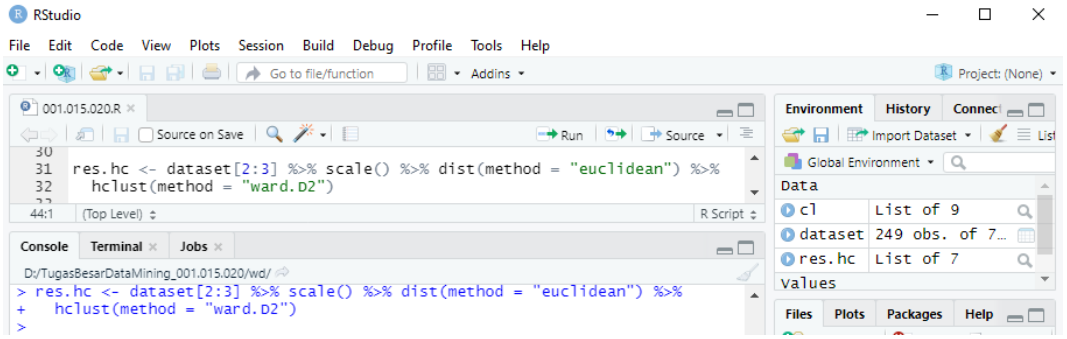
1. Instalasi dan penggunaan library

```
#install.packages("factoextra")
#install.packages("cluster")
#install.packages("magrittr")
library("cluster")
library("factoextra")
library("magrittr")
```



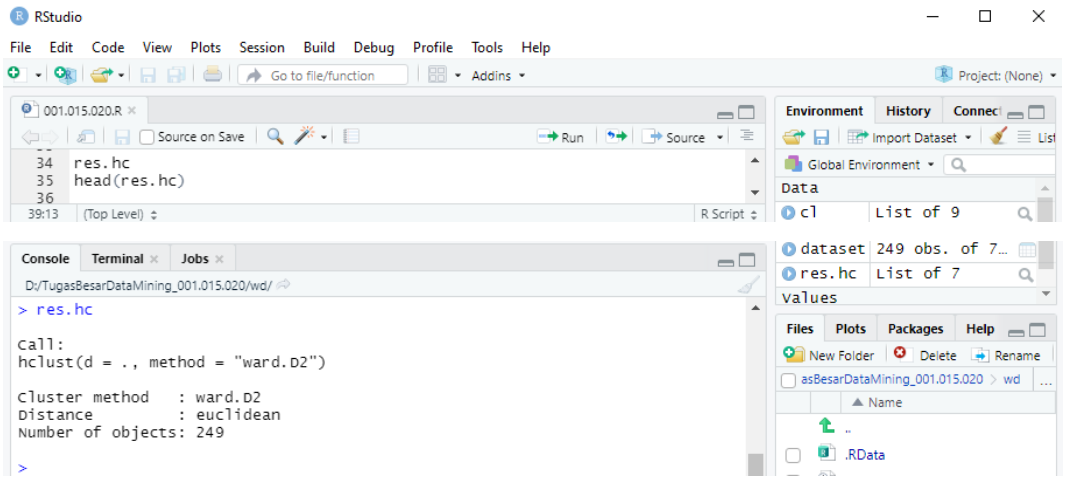
2. Dataset yang telah di import akan di scaling lalu di cluster

```
res.hc <- dataset[2:3] %>% scale() %>% dist(method
= "euclidean") %>% hclust(method = "ward.D2")
```



3. Melihat model yang telah dibangun

```
res.hc
```



head(res.hc)

```
> head(res.hc)
$merge
       [,1] [,2]
  [1,]   -7   -9
  [2,]  -13    1
  [3,]  -17  -41
  [4,]  -21  -35
  [5,]  -40    4
  [6,]  -44    5
  [7,]  -23  -27
  [8,]  -32    7
  [9,]  -25  -49
 [10,]  -29  -46
 [11,]  -36  -53
 [12,]  -37  -43
 [13,]  -38  -42
 [14,]  -45  -52
 [15,]  -55  -62
 [16,]  -92   15
 [17,]  -56  -65
 [18,]  -69   17
 [19,]  -70   18
 [20,]  -57  -85
 [21,]  -86   20
 [22,] -100   21
 [23,]  -58  -73
 [24,]  -87   23
 [25,]  -93   24
 [26,]  -94   25
 [27,]  -99   26
 [28,]  -59  -71
 [29,]  -60 -108
 [30,]  -61 -107
 [31,]  -64  -81
 [32,]  -89   31
 [33,]  -95   32
 [34,]  -67  -68
 [35,]  -74  -75
 [36,]  -76  -83
 [37,]  -77 -104
 [38,]  -79  -96
 [39,]  -97 -103
 [40,] -110 -119
 [41,] -121   40
 [42,] -141   41
 [43,] -111 -120
 [44,] -112 -154
 [45,] -115 -158
 [46,] -116 -137
 [47,] -117 -163
 [48,] -122 -132
 [49,] -147   48
 [50,] -164   49
 [51,] -127 -139
 [52,] -157   51
 [53,] -130 -159
 [54,] -131 -138
 [55,] -133 -150
 [56,] -165   55
 [57,] -135 -151
 [58,] -152   57
 [59,] -153   58
 [60,] -136 -143
 [61,] -142 -148
 [62,] -145 -156
 [63,] -162 -178
 [64,] -170 -249
 [65,] -171 -188
 [66,] -196   65
 [67,] -219   66
 [68,] -175 -198
 [69,] -177 -185
 [70,] -189   69
 [71,] -194   70
 [72,] -182 -202
 [73,] -206   72
 [74,] -220   73
 [75,] -221   74
 [76,] -240   75
 [77,] -184 -241
 [78,] -186 -205
 [79,] -190 -230
 [80,] -199 -212
 [81,] -246   80
 [82,] -211 -222
 [83,] -224   82
 [84,] -216 -226
 [85,] -223 -229
 [86,] -183 -217
 [87,] -167 -168
 [88,]  -78  -80
 [89,] -172 -193
 [90,] -187 -227
 [91,]  -33  -48
 [92,] -225 -231
 [93,] -125   60
 [94,]  -34   12
 [95,] -204   84
 [96,] -238   16
 [97,]  -15    2
 [98,] -105   33
 [99,]   43   62
[100,]   22   38
[101,] -123 -128
[102,] -140 -155
[103,] -109 -160
[104,]   -6   -8
[105,]  -50  -54
[106,] -124   54
[107,]  -39    6
[108,]  -19   94
[109,] -134   99
[110,]  -18    3
[111,]  -26  -30
[112,]  -11  -12
[113,] -173 -247
[114,]  -10  -16
[142,]  -31  104
[143,] -129   56
[144,]  -90  110
[145,]    8   13
[146,] -243   71
[147,] -210   92
[148,]  -14  107
[149,]   63   86
[150,]   59  101
[178,]  -47  153
[179,] -114  103
[180,] -248   64
[181,] -180  139
[182,]   -5  167
[183,] -169  135
[184,] -179  121
[185,]   67  113
[186,]  155  156
```

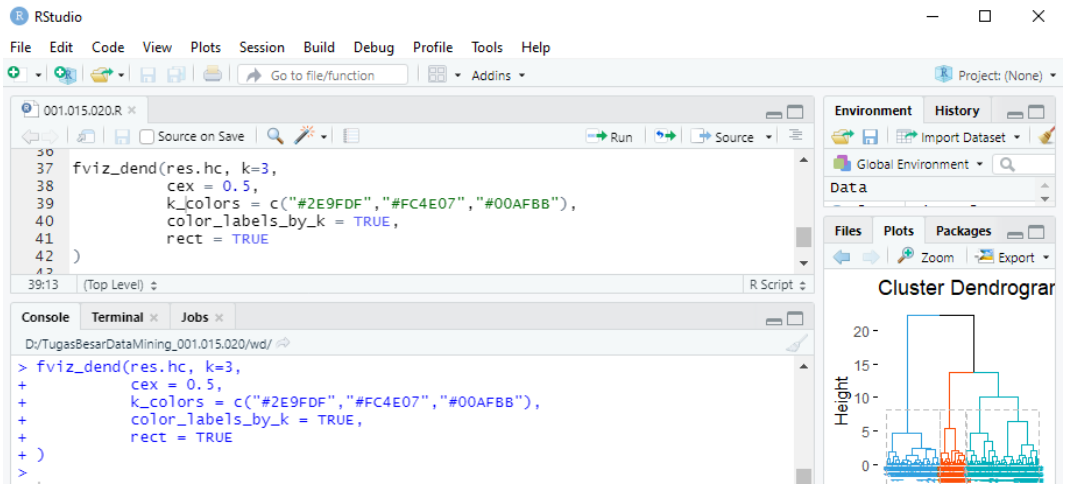
```
[115,] -200 -203 [151,]   68   79 [187,]  117  147
[116,]  -24  105 [152,]   44   93 [188,]  119  137
[117,] -245   85 [153,]   36  125 [189,]  152  179
[118,] -192   83 [154,] -197  127 [190,]   98  136
[119,]  -63 -106 [155,]   11  111 [191,]   -1  168
[120,] -201 -233 [156,]   10  116 [192,]  124  128
[121,] -174 -208 [157,]   14  138 [193,]  122  134
[122,]  -51 -146 [158,]  -28   88 [194,]  148  157
[123,] -176 -195 [159,] -118  143 [195,]  158  178
[124,] -207 -209 [160,]   34  100 [196,]  -98  162
[125,]  -20  -82 [161,]   19   39 [197,]   61  161
[126,]  -22   91 [162,]  -91  130 [198,]  169  176
[127,] -239   87 [163,]   97  140 [199,]  149  180
[128,] -126  106 [164,]   -3  142 [200,]  102  141
[129,] -181   95 [165,]   81   90 [201,]  131  170
[130,] -101   28 [166,]   37  145 [202,]  133  197
[131,] -149   46 [167,]  112  126 [203,] -237  120
[132,] -113   45 [168,]  108  114 [204,]   47  174
[133,]  -88   30 [169,] -213  115 [205,]  144  166
[134,] -161   53 [170,]   42   52 [206,]  132  175
[135,] -214   78 [171,] -218 -232 [207,]  165  185
[136,] -102   29 [172,] -234 -242 [208,]  151  199
[137,]  -72   35 [173,] -235 -244 [209,]  150  159
[138,]  -66    9 [174,] -144 -166 [210,]  191  194
[139,] -191 -215 [175,]   50  109 [211,]  186  190
[140,]   -2   -4 [176,]   77  123 [212,] -228  171
[141,]  -84   96 [177,]   89  146 [213,]  154  173
```

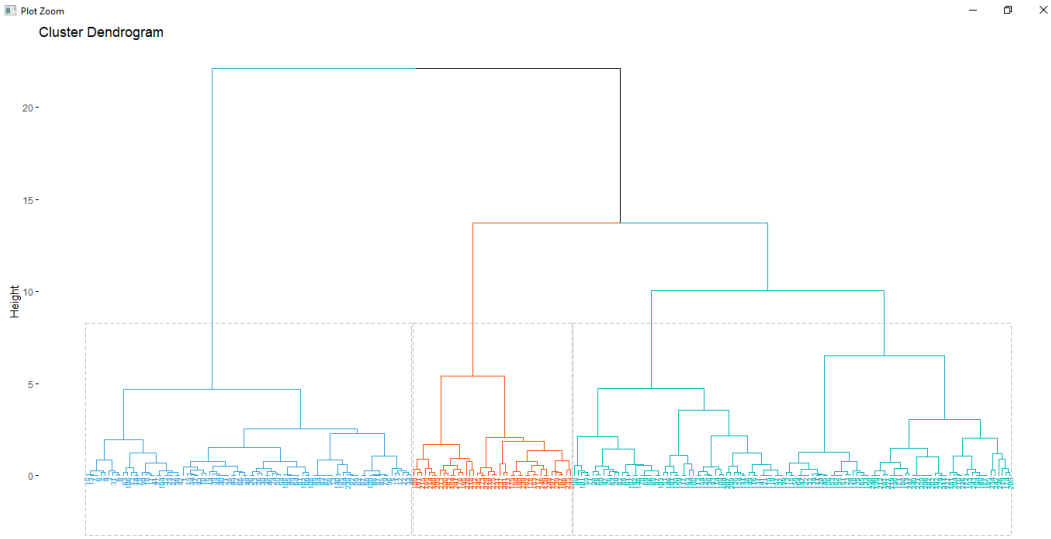
```
[214,]   27  200 [226,]  206  209
[215,]  163  164 [227,]  208  213 [238,]  214  221
[216,]  181  183 [228,]  195  202 [239,]  230  238
[217,] -236  184 [229,]  207  219 [240,]  229  234
[218,]  198  212 [230,]  210  211 [241,]  222  237
[219,]   76  118 [231,]  216  218 [242,]  233  239
[220,]  129  177 [232,]  203  227 [243,]  236  241
[221,]  160  182 [233,]  215  225 [244,]  231  235
[222,]  193  204 [234,]  220  223 [245,]  226  240
[223,]  172  217 [235,]  187  232 [246,]  243  245
[224,]  192  201 [236,]  196  228 [247,]  244  246
[225,]  188  205 [237,]  189  224 [248,]  242  247
```

```
$height
  [1]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
  [7]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [13]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [19]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [25]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [31]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [37]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [43]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [49]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [55]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [61]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [67]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [73]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [79]  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 [85]  0.00000000  0.03081273  0.03081273  0.03081273  0.03081273  0.03081273
 [91]  0.03081273  0.03081273  0.03557948  0.03557948  0.03557948  0.03773774
 [97]  0.03773774  0.03897537  0.04357579  0.05031698  0.06162547  0.06162547

[103]  0.06162547  0.06162547  0.06162547  0.07115896  0.07795074  0.08805472
[109]  0.09743842  0.10673844  0.12325094  0.12325094  0.12325094  0.12325094
[115]  0.12325094  0.14231792  0.14231792  0.15095095  0.15406367  0.15406367
[121]  0.15406367  0.15406367  0.15406367  0.15406367  0.15424625  0.16010766
[127]  0.16010766  0.16353020  0.17610944  0.17789740  0.17789740  0.17789740
[133]  0.17789740  0.17789740  0.17789740  0.17789740  0.17789740  0.17810823
[139]  0.18487640  0.18487640  0.18513300  0.18846814  0.18868869  0.18891231
[145]  0.19093953  0.19487684  0.19568714  0.19576512  0.19609104  0.20126793
[151]  0.21787893  0.22276278  0.22332249  0.22642643  0.23082539  0.23468092
[157]  0.23603937  0.23683561  0.24359605  0.24905636  0.25158492  0.25158492
[163]  0.25469437  0.26204373  0.26254185  0.26871657  0.27446225  0.28294878
[169]  0.28463584  0.28527048  0.30812734  0.30812734  0.30812734  0.31609294
[175]  0.32479473  0.32681840  0.33853967  0.34585908  0.35084536  0.35579480
[181]  0.35621646  0.37449787  0.37551408  0.40015001  0.40253587  0.40400388
[187]  0.40916402  0.43756975  0.45381151  0.48530463  0.48556211  0.49377260
[193]  0.50877027  0.51402901  0.53085236  0.54598518  0.55295322  0.56214195
[199]  0.56694696  0.57827914  0.60140298  0.61077116  0.62264091  0.65910968
[205]  0.66802750  0.68976684  0.70675966  0.74959988  0.75366166  0.77510372
[211]  0.78082348  0.78351530  0.79044701  0.83500056  0.84545165  0.87469636
[217]  0.88117776  0.91839924  1.01260976  1.06767317  1.07986411  1.09225303
[223]  1.20044635  1.23884646  1.24447354  1.26337648  1.35551864  1.45145756
[229]  1.46635503  1.50480172  1.69871783  1.84401516  1.94274188  2.05065520
[235]  2.06134290  2.11191495  2.15788648  2.29460507  2.52858637  3.05408611
[241]  3.56625207  4.69226723  4.73990383  5.42070157  6.49840197 10.02794947
[247] 13.70659732 22.10582050

$order
  [1]  15  13   7   9   2   4   3  31   6   8  63 106  72  74  75  90  18  17  41
 [20]  77 104  32  23  27  38  42   1  19  34  37  43  10  16  14  39  44  40  21
 [39]  35  45  52  66  25  49  36  53  26  30  29  46  24  50  54 105  95  89  64
 [58]  81 102  60 108  99  94  93  87  58  73 140 155  84 238  92  55  62  67  68
 [77] 100  86  57  85  79  96   5  11  12  22  33  48 180 191 215 169 214 186 205
 [96] 213 200 203 184 241 176 195 228 218 232 245 223 229 210 225 231 237 201 233
[115] 175 198 190 230 162 178 183 217 248 170 249 197 239 167 168 235 244  98  91
[134] 101  59  71  28  78  80  47  76  83  20  82  88  61 107 142 148  70  69  56
[153]  65  97 103  51 146 161 130 159 117 163 144 166 112 154 125 136 143 114 109
[172] 160 207 209 126 124 131 138 149 116 137 141 121 110 119 157 127 139 113 115
[191] 158 164 147 122 132 134 111 120 145 156 153 152 135 151 123 128 118 129 165
[210] 133 150 246 199 212 187 227 219 196 171 188 173 247 240 221 220 206 182 202
[229] 192 224 211 222 181 204 216 226 172 193 243 194 189 177 185 234 242 236 179
[248] 174 208

$labels
NULL

$method
[1] "ward.D2"

$call
hclust(d = ., method = "ward.D2")

>
```

4. Visualisasi data menggunakan fviz_dend

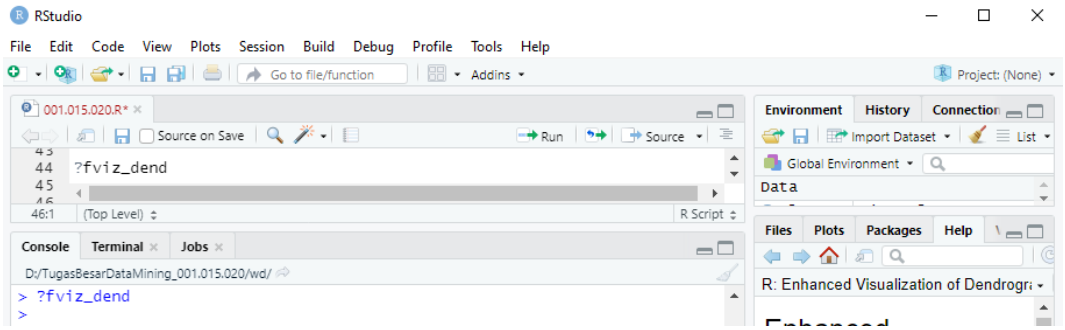
```
fviz_dend(res.hc, k = 3,
 cex = 0.5,
 k_colors = c("#2E9FDF", "#FC4E07", "#00AFBB"),
 color_labels_by_k = TRUE,
 rect = TRUE
)
```

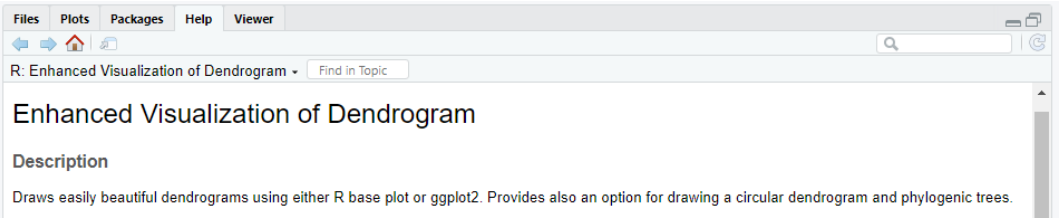




5. Untuk lebih lanjut informasi mengenai fviz_dend, Jalankan perintah

```
?fviz_dend
```





Usage

```
fviz_dend(
  x,
  k = NULL,
  h = NULL,
  k_colors = NULL,
  palette = NULL,
  show_labels = TRUE,
  color_labels_by_k = TRUE,
  label_cols = NULL,
  labels_track_height = NULL,
  repel = FALSE,
  lwd = 0.7,
  type = c("rectangle", "circular", "phylogenic"),
  phylo_layout = "layout.auto",
  rect = FALSE,
  rect_border = "gray",
  rect_lty = 2,
  rect_fill = FALSE,
  lower_rect,
  horiz = FALSE,
  cex = 0.8,
  main = "Cluster Dendrogram",
  xlab = "",
  ylab = "Height",
  sub = NULL,
  ggtheme = theme_classic(),
  ...
)
```

Arguments

x	an object of class dendrogram, hclust, agnes, diana, hcut, hkmeans or HCPC (FactoMineR).
k	the number of groups for cutting the tree.
h	a numeric value. Cut the dendrogram by cutting at height h. (k overrides h)
k_colors, palette	a vector containing colors to be used for the groups. It should contains k number of colors. Allowed values include also "grey" for grey color palettes; brewer palettes e.g. "RdBu", "Blues", ...; and scientific journal palettes from ggsci R package, e.g.: "npg", "aaas", "lancet", "jco", "ucscgb", "uchicago", "simpsons" and "rickandmorty".
show_labels	a logical value. If TRUE, leaf labels are shown. Default value is TRUE.
color_labels_by_k	logical value. If TRUE, labels are colored automatically by group when k != NULL.
label_cols	a vector containing the colors for labels.
labels_track_height	a positive numeric value for adjusting the room for the labels. Used only when type = "rectangle".
repel	logical value. Use repel = TRUE to avoid label overplotting when type = "phylogenic".
lwd	a numeric value specifying branches and rectangle line width.
type	type of plot. Allowed values are one of "rectangle", "triangle", "circular", "phylogenic".
phylo_layout	the layout to be used for phylogenic trees. Default value is "layout.auto". Allowed values include: layout.auto, layout_with_drl, layout_as_tree, layout.gem, layout.mds and layout_with_lgl.
rect	logical value specifying whether to add a rectangle around groups. Used only when k != NULL.
rect_border, rect_lty	border color and line type for rectangles.
rect_fill	a logical value. If TRUE, fill the rectangle.
lower_rect	a value of how low should the lower part of the rectangle around clusters. Ignored when rect = FALSE.
horiz	a logical value. If TRUE, an horizontal dendrogram is drawn.

cex	size of labels
main, xlab, ylab	main and axis titles
sub	Plot subtitle. If NULL, the method used hierarchical clustering is shown. To remove the subtitle use sub = "".
ggtheme	function, ggplot2 theme name. Default value is theme_classic(). Allowed values include ggplot2 official themes: theme_gray(), theme_bw(), theme_minimal(), theme_classic(), theme_void(), ...
...	other arguments to be passed to the function plot.dendrogram()

Value

an object of class fviz_dend which is a ggplot with the attributes "dendrogram" accessible using attr(x, "dendrogram"), where x is the result of fviz_dend().

Examples

```
# Load and scale the data
data(USArrests)
df <- scale(USArrests)

# Hierarchical clustering
res.hc <- hclust(dist(df))

# Default plot
fviz_dend(res.hc)

# Cut the tree
fviz_dend(res.hc, cex = 0.5, k = 4, color_labels_by_k = TRUE)
```

```
# Don't color labels, add rectangles
fviz_dend(res.hc, cex = 0.5, k = 4,
 color_labels_by_k = FALSE, rect = TRUE)

# Change the color of tree using black color for all groups
# Change rectangle border colors
fviz_dend(res.hc, rect = TRUE, k_colors ="black",
rect_border = 2:5, rect_lty = 1)

# Customized color for groups
fviz_dend(res.hc, k = 4,
 k_colors = c("#1B9E77", "#D95F02", "#7570B3", "#E7298A"))


 # Color labels using k-means clusters
 km.clust <- kmeans(df, 4)$cluster
 fviz_dend(res.hc, k = 4,
   k_colors = c("blue", "green3", "red", "black"),
   label_cols =  km.clust[res.hc$order], cex = 0.6)
```

[Package *factoextra* version 1.0.7 Index]

Kode Penuh

```
lokasi_kerja <- "D:/SHALIHA.001/3A/DATA MINING/Tugas Besar/wd"
setwd(lokasi_kerja)
getwd()

dataset <- read.csv("buddymove_holidayiq.csv", sep = ",")
head(dataset)


#K-Means
cl <- kmeans(dataset[2:3], 3)

cl

cl$size
```

```
cl$centers

plot(dataset[2:3], col = cl$cluster)

points(cl$centers, col = 1:2, pch = 8, cex =2)


#Agglomerative Hierarchical Clustering
install.packages("factoextra")
install.packages("cluster")
install.packages("magrittr")
library(factoextra)
library(cluster)
library(magrittr)

res.hc <- dataset[2:3] %>% scale() %>% dist(method = "euclidean")
%>%
  hclust(method = "ward.D2")

res.hc
head(res.hc)

fviz_dend(res.hc, k=3,
          cex = 0.5,
          k_colors = c("#2E9FDF","#FC4E07","#00AFBB"),
          color_labels_by_k = TRUE,
          rect = TRUE
)

?fviz_dend
```