SCRIPT COMPLETO (POPAJ – VILEI – PREMOLI)

1 - Analisi Univariata

```
> library(readxl)
> Dataset_1 <- read_excel("Desktop/Università/Fondamenti di Data Analytics/Tesina
- Importanza dei Ricercatori/Dataset_1.xlsx")
> View(Dataset_1)
>
> summary(Dataset_1)
  Countries         Researchers     Invest_on_GDP     Patent_App       Edu_Spending
 Length:44          Min.   : 0.500   Min.   :0.289   Min.   :  0.1763   Min.   :0.4180
 Class :character   1st Qu.: 6.134   1st Qu.:1.126   1st Qu.:  1.7233   1st Qu.:0.6880
 Mode  :character   Median : 8.800   Median :1.847   Median :  2.7125   Median
:0.9035
                    Mean   : 8.591   Mean   :2.006   Mean   : 11.0789   Mean   :0.9809
                    3rd Qu.:11.038   3rd Qu.:2.916   3rd Qu.:  5.1544   3rd Qu.:1.1487
                    Max.   :16.605   Max.   :5.706   Max.   :109.7341   Max.   :2.2000
   Education     Unemployment_Rate GDP_percapita
 Min.   :15.80   Min.   : 2.550    Min.   :  5363
 1st Qu.:30.65   1st Qu.: 4.594    1st Qu.: 17231
 Median :40.35   Median : 6.213    Median : 31757
 Mean   :39.85   Mean   : 7.375    Mean   : 36494
 3rd Qu.:49.52   3rd Qu.: 8.198    3rd Qu.: 49932
 Max.   :69.90   Max.   :28.700    Max.   :118084
>
> Calcolo del Range
> max(Dataset_1$Researchers)-min(Dataset_1$Researchers)
[1] 16.105
> max(Dataset_1$Invest_on_GDP)-min(Dataset_1$Invest_on_GDP)
[1] 5.417
> max(Dataset_1$Patent_App)-min(Dataset_1$Patent_App)
[1] 109.5578
> max(Dataset_1$Edu_Spending)-min(Dataset_1$Edu_Spending)
[1] 1.782
> max(Dataset_1$Education)-min(Dataset_1$Education)
[1] 54.1
> max(Dataset_1$Unemployment_Rate)-min(Dataset_1$Unemployment_Rate)
[1] 26.15
> max(Dataset_1$GDP_percapita)-min(Dataset_1$GDP_percapita)
[1] 112.721
>
> #Calcolo dello Scarto Interquartile
> IQR(Dataset_1$Researchers)
[1] 4.90425
> IQR(Dataset_1$Invest_on_GDP)
[1] 1.79025
> IQR(Dataset_1$Patent_App)
```

```
[1] 3.43106
> IQR(Dataset_1$Edu_Spending)
[1] 0.46075
> IQR(Dataset_1$Education)
[1] 18.875
> IQR(Dataset_1$Unemployment_Rate)
[1] 3.6045
> IQR(Dataset_1$GDP_percapita)
[1] 32700.57
>
> #Calcolo dello Scarto Interquartile
> IQR(Dataset_1$Researchers)
[1] 4.90425
> IQR(Dataset_1$Invest_on_GDP)
[1] 1.79025
> IQR(Dataset_1$Patent_App)
[1] 3.43106
> IQR(Dataset_1$Edu_Spending)
[1] 0.46075
> IQR(Dataset_1$Education)
[1] 18.875
> IQR(Dataset_1$Unemployment_Rate)
[1] 3.6045
> IQR(Dataset_1$GDP_percapita)
[1] 32700.57
>
> #Calcolo della Varianza
> var(Dataset_1$Researchers)
[1] 16.7784
> var(Dataset_1$Invest_on_GDP)
[1] 1.409743
> var(Dataset_1$Patent_App)
[1] 623.8605
> var(Dataset_1$Edu_Spending)
[1] 0.1651752
> var(Dataset_1$Education)
[1] 145.7402
> var(Dataset_1$Unemployment_Rate)
[1] 22.76318
> var(Dataset_1$GDP_percapita)
[1] 607142278
>
> #Calcolo dei Boxplot
> boxplot(Dataset_1$Researchers)
> boxplot(Dataset_1$Invest_on_GDP)
> boxplot(Dataset_1$Patent_App)
> boxplot(Dataset_1$Edu_Spending)
> boxplot(Dataset_1$Education)
```

```
> boxplot(Dataset_1$Unemployment_Rate)
> boxplot(Dataset_1$GDP_percapita)
> #Calcolo della Curva di Densità per Deviazioni Standard estremamente alte
> plot(density(Dataset_1$Patent_App))
> plot(density(Dataset_1$GDP_percapita))
>
> #Calcolo del Coefficiente di Gini e rappresentazione della Curva di Lorenz
> library(ineq)
> Gini(Dataset_1$Researchers)
[1] 0.2660657
> Gini(Dataset_1$Invest_on_GDP)
[1] 0.322032
> Gini(Dataset_1$Patent_App)
[1] 0.9183375
> Gini(Dataset_1$Edu_Spending)
[1] 0.2233691
> Gini(Dataset_1$Education)
[1] 0.1699172
> Gini(Dataset_1$Unemployment_Rate)
[1] 0.2978976
> Gini(Dataset_1$GDP_percapita)
[1] 0.3629871
>
> plot(Lc(Dataset_1$Researchers))
> plot(Lc(Dataset_1$Invest_on_GDP))
> plot(Lc(Dataset_1$Patent_App))
> plot(Lc(Dataset_1$Edu_Spending))
> plot(Lc(Dataset_1$Education))
> plot(Lc(Dataset_1$Unemployment_Rate))
> plot(Lc(Dataset_1$GDP_percapita))
>
> #Calcolo della Correlazione
> cor(Dataset_1[,2:8])
                  Researchers Invest_on_GDP Patent_App Edu_Spending  Education
Unemployment_Rate GDP_percapita
Researchers        1.0000000     0.6956306  0.2848665    0.4069525  0.4814344
-0.4158348    0.4942420
Invest_on_GDP      0.6956306     1.0000000  0.4405414    0.2527862  0.4930810
-0.3945079    0.3903168
Patent_App         0.2848665     0.4405414  1.0000000    0.1766262  0.3263952
-0.2236709   -0.1367245
Edu_Spending       0.4069525     0.2527862  0.1766262    1.0000000  0.3144050
-0.1879829    0.1658577
Education          0.4814344     0.4930810  0.3263952    0.3144050  1.0000000       -
0.4123959    0.5303689
Unemployment_Rate  -0.4158348    -0.3945079 -0.2236709   -0.1879829 -
0.4123959        1.0000000    -0.3322000
```

```
GDP_percapita      0.4942420    0.3903168 -0.1367245    0.1658577  0.5303689
-0.3322000    1.0000000
>
# Creazione dei Grafici di Dispersione
> plot(Dataset_1$Researchers ~ Dataset_1$Invest_on_GDP)
> plot(Dataset_1$Researchers ~ Dataset_1$Edu_Spending)
> plot(Dataset_1$Researchers ~ Dataset_1$Education)
> plot(Dataset_1$Researchers ~ Dataset_1$GDP_percapita)
> plot(Dataset_1$Invest_on_GDP ~ Dataset_1$Patent_App)
> plot(Dataset_1$Invest_on_GDP ~ Dataset_1$Edu_Spending)
> plot(Dataset_1$Education ~ Dataset_1$Unemployment_Rate)
> plot(Dataset_1$Unemployment_Rate ~ Dataset_1$GDP_percapita)
> plot(Dataset_1$Education ~ Dataset_1$GDP_percapita)
>
> # Regressioni Lineari Semplici Rilevanti
> r1<-lm(Dataset_1$Researchers~Dataset_1$Invest_on_GDP)
> summary(r1)

Call:
lm(formula = Dataset_1$Researchers ~ Dataset_1$Invest_on_GDP)

Residuals:
    Min      1Q  Median      3Q     Max
-8.6695 -1.5705  0.2347  1.7413  5.1317

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)             3.7760     0.8889   4.248 0.000117 ***
Dataset_1$Invest_on_GDP   2.3998     0.3824   6.275 1.59e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.977 on 42 degrees of freedom
Multiple R-squared:  0.4839,      Adjusted R-squared:  0.4716
F-statistic: 39.38 on 1 and 42 DF,  p-value: 1.595e-07
>
> r2<-lm(Dataset_1$Education~Dataset_1$GDP_percapita)
> summary(r2)

Call:
lm(formula = Dataset_1$Education ~ Dataset_1$GDP_percapita)

Residuals:
    Min      1Q  Median      3Q     Max
-18.524  -7.519  -1.247   5.286  23.690

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)           3.036e+01  2.812e+00  10.798 1.08e-13 ***
Dataset_1$GDP_percapita 2.599e-04  6.409e-05   4.054 0.000213 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.36 on 42 degrees of freedom
Multiple R-squared:  0.2813,	Adjusted R-squared:  0.2642 
F-statistic: 16.44 on 1 and 42 DF,  p-value: 0.000213

>
> r3<-lm(Dataset_1$Researchers~Dataset_1$Education)
> summary(r4)

Call:
lm(formula = Dataset_1$Researchers ~ Dataset_1$Education)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7174 -3.5168 -0.1662  2.7891  6.4829 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)          2.08134    1.90880    1.09 0.281757    
Dataset_1$Education  0.16335    0.04589    3.56 0.000938 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.633 on 42 degrees of freedom
Multiple R-squared:  0.2318,	Adjusted R-squared:  0.2135 
F-statistic: 12.67 on 1 and 42 DF,  p-value: 0.0009379

>
> r4<-lm(Dataset_1$Researchers~Dataset_1$Unemployment_Rate)
> summary(r4)

Call:
lm(formula = Dataset_1$Researchers ~ Dataset_1$Unemployment_Rate)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4391 -2.1415 -0.1519  2.3245  7.6562 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  11.2234     1.0547  10.642  1.7e-13 ***
Dataset_1$Unemployment_Rate  -0.3570     0.1205  -2.963    0.005 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.769 on 42 degrees of freedom
```

```
Multiple R-squared:  0.1729,	Adjusted R-squared:  0.1532 
F-statistic: 8.781 on 1 and 42 DF,  p-value: 0.004996
> 
> #Regressione Multipla
> #Prima Regressione
> modello <- lm(Dataset_1_2_$Researchers ~Dataset_1_2_$Invest_on_GDP + 
Dataset_1_2_$Patent_App + Dataset_1_2_$Edu_Spending, data = Dataset_1_2_)
> summary(modello)
> #Seconda Regressione
modello <- lm(Dataset_1_2_$Researchers ~Dataset_1_2_$Invest_on_GDP + 
Dataset_1_2_$Patent_App + Dataset_1_2_$Edu_Spending + 
Dataset_1_2_$Education+Dataset_1_2_$Unemployment_Rate + 
Dataset_1_2_$GDP_percapita ,data = Dataset_1_2_)
> summary(modello)
> 
#Realizzazione delle silhouette
> Dataset_touse <- Dataset_1[,2:8]
> View(Dataset_touse)
> Dataset_scale <- scale(Dataset_touse)
> set.seed(54)
> kclust_1 <- kmeans(Dataset_scale, centers=3)
> kclust_1
K-means clustering with 3 clusters of sizes 22, 19, 3

Cluster means:
  Researchers Invest_on_GDP Patent_App Edu_Spending  Education
1   0.5032282     0.4940436 -0.2182525    0.2571806  0.5107488
2  -0.7233751    -0.7851988 -0.2986530   -0.4450752 -0.7542580
3   0.8910355     1.3499392  3.4919871    0.9328181  1.0314755
  Unemployment_Rate GDP_percapita...8
1        -0.3356589         0.7823030
2         0.4910530        -0.8232534
3        -0.6485035        -0.5229508

Clustering vector:
 [1] 2 1 1 1 1 2 3 3 2 2 1 2 1 1 1 2 2 1 1 1 2 1 3 2 2 1 2 1 1 1 2 2 2 2 1 2 1 2
[39] 2 1 1 2 1 1

Within cluster sum of squares by cluster:
[1] 81.14595 64.55862 12.47190
 (between_SS / total_SS =  47.4 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"      
> library(cluster)
> set.seed(54)
```

```
> sil_1 <- silhouette(kclust_1$cluster,dist(Dataset_touse[,1:7]))
> sil_1
      cluster neighbor   sil_width
 [1,]       2        3  0.41829229
 [2,]       1        3  0.52037299
 [3,]       1        3  0.44372104
 [4,]       1        3  0.32581490
 [5,]       1        3  0.20342868
 [6,]       2        3  0.44227955
 [7,]       3        2 -0.62300181
 [8,]       3        2  0.16532543
 [9,]       2        3  0.37254955
[10,]       2        3  0.03106139
[11,]       1        3  0.55688563
[12,]       2        3 -0.03638288
[13,]       1        3  0.45165022
[14,]       1        3 -0.04425004
[15,]       1        3  0.37521130
[16,]       2        3  0.39947543
[17,]       2        3  0.43415817
[18,]       1        3  0.55371556
[19,]       1        3  0.44736991
[20,]       1        3  0.29130855
[21,]       2        3 -0.50130085
[22,]       1        3 -0.07144791
[23,]       3        2  0.21567788
[24,]       2        3  0.38167995
[25,]       2        3  0.26433343
[26,]       1        3  0.30901613
[27,]       2        3  0.41803646
[28,]       1        3  0.51017971
[29,]       1        3  0.04306842
[30,]       1        3  0.52602546
[31,]       2        3  0.43763771
[32,]       2        3  0.11005804
[33,]       2        3  0.44193487
[34,]       2        3  0.40863436
[35,]       1        3  0.55664932
[36,]       2        3  0.31400970
[37,]       1        3 -0.74502874
[38,]       2        3  0.37781209
[39,]       2        3 -0.35870376
[40,]       1        3  0.51674170
[41,]       1        3  0.44424424
[42,]       2        3  0.41826822
[43,]       1        3 -0.04798885
[44,]       1        3  0.54927083
attr(,"Ordered")
```

```
[1] FALSE
attr(,"call")
silhouette.default(x = kclust_1$cluster, dist = dist(Dataset_touse[,
    1:7]))
attr(,"class")
[1] "silhouette"
> plot(sil_1)
>
> set.seed(54)
> kclust_2 <- kmeans(Dataset_scale, centers=4)
> kclust_2
K-means clustering with 4 clusters of sizes 15, 7, 3, 19

Cluster means:
  Researchers Invest_on_GDP Patent_App Edu_Spending  Education
1  -0.4541806    -0.5182921 -0.2976016   -0.5210234 -0.4628556
2  -1.0996632    -0.9697988 -0.3231012   -0.3143159 -0.9861317
3   0.8910355     1.3499392  3.4919871    0.9328181  1.0314755
4   0.6230129     0.5533240 -0.1973805    0.3798478  0.5658594
  Unemployment_Rate GDP_percapita...8
1        -0.3333267        -0.5897190
2         1.8086034        -0.9837373
3        -0.6485035        -0.5229508
4        -0.3007797         0.9105683

Clustering vector:
 [1] 2 4 4 4 4 2 3 3 2 1 4 1 4 4 4 2 1 4 4 4 1 4 3 1 1 4 1 4 1 4 1 1 1 1 4 1 1 2
[39] 2 4 4 2 1 4

Within cluster sum of squares by cluster:
[1] 29.26060 21.20957 12.47190 69.94290
 (between_SS / total_SS =  55.9 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
> sil_2 <- silhouette(kclust_2$cluster,dist(Dataset_touse[,1:7]))
> sil_2
      cluster neighbor   sil_width
 [1,]       2        1  0.52665155
 [2,]       4        3  0.53839836
 [3,]       4        3  0.43750283
 [4,]       4        3  0.28784156
 [5,]       4        3  0.13497245
 [6,]       2        1  0.28328576
 [7,]       3        2 -0.70882749
 [8,]       3        1 -0.02790318
```

```
 [9,]       2        1  0.51566880
[10,]       1        3  0.11321566
[11,]       4        3  0.59557236
[12,]       1        3  0.07023920
[13,]       4        3  0.44772932
[14,]       4        3 -0.14423685
[15,]       4        3  0.35021967
[16,]       2        1 -0.18704584
[17,]       1        2 -0.15250330
[18,]       4        3  0.58911157
[19,]       4        3  0.48908411
[20,]       4        3  0.24456650
[21,]       1        3 -0.37443579
[22,]       4        3 -0.16860817
[23,]       3        1  0.00198283
[24,]       1        2  0.04988278
[25,]       1        3  0.22462353
[26,]       4        3  0.33909285
[27,]       1        2 -0.62134185
[28,]       4        3  0.52446685
[29,]       1        3 -0.14602141
[30,]       4        3  0.56850323
[31,]       1        2 -0.18106223
[32,]       1        3  0.15633524
[33,]       1        2 -0.42882220
[34,]       1        2 -0.55954077
[35,]       4        3  0.59568210
[36,]       1        2  0.18385784
[37,]       1        3 -0.11786614
[38,]       2        1  0.52204616
[39,]       2        3 -0.55584549
[40,]       4        3  0.53329907
[41,]       4        3  0.48577467
[42,]       2        1  0.52684939
[43,]       1        3 -0.15605385
[44,]       4        3  0.59017093
attr(,"Ordered")
[1] FALSE
attr(,"call")
silhouette.default(x = kclust_2$cluster, dist = dist(Dataset_touse[,
    1:7]))
attr(,"class")
[1] "silhouette"
> plot(sil_2)
>
> kclust_3 <- kmeans(Dataset_scale, centers=5)
> set.seed(54)
> kclust_3 <- kmeans(Dataset_scale, centers=5)
```

```
> kclust_3
K-means clustering with 5 clusters of sizes 7, 14, 3, 7, 13

Cluster means:
  Researchers Invest_on_GDP Patent_App Edu_Spending  Education
1  -1.0996632    -0.9697988 -0.3231012   -0.3143159 -0.9861317
2  -0.4152915    -0.6107688 -0.2888041   -0.4684502 -0.5524349
3   0.8910355     1.3499392  3.4919871    0.9328181  1.0314755
4   1.2591052     0.5877224 -0.2913600    1.4625337  0.5285528
5   0.1557601     0.5519601 -0.1639597   -0.3290520  0.6032858
  Unemployment_Rate GDP_percapita...8
1         1.8086034        -0.9837373
2        -0.3152205        -0.6430124
3        -0.6485035        -0.5229508
4        -0.3118221         0.7623899
5        -0.3168363         0.9323430

Clustering vector:
 [1] 1 5 4 4 5 1 3 3 1 2 4 2 4 5 5 1 2 5 5 5 2 5 3 2 2 5 2 5 2 4 2 2 2 2 4 2 2 1
[39] 1 4 5 1 5 5

Within cluster sum of squares by cluster:
[1] 21.20957 24.19999 12.47190 13.55557 41.81187
 (between_SS / total_SS =  62.4 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
> sil_3 <- silhouette(kclust_3$cluster,dist(Dataset_touse[,1:7]))
> sil_3
      cluster neighbor   sil_width
 [1,]       1        2  0.47522692
 [2,]       5        4 -0.61256395
 [3,]       4        5  0.47426759
 [4,]       4        5  0.32359939
 [5,]       5        4 -0.37304513
 [6,]       1        2  0.18036086
 [7,]       3        1 -0.70882749
 [8,]       3        2 -0.03781033
 [9,]       1        2  0.47406399
[10,]       2        3  0.19450264
[11,]       4        5  0.49449082
[12,]       2        3  0.14643601
[13,]       4        5  0.48302340
[14,]       5        4 -0.28128929
[15,]       5        4 -0.50344004
[16,]       1        2 -0.30245862
```

```
[17,]      2       1 -0.01725455
[18,]      5       4 -0.59777840
[19,]      5       4 -0.10571889
[20,]      5       4 -0.43205468
[21,]      2       3 -0.39068652
[22,]      5       4 -0.27672312
[23,]      3       2  0.02249040
[24,]      2       1  0.18056857
[25,]      2       3  0.32187186
[26,]      5       4 -0.01097687
[27,]      2       1 -0.58244475
[28,]      5       4 -0.61002061
[29,]      2       4 -0.37204966
[30,]      4       5  0.29961013
[31,]      2       1 -0.05148958
[32,]      2       3  0.24323115
[33,]      2       1 -0.35028978
[34,]      2       1 -0.50867374
[35,]      4       5  0.49002984
[36,]      2       1  0.29146954
[37,]      2       3 -0.07413568
[38,]      1       2  0.48017320
[39,]      1       3 -0.55584549
[40,]      4       5  0.51324795
[41,]      5       4 -0.09952451
[42,]      1       2  0.47548172
[43,]      5       4 -0.28044708
[44,]      5       4 -0.53621904
attr(,"Ordered")
[1] FALSE
attr(,"call")
silhouette.default(x = kclust_3$cluster, dist = dist(Dataset_touse[,
    1:7]))
attr(,"class")
[1] "silhouette"
> plot(sil_3)
>
>> #Utilizzo dell'Elbow Method per confermare la valutazione
> library(cluster)
> library(factoextra)
Loading required package: ggplot2
Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa
> fviz_nbclust(Dataset_scale, kmeans, method = "wss")
>> #KMeans Clustering
> View(Dataset_1)
> Dataset_touse <- Dataset_1[,2:8]
> Dataset_scale <- scale(Dataset_touse)
```

```
> set.seed(54)
> kclust_1 <- kmeans(Dataset_scale, centers=3, set.seed(54))
Error in if (is.na(iter.max) II iter.max < 1L) stop("'iter.max' must be positive") :
  missing value where TRUE/FALSE needed
> kclust_1 <- kmeans(Dataset_scale, centers=3)
> set.seed(54)
> kclust_1 <- kmeans(Dataset_scale, centers=3, set.seed(54))
Error in if (is.na(iter.max) II iter.max < 1L) stop("'iter.max' must be positive") :
  missing value where TRUE/FALSE needed
> set.seed(54)
> kclust_1 <- kmeans(Dataset_scale, centers=3)
> table(kclust_1$cluster)

 1  2  3
22 19  3

> plot(kclust_1$cluster)
> centerskclust_1 <- kclust_1$centers
> View(centerskclust_1)
> aggregate((Dataset_scale, by=list(kclust_1$cluster),mean))
Error: unexpected ',' in "aggregate((Dataset_scale,"
> aggregate(Dataset_scale, by=list(kclust_1$cluster),mean))
Error: unexpected ')' in "aggregate(Dataset_scale, by=list(kclust_1$cluster),mean))"
> aggregate(Dataset_scale,by=list(kclust_1$cluster),mean)
  Group.1 Researchers Invest_on_GDP Patent_App Edu_Spending  Education
1       1   0.5032282     0.4940436 -0.2182525    0.2571806  0.5107488
2       2  -0.7233751    -0.7851988 -0.2986530   -0.4450752 -0.7542580
3       3   0.8910355     1.3499392  3.4919871    0.9328181  1.0314755
  Unemployment_Rate GDP_percapita...8
1        -0.3356589         0.7823030
2         0.4910530        -0.8232534
3        -0.6485035        -0.5229508
>
> #Rappresentazione del grafico
> set.seed(54)
> kmodelS <- kmeans(Dataset_scale, centers=3)
> fviz_cluster(kmodelS, data = Dataset_scale, stand = FALSE, geom = "point")
>
> #Hierarchical Clustering
> Dataset_touse <- Dataset_1[,2:8]
> Dataset_scale <- scale(Dataset_touse)
> distmat_euclidean <- dist(Dataset_scale, method='euclidean')
> hclust_complete <- hclust(distmat_euclidean)
> plot(hclust_complete)
> hclust_single <- hclust(distmat_euclidean,method="single")
> plot(hclust_single)
> hclust_average <- hclust(distmat_euclidean,method="average")
> plot(hclust_average)
```

```
> cluster_1 <- cutree(hclust_complete,k=3)
> table(cluster_1)
cluster_1
 1  2  3
39  4  1
> plot(cluster_1)
> cluster_2 <- cutree(hclust_single,k=3)
> table(cluster_2)
cluster_2
 1  2  3
40  3  1
> plot(cluster_2)
> cluster_3 <- cutree(hclust_average,k=3)
> table(cluster_3)
cluster_3
 1  2  3
40  3  1
> plot(cluster_3)
```