Anexa 1- GRAFICE SI ANALIZE

1. Analiza Componentelor Principale

```
                         vars  n  mean    sd median trimmed   mad   min    max range  skew kurtosis   se
Country Name*               1 67 34.00 19.49  34.00   34.00 25.20  1.00  67.00 66.00  0.00    -1.25 2.38
ER_primary                  2 67 95.47  3.57  96.36   95.81  2.66 85.17  99.99 14.82 -0.90    -0.08 0.44
Repeaters_primary           3 67  1.56  1.25   1.27    1.44  1.09  0.00   4.63  4.63  0.80    -0.43 0.15
ER_secondary                4 67 96.29 13.47  97.54   96.53 13.47 67.31 129.12 61.81 -0.13    -0.27 1.65
Progression_secondary       5 67 96.48  3.45  98.01   96.90  2.33 87.54  99.97 12.43 -0.88    -0.43 0.42
Primary_completionRate      6 67 97.44  5.07  97.34   97.37  5.59 87.68 111.30 23.62  0.23    -0.16 0.62
Adolescents_OutofSchool     7 67  5.90  5.17   4.98    5.30  5.03  0.06  18.71 18.65  0.84    -0.26 0.63
Children_OutofSchool        8 67  2.96  3.08   1.85    2.53  2.12  0.00  10.42 10.42  1.11    -0.02 0.38
ER_tertiary                 9 67 54.19 19.87  51.08   54.60 24.14  8.46  88.18 79.72 -0.16    -0.85 2.43
Government_expenditure     10 67  4.37  0.68   4.43    4.38  0.61  2.76   5.86  3.10 -0.13    -0.04 0.08
Compulsory_education       11 67 10.75  1.97  10.00   10.59  1.48  6.00  15.00  9.00  0.66    -0.22 0.24
ER_preprimary              12 67 73.47 21.12  73.28   74.10 24.31 22.21 112.31 90.10 -0.27    -0.25 2.58
OverAge_primary            13 67  5.92  3.93   7.09    5.58  5.01  0.01  16.43 16.42  0.57    -0.27 0.48
```

Fig 1. Skewness&Kurtosis, Rstudio

```
> apply(education[,2:13], 2, cv)
             ER_primary       Repeaters_primary            ER_secondary   Progression_secondary  Primary_completionRate
               3.735592               79.935549               13.984484                3.575458                5.199899
Adolescents_OutofSchool    Children_OutofSchool             ER_tertiary  Government_expenditure    Compulsory_education
              87.719807              104.032496               36.676513               15.506489               18.338779
          ER_preprimary         OverAge_primary
              28.751521               66.450321
> sd(education$ER_primary)/mean(education$ER_primary)
[1] 0.03735592
```

Fig 2. Coeficientul de variație, Rstudio

```
> round(cov(education[,2:13]), 3)
                        ER_primary Repeaters_primary ER_secondary Progression_secondary Primary_completionRate Adolescents_OutofSchool
ER_primary                  12.720            -0.942       27.540                 3.969                 11.676                  -9.863
Repeaters_primary           -0.942             1.554       -3.514                -2.155                 -1.089                   1.493
ER_secondary                27.540            -3.514      181.342                26.713                 41.099                 -53.885
Progression_secondary        3.969            -2.155       26.713                11.901                  5.600                 -11.395
Primary_completionRate      11.676            -1.089       41.099                 5.600                 25.671                 -13.543
Adolescents_OutofSchool     -9.863             1.493      -53.885               -11.395                -13.543                  26.761
Children_OutofSchool        -8.660             1.085      -28.189                -5.052                -11.439                  11.200
ER_tertiary                 20.155            -2.848      187.870                32.878                 30.371                 -72.388
Government_expenditure       0.270            -0.001        2.882                 0.559                  0.350                  -1.181
Compulsory_education         0.027             1.023        1.570                -0.543                  2.312                  -0.580
ER_preprimary                3.906            -0.999      134.173                18.490                 11.038                 -38.876
OverAge_primary             -6.343             2.349      -16.880                -4.863                 -3.608                   5.303
                        Children_OutofSchool ER_tertiary Government_expenditure Compulsory_education ER_preprimary OverAge_primary
ER_primary                            -8.660      20.155                  0.270                0.027         3.906          -6.343
Repeaters_primary                      1.085      -2.848                 -0.001                1.023        -0.999           2.349
ER_secondary                         -28.189     187.870                  2.882                1.570       134.173         -16.880
Progression_secondary                 -5.052      32.878                  0.559               -0.543        18.490          -4.863
Primary_completionRate               -11.439      30.371                  0.350                2.312        11.038          -3.608
Adolescents_OutofSchool               11.200     -72.388                 -1.181               -0.580       -38.876           5.303
Children_OutofSchool                   9.457     -26.351                 -0.695                0.065       -18.059           3.167
ER_tertiary                          -26.351     394.987                  3.494               -0.725       217.810          -0.747
Government_expenditure                -0.695       3.494                  0.460                0.253         5.591           0.121
Compulsory_education                   0.065      -0.725                  0.253                3.889         7.352           1.688
ER_preprimary                        -18.059     217.810                  5.591                7.352       446.249           1.591
OverAge_primary                        3.167      -0.747                  0.121                1.688         1.591          15.457
```

Fig 3. Matricea de covarianță, Rstudio

```
P
                        ER_primary Repeaters_primary ER_secondary Progression_secondary Primary_completionRate
ER_primary                         0.0771            0.0000       0.0134                0.0000
Repeaters_primary       0.0771                       0.1061       0.0000                0.1811
ER_secondary            0.0000     0.1061                         0.0000                0.0000
Progression_secondary   0.0134     0.0000            0.0000                             0.0072
Primary_completionRate  0.0000     0.1811            0.0000       0.0072
Adolescents_OutofSchool 0.0000     0.0893            0.0000       0.0000                0.0000
Children_OutofSchool    0.0000     0.0266            0.0000       0.0000                0.0000
ER_tertiary             0.0259     0.2738            0.0000       0.0000                0.0130
Government_expenditure  0.3719     0.9637            0.0083       0.0245                0.2622
Compulsory_education    0.8971     0.0003            0.7242       0.5210                0.0628
ER_preprimary           0.7646     0.7839            0.0000       0.0265                0.4078
OverAge_primary         0.0002     0.0000            0.0100       0.0082                0.1390
                        Adolescents_OutofSchool Children_OutofSchool ER_tertiary Government_expenditure Compulsory_education
ER_primary              0.0000                  0.0000               0.0259      0.3719                 0.8971
Repeaters_primary       0.0893                  0.0266               0.2738      0.9637                 0.0003
ER_secondary            0.0000                  0.0000               0.0000      0.0083                 0.7242
Progression_secondary   0.0000                  0.0000               0.0000      0.0245                 0.5210
Primary_completionRate  0.0000                  0.0000               0.0130      0.2622                 0.0628
Adolescents_OutofSchool                         0.0000               0.0000      0.0025                 0.8053
Children_OutofSchool    0.0000                                       0.0002      0.0046                 0.8398
ER_tertiary             0.0000                  0.0002                           0.0185                 0.8023
Government_expenditure  0.0025                  0.0046               0.0185                             0.0609
Compulsory_education    0.8053                  0.8398               0.8023      0.0609
ER_preprimary           0.0031                  0.0172               0.0000      0.0005                 0.1676
OverAge_primary         0.0423                  0.0461               0.9967      0.9765                 0.0750
                        ER_preprimary OverAge_primary
ER_primary              0.7646        0.0002
Repeaters_primary       0.7839        0.0000
ER_secondary            0.0000        0.0100
Progression_secondary   0.0265        0.0082
Primary_completionRate  0.4078        0.1390
Adolescents_OutofSchool 0.0031        0.0423
Children_OutofSchool    0.0172        0.0461
ER_tertiary             0.0000        0.9967
Government_expenditure  0.0005        0.9765
Compulsory_education    0.1676        0.0750
ER_preprimary                         0.7887
OverAge_primary         0.7887
>
```

Fig 4. Funcția rcorr pe setul de date, Rstudio

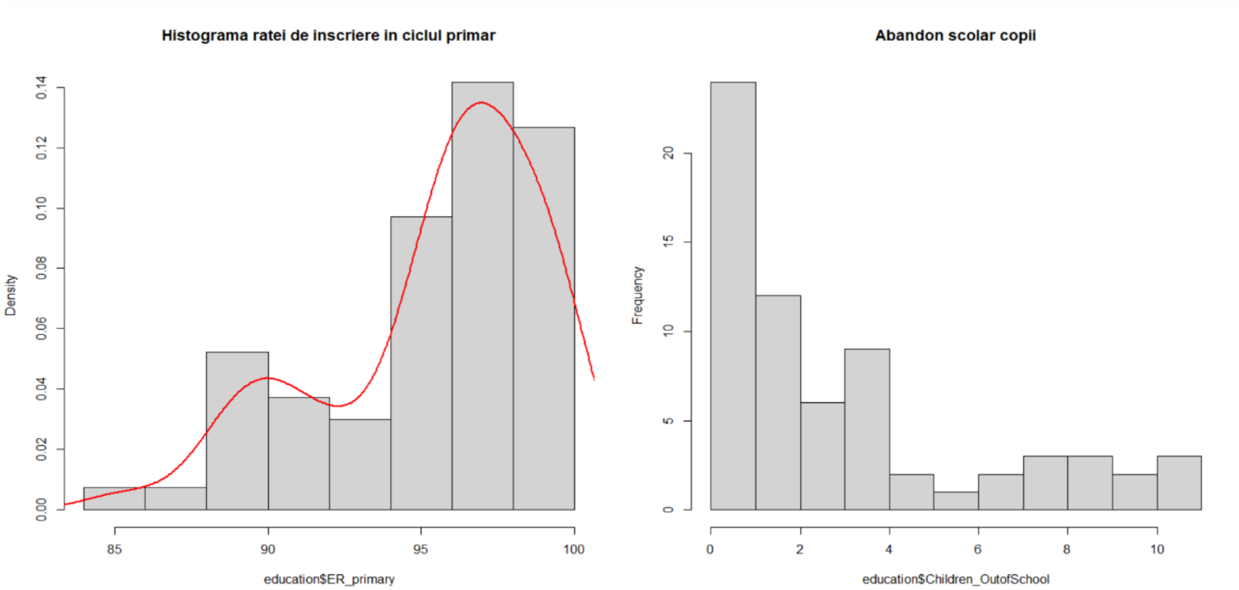


Fig 5. Histograme corespunzatoare ER_primary & Children_OutofSchool

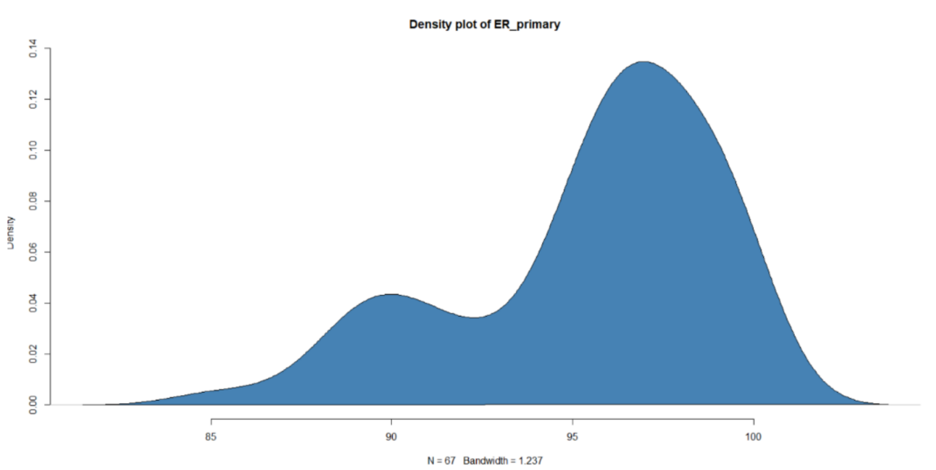


Fig 6. Densitatea de repartiție pentru rata de înscriere în ciclul primar

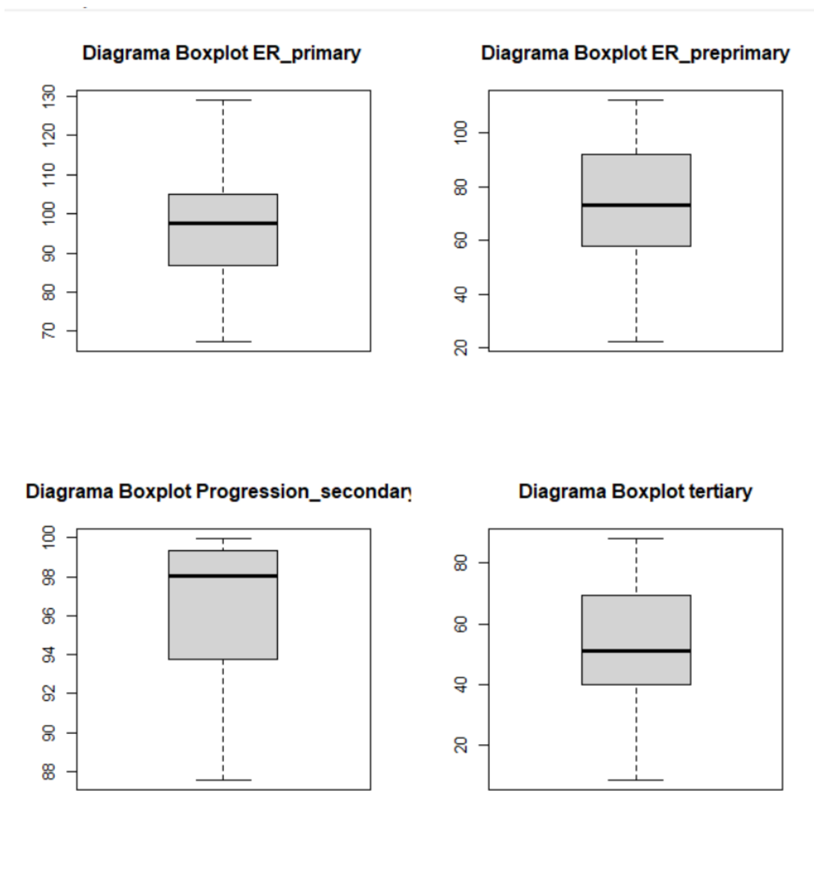


Fig 7. Diagrame Boxplot

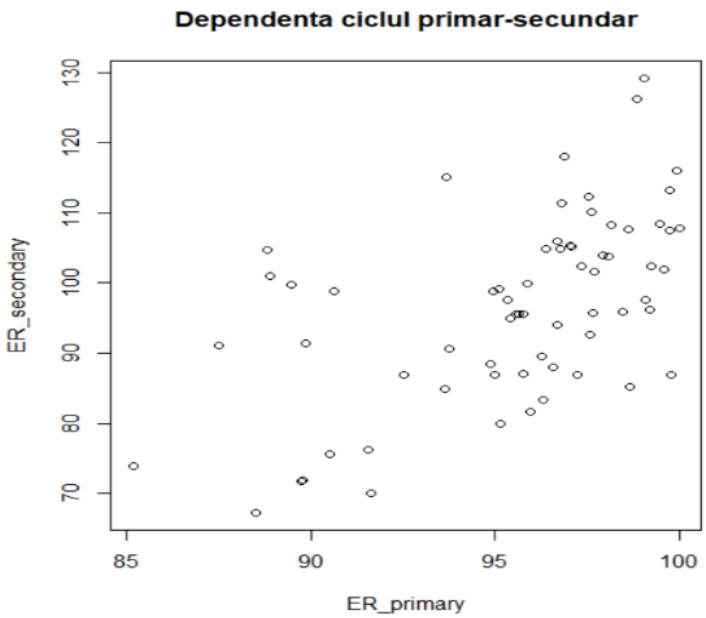


Fig 8. Functia PLOT

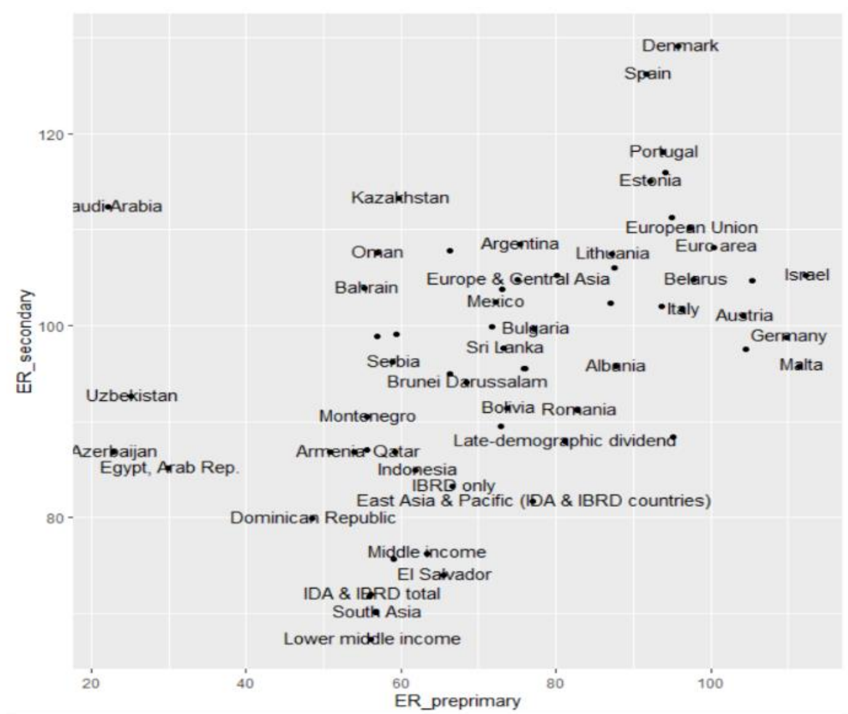


Fig 9. Funcția ggplot

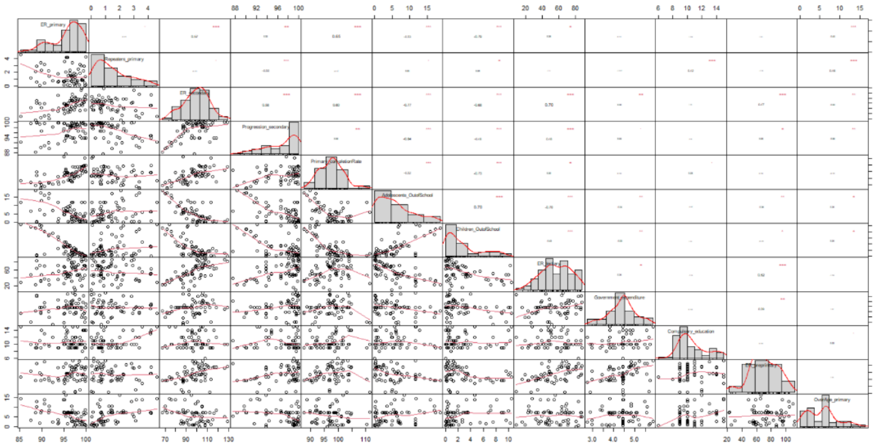


Fig 10. Funcția chart.Correlation

```
> stdev=pca$sdev
> valP=stdev^2
> procent_info=valP*100/12
> procent_cumulat=cumsum(procent_info)
> X=round(data.frame(stdev, valP, procent_info, procent_cumulat),4)
> X
         stdev   valP procent_info procent_cumulat
Comp.1  2.2282 4.9647      41.3722         41.3722
Comp.2  1.3913 1.9356      16.1299         57.5021
Comp.3  1.1852 1.4046      11.7049         69.2070
Comp.4  0.9142 0.8358       6.9651         76.1720
Comp.5  0.8397 0.7050       5.8754         82.0474
Comp.6  0.7915 0.6264       5.2203         87.2677
Comp.7  0.7337 0.5383       4.4857         91.7534
Comp.8  0.5441 0.2960       2.4669         94.2203
Comp.9  0.5103 0.2604       2.1698         96.3901
Comp.10 0.4501 0.2026       1.6886         98.0787
Comp.11 0.3822 0.1461       1.2171         99.2959
Comp.12 0.2907 0.0845       0.7041        100.0000
```

Fig 11. Statistici descriptive

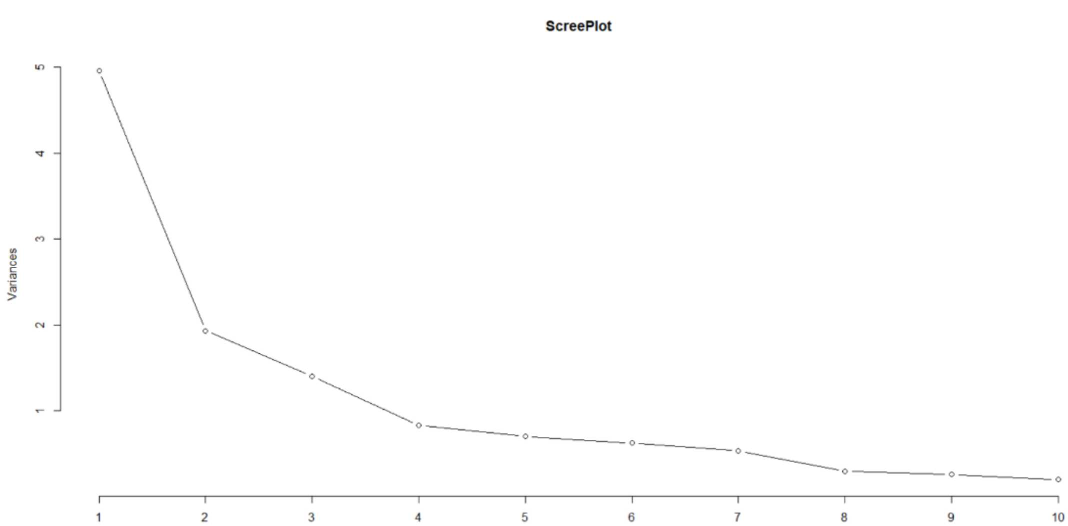


Fig 12. Scree Plot

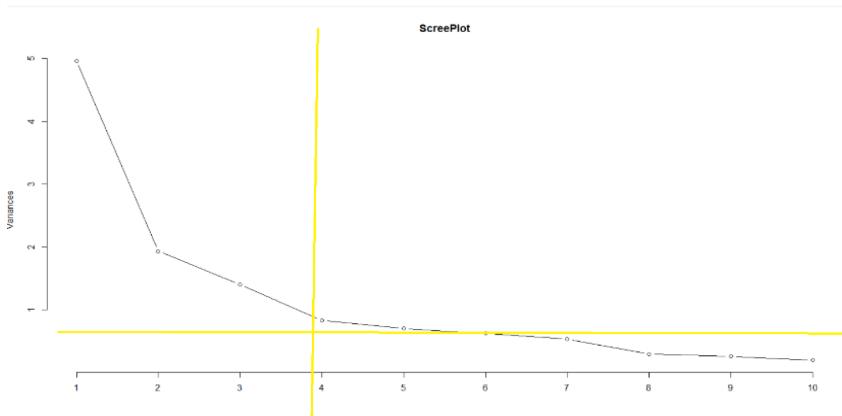


Fig 13. Realizarea tăieturii – criteriul pantei

```
> A = pca$loadings
> round(A,3)

Loadings:
                        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12
ER_primary               0.325  0.162  0.413         0.165  0.275         0.310  0.362   0.423   0.114   0.409
Repeaters_primary       -0.175 -0.477  0.289 -0.203         0.278  0.503  0.262 -0.412          -0.217
ER_secondary             0.401               -0.140 -0.109  0.148        -0.299 -0.386   0.172   0.697  -0.124
Progression_secondary    0.319  0.134 -0.279        -0.269 -0.509  0.175  0.497 -0.314   0.262  -0.130
Primary_completionRate   0.312         0.466               -0.209 -0.376 -0.351 -0.379          -0.390   0.270
Adolescents_OutofSchool -0.393                0.141         0.140 -0.321        -0.242   0.723          -0.318
Children_OutofSchool    -0.386        -0.234        -0.344         0.169 -0.316           0.206           0.708
ER_tertiary              0.311 -0.218 -0.297 -0.465                0.149 -0.307  0.343   0.339  -0.402  -0.196
Government_expenditure   0.173 -0.325 -0.196  0.675  0.484         0.208 -0.244           0.112
Compulsory_education           -0.471  0.357  0.319 -0.586 -0.209                0.338                  -0.178
ER_preprimary            0.202 -0.359 -0.371        -0.171  0.479 -0.530  0.296          -0.104           0.189
OverAge_primary         -0.182 -0.457        -0.353  0.401 -0.474 -0.285  0.184           0.110   0.297   0.158

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12
SS loadings     1.000  1.000  1.000  1.001  1.000  1.000  1.000  1.001  0.998   0.999   0.999   1.000
Proportion Var  0.083  0.083  0.083  0.083  0.083  0.083  0.083  0.083  0.083   0.083   0.083   0.083
Cumulative Var  0.083  0.167  0.250  0.333  0.417  0.500  0.583  0.667  0.750   0.833   0.917   1.000
> |
```

Fig 14. Vectorii proprii ai matricei de covarianță

```
> #verif daca sunt val proprii/vectorii proprii
> eigen(cov(ed_std))$values
 [1] 4.96466135 1.93558817 1.40458561 0.83580839 0.70504335 0.62643555 0.53828568 0.29603200 0.26037473 0.20263223 0.14605598
[12] 0.08449698
> X
         stdev   valP procent_info procent_cumulat
Comp.1  2.2282 4.9647      41.3722         41.3722
Comp.2  1.3913 1.9356      16.1299         57.5021
Comp.3  1.1852 1.4046      11.7049         69.2070
Comp.4  0.9142 0.8358       6.9651         76.1720
Comp.5  0.8397 0.7050       5.8754         82.0474
Comp.6  0.7915 0.6264       5.2203         87.2677
Comp.7  0.7337 0.5383       4.4857         91.7534
Comp.8  0.5441 0.2960       2.4669         94.2203
Comp.9  0.5103 0.2604       2.1698         96.3901
Comp.10 0.4501 0.2026       1.6886         98.0787
Comp.11 0.3822 0.1461       1.2171         99.2959
Comp.12 0.2907 0.0845       0.7041        100.0000
> round(eigen(cov(ed_std))$vectors,3)
        [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]   [,9]  [,10]  [,11]  [,12]
 [1,]  0.325  0.162  0.413 -0.015 -0.165  0.275 -0.089 -0.310  0.362 -0.423  0.114  0.409
 [2,] -0.175 -0.477  0.289  0.203 -0.033  0.278 -0.503 -0.262 -0.412  0.026 -0.217 -0.012
 [3,]  0.401 -0.092 -0.028  0.140  0.109  0.148 -0.083  0.299 -0.386 -0.172  0.697 -0.124
 [4,]  0.319  0.134 -0.279 -0.088  0.269 -0.509 -0.175 -0.497 -0.314 -0.262 -0.130  0.052
 [5,]  0.312 -0.010  0.466  0.078  0.002 -0.209  0.376  0.351 -0.379  0.046 -0.390  0.270
 [6,] -0.393  0.079  0.060 -0.141 -0.029  0.140  0.321  0.014 -0.242 -0.723 -0.095 -0.318
 [7,] -0.386 -0.026 -0.234  0.031  0.344 -0.005 -0.169  0.316 -0.039 -0.206  0.048  0.708
 [8,]  0.311 -0.218 -0.297  0.465  0.030  0.033 -0.149  0.307  0.343 -0.339 -0.402 -0.196
 [9,]  0.173 -0.325 -0.196 -0.675 -0.484 -0.064 -0.208  0.244 -0.037 -0.112 -0.086  0.094
[10,]  0.010 -0.471  0.357 -0.319  0.586 -0.209  0.032  0.030  0.338 -0.061  0.095 -0.178
[11,]  0.202 -0.359 -0.371 -0.082  0.171  0.479  0.530 -0.296 -0.075  0.104 -0.083  0.189
[12,] -0.182 -0.457 -0.012  0.353 -0.401 -0.474  0.285 -0.184  0.085 -0.110  0.297  0.158
```

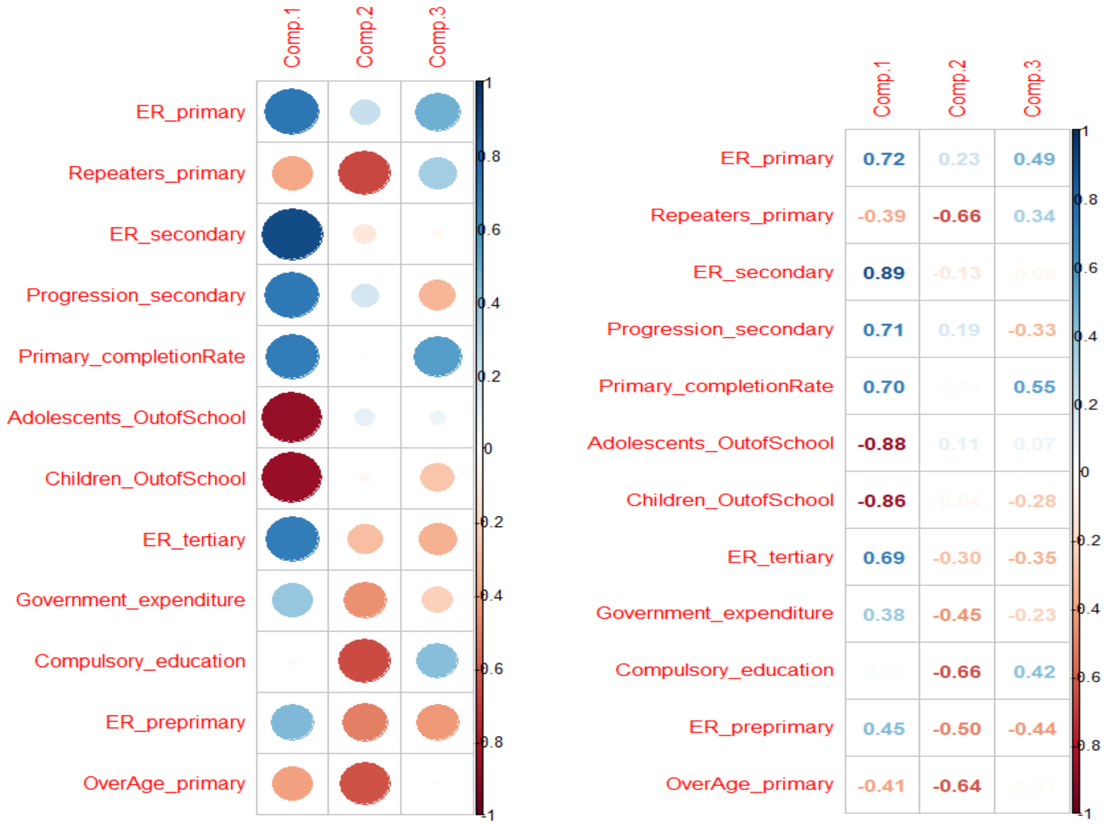
Fig 15. Valorile proprii ale matricei de covarianță

```
> round(eigen(cov(ed_std))$vectors,3)
        [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]   [,9]  [,10]  [,11]  [,12]
 [1,]  0.325  0.162  0.413 -0.015 -0.165  0.275 -0.089 -0.310  0.362 -0.423  0.114  0.409
 [2,] -0.175 -0.477  0.289  0.203 -0.033  0.278 -0.503 -0.262 -0.412  0.026 -0.217 -0.012
 [3,]  0.401 -0.092 -0.028  0.140  0.109  0.148 -0.083  0.299 -0.386 -0.172  0.697 -0.124
 [4,]  0.319  0.134 -0.279 -0.088  0.269 -0.509 -0.175 -0.497 -0.314 -0.262 -0.130  0.052
 [5,]  0.312 -0.010  0.466  0.078  0.002 -0.209  0.376  0.351 -0.379  0.046 -0.390  0.270
 [6,] -0.393  0.079  0.060 -0.141 -0.029  0.140  0.321  0.014 -0.242 -0.723 -0.095 -0.318
 [7,] -0.386 -0.026 -0.234  0.031  0.344 -0.005 -0.169  0.316 -0.039 -0.206  0.048  0.708
 [8,]  0.311 -0.218 -0.297  0.465  0.030  0.033 -0.149  0.307  0.343 -0.339 -0.402 -0.196
 [9,]  0.173 -0.325 -0.196 -0.675 -0.484 -0.064 -0.208  0.244 -0.037 -0.112 -0.086  0.094
[10,]  0.010 -0.471  0.357 -0.319  0.586 -0.209  0.032  0.030  0.338 -0.061  0.095 -0.178
[11,]  0.202 -0.359 -0.371 -0.082  0.171  0.479  0.530 -0.296 -0.075  0.104 -0.083  0.189
[12,] -0.182 -0.457 -0.012  0.353 -0.401 -0.474  0.285 -0.184  0.085 -0.110  0.297  0.158
> write.table(round(A,3))
"Comp.1" "Comp.2" "Comp.3" "Comp.4" "Comp.5" "Comp.6" "Comp.7" "Comp.8" "Comp.9" "Comp.10" "Comp.11" "Comp.12"
"ER_primary" 0.325 0.162 0.413 0.015 0.165 0.275 0.089 0.31 0.362 0.423 0.114 0.409
"Repeaters_primary" -0.175 -0.477 0.289 -0.203 0.033 0.278 0.503 0.262 -0.412 -0.026 -0.217 -0.012
"ER_secondary" 0.401 -0.092 -0.028 -0.14 -0.109 0.148 0.083 -0.299 -0.386 0.172 0.697 -0.124
"Progression_secondary" 0.319 0.134 -0.279 0.088 -0.269 -0.509 0.175 0.497 -0.314 0.262 -0.13 0.052
"Primary_completionRate" 0.312 -0.01 0.466 -0.078 -0.002 -0.209 -0.376 -0.351 -0.379 -0.046 -0.39 0.27
"Adolescents_OutofSchool" -0.393 0.079 0.06 0.141 0.029 0.14 -0.321 -0.014 -0.242 0.723 -0.095 -0.318
"Children_OutofSchool" -0.386 -0.026 -0.234 -0.031 -0.344 -0.005 0.169 -0.316 -0.039 0.206 0.048 0.708
"ER_tertiary" 0.311 -0.218 -0.297 -0.465 -0.03 0.033 0.149 -0.307 0.343 0.339 -0.402 -0.196
"Government_expenditure" 0.173 -0.325 -0.196 0.675 0.484 -0.064 0.208 -0.244 -0.037 0.112 -0.086 0.094
"Compulsory_education" 0.01 -0.471 0.357 0.319 -0.586 -0.209 -0.032 -0.03 0.338 0.061 0.095 -0.178
"ER_preprimary" 0.202 -0.359 -0.371 0.082 -0.171 0.479 -0.53 0.296 -0.075 -0.104 -0.083 0.189
"OverAge_primary" -0.182 -0.457 -0.012 -0.353 0.401 -0.474 -0.285 0.184 0.085 0.11 0.297 0.158
>
```

Fig 16. Vectorii proprii ai matricei de covarianță

```
> C = pca$scores[,1:3]
> rownames(C)=education$`Country Name`
> head(C)
               Comp.1     Comp.2     Comp.3
Albania     0.7959296  0.7239091 -0.3186771
Argentina   2.8970864 -2.0192088  0.7602075
Armenia    -1.8560885  1.6239169 -0.4593074
Austria     1.2733878 -2.8140783 -1.8654266
Azerbaijan -0.2845292  2.5264060  2.6001274
Bulgaria   -0.8185462  0.4895113 -2.6132113
> round(C.3)
```

Fig 17. Scorurile principale pentru primele 6 țări



	Comp.1	Comp.2	Comp.3
ER_primary	0.72337889	0.22564350	0.48982986
Repeaters_primary	-0.38958898	-0.66315393	0.34208638
ER_secondary	0.89248784	-0.12810333	-0.03335722
Progression_secondary	0.71032486	0.18674928	-0.33098153
Primary_completionRate	0.69557334	-0.01347711	0.55185383
Adolescents_OutofSchool	-0.87560826	0.10924500	0.07149181
Children_OutofSchool	-0.86080193	-0.03628574	-0.27735612
ER_tertiary	0.69400518	-0.30323734	-0.35218754
Government_expenditure	0.38483836	-0.45183912	-0.23260433
Compulsory_education	0.02252481	-0.65590463	0.42291422
ER_preprimary	0.44956970	-0.50011288	-0.43952782
OverAge_primary	-0.40577424	-0.63541033	-0.01433929

Fig 18. Matricea Factor

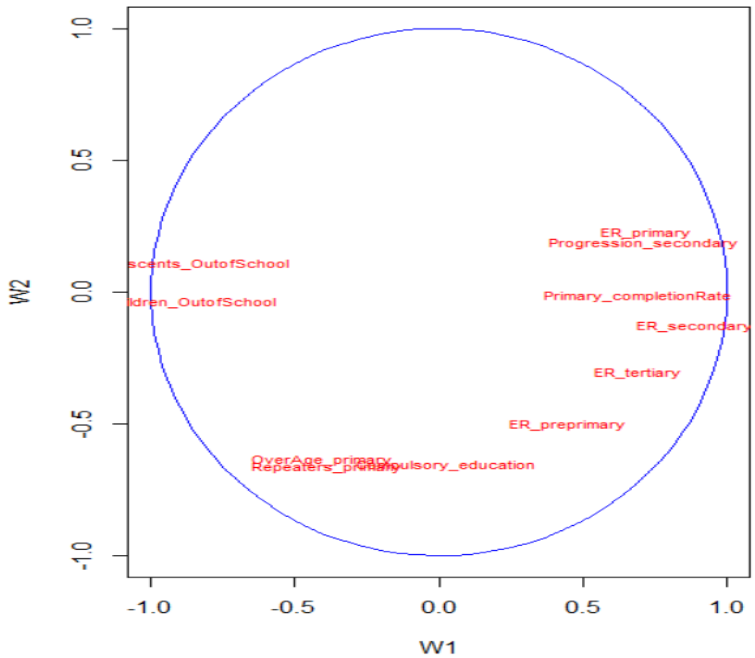


Fig 19. Cercul corelațiilor

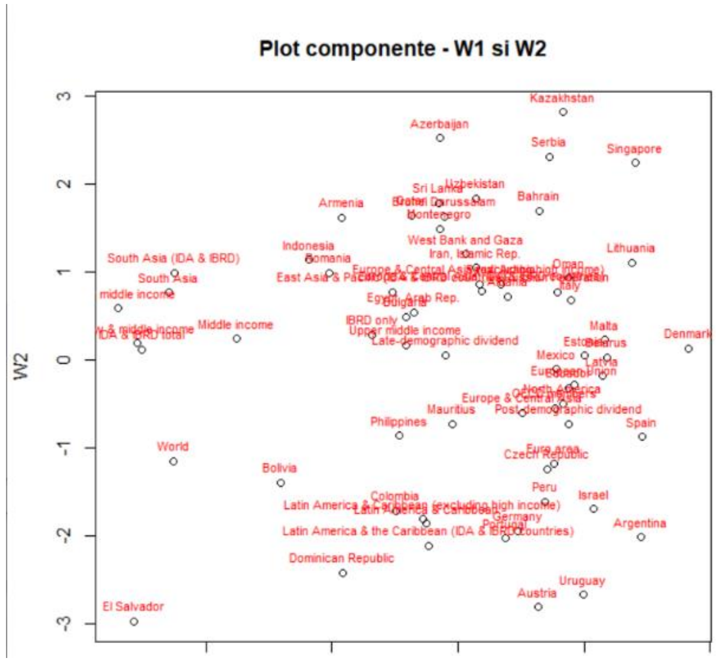


Fig 20. Grafic reprezentarea observațiilor în planul principal W1&W2

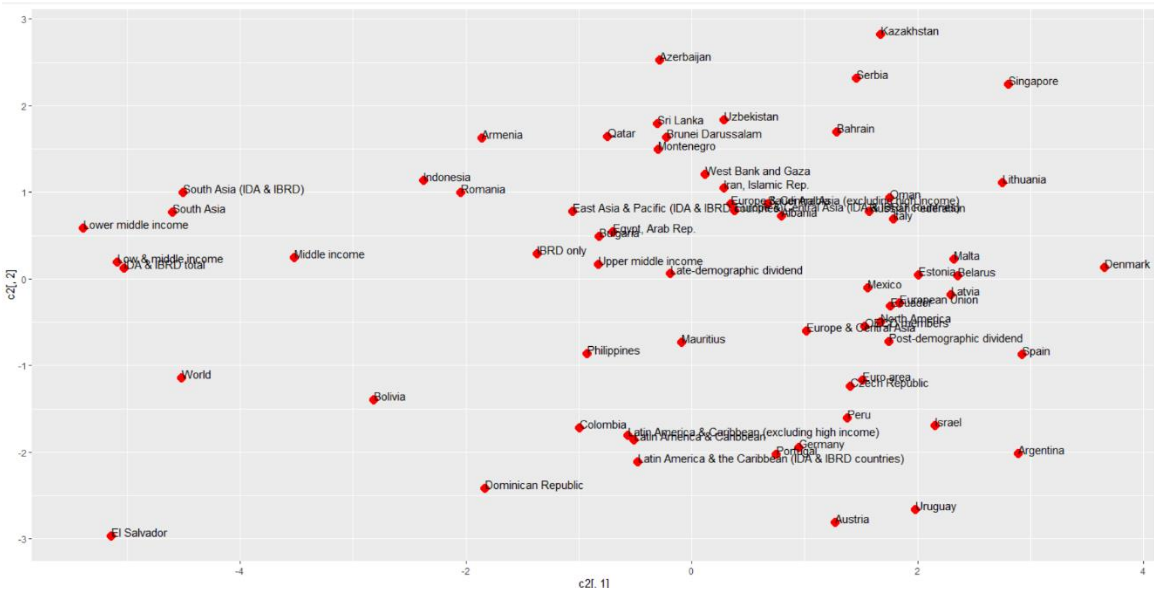


Fig 21. GGplot reprezentarea observațiilor în planul principal

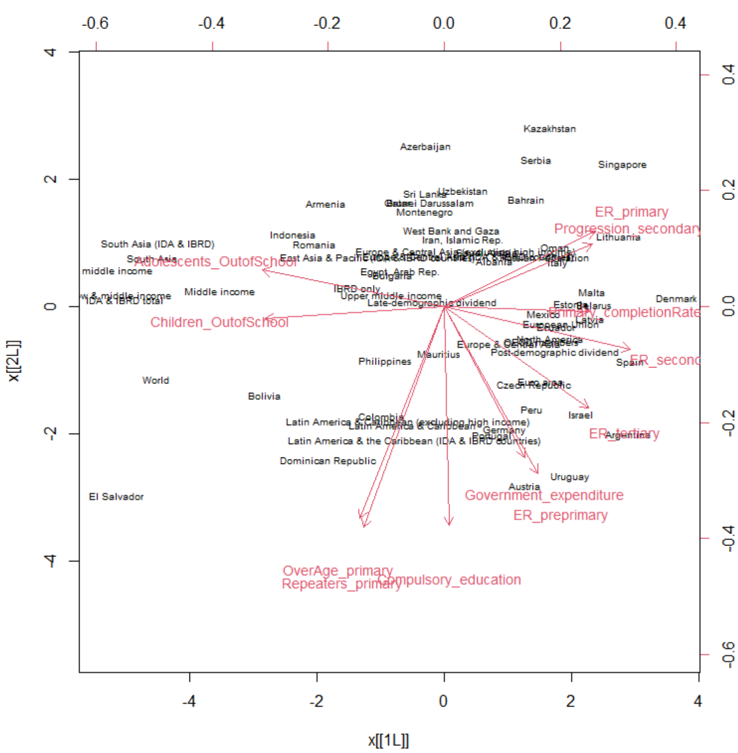


Fig 22. Biplot - reprezentarea observațiilor în planul principal, W1&W2

Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	Dim.10	Dim.11
Variance	4.832	1.647	1.404	0.799	0.672	0.557	0.326	0.265	0.220	0.186	0.093
% of var.	43.929	14.970	12.767	7.266	6.112	5.059	2.963	2.407	1.998	1.687	0.841
Cumulative % of var.	43.929	58.899	71.666	78.932	85.044	90.104	93.067	95.474	97.472	99.159	100.000

Individuals (the 10 first)

	Dist	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
1	1.736	0.808	0.202	0.217	-1.019	0.941	0.344	-0.286	0.087	0.027
2	3.941	3.069	2.909	0.606	1.721	2.686	0.191	0.725	0.559	0.034
3	3.646	-2.093	1.353	0.329	-1.097	1.090	0.090	-0.446	0.211	0.015
4	3.872	1.715	0.908	0.196	1.838	3.062	0.225	-1.886	3.779	0.237
5	4.783	-0.307	0.029	0.004	-3.052	8.443	0.407	2.692	7.705	0.317
6	3.579	-1.024	0.324	0.082	0.097	0.009	0.001	-2.634	7.371	0.542
7	2.187	1.116	0.384	0.260	-1.645	2.453	0.566	0.223	0.053	0.010
8	3.151	2.353	1.711	0.558	-0.280	0.071	0.008	-1.743	3.231	0.306
9	3.908	-2.724	2.291	0.486	1.727	2.705	0.195	-1.030	1.127	0.069
10	2.616	-0.375	0.043	0.021	-1.477	1.978	0.319	-0.396	0.167	0.023

Fig 23. Extragerea componentelor principale folosind funcția PCA – INDIVIDUALS

Variables (the 10 first)

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
ER_primary	0.703	10.226	0.494	-0.234	3.311	0.055	0.496	17.488	0.246
Repeaters_primary	-0.348	2.509	0.121	0.691	29.031	0.478	0.325	7.541	0.106
ER_secondary	0.895	16.587	0.802	0.088	0.472	0.008	-0.035	0.087	0.001
Progression_secondary	0.698	10.076	0.487	-0.236	3.379	0.056	-0.325	7.515	0.106
Primary_completionRate	0.701	10.173	0.492	-0.058	0.203	0.003	0.555	21.973	0.309
Adolescents_OutofSchool	-0.882	16.107	0.778	-0.040	0.095	0.002	0.071	0.359	0.005
Children_OutofSchool	-0.863	15.428	0.746	0.124	0.927	0.015	-0.283	5.686	0.080
ER_tertiary	0.720	10.725	0.518	0.193	2.272	0.037	-0.354	8.927	0.125
Government_expenditure	0.409	3.454	0.167	0.456	12.651	0.208	-0.245	4.264	0.060
Compulsory_education	0.054	0.060	0.003	0.730	32.376	0.533	0.403	11.569	0.162

```
> round(CP$var$cos2,3)
```

Fig 24. Extragerea componentelor principale folosind funcția PCA – VARIABLES

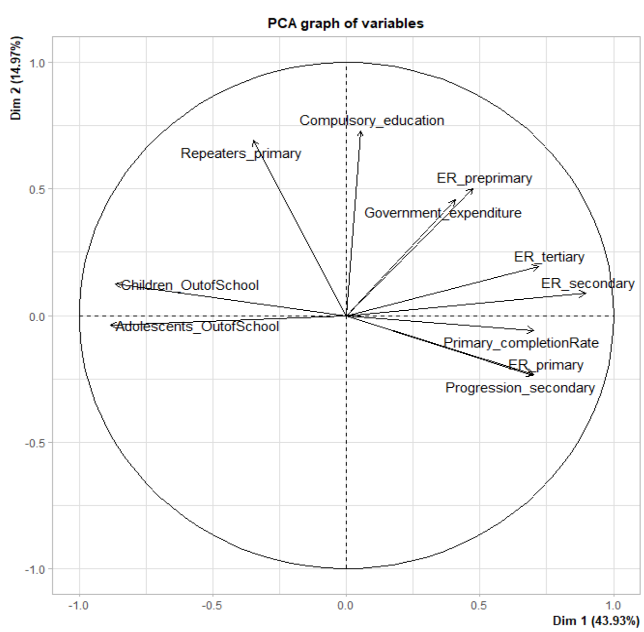


Fig 25. Extragerea componentelor principale folosind funcția PCA

2. *Analiza Cluster*

	Albania	Argentina	Armenia	Austria	Azerbaijan	Bulgaria	Bahrain	Belarus	Bolivia	Brunei Darussalam	Colombia	Czech Republic	Germany
Albania	0.000000	2.4059752	3.445773	3.160319	4.302529	3.910398	1.875260	2.1713272	4.260658	2.672937	2.0245800	3.055020	2.45359
Argentina	2.405975	0.0000000	4.993839	3.422266	4.871872	5.138244	2.554515	2.0403462	6.094561	4.437466	3.5524618	3.770055	3.50014
Armenia	3.445773	4.9938393	0.000000	4.658138	5.009945	2.286009	3.440010	4.4328302	2.031384	3.192325	2.9792018	4.341324	4.15594
Austria	3.160319	3.4222662	4.658138	0.000000	5.847037	4.088931	3.955895	2.2811524	5.154251	4.632819	4.0465338	1.152544	1.28398
Azerbaijan	4.302529	4.8718724	5.009945	5.847037	0.000000	6.273269	3.429149	5.8110888	6.144809	4.567567	3.6554254	5.771742	5.60846
Bulgaria	3.910398	5.1382444	2.286009	4.088931	6.273269	0.000000	4.445939	4.2952344	1.980160	4.461956	3.6749692	3.918695	3.75783
Bahrain	1.875260	2.5545146	3.440010	3.955895	3.429149	4.445939	0.000000	3.0662629	4.769660	2.543710	2.5428702	3.677462	3.54671
Belarus	2.171327	2.0403462	4.432830	2.281152	5.811089	4.295234	3.066263	0.0000000	5.243946	4.066953	3.4937163	2.538544	2.22978
Bolivia	4.260658	6.0945611	2.031384	5.154251	6.144809	1.980160	4.769660	5.2439463	0.000000	3.969395	3.5749459	4.785985	4.45340
Brunei Darussalam	2.672937	4.4374659	3.192325	4.632819	4.567567	4.461956	2.543710	4.0669532	3.969395	0.000000	3.2210531	4.114880	3.67235
Colombia	2.024580	3.5524618	2.979202	4.046534	3.655425	3.674969	2.542870	3.4937163	3.574946	3.221053	0.0000000	4.004787	3.49500
Czech Republic	3.055020	3.7700549	4.341324	1.152544	5.771742	3.918695	3.677462	2.5385442	4.785985	4.114880	4.0047866	0.000000	1.18541
Germany	2.453594	3.5001463	4.155946	1.283983	5.608463	3.757831	3.546717	2.2297804	4.453408	3.672356	3.4950052	1.185411	0.00000
Denmark	3.128009	1.8630021	5.839871	3.656383	5.858412	5.554599	3.368911	2.5021048	6.527471	4.915545	4.3014978	3.787367	3.63781
Dominican Republic	2.920821	4.3082329	2.666924	4.695670	3.630541	4.064413	3.060158	4.1738280	3.766153	3.594713	1.6164894	4.667073	4.29486
Europe & Central Asia (excluding high income)	1.888658	2.6749065	2.463772	3.206494	4.124590	3.185592	1.635120	2.5726736	3.815123	2.716557	2.2956615	3.141044	2.89927
Europe & Central Asia	1.243904	2.0460382	3.232905	2.925251	4.575810	3.622377	1.575664	1.7008739	4.273983	2.792766	2.3687162	2.818623	2.50087
Ecuador	1.548241	2.1370901	4.120261	3.607378	3.511767	4.758170	1.192161	2.8004799	5.196660	3.160988	2.5798427	3.414703	3.23128

Fig 26. Matricea de di(similaritate)

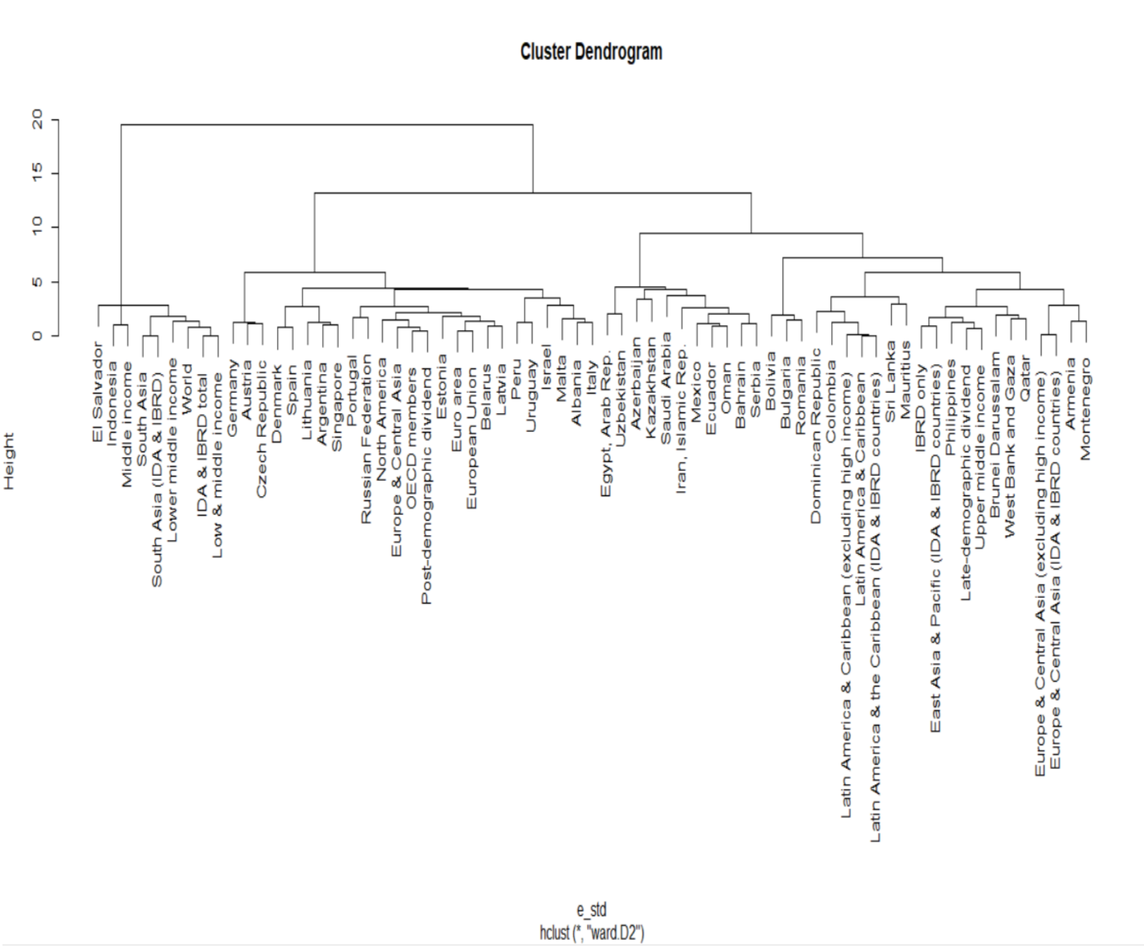


Fig 27. Dendrograma – Arbore al clasificării Metoda WARD

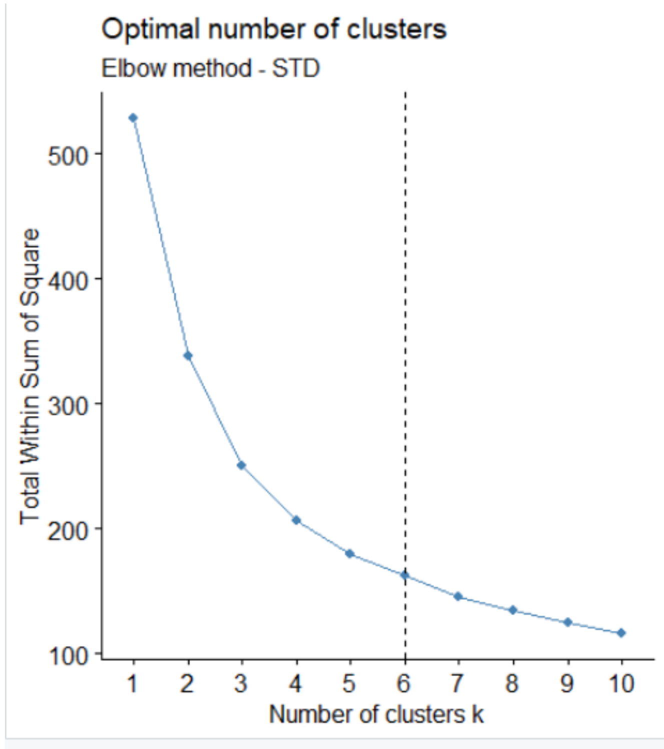


Fig 28. Plot – metoda cotului

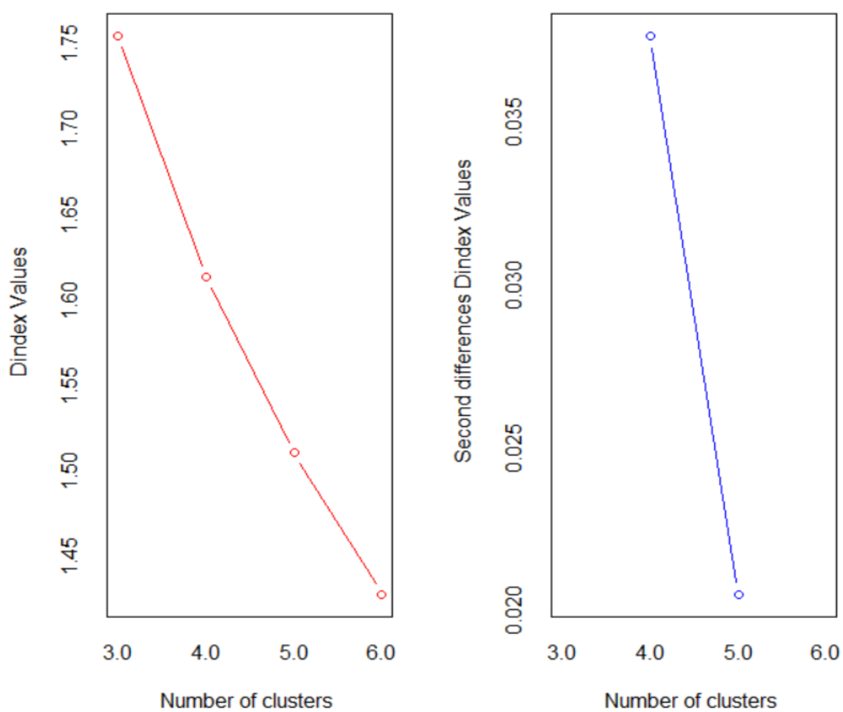
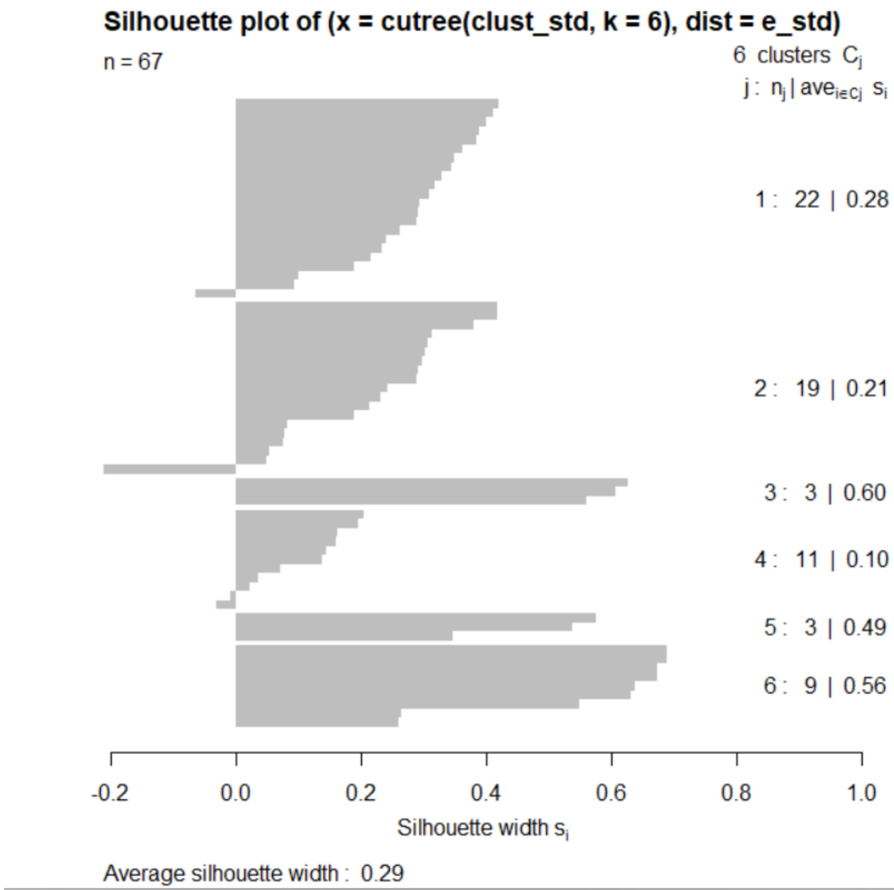


Fig 29. OUTPUT Metoda statisticii GAP


Silhouette plot of (x = cutree(clust_std, k = 6), dist = e_std)
n = 67
6 clusters C_j
j : n_j | $ave_{i \in Cj}$ s_i
1 : 22 | 0.28
2 : 19 | 0.21
3 : 3 | 0.60
4 : 11 | 0.10
5 : 3 | 0.49
6 : 9 | 0.56
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Silhouette width s_i
Average silhouette width : 0.29

	cluster	neighbor	sil_width
[1,]	1	2	0.098593435
[2,]	1	4	0.349056094
[3,]	2	5	-0.211032033
[4,]	3	1	0.606758353
[5,]	4	2	0.194878973
[6,]	5	2	0.537765147
[7,]	4	2	0.160100748
[8,]	1	3	0.216019444
[9,]	5	6	0.346957098
[10,]	2	4	0.230252183
[11,]	2	4	0.241891563
[12,]	3	1	0.625552259
[13,]	3	1	0.560502497
[14,]	1	3	0.328313621
[15,]	2	4	0.213136597
[16,]	2	1	0.082269913
[17,]	1	2	0.262602889
[18,]	4	1	-0.009955624
[19,]	4	2	0.020869231
[20,]	1	3	0.388935627
[21,]	1	3	0.361120943
[22,]	1	3	-0.065238948
[23,]	1	3	0.399427900
[24,]	2	6	0.379359416
[25,]	6	5	0.673430262
[26,]	6	2	0.259141687
[27,]	4	2	0.034215293
[28,]	1	3	0.092434009
[29,]	1	3	0.291113481
[30,]	4	1	0.143431414
[31,]	2	1	0.302469801
[32,]	2	1	0.307037076
[33,]	2	4	0.053091418
[34,]	6	5	0.636998412
[35,]	6	5	0.673037140
[36,]	2	1	0.312459231
[37,]	1	4	0.418718645
[38,]	1	3	0.343383047
[39,]	4	1	0.137910020
[40,]	6	5	0.548074598
[41,]	1	3	0.188804744
[42,]	2	5	0.047804899
[43,]	2	1	0.074327370
[44,]	1	3	0.233825464
[45,]	1	3	0.411708508
[46,]	4	1	0.204933951
[47,]	1	4	0.293703947
[48,]	2	5	0.288015910

Fig 30. Silhouette PLOT & OUTPUT

```
> round(centroizi_std,3)
```

	ER_primary	ER_secondary	Progression_secondary	Primary_completionRate	Adolescents_OutofSchool	Children_OutofSchool	ER_tertiary
1	0.639	0.887	0.422	0.408	-0.772	-0.628	0.948
2	0.056	-0.386	-0.051	-0.244	0.234	0.122	-0.391
3	-1.693	0.384	0.944	0.154	-0.813	-0.847	0.909
4	0.789	0.214	0.259	1.012	-0.336	-0.721	-0.485
5	-1.835	-0.164	0.582	-1.217	0.580	2.282	0.066
6	-1.468	-1.688	-1.749	-1.364	1.880	1.681	-1.224

	ER_preprimary
1	0.804
2	-0.192
3	1.562
4	-1.199
5	0.210
6	-0.685

Fig 31. Centroizii

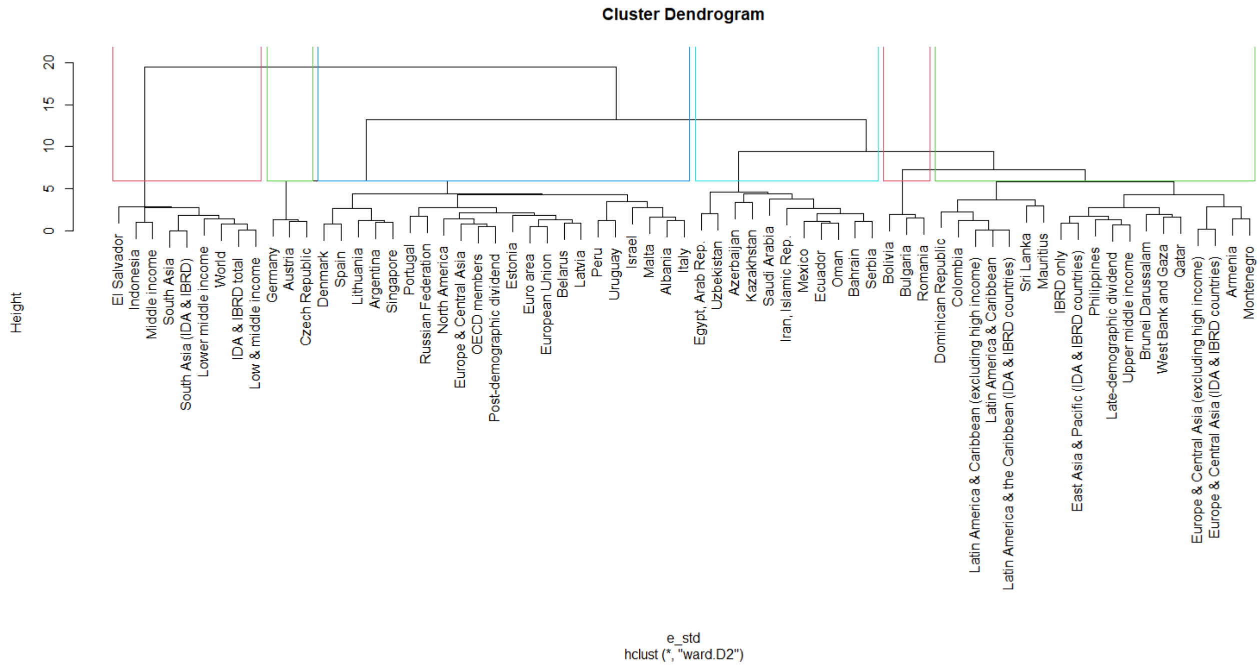


Fig 32. Componența claselor

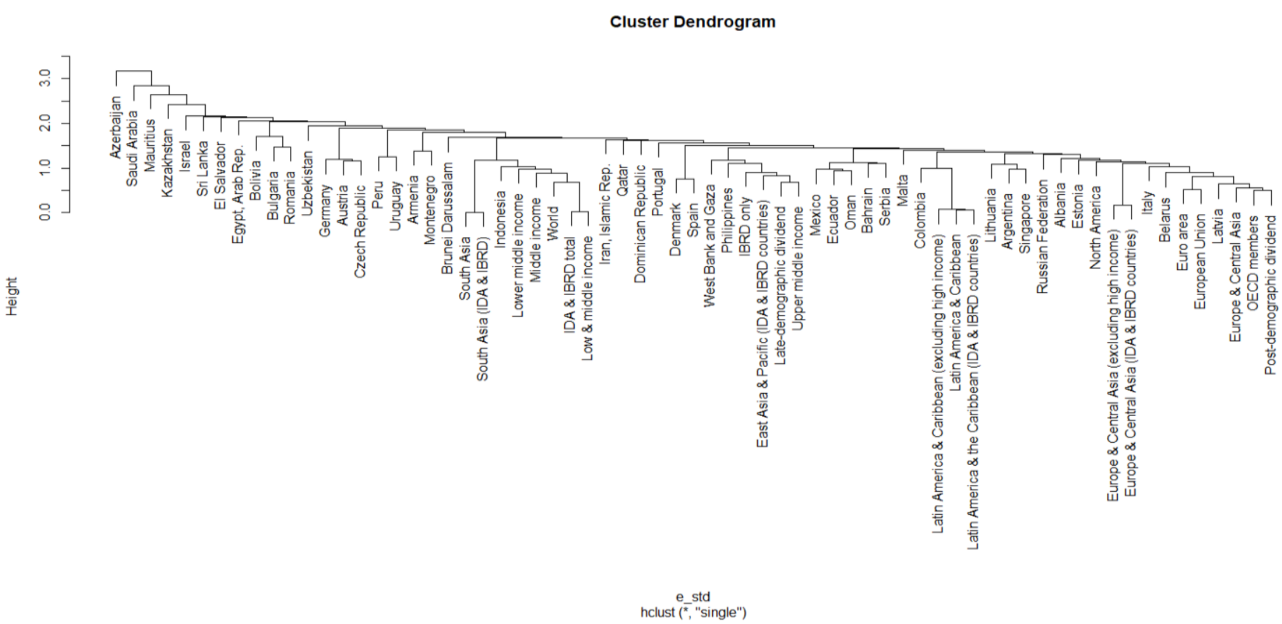


Fig 33. Reprezentarea dendrogramei, cazul unei legături simple

```
Clustering vector:
                                          Albania                                        Argentina
                                                5                                                5
                                          Armenia                                          Austria
                                                4                                                5
                                       Azerbaijan                                         Bulgaria
                                                6                                                4
                                          Bahrain                                          Belarus
                                                6                                                5
                                          Bolivia                                Brunei Darussalam
                                                4                                                1
                                         Colombia                                   Czech Republic
                                                1                                                5
                                          Germany                                          Denmark
                                                5                                                5
                               Dominican Republic       Europe & Central Asia (excluding high income)
                                                1                                                1
                            Europe & Central Asia                                          Ecuador
                                                5                                                6
                                 Egypt, Arab Rep.                                        Euro area
                                                6                                                5
                                            Spain                                          Estonia
                                                5                                                5
                                   European Union                                        IBRD only
                                                5                                                1
                                  IDA & IBRD total                                        Indonesia
                                                3                                                3
                               Iran, Islamic Rep.                                           Israel
                                                6                                                5
                                            Italy                                       Kazakhstan
                                                5                                                6
```

Fig 34. Algoritmul de partiționare K-means

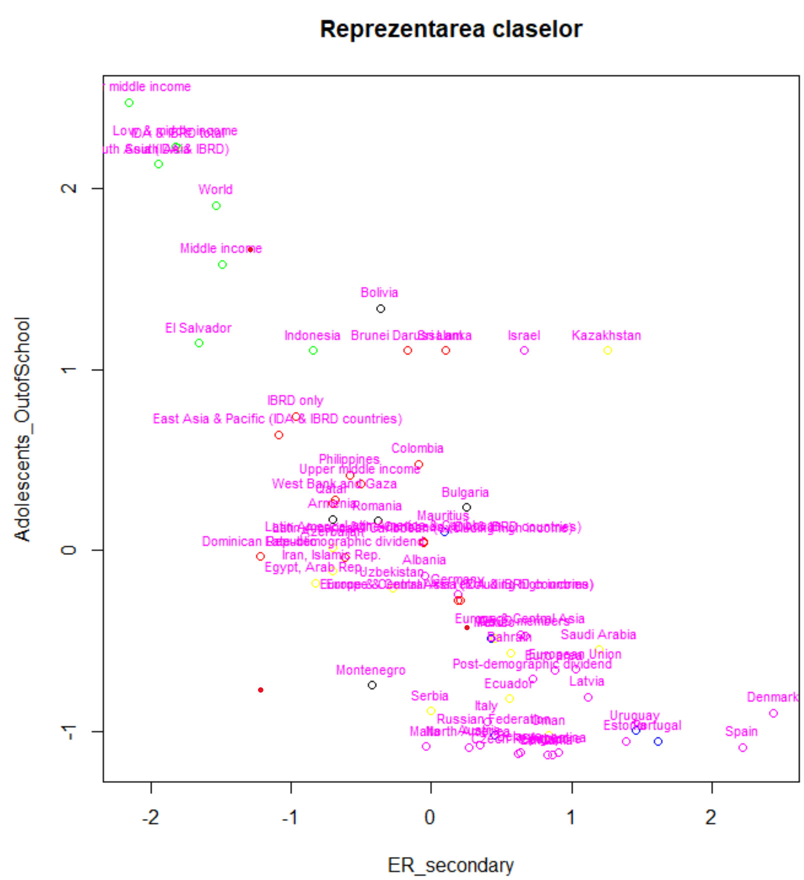


Fig 35. Algoritmul de partiționare K-means – reprezentarea grafică a claselor

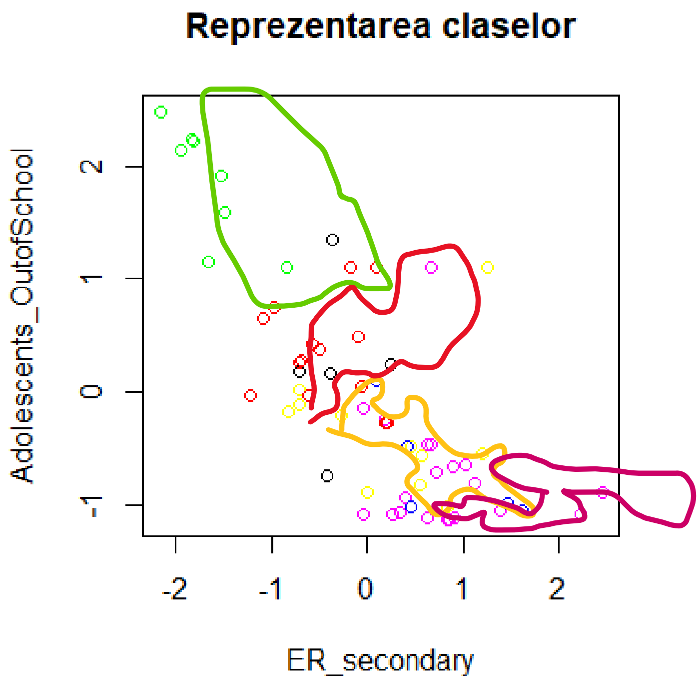


Fig 36. Algoritmul de partiționare K-means – reprezentarea grafică a claselor

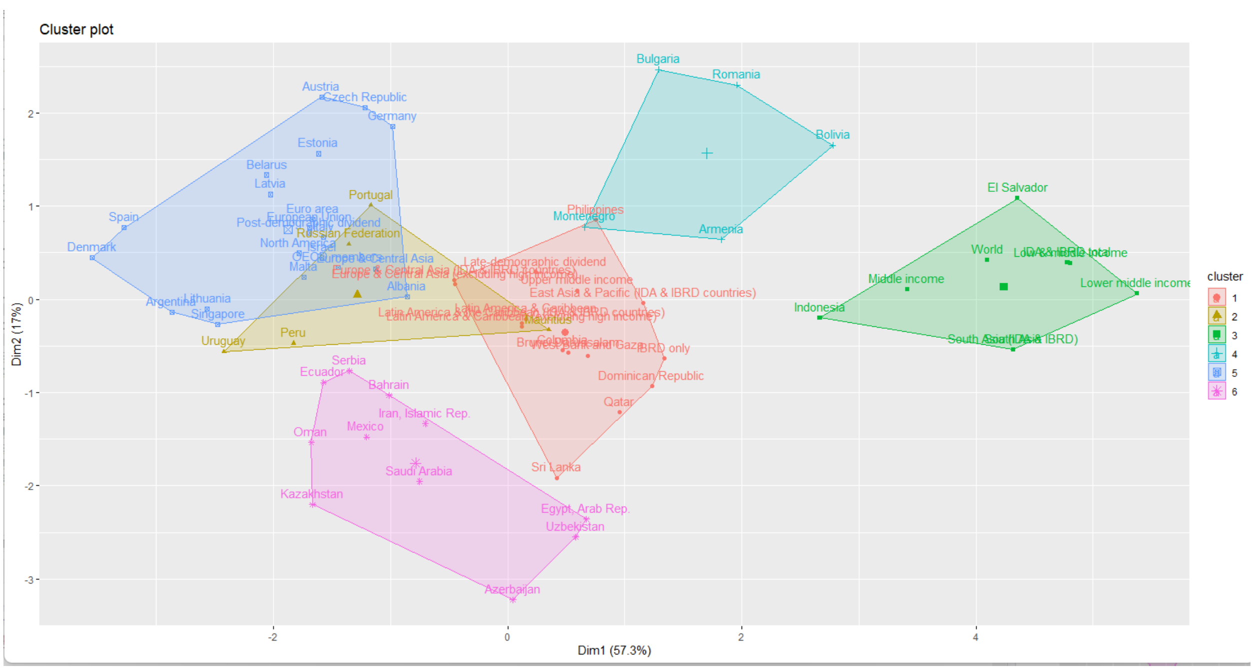


Fig 37. Reprezentarea claselor în funcție de combonentele principale

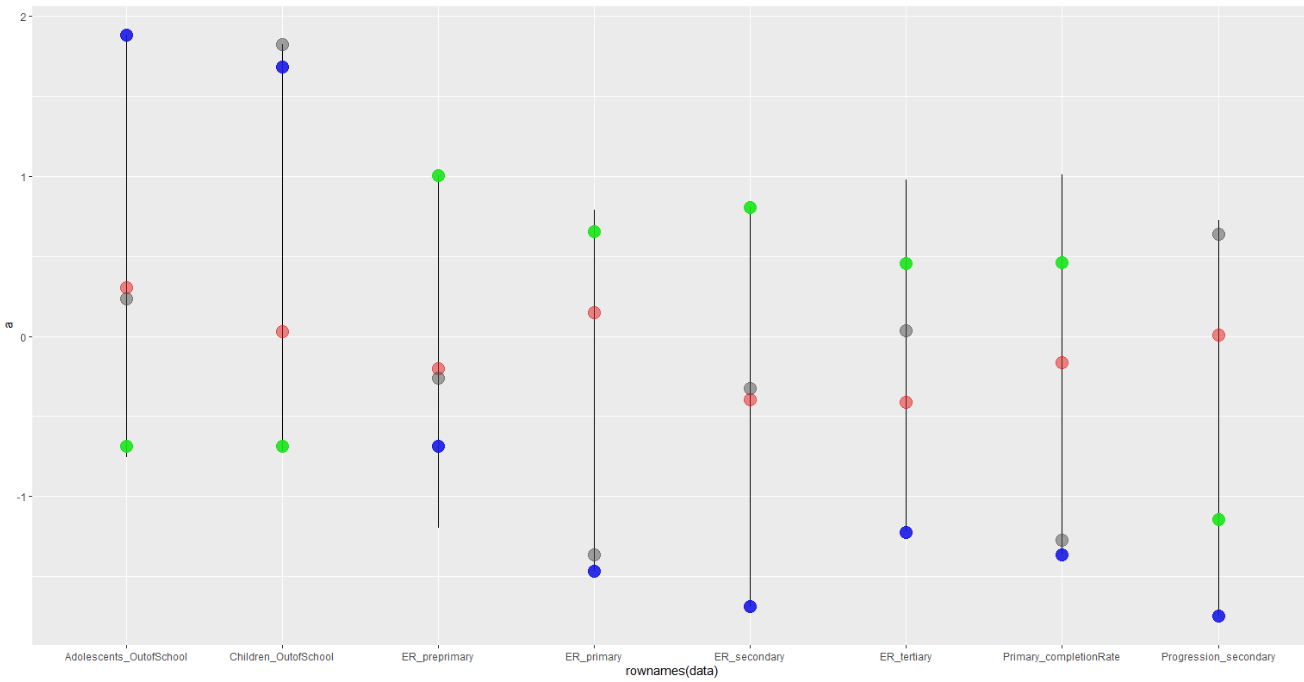


Fig 38. Evaluarea puterii de discriminarei

Anexa 2- SCRIPT R

1. ***Analiza Componentelor Principale***

#1: PRELUCRAREA DATELOR--

```
install.packages("Rcpp")
library(Rcpp)
library(readxl)
education <- read_excel("C:/Users/pc/Desktop/An I, sem I/SDA/education.xlsx")
View(education)
```

#Statistici descriptive--

```
 summary(education)
library(psych)
describe(education)
which.min(education$ER_primary)
which.max(education$ER_primary)

which.min(education$Children_OutofSchool)
which.max(education$Children_OutofSchool)

library(sp)
library(raster)
apply(education[,2:13], 2, cv)

#matricea de corelatie
round(cor(education[,2:13]), 3)

round(cov(education[,2:13]), 3)
```

 # Grafice-------------

```
#Histograma
windows()
```

```
par(mfrow = c(1, 2))
hist(education$ER_primary,freq=FALSE, main="Histograma ratei de inscriere in ciclul primar" )
lines(density(education$ER_primary), lwd = 2, col = 'red')
hist(education$Children_OutofSchool, main="Abandon scolar copii")
?hist
```

#**Densitatea de probabilitate**

```
d<-density(education$ER_primary)
windows()
plot(d, bw = "nrd0", frame = FALSE, col = "steelblue",
     main = "Density plot of ER_primary")
polygon(d, col = "steelblue")
```

#**Boxplot**

```
attach(education)
windows()
par(mfrow=c(2,2))
boxplot(education$ER_secondary, main="Diagrama Boxplot ER_primary")
boxplot(ER_preprimary, main="Diagrama Boxplot ER_preprimary")
boxplot(education$Progression_secondary, main="Diagrama Boxplot Progression_secondary")
boxplot(education$ER_tertiary, main="Diagrama Boxplot tertiary")
```

#**Functia PLOT(abline)**

```
plot(ER_primary, ER_secondary, main="Dependenta ciclul primar-secundar")
abline(lm(ER_primary~ER_secondary))
```

#ggplot

```
library(ggplot2)
windows()
ggplot(education, aes(x=ER_preprimary, y=ER_secondary)) +
  geom_point() + # Show dots
```

```
  geom_text(
    label=`Country Name`,
    nudge_x = 0.25, nudge_y = 0.25,
    check_overlap = T
  )
```

#CORELATII

```
round(cor(firme[,2:12]),3)
library(Hmisc)
M=rcorr(as.matrix(firme[,2:12]))
M #probabilitatile p-value coresp coef de semnificatie
#Daca p-value < 0.05 respingem ipoteza nula <=> nu avem o relatie de corelatiw semnificativa intre cele 2 vb

library(corrplot)
P=rcorr(as.matrix(round(education[,2:13]),3))
P
windows()
corrplot(P$r, type="lower", method="square")

windows()
corrplot(P$r, type="lower", method="number", p.mat=P$P,
         sig.level = 0.05, insig="blank")

library("PerformanceAnalytics")
chart.Correlation(education[,2:13], histogram=TRUE,pch=19)
```

 # Standardizarea Datelor---------

```
ed_std<- scale(education[,2:12], scale = TRUE)
apply(ed_std, 2, sd)
round(apply(ed_std, 2, mean), 3)
```

#Verificarea datelor standardizate

apply(ed_std, 2, sd) #1 OK

round(apply(ed_std, 2, mean), 3) #0 OK

round(cor(ed_std), 3)

round(cov(ed_std), 3) # matricea de corelatie = matricea de covarianta pt variabile standardizate (0,1) ok

#2. Funcția princomp pentru extragerea componentelor principale

pca=princomp(ed_std, cor = TRUE)

summary(pca)

stdev=pca$sdev

valP=stdev^2

procent_info=valP*100/12

procent_cumulat=cumsum(procent_info)

X=round(data.frame(stdev, valP, procent_info, procent_cumulat),4)

X

sum(valP)

#Scree Plot

windows()

scree_plot=prcomp(ed_std)

plot(scree_plot, type="l", main="ScreePlot")

#Vectorii proprii ai matricei de cov

A = pca$loadings

round(A,3)

write.table(round(A,3))

#verif daca sunt val proprii/vectorii proprii

eigen(cov(ed_std))$values

X

round(eigen(cov(ed_std))$vectors,3)

write.table(round(A,3))

#Scorurile principale

C = pca$scores[,1:3]

rownames(C)=education$`Country Name`

head(C) #primele 3 scoruri principale pt primele 6 tari

round(C,3)

View(ed_std)

ed_std

#Proprietățile CP

#1)verificam varianta totala si varianta generalizata

#varianta totala:

sum(valP) #=12 corect

cov(ed_std) # pe diag princ sa fie 1, suma tuturor sigmelor = 12 #ACEASTA ESTE **MATRICEA SIGMA**

round(cor(C),5)

valP

#2) Sunt corelate 2 cate 2

round(cor(C),6) #corelaatii = 0 intre doua cate 2 comp

#Matricea FACTOR

matrice_factor=cor(ed_std,C)

matrice_factor

library(corrplot)

corrplot(matrice_factor, method="number")

#Cercul corelațiilor

dev.new()

cerc = seq(0,2*pi,length=100)

plot(cos(cerc),sin(cerc),type="l",col="blue",xlab="W1",ylab="W2")

text(matrice_factor[,1],matrice_factor[,2],rownames(matrice_factor),col="red",cex=0.7)

#Reprezentarea observatiilor in planul principal

```
c2=data.frame(C)

dev.new()

plot(c2[,1],c2[,2],main="Plot componente - W1 si W2",xlab="W1",ylab="W2")

text(c2[,1],c2[,2],labels=rownames(c2),col="red",pos=3,cex=0.7)

dev.new()

plot(c2[,2],c2[,3],main="Plot componente - W2 si W3",xlab="W2",ylab="W3")

text(c2[,2],c2[,3],labels=rownames(c2),col="red",pos=3,cex=0.7)

install.packages("ggplot2")

library(ggplot2)

windows()

ggplot(c2,aes(x=c2[,1],y=c2[,2]))+ geom_point(shape=16,size=4,col="red") +
geom_text(label=rownames(c2),vjust=0,hjust=0,size=4)

windows()

ggplot(c2,aes(x=c2[,1],y=c2[,3]))+ geom_point(shape=16,size=4,col="red") +
geom_text(label=rownames(c2),vjust=0,hjust=0,size=4)
```

#Biplot

```
biplot(c2[,1:2], pca$loadings[,1:2], cex=c(0.6,0.9))

biplot(c2[,2:3], pca$loadings[,2:3], cex=c(0.6,0.9))
```

##12) extragerea comp principale folosind functia PCA

```
library(FactoMineR)


CP=PCA(ed_std)

summary(CP)

CP

round(CP$var$cos2,3)

X
```

#13) grafice - pachetul factoextra

```
library(factoextra)
fviz_pca_ind(pca,
        col.ind = "cos2",
        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
        repel = TRUE
)
fviz_pca_var(pca,
        col.var = "contrib",
        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
        repel = TRUE
)
fviz_pca_biplot(pca,
          repel=TRUE,
          col.var = "#2E9FDF", col.ind = "#696969")
```

2. *Analiza CLUSTER*

```
library(readxl)
ed <- read_excel("ed.xlsx")
View(ed)

educatie <- ed[,2:9]
educatie_std <- scale(educatie,scale=TRUE)
rownames(educatie_std)=ed$`Country Name`

#Matricea Distantelor
e_std = dist(as.matrix(educatie_std), method = "euclidian") #method=canverra, cebasev
View(as.matrix(e_std))
e_std[1] #=dist(Argentina, Albania) = 2.4059

d_euclid_1=sqrt((educatie_std[1,1]-educatie_std[2,1])^2+
          (educatie_std[1,2]-educatie_std[2,2])^2+
          (educatie_std[1,3]-educatie_std[2,3])^2+
          (educatie_std[1,4]-educatie_std[2,4])^2+
          (educatie_std[1,5]-educatie_std[2,5])^2+
          (educatie_std[1,6]-educatie_std[2,6])^2+
          (educatie_std[1,7]-educatie_std[2,7])^2+
          (educatie_std[1,8]-educatie_std[2,8])^2) #=2.4059

e_std[3] #=dist(Austria, Albania) = 3.16

d_euclid_2=sqrt((educatie_std[1,1]-educatie_std[4,1])^2+
        (educatie_std[1,2]-educatie_std[4,2])^2+
        (educatie_std[1,3]-educatie_std[4,3])^2+
        (educatie_std[1,4]-educatie_std[4,4])^2+
        (educatie_std[1,5]-educatie_std[4,5])^2+
        (educatie_std[1,6]-educatie_std[4,6])^2+
```

```
            (educatie_std[1,7]-educatie_std[4,7])^2+
            (educatie_std[1,8]-educatie_std[4,8])^2) #=3.16
#Am preluat din matricea initiala cu datele standardizate educatie_std si am aplicat formula SQRT(()^2)

library(ggplot2)
library(reshape2)
M <- melt(as.matrix(e_std))
windows()
ggplot(data = M, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() +
  theme(axis.text.x = element_text(angle = 45)) +
  scale_fill_gradient(low="white", high="black")
View(as.matrix(e_std))



#metoda ward
clust_std = hclust(e_std, method = "ward.D2")
cbind(clust_std$merge,clust_std$height)
windows()
plot(clust_std,labels=rownames(educatie_std))

#CRT 2
library(factoextra)
fviz_nbclust(educatie_std, hcut, method = "wss") +
  geom_vline(xintercept = 6, linetype = 2)+
  labs(subtitle = "Elbow method - STD")

#AL 3-LEA CRTERIU
install.packages("NbClust") #o met cu ft ft multe crt adunate >20
library(NbClust)
windows()
res<-NbClust(educatie_std, distance = "euclidean", min.nc=3, max.nc=6,
```

```
    method = "ward.D2", index = "all")

#++++++> vom analiza in continuare 6 clase<-CONCLUZIE

install.packages("cluster")
library(cluster)
si4_std <- silhouette(cutree(clust_std, k = 6), e_std)
windows()
plot(si4_std, cex.names = 0.5)
si4_std

library(MASS)
centroizi_std <- tapply(as.matrix(educatie_std), list(rep(cutree(clust_std, 6), ncol(educatie_std)), col(educatie_std)), mean)
colnames(centroizi_std)=colnames(educatie_std)
round(centroizi_std,3)
centroizi_std
plot(clust_std,labels=rownames(educatie_std))
rect.hclust(clust_std,k=6, border=2:5)

#Agegare simpla - metoda celor mai apropiati vecini
clust2_std = hclust(e_std, method = "single")
windows()
plot(clust2_std,labels=rownames(educatie_std))

#Algoritmul K-means
k_std=kmeans(educatie_std,6) #K=6 clase
k_std
clasa_std=k_std$cluster
c_std=cbind(clasa_std,round(educatie_std,6))
c_std
m_std=data.frame(c_std)
```

```
plot(m_std[,3], m_std[,6], col=c("red","blue","green","black","magenta","yellow")
     [m_std$clasa_std], main="Reprezentarea claselor", xlab=colnames(m_std[3]), ylab=colnames(m_std[6]))
text(m_std[,3],m_std[,6],labels=rownames(m_std),col="magenta",pos=3,cex=0.7)

library(factoextra)
fviz_cluster(list(data = educatie_std, cluster = clasa_std))

#Descompunerea variabilitatii----------------------------------
spat_std=k_std$totss
spaw_std=k_std$tot.withinss
spab_std=k_std$betweenss
r_cls_std=spab_std/spaw_std
variab_std=cbind(spat_std,spaw_std,spab_std,r_cls_std)
variab_std
#spat_std spaw_std spab_std r_cls_std
#[1,]      528 163.4311 364.5689   2.23072
k_std$withinss
sum(34.541061+14.224524+11.484663+9.956026 + 51.436683+41.788094) # = spaw = 163.4311
sum(17.69180+55.29219+30.75771+9.81517) # = spaw

#Evaluarea puterii de discriminare a variabilelor
library(psych)
library(ggplot2)

medii_std=data.frame(round(k_std$centers,3))
round(k_std$centers,3)
describe(medii_std) #preiau valoarea min si max
a=describe(medii_std)$min
b=describe(medii_std)$max
c1=t(medii_std[1,]) #centroizii in fiecare clasa
c2=t(medii_std[2,])
c3=t(medii_std[3,])
```

```
c4=t(medii_std[4,])

c5=t(medii_std[5,])

c6=t(medii_std[6,])

data=data.frame(c1,c2,c3,c4, c5, c6,a,b)

rownames(data)=colnames(medii_std)

data #X1-X4 centroizii claselor transpusi


ggplot(data) +

  geom_segment( aes(x=rownames(data), xend=rownames(data), y=a, yend=b), color="black") +

  geom_point( aes(x=rownames(data), y=c1), color=rgb(0.9,0.1,0.1,0.5), size=5 ) +

  geom_point( aes(x=rownames(data), y=c2), color=rgb(0.1,0.9,0.1,0.9), size=5 ) +

  geom_point( aes(x=rownames(data), y=c3), color=rgb(0.1,0.1,0.9,0.9), size=5 ) +

  geom_point( aes(x=rownames(data), y=c4), color=rgb(0.3,0.3,0.3,0.5), size=5 )

#cu cat "batul" este mai mare, cu atat puterea de discriminare este mai mare


install.packages("DiscriMiner")

library(DiscriMiner)


rez=rbind(round(discPower(c_std[,2:9], c_std[,1])$F_statistic,6),round(discPower(c_std[,2:9],
c_std[,1])$p_value,6))

colnames(rez)=colnames(c_std[,2:9])

rownames(rez)=c("F_statistic","p-value")

rez


ward_std=rep(cutree(clust_std,6))

kmeans_std=c_std[,1]


solutii=cbind(kmeans_std,ward_std)

solutii


r1 <-ifelse(kmeans_std==1,"
HIGHEST_ERtertiary",ifelse(kmeans_std==2,"mediumHigh",ifelse(kmeans_std==3,"LOWEST _ERprimary",
```

```
ifelse(kmeans_std==4,"Mediumlow", ifelse(kmeans_std==5, "HIGHEST_ChildrenOut", "low"))))) #!!! #am decodificat ce inseamna 1,2,3,4 pt k-means

r2 <- ifelse(ward_std==1,"HIGHEST_ERtertiary",ifelse(ward_std==2,"mediumHigh",ifelse(ward_std==3,"LOWEST_ERprimary", ifelse(ward_std==4,"Mediumlow", ifelse(ward_std==5, "HIGHEST_ChildrenOut", "low"))))) #!!! #la fel si pentru ward

solutii2=cbind(r1,r2)

solutii2

r1

r2

#Kmeans

k=kmeans(educatie_std,6) #16, 5, 9, 5, 21, 11

k

cls=k$cluster

cls=as.numeric(cls)

cls

set_date=cbind(educatie_std,cls)

df2=data.frame(set_date)

round(df2,3)

nr = round(nrow(df2)*.70) #70% din setul de date

nr #47

a <- sample(seq_len(nrow(df2)), size = nr) #un sample din setul de date total = 24 = nr

train <- df2[a, ] #setul de antrenare

test <- df2[-a, ]#setul de testare

round(train,3)

round(test,3)

df=data.frame(train)

df$cls[df$cls ==1] <- "clasa1"

df$cls[df$cls ==2] <- "clasa2"

cbind(round(df[,1:9],3),df[,9])


#K-NN = cel mai apropiat k vecin, k=sqrt(T), T=47

sqrt(47) #6<6.855<7
```

```
library(class)

pr <- knn(train[,-10],test[,-10],cl=train[,9],k=6) #am obtinut cls de apartenenta obtinuta pe baza clasificatorului pt setul de testare

pr

pr2 <- knn(train[,-10],test[,-10],cl=train[,10],k=7)

pr2

c1 <- table(pr,test[,9])

c1 #matrice de confuzie - confusion matrix - pe baza careia vom calcula gradul de clasificare corect(acuratetea modelului)


c2 <- table(pr2,test[,9])

c2

acc <- function(x){sum(diag(x)/(sum(rowSums(x)))) * 100} #acc calculeaza gradul de clasificare corect.

acc(c1) #95

acc(c2) #75.43


#Naive Bayes

install.packages("e1071", dependencies = TRUE)

install.packages("caTools", dependencies = TRUE)

install.packages("caret", dependencies = TRUE)


library(e1071)

library(caTools)

library(caret)

install.packages('tidyverse')

library(tidyverse)

install.packages('ggplot2')

library(ggplot2)

install.packages('caret')

library(caret)

install.packages('caretEnsemble')
```

```
library(caretEnsemble)
install.packages('psych')
library(psych)
install.packages('Amelia')
library(Amelia)
install.packages('mice')
library(mice)
install.packages('GGally')
library(GGally)
install.packages('rpart')
library(rpart)
install.packages('randomForest')
library(randomForest)



View(ed)
ed<-ed[1:20,]
ed<-data.frame(ed)
split <- sample.split(ed, SplitRatio = 0.75)
split
train_cl <- subset(ed, split == "TRUE")
rlang::last_error()
head(train_cl)
test_cl <- subset(ed, split == "FALSE")
head(test_cl)

train_scale <- scale(test_cl[,2:9])
head(train_scale)

test_scale <- scale(test_cl[,2:9])

set.seed(2020)
```

```
classifier_cl <- naiveBayes(Adolescents_OutofSchool ~ ., data = ed, laplace =  1)
classifier_cl

y_pred <- predict(classifier_cl, newdata = test_cl)
y_pred
u<-union(test_cl$Adolescents_OutofSchool, y_pred)
cm <- table(factor(test_cl$Adolescents_OutofSchool, u),factor(y_pred, u))

confusionMatrix(cm)
```