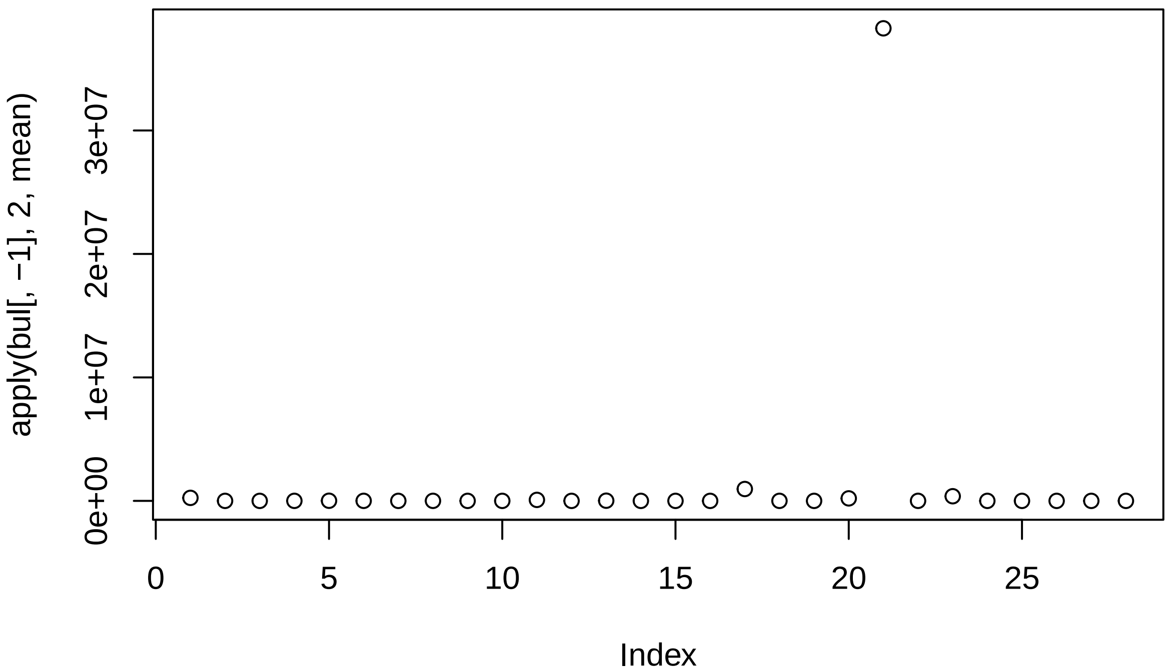
Analysis Bulgaria

```r
bul <- readxl::read_xlsx("C:/Users/User/Documents/UNITN/Computational social science/bulgaria/final_CIT

bul$activity_rate <- as.double(bul$activity_rate)
bul$labour_force_thousands <- as.double(bul$labour_force_thousands)
bul$employment_rate <- as.double(bul$employment_rate)
bul$foreign_direct_investment_euro <- as.double(bul$foreign_direct_investment_euro)

apply(bul[,-1], 2, mean)
```

```
##                    Population2021                     fertility_rate
##                      2.442478e+05                       1.618992e+00
##               unis_colleges_number                      teaching_staff
##                      1.928571e+00                       8.677089e+02
##                  students_enrolled                      graduates_uni
##                      9.206643e+03                       1.919709e+03
##           ratio_grad_enrolled_unis                       TotalSchools
##                      2.123366e-01                       5.880000e+02
##                        gradtotot57                       gradtotot812
##                      3.041211e-01                       2.057127e-01
##        employees_laborcontract_avg  employees_laborcontract_avg_topop
##                      7.899189e+04                       2.633891e-01
##                            avgwage                      activity_rate
##                      1.344071e+04                       7.038571e+01
##             labour_force_thousands                    employment_rate
##                      1.126893e+02                       6.527143e+01
##     foreign_direct_investment_euro                    books_pamphlets
##                      9.601554e+05                       3.965956e+02
##       accommodation_establishments              arrivals_accommodation
##                      1.191071e+02                       2.017012e+05
##                 revenue_accomm_lev                             crimes
##                      3.827540e+07                       1.014286e+03
##                 ecological_assets                   water_supply2020
##                      3.738583e+05                       9.903214e+01
## connected_to_wastewater_collecting             Pop_watersupplyregime
##                      6.878750e+01                       6.750714e+00
##                 waste_thousandtons                      hospitals2021
##                      1.009150e+02                       1.146429e+01
```

```r
plot(apply(bul[,-1], 2, mean))
```



Taking a look at the correlation plot.

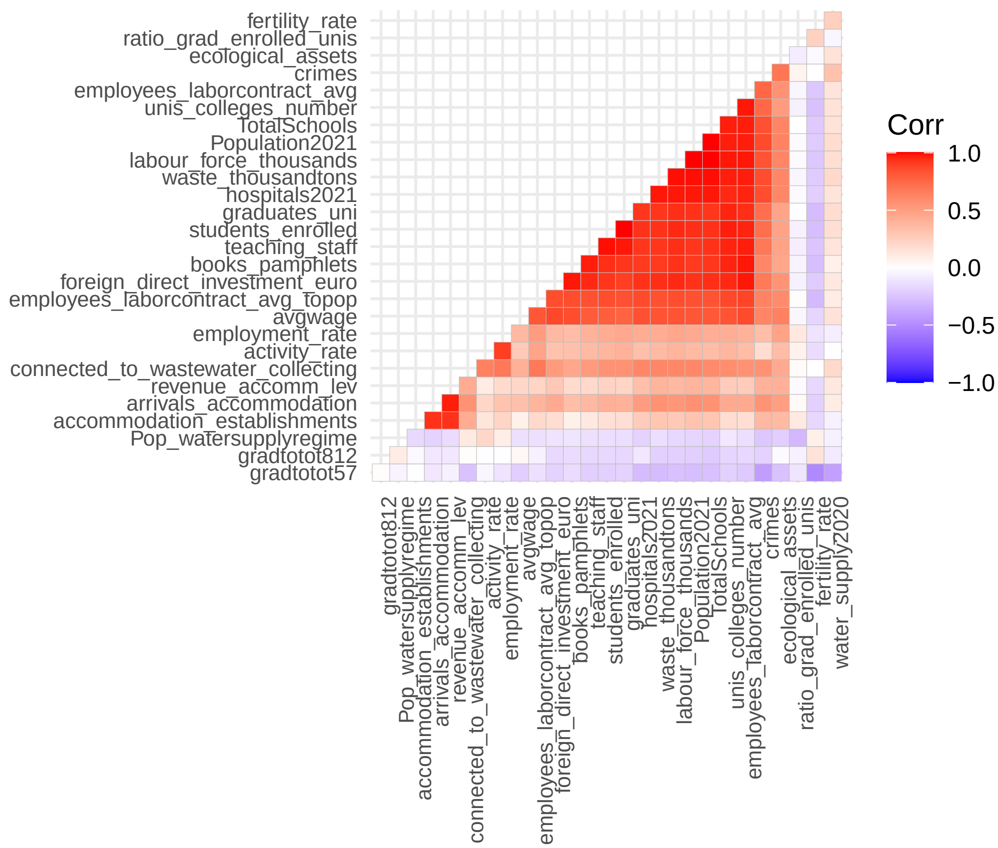
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(ggcorrplot)
```

```
## Warning: package 'ggcorrplot' was built under R version 4.1.3
```

```
ggcorrplot(cor(bul[,-1]), hc.order = TRUE, type = "lower", tl.srt = 90, tl.cex = 8)
```




```
cor_pmat(bul[,-1])
```

```
##                                   Population2021 fertility_rate
## Population2021                      0.000000e+00    0.250945796
## fertility_rate                      2.509458e-01    0.000000000
## unis_colleges_number                8.548964e-17    0.162013246
## teaching_staff                      2.811095e-12    0.196334207
## students_enrolled                   2.308166e-13    0.154160412
## graduates_uni                       1.616035e-12    0.134237146
## ratio_grad_enrolled_unis            9.174784e-01    0.222996596
## TotalSchools                        2.362230e-29    0.241087324
## gradtotot57                         1.418877e-01    0.006058911
## gradtotot812                        2.952747e-01    0.440660293
## employees_laborcontract_avg         3.450719e-17    0.219037081
## employees_laborcontract_avg_topop   3.207855e-08    0.116461484
## avgwage                             1.206978e-07    0.357018115
## activity_rate                       6.364915e-02    0.448764218
## labour_force_thousands              1.353251e-30    0.221271236
## employment_rate                     1.872476e-02    0.558539627
## foreign_direct_investment_euro      9.061431e-13    0.254730904
## books_pamphlets                     2.923013e-11    0.142785867
## accommodation_establishments        1.060308e-01    0.397044789
## arrivals_accommodation              1.723961e-03    0.298147499
## revenue_accomm_lev                  3.938970e-02    0.373134752
## crimes                              2.896434e-09    0.999296010
```

```
## ecological_assets                    3.074933e-04    0.870039025
## water_supply2020                     3.588740e-01    0.193202583
## connected_to_wastewater_collecting   6.150684e-04    0.980647984
## Pop_watersupplyregime                3.555547e-01    0.675779887
## waste_thousandtons                   5.810950e-27    0.262863431
## hospitals2021                        5.654075e-20    0.279043553
##                                    unis_colleges_number teaching_staff
## Population2021                             8.548964e-17   2.811095e-12
## fertility_rate                             1.620132e-01   1.963342e-01
## unis_colleges_number                       0.000000e+00   1.344727e-16
## teaching_staff                             1.344727e-16   0.000000e+00
## students_enrolled                          2.963349e-17   9.052569e-24
## graduates_uni                              7.621126e-16   1.695736e-19
## ratio_grad_enrolled_unis                   7.723807e-01   7.054611e-01
## TotalSchools                               5.813363e-17   1.524990e-12
## gradtotot57                                2.997456e-01   2.866066e-01
## gradtotot812                               3.694584e-01   3.912286e-01
## employees_laborcontract_avg                5.177584e-21   2.891258e-17
## employees_laborcontract_avg_topop          1.486333e-09   1.670977e-08
## avgwage                                    5.508042e-08   2.588742e-07
## activity_rate                              4.367627e-02   5.718552e-02
## labour_force_thousands                     7.430629e-19   1.547110e-13
## employment_rate                            2.730561e-02   3.373584e-02
## foreign_direct_investment_euro             6.483622e-15   2.281280e-14
## books_pamphlets                            3.378688e-16   1.202546e-16
## accommodation_establishments               3.461010e-01   5.396915e-01
## arrivals_accommodation                     1.820738e-02   5.164016e-02
## revenue_accomm_lev                         1.680250e-01   3.005860e-01
## crimes                                     5.074073e-06   7.564267e-05
## ecological_assets                          5.287059e-03   1.023470e-02
## water_supply2020                           4.360589e-01   4.689696e-01
## connected_to_wastewater_collecting         1.269101e-03   4.260336e-03
## Pop_watersupplyregime                      6.193130e-01   5.005677e-01
## waste_thousandtons                         1.387675e-17   4.264841e-12
## hospitals2021                              3.580988e-15   6.870403e-12
##                                    students_enrolled graduates_uni
## Population2021                          2.308166e-13  1.616035e-12
## fertility_rate                          1.541604e-01  1.342371e-01
## unis_colleges_number                    2.963349e-17  7.621126e-16
## teaching_staff                          9.052569e-24  1.695736e-19
## students_enrolled                       0.000000e+00  1.258211e-27
## graduates_uni                           1.258211e-27  0.000000e+00
## ratio_grad_enrolled_unis                7.565319e-01  9.563326e-01
## TotalSchools                            5.619835e-13  3.669794e-12
## gradtotot57                             3.003968e-01  3.201497e-01
## gradtotot812                            3.886736e-01  3.829993e-01
## employees_laborcontract_avg             1.205652e-15  8.571757e-14
## employees_laborcontract_avg_topop       1.010958e-08  2.532003e-08
## avgwage                                 5.230809e-07  2.036097e-06
## activity_rate                           4.697410e-02  3.275288e-02
## labour_force_thousands                  2.481015e-14  2.384885e-13
## employment_rate                         2.261861e-02  1.670390e-02
## foreign_direct_investment_euro          4.017206e-12  1.062732e-10
## books_pamphlets                         1.613908e-13  3.004326e-12
```

```
## accommodation_establishments          4.090528e-01  3.989867e-01
## arrivals_accommodation                3.160789e-02  3.432386e-02
## revenue_accomm_lev                    2.410499e-01  2.659196e-01
## crimes                                1.284867e-05  1.496406e-05
## ecological_assets                     8.930799e-03  1.209404e-02
## water_supply2020                      3.875272e-01  3.687161e-01
## connected_to_wastewater_collecting    1.863047e-03  1.926984e-03
## Pop_watersupplyregime                 4.765412e-01  4.816123e-01
## waste_thousandtons                    3.808107e-13  4.320566e-12
## hospitals2021                         1.332925e-12  8.246069e-12
##                                    ratio_grad_enrolled_unis TotalSchools
## Population2021                                    0.9174784 2.362230e-29
## fertility_rate                                    0.2229966 2.410873e-01
## unis_colleges_number                              0.7723807 5.813363e-17
## teaching_staff                                    0.7054611 1.524990e-12
## students_enrolled                                 0.7565319 5.619835e-13
## graduates_uni                                     0.9563326 3.669794e-12
## ratio_grad_enrolled_unis                          0.0000000 9.952233e-01
## TotalSchools                                      0.9952233 0.000000e+00
## gradtotot57                                       0.5661728 1.687816e-01
## gradtotot812                                      0.7773340 2.667091e-01
## employees_laborcontract_avg                       0.8296388 2.801606e-18
## employees_laborcontract_avg_topop                 0.9407662 3.575026e-08
## avgwage                                           0.8241284 1.561710e-07
## activity_rate                                     0.7279581 6.969806e-02
## labour_force_thousands                            0.9339930 2.023675e-29
## employment_rate                                   0.5405682 2.469519e-02
## foreign_direct_investment_euro                    0.7656116 1.085266e-13
## books_pamphlets                                   0.7995991 5.028625e-12
## accommodation_establishments                      0.5948895 1.178582e-01
## arrivals_accommodation                            0.8841564 2.054524e-03
## revenue_accomm_lev                                0.9095843 4.426235e-02
## crimes                                            0.7736133 1.959883e-08
## ecological_assets                                 0.6985980 4.419658e-04
## water_supply2020                                  0.8252628 4.413161e-01
## connected_to_wastewater_collecting                0.9128116 9.879394e-04
## Pop_watersupplyregime                             0.1143659 3.332932e-01
## waste_thousandtons                                0.8285402 1.372964e-22
## hospitals2021                                     0.9084748 6.089195e-19
##                                    gradtotot57 gradtotot812
## Population2021                     0.141887698    0.2952747
## fertility_rate                     0.006058911    0.4406603
## unis_colleges_number               0.299745563    0.3694584
## teaching_staff                     0.286606596    0.3912286
## students_enrolled                  0.300396829    0.3886736
## graduates_uni                      0.320149717    0.3829993
## ratio_grad_enrolled_unis           0.566172818    0.7773340
## TotalSchools                       0.168781606    0.2667091
## gradtotot57                        0.000000000    0.9673395
## gradtotot812                       0.967339496    0.0000000
## employees_laborcontract_avg        0.249296589    0.3918405
## employees_laborcontract_avg_topop  0.499404746    0.7780903
## avgwage                            0.284910052    0.8371293
## activity_rate                      0.838906739    0.9615401
```

```
## labour_force_thousands            0.172670229    0.3102730
## employment_rate                   0.526816958    0.9459506
## foreign_direct_investment_euro    0.336302111    0.4206455
## books_pamphlets                   0.435087034    0.4373995
## accommodation_establishments      0.961688540    0.8823993
## arrivals_accommodation            0.617980182    0.6007516
## revenue_accomm_lev                0.750597225    0.6186110
## crimes                            0.025874115    0.3262990
## ecological_assets                 0.174482585    0.9275258
## water_supply2020                  0.030624543    0.6586801
## connected_to_wastewater_collecting 0.181640485    0.9485631
## Pop_watersupplyregime             0.818751436    0.6027330
## waste_thousandtons                0.127778512    0.2954185
## hospitals2021                     0.138652510    0.6143097
##                                   employees_laborcontract_avg
## Population2021                                   3.450719e-17
## fertility_rate                                   2.190371e-01
## unis_colleges_number                             5.177584e-21
## teaching_staff                                   2.891258e-17
## students_enrolled                                1.205652e-15
## graduates_uni                                    8.571757e-14
## ratio_grad_enrolled_unis                         8.296388e-01
## TotalSchools                                     2.801606e-18
## gradtotot57                                      2.492966e-01
## gradtotot812                                     3.918405e-01
## employees_laborcontract_avg                      0.000000e+00
## employees_laborcontract_avg_topop                7.707822e-10
## avgwage                                          4.769939e-09
## activity_rate                                    7.004113e-02
## labour_force_thousands                           8.988365e-20
## employment_rate                                  3.689337e-02
## foreign_direct_investment_euro                   4.961713e-21
## books_pamphlets                                  1.641697e-19
## accommodation_establishments                     3.781604e-01
## arrivals_accommodation                           2.008526e-02
## revenue_accomm_lev                               1.697448e-01
## crimes                                           5.272764e-06
## ecological_assets                                1.776650e-03
## water_supply2020                                 4.737079e-01
## connected_to_wastewater_collecting               2.169007e-03
## Pop_watersupplyregime                            4.786344e-01
## waste_thousandtons                               3.411883e-17
## hospitals2021                                    2.143765e-16
##                                   employees_laborcontract_avg_topop
## Population2021                                         3.207855e-08
## fertility_rate                                         1.164615e-01
## unis_colleges_number                                   1.486333e-09
## teaching_staff                                         1.670977e-08
## students_enrolled                                      1.010958e-08
## graduates_uni                                          2.532003e-08
## ratio_grad_enrolled_unis                               9.407662e-01
## TotalSchools                                           3.575026e-08
## gradtotot57                                            4.994047e-01
## gradtotot812                                           7.780903e-01
```

```
## employees_laborcontract_avg                                 7.707822e-10
## employees_laborcontract_avg_topop                           0.000000e+00
## avgwage                                                     8.347491e-08
## activity_rate                                               1.078904e-02
## labour_force_thousands                                      8.458678e-09
## employment_rate                                             5.876543e-03
## foreign_direct_investment_euro                              8.777159e-09
## books_pamphlets                                             1.258685e-08
## accommodation_establishments                                3.180112e-01
## arrivals_accommodation                                      3.902443e-02
## revenue_accomm_lev                                          2.676039e-01
## crimes                                                      3.575331e-04
## ecological_assets                                           9.384480e-04
## water_supply2020                                            6.185624e-01
## connected_to_wastewater_collecting                          8.133561e-05
## Pop_watersupplyregime                                       5.217981e-01
## waste_thousandtons                                          2.196234e-08
## hospitals2021                                               5.841520e-09
##                                         avgwage activity_rate
## Population2021                     1.206978e-07  6.364915e-02
## fertility_rate                     3.570181e-01  4.487642e-01
## unis_colleges_number               5.508042e-08  4.367627e-02
## teaching_staff                     2.588742e-07  5.718552e-02
## students_enrolled                  5.230809e-07  4.697410e-02
## graduates_uni                      2.036097e-06  3.275288e-02
## ratio_grad_enrolled_unis           8.241284e-01  7.279581e-01
## TotalSchools                       1.561710e-07  6.969806e-02
## gradtotot57                        2.849101e-01  8.389067e-01
## gradtotot812                       8.371293e-01  9.615401e-01
## employees_laborcontract_avg        4.769939e-09  7.004113e-02
## employees_laborcontract_avg_topop  8.347491e-08  1.078904e-02
## avgwage                            0.000000e+00  1.429699e-01
## activity_rate                      1.429699e-01  0.000000e+00
## labour_force_thousands             4.202570e-08  4.039449e-02
## employment_rate                    5.280930e-02  6.367707e-11
## foreign_direct_investment_euro     1.405608e-09  9.453520e-02
## books_pamphlets                    1.104665e-08  8.302153e-02
## accommodation_establishments       6.795621e-01  5.216436e-01
## arrivals_accommodation             8.244060e-02  2.928131e-01
## revenue_accomm_lev                 2.782676e-01  5.992890e-01
## crimes                             4.245266e-04  3.577024e-01
## ecological_assets                  2.631720e-04  7.052540e-02
## water_supply2020                   4.354830e-01  9.766521e-01
## connected_to_wastewater_collecting 3.166612e-02  2.207454e-04
## Pop_watersupplyregime              4.976932e-01  2.840861e-01
## waste_thousandtons                 5.226491e-08  6.244443e-02
## hospitals2021                      3.258252e-08  8.721927e-02
##                                    labour_force_thousands employment_rate
## Population2021                               1.353251e-30    1.872476e-02
## fertility_rate                               2.212712e-01    5.585396e-01
## unis_colleges_number                         7.430629e-19    2.730561e-02
## teaching_staff                               1.547110e-13    3.373584e-02
## students_enrolled                            2.481015e-14    2.261861e-02
## graduates_uni                                2.384885e-13    1.670390e-02
```

```
## ratio_grad_enrolled_unis                   9.339930e-01    5.405682e-01
## TotalSchools                               2.023675e-29    2.469519e-02
## gradtotot57                                1.726702e-01    5.268170e-01
## gradtotot812                               3.102730e-01    9.459506e-01
## employees_laborcontract_avg                8.988365e-20    3.689337e-02
## employees_laborcontract_avg_topop          8.458678e-09    5.876543e-03
## avgwage                                    4.202570e-08    5.280930e-02
## activity_rate                              4.039449e-02    6.367707e-11
## labour_force_thousands                     0.000000e+00    1.334995e-02
## employment_rate                            1.334995e-02    0.000000e+00
## foreign_direct_investment_euro             2.669757e-14    5.607410e-02
## books_pamphlets                            8.989825e-13    6.397558e-02
## accommodation_establishments               1.426862e-01    2.716921e-01
## arrivals_accommodation                     2.782551e-03    1.099995e-01
## revenue_accomm_lev                         5.256686e-02    3.221385e-01
## crimes                                     4.605281e-08    7.418199e-02
## ecological_assets                          4.022472e-04    9.965846e-03
## water_supply2020                           3.921588e-01    7.231720e-01
## connected_to_wastewater_collecting         5.677852e-04    6.673639e-05
## Pop_watersupplyregime                      4.004869e-01    6.414645e-01
## waste_thousandtons                         5.224098e-25    2.178014e-02
## hospitals2021                              6.634480e-19    2.258016e-02
##                                    foreign_direct_investment_euro
## Population2021                                       9.061431e-13
## fertility_rate                                       2.547309e-01
## unis_colleges_number                                 6.483622e-15
## teaching_staff                                       2.281280e-14
## students_enrolled                                    4.017206e-12
## graduates_uni                                        1.062732e-10
## ratio_grad_enrolled_unis                             7.656116e-01
## TotalSchools                                         1.085266e-13
## gradtotot57                                          3.363021e-01
## gradtotot812                                         4.206455e-01
## employees_laborcontract_avg                          4.961713e-21
## employees_laborcontract_avg_topop                    8.777159e-09
## avgwage                                              1.405608e-09
## activity_rate                                        9.453520e-02
## labour_force_thousands                               2.669757e-14
## employment_rate                                      5.607410e-02
## foreign_direct_investment_euro                       0.000000e+00
## books_pamphlets                                      1.659138e-19
## accommodation_establishments                         3.555852e-01
## arrivals_accommodation                               1.815019e-02
## revenue_accomm_lev                                   1.302487e-01
## crimes                                               9.553139e-05
## ecological_assets                                    2.698625e-03
## water_supply2020                                     5.529411e-01
## connected_to_wastewater_collecting                   4.749950e-03
## Pop_watersupplyregime                                5.582831e-01
## waste_thousandtons                                   3.161996e-13
## hospitals2021                                        7.493611e-13
##                                    books_pamphlets accommodation_establishments
## Population2021                        2.923013e-11                 1.060308e-01
## fertility_rate                        1.427859e-01                 3.970448e-01
```

```
## unis_colleges_number                  3.378688e-16             3.461010e-01
## teaching_staff                        1.202546e-16             5.396915e-01
## students_enrolled                     1.613908e-13             4.090528e-01
## graduates_uni                         3.004326e-12             3.989867e-01
## ratio_grad_enrolled_unis              7.995991e-01             5.948895e-01
## TotalSchools                          5.028625e-12             1.178582e-01
## gradtotot57                           4.350870e-01             9.616885e-01
## gradtotot812                          4.373995e-01             8.823993e-01
## employees_laborcontract_avg           1.641697e-19             3.781604e-01
## employees_laborcontract_avg_topop     1.258685e-08             3.180112e-01
## avgwage                               1.104665e-08             6.795621e-01
## activity_rate                         8.302153e-02             5.216436e-01
## labour_force_thousands                8.989825e-13             1.426862e-01
## employment_rate                       6.397558e-02             2.716921e-01
## foreign_direct_investment_euro        1.659138e-19             3.555852e-01
## books_pamphlets                       0.000000e+00             6.906769e-01
## accommodation_establishments          6.906769e-01             0.000000e+00
## arrivals_accommodation                7.783529e-02             3.501127e-13
## revenue_accomm_lev                    3.425449e-01             3.621872e-13
## crimes                                5.219904e-04             5.478229e-02
## ecological_assets                     1.407395e-02             5.726460e-02
## water_supply2020                      6.856924e-01             7.621771e-01
## connected_to_wastewater_collecting    1.346303e-02             2.252809e-02
## Pop_watersupplyregime                 5.816912e-01             4.308826e-01
## waste_thousandtons                    2.055625e-11             1.153675e-01
## hospitals2021                         3.117799e-11             1.567887e-01
##                                    arrivals_accommodation revenue_accomm_lev
## Population2021                               1.723961e-03       3.938970e-02
## fertility_rate                               2.981475e-01       3.731348e-01
## unis_colleges_number                         1.820738e-02       1.680250e-01
## teaching_staff                               5.164016e-02       3.005860e-01
## students_enrolled                            3.160789e-02       2.410499e-01
## graduates_uni                                3.432386e-02       2.659196e-01
## ratio_grad_enrolled_unis                     8.841564e-01       9.095843e-01
## TotalSchools                                 2.054524e-03       4.426235e-02
## gradtotot57                                  6.179802e-01       7.505972e-01
## gradtotot812                                 6.007516e-01       6.186110e-01
## employees_laborcontract_avg                  2.008526e-02       1.697448e-01
## employees_laborcontract_avg_topop            3.902443e-02       2.676039e-01
## avgwage                                      8.244060e-02       2.782676e-01
## activity_rate                                2.928131e-01       5.992890e-01
## labour_force_thousands                       2.782551e-03       5.256686e-02
## employment_rate                              1.099995e-01       3.221385e-01
## foreign_direct_investment_euro               1.815019e-02       1.302487e-01
## books_pamphlets                              7.783529e-02       3.425449e-01
## accommodation_establishments                 3.501127e-13       3.621872e-13
## arrivals_accommodation                       0.000000e+00       7.767479e-17
## revenue_accomm_lev                           7.767479e-17       0.000000e+00
## crimes                                       2.267126e-03       3.641935e-02
## ecological_assets                            6.520253e-03       3.008247e-02
## water_supply2020                             6.180186e-01       5.459912e-01
## connected_to_wastewater_collecting           1.890579e-03       2.239572e-02
## Pop_watersupplyregime                        3.271437e-01       4.597803e-01
## waste_thousandtons                           2.053488e-03       4.125734e-02
```

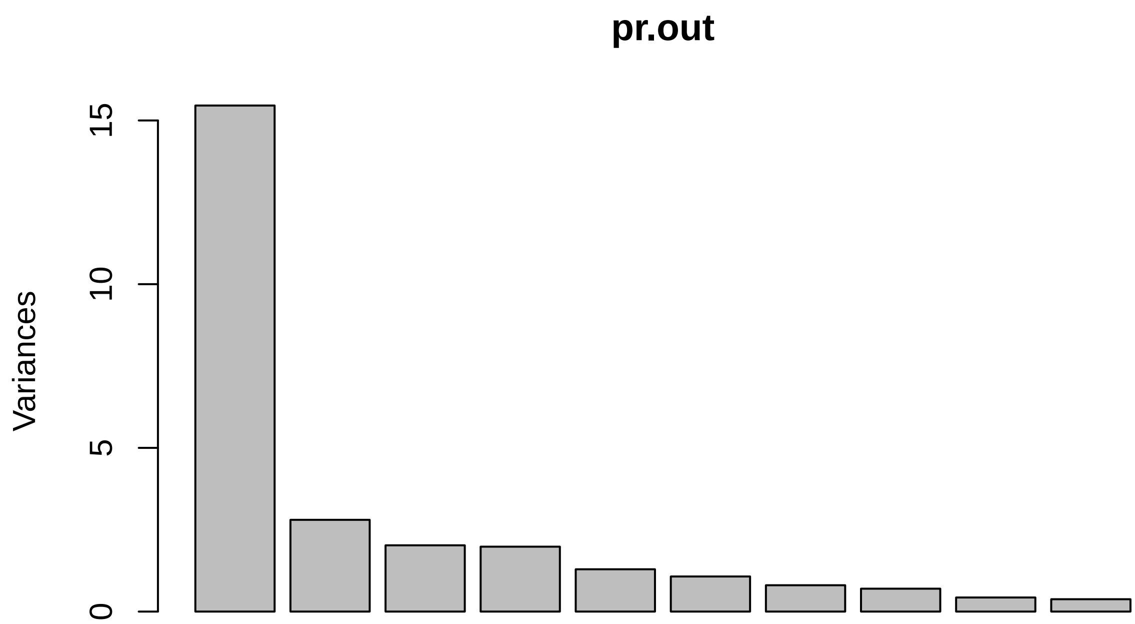
```
## hospitals2021                                 5.226827e-03       8.942890e-02
##                                         crimes ecological_assets
## Population2021                    2.896434e-09      3.074933e-04
## fertility_rate                    9.992960e-01      8.700390e-01
## unis_colleges_number              5.074073e-06      5.287059e-03
## teaching_staff                    7.564267e-05      1.023470e-02
## students_enrolled                 1.284867e-05      8.930799e-03
## graduates_uni                     1.496406e-05      1.209404e-02
## ratio_grad_enrolled_unis          7.736133e-01      6.985980e-01
## TotalSchools                      1.959883e-08      4.419658e-04
## gradtotot57                       2.587411e-02      1.744826e-01
## gradtotot812                      3.262990e-01      9.275258e-01
## employees_laborcontract_avg       5.272764e-06      1.776650e-03
## employees_laborcontract_avg_topop 3.575331e-04      9.384480e-04
## avgwage                           4.245266e-04      2.631720e-04
## activity_rate                     3.577024e-01      7.052540e-02
## labour_force_thousands            4.605281e-08      4.022472e-04
## employment_rate                   7.418199e-02      9.965846e-03
## foreign_direct_investment_euro    9.553139e-05      2.698625e-03
## books_pamphlets                   5.219904e-04      1.407395e-02
## accommodation_establishments      5.478229e-02      5.726460e-02
## arrivals_accommodation            2.267126e-03      6.520253e-03
## revenue_accomm_lev                3.641935e-02      3.008247e-02
## crimes                            0.000000e+00      4.909495e-05
## ecological_assets                 4.909495e-05      0.000000e+00
## water_supply2020                  9.173999e-02      4.478699e-01
## connected_to_wastewater_collecting 2.148812e-03      1.783444e-02
## Pop_watersupplyregime             2.115819e-01      2.952266e-01
## waste_thousandtons                1.060412e-08      5.035568e-04
## hospitals2021                     1.050377e-08      5.082737e-04
##                                   water_supply2020
## Population2021                          0.35887402
## fertility_rate                          0.19320258
## unis_colleges_number                    0.43605885
## teaching_staff                          0.46896957
## students_enrolled                       0.38752721
## graduates_uni                           0.36871610
## ratio_grad_enrolled_unis                0.82526279
## TotalSchools                            0.44131607
## gradtotot57                             0.03062454
## gradtotot812                            0.65868014
## employees_laborcontract_avg             0.47370794
## employees_laborcontract_avg_topop       0.61856240
## avgwage                                 0.43548300
## activity_rate                           0.97665210
## labour_force_thousands                  0.39215879
## employment_rate                         0.72317200
## foreign_direct_investment_euro          0.55294114
## books_pamphlets                         0.68569244
## accommodation_establishments            0.76217711
## arrivals_accommodation                  0.61801859
## revenue_accomm_lev                      0.54599119
## crimes                                  0.09173999
## ecological_assets                       0.44786989
```

```
## water_supply2020                         0.00000000
## connected_to_wastewater_collecting       0.30767670
## Pop_watersupplyregime                    0.74842448
## waste_thousandtons                       0.31990552
## hospitals2021                            0.46762118
##                                    connected_to_wastewater_collecting
## Population2021                                           6.150684e-04
## fertility_rate                                           9.806480e-01
## unis_colleges_number                                     1.269101e-03
## teaching_staff                                           4.260336e-03
## students_enrolled                                        1.863047e-03
## graduates_uni                                            1.926984e-03
## ratio_grad_enrolled_unis                                 9.128116e-01
## TotalSchools                                             9.879394e-04
## gradtotot57                                              1.816405e-01
## gradtotot812                                             9.485631e-01
## employees_laborcontract_avg                              2.169007e-03
## employees_laborcontract_avg_topop                        8.133561e-05
## avgwage                                                  3.166612e-02
## activity_rate                                            2.207454e-04
## labour_force_thousands                                   5.677852e-04
## employment_rate                                          6.673639e-05
## foreign_direct_investment_euro                           4.749950e-03
## books_pamphlets                                          1.346303e-02
## accommodation_establishments                             2.252809e-02
## arrivals_accommodation                                   1.890579e-03
## revenue_accomm_lev                                       2.239572e-02
## crimes                                                   2.148812e-03
## ecological_assets                                        1.783444e-02
## water_supply2020                                         3.076767e-01
## connected_to_wastewater_collecting                       0.000000e+00
## Pop_watersupplyregime                                    5.816404e-01
## waste_thousandtons                                       5.890266e-04
## hospitals2021                                            6.313287e-04
##                                    Pop_watersupplyregime waste_thousandtons
## Population2021                                 0.3555547       5.810950e-27
## fertility_rate                                 0.6757799       2.628634e-01
## unis_colleges_number                           0.6193130       1.387675e-17
## teaching_staff                                 0.5005677       4.264841e-12
## students_enrolled                              0.4765412       3.808107e-13
## graduates_uni                                  0.4816123       4.320566e-12
## ratio_grad_enrolled_unis                       0.1143659       8.285402e-01
## TotalSchools                                   0.3332932       1.372964e-22
## gradtotot57                                    0.8187514       1.277785e-01
## gradtotot812                                   0.6027330       2.954185e-01
## employees_laborcontract_avg                    0.4786344       3.411883e-17
## employees_laborcontract_avg_topop              0.5217981       2.196234e-08
## avgwage                                        0.4976932       5.226491e-08
## activity_rate                                  0.2840861       6.244443e-02
## labour_force_thousands                         0.4004869       5.224098e-25
## employment_rate                                0.6414645       2.178014e-02
## foreign_direct_investment_euro                 0.5582831       3.161996e-13
## books_pamphlets                                0.5816912       2.055625e-11
## accommodation_establishments                   0.4308826       1.153675e-01
```

```
## arrivals_accommodation                            0.3271437        2.053488e-03
## revenue_accomm_lev                                0.4597803        4.125734e-02
## crimes                                            0.2115819        1.060412e-08
## ecological_assets                                 0.2952266        5.035568e-04
## water_supply2020                                  0.7484245        3.199055e-01
## connected_to_wastewater_collecting                0.5816404        5.890266e-04
## Pop_watersupplyregime                             0.0000000        4.781744e-01
## waste_thousandtons                                0.4781744        0.000000e+00
## hospitals2021                                     0.3212510        6.523838e-19
##                                     hospitals2021
## Population2021                       5.654075e-20
## fertility_rate                       2.790436e-01
## unis_colleges_number                 3.580988e-15
## teaching_staff                       6.870403e-12
## students_enrolled                    1.332925e-12
## graduates_uni                        8.246069e-12
## ratio_grad_enrolled_unis             9.084748e-01
## TotalSchools                         6.089195e-19
## gradtotot57                          1.386525e-01
## gradtotot812                         6.143097e-01
## employees_laborcontract_avg          2.143765e-16
## employees_laborcontract_avg_topop    5.841520e-09
## avgwage                              3.258252e-08
## activity_rate                        8.721927e-02
## labour_force_thousands               6.634480e-19
## employment_rate                      2.258016e-02
## foreign_direct_investment_euro       7.493611e-13
## books_pamphlets                      3.117799e-11
## accommodation_establishments         1.567887e-01
## arrivals_accommodation               5.226827e-03
## revenue_accomm_lev                   8.942890e-02
## crimes                               1.050377e-08
## ecological_assets                    5.082737e-04
## water_supply2020                     4.676212e-01
## connected_to_wastewater_collecting   6.313287e-04
## Pop_watersupplyregime                3.212510e-01
## waste_thousandtons                   6.523838e-19
## hospitals2021                        0.000000e+00
```

PCA

```r
pr.out <- prcomp(bul[,-1] , scale = TRUE)
plot(pr.out)
```




```
pr.out$rotation
```

```
##                                           PC1          PC2          PC3
## Population2021                    0.251140229 -0.005046812  0.052035616
## fertility_rate                   -0.061933786  0.004236421  0.443056816
## unis_colleges_number              0.246977988 -0.100950594 -0.019287931
## teaching_staff                    0.239559270 -0.144548347 -0.017635288
## students_enrolled                 0.242020761 -0.116779470 -0.019543819
## graduates_uni                     0.239369553 -0.112562813 -0.028872779
## ratio_grad_enrolled_unis         -0.005262813  0.082745242  0.141708436
## TotalSchools                      0.250129175 -0.013680735  0.045314795
## gradtotot57                      -0.066483530 -0.011406212 -0.461880847
## gradtotot812                     -0.041952643  0.003572495 -0.058022670
## employees_laborcontract_avg       0.247939032 -0.105909278  0.009054059
## employees_laborcontract_avg_topop  0.226481280 -0.065514636 -0.110787032
## avgwage                           0.215023672 -0.118080001  0.024551632
## activity_rate                     0.110739683  0.056607059 -0.375160411
## labour_force_thousands            0.252312426 -0.023844442  0.024538549
## employment_rate                   0.129915368  0.111185559 -0.314809366
## foreign_direct_investment_euro    0.240654010 -0.101108177 -0.003687789
## books_pamphlets                   0.234898285 -0.170648060 -0.042960180
## accommodation_establishments      0.082140331  0.538786847 -0.041966539
## arrivals_accommodation            0.145251001  0.470908872  0.001131015
## revenue_accomm_lev                0.101725203  0.507921078  0.015530774
## crimes                            0.207845520  0.097844232  0.249667957
```

```
## ecological_assets                   0.165723426  0.160041012  0.068263378
## water_supply2020                    0.045534485  0.007033836  0.416155456
## connected_to_wastewater_collecting  0.168502577  0.194519974 -0.106008313
## Pop_watersupplyregime              -0.040172150 -0.077565432 -0.215900135
## waste_thousandtons                  0.250515857 -0.012492624  0.050882951
## hospitals2021                       0.247929971 -0.034340547  0.048643960
##                                            PC4           PC5          PC6
## Population2021                     -0.04016222  1.405881e-03 -0.026670081
## fertility_rate                      0.37475711 -3.609576e-02  0.103935000
## unis_colleges_number               -0.04174413  2.652159e-02 -0.034809698
## teaching_staff                     -0.04685187  1.025979e-02 -0.027435736
## students_enrolled                  -0.03959269  4.495466e-03 -0.045609725
## graduates_uni                      -0.02697869 -3.164669e-02 -0.065203828
## ratio_grad_enrolled_unis            0.11278802 -7.418846e-01  0.013598332
## TotalSchools                       -0.05623542 -1.941028e-02 -0.027090775
## gradtotot57                        -0.33887249 -8.712085e-02  0.041488149
## gradtotot812                        0.17981239  1.652511e-01  0.837895553
## employees_laborcontract_avg        -0.05237174  7.969481e-03  0.014758752
## employees_laborcontract_avg_topop   0.04736688 -5.897516e-02  0.105921510
## avgwage                            -0.01160603  3.114401e-02  0.271902878
## activity_rate                       0.44251861 -1.249495e-01 -0.141843819
## labour_force_thousands             -0.02995947 -6.659888e-03 -0.025058190
## employment_rate                     0.42421384 -1.925941e-01 -0.058761222
## foreign_direct_investment_euro     -0.07187571  3.242673e-02  0.039920946
## books_pamphlets                    -0.08357866  3.638601e-05  0.022444782
## accommodation_establishments       -0.12261989  2.032362e-02  0.039670130
## arrivals_accommodation             -0.11912516  7.387451e-02 -0.005329358
## revenue_accomm_lev                 -0.14676286  1.449268e-01 -0.025052242
## crimes                              0.02072875 -1.468744e-02 -0.027129525
## ecological_assets                   0.10532497  2.038307e-02  0.231521726
## water_supply2020                    0.15655341  2.256034e-01 -0.246563843
## connected_to_wastewater_collecting  0.33698932  5.638967e-02 -0.067358955
## Pop_watersupplyregime               0.30400210  5.196688e-01 -0.184599379
## waste_thousandtons                 -0.03224516  3.517184e-02 -0.036443026
## hospitals2021                      -0.02989101 -3.166684e-03  0.080688532
##                                            PC7          PC8         PC9
## Population2021                      0.02019425 -0.059980228  0.06275423
## fertility_rate                      0.23468104 -0.176219808 -0.03654949
## unis_colleges_number                0.09120326  0.051452197  0.01717751
## teaching_staff                      0.08183814  0.070820986  0.08810036
## students_enrolled                   0.05543924  0.110100735  0.14526978
## graduates_uni                       0.06115848  0.146203307  0.14627099
## ratio_grad_enrolled_unis            0.33129749  0.177025021 -0.36766992
## TotalSchools                        0.04673119 -0.071525175  0.03656808
## gradtotot57                        -0.14720072  0.210109707 -0.22699371
## gradtotot812                        0.09945341  0.327153104  0.14803643
## employees_laborcontract_avg         0.07032260 -0.008784486 -0.04867281
## employees_laborcontract_avg_topop  -0.06585973  0.109364831 -0.11445501
## avgwage                            -0.10297619 -0.049047887 -0.45033635
## activity_rate                      -0.16466121  0.080136405 -0.02264198
## labour_force_thousands              0.02609689 -0.040365356  0.01983792
## employment_rate                    -0.18778716 -0.101407142  0.09128334
## foreign_direct_investment_euro      0.10711059 -0.018436774 -0.18727399
## books_pamphlets                     0.12651227  0.030829900 -0.11940666
```

```
## accommodation_establishments        0.17436277  0.009984262  0.02076352
## arrivals_accommodation              0.09251497  0.056889976 -0.03328867
## revenue_accomm_lev                  0.07528513  0.066001173 -0.15934926
## crimes                             -0.12063551 -0.160887415  0.20309459
## ecological_assets                  -0.50307637 -0.437299225 -0.28397969
## water_supply2020                   -0.38587557  0.636274968 -0.25644061
## connected_to_wastewater_collecting  0.07506431  0.221115544  0.25165095
## Pop_watersupplyregime               0.44588949 -0.140510827 -0.40255136
## waste_thousandtons                  0.04099822 -0.049472640  0.02010311
## hospitals2021                       0.03824365 -0.041886813  0.13040802
##                                            PC10         PC11          PC12
## Population2021                     -0.0168836459 -0.093225382  0.1018160596
## fertility_rate                      0.6380286887  0.231486672  0.1769750710
## unis_colleges_number                0.0239611255 -0.016547933  0.0514600322
## teaching_staff                      0.0307476342  0.183442531  0.0886325014
## students_enrolled                   0.0004004975  0.047319915  0.1187740551
## graduates_uni                      -0.0519972274 -0.005348873  0.1794255872
## ratio_grad_enrolled_unis           -0.1729083306 -0.243545964  0.0309737448
## TotalSchools                        0.0004006340 -0.057788246  0.0874262960
## gradtotot57                         0.6080084393 -0.275306719  0.2913052930
## gradtotot812                       -0.1370303922 -0.081850435  0.1925188825
## employees_laborcontract_avg         0.0520653695  0.055411878 -0.0180460264
## employees_laborcontract_avg_topop   0.0743572849 -0.115109404 -0.6083186280
## avgwage                            -0.0737695150  0.131985939 -0.1103068998
## activity_rate                      -0.0552838685  0.208197337  0.1974823402
## labour_force_thousands             -0.0129917572 -0.034755837  0.0901811381
## employment_rate                    -0.0950562059  0.125241648  0.1968395523
## foreign_direct_investment_euro      0.1153390853  0.212229917 -0.0412527522
## books_pamphlets                     0.0219390341  0.214055832  0.0007111605
## accommodation_establishments       -0.0288290821  0.001390045  0.0014814354
## arrivals_accommodation              0.0015191893  0.098974325  0.0447057764
## revenue_accomm_lev                 -0.0305071498  0.237413108  0.0492522367
## crimes                              0.0055588757 -0.558573276  0.2035447627
## ecological_assets                   0.0362582972 -0.118389589 -0.0181203832
## water_supply2020                   -0.0395601061 -0.040884384  0.1182334671
## connected_to_wastewater_collecting  0.3340814174 -0.186116166 -0.4763904399
## Pop_watersupplyregime              -0.1151861273 -0.345360056  0.0969978774
## waste_thousandtons                 -0.0238408894 -0.076750683  0.0650278689
## hospitals2021                      -0.0146156027 -0.133726558  0.0104478511
##                                            PC13         PC14         PC15
## Population2021                     -0.1321722330  0.093474052  0.027638529
## fertility_rate                      0.0549335456 -0.076049052  0.190349517
## unis_colleges_number               -0.0266267429  0.084927756  0.170125690
## teaching_staff                      0.3296591037 -0.129434155 -0.131896564
## students_enrolled                   0.3223717184 -0.277026611 -0.009374319
## graduates_uni                       0.3849009164 -0.265605969  0.009411193
## ratio_grad_enrolled_unis            0.0724045786  0.052040906 -0.141319676
## TotalSchools                       -0.1279401565  0.216749460 -0.026143983
## gradtotot57                        -0.0371149842 -0.042851029 -0.003079070
## gradtotot812                        0.0093051158  0.120481589 -0.053755583
## employees_laborcontract_avg        -0.0231376124  0.178184628 -0.112819185
## employees_laborcontract_avg_topop   0.2170818491 -0.049027191  0.361194174
## avgwage                            -0.3529072840 -0.499788252  0.222815066
## activity_rate                      -0.0004909334  0.385773942  0.437424253
```

```
## labour_force_thousands            -0.1159618891  0.130856611 -0.020306794
## employment_rate                   -0.1840850669 -0.403153160 -0.266385818
## foreign_direct_investment_euro    -0.0864655962  0.183398640 -0.235351122
## books_pamphlets                   -0.0047737436  0.186571537 -0.194218466
## accommodation_establishments       0.2539483214 -0.035968977  0.315486202
## arrivals_accommodation            -0.0492190213 -0.006592734 -0.022878872
## revenue_accomm_lev                -0.0958276207 -0.069414221 -0.135651369
## crimes                            -0.1215066743 -0.125571328  0.107171134
## ecological_assets                  0.4243591003  0.189530657 -0.249992143
## water_supply2020                   0.0095962307  0.040162578 -0.031193869
## connected_to_wastewater_collecting -0.1314769442 -0.009570540 -0.339207667
## Pop_watersupplyregime              0.1155992369 -0.038622851 -0.047841014
## waste_thousandtons                -0.1791533385  0.062930152  0.175946231
## hospitals2021                     -0.1848429727  0.048135287  0.050536760
##                                          PC16         PC17         PC18
## Population2021                    -0.07710523 -0.129364489 -0.089701395
## fertility_rate                     0.06228573 -0.092578349 -0.035614218
## unis_colleges_number              -0.07346819 -0.429518433 -0.005901009
## teaching_staff                    -0.02108584  0.122618706  0.159107017
## students_enrolled                 -0.14704734  0.030368022 -0.017382224
## graduates_uni                     -0.17604598  0.057034704 -0.155396774
## ratio_grad_enrolled_unis          -0.06764702  0.004709447 -0.015114569
## TotalSchools                      -0.09965398 -0.140541976 -0.176573630
## gradtotot57                        0.03332089 -0.011601388 -0.021399049
## gradtotot812                       0.02071715 -0.078409544  0.023349072
## employees_laborcontract_avg        0.08927507 -0.016014067  0.083869565
## employees_laborcontract_avg_topop  0.34620346 -0.297320462  0.027079306
## avgwage                           -0.32614332  0.210215987  0.039521141
## activity_rate                     -0.21066731  0.218805084  0.163734938
## labour_force_thousands            -0.12483484 -0.167052338 -0.093584807
## employment_rate                    0.43301928 -0.224043524 -0.088335975
## foreign_direct_investment_euro     0.28397734  0.386075077 -0.048628070
## books_pamphlets                    0.15495621 -0.072691756  0.419308568
## accommodation_establishments       0.23366748  0.268961521  0.025040531
## arrivals_accommodation            -0.06783844 -0.100581051 -0.379495289
## revenue_accomm_lev                -0.07092829 -0.205087305  0.375331896
## crimes                             0.09112184  0.104875484  0.510061856
## ecological_assets                 -0.13413661 -0.008209166 -0.096416008
## water_supply2020                   0.16314088  0.009875505 -0.076672278
## connected_to_wastewater_collecting -0.34457527  0.155615575  0.035396129
## Pop_watersupplyregime              0.04406206 -0.009788759 -0.039430295
## waste_thousandtons                -0.02390299 -0.093289450 -0.178665815
## hospitals2021                      0.30159797  0.403833928 -0.286367833
##                                           PC19         PC20         PC21
## Population2021                    -0.154503748 -0.043549917  0.307144277
## fertility_rate                    -0.032124130  0.046450253  0.068674807
## unis_colleges_number               0.568141129  0.184904978 -0.523078719
## teaching_staff                    -0.242905710 -0.036666200 -0.400488598
## students_enrolled                 -0.052205906 -0.221050665  0.061306063
## graduates_uni                      0.103326338  0.205780106  0.281957120
## ratio_grad_enrolled_unis           0.022881560 -0.023021262 -0.025539591
## TotalSchools                      -0.408376680  0.267115180 -0.154543141
## gradtotot57                       -0.016054830  0.002050614  0.031113915
## gradtotot812                      -0.038771104 -0.064049526 -0.012397133
```

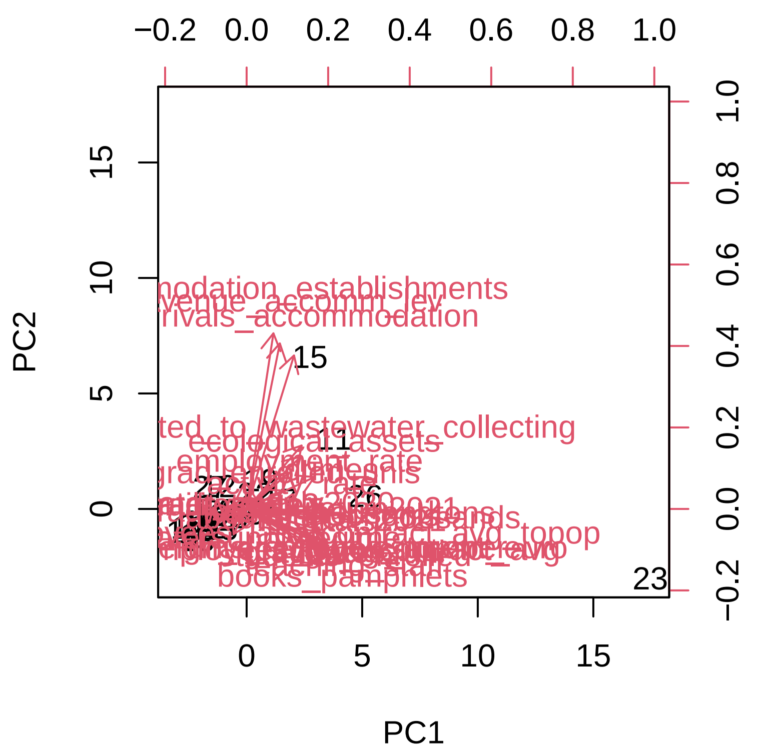
```
## employees_laborcontract_avg        -0.092580596 -0.128446268  0.099117046
## employees_laborcontract_avg_topop  -0.263475459  0.003936368  0.048354880
## avgwage                             0.007355372  0.072237077 -0.033152913
## activity_rate                      -0.083426973  0.025488912 -0.017970142
## labour_force_thousands             -0.173409321 -0.040669704  0.168902664
## employment_rate                     0.024620222 -0.031351816 -0.014868603
## foreign_direct_investment_euro     -0.032736532 -0.292591009 -0.283526928
## books_pamphlets                     0.242799215  0.224201731  0.356753583
## accommodation_establishments        0.193322154 -0.051909938  0.042794921
## arrivals_accommodation             -0.169503103  0.046030539 -0.164336350
## revenue_accomm_lev                 -0.062085362  0.078478970  0.104313837
## crimes                             -0.114871456 -0.038966560 -0.166695216
## ecological_assets                   0.156851129  0.018053772  0.010619666
## water_supply2020                    0.020320334  0.011650100  0.009907705
## connected_to_wastewater_collecting  0.142264473 -0.008082126  0.005050094
## Pop_watersupplyregime              -0.050101597  0.030905604  0.009886265
## waste_thousandtons                  0.280639863 -0.639668990  0.169182685
## hospitals2021                       0.159588451  0.457352796  0.107564439
##                                             PC22         PC23          PC24
## Population2021                      0.0654777324  0.129373772  0.2056324324
## fertility_rate                     -0.0147515380 -0.024175158  0.0088349658
## unis_colleges_number                0.0935155690  0.040487740  0.1601170253
## teaching_staff                     -0.4235288127  0.189444768 -0.2642041725
## students_enrolled                  -0.0749515978 -0.125239102  0.3446037987
## graduates_uni                       0.2969064949 -0.192220169 -0.0413222425
## ratio_grad_enrolled_unis           -0.0783260913 -0.026990899  0.0295704684
## TotalSchools                       -0.0974739992  0.257173572 -0.0490972126
## gradtotot57                        -0.0583458729  0.022290446 -0.0071334339
## gradtotot812                        0.0071975926 -0.009401281  0.0138215150
## employees_laborcontract_avg         0.0698125332  0.098498150 -0.2036714506
## employees_laborcontract_avg_topop  -0.0173954920 -0.155161872  0.0541945783
## avgwage                             0.0605070198  0.126376805 -0.0927488122
## activity_rate                      -0.0093185886 -0.154101205  0.0185061973
## labour_force_thousands              0.2084576409  0.321884996  0.1884395674
## employment_rate                     0.0350099916  0.121002701 -0.0275185552
## foreign_direct_investment_euro      0.3593413316 -0.167835736  0.3169568145
## books_pamphlets                     0.0794464067 -0.040725454 -0.3663331120
## accommodation_establishments        0.1344548124  0.532507548 -0.0448983418
## arrivals_accommodation              0.2046641657 -0.425390708 -0.4230215589
## revenue_accomm_lev                 -0.3052001492 -0.194673668  0.3861106586
## crimes                              0.1827975908 -0.199822382 -0.0917694112
## ecological_assets                  -0.0688416826 -0.002826552  0.0190181672
## water_supply2020                   -0.0069660492  0.115550268 -0.0261962050
## connected_to_wastewater_collecting  0.0008689237  0.101992718 -0.0298507629
## Pop_watersupplyregime              -0.0376494657 -0.012086685 -0.0002683304
## waste_thousandtons                 -0.3756266036 -0.099145235 -0.2242676573
## hospitals2021                      -0.4142926457 -0.165538552  0.1523689554
##                                             PC25          PC26          PC27
## Population2021                      0.0852439902  0.0791172178  0.5237957371
## fertility_rate                      0.0039106427 -0.0100638217 -0.0112871421
## unis_colleges_number                0.0940721568  0.0612522096  0.0449362052
## teaching_staff                      0.1595717161 -0.2937024411  0.1966910357
## students_enrolled                  -0.2195476356  0.4823372638  0.2154143092
## graduates_uni                       0.0468161826 -0.2505286199 -0.4165989338
```

```
## ratio_grad_enrolled_unis           0.0184181622  0.0003378177  0.0274883073
## TotalSchools                      -0.5061210677  0.2040150164 -0.3805876949
## gradtotot57                        0.0016714399 -0.0129735898  0.0048694753
## gradtotot812                      -0.0178018727 -0.0146793820 -0.0156542432
## employees_laborcontract_avg        0.6374265376  0.5066078164 -0.3091993501
## employees_laborcontract_avg_topop  -0.0199160070 -0.0901447228 -0.0160736699
## avgwage                           -0.0002416359  0.0497526207  0.0112651718
## activity_rate                      0.0304196767  0.0104378016  0.0151691322
## labour_force_thousands             0.2059455108 -0.4880843239  0.0715377025
## employment_rate                   -0.0236621077  0.0276649174 -0.0302060041
## foreign_direct_investment_euro    -0.1478529298 -0.1345584531 -0.0755563930
## books_pamphlets                   -0.3309351305  0.0005716673  0.2163856320
## accommodation_establishments      -0.0869148748  0.0788331992 -0.0041964401
## arrivals_accommodation             0.0491793992  0.0452399615  0.2645186511
## revenue_accomm_lev                 0.0842382901 -0.0894831851 -0.2268662044
## crimes                            -0.0169318943 -0.0260631710 -0.0123735264
## ecological_assets                 -0.0093271296  0.0069688993  0.0192815211
## water_supply2020                  -0.0235685186  0.0244925727  0.0239466505
## connected_to_wastewater_collecting -0.0369812432  0.0027659737 -0.0007149752
## Pop_watersupplyregime              0.0075788807  0.0007935117  0.0234186248
## waste_thousandtons                -0.1628121185 -0.1551360420 -0.1940579176
## hospitals2021                      0.1525697717 -0.0086488525  0.0585809637
##                                            PC28
## Population2021                     0.6128476922
## fertility_rate                    -0.0080424509
## unis_colleges_number               0.0448730022
## teaching_staff                     0.1287858709
## students_enrolled                 -0.3805720960
## graduates_uni                      0.2583817625
## ratio_grad_enrolled_unis          -0.0031092368
## TotalSchools                       0.0610317081
## gradtotot57                        0.0002864466
## gradtotot812                       0.0171326853
## employees_laborcontract_avg       -0.0301864303
## employees_laborcontract_avg_topop  0.0110645654
## avgwage                            0.0056638971
## activity_rate                     -0.0134870268
## labour_force_thousands            -0.5563890778
## employment_rate                    0.0157154941
## foreign_direct_investment_euro     0.1343258793
## books_pamphlets                   -0.1272072325
## accommodation_establishments       0.0135119023
## arrivals_accommodation            -0.1302307766
## revenue_accomm_lev                 0.0718153138
## crimes                            -0.0386269751
## ecological_assets                 -0.0044437954
## water_supply2020                  -0.0061153737
## connected_to_wastewater_collecting  0.0236414166
## Pop_watersupplyregime             -0.0134082821
## waste_thousandtons                 0.0380803984
## hospitals2021                     -0.1371227048
```

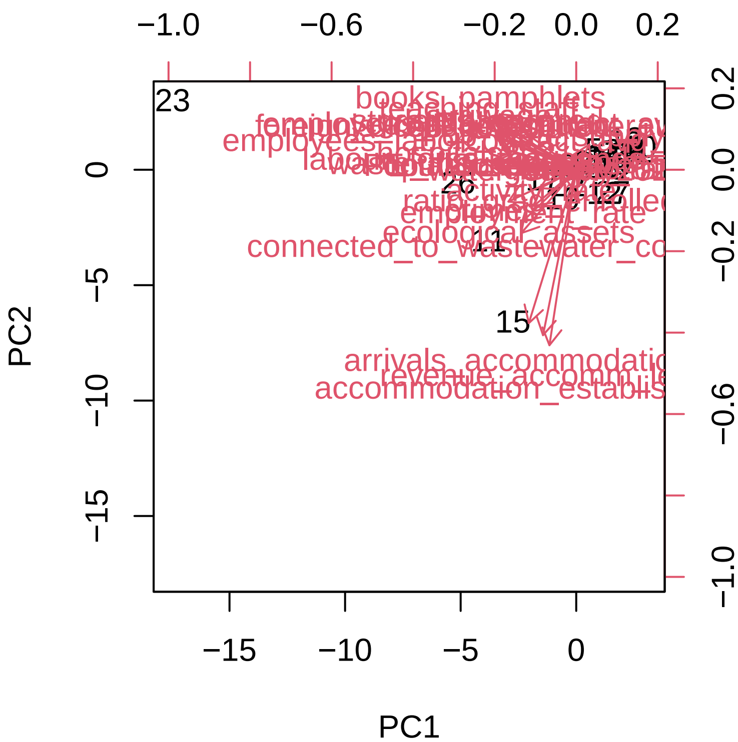
```
dim(pr.out$x)
```

```
## [1] 28 28
```

```
biplot(pr.out , scale = 0)
```




```
pr.out$rotation = -pr.out$rotation
pr.out$x = -pr.out$x
biplot(pr.out , scale = 0)
```




```
pr.out$sdev
```

```
##  [1] 3.931517e+00 1.673940e+00 1.422777e+00 1.408069e+00 1.136550e+00
##  [6] 1.035448e+00 8.969773e-01 8.363943e-01 6.559477e-01 6.144131e-01
## [11] 5.910615e-01 5.221696e-01 3.894370e-01 3.322988e-01 2.626092e-01
## [16] 1.989281e-01 1.607195e-01 1.192726e-01 1.019228e-01 8.704966e-02
## [21] 5.928014e-02 4.905328e-02 3.665443e-02 1.696045e-02 1.030414e-02
## [26] 5.546134e-03 2.585343e-03 1.228443e-15
```

```
#The variance explained by each principal component is obtained by squaring
#these:
pr.var <- pr.out$sdev^2
pr.var
```

```
##  [1] 1.545683e+01 2.802075e+00 2.024294e+00 1.982658e+00 1.291747e+00
##  [6] 1.072152e+00 8.045683e-01 6.995555e-01 4.302674e-01 3.775034e-01
## [11] 3.493538e-01 2.726611e-01 1.516612e-01 1.104225e-01 6.896358e-02
## [16] 3.957237e-02 2.583074e-02 1.422594e-02 1.038825e-02 7.577643e-03
## [21] 3.514135e-03 2.406225e-03 1.343547e-03 2.876568e-04 1.061753e-04
## [26] 3.075960e-05 6.684000e-06 1.509072e-30
```

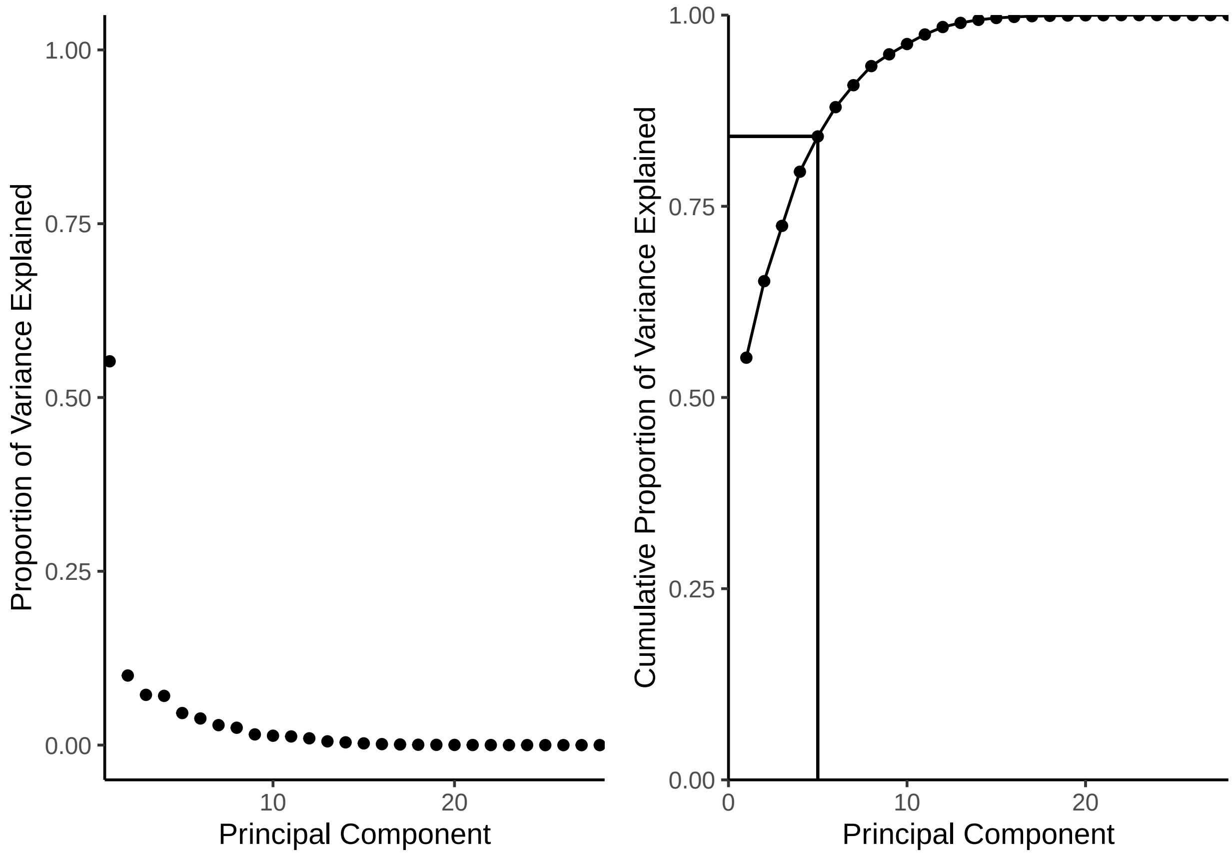
```
#explained var
pve <- pr.var / sum(pr.var)
pve
```

```
##  [1] 5.520296e-01 1.000741e-01 7.229622e-02 7.080921e-02 4.613381e-02
##  [6] 3.829113e-02 2.873458e-02 2.498412e-02 1.536669e-02 1.348227e-02
## [11] 1.247692e-02 9.737897e-03 5.416470e-03 3.943661e-03 2.462985e-03
## [16] 1.413299e-03 9.225266e-04 5.080694e-04 3.710089e-04 2.706301e-04
## [21] 1.255048e-04 8.593659e-05 4.798382e-05 1.027346e-05 3.791974e-06
## [26] 1.098557e-06 2.387143e-07 5.389545e-32
```

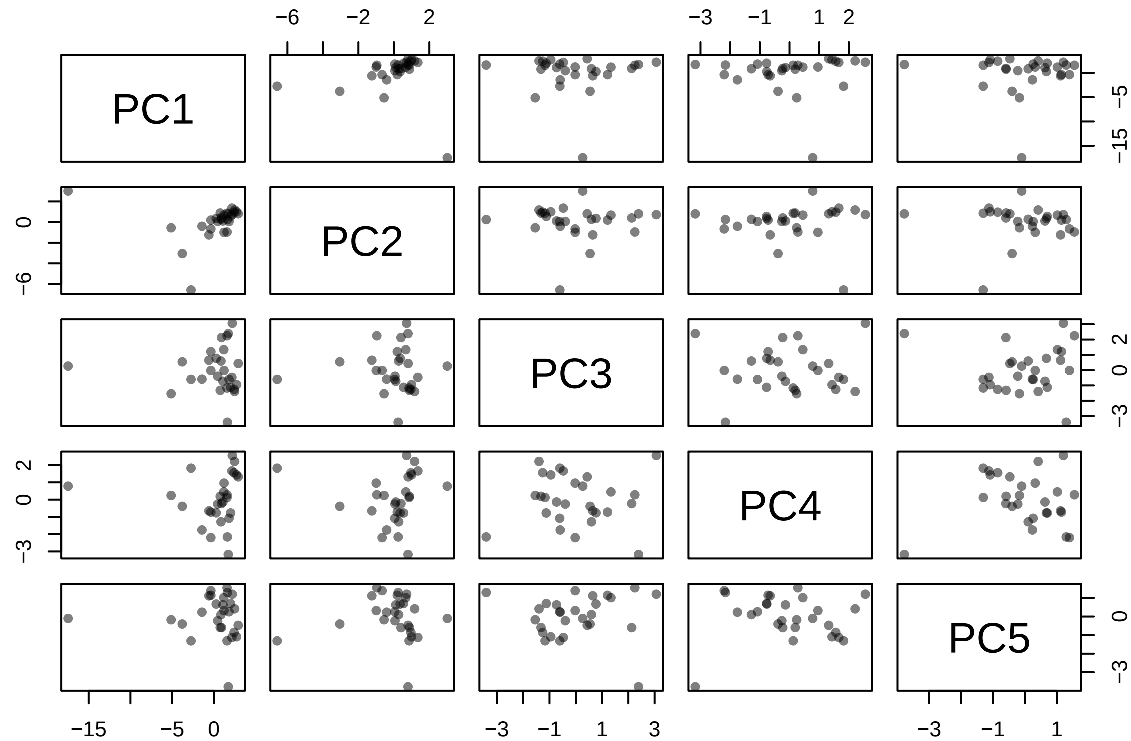
```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.1.3
```

```
p1 <- ggplot(data = as.data.frame(pve), aes(x = seq(1:28), y = pve)) +
        geom_point() +
        #geom_line() +
        #geom_segment(aes(x = 5, y = 0, xend = 5, yend = pve[5])) +
        #geom_segment(aes(x = 0, y = pve[5], xend = 5, yend = pve[5])) +
        xlab("Principal Component") +
        ylab("Proportion of Variance Explained") +
        scale_x_continuous(expand = c(0.01,0)) +
        scale_y_continuous(expand = c(0.05, 0), limits = c(0,1)) +
        theme_classic()

p2 <- ggplot(data = as.data.frame(cumsum(pve)), aes(x = seq(1:28), y = cumsum(pve))) +
        geom_point() +
        geom_line() +
        geom_segment(aes(x = 5, y = 0, xend = 5, yend = cumsum(pve)[5])) +
        geom_segment(aes(x = 0, y = cumsum(pve)[5], xend = 5, yend = cumsum(pve)[5])) +
        xlab("Principal Component") +
        ylab("Cumulative Proportion of Variance Explained") +
        scale_x_continuous(expand = c(0,0)) +
        scale_y_continuous(expand = c(0, 0)) +
        theme_classic()

ggarrange(p1, p2)
```




```
comp <- data.frame(pr.out$x[,1:5])

#write.csv(comp, "C:/Users/User/Documents/UNITN/Computational social science/bulgariafirst5PCs.csv")
plot(comp, pch=16, col=rgb(0,0,0,0.5))
```




```
#outlier in all the dimensions, and some grouping in the different projections.

summary(pr.out)
```

```
## Importance of components:
##                          PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     3.932 1.6739 1.4228 1.40807 1.13655 1.03545 0.89698
## Proportion of Variance 0.552 0.1001 0.0723 0.07081 0.04613 0.03829 0.02873
## Cumulative Proportion  0.552 0.6521 0.7244 0.79521 0.84134 0.87963 0.90837
##                            PC8     PC9    PC10    PC11    PC12    PC13    PC14
## Standard deviation     0.83639 0.65595 0.61441 0.59106 0.52217 0.38944 0.33230
## Proportion of Variance 0.02498 0.01537 0.01348 0.01248 0.00974 0.00542 0.00394
## Cumulative Proportion  0.93335 0.94872 0.96220 0.97468 0.98442 0.98983 0.99378
##                           PC15    PC16    PC17    PC18    PC19    PC20    PC21
## Standard deviation     0.26261 0.19893 0.16072 0.11927 0.10192 0.08705 0.05928
## Proportion of Variance 0.00246 0.00141 0.00092 0.00051 0.00037 0.00027 0.00013
## Cumulative Proportion  0.99624 0.99765 0.99858 0.99908 0.99945 0.99973 0.99985
##                           PC22    PC23    PC24   PC25     PC26     PC27
## Standard deviation     0.04905 0.03665 0.01696 0.0103 0.005546 0.002585
## Proportion of Variance 0.00009 0.00005 0.00001 0.0000 0.000000 0.000000
## Cumulative Proportion  0.99994 0.99998 0.99999 1.0000 1.000000 1.000000
##                             PC28
## Standard deviation     1.228e-15
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```r
bul_transform <- as.data.frame(-pr.out$x[,1:5])
```

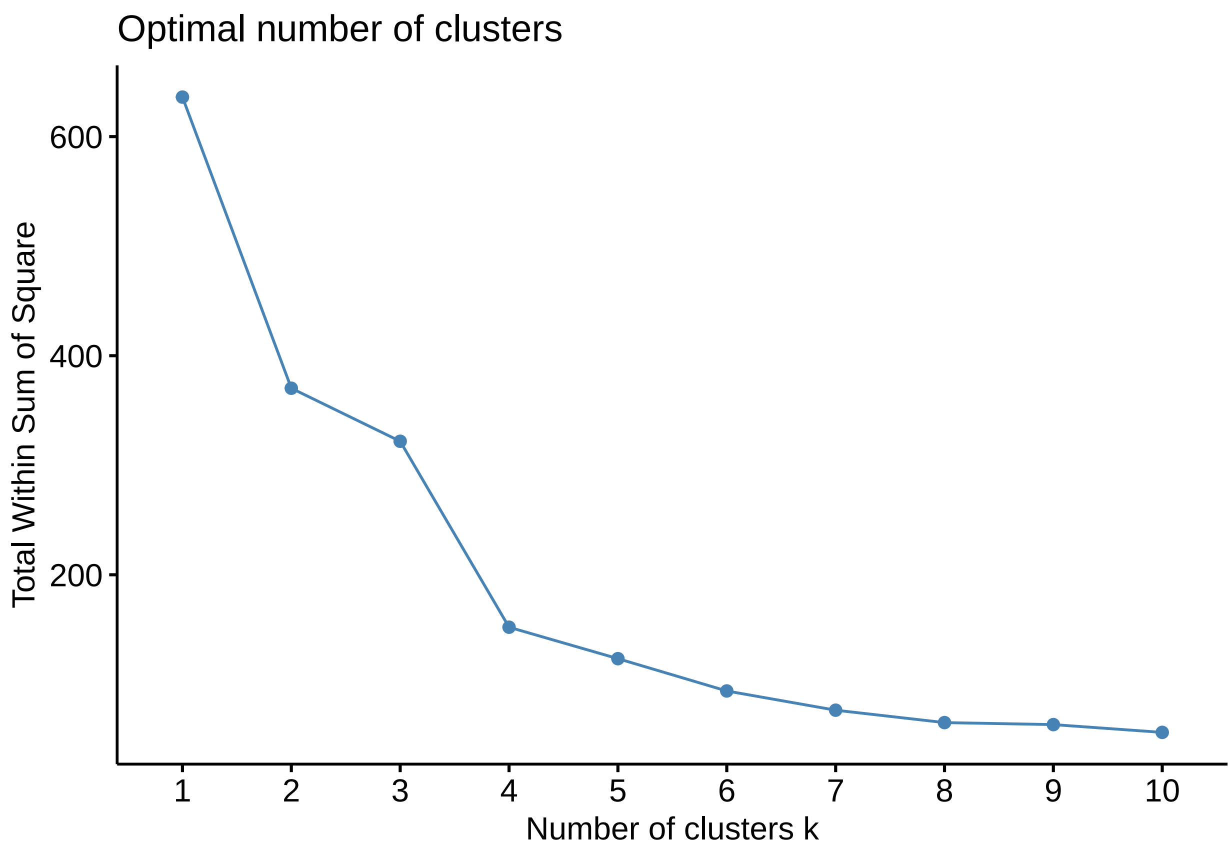
#Clustering k-means

```r
library(factoextra)
```

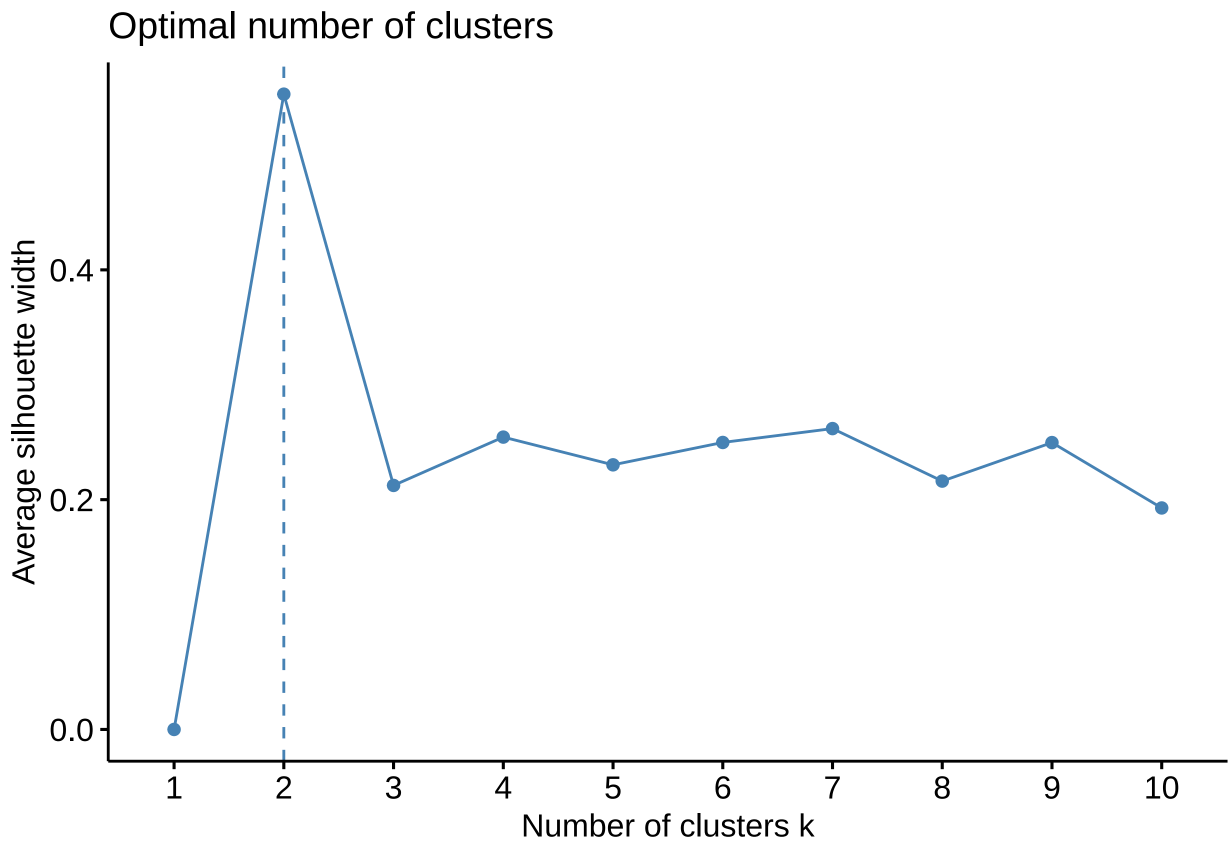
```
## Warning: package 'factoextra' was built under R version 4.1.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

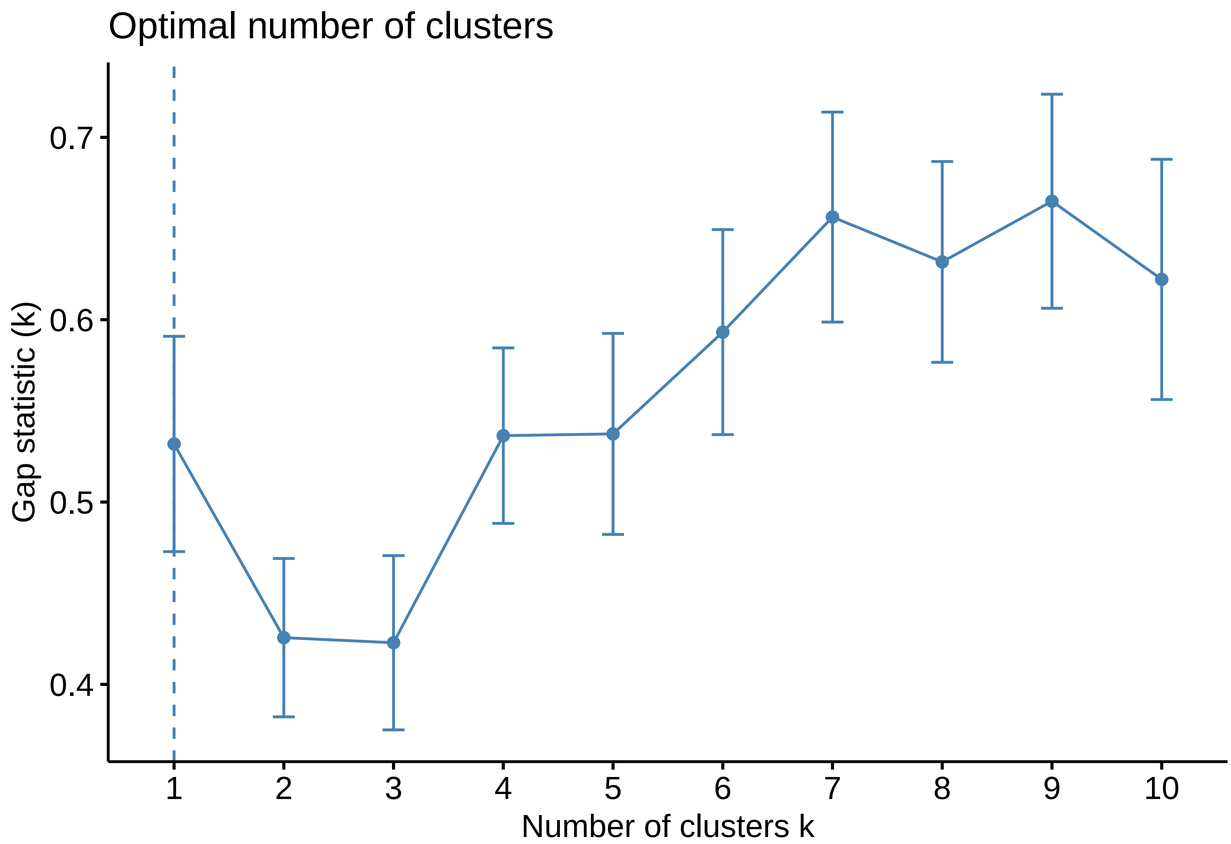
```r
fviz_nbclust(bul_transform, kmeans, method = 'wss')
```



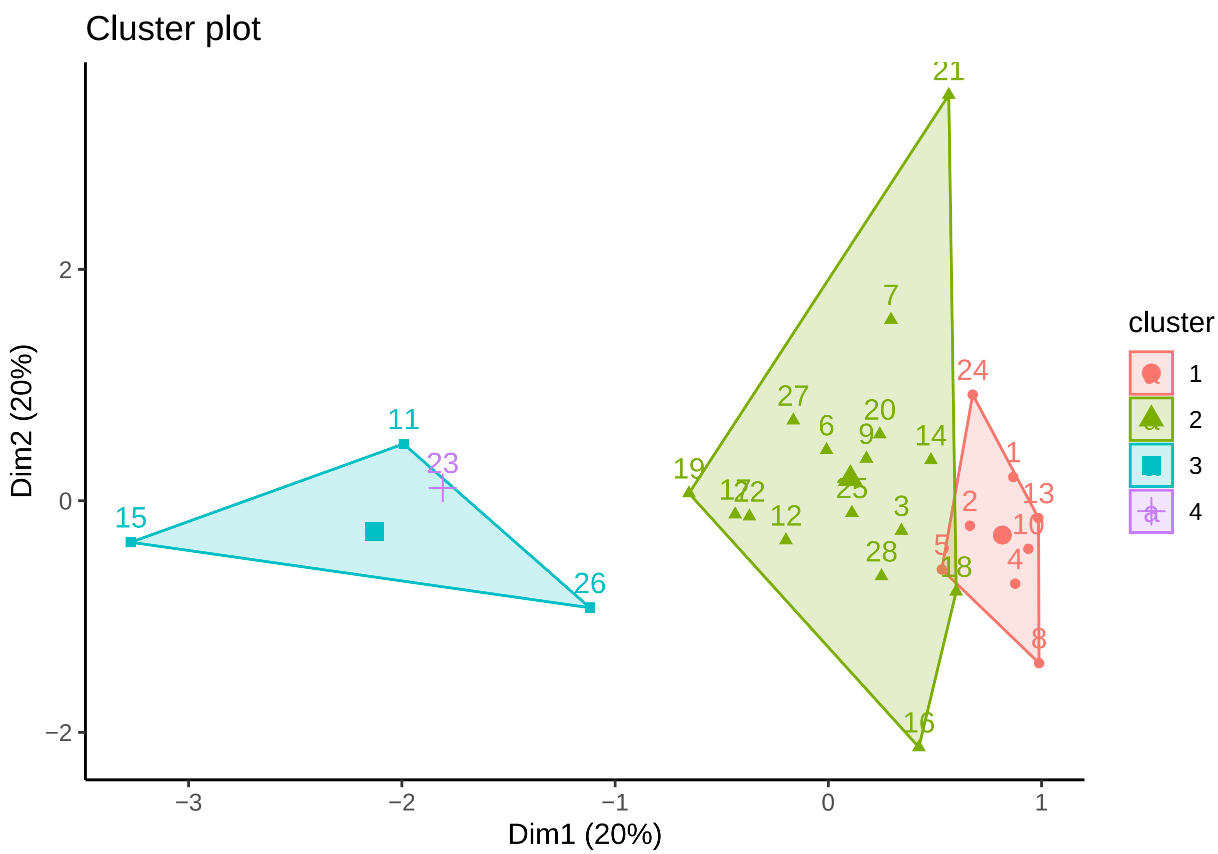

```r
fviz_nbclust(bul_transform, kmeans, method = 'silhouette')
```




```
fviz_nbclust(bul_transform, kmeans, method = 'gap_stat')
```



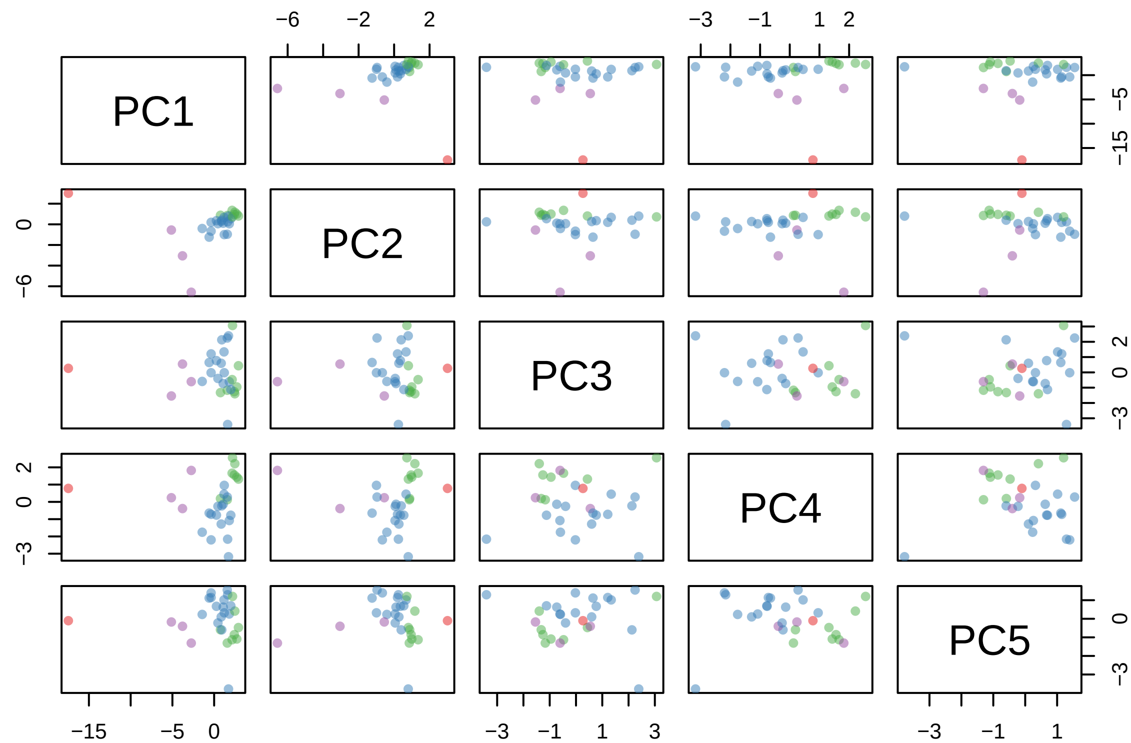

```
k = 4
kmeans_bul = kmeans(bul_transform, centers = k, nstart = 20)
fviz_cluster(kmeans_bul, data = bul_transform) + theme_classic()
```




```
set.seed(21)
k <- kmeans(comp, 4, nstart=25, iter.max=1000)
library(RColorBrewer)
library(scales)
```

```
## Warning: package 'scales' was built under R version 4.1.2
```

```
palette(alpha(brewer.pal(9,'Set1'), 0.5))
plot(comp, col=k$clust, pch=16)
```




```
clusters_kmeans4 <- k$cluster
```

```
bul_with_labels <- bul[,]
bul_with_labels$kmeans4 <- as.factor(clusters_kmeans4)

scaled <- as.data.frame(scale(bul_with_labels[,-c(1,30)]))
scaled$Region <- bul$Region
scaled$kmeans4 <- as.factor(clusters_kmeans4)
```

Random forest with previously found clusters. Turning to supervised method to see variable importance

```
library(randomForest)
```

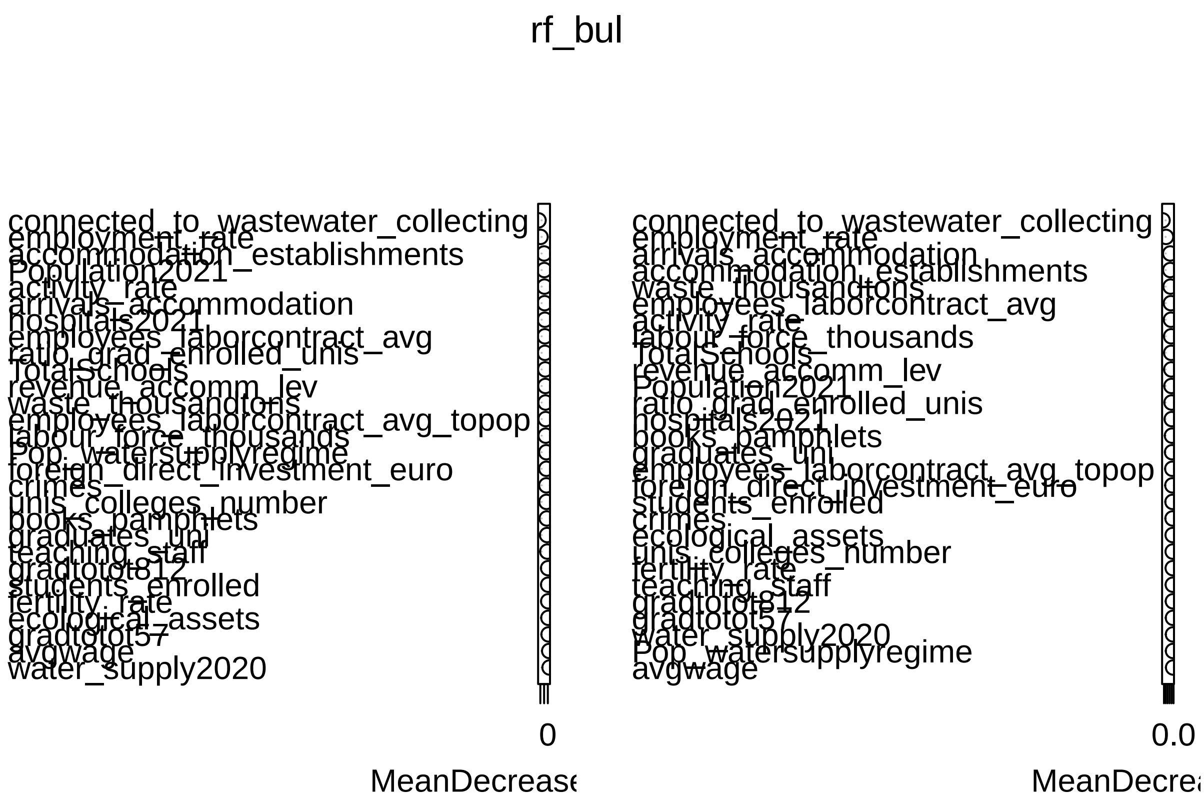
```
## Warning: package 'randomForest' was built under R version 4.1.3

## randomForest 4.7-1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
set.seed(21)
rf_bul <- randomForest(kmeans4 ~.-Region, data = scaled, mtry = 9, importance = TRUE)
importance(rf_bul)
```

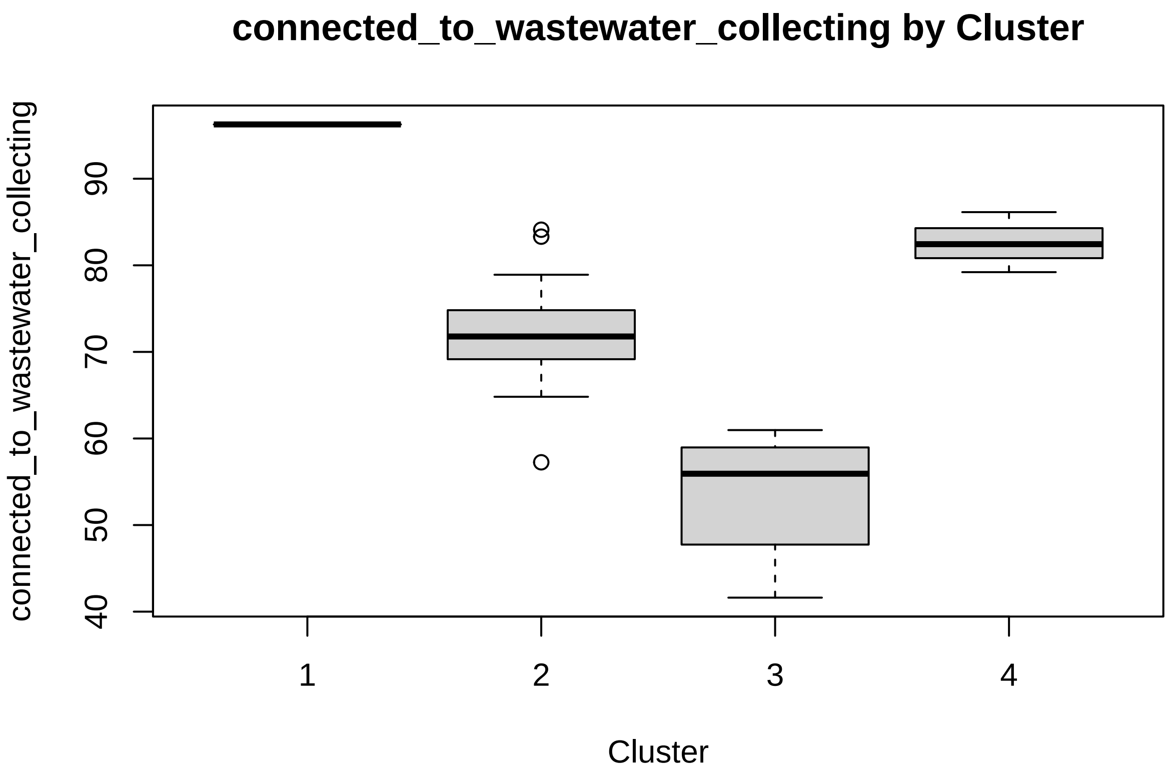
```
##                                     1          2          3          4
## Population2021                      0  3.0944974  2.2348611  4.9154188
## fertility_rate                      0  1.3806214 -1.7026641  0.8170415
## unis_colleges_number                0 -0.3044071  1.8447912  2.2107259
## teaching_staff                      0  1.2794980 -1.0010015  1.3440623
## students_enrolled                   0 -1.3137853 -0.2116132  1.7372705
## graduates_uni                       0 -1.7851431  2.2977510  1.9042342
## ratio_grad_enrolled_unis            0  3.7999063  3.4299788  0.7750618
## TotalSchools                        0  2.4348926 -0.5582108  4.5349450
## gradtotot57                         0 -0.2055544 -1.6759457 -0.7849477
## gradtotot812                        0 -0.8278552  0.6871270  0.8170415
## employees_laborcontract_avg         0  1.7473011  0.9631427  4.2273121
## employees_laborcontract_avg_topop   0 -0.4520301  3.5997366  0.0000000
## avgwage                             0 -1.8574279 -1.0010015 -0.4473031
## activity_rate                       0  2.8721927  4.7805602  0.5775428
## labour_force_thousands              0  2.1372957  0.3838362  3.3709993
## employment_rate                     0  7.8748633 10.1021484  3.3013910
## foreign_direct_investment_euro      0  0.9513106  0.4544046  2.8072960
## books_pamphlets                     0  0.6398717 -1.6373653  3.1168189
## accommodation_establishments        0 -0.7969141  5.5761435  3.5213750
## arrivals_accommodation              0  1.7293872  3.5332769  3.4417392
## revenue_accomm_lev                  0 -0.4388249  4.3726237  2.5461701
## crimes                              0  1.1874604  1.7372705  0.8534238
## ecological_assets                   0 -1.0268032 -1.8828449  2.3702273
## water_supply2020                    0 -1.7963290 -1.7213102 -1.5109947
## connected_to_wastewater_collecting  0  9.3916299 12.8024755  3.9666081
## Pop_watersupplyregime               0  1.3522689 -0.8170415  1.9488101
## waste_thousandtons                  0  1.9889544  0.4222991  4.6249729
## hospitals2021                       0  3.2940716  1.6973247  3.3709993
##                                     MeanDecreaseAccuracy MeanDecreaseGini
## Population2021                                 4.7267958        0.5494781
## fertility_rate                                -0.5044853        0.2170927
## unis_colleges_number                           1.8436906        0.2270247
## teaching_staff                                 0.6849340        0.2004856
## students_enrolled                             -0.4151720        0.2761430
## graduates_uni                                  0.9804711        0.3907602
## ratio_grad_enrolled_unis                       3.8596197        0.5313134
## TotalSchools                                   3.4875730        0.6096154
## gradtotot57                                   -0.9716918        0.1587595
## gradtotot812                                  -0.4082056        0.1773898
## employees_laborcontract_avg                    3.9413252        0.7016275
## employees_laborcontract_avg_topop              2.9257927        0.3366603
## avgwage                                       -1.8354708        0.1150994
## activity_rate                                  4.4702664        0.6711244
## labour_force_thousands                         2.8507184        0.6261772
## employment_rate                                9.7666767        1.9700773
## foreign_direct_investment_euro                 1.9833337        0.3112777
## books_pamphlets                                1.4089195        0.4152488
## accommodation_establishments                   4.9813727        0.8607068
```

```
## arrivals_accommodation                      4.2169647        0.8758904
## revenue_accomm_lev                          3.3142306        0.5875183
## crimes                                      1.9780300        0.2709768
## ecological_assets                          -0.7022895        0.2333642
## water_supply2020                           -2.3102545        0.1524638
## connected_to_wastewater_collecting         12.4128684        2.8783794
## Pop_watersupplyregime                       2.0780875        0.1480691
## waste_thousandtons                          3.1034616        0.7865394
## hospitals2021                               3.9628511        0.4757368
```

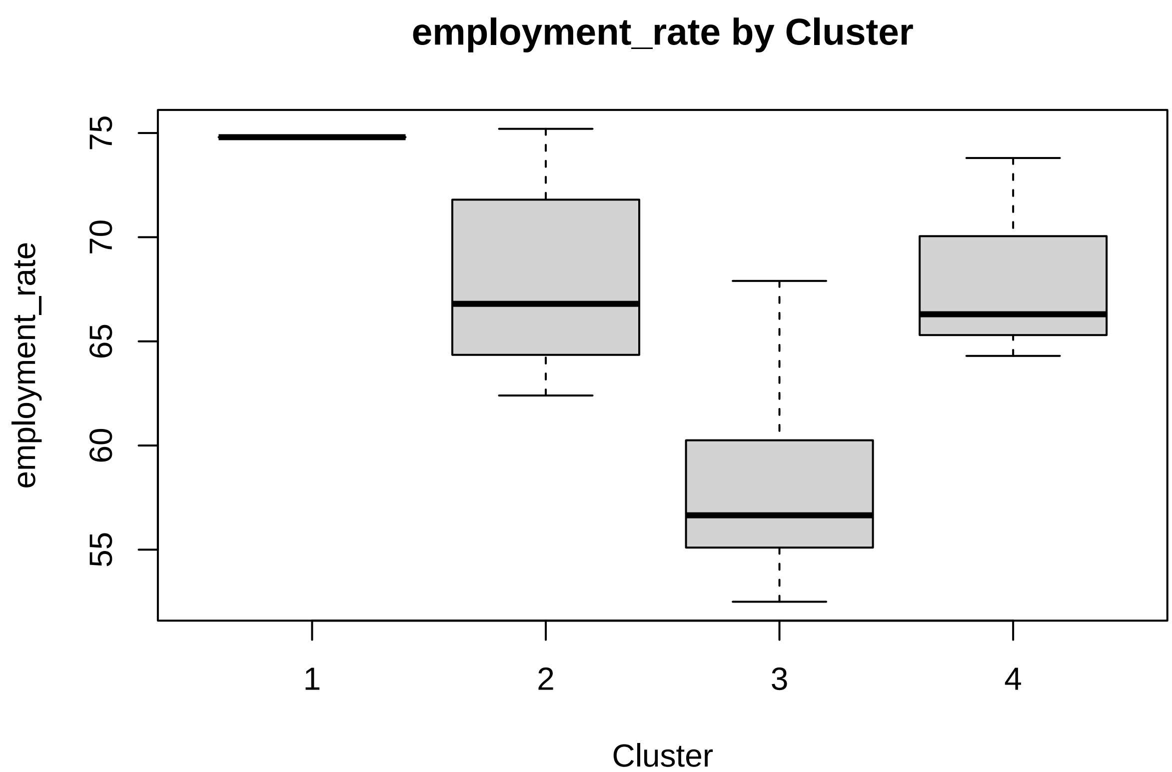
```r
varImpPlot(rf_bul)
```



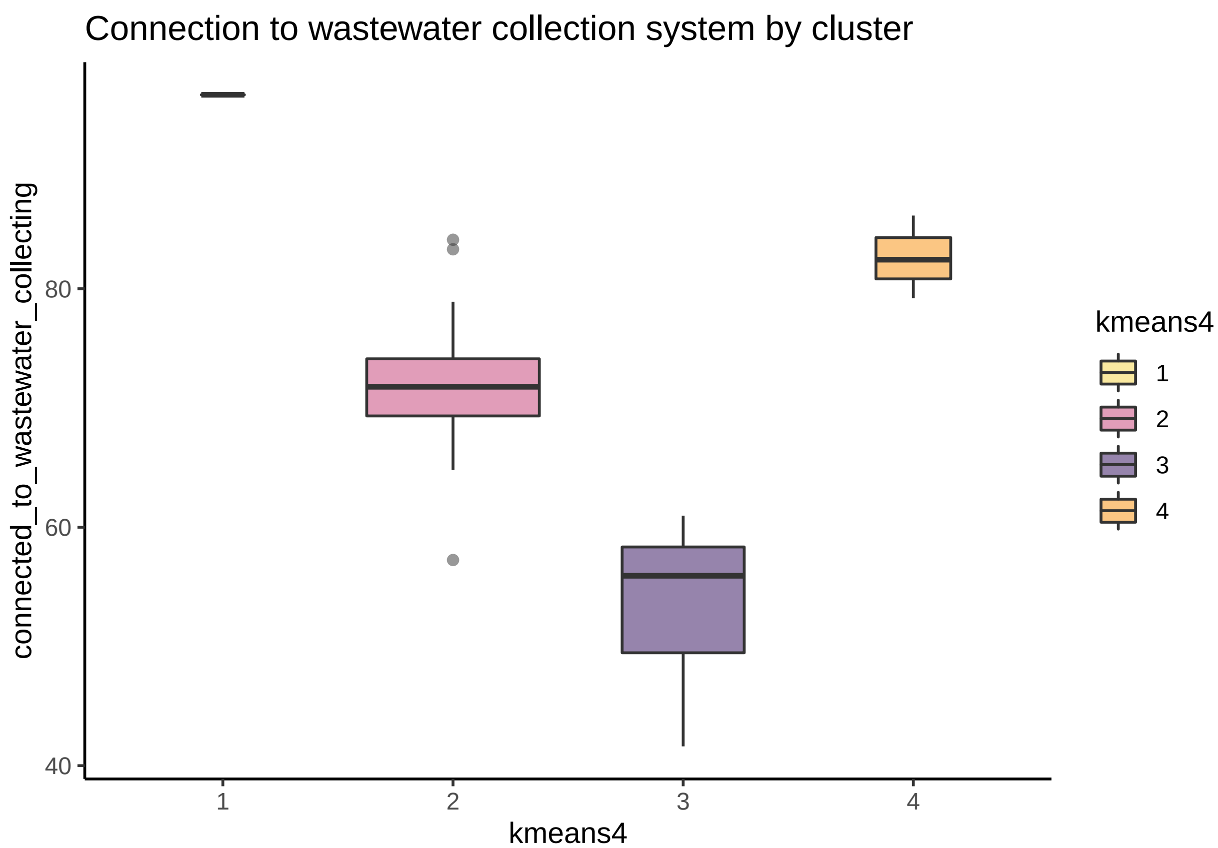

```r
boxplot(bul_with_labels$connected_to_wastewater_collecting ~ bul_with_labels$kmeans4,
        xlab='Cluster', ylab='connected_to_wastewater_collecting',
        main='connected_to_wastewater_collecting by Cluster')
```




```
boxplot(bul$employment_rate ~ bul_with_labels$kmeans4,
        xlab='Cluster', ylab='employment_rate',
        main='employment_rate by Cluster')
```



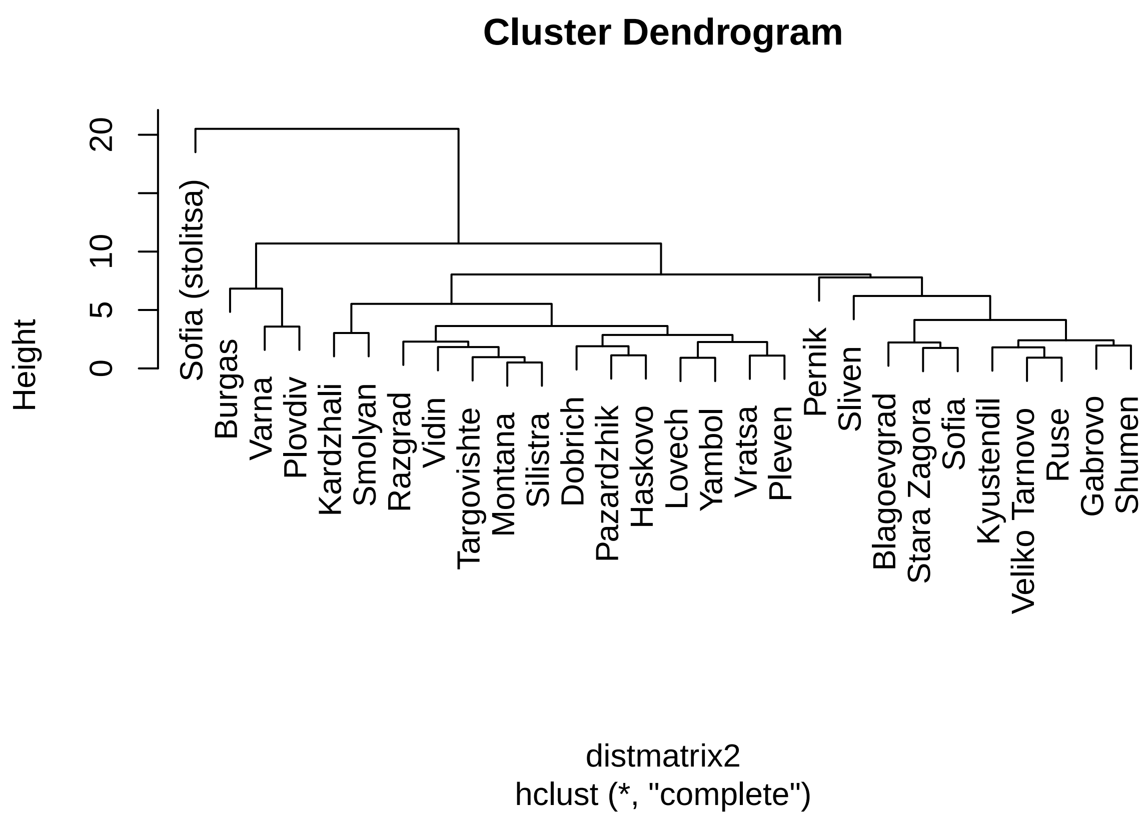

```
ggplot(bul_with_labels, aes(x=kmeans4, y=connected_to_wastewater_collecting, fill = kmeans4)) +
    geom_boxplot(varwidth = TRUE, alpha=0.5) +
    theme(legend.position="none") +
        scale_fill_manual(values = c("#f6d543", "#c53c74", "#2f0a5b", "#f98e09")) +
        theme_classic() +
        ggtitle("Connection to wastewater collection system by cluster")
```




```
rownames(scaled) <- scaled$Region
rownames(bul_transform) <- scaled$Region
```

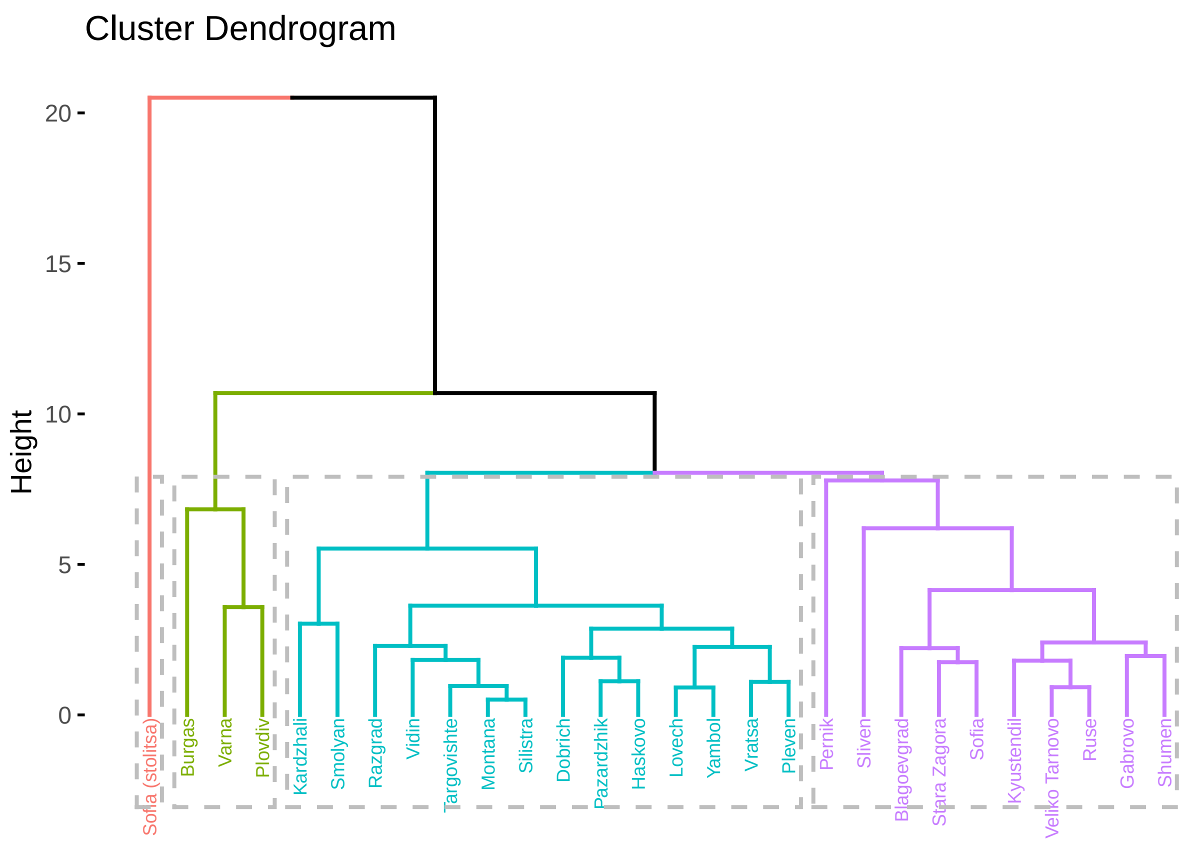
#Hierarchical clustering

```
set.seed(21)
distmatrix2 <- dist(bul_transform)
hc2 <- hclust(distmatrix2, method = "complete")
plot(hc2, labels = scaled$Region)
```




```
labs_pca_hc4 <- cutree(hc2, 4)


#hc2$labels <- as.vector(bul_with_labels[hc2$order, 'Region'])
fviz_dend(hc2,cex = 0.5, k = 4,rect = TRUE)
```

```
## Warning: ‘guides(<scale> = FALSE)‘ is deprecated. Please use ‘guides(<scale> =
## "none")‘ instead.
```




```r
#See var importance when hclust
set.seed(21)
rf_bul2 <- randomForest(labs_pca_hc4 ~.-Region - kmeans4, data = scaled, mtry = 9, importance = TRUE)
```

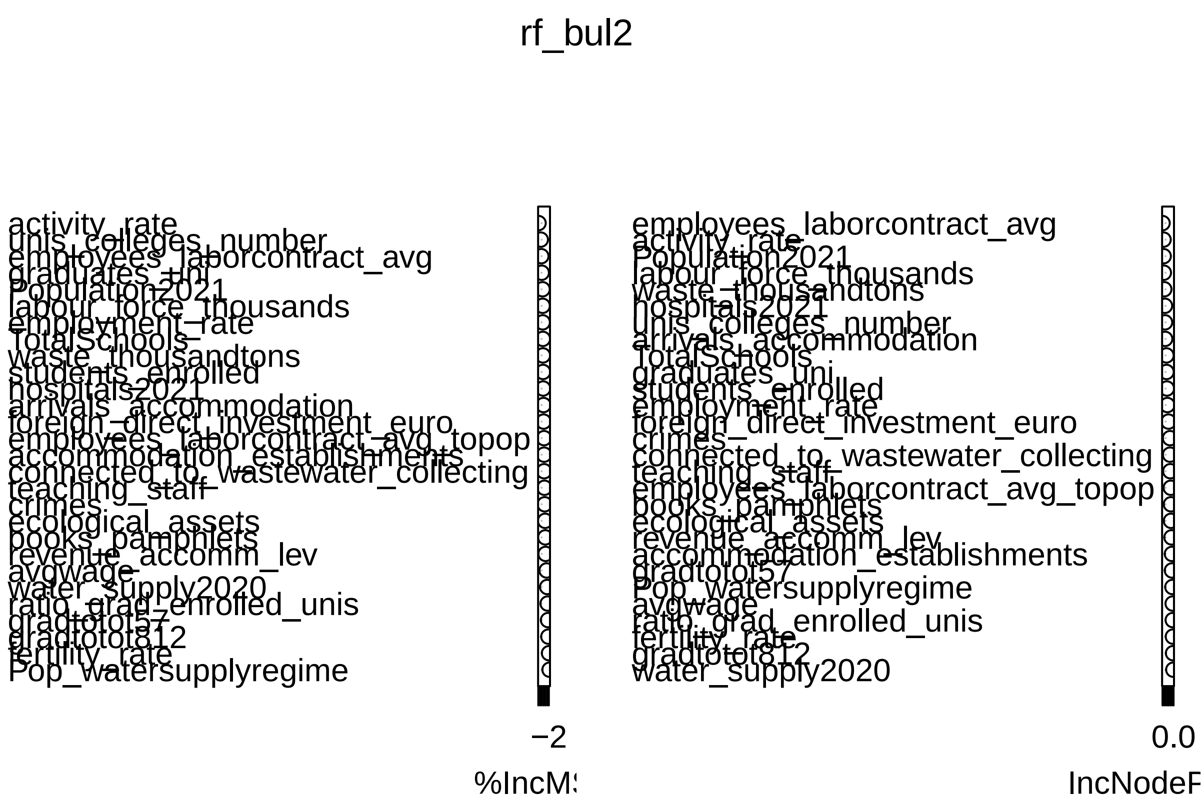
```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```

```r
importance(rf_bul2)
```

```
##                                       %IncMSE IncNodePurity
## Population2021                      5.2033142    1.24951975
## fertility_rate                     -2.2233084    0.08464824
## unis_colleges_number                7.2046531    1.07999555
## teaching_staff                      2.9962718    0.40954240
## students_enrolled                   4.6460614    0.84295702
## graduates_uni                       6.1058844    0.88974869
## ratio_grad_enrolled_unis           -0.6388594    0.10222527
## TotalSchools                        4.7879051    0.89658836
## gradtotot57                        -1.0865686    0.20081573
## gradtotot812                       -1.4547252    0.06458656
## employees_laborcontract_avg         6.5768174    1.39392990
## employees_laborcontract_avg_topop   3.4811170    0.35114744
## avgwage                             0.5529313    0.11122043
## activity_rate                       9.7289770    1.26258297
## labour_force_thousands              5.0603443    1.22931283
```

```
## employment_rate                       5.0000539    0.78126165
## foreign_direct_investment_euro        4.0084955    0.63736012
## books_pamphlets                       2.0376067    0.33069871
## accommodation_establishments          3.4290073    0.20594652
## arrivals_accommodation                4.1233131    1.06373374
## revenue_accomm_lev                    1.7112880    0.26007155
## crimes                                2.7466196    0.44704275
## ecological_assets                     2.2691104    0.32256053
## water_supply2020                     -0.1584301    0.03034158
## connected_to_wastewater_collecting    3.4168809    0.44510818
## Pop_watersupplyregime                -2.8136707    0.14672062
## waste_thousandtons                    4.7640144    1.17588837
## hospitals2021                         4.4972784    1.09049692
```

```
varImpPlot(rf_bul2)
```



Align clusters of the two methods

```
i = labs_pca_hc4 == 1
j = labs_pca_hc4 == 3
v = labs_pca_hc4 == 4
labs_pca_hc4[i] = 3
labs_pca_hc4[j] = 4
labs_pca_hc4[v] = 1
```

```r
labs_pca_hc4 == clusters_kmeans4
```

```
##            Vidin          Vratsa          Lovech         Montana 
##             TRUE            TRUE           FALSE            TRUE 
##           Pleven  Veliko Tarnovo         Gabrovo         Razgrad 
##             TRUE            TRUE            TRUE            TRUE 
##             Ruse        Silistra           Varna         Dobrich 
##             TRUE            TRUE            TRUE           FALSE 
##      Targovishte          Shumen          Burgas          Sliven 
##             TRUE            TRUE            TRUE            TRUE 
##     Stara Zagora          Yambol      Blagoevgrad      Kyustendil 
##             TRUE           FALSE            TRUE            TRUE 
##           Pernik           Sofia Sofia (stolitsa)       Kardzhali 
##             TRUE            TRUE            TRUE            TRUE 
##       Pazardzhik         Plovdiv         Smolyan         Haskovo 
##            FALSE            TRUE           FALSE           FALSE
```

```r
sum(labs_pca_hc4 == clusters_kmeans4)
```

```
## [1] 22
```

```r
table(labs_pca_hc4, clusters_kmeans4)
```

```
##             clusters_kmeans4
## labs_pca_hc4  1  2  3  4
##            1  1  0  0  0
##            2  0 10  0  0
##            3  0  6  8  0
##            4  0  0  0  3
```

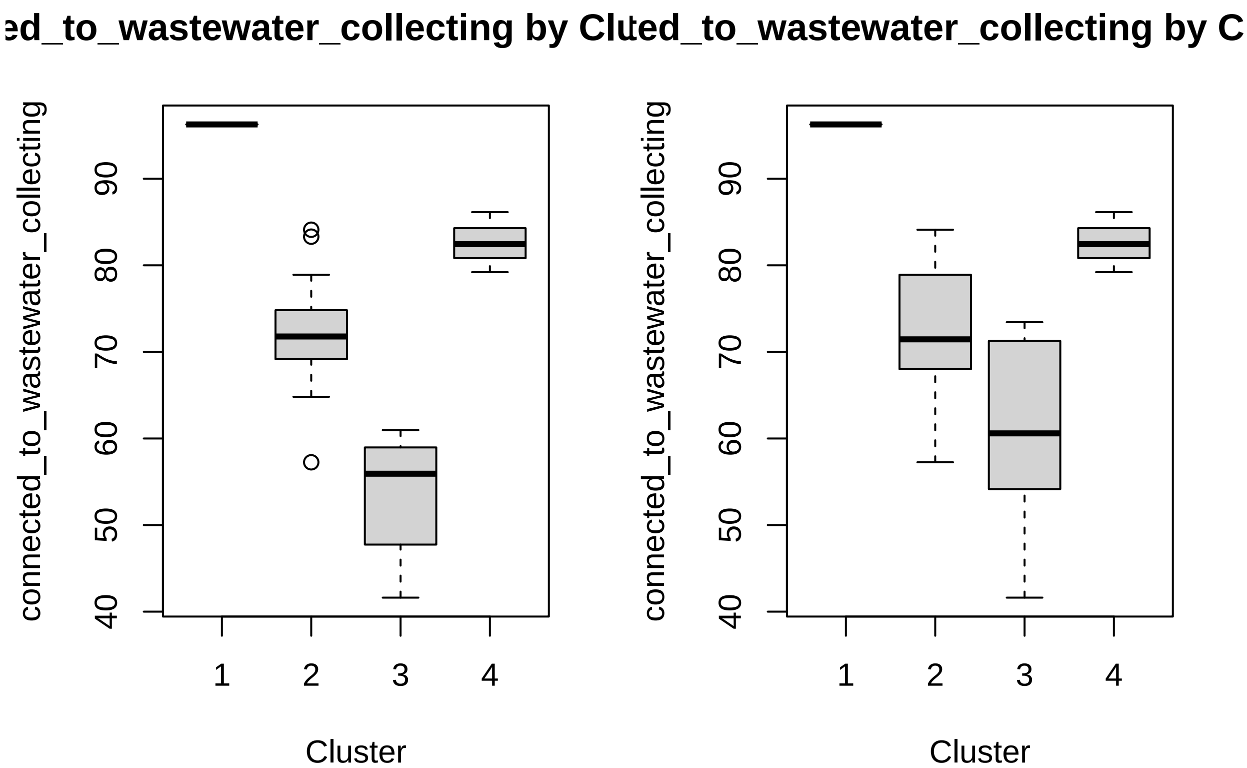
```r
bul_with_labels$labs_pca_hc4 <- as.factor(labs_pca_hc4)
scaled$labs_pca_hc4 <- as.factor(labs_pca_hc4)
```

```r
mclust::adjustedRandIndex(clusters_kmeans4, labs_pca_hc4)
```

```
## [1] 0.3964696
```

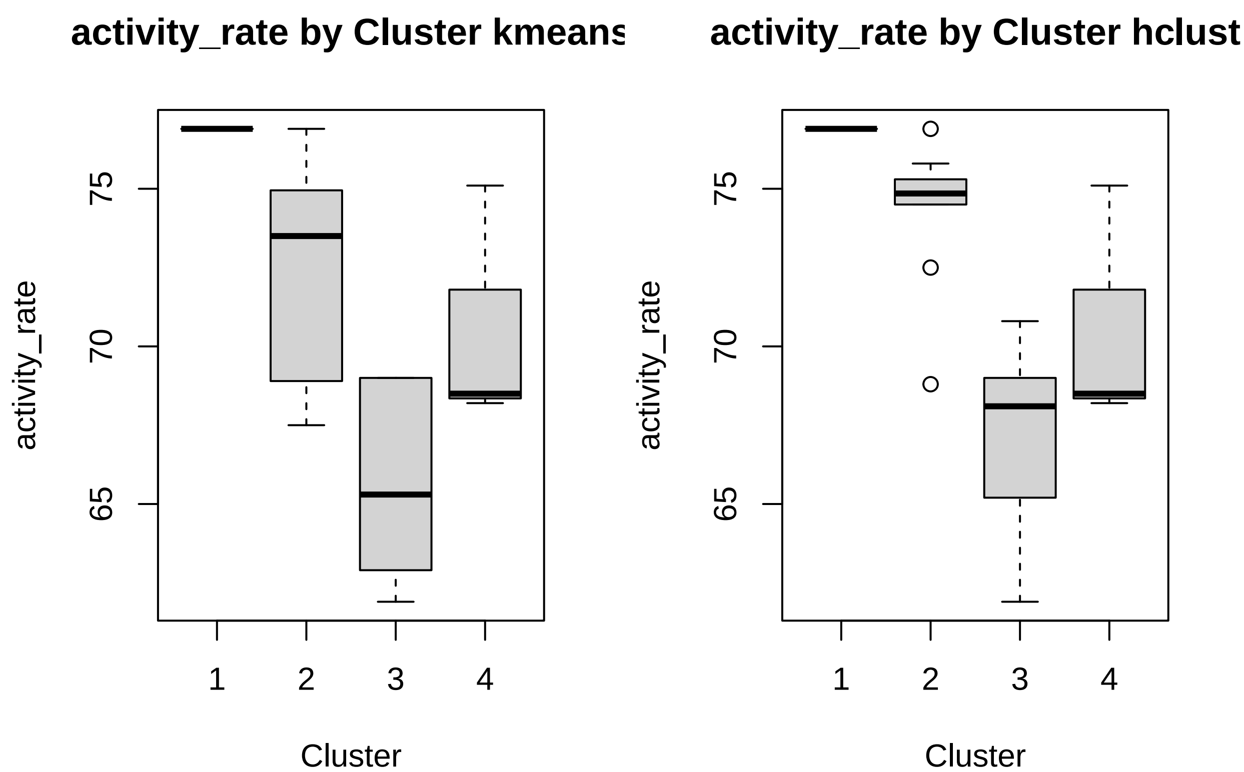
```r
#describe what it is and how it is calculated
```

```r
par(mfrow = c(1,2))

boxplot(bul_with_labels$connected_to_wastewater_collecting ~ bul_with_labels$kmeans4,
        xlab='Cluster', ylab='connected_to_wastewater_collecting',
        main='connected_to_wastewater_collecting by Cluster kmeans')
boxplot(bul_with_labels$connected_to_wastewater_collecting ~ bul_with_labels$labs_pca_hc4,
        xlab='Cluster', ylab='connected_to_wastewater_collecting',
        main='connected_to_wastewater_collecting by Cluster hclust')
```



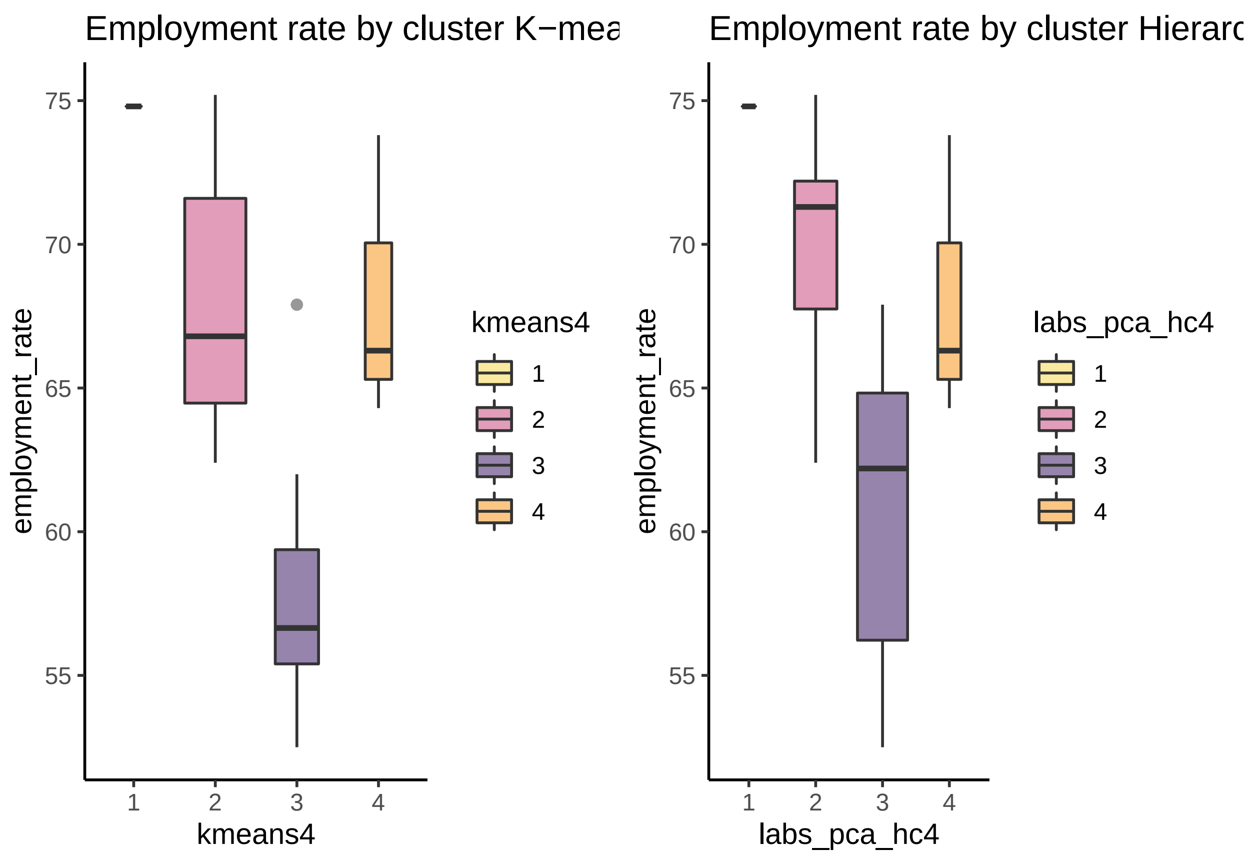

```
boxplot(bul$employment_rate ~ bul_with_labels$kmeans4,
        xlab='Cluster', ylab='employment_rate',
        main='employment_rate by Cluster kmeans')

boxplot(bul$employment_rate ~ bul_with_labels$labs_pca_hc4,
        xlab='Cluster', ylab='employment_rate',
        main='employment_rate by Cluster hclust')
```


```
boxplot(bul$activity_rate ~ bul_with_labels$kmeans4,
        xlab='Cluster', ylab='activity_rate',
        main='activity_rate by Cluster kmeans')

boxplot(bul$activity_rate ~ bul_with_labels$labs_pca_hc4,
        xlab='Cluster', ylab='activity_rate',
        main='activity_rate by Cluster hclust')
```

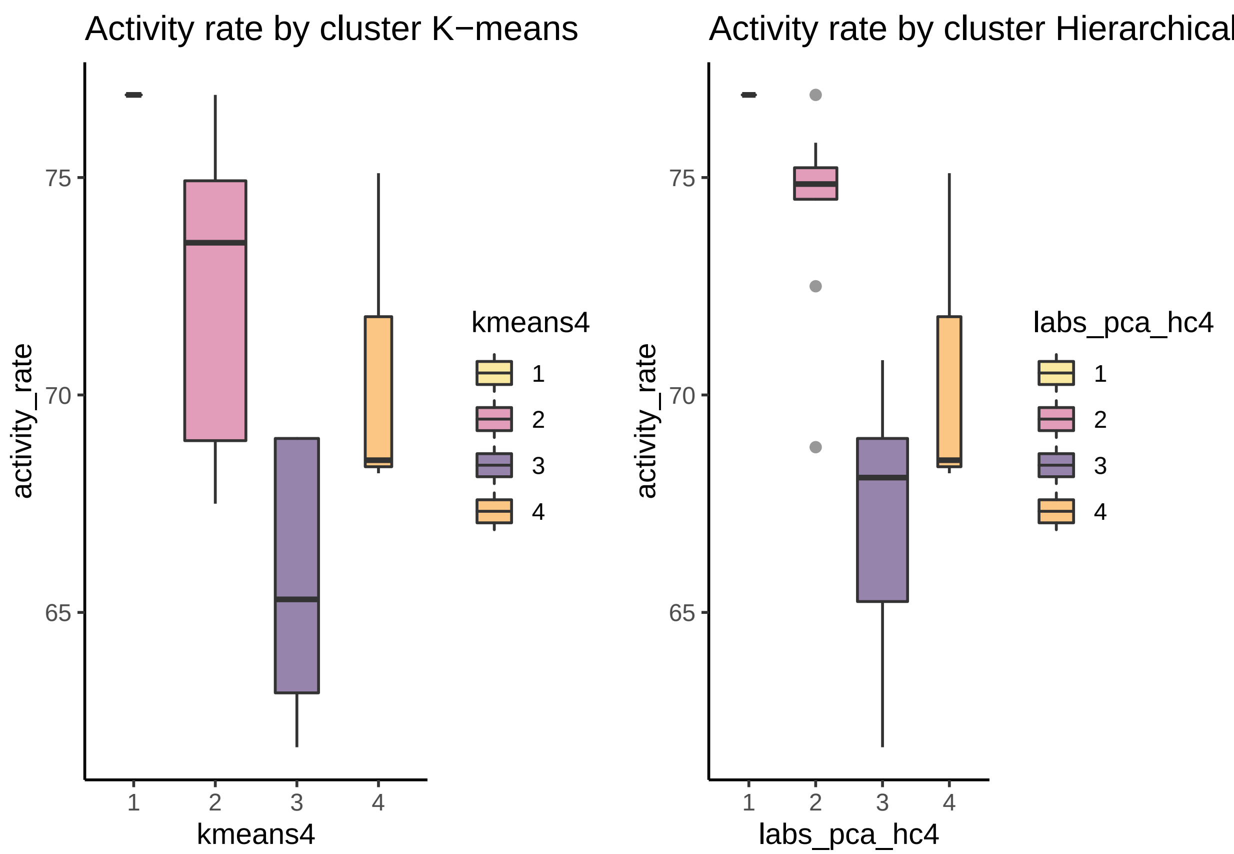



```
par(mfrow = c(1,1))
```

```
p1 <- ggplot(bul_with_labels, aes(x=kmeans4, y=employment_rate, fill = kmeans4)) +
    geom_boxplot(varwidth = TRUE, alpha=0.5) +
    theme(legend.position="none") +
        scale_fill_manual(values = c("#f6d543", "#c53c74", "#2f0a5b", "#f98e09")) +
        theme_classic() +
        ggtitle("Employment rate by cluster K-means")
p2 <- ggplot(bul_with_labels, aes(x=labs_pca_hc4, y=employment_rate, fill = labs_pca_hc4)) +
    geom_boxplot(varwidth = TRUE, alpha=0.5) +
    theme(legend.position="none") +
        scale_fill_manual(values = c("#f6d543", "#c53c74", "#2f0a5b", "#f98e09")) +
        theme_classic() +
        ggtitle("Employment rate by cluster Hierarchical clustering")
ggarrange(p1, p2)
```



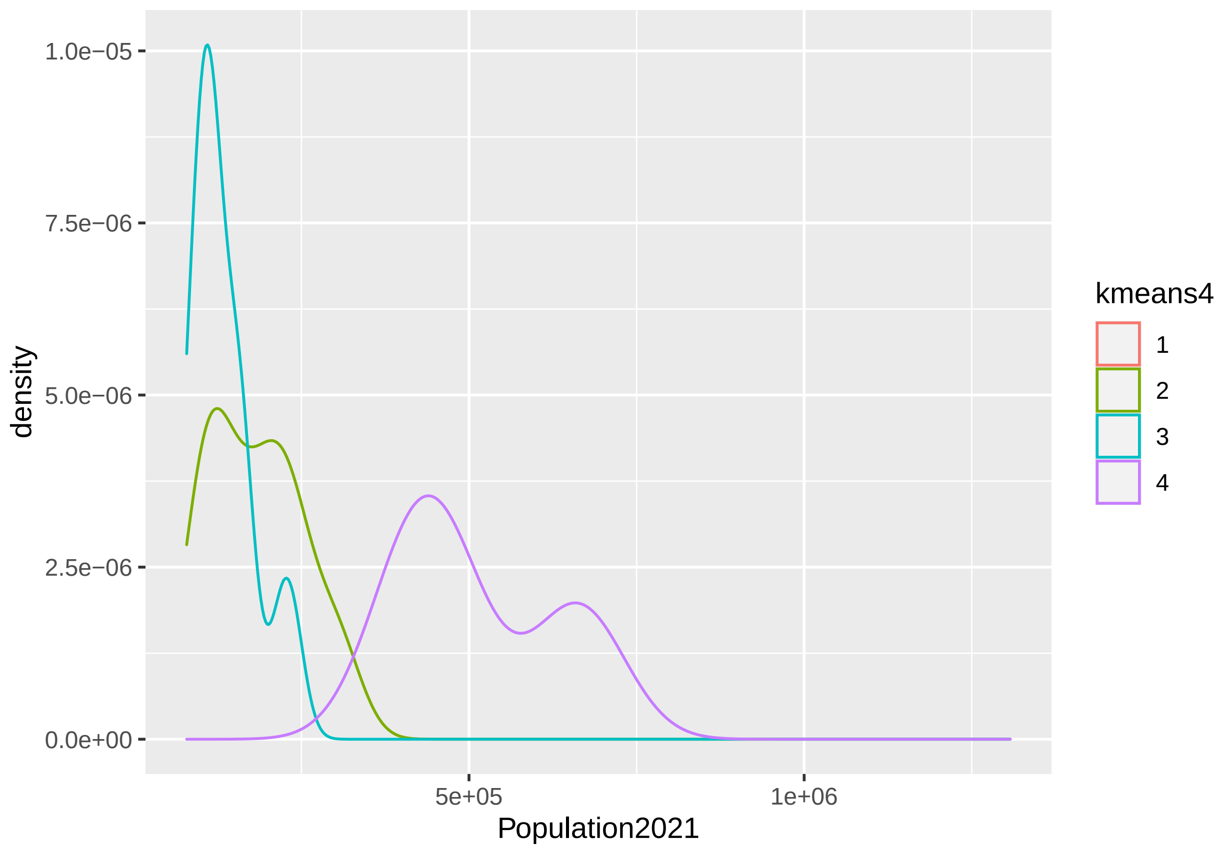

```r
p1 <- ggplot(bul_with_labels, aes(x=kmeans4, y=activity_rate, fill = kmeans4)) +
    geom_boxplot(varwidth = TRUE, alpha=0.5) +
    theme(legend.position="none") +
        scale_fill_manual(values = c("#f6d543", "#c53c74", "#2f0a5b", "#f98e09")) +
        theme_classic() +
        ggtitle("Activity rate by cluster K-means")
p2 <- ggplot(bul_with_labels, aes(x=labs_pca_hc4, y=activity_rate, fill = labs_pca_hc4)) +
    geom_boxplot(varwidth = TRUE, alpha=0.5) +
    theme(legend.position="none") +
        scale_fill_manual(values = c("#f6d543", "#c53c74", "#2f0a5b", "#f98e09")) +
        theme_classic() +
        ggtitle("Activity rate by cluster Hierarchical clustering")


ggarrange(p1, p2)
```




```
ggplot(bul_with_labels, aes(Population2021, colour = kmeans4)) +
        geom_density()
```

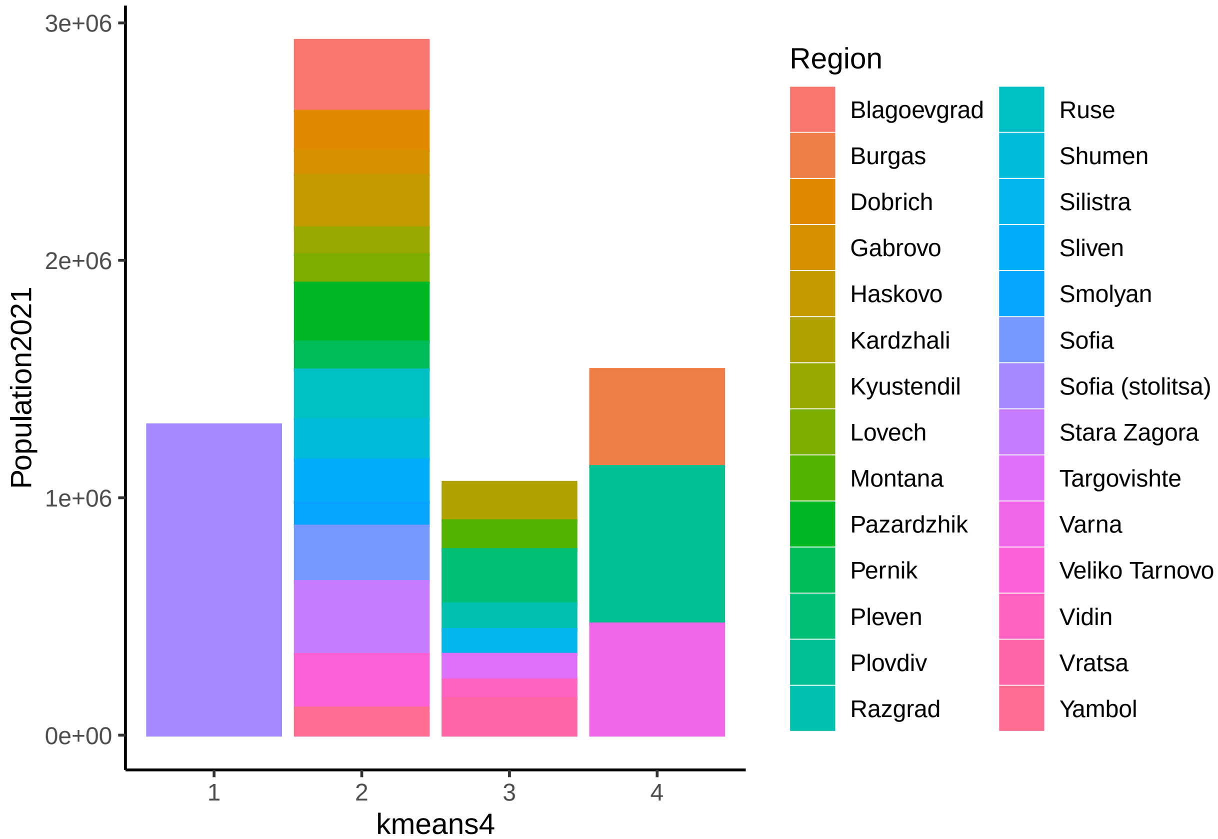
```
## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```



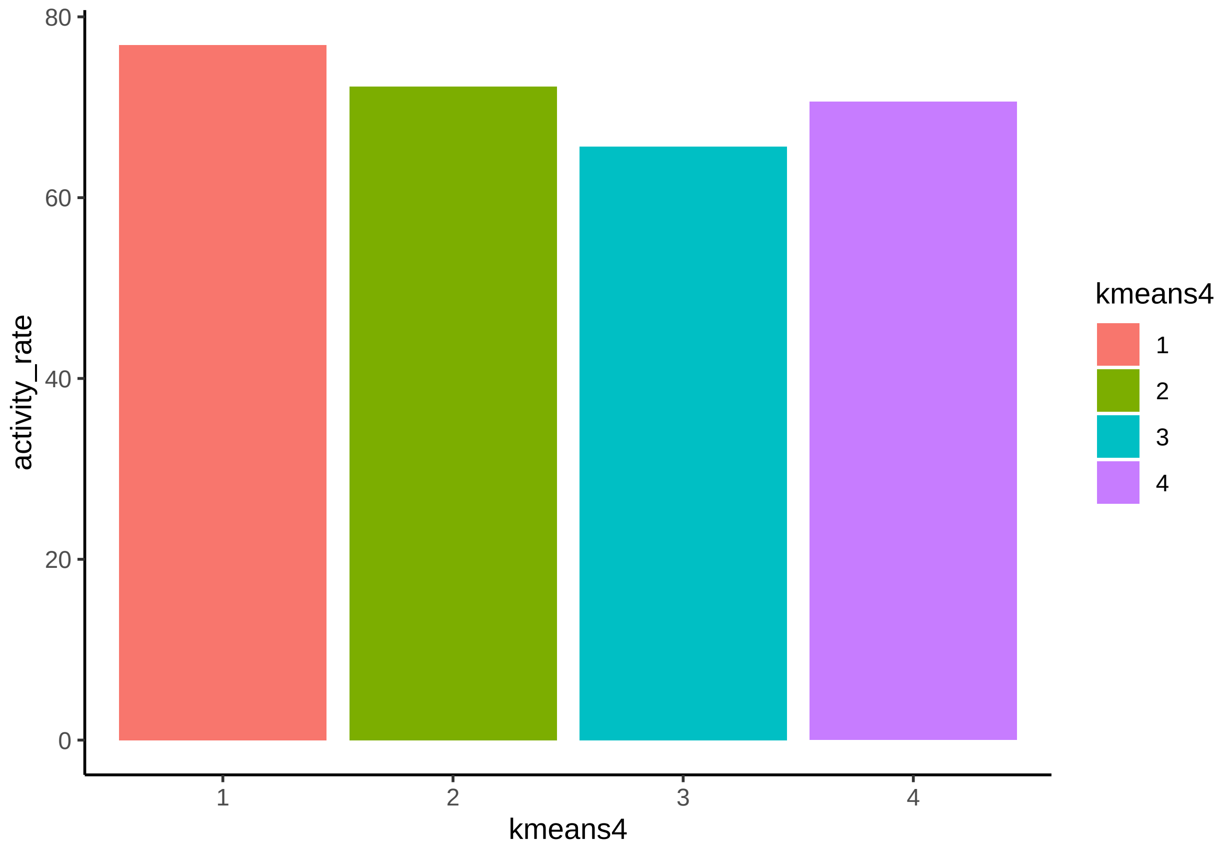

```
ggplot(bul_with_labels, aes(x = kmeans4, y = Population2021)) +
        geom_bar(
                aes(color = Region, fill = Region),
                stat = "summary", position = position_stack()
        ) +
        theme_classic()
```

```
## No summary function supplied, defaulting to `mean_se()`
```



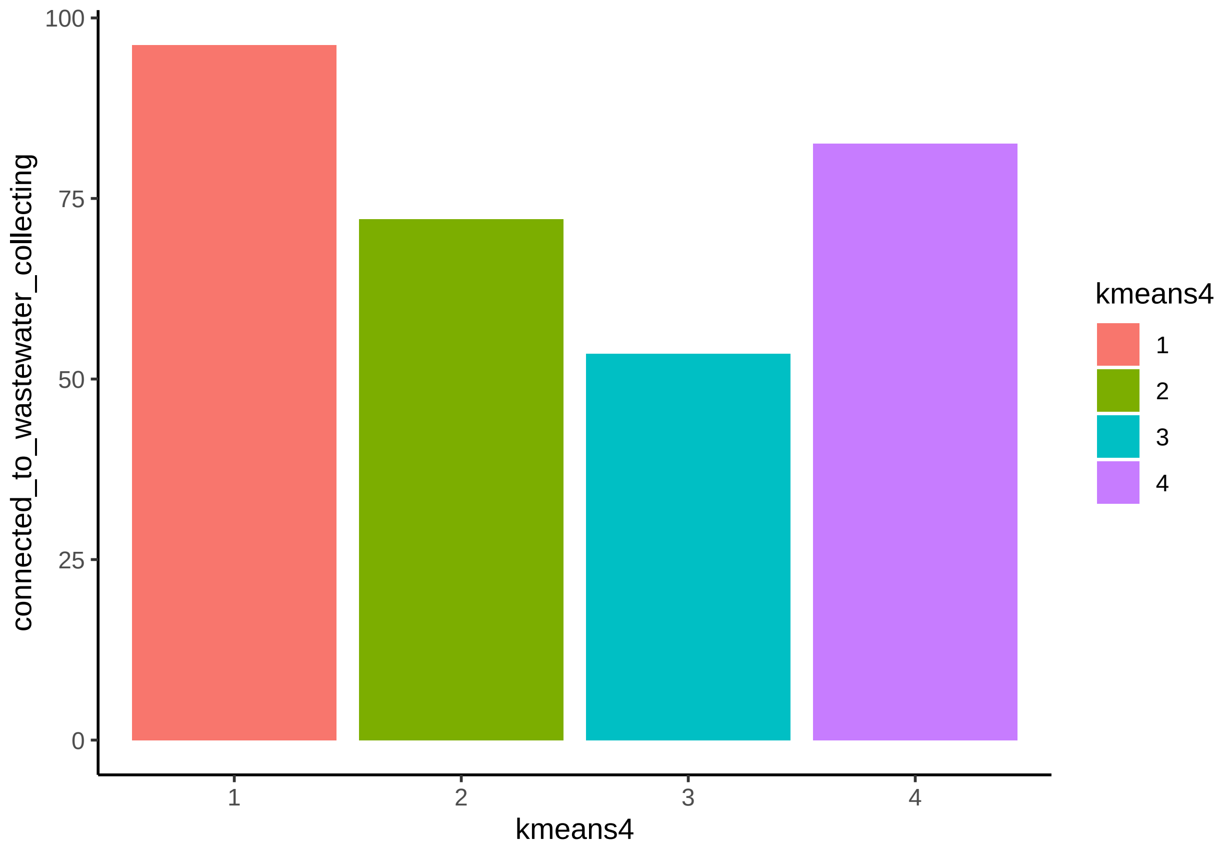

```
ggplot(bul_with_labels, aes(x = kmeans4, y = activity_rate)) +
        geom_bar(
                aes(fill = kmeans4),
                #aes(color = Region, fill = Region),
                stat = "summary", position = position_stack()
        ) +
        theme_classic()
```

```
## No summary function supplied, defaulting to `mean_se()`
```




```
ggplot(bul_with_labels, aes(x = kmeans4, y = connected_to_wastewater_collecting)) +
        geom_bar(
                aes(fill = kmeans4),
                #aes(color = Region, fill = Region),
                stat = "summary", position = position_stack()
        ) +
        theme_classic()
```

```
## No summary function supplied, defaulting to `mean_se()`
```




```
#To see which cities are classified differently
bul_with_labels$Region[bul_with_labels$kmeans4 == 2 & bul_with_labels$labs_pca_hc4 == 3]
```

```
## [1] "Lovech"     "Dobrich"    "Yambol"     "Pazardzhik" "Smolyan"
## [6] "Haskovo"
```

#Lasso and Ridge regression

Nested Cross Validation: model selection and model asessment. To choose the optimal lambda, and to have a picture of the medium test error rate.

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.1.3
```

```
## Loading required package: lattice
```

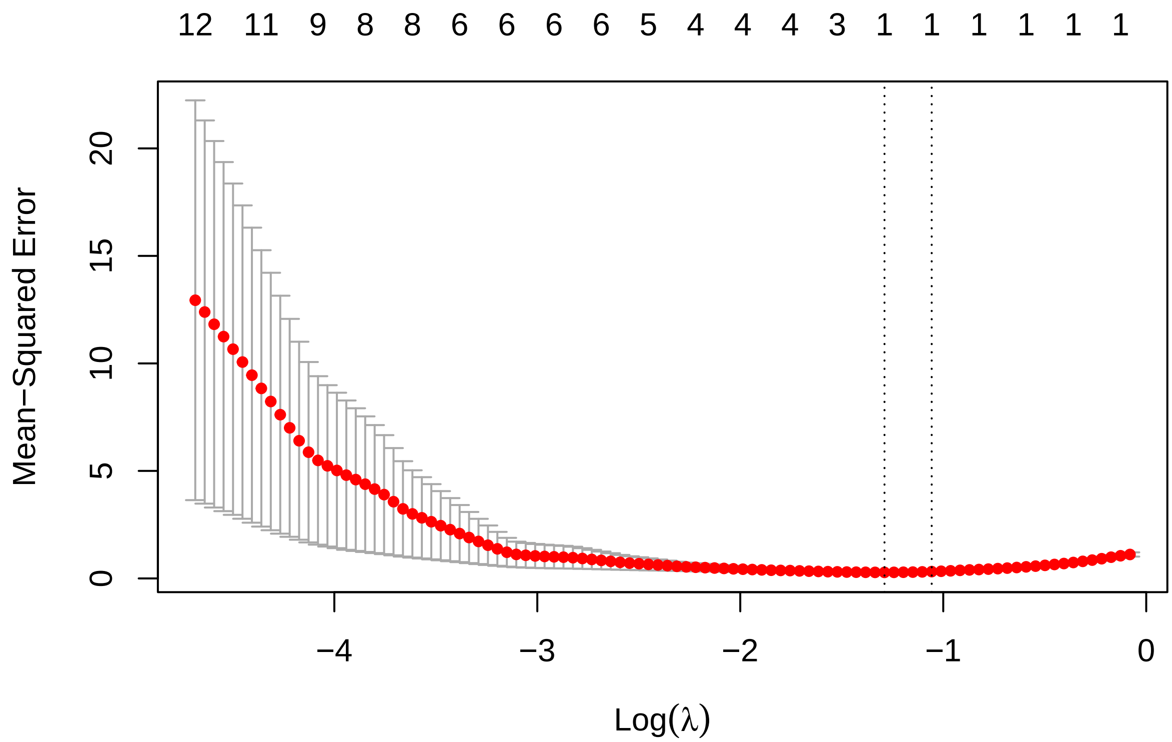
```
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.1.3
```

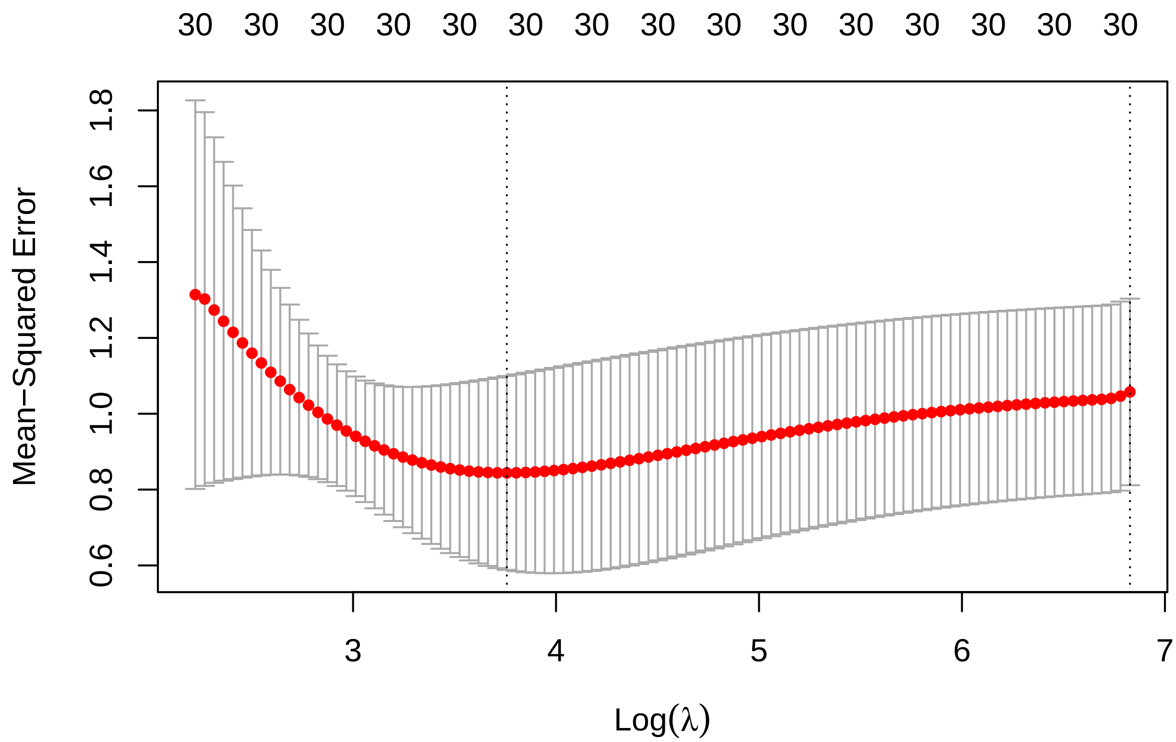
```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-4
```

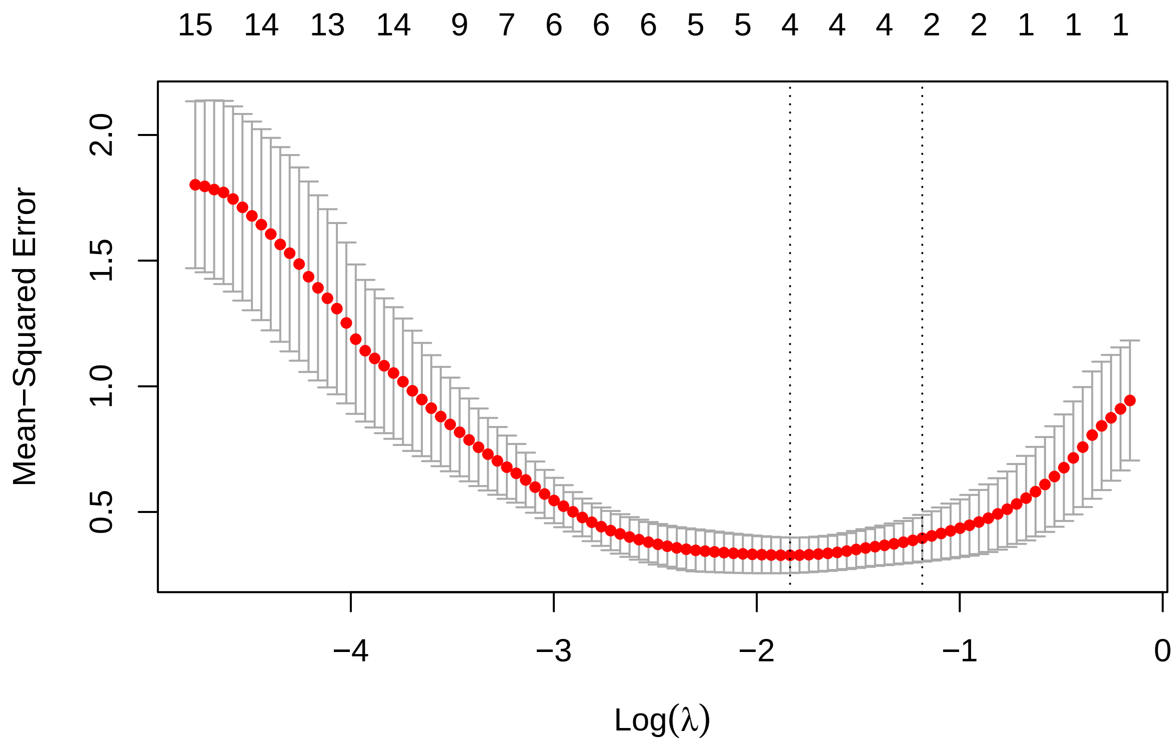
```
set.seed(21)
outfolds <- caret::createFolds(scaled$employment_rate ,k = 5, returnTrain = FALSE)
errors_folds_las <- c()
 errors_folds_rid <- c()

for (outfold in outfolds){
        tr <- scaled[-outfold,]
        ts <- scaled[outfold,]
        y_test <- scaled[outfold,16]
        y_train <- scaled[-outfold,16]

        #takinga as a variable the region according to kmeans
        x_train <- model.matrix(employment_rate ~ .-Region - labs_pca_hc4, data = tr) #[,-c(16,29,30,31
        x_test <- model.matrix(employment_rate ~ .-Region - labs_pca_hc4, data = ts) #[,-c(16,29,30,31)
        #print(colnames(x_train))
        cv.las <- cv.glmnet(x_train, y_train, alpha = 1, nfolds = 5)

        las <- glmnet(x_train, y_train, alpha = 1)

        plot(cv.las)
        bestlam = cv.las$lambda.min

        las_predict <- predict(las, s = bestlam, newx = x_test)
        #print(las_predict)
        #print(dim(las_predict))
        #print(y_test)
        #print(length(y_test))
        err <- mean((as.data.frame(las_predict)$s1 - y_test)^2)
        print(err)
        errors_folds_las <- c(errors_folds_las, err)

        #RIDGE
        cv.rid <- cv.glmnet(x_train, y_train, alpha = 0, nfolds = 5)

        rid <- glmnet(x_train, y_train, alpha = 0)

        plot(cv.rid)
        bestlam = cv.rid$lambda.min

        rid_predict <- predict(rid, s = bestlam, newx = x_test)
        #print(las_predict)
        #print(dim(las_predict))
        #print(y_test)
        #print(length(y_test))
        err2 <- mean((as.data.frame(rid_predict)$s1 - y_test)^2)
        print(err2)
        errors_folds_rid <- c(errors_folds_rid, err2)

}
```


12 11 9 8 8 6 6 6 6 5 4 4 4 3 1 1 1 1 1 1
Mean−Squared Error
0
5
10
15
20
−4
−3
−2
−1
0
Log(λ)

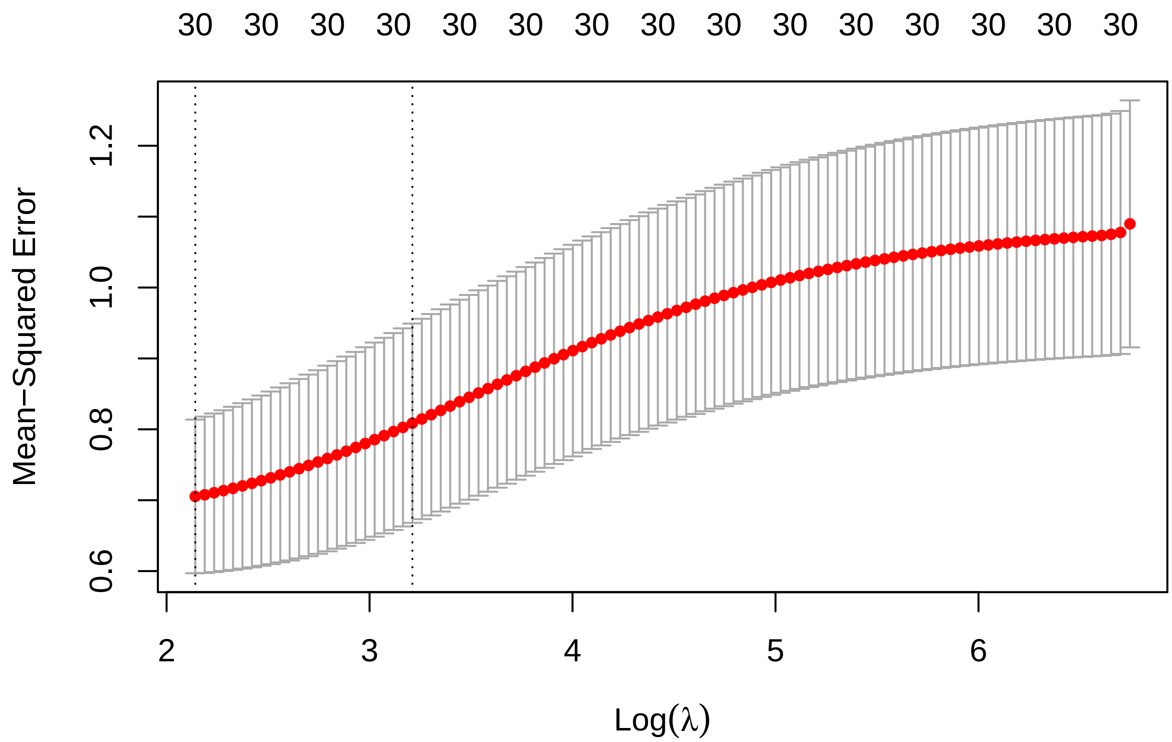
```
## [1] 0.3414291
```


30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
Mean−Squared Error
0.6
0.8
1.0
1.2
1.4
1.6
1.8
3
4
5
6
7
Log(λ)

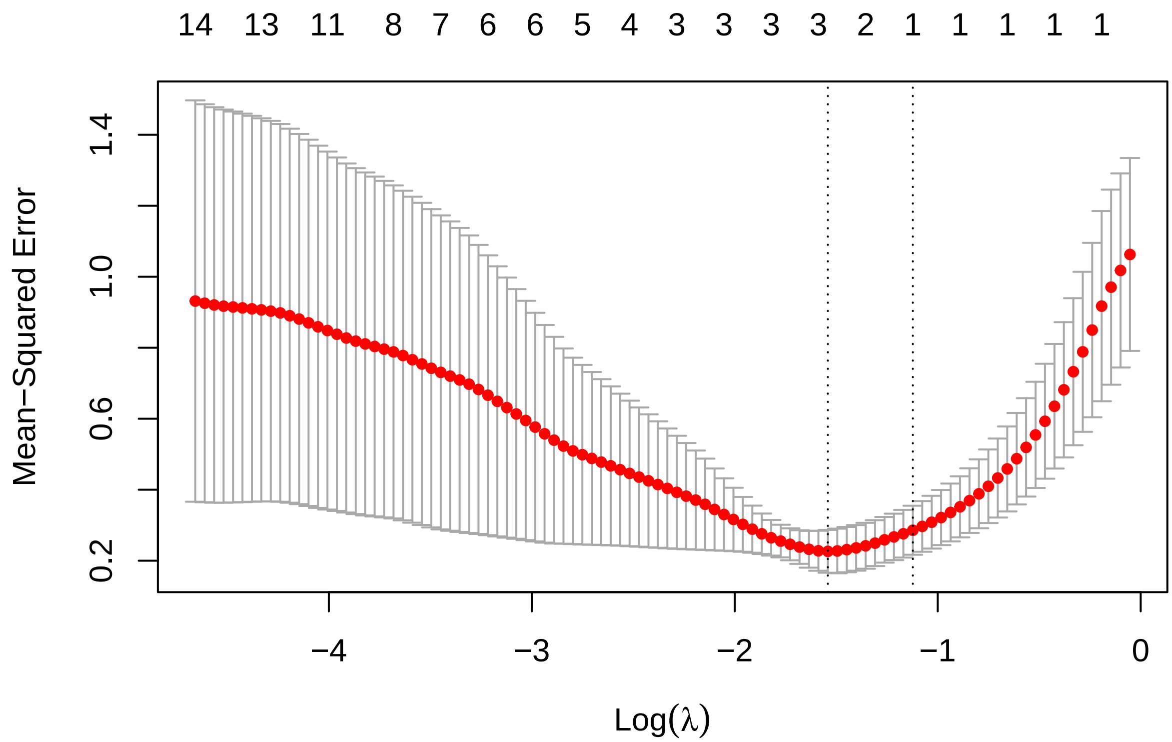
```
## [1] 0.7286646
```


15 14 13 14 9 7 6 6 6 5 5 4 4 4 2 2 1 1 1
Mean−Squared Error
0.5
1.0
1.5
2.0
−4
−3
−2
−1
0
Log(λ)

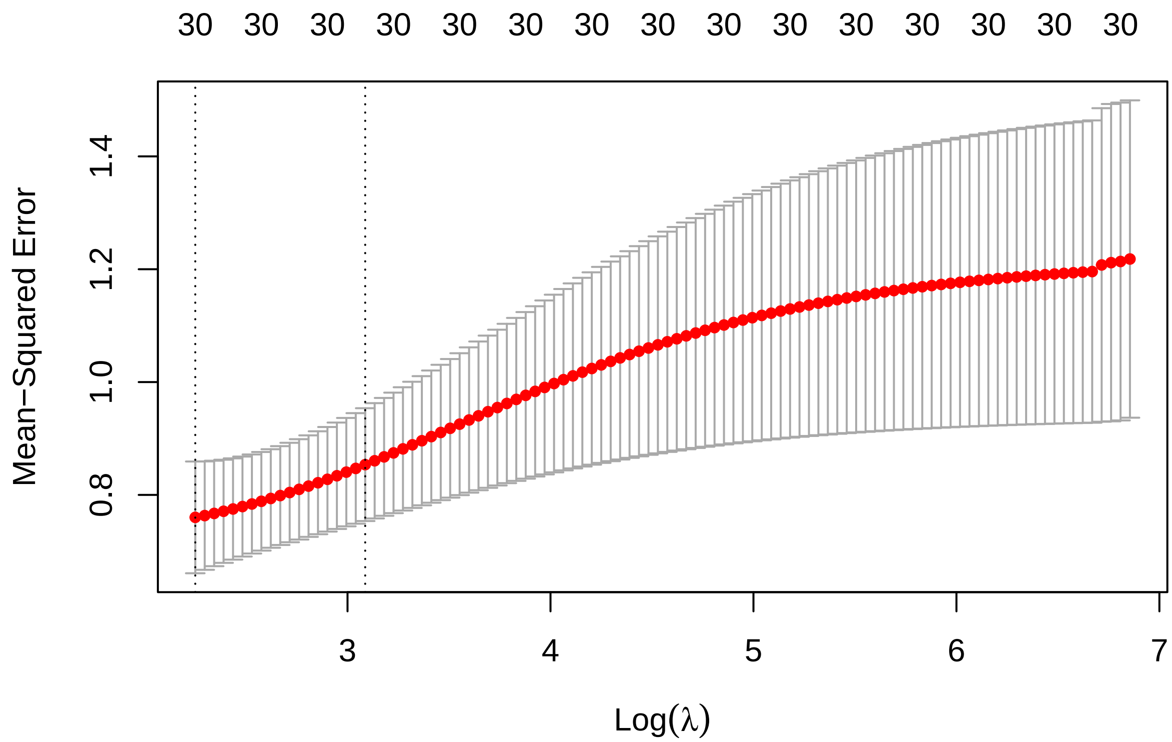
```
## [1] 0.2142669
```


30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
Mean−Squared Error
0.6
0.8
1.0
1.2
2
3
4
5
6
Log(λ)

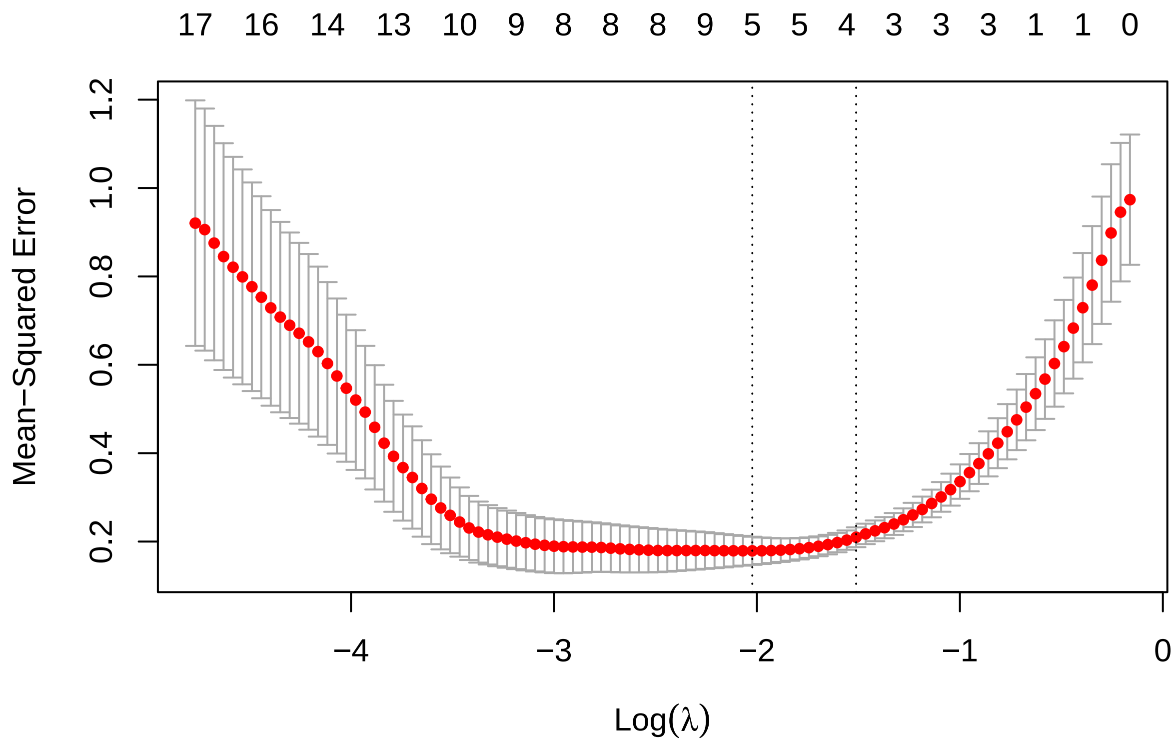
```
## [1] 0.7627827
```


14 13 11 8 7 6 6 5 4 3 3 3 3 2 1 1 1 1 1
Mean−Squared Error
0.2
0.6
1.0
1.4
−4
−3
−2
−1
0
Log(λ)

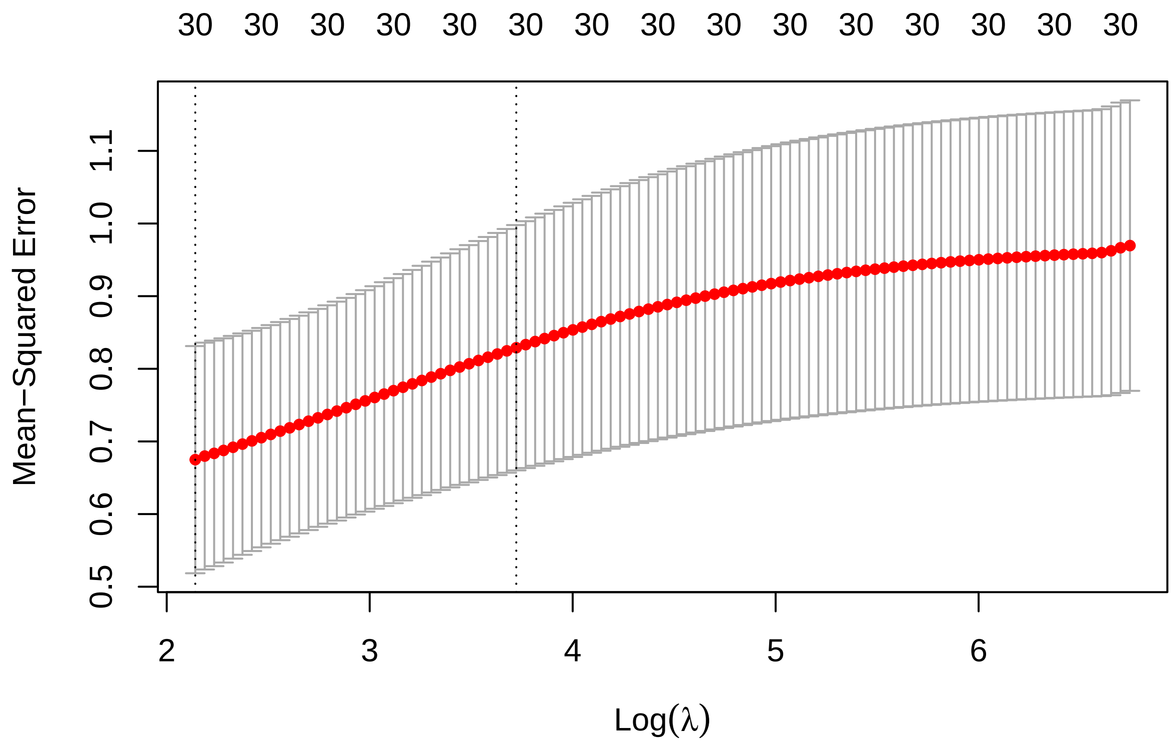
```
## [1] 0.2482378
```


30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
Mean−Squared Error
0.8
1.0
1.2
1.4
3
4
5
6
7
Log(λ)

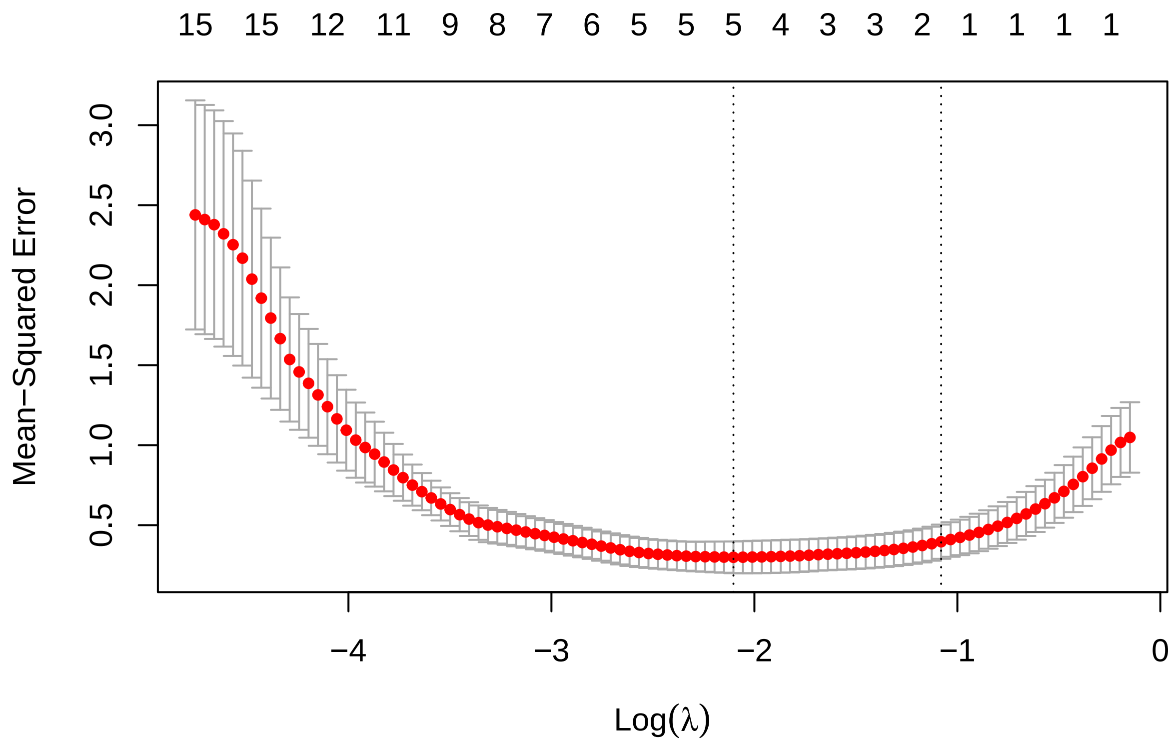
```
## [1] 0.4553804
```


17 16 14 13 10 9 8 8 8 9 5 5 4 3 3 3 1 1 0
Mean−Squared Error
0.2
0.4
0.6
0.8
1.0
1.2
−4
−3
−2
−1
0
Log(λ)

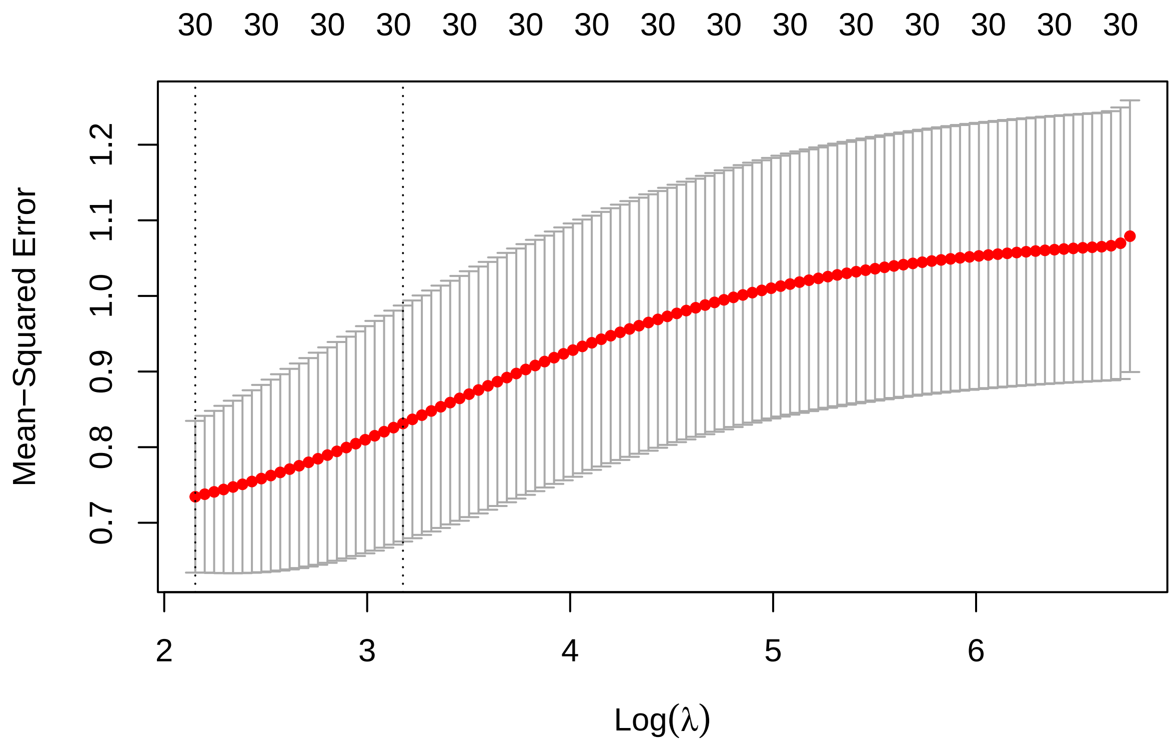
```
## [1] 0.4159981
```


30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
Mean−Squared Error
0.5 0.6 0.7 0.8 0.9 1.0 1.1
2 3 4 5 6
Log(λ)

```
## [1] 1.310944
```


15 15 12 11 9 8 7 6 5 5 5 4 3 3 2 1 1 1 1
Mean−Squared Error
0.5 1.0 1.5 2.0 2.5 3.0
−4 −3 −2 −1 0
Log(λ)

```
## [1] 0.1472002
```




```
## [1] 0.660005
```

```
#The estimate of the error of the models are:
mean(errors_folds_las)
```

```
## [1] 0.2734264
```

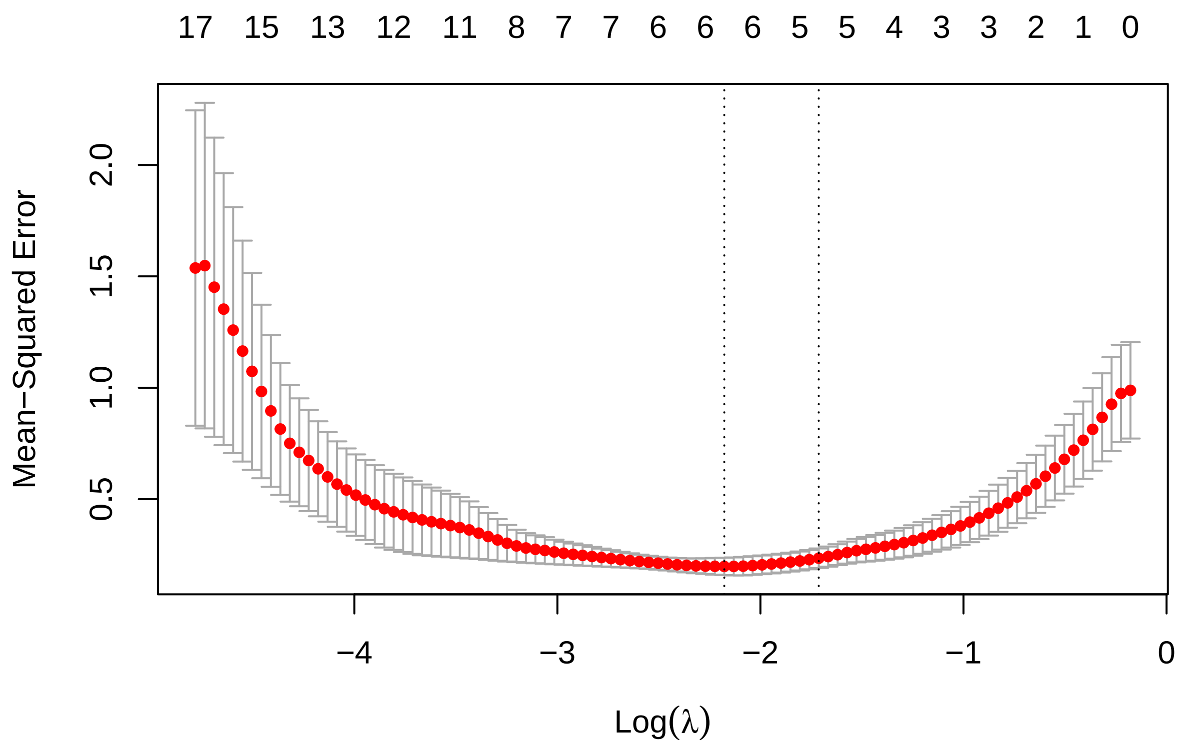
```
mean(errors_folds_rid)
```

```
## [1] 0.7835553
```

Just one train-test separation

```
y <- scaled$employment_rate

set.seed(21)
tr <- sample(1:nrow(scaled), 0.8*nrow(scaled))
ts <- -tr
y_test <- y[ts]
y_train <- y[tr]

x <- model.matrix(employment_rate ~ .-Region-labs_pca_hc4, data = scaled) #[,-c(16,29,30,31)]

x_train <- model.matrix(employment_rate ~ .-Region-labs_pca_hc4, data = scaled[tr,]) #[,-c(16,29,30,31).
x_test <- model.matrix(employment_rate ~ .-Region-labs_pca_hc4, data = scaled[ts,]) #[,-c(16,29,30,31)]
```

```r
set.seed(21)
cv.las <- cv.glmnet(x_train, y_train, alpha = 1, nfolds = 5)

las <- glmnet(x_train, y_train, alpha = 1)

plot(cv.las)
```




```r
bestlam = cv.las$lambda.min

las_predict <- predict(las, s = bestlam, newx = x_test)
mean((las_predict - y_test)^2) #test error based on just one split train-test
```

```
## [1] 0.4242374
```

Fit lasso on full dataset, final model to be used

```r
out.cv <- cv.glmnet(x, y, alpha = 1, nfolds = 5)
bestlam <- out.cv$lambda.min
out <- glmnet(x, y, alpha = 1)

las_pred <- predict(out, type = "coefficients", s = bestlam)[1:32,]
las_pred
```

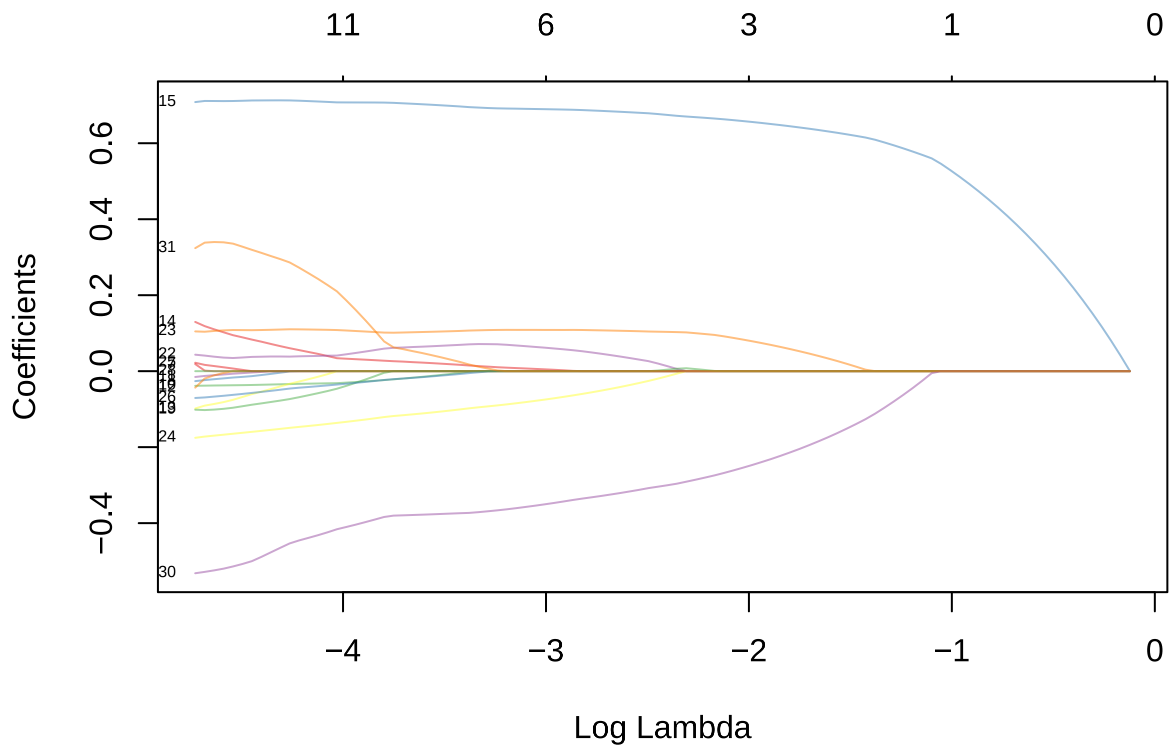
```
##                          (Intercept)                          (Intercept)
```

```
##                            0.06605187                             0.00000000
##                        Population2021                         fertility_rate
##                            0.00000000                             0.00000000
##                  unis_colleges_number                         teaching_staff
##                            0.00000000                             0.00000000
##                     students_enrolled                          graduates_uni
##                            0.00000000                             0.00000000
##             ratio_grad_enrolled_unis                           TotalSchools
##                            0.00000000                             0.00000000
##                           gradtotot57                           gradtotot812
##                            0.00000000                             0.00000000
##          employees_laborcontract_avg  employees_laborcontract_avg_topop
##                            0.00000000                             0.00000000
##                               avgwage                          activity_rate
##                            0.00000000                             0.65052093
##                labour_force_thousands     foreign_direct_investment_euro
##                            0.00000000                             0.00000000
##                      books_pamphlets       accommodation_establishments
##                            0.00000000                             0.00000000
##                arrivals_accommodation                     revenue_accomm_lev
##                            0.00000000                             0.00000000
##                                crimes                     ecological_assets
##                            0.00000000                             0.06891589
##                     water_supply2020 connected_to_wastewater_collecting
##                            0.00000000                             0.00000000
##              Pop_watersupplyregime                    waste_thousandtons
##                            0.00000000                             0.00000000
##                        hospitals2021                              kmeans42
##                            0.00000000                             0.00000000
##                              kmeans43                              kmeans44
##                           -0.23118154                             0.00000000
```

```r
print(las_pred[las_pred != 0]) # These are the most important variables
```

```
##       (Intercept)     activity_rate ecological_assets          kmeans43
##        0.06605187        0.65052093        0.06891589       -0.23118154
```

```r
#Again we see that activity rate is very important, but also ecological assets
#also region cluster!

plot(out, xvar = "lambda", label = TRUE)
```



Some predictions

```
one_x <- model.matrix(employment_rate ~ .-Region - labs_pca_hc4, data = scaled)[1,] #[1,-c(16,29,30,31)
prediction_new_observation <- predict(out, s = bestlam, newx = one_x)
#This is normalized, so to get employment rate, need to "de-normalize"
mean_empl_rate <- mean(bul_with_labels$employment_rate)
stdev_empl_rate <- sqrt(var(bul_with_labels$employment_rate))
one_y <- prediction_new_observation*stdev_empl_rate + mean_empl_rate
(prediction_new_observation - scaled[1,16])^2 #error for this one
```

```
##              s1
## [1,] 0.6266015
```

```
#Let's see what will happen if we increase the activity rate with say 10%
new_x <- bul_with_labels[c(1,2,3),]
new_x$activity_rate <- new_x$activity_rate + 10
new_x_scaled <- scaled[c(1,2,3),]
new_x_scaled$activity_rate <- (new_x$activity_rate - mean(bul_with_labels$activity_rate))/sqrt(var(bul_w
new_x_scaled_mm <- model.matrix(employment_rate ~ .-Region-labs_pca_hc4, data = new_x_scaled) #[,-c(16,2
new_scaled_predictions <- predict(out, s = bestlam, newx = new_x_scaled_mm)
new_predictions <- new_scaled_predictions*stdev_empl_rate + mean_empl_rate

#To see the change in the first three cities for example:
bul_with_labels[c(1,2,3),c(1,17)] #this would be before
```

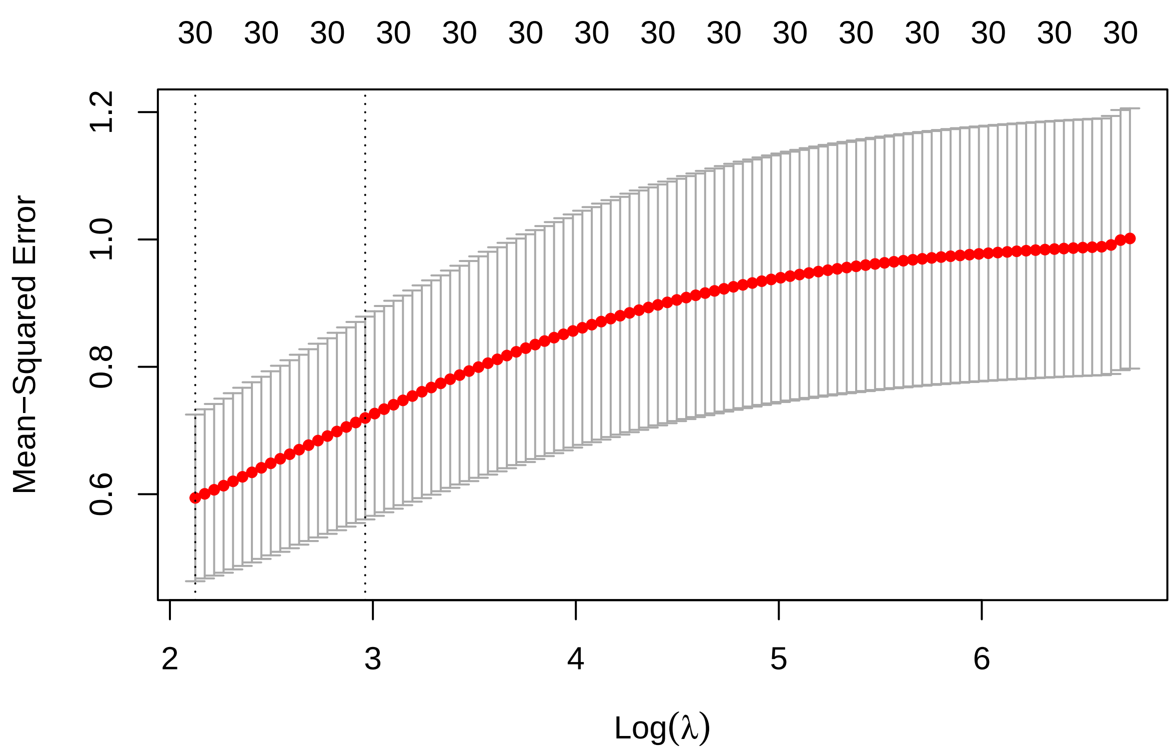
```
## # A tibble: 3 x 2
```

```
##   Region employment_rate
##   <chr>             <dbl>
## 1 Vidin              57.5
## 2 Vratsa             58.5
## 3 Lovech             65
```

```
new_predictions #after increase in activity rate with 10%
```

```
##              s1
## Vidin  71.96609
## Vratsa 68.86956
## Lovech 73.60174
```

```
set.seed(21)
cv.ridg <- cv.glmnet(x_train, y_train, alpha = 0, nfolds = 5)

ridg <- glmnet(x_train, y_train, alpha = 0)

plot(cv.ridg)
```



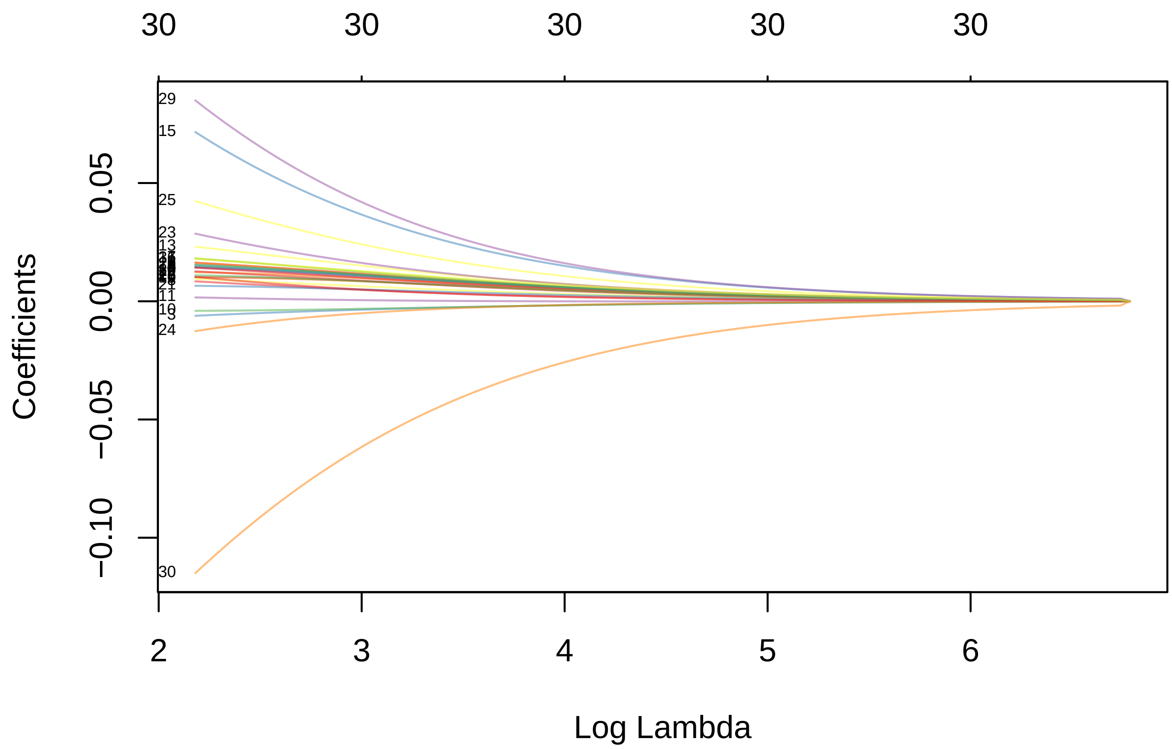

```
bestlam = cv.ridg$lambda.min

ridg_predict <- predict(ridg, s = bestlam, newx = x_test)
mean((ridg_predict - y_test)^2)
```

```
## [1] 1.075523
```

```r
out.cv <- cv.glmnet(x, y, alpha = 0, nfolds = 5)
bestlam <- out.cv$lambda.min
out <- glmnet(x, y, alpha = 0)
ridg_pred <- predict(out, type = "coefficients", s = bestlam)[1:32,]
ridg_pred
```

```
##                      (Intercept)                          (Intercept)
##                     -0.014789244                          0.000000000
##                   Population2021                       fertility_rate
##                      0.015362616                         -0.005438623
##             unis_colleges_number                       teaching_staff
##                      0.013832112                          0.013491261
##                students_enrolled                        graduates_uni
##                      0.015340629                          0.016959348
##          ratio_grad_enrolled_unis                         TotalSchools
##                      0.007554291                          0.013980163
##                      gradtotot57                         gradtotot812
##                     -0.003937786                          0.001309550
##       employees_laborcontract_avg   employees_laborcontract_avg_topop
##                      0.011953325                          0.021273243
##                          avgwage                        activity_rate
##                      0.013270576                          0.061892025
##             labour_force_thousands     foreign_direct_investment_euro
##                      0.016882316                          0.010119465
##                  books_pamphlets        accommodation_establishments
##                      0.010012730                          0.008819129
##            arrivals_accommodation                    revenue_accomm_lev
##                      0.011821126                          0.006309441
##                           crimes                     ecological_assets
##                      0.010624529                          0.025302893
##                water_supply2020 connected_to_wastewater_collecting
##                     -0.010337282                          0.037679052
##            Pop_watersupplyregime                   waste_thousandtons
##                      0.008863088                          0.014585569
##                   hospitals2021                            kmeans42
##                      0.014642704                          0.073110296
##                         kmeans43                            kmeans44
##                     -0.100762630                          0.016811707
```

```r
plot(out, xvar = "lambda", label = TRUE)
```




```
ridg_pred[order(abs(ridg_pred))] #ordering of the coefficients of variables
```

```
##                       (Intercept)                      gradtotot812
##                       0.000000000                       0.001309550
##                       gradtotot57                    fertility_rate
##                      -0.003937786                      -0.005438623
##                 revenue_accomm_lev          ratio_grad_enrolled_unis
##                       0.006309441                       0.007554291
##       accommodation_establishments             Pop_watersupplyregime
##                       0.008819129                       0.008863088
##                   books_pamphlets      foreign_direct_investment_euro
##                       0.010012730                       0.010119465
##                  water_supply2020                            crimes
##                      -0.010337282                       0.010624529
##            arrivals_accommodation          employees_laborcontract_avg
##                       0.011821126                       0.011953325
##                           avgwage                     teaching_staff
##                       0.013270576                       0.013491261
##               unis_colleges_number                      TotalSchools
##                       0.013832112                       0.013980163
##                 waste_thousandtons                    hospitals2021
##                       0.014585569                       0.014642704
##                       (Intercept)                 students_enrolled
##                      -0.014789244                       0.015340629
##                    Population2021                          kmeans44
```

```
##                         0.015362616                         0.016811707
##              labour_force_thousands                        graduates_uni
##                         0.016882316                         0.016959348
##   employees_laborcontract_avg_topop                   ecological_assets
##                         0.021273243                         0.025302893
## connected_to_wastewater_collecting                        activity_rate
##                         0.037679052                         0.061892025
##                            kmeans42                            kmeans43
##                         0.073110296                        -0.100762630
```

```r
write.csv(bul_with_labels, "C:/Users/User/Documents/UNITN/Computational social science/bulgaria/bul_with
```