Correlation

Maksym Rud'

27 09 2020

Correlation Analysis

Europe 2016 data_set

Data

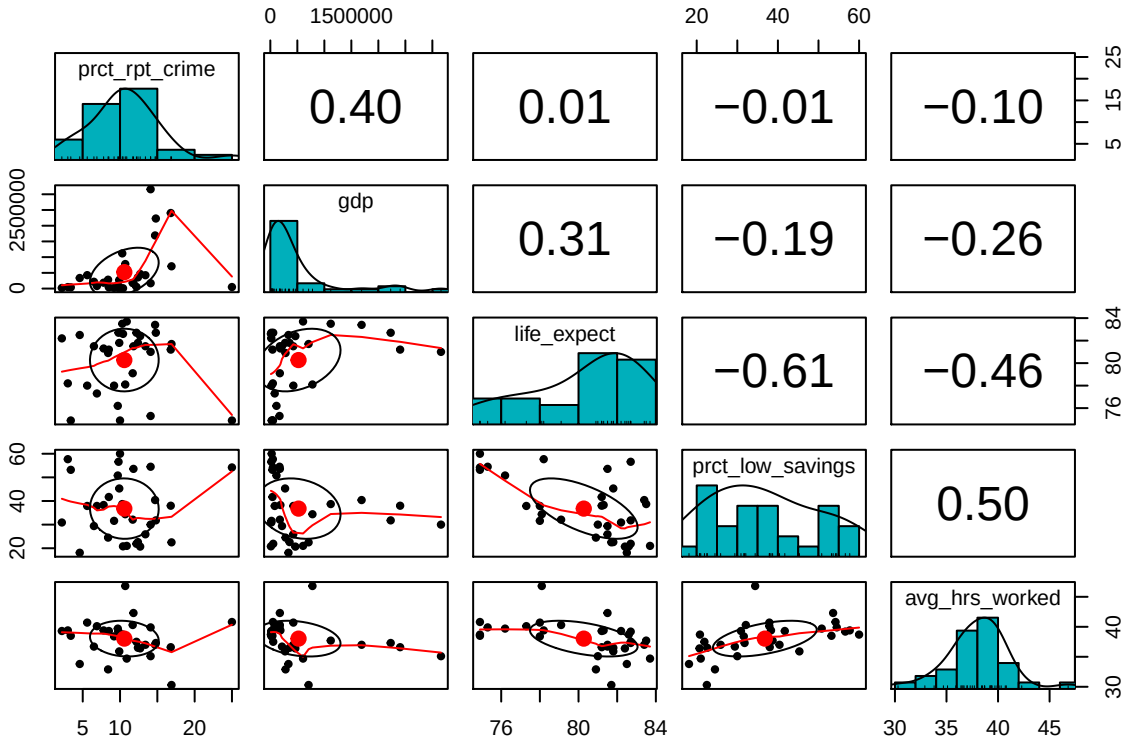
```
head(data_europe)
```

```
##    country prct_rpt_crime      gdp life_expect prct_low_savings avg_hrs_worked
## 1  Austria           12.4 356237.6        81.8             22.6           36.5
## 2  Belgium           13.4 424660.3        81.5             25.9           37.0
## 3 Bulgaria           25.0  48128.6        74.9             54.2           40.8
## 4  Croatia            3.0  46639.5        78.2             57.7           39.4
## 5   Cyprus            9.8  18490.2        82.7             56.6           39.2
## 6  Czechia           11.7 176370.1        79.1             32.1           40.3
```

Matrix plot

Pearson matrix plot

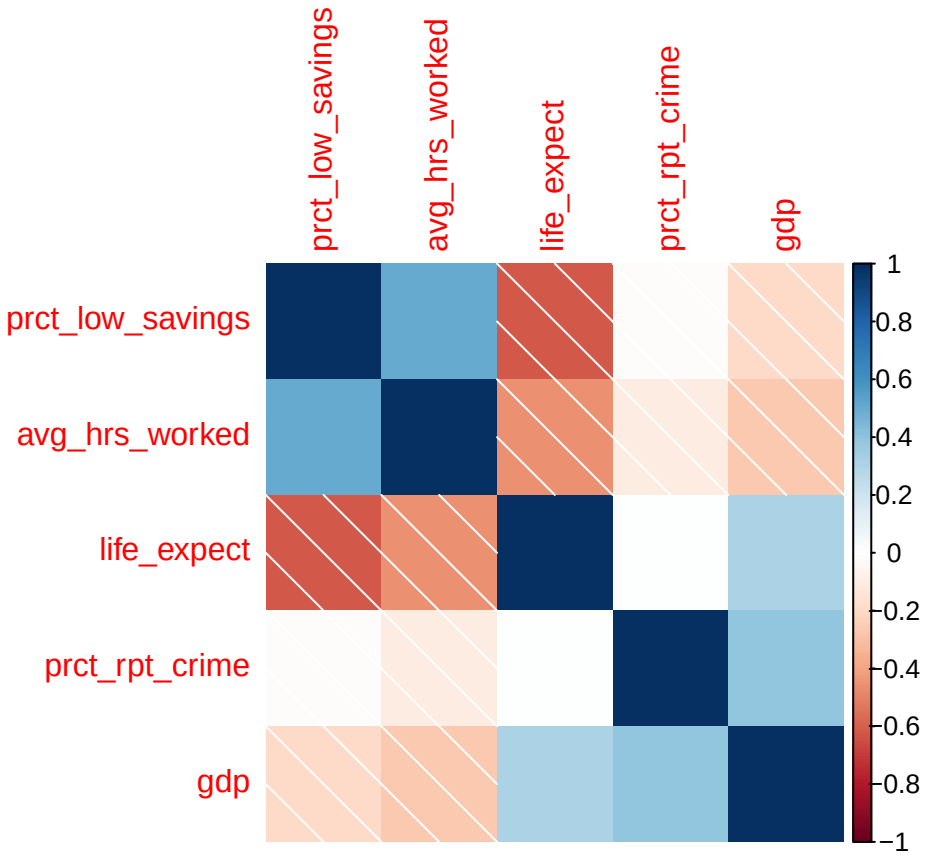
```
pairs.panels(data_europe[,2:6],
             method = "pearson", # correlation method
             hist.col = "#00AFBB",
             density = TRUE,  # show density plots
             ellipses = TRUE # show correlation ellipses
             )
```


prct_rpt_crime
gdp
life_expect
prct_low_savings
avg_hrs_worked
0.40
0.01
−0.01
−0.10
0.31
−0.19
−0.26
−0.61
−0.46
0.50
0
1500000
20
40
60
5
15
25
2500000
84
80
76
60
40
20
5
10
20
76
80
84
30
35
40
45

Corr Plot

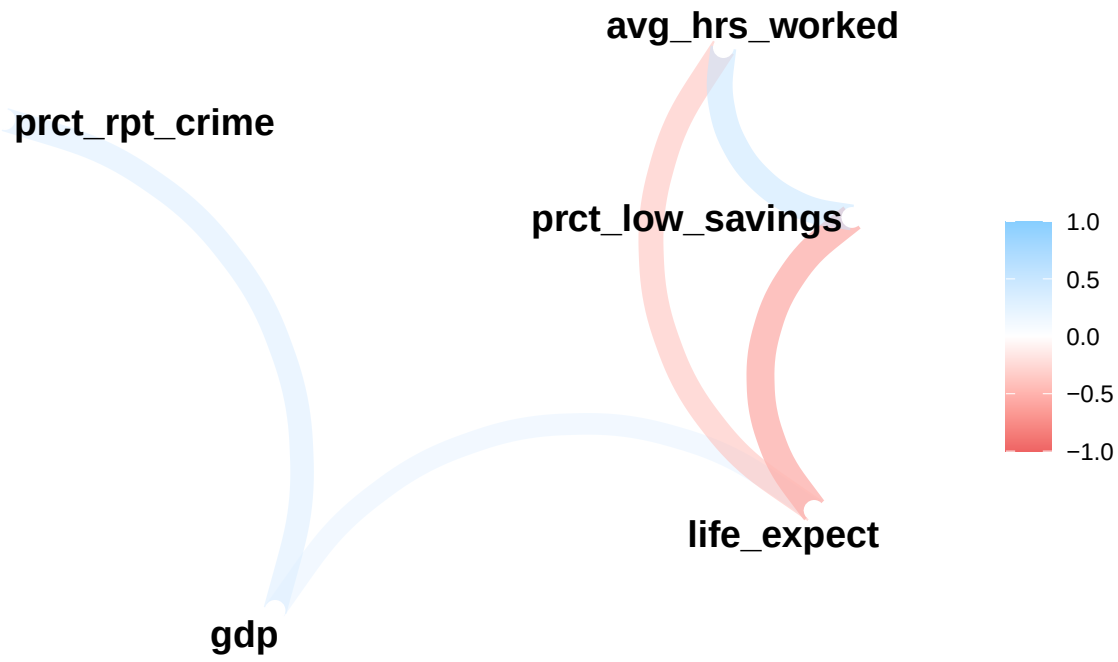

prct_rpt_crime
gdp
life_expect
prct_low_savings
avg_hrs_worked
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Cluster



Graph plot

```
network_plot(EU)
```



Tests for percent of low savings and avg hours of work

```
res <- cor.test(data_europe$avg_hrs_worked, data_europe$prct_low_savings, method = "pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  data_europe$avg_hrs_worked and data_europe$prct_low_savings
## t = 3.1883, df = 30, p-value = 0.003338
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1872249 0.7246518
## sample estimates:
##      cor
## 0.503076
```

The p-value of the test is 0.003338, which is less than the significance level alpha = 0.05. We can conclude that low savings and avg hours of work are significantly correlated with a correlation coefficient of 0.503076 and p-value of 0.003338 .

```r
res <- cor.test(data_europe$avg_hrs_worked, data_europe$prct_low_savings, method = "spearman")
res
```

```
##
##  Spearman's rank correlation rho
##
## data:  data_europe$avg_hrs_worked and data_europe$prct_low_savings
## S = 2236.5, p-value = 0.0003785
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.5900798
```

```r
res <- cor.test(data_europe$avg_hrs_worked, data_europe$prct_low_savings, method = "kendall")
res
```

```
##
##  Kendall's rank correlation tau
##
## data:  data_europe$avg_hrs_worked and data_europe$prct_low_savings
## z = 3.1967, p-value = 0.00139
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.3991913
```

Tests for percent of low savings and life excpectations

```r
res <- cor.test(data_europe$life_expect, data_europe$prct_low_savings, method = "pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  data_europe$life_expect and data_europe$prct_low_savings
## t = -4.2312, df = 30, p-value = 0.0002015
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.7913446 -0.3338069
## sample estimates:
##       cor
## -0.6113428
```

The p-value of the test is 0.0002015, which is less than the significance level alpha = 0.05. We can conclude that low savings and life excpectations are significantly correlated with a correlation coefficient of -0.6113428 and p-value of 0.0002015.

```r
res <- cor.test(data_europe$life_expect, data_europe$prct_low_savings, method = "spearman")
res
```

```
##
##  Spearman's rank correlation rho
##
## data:  data_europe$life_expect and data_europe$prct_low_savings
## S = 8111.4, p-value = 0.004733
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.4866928
```

```r
res <- cor.test(data_europe$life_expect, data_europe$prct_low_savings, method = "kendall")
res
```

```
##
##  Kendall's rank correlation tau
##
## data:  data_europe$life_expect and data_europe$prct_low_savings
## z = -2.9735, p-value = 0.002945
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##        tau
## -0.3738745
```