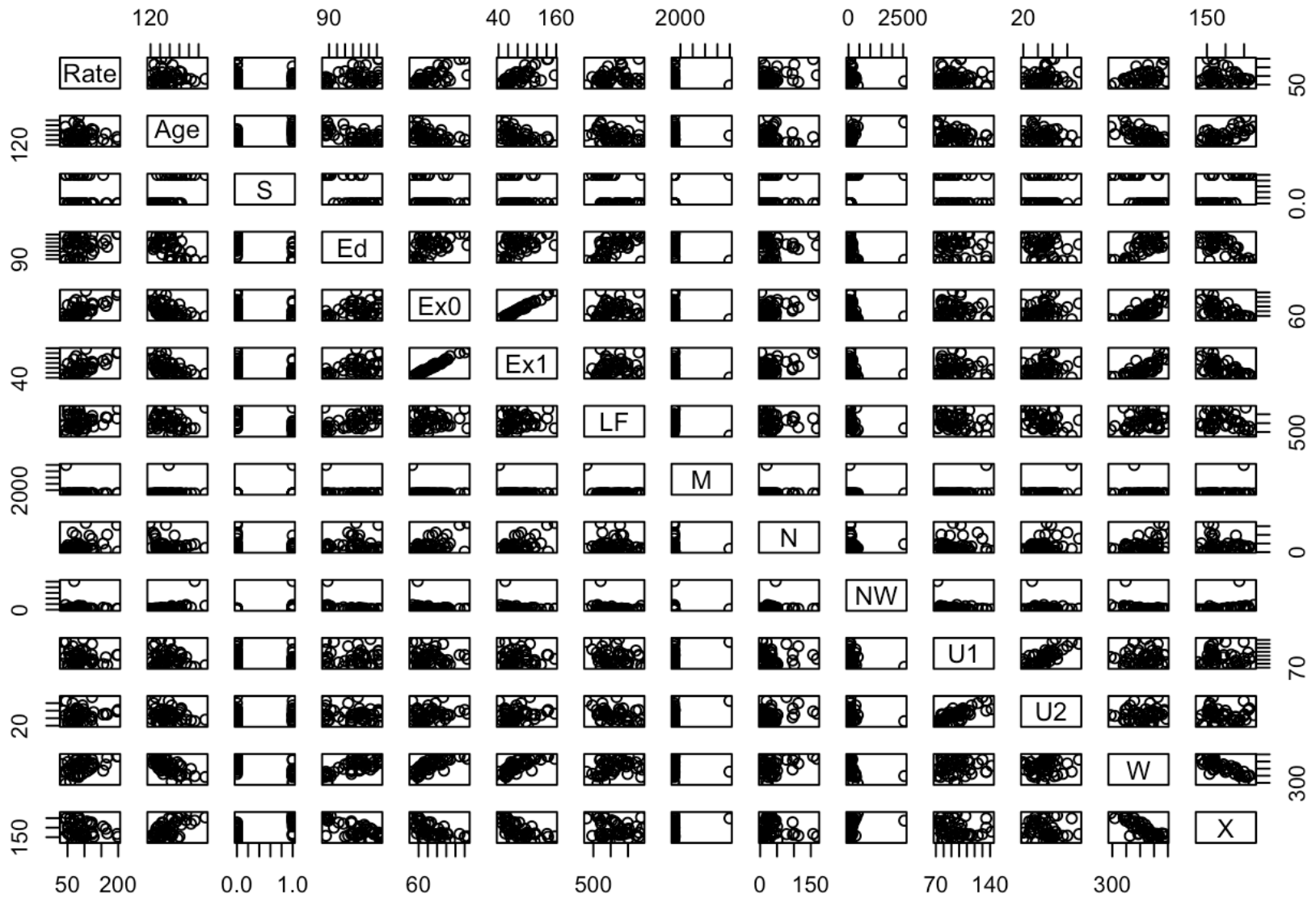
Raport 4 Statystyka

Jan Solarz

11 05 2020

z.1/2/3

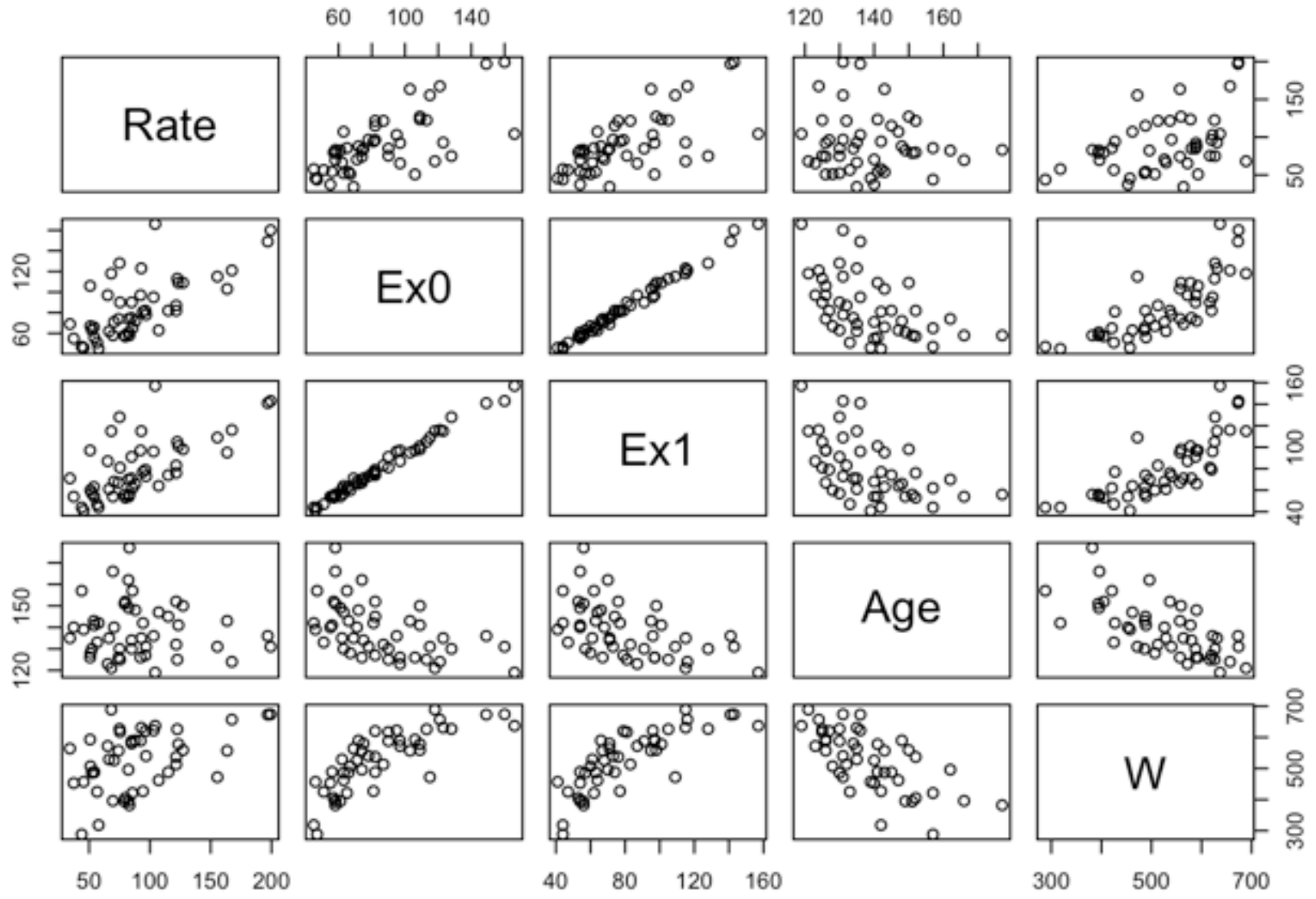
```
pairs(cbind(Rate, Age,  S,    Ed,   Ex0,  Ex1,  LF,   M,    N,    NW,   U1,   U2
,W,    X ))
```




```
#widzimy silna korelacje miedzy cechami Ex0 i Ex1, ale również miedzy U1 i U2
cor(cbind(Rate, Age,  S,    Ed,   Ex0,  Ex1,  LF,   M,    N,    NW,   U1,   U2   ,
W,    X ))
```

```
##              Rate          Age           S          Ed         Ex0         Ex1
## Rate  1.00000000 -0.089519902 -0.09087073  0.32310629  0.68761660  0.66674511
## Age  -0.08951990  1.000000000  0.58435534 -0.53023964 -0.50696949 -0.51317336
## S    -0.09087073  0.584355338  1.00000000 -0.70274132 -0.37354720 -0.37616753
## Ed    0.32310629 -0.530239642 -0.70274132  1.00000000  0.48326144  0.49940958
## Ex0   0.68761660 -0.506969490 -0.37354720  0.48326144  1.00000000  0.99359772
## Ex1   0.66674511 -0.513173356 -0.37616753  0.49940958  0.99359772  1.00000000
## LF    0.18913745 -0.160948824 -0.50546948  0.56117795  0.12217106  0.10634960
## M    -0.16902237  0.004461682  0.19845402 -0.22557612 -0.19533933 -0.20932621
## N     0.33749589 -0.280637618 -0.04991832 -0.01722740  0.52608277  0.51378940
## NW   -0.06372231  0.457234083  0.39708823 -0.38365025 -0.18123271 -0.18655758
## U1   -0.05067083 -0.224380599 -0.17241931  0.01810345 -0.04277496 -0.05171199
## U2    0.17706371 -0.244843390  0.07169289 -0.21568155  0.18564957  0.16922422
## W     0.44149947 -0.670055056 -0.63694543  0.73599704  0.78724481  0.79426205
## X    -0.18585829  0.629962825  0.67496961 -0.69788198 -0.63228425 -0.65072345
##               LF            M           N          NW          U1           U2
## Rate  0.1891374 -0.169022371  0.33749589 -0.06372231 -0.05067083  0.17706371
## Age  -0.1609488  0.004461682 -0.28063762  0.45723408 -0.22438060 -0.24484339
## S    -0.5054695  0.198454020 -0.04991832  0.39708823 -0.17241931  0.07169289
## Ed    0.5611780 -0.225576121 -0.01722740 -0.38365025  0.01810345 -0.21568155
## Ex0   0.1221711 -0.195339325  0.52608277 -0.18123271 -0.04277496  0.18564957
## Ex1   0.1063496 -0.209326206  0.51378940 -0.18655758 -0.05171199  0.16922422
## LF    1.0000000 -0.288299113 -0.12367222 -0.22796997 -0.22939968 -0.42076249
## M    -0.2882991  1.000000000 -0.07853513 -0.04361084  0.33547258  0.33600928
## N    -0.1236722 -0.078535135  1.00000000  0.05942204 -0.03811995  0.27042159
## NW   -0.2279700 -0.043610839  0.05942204  1.00000000 -0.21818218 -0.10460448
## U1   -0.2293997  0.335472578 -0.03811995 -0.21818218  1.00000000  0.74592482
## U2   -0.4207625  0.336009279  0.27042159 -0.10460448  0.74592482  1.00000000
## W     0.2946323 -0.101849407  0.30826271 -0.34330695  0.04485720  0.09207166
## X    -0.1923744  0.213000458 -0.12466424  0.31281286 -0.11491193 -0.05777706
##                W           X
## Rate  0.44149947 -0.18585829
## Age  -0.67005506  0.62996283
## S    -0.63694543  0.67496961
## Ed    0.73599704 -0.69788198
## Ex0   0.78724481 -0.63228425
## Ex1   0.79426205 -0.65072345
## LF    0.29463231 -0.19237444
## M    -0.10184941  0.21300046
## N     0.30826271 -0.12466424
## NW   -0.34330695  0.31281286
## U1    0.04485720 -0.11491193
## U2    0.09207166 -0.05777706
## W     1.00000000 -0.84417172
## X    -0.84417172  1.00000000
```

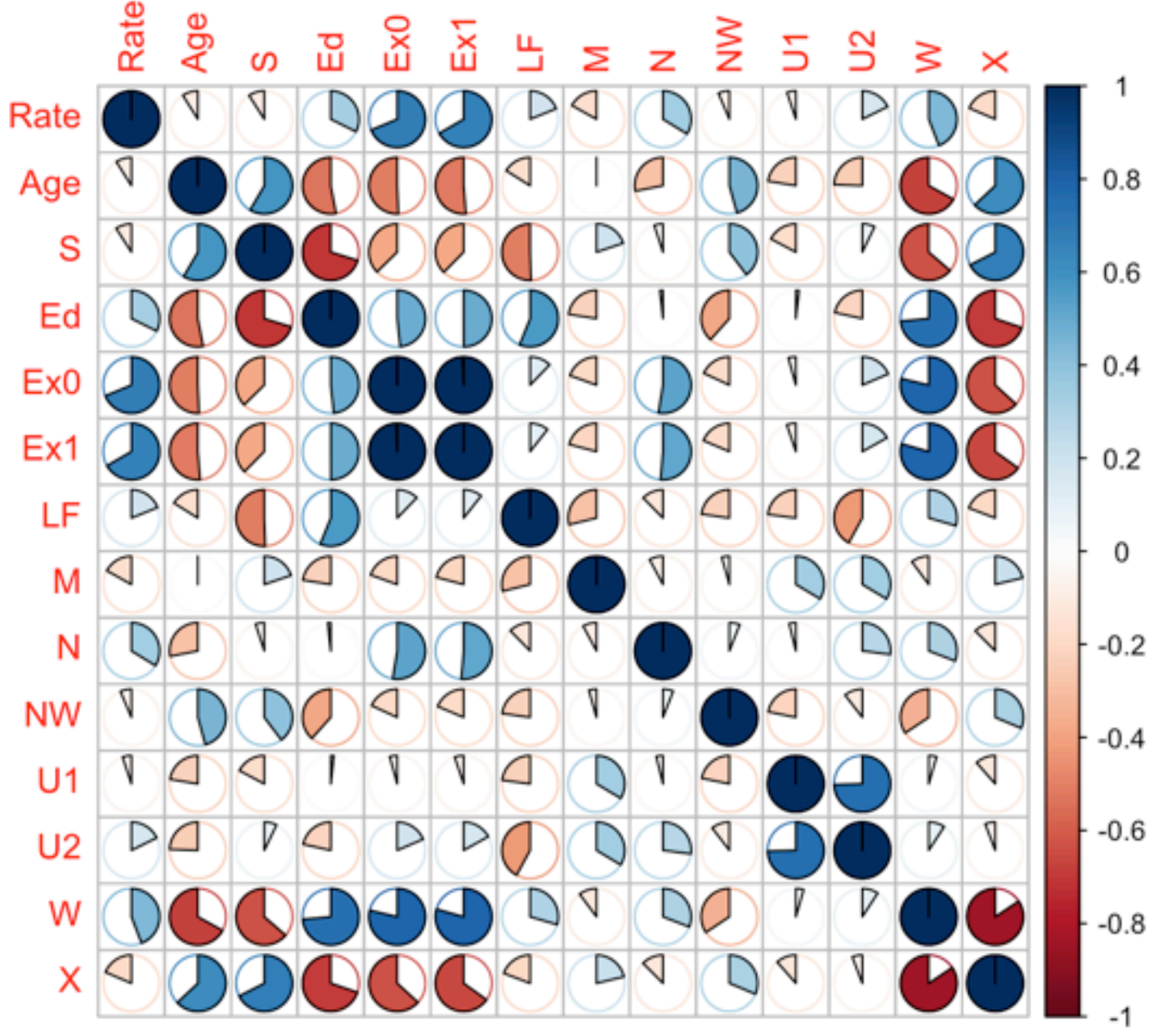
```
pairs(cbind(Rate,Ex0,Ex1,Age,W))
```



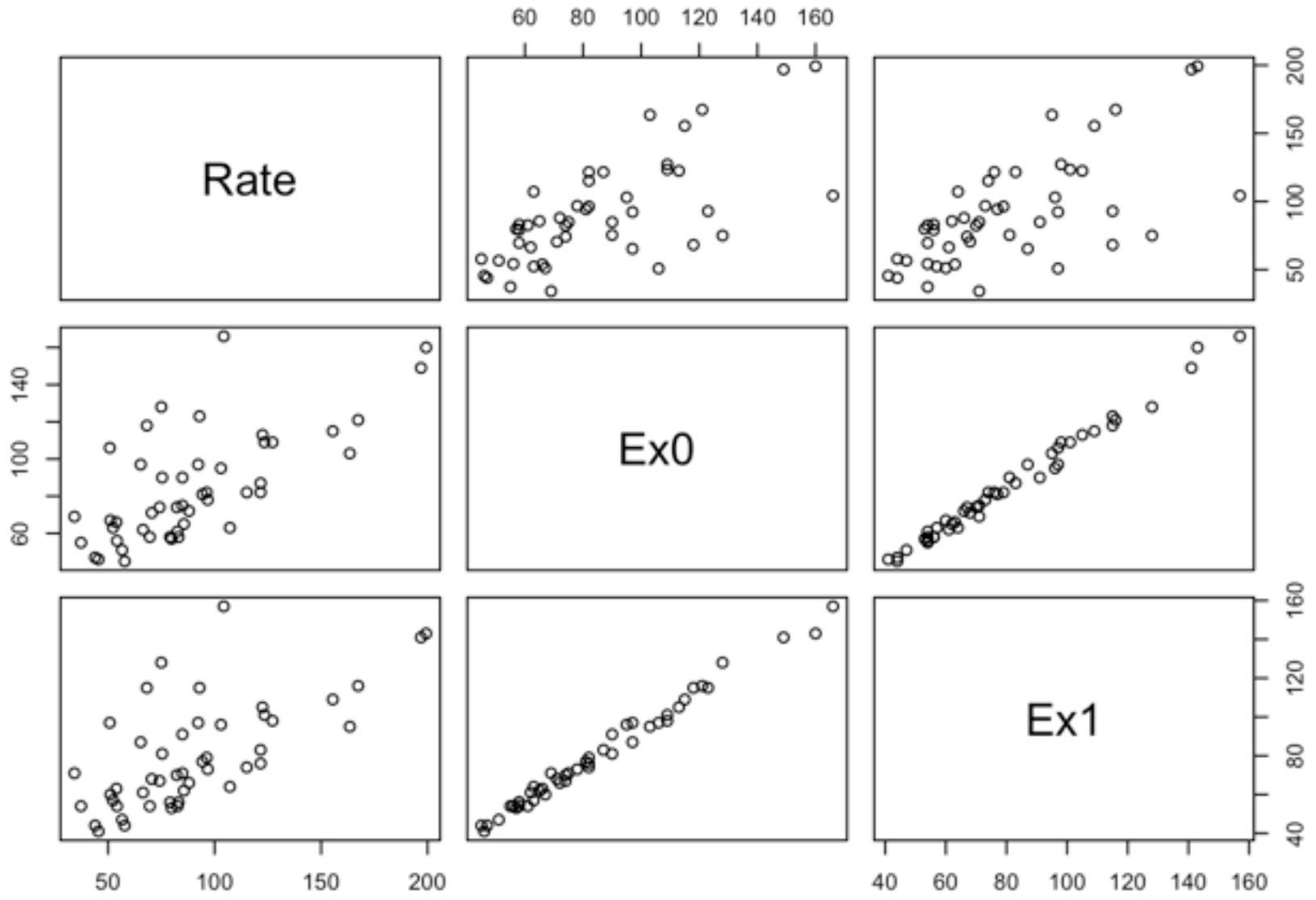

```
cor(cbind(Rate,Ex0,Ex1,Age,W))
```

```
##            Rate        Ex0        Ex1        Age          W
## Rate  1.0000000  0.6876166  0.6667451 -0.0895199  0.4414995
## Ex0   0.6876166  1.0000000  0.9935977 -0.5069695  0.7872448
## Ex1   0.6667451  0.9935977  1.0000000 -0.5131734  0.7942621
## Age  -0.0895199 -0.5069695 -0.5131734  1.0000000 -0.6700551
## W     0.4414995  0.7872448  0.7942621 -0.6700551  1.0000000
```

```
corrplot::corrplot(cor(wiatr),method = "pie")
```



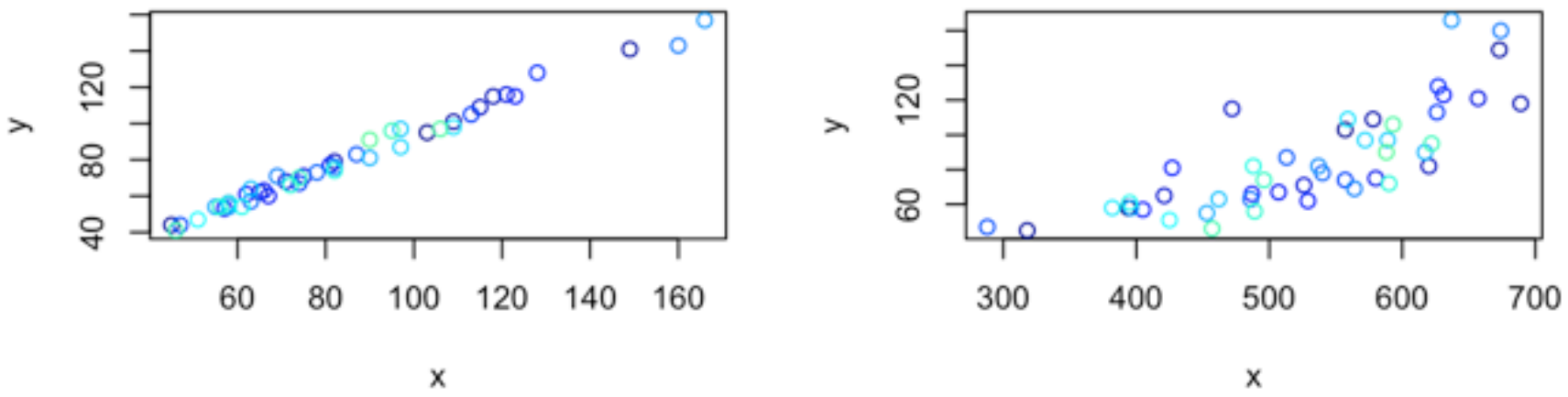

```
#po wygenerowaniu macierzy współczynników i wykresu obrazującego wielkości korelac
ji możemy #stiwerdzić że silna korelacja jest miedzy W,Ed | W,Ex0 | W,Ex1
#W odniesieniu do Rate: Rate,Ex0 i Rate,Ex1
pairs(cbind(Rate,Ex0,Ex1))
```




```
cor(cbind(Rate,Ex0,Ex1))
```

```
##           Rate       Ex0       Ex1
## Rate 1.0000000 0.6876166 0.6667451
## Ex0  0.6876166 1.0000000 0.9935977
## Ex1  0.6667451 0.9935977 1.0000000
```

```
#najwiekszy liniowy wplyw na zmienna Rate maja Ex0 i Ex1- najwyzszy wspolczynnik 0
.69 i 0.67
par(mfrow=c(2,2))
plot3D::points2D(Ex0,Ex1)
plot3D::points2D(W,Ex0)
plot3D::points2D(U1,U2)
plot3D::points2D(W,Ed)
```




```
#Istnieje prara Ex0 i Ex1 ktore sa silnie ze soba skorelowane- widac na wykresach
rozrzutu, wspolczyniki wielkosci 0.99
```

Czesc druga zad. 4/5/6/7

```
#zad4
#Nasz  model regresji liniowej opisujący zależność między zmienną Rate a trzynasto
ma #zmiennymi objaśniającymi Age,...,X
aa<-lm(Rate~Age+  S+    Ed+   Ex0+  Ex1+  LF+   M+    N+    NW+   U1+   U2   +W+
X,data=wiatr)
bb<-c(150, 1, 90, 50, 60, 500, 950, 30, 300, 100, 40, 400, 200)
cc<-c(Age,  S,    Ed,   Ex0,  Ex1,  LF,   M,    N,    NW,   U1,   U2   ,W,    X)
names((summary(aa)))
```

```
##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

```
summary(aa)$coefficients
```

```
##                  Estimate   Std. Error    t value     Pr(>|t|)
## (Intercept) -5.145976e+02 1.193486e+02 -4.3117196 0.0001377801
## Age          1.078183e+00 4.377478e-01  2.4630237 0.0191644210
## S            5.513618e+00 1.329101e+01  0.4148382 0.6809449633
## Ed           1.512464e+00 6.708269e-01  2.2546267 0.0309152825
## Ex0          1.876910e+00 1.138562e+00  1.6484911 0.1087416157
## Ex1         -8.788587e-01 1.232634e+00 -0.7129922 0.4808623885
## LF           5.758384e-02 1.364631e-01  0.4219738 0.6757806107
## M           -3.304402e-03 3.277325e-03 -1.0082621 0.3206648508
## N           -8.804831e-02 1.253989e-01 -0.7021456 0.4875116208
## NW          -2.112354e-03 1.115786e-02 -0.1893153 0.8510052926
## U1          -2.763975e-01 3.913121e-01 -0.7063352 0.4849371528
## U2           1.769364e+00 8.994953e-01  1.9670635 0.0576302765
## W            7.899337e-02 1.062441e-01  0.7435081 0.4624352019
## X            5.462586e-01 2.067025e-01  2.6427288 0.0124800018
```

```
Rate1=sum(aa$coefficients[2:14]*bb)+aa$coefficients[1]
#prognozowana wartość Rate przy 13 zmiennych objaśniających
Rate1
```

```
## (Intercept)
##    36.24069
```

```
names((summary(aa)))
```

```
##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

```
#współczynnik determinacji R^2
summary(aa)$r.squared
```

```
## [1] 0.7330418
```

```
#współczynnik determinacji adj R^2
summary(aa)$adj.r.squared
```

```
## [1] 0.6278765
```

```
cat("Rownanie regresji opisujace zaleznosc:","Nasze_Rate=cecha[i]*wspolczynnik[i]+
wyraz wolny")
```

```
## Rownanie regresji opisujace zaleznosc: Nasze_Rate=cecha[i]*wspolczynnik[i]+wyra
z wolny
```

```
cat("Rate=Age*1.078183e+00 +  S*5.513618e+00 + Ed*1.512464e+00 +   Ex0*1.876910e+0
0 +  Ex1*(-8.788587e-01) +  LF*5.758384e-02 +   M*-3.304402e-03 +    N*(-8.804831e
-02) +    NW*(-2.112354e-03) +   U1*(-2.763975e-01) +   U2*1.769364e+00 + W*7.8993
37e-02 + X*5.462586e-01 + -5.145976e+02")
```

```
## Rate=Age*1.078183e+00 +  S*5.513618e+00 + Ed*1.512464e+00 +   Ex0*1.876910e+00
+  Ex1*(-8.788587e-01) +  LF*5.758384e-02 +   M*-3.304402e-03 +    N*(-8.804831e-0
2) +    NW*(-2.112354e-03) +   U1*(-2.763975e-01) +   U2*1.769364e+00 + W*7.899337
e-02 + X*5.462586e-01 + -5.145976e+02
```

```
#zad5,6,7
#(a)
wektor_a<-c(Ex1,X,Ed, Age,U2)
wartosci_a<-c(60,200,90,150,40)
model_a<-lm(Rate~Ex1+X+Ed+ Age+U2,data=wiatr)
summary(model_a)
```

```
##
## Call:
## lm(formula = Rate ~ Ex1 + X + Ed + Age + U2, data = wiatr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.916 -13.952  -3.081  13.355  59.164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -479.5483   100.3883  -4.777 2.29e-05 ***
## Ex1            1.2689     0.1665   7.621 2.22e-09 ***
## X              0.5253     0.1469   3.576 0.000913 ***
## Ed             1.6652     0.4815   3.458 0.001281 **
## Age            1.1067     0.3819   2.898 0.006000 **
## U2             1.1204     0.4722   2.373 0.022426 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.21 on 41 degrees of freedom
## Multiple R-squared:  0.679,  Adjusted R-squared:  0.6399
## F-statistic: 17.35 on 5 and 41 DF,  p-value: 3.37e-09
```

```
summary(model_a)$adj.r.squared
```

```
## [1] 0.6398613
```

```
summary(model_a)$r.squared
```

```
## [1] 0.6790068
```

```
model_a$coefficients
```

```
##  (Intercept)          Ex1            X           Ed          Age           U2
## -479.5483284    1.2688878    0.5252569    1.6652227    1.1067425    1.1204328
```

```
Rate_a=sum(model_a$coefficients[2:6]*wartosci_a)+model_a$coefficients[1]
Rate_a
```

```
## (Intercept)
##    62.33506
```

```
#(b)
wektorb<-c(Ex0,LF,M,N,NW)
wartosci_b<-c(50, 500, 950, 30, 300)
model_b<-lm(Rate~Ex0+LF+M+N+NW,data=wiatr)
summary(model_b)
```

```
##
## Call:
## lm(formula = Rate ~ Ex0 + LF + M + N + NW, data = wiatr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.048 -16.267   2.798  18.395  53.936
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.443e+01  6.819e+01  -0.798    0.429
## Ex0          9.132e-01  1.808e-01   5.051 9.53e-06 ***
## LF           1.184e-01  1.170e-01   1.012    0.318
## M            1.813e-04  3.519e-03   0.052    0.959
## N           -2.187e-02  1.379e-01  -0.159    0.875
## NW           9.694e-03  1.230e-02   0.788    0.435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.2 on 41 degrees of freedom
## Multiple R-squared:  0.4918, Adjusted R-squared:  0.4298
## F-statistic: 7.936 on 5 and 41 DF,  p-value: 2.658e-05
```

```
summary(model_b)$adj.r.squared
```

```
## [1] 0.4298397
```

```
summary(model_b)$r.squared
```

```
## [1] 0.4918136
```

```
model_b$coefficients
```

```
##   (Intercept)           Ex0            LF             M             N
## -5.443313e+01  9.132311e-01  1.184325e-01  1.812538e-04 -2.186963e-02
##            NW
##  9.693557e-03
```

```
Rate_b=sum(model_b$coefficients[2:6]*wartosci_b)+model_b$coefficients[1]
Rate_b
```

```
## (Intercept)
##    52.86884
```

Porownanie trzech modeli

```
#model gdzie występują wszystkie cechy
summary(aa)$adj.r.squared
```

```
## [1] 0.6278765
```

```
#model z 5a
summary(model_a)$adj.r.squared
```

```
## [1] 0.6398613
```

```
#model z 5b
summary(model_b)$adj.r.squared
```

```
## [1] 0.4298397
```

```
#Najlepszy współczynnik determinacji ma model_a dla cech Ex1,X,Ed, Age,U2.
#Współczynik ten wynosi 0.64, ten model jest "najlepszy"

cat("Rownanie najlepszej regresji opisujace zaleznosc:","Nasze_Rate=cecha[i]*wspol
czynnik[i]+wyraz wolny")
```

```
## Rownanie najlepszej regresji opisujace zaleznosc: Nasze_Rate=cecha[i]*wspolczyn
nik[i]+wyraz wolny
```

```
cat("Rate=Age*1.1067 + Ed*1.6652  +  Ex1*1.2689  +  U2*1.1204 + X*0.5253 -479.5483
")
```

```
## Rate=Age*1.1067 + Ed*1.6652  +  Ex1*1.2689  +  U2*1.1204 + X*0.5253 -479.5483
```

Porownanie prognoz

```
#model gdzie występują wszystkie cechy
Rate1
```

```
## (Intercept)
##    36.24069
```

```
#5a
Rate_a
```

```
## (Intercept)
##    62.33506
```

```
#5b
Rate_b
```

```
## (Intercept)
##    52.86884
```

```
#Jak widać wartości te są zupełnie różne, skorygowane współczynniki determinacji są na podobnym poziomie (ok 0.63). Wpływ na odpowiednie dobranie współczynnika Rate mogą miec ilość cech objaśniających, ale również odopowiedni ich dobór(każda ma inną wagę)
```