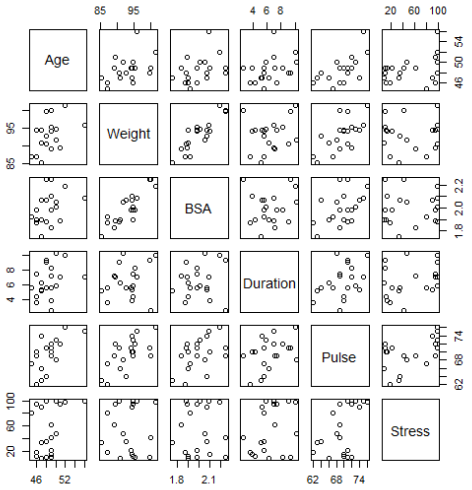
Detecting Multicollinearity in R

```
plot(BP1)
#BP1 is the design matrix
```




```
#calculate the correlation matrix
myCorr = cor(BP1)
myCorr
```

```
             Age      Weight         BSA     Duration      Pulse     Stress
Age       1.0000000 0.40734926 0.37845460 0.3437921 0.6187643 0.36822369
Weight    0.4073493 1.00000000 0.87530481 0.2006496 0.6593399 0.03435475
BSA       0.3784546 0.87530481 1.00000000 0.1305400 0.4648188 0.01844634
Duration  0.3437921 0.20064959 0.13054001 1.0000000 0.4015144 0.31163982
Pulse     0.6187643 0.65933987 0.46481881 0.4015144 1.0000000 0.50631008
Stress    0.3682237 0.03435475 0.01844634 0.3116398 0.5063101 1.00000000
```

```
# Eigensystem Analysis
eigen(cor(BP1))$values
```

```
3.01271192 1.38801608 0.70876081 0.51830290 0.30703939 0.06516889
```

```
# Condition Number: Ratio of max to min Eigen values of the correlation
matrix
max(eigen(cor(BP1))$values)/min(eigen(cor(BP1))$values)
kappa(cor(BP1), exact = TRUE)
```

```
46.2293
```

```
model = lm(BP~Age+Weight+BSA+Duration+Pulse+Stress, data = BP)


# variance inflation factors (Are any > 5 or 10?)

library(car)

vif(model)
```

```
    Age   Weight      BSA   Duration    Pulse   Stress
1.762807 8.417035 5.328751 1.237309   4.413575 1.834845
```

```
mean(vif(model))  #Is the mean VIF much bigger than 1?
```

```
3.832387
```

The following package, **mctest**, covers most of the methods described in the notes and many more.

https://cran.r-project.org/web/packages/mctest/mctest.pdf

Using the data set BP1,as an example, and contrasting the output with SPSS output, we have:

```
data(BP1)

x<-BP1[,-1]

y<-BP1[,1]

library (mctest)

mctest(x, y, type=c("b"), na.rm = TRUE, Inter=TRUE, method=NULL,

       corr=FALSE, detr=0.01, red=0.5, theil=0.5, cn=30, vif=10, tol=0.1,

       conf=0.95, cvif=10, leamer=0.1, all=FALSE)
```

```

Overall Multicollinearity Diagnostics

                       MC Results detection
Determinant |X'X|:         0.0307         0
Farrar Chi-Square:        56.2967         1
Red Indicator:             0.4424         0
Sum of Lambda Inverse:    22.9943         0
Theil's Method:           -1.4343         0
Condition Number:        201.4958         1

1 --> COLLINEARITY is detected
0 --> COLLINEARITY in not detected by the test
```

```
==================================
Eigvenvalues with INTERCEPT
                   Intercept    Age Weight     BSA Duration   Pulse   Stre
ss
Eigenvalues:          6.6558 0.2679 0.0714  0.0027   0.0011  0.0009   0.00
02
Condition Indeces:    1.0000 4.9842 9.6536 50.0668  77.3965 83.7514 201.49
58

==================================

Call:
imcdiag(x = x, y = y, method = method, corr = FALSE, vif = vif,
    tol = tol, conf = conf, cvif = cvif, leamer = leamer, all = all)


All Individual Multicollinearity Diagnostics Result

            VIF    TOL      Wi      Fi Leamer    CVIF Klein
Age      1.7628 0.5673  2.1359  2.8605 0.7532 -0.0039     0
Weight   8.4170 0.1188 20.7677 27.8139 0.3447 -0.0188     0
BSA      5.3288 0.1877 12.1205 16.2328 0.4332 -0.0119     0
Duration 1.2373 0.8082  0.6645  0.8899 0.8990 -0.0028     0
Pulse    4.4136 0.2266  9.5580 12.8009 0.4760 -0.0099     0
Stress   1.8348 0.5450  2.3376  3.1307 0.7382 -0.0041     0

1 --> COLLINEARITY is detected
0 --> COLLINEARITY in not detected by the test

Duration , Pulse , Stress , coefficient(s) are non-significant may be due
to multicollinearity

R-square of y on all x: 0.9962

* use method argument to check which regressors may be the reason of colli
nearity
==================================
```