Lesson 12

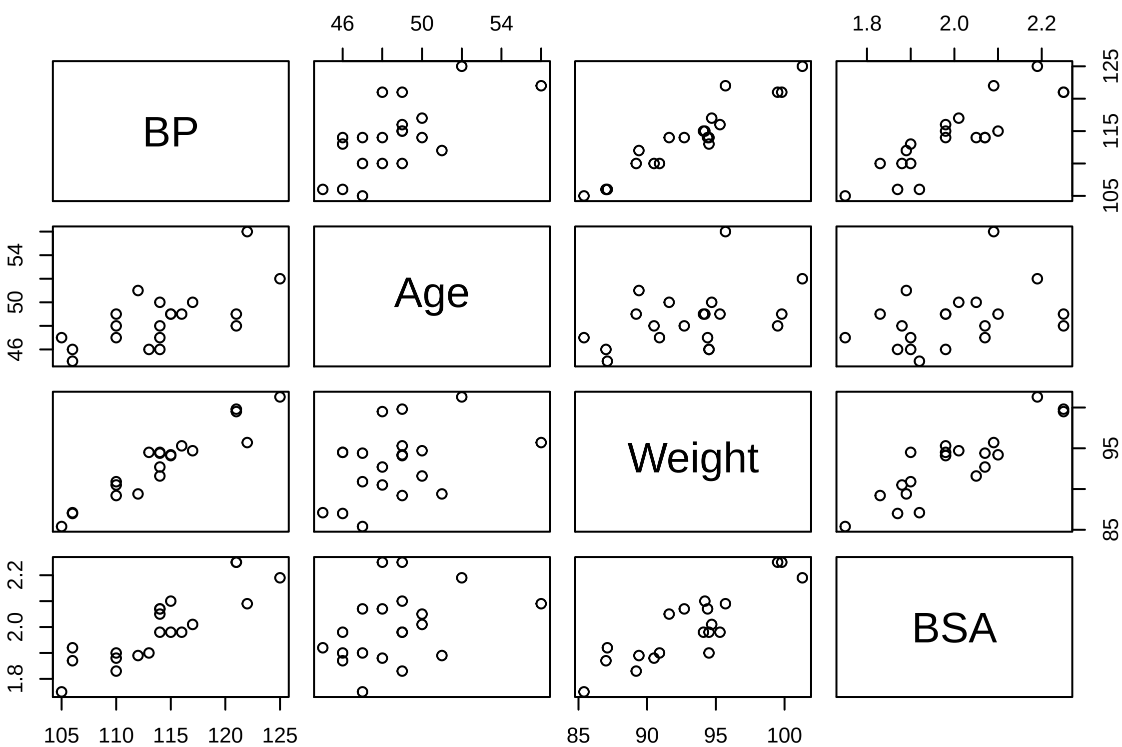
Christopher A. Swenson (chris@cswenson.com)

11/27/2021

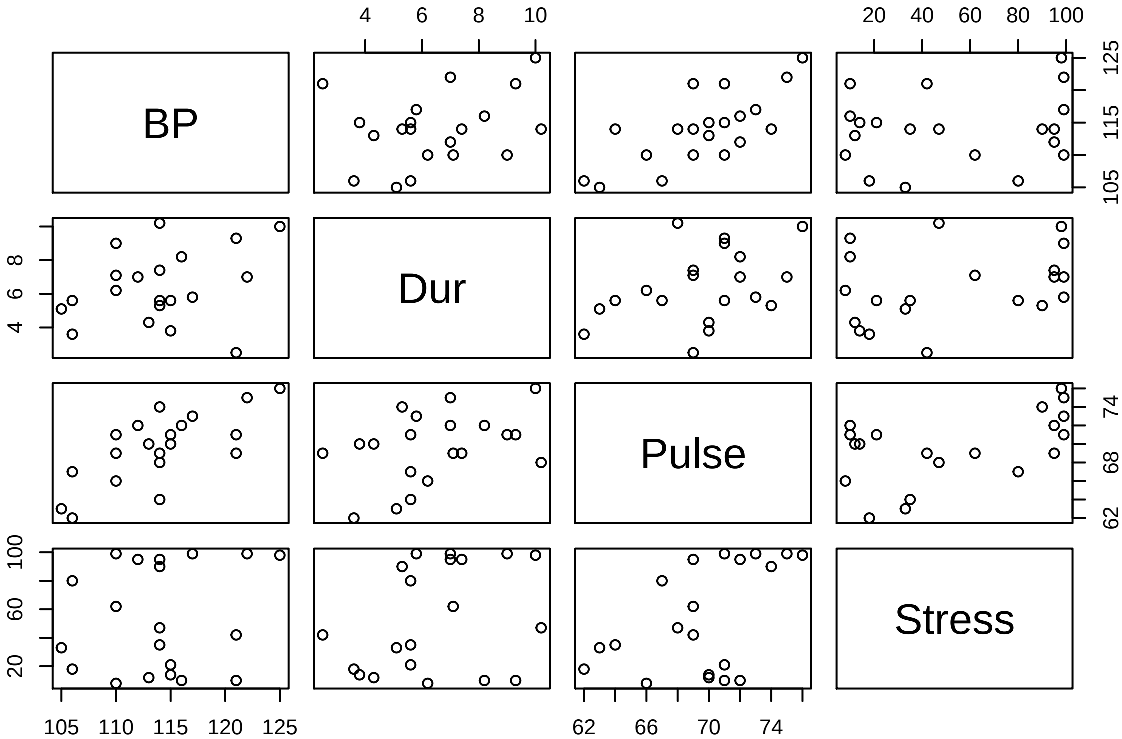
Blood pressure (multicollinearity)

Load the bloodpress data. Create scatterplot matrices of the data. Calculate correlations between the variables.

```
bloodpress <- read.table("./Data/bloodpress.txt", header=T)
attach(bloodpress)

pairs(bloodpress[,c(2:5)])
```




```
pairs(bloodpress[,c(2,6:8)])
```




```
round(cor(bloodpress[,c(2:8)]),3)
```

```
##            BP   Age Weight   BSA   Dur Pulse Stress
## BP      1.000 0.659  0.950 0.866 0.293 0.721  0.164
## Age     0.659 1.000  0.407 0.378 0.344 0.619  0.368
## Weight  0.950 0.407  1.000 0.875 0.201 0.659  0.034
## BSA     0.866 0.378  0.875 1.000 0.131 0.465  0.018
## Dur     0.293 0.344  0.201 0.131 1.000 0.402  0.312
## Pulse   0.721 0.619  0.659 0.465 0.402 1.000  0.506
## Stress  0.164 0.368  0.034 0.018 0.312 0.506  1.000
```

```
#           BP   Age Weight   BSA   Dur Pulse Stress
# BP     1.000 0.659  0.950 0.866 0.293 0.721  0.164
# Age    0.659 1.000  0.407 0.378 0.344 0.619  0.368
# Weight 0.950 0.407  1.000 0.875 0.201 0.659  0.034
# BSA    0.866 0.378  0.875 1.000 0.131 0.465  0.018
# Dur    0.293 0.344  0.201 0.131 1.000 0.402  0.312
# Pulse  0.721 0.619  0.659 0.465 0.402 1.000  0.506
# Stress 0.164 0.368  0.034 0.018 0.312 0.506  1.000

detach(bloodpress)
```

Uncorrelated predictors (no multicollinearity)

Load the uncorrpreds data. Create a scatterplot matrix of the data. Calculate the correlation between the predictors. Fit a simple linear regression model of y vs . Fit a simple linear regression model of y vs . Fit a multiple linear regression model of y vs + . Fit a multiple linear regression model of y vs + . Use the scatter3d function in the car package to create a 3D scatterplot of the data with the fitted plane for a multiple linear regression model of y vs + .

```r
library(car)
```

```
## Loading required package: carData
```

```r
uncorrpreds <- read.table("./Data/uncorrpreds.txt", header=T)
attach(uncorrpreds)

pairs(uncorrpreds)

cor(x1,x2) # 0
```

```
## [1] 0
```

```r
model.1 <- lm(y ~ x1)
summary(model.1)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.500 -2.500  0.500  1.938  3.750
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   52.750      3.346  15.764 4.13e-06 ***
## x1            -1.625      1.058  -1.536    0.176
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.993 on 6 degrees of freedom
## Multiple R-squared:  0.2821, Adjusted R-squared:  0.1625
## F-statistic: 2.358 on 1 and 6 DF,  p-value: 0.1755
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   49.500      4.655  10.634 4.07e-05 ***
# x1            -1.000      1.472  -0.679    0.522
anova(model.1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value Pr(>F)
## x1         1 21.125 21.1250  2.3581 0.1755
## Residuals  6 53.750  8.9583
```

```r
#           Df Sum Sq Mean Sq F value Pr(>F)
# x1         1      8   8.000  0.4615 0.5222
# Residuals  6    104  17.333

model.2 <- lm(y ~ x2)
summary(model.2)
```

```
##
## Call:
## lm(formula = y ~ x2)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.500 -1.688  0.125  1.000  3.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  62.1250     4.7888  12.973 1.29e-05 ***
## x2           -2.3750     0.7873  -3.017   0.0235 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.227 on 6 degrees of freedom
## Multiple R-squared:  0.6027, Adjusted R-squared:  0.5364
## F-statistic: 9.101 on 1 and 6 DF,  p-value: 0.02349
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   57.000      8.213   6.940 0.000444 ***
# x2            -1.750      1.350  -1.296 0.242545
anova(model.2)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value  Pr(>F)
## x2         1 45.125  45.125  9.1008 0.02349 *
## Residuals  6 29.750   4.958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value Pr(>F)
# x2         1   24.5  24.500    1.68 0.2425
# Residuals  6   87.5  14.583

model.12 <- lm(y ~ x1 + x2)
summary(model.12)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      1      2      3      4      5      6      7      8
##  0.125 -0.875  1.875 -1.125  1.375 -0.625  0.125 -0.875
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  67.0000     3.1494  21.274 4.25e-06 ***
## x1           -1.6250     0.4644  -3.499  0.01729 *
## x2           -2.3750     0.4644  -5.115  0.00372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.313 on 5 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8387
```

```
## F-statistic:  19.2 on 2 and 5 DF,  p-value: 0.004504
```

```r
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)   60.000      9.562   6.275  0.00151 **
# x1            -1.000      1.410  -0.709  0.50982
# x2            -1.750      1.410  -1.241  0.26954
anova(model.12)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## x1         1 21.125  21.125  12.246 0.017294 *
## x2         1 45.125  45.125  26.159 0.003724 **
## Residuals  5  8.625   1.725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#           Df Sum Sq Mean Sq F value Pr(>F)
# x1         1    8.0     8.0  0.5031 0.5098
# x2         1   24.5    24.5  1.5409 0.2695
# Residuals  5   79.5    15.9

model.21 <- lm(y ~ x2 + x1)
summary(model.21)
```

```
##
## Call:
## lm(formula = y ~ x2 + x1)
##
## Residuals:
##      1      2      3      4      5      6      7      8
##  0.125 -0.875  1.875 -1.125  1.375 -0.625  0.125 -0.875
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  67.0000     3.1494  21.274 4.25e-06 ***
## x2           -2.3750     0.4644  -5.115  0.00372 **
## x1           -1.6250     0.4644  -3.499  0.01729 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.313 on 5 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8387
## F-statistic:  19.2 on 2 and 5 DF,  p-value: 0.004504
```

```r
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)   60.000      9.562   6.275  0.00151 **
# x2            -1.750      1.410  -1.241  0.26954
# x1            -1.000      1.410  -0.709  0.50982
anova(model.21)
```

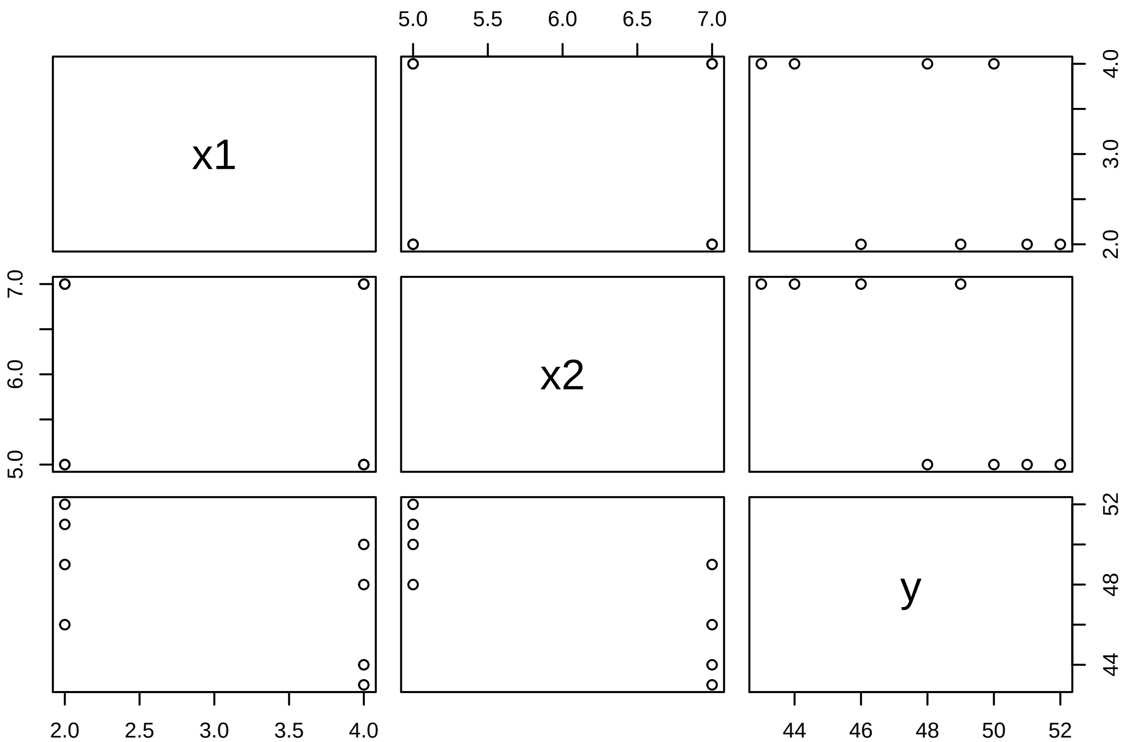
```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## x2          1 45.125  45.125  26.159 0.003724 **
## x1          1 21.125  21.125  12.246 0.017294 *
## Residuals   5  8.625   1.725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value Pr(>F)
# x2         1   24.5    24.5  1.5409 0.2695
# x1         1    8.0     8.0  0.5031 0.5098
# Residuals  5   79.5    15.9

# library(car)
scatter3d(y ~ x1 + x2)
```

```
## Loading required namespace: rgl
```

```
## Loading required namespace: mgcv
```

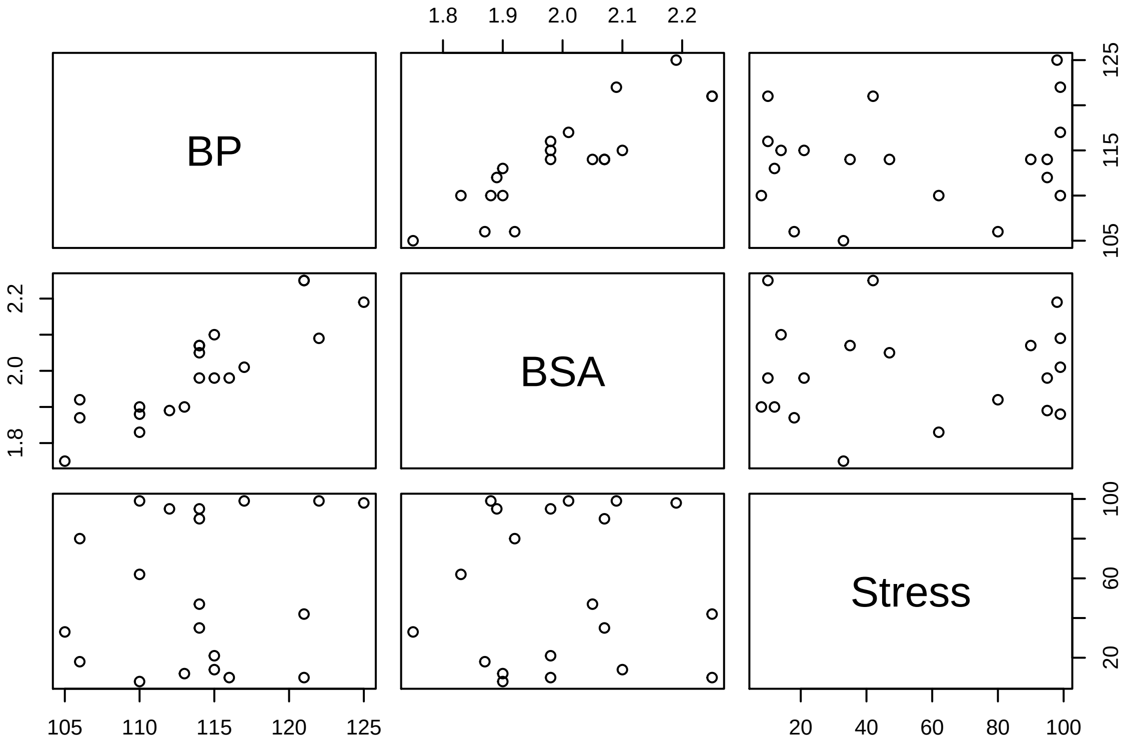



```
detach(uncorrpreds)
```

Blood pressure (predictors with almost no multicollinearity)

Load the bloodpress data. Create a scatterplot matrix of the data. Fit a simple linear regression model of BP vs Stress. Fit a simple linear regression model of BP vs BSA. Fit a multiple linear regression model of BP vs Stress + BSA. Fit a multiple linear regression model of BP vs BSA + Stress. Use the scatter3d function in the car package to create a 3D scatterplot of the data with the fitted plane for a multiple linear regression model of BP vs Stress + BSA.

```
bloodpress <- read.table("./Data/bloodpress.txt", header=T)
attach(bloodpress)

pairs(bloodpress[,c(2,5,8)])
```




```r
model.1 <- lm(BP ~ Stress)
summary(model.1)
```

```
##
## Call:
## lm(formula = BP ~ Stress)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6394 -3.3014  0.0722  2.2181  9.9287
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 112.71997    2.19345  51.389   <2e-16 ***
## Stress        0.02399    0.03404   0.705     0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.502 on 18 degrees of freedom
## Multiple R-squared:  0.02686,    Adjusted R-squared:  -0.0272
## F-statistic: 0.4969 on 1 and 18 DF,  p-value: 0.4899
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 112.71997    2.19345  51.389   <2e-16 ***
# Stress        0.02399    0.03404   0.705     0.49
anova(model.1)
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## Stress     1  15.04  15.044  0.4969 0.4899
## Residuals 18 544.96  30.275
```

```r
#           Df Sum Sq Mean Sq F value Pr(>F)
# Stress     1  15.04  15.044  0.4969 0.4899
# Residuals 18 544.96  30.275

model.2 <- lm(BP ~ BSA)
summary(model.2)
```

```
##
## Call:
## lm(formula = BP ~ BSA)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.314 -1.963 -0.197  1.934  4.831
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   45.183      9.392   4.811  0.00014 ***
## BSA           34.443      4.690   7.343 8.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.79 on 18 degrees of freedom
## Multiple R-squared:  0.7497, Adjusted R-squared:  0.7358
## F-statistic: 53.93 on 1 and 18 DF,  p-value: 8.114e-07
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   45.183      9.392   4.811  0.00014 ***
# BSA           34.443      4.690   7.343 8.11e-07 ***
anova(model.2)
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## BSA        1 419.86  419.86  53.927 8.114e-07 ***
## Residuals 18 140.14    7.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#           Df Sum Sq Mean Sq F value Pr(>F)
# BSA        1 419.86  419.86  53.927 8.114e-07 ***
# Residuals 18 140.14    7.79

model.12 <- lm(BP ~ Stress + BSA)
summary(model.12)
```

```
##
## Call:
## lm(formula = BP ~ Stress + BSA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -5.8992 -1.6483 -0.1643  1.7790  3.8524
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 44.24452    9.26104   4.777 0.000175 ***
## Stress       0.02166    0.01697   1.277 0.218924
## BSA         34.33423    4.61110   7.446 9.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.743 on 17 degrees of freedom
## Multiple R-squared:  0.7716, Adjusted R-squared:  0.7448
## F-statistic: 28.72 on 2 and 17 DF,  p-value: 3.534e-06
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 44.24452    9.26104   4.777 0.000175 ***
# Stress       0.02166    0.01697   1.277 0.218924
# BSA         34.33423    4.61110   7.446 9.56e-07 ***
anova(model.12)
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Stress     1  15.04   15.04  1.9998    0.1754
## BSA        1 417.07  417.07 55.4430 9.561e-07 ***
## Residuals 17 127.88    7.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value Pr(>F)
# Stress     1  15.04   15.04  1.9998    0.1754
# BSA        1 417.07  417.07 55.4430 9.561e-07 ***
# Residuals 17 127.88    7.52

model.21 <- lm(BP ~ BSA + Stress)
summary(model.21)
```

```
##
## Call:
## lm(formula = BP ~ BSA + Stress)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8992 -1.6483 -0.1643  1.7790  3.8524
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 44.24452    9.26104   4.777 0.000175 ***
## BSA         34.33423    4.61110   7.446 9.56e-07 ***
## Stress       0.02166    0.01697   1.277 0.218924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.743 on 17 degrees of freedom
```

```
## Multiple R-squared:  0.7716, Adjusted R-squared:  0.7448
## F-statistic: 28.72 on 2 and 17 DF,  p-value: 3.534e-06
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 44.24452    9.26104   4.777 0.000175 ***
# BSA         34.33423    4.61110   7.446 9.56e-07 ***
# Stress       0.02166    0.01697   1.277 0.218924
anova(model.21)
```

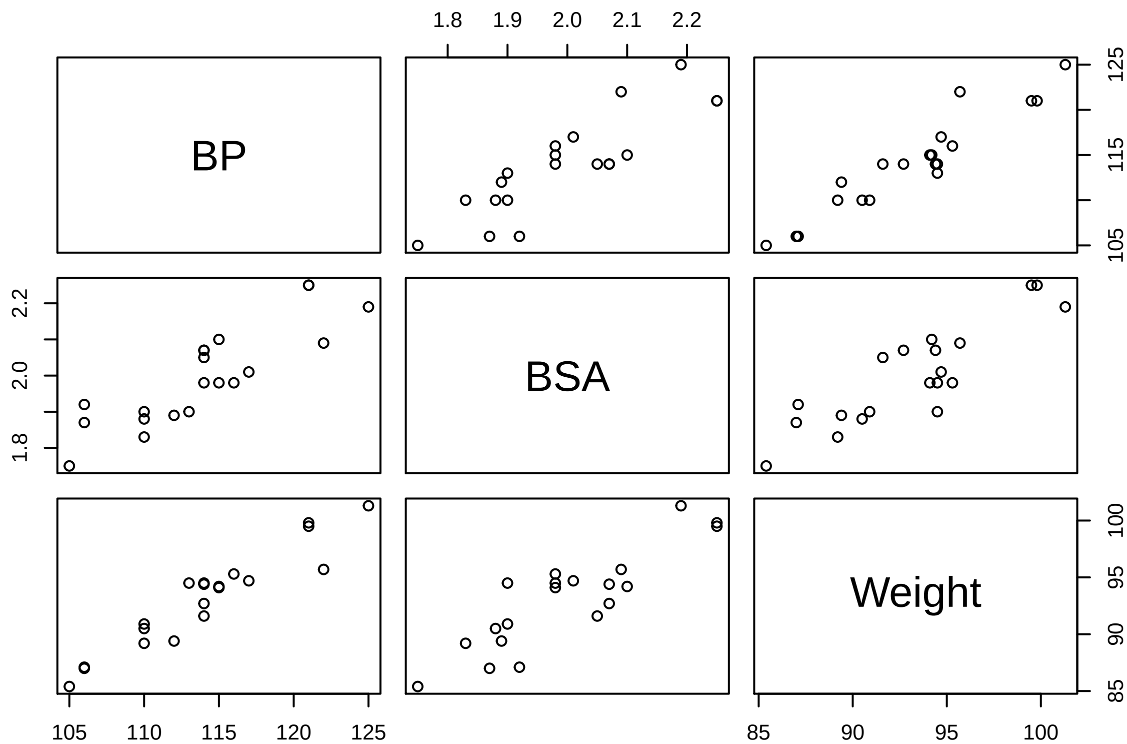
```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## BSA        1 419.86  419.86 55.8132 9.149e-07 ***
## Stress     1  12.26   12.26  1.6296    0.2189
## Residuals 17 127.88    7.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value Pr(>F)
# BSA        1 419.86  419.86 55.8132 9.149e-07 ***
# Stress     1  12.26   12.26  1.6296    0.2189
# Residuals 17 127.88    7.52

scatter3d(BP ~ Stress + BSA)

detach(bloodpress)
```

Blood pressure (predictors with high multicollinearity)

Load the bloodpress data. Create a scatterplot matrix of the data. Fit a simple linear regression model of BP vs Weight. Fit a simple linear regression model of BP vs BSA. Fit a multiple linear regression model of BP vs Weight + BSA. Fit a multiple linear regression model of BP vs BSA + Weight. Use the scatter3d function in the car package to create a 3D scatterplot of the data with the fitted plane for a multiple linear regression model of BP vs Weight + BSA. Predict BP for Weight=92 and BSA=2 for the two simple linear regression models and the multiple linear regression model.

```
bloodpress <- read.table("./Data/bloodpress.txt", header=T)
attach(bloodpress)

pairs(bloodpress[,c(2,5,4)])
```




```r
model.1 <- lm(BP ~ Weight)
summary(model.1)
```

```
##
## Call:
## lm(formula = BP ~ Weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6933 -0.9318 -0.4935  0.7703  4.8656
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.20531    8.66333   0.255    0.802
## Weight       1.20093    0.09297  12.917 1.53e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.74 on 18 degrees of freedom
## Multiple R-squared:  0.9026, Adjusted R-squared:  0.8972
## F-statistic: 166.9 on 1 and 18 DF,  p-value: 1.528e-10
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  2.20531    8.66333   0.255    0.802
# Weight       1.20093    0.09297  12.917 1.53e-10 ***
anova(model.1)
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## Weight     1 505.47  505.47  166.86 1.528e-10 ***
## Residuals 18  54.53    3.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#           Df Sum Sq Mean Sq F value Pr(>F)
# Weight     1 505.47  505.47  166.86 1.528e-10 ***
# Residuals 18  54.53    3.03

model.2 <- lm(BP ~ BSA)
summary(model.2)
```

```
##
## Call:
## lm(formula = BP ~ BSA)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.314 -1.963 -0.197  1.934  4.831
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   45.183      9.392   4.811  0.00014 ***
## BSA           34.443      4.690   7.343 8.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.79 on 18 degrees of freedom
## Multiple R-squared:  0.7497, Adjusted R-squared:  0.7358
## F-statistic: 53.93 on 1 and 18 DF,  p-value: 8.114e-07
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   45.183      9.392   4.811  0.00014 ***
# BSA           34.443      4.690   7.343 8.11e-07 ***
anova(model.2)
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## BSA        1 419.86  419.86  53.927 8.114e-07 ***
## Residuals 18 140.14    7.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#           Df Sum Sq Mean Sq F value Pr(>F)
# BSA        1 419.86  419.86  53.927 8.114e-07 ***
# Residuals 18 140.14    7.79

model.12 <- lm(BP ~ Weight + BSA)
summary(model.12)
```

```
##
## Call:
## lm(formula = BP ~ Weight + BSA)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -1.8932 -1.1961 -0.4061  1.0764  4.7524 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   5.6534     9.3925   0.602    0.555    
## Weight        1.0387     0.1927   5.392 4.87e-05 ***
## BSA           5.8313     6.0627   0.962    0.350    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.744 on 17 degrees of freedom
## Multiple R-squared:  0.9077, Adjusted R-squared:  0.8968 
## F-statistic: 83.54 on 2 and 17 DF,  p-value: 1.607e-09
```

```
#             Estimate Std. Error t value Pr(>|t|)    
# (Intercept)   5.6534     9.3925   0.602    0.555    
# Weight        1.0387     0.1927   5.392 4.87e-05 ***
# BSA           5.8313     6.0627   0.962    0.350 
anova(model.12)
```

```
## Analysis of Variance Table
## 
## Response: BP
##           Df Sum Sq Mean Sq  F value    Pr(>F)    
## Weight     1 505.47  505.47 166.1648 3.341e-10 ***
## BSA        1   2.81    2.81   0.9251    0.3496    
## Residuals 17  51.71    3.04                       
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value Pr(>F)    
# Weight     1 505.47  505.47 166.1648 3.341e-10 ***
# BSA        1   2.81    2.81   0.9251    0.3496    
# Residuals 17  51.71    3.04  

model.21 <- lm(BP ~ BSA + Weight)
summary(model.21)
```

```
## 
## Call:
## lm(formula = BP ~ BSA + Weight)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -1.8932 -1.1961 -0.4061  1.0764  4.7524 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   5.6534     9.3925   0.602    0.555    
## BSA           5.8313     6.0627   0.962    0.350    
## Weight        1.0387     0.1927   5.392 4.87e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.744 on 17 degrees of freedom
## Multiple R-squared:  0.9077, Adjusted R-squared:  0.8968
## F-statistic: 83.54 on 2 and 17 DF,  p-value: 1.607e-09
```

```
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)    5.6534     9.3925   0.602    0.555
# BSA            5.8313     6.0627   0.962    0.350
# Weight         1.0387     0.1927   5.392 4.87e-05 ***
anova(model.21)
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## BSA        1 419.86  419.86 138.021 1.391e-09 ***
## Weight     1  88.43   88.43  29.069 4.871e-05 ***
## Residuals 17  51.71    3.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value Pr(>F)
# BSA        1 419.86  419.86 138.021 1.391e-09 ***
# Weight     1  88.43   88.43  29.069 4.871e-05 ***
# Residuals 17  51.71    3.04

scatter3d(BP ~ Weight + BSA)

predict(model.1, interval="prediction",
        newdata=data.frame(Weight=92))
```

```
##       fit     lwr     upr
## 1 112.691 108.938 116.444
```

```
#       fit     lwr     upr
# 1 112.691 108.938 116.444

predict(model.2, interval="prediction",
        newdata=data.frame(BSA=2))
```

```
##        fit      lwr      upr
## 1 114.0689 108.0619 120.0758
```

```
#        fit      lwr      upr
# 1 114.0689 108.0619 120.0758

predict(model.12, interval="prediction",
        newdata=data.frame(Weight=92, BSA=2))
```

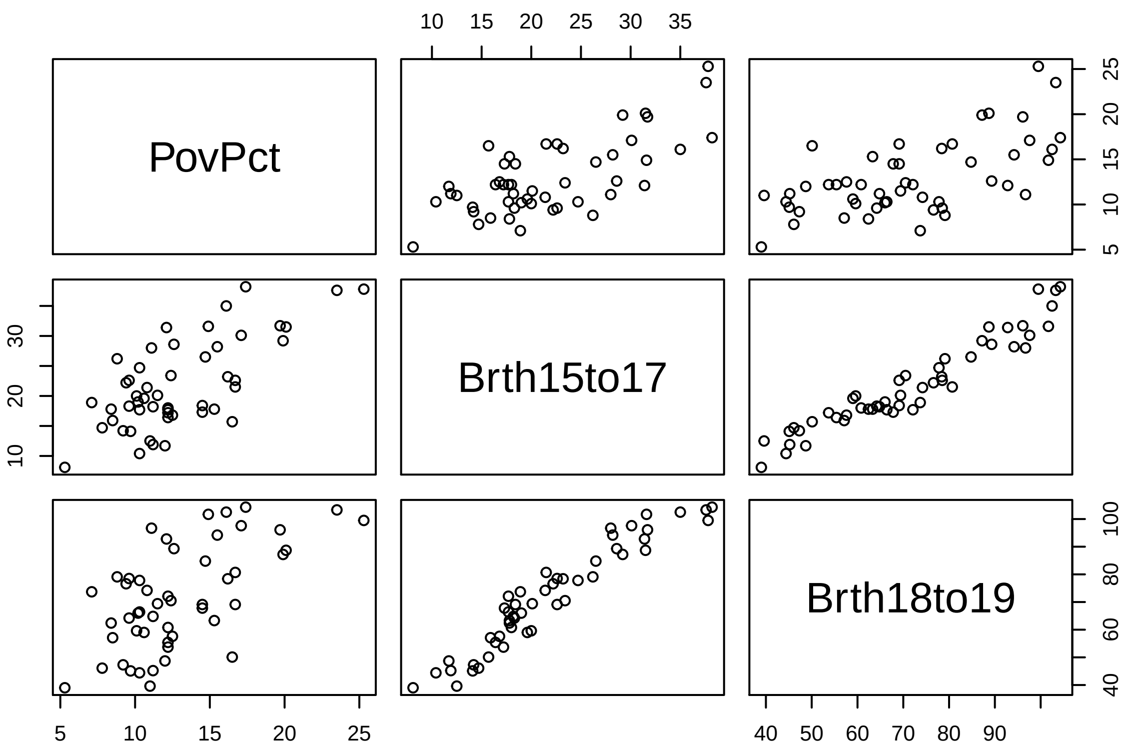
```
##        fit      lwr      upr
## 1 112.8794 109.0801 116.6787
```

```
#        fit      lwr      upr
# 1 112.8794 109.0801 116.6787

detach(bloodpress)
```

Poverty and teen birth rate (high multicollinearity)

Load the poverty data and remove the District of Columbia. Create a scatterplot matrix of the data. Fit a simple linear regression model of PovPct vs Brth15to17. Fit a simple linear regression model of PovPct vs Brth18to19. Fit a multiple linear regression model of PovPct vs Brth15to17 + Brth18to19.

```
poverty <- read.table("./Data/poverty.txt", header=T)
poverty <- poverty[poverty$Location!="District_of_Columbia",]
attach(poverty)

pairs(poverty[,c(2:4)])
```




```
model.1 <- lm(PovPct ~ Brth15to17)
summary(model.1)
```

```
##
## Call:
## lm(formula = PovPct ~ Brth15to17)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8312 -2.0912 -0.2901  2.5741  6.1775
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.4871     1.3181   3.404  0.00135 **
## Brth15to17    0.3872     0.0572   6.768 1.67e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.982 on 48 degrees of freedom
## Multiple R-squared:  0.4883, Adjusted R-squared:  0.4777
```

```
## F-statistic: 45.81 on 1 and 48 DF,  p-value: 1.666e-08
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   4.4871     1.3181   3.404  0.00135 **
# Brth15to17    0.3872     0.0572   6.768 1.67e-08 ***
# ---
# Residual standard error: 2.982 on 48 degrees of freedom
# Multiple R-squared:  0.4883,  Adjusted R-squared:  0.4777
# F-statistic: 45.81 on 1 and 48 DF,  p-value: 1.666e-08

model.2 <- lm(PovPct ~ Brth18to19)
summary(model.2)
```

```
##
## Call:
## lm(formula = PovPct ~ Brth18to19)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1542 -2.3119 -0.4056  2.0195  8.4746
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.05279    1.83169   1.667    0.102
## Brth18to19   0.13842    0.02482   5.576 1.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.248 on 48 degrees of freedom
## Multiple R-squared:  0.3931, Adjusted R-squared:  0.3805
## F-statistic: 31.09 on 1 and 48 DF,  p-value: 1.106e-06
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  3.05279    1.83169   1.667    0.102
# Brth18to19   0.13842    0.02482   5.576 1.11e-06 ***
# ---
# Residual standard error: 3.248 on 48 degrees of freedom
# Multiple R-squared:  0.3931,  Adjusted R-squared:  0.3805
# F-statistic: 31.09 on 1 and 48 DF,  p-value: 1.106e-06

model.12 <- lm(PovPct ~ Brth15to17 + Brth18to19)
summary(model.12)
```

```
##
## Call:
## lm(formula = PovPct ~ Brth15to17 + Brth18to19)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1177 -2.2548 -0.3315  2.5948  5.2562
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.43963    1.95904   3.287  0.00192 **
## Brth15to17   0.63235    0.19178   3.297  0.00186 **
```

```
## Brth18to19  -0.10227    0.07642  -1.338  0.18724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.958 on 47 degrees of freedom
## Multiple R-squared:  0.5071, Adjusted R-squared:  0.4861
## F-statistic: 24.18 on 2 and 47 DF,  p-value: 6.017e-08
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  6.43963    1.95904   3.287  0.00192 **
# Brth15to17   0.63235    0.19178   3.297  0.00186 **
# Brth18to19  -0.10227    0.07642  -1.338  0.18724
# ---
# Residual standard error: 2.958 on 47 degrees of freedom
# Multiple R-squared:  0.5071,  Adjusted R-squared:  0.4861
# F-statistic: 24.18 on 2 and 47 DF,  p-value: 6.017e-08

detach(poverty)
```

Blood pressure (high multicollinearity)

Load the bloodpress data. Fit a multiple linear regression model of BP vs Age + Weight + BSA + Dur + Pulse + Stress. Use the vif function in the car package to calculate variance inflation factors. Fit a multiple linear regression model of Weight vs Age + BSA + Dur + Pulse + Stress and confirm the VIF value for Weight as 1/(1-) for this model. Fit a multiple linear regression model of BP vs Age + Weight + Dur + Stress. Use the vif function in the car package to calculate variance inflation factors.

```
bloodpress <- read.table("./Data/bloodpress.txt", header=T)

attach(bloodpress)

model.1 <- lm(BP ~ Age + Weight + BSA + Dur + Pulse + Stress)
summary(model.1)
```

```
##
## Call:
## lm(formula = BP ~ Age + Weight + BSA + Dur + Pulse + Stress)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93213 -0.11314  0.03064  0.21834  0.48454
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12.870476   2.556650  -5.034 0.000229 ***
## Age           0.703259   0.049606  14.177 2.76e-09 ***
## Weight        0.969920   0.063108  15.369 1.02e-09 ***
## BSA           3.776491   1.580151   2.390 0.032694 *
## Dur           0.068383   0.048441   1.412 0.181534
## Pulse        -0.084485   0.051609  -1.637 0.125594
## Stress        0.005572   0.003412   1.633 0.126491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4072 on 13 degrees of freedom
## Multiple R-squared:  0.9962, Adjusted R-squared:  0.9944
## F-statistic: 560.6 on 6 and 13 DF,  p-value: 6.395e-15
```

```
#                Estimate Std. Error t value Pr(>|t|)
# (Intercept) -12.870476   2.556650  -5.034 0.000229 ***
# Age           0.703259   0.049606  14.177 2.76e-09 ***
# Weight        0.969920   0.063108  15.369 1.02e-09 ***
# BSA           3.776491   1.580151   2.390 0.032694 *
# Dur           0.068383   0.048441   1.412 0.181534
# Pulse        -0.084485   0.051609  -1.637 0.125594
# Stress        0.005572   0.003412   1.633 0.126491
# ---
# Residual standard error: 0.4072 on 13 degrees of freedom
# Multiple R-squared:  0.9962,  Adjusted R-squared:  0.9944
# F-statistic: 560.6 on 6 and 13 DF,  p-value: 6.395e-15

# library(car)
vif(model.1)
```

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

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

model.2 <- lm(Weight ~ Age + BSA + Dur + Pulse + Stress)
summary(model.2)
```

```
##
## Call:
## lm(formula = Weight ~ Age + BSA + Dur + Pulse + Stress)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7697 -1.0120  0.1960  0.6955  2.7035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.674438   9.464742   2.079  0.05651 .
## Age         -0.144643   0.206491  -0.700  0.49510
## BSA         21.421654   3.464586   6.183 2.38e-05 ***
## Dur          0.008696   0.205134   0.042  0.96678
## Pulse        0.557697   0.159853   3.489  0.00361 **
## Stress      -0.022997   0.013079  -1.758  0.10052
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.725 on 14 degrees of freedom
## Multiple R-squared:  0.8812, Adjusted R-squared:  0.8388
## F-statistic: 20.77 on 5 and 14 DF,  p-value: 5.046e-06
```

```
# Residual standard error: 1.725 on 14 degrees of freedom
# Multiple R-squared:  0.8812,  Adjusted R-squared:  0.8388
# F-statistic: 20.77 on 5 and 14 DF,  p-value: 5.046e-06
```

```r
1/(1-summary(model.2)$r.squared) # 8.417035
```

```
## [1] 8.417035
```

```r
model.3 <- lm(BP ~ Age + Weight + Dur + Stress)
summary(model.3)
```

```
##
## Call:
## lm(formula = BP ~ Age + Weight + Dur + Stress)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11359 -0.29586  0.01515  0.27506  0.88674
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.869829   3.195296  -4.967 0.000169 ***
## Age           0.683741   0.061195  11.173 1.14e-08 ***
## Weight        1.034128   0.032672  31.652 3.76e-15 ***
## Dur           0.039889   0.064486   0.619 0.545485
## Stress        0.002184   0.003794   0.576 0.573304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5505 on 15 degrees of freedom
## Multiple R-squared:  0.9919, Adjusted R-squared:  0.9897
## F-statistic: 458.3 on 4 and 15 DF,  p-value: 1.764e-15
```

```r
#               Estimate Std. Error t value Pr(>|t|)
# (Intercept) -15.869829   3.195296  -4.967 0.000169 ***
# Age           0.683741   0.061195  11.173 1.14e-08 ***
# Weight        1.034128   0.032672  31.652 3.76e-15 ***
# Dur           0.039889   0.064486   0.619 0.545485
# Stress        0.002184   0.003794   0.576 0.573304
# ---
# Residual standard error: 0.5505 on 15 degrees of freedom
# Multiple R-squared:  0.9919,  Adjusted R-squared:  0.9897
# F-statistic: 458.3 on 4 and 15 DF,  p-value: 1.764e-15

vif(model.3)
```

```
##      Age   Weight      Dur   Stress
## 1.468245 1.234653 1.200060 1.241117
```

```r
#      Age   Weight      Dur   Stress
# 1.468245 1.234653 1.200060 1.241117

detach(bloodpress)
```

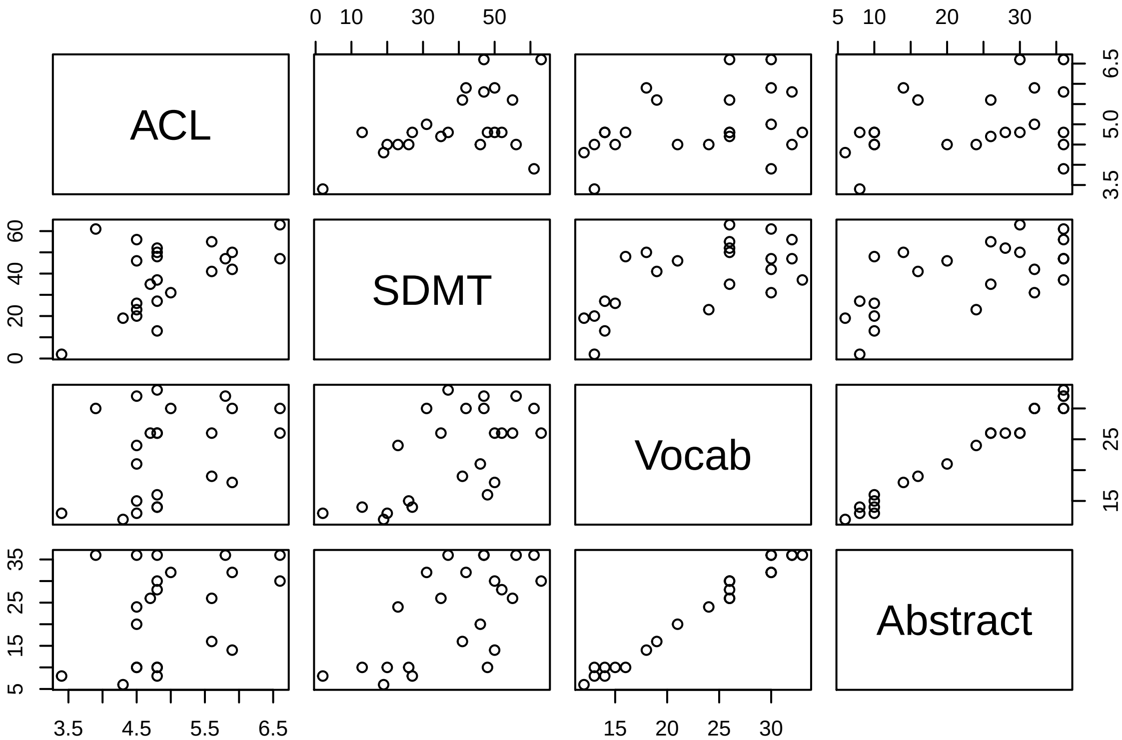
Allen Cognitive Level study (reducing data-based multicollinearity)

Load the sampled allentestn23 data. Create a scatterplot matrix of the data. Calculate the correlation between Vocab and Abstract. Fit a multiple linear regression model of ACL vs SDMT + Vocab + Abstract. Use the vif function in the car package to calculate variance inflation factors. Repeat for the full allentest data.

```r
allentestn23 <- read.table("./Data/allentestn23.txt", header=T)
attach(allentestn23)
```

```
## The following object is masked from package:carData:
##
##     Vocab
```

```r
pairs(allentestn23[,2:5])
```




```r
cor(Vocab, Abstract) # 0.9897771
```

```
## [1] 0.9897771
```

```r
model.1 <- lm(ACL ~ SDMT + Vocab + Abstract)
summary(model.1)
```

```
##
## Call:
## lm(formula = ACL ~ SDMT + Vocab + Abstract)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6172 -0.4044 -0.1293  0.5224  1.4084
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74711    1.34237   2.791   0.0116 *
## SDMT         0.02326    0.01273   1.827   0.0834 .
## Vocab        0.02825    0.15239   0.185   0.8549
## Abstract    -0.01379    0.10055  -0.137   0.8924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.7344 on 19 degrees of freedom
## Multiple R-squared:  0.2645, Adjusted R-squared:  0.1484
## F-statistic: 2.278 on 3 and 19 DF,  p-value: 0.1124
```

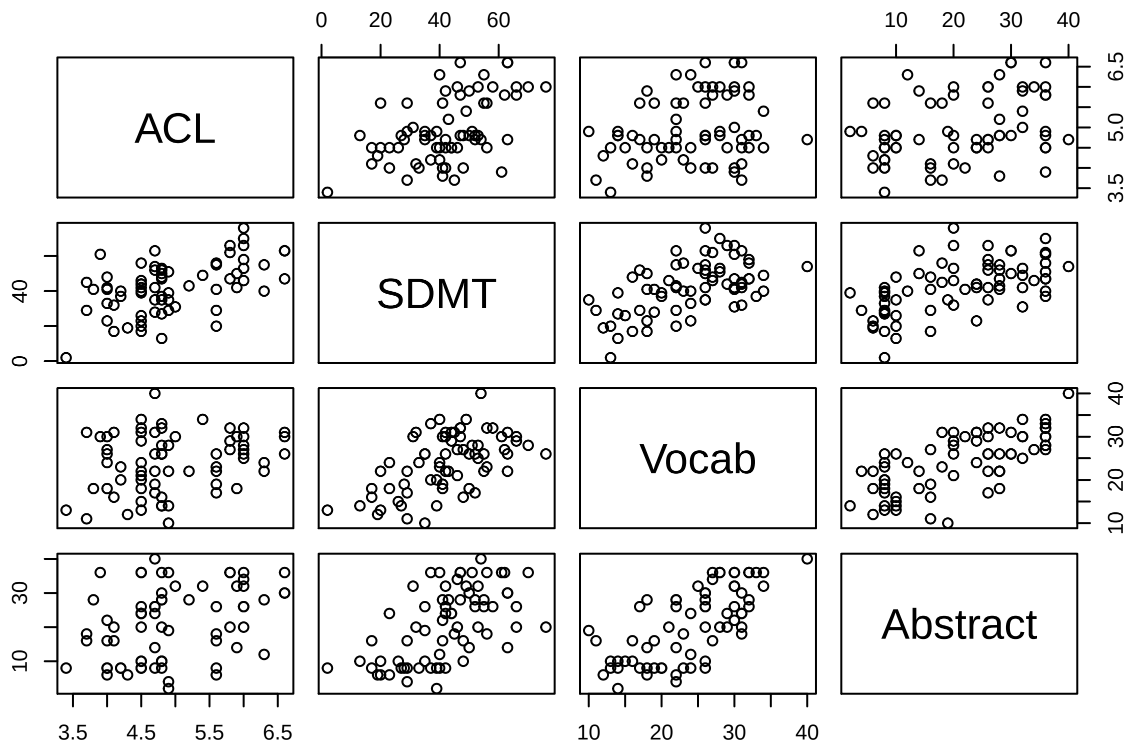
```r
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  3.74711    1.34237   2.791   0.0116 *
# SDMT         0.02326    0.01273   1.827   0.0834 .
# Vocab        0.02825    0.15239   0.185   0.8549
# Abstract    -0.01379    0.10055  -0.137   0.8924
# ---
# Residual standard error: 0.7344 on 19 degrees of freedom
# Multiple R-squared:  0.2645,  Adjusted R-squared:  0.1484
# F-statistic: 2.278 on 3 and 19 DF,  p-value: 0.1124

vif(model.1)
```

```
##      SDMT     Vocab  Abstract
##  1.726185 49.286239 50.603085
```

```r
#     SDMT     Vocab  Abstract
# 1.726185 49.286239 50.603085

detach(allentestn23)

allentest <- read.table("./Data/allentest.txt", header=T)
attach(allentest)
```

```
## The following object is masked from package:carData:
##
##     Vocab
```

```r
pairs(allentest[,2:5])
```




```r
cor(Vocab, Abstract) # 0.6978405
```

```
## [1] 0.6978405
```

```r
model.1 <- lm(ACL ~ SDMT + Vocab + Abstract)
summary(model.1)
```

```
##
## Call:
## lm(formula = ACL ~ SDMT + Vocab + Abstract)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63459 -0.46009 -0.02471  0.33624  1.52886
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.946347   0.338069  11.673  < 2e-16 ***
## SDMT         0.027404   0.007168   3.823 0.000298 ***
## Vocab       -0.017397   0.018077  -0.962 0.339428
## Abstract     0.012182   0.011585   1.051 0.296926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6878 on 65 degrees of freedom
## Multiple R-squared:  0.2857, Adjusted R-squared:  0.2528
## F-statistic: 8.668 on 3 and 65 DF,  p-value: 6.414e-05
```

```r
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  3.946347   0.338069  11.673  < 2e-16 ***
# SDMT         0.027404   0.007168   3.823 0.000298 ***
# Vocab       -0.017397   0.018077  -0.962 0.339428
```

```
# Abstract     0.012182   0.011585   1.051 0.296926
# ---
# Residual standard error: 0.6878 on 65 degrees of freedom
# Multiple R-squared:  0.2857,  Adjusted R-squared:  0.2528
# F-statistic: 8.668 on 3 and 65 DF,  p-value: 6.414e-05

vif(model.1)
```

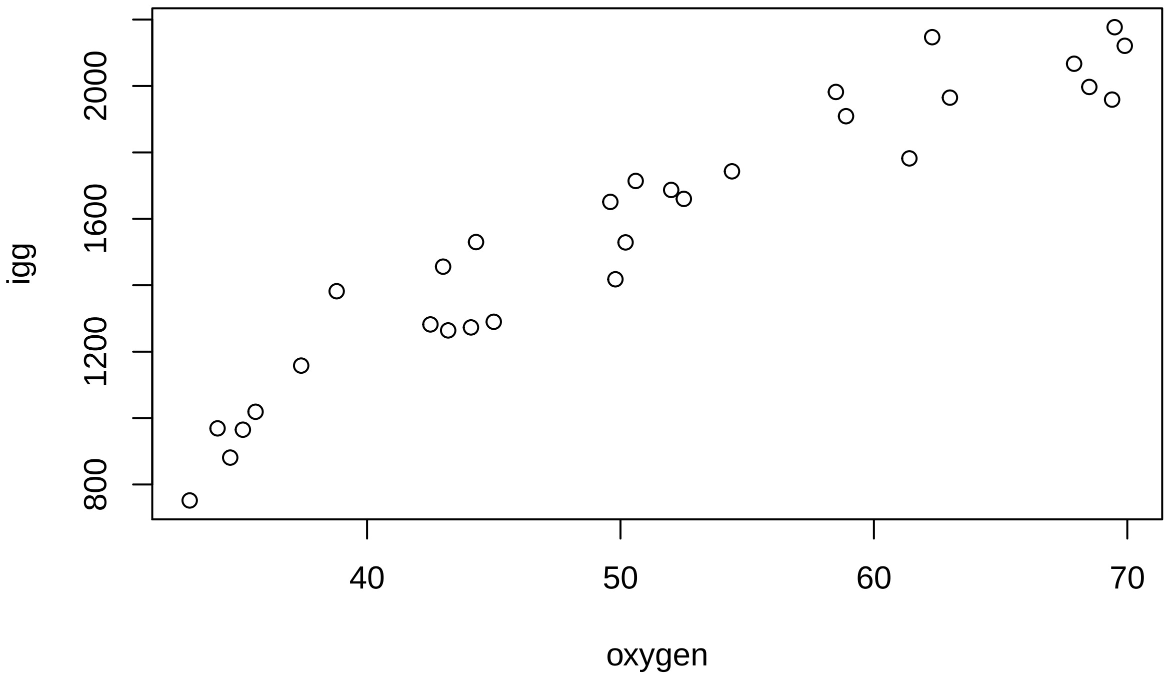
```
##     SDMT    Vocab Abstract
## 1.609662 2.093297 2.167428
```

```
#     SDMT    Vocab Abstract
# 1.609662 2.093297 2.167428

detach(allentest)
```

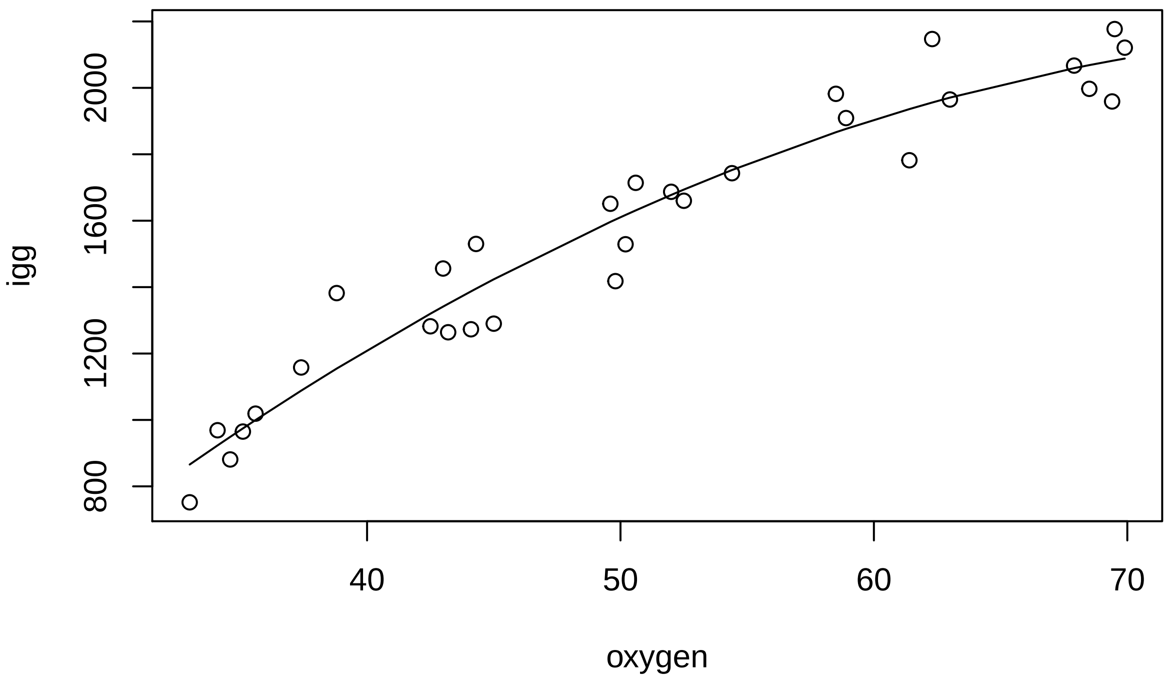
Exercise and immunity (reducing structural multicollinearity)

Load the exerimmun data. Create a scatterplot of igg vs oxygen. Calculate an oxygen-squared variable named oxygensq. Fit a quadratic regression model of igg vs oxygen + oxygensq. Add a quadratic regression line to the scatterplot. Use the vif function in the car package to calculate variance inflation factors. Create a scatterplot of oxygensq vs oxygen and calculate the correlation. Calculate a centered oxygen variable named oxcent and an oxcent-squared variable named oxcentsq. Fit a quadratic regression model of igg vs oxcent + oxcentsq. Use the vif function in the car package to calculate variance inflation factors. Create a scatterplot of igg vs oxcent with the quadratic regression line added. Fit a simple linear regression model of igg vs oxcent. Confirm the equivalence of the original quadratic and centered quadratic models by transforming the regression parameter estimates. Create a residual vs fits plot for the centered quadratic model. Create a normal probability plot of the residuals for the centered quadratic model. Predict igg for oxygen = 70 using the centered quadratic model.

```
exerimmun <- read.table("./Data/exerimmun.txt", header=T)
attach(exerimmun)

plot(oxygen, igg)
```




```r
oxygensq <- oxygen^2

model.1 <- lm(igg ~ oxygen + oxygensq)

plot(x=oxygen, y=igg,
     panel.last = lines(sort(oxygen), fitted(model.1)[order(oxygen)]))
```




```r
summary(model.1)
```

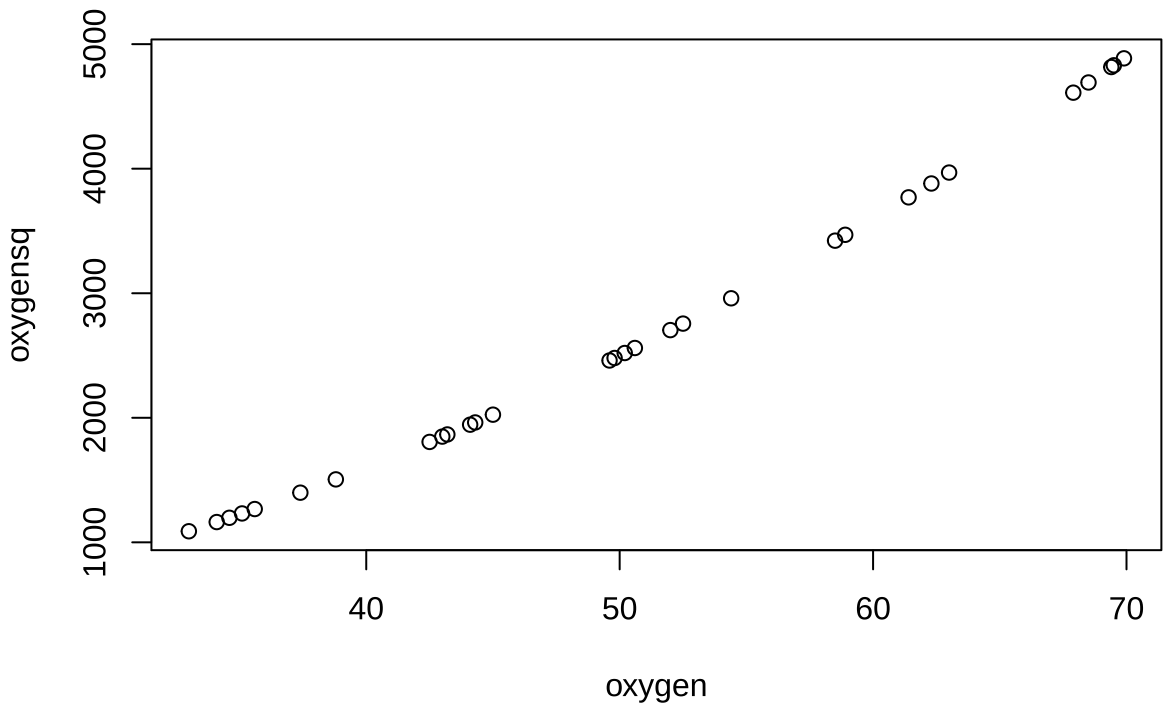
```
##
## Call:
```

```
## lm(formula = igg ~ oxygen + oxygensq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -185.375  -82.129    1.047   66.007  227.377
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1464.4042   411.4012  -3.560  0.00140 **
## oxygen         88.3071    16.4735   5.361 1.16e-05 ***
## oxygensq       -0.5362     0.1582  -3.390  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 106.4 on 27 degrees of freedom
## Multiple R-squared:  0.9377, Adjusted R-squared:  0.9331
## F-statistic: 203.2 on 2 and 27 DF,  p-value: < 2.2e-16
```

```
#               Estimate Std. Error t value Pr(>|t|)
# (Intercept) -1464.4042   411.4012  -3.560  0.00140 **
# oxygen         88.3071    16.4735   5.361 1.16e-05 ***
# oxygensq       -0.5362     0.1582  -3.390  0.00217 **
# ---
# Residual standard error: 106.4 on 27 degrees of freedom
# Multiple R-squared:  0.9377,  Adjusted R-squared:  0.9331
# F-statistic: 203.2 on 2 and 27 DF,  p-value: < 2.2e-16

vif(model.1)
```

```
##   oxygen oxygensq
## 99.94261 99.94261
```

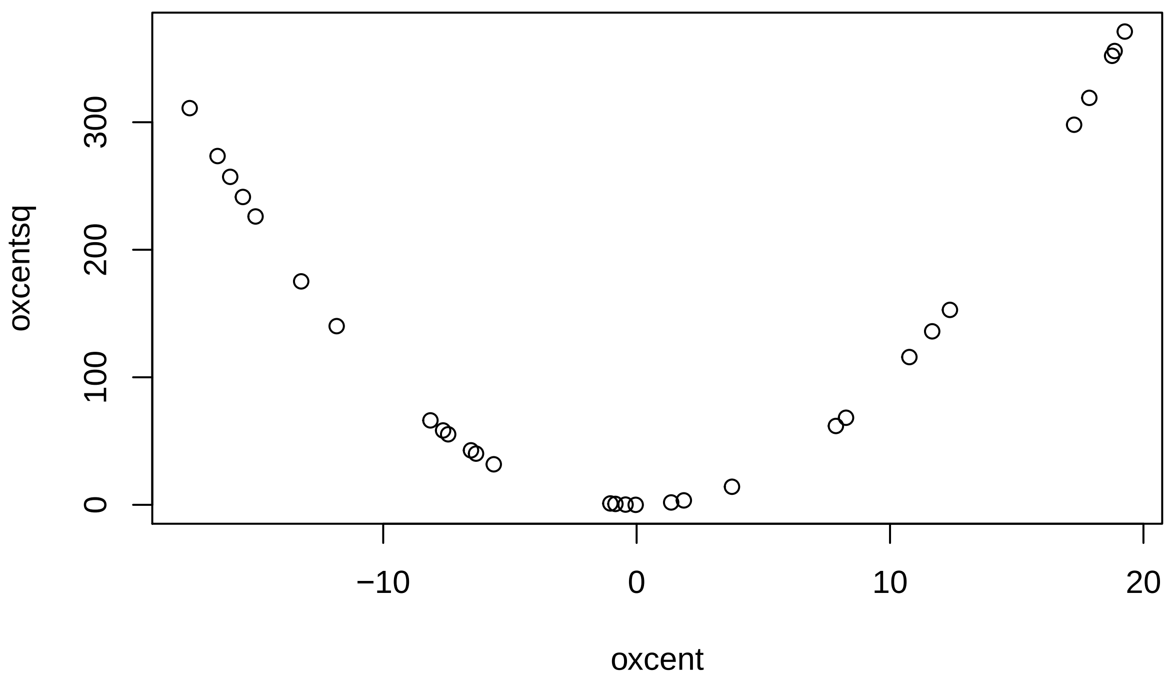
```
#   oxygen oxygensq
# 99.94261 99.94261

plot(oxygen, oxygensq)
```




```
cor(oxygen, oxygensq) # 0.9949846
```

```
## [1] 0.9949846
```

```
oxcent <- oxygen-mean(oxygen)
oxcentsq <- oxcent^2

plot(oxcent, oxcentsq)
```



```r
cor(oxcent, oxcentsq) # 0.2195179
```

```
## [1] 0.2195179
```

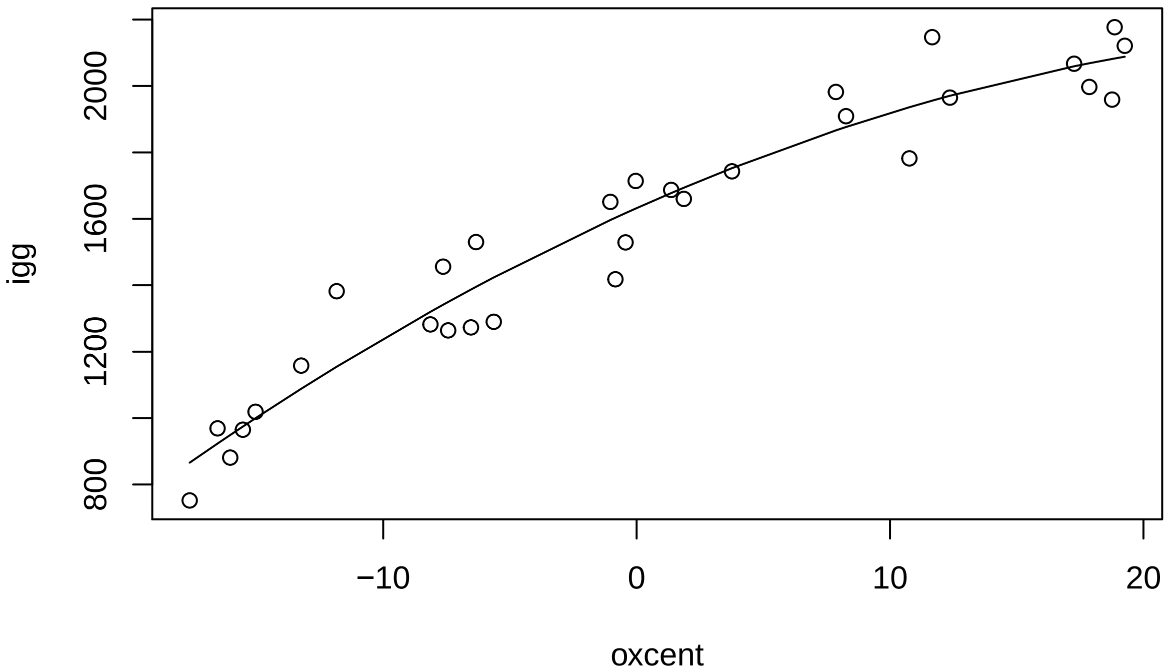
```r
model.2 <- lm(igg ~ oxcent + oxcentsq)
summary(model.2)
```

```
##
## Call:
## lm(formula = igg ~ oxcent + oxcentsq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -185.375  -82.129    1.047   66.007  227.377
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1632.1962    29.3486   55.61  < 2e-16 ***
## oxcent        33.9995     1.6890   20.13  < 2e-16 ***
## oxcentsq      -0.5362     0.1582   -3.39  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 106.4 on 27 degrees of freedom
## Multiple R-squared:  0.9377, Adjusted R-squared:  0.9331
## F-statistic: 203.2 on 2 and 27 DF,  p-value: < 2.2e-16
```

```r
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept) 1632.1962    29.3486   55.61  < 2e-16 ***
# oxcent        33.9995     1.6890   20.13  < 2e-16 ***
# oxcentsq      -0.5362     0.1582   -3.39  0.00217 **
# ---
# Residual standard error: 106.4 on 27 degrees of freedom
# Multiple R-squared:  0.9377,  Adjusted R-squared:  0.9331
# F-statistic: 203.2 on 2 and 27 DF,  p-value: < 2.2e-16

vif(model.2)
```

```
##   oxcent oxcentsq
## 1.050628 1.050628
```

```r
#   oxcent oxcentsq
# 1.050628 1.050628

plot(x=oxcent, y=igg,
     panel.last = lines(sort(oxcent), fitted(model.2)[order(oxcent)]))
```




```
model.3 <- lm(igg ~ oxcent)
summary(model.3)
```

```
##
## Call:
## lm(formula = igg ~ oxcent)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -228.16  -79.96  -11.78   83.75  211.93
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1557.633     22.782   68.37  < 2e-16 ***
## oxcent        32.743      1.932   16.95 2.97e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 124.8 on 28 degrees of freedom
## Multiple R-squared:  0.9112, Adjusted R-squared:  0.908
## F-statistic: 287.2 on 1 and 28 DF,  p-value: 2.973e-16
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) 1557.633     22.782   68.37  < 2e-16 ***
# oxcent        32.743      1.932   16.95 2.97e-16 ***

coef(model.2)[1]-coef(model.2)[2]*mean(oxygen)+coef(model.2)[3]*mean(oxygen)^2 # -1464.404
```

```
## (Intercept)
##   -1464.404
```

```
coef(model.2)[2]-2*coef(model.2)[3]*mean(oxygen) # 88.3071
```

```
##  oxcent
```

```
## 88.3071
```

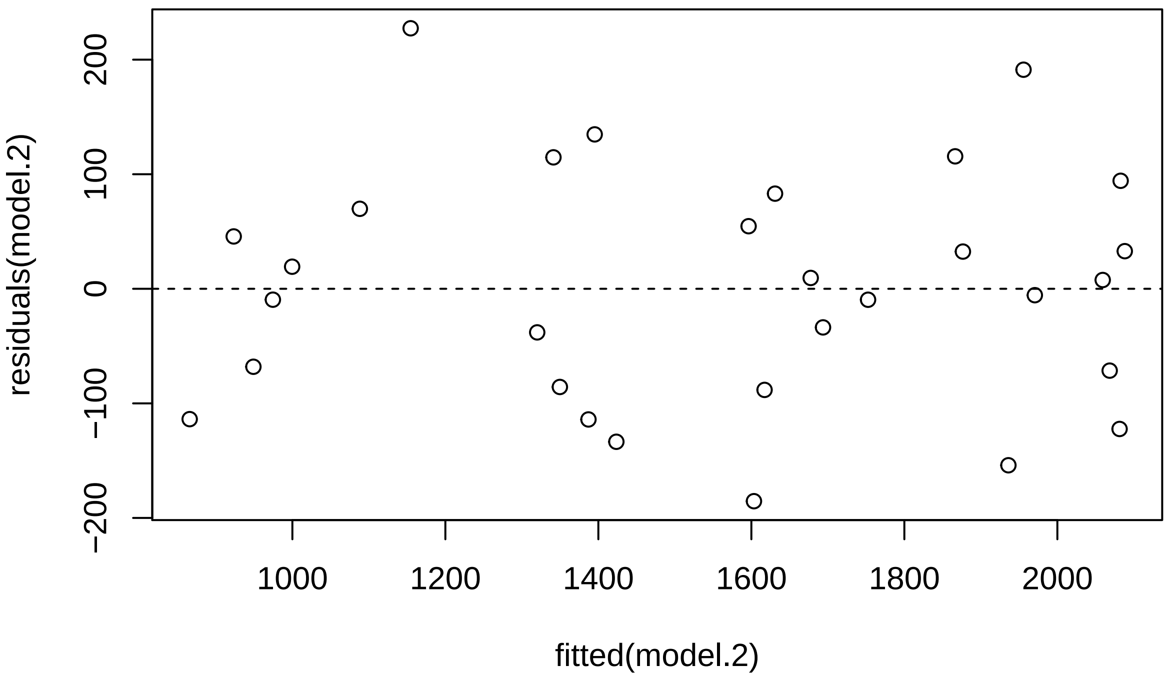
```
coef(model.2)[3] # -0.5362473
```

```
##   oxcentsq
## -0.5362473
```

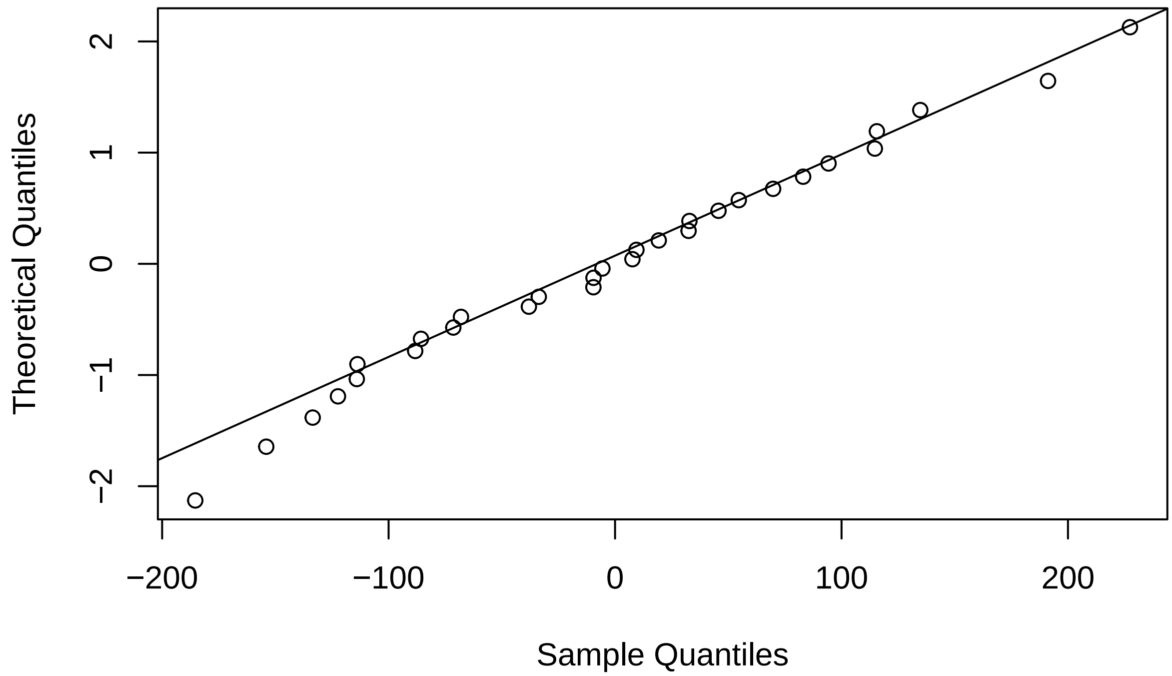
```
coef(model.1)
```

```
##   (Intercept)        oxygen      oxygensq
## -1464.4042284    88.3070970    -0.5362473
```

```
#   (Intercept)        oxygen      oxygensq
# -1464.4042284    88.3070970    -0.5362473

plot(x=fitted(model.2), y=residuals(model.2),
     panel.last = abline(h=0, lty=2))
```




```
qqnorm(residuals(model.2), main="", datax=TRUE)
qqline(residuals(model.2), datax=TRUE)
```




```
predict(model.2, interval="prediction",
        newdata=data.frame(oxcent=70-mean(oxygen), oxcentsq=(70-mean(oxygen))^2))
```

```
##        fit      lwr      upr
## 1 2089.481 1849.931 2329.031
```

```
#        fit      lwr      upr
# 1 2089.481 1849.931 2329.031

detach(exerimmun)
```