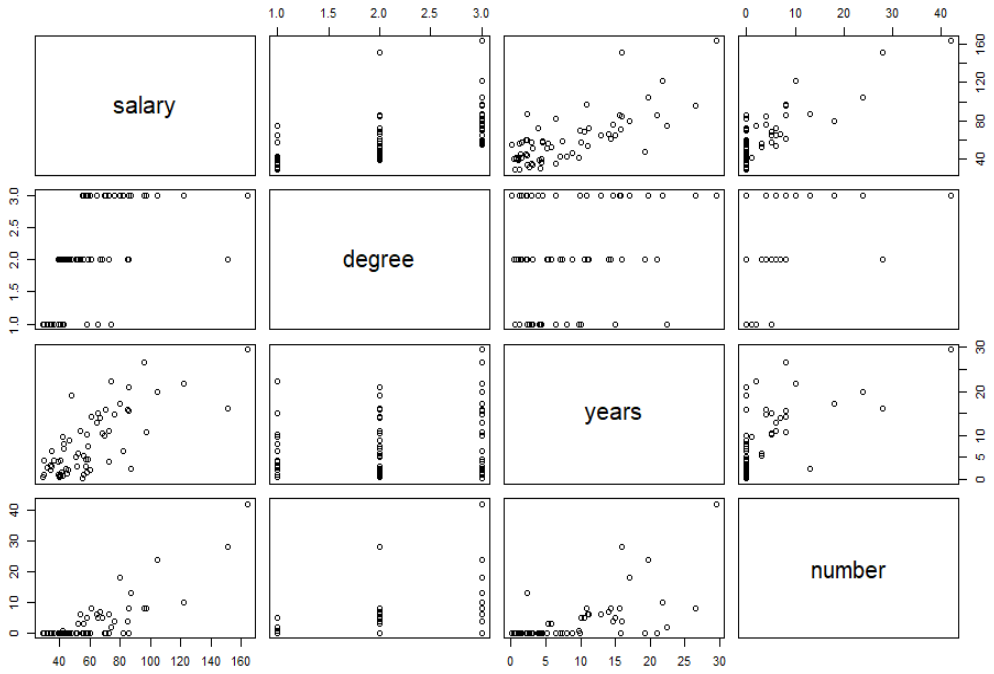
R code and output

```
salaryData=read.table("salary.txt",header=T)
attach(salaryData)
degree=factor(degree)
salaryData
```

```
##    salary degree years number
## 1    58.8      3  4.49      0
## 2    34.8      1  2.92      0
## 3   163.7      3 29.54     42
## 4    70.0      3  9.92      0
## 5    55.5      3  0.14      0
## 6    85.0      2 15.96      4
## 7    34.0      1  2.27      0
## 8    29.7      1  1.20      0
## 9    56.1      2  5.33      3
## 10   70.6      3 15.74      0
## 11   74.2      1 22.46      2
## 12   34.1      1  3.16      0
## 13   31.6      1  2.62      0
## 14   65.5      1 15.06      5
## 15   57.2      3  2.92      0
## 16   60.3      3  2.26      0
## 17   41.8      1  9.76      1
## 18   76.5      3 14.71      4
## 19  122.1      3 21.76     10
## 20   85.9      3 15.63      8
## 21   55.9      3  1.17      0
## 22   44.3      2  2.33      0
## 23   79.9      3 17.10     18
## 24   58.5      2  7.45      0
## 25   57.3      3  4.55      0
## 26   61.0      2 14.39      8
## 27   52.2      2  5.78      3
## 28   45.7      2  2.08      0
## 29   44.8      2  1.44      0
## 30   39.1      2  1.00      0
## 31   68.1      2 10.53      5
## 32   48.2      2 19.23      0
## 33   51.0      2  5.18      0
## 34   40.7      1  4.43      0
## 35   51.4      2  3.04      0
## 36   40.9      2  1.02      0
## 37   57.7      1 10.14      5
## 38   95.5      3 26.53      8
## 39   34.9      1  6.49      0
## 40   66.6      2 13.97      7
```

```
## 41   30.0      1  4.18      0
## 42   64.9      3 12.88      6
## 43  151.2      2 16.01     28
## 44   72.4      2 11.13      6
## 45   41.8      2  0.71      0
## 46   57.8      3  1.55      0
## 47   72.7      3  3.92      0
## 48   36.1      1  4.37      0
## 49   39.8      2  0.79      0
## 50   29.0      1  0.65      0
## 51   40.4      2  0.69      0
## 52   40.7      2  1.09      0
## 53   41.7      2  1.58      0
## 54   97.2      3 10.89      8
## 55   85.3      2 21.08      0
## 56   42.6      2  7.00      0
## 57   39.1      1  4.09      0
## 58   46.6      2  8.86      0
## 59   53.9      2 11.05      6
## 60   87.4      3  2.37     13
## 61   81.7      3  6.37      0
## 62   42.5      1  8.00      0
## 63   40.0      2  0.44      0
## 64   60.5      3  2.10      0
## 65  104.8      3 19.81     24
```

```
pairs(salaryData)
```



```r
round(cor(salaryData),3)
```

```
##        salary degree years number
## salary  1.000  0.554 0.742  0.830
## degree  0.554  1.000 0.217  0.297
## years   0.742  0.217 1.000  0.635
## number  0.830  0.297 0.635  1.000
```

```r
salary.fit.second=lm(salary~degree+years+number+degree:years+degree:number+ye
ars:number)
summary(salary.fit.second)
```

```
##
## Call:
## lm(formula = salary ~ degree + years + number + degree:years +
##     degree:number + years:number)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.089  -3.957  -0.424   2.248  25.737
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    28.10465    3.53708   7.946 1.07e-10 ***
## degree2        12.01821    4.39806   2.733  0.00843 **
## degree3        31.04306    4.81458   6.448 3.02e-08 ***
## years           1.71939    0.54403   3.160  0.00256 **
## number          2.00896    1.82592   1.100  0.27602
## degree2:years  -0.54176    0.62974  -0.860  0.39336
## degree3:years  -0.68243    0.63395  -1.076  0.28641
## degree2:number  0.53029    1.84800   0.287  0.77523
## degree3:number -1.13100    1.85374  -0.610  0.54430
## years:number    0.02698    0.02264   1.192  0.23845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.046 on 55 degrees of freedom
## Multiple R-squared:  0.9004, Adjusted R-squared:  0.8841
## F-statistic: 55.25 on 9 and 55 DF,  p-value: < 2.2e-16
```

```r
salary.fit.first=lm(salary~degree+years+number)
anova(salary.fit.first,salary.fit.second)
```

```
## Analysis of Variance Table
##
## Model 1: salary ~ degree + years + number
## Model 2: salary ~ degree + years + number + degree:years + degree:number +
##     years:number
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     60 6175.1
## 2     55 4500.2  5    1674.9 4.0941 0.003128 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
round(cor(model.matrix(salary.fit.second)),3)
```

```
## Warning in cor(model.matrix(salary.fit.second)): the standard deviation is
zero
```

```
##                (Intercept) degree2 degree3  years number degree2:years
## (Intercept)              1      NA      NA     NA     NA            NA
## degree2                 NA   1.000  -0.603 -0.112 -0.098         0.650
## degree3                 NA  -0.603   1.000  0.232  0.289        -0.392
## years                   NA  -0.112   0.232  1.000  0.635         0.354
## number                  NA  -0.098   0.289  0.635  1.000         0.116
## degree2:years           NA   0.650  -0.392  0.354  0.116         1.000
## degree3:years           NA  -0.421   0.698  0.659  0.618        -0.274
## degree2:number          NA   0.334  -0.201  0.216  0.432         0.560
## degree3:number          NA  -0.275   0.456  0.551  0.850        -0.179
## years:number            NA  -0.125   0.277  0.626  0.940         0.032
##                degree3:years degree2:number degree3:number years:number
## (Intercept)               NA             NA             NA           NA
## degree2               -0.421          0.334         -0.275       -0.125
## degree3                0.698         -0.201          0.456        0.277
## years                  0.659          0.216          0.551        0.626
## number                 0.618          0.432          0.850        0.940
## degree2:years         -0.274          0.560         -0.179        0.032
## degree3:years          1.000         -0.140          0.777        0.664
## degree2:number        -0.140          1.000         -0.092        0.259
## degree3:number         0.777         -0.092          1.000        0.887
## years:number           0.664          0.259          0.887        1.000
```

```
vif <- function(object, ...)
UseMethod("vif")
```

```
vif.default <- function(object, ...)
stop("No default method for vif. Sorry.")
```

```
vif.lm <- function(object, ...) {
  V <- summary(object)$cov.unscaled
  Vi <- crossprod(model.matrix(object))
        nam <- names(coef(object))
  if(k <- match("(Intercept)", nam, nomatch = F)) {
                v1 <- diag(V)[-k]
                v2 <- (diag(Vi)[-k] - Vi[k, -k]^2/Vi[k,k])
                nam <- nam[-k]
        } else {
                v1 <- diag(V)
                v2 <- diag(Vi)
                warning("No intercept term detected. Results may surprise.")
        }
        structure(v1*v2, names = nam)
```

```
}
vif(salary.fit.second)
```

```
##        degree2        degree3          years         number  degree2:years
##       3.731533       4.123107      12.148977     142.126679       8.889176
##  degree3:years degree2:number degree3:number   years:number
##      15.531866      39.760514     120.581404      12.502437
```

```
step(lm(salary ~ 1), data=salaryData,
     scope = ~ degree+years+number+degree:years+degree:number+years:number)
```

```
## Start:  AIC=427.37
## salary ~ 1
##
##          Df Sum of Sq   RSS    AIC
## + number  1     31148 14037 353.38
## + years   1     24852 20332 377.46
## + degree  2     14115 31069 407.02
## <none>                45184 427.37
##
## Step:  AIC=353.38
## salary ~ number
##
##          Df Sum of Sq   RSS    AIC
## + degree  2    4708.2  9328 330.82
## + years   1    3472.3 10564 336.90
## <none>                14037 353.38
## - number  1   31147.5 45184 427.37
##
## Step:  AIC=330.82
## salary ~ number + degree
##
##                 Df Sum of Sq     RSS    AIC
## + years          1    3153.2  6175.1 306.00
## + degree:number  2    1766.1  7562.3 321.17
## <none>                        9328.4 330.82
## - degree         2    4708.2 14036.5 353.38
## - number         1   21740.7 31069.0 407.02
##
## Step:  AIC=306
## salary ~ number + degree + years
##
##                 Df Sum of Sq     RSS    AIC
## + degree:number  2    1471.3  4703.9 292.31
## + degree:years   2     697.6  5477.6 302.21
## <none>                        6175.1 306.00
## + years:number   1     132.9  6042.2 306.59
## - years          1    3153.2  9328.4 330.82
## - degree         2    4389.1 10564.2 336.90
## - number         1    6814.3 12989.4 352.34
```

```
##
## Step:  AIC=292.31
## salary ~ number + degree + years + number:degree
##
##                 Df Sum of Sq    RSS    AIC
## <none>                       4703.9 292.31
## + years:number   1    108.03 4595.8 292.80
## + degree:years   2     87.44 4616.4 295.09
## - number:degree  2   1471.26 6175.1 306.00
## - years          1   2858.39 7562.3 321.17

##
## Call:
## lm(formula = salary ~ number + degree + years + number:degree)
##
## Coefficients:
##    (Intercept)          number         degree2         degree3
years
##         30.337           3.431           9.323          26.641
1.235
## number:degree2  number:degree3
##         -0.504          -1.990
```

```
salary.fit=lm(salary~degree+years+number+degree:number)
summary(salary.fit)
```

```
##
## Call:
## lm(formula = salary ~ degree + years + number + degree:number)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.1324  -3.4510  -0.1389   3.3618  23.8415
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      30.337      2.679  11.323 2.54e-16 ***
## degree2           9.323      3.157   2.953  0.00454 **
## degree3          26.641      3.401   7.832 1.17e-10 ***
## years             1.235      0.208   5.937 1.73e-07 ***
## number            3.431      1.427   2.405  0.01939 *
## degree2:number   -0.504      1.429  -0.353  0.72554
## degree3:number   -1.990      1.404  -1.417  0.16171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.006 on 58 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8851
## F-statistic: 83.19 on 6 and 58 DF,  p-value: < 2.2e-16
```

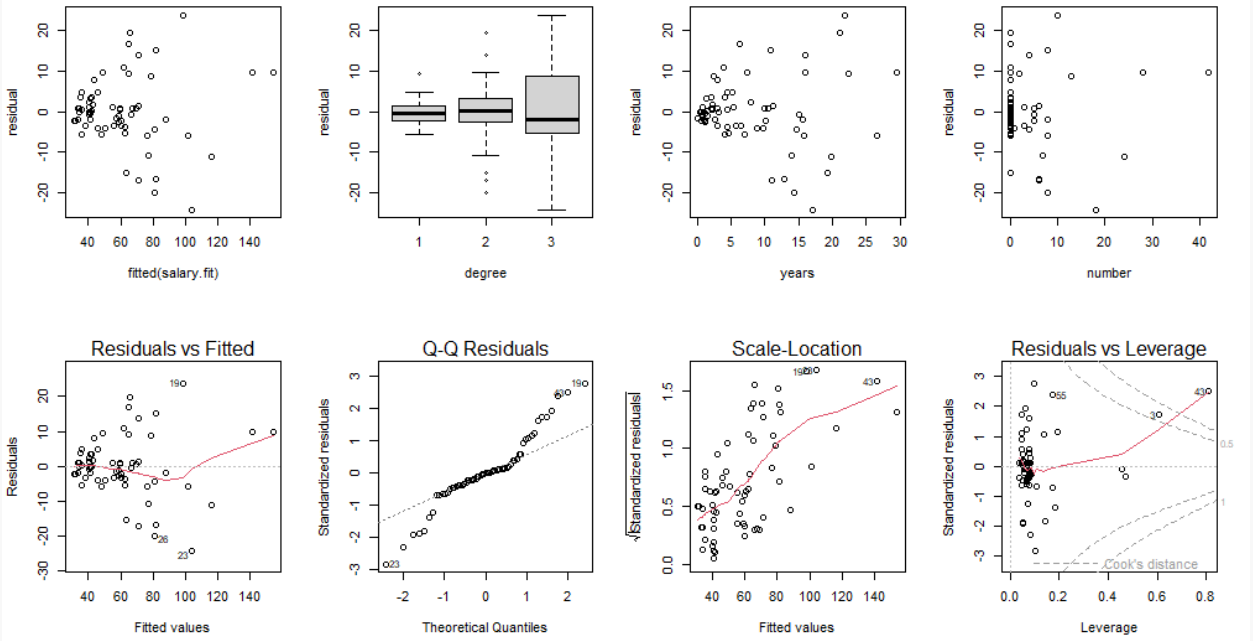
```r
round(vcov(salary.fit),3)

#estimated covariance matrix of estimates of regression coefficients
```

```
##                 (Intercept) degree2 degree3  years number degree2:number
## (Intercept)           7.178  -6.067  -5.914 -0.198 -1.047           1.153
## degree2              -6.067   9.969   6.356 -0.045  1.583          -1.807
## degree3              -5.914   6.356  11.569 -0.079  1.657          -1.614
## years                -0.198  -0.045  -0.079  0.043 -0.095           0.072
## number               -1.047   1.583   1.657 -0.095  2.035          -1.984
## degree2:number        1.153  -1.807  -1.614  0.072 -1.984           2.041
## degree3:number        1.167  -1.555  -1.837  0.069 -1.977           1.940
##                 degree3:number
## (Intercept)              1.167
## degree2                 -1.555
## degree3                 -1.837
## years                    0.069
## number                  -1.977
## degree2:number           1.940
## degree3:number           1.971
```

```r
residual=resid(salary.fit)

par(mfrow=c(2,4))
plot(fitted(salary.fit),residual);
plot(degree,residual,xlab="degree",ylab="residual");
plot(years,residual,xlab="years",ylab="residual");
plot(number,residual,xlab="number",ylab="residual")
plot(salary.fit)
```



```r
library(lmtest); bptest(salary~degree+years+number+degree:number)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  salary ~ degree + years + number + degree:number
## BP = 22.136, df = 6, p-value = 0.001144
```

```r
shapiro.test(resid(salary.fit))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(salary.fit)
## W = 0.94558, p-value = 0.006423
```

```r
round(sort(rstandard(salary.fit)),2)
```

```
##    23    26    59    42    32    65    40    38    10    56    41    25    18
## -2.83 -2.30 -1.94 -1.89 -1.82 -1.38 -1.24 -0.70 -0.68 -0.65 -0.64 -0.61 -0.51
##    17    58     1    39    15    27    37    21    50     8    13    20    30
## -0.46 -0.46 -0.43 -0.40 -0.39 -0.38 -0.35 -0.29 -0.25 -0.25 -0.23 -0.22 -0.21
##     5    46    49    14    52    63    12    51    36    53    48    16     4
## -0.19 -0.13 -0.10 -0.09 -0.03 -0.02 -0.02 -0.01  0.00  0.01  0.04  0.06  0.09
##    31     2     7    64     9    45    44    22    62    29    28    57    33
##  0.09  0.10  0.10  0.11  0.12  0.14  0.16  0.20  0.26  0.38  0.40  0.43  0.56
##    34    35    60    24    11    47     6     3    54    61    55    43    19
##  0.57  0.91  1.05  1.10  1.14  1.25  1.61  1.72  1.73  1.93  2.39  2.51  2.79
```

```r
round(sort(rstudent(salary.fit)),2)
```

```
##    23    26    59    42    32    65    40    38    10    56    41    25    18
## -3.02 -2.39 -1.98 -1.94 -1.86 -1.39 -1.25 -0.70 -0.68 -0.65 -0.63 -0.60 -0.50
##    17    58     1    39    15    27    37    21    50     8    13    20    30
## -0.46 -0.45 -0.42 -0.40 -0.39 -0.38 -0.35 -0.29 -0.25 -0.24 -0.23 -0.22 -0.20
##     5    46    49    14    52    63    12    51    36    53    48    16     4
## -0.19 -0.12 -0.09 -0.09 -0.03 -0.02 -0.02 -0.01  0.00  0.01  0.04  0.06  0.09
##    31     2     7    64     9    45    44    22    62    29    28    57    33
##  0.09  0.10  0.10  0.11  0.12  0.14  0.16  0.20  0.26  0.38  0.39  0.43  0.56
##    34    35    60    24    11    47     6     3    54    61    55    43    19
##  0.56  0.91  1.05  1.10  1.15  1.26  1.63  1.75  1.77  1.98  2.50  2.63  2.97
```

```r
inf=(influence.measures(salary.fit)[[1]])
round(inf[,8:11],4)
```

```
##      dffit  cov.r cook.d    hat
## 1  -0.1123 1.1822 0.0018 0.0655
## 2   0.0288 1.2246 0.0001 0.0788
## 3   2.1681 1.9865 0.6485 0.6057
## 4   0.0242 1.2135 0.0001 0.0702
## 5  -0.0576 1.2281 0.0005 0.0844
## 6   0.4651 0.8877 0.0300 0.0752
## 7   0.0292 1.2264 0.0001 0.0802
## 8  -0.0735 1.2233 0.0008 0.0834
```

```
## 9   0.0244 1.1734 0.0001 0.0391
## 10 -0.2397 1.1992 0.0083 0.1101
## 11  0.5577 1.1902 0.0442 0.1912
## 12 -0.0046 1.2255 0.0000 0.0784
## 13 -0.0665 1.2192 0.0006 0.0794
## 14 -0.0808 2.0790 0.0009 0.4573
## 15 -0.1060 1.1924 0.0016 0.0700
## 16  0.0170 1.2174 0.0000 0.0726
## 17 -0.1241 1.1810 0.0022 0.0680
## 18 -0.1343 1.1730 0.0026 0.0665
## 19  0.9735 0.4575 0.1193 0.0972
## 20 -0.0530 1.1906 0.0004 0.0568
## 21 -0.0842 1.2126 0.0010 0.0781
## 22  0.0461 1.1841 0.0003 0.0507
## 23 -1.0335 0.4451 0.1338 0.1046
## 24  0.2433 1.0233 0.0084 0.0468
## 25 -0.1599 1.1555 0.0037 0.0653
## 26 -0.7165 0.6300 0.0678 0.0823
## 27 -0.0758 1.1539 0.0008 0.0383
## 28  0.0917 1.1686 0.0012 0.0516
## 29  0.0913 1.1732 0.0012 0.0543
## 30 -0.0497 1.1908 0.0004 0.0563
## 31  0.0200 1.1834 0.0001 0.0465
## 32 -0.7638 0.8721 0.0799 0.1438
## 33  0.1213 1.1385 0.0021 0.0451
## 34  0.1628 1.1777 0.0038 0.0774
## 35  0.2050 1.0736 0.0060 0.0485
## 36 -0.0005 1.1968 0.0000 0.0562
## 37 -0.3327 2.1109 0.0161 0.4730
## 38 -0.3181 1.2832 0.0146 0.1710
## 39 -0.1164 1.2033 0.0020 0.0793
## 40 -0.3464 1.0063 0.0170 0.0713
## 41 -0.1832 1.1658 0.0048 0.0774
## 42 -0.4440 0.7602 0.0269 0.0498
## 43  5.4498 2.6783 3.8497 0.8110
## 44  0.0386 1.1895 0.0002 0.0536
## 45  0.0355 1.1958 0.0002 0.0578
## 46 -0.0359 1.2201 0.0002 0.0760
## 47  0.3363 0.9994 0.0160 0.0668
## 48  0.0122 1.2239 0.0000 0.0774
## 49 -0.0234 1.1969 0.0001 0.0574
## 50 -0.0754 1.2260 0.0008 0.0856
## 51 -0.0032 1.1989 0.0000 0.0579
## 52 -0.0084 1.1962 0.0000 0.0559
## 53  0.0024 1.1935 0.0000 0.0537
## 54  0.3904 0.8162 0.0210 0.0466
## 55  1.1414 0.6590 0.1707 0.1725
## 56 -0.1417 1.1251 0.0029 0.0460
## 57  0.1235 1.1974 0.0022 0.0775
## 58 -0.1045 1.1601 0.0016 0.0506
```

```
## 59 -0.4713 0.7473 0.0302 0.0534
## 60  0.4246 1.1490 0.0257 0.1404
## 61  0.5163 0.7563 0.0363 0.0635
## 62  0.0794 1.2221 0.0009 0.0836
## 63 -0.0058 1.2005 0.0000 0.0593
## 64  0.0299 1.2172 0.0001 0.0734
## 65 -0.6561 1.0940 0.0605 0.1823
```

```r
par(mfrow=c(2,3))
plot(resid(salary.fit))
plot(rstandard(salary.fit))
plot(rstudent(salary.fit))
plot(inf[,11],ylab="leverage")
plot(inf[,10],ylab="cook's distance")
plot(inf[,8],ylab="dffits")
```