```
> dim(swiss)
[1] 47  6
> #stepwise regression model
> step.model <- stepAIC(full.model, direction = 'both', trace = TRUE)
Start:  AIC=190.69
Fertility ~ Agriculture + Examination + Education + Catholic +
    Infant.Mortality

                   Df Sum of Sq    RSS    AIC
- Examination       1     53.03 2158.1 189.86
<none>                         2105.0 190.69
- Agriculture       1    307.72 2412.8 195.10
- Infant.Mortality  1    408.75 2513.8 197.03
- Catholic          1    447.71 2552.8 197.75
- Education         1   1162.56 3267.6 209.36

Step:  AIC=189.86
Fertility ~ Agriculture + Education + Catholic + Infant.Mortality

                   Df Sum of Sq    RSS    AIC
<none>                         2158.1 189.86
+ Examination       1     53.03 2105.0 190.69
- Agriculture       1    264.18 2422.2 193.29
- Infant.Mortality  1    409.81 2567.9 196.03
- Catholic          1    956.57 3114.6 205.10
- Education         1   2249.97 4408.0 221.43
```

```
> summary(step.model)

Call:
lm(formula = Fertility ~ Agriculture + Education + Catholic +
    Infant.Mortality, data = swiss)

Residuals:
     Min       1Q   Median       3Q      Max
-14.6765  -6.0522   0.7514   3.1664  16.1422

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       62.10131    9.60489   6.466 8.49e-08 ***
Agriculture       -0.15462    0.06819  -2.267  0.02857 *
Education         -0.98026    0.14814  -6.617 5.14e-08 ***
Catholic           0.12467    0.02889   4.315 9.50e-05 ***
Infant.Mortality   1.07844    0.38187   2.824  0.00722 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> summary(models)
Subset selection object
Call: regsubsets.formula(Fertility ~ ., data = swiss, nvmax = 5, method = "seqrep")
5 Variables  (and intercept)
                 Forced in Forced out
Agriculture          FALSE      FALSE
Examination          FALSE      FALSE
Education            FALSE      FALSE
Catholic             FALSE      FALSE
Infant.Mortality     FALSE      FALSE
1 subsets of each size up to 5
Selection Algorithm: 'sequential replacement'
         Agriculture Examination Education Catholic Infant.Mortality
1  ( 1 ) " "         " "         "*"       " "      " "
```

```
> step.model
Linear Regression with Backwards Selection

47 samples
 5 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 42, 42, 42, 42, 42, 44, ...
Resampling results across tuning parameters:

  nvmax  RMSE      Rsquared   MAE
  1      9.227550  0.4366973  7.857870
  2      8.710164  0.4781831  7.513855
  3      8.085123  0.6207448  7.038078
  4      7.376207  0.6699474  6.282399
  5      7.424916  0.6922072  6.312180

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was nvmax = 4.
> step.model$results
  nvmax     RMSE  Rsquared      MAE   RMSESD RsquaredSD    MAESD
1     1 9.227550 0.4366973 7.857870 1.777707  0.3218266 1.939873
2     2 8.710164 0.4781831 7.513855 1.660346  0.2653541 1.644553
3     3 8.085123 0.6207448 7.038078 2.616606  0.2208241 2.785060
```

```
> step

Call:
lm(formula = Fertility ~ Agriculture + Education + Catholic +
    Infant.Mortality, data = swiss)

Coefficients:
     (Intercept)       Agriculture         Education          Catholic  Infant.Mortality
         62.1013           -0.1546           -0.9803            0.1247            1.0784
```

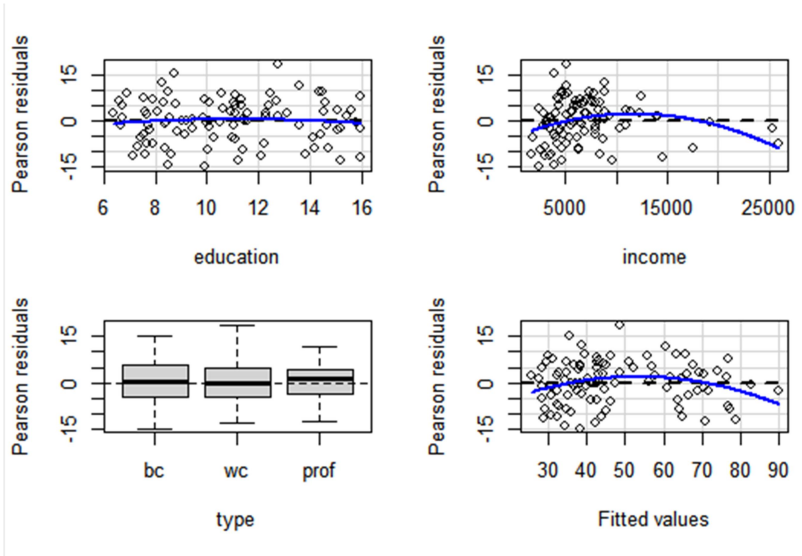
LAB 8

```
> step.model$results
  parameter      RMSE Rsquared      MAE    RMSESD RsquaredSD    MAESD
1      none 7.539194 0.665358 6.404781 2.826861  0.2707927 2.779205
> #final model coefficients
> step.model$finalModel

Call:
lm(formula = .outcome ~ Agriculture + Education + Catholic +
    Infant.Mortality, data = dat)

Coefficients:
     (Intercept)       Agriculture         Education          Catholic  Infant.Mortality
         62.1013           -0.1546           -0.9803            0.1247            1.0784
```

```
> summary(reg3)

Call:
lm(formula = prestige ~ type * (education + log2(income)), data = Pr

Residuals:
    Min      1Q  Median      3Q     Max
-13.970  -4.124   1.206   3.829  18.059

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -120.0459    20.1576  -5.955 5.07e-08 ***
typewc                  30.2412    37.9788   0.796  0.42800
typeprof                85.1601    31.1810   2.731  0.00761 **
education                2.3357     0.9277   2.518  0.01360 *
log2(income)            11.0782     1.8063   6.133 2.32e-08 ***
typewc:education         3.6400     1.7589   2.069  0.04140 *
typeprof:education       0.6974     1.2895   0.541  0.58998
typewc:log2(income)     -5.6530     3.0519  -1.852  0.06730 .
typeprof:log2(income)   -6.5356     2.6167  -2.498  0.01434 *
---
```

```
> summary(reg2)

Call:
lm(formula = prestige ~ education + log2(income) + type, data = Prestige)

Residuals:
    Min      1Q  Median      3Q     Max
-13.511  -3.746   1.011   4.356  18.438

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -81.2019    13.7431  -5.909 5.63e-08 ***
education      3.2845     0.6081   5.401 5.06e-07 ***
log2(income)   7.2694     1.1900   6.109 2.31e-08 ***
typeprof       6.7509     3.6185   1.866   0.0652 .
typewc        -1.4394     2.3780  -0.605   0.5465
```

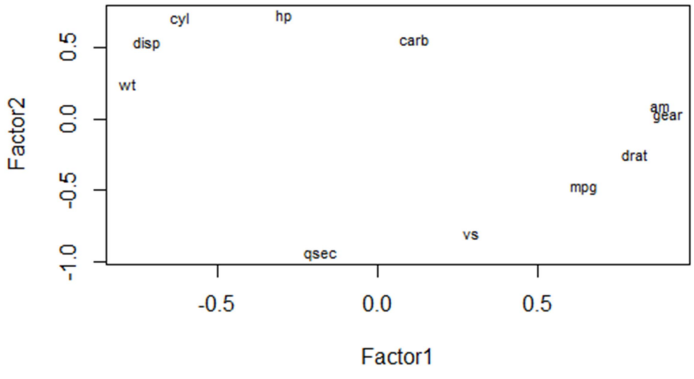
```
> summary(reg1)

Call:
lm(formula = prestige ~ education + log2(income) + women, data = Prestige)

Residuals:
    Min      1Q  Median      3Q     Max
-17.364  -4.429  -0.101   4.316  19.179

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -110.9658    14.8429  -7.476 3.27e-11 ***
education      3.7305     0.3544  10.527  < 2e-16 ***
log2(income)   9.3147     1.3265   7.022 2.90e-10 ***
women          0.0469     0.0299   1.568     0.12
```

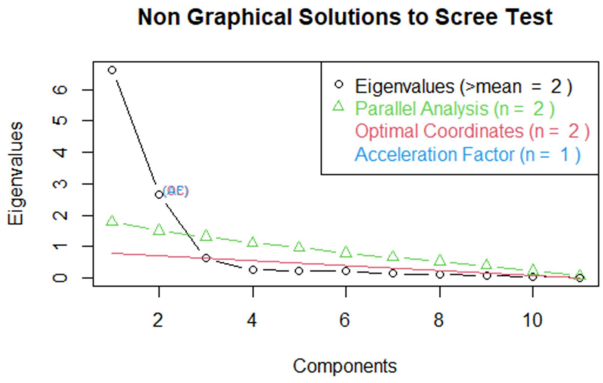


```
> fit # print results
Principal Components Analysis
Call: principal(r = mtcars, nfactors = 5, rotate = "v
Standardized loadings (pattern matrix) based upon corr
       RC2   RC5   RC3   RC1   RC4   h2    u2 com
mpg   0.68 -0.43 -0.48 -0.13  0.09 0.90 0.096 2.7
cyl  -0.60  0.68  0.27  0.18 -0.15 0.95 0.051 2.6
disp -0.72  0.53  0.22  0.33 -0.01 0.96 0.043 2.5
hp   -0.30  0.62  0.54  0.45 -0.09 0.96 0.037 3.4
drat  0.76 -0.22 -0.04 -0.04  0.60 0.99 0.014 2.1
wt   -0.82  0.25  0.43  0.12  0.02 0.94 0.063 1.8
qsec -0.21 -0.88 -0.27 -0.21  0.08 0.94 0.056 1.5
vs    0.29 -0.90 -0.24  0.14  0.02 0.97 0.029 1.4
am    0.92  0.12 -0.04 -0.01  0.11 0.88 0.118 1.1
gear  0.90 -0.06  0.34  0.04  0.05 0.92 0.078 1.3
carb  0.05  0.44  0.88  0.04 -0.01 0.96 0.035 1.5

                       RC2  RC5  RC3  RC1  RC4
SS loadings           4.47 3.21 1.84 0.44 0.42
Proportion Var        0.41 0.29 0.17 0.04 0.04
Cumulative Var        0.41 0.70 0.87 0.91 0.94
Proportion Explained  0.43 0.31 0.18 0.04 0.04
```

```
> fit <- factanal(mtcars, 3, rotation='varimax')
> print(fit, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mtcars, factors = 3, rotation = "varimax")

Uniquenesses:
 mpg  cyl disp   hp drat   wt qsec   vs   am gear carb
0.13 0.06 0.09 0.13 0.29 0.06 0.05 0.22 0.21 0.12 0.16

Loadings:
     Factor1 Factor2 Factor3
mpg   0.64   -0.48   -0.47
disp -0.72    0.54    0.32
drat  0.80
wt   -0.78            0.52
am    0.88
gear  0.91
cyl  -0.62    0.70
hp            0.72    0.51
qsec         -0.95
vs           -0.80
carb          0.56    0.72
```

```
> fit # print results

Call:
factanal(x = mtcars, factors = 3, rotation = "varimax")

Uniquenesses:
  mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
0.135 0.055 0.090 0.127 0.290 0.060 0.051 0.223 0.208 0.125 0.158

Loadings:
     Factor1 Factor2 Factor3
mpg   0.643  -0.478  -0.473
cyl  -0.618   0.703   0.261
disp -0.719   0.537   0.323
hp   -0.291   0.725   0.513
drat  0.804  -0.241
wt   -0.778   0.248   0.524
qsec -0.177  -0.946  -0.151
vs    0.295  -0.805  -0.204
am    0.880
gear  0.908          0.224
carb  0.114   0.559   0.719

                Factor1 Factor2 Factor3
SS loadings       4.380   3.520   1.578
Proportion Var    0.398   0.320   0.143
Cumulative Var    0.398   0.718   0.862

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 30.53 on 25 degrees of freedom.
The p-value is 0.205
> library(nFactors)
```





LAB 10

```
> summary(model)

Call:
lm(formula = sales ~ ., data = train.data)

Residuals:
     Min       1Q   Median       3Q      Max
-10.7142  -0.9939   0.3684   1.4494   3.3619

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.594142   0.420815   8.541 1.05e-14 ***
youtube     0.044636   0.001552  28.758  < 2e-16 ***
facebook    0.188823   0.009529  19.816  < 2e-16 ***
newspaper   0.002840   0.006442   0.441     0.66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 2.043 on 158 degrees of freedom
Multiple R-squared:  0.8955,    Adjusted R-squared:  0.8935
F-statistic: 451.2 on 3 and 158 DF,  p-value: < 2.2e-16
```

```
> predictions <- model %>% predict(test.data)
> RMSE(predictions, test.data$sales)
[1] 1.965508
> R2(predictions, test.data$sales)
[1] 0.9049049
```

LAB 11

```
> head(ageandheight)
  age height no_siblings
1  18   76.1           1
2  19   77.0           2
3  20   78.1           4
4  21   78.2           5
5  22   78.8           3
6  23   79.7           2
```

```
> summary(lmHeight)

Call:
lm(formula = height ~ age, data = ageandheight)

Residuals:
      Min       1Q   Median       3Q      Max
-0.27238 -0.24248 -0.02762  0.16014  0.47238

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.9283     0.5084  127.71  < 2e-16 ***
age           0.6350     0.0214   29.66 4.43e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.256 on 10 degrees of freedom
```

```
> residuals(t_line)
            1             2             3             4             5
 5.707281e-18  4.842801e-15 -6.171472e-15 -2.208288e-15  3.531252e-15
> residuals(t_line)^2
            1             2             3             4             5
3.257306e-35 2.345272e-29 3.808706e-29 4.876537e-30 1.246974e-29
> sum(residuals(t_line)^2)
[1] 7.888609e-29
```

LAB 12

```
> res.ftest

        F test to compare two variances

data:  y and x
F = 3.1088, num df = 4, denom df = 4, p-value = 0.2977
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.3236806 29.8585625
sample estimates:
ratio of variances
            3.1088
```