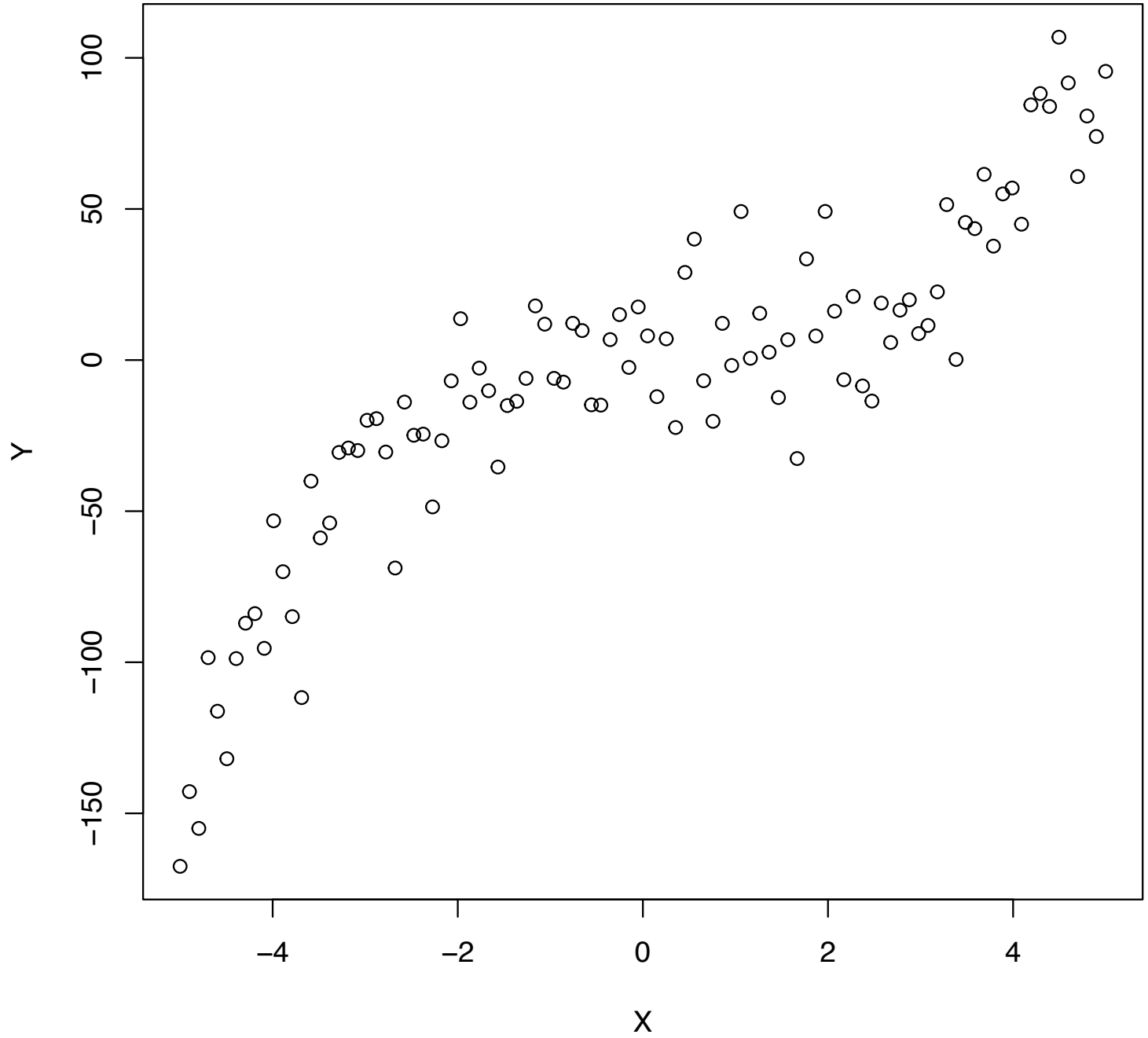
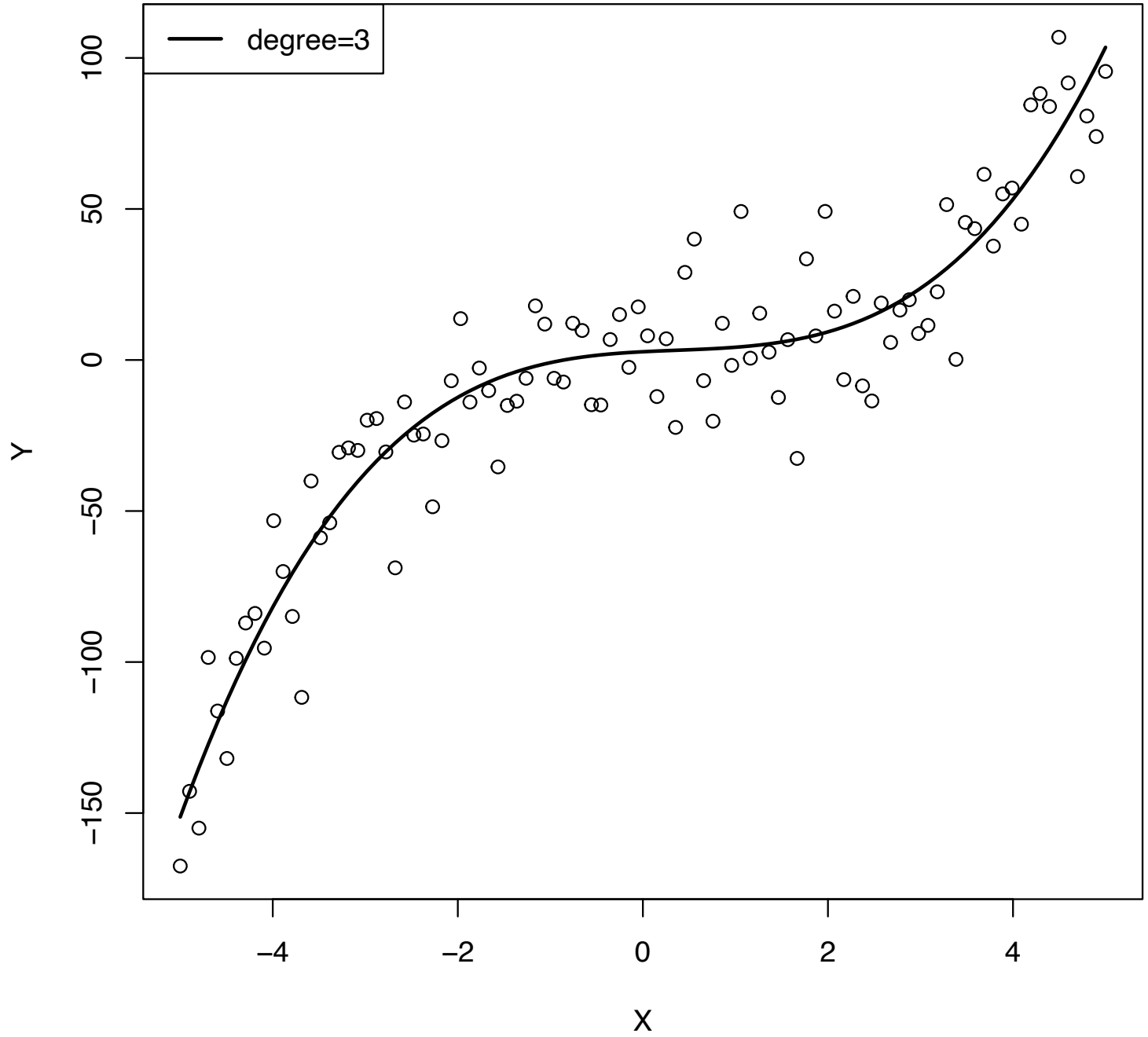
```
> df
    life  rpm brand
1  18.73  610     A
2  14.52  950     A
3  17.43  720     A
4  14.54  840     A
5  13.44  980     A
6  24.39  530     A
7  13.34  680     A
8  22.71  540     A
9  12.68  890     A
10 19.32  730     A
11 30.16  670     B
12 27.09  770     B
13 25.40  880     B
14 26.05 1000     B
15 33.49  760     B
16 35.62  590     B
17 26.07  910     B
18 36.78  650     B
19 34.95  810     B
20 43.67  500     B
```

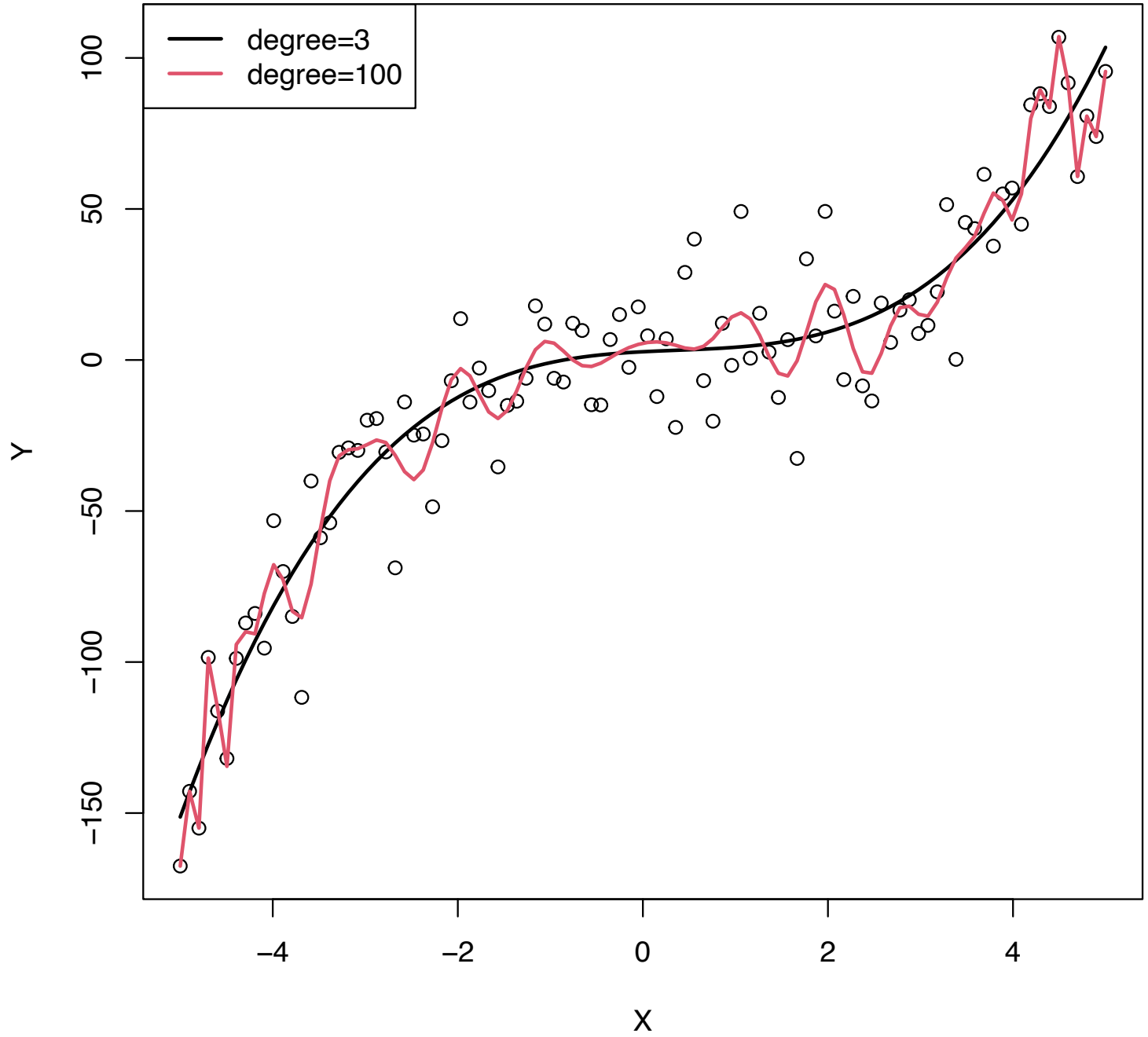
```
> df
    life  rpm brand
1  18.73  610     A
2  14.52  950     A
3  17.43  720     A
4  14.54  840     A
5  13.44  980     A
6  24.39  530     A
7  13.34  680     A
8  22.71  540     A
9  12.68  890     A
10 19.32  730     A
11 30.16  670     B
12 27.09  770     B
13 25.40  880     B
14 26.05 1000     B
15 33.49  760     B
16 35.62  590     B
17 26.07  910     B
18 36.78  650     B
19 34.95  810     B
20 43.67  500     B
```

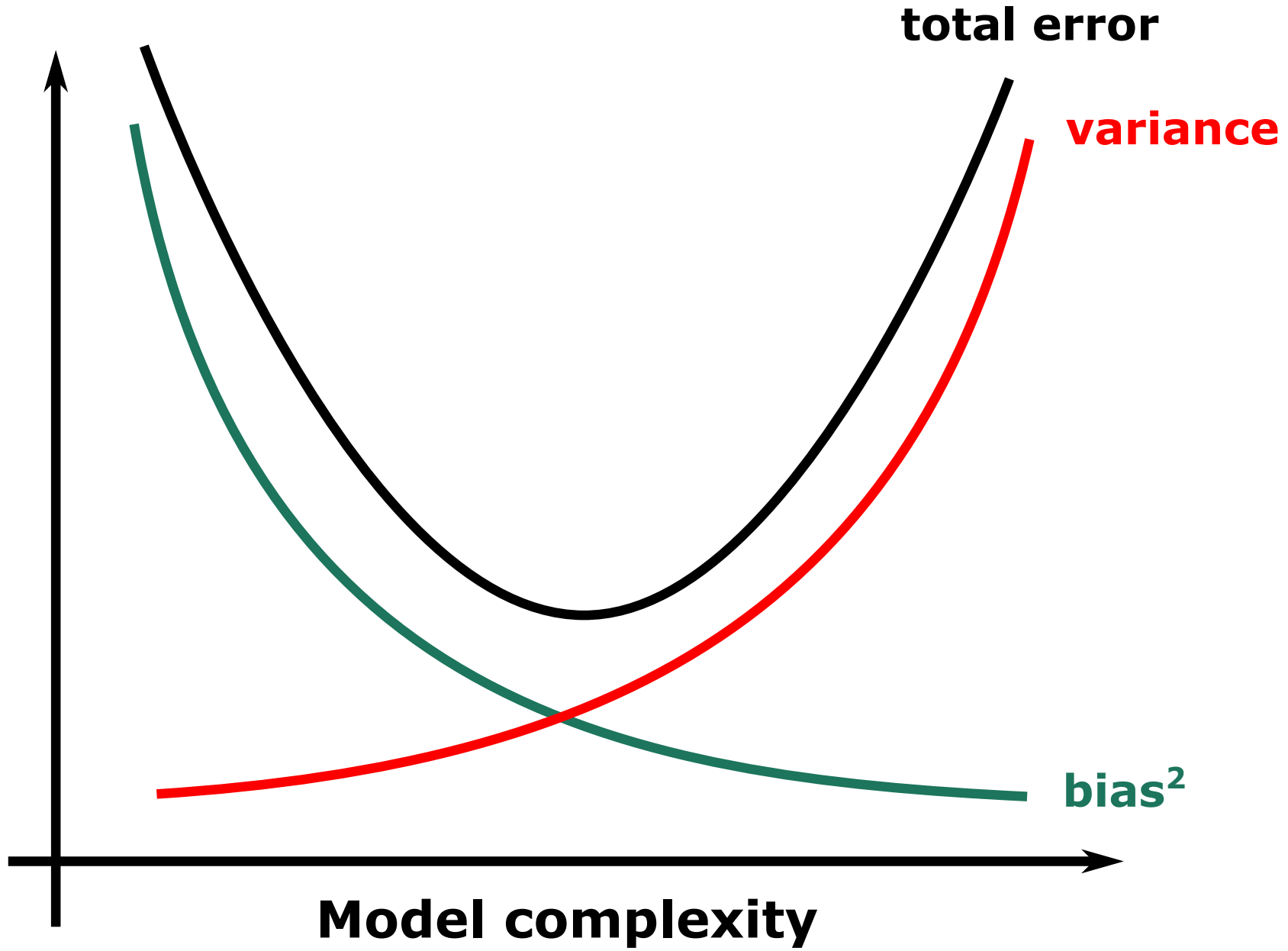
```
Call:
lm(formula = life ~ rpm + brand, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-5.5527 -1.7868 -0.0016  1.8395  4.9838

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.98560    3.51038  10.536 7.16e-09 ***
rpm         -0.02661    0.00452  -5.887 1.79e-05 ***
brandB      15.00425    1.35967  11.035 3.59e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.039 on 17 degrees of freedom
Multiple R-squared:  0.9003, Adjusted R-squared:  0.8886
F-statistic: 76.75 on 2 and 17 DF,  p-value: 3.086e-09
```

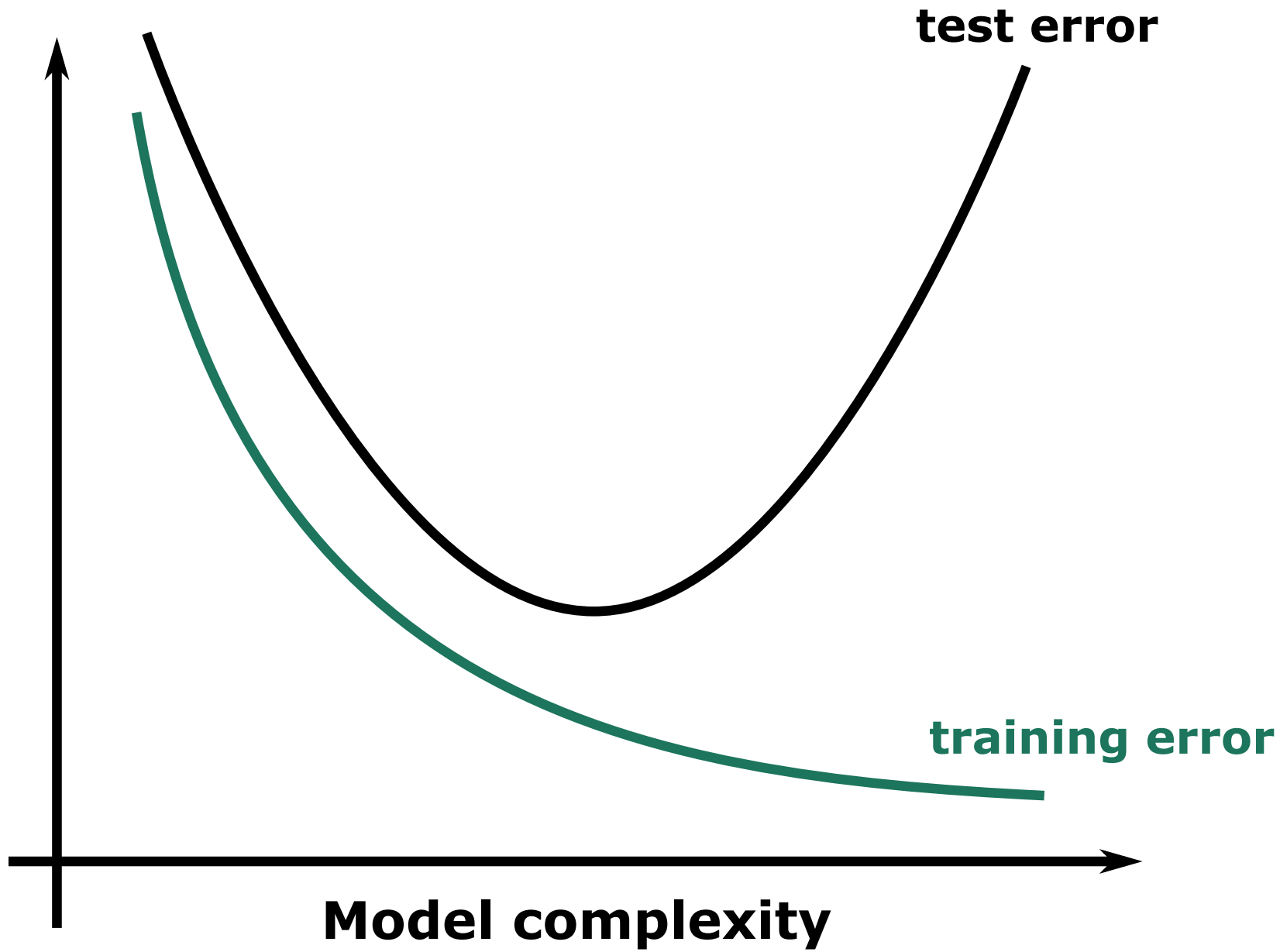

Y
100
50
0
−50
−100
−150
−4
−2
0
2
4
X

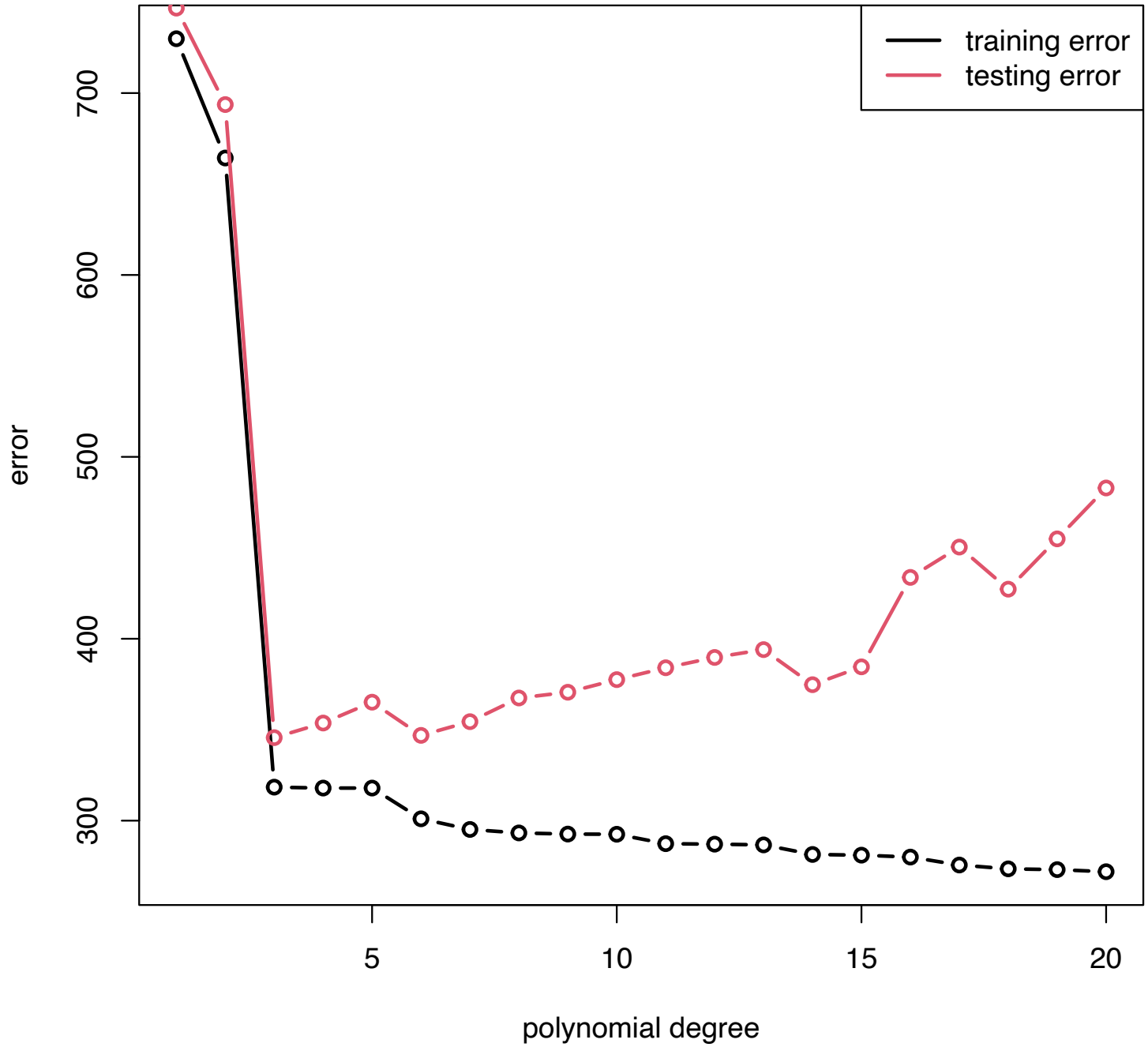
```
Call:
lm(formula = yy ~ xx + I(xx^2) + I(xx^3))

Residuals:
    Min      1Q  Median      3Q     Max
-61.339 -12.227   0.612  13.944  48.409

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.6731     3.1008  -1.507   0.1351
xx            2.5517     1.7729   1.439   0.1533
I(xx^2)      -0.6901     0.2719  -2.538   0.0128 *
I(xx^3)       0.9374     0.1062   8.826 4.93e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.67 on 96 degrees of freedom
Multiple R-squared:  0.8717,    Adjusted R-squared:  0.8677
F-statistic: 217.4 on 3 and 96 DF,  p-value: < 2.2e-16
```


degree=3
100
50
0
−50
−100
−150
Y
−4
−2
0
2
4
X


degree=3
degree=100
100
50
0
−50
−100
−150
Y
−4
−2
0
2
4
X


total error
variance
bias^2
Model complexity

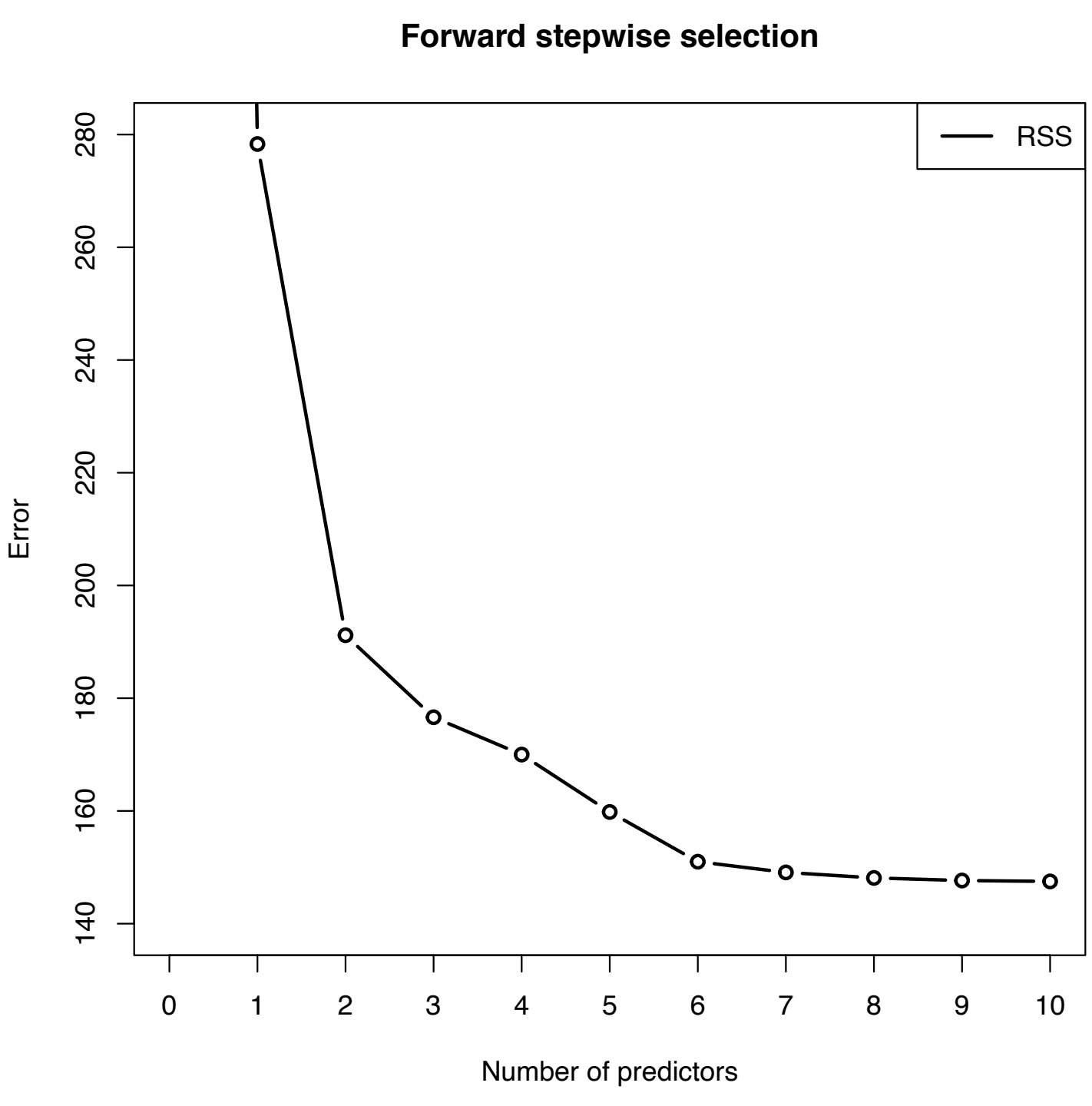

test error
training error
Model complexity

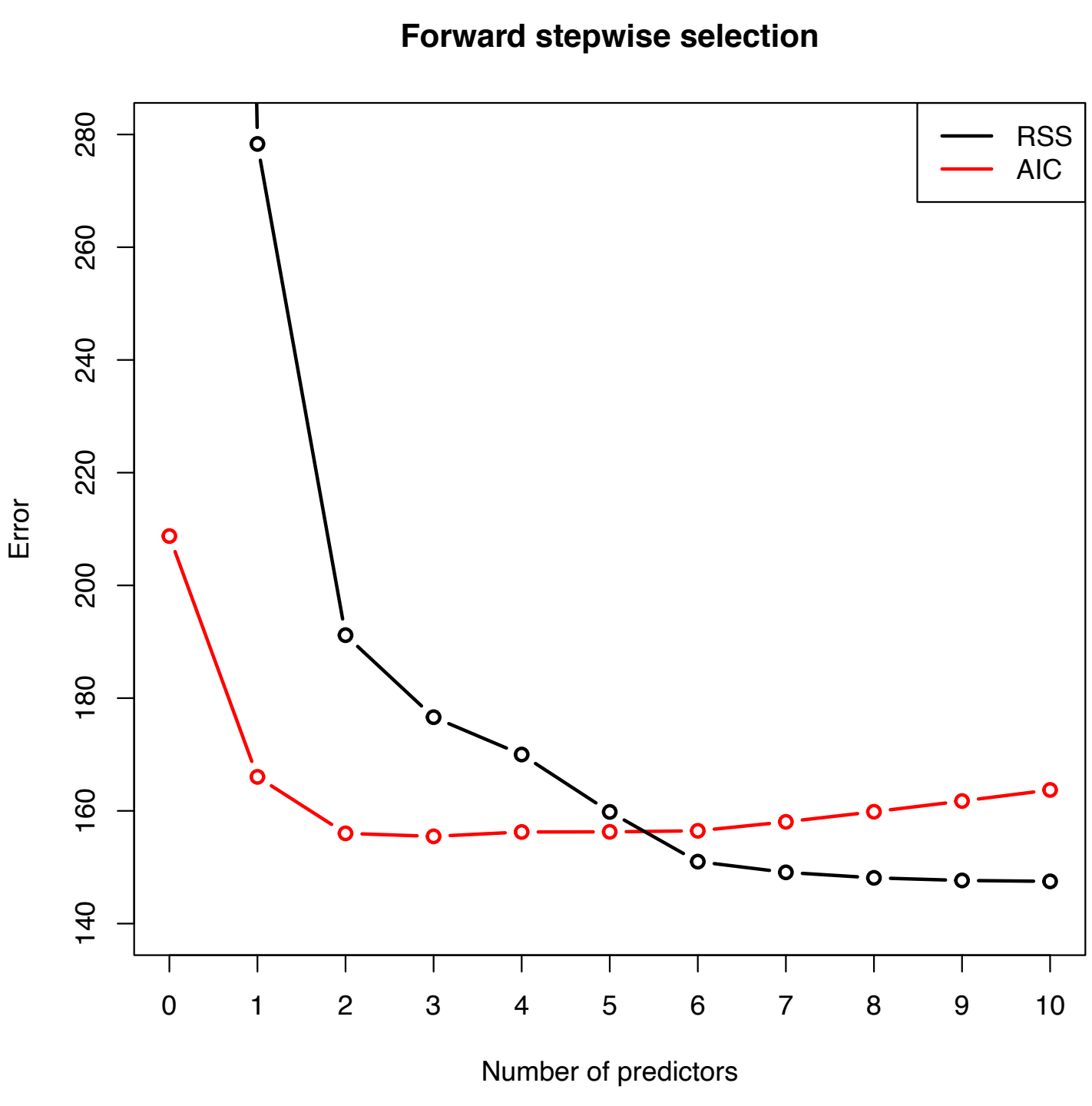

training error
testing error
error
700
600
500
400
300
5
10
15
20
polynomial degree

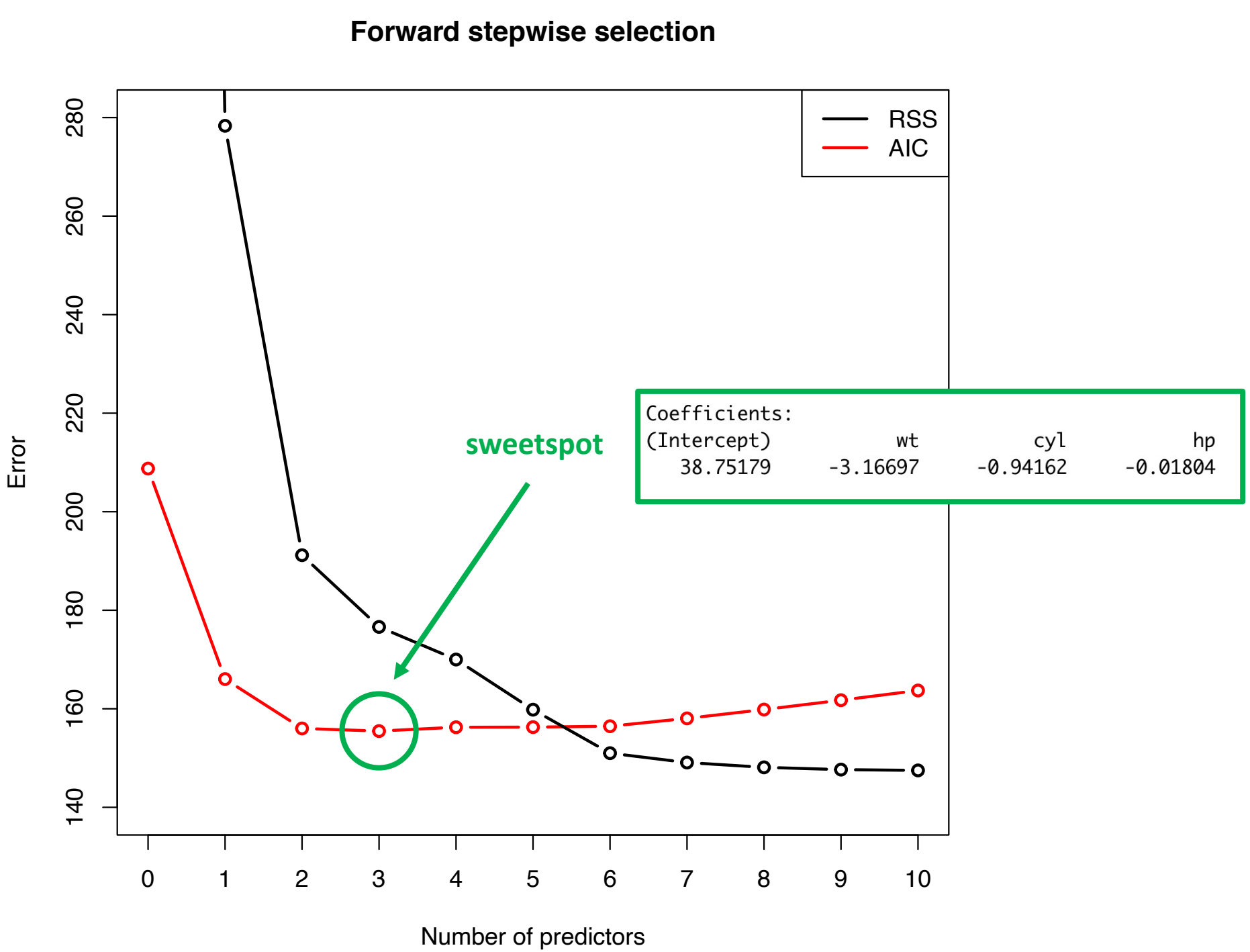
The dataset `mtcars` available in `R` contains the following information :

```
Description:

     The data was extracted from the 1974 _Motor Trend_ US magazine,
     and comprises fuel consumption and 10 aspects of automobile design
     and performance for 32 automobiles (1973-74 models).

Format:

     A data frame with 32 observations on 11 variables.

       [, 1]  mpg   Miles/(US) gallon
       [, 2]  cyl   Number of cylinders
       [, 3]  disp  Displacement (cu.in.)
       [, 4]  hp    Gross horsepower
       [, 5]  drat  Rear axle ratio
       [, 6]  wt    Weight (lb/1000)
       [, 7]  qsec  1/4 mile time
       [, 8]  vs    V/S
       [, 9]  am    Transmission (0 = automatic, 1 = manual)
       [,10]  gear  Number of forward gears
       [,11]  carb  Number of carburetors
```

```
> summary(lm(mpg ~ ., data=mtcars))

Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4506 -1.6044 -0.1196  1.2193  4.6271

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.30337   18.71788   0.657   0.5181
cyl         -0.11144    1.04502  -0.107   0.9161
disp         0.01334    0.01786   0.747   0.4635
hp          -0.02148    0.02177  -0.987   0.3350
drat         0.78711    1.63537   0.481   0.6353
wt          -3.71530    1.89441  -1.961   0.0633 .
qsec         0.82104    0.73084   1.123   0.2739
vs           0.31776    2.10451   0.151   0.8814
am           2.52023    2.05665   1.225   0.2340
gear         0.65541    1.49326   0.439   0.6652
carb        -0.19942    0.82875  -0.241   0.8122
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.65 on 21 degrees of freedom
Multiple R-squared:  0.869,     Adjusted R-squared:  0.8066
F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```


Forward stepwise selection
RSS
Error
140
160
180
200
220
240
260
280
0
1
2
3
4
5
6
7
8
9
10
Number of predictors


Forward stepwise selection
RSS
AIC
Error
280
260
240
220
200
180
160
140
0
1
2
3
4
5
6
7
8
9
10
Number of predictors


Forward stepwise selection
RSS
AIC
sweetspot
Coefficients:
(Intercept) wt cyl hp
38.75179 -3.16697 -0.94162 -0.01804
Error
Number of predictors
140
160
180
200
220
240
260
280
0
1
2
3
4
5
6
7
8
9
10