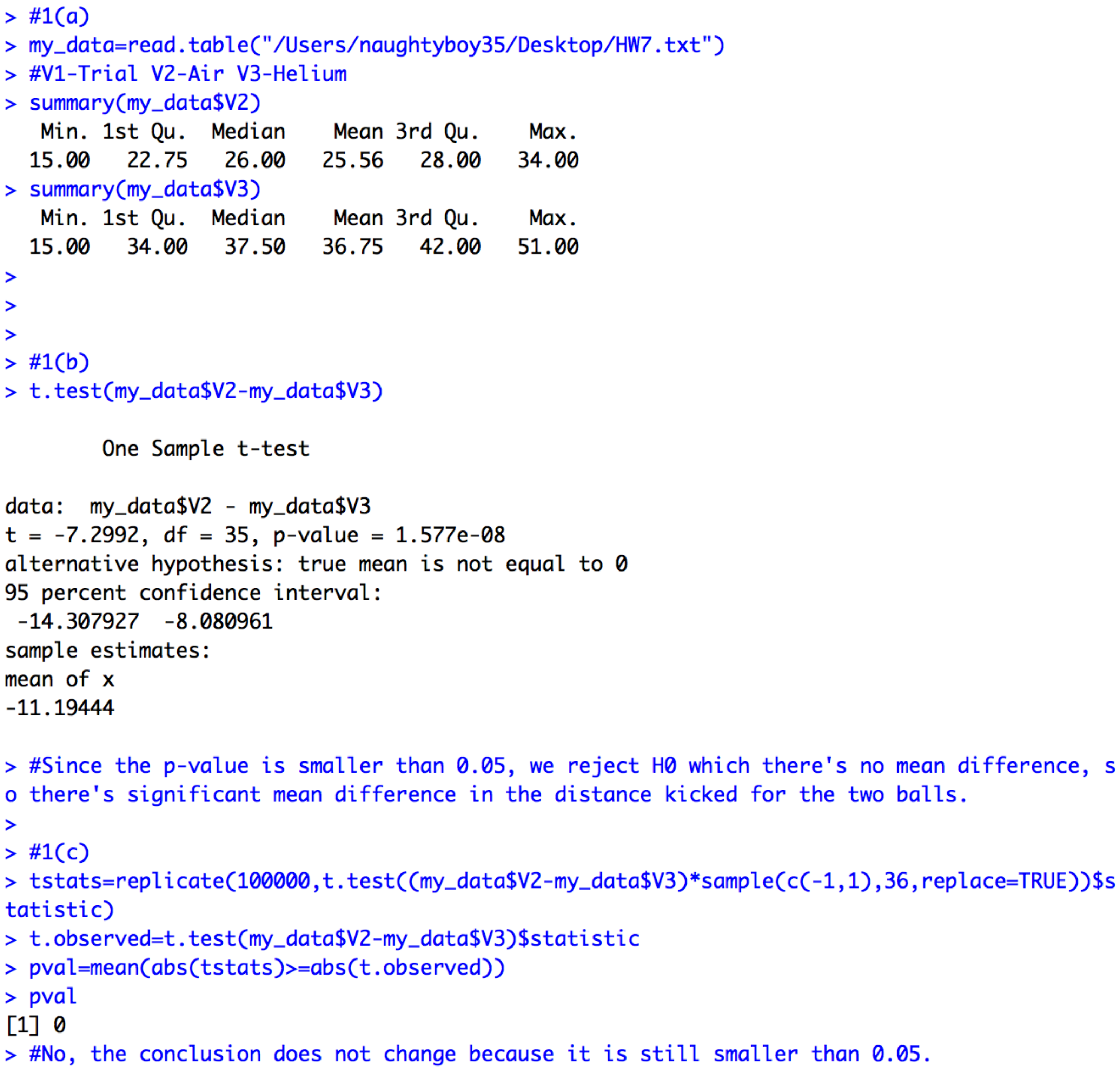
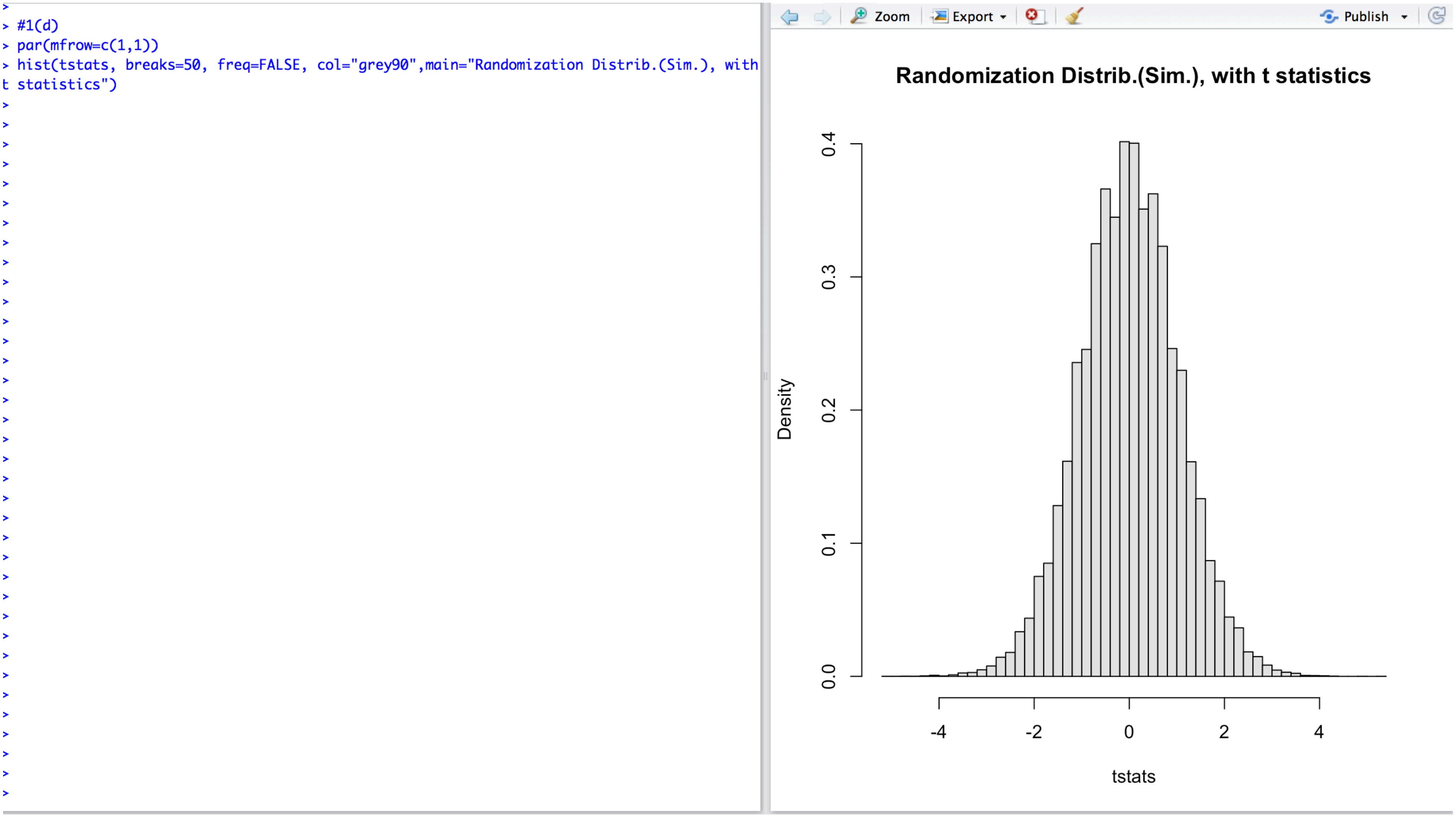
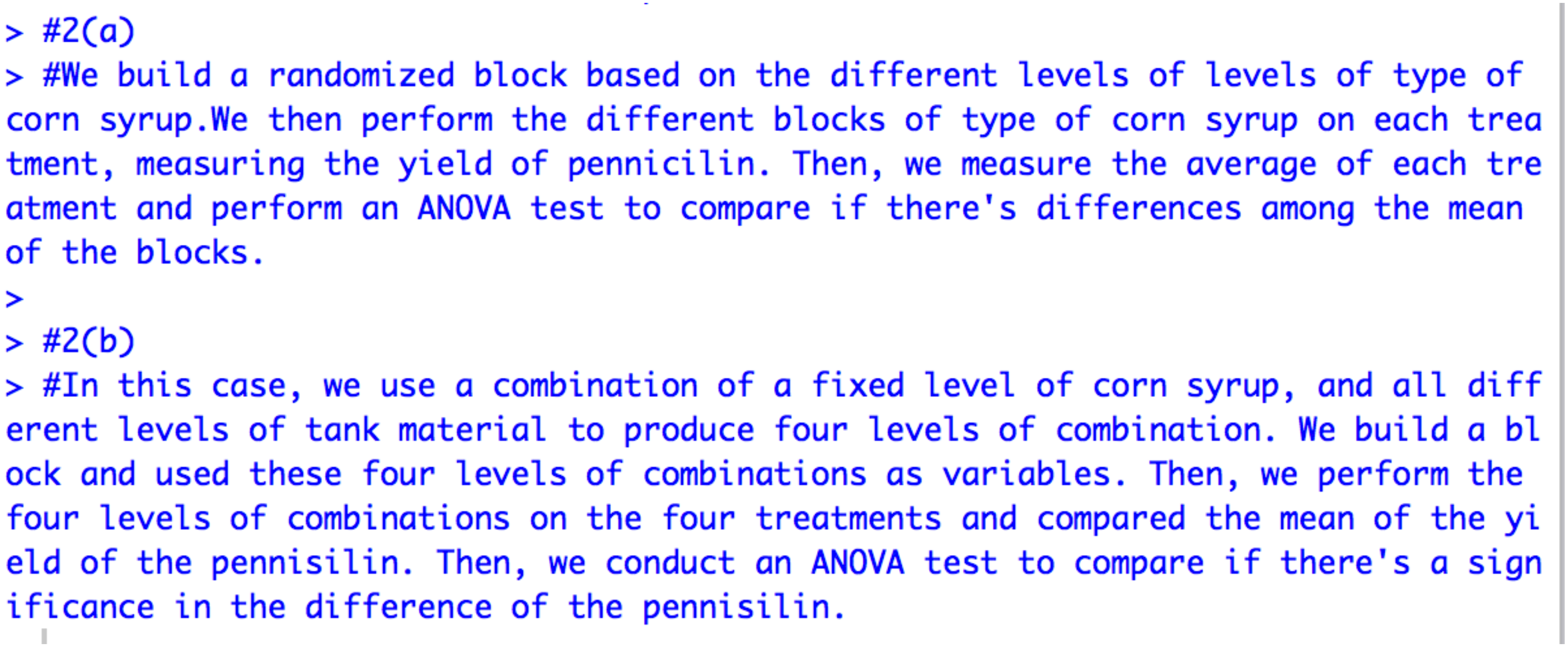


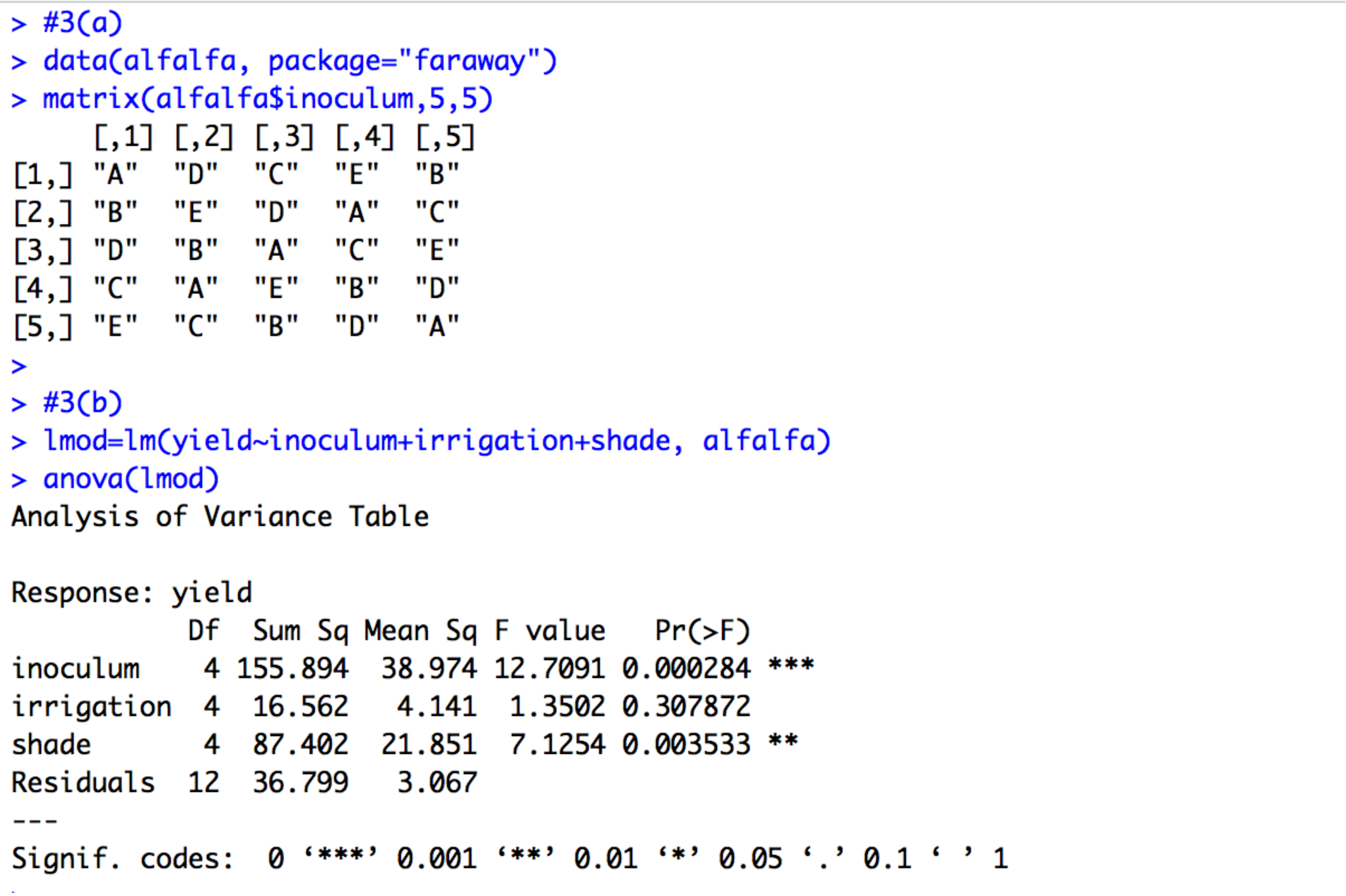
```
> #1(a)
> my_data=read.table("/Users/naughtyboy35/Desktop/HW7.txt")
> #V1-Trial V2-Air V3-Helium
> summary(my_data$V2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.00   22.75   26.00   25.56   28.00   34.00
> summary(my_data$V3)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.00   34.00   37.50   36.75   42.00   51.00
>
>
>
> #1(b)
> t.test(my_data$V2-my_data$V3)

	One Sample t-test

data:  my_data$V2 - my_data$V3
t = -7.2992, df = 35, p-value = 1.577e-08
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -14.307927  -8.080961
sample estimates:
mean of x
-11.19444

> #Since the p-value is smaller than 0.05, we reject H0 which there's no mean difference, s
o there's significant mean difference in the distance kicked for the two balls.
>
> #1(c)
> tstats=replicate(100000,t.test((my_data$V2-my_data$V3)*sample(c(-1,1),36,replace=TRUE))$s
tatistic)
> t.observed=t.test(my_data$V2-my_data$V3)$statistic
> pval=mean(abs(tstats)>=abs(t.observed))
> pval
[1] 0
> #No, the conclusion does not change because it is still smaller than 0.05.
```



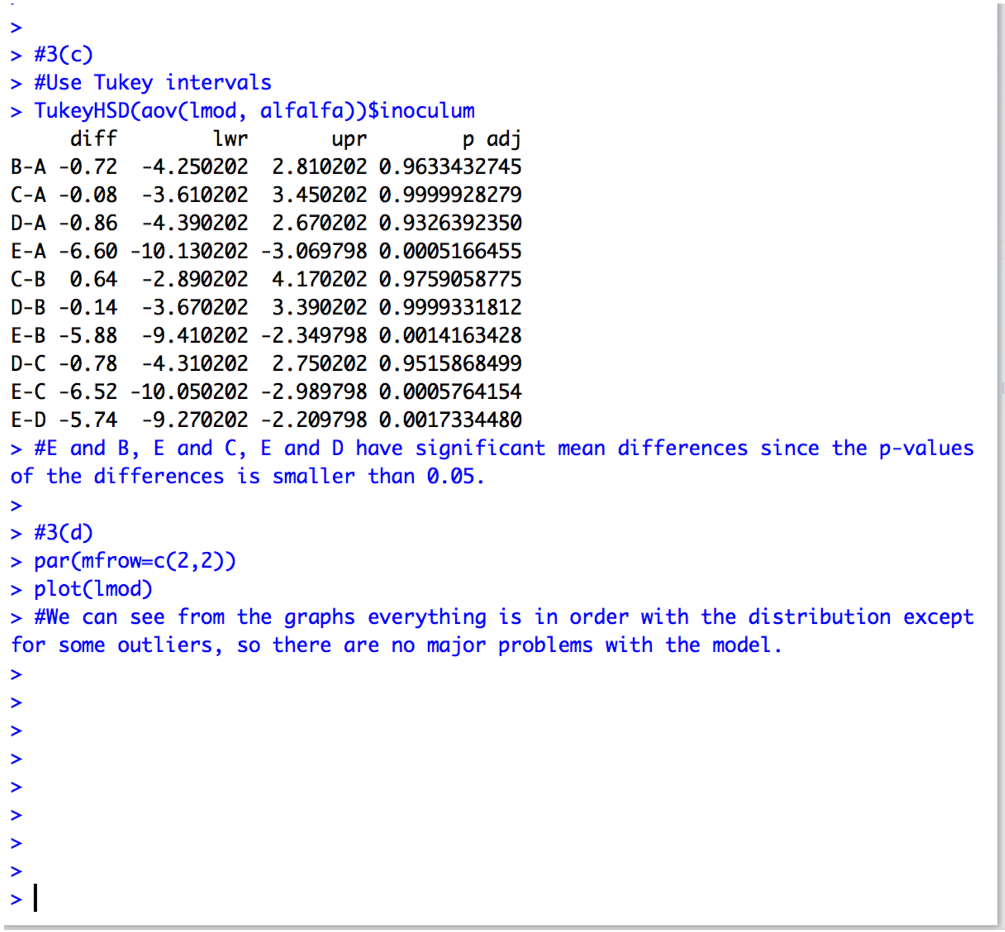
```
>
> #1(d)
> par(mfrow=c(1,1))
> hist(tstats, breaks=50, freq=FALSE, col="grey90",main="Randomization Distrib.(Sim.), with
t statistics")
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
```



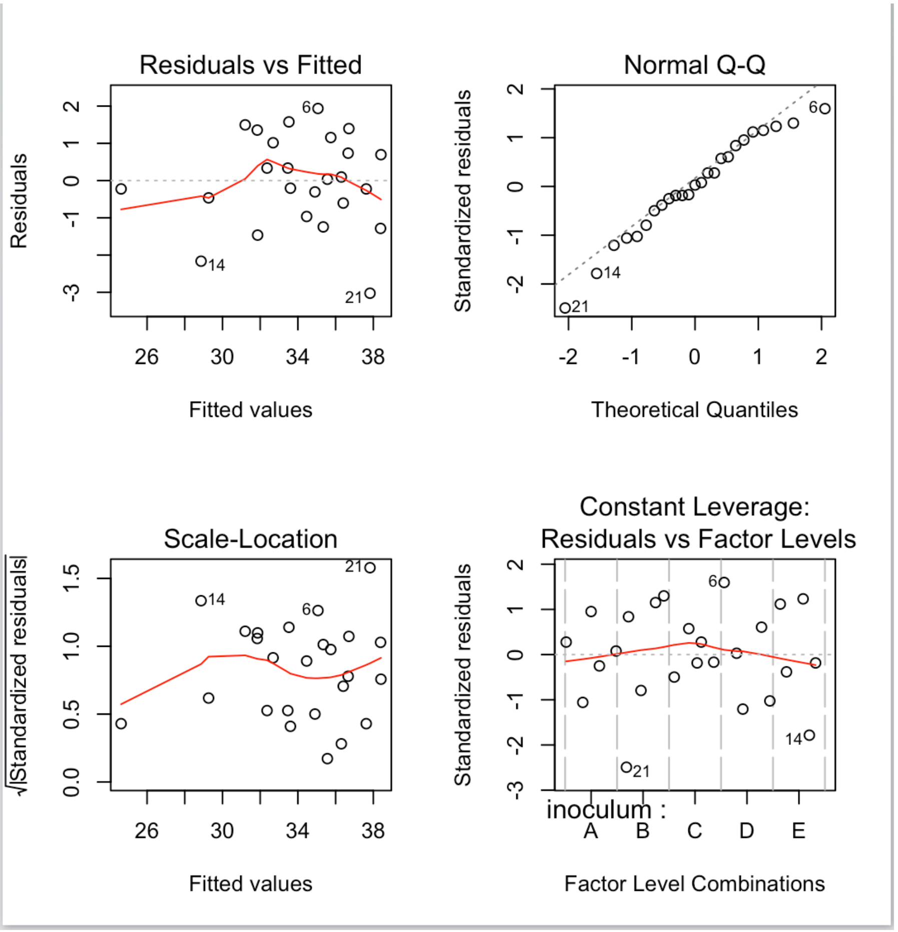
```
> #2(a)
> #We build a randomized block based on the different levels of levels of type of
corn syrup.We then perform the different blocks of type of corn syrup on each trea
tment, measuring the yield of pennicilin. Then, we measure the average of each tre
atment and perform an ANOVA test to compare if there's differences among the mean
of the blocks.
>
> #2(b)
> #In this case, we use a combination of a fixed level of corn syrup, and all diff
erent levels of tank material to produce four levels of combination. We build a bl
ock and used these four levels of combinations as variables. Then, we perform the
four levels of combinations on the four treatments and compared the mean of the yi
eld of the pennisilin. Then, we conduct an ANOVA test to compare if there's a sign
ificance in the difference of the pennisilin.
```

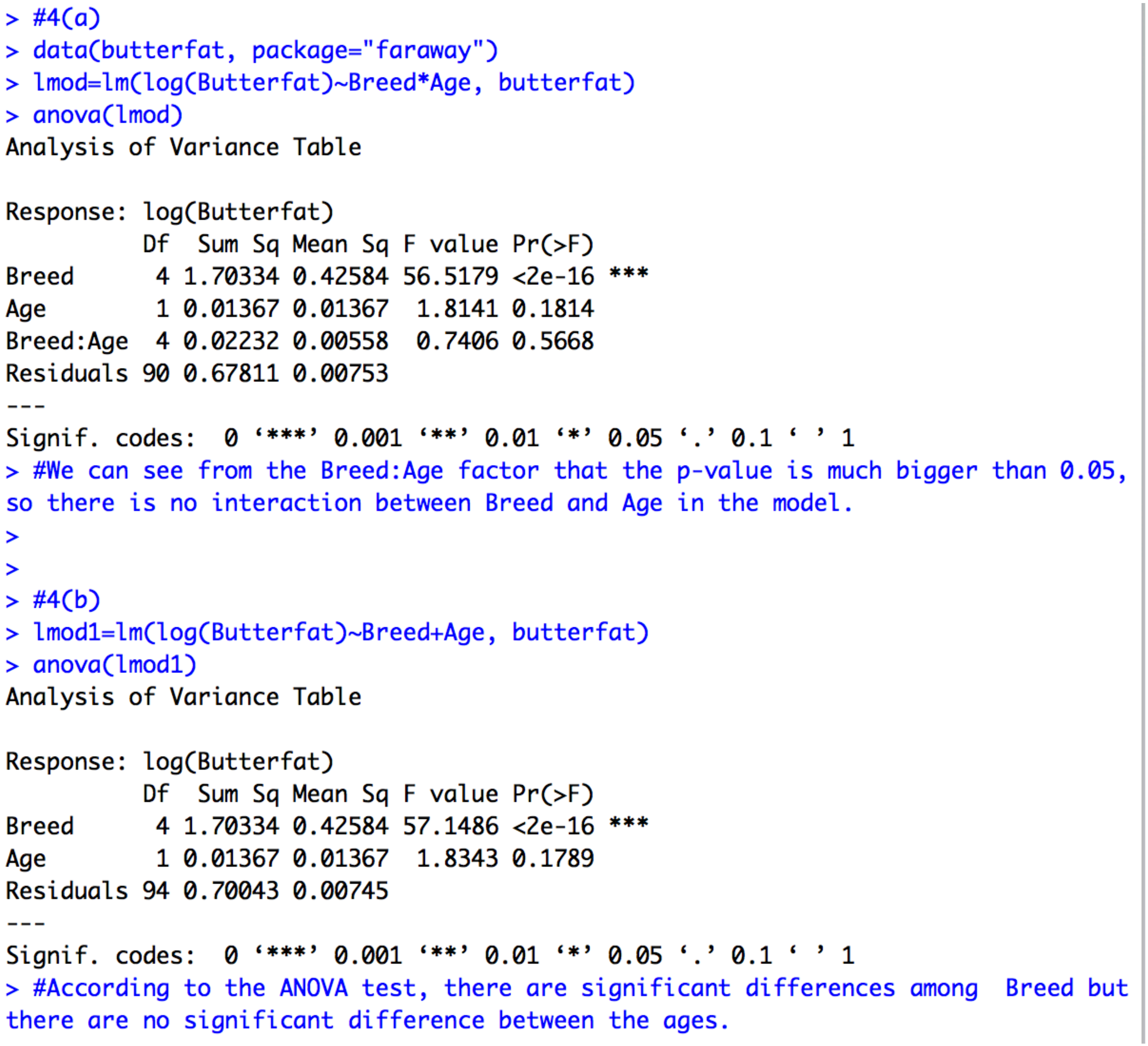


```
> #3(a)
> data(alfalfa, package="faraway")
> matrix(alfalfa$inoculum,5,5)
     [,1] [,2] [,3] [,4] [,5]
[1,] "A"  "D"  "C"  "E"  "B" 
[2,] "B"  "E"  "D"  "A"  "C" 
[3,] "D"  "B"  "A"  "C"  "E" 
[4,] "C"  "A"  "E"  "B"  "D" 
[5,] "E"  "C"  "B"  "D"  "A" 
> 
> #3(b)
> lmod=lm(yield~inoculum+irrigation+shade, alfalfa)
> anova(lmod)
Analysis of Variance Table

Response: yield
           Df  Sum Sq Mean Sq F value   Pr(>F)    
inoculum    4 155.894  38.974 12.7091 0.000284 ***
irrigation  4  16.562   4.141  1.3502 0.307872    
shade       4  87.402  21.851  7.1254 0.003533 ** 
Residuals  12  36.799   3.067                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

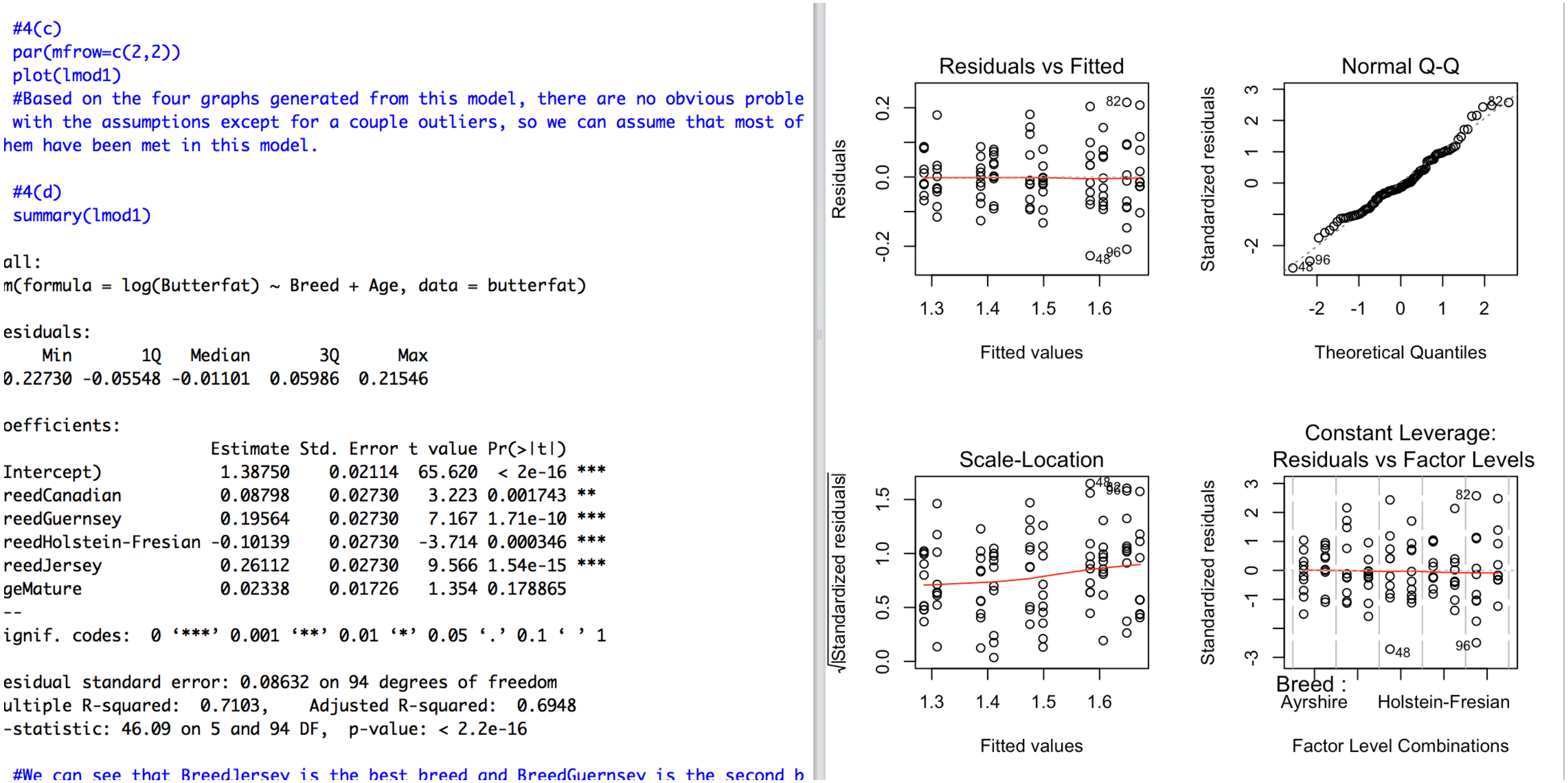


```
>
> #3(c)
> #Use Tukey intervals
> TukeyHSD(aov(lmod, alfalfa))$inoculum
     diff        lwr       upr        p adj
B-A -0.72  -4.250202  2.810202 0.9633432745
C-A -0.08  -3.610202  3.450202 0.9999928279
D-A -0.86  -4.390202  2.670202 0.9326392350
E-A -6.60 -10.130202 -3.069798 0.0005166455
C-B  0.64  -2.890202  4.170202 0.9759058775
D-B -0.14  -3.670202  3.390202 0.9999331812
E-B -5.88  -9.410202 -2.349798 0.0014163428
D-C -0.78  -4.310202  2.750202 0.9515868499
E-C -6.52 -10.050202 -2.989798 0.0005764154
E-D -5.74  -9.270202 -2.209798 0.0017334480
> #E and B, E and C, E and D have significant mean differences since the p-values
of the differences is smaller than 0.05.
>
> #3(d)
> par(mfrow=c(2,2))
> plot(lmod)
> #We can see from the graphs everything is in order with the distribution except
for some outliers, so there are no major problems with the model.
>
>
>
>
>
>
>
>
> |
```

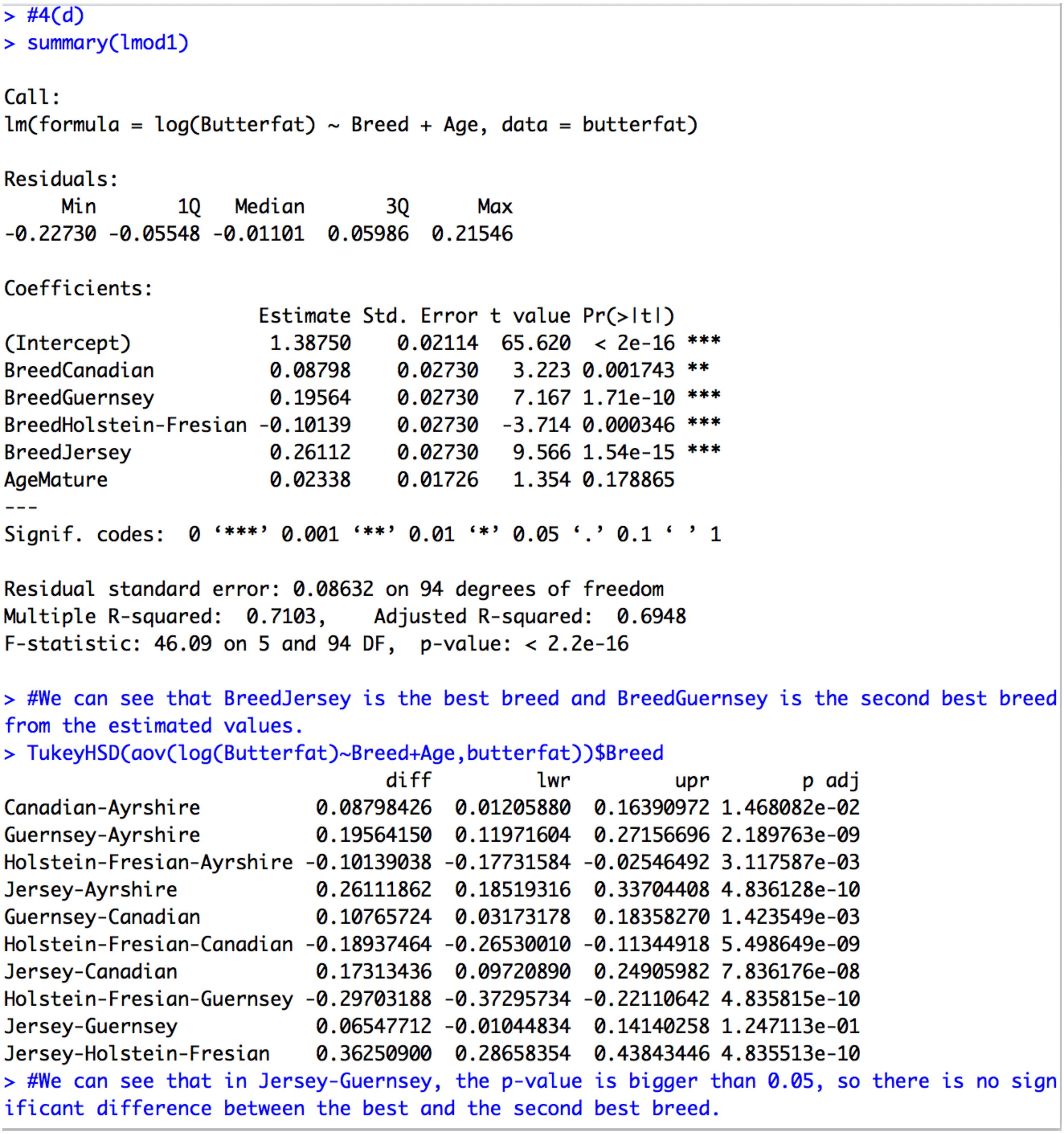




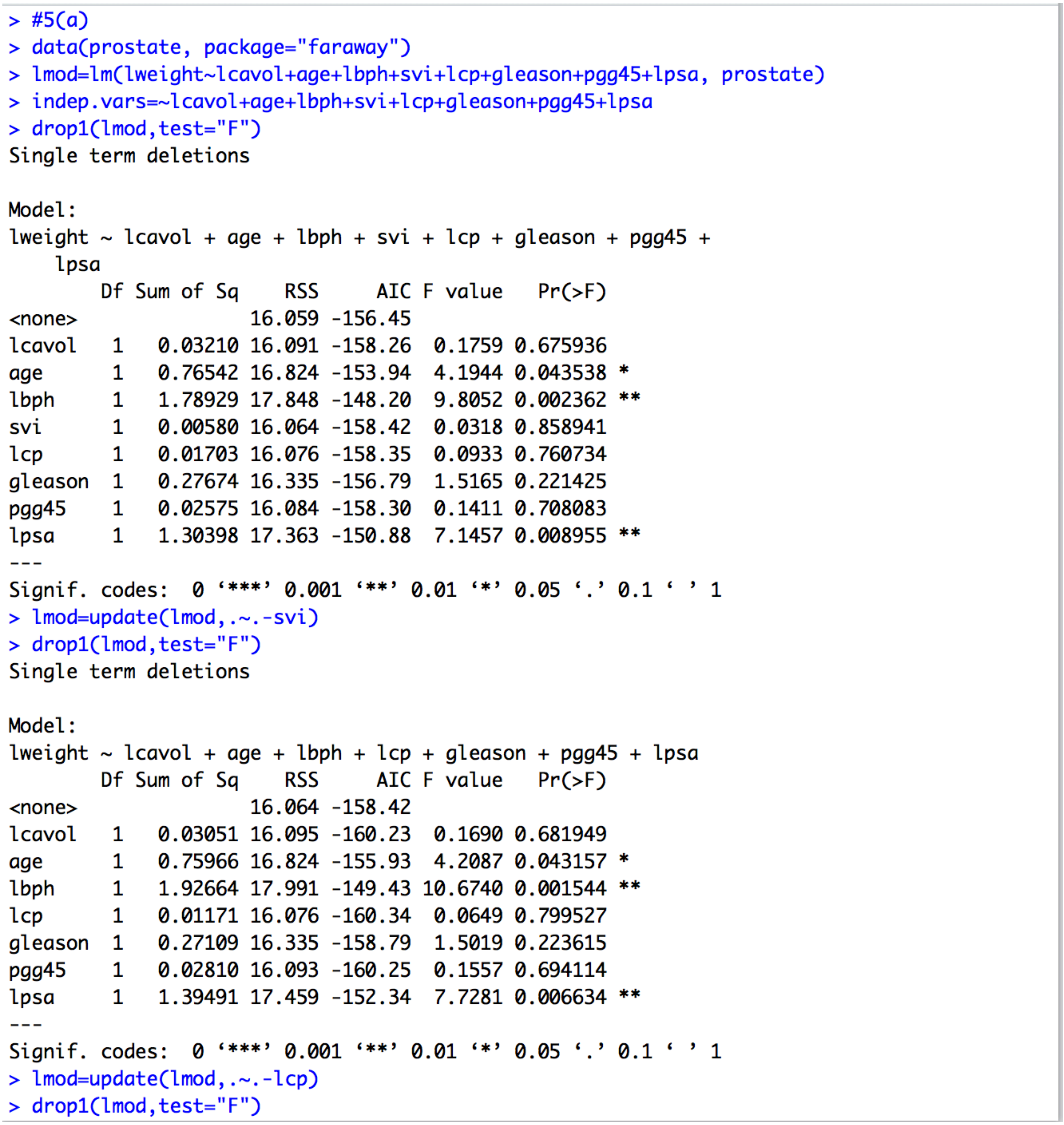
```
> #4(a)
> data(butterfat, package="faraway")
> lmod=lm(log(Butterfat)~Breed*Age, butterfat)
> anova(lmod)
Analysis of Variance Table

Response: log(Butterfat)
          Df  Sum Sq Mean Sq F value Pr(>F)    
Breed      4 1.70334 0.42584 56.5179 <2e-16 ***
Age        1 0.01367 0.01367  1.8141 0.1814    
Breed:Age  4 0.02232 0.00558  0.7406 0.5668    
Residuals 90 0.67811 0.00753                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #We can see from the Breed:Age factor that the p-value is much bigger than 0.05,
so there is no interaction between Breed and Age in the model.
> 
> 
> #4(b)
> lmod1=lm(log(Butterfat)~Breed+Age, butterfat)
> anova(lmod1)
Analysis of Variance Table

Response: log(Butterfat)
          Df  Sum Sq Mean Sq F value Pr(>F)    
Breed      4 1.70334 0.42584 57.1486 <2e-16 ***
Age        1 0.01367 0.01367  1.8343 0.1789    
Residuals 94 0.70043 0.00745                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #According to the ANOVA test, there are significant differences among  Breed but
there are no significant difference between the ages.
```



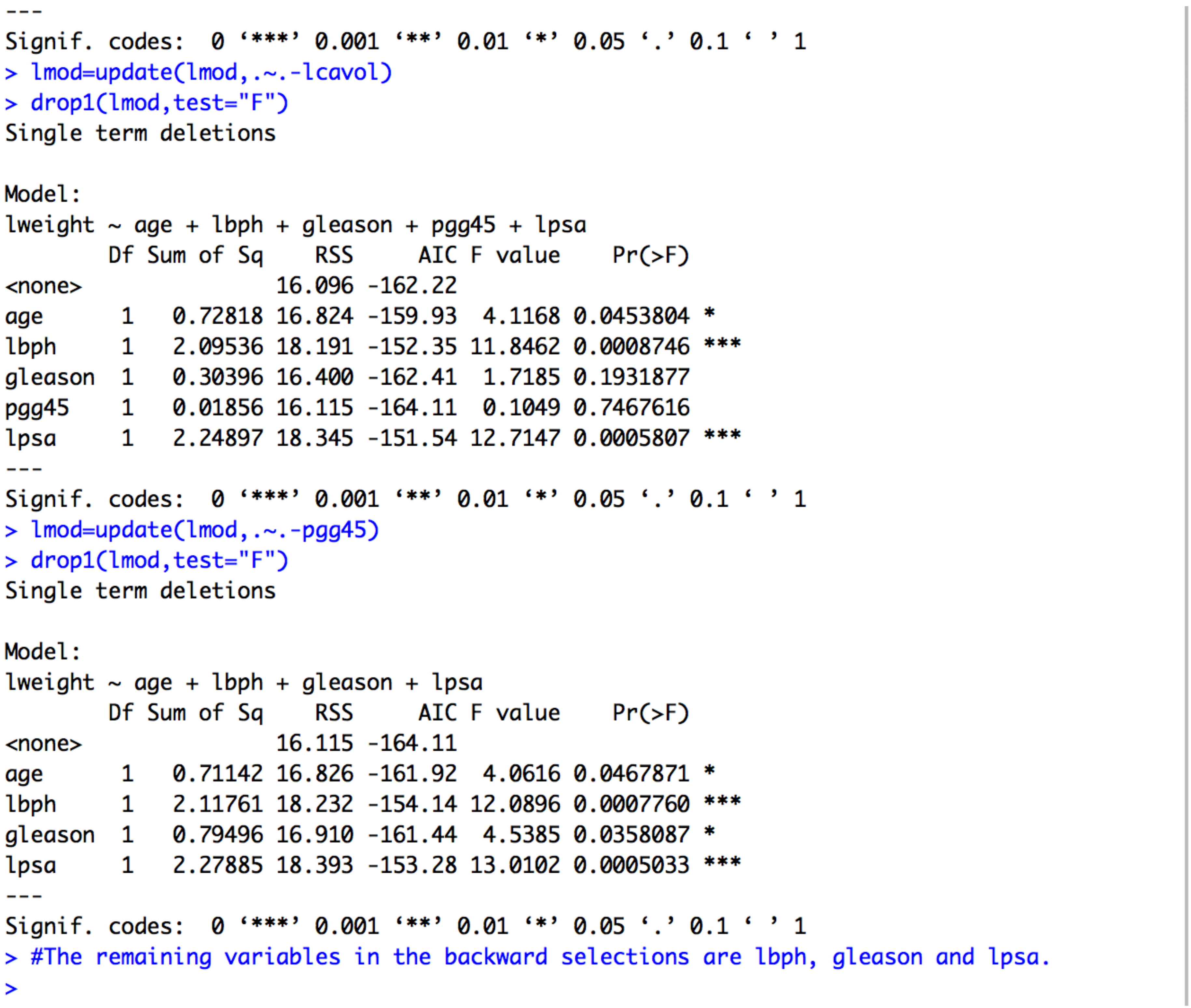
```
 #4(c)
 par(mfrow=c(2,2))
 plot(lmod1)
 #Based on the four graphs generated from this model, there are no obvious proble
 with the assumptions except for a couple outliers, so we can assume that most of
hem have been met in this model.

 #4(d)
 summary(lmod1)

all:
m(formula = log(Butterfat) ~ Breed + Age, data = butterfat)

esiduals:
    Min       1Q   Median       3Q      Max 
0.22730 -0.05548 -0.01101  0.05986  0.21546 

oefficients:
                     Estimate Std. Error t value Pr(>|t|)    
Intercept)            1.38750    0.02114  65.620  < 2e-16 ***
reedCanadian          0.08798    0.02730   3.223 0.001743 ** 
reedGuernsey          0.19564    0.02730   7.167 1.71e-10 ***
reedHolstein-Fresian -0.10139    0.02730  -3.714 0.000346 ***
reedJersey            0.26112    0.02730   9.566 1.54e-15 ***
geMature              0.02338    0.01726   1.354 0.178865    
--
ignif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

esidual standard error: 0.08632 on 94 degrees of freedom
ultiple R-squared:  0.7103,	Adjusted R-squared:  0.6948 
-statistic: 46.09 on 5 and 94 DF,  p-value: < 2.2e-16

 #We can see that BreedJersey is the best breed and BreedGuernsey is the second b
```



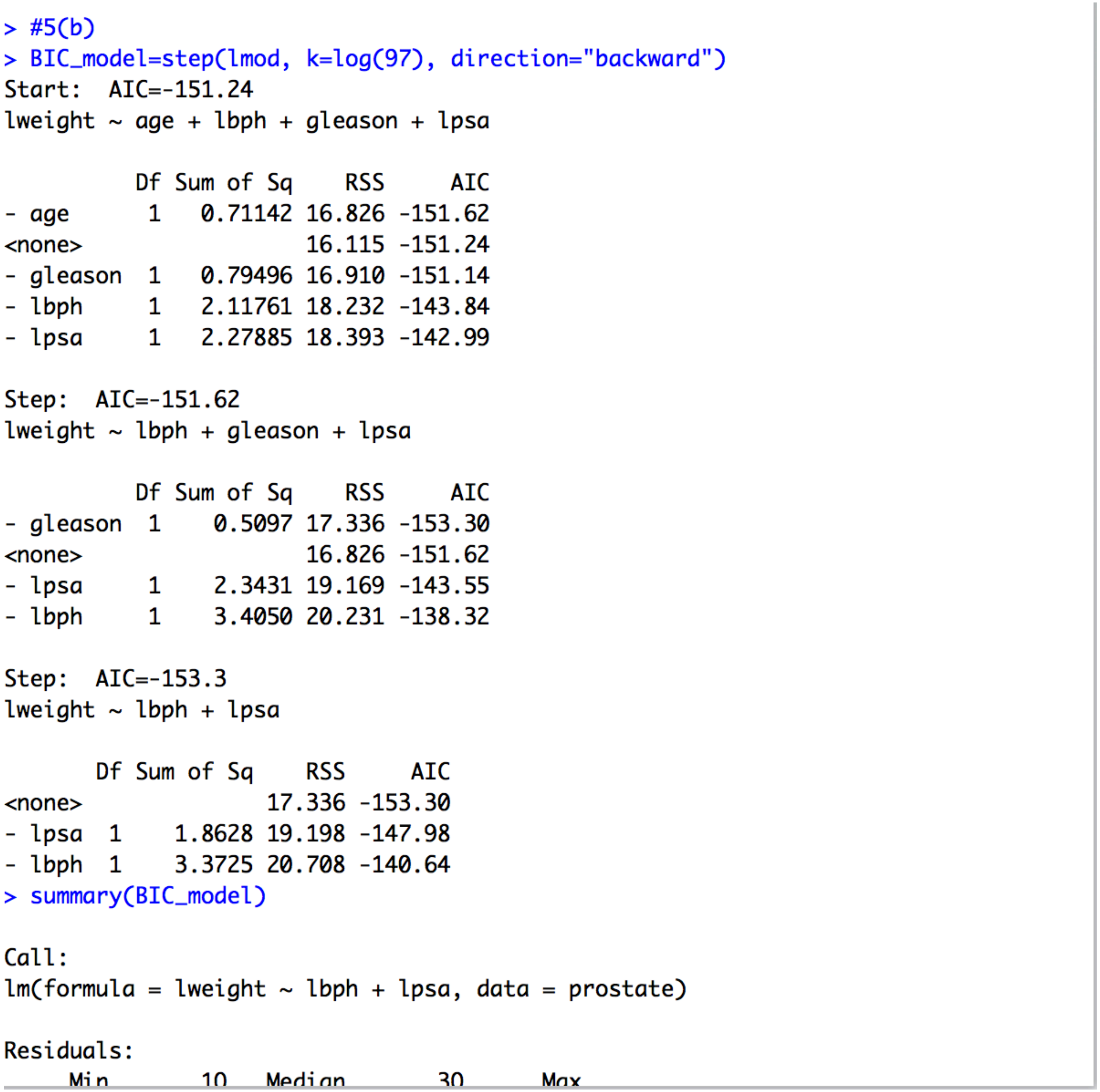
```
> #4(d)
> summary(lmod1)

Call:
lm(formula = log(Butterfat) ~ Breed + Age, data = butterfat)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.22730 -0.05548 -0.01101  0.05986  0.21546 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)            1.38750    0.02114  65.620  < 2e-16 ***
BreedCanadian          0.08798    0.02730   3.223 0.001743 ** 
BreedGuernsey          0.19564    0.02730   7.167 1.71e-10 ***
BreedHolstein-Fresian -0.10139    0.02730  -3.714 0.000346 ***
BreedJersey            0.26112    0.02730   9.566 1.54e-15 ***
AgeMature              0.02338    0.01726   1.354 0.178865    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08632 on 94 degrees of freedom
Multiple R-squared:  0.7103,	Adjusted R-squared:  0.6948 
F-statistic: 46.09 on 5 and 94 DF,  p-value: < 2.2e-16

> #We can see that BreedJersey is the best breed and BreedGuernsey is the second best breed
from the estimated values.
> TukeyHSD(aov(log(Butterfat)~Breed+Age,butterfat))$Breed
                                diff         lwr         upr        p adj
Canadian-Ayrshire          0.08798426  0.01205880  0.16390972 1.468082e-02
Guernsey-Ayrshire          0.19564150  0.11971604  0.27156696 2.189763e-09
Holstein-Fresian-Ayrshire -0.10139038 -0.17731584 -0.02546492 3.117587e-03
Jersey-Ayrshire            0.26111862  0.18519316  0.33704408 4.836128e-10
Guernsey-Canadian          0.10765724  0.03173178  0.18358270 1.423549e-03
Holstein-Fresian-Canadian -0.18937464 -0.26530010 -0.11344918 5.498649e-09
Jersey-Canadian            0.17313436  0.09720890  0.24905982 7.836176e-08
Holstein-Fresian-Guernsey -0.29703188 -0.37295734 -0.22110642 4.835815e-10
Jersey-Guernsey            0.06547712 -0.01044834  0.14140258 1.247113e-01
Jersey-Holstein-Fresian    0.36250900  0.28658354  0.43843446 4.835513e-10
> #We can see that in Jersey-Guernsey, the p-value is bigger than 0.05, so there is no sign
ificant difference between the best and the second best breed.
```



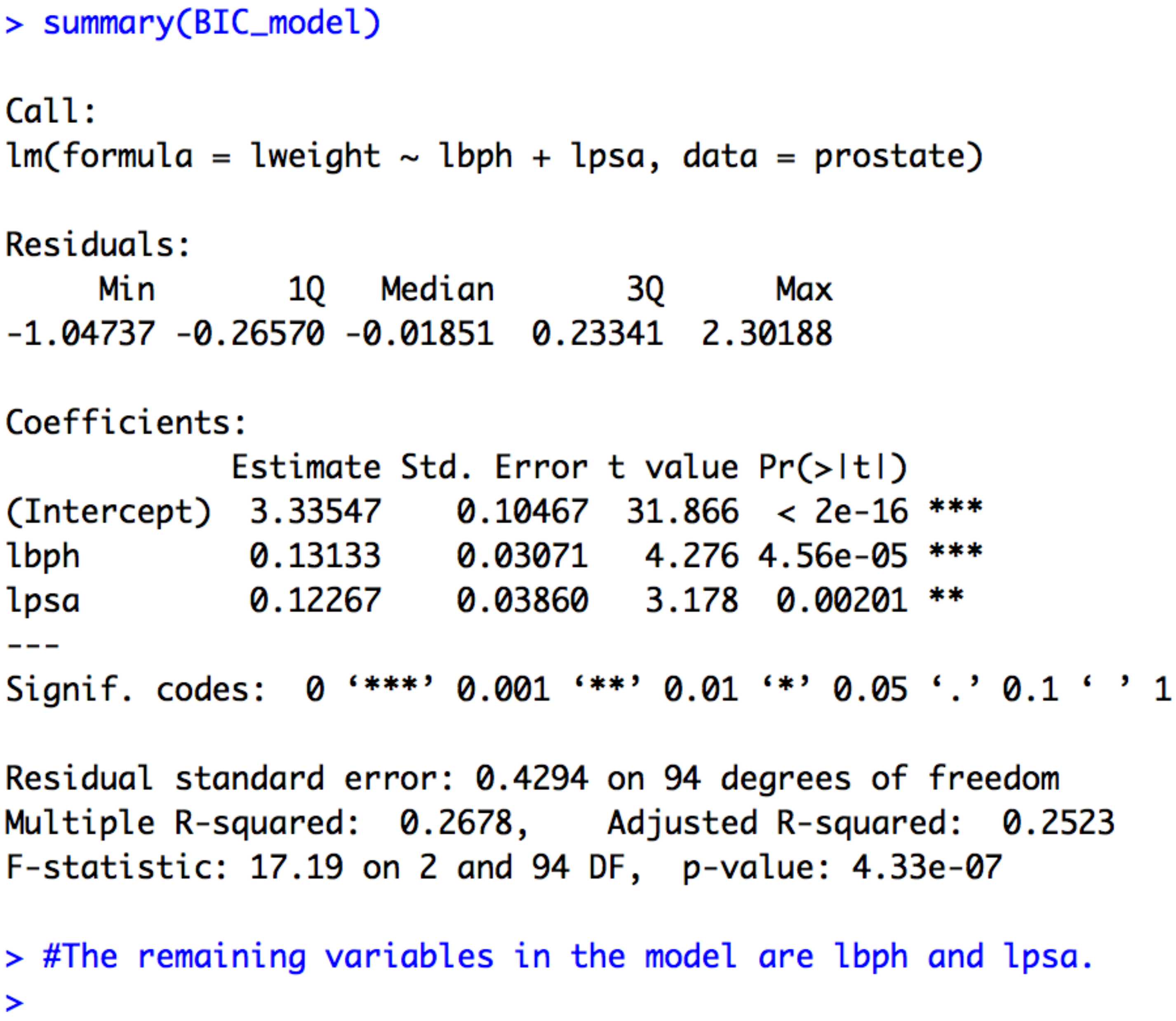
```
> #5(a)
> data(prostate, package="faraway")
> lmod=lm(lweight~lcavol+age+lbph+svi+lcp+gleason+pgg45+lpsa, prostate)
> indep.vars=~lcavol+age+lbph+svi+lcp+gleason+pgg45+lpsa
> drop1(lmod,test="F")
Single term deletions

Model:
lweight ~ lcavol + age + lbph + svi + lcp + gleason + pgg45 + 
    lpsa
        Df Sum of Sq    RSS     AIC F value   Pr(>F)   
<none>               16.059 -156.45                    
lcavol   1   0.03210 16.091 -158.26  0.1759 0.675936   
age      1   0.76542 16.824 -153.94  4.1944 0.043538 * 
lbph     1   1.78929 17.848 -148.20  9.8052 0.002362 **
svi      1   0.00580 16.064 -158.42  0.0318 0.858941   
lcp      1   0.01703 16.076 -158.35  0.0933 0.760734   
gleason  1   0.27674 16.335 -156.79  1.5165 0.221425   
pgg45    1   0.02575 16.084 -158.30  0.1411 0.708083   
lpsa     1   1.30398 17.363 -150.88  7.1457 0.008955 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod=update(lmod,.~.-svi)
> drop1(lmod,test="F")
Single term deletions

Model:
lweight ~ lcavol + age + lbph + lcp + gleason + pgg45 + lpsa
        Df Sum of Sq    RSS     AIC F value   Pr(>F)   
<none>               16.064 -158.42                    
lcavol   1   0.03051 16.095 -160.23  0.1690 0.681949   
age      1   0.75966 16.824 -155.93  4.2087 0.043157 * 
lbph     1   1.92664 17.991 -149.43 10.6740 0.001544 **
lcp      1   0.01171 16.076 -160.34  0.0649 0.799527   
gleason  1   0.27109 16.335 -158.79  1.5019 0.223615   
pgg45    1   0.02810 16.093 -160.25  0.1557 0.694114   
lpsa     1   1.39491 17.459 -152.34  7.7281 0.006634 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod=update(lmod,.~.-lcp)
> drop1(lmod,test="F")
```



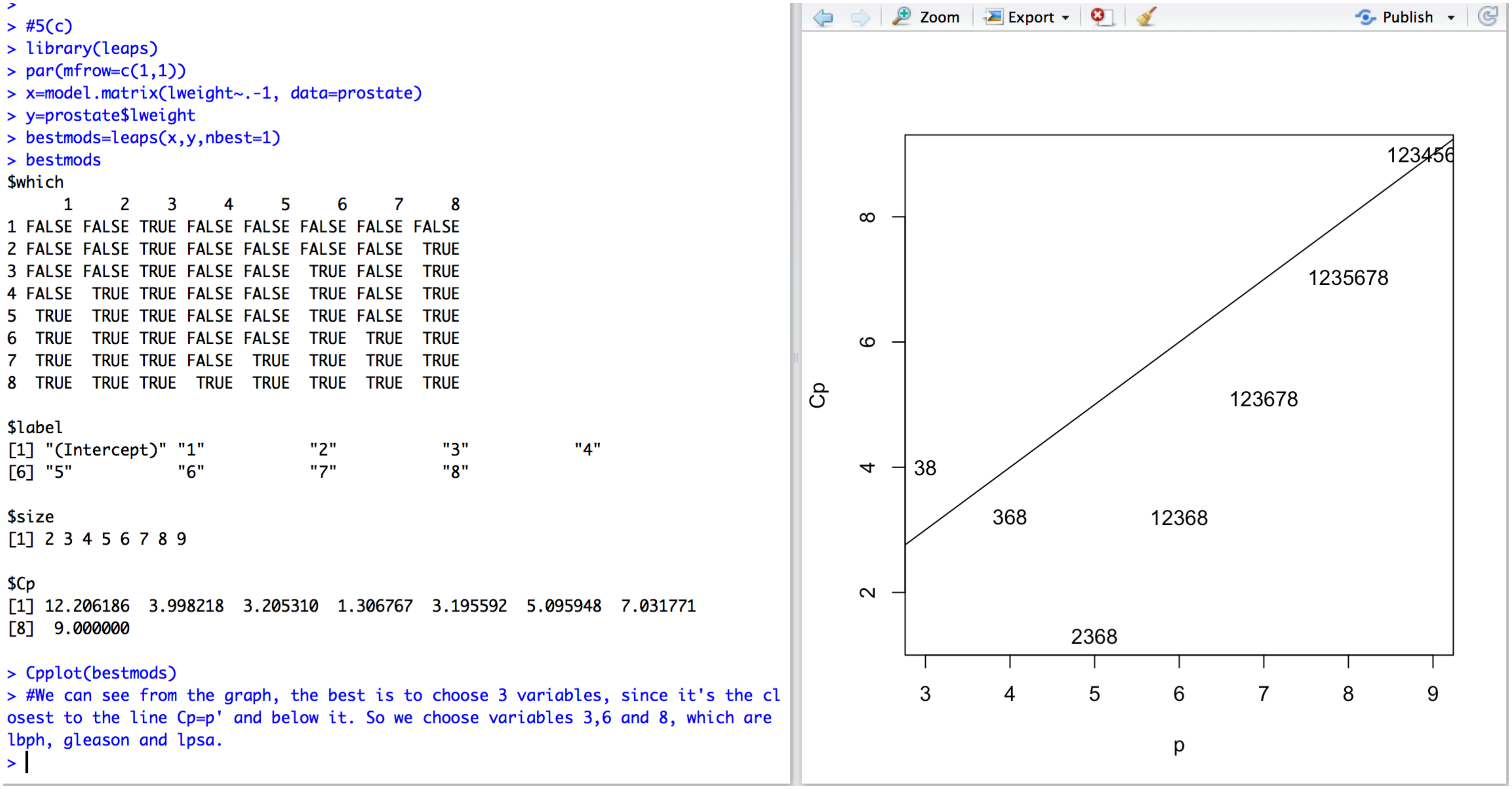
```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod=update(lmod,.~.-lcavol)
> drop1(lmod,test="F")
Single term deletions

Model:
lweight ~ age + lbph + gleason + pgg45 + lpsa
        Df Sum of Sq    RSS     AIC F value    Pr(>F)    
<none>                16.096 -162.22                      
age      1   0.72818 16.824 -159.93  4.1168 0.0453804 *  
lbph     1   2.09536 18.191 -152.35 11.8462 0.0008746 ***
gleason  1   0.30396 16.400 -162.41  1.7185 0.1931877    
pgg45    1   0.01856 16.115 -164.11  0.1049 0.7467616    
lpsa     1   2.24897 18.345 -151.54 12.7147 0.0005807 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod=update(lmod,.~.-pgg45)
> drop1(lmod,test="F")
Single term deletions

Model:
lweight ~ age + lbph + gleason + lpsa
        Df Sum of Sq    RSS     AIC F value    Pr(>F)    
<none>                16.115 -164.11                      
age      1   0.71142 16.826 -161.92  4.0616 0.0467871 *  
lbph     1   2.11761 18.232 -154.14 12.0896 0.0007760 ***
gleason  1   0.79496 16.910 -161.44  4.5385 0.0358087 *  
lpsa     1   2.27885 18.393 -153.28 13.0102 0.0005033 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #The remaining variables in the backward selections are lbph, gleason and lpsa.
> 
```



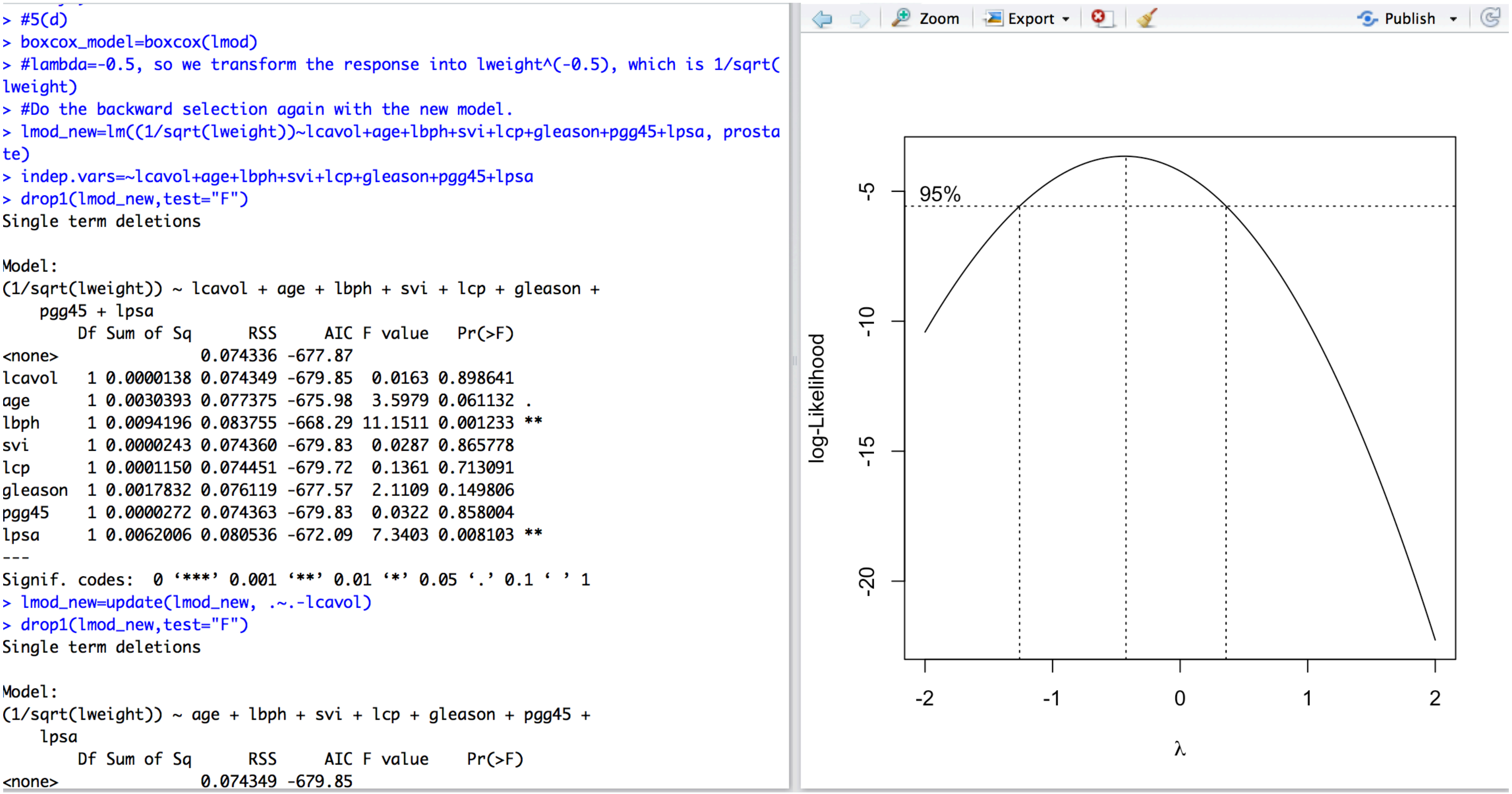
```
> #5(b)
> BIC_model=step(lmod, k=log(97), direction="backward")
Start:  AIC=-151.24
lweight ~ age + lbph + gleason + lpsa

          Df Sum of Sq    RSS     AIC
- age      1   0.71142 16.826 -151.62
<none>                 16.115 -151.24
- gleason  1   0.79496 16.910 -151.14
- lbph     1   2.11761 18.232 -143.84
- lpsa     1   2.27885 18.393 -142.99

Step:  AIC=-151.62
lweight ~ lbph + gleason + lpsa

          Df Sum of Sq    RSS     AIC
- gleason  1    0.5097 17.336 -153.30
<none>                 16.826 -151.62
- lpsa     1    2.3431 19.169 -143.55
- lbph     1    3.4050 20.231 -138.32

Step:  AIC=-153.3
lweight ~ lbph + lpsa

       Df Sum of Sq    RSS     AIC
<none>              17.336 -153.30
- lpsa  1    1.8628 19.198 -147.98
- lbph  1    3.3725 20.708 -140.64
> summary(BIC_model)

Call:
lm(formula = lweight ~ lbph + lpsa, data = prostate)

Residuals:
     Min       1Q   Median       3Q      Max
```



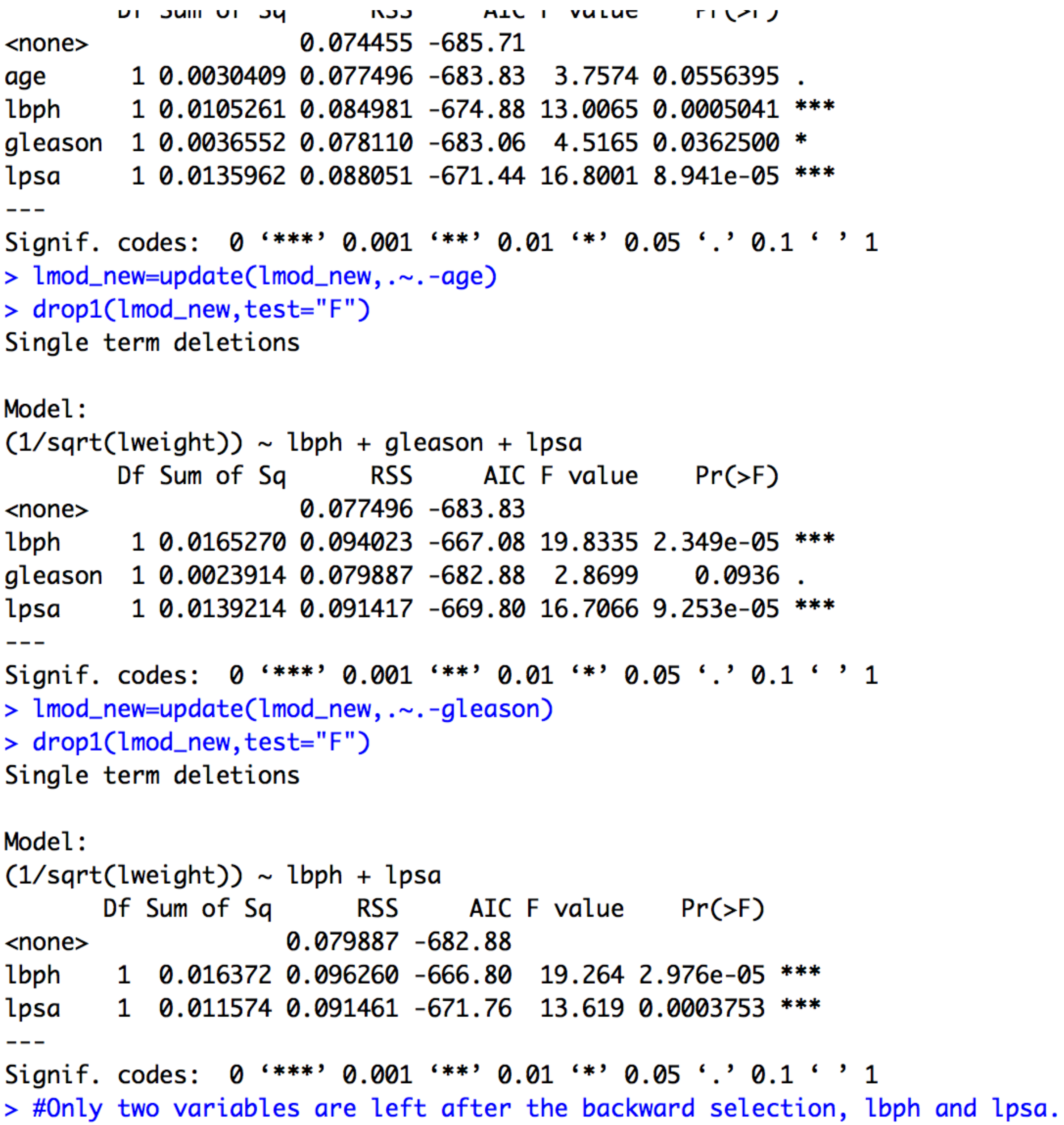
```
> summary(BIC_model)

Call:
lm(formula = lweight ~ lbph + lpsa, data = prostate)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.04737 -0.26570 -0.01851  0.23341  2.30188 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.33547    0.10467  31.866  < 2e-16 ***
lbph         0.13133    0.03071   4.276 4.56e-05 ***
lpsa         0.12267    0.03860   3.178  0.00201 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4294 on 94 degrees of freedom
Multiple R-squared:  0.2678,	Adjusted R-squared:  0.2523 
F-statistic: 17.19 on 2 and 94 DF,  p-value: 4.33e-07

> #The remaining variables in the model are lbph and lpsa.
> 
```



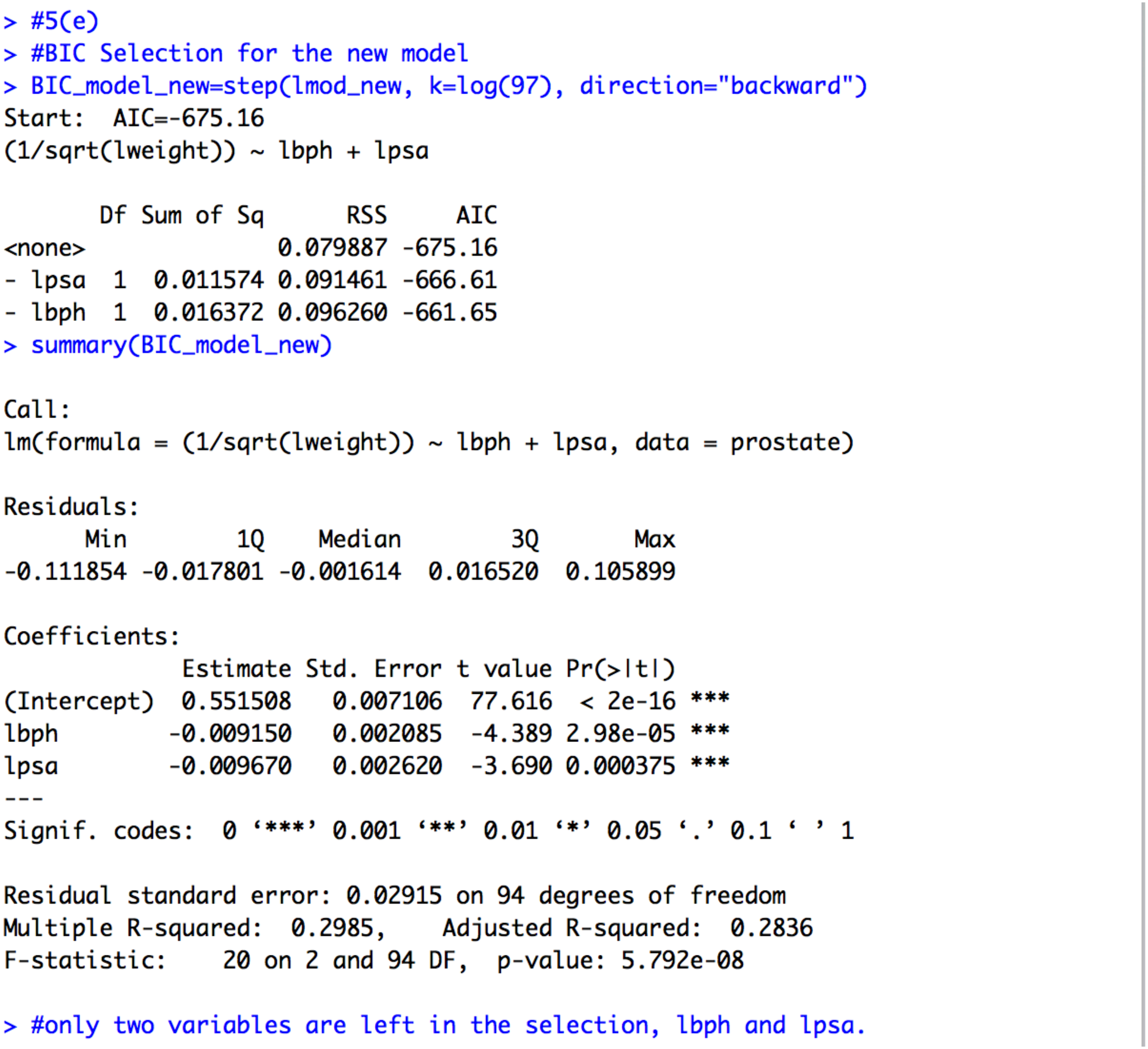
```
>
> #5(c)
> library(leaps)
> par(mfrow=c(1,1))
> x=model.matrix(lweight~.-1, data=prostate)
> y=prostate$lweight
> bestmods=leaps(x,y,nbest=1)
> bestmods
$which
      1     2     3     4     5     6     7     8
1 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
2 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE
3 FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
4 FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE
5  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE
6  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
7  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
8  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE

$label
[1] "(Intercept)" "1"           "2"           "3"           "4"
[6] "5"           "6"           "7"           "8"

$size
[1] 2 3 4 5 6 7 8 9

$Cp
[1] 12.206186  3.998218  3.205310  1.306767  3.195592  5.095948  7.031771
[8]  9.000000

> Cpplot(bestmods)
> #We can see from the graph, the best is to choose 3 variables, since it's the cl
osest to the line Cp=p' and below it. So we choose variables 3,6 and 8, which are
lbph, gleason and lpsa.
> 
```



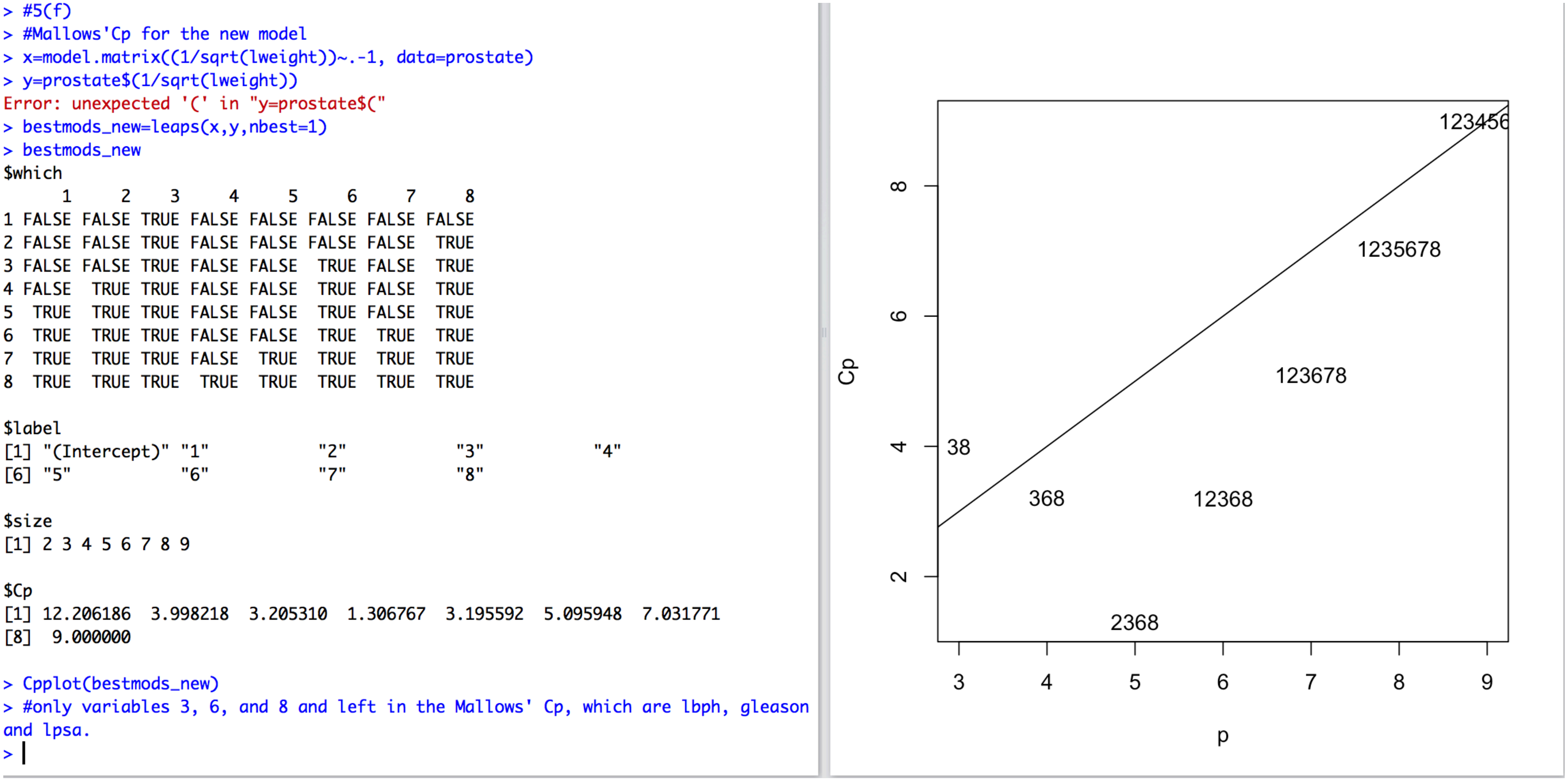
```
> #5(d)
> boxcox_model=boxcox(lmod)
> #lambda=-0.5, so we transform the response into lweight^(-0.5), which is 1/sqrt(
lweight)
> #Do the backward selection again with the new model.
> lmod_new=lm((1/sqrt(lweight))~lcavol+age+lbph+svi+lcp+gleason+pgg45+lpsa, prosta
te)
> indep.vars=~lcavol+age+lbph+svi+lcp+gleason+pgg45+lpsa
> drop1(lmod_new,test="F")
Single term deletions

Model:
(1/sqrt(lweight)) ~ lcavol + age + lbph + svi + lcp + gleason +
    pgg45 + lpsa
        Df Sum of Sq      RSS     AIC F value    Pr(>F)
<none>               0.074336 -677.87
lcavol   1 0.0000138 0.074349 -679.85  0.0163 0.898641
age      1 0.0030393 0.077375 -675.98  3.5979 0.061132 .
lbph     1 0.0094196 0.083755 -668.29 11.1511 0.001233 **
svi      1 0.0000243 0.074360 -679.83  0.0287 0.865778
lcp      1 0.0001150 0.074451 -679.72  0.1361 0.713091
gleason  1 0.0017832 0.076119 -677.57  2.1109 0.149806
pgg45    1 0.0000272 0.074363 -679.83  0.0322 0.858004
lpsa     1 0.0062006 0.080536 -672.09  7.3403 0.008103 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod_new=update(lmod_new, .~.-lcavol)
> drop1(lmod_new,test="F")
Single term deletions

Model:
(1/sqrt(lweight)) ~ age + lbph + svi + lcp + gleason + pgg45 +
    lpsa
        Df Sum of Sq      RSS     AIC F value     Pr(>F)
<none>               0.074349 -679.85
```



```
        Df Sum of Sq      RSS     AIC F value    Pr(>F)
<none>                0.074455 -685.71
age      1 0.0030409 0.077496 -683.83  3.7574 0.0556395 .
lbph     1 0.0105261 0.084981 -674.88 13.0065 0.0005041 ***
gleason  1 0.0036552 0.078110 -683.06  4.5165 0.0362500 *
lpsa     1 0.0135962 0.088051 -671.44 16.8001 8.941e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod_new=update(lmod_new,.~.-age)
> drop1(lmod_new,test="F")
Single term deletions

Model:
(1/sqrt(lweight)) ~ lbph + gleason + lpsa
        Df Sum of Sq      RSS     AIC F value    Pr(>F)
<none>                0.077496 -683.83
lbph     1 0.0165270 0.094023 -667.08 19.8335 2.349e-05 ***
gleason  1 0.0023914 0.079887 -682.88  2.8699    0.0936 .
lpsa     1 0.0139214 0.091417 -669.80 16.7066 9.253e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> lmod_new=update(lmod_new,.~.-gleason)
> drop1(lmod_new,test="F")
Single term deletions

Model:
(1/sqrt(lweight)) ~ lbph + lpsa
       Df Sum of Sq      RSS     AIC F value    Pr(>F)
<none>               0.079887 -682.88
lbph    1  0.016372 0.096260 -666.80  19.264 2.976e-05 ***
lpsa    1  0.011574 0.091461 -671.76  13.619 0.0003753 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #Only two variables are left after the backward selection, lbph and lpsa.
```



```
> #5(e)
> #BIC Selection for the new model
> BIC_model_new=step(lmod_new, k=log(97), direction="backward")
Start:  AIC=-675.16
(1/sqrt(lweight)) ~ lbph + lpsa

       Df Sum of Sq      RSS     AIC
<none>               0.079887 -675.16
- lpsa  1  0.011574 0.091461 -666.61
- lbph  1  0.016372 0.096260 -661.65
> summary(BIC_model_new)

Call:
lm(formula = (1/sqrt(lweight)) ~ lbph + lpsa, data = prostate)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.111854 -0.017801 -0.001614  0.016520  0.105899 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.551508   0.007106  77.616  < 2e-16 ***
lbph        -0.009150   0.002085  -4.389 2.98e-05 ***
lpsa        -0.009670   0.002620  -3.690 0.000375 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02915 on 94 degrees of freedom
Multiple R-squared:  0.2985,	Adjusted R-squared:  0.2836 
F-statistic:    20 on 2 and 94 DF,  p-value: 5.792e-08

> #only two variables are left in the selection, lbph and lpsa.
```



```
> #5(f)
> #Mallows'Cp for the new model
> x=model.matrix((1/sqrt(lweight))~.-1, data=prostate)
> y=prostate$(1/sqrt(lweight))
Error: unexpected '(' in "y=prostate$("
> bestmods_new=leaps(x,y,nbest=1)
> bestmods_new
$which
      1     2     3     4     5     6     7     8
1 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
2 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE
3 FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
4 FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE
5  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE
6  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
7  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
8  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE

$label
[1] "(Intercept)" "1"           "2"           "3"           "4"
[6] "5"           "6"           "7"           "8"

$size
[1] 2 3 4 5 6 7 8 9

$Cp
[1] 12.206186  3.998218  3.205310  1.306767  3.195592  5.095948  7.031771
[8]  9.000000

> Cpplot(bestmods_new)
> #only variables 3, 6, and 8 and left in the Mallows' Cp, which are lbph, gleason
and lpsa.
> 
```