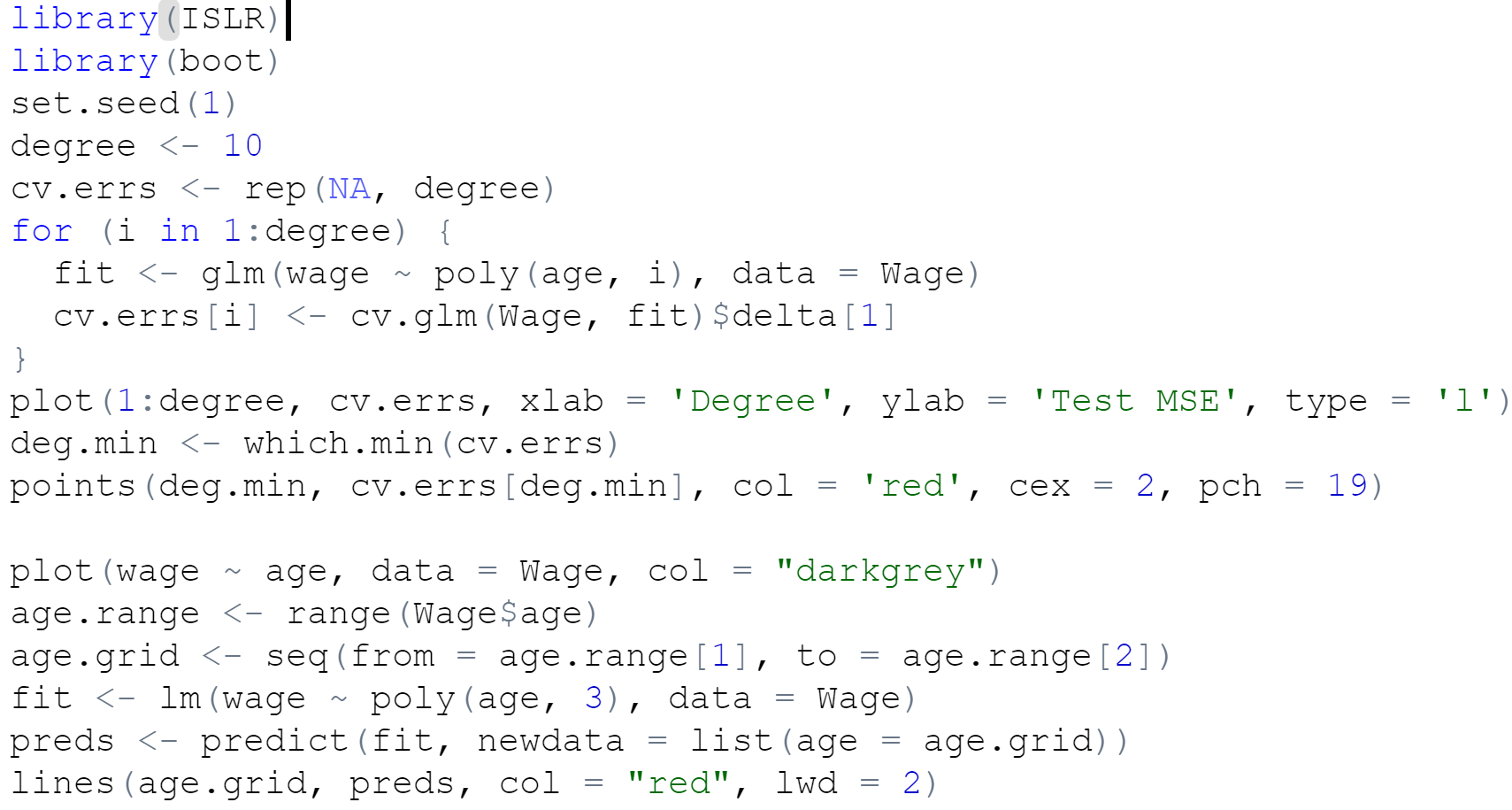
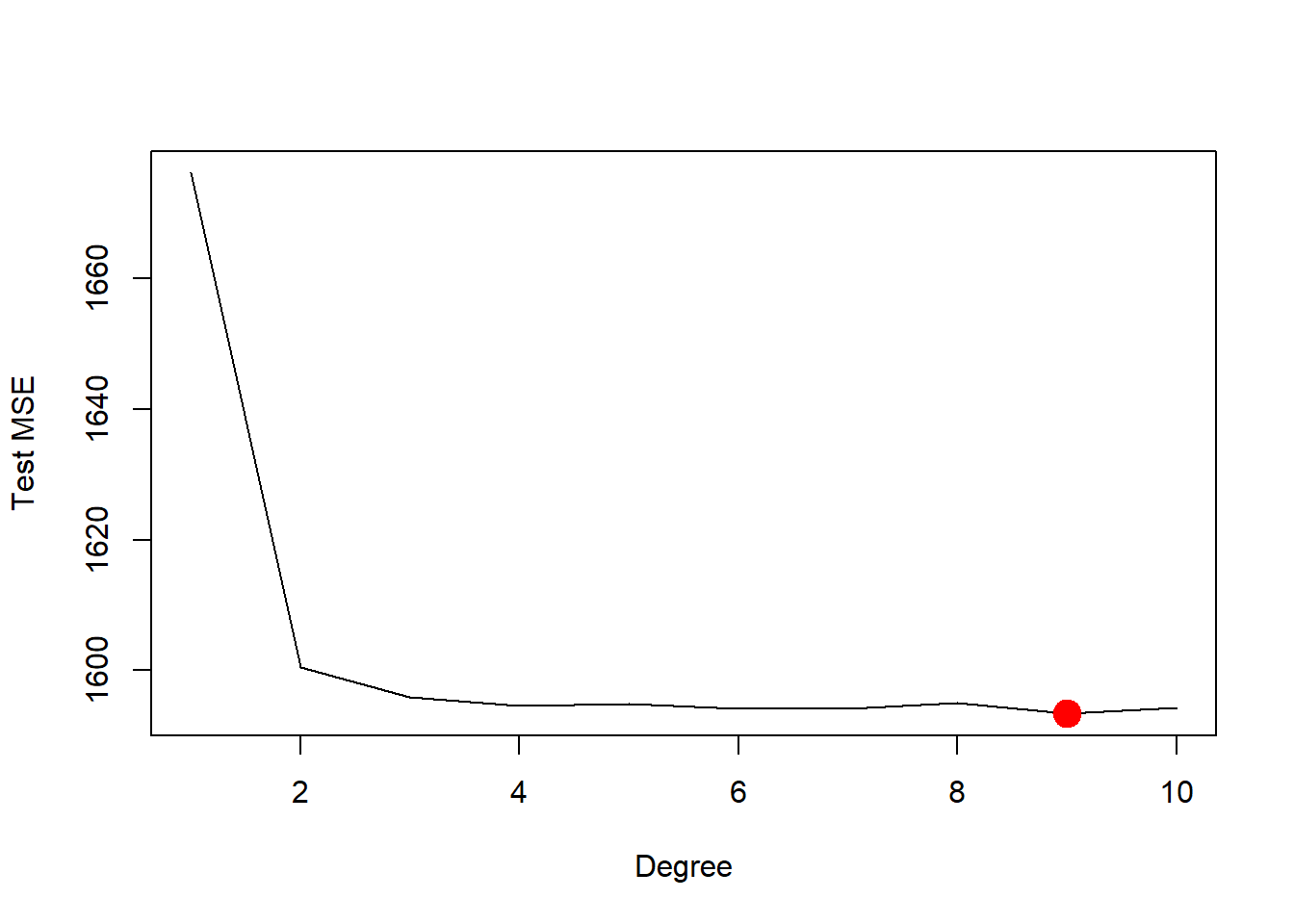
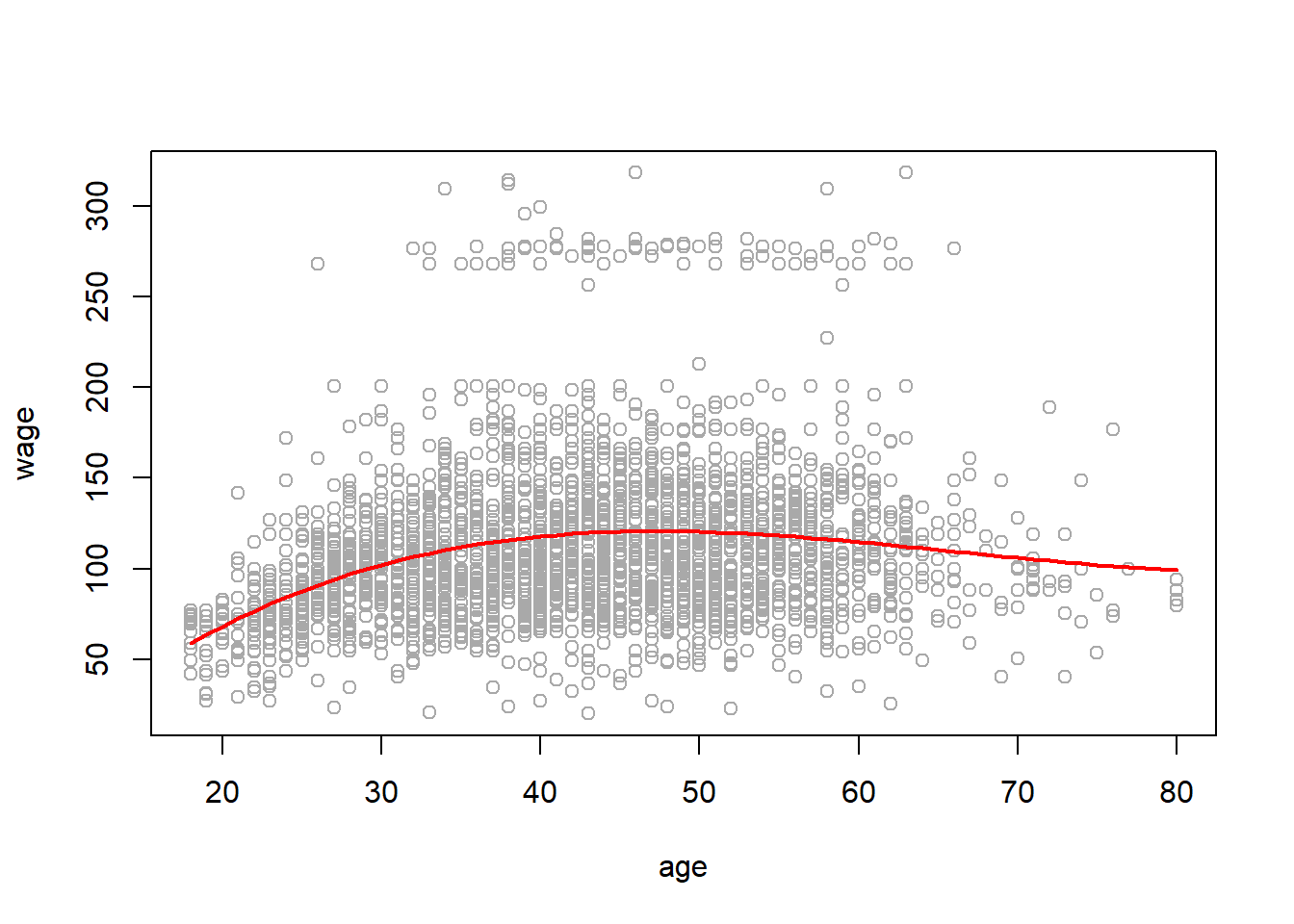
6a)



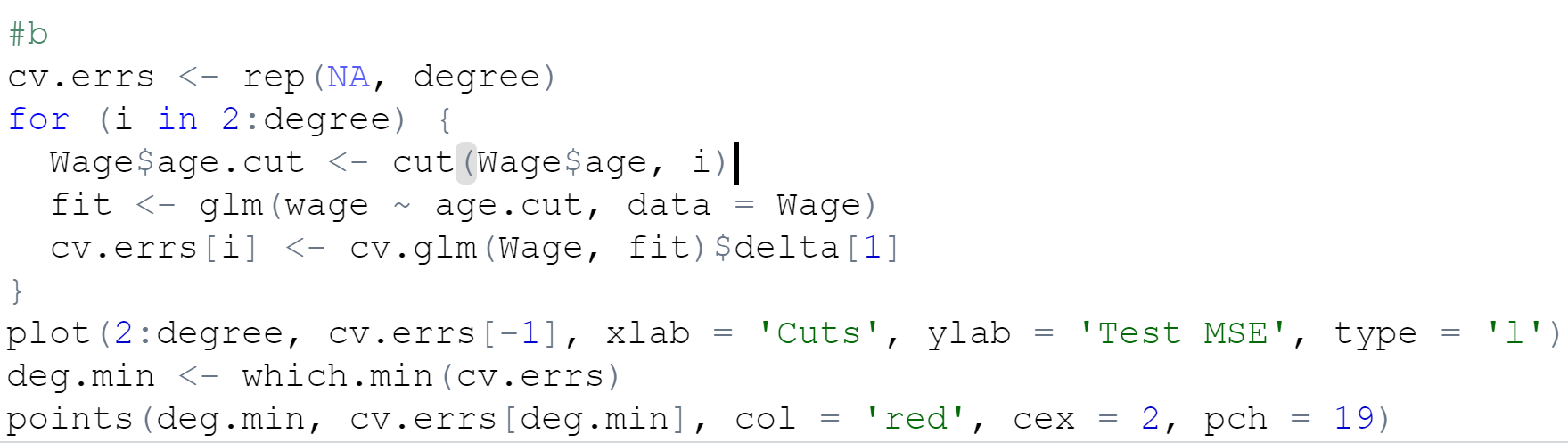


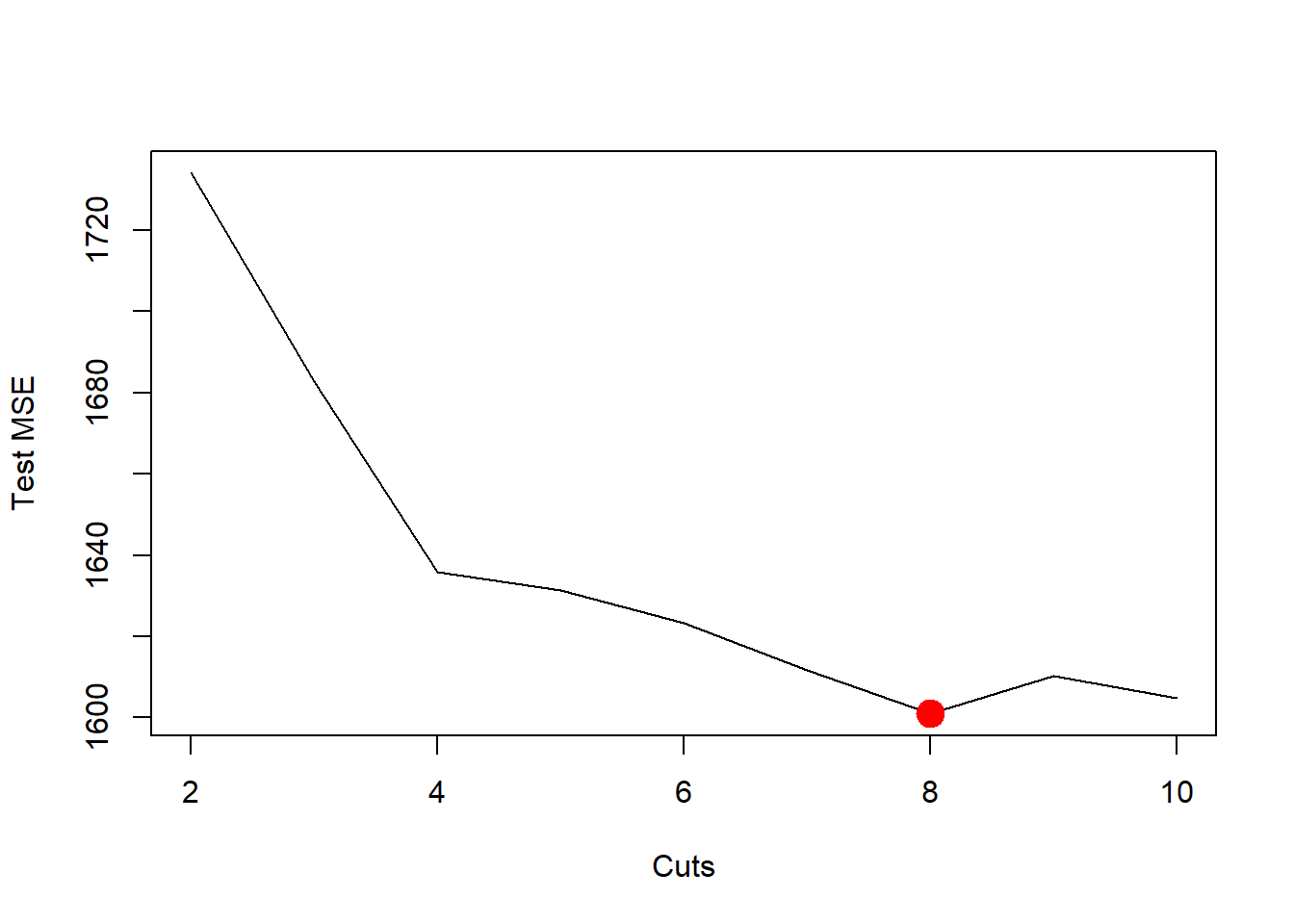
The minimum of test MSE is at degree 9. But test MSE of degree 4 is small enough. The comparison by ANOVA suggest degree 4 is enough.



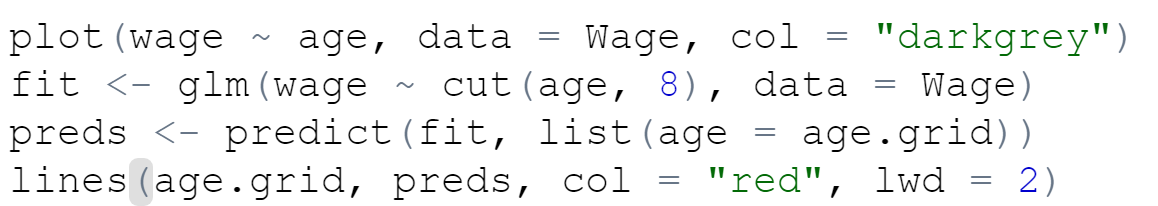
This graph shows that there is a plateau as wage increases, age increases.

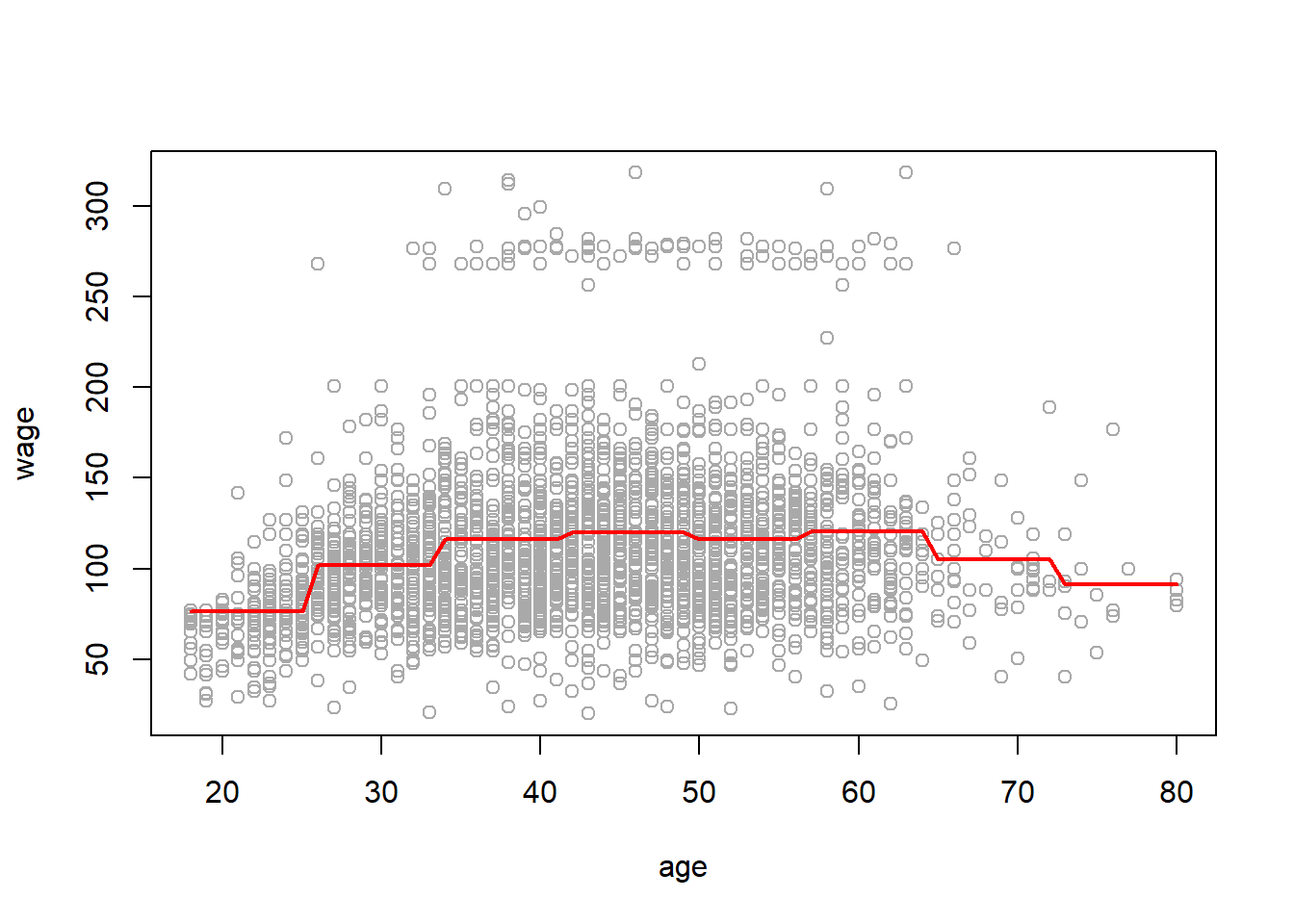
b)



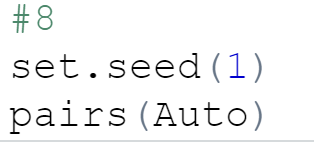


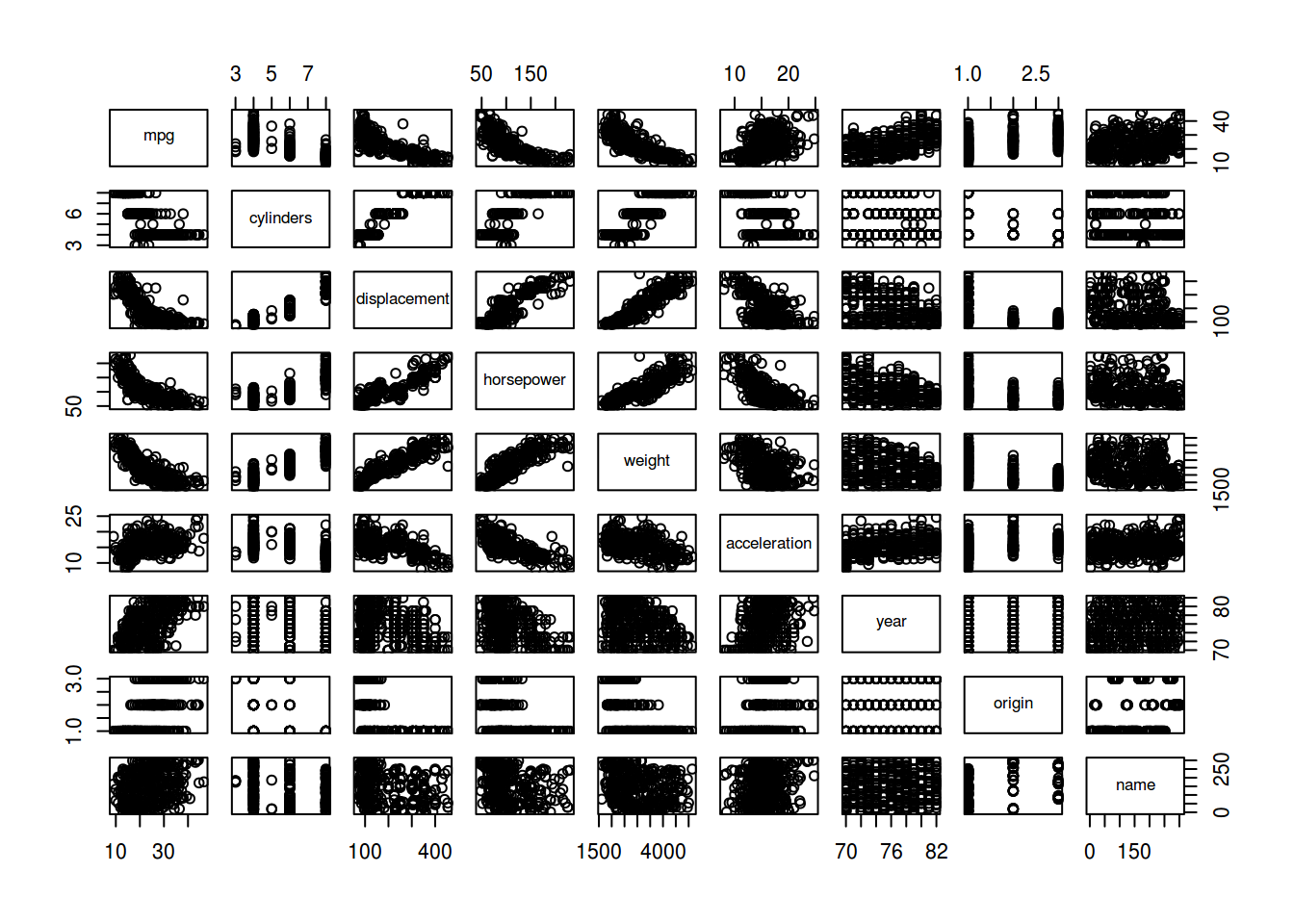
The model above shows that 8 cuts is the minimum test MSE.



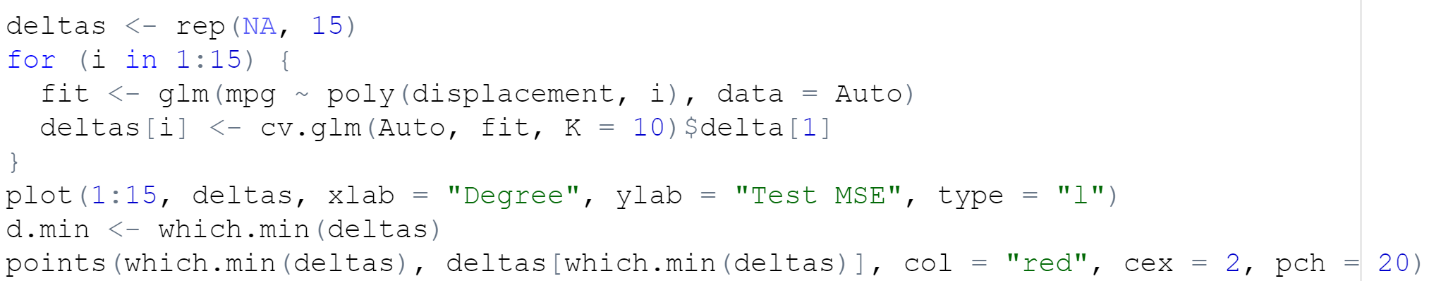


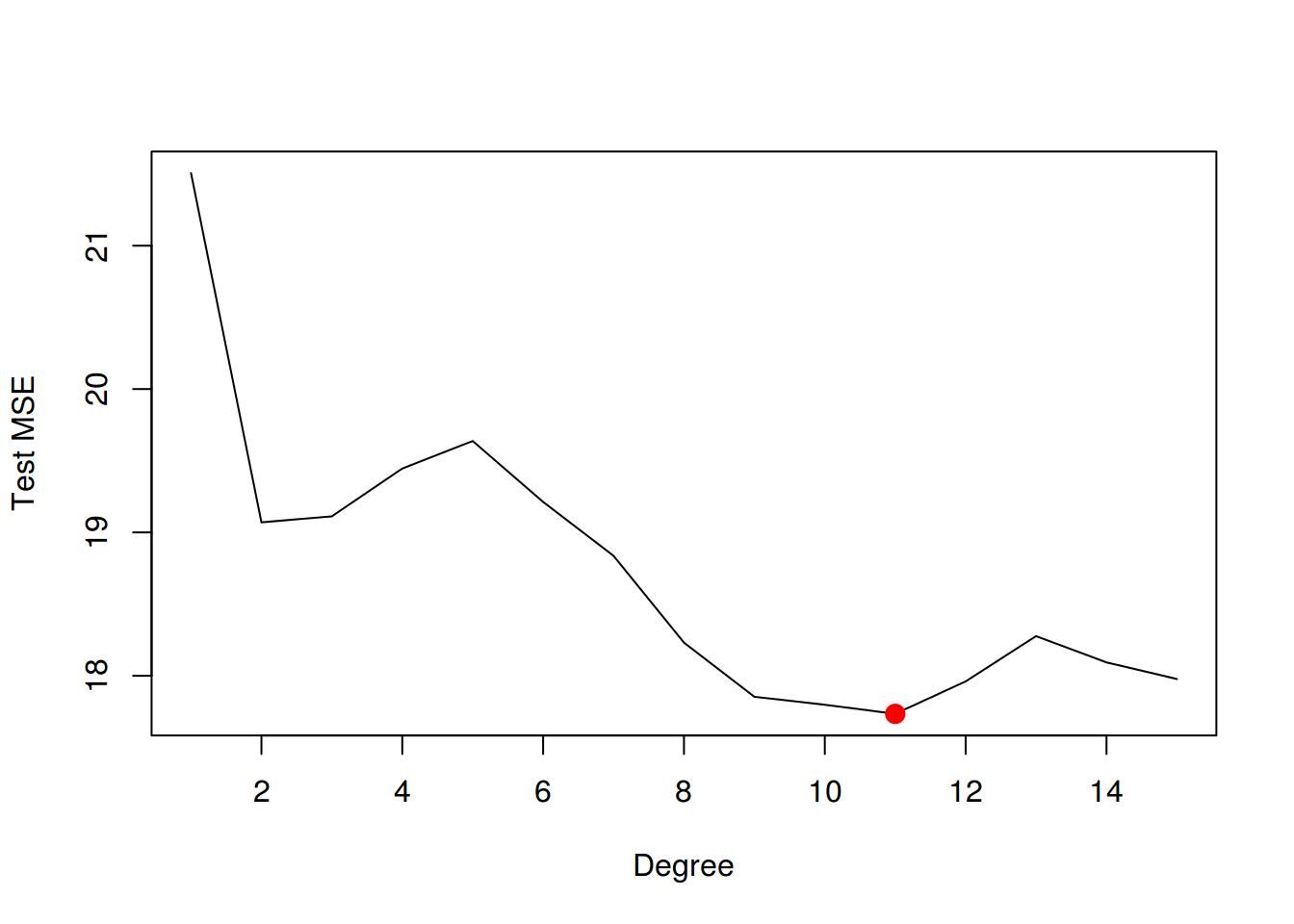
8)



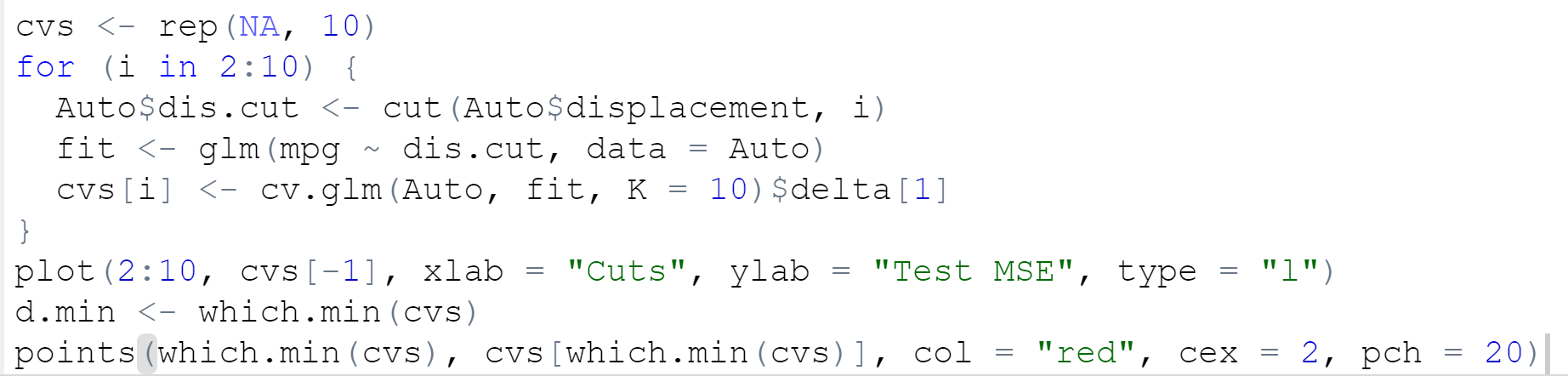


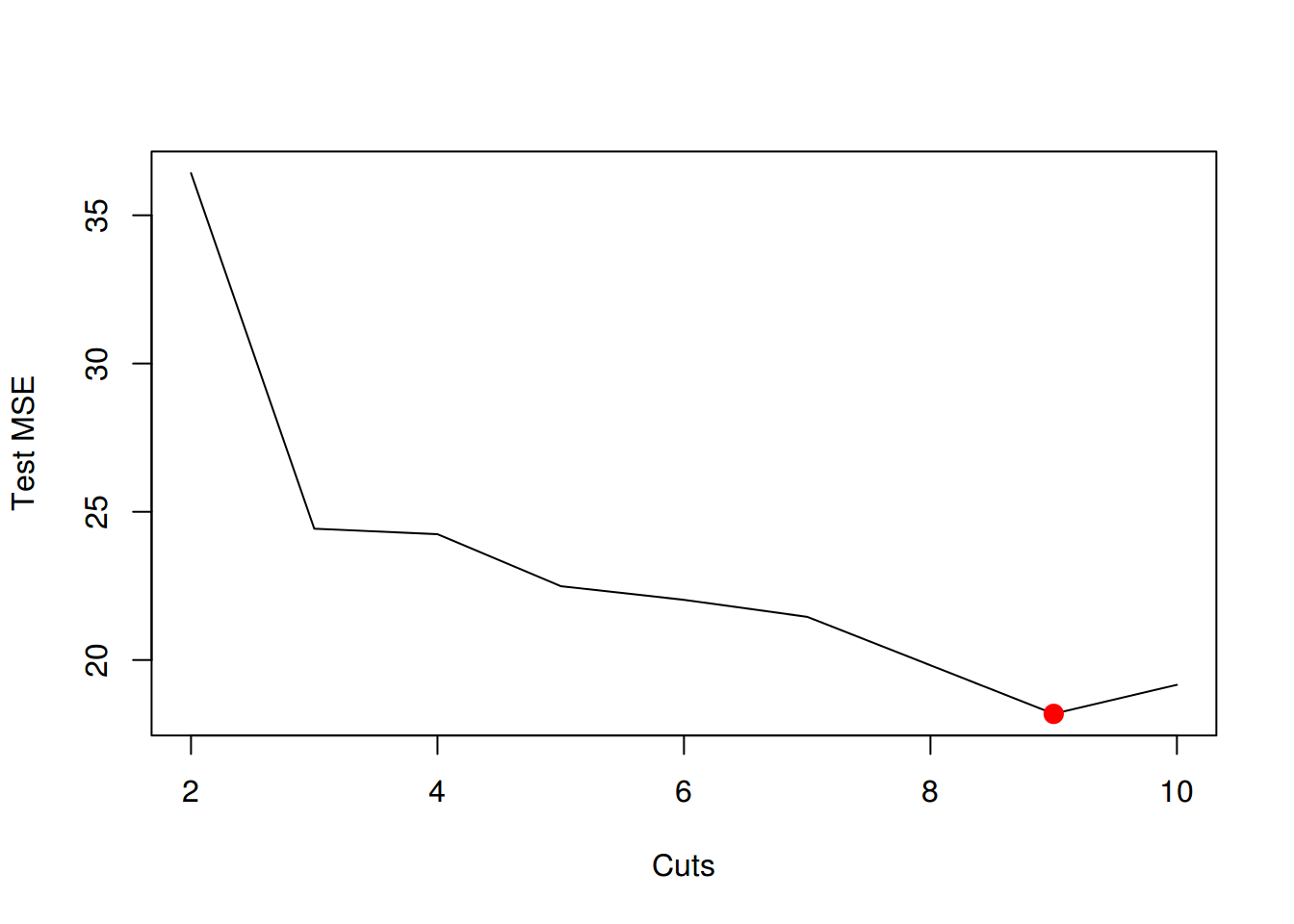
We may see that “mpg” is negatively correlated to “cylindes”, “displacement”, “horsepower” and “weight”. We begin by performing polynomial regression of “wage” vs “displacement”.



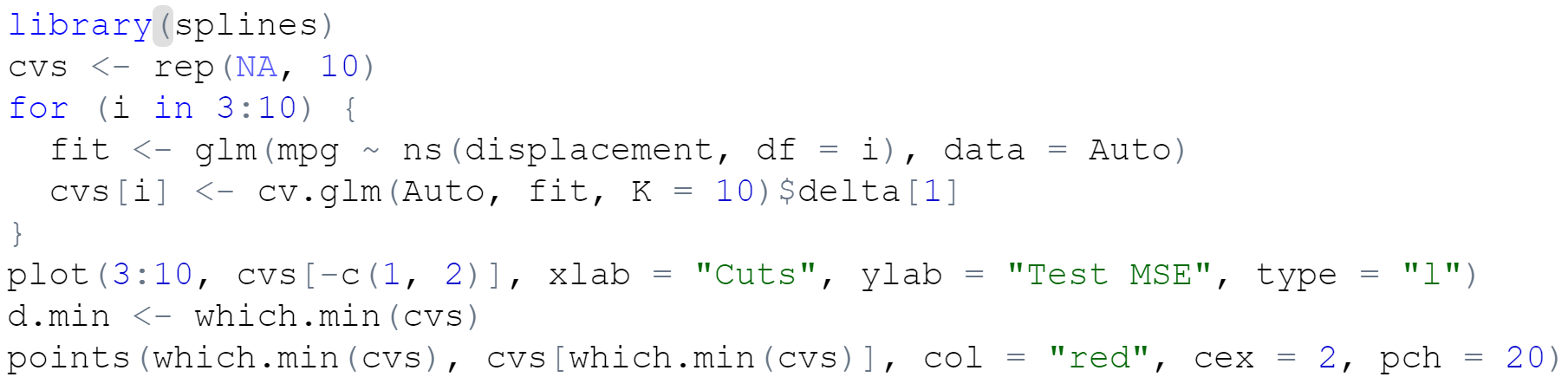


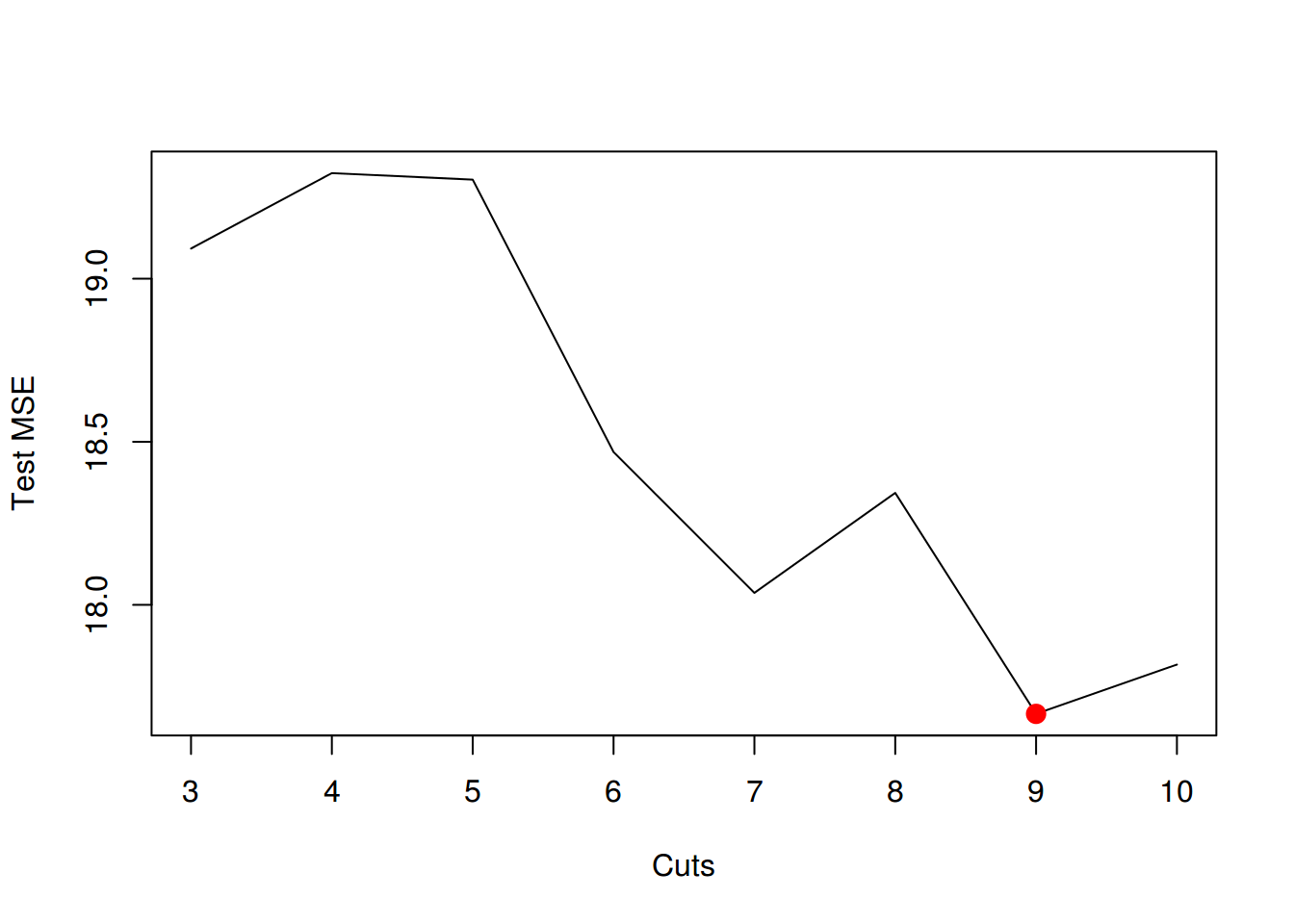
We may see that *d=11* is the optimal degree for the polynomial. Now we use step functions.



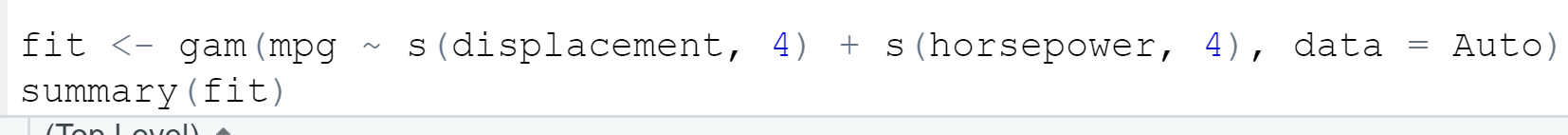


We may see that the error is minimum for 9 cuts.





We may see that the error is minimum for 9 degrees of freedom. It remains to use GAM.



Call: gam(formula = mpg ~ s(displacement, 4) + s(horsepower, 4), data = Auto)

Deviance Residuals:

Min 1Q Median 3Q Max

-11.2982 -2.1592 -0.4394 2.1247 17.0946

(Dispersion Parameter for gaussian family taken to be 15.3543)

Null Deviance: 23818.99 on 391 degrees of freedom

Residual Deviance: 5880.697 on 382.9999 degrees of freedom

AIC: 2194.05

Number of Local Scoring Iterations: 2

Anova for Parametric Effects

Df Sum Sq Mean Sq F value Pr(>F)

s(displacement, 4) 1 15254.9 15254.9 993.524 < 2e-16 \*\*\*

s(horsepower, 4) 1 1038.4 1038.4 67.632 3.1e-15 \*\*\*

Residuals 383 5880.7 15.4

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects

Npar Df Npar F Pr(F)

(Intercept)

s(displacement, 4) 3 13.613 1.863e-08 \*\*\*

s(horsepower, 4) 3 15.606 1.349e-09 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1