A "Testing-Based" Procedure for Model Selection in R

Ref. Chapter 10 in Faraway (2e) [Ch. 8 in 1st edn]

All the methods in the last lecture can be applied "manually" (and require only a little effort with relatively small datasets), but will very large datasets they can be automated.

Various approaches have been suggested, most of which initially tried to use model selection criteria discussed earlier (such as adjusted R^2), but this is an area of current research & is changing rapidly.

The stepwise() command from S-Plus was used in the example in the lecture notes, but this has not been implemented, as it has been superseded by approaches which use a set of "information criteria" (AIC, BIC, DIC, FIC); & the calculation of these measures is beyond the scope of this course.

One approach, currently available in R is the step() function, which uses Akaike's Information Criteria (AIC)

For ordinary linear models, AIC can be shown to be equivalent to Mallow's C_p (and both of them are only really valid for large samples). AIC is definitely measured on a different scale to Mallow's C_p, but we are still aiming for as small an AIC as possible (typically this will be a more "negative" value).