# Instructions

We are already familiar with our Toyota dataset. It contains data on used cars on sale during the late summer of 2004 in the Netherlands.

1. Make sure your Toyota.csv file is in the same folder as your Toyota.R file. Change the working directory use *Session -> Set Working Directory -> To Source File Location* from the Menu Bar.
2. Rerun the read.csv command to import the dataset “Toyota.csv”.
3. Move down to the end of the file, add “##########Assignment 5##########”.
4. Import libraries ‘rpart’ and ‘rpart.plot’
5. Select variables c(3,4,7,8,9,10,12,13,14,17,18) as we have done.
6. Categorize Price variable into a categorical variable with cutoff of 10000. (I will show how to do this in the video). (5 points)
7. Drop variable “Price”. (3 points)
8. Build a decision tree with all independent variables using rpart(). Do not specify any parameters. (5 points)
9. Use prp() to plot the tree. (5 points)
10. How many leaves does the tree have? (5 points)
11. What is the classification of the right most leaf? What is the IF-THEN rule to describe this leaf? (8 points)
12. What is the misclassification rate (error rate) of this tree? (5 points)
13. Now we can specify a complexity parameter (cp) to change the size of the tree. Use rpart() and specify “cp = 0.005” to build another tree. Use prp() to plot the tree. (5 points)
14. How many leaves does this tree have? What will be a potential problem of this tree? How can we address it? (9 points)