**Problem Statement:**

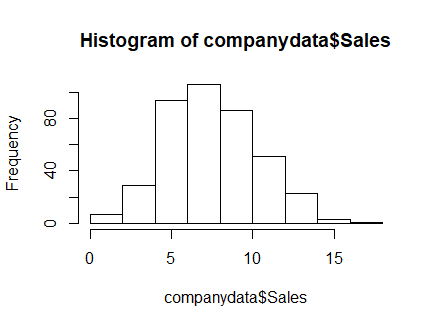
A cloth manufacturing company is interested to know about the segment or attributes contributing to high sale.

**Approach** - A decision tree can be built with target variable 'Sale' (we will first convert it into categorical variable) & all other variables will be independent in the analysis.

**Pre-processing:**

=>we will first convert it into categorical variable.

=>split the data into test and train data.



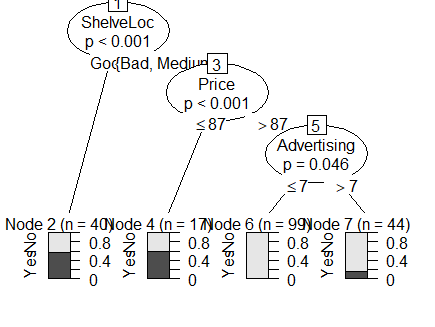
Using Party function:

op\_tree=ctree(high~CompPrice+Income+Advertising+Population+Price+ShelveLoc+Age+Education+Urban+US,data=cd\_train)

summary(op\_tree)

# Length Class Mode

# 1 Binary Tree S4



=> On looking into the Above tree, I see that if the Location of the Shelv is good, then there is a probability of 60% chance that the customer will buy.

=> With ShelveLoc having a Bad or Medium and Price <= 87, the probability of High sales could be 60%.

=> If ShelveLoc is Bad or Medium, With Price >= 87 and Advertising less then <= 7 then there is a zero percent chance of high sales.

=>If ShelveLoc is Bad or Medium, With Price >= 87 and Advertising less then > 7 then there is a 20 % percent chance of high sales.

=>mean(pred\_test\_df==CD$high)#accuracy

=>accuracy 0.6875=>68%

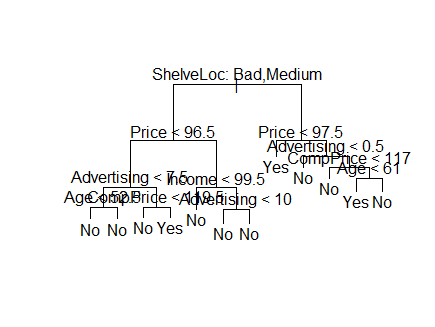
#Confusion Matrix and Statistics

#Prediction No Yes

# No 131 31

# Yes 10 28

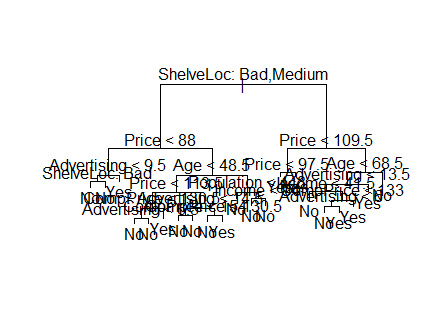
#Accuracy: 0.795



Number of terminal nodes: 12

Residual mean deviance: 0.2927 = 55.02 / 188

Misclassification error rate: 0.08 = 16 / 200



"""Confusion Matrix and Statistics

Reference

Prediction No Yes

No 153 9

Yes 19 19

Accuracy : 0.86