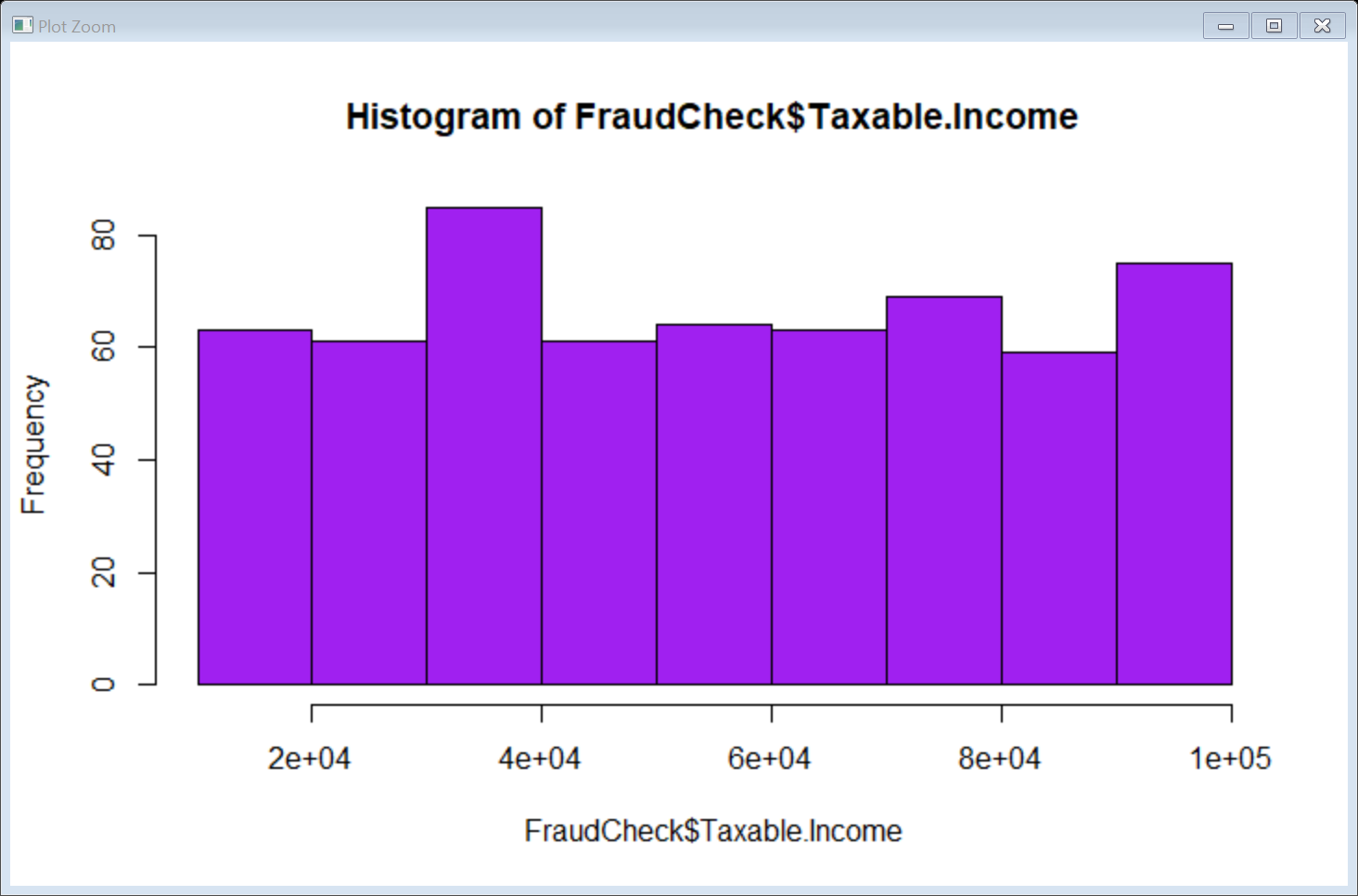
# Decision Tree

**Business Objective:** prepare a model on fraud data treating those who have taxable\_income <= 30000 as "Risky" and others are "Good

Step 1: Insall all the packages and read the file.

Step 2: histogram below shows the informatopn of taxable income for dataset fraudcheck.

****

Step 3:Using part function make a decision tree and summary of it is below

summary(opall\_tree)

Length Class Mode

1 BinaryTree S4

Step 4: using the training Data

summary(op\_tree)

Length Class Mode

1 BinaryTree S4

Step 5: Accuracy of the predict data is 0.82

Step 6: Crosstable of actual data with predicted

| pred\_test\_df

FC\_test$Risky\_Good | Good | Row Total |

-------------------|-----------|-----------|

Good | 246 | 246 |

| 0.820 | |

-------------------|-----------|-----------|

Risky | 54 | 54 |

| 0.180 | |

-------------------|-----------|-----------|

Column Total | 300 | 300 |

-------------------|-----------|-----------|

Step 7: Confusion Matrix and Statistics Accuracy =82%

Reference

Prediction Good Risky

Good 246 0

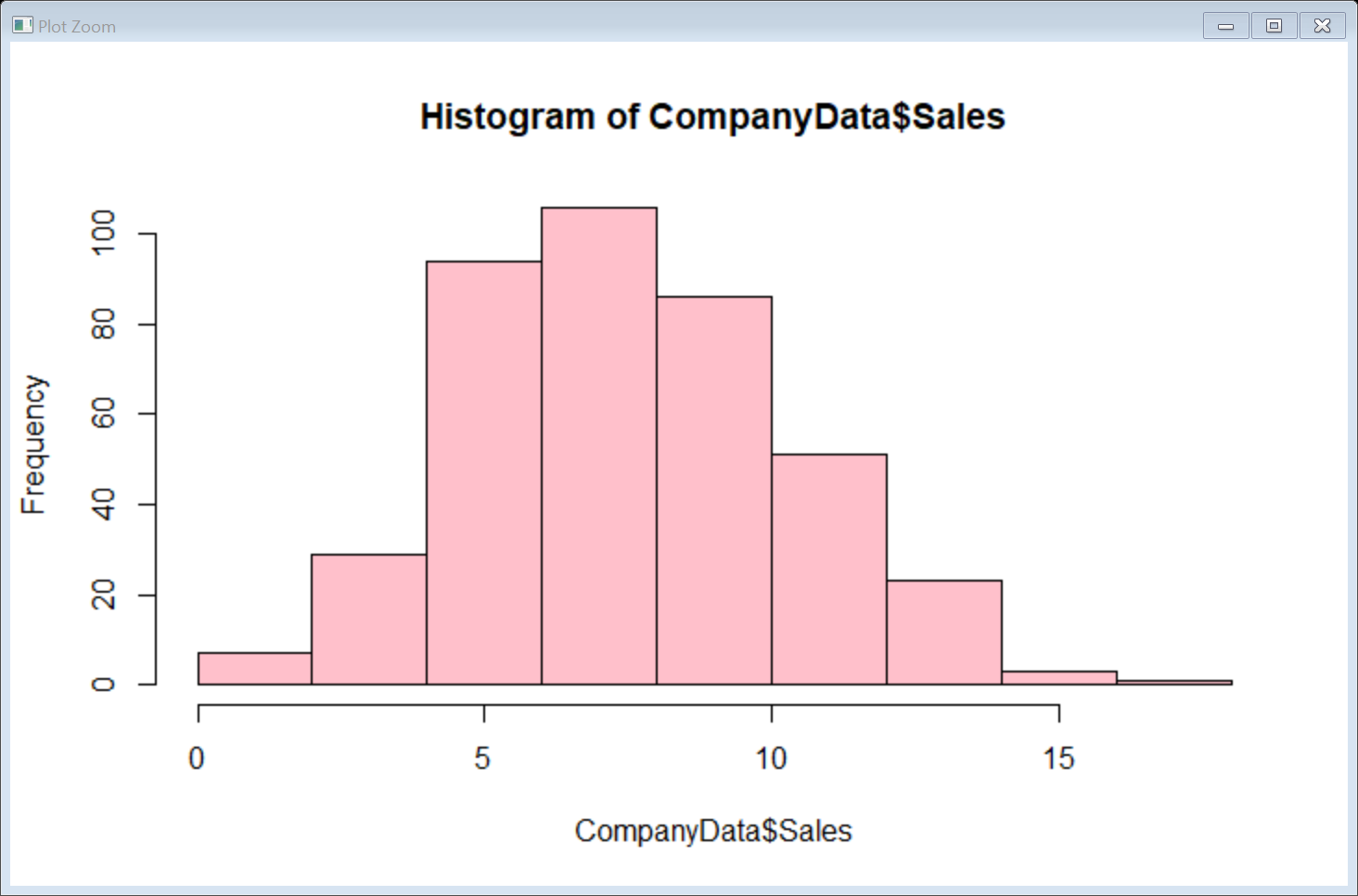
Risky 54 0

Conclusion: Accuracy =82% with Good = 246 Risky =0

**Business Objective**: A cloth manufacturing company is interested to know about the segment or attributes causes high sale.

Step 1:Install all the packages and read the file.

Step 2: Create a histogram to show the company data with sales.

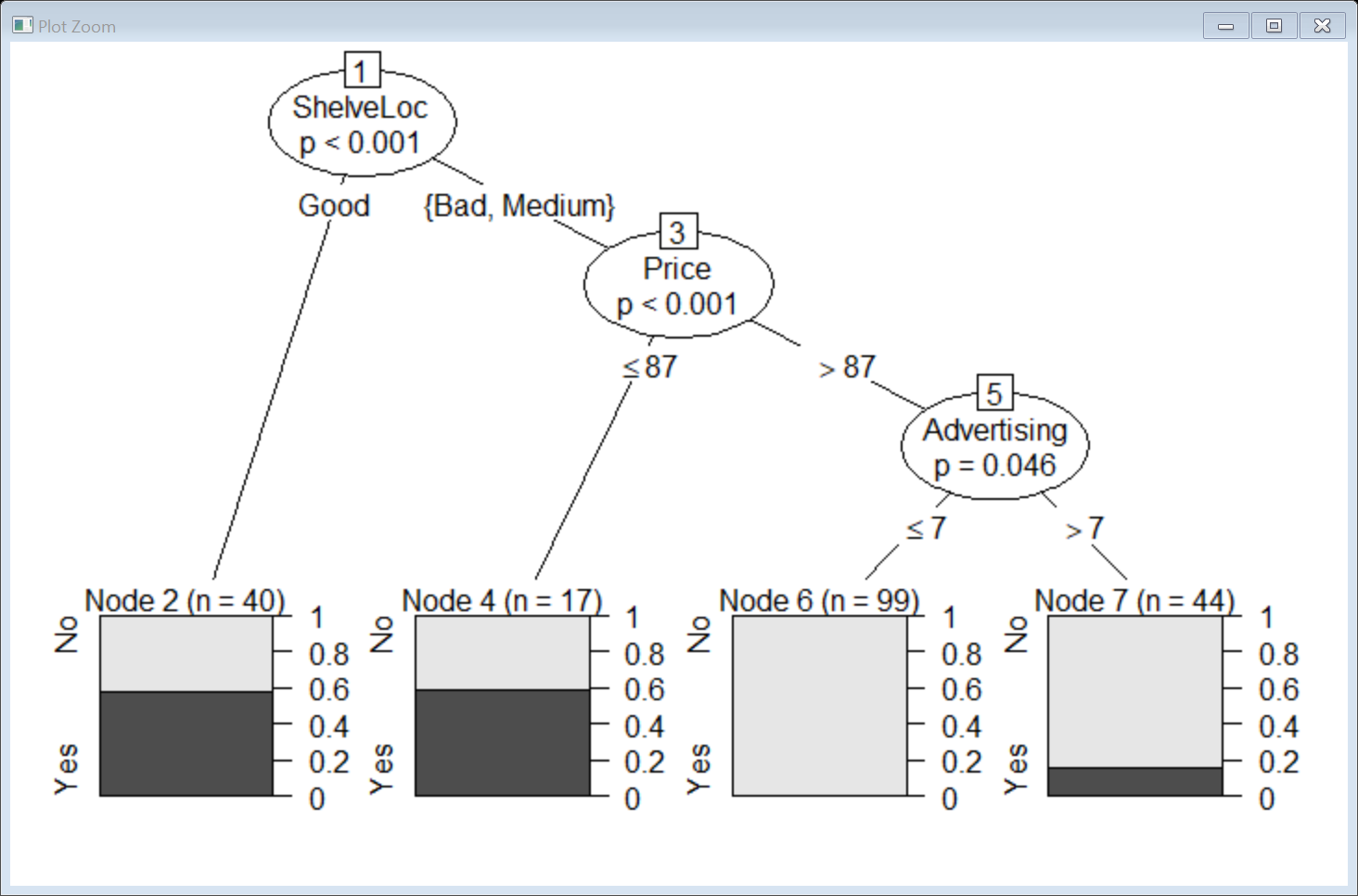


Step 3: Apply decision tree function on the sales output variable with respect to rest of the input variable.

Step 4: summary(op\_tree)

Length Class Mode

1 BinaryTree S4

Step 5:

Step 6:

Mean of predicted value is 0.6875

Step 7: | pred\_test\_df

CD\_test$High | No | Yes | Row Total |

-------------|-----------|-----------|-----------|

No | 131 | 31 | 162 |

| 2.468 | 5.899 | |

| 0.809 | 0.191 | 0.810 |

| 0.929 | 0.525 | |

| 0.655 | 0.155 | |

-------------|-----------|-----------|-----------|

Yes | 10 | 28 | 38 |

| 10.523 | 25.148 | |

| 0.263 | 0.737 | 0.190 |

| 0.071 | 0.475 | |

| 0.050 | 0.140 | |

-------------|-----------|-----------|-----------|

Column Total | 141 | 59 | 200 |

| 0.705 | 0.295 | |

-------------|-----------|-----------|-----------|

Step 8: Confusion Matrix and Statistics

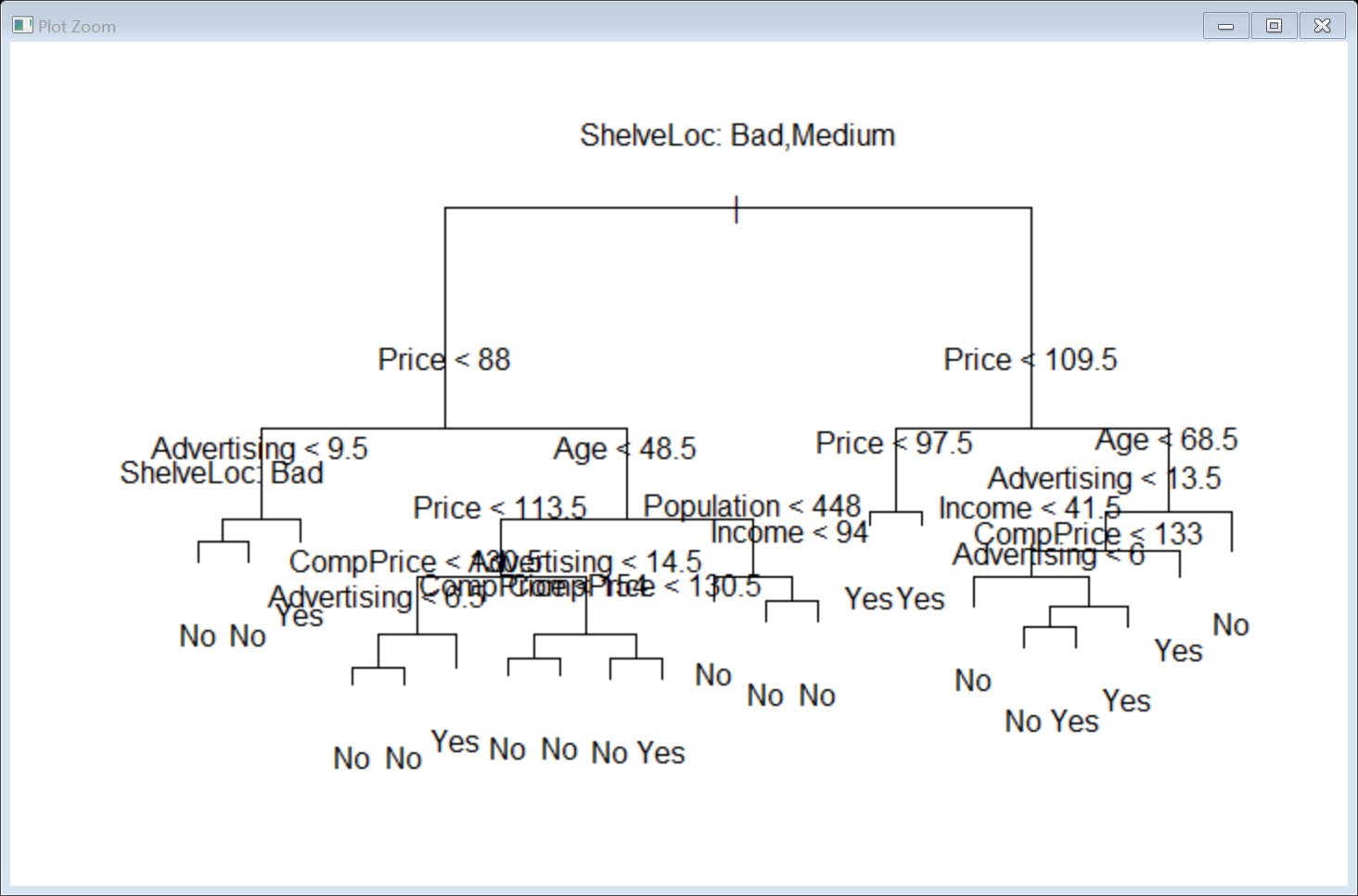
Reference

Prediction No Yes

No 131 31

Yes 10 28

Step 9:



Step 10: summary(pred\_tree$final)

No Yes

172 28

Step 11:Mean of train data 0.7725

Step 12: Test data

| pred\_tree$final

CD\_test$High | No | Yes | Row Total |

-------------|-----------|-----------|-----------|

No | 153 | 9 | 162 |

| 1.343 | 8.251 | |

| 0.944 | 0.056 | 0.810 |

| 0.890 | 0.321 | |

| 0.765 | 0.045 | |

-------------|-----------|-----------|-----------|

Yes | 19 | 19 | 38 |

| 5.727 | 35.177 | |

| 0.500 | 0.500 | 0.190 |

| 0.110 | 0.679 | |

| 0.095 | 0.095 | |

-------------|-----------|-----------|-----------|

Column Total | 172 | 28 | 200 |

| 0.860 | 0.140 | |

-------------|-----------|-----------|-----------|

Step 12:Test data

Confusion Matrix and Statistics

Reference

Prediction No Yes

No 153 9

Yes 19 19

Conclusion: Accuracy = 86%