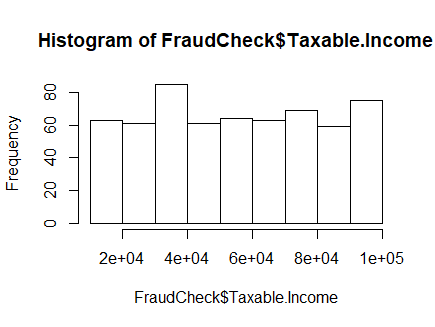
**Problem Statement:**

To Prepare a model on fraud data to check on the probability of Risky Vs Good. Risky patients -Taxable Income <= 30000.

Histogram:



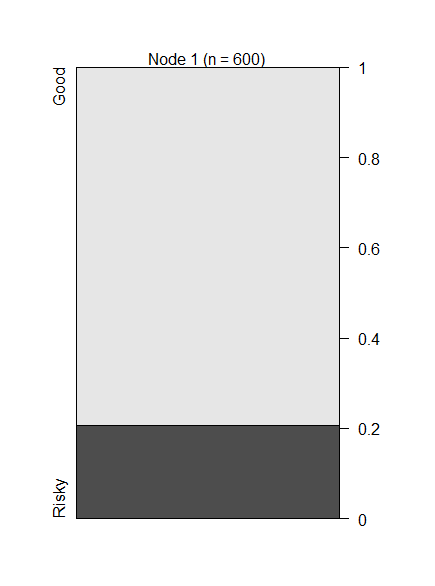
=>Split the data into train and test data

opall\_tree = ctree(Risky\_Good ~ Undergrad + Marital.Status + City.Population + Work.Experience + Urban, data = FC)

=>summary(opall\_tree)

Length Class Mode

1 Binary Tree S4



From the above tree, It looks like the data has 20 % of Risky patients and 80 % good patients.

Plot using training data:

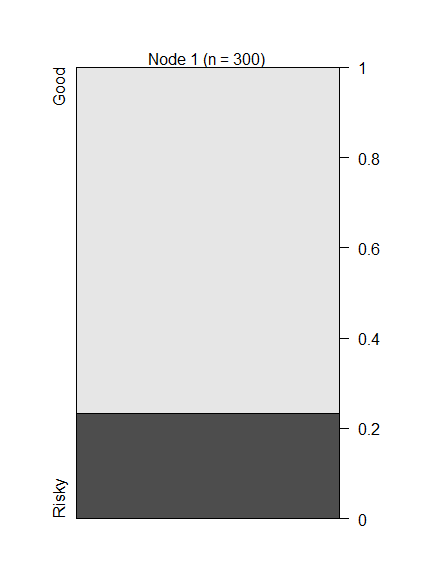
op\_tree = ctree(Risky\_Good ~ Undergrad + Marital.Status + City.Population + Work.Experience + Urban, data = FC\_train)

summary(op\_tree)

# Length Class Mode

# 1 BinaryTree S4

plot(op\_tree)



mean(pred\_test\_df==FC\_test$Risky\_Good)

# 0.82=>accuracy=82%

Confusion Matrix and Statistics

#Reference

#Prediction Good Risky

# Good 246 0

# Risky 54 0

#Accuracy : 0.82