R Notebook

\*\* I used the data set “earnings for three age groups” from \*\*<https://vincentarelbundock.github.io/Rdatasets/datasets.html>

\*\* the data included data from 1988-1989 of 4266 people and how much money they make in comparison to how old they are. I wanted to use this data to predict salary based on belonging to one of the three groups.

install.packages(“rpart”) install.packages(“rpart.plot”) library(rpart) library(rpart.plot) library(randomForest) library(partykit) library(caret) library(gdata) library(read.csv) mydata=read.csv(“Earnings.csv”) set.seed(1000) str(mydata) m1 <- rpart(age ~ ., data=mydata) m1 rpart.plot(m) predictions.ct <- predict(m1, mydata) plot(predictions.ct,mydataOzone)

**Random Forest** set.seed(8675309) ozone.rf <- randomForest(Ozone ~ ., data=mydata, mtry=3, importance=TRUE, na.action=na.omit) print(ozone.rf) round(importance(ozone.rf), 2) # Variable Importances varImpPlot(ozone.rf)

predictions.rf <- predict(ozone.rf, mydata) plot(predictions.rf,test\_data$Ozone)

\*\* My prediction is that the age group would be able to predict the range of earnings. In reality the curve via each age group was pretty consistent.