

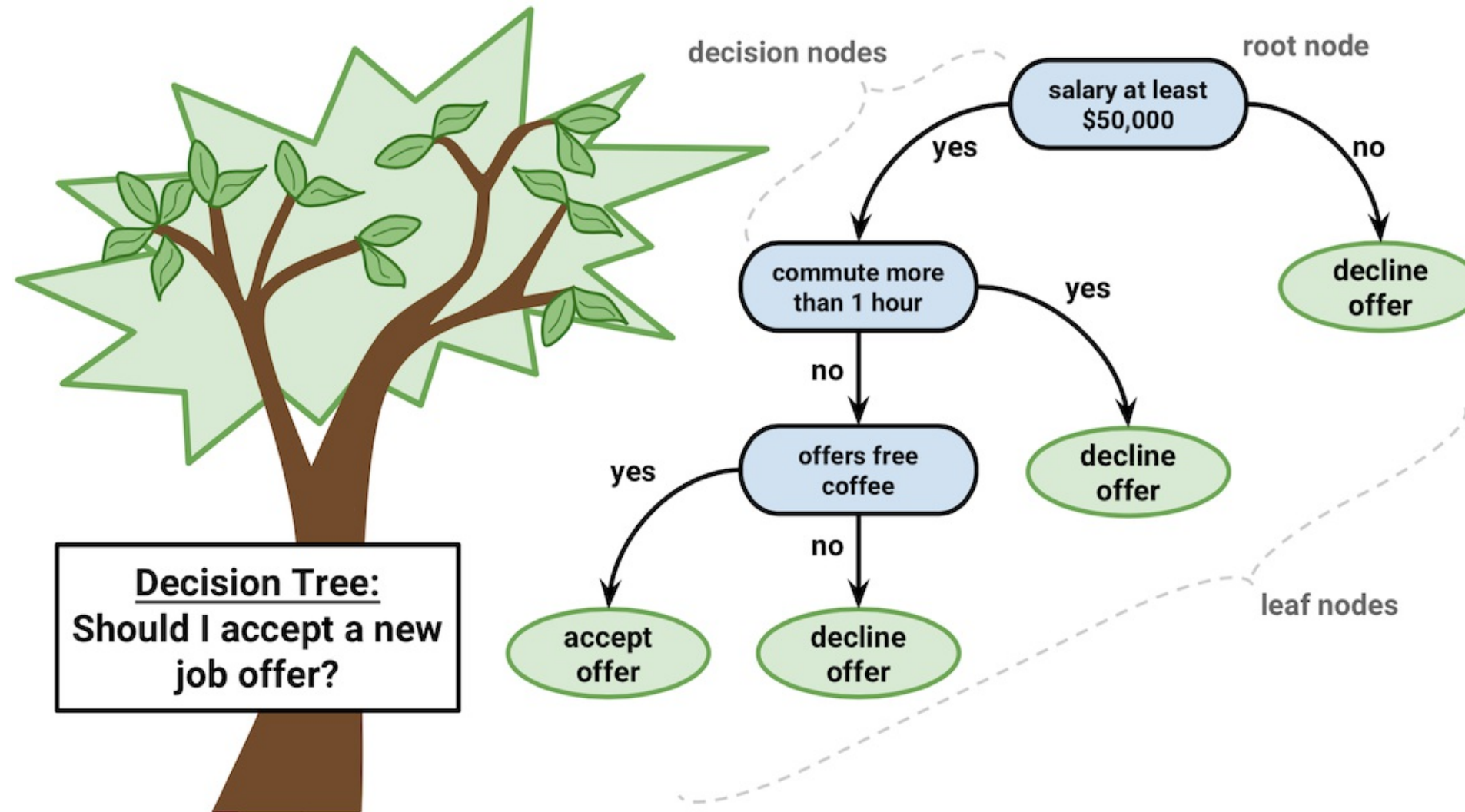


SUPERVISED LEARNING IN R: CLASSIFICATION

# Making decisions with trees

Brett Lantz  
Instructor

# A decision tree model





# Decision trees for prediction



## Check Your Rate

Get a custom rate for your **\$35,000** loan in **1 click**

First Name

Last Name

Street Address

City

State

Choose One

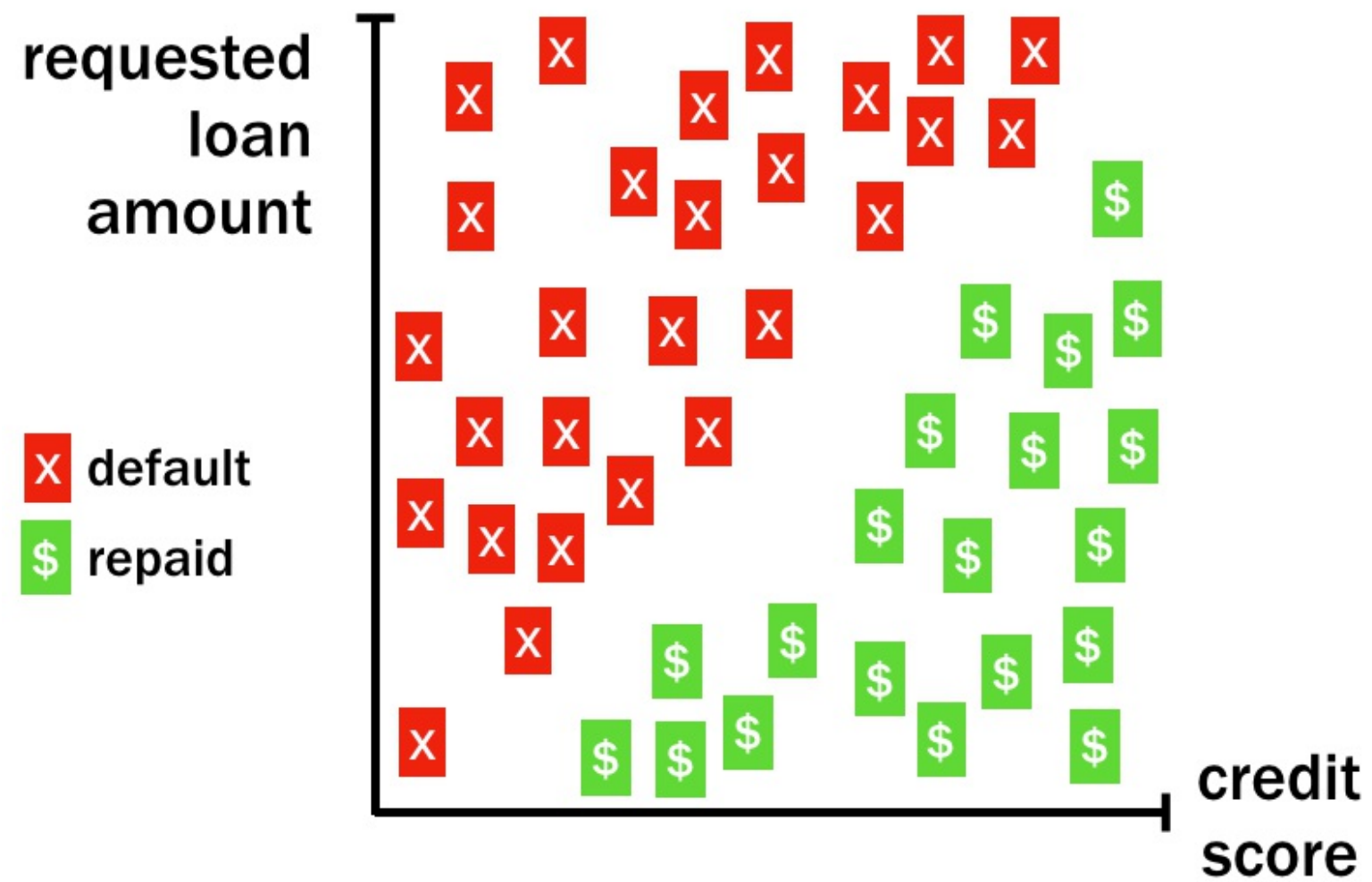
Zip Code

Date of Birth

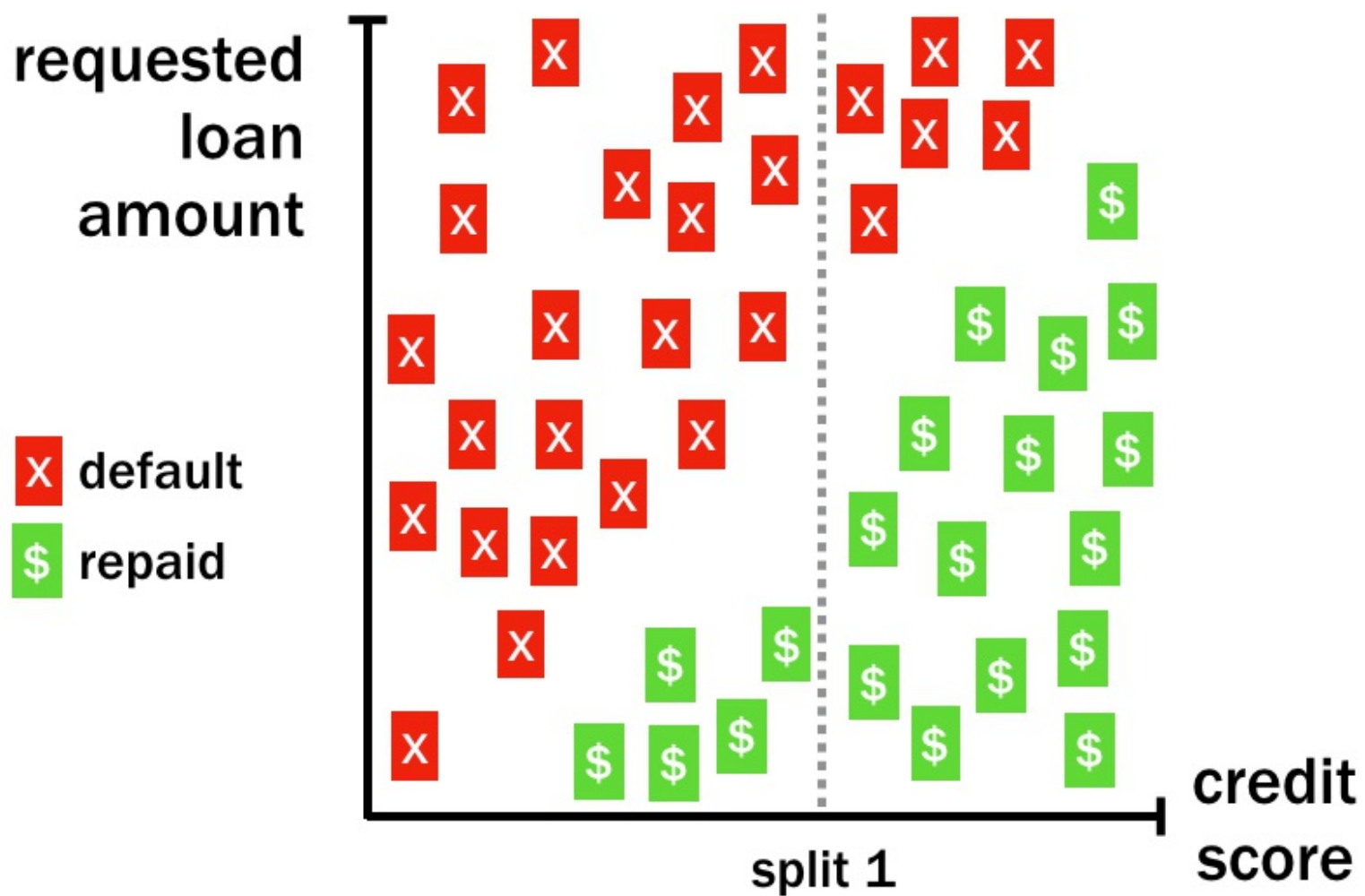
Month Day Year



# Divide-and-conquer

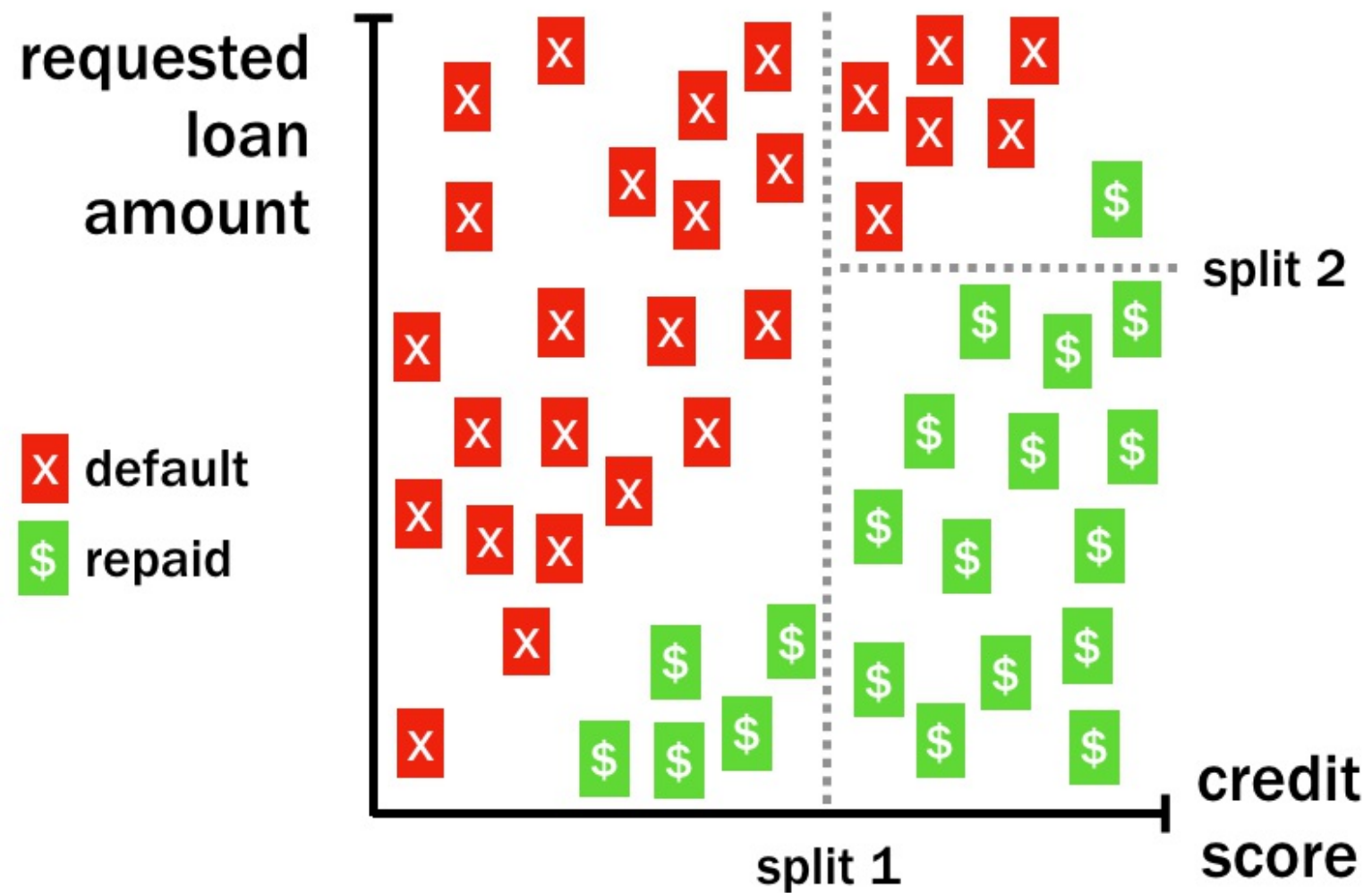


# Divide-and-conquer





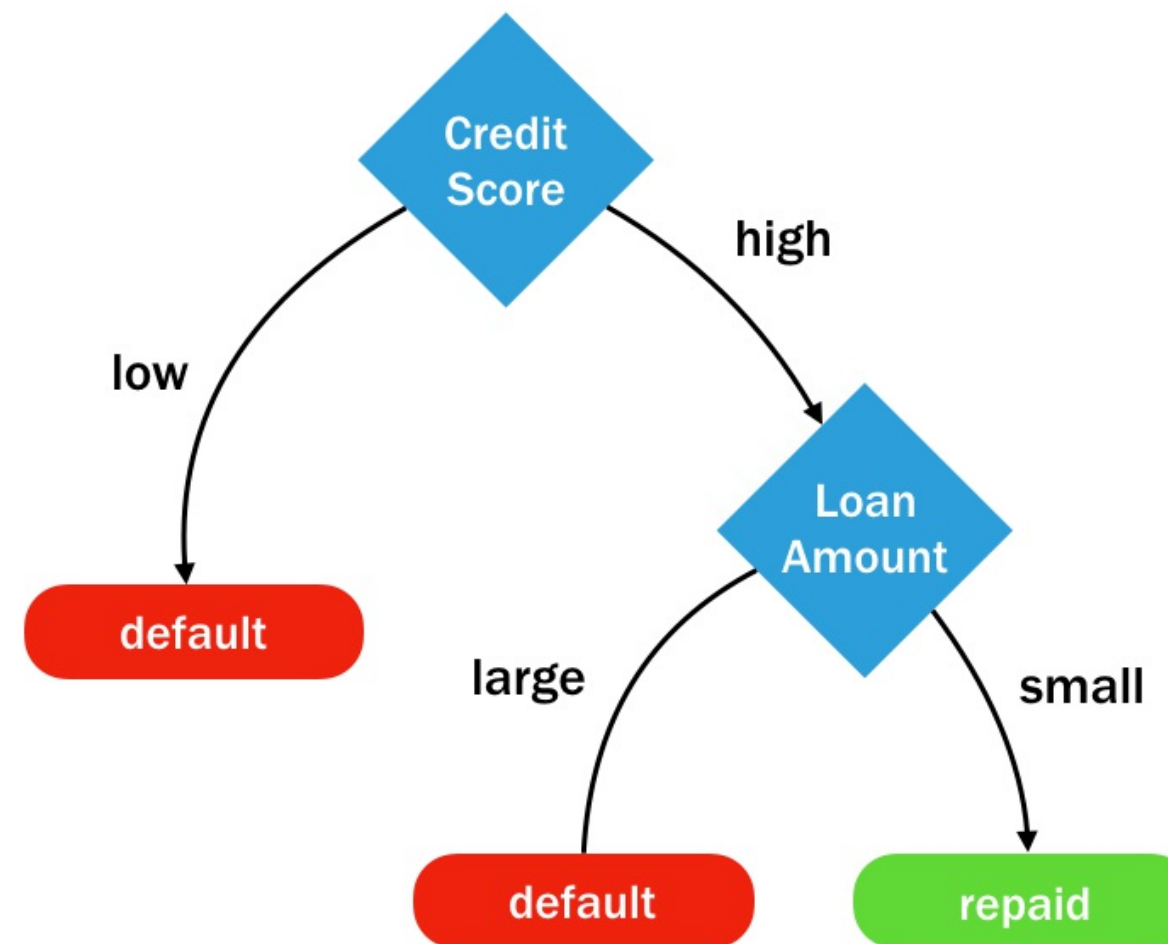
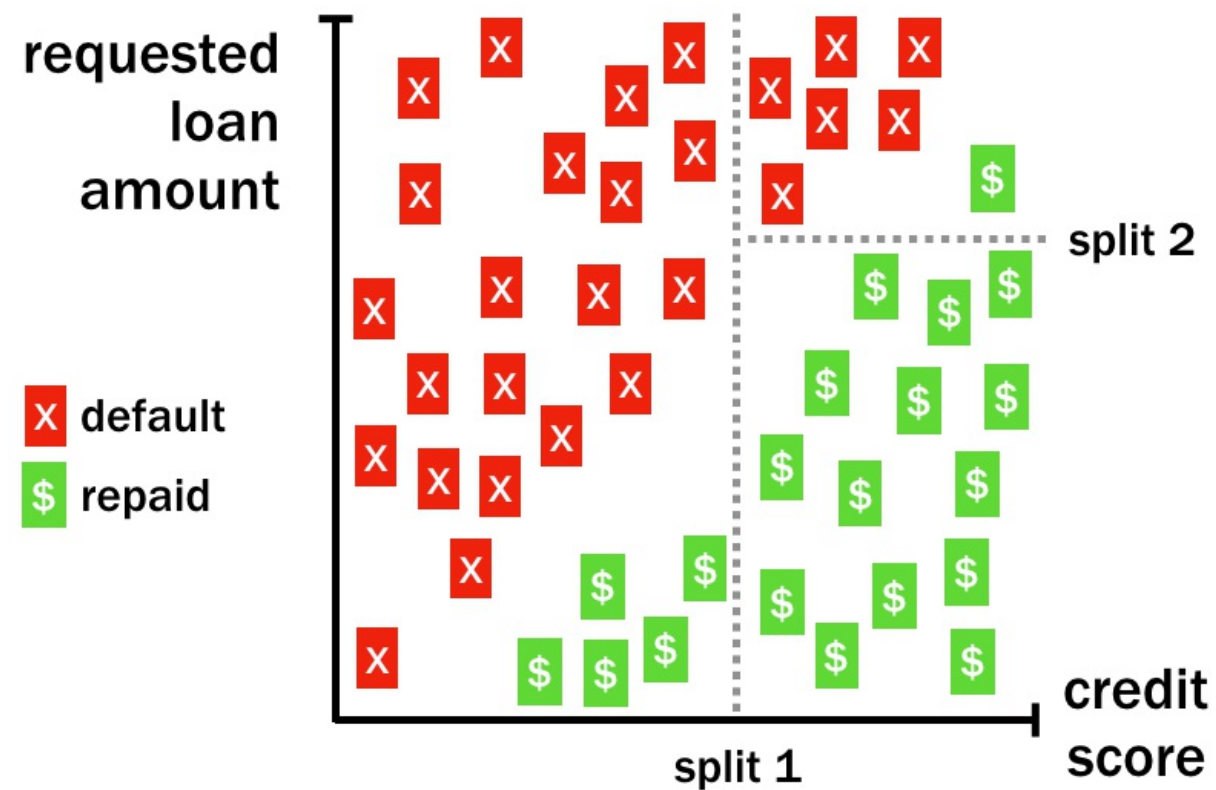
# Divide-and-conquer







# The resulting tree





# Building trees in R

```
# building a simple rpart classification tree
library(rpart)
m <- rpart(outcome ~ loan_amount + credit_score, data = loans,
           method = "class")
```

```
# making predictions from an rpart tree
p <- predict(m, test_data, type = "class")
```





## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**



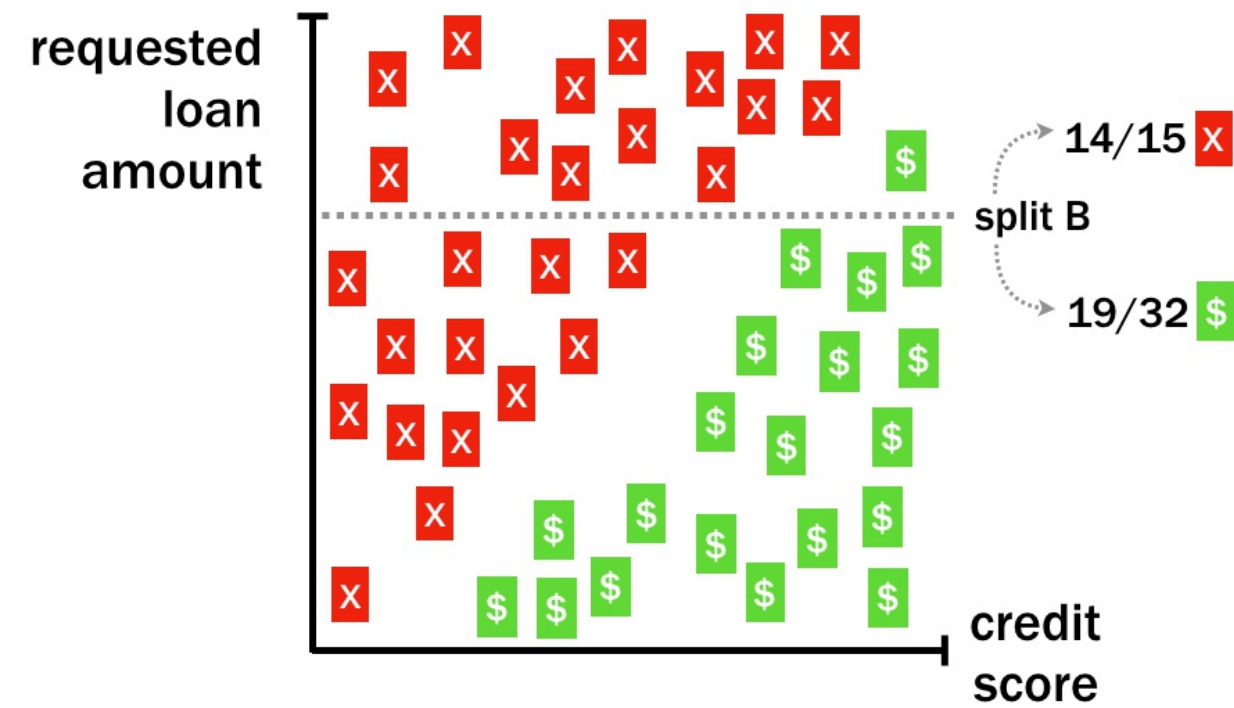
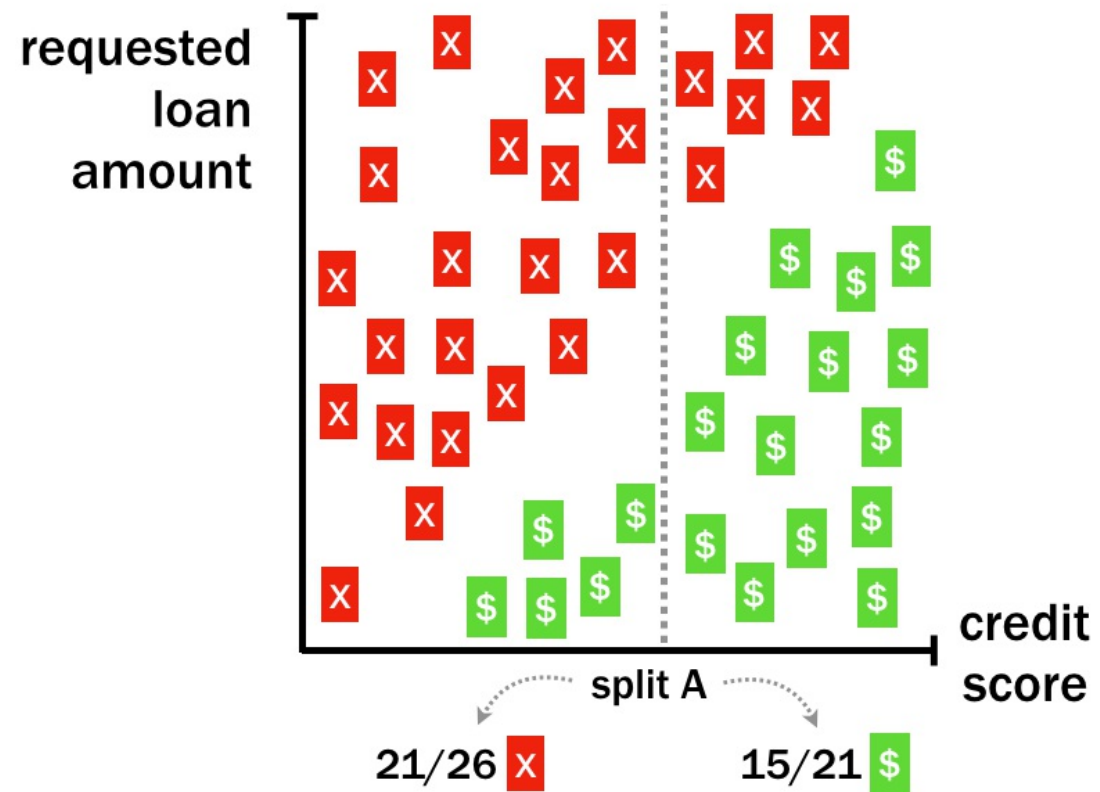
## SUPERVISED LEARNING IN R: CLASSIFICATION

# Growing larger classification trees

Brett Lantz  
Instructor

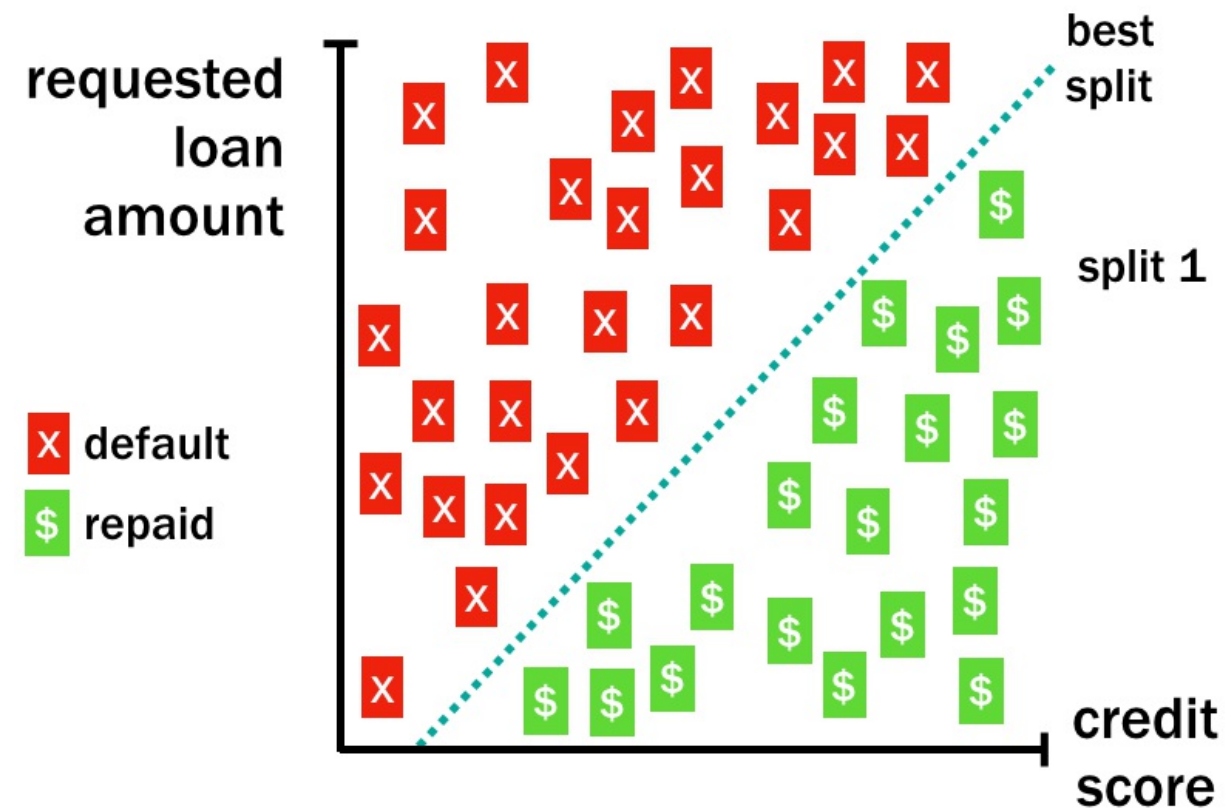
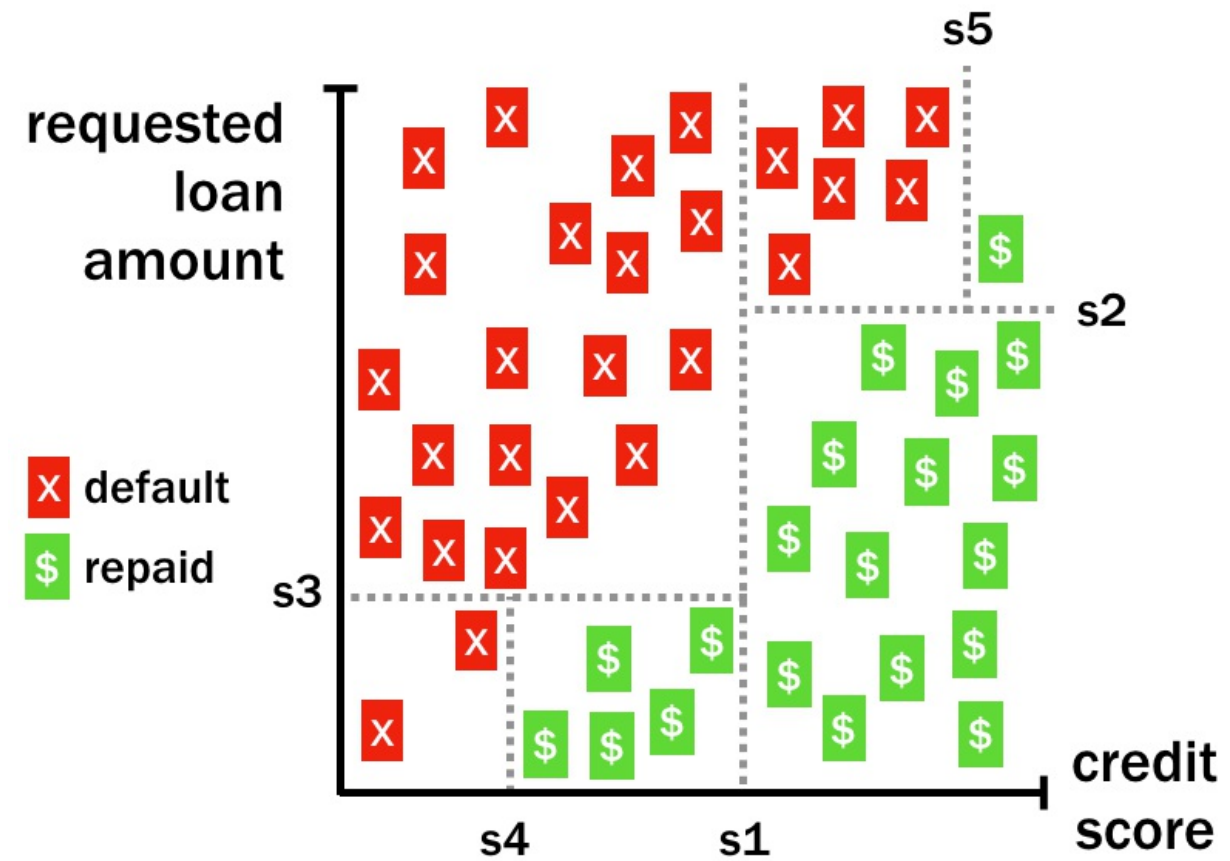


# Choosing where to split



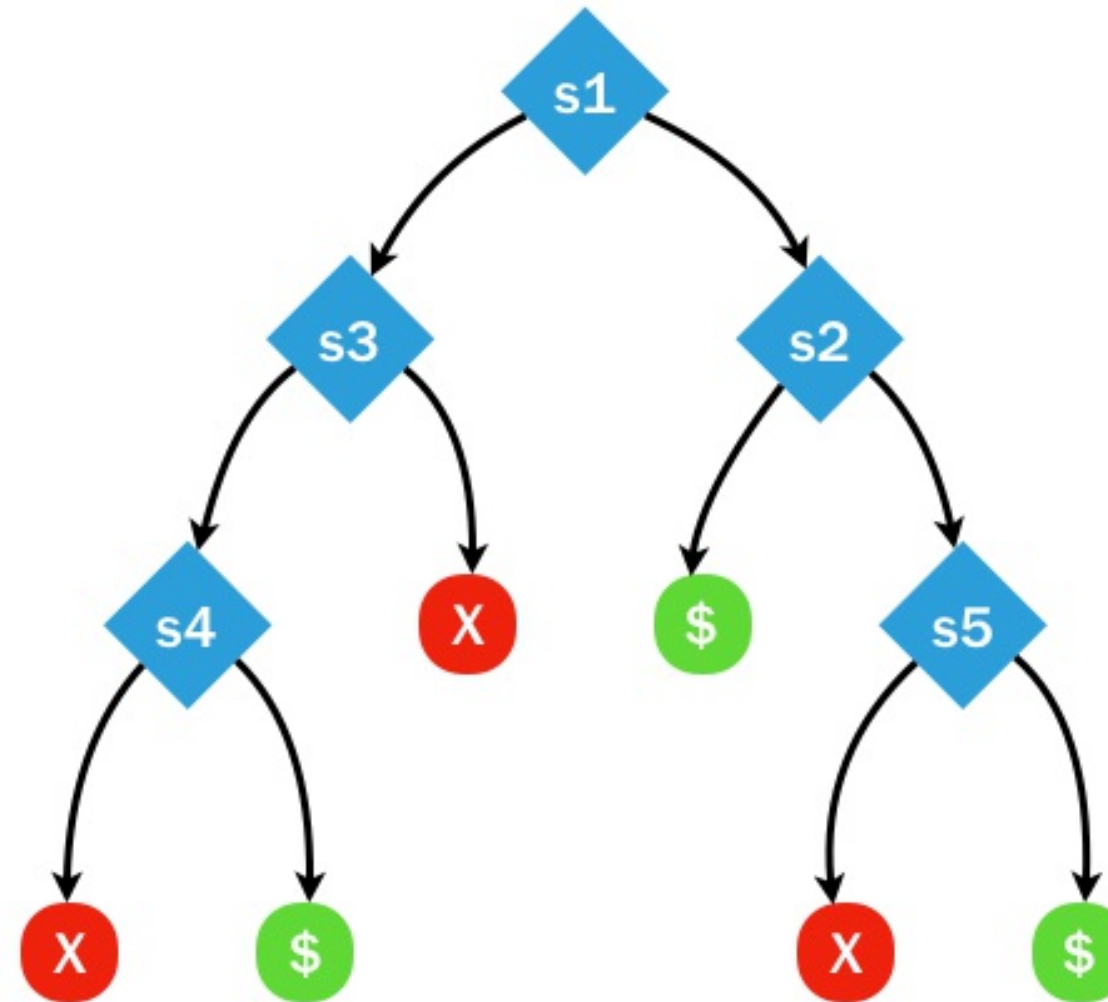
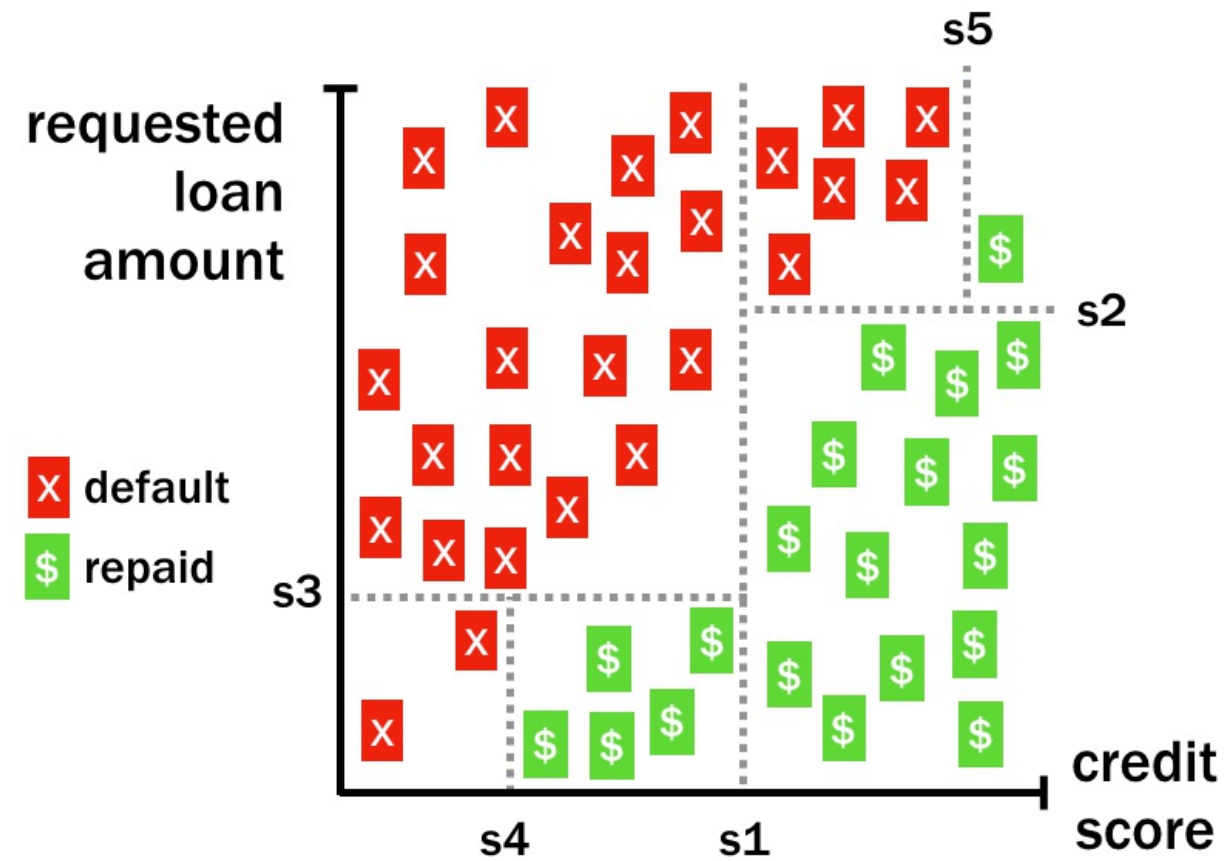


# Axis-parallel splits

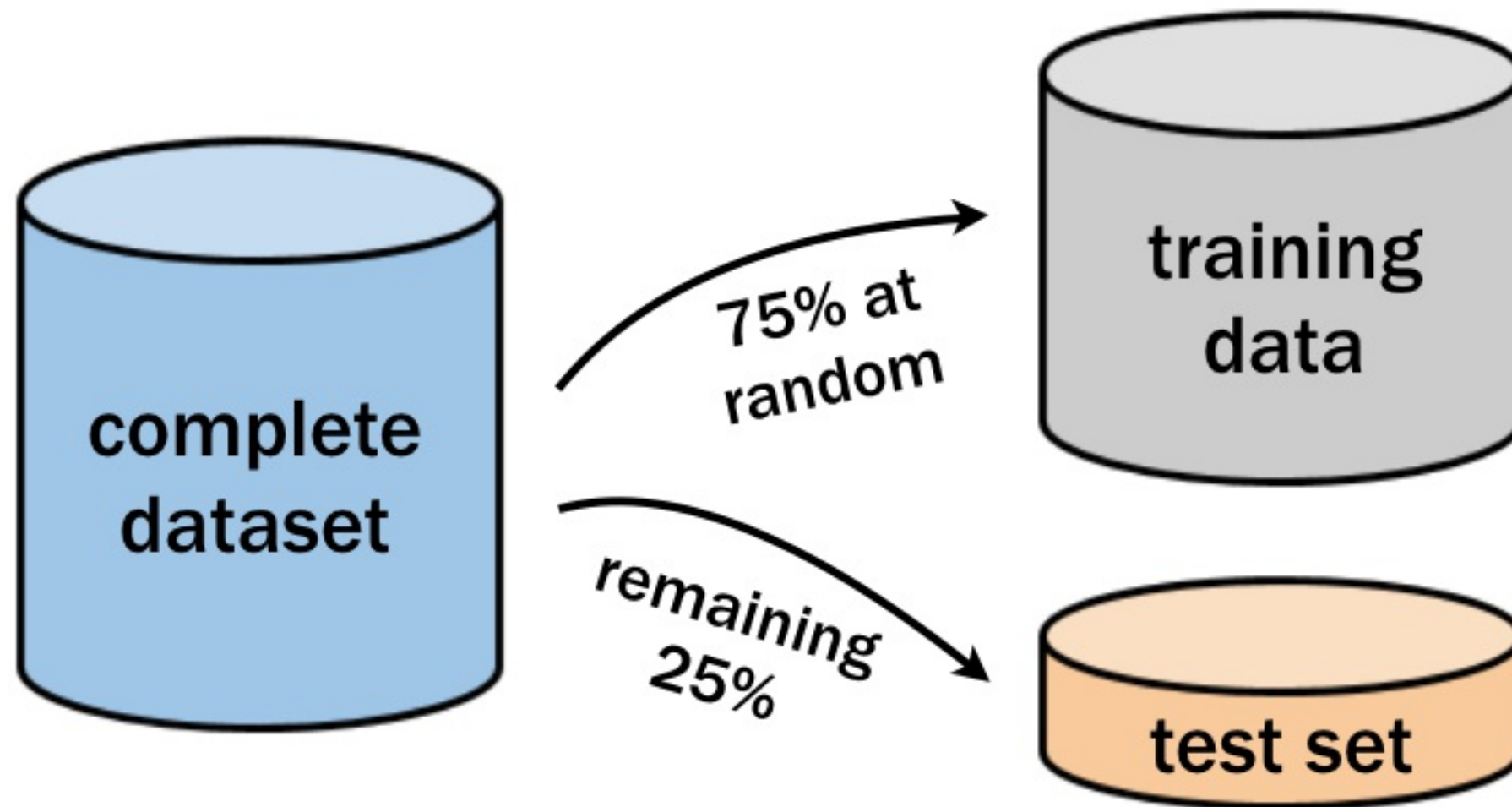




# The problem of overfitting



# Evaluating model performance





## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**





## SUPERVISED LEARNING IN R: CLASSIFICATION

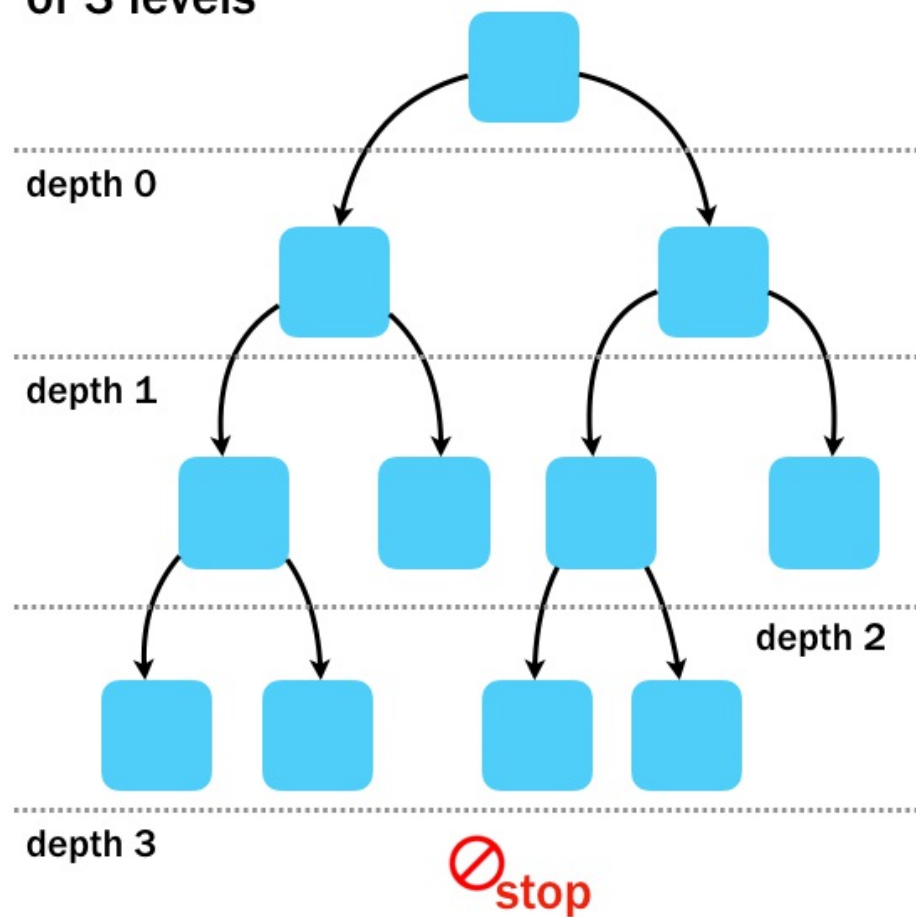
# Tending to classification trees

Brett Lantz  
Instructor

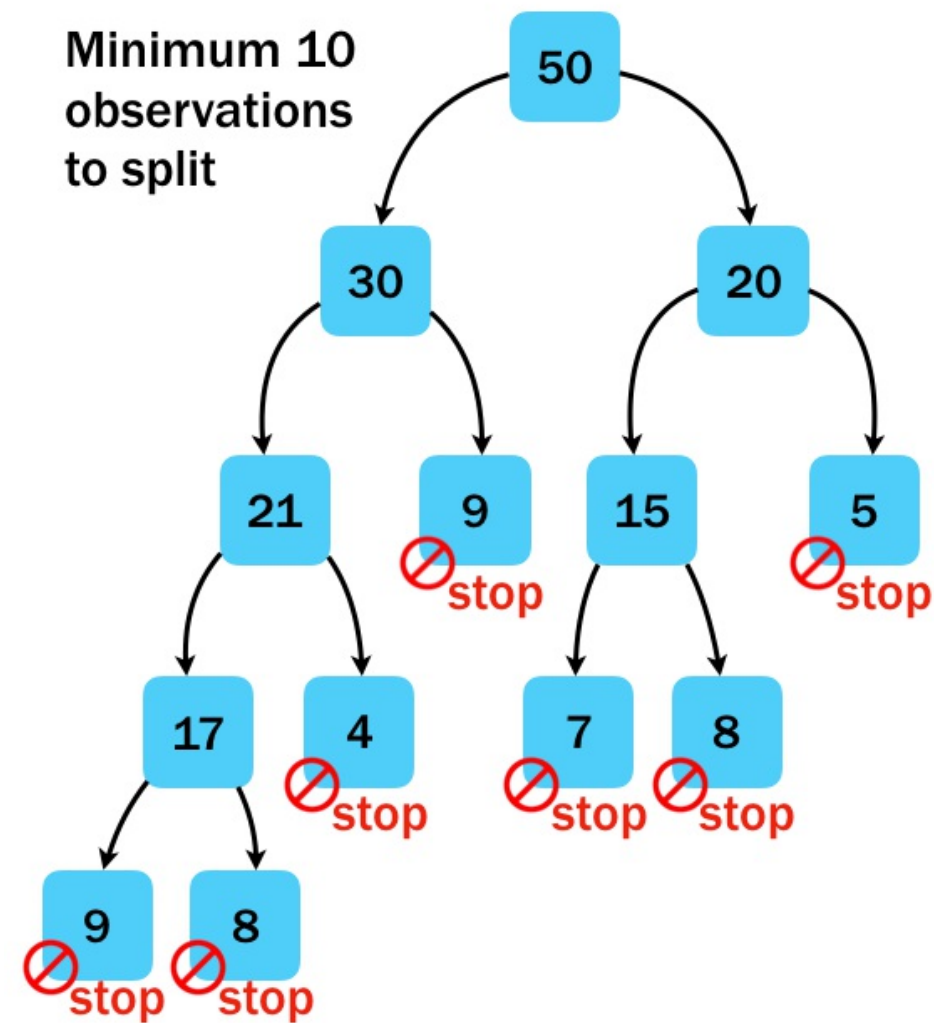


# Pre-pruning

Maximum depth  
of 3 levels

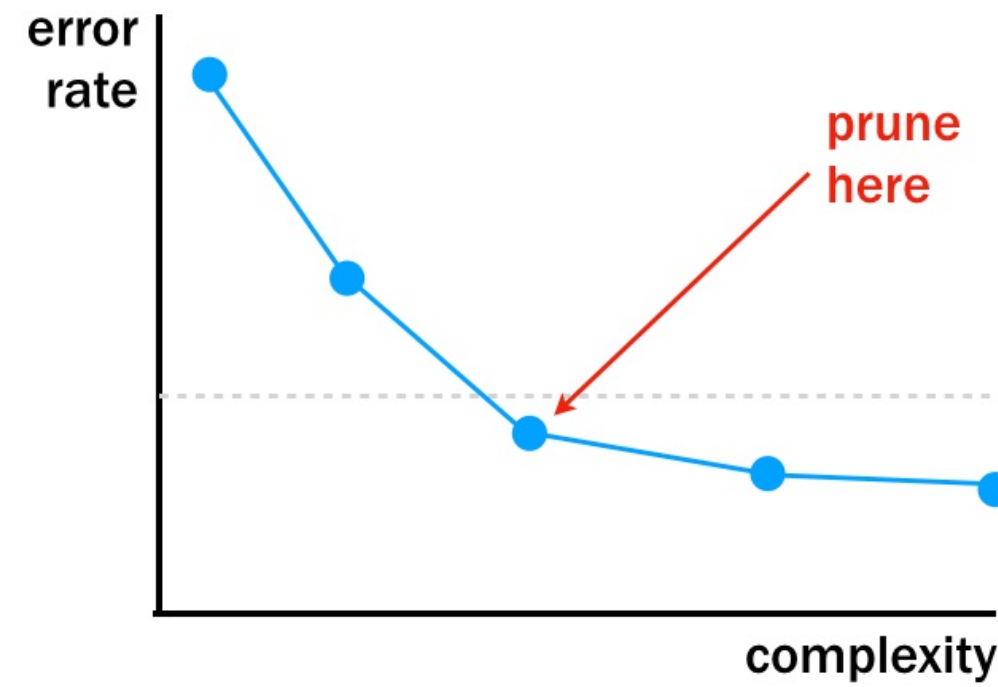
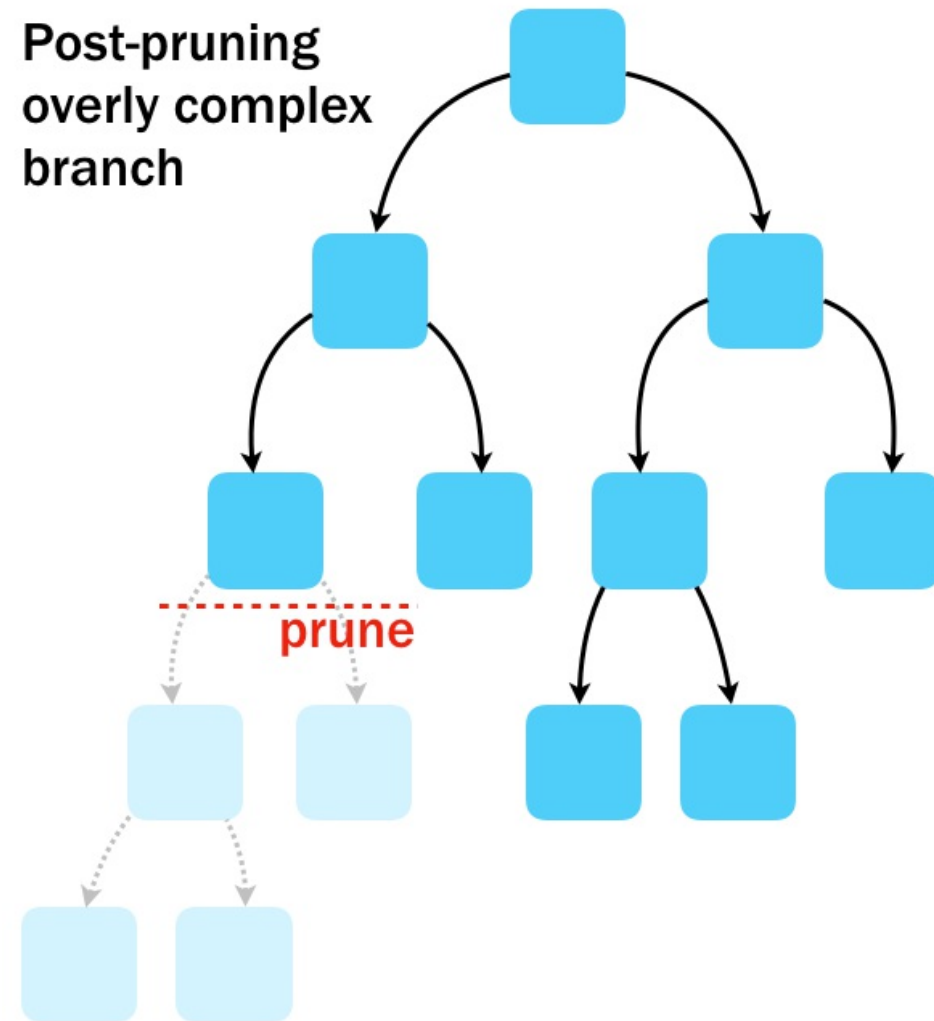


Minimum 10  
observations  
to split





# Post-pruning



# Pre- and post-pruning with R

```
# pre-pruning with rpart
library(rpart)
prune_control <- rpart.control(maxdepth = 30, minsplit = 20)

m <- rpart(repaid ~ credit_score + request_amt,
           data = loans,
           method = "class",
           control = prune_control)
```

```
# post-pruning with rpart
m <- rpart(repaid ~ credit_score + request_amt,
           data = loans,
           method = "class")

plotcp(m)

m_pruned <- prune(m, cp = 0.20)
```



## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**



## SUPERVISED LEARNING IN R: CLASSIFICATION

# Seeing the forest from the trees

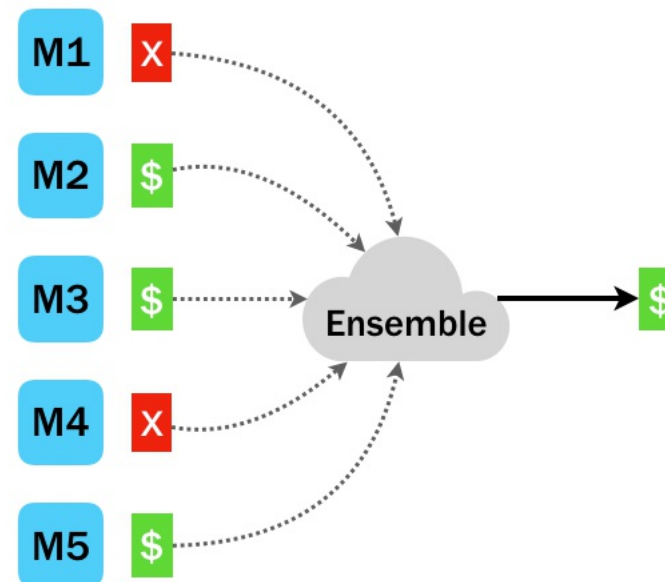
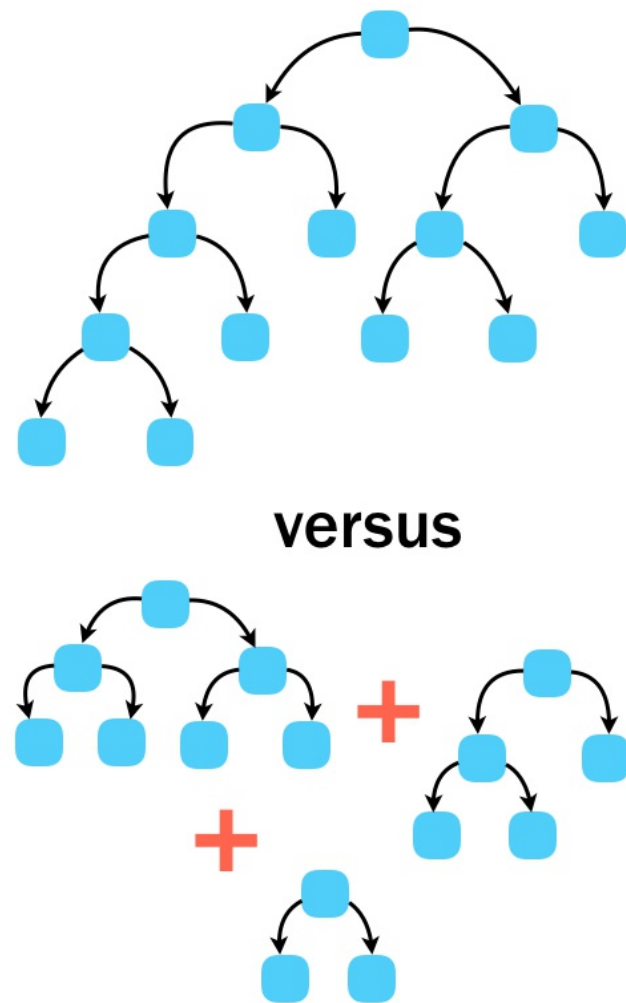
Brett Lantz  
Instructor

# Understanding random forests





# Making decisions as an ensemble





# Random forests in R

```
# building a simple random forest
library(randomForest)
m <- randomForest(repaid ~ credit_score + request_amt, data = loans,
                  ntree = 500,      # number of trees in the forest
                  mtry = sqrt(p)) # number of predictors (p) per tree

# making predictions from a random forest
p <- predict(m, test_data)
```



## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**