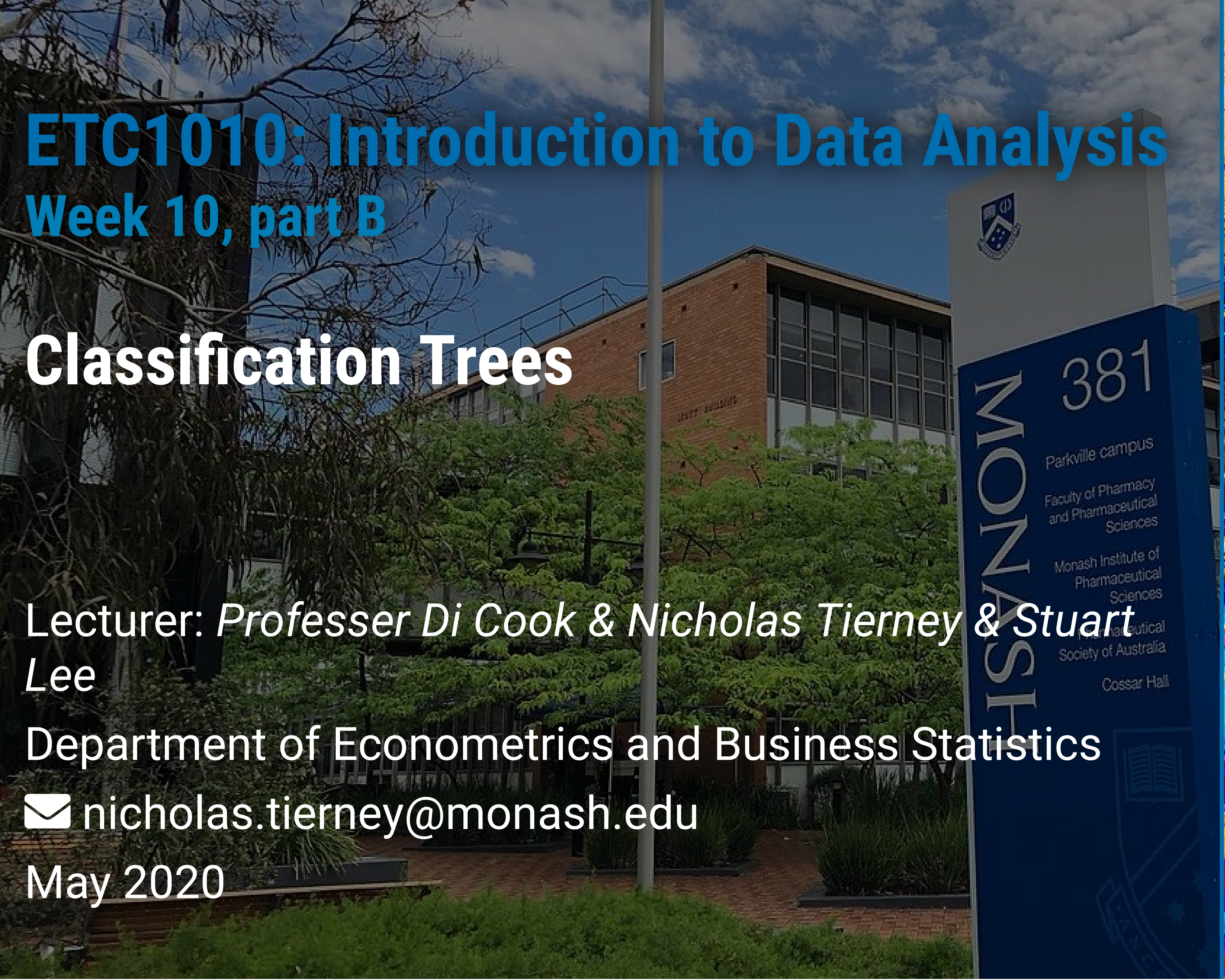


381
MONASH
Parkville campus
Faculty of Pharmacy and Pharmaceutical Sciences
Monash Institute of Pharmaceutical Sciences
Society of Australia
Cossar Hall

ETC1010: Introduction to Data Analysis

Week 10, part B

Classification Trees

Lecturer: *Professer Di Cook & Nicholas Tierney & Stuart Lee*

Department of Econometrics and Business Statistics

✉ nicholas.tierney@monash.edu

May 2020

recap

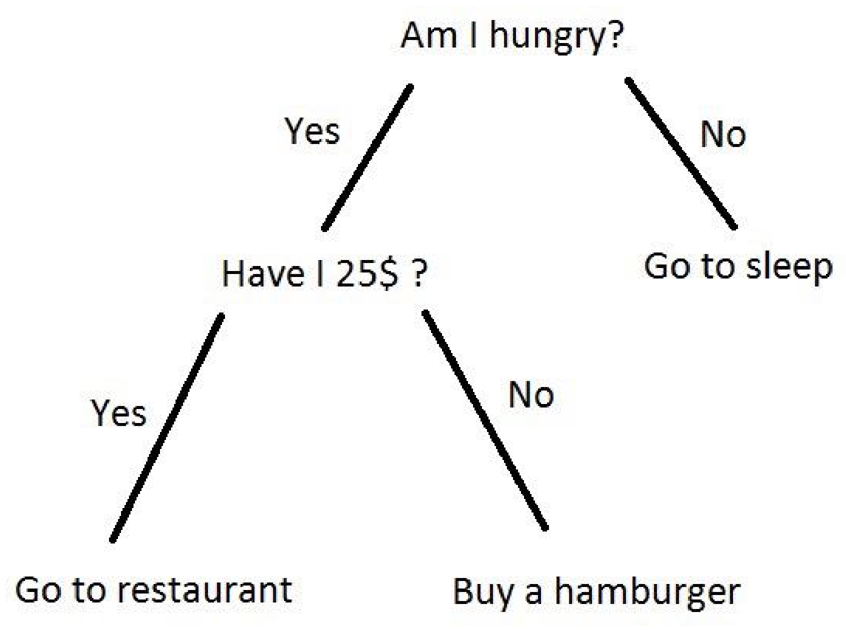
- Decision Tree

Admin

- Project
  - Use of data
  - Don't use kaggle datasets
  - Talk to us about your data in class and at consults
- Practical exam
  - Next Wednesday from 12pm Wednesday, closing 12pm Thursday

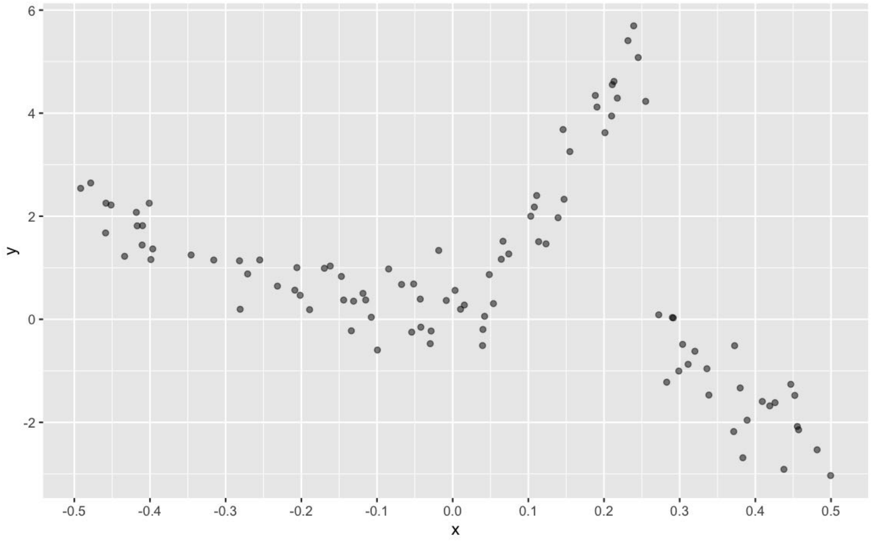
What is a decision tree?

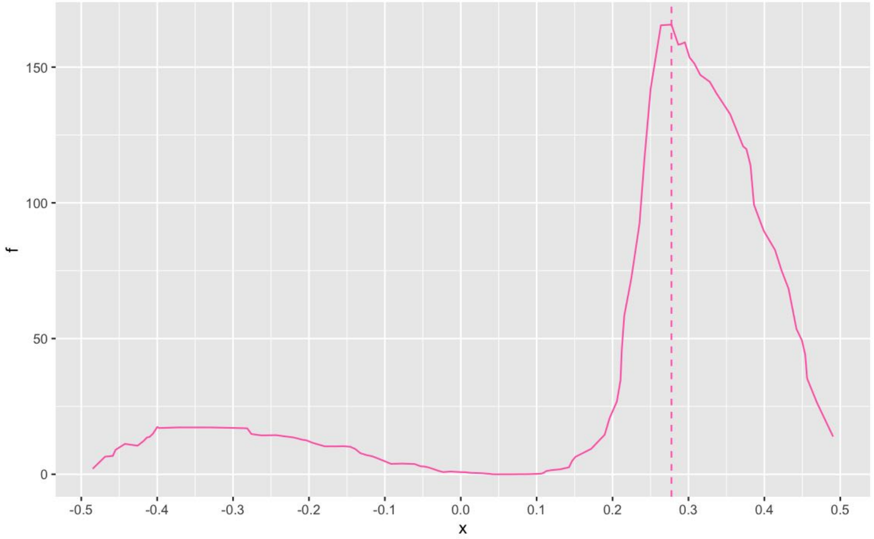
Tree based models consist of one or more of nested `if-then` statements for the predictors that partition the data. Within these partitions, a model is used to predict the outcome.

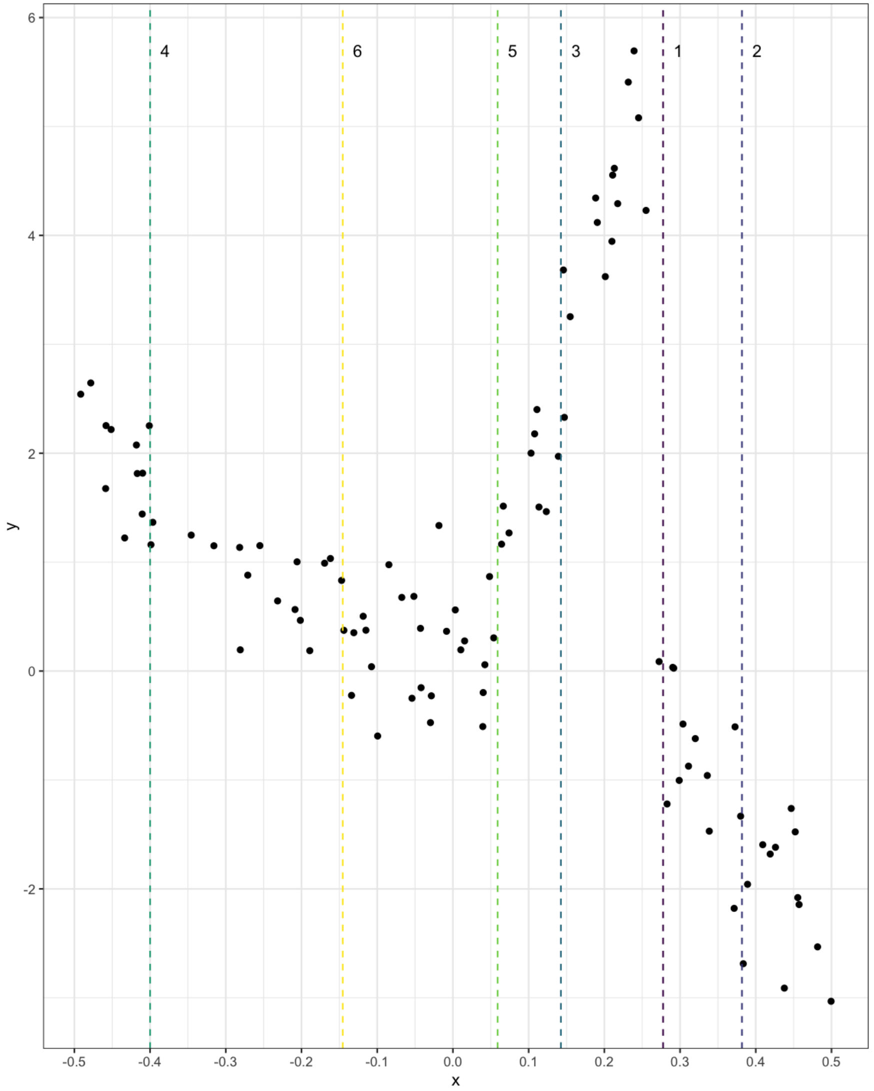


Source: Egor Dezhic

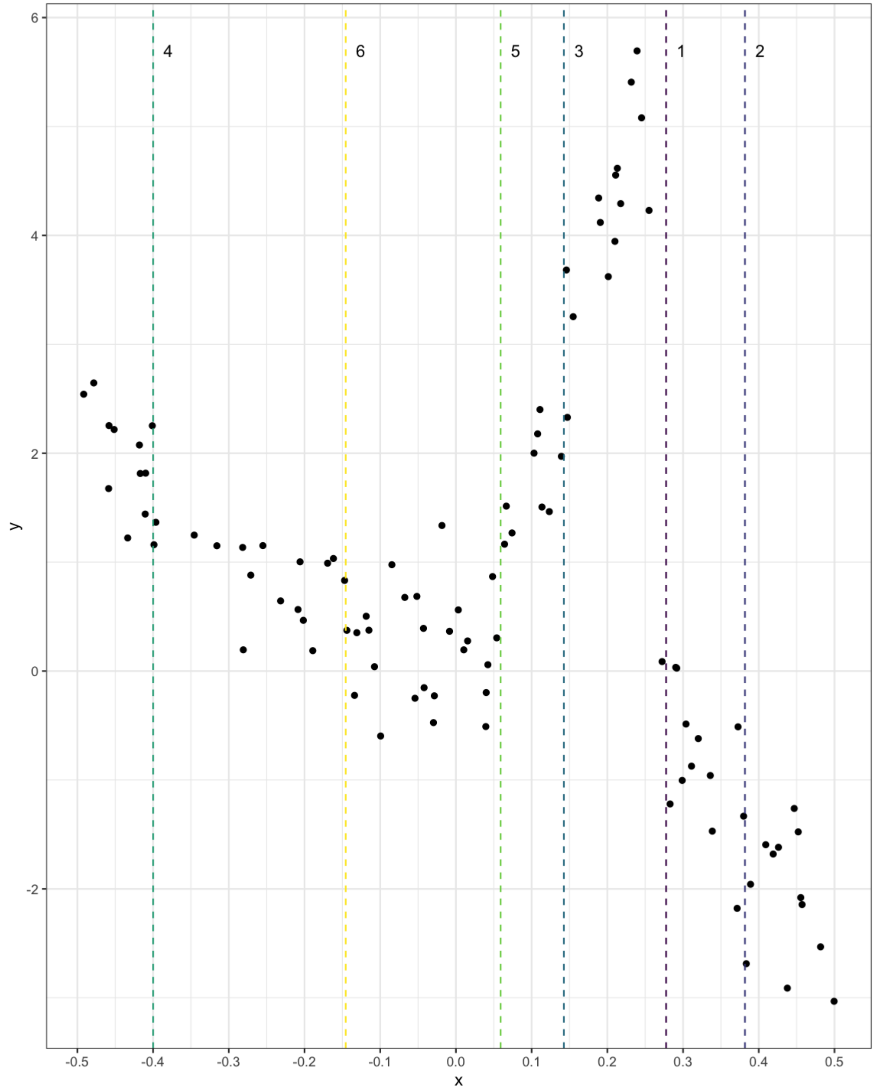
Regression Tree

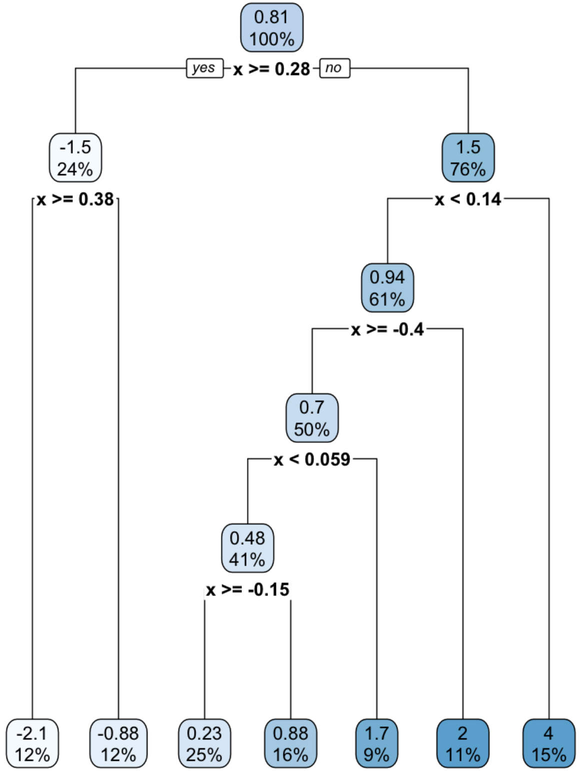






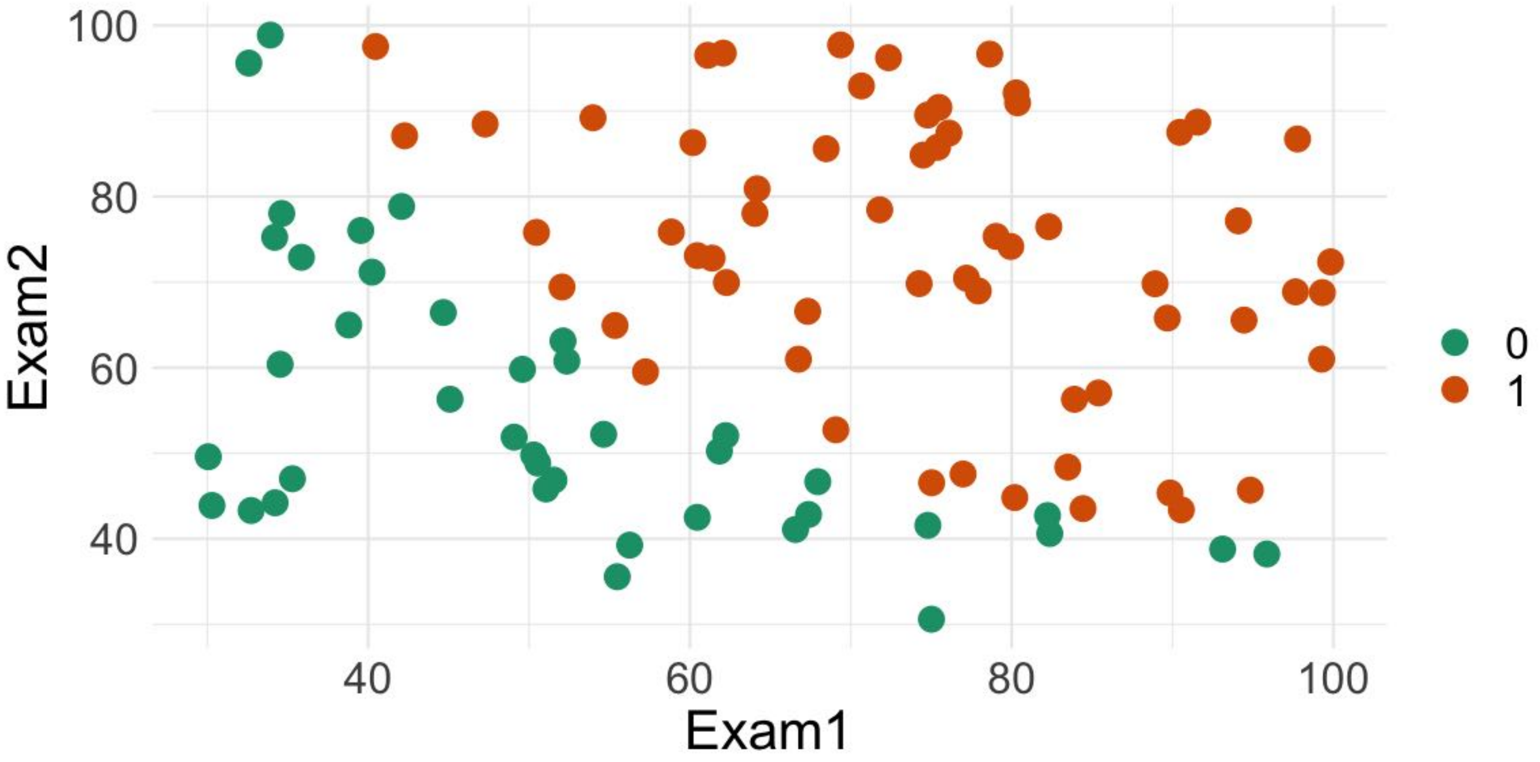
Regression Tree





Regression tree

- What if we want to predict something being in a particular group? Say,predicting whether someone passes a course based on two exam scores:
- Moving from continuous to categorical response.



Regression? Classification?

- Regression trees give the predicted response for an observation by using the mean response of the observations that belong to the same terminal node:

Classification

A classification tree predicts each observation belonging to the most commonly occurring class of observations.

However, when we interpret a classification tree, we are often interested not only in the class prediction (what is most common), but also the proportion of correct classifications.

Building a classification tree

- Similar approach to building a classification tree as for regression trees
- We use this "recursive binary splitting" approach
- But we don't use the residual sums of squares

$$SS_T = \sum(y_i - \bar{y})^2$$

Since we now have a category, we need some way to describe that.

We need something else!

Classification tree

- We can use the "classification error".
- Where we count up the number of mis-classified things, and choose the split that has the lowest number of mis-classified things.
- We can represent this in an equation as the fraction of observations in a region which don't belong to the most common class.

$$ E = 1 - \text{max}*{k}(\hat{p}*{mk}) $$

Here, $\hat{p}_{mk}$ refers to the proportion of observations in the mth region, from the kth class.

Understanding classification

Another way to think about this is to understand when E is zero, and when E is large

$$E = 1 - \max_k(\hat{p}_{mk})$$

E is zero when $\max_k(\hat{p}_{mk})$ is 1, which is 1 when observations are the same class:

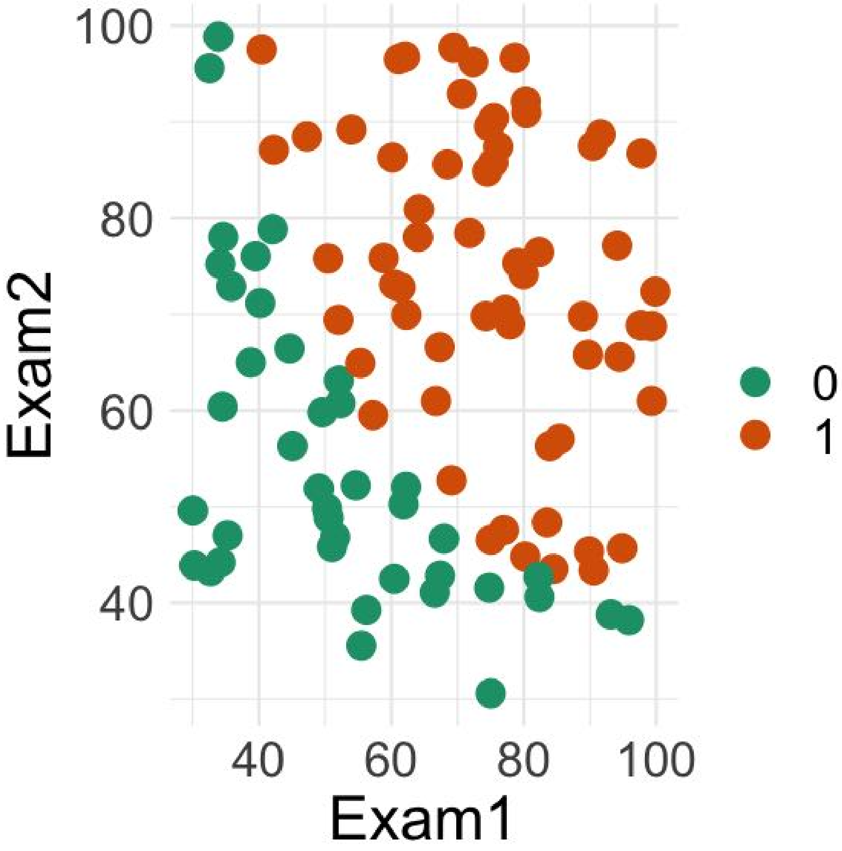
Classification trees

- A classification tree is used to predict a categorical response and regression tree is used to predict a quantitative response
- Use a recursive binary splitting to grow a classification tree. That is, sequentially break the data into two subsets, typically using a single variable each time.
- The predicted value for a new observation, x_0, will be the most commonly occurring class of observations in the sub-region in which x_0 falls

Predicting pass or fail ?

Consider the dataset `Exam` where two exam scores are given for each student, and a class `Label` represents whether they passed or failed the course.

```
##        Exam1    Exam2 Label
## 1 34.62366 78.02469     0
## 2 30.28671 43.89500     0
## 3 35.84741 72.90220     0
## 4 60.18260 86.30855     1
```

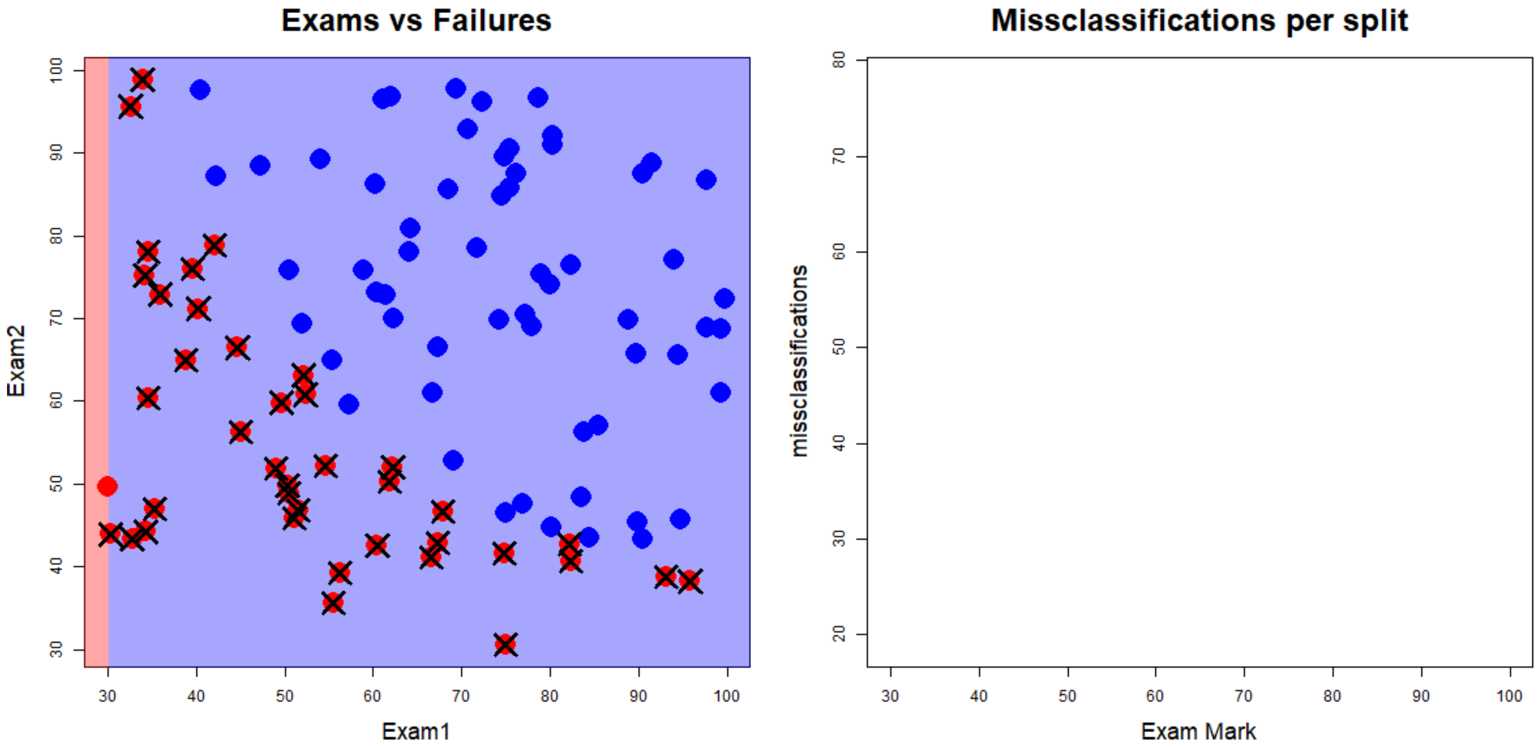


Your turn:

Open "10b-exercise-intro.Rmd" and let's decide a point to split the data.

Calculate the number of misclassifications

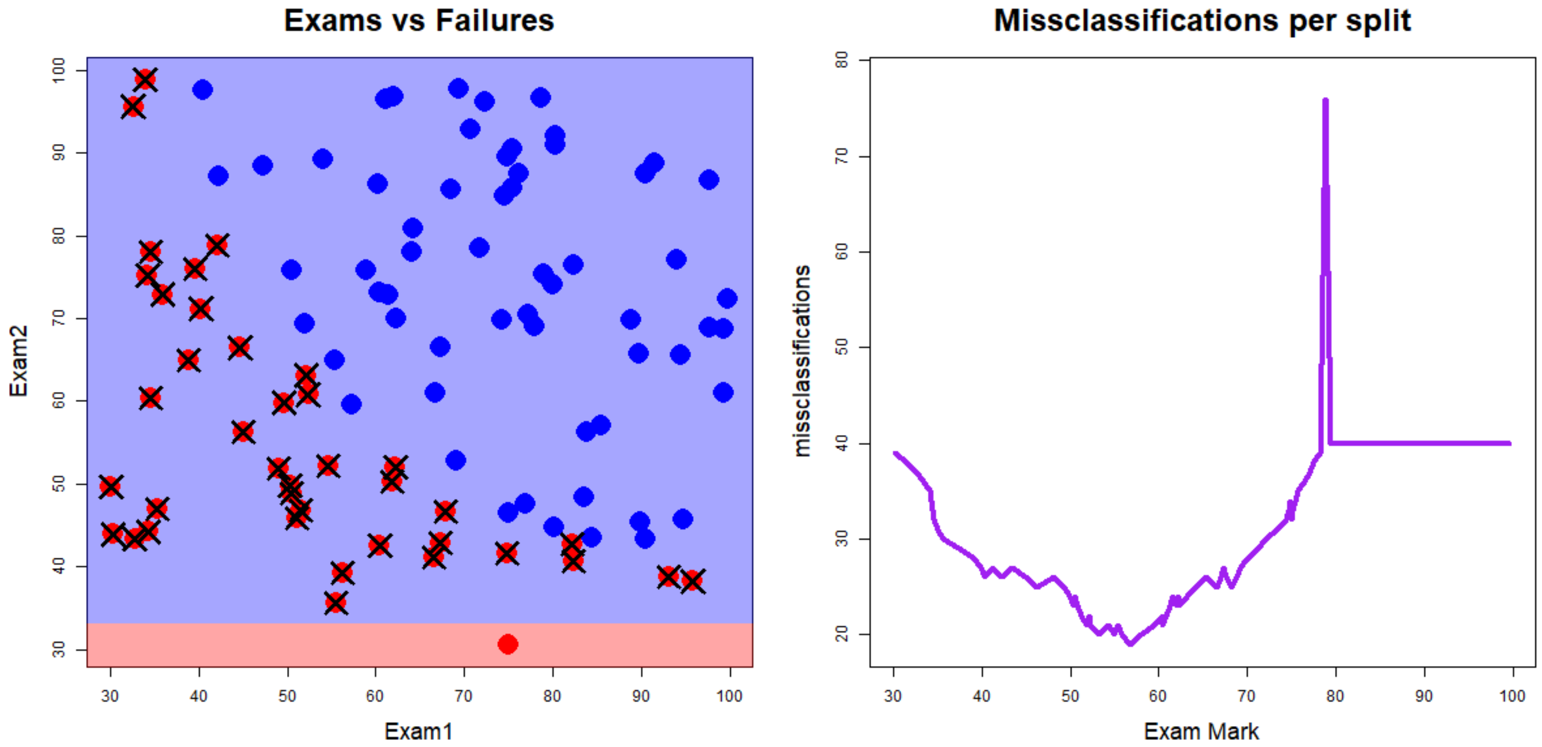
Along all splits for `Exam1` classifying according to the majority class for the left and right splits



Red dots are "fails", blue dots are "passes", and crosses indicate misclassifications. Source: John Ormerod, U.Syd

Calculate the number of misclassifications

Along all splits for Exam2 classifying according to the majority class for the top and bottom splits



Red dots are "fails", blue dots are "passes", and crosses indicate misclassifications. Source: John Ormerod, U.Syd

Combining the results from Exam1 and Exam2 splits

- The minimum number of misclassifications from using all possible splits of Exam1 was 19 when the value of Exam1 was **56.7**
- The minimum number of misclassifications from using all possible splits of Exam2 was 23 when the value of Exam2 was 52.5

So we split on the best of these, i.e., split the data on Exam1 at 56.7.

Split criteria - purity/impurity metrics

It turns out that classification error is not sufficiently sensitive for tree-growing.

In practice two other measures are preferable, as they are more sensitive:

- The Gini Index and
- Information Entropy.

They are both quite similar numerically.

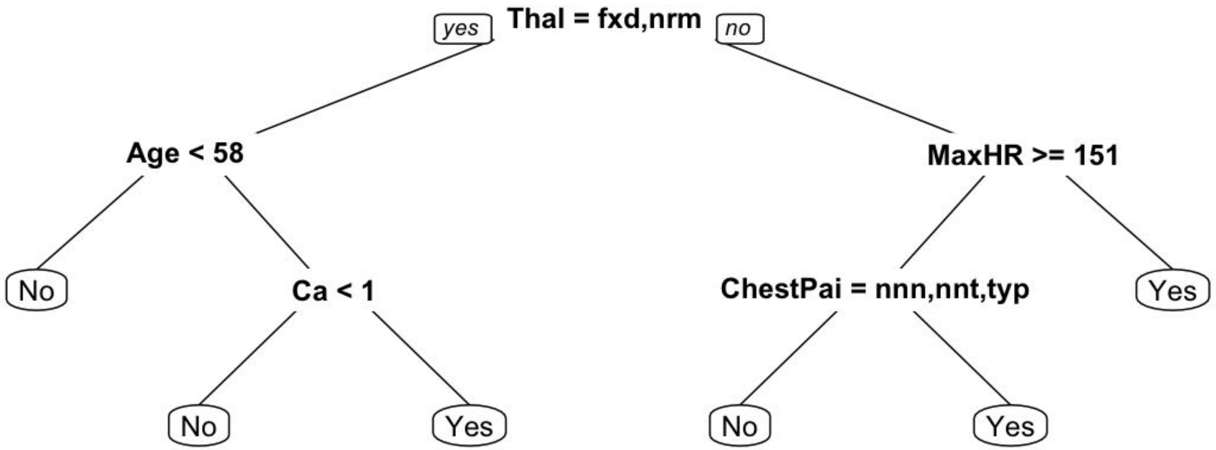
Small values mean that a node contains mostly observations of a single class, referred to as node purity.

Example - predicting heart disease

Y: presence of heart disease (Yes/No)

X: heart and lung function measurements

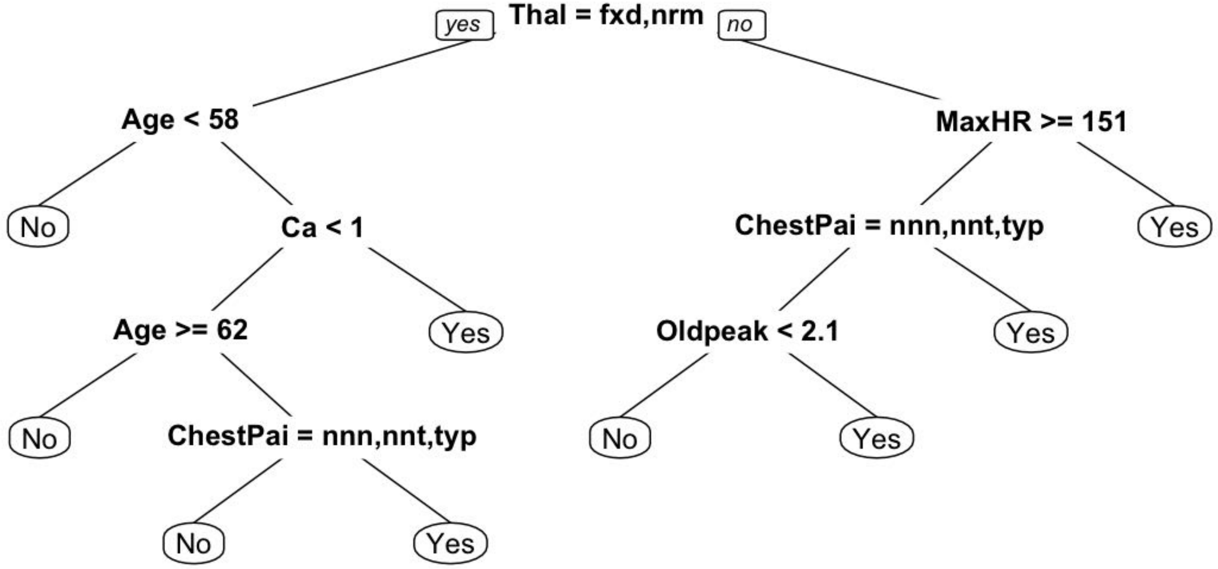
```
##  [1] "Age"       "Sex"       "ChestPain" "RestBP"    "Chol"      "Fbs"
##  [7] "RestECG"   "MaxHR"     "ExAng"     "Oldpeak"   "Slope"     "Ca"
## [13] "Thal"      "AHD"
```



Deeper trees

Trees can be built deeper by:

- decreasing the value of the complexity parameter cp, which sets the difference between impurity values required to continue splitting.
- reducing the `minsplit` and `minbucket` parameters, which control the number of observations below splits are forbidden.



Tabulate true vs predicted to make a confusion table.

		true	
		C1 (positive)	C2 (negative)
pred-	C1	*a*	*b*
icted	C2	*c*	*d*

- Accuracy: *(a+d)/(a+b+c+d)*
- Error: *(b+c)/(a+b+c+d)*
- Sensitivity: *a/(a+c)* (true positive, recall)
- Specificity: *d/(b+d)* (true negative)
- Balanced accuracy: *(sensitivity+specificity)/2*

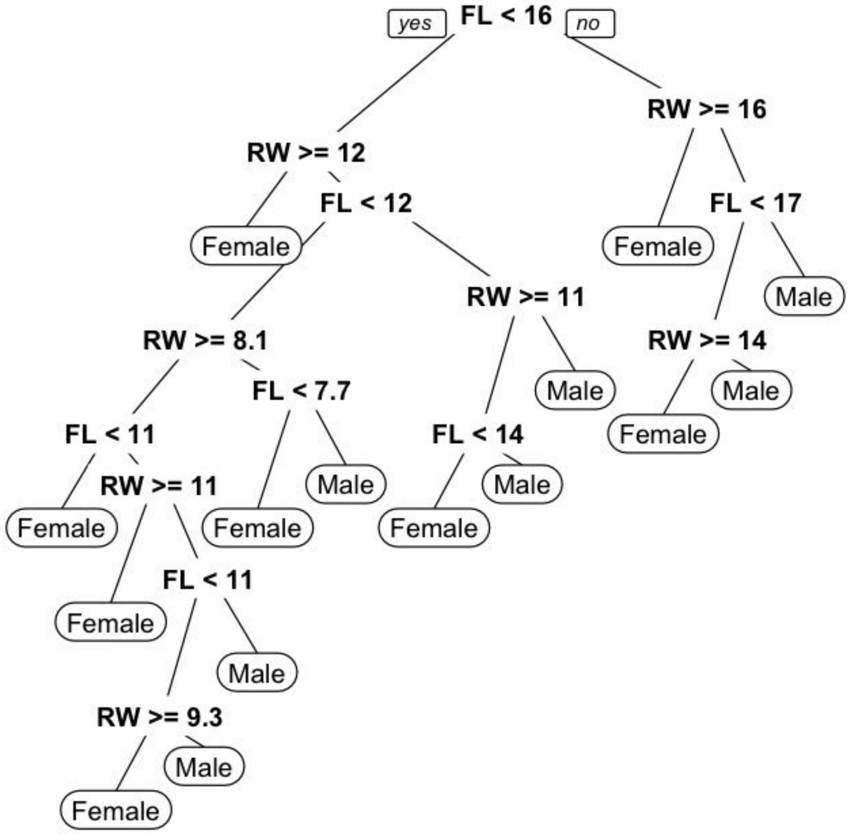
Confusion and error

```
##              Reference
## Prediction No Yes
##        No  75   5
##        Yes 11  58
##  Accuracy
## 0.8926174
```

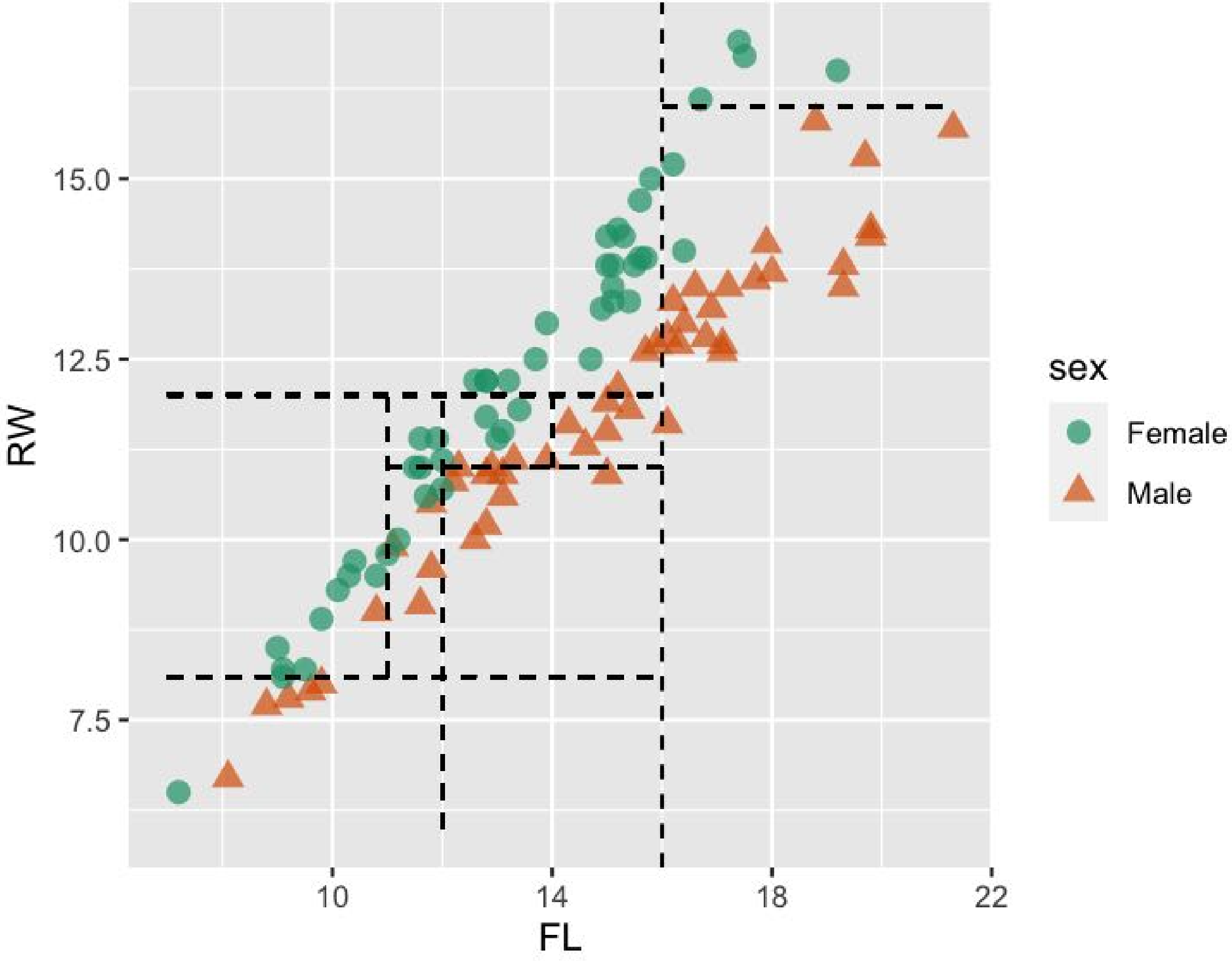
Example - Crabs

Physical measurements on WA crabs, males and females.

Data source: Campbell, N. A. & Mahon, R. J. (1974)

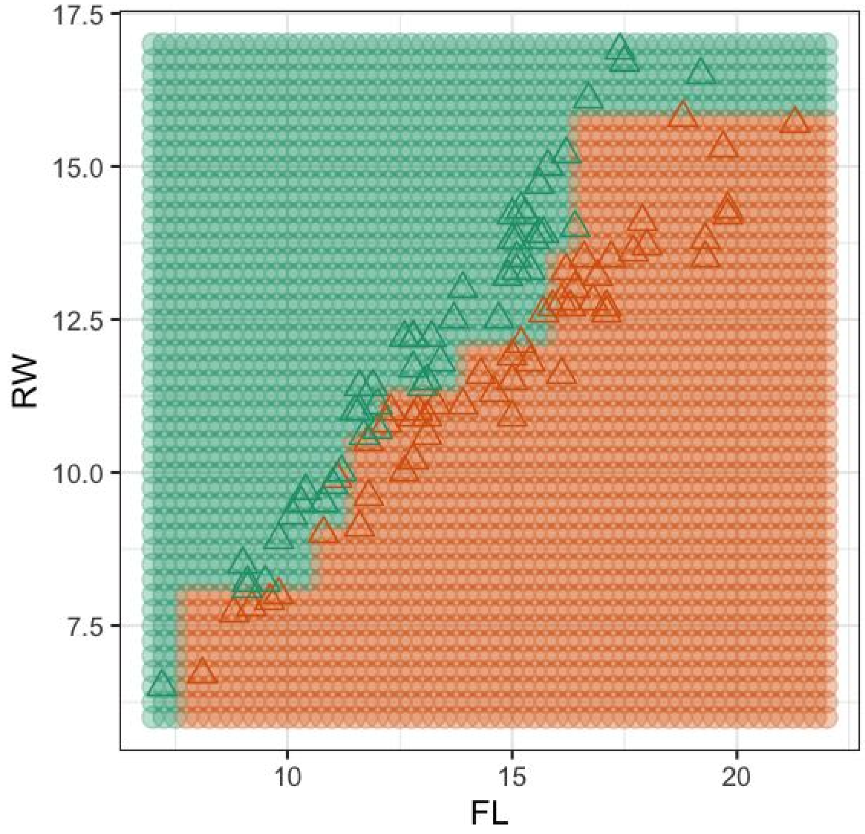


Example - Crabs

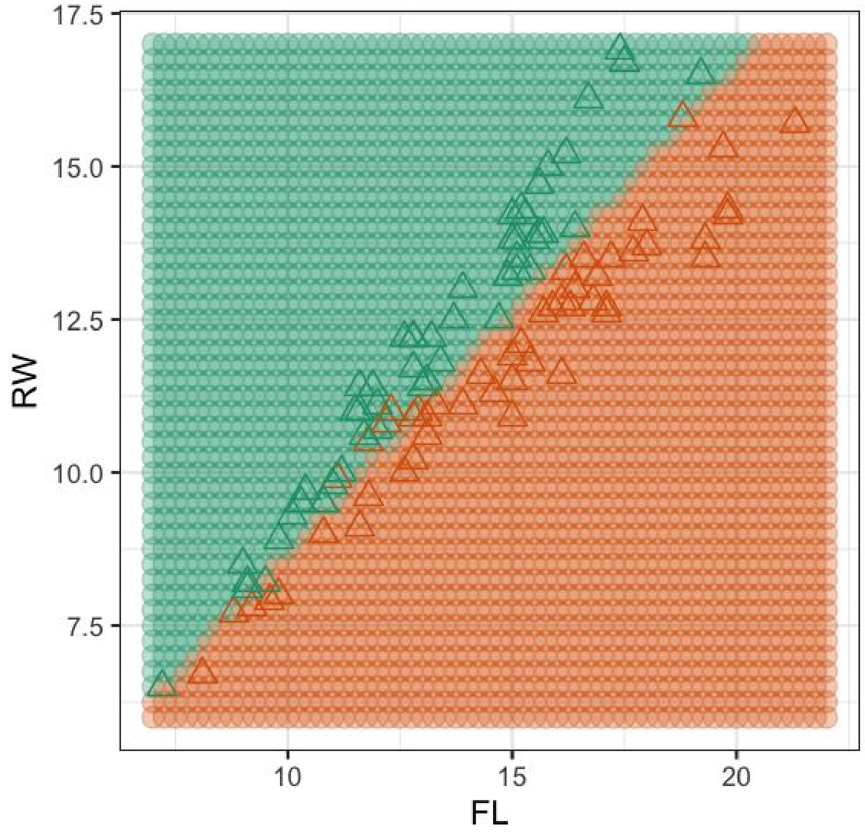


Comparing models

Classification tree



Linear discriminant classifier



Strengths and Weaknesses

Strengths:

- The decision rules provided by trees are very easy to explain, and follow. A simple classification model.
- Trees can handle a mix of predictor types, categorical and quantitative.
- Trees efficiently operate when there are missing values in the predictors.

Weaknesses:

- Algorithm is greedy, a better final solution might be obtained by taking a second best split earlier.
- When separation is in linear combinations of variables trees struggle to provide a good classification

👩‍💻 Made by a human with a computer

- Slides inspired by https://iml.numbat.space, https://github.com/numbats/iml.
- Created using R Markdown with flair by **xaringan**, and **kunoichi (female ninja) style**.