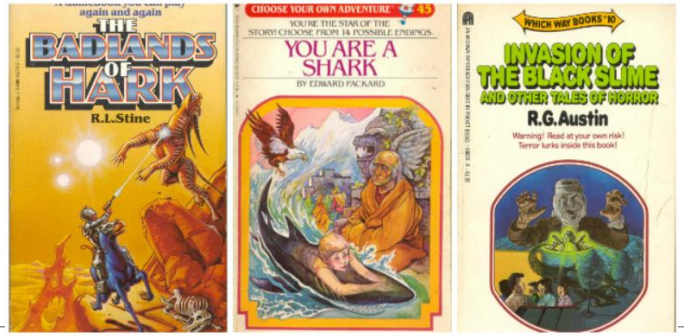
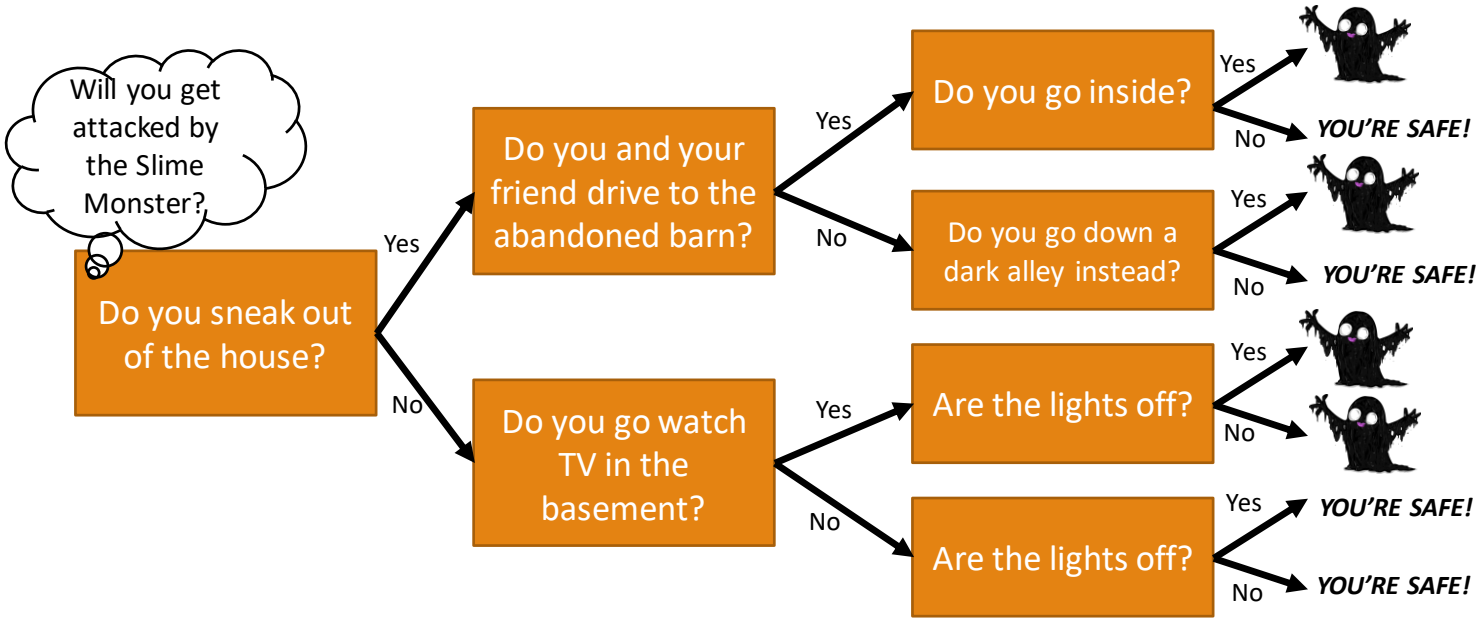
Decision Trees



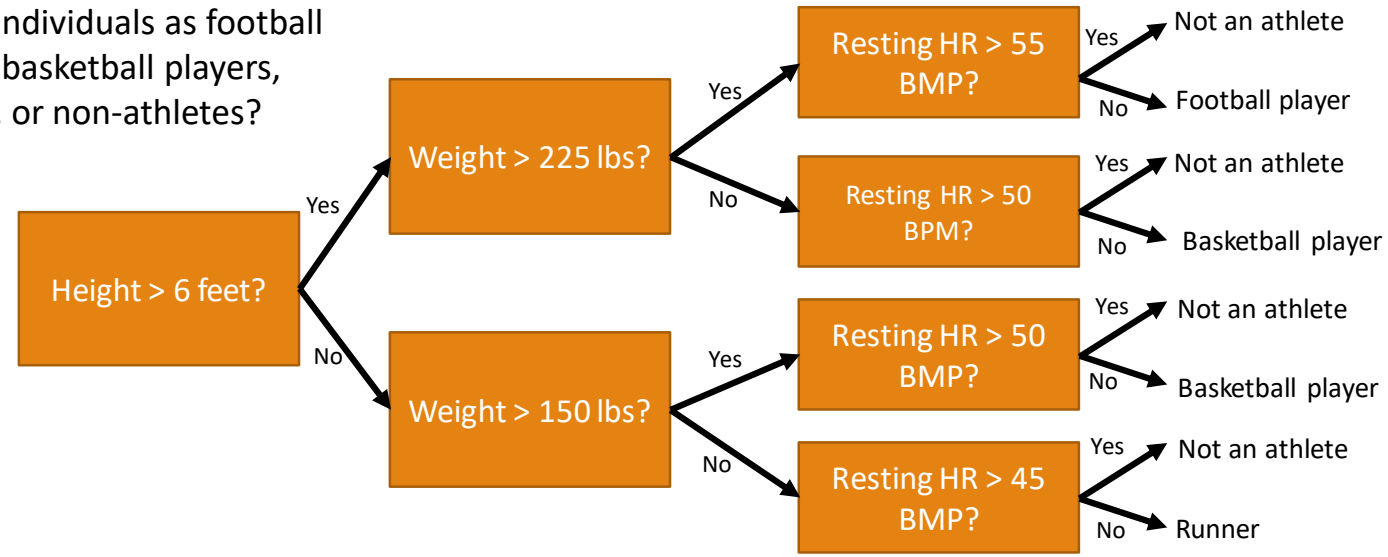
A decision tree is a method of classification similar to a "Choose Your Own Adventure" book. As we traverse the tree, we made a series of decisions (based on the data) that leads us to the conclusion (outcome)



Decision Tree Model

Our **decision tree model** is a "Choose Your Own Adventure" – but for data! Instead of making decisions about walking down a dark alley or leaving your flashlight at home, the decision tree model make decisions based on the values of the predictor data.

For example: Based on height, weight, and HR data, can we classify individuals as football players, basketball players, runners, or non-athletes?



How a Decision Tree Model Works

- Start with an empty node – this will become the ROOT of the tree
- Split the data, based on the threshold value of one attribute
  - Each split is based on the OPTIMAL attribute – which attribute has the best predictive power at this point?
  - Think: If we can only do one split, which attribute classifies the most data correctly?
- Now look at the resulting subsets at the new INTERNAL NODES.
- Analyze each internal node, and split the data again based on the next most optimal/predictive attribute
- Continue splitting until a stopping criteria is met
- The final split, which determines class assignment, are the LEAVES of the tree