**In-Class Activity: Decision Trees in R (Answer Key)**

**Question 1:**

Diagram, engineering drawing

Description automatically generated

**Question 1:**

Based on the tree you’ve generated, how likely is it that the following passengers will survive:

A Male with a ticket fare of 15? 0.10=10/(171+19)  
B 30 year-old male who pays 35 for a ticket of class 1? 0.43=25/(33+25)C 30 year-old female with a ticket of fare 20 and ticket class of 3? 27/(27+28)=0.49

D 23 year-old female who pays 15 for the ticket of class 4? 27/(27+28)=0.48

**Question 2:**

Which tree has higher accuracy rate, Tree #1 or #2?

**#2 has the higher accuracy rate for training set, and #1 has the higher accuracy rate for the testing set.**

Does increasing max-depth always lead to higher accuracy rate of the testing set?

**No**

**Question 3:**

Based on the analyses we have done so far, use your own words to summarize how **max\_depth** and **min\_samples\_split** can alter the decision tree.

1. **Larger max\_depth → more complex tree:**

A larger max\_depth means that more layers are allowed in the tree, thus may result in more splits, i.e. a more complex tree.

1. **Smaller min\_samples\_split → more complex tree:**

A smaller min\_samples\_split means that fewer observations are needed for an additional split to be allowed, thus may result in more splits, i.e. a more complex tree.