Name:_____ Date:_____

LAB 4G: Growing trees
Response Sheet

Directions: Record your responses to the lab questions in the spaces provided.

Trees vs. Lines

Our first tree

(1) Write and run code using the `tree()` function to create a *classification* tree that predicts whether a person `survived` the Titanic based on their `sex`.

(2) Why can't we just use a *linear model* to predict whether a passenger on the Titanic `survived` or not based on their `sex`?

Viewing trees

(3) To actually look at and interpret our `tree1`, place the model into the `treeplot` function.

- **(4) Write down the labels of the two *branches*.**

- **(5) Write down the labels of the two *leaves*.**

Answer the following, based on the `treeplot`:

- **(6) Which `sex` does the model predict will survive?**

- **(7) Where does the plot tell you the number of people that get sorted into each leaf? How do you know?**

- **(8) Where does the plot tell you the number of people that have been sorted *incorrectly* in each leaf?**

Name:_____ Date:_____

LAB 4G: Growing trees
Response Sheet

Leafier trees

(9) Similar to how you included multiple variables for a linear model, create a `tree` that predicts whether a person `survived` based on their `sex`, age, `class`, and where they `embarked`.

(10) Write and run code creating a `treeplot` for this model and answer the following questions:

- **(11) Mrs. Baxter was a 50-year-old female with a 1st class ticket from Cherbourg. Does the model predict that she survived?**

- **(12) Which variable ended up not being used by `tree2`?**

Tree complexity

(13) Using the same variables that you used in `tree2`, create a model named `tree3` but include `cp = 0.005` and `minsplit = 10` as arguments.

- **(14) How is `tree3` different from `tree2`?**

Predictions and Cross-validation

(15) Fill in the blanks below to predict whether people in the `titanic_test` data survived or not using `tree1`.

```
titanic_test <- mutate(________, prediction = predict(________, newdata = ________, type = "class"))
```

Measuring model performance

(16) Where does the first misclassification occur?

Name:_____ Date:_____

LAB 4G: Growing trees
Response Sheet

Misclassification rate

(17) Fill in the blanks to create a function to calculate the MCR.

```
calc_mcr <- function(actual, predicted) {

  sum(___________ != ___________) / length(actual)

}
```

On your own

(18) In your own words, explain what the *misclassification rate* is.

(19) Which model (`tree1`, `tree2`, or `tree3`) had the lowest misclassification rate for the `titanic_test` data?

(20) Write and run code creating a 4th model using the same variables used in `tree2`. This time though, change the *complexity parameter* to `0.0001`. Then answer the following.

- **(21) Does creating a more complex *classification tree* always lead to better predictions? Why not?**

(22) Write and run code creating a *regression tree* model to predict the Titanic's passengers' ages and calculate the MSE.