

Problem: You got lazy during your data entry. You only filled in the site, plot, and year data for the first relevant row each time. Now, your data set is full of empty cells, and R won’t run stats properly on data that looks like this.

Goal: Sure, you could fill in all the missing info by hand, but you have 170 other data sheets just like this one. Yikes! Write an R function that will fill in all the missing information for you.

New Functions to know:

1. dim(). Returns the dimensions of a matrix/array/data frame (*i.e*. it returns a vector with two entries—the number of rows in the object given, and the number of columns in the object given).

2. View(). Opens up a viewer that allows you to see an object or data file. Note the capital V!

Steps to take:

1. Start with the basic function skeleton we’ve already shown you.

2. Set up the function so that it takes a “dirty” data sheet as input and a returns a “clean” data sheet as output.

3. Use dim() to figure out how many rows and columns there are in the data sheet given.

4. Use a for loop to go column by column through the data sheet, and then a **second for loop** INSIDE of the first one to go row by row through the data sheet. Essentially, these two for loops will allow us to check each individual cell in the data sheet, top to bottom within a column and then left to right between columns.

5. For each cell, use an if call to check if the cell is empty. If it is, use what’s in the cell **directly above it** to fill it in with the correct data.

Function testing:

Use the data sheet “missing.data.csv” to test your function. Save the data sheet to your working directory (ask for help if you don’t know how to do this). Then, load the data sheet by running this code:

data1 = as.matrix(read.csv(“missing.data.csv”))

You can view your data sheet before and after you run your function on it by using View()

View(data1)