



Working in Program R – A skill assessment

R has become the universal quantitative interface for ecologists – it’s free, powerful, flexible enough to do everything but write manuscripts for you, and is relatively user friendly. Having a reasonable understanding of R will make working with your data, from manipulation to analysis, much easier. We will be using Program R extensively throughout this course and your skill level will build over our two weeks together rather quickly. In this exercise, you will be asked to complete some simple data handling tasks on an example dataset so that we may gauge your current level of R-mastery. Below each prompt, please paste the code (i.e. lines of R script) you used to best serve your needs. **You are not required to provide an “answer” to any of the questions nor a verbal description of the steps taken.** . If you reach a question that you just can’t figure out, don’t sweat it too much, this is simply our way to figure out how to best serve you. Prior to completing this exercise, we strongly suggest that you complete the Swirl exercises (see instructions [link here](#)). After completing this exercise, please save it as an R file with the title “yourName_RAassessment.R” and email as an attachment to rydert@si.edu and bsevensunc@gmail.com.

Before you begin:

- If you do not yet have R on your computer, download and install Program R ([link here](#)).
- Download and install R Studio onto your computer ([link here](#)).
- Download the file ([link here](#)) to an accessible location on your hard drive.
- Open R Studio and copy-and-paste questions 1-20 into an R script. Place a “#” before each question – the “#” is used to comment out sections of your script (sections of code not evaluated in R). Following each question prompt, place the code you used to answer the question (don’t provide the answers!).

Stumped? It happens to everyone! Some quick search engine queries may help you solve problems (Stack Overflow is often a great source). There’s also a good “R for Beginners” guide ([link here](#)) and a great cheat sheet for summarizing data using dplyr ([link here](#)) available. If you find any of these operations difficult, be sure to revisit the Swirl exercises!

1. Provide the script below that you would use to set your working directory to the folder that contains exampleBirdData.csv.
2. Install the R library (i.e., package) “dplyr” and load the library into R’s global environment.
3. Load exampleBirdData.csv into R’s global environment and assign it the name “birdCounts”.
4. What type of object is birdCounts (i.e. data frame, matrix, factor)?

5. What type of object is the birdCounts field (i.e. column) labeled “count”?
6. Display the structure of birdCounts and all associated fields (object types).
7. How many records (i.e. rows) does birdCounts contain?
8. Using one line of code, display the number of records and fields in birdCounts.
9. Display the first 10 records of birdCounts.
10. Display the names of all of the fields in birdCounts.
11. Change the name of the “count” field to “abundance”.
12. Determine how many species were observed in this study.
13. Determine the total observed abundance (i.e., across species) in this study.
14. You want to explore the abundance the Gray Catbird (GRCA, *Dumetella carolinensis*). Make a reduced data frame containing just this species and call this data frame “grcaCounts”.
15. Display the summary information for grcaCounts (i.e. factor counts and summary statistics such as mean and median abundance).
16. Save grcaCounts as a .csv file to your hard drive.
17. Determine the number of sites in which Carolina Chickadees (CACH, *Poecile carolinensis*) was observed.
18. Construct a table that summarizes the number of sites in which each dietary guild (field name = “diet”) was observed.
19. Summarize the abundance of birds across sites by making a new data frame called “birdSpeciesCounts”. Your data frame should have only “species” and “abundance” fields.
20. You want to explore patterns of abundance by species. To do so, sort birdSpeciesCounts from highest to lowest abundance.

A challenge! Write a function that calculates the standard error of Gray Catbird abundance. Standard error can be defined by the mathematical formula below, where x would represent the vector of GRCA abundances across sites and n the number of sites. *Hint: If you’re stumped on this one, visit [this web](#) page that describes writing functions in R.*

$$\text{Standard error } (x) = \frac{\text{StDev}(x)}{\sqrt{n}}$$