R Assignment 1: Introduction to R, RStudio, and the Global Biodiversity Information Facility (GBIF)

#### Motivation

As taxonomists it is useful to know where species occur and co-occur on continental scales to help inform species delimitations. For example, two sister species separated by a major geographic barrier (such as a mountain range) are not likely to interbreed. This type of information is one piece of the puzzle of delimiting species. To make preliminary estimates of a species’ distribution, we can harness the data resources organized by the [Global Biodiversity Information Facility](https://www.gbif.org/what-is-gbif). GBIF is an open-access database housing information about where and when species have been observed or collected. To the modern plant taxonomist, it is indispensable: one can search GBIF for any plant taxon of interest, and immediately have access to all global records of where and when their plant has been observed.

To scratch the surface of GBIF reveals a massive database spanning the entire tree of life. How does one go about efficiently collecting, organizing, validating, and analysing so much information? The solution is made tractable by employing data analysis tools like those found in R (Note: For an introduction to R, see this week’s [pre-lab reading in the section below](#prelab)). By combining the biodiversity resources of GBIF with the analytical tools in R, we can collect, process, and visualize species distribution data on any laptop or tablet. In this week’s lab we’ll first gather the required tools (R and GBIF) and develop some basic skills using them. In the following lab we’ll use our new skills to create research-grade distribution maps.

#### Pre-lab

Before starting this week’s lab you will have:

1. completed pre-lab [Activity 1: Exploring the GBIF website](#A1)
2. completed the pre-lab reading: [Introduction for Students](https://moderndive.com/preface.html#introduction-for-students) and [Chapters 1.1–1.3](https://moderndive.com/1-getting-started.html) from ModernDive—a textbook for learning data science using R. We will work through Chapters 1–4 of this textbook over the semester, and we encourage you to read ahead (through Ch.4) on your own time!
3. R and RStudio downloaded onto your computer. [Chapter 1.1 in ModernDive](https://moderndive.netlify.app/1-getting-started.html) provides instructions for downloading R and RStudio. If you have trouble with installation, contact the TAs as soon as possible for assistance,
4. [OPTIONAL] created an RStudio Cloud account. Although ModernDive provides an excellent introduction to R, RStudio Cloud has some [useful interactive primers](https://rstudio.cloud/learn/primers/) that make sense of the R language by getting you to write some simple code. **This pre-lab activity is not required**, but recommended for anyone that is completely new to R, or wants some hands-on experience with R right away.

#### Outcomes

By the end of this lab you will:

1. be familiar with GBIF and the types of data that can be obtained,
2. know the difference between R and RStudio,
3. be familiar with the layout of RStudio, and understand how to install packages and navigate the iris dataset (could also do our own exercises with the iris dataset, but see <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/iris.html> for problems),
4. have a sense of how R and GBIF can be used together to create species distribution maps,

#### Activity 1 (Pre-lab): Exploring the GBIF website

Let’s jump into GBIF! First go to the [GBIF website](https://www.gbif.org). In another tab/window follow along with [this video tutorial](https://www.youtube.com/watch?v=40yPnJdeM_A&t=3s) to learn how to navigate through the GBIF website (NOTE: creating an account and downloading the data onto your computer, as suggested in the video, is **NOT necessary**).

By following the video tutorial, you’ll learn how to search for GBIF data, what types of data your search will return, and how that data is structured. Notice that GBIF is a biodiversity database aggregator, pulling data from other databases such as eBird and iNaturalist. That means that you can choose to share your BIOL324 iNaturalist observations so that they appear on GBIF!

**CANVAS QUIZ QUESTION**

#### Activity 2: An Introduction to R

**MB: I’m imagining doing this live on Zoom**

Now that we have a sense of what kinds of data we could get from GBIF (namely, observation records of a species), we need a tool that will allow us to organize, filter, and visualize this data. While there are many options for handling biodiversity data, in BIOL324 we will focus on using R. The layout and style of R can be daunting if this is your first time using it. Don’t worry! In this course, we don’t expect you to know how to use R—we’ll provide the resources needed to learn the basics, and there will be plenty of lab time to work together so that we’re all tackling the learning curve together. And the payoff is worth it: knowing R can help you land summer research jobs, and is increasingly used for teaching in BIOL classes (e.g. BIOL300, BIOL406, BIOL413, BIOL416).

From reading [Chapter 1.1 of ModernDive](https://moderndive.netlify.app/1-getting-started.html), you will know that R is a programming language, while RStudio is simply a user interface that ‘wraps around’ R. RStudio is not *needed* to use R, but it is convienient! While RStudio is an app that you download to locally to you computer, RStudio Cloud is web-based, meaning that you can use RStudio within your web browser, without downloading anything. If you created an RStudio Cloud account to complete the primers (see: [pre-lab activities](#prelab)) then you can use your account to use RStudio online (**not required for this class**).

During this activity, we’ll work through [Chapter 1.4](https://moderndive.com/1-getting-started.html) to explore a dataset in R. Along the way, we’ll learn some basic R skills like creating an object, determining classes, using functions, and dissecting datasets. All of these skills will come in handy as we move towards using R and GBIF together to solve taxonomic problems.

To follow along, open RStudio cloud and attach the required packages (see: [Chapter 1.4](https://moderndive.com/1-getting-started.html)). The TA will step through this activity while you follow along. If something isn’t working for you, please let us know in the chat! It is **very likley that someone is having or will have the same problem as you!**.

**CANVAS QUIZ QUESTIONS** - What are the different windows (console, script, environment, plot) and what do they do? - what is an object? - what is a function?