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

Mannfred Boehm

#### Outcomes

By the end of this tutorial you will: 1) be familiar with GBIF and the types of data that can be obtained, 2) have some basic skills in R and understand why/how it is useful to taxonomic research, 3) know the difference between R, RStudio, and RStudio Cloud, 4) have created an RStudio Cloud account, and be familiar with the BIOL324 workspace, 5) have a sense of how R and GBIF can be used together to create species distribution maps, 6) have chosen some plant group(s) to map.

#### Motivation

As taxonomists we need to estimate where related 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.

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 such as R. To bring yourself up to speed, this week’s required reading is [Introduction for Students](https://moderndive.netlify.app/preface.html#introduction-for-students) and [Getting Started With Data in R](https://moderndive.netlify.app/1-getting-started.html) from ModernDive - a textbook for learning data science in R. In Section 1.1, they will ask you to open RStudio, please see the instructions in Activity 2 (below) to follow along.

By combining the biodiversity resources of GBIF with the analytical tools of R, we can collect, process, and visualize species distribution data on any laptop or tablet. In this week’s tutorial we’ll first gather the required tools (R and GBIF) and develop some basic skills using them. In the following tutorial we’ll use our newfound skillset to create research-grade distribution maps.

#### Activity 1: Exploring the GBIF website

MB: Add some instructions or an activity for exploring the GBIF website.

#### Activity 2: An Introduction to the BIOL324 RStudio Workspace (15 minutes + 45 minutes reading textbook)

From reading [Getting Started With Data in R](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 typically downloaded locally, RStudio Cloud allows you to use RStudio within your web browser, without downloading anything.

To get started go to [RStudio Cloud](https://rstudio.cloud/) and create an account.

Once logged on, use [this invitation link](https://rstudio.cloud/spaces/73822/join?access_code=eGVBWIOKYgPYkF6jFV6zKvwz180tjCqCGIM67YwC) to join the BIOL324 RStudio Workspace. In the left-side panel there is a project in our workspace named ‘Mapping Species Distributions’. Click *Begin* to open it. Take a look around and refer to [Getting Started With Data in R](https://moderndive.netlify.app/1-getting-started.html) to make sense of the various windows, buttons, menus, and features. We’ll return to this project next week to better understand what is going here.

#### Activity 3: First steps in R (30 minutes)

R is best learned by doing. To develop some basic skills in R, log onto RStudio Cloud and navigate to **Primers** on the left-side panel. Required for this week are the exercises *The Basics* and *Work With Data*. We strongly recommend completing *Visualize Data*, and *Tidy Your Data*, though these are optional.

MB: Could add in some open-ended questions here to help motivate completing these tutorials. Maybe some questions that get them thinking about how basic functions like creating objects or subsetting could be useful for handling species distribution data.

#### Activity 4: Selecting an interesting plant

MB: Some activity or brainstorm that gets the students thinking about which plant groups to choose for mapping. This is a good opportunity to teach nomenclature, synonyms, authorities, etc. Could also be it’s own tutorial!