



# Intro to R and Open Science Practices for Biologists Workshop

مقدمة في لغة الأَر وممارسات  
العلوم المفتوحة لعلماء الأحياء

د. مُنح أبو العز 

2022-31-10



# What is R?



- R is a statistical programming language
- R has been around since 1995, and was created by *Ross Ihaka* and *Robert Gentleman* at the *University of Auckland, New Zealand*
- R is based off the **S programming language** developed at *Bell Labs* and was developed to teach intro *statistics*

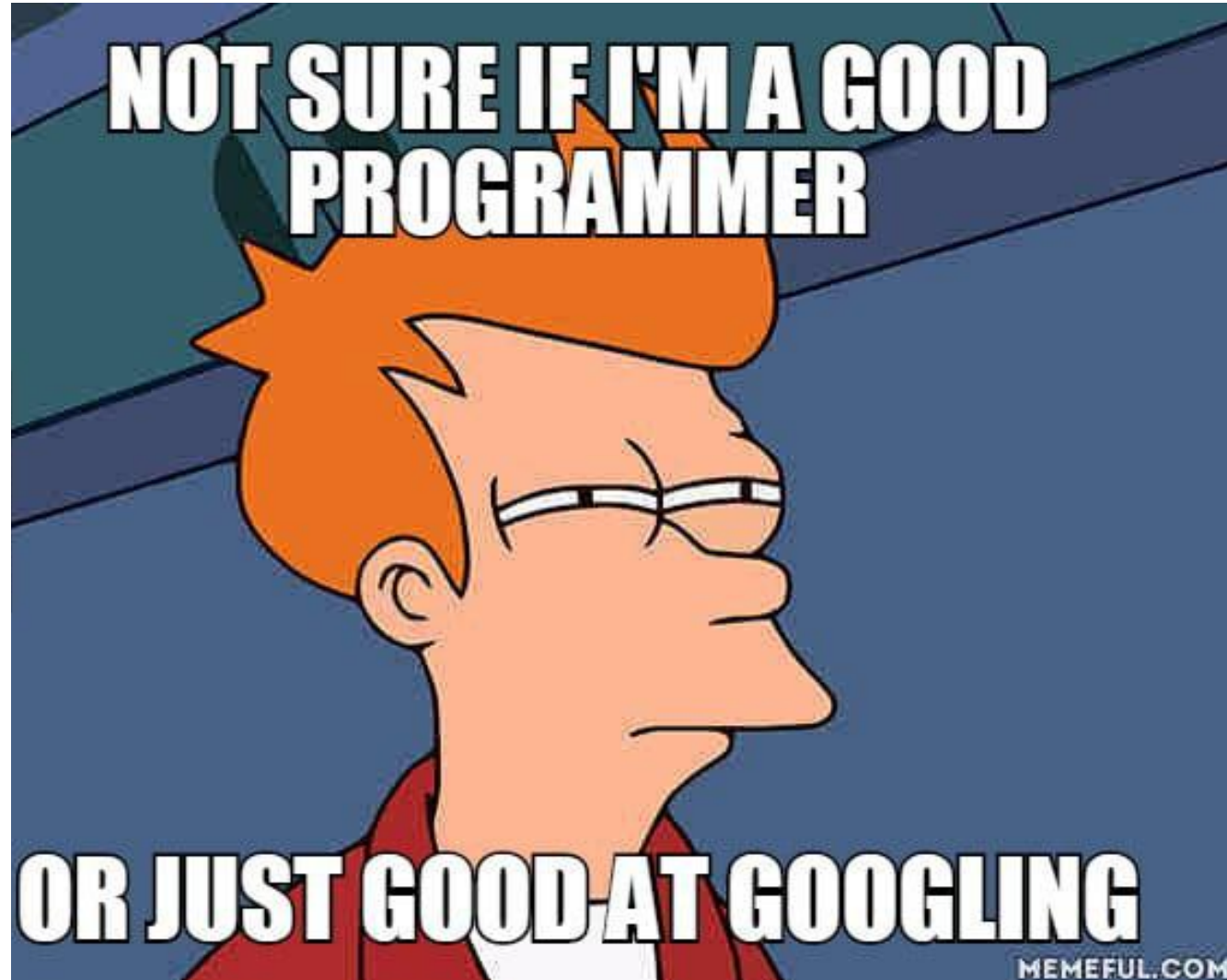


# Advantages of using R...



- R is *open source!!*
- R is **actively developed** by a community
- R runs on **multiple platforms** (Windows/MacOS/Linux)
- R can work with much **larger datasets**
- R is far more **reproducible** because of its **scripting** capabilities
- R is **widely used**

You will always get help...



# What is Rstudio?



- RStudio is a convenient interface for R (an *integrated development environment, IDE*)
- At its simplest:
  - R is like a car engine
  - Rstudio is like a car's dashboard

R: Engine



RStudio: Dashboard



**RSTUDIO**



# What is Rstudio?



- RStudio, like most IDEs, provides a **graphical interface to R**, making it more user-friendly,
- Rstudio provides dozens of useful features (can clearly view graphs, data tables, R code, and output all at the same time...)



**RSTUDIO**

# Why learn R and Rstudio?



Imagine it is a container with many different tools (**functions**)



Each function is suited for one task



We can combine many functions to perform less trivial tasks



# Why learn R/Rstudio in Bioinformatics?



- R is one of the most widely-used and powerful programming languages in bioinformatics.
- R especially shines where a variety of statistical tools are required (e.g. RNA-Seq, population genomics, etc.)
- R is very handy in the generation of publication-quality graphs and figures.





# Rstudio Interface



The screenshot shows the RStudio interface with the following components:

- Script editor pane:** A text editor containing R code for loading libraries, reading data, and defining a task. A red underline is under the text "Script editor pane: suite of commands".
- Environment pane:** Displays the current R environment, showing the 'Data' environment with 1000 observations of 21 variables. A purple underline is under the text "Environment pane: R objects in memory".
- Console pane:** Shows the execution of the code from the script editor, including an error message: "Error in lapply(X = X, FUN = FUN, ...) : object 'german\_short' not found". A red underline is under the text "Console pane: sandbox, interactive area".
- Files, plots, help pane:** A pane at the bottom right for managing files, plots, packages, help, and viewer. A green underline is under the text "Files, plots, help pane".



*Panes*

Now let's login  
to Rstudio Cloud



<https://rstudio.cloud/>