# 

**Getting Started with R using Posit Cloud**

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_ Period: \_\_\_\_\_\_

**For each activity below:**

✔️ means to complete this task

✏️ means to write an answer here

# **What is R and Posit Cloud?**

# 

**R** is a programming language that allows you to perform various statistical analyses and create graphs and plots to better understand your data. It's like having a calculator for statistics that can do much more!

**Posit Cloud** is a platform that provides a user-friendly interface to work with R using a website. You do not need to install any software on a computer to use R through Posti Cloud.

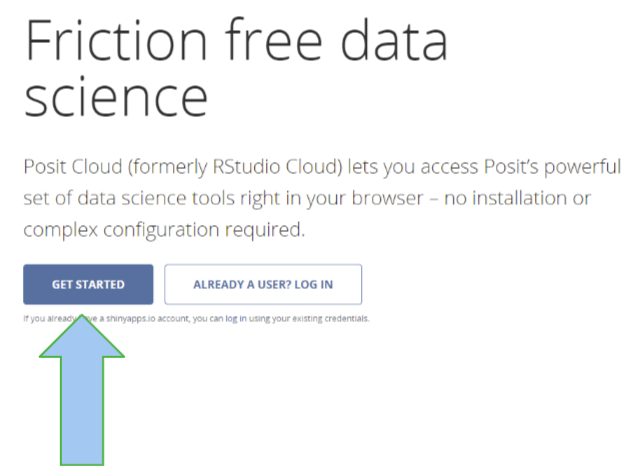
With R and Posit Cloud, you can explore real-world data, make discoveries, and solve problems in exciting fields like biology, data science, and computer science.

# **Activity #1: Accessing Posit Cloud and Creating an Account**

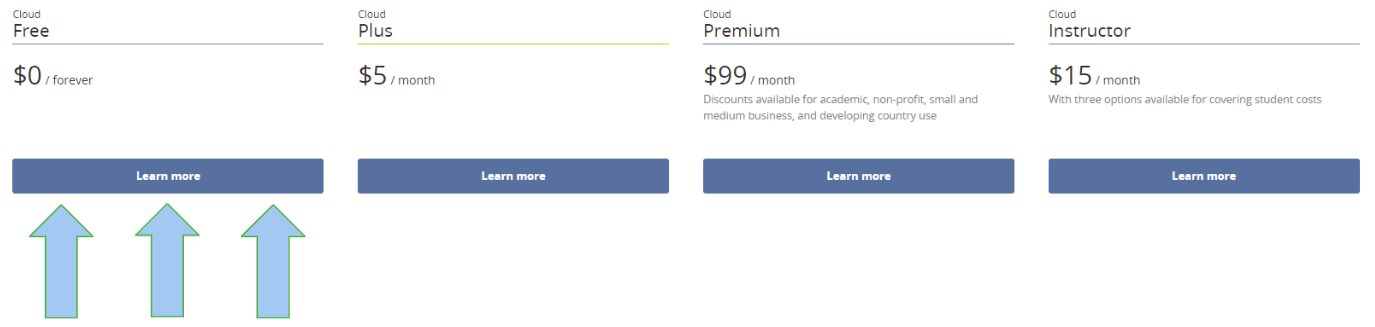
Before we can get started with code, we need to access the Posit Cloud as a programming language IDE to do the functions we want it to do.

Let’s start!

1. ✔️ Go to <https://posit.cloud/>
2. ✔️ When you get to the page, click on the button “**Get Started**”



1. From there you come to a page that lists pricing options. Click on that “**Learn More**” button **UNDER THE CLOUD FREE OPTION**!



# 

1. ✔️ Then click on the navy blue “**Sign Up**” button which brings you to a page to create an account.
2. ✔️ Create an account either through the “Sign up with Google” button if you have a Google account or by filling in credentials (Email, Password, First Name, Last Name)
3. ✔️ Once you have created your account, you will be taken to your Workspace which will store all your projects

# 

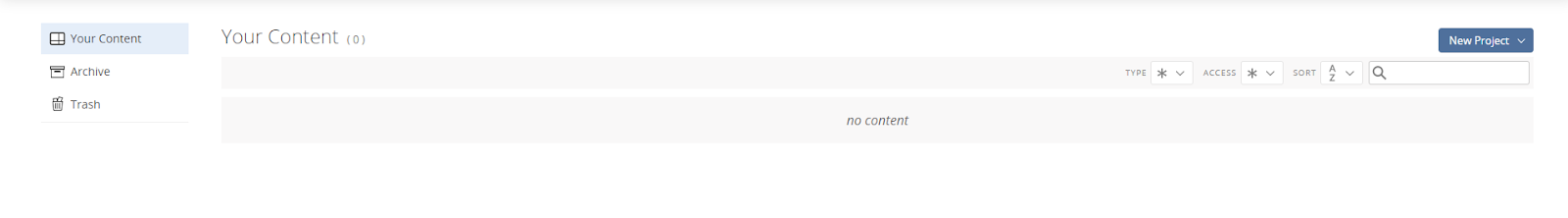
# **Activity #2: Getting Comfortable with the Workspace and Working Directory**

Now it is time to understand the Posit Cloud workspace and a few basic mechanisms that will be built on for later exercises

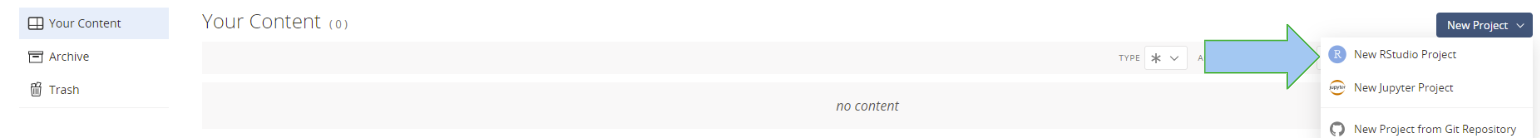
**Workspace and Working Directory**

Let’s start by creating your first project.

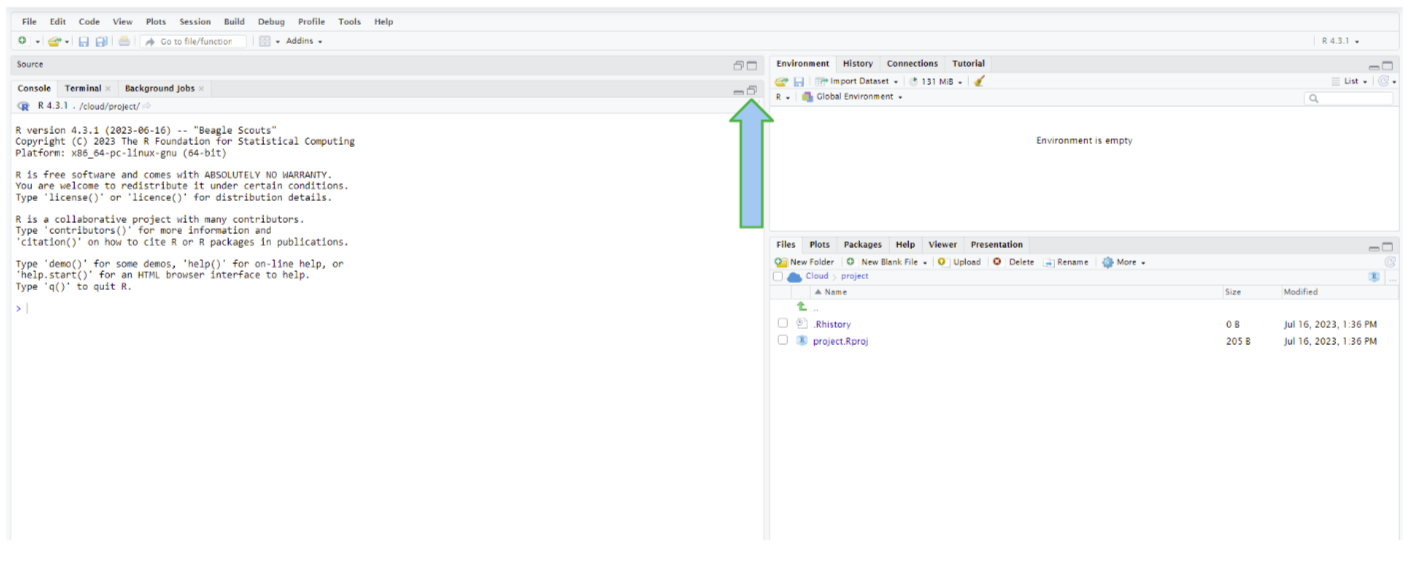
1. ✔️ After creating your account, you will have come across a somewhat blank page. This is your **workspace** where all your Posit Cloud projectsare stored

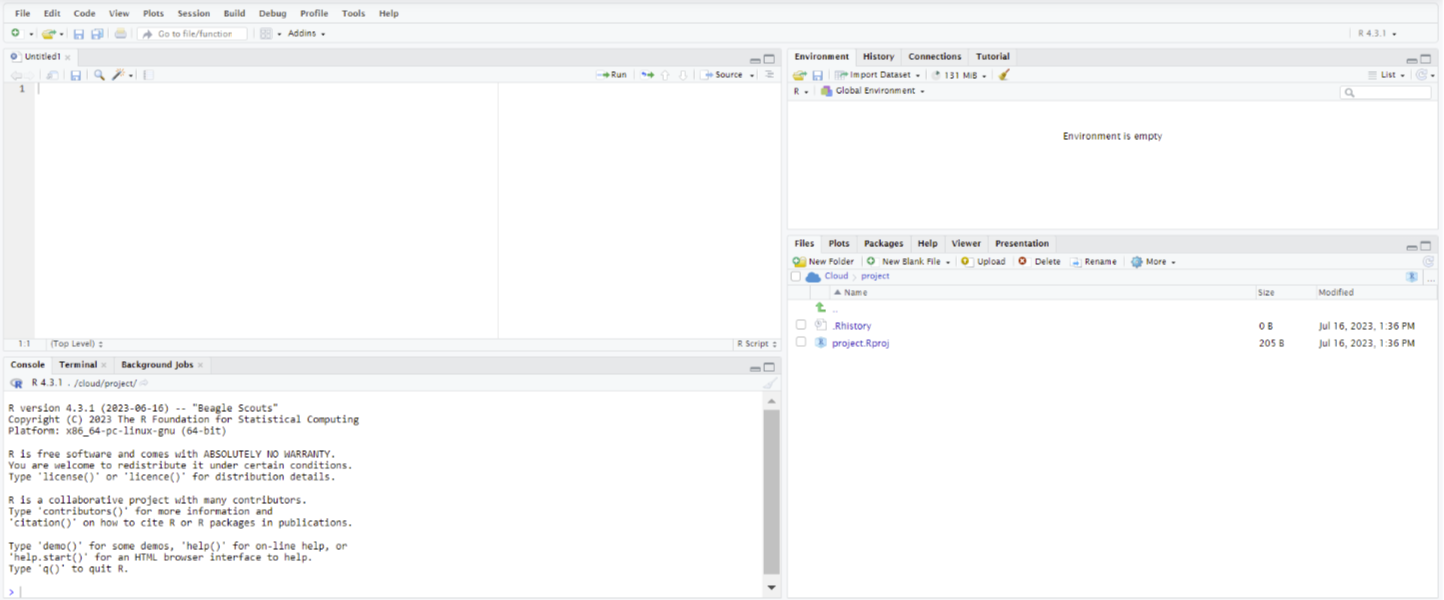


1. ✔️ To create a project, at the top right of the page, click on navy blue “New Project” and then the option “New RStudio Project”



1. ✔️ Here you come to your **working directory.** Click on thebutton that the arrow below points to so that all 4 major panels are visible



1. ✔️ Your working directory should look something like this. 
2. ✔️ The panel in the upper left is the **Source panel**, the panel where you will write your code to perform certain functions. You see in the source panel a blank tab. The tab above is your **script**, a file that contains all your code. The panel on the lower left is the **Console panel** which shows the results of the code you ran in the source panel. This helps show your code does not have any errors after you run it. The panel on the upper right is the **Environment panel** which shows the variables and datasets that you have created and assigned. The panel on the lower right has different tabs that are used in different circumstances. We will mainly use the **plots** tab to show us the visualizations we will create from the datasets.

# 

1. ✔️ To create a new script, go to the top left, select **File** → **New File** → **R Script**
2. ✔️ After you do that, go to the top left, select **File** → **Save As** to save your file. A Save As window opens, allowing you to choose a name and file extension for your document. Call the file “firstProgramOnBasics”

* Tip: When naming programs it is helpful to be consistent. For our naming convention, the first letter of the first word is lowercase (“first”) but other words after it starts with an uppercase letter ( “Program”, “On”, “Basics”).

# **Activity #3: Creating Print Statements**

In Posit Cloud, a **print statement** is a command that displays or prints a specific message or the value of a variable on the screen. Print statements are important because they allow you to see and verify the output of your code, making it easier to understand how your program is working and helping you identify and fix any issues or errors in your code. They are set up like this: **print(“ “)** where text is placed between the quotes.

Let’s try out a **print statement**

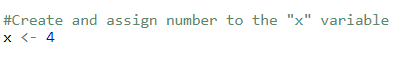
1. ✔️ In the Source panel, type **print(“ Hello World “)**
2. ✔️ Place your mouse cursor at the end of the closed parentheses, HOLD down the “Ctrl” key (on the bottom left of your keyboard), and then click the “Enter” key to run the line of code
3. ✏️ What do you see happen? Write it down in the space below.
4. ✏️ Highlight the same line of code of **print(“Hello World“)** and then click the green “run” button at the top of the Source panel. What happened? Was the result the same?

**Congratulations, you ran your first line of code!**

# **Activity #4: Creating and Assigning Variables**

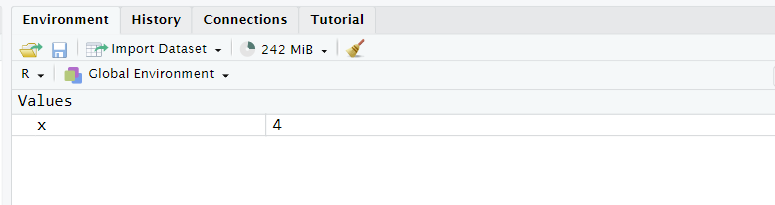
**Variables** are essential elements in Posit Cloud that store and manipulate data. They allow users to efficiently organize, manipulate, and analyze data by providing a means of data storage, organization, and computation. Variables play a crucial role in enabling reproducibility and transparency in data analysis, as they facilitate code readability and the ability to share and reproduce analyses.

1. ✔️ In the Source panel, choose the letter “x” in the alphabet and type it in lowercase
2. ✔️ Give it a space and type the characters **<-** after it. This lets the program know you are assigning this variable a value.
3. ✔️ Give the variable a numerical value between 1-9. You can choose the value. Run the code.

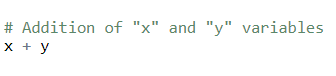


* In R, comments are lines of text that provide explanations or reminders within the code but are not executed during runtime. In R, a comment is created using a **#**. They enhance code readability and promote effective collaboration among programmers.

1. ✔️ After you have created and assigned your “x” variable, you will notice that the variable gets added to the Environment panel.



1. ✔️ Repeat steps 1-3 for creating another variable. Call the variable “y”. Give the variable “y” a different number from your “x” variable.
2. ✔️ Let’s perform some basic mathematical functions using the variables you created. Let’s add the variables. To do this, run this code:



1. ✔️ Great, you were able to perform addition. Redo steps 1-5 but assign different numbers for “x” and “y” and try different mathematical operations with the variables

| R command | Mathematical operator | Example |
| --- | --- | --- |
| + | +(addition) | 4 + 2 = |
| - | - (subtraction) | 4 - 2 = 2 |
| \* | X (multiply) | 4 x 2 = 8 |
| / | ÷ (divide) | 4 / 2 = 2 |
| ^ | exponent | 4 ^ 2 = 16 |

1. ✏️ In Posit Cloud, write up some code to perform a mathematical function between your variables. Run the code. Paste the code and the results of the code below. What happens? Does the result match your prediction?
2. ✏️ Try changing the numerical value of the variables and run the code with different mathematical functions. What happens? Does the result match your prediction?

# **Activity #5: Using Datasets and Dataframes**

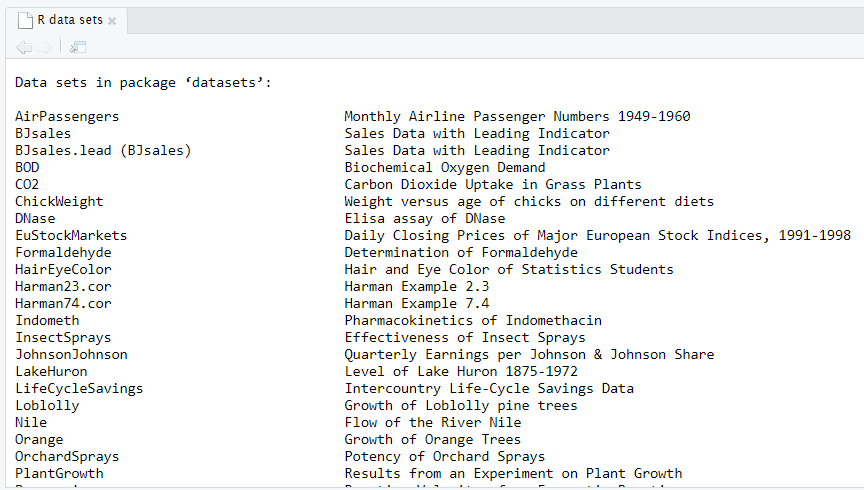
A **dataset** refers to a structured collection of organized data points or observations. It can include various types of information, such as numerical measurements, categorical variables, textual descriptions, or a combination of these. Datasets are used in research, analysis, and various fields of study to examine patterns, relationships, and trends within the data.

In R, we use something called a dataframe to store our dataset. A **dataframe** is a table that helps us organize the data into rows and columns. It is a useful way to work with data, allowing us to easily analyze and visualize the information we have.

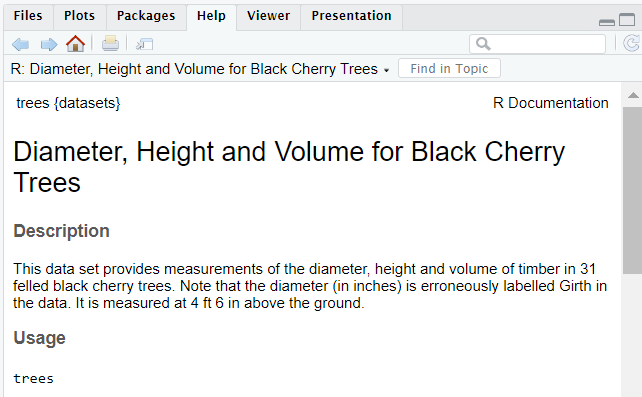
In R, a **vector** is a data structure that represents a single variable and can be thought of as a column in a table. Just like a column holds the same type of information for each observation (row) in a dataset, a vector contains a sequence of values of the same data type, all of the same variable of interest. It provides a convenient way to store and manipulate a collection of related data points for analysis and computations.

Let’s use a built-in dataset in R as an example.

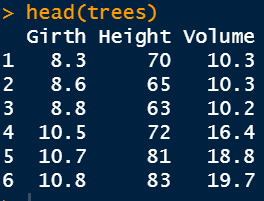
1. ✔️ In the Source panel, type **data()** and run this line. You will then see a new tab on the source panel that looks like this:



1. ✔️ Scroll to the bottom of the page until you come across a dataset named trees.
2. ✔️ You do not know what the “trees” dataset is. Since it is built in R, you can use the **help()** command to help you understand what the “trees” dataset means. Type **help(trees)** into the Source Panel and run this line. A pop-up in the bottom right panel should appear. It should look like this:

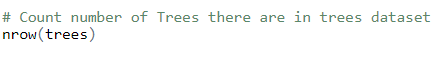


1. ✏️ Read what the “trees” dataset is about. How many variables are in the dataset? What does each variable represent?
2. ✔️ You want to take a look at the dataset. To do so, type in the command **head(trees)**. You should get this in your console panel:



* **Dataframe** refers to a tabular structure that holds data in rows and columns. In this case, the Dataframe consists of three columns: "Girth," "Height," and "Volume."
* The columns are **vectors**, which are essentially sequences of values representing specific attributes or measurements. Together, these vectors collectively form the dataset, with each vector contributing specific information about the observations.

1. ✔️ So you now understand what the “trees” dataset is about. But we do not know how many trees were recorded in the dataset. To do this, type the following **nrow(trees)** command into the Source panel which gives the number of rows in the dataset that is equal to the number of trees in the “trees” dataset.



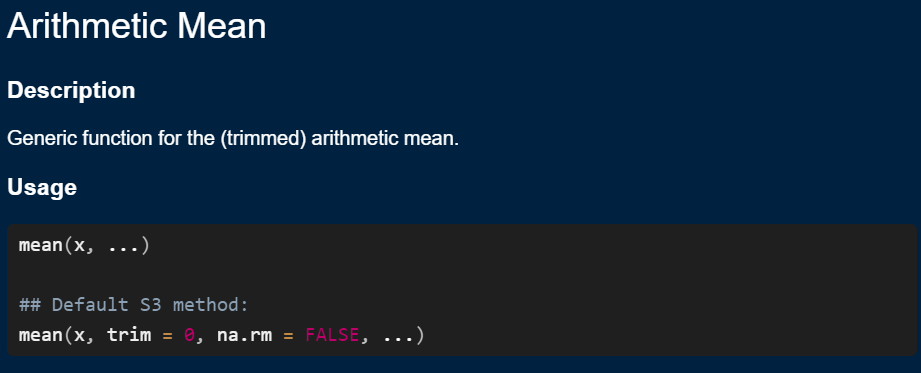
1. ✏️ How many trees are in the dataset?

# **Activity #6: Calculating Descriptive Statistics**

Let’s perform some descriptive statistical analysis of a variable from the dataset, - let’s find the mean and standard deviation of a variable

* **Mean** - the sum of values divided by the number of values in a dataset
* **Standard Deviation** - the amount of variation or spread in a dataset

1. ✔️ You want to find the mean of the “Girth” variable of the “trees” dataset. But you do not know what the command is to find the mean. To find the command for mean, type **?mean** command into the Source panel. You should see this pop-up in the bottom right panel:



1. ✔️ Read the description in the bottom right panel.
2. ✔️ Based on the description, write a command for finding the mean of the “weight” variable of the “trees” dataset. Here’s what the code should look like:



* trees$Girth is a reference to a vector (column) in the trees dataset that represents the Girth measurements of different trees

1. ✏️ What is the mean of the “Girth” variable of the “trees” dataset?
2. ✔️ Repeat steps 4-6 but this time, find the **standard deviation** of the “Girth” and “volume” variables of the “trees” dataset.
3. ✏️ What is the standard deviation of the “Girth” variable of the “trees” dataset?
4. ✔️ Great! We found the mean and standard deviation of one variable. If you wish to find descriptive statistics for all the variables of the dataset you can use the

**summary ()** command.

1. ✔️ Type the code **summary(trees)** into the Source panel and run the line. Look at the variables “weight” and “Time” in the Console panel.
2. ✏️ What are the minimum, median, mean, and maximum of the “Girth” variable?
3. ✏️ What are the minimum, median, mean, and maximum of the weight variable of the “Height” variable?
4. ✏️ What are the minimum, median, mean, and maximum of the weight variable of the “Time” variable?