

Preliminary tasks

Mathieu Ferry

Preliminary tasks for W2

Welcome to the computational/quantitative/mixed methods workshop!

As part of this workshop, we (Anupam Das & I) are going to work on textual analysis. In my case, I will focus on automatic exploratory statistical tools. To this end, we will build from a research work I have been conducting with Jeanne Subtil (Sciences Po/CRIS) on online matrimonial advertisements in India. We explored online marriage-making strategies of Uttar Pradesh-based Hindu spouse-to-be individuals registered on the [Jeevansathi platform](#). Our working paper is currently available here: INSERTLINK.

Before the workshop, I request you to complete two tasks:

1. Read the working paper

The paper explains our research questions, how we constructed our database and how we conducted our analysis. We will expand on this research during the workshop and we will analyse a subset of the data. It is important to read the paper beforehand so as to have a better understanding of what we are working on.

2. Get ready to conduct statistical textual analysis during the workshop

This involves two points:

- Download the sample database used during the workshop (available on our Google Drive or from here (INSERT LINK)). There are two files where our data are stored: “sample.csv” and “sample.rdata”. These are exactly the same files except that the “sample.rdata” is a format native to R (the software we use) and allows to save more pre-formatting characteristics of the data so it will be this one we will use in our session, but you can start to have a look at it by opening the csv file.


- Download the “TextualAnalysis.R” file. It contains the R script we will use during the workshop. If you are not familiar with programming and/or R language, don’t worry! Though we will need these lines of code, we will not spend time actually writing code lines during the workshop and we will rather play with apps (“shiny apps”) that are more user-friendly.

If you are already an R user and you are familiar with importing data in RStudio from your computer, feel free to use RStudio from your computer during the workshop and ignore the following instructions.

If you are *not*, do not worry! To avoid any technological complications, we are going to use R/RStudio from a web server for which you need to create an account. Note that you *don’t* need to know R to follow this workshop and this workshop is *not* a R workshop. Still, if you are curious about R, you can check this webpage which has a lot of resources: [RStudio Education](#).

For our workshop, follow the following steps:

1. Go to the [Posit server website](#) and click on “sign up”:



Log In

Friction free data science

Posit Cloud (formerly RStudio Cloud) lets you access Posit’s powerful set of data science tools right in your browser – no installation or complex configuration required.

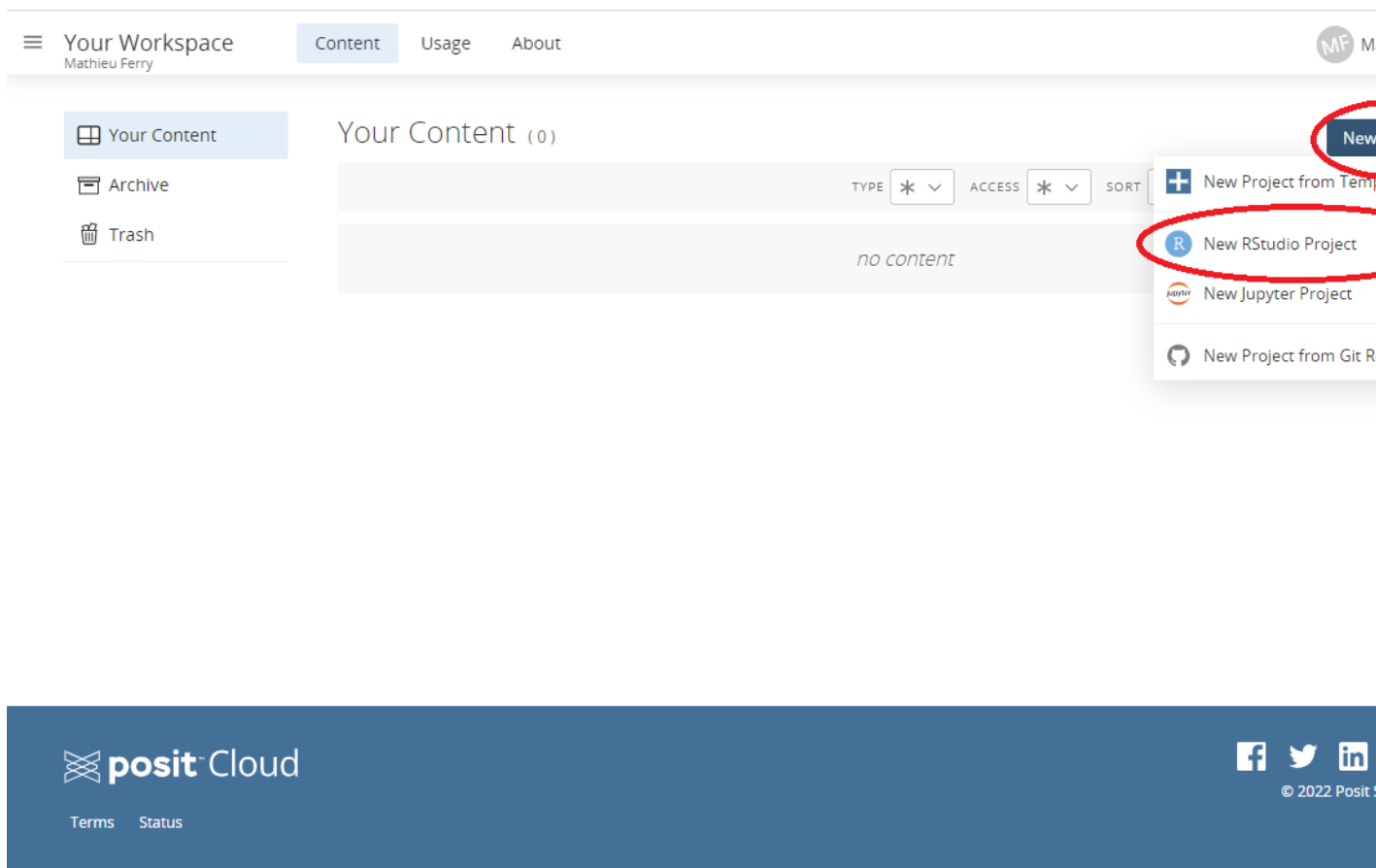
GET STARTED
ALREADY A USER? LOG IN

If you already have a shinyapps.io account, you can log in using your existing credentials.



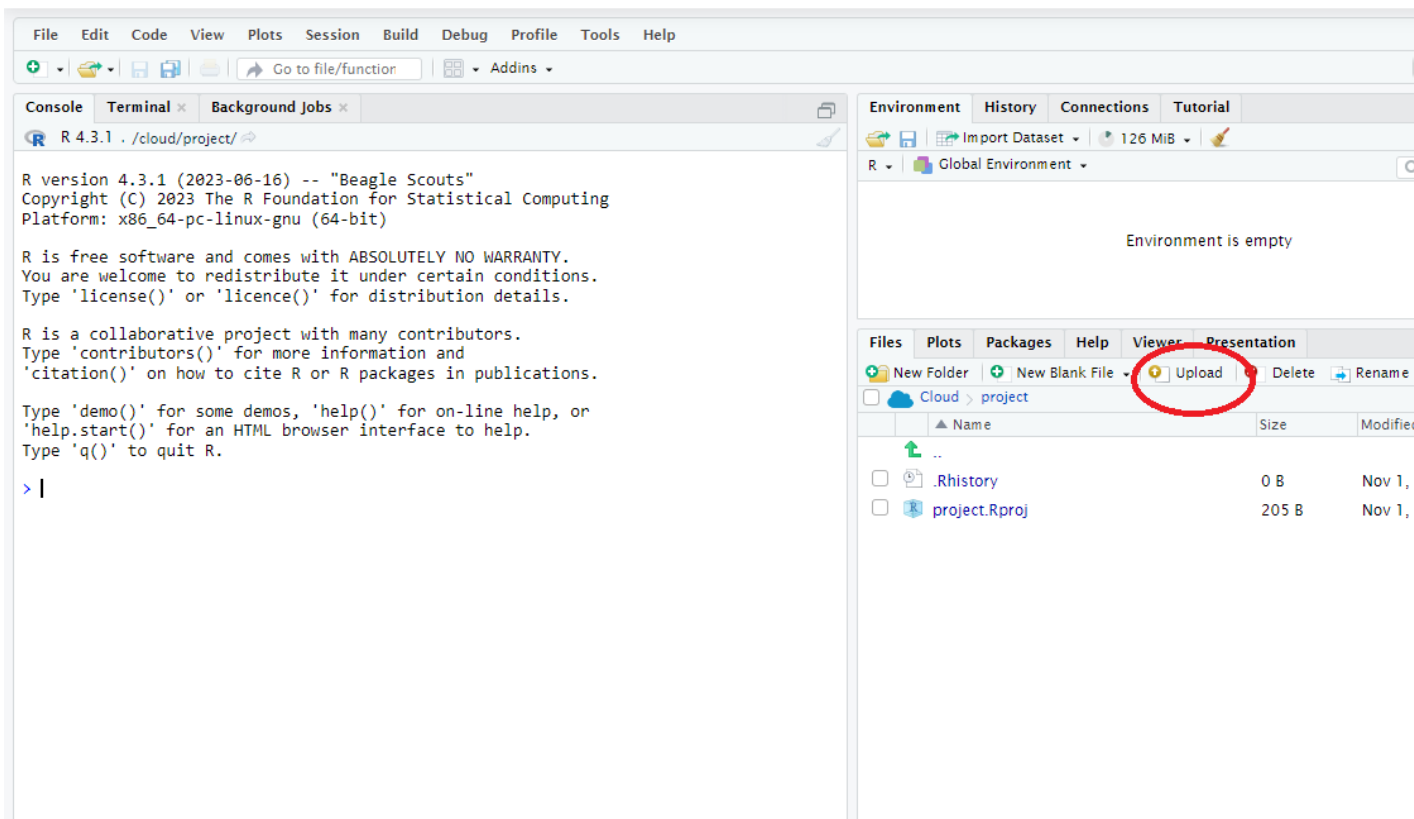
There, you can choose between several “plans”. Unless you are willing to spend money, choose the “free plan” and set up an account with your email ID.

2. After that is done, you should be able to access the cloud from where we will conduct our analysis. Click on “New project” and choose “New RStudio Project”:



You can entitle this project as “Pondicherry workshop” if you wish.

3. If everything has worked fine, you are now in the RStudio cloud. The left panel is the “console”: we run commands from there to ask the software to conduct our analyses. The top right panel is your “Global environment”, that is where the data will appear once you have uploaded it in RStudio. The bottom right panel is your “File explorer” on the cloud.
4. You now need to upload the data (“sample.rdata”) and the R script (“TextualAnalysis.R”) to the cloud. To this end, click on the “Upload” button and select the files from your computer. Files need to be added one by one.



The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu bar is a toolbar with icons for saving, opening, and navigating files. The main window is divided into three panes: Console, Environment, and Files.

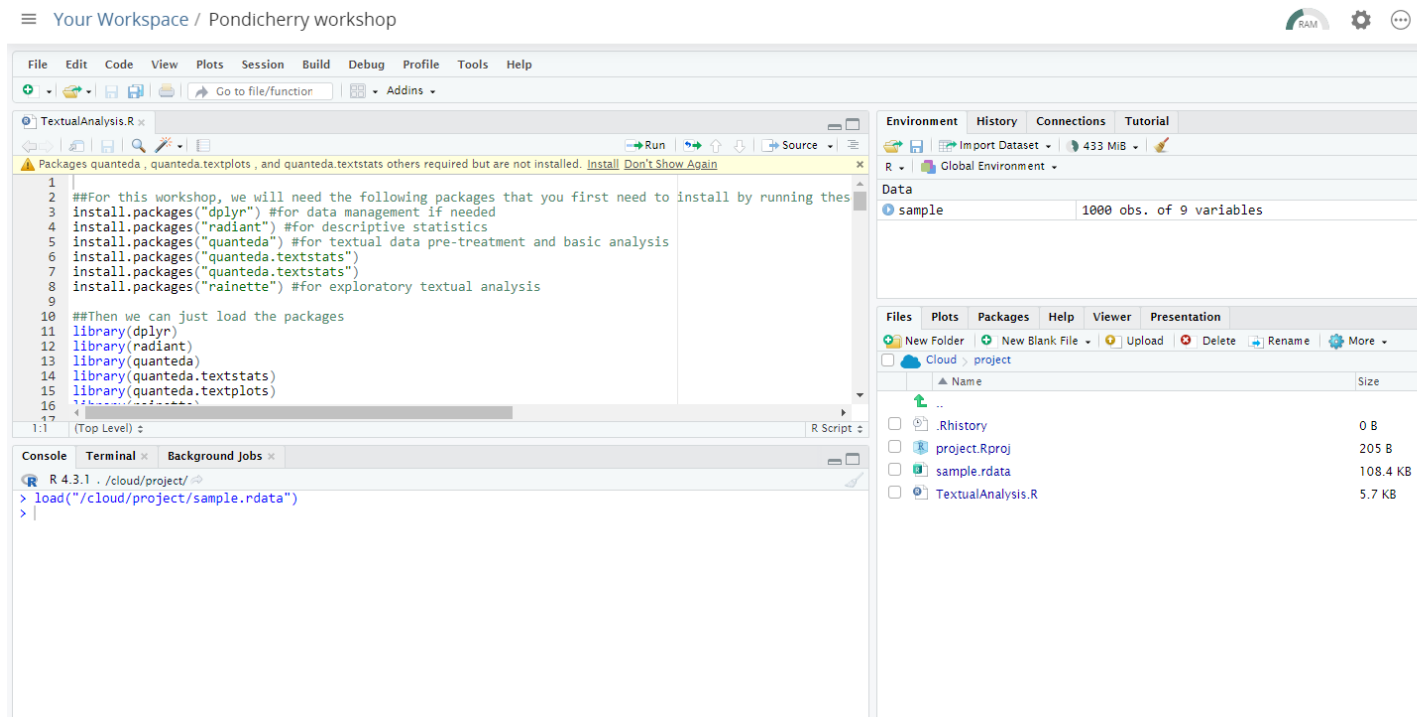
The Console pane shows the R version 4.3.1 (2023-06-16) and the R Foundation for Statistical Computing logo. It also displays the R license and a list of contributors. The prompt is `> |`.

The Environment pane shows the Global Environment and indicates that the environment is empty.

The Files pane shows the project structure. The 'Upload' button is circled in red. The file list includes:

Name	Size	Modified
..		
.Rhistory	0 B	Nov 1, 2023
project.Rproj	205 B	Nov 1, 2023

- Once that is done, you can click on your two loaded files to open the R script and load the database into the R environment. Once that is done, you should see something like this:



We are done for now! We will start from there in our workshop sessions. If you have any issue running any of these steps, please contact me at mathieukferry@gmail.com.