Importing Data Into R

Dr. Aline Coutinho

In this tutorial, you will learn how to import all sorts of data into R. The document is divided into four sections that present the major commands to import different data files.

Importing rectangular data

Rectangular data refers to a dataset that fits inside a rectangle of rows and columns. The difference between rows and columns must be clear, as the latter usually refer to variables and the former to observations. The most common format files for rectangular data are csv and tsv.

To read and import rectangular data into R, you will need the package `readr`.

This package is included in the `tidyverse` package, but below is the code necessary to install it in R:

```
install.packages("readr")
library(readr)
```

To import a dataset from a csv or tsv file, click on *Import Dataset* in your environment window. Choose *From Text (readr)*.

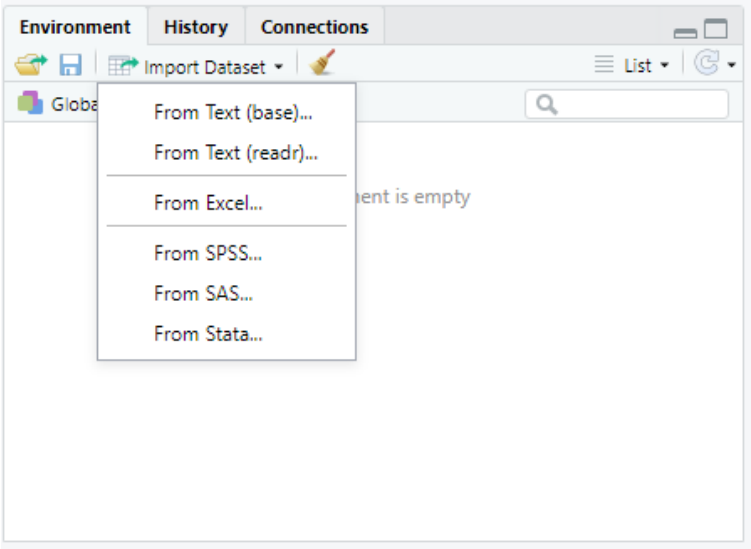

Figure 1: Import Dataset command

A new window *Import Text Data* will appear. There, you can browse the datafile in your device.

On the *Import Options* section, you can set the parameters necessary to import the data you want. For example:

- You can name the object that will be created, that is, the name of the dataset once you import into R;
- You can tell RStudio to remove certain varibles in the dataset, either by using the *skip* function or by right-clicking on the arrow beside each variable and determining it to skip that variable;
- You can delimiter the dataset by Comma (for csv files) or Tab (for tsv files).

For the purpose of reproducibility (see Tutorial), copy the code at the bottom of this window (the *Code Preview* section), and save it in your script. Click on *Import* and RStudio will import the data for you.

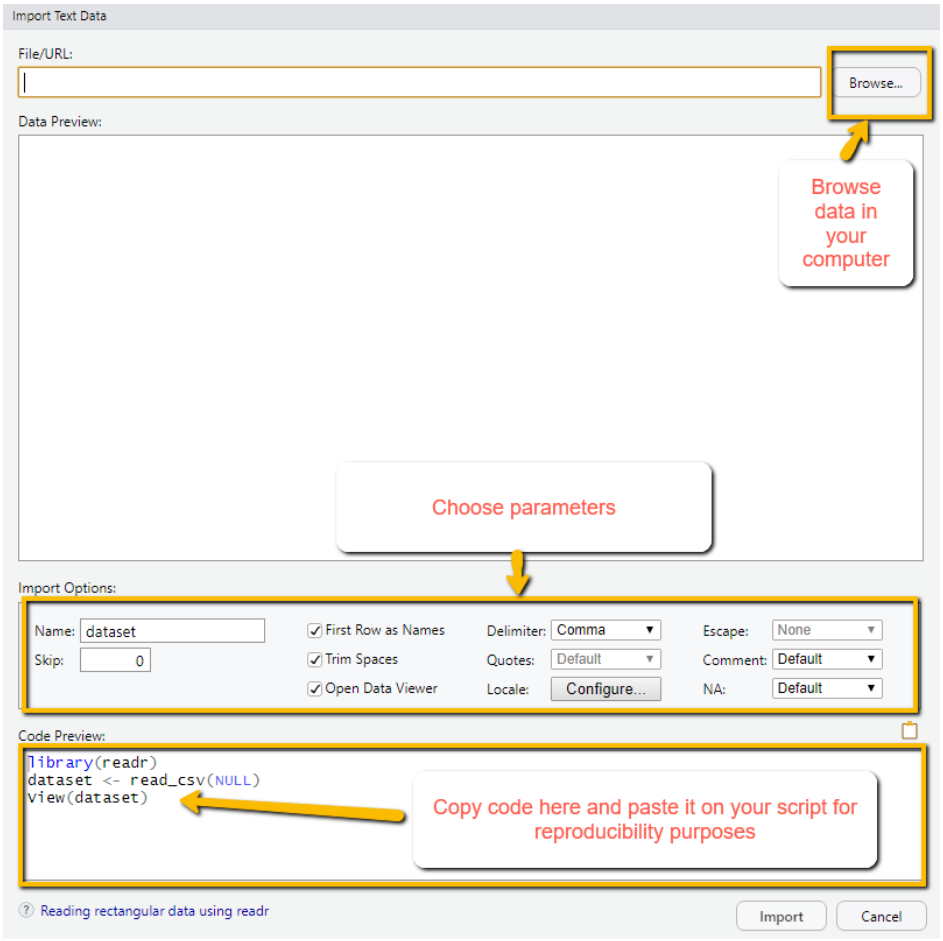

Figure 2: Specify parameters and copy code

Importing Excel files

A similar approach will be done for Excel files. However, you will need to download and load the `readxl` package:

```
install.packages("readxl")
library(readxl)
```

Click on *Import Dataset* in your environment window. Choose *From Excel* (see Figure 1).

After browsing the Excel file on yur computer, a preview of the dataset will appear in this window. You will note that RStudio is guessing the data type of each variable: character, double (the equivalent of "float"

in other programming languages such as Python or C). You can change the type of data by clicking on the little arrow.

In the *import options* section, you can define the Excel tabs you want to import, and the range of values.

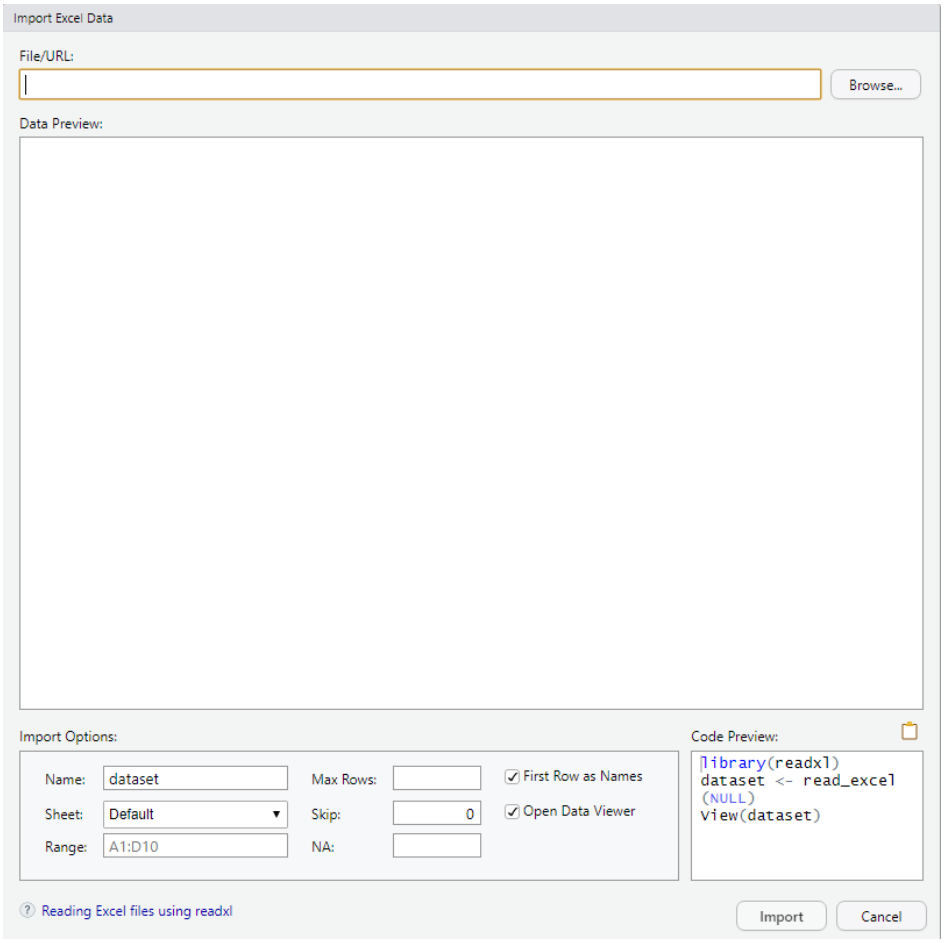

Figure 3: Specify parameters to import Excel files

Importing data from the Web

The package `rvest` allows to get html data from the web and import it into R. It is good to scrape data from websites.

As an example, suppose you want to import data on the cast of Mandalorian becase *this is the way* - see what I did here? :)

A good source for this data is the IMDB website. You can write the following command on your command window which will allow you to scrape the data for the cast of this amazing show.

```
library(rvest)
```

```
## Loading required package: xml2
```

```r
This_is_the_way <- read_html("https://www.imdb.com/title/tt8111088/")

cast <- html_nodes(This_is_the_way, ".primary_photo+ td a")

html_text(cast, trim = TRUE)
```

```
## [1] "Pedro Pascal"       "Carl Weathers"      "Gina Carano"
## [4] "Giancarlo Esposito" "Sasha Banks"
```

Let's unpack the code above:

- You first load the library `rvest`;
- You create an object called `this_is_the_way` that hosts the html page you will scrap by using the `read_html()` function;
- You create another object with the specification tag that the SelectorGadget gives you and include it in the function `html_nodes()`;
- Finally, you can trim the data you are collecting by removing extra spaces and other nuisance with the function `html_text(trim = TRUE)`.

> **The SelectorGadget:** The SelectorGadget is a Google Chrome extension that allows you to figure out what `css` (css is a web development language in charge of styling webpages) matches the data you want to scrape from a webpage.

This is a very simple a web scraping code, but it suffices for the purposes of introducing web scraping.

Importing data from other statistical software (SPSS, SAS, Stata)

There are two ways to import data from other statistical software. You can either export data from your software and save it as a csv file (see first part of this tutorial to import rectangular data into R) or you can use the built-in function in RStudio.

To import data from SPSS, SAS or Stata, you will need the `haven` package. The steps are similar to the ones outlined above. See Figure 1 for a visual guide.

Additional resources

- Click here for more information on the SelectorGadget