DATA SOCIETY™

"If you can't explain it simply, you don't understand it well enough."

- Albert Einstein

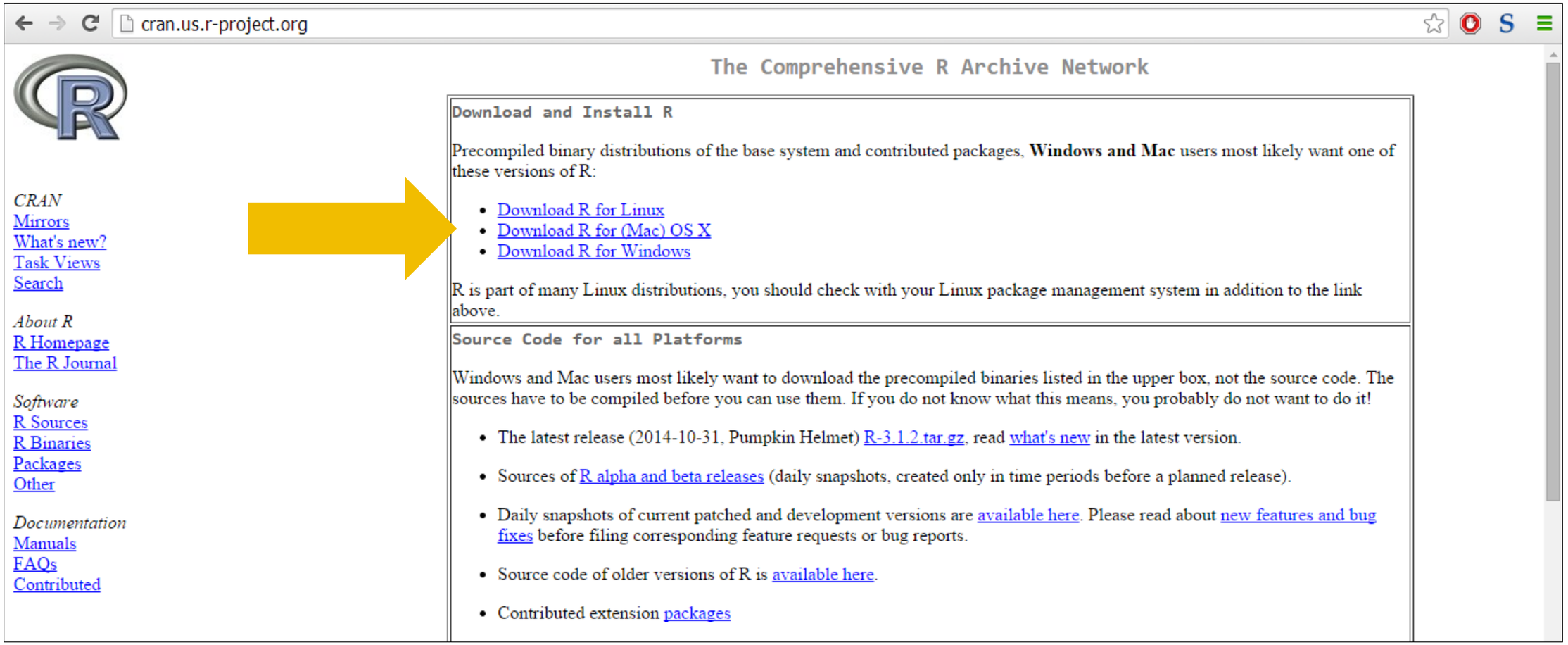
Setting up R: overview

1. What is R?
2. Download R from the CRAN website (http://cran.r-project.org/)
   - R for Windows
   - R for Mac
3. Install R Studio (http://www.rstudio.com/products/rstudio/download/)
   - RStudio a brief tour
4. Running a script
   - Variables
5. Reading in a data
   - Manually
   - Through the script

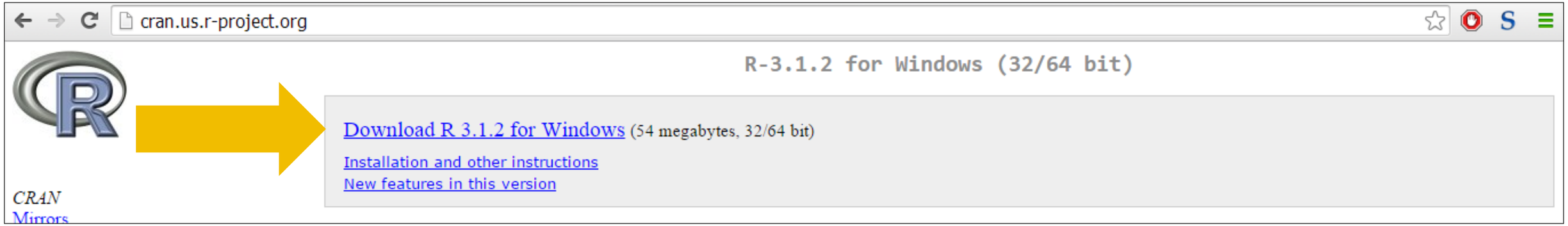
What is R?

- R is a statistical programming software
  - Has many similar features to SAS, Excel and SPSS
  - Scripting language
- It's free and open source
- Has lots of helpful pre-built functions
  - You can build your models quicker
- Easy to learn

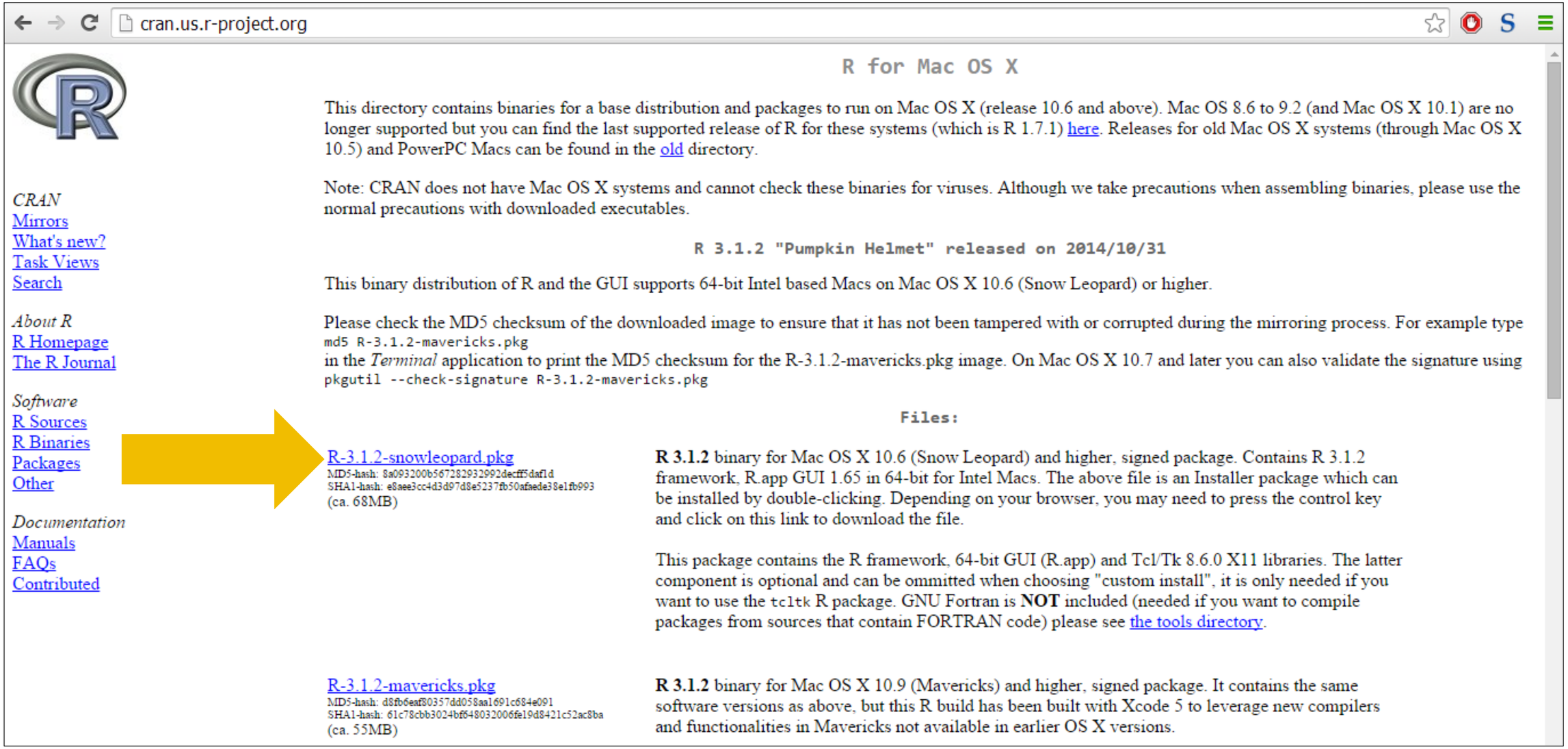
Install R


cran.us.r-project.org
The Comprehensive R Archive Network
CRAN
Mirrors
What's new?
Task Views
Search
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
Packages
Other
Documentation
Manuals
FAQs
Contributed
Download and Install R
Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:
Download R for Linux
Download R for (Mac) OS X
Download R for Windows
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.
Source Code for all Platforms
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!
The latest release (2014-10-31, Pumpkin Helmet) R-3.1.2.tar.gz, read what's new in the latest version.
Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
Source code of older versions of R is available here.
Contributed extension packages

R for Windows

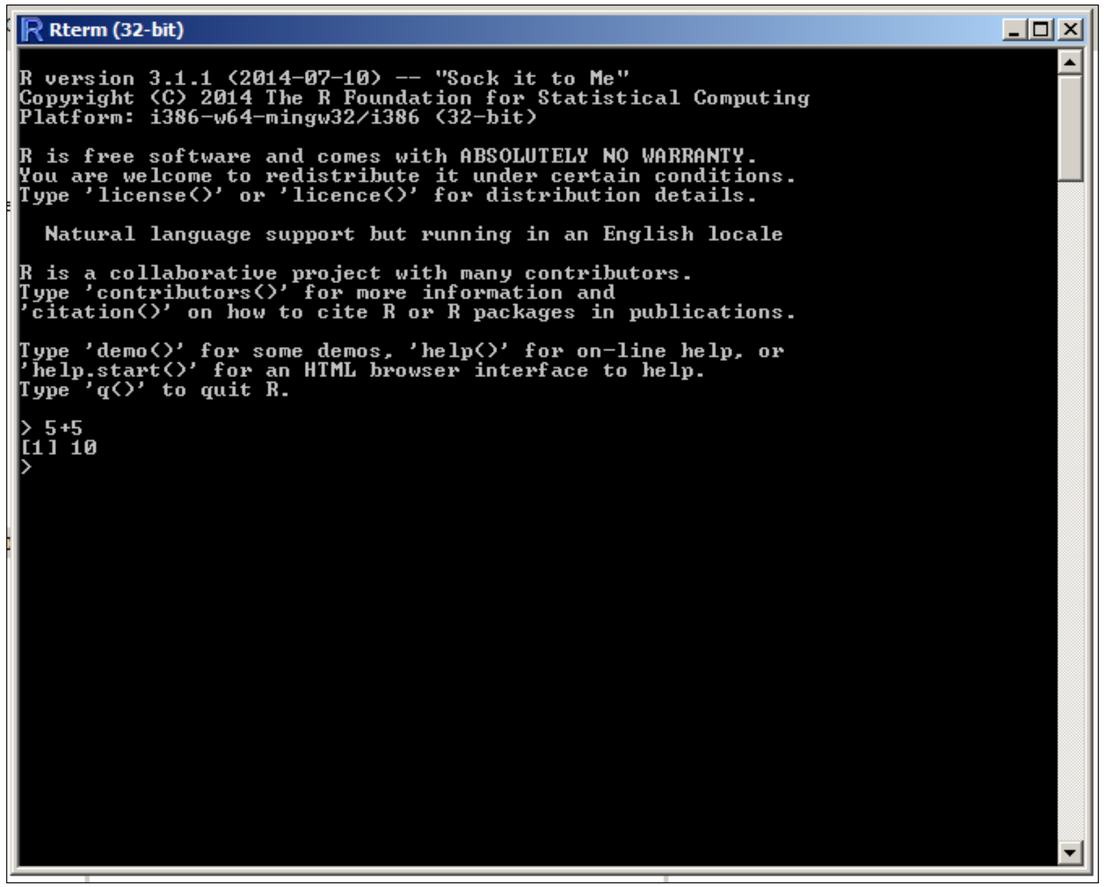


R for Mac

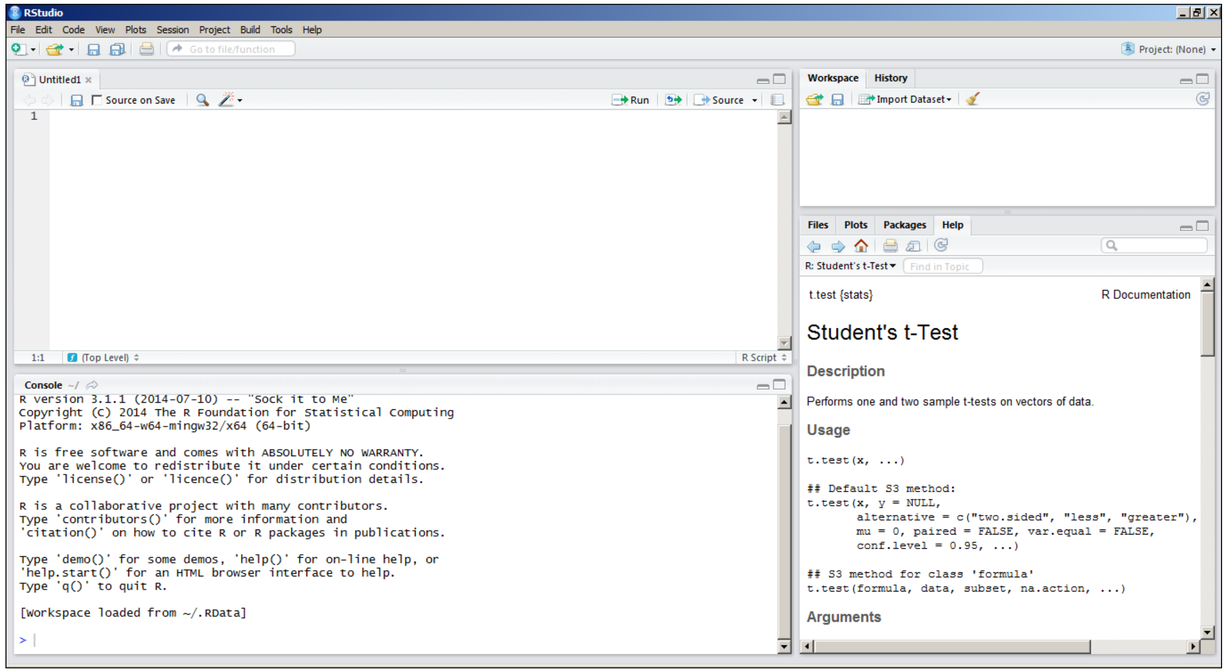

cran.us.r-project.org
CRAN
Mirrors
What's new?
Task Views
Search
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
Packages
Other
Documentation
Manuals
FAQs
Contributed
R for Mac OS X
This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) here. Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the old directory.
Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.
R 3.1.2 "Pumpkin Helmet" released on 2014/10/31
This binary distribution of R and the GUI supports 64-bit Intel based Macs on Mac OS X 10.6 (Snow Leopard) or higher.
Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type
md5 R-3.1.2-mavericks.pkg
in the Terminal application to print the MD5 checksum for the R-3.1.2-mavericks.pkg image. On Mac OS X 10.7 and later you can also validate the signature using
pkgutil --check-signature R-3.1.2-mavericks.pkg
Files:
R-3.1.2-snowleopard.pkg
MD5-hash: 8a093200b567282932992decff5daf1d
SHA1-hash: e8aee3cc4d3d97d8e5237fb50afaede38e1fb993
(ca. 68MB)
R 3.1.2 binary for Mac OS X 10.6 (Snow Leopard) and higher, signed package. Contains R 3.1.2 framework, R.app GUI 1.65 in 64-bit for Intel Macs. The above file is an Installer package which can be installed by double-clicking. Depending on your browser, you may need to press the control key and click on this link to download the file.
This package contains the R framework, 64-bit GUI (R.app) and Tcl/Tk 8.6.0 X11 libraries. The latter component is optional and can be ommitted when choosing "custom install", it is only needed if you want to use the tcltk R package. GNU Fortran is NOT included (needed if you want to compile packages from sources that contain FORTRAN code) please see the tools directory.
R-3.1.2-mavericks.pkg
MD5-hash: d8fb6eaf80357dd058aa1691c684e091
SHA1-hash: 61c78cbb3024bf648032006fe19d8421c52ac8ba
(ca. 55MB)
R 3.1.2 binary for Mac OS X 10.9 (Mavericks) and higher, signed package. It contains the same software versions as above, but this R build has been built with Xcode 5 to leverage new compilers and functionalities in Mavericks not available in earlier OS X versions.

What is RStudio?

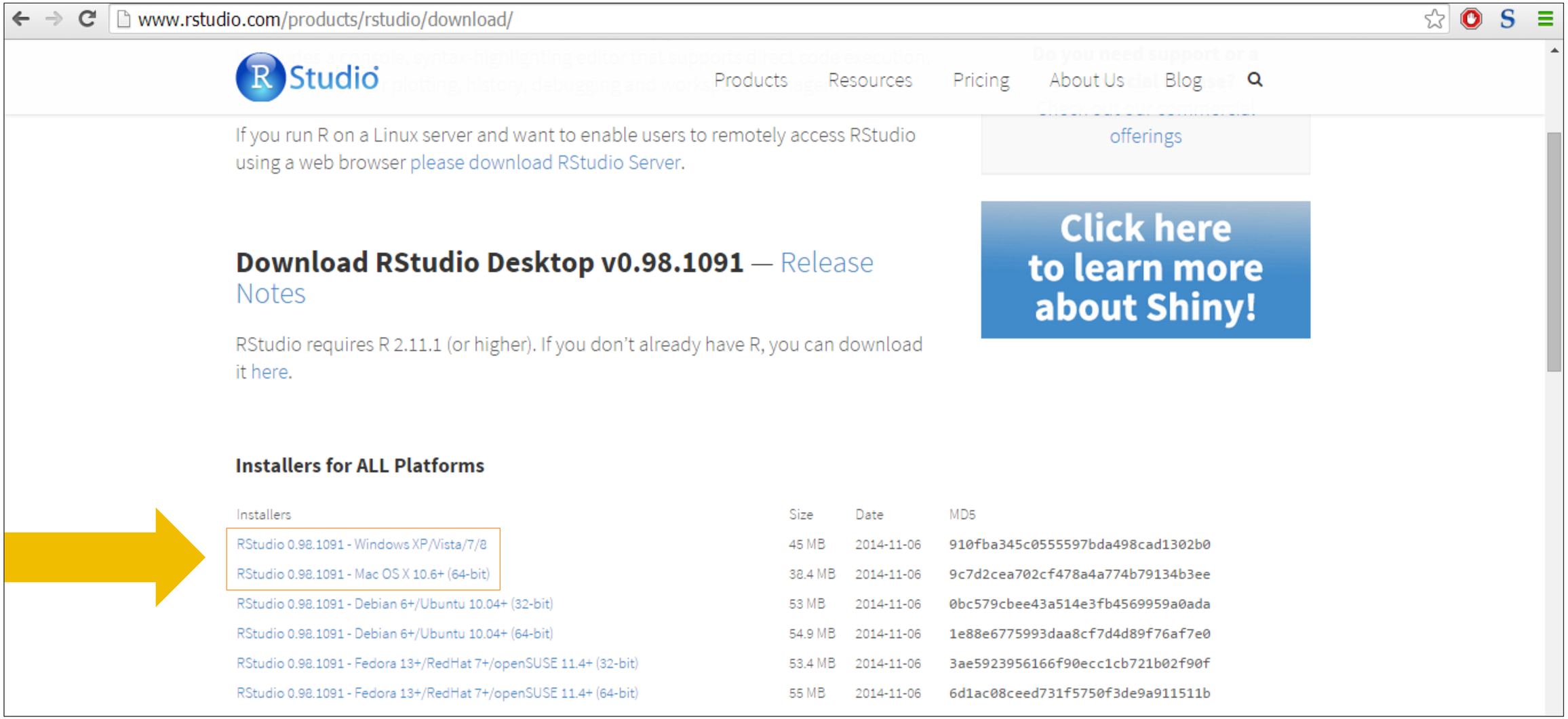
The user interface in the R terminal is bulky and non-intuitive



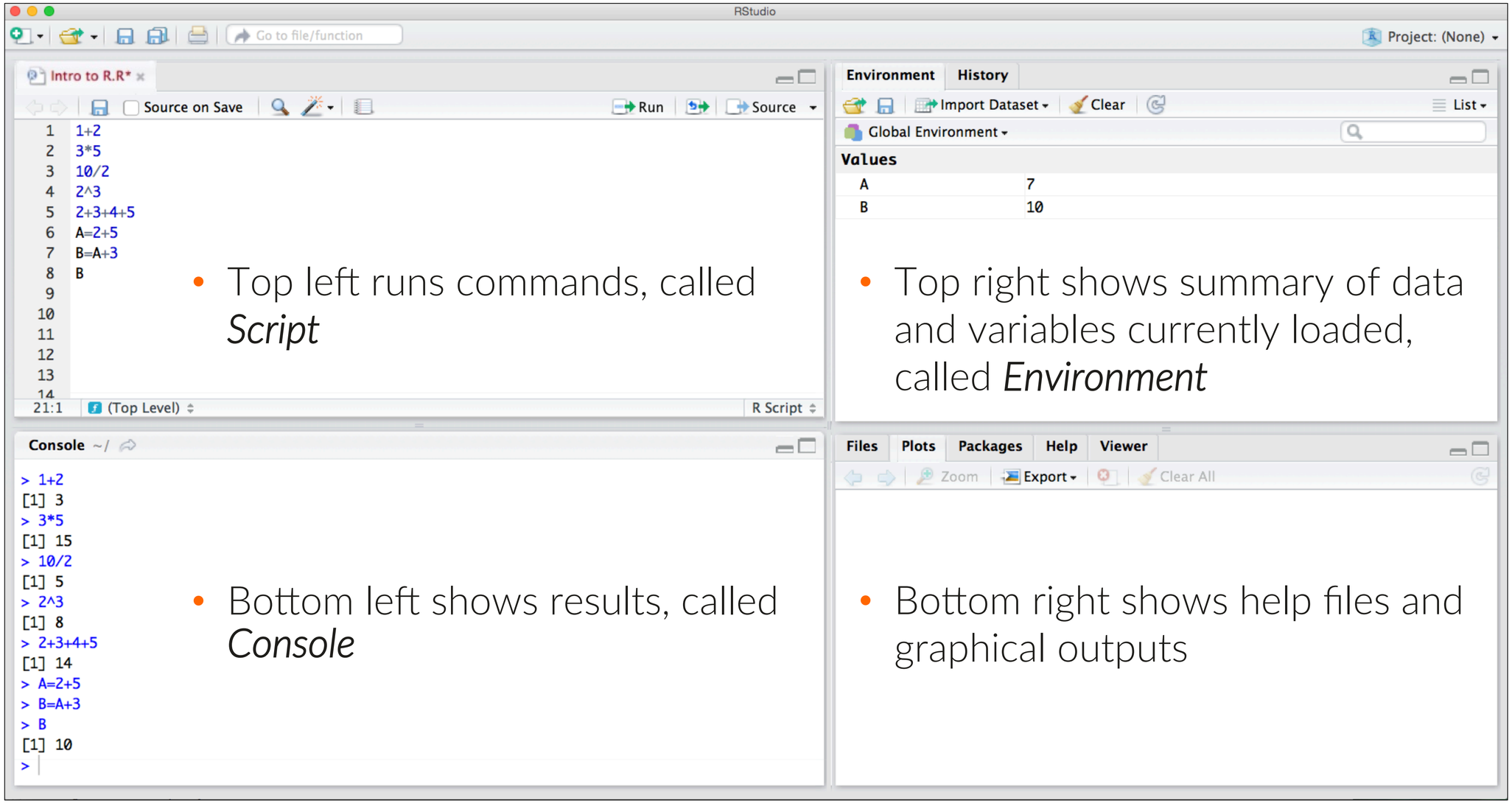
RStudio provides a better interface for a more intuitive user experience



Install RStudio



RStudio

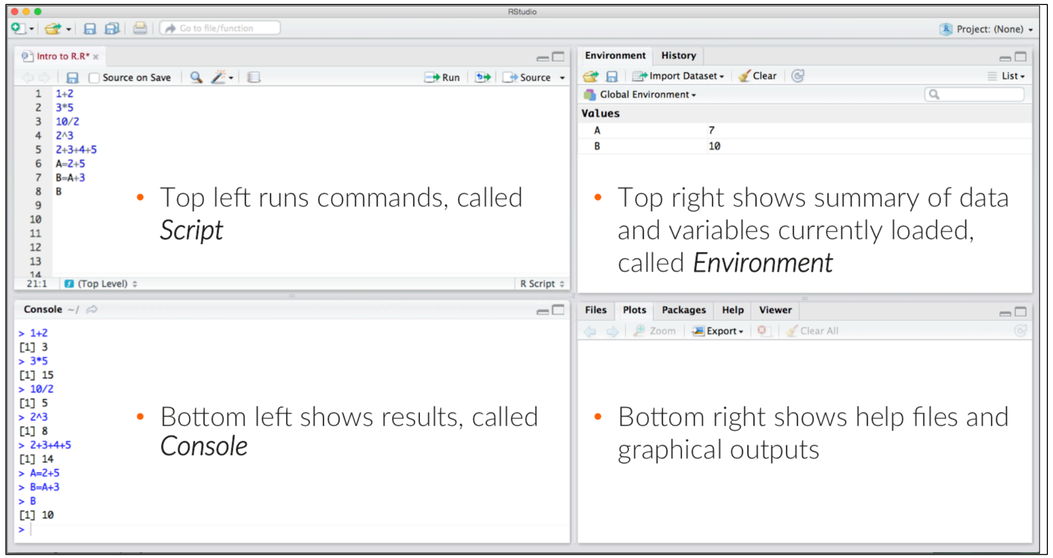


- Top left runs commands, called *Script*
- Top right shows summary of data and variables currently loaded, called *Environment*
- Bottom left shows results, called *Console*
- Bottom right shows help files and graphical outputs

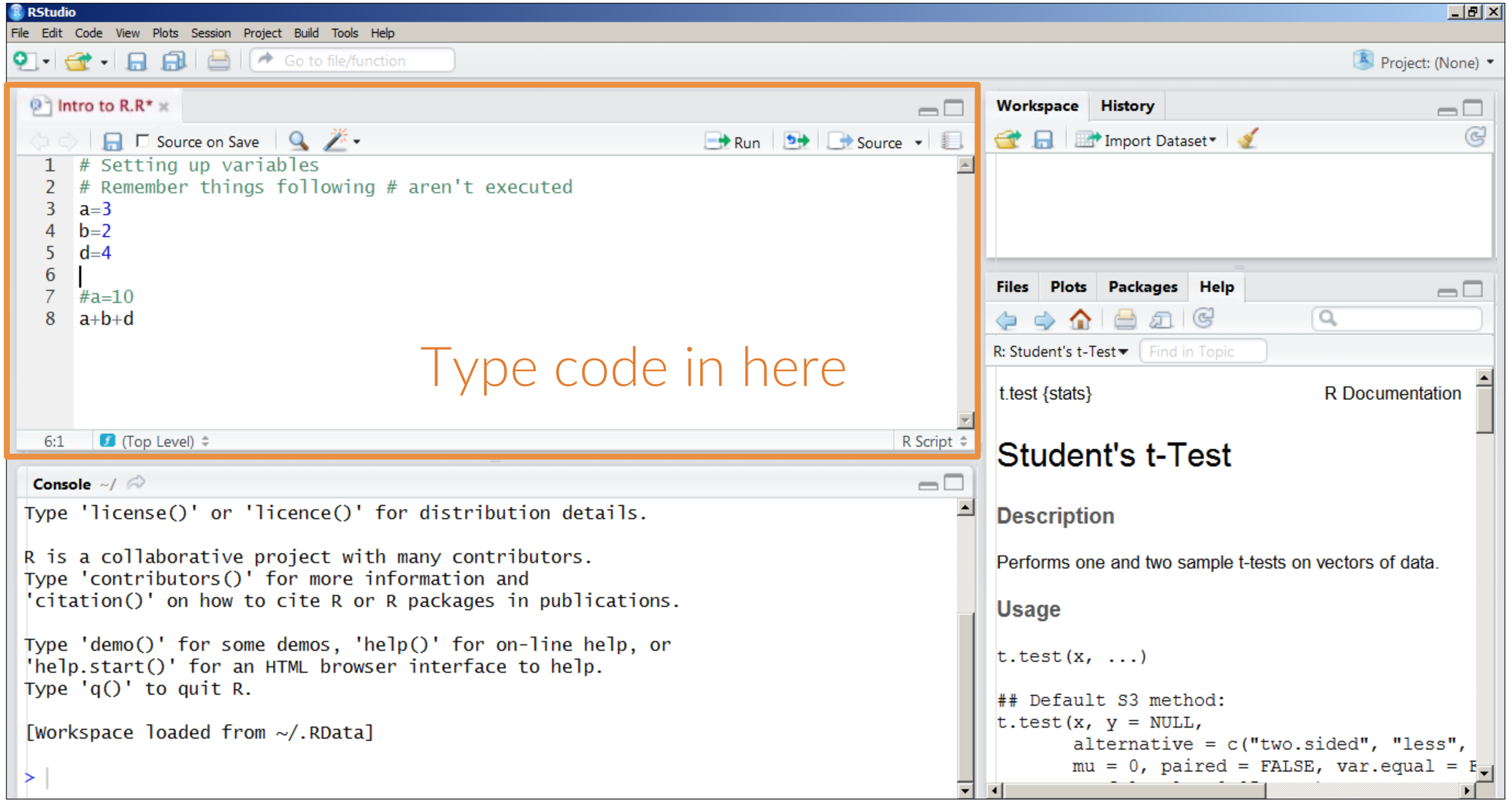
Script and Console in RStudio

- Script
  - We use scripts to generate reusable code
  - Scripts are like macros
  - Code is not executed here unless you press "Run" (see next slide)
  - Good coding practice:
    - When saving a script use a name that describes what the script does (i.e. counter, graph)
    - Comment your code using "#"

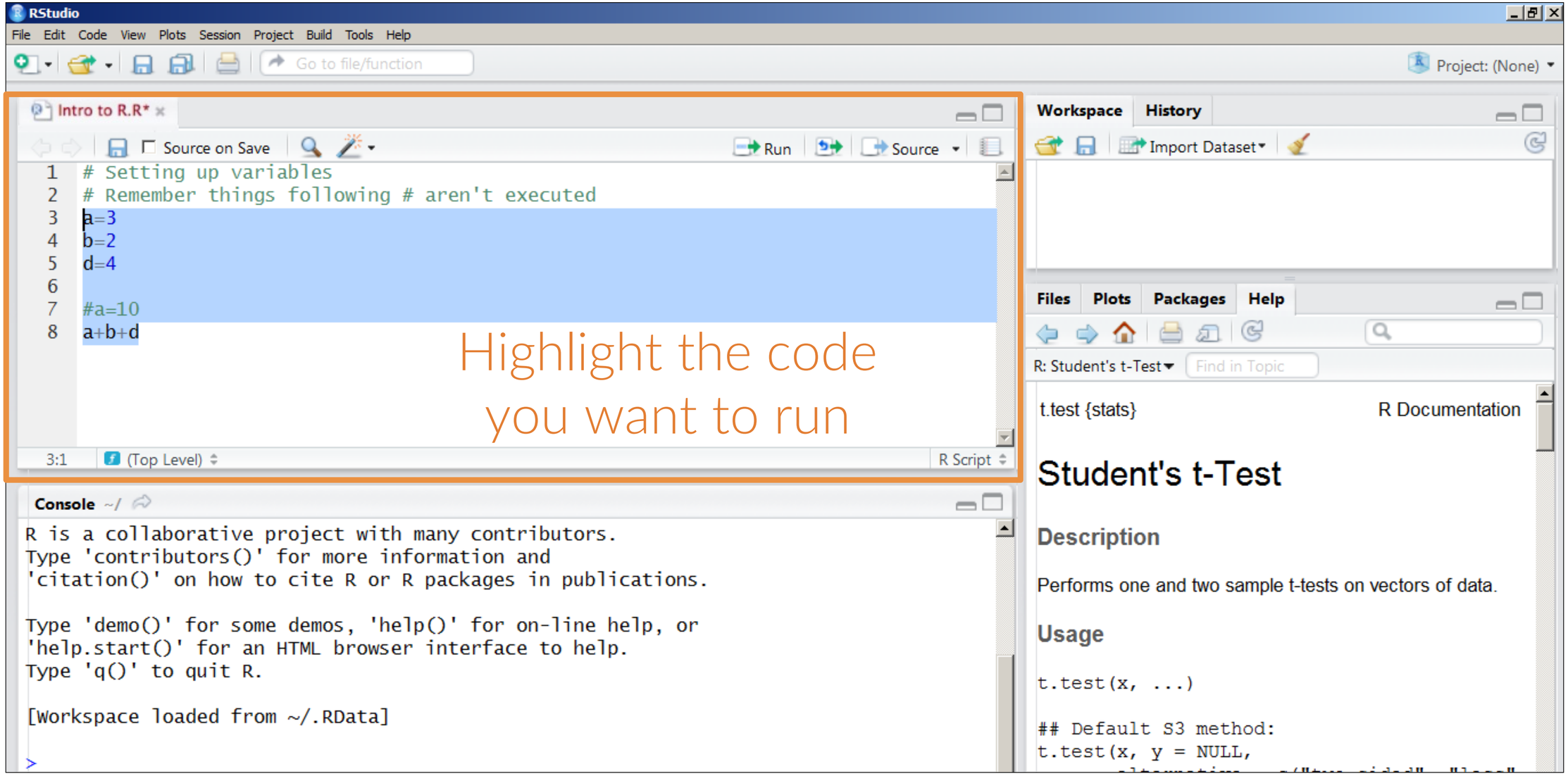
- Console
  - This is where the code executes
  - It is the actual R program (backend)
  - Anything that you type in here will be executed



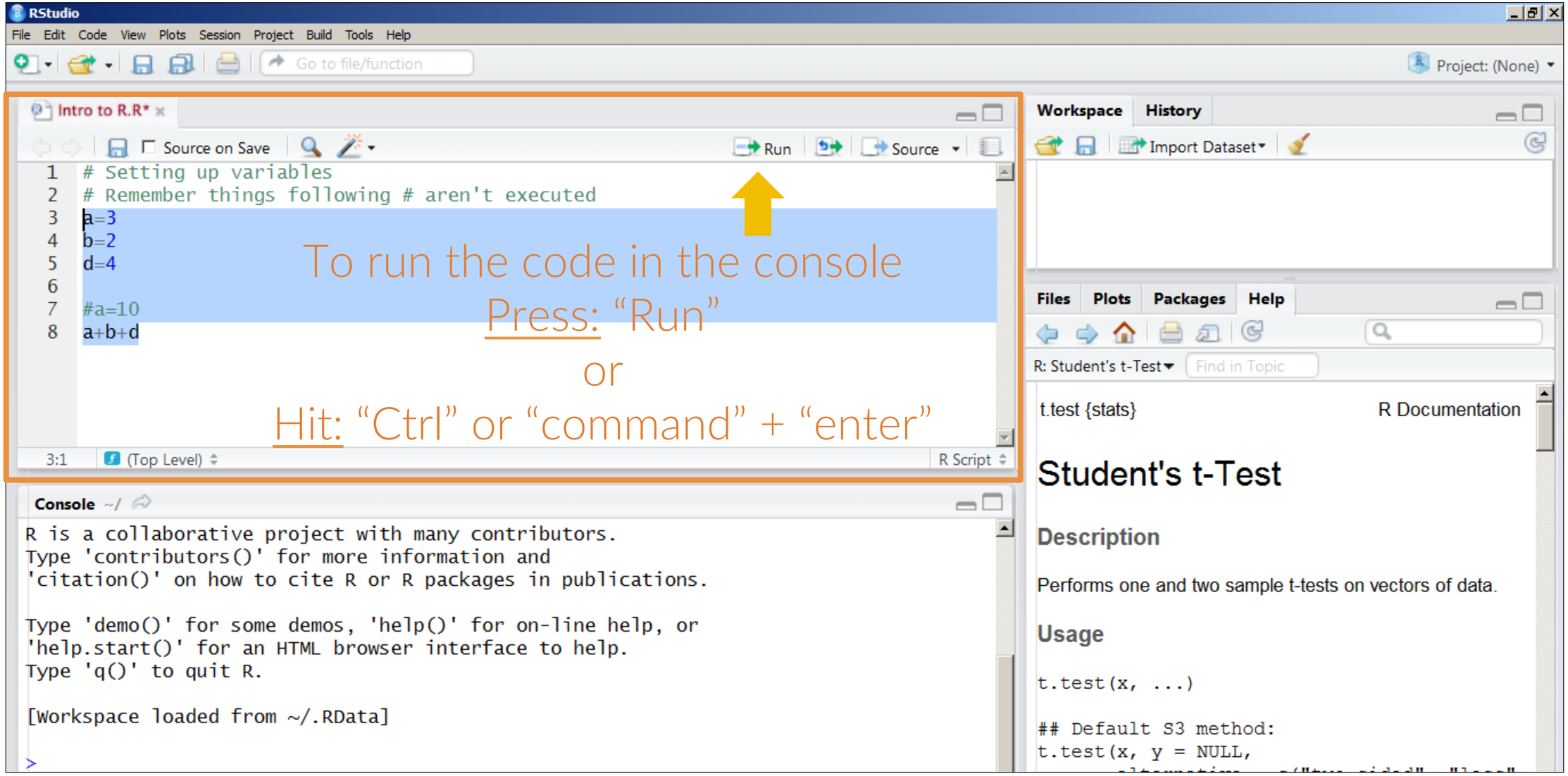
Running code from script



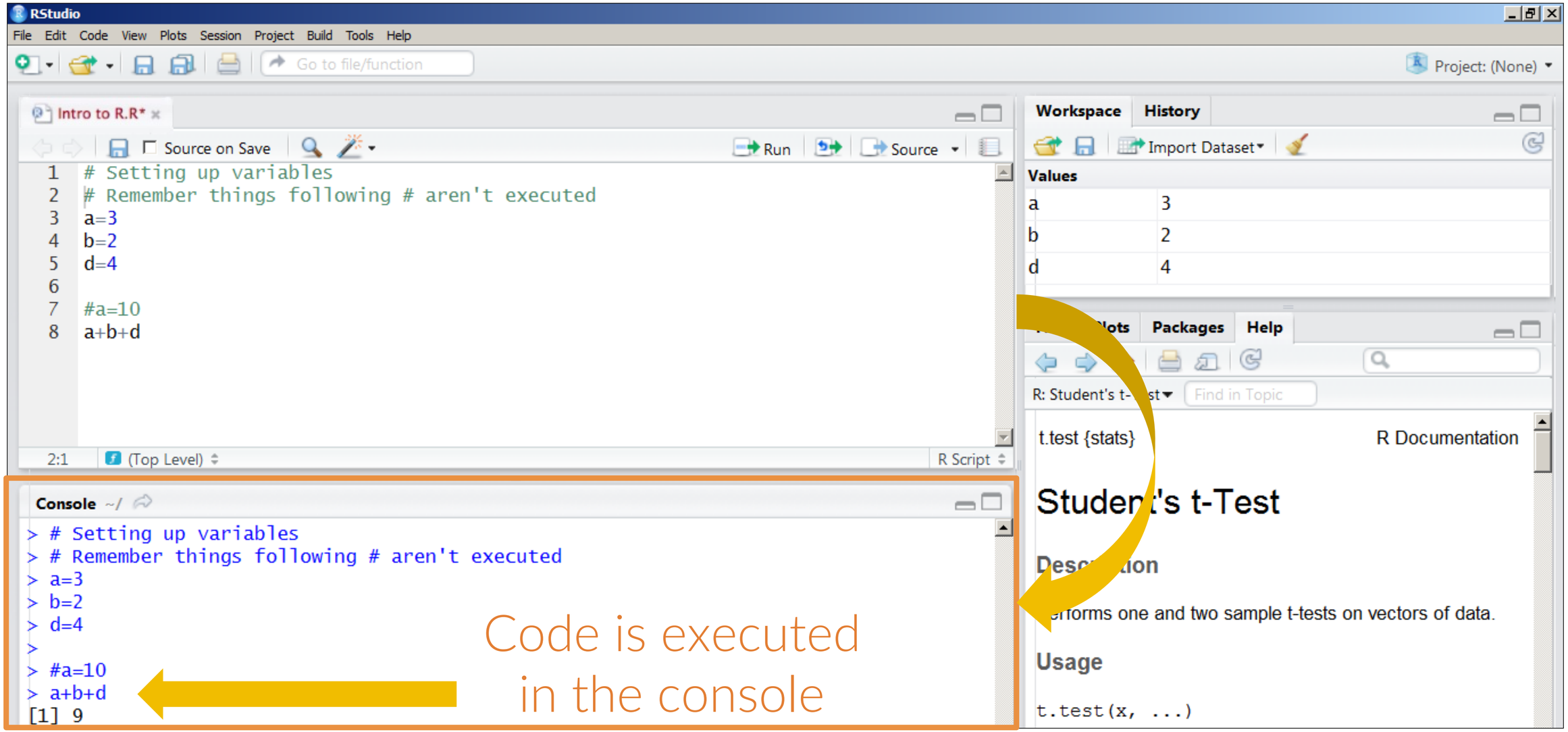
Running code from a script



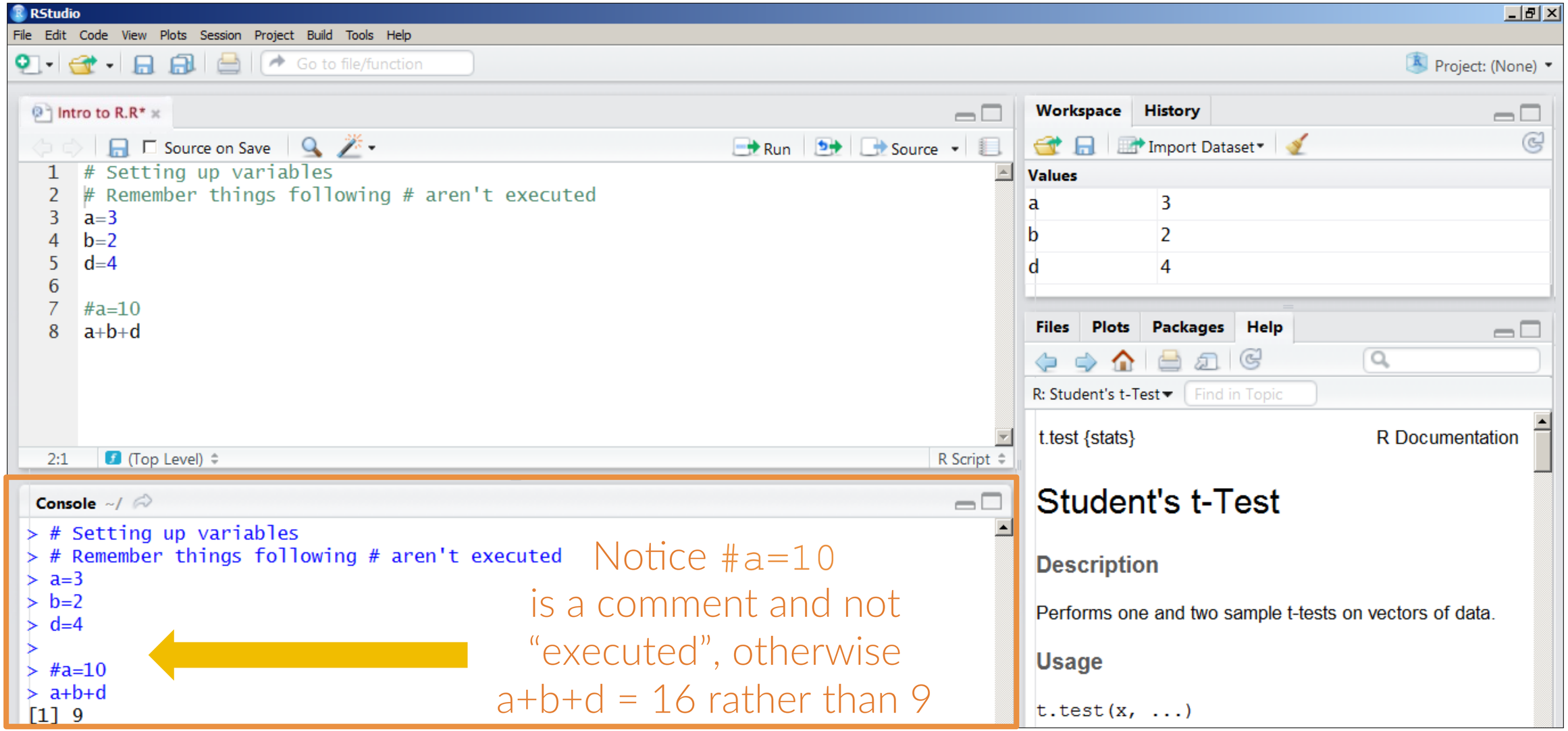
Running code from a script



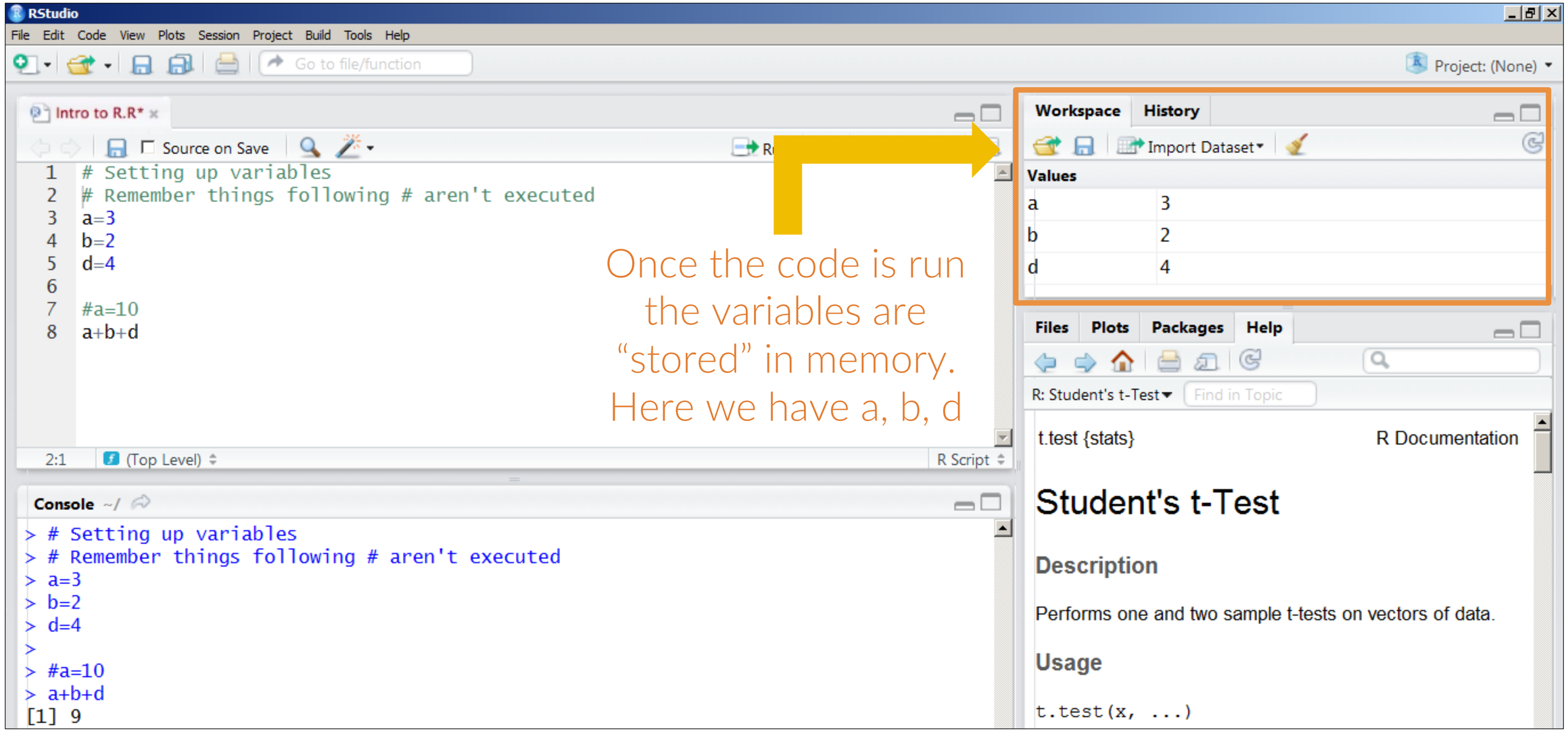
Running code from a script



Running code from a script



Variables



Once the code is run the variables are "stored" in memory. Here we have a, b, d

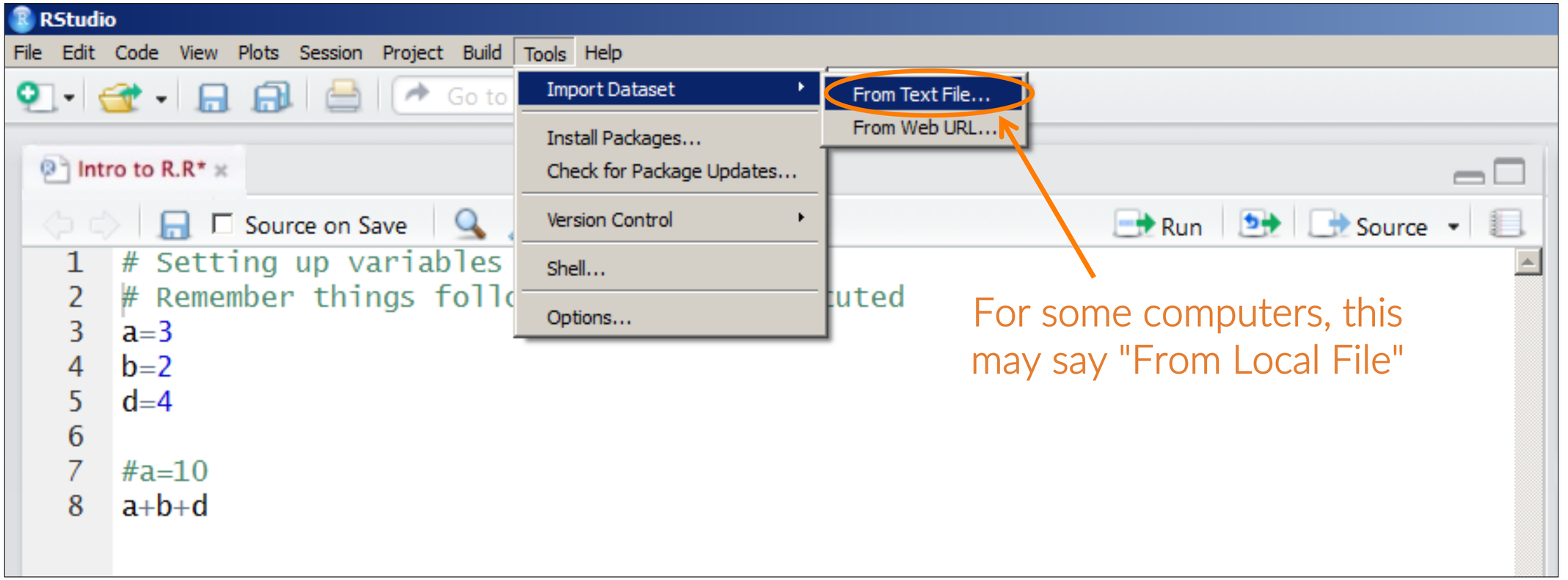
Reading in data

- Data comes in many forms
  - R can read in most formats

- The most common to work with now days are
  - `csv` files – comma separated value files
  - `tsv` files – tab separated value files

- R can also read in others (excel and json) but that is not covered here

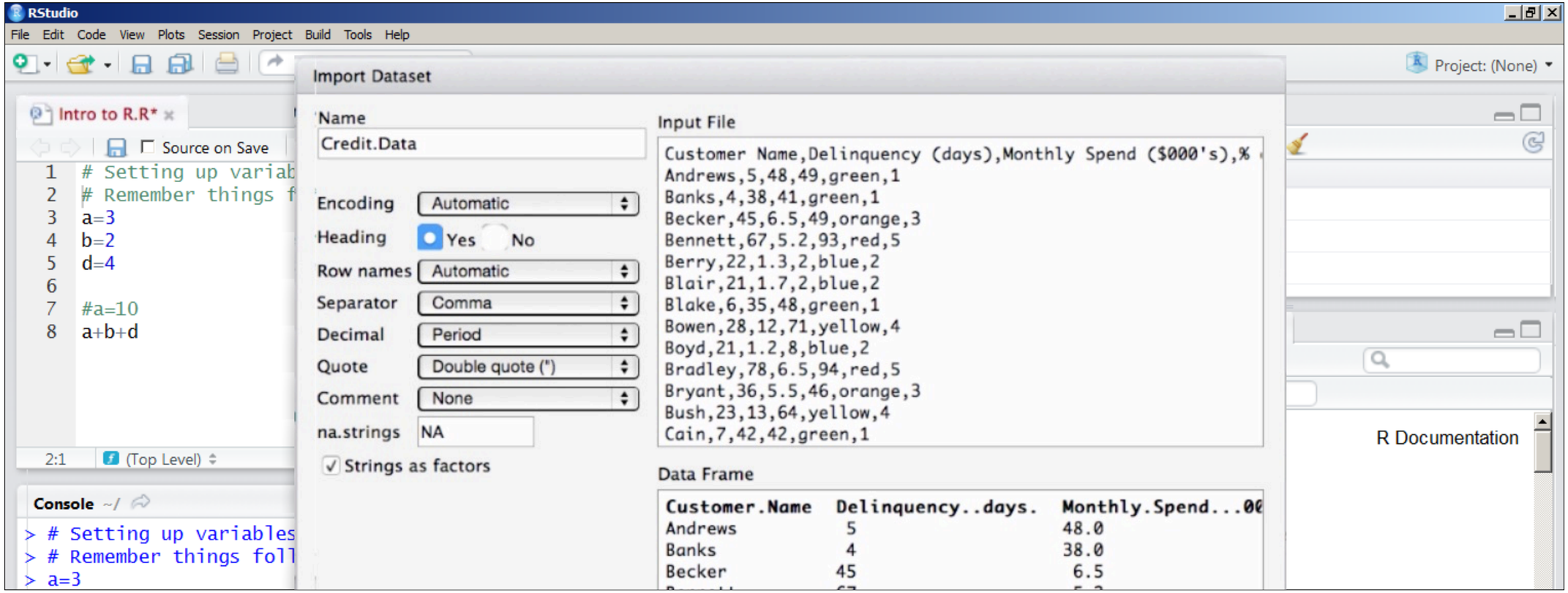
Reading in data: manually



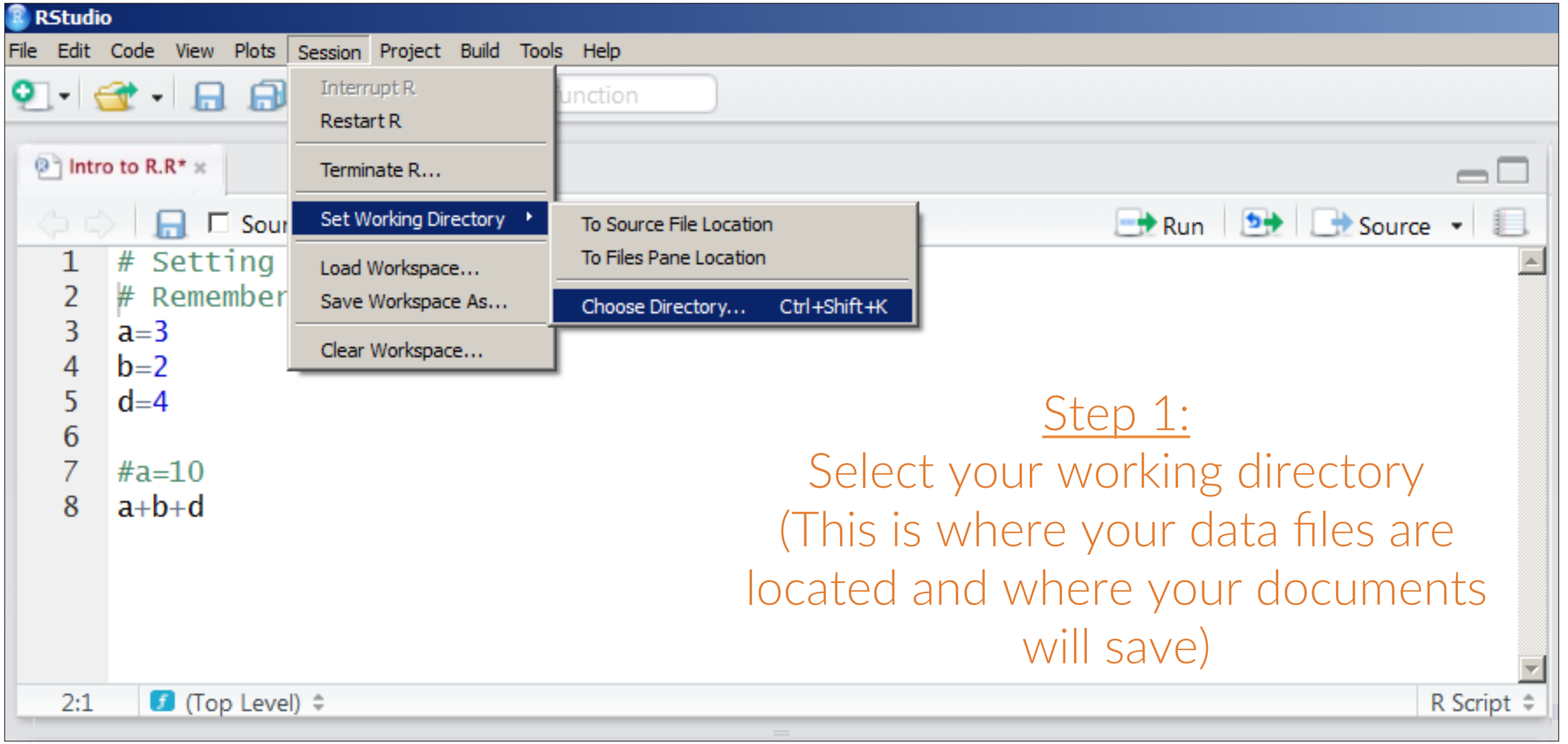
Tools > Import Dataset > From Text File...

Reading in data: manually

- RStudio will automatically select options for you such as heading, sometimes it makes mistakes so be sure to double check

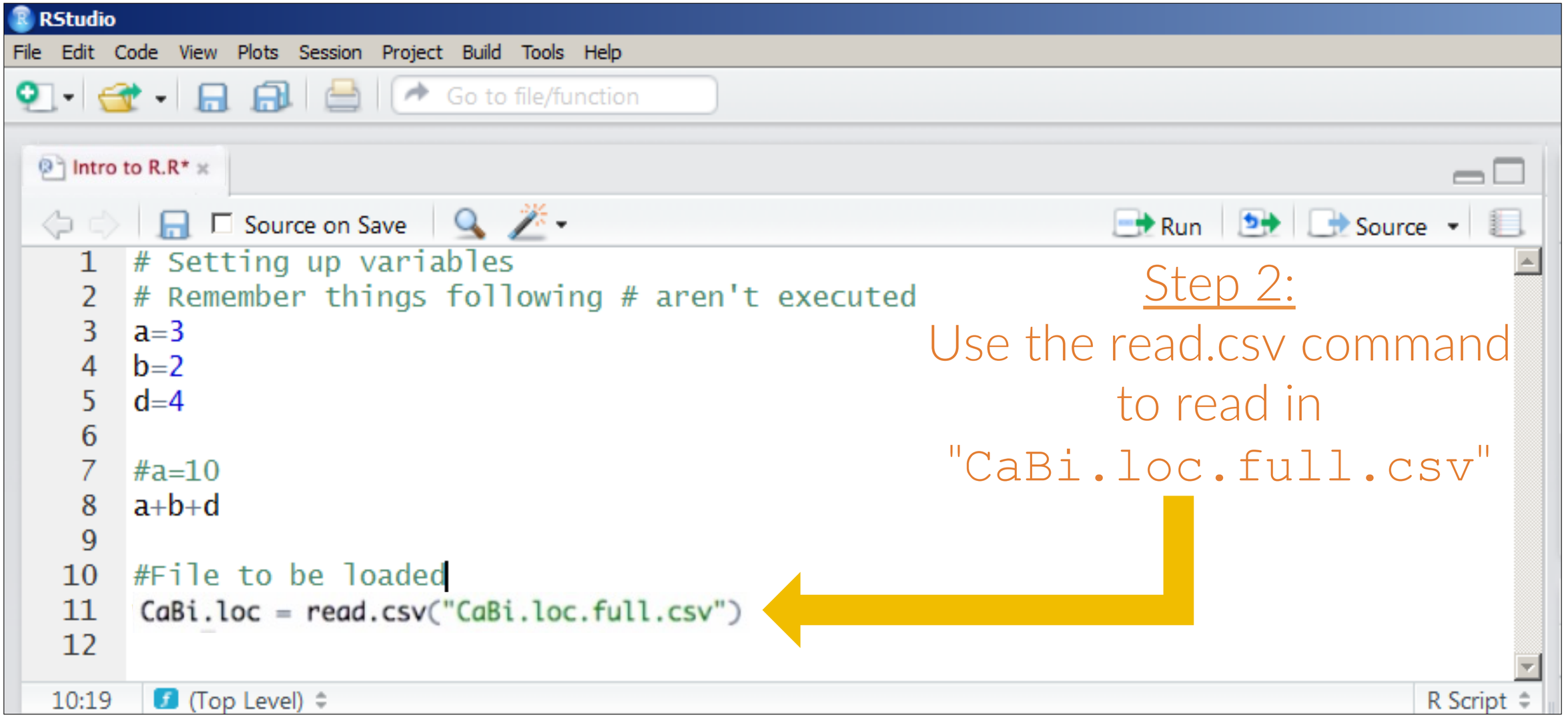


Reading in data: through the script

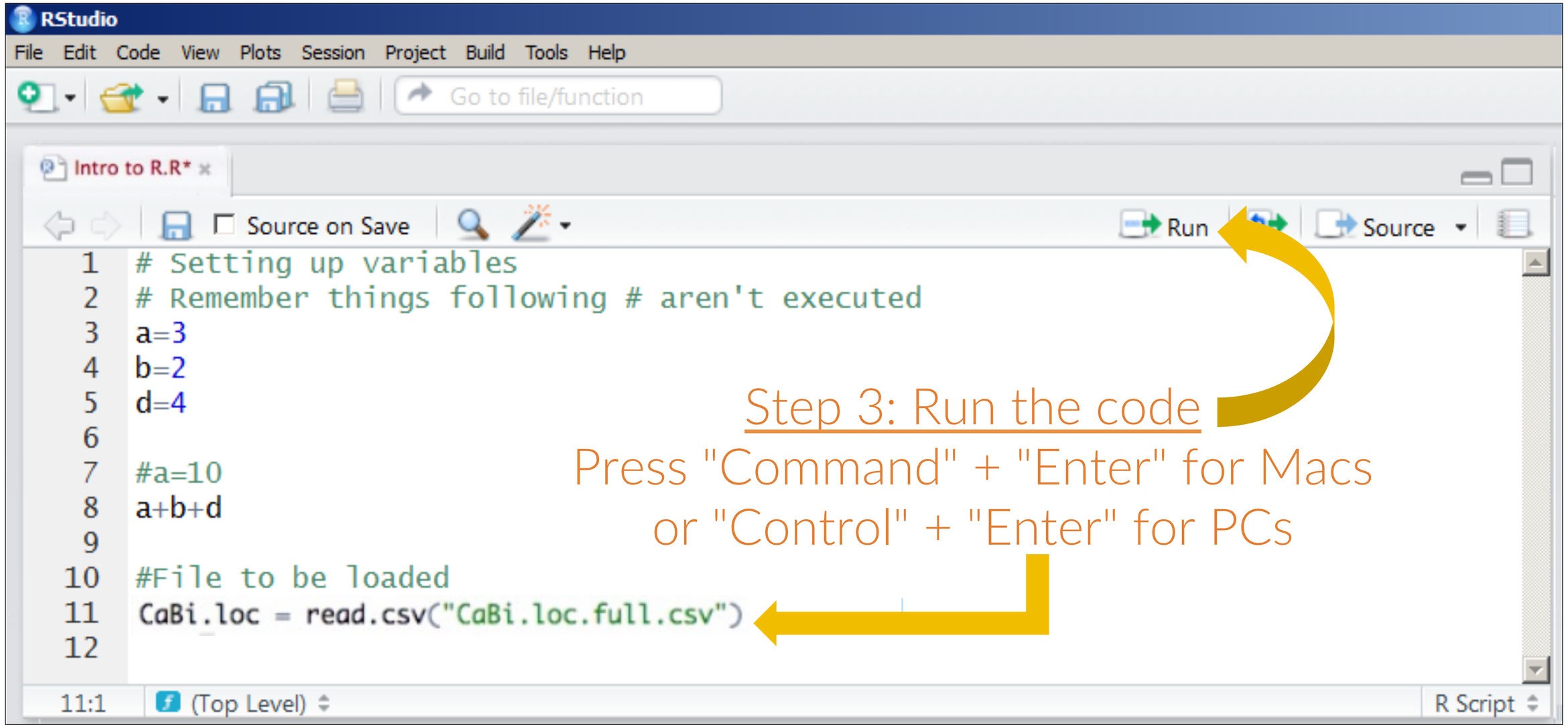


Step 1:
Select your working directory
(This is where your data files are located and where your documents will save)

Reading in data: through the script



Reading in data: through the script

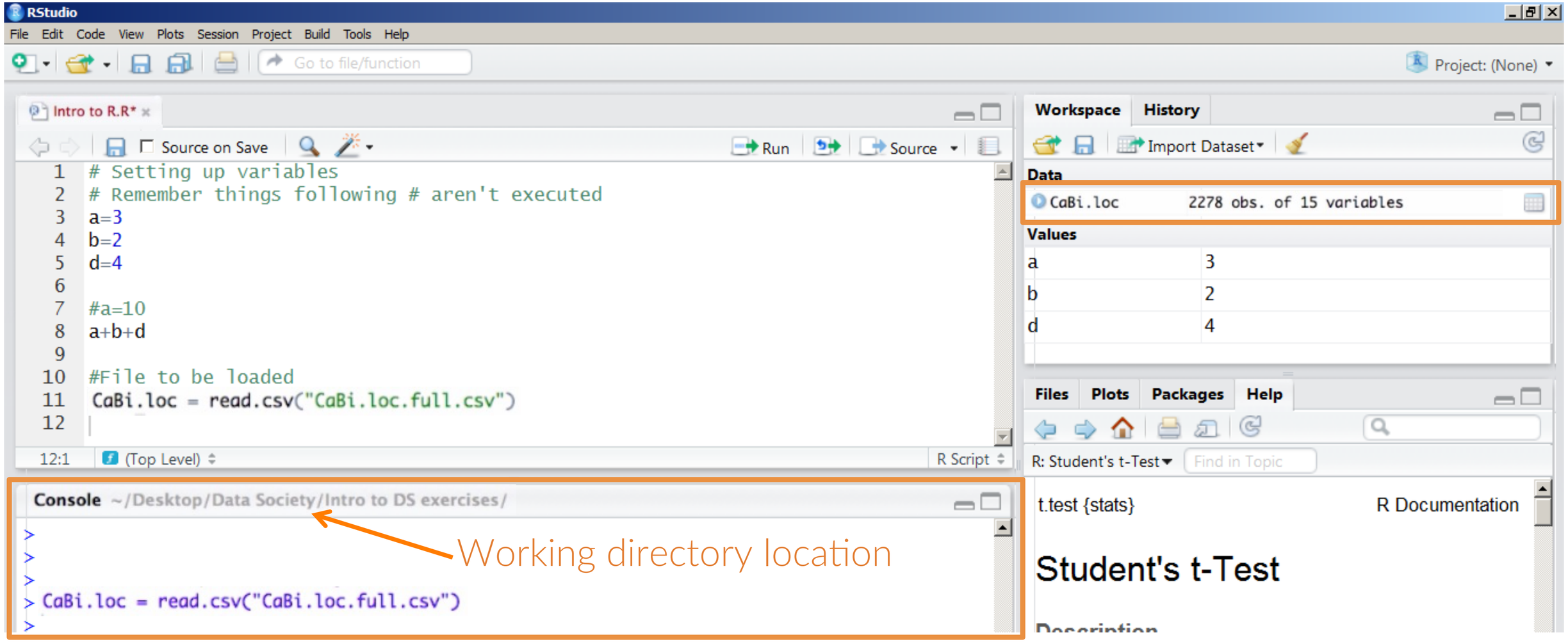


Step 3: Run the code

Press "Command" + "Enter" for Macs

or "Control" + "Enter" for PCs

Reading in data: through the script



Setting up R: overview

1. What is R?
2. Download R from the CRAN website (http://cran.us.r-project.org/)
   - R for Windows
   - R for Mac
3. Install R Studio (http://www.rstudio.com/products/rstudio/download/)
   - RStudio a brief tour
4. Running a script
   - Variables
5. Reading in a data
   - Manually
   - Through the script