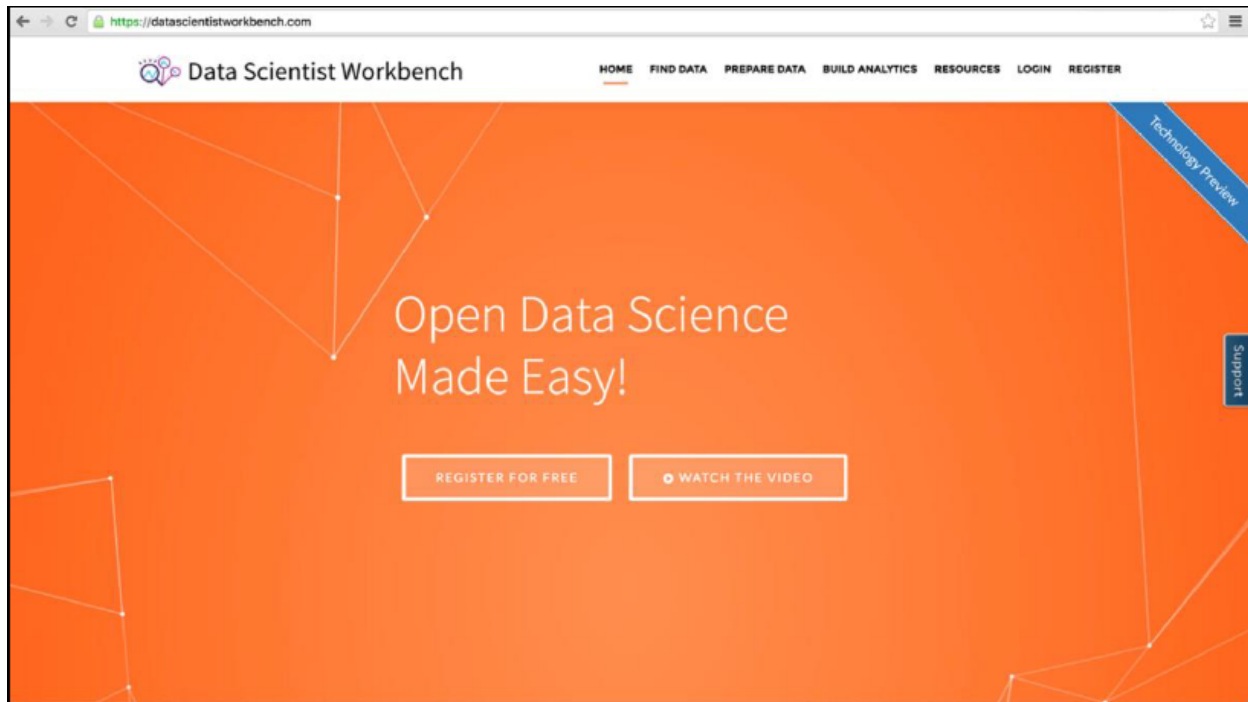




Lab: Introducing RStudio IDE

This Lab must be done on the Data Scientist Workbench. If you have not signed up yet, go to datascientistworkbench.com and click SIGN UP FOR FREE. Then login (2nd link from the right):





Contents

LAB 1	INTRODUCING RSTUDIO IDE	3
	1.1 GETTING STARTED WITH RSTUDIO IDE.....	4
	1.2 SUMMARY	4
	1.3 GETTING STARTED WITH RSTUDIO IDE - SOLUTIONS	5



Lab 1 Introducing RStudio IDE

After completing this hands-on lab, you will know:

- How to open and navigate within RStudio IDE
- How to create and save R scripts
- How to run R commands and check your environment variables
- How to upload files to RStudio
- How to install new packages
- How to load libraries
- How to restart RStudio

Allow 30 minutes to complete this section of the lab.

NOTE: This is a guided Lab. Solutions are found at the end of this document



1.1 Getting Started with RStudio IDE

- __1. Open RStudio
- __2. Create a new R script
- __3. Assign x to 1
- __4. Run the command to see the output in the console
- __5. Save the R Script
- __6. Check your environment for “x”
- __7. Upload a dataset to RStudio
- __8. Install a new package, such as “tm”
- __9. Load the library using library(tm)
- __10. Restart RStudio

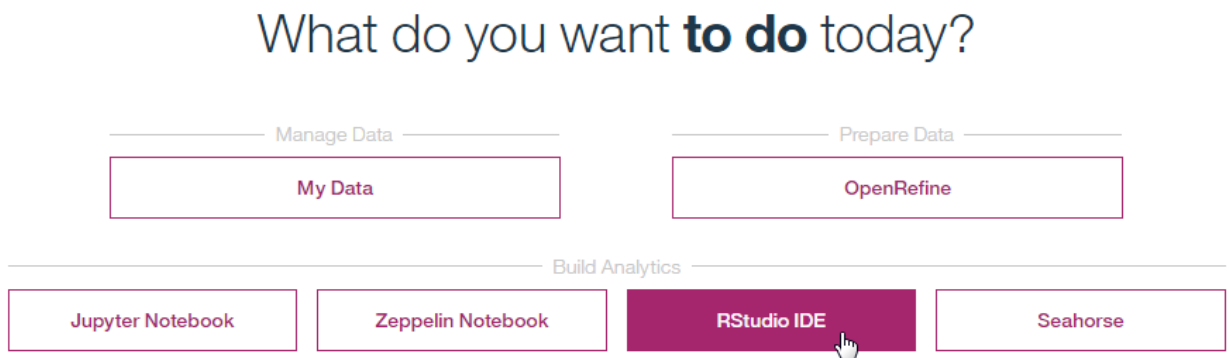
1.2 Summary

Congratulations! ...

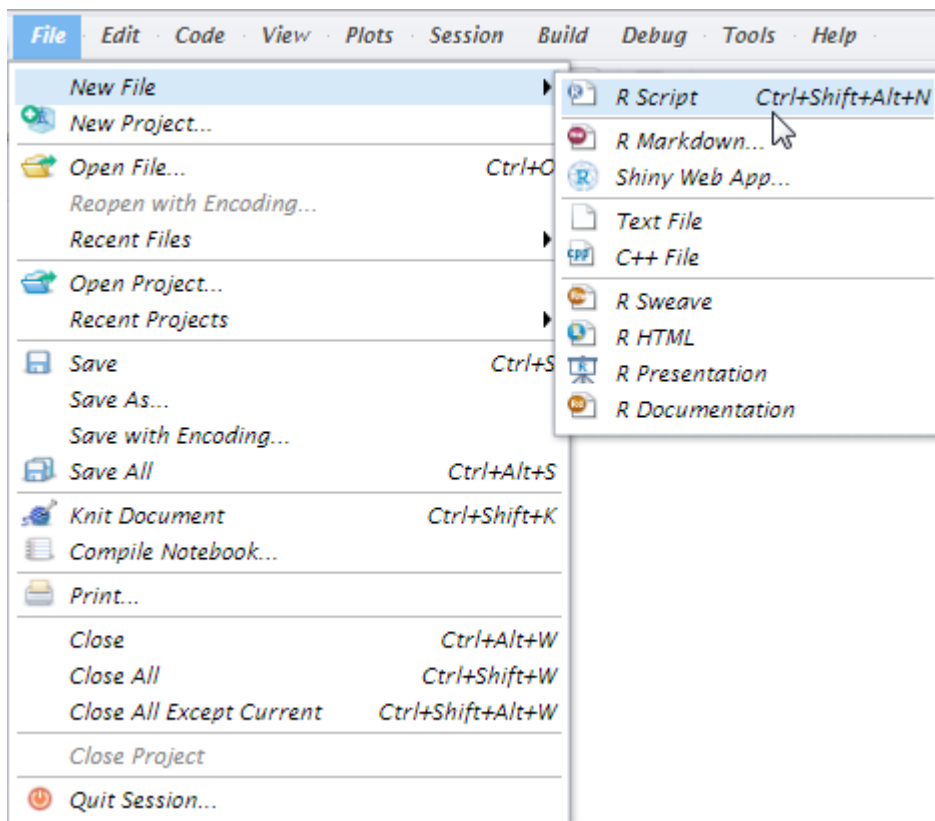


1.3 Getting Started with RStudio IDE - Solutions

- ___1. **Open RStudio:** To open RStudio IDE, login to datascientiseworkbench.com and click **RStudio IDE** on the DSWB workspace:

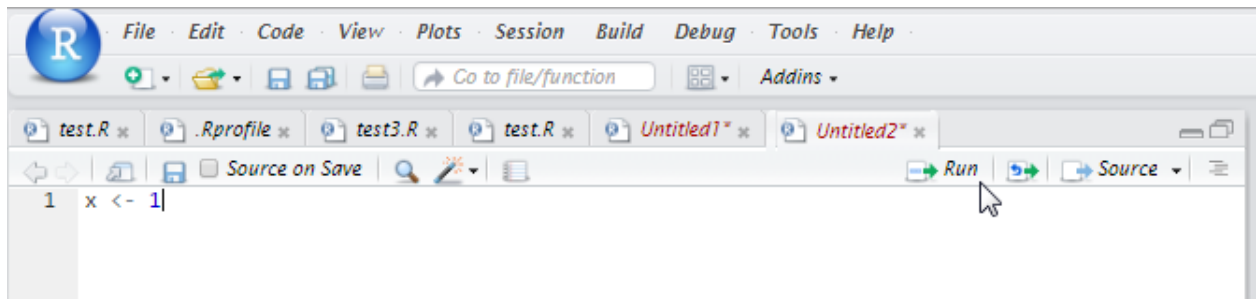


- ___2. **Create a new R script:** From the toolbar, select **File > New File > R Script**

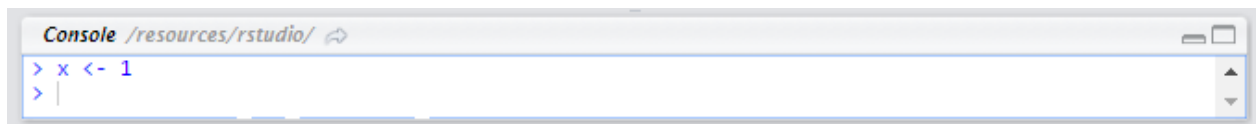




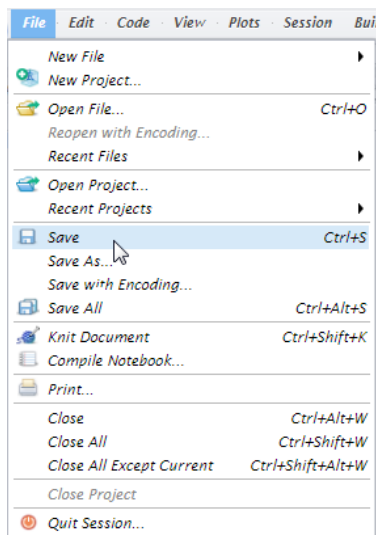
- __3. **Assign x to 1:** In the script that you created in Step 2, enter `x <- 1` or `x = 1`.



- __4. **Run the command to see the output in the console:** To run the command, click the Run icon on the toolbar. (If you run this command from the Console, press enter to run it.)



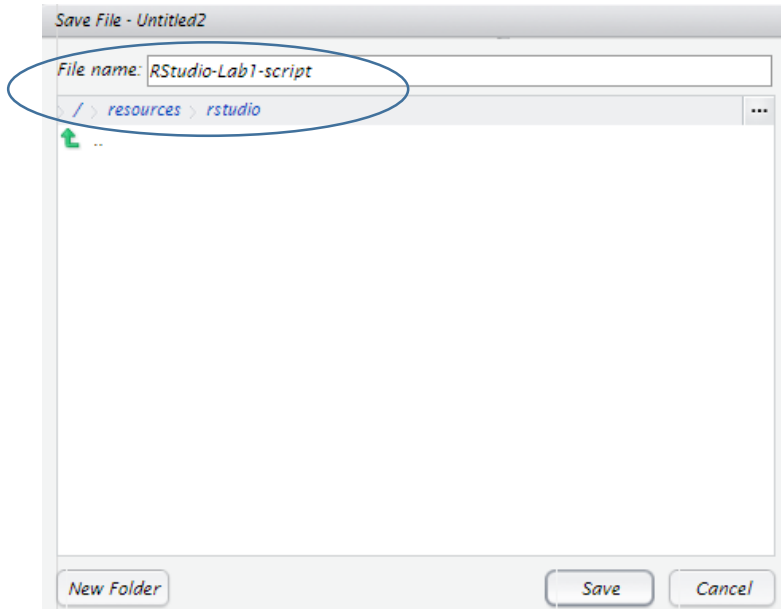
- __5. **Save the R Script:** Select **File > Save**



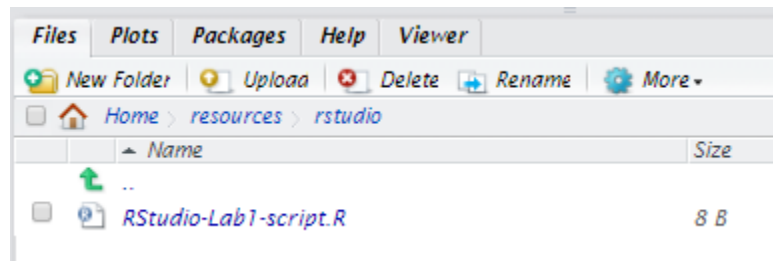


On the Save File window, type the file name and click **Save**

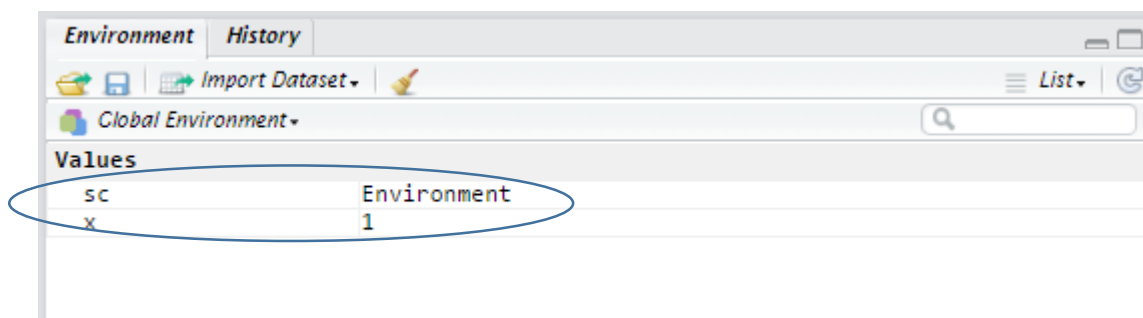
Note that the file is saved in the default resources/rstudio directory:



The saved file displays in the File section (bottom right quadrant):

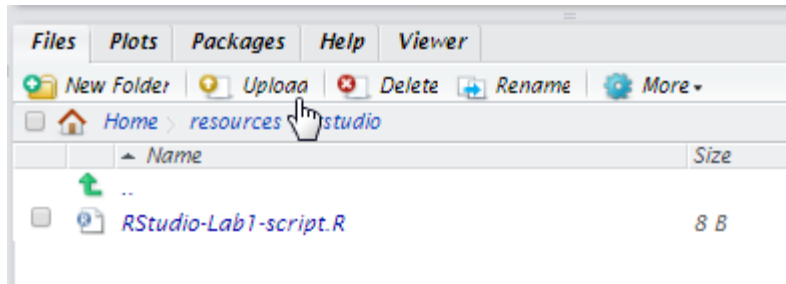


- ___6. **Check your environment for “x”:** In the top-right quadrant, look at the values under the Environment tab. Note the sc variable for the pre-installed Spark environment.

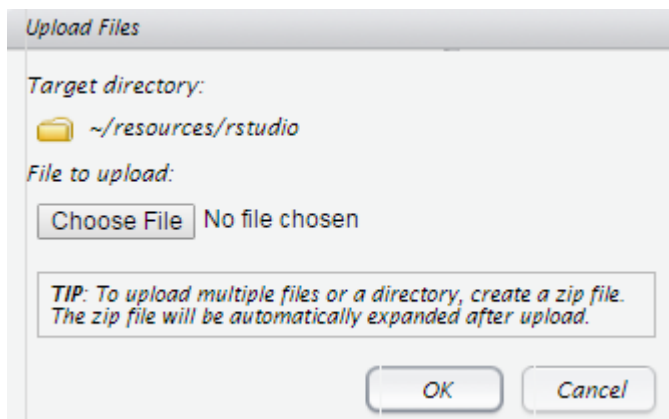




- ___7. **Upload a dataset to RStudio:** From the Files tab (bottom right quadrant); navigate to the directory where you want to upload the file and click Upload:

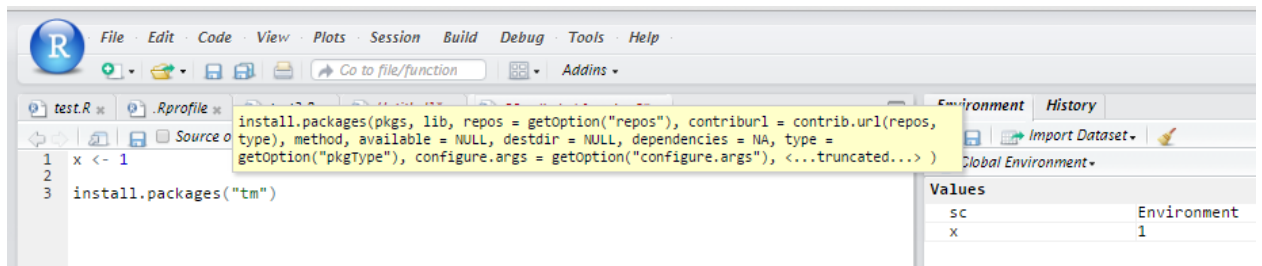


Choose the file to upload (it can be either from your local computer or from within any RStudio directories) and click OK;

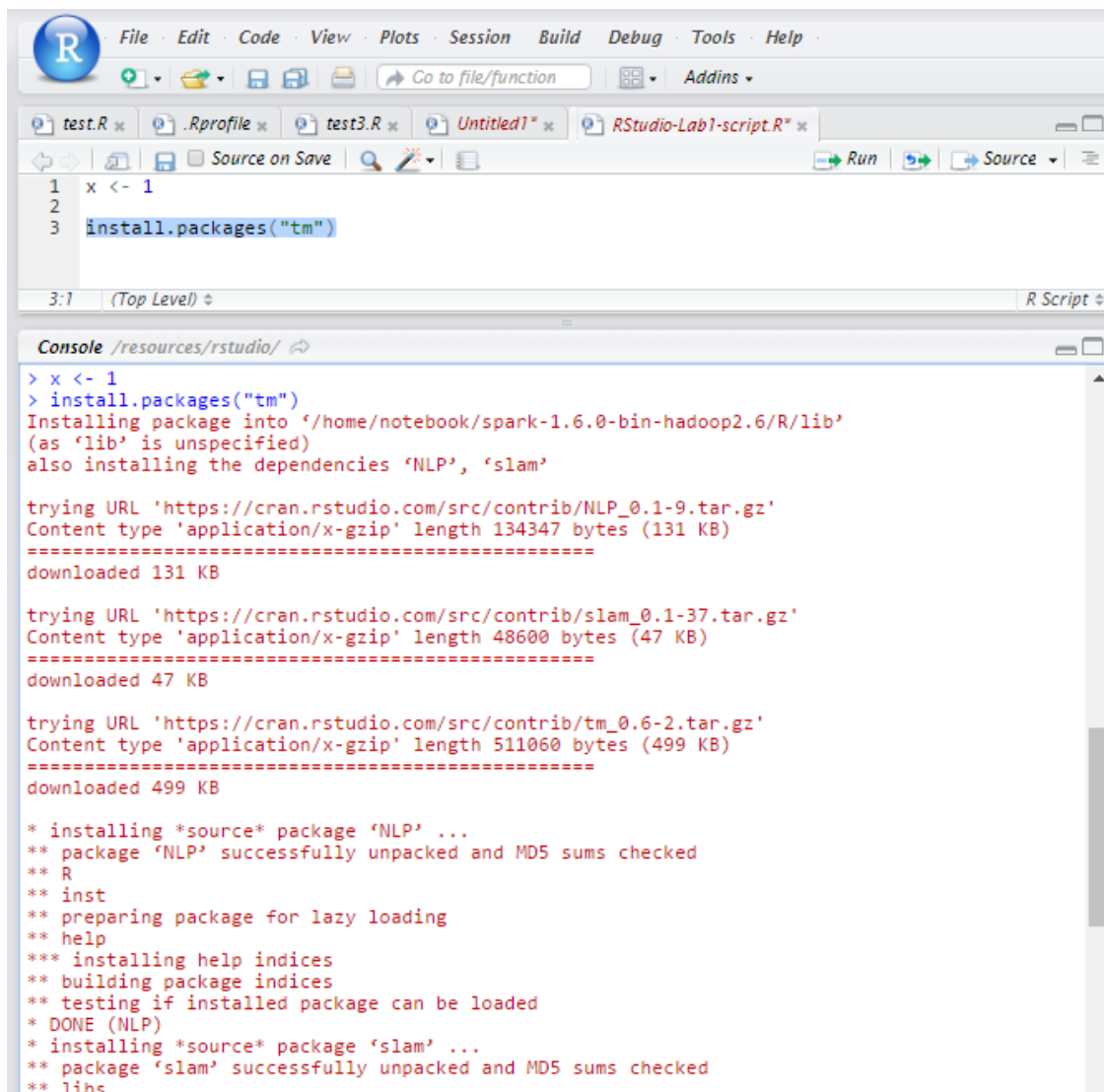




8. **Install a new package, such as “tm”:** In the R Editor, type `install.packages("tm")` and run the command:



In the Console, you will see the status of the package that you installed:





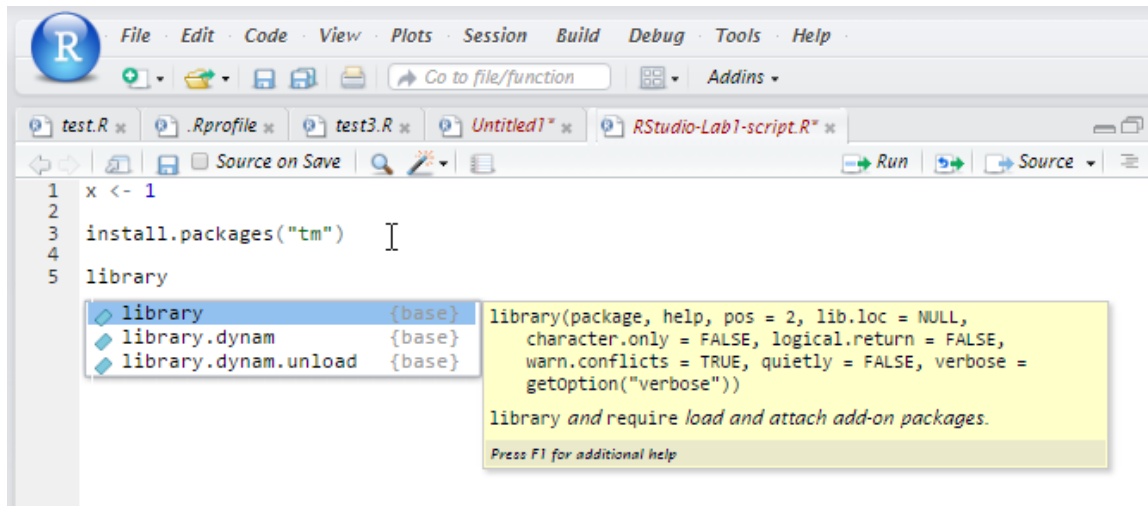
continuation of the console log for installed package “tm”- note the directory where the “tm” package was installed (last line of the console log):

```
Console /resources/rstudio/
** libs
gcc -std=gnu99 -I/usr/share/R/include -DNDEBUG -fpic -g -O2 -fstack-protector-strong -Wformat -Werror=format-security -Wdate-time -D_FORTIFY_SOURCE=2 -g -c apply.c -o apply.o
gcc -std=gnu99 -I/usr/share/R/include -DNDEBUG -fpic -g -O2 -fstack-protector-strong -Wformat -Werror=format-security -Wdate-time -D_FORTIFY_SOURCE=2 -g -c grouped.c -o grouped.o
gcc -std=gnu99 -I/usr/share/R/include -DNDEBUG -fpic -g -O2 -fstack-protector-strong -Wformat -Werror=format-security -Wdate-time -D_FORTIFY_SOURCE=2 -g -c sparse.c -o sparse.o
gcc -std=gnu99 -I/usr/share/R/include -DNDEBUG -fpic -g -O2 -fstack-protector-strong -Wformat -Werror=format-security -Wdate-time -D_FORTIFY_SOURCE=2 -g -c util.c -o util.o
gcc -std=gnu99 -shared -L/usr/lib/R/lib -Wl,-z,relro -o slam.so apply.o grouped.o sparse.o util.o -lblas -lgfortran -lm -lquadmath -L/usr/lib/R/lib -lR
installing to /home/notebook/spark-1.6.0-bin-hadoop2.6/R/lib/slam/libs
** R
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded
* DONE (slam)
* installing *source* package 'tm' ...
** package 'tm' successfully unpacked and MD5 sums checked
** libs
gcc -std=gnu99 -I/usr/share/R/include -DNDEBUG -fpic -g -O2 -fstack-protector-strong -Wformat -Werror=format-security -Wdate-time -D_FORTIFY_SOURCE=2 -g -c copy.c -o copy.o
gcc -std=gnu99 -shared -L/usr/lib/R/lib -Wl,-z,relro -o tm.so copy.o -L/usr/lib/R/lib -lR
installing to /home/notebook/spark-1.6.0-bin-hadoop2.6/R/lib/tm/libs
** R
** data
** inst
** preparing package for lazy loading
** help
*** installing help indices
** building package indices
** installing vignettes
** testing if installed package can be loaded
* DONE (tm)

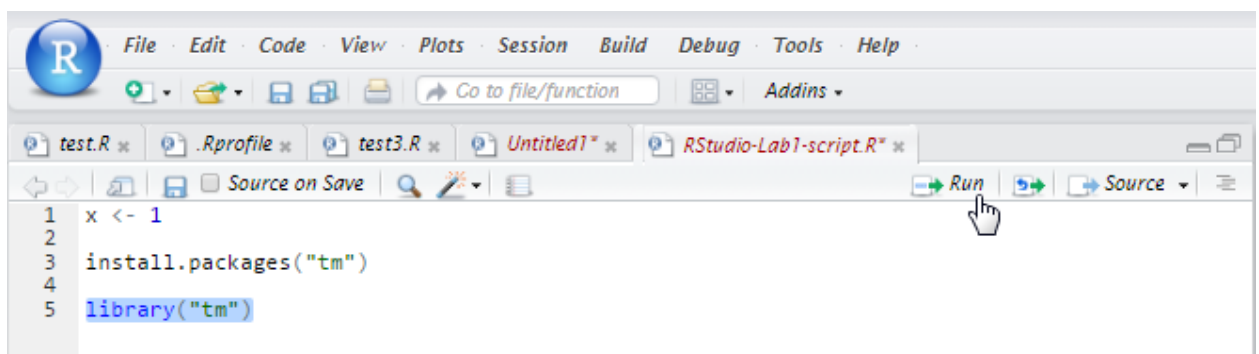
The downloaded source packages are in
'/tmp/Rtmphfpufu/downloaded_packages'
> |
```



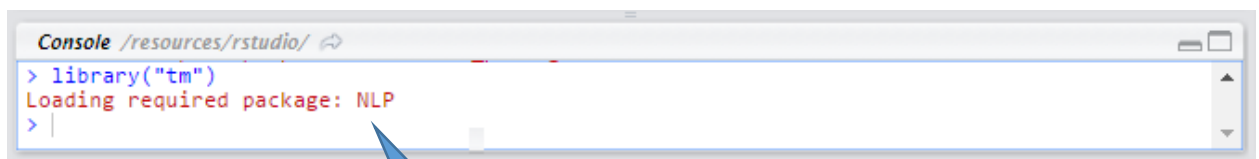
__9. **Load the library using library(tm):** To *load* a library, use the library command:



Type `library("tm")` and run the command:



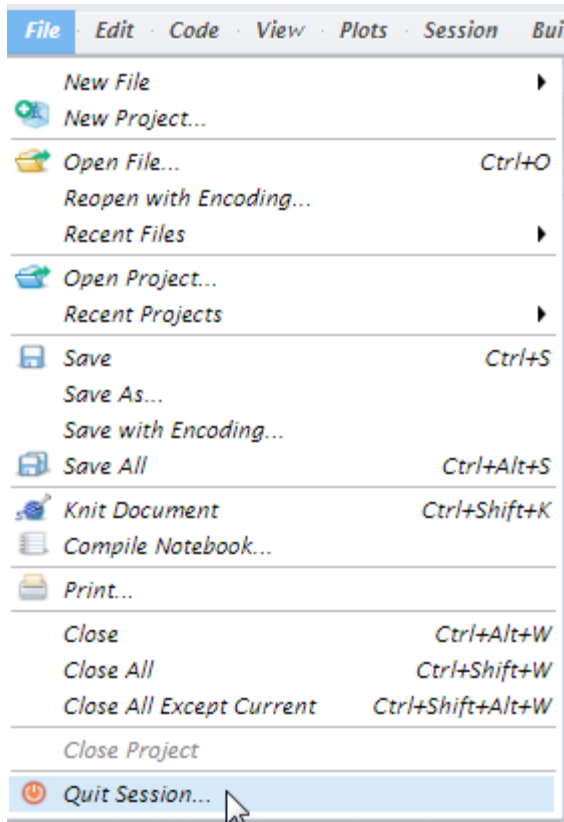
The console log shows that status of the library load command:



Not sure if the
tm library was
loaded; what is
NLP?



- ___10. **Start a new R session:** *To start a new RStudio session, select File > Quit Session:*



This restarts RStudio IDE. If you are getting errors that you cannot resolve, such as the following, it's recommended that you quit your current session and restart RStudio IDE:.

```
> SC
Error in callJMethod(x, "getClass") :
  Invalid jobj 0. If SparkR was restarted, Spark operations need to be re-executed.

+ data <- read.csv2("/resources/")
Error: unexpected string constant in:
"data <- read.csv2("/resources/")
data <- read.csv2("")
```