



# Introduction to RStudio Server

The CCDL

The following two analysis examples are both microarray differential expression analyses.



# Differential Analysis Example 1: “100% Up to Date Analysis”

**Methods Documentation:** ~200 words describe the general methods in a publication. The step-by-step recount of how these data came to be were handwritten, with one copy located in a lab closet.

**Data availability:** The original data are on a flashdrive, in a desk drawer, in a lab. The already processed version is on [GEO](#).

**Project organization:** Many files with various nebulous terms like “MasterSheet” and “Sorted” and “Edit” and “100% Up to Date”.

**Software versions:** No idea what software package(s), regardless of version.

| 1     | ProbeName                | Sequence            | GeneName                 | SystematicName           | Description | RName   | RSavedName  |
|-------|--------------------------|---------------------|--------------------------|--------------------------|-------------|---|---|
| 881   | CHIP_JH_SE3_19777        | GTCTGTATGACAGCCACG  | CHIP_JH_SE3_19777        | CHIP_JH_SE3_19777        | Unknown     | { 880 } - { 880 }_AK056136 -> CHIP_JH_SE3_19777 < Unknown . | { 880 } - { 880 }_AK056136 -> CHIP_JH_SE3_19777 < Unknown . |
| 26316 | CHIP_JH_SE3_19780        | ATAGACAGAAAGCAGCAGT | CHIP_JH_SE3_19780        | CHIP_JH_SE3_19780        | Unknown     | { 26315 }_AK056136 -> CHIP_JH_SE3_19780 < Unknown .         | { 26315 }_AK056136 -> CHIP_JH_SE3_19780 < Unknown .         |
| 27565 | CHIP_CON_ALB_196         | CTAGATGAAAATAAAGTCA | CHIP_CON_ALB_196         | CHIP_CON_ALB_196         | Unknown     | { 27564 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .         | { 27564 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .         |
| 28964 | CHIP_CON_ALB_196         | CTAGATGAAAATAAAGTCA | CHIP_CON_ALB_196         | CHIP_CON_ALB_196         | Unknown     | { 28963 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .         | { 28963 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .         |
| 19195 | CHIP_CON_ALB_193         | GTTCCATTTTACATTTCCC | CHIP_CON_ALB_193         | CHIP_CON_ALB_193         | Unknown     | { 19194 }_NM_014178 -> CHIP_CON_ALB_193 < Unknown .         | { 19194 }_NM_014178 -> CHIP_CON_ALB_193 < Unknown .         |
| 32499 | CHIP_CON_ALB_193         | GTTCCATTTTACATTTCCC | CHIP_CON_ALB_193         | CHIP_CON_ALB_193         | Unknown     | { 32498 }_NM_014178 -> CHIP_CON_ALB_193 < Unknown .         | { 32498 }_NM_014178 -> CHIP_CON_ALB_193 < Unknown .         |
| 6938  | CHIP_CON_ALB_196         | CTAGATGAAAATAAAGTCA | CHIP_CON_ALB_196         | CHIP_CON_ALB_196         | Unknown     | { 6937 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .          | { 6937 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .          |
| 773   | CHIP_CON_ALB_196         | CTAGATGAAAATAAAGTCA | CHIP_CON_ALB_196         | CHIP_CON_ALB_196         | Unknown     | { 772 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .           | { 772 }_NM_014178 -> CHIP_CON_ALB_196 < Unknown .           |
| 33424 | CHIP_JH_SE3_19779        | TTATGAGAATGGGGCTCAC | CHIP_JH_SE3_19779        | CHIP_JH_SE3_19779        | Unknown     | { 33423 }_AK056136 -> CHIP_JH_SE3_19779 < Unknown .         | { 33423 }_AK056136 -> CHIP_JH_SE3_19779 < Unknown .         |
| 9632  | CHIP_CON_ALB_193         | GTTCCATTTTACATTTCCC | CHIP_CON_ALB_193         | CHIP_CON_ALB_193         | Unknown     | { 9631 }_NM_014178 -> CHIP_CON_ALB_193 < Unknown .          | { 9631 }_NM_014178 -> CHIP_CON_ALB_193 < Unknown .          |
| 9658  | CHIP_JH_SE3_13875        | TGTAGAAATGTCTCCAG   | CHIP_JH_SE3_13875        | CHIP_JH_SE3_13875        | Unknown     | { 9657 }_AK056431 -> CHIP_JH_SE3_13875 < Unknown .          | { 9657 }_AK056431 -> CHIP_JH_SE3_13875 < Unknown .          |
| 10134 | CHIP_JH_SE3_19778        | AAGAGAAATGATCTGTGT  | CHIP_JH_SE3_19778        | CHIP_JH_SE3_19778        | Unknown     | { 10133 }_AK056136 -> CHIP_JH_SE3_19778 < Unknown .         | { 10133 }_AK056136 -> CHIP_JH_SE3_19778 < Unknown .         |
| 791   | CHIP_JH_SE3_23399        | AGAGCTAAACACAGCTCAA | CHIP_JH_SE3_23399        | CHIP_JH_SE3_23399        | Unknown     | { 790 }_AK074776 -> CHIP_JH_SE3_23399 < Unknown .           | { 790 }_AK074776 -> CHIP_JH_SE3_23399 < Unknown .           |
| 2421  | CHIP_JH_SE3_16148        | TCTTACCCTCTTCAAHA   | CHIP_JH_SE3_16148        | CHIP_JH_SE3_16148        | Unknown     | { 2420 }_AL832845 -> CHIP_JH_SE3_16148 < Unknown .          | { 2420 }_AL832845 -> CHIP_JH_SE3_16148 < Unknown .          |
| 35094 | CHIP_JH_SE3_34118        | CAGCTCTCTGATTAECTCC | CHIP_JH_SE3_34118        | CHIP_JH_SE3_34118        | Unknown     | { 35093 }_CR598488 -> CHIP_JH_SE3_34118 < Unknown .         | { 35093 }_CR598488 -> CHIP_JH_SE3_34118 < Unknown .         |
| 9840  | CONTROL_EPILEPSY_SE3_382 | TGTTTTGGGGATGAGAAGG | CONTROL_EPILEPSY_SE3_382 | CONTROL_EPILEPSY_SE3_382 | Unknown     | { 9839 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown .  | { 9839 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown .  |
| 12692 | CONTROL_EPILEPSY_SE3_382 | TGTTTTGGGGATGAGAAGG | CONTROL_EPILEPSY_SE3_382 | CONTROL_EPILEPSY_SE3_382 | Unknown     | { 12691 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown . | { 12691 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown . |
| 34255 | CONTROL_EPILEPSY_SE3_382 | TGTTTTGGGGATGAGAAGG | CONTROL_EPILEPSY_SE3_382 | CONTROL_EPILEPSY_SE3_382 | Unknown     | { 34254 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown . | { 34254 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown . |
| 29946 | CHIP_CON_ALB_194         | GAATATGAAATTTGGACA  | CHIP_CON_ALB_194         | CHIP_CON_ALB_194         | Unknown     | { 29945 }_NM_014178 -> CHIP_CON_ALB_194 < Unknown .         | { 29945 }_NM_014178 -> CHIP_CON_ALB_194 < Unknown .         |
| 26892 | CHIP_JH_SE3_23396        | CTAACACACCTCAAAAG   | CHIP_JH_SE3_23396        | CHIP_JH_SE3_23396        | Unknown     | { 26891 }_AK074776 -> CHIP_JH_SE3_23396 < Unknown .         | { 26891 }_AK074776 -> CHIP_JH_SE3_23396 < Unknown .         |
| 40335 | CHIP_JH_SE3_23398        | GAGCTAAACACACCTCAA  | CHIP_JH_SE3_23398        | CHIP_JH_SE3_23398        | Unknown     | { 40334 }_AK074776 -> CHIP_JH_SE3_23398 < Unknown .         | { 40334 }_AK074776 -> CHIP_JH_SE3_23398 < Unknown .         |
| 31145 | CHIP_JH_SE3_23400        | GAGAGCTAAACACACCTCA | CHIP_JH_SE3_23400        | CHIP_JH_SE3_23400        | Unknown     | { 31144 }_AK074776 -> CHIP_JH_SE3_23400 < Unknown .         | { 31144 }_AK074776 -> CHIP_JH_SE3_23400 < Unknown .         |
| 3994  | CHIP_JH_SE3_969          | TTGTGAATGTGTGTGTTGA | CHIP_JH_SE3_969          | CHIP_JH_SE3_969          | Unknown     | { 3993 }_AK130883 -> CHIP_JH_SE3_969 < Unknown .            | { 3993 }_AK130883 -> CHIP_JH_SE3_969 < Unknown .            |
| 14335 | CONTROL_EPILEPSY_SE3_382 | TGTTTTGGGGATGAGAAGG | CONTROL_EPILEPSY_SE3_382 | CONTROL_EPILEPSY_SE3_382 | Unknown     | { 14334 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown . | { 14334 }_NM_004233 -> CONTROL_EPILEPSY_SE3_382 < Unknown . |
| 13000 | CONTROL_EPILEPSY_SE3_275 | TAGAAGGCTCAACACCC   | CONTROL_EPILEPSY_SE3_275 | CONTROL_EPILEPSY_SE3_275 | Unknown     | { 12999 }_NM_032895 -> CONTROL_EPILEPSY_SE3_275 < Unknown . | { 12999 }_NM_032895 -> CONTROL_EPILEPSY_SE3_275 < Unknown . |
| 32032 | CONTROL_EPILEPSY_SE3_381 | AGGTACCAAGAAGCTTTT  | CONTROL_EPILEPSY_SE3_381 | CONTROL_EPILEPSY_SE3_381 | Unknown     | { 32031 }_NM_004233 -> CONTROL_EPILEPSY_SE3_381 < Unknown . | { 32031 }_NM_004233 -> CONTROL_EPILEPSY_SE3_381 < Unknown . |
| 25796 | CHIP_JH_SE3_972          | GGTATAGTGTCAAAAGTAT | CHIP_JH_SE3_972          | CHIP_JH_SE3_972          | Unknown     | { 25795 }_AK130883 -> CHIP_JH_SE3_972 < Unknown .           | { 25795 }_AK130883 -> CHIP_JH_SE3_972 < Unknown .           |

- 100% Up to Date Log 3-3-15 CS.xlsx
- ▼ CAGE Data
  - CorrelationREV3LTRAF.svp
  - FANTOMCAGEData3-4-15CS.xlsx
  - PRKCQ-ASTCAGEData.txt
  - PRKCQCAGEData.txt
  - REV3LTRAF3IP2-AS1PromoterDataFANTOM.txt
  - TRAF3IP2 Promoter Data.txt
  - TRAFREVTRAFASCTSFANTOMData3-4-15CS.sav
  - ListofMicroRNAs3-5-15CS.xlsx
  - lncRNAMicroarrayMasterSheet11-4-14CS.xlsx
  - lncRNASortedDataEdit3-28-14.xlsx
- ▼ Mice LncRNAs
  - Conservation\_DataTable.csv
  - GenomeCoordinateOutput.txt
  - GenomeCoordinateOutput2.txt
  - LincRNAs NFKB Mouse Guttman 2-27-15 CS.xls
  - lncRNA in Mouse Brain.txt
  - lncRNA Lists 2-26-15 CS.xlsx
  - lncRNAs enriched in Mouse and Human 2-24-15 CS.xls
  - mmc1 (1).pdf
  - mouse\_long\_ncRNAs\_with\_orthologous\_human\_transcripts.xls
  - pnas\_0706729105\_06729Table5.xls
  - pnas\_0706729105\_06729Table7.xls
  - New list of LncRNAs of Interest 1-20-15 CS.xlsx
  - NFKB Signaling.xlsx
  - NPP Format Tables and Figures lncRNA Paper
  - Redone P2 P3 westerns
  - Related Journal Articles
  - Tox Data Analysis
    - DrugsbyMonth.xls
    - NMS-ToxData2014Original
    - NMSData2-3-15CS.sav
    - NMSData2-5-15CS.sav
    - NMSDatabyMonth2-6-15CS (2).sav
    - NMSDatabyMonth2-6-15CS.sav
    - NMSDatazeros2-6-15CS.sav
    - NMSTOXAccess.accdb
    - ToxDataAnalysisSPSOutput.xlsx
    - WCME-NMSDataSort2-3-15CS.xlsx
    - WCME-NMSDataSort2-5-15CS.xlsx

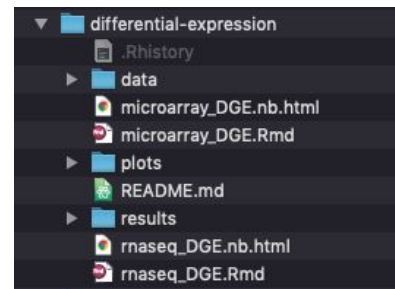
# Data Analysis Example 2: “GitHubbed Analysis”

**Methods Documentation:** Publicly available, online notebooks show step-by-step process that can be re-run.

**Data availability:** Data that was originally used is [available for download online](#).

**Project organization:** Folders labeled, most recent is present, with [prior versions tracked on GitHub](#).

**Software versions:** Specific requirements and version numbers shown in notebook and [README](#).



- 1) Install libraries
- 2) Import and set up data
- 3) Set up design matrix
- 4) Apply linear model
- 5) Explore fitness of model
- 6) Write statistics to output results file

## Differential Expression Analysis: Microarray

ALSF CCCL - Candace Savonen

Purpose: This notebook takes data and metadata from refine.bio and identifies differentially expressed genes. This script is generally applicable to microarray data.

### 1) Install libraries

This script uses the bioconductor R package limma to identify differentially expressed genes.

The full guide on [limma](#) shows examples of limma functions. Citation: Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015). “limma powers differential expression analyses for RNA-sequencing and microarray studies.” Nucleic Acids Research, 43(7), e47.

```
if (!("limma" %in% installed.packages())) {  
  # Install limma  
  BiocManager::install("limma", update = FALSE)  
}
```

Attach the `limma` library:

```
# Magrittr pipe  
"R" <- dplyr::`>`  
# Attach library  
library(limma)
```

Create output folders.

```
# Create the results folder if it doesn't exist  
if (!dir.exists("results")) {  
  dir.create("results")  
}  
# Create the plots folder if it doesn't exist  
if (!dir.exists("plots")) {  
  dir.create("plots")  
}
```

## refine.bio Example Workflow: Differential expression

refine.bio includes both microarray and RNA-seq experiments. The methods for performing differential gene expression (DGE) analysis for each technology differ. In this module, we include simple two-group comparison examples for each.

### Contents

- **Microarray notebook:** takes microarray data and metadata from refine.bio and identifies genes that are differentially expressed between two groups.
- **RNA-seq notebook:** takes RNA-seq data without quantile normalization from refine.bio and identifies genes that are differentially expressed between two groups. Read more about skipping quantile normalization [here](#) in our documentation.
- **GenePattern differential expression analysis:** `GenePattern` modules can be run via a GUI. To use refine.bio data with `GenePattern`, you will need to change the format as described in this section.

### Requirements and usage

This module requires you to install the following software to run examples yourself:

- **R**
- **RStudio** for working with [R Notebooks](#).
- **Bioconductor**
- **tidyverse**

These requirements can be installed by following the instructions at the links above. The example R Notebooks are designed to check if additional required packages are installed and will install them if they are not.

“100% Up to Date Analysis” OR “GitHubbed Analysis”

Which analysis would you...

...be more inclined to borrow the methods from?

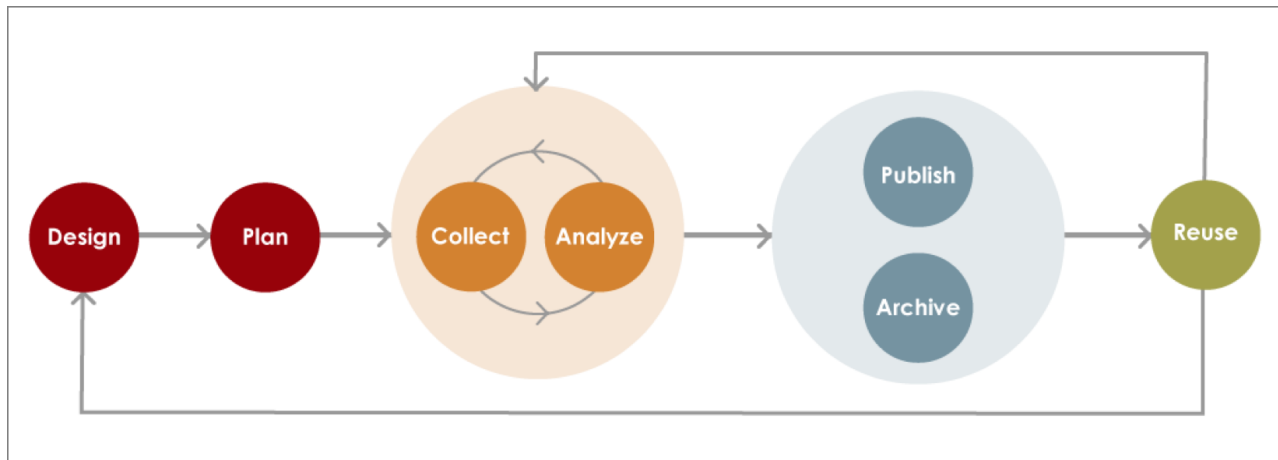
...trust the methods of more?

...feel would be *easier to reproduce*?



# A word about research reproducibility

- Thoughtful **data management** and **documentation** throughout the research life cycle:
  - Project organization, file naming conventions, file back up (3+2+1)
  - Document all steps, metadata, and computational environment
  - Version control and interoperability
  - Make data, protocol and code accessible, reusable and citable

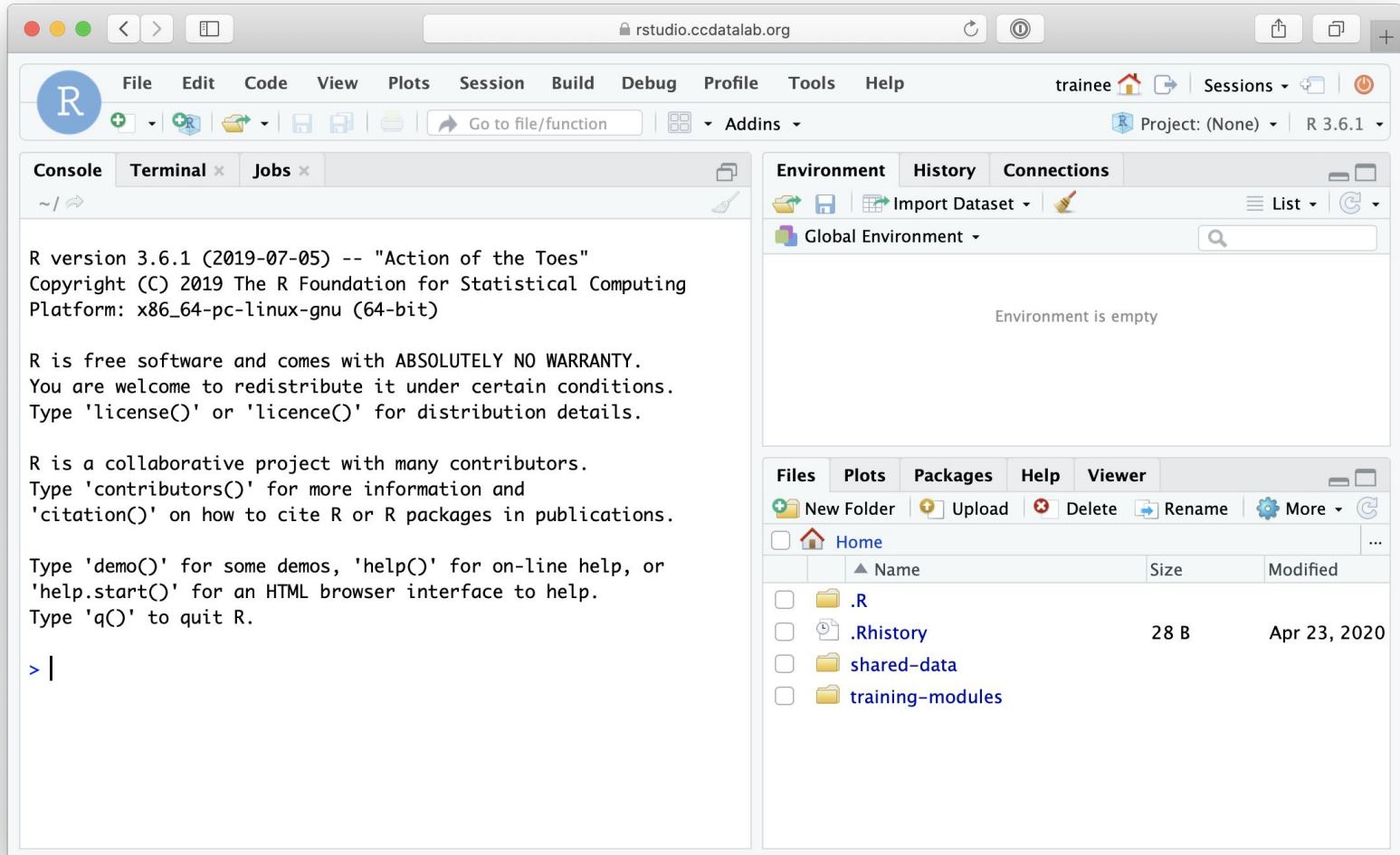


More resources: <https://guides.library.cmu.edu/reproducibility> or join our relevant workshops

# Command line vs GUI (graphics user interface)

- An interface is how you interact with a program
- GUI's have buttons you can *click* to do things,
- Command-line interfaces have you *type* out things to do them

# RStudio Server: A basic guide



The screenshot displays the RStudio Server web interface in a browser window. The address bar shows `rstudio.ccdatalab.org`. The top navigation bar includes menus for File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The user is logged in as 'trainee' and the R version is 3.6.1.

The interface is divided into several panels:

- Console:** Shows the R startup message and a prompt `> |`. The text in the console is:

```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```
- Environment:** Shows the 'Global Environment' and indicates that the environment is empty.
- Files:** A file browser showing the current directory structure:

|                          | Name               | Size | Modified     |
|--------------------------|--------------------|------|--------------|
| <input type="checkbox"/> | Home               |      |              |
| <input type="checkbox"/> | ▲ .R               |      |              |
| <input type="checkbox"/> | 📄 .Rhistory        | 28 B | Apr 23, 2020 |
| <input type="checkbox"/> | 📁 shared-data      |      |              |
| <input type="checkbox"/> | 📁 training-modules |      |              |





Console Terminal x Jobs x

~/

R version 3.6.1 (2019-07-05) -- "Action of the Toes"  
Copyright (C) 2019 The R Foundation for Statistical Computing  
Platform: x86\_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

&gt; |

**The Console:**  
where you tell R what to do through  
command line instructions

Environment History Connections

Import Dataset

Global Environment

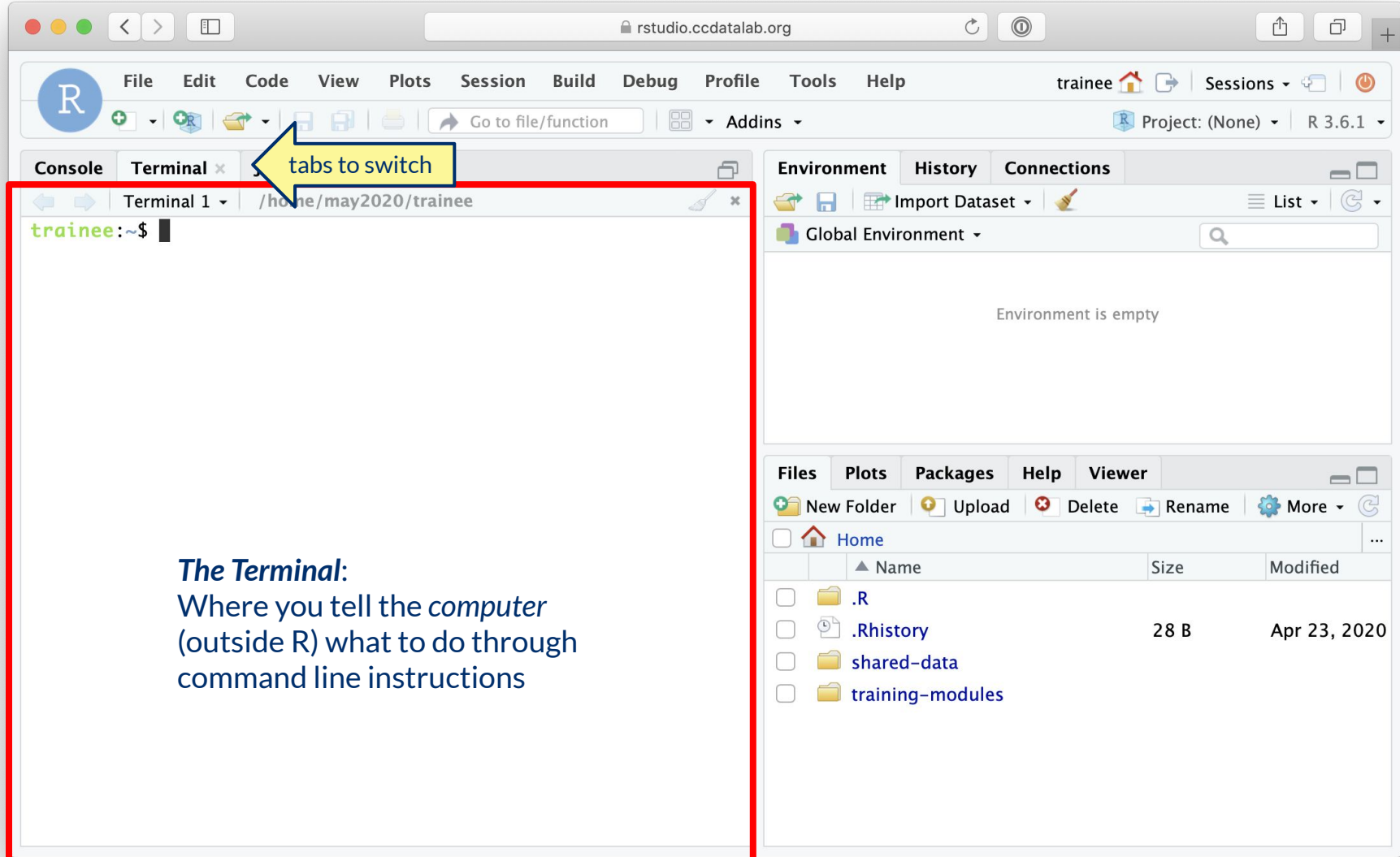
Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Home

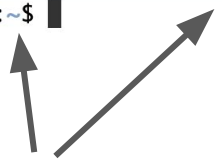
|                          | Name             | Size | Modified     |
|--------------------------|------------------|------|--------------|
| <input type="checkbox"/> | .R               |      |              |
| <input type="checkbox"/> | .Rhistory        | 28 B | Apr 23, 2020 |
| <input type="checkbox"/> | shared-data      |      |              |
| <input type="checkbox"/> | training-modules |      |              |



tabs to switch

**The Terminal:**  
Where you tell the *computer* (outside R) what to do through command line instructions

```
trainee:~$
```



These indicate what **directory** you are **currently** carrying out a command in

This is called your "current directory"

"~" is a shortcut for your "Home" directory, so these mean the same thing.

Global Environment

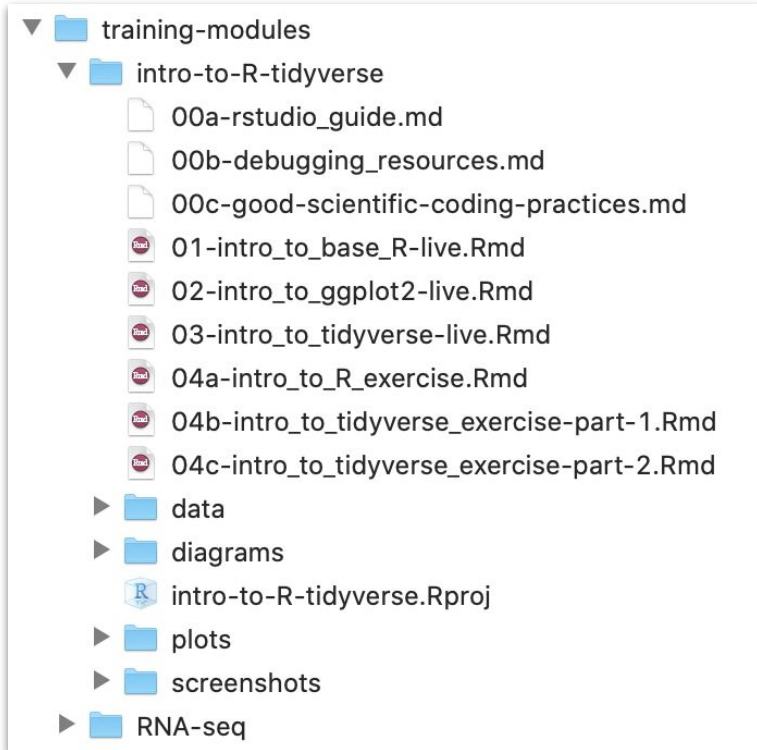
Environment is empty

New Folder Upload Delete Rename More

|                          | Name             | Size | Modified     |
|--------------------------|------------------|------|--------------|
| <input type="checkbox"/> | Home             |      |              |
| <input type="checkbox"/> | .R               |      |              |
| <input type="checkbox"/> | .Rhistory        | 28 B | Apr 23, 2020 |
| <input type="checkbox"/> | shared-data      |      |              |
| <input type="checkbox"/> | training-modules |      |              |

# Directories = Folders

When we are doing things through command line, we have to look at where the files we are using are being kept.



The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top right shows the user 'trainee', a home icon, a refresh icon, and a power icon. Below the menu bar is a toolbar with icons for adding files, saving, and navigating. The main workspace is divided into three panes: Console, Terminal, and Jobs. The Console pane shows the following output:

```
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
```

The Environment pane on the right shows 'Global Environment' and 'Environment is empty'. The Files pane at the bottom right shows a file explorer view of the home directory:

|                          | Name             | Size | Modified     |
|--------------------------|------------------|------|--------------|
| <input type="checkbox"/> | ▲ .R             |      |              |
| <input type="checkbox"/> | .Rhistory        | 28 B | Apr 23, 2020 |
| <input type="checkbox"/> | shared-data      |      |              |
| <input type="checkbox"/> | training-modules |      |              |

Below the terminal output, there is explanatory text:

**ls** - how to tell the computer to **list** the files and folders in a directory (files that start with a '.' are not shown by default)

**cd** - how to **change** directories

The screenshot shows the RStudio interface with the following components:

- Terminal Pane:** Shows a sequence of commands and their outputs:

```
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/training-modules$
```

An arrow points from the text below to the prompt `trainee:~/training-modules$`.
- Environment Pane:** Shows "Global Environment" with the message "Environment is empty".
- Files Pane:** Shows a file browser view of the current directory with the following table:

|                          | Name             | Size | Modified     |
|--------------------------|------------------|------|--------------|
| <input type="checkbox"/> | .R               |      |              |
| <input type="checkbox"/> | .Rhistory        | 28 B | Apr 23, 2020 |
| <input type="checkbox"/> | shared-data      |      |              |
| <input type="checkbox"/> | training-modules |      |              |

Note that the words before where our cursor is has changed to reflect that we are in the "training-modules" directory

**ls** - how to tell the computer to **list** the files and folders in a directory (files that start with a '.' are not shown by default)

**cd** - how to **change** directories



Console Terminal x Jobs x

Terminal 1 /home/may2020/trainee/training-modules

```
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/training-modules$
```

Environment History Connections

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

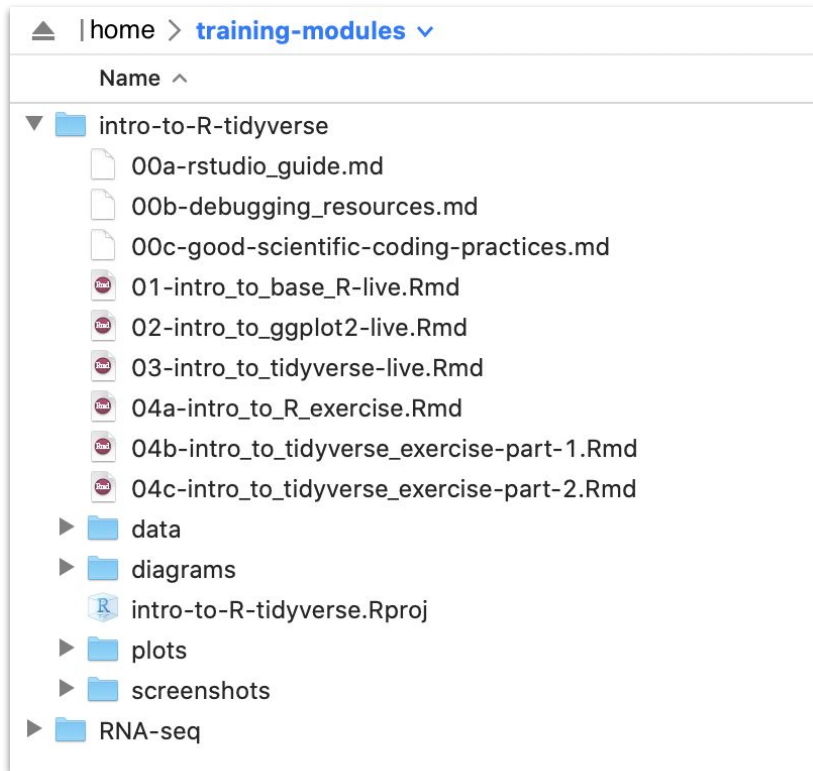
Home

|                          | Name             | Size | Modified     |
|--------------------------|------------------|------|--------------|
| <input type="checkbox"/> | .R               |      |              |
| <input type="checkbox"/> | .Rhistory        | 28 B | Apr 23, 2020 |
| <input type="checkbox"/> | shared-data      |      |              |
| <input type="checkbox"/> | training-modules |      |              |

The files tab over here  
does **NOT** reflect your current  
directory or any changes within it

# File paths: Directions to a file or folder

Let's say we want access to "01-intro\_to\_base\_R-live.Rmd"



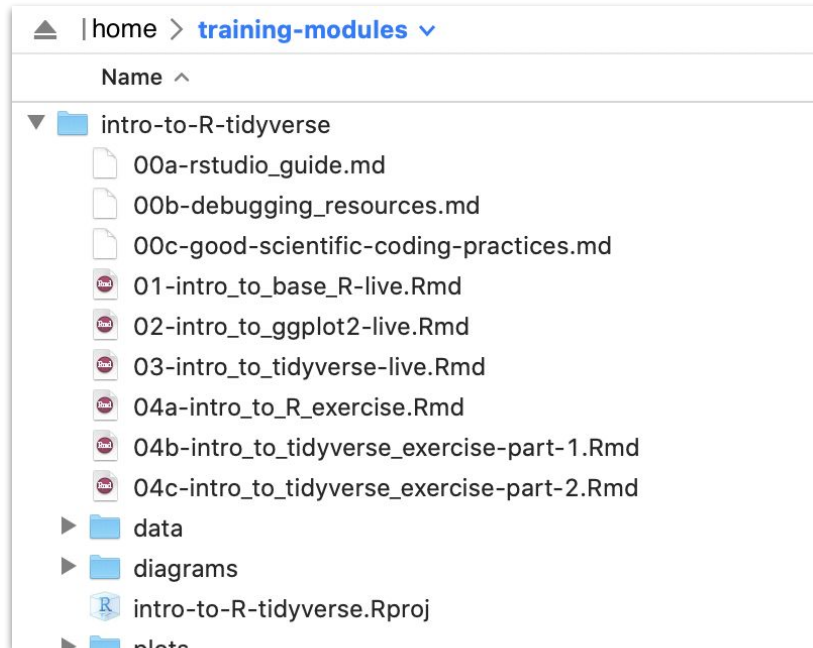
Current directory = "training-modules"

File path = "intro-to-R-tidyverse/01-intro\_to\_base\_R-live.Rmd"



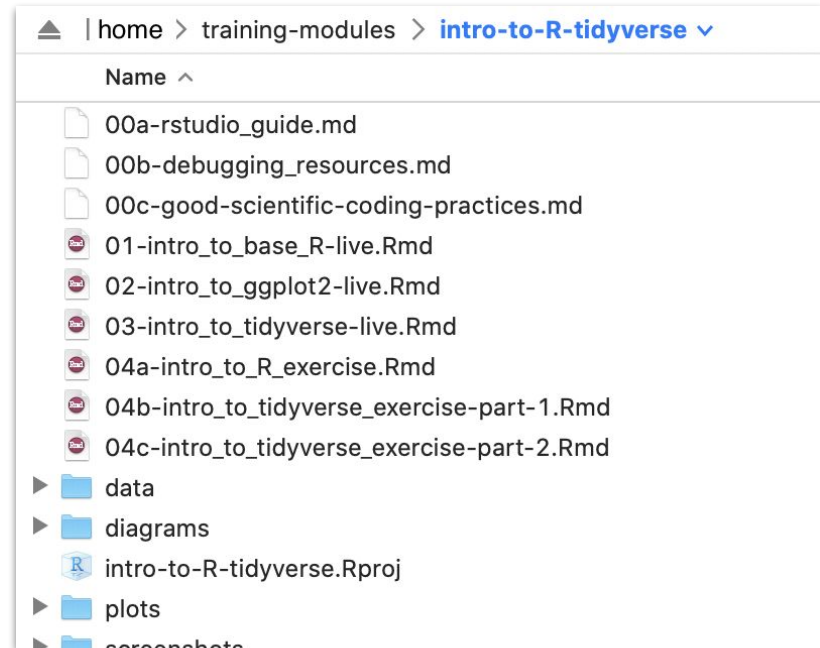
# File Paths can be *relative*

Let's say we want access to “01-intro\_to\_base\_R-live.Rmd”



Current directory = “training-modules”

Relative file path =  
“intro-to-R-tidyverse/01-intro\_to\_base\_R-live.Rmd”



Current directory = “training-modules/intro-to-R-tidyverse”

Relative file path = “01-intro\_to\_base\_R-live.Rmd”



# Introduction to R

The CCDL

# R programming

Programming: making executable scripts for accomplishing a task  
(in this case, data analysis is our task)

Scripts allow others to see, step-by-step, what you did.

## Why we use R:

- It's free
- People make cool packages that do stuff for us
- Many researchers in genomics use it (as well as Python)

## One in five genetics papers contains errors thanks to Microsoft Excel

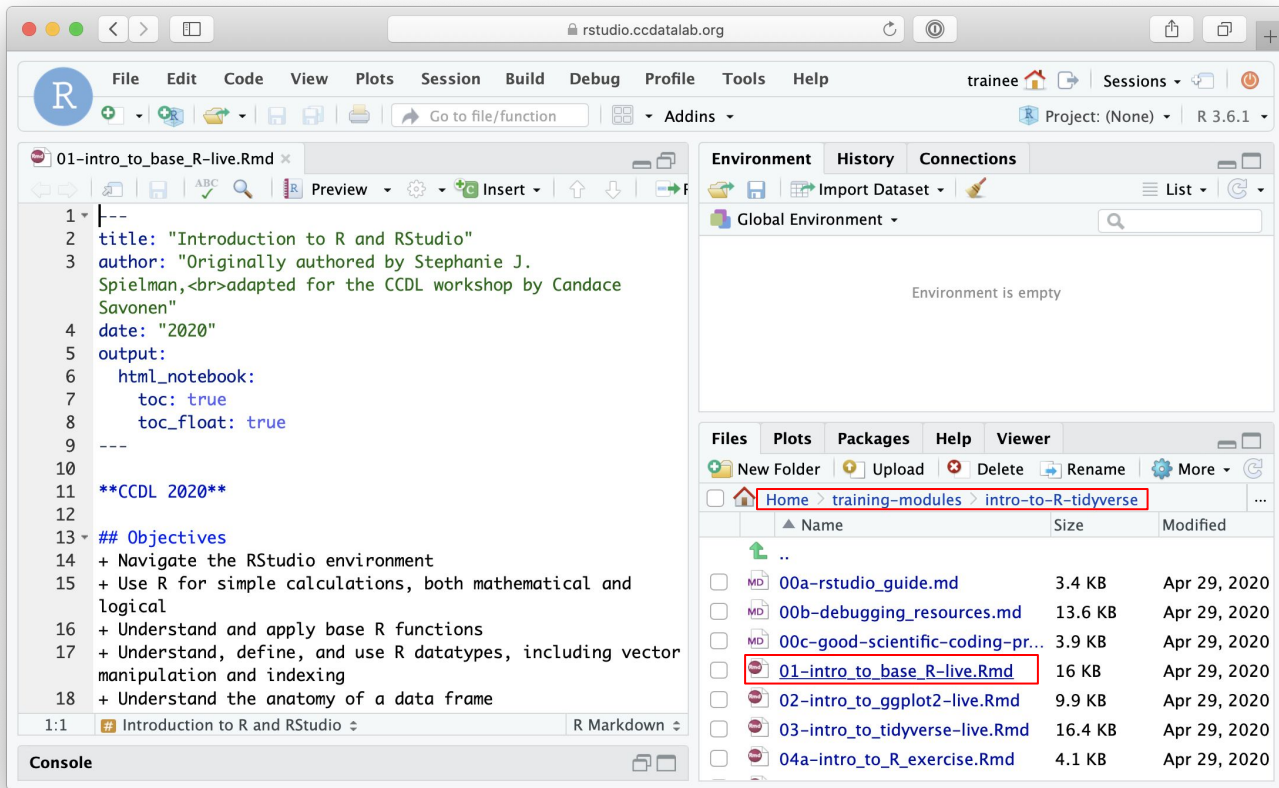
By [Jessica Boddy](#) | Aug. 29, 2016, 1:45 PM

| <b>What you type</b> | <b>What you see</b> | <b>How Excel stores it</b> |
|----------------------|---------------------|----------------------------|
| MARCH1               | 1-MAR               | 42430                      |
| SEPT2                | 2-SEP               | 42615                      |

<https://www.sciencemag.org/news/2016/08/one-five-genetics-papers-contains-errors-thanks-microsoft-excel>  
Ziemann et al. Genome Biology (2016) 17:177 DOI 10.1186/s13059-016-1044-7

# R Notebooks

Use the "Files" tab to open: [training-modules/intro-to-R-tidyverse/01-intro\\_to\\_base\\_R-live.Rmd](#)



The screenshot displays the RStudio web interface in a browser window. The main editor shows the content of the R Notebook file '01-intro\_to\_base\_R-live.Rmd'. The file content includes a title, author information, a date, output options, and a list of objectives for the workshop.

```
1 |---
2 |title: "Introduction to R and RStudio"
3 |author: "Originally authored by Stephanie J.
  |Spielman, <br>adapted for the CCDL workshop by Candace
  |Savonen"
4 |date: "2020"
5 |output:
6 |  html_notebook:
7 |    toc: true
8 |    toc_float: true
9 |---
10 |
11 |**CCDL 2020**
12 |
13 |## Objectives
14 |+ Navigate the RStudio environment
15 |+ Use R for simple calculations, both mathematical and
  |+ logical
16 |+ Understand and apply base R functions
17 |+ Understand, define, and use R datatypes, including vector
  |+ manipulation and indexing
18 |+ Understand the anatomy of a data frame
19 |
20 |1:1 Introduction to R and RStudio
```

The 'Files' pane on the right shows the directory structure: Home > training-modules > intro-to-R-tidyverse. The file '01-intro\_to\_base\_R-live.Rmd' is highlighted with a red box, indicating it is the selected file.

| Name                             | Size    | Modified     |
|----------------------------------|---------|--------------|
| ..                               |         |              |
| 00a-rstudio_guide.md             | 3.4 KB  | Apr 29, 2020 |
| 00b-debugging_resources.md       | 13.6 KB | Apr 29, 2020 |
| 00c-good-scientific-coding-pr... | 3.9 KB  | Apr 29, 2020 |
| 01-intro_to_base_R-live.Rmd      | 16 KB   | Apr 29, 2020 |
| 02-intro_to_ggplot2-live.Rmd     | 9.9 KB  | Apr 29, 2020 |
| 03-intro_to_tidyverse-live.Rmd   | 16.4 KB | Apr 29, 2020 |
| 04a-intro_to_R_exercise.Rmd      | 4.1 KB  | Apr 29, 2020 |

# R Notebooks

- R Notebooks allow you to have files that show both your code and results

Executable **code chunk**

Can **click** here to run a code chunk

The screenshot shows the RStudio interface with a code editor window titled '01-intro\_to\_base\_R-live.Rmd'. The code editor contains the following text:

```
82 For example, we can do some simple multiplication like
83 this.
84 When you execute code within the notebook, the results
85 appear beneath the code.
86 Try executing this chunk by clicking the *Run* button
87 within the chunk or by
88 placing your cursor inside it and pressing
89 *Cmd+Shift+Enter*.
90
91 Use the console to calculate other expressions. Standard
92 order of operations applies (mostly), and you can use
93 parentheses `()` as you might expect (but not brackets `[]`
94 or braces `{}`, which have special meanings). Note however,
95 that you must **always** specify multiplication with `*`;
96 implicit multiplication such as `10(3 + 4)` or `10x` will
```

A blue box highlights the code chunk starting at line 87. A blue arrow points from the text 'Executable code chunk' to this box. A green arrow points from the text 'Can click here to run a code chunk' to the green 'Run' button (a play icon) in the top right corner of the code chunk box.

Below the code chunk, the output is displayed in a separate box:

```
[1] 30
```

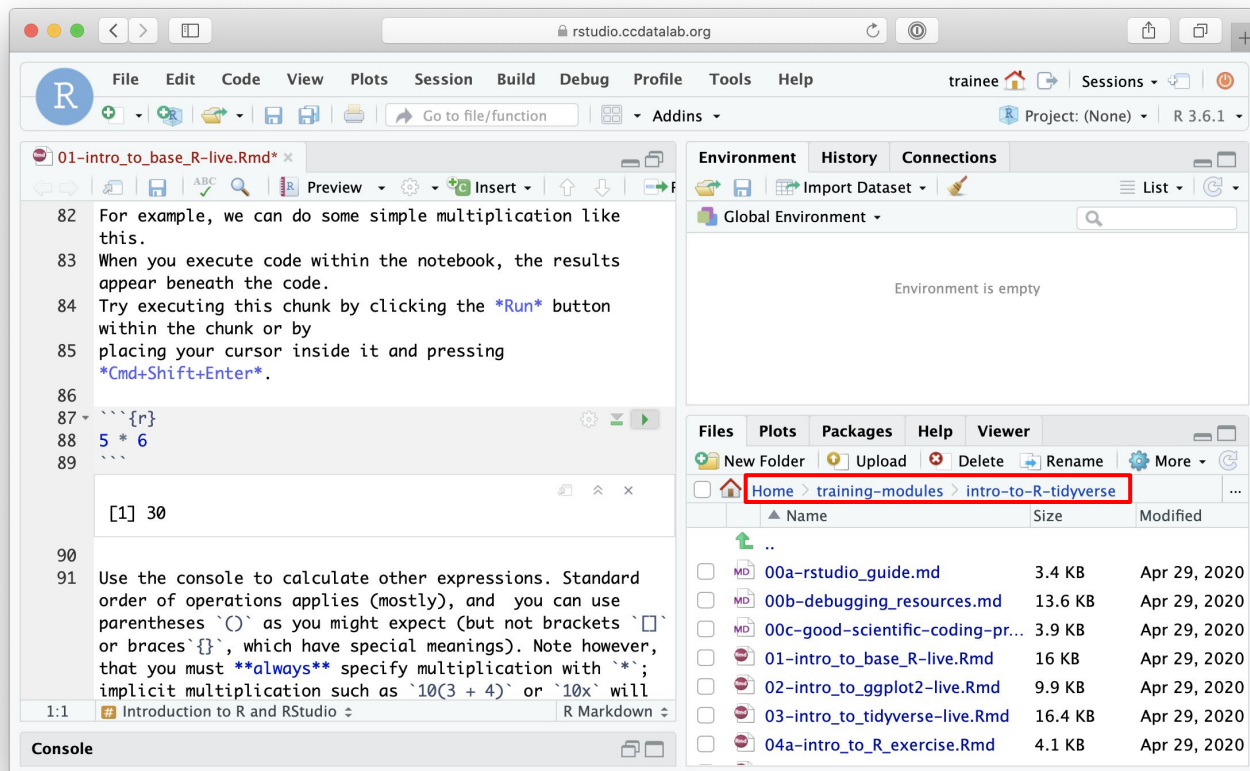
A red arrow points from the text 'Output from above code chunk' to this output box.

The RStudio interface also shows the 'Environment' pane on the right, which is currently empty, and the 'Files' pane at the bottom right, which shows a file list including '01-intro\_to\_base\_R-live.Rmd'.

Output from above code chunk

# R Notebooks

- Code that runs in R Notebooks uses wherever the file is saved as its *current directory*





01-intro\_to\_base\_R-live.Rmd\*

```
82 For example, we can do some simple multiplication like
83 this.
84 When you execute code within the notebook, the results
85 appear beneath the code.
86 Try executing this chunk by clicking the *Run* button
87 within the chunk or by
88 placing your cursor inside it and pressing
89 *Cmd+Shift+Enter*.
```

```
87 ```{r}
88 5 * 6
89 ```
```

```
[1] 30
```

```
90
91 Use the console to calculate other expressions. Standard
order of operations applies (mostly), and you can use
parentheses `()` as you might expect (but not brackets `[]`
or braces `{}`), which have special meanings). Note however,
that you must **always** specify multiplication with `*`;
implicit multiplication such as `10(3 + 4)` or `10x` will
```

```
1:1 # Introduction to R and RStudio
```

R Markdown

Console



Environment History Connections

Import Dataset

List

Global Environment

Environment is empty

Files Plots Packages Help Viewer

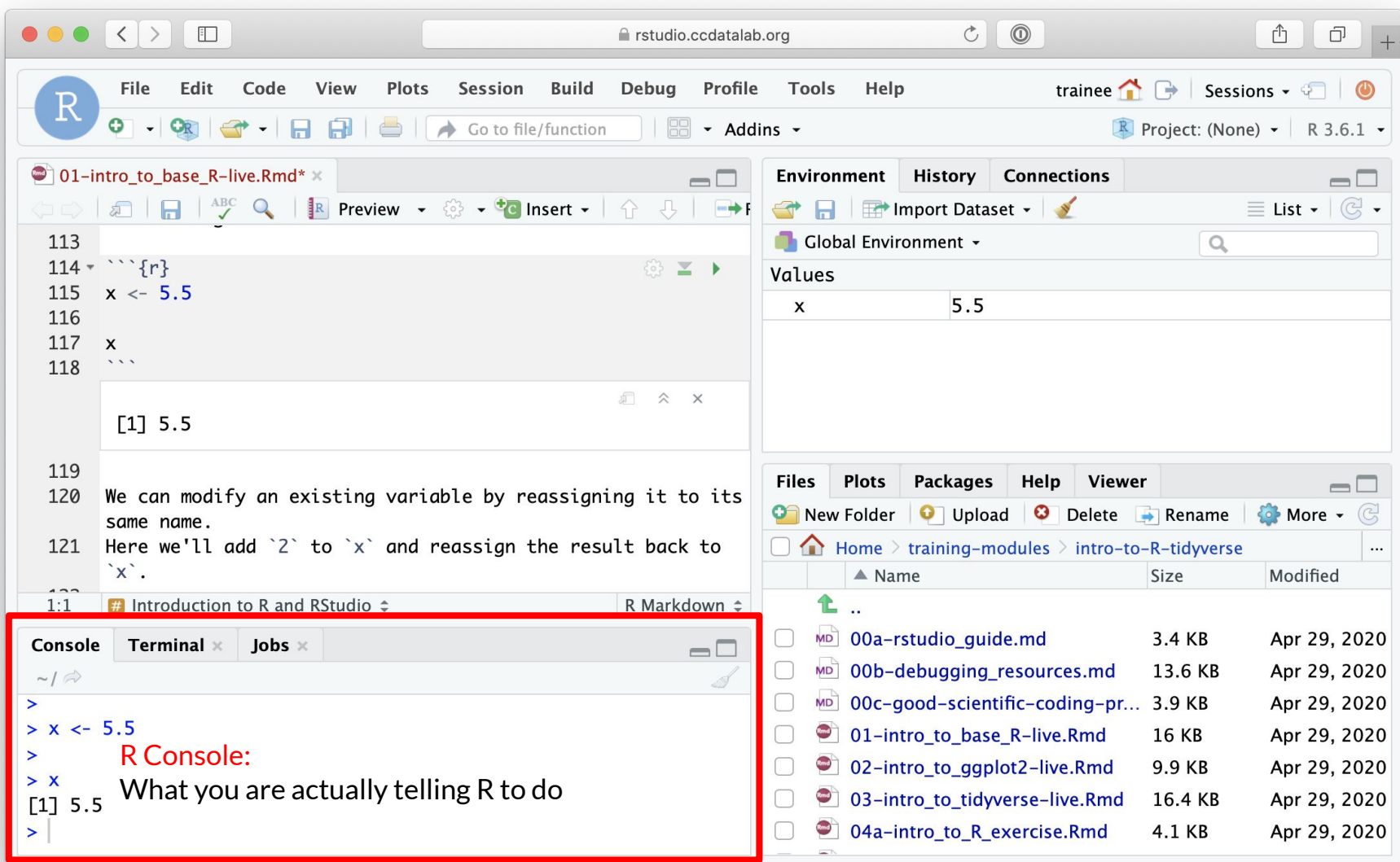
New Folder Upload Delete Rename More

Home &gt; training-modules &gt; intro-to-R-tidyverse

|                          | Name                           | Size    | Modified     |
|--------------------------|--------------------------------|---------|--------------|
|                          | ..                             |         |              |
| <input type="checkbox"/> | 00a-rstudio_guide.md           | 3.4 KB  | Apr 29, 2020 |
| <input type="checkbox"/> |                                | 3.6 KB  | Apr 29, 2020 |
| <input type="checkbox"/> |                                | 9 KB    | Apr 29, 2020 |
| <input type="checkbox"/> |                                | 6 KB    | Apr 29, 2020 |
| <input type="checkbox"/> |                                | 9 KB    | Apr 29, 2020 |
| <input type="checkbox"/> | 03-intro_to_tidyverse-live.Rmd | 16.4 KB | Apr 29, 2020 |
| <input type="checkbox"/> | 04a-intro_to_R_exercise.Rmd    | 4.1 KB  | Apr 29, 2020 |

Click here to show the Console







01-intro\_to\_base\_R-live.Rmd\*

Preview Insert

```
113
114 ` ``{r}
115 x <- 5.5
116
117 x
118 ` ``
```

**R Script or Notebook:**  
Where you are writing and editing  
what you will tell R or Terminal

```
[1] 5.5
```

```
119
120 We can modify an existing variable by reassigning it to its
    same name.
121 Here we'll add `2` to `x` and reassign the result back to
    `x`.
122
```

1:1 Introduction to R and RStudio

R Markdown

Console Terminal Jobs

```
>
> x <- 5.5
>
> x
[1] 5.5
> |
```

Environment History Connections

Import Dataset

Global Environment

Values

|   |     |
|---|-----|
| x | 5.5 |
|---|-----|

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Home &gt; training-modules &gt; intro-to-R-tidyverse

|  | Name                             | Size    | Modified     |
|--|----------------------------------|---------|--------------|
|  | ..                               |         |              |
|  | 00a-rstudio_guide.md             | 3.4 KB  | Apr 29, 2020 |
|  | 00b-debugging_resources.md       | 13.6 KB | Apr 29, 2020 |
|  | 00c-good-scientific-coding-pr... | 3.9 KB  | Apr 29, 2020 |
|  | 01-intro_to_base_R-live.Rmd      | 16 KB   | Apr 29, 2020 |
|  | 02-intro_to_ggplot2-live.Rmd     | 9.9 KB  | Apr 29, 2020 |
|  | 03-intro_to_tidyverse-live.Rmd   | 16.4 KB | Apr 29, 2020 |
|  | 04a-intro_to_R_exercise.Rmd      | 4.1 KB  | Apr 29, 2020 |

The image shows the RStudio interface with the following components:

- Code Editor:** Contains R code for a live session:

```
113  
114 ```{r}  
115 x <- 5.5  
116  
117 x  
118 ```  
[1] 5.5  
119  
120 We can modify an existing variable by reassigning it to its  
    same name.  
121 Here we'll add `2` to `x` and reassign the result back to  
    `x`.
```
- Environment Pane:** Shows the current R environment with a search bar and a table of values:

| Global Environment |     |
|--------------------|-----|
| x                  | 5.5 |
- Console:** Shows the execution of the code:

```
>  
> x <- 5.5  
>  
> x  
[1] 5.5  
> |
```
- Files Pane:** Shows a file explorer view of the current project directory:

| Name                             | Size    | Modified     |
|----------------------------------|---------|--------------|
| ..                               |         |              |
| 00a-rstudio_guide.md             | 3.4 KB  | Apr 29, 2020 |
| 00b-debugging_resources.md       | 13.6 KB | Apr 29, 2020 |
| 00c-good-scientific-coding-pr... | 3.9 KB  | Apr 29, 2020 |
| 01-intro_to_base_R-live.Rmd      | 16 KB   | Apr 29, 2020 |
| 02-intro_to_ggplot2-live.Rmd     | 9.9 KB  | Apr 29, 2020 |
| 03-intro_to_tidyverse-live.Rmd   | 16.4 KB | Apr 29, 2020 |
| 04a-intro_to_R_exercise.Rmd      | 4.1 KB  | Apr 29, 2020 |

**R environment:**  
What R knows and remembers for you

The image shows the RStudio interface with the following components:

- Code Editor:** Contains R code for assigning the value 5.5 to variable x. The output shows [1] 5.5.
- Environment Pane:** Shows the variable x with the value 5.5.
- Files Pane:** A red box highlights the Files pane, which shows a directory listing of files in the 'intro-to-R-tidyverse' folder.

**Other Assistance Tabs:**  
Things that help you in your coding

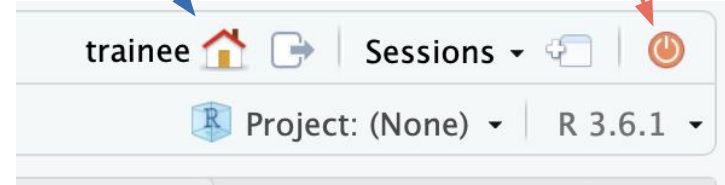
| Name                             | Size    | Modified     |
|----------------------------------|---------|--------------|
| ..                               |         |              |
| 00a-rstudio_guide.md             | 3.4 KB  | Apr 29, 2020 |
| 00b-debugging_resources.md       | 13.6 KB | Apr 29, 2020 |
| 00c-good-scientific-coding-pr... | 3.9 KB  | Apr 29, 2020 |
| 01-intro_to_base_R-live.Rmd      | 16 KB   | Apr 29, 2020 |
| 02-intro_to_ggplot2-live.Rmd     | 9.9 KB  | Apr 29, 2020 |
| 03-intro_to_tidyverse-live.Rmd   | 16.4 KB | Apr 29, 2020 |
| 04a-intro_to_R_exercise.Rmd      | 4.1 KB  | Apr 29, 2020 |

# RStudio Sessions

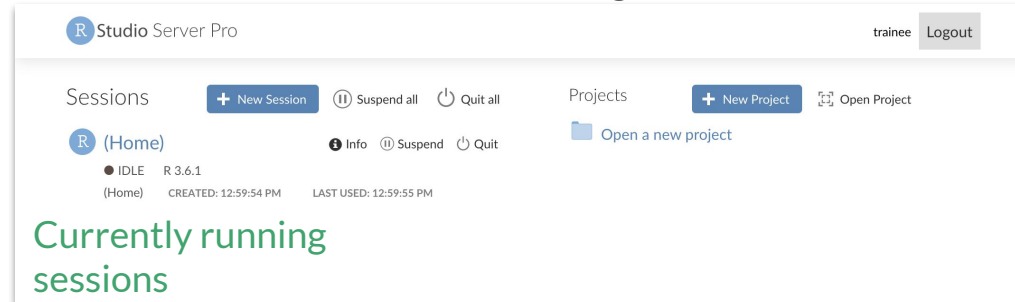
- On the server, R run is running many times at once
  - Each user has their own “**Session**” running, with its own memory and processes
  - It is possible for a user to have more than one session at a time
- We will usually want to start new sessions between notebooks to keep the environment clean

Go to the sessions page

End the current session



## Session Page

A screenshot of the RStudio Server Pro interface. At the top, it says 'Studio Server Pro' and 'trainee Logout'. Below this, there are two main sections: 'Sessions' and 'Projects'. The 'Sessions' section has buttons for '+ New Session', 'Suspend all', and 'Quit all'. Below it, there is a session entry for '(Home)' with status 'IDLE', version 'R 3.6.1', and timestamps 'CREATED: 12:59:54 PM' and 'LAST USED: 12:59:55 PM'. The 'Projects' section has buttons for '+ New Project' and 'Open Project', and a link 'Open a new project'. At the bottom left, the text 'Currently running sessions' is displayed in green.