

ANT 6973

DATA VISUALIZATION AND EXPLORATION

PLAN FOR TODAY

1. Introductions

2. Orientation

- Syllabus and course website

- R and RStudio

3. Complete your first in-class activity!

WHO AM I?

Dr. Fernando Campos

- Department of Anthropology
- Email: fernando.campos@utsa.edu
- Office Hours:
 - Tuesdays and Thursdays, 10:00 to 11:00
 - By appointment



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.

Taught at Tulane University, New Orleans.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.

Taught at Tulane University, New Orleans.

Postdoc at Duke University, NC.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.

Taught at Tulane University, New Orleans.

Postdoc at Duke University, NC.

Started at UTSA in 2018!



WHAT DO I DO?

Dr. Fernando Campos

I'm a biological anthropologist, studying the behavioral ecology of wild primates.





Nicaragua

Costa Rica

Panama



Nicaragua

Área de Conservación
Guanacaste

Costa Rica

Panama





WHITE-FACED CAPUCHIN



MANTLED HOWLER



BLACK-HANDED SPIDER MONKEY



Kenya

Amboseli
National Park

Tanzania

AMBOSELI BABOON RESEARCH PROJECT



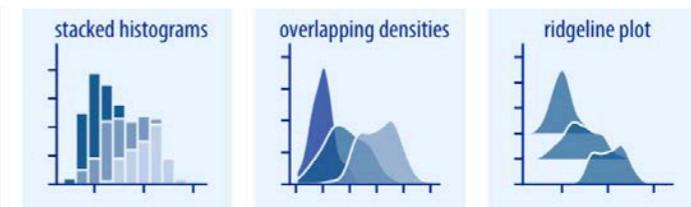
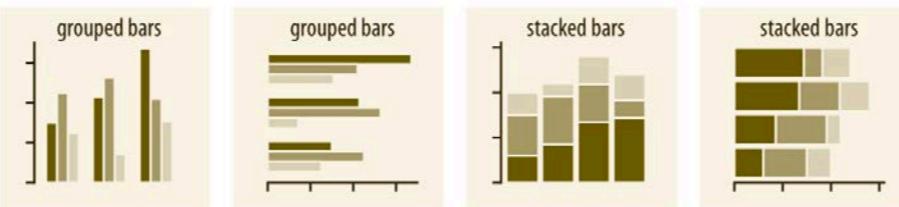
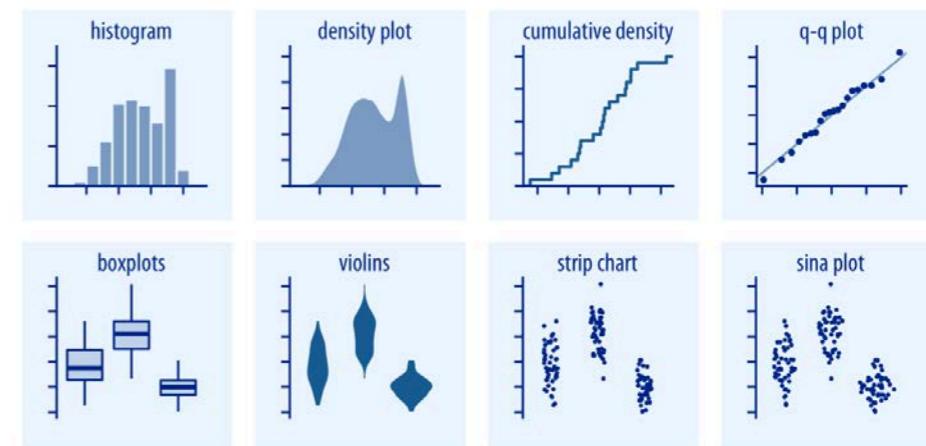




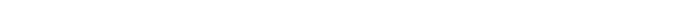
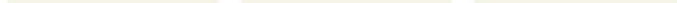
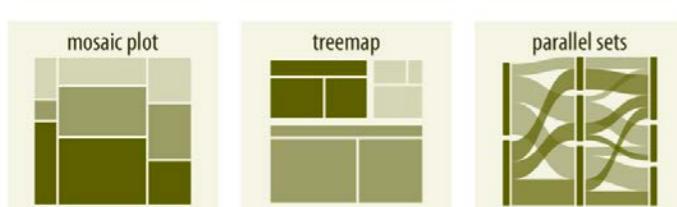
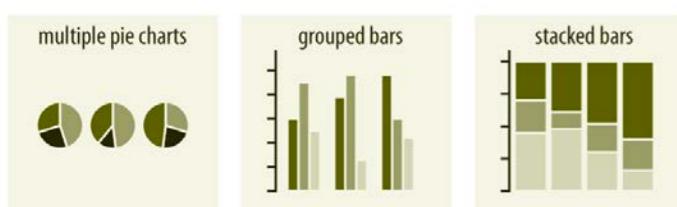
WHY AM I TEACHING
THIS COURSE?

DIRECTORY OF VISUALIZATIONS

Amounts



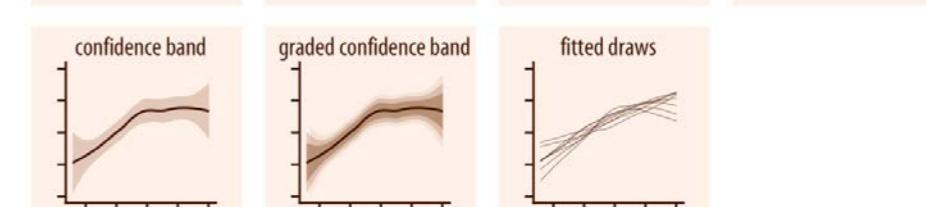
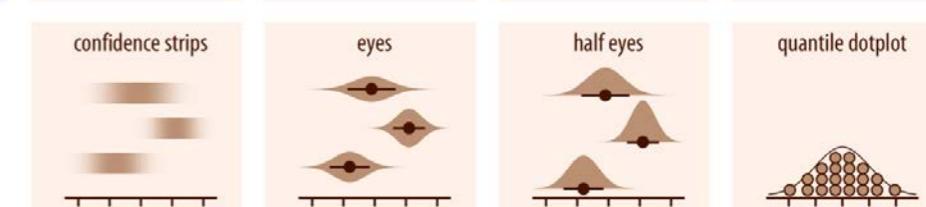
Proportions



X-Y Relationships

Distributions

Uncertainty



YOUR TURN

1. Who are you?
2. What are your research interests?
3. What is your experience with data?
4. Do you have any experience using R, and if so, how much?
5. Is there anything specific that you hope to learn in this course?

**HELLO
my name is**

15:00

ORIENTATION

ORIENTATION

- Syllabus and course website
- R and RStudio

COURSE WEBSITE

- <https://www.campos-lab.net/courses/data-viz-course/>
- Bookmark it!
- Go here for up-to-date information, schedule, assignments, readings, etc.
- I'll use Blackboard only for posting grades.
- Let's explore this a bit...

ANT 6937, Spring 2019

HOME DATA CAMP CHEAT SHEETS SCHEDULE POLICIES

Instructor Dr. Fernando Campos

Location McKinney Humanities (MH) 3.02.52

Times Thu 9:00 am - 11:45 am

Course ANT 6937.002

CRN 34610

Office Hours Tue/Thu, 2:00-3:30 pm, MH 4.03.44

Overview

This course is designed for graduate students in social or natural sciences who would like to gain experience in data visualization and analysis using R. It aims to be accessible to those who are completely new to programming, but also to provide new insights and skills to those who already have experience using R.

Approach

This course will use a mix of traditional and [flipped classroom](#) approaches, with new material introduced in reading assignments prior to class while class time will focus on applying those skills to data sets. The course will be activity-focused. Most concepts will be first introduced in assigned readings outside of class, leaving class time to focus on developing practical skills, completing activities, making progress on individual projects. We will cover topics in both data visualization as well as data manipulation. Students are encouraged to work collaboratively in and out of class. The [online course schedule](#) provides an overview of the topics covered as well as links to weekly readings, in-class activities, and any lecture material. This schedule is preliminary and always subject to change.

Texts

We will read sections from four openly licensed, online textbooks for this course. Hard copies of the books are *not* required, but some can be found at the links below.

- [Fundamentals of Data Visualization](#) by Claus Wilke.
- [Data Visualization: A practical introduction](#) by Kieran Healy. | [Hard copy](#).
- [ModernDive: An Introduction to Statistical and Data Sciences via R](#) by Chester Ismay and Albert Y. Kim.
- [R For Data Science](#) by Hadley Wickham and Garrett Grolemund. | [Hard copy](#).

Additional reading material will be linked from the [schedule](#).

GITHUB TEXTBOOK: WILKE TEXTBOOK: HEALY TEXTBOOK: MODERNDIVE TEXTBOOK: R4DS CC BY

ORIENTATION

- Syllabus and course website
- R and RStudio

WHAT'S R VS. RSTUDIO?

- R is a programming language.
- RStudio is a nice user interface for R (an *integrated development environment* or IDE) that provides many convenient features and tools for using R.



Engine



Dashboard

WHY R?



- R is the premier data analysis and visualization platform.

```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r

R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

> █
```

WHY LEARN R?

- **Pros (1 of 2):**

- Powerful, state-of-the-art, and used by many researchers, R is becoming a standard for data analysis
- Easy to extend, modify, and improve with add-on packages
 - **Programmable:** if R can't do a particular task right now, you (or someone) can program it
- Freely available for Windows, Mac, Unix



```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.
> |
```

WHY LEARN R?

- **Pros (2 of 2):**

- Reproducibility: R relies on a series of *written commands* rather than pointing and clicking (and this is a good thing!)
- Powerful plotting packages produce publication-quality graphics.
- R has a large and welcoming community.



```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.

> █
```

WHY LEARN R?

- **Cons:**

- Command-based rather than interactive in the point-and-click sense (like SPSS & Excel)
 - Can be significant learning hurdle
 - Specific commands easy to forget
- Documentation sometimes poor
- Still evolving: backward-compatibility has been an issue



```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

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.
> |
```

- Let's make sure we all have a current installation of R.
- <https://cloud.r-project.org/>
- If you already have R installed, it's still a good idea to update.

ANT 6973, Spring 2021 The Comprehensive R Archive Network cloud.r-project.org



The Comprehensive R Archive Network

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 (2020-10-10, Bunny-Wunnies Freak Out) [R-4.0.3.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](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.

Submitting to CRAN

To "submit" a package to CRAN, check that your submission meets the [CRAN Repository Policy](#) and then use the [web form](#).

If this fails, upload to <ftp://CRAN.R-project.org/incoming/> and send an email to CRAN-submissions@R-project.org following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

Note that we generally do not accept submissions of precompiled binaries due to security reasons. All binary distribution listed above are compiled by selected maintainers, who are in charge for all binaries of their platform, respectively.

For queries about this web site, please contact [the webmaster](#).

- And a current installation of RStudio.
- <https://www.rstudio.com/>
- If you already have RStudio installed, check for a new version using menus:
 - Help > Check for Updates

Screenshot of the RStudio download page (rstudio.com/products/rstudio/download/#download) showing the download process and available installers.

RStudio Desktop 1.4.1103 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:

DOWNLOAD RSTUDIO FOR MAC
1.4.1103 | 152.77MB
Requires macOS 10.13+ (64-bit)

All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy. RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

| OS | Download | Size | SHA-256 |
|---------------------|---|-----------|-----------|
| Windows 10/8/7 | RStudio-1.4.1103.exe | 156.96 MB | c3384189 |
| macOS 10.13+ | RStudio-1.4.1103.dmg | 152.77 MB | 20148bd6 |
| Ubuntu 16 | rstudio-1.4.1103-amd64.deb | 119.26 MB | f0857e27 |
| Ubuntu 18/Debian 10 | rstudio-1.4.1103-amd64.deb | 120.30 MB | 76864349 |
| Fedora 19/Red Hat 7 | rstudio-1.4.1103-x86_64.rpm | 138.02 MB | 8fcfb2d29 |
| Fedora 28/Red Hat 8 | rstudio-1.4.1103-x86_64.rpm | 138.01 MB | e2bf11e9 |
| Debian 9 | rstudio-1.4.1103-amd64.deb | 120.45 MB | 4a4d159c |
| OpenSUSE 15 | rstudio-1.4.1103-x86_64.rpm | 122.02 MB | fdc33f7a |

Zip/Tarballs

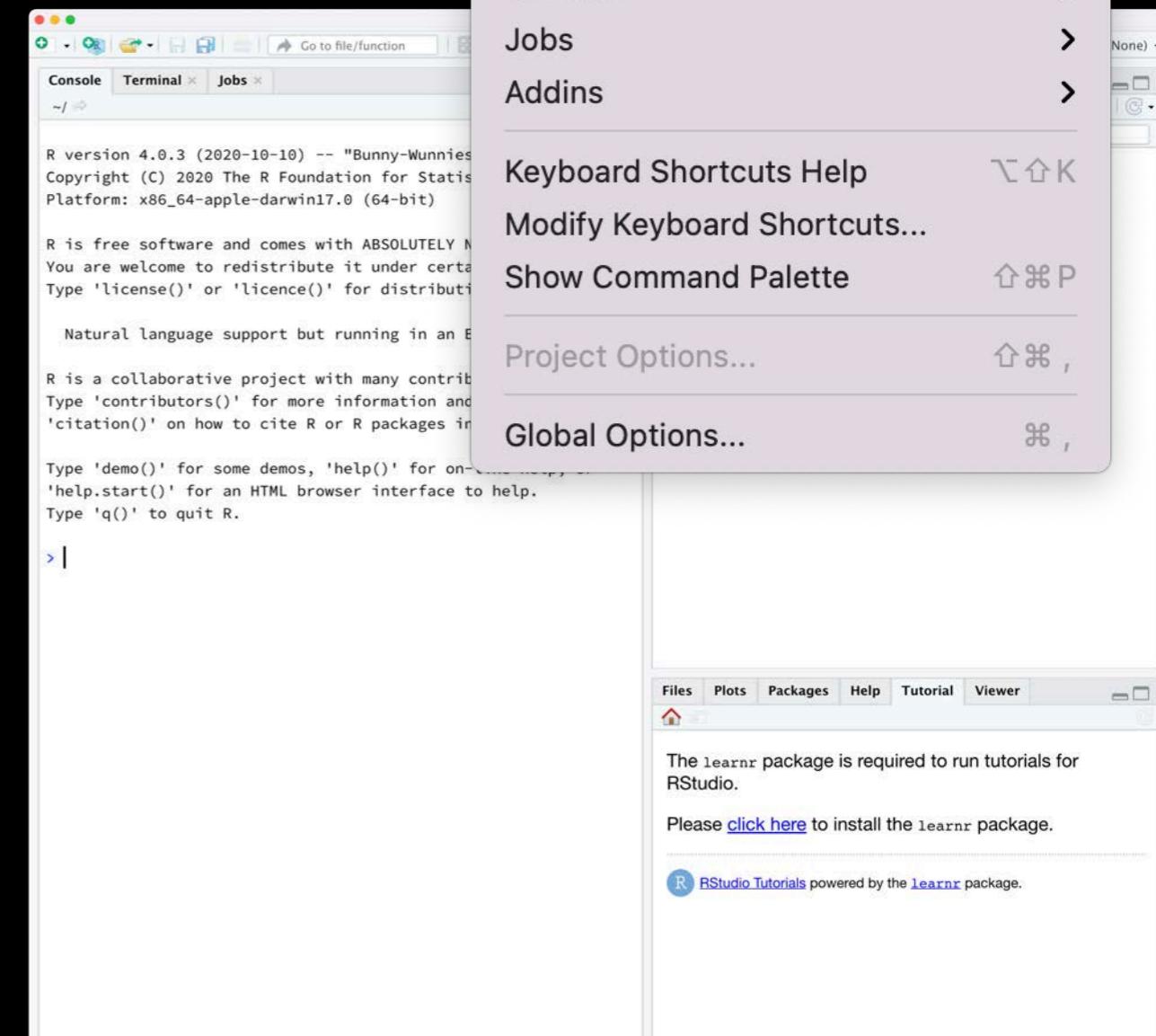
| OS | Zip/tar | Size | SHA-256 |
|---------------------|--|-----------|----------|
| Windows 10/8/7 | RStudio-1.4.1103.zip | 229.08 MB | 8e0d97c7 |
| Ubuntu 16 | rstudio-1.4.1103-amd64-debian.tar.gz | 176.95 MB | f45b37e0 |
| Ubuntu 18/Debian 10 | rstudio-1.4.1103-amd64-debian.tar.gz | 176.57 MB | b139d776 |
| Fedora 19/Red Hat 7 | rstudio-1.4.1103-x86_64-fedorat.tar.gz | 176.68 MB | 64017897 |
| Debian 9 | rstudio-1.4.1103-amd64-debian.tar.gz | 176.92 MB | 79f29fd1 |

Source Code

A tarball containing source code for RStudio 1.4.1103 can be downloaded from [here](#).



- If you already have RStudio installed, now would be a good time to **update all your packages**.
- In RStudio, Tools > Check for Package Updates...





RStudio

Project: (None)

Console ~/ >

Environment History

Import Dataset List Global Environment

Environment is empty

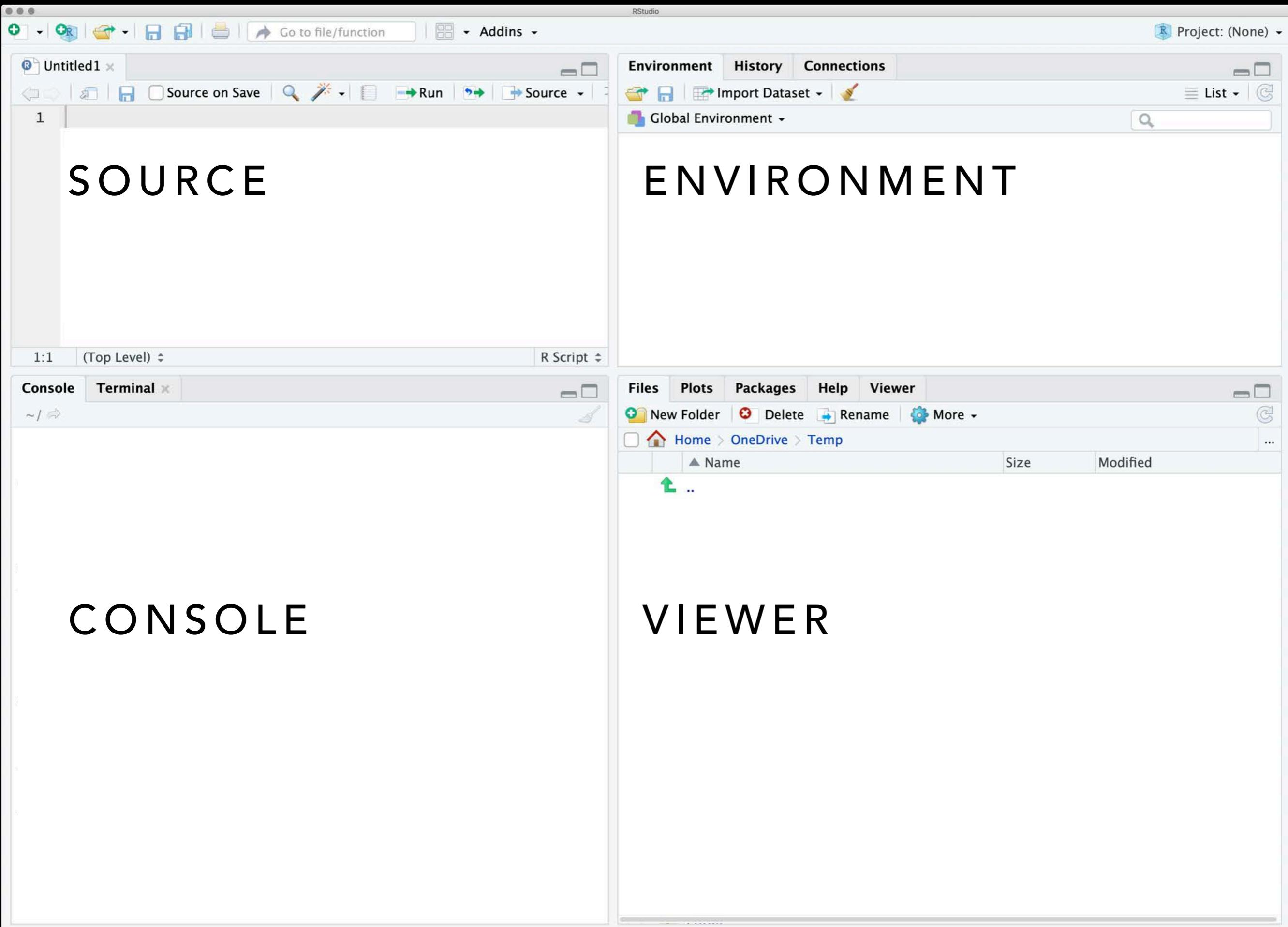
Files Plots Packages Help Viewer

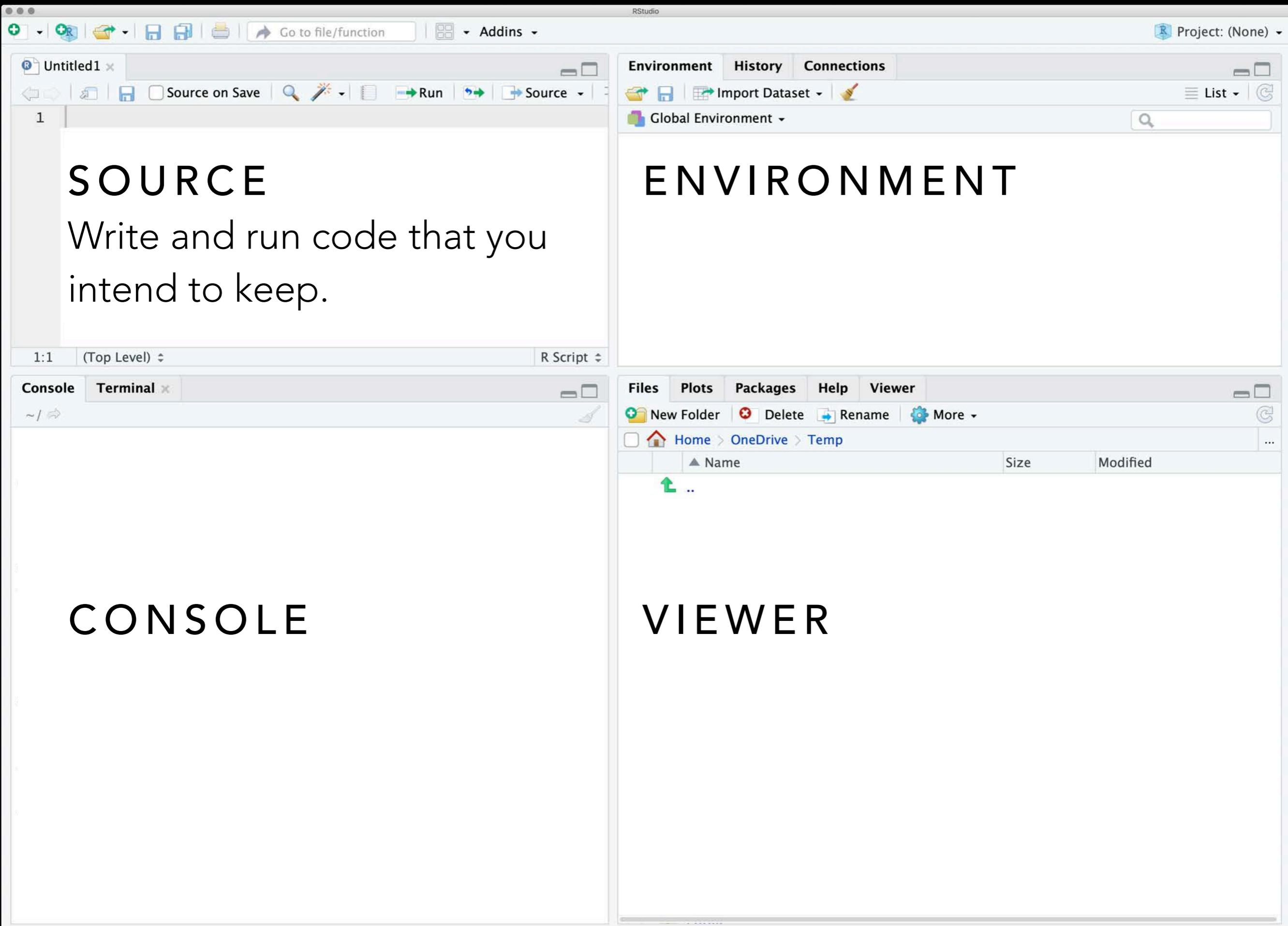
New Folder Delete Rename More

Home > Public

| Name | Size | Mo |
|------|------|----|
| .. | | |

This screenshot shows the RStudio interface. At the top is the RStudio logo. Below it is the main window with several panes. The leftmost pane is the 'Console' showing a command prompt. To its right is the 'Environment' pane, which displays the message 'Environment is empty'. Further right is the 'Files' pane, showing a directory structure starting with 'Home > Public'. The bottom right corner of the 'Files' pane contains a small table with three columns: 'Name', 'Size', and 'Mo'. A single row is present in the table, labeled with a green double-headed arrow icon and the text '..'. The top menu bar includes 'File', 'Edit', 'View', 'Code', 'Tools', 'Help', and 'About'. The top toolbar contains various icons for file operations like opening, saving, and printing, along with a 'Go to file/function' search bar and an 'Addins' dropdown.





The image shows the RStudio IDE interface with four main sections labeled from top-left to bottom-right:

- SOURCE**: A panel for writing and running code, featuring a text editor window titled "Untitled1 x".
- ENVIRONMENT**: A panel showing the global environment, including the "Global Environment" dropdown and a search bar.
- CONSOLE**: A panel for viewing executed code, with tabs for "Console" and "Terminal".
- VIEWER**: A panel for viewing files and plots, showing a file browser for "OneDrive > Temp".

The "Source" panel contains the following text:

Write and run code that you intend to keep.

The "Environment" panel contains the following text:

Global Environment

The "Console" panel contains the following text:

1:1 (Top Level) R Script

Console Terminal

~/

The "Viewer" panel contains the following text:

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > OneDrive > Temp

| Name | Size | Modified |
|------|------|----------|
| .. | | |

The image shows the RStudio interface with four main sections labeled: SOURCE, ENVIRONMENT, CONSOLE, and VIEWER.

SOURCE
Write and run code that you intend to keep.

ENVIRONMENT
View objects, functions, etc.
View command history.

CONSOLE
View executed code. Use sparingly for experimentation.

VIEWER

The RStudio interface includes a top navigation bar with tabs for Environment, History, and Connections. It also features a Global Environment pane, a search bar, and various toolbars for file operations like Import Dataset and Source.

The image shows the RStudio IDE interface with four main sections labeled: SOURCE, ENVIRONMENT, CONSOLE, and VIEWER.

SOURCE
Write and run code that you intend to keep.

ENVIRONMENT
View objects, functions, etc.
View command history.

CONSOLE
View executed code. Use sparingly for experimentation.

VIEWER
View files in working directory, packages, image outputs, etc.





1. ORGANIZE YOUR WORK INTO PROJECTS

(more on this next week)



2. ALWAYS START WITH
A CLEAN SLATE

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

Install Packages...
Check for Package Updates...

Version Control >

Shell...
Terminal
Jobs
Addins

Keyboard Shortcuts Help ⌘K
Modify Keyboard Shortcuts...
Show Command Palette ⌘P

Project Options... ⌘,

Global Options... ⌘,

RStudio

Console Terminal Jobs

~/Documents/academic-website/ ↵

R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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.

Natural language support but running in an English locale

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 History Connections

R Global Environment

Env

Files Plots Packages Help Tutorial Viewer

New Folder Delete Rename More

Home > Documents > academic-website

| | Name | Size | Modified |
|----|----------------|--------|------------------------|
| .. | | | |
| | .gitignore | 53 B | Apr 12, 2020, 9:53 AM |
| | .Rhistory | 2 KB | Jan 14, 2021, 10:44 AM |
| | academic.Rproj | 258 B | Jan 14, 2021, 10:44 AM |
| | assets | | |
| | config | | |
| | content | | |
| | data | | |
| | LICENSE.md | 1.1 KB | Apr 12, 2020, 9:51 AM |
| | public | | |
| | README.md | 3.3 KB | Apr 12, 2020, 9:51 AM |
| | resources | | |

R version 4.0.3 (2020-10-10) -- "Bunny"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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 details.

Natural language support but running

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

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

>

Project: (None)

Console Terminal Jobs

~/Documents/academic-website/

Options General

Code Console Appearance Pane Layout Packages R Markdown Sweave Spelling Git/SVN Publishing Terminal Accessibility Python

Basic Graphics Advanced

R Sessions

Default working directory (when not in a project):
~/Documents/academic-website

Restore most recently opened project at startup

Restore previously open source documents at startup

Workspace

Restore .RData into workspace at startup

Save workspace to .RData on exit:

History

Always save history (even when not saving .RData)

Remove duplicate entries in history

Other

Wrap around when navigating to previous/next tab

Automatically notify me of updates to RStudio

Send automated crash reports to RStudio

OK Cancel Apply

resources

| Size | Modified |
|--------|------------------------|
| 53 B | Apr 12, 2020, 9:53 AM |
| 2 KB | Jan 14, 2021, 10:44 AM |
| 258 B | Jan 14, 2021, 10:44 AM |
| 1.1 KB | Apr 12, 2020, 9:51 AM |
| 3.3 KB | Apr 12, 2020, 9:51 AM |



3. TWO WAYS OF INTERACTING WITH R: CONSOLE AND SCRIPTS

(most of the time, use scripts)

RStudio

Project: (None)

Untitled1*

```
1 2 > 3
2
3 2 + 2 == 4
4
5 isTRUE(2 + 2 == 5)
6
```

Source on Save | Run | Source | Environment | History | Connections | Import Dataset | Global Environment | List | Search

Environment is empty

6:1 (Top Level) R Script

Console Terminal Jobs

```
~/Documents/academic-website/
> 2 > 3
[1] FALSE
> 2 + 2 == 4
[1] TRUE
> isTRUE(2 + 2 == 5)
[1] FALSE
>
```

Files Plots Packages Help Tutorial Viewer

New Folder Delete Rename More

Home > Documents > academic-website

| | Name | Size | Modified |
|----|----------------|--------|------------------------|
| .. | | | |
| | .gitignore | 53 B | Apr 12, 2020, 9:53 AM |
| | .Rhistory | 2.2 KB | Jan 19, 2021, 12:51 PM |
| | academic.Rproj | 258 B | Jan 14, 2021, 10:44 AM |
| | assets | | |
| | config | | |
| | content | | |
| | data | | |
| | LICENSE.md | 1.1 KB | Apr 12, 2020, 9:51 AM |
| | public | | |
| | README.md | 3.3 KB | Apr 12, 2020, 9:51 AM |
| | resources | | |



4. BECOME COMFORTABLE WITH ERROR MESSAGES



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.

```
> my_data <- read.csv("my_data.csv")
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file 'my_data.csv': No such file or directory
> Can you say that in plain English please?
Error: unexpected symbol in "Can you"
> :(
Error: unexpected ':' in ":"
```

> |



5. USE R STUDIO'S TOOLS
TO MAKE LIFE EASIER

YOUR TURN

Please complete the activity
for this week listed on the
course website under:

Schedule > Assignments

It's short!

