



BIOINFORMATICS

Hands-on Lab Session

Class 04

Barry Grant
UC San Diego

<http://thegrantlab.org/teaching>



What is R?

R is a freely distributed and widely used programing **language** and **environment** for statistical computing, data analysis and graphics.



What is R?

R is a freely distributed and widely used programing **language** and **environment** for statistical computing, data analysis and graphics.

R provides an unparalleled interactive environment for data analysis.



What is R?

R is a freely distributed and widely used programming **language** and **environment** for statistical computing, data analysis and graphics.



R provides an unparalleled interactive environment for data analysis.

It is script-based (*i.e.* driven by computer code) and not GUI-based (point and click with menus).



zico:~> R

R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 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.

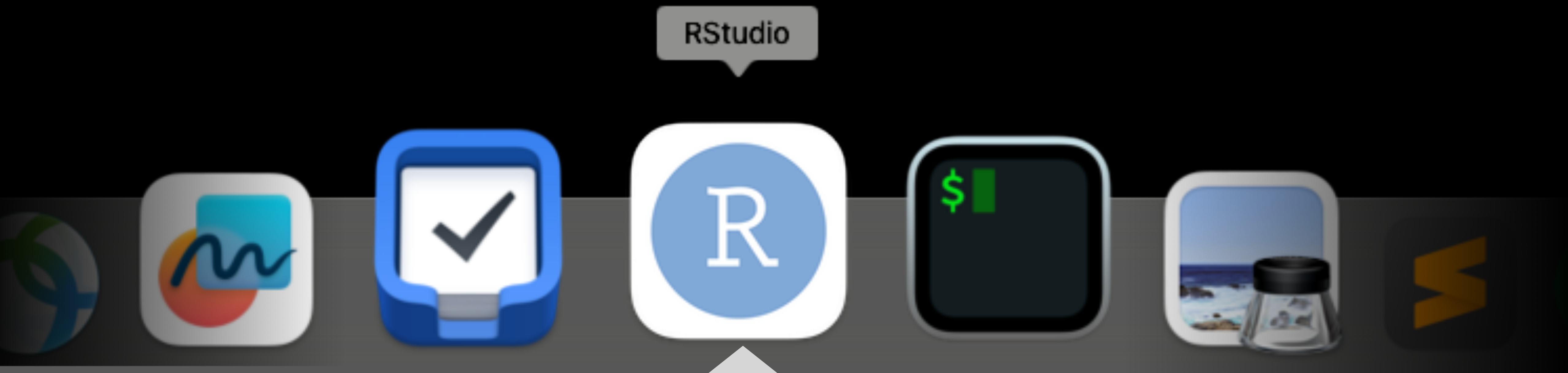
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> █

Getting Started

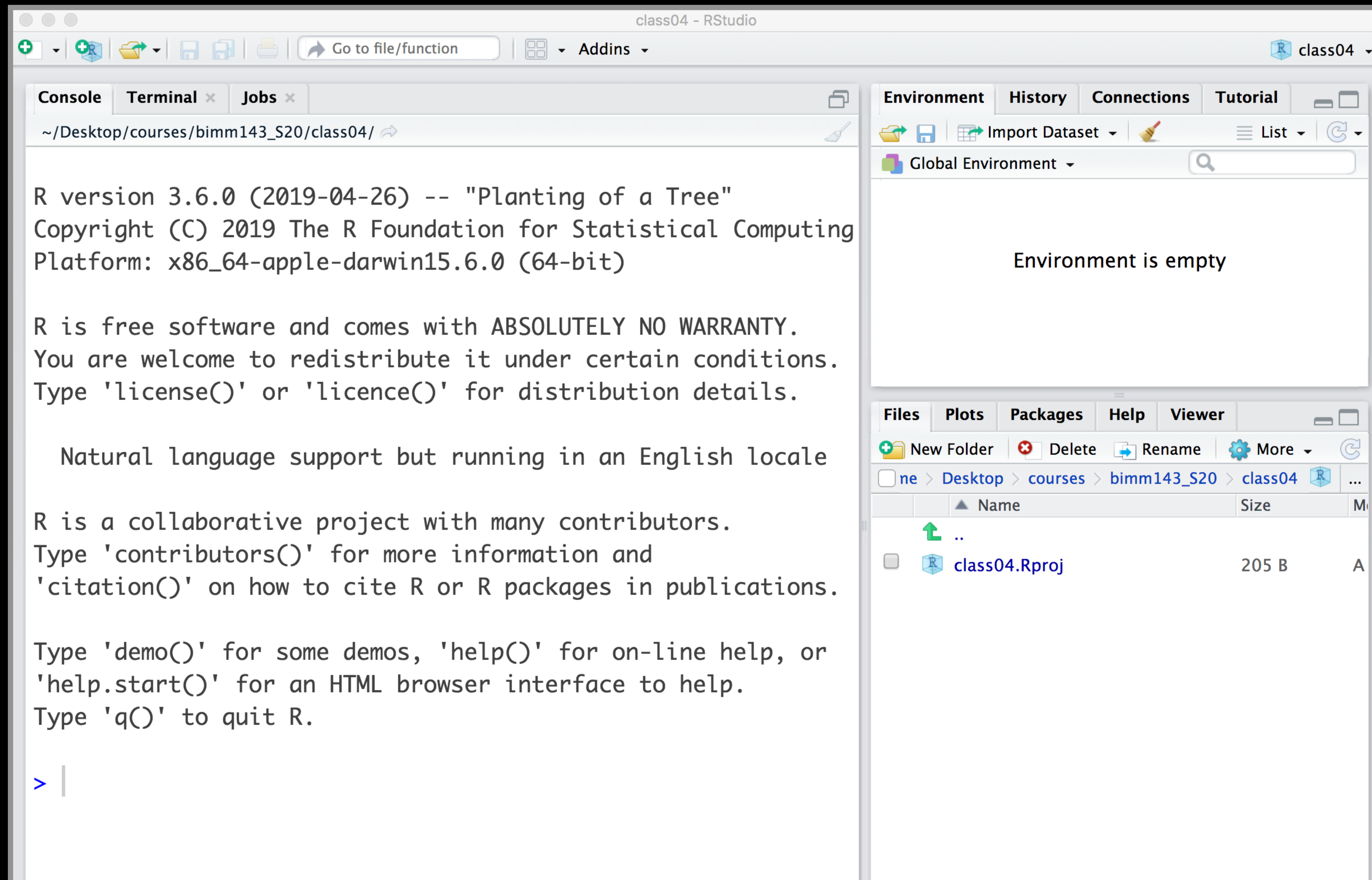
Open RStudio Now!

- Open **RStudio** as you would any program. It's icon looks like this:

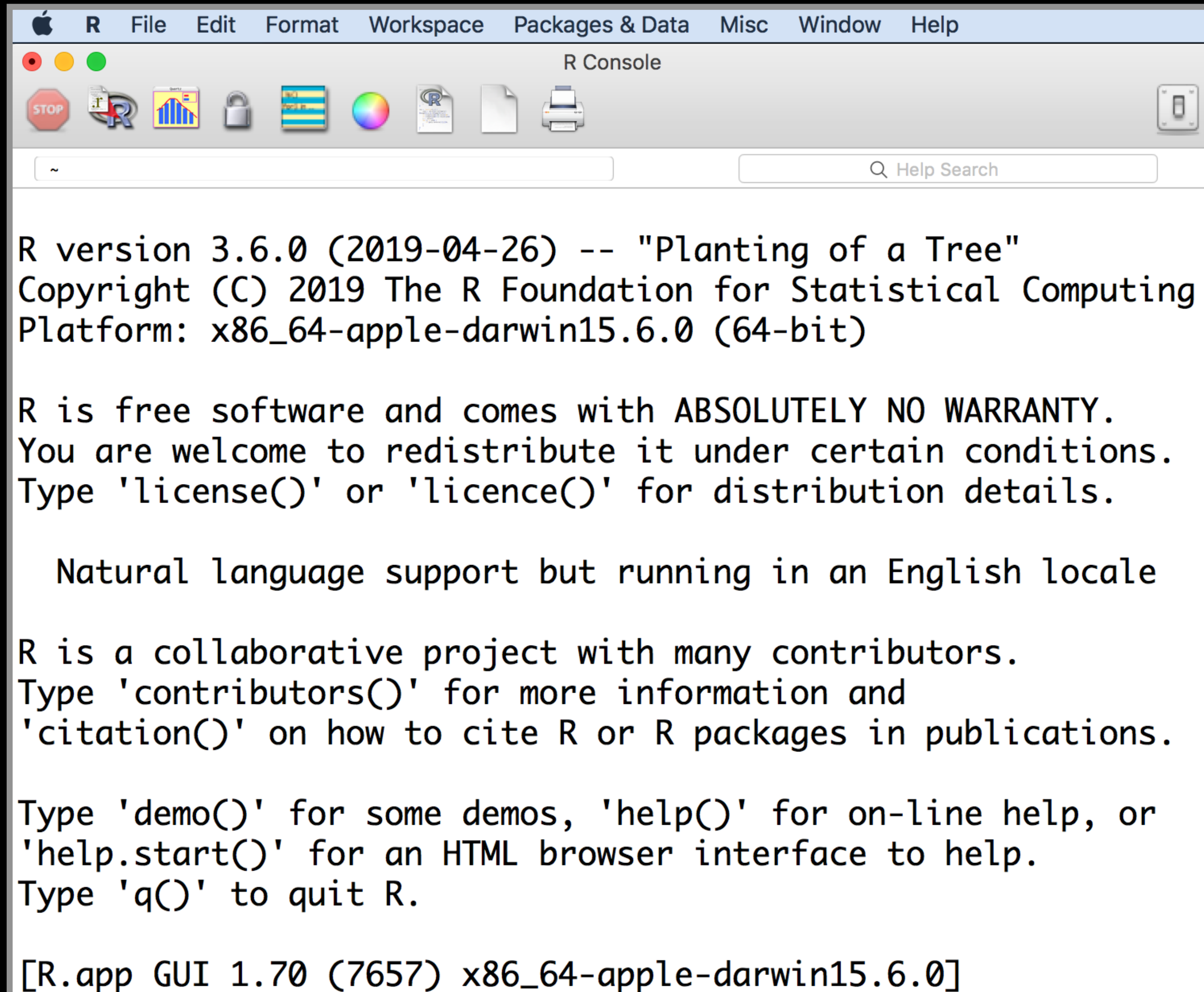


RStudio is an IDE for R!

Open RStudio Now!



R.app GUI is NOT what we want!

A screenshot of the R Console window on a Mac. The window has a menu bar with 'R', 'File', 'Edit', 'Format', 'Workspace', 'Packages & Data', 'Misc', 'Window', and 'Help'. Below the menu bar is a toolbar with icons for stopping, running, saving, and other functions. The main area of the window displays the R version 3.6.0 (2019-04-26) -- "Planting of a Tree" and copyright information. It also shows the license text: '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.' and instructions for using 'demo()', 'help()', 'help.start()', and 'q()'. At the bottom, it shows the command '[R.app GUI 1.70 (7657) x86_64-apple-darwin15.6.0]'.

```
R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

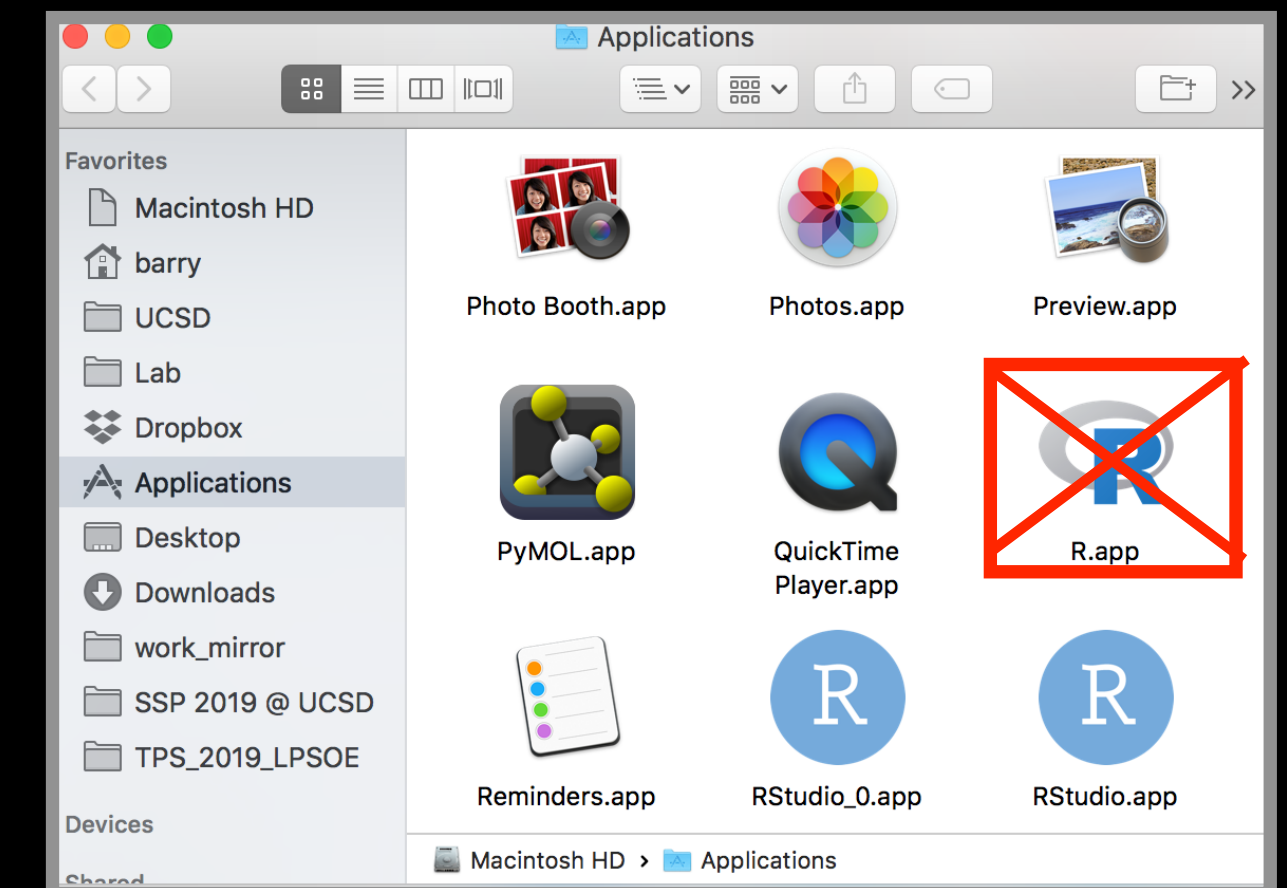
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

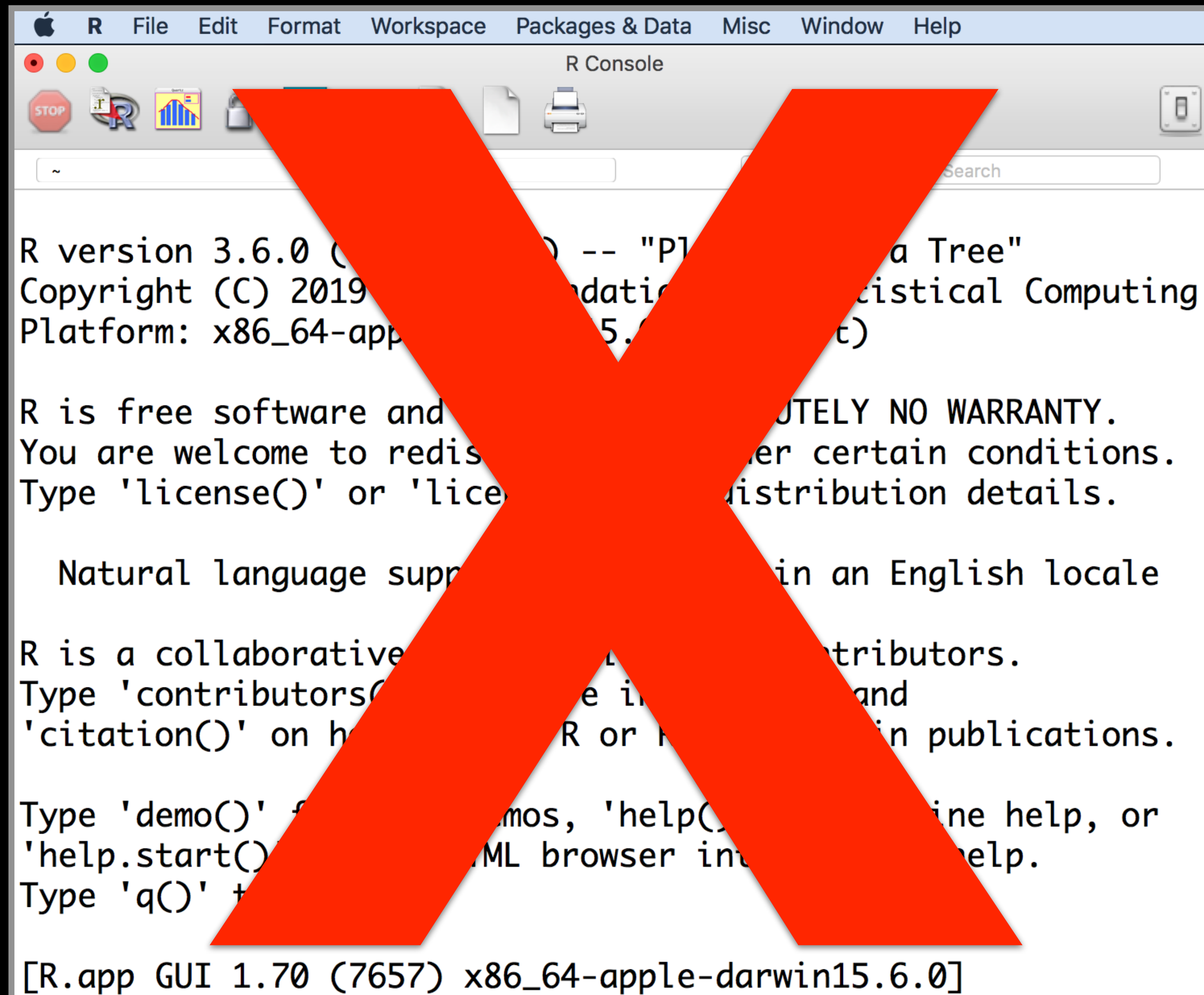
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

[R.app GUI 1.70 (7657) x86_64-apple-darwin15.6.0]
```



R.app GUI is NOT what we want!



R Console

```
R version 3.6.0 (2019-10-11) -- "Pleasant Plums"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0

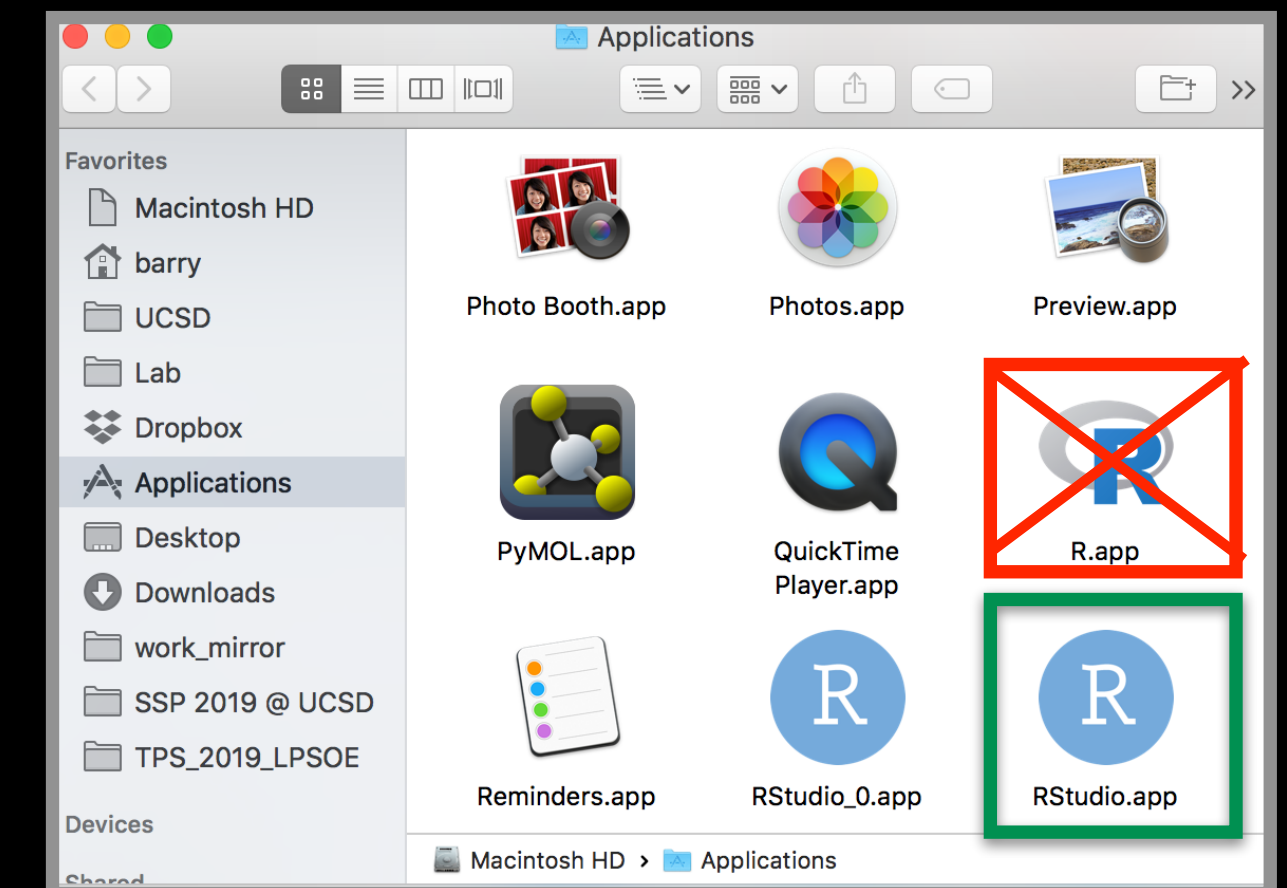
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

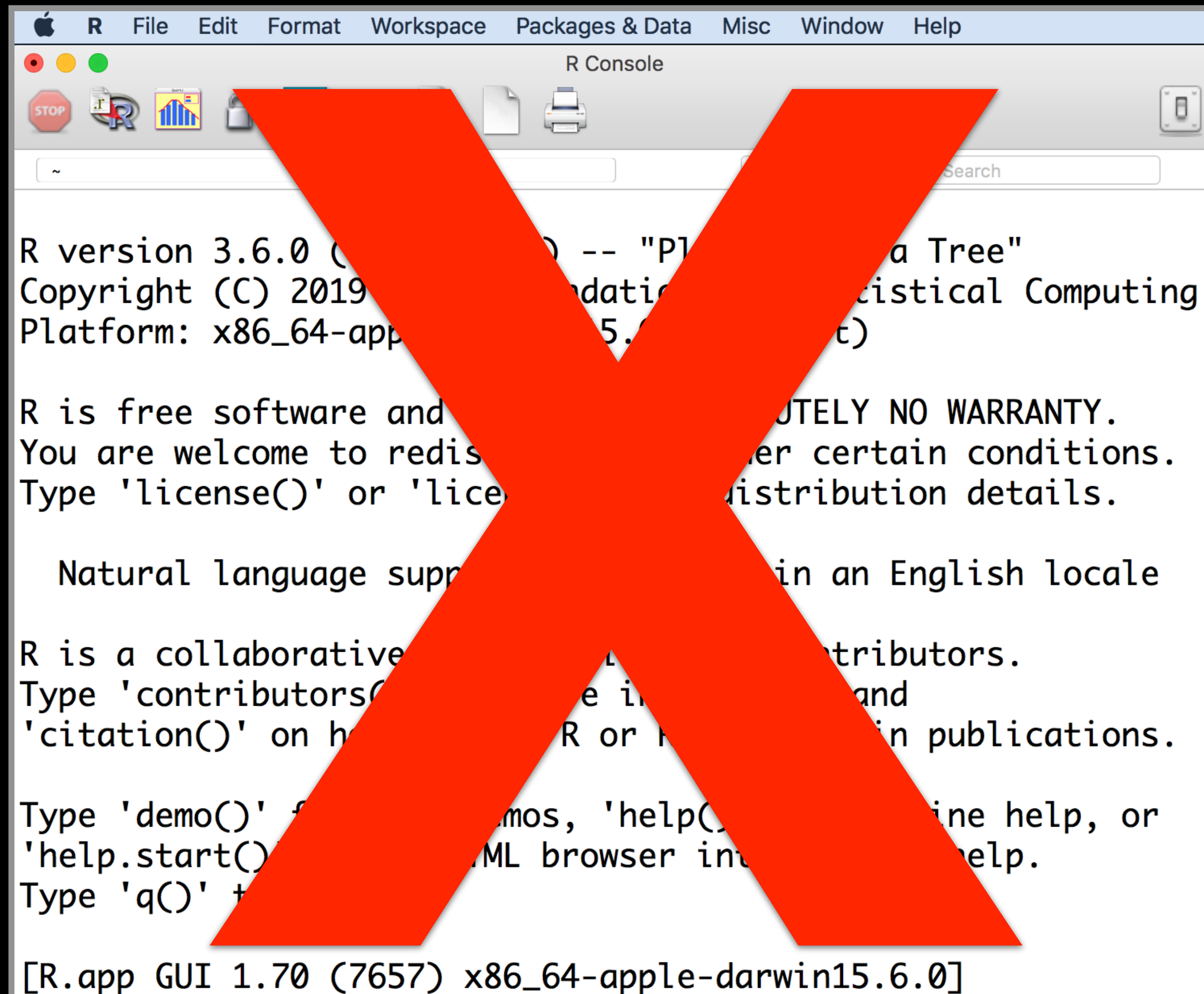
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

[R.app GUI 1.70 (7657) x86_64-apple-darwin15.6.0]
```

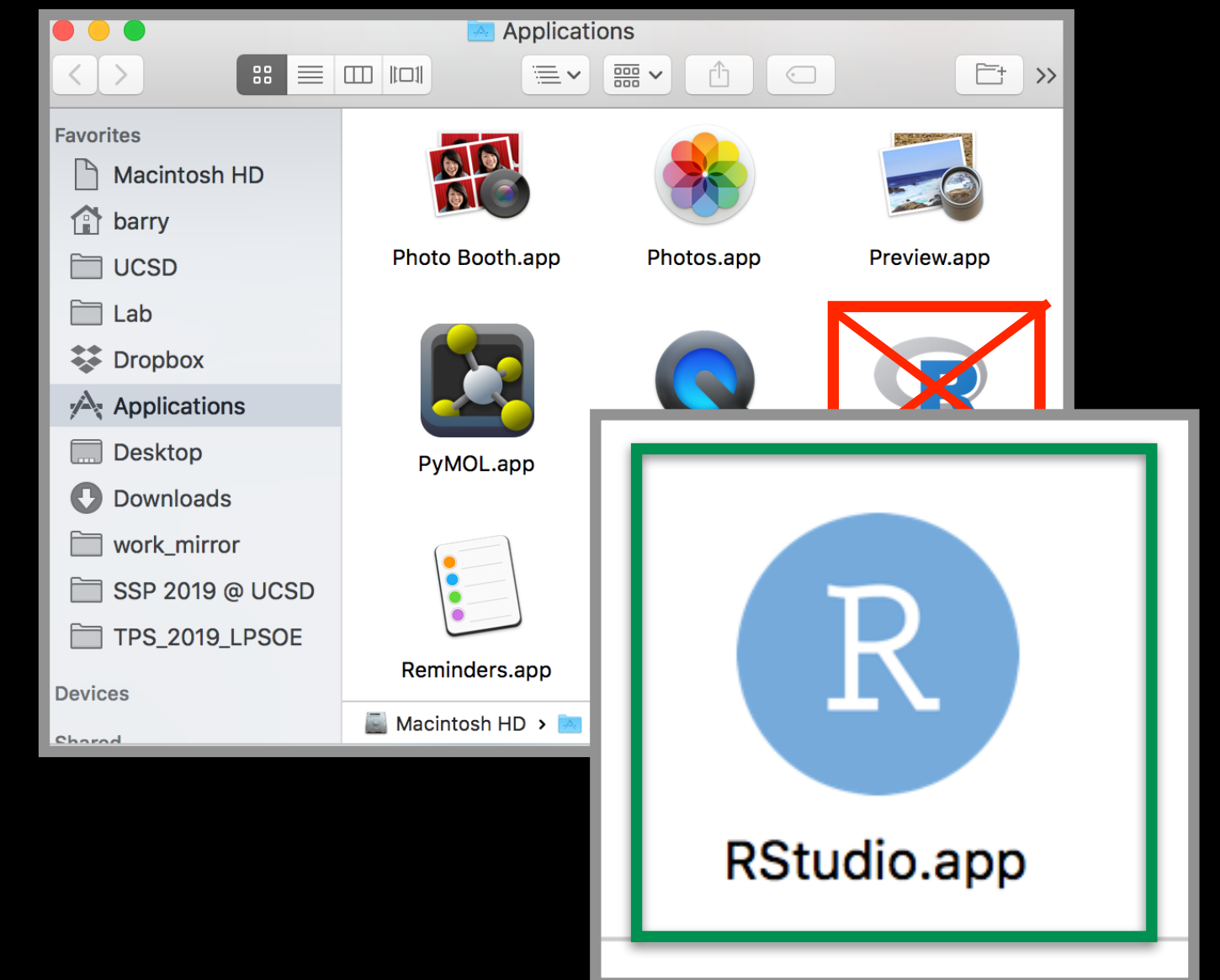


R.app GUI is NOT what we want!

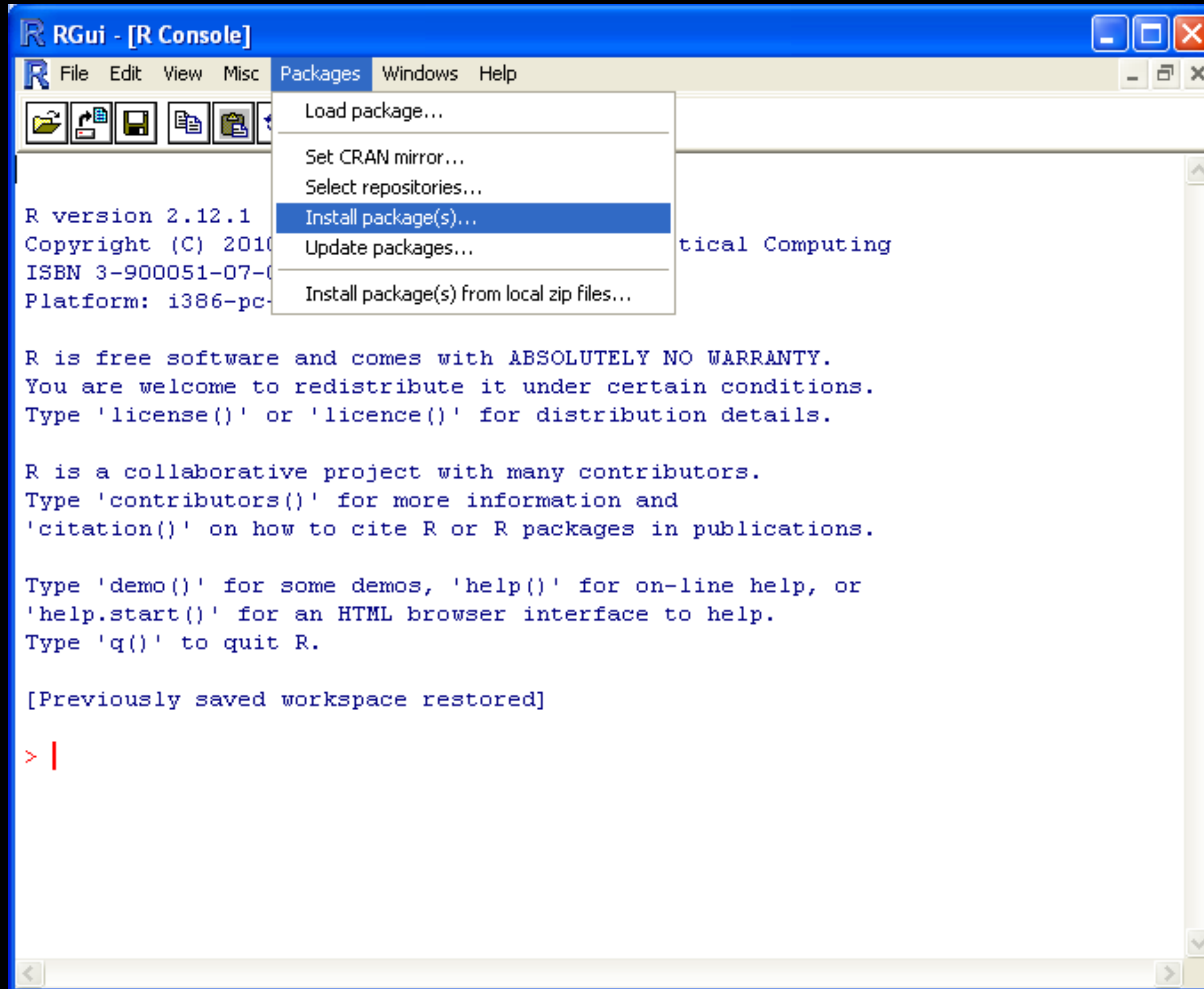


The screenshot shows the R Console window with a large red X overlaid on it. The console text reads:

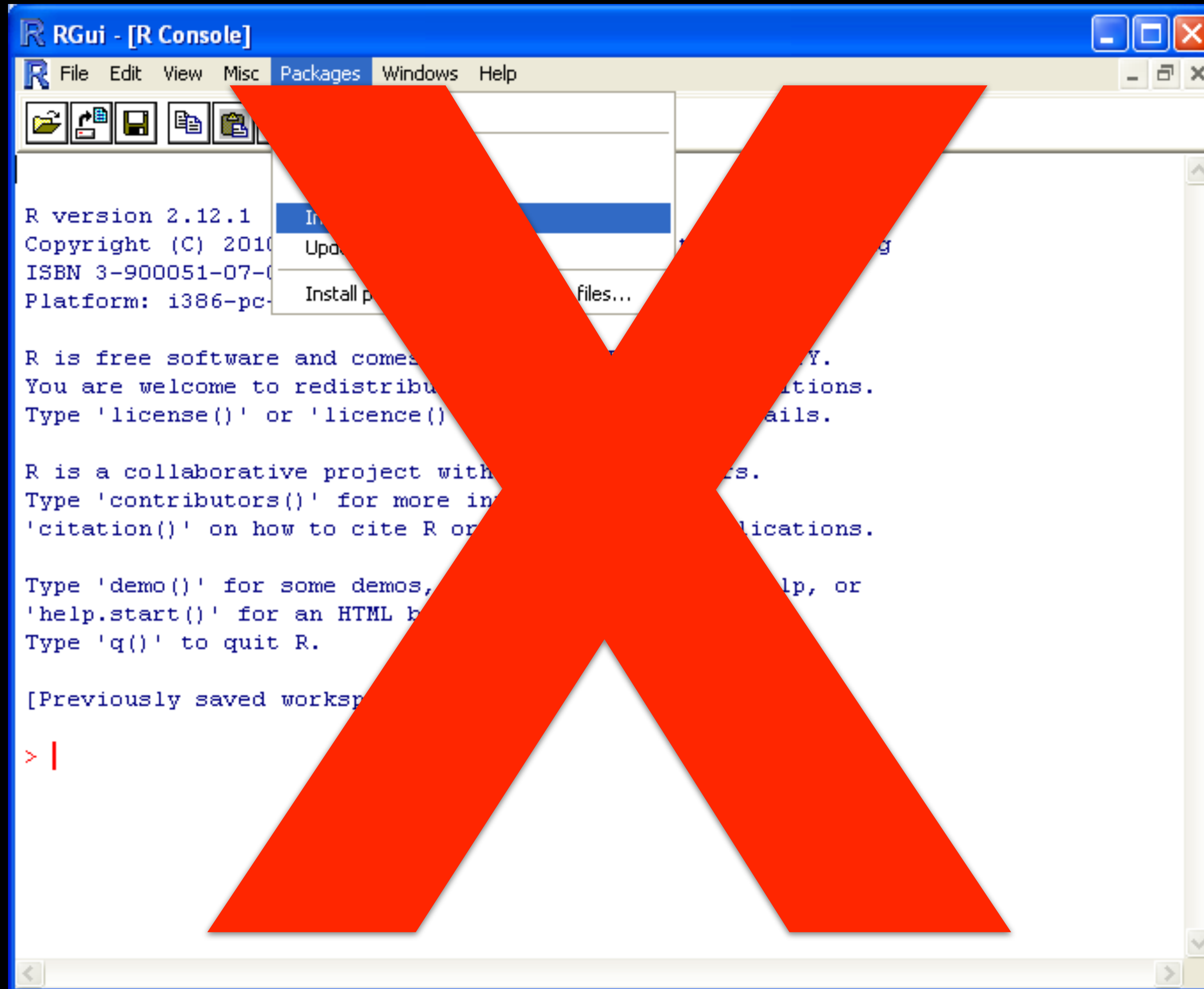
```
R version 3.6.0 (2019-10-10) -- "Pleasant Plums and a Tree"  
Copyright (C) 2019 The R Foundation for Statistical Computing  
Platform: x86_64-apple-darwin15.6.0 (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
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.  
  
[R.app GUI 1.70 (7657) x86_64-apple-darwin15.6.0]
```



RGui is NOT what we want!

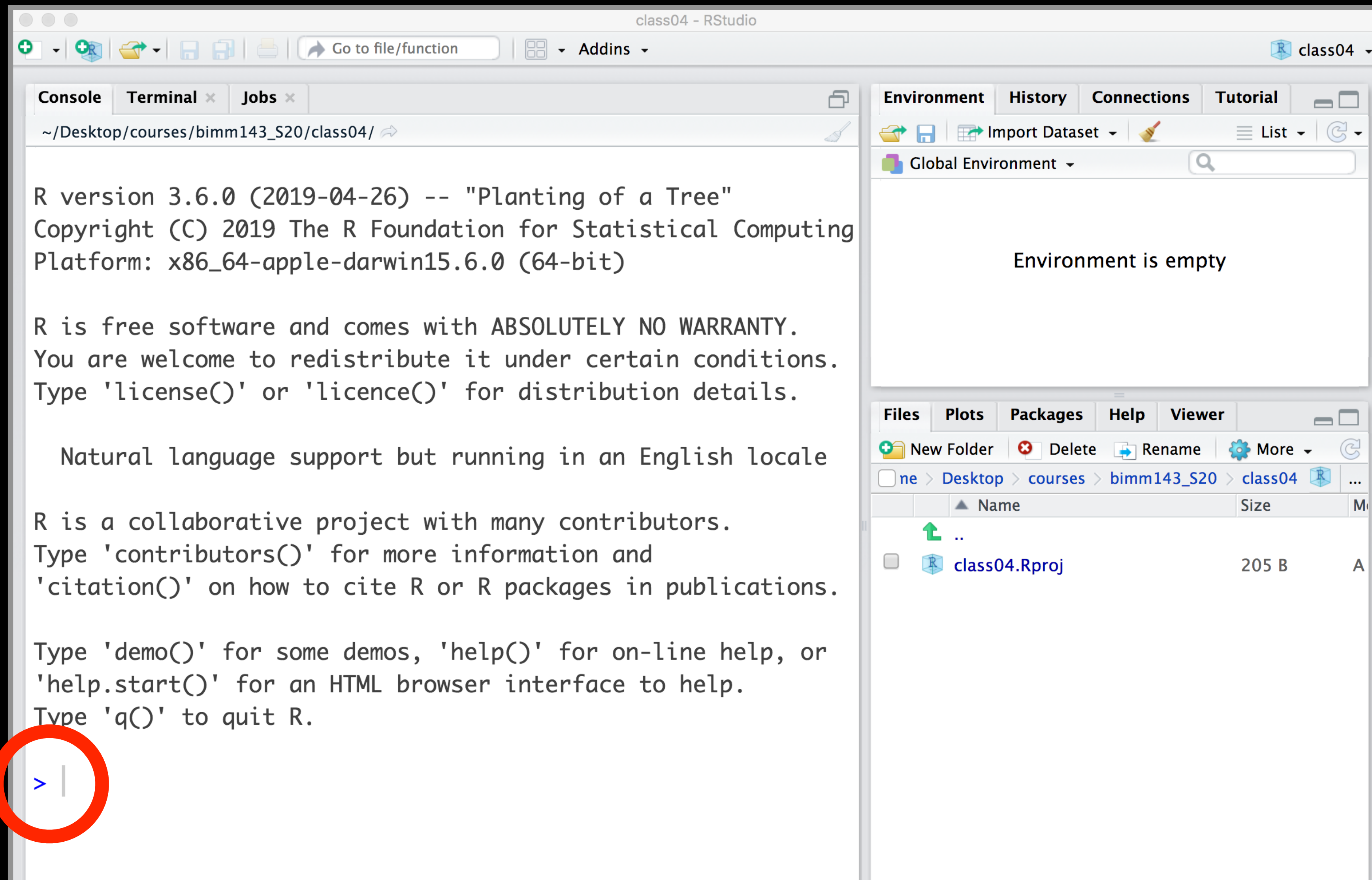


RGui is NOT what we want!



We want RStudio a separate IDE for R!

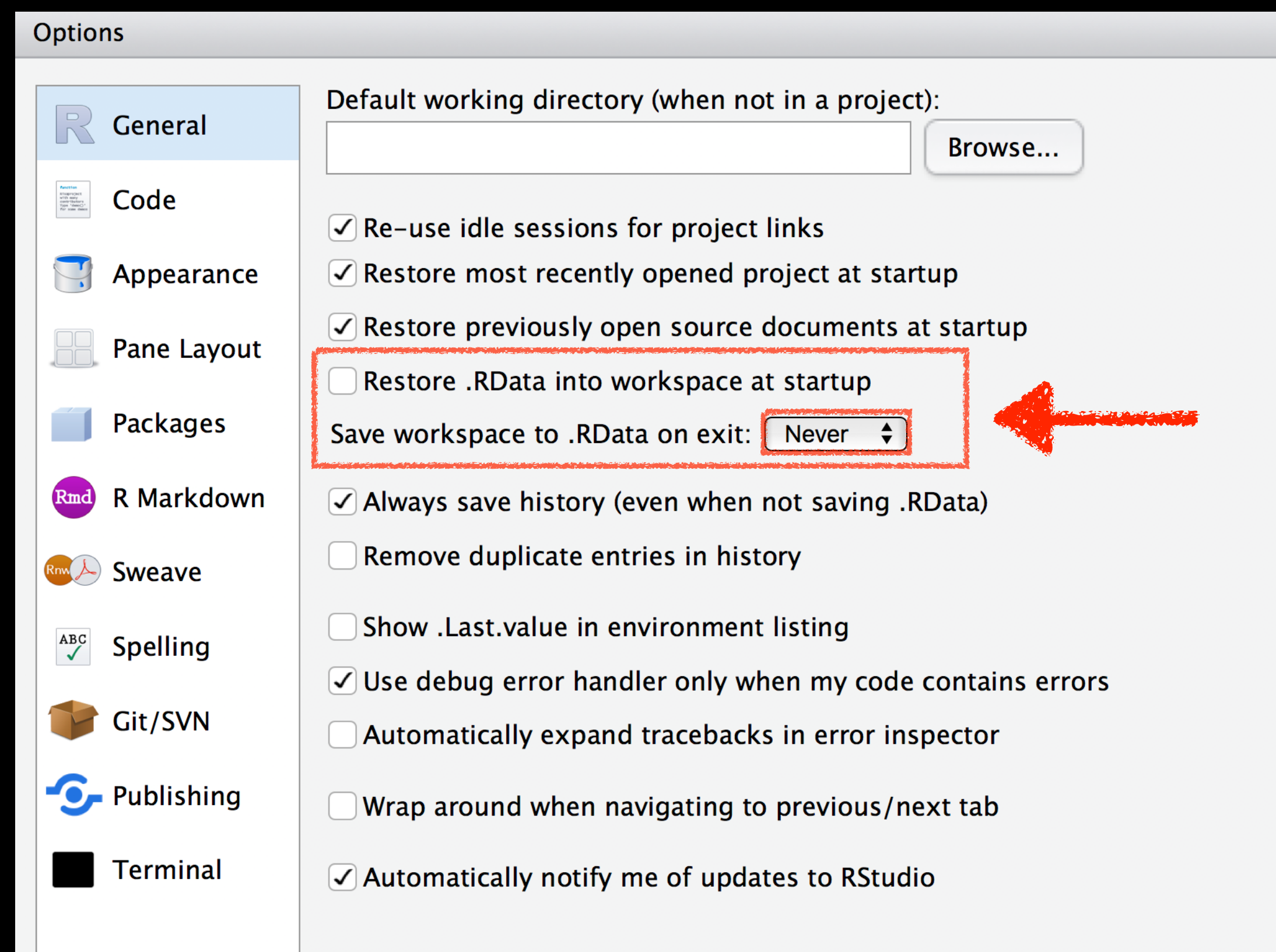
Open RStudio Now!



Tools > Global Options

Do it Now!

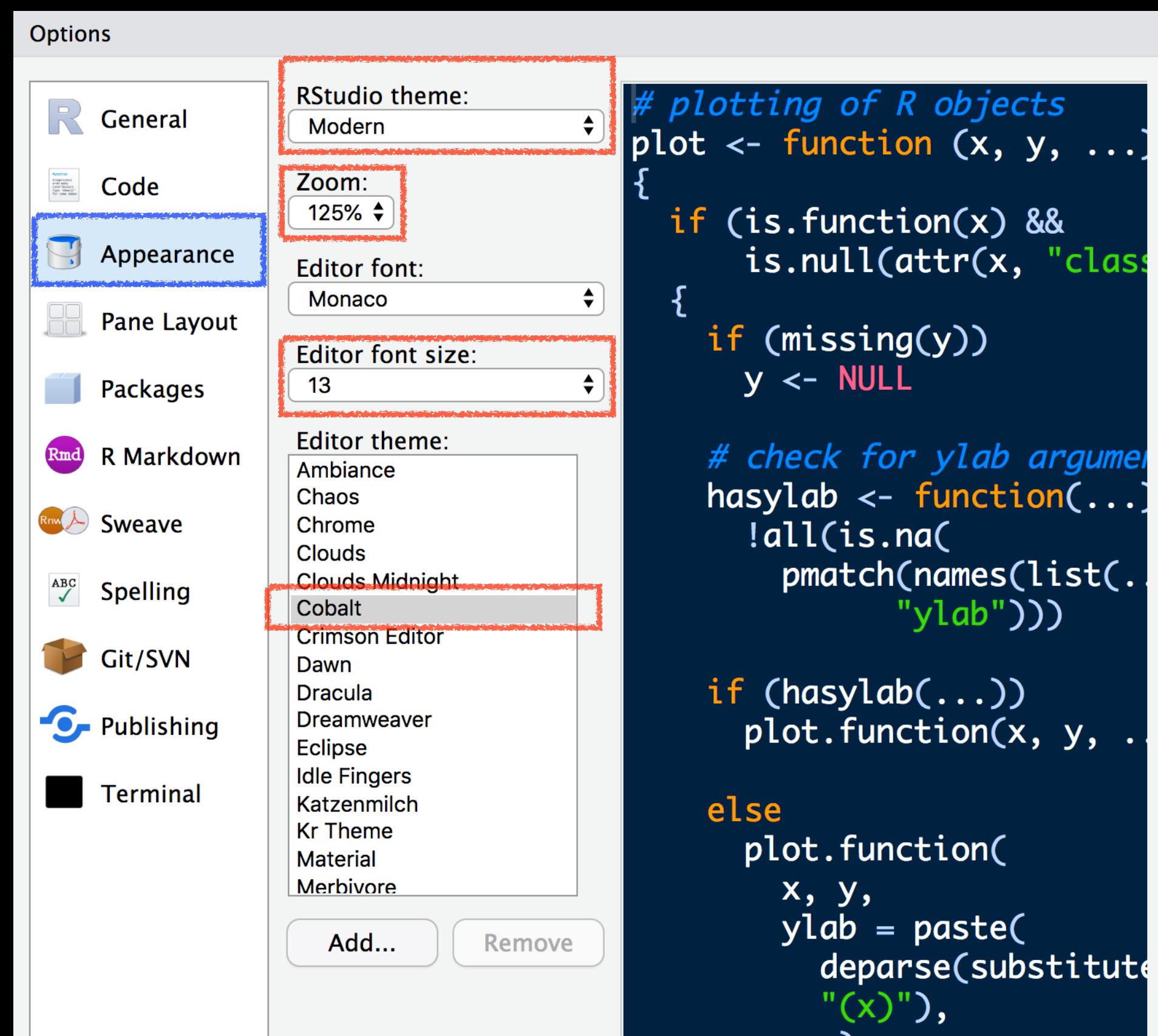
- We will **NOT** save our workspace...



Tools > Global Options

Optional!

- [Optional] Change to dark Appearance and increase font size, etc. ...



R Studio®

Follow Along!

Source

Console Terminal x

R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
Copyright (C) 2019 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.

```
> library(ggplot2)
> ggplot(mpg, aes(displ, hwy, color=class)) +
+ geom_point()
> |
```

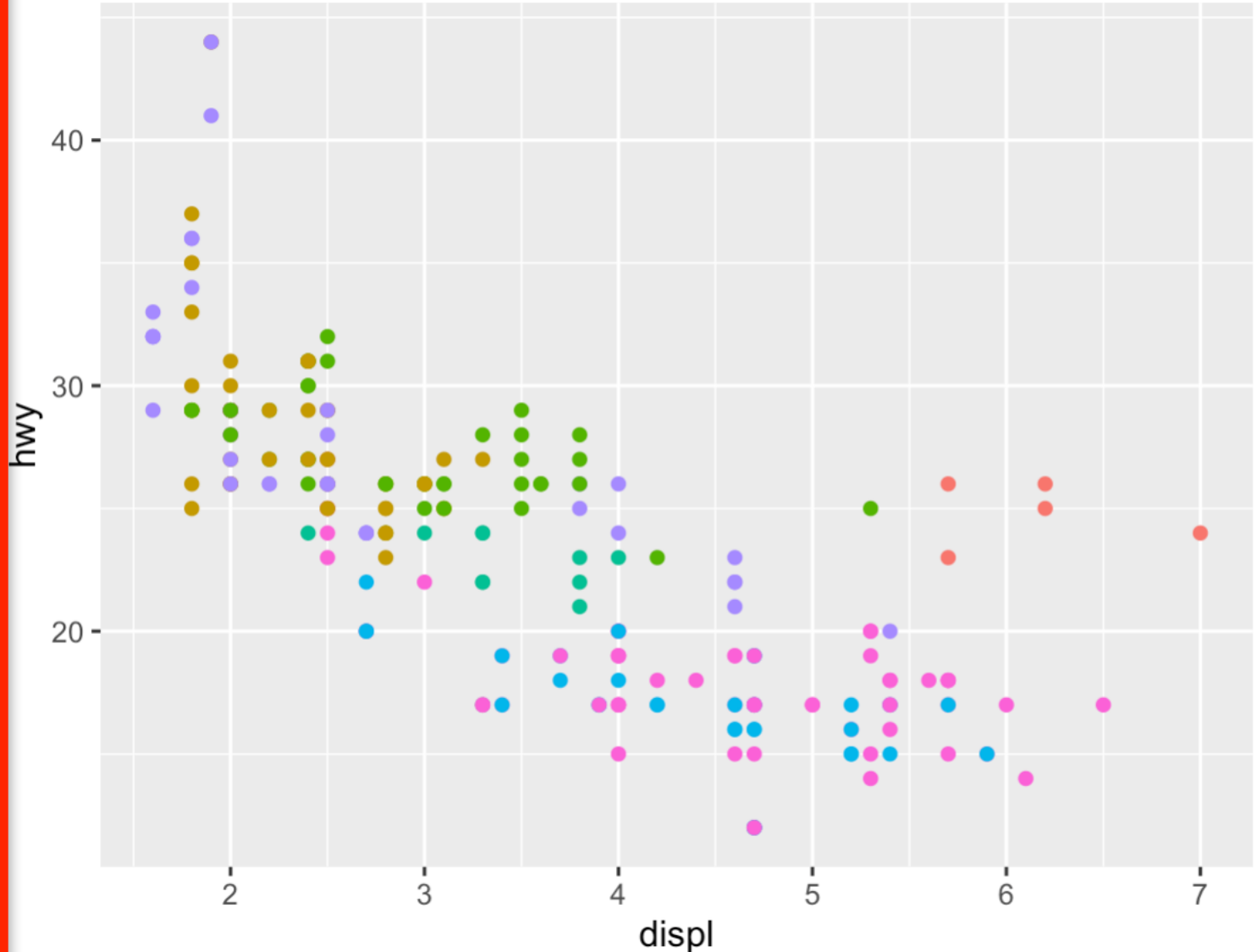
Environment History Connections Tutorial

Global Environment

Environment is empty

Files Plots Packages Help Viewer

Zoom Export



A scatter plot showing highway mileage (hwy) on the y-axis (ranging from 20 to 40) versus engine displacement (displ) on the x-axis (ranging from 2 to 7). The data points are colored by car class: 2seater (red), compact (yellow), midsize (green), minivan (teal), pickup (blue), subcompact (purple), and suv (pink). The plot shows a general negative correlation between engine displacement and highway mileage, with different car classes clustered at various points along this trend.

class

- 2seater
- compact
- midsize
- minivan
- pickup
- subcompact
- suv

Hands-on Topics (R Basics):

Vectors and vectorization

Data frames

Calling Functions

Getting help in R

Workspace and RStudio projects

Hands-on Topics (R Basics):

Vectors and vectorization

Data frames

Calling Functions

Getting help in R

Workspace and RStudio projects

R has many data structures

These include:

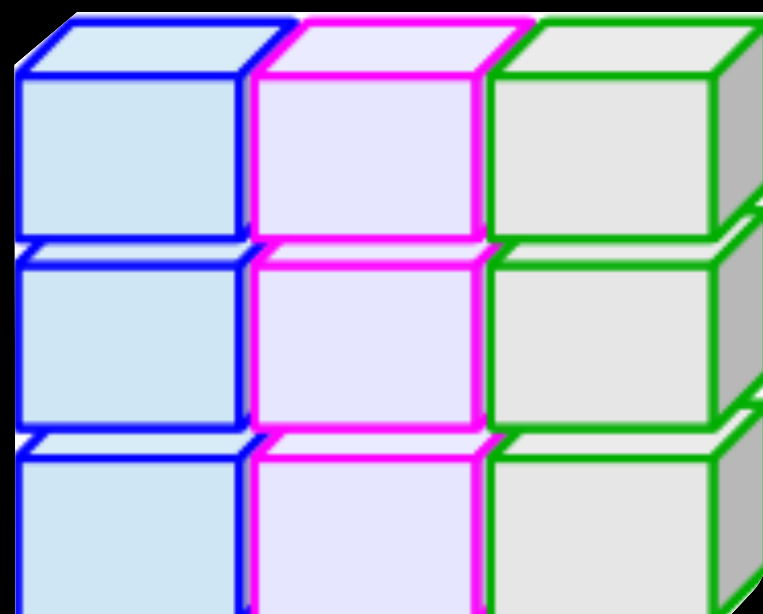
- vector
- data frame
- list
- matrix
- arrays

R has many data structures

1D all elements of the **same "type"**



2D column elements can be **different "types"**



These include:

- **vector**
- **data frame**
- list
- matrix
- arrays

RStudio >

Set appearance and demo **vectors** and **data.frames**

Tips for learning

- Ask yourself questions.
- Work with frustration.
- **Practice!**
 - ➔ And practice consciously:
 - ➔ Make a **prediction**, then **test it**, then **reflect**...

bioboot.github.io

Home

Gmail

GDrive

Gcal

Blink

GitHub

BIMM143

BGGN213

BIMM194

BGGN239

+ v

40+

31

UC San Diego

BIMM 143

A hands-on introduction to the computer-based analysis of genomic and biomolecular data from the School of Biological Sciences, UCSD [↗](#).

Overview

Schedule

Computer Setup

Learning Goals

Assignments & Grading

Ethics Code

Videos:

- 4.1 [Why R and RStudio](#) [↗](#),
- 4.2 [Major R data structures, data types, and using functions](#) [↗](#),
- 4.3 [Working with DataCamp](#) [↗](#) **N.B.** Use your UCSD email invite to sign up and visit our class group/organization.

Supporting Material:

- Lecture Slides: [Large PDF](#) [↗](#), [Small PDF](#) [↗](#),
- Cheat Sheet: [Base R overview](#) [↗](#),
- Lab: [Hands-on section](#) [↗](#),
- Lab: [Video walk-through](#) [↗](#) focusing on introducing R data structures and core syntax,
- Extra credit lab: [Introduction to data in R](#) [↗](#),
- Optional extension: [Advanced conservation analysis of globins with R](#) [↗](#) this demonstrates where we are going on our R learning journey. You should be able to do analysis like this on your own at the end of the course.
- Student Hours: [After class on Thur @ 12:10pm](#) [↗](#),
- Feedback: [Muddy point assessment](#) [↗](#),

Homework:

- Answer the `04_HW_quiz` questions on [GradeScope](#) [↗](#),
- Submit your rendered `04_Lab_report` (i.e. a PDF report generated from R) to [GradeScope](#) [↗](#),
- DataCamp:** Sign-up to our **25_bioinformatic group/organization** via the link in your UCSD email and **complete [Introduction to R!](#)** [↗](#) (4hrs).

40+

31

UC San Diego

BIMM 143

A hands-on introduction to the computer-based analysis of genomic and biomolecular data from the School of Biological Sciences, UCSD 

Overview

Schedule

Computer Setup

Learning Goals

Assignments & Grading

Ethics Code



bioboot.github.io

Home Gmail GDrive Gcal Blink GitHub BIMM143 BGGN213 BIMM194 BGGN239 +

Schedule · BIMM 143

Videos:

- 4.1 [Why R and RStudio](#) ,
- 4.2 [Major R data structures, data types, and using functions](#) ,
- 4.3 [Working with DataCamp](#)  **N.B.** Use your UCSD email invite to sign up and visit our class group/organization.

Supporting Material:

- Lecture Slides: [Large PDF](#) , [Small PDF](#) ,
- Cheat Sheet: [Base R overview](#) ,
- Lab: [Hands-on section](#) ,
- Lab: [Video walk-through](#)  focusing on introducing R data structures and core syntax,
- Extra credit lab: [Introduction to data in R](#) ,
- Optional extension: [Advanced conservation analysis of globins with R](#)  this demonstrates where we are going on our R learning journey. You should be able to do analysis like this on your own at the end of the course.
- Student Hours: [After class on Thur @ 12:10pm](#) ,
- Feedback: [Muddy point assessment](#) ,

Homework:

- Answer the `04_HW_quiz` questions on [GradeScope](#) ,
- Submit your rendered `04_Lab_report` (i.e. a PDF report generated from R) to [GradeScope](#) ,
- DataCamp:** Sign-up to our **25_bioinformatic group/organization** via the link in your UCSD email and **complete Introduction to R!**  (4hrs).

40+

31

UC San Diego

BIMM 143

A hands-on introduction to the computer-based analysis of genomic and biomolecular data from the School of Biological Sciences, UCSD [↗](#).

Overview

Schedule

Computer Setup

Learning Goals

Assignments & Grading

Ethics Code

bioboot.github.io

Home Gmail GDrive Gcal Blink GitHub BIMM143 BGGN213 BIMM194 BGGN239 +

Schedule · BIMM 143

Videos:

- 4.1 [Why R and RStudio](#) [↗](#),
- 4.2 [Major R data structures, data types, and using functions](#) [↗](#),
- 4.3 [Working with DataCamp](#) [↗](#) **N.B.** Use your UCSD email invite to sign up and visit our class group/organization.

Supporting Material:

- Lecture Slides: [Large PDF](#) [↗](#), [Small PDF](#) [↗](#),
- Cheat Sheet: [Base R overview](#) [↗](#),
- Lab: [Hands-on section](#) [↗](#),
- Lab: [Video walk-through](#) [↗](#) focusing on introducing R data structures and core syntax,
- Extra credit lab: [Introduction to data in R](#) [↗](#),
- Optional extension: [Advanced conservation analysis of globins with R](#) [↗](#) this demonstrates where we are going on our R learning journey. You should be able to do analysis like this on your own at the end of the course.
- Student Hours: [After class on Thur @ 12:10pm](#) [↗](#),
- Feedback: [Muddy point assessment](#) [↗](#),

Homework:

- Answer the `04_HW_quiz` questions on [GradeScope](#) [↗](#),
- Submit your rendered `04_Lab_report` (i.e. a PDF report generated from R) to [GradeScope](#) [↗](#),
- DataCamp:** Sign-up to our **25_bioinformatic group/organization** via the link in your UCSD email and **complete Introduction to R!** [↗](#) (4hrs).

R Language Basics

(Part 1)

Barry Grant
UC San Diego

Like a presentation
Push the Down arrow to advance



Checking your understanding

We will use these types of “code chunks” to ask you to complete some missing code, typically indicated by the blank _____ place holder.

For example, complete the code below so your answer sums to 10:

Exercise fill in the blank

↺ Start Over

💡 Show Hint

▶ Run Code

1 5 + 3 + _____

Hands-on Topics:

Vectors and vectorization

Data frames

Calling Functions

Getting help in R

Workspace and RStudio projects

RStudio >

Demo with `log()` and `plot()` functions

Hands-on Topics:

Vectors and vectorization

Data frames

Calling Functions

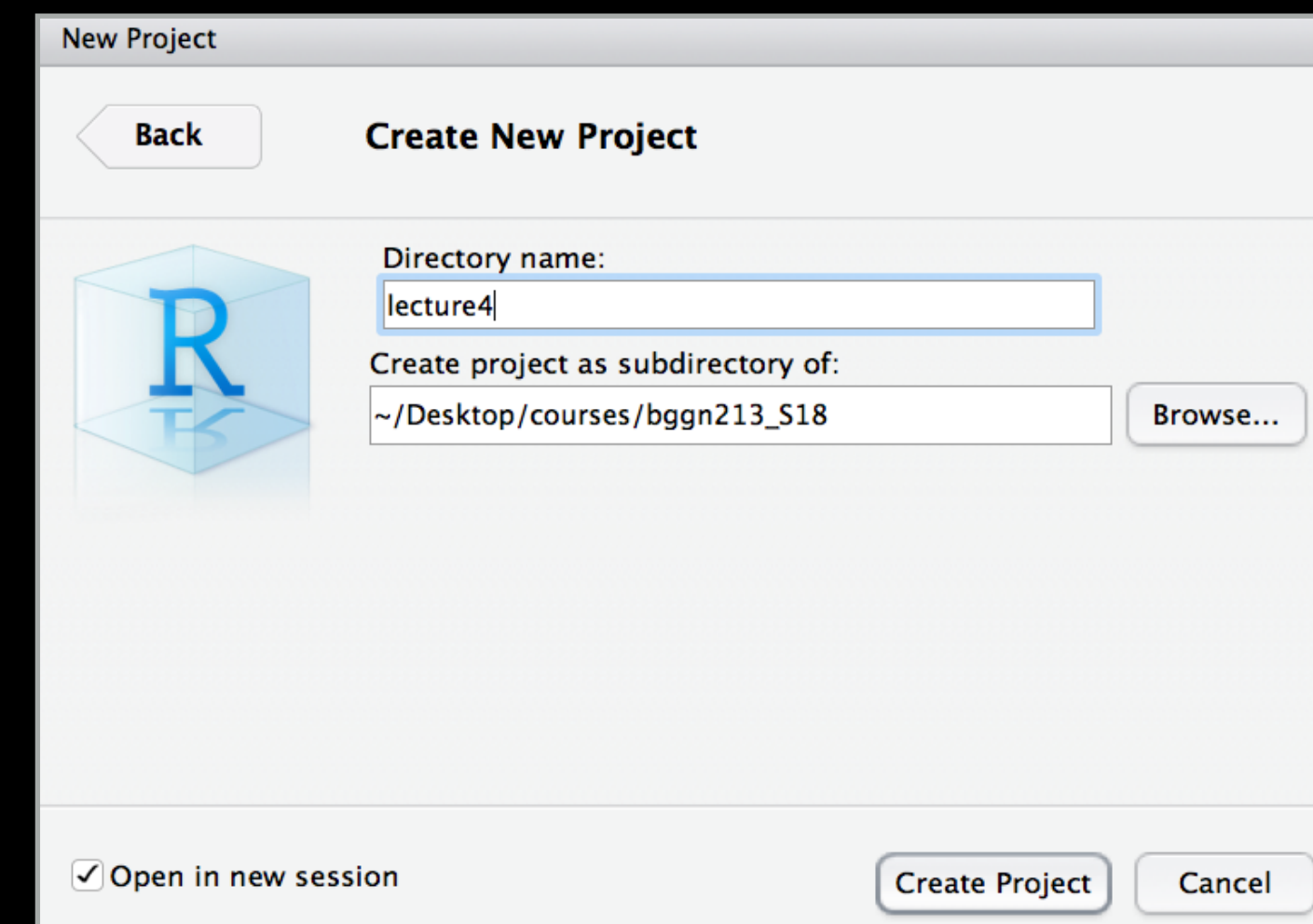
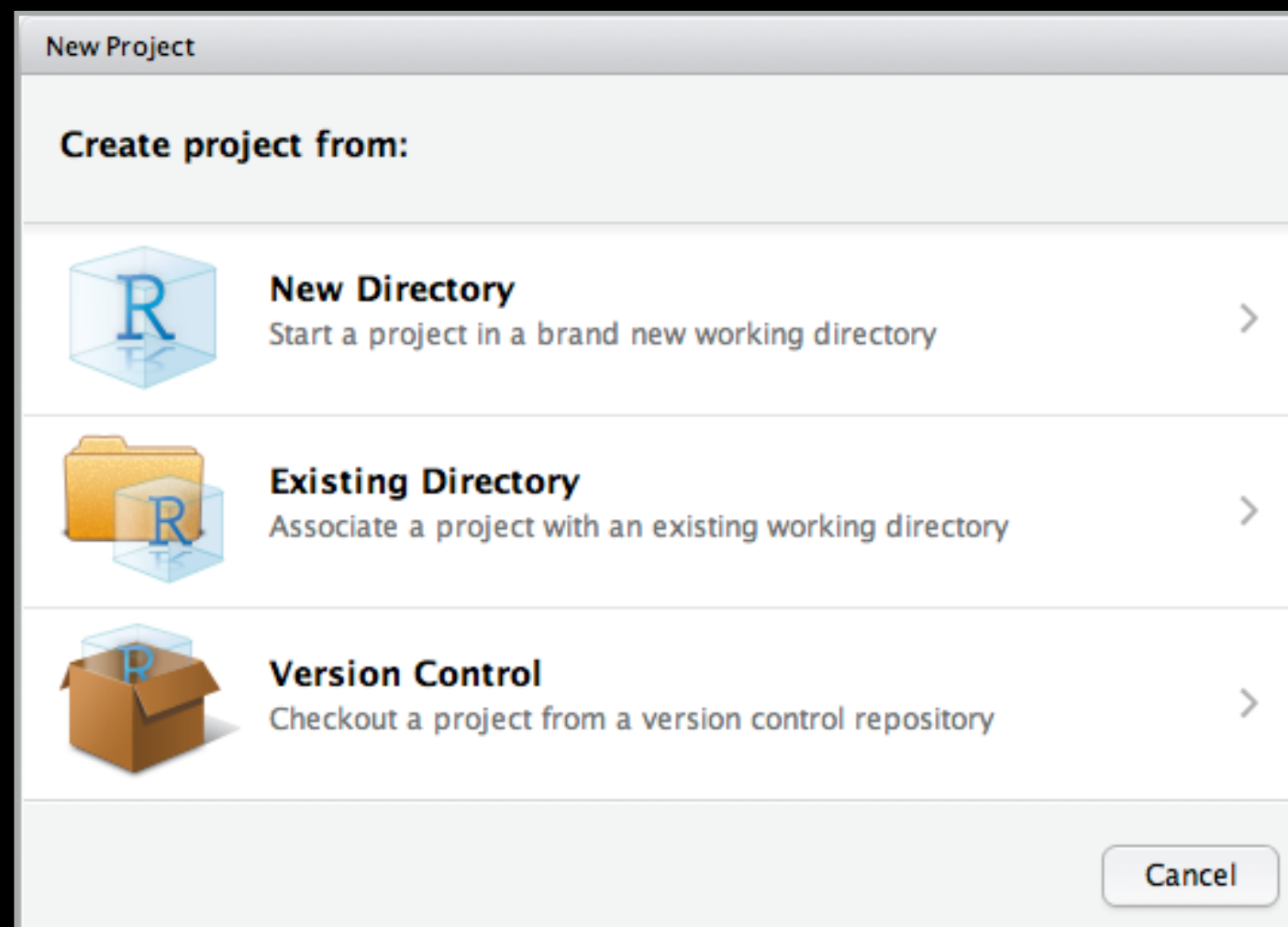
Getting help in R

Workspace and **RStudio projects**

RStudio Projects

- We will use a new RStudio **project** for each new class going forward.

File > New Project > New Directory > New Project...



- **Projects** will help keep us **organized**:
 - Store our work into separate directories (folders) with their own workspace, history, and source documents.

RStudio >

Create a **BIMM143** working folder and create a **class04** New Project
(create and save a simple Rscript...)

Terminal

```
quarto install tinytex
```

Homework: DataCamp Courses!

Complete the **Introduction to R** course on **DataCamp**
(Check Piazza for your DataCamp invite and sign up with your
UCSD email (i.e. first part of your email address) please.

Let me know **NOW** if you don't have access to DataCamp!

DataCamp

campus.datacamp.com/courses/free-introduction-to-r/chapter-1-intro-to-basics-1?ex=1

Home Gmail Gcal GitHub BIMM143 BGGN213 GDrive Atmosphere CloudLaunch BIMM194 Blink News +

Schedule · BIMM 143

How it works | R

DataCamp

Exercise

How it works

In the editor on the right you should type R code to solve the exercises. When you hit the 'Submit Answer' button, every line of code is interpreted and executed by R and you get a message whether or not your code was correct. The output of your R code is shown in the console in the lower right corner.

R makes use of the `#` sign to add comments, so that you and others can understand what the R code is about. Just like Twitter! Comments are not run as R code, so they will not influence your result. For example, *Calculate 3 + 4* in the editor on the right is a comment.

You can also execute R commands straight in the console. This is a good way to experiment with R code, as your submission is not checked for correctness.

Instructions100 XP

- In the editor on the right there is already some sample code. Can you see which lines are actual R code and which are comments?
- Add a line of code that calculates the sum of 6 and 12, and hit the 'Submit Answer' button.

Take Hint (-30 XP)

Course Outline

script.R

```
1 # Calculate 3 + 4
2 3 + 4
3
4 # Calculate 6 + 12
5
```

↺

Run Code

Submit Answer

R Console

> |

DataCamp

campus.datacamp.com/courses/visualization-best-practices-in-r/comparing-distributions?

Home Gmail Gcal GitHub BIMM143 BGGN213 GDrive Atmosphere CloudLaunch BIMM194 Blink News +

Schedule · BIMM 143

A simple boxplot | R

DataCamp

Exercise

A simple boxplot

Let's get started by simply making a box plot similar to the ones we made in the slides.

Modify the `filter()` to look at `'RED'` cars instead of blue and then map the x-axis to `gender` and the y-axis to driver `speed`. Add your boxplot geometry and give it a title.

Instructions

100 XP

- Modify `filter()` command to filter to `'RED'` vehicles.
- Fill in `aes()` by mapping `x` to `gender` and `y` to `speed`.
- Add a boxplot geometry to the plot.
- Title plot `'Speed of red cars by gender of driver'`.

Take Hint (-30 XP)

script.R

```
1 md_speeding %>%
2   filter(vehicle_color == 'BLUE') %>%
3   # Map x and y to gender and speed
   columns respectively
4   ggplot(aes(x=gender, y=speed)) +
5   # add a boxplot geometry
6   geom_boxplot() +
7   # give plot supplied title
8   labs(title = 'My fun subtitle')
```

Run Code

Submit Answer

Plots

My fun subtitle

Previous Plot

1/1

Next Plot

R Console

Slides

```
> md_speeding %>%
  filter(vehicle_color == 'BLUE') %>%
  # Map x and y to gender and speed columns respectively
  ggplot(aes(x=gender, y=speed)) +
  # add a boxplot geometry
  geom_boxplot() +
  # give plot supplied title
  labs(title = 'My fun subtitle')
> |
```



BIOINFORMATICS

Hands-on Lab Session

Class 04

Barry Grant
UC San Diego

<http://thegrantlab.org/teaching>