



# Introduction to R/RStudio

**Patrick Kimes, PhD**

Postdoctoral Fellow

Dana-Farber Cancer Institute

Harvard TH Chan School of Public Health



Top Ten Seminars  
October 1, 2019

# top ten seminars in data science

**October 1, 2019**

October 22, 2019

November 12, 2019

December 10, 2019

January 21, 2020

February 11, 2020

March 17, 2020

April 14, 2020

May 19, 2020

June 16, 2020

**Introduction to R and RStudio**

Data visualization with ggplot2

Data visualization principles and plots to avoid

Design of Clinical Trials Basics

Correlation: you are probably using it wrong

How to detect and deal with batch effects

Brief introduction to machine learning

Culprits of the reproducibility crisis: multiple testing and p-hacking

Experimental Design: How many size and should I pool?

Detecting differentially expressed genes with RNA-seq

**R and RStudio?**

# R and RStudio?



**R**

**programming  
language**  
think Java, C,  
**C++, Python, ...**



**R and RStudio?**

**RStudio**



**software to make data  
analysis with R easier**

**R and RStudio?**



**RStudio**

**R**

**programming  
language**

**software to make data  
analysis with R easier**

**Studio<sup>®</sup>**

RStudio

Console ~/ ↗

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.

>

Environment History Connections

Import Dataset | Global Environment | List | C

Environment is empty

R Studio®

Files Plots Packages Help Viewer

Home Find in Topic

R Resources

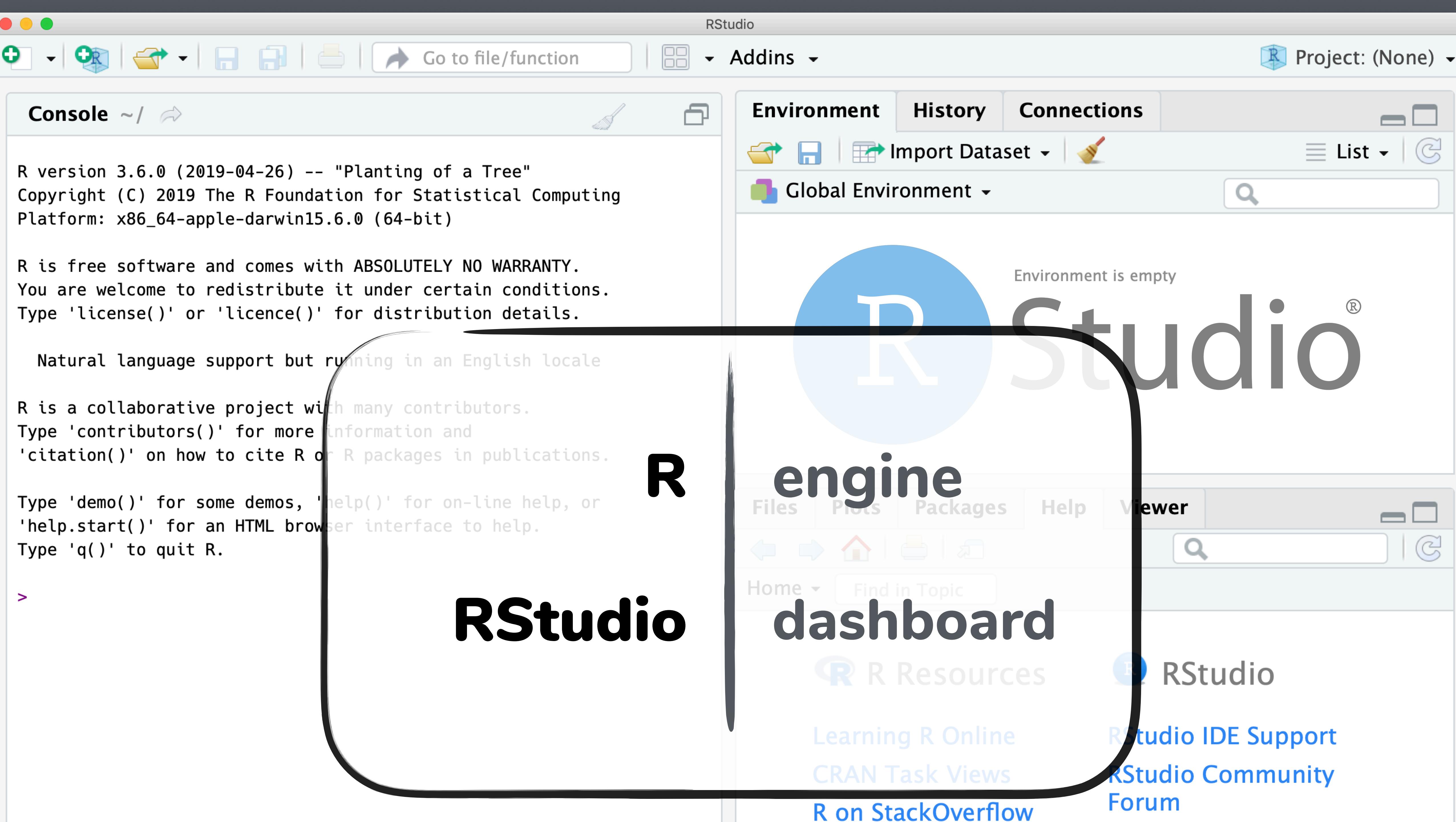
Learning R Online

CRAN Task Views

R on StackOverflow

RStudio IDE Support

RStudio Community Forum



**why  
R and RStudio?**

**why  
R and RStudio?**

**what about SAS?**

# why R and RStudio?

- R is free, open source
- R is the home of new methods
- R has a large, active community
- R is highly interoperable, extensible



what about SAS?

**why  
R and RStudio?**

**what about Python?**

# why R and RStudio?

- Good question! Up to you!
- R is arguably easier to learn
- R has more statistical tools
- R makes exploration and visualization easier



what about Python?

**why  
R and RStudio?**



**why**

**R and RStudio?**

**it gets you to  
the data *fast!***



アボガド6  
@avogado6

## 現実逃避

[Translate Tweet](#)

8:03 AM · Aug 25, 2019 · Twitter Web Client

115.4K Retweets 384.4K Likes

it gets you to  
the data **fast!**  
and that's **fun!**

<https://twitter.com/avogado6/status/1165595520967954432>



who already has  
R / RStudio installed?

# how to install R and RStudio

---

# How to draw an Owl.

---

*"A fun and creative guide for beginners"*

---

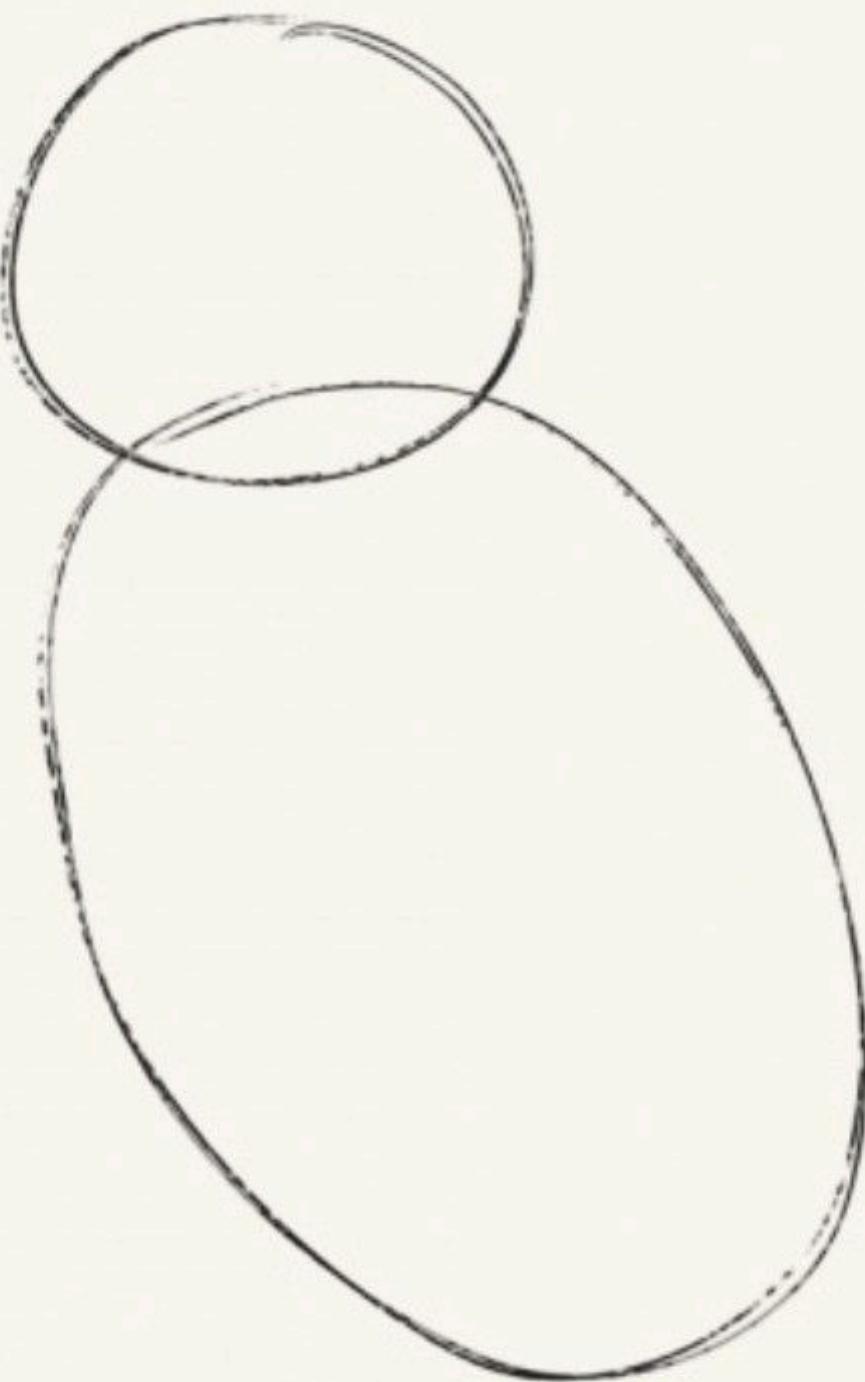


Fig 1. Draw two circles



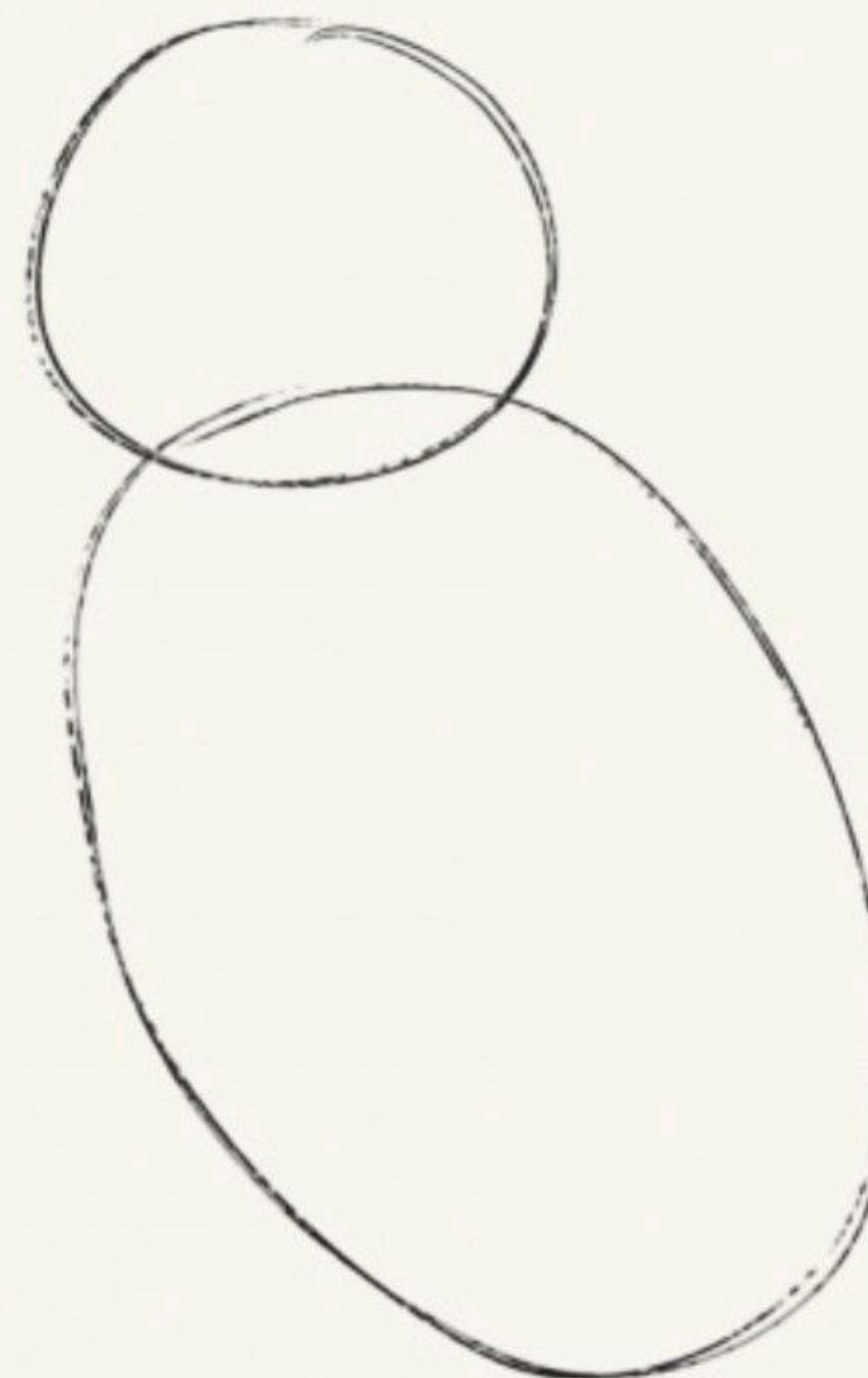
Fig 2. Draw the rest of the damn Owl

---

---

# how to install R and RStudio

---



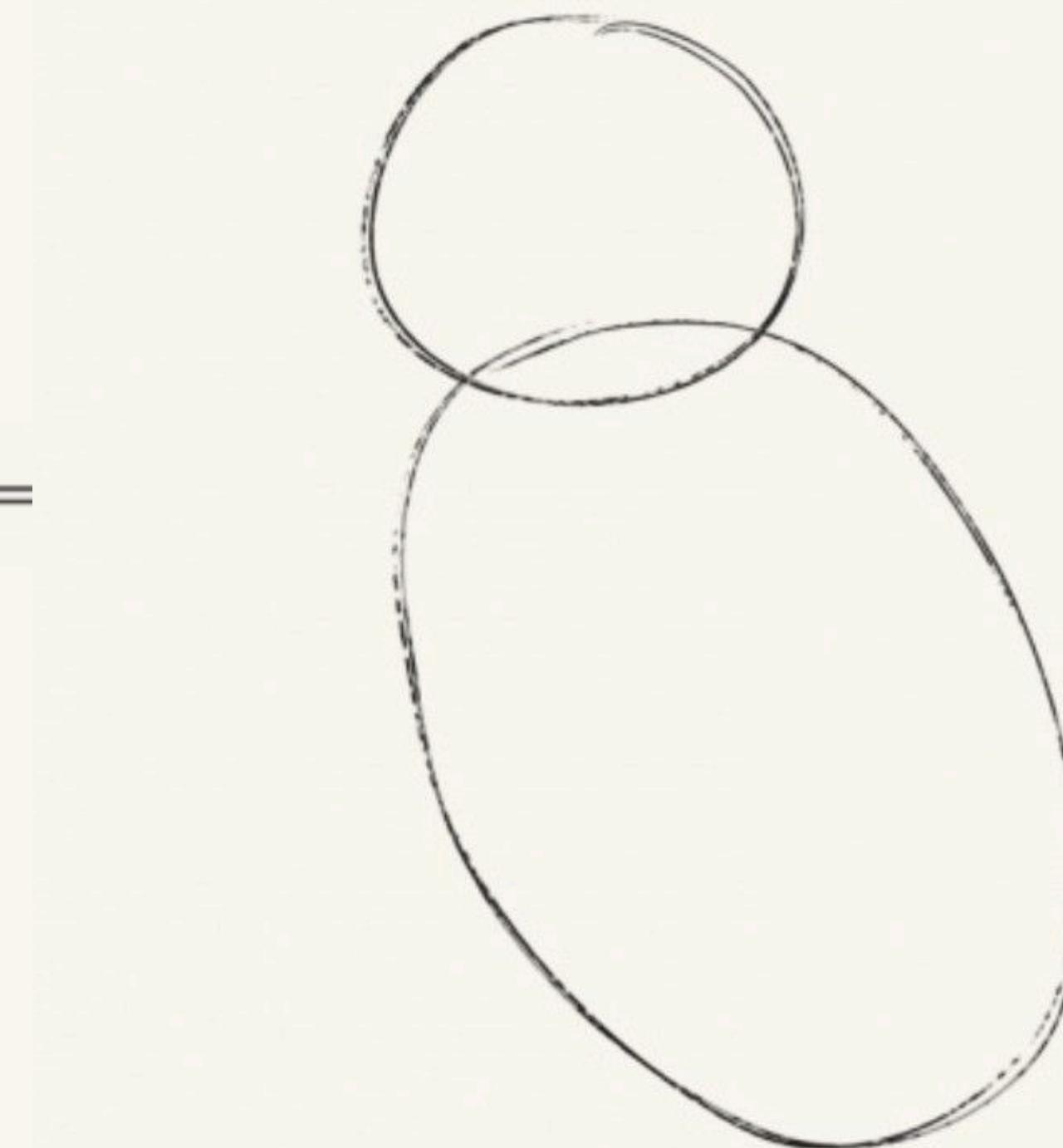
1. Search “R”,  
Search “RStudio”
2. Install “R”,  
Install “RStudio”

# how to install R and RStudio

maybe a few more steps  
so please do this later

a much better guide:

**[rafaelab.github.io/dsbook/installing-r-rstudio](https://rafaelab.github.io/dsbook/installing-r-rstudio)**



1. Search “R”,  
Search “RStudio”
2. Install “R”,  
Install “RStudio”

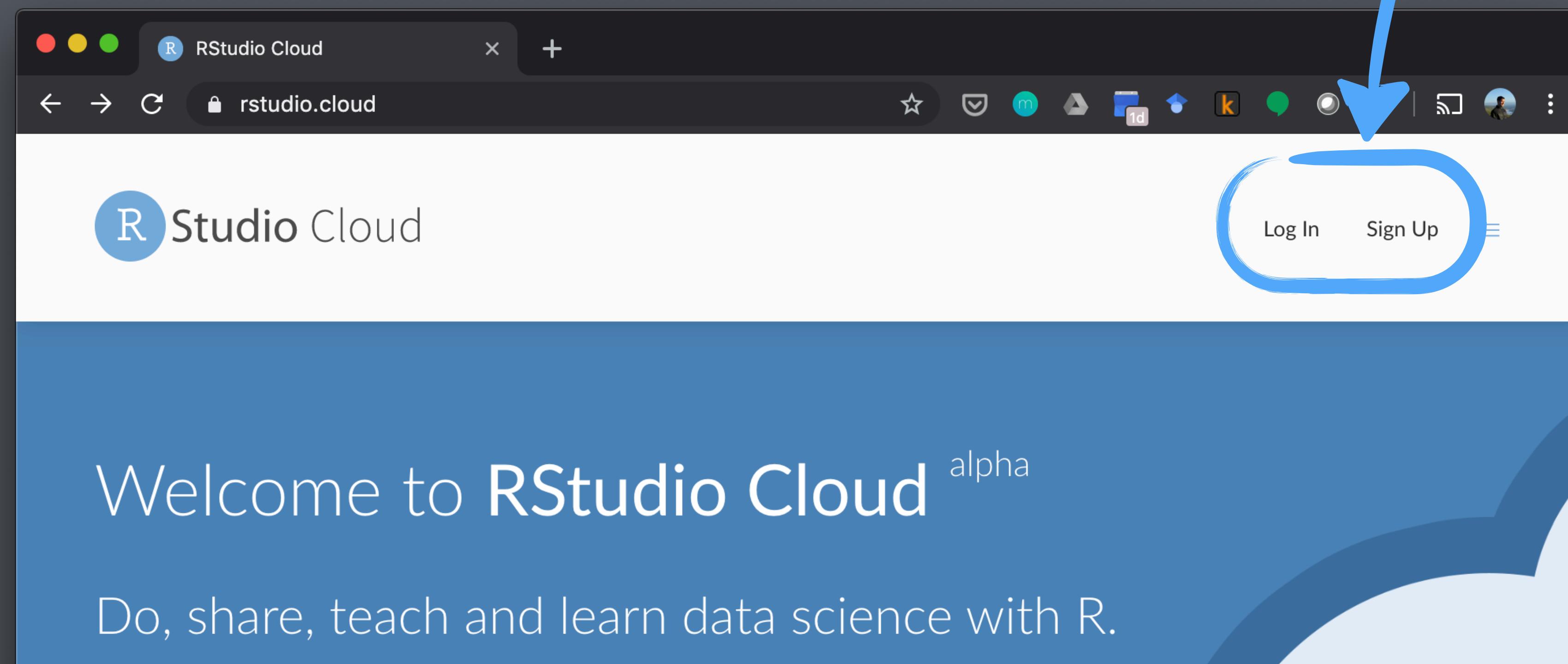
lucky us!  
we have a workaround!

<https://rstudio.cloud>

lucky us!  
we have a workaround!

do this

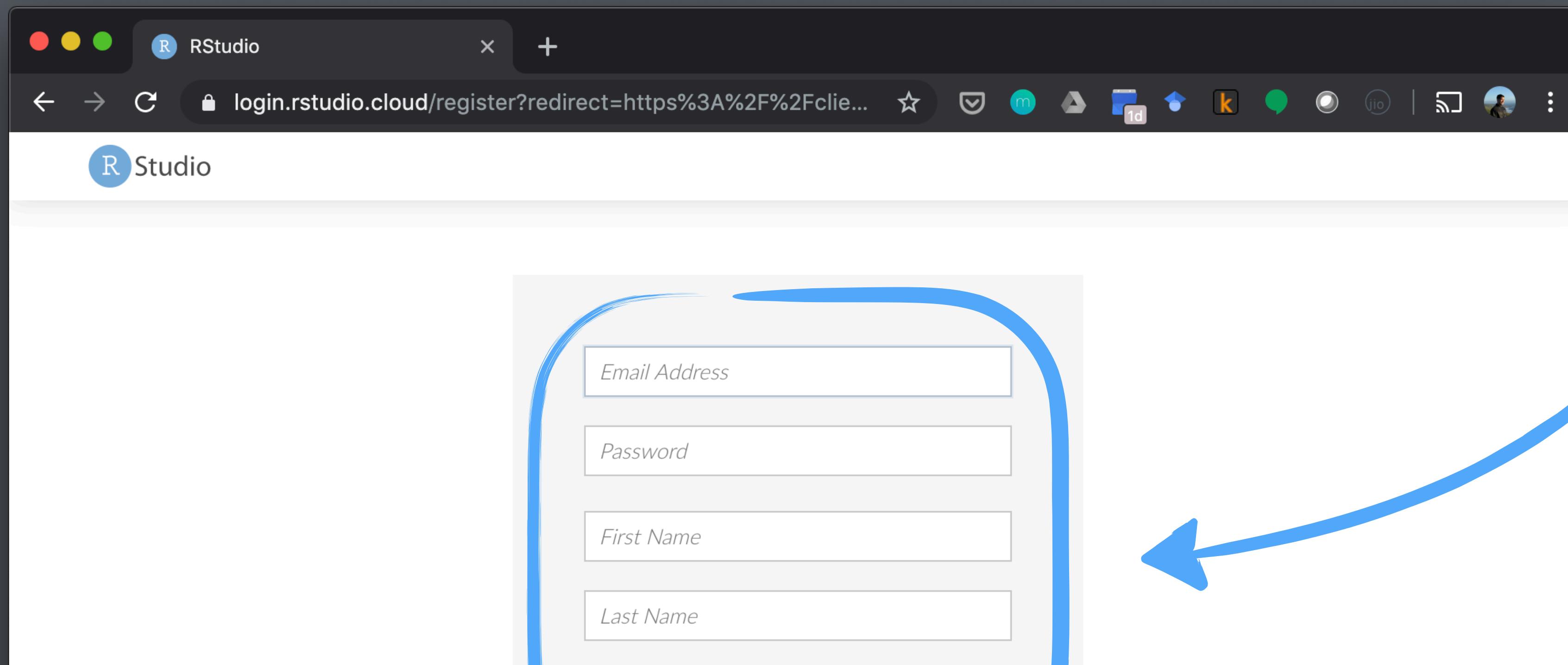
<https://rstudio.cloud>



lucky us!  
we have a workaround!

<https://rstudio.cloud>

this too

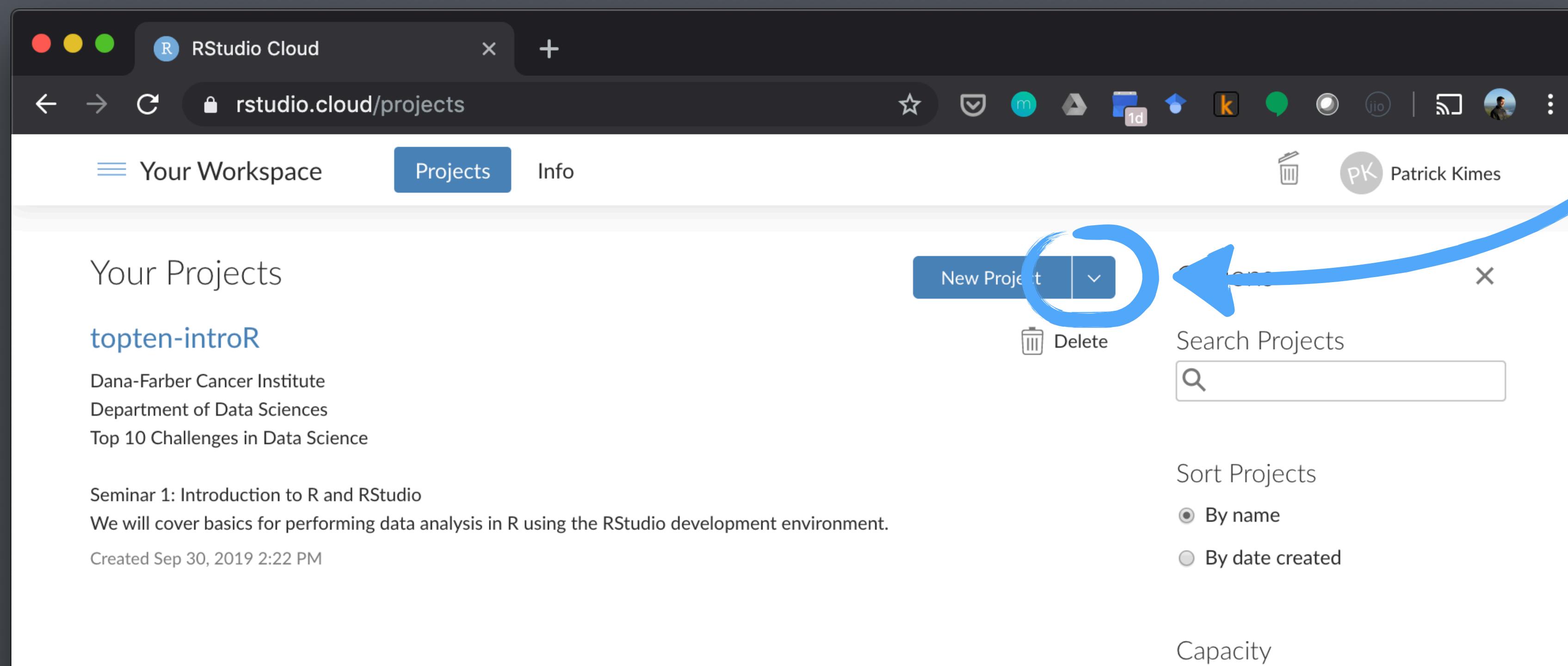


# lucky us!

# we have a workaround!

<https://rstudio.cloud>

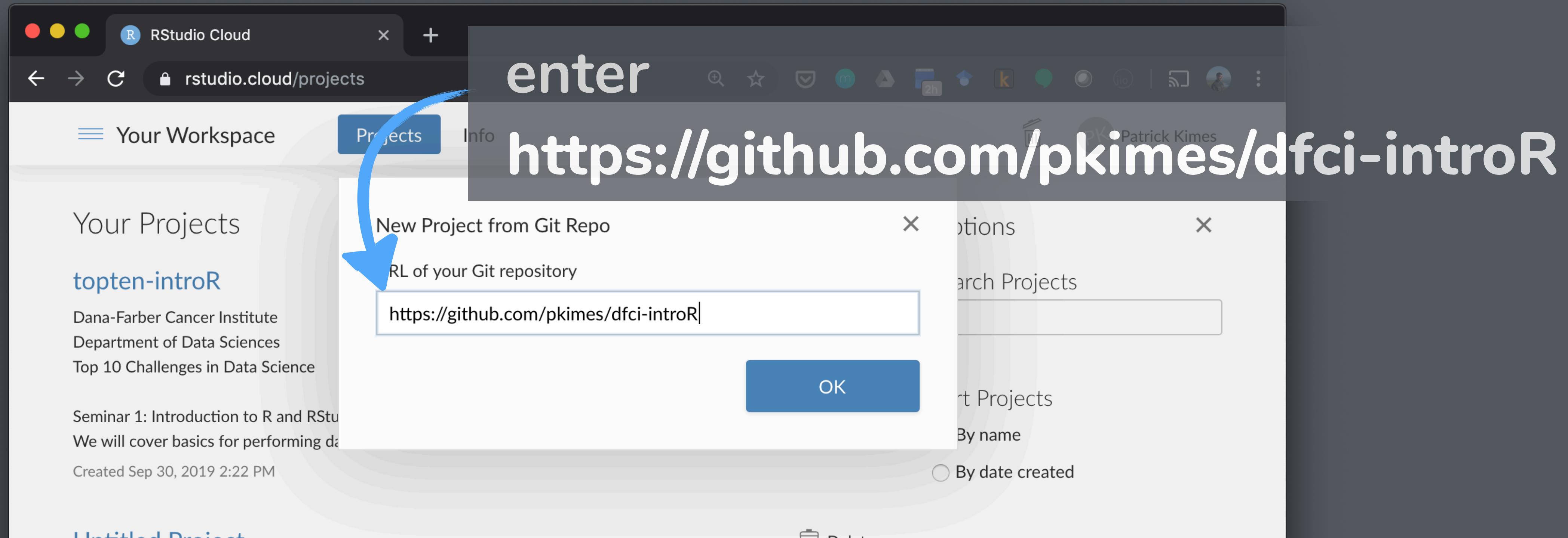
select  
[from Git repo]



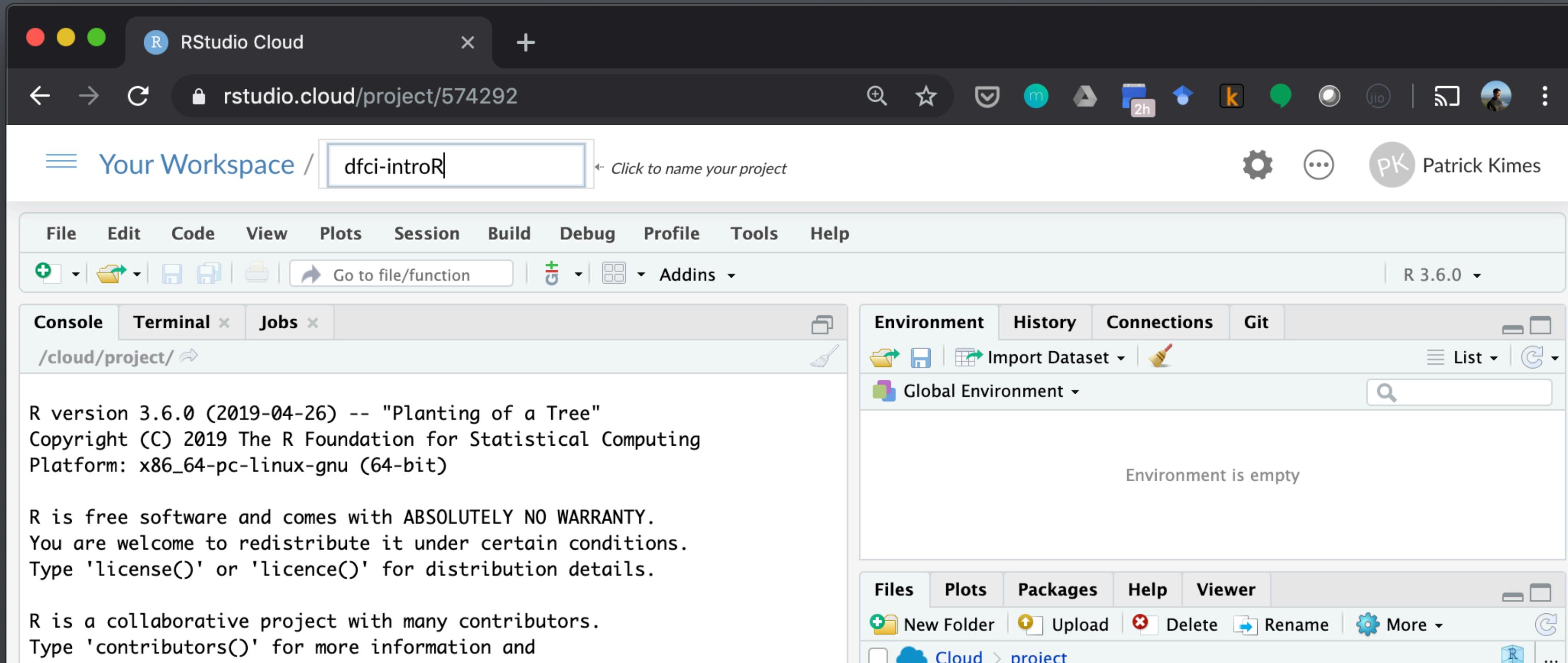
# lucky us!

# we have a workaround!

<https://rstudio.cloud>

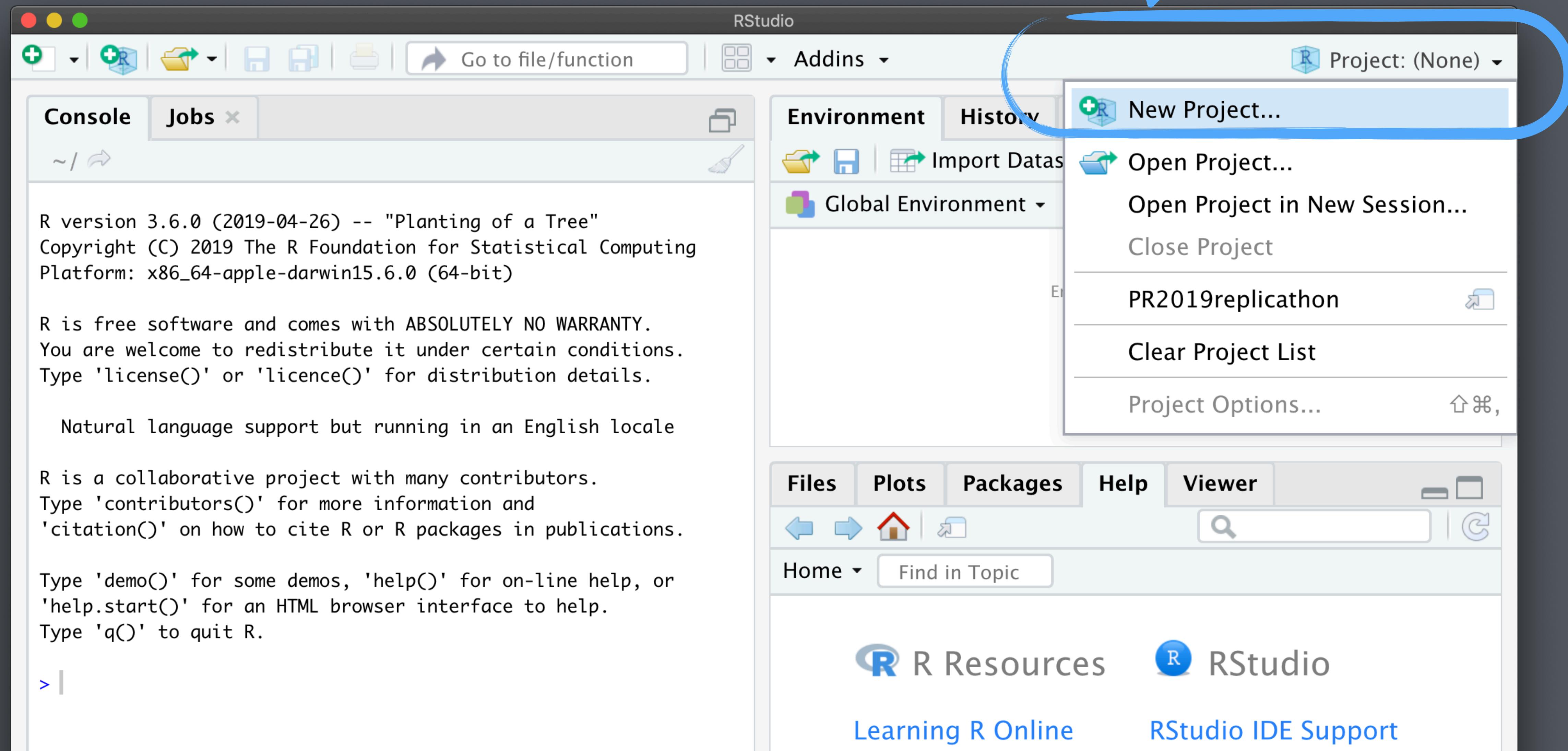


# welcome to RStudio!

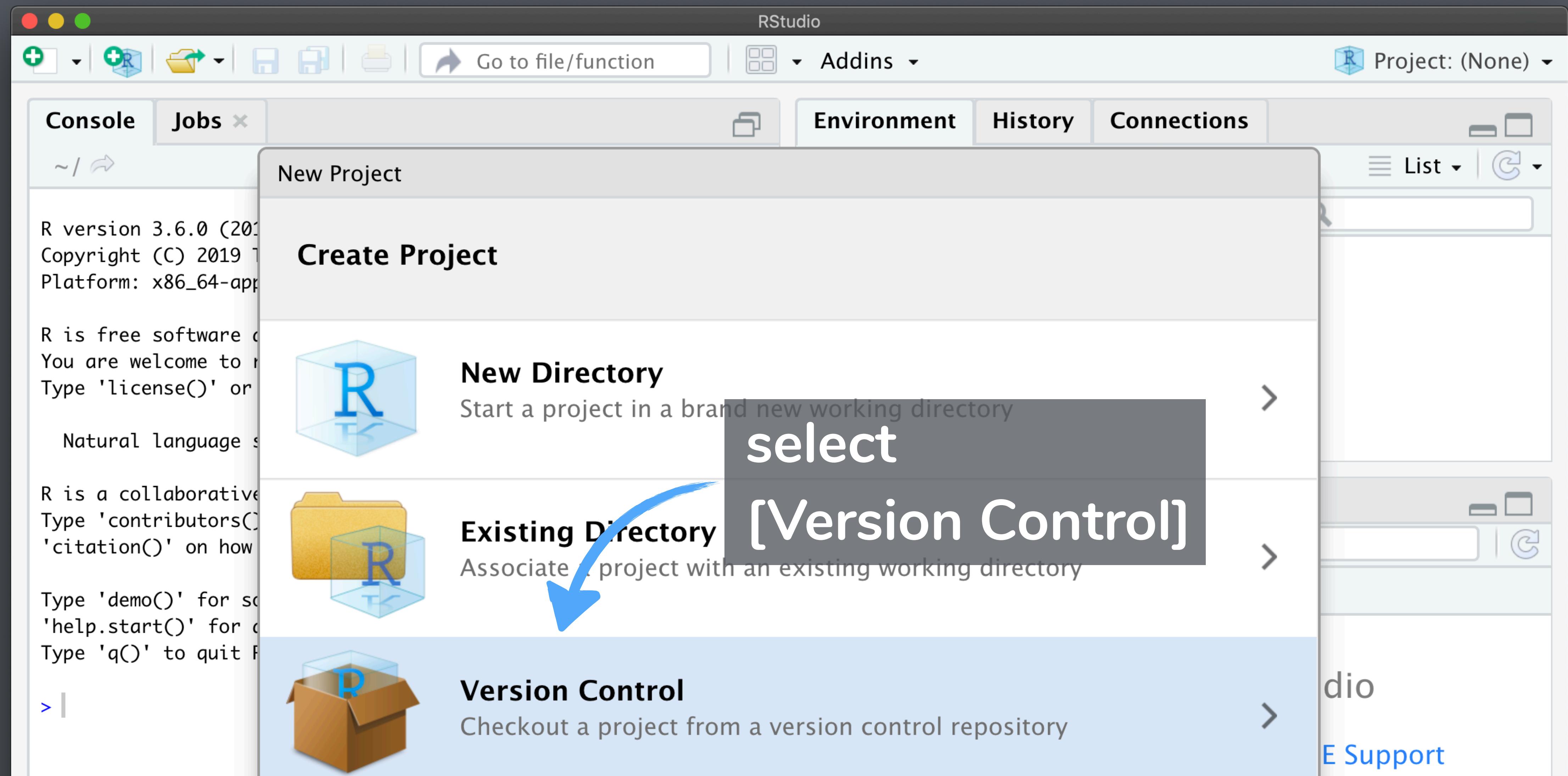


# local RStudio

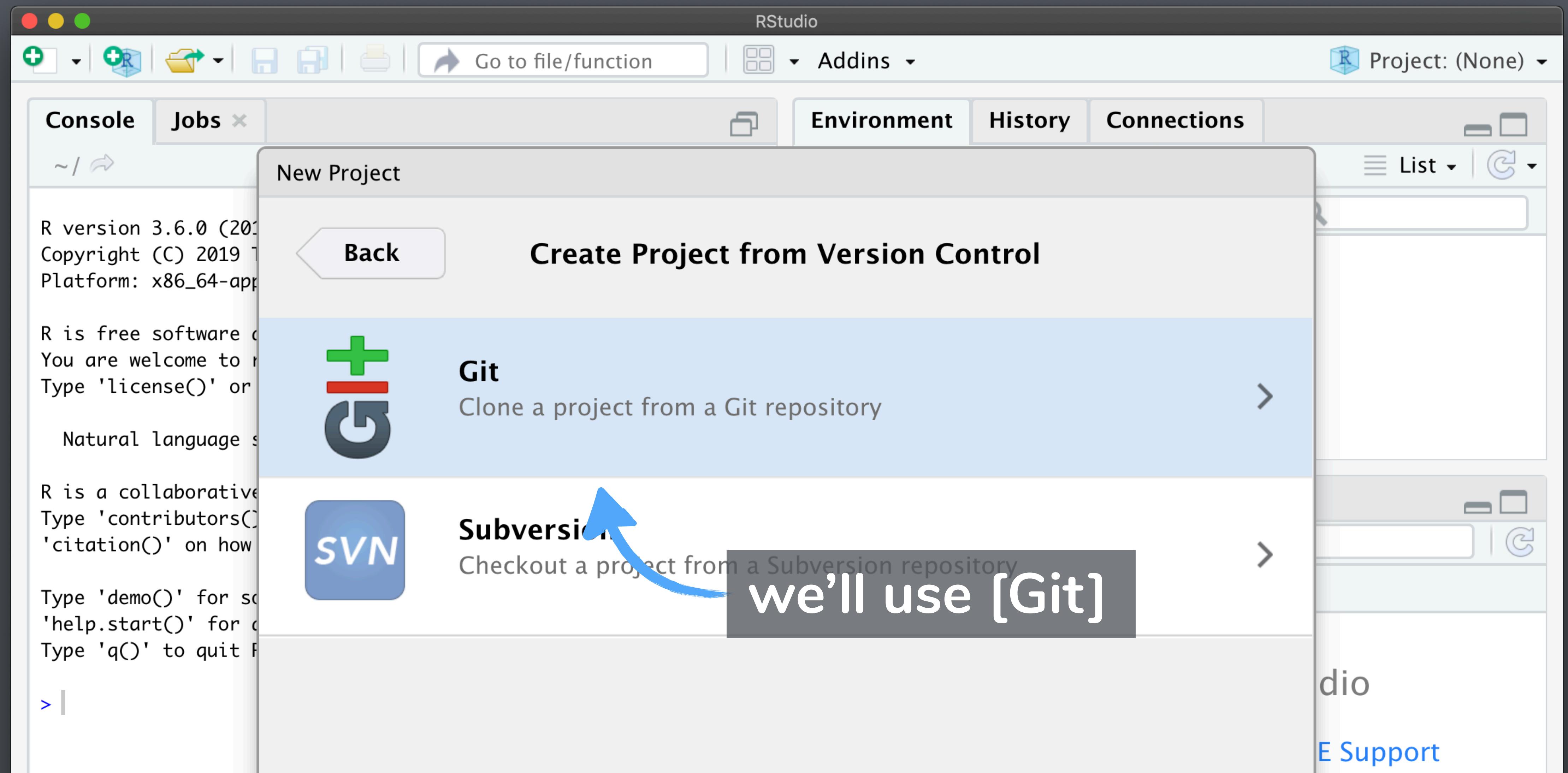
select  
[New Project]



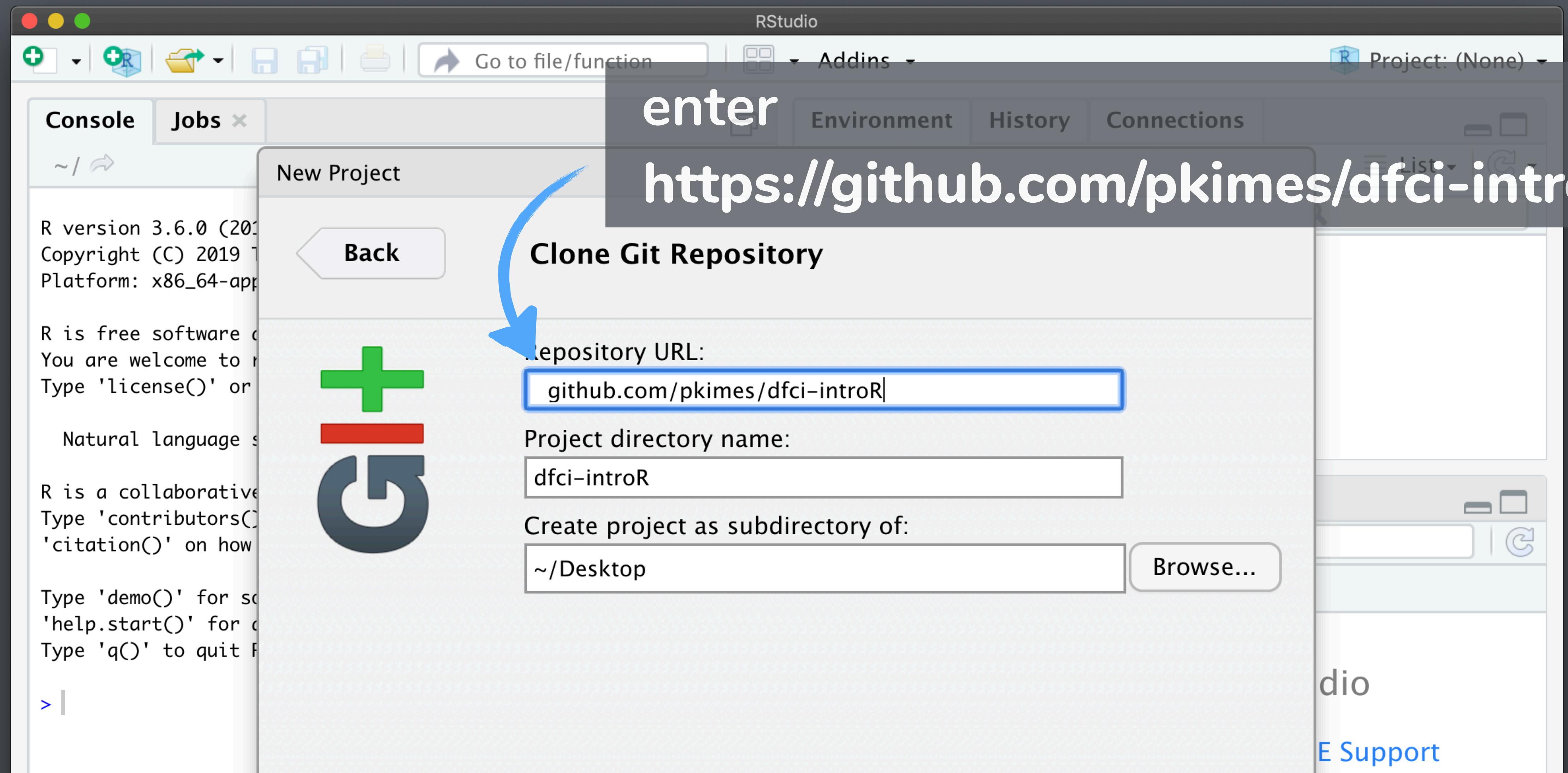
# local RStudio



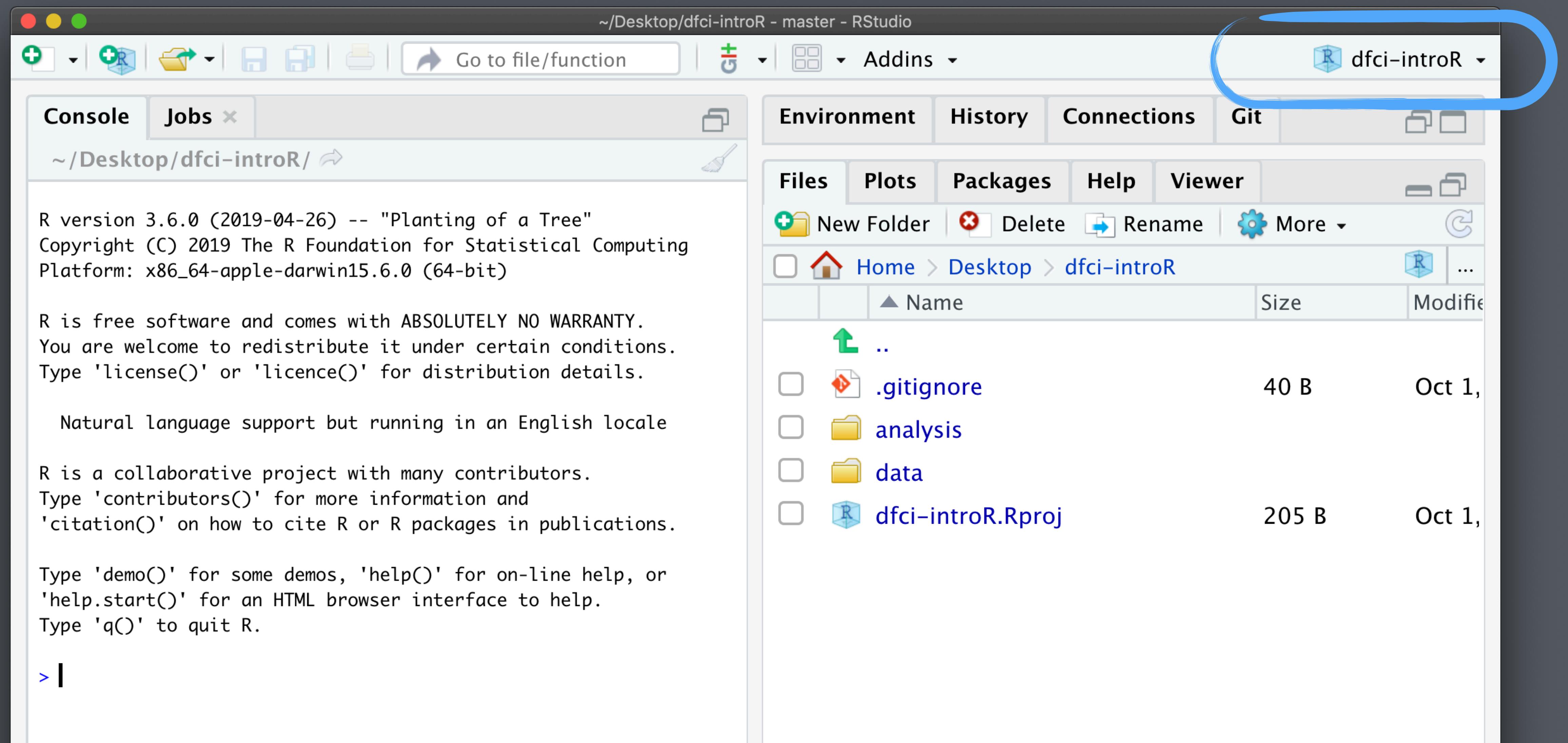
# local RStudio



# local RStudio



# we're good to go!



The screenshot shows the RStudio interface with the following details:

- Console Tab:** Shows the R startup message and basic information about the R version and platform.
- Environment Tab:** Shows the file structure of the current project directory (`dfci-introR`) located at `~/Desktop/dfci-introR`. The files listed are `..`, `.gitignore`, `analysis`, `data`, and `dfci-introR.Rproj`.
- Project Bar:** Shows the project name `dfci-introR` selected.
- Toolbar:** Includes standard icons for file operations like Open, Save, Print, and Go to file/function.
- Top Bar:** Shows the current working directory as `~/Desktop/dfci-introR - master - RStudio` and the project name `dfci-introR`.

A blue circle highlights the `dfci-introR` package icon in the Environment tab.

```
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

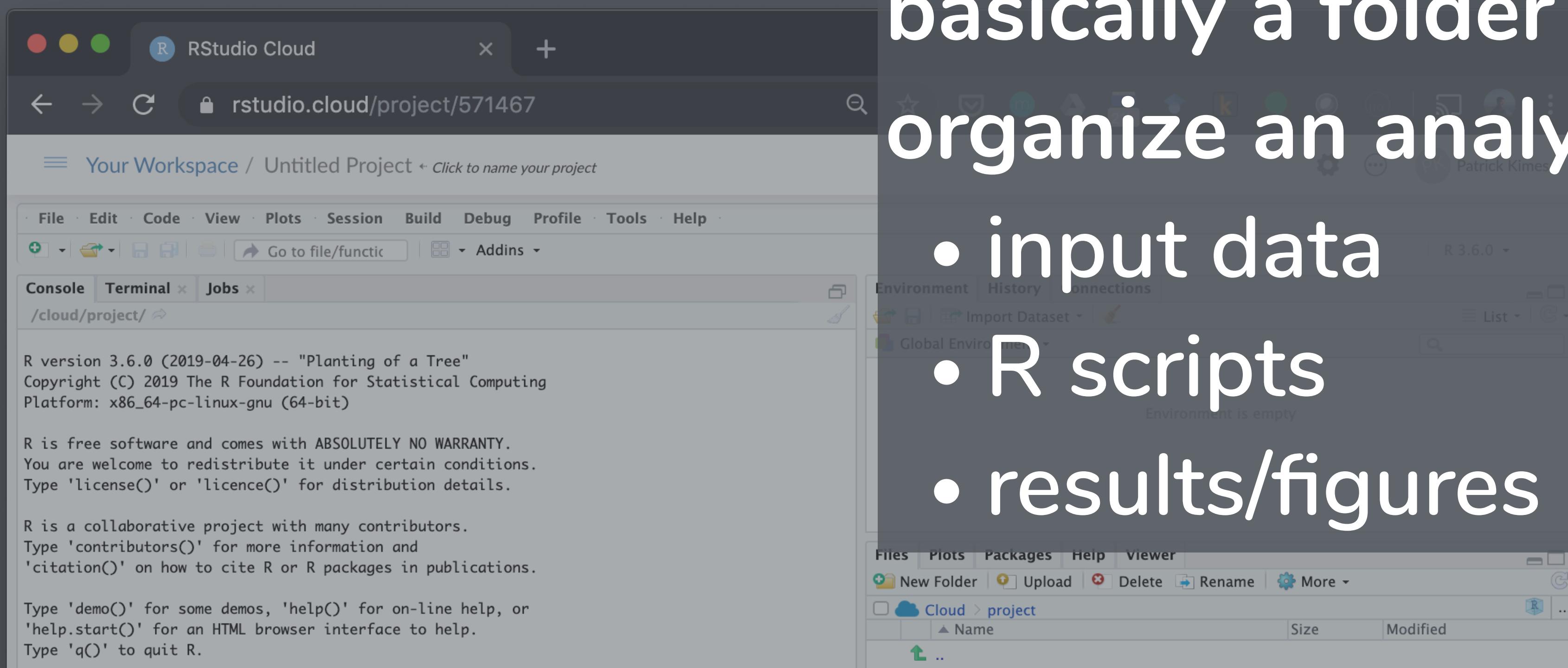
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

> |
```

# you now have a project!

# what's an RStudio project?



basically a folder to  
organize an analysis

- input data
- R scripts
- results/figures

let's give it a try!

**what did we  
(hopefully) cover?**

arithmetic  
variables  
functions  
help  
installing packages  
loading packages  
for-loops

# **some pieces in the modern (R) data scientist's toolbox**

# **some pieces in the modern (R) data scientist's toolbox**

**rmarkdown**

documentation, communication

**tidyverse**

data manipulation, visualization

**shiny**

web application framework

**[bioconductor]**

community of genomics packages

# **some pieces in the modern (R) data scientist's toolbox**

**rmarkdown**

documentation, communication

**tidyverse**

data manipulation, visualization

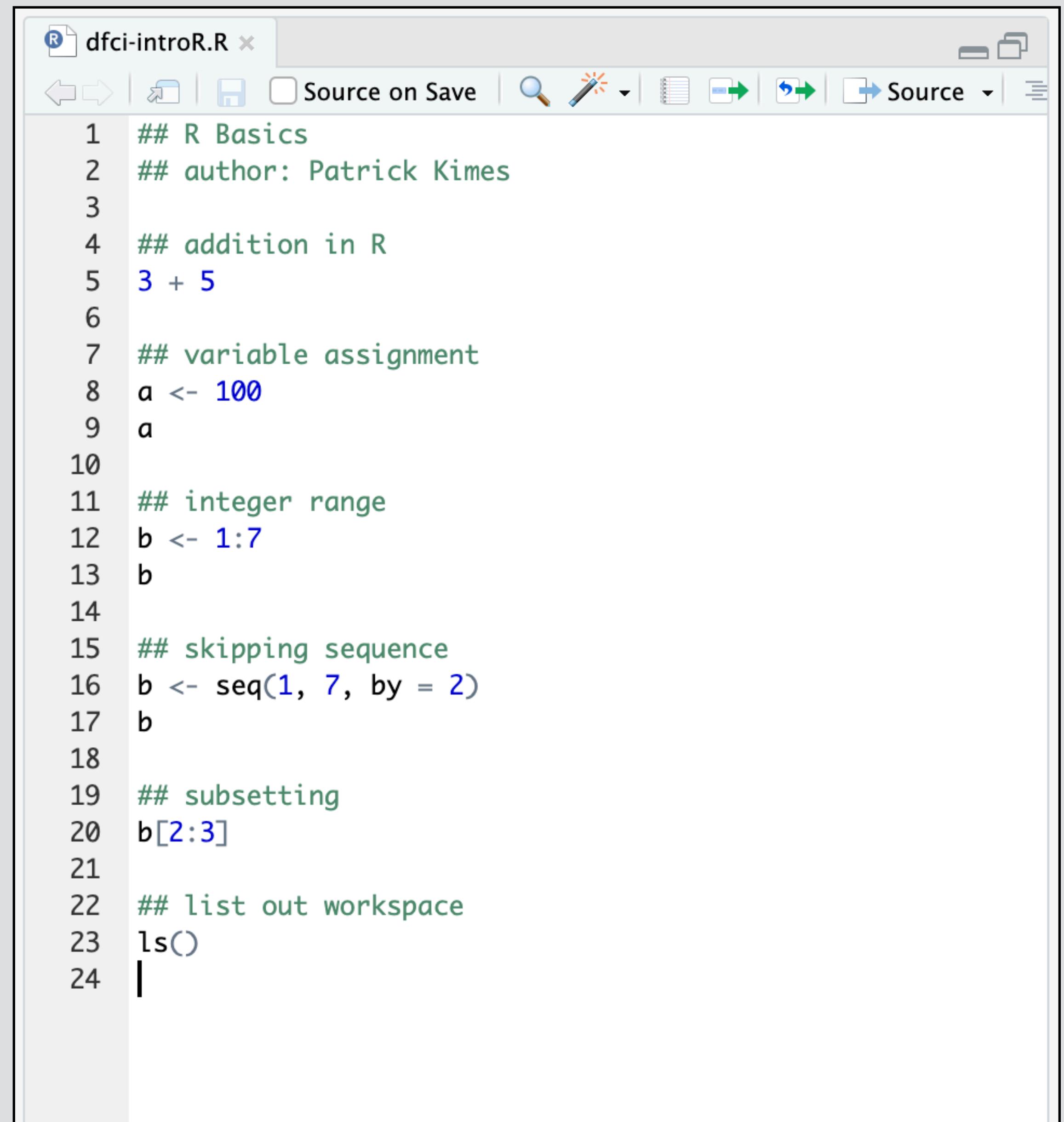
**shiny**

web application framework

**[bioconductor]**

community of genomics packages

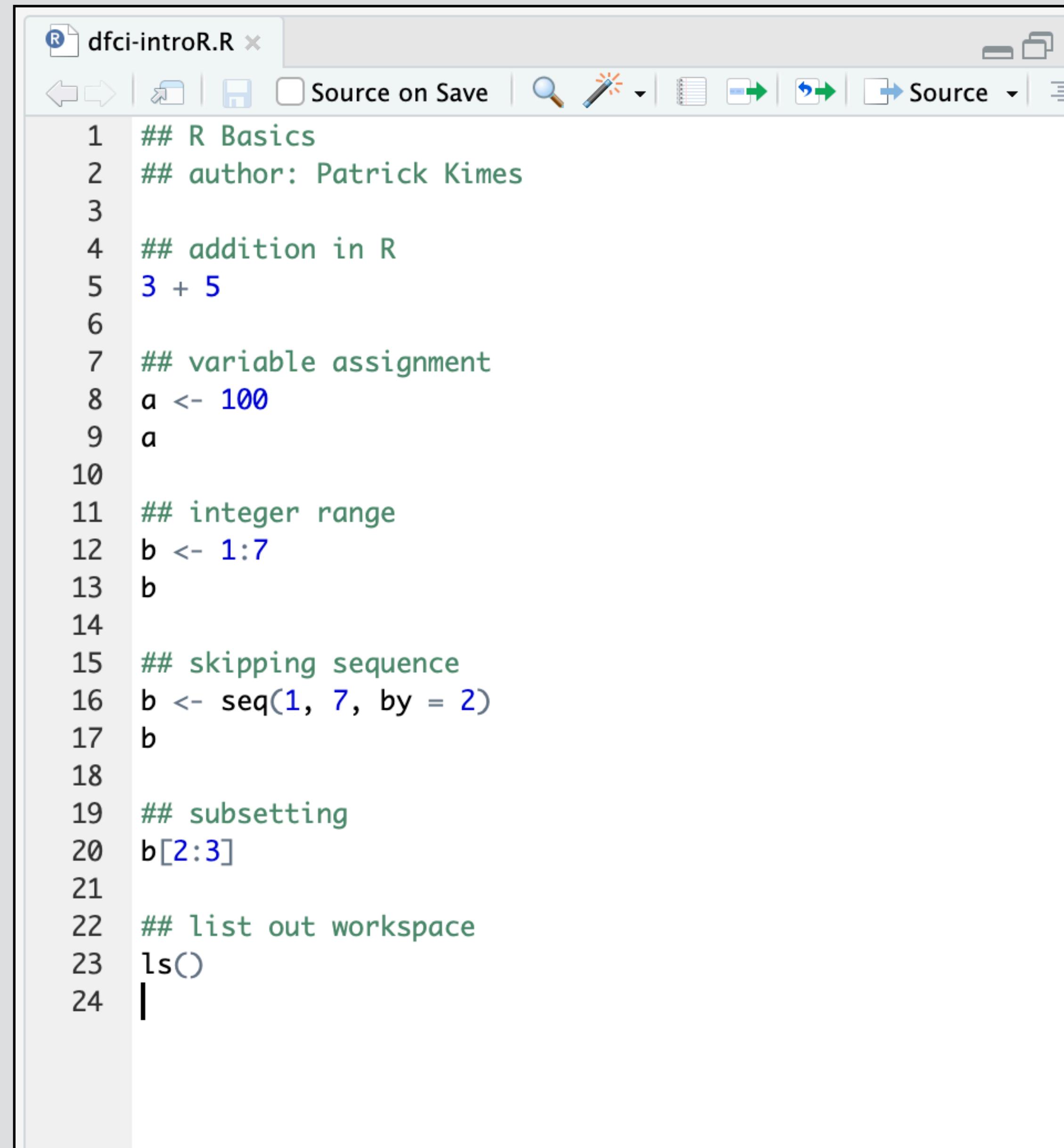
# .R file



The screenshot shows the RStudio interface with an open R script file titled "dfci-introR.R". The code in the editor window is as follows:

```
1 ## R Basics
2 ## author: Patrick Kimes
3
4 ## addition in R
5 3 + 5
6
7 ## variable assignment
8 a <- 100
9 a
10
11 ## integer range
12 b <- 1:7
13 b
14
15 ## skipping sequence
16 b <- seq(1, 7, by = 2)
17 b
18
19 ## subsetting
20 b[2:3]
21
22 ## list out workspace
23 ls()
```

# .R file



```
## R Basics
## author: Patrick Kimes
## addition in R
3 + 5

## variable assignment
a <- 100
a

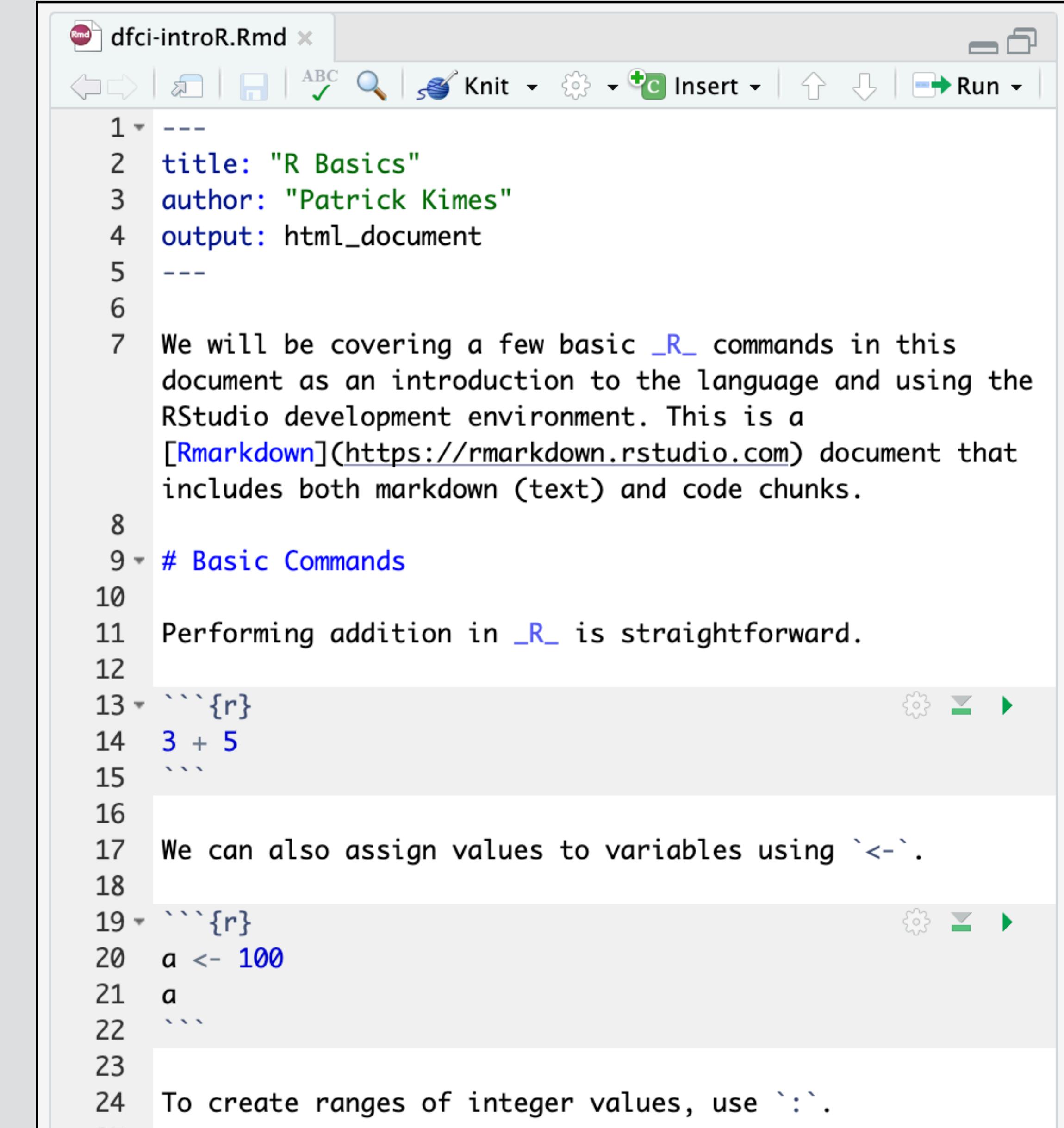
## integer range
b <- 1:7
b

## skipping sequence
b <- seq(1, 7, by = 2)
b

## subsetting
b[2:3]

## list out workspace
ls()
```

# .Rmd file



```
---
title: "R Basics"
author: "Patrick Kimes"
output: html_document
---

We will be covering a few basic _R_ commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](https://rmarkdown.rstudio.com) document that includes both markdown (text) and code chunks.

# Basic Commands

Performing addition in _R_ is straightforward.

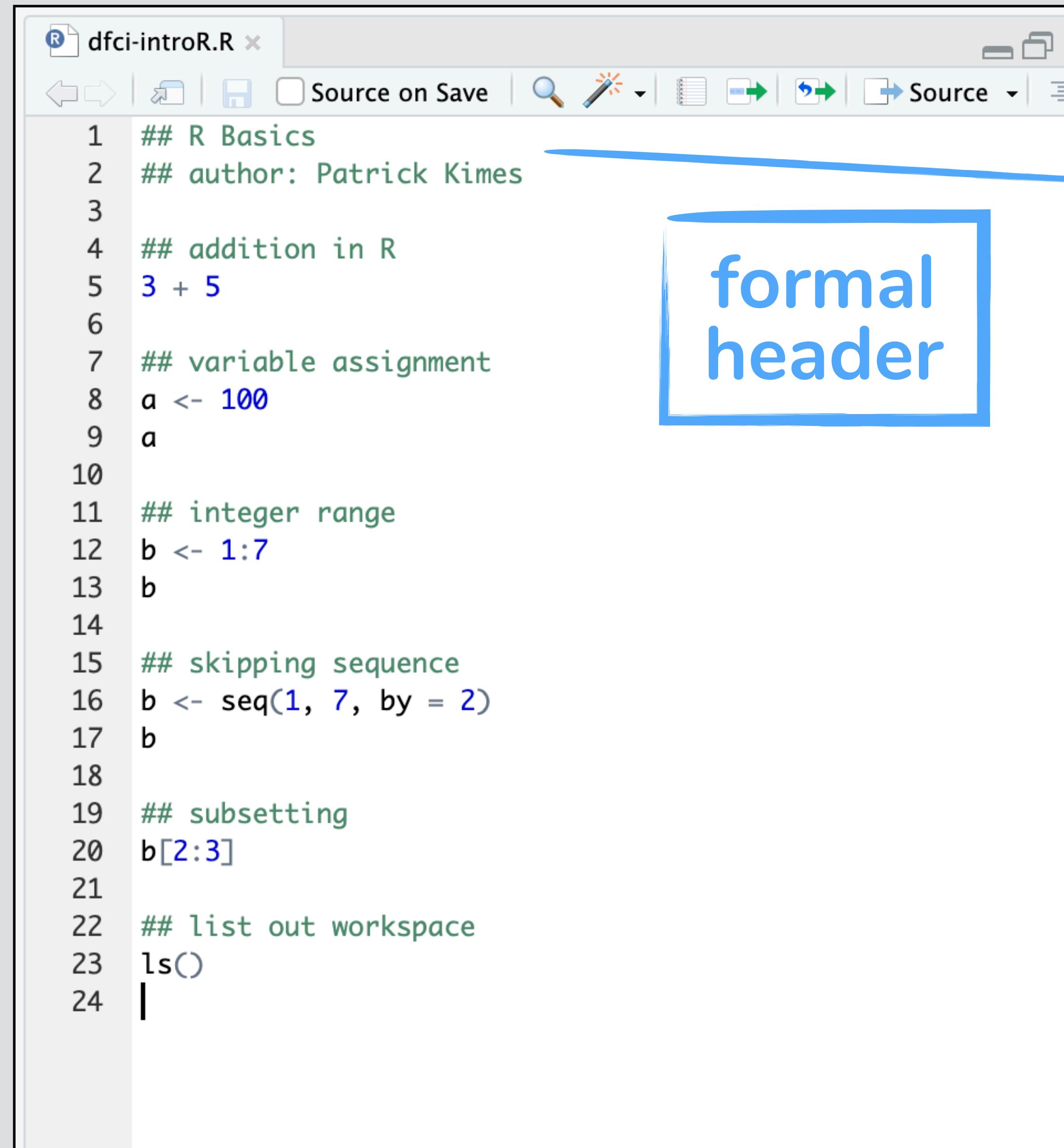
```{r}
3 + 5
```

We can also assign values to variables using `<-`.

```{r}
a <- 100
a
```

To create ranges of integer values, use `:`.
```

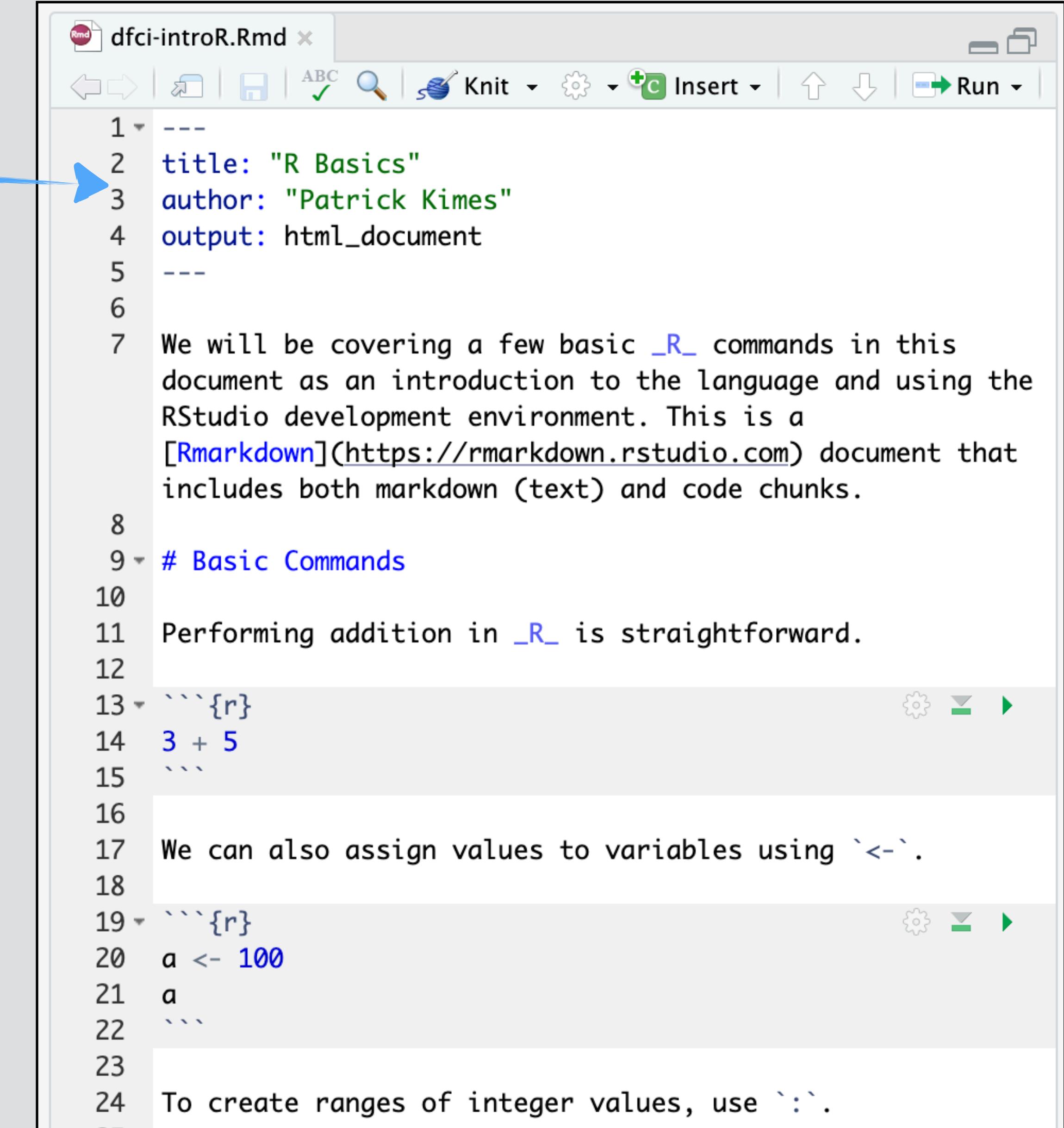
# .R file



```
## R Basics
## author: Patrick Kimes
## addition in R
3 + 5
## variable assignment
a <- 100
a
## integer range
b <- 1:7
b
## skipping sequence
b <- seq(1, 7, by = 2)
b
## subsetting
b[2:3]
## list out workspace
ls()
```

formal header

# .Rmd file



```
---
title: "R Basics"
author: "Patrick Kimes"
output: html_document
---

We will be covering a few basic R commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](https://rmarkdown.rstudio.com) document that includes both markdown (text) and code chunks.

# Basic Commands

Performing addition in R is straightforward.

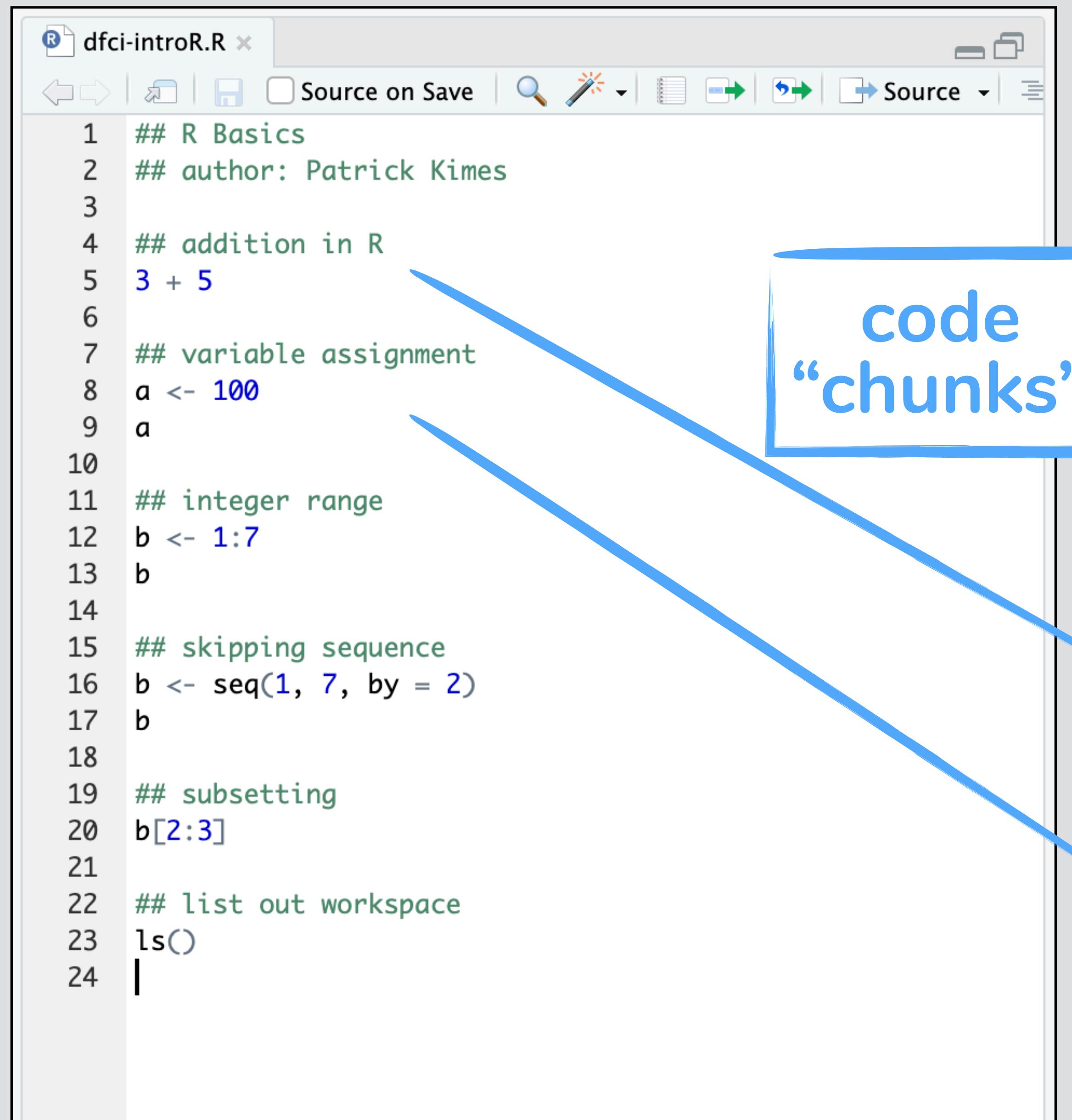
```{r}
3 + 5
```

We can also assign values to variables using <-.

```{r}
a <- 100
a
```

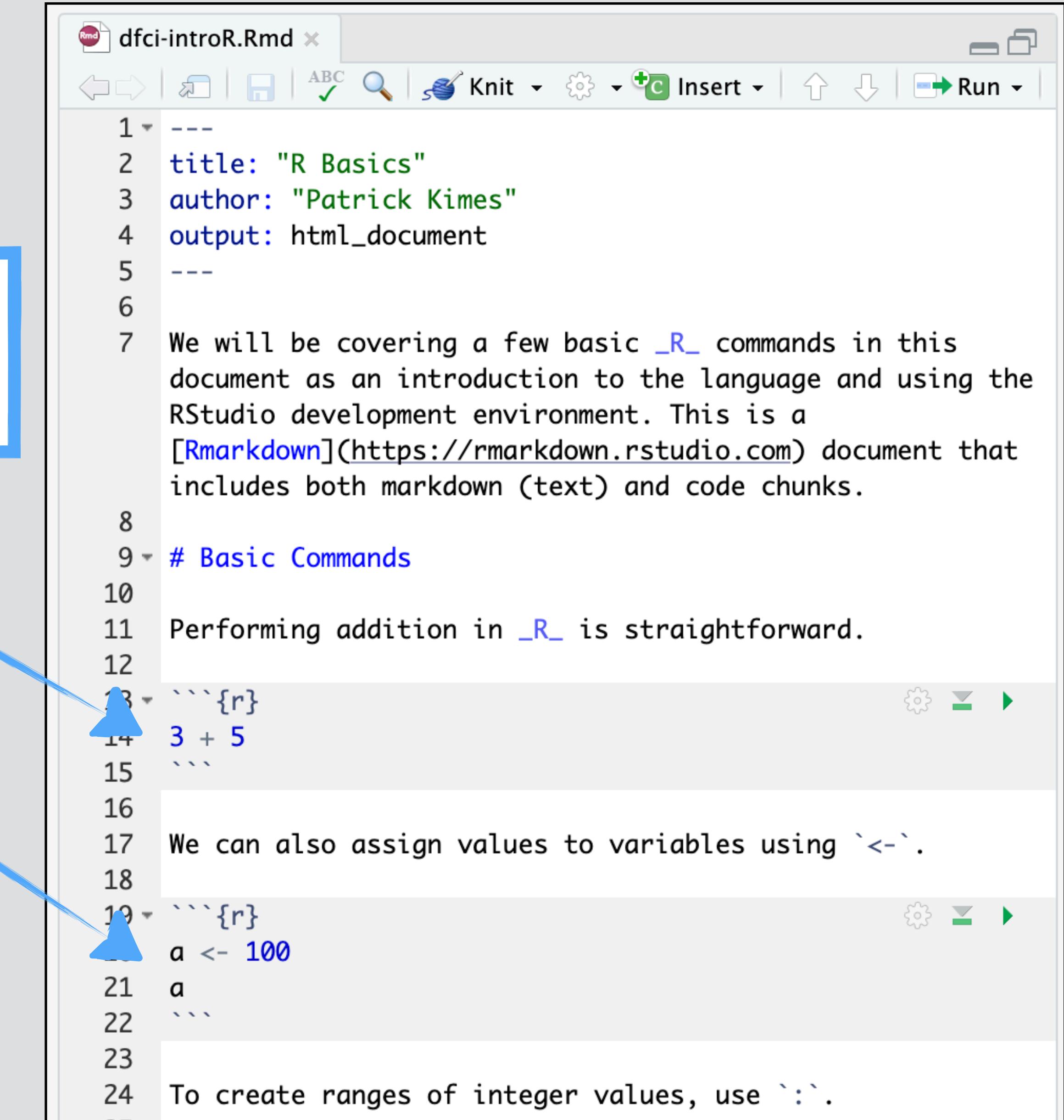
To create ranges of integer values, use :::.
```

# .R file



```
1 ## R Basics
2 ## author: Patrick Kimes
3
4 ## addition in R
5 3 + 5
6
7 ## variable assignment
8 a <- 100
9 a
10
11 ## integer range
12 b <- 1:7
13 b
14
15 ## skipping sequence
16 b <- seq(1, 7, by = 2)
17 b
18
19 ## subsetting
20 b[2:3]
21
22 ## list out workspace
23 ls()
```

# .Rmd file

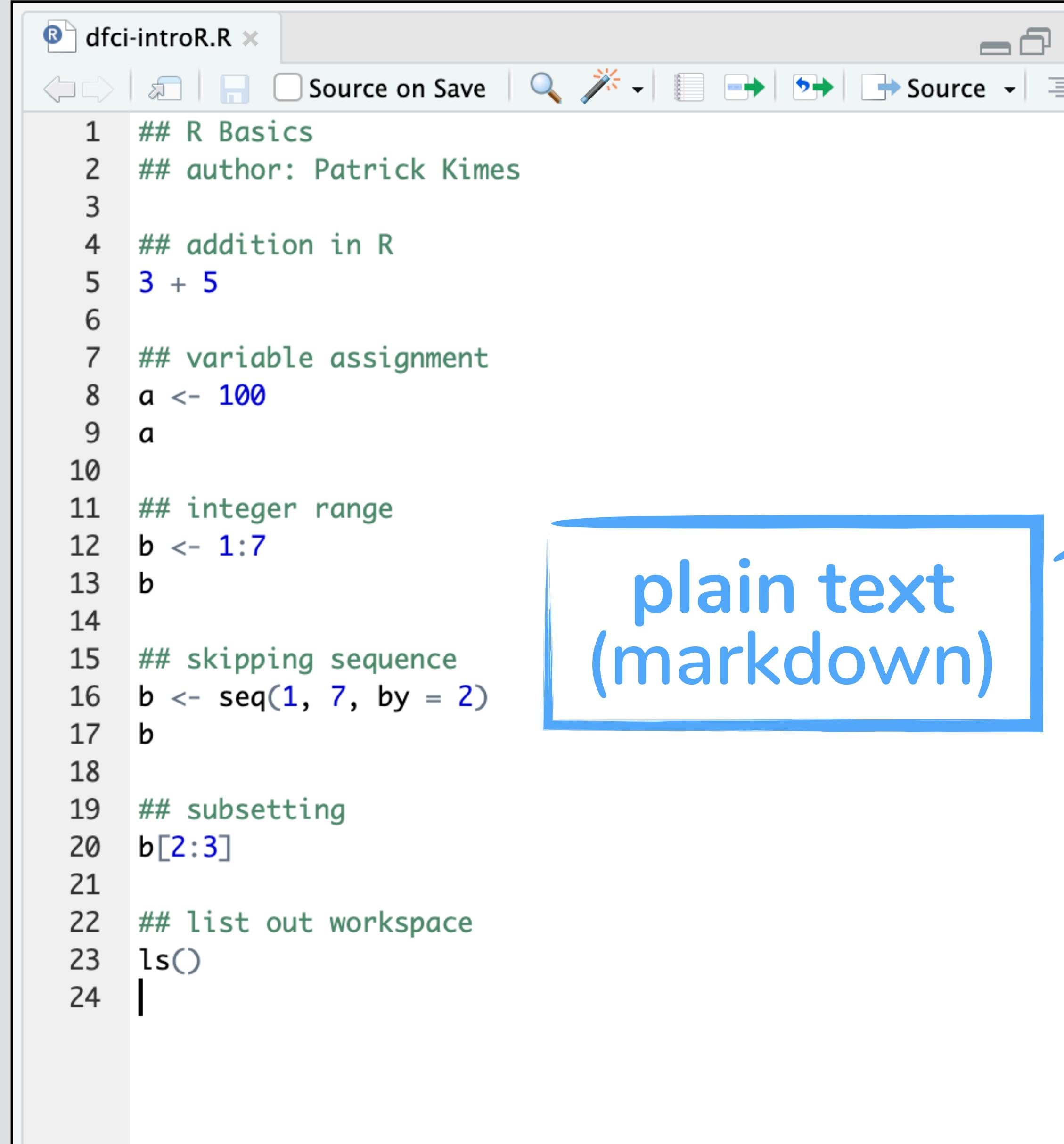


```
1 ---
2 title: "R Basics"
3 author: "Patrick Kimes"
4 output: html_document
5 ---
6
7 We will be covering a few basic R commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](https://rmarkdown.rstudio.com) document that includes both markdown (text) and code chunks.
8
9 # Basic Commands
10
11 Performing addition in R is straightforward.
12
13 ````{r}
14 3 + 5
15 ````

16
17 We can also assign values to variables using <-.
18
19 ````{r}
20 a <- 100
21 a
22 ````

23
24 To create ranges of integer values, use :::.
```

# .R file



```
## R Basics
## author: Patrick Kimes
## addition in R
3 + 5

## variable assignment
a <- 100
a

## integer range
b <- 1:7
b

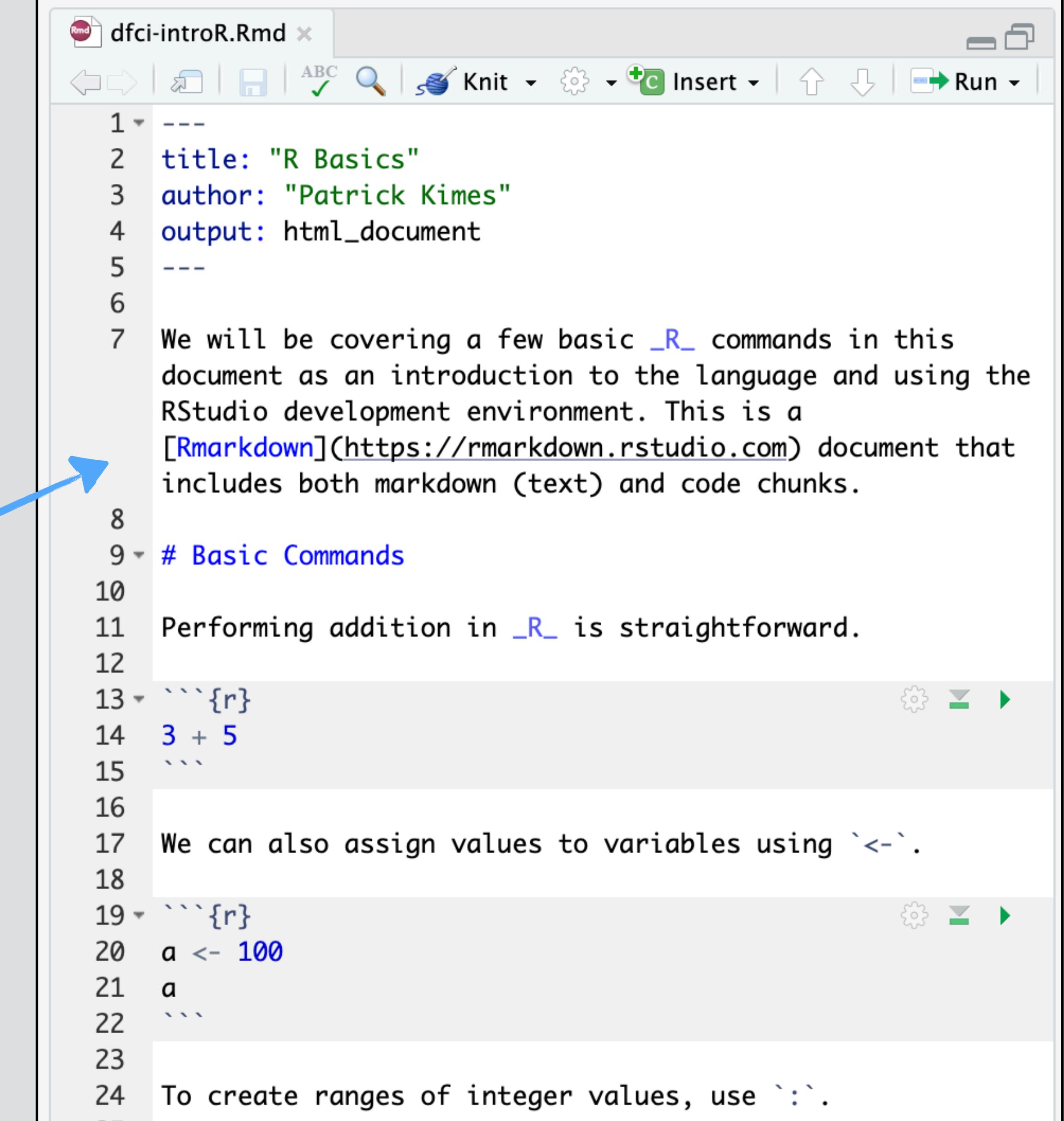
## skipping sequence
b <- seq(1, 7, by = 2)
b

## subsetting
b[2:3]

## list out workspace
ls()
```

plain text (markdown)

# .Rmd file



```
---
title: "R Basics"
author: "Patrick Kimes"
output: html_document
---

We will be covering a few basic R commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](https://rmarkdown.rstudio.com) document that includes both markdown (text) and code chunks.

# Basic Commands

Performing addition in R is straightforward.

```{r}
3 + 5
```

We can also assign values to variables using <-.

```{r}
a <- 100
a
```

To create ranges of integer values, use :::.
```

# .R file

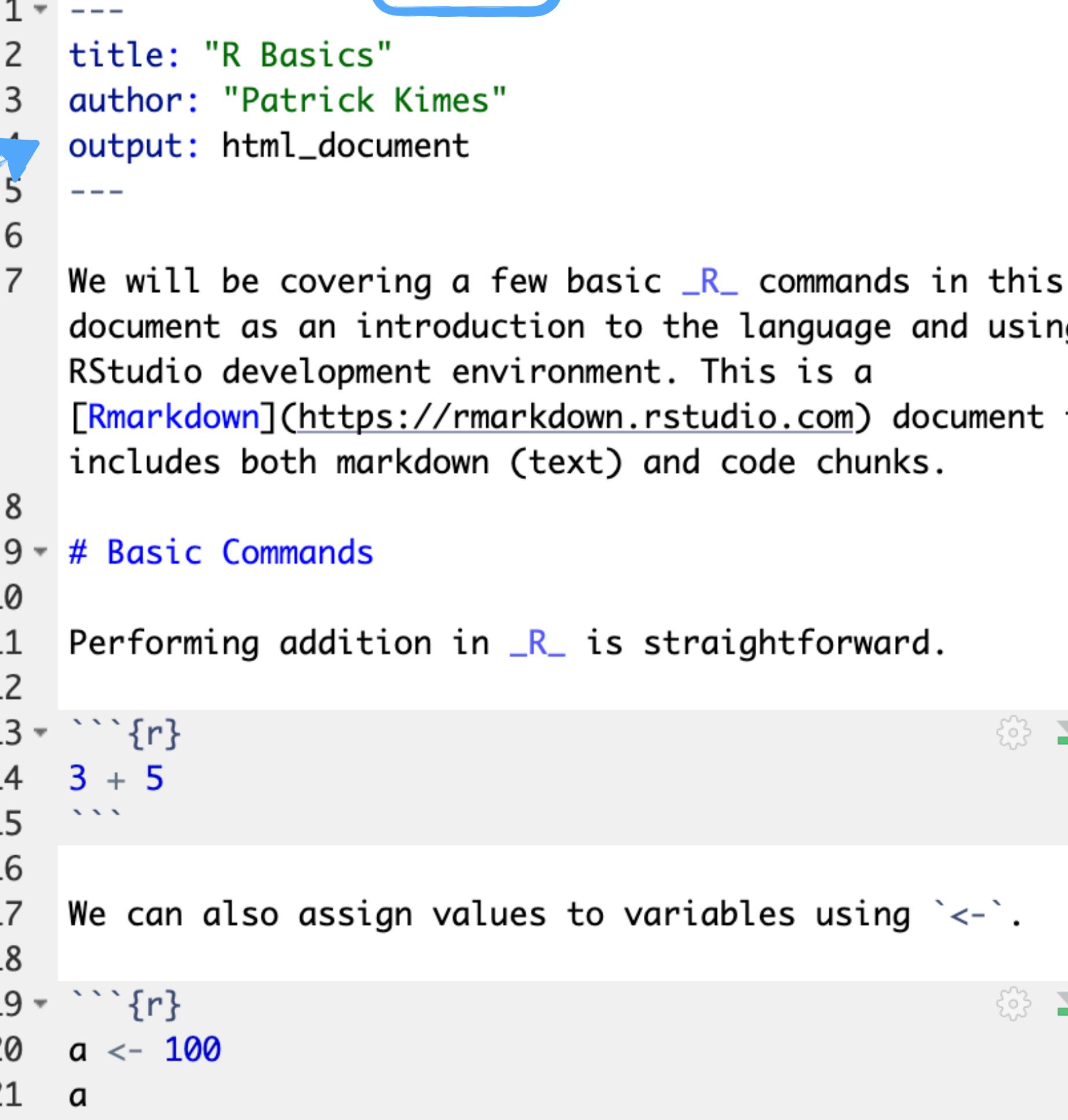
The screenshot shows the RStudio interface with a code editor containing an R script. The script includes comments like '## R Basics' and '## addition in R', and code such as '3 + 5'. A blue callout box highlights the text 'specified output format'.

```
1 ## R Basics
2 ## author: Patrick Kimes
3
4 ## addition in R
5 3 + 5
6
7 ## variable assignment
8 a <- 100
9 a
10
11 ## integer range
12 b <- 1:7
13 b
14
15 ## skipping sequence
16 b <- seq(1, 7, by = 2)
17 b
18
19 ## subsetting
20 b[2:3]
21
22 ## list out workspace
23 ls()
24
```

specified  
output format

**specified  
output format**

# .Rmd file



The screenshot shows the RStudio interface with the following details:

- Title Bar:** The file name is "dfci-introR.Rmd".
- Toolbar:** Includes icons for back, forward, file, ABC (checkmark), search, and a Knit button (highlighted with a blue box).
- Code Editor:** Displays R Markdown code. Lines 1-4 define the document's metadata. Line 7 contains explanatory text and a link to RMarkdown. Lines 9-11 show basic R code for addition. Lines 17 and 20 demonstrate variable assignment. Lines 24 and 25 show how to create integer ranges.
- Code Chunks:** Indicated by three backticks followed by '{r}' at the start of lines 13, 19, and 24.
- Run Buttons:** Located in the top right corner, with "Run" highlighted.

# .Rmd file

The screenshot shows the RStudio interface with the following details:

- Title Bar:** /cloud/project/dfci-introR.html
- Address Bar:** pkimes.rstudio.cloud/fb549fbb697541bcb3b1cff9c01a57b8/?view=rmarkdown
- Toolbar:** Publish, Find
- Code Editor:** dfci-introR.Rmd
- Content:**
  - # R Basics
  - Author:** Patrick Kimes
  - We will be covering a few basic *R* commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](#) document that includes both markdown (text) and code chunks.
  - ## Basic Commands
  - Performing addition in *R* is straightforward.

```
3 + 5
## [1] 8
```
  - We can also assign values to variables using `<-`.

```
a <- 100
a
## [1] 100
```
  - To create ranges of integer values, use `:`.

```
b <- 1:7
b
```

**dfci-introR.Rmd**

---  
title: "R Basics"  
author: "Patrick Kimes"  
output: html\_document  
---

We will be covering a few basic `R` commands in this document as an introduction to the language and using the RStudio development environment. This is a [\[Rmarkdown\]\(https://rmarkdown.rstudio.com\)](#) document that includes both markdown (text) and code chunks.

# Basic Commands

Performing addition in `R` is straightforward.

```
```{r}
3 + 5
```
We can also assign values to variables using `<-`.
```

We can also assign values to variables using `<-`.

```
```{r}
a <- 100
a
```
To create ranges of integer values, use `:`.
```

To create ranges of integer values, use `:`.

# R Basics

Patrick Kimes

We will be covering a few basic *R* commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](#) document that includes both markdown (text) and code chunks.

## Basic Commands

Performing addition in *R* is straightforward.

```
3 + 5
```

```
## [1] 8
```

We can also assign values to variables using `<-`.

```
a <- 100  
a
```

```
## [1] 100
```

To create ranges of integer values, use `:`.

```
b <- 1:7  
b
```

# .Rmd file

```
1 ---  
2 title: "R Basics"  
3 author: "Patrick Kimes"  
4 output: html_document  
---
```

```
7 We will be covering a few basic R commands in this document as an introduction to the language and using the RStudio development environment. This is a [Rmarkdown](https://rmarkdown.rstudio.com) document that includes both markdown (text) and code chunks.
```

```
8  
9 # Basic Commands
```

```
11 Performing addition in R is straightforward.
```

```
13 ````{r}  
14 3 + 5  
15 ````
```

```
17 We can also assign values to variables using `<-`.
```

```
19 ````{r}  
20 a <- 100  
21 a  
22 ````
```

```
24 To create ranges of integer values, use `:`.
```

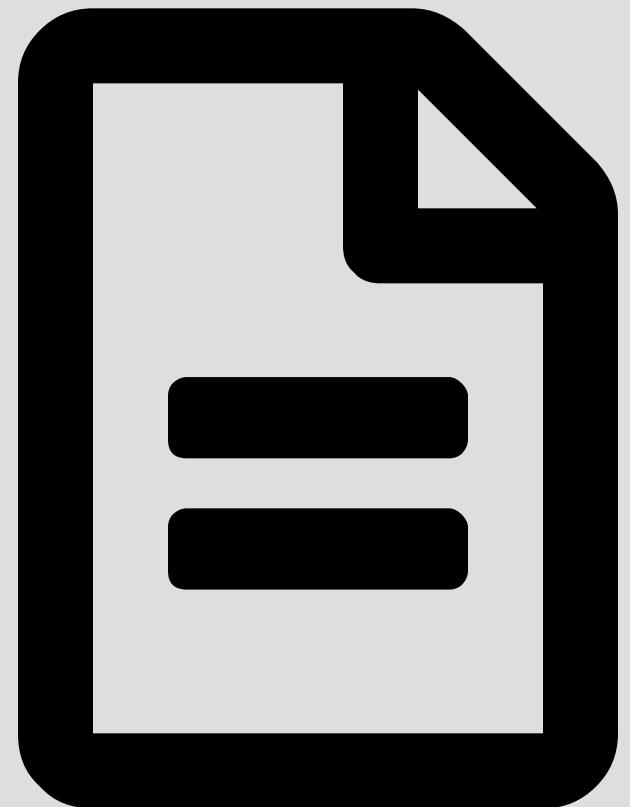
formatted text

R code

output!

rmarkdown

documentation, communication



**myfile.Rmd**

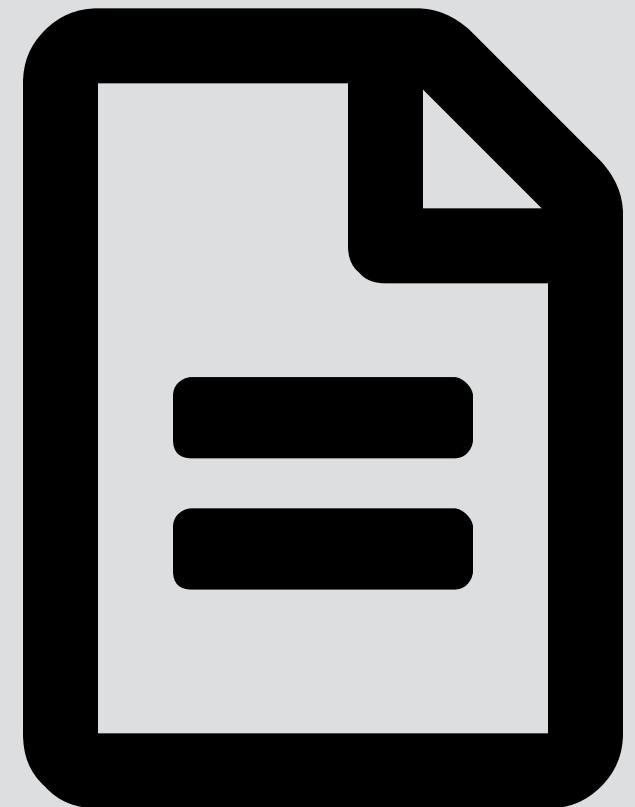
**markdown**

**+**

**R code chunks**

# rmarkdown

# documentation, communication



**myfile.Rmd**

markdown  
+  
R code chunks

execute  
R code

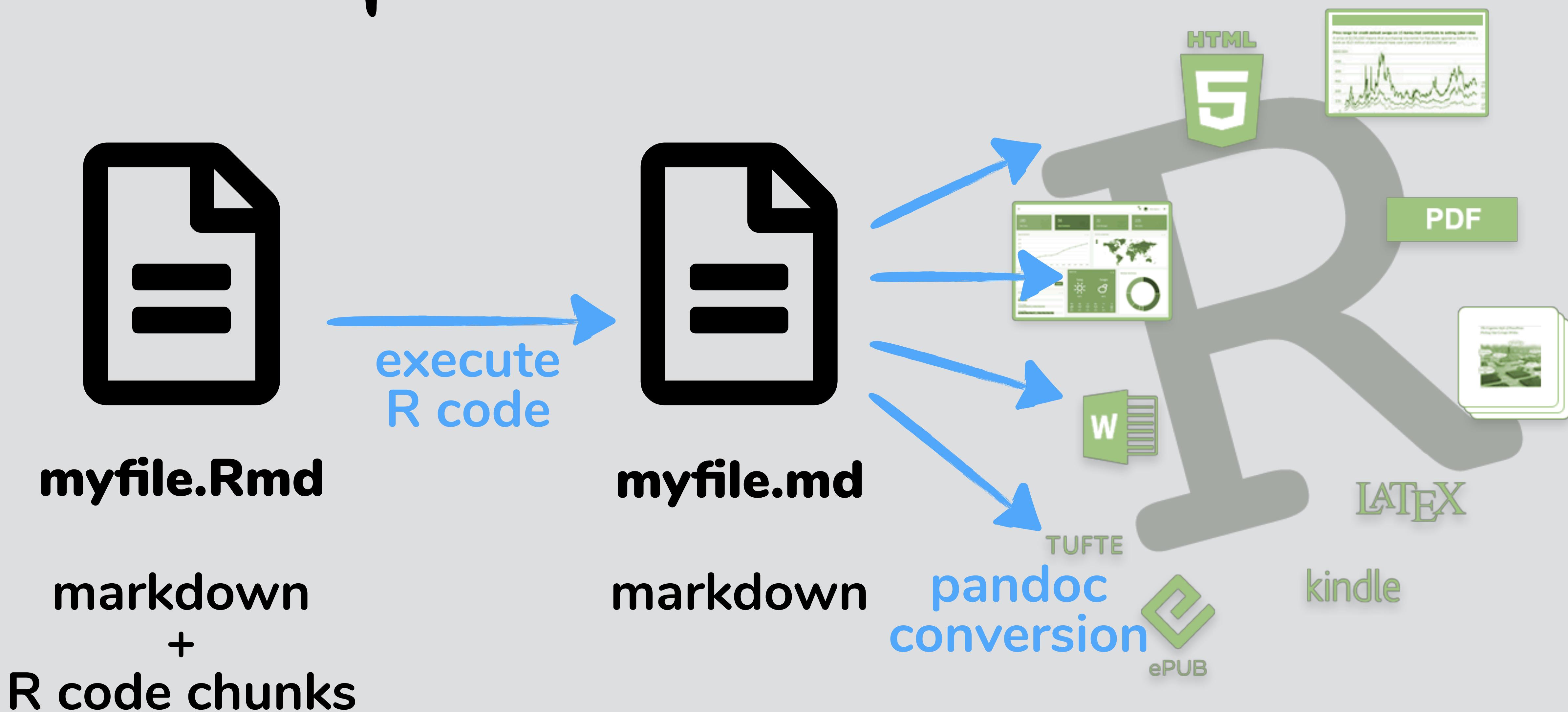


**myfile.md**

markdown

# rmarkdown

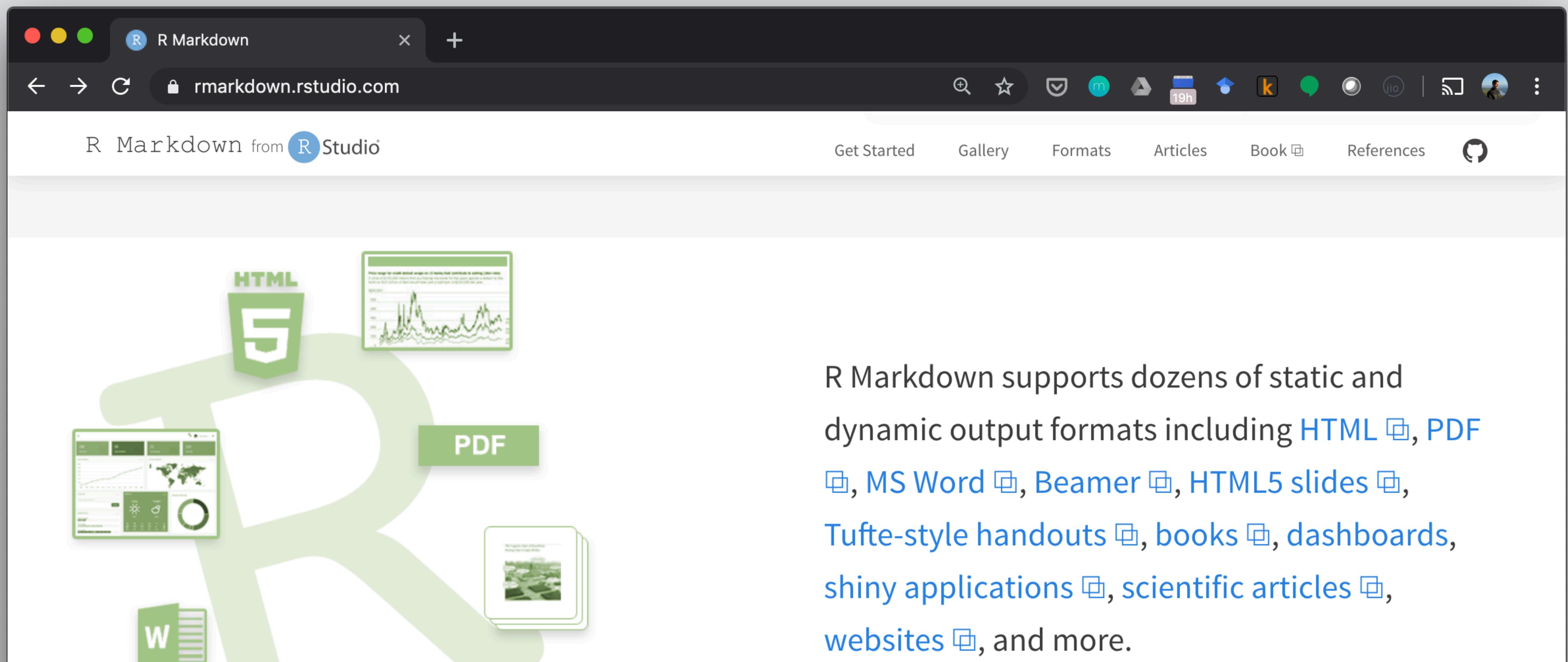
documentation, communication



# rmarkdown

# documentation, communication

[rmarkdown.rstudio.com](https://rmarkdown.rstudio.com)



The screenshot shows the homepage of rmarkdown.rstudio.com. The top navigation bar includes links for Get Started, Gallery, Formats, Articles, Book, and References. Below the navigation, there's a large graphic featuring a green silhouette of a person holding a smartphone. Overlaid on the silhouette are icons representing different output formats: an HTML5 logo, a PDF icon, a Beamer presentation icon, a shiny dashboard icon, and a scientific article icon. To the right of the graphic, a block of text describes the versatility of R Markdown.

R Markdown supports dozens of static and dynamic output formats including [HTML](#), [PDF](#), [MS Word](#), [Beamer](#), [HTML5 slides](#), [Tufte-style handouts](#), [books](#), [dashboards](#), [shiny applications](#), [scientific articles](#), [websites](#), and more.

let's give it a try!

**what did we  
(hopefully) cover?**

create a new Rmd file  
writing simple markdown  
creating code chunks  
executing code  
knitting documents

# **some pieces in the modern (R) data scientist's toolbox**

**rmarkdown**

documentation, communication

**tidyverse**

data manipulation, visualization

**shiny**

web application framework

**[bioconductor]**

community of genomics packages

# some pieces in the modern (R) data scientist's toolbox

rmardown

documentation, communication

tidyverse

data manipulation, visualization

shiny

web application framework

[bioconductor]

community of genomics packages

# tidyverse | data manipulation, visualization

The screenshot shows the tidyverse.org homepage. At the top, there's a navigation bar with links for Packages, Articles, Learn, Help, and Contribute. Below the navigation, there's a large graphic featuring six hexagonal icons representing different packages: dplyr (orange, with a pliers icon), ggplot2 (grey, with a network graph icon), readr (blue, with a document icon), purrr (black, with a jagged line icon), tibble (dark blue, with a grid icon), and tidyr (orange, with a circular arrow icon). To the right of the graphic, the text reads "R packages for data science" and "The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures." Below this, there's a section titled "Install the complete tidyverse with:" followed by a code snippet: `install.packages("tidyverse")`. At the bottom left, it says "The tidyverse is proudly supported by R Studio". On the bottom right, there's a large "tidyverse.org" logo with social media icons.

Tidyverse

Packages Articles Learn Help Contribute

R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

The tidyverse is proudly supported by R Studio

tidyverse.org

# some pieces in the modern (R) data scientist's toolbox

rmardown

documentation, communication

tidyverse

data manipulation, visualization

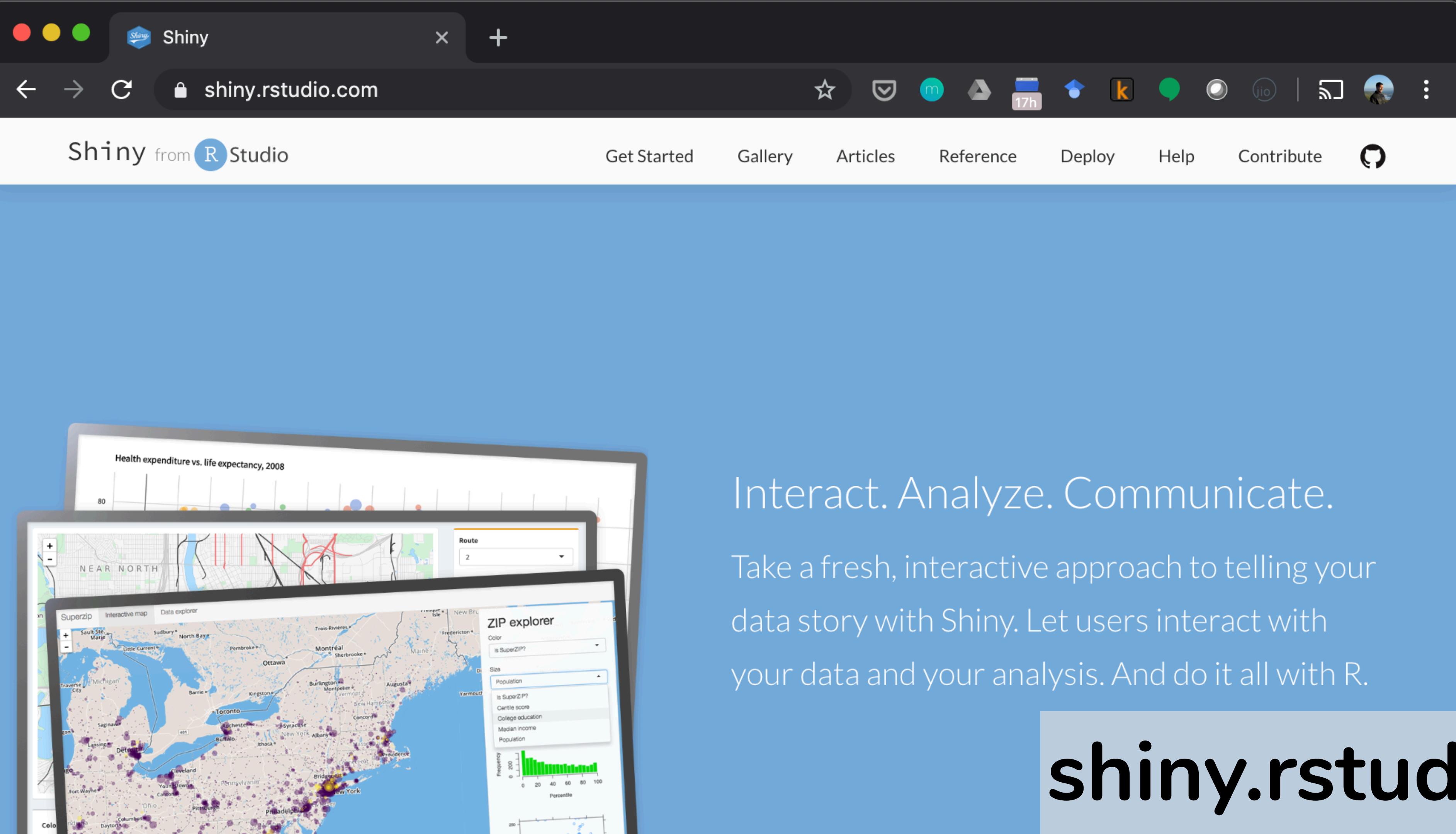
shiny

web application framework

[bioconductor]

community of genomics packages

# shiny | web application framework



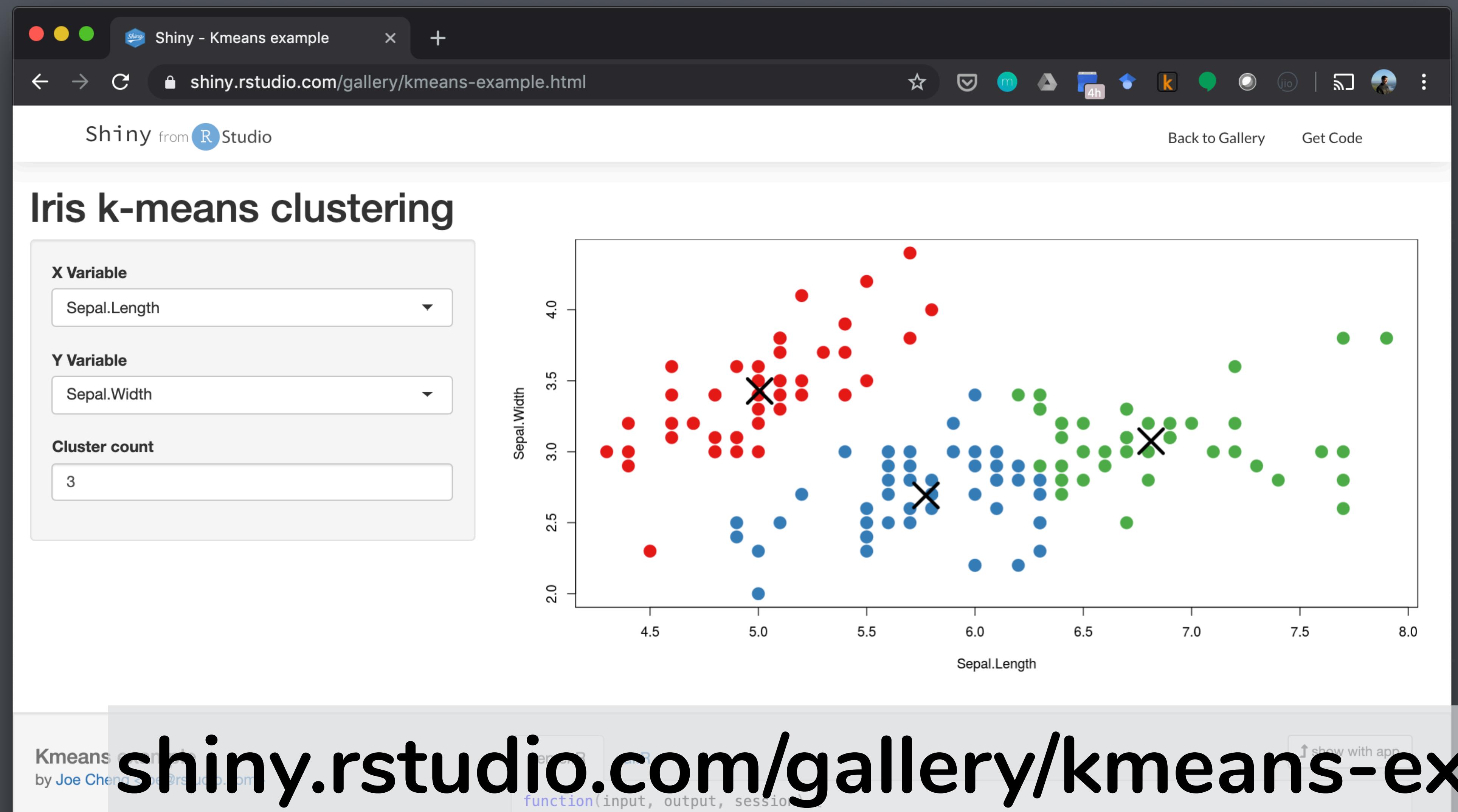
The screenshot shows the official Shiny website at [shiny.rstudio.com](https://shiny.rstudio.com). The top navigation bar includes links for "Get Started", "Gallery", "Articles", "Reference", "Deploy", "Help", "Contribute", and a GitHub icon. Below the navigation is a large blue section featuring three overlapping screenshots of Shiny applications: a scatter plot titled "Health expenditure vs. life expectancy, 2008", a map titled "ZIP explorer" showing SuperZIPs across North America, and a histogram titled "Frequency" showing population distribution by percentile.

Interact. Analyze. Communicate.

Take a fresh, interactive approach to telling your data story with Shiny. Let users interact with your data and your analysis. And do it all with R.

[shiny.rstudio.com](https://shiny.rstudio.com)

# shiny | web application framework



# some pieces in the modern (R) data scientist's toolbox

rstudio

documentation, communication

tidyverse

data manipulation, visualization

shiny

web application framework

[bioconductor]

community of genomics packages

# bioconductor | community of genomics packages

The screenshot shows the Bioconductor website ([bioconductor.org](http://bioconductor.org)) displayed in a web browser. The header features the Bioconductor logo and navigation links for Home, Install, Help, Developers, and About. A search bar is also present. The main content area includes sections for 'About Bioconductor', 'News', 'Install', 'Learn', 'Use', and 'Develop'.

**About Bioconductor**

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, and an active user community. Bioconductor is also available as an [AMI](#) (Amazon Machine Image) and a series of [Docker](#) images.

**News**

- Bioconductor [3.9](#) is available.
- Core team **job opportunities** for scientific programmer / analyst and senior programmer / analyst! contact Martin.Morgan at RoswellPark.org
- Bioconductor [F1000 Research Channel](#) available.

**Install »**

- Discover [1741 software packages](#) available in Bioconductor release 3.9.

Get started with Bioconductor

- [Install Bioconductor](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

**Learn »**

Master Bioconductor tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

**Use »**

Create bioinformatic solutions with Bioconductor

- [Software](#), [Annotation](#), and [Experiment](#) packages
- [Docker](#) and [Amazon](#) machine images
- [Latest release announcement](#)

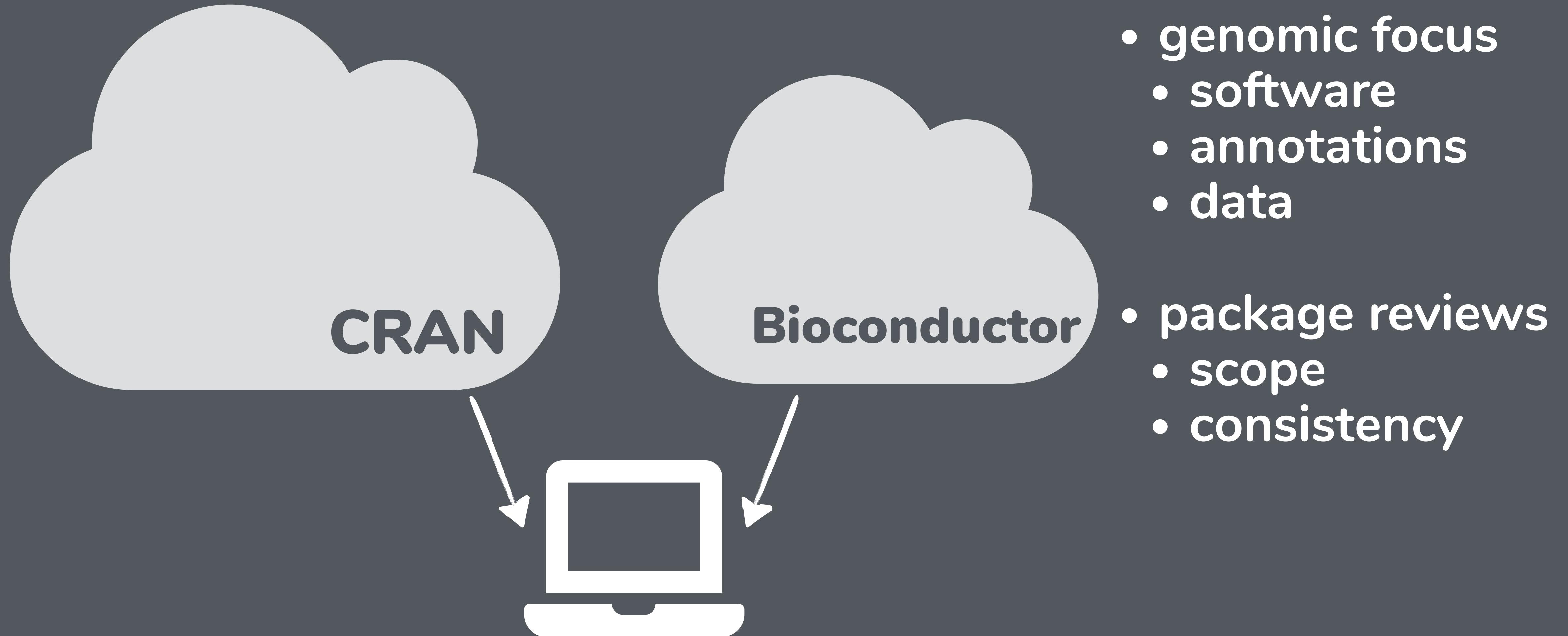
**Develop »**

Contribute to Bioconductor

- [Developer resources](#)
- [Use Bioc 'dev'](#)
- 'Devel' package
- [Package guidelines](#)
- [New package submission](#)

bioconductor.org

# bioconductor | community of genomics packages





**awesome!**

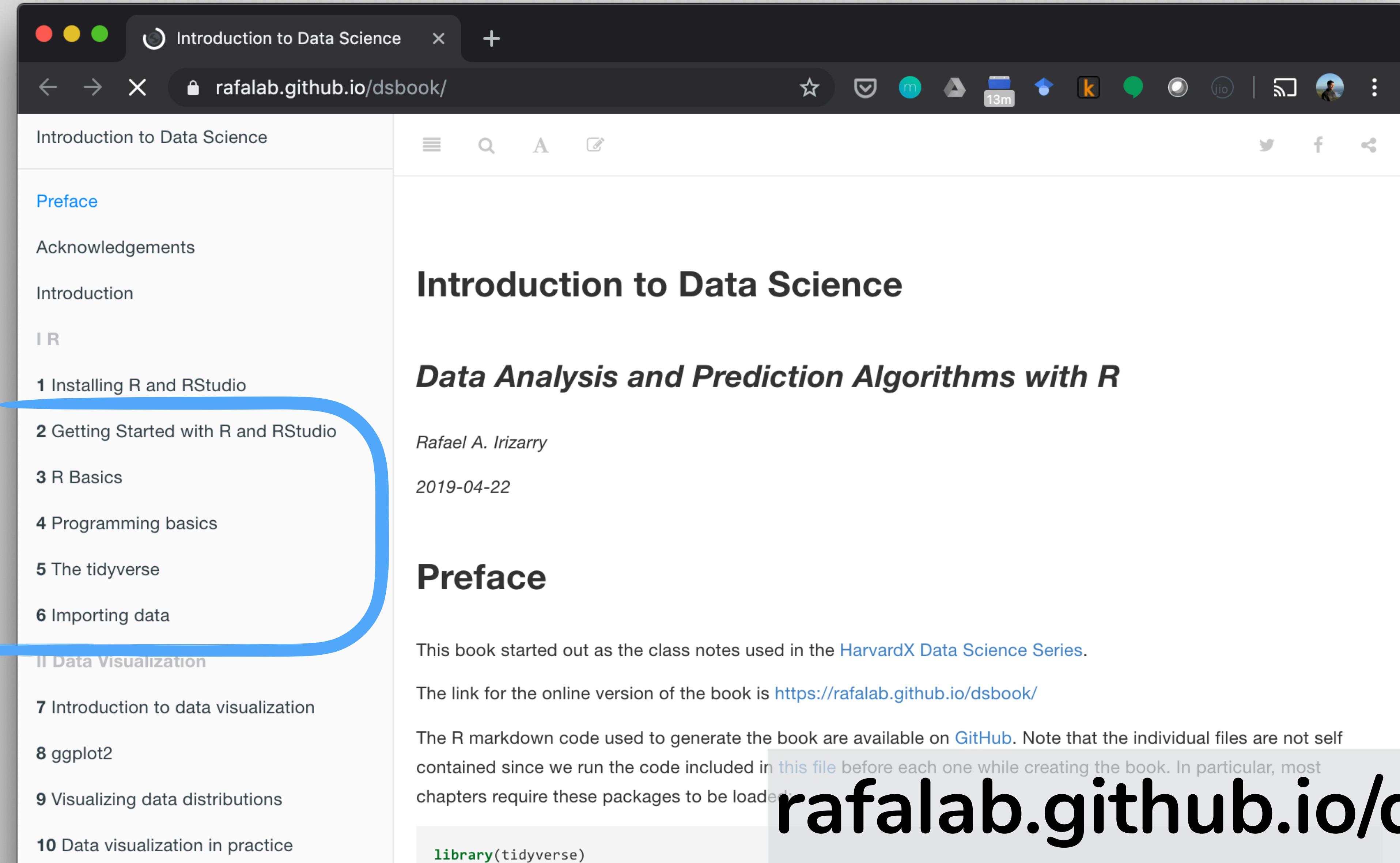


**where do we go  
from here?**

**where do we go  
from here?**

**wait, I'm lost**

# introduction to data science



Introduction to Data Science

rafalab.github.io/dsbook/

Preface

Acknowledgements

Introduction

I R

- 1 Installing R and RStudio
- 2 Getting Started with R and RStudio
- 3 R Basics
- 4 Programming basics
- 5 The tidyverse
- 6 Importing data

II Data Visualization

- 7 Introduction to data visualization
- 8 ggplot2
- 9 Visualizing data distributions
- 10 Data visualization in practice

## Introduction to Data Science

### ***Data Analysis and Prediction Algorithms with R***

Rafael A. Irizarry

2019-04-22

## Preface

This book started out as the class notes used in the [HarvardX Data Science Series](#).

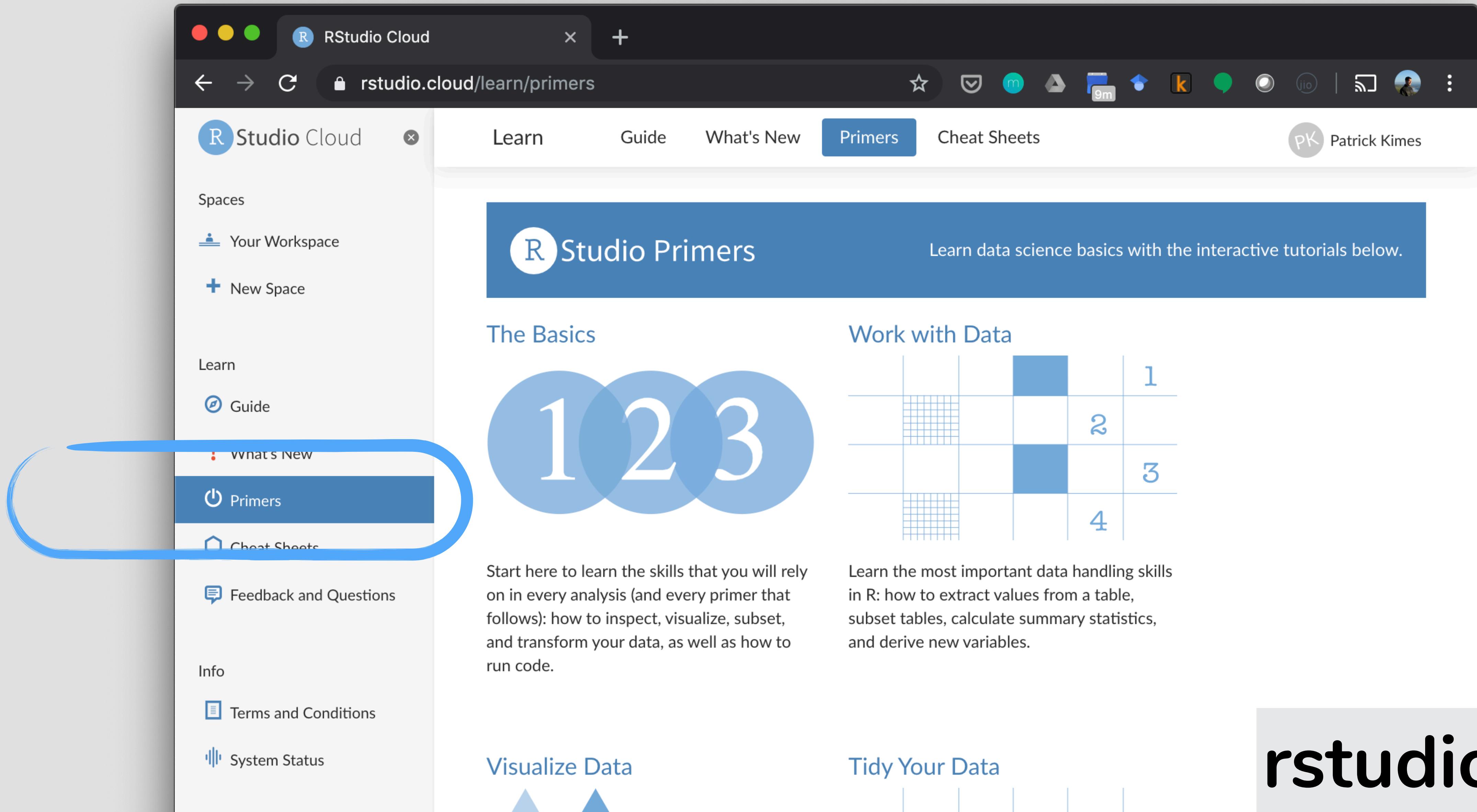
The link for the online version of the book is <https://rafalab.github.io/dsbook/>

The R markdown code used to generate the book are available on [GitHub](#). Note that the individual files are not self contained since we run the code included in [this file](#) before each one while creating the book. In particular, most chapters require these packages to be loaded:

```
library(tidyverse)
```

**rafalab.github.io/dsbook**

# RStudio Cloud tutorials



A screenshot of a web browser showing the RStudio Cloud 'Primers' page. The browser window has a dark theme with a light gray header bar. The title bar says 'RStudio Cloud' and the address bar shows 'rstudio.cloud/learn/primers'. The main content area features a large blue banner with the text 'Studio Primers' and a subtext 'Learn data science basics with the interactive tutorials below.' Below the banner, there are two main sections: 'The Basics' and 'Work with Data'. The 'The Basics' section contains three large blue circles with the numbers 1, 2, and 3 inside them. The 'Work with Data' section features a 4x4 grid with colored squares. A blue hand-drawn style circle highlights the 'Primers' menu item in the left sidebar.

- Spaces
- Your Workspace
- New Space
- Learn
  - Guide
  - What's New
  - Primers
  - Cheat Sheets
- Feedback and Questions
- Info
  - Terms and Conditions
  - System Status

**Studio Primers**

Learn data science basics with the interactive tutorials below.

### The Basics

Start here to learn the skills that you will rely on in every analysis (and every primer that follows): how to inspect, visualize, subset, and transform your data, as well as how to run code.

### Work with Data

Learn the most important data handling skills in R: how to extract values from a table, subset tables, calculate summary statistics, and derive new variables.

### Visualize Data

### Tidy Your Data

rstudio.cloud

# R for data science

The screenshot shows a web browser window with the title bar "R for Data Science" and the URL "r4ds.had.co.nz". The page content is the "Welcome" chapter of the book. On the left, there's a sidebar with chapters grouped by section: I Explore (1 Introduction, 2 Introduction), II Wrangle (3 Data visualisation, 4 Workflow: basics, 5 Data transformation, 6 Workflow: scripts, 7 Exploratory Data Analysis, 8 Workflow: projects), and III Tidy (9 Introduction, 10 Tibbles, 11 Data import, 12 Tidy data, 13 Relational data). The main content area features the title "R for Data Science" and authors "Garrett Grolemund" and "Hadley Wickham". Below this is a large "Welcome" heading. A paragraph describes the book's purpose: "This is the website for “R for Data Science”. This book will teach you how to do data science with R: You’ll learn how to get your data into R, get it into the most useful structure, transform it, visualise it and model it. In this book, you will find a practicum of skills for data science. Just as a chemist learns how to clean test tubes and stock a lab, you’ll learn how to clean data and draw plots—and many other things besides. These are the skills that allow data science to happen, and here you will find the best practices for doing each of these things with R. You’ll learn how to use the grammar of graphics, literate programming, and reproducible research to save time. You’ll also learn how to manage cognitive resources to facilitate discoveries when wrangling, visualising, and exploring data." To the right of the text is a graphic of the book cover, which features a green parrot-like bird and the title "R for Data Science" in white.

**R for Data Science**

*Garrett Grolemund*  
*Hadley Wickham*

## Welcome

This is the website for “**R for Data Science**”. This book will teach you how to do data science with R: You’ll learn how to get your data into R, get it into the most useful structure, transform it, visualise it and model it. In this book, you will find a practicum of skills for data science. Just as a chemist learns how to clean test tubes and stock a lab, you’ll learn how to clean data and draw plots—and many other things besides. These are the skills that allow data science to happen, and here you will find the best practices for doing each of these things with R. You’ll learn how to use the grammar of graphics, literate programming, and reproducible research to save time. You’ll also learn how to manage cognitive resources to facilitate discoveries when wrangling, visualising, and exploring data.



**O'REILLY**

**R for Data Science**

VISUALIZE, MODEL, TRANSFORM, TIDY, AND PORT DATA

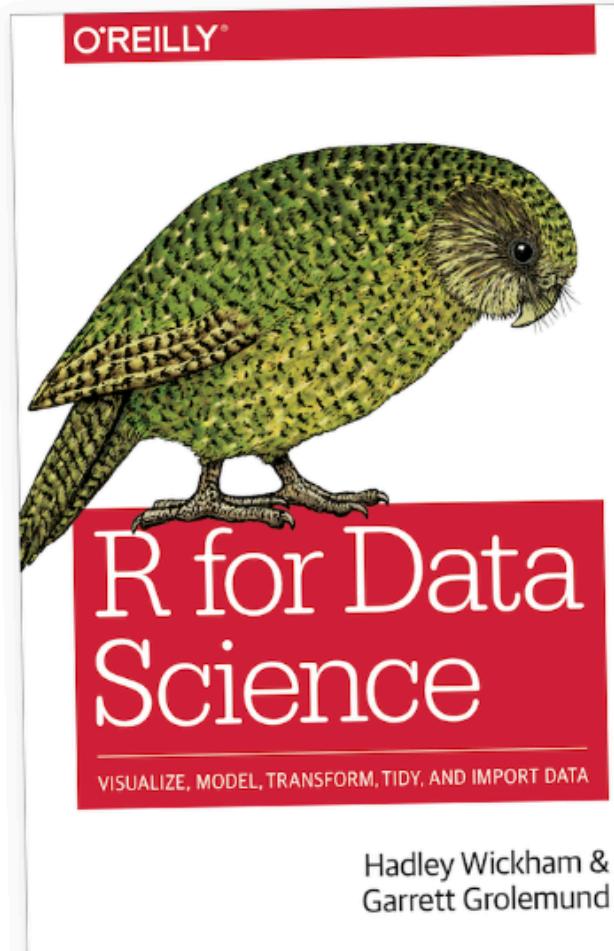
Hadley Wickham & Garrett Grolemund

r4ds.had.co.nz

# learn the tidyverse

The screenshot shows a web browser window with the title "Learn the tidyverse - Tidyverse". The URL in the address bar is "tidyverse.org/learn/". The page content includes a header with "Tidyverse" and navigation links for "Packages", "Articles", "Learn" (which is highlighted), "Help", and "Contribute". Below the header, there's a section titled "Learn the tidyverse" featuring a book cover for "R for Data Science" by Hadley Wickham & Garrett Grolemund. To the right of the book cover, there's a section titled "R for data science" with a description of the book and its benefits. Further down, there's a "Books" section with a list item about "Statistical Inference via Data Science: A ModernDive into R and the tidyverse". On the right side of the page, there's a "Upcoming events" sidebar with two entries: "Building Tidy Tools Workshop" in Atlanta, GA on Oct 14-15, and "rstudio::conf 2020" in San Francisco, CA on Jan 27-30.

## Learn the tidyverse



**R for data science**

The best place to start learning the tidyverse is [R for Data Science](#) (R4DS for short), an O'Reilly book written by Hadley Wickham and Garrett Grolemund. It's designed to take you from knowing nothing about R or the tidyverse to having all the basic tools of data science at your fingertips. You can read it online for free, or [buy a physical copy](#).

We highly recommend pairing R4DS with the [RStudio cheatsheets](#). These cheatsheets have been carefully designed to pack a lot of information into a small amount of space. You can keep them handy at your desk and quickly jog your memory when you get stuck. Most of the cheatsheets have been translated into multiple languages.

## Books

- [Statistical Inference via Data Science: A ModernDive into R and the tidyverse](#) by Chester Ismay and Albert Y. Kim.  
"Help! I'm new to R and RStudio and I need to learn them! What do I do?" If you're asking yourself this, this book is for you.

## Upcoming events

**Building Tidy Tools Workshop**  
Atlanta, GA  
Oct 14-15

You should take this workshop if you have experience programming in R and want to learn how to tackle larger scale problems. The class is taught by Hadley Wickham, Chief Scientist at RStudio.

**rstudio::conf 2020**  
San Francisco, CA  
Jan 27-30

rstudio::conf 2020 covers all things RStudio, including workshops to teach you the tidyverse, and talks to show you the latest and greatest.

[tidyverse.org/learn/](https://tidyverse.org/learn/)

# advanced R

The screenshot shows a Mac OS X desktop with a dark-themed browser window open to [adv-r.hadley.nz](http://adv-r.hadley.nz). The title bar says "Advanced R". The page content is the "Welcome" page of the book "Advanced R" by Hadley Wickham. The sidebar on the left lists chapters from 1 to 8, along with sections for Foundations, Names and values, Vectors, Subsetting, Control flow, Functions, Environments, Conditions, and Functional programming. A Creative Commons BY-NC-SA license logo is at the bottom. To the right of the main content is a thumbnail image of the book cover.

**Advanced R**

Hadley Wickham

## Welcome

This is the website for 2nd edition of “[Advanced R](#)”, a book in Chapman & Hall’s R Series. The book is designed primarily for R users who want to improve their programming skills and understanding of the language. It should also be useful for programmers coming to R from other languages, as help you to understand why R works the way it does.

If you’re looking for the electronic version of the 1st edition, you can find it online at <http://adv-r.had.co.nz/>.

## License

This work, as a whole, is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0](#)

adv-r.hadley.nz

# biomedical data science open online training

The screenshot shows a web browser window with the title bar "harvardx". The address bar indicates the page is "Not Secure" and shows the URL "rafalab.github.io/pages/harvardx.html". The browser interface includes standard controls like back, forward, and search, along with various icons for extensions or add-ons.

The main content area of the website has a header with the logo "rafalab" and navigation links for Home, About, People, Research, Publications, Software, and Teaching. Below this, a large section header reads "HarvardX Biomedical Data Science Open Online Training".

A text block explains that in 2014 funding was received from the NIH BD2K initiative to develop MOOCs for biomedical data science, divided into three series: "Data Analysis for the Life Sciences", "Genomics Data Analysis", and "Using Python for Research".

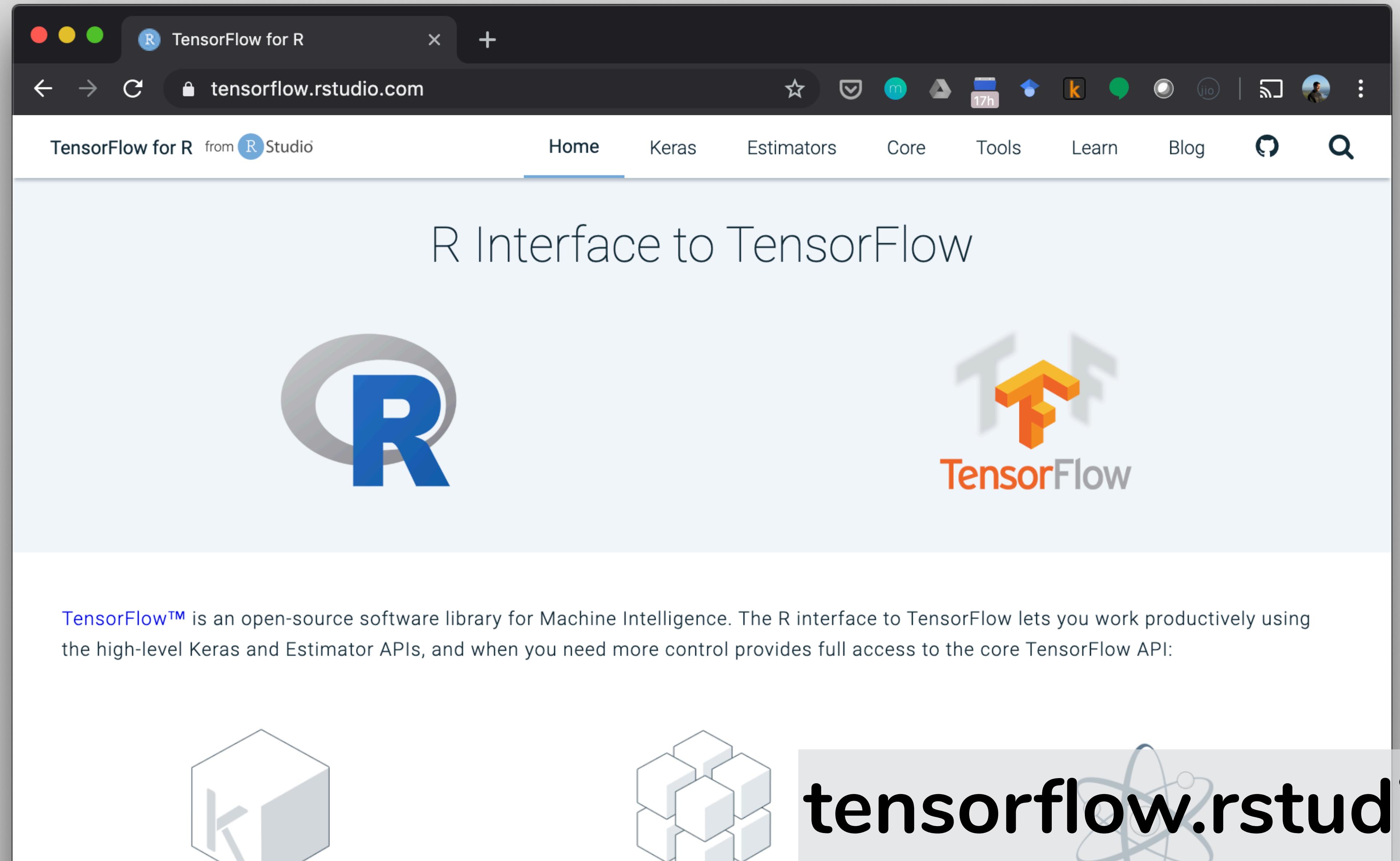
Links are provided for course material, including R markdown documents and video lectures. A note states that access requires EdX login.

Below this, a section titled "Data Analysis for the Life Sciences Series" is shown, with a sub-section "Statistics and R". A table lists "Lecture Title", "Time", "Video", "Material", and "Course" for the "Week 1: R" lecture, which is "Getting Started with R".

At the bottom of the page, there is a footer with links to "GitHub" and "rafalab.github.io/pages/harvardx".

rafalab.github.io/pages/harvardx

# ... and because it's 2019, deep learning



The screenshot shows a web browser window with the title bar "TensorFlow for R". The address bar contains "tensorflow.rstudio.com". The page itself is titled "TensorFlow for R from R Studio" and has a navigation bar with links for Home, Keras, Estimators, Core, Tools, Learn, and Blog. The main content area features the R logo and the TensorFlow logo. Below them, a text block states: "TensorFlow™ is an open-source software library for Machine Intelligence. The R interface to TensorFlow lets you work productively using the high-level Keras and Estimator APIs, and when you need more control provides full access to the core TensorFlow API:". At the bottom, there are icons for Keras and Estimators, and a large "tensorflow.rstudio.com" watermark.

R Interface to TensorFlow

TensorFlow™ is an open-source software library for Machine Intelligence. The R interface to TensorFlow lets you work productively using the high-level Keras and Estimator APIs, and when you need more control provides full access to the core TensorFlow API:

**tensorflow.rstudio.com**



wait, I'm still lost

questions?