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	**Introduction to R and RStudio**
October 22, 2019	Data visualization with ggplot2
November 12, 2019	Data visualization principles and plots to avoid
December 10, 2019	Design of Clinical Trials Basics
January 21, 2020	Correlation: you are probably using it wrong
February 11, 2020	How to detect and deal with batch effects
March 17, 2020	Brief introduction to machine learning
April 14, 2020	Culprits of the reproducibility crisis: multiple testing and p-hacking
May 19, 2020	Experimental Design: How many size and should I pool?
June 16, 2020	Detecting differentially expressed genes with RNA-seq

R ***and*** RStudio**?**

R ***and*** RStudio?

R | **programming language**

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

RStudio®

R ***and*** RStudio**?**

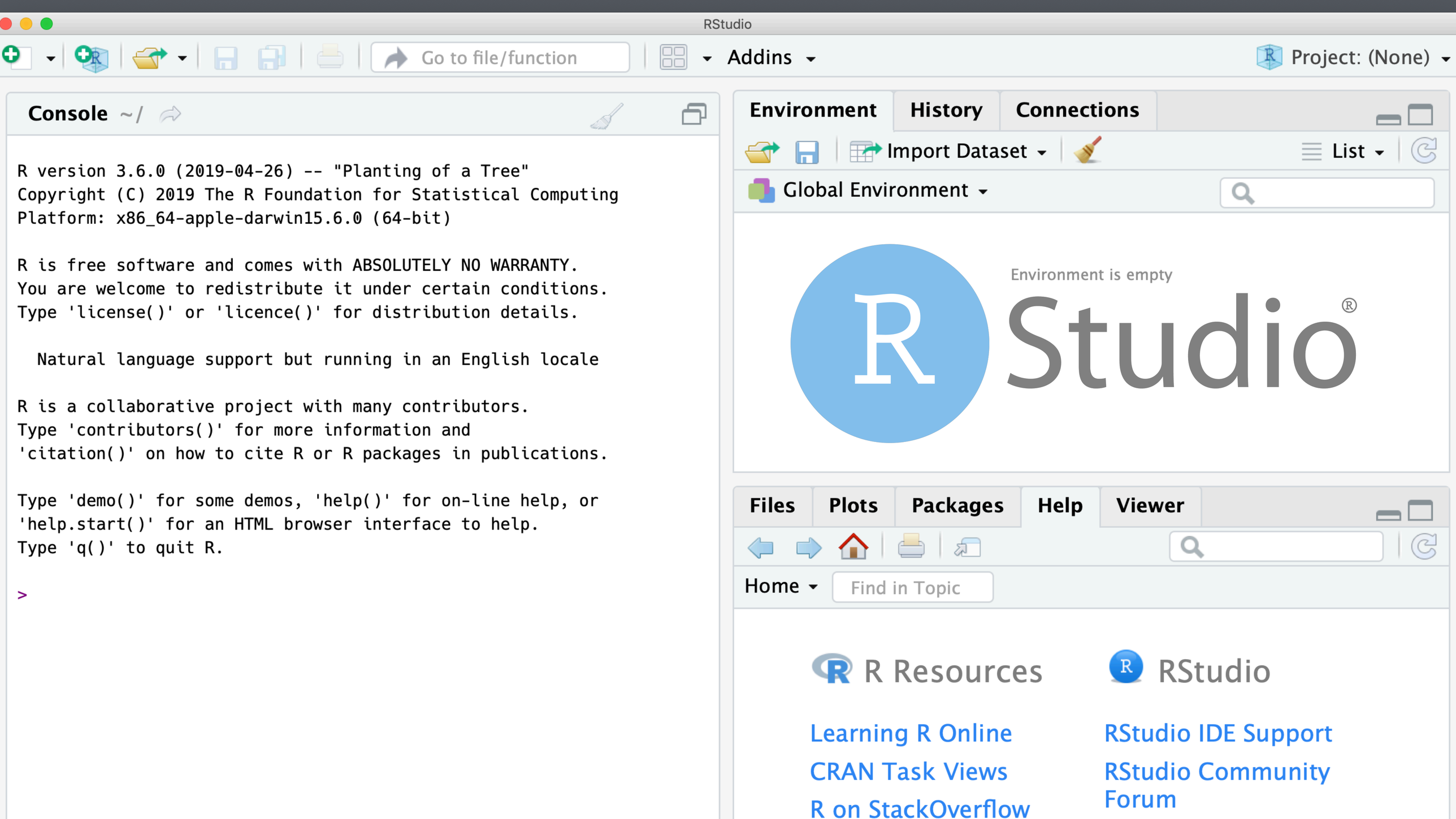
RStudio | **software to make data analysis with R easier**

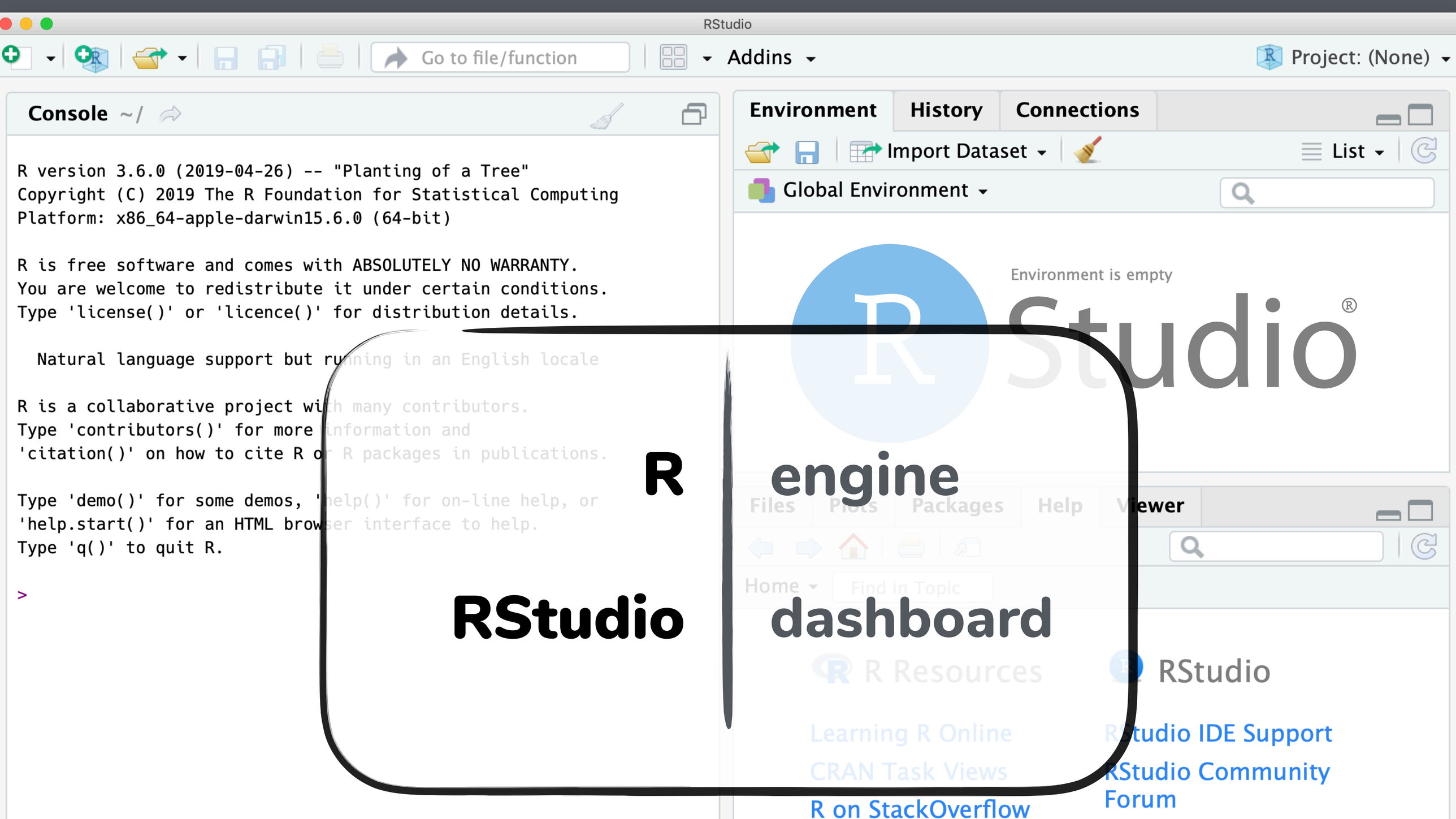
RStudio®

R *and* RStudio?

R | programming language

RStudio | software to make data analysis with R easier


RStudio
Go to file/function
Addins
Project: (None)
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
List
Global Environment
Environment is empty
RStudio
Files
Plots
Packages
Help
Viewer
Home
Find in Topic
R Resources
RStudio
Learning R Online
CRAN Task Views
R on StackOverflow
RStudio IDE Support
RStudio Community Forum


RStudio
Go to file/function
Addins
Project: (None)
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
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Find in Topic
R Resources
Learning R Online
CRAN Task Views
R on StackOverflow
RStudio
RStudio IDE Support
RStudio Community Forum
R
engine
RStudio
dashboard

why
R and RStudio?

why
R and RStudio?

what about SAS?

why
R and RStudio?

what about SAS?

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

why
R and RStudio?

what about Python?

why
R and RStudio?

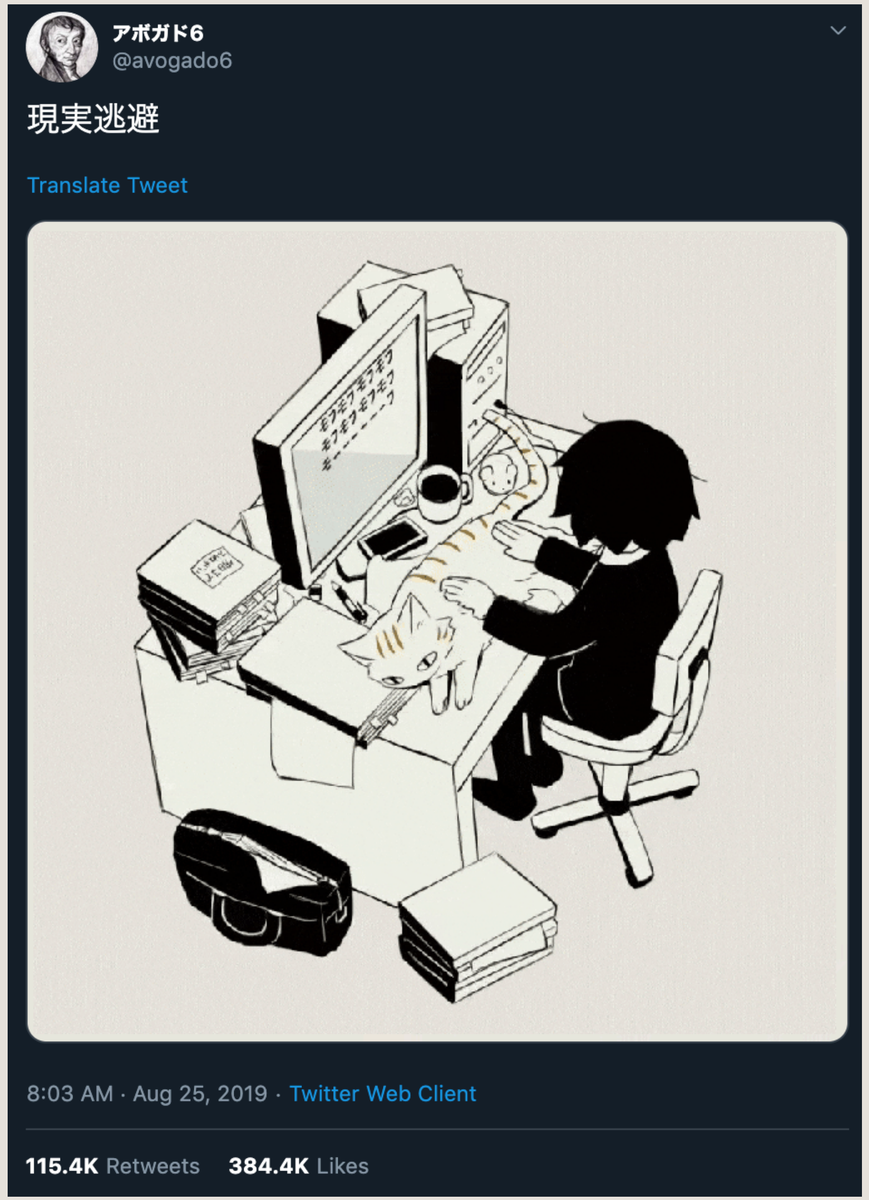
what about Python?

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

why
R and RStudio?

why
R and RStudio?

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



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

and that's ***fun!***

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


RStudio®
who already has
R / RStudio installed?

how to install R and RStudio

How to draw an Owl.

"A fun and creative guide for beginners"

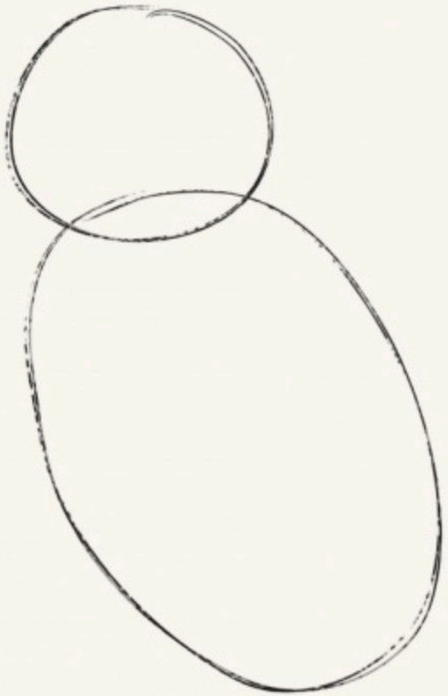
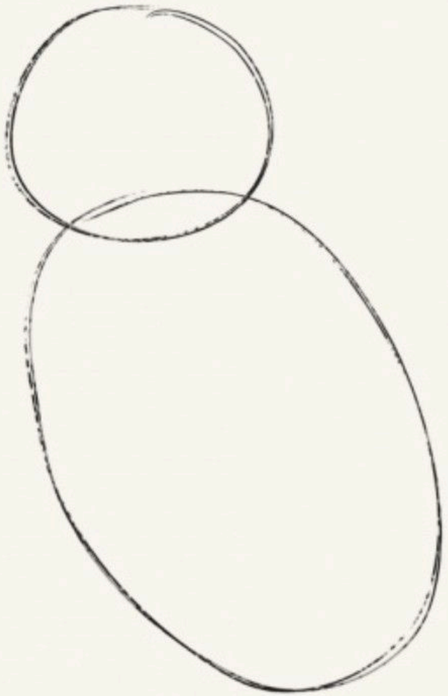


Fig 1. Draw two circles

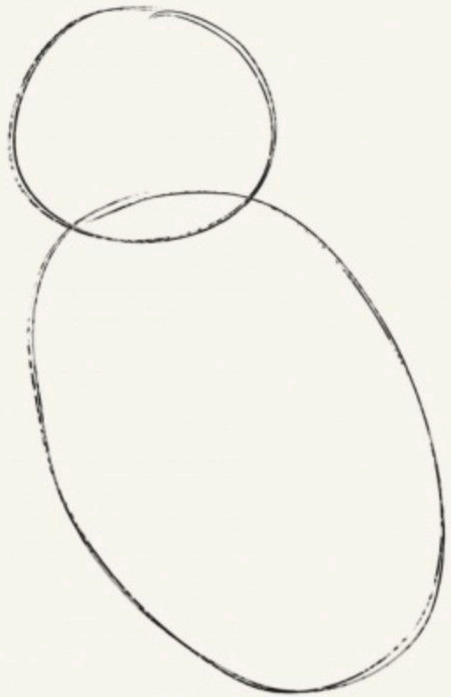
Fig 2. Draw the rest of the damn Owl

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**

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

a much better guide:

rafalab.github.io/dsbook/installing-r-rstudio

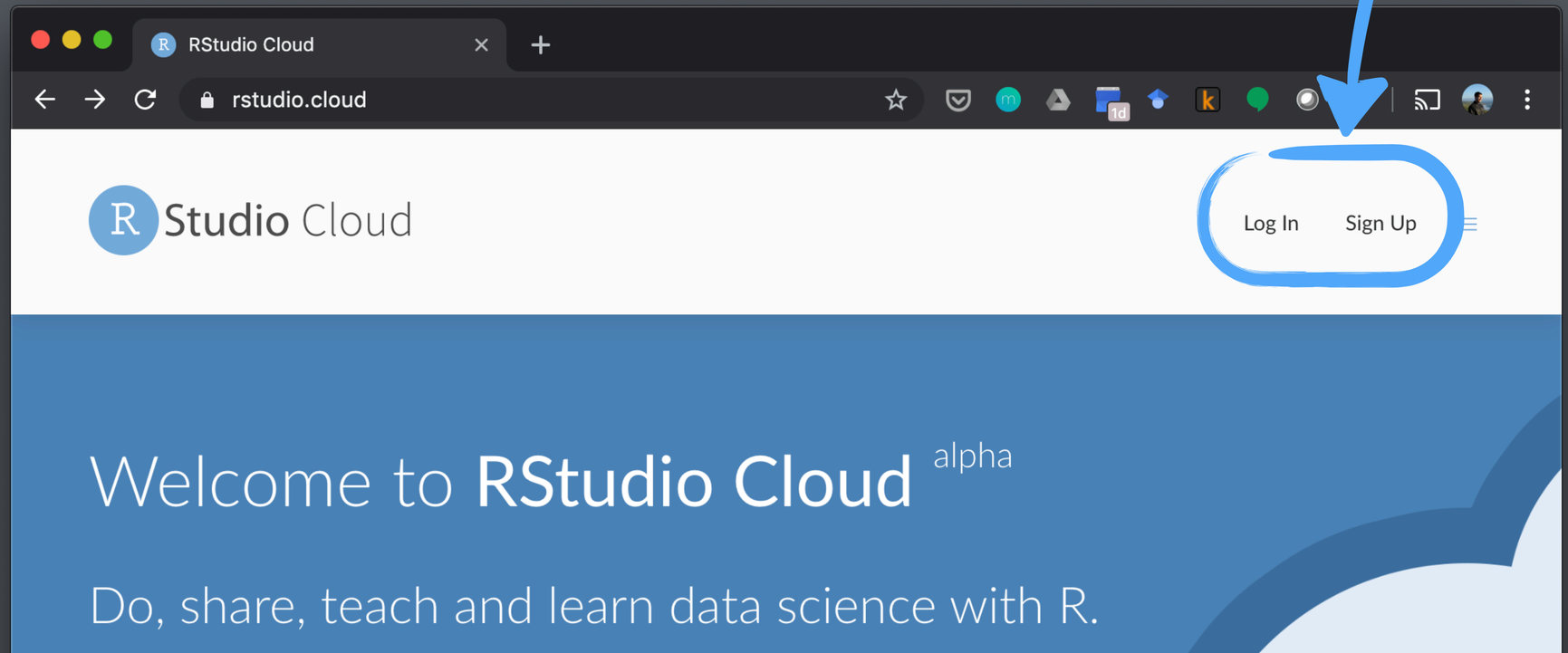
lucky us!
we have a workaround!

https://rstudio.cloud

lucky us!
we have a workaround!

do this

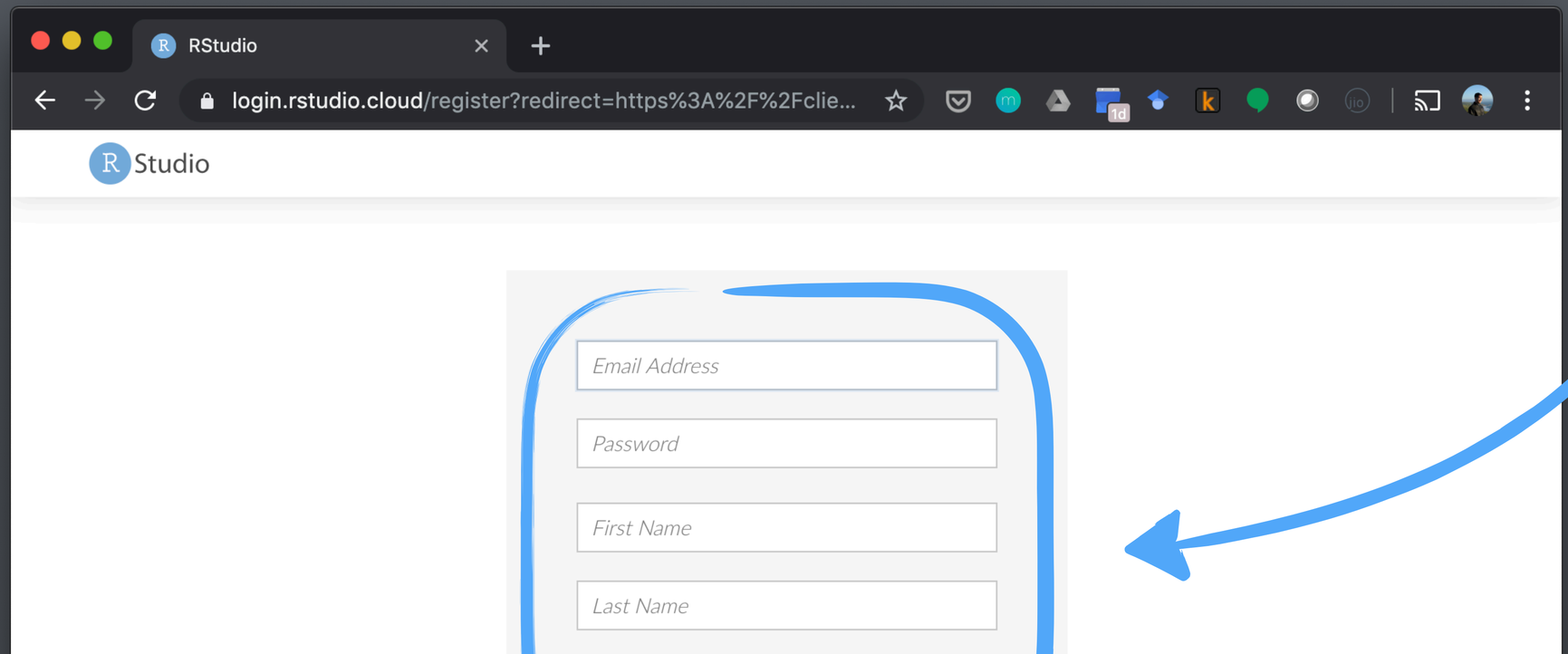
https://rstudio.cloud



lucky us!
we have a workaround!

https://rstudio.cloud

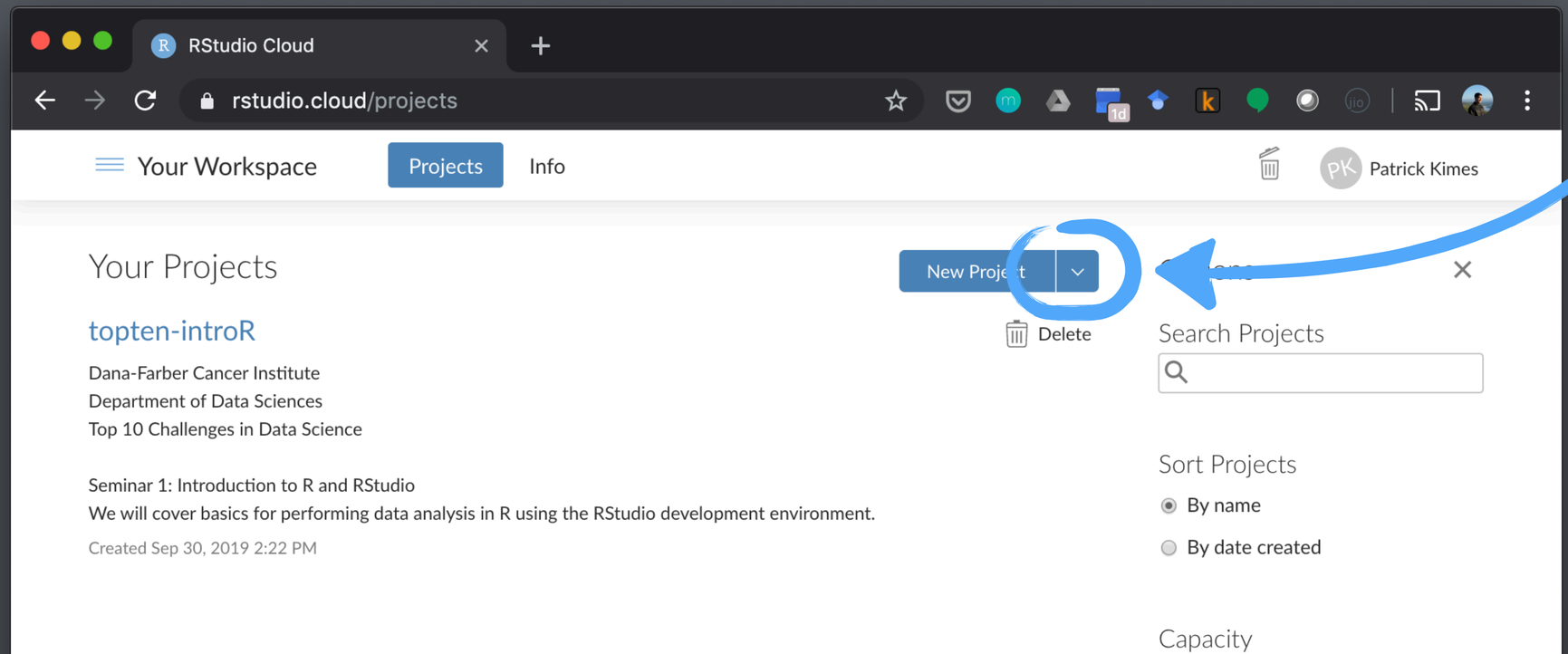
this too



lucky us!
we have a workaround!

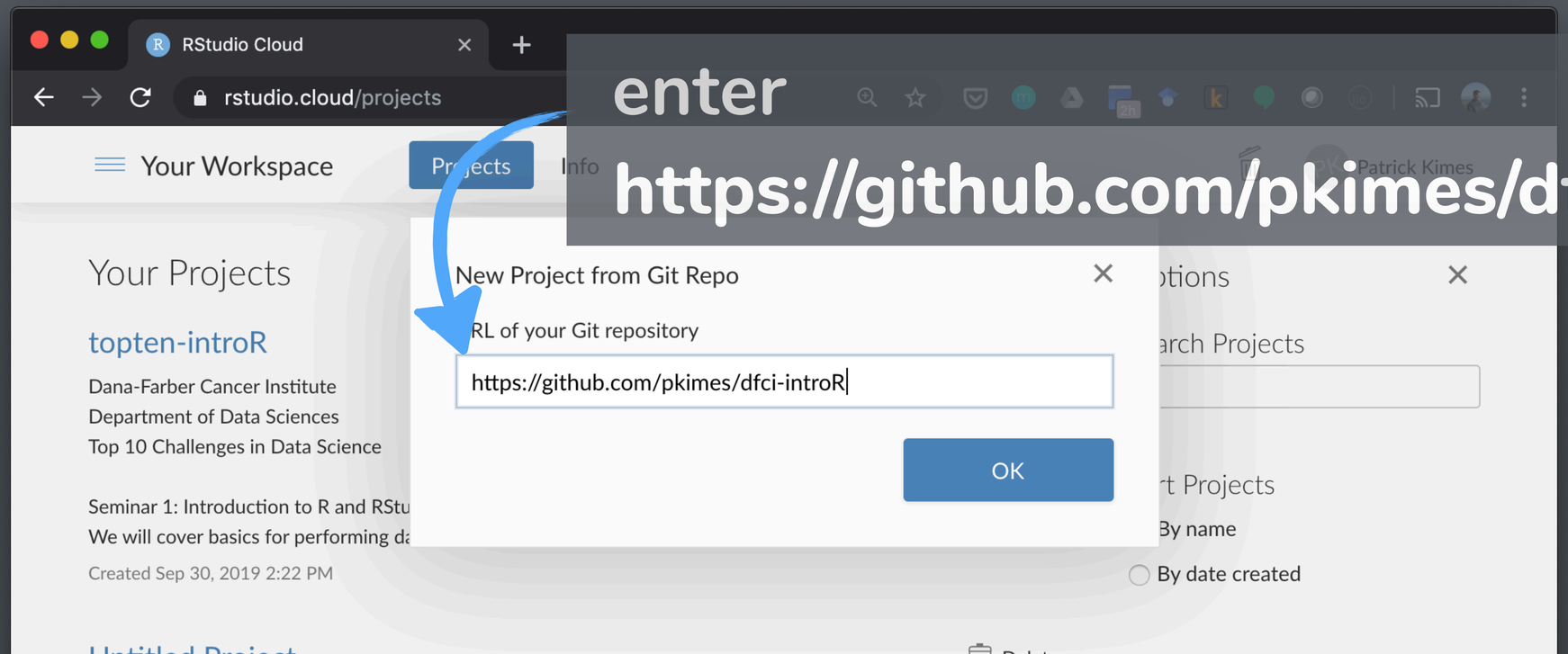
https://rstudio.cloud

select
[from Git repo]

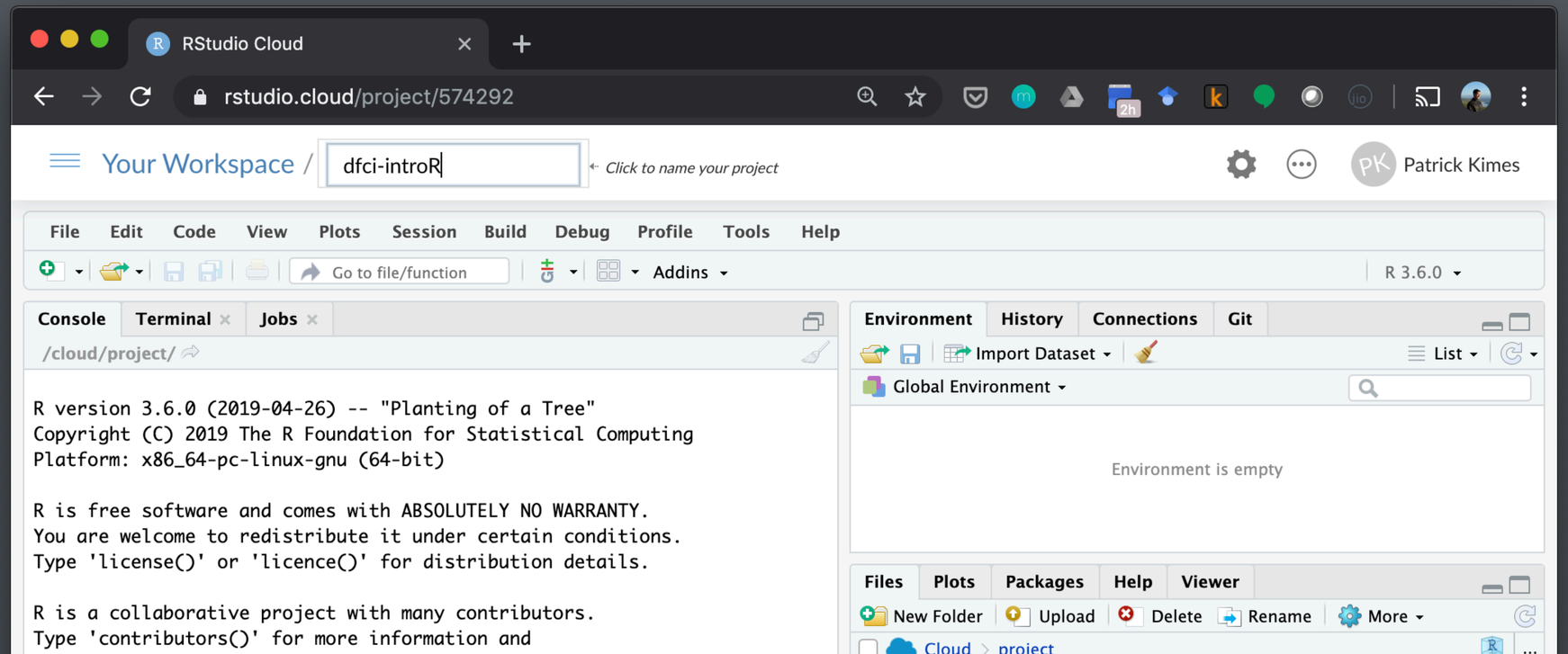


lucky us!
we have a workaround!

https://rstudio.cloud

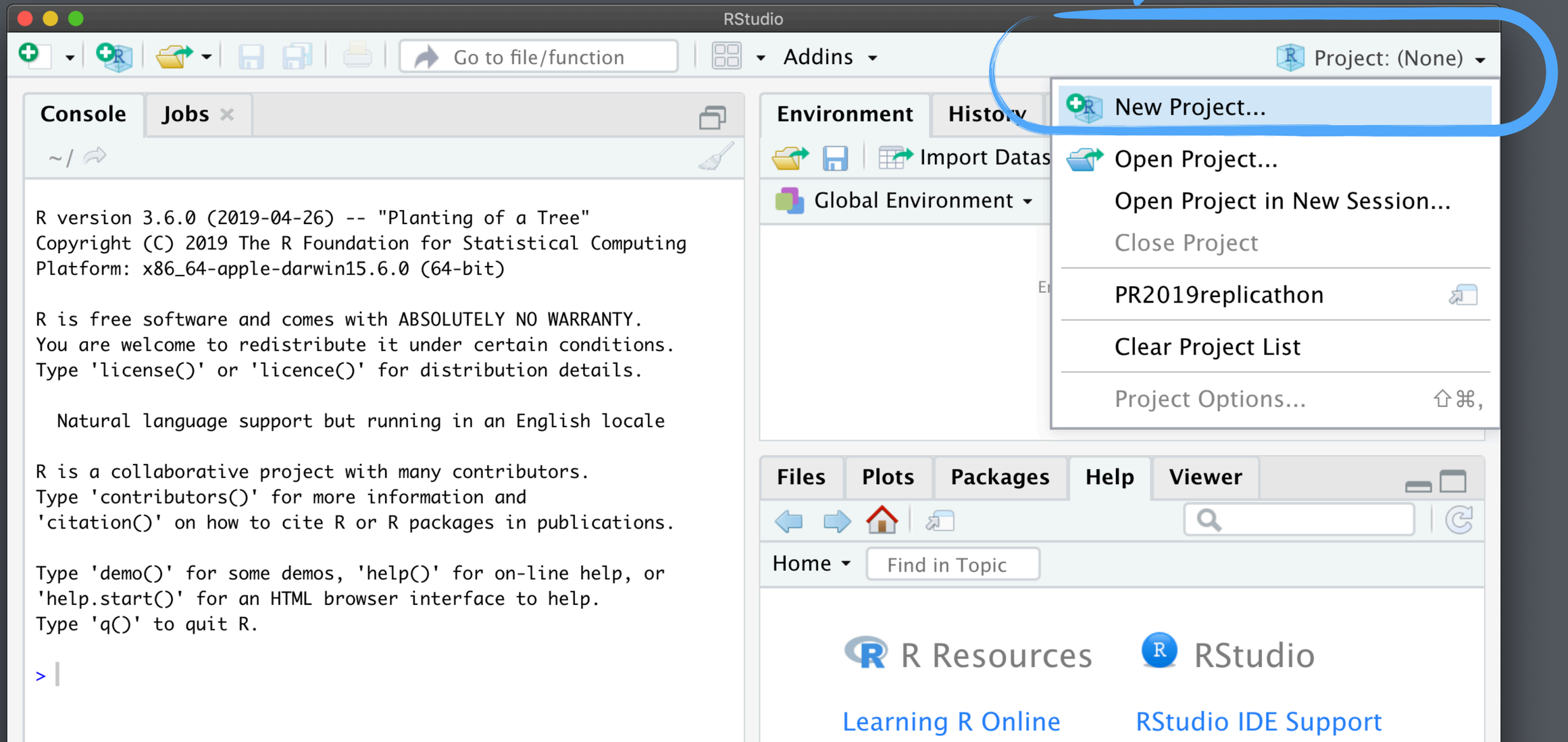


welcome to RStudio!

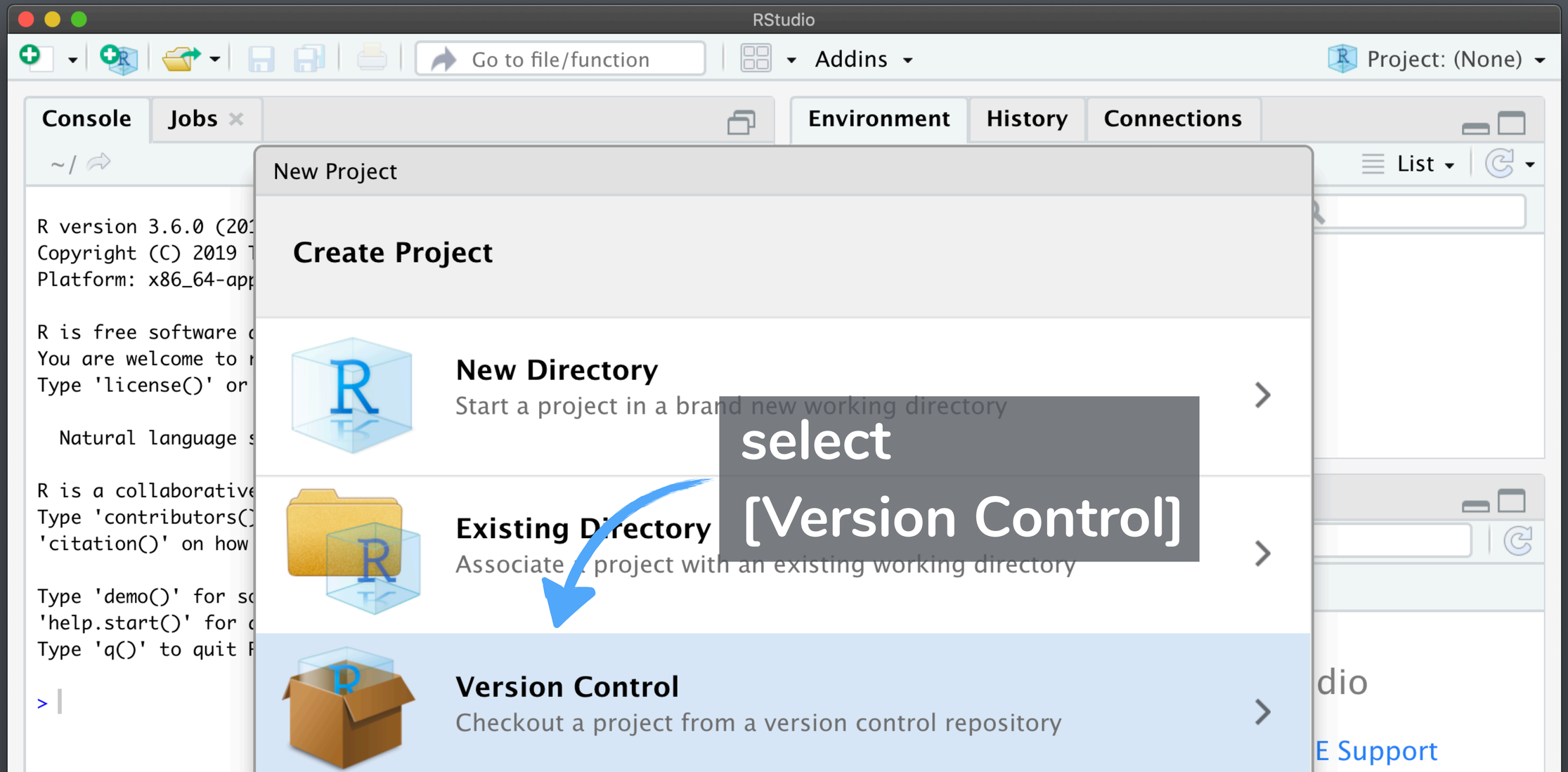


local RStudio

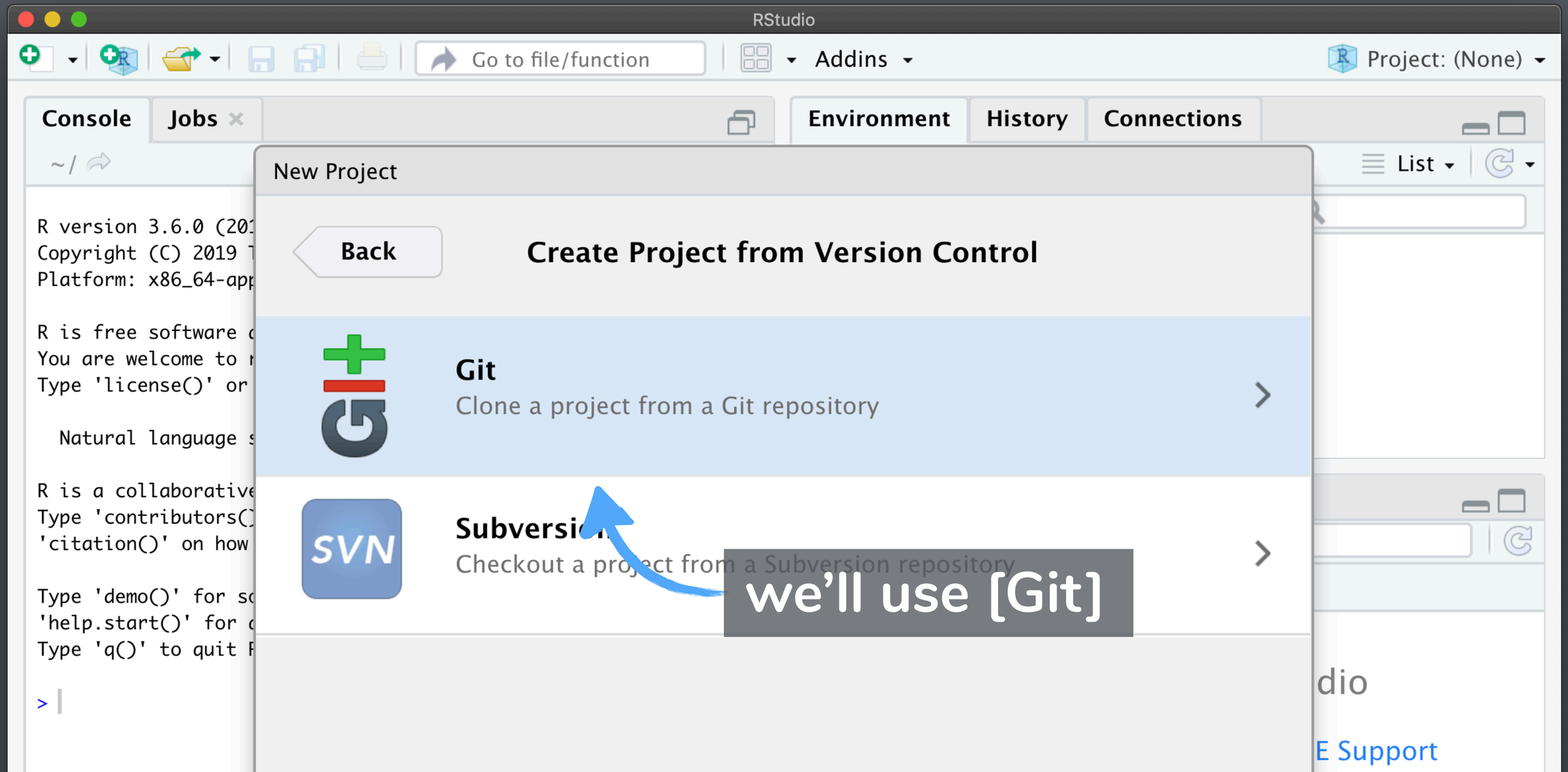
select
[New Project]



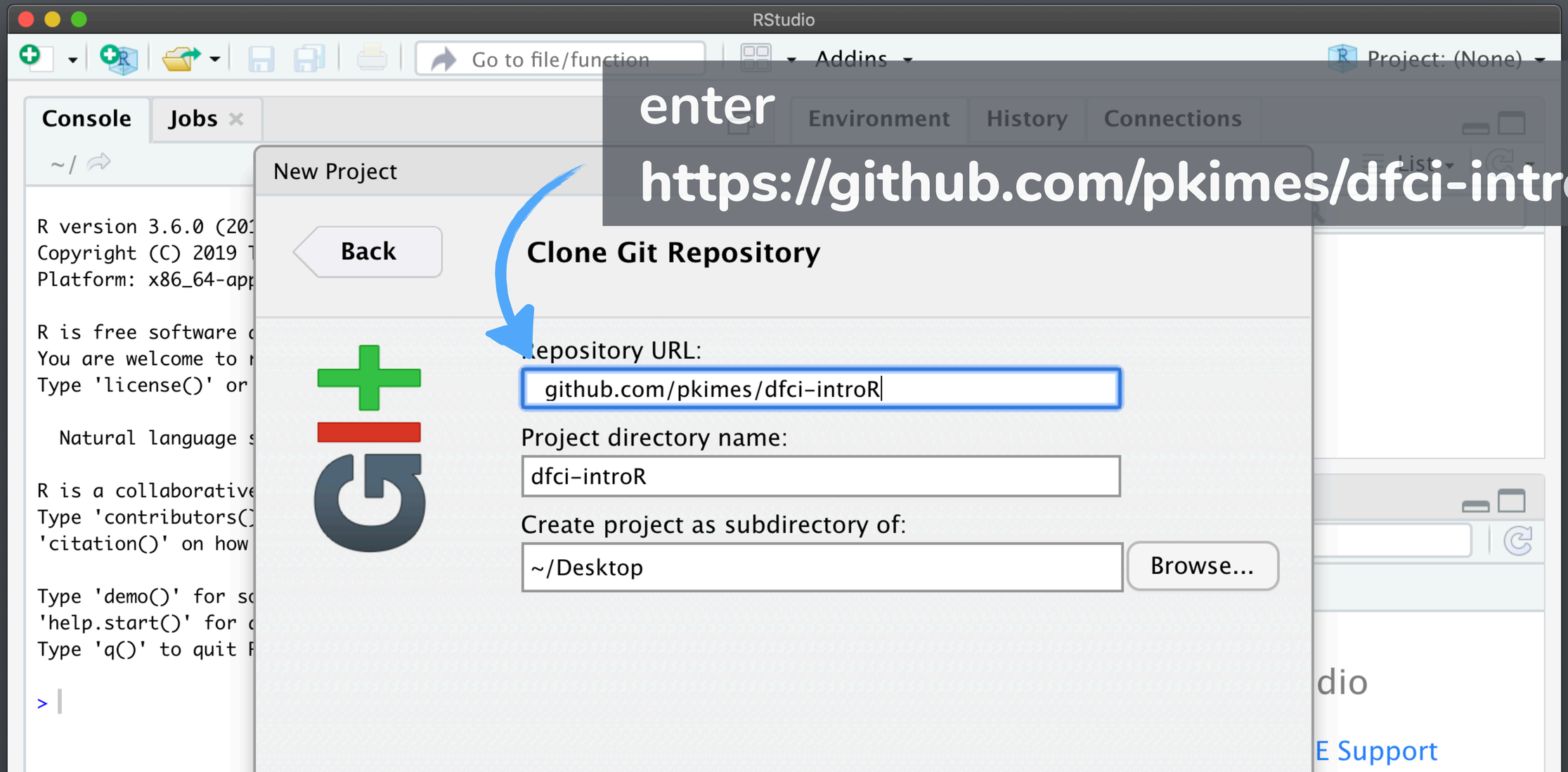
local RStudio



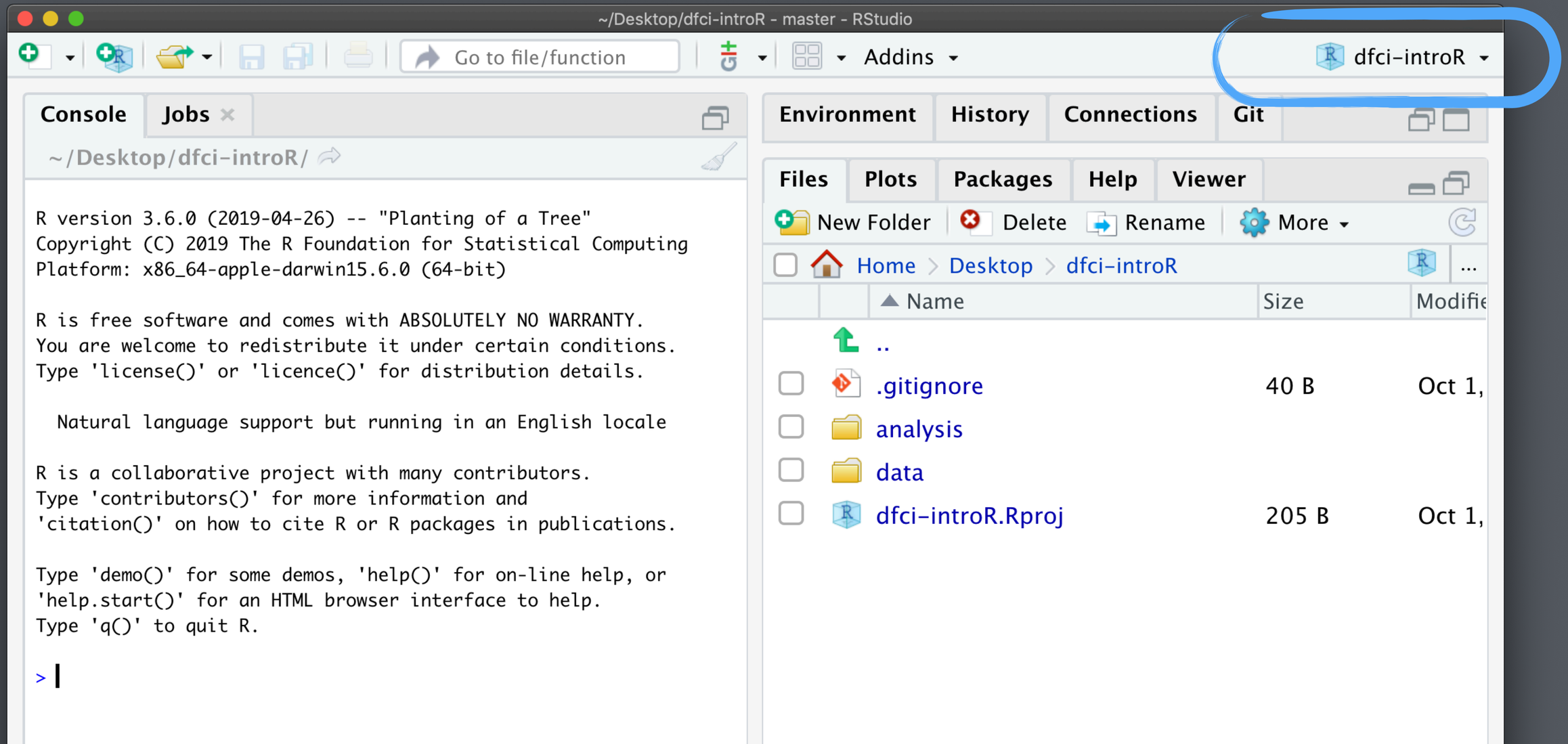
local RStudio



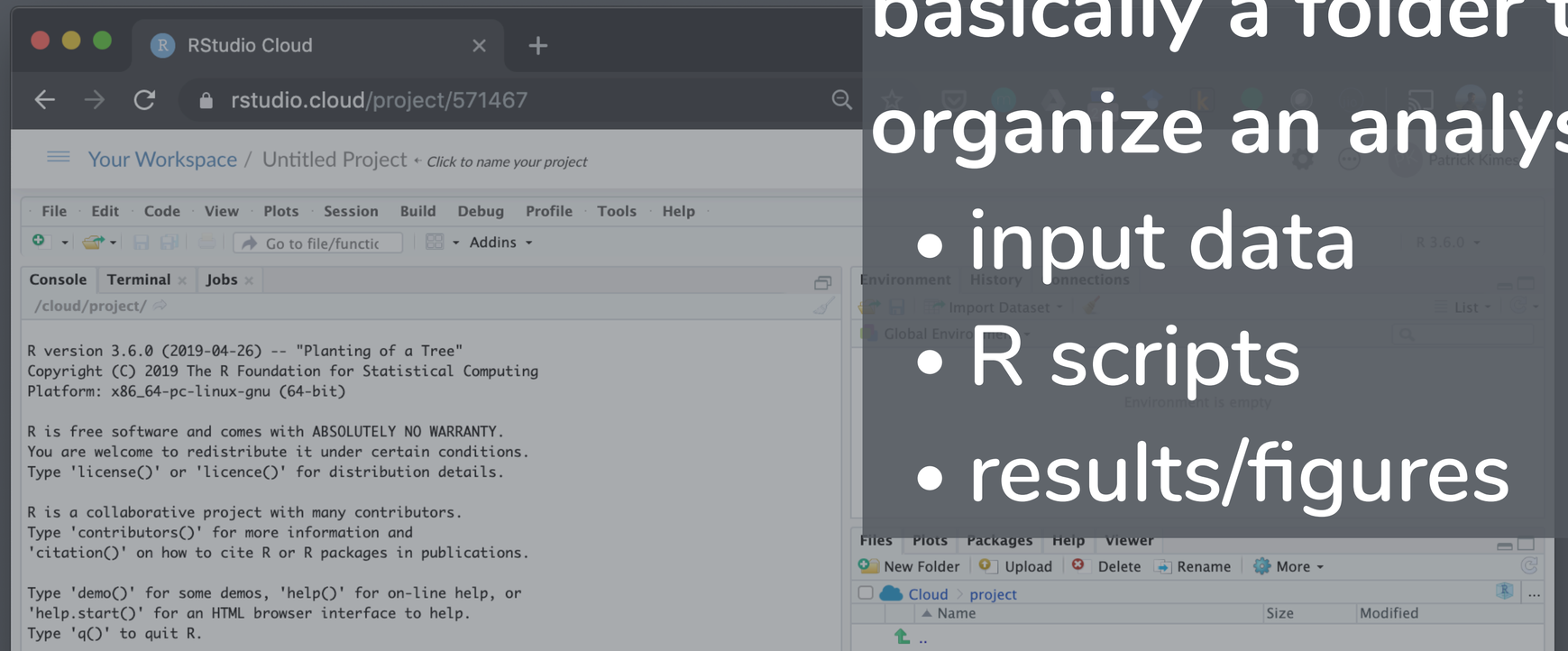
local RStudio



we're good to go!



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

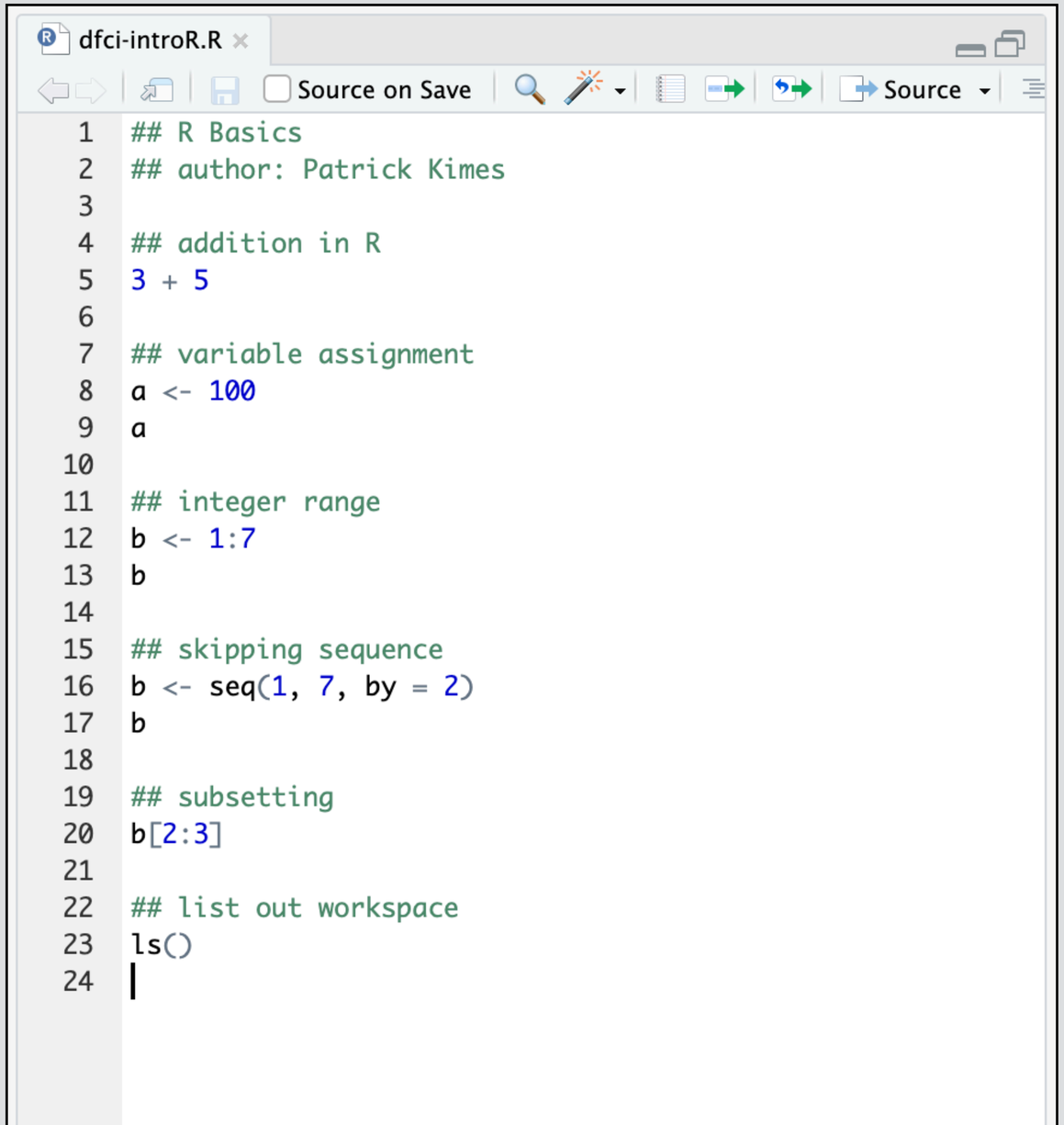
some pieces in the modern (R) data scientist's toolbox

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


dfci-introR.R
Source on Save
Source
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()

.R file

dfci-introR.R

Source on Save | Source

```r
## 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

dfci-introR.Rmd

Knit | Insert | Run

````
---
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

dfci-introR.R

Source on Save | Source

```r
## 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

dfci-introR.Rmd

Knit | Insert | Run

```
---
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

dfci-introR.R

Source on Save | Source

```r
## 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()
```

code "chunks"

.Rmd file

dfci-introR.Rmd

Knit | Insert | Run

````
---
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

dfci-introR.R

Source on Save | Source

```r
## 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

dfci-introR.Rmd

Knit | Insert | Run

```
---
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

dfci-introR.R

Source on Save | Source

```r
## 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()
```

specified output format

.Rmd file

dfci-introR.Rmd

Knit | Insert | Run

````
---
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 `:`.
````

/cloud/project/dfci-introR.html

pkimes.rstudio.cloud/fb549fbb697541bcb3b1cff9c01a57b8/?view=rmarkdown

dfci-introR.html | Open in Browser | Find | Publish

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

dfci-introR.Rmd

Knit | Insert | Run

```
---
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

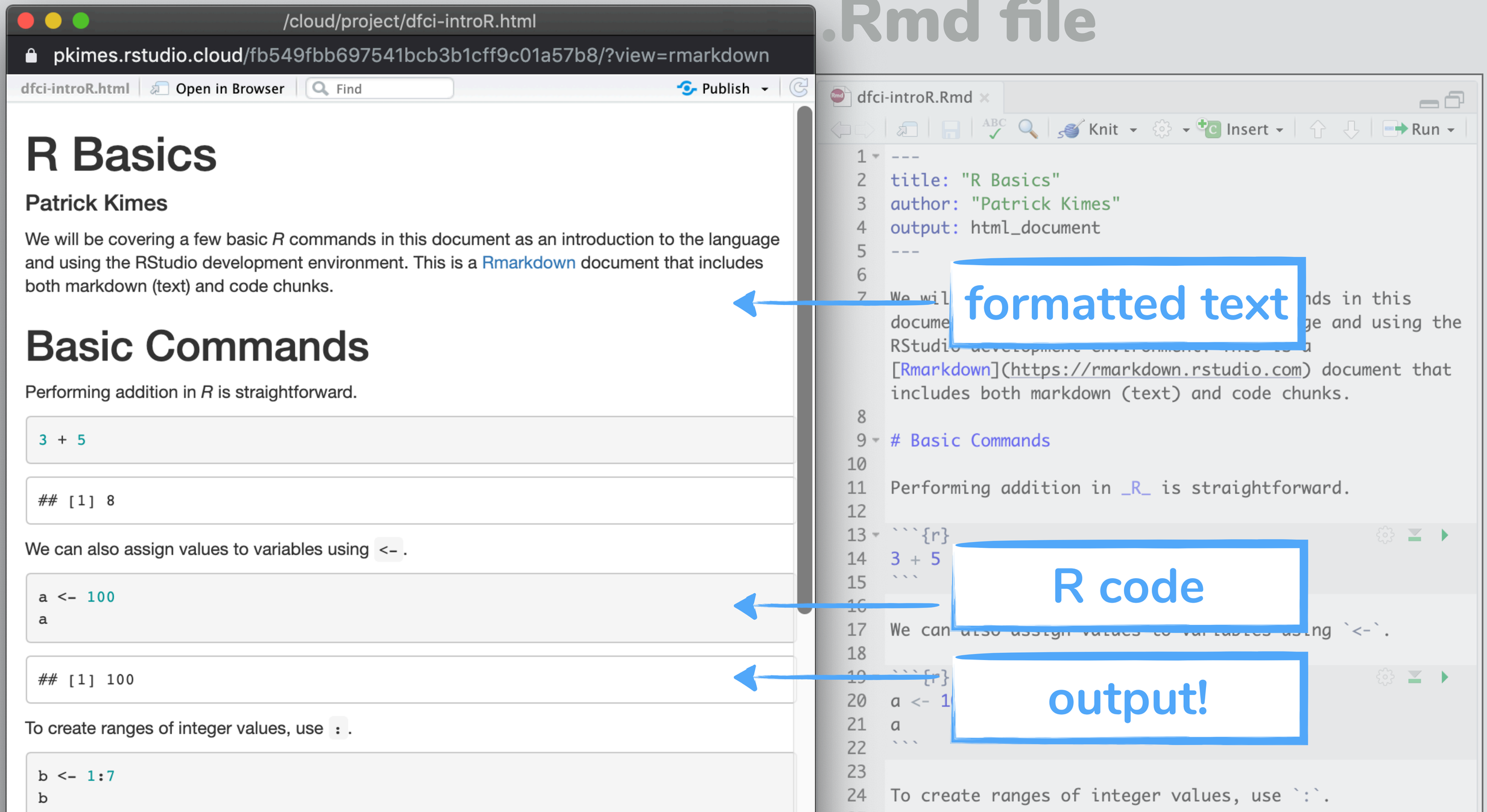
Performing addition in _R_ is straightforward.

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

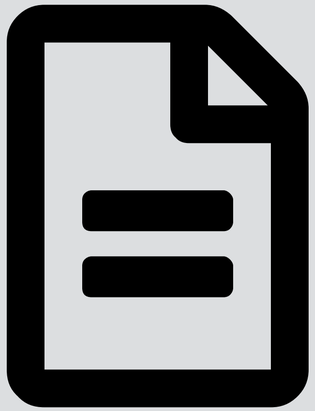
We can also assign values to variables using `<-`.

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

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


/cloud/project/dfci-introR.html
pkimes.rstudio.cloud/fb549fbb697541bcb3b1cff9c01a57b8/?view=rmarkdown
dfci-introR.html
Open in Browser
Find
Publish
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
dfci-introR.Rmd
Knit
Insert
Run
1 ---
2 title: "R Basics"
3 author: "Patrick Kimes"
4 output: html_document
5 ---
6
7 We wil nds in this
docume ge and using the
RStudi
[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 ng `<-`.
18
19 ```{r}
20 a <- 1
21 a
22 ```
23
24 To create ranges of integer values, use `:`.
formatted text
R code
output!

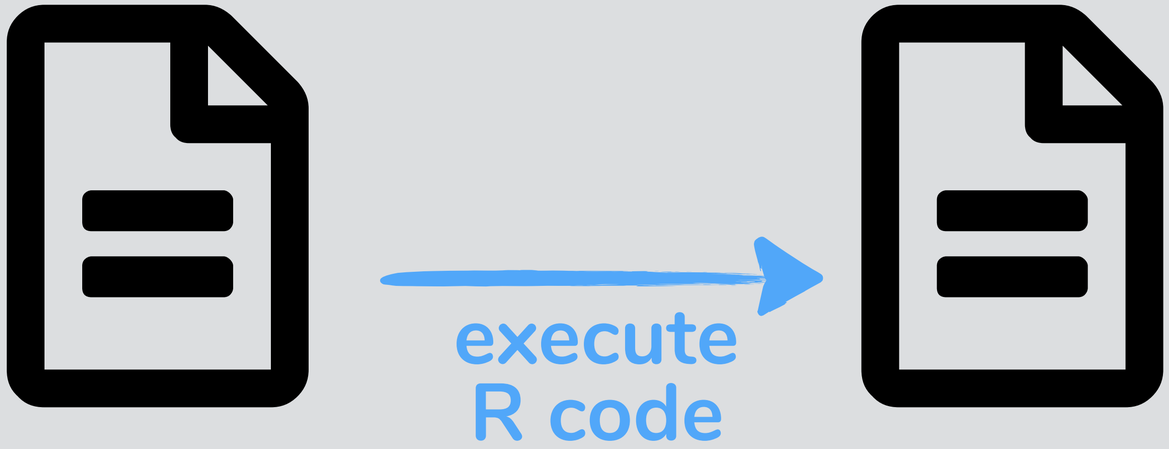
rmarkdown | documentation, communication



myfile.Rmd

markdown
+
R code chunks

rmarkdown | documentation, communication



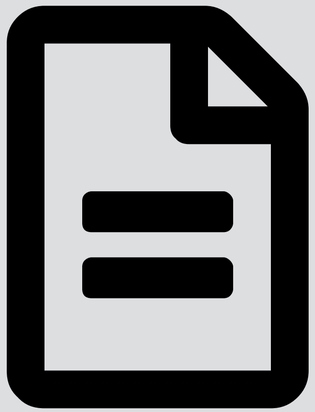
myfile.Rmd

markdown
+
R code chunks

myfile.md

markdown

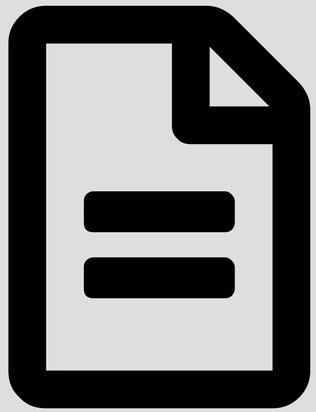
rmarkdown | documentation, communication



myfile.Rmd

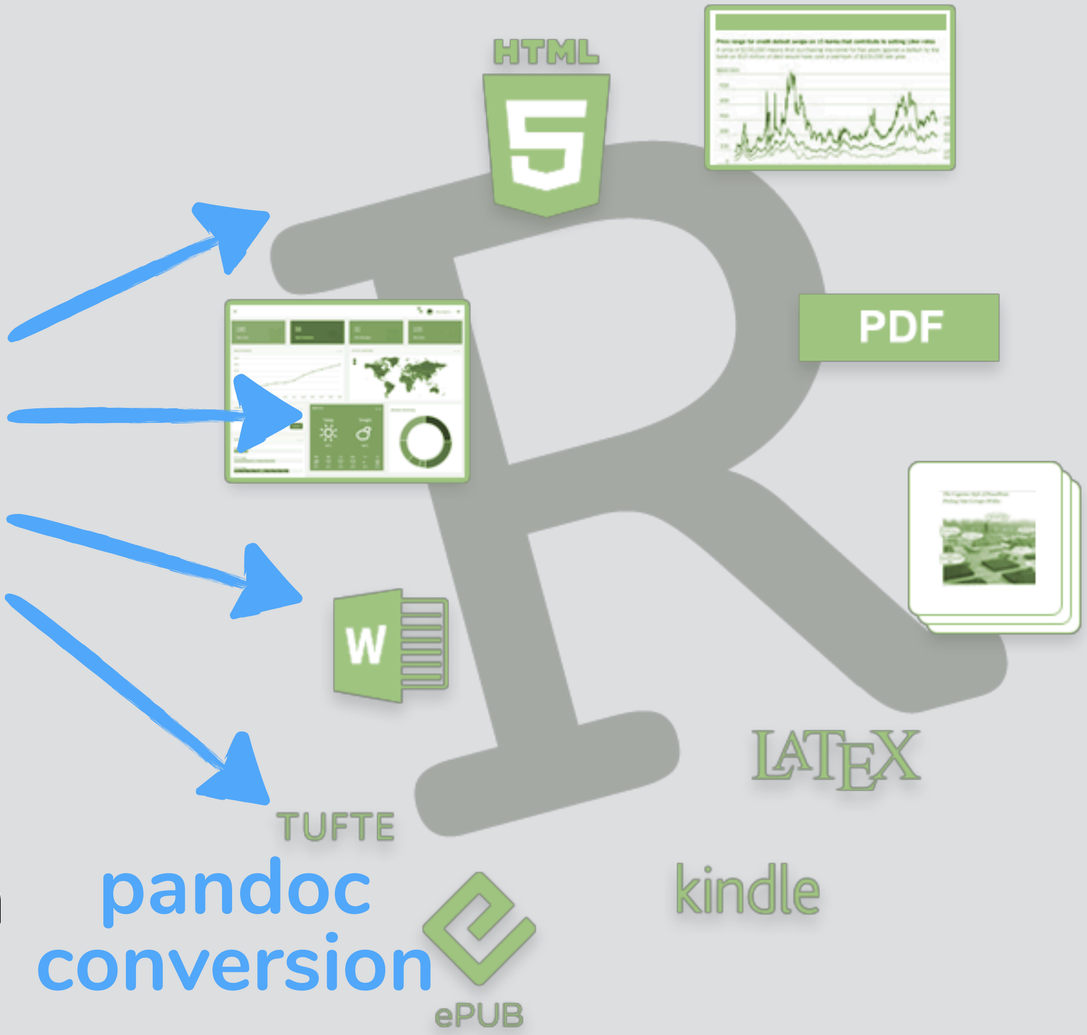
markdown
+
R code chunks

execute
R code



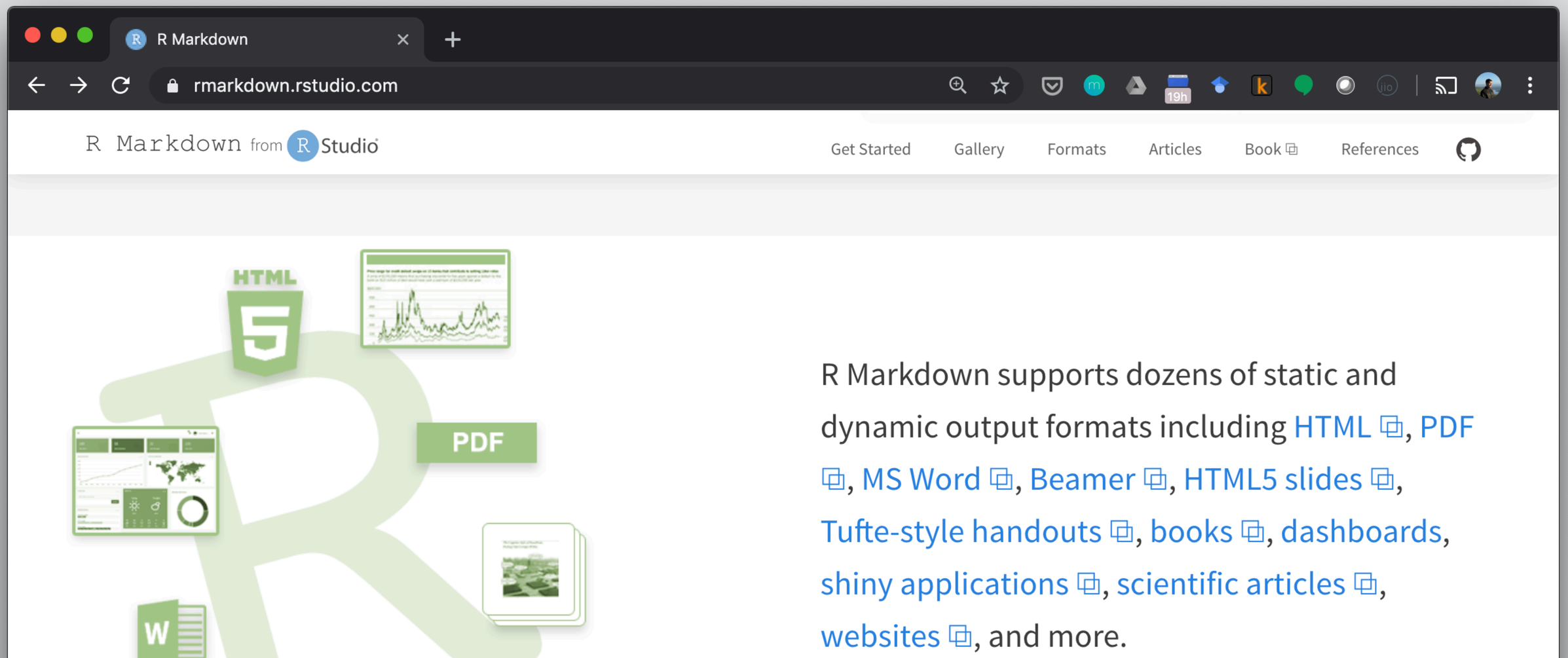
myfile.md

markdown



rmarkdown | documentation, communication

rmarkdown.rstudio.com



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

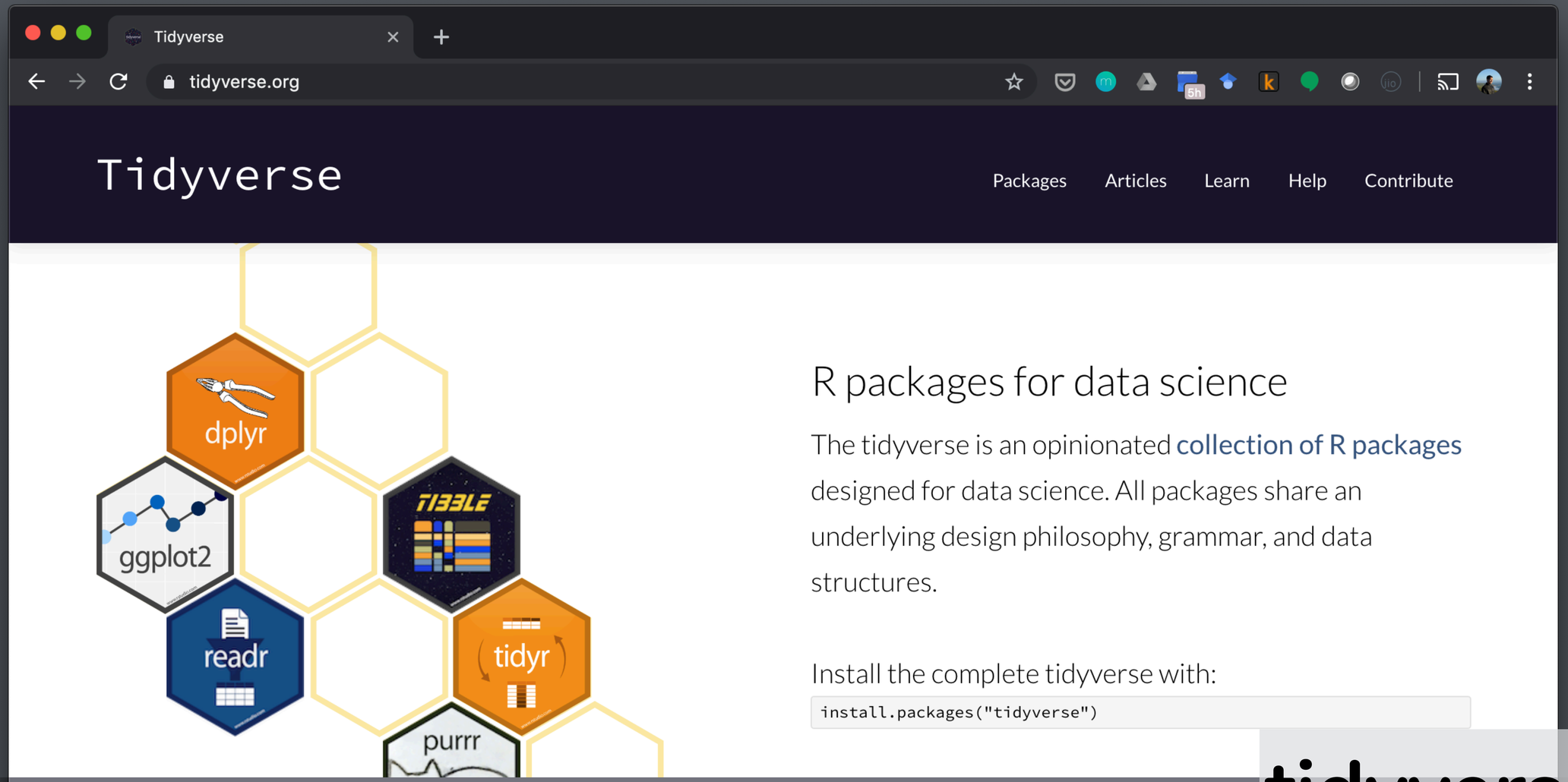
some pieces in the modern (R) data scientist's toolbox

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

rmarkdown	documentation, communication
tidyverse	data manipulation, visualization
shiny	web application framework
[bioconductor]	community of genomics packages

tidyverse | data manipulation, visualization

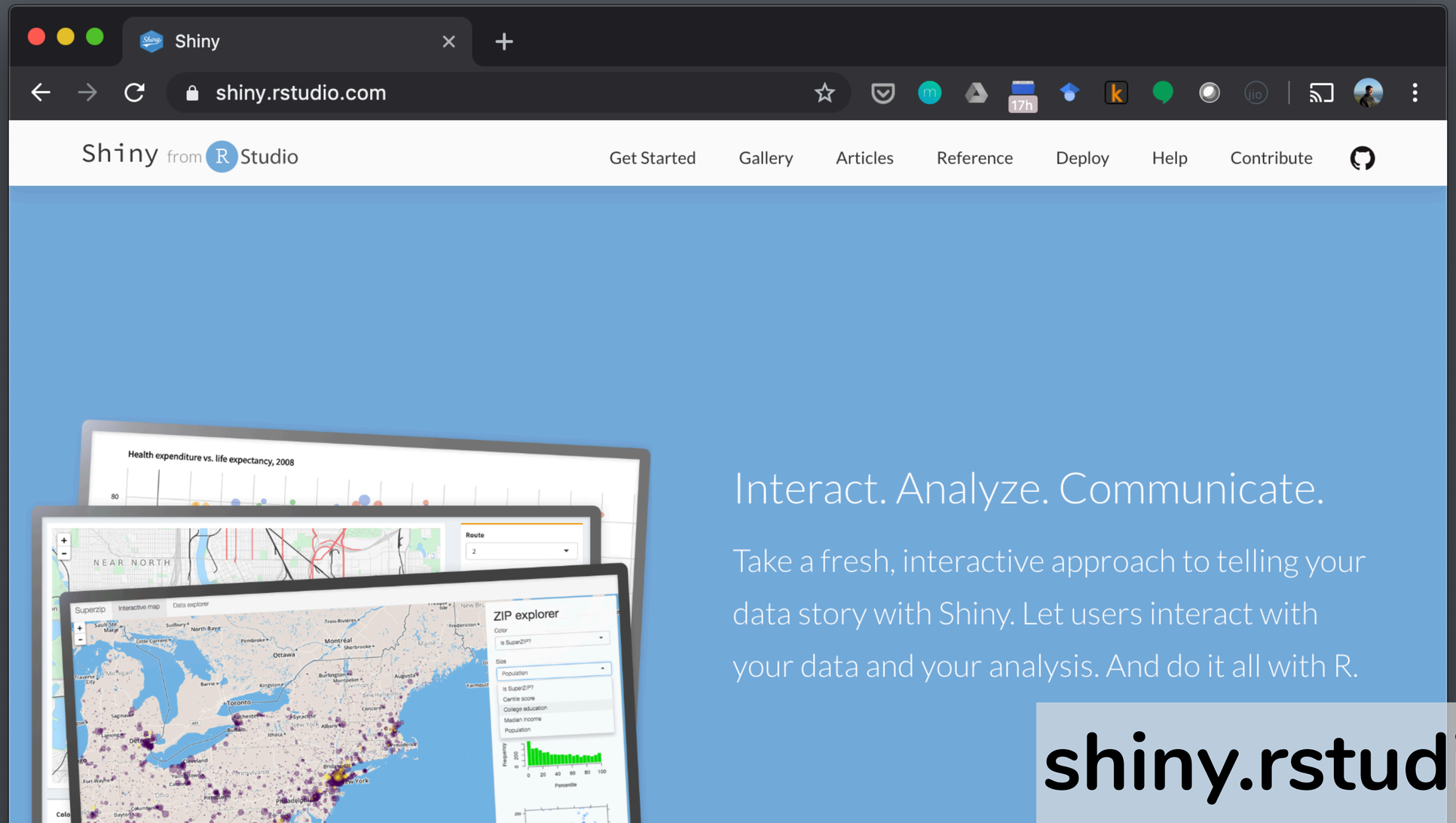


tidyverse.org

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

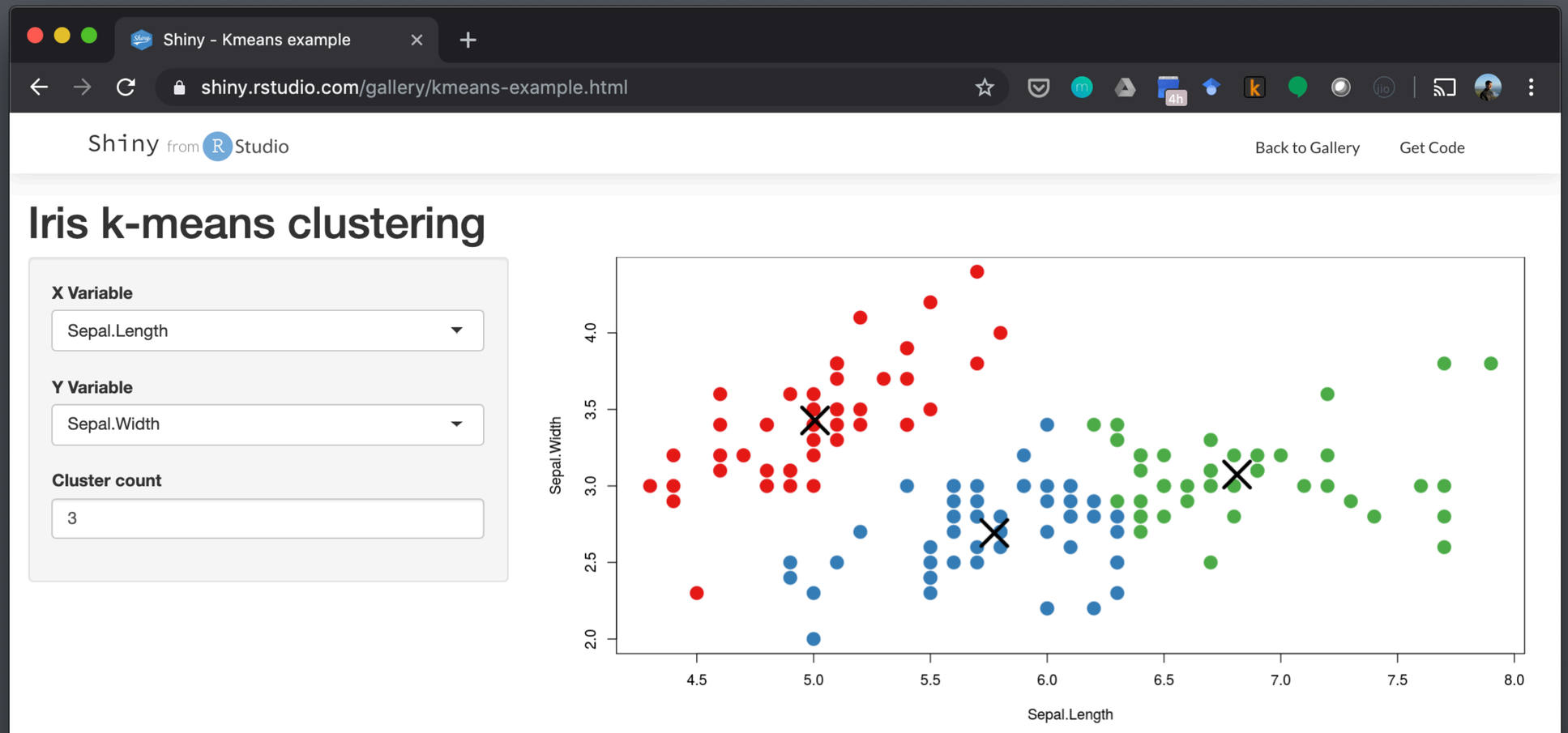
rmarkdown	documentation, communication
tidyverse	data manipulation, visualization
shiny	web application framework
[bioconductor]	community of genomics packages

shiny | web application framework



shiny.rstudio.com

shiny | web application framework

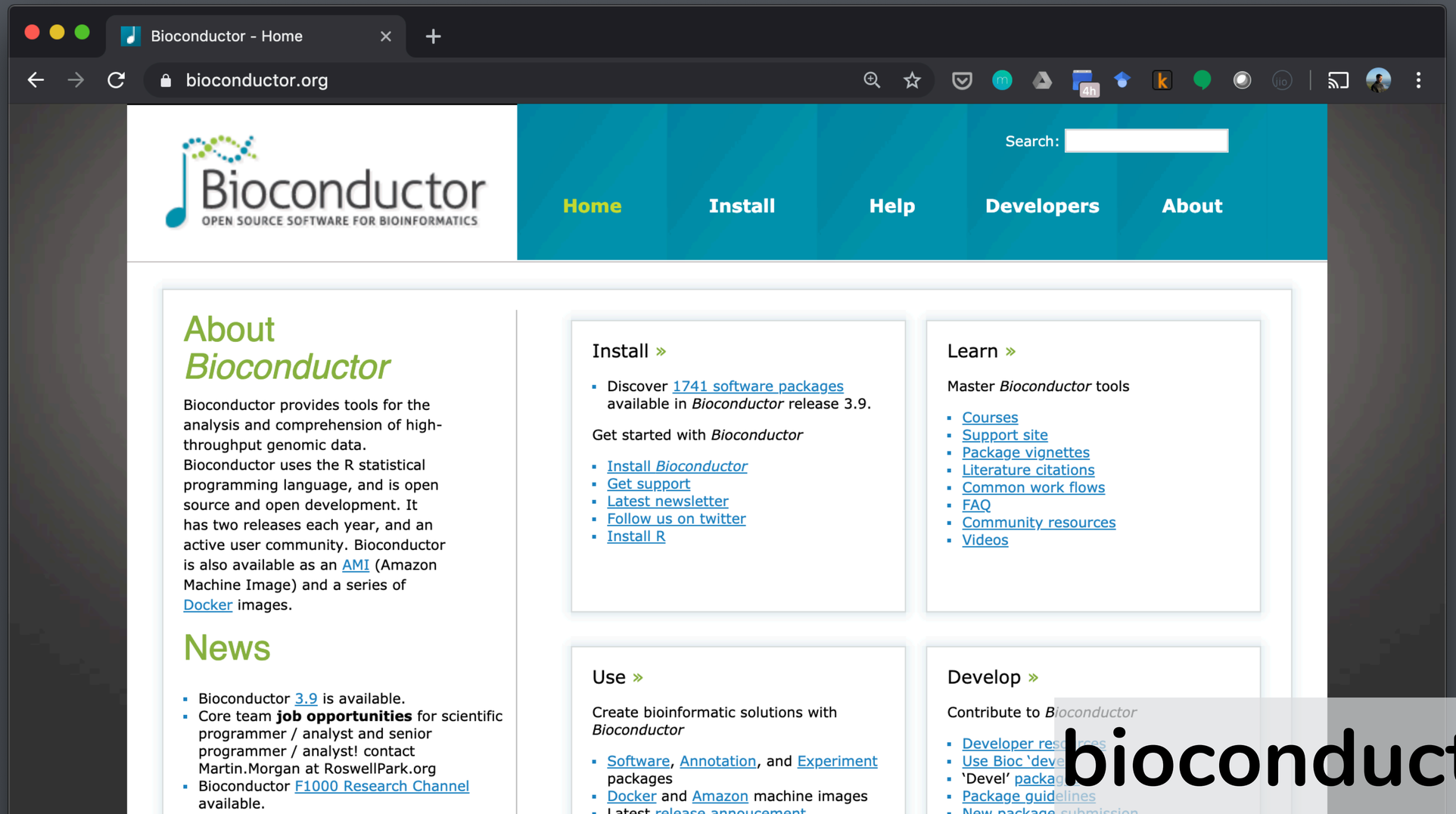


shiny.rstudio.com/gallery/kmeans-example

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

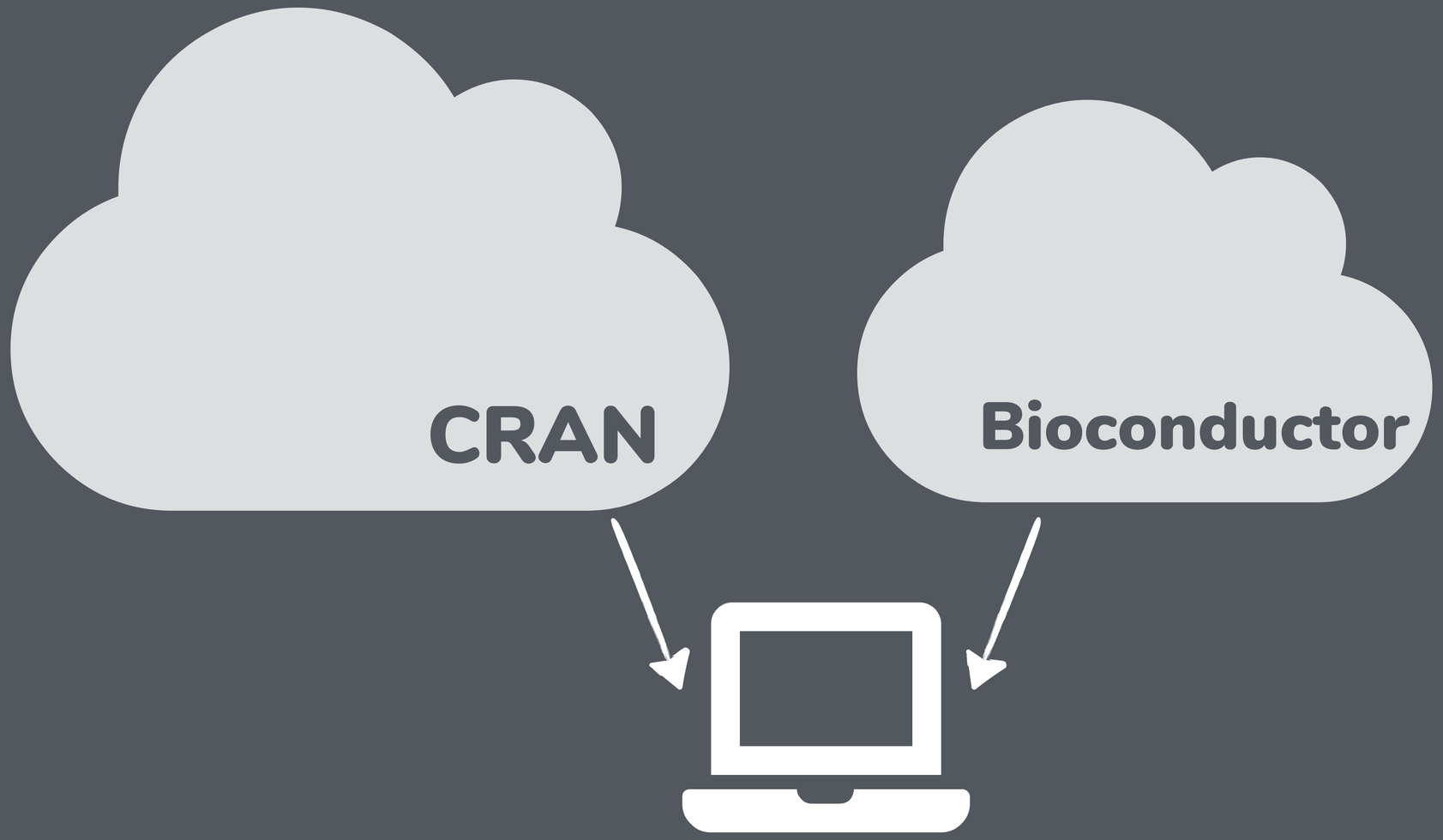
rmarkdown	documentation, communication
tidyverse	data manipulation, visualization
shiny	web application framework
[bioconductor]	community of genomics packages

bioconductor | community of genomics packages



bioconductor.org

bioconductor | community of genomics packages



- genomic focus
  - software
  - annotations
  - data
- package reviews
  - scope
  - consistency

RStudio®

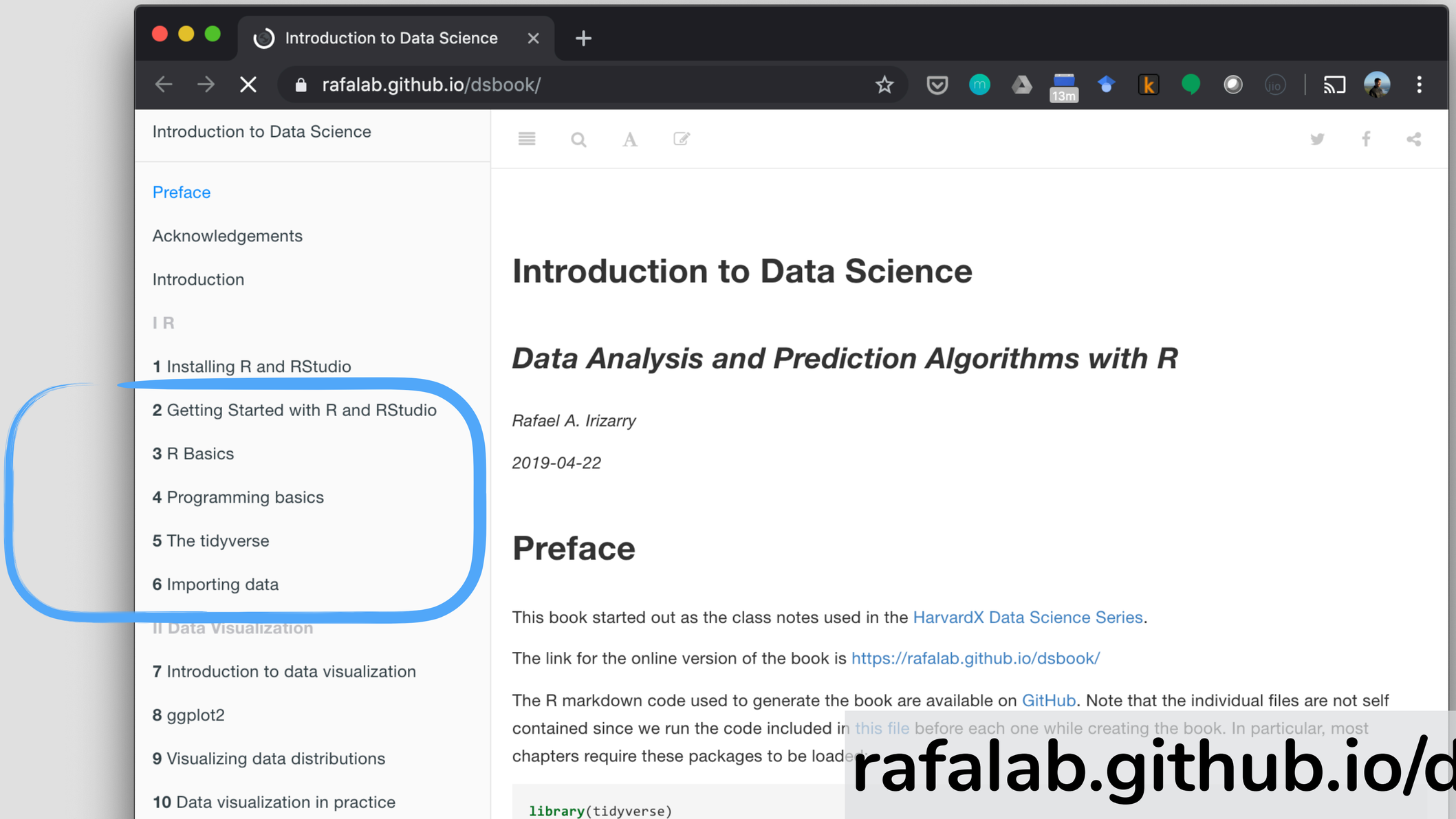
awesome!

where do we go from here?

where do we go
from here?

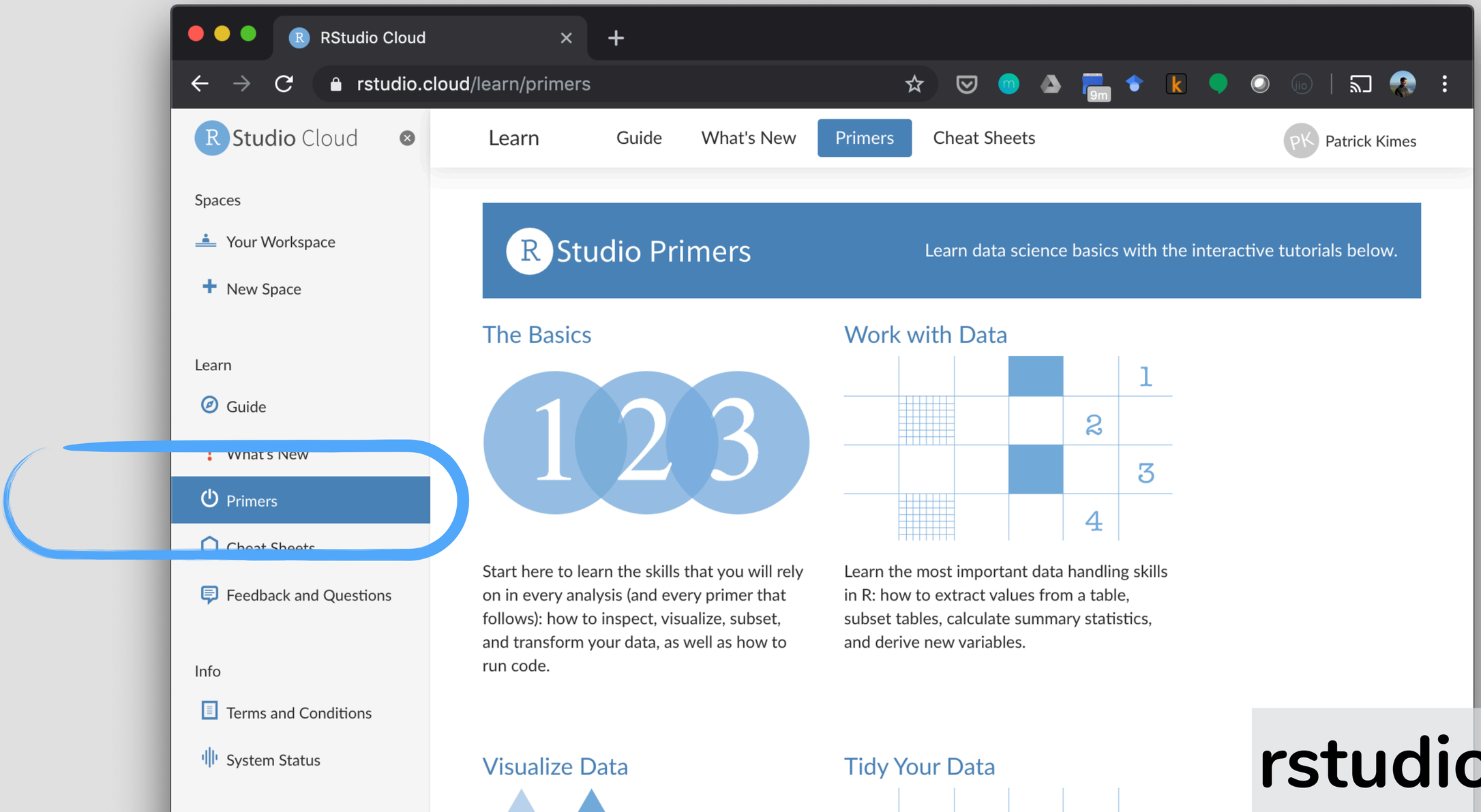
wait, I’m lost

introduction to data science



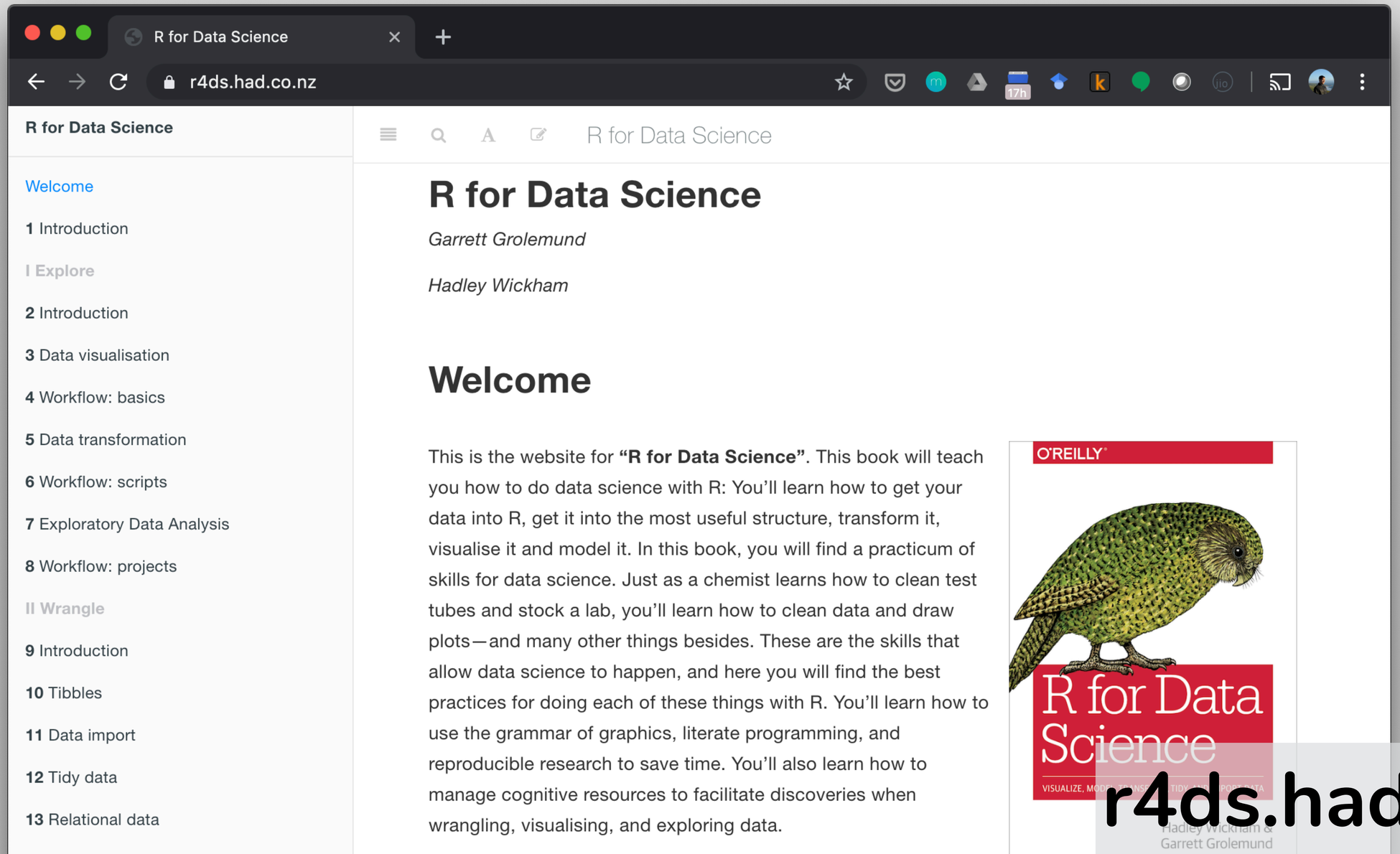
rafalab.github.io/dsbook

RStudio Cloud tutorials



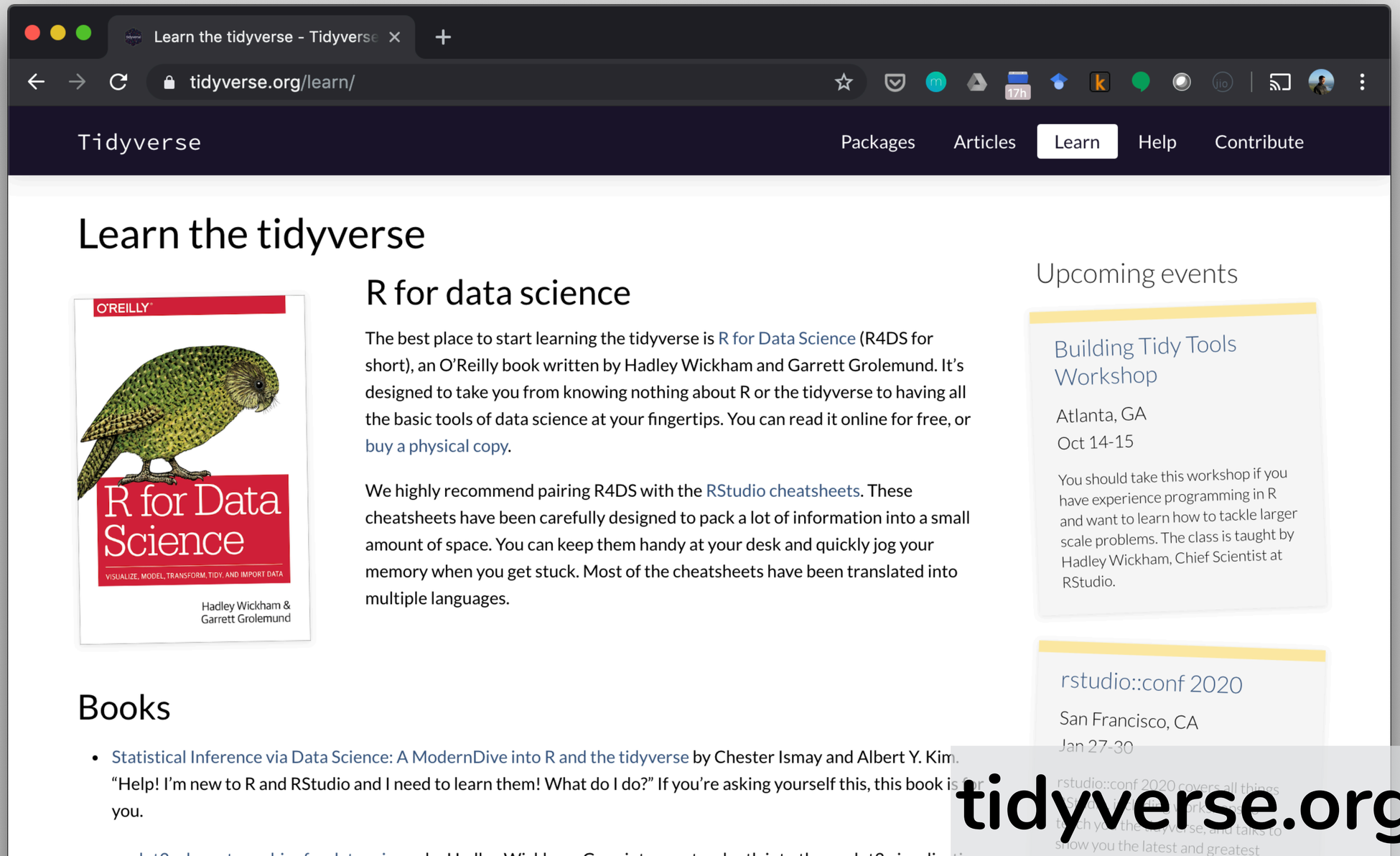
rstudio.cloud

R for data science



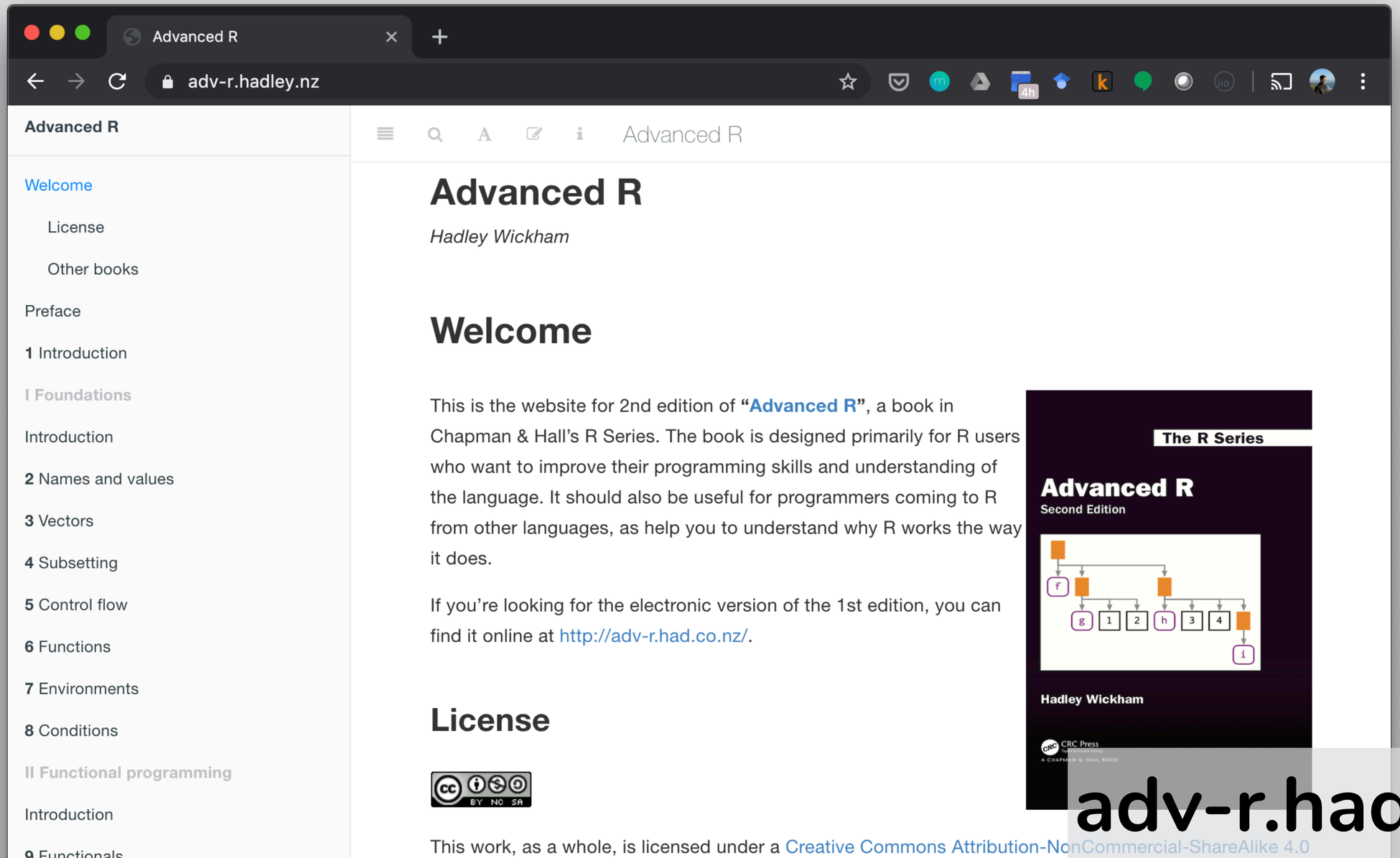
r4ds.had.co.nz

learn the tidyverse



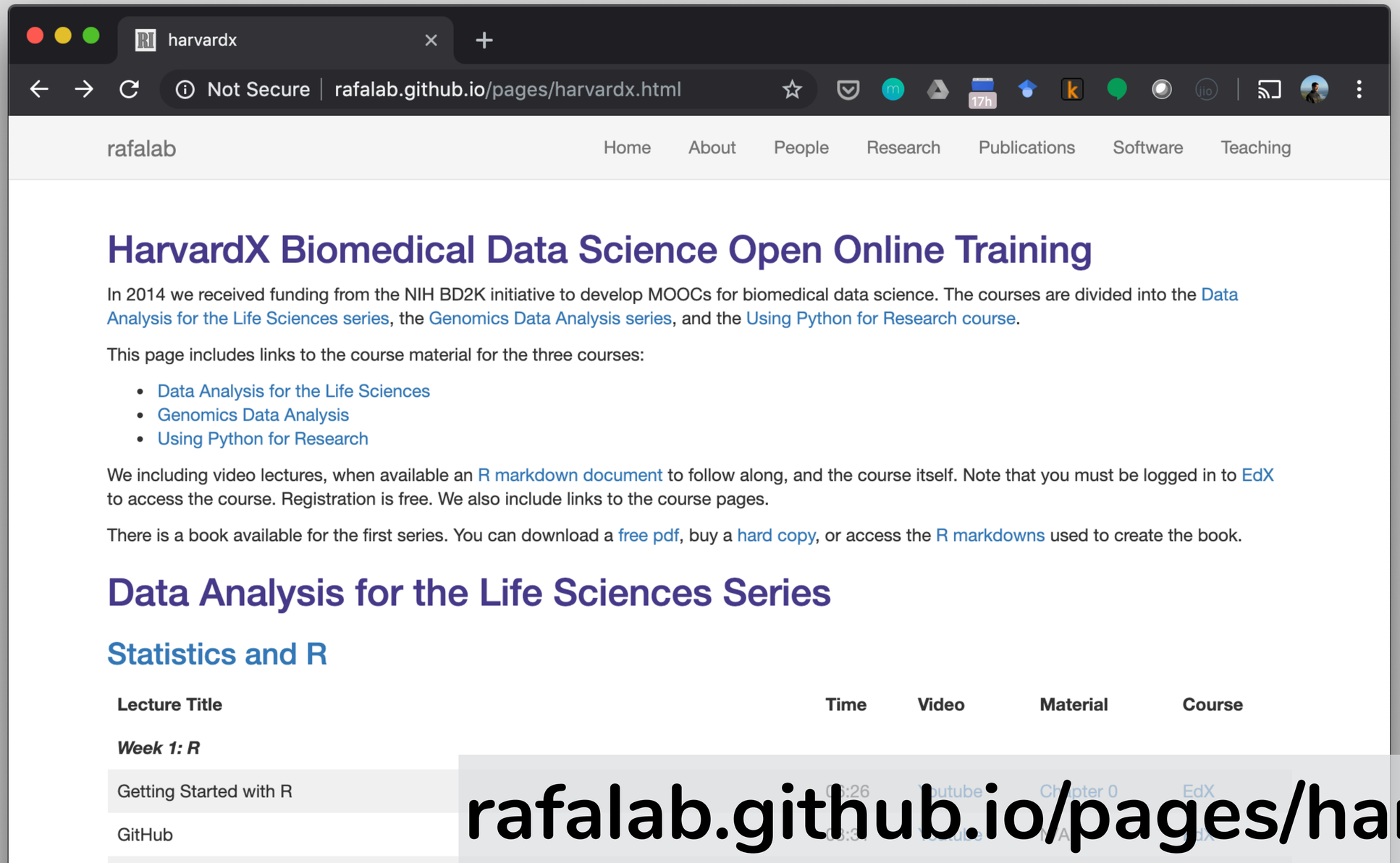
tidyverse.org/learn

advanced R



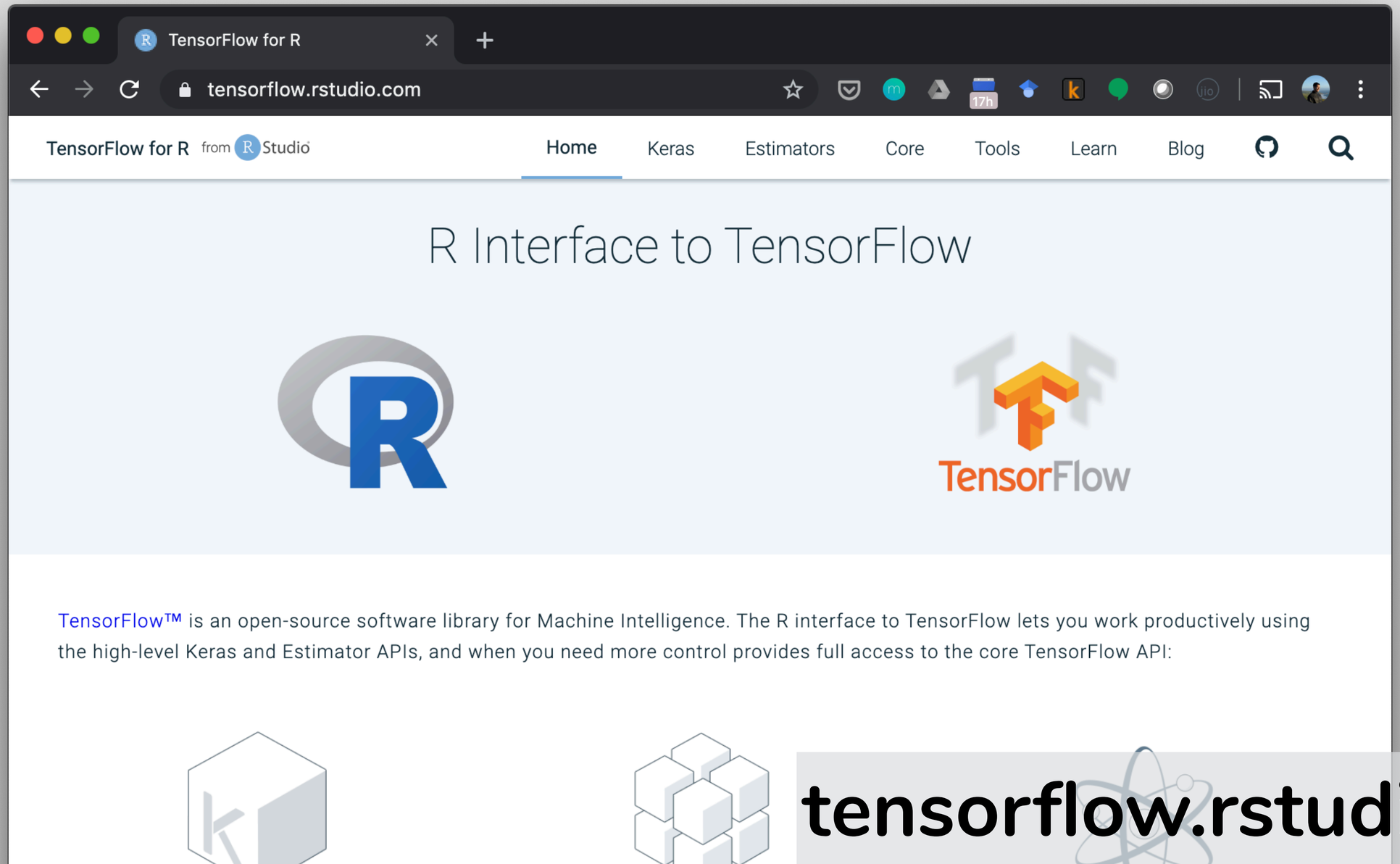
adv-r.hadley.nz

biomedical data science open online training



rafalab.github.io/pages/harvardx

… and because it's 2019, deep learning



tensorflow.rstudio.com

RStudio®

wait, I'm still lost

questions?