Data Science in the tidyverse

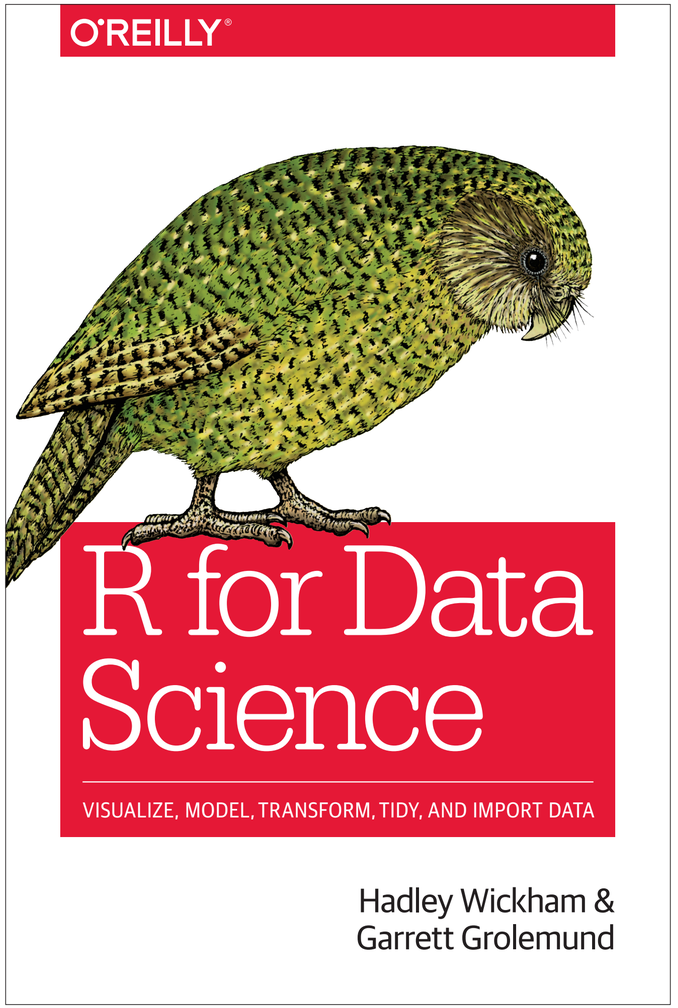
Amelia McNamara

amelia.mn

@AmeliaMN

Online at:

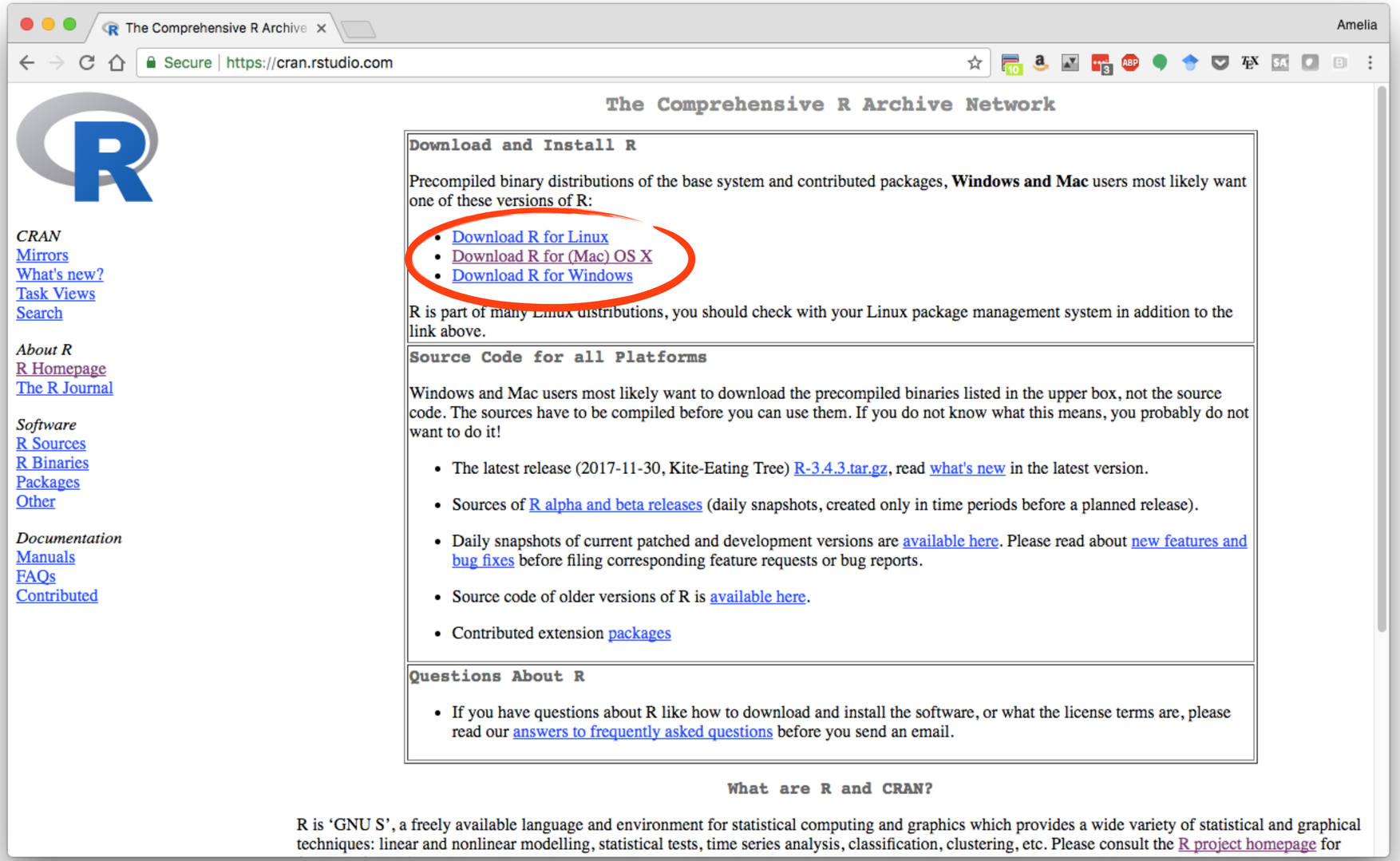
http://r4ds.had.co.nz/



Installing locally

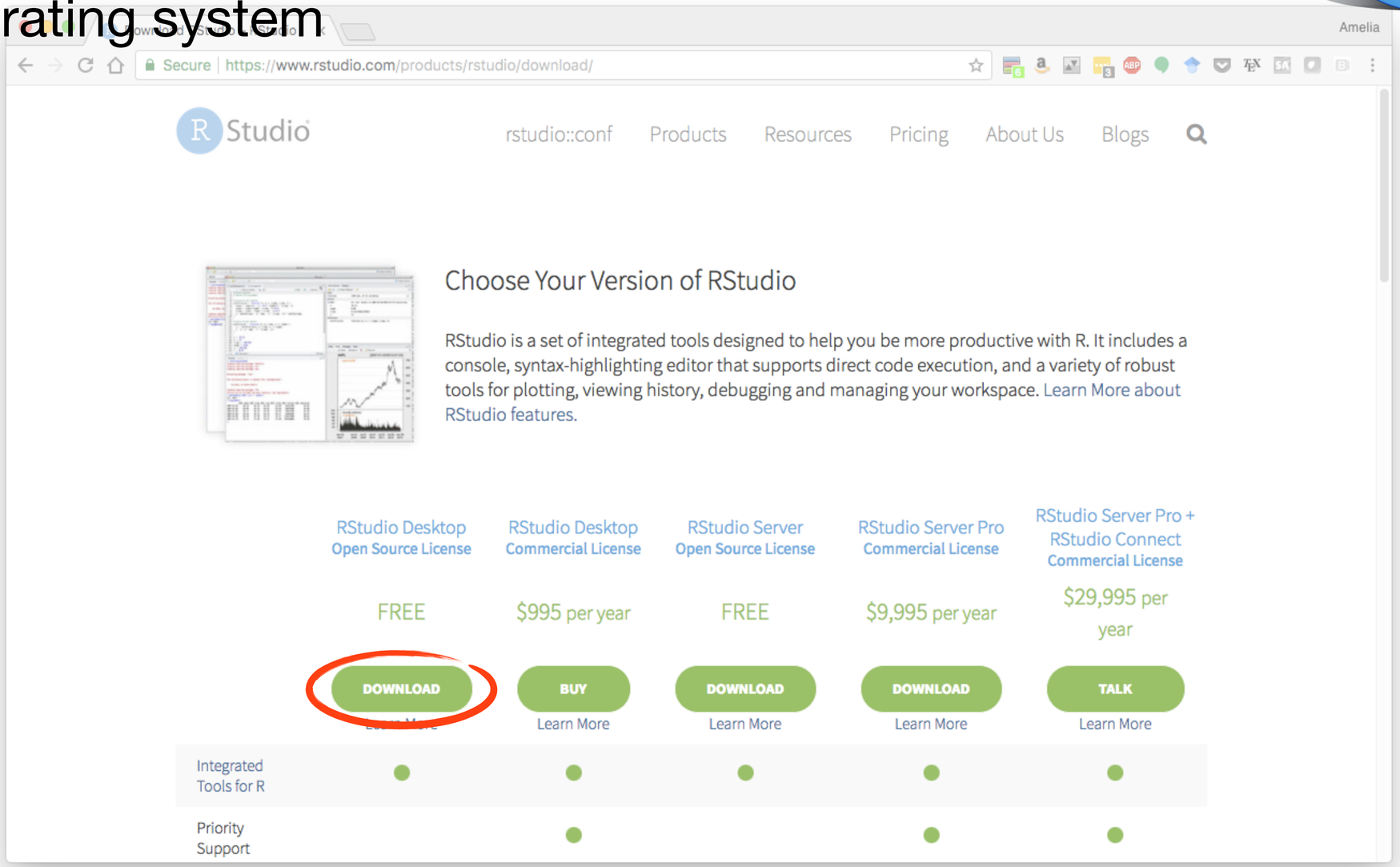
First, you will need to install R (the programming language).

1. Go to https://cran.rstudio.com/
2. Select your operating system



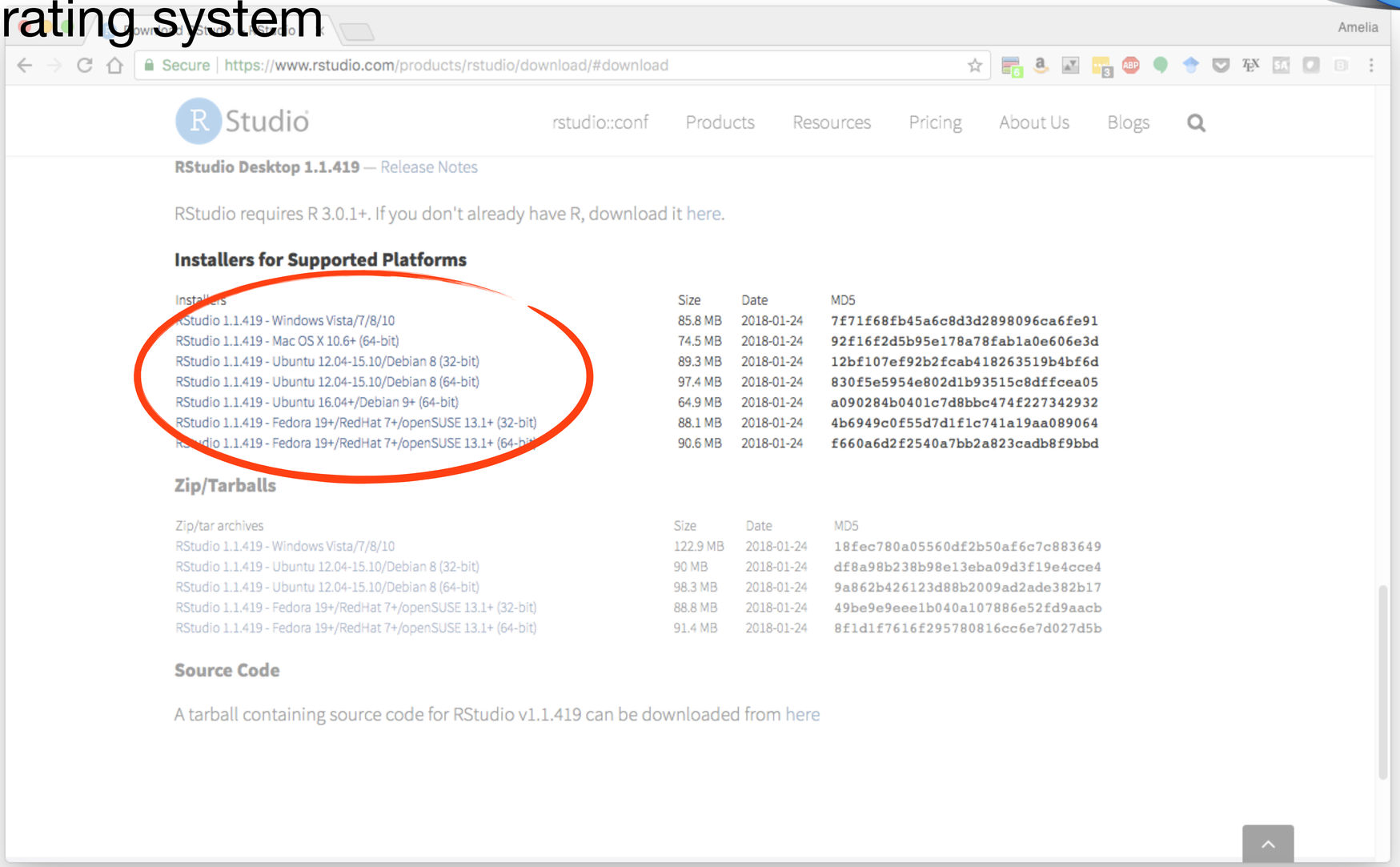
Then, install RStudio (the application).

1. Go to https://www.rstudio.com/products/rstudio/download/
2. Select RStudio desktop
3. Select your operating system



Then, install RStudio (the application).

1. Go to https://www.rstudio.com/products/rstudio/download/
2. Select RStudio desktop
3. Select your operating system



Installing packages

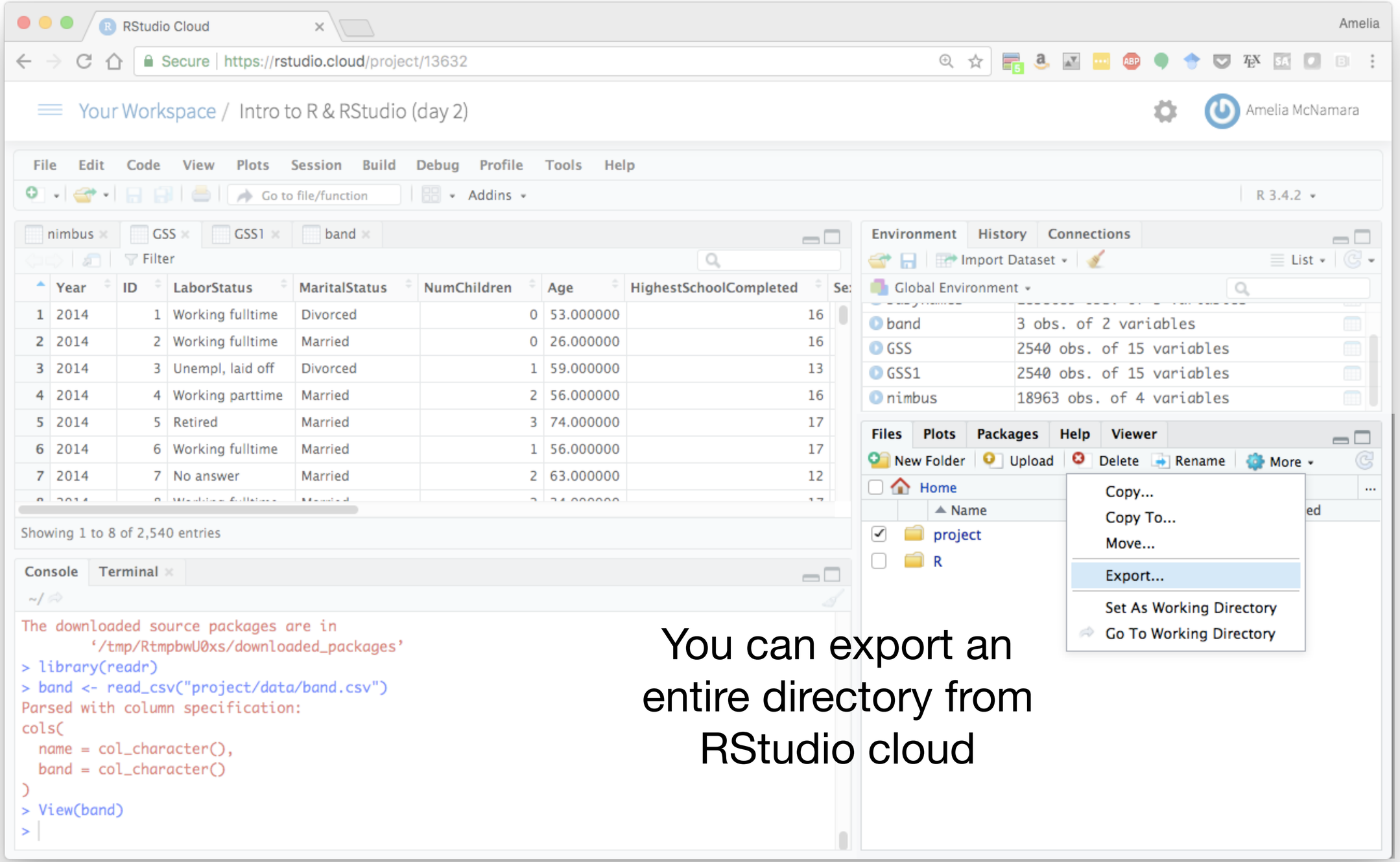
Shortcut to install

- ggplot2, for data visualisation.
- dplyr, for data manipulation.
- tidyr, for data tidying.
- readr, for data import.
- purrr, for functional programming.
- tibble, for tibbles, a modern re-imagining of data frames.

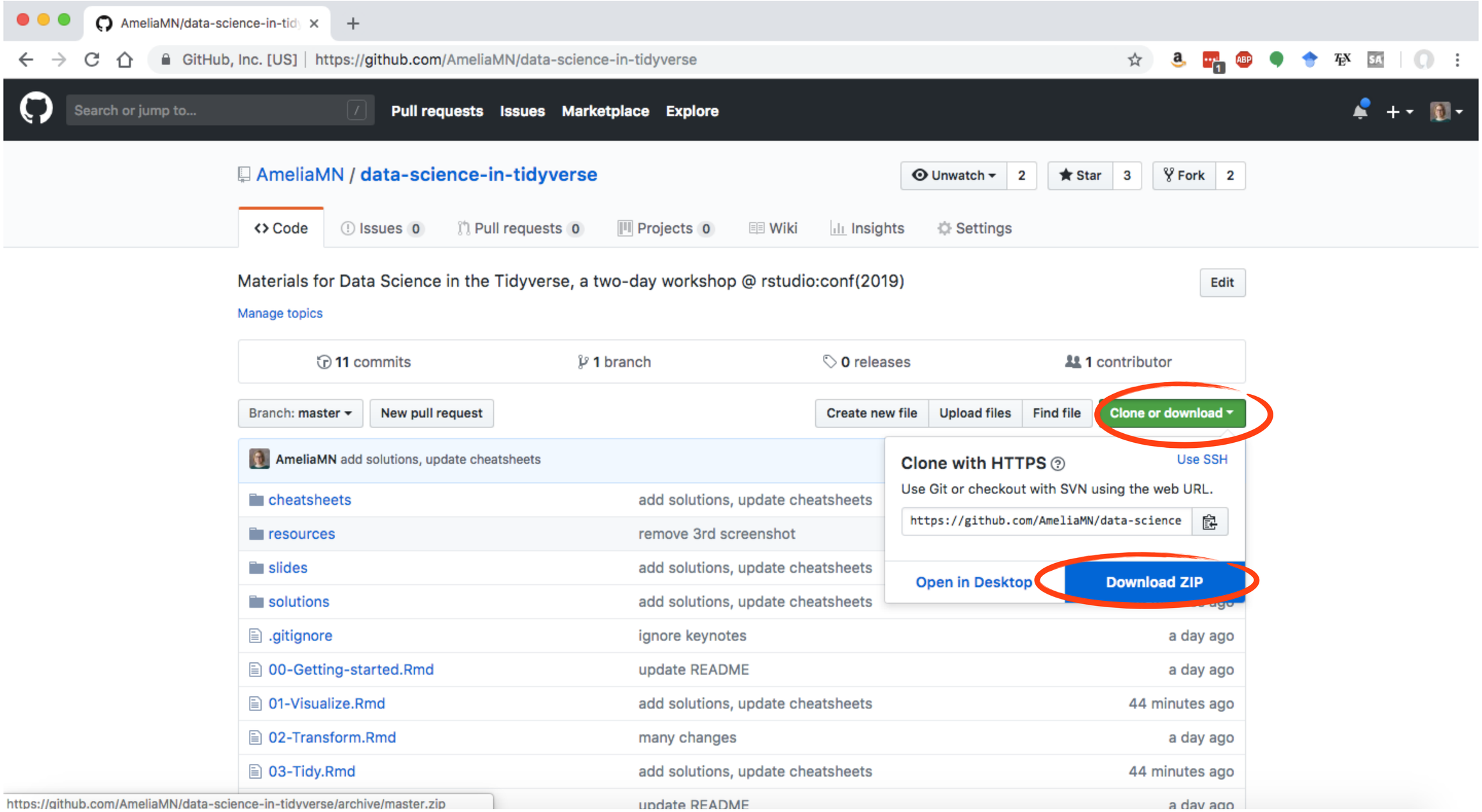
And more

```
install.packages(c("babynames", "fivethirtyeight", "formatR",
"gapminder", "hexbin", "mgcv", "maps", "mapproj", "nycflights13",
"rmarkdown", "skimr", "tidyverse", "viridis"))
```

Getting our code


RStudio Cloud
Amelia
Secure | https://rstudio.cloud/project/13632
Your Workspace / Intro to R & RStudio (day 2)
Amelia McNamara
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 3.4.2
nimbus
GSS
GSS1
band
Filter
Year ID LaborStatus MaritalStatus NumChildren Age HighestSchoolCompleted
1 2014 1 Working fulltime Divorced 0 53.000000 16
2 2014 2 Working fulltime Married 0 26.000000 16
3 2014 3 Unempl, laid off Divorced 1 59.000000 13
4 2014 4 Working parttime Married 2 56.000000 16
5 2014 5 Retired Married 3 74.000000 17
6 2014 6 Working fulltime Married 1 56.000000 17
7 2014 7 No answer Married 2 63.000000 12
Showing 1 to 8 of 2,540 entries
Console Terminal
~/
The downloaded source packages are in
'/tmp/RtmpbwU0xs/downloaded_packages'
> library(readr)
> band <- read_csv("project/data/band.csv")
Parsed with column specification:
cols(
name = col_character(),
band = col_character()
)
> View(band)
>
Environment History Connections
Import Dataset
List
Global Environment
band 3 obs. of 2 variables
GSS 2540 obs. of 15 variables
GSS1 2540 obs. of 15 variables
nimbus 18963 obs. of 4 variables
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Home
Name
project
R
Copy...
Copy To...
Move...
Export...
Set As Working Directory
Go To Working Directory
You can export an entire directory from RStudio cloud

Or, download a clean version from https://github.com/AmeliaMN/data-science-in-tidyverse



More packages

(Applied) **Data Science**

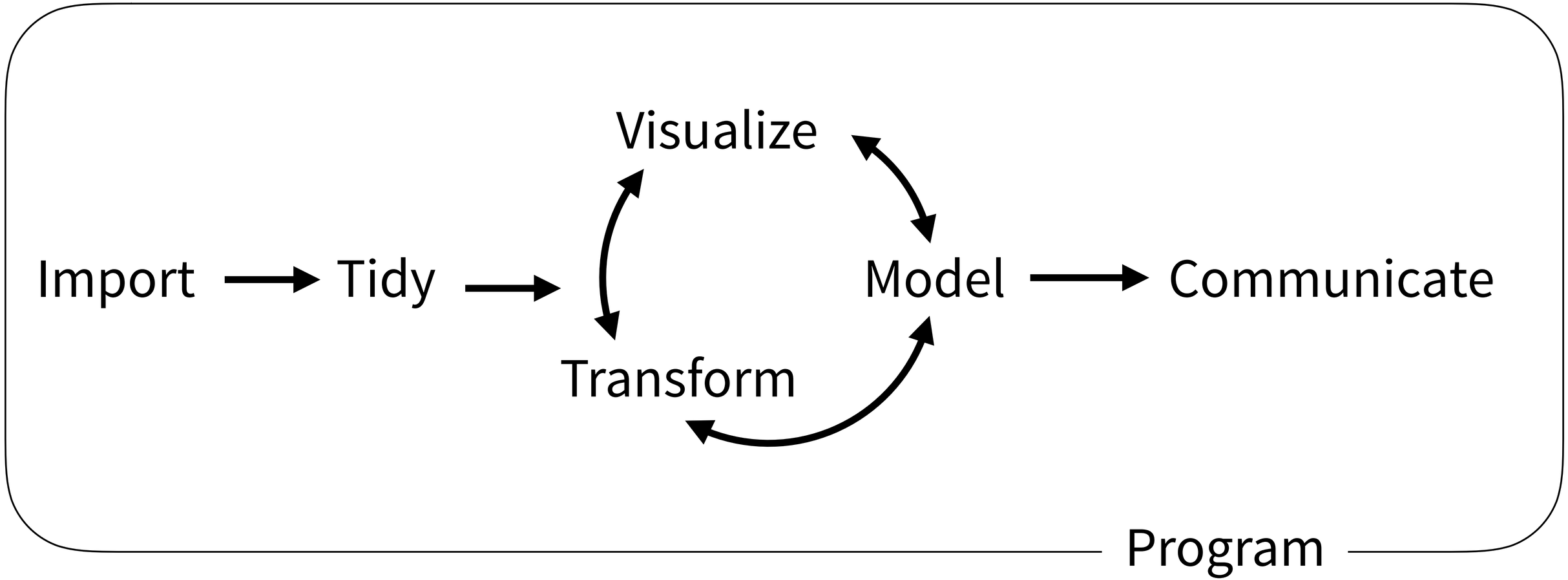


Table of contents

Review things we've covered

Table of contents

Generally useful things

Getting help

Searching for help

- Official word on learning more: https://www.tidyverse.org/learn/
- We've seen the R help functions `?` and `help()`
- Google, putting in R as a search term (Google recognizes it now!)
- Search on http://stackoverflow.com/ (add keywords like tidyverse)

Physical communities

- There are R meetups in many major cities
- If you are a gender minority, check out R-ladies meetups

Online communities

- R4DS learning community: https://medium.com/@kierisi/r4ds-the-next-iteration-d51e0a1b0b82
- https://community.rstudio.com/ is intentionally friendly to beginners!
- Asking on http://stackoverflow.com/ is perhaps an intermediate skill
- I don't recommend asking on R-help
- Official word on asking for help: https://www.tidyverse.org/help/

!!

Thanks to my fantastic TAs

@kierisi

@i_steves

@baumerben

Thank you!