

# Data Science in the tidyverse

Amelia McNamara

[amelia.mn](#)

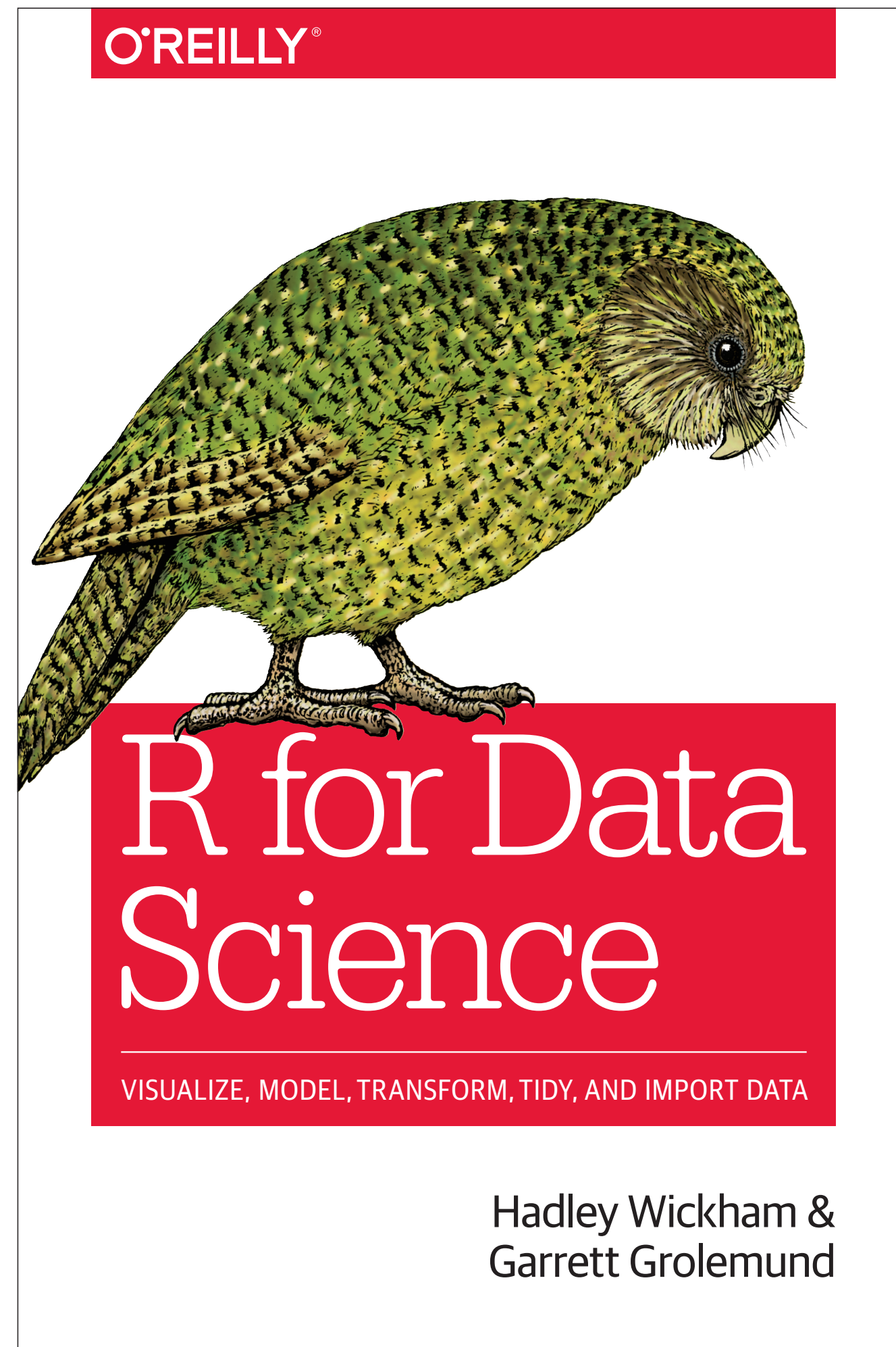
@AmeliaMN



*Data Science in the tidyverse* is licensed under a [Creative Commons Attribution 4.0 International License](#). Based on work at <https://github.com/cwickham/data-science-in-tidyverse> and <https://github.com/rstudio-education/master-the-tidyverse>

Online at:

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

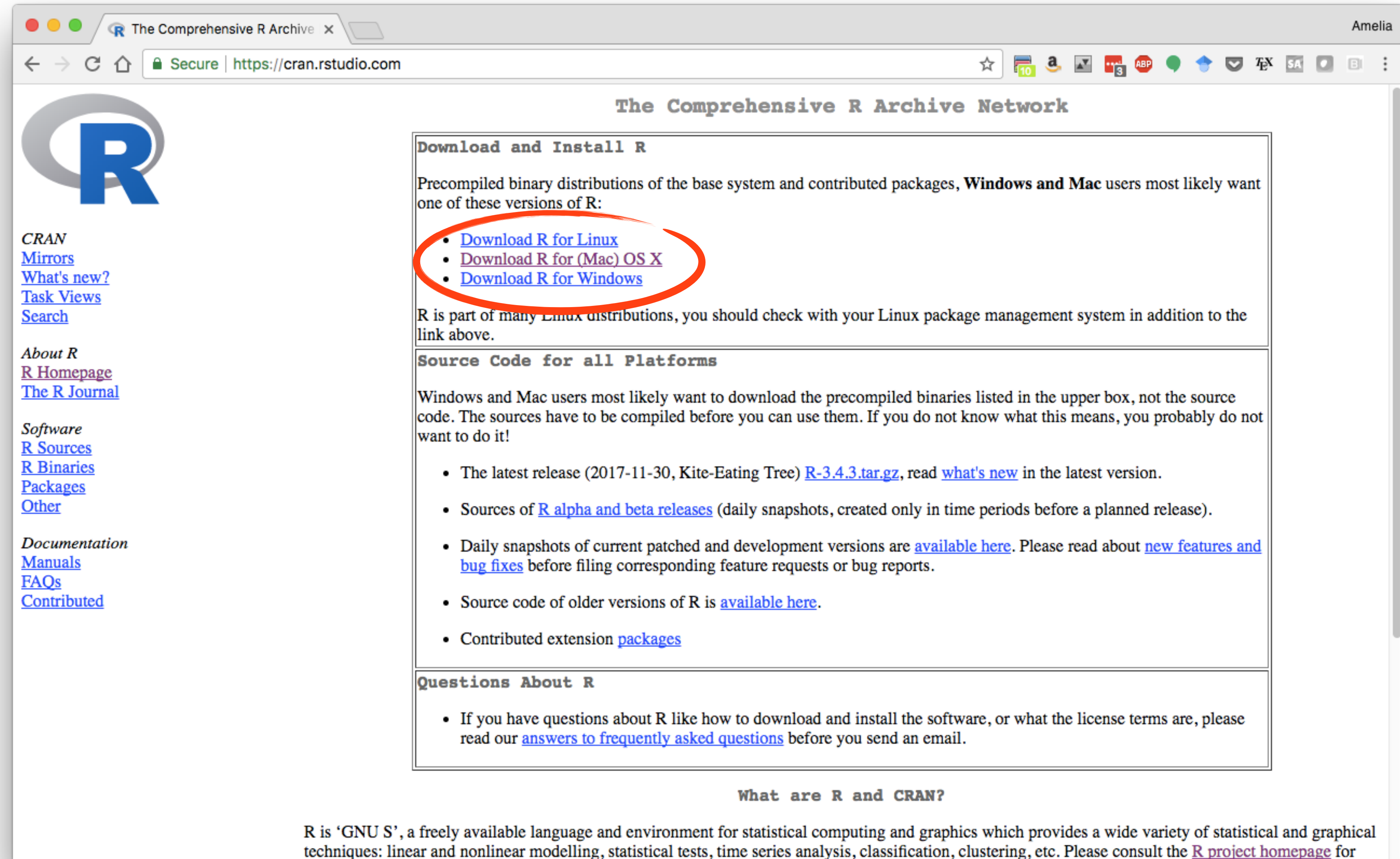


Installing  
locally



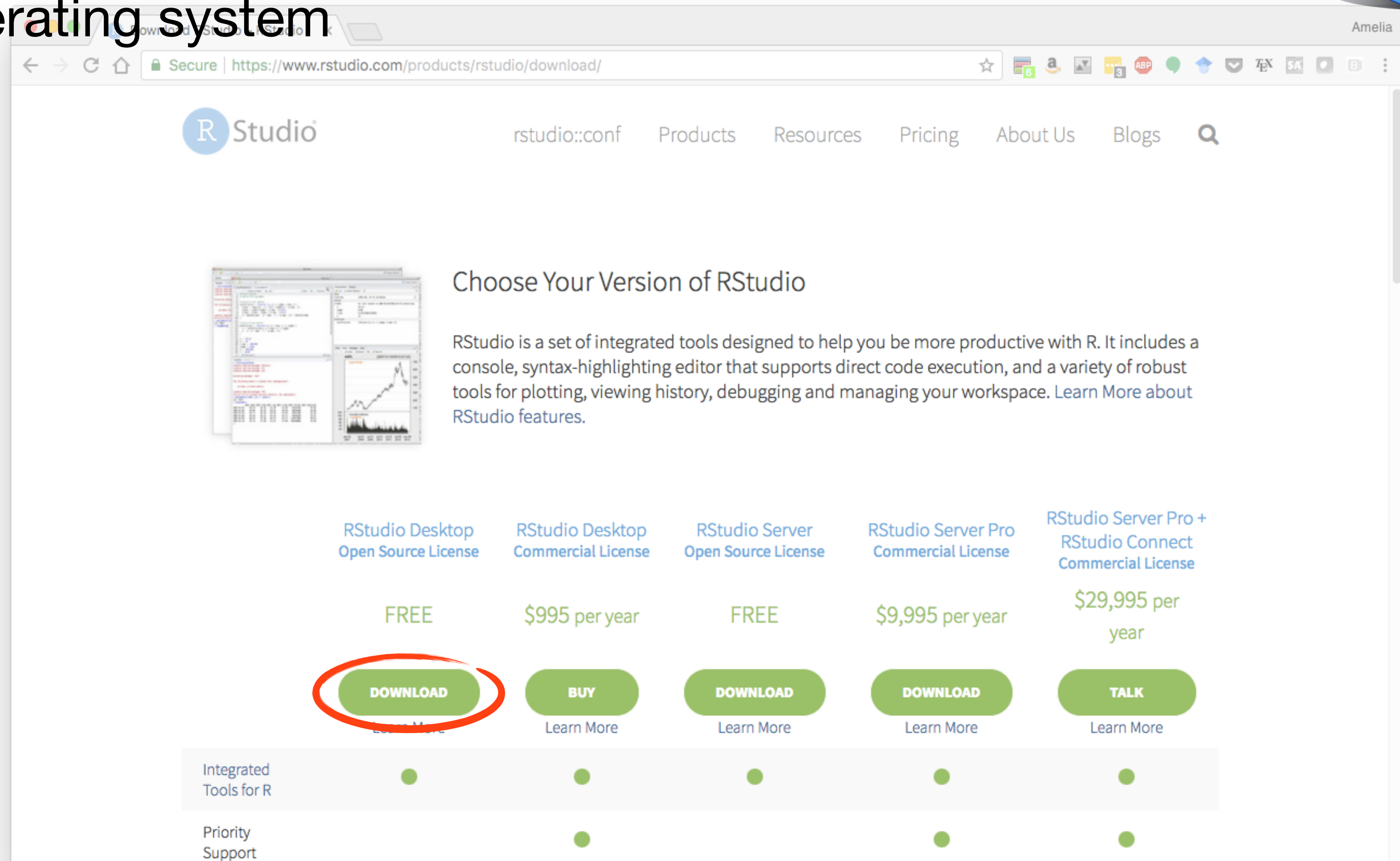
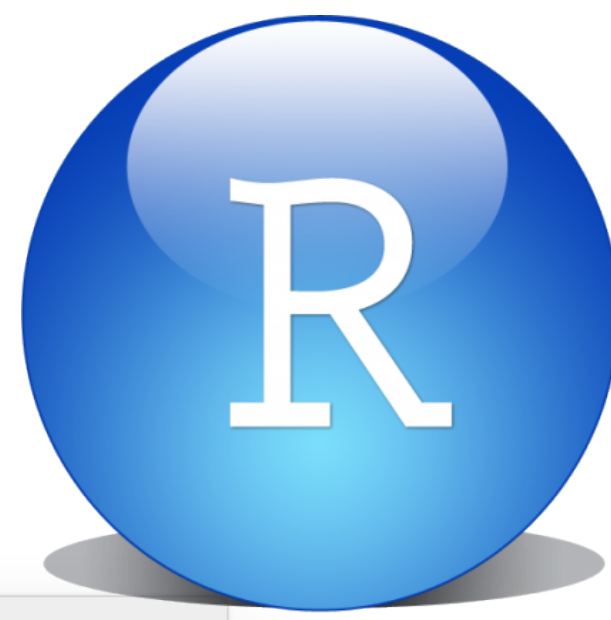
First, you will need to install R (the programming language).

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

A screenshot of the CRAN (Comprehensive R Archive Network) website. The browser window shows the URL https://cran.rstudio.com/. The page has a sidebar on the left with links for CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Other, Documentation, Manuals, FAQs, and Contributed. The main content area is titled 'The Comprehensive R Archive Network' and contains several sections: 'Download and Install R' with a list of links for Linux, Mac OS X, and Windows (the first three are circled in red); 'Source Code for all Platforms' with a list of links for the latest release, alpha and beta releases, daily snapshots, older versions, and contributed packages; and 'Questions About R' with a link to frequently asked questions. At the bottom, there is a section titled 'What are R and CRAN?' with a paragraph explaining R as a freely available language and environment for statistical computing and graphics.

Then, install RStudio (the application).

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

A screenshot of the RStudio website's download page. The browser address bar shows 'https://www.rstudio.com/products/rstudio/download/'. The page features the RStudio logo and navigation links. A section titled 'Choose Your Version of RStudio' includes a description of the software and a 'Learn More' link. Below this, five product options are listed with their respective prices and download buttons. The 'RStudio Desktop Open Source License' option is highlighted with a red circle around its 'DOWNLOAD' button. A table at the bottom compares features across the different versions.

	RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
	<b>DOWNLOAD</b> <a href="#">Learn More</a>	<b>BUY</b> <a href="#">Learn More</a>	<b>DOWNLOAD</b> <a href="#">Learn More</a>	<b>DOWNLOAD</b> <a href="#">Learn More</a>	<b>TALK</b> <a href="#">Learn More</a>
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●



Then, install RStudio (the application).

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



**RStudio Desktop 1.1.419 — Release Notes**

RStudio requires R 3.0.1+. If you don't already have R, download it [here](#).

**Installers for Supported Platforms**

Installers	Size	Date	MD5
RStudio 1.1.419 - Windows Vista/7/8/10	85.8 MB	2018-01-24	7f71f68fb45a6c8d3d2898096ca6fe91
RStudio 1.1.419 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-01-24	92f16f2d5b95e178a78fab1a0e606e3d
RStudio 1.1.419 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-01-24	12bf107ef92b2fcab418263519b4bf6d
RStudio 1.1.419 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-01-24	830f5e5954e802d1b93515c8dffcea05
RStudio 1.1.419 - Ubuntu 16.04+/Debian 9+ (64-bit)	64.9 MB	2018-01-24	a090284b0401c7d8bbc474f227342932
RStudio 1.1.419 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-01-24	4b6949c0f55d7d1f1c741a19aa089064
RStudio 1.1.419 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-01-24	f660a6d2f2540a7bb2a823cadb8f9bbd

**Zip/Tarballs**

Zip/tar archives	Size	Date	MD5
RStudio 1.1.419 - Windows Vista/7/8/10	122.9 MB	2018-01-24	18fec780a05560df2b50af6c7c883649
RStudio 1.1.419 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-01-24	df8a98b238b98e13eba09d3f19e4cce4
RStudio 1.1.419 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-01-24	9a862b426123d88b2009ad2ade382b17
RStudio 1.1.419 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-01-24	49be9e9eee1b040a107886e52fd9aacb
RStudio 1.1.419 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-01-24	8f1d1f7616f295780816cc6e7d027d5b

**Source Code**

A tarball containing source code for RStudio v1.1.419 can be downloaded from [here](#)

# 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

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	Sex
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
8	2014	8	Working fulltime	Married	2	74.000000		17

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>

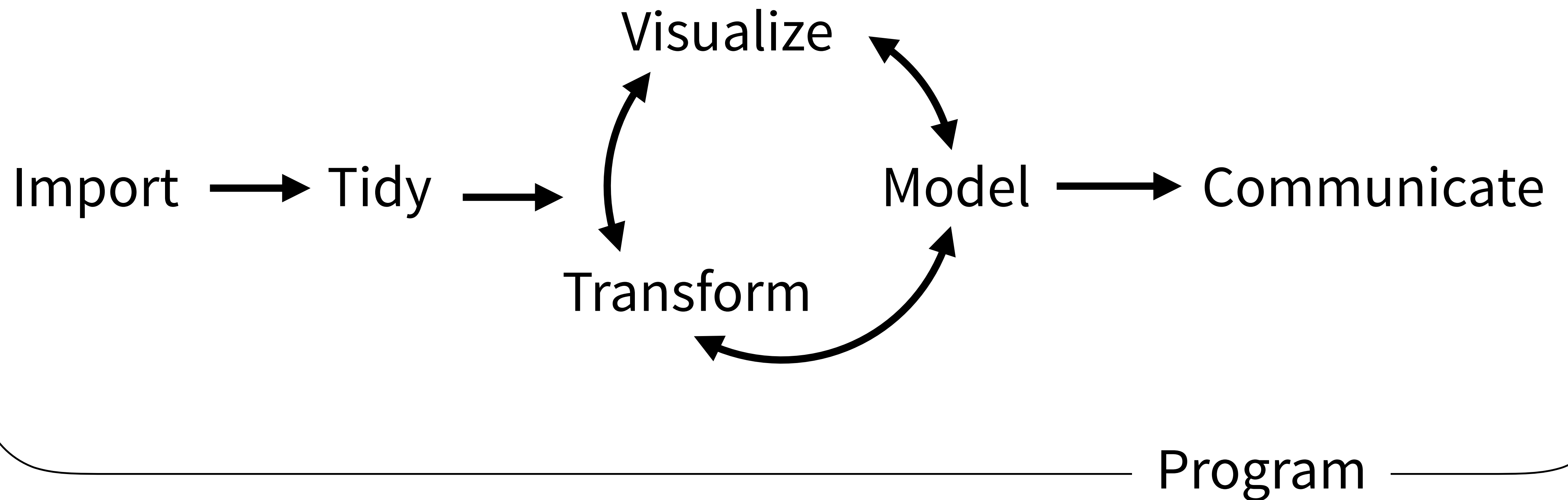
The screenshot shows the GitHub repository page for 'AmeliaMN / data-science-in-tidyverse'. The repository is described as 'Materials for Data Science in the Tidyverse, a two-day workshop @ rstudio:conf(2019)'. It has 11 commits, 1 branch, 0 releases, and 1 contributor. The 'Clone or download' button is highlighted with a red circle, and the 'Download ZIP' button in the dropdown menu is also highlighted with a red circle. The file list includes folders like 'cheatsheets', 'resources', 'slides', and 'solutions', and files like '.gitignore', '00-Getting-started.Rmd', '01-Visualize.Rmd', '02-Transform.Rmd', and '03-Tidy.Rmd'.

File/Folder	Commit Message	Time Ago
cheatsheets	add solutions, update cheatsheets	
resources	remove 3rd screenshot	
slides	add solutions, update cheatsheets	
solutions	add solutions, update cheatsheets	
.gitignore	ignore keynotes	a day ago
00-Getting-started.Rmd	update README	a day ago
01-Visualize.Rmd	add solutions, update cheatsheets	44 minutes ago
02-Transform.Rmd	many changes	a day ago
03-Tidy.Rmd	add solutions, update cheatsheets	44 minutes ago
	update README	a day ago



More  
packages

# (Applied) Data Science



## R for Data Science

### Welcome

**1** Introduction

### I Explore

**2** Introduction

**3** Data visualisation

**4** Workflow: basics

**5** Data transformation

**6** Workflow: scripts

**7** Exploratory Data Analysis

**8** Workflow: projects

# Table of contents

## II Wrangle

**9** Introduction

**10** Tibbles

**11** Data import

**12** Tidy data

**13** Relational data

**14** Strings

**15** Factors

**16** Dates and times

## III Program

**17** Introduction

**18** Pipes

**19** Functions

**20** Vectors

**21** Iteration

## IV Model

**22** Introduction

**23** Model basics

**24** Model building

**25** Many models

## V Communicate

**26** Introduction

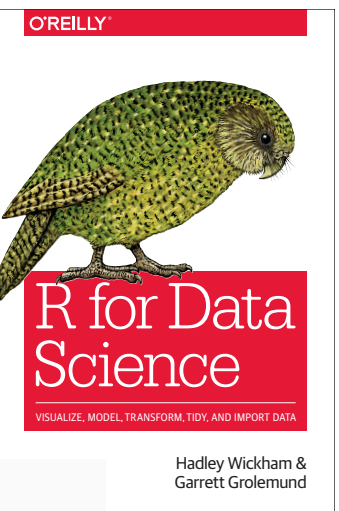
**27** R Markdown

**28** Graphics for communication

**29** R Markdown formats

**30** R Markdown workflow

**Review things  
we've covered**





## R for Data Science

### Welcome

#### 1 Introduction

### I Explore

#### 2 Introduction

#### 3 Data visualisation

#### 4 Workflow: basics

#### 5 Data transformation

#### 6 Workflow: scripts

#### 7 Exploratory Data Analysis

#### 8 Workflow: projects

# Table of contents

## II Wrangle

### 9 Introduction

### 10 Tibbles

### 11 Data import

### 12 Tidy data

### 13 Relational data

### 14 Strings

### 15 Factors

### 16 Dates and times

## III Program

### 17 Introduction

### 18 Pipes

### 19 Functions

### 20 Vectors

### 21 Iteration

## IV Model

### 22 Introduction

### 23 Model basics

### 24 Model building

### 25 Many models

## V Communicate

### 26 Introduction

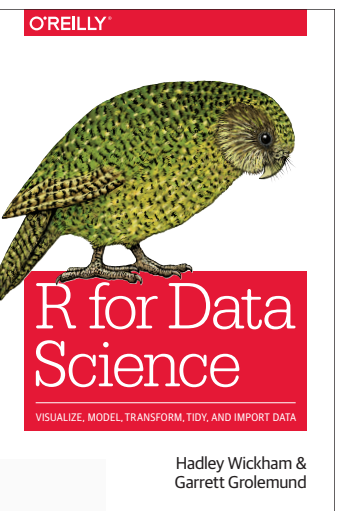
### 27 R Markdown

### 28 Graphics for communication

### 29 R Markdown formats

### 30 R Markdown workflow

**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!