Meet the Toolkit

K Arnold, based on IntroDS.org

Everyone should be
  quick to listen,
  slow to speak and
  slow to become angry,
because human anger does not produce the righteousness that God desires.

James 1:19-20

What disagreements have you had recently?

So far...

- Hands on practice with R, RStudio, Git, GitHub
- First look at visualizing and summarizing data in R
- Why summary statistics alone are not sufficient for data exploration

Logistics

- Discussion 1 due tomorrow (responses due next week)
- Moodle activity completion checklist
- Try a different technology for lab on Friday
- Log your daily temperature?

Any questions?

Reproducible data analysis

Reproducibility checklist

What does it mean for a data analysis to be "reproducible"?

Reproducibility checklist

What does it mean for a data analysis to be "reproducible"?
Near-term goals:

- Can you re-make all tables and figures easily?
- Does the code actually do what you think it does?
- Is it clear **why** decisions were made? (e.g., how were parameter settings chosen?)

Long-term goals:

- Can the code be used for other data?
- Can you extend the code to do other things?

Toolkit

- Scriptability $\rightarrow$ R
- Literate programming (code, narrative, output in one place) $\rightarrow$ R Markdown
- Version control $\rightarrow$ Git / GitHub

Toolkit overview

The whole game

🧶 **knit**
✅ **commit**
⬆️ **push**

R

R

R can be used as a calculator.

```
8738787213 / 1653
```

```
## [1] 5286623
```

Most common data type: "data frames". Each row = one *observation*. Each *column* = one *variable*.

```
mtcars
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1

R

- We use the $ operator to access a variable within a data frame.

```
mtcars$mpg
```

```
##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7
## [31] 15.0 21.4
```

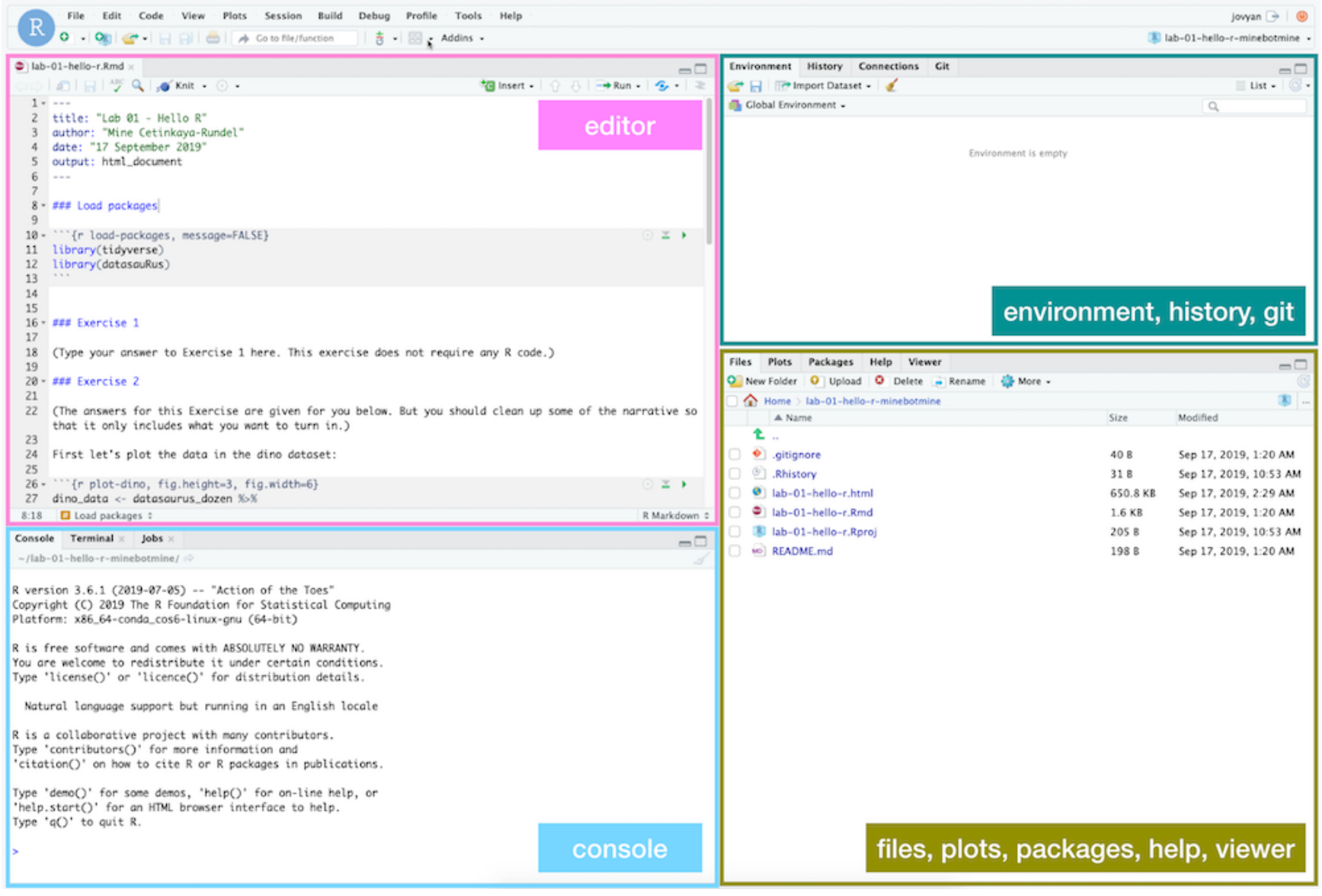
- Functions are (often) verbs, followed by what they will be applied to in parentheses.

```
do_this(to_this)
do_that(to_this, to_that, with_those)
```

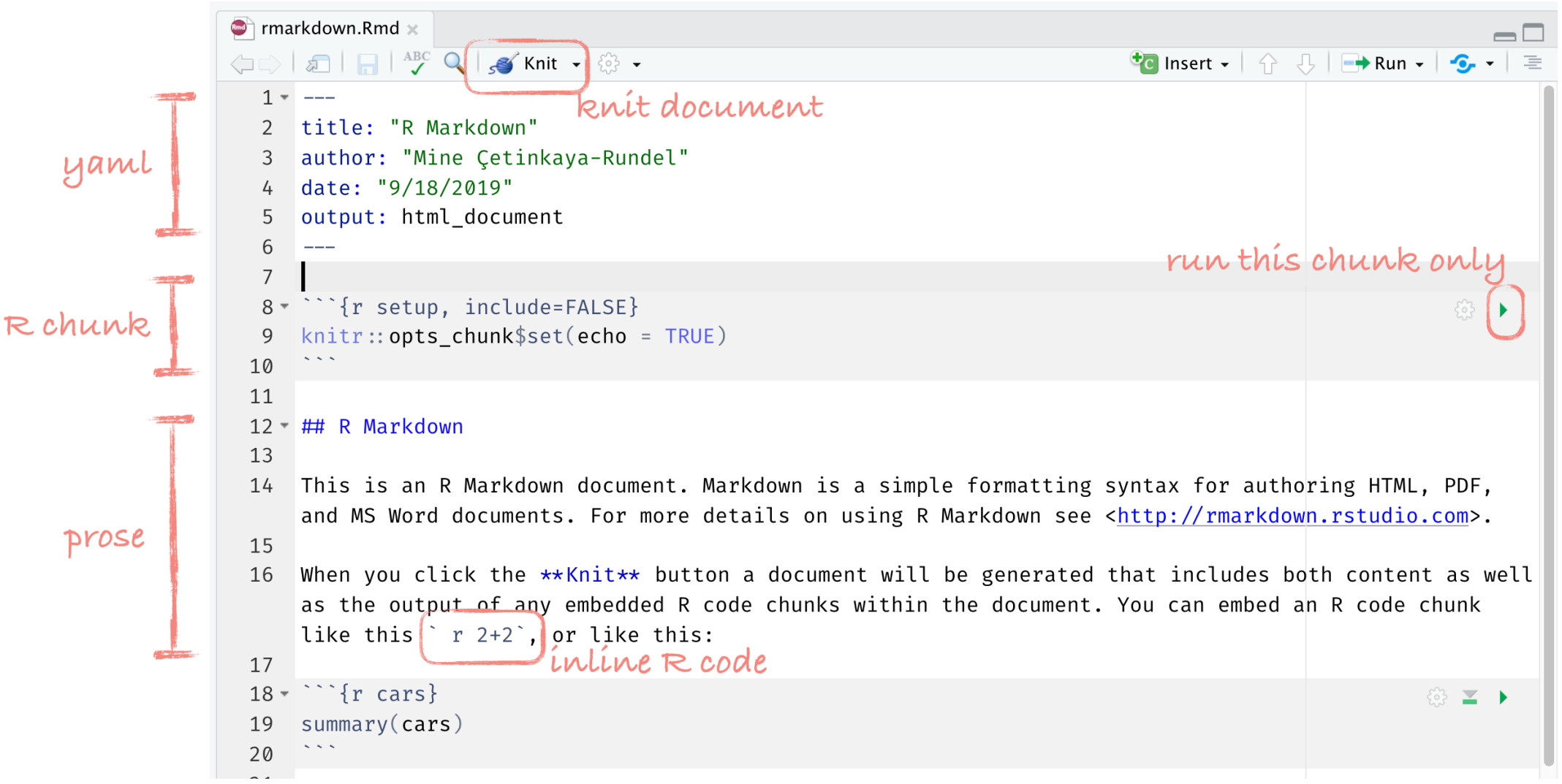
R

- "Package" = shareable code (aka *library* in other languages)
- Over 16k packages on CRAN (**C**omprehensive **R** **A**rchive **N**etwork)
  - Good chance someone already solved a problem that you're working on, and you can benefit from their work by downloading their package.
- Using R packages:
  - Install them from CRAN with `install.packages("x")`
  - Use them in R with `library(x)`
  - Get help on them by clicking on their name in the Packages list in RStudio

RStudio


File Edit Code View Plots Session Build Debug Profile Tools Help
jovyan
Go to file/function
Addins
lab-01-hello-r-minebotmine
lab-01-hello-r.Rmd
Knit
Insert
Run
editor
1 ---
2 title: "Lab 01 - Hello R"
3 author: "Mine Cetinkaya-Rundel"
4 date: "17 September 2019"
5 output: html_document
6 ---
7
8 ### Load packages
9
10 ```{r load-packages, message=FALSE}
11 library(tidyverse)
12 library(datasauRus)
13 ```
14
15
16 ### Exercise 1
17
18 (Type your answer to Exercise 1 here. This exercise does not require any R code.)
19
20 ### Exercise 2
21
22 (The answers for this Exercise are given for you below. But you should clean up some of the narrative so that it only includes what you want to turn in.)
23
24 First let's plot the data in the dino dataset:
25
26 ```{r plot-dino, fig.height=3, fig.width=6}
27 dino_data <- datasaurus_dozen %>%
8:18
Load packages
R Markdown
Environment History Connections Git
Import Dataset
List
Global Environment
Environment is empty
environment, history, git
Console Terminal Jobs
~/lab-01-hello-r-minebotmine/
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-conda_cos6-linux-gnu (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.
>
console
Files Plots Packages Help Viewer
New Folder
Upload
Delete
Rename
More
Home
lab-01-hello-r-minebotmine
Name
Size
Modified
.gitignore 40 B Sep 17, 2019, 1:20 AM
.Rhistory 31 B Sep 17, 2019, 10:53 AM
lab-01-hello-r.html 650.8 KB Sep 17, 2019, 2:29 AM
lab-01-hello-r.Rmd 1.6 KB Sep 17, 2019, 1:20 AM
lab-01-hello-r.Rproj 205 B Sep 17, 2019, 10:53 AM
README.md 198 B Sep 17, 2019, 1:20 AM
files, plots, packages, help, viewer

R Markdown


rmarkdown.Rmd
Knit
Insert
Run
knit document
yaml
R chunk
prose
run this chunk only
inline R code
1 ---
2 title: "R Markdown"
3 author: "Mine Çetinkaya-Rundel"
4 date: "9/18/2019"
5 output: html_document
6 ---
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
12 ## R Markdown
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16 When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this `r 2+2`, or like this:
18 ```{r cars}
19 summary(cars)
20 ```

R Markdown tips

- Most importantly: environment of your R Markdown document is separate from that of the Console
- Help:
  - R Markdown cheat sheet
  - Markdown Quick Reference (Help -> Markdown Quick Reference)

How will we use R Markdown?

- Every assignment / report / project / etc. is an R Markdown document
- You'll always have a template R Markdown document to start with
- The amount of scaffolding in the template will decrease over the semester

Getting help in R

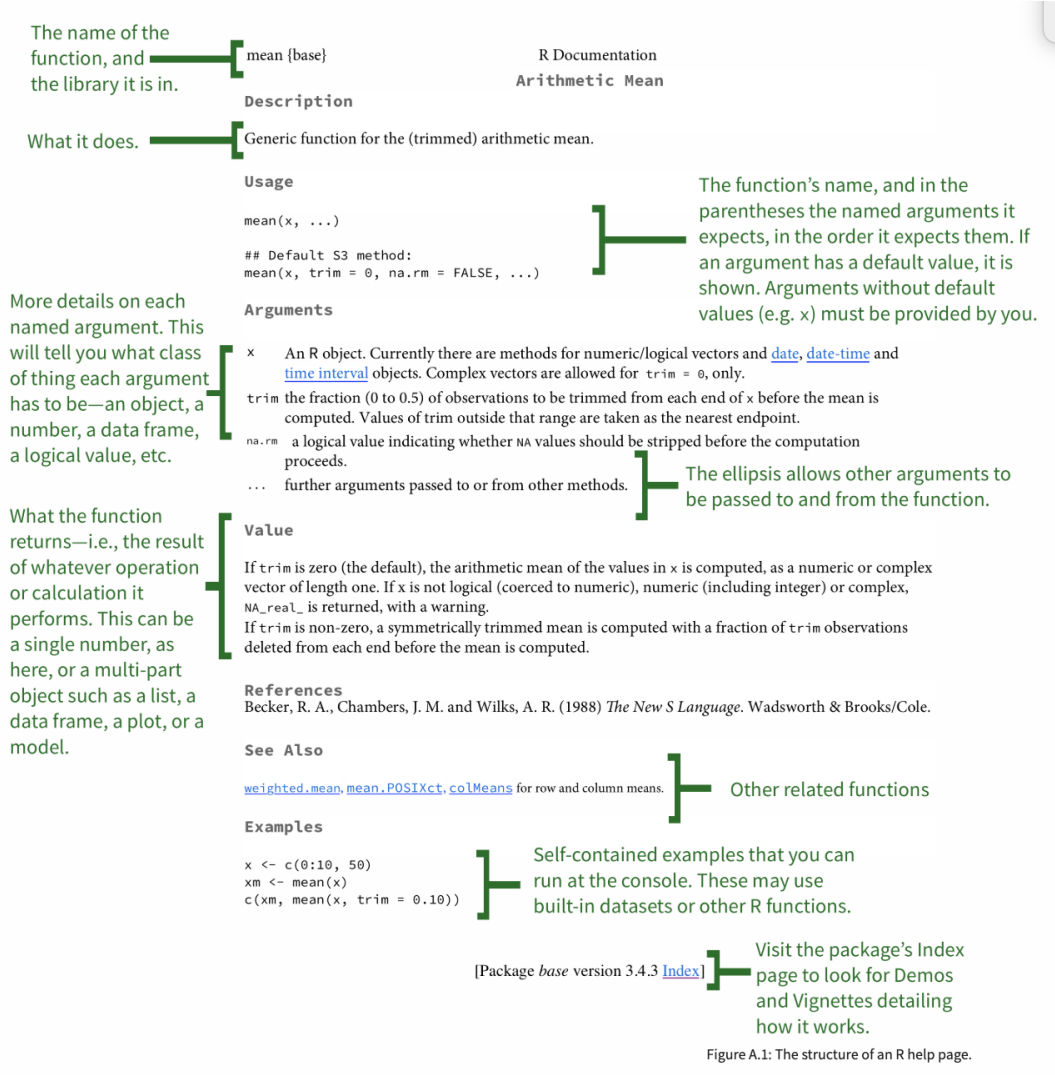
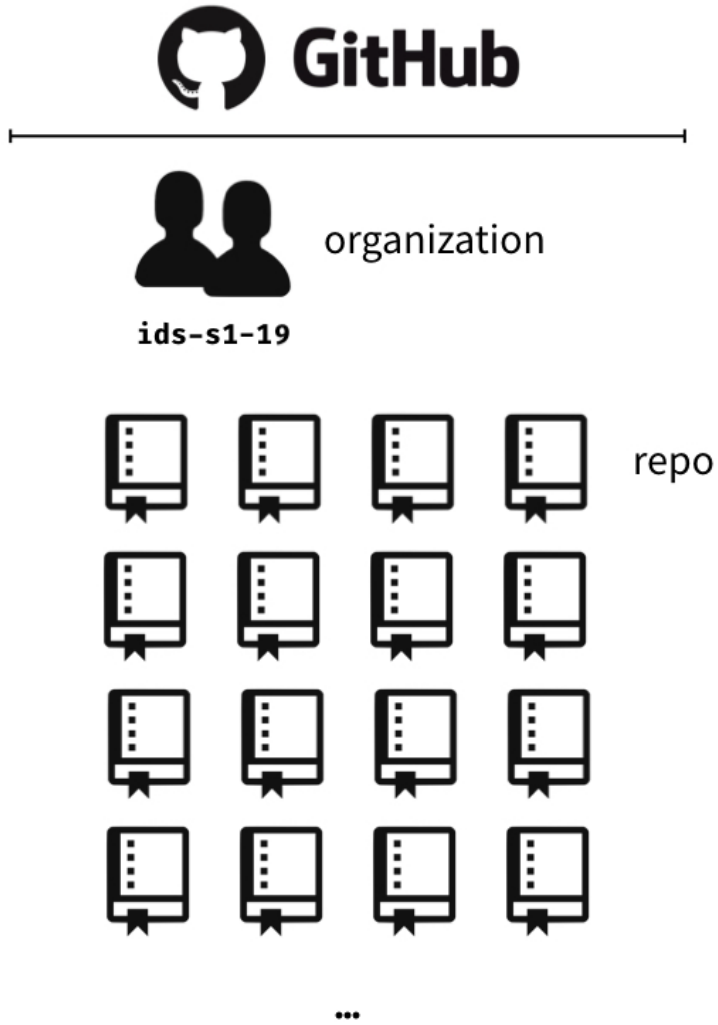
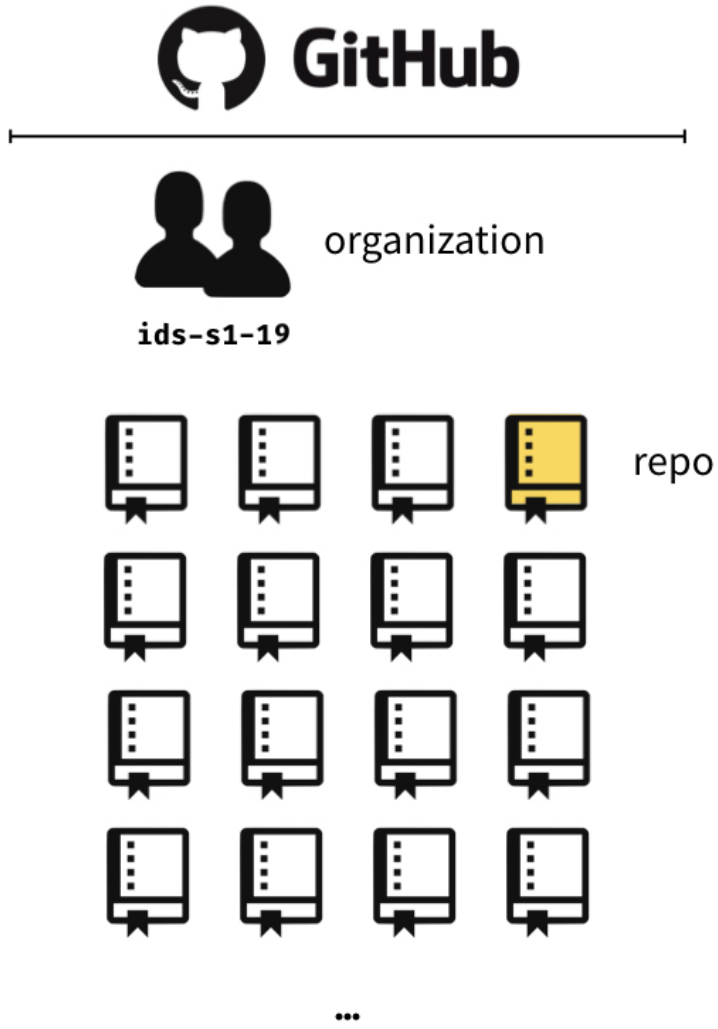
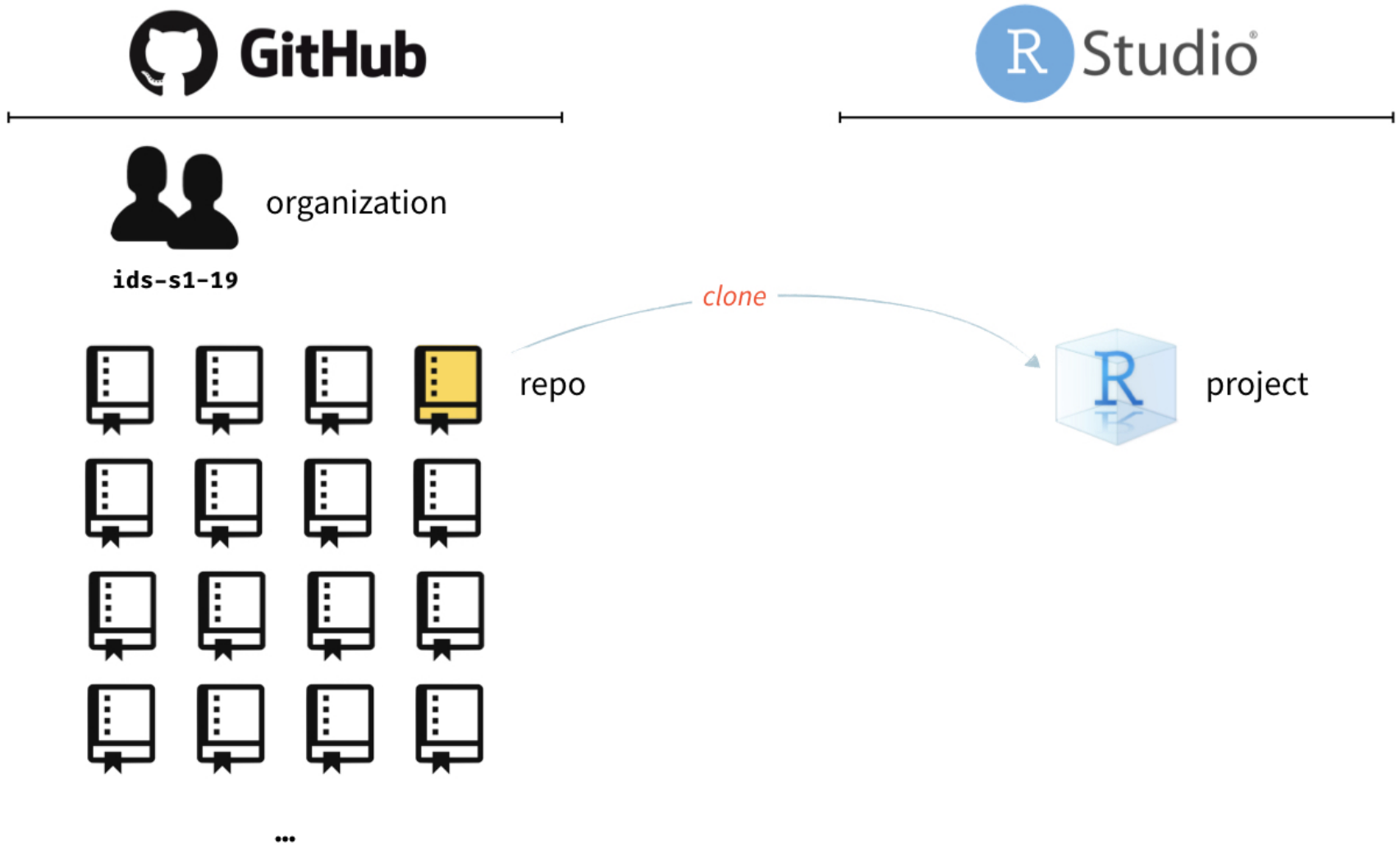


Figure A.1: The structure of an R help page.

Source: http://socviz.co/appendix.html#a-little-more-about-r

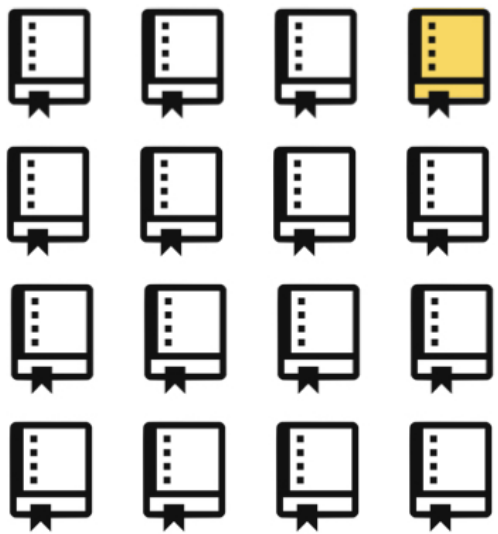

GitHub
organization
ids-s1-19
repo
...


GitHub
organization
ids-s1-19
repo
...


GitHub
RStudio
organization
ids-s1-19
clone
repo
project
...

organization

ids-s1-19



...

repo

clone / pull

push

project

commit

- using the R programming language
- leveraging functionality provided by R packages
- version controlling files with Git

Asking good questions

- Always include your code and the error
- Create a minimum working example (we'll keep working on this throughout the semester)
- Use code formatting