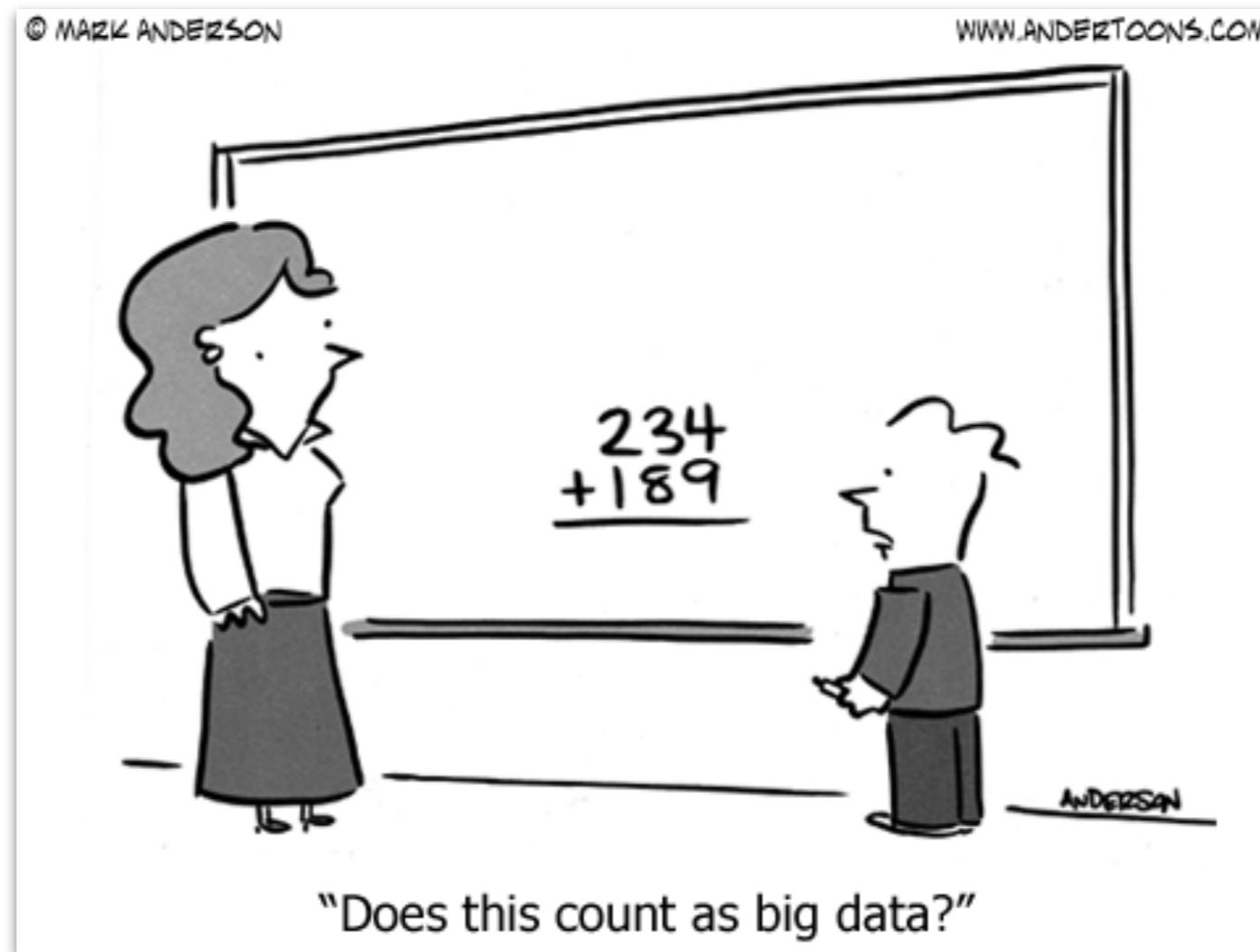


Wrangling data 1



O COLLABORATIVE PLAYLIST

psych252

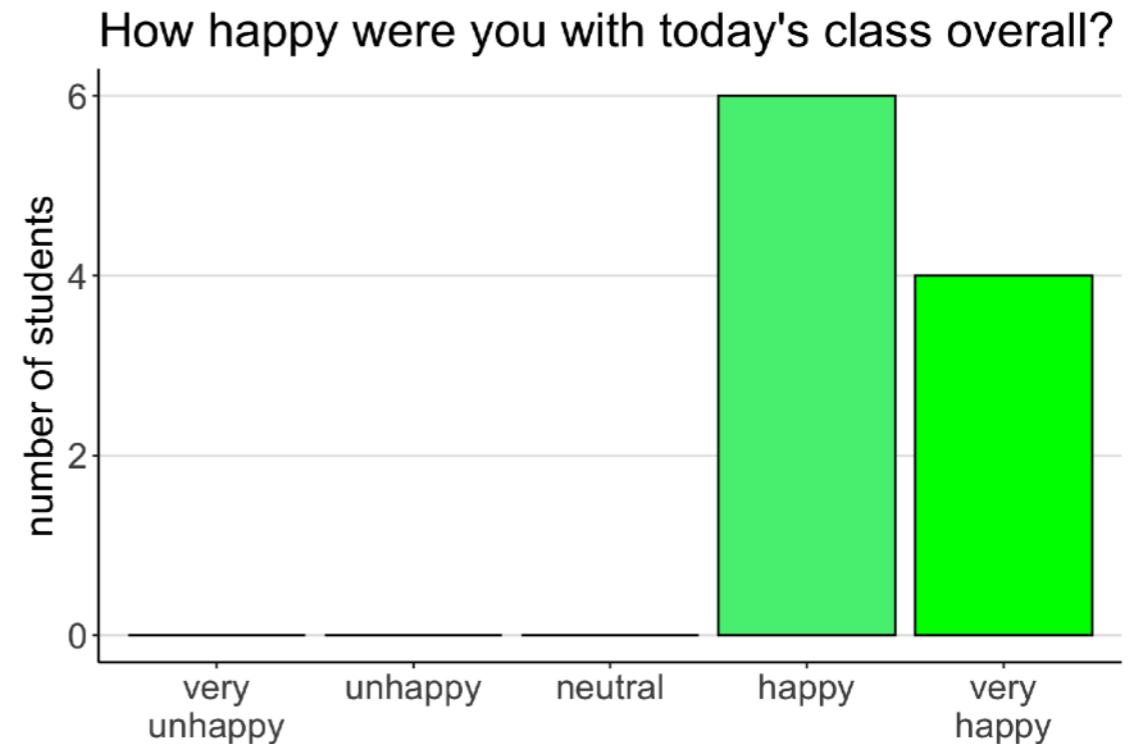
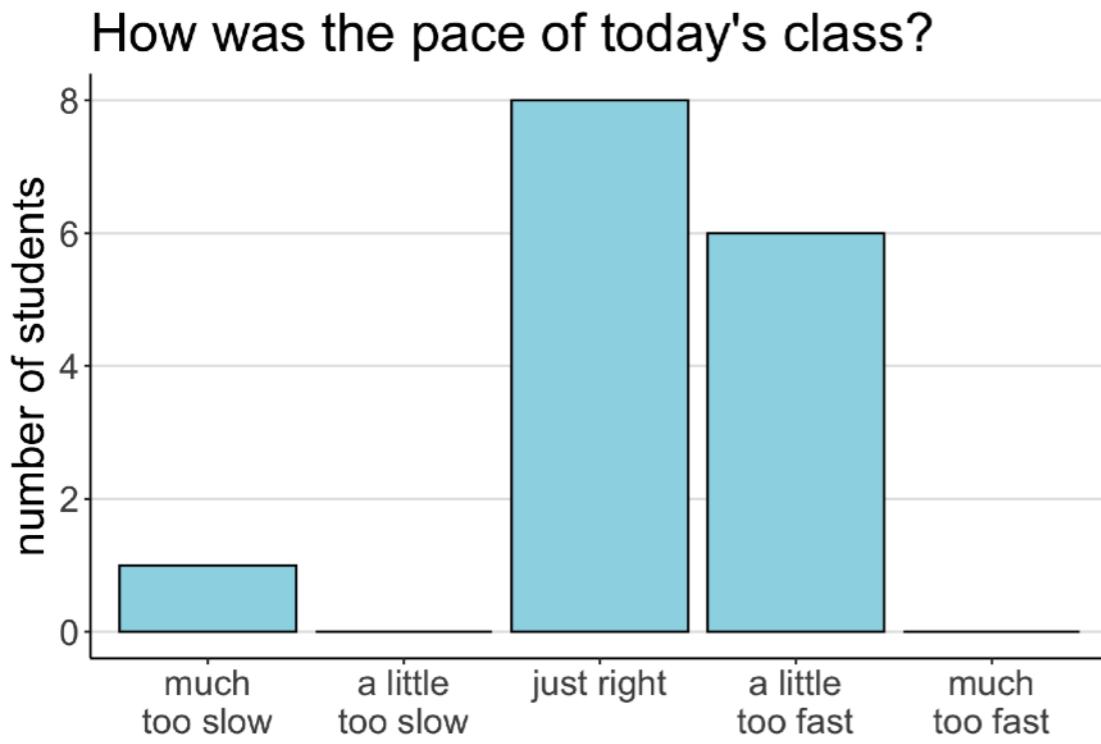
<https://tinyurl.com/psych252spotify24>

PLAY

01/17/2024

Your feedback

Your feedback



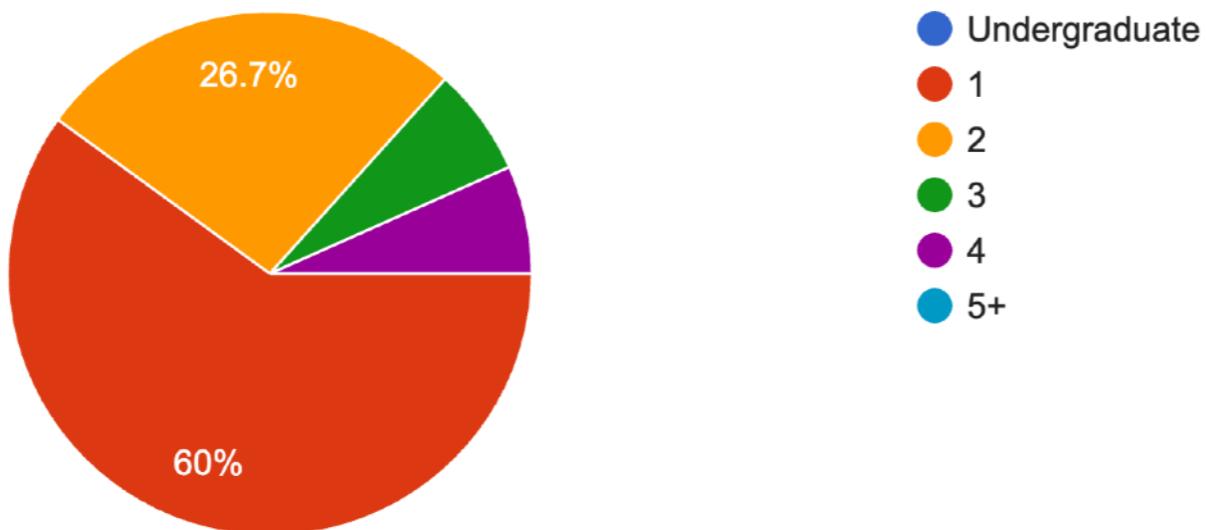
I liked the pace of the class for most of the time except for the last like 10 minutes when we were just jumping thru the markdown file really quickly, I didn't really understand the shiny parts etc. At the end

I think maybe it's better to blurb us on what are the more important contents in the rmd files so that we can better preview contents before class!

Introductory survey

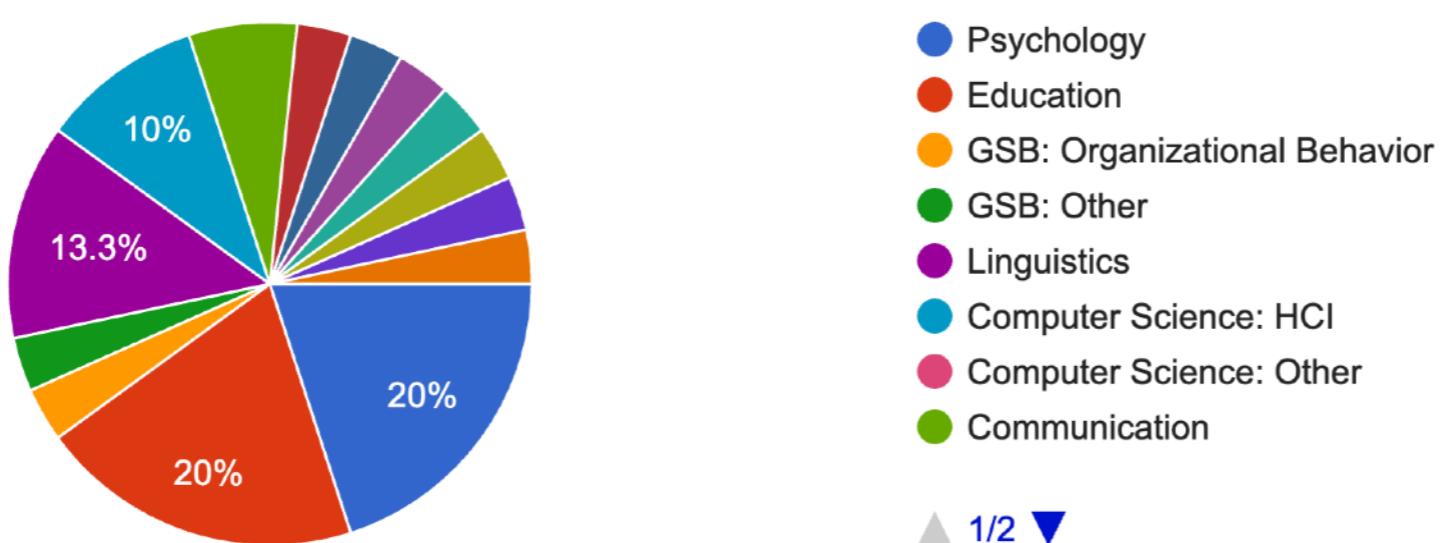
What year of graduate school are you in?

30 responses



What department are you in?

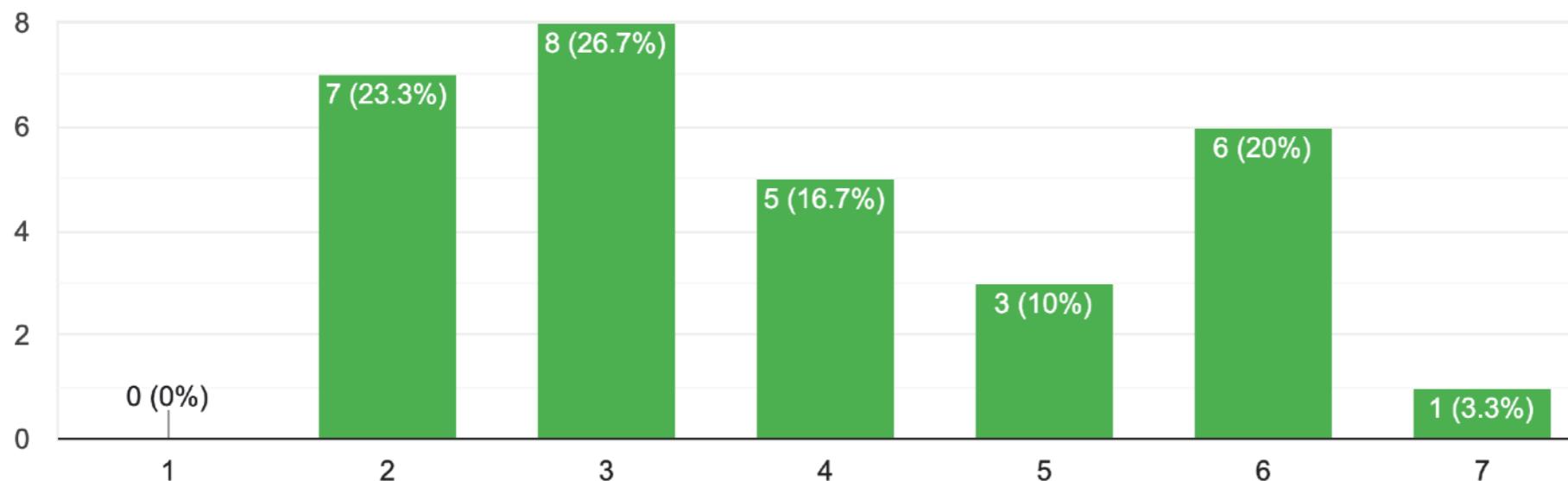
30 responses



Introductory survey

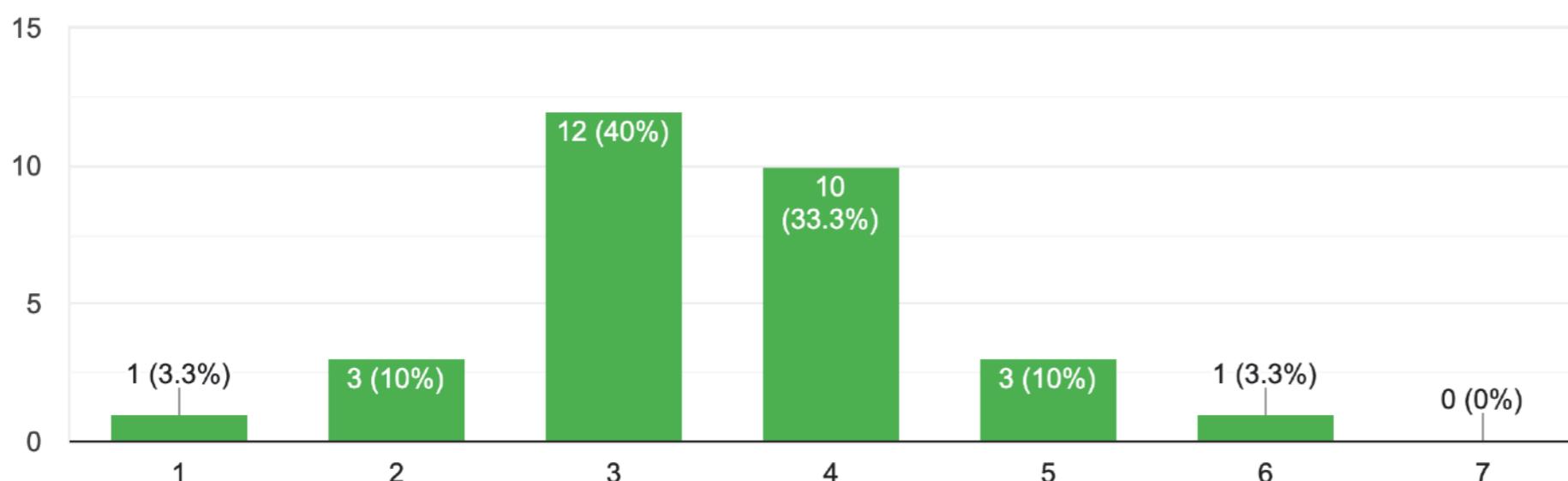
Please rate your level of experience with computer programming

30 responses



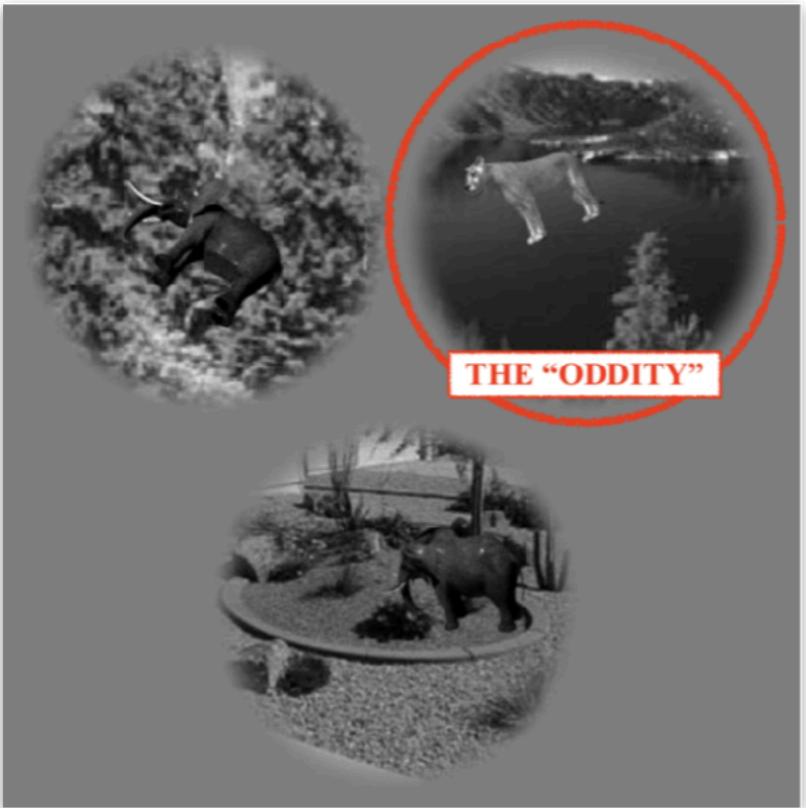
Please rate your level of experience with statistics

30 responses



Logistics

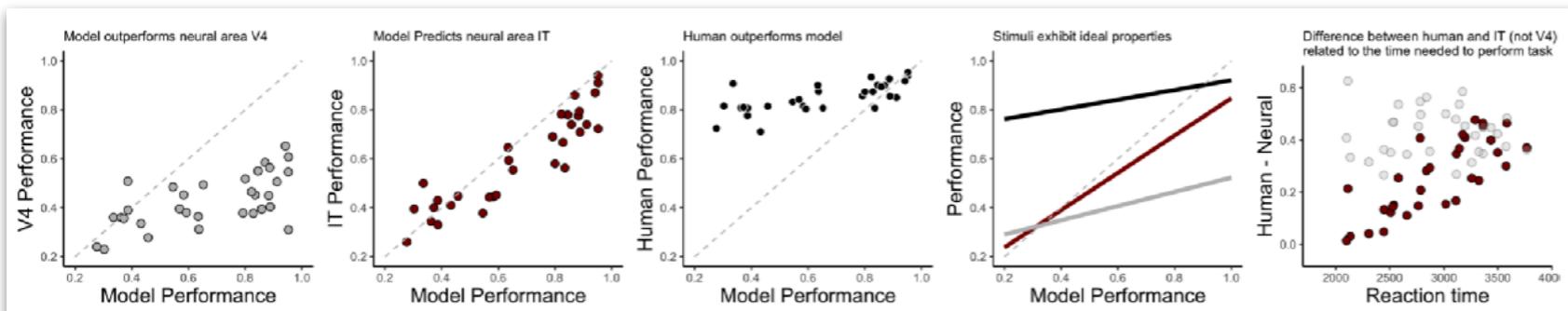
Homework 2



cool experiment

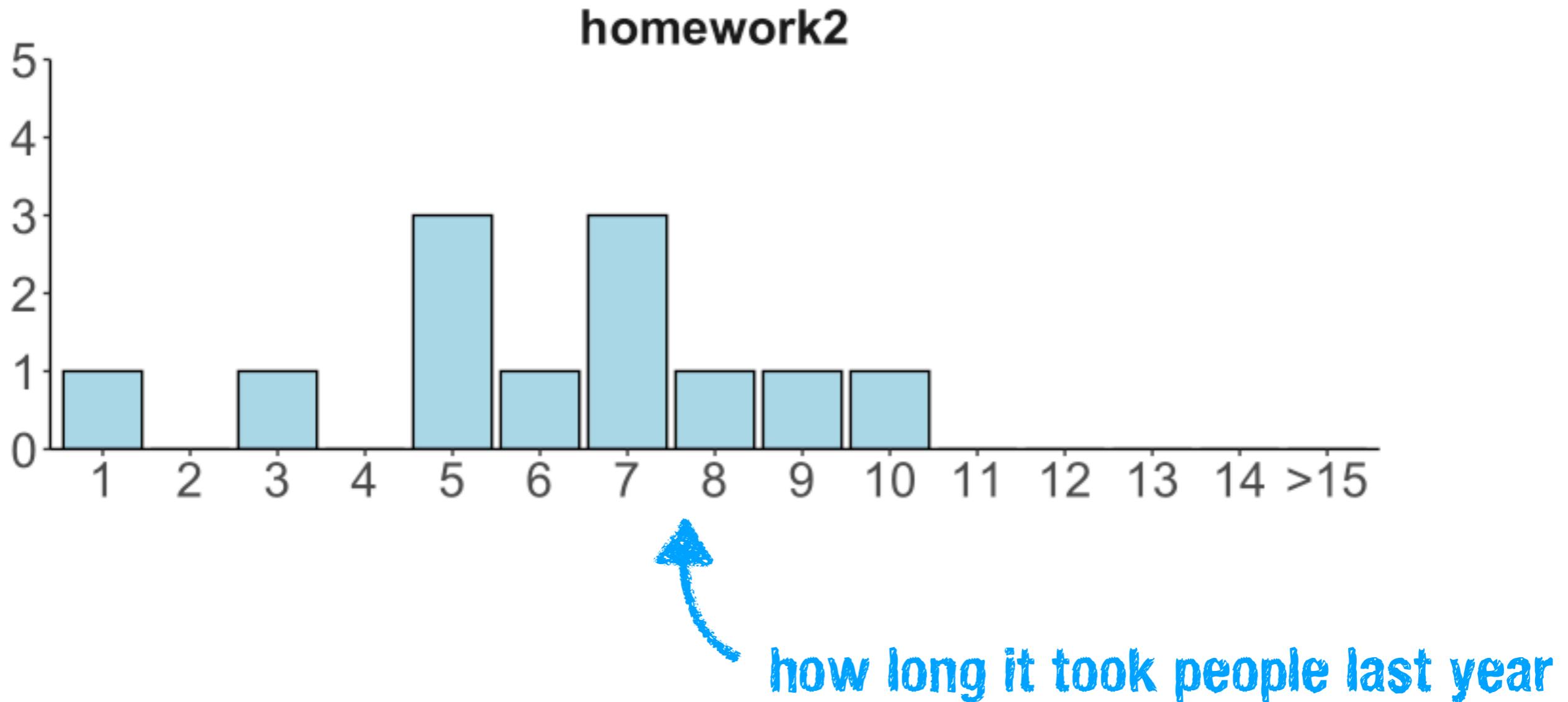
NAME		RELATION Relationship of the member to the head of the family	HOME DATA		FARMING INSCRIPTION		EDUCATION		Place of birth of each of the United States in which he or she was born
NAME	ADDRESS		AGE	SEX	FEAR OF ANIMALS	AGRICULTURE	MANUFACTURE	TRADE	
Ward, George	Head ① Hwy. 29 Rd.	30 M	Male	Yes	Yes	Yes	Yes	No	Taylor
Willie J.	Wife - H	27		Yes	Yes	Yes	Yes	No	Kentucky
George E.	Son	27		Yes	Yes	Yes	Yes	No	Georgia
Lester	Son	6		Yes	Yes	Yes	Yes	No	Georgia
Doris	Son	15		Yes	Yes	Yes	Yes	No	Georgia
Cathleen Elizabeth M.	Daughter	7		Yes	Yes	Yes	Yes	No	Scotland
Tucker, Georgia C.	Head ② Hwy. 29	26 F		Yes	Yes	Yes	Yes	No	Georgia
Norman	Wife of	26		Yes	Yes	Yes	Yes	No	Georgia
Ledgey	Daughter	11		Yes	Yes	Yes	Yes	No	Georgia
Humbry, Earl R.	Head ③ Hwy. 29	30 M		Yes	Yes	Yes	Yes	No	Georgia
Doris B.	Wife - H	27		Yes	Yes	Yes	Yes	No	Georgia
Earl Jr.	Son	10		Yes	Yes	Yes	Yes	No	Georgia
Linda G.	Son	6		Yes	Yes	Yes	Yes	No	Georgia
Johnnie L.	Son	3		Yes	Yes	Yes	Yes	No	Georgia
Mabeline	Daughter	2		Yes	Yes	Yes	Yes	No	Georgia
Linda	Son	1		Yes	Yes	Yes	Yes	No	Georgia
Marion M.	Head ④ Hwy. 29	26 M		Yes	Yes	Yes	Yes	No	Georgia
Lester	Wife - H	26		Yes	Yes	Yes	Yes	No	Georgia
Marion M.	Son	10		Yes	Yes	Yes	Yes	No	Georgia
Lester L.	Daughter	8		Yes	Yes	Yes	Yes	No	Georgia
Cynthia	Son	5		Yes	Yes	Yes	Yes	No	Georgia
Gwendolyn	Son	3		Yes	Yes	Yes	Yes	No	Georgia
Howard, William	Head ⑤ Hwy. 29	26 M		Yes	Yes	Yes	Yes	No	Georgia
Lester	Wife - H	26		Yes	Yes	Yes	Yes	No	Georgia
Edward, George W.	Son	23		Yes	Yes	Yes	Yes	No	Georgia
Doris	Daughter	21		Yes	Yes	Yes	Yes	No	Georgia
Cynthia	Daughter	16		Yes	Yes	Yes	Yes	No	Georgia
Howard, Joe S.	Son	17		Yes	Yes	Yes	Yes	No	Georgia

messy dataset



neat plot!

Homework 2



Homework 2

- Due **Thursday 25th, at 8pm**
- Don't wait until the very last moment to knit your RMarkdown file into a pdf. It may not compile and debugging takes time ...
- You can upload earlier versions of your homework on Canvas and still update until the deadline.
- Get and give help via Ed Discussion!

Homework 2

- **We encourage you to work in groups!**
- It's more fun
- You learn more (learning through explaining)

Homework 2

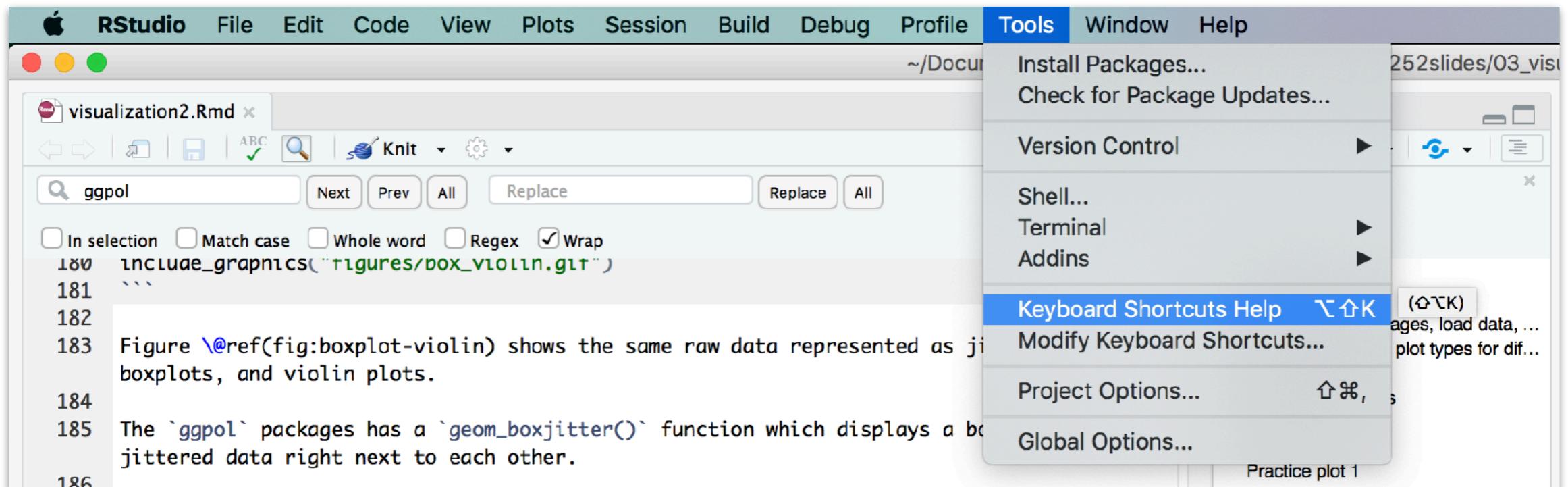
- You can adjust the figure size in your output by using the code chunk options



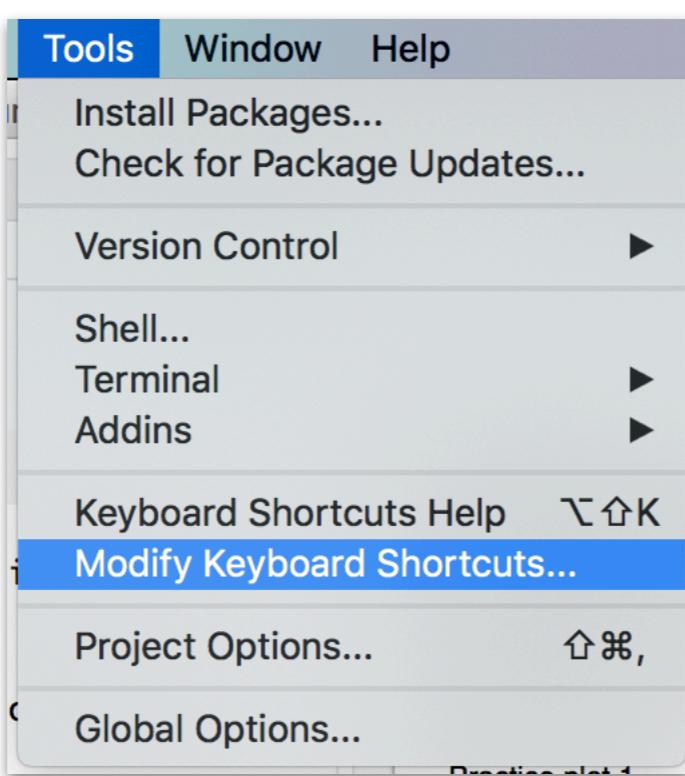
<https://yihui.org/knitr/options/>

Things that came up

Learn the keyboard shortcuts!



... and make
your own

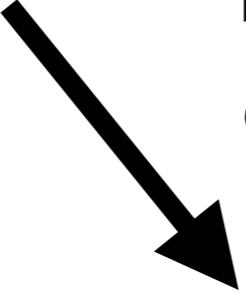


Name	Shortcut	Scope
Browse Addins		Workbench
Check Spelling	F7	Workbench
Check for RStudio Updates		Workbench
Clear All Breakpoints...		Workbench
Clear All Plots...		Workbench
Clear Console	Ctrl+L	Workbench
Clear Knitr Cache		Workbench
Clear Prerendered Output		Workbench
Clear Terminal Buffer		Workbench
Clear Workspace		Workbench
Close All Documents		Workbench
Close Current Document	Cmd+W	Workbench
Close Current Project		Workbench
Close Other Documents	Shift+Alt+Cmd+W	Workbench
Close Terminal		Workbench
Compile Notebook	Shift+Cmd+K	Workbench
Console on Left		Workbench
Console on Right		Workbench
Copy Current Plot to Clipboard...		Workbench

Reformatting code

```
1 ggplot(data = df.diamonds[1:150,], mapping = aes(x = color, y = price)) +  
2   # individual data points (jittered horizontally)  
3   geom_point(alpha = 0.2,  
4             position = position_jitter(width = 0.1, height = 0),  
5             size = 2) +  
6   # error bars  
7   stat_summary(fun.data = "mean_cl_boot",  
8                 geom = "linerange",  
9                 color = "black",  
10                size = 1) +  
11   # means  
12  stat_summary(fun.y = "mean",  
13                 geom = "point",  
14                 shape = 21,  
15                 fill = "red",  
16                 color = "black",  
17                 size = 4)
```

highlight code and press
cmd + i

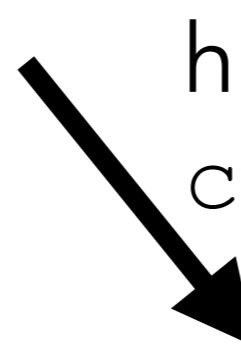


```
1 ggplot(data = df.diamonds[1:150,], mapping = aes(x = color, y = price)) +  
2   # individual data points (jittered horizontally)  
3   geom_point(alpha = 0.2,  
4             position = position_jitter(width = 0.1, height = 0),  
5             size = 2) +  
6   # error bars  
7   stat_summary(fun.data = "mean_cl_boot",  
8                 geom = "linerange",  
9                 color = "black",  
10                size = 1) +  
11   # means  
12  stat_summary(fun.y = "mean",  
13                 geom = "point",  
14                 shape = 21,  
15                 fill = "red",  
16                 color = "black",  
17                 size = 4)
```

Commenting code

```
1 ggplot(data = df.diamonds,  
2         mapping = aes(x = color, y = price)) +  
3         stat_summary(fun.y = "mean", geom = "bar")
```

highlight code and press
cmd + shift + c



```
1 # ggplot(data = df.diamonds,  
2 #           mapping = aes(x = color, y = price)) +  
3 #         stat_summary(fun.y = "mean", geom = "bar")
```

Quickly copying code

```
1 ggplot(mapping = aes(x = color, y = price), data = df.diamonds) +  
2   stat_summary(fun.y = "mean", geom = "point")
```



put cursor anywhere in line 1
cmd + shift + d

```
1 ggplot(mapping = aes(x = color, y = price), data = df.diamonds) +  
2 ggplot(mapping = aes(x = color, y = price), data = df.diamonds) +  
3   stat_summary(fun.y = "mean", geom = "point")
```

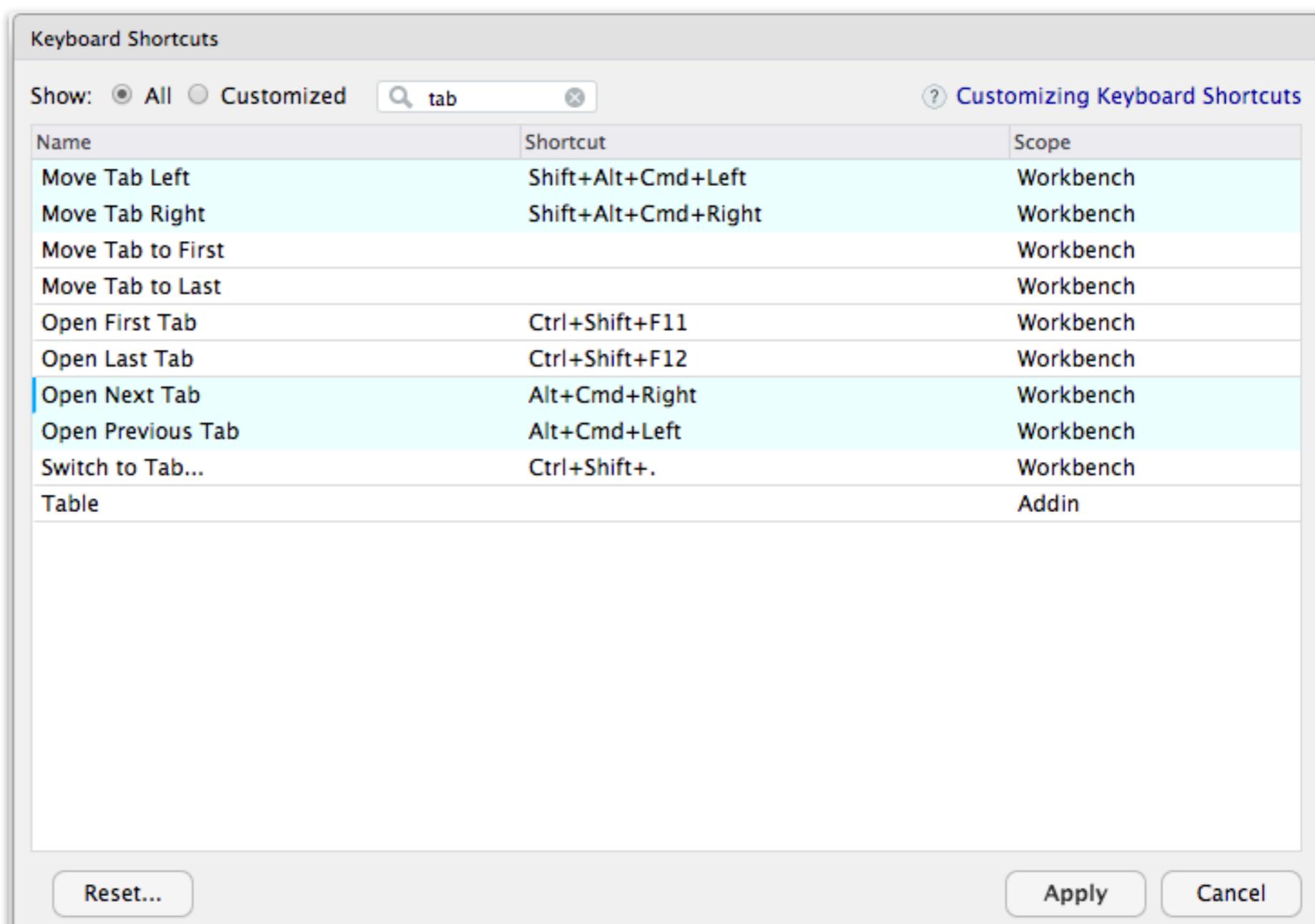
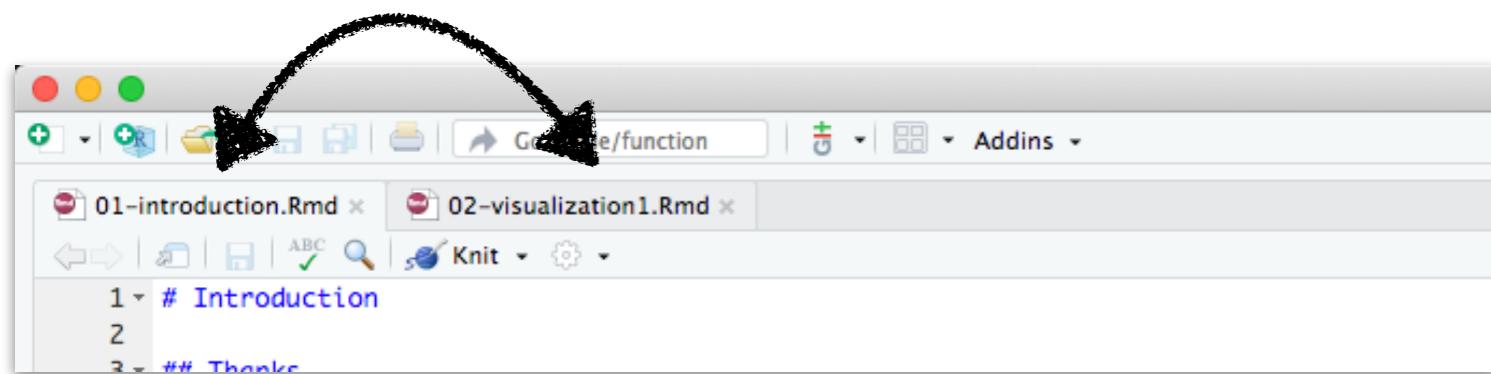


comment
cmd + shift + c

```
1 # ggplot(mapping = aes(x = color, y = price), data = df.diamonds) +  
2 ggplot(mapping = aes(x = cut, y = price), data = df.diamonds) +  
3   stat_summary(fun.y = "mean", geom = "point")
```

change

Navigating between tabs



Jumping between code chunks

```
190 Here is the help file for the `print()` function:  
191  
192 ```{r visualization1-17, echo=FALSE, fig.cap="Help file for the print()  
function.", out.width="95%"}  
193 include_graphics("figures/help_print.png")  
194 ```  
195  
196 ## Data visualization using `ggplot2`  
197  
198 We will use the `ggplot2` package to visualize data. By the end of next class,  
you'll be able to make a figure like this:  
199  
200 ```{r visualization1-18, echo=FALSE, fig.cap="What a nice figure!",  
out.width="95%"}  
201 include_graphics("figures/combined_plot.png")  
202 ```  
...
```



Keyboard Shortcuts		
Show:	All	Customized
<input type="text"/> chunk		Customizing Keyboard Shortcuts
Name	Shortcut	Scope
Restart R Session and Clear Chunk Output		Workbench
Restart R Session and Run All Chunks		Workbench
Go to Next Chunk	Alt+Cmd+Down	Editor
Go to Previous Chunk	Alt+Cmd+Up	Editor
Insert Chunk	Alt+Cmd+I	Editor

Data wrangling 1

Two styles of coding in R

Base R
Cheat Sheet

Getting Help

`?mean`
Get help of a particular function.
`help.search('weighted mean')`
Search the help files for a word or phrase.
`help(package = 'dplyr')`
Find help for a package.

[More about an object](#)

`str(iris)`
Get a summary of an object's structure.
`class(iris)`
Find the class an object belongs to.

Using Packages

`install.packages('dplyr')`
Download and install a package from CRAN.

`library(dplyr)`
Load the package into the session, making all its functions available to use.

`dplyr::select`
Use a particular function from a package.

`data(iris)`
Load a built-in dataset into the environment.

Working Directory

`getwd()`
Find the current working directory (where inputs are found and outputs are sent).

`setwd('C://file/path')`
Change the current working directory.

Use projects in RStudio to set the working directory to the folder you are working in.

Vectors

Creating Vectors

<code>c(2, 4, 6)</code>	<code>2 4 6</code>	Join elements into a vector
<code>2:6</code>	<code>2 3 4 5 6</code>	An integer sequence
<code>seq(2, 3, by=0.5)</code>	<code>2.0 2.5 3.0</code>	A complex sequence
<code>rep(1:2, times=3)</code>	<code>1 2 1 2 1 2</code>	Repeat a vector
<code>rep(1:2, each=3)</code>	<code>1 1 1 2 2 2</code>	Repeat elements of a vector

Vector Functions

<code>sort(x)</code>	<code>rev(x)</code>
Return x sorted.	Return x reversed.
<code>table(x)</code>	<code>unique(x)</code>
See counts of values.	See unique values.

Selecting Vector Elements

By Position

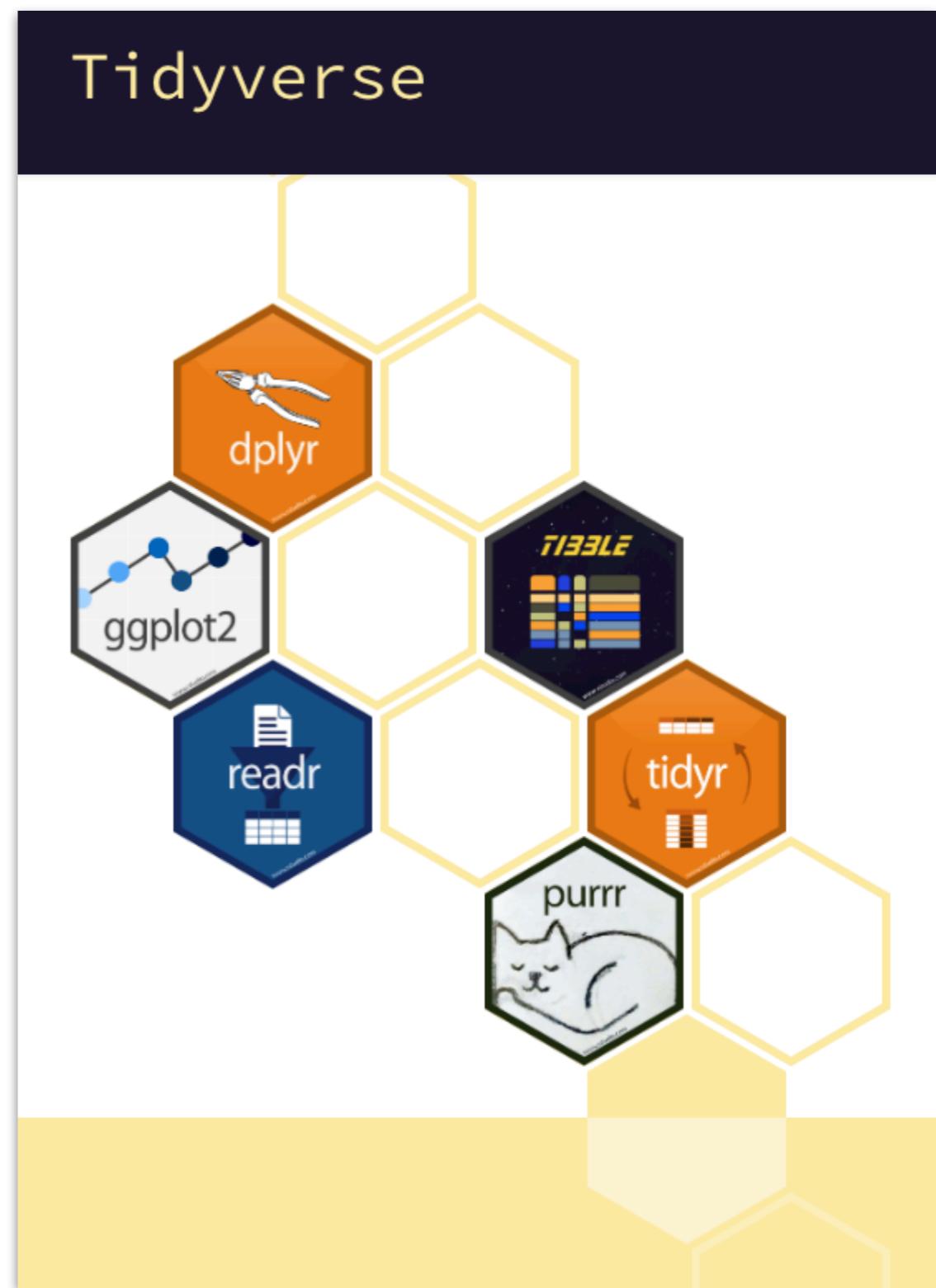
<code>x[4]</code>	The fourth element.
<code>x[-4]</code>	All but the fourth.
<code>x[2:4]</code>	Elements two to four.
<code>x[-(2:4)]</code>	All elements except two to four.
<code>x[c(1, 5)]</code>	Elements one and five.

By Value

<code>x[x == 10]</code>	Elements which are equal to 10.
<code>x[x < 0]</code>	All elements less than zero.
<code>x[x %in% c(1, 2, 5)]</code>	Elements in the set 1, 2, 5.

Named Vectors

<code>x['apple']</code>	Element with name 'apple'.
-------------------------	----------------------------



Software can be chaotic, but we make it work



Expert

Trying Stuff Until it Works

O RLY?

The Practical Developer
@ThePracticalDev

RStudio time

**STAR
WARS**



Reprex

- best way to get help is by posting a **reprex**
- **reprex** = reproducible example

reprex

CRAN 0.2.1 build passing  build passing  78% lifecycle stable



Overview

Prepare reprexes for posting to [GitHub issues](#), [StackOverflow](#), or [Slack snippets](#). What is a `reprex`? It's a **reproducible example**, as coined by [Romain Francois](#).

Given R code on the clipboard, selected in RStudio, as an expression (quoted or not), or in a file ...

- run it via `rmarkdown::render()`,
- with deliberate choices re: arguments and setup chunk.

Get resulting runnable code + output as

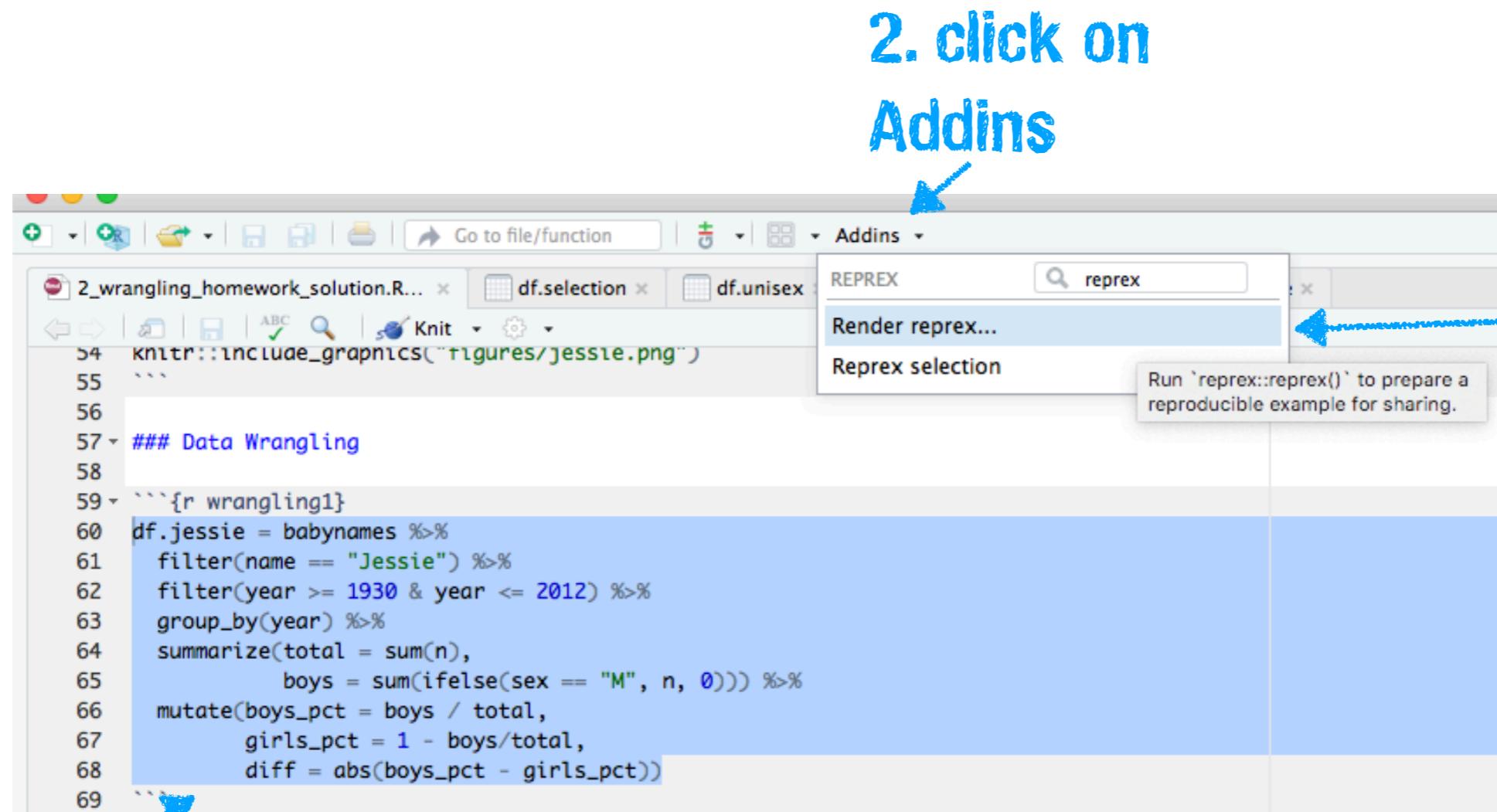
- Markdown, formatted for target venue, e.g. `gh` or `so`, or as
- R code, augmented with commented output.

Result is returned invisibly, placed on the clipboard, and written to a file. Preview an HTML version in RStudio viewer or default browser.

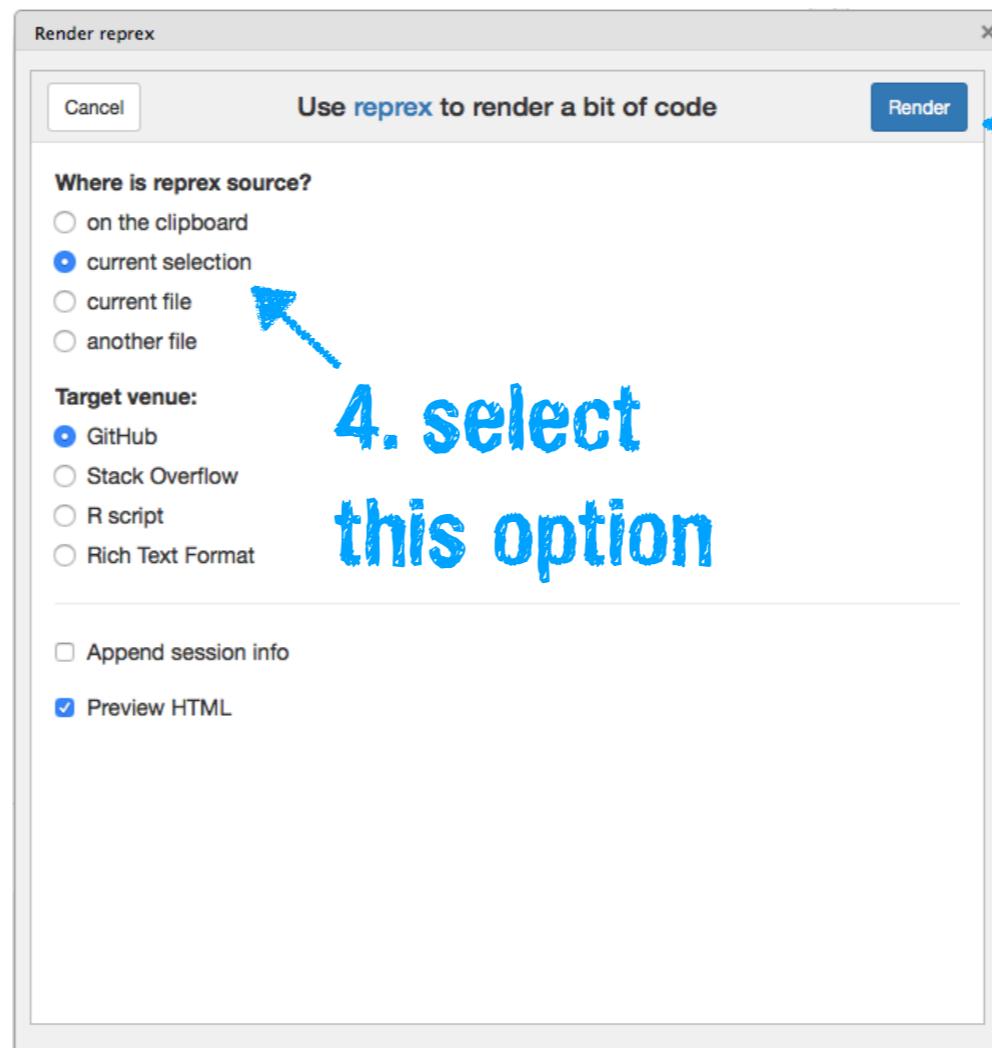


Reprex

```
install.package("reprex")
```



Reprex



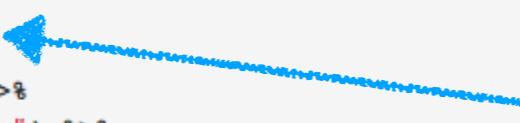
4. select
this option

5. click
render

6. copy and paste from the viewer

```
df.jessie = babynames %>%
  filter(name == "Jessie") %>%
  filter(year >= 1930 & year <= 2012) %>%
  group_by(year) %>%
  summarize(total = sum(n),
            boys = sum(ifelse(sex == "M", n, 0))) %>%
  mutate(boys_pct = boys / total,
        girls_pct = 1 - boys/total,
        diff = abs(boys_pct - girls_pct))
#> Error in babynames %>% filter(name == "Jessie") %>% filter(year >= 1930 & : could not find function "%>%"
```

Reprex



```
library("babynames")
library("tidyverse")
df.jessie = babynames %>%
  filter(name == "Jessie") %>%
  filter(year >= 1930 & year <= 2012) %>%
  group_by(year) %>%
  summarize(total = sum(n),
            boys = sum(ifelse(sex == "M", n, 0))) %>%
  mutate(boys_pct = boys / total,
        girls_pct = 1 - boys/total,
        diff = abs(boys_pct - girls_pct)) %>%
  print()
#> # A tibble: 83 x 6
#>   year  total  boys boys_pct girls_pct   diff
#>   <dbl> <int> <dbl>     <dbl>     <dbl> <dbl>
#> 1 1930    3525  1329     0.377     0.623  0.246
#> 2 1931    3196  1267     0.396     0.604  0.207
#> 3 1932    3178  1282     0.403     0.597  0.193
#> 4 1933    2886  1079     0.374     0.626  0.252
#> 5 1934    2883  1090     0.378     0.622  0.244
#> 6 1935    2721  1103     0.405     0.595  0.189
#> 7 1936    2599  1012     0.389     0.611  0.221
#> 8 1937    2589  1041     0.402     0.598  0.196
#> 9 1938    2446   970     0.397     0.603  0.207
#> 10 1939   2454  1058     0.431     0.569  0.138
#> # ... with 73 more rows
```

Created on 2019-01-24 by the [reprex package](#) (v0.2.1)

7. make sure to load necessary packages in your reprex

Using functions with tidyverse verbs

```
1 df.starwars %>%  
2   select(where(fn = is.numeric))
```

my
recommendation

- fn = is.numeric
- fn = "is.numeric"
- fn = function(x) {is.numeric(x)}
- fn = ~ is.numeric(.)

flexible, short, works well with
other verbs we'll learn about later

- fn = ~ !is.numeric(.)

select all
variables that
are not numeric

Feedback

How was the pace of today's class?

much a little just a little much
too too right too too
slow slow

How happy were you with today's class overall?



What did you like about today's class? What could be improved next time?