

# **weaving reproducibility through the undergraduate statistics curriculum**

mine çetinkaya-rundel

**#1**

**two-pronged approach**

Convince  
researchers to adopt  
a reproducible  
research workflow

**#2**

Train new  
researchers who  
don't have any  
other workflow

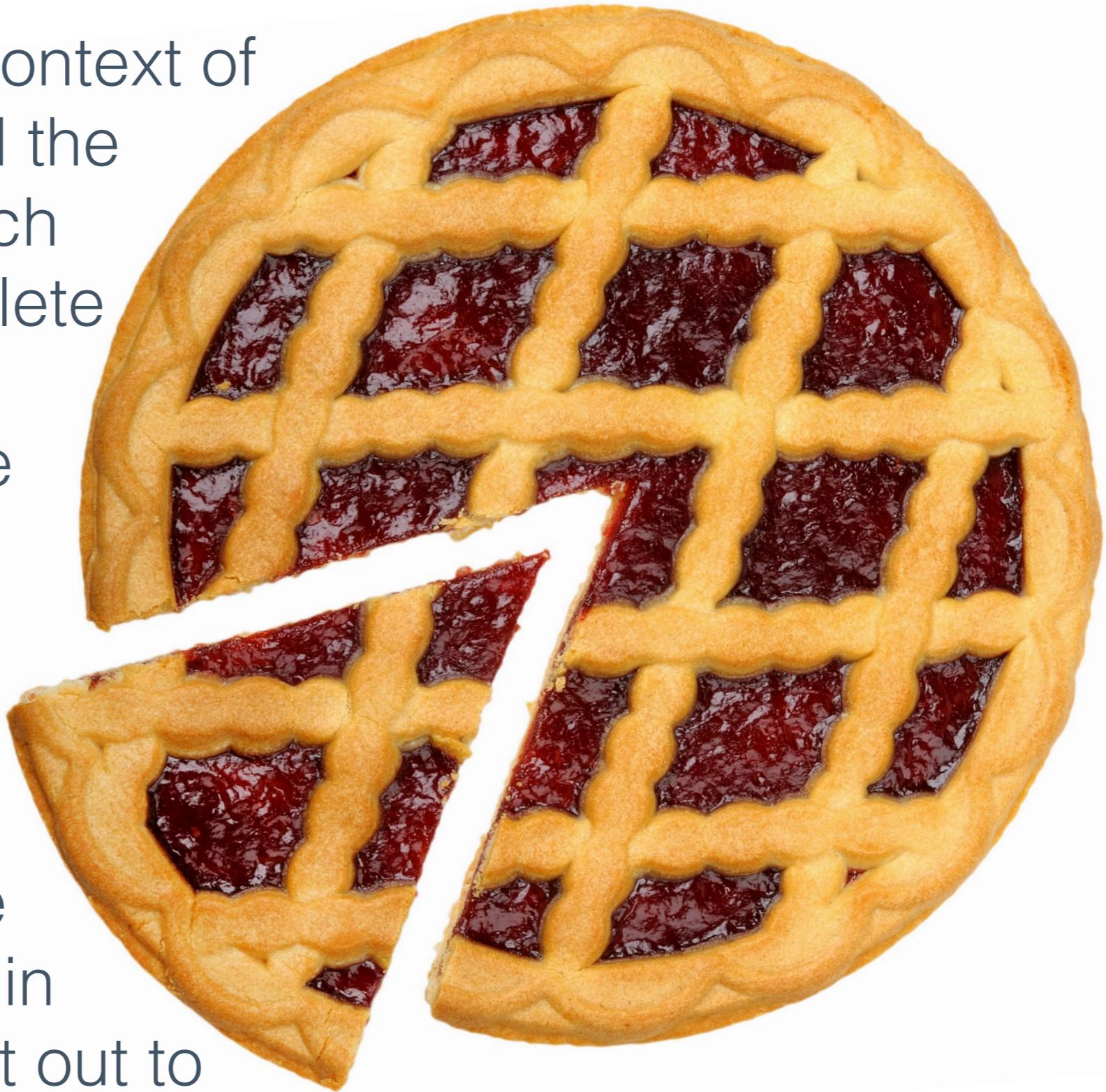


# **reproducibility**

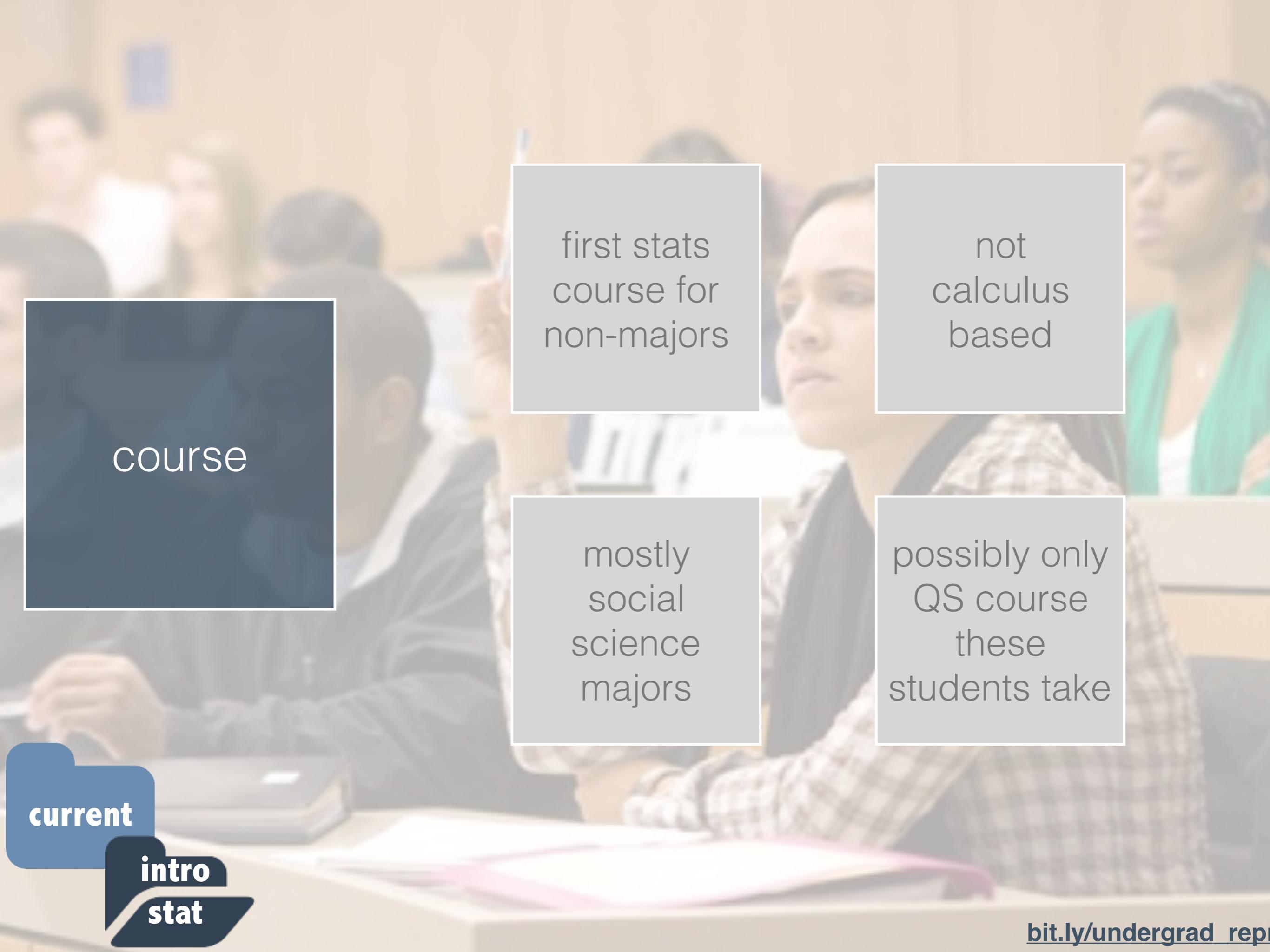
often comes up in the context of published research and the need to accompany such research with the complete data and analyses, including software/code

# **statistics**

educators who teach data analysis should be instilling best practices in students before they set out to do research





A blurred background image of a classroom with several students of diverse ethnicities sitting at their desks, facing forward. The room appears to be a lecture hall or classroom setting.

course

first stats  
course for  
non-majors

not  
calculus  
based

mostly  
social  
science  
majors

possibly only  
QS course  
these  
students take

current

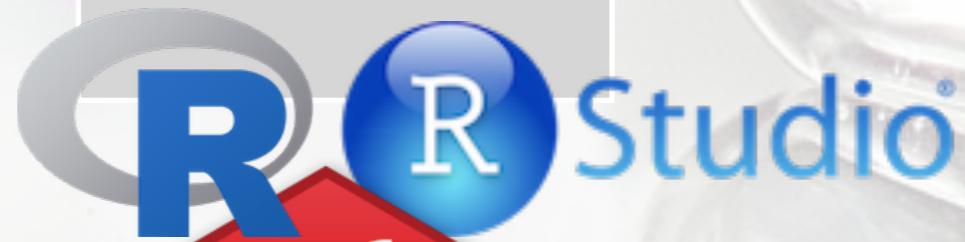
intro  
stat

[bit.ly/undergrad\\_reps](http://bit.ly/undergrad_reps)

A background image showing a collection of clear plastic petri dishes containing white bacterial cultures, arranged in a grid pattern.

reproducibility

literate  
programming

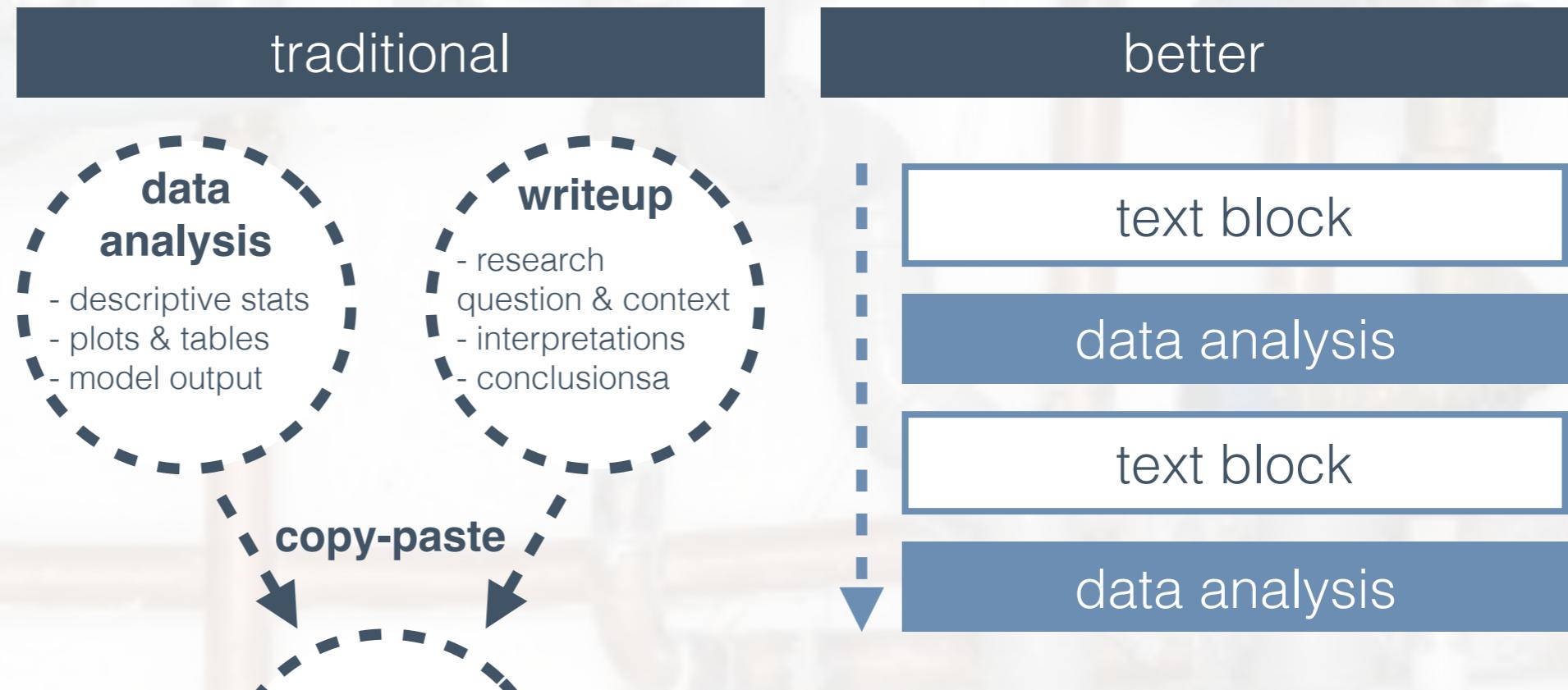
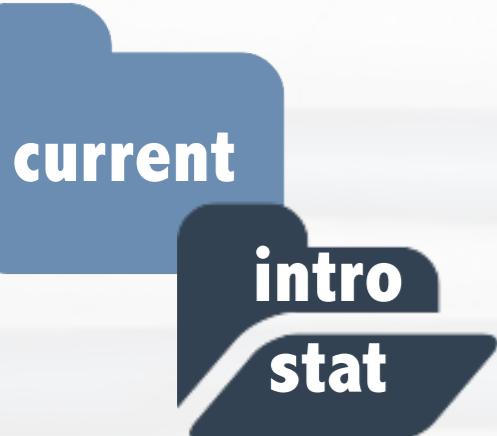


current

intro  
stat

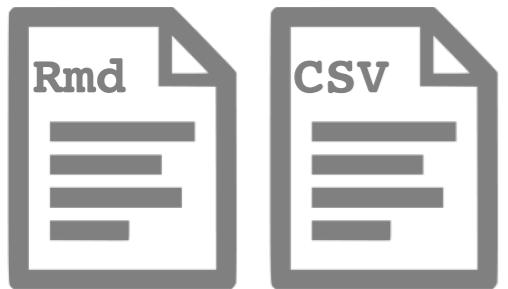
[bit.ly/undergrad\\_repro](http://bit.ly/undergrad_repro)

# workflow



→ CMS → ✓

[bit.ly/undergrad\\_repor](http://bit.ly/undergrad_repor)



→ CMS → ✓



#### ▶ Assignment Instructions

### Assignment Submission

There is no student submitted text.

### Submitted Attachments

[lab10.html](#) ( 2 MB; Apr 14, 2016 6:05 am )

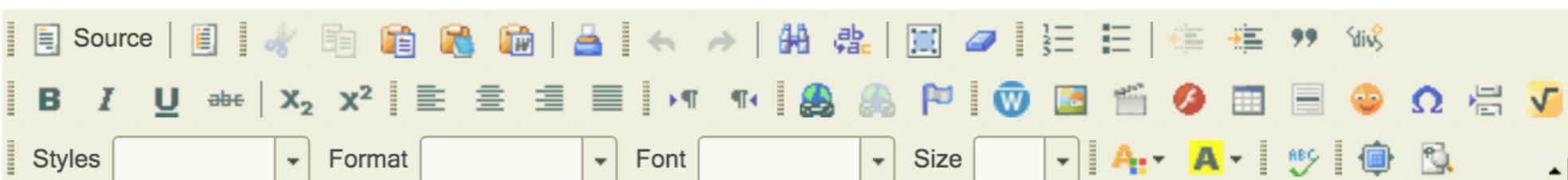
[lab10.Rmd](#) ( 12 KB; Apr 14, 2016 6:05 am )

Grade:  (max 100.0)

Assign Grade Overrides

### Instructor Summary Comments

Use the box below to enter additional summary comments about this submission.



Ex17) Collinearity refers to relationships between predictors. -2 Here this is just dependence between cases.

Ex19) Probably generalizable to other large state schools -1

current

intro  
stat

[bit.ly/undergrad\\_repor](http://bit.ly/undergrad_repor)



can  
students  
handle it?

yes!

point-and-  
click is no  
less  
overhead  
than scripting

current

intro  
stat

[bit.ly/undergrad\\_repl](http://bit.ly/undergrad_repl)

### **III. Adding Proportions to Summary Table**

For categorical variables, you should see the counts of each possible outcome of that variable in the **Summary Table**. To see the breakdown of proportions or percentages, follow these steps:

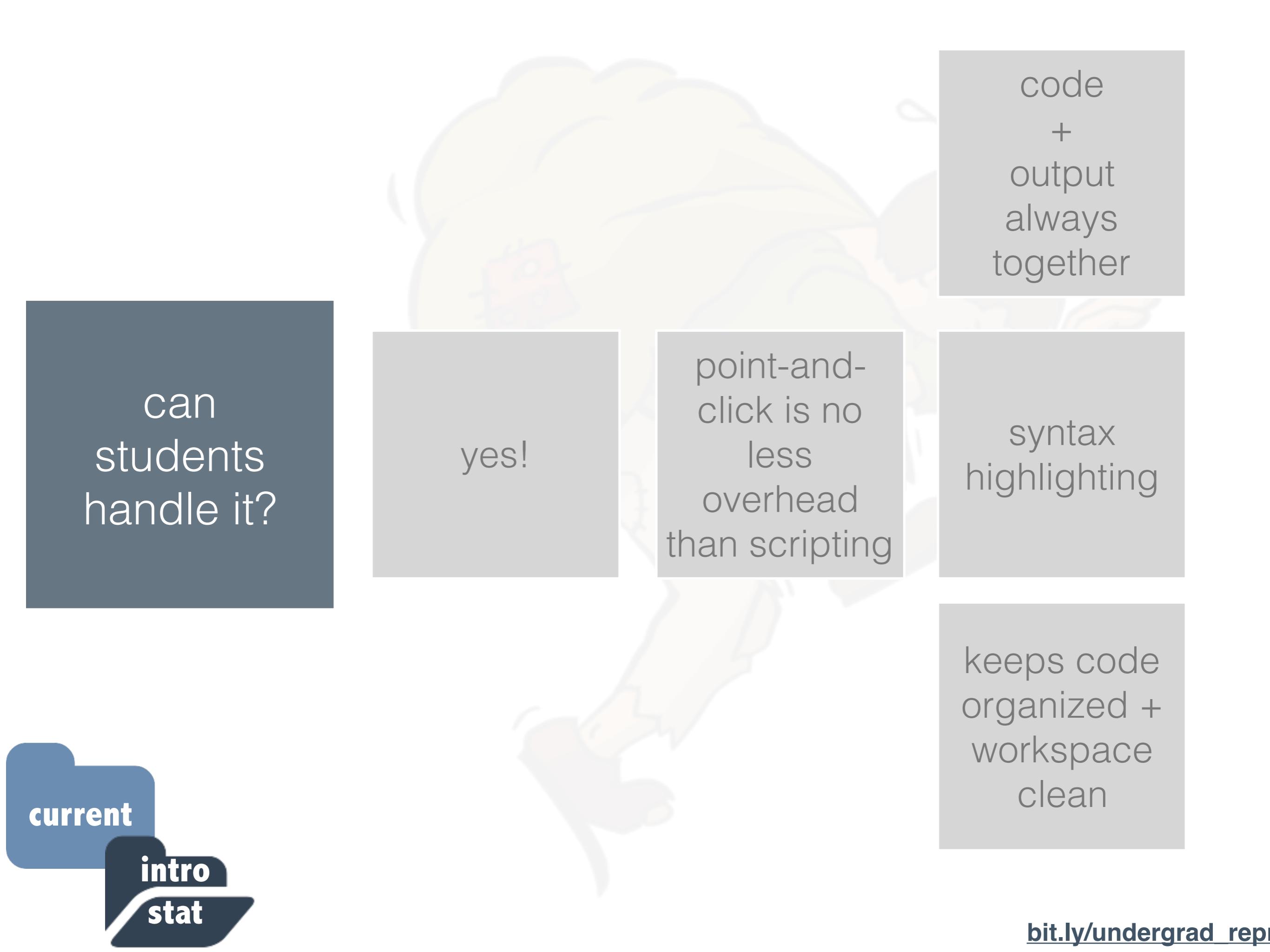
The screenshot shows the Fathom software interface with a title bar "Fathom - [Sleep Data]". The menu bar includes File, Edit, Object, Collection, Summary, Window, Help, and a toolbar with icons for Collection, Table, Graph, Summary, Estimate, Test, Model, and Slide. A central window displays a "Collection 1" table with four columns: Gender, Grade, Sleep, and a formula column "<new>". The "Sleep" column contains numerical values ranging from 5.5 to 8.0. The bottom status bar reads "Fathom Dynamic Data Software".

	Gender	Grade	Sleep	<new>
1	F	R	5.5	
2	M	R	6.0	
3	M	R	6.0	
4	F	R	7.0	
5	F	R	6.0	
6	M	R	4.0	
7	M	R	8.0	
8	F	R	7.0	
9	F	R	5.0	
10	F	R	6.0	
11	M	R	7.5	
12	M	R	7.0	
13	M	R	6.5	
14	M	R	6.5	
15	M	R	6.0	
16	M	R	7.0	
17	F	S	6.5	
18	F	S	8.0	
19	F	S	9.0	
20	M	S	7.0	
21	F	S	7.0	
22	M	F	7.0	
23	F	S	7.0	
24	F	F	8.0	

- Click on the **Summary Table** to highlight it, click on the “**Summary**” drop-down menu and select “**Add Formula**”. In general, whenever you click and select a *Fathom* object (such as a **Table**, **Graph**, or **Summary**) the menu at the top of the screen will change to give you options for working on that object.
- In the formula editor that pops up, type “*rowproportion*” (without the quotes) to see the row proportions or “*columnproportion*” to see the column proportions. Be sure to spell the names of the formulas correctly or else *Fathom* will give you an error. (If you spell the names correctly, they should change to a purplish color in your editor.)
- You will see that each cell in the **Summary Table** now includes numbers for multiple statistics. To see which numbers correspond with which statistics, simply look at the bottom of your summary table to see the order of the statistics or formulas within each cell.
- To delete (or change) a particular statistic from the table, you can double click on its name at the bottom of the **Summary Table**. In the formula editor, press delete (or make your changes) and then click “**OK**”.

current

intro  
stat



can  
students  
handle it?

yes!

point-and-  
click is no  
less  
overhead  
than scripting

syntax  
highlighting

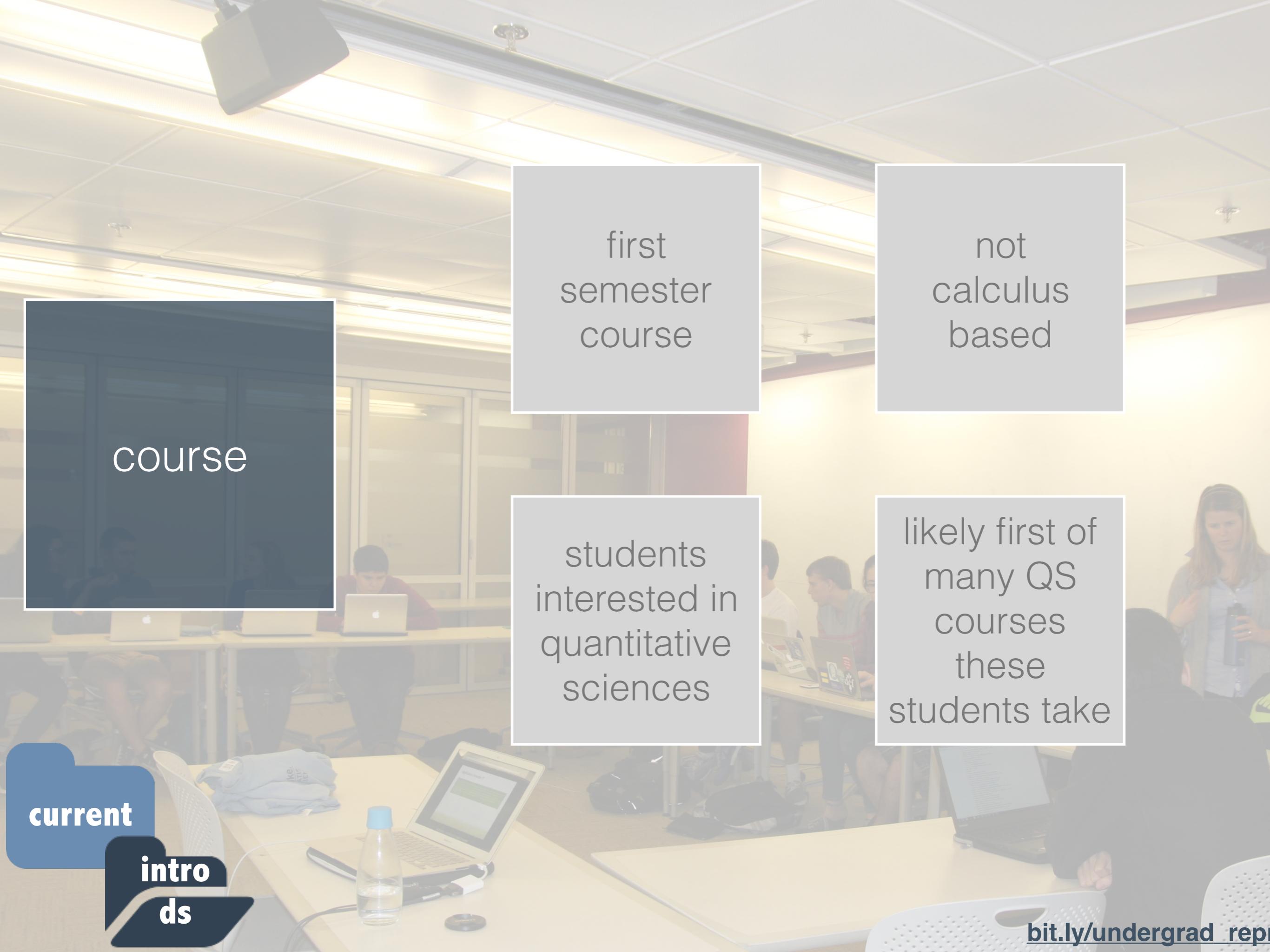
keeps code  
organized +  
workspace  
clean

current

intro  
stat

code  
+  
output  
always  
together

[bit.ly/undergrad\\_repl](http://bit.ly/undergrad_repl)

A photograph of a classroom environment. Students are seated at wooden desks, working on laptops. The room has a modern feel with recessed ceiling lights and large windows in the background.

course

first  
semester  
course

not  
calculus  
based

students  
interested in  
quantitative  
sciences

likely first of  
many QS  
courses  
these  
students take

current

intro  
ds

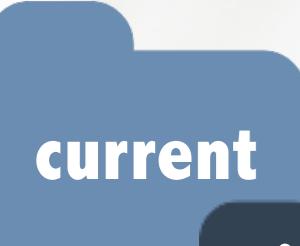
[bit.ly/undergrad\\_repl](http://bit.ly/undergrad_repl)

The background of the slide features a close-up photograph of several clear plastic petri dishes arranged in a grid. Each dish contains a different type of bacterial culture, showing various growth patterns and colors like white, yellow, and green.

reproducibility

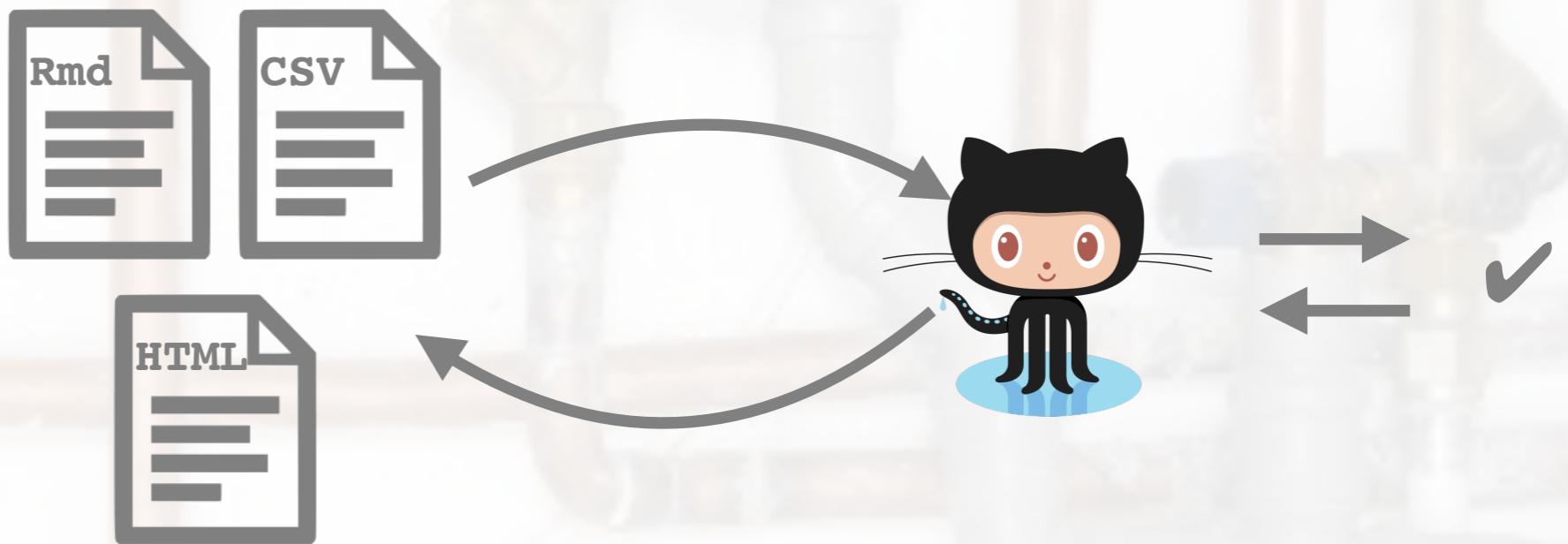
literate  
programming

version  
control

The R logo, consisting of two interlocking circles, is positioned next to the word "Studio" in a blue, sans-serif font. The "S" in "Studio" has a registered trademark symbol (®) at its top right corner.A blue icon shaped like a folder or a cloud, containing the word "current" in white.A dark blue icon shaped like a folder or a cloud, containing the word "intro" in white.A dark blue icon shaped like a folder or a cloud, containing the letters "ds" in white.

[bit.ly/undergrad\\_repro](http://bit.ly/undergrad_repro)

workflow

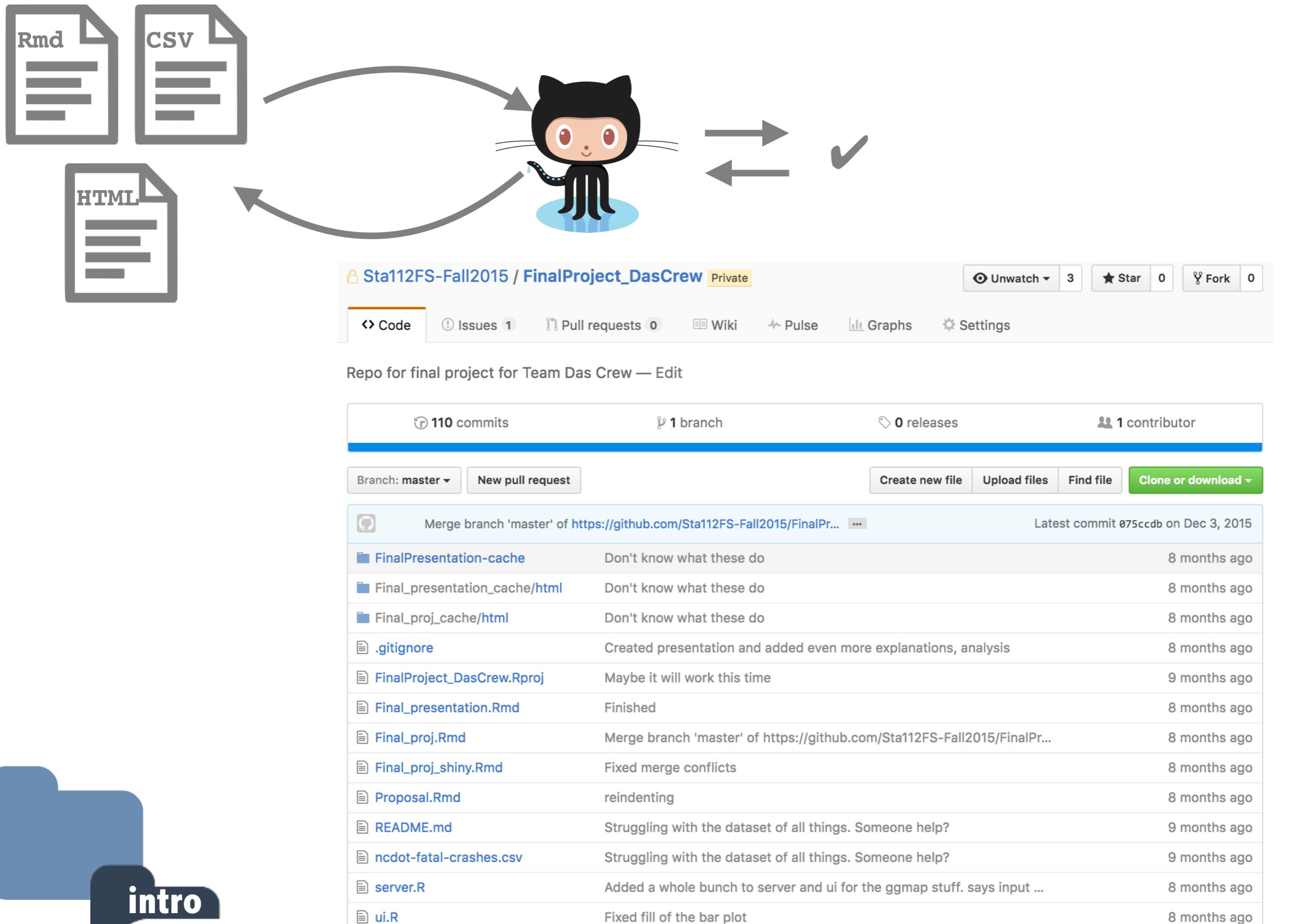


current

intro

ds

[bit.ly/undergrad\\_repl](https://bit.ly/undergrad_repl)



[bit.ly/undergrad\\_repor](http://bit.ly/undergrad_repor)



can  
students  
handle it?

yes!

but...

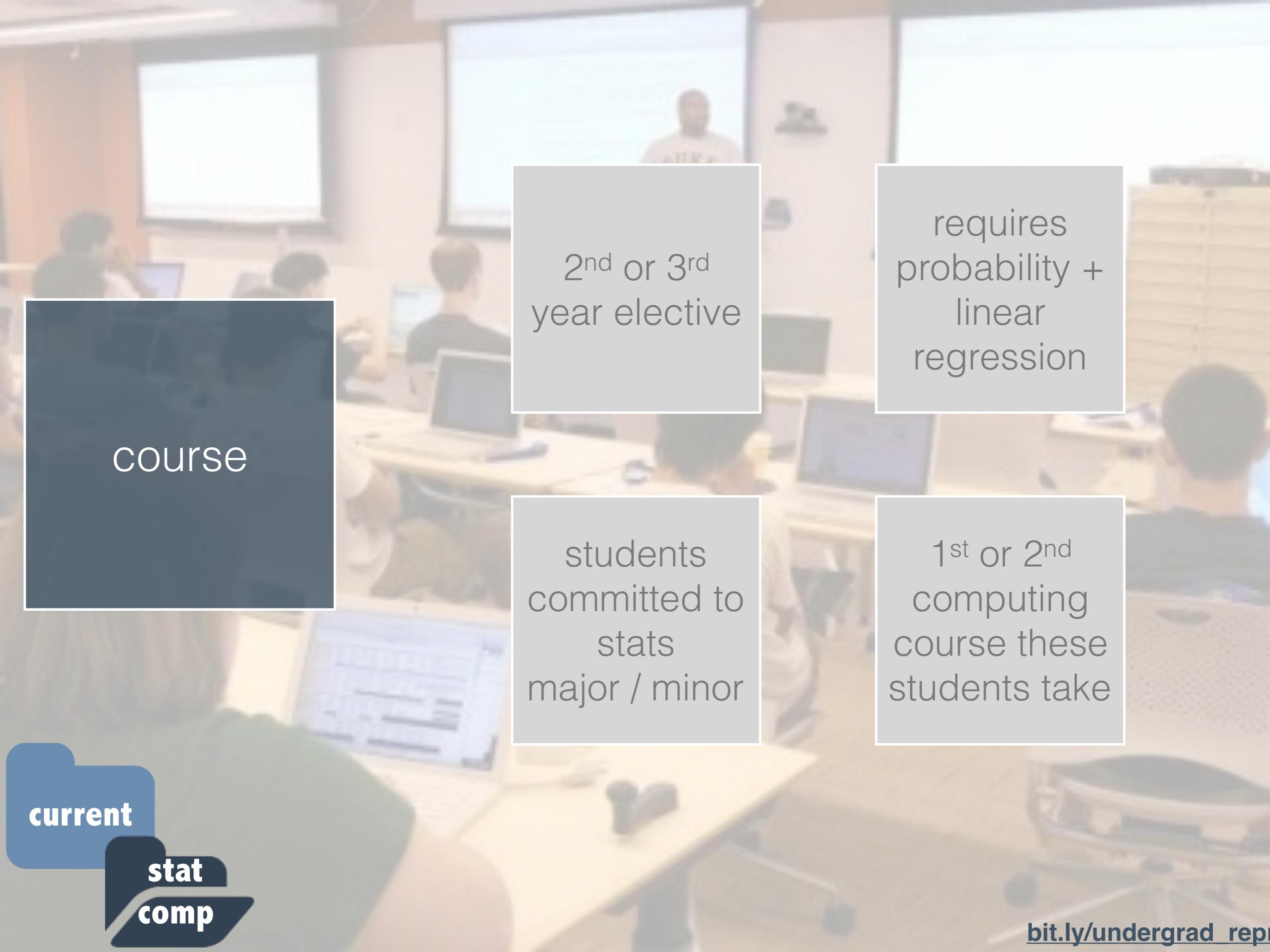
instruction of  
workflow  
requires time  
and care

current

intro

ds

[bit.ly/undergrad\\_repro](http://bit.ly/undergrad_repro)

The background image shows a classroom environment. In the foreground, several students are seated at wooden desks, facing a teacher who is standing at the front of the room. The teacher is wearing a light-colored shirt. The room has a warm, wooden interior.

course

current

stat

comp

2<sup>nd</sup> or 3<sup>rd</sup>  
year elective

students  
committed to  
stats  
major / minor

requires  
probability +  
linear  
regression

1<sup>st</sup> or 2<sup>nd</sup>  
computing  
course these  
students take

reproducibility

literate  
programming

version  
control

build  
tools

**make**

current

stat

comp

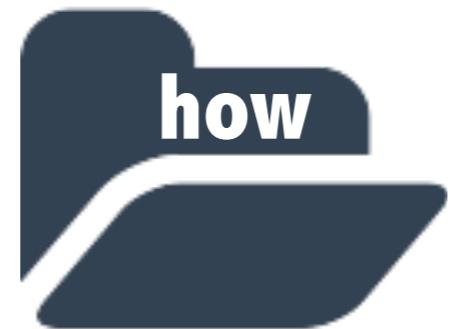
R R Studio<sup>®</sup>



[bit.ly/undergrad\\_repro](https://bit.ly/undergrad_repro)

details at [bit.ly/Sta323\\_2016](https://bit.ly/Sta323_2016)

grow toolkit along with the complexity of computation



what

capstone  
course

senior  
thesis /  
independent  
study

how

need  
instructor  
buy-in

needs  
to be  
part of  
assessment

easily  
adoptable  
framework  
will help

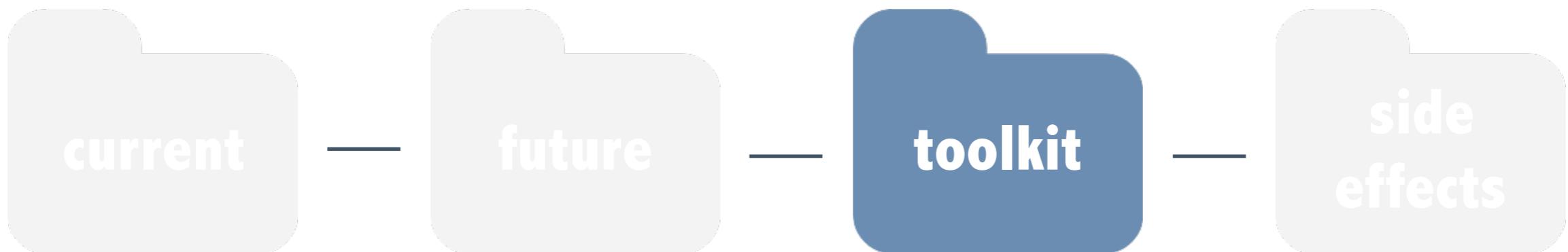
future

Karl's  
steps  
2RR

Project  
TIER

Reproducible  
Science  
Curriculum

[bit.ly/undergrad\\_repre](http://bit.ly/undergrad_repre)



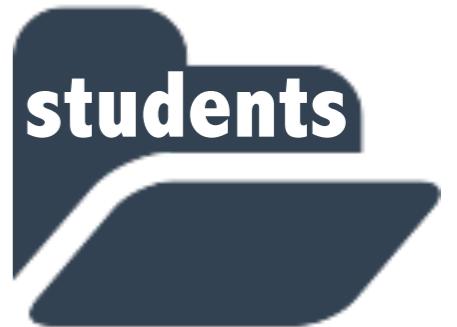
R

other

built-in  
seamless  
ecosystem  
with RStudio

any  
scripting  
language

more  
overhead in  
some than  
others



for instructors

easy  
Q&A

easy  
grading

for students

easy  
collaboration

self-  
promotion

side  
effects

[bit.ly/undergrad\\_reps](http://bit.ly/undergrad_reps)

# thank you!

[bit.ly/undergrad\\_repro](https://bit.ly/undergrad_repro)

**README with links to resources + course pages**



minebocek



[mine@stat.duke.edu](mailto:mine@stat.duke.edu)



mine-cetinkaya-rundel