

integrating reproducibility into the undergraduate statistics curriculum

mine çetinkaya-rundel

weaving reproducibility through the undergraduate statistics curriculum

mine çetinkaya-rundel

#1

two-pronged approach

Convince
researchers to adopt
a reproducible
research workflow

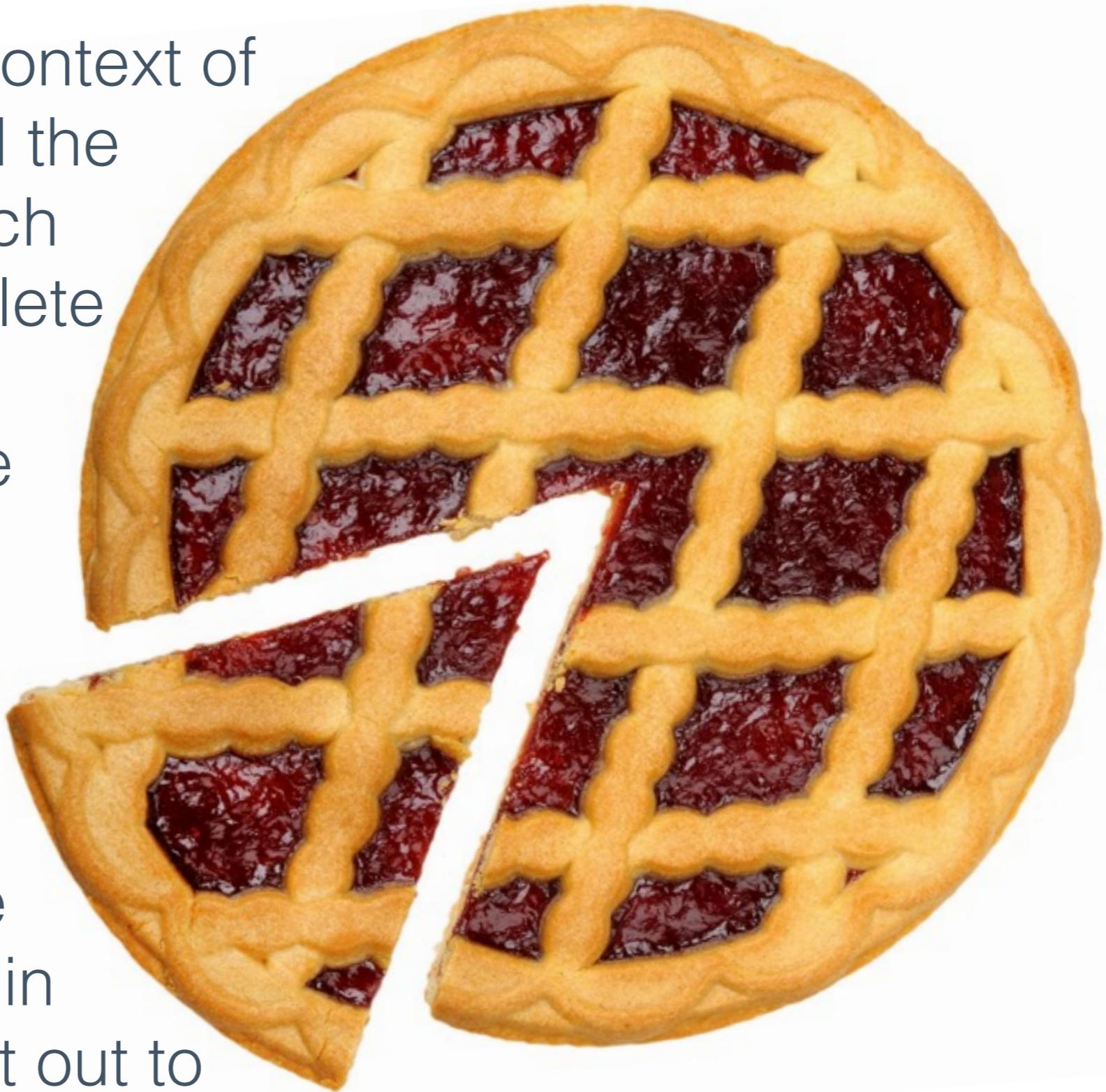
#2

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

current

future

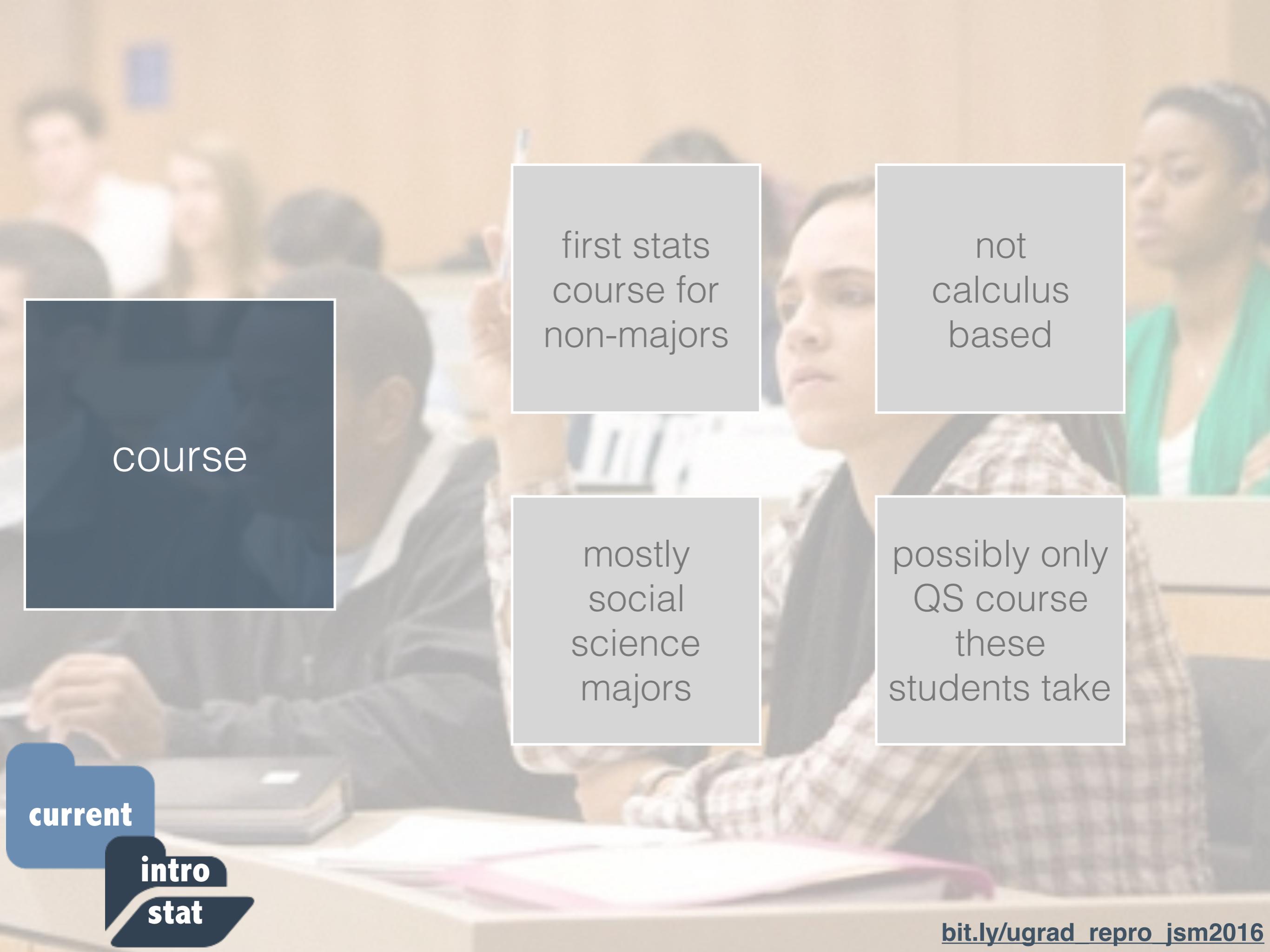
toolkit

**side
effects**

**intro
stat**

**intro
ds**

**stat
comp**



course

first stats
course for
non-majors

not
calculus
based

mostly
social
science
majors

possibly only
QS course
these
students take

current

intro
stat

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

reproducibility

literate
programming

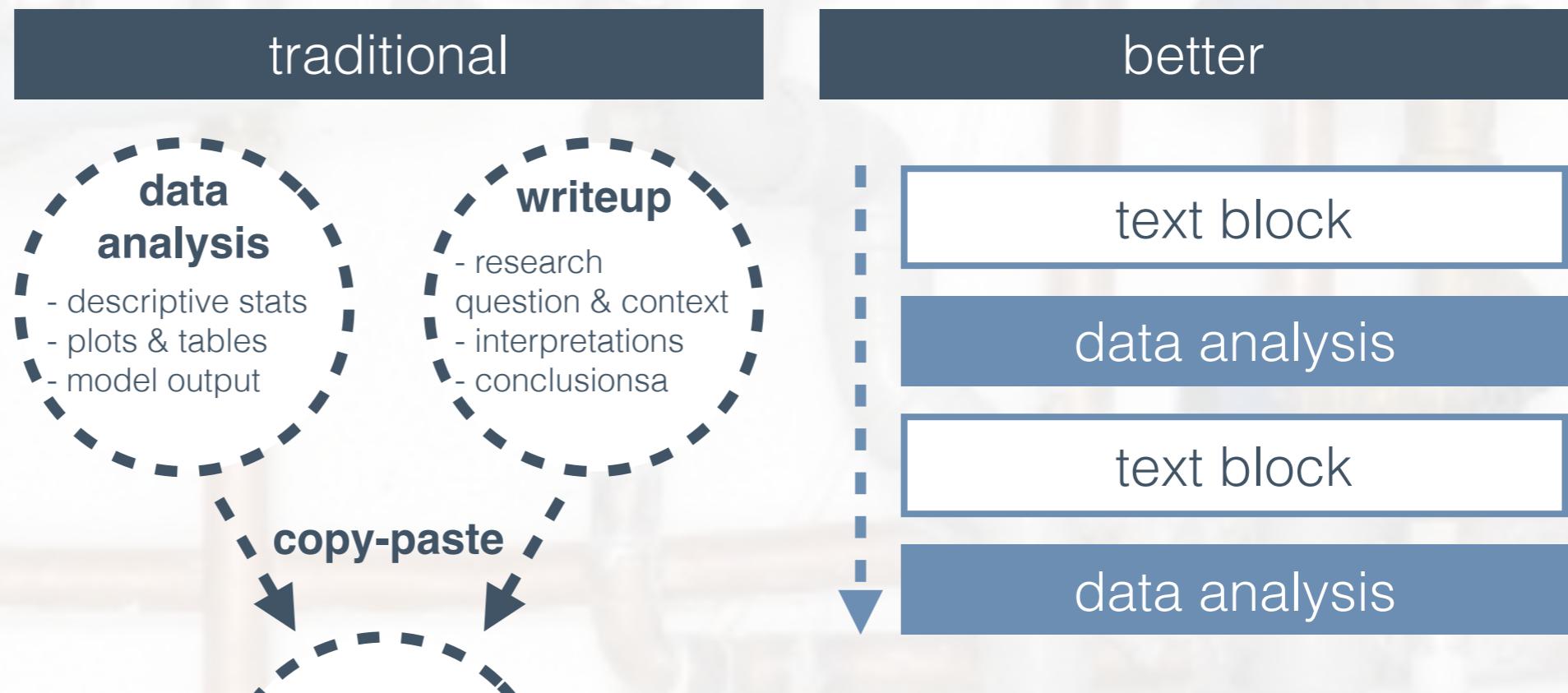
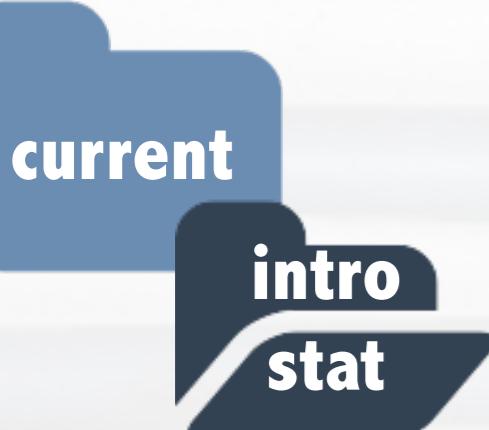


current

intro
stat

[bit.ly/ugrad repro jsm2016](http://bit.ly/ugrad_repro_jsm2016)

workflow



→ CMS → ✓

can
students
handle it?

yes!
actually
makes
learning R
easier

code
+
output
always
together

syntax
highlighting

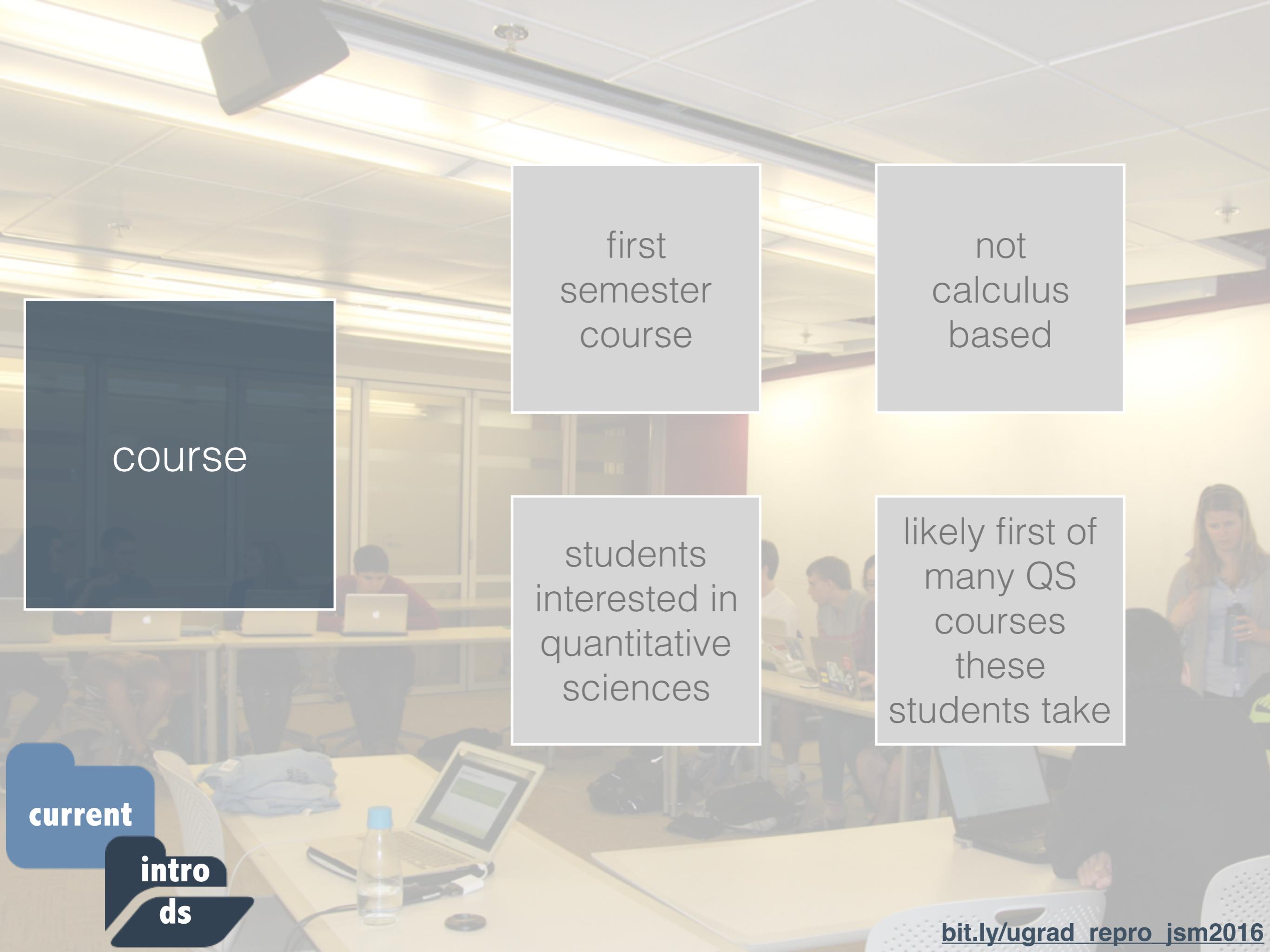
keeps
code
organized

keeps
workspace
clean

current

intro
stat

bit.ly/ugrad_repro_jsm2016

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

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/ugrad_repro_jsm2016

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

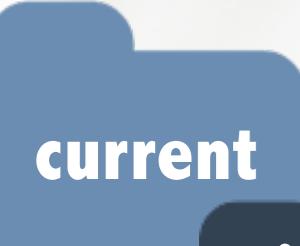
reproducibility

literate
programming

version
control

The R logo, consisting of two interlocking 'R's, is positioned next to the word "Studio" in a blue sans-serif font.

R Studio

A blue speech bubble icon with the word "current" inside.

current

A dark blue speech bubble icon with the word "intro" inside.

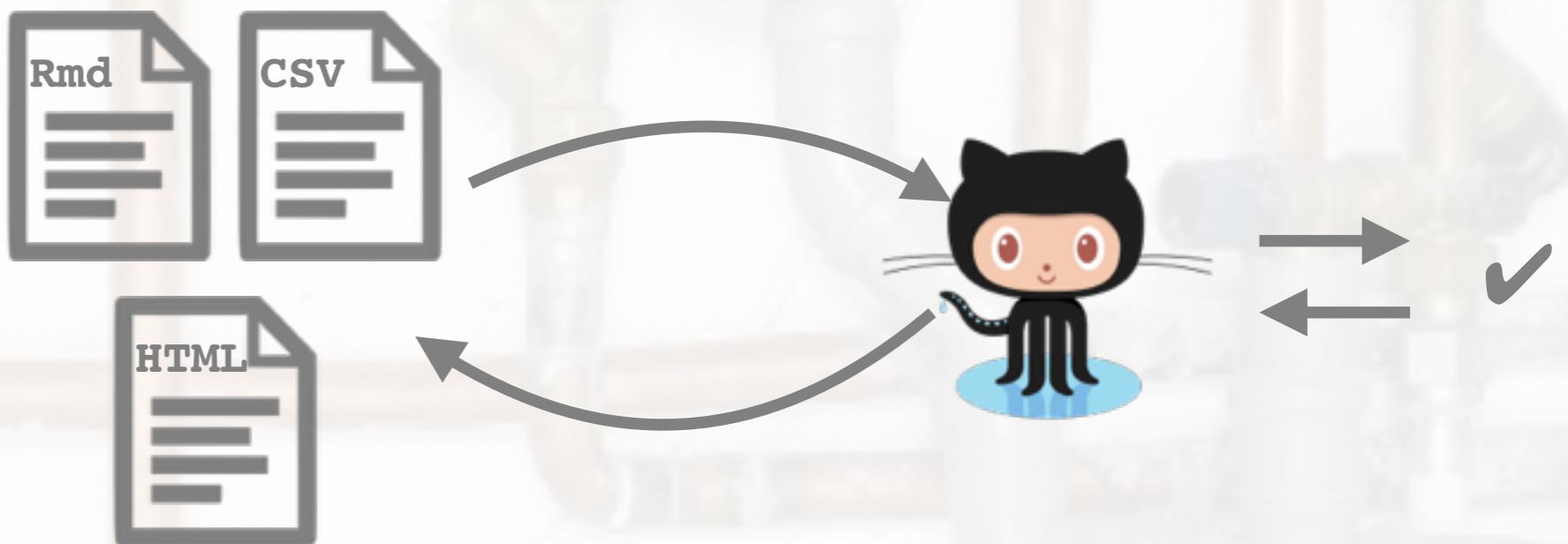
intro

A dark blue speech bubble icon with the letters "ds" inside.

ds

[bit.ly/ugrad repro jsm2016](http://bit.ly/ugrad_repro_jsm2016)

workflow



current

intro

ds

bit.ly/ugrad_repro_jsm2016



can
students
handle it?

yes!

but...

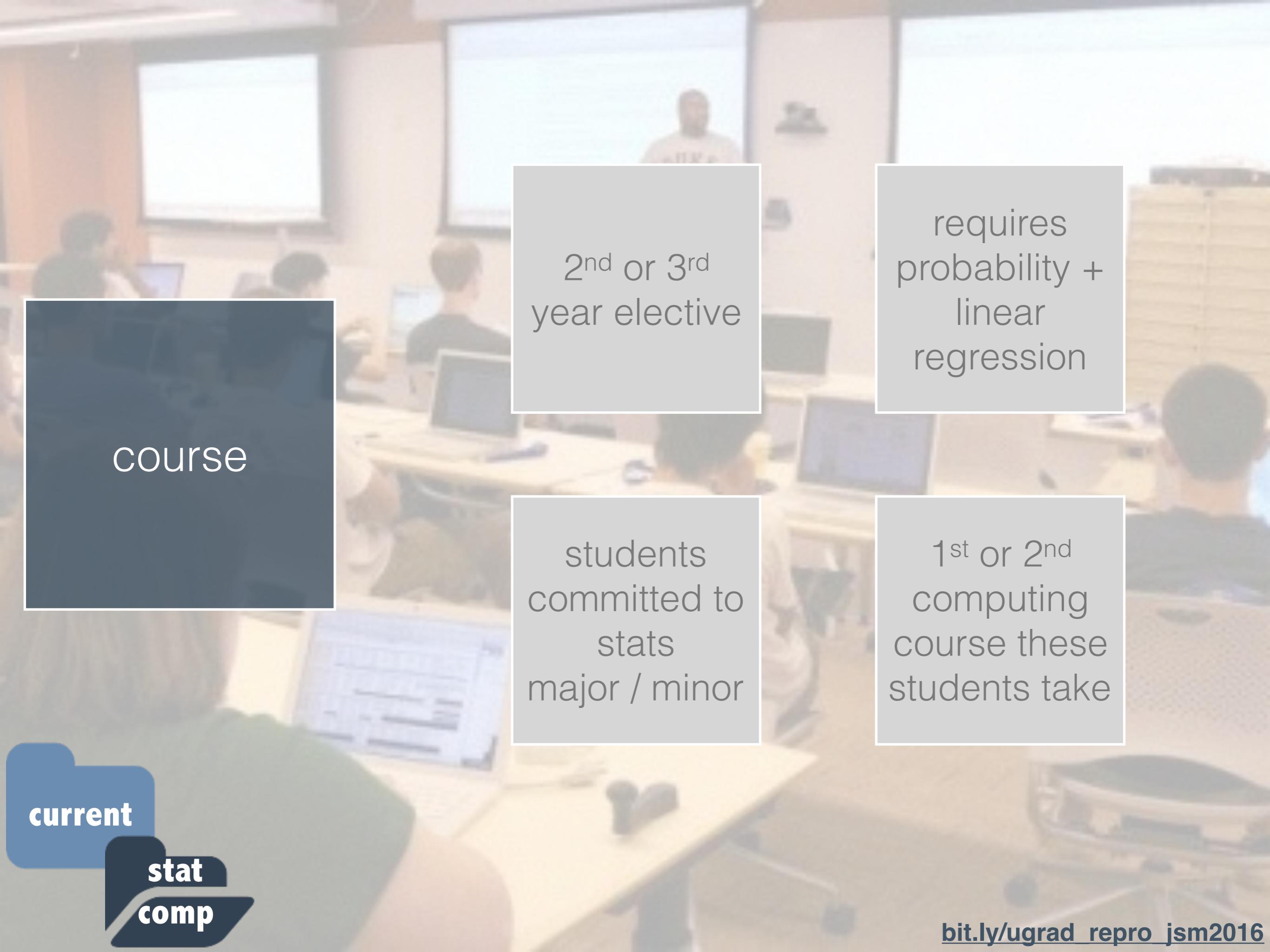
instruction of
workflow
requires time
and care

current

intro

ds

bit.ly/ugrad_repro_jsm2016



course

2nd or 3rd
year elective

requires
probability +
linear
regression

students
committed to
stats
major / minor

1st or 2nd
computing
course these
students take

current

stat

comp

The background of the slide features a collection of clear plastic petri dishes arranged in a grid pattern. 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

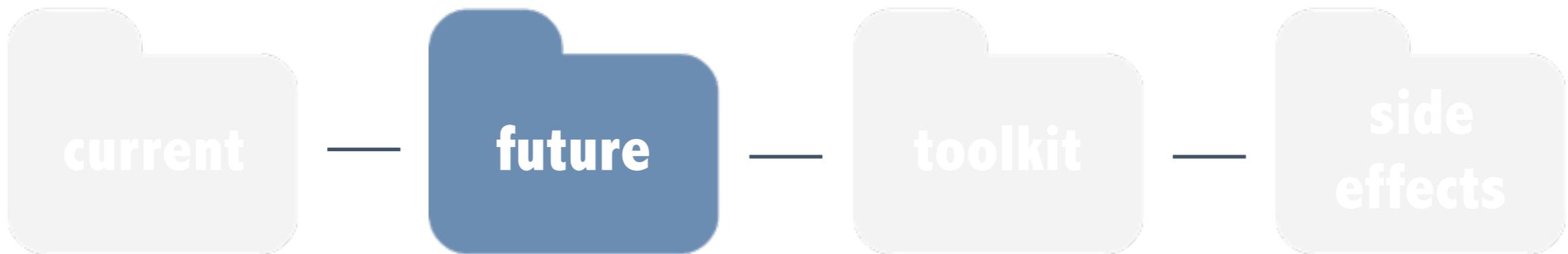
build
tools

make

The R logo, consisting of a grey circle with a white 'R' inside, is positioned next to the word 'Studio' in a blue sans-serif font. The 'R' is partially overlapping the circle.A blue speech bubble icon with a white outline and a slightly rounded bottom. The word 'current' is written in white, bold, sans-serif font inside it.A blue speech bubble icon with a white outline and a slightly rounded bottom. The word 'stat' is written in white, bold, sans-serif font inside it.A dark blue speech bubble icon with a white outline and a slightly rounded bottom. The word 'comp' is written in white, bold, sans-serif font inside it.

details at http://bit.ly/statcomp_jsm2016

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

bit.ly/ugrad_repro_jsm2016



R

other

built-in
seamless
ecosystem
with RStudio

any
scripting
language
(more overhead in
some than others)

e.g.
Stata /
Project
TIER



for instructors

easy
Q&A

easy
grading

for students

easy
collaboration

self-
promotion

side
effects

bit.ly/ugrad_repro_jsm2016

thank you!

questions?



minebocek



mine@stat.duke.edu



mine-cetinkaya-rundel

bit.ly/ugrad_repro_jsm2016