

integrating reproducibility into the undergraduate statistics curriculum

mine çetinkaya-rundel

-  [minebocek](#)
-  mine@stat.duke.edu
-  [mine-cetinkaya-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

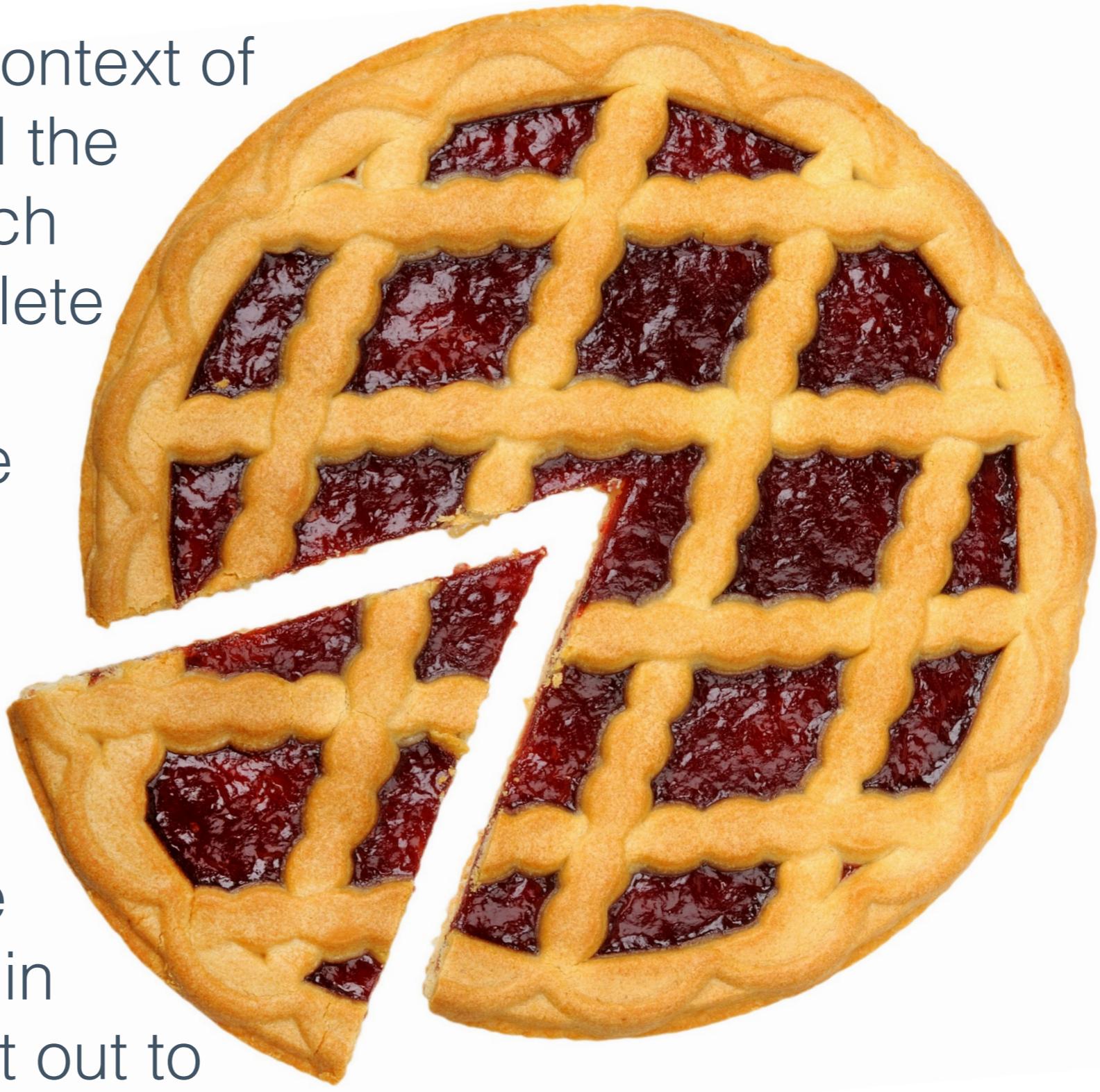


reproducibility

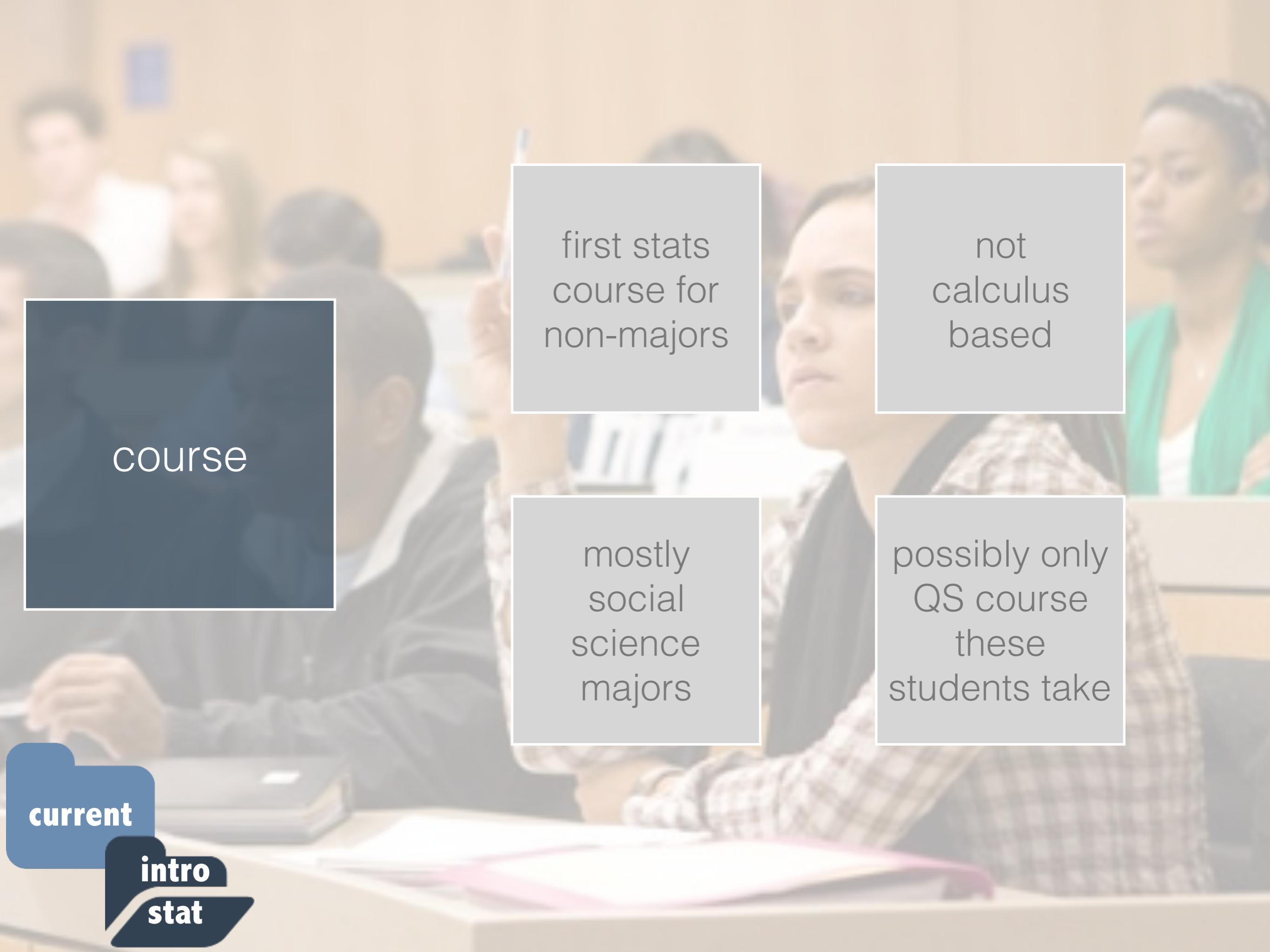
often comes up in the context of published research and the need to accompany such research with the complete data and analyses, including software/code

statistics

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





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

current

intro
stat

first stats
course for
non-majors

not
calculus
based

mostly
social
science
majors

possibly only
QS course
these
students take

The background of the slide features a collection of clear plastic petri dishes arranged in a grid pattern. Each dish contains a white agar medium with small, circular bacterial colonies of varying sizes, suggesting a controlled experiment or a demonstration of growth conditions.

reproducibility

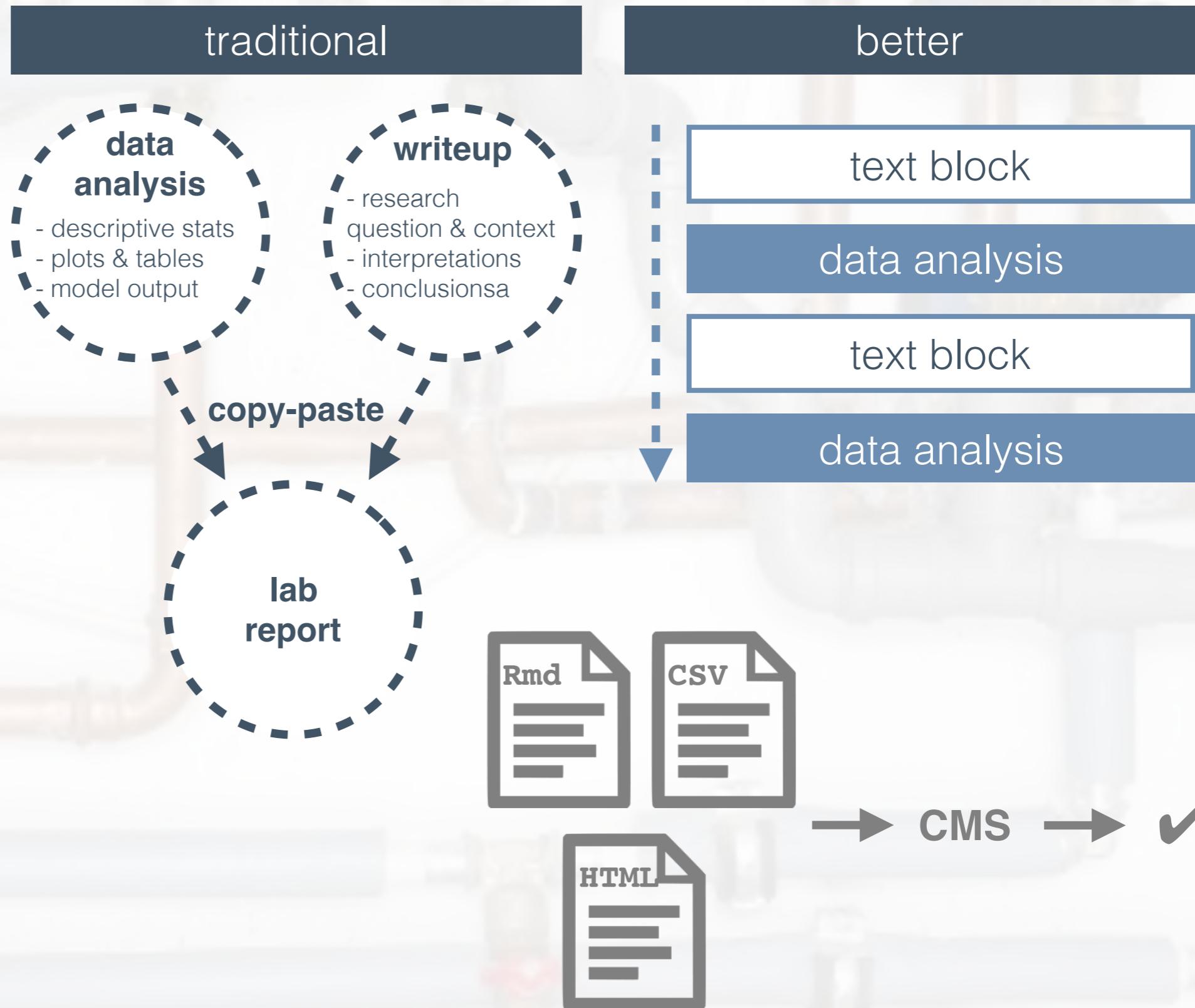
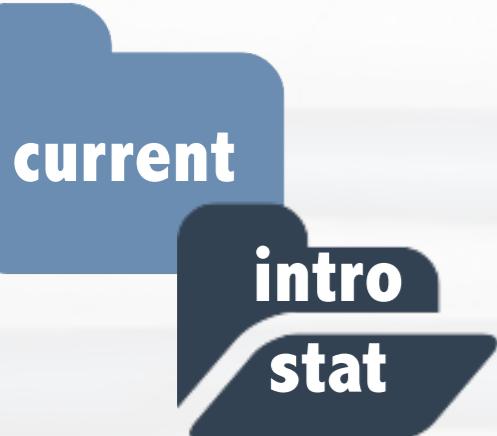
literate
programming

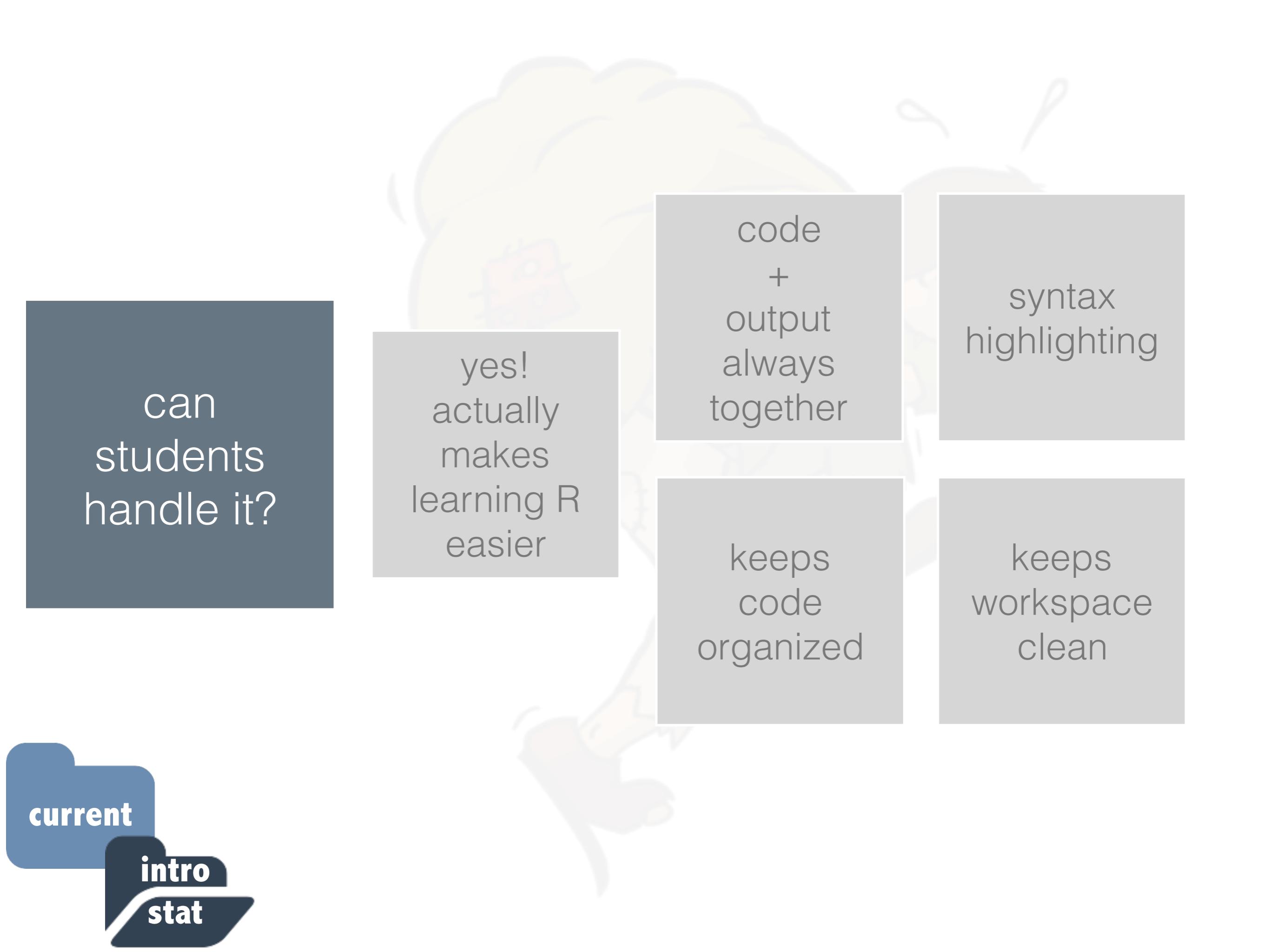


current

intro
stat

workflow





can
students
handle it?

yes!
actually
makes
learning R
easier

code
+
output
always
together

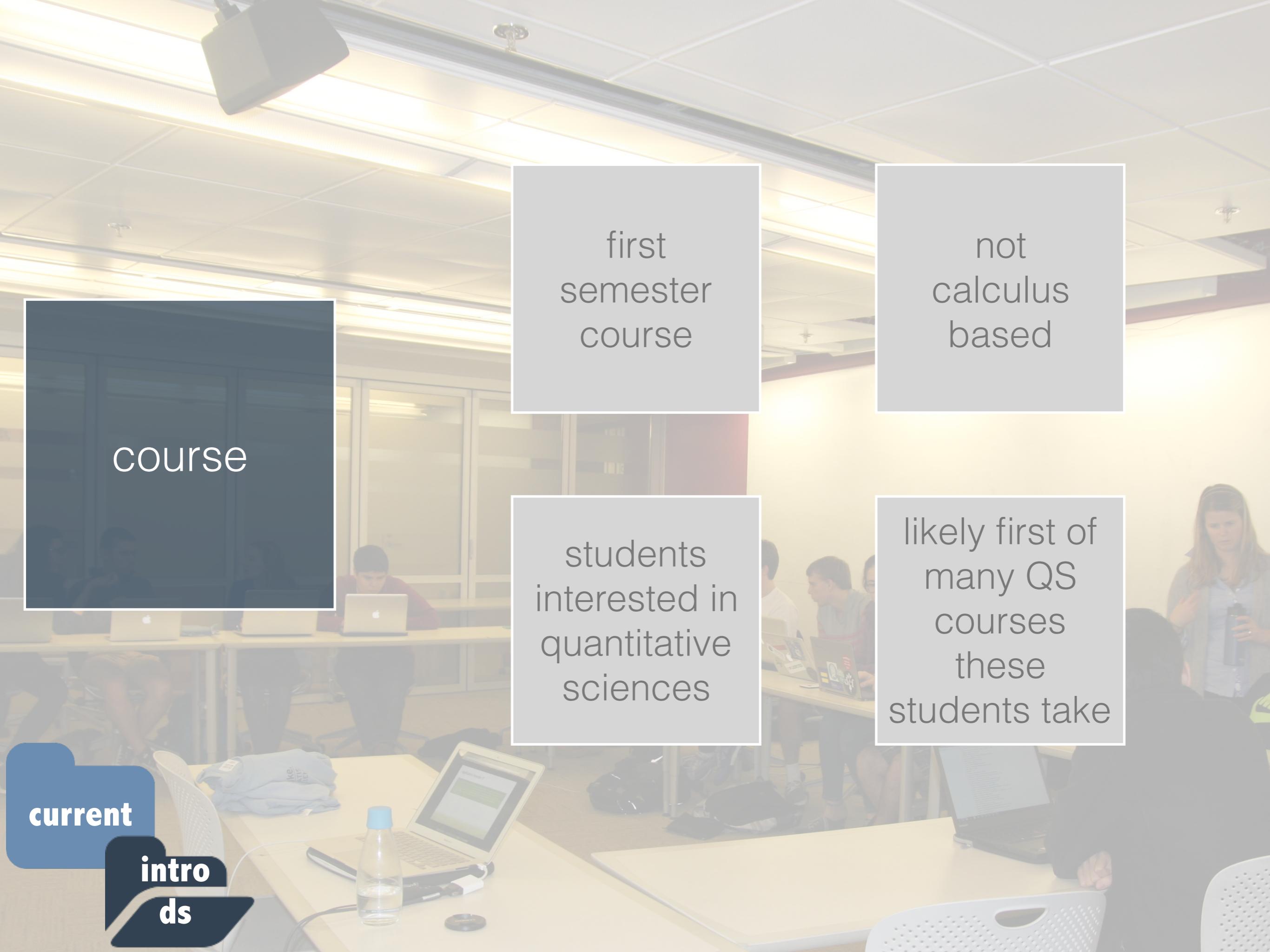
syntax
highlighting

keeps
code
organized

keeps
workspace
clean

current

intro
stat



course

first
semester
course

not
calculus
based

students
interested in
quantitative
sciences

likely first of
many QS
courses
these
students take

current

intro
ds

reproducibility

literate
programming

version
control

current

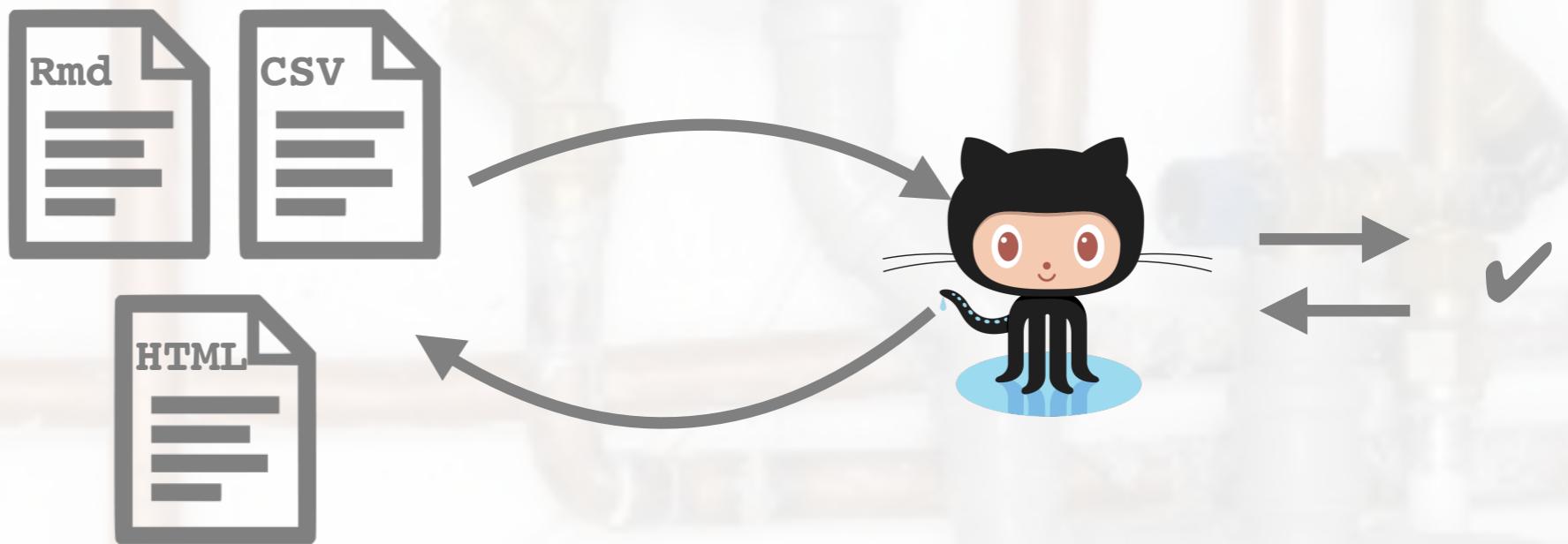
intro

ds

R R Studio[®]



workflow



current

intro

ds



can
students
handle it?

yes!

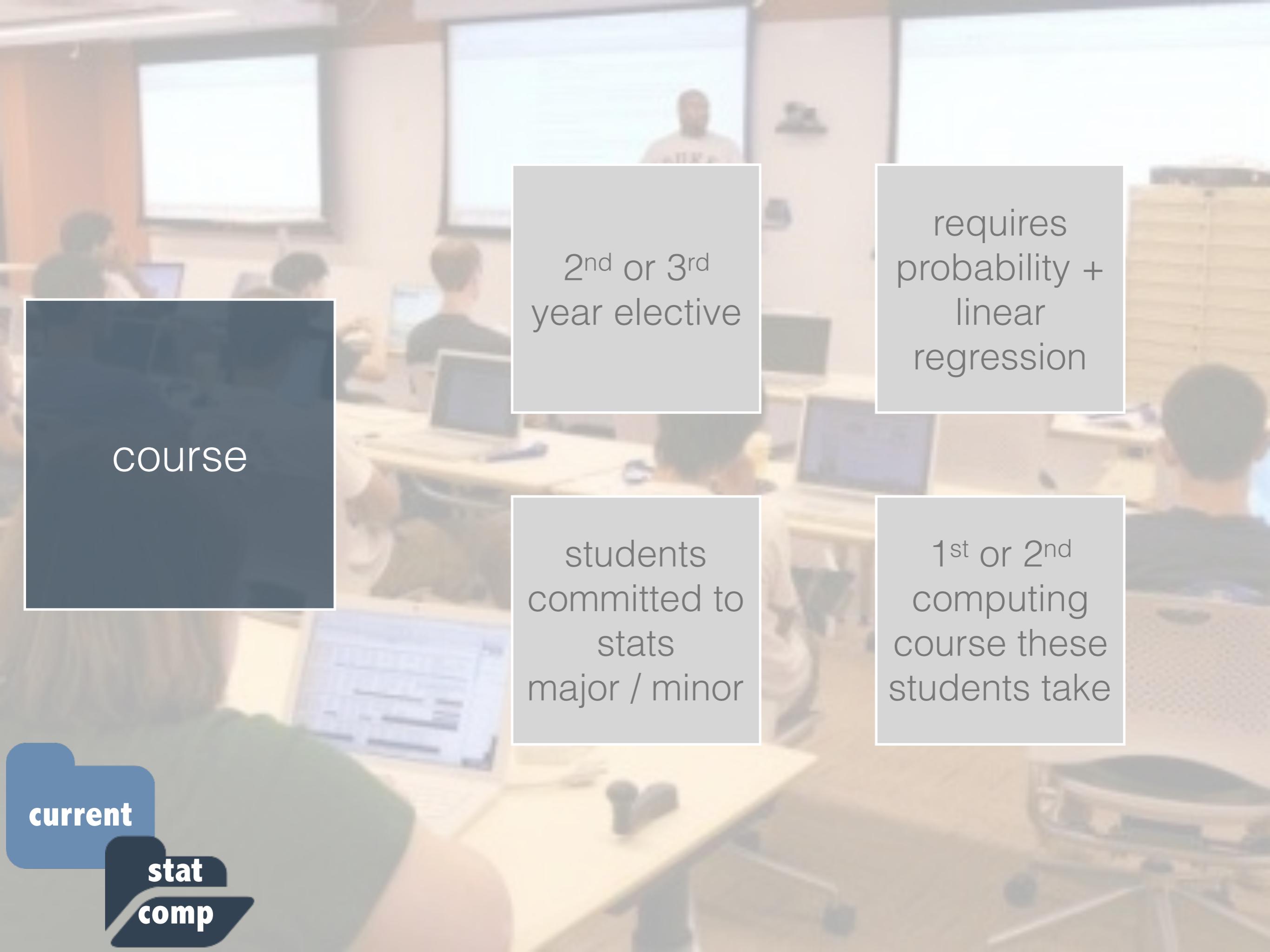
but...

instruction of
workflow
requires time
and care

current

intro

ds



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

reproducibility

literate
programming

version
control

build
tools

make

current

stat

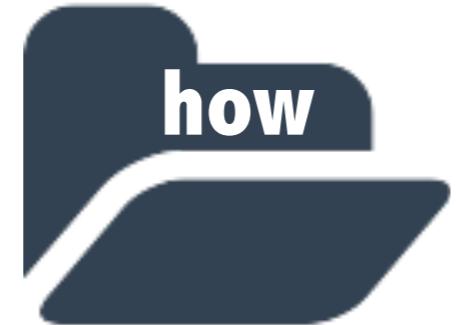
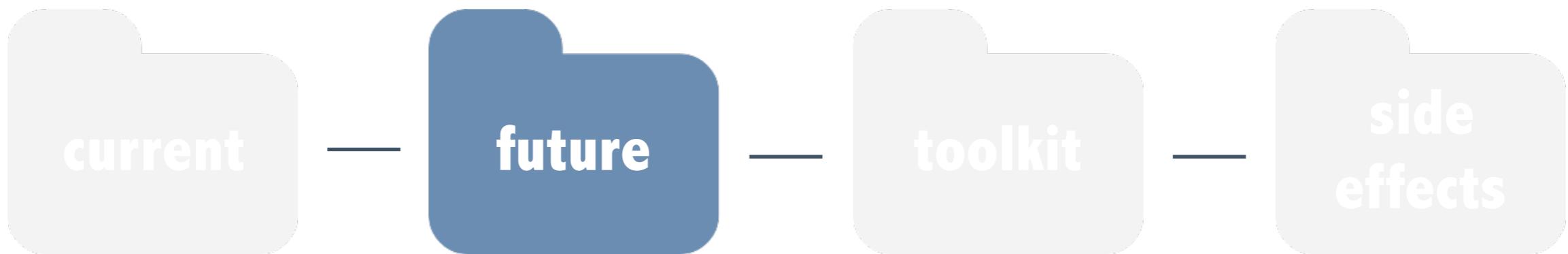
comp

R R Studio[®]



details at http://bit.ly/statcomp_jsm2016

grow toolkit along with the complexity of computation



what

capstone
course

senior
thesis /
independent
study

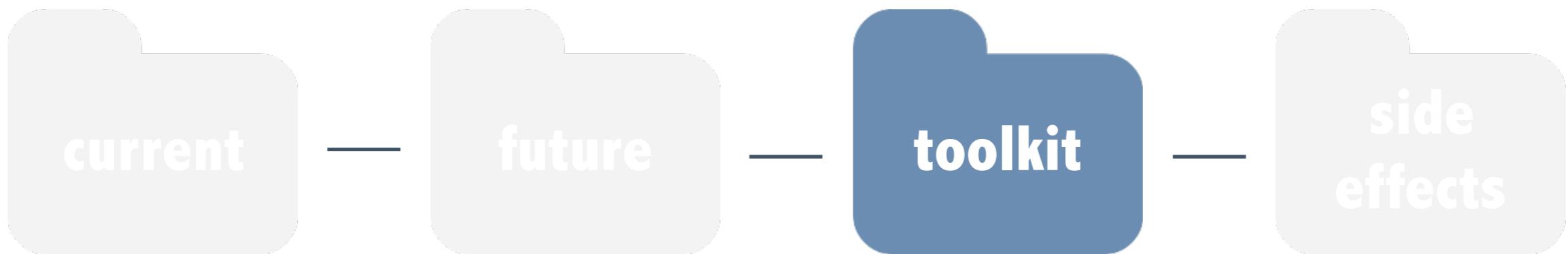
how

need
instructor
buy-in

needs
to be
part of
assessment

easily
adoptable
framework
will help

future



R

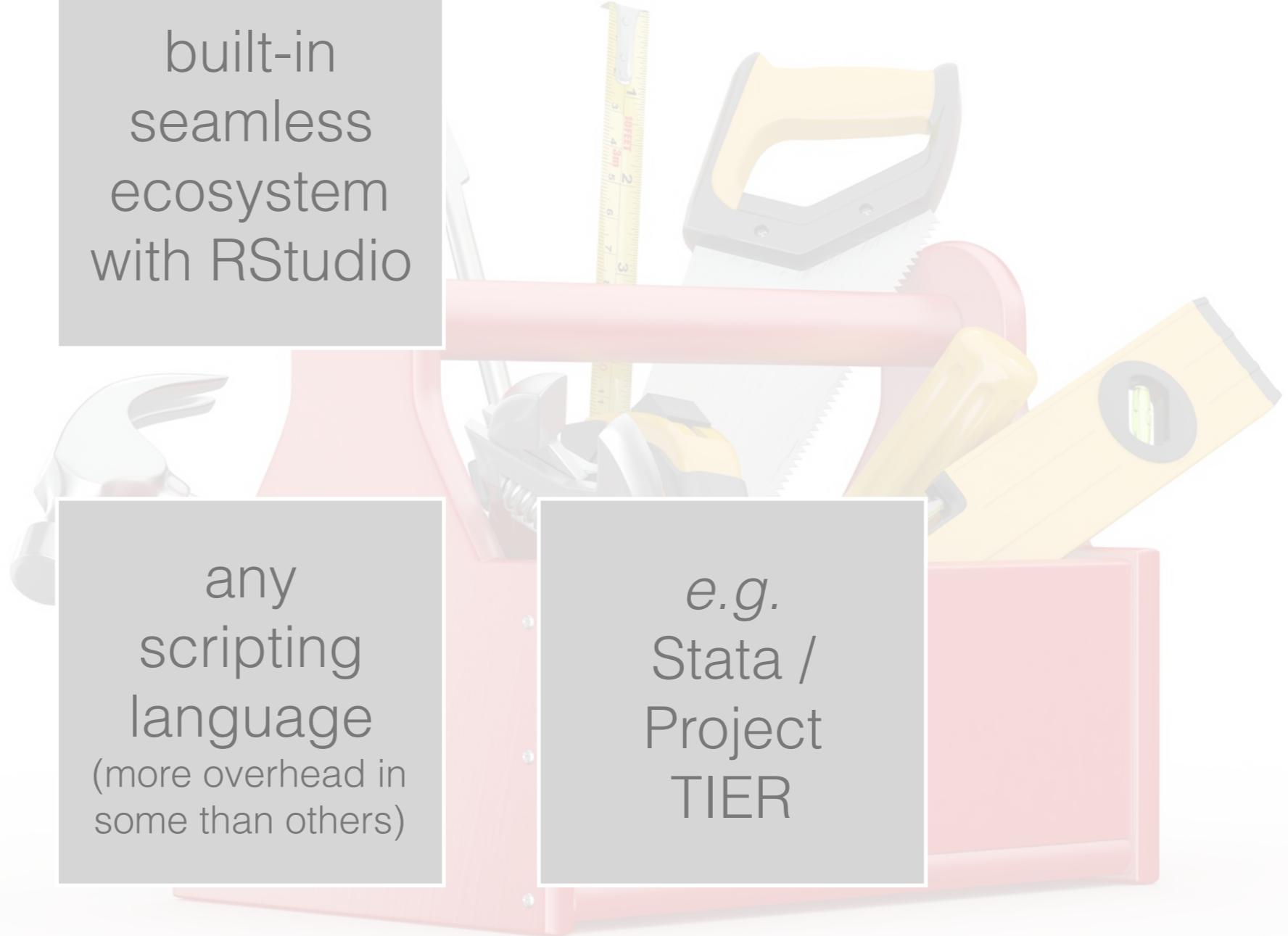
other



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