weaving reproducibility through the undergraduate statistics curriculum

mine çetinkaya-rundel

minebocek

mine@stat.duke.edu

mine-cetinkaya-rundel

bit.ly/ugrad_repro_jsm2016

two-pronged approach

#1

Convince researchers to adopt a reproducible research workflow

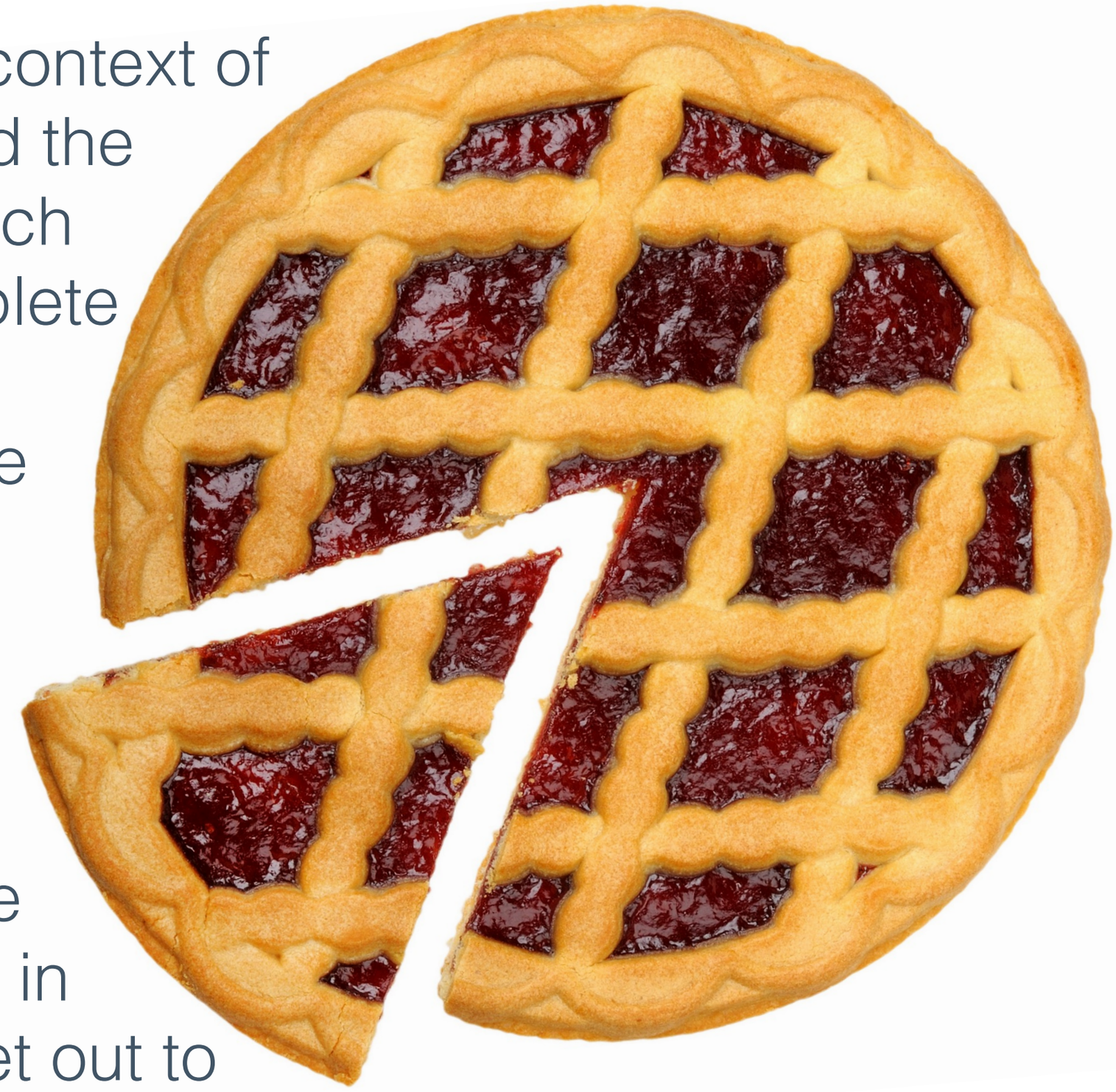
#2

Train new researchers who don't have any other workflow

bit.ly/ugrad_repro_jsm2016

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

bit.ly/ugrad_repro_jsm2016

current — future — toolkit — side effects

intro stat

intro ds

stat comp

bit.ly/ugrad_repro_jsm2016

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/ugrad_repro_jsm2016

reproducibility

literate programming

current

intro stat

bit.ly/ugrad_repro_jsm2016

workflow

traditional

data analysis
- descriptive stats
- plots & tables
- model output

writeup
- research question & context
- interpretations
- conclusionsa

copy-paste

lab report

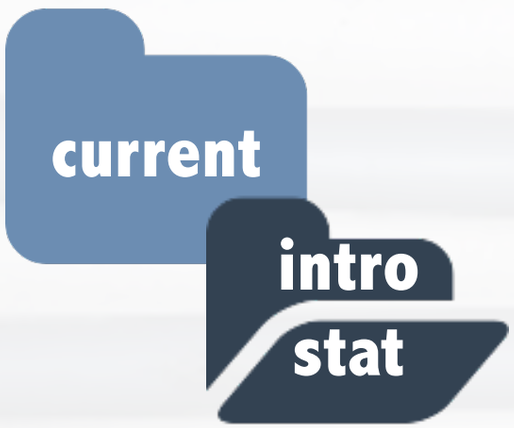
better

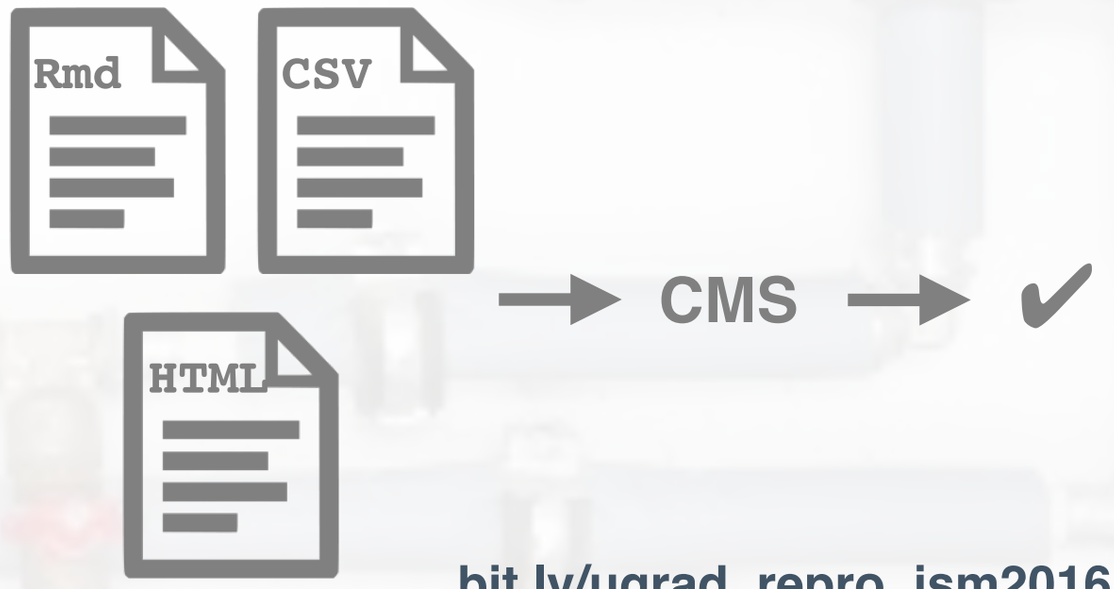
text block

data analysis

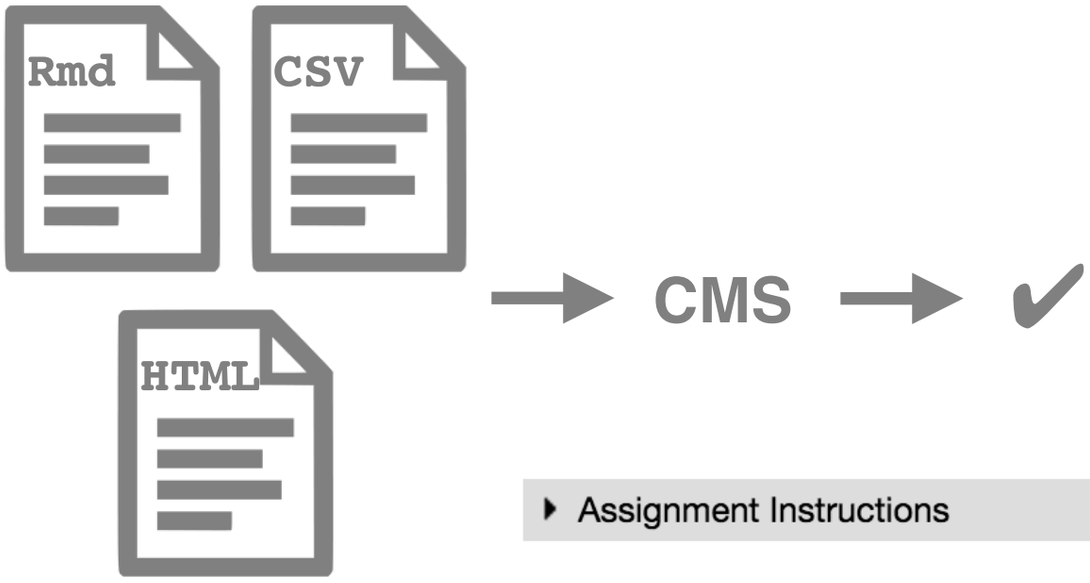
text block

data analysis





bit.ly/ugrad_repro_jsm2016



▸ 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: 97.0 (max 100.0)

Assign Grade Overrides

Instructor Summary Comments

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

Source | Styles | Format | Font | Size

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

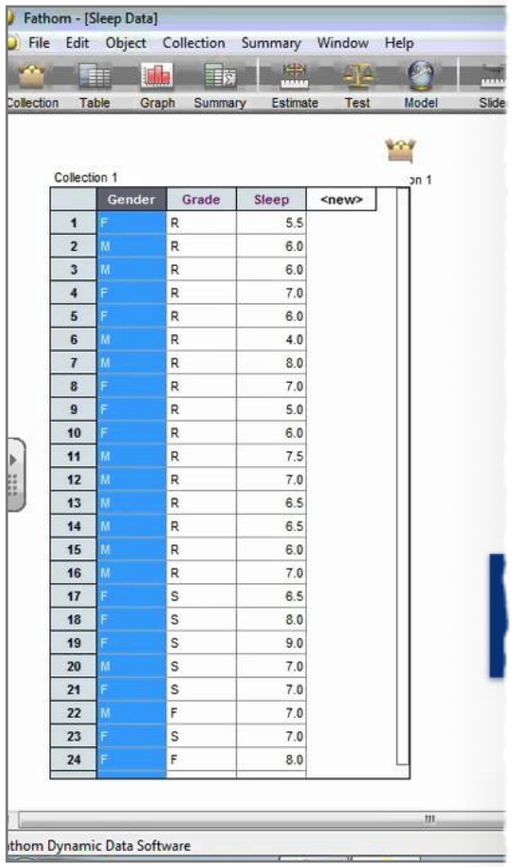
can students handle it?

yes!

point-and-click is no less overhead than scripting

current intro stat

bit.ly/ugrad_repro_jsm2016



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:

- 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**".

can students handle it?

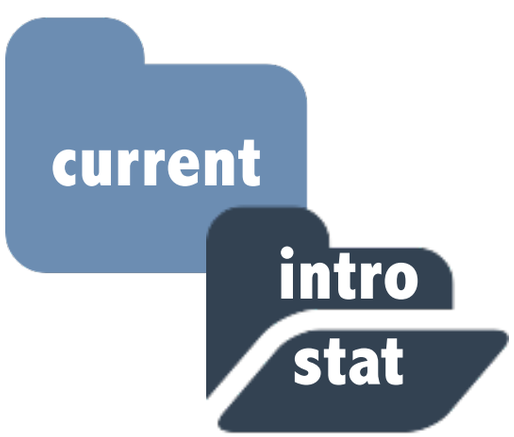
yes!

point-and-click is no less overhead than scripting

code + output always together

syntax highlighting

keeps code organized + workspace clean



bit.ly/ugrad_repro_jsm2016

course

first semester course

not calculus based

students interested in quantitative sciences

likely first of many QS courses these students take

current

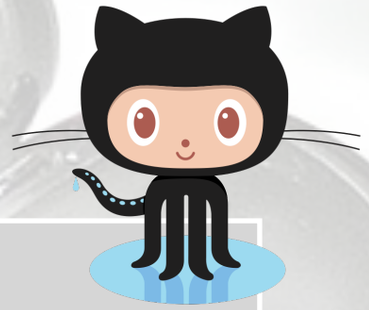
intro ds

bit.ly/ugrad_repro_jsm2016

reproducibility

literate programming

version control

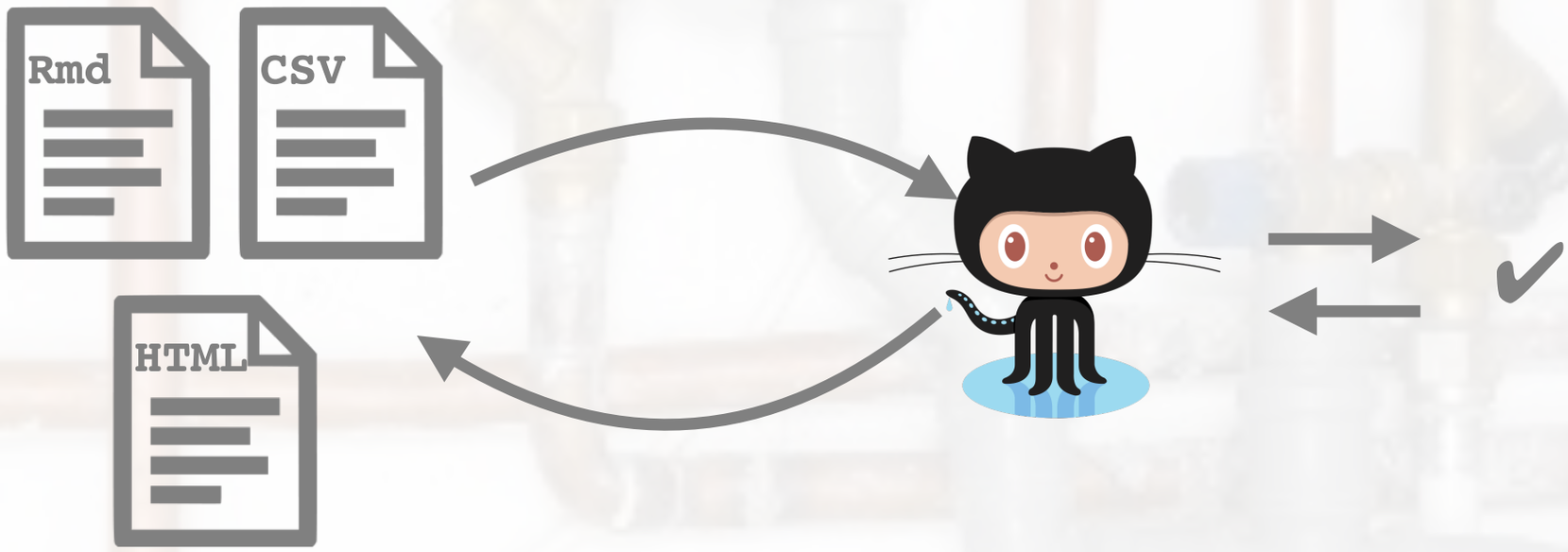


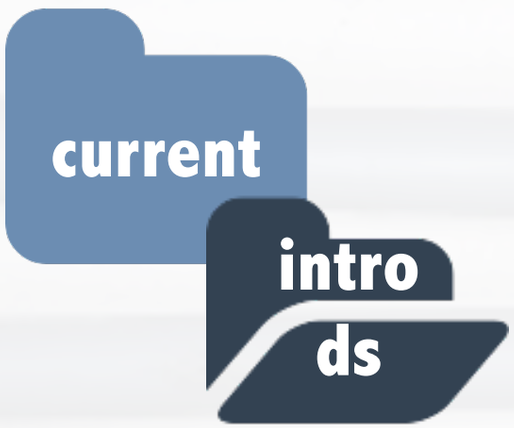
current

intro ds

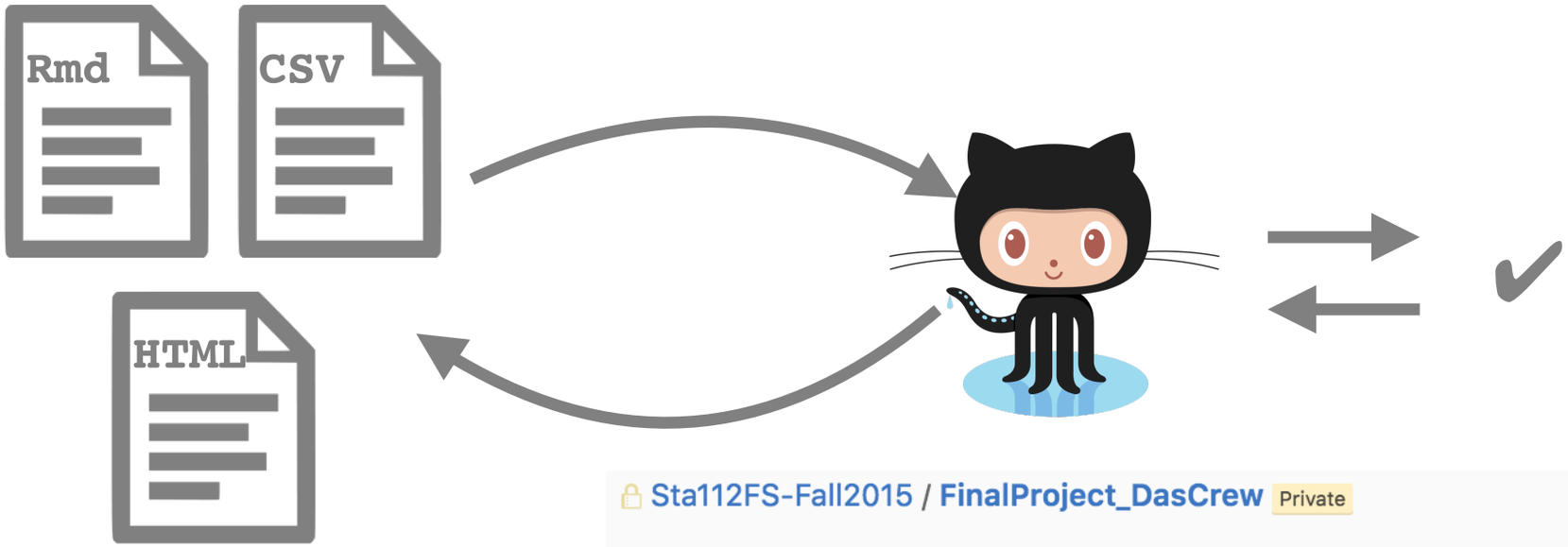
bit.ly/ugrad_repro_jsm2016

workflow





bit.ly/ugrad_repro_jsm2016



Sta112FS-Fall2015 / FinalProject_DasCrew Private

Unwatch 3 | Star 0 | Fork 0

Code | Issues 1 | Pull requests 0 | Wiki | Pulse | Graphs | Settings

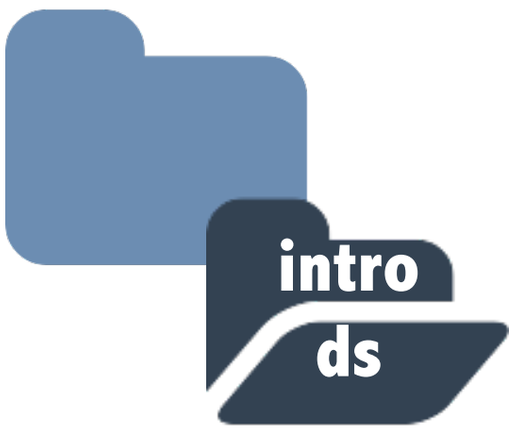
Repo for final project for Team Das Crew — Edit

110 commits | 1 branch | 0 releases | 1 contributor

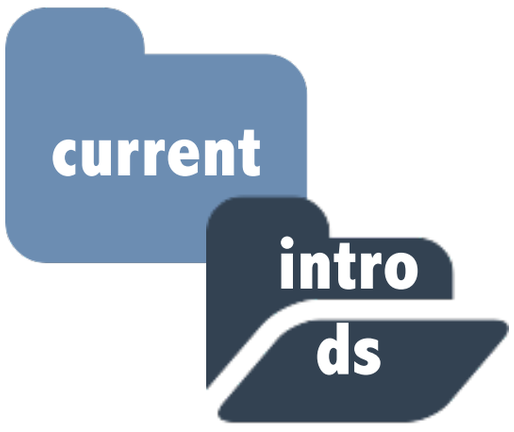
Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

Merge branch 'master' of https://github.com/Sta112FS-Fall2015/FinalPr... Latest commit 075ccdb on Dec 3, 2015

FinalPresentation-cache	Don't know what these do	8 months ago
Final_presentation_cache/html	Don't know what these do	8 months ago
Final_proj_cache/html	Don't know what these do	8 months ago
.gitignore	Created presentation and added even more explanations, analysis	8 months ago
FinalProject_DasCrew.Rproj	Maybe it will work this time	9 months ago
Final_presentation.Rmd	Finished	8 months ago
Final_proj.Rmd	Merge branch 'master' of https://github.com/Sta112FS-Fall2015/FinalPr...	8 months ago
Final_proj_shiny.Rmd	Fixed merge conflicts	8 months ago
Proposal.Rmd	reindenting	8 months ago
README.md	Struggling with the dataset of all things. Someone help?	9 months ago
ncdot-fatal-crashes.csv	Struggling with the dataset of all things. Someone help?	9 months ago
server.R	Added a whole bunch to server and ui for the ggmap stuff. says input ...	8 months ago
ui.R	Fixed fill of the bar plot	8 months ago



can students handle it?	yes!	but…	instruction of workflow requires time and care



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

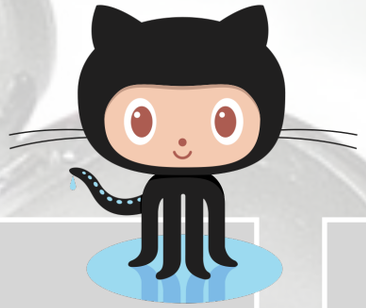
stat comp

bit.ly/ugrad_repro_jsm2016

reproducibility

literate programming

version control



build tools

make

current

stat comp

bit.ly/ugrad_repro_jsm2016

details at bit.ly/Sta323_2016

grow toolkit along with the complexity of computation

bit.ly/ugrad_repro_jsm2016

current — future — toolkit — side effects

bit.ly/ugrad_repro_jsm2016

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/ugrad_repro_jsm2016

current — future — **toolkit** — side effects

bit.ly/ugrad_repro_jsm2016

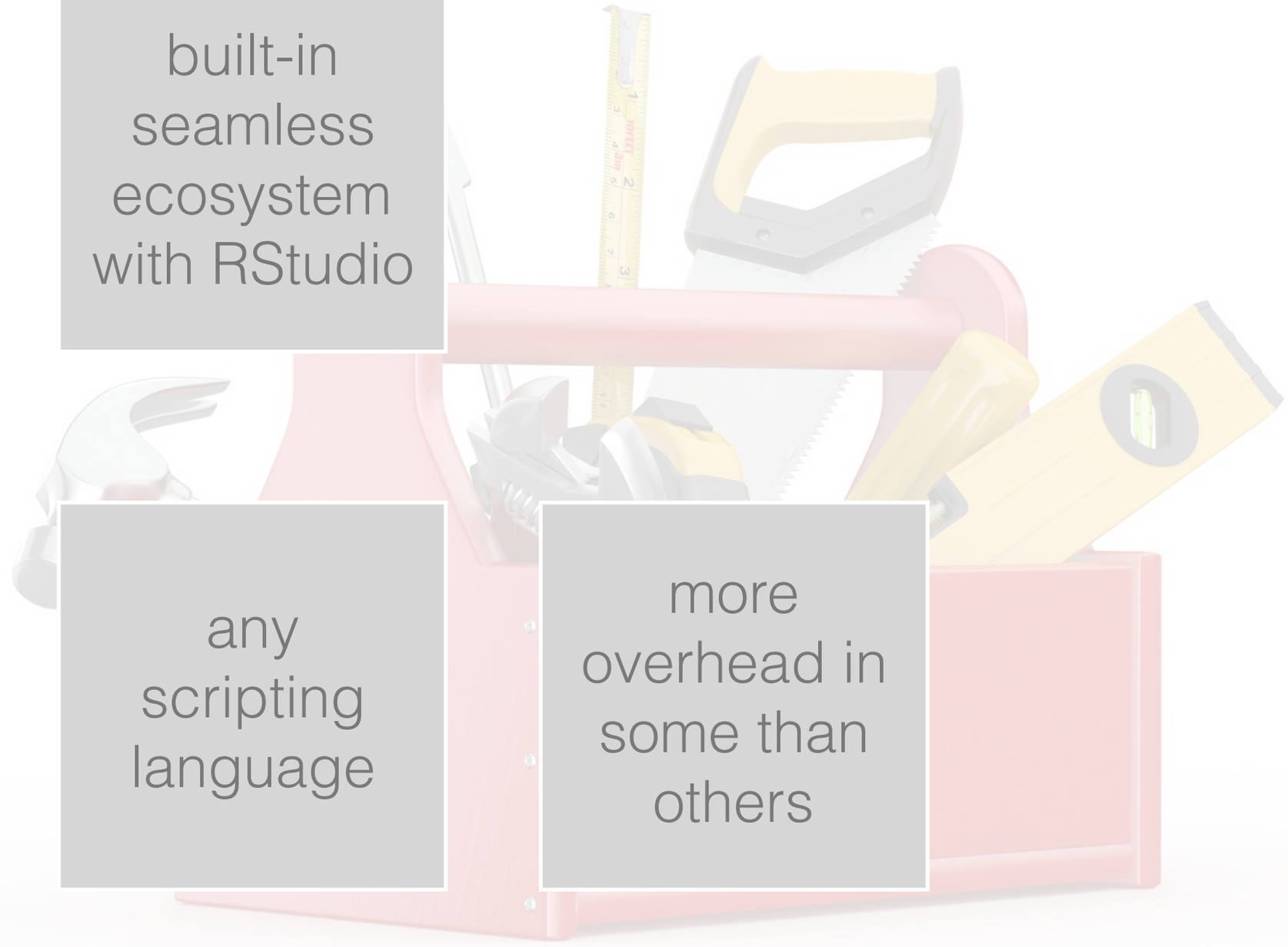
R

other

built-in seamless ecosystem with RStudio

any scripting language

more overhead in some than others



toolkit

bit.ly/ugrad_repro_jsm2016

current — future — toolkit — side effects

instruc-tors

students

bit.ly/ugrad_repro_jsm2016

for instructors

easy Q&A

easy grading

for students

easy collaboration

self-promotion

side effects

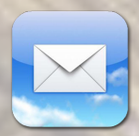
bit.ly/ugrad_repro_jsm2016

thank you!

bit.ly/ugrad_repro_jsm2016

README with links to resources + course pages


minebocek


mine@stat.duke.edu


mine-cetinkaya-rundel