Building a gateway between classrooms and data science using QUBESHub 12th Gateway Computing Environments Conference M. Drew LaMar and Sam Donovan 2017/10/25

Gateway challenges

Core challenges

- Audience (*faculty and students*): Gateway must address two different audiences.
- **Context (***research and education***)**: Faculty themselves have very different needs.
- **Content (***faculty training***)**: Faculty need training not only in computational tools and quantitative/statistical concepts (which are underdeveloped in undergraduate curricula), but also in the pedagogical strategies in bringing these to the classroom.

Cultural challenges

- **Isolation**: Difficult for faculty to share teaching materials and strategies, due to fear of public scrutiny.
- **Recognition**: Recognition of teaching scholarship in hiring, promotion and tenure is underdeveloped and widely variable.
- Adoption: Adopting existing teaching materials from repositories is challenging as modifications are almost always necessary.

Educational Gateway

QUBES: Quantitative Undergraduate Biology Education and Synthesis



QUBES is a community of math and biology educators who share resources and methods for preparing students to use quantitative approaches to tackle real, complex, biological problems.

QUBES: Quantitative Undergraduate Biology Education and Synthesis



https://qubeshub.org 4/29

Data Science challenges (in Biology)

- **Faculty training**: Biology faculty are not domain experts in statistics and data analysis
- Messy data: Many opportunities in using real data, but *real data is messy*
- **Complex software**: Many opportunities in using research tools, but *software is complex*
- Accessibility:
 - Lack of user-friendly *open-source* software
 - Computers in labs might require substantial setup prior to use of software
 - Student laptops are highly variable in computational ability, setup and maintenance
- **Cognitive overload**: Biology students need to learn biology, math, physics, chemistry, statistics, experimental design, data skills, etc.
 - **Scripting/Programming**: Reproducibility is becoming more important in science.



Image credit: http://cgrb.oregonstate.edu/service-spotlight/software-carpentry

Faculty Training Solutions QUBES Faculty Mentoring Networks

Faculty Mentoring Networks are:

- Online groups, typically 10-15 faculty members
- Focused on a specific topic or material
- Typically meet every two weeks over a period of several months
- Led by teams of expert content and pedagogy mentors



Literacy Using Botany

January 15-May 01, 2018



Accessibility Solutions Software "in the cloud"



Accessibility Solutions Jupyter and R Notebooks



http://jupyter.org/ Why I love R Notebooks

The Bridge

How can we get student's manipulating and analyzing data as fast as possible (i.e. *doing science*), while at the same time creating a scaffold to scripting skills?

The Bridge

How can we focus students' attention on meaningful disciplinary work while reducing the technical overhead to do that work?



tional/large/e778_firefly_serenity_cutaway_poster_set_21pg / 29 Image credit:







1. Highly-accessible as a free, open-source web application

This organization Search	Pull requests Issues Marketplace Ex	plore 🌲 🕂 🛪 🐺 🕶	
Serenity IN DEVELOPMENT. To follow development of Serenity, please visit the main repository at https://github.com/serenity-r/serenity. • https://github.com/serenity-r/serenity • https://github.com/serenity-r/serenity • Repositories 34			
Pinned repositories Customize pinned repositories			
 ≡ serenity Data science in the classroom ■ R 	 ≡ serenity.qubes Serenity code for installation and execution on QUBES. ■ R	 gridstackr Forked from jbkunst/gridstackr An implementation of gridstack.js for R JavaScript 	
≡ mxgraphr An R htmlwidget for the mxGraph JavaScript library	phosphorr R HTMLWidget implementation of PhosphorJS		
JavaScript	🔵 JavaScript 🛛 ★ 1		

https://github.com/serenity-r



2. Design keeps the focus on the data and the data life cycle



Program

NJASP: Not Just Another Statistics Package



3. Streamlined communication and reporting with R Markdown



Image credit: http://rmarkdown.rstudio.com/authoring_quick_tour.html#output_formats



4. Reports will include workflows that can be reproduced or repurposed



1	library(magrittr)
2	library(tidyverse)
3	
4	filtered_storms <- dplyr::storms %>%
5	filter(category = 5, year >= 2000) %>%
6	<pre>unite("date", year:day, sep = "-") %>%</pre>
7	group_by(name) %>%
8	filter(pressure = max(pressure)) %>%
9	<pre>mutate(date = as.Date(date)) %>%</pre>
10	arrange(desc(date)) %>%
11	ungroup() %T>%
12	<pre>print()</pre>

GUI -> Code

Image credit (left): http://rstudio.github.io/shiny/tutorial/#hello-shiny Image credit (right): https://benjaminlmoore.wordpress.com/



5. Workflows will follow best practices in data science



Follow V

New blog post: "Don't teach students the hard way first" varianceexplained.org/r/teach-hard-w... #rstats

Imagine you were going to a party in an unfamiliar area, and asked the host for directions to their house. It takes you thirty minutes to get there, on a path that takes you on a long winding road with slow traffic. As the party ends, the host tells you "You can take the highway on your way back, it'll take you only ten minutes. I just wanted to show you how much easier the highway is."

Wouldn't you be annoyed? And yet this kind of attitude is strangely common in programming education.



5. Workflows will follow best practices in data science





5. Workflows will follow best practices in data science





6. Simultaneous exploration of multiple representations of data



Crosstalk: Using Crosstalk



7. Multiple formats for communication and dissemination



Storyboard: htmlwidgets showcase



Shiny: ggplot2 linked brushing

flexdashboard for R: flexdashboard Examples

8. Multiple modes of data import







8. Multiple modes of data import





9. Integration with built-in learning management systems





10. Link computational modeling with analysis of the resulting data



http://www.netlogoweb.org/launch



- 1. Highly-accessible as a free, open-source web application
- 2. User-interface design keeps the focus on the data and the data life cycle
- 3. Communication and reporting will be streamlined
- 4. Reports of results will include workflows that can be reproduced or repurposed
- 5. Workflows will follow best practices in data science
- 6. Simultaneous exploration of multiple representations of data
- 7. Multiple formats for communication and dissemination
- 8. Multiple modes of data import
- 9. Integration with built-in learning management systems
- 10. Link computational modeling with analysis of the resulting data



Thank you!

QUBES Leadership Team

Kristin Jenkins Carrie Diaz-Eaton Jeremy Wojdak Alison Hale Gaby Hamerlinck Arietta Fleming-Davies

Inspired by Radiant

Vincent Nijs





Follow Serenity development at https://github.com/serenity-r Slides created via the R package xaringan.

This material is based upon work supported by the National Science Foundation under DBI 1346584, DUE 1446269, DUE 1446258, and DUE 1446284. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Science Gateways Community Institute