Serenity now! Keep calm and do science with real data in the classroom

M. Drew LaMar, Sam Donovan and Hayley Orndorf

Tuesday, June 19, 2018, 2:00 pm
Discussion

What types of **data literacy** do you want your students to gain competency in?

- **Plan**: Design for data collection
- **Collect**: Collect data
- **Assure**: Check & inspect
- **Describe**: Assign metadata
- **Preserve**: Long-term archiving
- **Discover**: Find relevant data
- **Integrate**: Combine data sets
- **Analyze**: Explore data

[Image from DataONE](http://www.dataone.org)
Discussion: Usability vs. Flexibility

Image credit: "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners" by Catherine D'Ignazio and Rahul Bhargava
Discussion: Usability vs. Flexibility
Where do you feel like you fall in this space in the context of teaching?

Image credit: "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners" by Catherine D'Ignazio and Rahul Bhargava
Discussion: Usability vs. Flexibility

Image credit: "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners" by Catherine D'Ignazio and Rahul Bhargava
Data Science pedagogical challenges
Data Science pedagogical challenges

- **Messy data**: Many opportunities in using real data, but *real data is messy*
Data Science pedagogical challenges

- **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*
Data Science pedagogical challenges

- **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
Data Science pedagogical challenges

- **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
Data Science pedagogical challenges

- **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
Data Science pedagogical challenges

- **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.
Data Science pedagogical challenges

- **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.
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?
Serenity: Data Science in the Classroom
Serenity
Data science in the classroom
Serenity
Data science in the classroom

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

https://github.com/serenity-r
Serenity
Data science in the classroom

2. Design keeps the focus on the data and the data life cycle
Serenity
Data science in the classroom

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

NJASP: Not Just Another Statistics Package

Image from R for Data Science, by Garrett Grolemund and Hadley Wickham
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
4. Reports will include workflows that can be reproduced or repurposed

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

David Robinson
@drob

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.
Serenity
Data science in the classroom

5. Workflows will follow best practices in data science

Image credit: https://www.rstudio.com/about/gear/
Serenity
Data science in the classroom

5. Workflows will follow best practices in data science

Image credit: https://rviews.rstudio.com/2017/06/08/what-is-the-tidyverse/
Serenity
Data science in the classroom

6. Simultaneous exploration of multiple representations of data

Crosstalk: Using Crosstalk
Serenity
Data science in the classroom

7. Multiple formats for communication and dissemination

Storyboard: htmlwidgets showcase

Shiny: ggplot2 linked brushing

flexdashboard for R: flexdashboard Examples
Serenity
Data science in the classroom

8. Multiple modes of data import

Image credit: https://twitter.com/datacamp/status/846256025472847872
Serenity
Data science in the classroom

8. Multiple modes of data import

https://www.re3data.org/
9. Integration with built-in learning management systems

Serenity
Data science in the classroom

10. Link computational modeling with analysis of the resulting data

http://www.netlogoweb.org/launch
Serenity
Data science in the classroom

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
Gaby Hamerlinck
Deb Rook
Pam Bishop

Arietta Fleming-Davies
Nicole Chodkowski
Elia Crisucci
Jenny Kwan
Sondra LoRe
Kevin Kidder

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.