

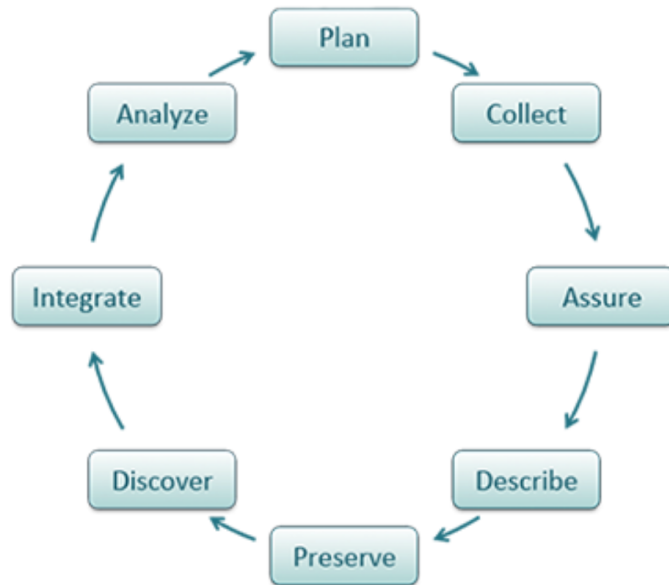
Serenity now! Keep calm and do science with real data in the classroom

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

Discussion: Usability vs. Flexibility

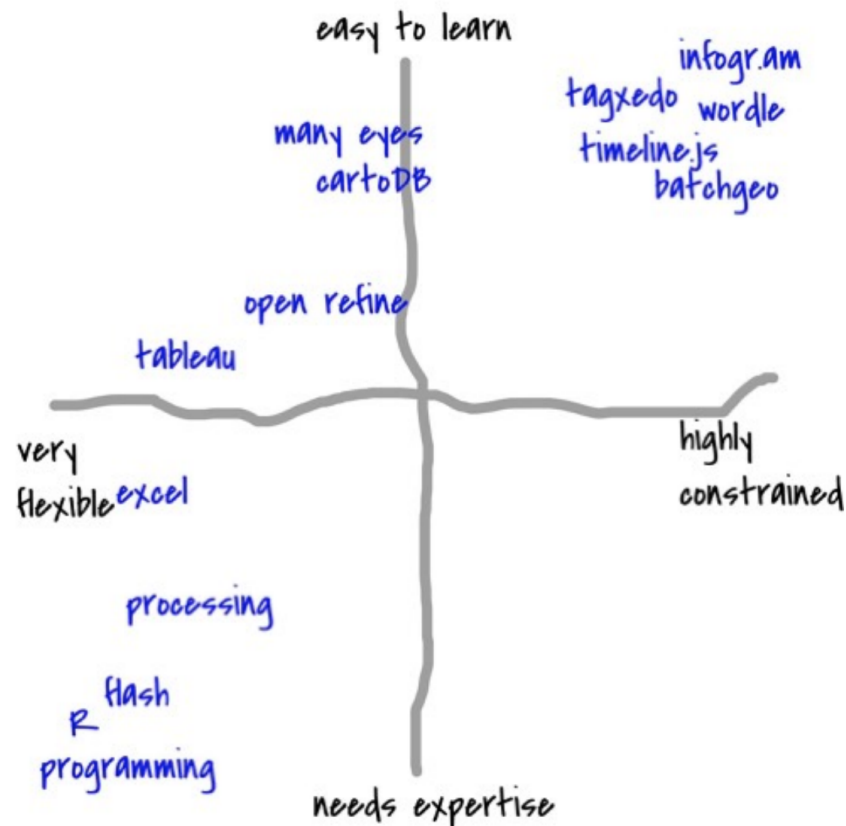


Image credit: "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners" by Catherine D'Ignazio and Rahul Bhargava

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

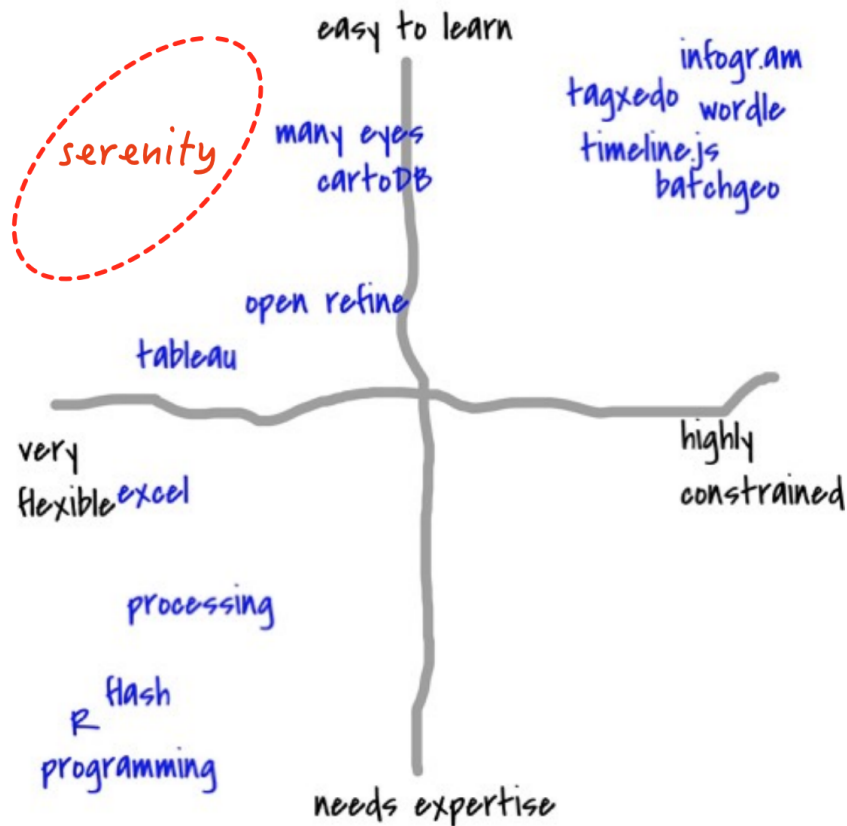


Image credit: "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners" by Catherine D'Ignazio and Rahul Bhargava

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

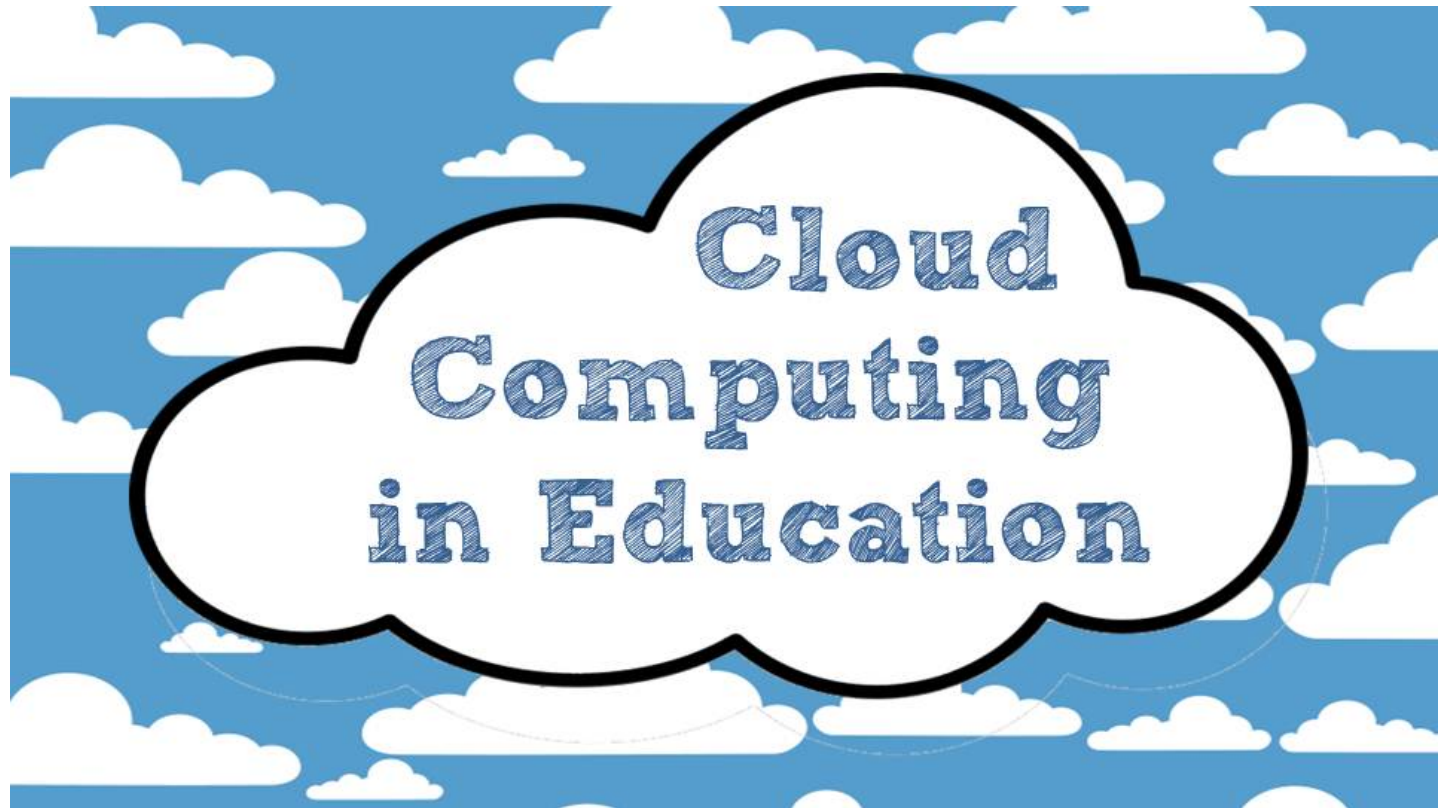
- **Messy data:** Many opportunities in using real data, but *real data is messy*
- **Complex software:** Many opportunities in using research tools, but *software is complex*
- **Accessibility:**
 - Lack of user-friendly *open-source* software
 - Computers in labs might require substantial setup prior 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.

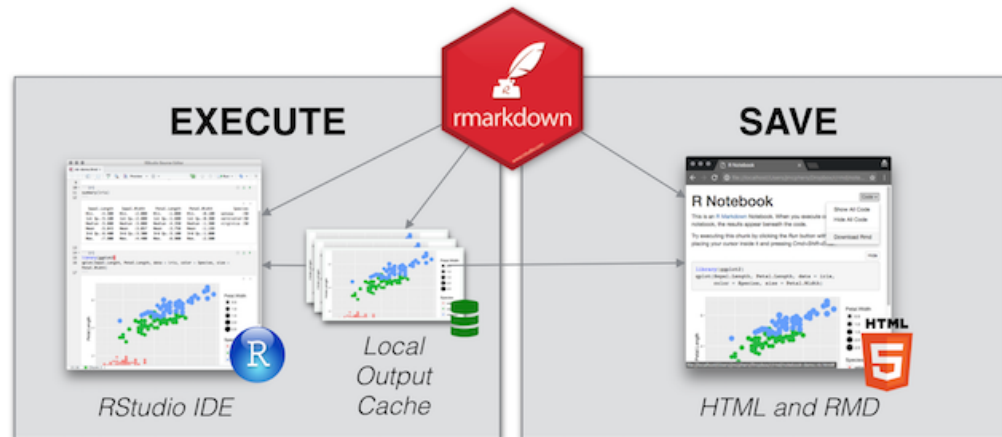
Accessibility Solutions

Software "in the cloud"



Accessibility Solutions

Jupyter and R Notebooks



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

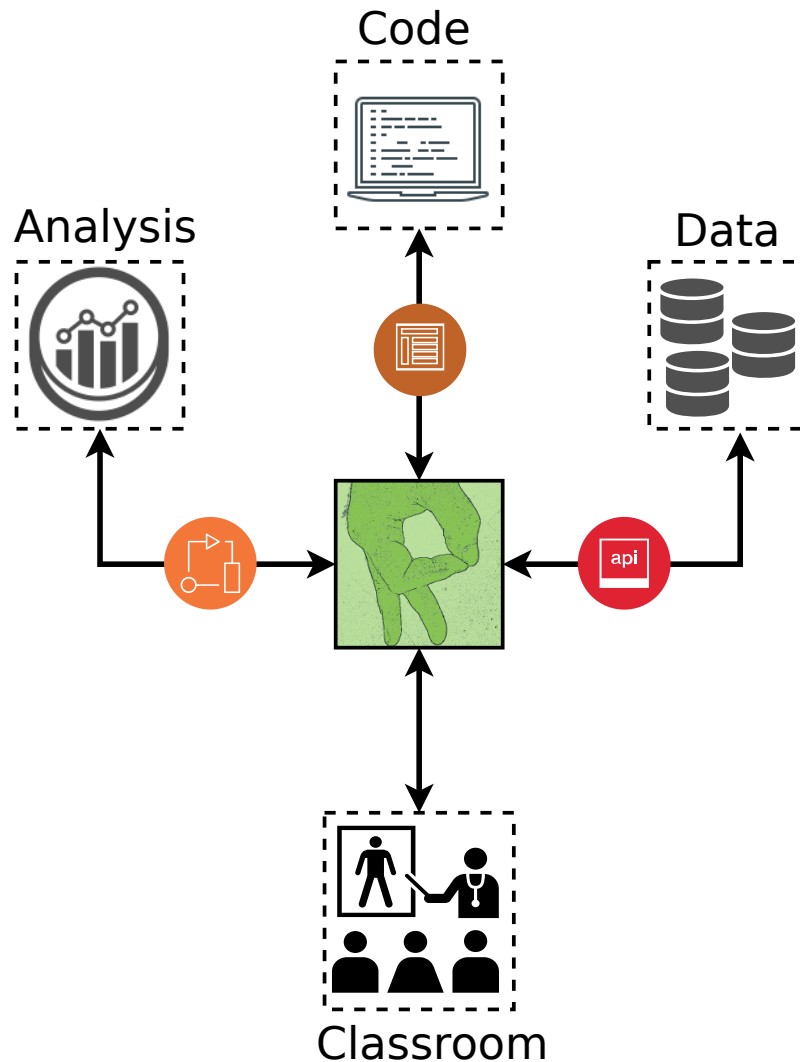
The Bridge

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

The screenshot shows the GitHub organization page for 'Serenity'. The header includes the GitHub logo, a search bar, and navigation links for 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. The organization's profile section features a green hand icon, the name 'Serenity', and a description: 'IN DEVELOPMENT. To follow development of Serenity, please visit the main repository at <https://github.com/serenity-r/serenity>'. Below this, there are links to the repository and the organization's email. The 'Repositories' tab is selected, showing 34 repositories. A 'Pinned repositories' section displays five repositories: 'serenity' (Data science in the classroom, R), 'serenity.qubes' (Serenity code for installation and execution on QUBES, R, 2 forks), 'gridstackr' (Forked from jbkunst/gridstackr, An implementation of gridstack.js for R, JavaScript), 'mxgraphr' (An R HTMLWidget for the mxGraph JavaScript library, JavaScript), and 'phosphor' (R HTMLWidget implementation of PhosphorJS, JavaScript, 1 star).

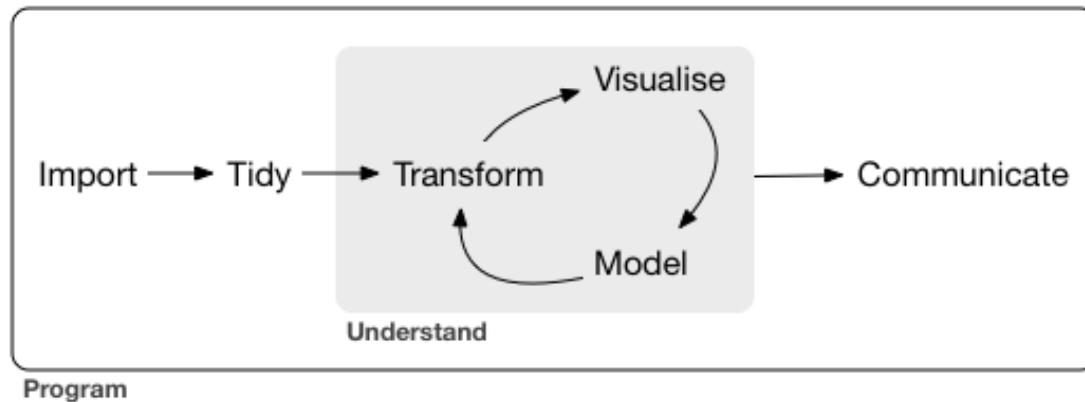
<https://github.com/serenity-r>

Serenity

Data science in the classroom



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

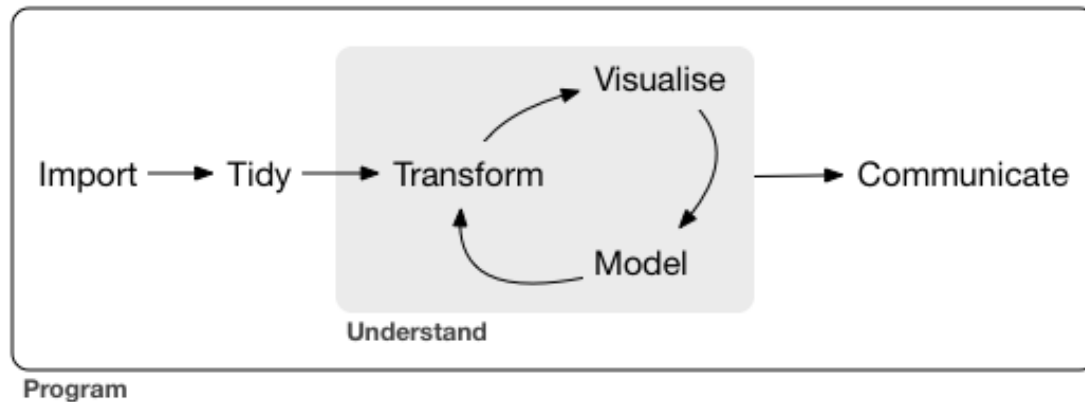


Serenity

Data science in the classroom



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



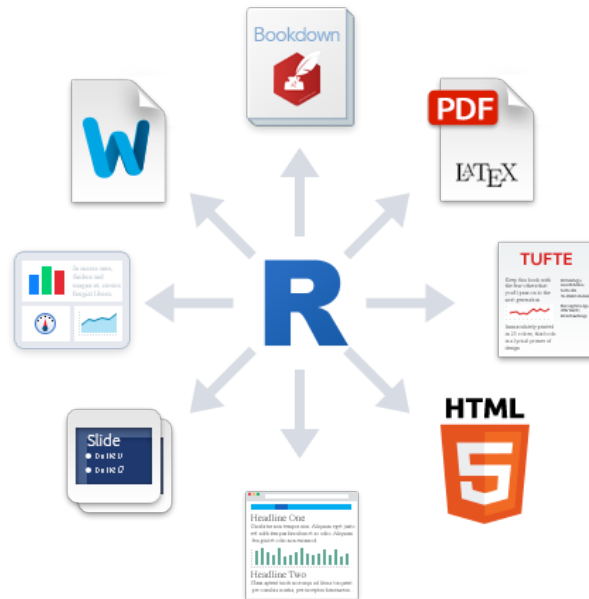
NJASP: Not Just Another Statistics Package

Serenity

Data science in the classroom



3. Streamlined communication and reporting with R Markdown

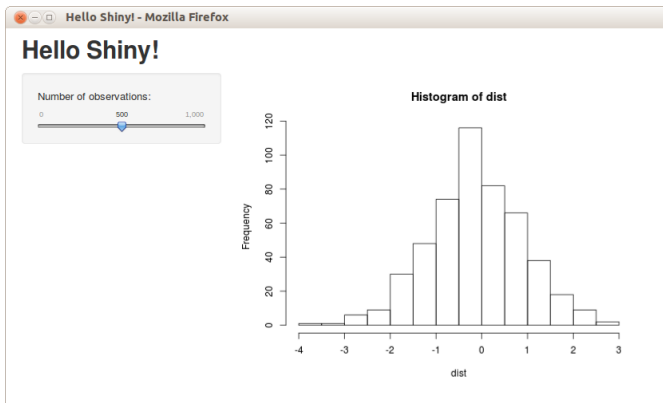


Serenity

Data science in the classroom



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



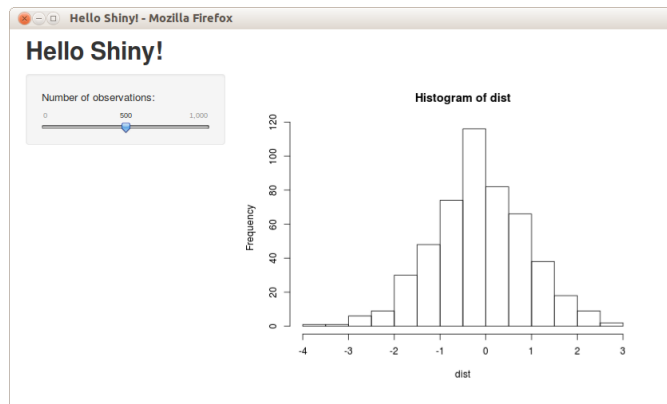
```
1 library(magrittr)
2 library(tidyverse)
3
4 filtered_storms <- dplyr::storms %>%
5   filter(category == 5, year >= 2000) %>%
6   unite("date", year:day, sep = "-") %>%
7   group_by(name) %>%
8   filter(pressure == max(pressure)) %>%
9   mutate(date = as.Date(date)) %>%
10  arrange(desc(date)) %>%
11  ungroup() %T>%
12  print()
13
```

Serenity

Data science in the classroom



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   unite("date", year:day, sep = "-") %>%
7   group_by(name) %>%
8   filter(pressure = max(pressure)) %>%
9   mutate(date = as.Date(date)) %>%
10  arrange(desc(date)) %>%
11  ungroup() %T>%
12  print()
13
```

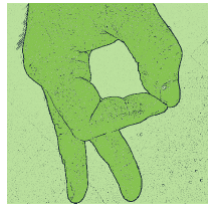
GUI -> Code

Image credit (left): <http://rstudio.github.io/shiny/tutorial/#hello-shiny>

Image credit (right): <https://benjaminlmoore.wordpress.com/>

Serenity

Data science in the classroom



5. Workflows will follow best practices in data science



David Robinson

@drob

Follow



New blog post: "Don't teach students the hard way first"

[varianceexplained.org/r/teach-hard-w...](https://varianceexplained.org/r/teach-hard-way-first/)
[#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

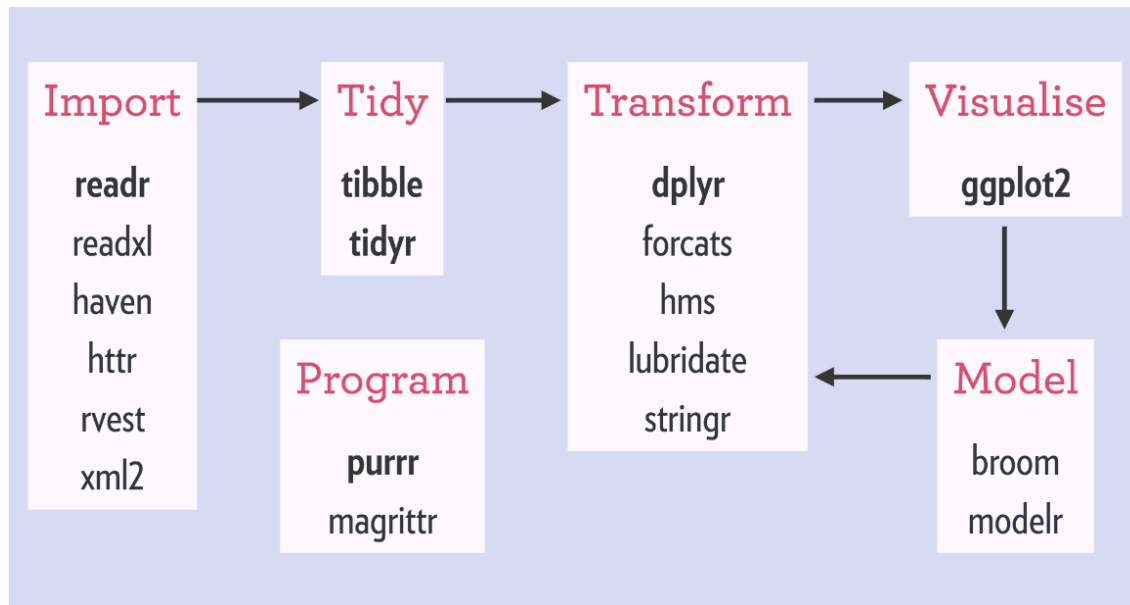


Serenity

Data science in the classroom



5. Workflows will follow best practices in data science

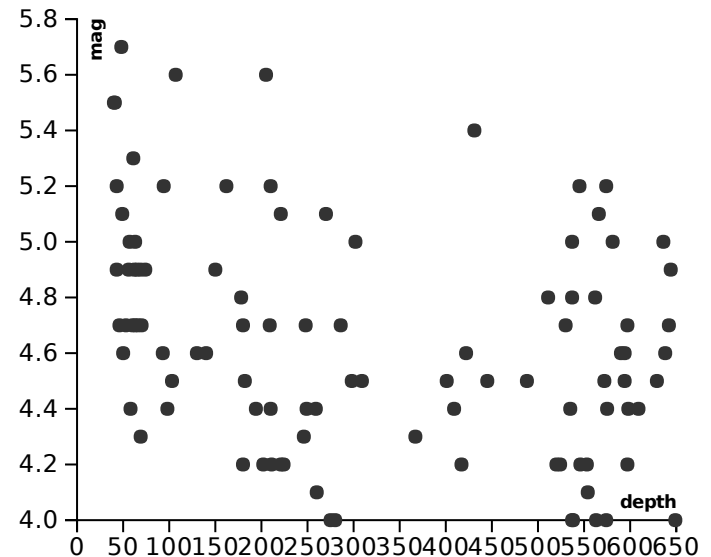
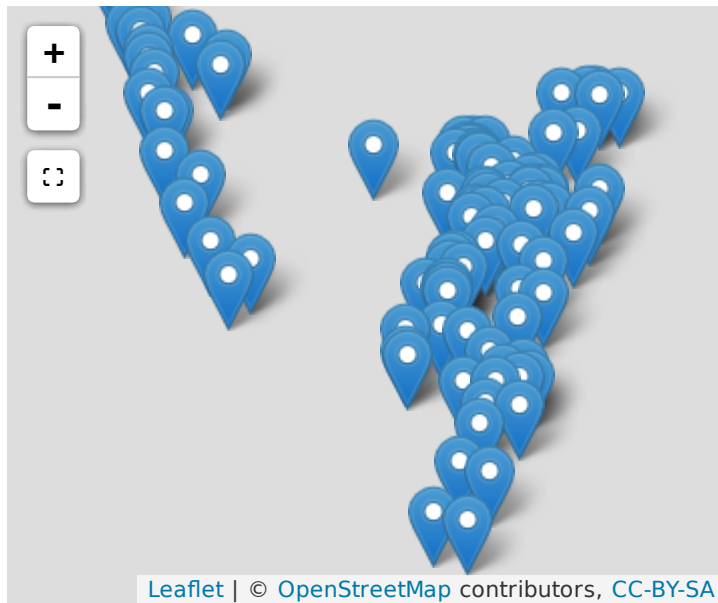


Serenity

Data science in the classroom



6. Simultaneous exploration of multiple representations of data



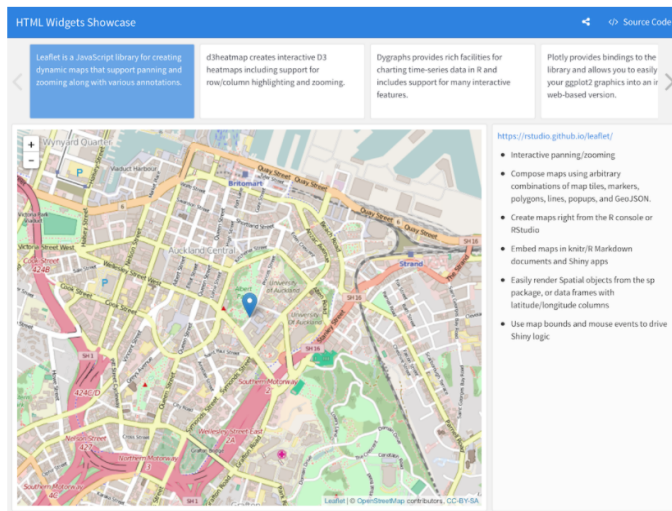
Crosstalk: Using Crosstalk

Serenity

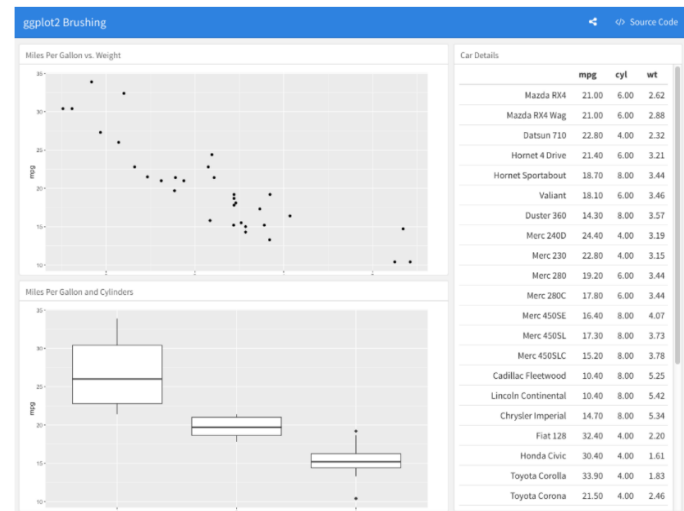
Data science in the classroom



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

enterprise infrastructure
technology operation
information
scorecards
analyze text mining
metrics
applications
connection technique
solution stakeholder

R Data Import Tutorial

Minitab SPSS®
 sas SYSTAT®

{JSON}
JavaScript Object Notation

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES



LMS Market Share for All Institutions

Microsoft, Google, Oracle/PeopleSoft, Cengage Learning, Pearson, SunGard, Datatel & Jenzabar

1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

* A consortium of the U. Michigan, Indiana U., MIT, Stanford and uPortal.

Consortium* 2004

Sakai

moodle

Open Source

1996

1999

U. Guelph 1999

Pearson acquires eCollege

eCollege

eCollege becomes Pearson LearningStudio

LearningStudio

Desire2Learn

Innovative Learning Technology

ANGEL LEARNING

Indiana U. (IUPUI) 2002

Prometheus

George Washington 1998

Cornell* 1997

*Cornell's early application called CourseInfo

Virginia C. U. 1997

WEB COURSE in a BOX

WebCT

CE

Vista

Bb Learn* R9.1

Blackboard

Enable. Innovate. Everywhere.

The thickness of the line represents marketshare.

Supports Select Academic Departments

\$40-60K

Standard System Cost of Software License and Hardware Infrastructure From Vendor Point of View.

\$100-400K

Supports Entire University

The 2005 - 2009 data is from a Campus Computing project <http://www.campuscomputing.net/survey>. This data represents how many U.S. campuses use a particular LMS as their campus standard.

www.deltainitiative.com

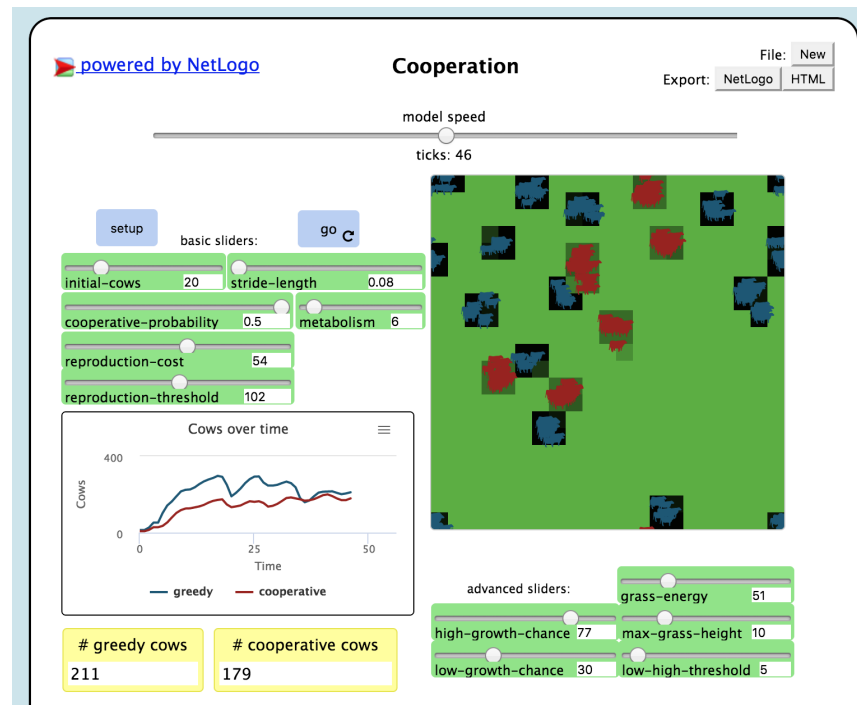
delta initiative

Serenity

Data science in the classroom



10. Link computational modeling with analysis of the resulting data



<http://www.netlogoweb.org/launch>

Serenity

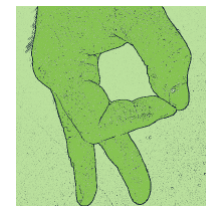
Data science in the classroom



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.