Scaling Data Science Education

Sean Kross

May 10, 2016

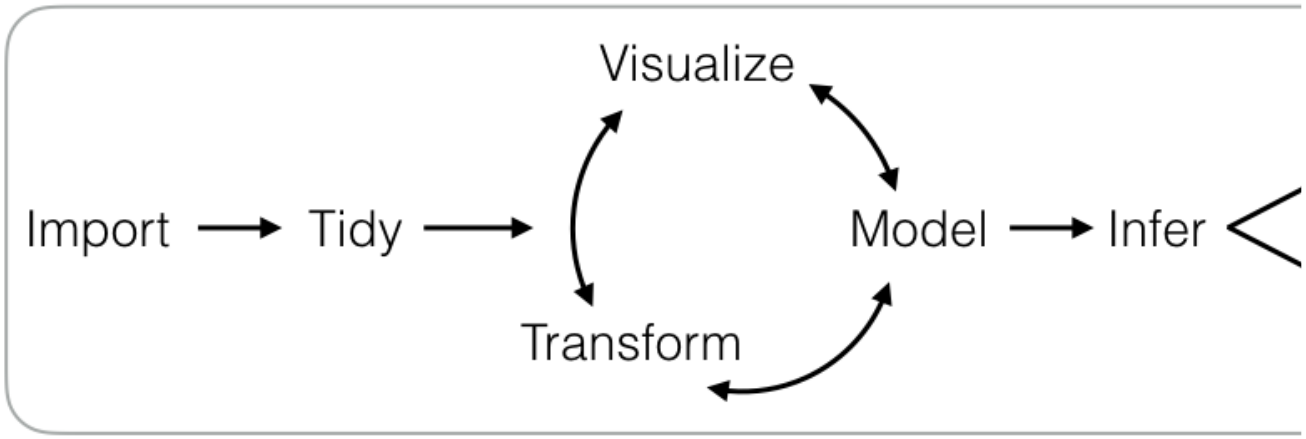
Who am I?

- Genomics at New York University
- Computer Science at The University of Maryland
- Statistics at Johns Hopkins University
- Currently I work in the Department of Biostatistics at JHU
- The path I've taken is problematic.

What is this talk about?

- What is data science?
- What is data science education?
- How can we scale data science education? (And by "we" I mean you and I because I need your help!)

What is Data Science?: Process



https://github.com/rstudio/RStartHere

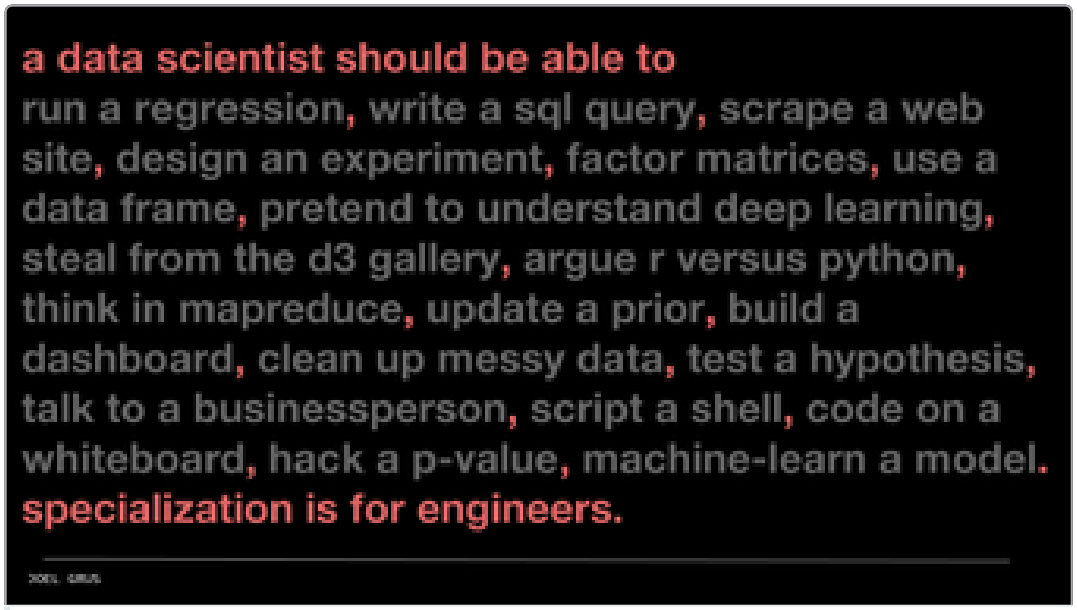
What is Data Science?: Skills

The tip of the iceberg:

(Probably and American colloquialism (sorry))

- Computer Science
  - Machine Learning
  - Distributed Computing
  - Parallel Computing
  - Algorithms
- Statistics
  - Inference
  - Regression
- + the scientific methods from a specific field (genomics, sociology, molecular gastronomy, etc)

Data Science Education: Motivation



Jenny Bryan
@JennyBryan

Follow

A data scientist should be able to ...
My favourite slide from @joelgrus's fun talk #ddtx16
8:33 PM - 17 Jan 2016

520 694

What is the underlying concept here?

What is Data Science Education?

- How can we most effectively teach students to manipulate and interrogate data using a computer?
- How can we scale this kind of education to millions of people?
- How can we improve upon issues that are currently present in computer science and mathematics education?

Attempts at Scaling Data Science Education

What is swirl?

What is swirl?

- swirl is an R package that turns the R console into an interactive learning environment for data science.
- swirl takes advantage of the R console's call-and-response behavior.
- swirl provides an authentic environment, there is no separation between where a student is learning data science and where they will go on to practice data science.

What problems does swirl solve?

- swirl is free and open source, and it always will be.
- swirl is available in multiple langauges (we're always looking for volunteer translators!)
- Anyone can make and distribute their own swirl course for free.

What is a swirl course?

- A swirl course is a collection of lessons.
- Each lesson contains about 20 minutes of instruction about a particular topic in data science.
- Lessons are written and structured using yaml and R.

Demo

swirl

```
> library(swirl)
> swirl()

| Welcome to swirl! Please sign in. If you've been here before, u
| as you did then. If you are new, call yourself something unique

What shall I call you? sean

| Thanks, sean. Let's cover a couple of quick housekeeping items
| our first lesson. First of all, you should know that when you s
| means you should press Enter when you are done reading and read

...  <-- That's your cue to press Enter to continue

| Also, when you see 'ANSWER:', the R prompt (>), or when you are
| from a list, that means it's your turn to enter a response, the
| continue.
```

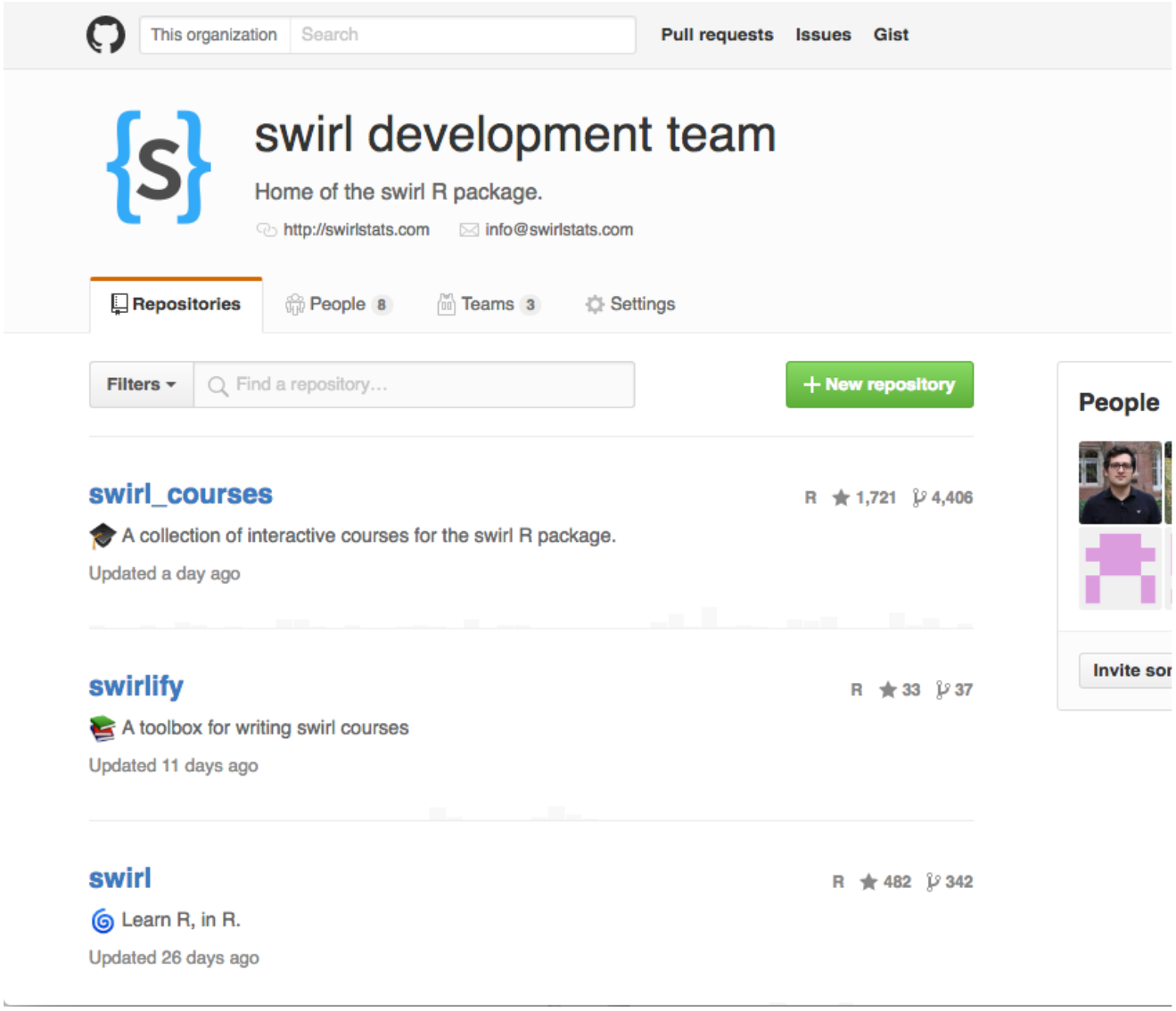
swirl

```
Selection: 1

  |

| In this lesson, we will explore some basic building blocks of the R programming l

...

  |===

| If at any point you'd like more information on a particular topic related to R, y
| help.start() at the prompt, which will open a menu of resources (either within RS
| web browser, depending on your setup). Alternatively, a simple web search often y
| you're looking for.

...

  |=====

| In its simplest form, R can be used as an interactive calculator. Type 5 + 7 and

>
```

Coming to swirl this summer:

- Improved lessons
- More courses
- Better channels for distributing swirl courses
- Feedback for course instrcutors to assess student's performance

We want you!


This organization
Search
Pull requests
Issues
Gist
swirl development team
Home of the swirl R package.
http://swirlstats.com
info@swirlstats.com
Repositories
People 8
Teams 3
Settings
Filters
Find a repository...
New repository
People
swirl_courses
R 1,721 4,406
A collection of interactive courses for the swirl R package.
Updated a day ago
swirlify
R 33 37
A toolbox for writing swirl courses
Updated 11 days ago
swirl
R 482 342
Learn R, in R.
Updated 26 days ago

Acknowledgements

- Co-Creators:
  - Nick Carchedi, Bill Bauer, Gina Grdina
- Generous support from:
  - Brian Caffo, Roger Peng, Jeff Leek
- Financial support from:
  - The United States National Institutes of Health

Questions?

- Find me:
  - On Twitter: @seankross
  - On GitHub: seankross
  - Online: http://seankross.com
  - Email me: sean@seankross.com
- Find swirl:
  - On Twitter: @swirlstats
  - On GitHub: swirldev
  - Online: http://swirlstats.com
  - Email swirl: info@swirlstats.com