# The democratization of data science education

Sean Kross UC San Diego Design Lab 2017-10-11

## A little about me

- New! Improved!
- Advised by Philip Guo
- Main interests:
  - Data Science
  - Online Education
  - Open Science











#### **The Johns Hopkins Data Science Lab**







#### Jeff Leek

**Roger Peng** 

#### **Brian Caffo**

#### < All Collections

#### Collection idea for us?

#### Practical Data Science for Stats - a PeerJ Collection

Data Science	Statistics	Scientific Computing and Simulation
Computer Education Com		mputational Science Social Computing
Software Engine	Ionce and Medical Education Computational Biology	
Human-Compu	Anthropology Programming Languages	
Visual Analytics	Graphics	Bata Mining and Machine Learning

#### September 27, 2017 proprint Forecasting at scale

1.145 downloads 3 615 views Sean J Taylor, Benjamin Letham https://doi.org/10.7287/peerj.preprints.3190v2

September 1, 2017 preprint How to share data for collaboration 768 downloads 3,490 views Shannon E Ellis, Jeffrey T Leek https://doi.org/10.7287/peerj.preprints.3139v5



Practical Data Science for Stats

The "Practical Data Science for Stats" Collection contains preprints focusing on the practical side of data science workflows and statistical analysis. Curated by Jennifer Bryan and Hadley Wickham.

There are many aspects of dayto-day analytical work that are almost absent from the conventional statistics literature and curriculum. And yet these activities account for a considerable share of the time and effort of data analysts and applied statisticians.

The goal of this collection is to

#### https://peerj.com/collections/50-practicaldatascistats/

## Rationale: "Let's put in-person courses online to augment in-person teaching."

Course	Enrollment	Completions
Mathematical Biostatistics Boot Camp 1	109,789	4,150
Mathematical Biostatistics Boot Camp 2	23,842	944
Computing for Data Analysis	243,987	21,069
Data Analysis	193,126	6,500
Biostatistics Case Study	39,140	3,322

#### We were on to something.

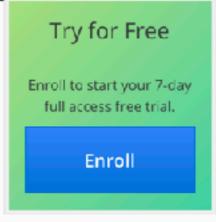
#### coursera

Catalog Search catalog Q

For Enterprise



#### Data Science About this Specialization Specialization Courses Pricing Launch Your Career in Data Science. A nine-course introduction to data science, developed and taught by leading professors. Creators



FAO

Financial Aid is available for learners who cannot afford the fee. Learn more and apply.

About This Specialization

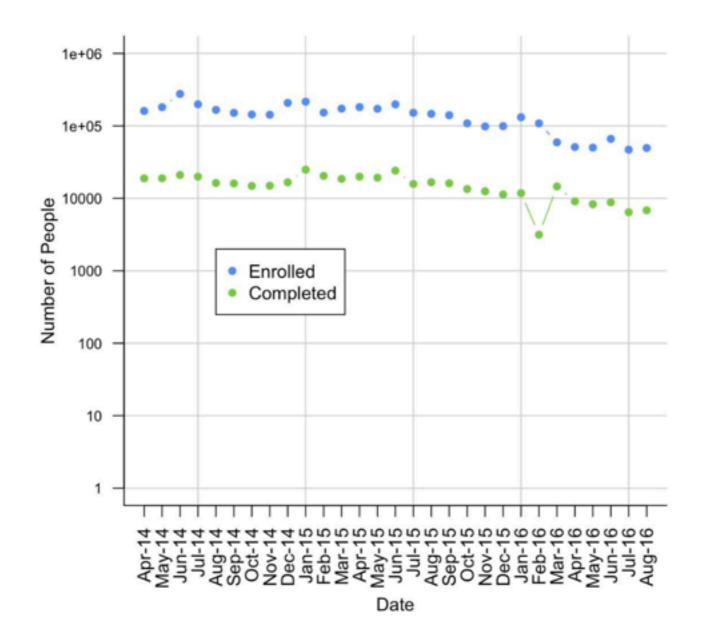
#### Ask the right questions, manipulate data sets, and create visualizations to communicate results.

This Specialization covers the concepts and tools you'll need throughout the entire data science pipeline, from asking the right kinds of questions to making inferences and publishing results. In the final Capstone Project, you'll apply the skills learned by building a data product using real-world data. At completion, students will have a portfolio demonstrating their mastery of the material.

## Nine Courses

- 1. The Data Scientist's Toolbox
- 2. R Programming
- 3. Getting and Cleaning Data
- 4. Exploratory Data Analysis
- 5. Reproducible Research
- 6. Statistical Inference
- 7. Regression Models
- 8. Practical Machine Learning
- 9. Developing Data Products

#### **Enrollment and Completions of the Data Science Specialization**



## Key innovations

### Give everything away for free



#### Capstones -> Portfolios -> Jobs







## Run every course every month

## Integrate content



https://ubc-mds.github.io/

# {Swirl}

Learn R, in R.

swirl teaches you R programming and data science interactively, at your own pace, and right in the R console!

Owndependidt/Lysankrose github to * -master - Ritudo         Image: Ima	🗯 RStudio File Edit Code View Plots Session Build Debug Profile Tools Window	Help 🕢 🕴 🛜 🔽 🕪 71% 🔟 Tue 10:20 PM Sean Kross Q 🧔 😑			
Untitled1x       Console Terminal x R Markdown x         1       -/Developer/GitHub/seankross.gitHub.lof >         1       > library(swirl)         1       > wirl()         1       Welcome to swirl! Please sign in. If you've been here         1       before, use the same name as you did then. If you are new,         1       call yourself something unique.         1       Thanks, sean. Let's cover a couple of quick housekeeping         1       terms before we begin our first lesson. First of all, you         1       terms before we begin our first lesson. First of all, you         1       terms before we begin our first lesson. First of all, you         1       terms when you are done reading and ready to continue.         1       terms when you see 'ANSWER:', the prompt (>), or when you         1       press Enter to continue         1       Also, when yous se'ANSWER:', the means it's your turn         iere ablad Environment is empty       Files Plots Packages Help Viewer         1       Name       Siz	//Developer/GitHub/seankross.github.io - master - RStudio				
1       -/Developer/CitHub/seankross.github.io/         1       -/Developer/CitHub/seankross.github.io/         1       > library(swir1)         1       I Hi Type swir1() when you are ready to begin.         > swir1()       I Hi Type swir1() when you are ready to begin.         I Welcome to swir1! Please sign in. If you've been here       before, use the same name as you did then. If you are new, i call yourself something unique.         1       I Welcome to swir1! Please sign in. If you've been here         1       I welcome to swir1! Please sign in. If you've been here         1       I welcome to swir1! Please sign in. If you are new, i call yourself something unique.         What shall I call you? sean       I Thanks, sean. Let's cover a couple of quick housekeeping i tems before we begin our first lesson. First of all, you i should how that when you see '', that means you should i press Enter when you are done reading and ready to continue.         I mipport Dataset - I List -	🔍 🚽 🔩 🚽 🚽 🔚 📑 🛛 📥 🛛 📣 Co to file/function 👘 🗄 🗧 🗄 👻 Addins	- 🤹 seankross.github.io -			
<pre>1 &gt; library(swir1) 1 Hi! Type swir1() when you are ready to begin. 5 swir1() 1 Hi! Type swir1() when you are ready to begin. 5 swir1() 1 Welcome to swir1! Please sign in. If you've been here 1 before, use the same name as you did then. If you are new, 1 call yourself something unique. What shall I call you? sean 1 Thanks, sean. Let's cover a couple of quick housekeeping 1 thanks, sean. Let's cover a couple o</pre>	Untitled1 ×	Console Terminal × R Markdown ×			
1:1       (Top Level):         1:1       (Top Level):         Interviewent       R Script:         Image: Script:       Image: Script:	🗢 🗆 🖉 🔚 🖸 Source on Save 🔍 🎢 🚽 📄 🕞 Run 🖻 🔿 Sourc	~/Developer/GitHub/seankross.github.io/ 🖘 🔬			
<pre>&gt; swirlQ   Welcome to swirl! Please sign in. If you've been here   before, use the same name 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 before we begin our first lesson. First of all, you   should know that when you see '', that means you should   press Enter when you are done reading and ready to continue &lt; That's your cue to press Enter to continue   Also, when you see 'ANSWER:', the R prompt (&gt;), or when you   are asked to select from a list, that means it's your turn   Files Plots Packages Help Viewer   Name   Size Modified   Name</pre>	1	> library(swirl)			
I:1       (Top Level) :         I:1       (Top Level) :         I:1       (Top Level) :         RScript :       What shall I call you? sean         I Thanks, sean. Let's cover a couple of quick housekeeping i items before we begin our first lesson. First of all, you is should i press Enter when you are done reading and ready to continue.         Environment History Connections Git       < That's your cue to press Enter to continue.         < That's your cue to press Enter to continue       < That's your cue to press Enter to continue.         < That's your cue to select from a list, that means it's your turn       < That's Plots Packages Help Viewer         Files Plots Packages Help Viewer       < @         < Name       Size Modified         < Name       < @		Hi! Type swirl() when you are ready to begin.			
<pre>1:1 (Top Level): Environment History Connections Git Global Environment is empty Environment is empty </pre> <pre>     before, use the same name as you did then. If you are new,     call yourself something unique. What shall I call you? sean </pre> I Thanks, sean. Let's cover a couple of quick housekeeping items before we begin our first lesson. First of all, you ishould know that when you are done reading and ready to continue.  I also, when you are done reading and ready to continue.  I also, when you see 'ANSWER:', the R prompt (>), or when you is are asked to select from a list, that means it's your turn  Files Plots Packages Help Viewer  I Also, New Folder I Delete Rename I More -  I Also I Delete Rename I More -  I are asked to select from a list, that means it's your turn  Files Plots Packages Help Viewer  I Aname  Size Modified  I codefinch		> swirl()			
1:1       (Top Level):       R Script:         Environment History Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Connections Git       Image: Connections Git         Image: Connections Git       Image: Conn		before, use the same name as you did then. If you are new,			
1:1       (Top Level):       R Script:         Environment History Connections Git       Image: Script:         Image: Script:       Items before we begin our first lesson. First of all, you is should know that when you see '', that means you should is press Enter when you are done reading and ready to continue.         Image: Script:       Items before we begin our first lesson. First of all, you is should know that when you see '', that means you should is press Enter when you are done reading and ready to continue.         Image: Script:       Items before we begin our first lesson. First of all, you is should know that when you see '', that means you should is press Enter when you are done reading and ready to continue.         Image: Script:       Items before we begin our first lesson. First of all, you is should know that when you see '', that means you should is press Enter to continue.         Image: Script:       Items before we begin our first lesson. First of all, you is should know that when you are done reading and ready to continue.         Image: Script:       Items before we help viewer         Image: Script:       Ite		What shall I call you? sean			
Environment History   Connections Git	1:1 (Top Level) : R Script :	<pre>items before we begin our first lesson. First of all, you should know that when you see '', that means you should</pre>			
Clobal Environment •  Environment is empty  List • C New Folder © Delete Rename  More •  C New Folder © Delete More •  C Modified  C Modified C Modified  C Modif	Environment History Connections Git				
Environment is empty  Files Plots Packages Help Viewer  Files Plots Packag		<pre> &lt; That's your cue to press Enter to continue</pre>			
Environment is empty	Global Environment -				
Environment is empty		Files Plots Packages Help Viewer			
Image: Image	Environment is empty				
▲ Name Size Modified					
		🗌 🧰 codefinch			
Conegescorecard		Collegescorecard			

Kan	Help 🕑 🕙 🕴 🛜 🖵 🜒 71% 🔟 Tue 10:21 PM Sean Kross 🔍 🧔 😑			
/Developer/GitHub/seankross.github.io - master - RStudio				
🔍 🚽 🕸 🔤 😴 🕶 🔚 🔚 🔚 🧼 Co to file/function 🔢 🗧 🛨 🔛 👻 Addins	- 🔹 seankross.github.io -			
Untitled1 ×	Console Terminal × R Markdown ×			
🗢 🖘 🖉 🔚 🗌 Source on Save 🔍 🎢 🚽 📄 🕞 Run 📑 🔂 🕞 Sourc	~/Developer/GitHub/seankross.github.io/ 🔅 💰			
1	<pre>values 0.5, 55, -10, and 6. &gt; num_vect &lt;- c(0.5, 55, -10, 6)</pre>			
	That's the answer I was looking for.			
	<pre>  13%   Now, create a variable called tf that gets the result of   num_vect &lt; 1, which is read as 'num_vect is less than 1'.</pre>			
	> tf <- num_vect < 1			
	That's correct!			
1:1 (Top Level) : R Script :	Image:			
Environment History Connections Git	1: a single logical value			
🚰 🔒 📅 Import Dataset 🗸 💰 🔤 List 🗸 🎯	2: a vector of 4 logical values			
🐴 Global Environment 🗸 🔍				
Values	Selection:			
num_vect num [1:4] 0.5 55 -10 6	Files Plots Packages Help Viewer			
tf logi [1:4] TRUE FALSE TRUE FALSE	💁 New Folder 🚱 Delete 📮 Rename 🚯 More - 📿			
	Home > Developer > GitHub			
	▲ Name Size Modified			
	C codefinch			
	Collegescorecard			

# Three areas I would like to concentrate on:

- 1. How can we develop new interactive learning systems for data science?
- 2. The great courses of the world are sitting in professor's file cabinets. How can we make online course content creation easier?
- 3. How do folks do data analysis? Why do they make certain choices during an analysis? How do upstream decisions made during an analysis affect downstream results?

## Thank you!

#### Questions?

Link to these slides: <a href="mailto:seankross.com/dlab-talk-dss/">seankross.com/dlab-talk-dss/</a>

Let's talk: <a href="mailto:seankross@ucsd.edu">seankross@ucsd.edu</a>

Find me on Twitter: @seankross