

# Open Education Resources

## + Statistics Education

Mine Çetinkaya-Rundel  
Duke University + RStudio

@minebocek 

mine-cetinkaya-rundel 

mine@stat.duke.edu 



Mine Çetinkaya-Rundel  
Duke University + RStudio

- ▶ OpenIntro: Partner & overseer statistics content + co-author of textbooks + developer of labs and other supplemental materials
- ▶ MOOCs: Statistics with R on Coursera (specialization of 4 courses + capstone)
- ▶ All teaching materials available on GitHub (or by request if they pre-date my usage of GitHub for this purpose)



**Open-access:** Refers to online research outputs that are free of all restrictions on access (e.g. access tolls) and free of many restrictions on use (e.g. certain copyright and license restrictions).

**Open-source (software):** Computer software with its source code made available with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose. Open-source software may be developed in a collaborative public manner.

... in this talk we'll focus resources that are free of all  
(well, most) restrictions on access,  
and note how the resources mentioned rank  
on other axes of openness

links for all references at  
[http://bit.ly/open\\_stated](http://bit.ly/open_stated)  
(and also on the TANGO website)



Do you currently use (or plan to use)  
**open** resources in your teaching?



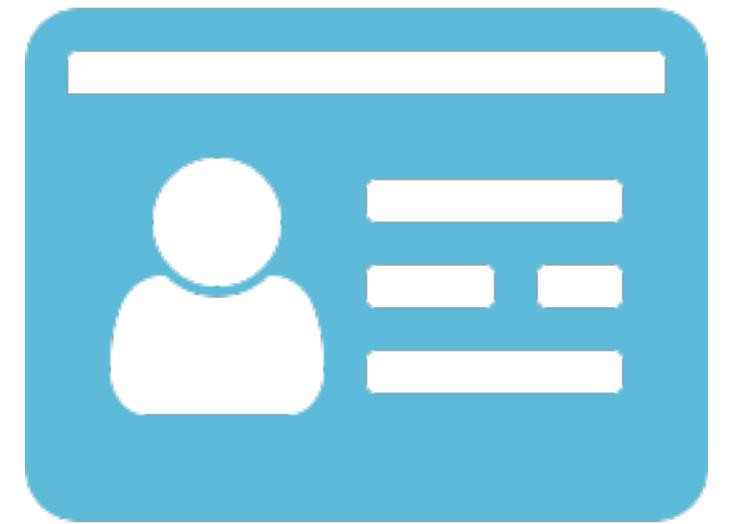
If yes, what types of **open** resources do you use (textbook, software, etc.)?



**outline**



**courses**



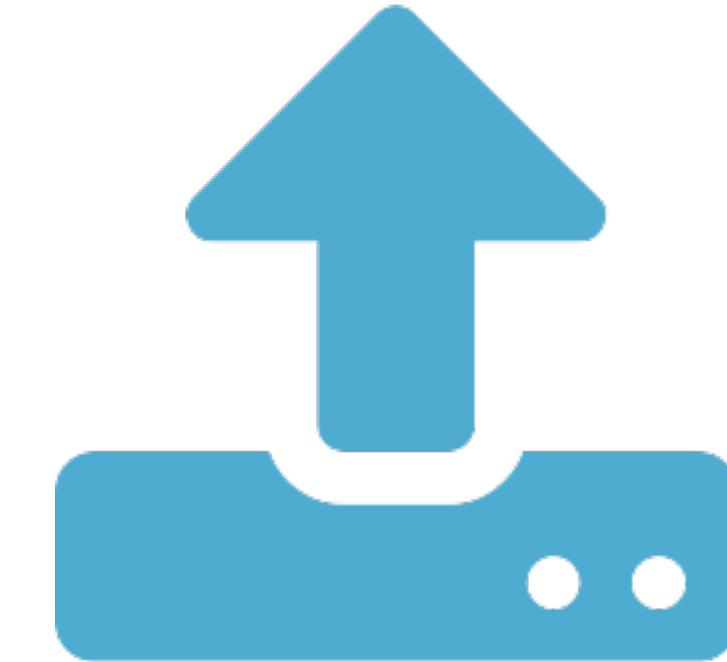
**licensing**



**OpenIntro**



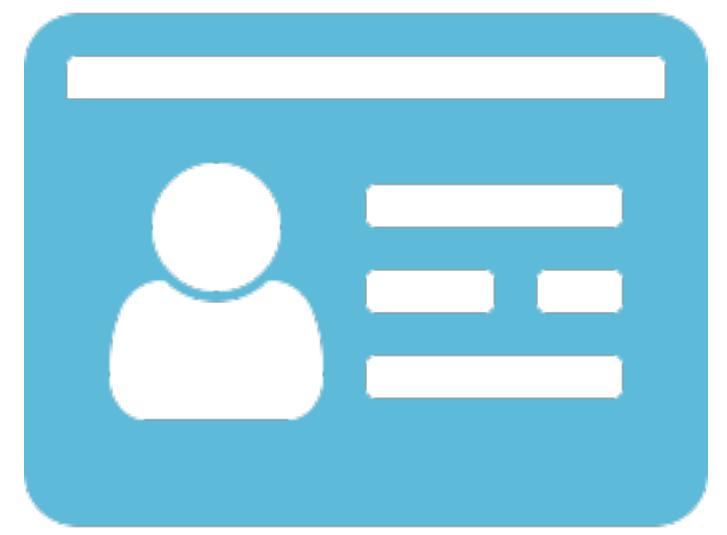
**textbooks**



**create**



**software**



licensing

# Why should I care about licensing of open resources?

- ▶ If you want to **use** an open resources, you should make sure that the resources are indeed allowed to be used in the your medium in the way you intend to use them.
- ▶ If you want to **create** open resources, you need to familiarize yourself with different types of licenses and choose the one that works for you.
- ▶ If you want to **contribute** to an existing open resource, you need to understand the stipulations of its license and make sure you are okay with it.

# Creative Commons Licenses



- ▶ CC licenses allow creators to communicate which rights they reserve, and which rights they waive for the benefit of recipients or other creators.
- ▶ An easy-to-understand one-page explanation of rights, with associated visual symbols, explains the specifics of each CC license.



## Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

### You are free to:



**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

e.g. OpenIntro PDFs  
released under  
CC BY-SA 3.0

### Under the following terms:



**Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



**ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

## GNU General Public License



- ▶ Widely used free software license, which guarantees end users the freedom to run, study, share and modify the software.
- ▶ This is a copyleft license, which means that derivative work can only be distributed under the same license terms.

## MIT License



- ▶ Permissive free software license that puts very limited restriction on reuse.
- ▶ Permits reuse within proprietary software provided that all copies of the licensed software include a copy of the MIT License terms and the copyright notice.
- ▶ MIT licensed software can be integrated into GPL software, but not the other way around.

# Considerations when choosing / using a license:

- ▶ Attribution
- ▶ ShareAlike
- ▶ Non-Commercial
- ▶ No Derivatives



**textbooks**



Do you have any hesitations about using an **open** textbook for your course? If so, what are they?



*There is a lot of free stuff online, and the quality varies.*



*Some projects get orphaned.*



*Textbook prices are a “drop in the bucket”  
and “do not matter”.*

2013 survey by US PIRG in 33 states and 156 different campuses (n = 2,039):



Due to the high cost of textbooks, 65 percent of students said they decided against buying a book required for class.



Of those students, nearly all (94 percent) said they were concerned that doing so would hurt their grade in a class.

*Portland Community College recently selected one of our free statistics textbooks, and it is estimated that their students will save \$250,000 each year!*



# Why should I consider using an open textbook?

- ▶ Cost
- ▶ Student access on day one of course as well as after the course
- ▶ Fresh content
- ▶ If you are so inclined, possibility to adapt the textbook for your particular course

# I'm interested in using an open textbook for my intro stats course. What options are out there?

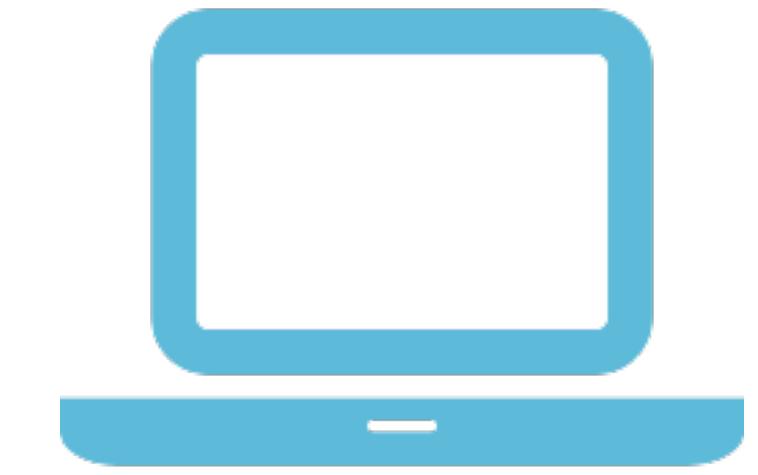
- ▶ [Think Stats \(Downey\)](#)
- ▶ [CK-12 Probability and Statistics](#)
- ▶ [Online Statistics Education \(Lane\)](#)
- ▶ [Intro to Statistical Thought \(Lavine\)](#)
- ▶ [Intro to Probability and Statistics using R \(Kerns\)](#)
- ▶ [Introduction to Statistical Thinking \(Yakir\)](#)
- ▶ [Probability and Statistics EBook \(SOCR\)](#)
- ▶ [Statistics Wikibook](#)
- ▶ OpenIntro (more on this later)

Note:  
This is not a  
comprehensive  
list.

I'm interested in using an open textbook for my intermediate / advanced stats course. What options are out there?

- ▶ Practical Regression and ANOVA using R (Faraway)
- ▶ A first course in design and analysis of experiments (Oehlert)
- ▶ An Introduction to Statistical Learning (James et al.)
- ▶ Elements of Statistical Learning (Hastie et al.)
- ▶ Introduction to Probability (Grinstead and Snell)
- ▶ The Elements of Data Analytic Style (Leek)
- ▶ Regression models for Data Science in R (Caffo)
- ▶ R for Data Science (Grolemund and Wickham)

Note:  
This is not a  
comprehensive  
list.



**software**



How do you decide which software to  
use in your course?

# Why should I consider using open software?

- ▶ Active development makes modern tools accessible
- ▶ Free access during class
- ▶ Free access beyond class / school
- ▶ As a statistician, R or Python!

# Aside from programming languages, what open tools are available for statistical exploration?

- ▶ [CalPoly Shiny Apps Collection](#)
- ▶ [Rossmann/Chance Applet Collection](#)
- ▶ [ShinyEd Shiny Apps Collection](#)
- ▶ [TinkerPlots](#)
- ▶ [StatKey](#)

**Note:**  
This is not a  
comprehensive  
list.



courses



Will massive open online courses put  
most teachers out of a job?

Very doubtful.

# What are some open online statistics courses?

- ▶ MIT OpenCourseware:
  - ▶ Statistical Thinking and Data Analysis
  - ▶ Introduction to Probability and Statistics
  - ▶ Statistics for Applications
- ▶ Coursera:
  - ▶ Statistics with R, Duke University (5 courses)
  - ▶ Data Science, Johns Hopkins University (10 courses)

Note:  
This is not a  
comprehensive  
list.

# How can I use these open courses to complement my teaching?

- ▶ Inspiration (for you)
- ▶ Supplementation (for your students)
- ▶ Flipped classroom



Want to use a resource that appears to be open, but is behind a pay wall or somewhere inconvenient for your students? Ask the creator for access or permission to rehost!

## **Another resource: Statistical Commons (for ASA members only)**

A repository for lecture notes, problem sets, exams, educational data sets, and code containing complete sets of course materials as well as other contributed class materials, e.g., projects and exams.

- ▶ Data Analysis and Statistical Inference (Çetinkaya-Rundel)
- ▶ Introduction to Biostatistics (Lock-Morgan)
- ▶ Probability and Statistical Inference (Banks)



**OpenIntro**

[www.openintro.org](http://www.openintro.org)

OpenIntro was started with one goal  
in mind: create a free and open-source  
introductory textbook.

The mission of OpenIntro is to make educational products  
that are free, transparent, and lower barriers to education.

- 2009 Officially started working on an open-source textbook
- 2010 Launch Preliminary Edition of OS
- 2011 Launch First Edition of OS
- 2012 Second Edition of OS, R labs launch
- 2013 Videos + slides start becoming available
- 2014 Coursera + two new intro stat textbooks (ISRS + AHSS)
- 2015 Third Edition of OS
- 2016 First batch of companion DataCamp courses for ISRS
- 2017 Second batch of companion DataCamp courses for ISRS + ???



David

PARTNER



Chris

PARTNER (SABBATICAL)



Mine

PARTNER



Andrew

FELLOW



Meenal

FELLOW



Leah

FELLOW



Shannon

FELLOW



Rebecca

VOLUNTEER



Edwin

VOLUNTEER



Matt

VOLUNTEER



Albert

VOLUNTEER



Curry

VOLUNTEER

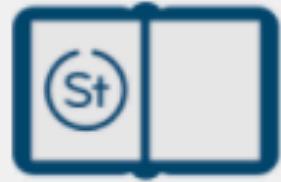


Matt

ADVISOR



- ▶ Free PDF online
- ▶ Paperback b&w print available for under \$15
- ▶ Hardcover color copy available for under \$20



## Textbook

Download the latest book and more.



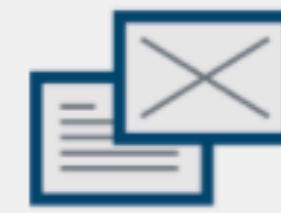
## Labs

Practice using labs for R or SAS.



## Videos

Watch and learn.



## Teachers

Making teachers' lives a little easier.



## Forums

Ask questions, find answers.



## Extras

Data, more free books.

LABS FOR R

← HOVER HERE TO SWITCH SOFTWARES X

LABS WITH SAS

SUGGEST A SOFTWARE

OpenIntro Labs promote the understanding and application of statistics through applied data analysis. The statistical software R is a widely used and stable software that is free. RStudio is a user-friendly interface for R.

[Download R](#)

[Download RStudio, after you install R \(optional\)](#)

[DataCamp - interactive versions of OpenIntro Labs](#)

[Access the markdown source files on GitHub](#)

[PDF + Google Doc Versions of Labs](#)

[Labs in Brazilian Portuguese](#)

[Provide feedback or report typos](#)

[Intro to R and RStudio](#)

[Introduction to Data](#)

[Probability](#)

[Distributions](#)

[Intro to Inference](#)

[Confidence Levels](#)

[Inf. for Numerical Data](#)

[Inf. for Categorical Data](#)

[Linear Regression](#)

[Multiple Regression](#)

# Multiple linear regression

## Grading the professor

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. The article titled, “Beauty in the classroom: instructors’ pulchritude and putative pedagogical productivity” by Hamermesh and Parker found that instructors who are viewed to be better looking receive higher instructional ratings.

In this lab we will analyze the data from this study in order to learn what goes into a positive professor evaluation.

## Getting Started

### Load packages

In this lab we will explore the data using the `dplyr` package and visualize it using the `ggplot2` package for data visualization. The data can be found in the companion package for the OpenIntro labs, `oilabs`.

Let's load the packages.

```
library(dplyr)
library(ggplot2)
library(oilabs)
library(GGally)
```



This is the first time we're using the `GGally` package. We will be using the `ggpairs` function from this package later in the lab.



Want to use a OpenIntro resources?  
Create a teacher account!

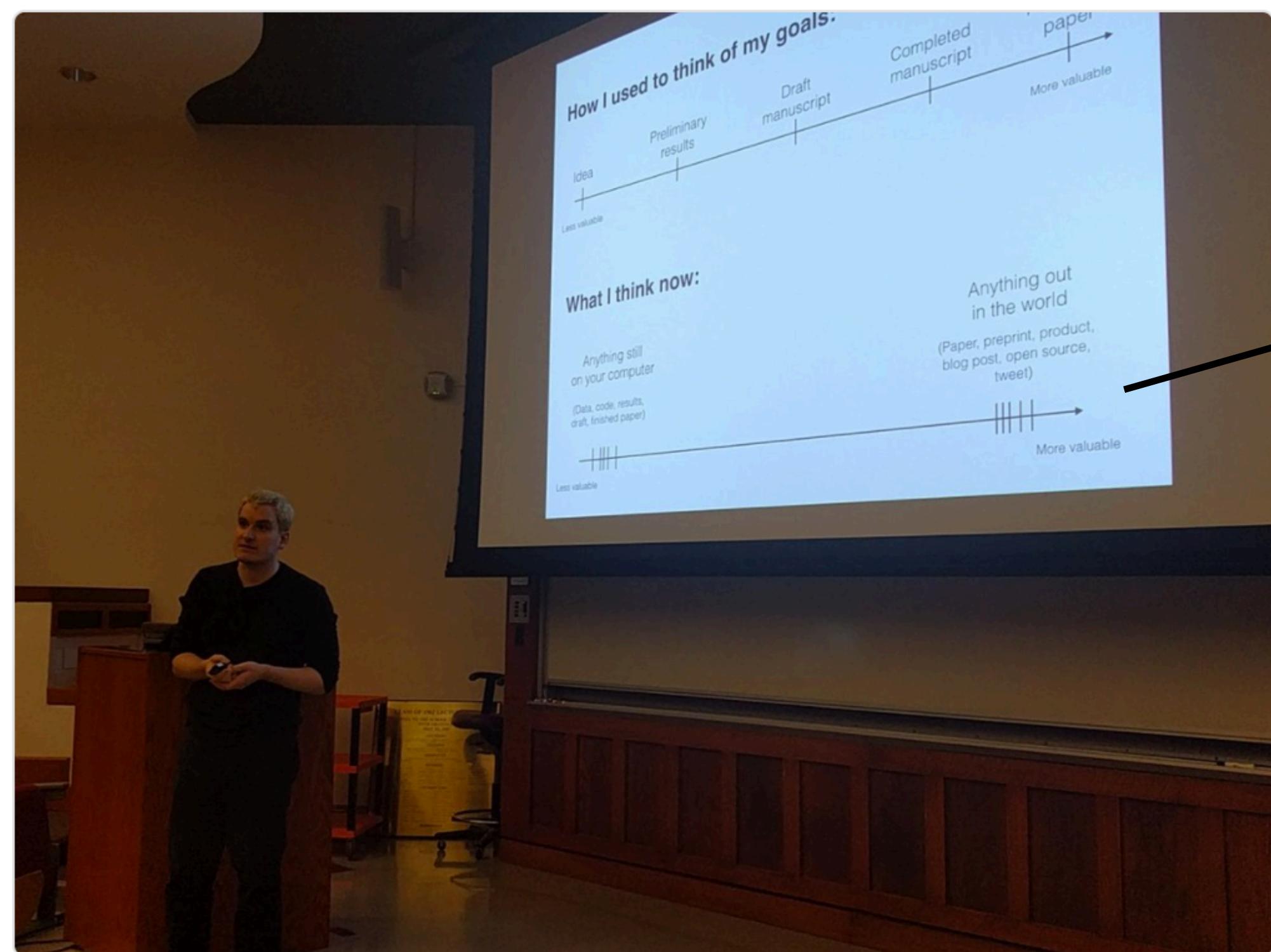


create

 David Robinson Retweeted

**Casey Greene** @GreeneScientist · May 12

@drob advice on what has value. Loving where #preprints end up on this scale!  
#ASAPbio



4

27

51

Anything still  
on your computer

(data, code, results,  
draft, finished paper)

Less valuable

Anything out  
in the world

(paper, preprint, product,  
blog post, open source, tweet)

More valuable

Source: David Robinson (@drob)

# Why are some tools that are helpful for getting my materials out there in the open?

- ▶ Course materials, slides, code, etc. – GitHub
- ▶ What about Blackboard, Sakai, Moodle, etc.?
- ▶ What about just a standard public website?

<https://github.com/mine-cetinkaya-rundel>



## Mine Cetinkaya-Rundel

mine-cetinkaya-rundel

[Add a bio](#)

 Duke University, @rstudio

 Durham, NC

 mine@stat.duke.edu

 <https://stat.duke.edu/~mc301>

Overview

Repositories 52

Stars 0

Followers 336

Following 0

sta

Type: All ▾

Language: All ▾

New

10 results for repositories matching sta

 Clear filter

### sta104\_su15

Course materials for Sta 104 - Summer 2015 semester at Duke University

 HTML  21  12 Updated on Jun 18, 2015

### sta101-s16

Course materials for Sta 101 - Spring 2016 semester at Duke University

 HTML  13  14 Updated on Jun 30, 2016

[Add topics](#)

142 commits

1 branch

Branch: master ▾

[New pull request](#) mine-cetinkaya-rundel add pretest + getting to know you info[Application Exercises](#)

typo in bayes rule

6

[Discussion Sections](#)

lab updates

a year ago

[Labs](#)

update link to dl Rmd file + number of questions on template

a year ago

[Logistics](#)

add pretest + getting to know you info

11 months ago

[Project](#)

add eval forms

a year ago

[Review](#)

add review materials for exams

a year ago

[Slides](#)

.

a year ago

[Sta101\\_S16\\_web](#)

incorrect answer noted for clicker question

a year ago

[.gitignore](#)

add rt\_key to gitignore

a year ago

[LICENSE.md](#)

add readme and license

a year ago

[README.md](#)

update readme title

a year ago

<a href="#">Application Exercises</a>	►	<a href="#">Absences</a>
<a href="#">Discussion Sections</a>	►	<a href="#">Accomodations</a>
<a href="#">Labs</a>	►	<a href="#">Assessments</a>
<a href="#">LICENSE.md</a>		<a href="#">Grades</a>
<a href="#">Logistics</a>	►	<a href="#">Keys</a>
<a href="#">Private</a>	►	<a href="#">Midterm eval</a>
<a href="#">Project</a>	►	<a href="#">Misc</a>
<a href="#">README.md</a>		<a href="#">Photo rosters</a>
<a href="#">Review</a>	►	<a href="#">Teams</a>
<a href="#">Slides</a>	►	
<a href="#">Sta101_S16_web</a>	►	

# Why are some tools that are helpful for getting my materials out there in the open?

- ▶ Course materials, slides, code, etc. – GitHub
  - ▶ What about Blackboard, Sakai, Moodle, etc.?
  - ▶ What about just a standard public website?
- ▶ Textbook – blogdown
  - ▶ What about LaTeX?
  - ▶ What about Word or some other word processor?



Welcome

1 Introduction

I Explore

2 Introduction

3 Data visualisation

3.1 Introduction

3.2 First steps

3.3 Aesthetic mappings

3.4 Common problems

3.5 Facets

3.6 Geometric objects

3.7 Statistical transformations

3.8 Position adjustments

3.9 Coordinate systems

3.10 The layered grammar of graphics

## 3 Data visualisation

### 3.1 Introduction

“The simple graph has brought more information to the data analyst’s mind than any other device.”

— John Tukey

This chapter will teach you how to visualise your data using ggplot2. R has several systems for making graphs, but ggplot2 is one of the most elegant and most versatile. ggplot2 implements the **grammar of graphics**, a coherent system for describing and building graphs. With ggplot2, you can do more faster by learning one system and applying it in many places.

If you’d like to learn more about the theoretical underpinnings of ggplot2 before you start, I’d recommend reading “The Layered Grammar of Graphics”, <http://vita.had.co.nz/papers/layers-grammar.pdf>.

#### 3.1.1 Prerequisites

# Why are some tools that are helpful for getting my materials out there in the open?

- ▶ Course materials, slides, code, etc. – GitHub
  - ▶ What about Blackboard, Sakai, Moodle, etc.?
  - ▶ What about just a standard public website?
- ▶ Textbook – blogdown
  - ▶ What about LaTeX?
  - ▶ What about Word or some other word processor?
- ▶ Computing labs – rmarkdown
  - ▶ What about LaTeX?
  - ▶ What about Word or some other word processor?

RStudio

multiple\_regression.Rmd

```
## Simple linear regression
The fundamental phenomenon suggested by the study is that better looking teachers are evaluated more favorably. Let's create a scatterplot to see if this appears to be the case:
```{r scatter-score-bty_avg}
qplot(data = evals, x = bty_avg, y = score)
```
Before we draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?
4. Replot the scatterplot, but this time use `geom = "jitter"`. What was misleading about the initial scatterplot?
```{r scatter-score-bty_avg-jitter}
qplot(data = evals, x = bty_avg, y = score, geom = "jitter")
```
5. Let's see if the apparent trend in the plot is something more than
1:1 # Multiple linear regression
```

Environment History Presentation

Import Dataset Global Environment

Environment is empty

Console Terminal R Markdown

The downloaded binary packages are in  
/var/folders/p7/r2fscg6d6dqg1xbncwf\_51yc0000gq/T//RtmpKz2qbW/downloaded\_packages

## Simple linear regression

The fundamental phenomenon suggested by the study is that better looking teachers are evaluated more favorably. Let's create a scatterplot to see if this appears to be the case:

```
qplot(data = evals, x = bty_avg, y = score)
```

Before we draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

**Exercise 4**

Replot the scatterplot, but this time use `geom = "jitter"`. What was misleading about the initial scatterplot?

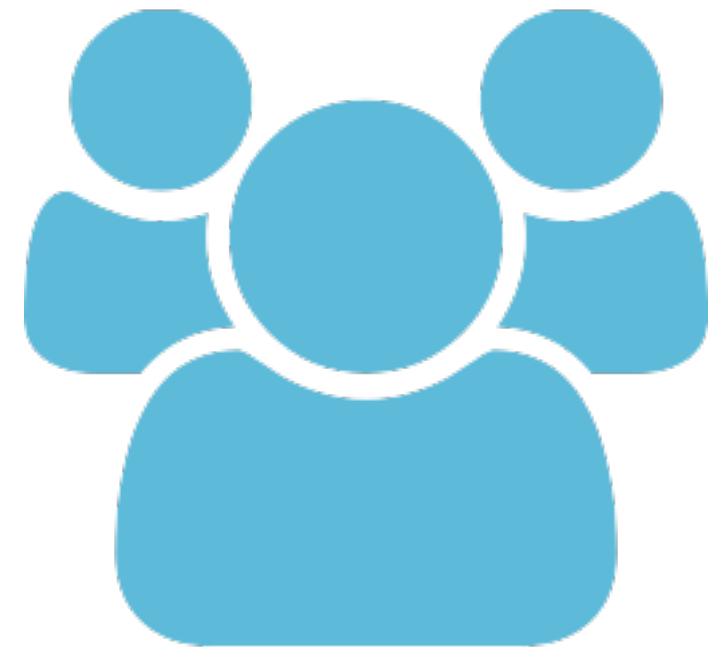
```
qplot(data = evals, x = bty_avg, y = score, geom = "jitter")
```

**Exercise 5**

Let's see if the apparent trend in the plot is something more than natural variation. Fit a linear model called `m_bty` to predict average professor score by average beauty rating. Write out the equation for the linear model and interpret the slope. Is average beauty score a statistically significant predictor? Does it appear to be a practically significant predictor?



Have you used other tools and resources that you have found useful for creating **open** education resources? If so, what are they?



**discussion**



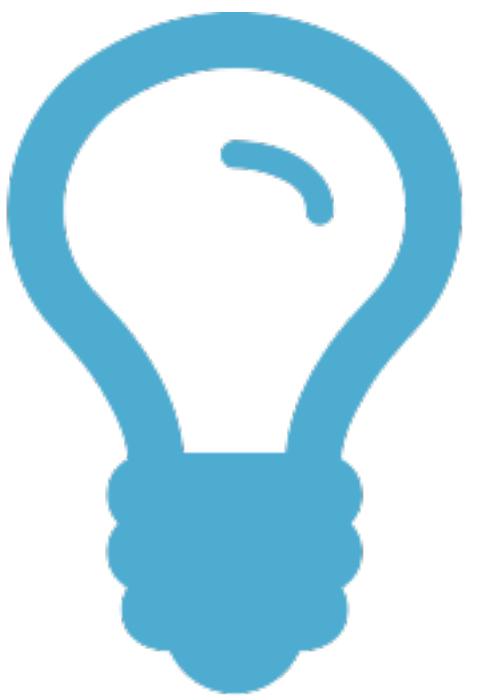
Do you currently work on / contribute to or, in the near future, do you see yourself working on / contributing to any **open** education resources? If so, what are they?



Do you have any hesitations about working on / contributing to **open** education resources? If so, what are they?



Other questions?



parting thoughts

- ▶ When teaching and choosing a textbook, ask...  
Would I think this is a fair price if I was a student?  
... and consider free options
- ▶ When trying out free resources....  
Provide constructive and critical feedback
- ▶ When developing classroom materials....  
Take an extra 25% of development time to put the resources into good shape for sharing

# Thank You +

@minebocek 

mine-cetinkaya-rundel 

mine@stat.duke.edu 