

# Course wrap-up

Jeffrey Leek, Assistant Professor of Biostatistics  
Johns Hopkins Bloomberg School of Public Health

# Why we do applied statistics

"It is not the critic who counts: not the man who points out how the strong man stumbles or where the doer of deeds could have done better. The credit belongs to the man who is actually in the arena, whose face is marred by dust and sweat and blood, who strives valiantly, who errs and comes up short again and again, because there is no effort without error or shortcoming, but who knows the great enthusiasms, the great devotions, who spends himself for a worthy cause; who, at the best, knows, in the end, the triumph of high achievement, and who, at the worst, if he fails, at least he fails while daring greatly, so that his place shall never be with those cold and timid souls who knew neither victory nor defeat."



*Theodore Roosevelt, 26th President of the United States*

[Statistics and the science game](#)

# The key challenge in applied statistics

Ask yourselves, what problem have you solved, ever, that was worth solving, where you knew all of the given information in advance? Where you didn't have a surplus of information and have to filter it out, or you didn't have insufficient information and have to go find some?



[Dan Myer, Mathematics Educator](#)

# Why applied statistics?



# Why applied statistics?


## The New York Times

### For Today's Graduate, Just One Word: Statistics


By **STEVE LOHR**


Published: August 5, 2009

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

 TWITTER

 LINKEDIN

 COMMENTS  
(58)

 SIGN IN TO E-MAIL

# Why applied statistics?

McKinsey Global Institute



June 2011

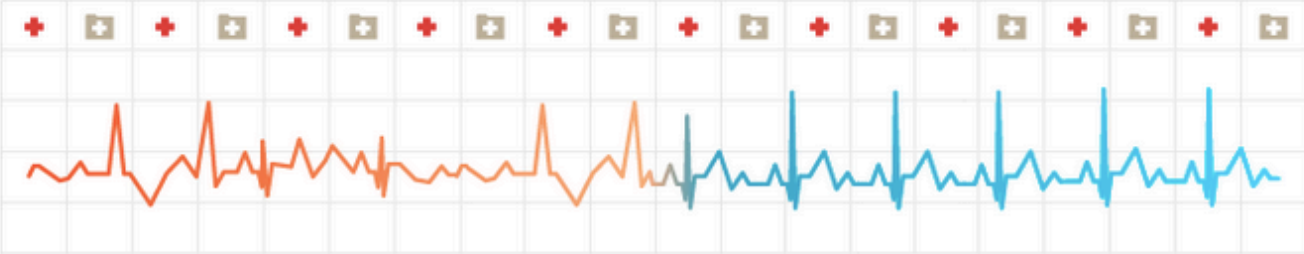
Big data: The next frontier  
for innovation, competition,  
and productivity

# Why are you lucky?



# Why are you lucky

Competition Details » [Get the Data](#) » [Make a submission](#)



## Improve Healthcare, Win \$3,000,000.

**Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)**

Please note: Deadline is 06:59:59 UTC on October 4, 2012 for new registrations and team mergers.

- **Description**
- [Evaluation](#)
- [Rules](#)
- [Dos and Don'ts](#)
- [FAQ](#)
- [Milestone Winners](#)
- [Timeline](#)

## [Heritage Health Prize](#)



# New data drives new statistical ideas

- How do we make better beer?
  - **Data:** Measures of beer quality
  - **Statistic:** The [t-statistic](#)
- What characteristics of field lead to better crops?
  - **Data:** Field characteristics, crop yields
  - **Statistic:** [Analysis of variance \(ANOVA\)](#)
- How long do people live?
  - **Data:** Survival times of people (censored)
  - **Statistic:** [Kaplan-Meier Estimator](#)
- What movies will you like?
  - **Data:** Lots of other peoples movie ratings
  - **Statistic(s):** [Recommender systems](#)

# Who is an applied statistician?

Daryl Morey



Hilary Mason



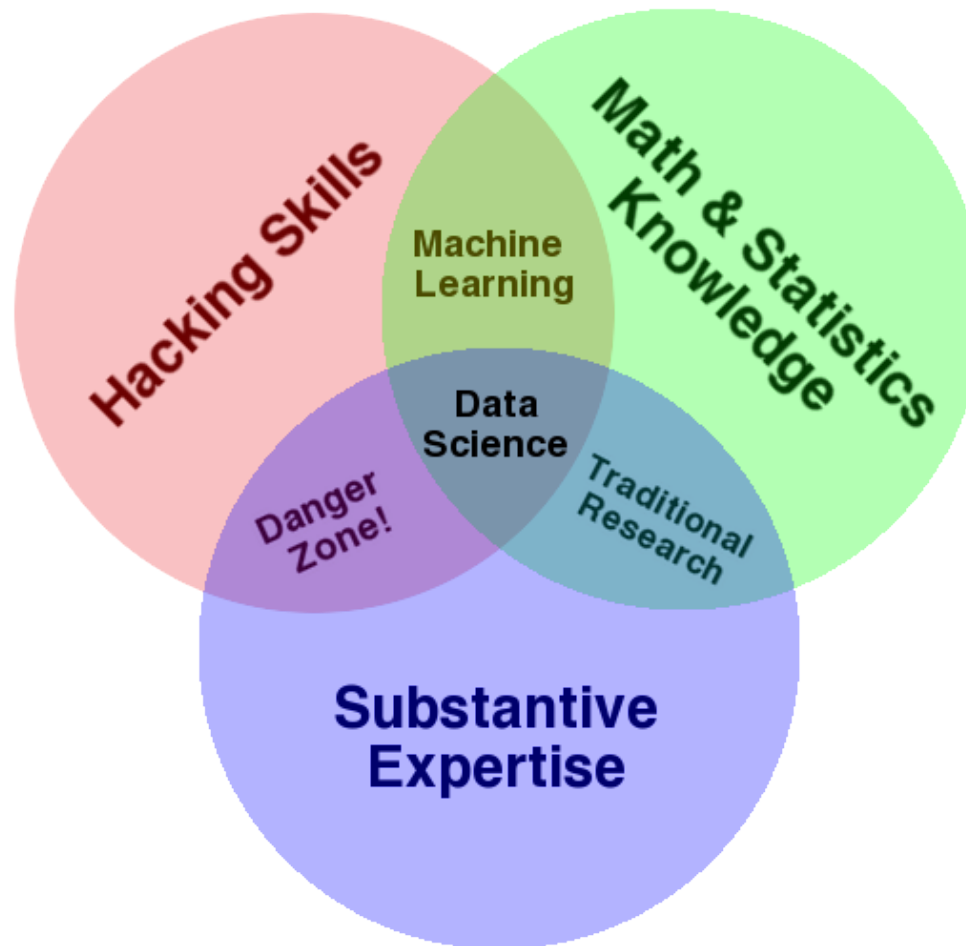
Daphne Koller



Nate Silver



# An important goal



[Drew Conway](#)

11/16

# These might be useful

The screenshot shows the OpenIntro website's Statistics page. The browser address bar displays [www.openintro.org/stat/textbook.php](http://www.openintro.org/stat/textbook.php). The page features a dark blue header with the OpenIntro logo and navigation links: SUBJECTS, CONTRIBUTE, RIGHTS, BLOG, and LOGIN. Below this is a sub-header for Statistics with tabs for OVERVIEW, TEXTBOOK (selected), SUPPLEMENTS, LABS, and LINKS. The main content area is divided into three columns. The left column, titled 'SECOND EDITION', describes the free textbook and provides a 'Download Second Edition' link. The middle column, titled 'FIRST EDITION', provides information about the differences between editions and includes a table of contents. The right column, titled 'DATA SETS', describes the data sets included and provides a 'Data Sets' link. Below the 'DATA SETS' section is a 'PROBABILITY TABLES' section with links for a 'Printer-friendly PDF' and 'Download the LaTeX source'.

**OpenIntro**

SUBJECTS CONTRIBUTE RIGHTS BLOG LOGIN

**Statistics**

OVERVIEW **TEXTBOOK** SUPPLEMENTS LABS LINKS

**SECOND EDITION**

*OpenIntro Statistics* is a free textbook for introductory statistics. We've spent thousands of hours to make this textbook ready to compete on any stage. The book can be downloaded for free as a PDF or purchased on Amazon.com for \$9.94 (get 2-day shipping with a free student trial of Amazon Prime).

Download Second Edition

Click links to download individual chapters, appendices, or the textbook's source files.

FRONT MATERIAL OF BOOK

REVIEW OR SUBMIT TYPOS

CHAPTER 1	CHAPTER 6
CHAPTER 2	CHAPTER 7
CHAPTER 3	CHAPTER 8
CHAPTER 4	APPENDICES
CHAPTER 5	SOURCE

READ ONLINE REVIEWS

**FIRST EDITION**

For information about the differences between the First and Second Edition, please see [this blog post](#). The First Edition can be downloaded for free as a PDF or purchased on Amazon.com for \$9.02.

Click links to download individual chapters, appendices, or the textbook's source files.

FIRST EDITION, FULL TEXT

FRONT MATERIAL OF BOOK

REVIEW OR SUBMIT TYPOS

CHAPTER 1	CHAPTER 6
CHAPTER 2	CHAPTER 7
CHAPTER 3	CHAPTER 8
CHAPTER 4	APPENDICES
CHAPTER 5	SOURCE

READ ONLINE REVIEWS

**DATA SETS**

Data sets used in the textbook are included in the zipped file below. Each data set is saved as a tab-delimited text file so it can be easily loaded into any statistical software. These data sets can also be found in the OpenIntro R packages ([openintro](#), [Oldata](#)).

Data Sets

**PROBABILITY TABLES**

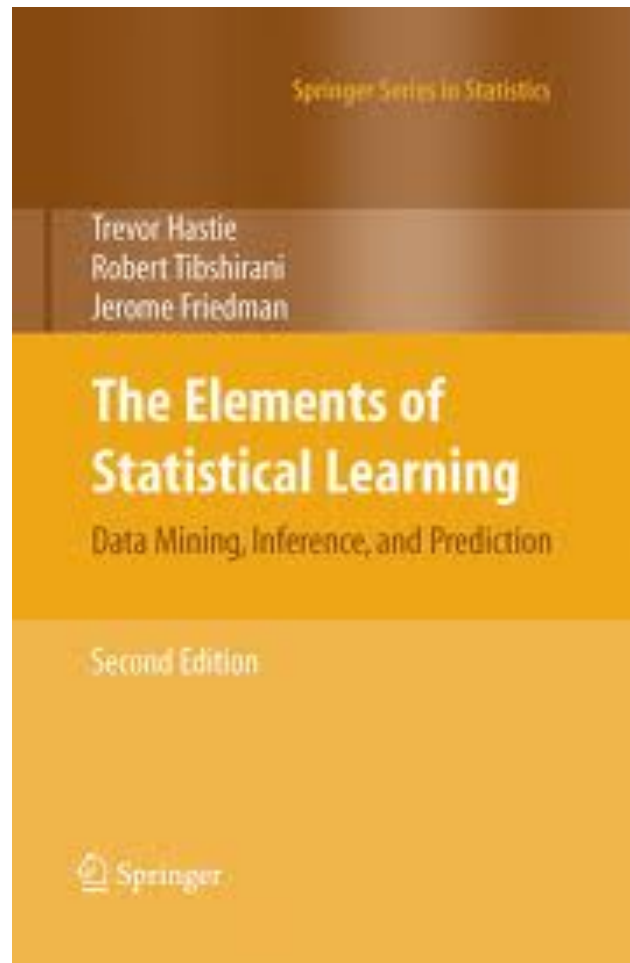
Normal, t, and chi-square probability distribution tables in a printer-friendly PDF. Or if you'd rather customize the tables, download the LaTeX source.

Printer-friendly PDF

Download the LaTeX source

<http://www.openintro.org/>

# These might be useful



<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

# These might be useful

## Advanced Data Analysis from an Elementary Point of View

Cosma Rohilla Shalizi

Spring 2013  
Last L<sup>A</sup>T<sub>E</sub>X'd Friday 18<sup>th</sup> January, 2013

[Advanced Data Analysis from An Elementary Point of View](#)

14/16

# Also check out

- [Andrew Gelman's blog](#)
- [Larry Wasserman's blog](#)
- [Statsblogs](#)
- [Flowing Data](#)
- [junkcharts](#)
- [Hilary Mason's Blog](#) and [@hmason](#)
- [Cosma Shalizi's Blog](#)
- [Some other guys' blog](#)

[The top Biostatistics Department in the World](#) - No bias here :-)

# It has been my exteme pleasure

Thank you so much for all of your dedication, time, and enthusiasm. This has been a wonderful experience for me and I hope it has been for you too.