Motivation and pre-requisites

Jeffrey Leek
Johns Hopkins Bloomberg School of Public Health

About this course

- This course covers the basic ideas behind machine learning/prediction
  - Study design - training vs. test sets
  - Conceptual issues - out of sample error, ROC curves
  - Practical implementation - the caret package
- What this course depends on
  - The Data Scientist's Toolbox
  - R Programming
- What would be useful
  - Exploratory analysis
  - Reporting Data and Reproducible Research
  - Regression models

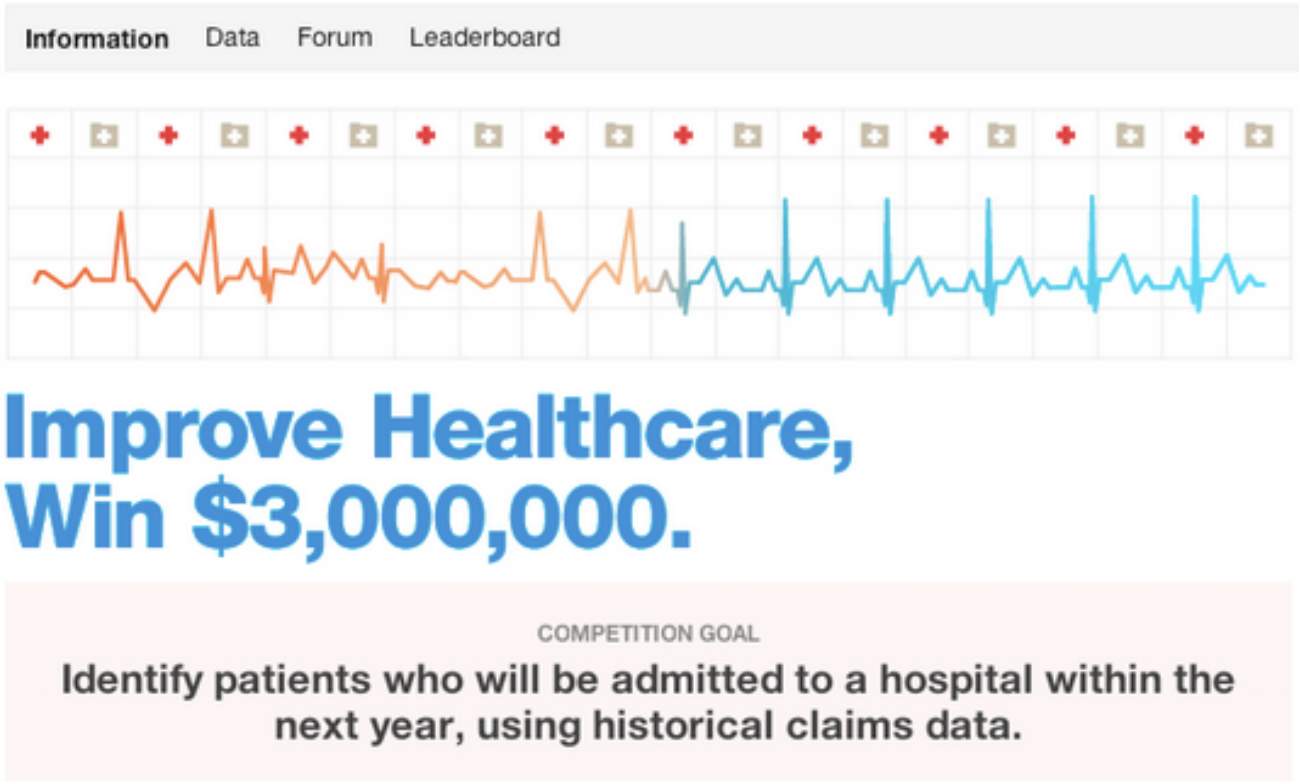
Who predicts?

- Local governments -> pension payments
- Google -> whether you will click on an ad
- Amazon -> what movies you will watch
- Insurance companies -> what your risk of death is
- Johns Hopkins -> who will succeed in their programs

Why predict? Glory!

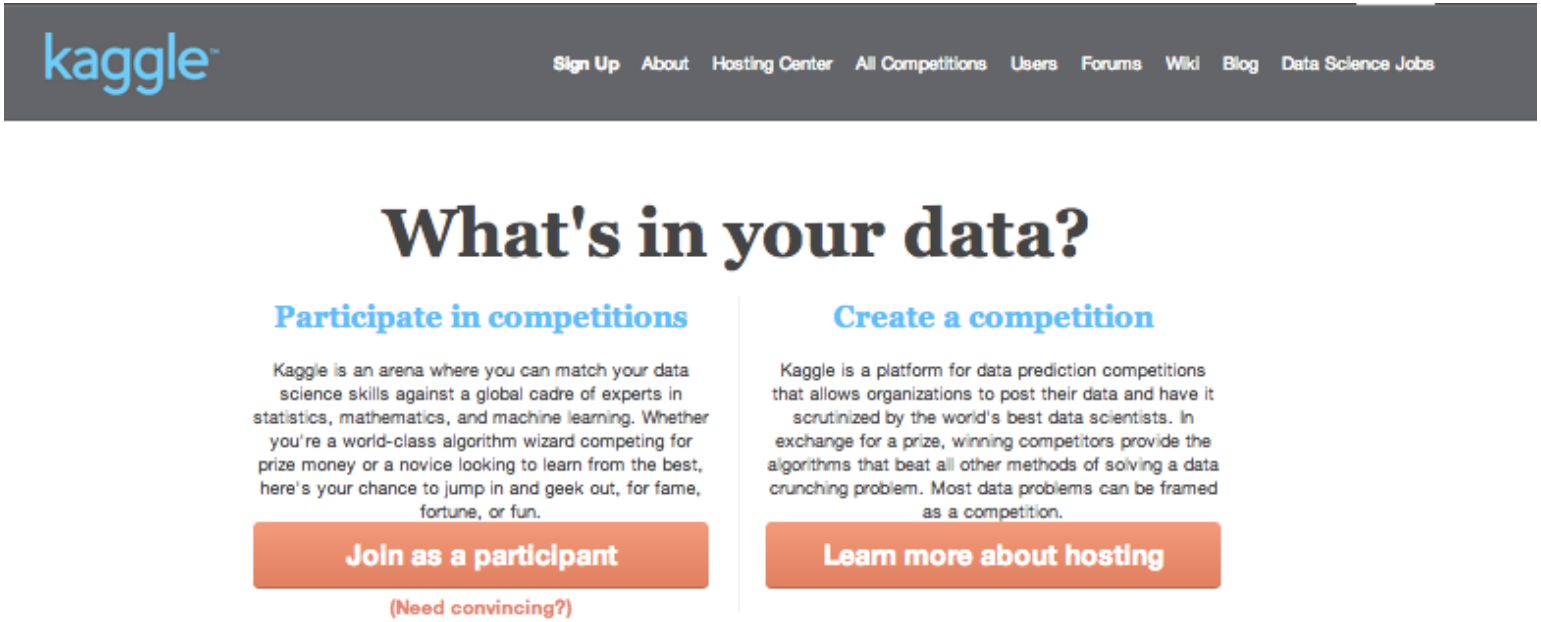
http://www.zimbio.com/photos/Chris+Volinsky

Why predict? Riches!



http://www.heritagehealthprize.com/c/hhp

Why predict? For sport!

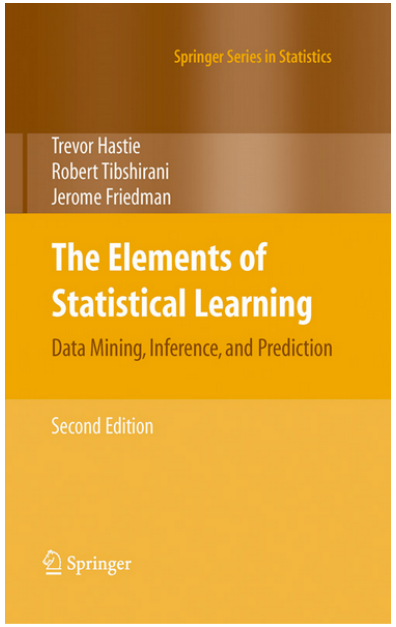


http://www.kaggle.com/

Why predict? To save lives!

http://www.oncotypedx.com/en-US/Home

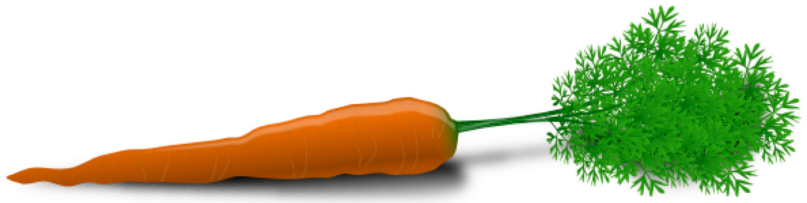
A useful (if a bit advanced) book



The elements of statistical learning

A useful package

the caret package



The **caret** package (short for *C*lassification *A*nd *RE*gression *T*raining) is a set of functions that attempt to streamline the process for creating predictive models. The package contains tools for:

Links

`train` Model List

Topics

- Main Page
- Data Sets
- Visualizations
- Pre-Processing

http://caret.r-forge.r-project.org/

Machine learning (more advanced material)

Stanford

Machine Learning

Andrew Ng

Taught In: English
Subtitles Available In: English

Sessions:
Oct 14th 2013 (10 weeks long)
Learn for Free

3,794 Tweet
12k +1
14k Like

https://www.coursera.org/course/ml

Even more resources

- List of machine learning resources on Quora
- List of machine learning resources from Science
- Advanced notes from MIT open courseware
- Advanced notes from CMU
- Kaggle - machine learning competitions