

# How to Process, Visualize, and Analyze Data

Eugene Wu, Adam Marcus

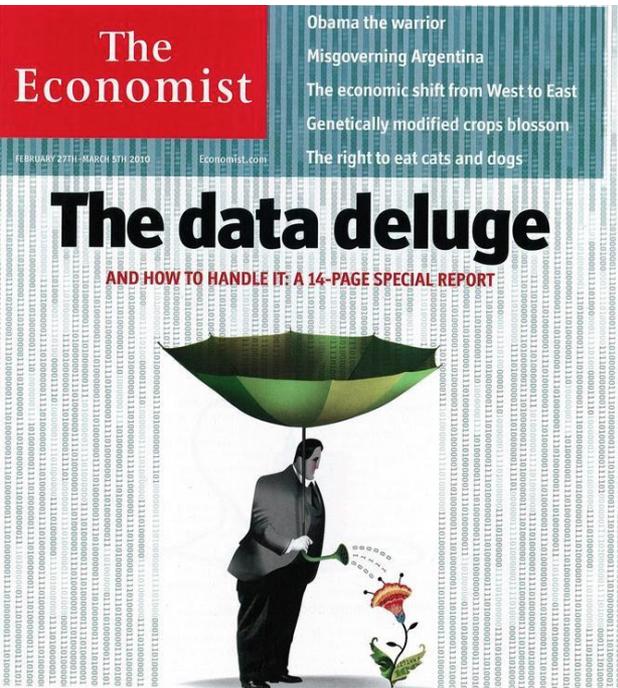
# Course Basics

- 5 lab days: ~30min background, 2.5h lab
- 1 presentation madness day
- No grades
- No homework, unless you don't finish lab

# Day 6 Madness

- After lab 4, you will have all the skills you need
- Find your own dataset + questions
- Tell us a story: 2 slides, 1 minute

Why is this important?



© The Economist, O'Reilly Media, and Nature Publishing Group. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/fairuse>.

“I keep saying that the sexy job in the next 10 years will be statisticians”

Hal Varian, Chief Economist Google

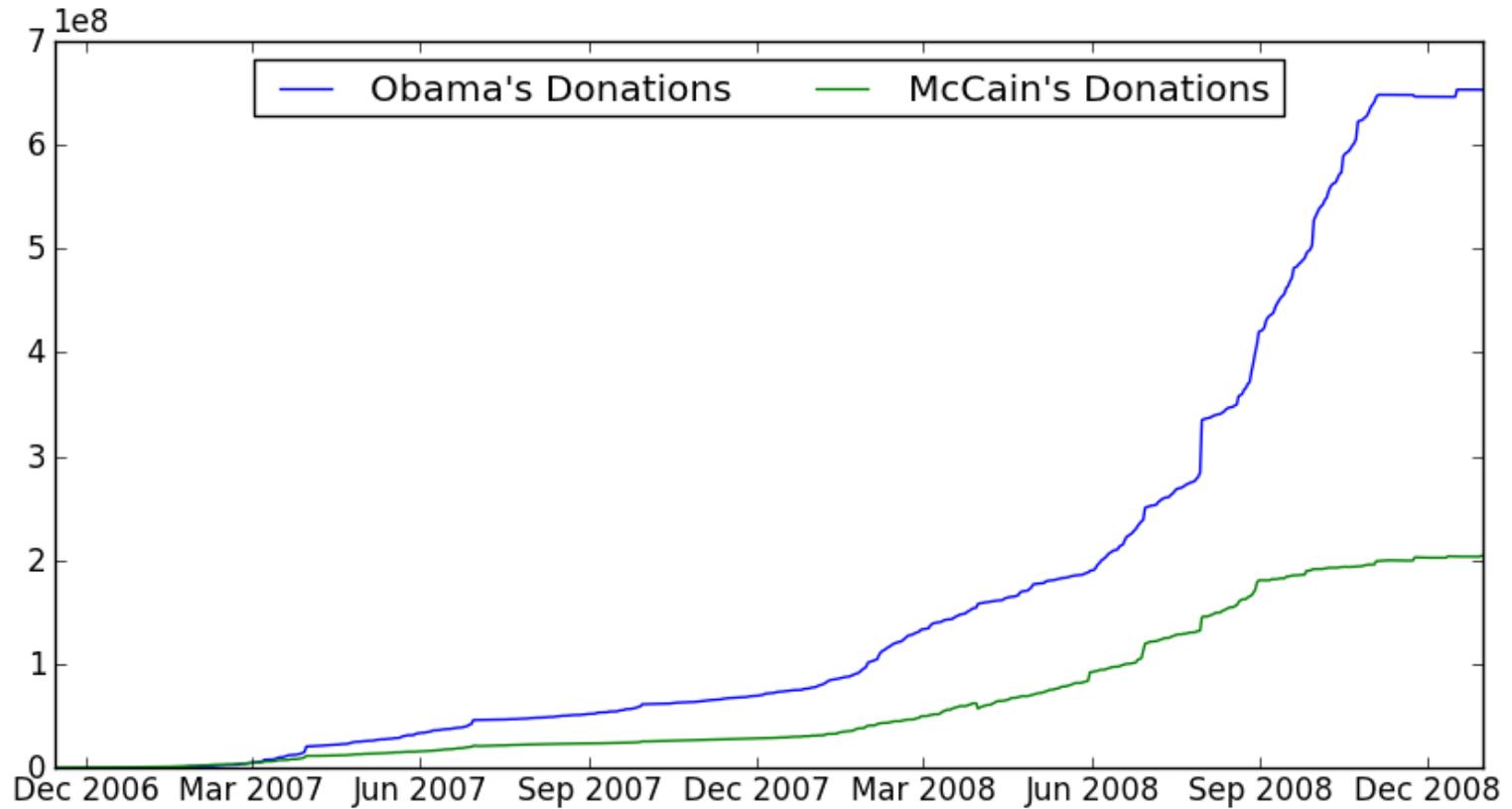
Statisticians will never have a sexy job.  
Data-powered storytellers will.

# Schedule

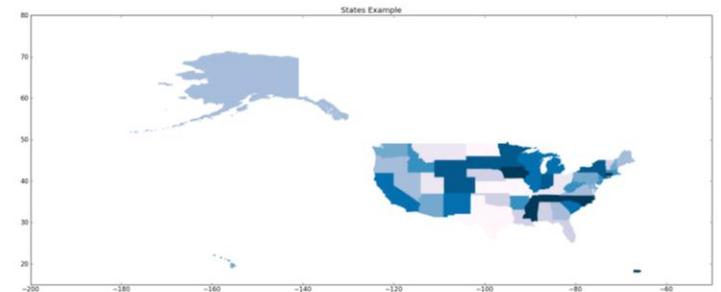
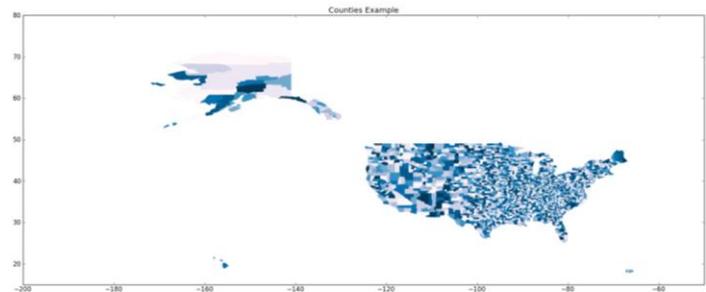
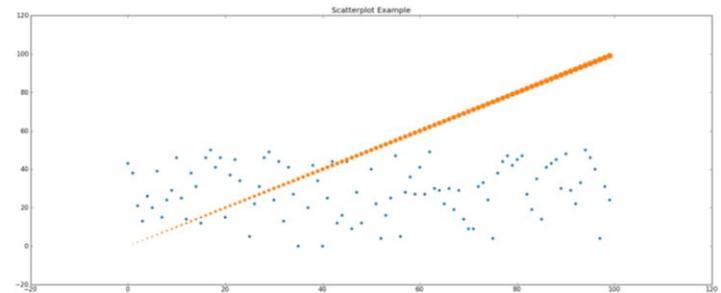
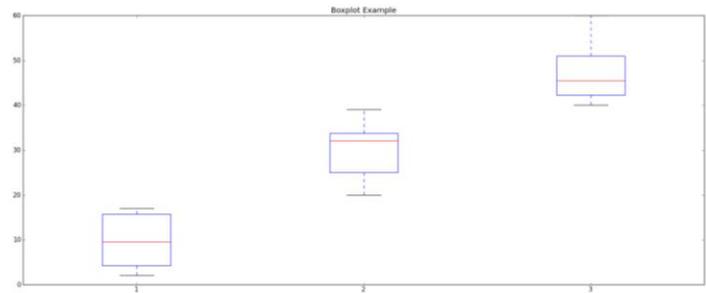
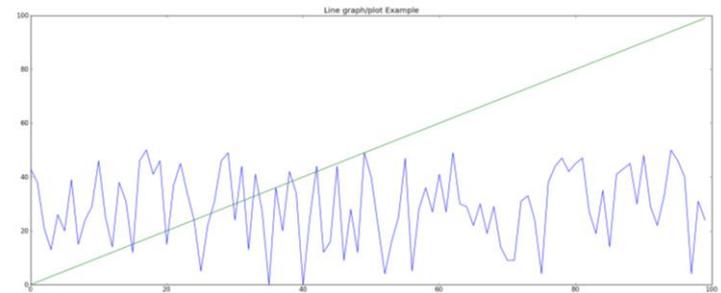
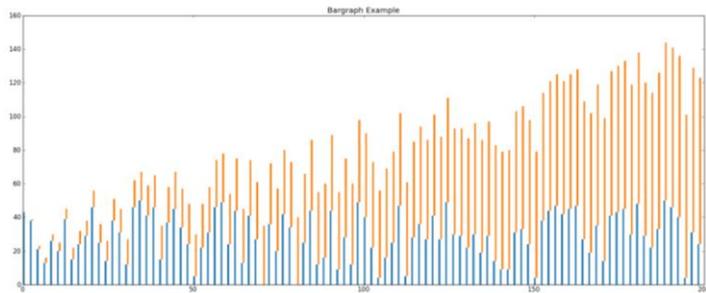
# Day 0

- Setup
- Optimistically, you've already done this

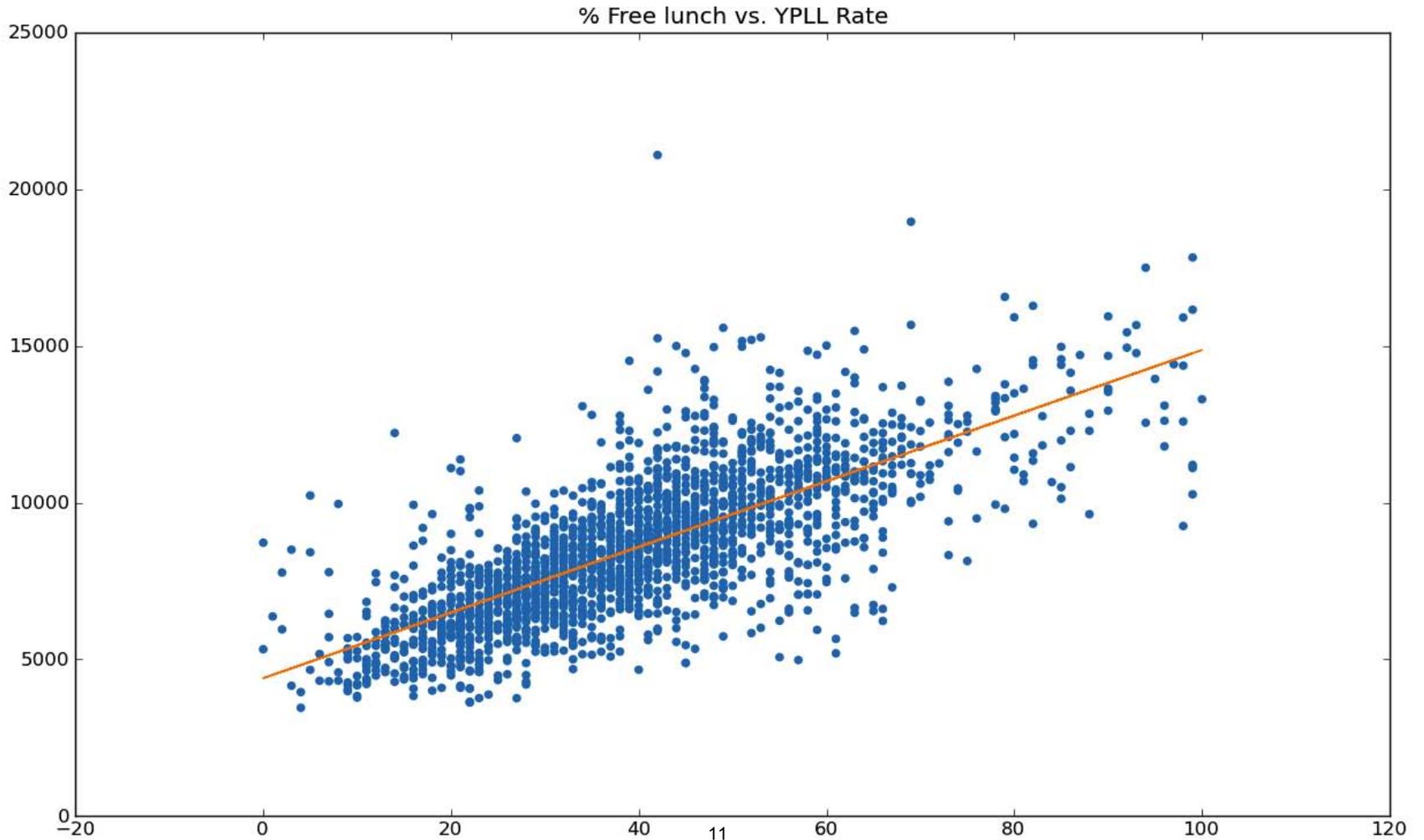
# Today



# Day 2: Visualizations



# Day 3: Statistics



# Day 4: Text Analysis with Kenneth Lay

[Images of Kenneth Lay](#) removed due to copyright restrictions.

## Day 5: Scaling up with Hadoop/MapReduce

# Day 6: Storytelling Madness!

biology

sociology

economics

stackoverflow

web scraping

source code

healthcare

finance

MIT OpenCourseWare  
<http://ocw.mit.edu>

Resource: How to Process, Analyze and Visualize Data  
Adam Marcus and Eugene Wu

The following may not correspond to a particular course on MIT OpenCourseWare, but has been provided by the author as an individual learning resource.

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.