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?
“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
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

stackoverflow

web scraping

sociology

source code

economics

finance

healthcare