5 of 93

Big Data in Context

- Evolution of technology
- Corporate examples
- McKinsey Big Data report (2011)

15 of 93

Historical Perspective

- Data gathering and storage expensive
- Limited database flexibility
- Analytical tools less flexible, required more coding skill
- Organizations used data when need was acute and resources were available
    - Big banks, logistics, government
    - Labeled IT, data warehousing, business intelligence/analytics
    - *Business analysts* typically not highly technical
    - *Scientists, developers, mathematicians* employed in hard-core use cases

27 of 93

Classic Examples: Finance

- D.E. Shaw, Renaissance Technologies
    - Hedge funds
    - Founded in 1980s
    - Among first to use quantitative methods and automated trading
    - Used mathematical and statistical analysis
    - Hired primarily mathematicians and scientists
- Financial industry at forefront of data science use and application

33 of 93

Classic Examples: Walmart

- Pioneered gathering data about customers
  - Point-of-sale interactions
  - Barcode scanners
  - Supply chain logistics
- Tracked data to improve efficiency and lower costs

41 of 93

Classic Examples: CompStat

- New York City system for tracking and intervening in crime
- Started as low-tech "Charts of the Future" by Transit Police
- Tracked locations of subway crimes
- Adopted by NYPD
- Created dashboards and reports for real-time crime tracking
- Effective, but controversial

44 of 93

Classic Examples: Amazon.com

Recommender systems

53 of 93

Contemporary Perspective

- Overarching trend: more data and more flexibility
- Changes in data: amount, variety, tracking
  - Purchases, websites, location
- Changes in storage: inexpensive, more flexible and accessible
  - Programming languages such as Python
  - Visualization and library tools
- Labeled "big data" and "data science"

58 of 93

Business Technology Trends

- IT moving from back-office to front lines
  - Integration with marketing and sales via the cloud and SaaS
- Big data more visible in organizations, up to C-level and board of directors
- Tech and data fluency increasing

63 of 93

Recent Examples: Luminar

- Gather and correlate detailed consumer profiles
- Improve efficiency or offer new products/services
- Use variety of data: social media, purchasing, TV watching

70 of 93

Recent Examples: LinkedIn

- "People you may know"
- Data science team charged with developing new features
- Similar to Amazon's recommendations but more sophisticated use of data
- Split: Product team (engineering) and decision sciences team (CFO)
- Questions surrounding how to organize data science teams

75 of 93

Recent Examples: *New York Times*

- Data science and visualization used to engage public
- Upshot blog
    - Statistics and visualization tell stories on developing topics
- Showcase for breadth of uses for data science

87 of 93

Challenges in Using Data Science

- Which data?
- How much data?
- How to use data
    - "If you build it, they will come" approach generally ineffective
- Standardization/interoperability
- Opening up
- Information security

Challenges in Using Data Science (cont.)

- Make or buy?
- Talent
  - Shortage of skilled data scientists
- Organization
  - Who owns data science?
  - Where does it sit?