

VISUAL ANALYTICS INTRODUCTION

LECTURE 1

Petra Isenberg

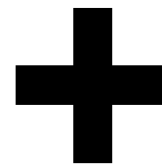
INSTRUCTOR

DR. PETRA ISENBERG

petra.Isenberg@inria.fr

OFFICE — Digiteo Moulon Building

OFFICE HOURS — By appointment





YOU!

QUICK INTROS

Any particular interests?

Special Report | Data, data everywhere

Information has gone from scarce to superabundant. That brings huge new benefits, says Kenneth Cukier (interviewed here)—but also big headaches

SLOAN DIGITAL SKY SURVEY

started in 2000

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

in first weeks, collected more data than entire history of astronomy before

WALMART

WAL★MART

1 million customer transactions per hour

100

likely has information on >145 million Americans

[1]

...AND MORE

- YouTube users upload 300 hours of new video every minute of the day <http://expandedramblings.com/index.php/youtube-statistics/>
- Facebook has currently on average 1.04 billion active users daily <http://newsroom.fb.com/company-info/>
- the Library of Congress adds 12,000 items to their collection every day <http://www.loc.gov/about/fascinating-facts/>

WHAT IS USEFUL?

data \neq useful information

you want insight

→ analysis is needed

ANALYSIS IS NOT SIMPLE



Gary King, Harvard

research project: predict U.S. unemployment rate

method: Twitter & social media analysis
→ sentiment analysis by word count

unemployment
jobs
classifieds



Look for counts of those words & correlate to monthly unemployment rate

ANALYSIS IS NOT SIMPLE



spike in people looking for jobs?

lots of people going to get laid off?

HUMAN-IN-THE LOOP

- it is sometimes dangerous to rely on purely automated analyses
- human judgment and intervention often needed
 - for: background information, flexible analysis (unintended directions), creativity
 - because: data can be incomplete, inconsistent, or deceptive

COURSE OBJECTIVES

learn about data, its properties, and its problems

practical data analysis



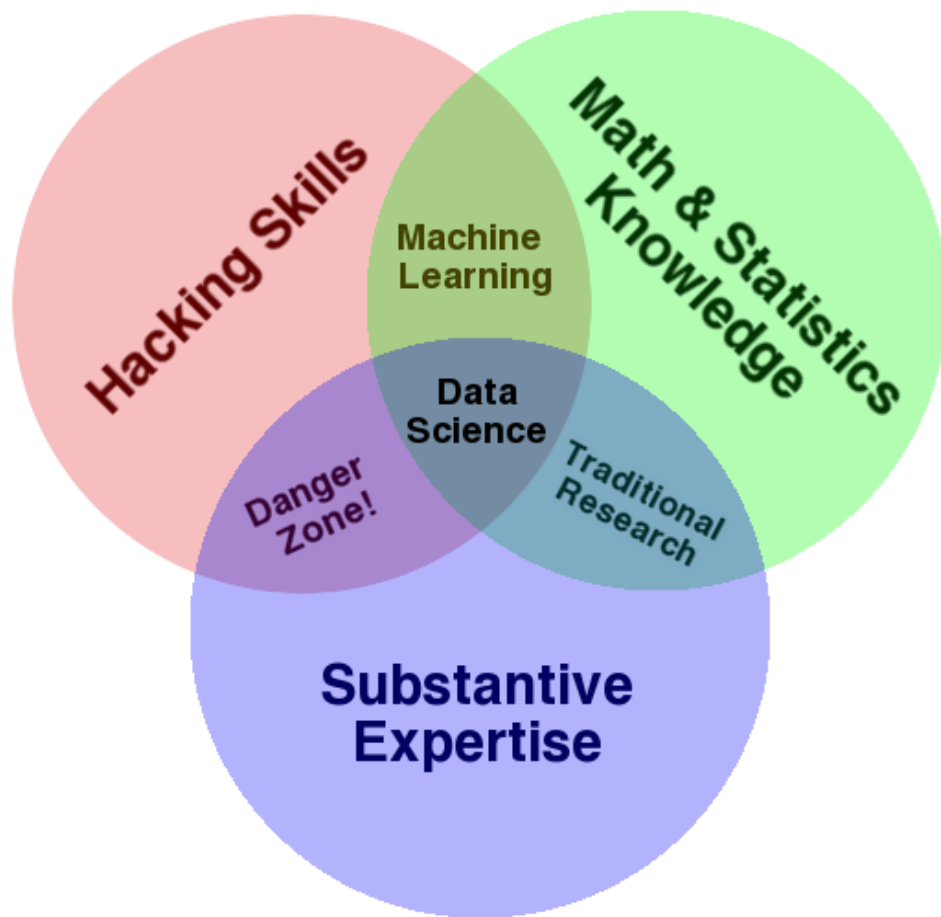
LOTS OF ANALYSIS PRACTICE WITH REAL DATA

“CRITICAL THINKING WITH DATA”

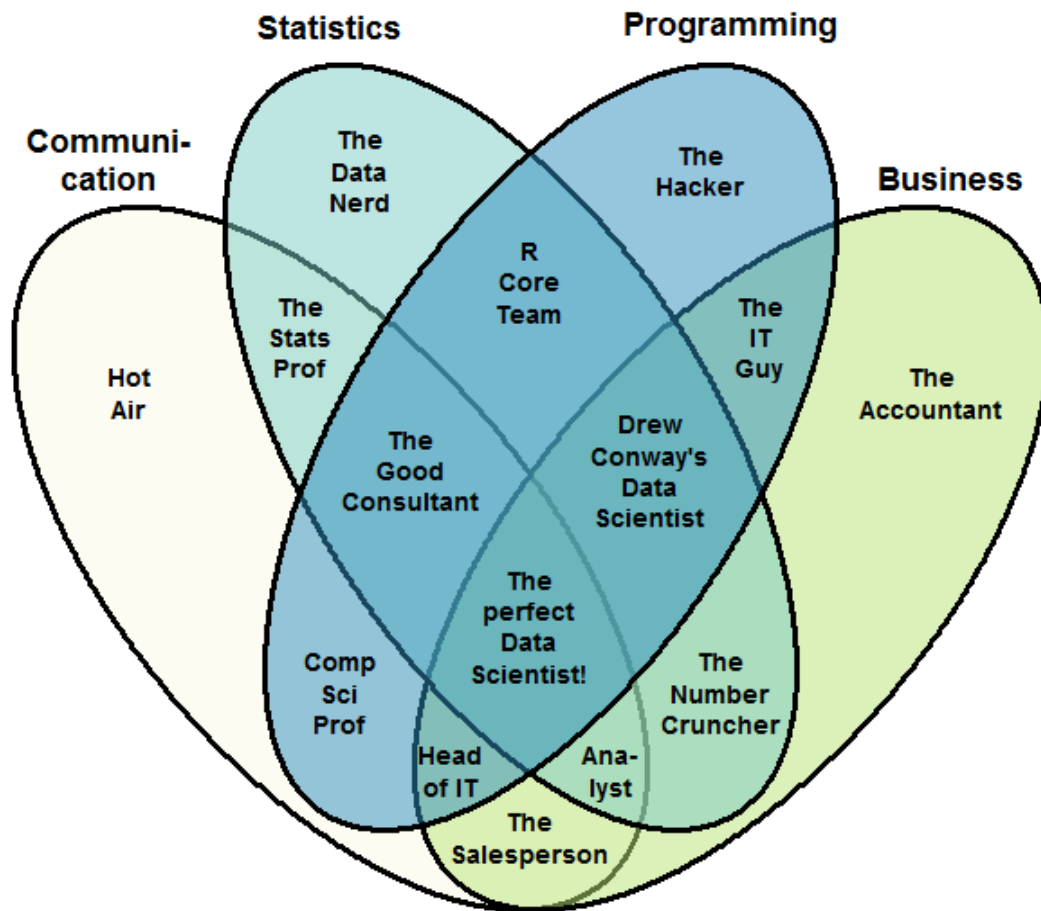
(And building competency actually doing data analysis.)

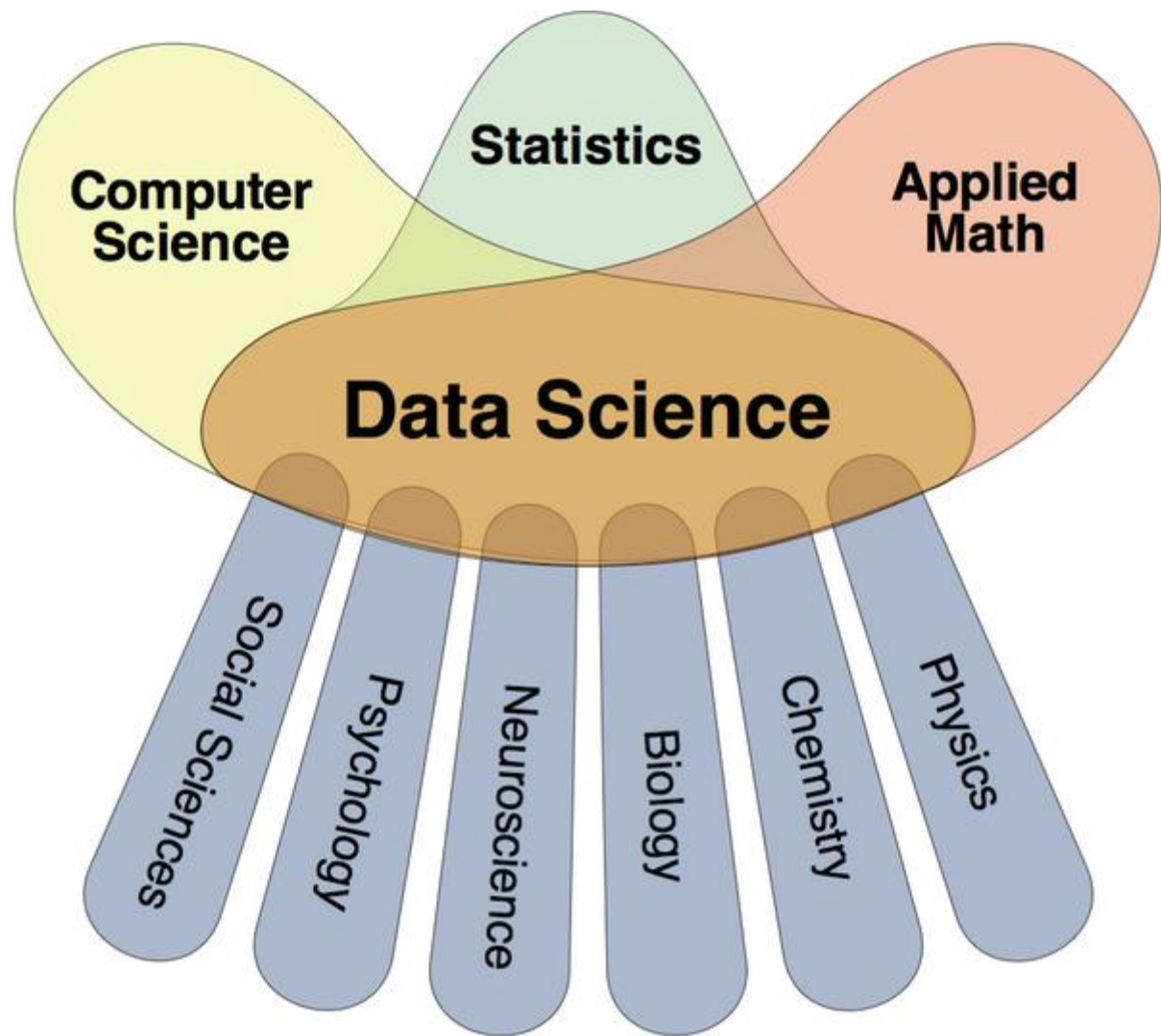
**SO IS THIS A
“DATA SCIENCE”
CLASS?**

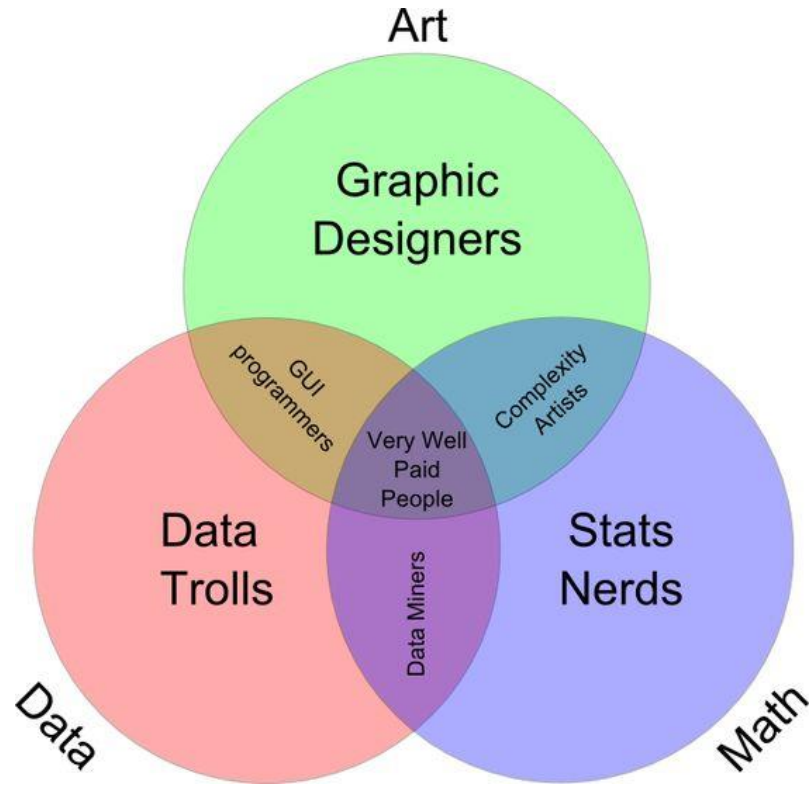
(Sort of.)

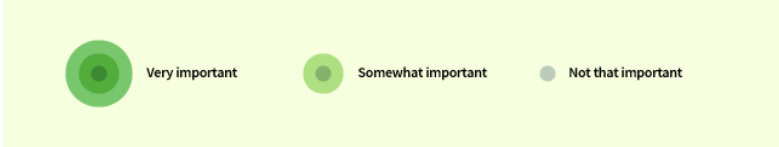
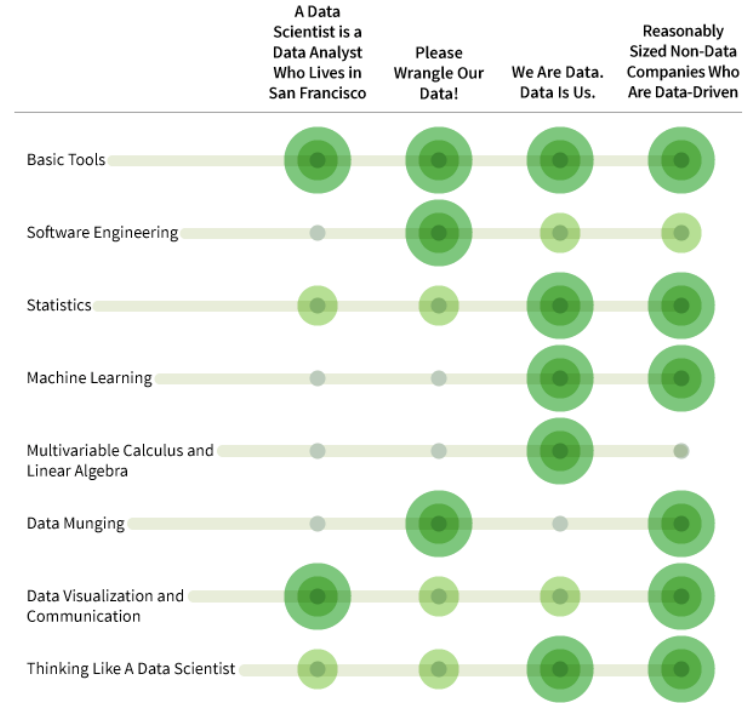


The Data Scientist Venn Diagram









4 types of data science jobs with a breakdown of the 8 skills you need to get the job

DATA ANALYSIS HAS BEEN AROUND FOR A WHILE

1935: "The Design of Experiments"

R.A. Fisher



1939: "Quality Control"

W.E. Demming

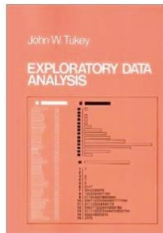


1958: "A Business Intelligence System"



Peter Luhn

1977: "Exploratory Data Analysis"



1989: "Business Intelligence"

Howard
Dresner



1997: "Machine Learning"

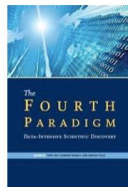


2010: "The Data Deluge"

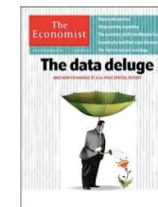
1996: Google



2007: "The Fourth Paradigm"



2009: "The Unreasonable Effectiveness of Data"



Abridged Version of Jeff Hammerbacher's timeline for Berkeley CS 194, 2012

WHAT'S DIFFERENT NOW?

More (accessible) data

Better tools!

More opportunities

“CRITICAL THINKING WITH DATA”

(And building competency actually doing data analysis.)

COURSE INFO

Part 1: Analytics

Part 2: Visualization

Project

September / October

November / December

Class website:

Slack channel – everyone should have gotten an invite

LECTURES



TUTORIALS

You will learn about:

- Data scraping
- Data cleaning
- Simple statistical analysis with R
- Analysis with Tableau
- Making reports
- how to design visuals to answer questions about data



GRADING SCHEME

- Assignments: 50%
 - check the website for due dates of assignments and how to submit them
- Project: 40% (presentation + report)
- Slack participation: 10%

READINGS

mostly for additional interest

will announce readings on a per-lecture basis

ELECTRONICS POLICY

Laptops and devices okay
(in fact you'll need them)
...but use them for work!



BEHAVIOR & SOCIETY

Students are Better Off without a Laptop in the Classroom

What do you think they'll actually use it for?

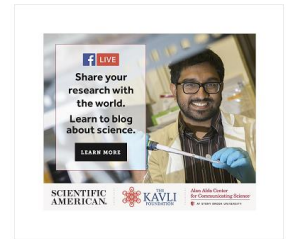
By Cindi May on July 11, 2017



Credit: Getty Images

As recent high school graduates prepare for their migration to college in the fall, one item is sure to top most students' shopping wish lists: a laptop computer. Laptops are ubiquitous on university campuses, and are viewed by most students as absolute must-have items, right alongside laundry detergent, towels, and coffee pots.

Without question, personal laptops can enhance the college experience by facilitating engagement with online course material, providing access to sources for research, maximizing internship searches, and even improving communication with friends and parents. Many students also opt to bring their laptops to class so that they can take notes, view online lecture slides, and search the web for course-related material. This practice, it



ADVERTISEMENT

LATEST NEWS



Puerto Rico Looks to Alphabet's X Project Loon Balloons to Restore Cell Service



The Ethical Minefields of Technology



Astronomers Are Finally Mapping the "Dark Side" of the Milky Way

QUESTIONS

WHAT IS VISUAL ANALYTICS

And where does it come from?

WHAT IS DATA ANALYSIS?

- traditionally: data analysis = statistics
- generally: data analysis = careful thinking about evidence (data)
- data analysis now covers a range of activities and skills
 - defining your problem
 - disassembling problems and data into analyzable pieces
 - evaluate the data & draw conclusions
 - make or recommend a decision

DATA ANALYSIS EXAMPLE

What might we be interested in analyzing?

What do you notice in the data?

	September	October	November	December	January	February
Gross sales	\$5,280,000	\$5,501,000	\$5,469,000	\$5,480,000	\$5,533,000	\$5,554,000
Target sales	\$5,280,000	\$5,500,000	\$5,729,000	\$5,968,000	\$6,217,000	\$6,476,000
Ad costs	\$1,056,000	\$950,400	\$739,200	\$528,000	\$316,800	\$316,800
Social network costs	\$0	\$105,600	\$316,800	\$528,000	\$739,200	\$739,200
Unit prices	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

What has been happening during the last six months with sales?

How do their gross sales figures compare to their target sales figures?

	September	October	November	December	January	February
Gross sales	\$5,280,000	\$5,501,000	\$5,469,000	\$5,480,000	\$5,533,000	\$5,554,000
Target sales	\$5,280,000	\$5,500,000	\$5,729,000	\$5,968,000	\$6,217,000	\$6,476,000
Ad costs	\$1,056,000	\$950,400	\$739,200	\$528,000	\$316,800	\$316,800
Social network costs	\$0	\$105,600	\$316,800	\$528,000	\$739,200	\$739,200
Unit prices (per oz.)	\$2.00	\$2.00	\$2.00	\$1.90	\$1.90	\$1.90

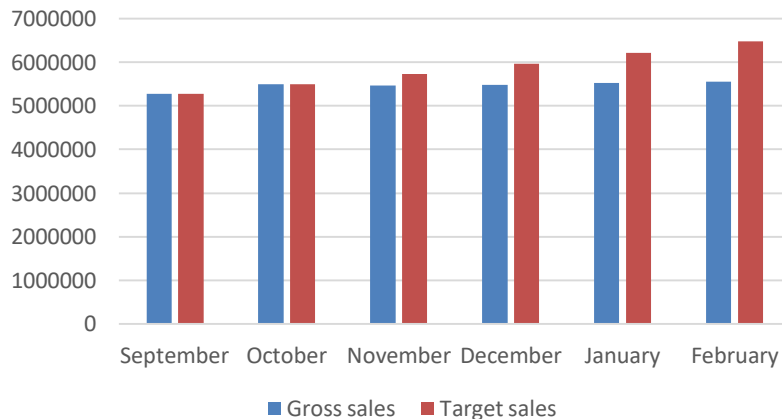
Do you see a pattern in Acme's expenses?

What do you think is going on with these unit prices? Why are they going down?

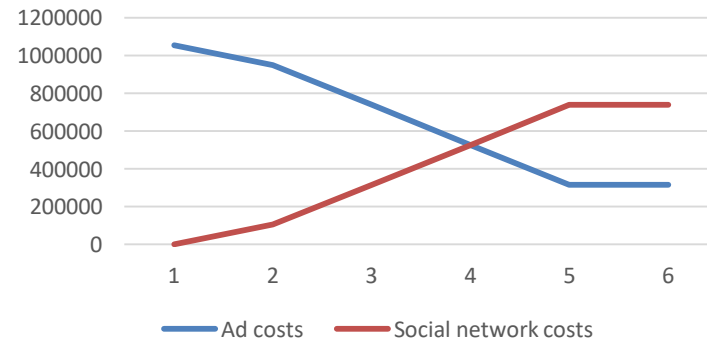
VISUAL ANALYTICS

“the science of analytical reasoning facilitated by interactive visual interfaces” [1]

Gross Sales vs. Target Sales



Ad costs vs. social network costs



VISUAL ANALYTICS

Visual analytics combines **automated analysis** techniques with **interactive visualizations** for an effective understanding, reasoning and decision making on the basis of **very large and complex data sets** [5].

GRAND CHALLENGE

Enable profound insight

– allow an analyst to examine

- massive, multi-dimensional, multi-source, time-varying information
- to make the right decisions (in time-critical manner)

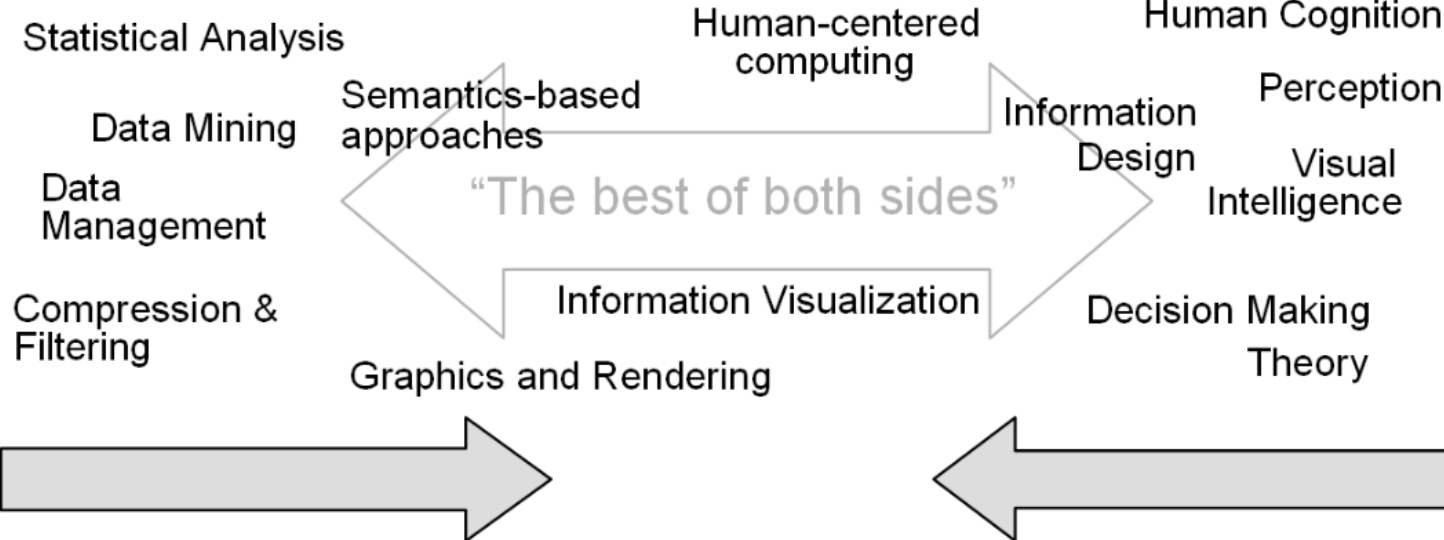
METHOD

- combine automated analysis with human intervention
- represent data visually to
 - allow interaction
 - insight generation
 - drawing of conclusions
 - make better decisions

SCOPE

automated analysis

human analysis



CONFIRM VS. EXPLORE

confirmatory analysis

- start with a hypothesis about the data
- confirm that it is true

focus of fully automated analysis methods

exploratory analysis

- likely no a-priori information about the data
- not sure about patterns and information present
- explore to create hypotheses & confirm later

focus of visual analytics

SCOPE

visual analytics = an iterative process that involves

- information gathering
- data preprocessing
- knowledge representation
- interaction
- decision making.

EXAMPLES

<https://www.youtube.com/watch?v=K9PvskathGI>

EXAMPLES

Baseball4D

A Tool for Baseball Game Reconstruction & Visualization

Carlos Dietrich¹, David Koop², Huy T. Vo², and Cláudio T. Silva²

¹Independent Consultant, E-mail: cadietrich@gmail.com

²New York University, E-mail: {dakoop, huy.vo, csilva}@nyu.edu

For this and the following videos, see:

<http://ieevis.org/year/2014/info/overview-amp-topics/paper-sessions>

EXAMPLES

Integrating Predictive Analytics and Social Media

Yafeng Lu, Robert Krüger, Dennis Thom, Feng Wang,
Steffen Koch, Thomas Ertl, Ross Maciejewski

ASU VADER

USTUTT VIS

online demo: <https://www.youtube.com/watch?v=Zwjg8w8Xigo>

EXAMPLES

LoyalTracker: Visualizing Loyalty Dynamics in Search Engines

Conglei Shi, Yingcai Wu, Shixia Liu, Hong Zhou and Huamin Qu

EXAMPLES

PEARL: An Interactive Visual Analytic Tool for Understanding
Personal Emotional Style Derived from Social Media

Jian Zhao, Liang Gou, Fei Wang, and Michelle Zhou

University of Toronto

IBM Research

EXAMPLES

A System for Visual Analysis of Radio Signal Data

Tarik Crnovrsanin (tecnovr@ucdavis.edu)

Chris Muelder (cwmuelder@ucdavis.edu)

Kwan-Liu Ma (ma@cs.ucdavis.edu)

VIDI lab @ University California, Davis



EXAMPLES

#FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media

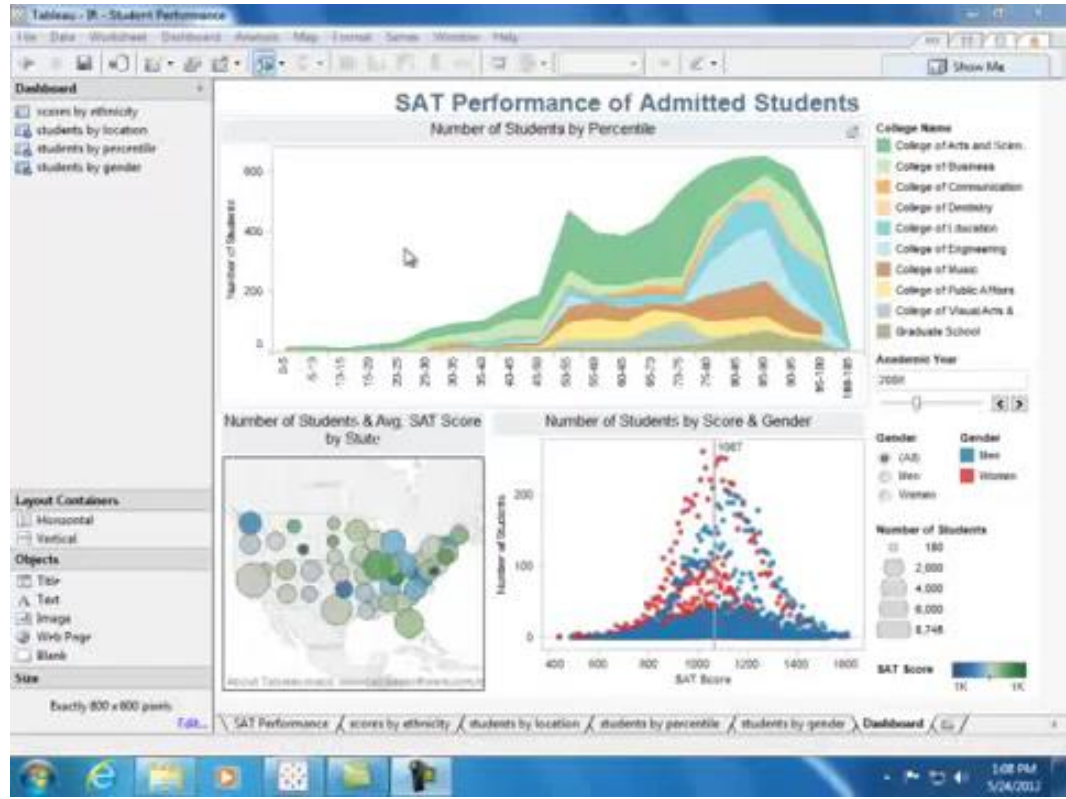
*Jian Zhao, Nan Cao, Zhen Wen, Yale Song,
Yu-Ru Lin, Christopher Collins*



UNIVERSITY OF
TORONTO



EXAMPLES



https://www.youtube.com/watch?v=_Ytz8op5lig&list=PL722C2D5AE0BF7E99

REQUIREMENTS

development & understanding of

- data transformations & analysis algorithms
- analytical reasoning techniques
- visual representations and interactions
- techniques for production, presentation, and dissemination

CHALLENGES

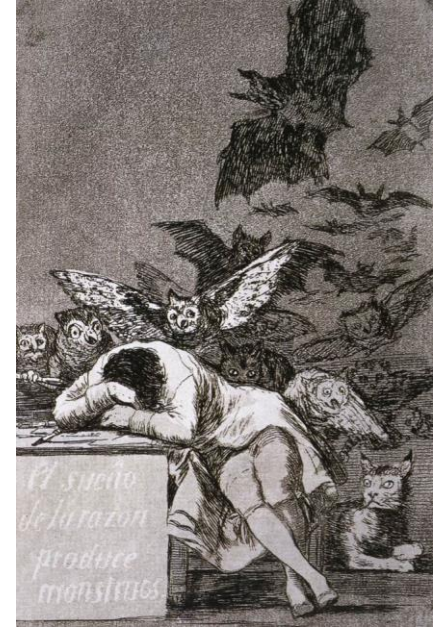
data

- help machines understand semantics
- quality of data is often low
- dealing with uncertainty in the data
- understanding the history or trustworthiness of data
- quantity (e.g. large and streaming data)

CHALLENGES

human reasoning & decision making

- understanding and supporting how humans reason about data
- support convergent & divergent thinking
- create interfaces that are meaningful, clear, effective, and efficient

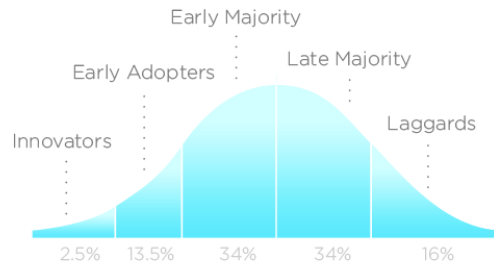


Francisco de Goya, *The Sleep of Reason Produces Monsters* (*El sueño de la razón produce monstruos*), c. 1797

CHALLENGES

adoption

- communicate benefits of developed tools to drive frequent use
- make tools accepted by users



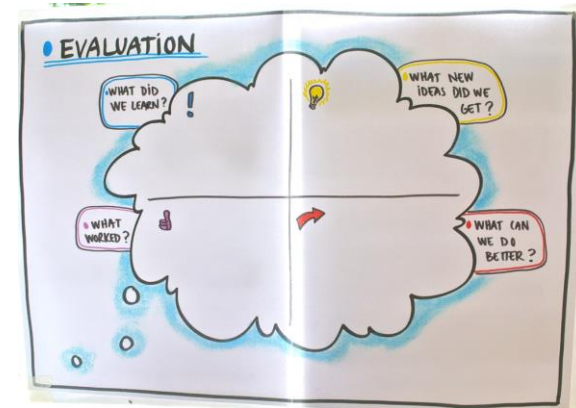
INNOVATION ADOPTION LIFECYCLE

https://en.wikipedia.org/wiki/Technology_adoption_life_cycle

CHALLENGES

evaluation

- develop methods to compare novel tools to existing ones
- assess how good (effective, efficient, etc.) a tool is
 - very difficult for measures other than time & error, e.g. how many insights a tool generates



<https://www.flickr.com/photos/co-laborate/7269016072>

CHALLENGES

problem interdependence

- analysis in the “real world” often does not consist of isolated problems or questions
- problems are often correlated and how one is solved influences how one should approach another
- synthesis of analyses is needed

CHALLENGES

integration of analysis methods

- it is simple to do many isolated analyses
- it is hard to integrate them well into one tool, interface for human analysis

CHALLENGES

SCALABILTY

SCALABILTY

SCALABILTY

SCALABILITY TYPES

information scalability

- capability to extract relevant information from massive (possibly dynamically changing) data streams
- methods: abstract data sets, filter & reduce data, represent data in multi-resolution

SCALABILITY TYPES

display scalability

- capability to of visualizations and tools to scale to different types of displays



Sony SmartWatch

MOORE'S LAW...

Microprocessor Transistor Counts 1971-2011 & Moore's Law

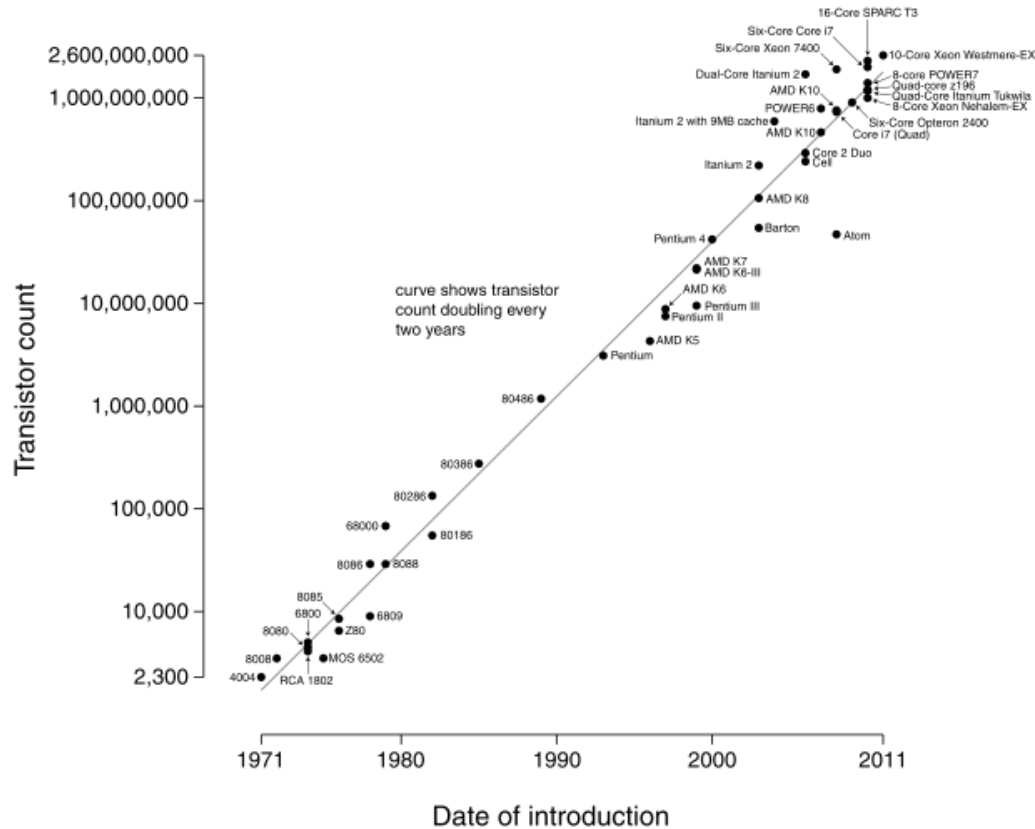
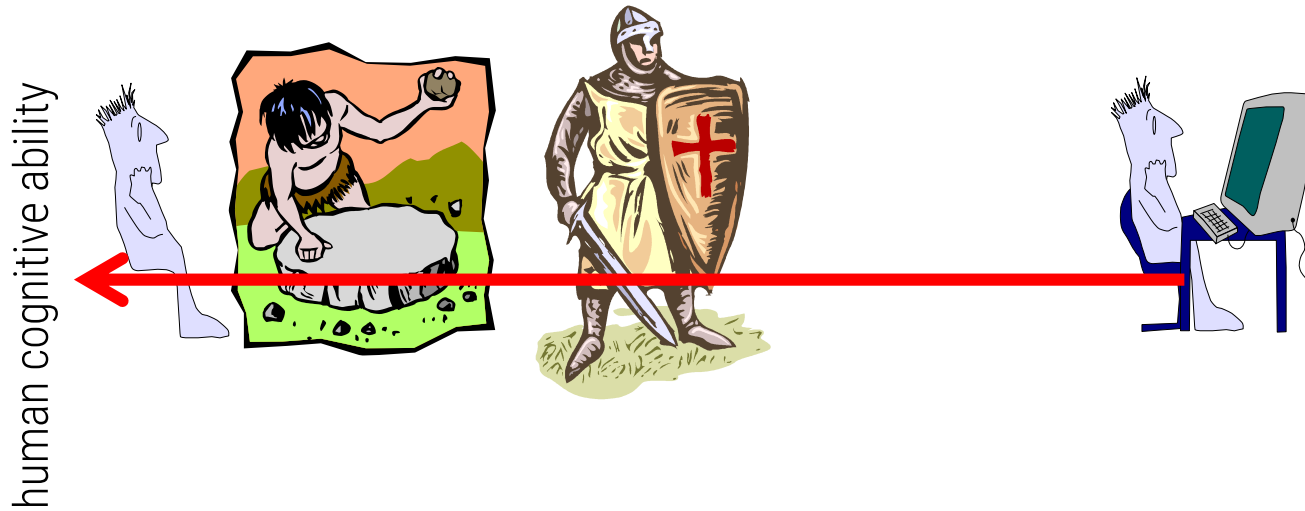


image source wikipedia

PEOPLE STAY ~THE SAME ...

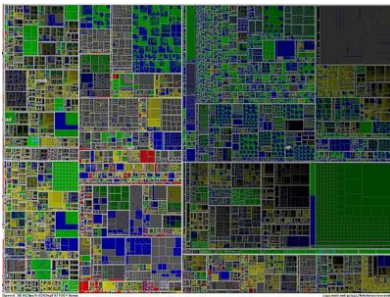


information glut = we can access more information than we can process

SCALABILITY TYPES

visual scalability

- capability to of visualizations to effectively display massive data sets in terms of number of data items or dimensions
- depends on quality of layout, interaction techniques, perceptual capabilities



Treemap of a million items

<http://www.cs.umd.edu/hcil/millionvis/>

SCALABILITY TYPES

human scalability

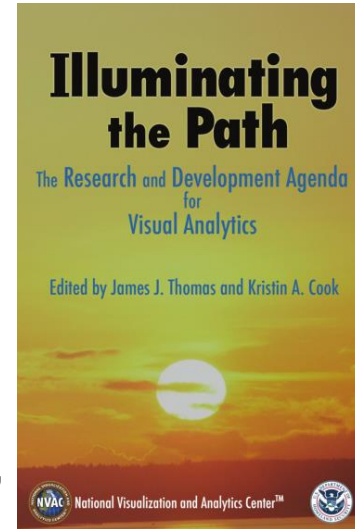
- human skills don't scale but numbers of humans involved in analysis can
- techniques must scale from a single to multiple users

SCALABILITY TYPES

- software scalability
 - software systems and algorithms must scale to larger data & different data
- others
 - privacy and security in multi-user settings
 - collaboration across languages and borders

HISTORY

- outgrowth of the Scientific & Information Visualization community
- started with US National Visualization and Analytics Center (NVAC) at PNNL in 2004
- developed the first research and development agenda “Illuminating the Path”
- sponsored initially by DHS (US Department of Homeland Security)

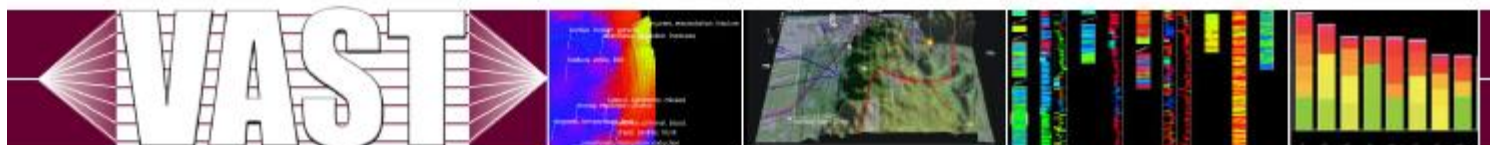


ORIGINAL GOALS

- analyzing terrorist threats
 - safeguarding borders and ports
 - preparing for and responding to emergencies
- now only part of the larger research goals

HISTORY

- VAST symposium ☒ conference
 - visual analytics, science, and technology
- part of the IEEE Visualization conference
- started Visual Analytics as its own research area in 2006



HISTORY

- 2008 EU funds VisMaster, a Coordination Action to join European academic and industrial R&D
- in Europe initial focus not on “homeland” security, rather broad applicability
 - physics, astronomy, climate monitoring, weather, etc.

HISTORY

- many centers in Europe
- In France mainly Inria
- web: visual-analytics.eu
- book: Mastering the information age – solving problems with visual analytics
- YouTube: you saw it already

FUTURE

The Sexiest Job of the 21st Century: Data Analyst

Chris Morris, Special to CNBC.com
Wednesday, 5 Jun 2013 | 1:00 PM ET



Photo: Biddiboo | Getty Images

In tech jobs market, data analysis is tops

Jon Swartz, USA TODAY 10:20 a.m. EDT October 5, 2012

Second of five reports this week on the job outlook in key industries.



(Photo: Elaine Thompson, AP)

f 256 CONNECT
t 215 TWEET
in 47 LINKEDIN
3 COMMENT
EMAIL
MORE

SAN FRANCISCO -- Like a coveted free agent in sports, Kelly Halfin had a multitude of choices when she decided to take a job in tech in the U.S.

The Belgian had five American companies lined up, eager to sign her on to lead their data analysis

READINGS

1. Illuminating the Path: The Research and Development Agenda for Visual Analytics Paperback – January 1, 2005 by James J. Thomas (Editor), Kristin A. Cook (Editor)
2. Daniel A. Keim and Florian Mansmann and Jörn Schneidewind and Hartmut Ziegler and Jim Thomas, *Visual Analytics: Scope and Challenges*, 2008, Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, Springer, Lecture Notes In Computer Science (Incs)
3. Michael Milton. Head First Data Analysis: A learner's guide to big numbers, statistics, and good decisions.
4. Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges (pp. 154-175). Springer Berlin Heidelberg.