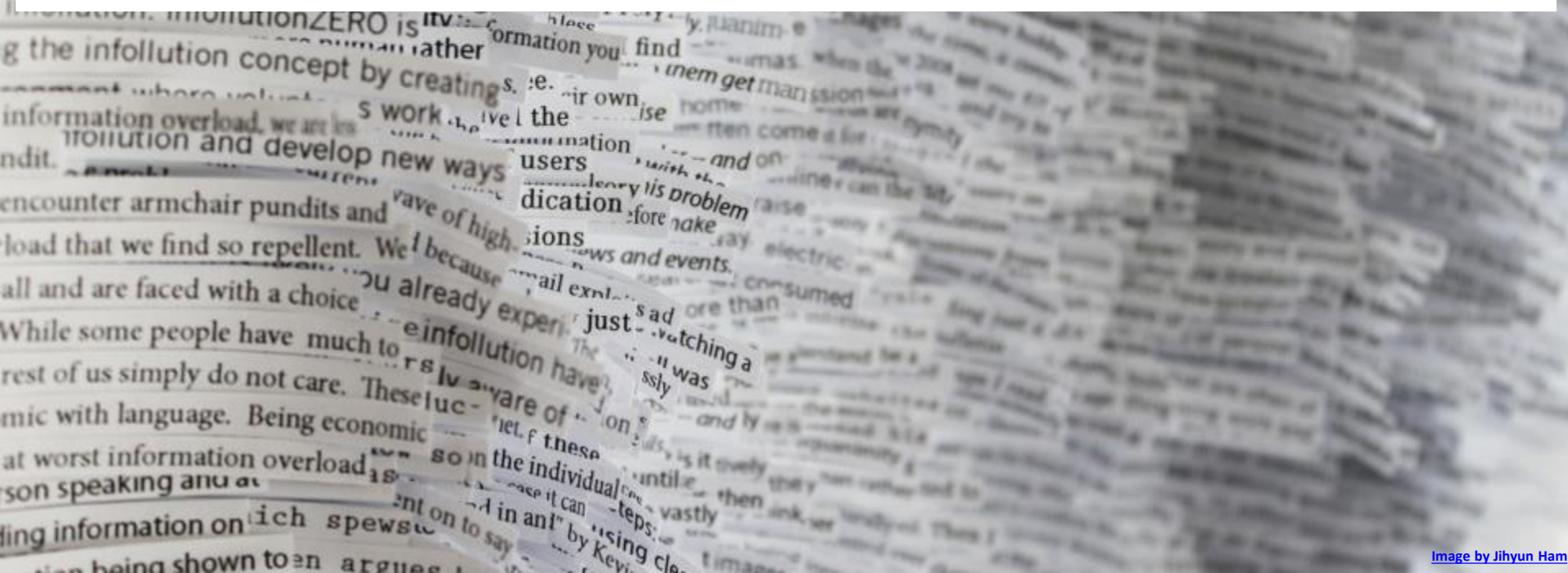# VISUAL ANALYTICS INTRODUCTION

LECTURE 1

Petra Isenberg

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

SDSS

WALMART

- 1 million customer transactions per hour

- likely has information on >145 million Americans [1]

Image http://harryhammer.wordpress.com/2010/03/20/wal-mart-a-true-champion/
[1] http://centerformediajustice.org/wp-content/files/WALMART_PRIVACY_MB.pdf

...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
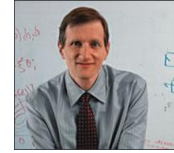 http://www.loc.gov/about/fascinating-facts/

WHAT IS USEFUL?

- data != useful information
- you want insight

→ analysis is needed

ANALYSIS IS NOT SIMPLE

- research project: predict U.S. unemployment rate

Gary King, Harvard

- 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

- learn how to analyze (& visualize) data
  - Getting data
  - Cleaning data
  - Analyzing data
  - Visualizing data (with existing & new tools)

INSTRUCTORS

Petra Isenberg

petra.isenberg@inria.fr

Acknowledgements

- Wesley Willett co-designed the course and made many of the original slide decks

OFFICE HOURS

- offices: at Université Paris Sud / Bâtiment 660 (plateau de Saclay)
- email me for appointments

COURSE INFO

Part 1: Analytics	Part 2: Visualization	Exam
September / October	November	TBD

Class website:
http://tinyurl.com/VisualAnalytics2016

LESSON PLAN

- Part 1 "Analysis" (roughly)
  – Lecture 1: Introduction
  – Lecture 2: Data Collection + Data and Ethics
  – Lecture 3: Data Cleaning / Wrangling
  – Lecture 4: Sensemaking
  – Lecture 5: Basic Statistics
  – Lecture 6: Reproducible Research
  – Lecture 7: Analysis Tools (you!)
  – Lecture 8: Big Data Analytics
  – Lecture 9: Application Areas

TUTORIALS

- ## You will learn about:
  - Data scraping
  - Data cleaning
  - Simple statistical analysis with R
 - Analysis with Tableau
  - Making reports
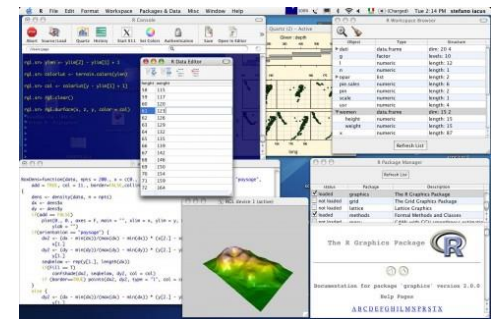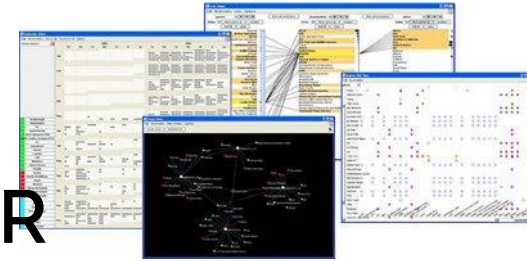
LESSON PLAN

- Part 2 "Information Visualization"
  - Lecture 1: Introduction
  - Lecture 2: Perception & Color
  - Lecture 3: Multi-Dimensional Data
  - Lecture 4: Interaction
  - Lecture 5: Time and Animation
  - Lecture 6: Graphs and Trees

TUTORIALS

- ## You will learn
  - how to design visuals to answer questions about data
  - how to apply visualization guidelines to design
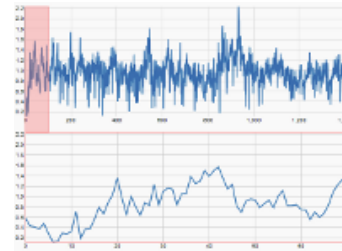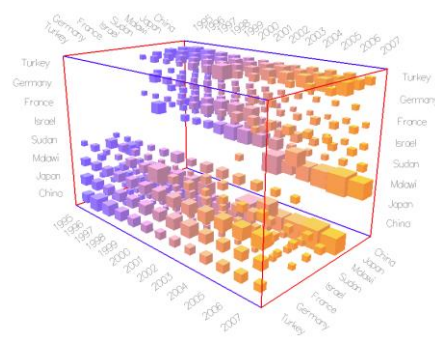
GRADING SCHEME

- Assignments: 70%
  - check the website for due dates of assignments and how to submit them
- Exam: 30%

READINGS

- will announce readings on a per-lecture basis

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

reference [3]

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

reference [3]

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?

reference [3]

VISUAL ANALYTICS

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

Gross Sales vs. Target Sales
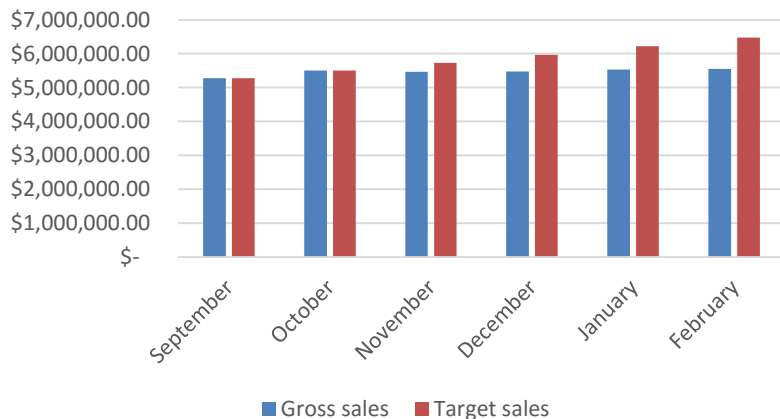
Ad costs vs. social network costs
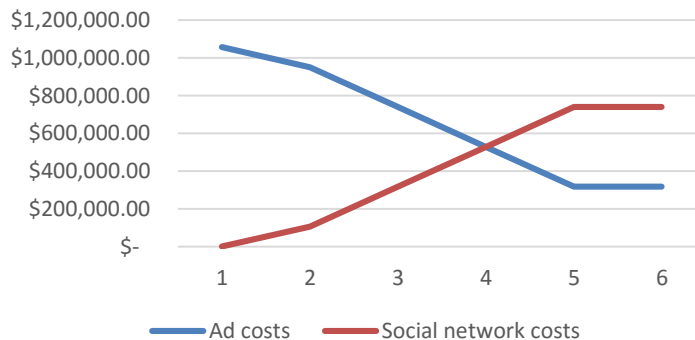
VISUAL ANALYTICS

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
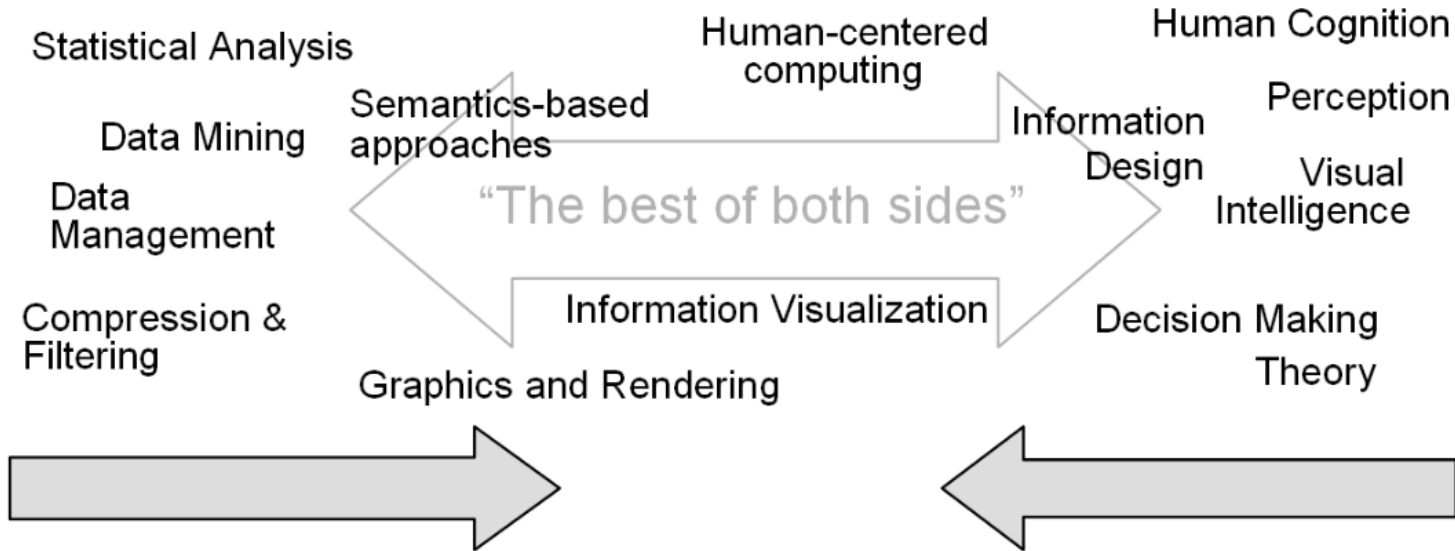
- represent data visually to
  - allow interaction
  - insight generation
  - drawing of conclusions
  - make better decisions

Reference [2]

SCOPE

automated analysis   human analysis

Statistical Analysis   Human-centered   Human Cognition
                                  computing
Semantics-based   Perception
approaches   Information
Data Mining   Design   Visual
                "The best of both sides"   Intelligence
Data
Management

Compression &   Information Visualization   Decision Making
Filtering   Theory

Graphics and Rendering

Reference [5]

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.

Reference [2]

EXAMPLES

EXAMPLES

EXAMPLES

Baseball4D

A Tool for Baseball Game Reconstruction & Visualization

Carlos Dietrich[1], David Koop[2], Huy T. Vo[2], and Cláudio T. Silva[2]

[1] Independent Consultant, E-mail: cadietrich@gmail.com
[2] New York University, E-mail: {dakoop, huy.vo, csilva}@nyu.edu

For this and the following videos, see:
http://ieeevis.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

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 (tecrnovr@ucdavis.edu)
Chris Muelder (cwmuelder@ucdavis.edu)
Kwan-Liu Ma (ma@cs.ucdavis.edu)

VIDI lab @ University California, Davis

ViDi
Visualization & Interface
Design Innovation

EXAMPLES

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

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

EXAMPLES

REQUIREMENTS

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

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

CHALLENGES

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

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

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

- scalability
  - data quantity (e.g. large and streaming data)
  - visualization of data
  - complexity and urgency of tasks
  - collaboration

MOORE'S LAW…

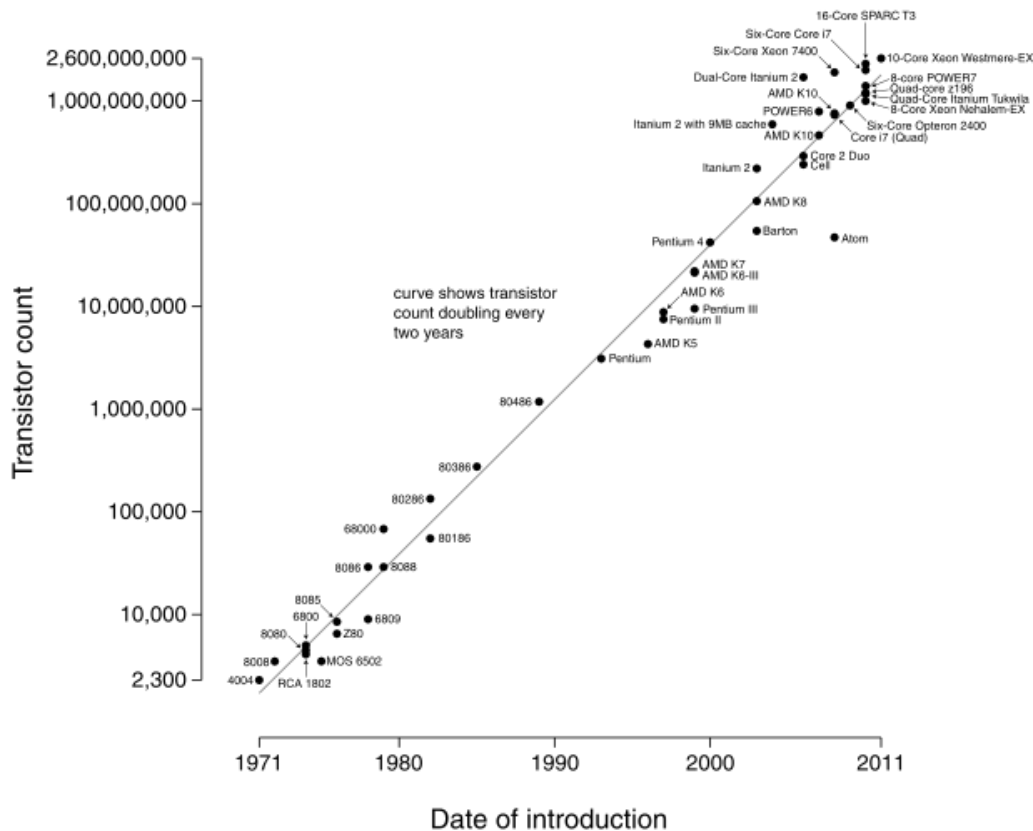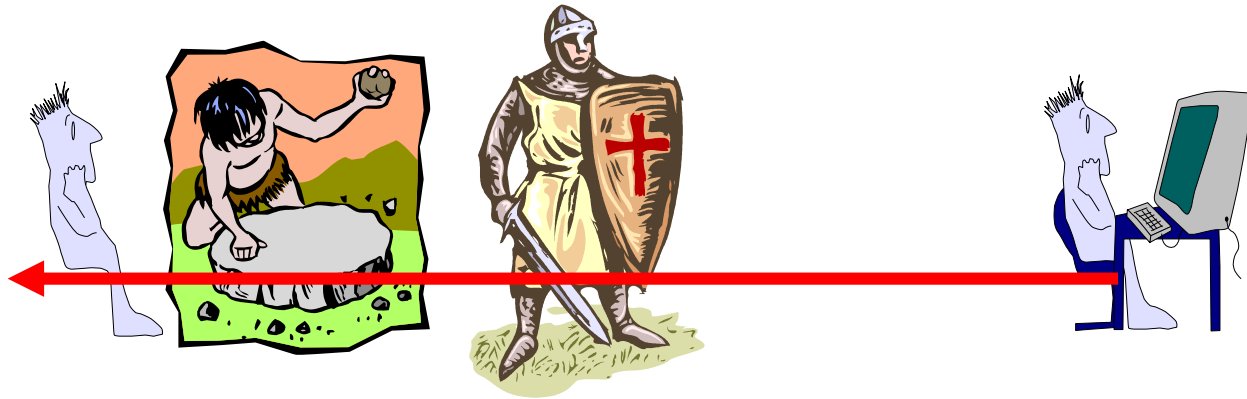Microprocessor Transistor Counts 1971-2011 & Moore's Law

image source wikipedia

PEOPLE STAY ~THE SAME ...

human cognitive ability

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

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)
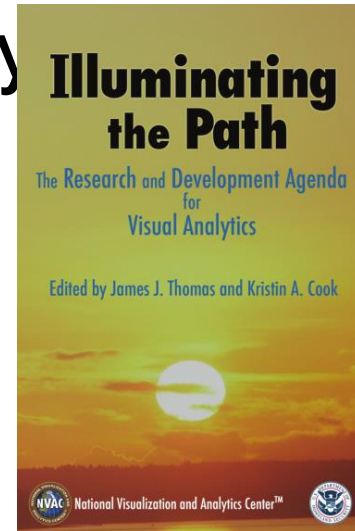
ORIGINAL GOALS

- analyzing terrorist threats
- safeguarding boarders 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

IEEE Symposium on Visual Analytics Science and Technology 2006   •   October 31 - November 2, 2006
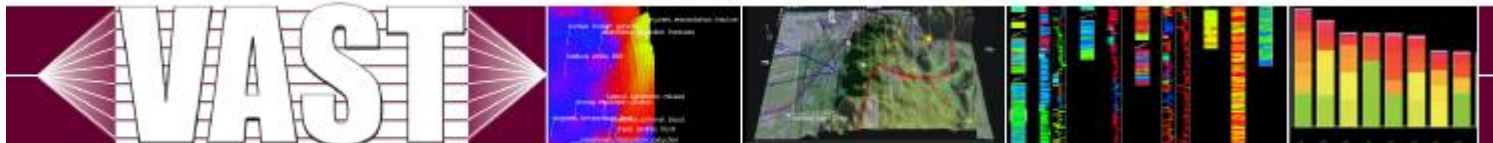
HISTORY

- 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

CNBC

Photo: Biddlboo | 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   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 (lncs)

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.