

VISUAL ANALYTICS INTRODUCTION

LECTURE 1

Petra Isenberg

INSTRUCTOR

DR. PETRA ISENBERG

petra.isenberg@inria.fr

DR. NATKAMON TOVANICH

natkamon.tov@gmail.com

OFFICE — Digiteo Moulon Building & Ecole Polytechnique

OFFICE HOURS — By appointment

SOME CLASS RULES

- Anyone may ask questions at any time
- Don't judge others' questions
- We'll use first names

If you are uncomfortable you can also call me Dr. Isenberg. Do not call me: Miss, Mrs Isenberg, Professor Isenberg, Esteemed professor, ...

- Be on time
- Be responsible in team work

VISUAL ANALYTICS

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

We're measuring more than ever before...

SLOAN DIGITAL SKY SURVEY

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

started in 2000

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

<https://www.youtube.com/watch?v=08LBItPDZw>

WALMART

Collects 2.5 petabytes of unstructured data from 1 million customers every hour [1]



...AND MORE

Big Time Big Data Statistics

- The big data analytics market is set to reach **\$103 billion by 2023**.
- Poor data quality costs the US economy **up to \$3.1 trillion yearly**.
- In 2020, **every person generated 1.7 megabytes** in just a second.
- Internet users generate about **2.5 quintillion bytes of data each day**.
- **95% of businesses** cite the need to manage unstructured data as a problem for their business.
- **97.2% of organizations** are investing in big data and AI.
- Using big data, **Netflix saves \$1 billion per year** on customer retention.
- Predictions estimate the world will generate **181 zettabytes** of data by 2025.

...BUT YOU KNOW THIS

WHY THIS COURSE?

WHAT IS USEFUL?

data != useful information

→ analysis is needed

you know what you ask → you query

you don't know what to ask → you explore

SOMETIMES QUERIES ARE HARD, TOO

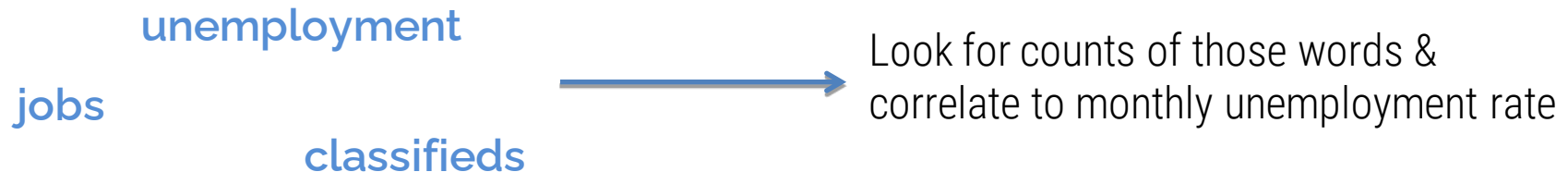


Gary King, Harvard

research project: predict U.S. unemployment rate

method: Twitter & social media analysis

→ sentiment analysis by word count



Jobs



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

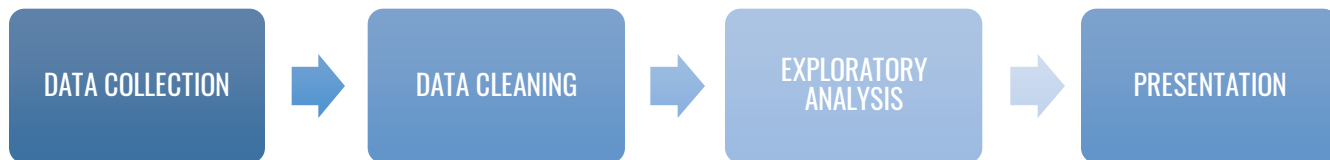
human judgment and intervention often needed

- 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 with real data



“CRITICAL THINKING WITH DATA”

(And building competency actually doing data analysis.)

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

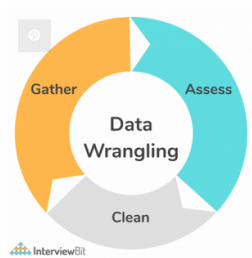
(Sort of.)

Top Data Scientist Skills You Must Have in 2022

May 13, 2022

in  

5. Knowledge of Data Wrangling and Data Exploration:



Data Wrangling is the process of cleaning and unifying messy and corrupt collections for easy access and analysis. Take, for example, the act of luggage. What happens if you stuff your entire wardrobe into your bag

6. Knowledge of Data Visualisation:



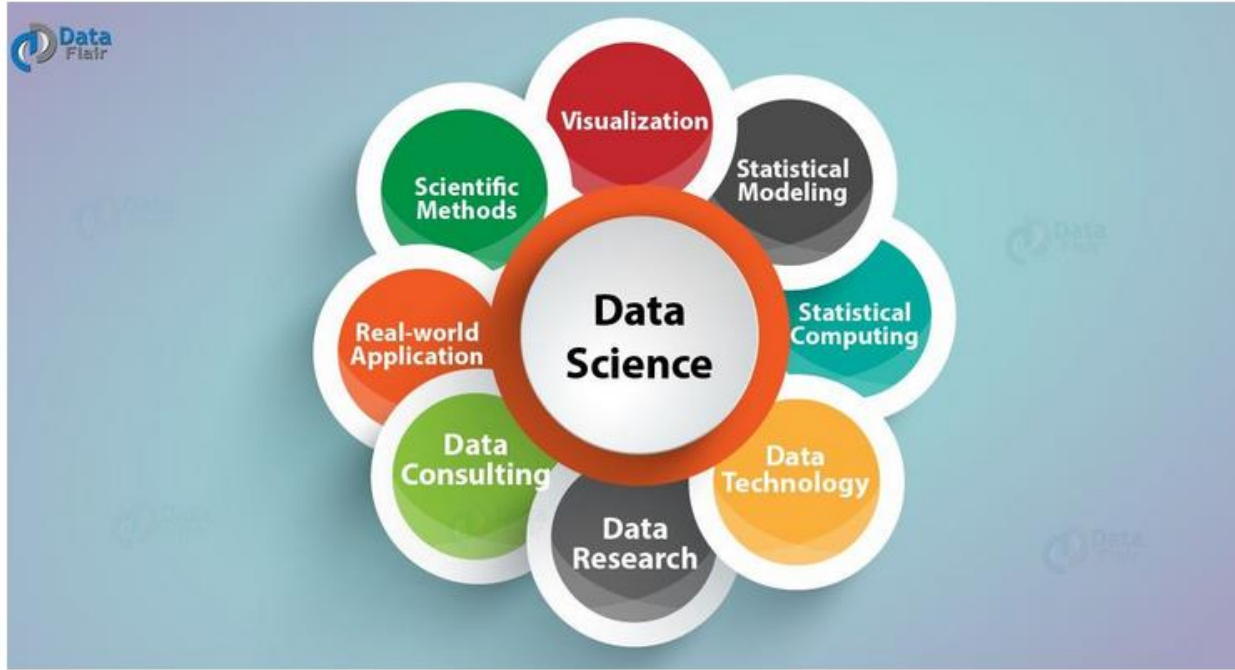
One of the most significant aspects of data analysis is data visualization. It has always been critical to convey information in a way that is both understandable and pleasant to the eye. One of the skills that Data Scientists must acquire in order to connect more effectively with end-users is data visualization. There are programs available, including Tableau, Power BI, Qlik Sense, and many others, that have a user-friendly interface.

Data visualization is more of an art than a pre-programmed procedure. There is no such thing as a "one-size-fits-all" solution here. A Data Visualization expert

Technical Skills Required to Become a Data Scientist

One of the most important technical data scientist skills are:

- [Statistical analysis and computing](#)
- [Machine Learning](#)
- [Deep Learning](#)
- Processing large data sets
- [Data Visualization](#)
- Data Wrangling
- Mathematics
- Programming
- Statistics
- [Big Data](#)



Data Science Skills That Are In High Demand In 2022 ...

These were just 3 highly ranked links---there are hundreds more like them

“CRITICAL THINKING WITH DATA”

(And building competency actually doing data analysis.)

COURSE INFO

Part 1: Analytics

September / October

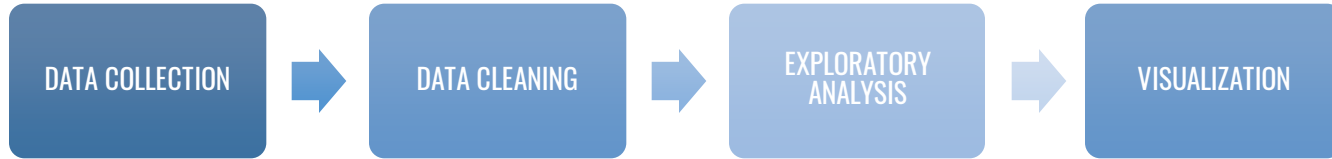
Part 2: Visualization

November / December

Project

Class website:
Edunado

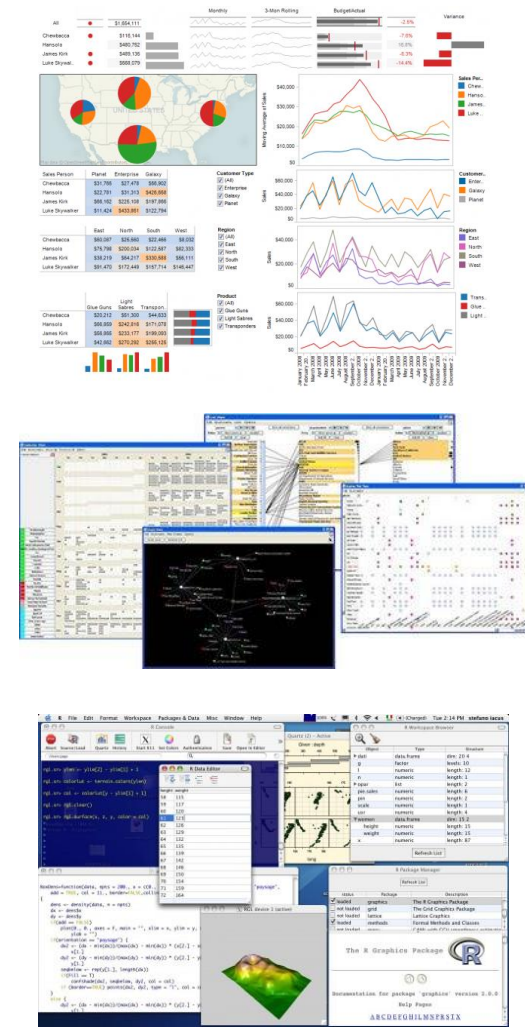
LECTURES



TUTORIALS

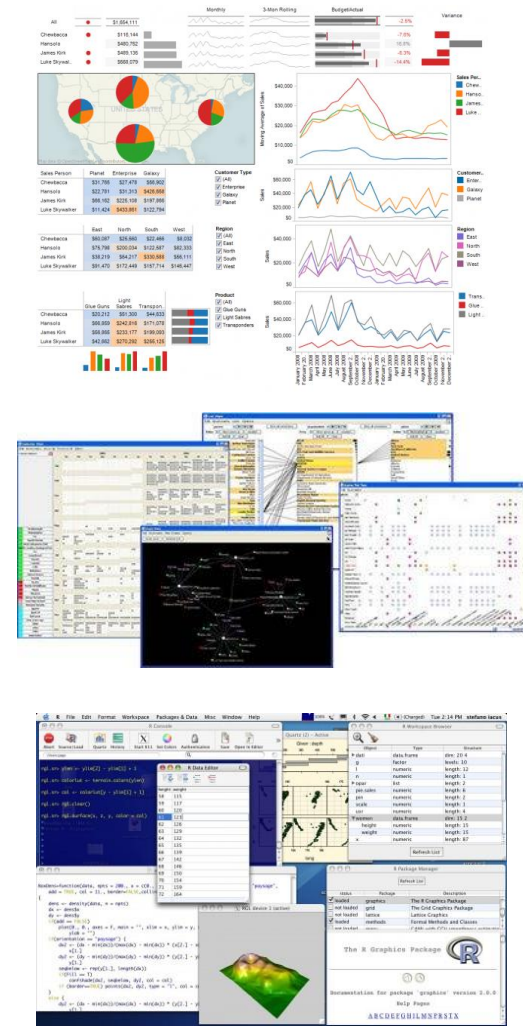
You will learn about:

- Data scraping
- Data cleaning
- Analysis with Tableau
- Programming visualizations



WORKSHOPS

- Building a project team and proposal
- Visualization ideation
- Vis implementation
- Project midterm presentation & discussion



GRADING SCHEME

- 50% Individual work
 - Quizzes at the beginning of class - about the content of the last class. (15%)
 - 3 Checkpoint submissions will form 5% of your grade (submit a topic, submit data for your research question, hand in 2 sketches)
 - Peer review quality (10%)
 - Individual Sketch (10%)
 - One-view visualization implemented in D3 (10%)
- 50% Group work
 - Exploratory data analysis with Tableau (10%)
 - Final project (40%)

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!



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

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



ADVERTISEMENT

LATEST NEWS



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



The Ethical Minefields of Technology



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

QUESTIONS

VISUALIZATION WARMUP

How to participate?



1

Go to wooclap.com

2

Enter the event code in the top banner

Event code
OKZCHH

 Enable answers by SMS

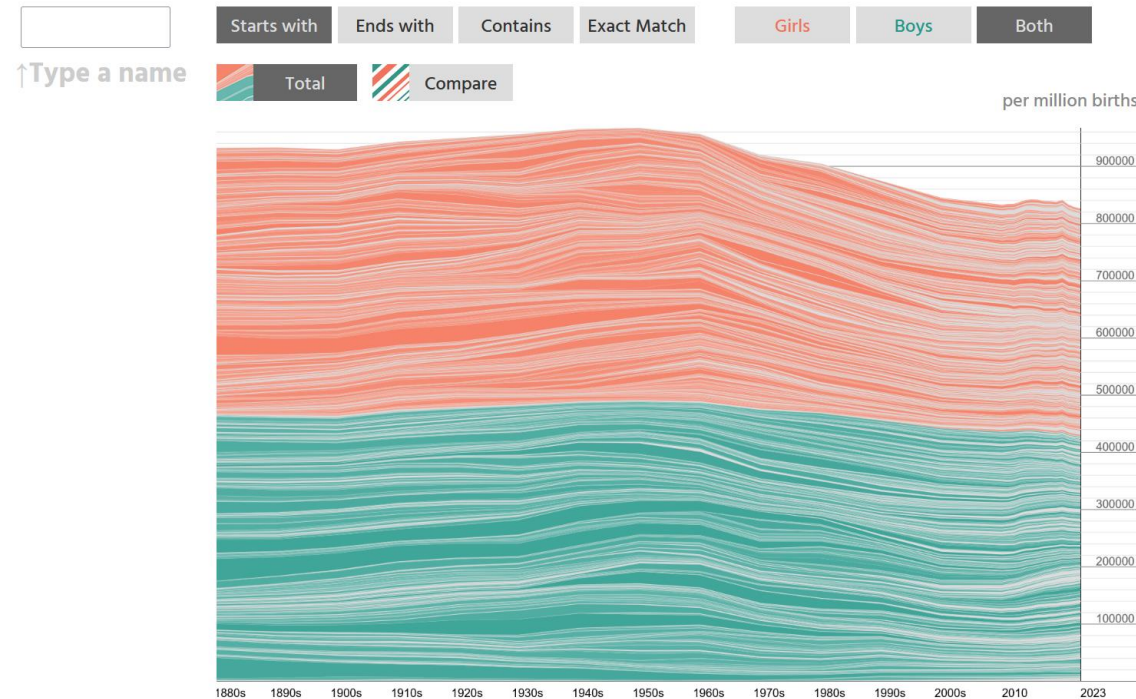
<https://namerology.com/baby-name-grapher/>

NameGrapher

Explore the historical popularity of United States baby names

Start typing in the lefthand text box and the graph will update. Click the option buttons for different types of results.

Tips: Be sure to check out both the "total" and "compare" views, and try multiple search terms separated by commas.



Acknowledgements: Evan Peck

Look for your name
+
the name of people
sitting next to you

Write down one
observation

How to participate?



1

Go to wooclap.com

2

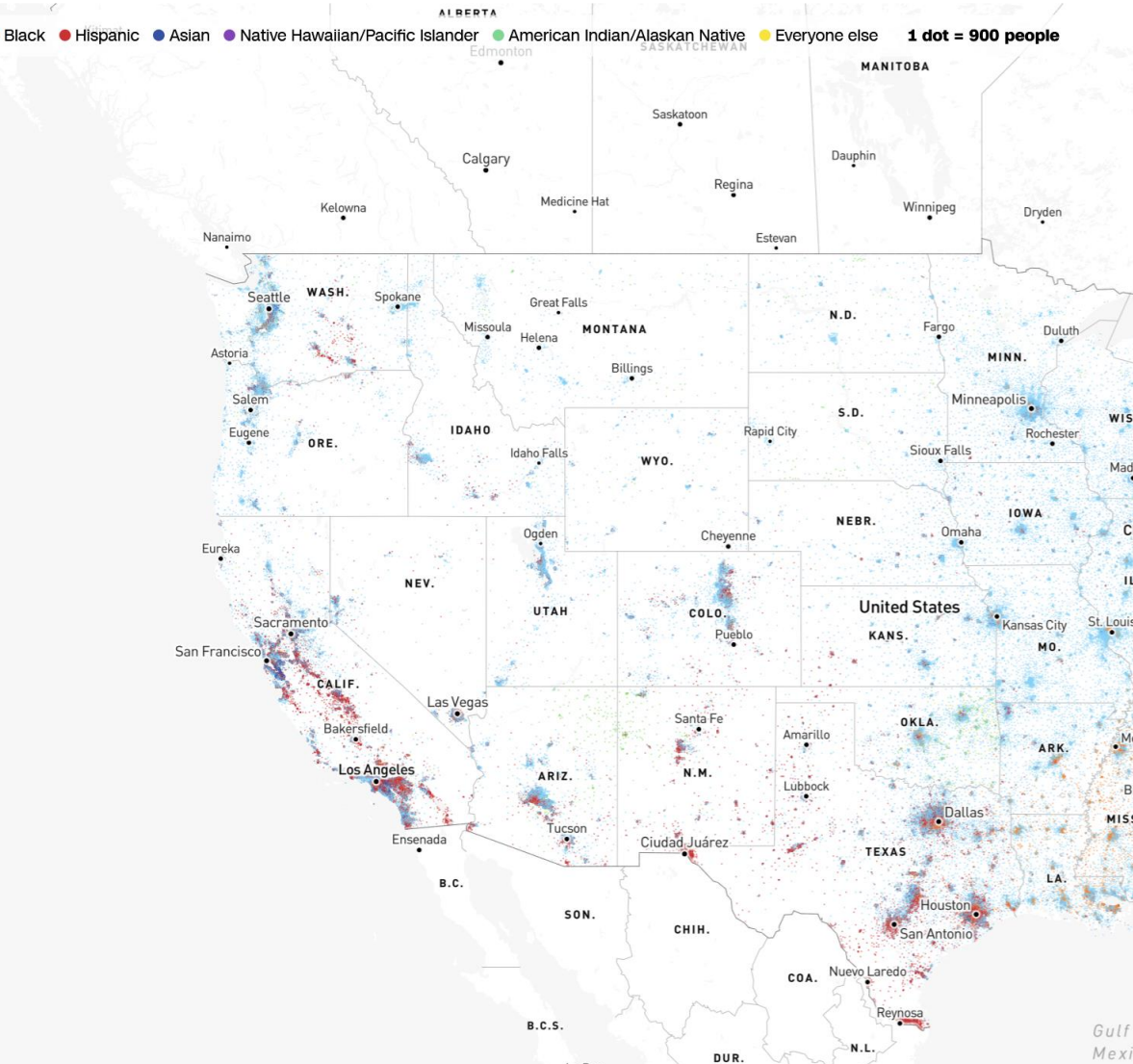
Enter the event code in the top banner

Event code
OKZCHH

 Enable answers by SMS

Q Decatur, Illinois, United States

Or try: [Los Angeles](#) | [Chicago](#) | [Miami](#)



Zoom to an area of the map where you would like to live (or have lived)
+ areas the people next to you would like to live in

Write down one observation

WHAT IS VISUAL ANALYTICS

And where does it come from?

WHAT IS DATA ANALYSIS?

Traditionally

DATA ANALYSIS = STATISTICS

WHAT IS DATA ANALYSIS?

data analysis = careful thinking about evidence (data)

WHAT IS DATA ANALYSIS?

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

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].

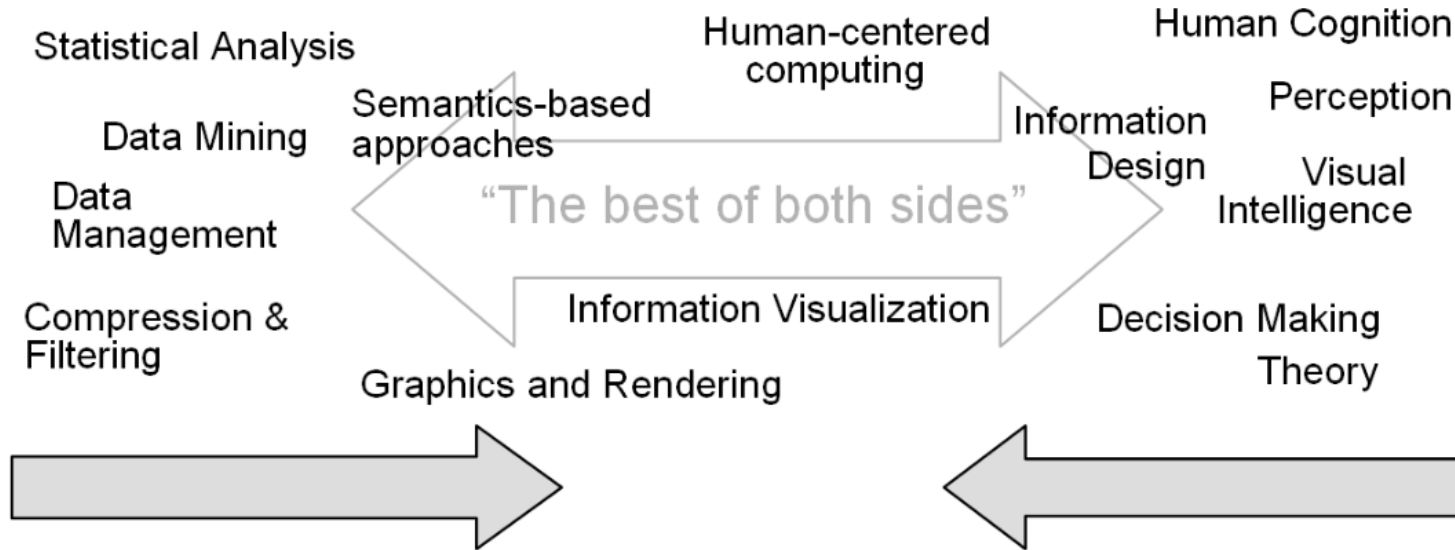
METHOD

- 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



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)

CONFIRM VS. EXPLORE

confirmatory analysis

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

focus of fully automated analysis methods

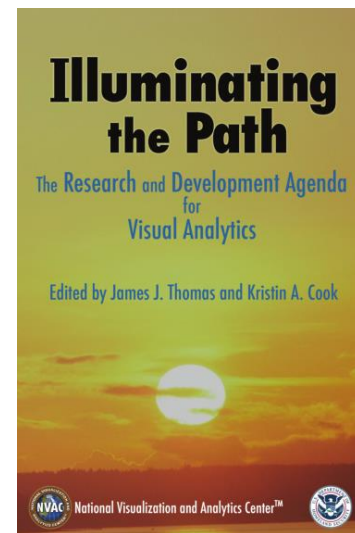
exploratory analysis

- 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

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)



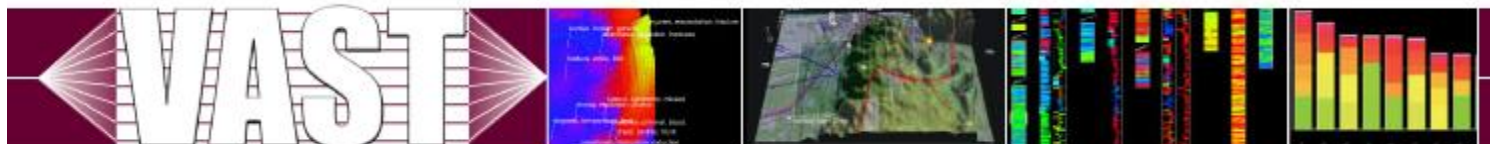
ORIGINAL GOALS

- 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 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

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 **215** **in 47** **3**
CONNECT TWEET LINKEDIN 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

HERE

- I expect that you learn the automation elsewhere (e.g. machine learning, DB classes)
- Focus
 - Getting data ready for visualization
 - Asking the right questions
 - Creating information visualizations

EXAMPLES

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

RESEARCH 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](mailto:dakoop@nyu.edu), [huy.vo](mailto:huy.vo@nyu.edu), [csilva](mailto:csilva@nyu.edu)}@nyu.edu

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>

VIS FULL PAPERS

Visualization of Human Spine Biomechanics for Spinal Surgery

Pepe Eulzer, Sabine Bauer, Francis Kilian, Kai Lawonn

Health & Disease: Friday, 0830-0845



25–30 October 2020
Virtual Conference

ieevis.org

VIS FULL PAPERS

Gender in 30 Years of IEEE Visualization

Natkamon Tovanich, Pierre Dragicevic, Petra Isenberg

Perspectives and Reflections: Wednesday,
1045-1100



24 - 29 October 2021
Virtual Conference

ieevis.org

Untidy Data: The Unreasonable Effectiveness of Tables

Lyn Bartram, Michael Correll, Melanie Tory

Studies and Evaluation Methodology: Thursday,
0915-0930

VIS FULL PAPER

Visual Analytics for RNN-Based Deep Reinforcement Learning

Junpeng Wang, Wei Zhang, Hao Yang, Chin-Chia Michael Yeh, Liang Wang

Explainable AI and Machine Learning: Wednesday,
0915-0930



24 - 29 October 2021
Virtual Conference

ieevis.org

VIS FULL PAPER

InfoColorizer: Interactive Recommendation of Color Palettes for Infographics

Lin-Ping YUAN, Ziqi Zhou, Jian Zhao, Yiqiu Guo, Fan Du, Huamin Qu

Recommendation and Automation: Wednesday,
1200-1215

IS THIS WORK DIFFICULT?

CHALLENGES

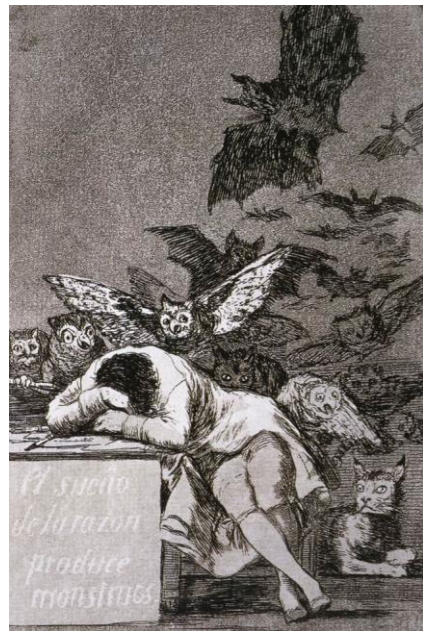
data

- 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

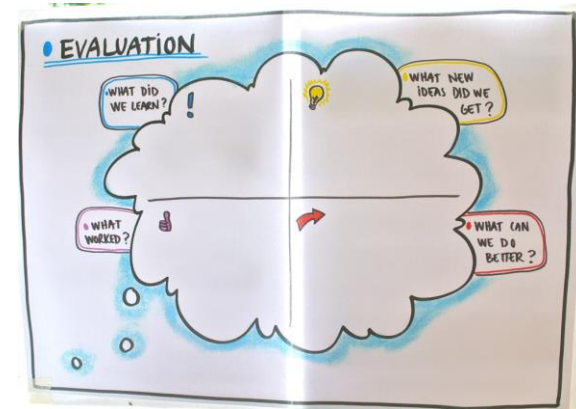


Francisco de Goya, *The Sleep of Reason Produces Monsters* (*El sueño de la razón produce monstruos*), c. 1797

CHALLENGES

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

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

MOORE'S LAW...

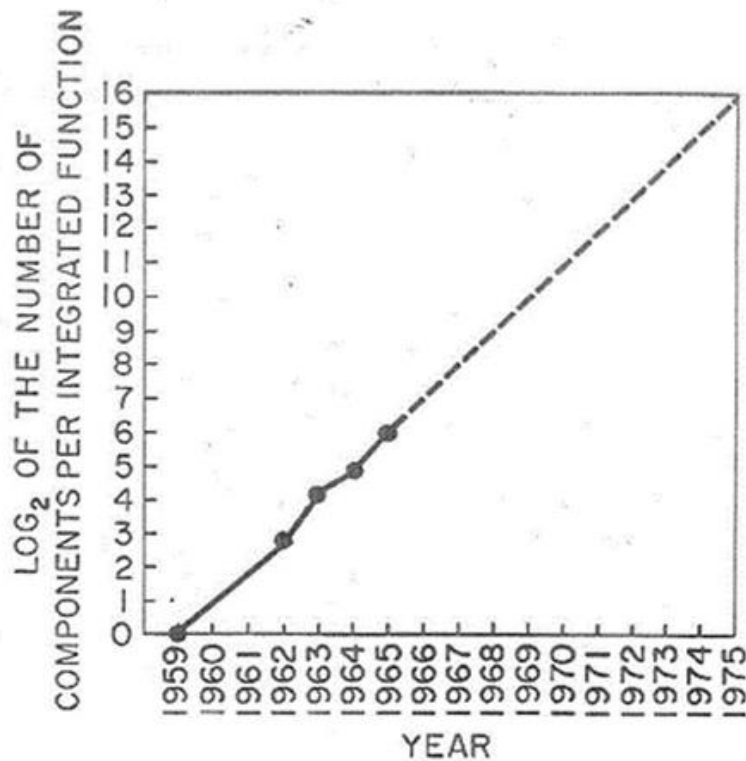
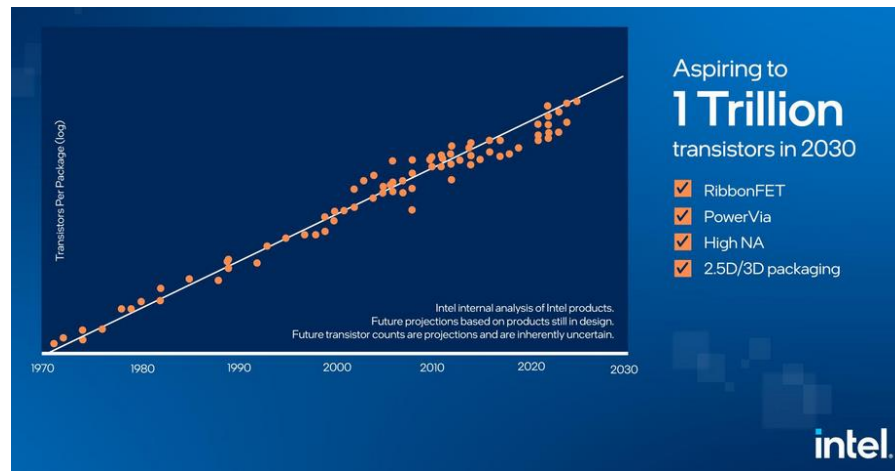
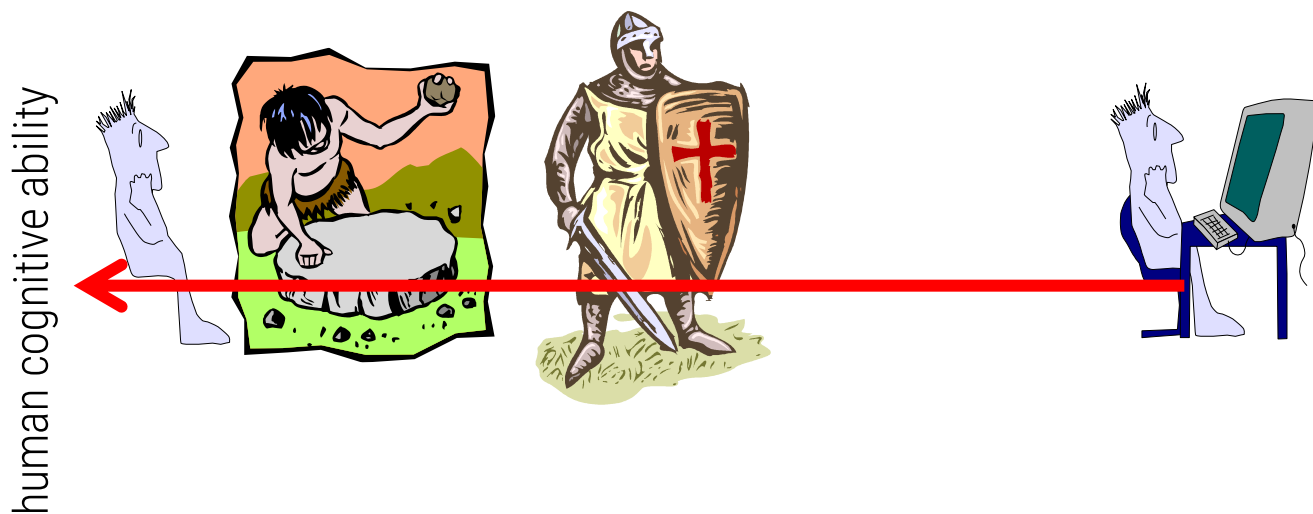


Fig. 2 Number of components per integrated function for minimum cost per component extrapolated vs time.



PEOPLE STAY ~THE SAME ...



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

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

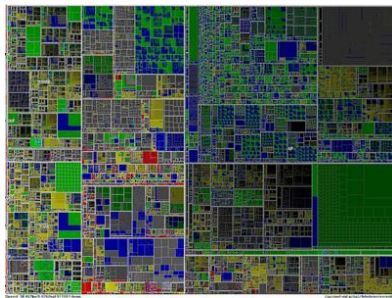


Sony SmartWatch

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

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.