



Universiteit Leiden

# The all pervasive impact of Big Data in daily life

From Causality to Correlation



H.Jaap van den Herik<sup>1,2,3</sup>

Rob van Eijk<sup>2,3</sup>

(1) Leiden University, LCDS

(2) Leiden University, eLaw

(3) Leiden University, CRK

Launch Future Center – Smart Systems

Leiden Centre of Data Science

The Hague, 23 April 2014

Sessie 13.30-14.00



# Acknowledgements

With much pleasure I would like  
to thank Hans Konstapel and Jeroen van der Leijé.

Research is team work over a long period. Thus, the list of collaborators  
would be long. Below I provide a selection. So, I would like to  
acknowledge for their help and inspiration:

Stan Bentvelsen, Yvo van Schulpen, Jos Vermaseren, Aske Plaat,  
Ben Ruijl, Peter de Kock, Ron Boelsma, Roel in 't Veld,  
Joke Hellemons and Eric Postma



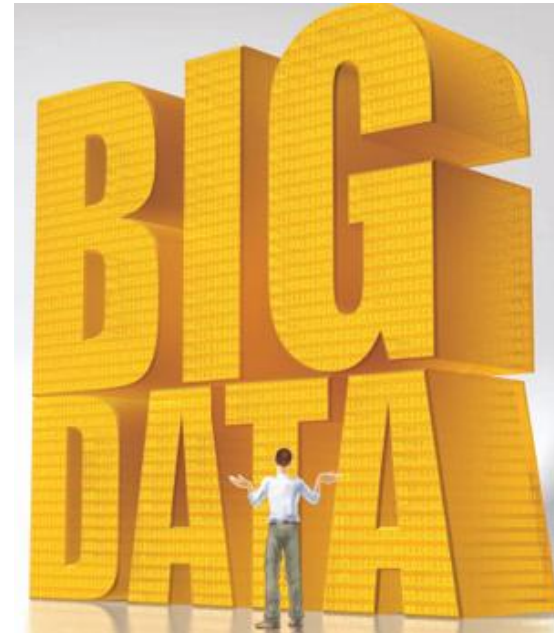
# Contents

1. What is Big Data?
2. Where do we find Big Data?
3. Role of Big Data
4. Big Data in action: NOS Zomercolumn
5. Social Innovation
6. New developments
7. Narrative Science
8. From correlations to causation?
9. Leiden Centre of Data Science

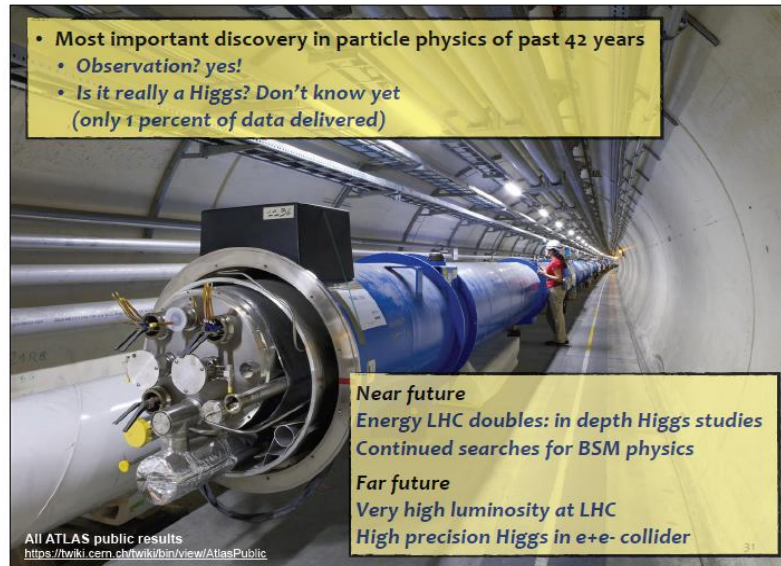


# BIG DATA

- Definition van Tom White (2012) :  
“Big Data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand databases management tools or traditional data processing applications.”
- The challenges capture:
  1. curation,
  2. storage,
  3. search,
  4. sharing,
  5. transfer,
  6. analysis,
  7. visualization,
  8. interpretation,
  9. real-time (van Eijk, 2013)



# The breakthrough of the century



Higgs particle





# Three perspectives on the development of Big Data

- (1) An overflow of Data Results  
(e.g., in particle physics)
- (2) A lack of coordination among the information items  
(as in the 9/11 events, Boston 15 April, 2013)
- (3) The power of Big Data  
(by using concepts as visualization and  
narrative science)





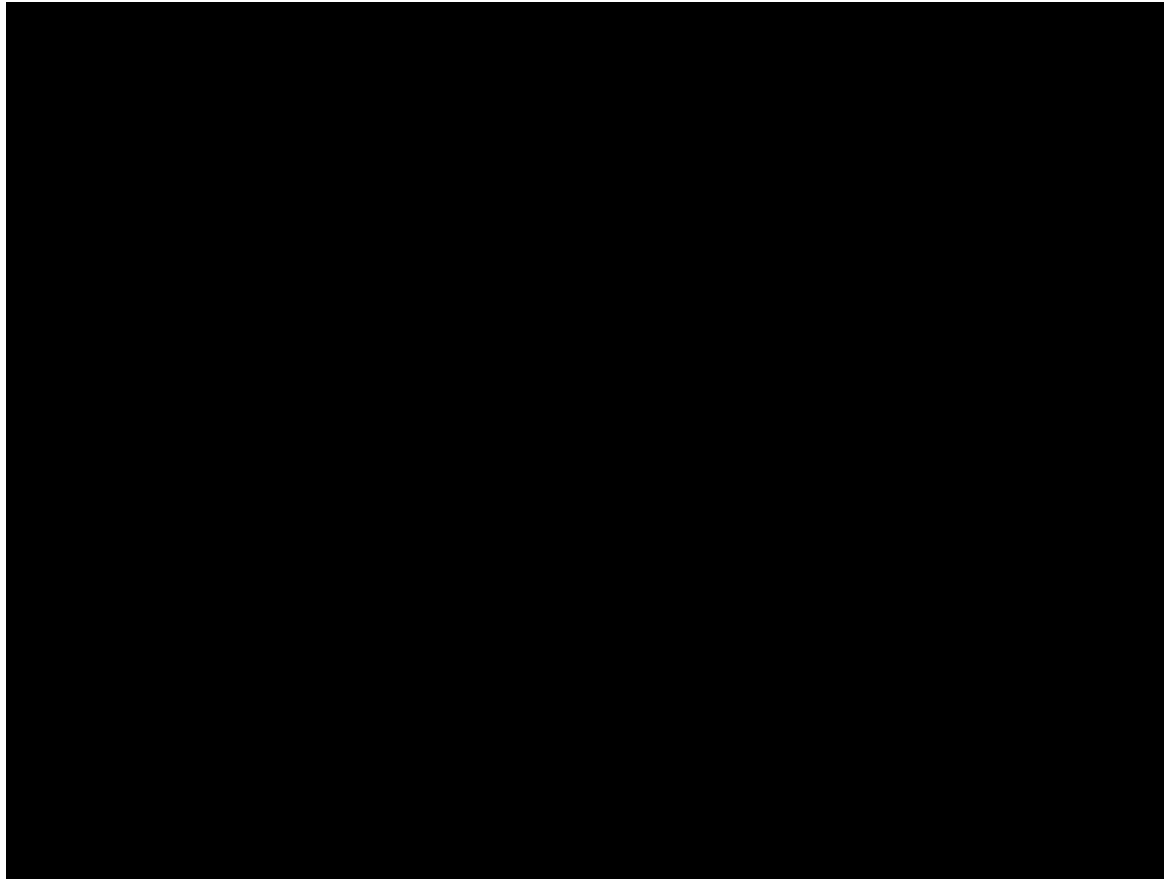
# The role of BIG DATA

- Social-economic Ph.D theses from 1970 to 2000 are frequently “outdated” by BIG DATA developments.
  - Deep Knowledge vs. Partial Knowledge
- Real-time bidding (RTB) happens in 30 milliseconds (0,030 sec.)





# NOS Journal interview met Rob van Eijk



<http://nos.nl/video/527311-handel-in-een-fractie-van-een-seconde.html>





## Ideas on INNOVATION

### Technological Innovation

Cloud

Crowd

Narrative Science

### Social Innovation

The new way of working

Communication via Social Media

Tracking & Tracing of the Individual

# From Social Change to Social Innovation

Two difficult social changes:

- To fill up without paying
- Cracking Thud
- Liability at the Oil Companies (not at the Police or Public Prosecutor)
- Liability at the Banks (not at the Police or Public Prosecutor)





# Social Innovation

Definition: Social Innovation is the development and implementation of novel ideas (products, services, and models) to satisfy social needs, and to create new social relations or partnerships.

## Seen as a process:

1. Identification of social needs
2. Invention of new solutions
3. Evaluation of effectiveness
4. Rescaling of effective social innovations



# Social Innovation EU Report

[http://ec.europa.eu/bepa/pdf/publications\\_pdf/social\\_innovation.pdf](http://ec.europa.eu/bepa/pdf/publications_pdf/social_innovation.pdf)

[http://ec.europa.eu/regional\\_policy/newsroom/detail.cfm?id=597&LAN=EN](http://ec.europa.eu/regional_policy/newsroom/detail.cfm?id=597&LAN=EN)



# Applications

1. Safety (politics, military)
2. Public Safety (Live View)      Example 1
3. Commerce (ads)
4. Banking (money streams)
5. Health care      Example 2
6. Judiciary (CODR)
7. Waterway transport      Example 3
8. Communication (twitter, phablet)
9. Education (MOOC)
10. Public governance
11. Warfare (Multi Agent Systems, Socio Cognitive Models)



MOOC = Massive Open Online Courses

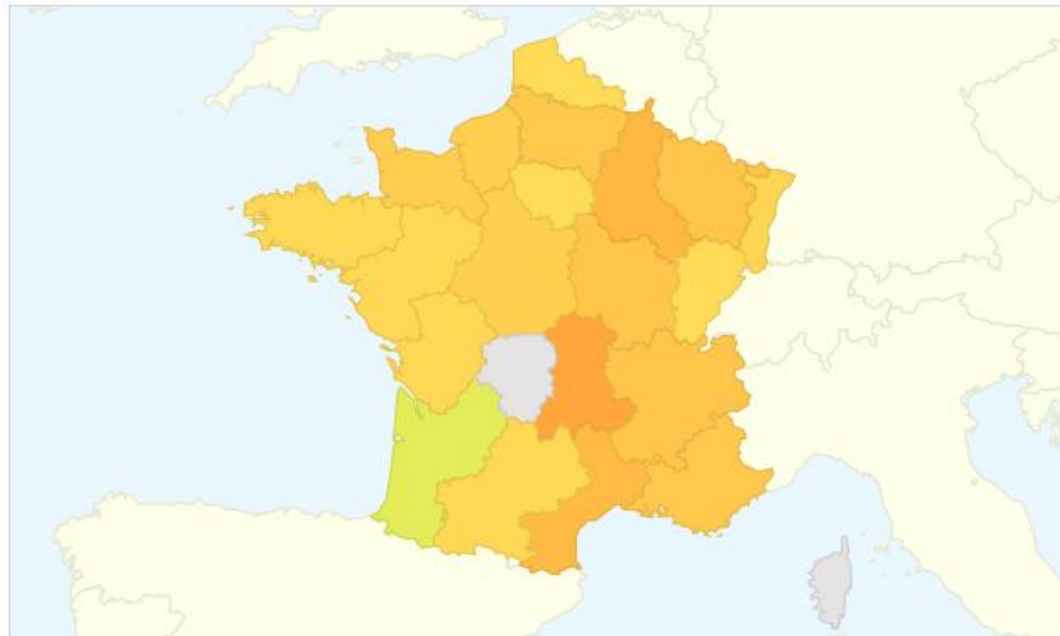
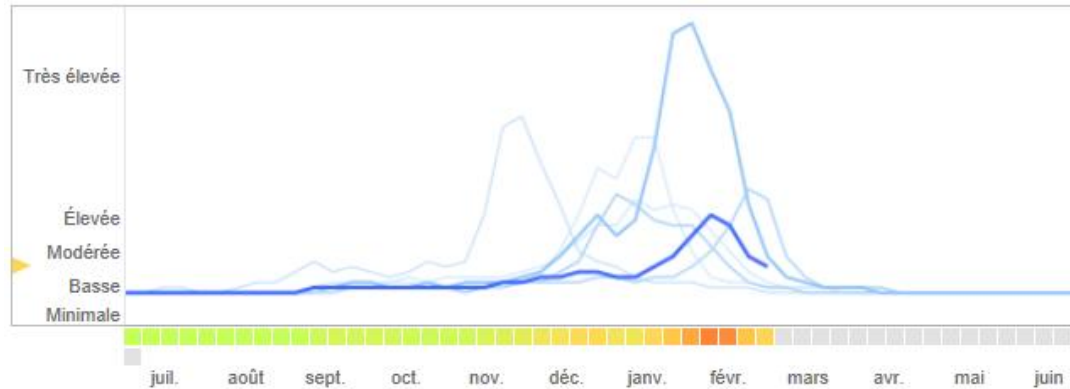


# Public Safety





## Flu prediction from correlation with Google search results







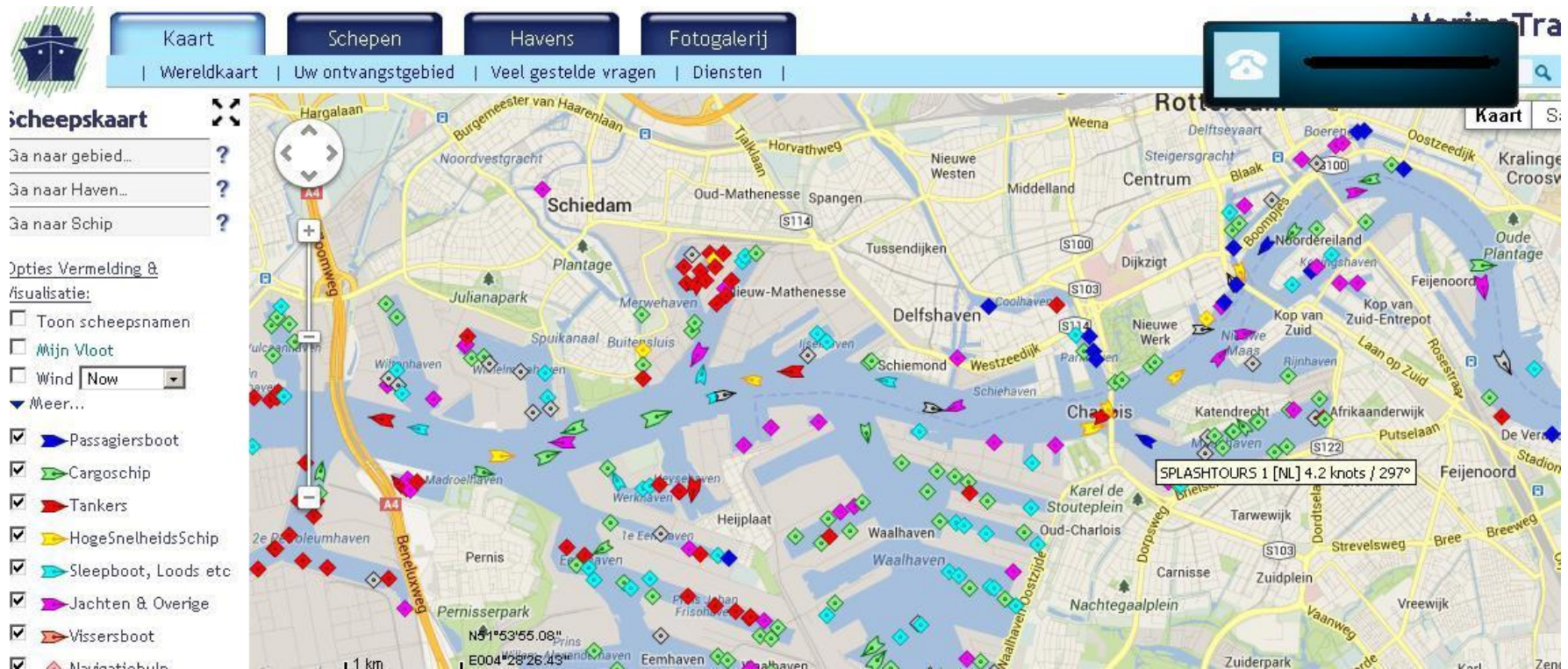
## Transport behavior by one vessel







# The places of all vessels at one moment





# New Developments

1. Computational Turn

2. Real-time Bidding  
(Sentiment mining)

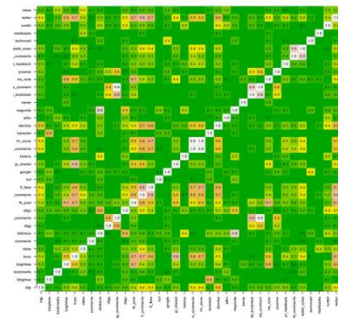
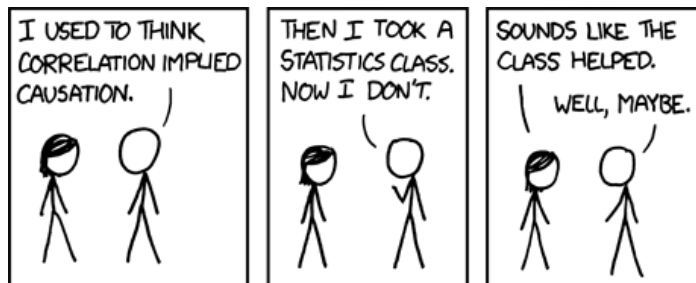
3. Narrative Science



# Computational Turn:

From causality to correlation

- Sampling is no longer at stake. Nowadays data from big populations (Twitter feeds, clicking behavior, Facebook data) are important.
- Insight into causal relations has lost its importance at many places.
- Correlation (what works well and what not) has taken over priority. This development is called Computational Turn.
- Computational Turn asks for reflection from economics, law, social sciences, behavioral sciences, and philosophical perspectives.





# Narrative Science

- BIG DATA:
- collection
  - awareness
  - usage



How did it happen that way?

- generation of data (collection)
- visualization of data (Napoleon)
- narrative science (which story is in BIG DATA?  
e.g., Wiki Leaks?)



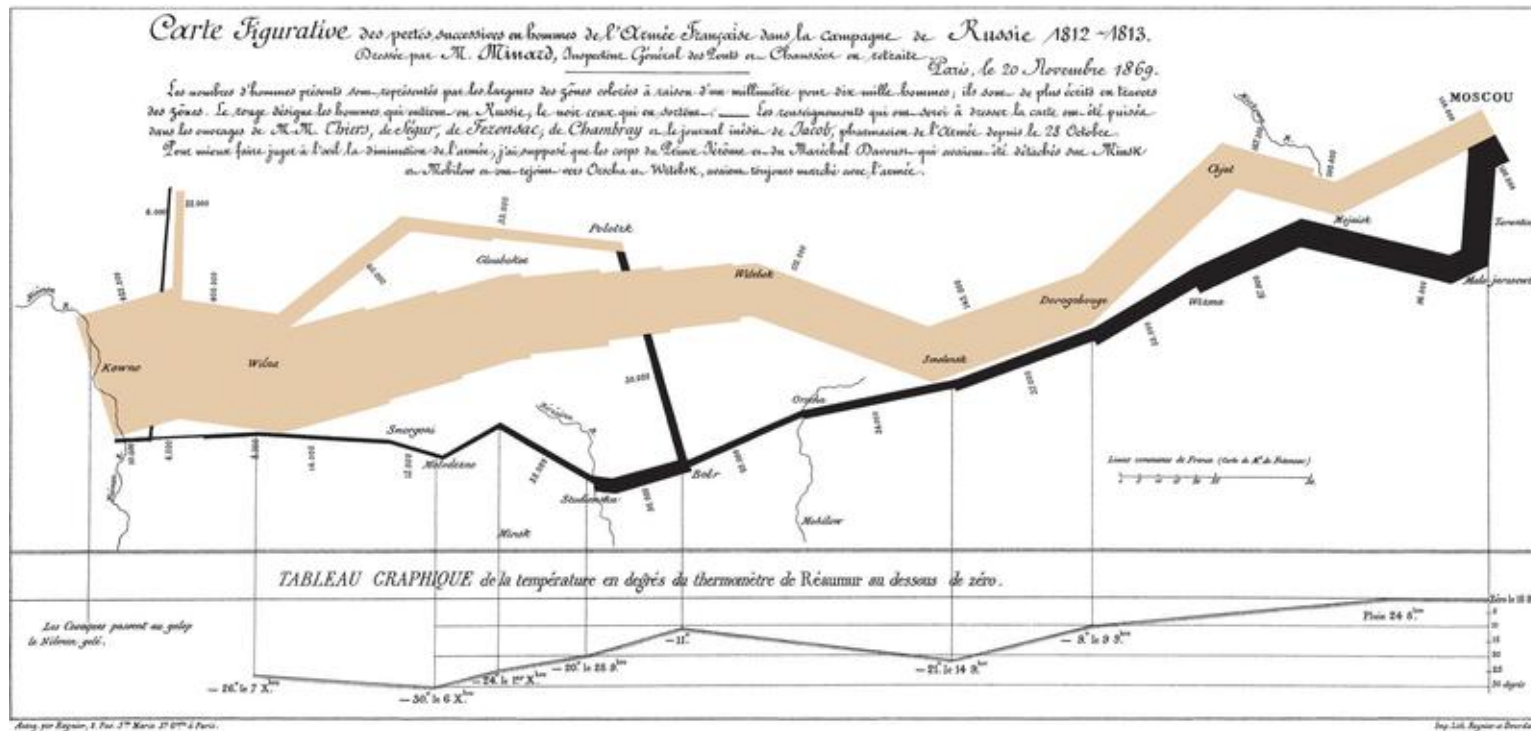
# Verdict collection

A “First” Collection by Ch.J. Enschedé

Collecting verdicts by the Dutch Courts is important for:

- Legal Certainty
- Equality of Treatment

# Napoleon







## Narrative science

- Finding the causations behind the correlations: make a story
- Examples:
  - Boston April 2013
  - Google Flu chart

Future for AI:

Reason about correlations to predict causations





# Leiden Data Science Center

- The new center will focus on multidisciplinary research
- Emphasis on data science: big data and small data
- We start with:
  - Bioscience
  - Human genome
  - Physics
  - Mathematics
  - Advanced computer science
  - Aviation
  - Law

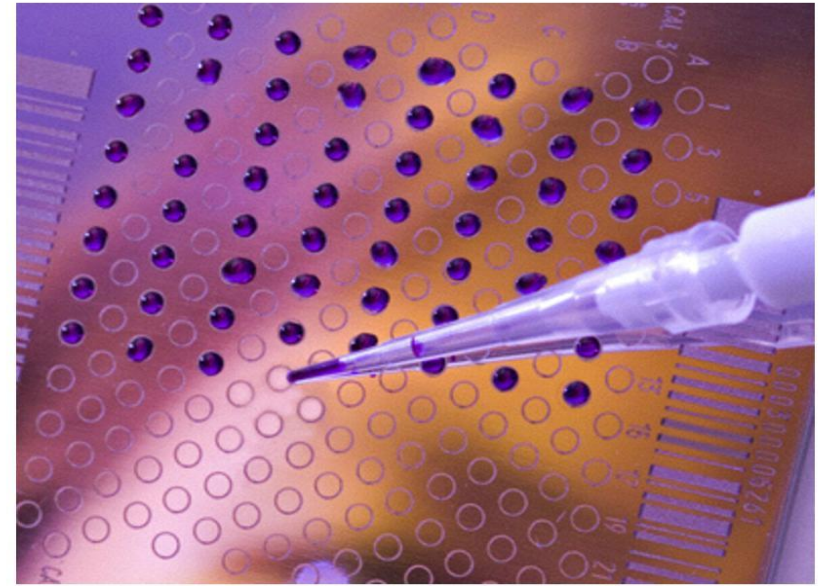
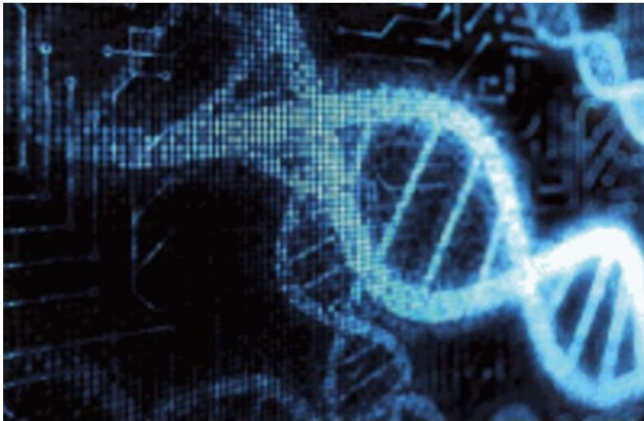


# Leiden Centre of Data Science Bio





# Leiden Centre of Data Science Life



DUTCH TECHCENTRE FOR LIFE SCIENCES



LEIDEN UNIVERSITY MEDICAL CENTER





## Theoretical physics



Higgs boson found at LHC  
4 July 2012



Leiden Centre of Data Science

# Multicriteria Optimization



LIACS  
Universiteit Leiden



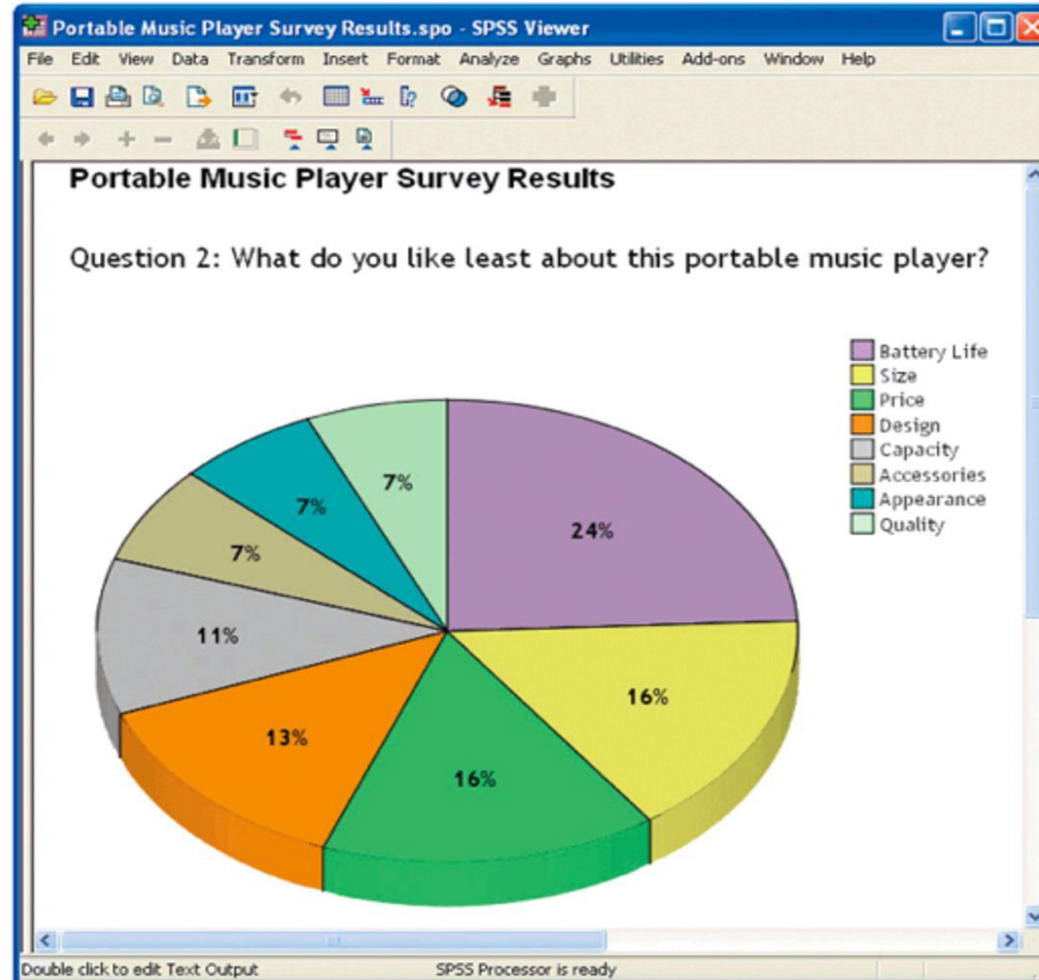
TATA  
TATA STEEL

monetdb





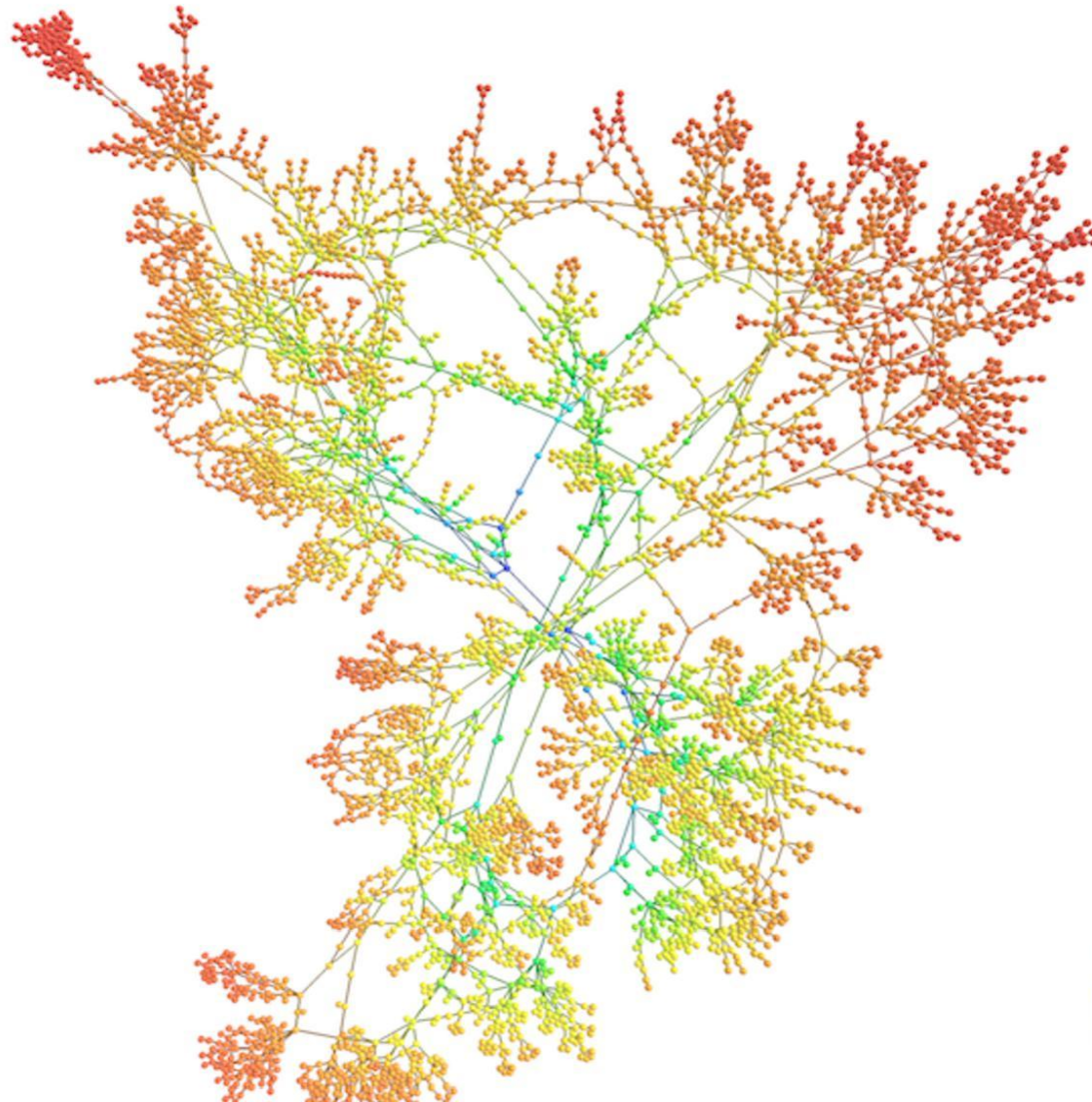
# Statistics: Categories





Leiden Centre of Data Science

# Random Graphs

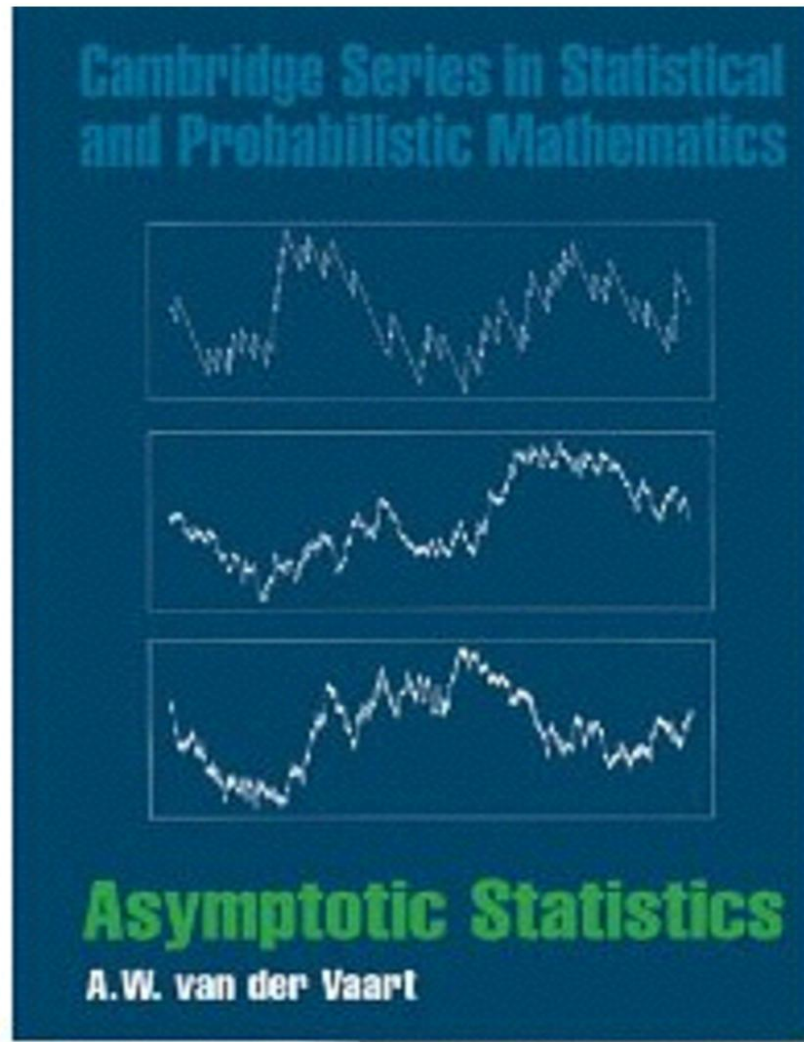






Leiden Centre of Data Science

# Asymptotic Statistics







Leiden Centre of Data Science

# Aviation







Leiden Centre of Data Science

# Legal





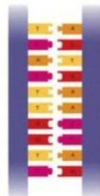
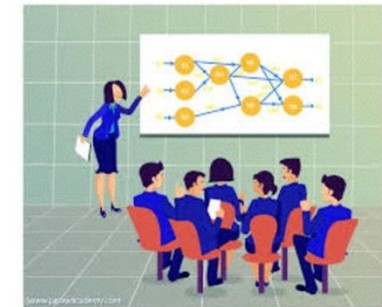
Leiden Centre of Data Science

# Activities

Education



Research



Business Optimization





Leiden Centre of Data Science

# Future Plans

- Grow “Virtual Lab”
- Increase Collaborations with interested partners

