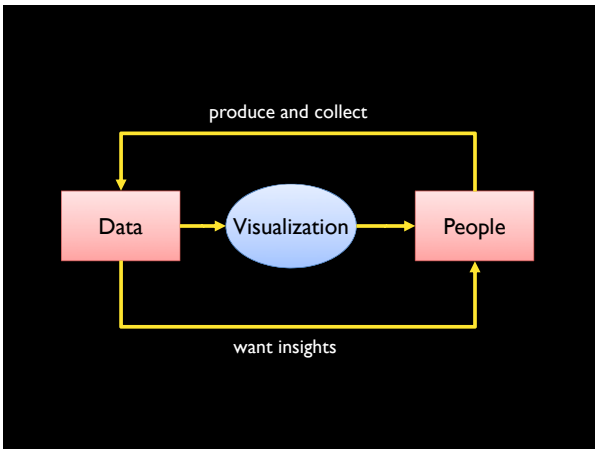
Visualization for Transparency

Jack van Wijk

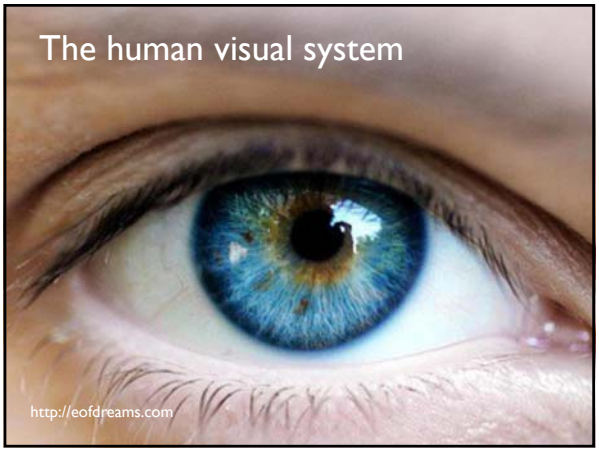
Responsible Data Science Seminar
Amsterdam
December 14th, 2016

TU/e Technische Universiteit Eindhoven University of Technology

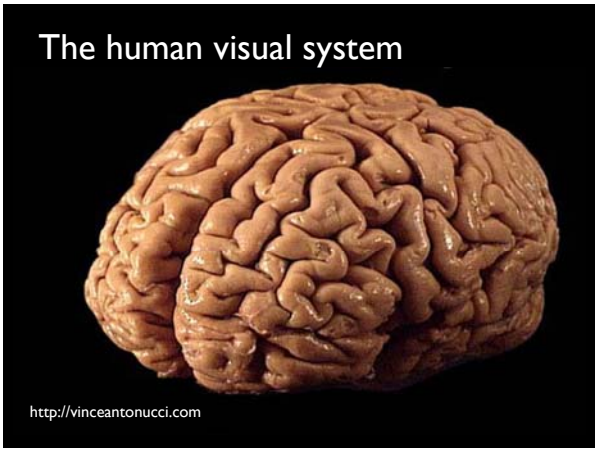
Where innovation starts



The human visual system



The human visual system

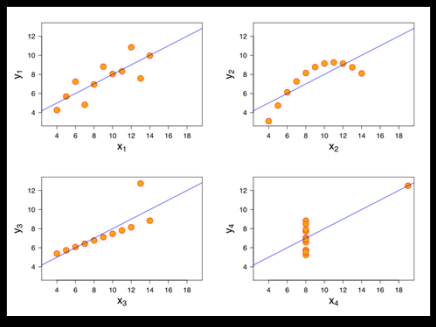


Anscombe's quartet

I		II		III		IV	
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

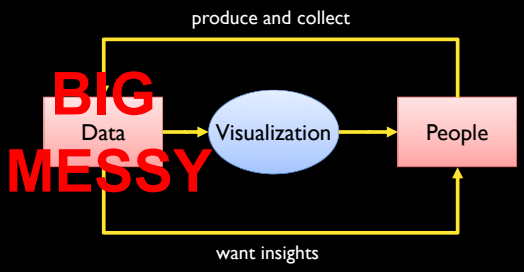
Four data-sets: same average x and y, same variance, same correlation, …

Anscombe's quartet



Francis Anscombe, 1973

Visualization

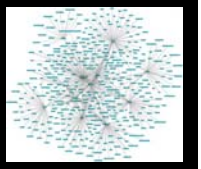


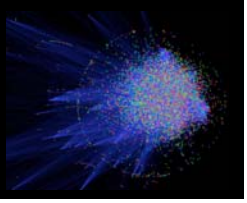
Standard diagrams don't scale

Multivariate visualization:
scatterplot

Tree visualization:
tree diagram

Graph visualization:
node link diagram





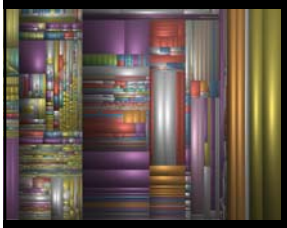
SequoiaView

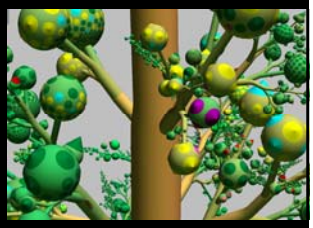
Van Wijk and Van de Wetering, 1999

Botanically inspired treevis

Kleiberg, Van de Wetering, van Wijk, 2001

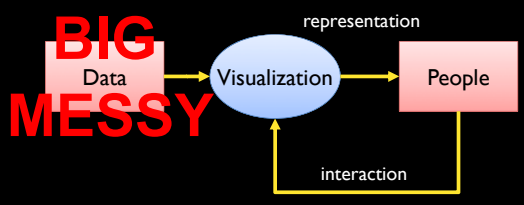
One picture is worth a lot of pixels





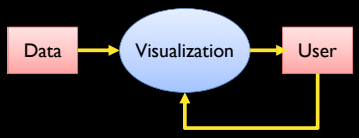
- Focus on visual presentation
- Show everything in one picture?

Visualization

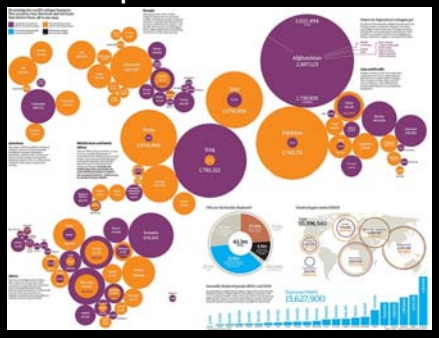


Information Visualization

- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition (*Card et al., 1999*)



InfoGraphic ≠ InfoVis

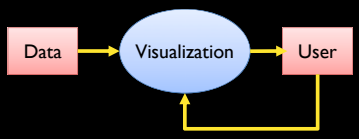


The Guardian, 2010

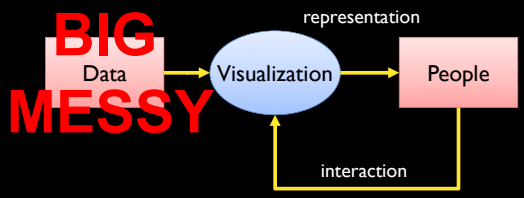
Information Visualization

- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition (*Card et al., 1999*)

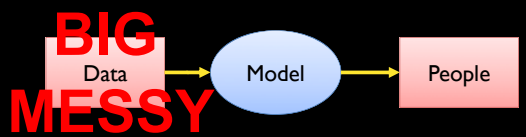
Visualization: exploration process, not a picture.



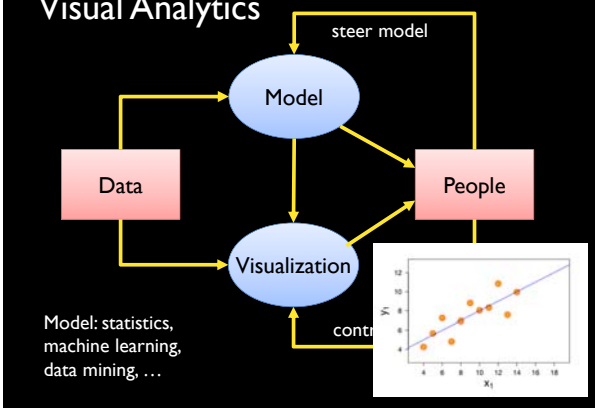
Visualization



Statistics, machine learning, data mining, …



Visual Analytics



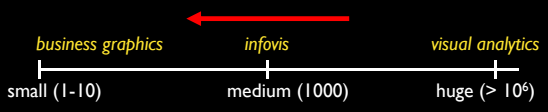
Visual Analytics

Computers
- fast
- precise
- computation
- search and store data
- graphics

Humans
- Flexible, inventive
- Solve problems
- Handle new situations
- Handle incomplete and/or inconsistent information
- Have domain knowledge and experience
- Can **see** things that are hard to compute

Visual Analytics: Synergy between computers and humans

Data size



Try to move to the left:
- Filter, aggregate, statistics, machine learning, … without loosing essential information

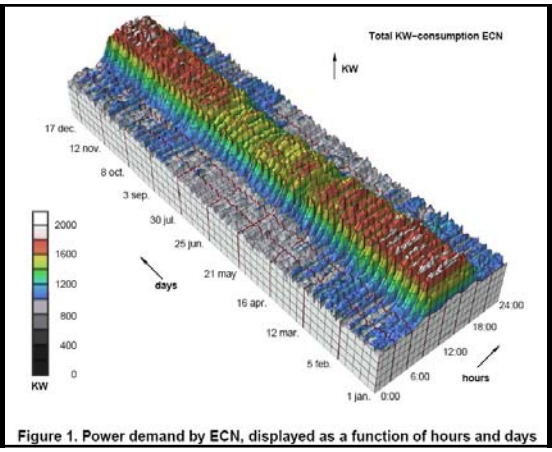
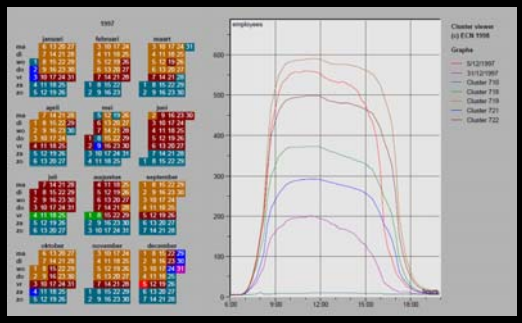


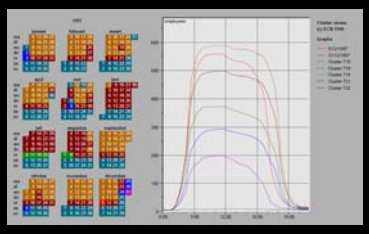
Figure 1. Power demand by ECN, displayed as a function of hours and days

Visual Analytics Time Series Data



Van Wijk and Van Selow, IEEE InfoVis 1999

The pattern

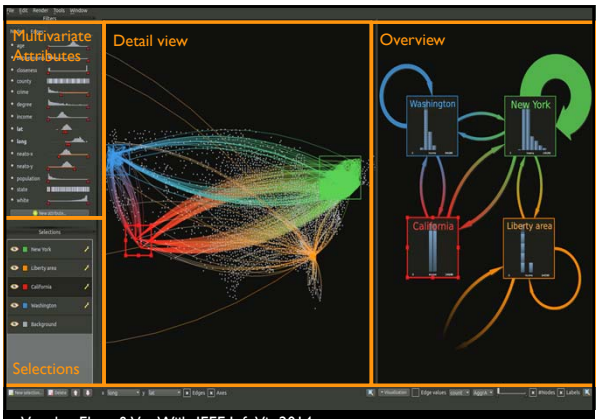


- combination of simple presentations,
- use interaction and statistics

Challenges

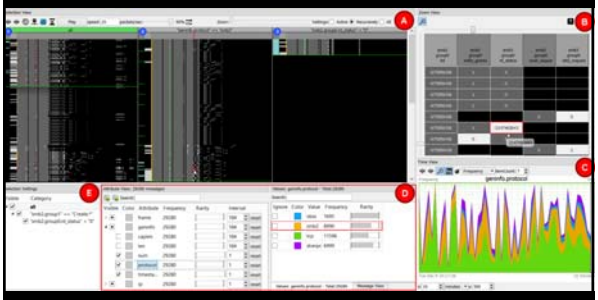
- Handling big, messy data
- Dealing with models

- Examples from Eindhoven
- Lessons learned, challenges ahead



Van den Elzen & Van Wijk, IEEE InfoVis 2014

Network protocol data



Cappers & Van Wijk, IEEE VizSec 2015

Image categorization

Van der Corput & Van Wijk, IEEE PacificVis, 2016

Dealing with big, messy data

Custom cases can be solved, using a combination of simple presentations, interaction and statistics

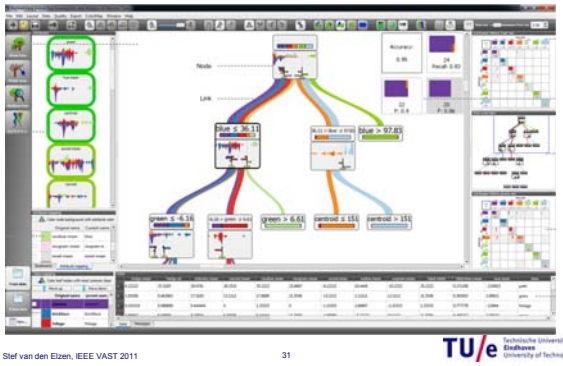
Generic challenges:

- How to scale up?
- How to generalize?

Dealing with models

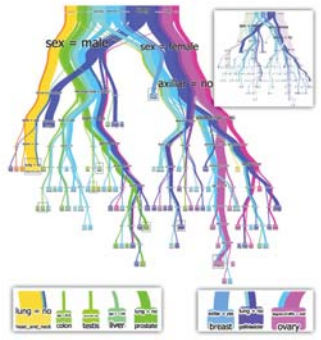
- How to define ML/DM models?
  - What model, what parameter settings?
  - Which features of the data?
- How to understand the result?
- *How to make the black box transparent?*

BaobabView



Stef van den Elzen, IEEE VAST 2011 31 TU/e Technische Universiteit Eindhoven University of Technology

BaobabView



Decision tree
tumor location

32

TU/e

EuroVis 2015
The EG / VGTC Conference on Visualization

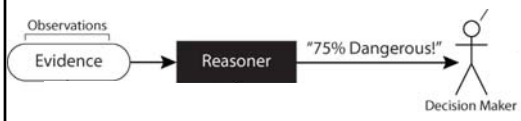
25 - 29 May
Cagliari / Sardinia / Italy

Rationale Visualization for Safety and Security

Roeland Scheepens
Steffen Michels
Huub van de Wetering
Jarke J. van Wijk

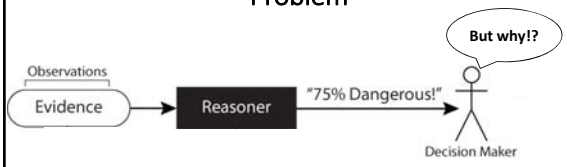
TU/e

Context



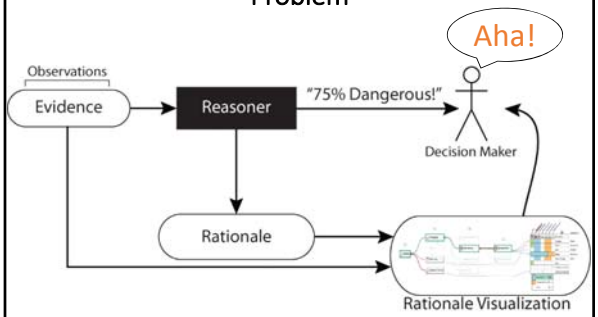
AIS-data, radar data, web data, reports… on vessels → *Probabilistic first order logic inference engine* → *Coast guard*

Problem

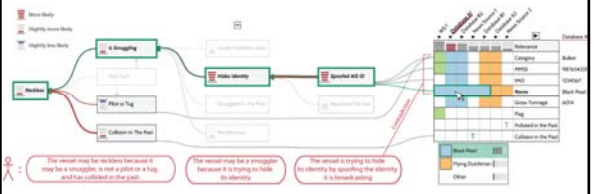


AIS-data, radar data, web data, reports… on vessels → *Probabilistic first order logic inference engine* → *Coast guard*

Problem



Example



COMMIT/ A public- private research community
THALES
Embedded Systems Innovation BY TNO
TU/e

How to understand a model?

Custom cases can be solved, using a combination of simple presentations, interaction and statistics

Challenges:
- How to scale?
- How to generalize?
- Can we understand neural networks…?

Conclusions

Data and models can be made transparent, using a combination of simple presentations, interaction and statistics

and with smart students and a lot of effort!

Challenges:
- How to scale?
- How to generalize?
- *How to develop custom solutions efficiently?*

Thank you!