Information Visualisation (1) Introduction to Infovis

Prof. Bruno Dumas

UNIVERSITÉ DE NAMUR

Why Study Information Visualisation in a Data Science Program?

Information Overload

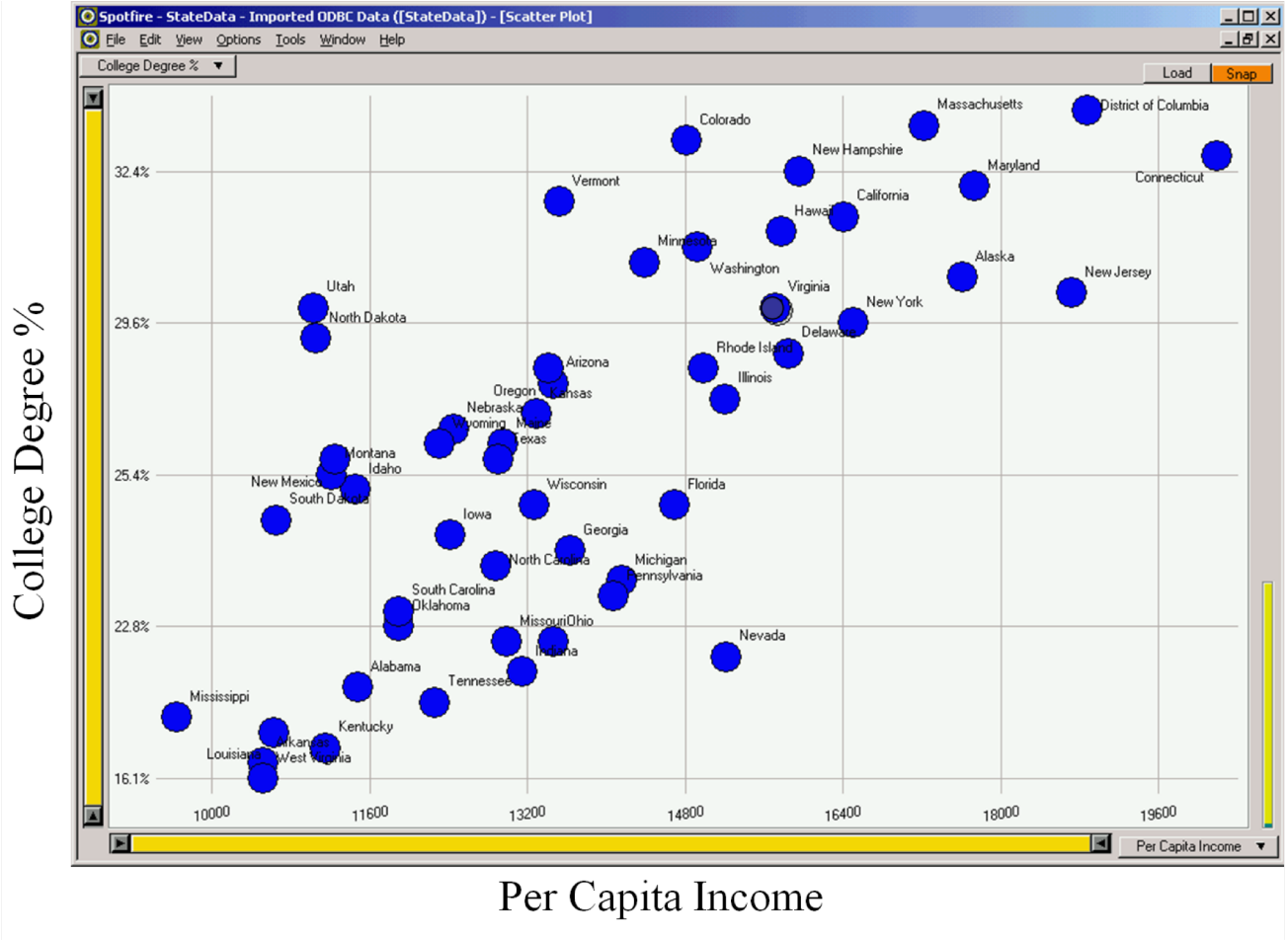
- Between 1 and 2 exabytes of unique info produced per year
  - 1000000000000000000 (1018) bytes
  - 250 Mo for every man, woman and child
- How to make use of the data?
- How do we make sense of the data?
- How do we employ this data in decision making processes?
- How do we avoid being overwhelmed?

Human Vision

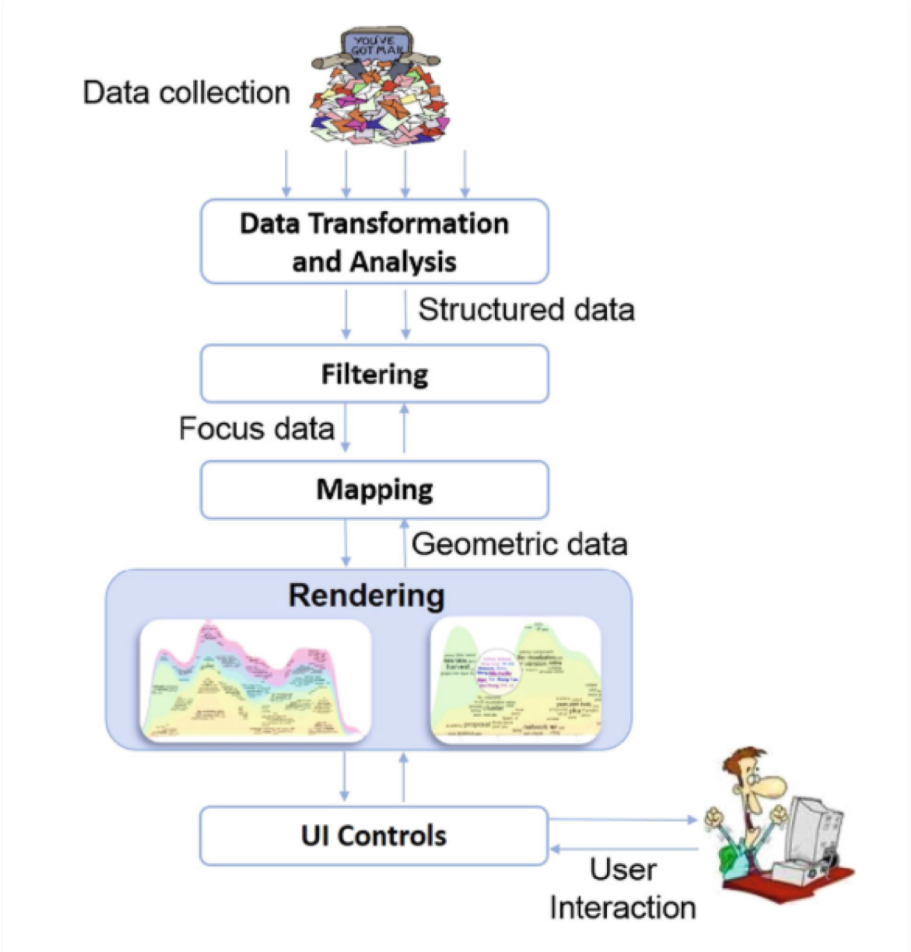
- Highest bandwidth sense
  - Fast, parallel
  - Pattern recognition
  - Extends memory and cognitive capacity
  - People think visually
- The Challenge
  - Transform the data into information (understanding, insight) thus making it useful to people
- Example:
  - Which state has the highest income?
  - Is there a relationship between income and education?

State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	24.1%	14154
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923
West Virginia	16.1%	10520
Wisconsin	24.9%	13276
Wyoming	25.7%	12311

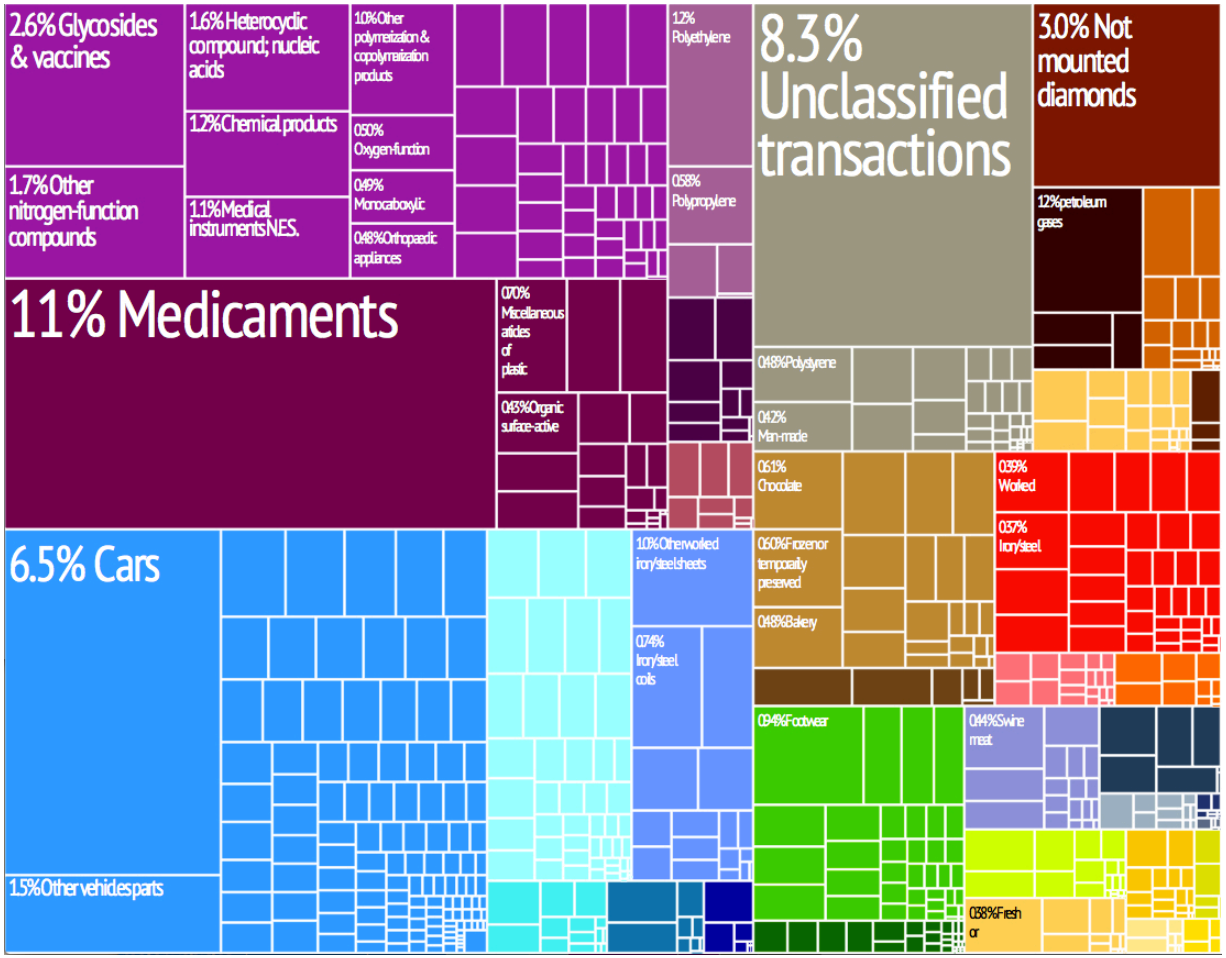
Human Vision (cont.)



Information Visualisation Pipeline

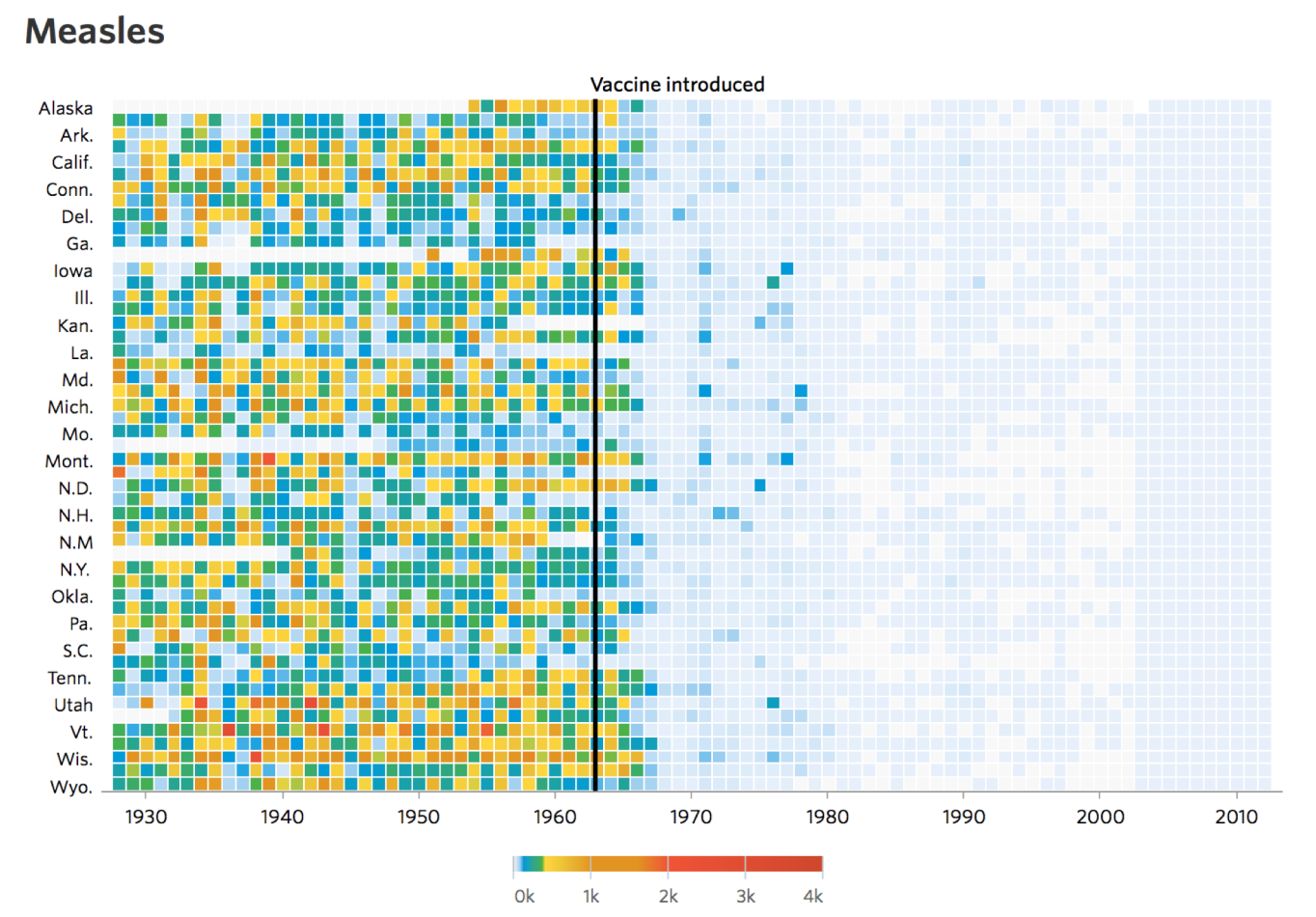


Use Case: 2009 Exports of Belgium



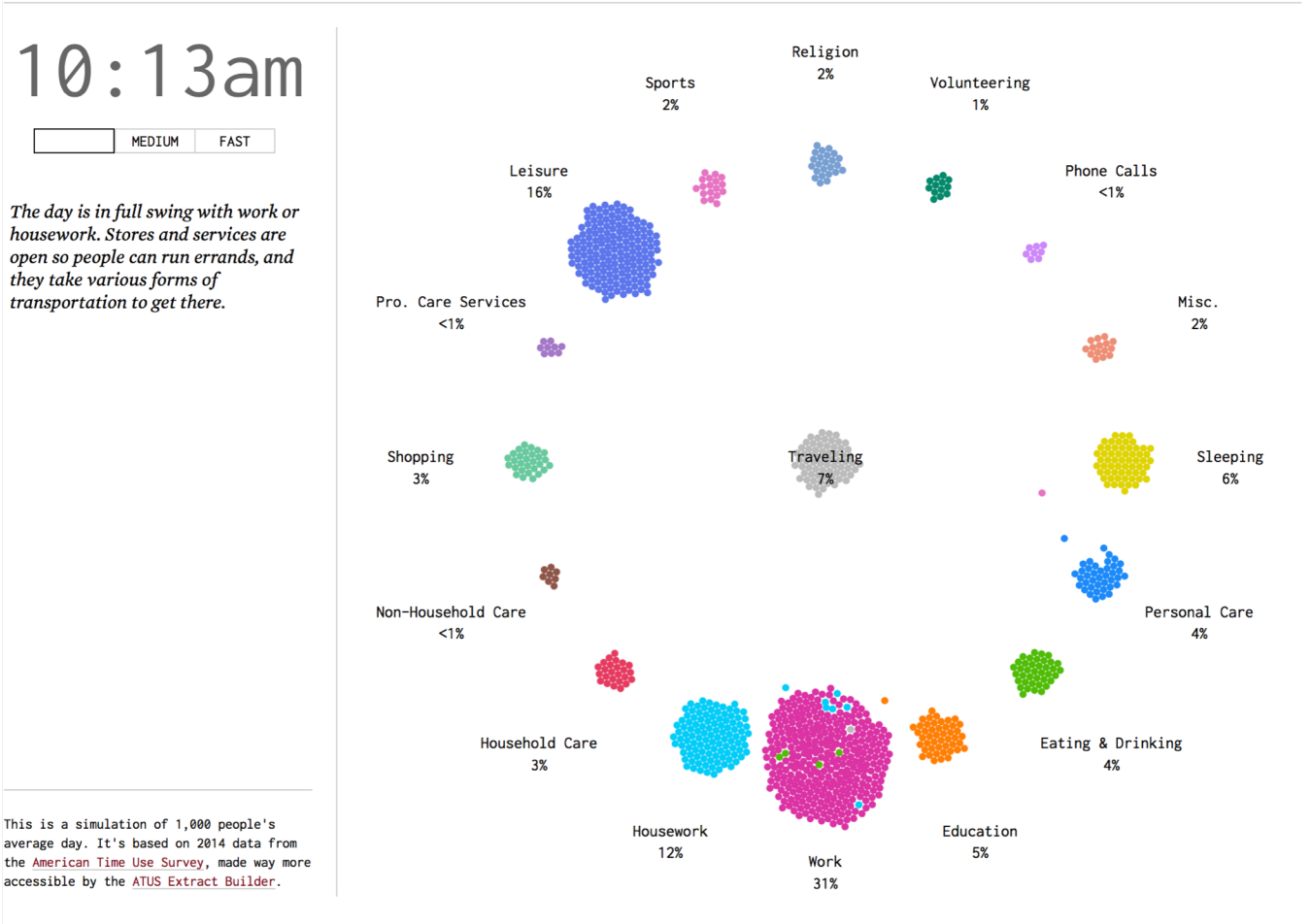
https://commons.wikimedia.org/wiki/File:Tree_map_export_2009_Belgium.jpeg

Use Case: WSJ Infectious Diseases and Vaccines



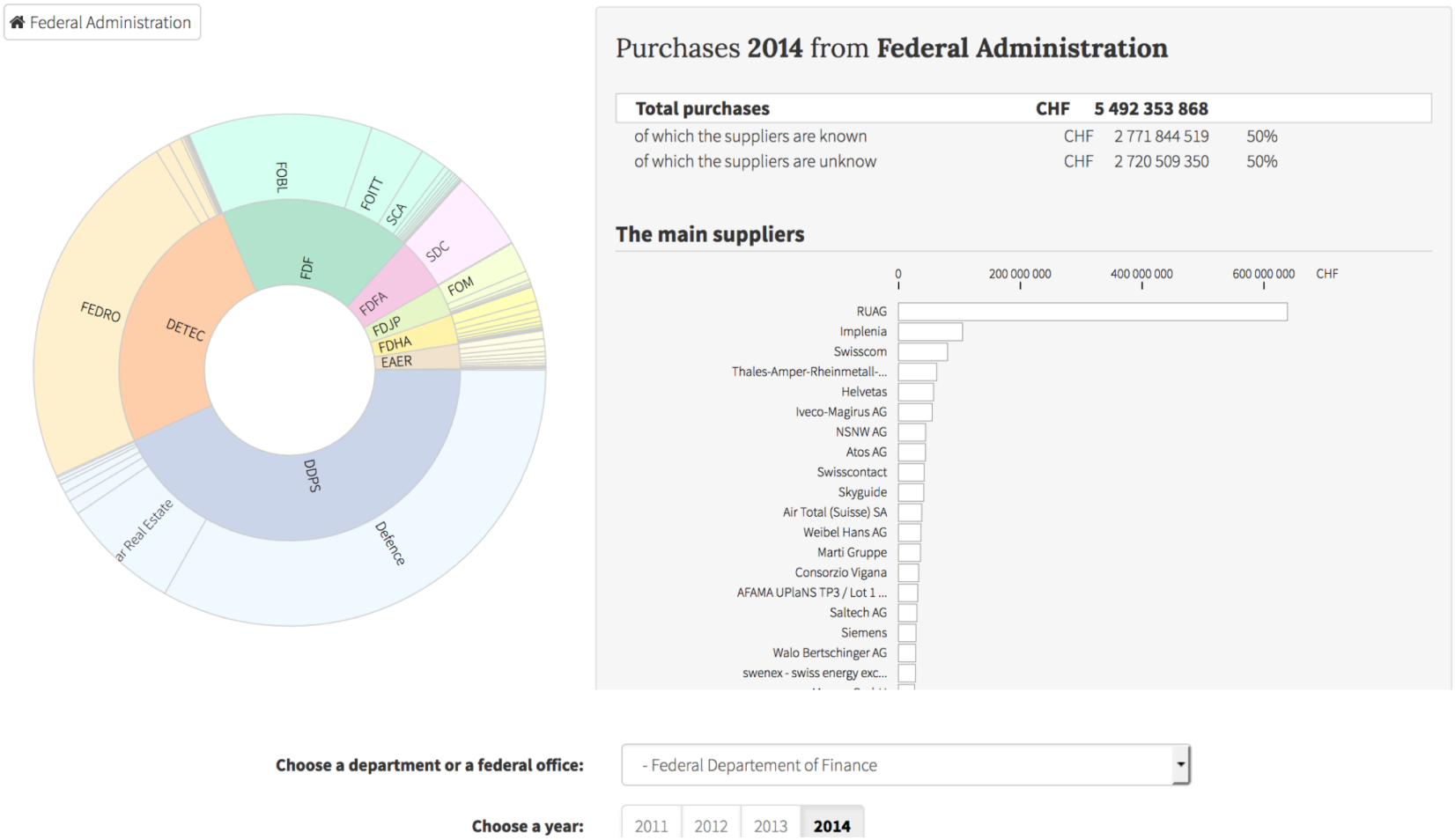
http://graphics.wsj.com/infectious-diseases-and-vaccines/

Use Case: A Day In the Life of Americans



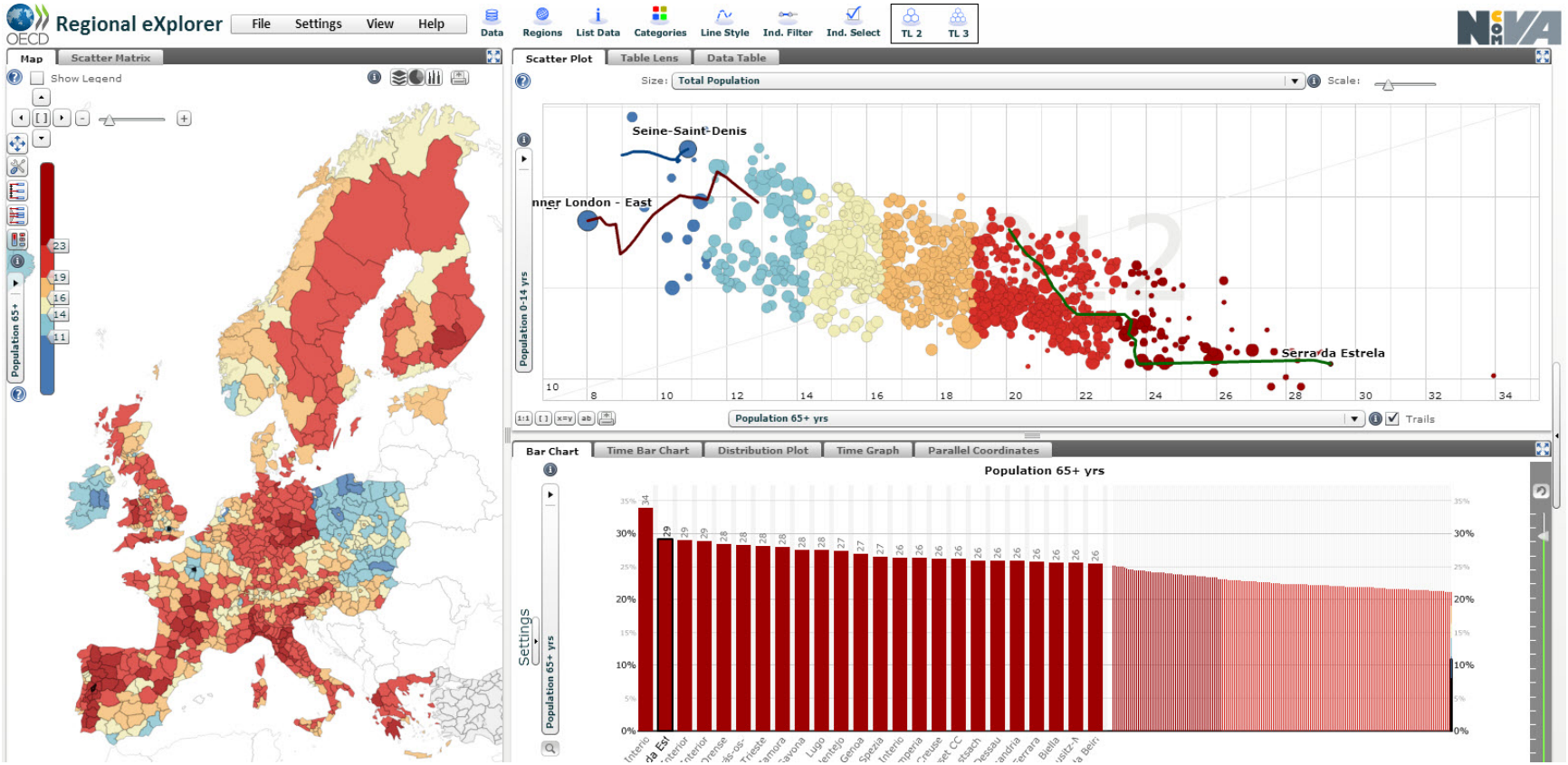
http://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/

Use Case: Purchases in 2014 from Swiss Federal Administration



http://enquete.lematindimanche.ch/interactif/achats/indexen.html?lang=en

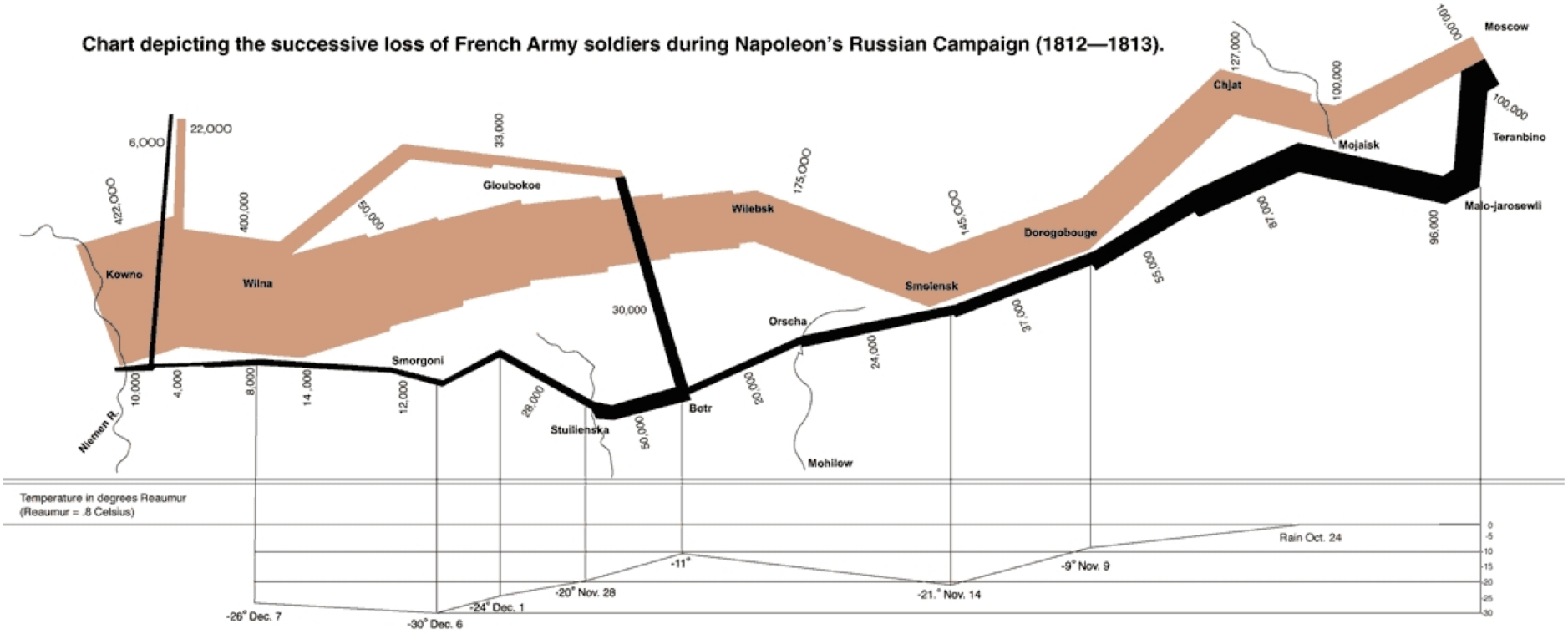
A Needed Practice



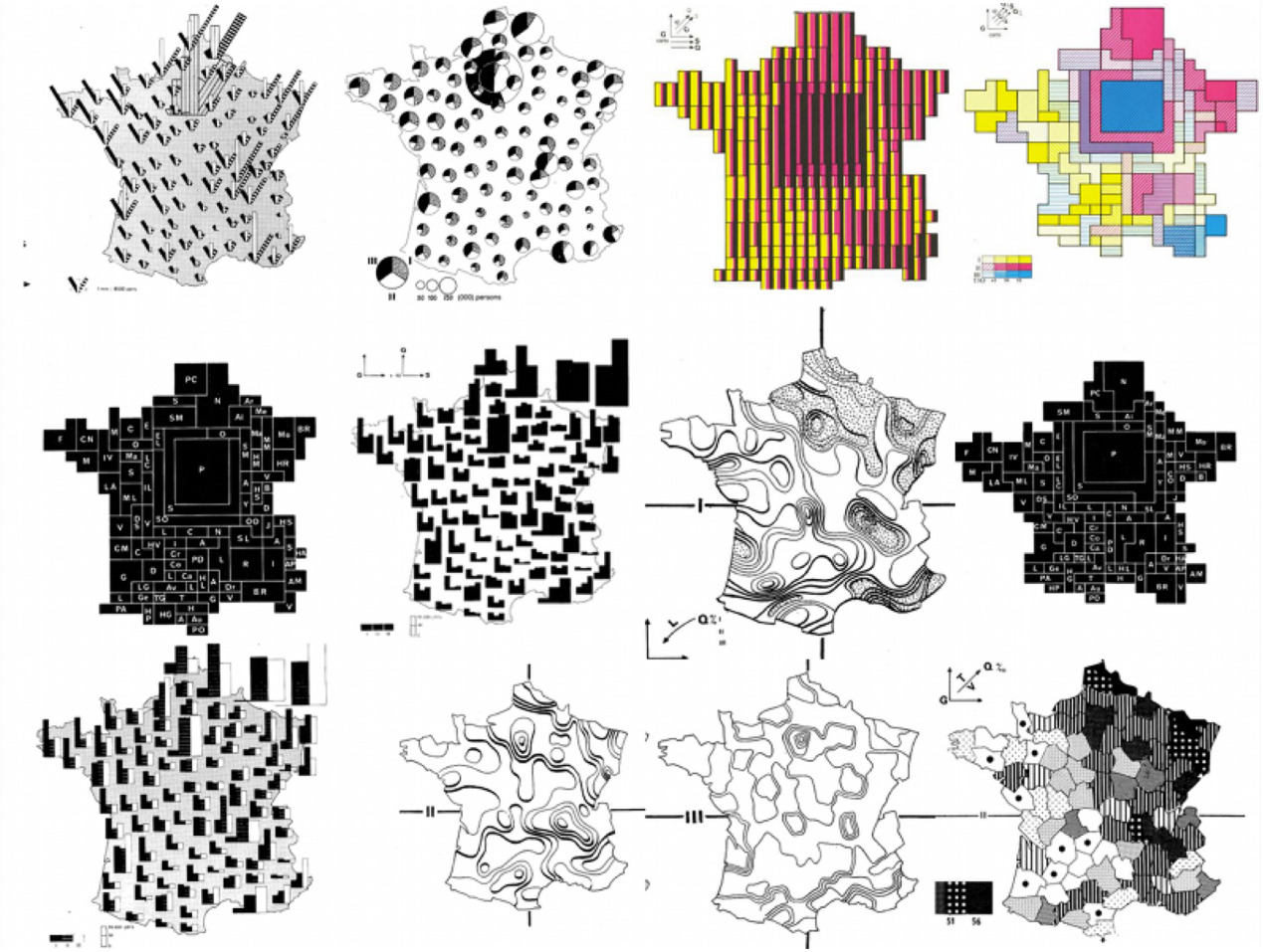
A Needed Practice

A Needed – and Old – Practice

- This map shows the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path (Charles Minard, 1869)



A Needed – and Old – Practice



Jacques Bertin, Sémiologie graphique, 1983

Characteristics of Information Visualisation

- Human in the loop
- Computer in the loop
- External representation
- Depend on vision
- Can show the data in detail (compared to statistics)
- Interactive

Characteristics: Human in the Loop

- Vis provides analysis of data when you don't know what question to ask in advance
  - … Compared to statistics or machine learning
- (Good) vis systems augment human capabilities
- Vis tools can be used in many stages
  - To gain a clearer understanding of analysis requirements before developing models
  - To refine, analyse, monitor, debug or extend an existing computational solution
  - For exploratory analysis, to help users generate and check hypotheses
  - For presentation, to explain something that you already know

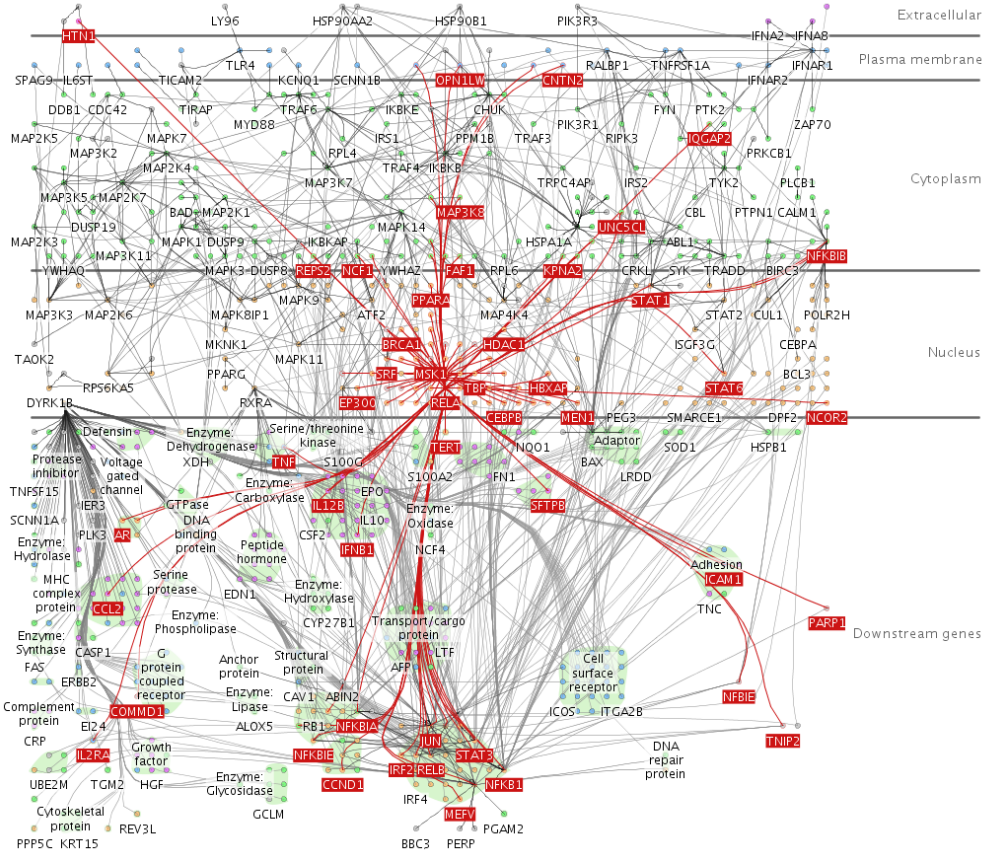
Characteristics: Human in the Loop – Roles of Infovis

- Communication
  - Prerequisite: confirmed hypothesis
  - Outcome: clear visualisation
- Confirmation
  - Prerequisite: hypothesis
  - Outcome: confirmation/rejection (new hypothesis)
- Exploration
  - Prerequisite: domain knowledge
  - Outcome: new hypothesis
- Visual Analytics' Motto: *detect the expected and discover the unexpected*

Keim, Daniel A. "Information visualization and visual data mining." IEEE transactions on Visualization and Computer Graphics 8, no. 1 (2002): 1-8.

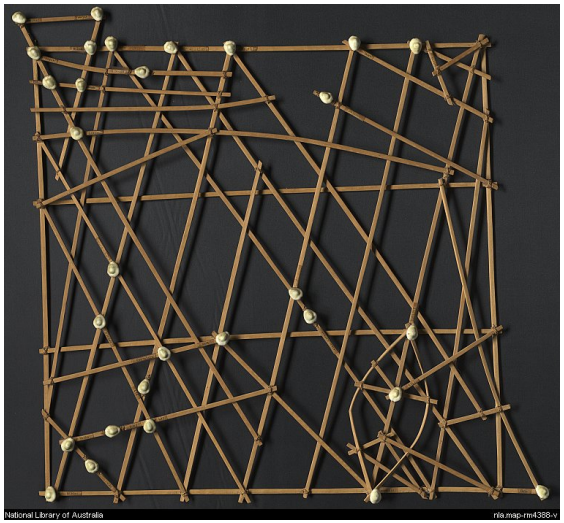
Characteristics: Computer in the Loop

- Advantages compared to a representation drawn by hand:
  - Generation of automatic representations
  - Can follow dynamically updated datasets
  - Interactivity (more on this later)



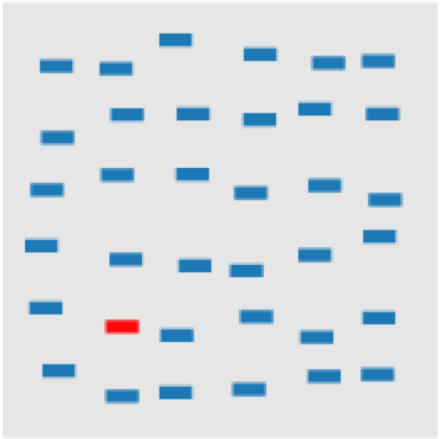
Characteristics: External Representation

- Augmentation of human capacity
- Offload internal cognition
  - Use of multiple senses in parallel
  - Concept of « external memory »
- External representations can take many forms, not only diagrams
  - Physical representations for example…
  - Field of « physicalisation », not mentioned further in this class



Characteristics: Depends on Vision

- Very high-bandwidth channel
- Parallel processing of a significant amount of visual information
  - E.g. « visual popout » (red item in the middle of blue ones)
  - MHP processor model describes this well
- However, only a tiny portion of what we see is perceived at a high resolution



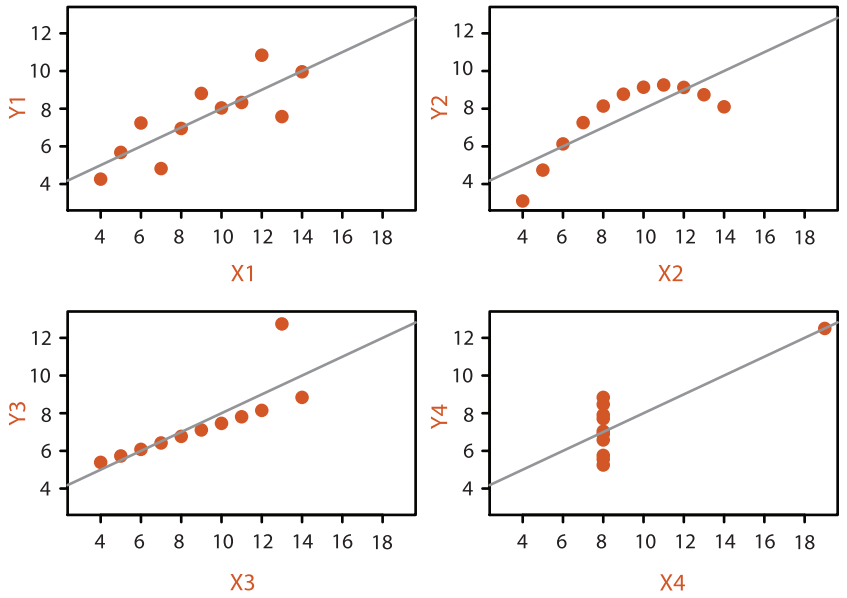
Characteristics: Can Show Data in Detail

- Visualisation helps when needing to see a dataset structure in detail instead of only a brief summary of it
- Typical situations
  - Find unexpected patterns
  - Confirm expected patterns
  - Assess the validity of a statistical model
  - Judge whether a model fits data
  - …
- Machine learning works well when you know what question to ask and data is clearly structured and with proper semantics
- Statistics have limitations because they summarize
  - Example: Anscombe's quartet

Limits of Statistics: an Illustration (Anscombe's Quartet)

Anscombe's Quartet: Raw Data

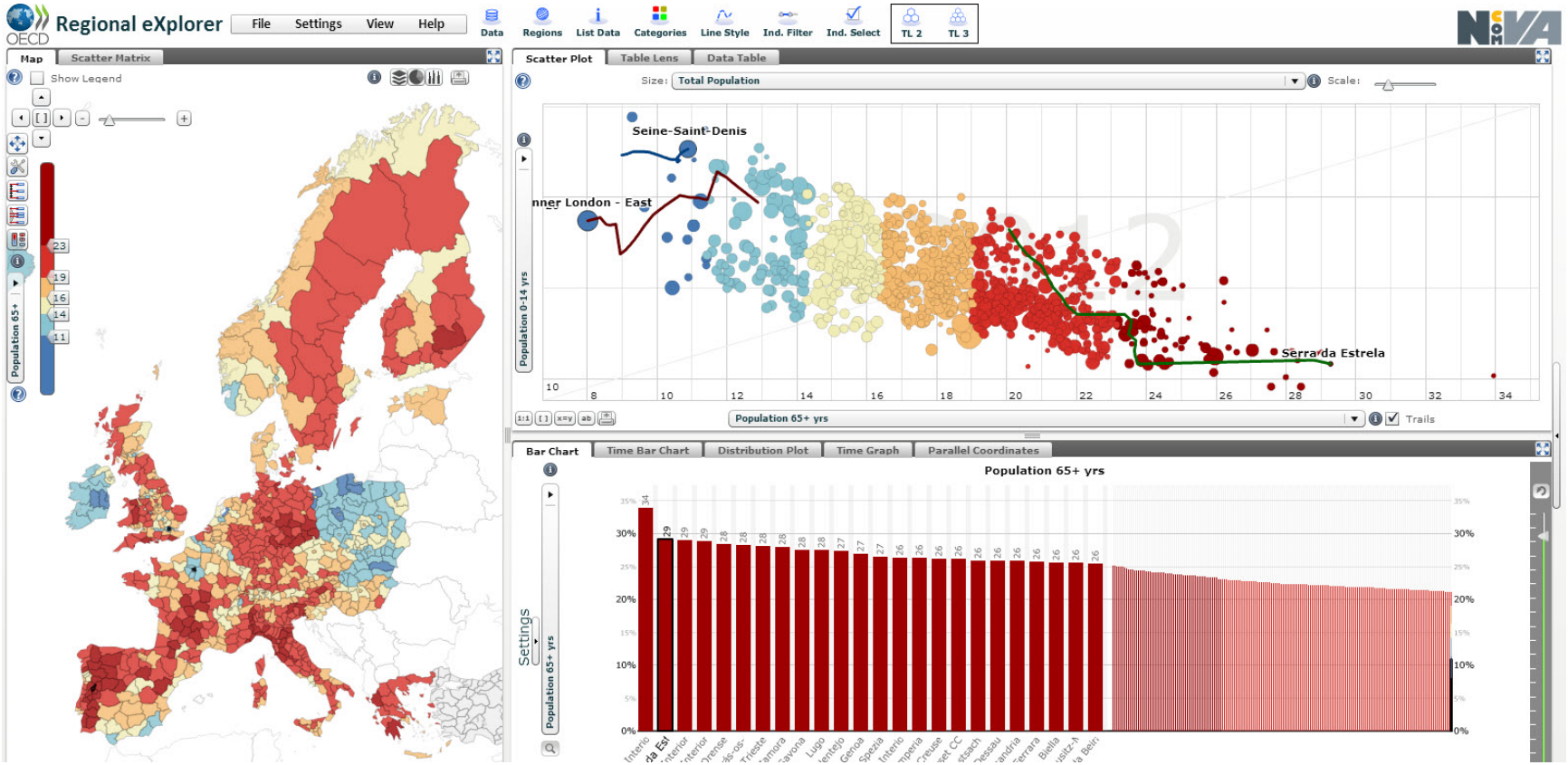
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



Characteristics: Interactivity

- Interactivity is crucial for handling complexity
  - Typically when data is large enough…
- Interactivity involves
  - Changing representations
  - Zooming
  - Filtering
  - Getting details
  - Summarizing
  - Combining views
  - …

Linking and Brushing Example

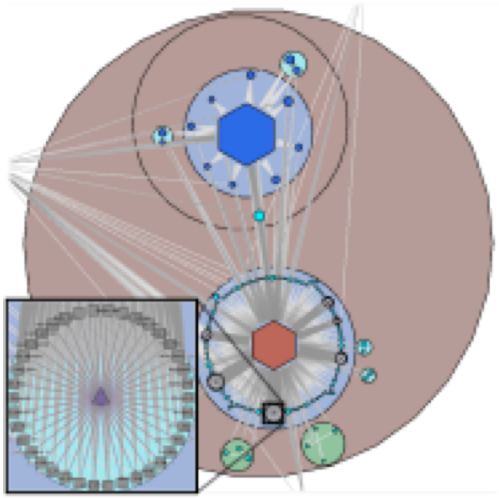


Challenges When Designing Visualisations

- The Vis idiom design space is huge
- Tasks need to be well defined
- A visualisation needs to be effective
- Most designs are ineffective…
- Validation is difficult
- Resources are limited

Challenges: the Vis Idiom Space is Huge

- Vis Idiom: “a distinct approach to creating and manipulating visual representations”
  - Or: a visual encoding of data as a single picture.
- Lots of different vis idioms for a given data, gets even bigger when one considers interactivity…
- Goal: select the right vis idiom(s) considering the data and tasks of the user



Challenges: Tasks Need to Be Well Defined

- Data is not everything, select vis idioms must also correspond well to users' tasks
- A selection of (generic) tasks a vis tool can support:
  - Presentation
  - Discovery
  - Generating new hypotheses
  - Exploring unknown datasets
  - Confirming existing hypotheses
  - Enjoyment of information
  - Producing more information for further use…

Challenges: A Visualisation Needs to Be Effective

- Visualisations need to support user tasks
- Correctness, accuracy, truth… are important concerns
- The challenge: any depiction of data is an abstraction where choices are made…
  - Which choices to make and which aspects to emphasize = goal of this class

- *"It's not just about making pretty pictures!"*
  - A visualisation can be beautiful, but must be effective!

Challenges: Most Designs Are Ineffective

- The problem: *the vast majority of possibilities in the design space will be ineffective for any specific usage context* !
  - One potential design may be a poor match with the properties of the human perceptual and cognitive systems, another may be a bad match for the task…
- Broad view of the design space necessary
  - Better to look for a “good match” while considering the whole design space than struggling to get to the “best match” while focusing too much!

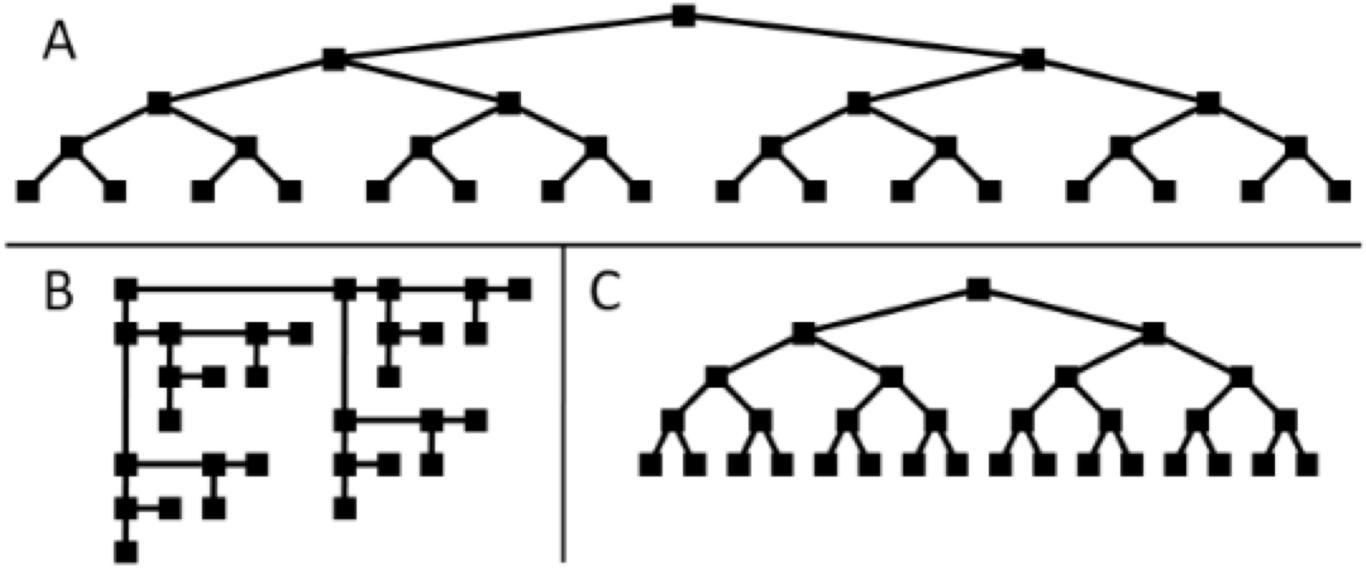
Challenges: Validation is difficult

- Validation = making sure that your design meets its design goal
- Problems:
  - How do you know if it works?
  - How do you argue that one design is worse or better than the other?
    - What does “better” even mean…?
  - Do users get something done faster?
    - More effectively? With more fun?
    - Who is the user, anyway…?
  - What sort of data will you use? What kind of data is benchmark data?
  - Etc. etc. etc.
- In the end, knowing what your task, design and user goals are is paramount

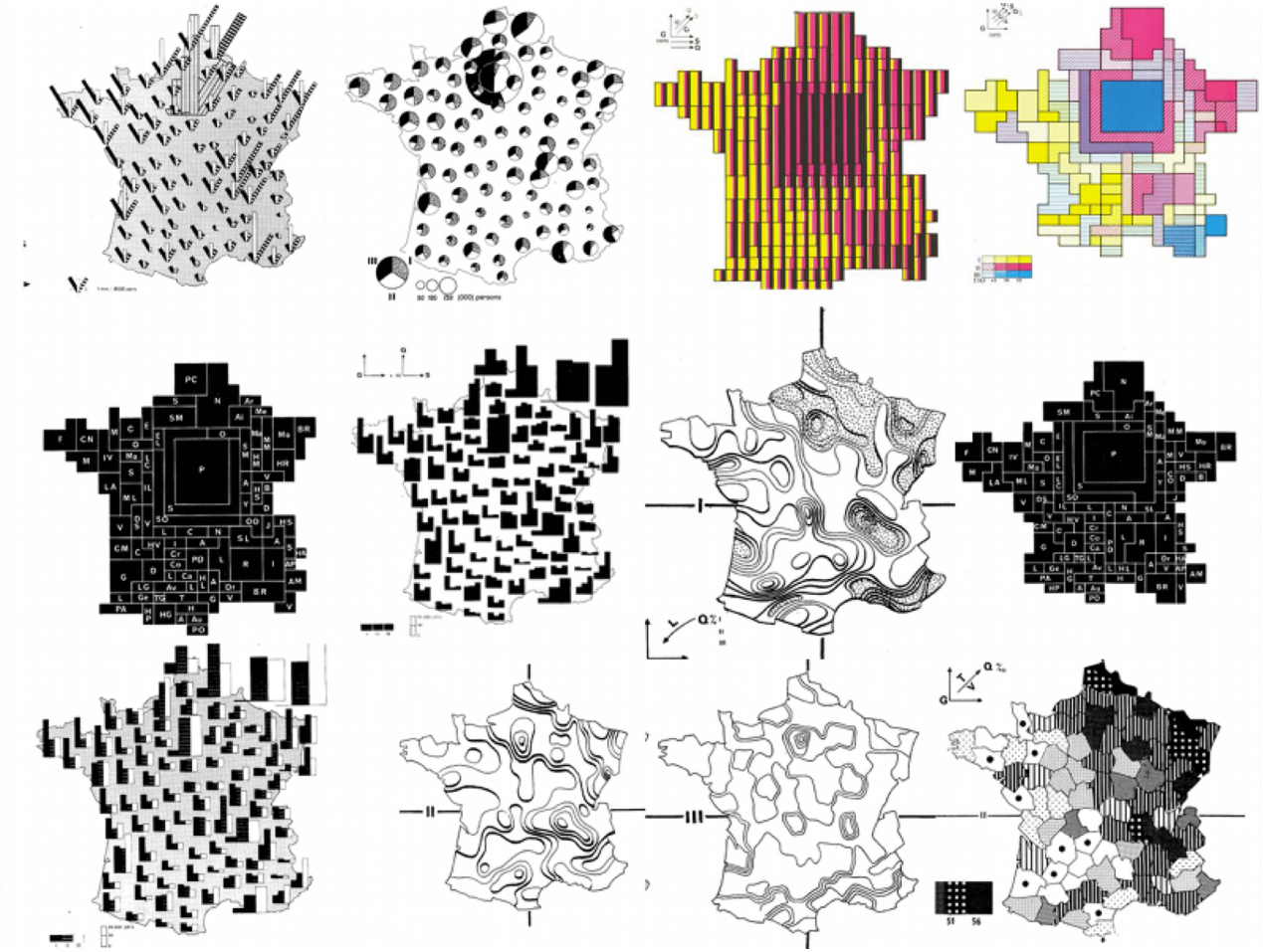
Challenges: Resources are Limited

- Three kind of resource limitations:
- Computational capacity
  - Issue arises quickly with larger datasets -> scalability
  - Balance between memory usage, responsiveness, precision…
- Human Perceptual and cognitive capacity
  - Human memory and attention are also finite resources !
  - Issues for long-term recall as well as short-term working memory
- Display capacity
  - Not enough pixels on screen…
  - Concept of information density (see next slide)
  - Trade-off between showing as much as possible at once (less navigation necessary) vs. showing too much at once (visual clutter)

Resource Limitation: Information Density Example



(Resource Limitation: Information Density – The Ultimate Reference)



Jacques Bertin, Sémiologie graphique, 1983

Resource Limitation: Pixel Density (Extreme) Example

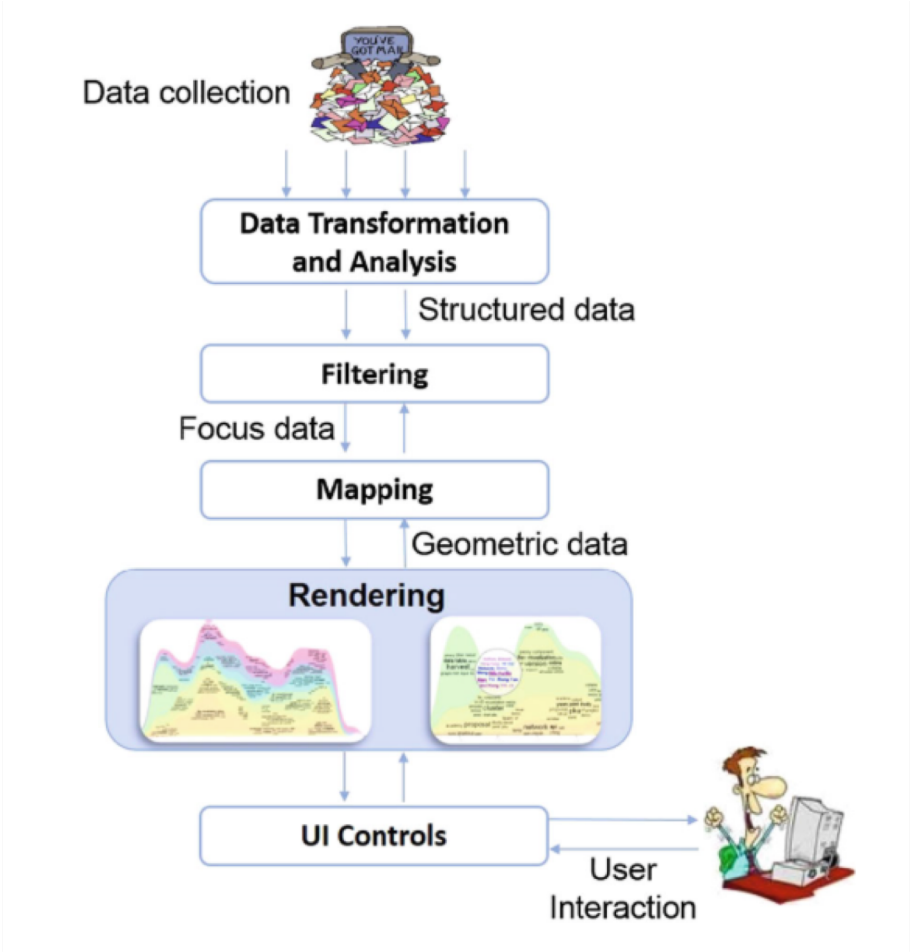
Powerwall, Université de Constance (DE)
http://www.vis.uni-konstanz.de/powerwall/

Getting Visualisation Design Right: Basic Tools

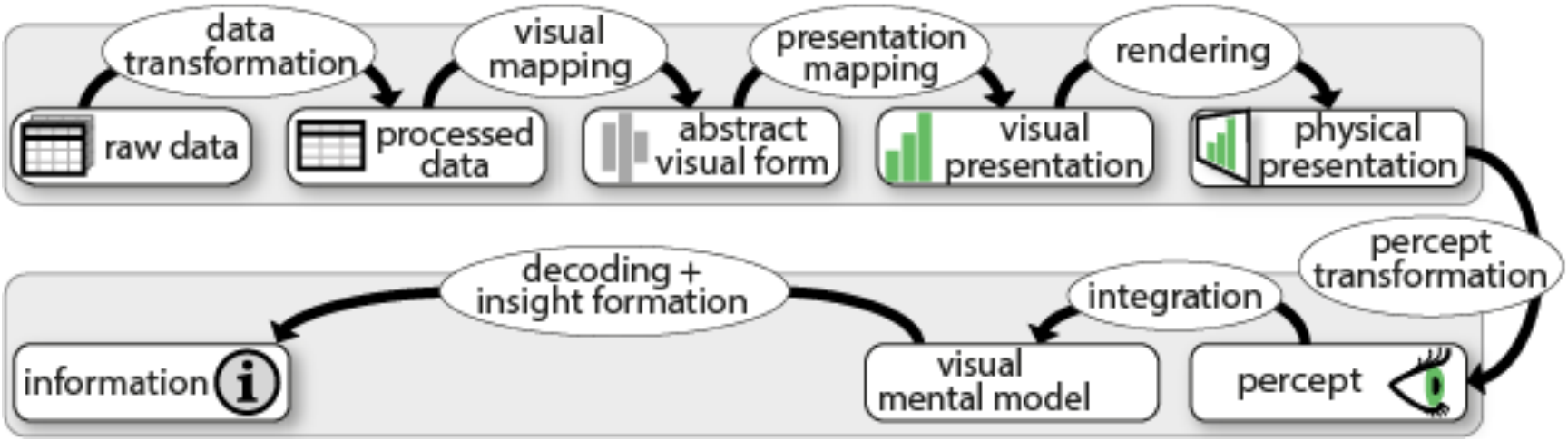
- Tools to help analysis and design of infovis systems

- Two main tools:
- The information visualisation pipeline (again)
  - … and a refined version of it
  - Gives background on the main constituents of a visualisation
- A Three-Part Analysis Framework
  - Helps with the analysis and design

Information Visualisation Pipeline

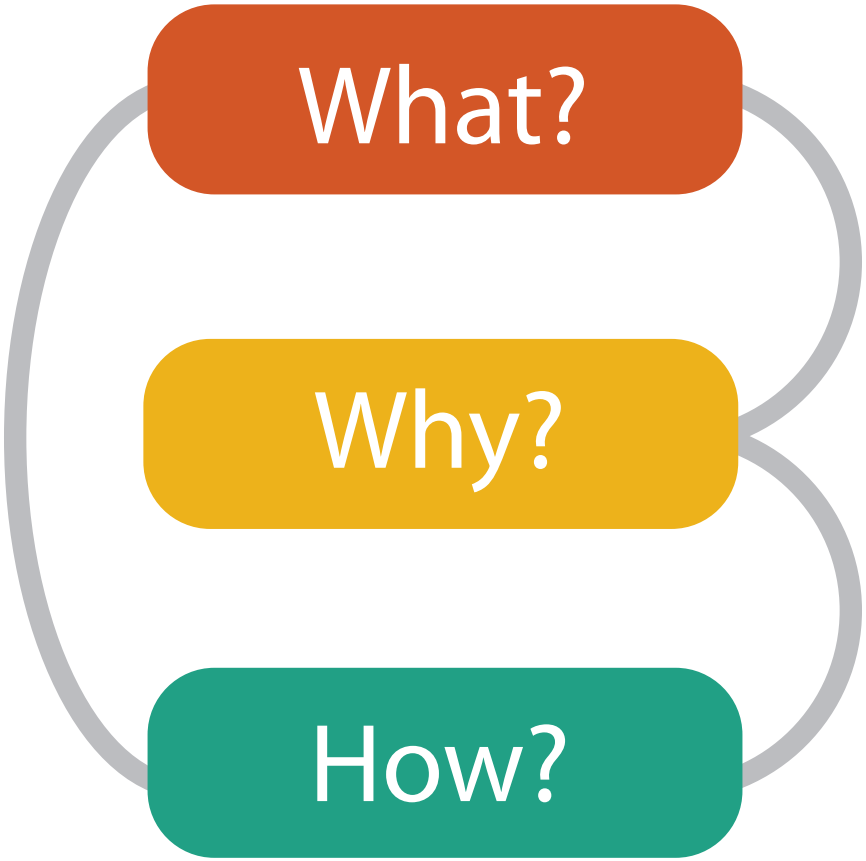


A More Advanced Information Visualisation Pipeline (Jansen et al. 2013)



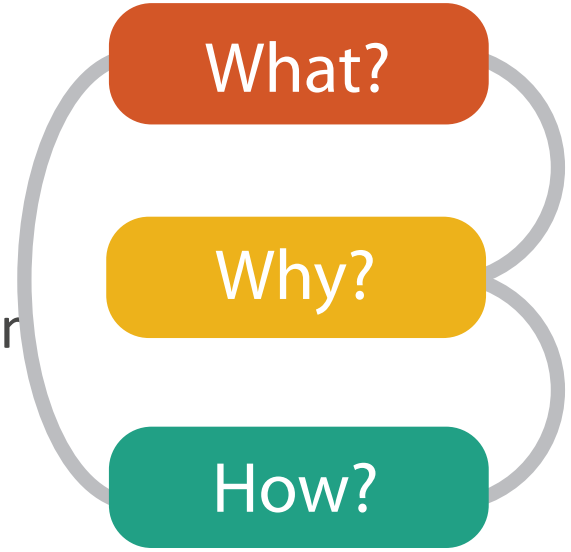
Jansen, Y., & Dragicevic, P. (2013). An interaction model for visualizations beyond the desktop. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2396-2405.

A Three-Part Analysis Framework

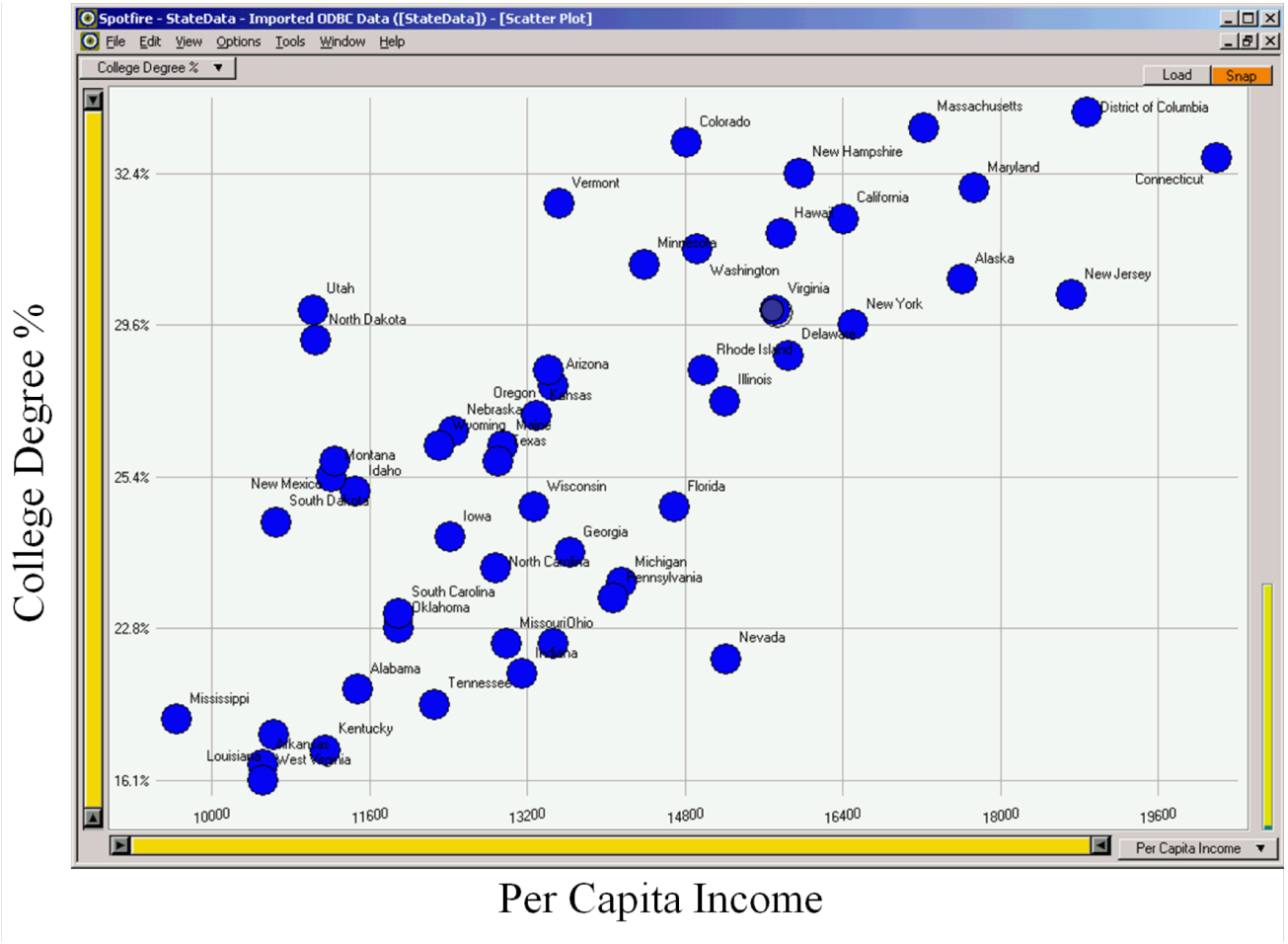


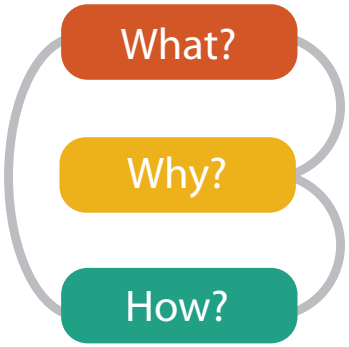
A Three-Part Analysis Framework

- WHAT data the user sees
  - Aka the data
- WHY the user intends to use a vis tool
  - Aka the task
- HOW the visual encoding and interaction idiom are constructed
  - Aka the visual idiom and interaction

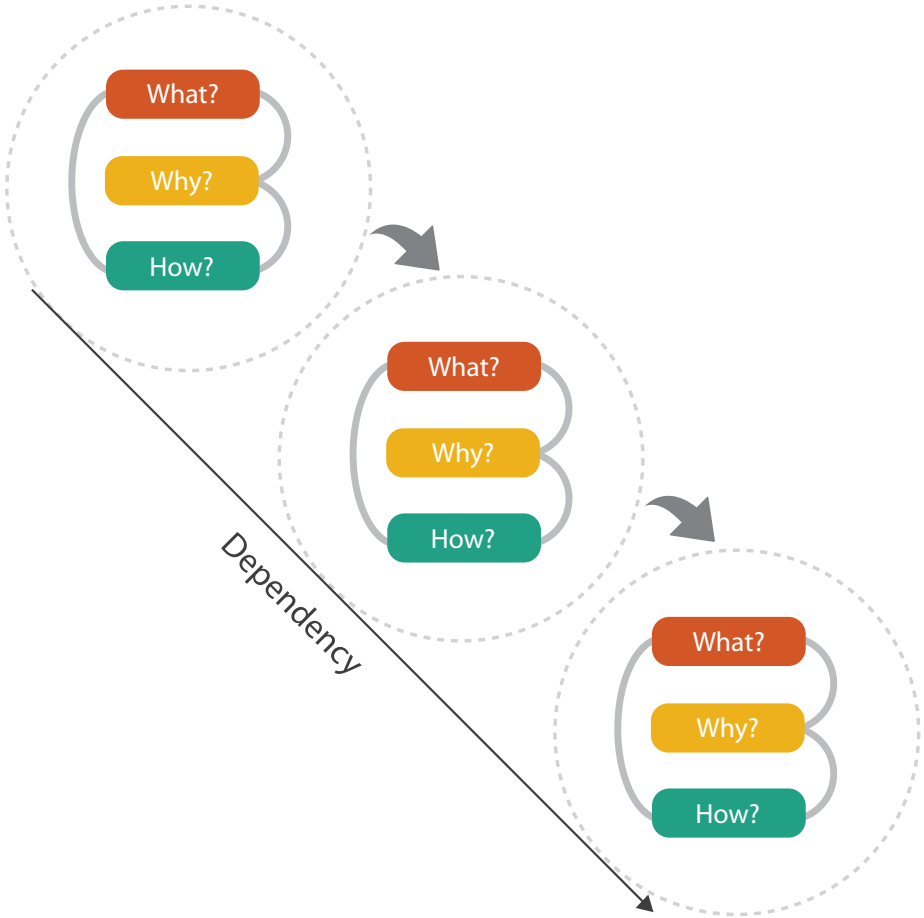


A Simple Example:





Sometimes, Multiple Levels of Analysis Needed



Program for the Next Weeks

- Data, Task and Validation
- Marks and Channels, Color Mapping
- Tables, Spatial data, Networks and Trees
- Manipulating View, Facetting, Focus + Context
- Reduce Items and Attributes + Some Cases Analysis