

Data Visualization

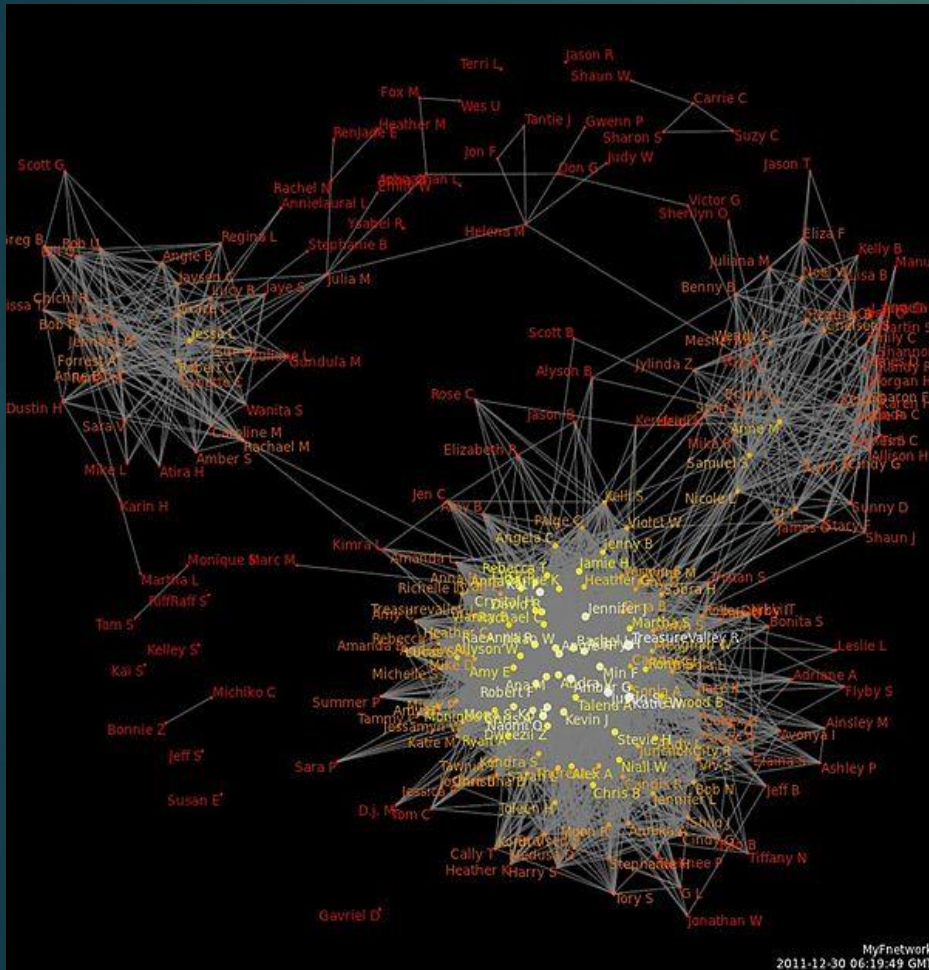
VINH PHAN
 AW1 – 06/01/2014



Agenda

- ▶ 1. Dealing with Data
- ▶ 2. Foundations of Visualization
- ▶ 3. Some Visualization Techniques
- ▶ 4. Life Cycle of Visualizations
- ▶ 5. Conclusion
- ▶ 6. Key Persons & Conferences

I. Dealing with Data: Explosion of Electronic Information

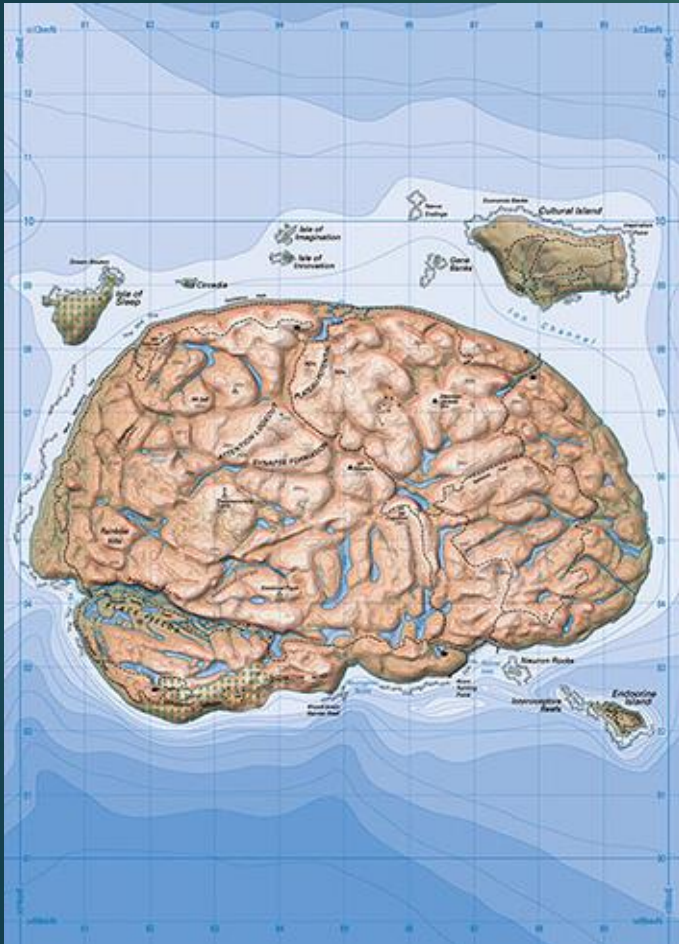


- ▶ 2012 estimate: 2.5 exabytes/day new info (1 exabyte = 1 million terabytes) [IBM: What is big data?]
- ▶ Open government / transparency movements.
 - ▶ e.g. Open Data Hamburg
- ▶ Digitization of media (photos, music, books...).
- ▶ E-commerce, electronic archives.
- ▶ Remote sensors, RFID tags, POS systems.
- ▶ Social media
- ▶ And much more...

[Data visualization from social media (facebook)]
→ http://en.wikipedia.org/wiki/Data_visualization#Data_analysis

II. Foundations of Visualization

History/Definition/Taxonomy



- ▶ Amplify cognition using the human vision (eye-brain connection, visual perception...)
 - ▶ which is the “highest-bandwidth” sense.
 - ▶ which is fast, parallel, bypass the language center, go direct to the visual cortex.
 - ▶ people think visually (visual memory: 3-4 simple shapes) [Vogel et al 2001]
- ▶ Leverage ability to recognize patterns, visual sense-making
 - ▶ attentive attributes (texts, numbers) → proceeds in serial (slow)
 - ▶ pre-attentive attributes (color, size, shape, orientation...) → proceeds in parallel (fast)

[“Map of new Brainland”, Unit Seven (2007)]

→ <http://www.unitseven.co.nz/>

II. Foundations of Visualization: Pre-attentive Attributes Examples

- ▶ Attentive processing → serially with conscious effort.
- ▶ Answer in 5 seconds: Count the number of 3s? [K. Andrews, 2013]

175496490872545628327267094621
635280462905702676727325929055
561548569586711934907152874596
596289748716229184490082538851
180265490932887579802909278921
872634890928895000283058985889
927756990049828005987761883115

II. Foundations of Visualization: Pre-attentive Attributes Examples

- ▶ Pre-attentive processing: parallel without conscious effort.
- ▶ Color is pre-attentive

175496490872545628**3**27267094621
6**3**5280462905702676727**3**25929055
5615485695867119**3**4907152874596
5962897487162291844900825**3**8851
1802654909**3**2887579802909278921
8726**3**489092889500028**3**058985889
92775699004982800598776188**3**115

II. Foundations of Visualization:

Visualization Example

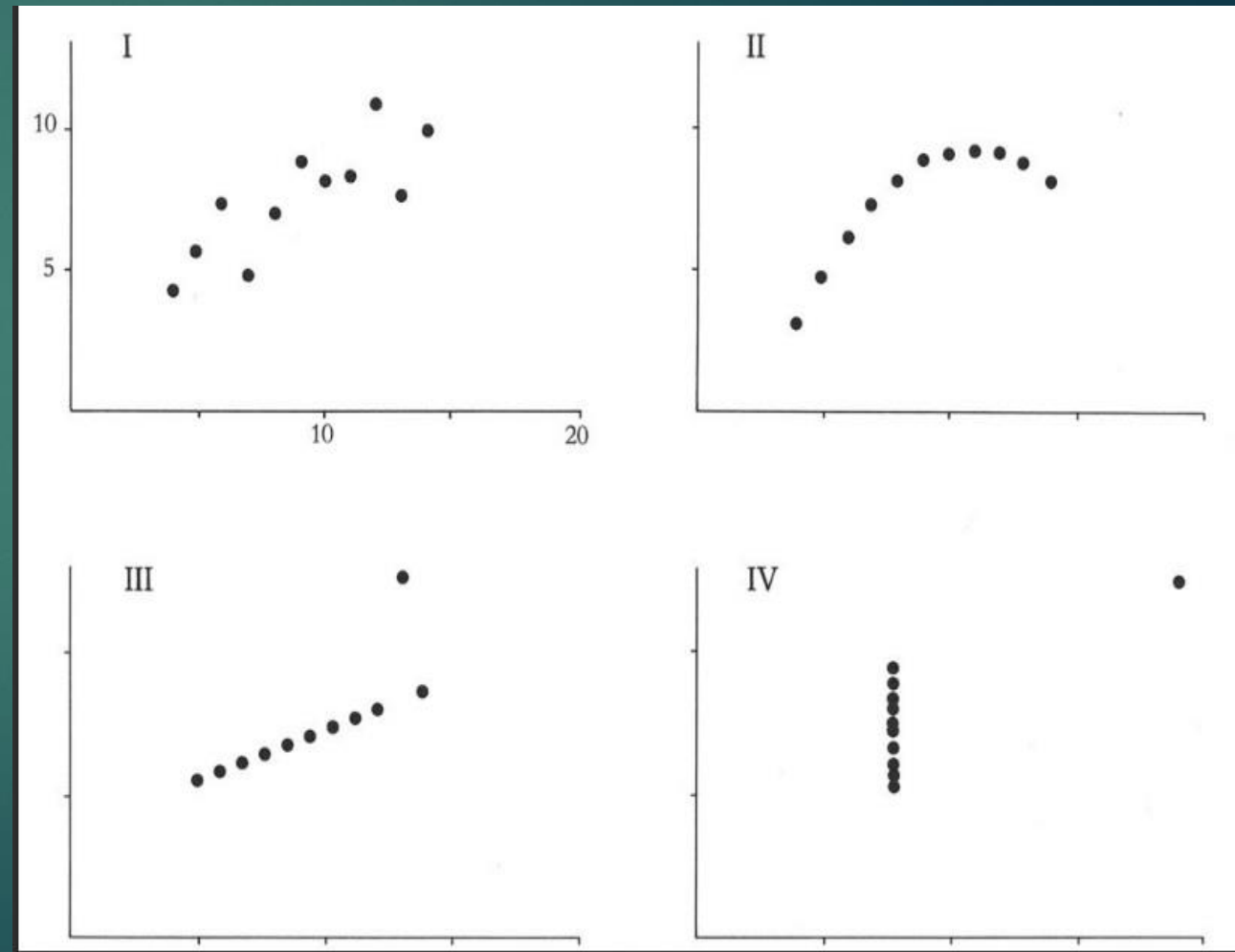
- ▶ Try to find the relationship between these numbers [E. Tufte, 2007]

I		II		III		IV	
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

[The Visual Display of Quantitative Information, E. Tufte (2007)]

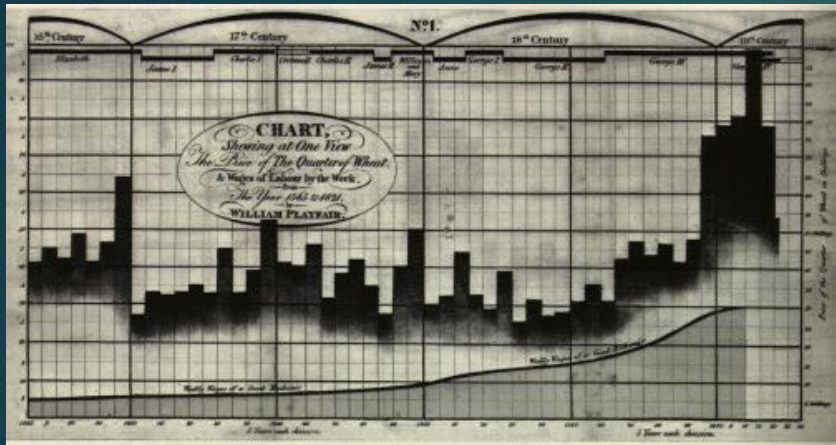
II. Foundations of Visualization: Visualization Example

- ▶ Quite hard?
- ▶ Try using scatter-plots [E.Tufte, 2007]

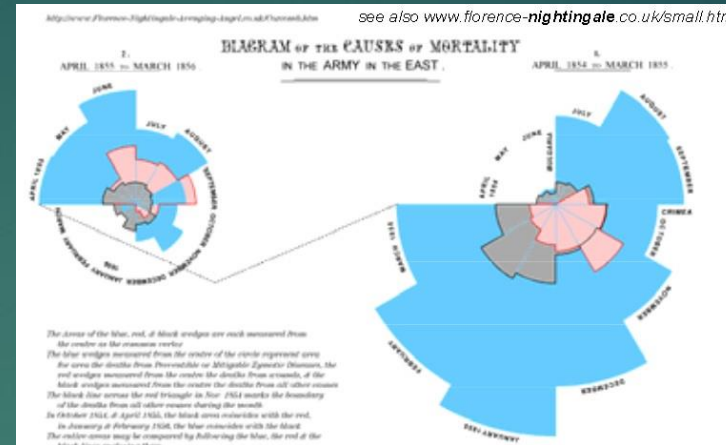


[The Visual Display of Quantitative Information,
E. Tufte (2007)]

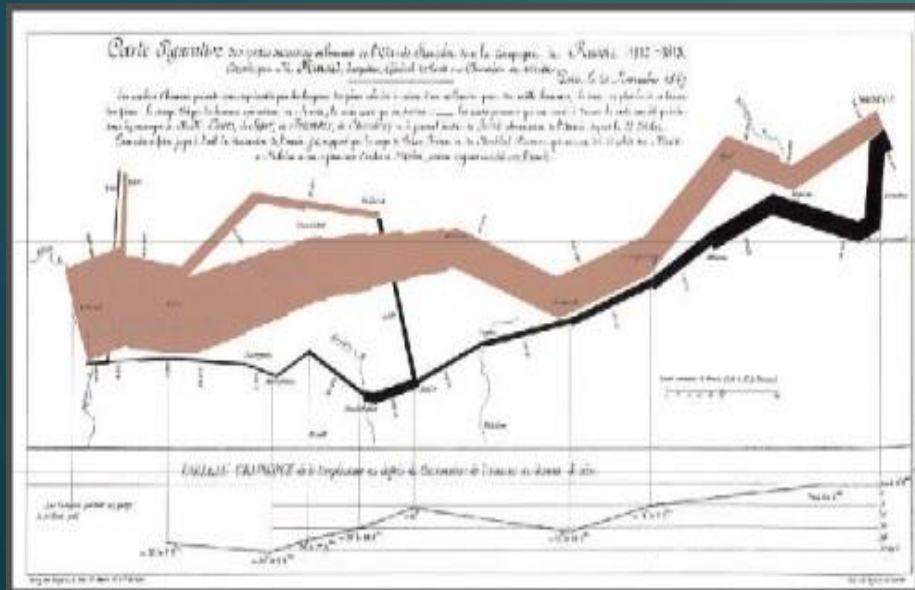
II. Foundations of Visualization: History/Definition/Taxonomy



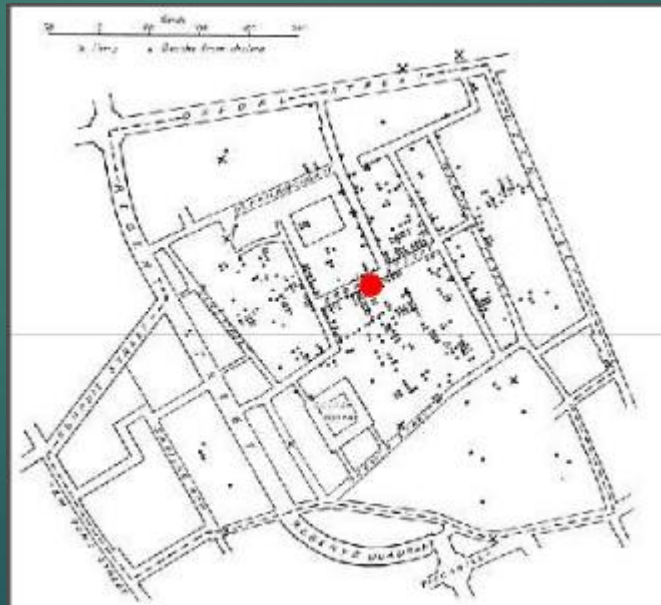
William Playfair (1786)



F.Nightingale (1858) – Causes of mortality In the army



Charles Minard (1869)
Napoleon's March



John Snow (1854)
London's cholera map

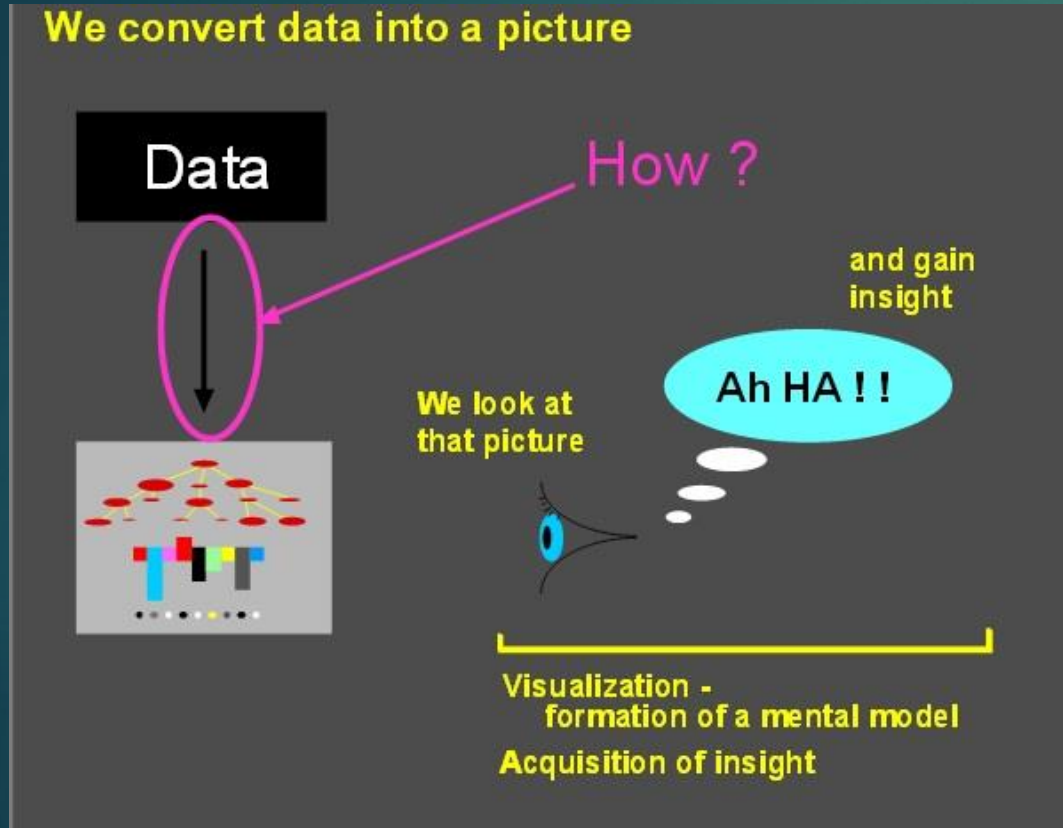


Harry Beck (1931)
London Underground map

II. Foundations of Visualization:

History/Definition/Taxonomy

10

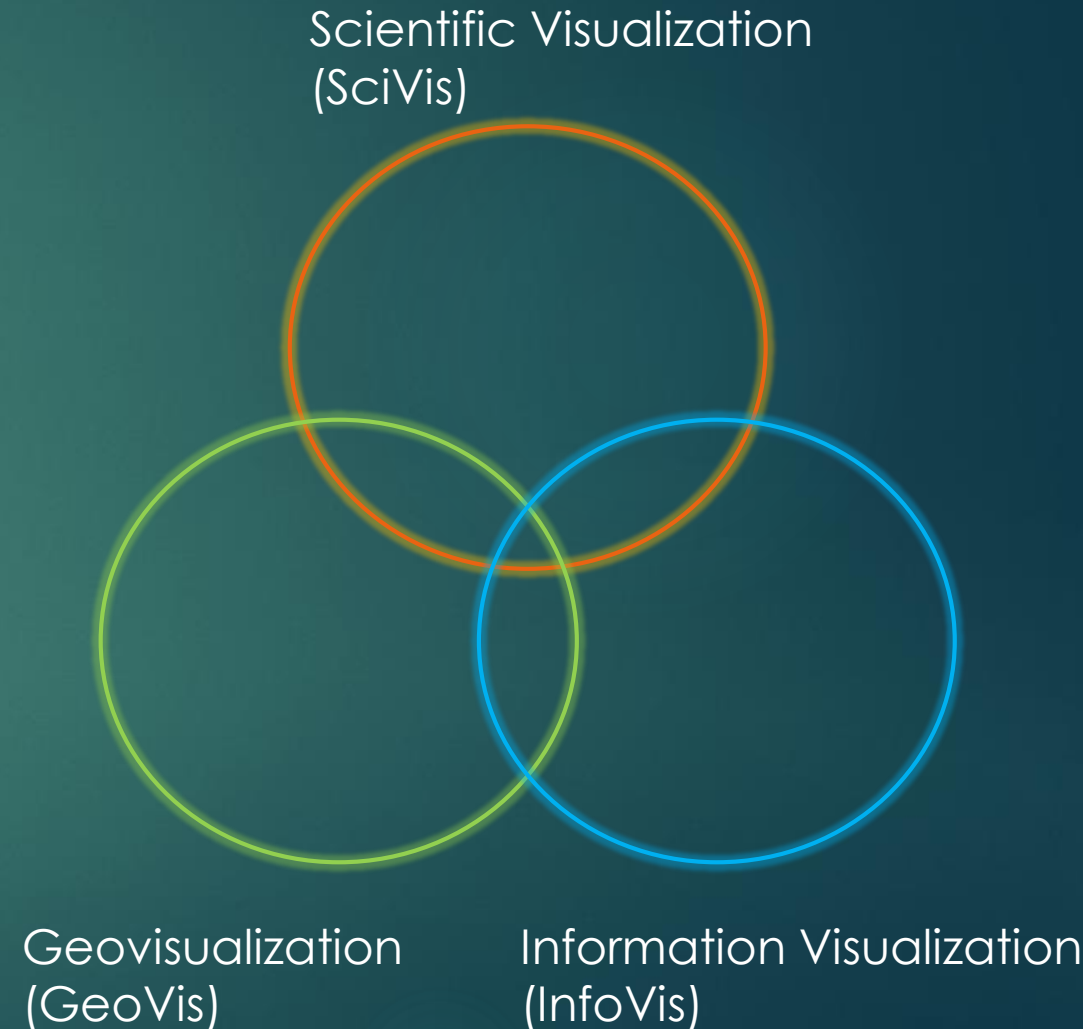


- ▶ Visualize: to form a mental model or mental image of something [Spence]
 - ▶ is a human cognitive activity.
 - ▶ Independent of computer.
 - ▶ May involve other senses: hearing, touch, ...
- ▶ How can the computer help?
 - ▶ facilitates interaction, and rearrangement of a display.
- ▶ Modern day: powerful graphic chips enable animation, real-time data processing

II. Foundations of Visualization:

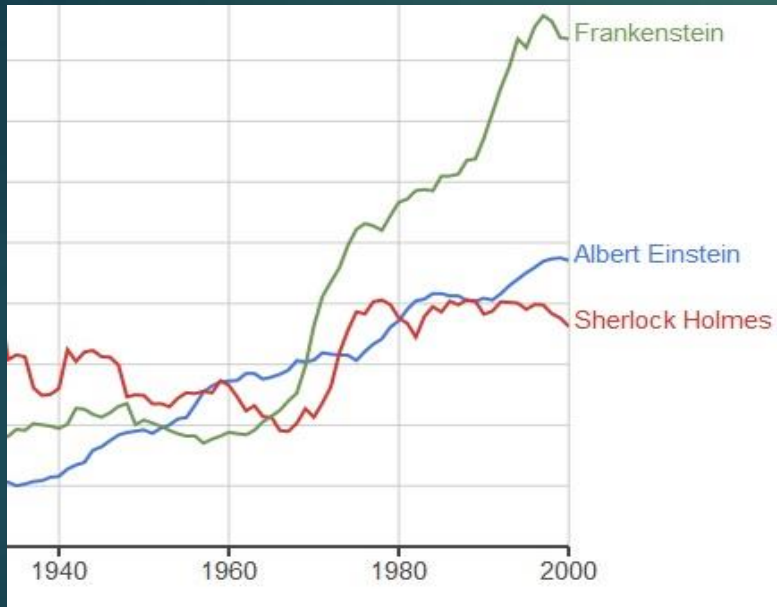
History/Definition/Taxonomy

- ▶ SciVis (flows, volumes, surfaces...)
- ▶ GeoVis (maps, geospatial data)
- ▶ InfoVis (abstract structures)
- ▶ DataVis = InfoVis + GeoVis
- ▶ Principal task of InfoVis: allow information to be derived from data.
- ▶ Definition: “InfoVis is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition.” [Card et al. 1999]



III. Visualization Techniques: Multivariate Analysis

12

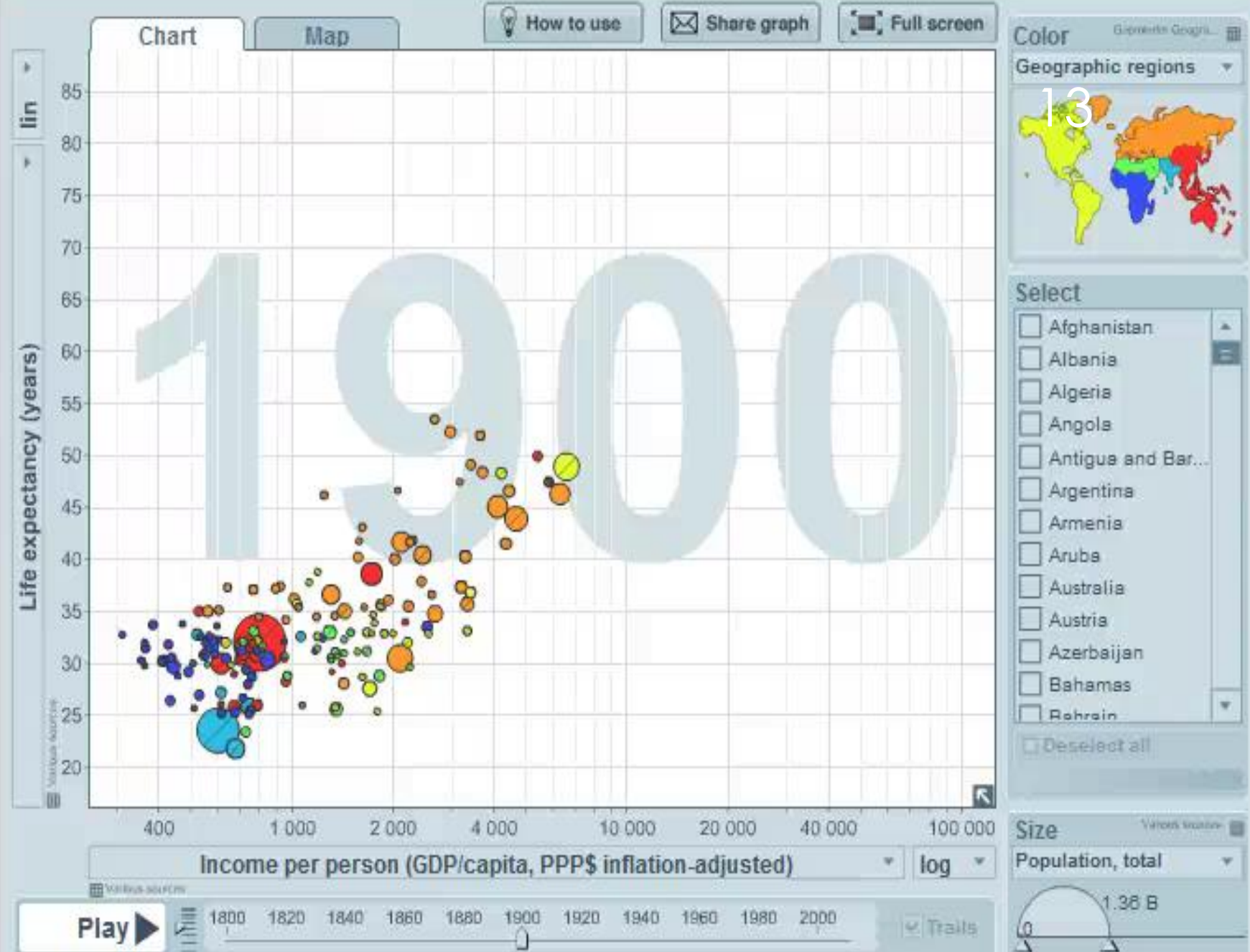


Google Ngrams Viewer
[<http://books.google.com/ngrams>]

- ▶ Problem: most real-world problems often involve many dependent attributes to analyze → many techniques become less effective.
- ▶ Multivariate data: data involve many attributes.
- ▶ *Geometric techniques*: scatterplot matrix, parallel sets...
- ▶ *Icon techniques*: star plots, Chernoff faces...
- ▶ *Pixel-oriented techniques*.

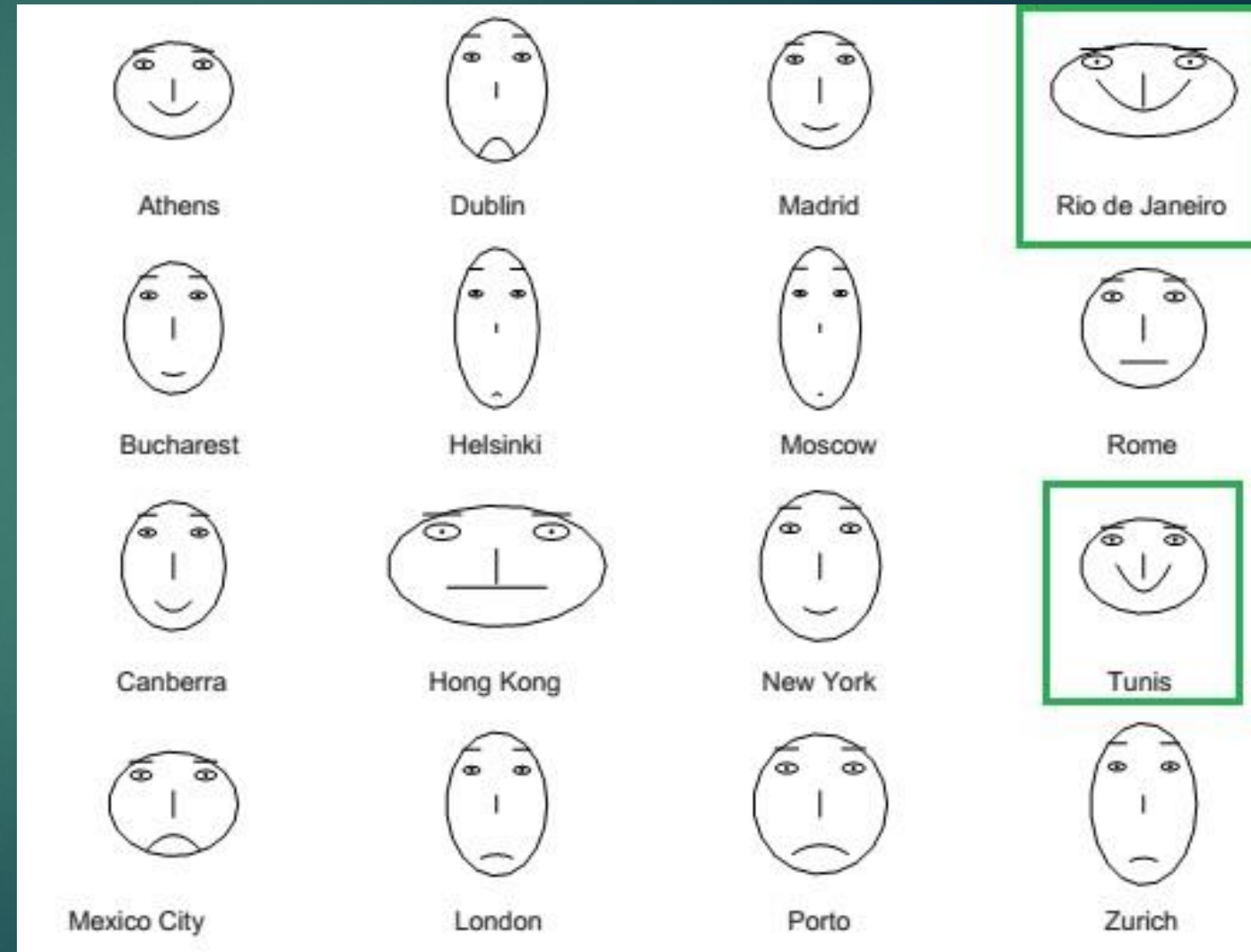
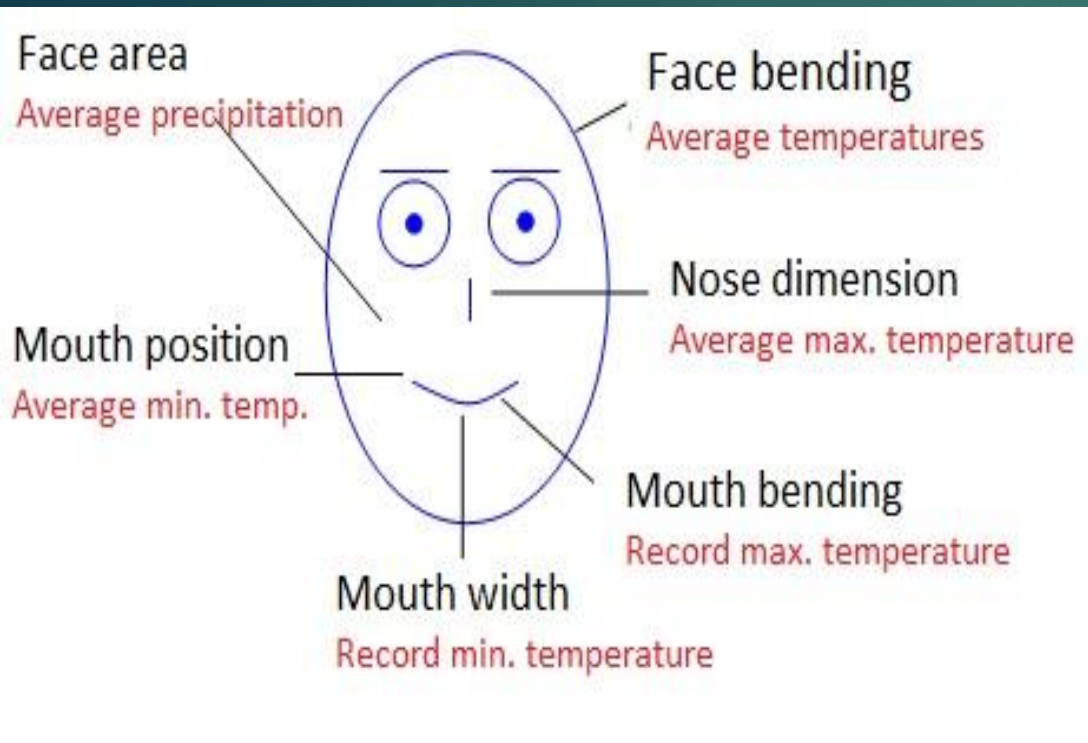
[Extended Scatterplot]

Relationship:
Wealth – Health – Population –
Geographical location



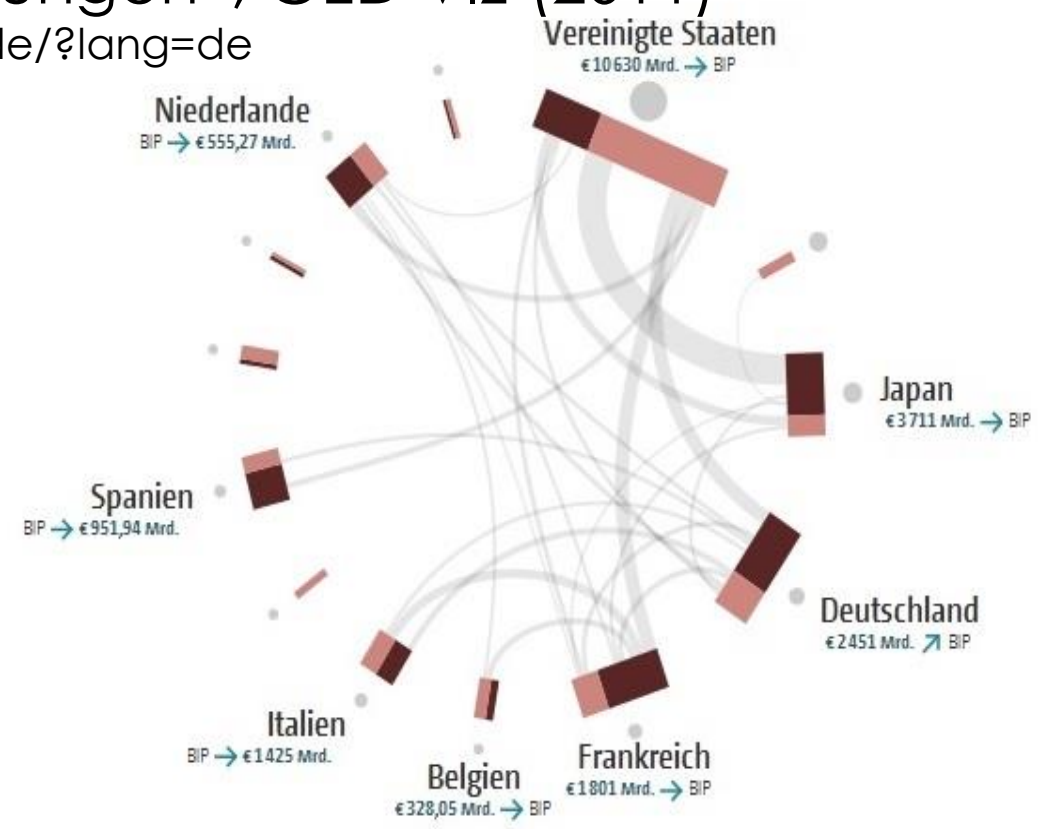
“Gapminder”
→ <http://www.gapminder.org>

[Chernoff Faces]



“Schuldenbeziehungen”, GED VIZ (2011)
 → <http://viz.ged-project.de/?lang=de>

[Multivariate Analysis]



QUELLEN
 Bank for International Settlements (BIS),
 OECD/European Commission

LEGENDE
 Forderungen Schulden
 2011 in Milliarden EUR (laufende Preise)

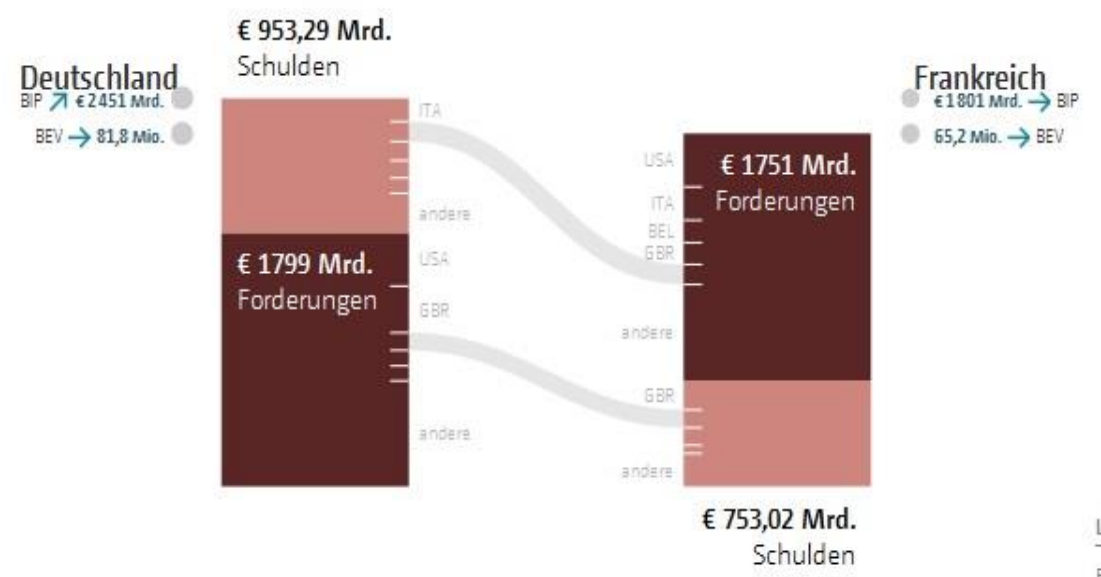
SLIDESHOW Hinzufügen
 ohne Titel 1

16

“Schuldenbeziehungen”, GED VIZ (2011)

→ <http://viz.ged-project.de/?lang=de>

[Multivariate Analysis]



QUELLEN | mehr

Bank for International Settlements (BIS),
 OECD/European Commission

LEGENDE | mehr

Forderungen (dark red) | Schulden (light red)

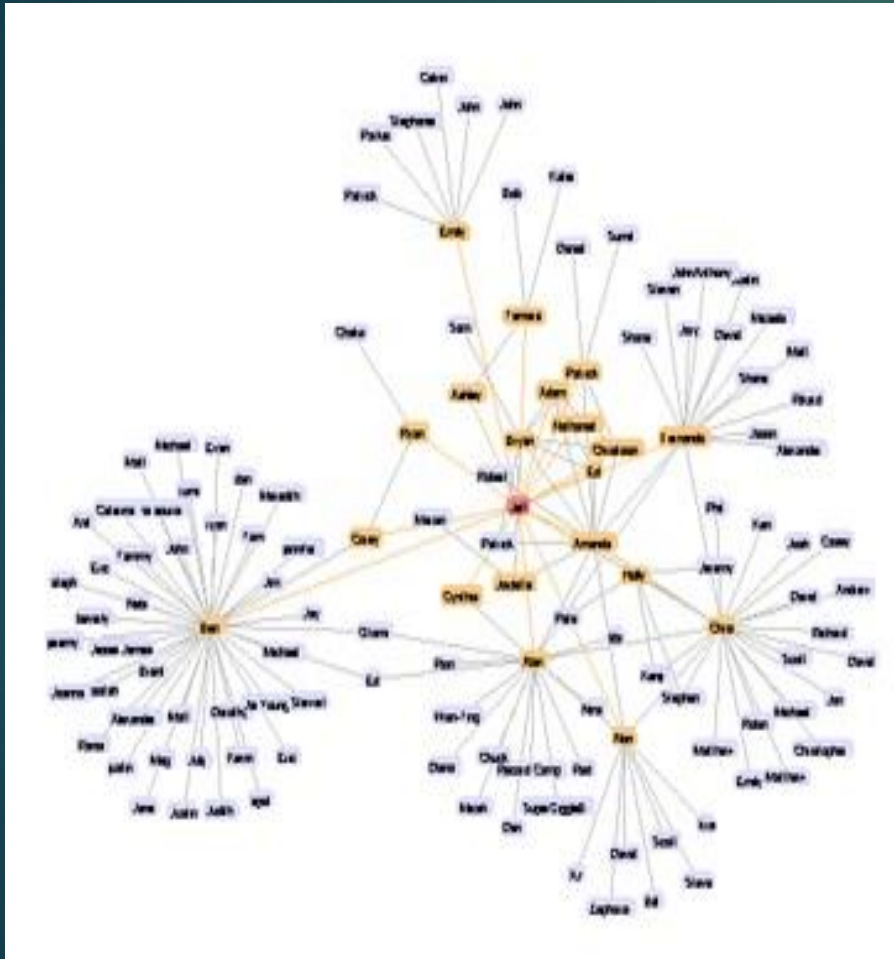
2011 in Milliarden EUR (laufende Preise)

SLIDESHOW | Hinzufügen

ohne Titel | 1

TEILEN, EINBETTEN, EXPORTIEREN

III. Visualization Techniques: Networks and Hierarchies

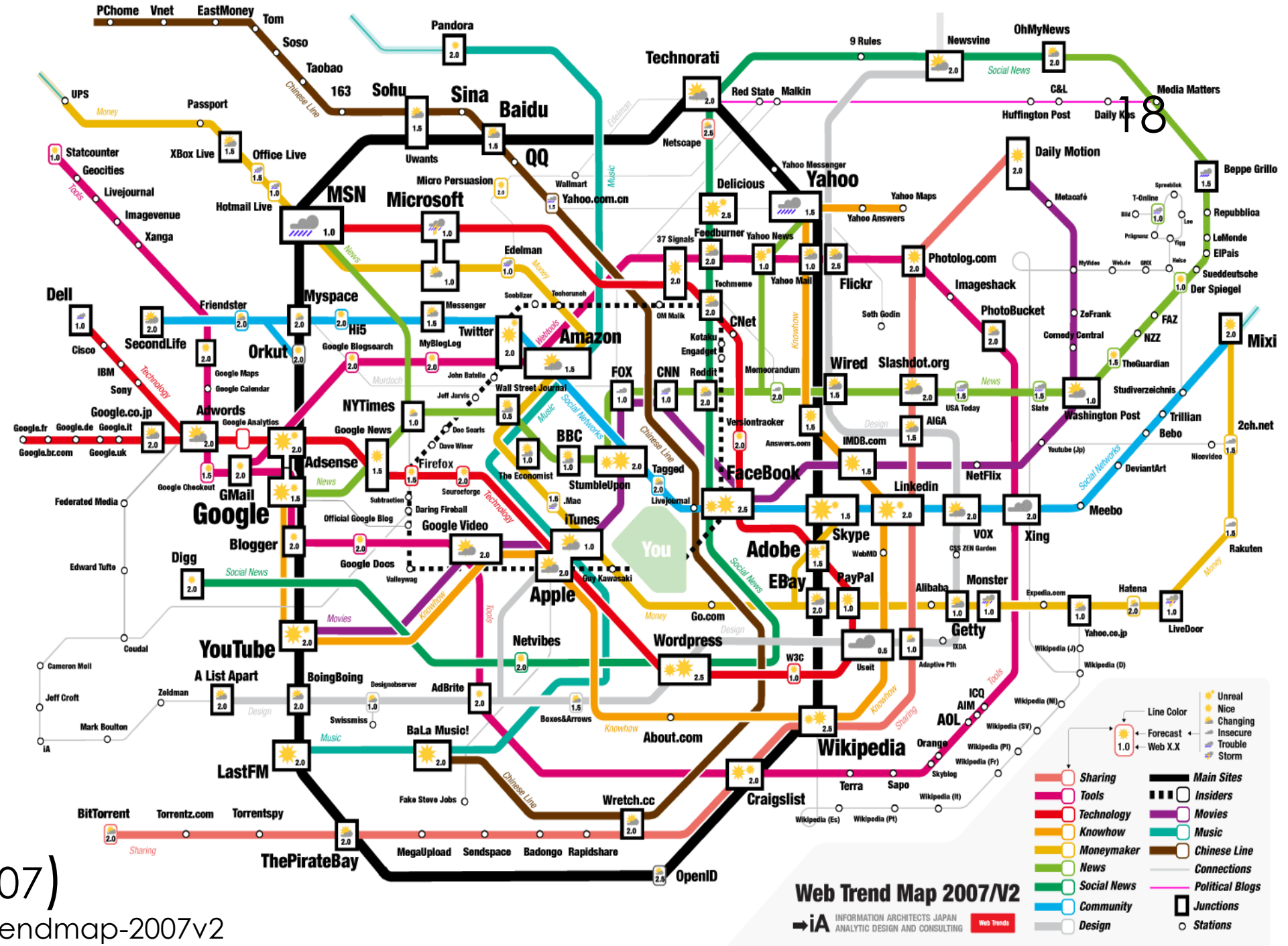


- ▶ Examples: transport network, company organization, social network ...
- ▶ Important characteristic: relation and enclosure.
- ▶ Can be represented by graphs and trees.
- ▶ Network data: concept maps, mind maps, geographical representation, 3D graphs ...
- ▶ Hierarchical data: file system, treemap, newsmap ...

Graph of social network using force-directed algorithm, Jeffrey Heer (UC Berkeley)
→ <http://vw.indiana.edu/ivsi2004/jherr/>

[Network Data] → mind maps

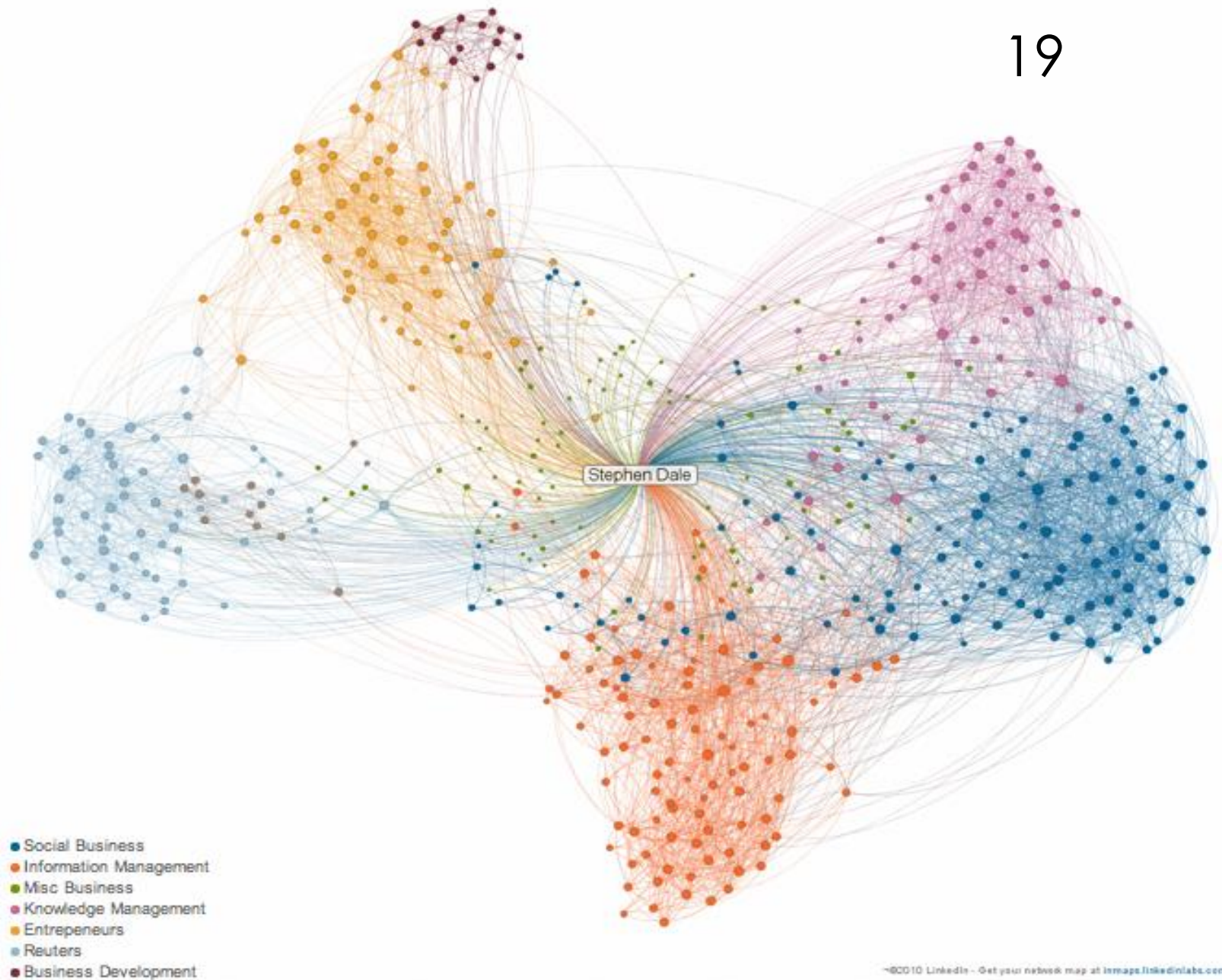
- 200 most successful websites
- Inspired by Tokyo Metro Map.



“Trend Map” (2007)
→ <http://ia.net/blog/ia-trendmap-2007v2>

[Complex Network Data] → social graphs

- Social networks
- Each node: an individual.
- Problem: density of nodes and edges in small space.
- Not very scalable



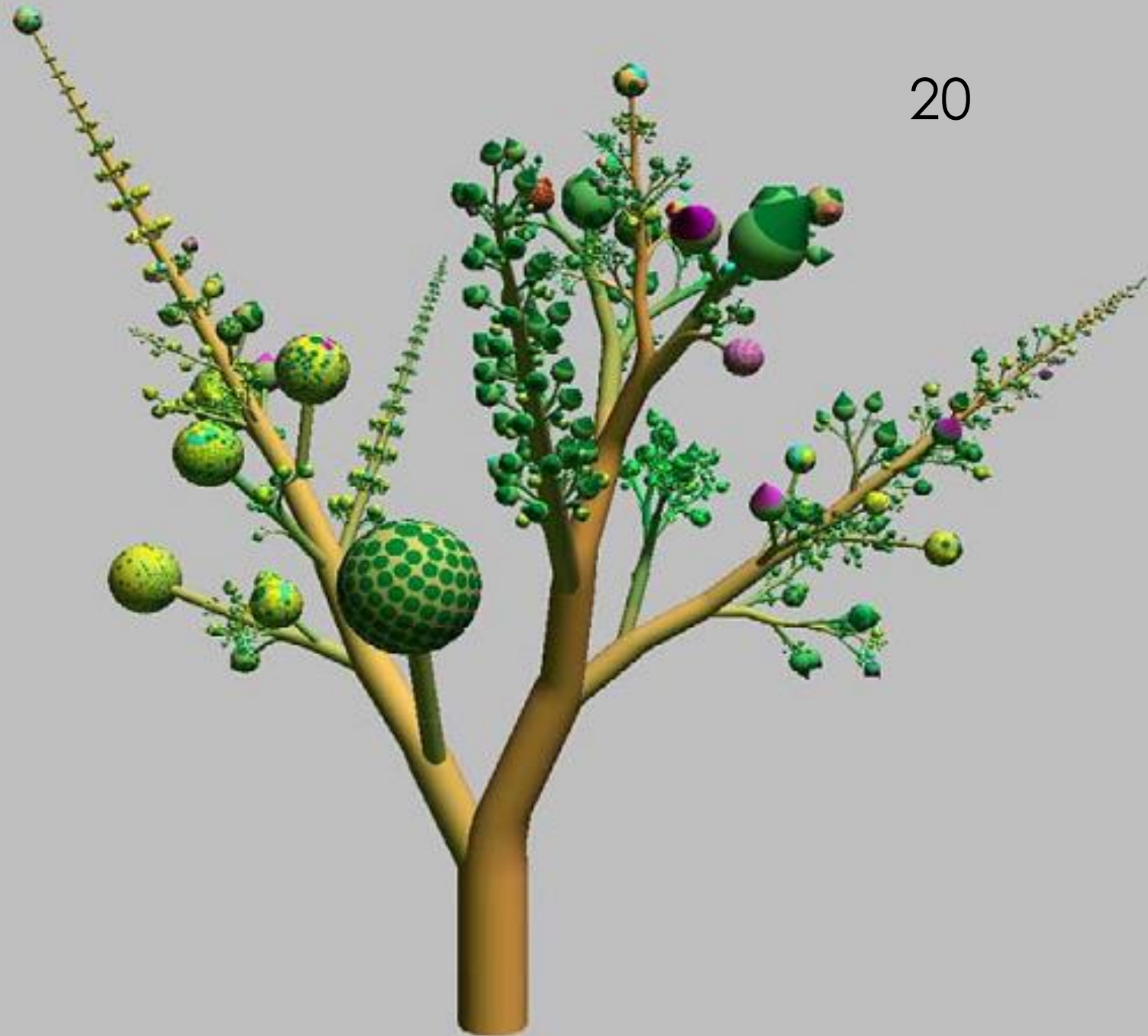
“LinkedIn Social Graph”

→ <http://steve-dale.net/2011/03/10/knowledge-hub-4-social-graph-and-activity-stream/>

[Hierarchical Data]

→ trees

- *Botanical tree*: represents file system.
- *Fruit* represents a collection of files.
- *Colored spot* represents a file.
- *Area & color of spot* → dimension & file type.

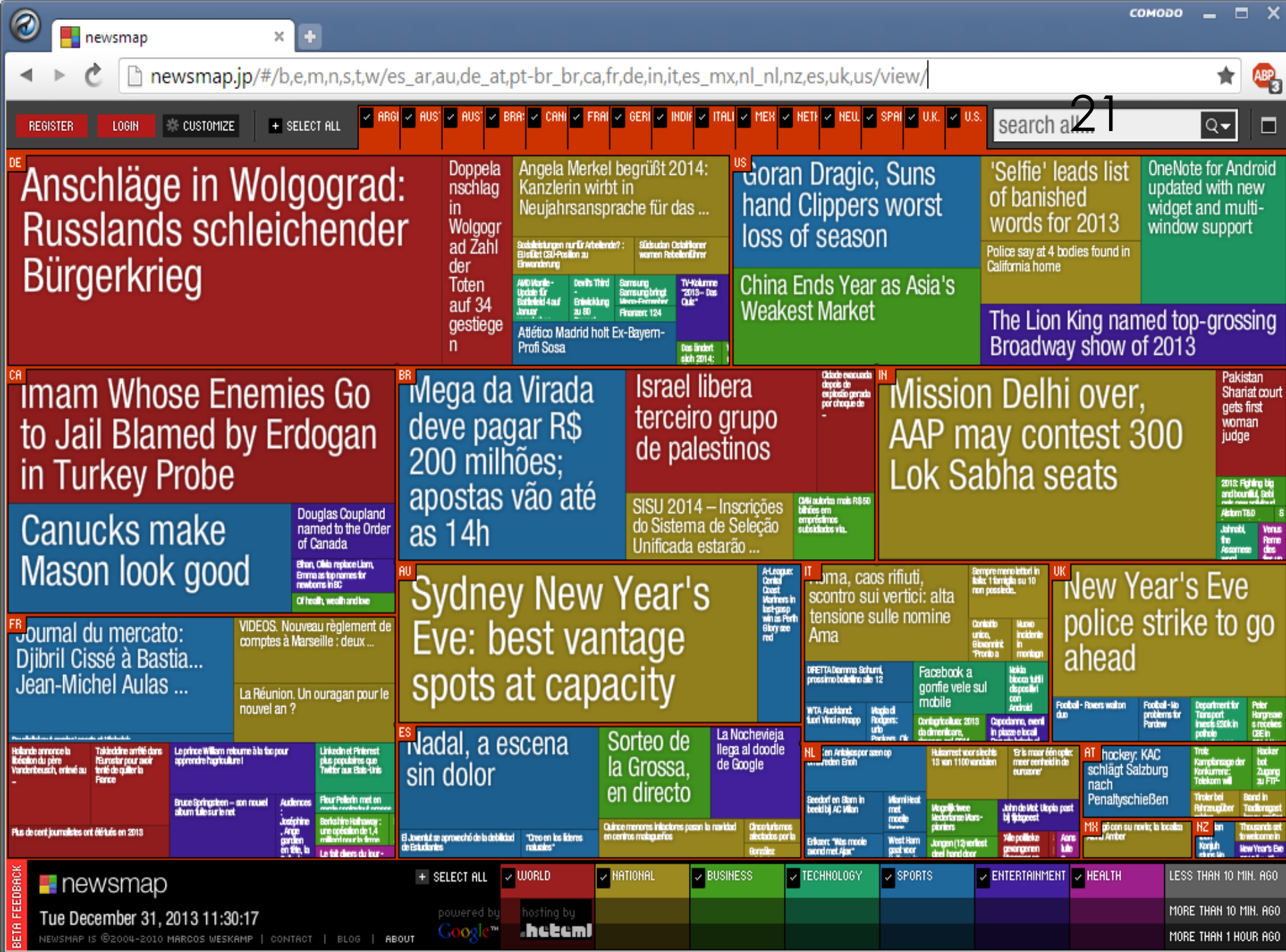


“Botanical Tree”, Jack van Wijk
[Eindhoven University of Technology, NL]

[Hierarchical Data]
 → tree maps

- Newsmap: visualizes news from Google News.
- Color: type/freshness of news.
- Size: number of articles per story.

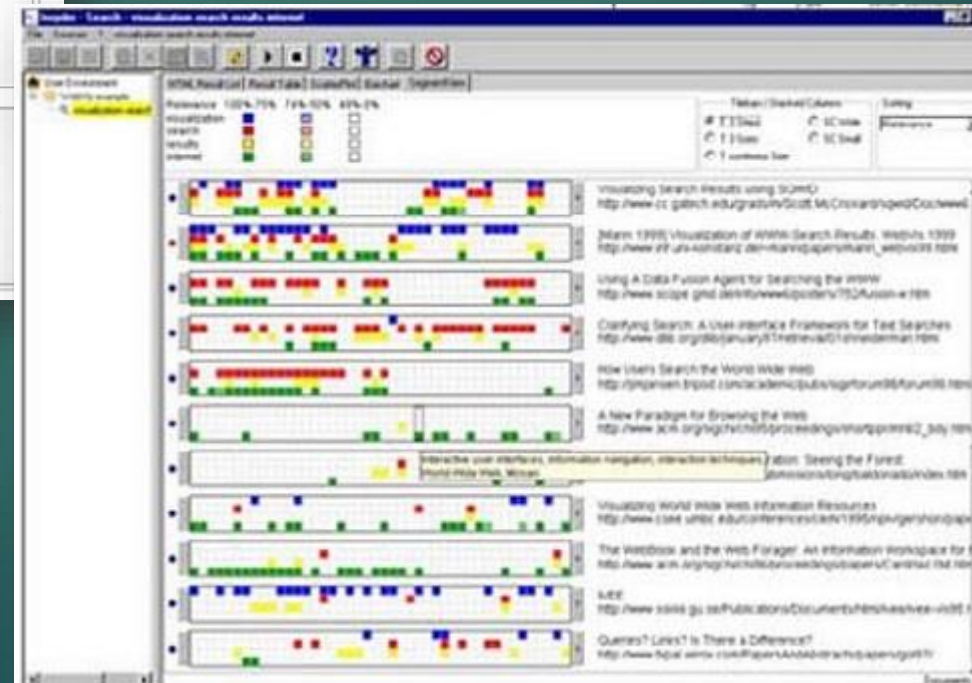
Newsmap (2013)
 → <http://marumushi.com/apps/newsmap>



III. Visualization Techniques: World Wide Web



“Google Analytics”
[<http://google.com/analytics>]



“INSYDER Tool (Tile bar)”
[Harald Reiterer, Uni. Konstanz (Springer-Verlag, 2005)]

- ▶ Website maps (site maps) → Dynamic Diagrams
[<http://dynamicdiagrams.com>]
- ▶ Website log data → Google Analytics
[<http://google.com/analytics>].
- ▶ Visualizing search engine results → KartOO
[<http://www.kartoo.com>].
- ▶ Analysis of interaction in blogs → blogviz
[<http://blogviz.com>]

IV. Life Cycle of Visualizations

Gathering Data

Discovery/
Acquisition

Cleaning/
“Munging”

Analyzing It

Analysis/Exploratory
Visualization

Sharing Findings

Publication

Own Data

- Spreadsheets
- Databases
- Digitized media

Other Sources

- Public data
- Archives/Libraries
- Academic partners
- Purchase

Scraping

- Junar
- OutwitHub
- ScraperWiki

IV. Life Cycle of Visualizations

Gathering Data

Discovery/
Acquisition

Cleaning/
“Munging”

Analyzing It

Analysis/Exploratory
Visualization

Sharing Findings

Publication

Original Research

- Spreadsheets
- Databases
- Digitized media

Other Downloads

- Public data
- Archives/Libraries
- Academic partners
- Purchase

Scraping

- Junar
- OutwitHub
- etc ...

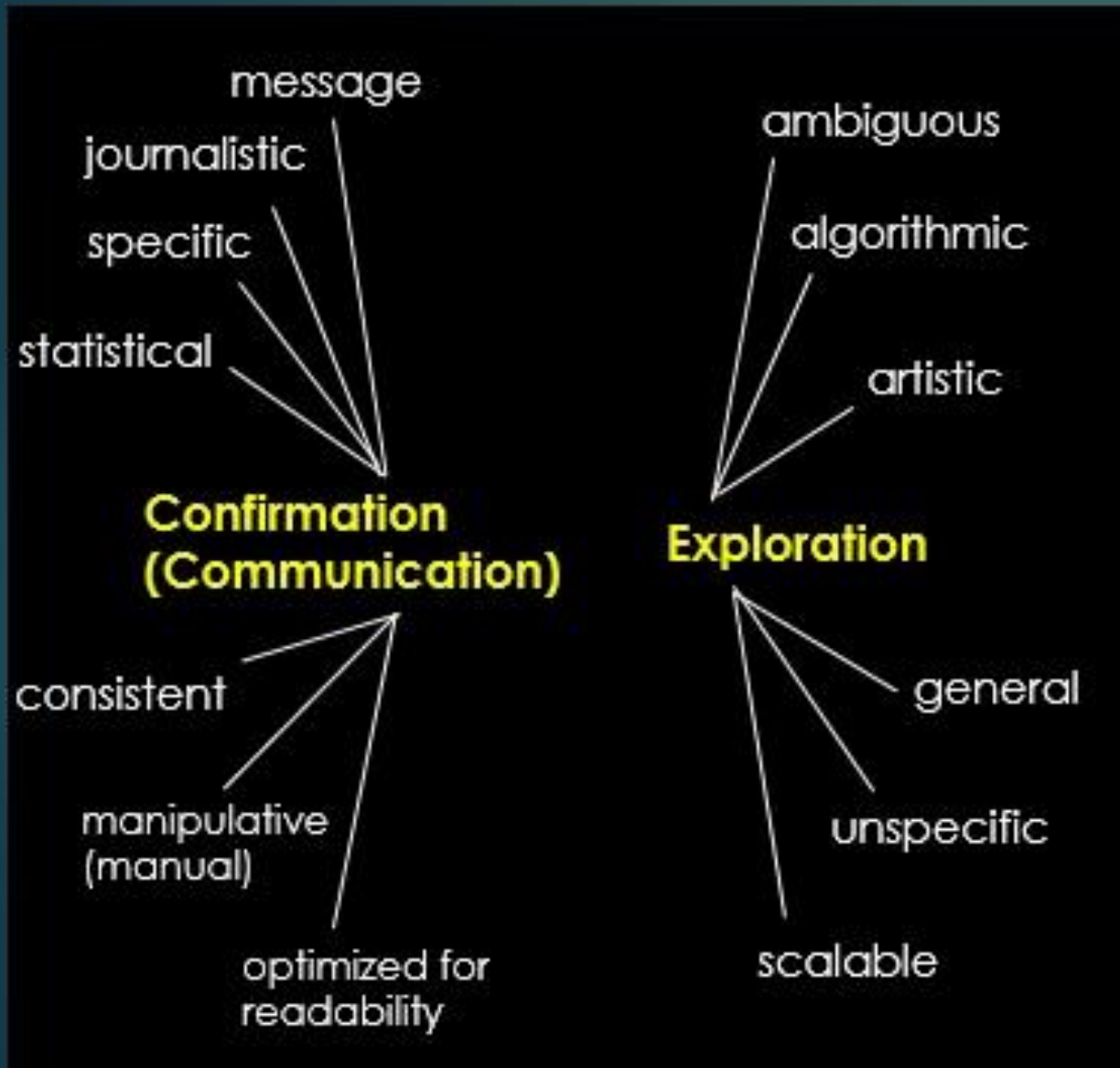
Normalization/Format/ Conversion

- Google Refine
- Data Wrangler
- Mr. Data Converter
→ <http://shancarter.github.io/mr-data-converter/>
- etc ...

Many types of data

- ✓ Linear structures
- ✓ Hierarchies
- ✓ Networks, graphs
- ✓ Multi-dimensional metadata
- ✓ Univariate, bivariate, multivariate
- ✓ Ordinal
- ✓ Quantitative
- ✓ Categorical
- ✓ Relationships
- ...

IV. Life Cycle of Visualizations

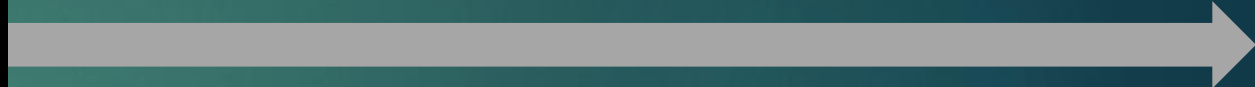


Analyzing It

Analysis/Exploratory Visualization

Sharing Findings

Publication



Web Services

- Google Analytics
- Google Fusion Tables
- IBM ManyEyes

Application

- Tableau/Tableau Public
- OfficeSuite (MS/Open)
- Gephi
- Node XL
- Spotfire
- R
- Processing

IV. Life Cycle of Visualizations

Gathering Data

Discovery/
Acquisition

Cleaning/
“Munging”

Analyzing It

Analysis/Exploratory
Visualization

Sharing Findings

Publication

Original Research

- Spreadsheets
- Databases
- Digitized media

Other Downloads

- Public data
- Archives/Libraries
- Academic partners
- Purchase

Scraping

- Junar
- OutwithHub
- ScraperWiki

Normalization/Format/ Conversion

- Google Refine
- Data Wrangler
- ...

Web Services

- Google Analytics
- Google Fusion Tables
- IBM ManyEyes

Application

- Tableau/Tableau Public
- OfficeSuite (MS/Open)
- Gephi
- Node XL
- Spotfire
- R
- Processing

Static

- Photoshop/Illustrator
- Previous tools...

Animated

- Processing
- Flash
- After Effects

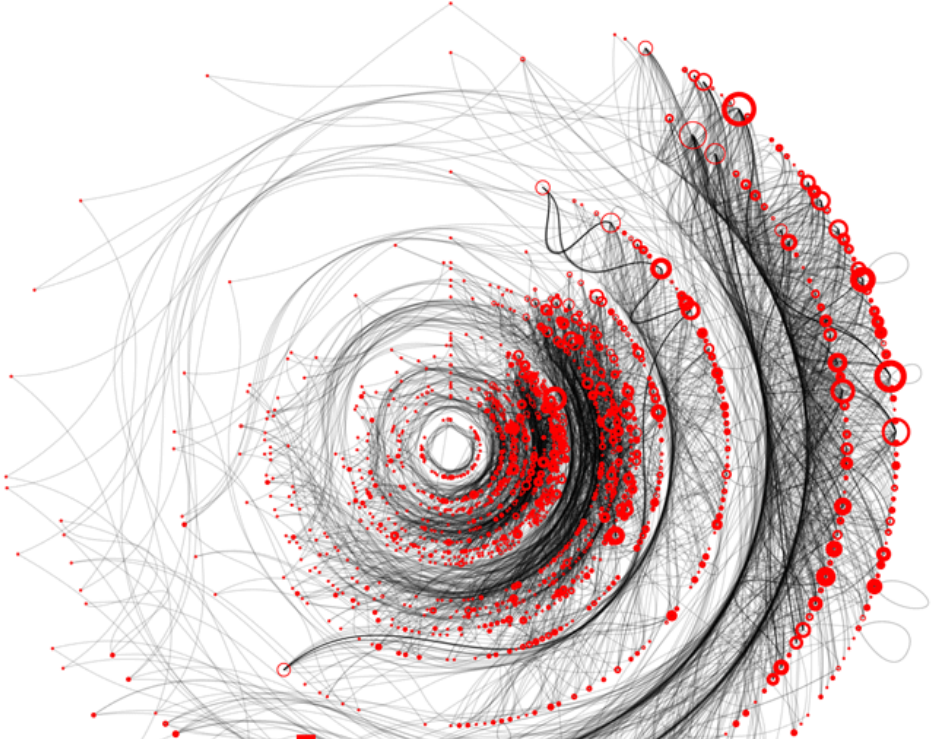
Interactive/Web

- HTML5
 - ProtoVis
 - D3
 - ProcessingJS
- Adobe Flash/Flex
- Processing

IV. Life Cycle of Visualizations

- ▶ Static representations
 - ▶ Code Poetry

27



poetry ON THE ROAD

7. INTERNATIONALES LITERATURFESTIVAL BREMEN
11. – 19. MAI 2006

VERANSTALTET VON: HOCHSCHULE BREMEN UNIVERSITY OF APPLIED SCIENCES radiobremen Goethebund in Bremen e.V.

GEFÖRDERT VON: Bremen Marketing, Senator für Kultur, Karin und Uwe Holweg Stiftung, Bernd und Eva Hockemeyer Stiftung, Wolfgang Ritter-Stiftung, DAAD, Waldemar Koch Stiftung, Bremer Literaturstiftung, pro helvetia

WWW.POETRY-ON-THE-ROAD.COM | Programmheft und Karten bei: Buchladen im Ostertor, Fehrfeld 60, Fon: 0421-785 28 | Die Sparkasse Bremen

GESTALTUNG: jung und pfeffer: visuelle kommunikation Bremen, Amsterdam | mit Boris Müller, esono.com

Boris Müller (2006)

[<http://esono.com/boris/projects/poetry06/>]

IV. Life Cycle of Visualizations

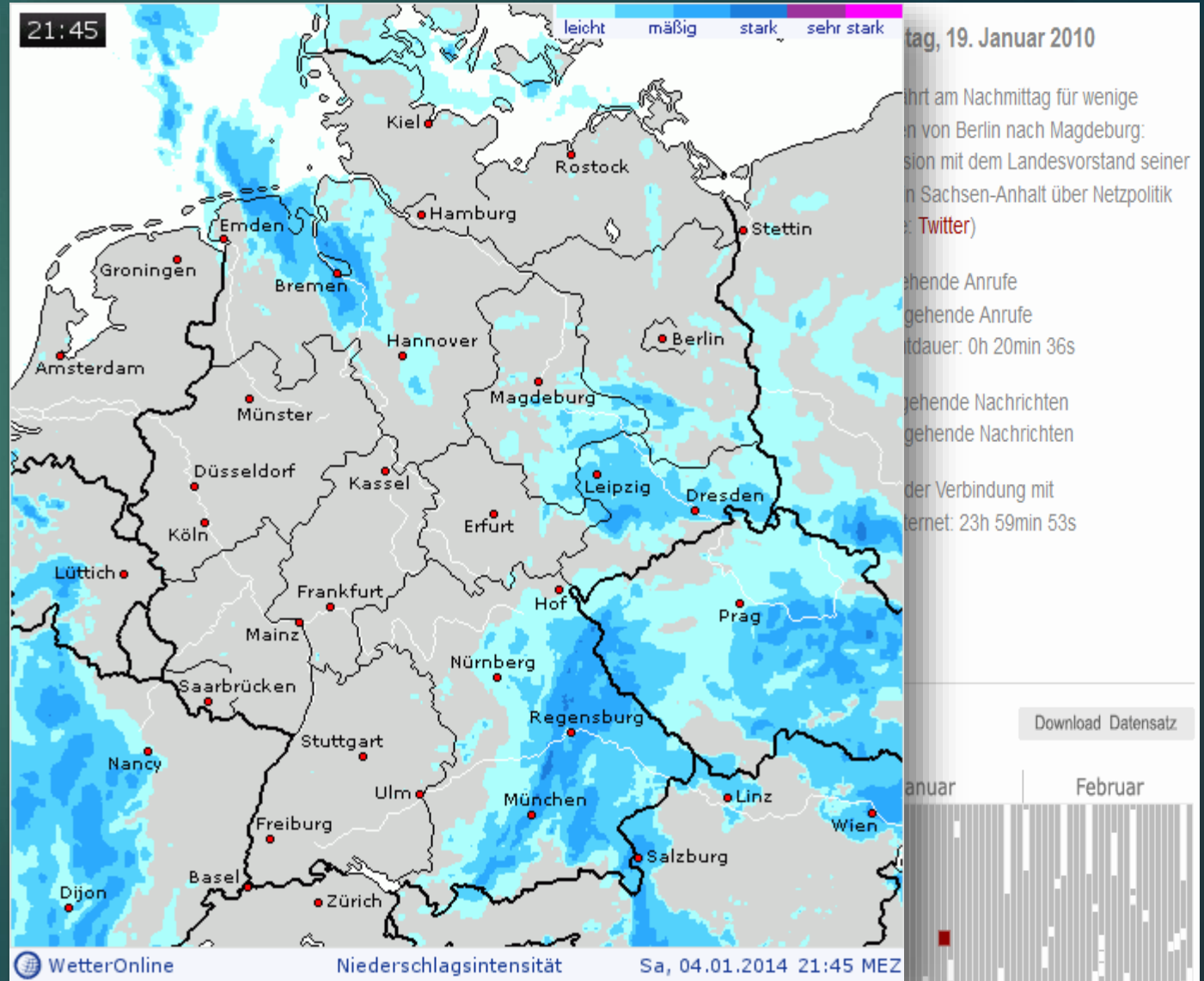
28

- ▶ Static representations:
 - ▶ Code Poetry
- ▶ Manipulable representations:
 - ▶ LiquiData



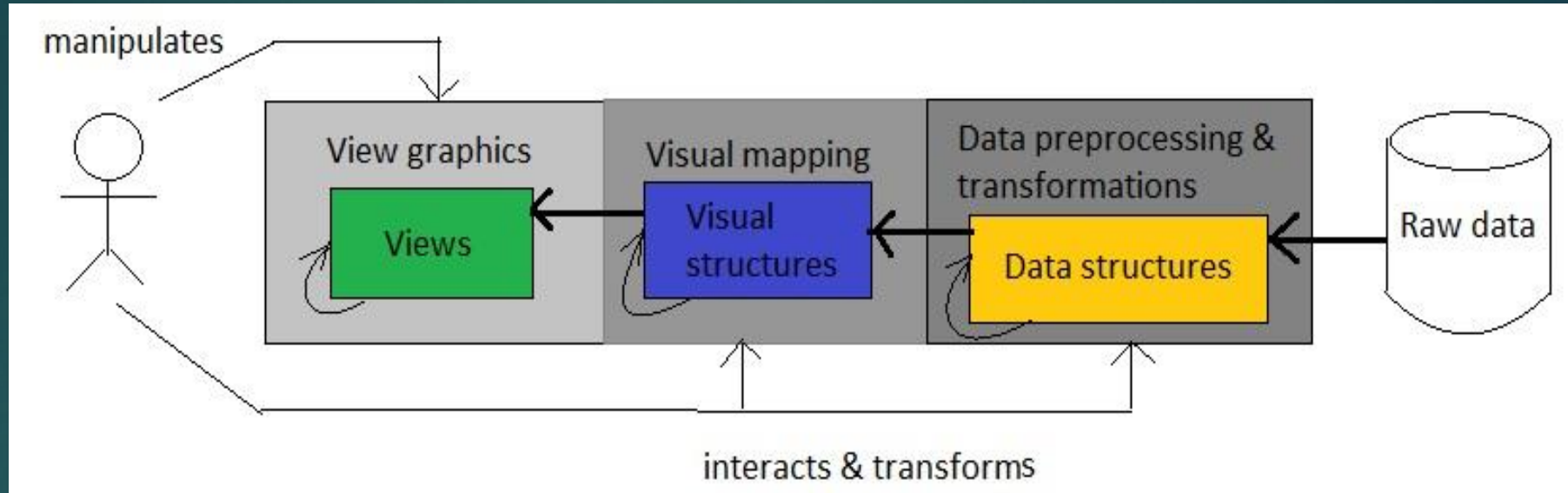
IV. Life Cycle of Visualizations

- ▶ Static representations:
 - ▶ Code Poetry
- ▶ Manipulable representations:
 - ▶ LiquiData
- ▶ Transformable representations:
 - ▶ Regenradar (Wetteronline)
 - ▶ Malte Spitz (Zeit)



V. Conclusion

30



The 4th V of Big Data

Volume

Variety

Velocity

Data Visualization Tools

Enable your employees to:

- Analyze and visualize real-time data on their own
- Collaborate using online graphics to generate ideas and identify trends

Value

Empower your business to:

- Achieve greater revenue opportunities
- Create market-leading innovations
- Improve customer experiences

[Big Data Visualization: Turning Big Data into Big Insights, Intel (2013)]

VI. Key Persons, Sources & Conferences

31

- ✓ IEEE Symposium on Information Visualization (InfoVis). Since 1995. The main conference in the field, quite low acceptance rate (23% in 2006), very focussed, high quality papers, single-track. → <http://ieevis.org>
Proceedings published with IEEE:
→ <http://conferences.computer.org/infovis/Eurographics/IEEE>
 - ✓ Symposium on Visualization (EuroVis). Formerly VisSym. Fairly high quality. → <http://eurovis.org>
Proceedings published with Eurographics:
→ <http://www.eg.org/EG/DL/WS/VisSym>
 - ✓ International Conference on Information Visualisation (IV). Since 1997, usually in London. Broad in scope, fairly high acceptance rate (57% in 2007), papers of mixed quality, multi-track.
→ <http://www.graphicslink.co.uk/IV2013/>
Proceedings published with IEEE:
→ <http://ieeexplore.ieee.org/servlet/opac?punumber=1000370>
 - ✓ The First IEEE Big Data Visualization Workshop.
→ <http://vis.ucdavis.edu/Workshops/BigDataVis2013/>
- ▶ Hans Rosling (Gapminder Foundation, Karolinska Institute, Sweden)
 - ▶ Boris Müller (FH Potsdam)
 - ▶ Edward Tufte
 - ▶ Moritz Stefaner (FH Potsdam)
 - ▶ William S. Cleveland
 - ▶ Robert Spence (Imperial College London)
 - ▶ Stephen Few
 - ▶ Ben Fry
 - ▶ Ricardo Mazza