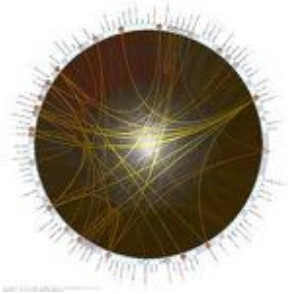
Information Visualization

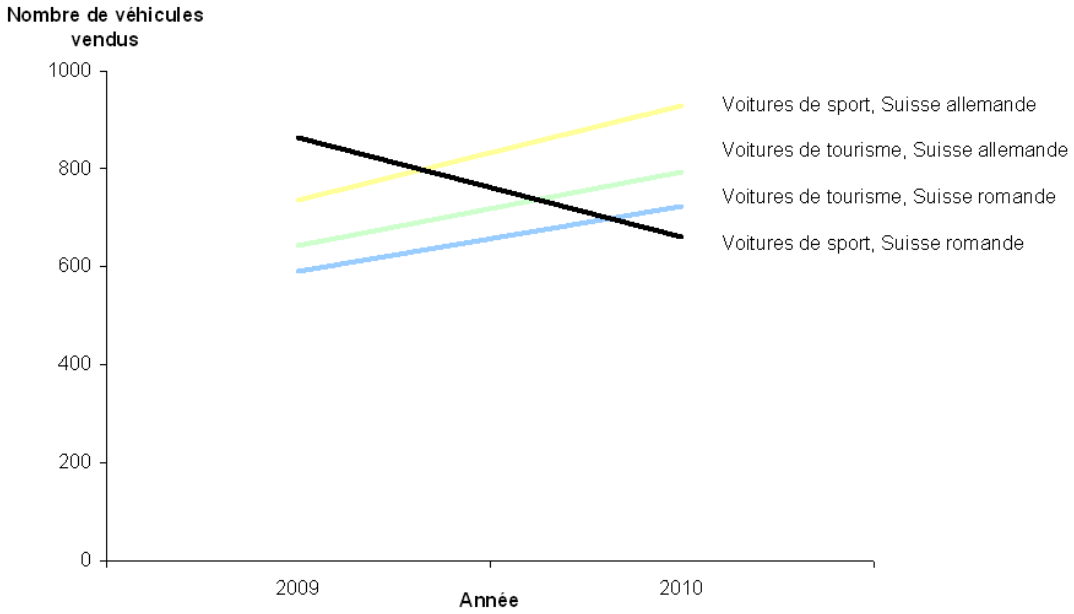


Denis Lalanne
University of Fribourg
07/05/2019

Nombre de véhicules vendus par type, région et année

	Suisse allemande		Suisse romande	
Type	2009	2010	2009	2010
Voitures de tourisme	643	793	590	724
Voitures de sport	735	928	863	662

Nombre de véhicules vendus par type, région et année



Motivation - Data Overload

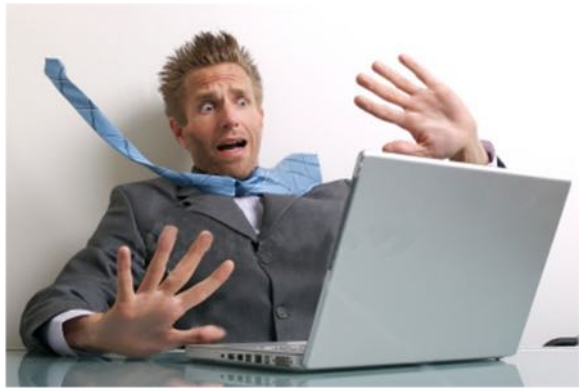
- How much data?
  - 2.5 exabytes of unique data produced every day!
    - ✓ 2.5* 10000000000000000 (10^{18}) bytes
    - ✓ Printed documents only .003% of total
- How to make use of the data?

Web,
Books,
Papers,
emails,
Scientific data,
Biotech,
Shopping
People
Stock/finance
News

Data transfert

How??

Vision: 100 MB/s
Ears: <100 b/s
Haptic/tactile
Smell
Taste



Human Vision

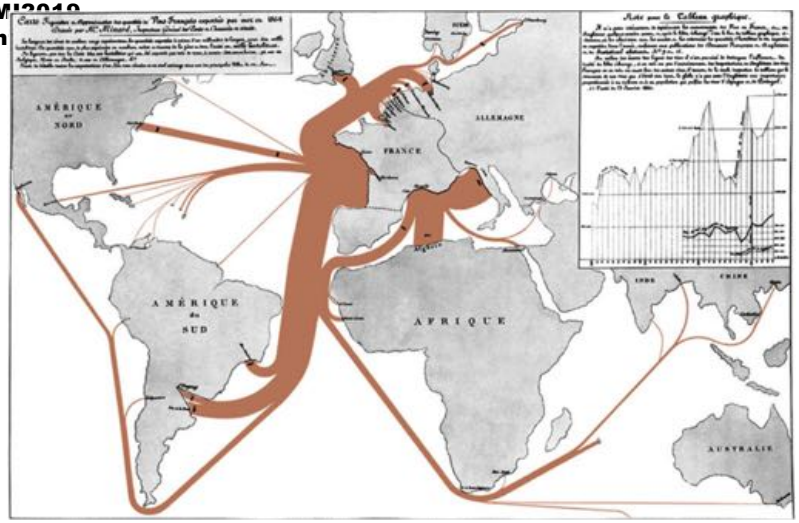
- Highest bandwidth sense
  - Fast, parallel
  - Pattern recognition
  - Extends memory and cognitive capacity
  - People think visually
- The Challenge
  - Transform the data into information (understanding, insight) thus making it useful to people
- Example:
  - Which state has the highest income?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

College degree %

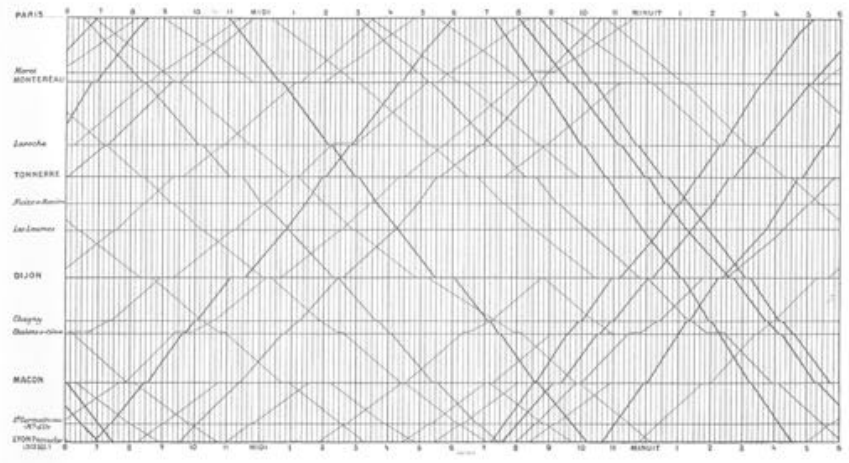
Per capita income

Visualization is an old practice …

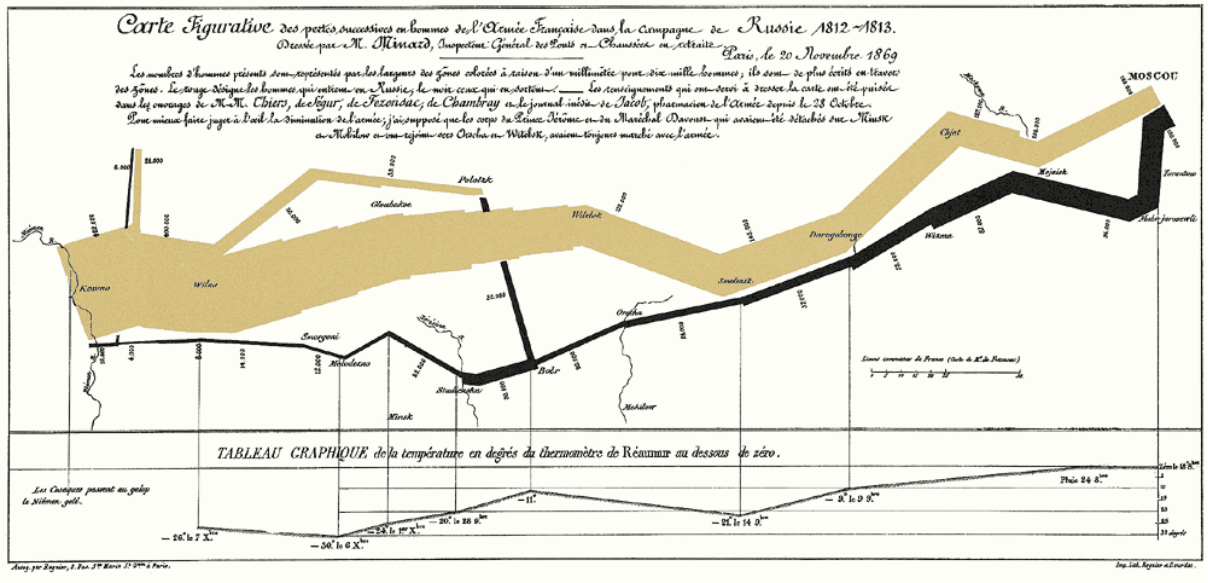
MMI2019
In



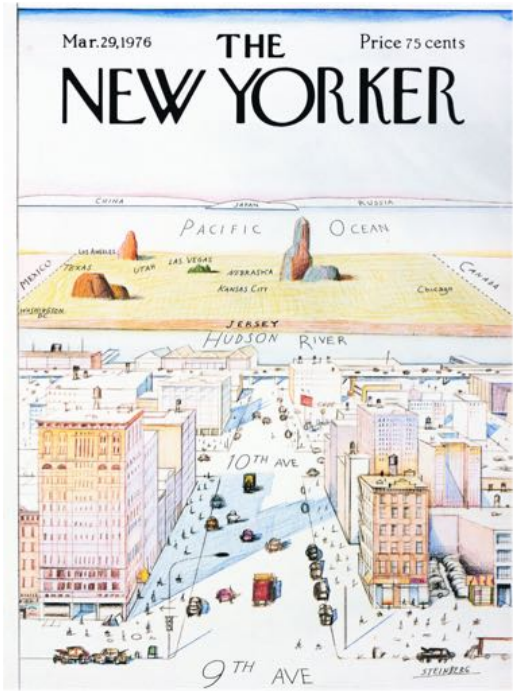
FRENCH WINE EXPORT (CHARLES MINARD, 1864)



TRAIN SCHEDULE BY E.J. MAREY (1885)



C.J.Minard's Napoleon's army in Russia (drawn in 1861)



VIEW OF THE WORLD FROM 9TH AVENUE

[STEINBERG 1976]

The classic definition

Information visualization is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition. [Card, Mackinlay, Shneiderman]

- "The purpose of visualization is insight, not pictures"
  - Insight: discovery, decision making, explanation

Information visualization

- What is “information”?
  - ✓ Items, entities, things which do not have a direct physical correspondence
  - ✓ Examples: baseball statistics, stock trends, connections between criminals, car attributes...

- What It’s Not InfoViz?
  - ✓Scientific Visualization
    - ✓ Primarily relates to and represents something physical or geometric
    - ✓ Examples: Air flow over a wing, weather over Pennsylvania

Rough operational definition

get some data and give them a visual representation

(data items -> visual items, data features -> visual features)

+

provide some interaction facilities to *explore* and *focus* on interesting subsets

From raw data …

```
echo cars.data 1>&2
sed 's/.//' >cars.data <<'//GO.SYSIN DD cars.data'
-18.0   8.   307.0      130.0      3504.      12.0   70.  1.
-15.0   8.   350.0      165.0      3693.      11.5   70.  1.
-18.0   8.   318.0      150.0      3436.      11.0   70.  1.
-16.0   8.   304.0      150.0      3433.      12.0   70.  1.
-17.0   8.   302.0      140.0      3449.      10.5   70.  1.
-15.0   8.   429.0      198.0      4341.      10.0   70.  1.
-14.0   8.   454.0      220.0      4354.       9.0   70.  1.
-14.0   8.   440.0      215.0      4312.       8.5   70.  1.
-14.0   8.   455.0      225.0      4425.      10.0   70.  1.
-15.0   8.   390.0      190.0      3850.       8.5   70.  1.
-NA     4.   133.0      115.0      3090.      17.5   70.  2.
-NA     8.   350.0      165.0      4142.      11.5   70.  1.
-NA     8.   351.0      153.0      4034.      11.0   70.  1.
-NA     8.   383.0      175.0      4166.      10.5   70.  1.
-NA     8.   360.0      175.0      3850.      11.0   70.  1.
-15.0   8.   383.0      170.0      3563.      10.0   70.  1.
-14.0   8.   340.0      160.0      3609.       8.0   70.  1.
-NA     8.   302.0      140.0      3353.       8.0   70.  1.
-15.0   8.   400.0      150.0      3761.       9.5   70.  1.
-14.0   8.   455.0      225.0      3086.      10.0   70.  1.
-24.0   4.   113.0      95.00      2372.      15.0   70.  3.
-22.0   6.   198.0      95.00      2833.      15.5   70.  1.
-18.0   6.   199.0      97.00      2774.      15.5   70.  1.
-21.0   6.   200.0      85.00      2587.      16.0   70.  1.
-27.0   4.   97.00      88.00      2130.      14.5   70.  3.
-26.0   4.   97.00      46.00      1835.      20.5   70.  2.
-25.0   4.   110.0      87.00      2672.      17.5   70.  2.
-24.0   4.   107.0      90.00      2430.      14.5   70.  2.
-25.0   4.   104.0      95.00      2375.      17.5   70.  2.
-26.0   4.   121.0      113.0      2234.      12.5   70.  2.
-21.0   6.   199.0      90.00      2648.      15.0   70.  1.
-10.0   8.   360.0      215.0      4615.      14.0   70.  1.
-10.0   8.   307.0      200.0      4376.      15.0   70.  1.
-11.0   8.   318.0      210.0      4382.      13.5   70.  1.
-9.0    8.   304.0      193.0      4732.      18.5   70.  1.
-27.0   4.   97.00      88.00      2130.      14.5   71.  3.
-28.0   4.   140.0      90.00      2264.      15.5   71.  1.
-25.0   4.   113.0      95.00      2228.      14.0   71.  3.
-25.0   4.   98.00      NA         2046.      19.0   71.  1.
-NA     4.   97.00      48.00      1978.      20.0   71.  2.
-19.0   6.   232.0      100.0      2634.      13.0   71.  1.
-16.0   6.   225.0      105.0      3439.      15.5   71.  1.
-17.0   6.   250.0      100.0      3329.      15.5   71.  1.
-19.0   6.   250.0      88.00      3302.      15.5   71.  1.
-18.0   6.   232.0      100.0      3288.      15.5   71.  1.
-14.0   8.   350.0      165.0      4209.      12.0   71.  1.
-14.0   8.   400.0      175.0      4464.      11.5   71.  1.
-14.0   8.   351.0      153.0      4154.      13.5   71.  1.
```

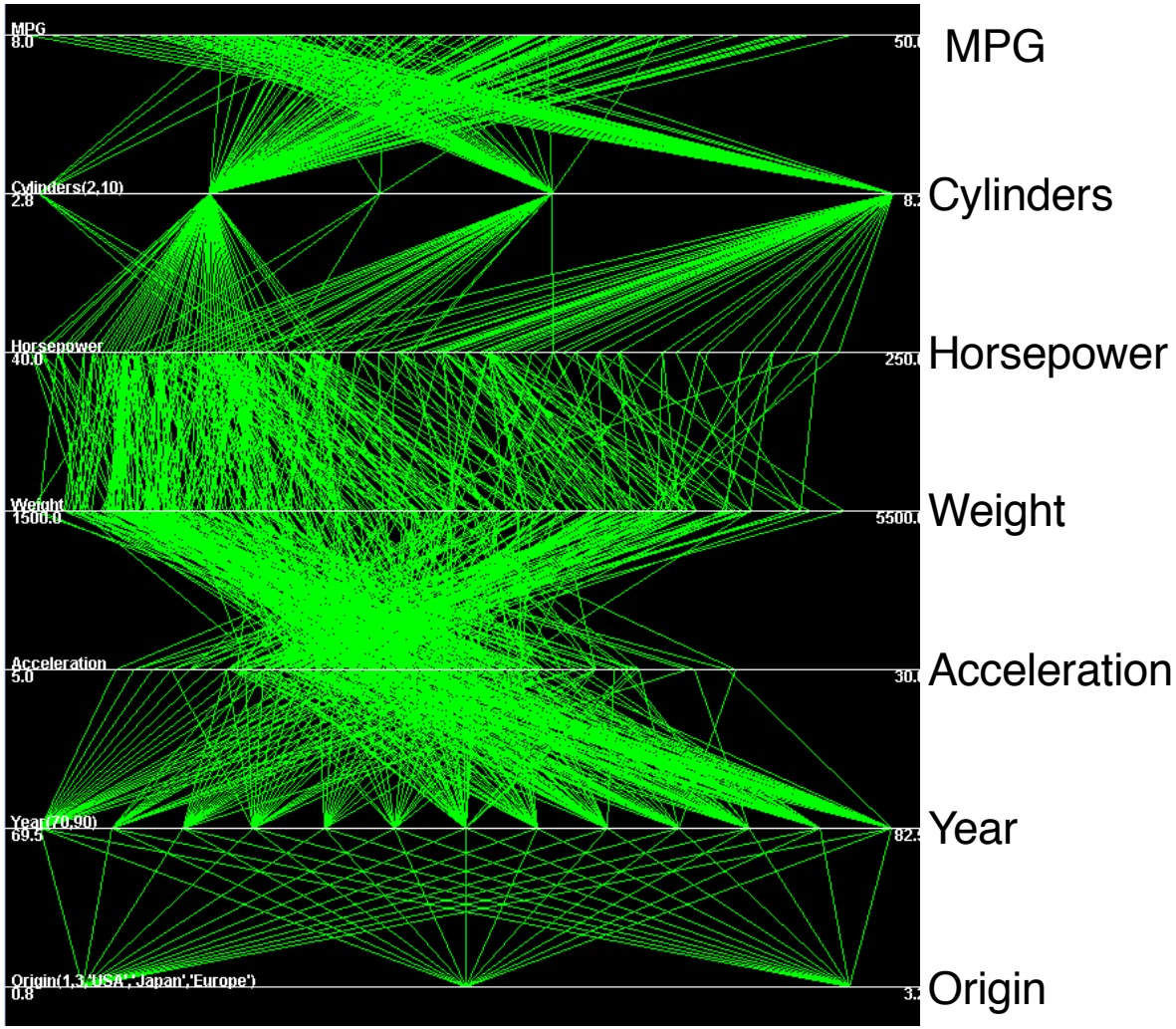
- **Car Dataset**

- 7 dimensions (391 cars)
  - miles per gallon (M.P.G.)
  - number of cylinders
  - horsepower
  - weight
  - acceleration (time from 0 to 60 mph)
  - year
  - origin (USA, Europe, Japan)

… to visualization

- Cars dataset: **parallel coordinates** depicting *trends* over 7 dimensions.

- 7 dimensions (391 cars):
  - miles per gallon (M.P.G.)
  - number of cylinders
  - horsepower
  - weight
  - acceleration (time from 0 to 60 mph)
  - year
  - origin (USA, Europe, Japan)
- Gain meaning,
  - e.g. weight inversely proportional to acceleration



Why is Infovis useful?

- We need to make sense of phenomena described by data, but raw data is hard/impossible to handle by humans
- Infovis is perfect for exploration, perfect when the goal is vague
- Other approaches to data analysis make *exploration* difficult
  - Statistics:
    (+) strong verification
    (-) does not support exploration and vague goals
  - Data mining:
    (+) actionable and reliable
    (-) black box style, not interactive (question-response)

Roles of InfoVis

- Exploration
  - Prerequisite: domain knowledge
  - Outcome: new hypothesis
- Confirmation
  - Prerequisite: hypothesis
  - Outcome: confirmation/rejection (new hypothesis)
- Communication
  - Prerequisite: confirmed hypothesis
  - Outcome: clear visualization

[Keim KDD'02]

Visual Analytics' Motto: *detect the expected and discover the unexpected*

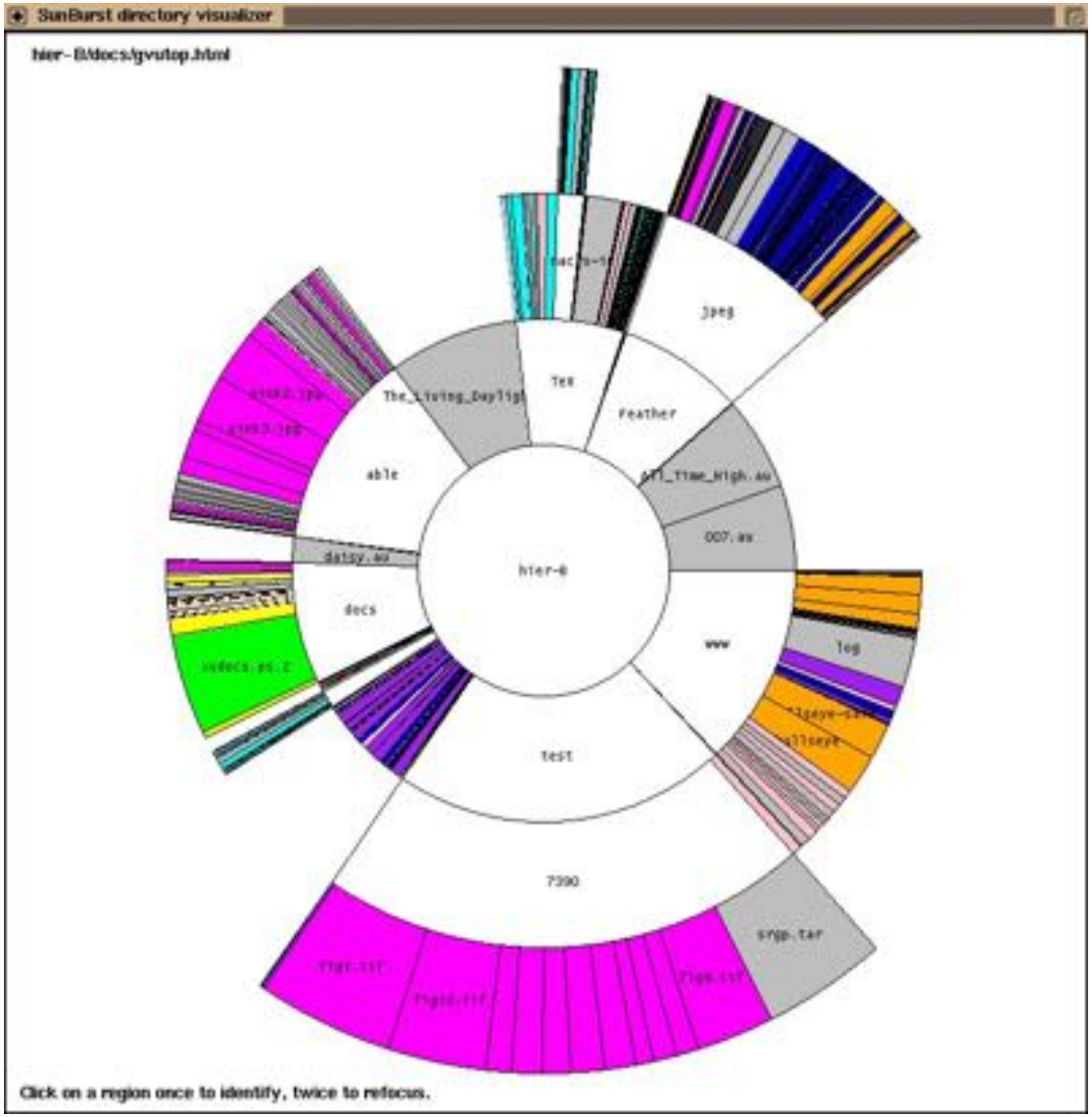
Tasks in InfoVis

- Search
  - Finding a specific piece of information
    - ✓ How many games did the Braves win in 1995?
    - ✓ What novels did Ian Fleming author?
- Browsing
  - Look over or inspect something in a more casual manner, seek interesting information
    - ✓ Learn about crystallography
- Analysis
  - Comparison-Difference
  - Extremes
  - Patterns
- Monitoring

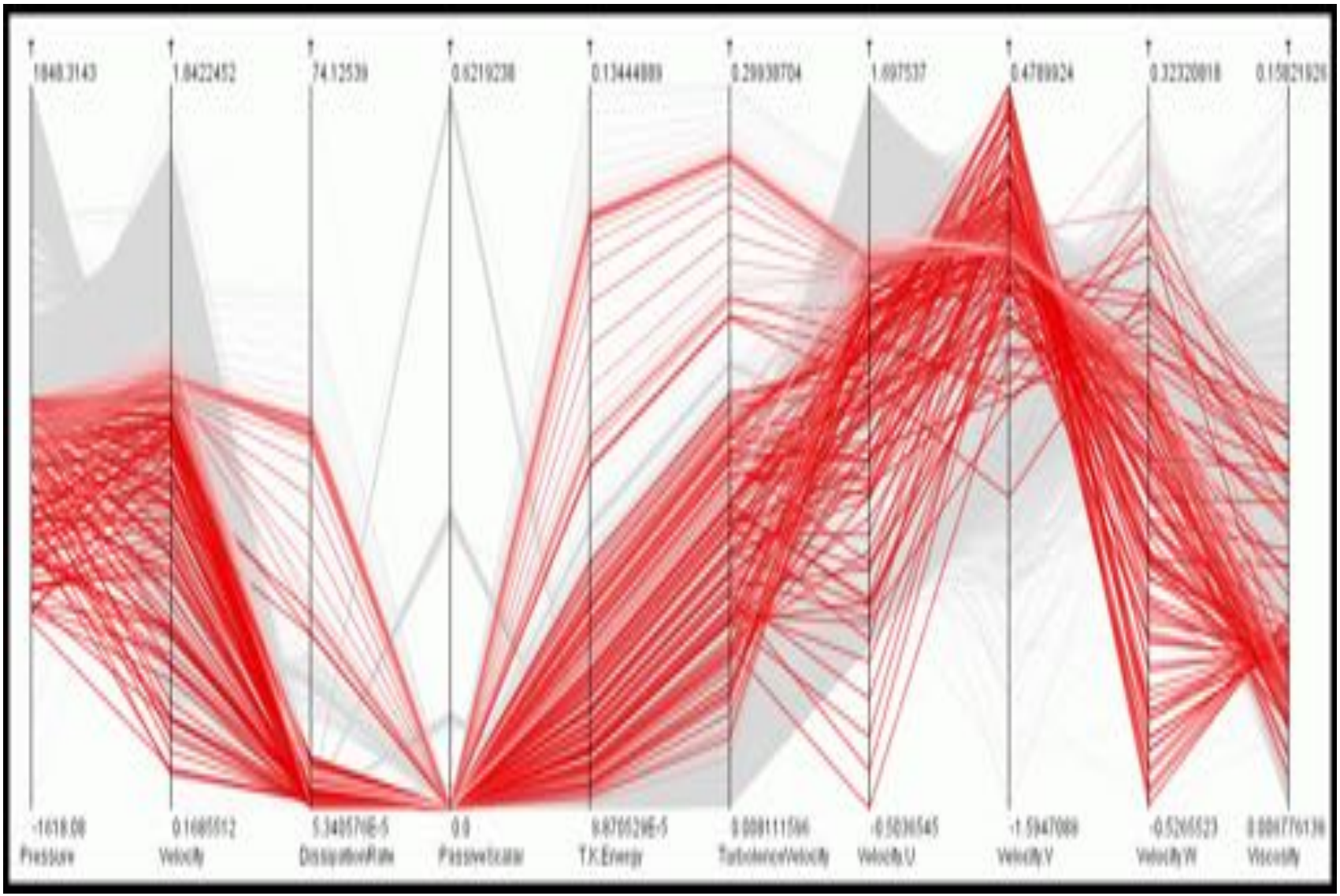
InfoVis Gallery …

Treemaps

Sunburst


SunBurst directory visualizer
hier-B/docs/gvutop.html
jpeg
tex
Feather
The_Living_Daylig
able
hier-0
docs
www
test
7390
rgp.tar
Click on a region once to identify, twice to refocus.

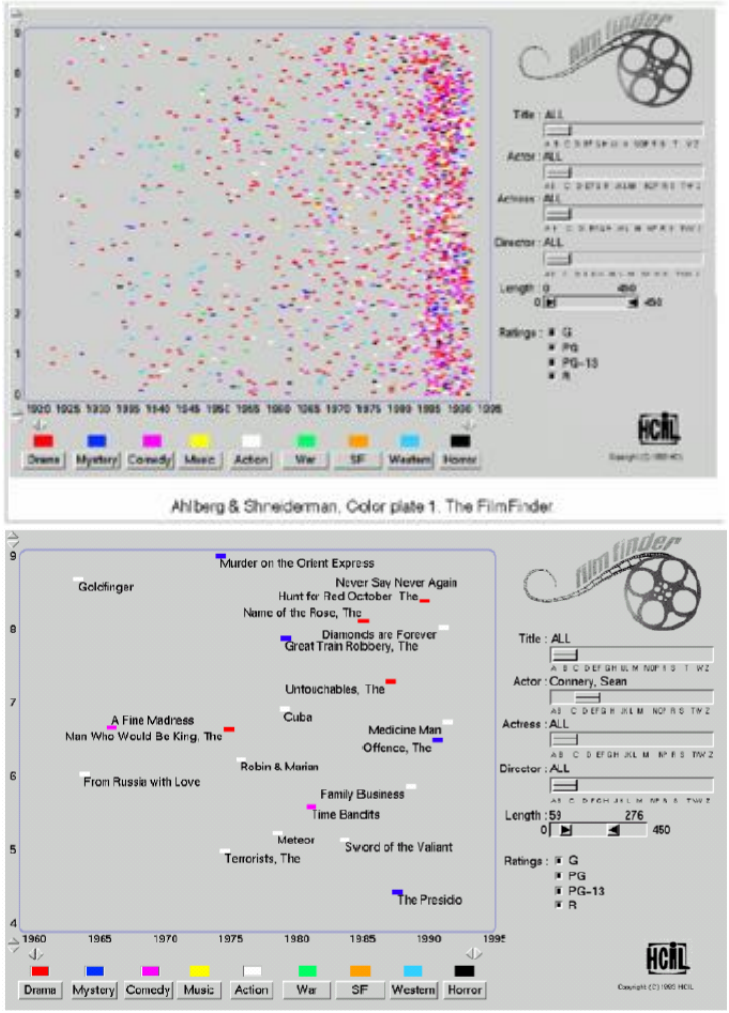
Parallel Coordinates



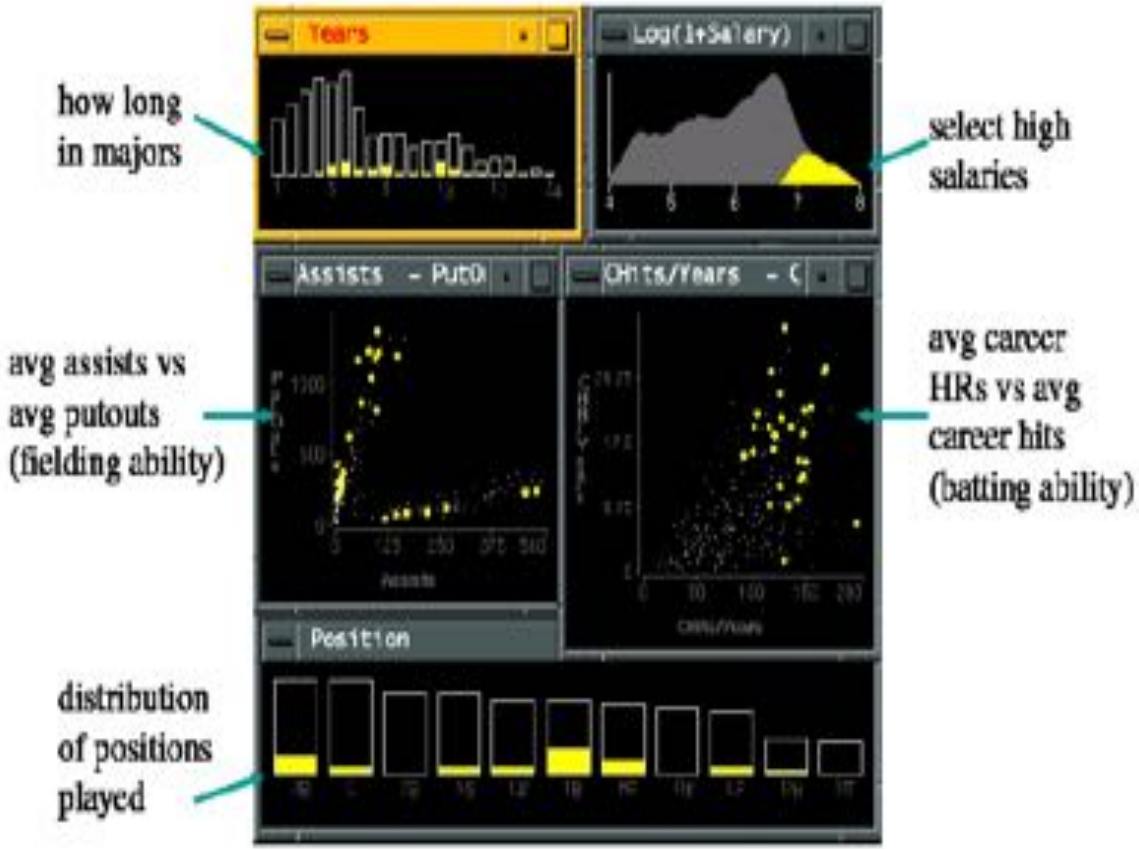
Graphs

Interaction …

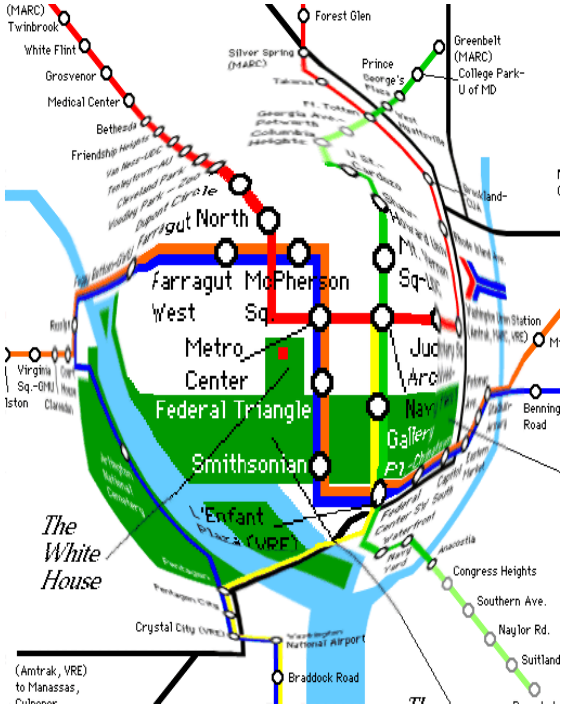
Filtering / Link & Brush



Ahlberg & Shneiderman, Color plate 1. The FilmFinder



Focus + Context



Unit	State	County	Output	Problems	Health
Unit45	Arizona	L	50	0	9
Unit46	Arizona	L	50	0	9
Unit47	Nebraska	V	90	2	9
Unit48	Nebraska	V	90	1	9
Unit49	Nebraska	V	50	2	8
Unit50	Nebraska	F	50	3	7
Unit51	Nebraska	F	70	0	9
Unit52	Nebraska	P	60	1	9
Unit53	Nebraska	P	50	1	8
Unit54	Nebraska	P	90	0	9
Unit55	Nebraska	P	90	0	9
Unit56	Nebraska	Q	90	0	9
Unit57	Nebraska	Q	90	1	9

How to represent high dimensional data?

- **Paper / display**
  - Spatial coordinates (2D or 3D)
  - Shape
  - Color
  - Size
  - Angle / Slope
  - Style / Texture
  - Time (motion)

8-D?

Human perception

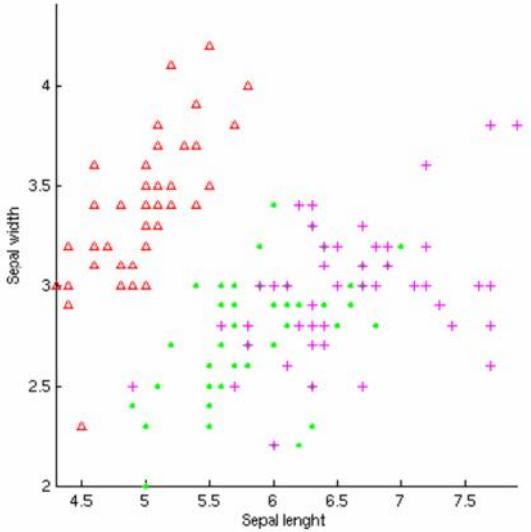
Limited!

Human analysis capabilites

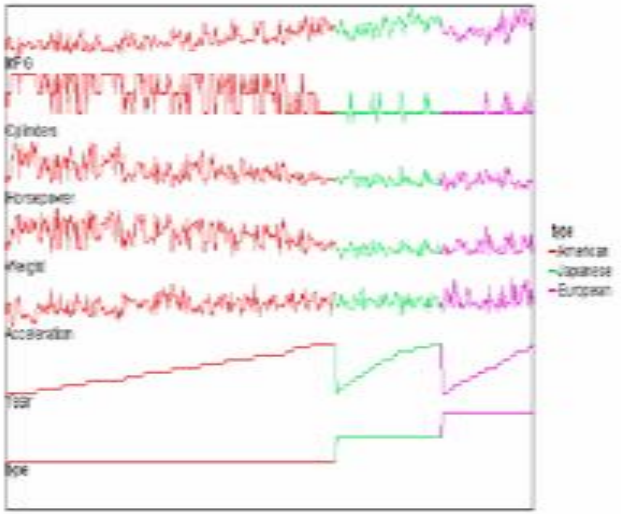
Limited!

Visualization techniques

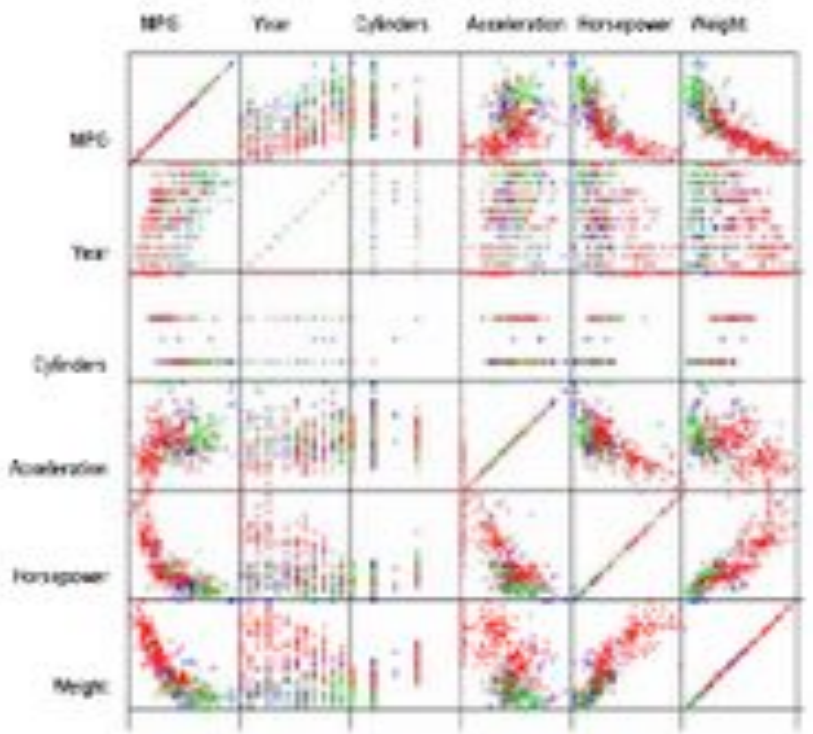
- Scatterplots



- Multi-line graphs

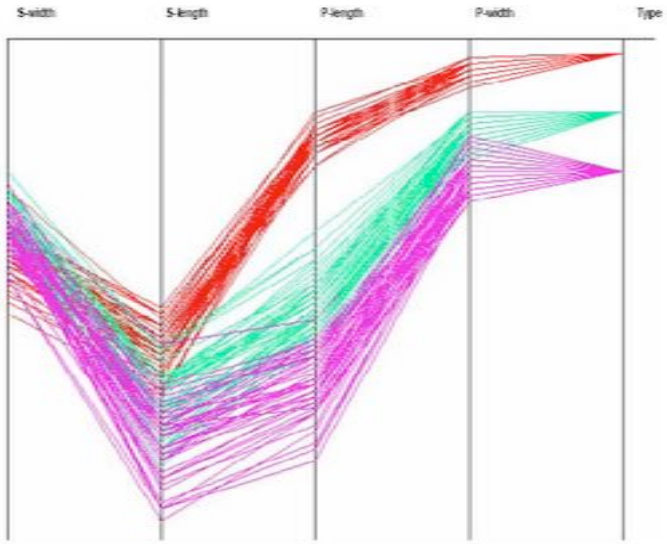


- Matrices of scatterplots

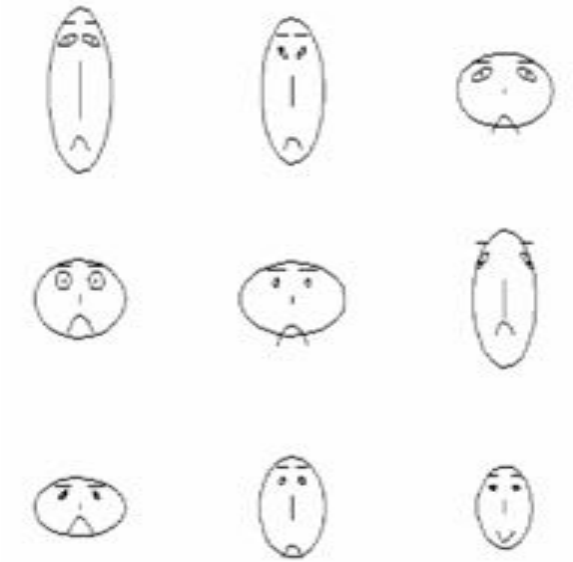


Visualization techniques

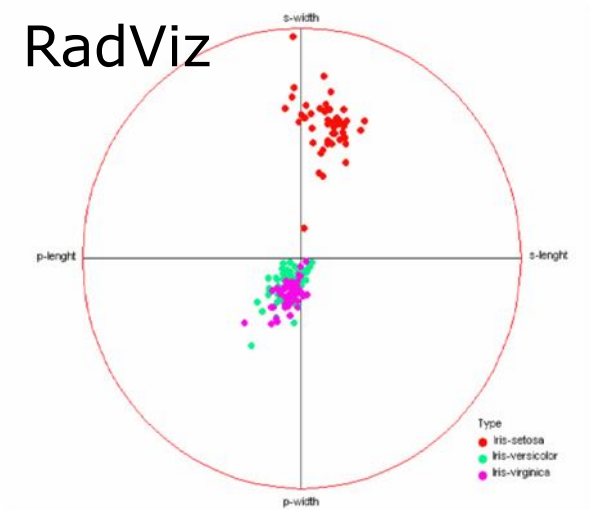
parallel coordinates



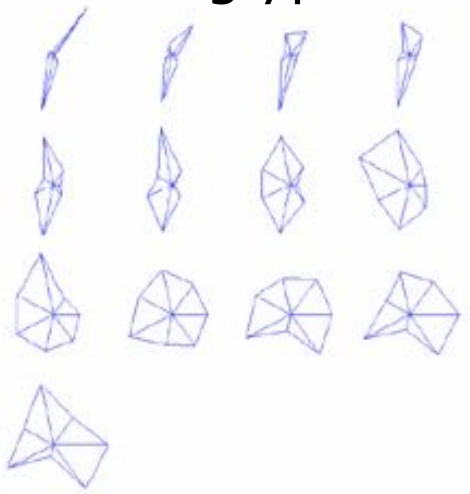
Chernoff faces



RadViz

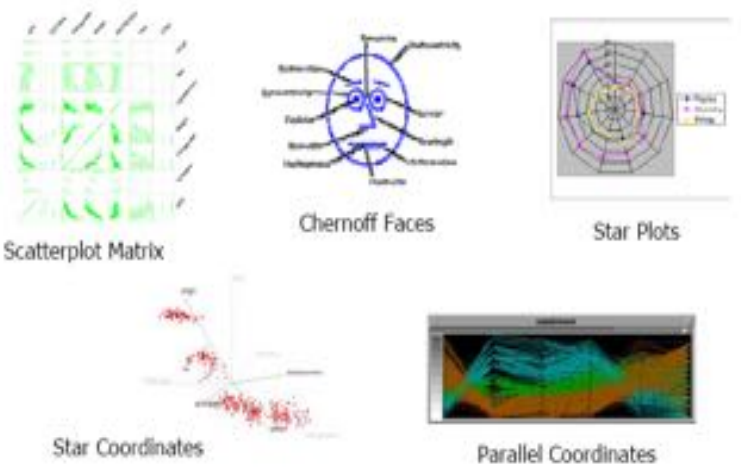


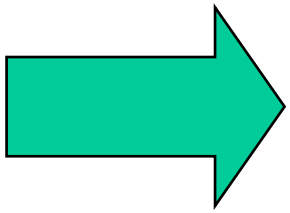
Star glyphs

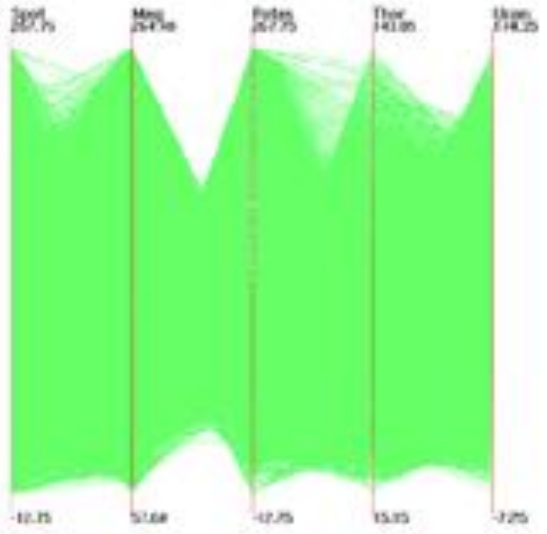


Why mining is necessary?

- What happens to the following visualizations when you have lots and lots of data cases?



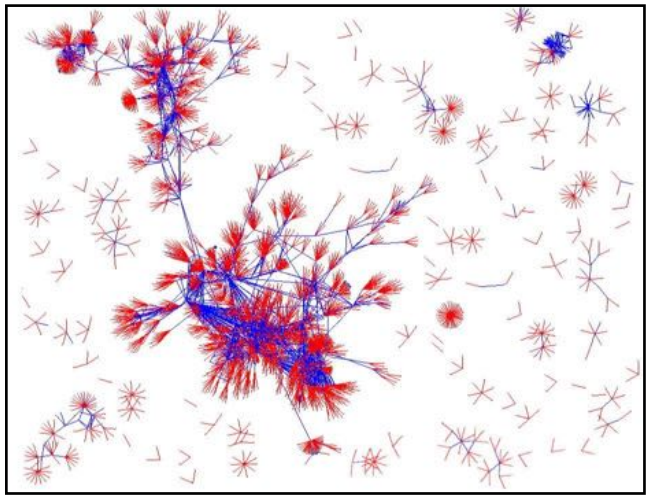




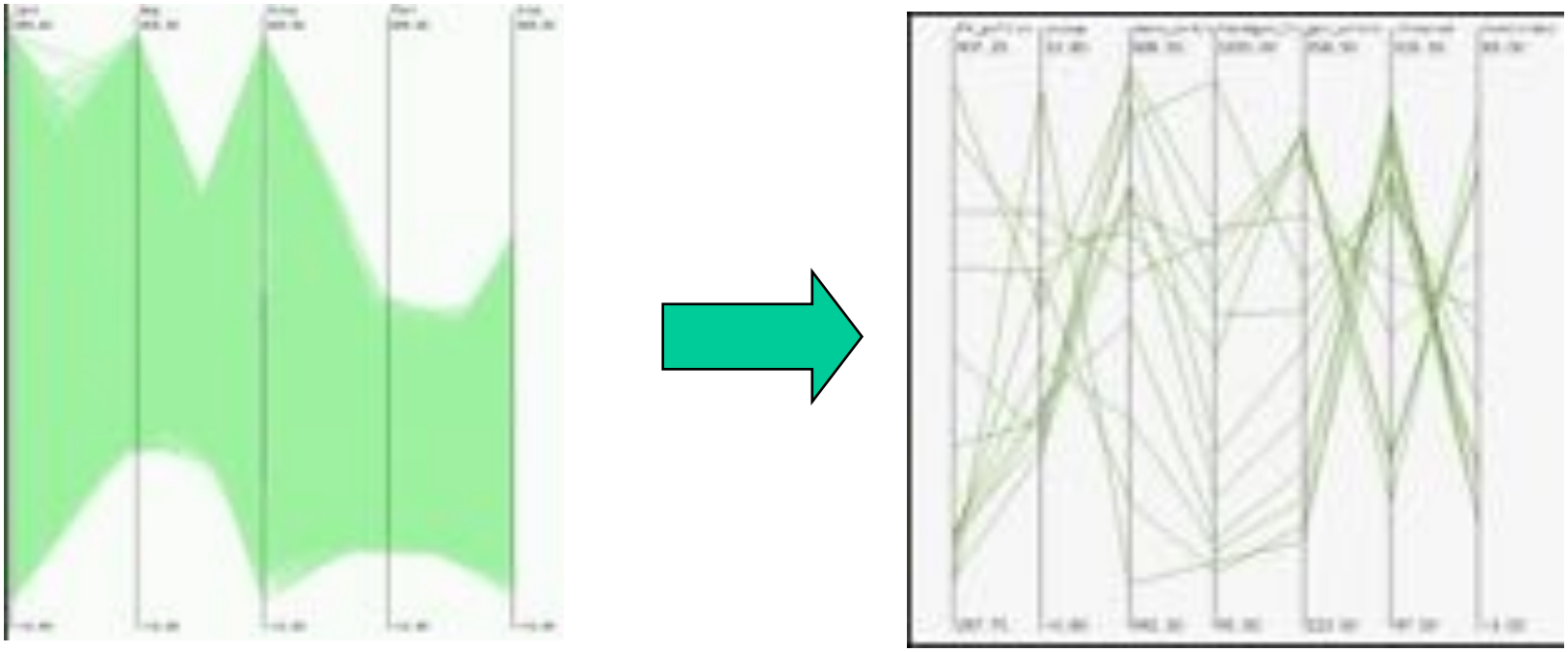
- => Becomes unreadable
  - (e.g. parallel coordinates)

Data mining necessary

- Reduce dimensions
  - Multi-dimensional scaling (MDS)
  - Principal component analysis
- Reduce data
  - Sampling – We only include every so many data cases or variables
  - Aggregation – We cluster together many data cases or variables
- Cluster data
  - Force Directed Algorithm (Spring Algorithm)
  - Hierarchical clustering methods
  - K-means

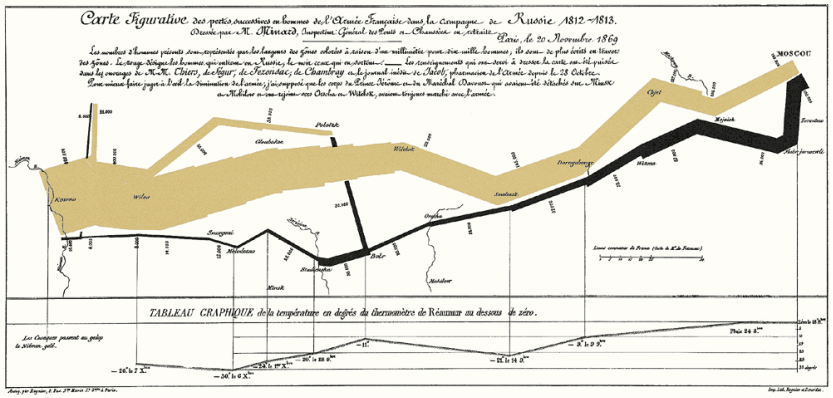


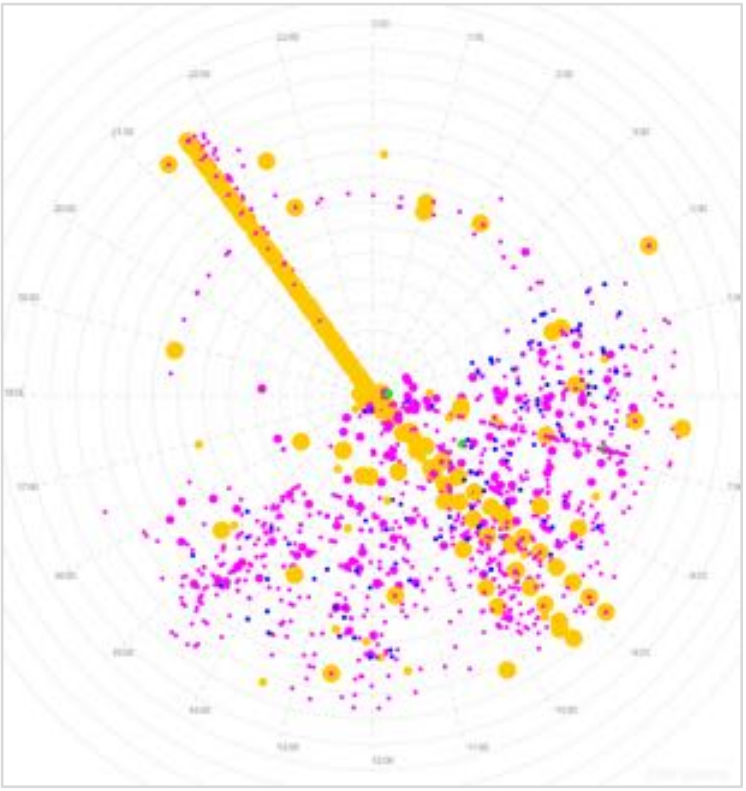
Hierarchical Parallel Coordinates

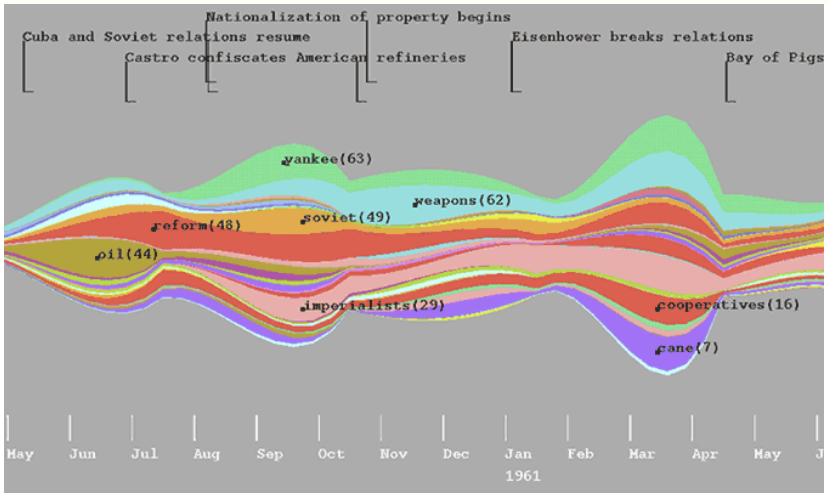


Temporal visualizations

- ✓ When was something greatest/least?
- ✓ Is there a trend?
- ✓ Are two series similar?
- ✓ Do any of the series match a pattern?







Humanitics:
An approach for visual exploration of temporal origin-destination data

Ilya Boyandin, Enrico Bertini, Denis Lalanne

UNHCR Refugee Dataset

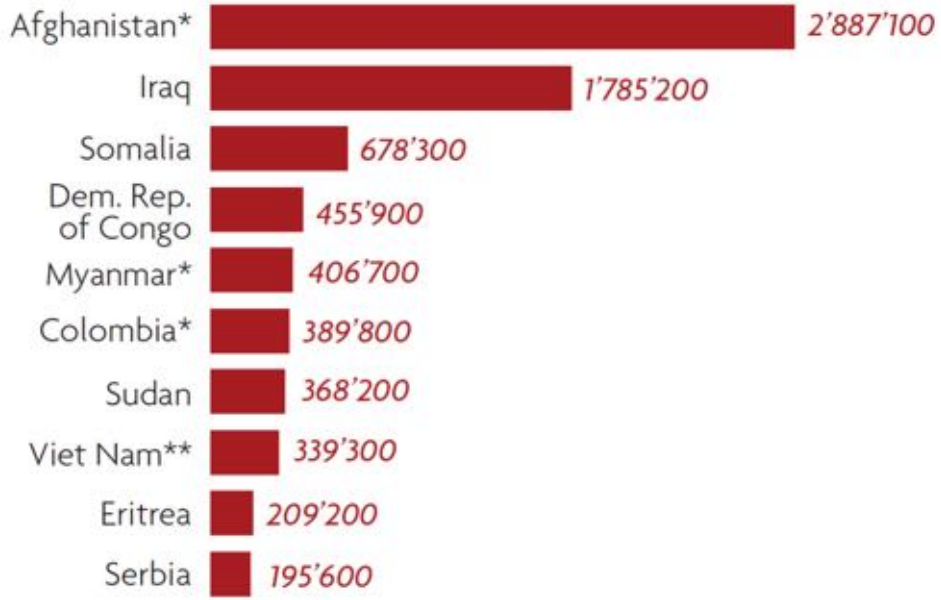
240 countries

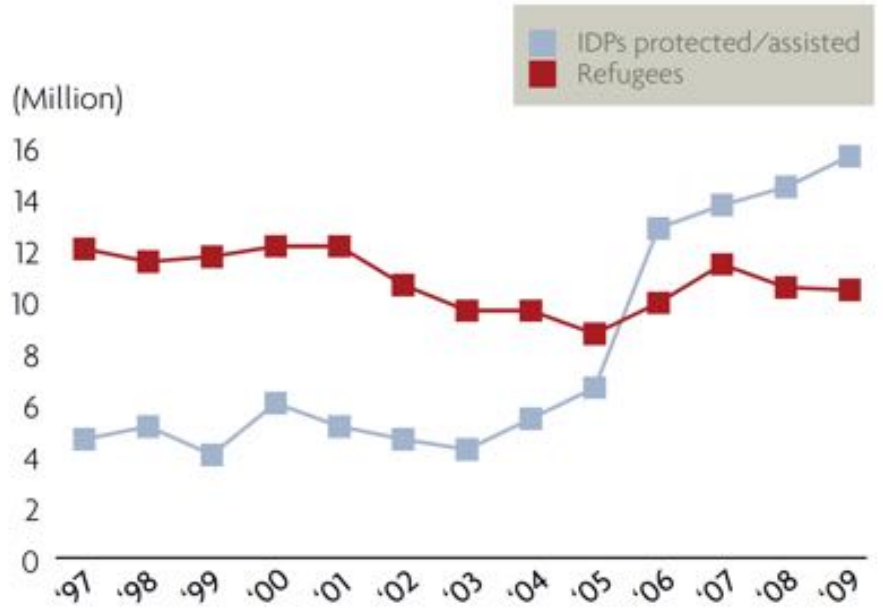
6000 flows (OD-pairs)

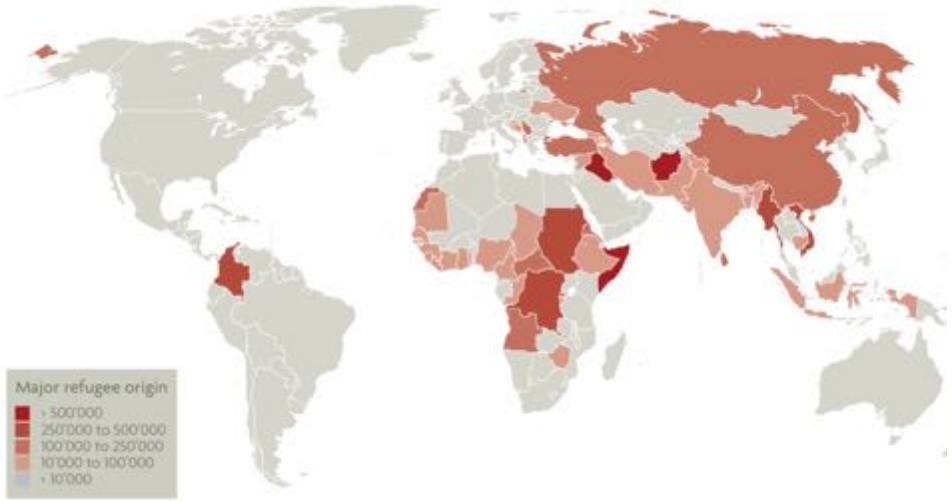
1975 to 2009

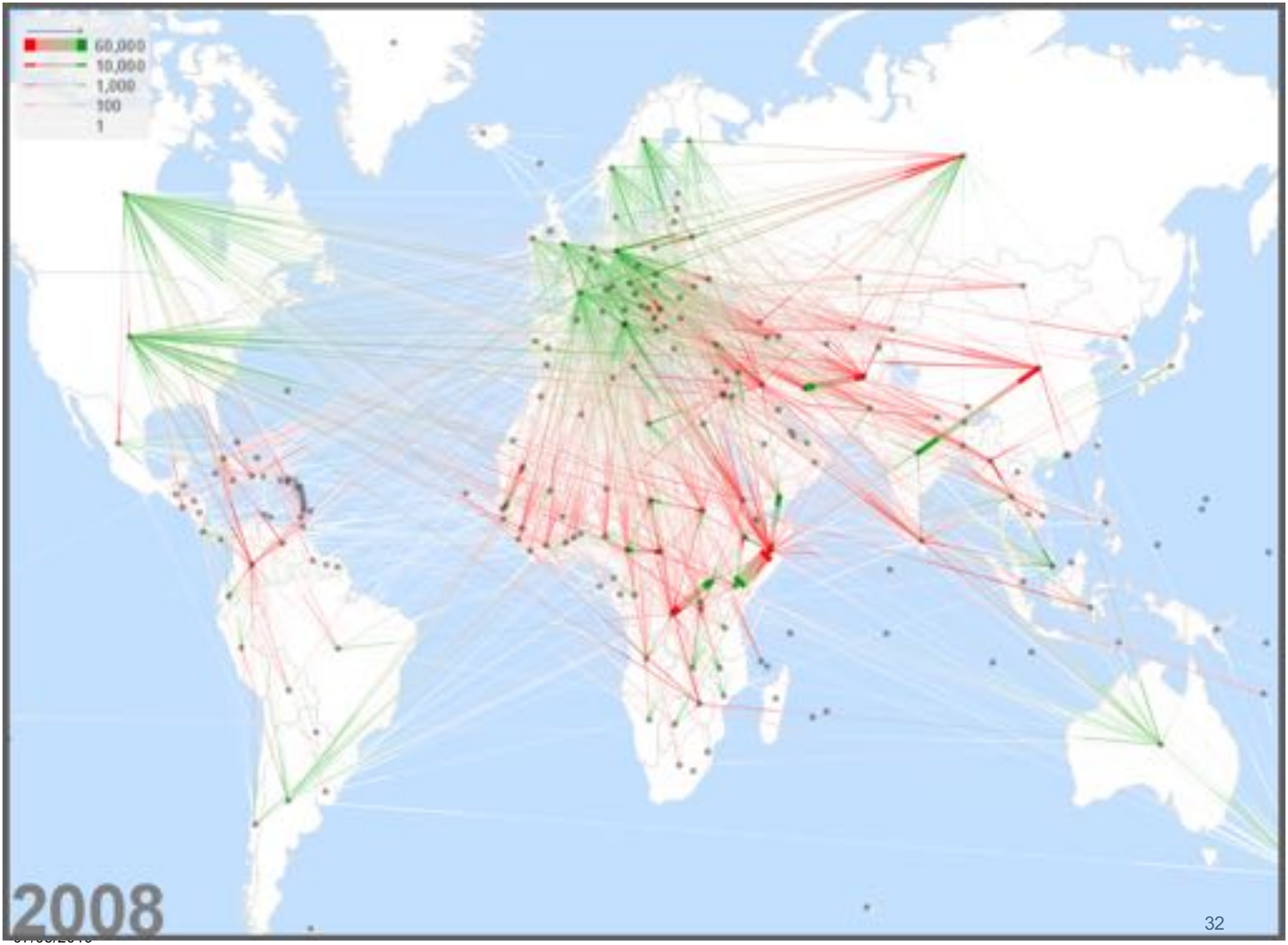
Millions of people

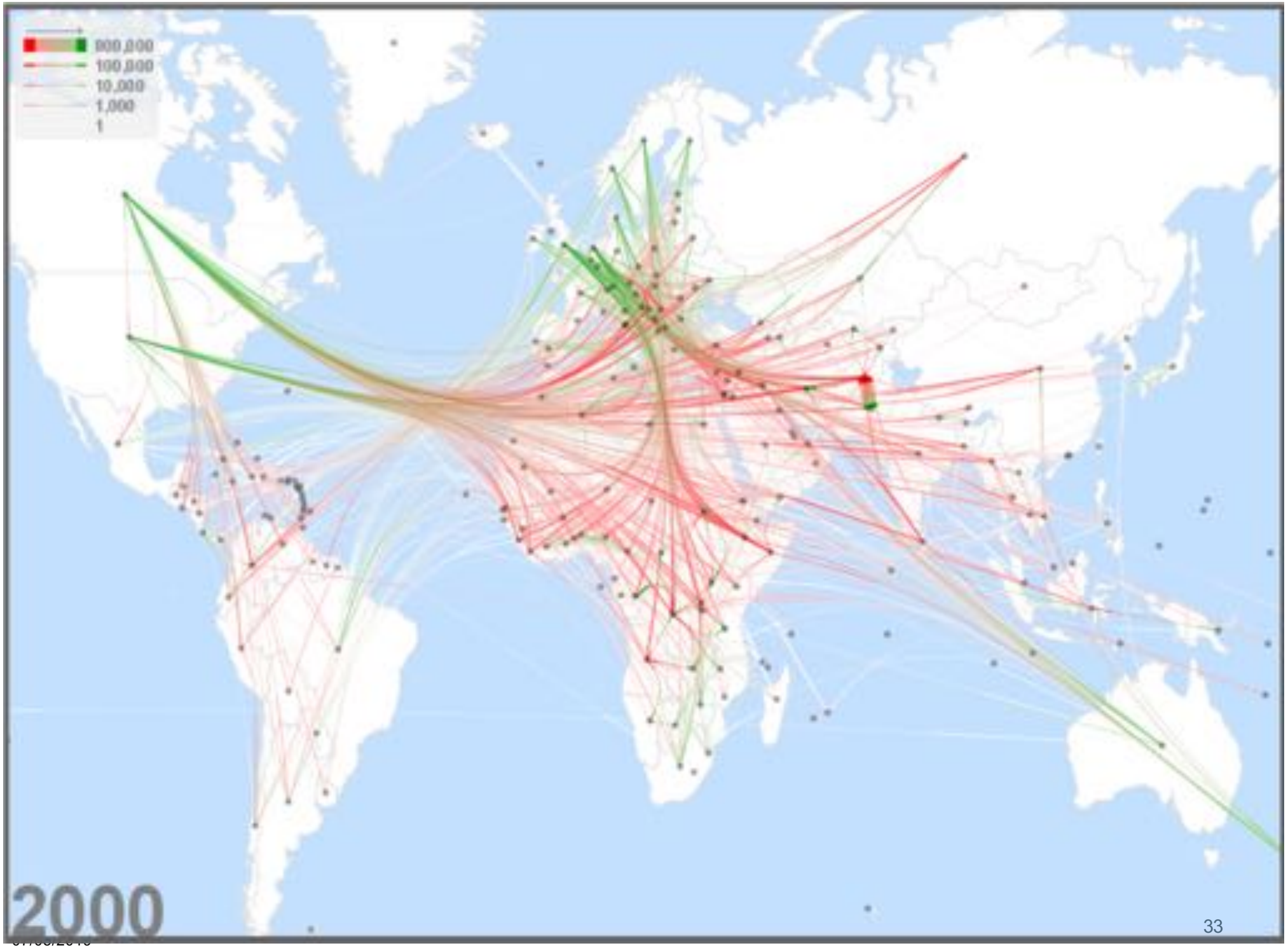
CHARTS THEY USE AT THE UN

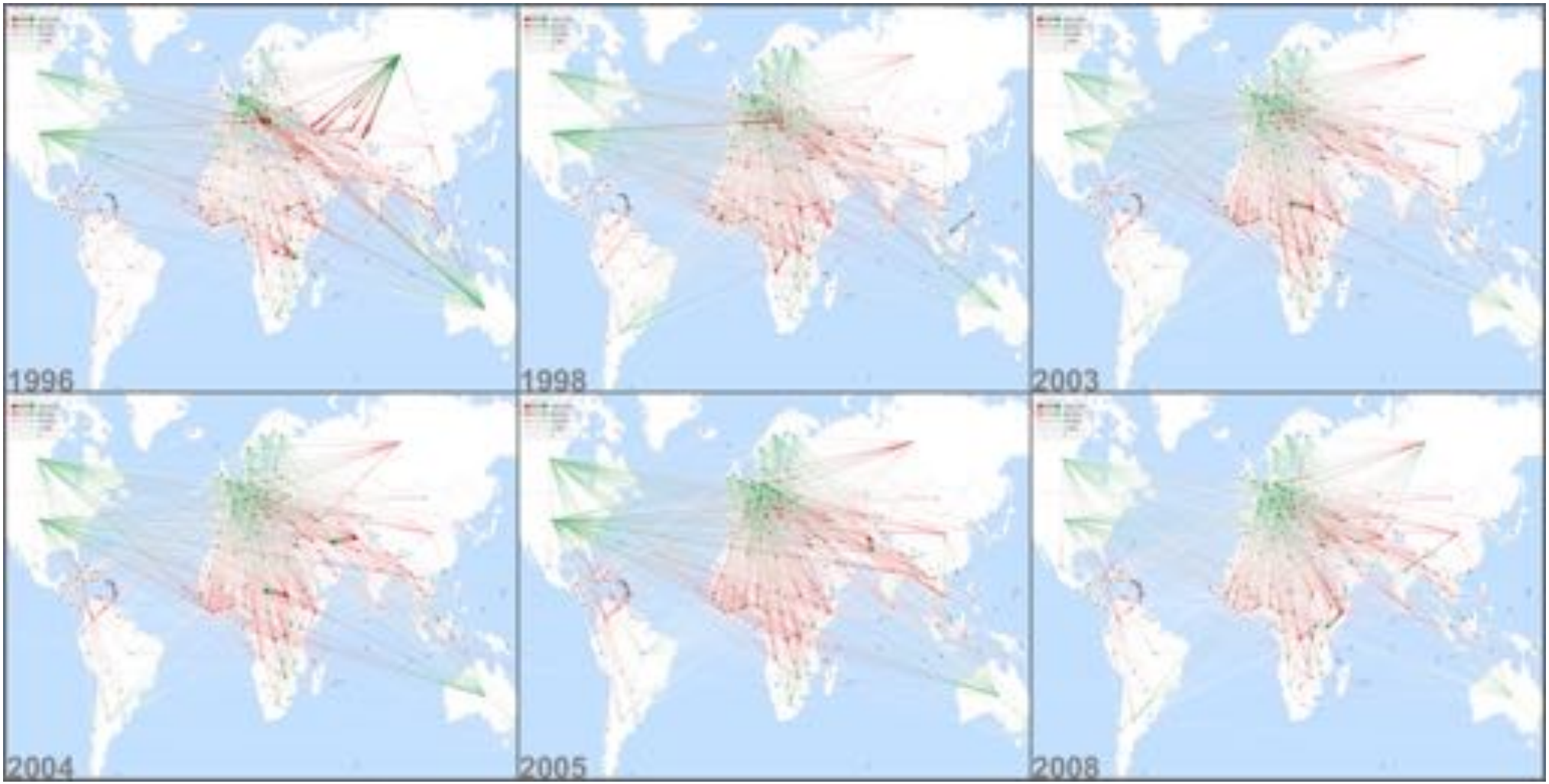


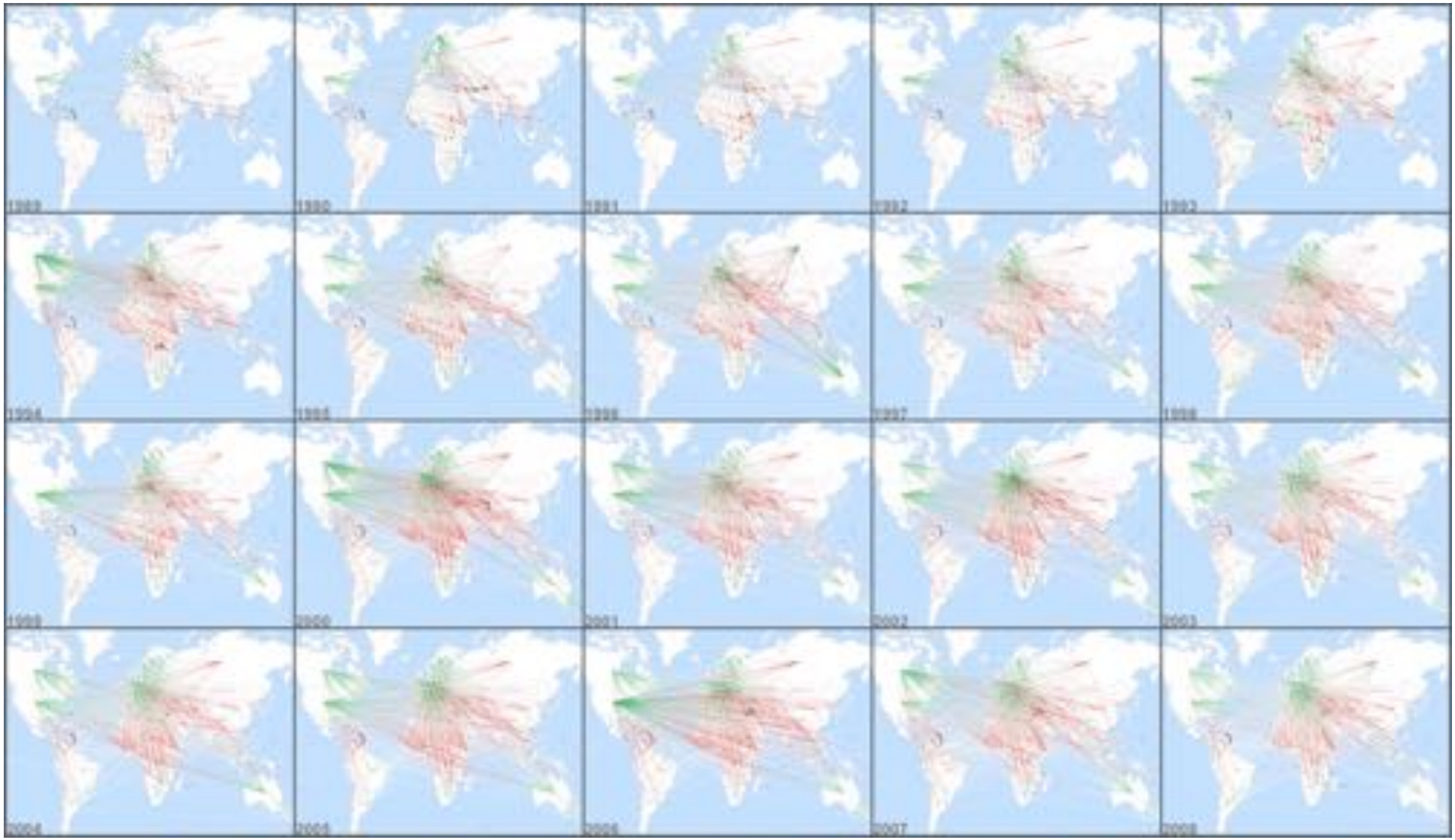




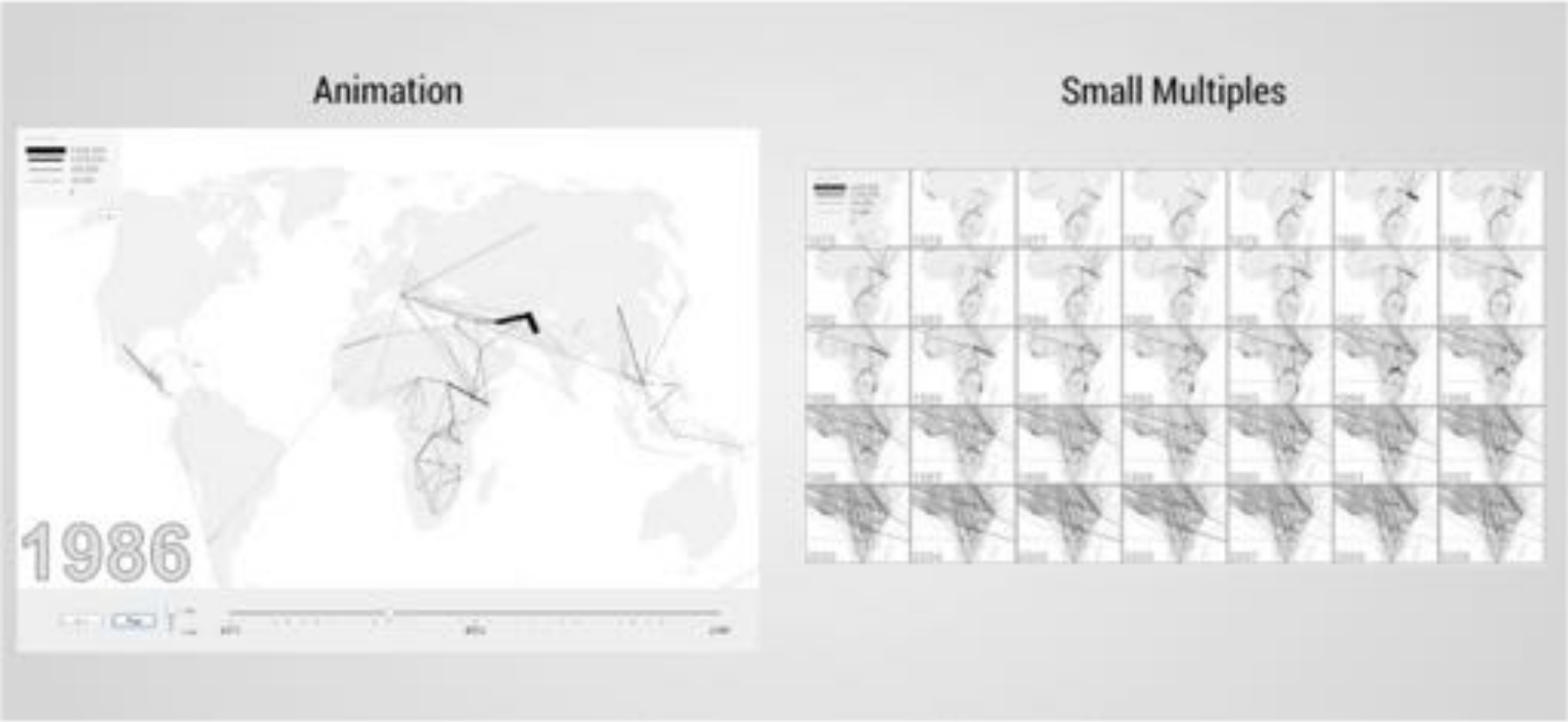

60,000
10,000
1,000
100
1
2008

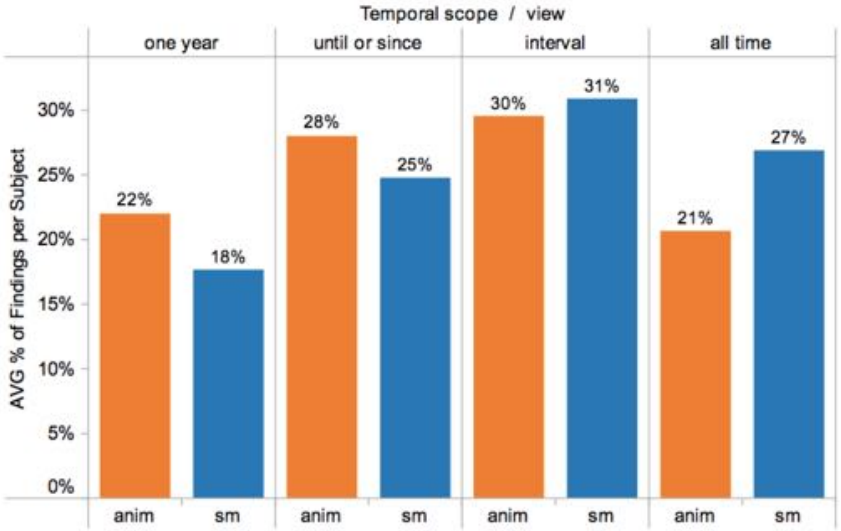

900,000
100,000
10,000
1,000
1
2000

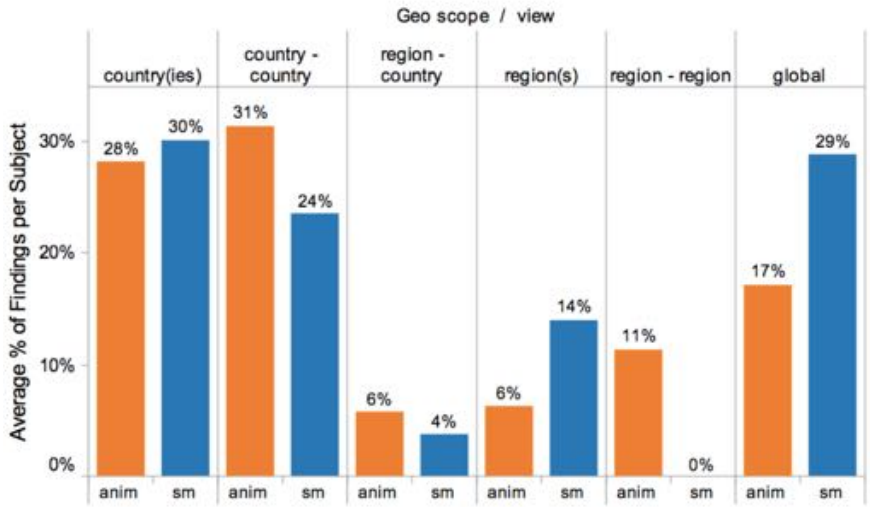

1996
1998
2003
2004
2005
2008



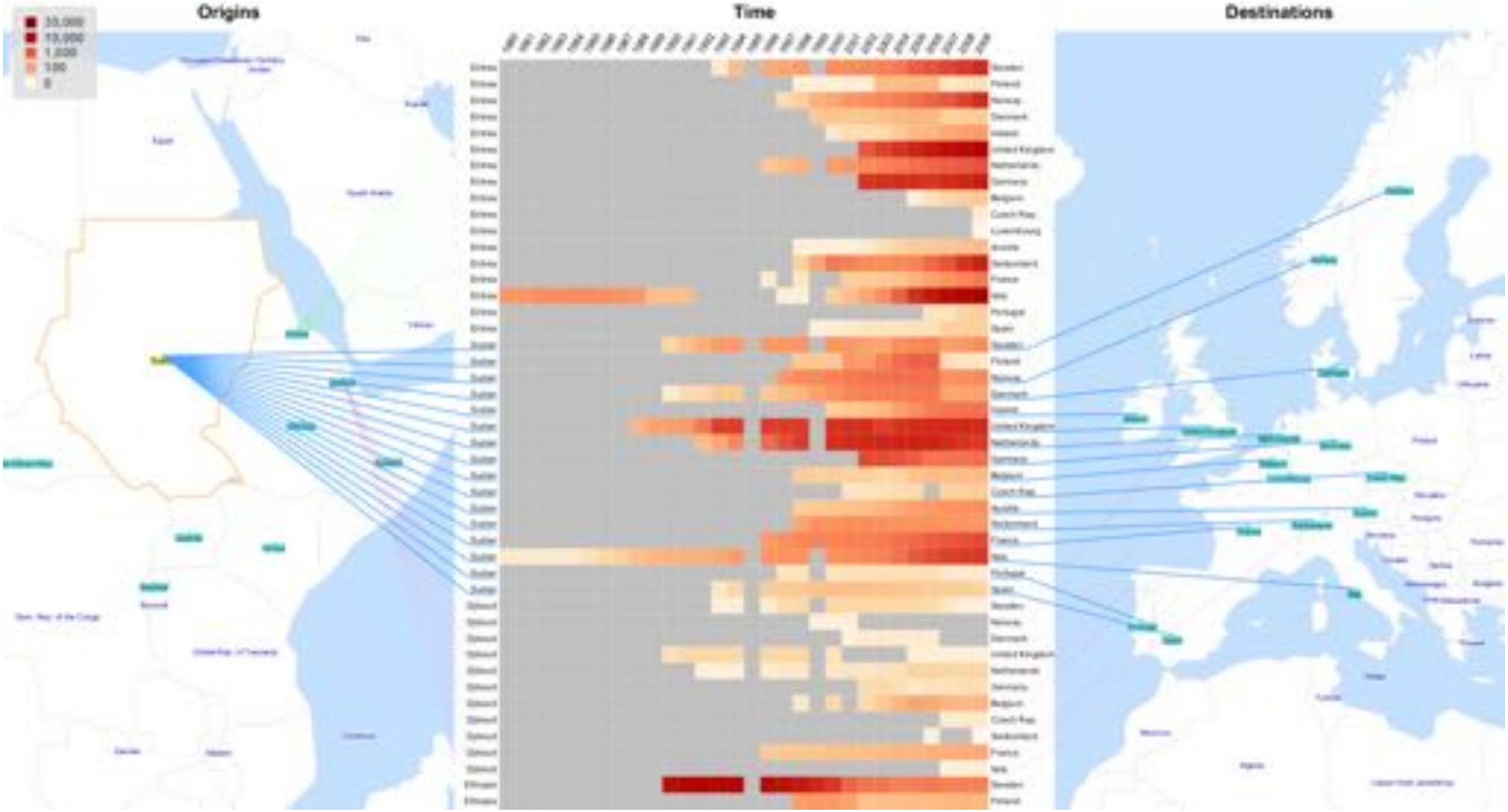
User Studies




Temporal scope / view
one year
until or since
interval
all time
AVG % of Findings per Subject
30%
25%
20%
15%
10%
5%
0%
22%
18%
28%
25%
30%
31%
21%
27%
anim
sm

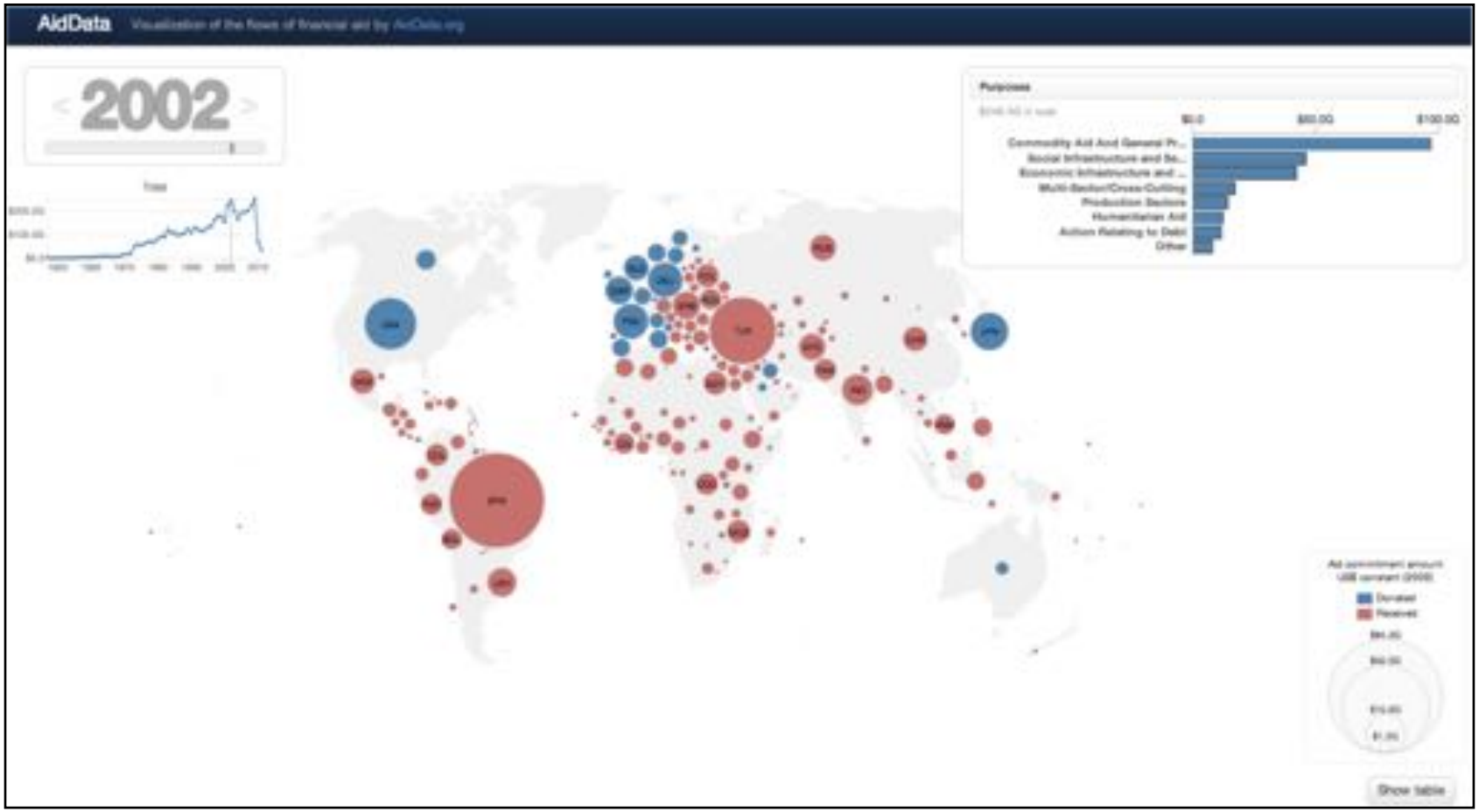

Geo scope / view
country(ies)
country - country
region - country
region(s)
region - region
global
Average % of Findings per Subject
30%
20%
10%
0%
28%
30%
31%
24%
6%
4%
6%
14%
11%
0%
17%
29%
anim
sm

FlowStrates



http://code.google.com/p/jflowmap/

Alddata



http://aiddata.herokuapp.com

What you should be able to answer

- What is InfoVis?
- Why is InfoVis useful?
- What are the main tasks that InfoVis tries to solve?
- What are the techniques for representing high dimensional data?
- What are the basic interaction techniques?