VISUALIZING DATA

Curse of dimensionality
Dynamic graphics
Subjectivity and interpretation
“Do not disturb my circles.” — Archimedes of Syracuse (287–212 BC), last words

PROBLEM 1: REPRESENTING HIGH-DIMENSIONAL DATA IN TWO DIMENSIONS. THE PROBLEM OF BIASED REPRESENTATIONS OF HIGH-DIMENSIONAL DATA
“Do not disturb my circles.”
There is no such thing as a “best projection”. Something will always get distorted.
The problem of biased representations

James Cheshire (www.spatialanalysis.co.uk). Appeared in Significance

“The earth is round; representing it on flat paper has to involve distortion. The only “true” map of the earth is therefore a globe.”

A conic projection places an imaginary cone as a hat over the earth and projects the earth’s features upon it. A polyconic projection can be thought of doing the same, but with a different cone for each line of latitude. Each cone touches its line of latitude as a tangent. Lines of latitude appear as arcs of circles on the projected map, but the circles are not concentric. The projection was invented in 1825 and was in common use, especially in the United States, until the middle of the twentieth century. Image: Stiebe
The problem of biased representations

Polyconic man
by L. LoCurto and B. Outcault.
Appeared in Significance

“This is a map of the human body, distorted as a map of the earth is distorted. Most of us would call the distortion extreme. We accept such distortion in maps, because we are used to it; we reject it in images of our bodies, because we are not.”
Representing Functions: The SteamGraph

The SteamGraph: nice representation of functional data

Songs listened by a particular last.fm user over an 18 month period of time
(source: http://infosthetics.com)
Twitter activity during the 2012 European Football Tournament
(http://philogb.github.com/blog/2012/07/02/euro2012-streamgraph-bsides)
2008 movies, by Graham Wills (2012)
Appeared in Significance

“I like this graphic because it looks different. On any page the eye would be drawn to it. That makes me want to look at it and see what it is saying”.

“The circularity does lead to distortion, though. Movies out on the edge look thinner, more smeared out and more anaemic. Is this in the data, or an artefact of their being out at the circumference so that a similar success has to stretch out over a longer distance?”
Italian 2012 box office revenues: a comparison among different approaches
**Representing Functions: The SteamGraph**

**SteamGraph:** Visualizing the individual proportions, and how distributions have changed over time

**Line plot:** Visualizing the trends for each group, losing the sense of a whole and distributions

“By revealing patterns in data, we create meaning, we welcome connections, we evoke the encoding channels of symmetry and color, and we tap the visceral mechanics of memory.”

PROBLEM 2: HOW TO EFFECTIVELY REPRESENT COMPLEX AND BIG DATA?
HOW TO EFFECTIVELY REPRESENT DATA?

Psycology: Draw the attention to the issue and stimulate curiosity
   - Make easy to remember graphs

Simplicity: Draw simple, “cognitive efficient” graphs
   - Make easy to read graphs

“Make ’em laugh, make ’em cry, and they might remember your message.”

Examples

Bateman et al. (2010)
HOW TO EFFECTIVELY REPRESENT DATA?

“Make ’em laugh, make ’em cry, and they might remember your message.”

Psychology: Draw the attention to the issue and stimulate curiosity
- Make easy to remember graphs

Simplicity: Draw simple, “cognitive efficient” graphs
- Make easy to read graphs

Examples

Bateman et al. (2010)
HOW TO EFFECTIVELY REPRESENT DATA?

Psycology: Draw the attention to the issue and stimulate curiosity
- Make easy to remember graphs

Simplicity: Draw simple, “cognitive efficient” graphs
- Make easy to read graphs

Examples

“Make ’em laugh, make ’em cry, and they might remember your message.”

Bateman et al. (2010); Gelman Unwin (2012)
"Thousands of years of human history oversimplified into 100-some words”

Martin Elmer
“Thousands of years of human history oversimplified into 100-some words”

Martin Elmer
HOW TO EFFECTIVELY REPRESENT DATA?

Interstellar memes (http://xkcd.com/1212/), appeared in Significance (June 2013)

Alessandra Menafoglio and Alessia Pini
Real-time data: data become available at the same time of their representation (Examples: internet data, social networks, etc.).

- Representation of Big data
- Synthesis of complex data in real time
HOW TO VISUALIZE REAL-TIME DATA?

**Immaterials**: light painting WiFi (by Timo Arnall, Jørn Knutsen and Einar Sneve Martinussen at Oslo University) [The taller the blue line, the stronger the signal]

“The city is filled with an invisible landscape of networks that is becoming an interwoven part of daily life.”

Appeared in Significance (August 2013)
**Immaterials**: light painting WiFi (by Timo Arnall, Jørn Knutsen and Einar Sneve Martinussen at Oslo University) [The taller the blue line, the stronger the signal]

“The city is filled with an invisible landscape of networks that is becoming an interwoven part of daily life.”

WiFi networks are both physically invisible and technically obscure, which makes them blackboxed on multiple levels. […] Through visualizations and the process of creating them we have unpacked some of the qualities of WiFi networks and made them understandable as phenomena in space and in the contexts that surround them.

Appeared in Significance (August 2013)
PROBLEM 3: HOW TO REPRESENT DYNAMICS THROUGH STATIC GRAPHICS? eARTICLE: A SOLUTION TO REPRESENT COMPLEX DATA?
DYNAMICS GRAPHICS

Courtesy of Paolo Zanini
DYNAMICS GRAPHICS

Courtesy of Paolo Zanini
“Sometimes I think that the only way objective data can be displayed effectively is with a subjective interpretation.”

“Visualization is merely another tool of communication, and it doesn’t mean it’s going to be the only pathway to truth. We like to think that our data is empirical, but much of it is open to interpretation.”

**PROBLEM 4: VISUALIZING DATA: AN OBJECTIVE OR SUBJECTIVE ISSUE? WHICH ETHIC RULES VISUALIZERS SHOULD FOLLOW?**
Which ethic rules visualizers should follow?

“Tolerance is rising – but it cannot get much higher than 100%”

Gallup (http://www.gallup.com/poll/149390/record-high-approveblack-white-marriages.aspx), appeared in Significance (April 2013)
Krugman’s graph plots proportion against time, and the fact that the proportion can’t go over 100% is relevant. There is no sign in his graph that the line which has already reached 86% is already almost as high as it can go.

Ratios suggest logs. A natural way to look at changes in proportions is odds ratios since it seems logical that the rate of change depends on (a) how much has been achieved and (b) how much remains to be done. Both of these are automatically incorporated in an odds ratio. The log odds are plotted in the graph below (Figure 4). It resembles Figure 3, but it does not have the positive–negative change of sign around 1990. Now the overall trend looks pretty straight, with obvious levelling-off around the year 2000 and steep acceleration in toleration immediately after that. Remember of course that there are sampling errors in these data.

“\textbf{The same data can be graphed in different ways. Looking at data the wrong way can suggest wrong inferences}”
Which ethical rules visualizers should follow?

2012 Presidential Run

GOP Candidates

- 70% Back Palin
- 63% Back Huckabee
- 60% Back Romney

Source: Opinions Dynamic
WHICH ETHIC RULES VISUALIZERS SHOULD FOLLOW?

**Pie chart**  
Central angle proportional to the quantity it represents

**Donut chart**  
No more central angle!!

WHAT DATA-RELATED AREA ARE YOU MOST INTERESTED IN?

Below are results of a poll on FlowingData in May 2009. Readers come from a variety of fields, but Statistics, Design, and Business led the way.

http://book.flowingdata.com/ch05/donut.html
Which ethic rules visualizers should follow?
WHICH ETHIC RULES VISUALIZERS SHOULD FOLLOW?

The government spends billions of dollars to support the energy industry, which allows it to make energy cheaper than it should cost on the open market. These subsidies—either in the form of tax breaks or direct funding—favor some types of energy over others, giving our country a skewed sense of what each gallon of gas or wind-powered electron costs. This is a look at where the government directed its subsidy dollars from 2002 to 2008.

**Carbon Capture and Storage**
- Tax Breaks: $3 billion
- Direct Spending: $2 billion

**Renewable Energy**
- Tax Breaks: $6.2 billion
- Direct Spending: $6 billion

**Corn Ethanol**
- Tax Breaks: $11 billion
- Direct Spending: $3 billion

**Fossil Fuels**
- Tax Breaks: $53.9 billion
- Direct Spending: $16.3 billion

Source: “Estimating U.S. Government Subsidies to Energy Sources” by the Environmental Law Institute

A collaboration between GOOD and DeepLocal
Which ethic rules visualizers should follow?

David McCandless (2009)
WHICH ETHIC RULES VISUALIZERS SHOULD FOLLOW?

The Middle Class Tax Target

The amount of total taxable income (left scale) for all filers by adjusted gross income level for 2008

$1.4 trillion

Source: IRS

http://online.wsj.com/article/SB10001424052748704621304576267113524583554.html


All the material will be available soon on the website

(Please, interact and send links or papers you’d like to add)

Contact us @ Polimi!

alessandra.menafoglio@polimi.it
alessia.pini@polimi.it