Interactive Data Visualization

Basic Introduction to Visualization and Tools

Benjamin WEYERS
University of Trier
Forthnightly Video Conference on Visualization

• Send your questions, topics, or requests to weyers@uni-trier.de

• They will move into the agenda and discussed or answered directly

• Search the Collaboratory for the „Visualization Video Conference“ collab

• Takes place each 1st and 3rd Thursday
  • 9-10am

• Is announced over Twitter!
Literature and References

• Scott Murray, (2013), *Interactive Data Visualization for the Web*, O’Reilly

All material not equipped with additional references (URL) on the slides is taken from the above books.
Graph drawing by Charles Joseph Minard showing the decreasing size of the Grande Armée as it marches to Moscow and back with the size of the army equal to the width of the line.
Harry Beck: „When you are underground it does not matter where your are“
Information Visualization

• “to visualize” [Oxford, 2010]
  – form a mental image of; imagine
  – make (something) visible to the eye

• not only produce pretty images, but aid the understanding of data

• visualization is interdisciplinary by definition

• the difference between InfoVis and SciVis is the type of data being visualized: Abstract Data vs. Measured Spatial Data
  – Nevertheless, the differentiation is not 100% clear…
  – Very different definitions out there

• in general: Information visualization‘s objective (as well as scientific visualization) is to represent data in a way that the human is able to gain insight and get enabled to understand its structure easily.
Information Visualization Guidelines

• A good visualization considers

1. The user (cognitive skills, pre-knowledge – expert vs. novice, disabilities, …)
2. The task (goals, operations, information, context, …)
3. The data

• The creation process of visualization applications and data analysis tools need to involve an incremental development approach:
  – identify the user group and investigate the/their pre...
  – identify the task and to investigate the task’s specific characteristics and derive the relevant analysis steps from it
  – create a prototype and to evaluate it in a user study
  – repeat
The Context

• What you need:
  – Understanding **what context is**
  – A description of context that is **understandable by the computer**
  – A system that brings the description and data together, which characterises the context

→ Makes context understandable by the computer-based system
The Data

• Before starting to discuss the visual representation of data and values, the different types of data should be classified first

• The goal of this classification is to describe concepts in InfoVis not like

„Color encoding is well suited for the representation of the development in the stock market“

but

„Color encoding is well suited for the representation of categories“
Visual Representation of Data and Values – Data Classification

- Stevens (1946) proposed a qualitative classification of data attributes along numeric scales:

1. **Nominal**: Elements that could not be logically ordered, e.g., apple, orange, grape, banana
2. **Ordinal**: Elements that could be logically ordered but where the ordering has no distance metric, e.g., weather situations along a scale of favor
3. **Interval**: Ordinal attributes that can be equipped with a (discrete) metric, e.g., arrival time of planes
4. **Ratio**: Extended Interval scale to the whole range of real numbers, e.g., the mass of solid physical objects
**Visual Representation of Data and Values – Data Classification**

1. **Nominal ↔ Category Data ↔ char**
   - Visual representation with colored dots and labels.

2. **Ordinal ↔ Discrete Data ↔ enum**
   - Example with numbers and colors.

3. **Ratio + Interval ↔ Continuous Data ↔ int, float**
   - Graph with values indicated on a line.

4. Similarly, another example illustrating the relationship between ratio, interval, continuous data, and types such as int, float.
Visual Representation of Data and Values – Data Classification

• The attribution of entities can be of higher dimension
• In general, an entity can be represented as field of attributes of different dimensions -> **Data Objects**

• **Operations** applied to data objects or more general than entities, attributes, or relations, which can not be defined as operations
  - Examples of operations are:
    - Mathematical operations
    - Merge of two lists
    - Invert values
    - Instantiation of entities or relations
    - …
Visual Representation of Data and Values

• Central **requirements of the visualization** of data and values are:
  – Present **more than one value** at a time
  – Relevant **dependencies and correlation** should be visible at one glance
  – Should be **intuitive and simple** to understand
  – Should **match the basic visual and perceptual characteristics** of the human visual cognition

• In the following, the visual representation of **single values** will be followed by the representation of **multiple values** up to **multi-dimensional data representations**
Single Value Representation
(0-dimensional)
Single Value Representation

- A simple example for the representation of a single value is the altimeter used in airplanes.
- This instrument is responsible for various accidents.
- By changes in attention and focusing to other contexts, changes in the altimeter can be overseen very easily.

http://www.m0a.com/altimeter/

Video
Change Blindness

- This effect is known as change blindness. Human perception is unaware of small changes in complex environments or in more or less complex representations.
Single Value Representation
1-Dimensional Value Representation
Representation of Value Sets – One Dimension

• **Example:** Visualize the prize of various cars, which are related to a linear scale

• **Question:** What is most effective? What should be identified in the data?

• **Focusing on**
  – Mean values
  – Distribution
  – Min and Max (price/s)
  – …

• Possible Visualization Methods: **Dot Plot, Box Plot, Histogram**
Representation of Value Sets – One Dimension

- Dot Plot
- Box Plot
- Histogram

Price in k€:

<table>
<thead>
<tr>
<th>00 - 20</th>
<th>20 - 30</th>
<th>30 - 40</th>
<th>40 - 50</th>
<th>50 - 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Jun.-Prof. Dr.-Ing. Benjamin Weyers | Trier University | Dept. IV – Computer Science
HPAC Training Workshop – Visualization Tools
Box Plot and Histogram

• Box Plot:
  – Central Line specifies the Median
  – End of Boxes specify 25 and 75 percentile
  – End of Lines specify 5 and 95 percentile
  – Dots show outliers

• Histogram:
  – Represent frequency/occurrence of values in a value set or of specific characteristics

[Diagram of Box Plot and Histogram]
2-Dimensional Value Representation
Presentation of Values Sets – Two Dimension

• Example: 2 dimensions \((\text{numberOfBedrooms}, \text{price [k€]})\)
  
  \((1, 108), (1, 115), (1, 135), (1, 150), \ldots, (4, 150), \ldots, (5, 195)\)
Scatter Plot

• Basic **Scatter Plot** can visualize two-dimensional data

• Possible interpretations are
  – Identification of trends
  – Local trade-offs
  – Outliers

• A specific area for scatter plots is in time-dependent data, such as **Spiking Plots** in Neuroscience

![Scatter Plot Example](https://capocaccia.ethz.ch/capo/wiki/2013/spinnaker13)
Spiking Plots

![Spiking Plots Image]

Jun.-Prof. Dr.-Ing. Benjamin Weyers | Trier University | Dept. IV – Computer Science
HPAC Training Workshop – Visualization Tools
Presentation of Value Sets – Two Dimension

Function Plot

Bar Chart

Scatter Plot

Heat Maps

https://capocaccia.ethz.ch/capo/wiki/2013/spinnaker13

http://www.infovis.info/

jcharts.sourceforge.net
In 1990, there were nearly 6 billion people in the world. Another 80 million are added each year. During the 1990s, population growth slowed down considerably to an annual rate of 1.4 percent.

World population doubled between the late-1950s and the late-1990s. It is projected to reach 9.4 billion by 2050.

The fastest rates of population growth are concentrated in Africa, together with some parts of the Middle East, and a few countries in Central and South America, and Southeast Asia.

Growth has slowed significantly in the two most heavily-populated countries: China and India. In the richest countries of the world, population growth has almost stopped. In a few countries, the number of people has begun to decline.
Presentation of Values Sets – Two Dimension
Presentation of Values Sets – Two Dimension

Australia

New Zealand
3-Dimensional Value Representation
3D Data

http://stats.stackexchange.com/questions/70569/interpreting-3d-scatter-plot
N-Dimensional Value Representation & Interaction Concepts for Visualization
Presentation of Value Sets – N Dimensions

• A well known and nice concept is the so called „Small Multiples“ described by Tufte, 2013

• In general: A data set is presented in many small drawings, where one dimension is altered between all drawings, such as time, position, etc.

• Small Multiples are very good in making changes visible along the altered dimension
  – It enables the user to compare different views to the data with each other in a very simple and convenient way.
Small Multiples – Historic Example

Small Multiples – Examples

- Double Bar Charts
Small Multiples – Examples

2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support

Income under $20,000 | $20-40,000 | $40-75,000 | $75-150,000 | Over $150,000

All voters

White Catholics

White evangelicals

White non-evang. Protestants

White other/no religion

Blacks

Hispanics

Other races

Orange and green colors correspond to states where support for vouchers was greater or less than the national average. Where a category represents less than 1% of the voters of a state, the state is left blank.

http://media.juiceanalytics.com/images/smallmultiples1.png
Small Multiples – Examples

- **Trouble Spots**
  - Air-conditioning
  - Body exterior (paint)
  - Body exterior (rust)
  - Body hardware
  - Body integrity
  - Brakes
  - Clutch
  - Driveline
  - Electrical system (d泅ionic)
  - Engine cooling
  - Engine mechanical
  - Exhaust system
  - Fuel system
  - Ignition system
  - Suspension
  - Transmission (manual)
  - Transmission (automated)
  - Trouble Index
  - Gear Index

- **Chevrolet Malibu** (Chevelle, 6, V8)
- **Chevrolet Monte Carlo**
- **Datsun 210, B210**
- **Ford Granada 6**
- **Ford pickup truck (EWD)**
- **Honda Accord**
- **Mercedes-Benz 300D (Diesel)**
- **Plymouth Valiant**
- **Saharu (except Toyota)**
- **Toyota Corolla (except Toyota)**
- **Volkswagen Rabbit (diesel)**
- **Volvo 240 series**

- ![Graph showing small multiples examples](http://www.infovis-wiki.net/index.php?title=Teaching:TUW_-_UE_InfoVis_WS_2008/09_-_Gruppe_06_-_Aufgabe_1_-_Small_Multiples)
Visual Information-Seeking Mantra (Shneiderman Mantra)

Overview first, zoom and filter, then details-on-demand

- **Overview**: Gain an overview of the entire collection.
- **Zoom**: Zoom in on items of interest
- **Filter**: Filter out uninteresting items.
- **Details-on-demand**: Select an item or group and get details when needed
- **Relate**: View relationships among items
- **History**: Keep a history of actions to support undo, replay, and progressive refinement
- **Extract**: Allow extraction of sub-collections and of the details when needed.

**Visual Information-Seeking Mantra (Shneiderman Mantra)**

- **Zoom**
- **Details**
- **Filter**
Interactive Data Analysis – Selection

- Adding **interaction capabilities** to information visualizations extends the possible **interpretation and working scope** of these kinds of visualization massively.

- One well known interaction paradigm in information visualization is the **multi-view paradigm** or **multiple coordinated views**.

- The multi-view paradigm is (mainly comprised) of two interaction metaphors:
  1. **Linking** – Selection in one view gets propagated to other connected representations
  2. **Brushing**
Multiple View – Brushing


Edgar Anderson’s Iris data set scatterplot matrix
Interactive Data Analysis – Filtering
Interactive Data Analysis – Filtering

(a) Interactive Data Analysis – Filtering

(b) Interactive Data Analysis – Filtering

(c) Interactive Data Analysis – Filtering

(d) Interactive Data Analysis – Filtering
Presentation of Values Sets – N Dimension

Scatter Plot Matrix

Extended Scatter Plots

Geographical Hexbins

Trees & Graphs
Parallel Coordinates

• Is a simple solution to present multi-dimensional data
• There are some problems with this type of representation but is still a way to present multi dimensions in one chart
Coordinate Plots

• It does not become clear that there is a trade-off between B and E, where C and G show a strong correlation
• Interaction can be solution to this draw back
Coordinate Plots – Parallel Coordinates

- Brushing, Filtering, Sorting…
Star Plots

- Close related to parallel coordinates
- In case of star plots, parameter axes are arranged in radial fashion

- They can be used for comparing different data sets, e.g., the performance of students
- They support the visibility of object and structures in the data
Dust & Magnet: Interactive Visualization for Everyday Data

Ji Soo Yi¹, Rachel Melton Ponder², John Stasko², & Julie A. Jacko¹

¹Industrial and Systems Engineering, Georgia Tech
²GVU Center & College of Computing, Georgia Tech
Relational Data
Literature


Representation of Relations

- A relation can be defined as logical or natural association between two or more things; as relevance of one to something else; as connection.
<table>
<thead>
<tr>
<th>Originator</th>
<th>Receiver</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>H</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>K</td>
</tr>
<tr>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td>G</td>
<td>K</td>
</tr>
<tr>
<td>K</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>H</td>
</tr>
</tbody>
</table>

(a) (b) (c)
• The resulting representation is less meaningful
Friday, 17 August, 2001 16:17:50

pierlauder

bverseandlaine

jeffers

loves

twouters
Node-Link Diagrams

- **Node-Link Diagrams** are visual presentations of relational data structures.

- Relational data structures refer to graphs.

- Special types of graphs are **trees**.
Trees

- Trees are a sub-group of graphs with a specific structure.

[Diagram of a tree structure with labels for designated root node, parent of A, sibling of A, child of A, and leaf nodes.]
3D Cone Trees
3D Cone Trees

- Cockburn et al. 2000 evaluated that Cone Trees are really attractive to work with, but are less efficient than standard 2D tree (file browsers)

http://www.java2s.com/Code/JavaImages/FileTreeDragSource.PNG
Tree Maps

- Depth of encapsulation directly relates to depth in the tree
Tree Maps
Force-Directed Approach

http://bl.ocks.org/couchand/raw/6420534/

https://vida.io/documents/N4jSip7n68yQ48DXp
Edge Bundling

Ersoy et al.
Skeleton-based edge bundling for graph visualization

Selassie et al.
Divided Edge Bundling for Directional Network Data
Edge Bundling

Holten & van Wijk  
*Force-Directed Edge Bundling for Graph Visualization*

Selassie et al.  
*Divided Edge Bundling for Directional Network Data*

Force Directed Edge Bundling

Divided Edge Bundling
Holten
Hierarchical Edge Bundles:
Visualization of Adjacency Relations in Hierarchical Data
Edge Bundling – short view to 3D graph drawing
Spatial Data
Scientific Visualization
Data Types

• Scalar Fields (structured/unstructured grid)
  – Explicit Volume Rendering → Ray Casting
  – Implicit Volume Rendering → Extraction of Iso Surface/Marging Cubes

• Vector Fields (structured/unstructured grid)
  – Glyphs
  – Streamlines
  – Particle Tracing
  – Line Integral Convolution
  – Topology Methods

• Tensor Fields
Iso Surface – Implicit Volume Rendering

- Use of ray tracing or geometry rendering

- Extraction of Iso Surfaces: Marching Cubes

https://upload.wikimedia.org/wikipedia/commons/thumb/a/a7/MarchingCubes.svg/350px-MarchingCubes.svg.png

http://www.cs.utah.edu/~shirley/papers/iso/both.jpg
Volume Rendering – Ray Casting

Month 26
Threshold: 0.1
Ray Casting

1. Ray Casting
2. Sampling
3. Shading
4. Compositing
Vector Field - Glyphs
Vector Field – Stream Lines

Examples from Neuroscience
NeuroScheme

(this demo uses cortex data)

http://gmrv.es

Universidad Rey Juan Carlos

POLITÉCNICA
VIOLA – Life Demo

https://github.com/HBPVIS/VIOLA

http://hbpvis.github.io/VIOLA/
Tools and Libraries
• The following content is based on the Book „Interactive Data Visualization for the Web“ written by Scott Murray

and

• his tutorial to find on http://alignedleft.com/tutorials/d3

• The home of D3 is http://d3js.org/

and

https://github.com/mbostock/d3
Add-On Libs

Circle Packing

Hierarchical Edge Bundling

Chord Diagram

Interactive Bible Contradictions & Bar Charts

Interactive Function Plots
ParaView
Geppetto
Visual Integrity
Graphical Integrity

• Graphical integrity specifies the aspect in how far graphical representations lie about the underlying data

• Tufte specifies a „lie factor“ as given below

\[
\text{lie factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}
\]

• If the Lie Factor is 1, the graphic is valid
Lie Factor – Example

• The graphic was intended to present a regulation by the US Dept. of Transportation defining an increase of miles per gallon a car has to meet until 1985 from 18.0 miles per gallon in 1978 to 27.5 in 1985

• Real increase is:

\[
\frac{27.5 - 18.0}{18.0} \times 100 = 53\%
\]

Shown is (relative lengths of lines):

\[
\frac{5.3 - 0.6}{0.6} \times 100 = 783\%
\]

\[\text{Lie Factor} = \frac{783}{53} = 14.8\]
Lie Factor – Examples

• Lie Factor of 9.4:

  ![Graph showing the price per barrel of light crude oil from 1973 to 1979, with April 1, 1979, priced at $14.55.](image1)

• Lie Factor of 2.8:

  ![Graph showing the percentage of doctors devoted solely to family practice from 1964 to 1990.](image2)
Forthnightly Video Conference on Visualization

• Send your questions, topics, or requests to weyers@uni-trier.de

• They will move into the agenda and discussed or answered directly

• Search the Collaboratory for the „Visualization Video Conference“ collab

• Takes place each 1st and 3rd Thursday
• 9-10am

• Is announced over Twitter!
1 minute paper
Contactinformation

The video conference will be held online. Connection details can be found below.
Telephone Number: 004910097956411 Conference Number: 97956411
In order to enter the conference use one of the following numbers/addresses:
1) H.323 dial-in using a local gatekeeper (Global Dialling Scheme, GDS) - dial 004910097956411
2) H.323 dial-in using DFN gatekeeper - configure your email address as H.323 alias and your phone number as E.164 number - configure gk.vc.dfn.de as the gatekeeper - dial 004910097956411
3) H.323 dial-in without any gatekeeper (URI dialling) - for Tandberg, Mirial dial 97956411@vc.dfn.de - for Polycom, Lifesize dial vc.dfn.de##97956411 Additional information about URI dialling and other end systems can be found at: http://vcc.zih.tu-dresden.de/index.php?linkid=11100&lang=en
4) SIP dial-in (For clients such as Cisco SX10 or Polycom Real Presence Mobile Client at Android or iOS) - dial 97956411@vc.dfn.de Registration to a SIP proxy is not required. If your endpoint supports BFCP protocol, you can use the second video stream for presentations.
5) Telephone/Mobile (POTS/ISDN) dial-in: - dial +49-30-20097956411 You will be directed to the conference.
You can test all of this anytime as the only participant to the conference to see if your VC system is working well with the MCUs. If you need a peer for testing two-way video and audio with more than your own signals give us a shout and we can schedule a common test slot some days before the meeting for anybody who wants to test that thoroughly in advance. For troubleshooting please contact DFNVC Hotline at +49-711-63314-214.