Information Visualization

Introduction

Inspired from Petra Isenberg
petra.isenberg@inria.fr
Why

INFORMATION VISUALIZATION
It is estimated that 800 exabyte ($800 \times 10^{19}$) of digital information will be generated this year.

[source: The Diverse and Exploding Digital Universe, IDC, 2008]
[credit: Did You Know; Fisch, McLeod, Brenman]
Question

how can we effectively access data?
- understand its structure?
- make comparisons?
- make decisions?
- gain new knowledge?
- convince others?
-...
Many possible ways to address…

Information Visualization
## Example

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td></td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td></td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td></td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td></td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td></td>
</tr>
</tbody>
</table>

Raw Data from Anscombe’s Quartet

[Source: Anscombe's quartet, Wikipedia]
**Statistical Analysis**

For all four columns, the statistics are identical

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td></td>
<td>y</td>
<td></td>
<td>x</td>
<td></td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td></td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
<td>6.58</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td></td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>5.76</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td></td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
<td>7.71</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td></td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
<td>8.84</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td></td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
<td>8.47</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td></td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
<td>7.04</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td></td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
<td>5.25</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td></td>
<td>4.0</td>
<td>5.39</td>
<td>19.0</td>
<td>12.50</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td></td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
<td>5.56</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td></td>
<td>7.0</td>
<td>6.42</td>
<td>8.0</td>
<td>7.91</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td></td>
<td>5.0</td>
<td>5.73</td>
<td>8.0</td>
<td>6.89</td>
</tr>
</tbody>
</table>

Mean of $x$ 9.0  
Variance of $x$ 11.0  
Mean of $y$ 7.5  
Variance of $y$ 4.12  
Correlation between $x$ and $y$ 0.816  
Linear regression line $y = 3 + 0.5x$

[Source: Anscombe's quartet, Wikipedia]
Visual Representation of the Data

Visual representation reveals 4 different stories

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>5.39</td>
<td>19.0</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>5.39</td>
<td>7.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>5.73</td>
<td>8.0</td>
</tr>
</tbody>
</table>

[Source: Anscombe's quartet, Wikipedia]
Why visual data representations?

- Vision is our most dominant sense
- We are very good at recognizing visual patterns
- We need to see and understand in order to explain, reason, and make decisions

common examples:

graphs / hierarchies  charts  maps

all examples from: http://vis.stanford.edu/protovis/
Other benefits of visualization

• expand human working memory
  – offload cognitive resources to the visual system,
• reduce search
  – by representing a large amount of data in a small space,
• enhance the recognition of patterns
  – by making them visually explicit
• aid monitoring of a large number of potential events
• provides a manipulable medium & allows exploration of a space of parameter values.
百聞不如一見

"One hundred rumors are not comparable to one look."

An Old Chinese Inscription
Information visualization

- Create visual representation
- Concentrates on abstract data
- Includes interaction

Official Definition:

*The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.*

[Card et al., 1999]
Functions of Visualizations

• Recording information
  – Tables, blueprints, satellite images

• Processing information
  – needs feedback and interaction

• Presenting information
  – share, collaborate, revise
  – for oneself, for one’s peers and to teach

• Seeing the unseen
Visualization of abstract data has been practiced for hundreds of years...

**HISTORICAL EXAMPLES**
The Broadway Street Pump

• In 1854 cholera broke out in London
  – 127 people near Broad Street died within 3 days
  – 616 people died within 30 days
• “Miasma in the atmosphere”
• Dr. John Snow was the first to link contaminated water to the outbreak of cholera
• How did he do it?
  – he talked to local residents
  – identified a water pump as a likely source
  – used maps to illustrate his theory
  – convinced authorities to disable the pump

More info here: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak
Napoleon’s March on Moscow  Charles Minard, 1869

Named the best statistical graphic ever drawn (by Edward Tufte)

- Includes: spatial layout linked with stats on: army size, temperature, time
- Tells a story in one overview

More info: The Visual Display of Quantitative Information (Tufte)
CARTE FIGURATIVE des pertes successives en hommes de l’Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

-26° le 7 X°
-30° le 5 X°
11°
-21° le 14 X°
-24° le 10 X°
-27° le 28 X° Pluie 25 8°
... AND VERY RECENTLY
TrashTrack

Winner of the NSF International Science & Engineering Visualization Challenge!
http://senseable.mit.edu/trashtrack/
Artificial Intelligence

http://www.turbulence.org/spotlight/thinking/chess.html
Open Data

- Movement making government data freely available
- Encourage participation by everyone

OECD Better Life Index: http://www.oecdbetterlifeindex.org/
Many Eyes

- Upload data, create visualizations, discuss
- Distributed asynchronous collaboration

http://www-958.ibm.com/software/data/cognos/manyeyes/
Software Visualization

EZEL: a Visual Tool for Performance Assessment of Peer-to-Peer File-Sharing Networks (Voinea et al., InfoVis, 2004)
Text Visualization

Parallel Tag Clouds to Explore Faceted Text Corpora (Collins et al., VAST 2009)
Graphs

Here Wikipedia

http://sepans.com/sp/psots/wiki_category/
Family Trees

http://www.aviz.fr/geneaquilts/
Geographic Visualization

http://data-arts.appspot.com/globe
Weather
Data Dashboards

[Image of a data dashboard showing various sectors, markets, and stock data.]

**NYT.COM**
The rights hedgehog provocateur
Mon, 07 Jun 2011 14:30:51 GMT
Andrew Breitbart uses his network of Web sites and their legions of followers to bring conservative media red meat.

**TWITTER**
pennystockchat
Mon, 07 Jun 2011 14:30:52 +0000

[Image of a Twitter feed with related stock market data.]
Resources for more examples

• Visualization conferences
• Blogs
• Books
  – Textbooks
    • Readings in Information Visualization: Using Vision to Think (a bit old now but good intro)
    • Information Visualization (Robert Spence – a light intro, I recommend as a start)
    • Information Visualization Perception for Design (Colin Ware, focused on perception and cognition)
    • Interactive Data Visualization: Foundations, Techniques, and Applications (Ward et al. – most recent)
  – Examples
    • Beautiful Data (McCandless)
    • Now You See it (Few)
    • Tufte Books: Visual Display of Quantitative Information (and others)
    • ... (many more, ask me for details)
It is difficult to create

CREATE VISUALIZATIONS

GOOD
What is a representation?

- A representation is
  - a formal system or mapping by which the information can be specified (D. Marr)
  - a sign system in that it stands for something other than its self.

- for example: the number thirty-four

34 100010 XXXIV

decimal binary roman
Presentation

• different representations reveal different aspects of the information
  - decimal: counting & information about powers of 10,
  - binary: counting & information about powers of 2,
  - roman: counting & adding and subtracting

• presentation
  how the representation is placed or organized on the screen

34, 34, 34
Principles of Graphical Excellence

• Well-designed presentation of interesting data – a matter of *substance, statistics, design*

• Complex ideas communicated with clarity, precision, efficiency

• Gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space

• Involves almost always multiple variables

• Tell the truth about the data

---

The Visual Display of Quantitative Information, Tufte
Or a bit more simply...

- Solving a problem simply means representing it so as to make the solution transparent ... (Simon, 1981)
- Good representations:
  - allow people to find relevant information
    - information may be present but hard to find
  - allow people to compute desired conclusions
    - computations may be difficult or “for free” depending on representations
How do we arrive at a visualization?

The Visualization Pipeline

From [Spence, 2000]
Visualization Reference Model

Also a visualization pipeline a bit expanded

---

Data Abstraction → Spatial Layout → Presentation → View

Data Transformation → Spatial Mapping Transformation → Presentation Transformation → View Transformation

From [Card et al., Readings in Information Visualization]
Visualization pipeline in an image

[Tobiasz et al., 2009]
Knowledge Crystallization Cycle

[Card et al., 1999]

Working with visualizations in NOT a linear process
Pitfalls

• Selecting the wrong data
• Selecting the wrong data structure
• Filtering out important data
• Failed understanding of the types of things that need to be shown
• Choosing the wrong representation
• Choosing the wrong presentation format
• Inappropriate interactions provided to explore the data
Data

• Data is the foundation of any visualization
• The visualization designer needs to understand
  – the data properties
  – know what meta-data is available
  – know what people want from the data
Nominal, Ordinal and Quantitative

• Nominal (labels)
  – Fruits: apples, oranges

• Ordered
  – Quality of meat: grade A, AA, AAA
  – Can be counted and ordered, but not measured

• Quantitative: Interval
  – no clear zero (or arbitrary)
  – e.g. dates, longitude, latitude
  – usually compare differences (intervals)

• Quantitative: Ratio
  – meaningful origin (zero)
  – physical measurements (temperature, mass, length)
  – counts and amounts

S.S. Stevens, On the theory of scales of measurements, 1946
Nominal, Ordinal and Quantitative

- Nominal (labels)
  - Operations: $=, \neq$

- Ordered
  - Operations: $=, \neq, <, >$

- Quantitative: Interval
  - Operations: $=, \neq, <, >, -, +$
  - Can measure distances or spans

- Quantitative: Ratio
  - Operations: $=, \neq, <, >, -, +, \cdot, \div$
  - Can measure ratios or proportions

S.S. Stevens, On the theory of scales of measurements, 1946
Data-Type Taxonomy

- 1D (linear)
- Temporal
- 2D (maps)
- 3D
- nD (relational)  
  vis examples later
- Trees (hierarchies)
- Networks (graphs)

Shneiderman: The Eyes Have It
Why is this important?

• Nominal, ordinal, and quantitative data are best expressed in different ways visually

• Data types often have inherent tasks
  – temporal data (comparison of events)
  – trees (understand parent-child relationships)
  – ...

• But:
  – any data type (1D, 2D,...) can be expressed in a multitude of ways!
Visualization’s Main Building Blocks

Marks which represent:

- Points
- Lines
- Areas

From Semiology of Graphics (Bertin)
Points

• “A point represents a location on the plane that has no theoretical length or area. This signification is independent of the size and character of the mark which renders it visible.”
• a location
• marks that indicate points can vary in all visual variables

From Semiology of Graphics (Bertin)
“A line signifies a phenomenon on the plane which has **measurable length but no area**. This signification is independent of the width and characteristics of the mark which renders it visible.”

• a boundary, a route, a connection

From Semiology of Graphics (Bertin)
Areas

- “An area signifies something on the plane that **has measurable size**. This signification applies to the entire area covered by the visible mark.”
- an area can change in position but not in size, shape or orientation without making the area itself have a different meaning

From Semiology of Graphics (Bertin)
Visual Variables Applicable to Marks

From Semiology of Graphics (Bertin)
Additional Variables for Computers

• **motion**
  - direction, acceleration, speed, frequency, onset, ‘personality’

• **saturation**
  - colour as Bertin uses largely refers to hue, saturation != value

Extending those from Semiology of Graphics (Bertin)
Additional Variables for Computers

- **flicker**
  - frequency, rhythm, appearance

- **depth? ‘quasi’ 3D**
  - depth, occlusion, aerial perspective, binocular disparity

- **Illumination**

- **transparency**

From Semiology of Graphics (Bertin)
**Characteristics of Visual Variables**

- **Selective:**
  Is a change in this variable enough to allow us to select it from a group?

- **Associative:**
  Is a change in this variable enough to allow us to perceive them as a group?

- **Quantitative:**
  Is there a numerical reading obtainable from changes in this variable?

- **Order:**
  Are changes in this variable perceived as ordered?

- **Length (resolution):**
  Across how many changes in this variable are distinctions possible?

From *Semiology of Graphics* (Bertin)
Visual Variable: Position

- Selective
- Associative
- Quantitative
- Order
- Length

From Semiology of Graphics (Bertin)
Visual Variable: Size

- selective
- associative
- quantitative
- order
- Length

- theoretically infinite but practically limited
- association and selection ~ 5 and distinction ~ 20
Size

points  lines  areas
Visual Variable: Shape

- • selective
  ![Selective Shapes]

- • associative
  ![Associative Shapes]

- • quantitative
  ![Quantitative Shapes]

- • order
  ![Order Shapes]

- • length
  - infinite
  ![Length Shapes]
Shape

points  lines  areas
**Visual Variable: Value**

- **selective**
- **associative**
- **quantitative** (not selected)
- **order**
- **length**
  - theoretically infinite but practically limited
  - association and selection ~ < 7 and distinction ~ 10
Value

points

lines

areas
Value

• Ordered, cannot be reordered

Values not ordered correctly according to scale
Information has to be read point by point

Values ordered correctly
Image much more useful

annual deaths per 1000 inhabitants, Paris
Visual Variable: Colour

- **selective**
- **associative**
- **quantitative**
- **order**
- **length**
  - theoretically infinite but practically limited
  - association and selection ~ < 7 and distinction ~ 10
Visual Variable: Orientation

- **selective**
- **associative**
- **quantitative**
- **order**
- **length**
  - ~5 in 2D; ? in 3D
Orientation

points  lines  areas
Visual Variable: Texture

- selective
- associative
- theoretically infinite

≠ quantitative
≠ order
≠ length
Texture

points

lines

areas
Visual Variable: Motion

- **selective**
  - motion is one of our most powerful attention grabbers

- **associative**
  - moving in unison groups objects effectively

- **quantitative**
  - subjective perception

- **order**

- **length**
  - distinguishable types of motion?
Motion
# Visual Variables

<table>
<thead>
<tr>
<th>Visual Variable</th>
<th>Selective</th>
<th>Associative</th>
<th>Quantitative</th>
<th>Order</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Dependant on resolution</td>
</tr>
<tr>
<td>Size</td>
<td>Yes</td>
<td>Yes</td>
<td>Approximate</td>
<td>Yes</td>
<td>Association: 5; Distinction: 20</td>
</tr>
<tr>
<td>Shape</td>
<td>With Effort</td>
<td>With Effort</td>
<td>No</td>
<td>No</td>
<td>Infinite</td>
</tr>
<tr>
<td>Value</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Association: 7; Distinction: 10</td>
</tr>
<tr>
<td>Hue</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Association: 7; Distinction: 10</td>
</tr>
<tr>
<td>Orientation</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Grain</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Texture</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Infinite</td>
</tr>
<tr>
<td>Motion</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Carpendale, 2003
Summary

– Now you know the main building blocks are **marks**
– Marks are modified by **visual variables**
– Visual variables have **specific characteristics**
– These characteristics influence how the data will be perceived