DATA VISUALIZATION AND VISUAL ANALYTICS

S. Rinzivillo – rinzivillo@isti.cnr.it
TAXONOMY OF VISUAL VARIABLES
Cleveland / McGill, 1984

Figure 3. Graphs from position–angle experiment.

Figure 4. Graphs from position–length experiment.
CLEVELAND & MCGILL: GRAPHICAL ENCODINGS

- Angle
- Area
- Color Hue
- Color Saturation
- Density
- Length
- Position on a common scale
- Position on non aligned scale
- Slope
- Volume
• It is difficult to compare angles
  • Underestimation of acute angles
  • Overestimation of obtuse angles
  • Easier if bisectors are aligned

• Area estimation helps
ANGLE DECODING

- It is difficult to compare angles
  - Underestimation of acute angles
  - Overestimation of obtuse angles
  - Easier if bisectors are aligned
• Same difficulties as angles
• Easier task since one branch is aligned with x-axis
**AREA DECODING**

- Area is not well decoded
  - Different regular shapes
  - Irregular shapes
  - Context influences (thin area within compact thick area)
LENGTH DECODING

• Straight forward to encode numerical values

• Difficulties with relative lengths
POSITION ON A COMMON SCALE

- Widely used in statistical charts
POSITION ON NON-ALIGNED SCALE

• Not as bas as common scale
• Still acceptable
DESIGNING EFFECTIVE VISUALIZATIONS

• If possible, use graphical encoding that are easily decoded

• Graphical Attributes ordered (Cleveland & McGill):
  • Position along a common scale
  • Position on non aligned scales
  • Length
  • Angle and Slope
  • Area
  • Volume, density, color saturation
  • Color Hue
Most Efficient

Least Efficient

Position

Length

Slope

Angle

Area

Intensity

Color

Shape

Quantitative

Ordinal

Nominal
TAKEAWAY MESSAGES

• Data type for entities and relationships
• Visual variables for representation
• Mapping of types to VVs
• Some VVs are more appropriate for specific data types
VISUAL ANALYTICS

DOS AND DON’TS FOR VISUAL CHARTS
CRASH COURSE ON EFFECTIVE CHARTING

Dona M. Wong
Guide to Information Graphics
The Dos and Don’ts of Presenting Data, Facts, and Figures
W. W. Norton & Company
CHARTING PIPELINE

**Research**
- Found pertinent and authoritative data
- Integrate disputable sources

**Edit**
- Select your key message
- Filter, transform, and simplify data to deliver your message

**Plot**
- Choose the right chart type
- Choose the right chart properties
- Use opportune labelling
- Add colors (if needed)

**Review**
- Look at the chart from reader perspective
- Compare with independent sources
CHARTING EXAMPLES

- May these charts be improved? Why? How?
May these charts be improved? Why? How?
FONT S
**Fonts**

Typographic parts of a glyph:
1) x-height; 2) **ascender line**; 3) apex; 4) **baseline**; 5) ascender; 6) crossbar; 7) stem; 8) **serif**; 9) leg; 10) bowl; 11) counter; 12) collar; 13) loop; 14) ear; 15) tie; 16) horizontal bar; 17) arm; 18) vertical bar; 19) cap height; 20) **descender line**.

Font size: 
\[ \text{Font size} = (1) + (2) + (20) = (19) + (20) \]
**Fonts: General Rules**

- Leading should be 2 points larger than type size
- Avoid too small or condensed typefaces
- Keep style simple: use **bold** or *italic* to emphasize a word (better not both)
- Avoid **ALL CAPS**
- Avoid *styled fonts*
- Avoid C***C Sans Serif
- Reduce type at an angle
- Avoid tracking

Fonts are meant to describe, not to adorn
**TYPOGRAPHY IN CHARTS**

**Don’t**

- **HEADLINE OF THE CHART**
  - Don’t use all caps or high contrast white type out of black
  - Don’t use tilted text
  - Don’t use bold for axis

- A brief description that outlines what the data shows
  - Don’t use bold and italic

**Do**

- Headline of the chart

- A brief description that outlines what the data shows

---

**Graph:**

- **Y-axis:** Numbers 0 to 10
- **X-axis:** Town A, Town B, Town C, Town D
- **Bars:** Heights representing different values for each town.
TYPOGRAPHY IN CHARTS

Don’t

Headline of the chart

Title of y-axis

Title of x-axis

Do

Headline of the chart

Title of y-axis

Title of x-axis
## Typography in Charts

Many elements in **bold**. Which part is highlighted?

<table>
<thead>
<tr>
<th>Name</th>
<th>Data</th>
<th>Data</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company A</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Company B</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Company C</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Company D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Give emphasis to relevant results.
DATA-INK RATIO

Visual Display of Quantitative Data
Edward Tufte, 1983
DATA-INK RATIO

Data-Ink Ratio = \frac{\text{Data ink}}{\text{Total ink used in graphic}}

Categoria 1 | Categoria 2
---|---
Serie 1 | Serie 2
Serie 3 | Serie 4

Legend: Serie 1 | Serie 2 | Serie 3 | Serie 4
DATA-INK RATIO

Data-Ink Ratio = \frac{\text{Data ink}}{\text{Total ink used in graphic}}

Categoria 1
- Serie 1
- Serie 2
- Serie 3
- Serie 4

Categoria 2
- Serie 1
- Serie 2
- Serie 3
- Serie 4
**BAR CHARTS**

Represent discrete quantities

Town A: 8
Town B: 6
Town C: 4
Town D: 2
Avoid non-functional adornation
BAR CHARTS: BASELINE

Chart Title

Chart Title
BAR CHARTS: BASELINE

Chart Title

A  B  C  D
BAR CHARTS: ORDERING

France 1,1
Germany 4,1
Italy 6,1
China 9,1
USA 7,1

France 1,1
Germany 4,1
Italy 6,1
US 7,1
China 9,1
Visual Analytics

Series 1

Year
- 2006
- 2007
- 2008
- 2009
- 2010

Values
- 0
- 10
- 20

Series 1

Categories
- Category 1: -9.1
- Category 2: -7.1
- Category 3: -6.1
- Category 4: -4.1
- Category 5: -2.1

Visual Analytics
va602aa
Pie Charts compares relative sizes and contributions
PIE CHARTS: ORDERING SLICES
May these charts be improved? Why? How?
TAKEAWAY MESSAGES

• Charts exploit position on scale VV
• Best practice to reduce biases and misinterpretation of charts
Information Graphics
A Comprehensive Illustrated Reference

Visual Tools for Analyzing, Managing, and Communicating

Robert L. Harris
Chart Suggestions—A Thought-Starter

**Comparison**

What would you like to show?

**Relationship**

Two Variables

Three Variables

**Distribution**

Over Time

Changing Over Time

Static

**Composition**

Few Periods

Many Periods

Three Variables

One Variable per Item

Many Categories

Few Categories

Many Items

Few Items

Cyclical Data

Non-Cyclical Data

Single or Few Categories

Many Categories

Few Periods

Over Time

Few Data Points

Many Data Points

Two Variables

Simple Share of Total

Accumulation or Subtraction to Total

Components of Components

**Visual Analytics**

va602aa
Bars vs. Lines

Line implies trends. Do not use for categorical data.
TREND OVER TIME

WILLIAM PLAYFAIR
1759-1823
TREND OVER TIME

Visual Analytics
va602aa
TREND OVER TIME

Budget Forecasts, Compared With Reality
Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

Latest forecast
Today, with a better understanding of the severity of the economic downturn, the deficit situation is much more dire.

The 1995 forecast for 1999 did not predict a surplus...

...but the 2008 forecast for 2012 did.

Make clear distinction between data and prediction
STREAMGRAPHS
Six Ways to Find Value in Twitter’s Noise

1. Look for unexpected loves. A prototype created by the design firm frog Design in Germany shows how Twitter has the potential to be a valuable tool for fostering innovative ideas. When the company’s CEO, Brad Ho, asked the company’s employees to tweet about what they liked on the Web, it quickly became apparent that the most popular topics were those that were not expected. For example, the company’s employees were more interested in the latest news about the stock market than in new technology products. This unexpected result led to a series of brainstorming sessions that resulted in new product ideas.

2. Look for unexpected passions. A group of researchers at the University of California, San Diego, created a tool that allows users to search for specific words on Twitter. When the researchers searched for the word “love,” they found that it was used more than any other word on the site. This unexpected result led to a series of studies that explored the role of love in social media.

3. Look for unexpected connections. A team of researchers at the University of California, Los Angeles, created a tool that allows users to search for specific words on Twitter. When the researchers searched for the word “friend,” they found that it was used more than any other word on the site. This unexpected result led to a series of studies that explored the role of friendship in social media.

4. Look for unexpected emotions. A team of researchers at the University of California, Berkeley, created a tool that allows users to search for specific words on Twitter. When the researchers searched for the word “happy,” they found that it was used more than any other word on the site. This unexpected result led to a series of studies that explored the role of happiness in social media.

5. Look for unexpected trends. A group of researchers at the University of California, Los Angeles, created a tool that allows users to search for specific words on Twitter. When the researchers searched for the word “trend,” they found that it was used more than any other word on the site. This unexpected result led to a series of studies that explored the role of trends in social media.

6. Look for unexpected insights. A team of researchers at the University of California, Berkeley, created a tool that allows users to search for specific words on Twitter. When the researchers searched for the word “insight,” they found that it was used more than any other word on the site. This unexpected result led to a series of studies that explored the role of insights in social media.

The Pad Launch by the Numbers

547,898

The total number of tweets that were sent during the launch weekend of the Pad.
65% of the market is controlled by companies B and C.
Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.
Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.
Furthermore, we present the distribution of attacks towards employees in detail in Fig. 10 right. The blue employees are secretaries, the green ones are administrators and the red ones are scientific employees. The number following the name is the number of times that person was attacked. All of the names are pseudonyms for real people. The person that suffered the most attacks is Monja a secretary with overall 8 attacks. In contrast, all other victims suffered between 1 and 3 attacks.
SHOWING CHANGES
SHOWING CHANGES

Comparison of EDV, P1, and P2 across 2010 and 2011.
DENSITY PLOT
2D DENSITY PLOTS
BOX PLOTS

- High value
- 75th percentile
- Median (50th percentile)
- 25th percentile
- Low value

Spread (100% of the values)
Midspread (50% of the values)
SCATTERPLOT
CLUTTERING, OVERPLOTTING
alpha = 1/10

alpha = 1/100
A FEW EXAMPLES AND CASE STUDIES
The Office for National Statistics (ONS) said gross domestic product (GDP) expanded by 0.7 per cent in the fourth quarter - an increase from the 0.6 per cent calculated on the watchdog’s first look at the economy.

Source: http://www.dailymail.co.uk/news/article-4248690/Economy-grew-0-7-final-three-months-2016.html
Procent użytkowników rolnych w gospodarstwach > niż 50 ha:

1989

25%

Source: http://weekend.gazeta.pl/weekend/1,152121,20528386,wierza-mocniej-niz-miastowi-za-to-zarabiaja-mniej-i-calym.html
VISUALIZATION TAXONOMY

Borkin MA, VoAA, Bylinskii Z, Isola P, Sunkavalli S, OlivaA, Pfister H.
What Makes a Visualization Memorable?

http://vcg.seas.harvard.edu/publications/what-makes-visualization-memorable
AREA
<table>
<thead>
<tr>
<th>Category</th>
<th>1234</th>
<th>X45</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>543.2109</td>
<td>7%</td>
</tr>
<tr>
<td>Group</td>
<td>45.67</td>
<td>45%</td>
</tr>
<tr>
<td>Unit</td>
<td>9876</td>
<td>98%</td>
</tr>
<tr>
<td>Class</td>
<td>123.78</td>
<td>12%</td>
</tr>
</tbody>
</table>

Text Based

Phrase Net

Word Cloud

Word Tree

Title

• Sed digerum vehicula
• Noli qua comque
• Sed ut et rhincus edio
• Integer et adis

Heading 1

• Nemo aequip river et habere
• Serious itur non temptum quae super
• In est, semper et

Heading 2

• Sed etur quae vehicula
• Frigida non
• Sed etur aequip

Heading 3

• Chart 1
• Chart 2
• Chart 3
• Chart 4
VISUAL TAXONOMY

The Data Visualisation Catalogue

Search by Function

Arc Diagram, Area Graph, Bar Chart, Box & Whisker Plot, Brainstorm, Bubble Chart

View by List

Bubble Map, Calendar, Chord Diagram, Choropleth Map, Circle Packing, Connection Map

http://www.datavizcatalogue.com/
TAKEAWAY MESSAGES

• Appropriate chart type for specific data type and visualization task