Overview

- Fundamental principles of graphic design and visual communication
  - help you create more effective information visualizations.

- Use of salience, colour, consistency and layout
  - communicate large data sets and complex ideas with greater immediacy and clarity.
Why Visualise? To see what’s in the data

Anscombe’s quartet

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

Regression between x and y in each case: $y = 3 + 0.5x$
Information Visualization

- 2 main objectives
- Data analysis
  - understand the data
  - derive information from them
  - involves comprehensivity
- Communication
  - of information
  - involves simplification

How do we get from Data to Visualization?

Need to understand

- the properties of the data or information
- the properties of the image
- the rules mapping data to images
How do we get from Data to Visualization?

Need to understand

- the properties of the data or information
- the properties of the image
- the rules mapping data to images
Types of Data

- Nominal (labels or types)
  - Sex: Male, Female,
  - Genotype: AA, AT, AG...

- Ordinal
  - Days: Mon, Tue, Wed, Thu, Fri, Sat, Sun
  - Abundance: abundant - common – rare

- Quantitative
  - Physical measurements: temperature, expression level

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

- **1D** e.g. DNA sequences
- **Temporal** e.g. time series microarray expression
- **2D** e.g. distribution maps
- **3D** e.g. Anatomical structures
- **nD** e.g. Fisher’s Iris data set
- **Trees** e.g. Linnean taxonomies, phylogenies
- **Networks** e.g. Metabolic pathways
- **Text and documents** e.g. publications

B. Shneiderman, The eyes have it: A task by data type taxonomy for information visualization, 1996
Types of Data

- Tabular
  - Categorical/Nominal
    - Male
    - Female
  - Ordinal
    - Abundant
    - Common
    - Rare
  - Quantitative
    - 10 mm
    - 15.5 mm
    - 23 mm
  - Relational
    - Trees
    - Networks
  - Spatial
    - Maps

slide adapted from Munzner 2011, Visualization Principles
How do we get from Data to Visualization?

Need to understand

- the properties of the data or information
- the properties of the image
- the rules mapping data to images
Theory of Graphics

- Application of human perception
  - understand and memorize forms in an image
  - XY dimensions of the plane and variation in Z dimension

- Correspondence between data and image

- Level of perception required by objective

- Mobility or immobility of the image

Semiology of Graphics

- visual encoding
  - points, lines, areas
    - patterns, trees/networks, grids
  - positional: XY
    - 1D, 2D, 3D
  - retinal: Z
    - size, lightness, texture, colour, orientation, shape,
  - temporal:
    - animation
Language of Graphics

- Graphics can be thought of as forming a sign system:
  - Each mark (point, line, or area) represents a data element.
  - Choose visual variables to encode relationships between data elements
    - difference, similarity, order, proportion
    - only position supports all relationships

- Huge range of alternatives for data with many attributes
  - find images that express and effectively convey the information.
Accuracy of Quantitative Perceptual Tasks

More accurate

position

length

angle

tslope

area

volume

density

colour

Less accurate

Gestalt Effects

- Visual system tries to structure what we see into patterns
- Gestalt is the interplay between the parts and the whole
  - “The whole is ‘other’ than the sum of its parts.” – Kurt Koffka

Gestalt Laws/Principles
Principle of Simplicity

- Every pattern is seen such that the resulting structure is as simple as possible.
  - Different projections of same cube
  - Perceived as 2 or 3 D
  - Depending on the simpler interpretation
Principle of Proximity

Things that are near to each other appear to be grouped together.
Principle of Similarity

Similar things appear to be grouped together
Variable Opacity for Clarity

- Use of similarity of stroke and opacity to clarify image
  - Layers in the image

M. Krzynski, behind every great visualization is a design principle, 2012
The law of closure posits that we perceptually close up, or complete, objects that are not, in fact, complete.
Principle of Connectedness

- Things that are physically connected are perceived as a unit

- Stronger than colour, shape, proximity, size
Principle of Good Continuation

Points connected in a straight or smoothly curving line are seen as belonging together

- lines tend to be seen as to follow the smoothest path
Principle of Common Fate

Things that are moving in the same direction appear to be grouped together
Principle of Familiarity

- Things are more likely to form groups if the groups appear familiar or meaningful.
Figure-Ground & Smallness

- Smaller areas seen as figures against larger background
- Surroundedness
Principle of Symmetry

The principle of symmetry is that, the symmetrical areas tend to be seen as figures against the asymmetrical background.
3D Effect
Context affects perceptual tasks

- Comparing values
  - Length
  - Curvature
  - Area
  - 2.5D shape
  - Position in 2.5D
Ambiguous Information: Length
Ambiguous Information: Length

Muller-Lyer
Horizontal-Vertical Illusion
Ambiguous Information: Curvature
Ambiguous Information: Area (Context)
2.5D Shape

Adapted from Shepard R N, 1990 Mind Sights: Original Visual Illusions, Ambiguities, and other Anomalies
2.5D Shape

Adapted from Shepard R N, 1990 Mind Sights: Original Visual Illusions, Ambiguities, and other Anomalies ( 
Ambiguous Information: Position in 2.5D space
Preattentive Visual Features

- the ability of the low-level human visual system to rapidly identify certain basic visual properties
- a unique visual property e.g., colour red allows it to "pop out"
- aids visual searching

Adapted from Perception in visualization C. Healey, : http://www.csc.ncsu.edu/faculty/healey/PP/
Preattentive Visual Features

- Some more effective than others

- closure
- curvature
- length

Adapted from Perception in visualization C. Healey, : http://www.csc.ncsu.edu/faculty/healey/PP/
Preattentive Visual Features

- Flicker
- Direction of movement
- Enclosure/containment

Adapted from Perception in visualization C. Healey, : http://www.csc.ncsu.edu/faculty/healey/PP/
More than 2 Preattentive visual features

A target made up of a combination of non-unique features normally cannot be detected preattentively

- spot the red square
- difficult to detect
- serial search required

Adapted from Perception in visualization C. Healey, : http://www.csc.ncsu.edu/faculty/healey/PP/
Boundary detection

Horizontal boundary

Vertical boundary

Adapted from Perception in visualization C. Healey, : http://www.csc.ncsu.edu/faculty/healey/PP/
Region tracking
Use of preattentive features

- target detection:
  - users rapidly and accurately detect the presence or absence of a "target" element with a unique visual feature within a field of distractor elements

- boundary detection:
  - users rapidly and accurately detect a texture boundary between two groups of elements, where all of the elements in each group have a common visual property

- region tracking:
  - users track one or more elements with a unique visual feature as they move in time and space, and

- counting and estimation:
  - users count or estimate the number of elements with a unique visual feature.
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
COLOUR
Colour

“Colour used poorly is worse than no colour at all” - Edward Tufte

- “Above all, do no harm”
- colour can cause the wrong information to stand out and
- make meaningful information difficult to see.
Colour space

A colour space is a mathematical model for describing colour.

- RGB, HSB, HSL, Lab and LCH
- RGB is the most common in computer use,
  - but least useful for design
  - our eyes do not decompose colours into RGB constituents
- HSV, describes a colour in terms of its hue, saturation and value (lightness),
  - models colour based on intuitive parameters
  - more useful.
Colourimetry

- **Hue (colour)**
  - around the circle

- **Saturation**
  - Inside to outside
  - Colour to grey scale

- **Lightness (value)**
  - top to bottom

Figs. Courtesy of S Rogers, ONS
Brewer Palettes

- Brewer palettes (colorbrewer.org) provide a range of palettes based on HSV model which make life easier for us.

Avoid the use of hue to encode quantitative variables

Quantitative encoding e.g. heat maps

Two-sided quantitative encodings

**QUALITATIVE**
- set1
- set2
- pastel2
- dark2

**SEQUENTIAL**
- blues
- greens
- reds
- ylorbr

**DIVERGING**
- spectral
- rdylbu
- rdylgn
- piyg

Fig. Courtesy of M. Krzywinski,
Examples

Poor use of colour

**one color dominates**

**difficult to distinguish**

**murky**

Brewer colours

**recolored with Brewer palettes**

M. Krzewinski, behind every great visualization is a design principle, 2012
Conversion to Grey scale

- Ensure chosen colour set works well in grey scale
  - Sequential palette works well here

Fig. Courtesy of M Krzywinski
Trouble with perceptual colour....

Figs. Courtesy of S Rogers, ONS
Context Affects Perceived Colour

Figs. Courtesy of S Rogers, ONS
Accessibility (W3C): 10-20% of population are red/green colour blind. (74? 21? No number at all?)....
Colour Blindness

8% males of USA descent

Red-green

Red-green

Blue-yellow

Fig. Courtesy of M Krzywinski
BioVis Example: Immunofluorescence images

red-green image of P2Y1 receptor and migrating granule neurons, effectively remapped to magenta-green using the channel mixing method.
Blue-Yellow
- might be better than

Green-Magenta
- talk about same colours
From Data to Visualization...

- The properties of the data or information
- The properties of the image
- The rules mapping data to images
Encoding Schemes

Adapted from Mackinlay J (1986) Automating the design of graphical presentations of relational information.
Mapping data types to encoding

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density</td>
<td>Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Saturation</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density</td>
</tr>
<tr>
<td>Density</td>
<td>Containment</td>
<td>Saturation</td>
</tr>
<tr>
<td>Saturation</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Volume</td>
</tr>
</tbody>
</table>

Mackinlay J (1986) Automating the design of graphical presentations of relational information.
Don’t forget Salience…

- Physical properties that set an object apart from its surroundings
  - Distinct features have high salience
- Encodings have differences in discrimination and accuracy
- Context affects salience
- Choose salient encodings for primary navigation
  - Colour is good for categories - salience decreases with more hues.
- Focus attention by increasing salience of interesting patterns
- Unexpected or bad things can happen when unimportant elements in a figure are salient.
  - The reader will use salience to suggest what is important.
Example Encodings in BioVis…

**DNA sequence** – 1D, Nominal data, colour

**Phylogeny** – Tree, Nominal data, position, colour

**Microarray gene expression** – 2D, Ordinal data, colour, position

**Microarray time-series** – temporal data, quantitative data, Position, height, colour
Examples

Sequence alignments – matrix, colour, position, length

Use of Symmetry, hue, saturation, length


Borkin et al, 2011 Evaluation of artery visualizations for heart disease diagnosis
Examples

- Circos
  - human genome
  - location of genes implicated in disease
  - regions of self-similarity
    structural variation within populations
  - Uses:
    links, heat maps, tiles, histograms
    Use of colour, good continuity, length, transparency, ..
Which Encoding?

▲ Challenge:
  ➢ Pick the best encoding from the large number of possibilities.
  ➢ Wrong visual encoding can mislead or confuse user

▲ Visual Representation should be expressive

▲ Principle of Consistency:
  ➢ The properties of the representation should match the properties of the data.

▲ Principle of Importance Ordering:
  ➢ Encode the most important information in the most “effective” way.
Expressiveness

Visual Representation encodes all the facts

- An nD (or 1:N) data set e.g. Iris data set

Fig. Courtesy of M Krzywinski
Expressiveness

- Encodes only the facts
  - Wrong use of a bar chart implies something better about Swedish cars than USA ones...

Adapted from Mackinlay J (1986) Automating the design of graphical presentations of relational information.
Consistency

- Visual variation in a figure should always reflect and enhance any underlying variation in the data.

- Avoid using more than one encoding to communicate the same information.

- Do not use visually similar encodings for independent variables.
Consistency

- red - processed genes, but salience attenuated
- other genes encoded with competing glyph - red star.
**Consistency**

Uniform size and alignment of exons and introns reduces complexity and aids interpreting their complex arrangement.

*spacing variation is implied*

![Diagram showing spacing variation in exons and introns](image1)

*variation refactored*

![Diagram showing variation refactored in exons and introns](image2)


M. Krzynski, behind every great visualization is a design principle, 2012
Consistency

Order items in a legend according to order of appearance in the plot

M. Krzwiniski, behind every great visualization is a design principle, 2012
Consistency - Navigational aids

- Use consistent axes when comparing charts

misleading

improved


M. Krzwinski, behind every great visualization is a design principle, 2012
Increase data:ink ratio

- **Navigational aids**
  - should not compete with the data for salience.

- **Avoid**
  - heavy axes,
  - error bars and
  - glyphs

Increase data:ink ratio

Avoid unnecessary containment

Fig. Courtesy of M Krzywinski
Increase data:ink ratio

Avoid “Chart junk”


M. Krzwinski, behind every great visualization is a design principle, 2012
Keep things simple - Avoid 3D

3D scatter plots are better as a series of 2D projections.


M. Krzwiniski, behind every great visualization is a design principle, 2012
The potential to overcome well known problems with static imagery....
Change Blindness....

SPOT THE DIFFERENCE

Fig. courtesy of S Rogers, ONS, UK
Interaction

- Supports the user in exploring data

- Shneiderman’s Information Seeking Mantra:
  - Overview first, zoom and filter, then details on demand
Interaction: operations on the data

- sorting
- filtering
- browsing / exploring
- comparison
- characterizing trends and distributions
- finding anomalies and outliers
- finding correlation
- following path
Interaction: Techniques to support operations

- Re-orderable matrices - sorting
- Brushing - browsing
- Linked views – comparison, correlation, different perspectives
  - Linking
- Overview and detail -
  - Excentric labelling
- Zooming – dealing with complexity/amount of data
- Focus & context - dealing with complexity/amount of data
  - Fisheye....
  - Hyperbolic
- Animated transitions - keeping context
- Dynamic queries - exploring
Sortable tables, matrices

<table>
<thead>
<tr>
<th>Club</th>
<th>Total</th>
<th>High</th>
<th>Avg</th>
<th>Capacity</th>
<th>Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFC Bournemouth</td>
<td>53513</td>
<td>6924</td>
<td>5361</td>
<td>9938</td>
<td></td>
</tr>
<tr>
<td>Boston United</td>
<td>30274</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bristol Rovers</td>
<td>56840</td>
<td>9447</td>
<td>6684</td>
<td>11975</td>
<td></td>
</tr>
<tr>
<td>Bury</td>
<td>29705</td>
<td>6044</td>
<td>11669</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambridge United</td>
<td>43795</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carlisle United</td>
<td>50744</td>
<td>10684</td>
<td>5024</td>
<td>16651</td>
<td></td>
</tr>
<tr>
<td>Darlington</td>
<td>39200</td>
<td>5163</td>
<td>3895</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exeter City</td>
<td>35012</td>
<td>6065</td>
<td>3903</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hartlepool United</td>
<td>48251</td>
<td>6360</td>
<td>4835</td>
<td>7229</td>
<td></td>
</tr>
<tr>
<td>Hull City</td>
<td>83322</td>
<td>11885</td>
<td>9323</td>
<td>15160</td>
<td></td>
</tr>
<tr>
<td>Kidderminster Harriers</td>
<td>24371</td>
<td>3772</td>
<td>5622</td>
<td>12573</td>
<td></td>
</tr>
<tr>
<td>Leyton Orient</td>
<td>13570</td>
<td>6271</td>
<td>3009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lincoln City</td>
<td>44275</td>
<td>5627</td>
<td>4394</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macclesfield Town</td>
<td>20879</td>
<td>8335</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford United</td>
<td>52071</td>
<td>5658</td>
<td>5207</td>
<td>12500</td>
<td></td>
</tr>
<tr>
<td>Rochdale</td>
<td>28404</td>
<td>10249</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rushden &amp; Diamonds</td>
<td>39302</td>
<td>4329</td>
<td>39635</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scunthorpe United</td>
<td>31770</td>
<td>3873</td>
<td>2183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shrewsbury Town</td>
<td>52181</td>
<td>3738</td>
<td>8004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southend United</td>
<td>38592</td>
<td>5168</td>
<td>38123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swansea City</td>
<td>38920</td>
<td>6327</td>
<td>3911742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torquay United</td>
<td>30589</td>
<td>4937</td>
<td>6283</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrexham</td>
<td>56272</td>
<td>443</td>
<td>5500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>York City</td>
<td>27211</td>
<td>433</td>
<td>9034</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Linked Views, Linking, Brushing, Excentric Labels

Time-series Microarray Data
Focus and Context – Fisheye tree
Animated transitions, brushing
Drill Down, Select, Zooming, Focus & Context
Brushing, Filtering, Principle of Continuity
Linked Views, Filtering
Multiple Trees Comparison

- Brushing
- Focus & context
Zooming, filter
Smoothly Animated Transitions
Principles of Information Visualisation Tutorial: Part 2 - Design Process

Cydney Nielsen