Chapter 2 - Visual Perception
Optimizing Information Visualization regarding the human visual system

Vorlesung „Informationsvisualisierung”
Prof. Dr. Florian Alt, WS 2013/14

Konzept und Folien (4th revised edition):
Thorsten Büring, Andreas Butz, Michael Burch
Outline

• Definition & Context
• Preattentive Processing
• Visual Memory
• Gestalt Theory
• Data encoding
  – glyphs
  – color
Preattentive Processing

- Length / Width (=Size)

Images taken from http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Data Encoding
Characteristics of Visual Properties

- Some properties possess intrinsic meaning
  - Density with Grayscale: the darker the more
  - Size / Length / Area: the larger the more
  - Position: depending on culture, in Europe the leftmost / topmost are first
  - Color: depending on culture, e.g. white associated with death in Japan

- Accuracy of representations for quantitative measures (empirically verified by Cleveland & McGill, 1985)

(Mackinlay, 1988)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line length</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>2D position</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Line width</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Enclosure</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Hue</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Intensity</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
Encoding Data with Glyphs

- **Glyph**: graphical object designed to convey multiple data values
- **Multidimensional discrete data**, e.g., a collection of cars with several attributes each, e.g., horsepower, weight, acceleration etc.
- **What visual properties can be mapped to the attributes?**
- **FilmFinder example**
  - Color to film genre
  - X-position to year of release
  - Y-position to popularity
- **Additional properties**
  - Lightness
  - Shape
  - Orientation
  - Texture
  - Motion
  - Blinking

FilmFinder (www.cs.umd.edu/)
Encoding Data with Glyphs

• Limitations of low-level graphical attributes for glyph design
  • Easily resolvable steps of a visual property
    – 12 colors (for preattentive processing only 8 colors)
    – About 4 orientation steps
    – At most 4 size steps
    – Binary blink coding (on / off)
    – Texture unknown
    – Shape unknown

• Mixing visual properties
  – Properties are not independent from each other, e.g. blink coding interferes with motion coding
  – Conjunctions are usually non-preattentive
  – Some dimensions are integral
  – Best to restrict the mapping to color, shape, spatial position (and motion)
Color Vision & Model

• Human color vision
  – Sensory response to electromagnetic radiation in the spectrum 0.4 – 0.7 micrometers
  – Based on three dimensions (three different types of color receptors in human retina)

• Powerful encoding potential: compared to gray scales the number of just noticeable differences is much higher

• About 8% of the male and 1% of the female population are color-blind

• Color Model HSV (aka HSB)
  – Hue - blue, green, etc. (X axis)
  – Saturation – intensity of color (Y axis)
  – Value – light/dark (slider)
Color Scales

• Definition: pictorial representation of a set of distinct categorical or numerical values, where each value is assigned its own color (Levkovitz 1996)

• Desired properties of perception
  – Preserve the order of the data values, if any
  – Uniform distance between adjacent values (i.e. equally spaced numerical steps are perceived as equally spaced perceptual steps)
  – No artificial boundaries that do not exist in the data (i.e. continuously present continuous values)
Color Rules I

• Always ensure a reasonable luminance contrast between foreground and background color – chromatic variation may not be enough!

• Black and white borders around colored symbols can reduce contrast effects

• Canonical colors (close to an ideal) are easier to remember

• Only a small set of basic colors should be used for nominal (distinct) labeling
  – At most 12 colors: red, green, yellow, blue, black, white, pink, cyan, gray, orange, brown, purple
  – The first four colors are “hard-wired” into the human brain – should be used with priority
Grayscale

- Usually not considered a color scale, but very common
- Provides simple and natural sense of order
- Disadvantages
  - Limited number of just-noticeable-differences (JNDs) about 60 to 90
  - Contrast effects can significantly reduce accuracy
  - Luminance channel is fundamental to much of perception – grayscale encoding may be considered “a waste of perceptual resources” (Ware, 2000)
Rainbow for Ordering Data?

• Most common: rainbow scale for ordinal and quantitative (spectral colors)
  – Continuous spectrum
  – Common arbitrary division in 8 or less named colors (red, orange, yellow, green, cyan, blue, indigo, violet)

• Problems with rainbow scale
  – Can you order the color blocks from low to high?
  – Yellow (in the middle of the scale) may draw too much attention, when users are seeking for extreme values
  – Perception of non-existing boundaries

John Stasko
Recommended Color Scales

• Ordinal data
  – Low saturation to high saturation (single hue) - also very limited JNDs
  – Dark to light (single hue)
  – Red to green, yellow to blue, red to blue

• Ratio (hardly feasible) / diverging data
  – Neutral value (e.g. white) to represent zero
  – Increases in saturation toward distinct colors for positive and negative values (double-ended multiple hue)

• Look up www.colorbrewer2.org for inspiration
colorbrewer2.org
Redundant Color Scales

• Use multiple color properties to redundantly represent data
• Visual reinforcement of steps
• Overcome visual deficiencies
• Redundant model components: data values are mapped to both hue and brightness
• Heated-object scale
  – Going from black to white passing through orange and yellow
  – Monotonic increase in brightness provides more natural ordering than rainbow scale
• Linearized optimal color scale
  – Scale maximizing the number of JNDs while preserving a (more or less) natural order
Color Scale

- US presidential elections - Bush & RNC's campaign funding

Color Scale

- Vote distribution of 2004 US presidential election - the darker the color, the more of a landslide it was for the winning party

http://fundrace.huffingtonpost.com/moneymap.php
Color Scale

Sheelagh Carpendale
Color Scale
Color Rules II

- For larger areas on a white background use low-saturation light colors
- Small color-coded objects should be given high saturation
- Use red and green in the center of the field of view (edges of retina not sensitive for these)
- Use black, white, yellow in periphery
- Use color for grouping and search
- Generation of color families
  - Use canonical colors
  - Family members should differ by saturation
  - Better: saturation and lightness
Bivariate Color Coding

• Recap: color is three-dimensional
• Two data dimensions may be mapped to different color dimensions (e.g. hue and saturation, hue and lightness)
• Problem: bivariate color coding has been found notoriously difficult to read (Wainer & Francolini, 1980)
• The same applies to multidimensional color coding
  – E.g. amount of red, amount of green, amount of blue for coding colored dots in scatterplot (Ware & Beatty 1988)
  – Clusters could be easily identified by the participants of a user test
  – Precise decoding of the color components difficult
Example: Bivariate Color Coding

Chapter 3 - Multidimensional Information Visualization I

Concepts for visualizing univariate to hypervariate data

Vorlesung „Informationsvisualisierung”
Prof. Dr. Florian Alt, WS 2013/14

Konzept und Folien (4th revised edition):
Thorsten Büring, Andreas Butz, Michael Burch
Outline

• Reference model and data terminology
• Visualizing data with < 4 variables
• Visualizing multivariable data
  – Geometric transformation
  – Glyphs
  – Pixel-based
  – Dimensional Stacking
  – Downscaling of dimensions
• Case studies: support for exploring multidimensional data
  – Rank-by-feature
  – Value & relation display
  – Dust & magnet
• Clutter reduction techniques
Reference Model and Terminology
Information Visualization

• The use of computer-supported, interactive, visual representations of abstract data to amplify cognition (Card et al. 1999)
• How to construct interactive visual representations?
• Reference Model for Visualization

Card et al. 1999

![Diagram](Image)
Multivariate Data

• Data collected on several variables for each sampling unit.

• For example, if we collect information on weight (w), height (h), and shoe size (s) from each of a random sample of individuals, then we would refer to the triples (w₁, h₁, s₁), (w₂, h₂, s₂),... as a set of multivariate data.

• Metadata (Data about data)
  – Descriptive information about the data
  – Units, for example mph or inches
  – Data types
Data Table

- **Cases (observations)**
- **Variables (aka attributes)**
- **Example car data set**
  - 406 cases
  - 8 variables for each case

- **Metadata**
  - Descriptive information about the data
  - Units, e.g. lbs., mph, inches
  - Constraints, e.g. if var₁ is ‘41’, then var₇ can only be ‘11’ or ‘3’
  - Data types

<table>
<thead>
<tr>
<th></th>
<th>Variableₓ</th>
<th>Variableᵧ</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caseᵢ</td>
<td>Valueᵢₓ</td>
<td>Valueᵢᵧ</td>
<td>...</td>
</tr>
<tr>
<td>Caseⱼ</td>
<td>Valueⱼₓ</td>
<td>Valueⱼᵧ</td>
<td>...</td>
</tr>
<tr>
<td>Caseₖ</td>
<td>Valueₖₓ</td>
<td>Valueₖᵧ</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>mpg</th>
<th>cylinders</th>
<th>engine displ.</th>
<th>horsepower</th>
<th>weight</th>
<th>acceleration</th>
<th>prod. year</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chevrolet C. M.</td>
<td>18</td>
<td>8</td>
<td>307</td>
<td>130</td>
<td>3504</td>
<td>12</td>
<td>70</td>
<td>USA</td>
</tr>
<tr>
<td>Datsun PL510</td>
<td>27</td>
<td>4</td>
<td>97</td>
<td>88</td>
<td>2130</td>
<td>14,5</td>
<td>70</td>
<td>Asia</td>
</tr>
<tr>
<td>Audi 100 LS</td>
<td>24</td>
<td>4</td>
<td>107</td>
<td>90</td>
<td>2430</td>
<td>14,5</td>
<td>70</td>
<td>Europe</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Dimensionality of Data

• On how many variables was a data case measured?
  • 1 variable – Univariate
  • 2 variables – Bivariate
  • 3 variables – Trivariate
  • > 3 variables – Hypervariate = multivariate = multivariable data

• Visualizations that encode multivariable data are called multidimensional visualizations

• Visualizing multivariable data is one of the most challenging tasks in Information Visualization
Data Types

• Nominal (categorical)
  – Unordered set
  – Operators: =, ≠
  – Example: car origin (Europe, USA, Asia)

• Ordinal
  – Possess a natural order
  – Operators: <, >
  – Example: ratings, school grades

• Quantitative
  – Allow for arithmetic operations
  – Operators: *, /, +, -
  – Example: acceleration in seconds

• Also subtypes exist: e.g., quantitative geographic (geographic coordinates), quantitative time
Data Transformation

• Transformation of raw data into data tables can involve loss or gain of information
  – Classing: quantitative to ordinal data by dividing values into ranges, e.g., acceleration of cars into <slow, medium, fast>
  – Nominal to ordinal data by sorting the values lexicographically, e.g., products in a supermarket
  – Derived values e.g., calculating statistical summaries (mean, median...)
  – Derived structures (e.g., sorting cases and / or variables)
  – Sampling (determining a representative subset of the data set)
  – Aggregation of data (e.g., determining frequencies, average sum, maximum)

• Deal with errors, missing values and duplicates
Objectives of Visual Structures

- Various mappings possible
- Quality factors of mapping
  - Expressiveness - all and only the data in the data table are represented in the structure
  - Increased effectiveness compared to another mapping
  - Faster to interpret
  - Can convey more distinctions
  - Leads to fewer errors in interpretation
  - See previous lecture on perception!

Card et al. 1999
Univariate Data
Univariate Data

• Example 1D Visualization for Distribution of Variable Values

Plot
Univariate Data

- Example 1D Visualization for Distribution of Variable Values

Boxplot
Univariate Data

- Example 1D Visualization for Distribution of Variable Values

Line graph - not very reasonable in this case
Univariate Data

• Example 1D Visualization for Distribution of Variable Values

Histogram
Univariate Data

• Example 1D Visualization for Distribution of Variable Values

Bargram
Histogram Distribution Analysis

Images from Field & Hole 2003
Interactive Bargrams

Bivariate Data
Bivariate Data

• Most common for displaying bivariate data is the scatterplot
• Each spatial dimension is assigned a (usually quantitative) axis variable
• Cases are mapped to a spatial position according to the data values for the axes
• Users can easily identify global trends, local trade-offs, outliers …
• Potential problems?
Bivariate Data

- Can provide answers to the following questions
  - Are variables X and Y related?
  - Are variables X and Y linearly related?
  - Are variables X and Y non-linearly related?
  - Does the variation in Y change depending on X?
  - Are there outliers?
Scatterplot Analysis

- No relationship
Scatterplot Analysis

• Exact linear (positive correlation)
Scatterplot Analysis

- Strong linear (positive / negative correlation)
Scatterplot Analysis

- Homoscedastic
- Heteroscedastic
Scatterplot Analysis

Quadratic relationship

Exponential relationship

Sinusoidal relationship (damped)

Outlier

Experiment

• Is there a correlation between shoe size and body height?
• Is it depending on the country a person is born in?

Find it out by a visualization!

• How do you display the correlation?
• How do you display the place of birth?
• Are there any outliers?
Algorithmical Scatterplot Analysis

• For example k-means clustering (also for >2 dimensions)
• In statistics and data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.
• Partitioning of the data space into Voronoi cells.
Time-Based Bivariate Data

• Plot of time series
  – X-axis represents time
  – Y-axis a function of time

• Closing prices of 1,430 individual stocks across 52 weeks

TimeSearcher, Hochheiser & Shneiderman 2004
Time Map

  - X-axis: month
  - Y-axis: years and weekdays (Sunday to Saturday)
  - 4 categories of ozone concentration mapped to distinct colors
- Reveals seasonal patterns
  - Ozone levels are much higher in summer months
  - High ozone days have steadily decreased
- How could this visualization be improved?

Image taken from Mintz et al. 1997
Geographic Bivariate Data

• Size of each territory shows relative proportion of the world population living there

• Potential problem with this visualization?

Image taken from worldmapper.org
Trivariate Data
Trivariate Data

• Tempting: map each variable to each dimension of a 3D scatterplot
• Occlusion of points with different positions
• Problem with static representation?

Occlusion of points with different positions and in general difficulty to visually map the (x,y,z)-coordinates of the single points
Trivariate Data

• Idea 1:
  – Map each variable to a dimension (axis) of a 3D scatterplot
  – Enhancement by using guiding lines

Still problematic:
Many lines lead to visual clutter and even more occlusions
Trivariate Data

• Idea 2:
  – Map each variable to a dimension (axis) of a 3D scatterplot
  – Rotation is used to explore the data but if the dataset is very dense we have to also look inside it

More complex interactive features needed to analyze the trivariate data