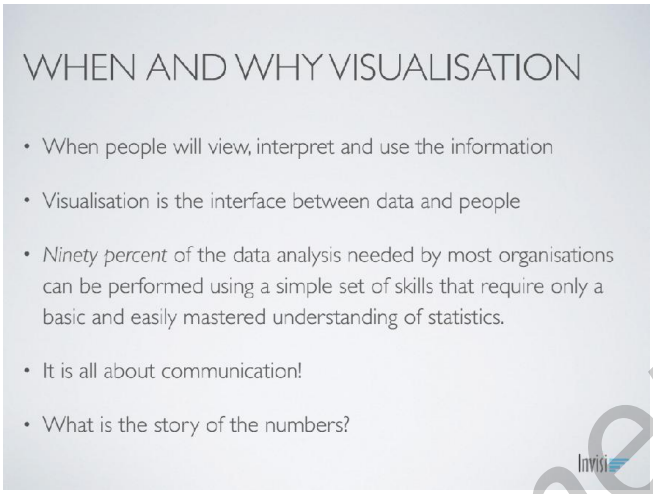
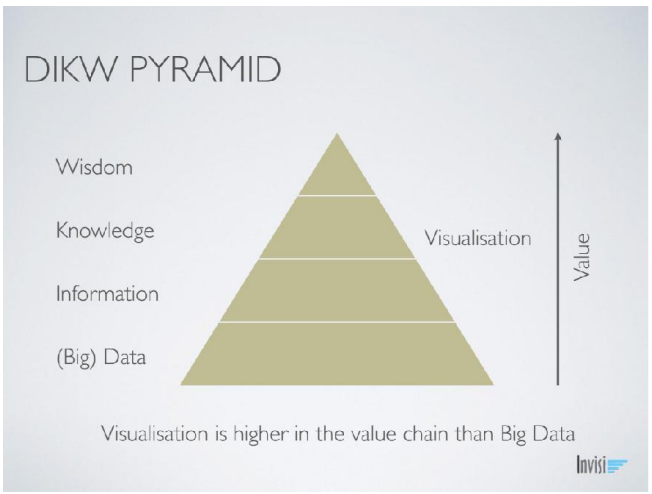
BIG INSIGHT: VISUALISATION IN RELATION TO BUSINESS INTELLIGENCE, BIG DATA AND DATA SCIENCE

Dirk Marc Guy STROO, PhD
Director Invisi BV

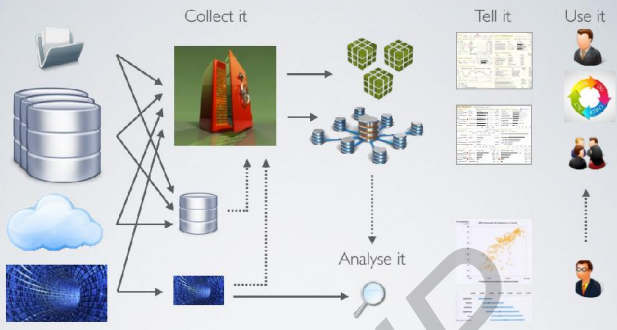
InvisiBV, Netherlands, nl.linkedin.com/in/marcstroo





CONSUMERS OF INFORMATION

- Role in organisation:
  - Board and senior management
  - Knowledge workers
  - *Analysts*
- Type of consumption:
  - Operator
  - Tourist
  - Farmer
  - *Discoverer*
  - *Miner*



A BI ARCHITECTURE

Including place of visualisation and Big Data

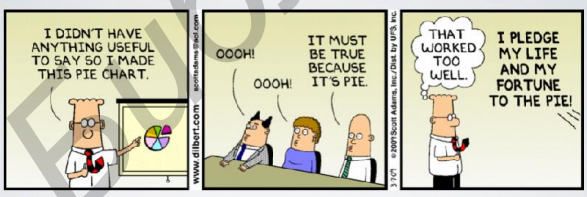
INFORMATION VISUALISATION

- According to Card, Mackinlay and Schneidermann, the characteristics of Information Visualisation are:
  - Computer-supported
  - Interactive
  - Visual representations
  - Abstract data
  - Amplify cognition

TRAITS OF MEANINGFUL DATA

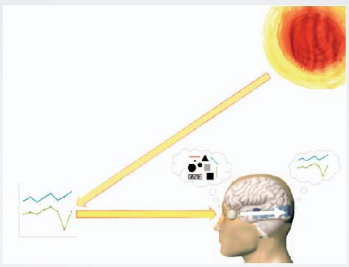
- High volume
- Historical
- Consistent
- Multivariate
- Atomic
- Clean
- Clear
- Dimensionally structured
- Richly segmented
- Of known pedigree

GOOD AND BAD VISUALISATION



THINKING WITH OUR EYES

- Approximately 70% of the body's sense receptors reside in our eyes

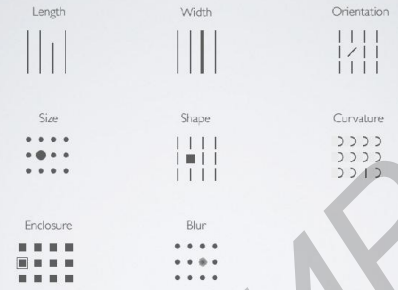


PRE-ATTENTIVE ATTRIBUTES OF VISUAL PERCEPTION

We perceive several basic attributes of visual images pre-attentively, that is, without the need for conscious awareness.

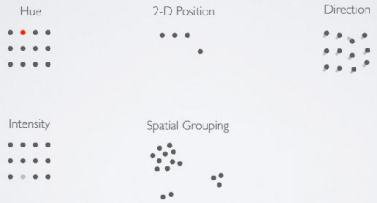
Invisi

FORM



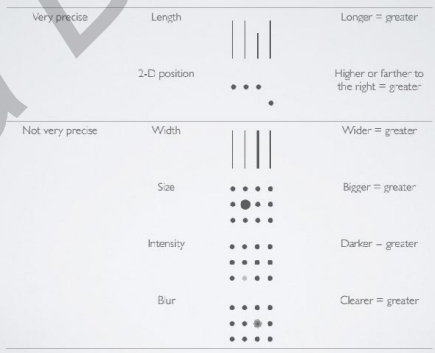
Invisi

COLOUR, SPATIAL POSITION AND MOTION



Invisi

PRE-ATTENTIVE ATTRIBUTES TO REPRESENT QUANTITATIVE VALUES



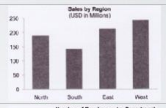
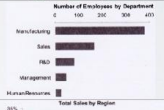
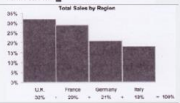
Invisi

ANALYTICAL INTERACTION

- Comparing
- Sorting
- Adding variables
- Filtering
- Highlighting
- Aggregating
- Re-expressing
- Re-visualising
- Zooming and panning
- Re-scaling
- Accessing details on demand
- Annotating
- Bookmarking

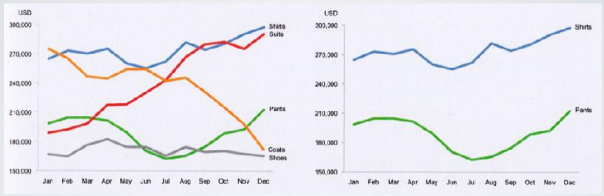
Invisi

EXAMPLES OF COMPARISONS

Type	Description	
Nominal	Comparing values that have no particular order	
Ranking	Comparing values that are arranged by magnitude, from low to high or high to low	
Part-to-whole	Comparing values that when combined make up parts of a whole	

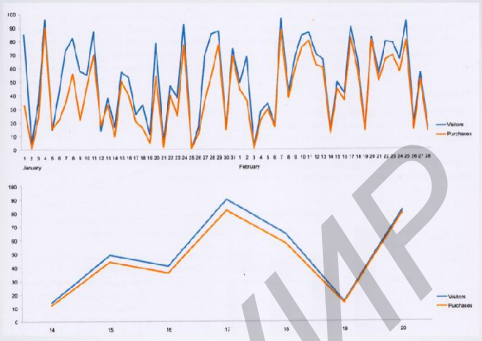
Invisi

EXAMPLE OF FILTERING



Invisi

EXAMPLE OF ZOOMING



Invisi

ANALYTICAL NAVIGATION

Directed

Exploratory

Invisi

ANALYTICAL NAVIGATION

- Directed analysis begins with a specific question that we hope to answer
- With exploratory analysis, we begin by simply looking at data without predetermining what we might find
- Information visualisation is ideal for exploratory data analysis. Our eyes are naturally drawn to trends, patterns, and exceptions that would be difficult or impossible to find using more traditional approaches, such as tables of text, including pivot tables

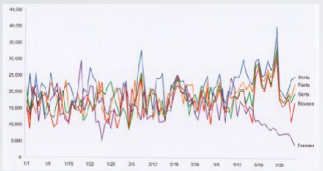
Invisi

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

– Shneiderman's Mantra

Invisi

SHNEIDERMAN'S MANTRA IN ACTION



Start

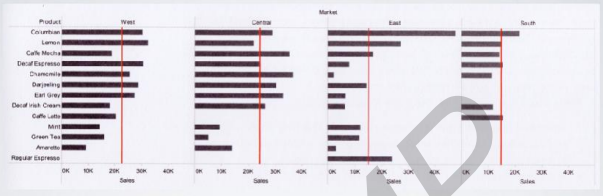
Result

Invisi

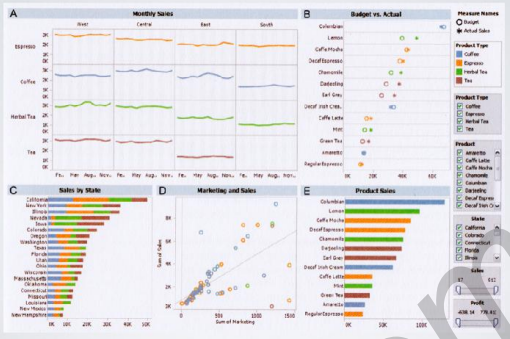
ANALYTICAL TECHNIQUES AND PRACTICES

- Optimal quantitative scales
- Reference lines and regions
- Trellises and crosstabs
- Multiple concurrent views and brushing
- Focus and context together
- Details on demand
- Over-plotting reduction

EXAMPLE OF REFERENCE LINES



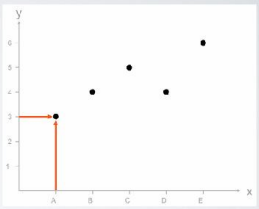
EXAMPLE OF MULTIPLE CONCURRENT VIEWS



GUIDELINES FOR REPRESENTING QUANTITATIVE DATA

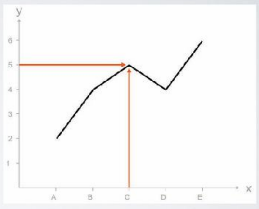
POINTS

Points encode individual values as 2-D position



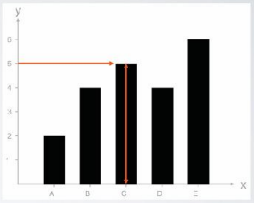
LINES

Lines encode individual values as 2-D position (at each data point connected by the line), but by connecting the data values the added characteristics of slope and direction also carry information



BARS

Bars encode individual values as 2-D position at the endpoint of the bad, and also the height (vertical bars) or length (horizontal bars)



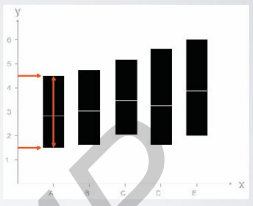
Invisi

BOXES

Boxes look and mostly function like bars, except that they encode two values, rather than one.

They represent a range of values by using the one end of the box to encode the lowest value in the range and the other end to encode the highest value.

Like bars, boxes encode values both as 2-D location based on the position of each end, and as line length based on the length of the box.



Invisi

PATTERNS

- High, low and in between
- Going up, going down and remaining flat
- Steep and gradual
- Steady and fluctuating
- Random and repeating
- Straight and curved
- Non-intersecting and intersecting
- Symmetrical and skewed
- Wide and narrow
- Clusters and gaps
- Tightly and loosely distributed
- Normal and abnormal

Invisi

Invisi

DIVERSE TYPES OF VISUAL ANALYTICS

- Time-Series Analysis
- Part-to-whole and Ranking Analysis
- Deviation Analysis
- Distribution Analysis
- Correlation Analysis
- Multivariate Analysis

Invisi

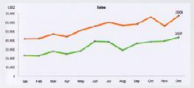
TIME-SERIES PATTERNS

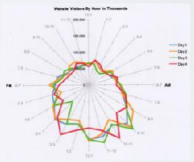
- Trend
- Variability
- Rate of change
- Co-variation
- Cycles
- Exceptions

Invisi

TIME-SERIES DISPLAYS

- Line graphs
- Bar graphs
- Dot plots
- Radar graphs
- Heatmaps





Invisi

ANIMATED SCATTERPLOTS FOR ANALYSING CORRELATION CHANGES

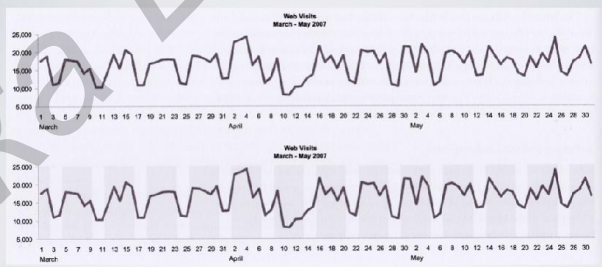
Data points can be animated (given motion) to display how the relationship between two quantitative values changed through time. Hans Rosling popularised this.

TIME-SERIES TECHNIQUES AND BEST PRACTICES

- Aggregating to various time intervals
- Viewing time periods in context
- Grouping related time intervals
- Using running averages to enhance perception of high-level patterns
- Omitting missing values from a display
- Optimising a graph's aspect ratio
- Using logarithmic scales to compare rates of change
- Overlapping time scales to compare cyclical patterns
- Using cycle plots to examine trends and cycles together
- Combining individual and cumulative values to compare actuals to a target
- Shifting time to compare leading and lagging indicators
- Stacking line graphs to compare multiple variables
- Expressing time as 0-100% to compare asynchronous processes

Invisi

GROUPING RELATED TIME INTERVALS



By clearly separating weekdays from weekends we can, for example, differentiate expected drops in Web traffic on weekends from unexpected drops on weekdays, which ought to be investigated.

PART-TO-WHOLE AND RANKING PATTERNS

- Uniform
- Uniformly different
- Non-uniformly different
- Increasingly different
- Decreasingly different
- Alternating differences
- Exceptional

Invisi

PART-TO-WHOLE AND RANKING DISPLAYS

- Pie charts
- Bar graphs
- Dot plots
- Pareto charts

Invisi

PART-TO-WHOLE AND RANKING TECHNIQUES AND BEST PRACTICES

- Grouping categorical items in an ad hoc manner
- Using Pareto charts with percentile scales
- Re-expressing values to solve quantitative scaling problems
- Using line graphs to view ranking changes through time

DEVIATION ANALYSIS DISPLAYS

- Deviation analysis doesn't require fancy visualisations. The two best are:
  - Bar graph
  - Line graph
- Both should display as a reference line the set of values to which the other values will be compared.

DEVIATION ANALYSIS TECHNIQUES AND BEST PRACTICES

- Expressing deviations as percentages
- Comparing deviations to other points of reference

DESCRIBING DISTRIBUTIONS

- Key visual characteristics of distributions:
  - Spread
  - Center Shape

DISTRIBUTION PATTERNS

- Shape
  - Curved or flat?
  - If curved, upward or downward?
  - If curved upward, single or multiple peaked?
  - If single peaked, symmetrical or skewed?
  - Concentrations?
  - Gaps?
- Outliers

DISTRIBUTION DISPLAYS

- Single distribution displays:
  - Histograms
  - Frequency polygons
  - Strip plots
  - Stem-and-Leaf plots
- Multiple distribution displays:
  - Box plots
  - Multiple strip plots
  - Frequency polygons
  - Distribution deviation graphs

DISTRIBUTION TECHNIQUES AND BEST PRACTICES

- Keeping intervals consistent
- Selecting the best interval
- Using measures that are resistant to outliers

DESCRIBING CORRELATIONS

- Key characteristics of correlations:
  - Direction
  - Strength
  - Shape
- Statistical summaries of correlations:
  - Linear correlation of coefficient (expressed in formulas as r)
  - Coefficient of determination (expressed in formulas as r^2)

CORRELATION PATTERNS

- Shape
  - Is it straight or curved?
  - If curved, is it curved in one direction only or both?
  - If curved in one direction only, is it logarithmic, exponential, or some other shape?
  - If curved in both positive and negative directions, does it curve upward or downward?
  - Are there concentrations of values?
  - Are there gaps in the values?
- Outliers

CORRELATION DISPLAYS

No graph is more useful for examining correlations than a scatterplot. No other graph displays the correlation of two variables as well. The limitation is that it can compare two quantitative variables only.

The table lens can be used to detect possible correlations between many variables all at once in one display.

CORRELATION TECHNIQUES AND BEST PRACTICES

- Optimising aspect ratio and and quantitative scales
- Removing fill colour to reduce over-plotting
- Comparing data to reference regions
- Visually distinguishing data sets when they're divided into groups
- Using trend lines to enhance perception of the correlation's shape, strength, and outliers
- Using multiple trend lines to see categorical differences
- Removing the rough to see the smooth more clearly
- Using trellis and cross tab displays to reduce complexity and over-plotting
- Using grid lines to enhance comparisons between scatterplots

MULTIVARIATE PATTERNS

- In multivariate analysis, we examine patterns formed by several values that measure different attributes of something, which exhibit its *multivariate profile*.
- Multivariate analysis revolves around the following questions:
  - Which items are most alike?
  - Which items are most exceptional?
  - How can these items be combined into logical groups based on similarity?
  - What multivariate profile corresponds best to a particular outcome?

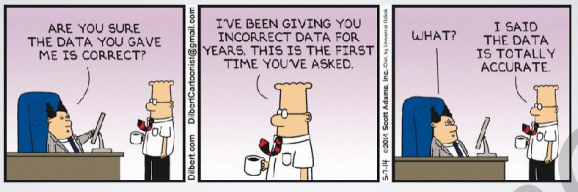
MULTIVATIATE DISPLAYS

- Glyphs
- Multivariate heatmaps
- Parallel coordinates plots

MULTIVARIATE TECHNIQUES AND BEST PRACTICES

- Ranking items by similarity
- Clustering items by similarity

RESUME: VISUALISE A GOOD STORY



NOISE

- A lot of data is noise
- It hase little to no informational value
- Collecting a lot of noise makes it harder to find valuable information

HOW TO BRING INSIGHT TO USERS

- Possibilities are:
  - Use an Agile development approach to deliver insights to users quickly
  - Use an entrepreneurial approach with concepting and prototyping to guide and educate users and sponsors
- Visualisation is only one of the success factors for implementing and using information in an organisation

LITERATURE: NOW YOU SEE IT

Now You See It does for data analysis what Stephen Few's book Show Me the Numbers does for data presentation: it teaches simple, fundamental, practical techniques that anyone can use. Only this time they're for making sense of information, not presenting it. These techniques rely primarily on something almost everyone has: vision.

LITERATURE: SIGNAL

In this age of so-called Big Data, organisations are scrambling to implement new software and hardware to increase the amount of data they collect and store. However, in doing so they are unwittingly making it harder to find the needles of useful information in the rapidly growing mounds of hay. If you don't know how to differentiate signals from noise, adding more noise only makes things worse.