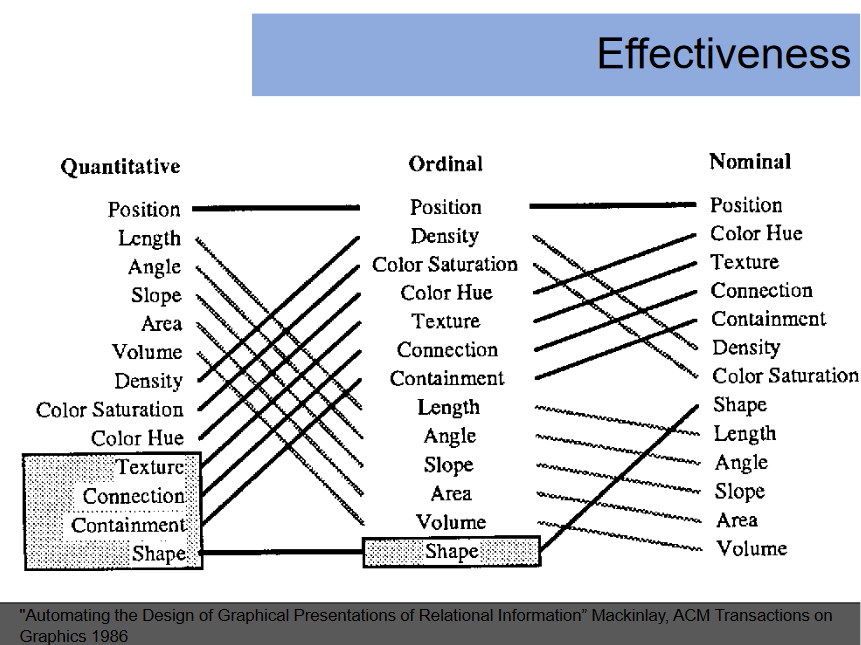
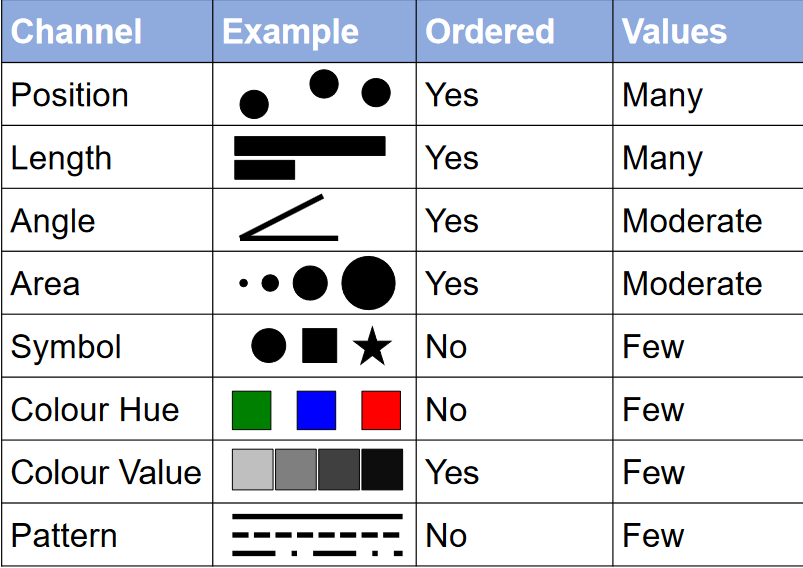
* Nominal Data: Labels and Categorical variables
  + Only deals in Binary operators
* Ordinal Data: Categorical variables which have an order
  + Binary operators and can also use > or <
* Quantitative (Ratio): Defined zero point.
  + Includes variables like mass, height, age
  + Operators: !=, = ,>,<, -, %
* Quantitative (Interval): Lacks absolute zero point
  + Operators: !=, = ,>,<, -, % etc
  + Includes variables like Year, temperature
* Column= Attribute
* Row= Item/observation
* Good visualization practices:
  + Left align text columns
  + Use consistent rounding
  + Right alight numerical columns
  + Show data values if precision is needed
  + If visualization doesn’t add anything, no need to add
  + Avoid angled/ vertical text
  + Remove background and chart background colour
  + No legend depending if data is labeled properly
  + Monochrome depending on graph type(most time keep monochrome), use colour deliberately
  + No need for grid lines
  + Sometimes one can remove vertical axis but add data values if doing so
  + Chart border is pointless
  + Don’t use 3D graphs, multiple graphs are better
  + For pie graphs, adding labels helps a lot especially if slices are similar in size
  + Bar chart axes must be at zero
  + Don’t obscure data with support
  + Consistent fonts
* Can filter data in tables by attributes
* Can group data together
* Can use aggregation on table such as min, max, count
* Can use data values from other columns to create a new one
* Bar Chart:
  + Each bar is different item
  + Uniform width as data isn’t encoded in width
  + Height represents data value
  + One quantitative data attribute for bar length
  + Horizontal ones can be used if length of text is long
  + Can have multiple attributes for y axis (Can either repeat graph for different regions or use colours)
  + Stacked bar charts are an option
* Dot Plots:
  + X: Values
  + Y: Categories
  + Dots are observation
  + Height/length is dependant on graph
* Line Chart:
  + Show trends/cycles over given time
  + Y is usually value and x is time/must be contniuous
  + Show points for emphasis
  + Can show multiple lines for comparisons
  + Can use slope graphs when only two times of interest exist
* Scatter graphs:
  + Point represents one observation
  + X and Y both quantitative
  + Used to show correlation between two quantitative variable
  + Can identify groups and outliers
  + Can plot every pair in a scatter plot matrix to compare all variables
* Histograms:
  + In depth view of numeric variable
  + Construct guide:
    - Divide data into bins
    - Count occurrence of each bin
    - Normalize counts
    - Pot on a bar graph of normalized counts
  + Width=range, height = frequency/probability
  + Shapes can reveal skewness/ modes
* Data-ink ratio = data-ink/total ink used in graphic
* Quantitative info is encoded in a graph via position (x,y), shape, size and colours
* Data visualisation is about mapping data dimensions to visual encodings
* 6 retinal variables in 2 dimensions:
  + Size
  + Value (brightness of colour)
  + Texture
  + Colour (Hue)
  + Orientation
  + Shape
* Marks are either points(0D), lines(1D) or areas(2D)
* Channels:
  + Position (horizontal, vertical or both)
  + Colour
  + Shape
  + Tilt
  + Size like length area and volumes
* Expressive: Encodes all and only facts in data
* Effective: If information is perceived more readily/easily
  + Visualization can be expressive but not necessarily effective
* A screenshot of a computer

  Description automatically generated



* Principle of Importance Ordering: Encode more important info more effectively
* Human perception of stimuli is non-linear
* Hypothesis: Educated guess about some aspect of world
* Position on common axis>length > angle as its easier to visualise information
* Discriminability: Choose channel which can accommodate number of distinct categories wish to encode
  + Ensure encoding has sufficient capacity to encode all values of data dimensions
    - Continuous data dimensions = continuous encodings
    - Categorical is hard as we are limited by possibilities
  + Limit colours as hard to tell apart
* Separability:
  + Channels can interfere with each other
* Redundancy: Avoid encoding redundant information or information already encoded
* Remember this for discriminability:  
  
* Stimuli: Features of environment which senses are capable of detecting
* Sensation: Physical response of a sense organ to stimuli
* Sense organs are passive and biologically understandable
* Perception: Psychological process of actively selecting and organising stimulus info detected by sensory organs so to create awareness
  + Active Process
  + Somewhat psychologically understandable
* Small subset of stimuli info is selected and organised to create conscious awareness
* Visual Attention: Various mechanisms which help determine which regions of an image are selected for detailed analysis
* Fixations: Detailed info from a small region is visible
* Saccades: Brief period which eyes flick to a new location
* The above two occur in a cycle and repeat three-four times per second
* 3 properties of visual attention:
  + Pre-attentive properties: Set of visual features which are detected by low level, fast-acting processes within period of single fixation
    - Low level vision systems can help make a quick summary of simple visual features are distributed across field of view
    - Colour>Shape here
    - Can use this to focus on areas of interest
  + Change Blindness: Major changes to a visual representation going unnoticed
  + Attention blindness
* Gestalt principles: Principles of how we organize disparate elements of sensory experience into a sensible whole
  + Key features include Proximity principle, similarity principle, connectedness principle and enclosure principle
  + All take the form of items related by (blank) tend to be grouped perceptually into higher order units
  + Can combine them
  + Tendencies
  + NO agreed ordering
  + Confounding Gestalt principles lessens the effect of any individual principle
* Proximity Principle: elements tend to be perceived as aggregated into  
  groups if they are near each other
* Similarity Principle: elements tend to be integrated into groups if they  
  are like each other
* Connectedness Principle: elements tend to be grouped together if they  
  are connected by other elements
* Enclosure principle: elements tend to be grouped together if they  
  are enclosed by other elements
* CIE Chromaticity Diagram:
  + - X Coordinate is a combination of red and green wavelengths while the Y coordinates is a combination of green and blue wavelengths.
    - Colour gamut is a subset of colours that can be represented by mixing colours at its corners
    - Spectral locus: The curved edge. Each point represents a pure hue/unsaturated color of specific wavelength
    - Chromaticity = Hue + Saturation
    - Entire colour range cant be displayed based on any 3 colours
    - 2D
    - Only perceivable colours
* Colour Models:
  + RGB (Red Green Blue)
    - Based on Cartesian coordinate system
    - X = Red, Y = Green, Z = Blue
    - Corners are RGB, CYM, BW values with black being at origin and white being (1,1,1)
  + HSL (Hue Saturation Lightness)
    - 3 Measures
      * Hue: Colour attribute which describes pure colour
      * Saturation: Gives a measure of how much a pure colour is diluted with white light
      * Lightness
  + The Munsell system
    - Perceptually uniform version of HSL space
    - 5 Principle Hues: Red, Yellow, Green, Blue and Purple with 5 intermediate Hues between principle hues
    - Value: Measures perceptual lightness (0 being black and white being 10)
    - Chroma: Purity of colour
* Colour blindness is a decreased ability to perceive colour differences under normal lighting conditions
  + Try using monochromatic , redundant encodings , colourblind safe palettes
  + Types include Protanopia, Deuteranopia and Tritanopia
* Limit colour to 6 colours and corresponding categories
  + If a lot of categories, use colour for things you wish to highlight and grey for everything else
  + Use qualitative scale hue for nominal and Sequential/Diverging scale for ordinal or quantitative
  + Avoid rainbow maps and maybe use viridis palette
* To visualize multivariate in 2D/3D we can do the following
  + Select Features
    - We can allow the user to choose different dimension via interacting with graph
  + Heatmaps
    - Each row = Row in data
    - Each column = Attribute in data
    - Each cell is a single value
    - Usually normalize values by point-min(x)/ Range of data
    - Data attributes must share common features or scales
  + Glyphs
    - Radar charts can be used: distance from center indicates quantitative attribute . Bad for nominal/Categorical data display
    - Glyphs are any kind of purposeful mark
    - Chenroff Faces uses a human like face for a glyph
      * Face is a whole unit or row of data
      * Parts of face are adjusted based on data they represent
      * Wide range of disadvantages such as subjectivity and complexity. They are also inefficient and prone to being misinterpreted. Lack of precision
  + Parallel Coordinates
    - Each line in this plot represents row in dataset
    - Each access scaled to min and max of each data attribute
    - Messy
    - Patterns are order dependant
    - Better if interactive
  + Brushing and linking (scatterplots)
  + Use dimension reduction technique
    - PCA (Principle Component Analysis:
      * Center data to mean
      * Build covariance matrix
      * Find eigenvalues and eigen vectors
      * Plot data along with eigenvectors
* Maps: Scaled down accurate version of physical world
  + Bottom layer is basemap showing geographical boundaries then add data layers ontop
* Projections:
  + Mercator
  + Gall-Peters
* You use latitude and longitude to map points
* GeoJSON: open standard format for representing simple geographical features using JSON
* TopoJSON is an extension of GeoJSON which exploits redundancy in geometry
* Useful encodings for spatial dimensions:
  + Size
  + Shape
  + Luminance
  + Hue
* Useful Graphs :
  + Dot maps(point mark) x encodes longitude while y encodes latitude
  + Line Maps (mark line) usually have encodings for size and colour
  + Choropleth maps(area mark): Used to shaw regional variation. Issue where data isn’t necessarily clear
  + Cartogram (Area mark): Cartograms are map based visualisations which distort geography of map to display data attributes
  + Tile grid maps: Constant area to represent different regions. Colour represents data attribute