

# Introduction to Data Journalism



# Learning Outcomes

By the end of the lesson you should be able to

- Describe the meaning is Data Journalism
- Define the meaning of 'Visualisation'
- Recall the visualisation process and the eight visual variables
- Identify the importance of human perception
- Identify good visual representation

# What is Data Journalism?

- **Data journalism** or **data-driven journalism (DDJ)** is a journalistic process based on analysing and filtering large data sets for the purpose of creating or elevating a news story.
- It reflects the increased role that numerical data is used in the production and the distribution of information, and the increased interaction between journalists and fields such as design, computer science, and statistics.
- The use of data journalism helps tell a complex story through the use of infographics and data visualisations.

# What is Data Journalism?

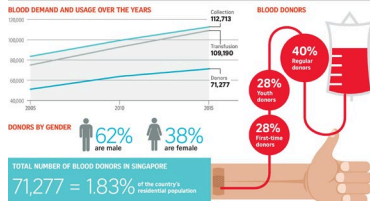
■ The areas covered by data journalism include:

- ▶ Computer-assisted reporting and data-driven journalism (journalists make use of large databases to produce stories)
- ▶ Infographics
- ▶ Data visualisation
- ▶ Interactive visualisation
- ▶ Database journalism (pieces of information are organised in a database)

# Example of Data Journalism

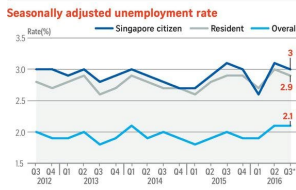
## Getting blood where it's needed

The Health Sciences Authority (HSA) aims to supply the safest possible blood to all patients who need blood transfusion.

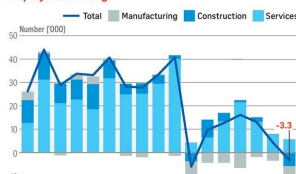


## Labour market at a glance

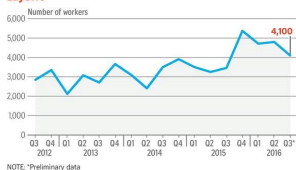
Latest estimates paint a gloomy picture of the labour market in the third quarter.



## Employment change



## Layoffs

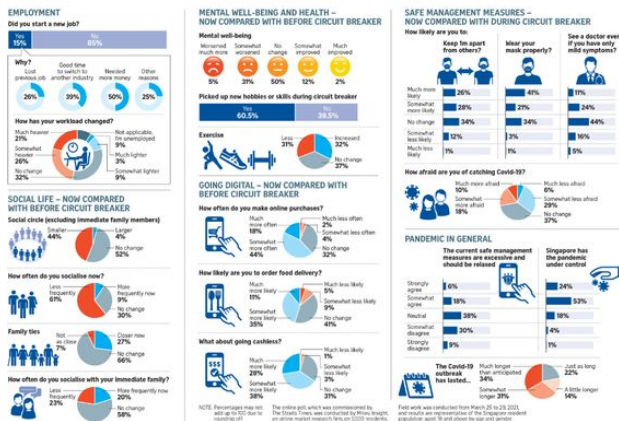


NOTE: \*Preliminary data

Source: MINISTRY OF MANPOWER  
STRAITS TIMES GRAPHICS

## How life has changed

On April 7 2020, much of Singapore came to a standstill. One year on, The Straits Times partnered market research firm Midas Insight to ask Singapore residents how the lockdown has changed their lives. **Timothy Goh** looks at the findings.

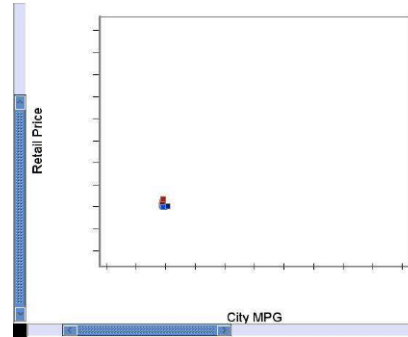


- <https://www.straitstimes.com/singapore/blood-donations-rise-to-match-surge-in-demand>
- <https://www.straitstimes.com/singapore/manpower/job-outlook-remains-gloomy-analysts>
- <https://www.straitstimes.com/singapore/health/one-year-after-circuit-breaker-people-in-spore-socialising-less-working-more-mental>

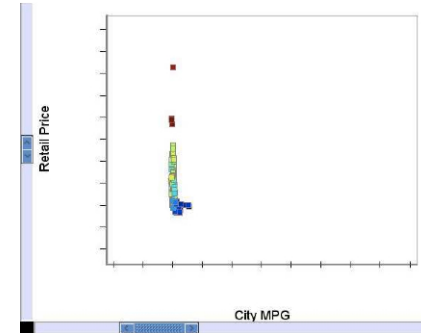
The same data *plotted with*  
different scales is perceived  
dramatically differently:

Exploration

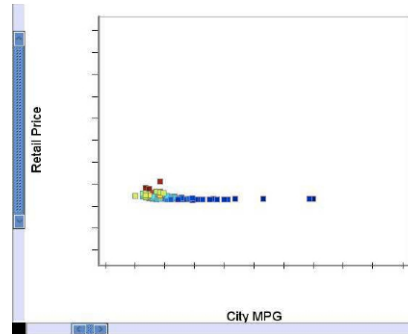
Sense-making



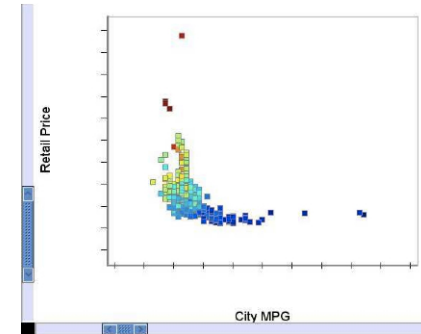
(a) Equally (uniformly) large scale  
in both x and y



(b) Large scale in x



(c) Large scale in y



(d) Scale determined by range of  
x- and y-values.

# What is 'Visualisation'?

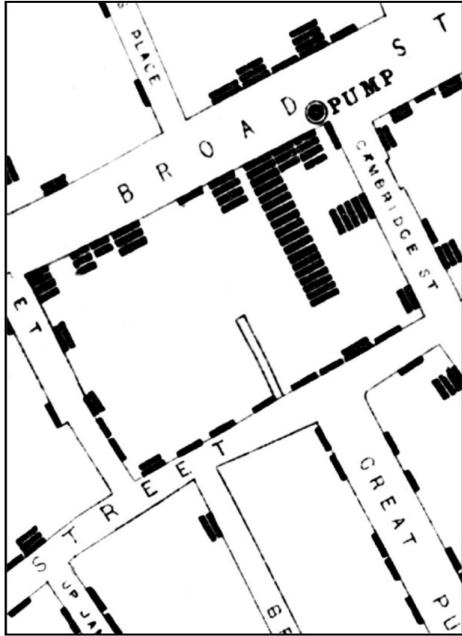
- Visualisation is communication of information using graphical representations.
- The purpose of information visualisation is not to make pictures, but to **help us think**.
- Information Visualisation vs. Scientific Visualisation
  - ▶ Computer-supported, interactive, visual representation
  - ▶ Visual representation of physical data – like X-ray, MRI

## Data Visualization

Activities	Exploration Sense-making	→	Communication
Technologies	Information Visualization Scientific Visualization	→	Graphical Presentation
Immediate Goal	Understanding		
End Goal	Good Decisions		



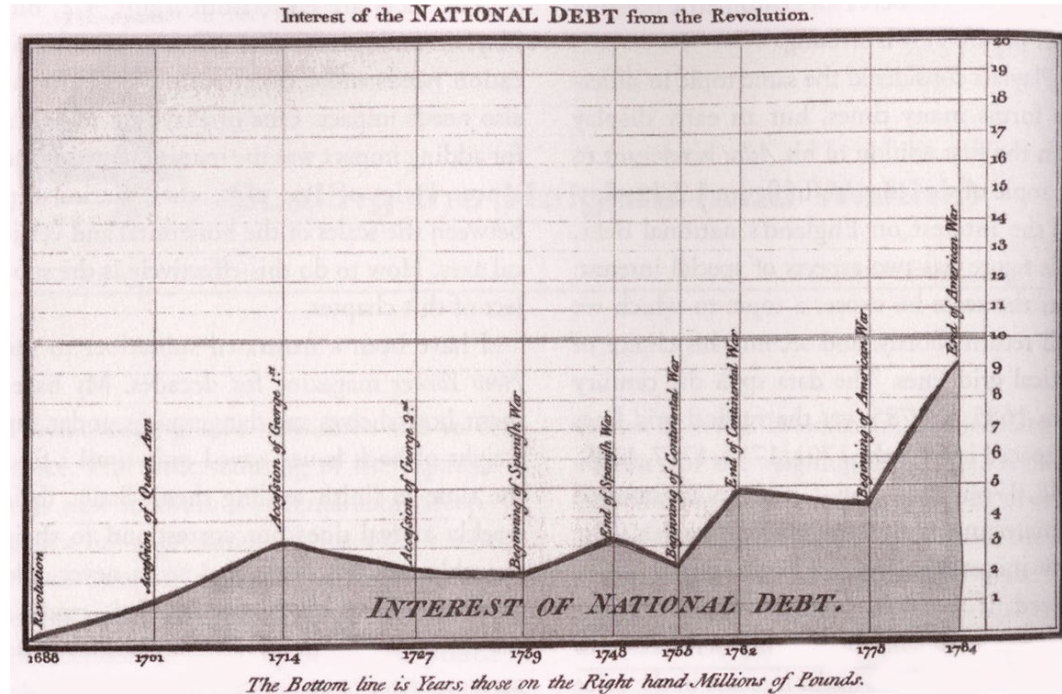
# Early Visualisation



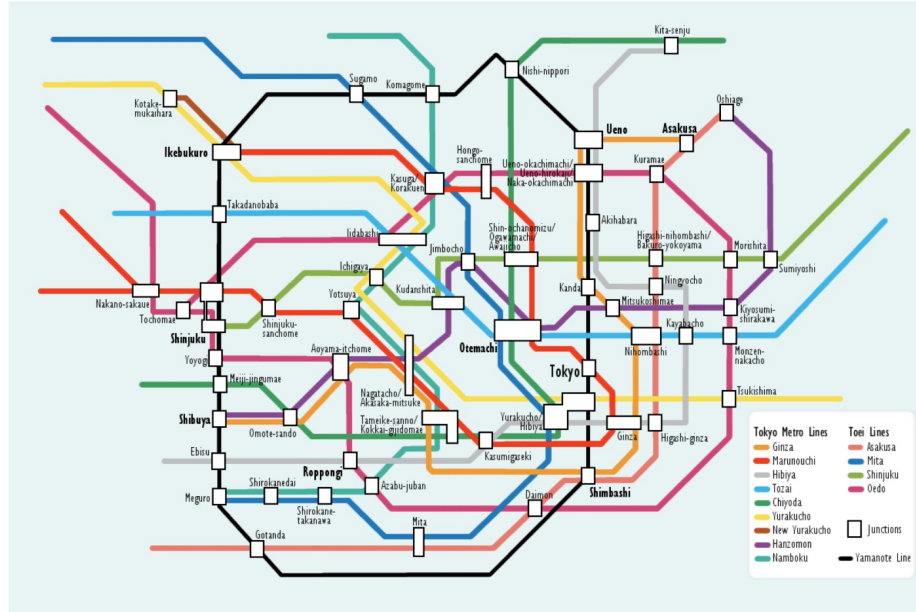
A section of John Snow's map of the deaths from cholera in London in 1663.

Each bar within the houses represents one deceased individual.

# Early Visualisation

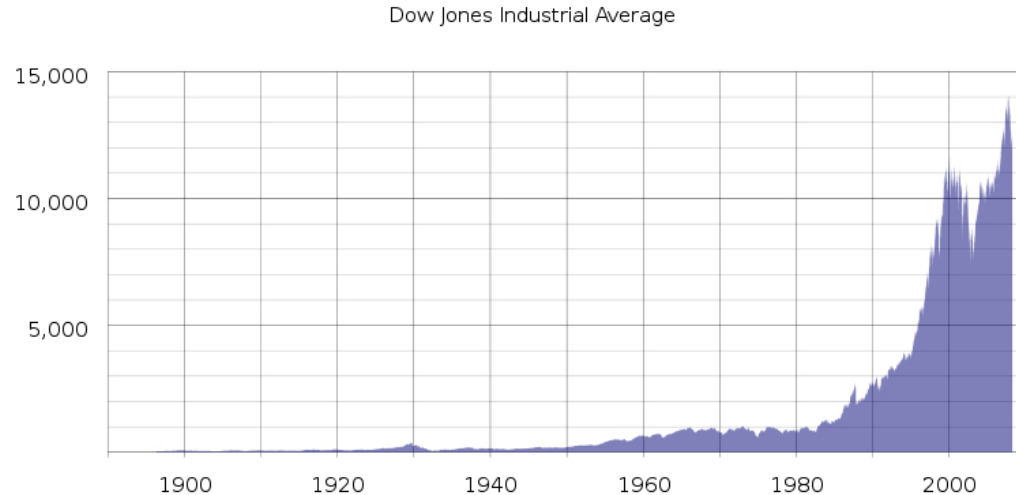


# Visualisation Today



The Tokyo Underground map. A logical representation of the metro highlighting qualitative relationships between the stops.

# Visualisation Today



Dow Jones Industrial Average (DJIA) from 1900 to 2000. The Dow Jones Industrial Average is a U.S. stock index based on the weighted average of the stock prices of 30 large and actively traded U.S. companies.

# Why data visualisation?

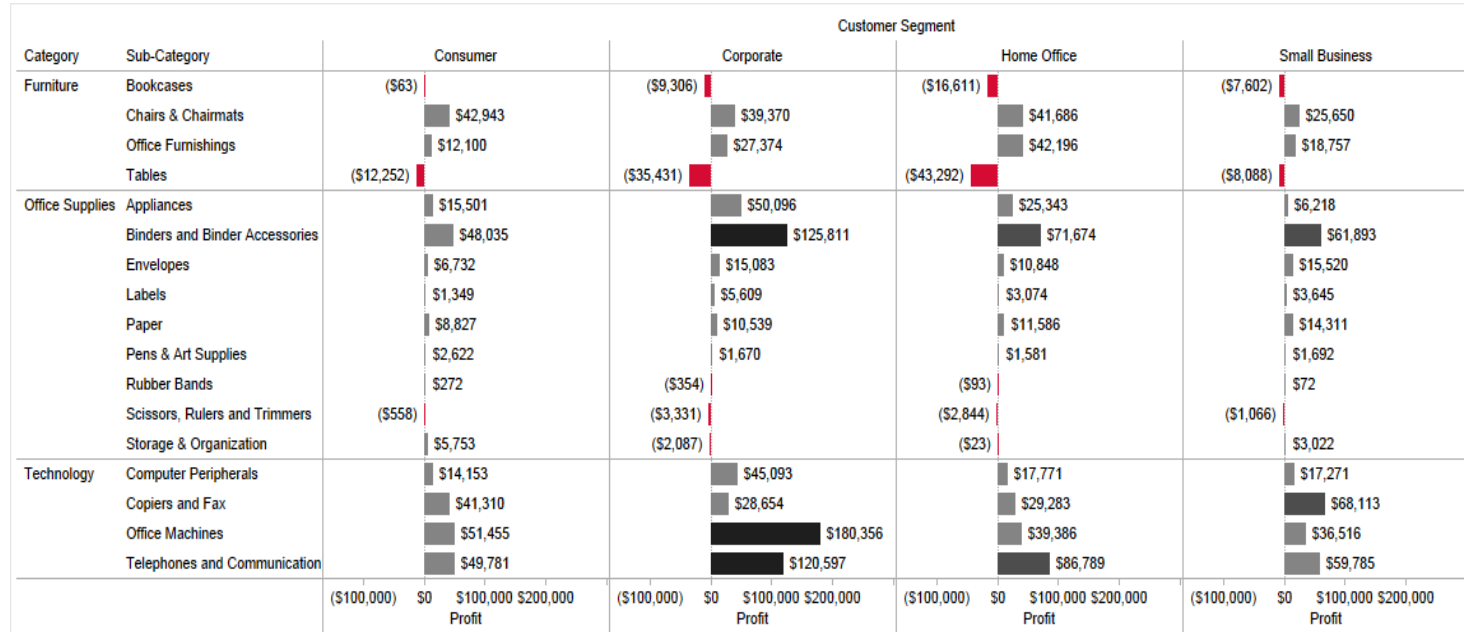
Spend 10 seconds with the following data

Category	Sub-Category (group)	Customer Segment			
		Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-63.02	-9,305.76	-16,610.95	-7,602.40
	Chairs & Chairmats	42,942.97	39,370.10	41,686.28	25,650.38
	Office Furnishings	12,099.80	27,374.47	42,196.25	18,757.40
	Tables	-12,251.51	-35,430.73	-43,292.40	-8,087.89
Office Supplies	Appliances	15,501.48	50,095.94	25,343.06	6,217.58
	Binders and Binder Ac..	48,035.27	125,811.27	71,674.19	61,892.69
	Envelopes, Labels, Pa..	16,907.52	31,230.67	25,508.13	33,476.65
	Pens & Art Supplies	2,621.68	1,670.40	1,580.82	1,691.88
	Rubber Bands	271.85	-353.54	-93.12	72.14
	Scissors, Rulers and ..	-558.10	-3,330.62	-2,844.06	-1,066.47
	Storage & Organization	5,752.65	-2,086.83	-23.24	3,021.57
Technology	Computer Peripherals	14,152.79	45,092.93	17,771.05	17,270.71
	Copiers and Fax	41,310.35	28,654.48	29,283.14	68,113.50
	Office Machines	51,454.78	180,356.22	39,386.23	36,515.70
	Telephones and Com..	49,781.48	120,596.92	86,788.72	59,784.52

# We're Faster When We Can "See" Data

Category	Sub-Category	Customer Segment			
		Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-63.02	-9,305.76	-16,610.95	-7,602.40
	Chairs & Chairmats	42,942.97	39,370.10	41,686.28	25,650.38
	Office Furnishings	12,099.80	27,374.47	42,196.25	18,757.40
	Tables	-12,251.51	-35,430.73	-43,292.40	-8,087.89
Office Supplies	Appliances	15,501.48	50,095.94	25,343.06	6,217.58
	Binders and Binder Ac..	48,035.27	125,811.27	71,674.19	61,892.69
	Envelopes	6,731.55	15,082.58	10,848.34	15,520.13
	Labels	1,349.23	5,608.87	3,073.87	3,645.20
	Paper	8,826.74	10,539.22	11,585.92	14,311.32
	Pens & Art Supplies	2,621.68	1,670.40	1,580.82	1,691.88
	Rubber Bands	271.85	-353.54	-93.12	72.14
	Scissors, Rulers and ..	-558.10	-3,330.62	-2,844.06	-1,066.47
	Storage & Organization	5,752.65	-2,086.83	-23.24	3,021.57
Technology	Computer Peripherals	14,152.79	45,092.93	17,771.05	17,270.71
	Copiers and Fax	41,310.35	28,654.48	29,283.14	68,113.50
	Office Machines	51,454.78	180,356.22	39,386.23	36,515.70
	Telephones and Com..	49,781.48	120,596.92	86,788.72	59,784.52

# Which category / customer segment is the most profiting for the company?



# How our brains work

- Long term memory c.f. Computer Hard disk
- Working memory c.f. Computer RAM
  - ▷ Very Limited ~ 3 slots
  - ▷ Very Volatile ~ 1-5 minutes

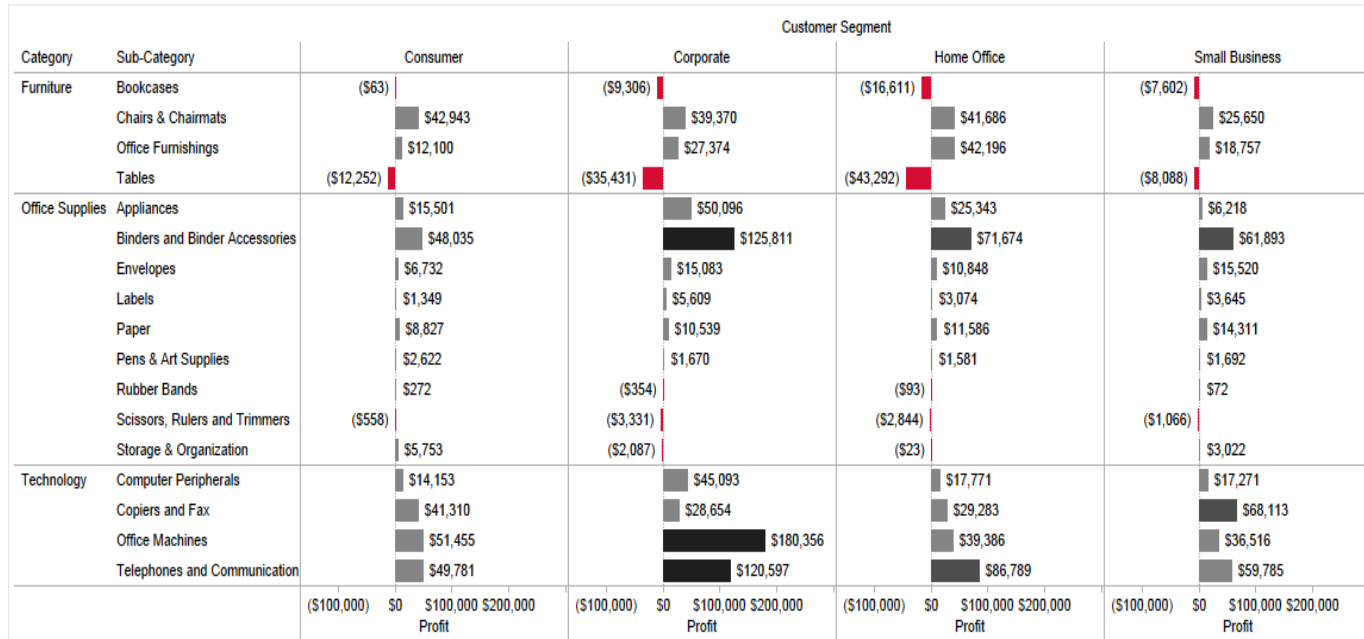




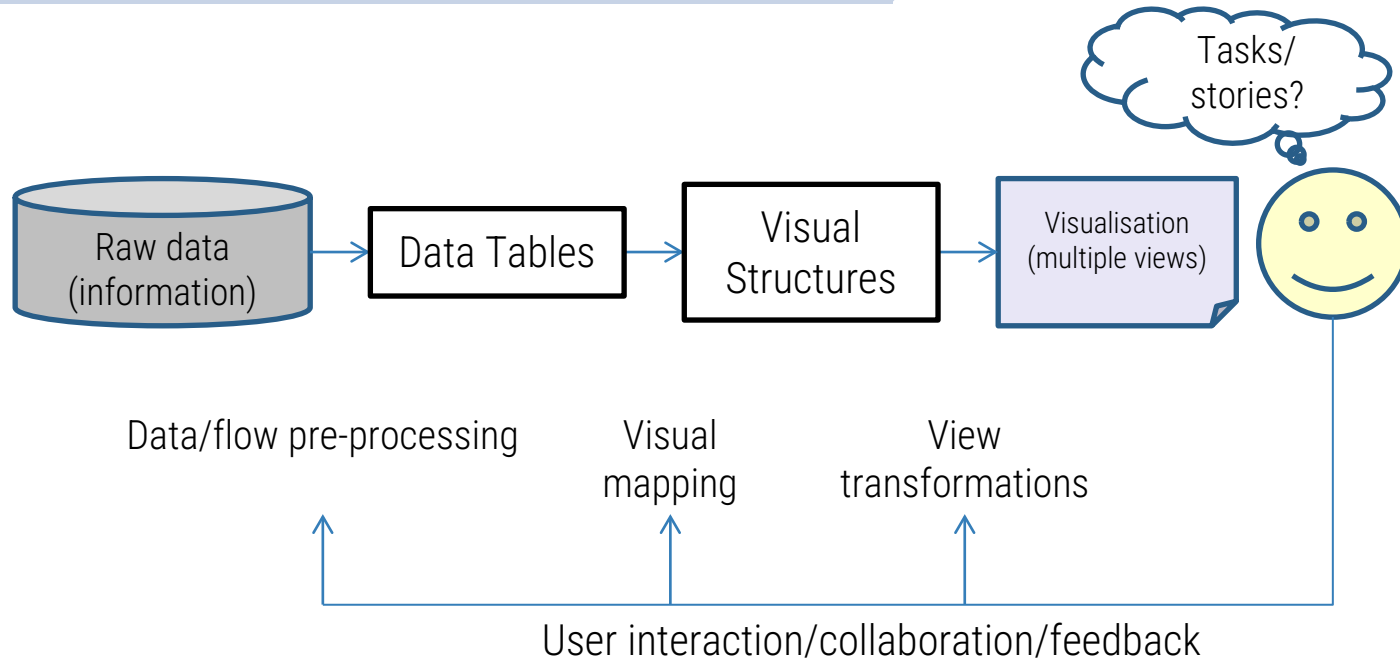
## We need at least 60 slots of working memory to store the following

Category	Sub-Category (group)	Customer Segment			
		Consumer	Corporate	Home Office	Small Business
Furniture	Bookcases	-63.02	-9,305.76	-16,610.95	-7,602.40
	Chairs & Chairmats	42,942.97	39,370.10	41,686.28	25,650.38
	Office Furnishings	12,099.80	27,374.47	42,196.25	18,757.40
	Tables	-12,251.51	-35,430.73	-43,292.40	-8,087.89
Office Supplies	Appliances	15,501.48	50,095.94	25,343.06	6,217.58
	Binders and Binder Ac..	48,035.27	125,811.27	71,674.19	61,892.69
	Envelopes, Labels, Pa..	16,907.52	31,230.67	25,508.13	33,476.65
	Pens & Art Supplies	2,621.68	1,670.40	1,580.82	1,691.88
	Rubber Bands	271.85	-353.54	-93.12	72.14
	Scissors, Rulers and ..	-558.10	-3,330.62	-2,844.06	-1,066.47
	Storage & Organization	5,752.65	-2,086.83	-23.24	3,021.57
Technology	Computer Peripherals	14,152.79	45,092.93	17,771.05	17,270.71
	Copiers and Fax	41,310.35	28,654.48	29,283.14	68,113.50
	Office Machines	51,454.78	180,356.22	39,386.23	36,515.70
	Telephones and Com..	49,781.48	120,596.92	86,788.72	59,784.52

# We need only 2 – 3 slots to differentiate the color and the length of the bars



# Visualisation Process



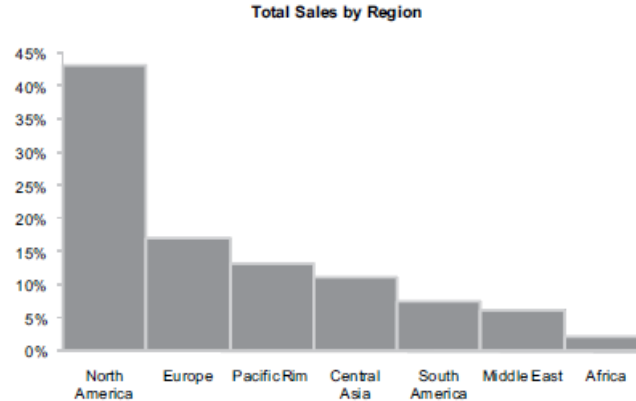
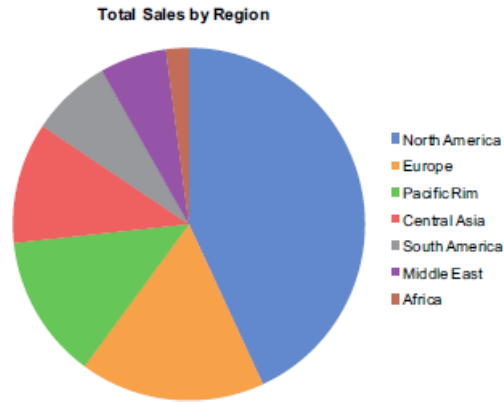
# Data pre-processing/handling

- Data is mapped to fundamental data types
- Specific application data issues – missing values, errors in input, large data
  - ▷ Removal of missing data? Interpolation?
  - ▷ Using different methods to extract relevant data – CSV, JSON, XML
  - ▷ Large data may require sampling, filtering, aggregation

Objective → clean data → meaningful visualisation

# Visual Mapping

Which visual representation to use?



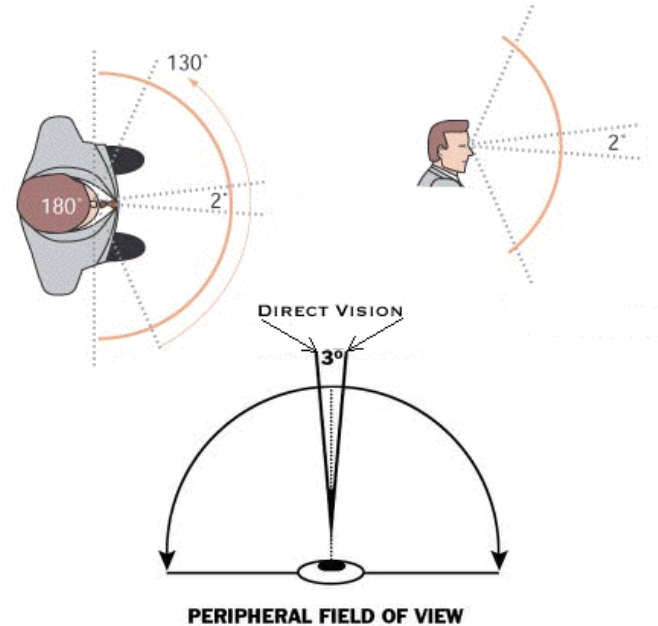
The pie chart doesn't work nearly as well as the bar graph because, to decode it, we must compare the 2-D areas or the angles formed by the slices, but we can easily compare the lengths of bars on the right.

# View Transformation

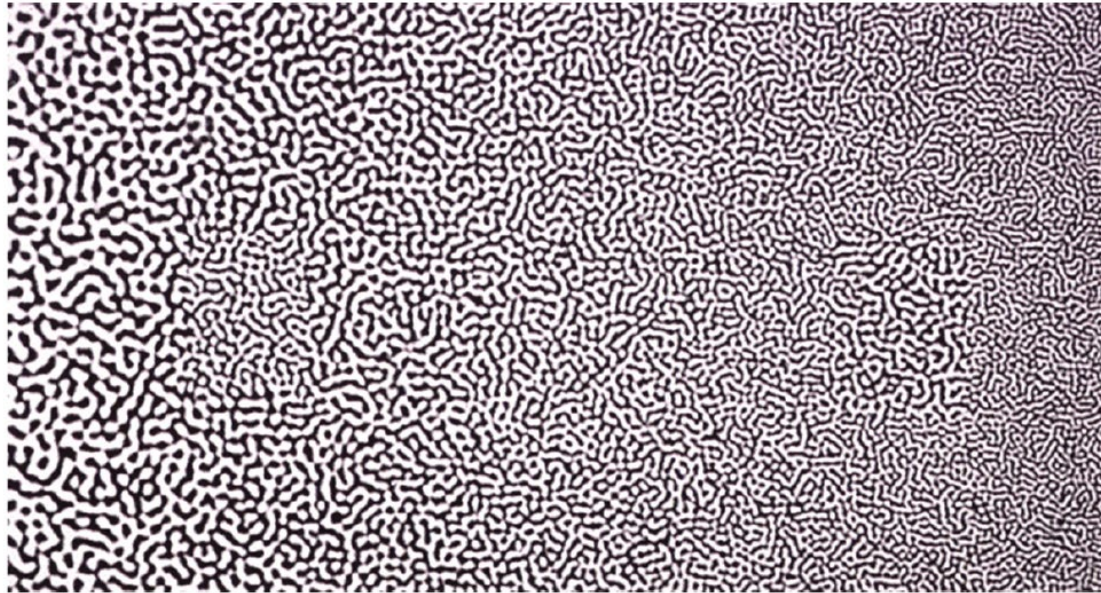
- Mapping of the visual to the final presentation (dashboard, report)
- Measure by expressiveness and effectiveness
- Expressiveness
  - An expressive visualisation presents all the information, and only the information
- Effectiveness
  - A visualisation is effective when it can be interpreted accurately and quickly

# How to make data visualization effective?

## Humans have limited vision



# Facts on Visual Perception





# Facts on Visual Perception



# Facts on Visual Perception



## Can you spot the differences?



## Can you spot the differences?



# Visual Perception

## Pre-Attentive Processing

### ■ Some Visual Properties Processed Pre-Attentively

- ▶ No need to focus attention
- ▶ You will notice whether you want to or not

### ■ Pre-Attentive Properties Important for Design of Visualisations

- ▶ Can be perceived immediately
- ▶ Can mislead viewer

# Pre-attentive Visual Attributes

Group	Attribute		
Form	Length	Width	Orientation
	Size	Shape	Curvature
	Enclosure	Spatial Grouping	Blur
Color	Hue	Intensity	
Spatial Position	2-D Position		
Motion	Direction		

# Visual Perception

## Pre-Attentive Processing

How many 3s ?

08028085080830802809850-802808  
567847298872ty4582020947577200  
21789843890r455790456099272188  
897594797902855892594573979209

# Visual Perception

Pre-Attentive Processing → Colour

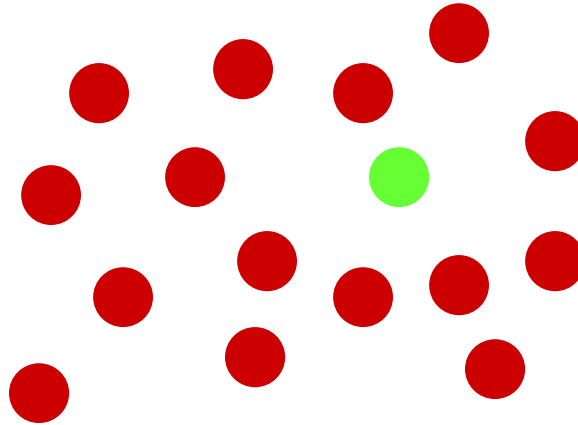
How many 3s ?

08028085080830802809850-802808  
567847298872ty4582020947577200  
21789843890r455790456099272188  
897594797902855892594573979209



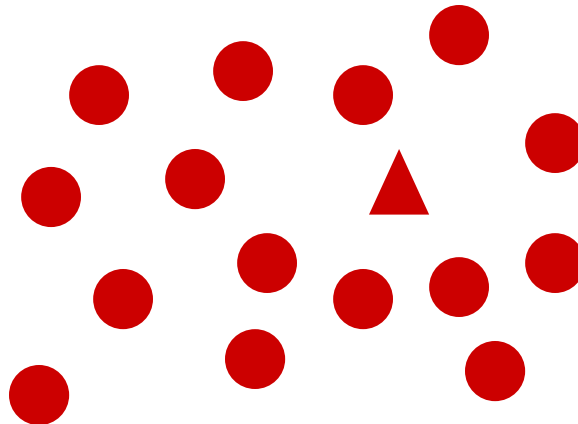
# Visual Perception

Pre-Attentive Processing → Colour



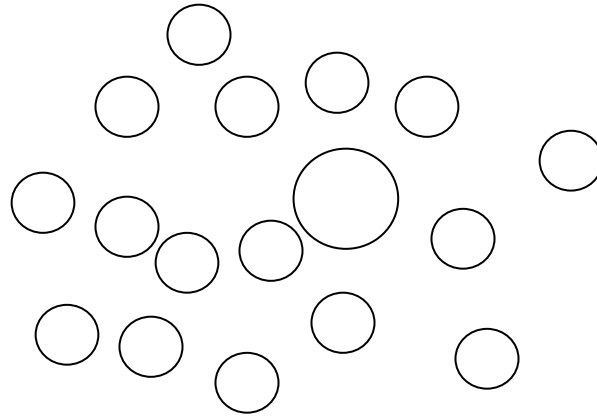
# Visual Perception

Pre-Attentive Processing → Shape



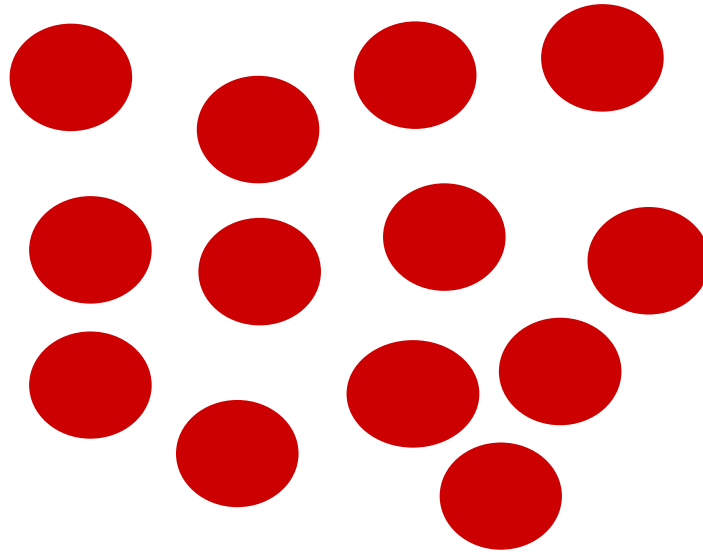
# Visual Perception

Pre-Attentive Processing → Size



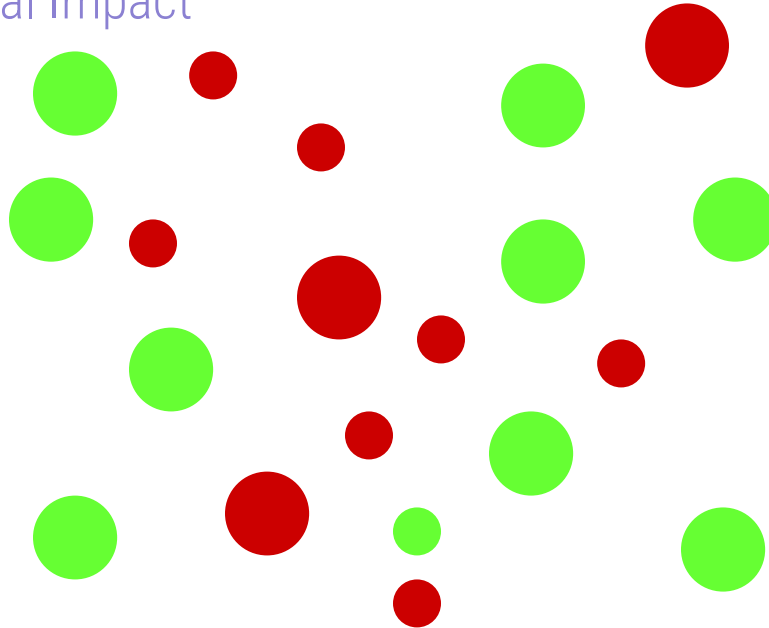
# Visual Perception

Pre-Attentive Processing → Motion



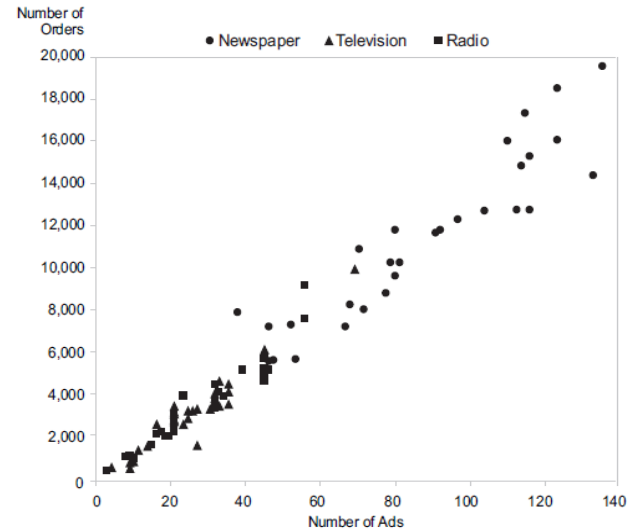
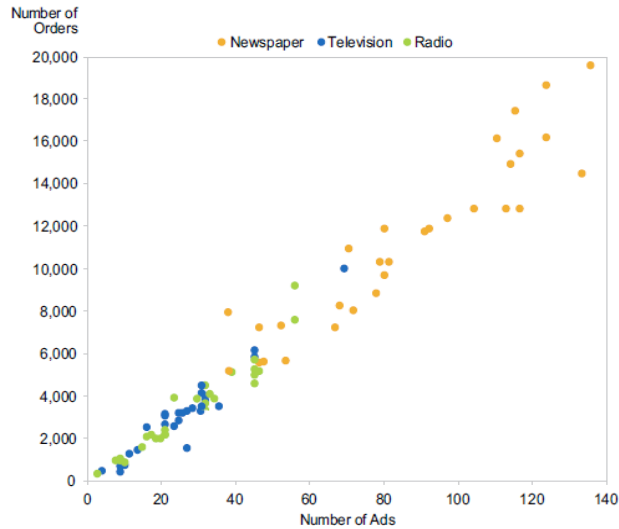
# Visual Perception

## Conjunction – Minimal Impact



# Visual Variables

## Position, Colour, Mark



# Eight Visual Variables

Position

Mark

Size (Length, Area and Volume)

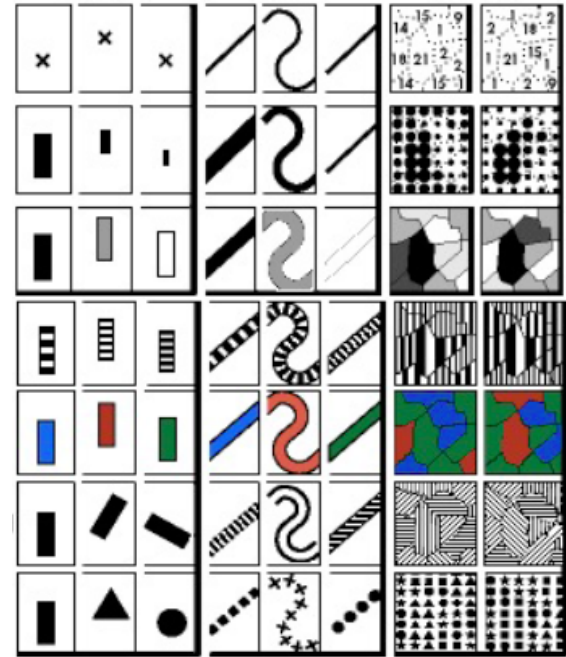
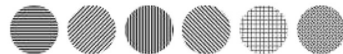
Brightness

Color

Orientation

Texture

Motion



# Colour map



standard linear gray scale



rainbow



heated



blue to cyan



blue to yellow

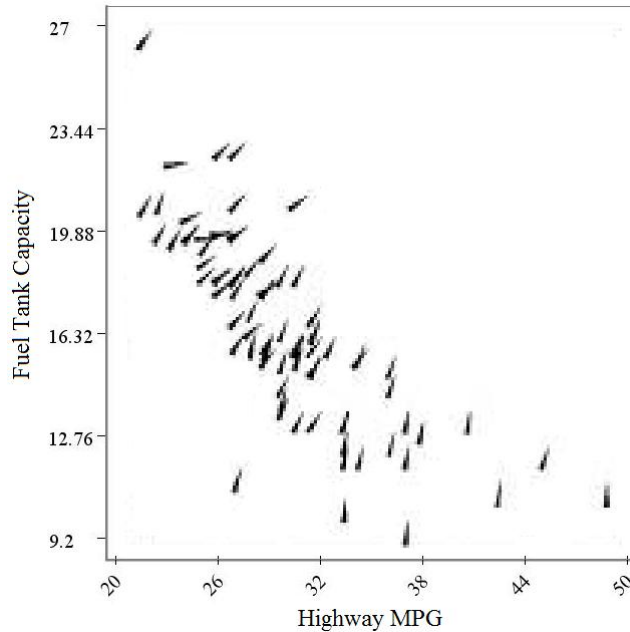


# Orientation









Example orientations of a representation graphic, where the lowest value maps to the mark pointing upward and increasing values rotate the mark in a clockwise rotation.

# Orientation

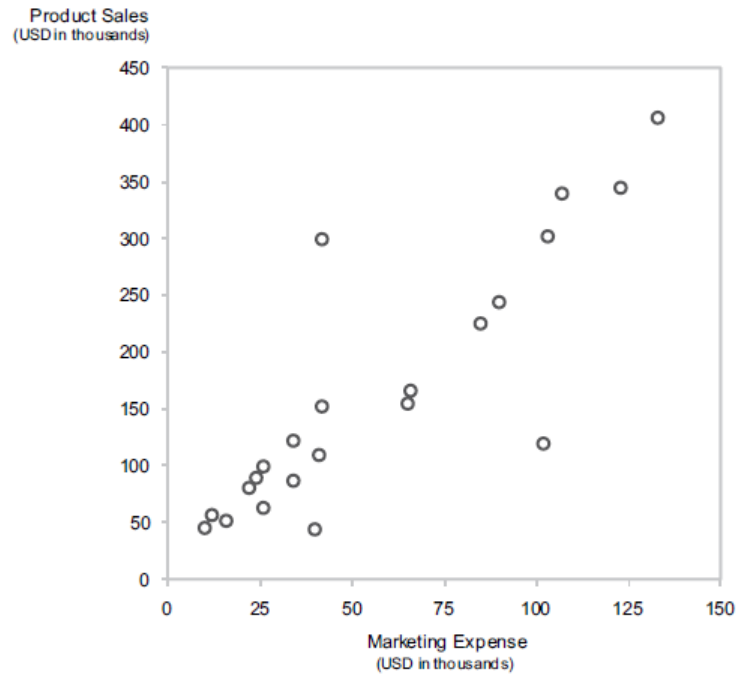


Sample visualization of the 1993 car models data set depicting using highway miles-per-gallon versus fuel tank capacity (position) with the additional data variable, midrange price, used to adjust mark orientation.

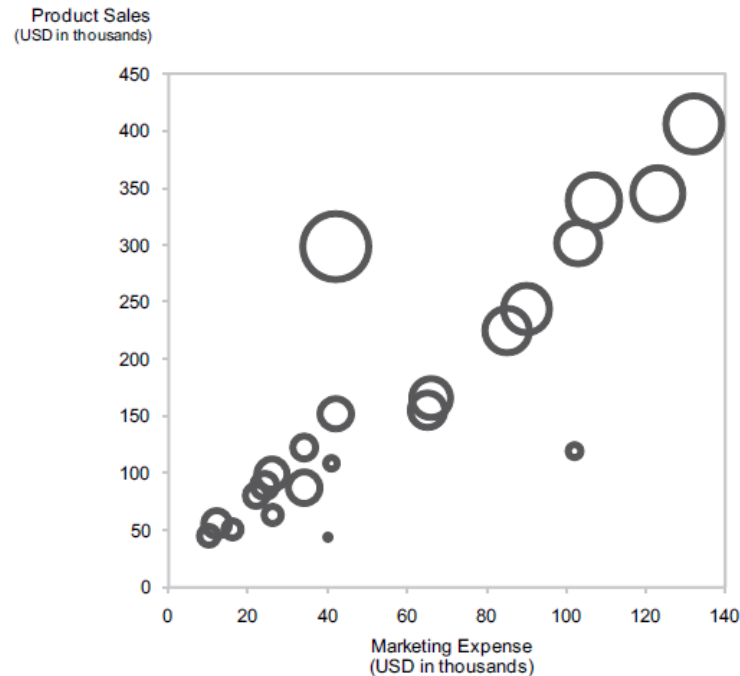
# Precision of Quantitative Perception

Precision of Quantitative Perception	Attribute	Example	Description
Very precise	Length		Longer = greater
	2-D Position		Higher or farther to the right = greater
Not very precise	Width		Wider = greater
	Size		Bigger = greater
	Intensity		Darker = greater
	Blur		Clearer = greater

# Precision of Quantitative Perception



# Precision of Quantitative Perception



# Gestalt Principles of Perception

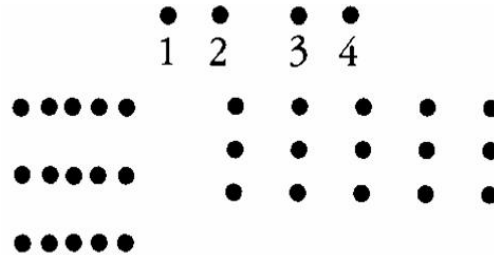
**Closure** - Closure occurs when an object is incomplete or a space is not completely enclosed. If enough of the shape is indicated, people perceive the whole by filling in the missing information.



Although the panda above is not complete, enough is present for the eye to complete the shape.

# Gestalt Principles of Perception

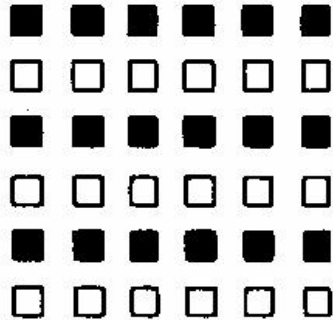
**Proximity** - When elements are placed close together, they tend to be perceived as a group



1 + 2 = as one group  
3 + 4 = as another group

# Gestalt Principles of Perception

**Similarity** - When objects look similar to one another., they are often perceived as a group or pattern.

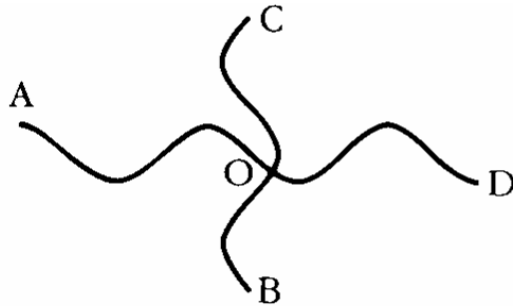


Similarity means there is a tendency to see groups which have the same characteristics so in this example, there are three groups of black squares and three groups of white squares arranged in lines.



# Gestalt Principles of Perception

**Continuity** - Continuation occurs when the eye is compelled to move through one object and continue to another object.



The principle of continuity predicts the preference for continuous figures. We perceive the figure as two crossed lines instead of 4 lines meeting at the center.

# Gestalt Principles of Perception

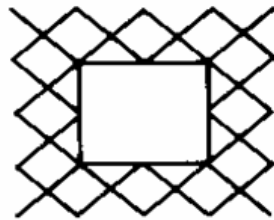
**Figure and Ground** - The eye differentiates an object from its surrounding area. A shape is naturally perceived as **figure** (object), while the surrounding area is perceived as **ground** (background).

Figure

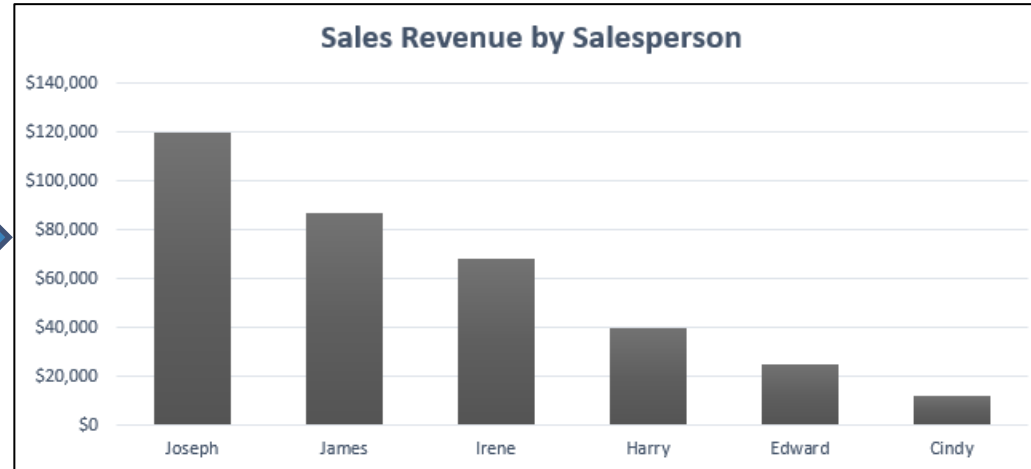
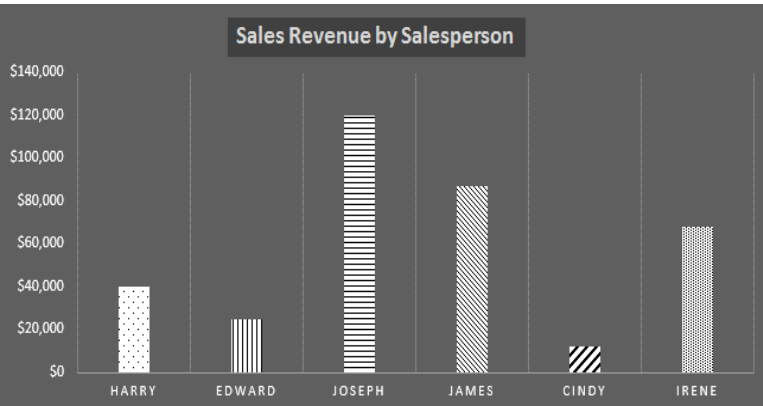
The word above is clearly perceived as figure with the surrounding white space ground.

# Gestalt Principles of Perception

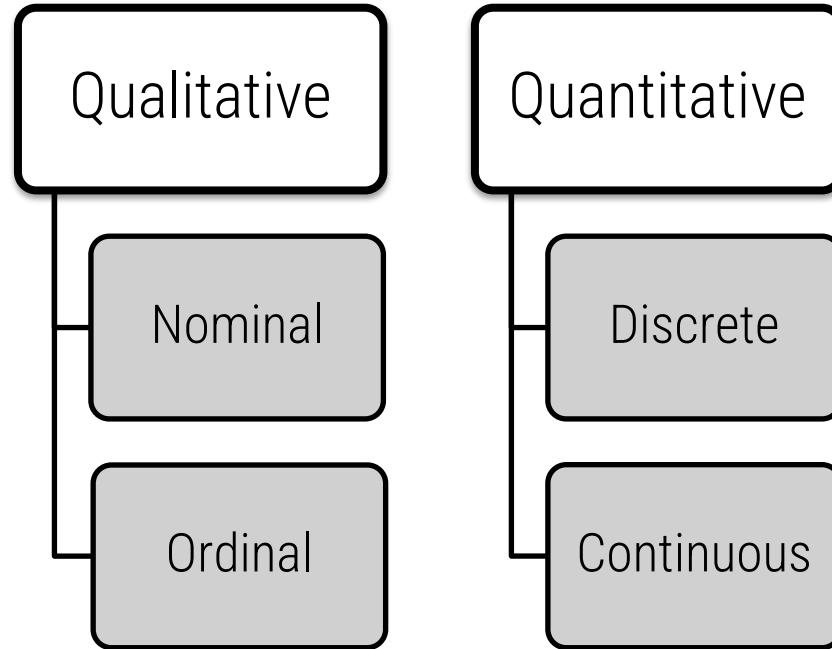
**Area and Symmetry** - The principle of area states that the smaller of two overlapping figures is perceived as figure while the larger is regarded as ground. The principle of the symmetrical figure is that it is seen as a closed figure. Symmetrical contours thus define a figure and isolate it from its ground.













# How can we improve the visualisation?



# Types of Data



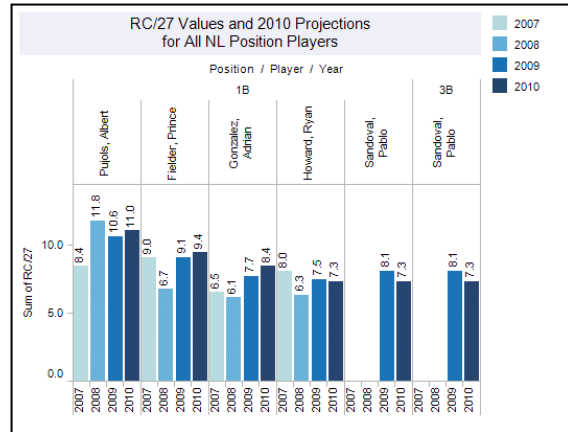
# Visualise Different Types of Data

	Position	Length	Size	Shape	Color
Quantitative					
Ordinal					
Nominal					

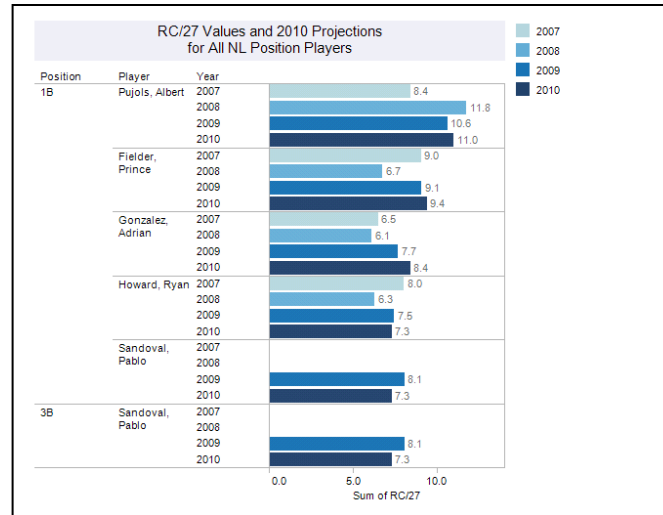
# How Do Humans Like Their Data?

Orientate data so people can read it easily

*Good*



*Better*



# Using Colors to Distinguish Data

+Form

+Qualitative

+Quantitative  
+Ordinal

+Quantitative  
(Diverging)

Gray scale



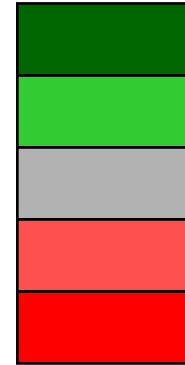
Full spectral  
scale



Single sequence  
single hue scale



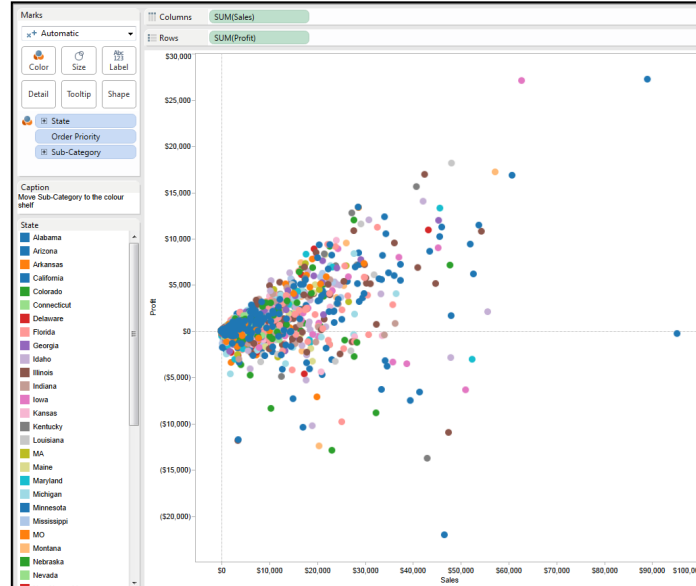
Double-ended  
multiple hue scale





# Using Colors to Distinguish Data

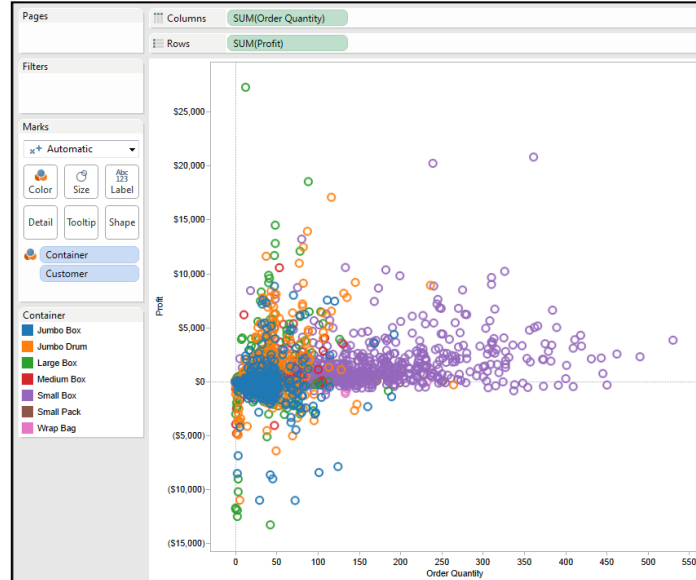
Humans can only distinguish ~8 colors



*This is NOT helpful.*

# Using Colors to Distinguish Data

Humans can only distinguish ~8 colors



*This is helpful.*

# What is good visual representation?

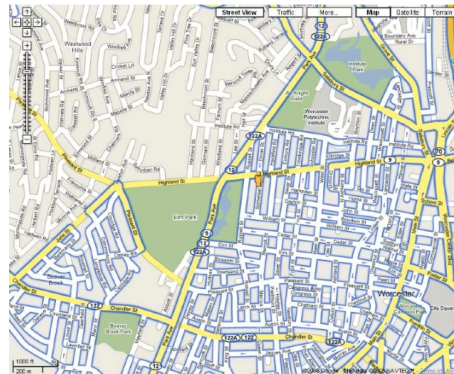
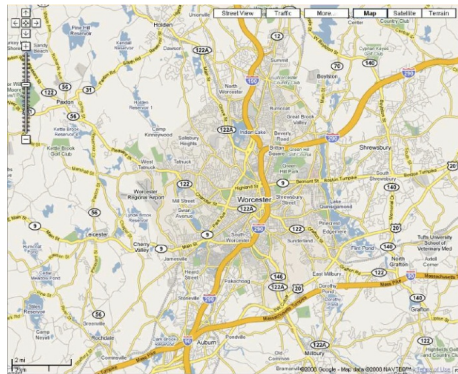
A successful visualisation is one that efficiently and accurately conveys the desired information to the target audience.

- Suitable mapping from data to visualisation
- Ability to select and modify view
- Sufficient information density – not too much or too little
- Importance of keys, labels and legends
- Using colour with care
- Importance of aesthetics

# Mapping Data to Visualisation

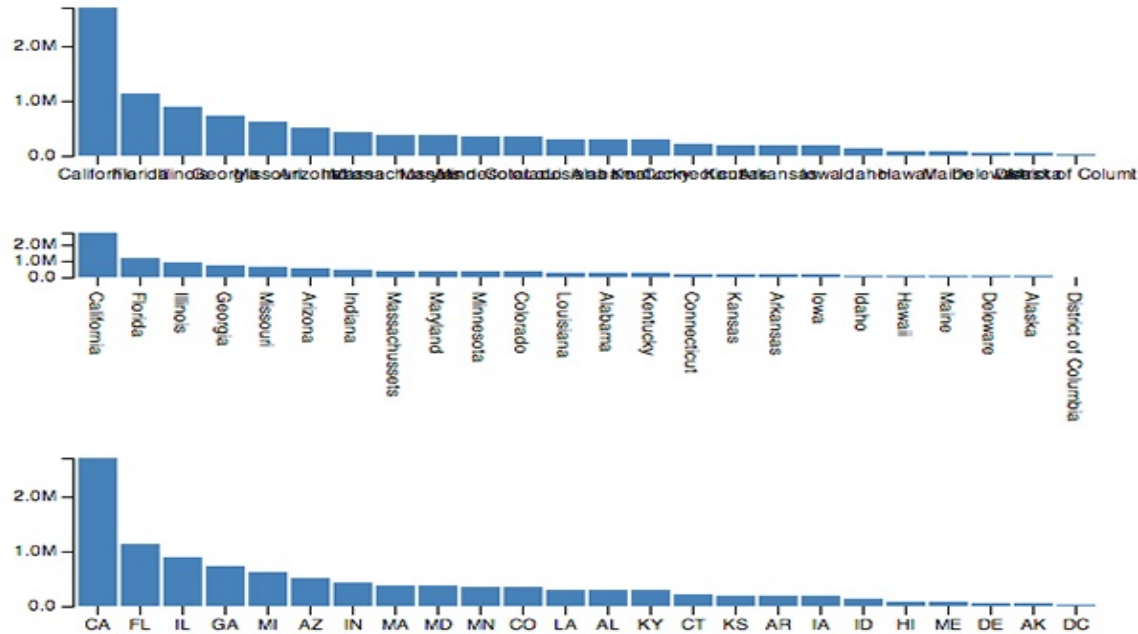
- Mapping **spatial data** (longitude and latitude) to position on **map**
- Mapping based on context – **temperature to color, blood pressure to height**
- Important consideration:
  - Compatibility between scale of data field and the attribute. *For example, ordered data attributes (e.g. age) should not be represented by un-ordered graphical attribute like shape*

# Ability to select and modify views

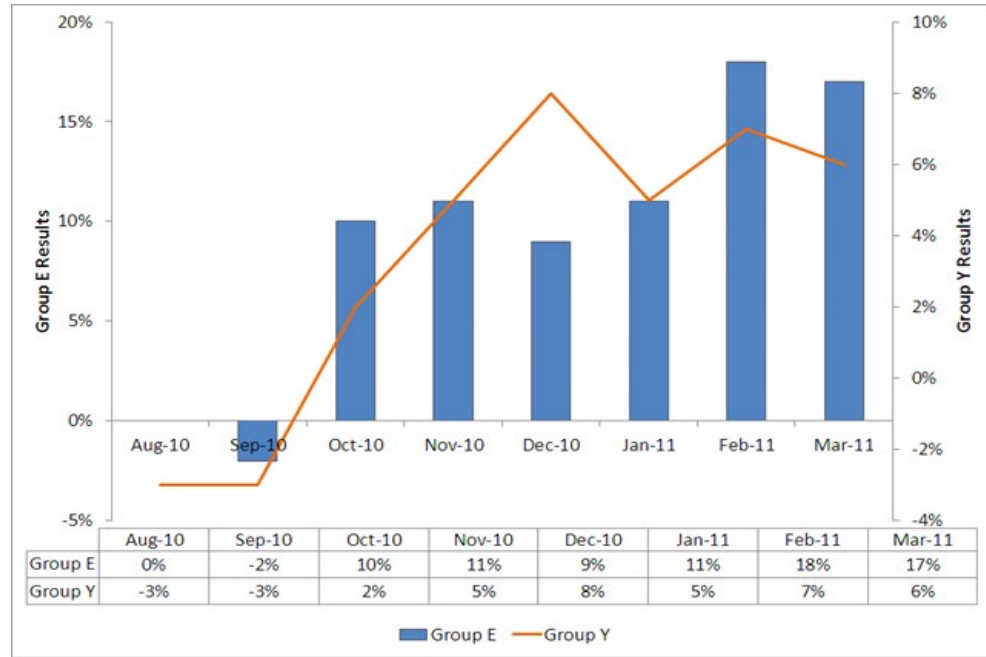


Levels of detail in maps.

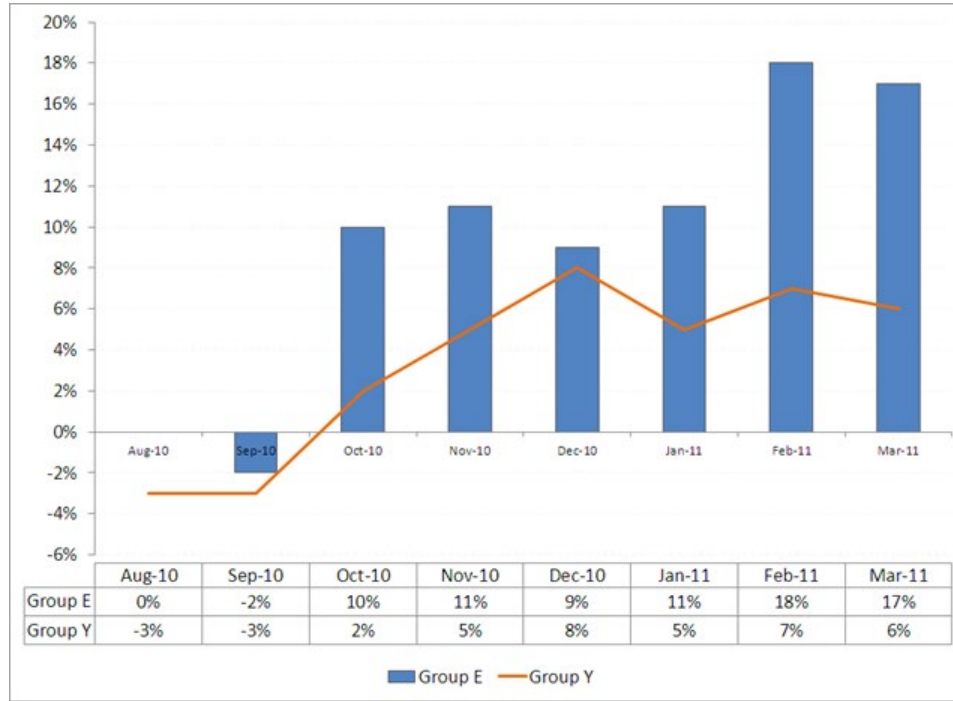
# Importance of keys, labels, and legends



# Importance of keys, labels, and legends

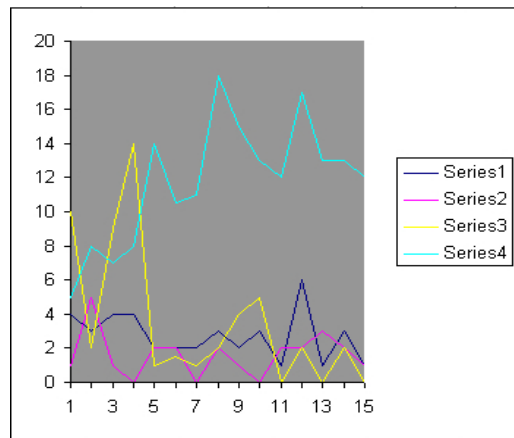
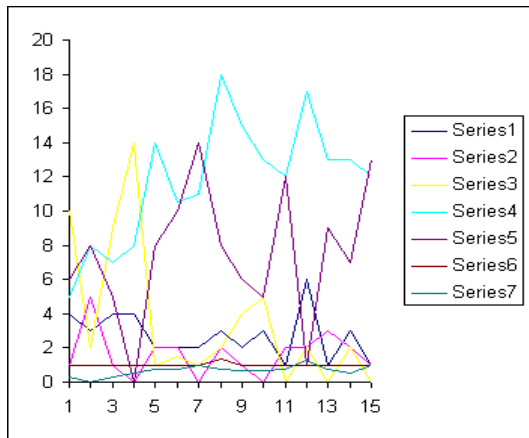


# Importance of keys, labels, and legends



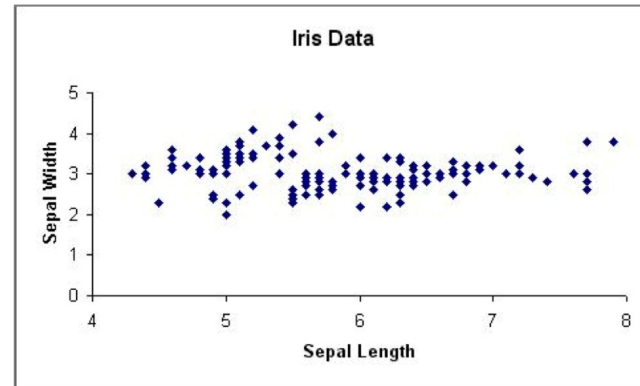
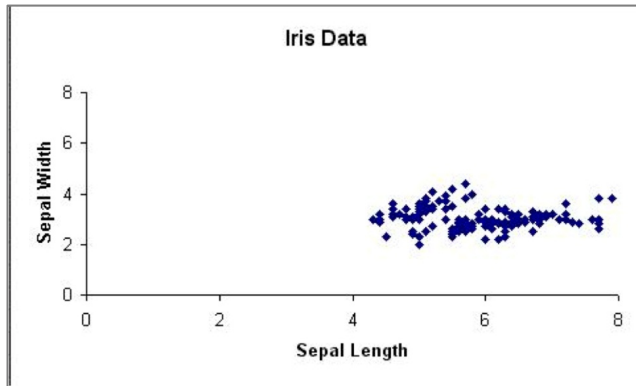


# Using Color with Care



Too many colors versus a moderate number of colors

# The Importance of Aesthetics

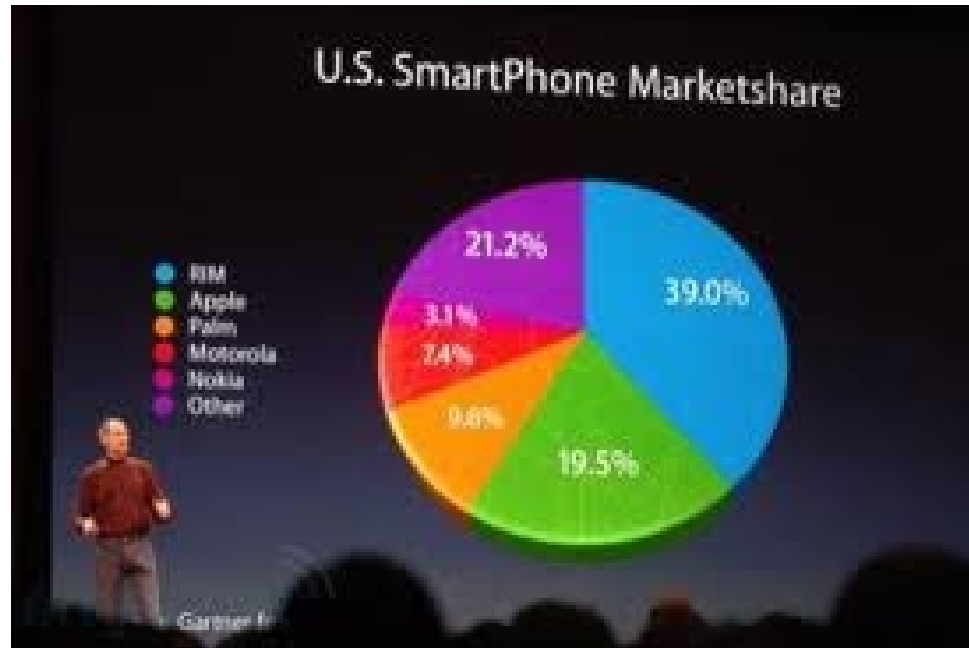


(left) Everything to one side vs. (right) balanced between left and right.

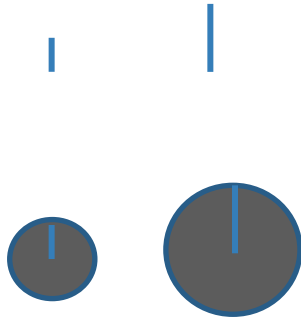
# Tell the truth, no distortion

- “Three Dimensional “effects
- Nonlinear data scaling (lie factor)
- Truncated graph

## “Three Dimensional” effects



# Nonlinear Data Scaling



Data is linear.  
Length is fine  
but not area

## THE ISSUE OF TRUST

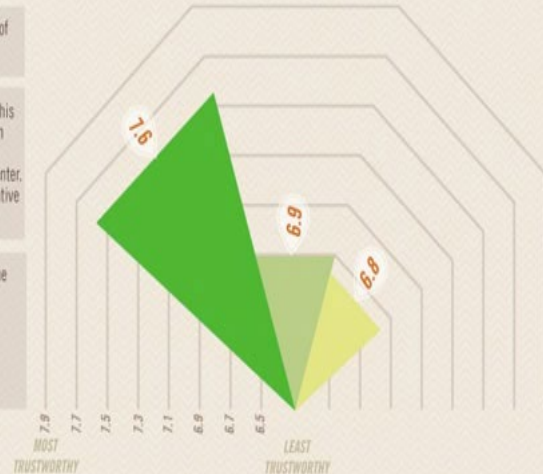
### ACCENTS AND DISTRUST

Another reason why accents affects customer service is the question of credibility. If I can not understand you, then I can not trust you.

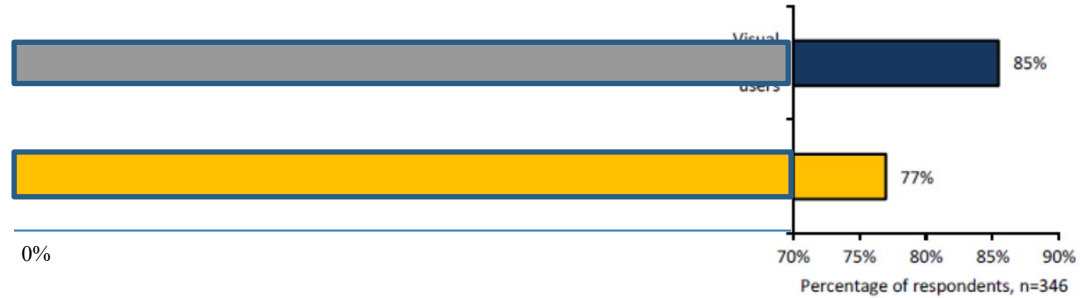
An experiment conducted by the University of Chicago demonstrated this aspect. The question posed, do trivia statements sound less true when spoken by a non-native speaker? Furthermore, listeners were told in advance that all of the trivia questions were provided by the experimenter. This way, even listeners who were knowingly prejudice against non-native accents should not have been affected.

The results showed that the heavier the accent the less trust worthy the person became.

- ▲ NATIVE ACCENT
- ▲ MILD ACCENT
- ▲ HEAVY ACCENT



# Truncated Graph



# Summary

- Pre-Attentive Processing
- Gestalt Principles – how human perceive data
- Techniques and best practices to efficiently and accurately convey the desired information to the target audience.