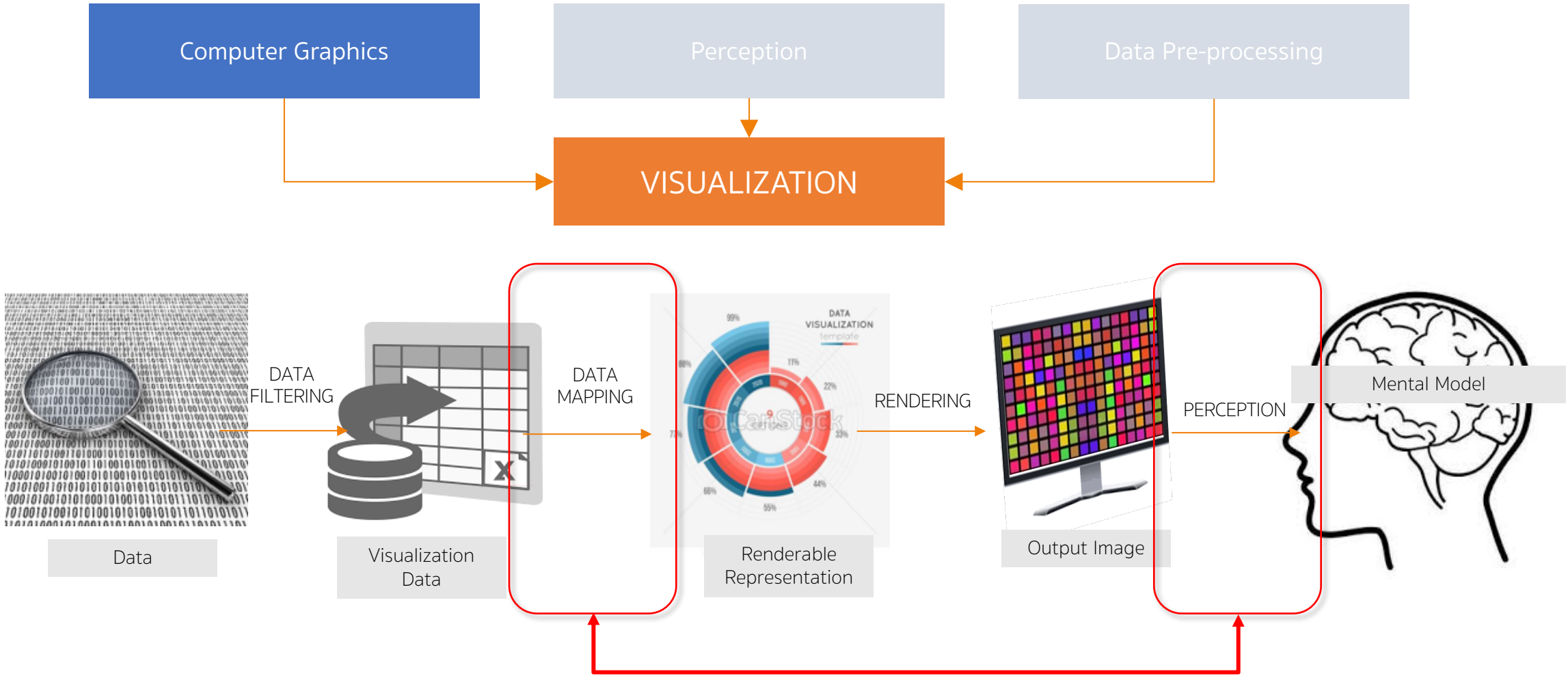
Visual Encoding

29/09/2022

Perceptual Principles for Visualization

Visual Encoding



How we map data to visual representations that will help us understand it

Perception

RECOGNITION

ORGANIZATION

INTERPRETATION

How we interpret signals from the world around us forming a mental representation of the environment

Perception is the process of **recognizing** (being aware of), **organizing** (gathering and storing) and **interpreting** (binding to knowledge) sensory information

Some reasons for studying visual perception

- What kind of visual signals do humans perceive?
- How sensitive are humans to different visual signals?
- How do interpretations differ across different contexts?

Human Vision

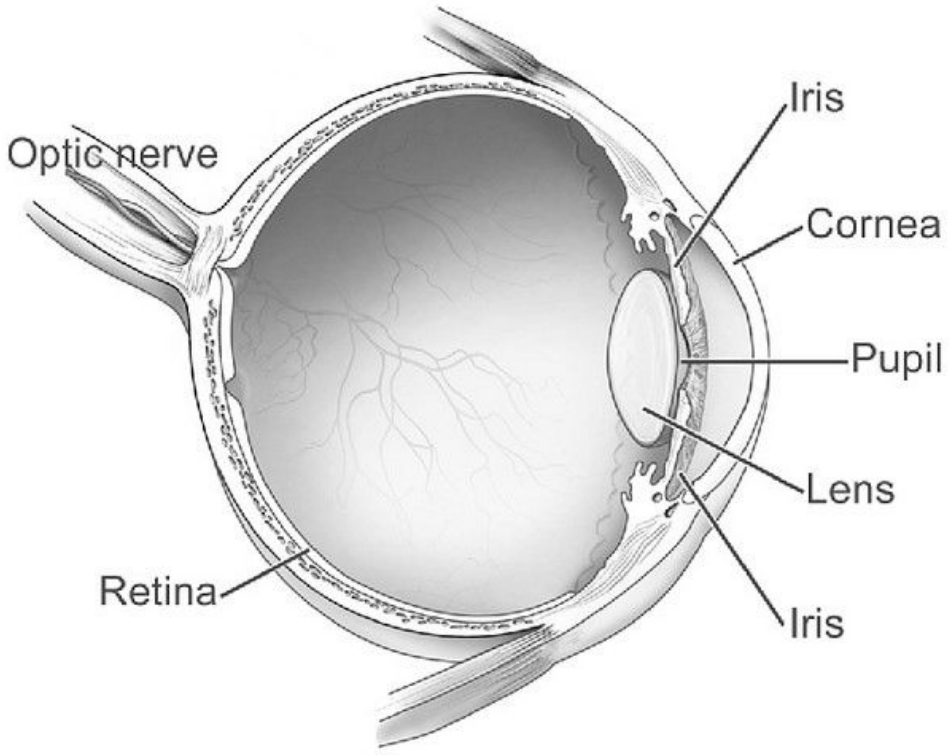


Image © National Eye Institute

The eye receives input in the form of **visible light**, which is focused on layer of photoreceptor (light receiving) cells in the **retina.**

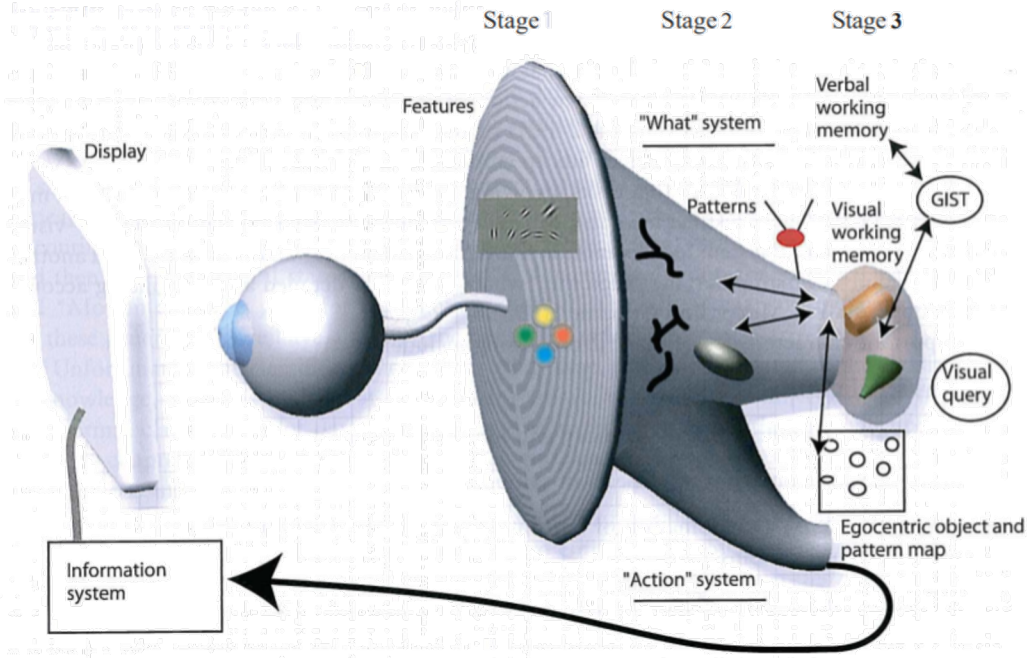
Cone cells respond to **red, green or blue light;** combined to allow **perception of a full range of color**

A network of neurons and ganglions detect the contrasts within an image (such as shadows or edges) and pass this electrochemical signals via the optic nerve, to the brain.

Read more: https://www.interaction-design.org/literature/topics/visual-perception

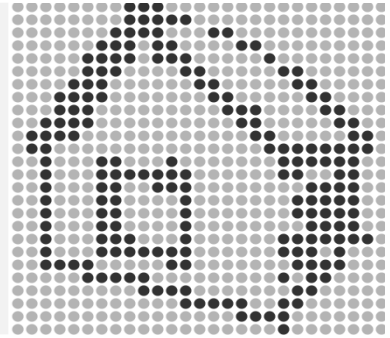
Visual Information Processing

Three-stage Model of Human Visual Information Processing (Ware 2004): How we visually interpret the world around us forming a mental representation of the environment.



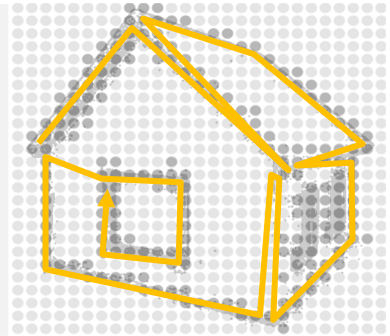
Stage 1: Sensing (Pre-attentive Processing)

- **Bottom-up** feature extraction e.g. edges, orientation, colour, texture, motion
- Involuntary processing of **transitory** visual information by large arrays of neurons in **parallel.**



Stage 2: Organization (Pattern Perception)

- Identify **patterns** e.g. contours, regions, motions
- Slow **serial** processing: with both working and long-term memory involved



Stage 3: Interpretation (Goal-Directed Processing)

- Construct sequence of visual queries, answered through visual search **strategies**.
- Limited objects held in visual working memory through active attention

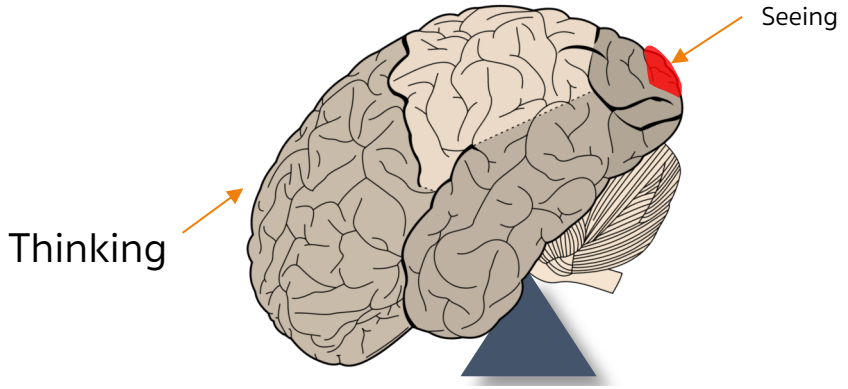
Data Visualization and Human Perception

"**Seeing (i.e., visual perception),** handled by the visual cortex located in the rear of the brain, is extremely fast and efficient. We see immediately, with little effort.

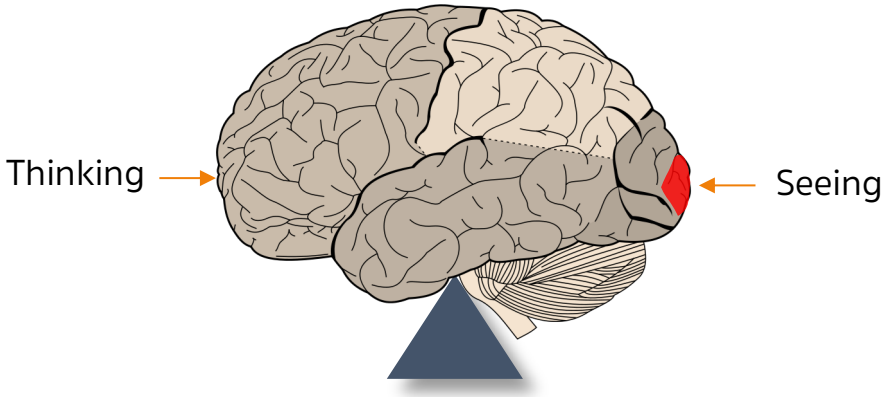
Thinking (i.e., cognition), handled primarily by the cerebral cortex in the front of the brain, is much slower and less efficient. Traditional data sensemaking and presentation methods require conscious thinking for almost all of the work.

Data visualization is effective because it shifts the balance toward greater use of visual perception, taking advantage of our powerful eyes whenever possible."

From the encyclopedia of Human-Computer Interaction (Chapter 35) https://nteraction-design.org



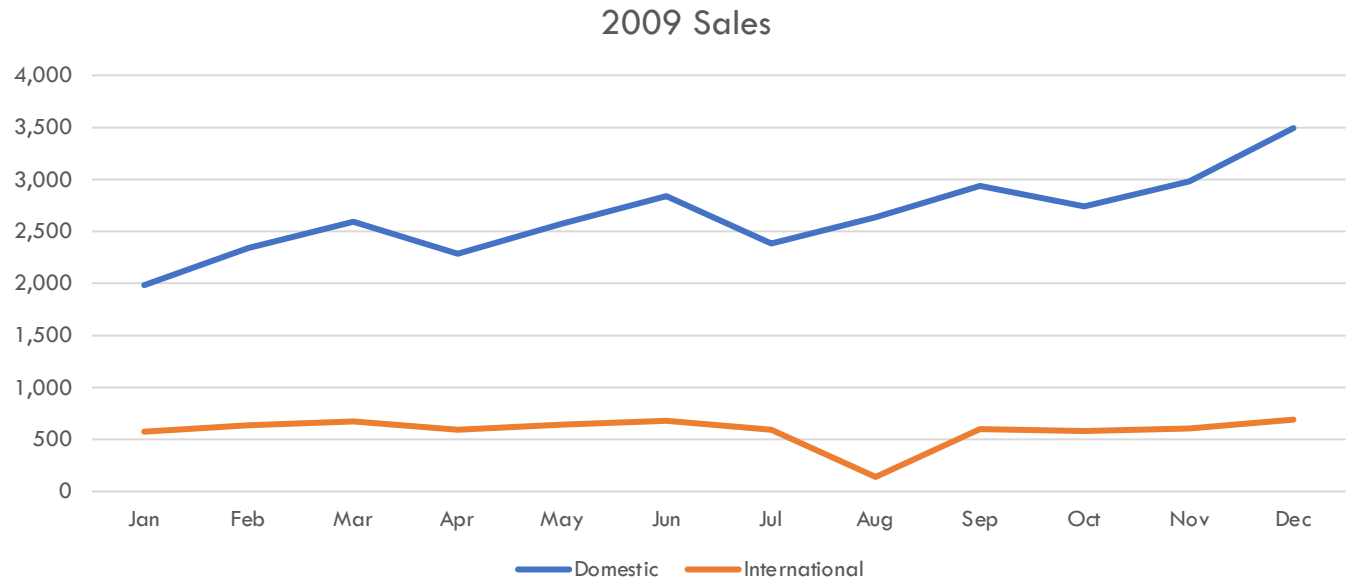
Out-of-balance

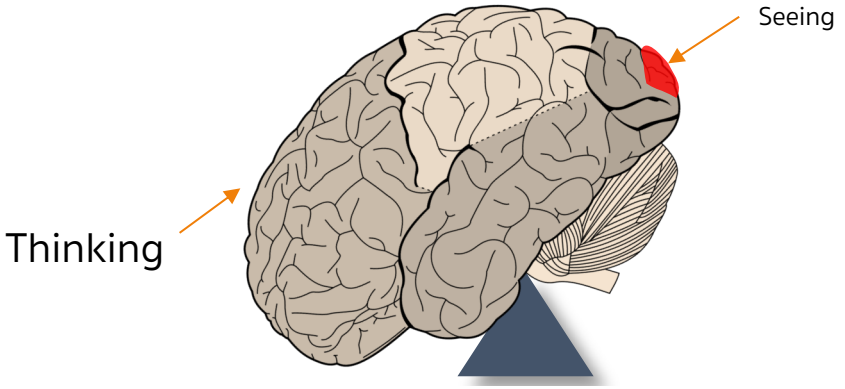


In-balance

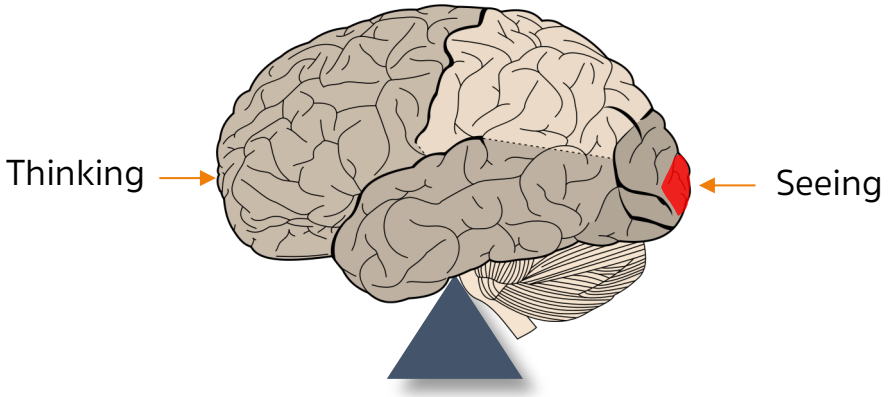
Data Visualization and Human Perception

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493
International	574	636	673	593	644	679	593	139	599	583	602	690
Total	2,557	2,979	3,266	2,876	3,218	3,517	2,975	2,773	3,537	3,322	3,585	4,183



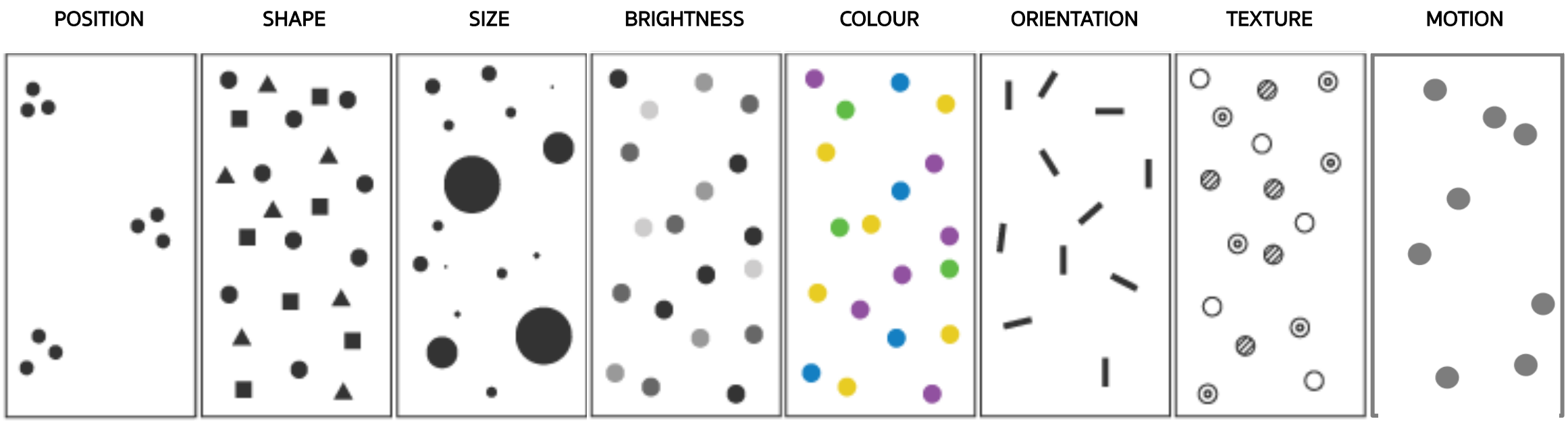


Out-of-balance



In-balance

Recap: Visual Encoding Channels



The Eight Visual Variables (Ward et al 2010)

Graphics primitives used to depict data in visualisation

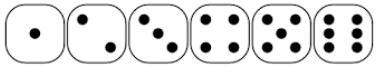
Which one to use?

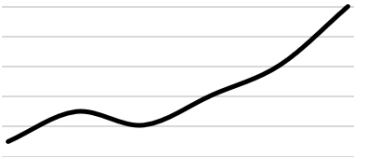
Preview : Common Data Attribute Types

N.B. A more comprehensive discussion of this in provided in future lectures; do not use this page as a reference in assignments

Quantitative

- Discrete : taking only integer values or from a specific subset of possible values e.g. 2, 4, 6, ..
- Continuous : real values e.g. numbers in the interval [0, 5]





Categorical

- Binary : true/false
- **Ordinal** / ranked: values that have an implied ordering, e.g. small, medium, large
- Nominal / Arbitrary : potentially infinite range and no ordering e.g. addresses

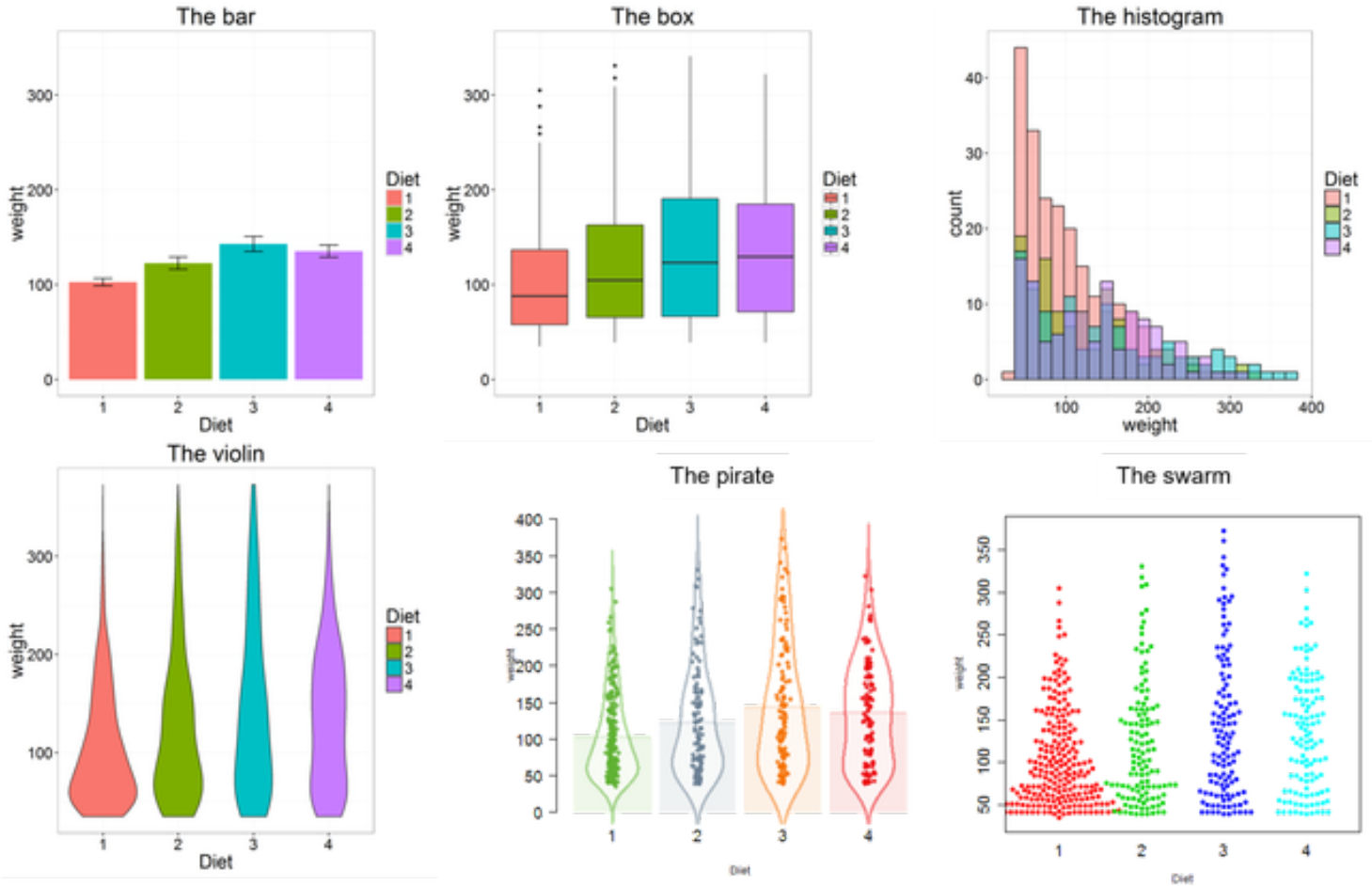
Expressing Data through Visualization

Various alternative encodings can be used to depict data. A particular approach may be more optimal for a specific objective

	Weight	Time	Diet
1	42	0	1
2	51	2	1
3	59	4	1
4	64	6	1
5	76	8	1
6	93	10	1
7	106	12	1
…	…	…	….
575	205	16	4
576	234	18	4
577	264	20	4
578	264	21	4

The R ChickWeight data relating four different diets to chicken weight, plotted six different ways.

Image by Sho Tsuji 2016 [URL]



Multi-Channel Visualizations

One channel may not be sufficient for complex data, or for a high-dimensional task

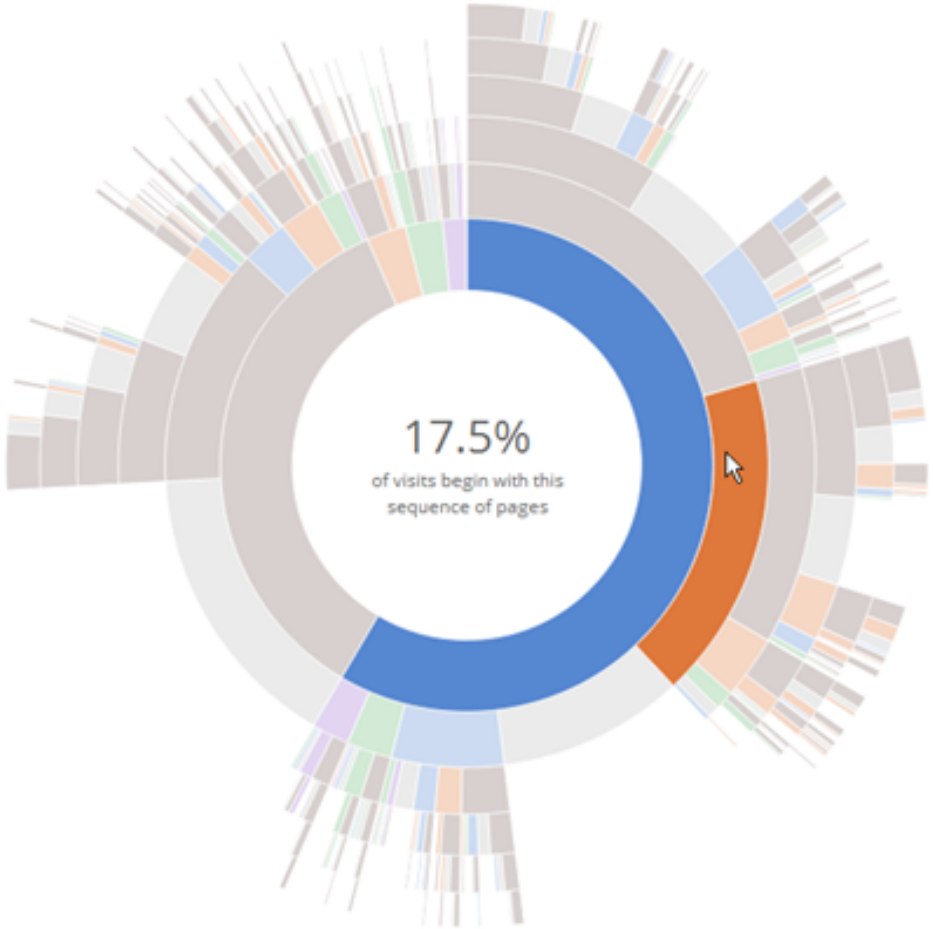
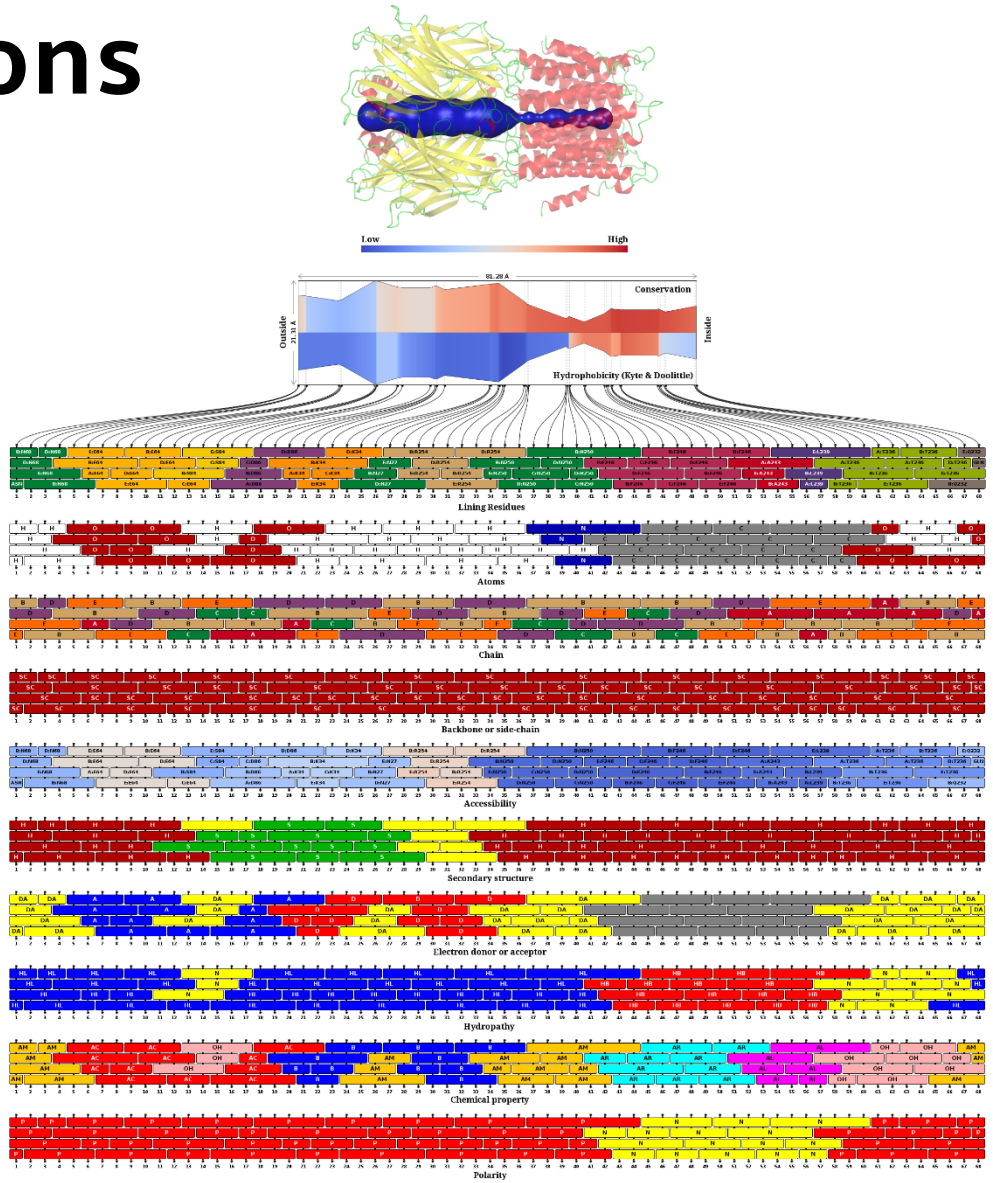


Image © A. Kaushik [URL]



MolecularVisualization in ChexVis. Image © Masood et al. 2015

Some Common Visualization Tasks

N.B. A more comprehensive discussion of tasks is provided in future lectures; do not use this page as a reference in assignments

Identify: recognize object based on characteristics presented

- *e.g. is there a three-bed house for sale?*

Locate: establish position of an object

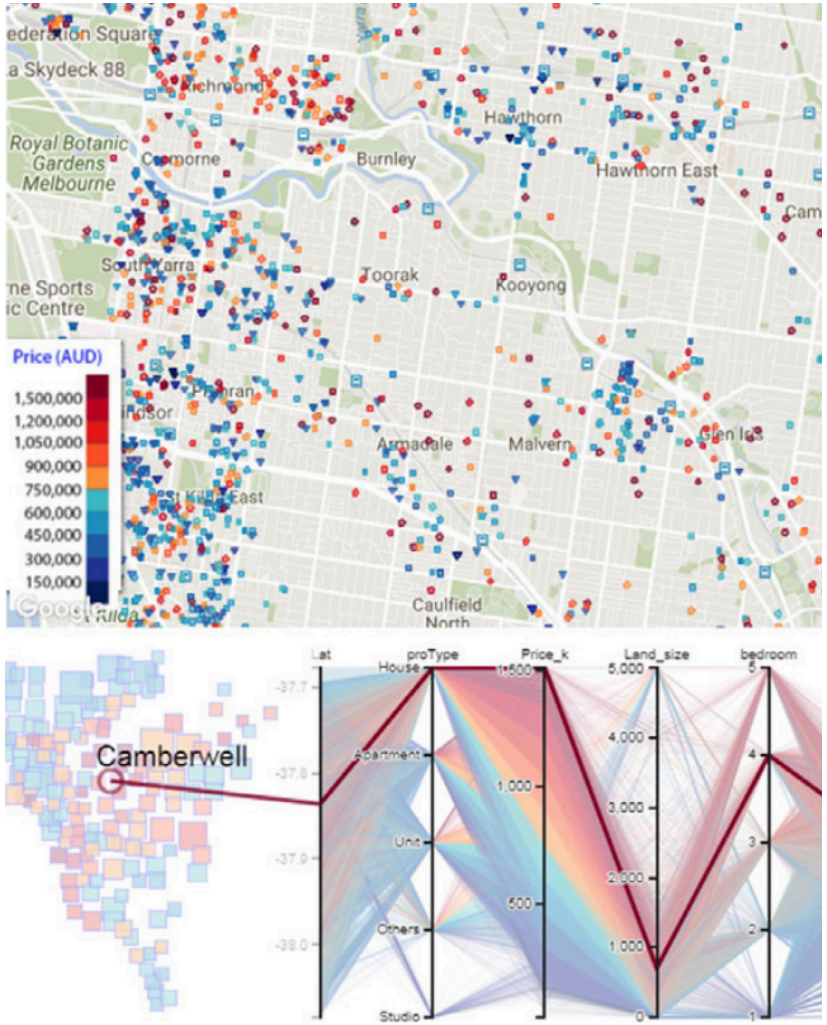
- *e.g. where is such a house located?*

Associate: group objects based on a relationship

- *e.g. which houses are in walking distance from rail line?*

Compare: examine the similarities/differences between objects

- *e.g. which of these two houses has better specs (price, size, etc.)*



HouseSeeker (Li et al 2015) [URL]

Characteristic of Channels [Ward 2010]

Is it Selective? *Important for Categorical Attributes*

- Is a mark easily distinguishable from other marks?
- Can we make out the difference between two marks?

Is it Associative?

- Does it support grouping?

Is it Ordinal

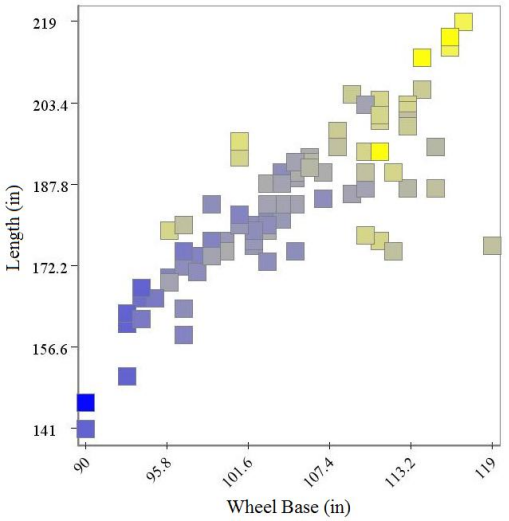
- Can we see a change in order?

Is it Quantitative

- Can we quantify the difference between two marks?

What is the Range/Discriminability?

- How many unique marks can we make?
- For quantitative attributes, this is also related to *accuracy*



1. Position

Placement of graphics within the display space

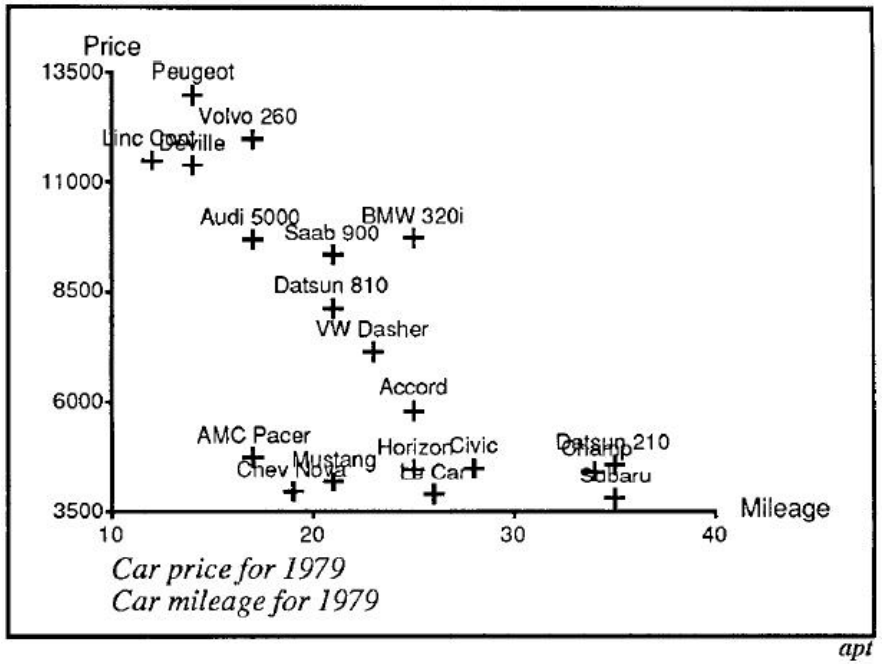
- Greatest impact in most visualizations.

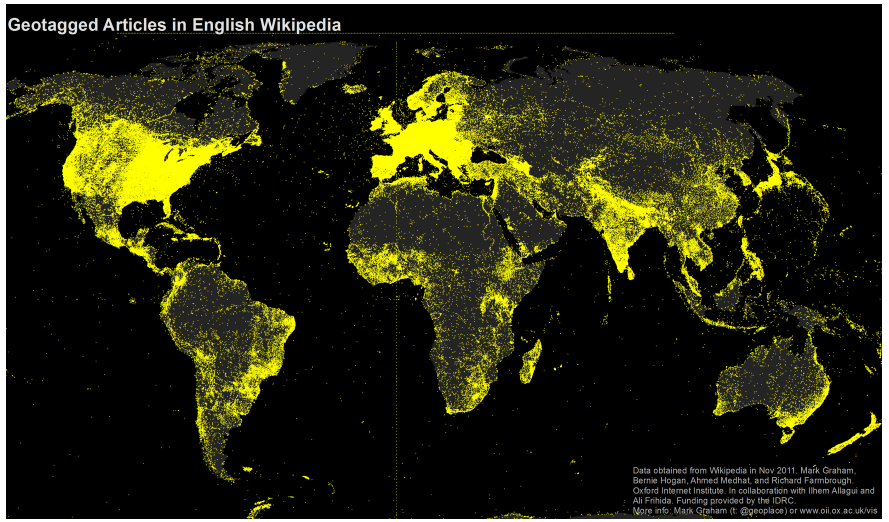
Position is used in ALL visualizations in some way

- Encoding: expressing values
- Encoding: categorizing classes
- Arranging the data for legibility e.g. sorting, alignment, layout
- In spatial data : position is a key attribute

Characteristics

SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	✓	✓	✓	High





Top Images from [Ward 2010]; bottom Mark Graham [2011]

Position

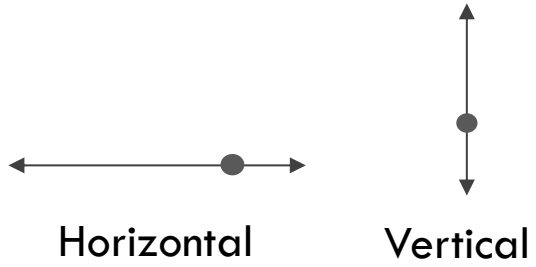
Examples of positional queries:

- Where do most values fall
- Does data fit a well-known distribution
- Are there significant trends
- Are there structures/clusters

Effectiveness can depend on

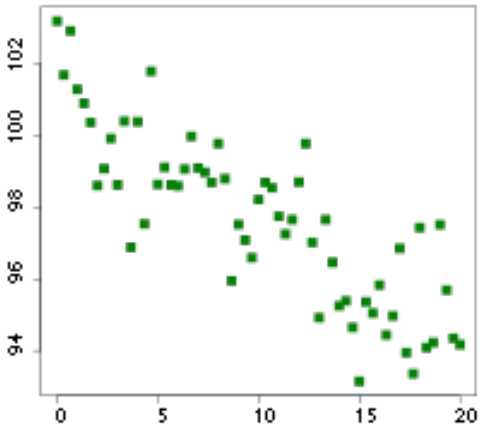
- Scale: linear, logarithmic etc.
- Supplementary graphics such as axes
- Projection: dimension reduction from *higher-D* space to 2D for conventional display

1-Dimensional

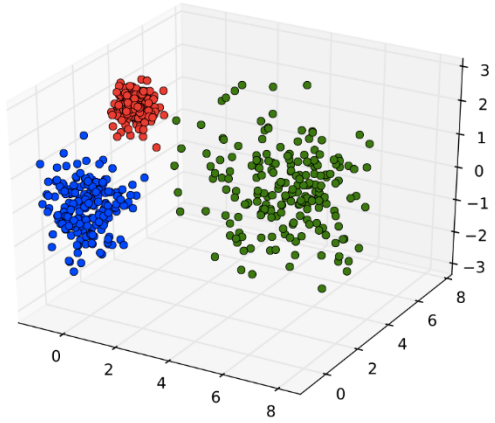


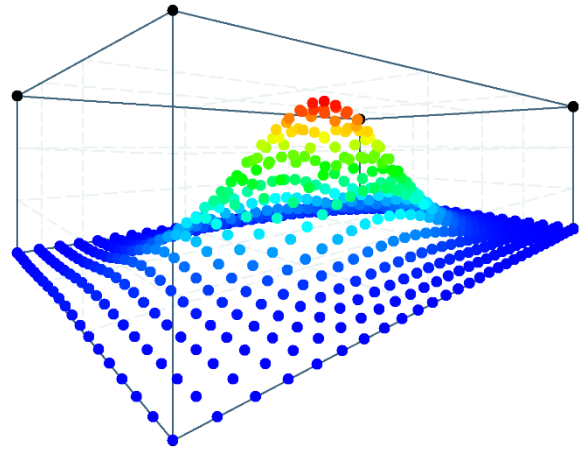
Horizontal Vertical

2-Dimensional



3-Dimensional





Generally, 3D should be avoided unless specifically relevant e.g., mathematical / scientific data sets with spatial arrangement in 3D

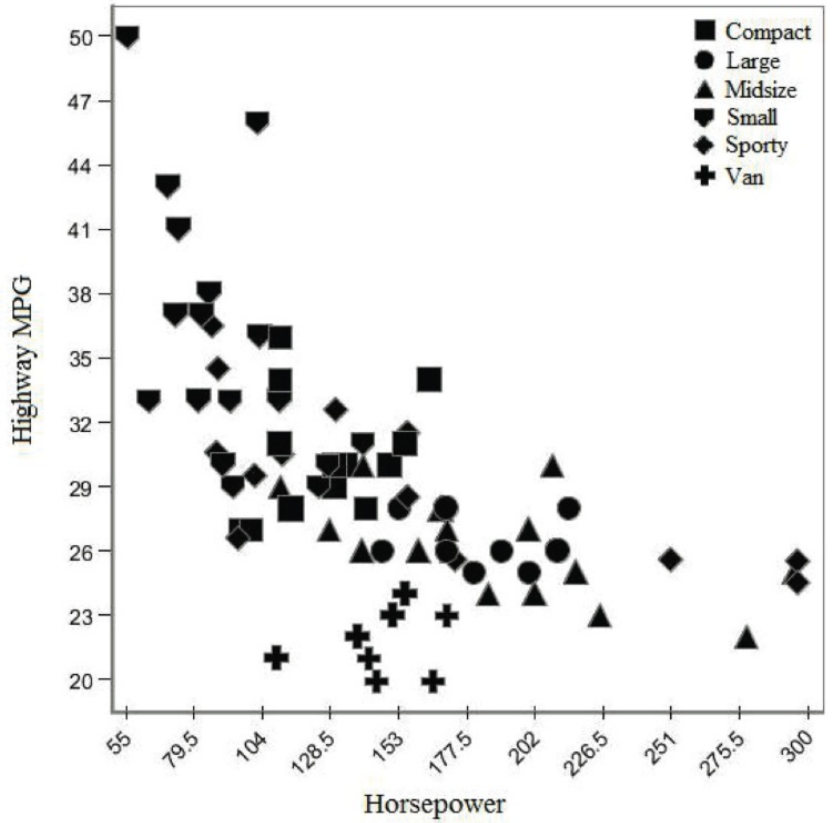
2. Shape / Mark

Mark: use of graphics primitives to create geometric elements in visualization, e.g. points, lines, areas, volumes and their compositions. Implicitly required in visualization but not always useful to express data

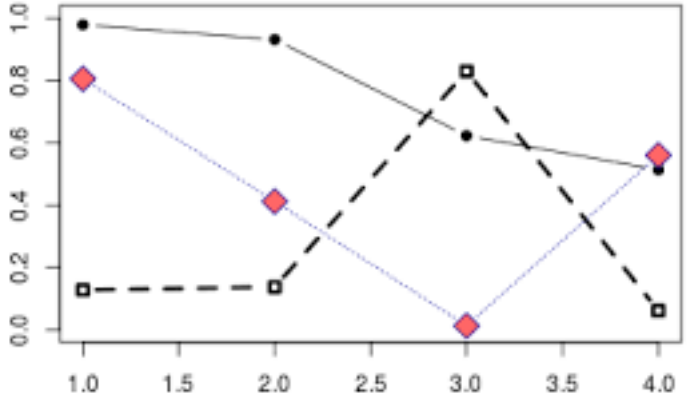
Representing attributes of data through shape

- Good for categorical data; can recognizably represent many classes.
- Shape alone doesn't usually allow reliable ordering or grouping.

SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	≈	✗	✗	High
	(somewhat)			



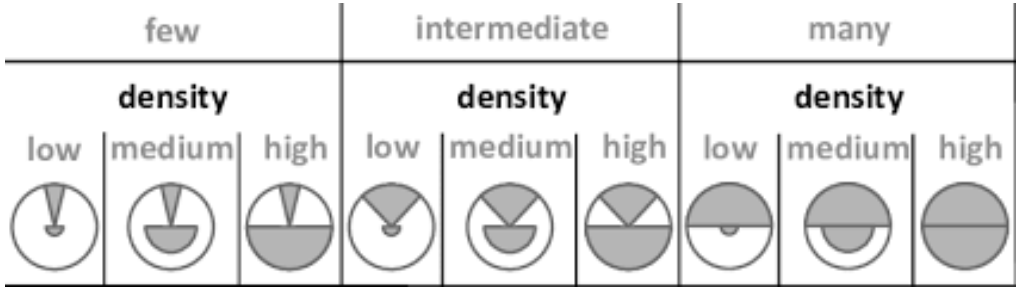
Marks and Shapes



Encoding using shape:

- Aim to maximize distinguishability from others
- BUT need to limit size and complexity; reduce mental load, occlusion

- Glyphs are more complex instances, where shapes are used for higher dimensional encoding



- In general, marks should have equivalent complexity/scale to avoid unfair emphasis

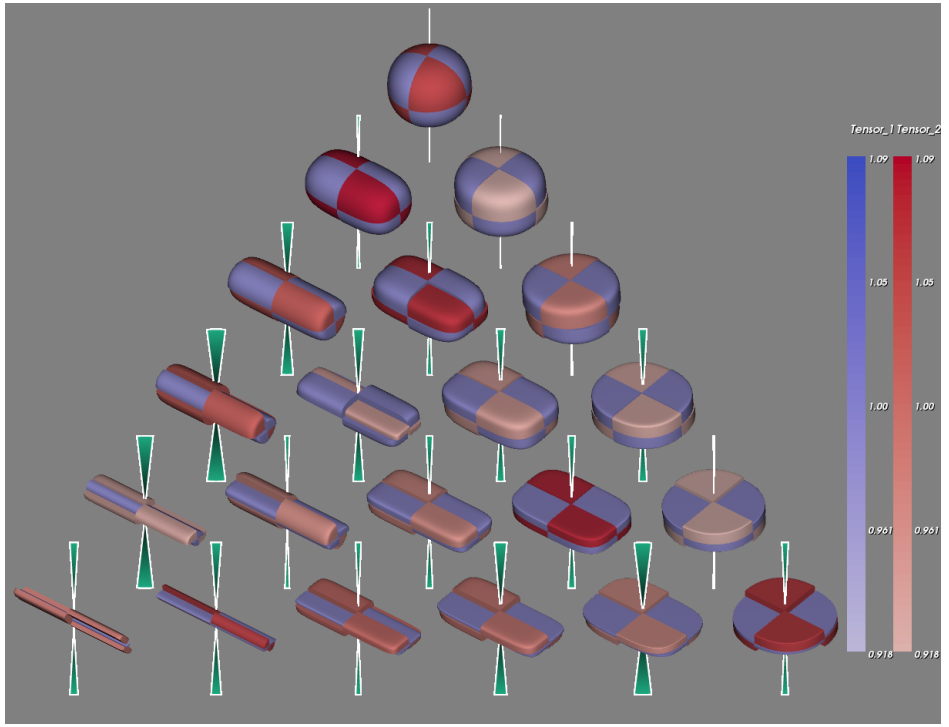
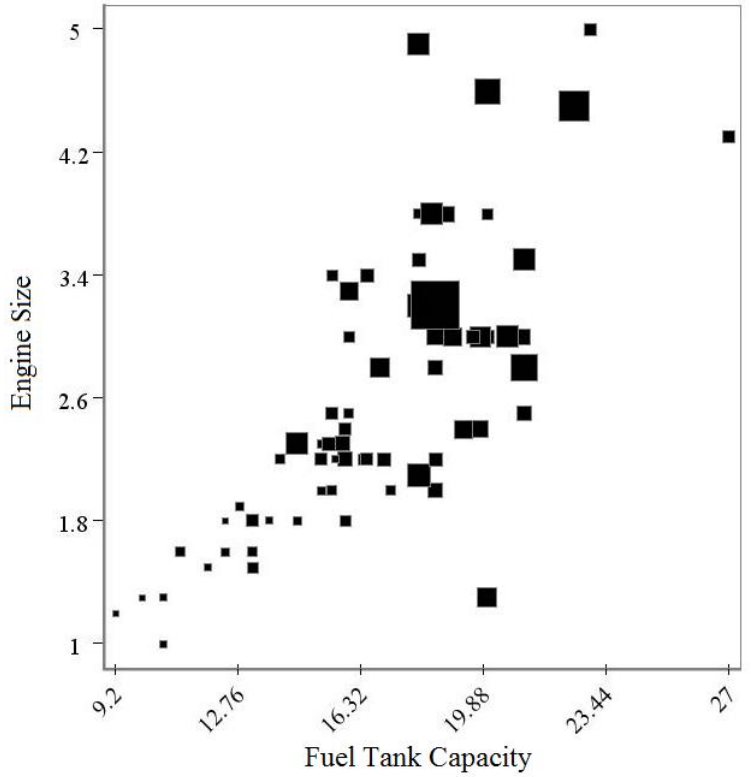


Image © Zhang 2015

3. Size

Represent data through how large/small a mark is drawn

- Data variables expressed as length, area or volume
- Easy to see whether one is bigger, but takes up more screen real-estate than position
- Can be applied to categorical data, but range/discriminability is limited in this case



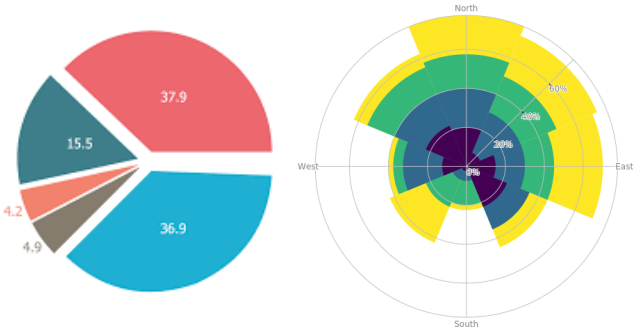
Characteristics (for 1D length):

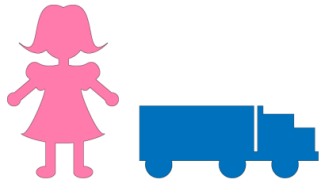
SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	✓	✓	✓	High

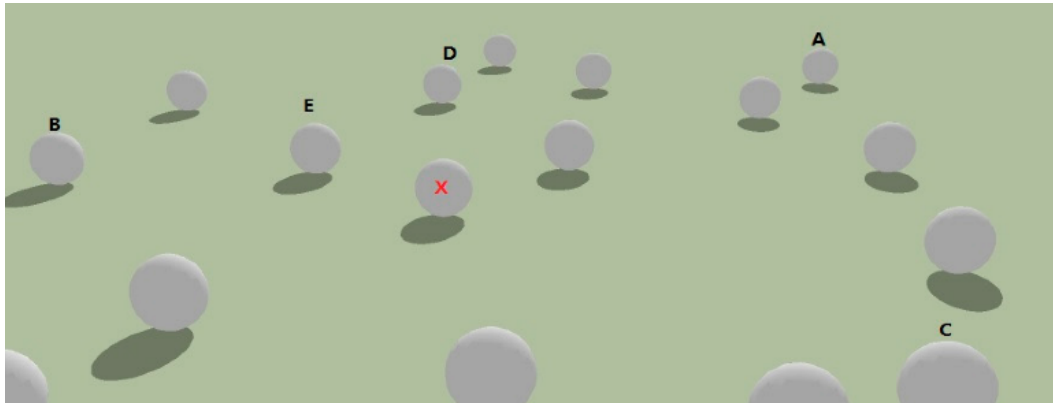
Size

Effectiveness is dependent on type of mark:

- Good for 1D; More suitable than position for comparing measurements e.g. length, mass
- Ok for 2D, but can be imprecise/ambiguous if comparisons are non-linear e.g. angles, non-uniform shapes
- Usually poor for 3D : due to perspective, occlusion, depth ambiguity etc..



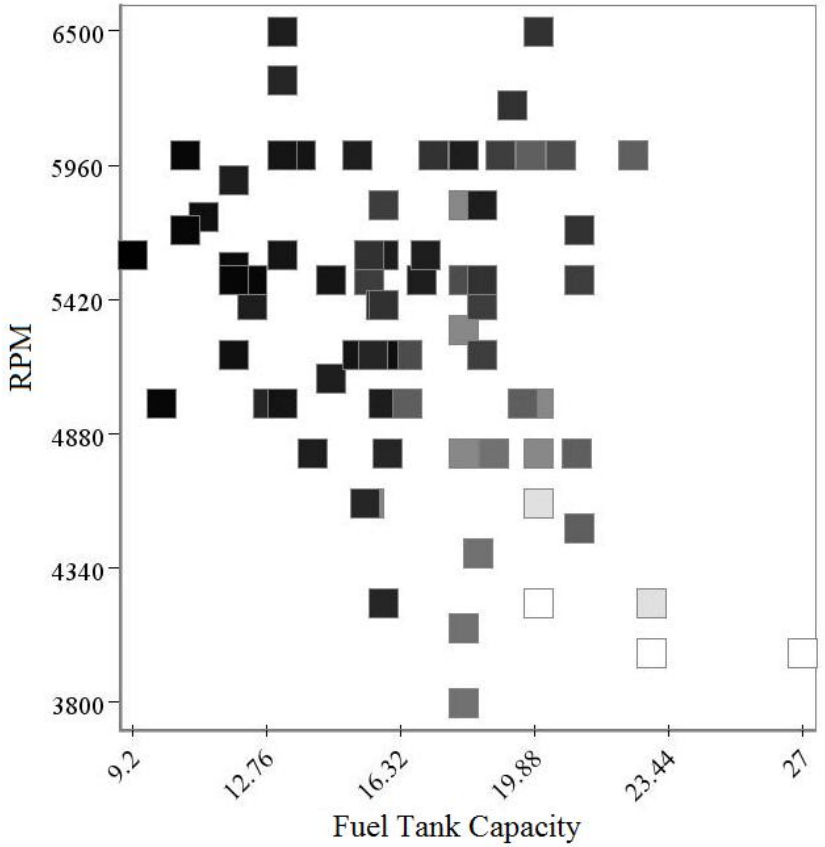




4. Brightness

Represent data through how a mark is shaded

- Also called Value/ Intensity / Luminance
- OK for quantitative data, limited for categorical data
- Disadvantages:
  - Not very many shades distinguishable to human eye
  - Sometimes subjective: Visual response is not linear



Characteristics

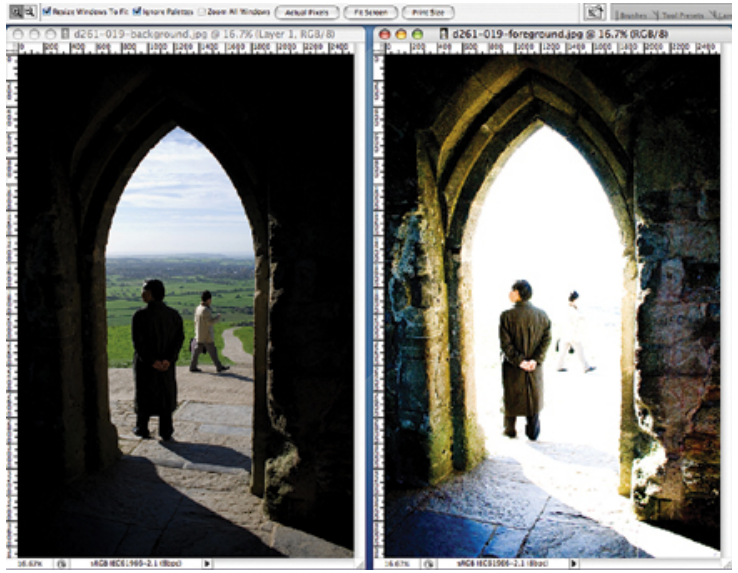
SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	✓	✓	≈	High

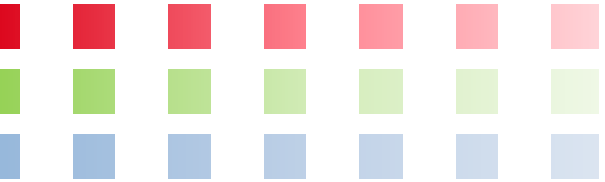
(somewhat)

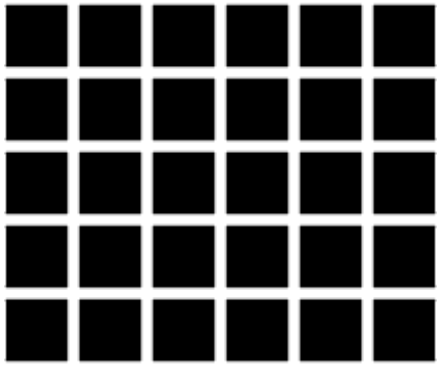
Subjective Brightness Perception

Perceived Brightness affected by

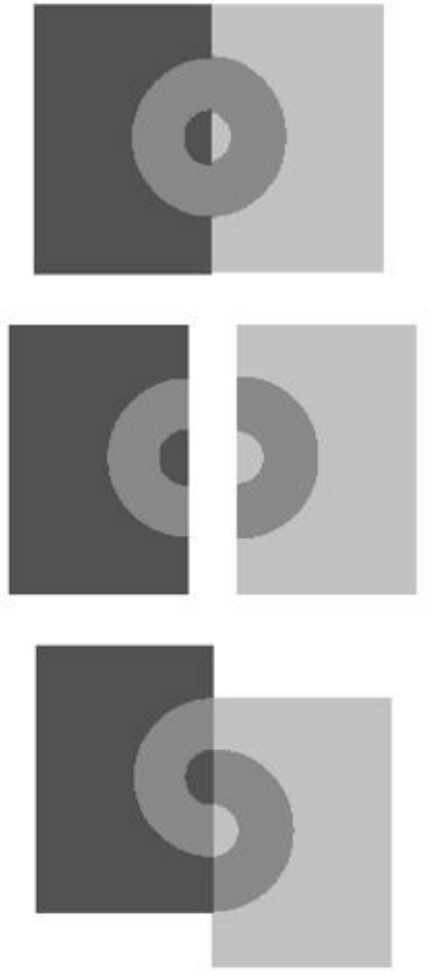
- Surrounding area
- Adaptation
- Size of mark – brightness perceived in an area not a point
- Hue





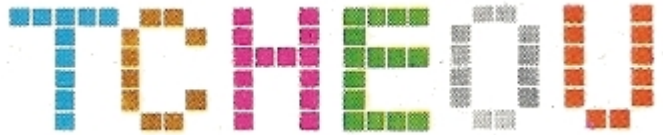


Hermann Grid Illusion



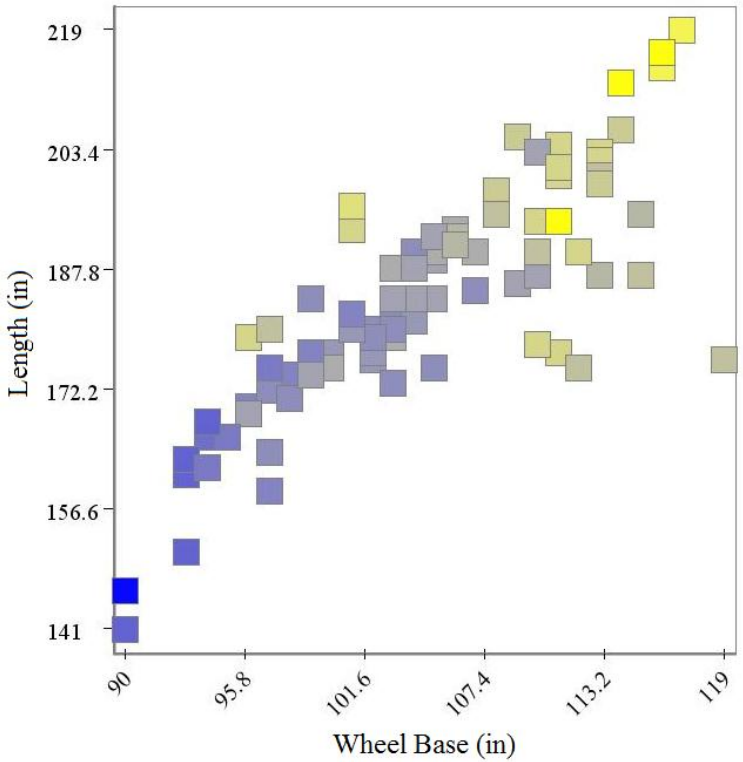
Koffka Ring Illusion

5. Colour/Hue



Represent variables through colour

- In visualisation typically refers to hue, saturation (n.b., brightness is also an element of colour)
- Very powerful visual cue; but if used incorrectly can mislead



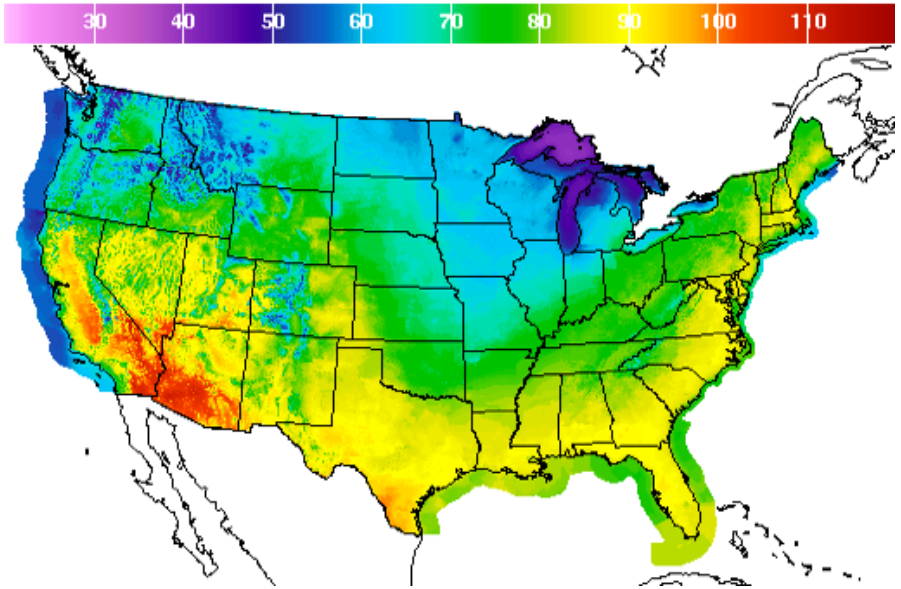
- Good for categorical attributes
- Does not generally work well for quantitative/ordinal attributes

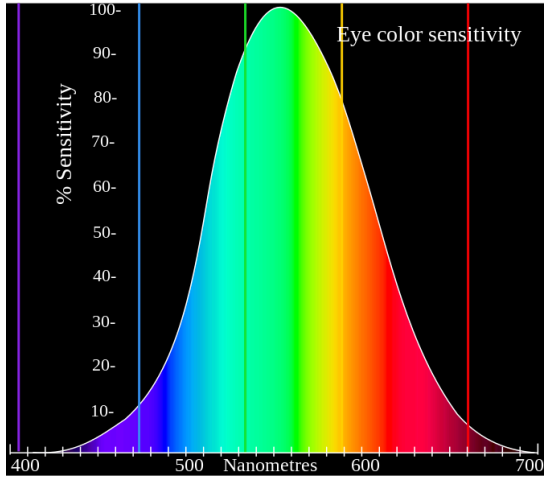
Characteristics

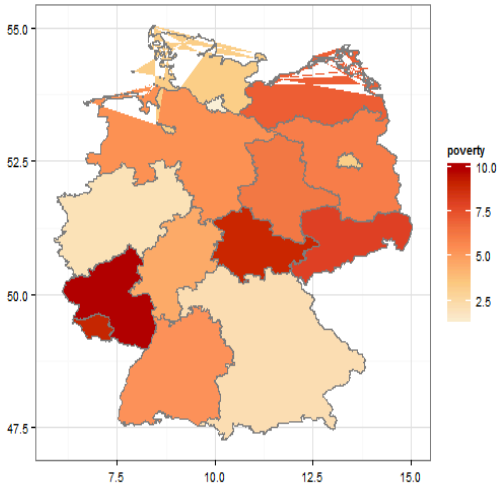
SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	✓	✗	✗	Limited

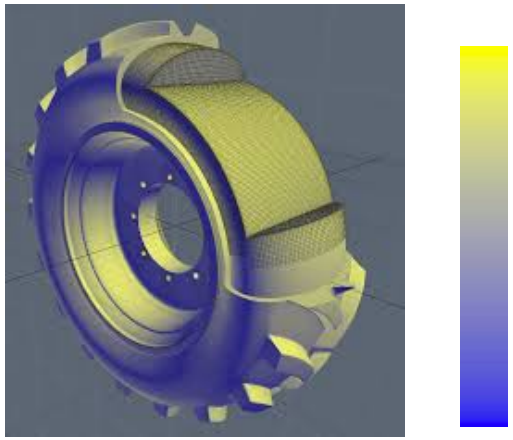
Colour in visualization

- Colour space is 3D, difficult to make reliable one-D mapping / comparisons
- Visual response is not uniform. This affects both quantifiability + categorization
- Effectiveness for quantitative/ordered data depends choice of colours





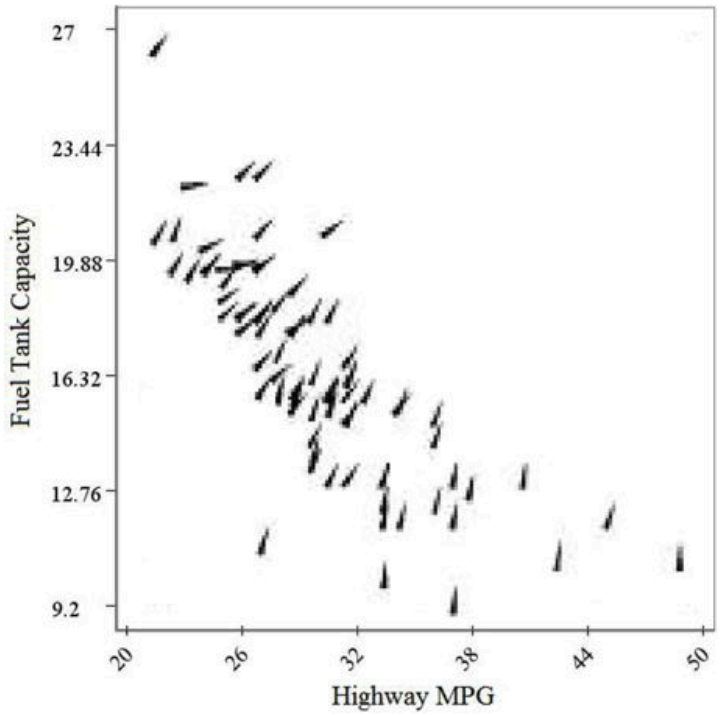




Due to it's importance in visual communication there will be a more intensive discussion of colour in a later lecture

6. Orientation

Image from Ward et al 2014



Mark is rotated in connection to the data variable

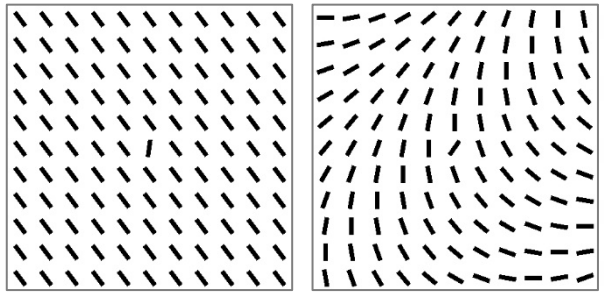
- Also called direction or angle
- Tied to pre-attentive processing, good pop-out effect and parallel processing
- But limited precisions and range (<45 degree changes unreliably perceived)

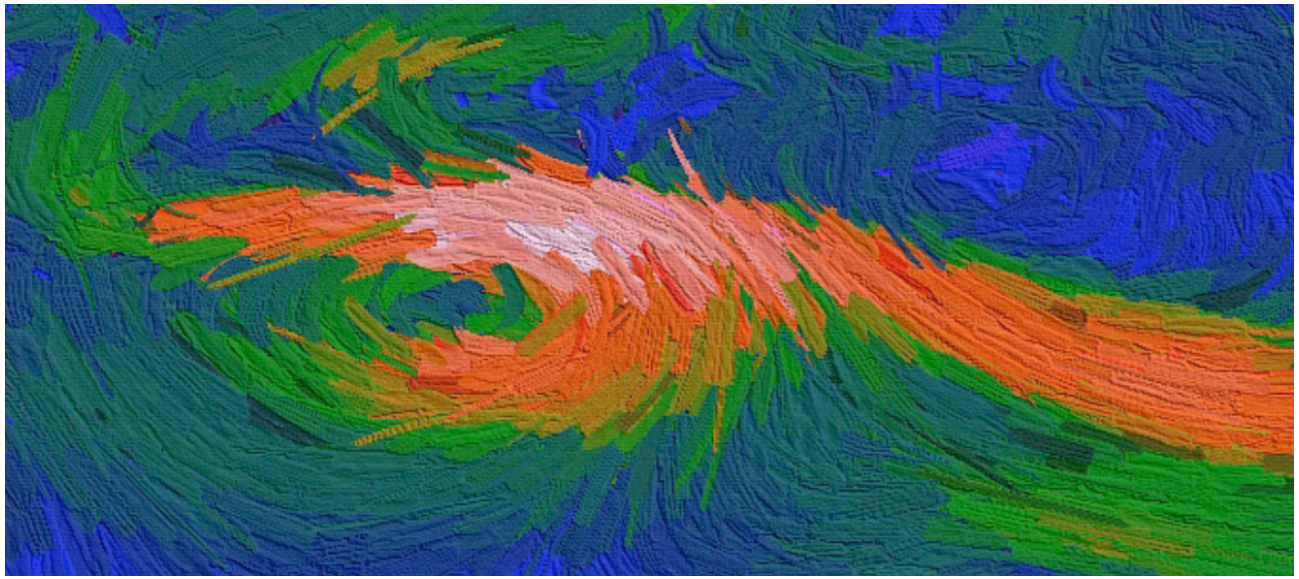
Characteristics

SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
≈	≈	≈	≈	Limited
(somewhat)	(somewhat)	(somewhat)	(somewhat)	

Orientation

- Orientation effect is salient but response is non-linear
- Also affected by heterogeneity of distractors





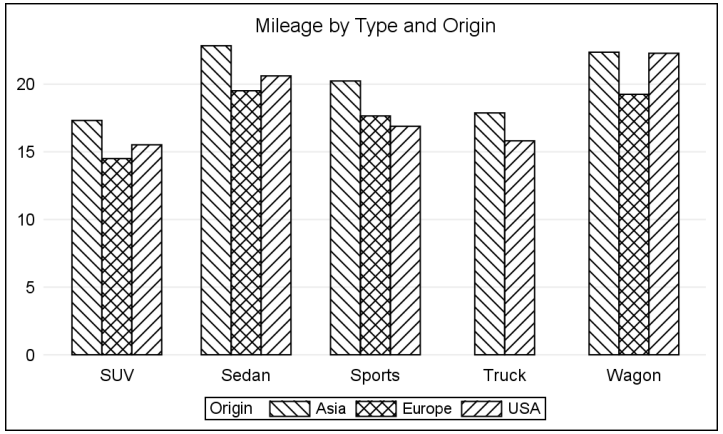
Painterly visualization of a simulated supernova collapse: pressure → luminance, velocity → hue, flow direction → orientation. (C. Healey)

Sea conditions around North American coast. Temperature is mapped to hue and wind speed is mapped to orientation. (C. Healey).

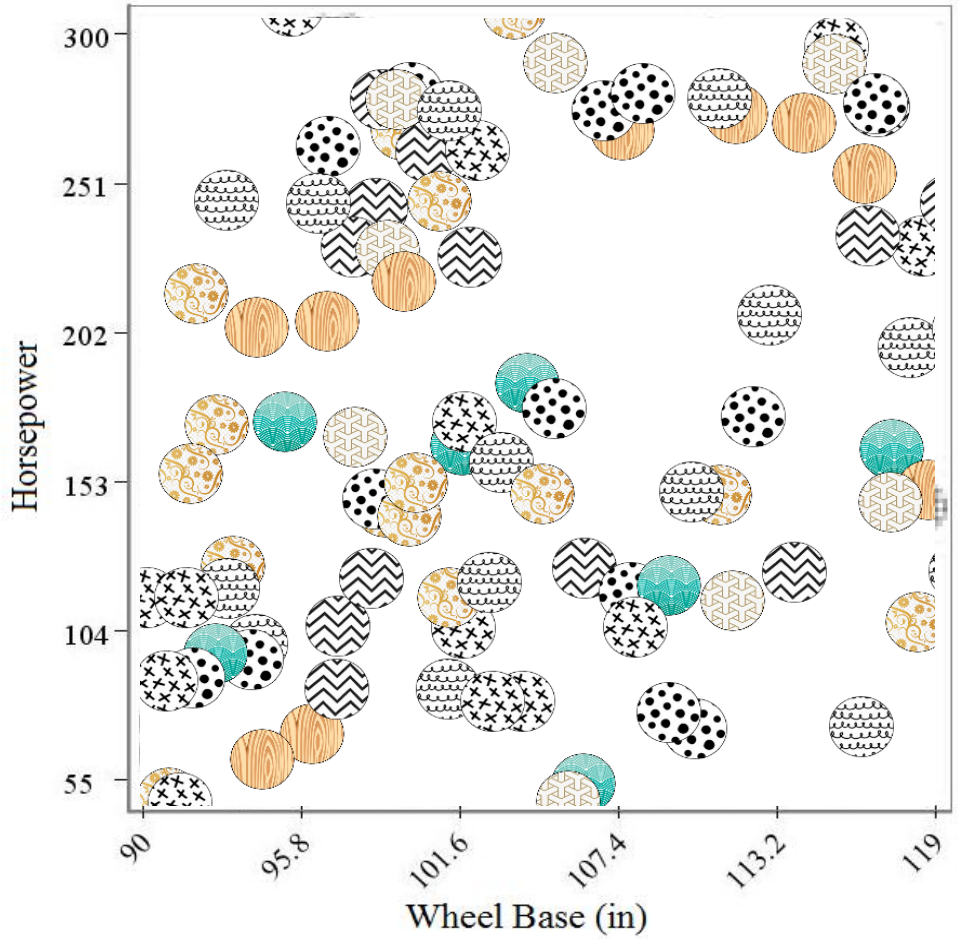
7. Texture / Pattern

Data encoded in visual patterns

- Can be considered a combination of other variables (mark, color, orientation, etc., of a shape, region or surface)



Sometimes used as a replacement for colour



Characteristics

SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	✓	✗	✗	High

Texture / Pattern

Can be decomposed into fundamental perceptual dimensions

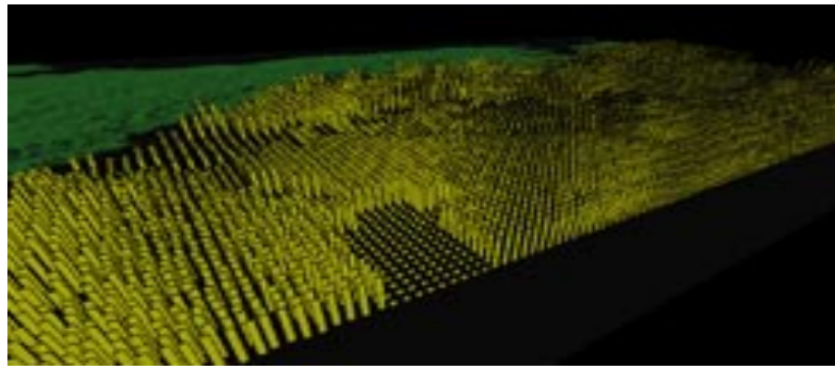
- Regularity
- Directionality
- Contrast
- Size / Coarseness / density

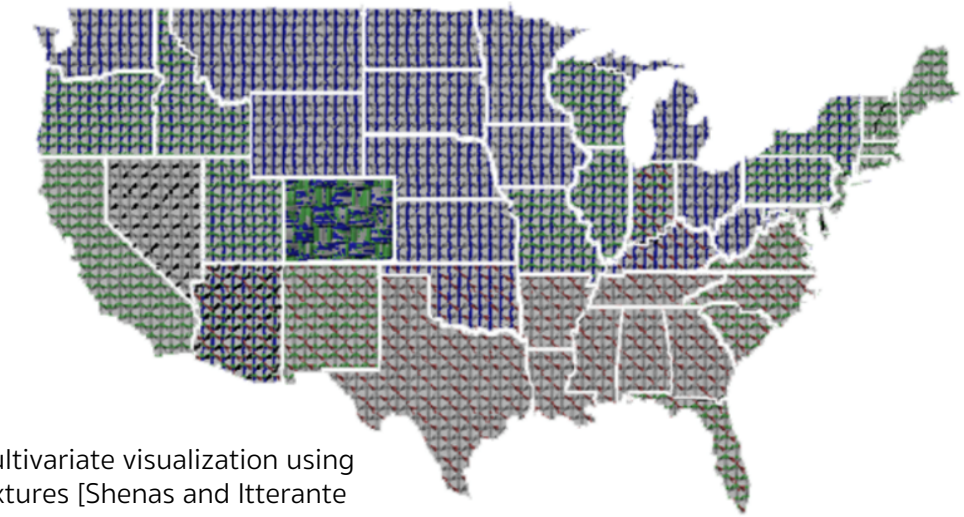
High-dimensionality provides opportunities for multi-variate visualization

Disadvantages

- More complex to design
- High cognitive load
- Space/time required to depict a pattern

Ocean conditions visualization {Healey and Enns 1998]: plankton density mapped to height, temperature mapped to density, current mapped to regularity;





Multivariate visualization using textures [Shenas and Itterante 2005]

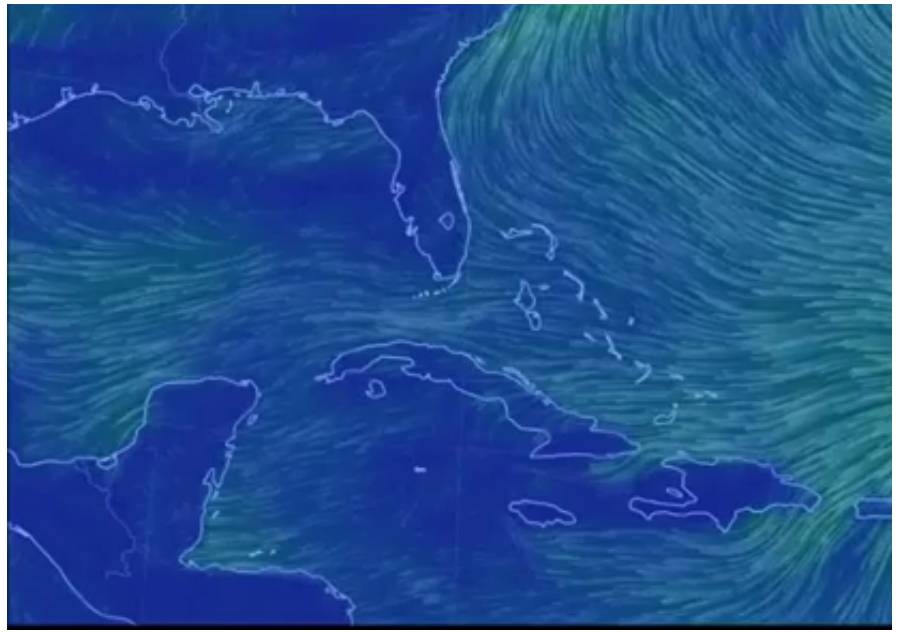
8. Motion

Information encoded through change in positions

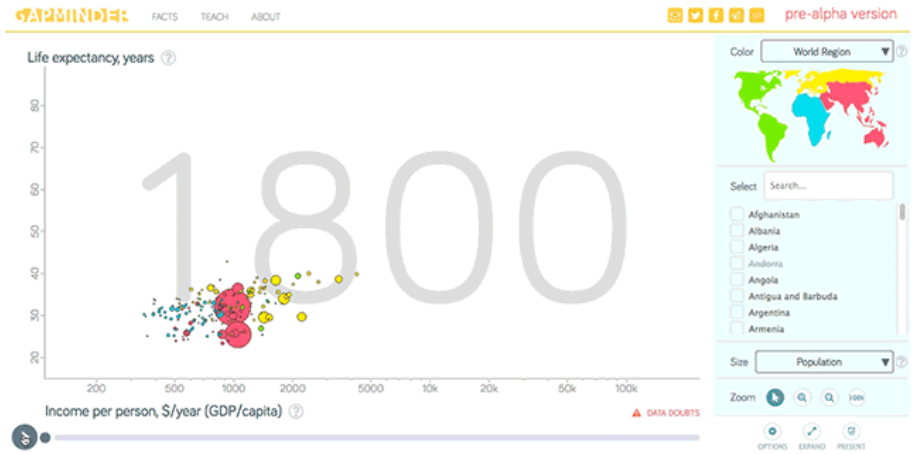
- A change in other visual variable e.g., moving (varying position); flashing (varying opacity)
- Data values encoded through the degree of change e.g. Speed or Direction of motion
- Dependence on another channel reduces separability
- Space/time interval required to perceive change; more complex to implement
- Good pop-out due to outliers, grouping due to similarities

Characteristics

SELECTIVE	ASSOCIATIVE	ORDINAL	QUANTITATIVE	RANGE
✓	✓	✓	✓	Limited



Direct data encoding to motion [https://earth.nullschool.net/]



Indirect encoding to motion © gapminder

Recommended Readings

- Chapter 1 – "Foundations for a science of Data Visualization" in **Information Visualization (4th Edn.)**, Colin Ware 2019
  - single chapter excerpt available at (click on preview):
    - https://www.ebooks.com/en-ie/book/209920967/information-visualization/colin-ware/?_c=1
    - http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/articles/book_information-visualization-perception-for-design_Ware_Chapter1.pdf

- Chapter 5 – "Marks and Channels" in **Visualization Analysis and Design**, Tamara Munzner 2014

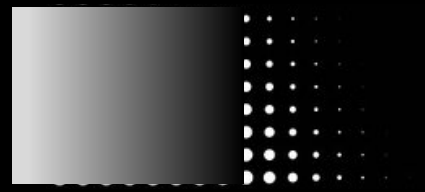
  [Available as e-book in Library Reading Rooms]

  [A free low-quality pre-publication draft of is available on Author's web page: https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=88]

- Chapter 4 – "Visualization Foundations" in **Interactive Data Visualization**, Ward et al 2010 [Available as e-book in Library Reading Rooms]

The End

For now



References

Mingzhao Li, Zhifeng Bao, Timos Sellis and Shi Yan. Visualization-Aided Exploration of the Real Estate Data. Australasian Database Conference. September 2016. DOI:10.1007/978-3-319-46922-5_34