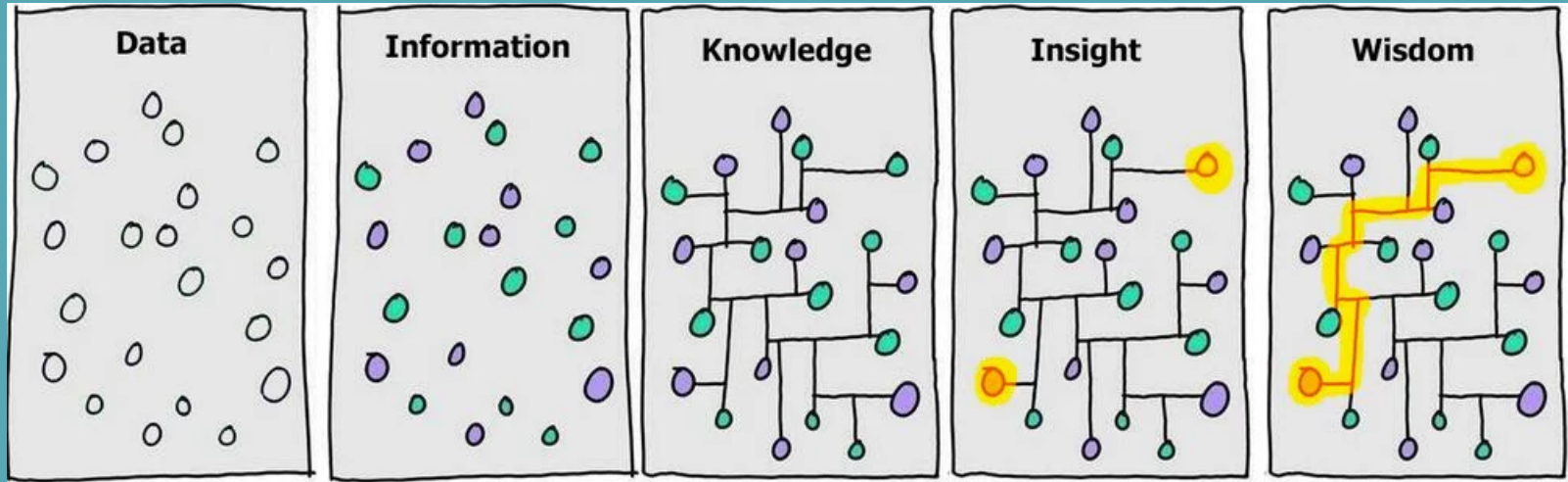


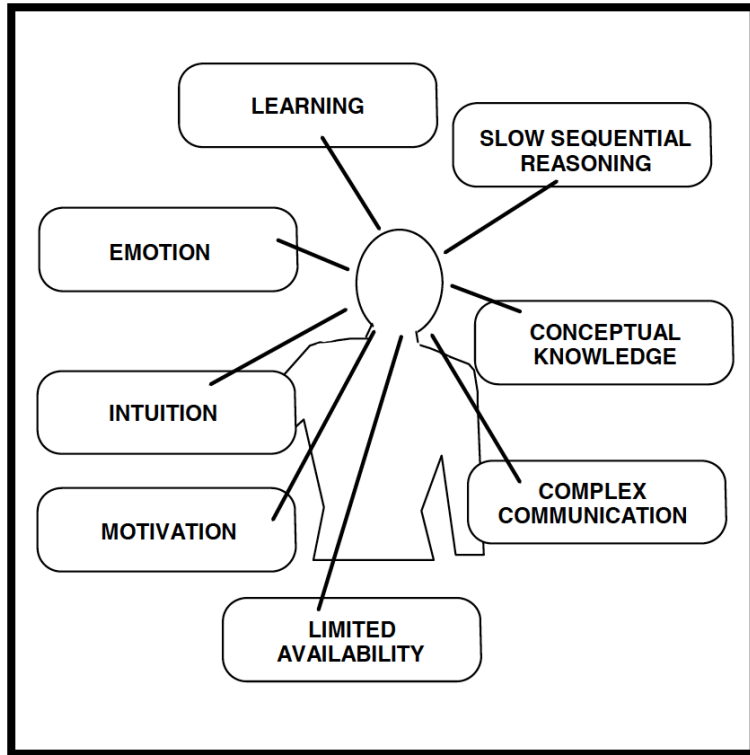
03

DATA-DRIVEN

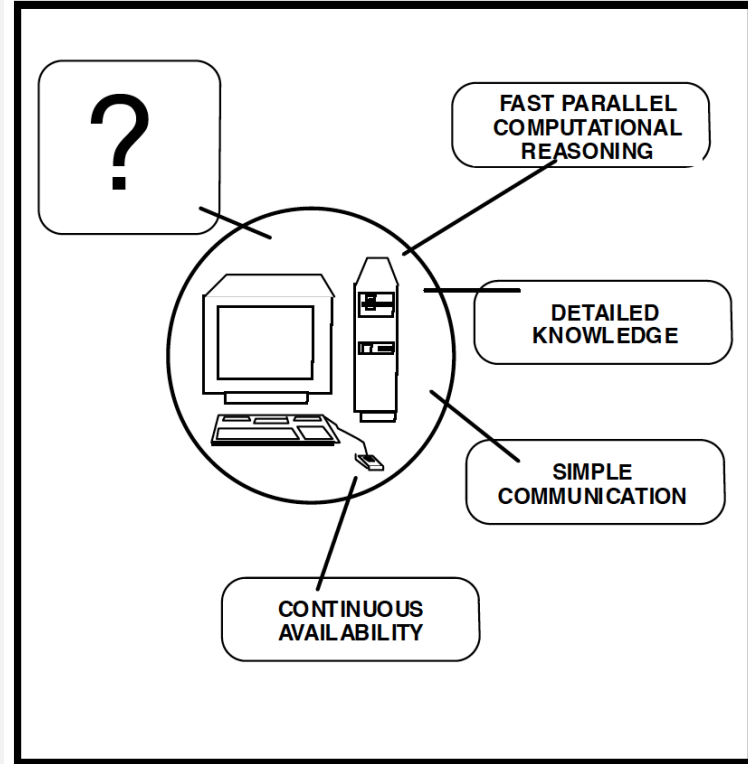


DATA-DRIVEN: Human versus Computer

Humans are pattern-driven



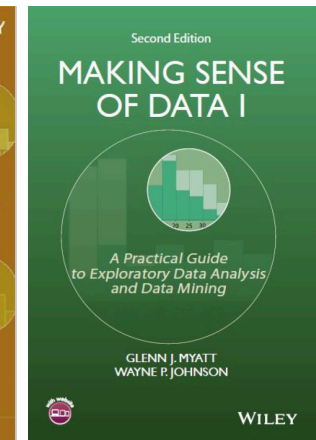
Computers are data-driven



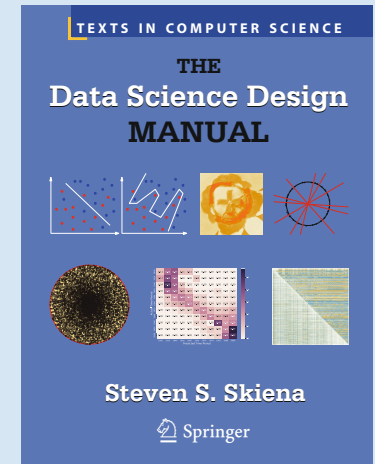
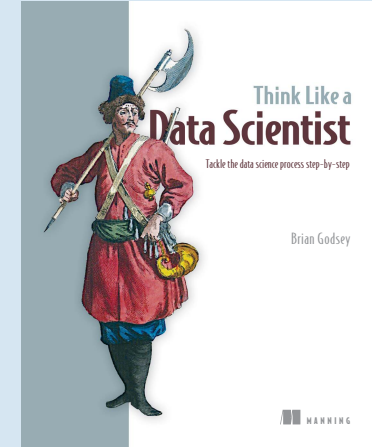
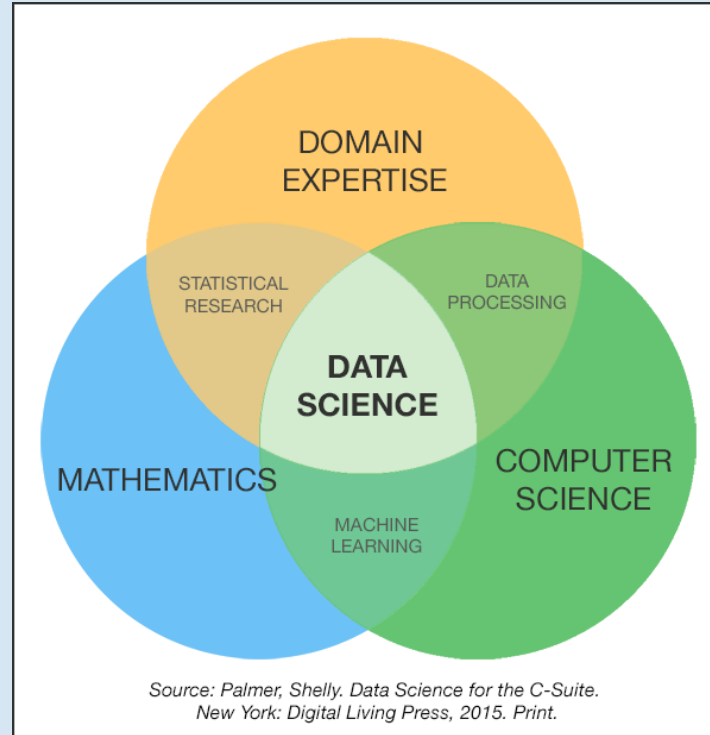
DATA-DRIVEN: Data wordt op steeds grotere schaal verzameld

Innovation type	The stages of computing innovation				
	Problem	Innovation	Proof/ recognition	Adoption	Refinement
Computing	Cracking codes High-powered physics Ship navigation	Pre-1950s • Purpose-built computing machines	~1950s • Enigma • ENIAC	~1970s • First PCs • Computers in schools and libraries	~1980s • Supercomputers • Consumer PCs
Networking	Communicating and sending text and files	~1970s • pre-internet • ARPANET	~1980s • Academic networks • IRC	1990s • Prodigy • Compuserve • AOL	2000s • Mobile devices • Social networks • Cloud services
Big data collection and use	Too much useful data being thrown away	~2000 • Web crawling • Click tracking • Early, big social networks	2000s • Google search • Big retailers tracking users	2010s • Twitter firehose • Hadoop	2015+ • Massive API development • Format standardization
Big data statistical analysis	Even basic statistics are hard to calculate on large data sets	2000s • Google search • Amazon streamlining processes	2010s • Netflix challenge • Kaggle.com	2015+ • Google Analytics • Budding analytics start-ups	2025+? • Ubiquitous intelligent, integrated systems

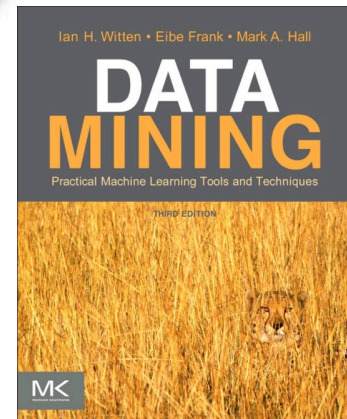
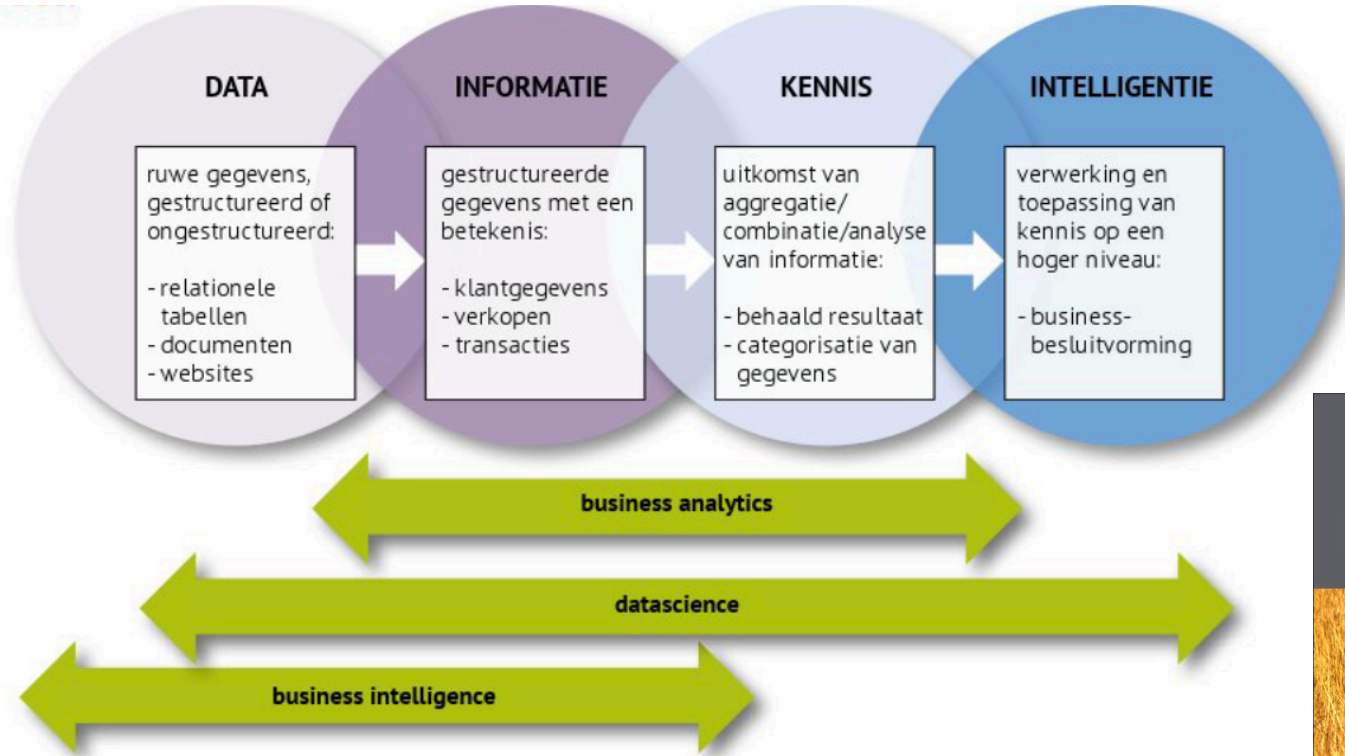
We're currently (2018) in the refinement phase of big-data *collection* & the adoption phase of *statistical analysis* of big data.



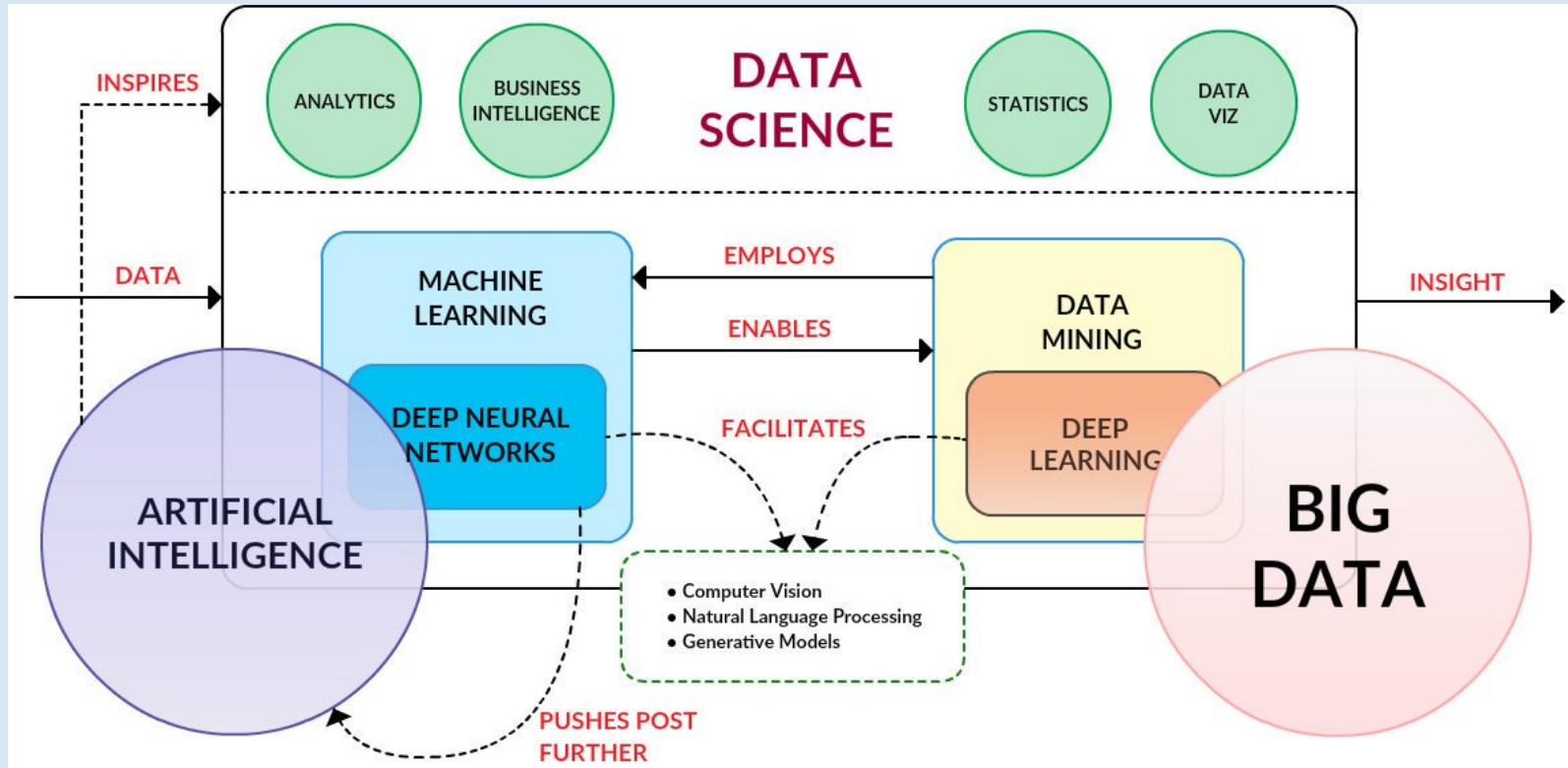
Wat geeft grip op DATA?



INFORMASIE



Wat geeft grip op DATA?

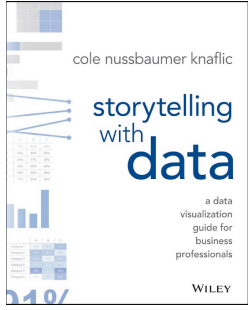
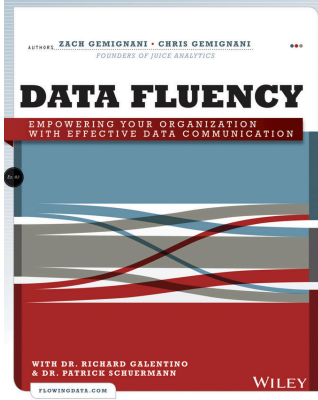
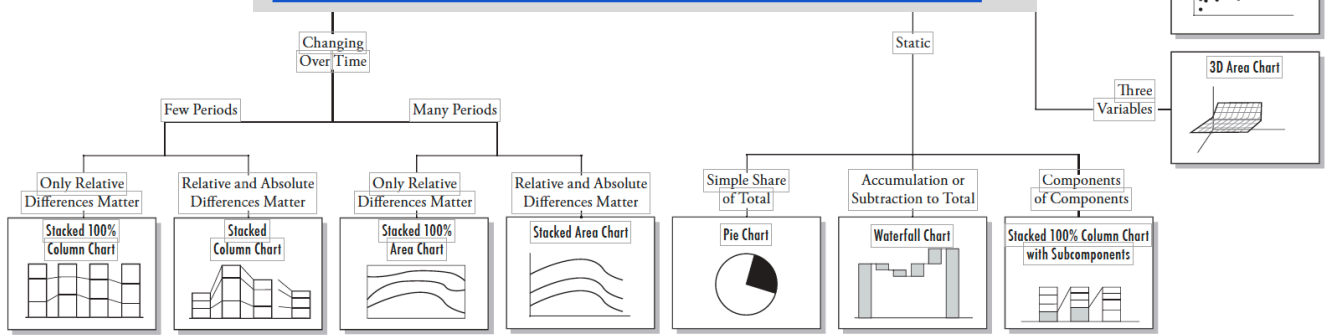
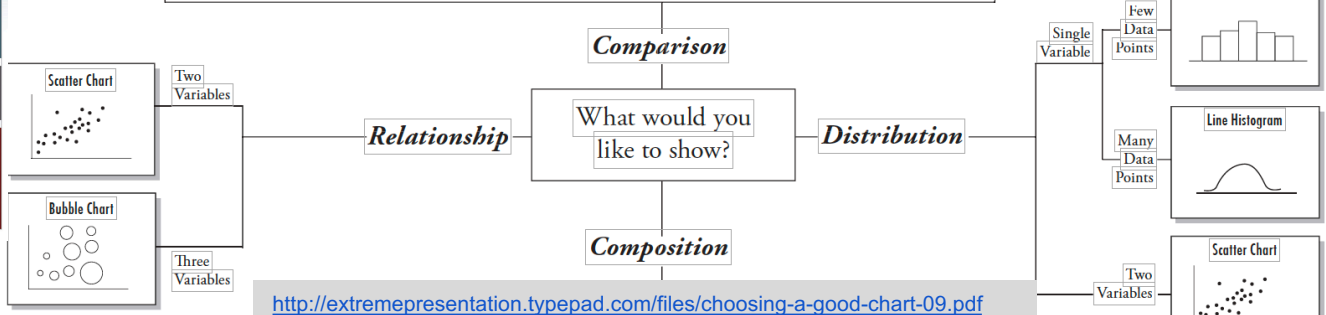
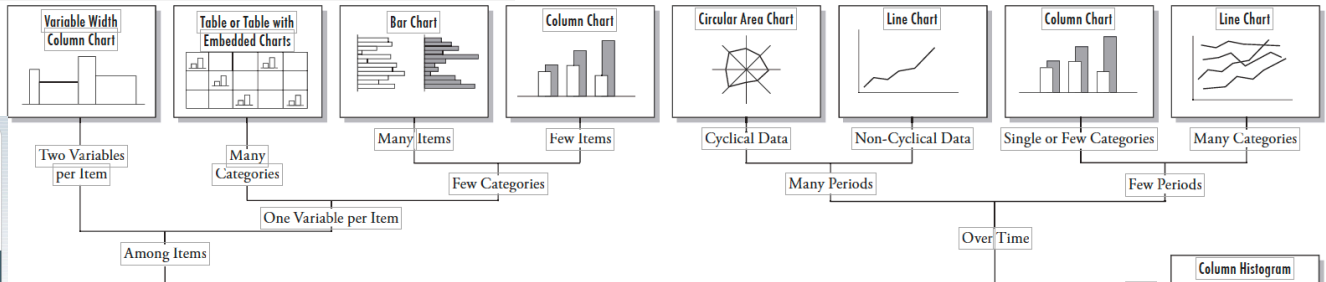


Wat geeft grip op DATA?



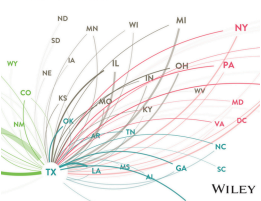
Wat geeft grip op DATA?

<http://labs.juiceanalytics.com/chartchooser/index.html>

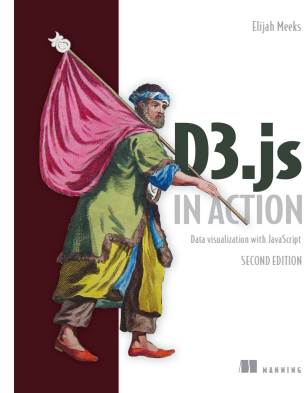
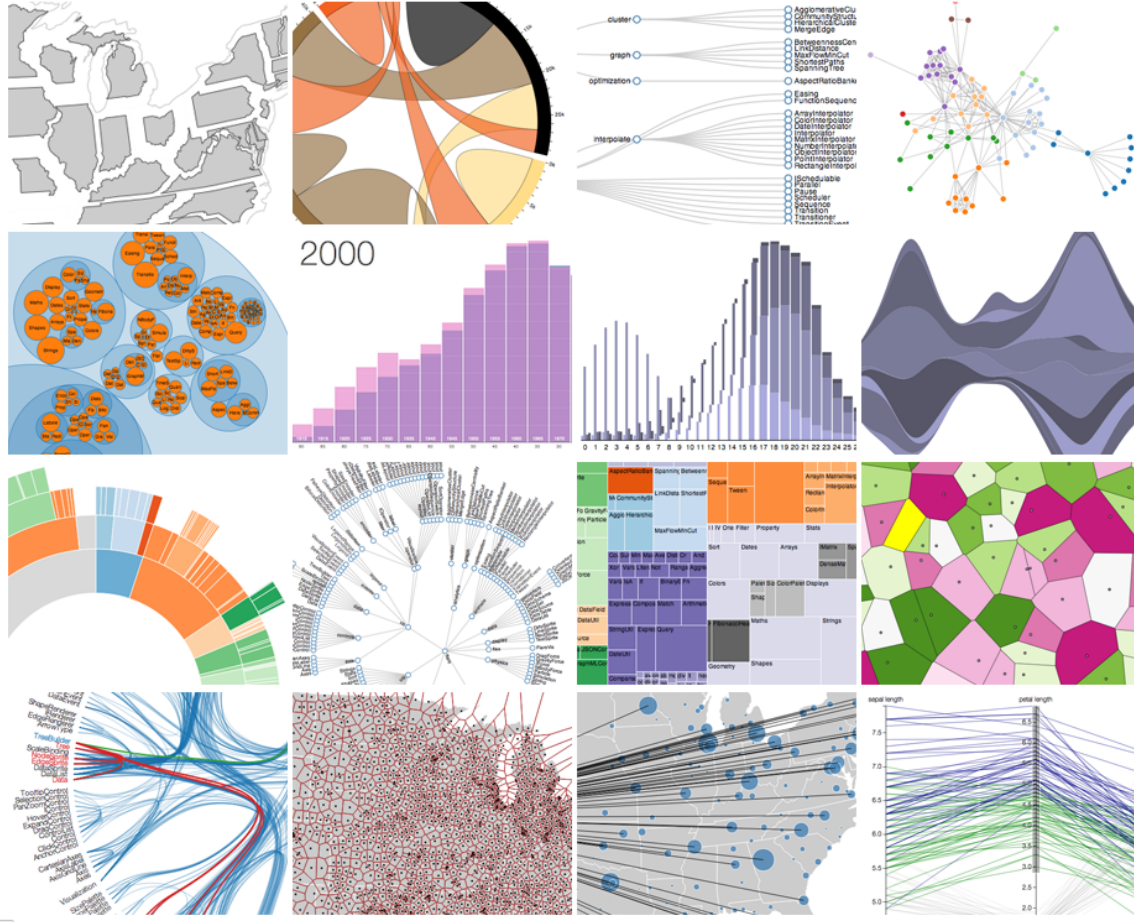
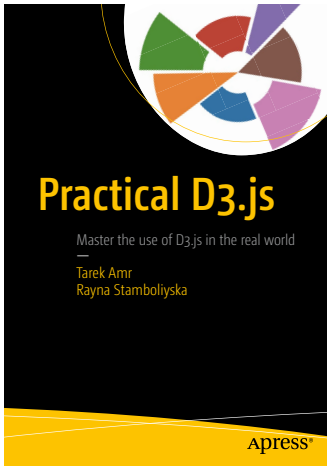
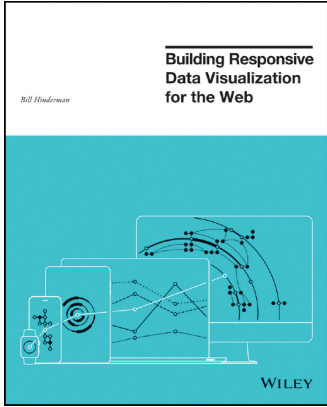


Graph Analysis and Visualization

Discovering Business Opportunity in Linked Data



Wat geeft grip op DATA?



Can you find the
the **mistake?**

1 2 3 4 5 6 7 8 9

DATA-DRIVEN: WHAT IS DATA?

Data [gegevens]

Raw Facts
No Context
Numbers
Symbols

Data comes from the Latin word, "datum," meaning a "thing given."

Although the term "data" has been used since as early as the 1500s, modern usage started in the 1940s and 1950s as practical electronic computers began to input, process, and output data.

98734975471894614398734578
20875980542158009258202908
12349823094823048002343423

98734975471894614398734578
20875980542158009258202908
12349823094823048002343423

TYPES OF DATA: Quantitative versus Qualitative [numerical vs categorical]

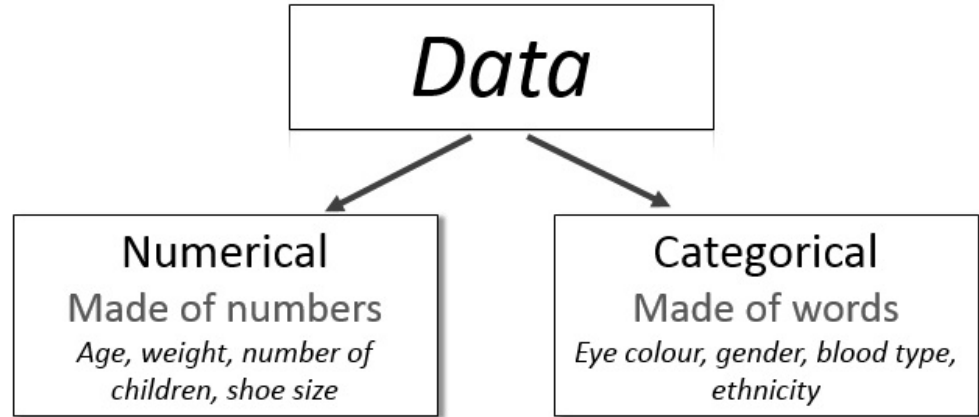
Data Quantification

Quantitative [Numerical] data:

This data can be described using **numbers**, and basic mathematical procedures, including addition, are possible on the set. It can be **discrete** (countable numbers) or **continuous** (infinitely large or small)

Qualitative [Categorical] data:

This data are categories. It cannot be described using numbers and basic mathematics. Is generally thought of as being described using "natural" **categories** and **language**.



- Quantitative values
 - **Measure** things
 - *Revenue, Units, Marketshare, Duration, Customer Satisfaction, Visits, Price, etc.*
- Categorical values
 - Subdivide things into **groups**
 - *Region, product, category, employee, etc.*

Data



Qualitative

Descriptive
information



“I drink coffee every day”



Quantitative

Numerical
information



Discrete
(Counted)



“I drink 4 coffees every day”



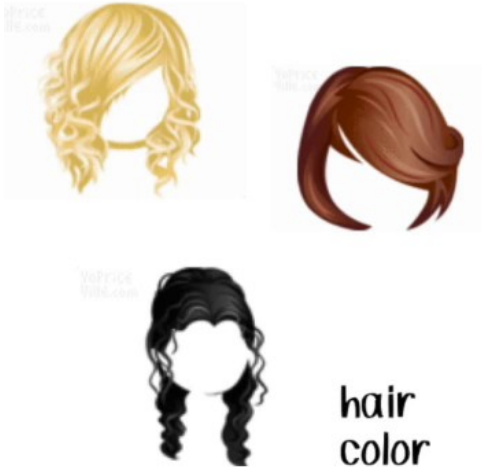
Continuous
(Measured)



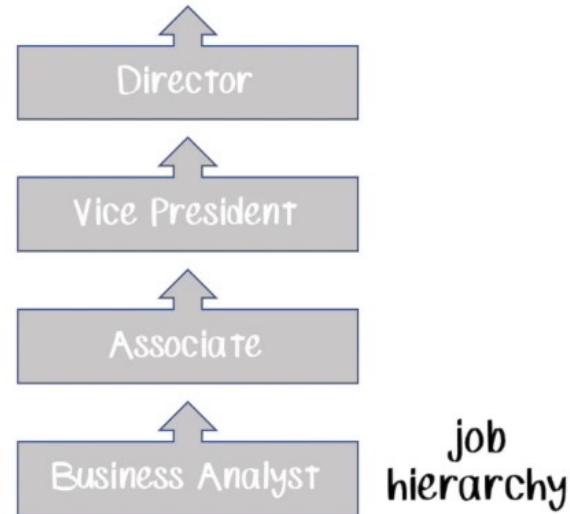
“I drink 80grs of coffee every day”

CATEGORICAL DATA

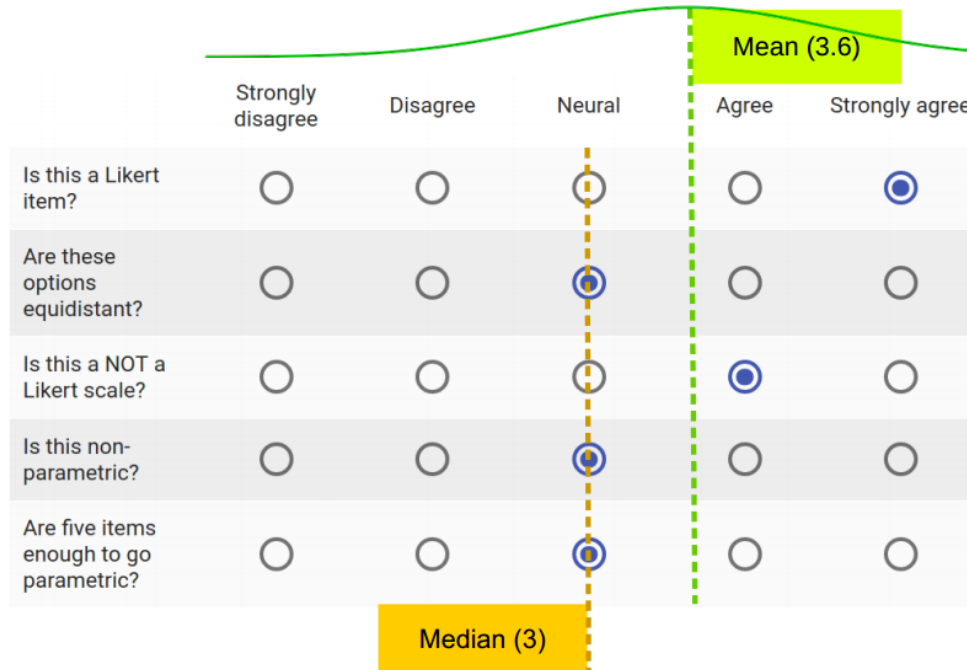
NOMINAL DATA



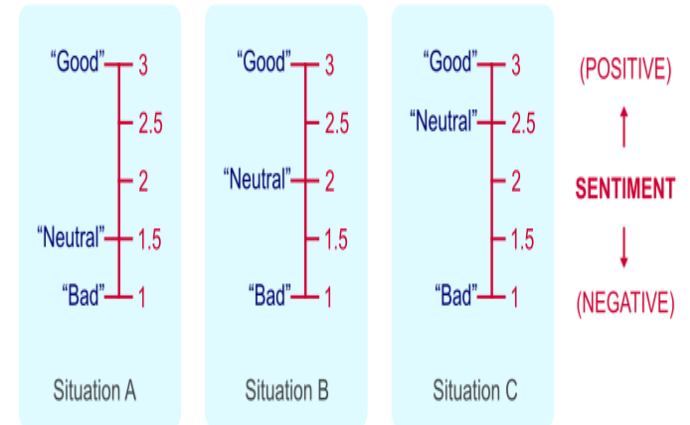
ORDINAL DATA



Likert-scale data



ORDINAL VARIABLE - INTERVALS ARE UNKNOWN

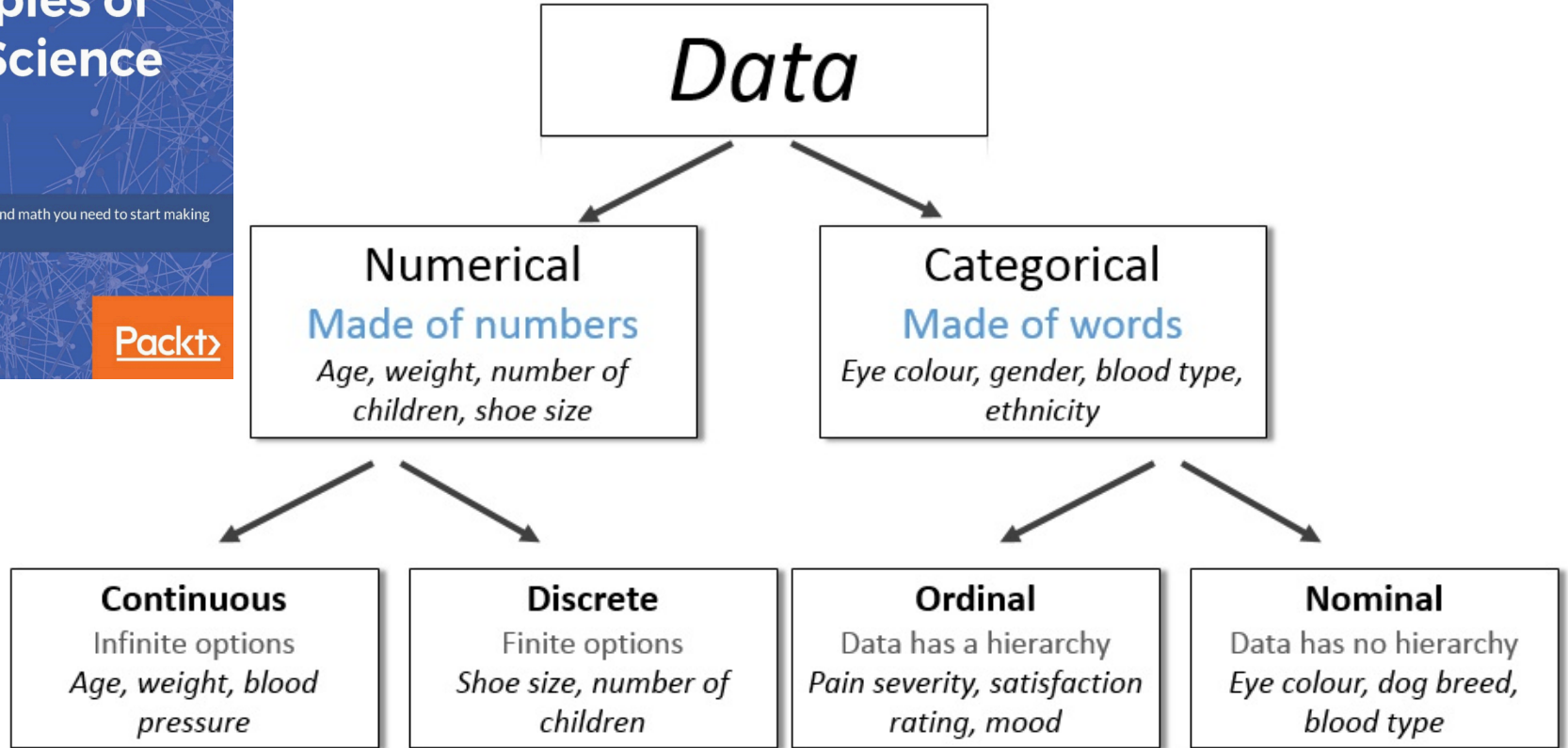


Principles of Data Science

Learn the techniques and math you need to start making sense of your data



Packt>



We onderscheiden 4 meetniveaus:

nominaal + ordinaal [discrete data]

interval + ratio [continue data]

Meetniveaus / Meetschalen:

Wanneer je onderzoek doet heb je vaak **variabelen** die je hierin moet verwerken.

Variabelen zijn elementen uit een onderzoek die verschillende waarden kunnen aannemen. Deze waarden kunnen worden gecategoriseerd in verschillende meetniveaus.

Meetniveaus kunnen iets vertellen over welke data-analyse geschikt is voor structurering.

LEVELS OF DATA: LEVELS OF MEASUREMENTS/OBSERVATIONS

Meetniveau	Wat je kunt berekenen met behulp van waarden op het meetniveau
Nominaal	Tellen, percentages berekenen
Ordinaal	Tellen, percentages berekenen en hoger/lager aangeven
Interval	Tellen, hoger/lager aangeven, verschillen in eenheden aangeven, gemiddelde, spreiding
Ratio	Tellen, hoger/lager aangeven, verschillen in eenheden aangeven, gemiddelde, spreiding en het berekenen van verhoudingen

Meetniveaus [level] /Meetschalen [scale]:

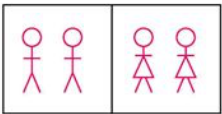
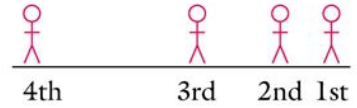
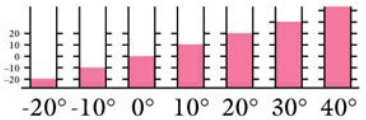
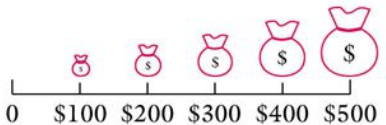
De hoogte van het meetniveau is bepalend voor:

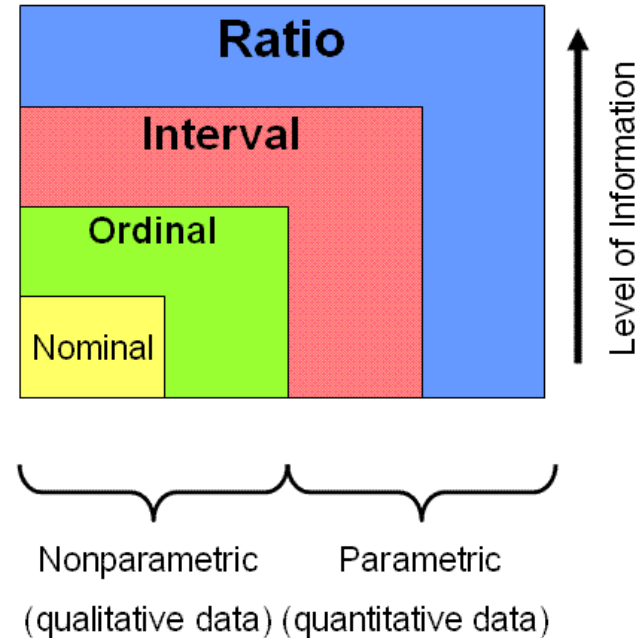
Statische-analyse + Grafische weergave

Meetniveaus & hun kenmerken

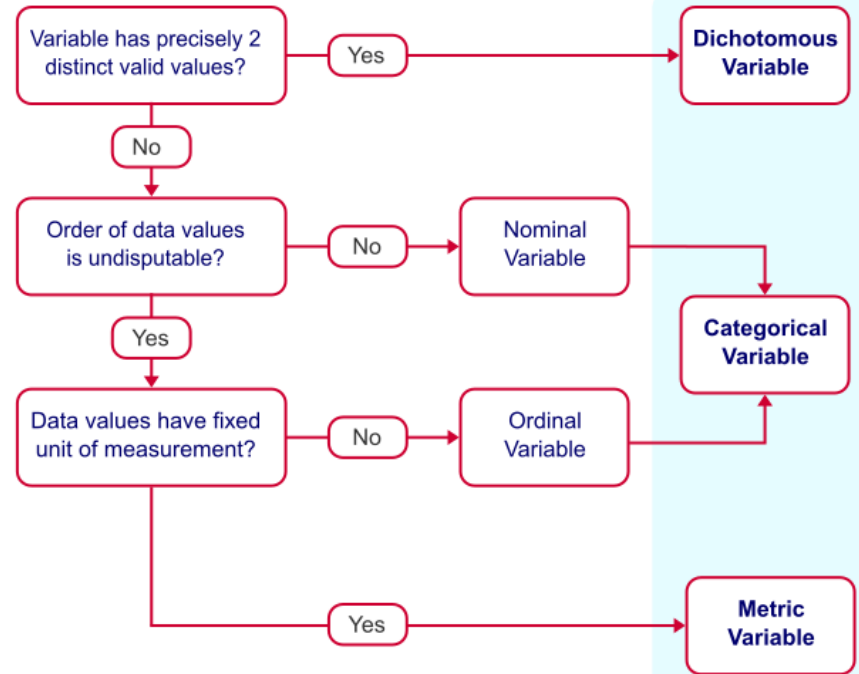
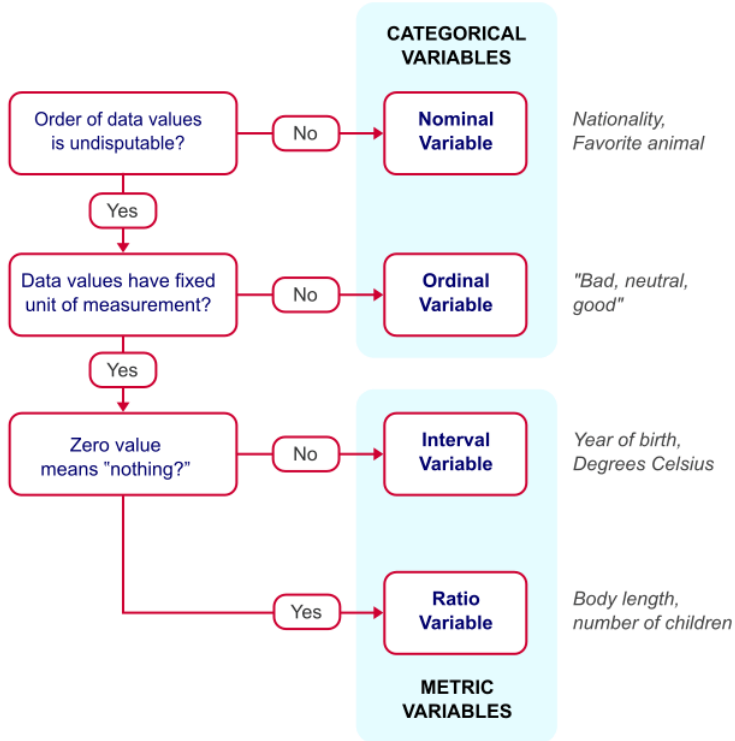
			Scale	Rationiveau
		Ordinaal niveau	Intervalniveau	Verhouding blijven gelijk
Nominale niveau	Ordering	Ordering	Gelijke verschillen	Gelijke verschillen
Onderscheid	Onderscheid	Onderscheid	Onderscheid	Onderscheid
<i>Geslacht</i>	<i>Opleidingsniveau</i>	<i>Intelligentie</i>	<i>Leeftijd</i>	<i>Leeftijd</i>

LEVELS OF DATA: LEVELS OF MEASUREMENTS/OBSERVATIONS

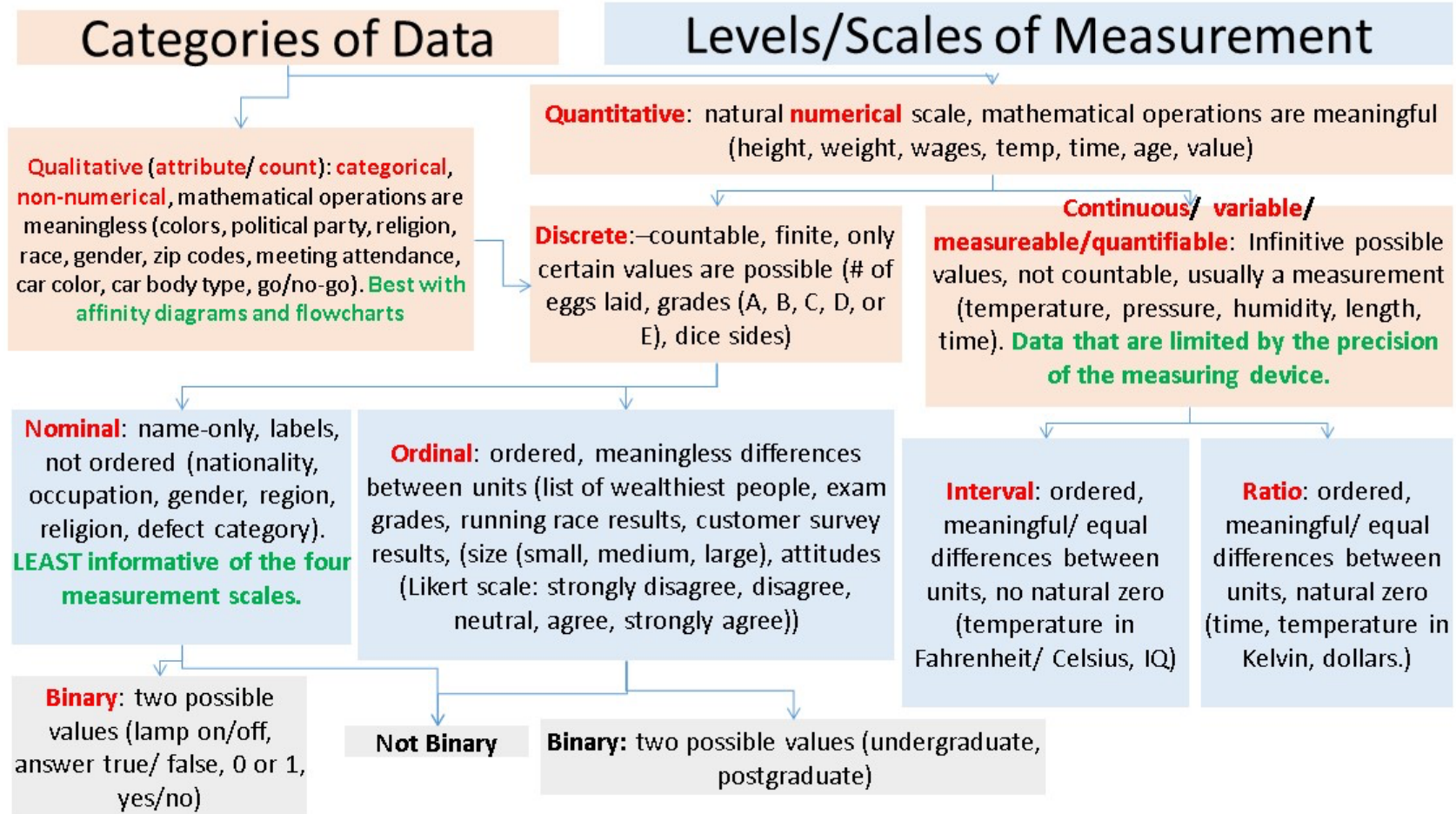
SCALE		EXAMPLE
Nominal		Gender
Ordinal		Position in race
Interval		Temperature (in Fahrenheit)
Ratio		Money



LEVELS OF DATA: What data to measure or observe?



LEVELS OF DATA: What data to measure or observe?

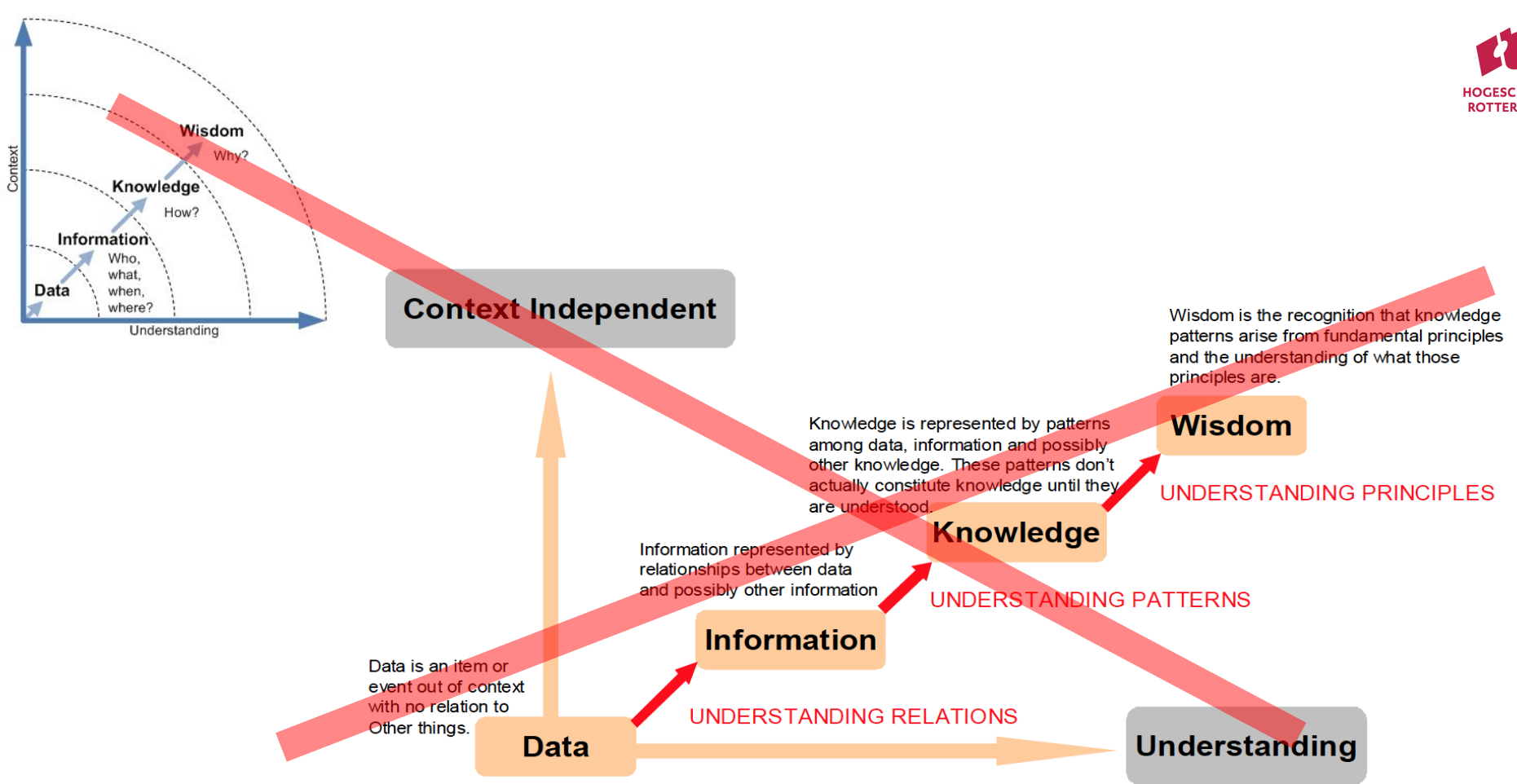


03

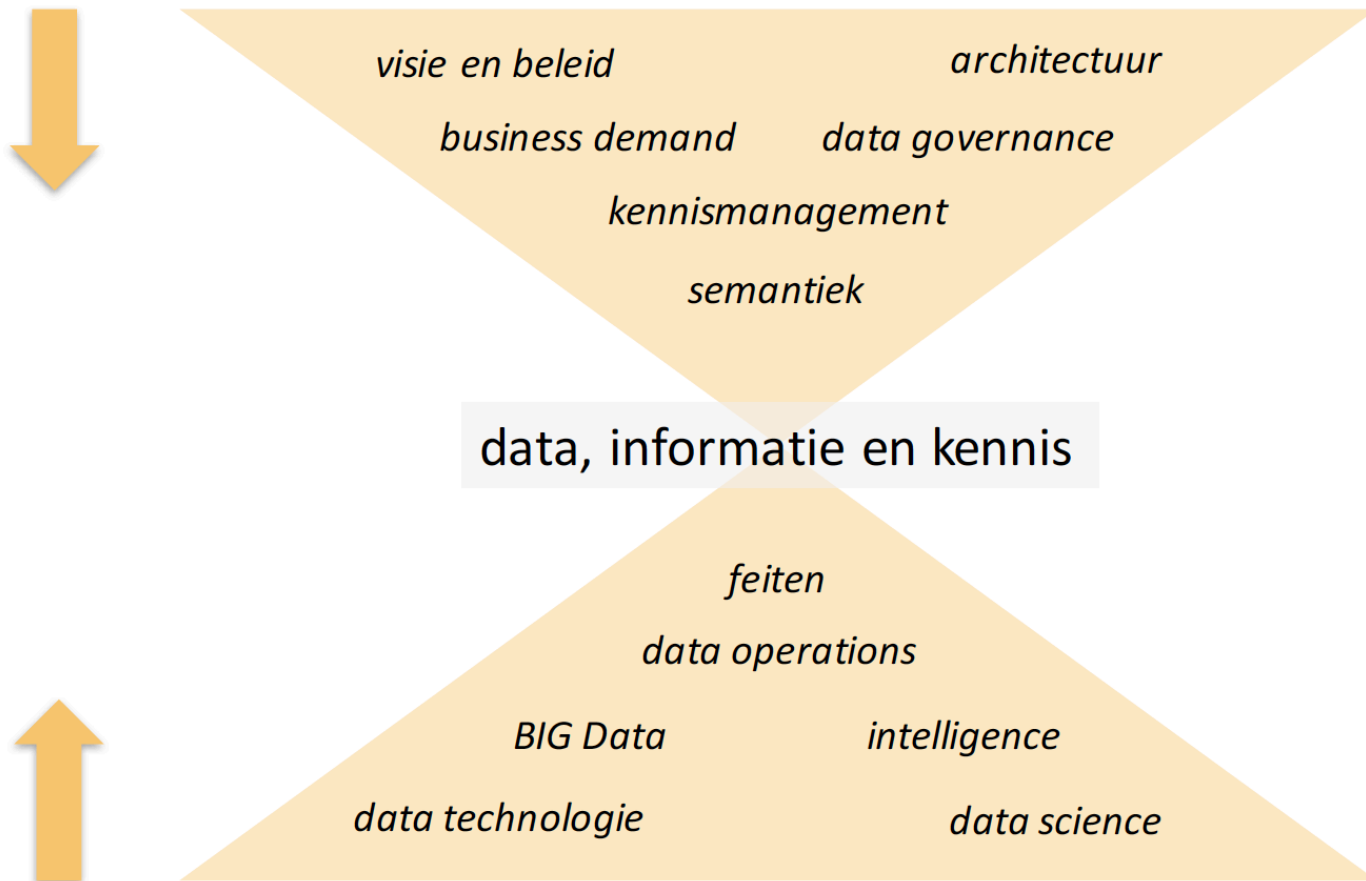
FUNDAMENTALS

Of

DATA + VISUALIZATION

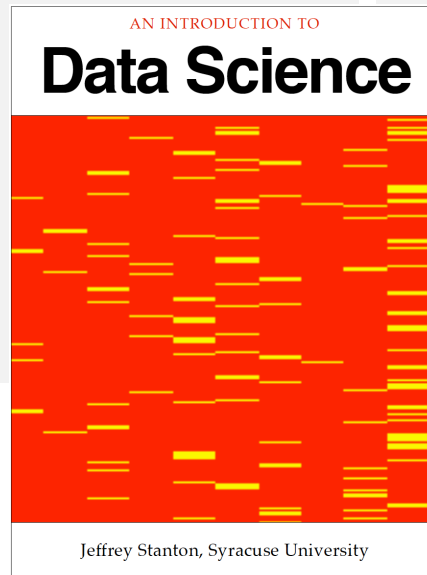


Data, informatie en kennis is wat ons verbindt



Data [gegevens]

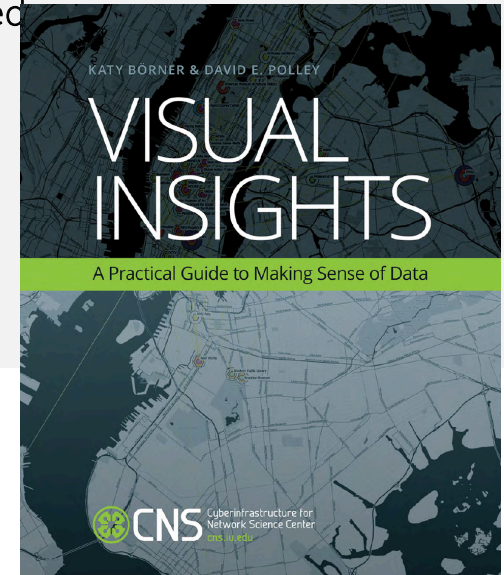
Raw Facts
No Context
Numbers
Symbols



Information

Data with structure = processed data
Value-added to Data

- Summarised
- Organised
- Analysed



TYPES OF DATA: Structured vs Unstructured [organized vs unorganized]

Data Structuring

Structured (organized) data: This is data that can be thought of as observations and characteristics. It is usually organized using a table method (rows and columns).

Unstructured (unorganized) data: This data exists as a free entity and does not follow any standard organization hierarchy.

Structured Data

High Degree of organization, such as a relational database

Column	Value
Patient	Joe Brown
Date of Birth	02/13/1972
Date Admitted	02/05/2014

Unstructured Data

Information that is difficult to organize using traditional mechanisms

“The patient came in complaining of chest pain, shortness of breath, and lingering headaches...smokes 2 packs a day... family history of heart disease...has been experiencing similar symptoms for the past 12 hours...”

DATA STRUCTURING: Generalized Form of a Data Table

Data Table [DATA MATRIX]

A generalized version of the data table is shown.

This table can represent any number of observations described over multiple variables.

This table describes a series of observations (from o_1 to o_n) where each observation is described using a series of variables (from x_1 to x_p). A value is provided for each variable of each observation.

	Variables					
Observations	x_1	x_2	x_3	\dots	x_p	
o_1	x_{11}	x_{12}	x_{13}	\dots	x_{1p}	
o_2	x_{21}	x_{22}	x_{23}	\dots	x_{2p}	
o_3	x_{31}	x_{32}	x_{33}	\dots	x_{3p}	
\dots	\dots	\dots	\dots	\dots	\dots	
o_n	x_{n1}	x_{n2}	x_{n3}	\dots	x_{np}	

Most data that exists in text form, including server logs and Facebook posts, is **unstructured**

Scientific observations, as recorded by careful scientists, are kept in a very neat and **organized (structured)** format: **THE DATA TABLE**

A **genetic sequence** of chemical nucleotides [ACGTATTGCA] is **unstructured** even if the order of the nucleotides matters

DATA STRUCTURING: Observations versus Variables

Data Table [DATA MATRIX]

A generalized version of the data table is shown.

This table can represent any number of **observations** described over multiple **variables**.

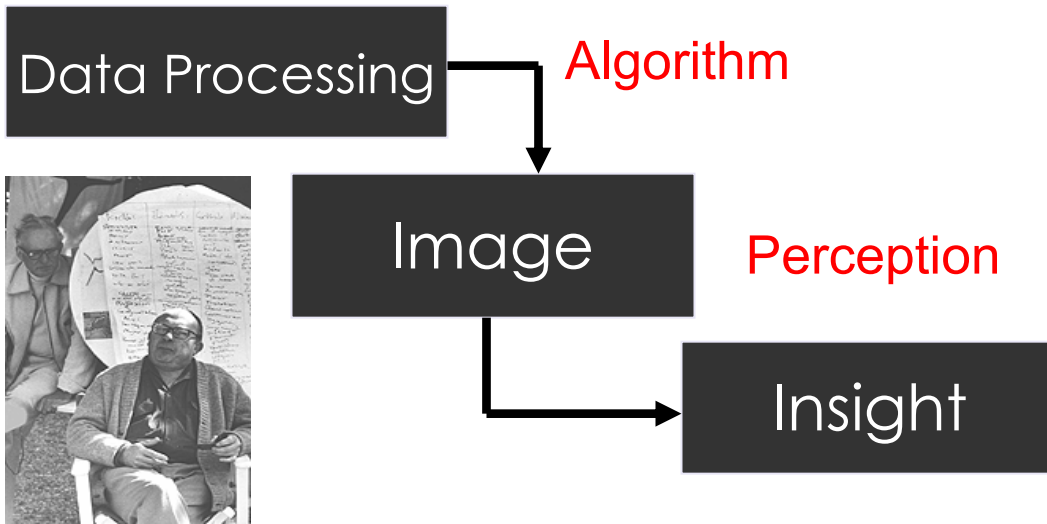
This table describes a series of observations (from o_1 to o_n) where each observation is described using a series of variables (from x_1 to x_p). A value is provided for each variable of each observation.

	Variables					
Observations	x_1	x_2	x_3	...	x_p	
o_1	x_{11}	x_{12}	x_{13}	...	x_{1p}	
o_2	x_{21}	x_{22}	x_{23}	...	x_{2p}	
o_3	x_{31}	x_{32}	x_{33}	...	x_{3p}	
...	
o_n	x_{n1}	x_{n2}	x_{n3}	...	x_{np}	

Patient ID	Treated	Age	Outcome	Random
1	Yes	Young	Positive	0.24
2	No	Young	Positive	0.85
3	Yes	Old	Negative	0.64
4	No	Old	Negative	0.70
5	No	Old	Negative	0.87
6	No	Old	Negative	0.72
7	No	Old	Negative	0.86
8	No	Young	Negative	0.16
9	No	Young	Positive	0.17

DATA STRUCTURING: data worden pas inzichtelijk als (beeld)figuur (graphical visualization) of GRAAF (graph)

Jacques Bertin who wrote the classic works of **graphical visualization** “*Semiology of Graphics*” states that the “*transformation from numbers to insight requires two stages*”



Practical Data Visualization

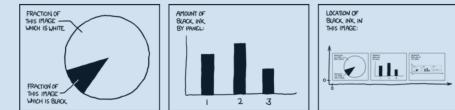
March 18, 2015

COMPSCI 216:
Everything Data



Angela Zoss
Data Visualization Coordinator
Data and Visualization Services

Communicating through infographics: visualizing scientific and engineering information

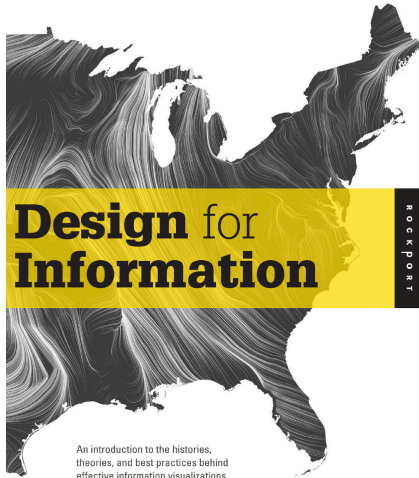


Christa Kelleher
Nicholas School of the Environment
Duke University

DATA STRUCTURING: Preattentive Processing [perceptual level]

THREE-STAGE MODEL OF PERCEPTUAL PROCESSING

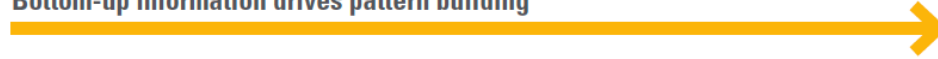
A schematic overview of the simplified information-processing model of human visual perception proposed by Collin Ware.¹⁴



An introduction to the histories, theories, and best practices behind effective information visualizations

Isabel Meirelles

Bottom-up information drives pattern building



Top-down attentional processes reinforce relevant information

STAGE 1

Billions of neurons work in parallel to extract millions of **features** that are processed rapidly and simultaneously, such as color, texture, orientation, and so on.

STAGE 2

Patterns are extracted serially and slowly, such as regions of the same color, and regions of the same texture. The pattern-finding process leads to two pathways: object perception, and locomotion and action.

STAGE 3

At the highest level of perception, we are able to hold between one and three **objects** at any instance in our working visual memory. Patterns that provide answers to the visual query construct the objects in conjunction with information stored in our long-term memory and that are related to the task at hand.

DATA STRUCTURING: Preattentive Processing [perceptual level]

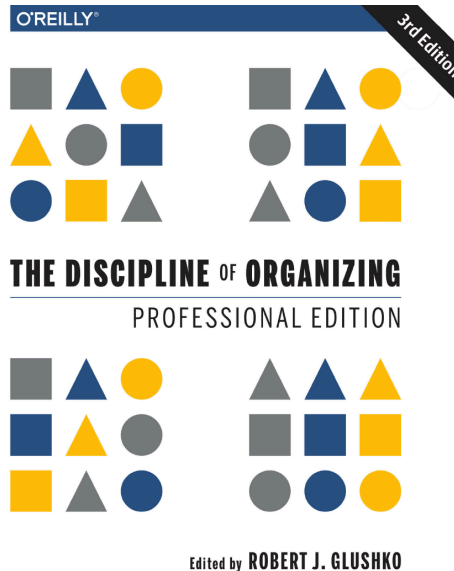
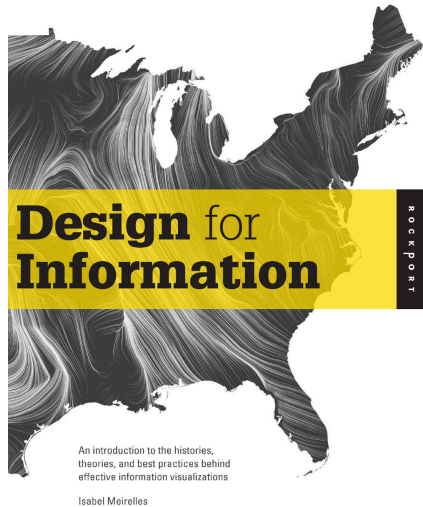
- Stage 1:** Rapid parallel processing to extract basic features;
- Stage 2:** Slow serial processing for extraction of patterns and structures;
- Stage 3:** Sequential goal-oriented processing with information reduced to a few objects and held in working visual memory to form the basis for visual thinking.

18596746321475030608030504090
70502769843010215346748950213
06057204020503090845064201040
70204070835061305080239245798

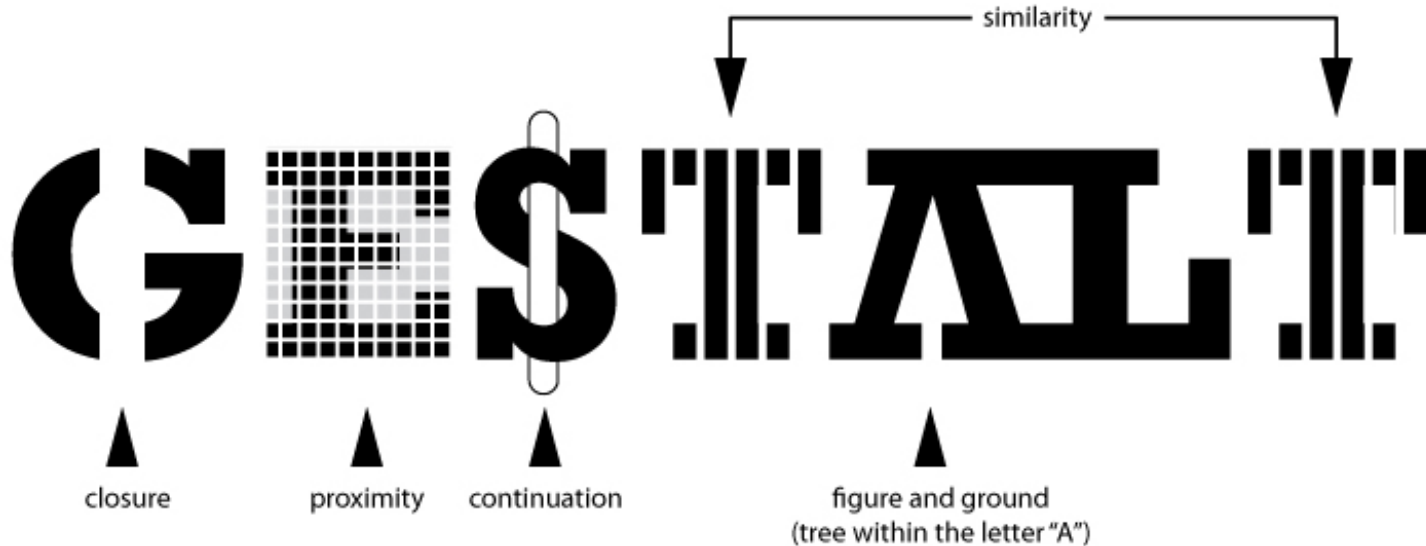
18596746321475030608030504090
70502769843010215346748950213
06057204020503090845064201040
70204070835061305080239245798

18596746321475030608030504090
70502769843010215346748950213
06057204020503090845064201040
70204070835061305080239245798

18596746321475030608030504090
70502769843010215346748950213
06057204020503090845064201040
70204070835061305080239245798



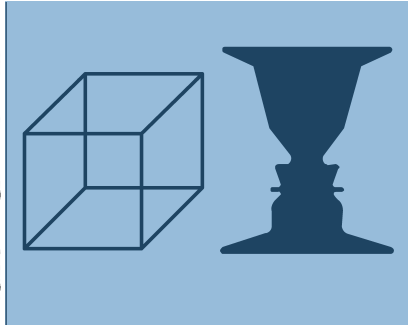
Gestalt psychologie classificeert perceptie volgens (top-down) groepering principes



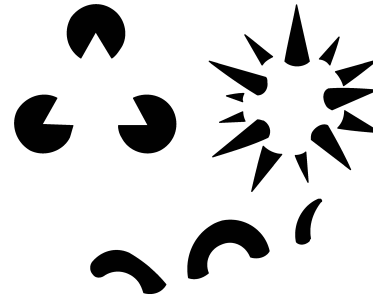
Gestalt classificeert Perceptie



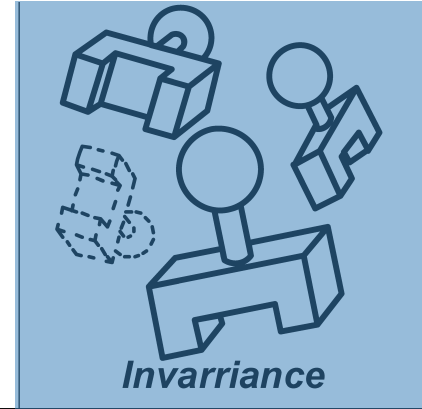
Emergence



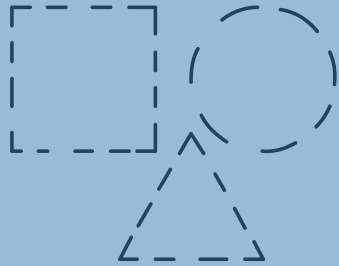
Multistability
Figure/Background selection



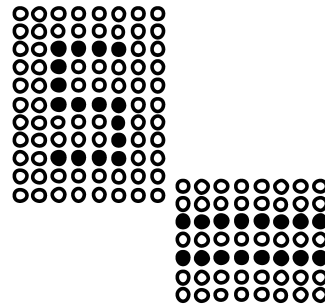
Reification
Illusory contours



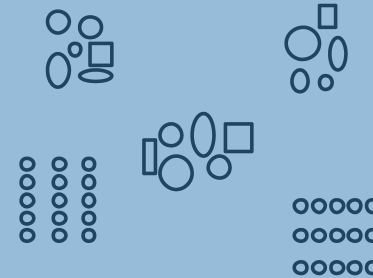
Invariance



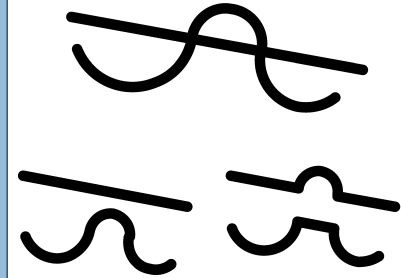
Closure



Similarity



Proximity



We see this...but not this
Continuity

Gestalt is geen PoP-out fenomeen



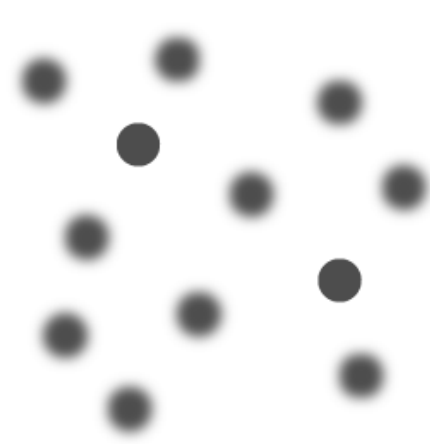
Color



Size

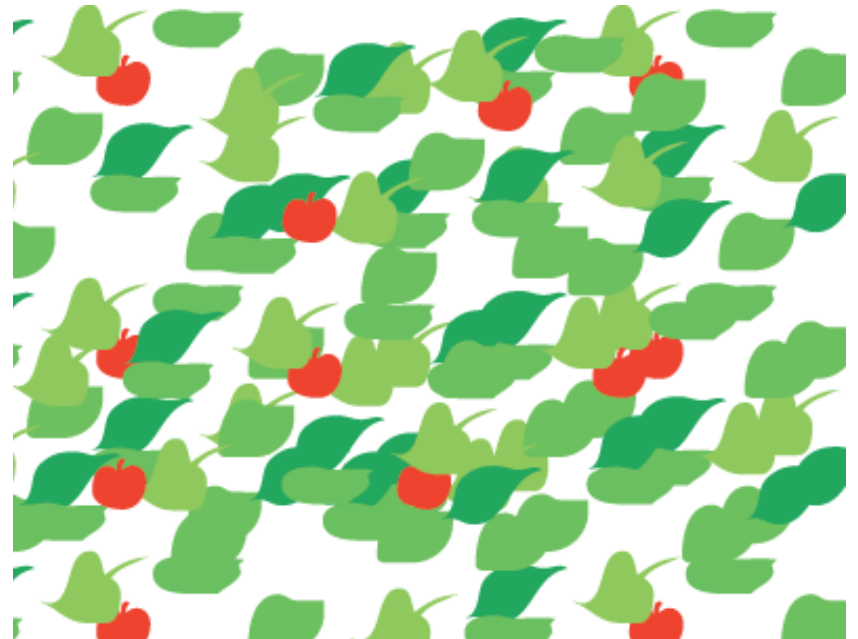
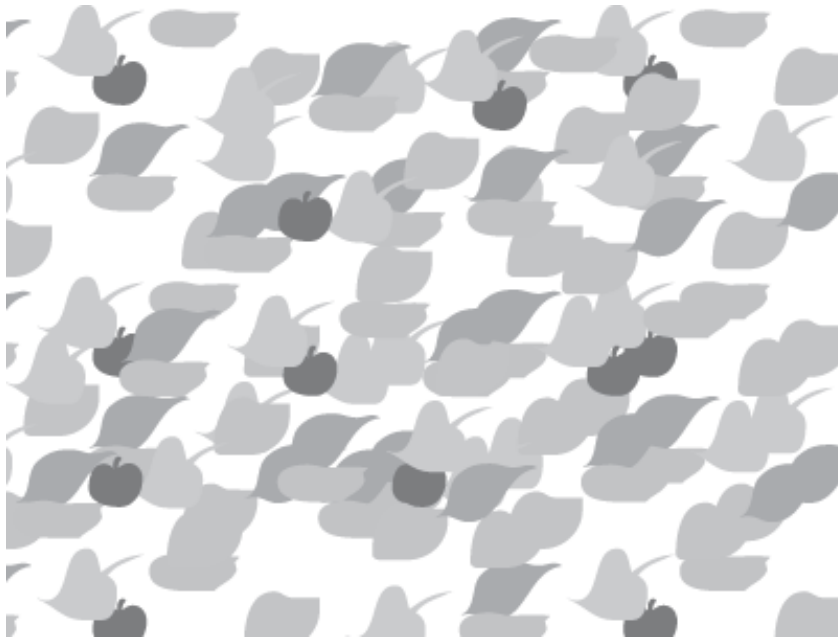


Curvature

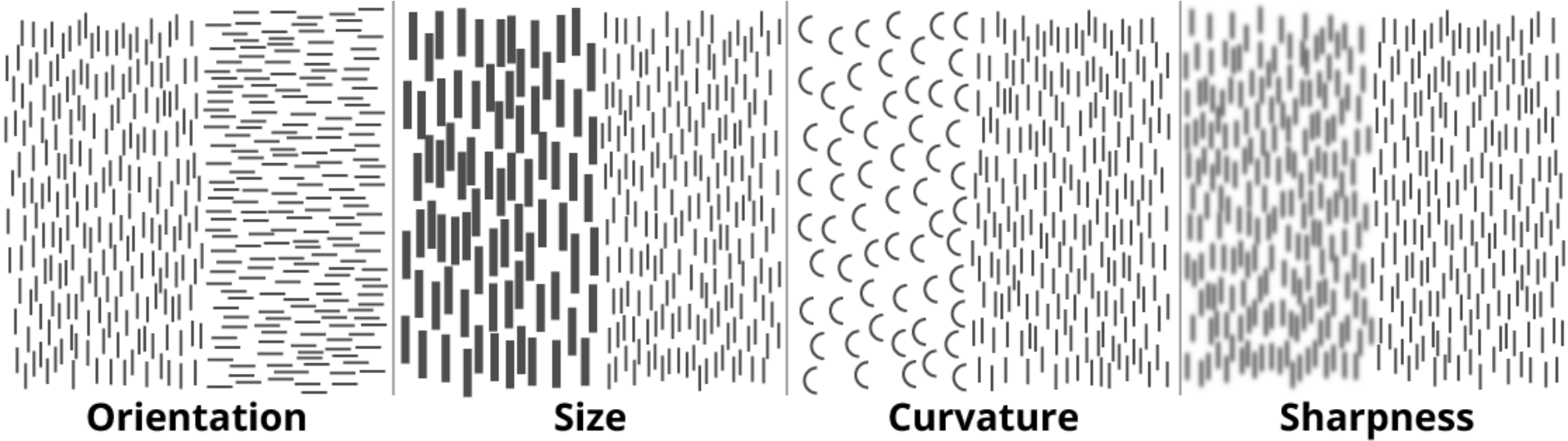


Sharpness

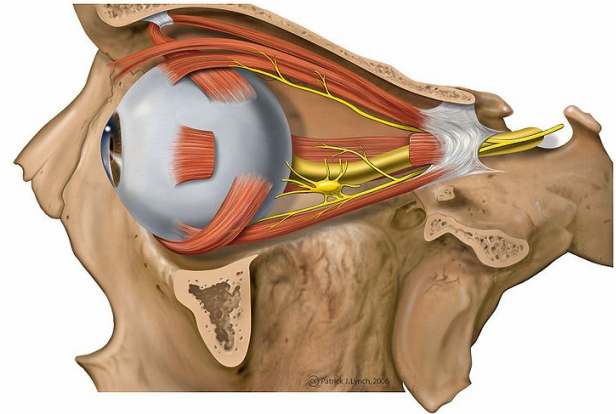
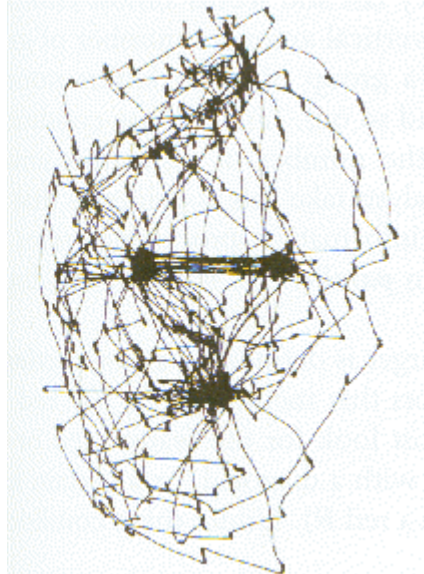
Gestalt is geen chromatisch fenomeen



Gestalt is geen Textuur fenomeen



Gestalt is een passief fenomeen



Gestaltprincipes zijn een bijzondere vorm van passieve visuele waarneming ze vereisen dan ook geen oogbewegingen

Gestalt & Design



1971



1987



1992



2011



1961



1978



1986



WWF®

WWF®



1891-1900



1900-1934



1934-1970



1970-1986



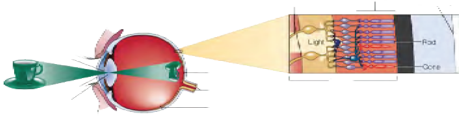
1986-2002



2002-2004



2004-present

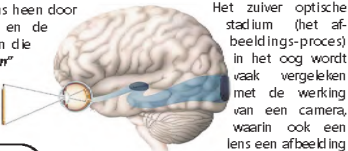


1 Paradox van het Zien

Zien is meer dan fotos maken

Het oog als camera

Wé zien de wereld om ons heen door middel van onze ogen en de daarmee verbonden delen die tezamen het **"visuele brein"** vormen. Het proces van de visuele gewaarwording delen we op in een aantal stadia.



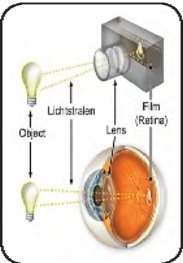
Het zuiver optische stadium (het afbeeldings-proces) in het oog wordt vaak vergeleken met de werking van een camera, waarin ook een lens een afbeelding vormt op een licht-gevoelige filmplaat.

Hoewel deze vergelijking onjuist is, leidt zij wel tot zinvolle vragen.

Hoe is het mogelijk dat we met ons oog scherp en "recht op staand" zien?

Bij het maken van een foto moeten we de camera zoveel mogelijk stil houden. Wanneer niet de camera, maar het gefotografeerde object beweegt, wordt de foto onscherp. Ook wordt het object "op zijn kop" afgebeeld. Ten slotte moet de belichting stijf zeer kort zijn.

Het oog beweegt in het hoofd, het hoofd beweegt op ons lichaam, dat zich weer verplaatst in de ruimte. Het netvlies is continu belicht. Toch hoefje niet muilstil te zitten om goed te kunnen zien!



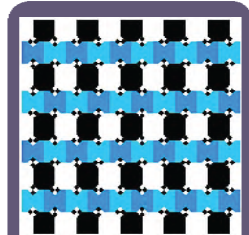
Het eerste stadium is zuiver **"optisch"**: het vormen van een scherpe afbeelding door de lens van het oog op het netvlies (**retina**), dat de lichtgevoelige cellen bevat. Volgende stadia zijn het omzetten van een afbeelding in zenuwsignalen en het verwerken van deze signalen door de hersenen.

Zien

We nemen onze omgeving scherp waar terwijl we lopen, fietsen of autorijden, en ook bewegende objecten kunnen we scherp zien. Terwijl retinale-afbeeldingen juist omgekeerd, instabiel en plat (tweedimensionaal, 2D) zijn.

In de wereld om ons heen vinden we aanwijzingen dat dieren (inclusief primaten zoals wijzelf) niet reageren op de afbeeldingen in hun ogen, maar op een "brein-veranderende-versie" ervan.

Wanneer de ogen van een kikker operatief worden omgedraaid t.o.w. zijn kop, dan kan hij geen vliegen meer vangen. Maar, na een paar dagen is de kikker hersteld. Dus het kikker-brein maakt dat wat de ogen aan beelden binnen halen geschikt om er iets mee te kunnen doen in de "buitenwereld".



Bij het kijken naar het "Blokpatroon", zoals hierboven afgebeeld, zullen velen de indruk hebben dat de horizontale blauwe balken scheef lopen.

Er is een prijs die we betalen voor deze "Brein-veranderde-versie". Zien is "niet natuurgetrouw". Dit fenomeen kennen we als "Gezichtsberog".

In dit cahier zullen we stap voor stap na-gaan hoe wetenschappers / kunstenaars "gezichtsbedrog" bestudeerd hebben. Beide concluderen dat ons brein zich gedraagt als een **"verhalen verteller"**.

Gezichtsbedrog reflecteert ons vermogen om te komen tot een (be)grijpbare realiteit.

Gezichtsbedrog wordt vaak omschreven als *"onverwachte valkuilen van het zien."*

Het zijn echter *"ogenschijnlijke"* weeffouten die alleen kunnen bestaan in de visuele ruimte van onze hersenen als gevolg meerduidigheid of incompleetheit van de zintuigelijke informatie.

Ons visuele brein kiest voor de meest voor hand liggende interpretatie ---of vult zelf aan--- door gebruik te maken van ingebouwde "kennis" in ons brein over de tastbare wereld om ons heen.

Het in kaart brengen van gezichtsbedrog geeft inzicht in hoe mensen waarnemen.

In de wetenschap weerspiegelt het "experiment" het doel van de onderzoeker om oorzaak en gevolg vast te leggen, en zo een "hypothese" (een veronderstelling) te kunnen testen door deze te aanvaarden of te verwerpen.

Tekenen van wat we om ons heen zien vormt de oudste methode om waarnemingen vast te leggen.

De waarnemer verklaart:

"Ik keek en dit is wat ik zag."

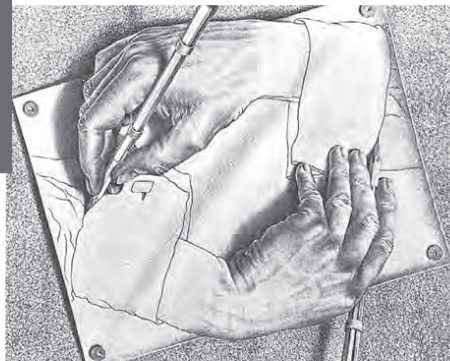
Zoals de grottekeningen van de Cro-Machnon (zie blauwe kader).



Prehistorische grottekeningen van Lascaux in Zuid-Frankrijk tonen kleurrijke afbeeldingen van dieren, die meer dan 15.000 jaar geleden werden gecreëerd.

Zien is Keuzes Maken

Hypothese Testen



Als een hand een potlood vasthoudt, en als een tweede hand ook een potlood vasthoudt ... en als dit afgebeeld wordt op een stukje papier, wat zien we dan?

De litho "Tekenen" (1948) van de graficus Escher, zoals weergegeven in het midden van deze pagina, doet ons geloven dat twee handen "elkaar tekenen".

De hypothese van twee zichzelf tekenende handen staat weliswaar niet los van het beeld op ons netvlies maar ze botst wel met de "natuurwet" die ons leert dat handen vast zitten aan een lichaam en het is dat lichaam wat maakt dat die handen tekenen.

Het brein interpreteert en maakt zo een keuze. Het vertelt een verhaal vanuit een bepaald uitgangspunt of hypothese: Een gedachte waar niet aan getwijfeld wordt.

We kunnen nu tevreden de armen over elkaar slaan en het verband tussen de retinale afbeelding van de ogen (oorzaak) en de "brein-veranderende-versie" ervan (gevolg) als opgelost beschouwen: Er moet hier sprake zijn van "gezichtsbedrog".

Deze hypothese is onjuist. Het was Escher's bedoeling ons te bedriegen. Toch zal elk kind van een jaar of 3-4 zich niet voor de gek laten houden; simpelweg omdat het jonge brein niet de verbanden "kent" die de meeste volwassenen wel zouden "zien".

Zo bezien is het maken van een tekening geen betrouwbare methode om oorzaak en gevolg vast te leggen.

Gezichtsbedrog classificatie

Classes

Example

Description

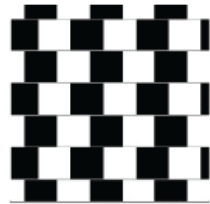
Ambiguity



Information is insufficient to result in a single interpretation.

Rubin's figure can be perceived either as a vase (black) or as two face-to-face characters (white).


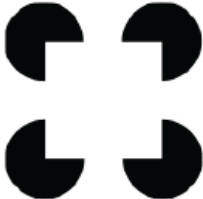
Distortion



















The visual context induces a distortion in size, contrast, motion or disposition appreciation.

In the Café Wall illusion, the lines, although parallel, appear to be convergent or divergent.









Gezichtsbedrog classificatie



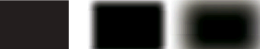


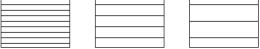

Classes	Example	Description
Paradox		<p>The figure appears to be an impossible object when viewed from a critical position.</p> <p>The Penrose triangle introduces a “mise en abyme,” which makes the figure implausible.</p>
Fiction		<p>The observer perceives visual elements absent in the figure because of the context.</p> <p>The Kanizsa square’s contour is reconstructed by the perceptual system.</p>

DATA STRUCTURING: Bertin's 7 Visual Variables (1967)

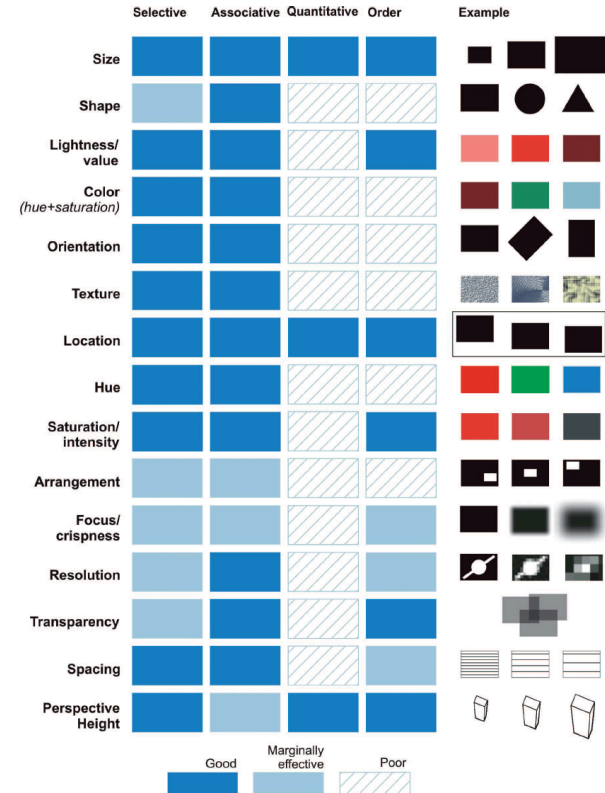
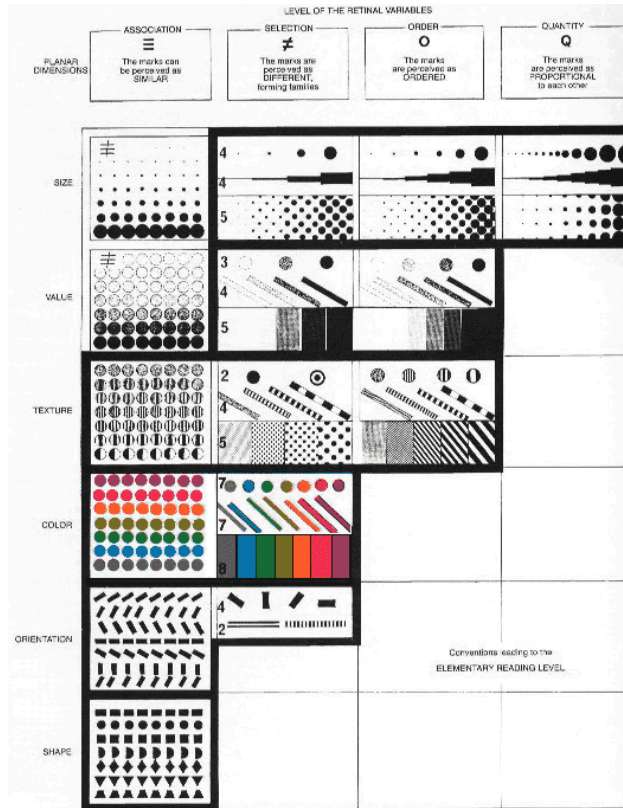
		Characteristics				
		<i>Selective</i>	<i>Associative</i>	<i>Quantitative</i>	<i>Order</i>	<i>Length</i>
Visual Variables	<i>Position</i>					Theoretically Infinite
	<i>Size</i>					Selection: ~5 Distinction: ~20
	<i>Shape</i>					Theoretically Infinite
	<i>Value</i>					Selection: <7 Distinction: ~10
	<i>Color</i>					Selection: <7 Distinction: ~10
	<i>Orientation</i>					Theoretically Infinite
	<i>Texture</i>					Theoretically Infinite

DATA STRUCTURING: Visual Variables

Visual Variable	Author	Example
Size	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Shape	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Lightness/ value	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Color (hue+saturation)	Bertin (1967/83).	
Orientation	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Texture	Bertin (1967/83), Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Tyner (2010).	
Location	Bertin (1967/83), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Hue	Morrison (1974), MacEachren (1995), Kraak & Ormeling (2003), Krygier & Wood (2005), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	

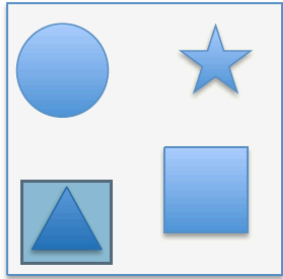
Visual Variable	Author	Example
Saturation/ intensity	Morrison (1974), MacEachren (1995), Krygier & Wood (2005), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Arrangement	Morrison (1974), MacEachren (1995), Dent et al. (2009), Slocum et al. (2010), Tyner (2010).	
Focus/ crispness	MacEachren (1995).	
Resolution	MacEachren (1995).	
Transparency	MacEachren (1995).	
Spacing	Slocum et al (2010).	
Perspective Height	Slocum et al (2010).	

DATA STRUCTURING: Visual Variables

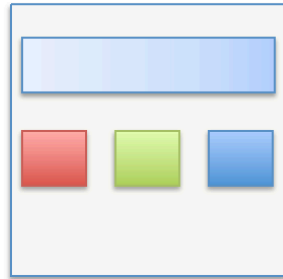


SEE ALSO: http://www.iag-aig.org/attach/30dee1f85f7bd479367f1f933d48b701/V61N1_2FT.pdf
<http://www3.sympatico.ca/blevis/thesis49prev.html>

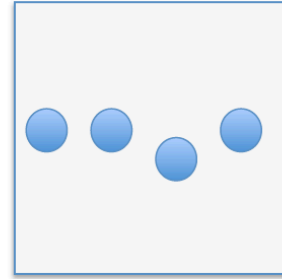
DATA STRUCTURING: Preattentive Processing [perceptual level]



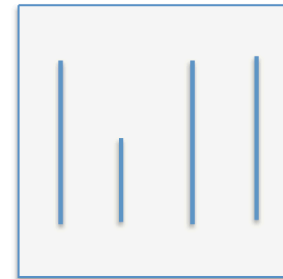
Shape



Color



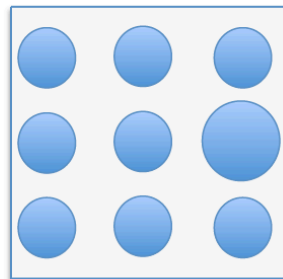
Position



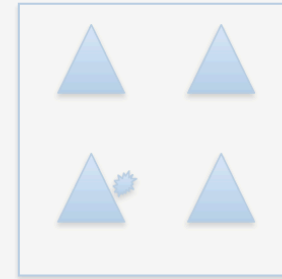
Line Length



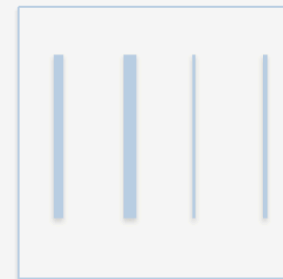
Orientation



Size



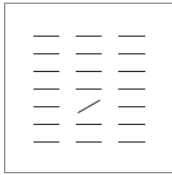
Added Marks



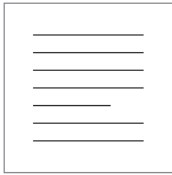
Line Width

DATA STRUCTURING: Preattentive Processing [data type]

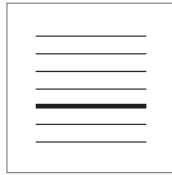
LINE ORIENTATION



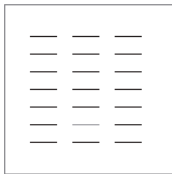
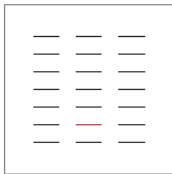
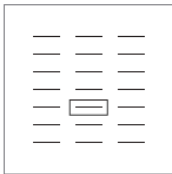
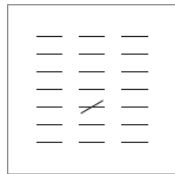
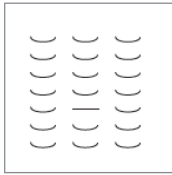
LINE LENGTH



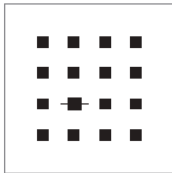
LINE WEIGHT



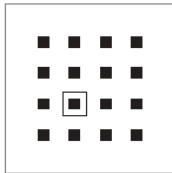
CURVATURE



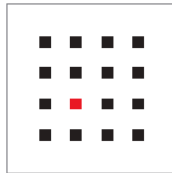
ADDED MARKS



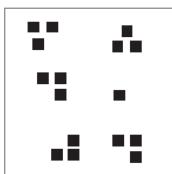
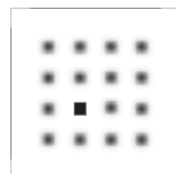
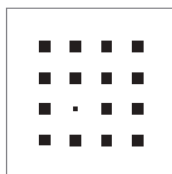
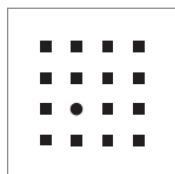
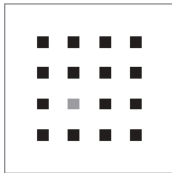
ENCLOSURE



COLOR/HUE



INTENSITY/VALUE



SHAPE

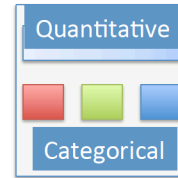
SIZE

SHARPNESS

NUMEROSITY



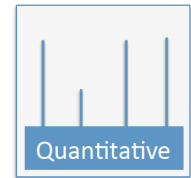
Shape



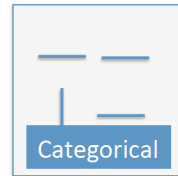
Color



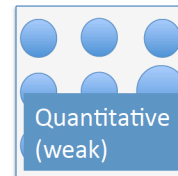
Position



Line Length



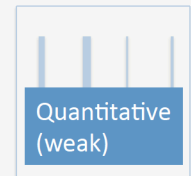
Orientation



Size



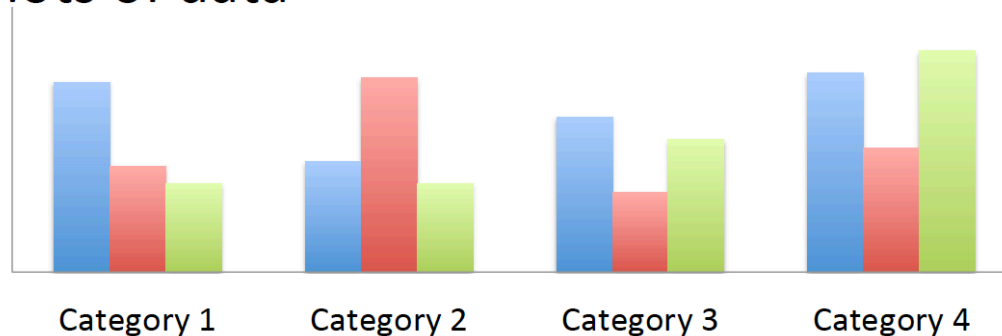
Added Marks



Line Width

- Tables:
 - Accuracy
 - Lookup
- Charts
 - Story
 - Summarize lots of data

Lender	Adjustable	Fixed
Bank 1	7%	5.25%
Bank 2	7.25%	5.5%



Hoofdstuk 12: Grafieken en tabellen

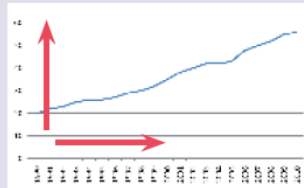
Breng cijfers beter in beeld

Grote hoeveelheden cijfers weer-
geven met behulp van grafieken en
tabellen biedt allerlei voordelen:
abstracte data worden samengevat
in een beeld of een lijst, en tekst
en data kunnen van elkaar worden
gescheiden. Een plaatje blijft beter
hangen dan een reeks cijfers. En
vergeet ook niet de retorische func-
tie van grafieken en tabellen.

KADER

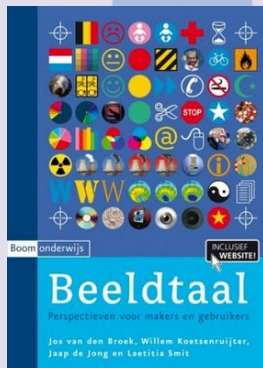
Semiotiek en de lijngrafiek

Traditioneel leest iedereen in de westerse wereld lijngrafieken zó, dat omhoog doorgaans iets positiefs betekent. Zonder enige toevoeging weten we dat het goed gaat met de zaken als we dit zien:

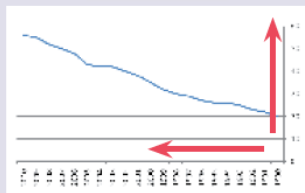


← Ook zonder as-titels associëren we deze lijn met iets positiefs (al zou het ook kunnen gaan om de stijging van het aantal diefstallen). (Fig. 12.8)

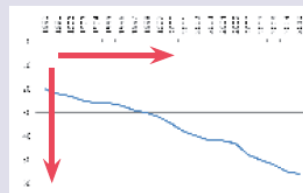
Die waarneming heeft met semiotiek te maken. We beheersen een tekensysteem waarbinnen we hebben afgesproken dat rechts omhoog iets positiefs aanduidt. Naar rechts is vooruit, en omhoog is meer. Iedere afwijking van dit patroon levert vanwege een verkeerde betekenisassociatie verwarring op bij de kijker. Kijk maar. Hieronder is nog drie keer dezelfde dataset weergegeven. Merk je hoe verwarrend dat werkt? Zo sterk zijn grafiekconventies. Wijk er alleen vanaf als je verwarring wilt stichten. ◀



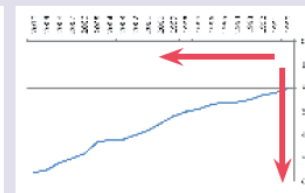
Drie grafieken waarbij de assenstelsels zijn omgedraaid.



↑ Richting x-as omgewisseld.
(Fig. 12.9)



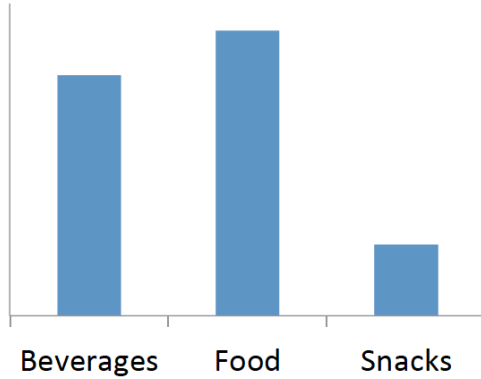
↑ Richting y-as omgewisseld.
(Fig. 12.10)



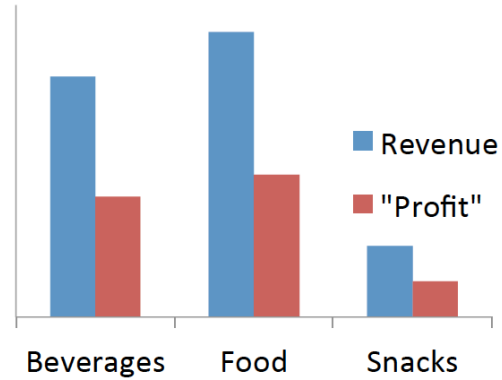
↑ Richting van x- én y-as omgewisseld. (Fig. 12.11)

DATA STRUCTURING: Relationships & Patterns [applied: CHARTS]

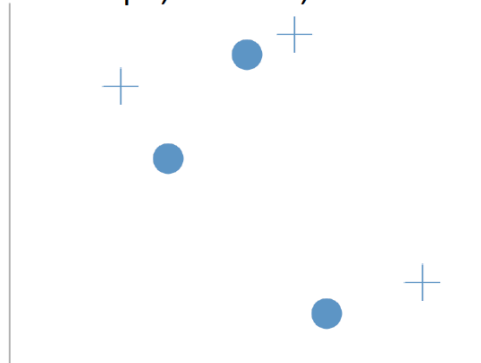
Length, Position



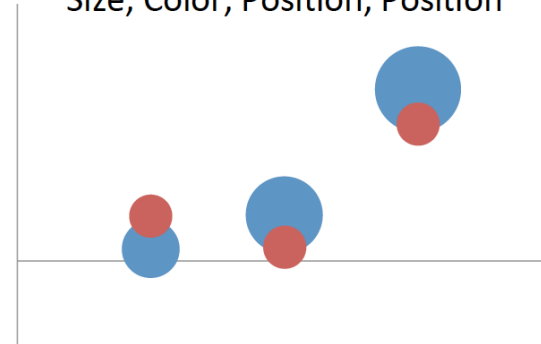
Length, Position, Color



Shape, Position, Position

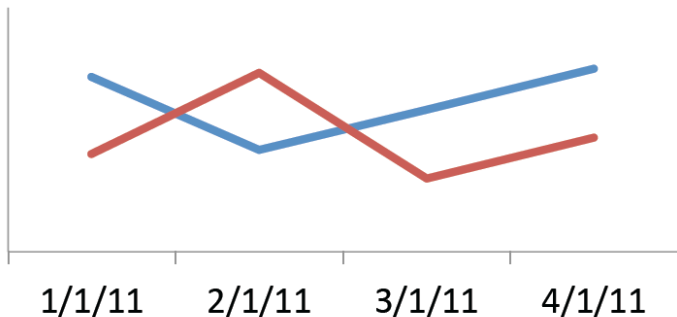


Size, Color, Position, Position

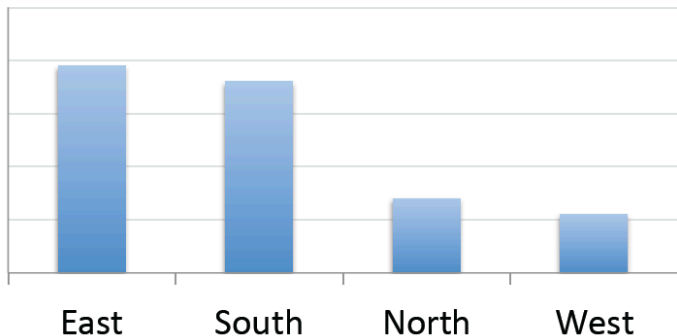


DATA STRUCTURING: Relationships & Patterns [applied: CHARTS]

Time Series



Part-to-whole



Numerical
Made of numbers
Age, weight, number of children, shoe size

Continuous
Infinite options
Age, weight, blood pressure

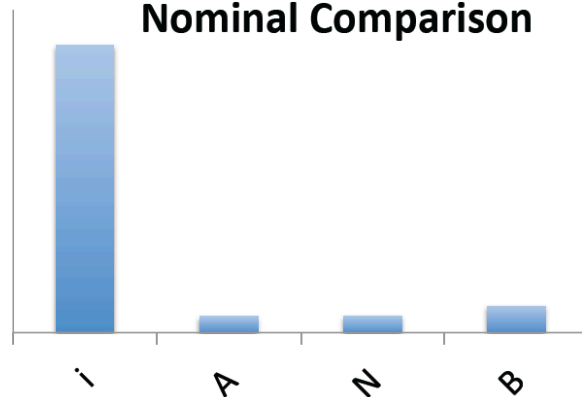
Discrete
Finite options
Shoe size, number of children

Categorical
Made of words
Eye colour, gender, blood type, ethnicity

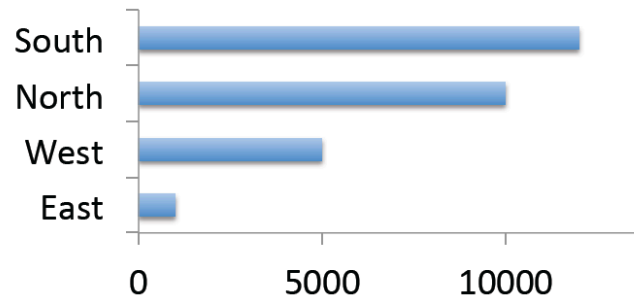
Ordinal
Data has a hierarchy
Pain severity, satisfaction rating, mood

Nominal
Data has no hierarchy
Eye colour, dog breed, blood type

Nominal Comparison



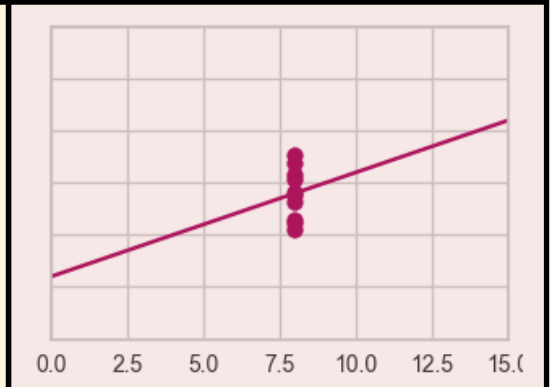
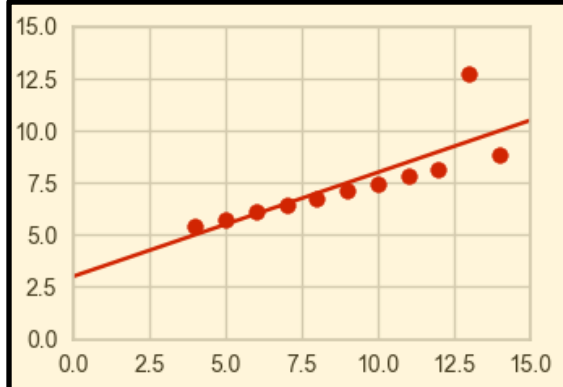
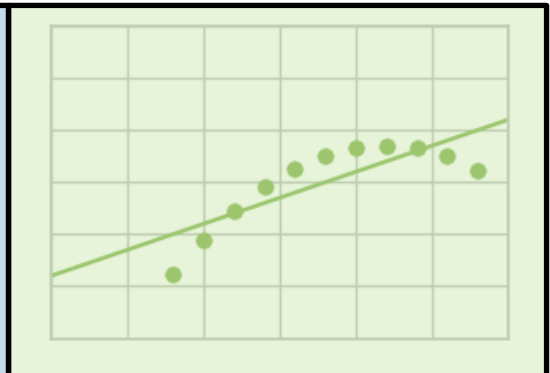
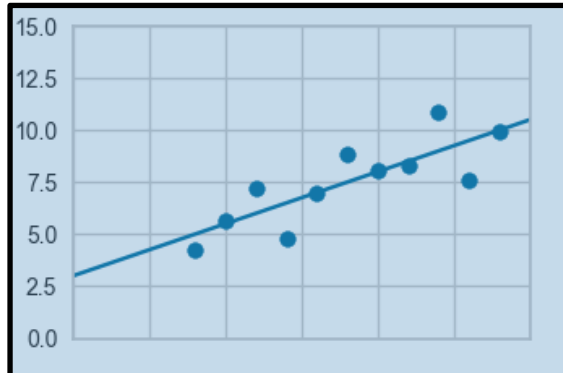
Ranking



DATA STRUCTURING: Relationships & Patterns [Anscombe Quartet]

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

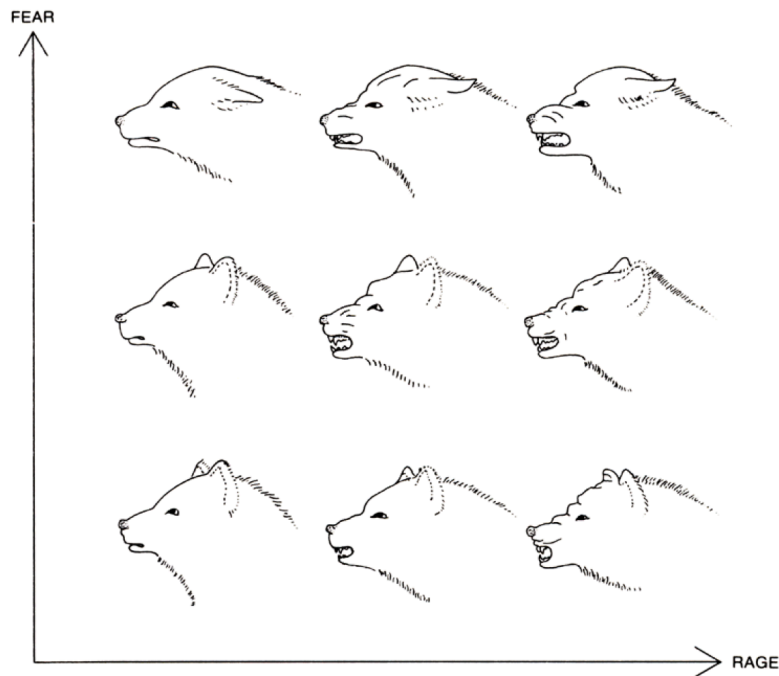


when the image is the data

the visual medium is ideal
for depicting multivariate
data

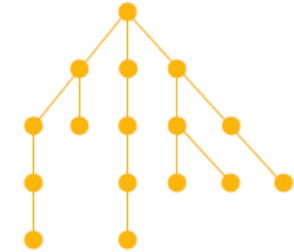
arguably univariate and
bivariate data should be
tabularized, within
reason

this example shows a plot
for a case where data
cannot be easily
parametrized



DATA STRUCTURING: Relationships & Patterns [HIERARCHIES]

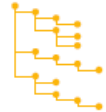
CARTESIAN SYSTEMS



node-link layout



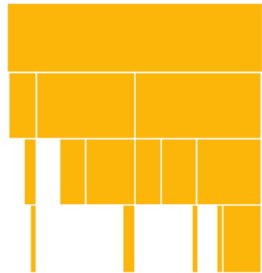
dendrogram



indented layout



cone-tree



icicle tree



treemap



Graph Drawing

4 Major tree visualizations

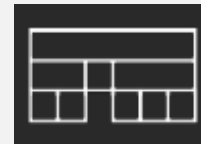
Indented lists



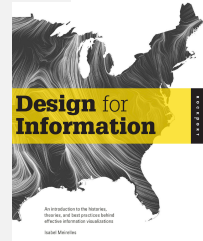
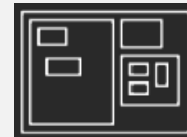
Node-link trees



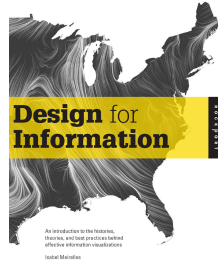
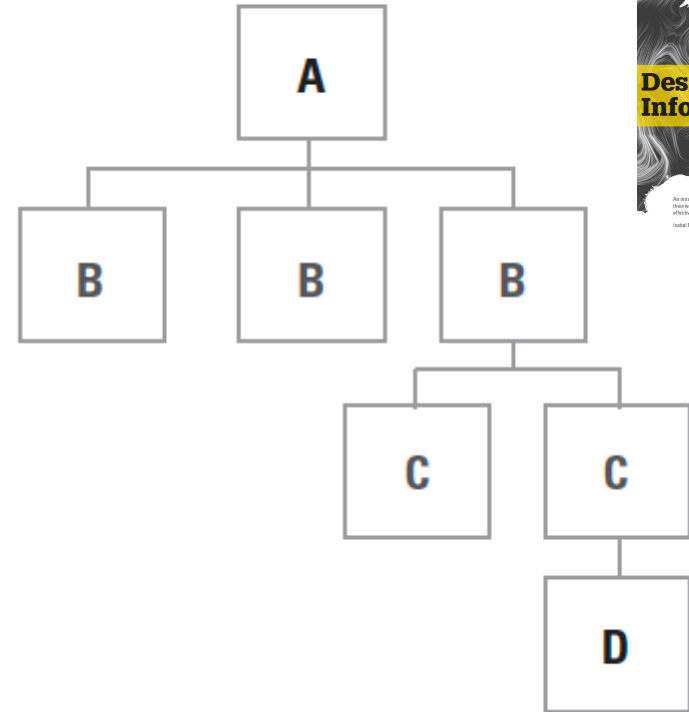
Layered diagrams



Treemaps



DATA STRUCTURING: Graph (graaf) [treemap versus intended list]



Wat is een graaf (graph)?



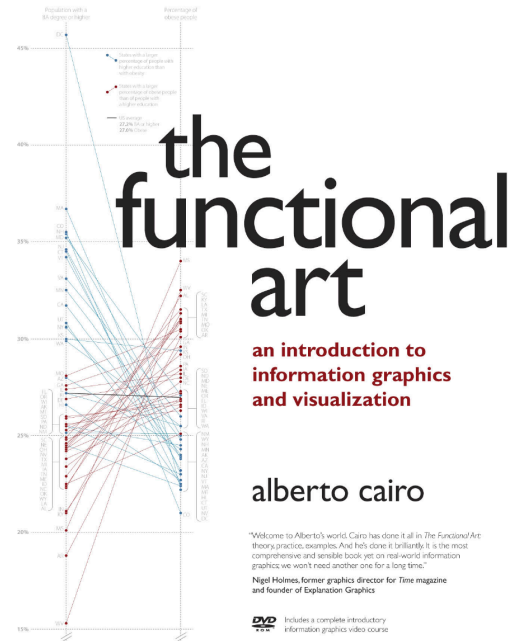
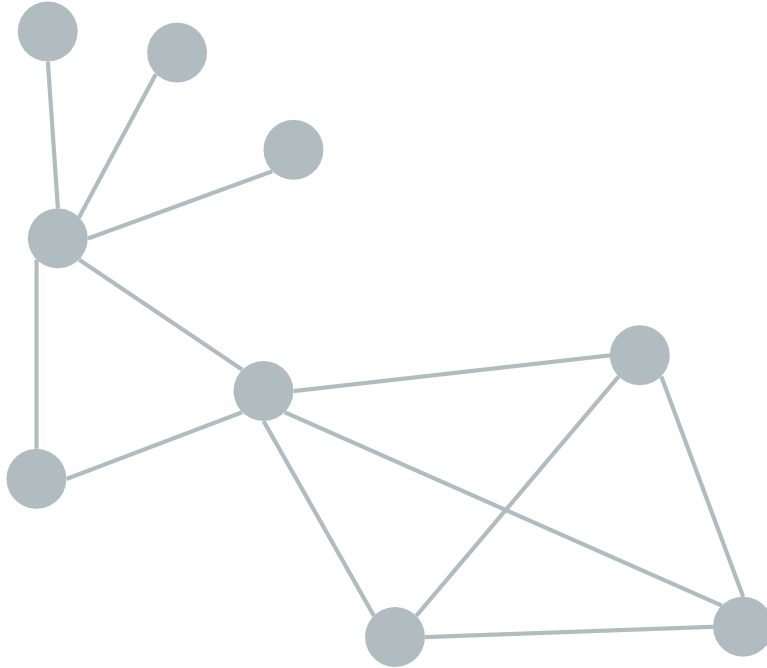
Graph Drawing

The primary concern of graph drawing is the spatial arrangement of nodes and links

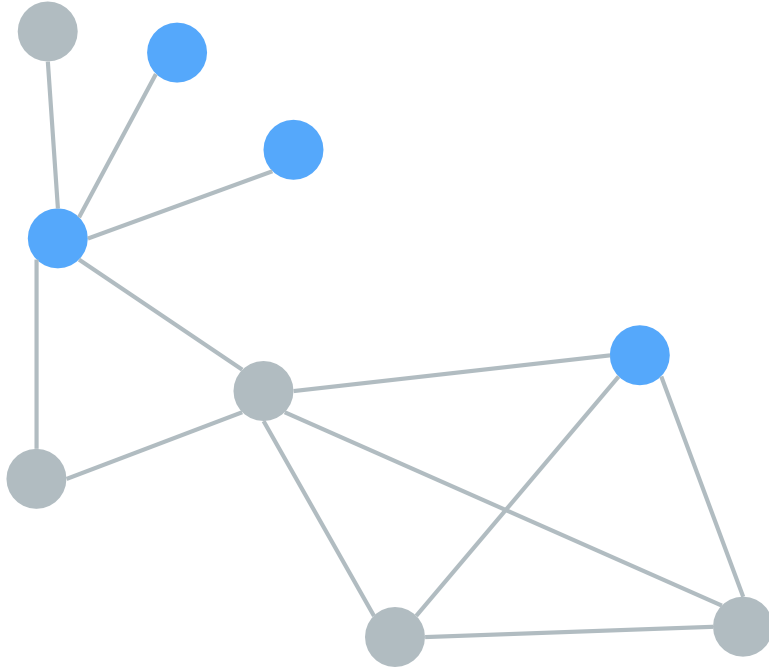
Often (but not always) the goal is to effectively depict the graph structure:

- Connectivity patterns
- Partitions / Clusters
- Outliers

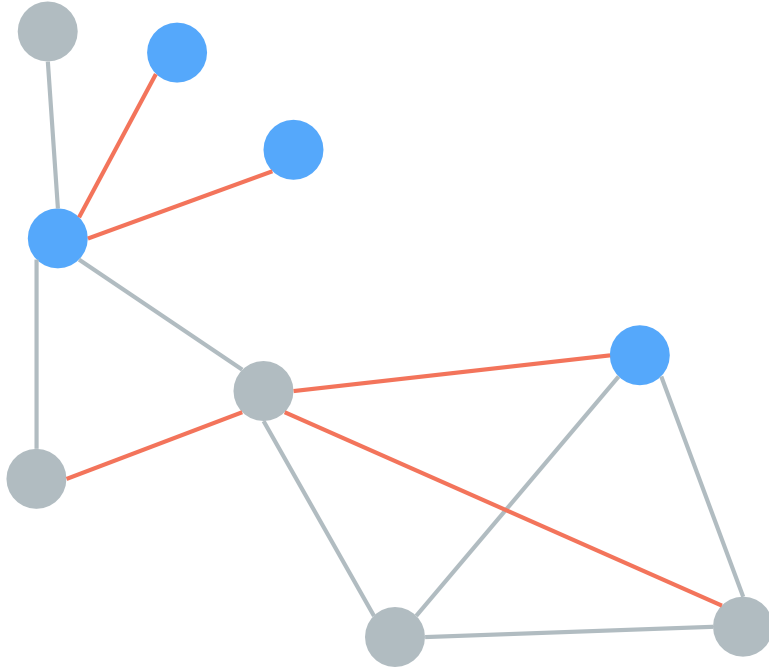
Wat is een graaf (graph)?



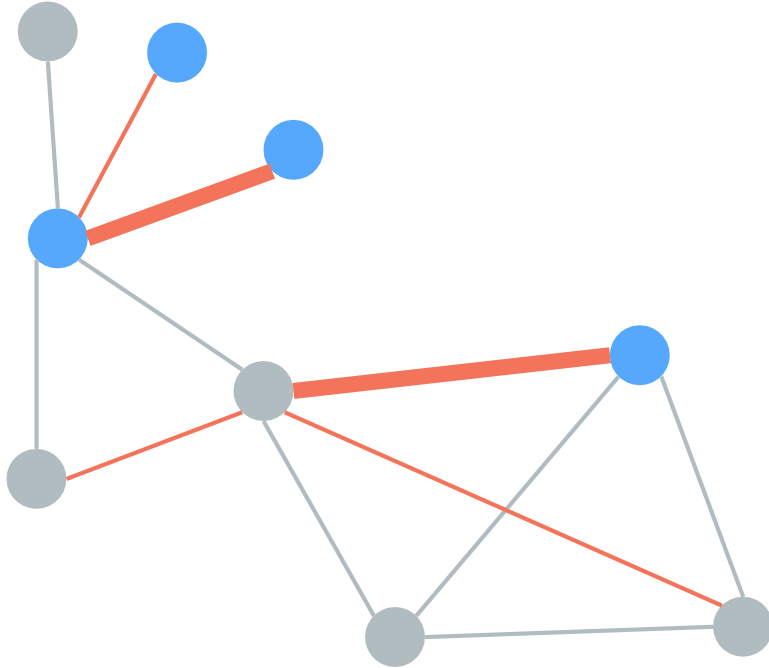
Wat is een **graaf**
(graph)?



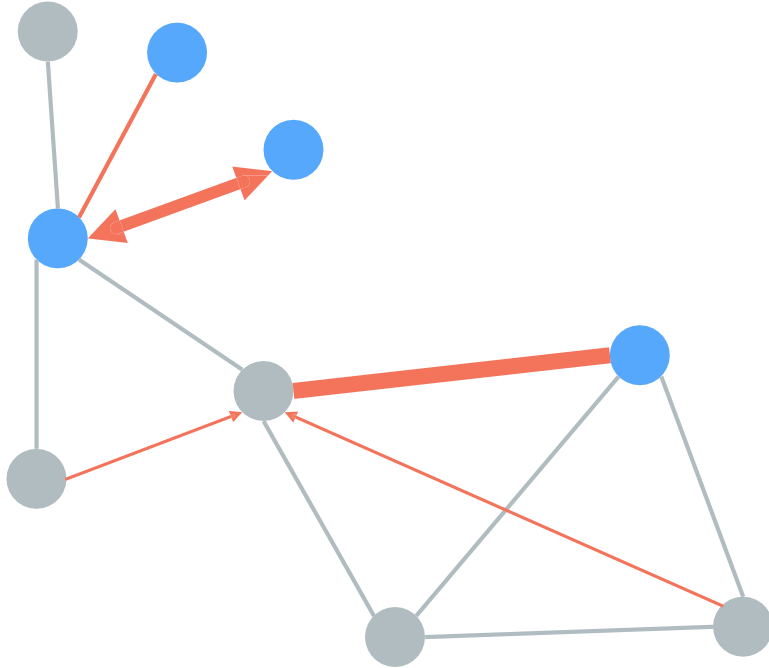
Wat is een **graaf**
(graph)?



Wat is een **graaf**
(graph)?



Wat is een **graaf**
(graph)?

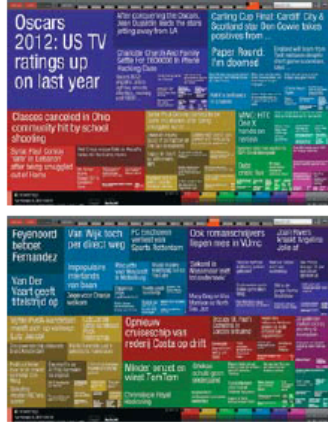


DATA STRUCTURING: TREE MAP [example]

AMERICAS



EUROPE



ASIA



- Canada
- United States
- Mexico
- Brazil
- Argentina
- United Kingdom
- Netherlands
- France
- Spain
- Germany
- Austria
- Italy
- India
- Australia
- New Zealand

AUTHOR Marcos Weskamp (concept, design, frontend and backend coding) and Dan Albritton (backend coding)

COUNTRY United States

DATE 2014

MEDIUM Online, real-time interactive application
<http://newsmap.jp>

DOMAIN News coverage aggregated by Google News API
To provide an overview of online news stories and reveal underlying patterns in news reporting around the world

TASK The visualization uses the treemap technique. The algorithm renders the inner-division shapes closer to rectangles, facilitating readability of text.

STRUCTURE

DATA TYPE AND VISUAL ENCODING

Categorical: News segments

Encoding: Color hues and spatial grouping

Categorical: Countries

Encoding: Label and enabled by selection

Temporal: News age: how old the news is

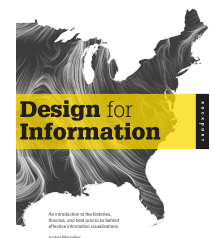
Encoding: Color value

Quantitative: Number of related stories

Encoding: Area size

Nominal: Title of news story

Encoding: Type size relative to the quantitative data

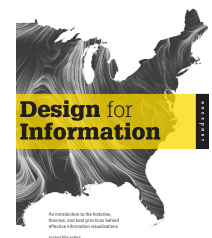


DATA STRUCTURING: TREE MAP [example]

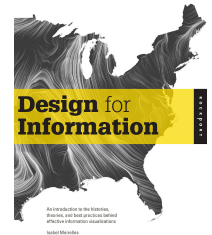
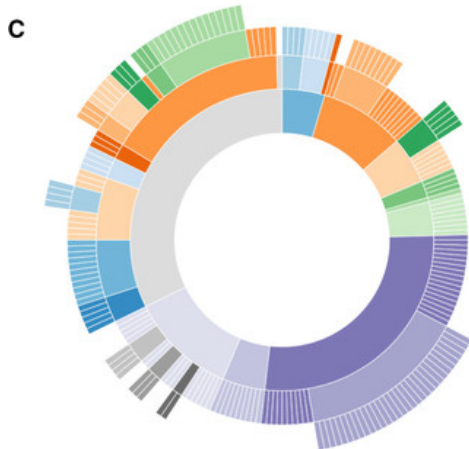
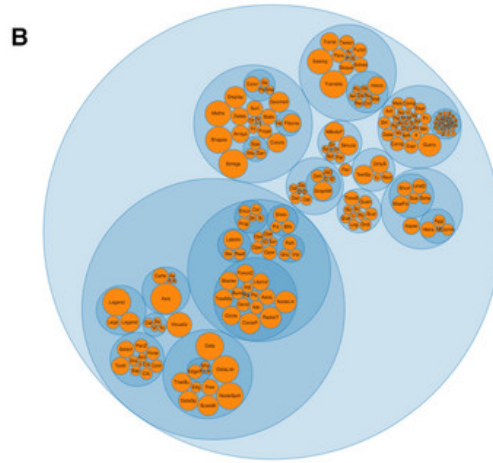
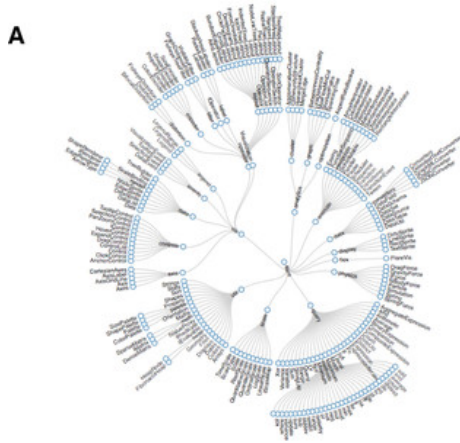
<https://github.com/IJMacD/newsmap-js>

<https://newsmap.ijmacd.com>

<p>Telegraaf.nl: William Spaaij: 'Breuk Noortje heeft mijn ogen geopend'</p> <h1>William Spaaij: 'Breuk Noortje Telegraaf.nl'</h1>	<p>Familie van overleden dj Avicii richt stichting op voor zelfmoordpreventie</p>	<p>Vriendin Dave Mantel: 'Eike ochtend als ik opsta mis ik hem'</p>	<p>'Ouders overleden GGZ-patiënt kregen zwijggeld aangeboden'</p>	<p>Baudet wordt geen Statenlid, Nanninga wel in Binnenland</p>	<p>11.348.000.000 euro begrotingsoverschot, dit doet de overheid ermee</p>	<p>Minister Blok brengt bezoek aan Australië en Maleisië om MH17</p>
	<p>Patty Brard pronkt met haar slanke figuur</p>	<p>Een 'boost voor creativiteit' of is het internet in gevaar?</p>	<p>A73 naar het zuiden korte tijd afgesloten na ongeluk bij Wijchen, weg is weer vrij</p>	<p>Celstraf van 18 jaar voor Groenstroom cokesmokkelaar in Engeland</p>	<p>'Audi-fabriek in België bouwt heilt minder elektronica's door tekort aan accu's</p>	<p>Werknemers opnieuw vaker ziek thuis</p>
	<p>Alle aanklachten tegen acteur Jossie Smollett vervallen</p>	<p>Sophie Turner verdedigt seks voor verpersimpel door ontroerende aflevering Spoofeers</p>	<p>Experts: Wolyven op Veluwe gaan niet naar Nationaal Park Hoge Veluwe</p>	<p>17:39 Uitslaande brand in woning Oosterbeek</p>	<p>Rechterhand Rob Jetten (D66) gaat naar Shell: 'Ongelooflijk dat mensen zo los gaan'</p>	<p>Angst op weg om Armsterc</p>
	<p>Boete voor talkshow direct na overtreiding Mediawet I Show</p>	<p>Keane komt naar het podium</p>	<p>Toch verblijfsvergunning voor uitgeprocedee</p>	<p>High Friet gaat viral: 'Boekingen stromen binnen'</p>	<p>Hoekstra: overschot van 11 miljard niet naar onderwijs, zorg of veiligheid</p>	<p>Angst op weg om Armsterc</p>
	<p>Justin Bieber trekt van leer tegen Halsey-haters</p>	<p>Average Brits: ongeveer 100 miljoen mensen in de wereld</p>	<p>Armeens gezin Den Haag</p>	<p>Brun water uit de kraan door waterstoring in Arnhem</p>	<p>Hoekstra: overschot van 11 miljard niet naar onderwijs, zorg of veiligheid</p>	<p>Angst op weg om Armsterc</p>
	<p>Thomas Acdá gaat diggs en schiedingen - rijk zijn wil hij niet</p>	<p>Camille is nu naar baby met Harry en Meghan</p>	<p>Stenengoiend kauen in Soest te lijf met pinnen</p>	<p>Zwaargewonde door steekpartij Amsterdam I longartsen</p>	<p>Flinke problemen op Schiphol door harde wind</p>	<p>Angst op weg om Armsterc</p>
	<p>Marko: "Verstappen de nieuwe technische directeur aan elkaar gekoppeld"</p>	<p>NAC heeft te veel 'Ruud en de nieuwe technische directeur aan elkaar gekoppeld'</p>	<p>Acht jaar cel voor aanval met mes op toptennisster Kvitová</p>	<p>Nederlandse export naar VK enorm gekrompen sinds Brexit-referendum</p>	<p>Democraten willen rapport Mueller op 2 april I Buitenland</p>	<p>Angst op weg om Armsterc</p>
<p>Ard van Peppen stopt per direct: 'Die aanbiedingen zijn niet voorbijgekomen'</p>	<p>Haerenveen licht optie Kobayashi niet; contracten Schears en Schmidt opgezegd</p>	<p>Gasly over nieuwe regels bandenwarmers: 'Zorgen voor penibele situaties'</p>	<p>Interland-Alert: Victor Jensen scoort voor Denemarken</p>	<p>Nutricia is een soort familiebedrijf I Cuijk</p>	<p>Waterlood Iran breidt zich uit; minstens 19 doden</p>	<p>Angst op weg om Armsterc</p>
	<p>Van der Sar verzet zich tegen FIFA: 'U pleegt roofbouw op de spelers'</p>	<p>Barney hunkert naar afscheid: 'Voel me soms een van de 85'</p>	<p>Negen kraamzusters tegelijk zwanger: 'Eijn om elkaars buik te zien'</p>	<p>Democraten willen rapport Mueller op 2 april I Buitenland</p>	<p>Waterlood Iran breidt zich uit; minstens 19 doden</p>	<p>Angst op weg om Armsterc</p>
	<p>Video: Michael Schumacher maakt debuut voor Ferrari Formule 1-team</p>	<p>Overwinning voor Jong Oranje: Galpko aan boze Tweente-Sas: '185 kaarten is in voldoende binnen'</p>	<p>Ronaldo komt met lichtte spierblessuur bovenbeen</p>	<p>Democraten willen rapport Mueller op 2 april I Buitenland</p>	<p>Waterlood Iran breidt zich uit; minstens 19 doden</p>	<p>Angst op weg om Armsterc</p>



DATA STRUCTURING: TREE MAP [example]



DATA STRUCTURING: Spatial relationships [MAPPING]



VISUAL ELEMENTS

SIGNIFYING PROPERTIES

VARIABLES OF THE IMAGE

XY | 2 dimensions of the plane

Z

POINT

LINE

AREA

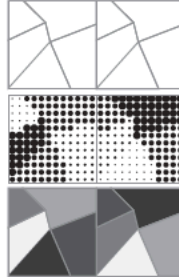
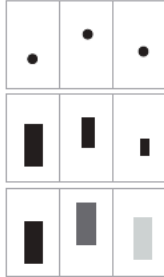
QUANTITATIVE

ORDERED

SELECTIVE

ASSOCIATIVE

DISSOCIATIVE



DIFFERENTIAL VARIABLES

Texture

Color

Orientation

Shape



Visual Encoding

The **data attribute** of **dimension** is one of the most important characteristics when considering how to conceptualize **visual marks** in **cartography**.

The basic graphic elements of visual representation are:

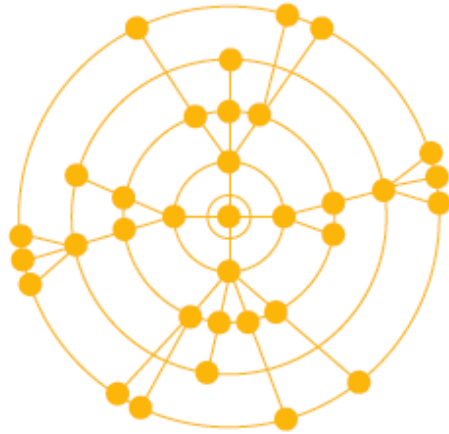
Point has no dimension provides a sense of place.

Line has one dimension [X] provides a sense of length and direction.

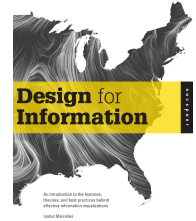
Plane has two dimensions [X,Y] provides a sense of shape and scale.

Volume has 3 dimensions [X,Y,Z] provides a sense of space, shape and scale.

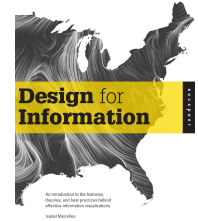
POLAR SYSTEMS



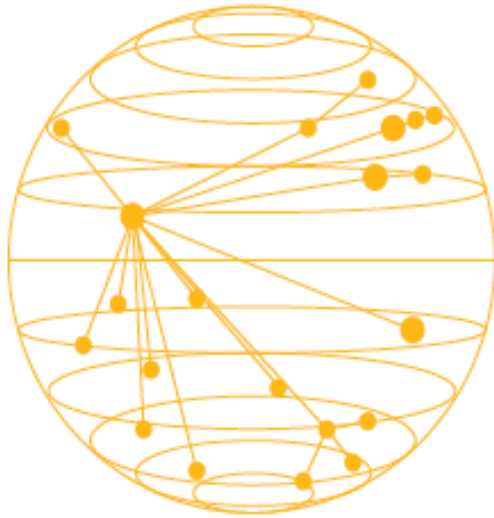
node-link radial layout



radial icicle or sunburst



OTHER GEOMETRIES

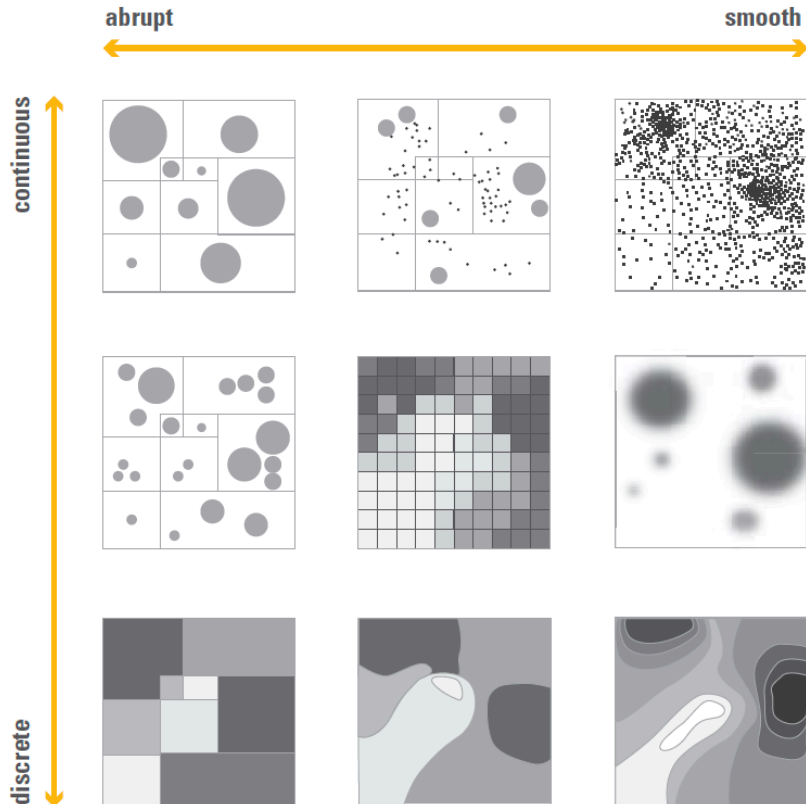


3D hyperbolic tree



Voronoi treemap

DATA STRUCTURING: Spatial relationships [MAPPING]



Visual Encoding

Visual encoding is the process of matching the phenomena to be visualized, which is provided by the dataset (data scale and attributes), to the most suitable type of representation (graphical elements and visual properties). Visual encoding in **cartography** is often called **symbolization**.

The **data attributes** highly significant for **spatial relationships** are **discrete** versus **continuous** data & **abrupt** (low resolution) versus **smooth** data (high resolution).



DATA STRUCTURING: Mapping

Graphical Methods



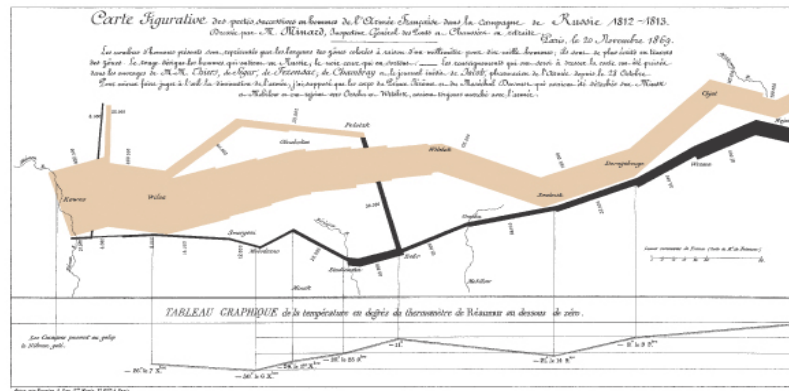
There are 6 graphical methods used primarily in thematic maps for representing all sorts of qualitative and quantitative data:

1. Dot Distribution Maps
2. Graduated Symbol Maps
4. Isometric & Isopleth Maps
5. Flow And Network Maps
3. Choropleth Maps
6. Area & Distance Cartograms

Maximization of useful information on a limited display

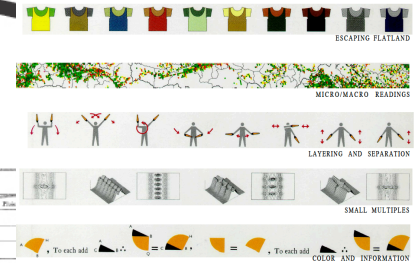
Probably the best statistical graphic ever drawn, this map by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the Polish-Russian border, the thick band shows the size of the army at each position. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. Exquisitely printed in two colors on fine archival paper, 22" by 15".

[Minard's sources.](#) [Minard's biography.](#)



Edward R. Tufte

Envisioning Information



DATA STRUCTURING: Mapping Dot distribution maps

The New York Times

Mapping the 2010 U.S. Census

Browse population growth and decline, changes in racial and ethnic concentrations and patterns of housing development.

View More Maps



<http://www.nytimes.com/projects/census/2010/map.html>

Distribution of racial and ethnic groups in 2010

One dot = 5,000 people

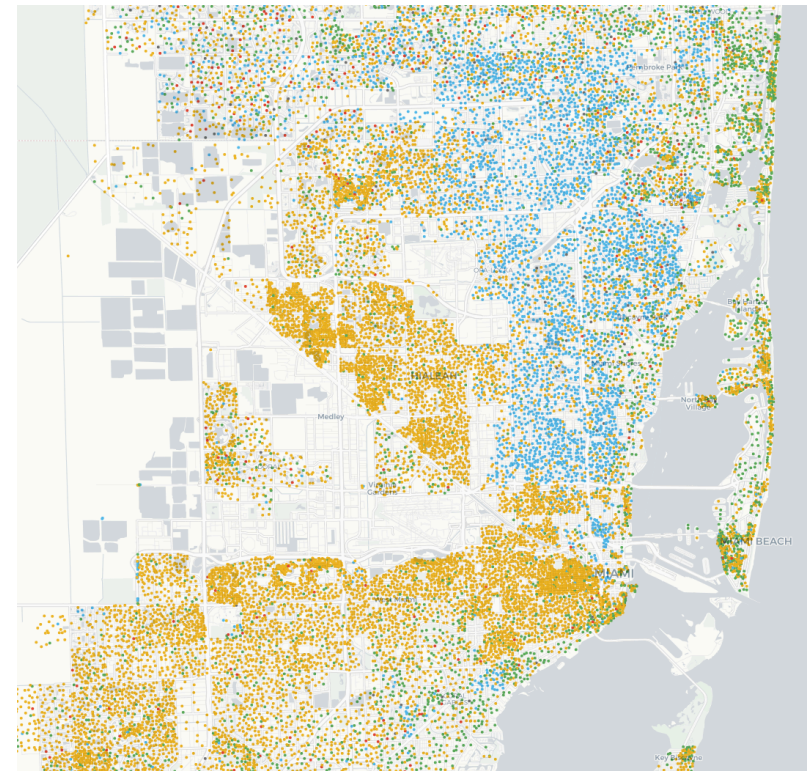
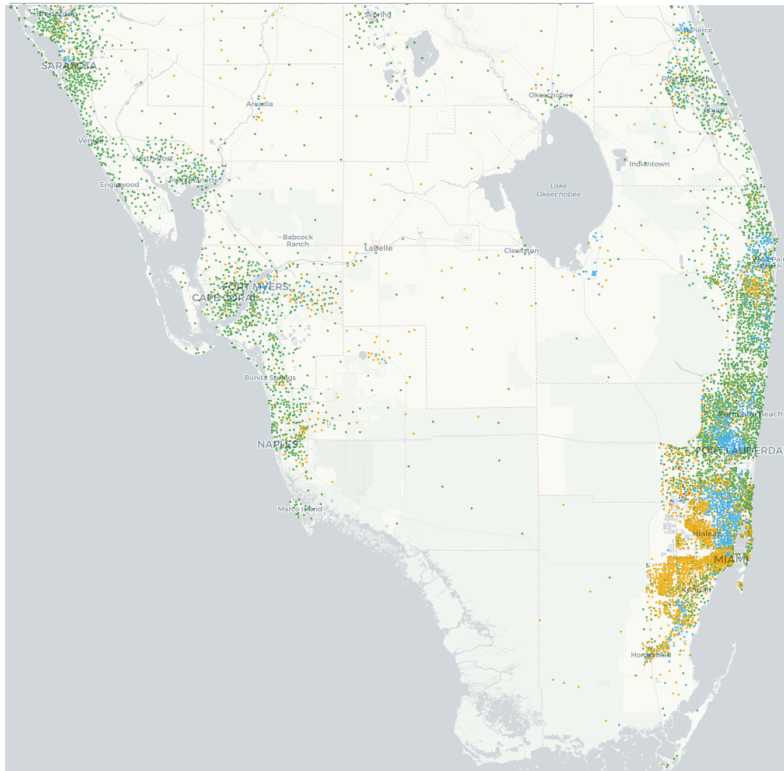
- White
- Black
- Hispanic
- Asian
- Native American
- Other

Zoom to a State

Missouri

2010 POPULATION	CHANGE FROM 2000
5,988,927	+7.0%

RACE/ETHNICITY	SHARE OF POP.	CHANGE FROM 2000
Whites:	81%	+4%
Blacks:	11%	+10%
Hispanics:	4%	+79%
Asians:	2%	+59%
Native Amer.:	0%	+3%
Multiracial:	2%	+48%
Other groups:	0%	+35%



DATA STRUCTURING: Mapping Graduated symbol maps

365 **Obama**
Electoral Votes
Projected Winner

0
undecided

173 **McCain**
Electoral Votes

Popular vote: 66,862,039

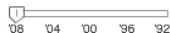
270 needed to win

Popular vote: 58,319,442

- State winners
- County bubbles
- County leaders
- Voting shifts

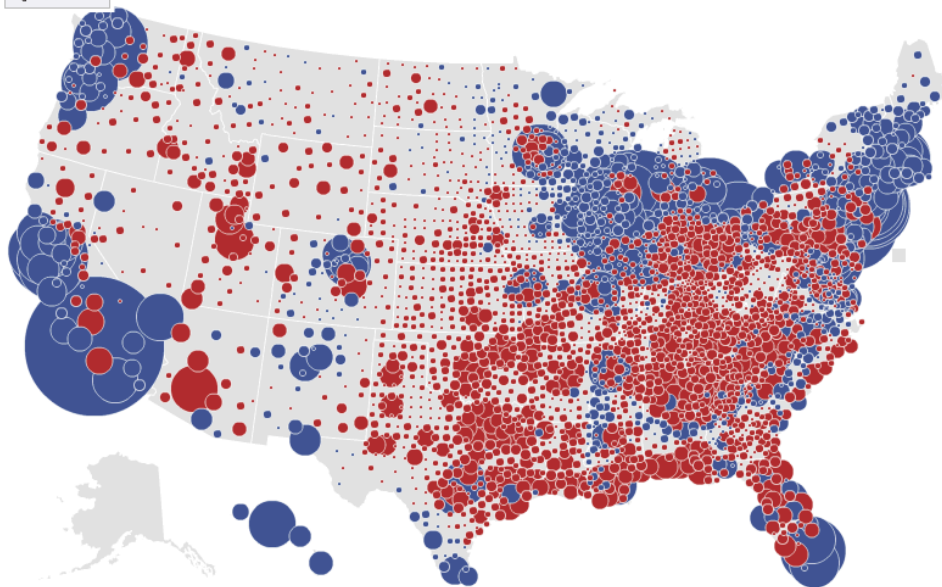
ZOOM IN

Year

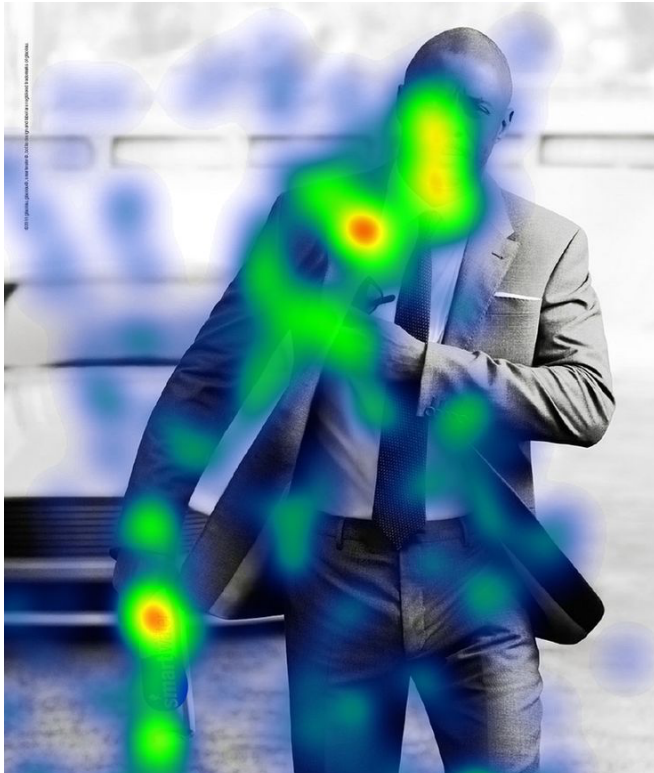


Map key

Circle size is proportional to the amount each county's leading candidate is ahead.



DATA STRUCTURING: Mapping Isometric and isopleth maps



smart because
good taste is non-negotiable

smart water
smart because it's made that way

Please fill out the information below.

Personal Information

First Name

Last Name

Contact Information

Address

City

County

Post Code

Country

An eye-tracking heatmap overlaid on a registration form. The heatmap shows high attention (red and yellow) on the form fields, particularly the name and address sections, and the 'Submit' button. The background is mostly green, indicating low attention.

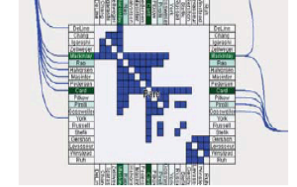
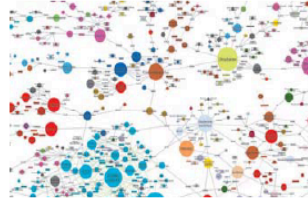
EYE TRACKING THE USER EXPERIENCE

A Practical Guide to Research

by **Aga Bejko** Foreword by Steve Krug

Rosenfeld

DATA STRUCTURING: Mapping Flow and network maps



LINEAR:

Nodes are organized linearly and the links are usually arcs connecting nodes.

Con: It's hard to identify clusters and is only feasible for small datasets.



FORCE DIRECTED:

There are many algorithms that use an iterative process to locate nodes according to physical forces.

Con: There are too many node occlusions and link crossings in dense areas.



CIRCULAR:

Nodes are organized around the circumference and usually grouped by categories. Links cross the circle and are usually bundled so as to simplify the crossings.

Con: It's hard to identify clusters.



COMMUNITY STRUCTURE:

The focus is on community structures.



GEOGRAPHY BASED:

Spatial location of a node is provided by its geo position.



MATRIX:

Grid of nodes with link information positioned within the cell.



SANKEY TYPE DIAGRAMS:

Nodes are organized vertically and the links horizontally.



FORCE DIRECTED:

Force directed graphs centered on a node.



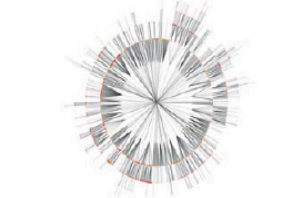
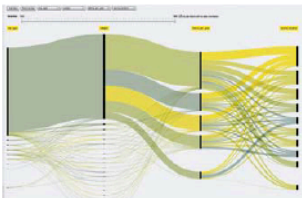
POLAR OR RADIAL:

Nodes are organized around a central node, with their position related to the number of hops it takes to reach it.



RADIAL COMMUNITY STRUCTURE:

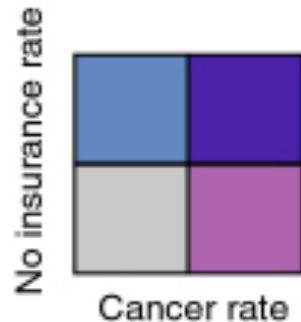
Nodes are organized around a central community



DATA STRUCTURING: Mapping Choropleth (Multivariate) maps

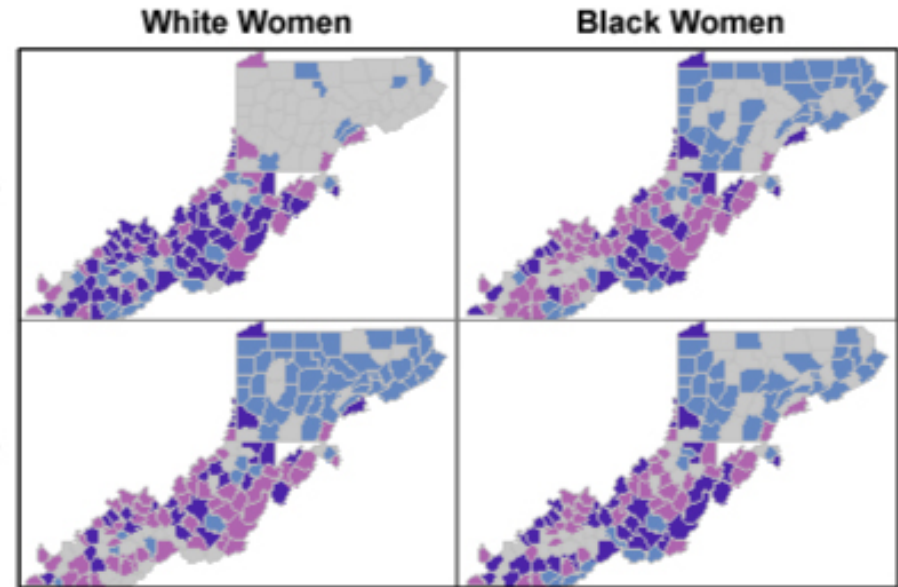
Multivariate Data

Thematic maps can depict several sets of **nonspatial** data simultaneously. When a thematic map portrays exclusively one set of data, it is called **univariate**. If it shows two distinct sets of data, it is called **bivariate**, and for more than two sets, maps are called **multivariate**.

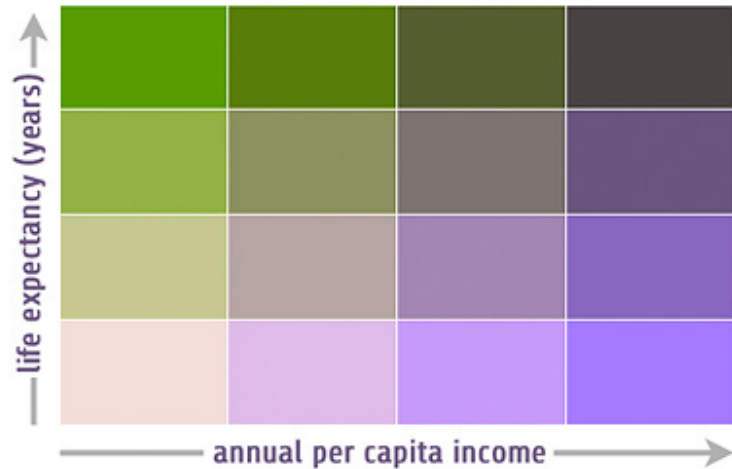


Cervical Cancer

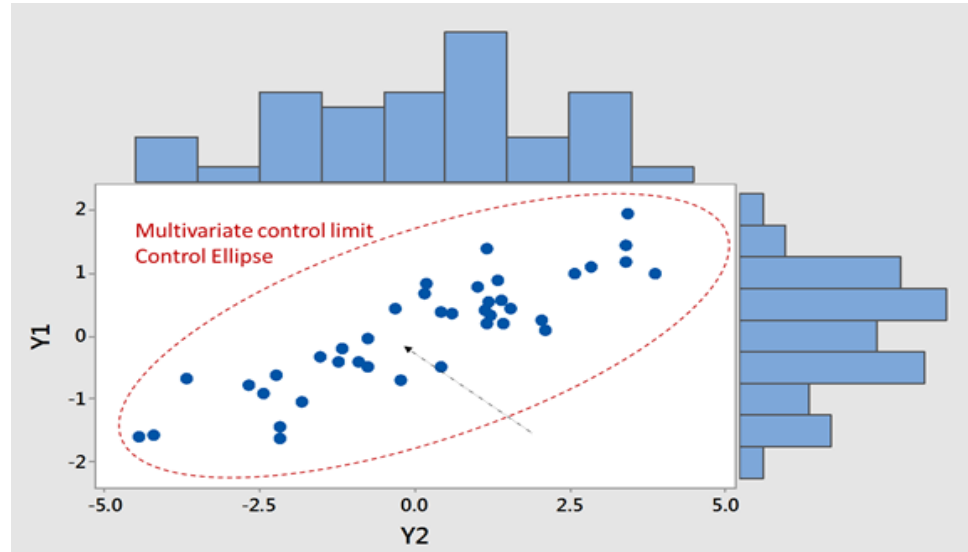
Breast Cancer



DATA STRUCTURING: Mapping Choropleth (Multivariate) maps



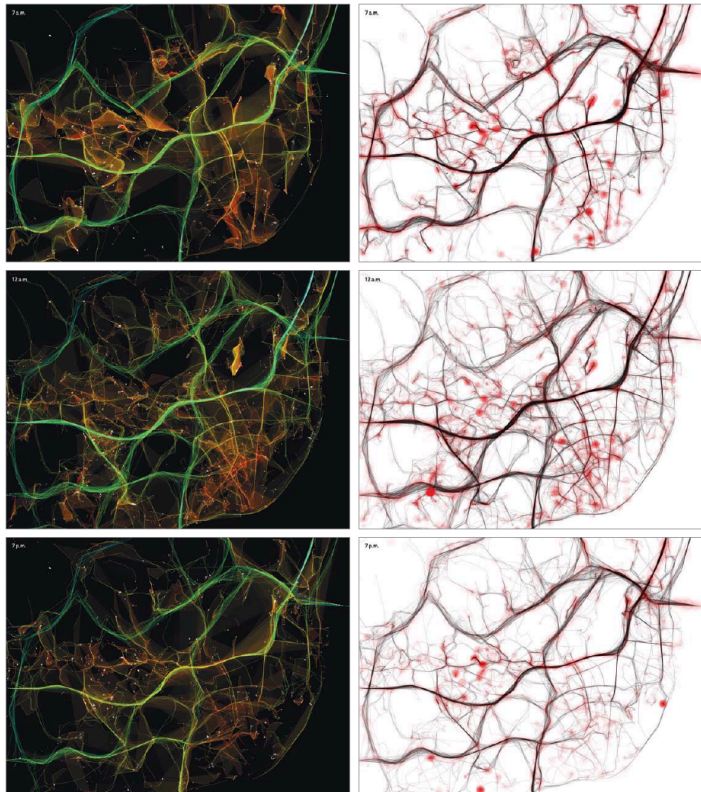
two-theme map legend (bivariate)



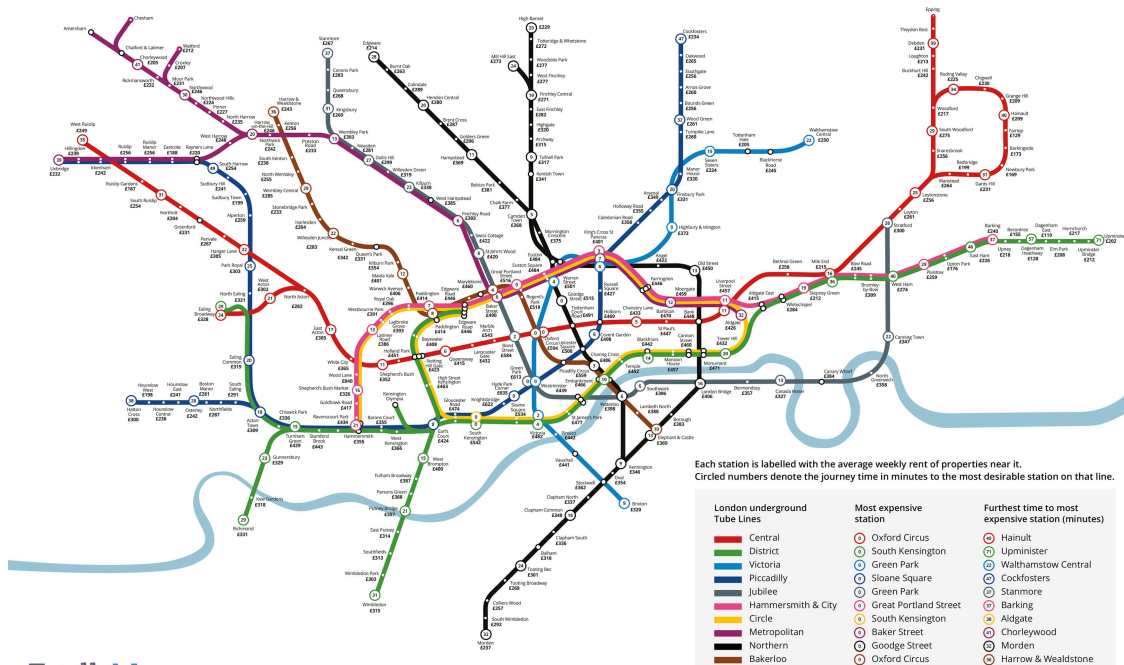
Multivariate Data

Thematic maps can depict several sets of **nonspatial** data simultaneously. When a thematic map portrays exclusively one set of data, it is called **univariate**. If it shows two distinct sets of data, it is called **bivariate**, and for more than two sets, maps are called **multivariate**.

DATA STRUCTURING: Mapping Area and distance cartograms



London Underground Map Weekly rent vs journey time



TotallyMoney.com

THE NATURE OF CODE

How can we capture the unpredictable evolutionary and emergent properties of nature in software?

How can understanding the mathematical principles behind our physical world help us to create digital worlds?

This book focuses on a range of programming strategies and techniques behind computer simulations of natural systems, from elementary concepts in mathematics and physics to more advanced algorithms that enable sophisticated visual results. Using the open-source language Processing, readers will progress from building a basic physics engine to creating intelligent moving objects and complex systems, setting the foundation for further experiments in generative design.

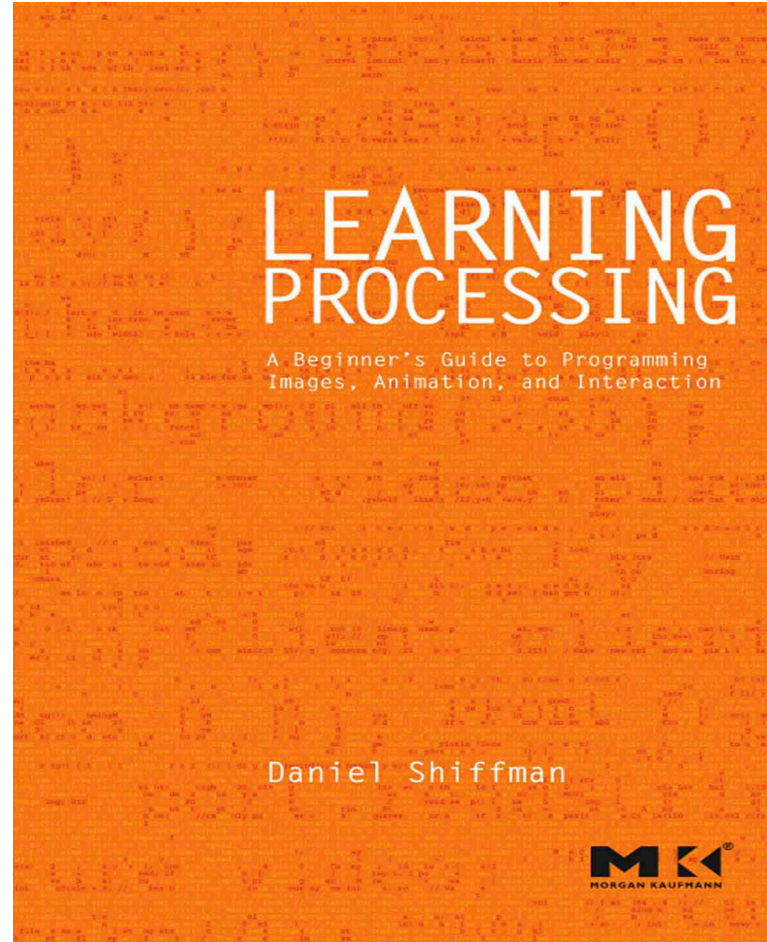
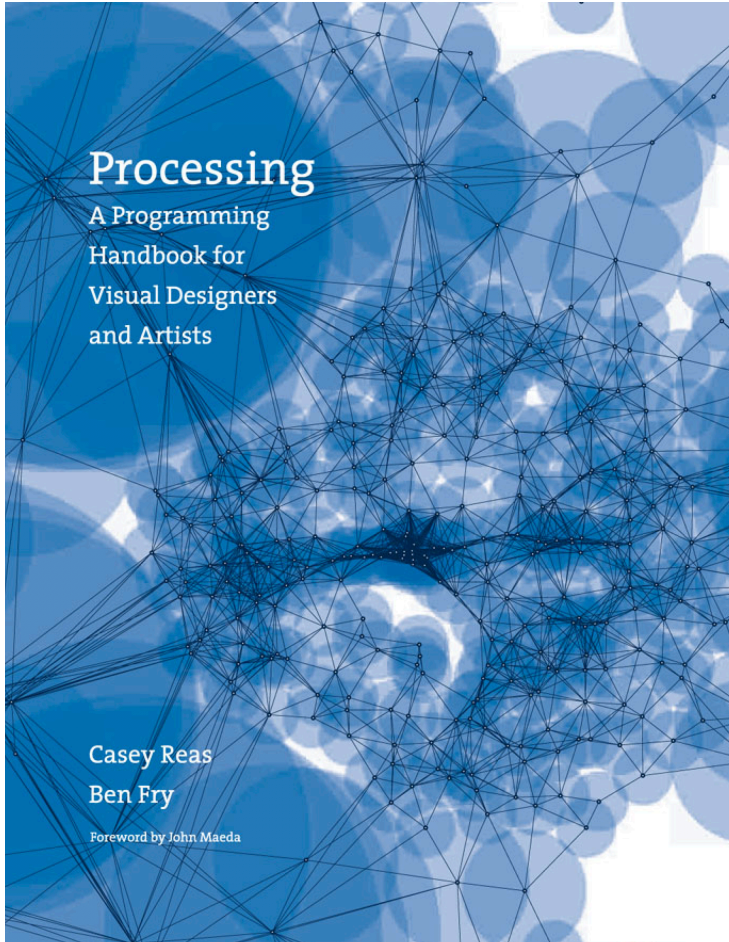
<http://natureofcode.com>

THE NATURE OF CODE
DANIEL SHIFFMAN

THE NATURE OF CODE

SIMULATING NATURAL SYSTEMS WITH PROCESSING

DANIEL SHIFFMAN



Structure

Statements and Comments

Coordinates

Width and Height

Setup and Draw

No Loop

Loop

Redraw

Functions

Recursion

CreateGraphics

Form

Points and Lines

Shape Primitives

Pie Chart

Regular Polygon

Star

Triangle Strip

Bezier

Data

Variables

Integers and Floats

True/False

Characters and Strings

Variable Scope

Datatype Conversion

Arrays

Array

Array 2D

Array Objects

Control

Iteration

Embedded Iteration

Conditionals 1

Conditionals 2

Logical Operators

Image

Load and Display Image

Background Image

Transparency

AlphaMask

CreateImage

Pointillism

Request Image

Color

Hue

Saturation

Brightness

Color Variables

Relativity

Linear Gradient

Radial Gradient

Math

Increment/Decrement

Operator Precedence

Distance 1D

Distance 2D

Map

Sine

Sine and Cosine

Sine Wave

Additive Wave

Polar to Cartesian

Arctangent

Graphing 2D Equation

Interpolate

Random

Double Random

Random Gaussian

Noise 1D

Noise 2D

Noise 3D

NoiseWave

Input

Mouse ID

Mouse 2D

MouseDown

Mouse Signals

Easing

Constrain

Storing Input

Mouse Functions

Keyboard

Keyboard Functions

Milliseconds

Clock

Transform

Translate

Scale

Rotate

Arm

Objects

Objects

Multiple Constructors

Composite Objects

Inheritance

Typography

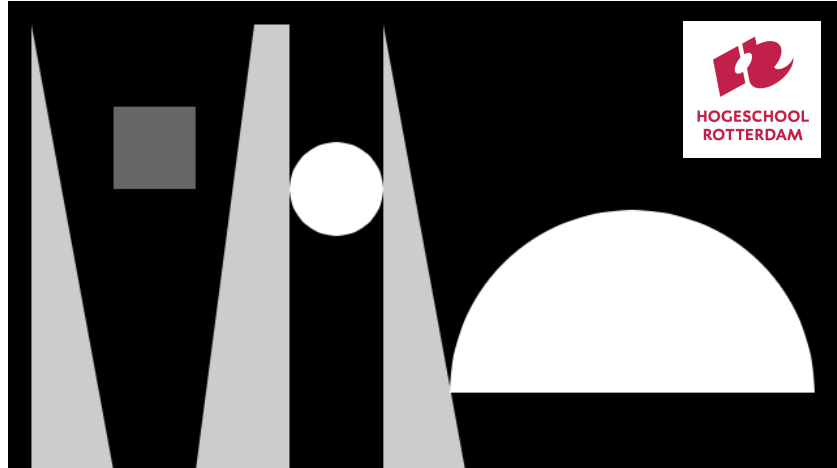
Letters

Words

Web

Embedded Links

Loading Images



```
size(640, 360);  
background(0);  
noStroke();
```

```
fill(204);  
triangle(18, 18, 18, 360, 81, 360);
```

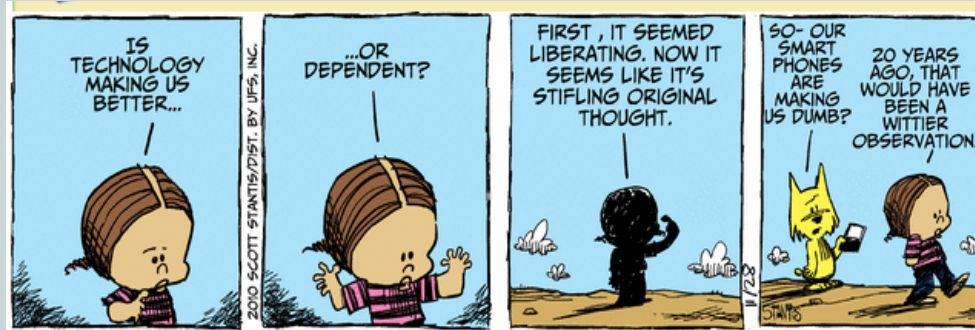
```
fill(102);  
rect(81, 81, 63, 63);
```

```
fill(204);  
quad(189, 18, 216, 18, 216, 360, 144, 360);
```

```
fill(255);  
ellipse(252, 144, 72, 72);
```

```
fill(204);  
triangle(288, 18, 351, 360, 288, 360);
```

```
fill(255);  
arc(479, 300, 280, 280, PI, TWO_PI);
```

This lesson was developed by:

Robert Frans van der Willigen
CMD, Hogeschool Rotterdam
MARCH 2019

Creative Commons License Types

	Can someone use it commercially?	Can someone create new versions of it?
Attribution		
Share Alike		Yup, AND they must license the new work under a Share Alike license.
No Derivatives		
Non-Commercial		Yup, AND the new work must be non-commercial, but it can be under any non-commercial license.
Non-Commercial Share Alike		Yup, AND they must license the new work under a Non-Commercial Share Alike license.
Non-Commercial No Derivatives		

SOURCE

<http://www.masternewmedia.org/how-to-publish-a-book-under-a-creative-commons-license/>

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

This lesson is licensed under a Creative Commons Attribution-Share-Alike license. You can change it, transmit it, show it to other people. Just always give credit to RFvdW.

<http://empoweringthenatives.edublogs.org/2012/03/15/creative-commons-licenses/>

<http://creativecommons.org/licenses/>

