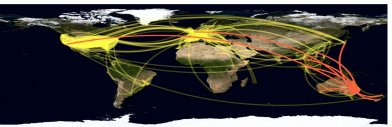
Aviz
Visual Analytics Project

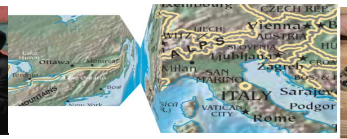
université PARIS-SACLAY

Inria

The Human Processor

Jean-Daniel Fekete
Jean-Daniel.Fekete@inria.fr
https://www.aviz.fr/~fekete



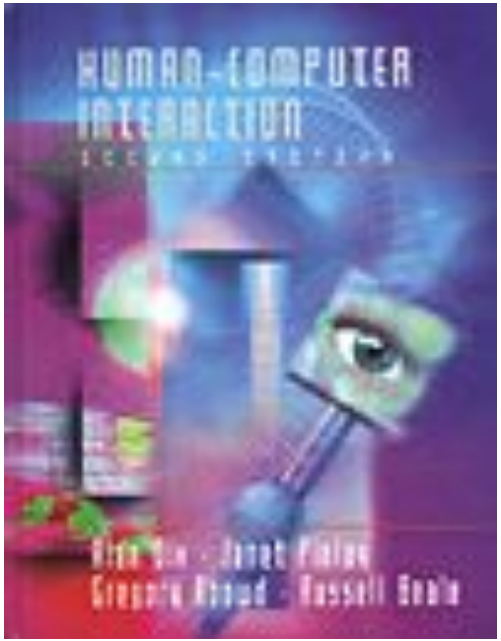
HCI inspired by

HUMAN-COMPUTER INTERACTION

third edition

Alan Dix - Janet Finlay - Gregory Abowd - Russell Beale

Prentice Hall, 2004. ISBN 0-13-046109-1

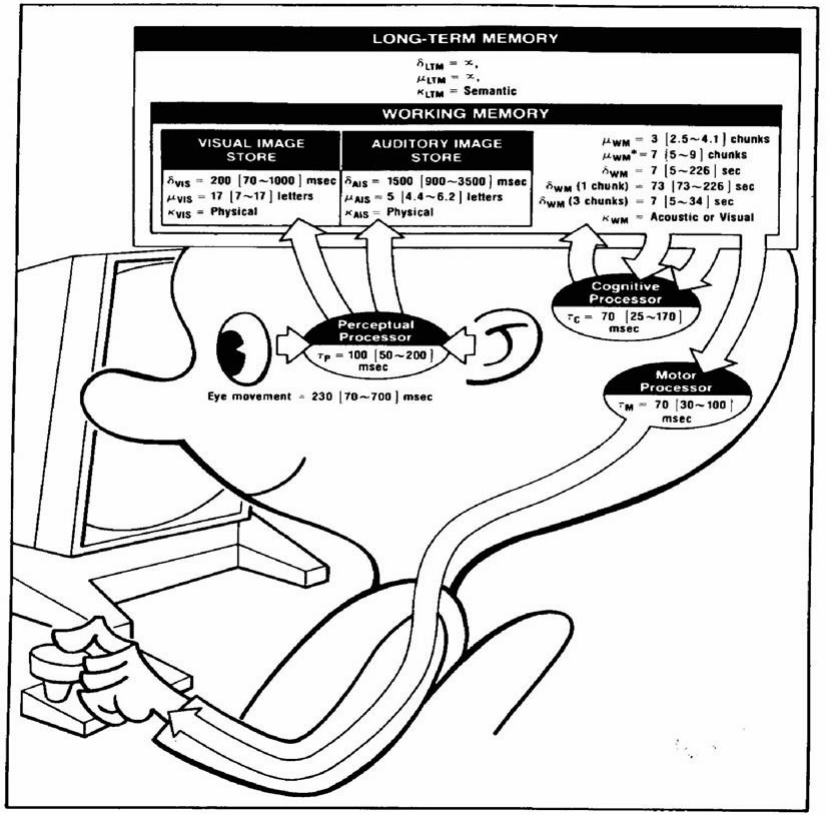


The Human Processor

Humans can be viewed as an information processing system:

- information received and responses given via input-output channels
- information stored in memory
- information processed and applied in various ways

Capabilities of humans in these areas are important to design, as are individual differences



Stuart K. Card, Allen Newell, and Thomas P. Moran. 1983. The Psychology of Human-Computer Interaction. L. Erlbaum Assoc. Inc., Hillsdale, NJ, USA.

Take Home Message

You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work

Humans input-output processing

A **cortical homunculus** is a distorted representation of the human body, based on a neurological "map" of the areas and proportions of the human brain dedicated to processing motor functions, or sensory functions, for different parts of the body.

[https://en.wikipedia.org/wiki/Cortical_homunculus]

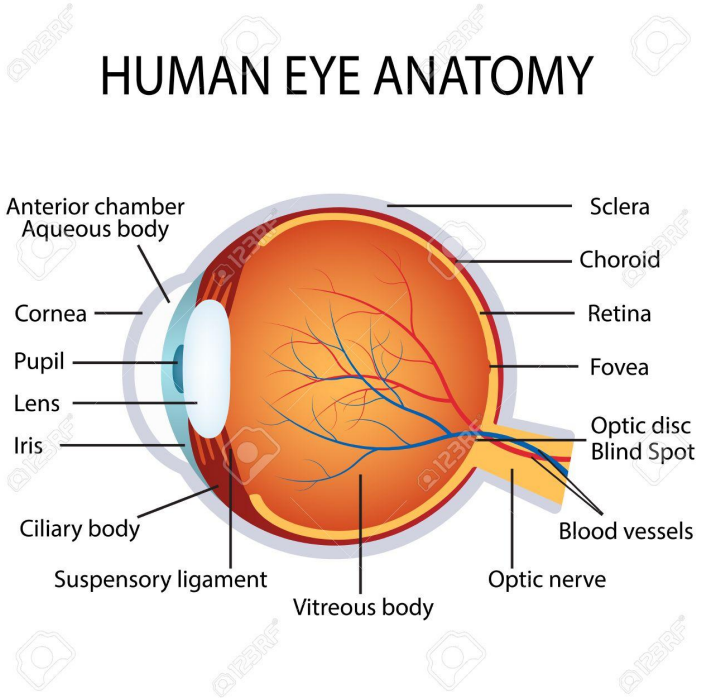
Vision

Two stages in vision :

- physical reception of stimulus
- processing and interpretation of stimulus

The Eye:

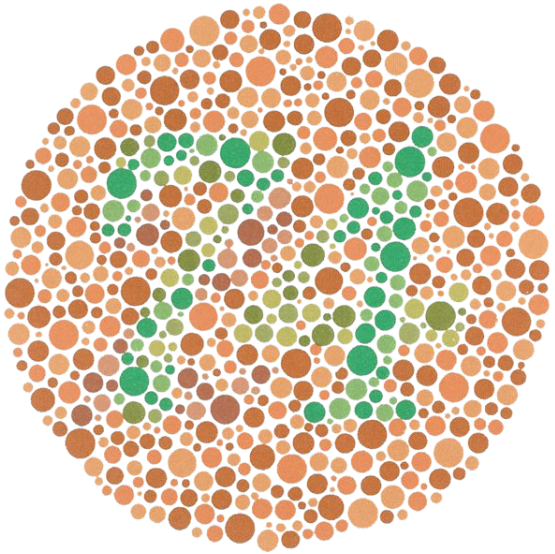
- mechanism for receiving light and transforming it into electrical energy
- light reflects from objects; their images are focused upside-down on retina
- retina contains rods for low light vision and cones for colour vision
- ganglion cells detect pattern and movement



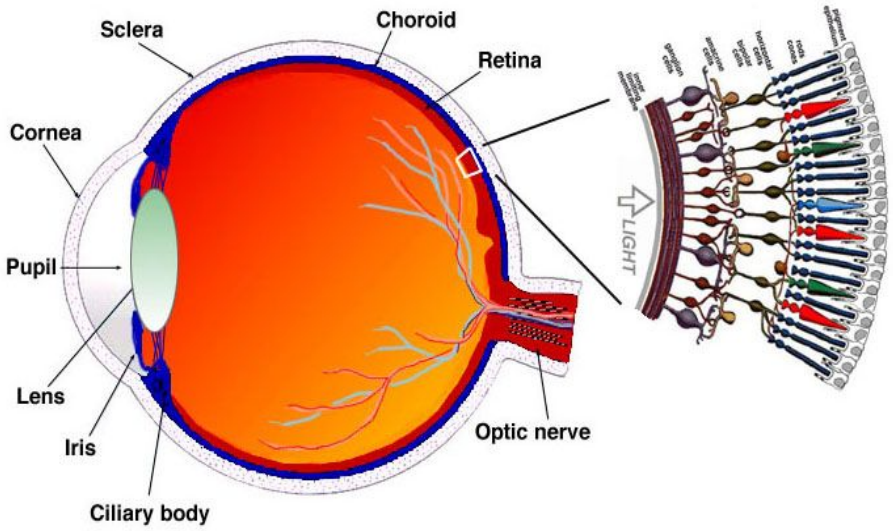
Vision: important issues

The vision system is very powerful but suffers from many problems and limitations:

- Some features can be perceived very quickly and accurately
  - Preattentive perception
- But limited in resolution, perception of color, lightness, movement
- The brain tries to compensate but cannot always
  - Blind spot experiment
- Deficiencies
  - Color blindness
    - 8% of males and 0.5% of females
  - Change blindness



Physical World➡**Visual System**



Simple Anatomy of the Retina, Helga Kolb

Rods

No color (sort of)

All over the retina

More sensitive

Cones

Three different kinds of “color receptors”

Mostly in the center

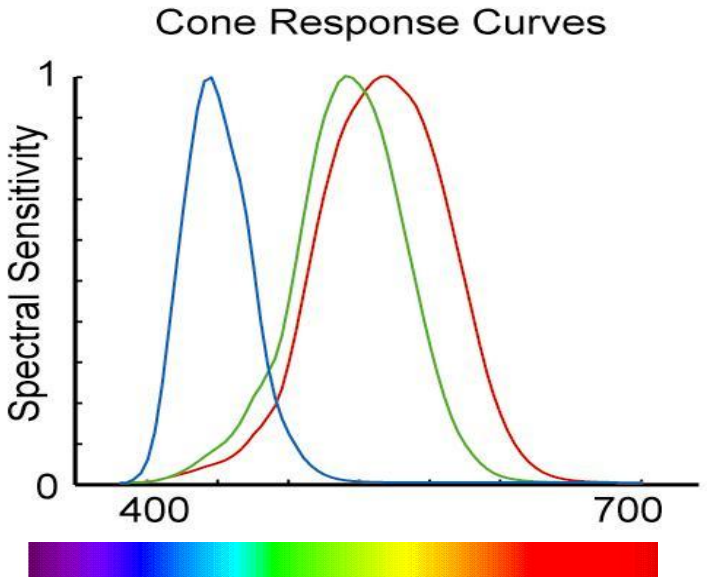
Less Sensitive

Perception of color is split in 3

- Old visual system: intensity
- Newer visual system: hue/saturation

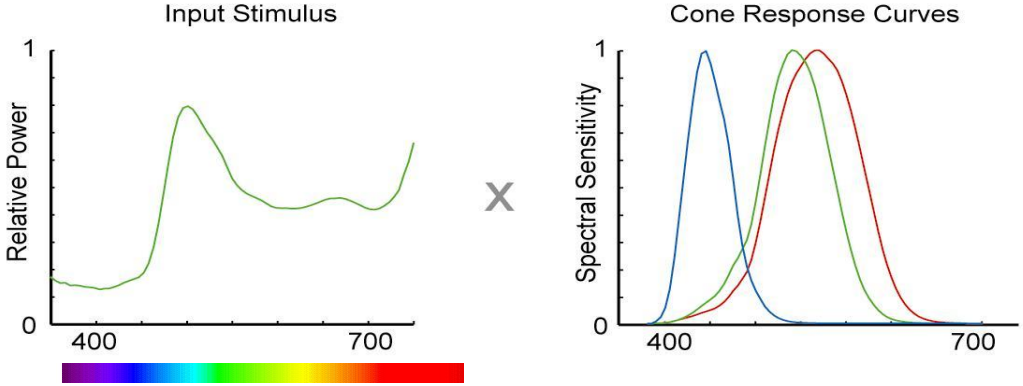
Cone response

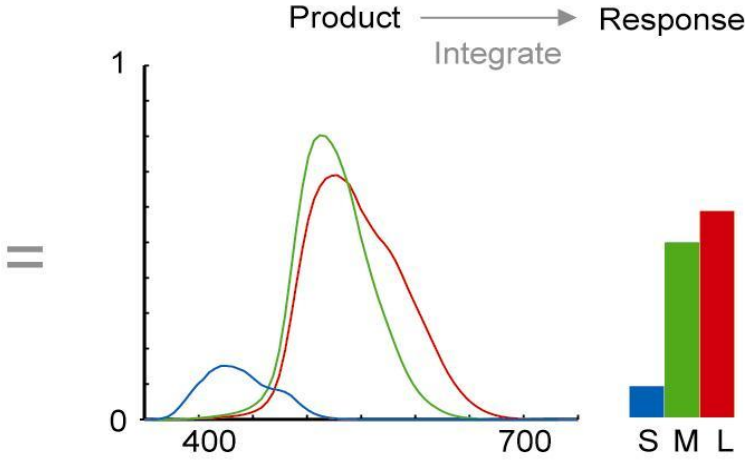
- LMS (Long, Middle, Short) cones
- Capture different wavelengths (some better than others)
- Transmit a signal to the brain



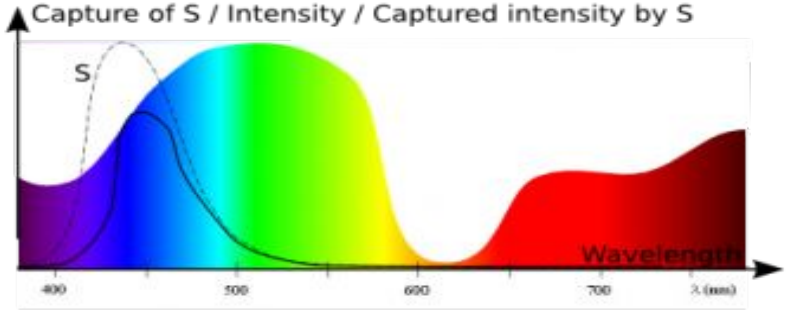
A Field Guide to Digital Color, Maureen Stone

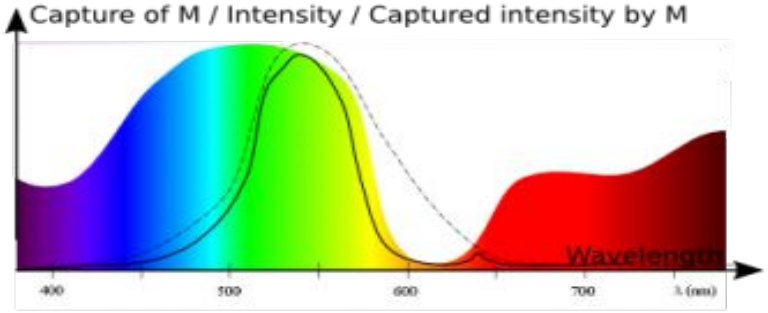
Cone response

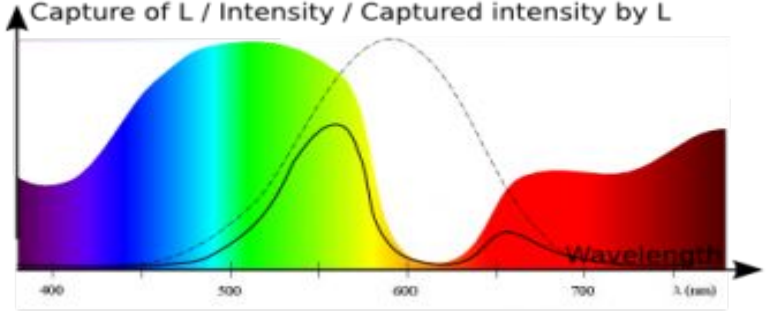




A Field Guide to Digital Color, Maureen Stone



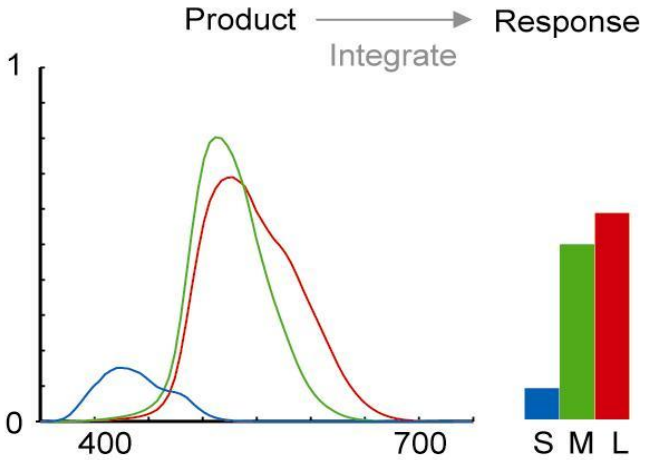


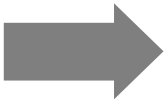


SML decomposition

http://www.science4all.org/article/colors/

Visual System ➡ **Color Models**





This is the color the eye sees

This is not necessarily the color the brain sees!

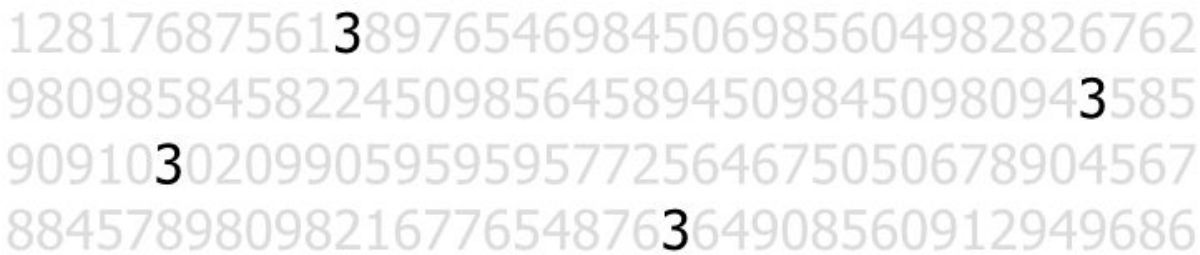
PREATTENTIVE PROCESSING

How many 3’s do you see?

128176875613897654698450698560498282676 2
980985845822450985645894509845098094358 5
909103020990595959577256467505067890456 7
884578980982167765487636490856091294968 6

From: Ware, Information Visualization using Vision to Think

How about now?



From: Ware, Information Visualization using Vision to Think

Preattentive Processing

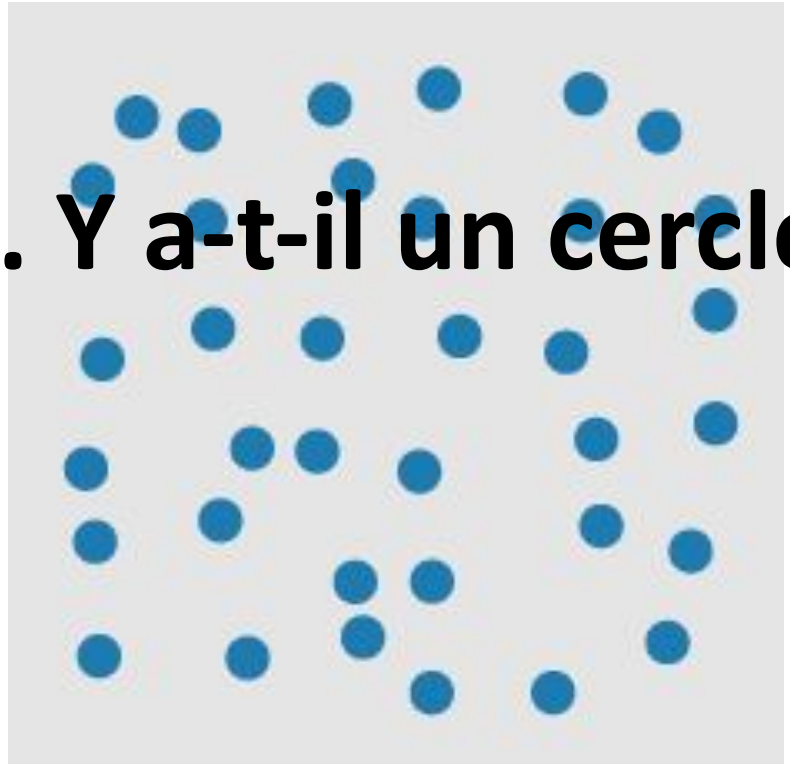
- Some stimuli can be perceived **without** the need for focused attention
- Generally within **200-250 ms**
- Seems to be done **in parallel** by the low-level vision system

Visual encoding has a **big** impact on this!

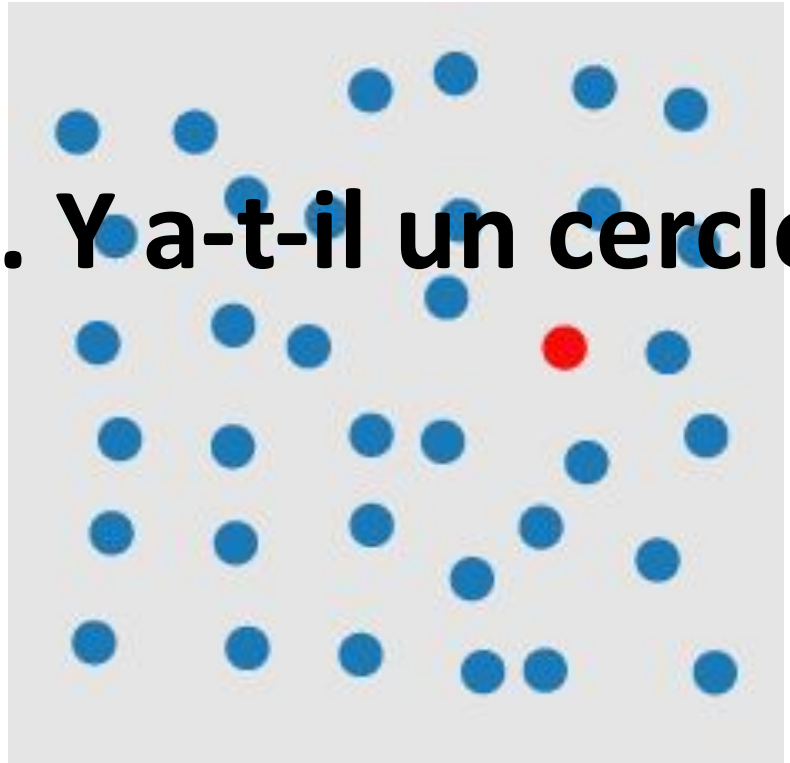
Visual encodings influence **preattentive** processing

DETERMINE IF A RED CIRCLE IS PRESENT

Q1. Y a-t-il un cercle rouge?



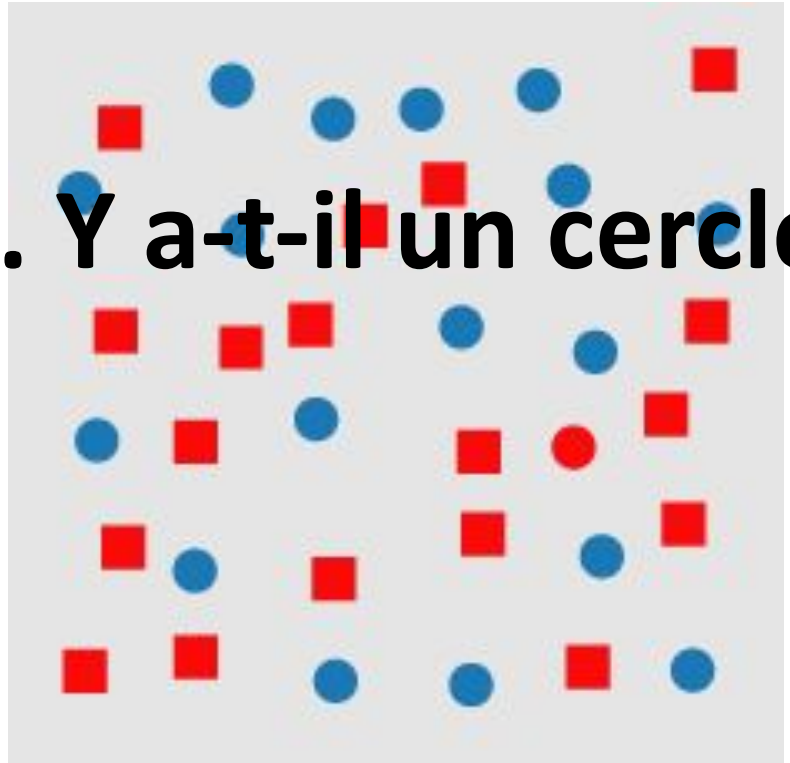
Q2. Y a-t-il un cercle rouge?



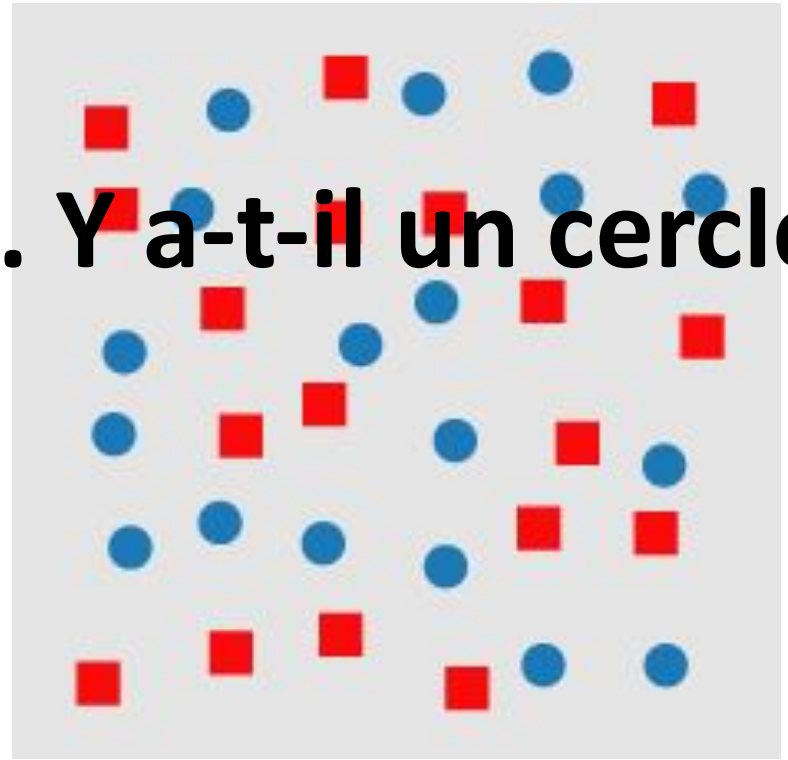
Q3. Y a-t-il un cercle rouge?

Q4. Y a-t-il un cercle rouge?

Q5. Y a-t-il un cercle rouge?

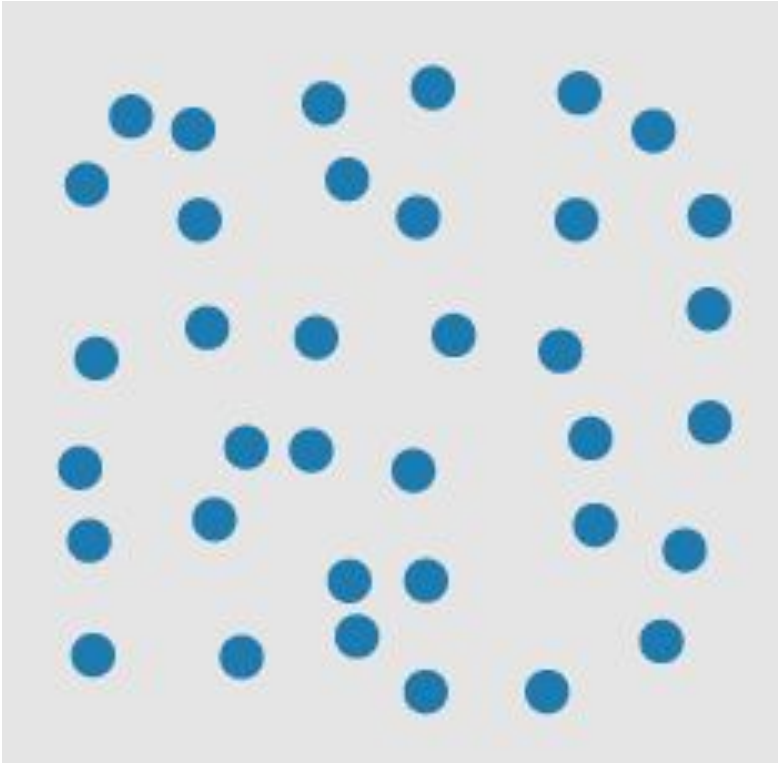


Q6. Y a-t-il un cercle rouge?



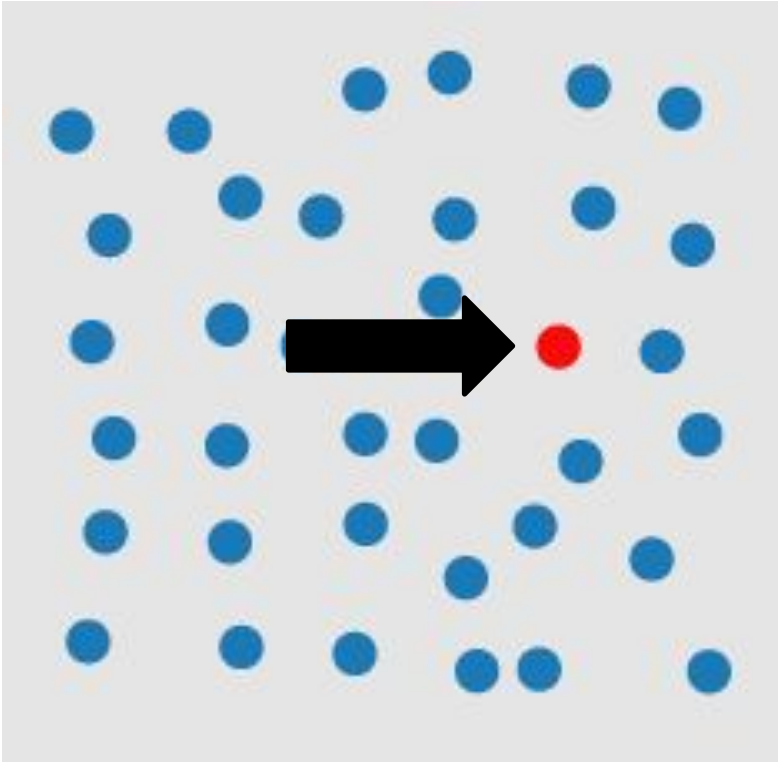
Q1. Y avait il un cercle rouge?

Non



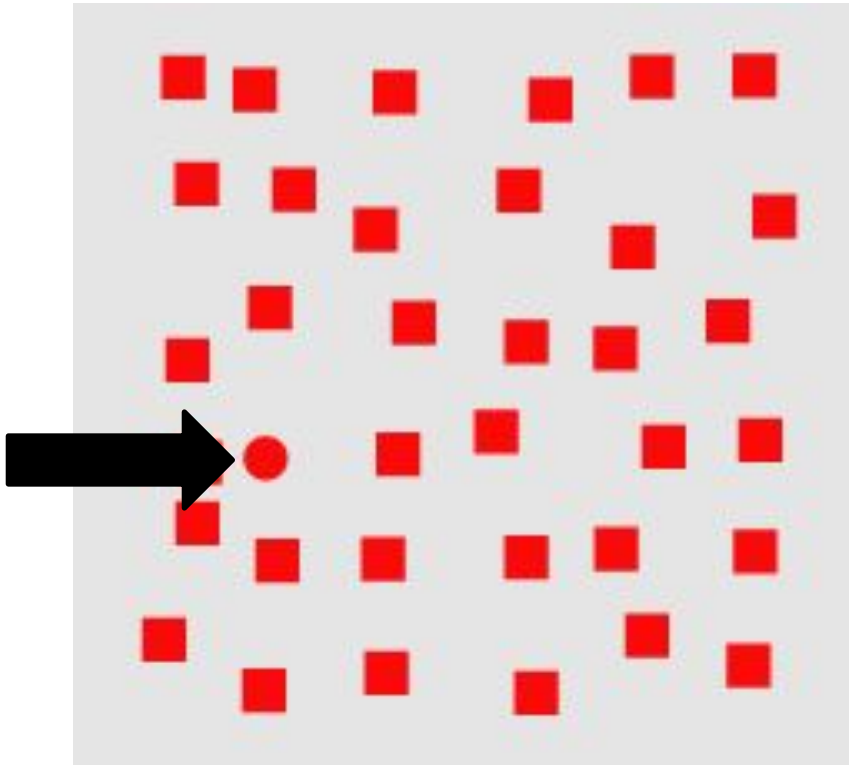
Q2. Y avait il un cercle rouge?

Oui



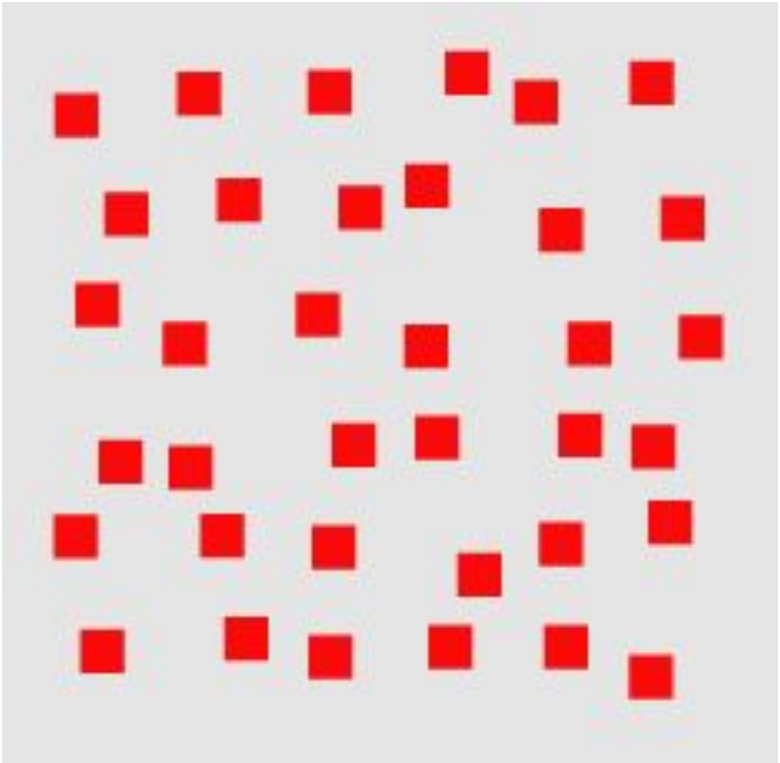
Q3. Y avait il un cercle rouge?

Oui



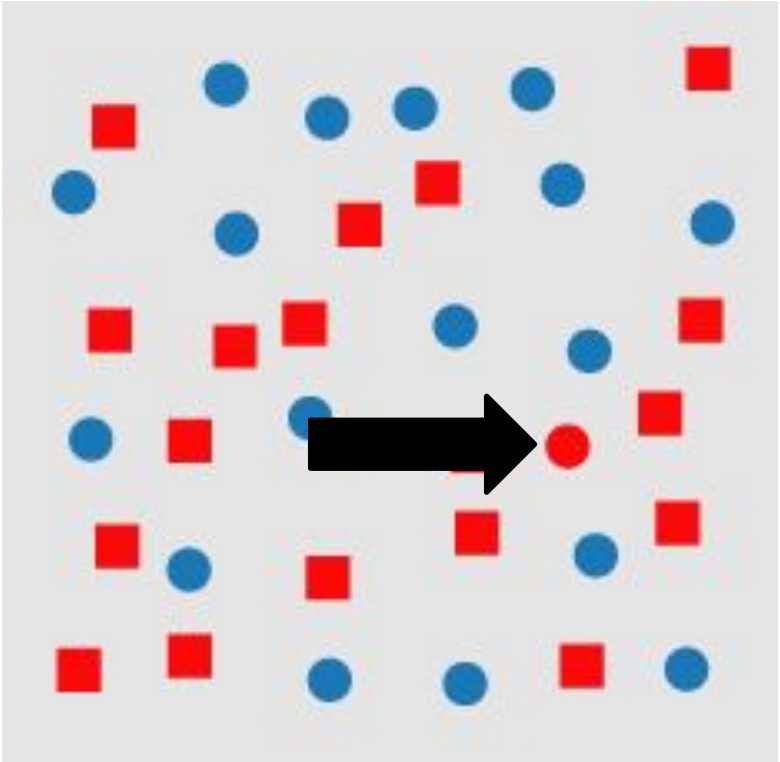
Q4. Y avait il un cercle rouge?

Non



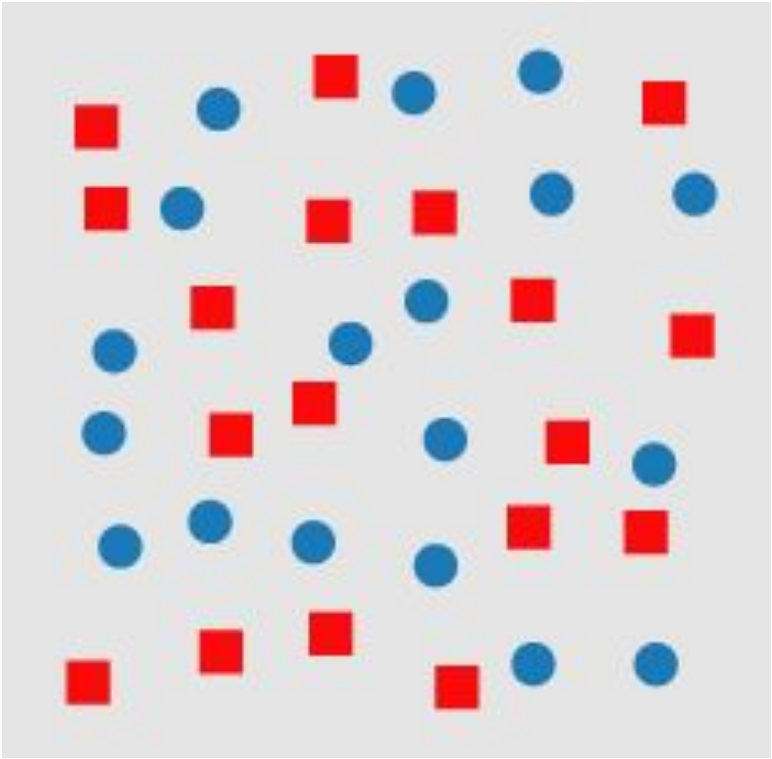
Q5. Y avait il un cercle rouge?

Oui

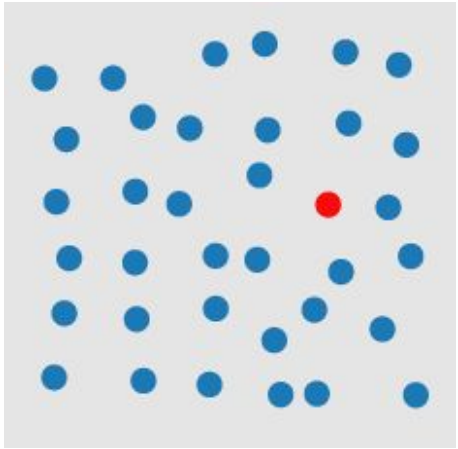


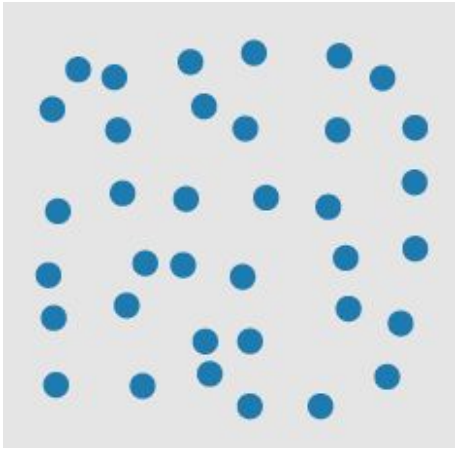
Q6. Y avait il un cercle rouge?

Non



Hue

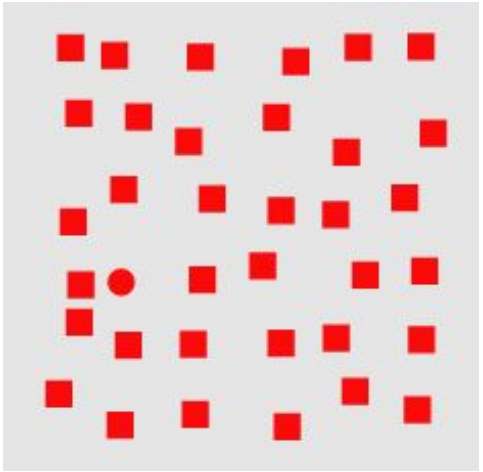


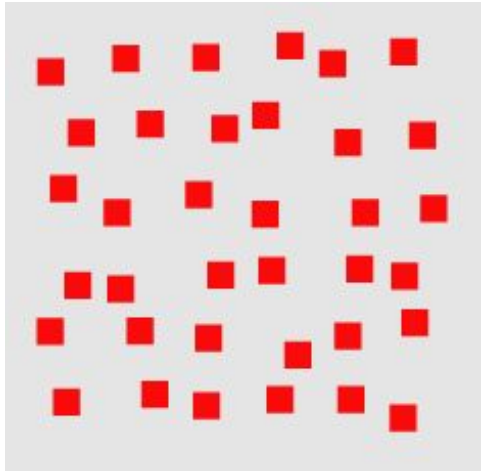


Yes, can be done preattentively

From: Healey, Perception in Visualization

Shape

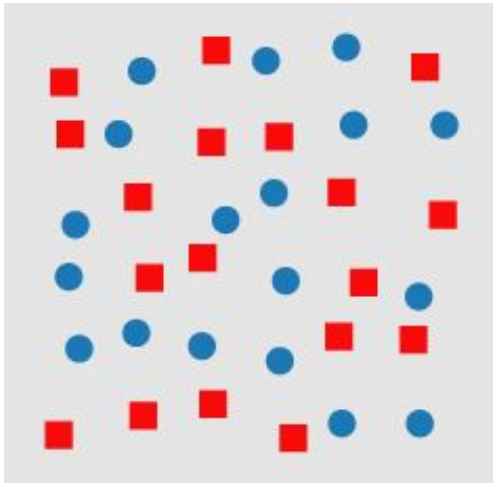


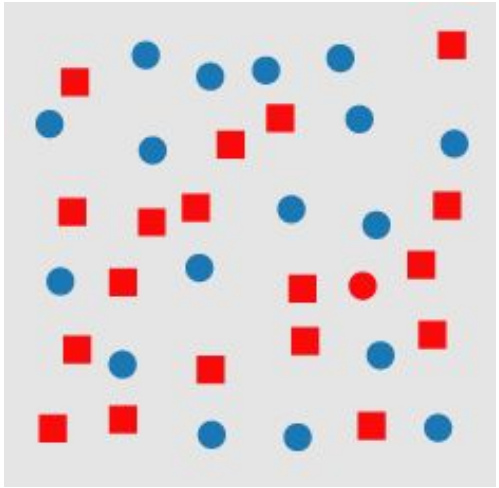


Yes, can be done preattentively

From: Healey, Perception in Visualization

Hue and Shape



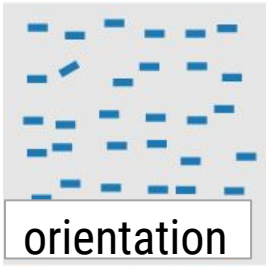


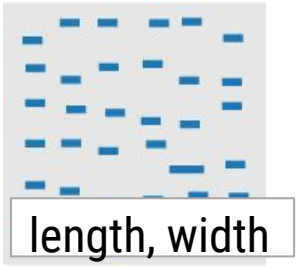
Cannot be done preattentively due to the **conjunction** of shape and hue

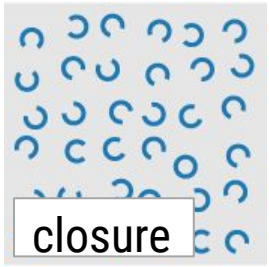
$\longrightarrow$ need to search

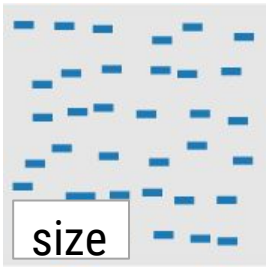
From: Healey, Perception in Visualization

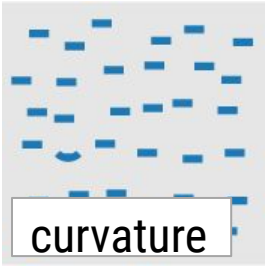
Preattentive visual features (some)

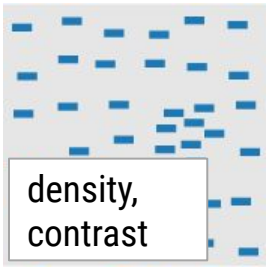


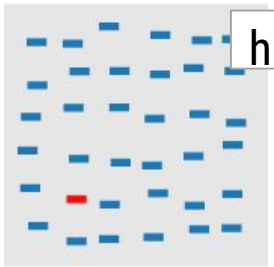


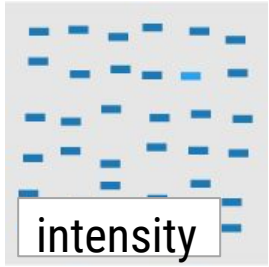


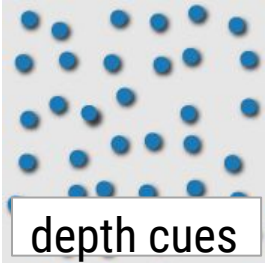


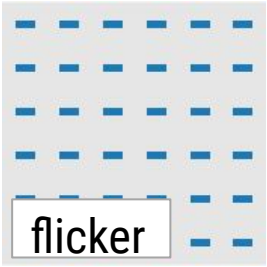


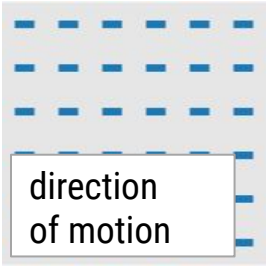


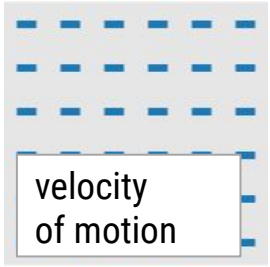


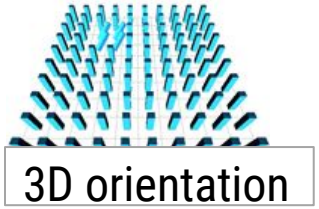


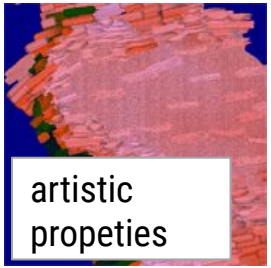








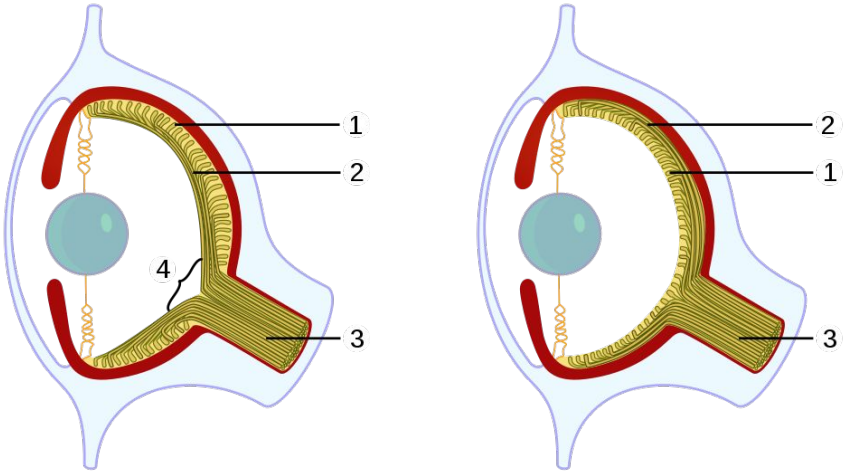




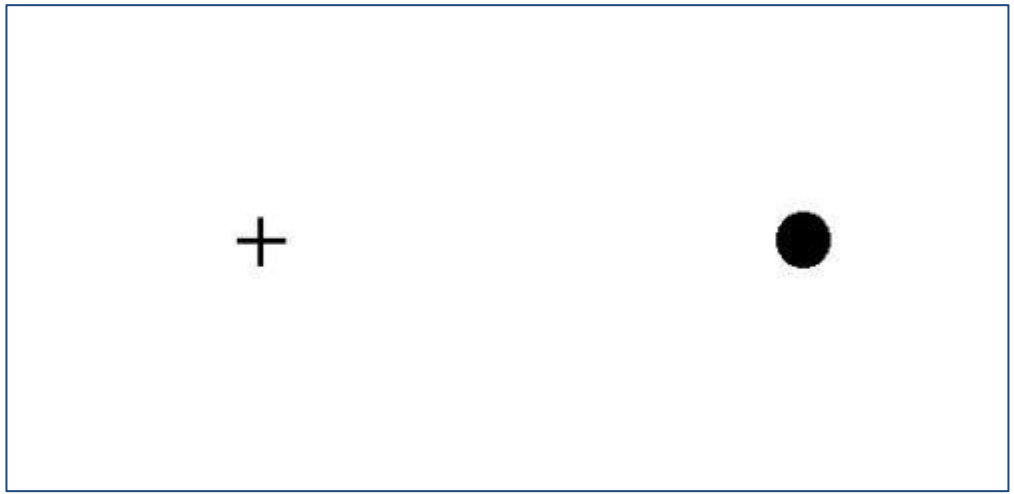
From: Healey, Perception in Visualization

The brain compensates somehow

Blind spot



- Look at graphic of the cross and the circle
- Cover your LEFT eye and stare at the cross with your RIGHT eye.
- SLOWLY move towards the computer screen while still staring at the cross with your RIGHT eye.
- At somewhere around 10-14 inches from the computer screen – the black circle will disappear



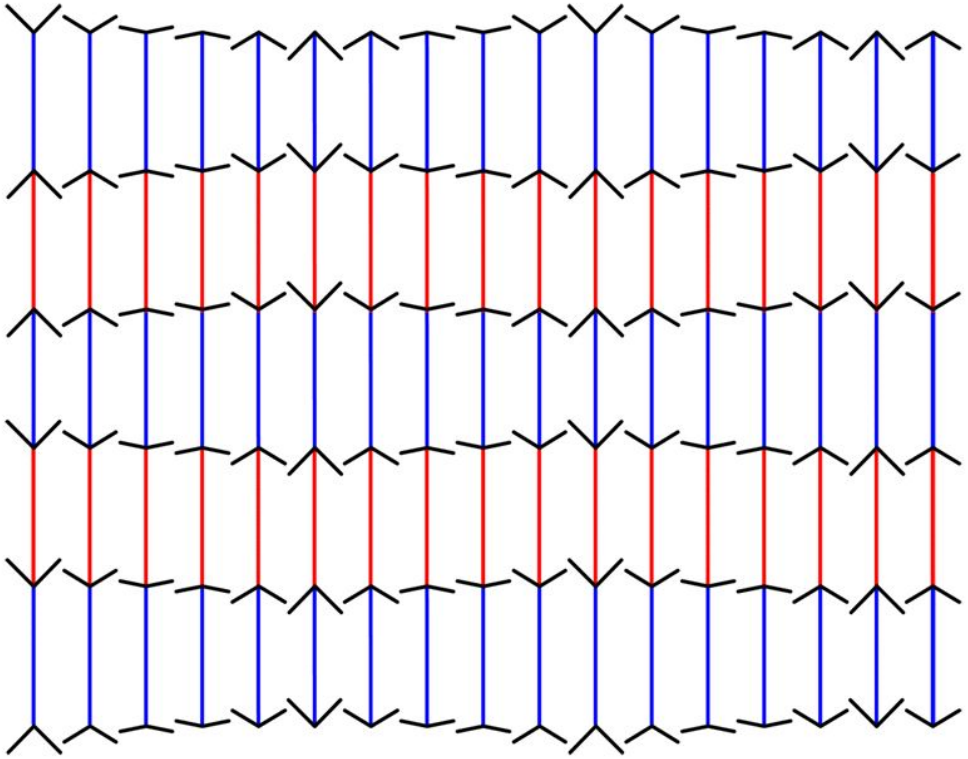
Change blindness

Change blindness

Müller-Lyer Sinusoidal Waves

New variant by Gianni A. Sarcone

Though the **blue** and **red** segments seem to oscillate, they are always the **same length**! **Nothing moves except the arrows** at the endpoints of each color segment...



Conclusion (temporary) on Vision

- We do not understand vision by ourselves
  - Even if we use it all the time
- Many super capabilities
  - Use them!
- Many super inabilities
  - Study them, avoid them!
- Intuition on visual feature to use will fail ALWAYS
  - Follow well-established guidelines

Memory

Three types:

- Sensory memories
- Short-term memory
  - Working memory
- Long-term memory

Sensory memory

- iconic | visual stimuli
- echoic | aural stimuli
- haptic | touch stimuli

- Constantly overwritten
- Info. passes from sensory to STM by attention

Short-term Memory (STM)

- Scratch-pad
- Rapid access
  - 70ms
- rapid decay
  - 200ms
- limited capacity
  - 7 +/- 2 chunks
- Recency effect

- UI should avoid flushing the STM
- Many operations flush it
  - Writing (e.g. SQL)
  - Errors
  - Interruptions
- Keep users in the flow!

Long-term Memory (LTM)

- Repository for all our knowledge
- slow access
  - 1-10 second
- slow decay, if any
- huge or unlimited capacity

Two types:

- Episodic
  - serial memory of events
- Semantic
  - structured memory of facts, concepts, skills
- Semantic LTM derived from episodic LTM

Individual differences

- Long-term
  - Physical and intellectual abilities
- Short-term
  - Stress or fatigue
- Changing
  - Age, illness

Ask: will design decision exclude section of user population?

Input devices

- Keyboard
- Mouse
- Pointer
- Finger Touch
- Voice
- Signal processing (image, video, sound)

- Different devices support different styles of interaction
  - Performance and capabilities differ
  - Compatibility may be possible but at a cost

Mouse / Pointer / Touch

- Handheld pointing device
- Mouse located on desktop
  - requires physical space
  - no arm fatigue
- Relative movement
- Screen cursor follows mouse
- Screen cursor oriented in (x, y) plane,
- mouse movement in (x, z) plane
- Multiple buttons (up to 5)
- Possible to track "proximity"

- Touch over the screen
  - No extra space
  - Arm fatigue
- Absolute movement
- Finger hides the point
- No buttons, just touch
- No proximity
  - But pressure arrives

Movements

- Time to respond to stimulus:
  - reaction time + movement time
- Movement time
  - dependent on age, fitness etc.
- Reaction time
  - dependent on stimulus

Type:

- Visual: 200ms
- Auditory: 150 ms
- Pain: 700ms
- Increasing reaction time decreases accuracy in the *unskilled operators* not in the *skilled* ones.

Movement: Fitts' Law

Movement time = $a + b \log(\text{distance}/\text{size}+1)$
Where a and b are cst dependent of the user and setting.

Practical use:

- Buttons/targets should not be too small
- Important targets should be close to the mouse

Interaction

Interaction: the communication between the user and the system

Donald Norman’s Interaction framework

- User establishes the goal
- Formulates intention
- Specifies actions at interface
- Executes action
- Perceives system state
- Interprets system state
- Evaluates system state with respect to goal

WIMP and Touch-Based

- Windows
- Icons
- Menus
- Pointers

Traditional now on desktop/laptop

Slightly different on tablets and telephones

Usability paradigms and principles

- Designing for maximum usability is the goal of Design
- History of interactive system design provides paradigms for usable designs
- Principles of usability are more general means of understanding usability

- If not sure about principles, stick to established rules

Direct Manipulation

1982 – Shneiderman describes graphically-based interaction

- Visibility of objects
- Incremental action and rapid feedback
- Reversibility encourages exploration
- Syntactic correctness of all actions
- Replace language with action

Principles of good interfaces

Usability

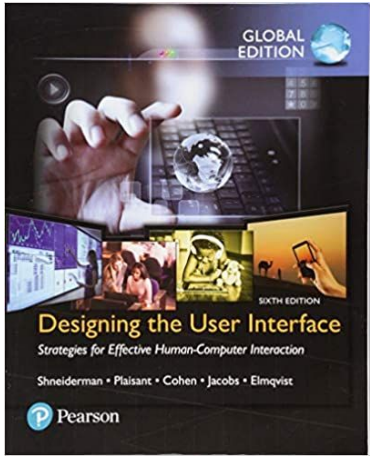
- Learnability
- Flexibility
- Robustness

Learnability

- Predictability
- Synthesizability
- Familiarity
- Generalizability
- Consistency

Take Home Message

You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work
You are not the user, introspection will not work

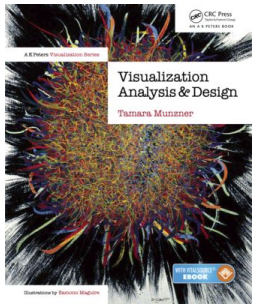


Designing the User Interface: Strategies for Effective Human-Computer Interaction

Ben Shneiderman

Introduction to Information Visualization

Jean-Daniel Fekete (Jean-Daniel.Fekete@inria.fr) +
Petra Isenberg+Tamara Munzner+Chris North

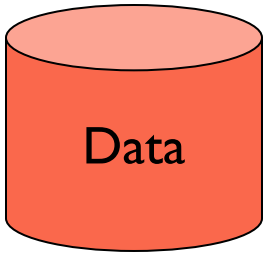


What is Information Visualization?

- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition [Card et al. 99]
- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively [Muzner 14]

The Big Problem

Web,books, papers, game scores,
Scientific data,
Biotech,
Shopping
People
Stock/finance
News
ML Models

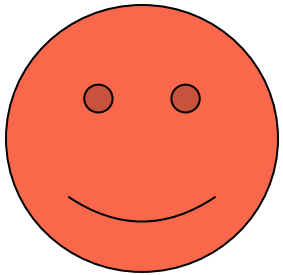


Data Transfer

How?

Vision: 100 MB/s
Ears: <100 b/s
Telepathy
Haptic/tactile
Smell
taste

Human



Human Vision

- Highest bandwidth sense
- Fast, parallel
- Pattern recognition
- Pre-attentive
- Extends memory and cognitive capacity
    - (Multiplication test)
- People think visually

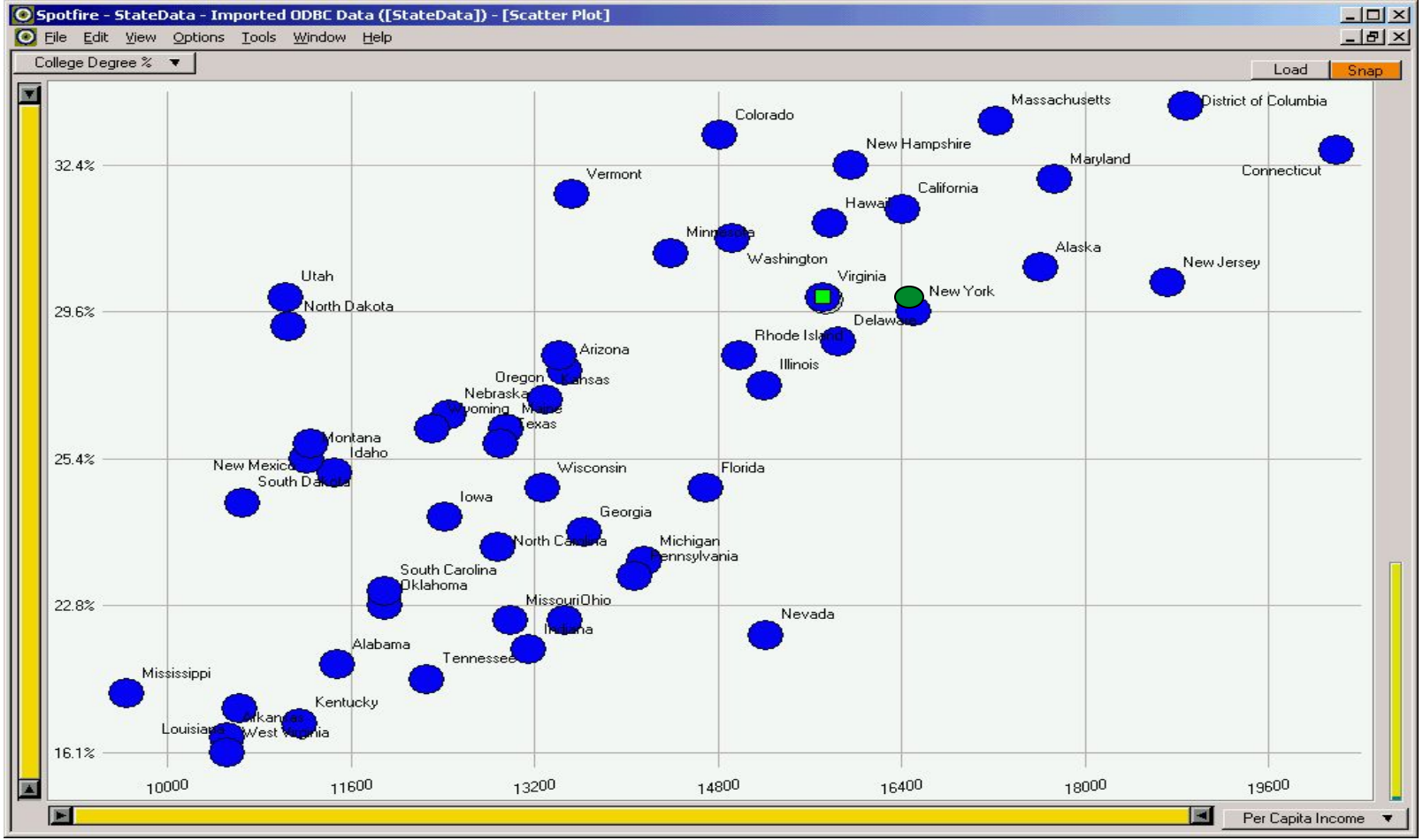
Impressive. Let's use it!

- Which state has highest Income?
- Relationship between Income and Education?
- Outliers?

Table - StateData ()

Load | Snap

State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	24.1%	14154
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923
West Virginia	16.1%	10520
Wisconsin	24.9%	13276
Wyoming	25.7%	12311



College Degree %

Per Capita Income

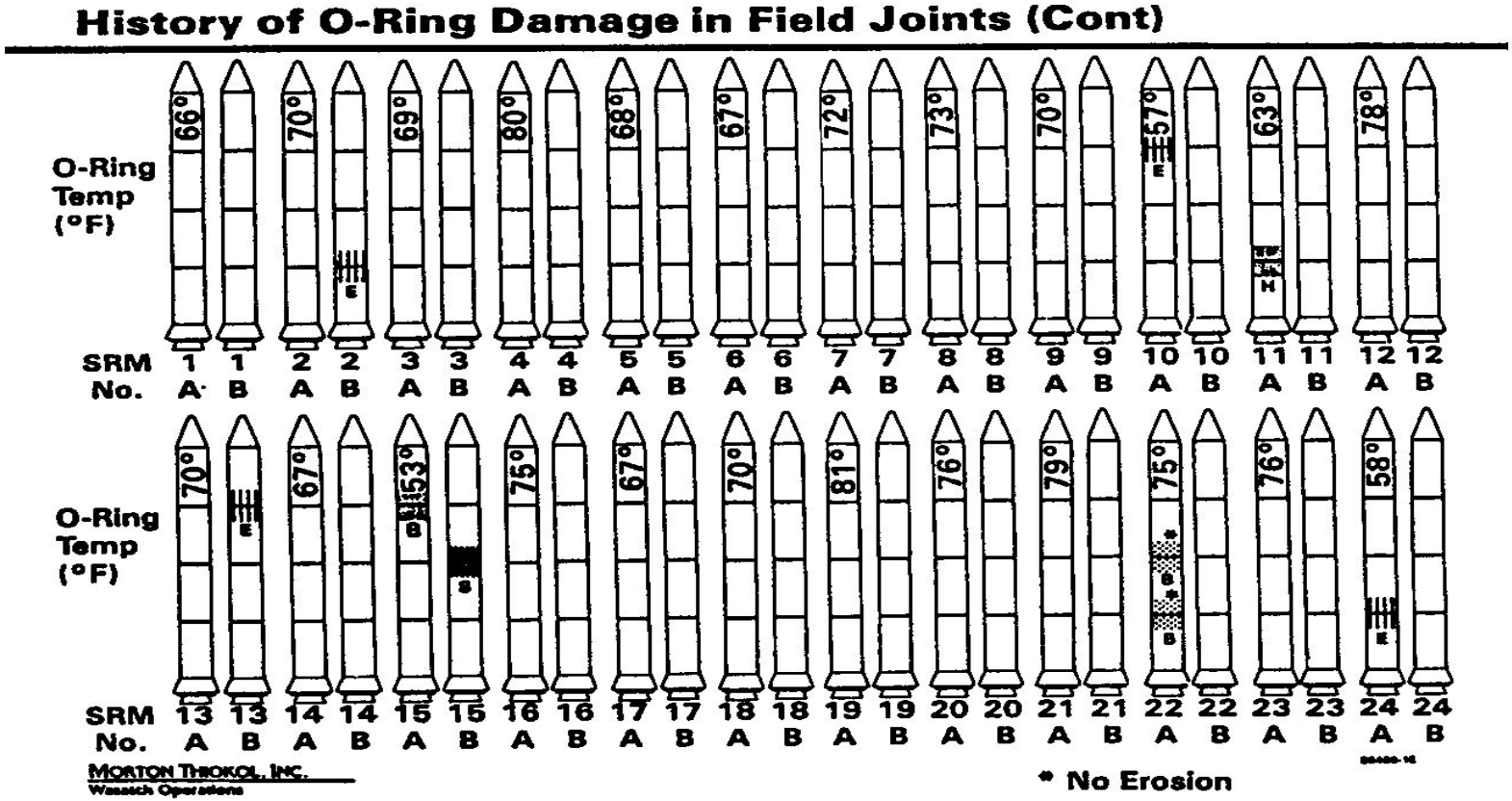
More than just "data transfer"

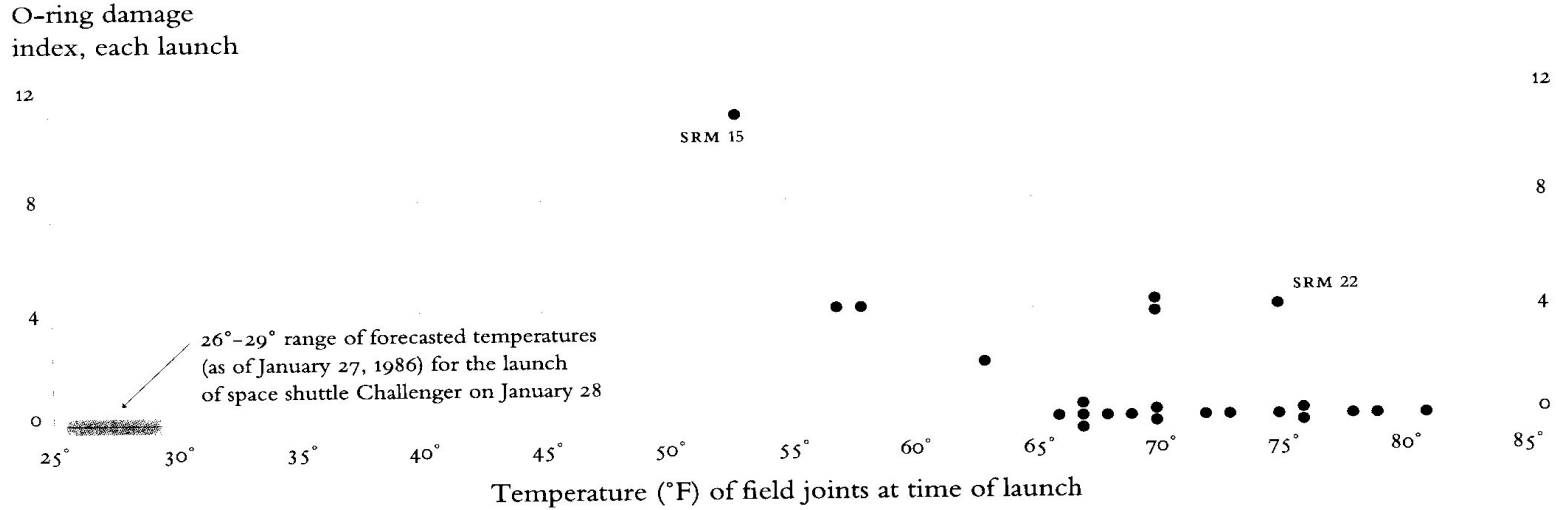
- Support the ability to glean higher level knowledge from the data

- Learn = data → knowledge

Insight!

What's the Big Deal?



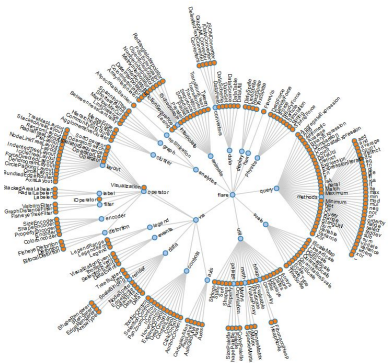


Presentation is everything!

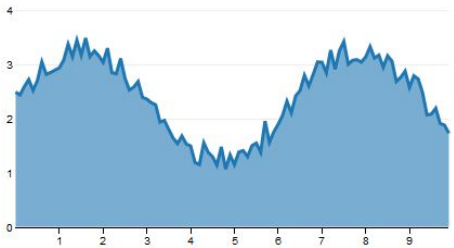
Why visual data representations?

- Vision is our most dominant sense
- We are very good at recognizing visual patterns
- We need to see and understand in order to explain, reason, and make decisions

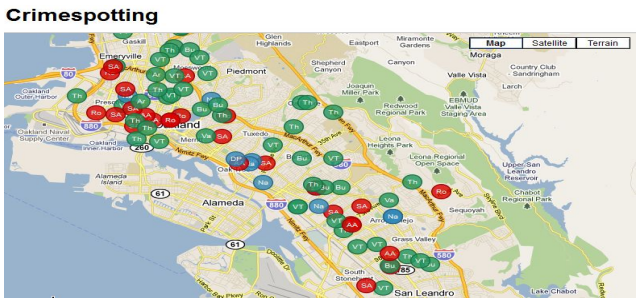
common examples:



graphs / hierarchies



charts

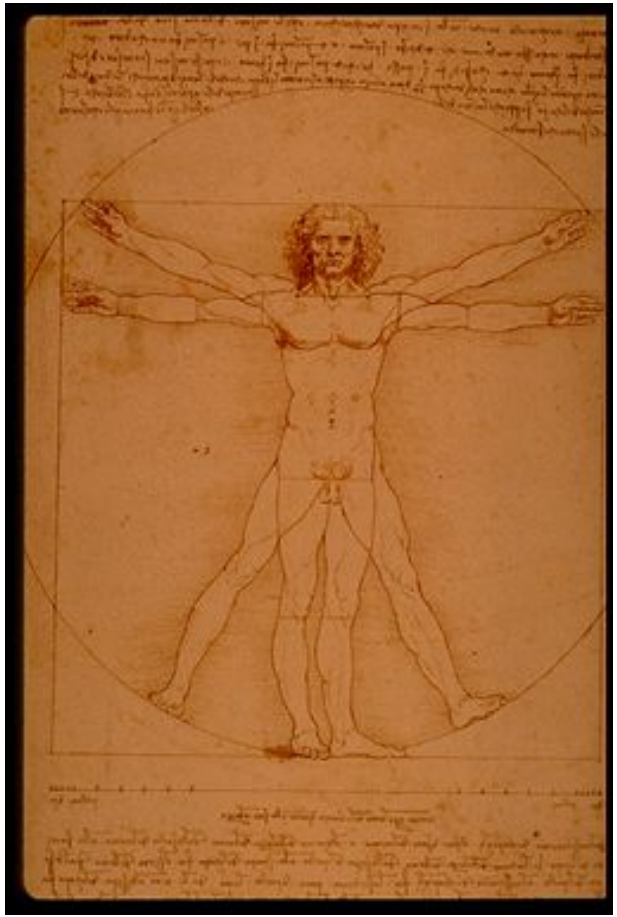


maps

all examples from: http://vis.stanford.edu/protovis/

Other benefits of visualization

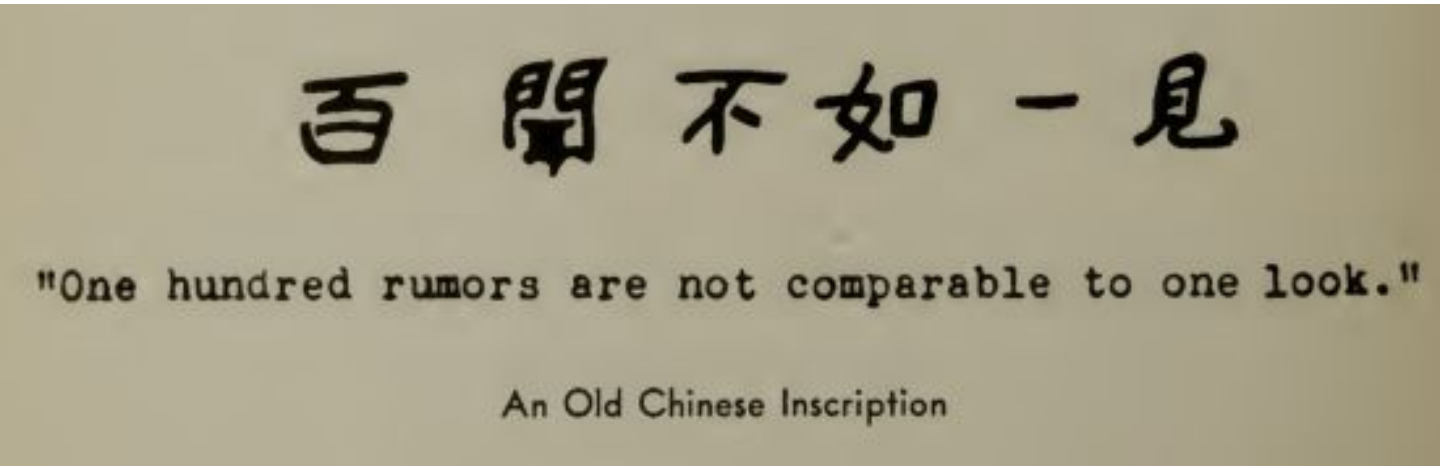
- expand human working memory
  - offload cognitive resources to the visual system,
- reduce search
  - by representing a large amount of data in a small space,
- enhance the recognition of patterns
  - by making them visually explicit
- aid monitoring of a large number of potential events
- provides a manipulable medium & allows exploration of a space of parameter values.



L’occhio,
che si dice finestra dell’anima,
è la principale via donde il comune senso può
piú copiosamente e magnificamente considerare
le infinite opere di natura.

Leonardo da Vinci
(1452 - 1519)

The eye...
the window of the soul,
is the principal means
by which the central sense
can most completely and
abundantly appreciate
the infinite works of nature.



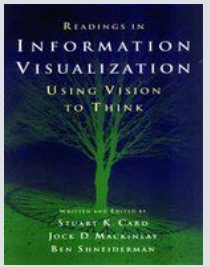
Via Brinton, Graphic Presentation, 1939

Information visualization

- Create visual representation
- Concentrates on abstract data
- Includes interaction

Official Definition:

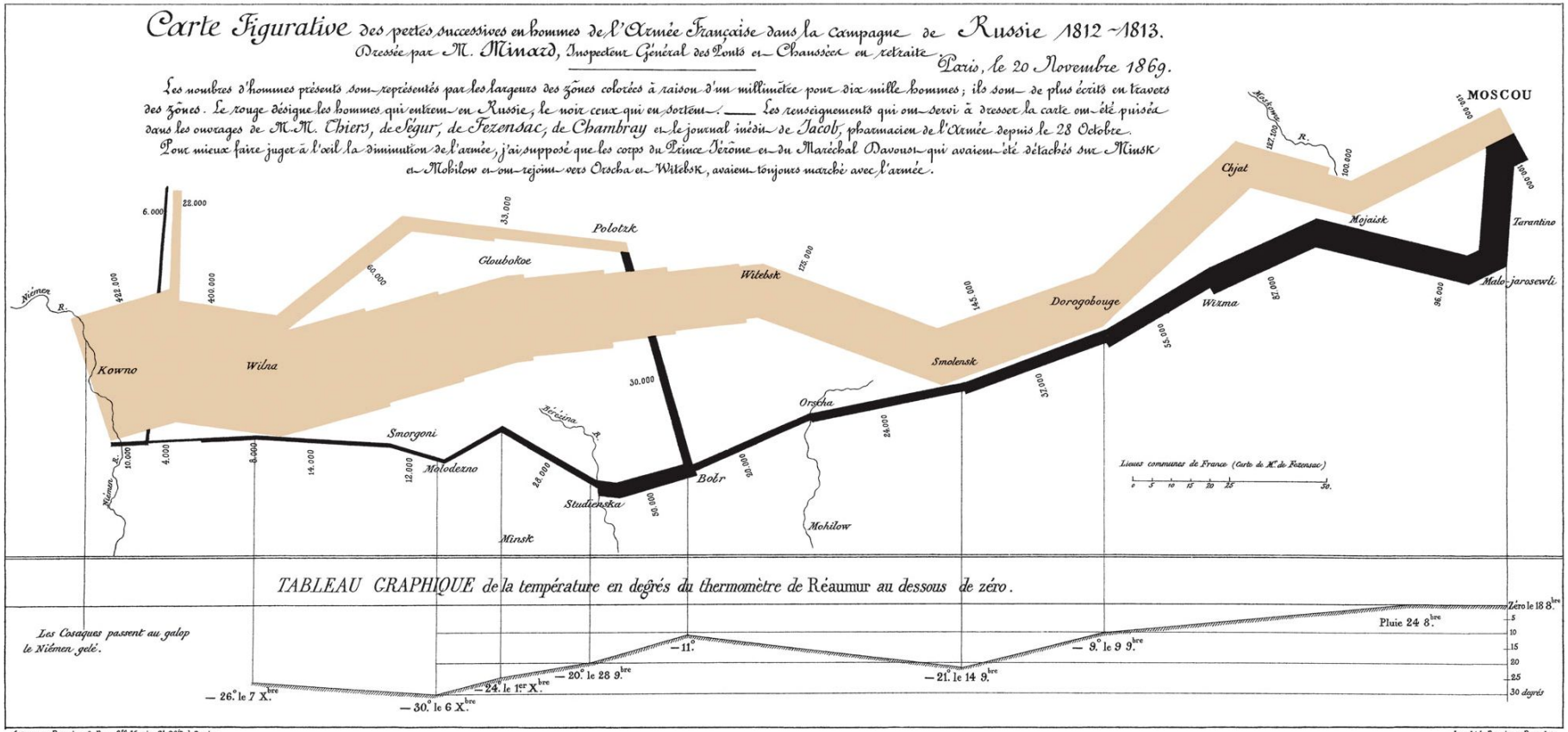
The use of computer-supported, interactive, visual representations of abstract data to amplify cognition. [Card et al., 1999]



Functions of Visualizations

- Recording information
  - Tables, blueprints, satellite images
- Processing information
  - needs feedback and interaction
- Presenting information
  - share, collaborate, revise
  - for oneself, for one's peers and to teach
- Seeing the unseen

History: Static Graphics



The Broadway Street Pump

- In 1854 cholera broke out in London
  - 127 people near Broad Street died within 3 days
  - 616 people died within 30 days
- "Miasma in the atmosphere"
- Dr. John Snow was the first to link contaminated water to the outbreak of cholera
- How did he do it?
  - he talked to local residents
  - identified a water pump as a likely source
  - used maps to illustrate his theory
  - convinced authorities to disable the pump

More info here: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

John Snow, 1854

… AND QUITE RECENTLY

TrashTrack

http://senseable.mit.edu/trashtrack/

Artificial Intelligence

http://www.bewitched.com/chess/

Specific Visualization Environments

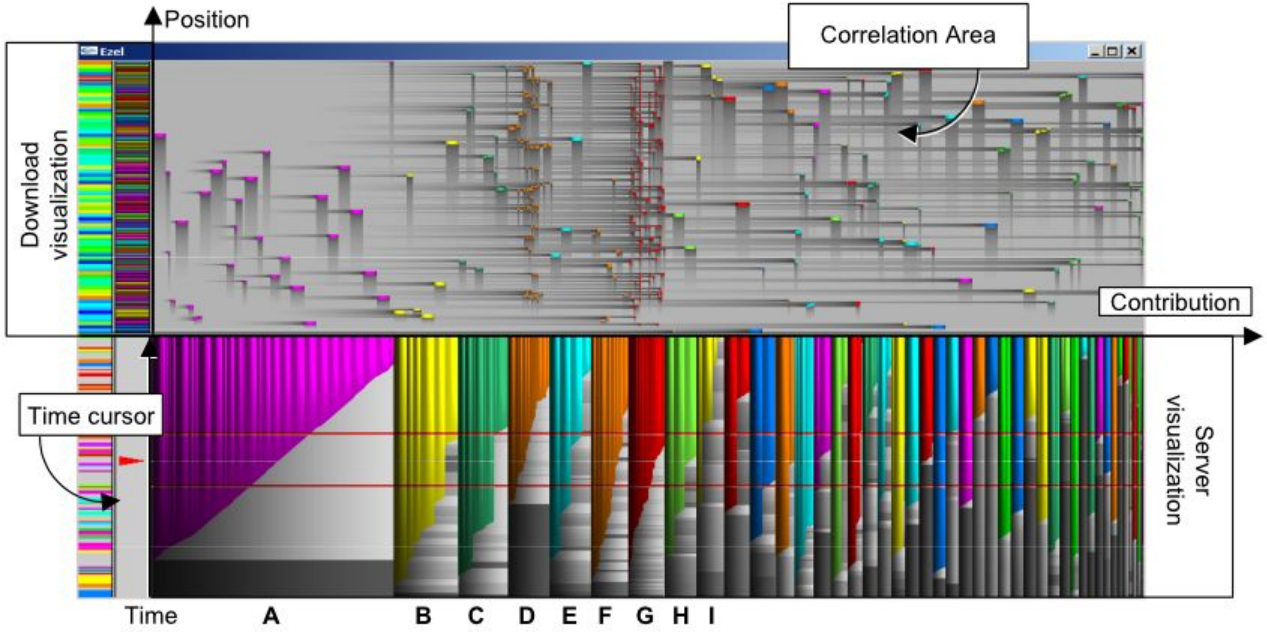
Molecular visualisation in the Reality Cube
University of Groningen, NL

Tabletops for Visualization
University of Calgary

WILD Wall, INRIA

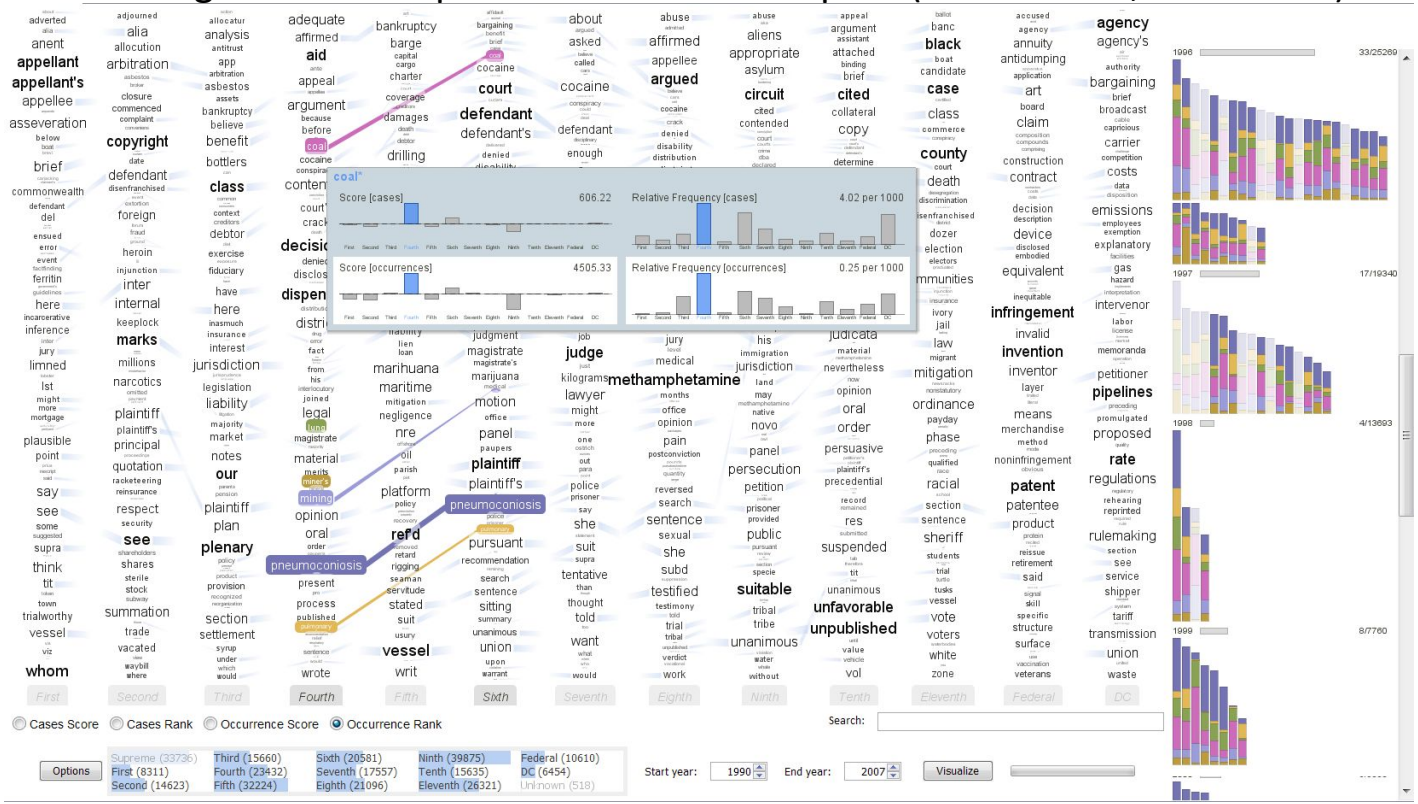
Software Visualization

EZEL: a Visual Tool for Performance Assessment of Peer-to-Peer File-Sharing Networks (Voinea et al., InfoVis, 2004)



Text Visualization

Parallel Tag Clouds to Explore Faceted Text Corpora (Collins et al., VAST 2009)

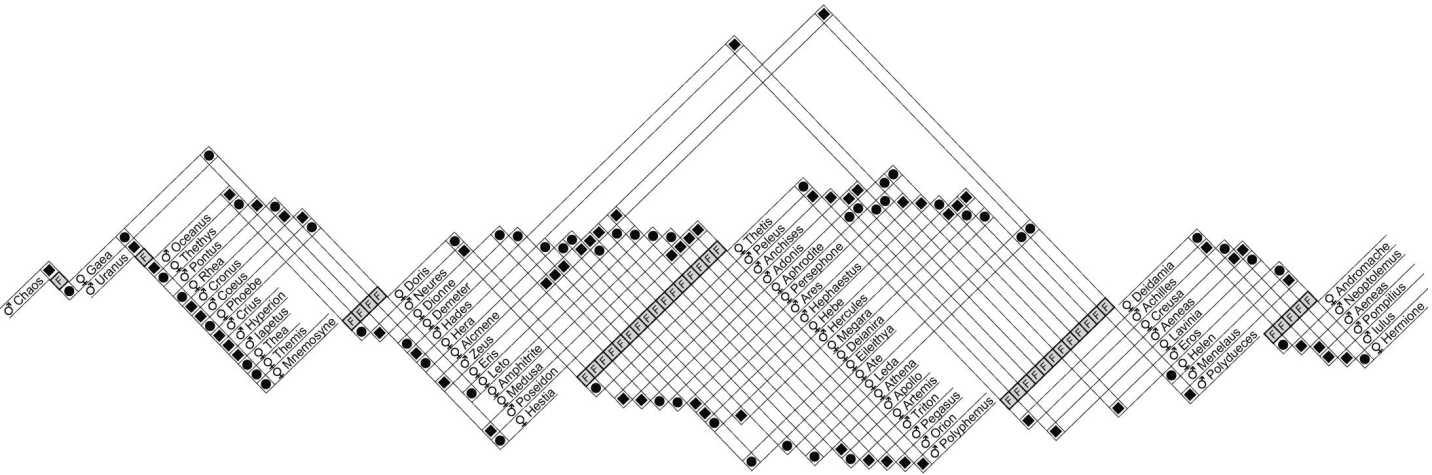


Graphs

http://www.facebook.com/note.php?note_id=469716398919

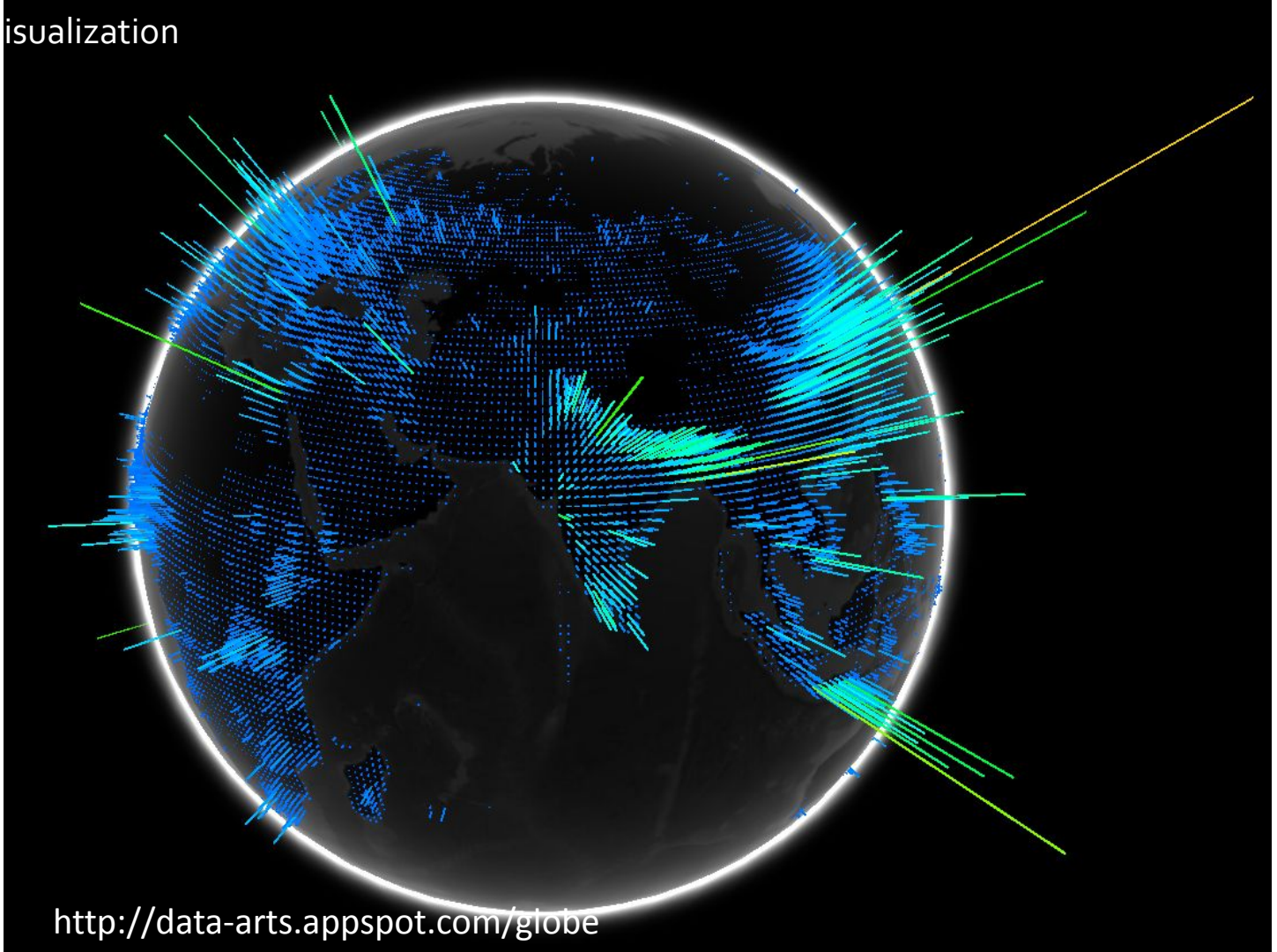
Visualizing Friendships by Paul Butler on Tuesday, December 14, 2010

Family Trees



http://www.aviz.fr/geneaquilts/

isualization

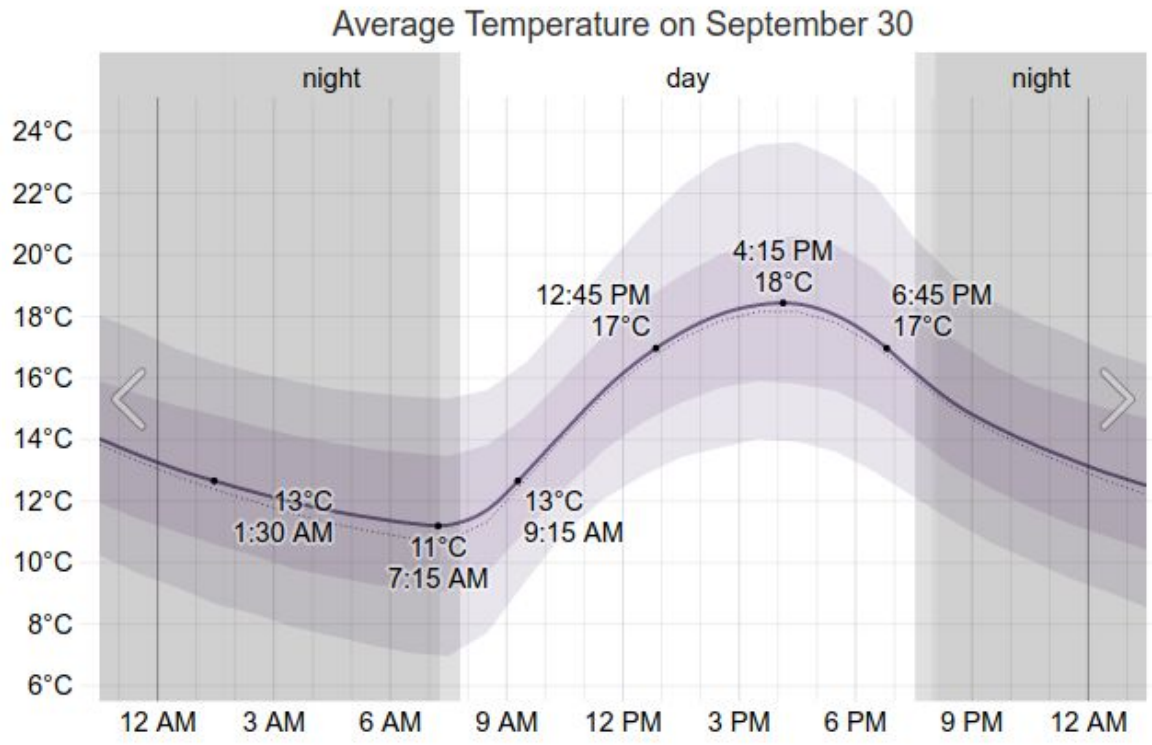


http://data-arts.appspot.com/globe

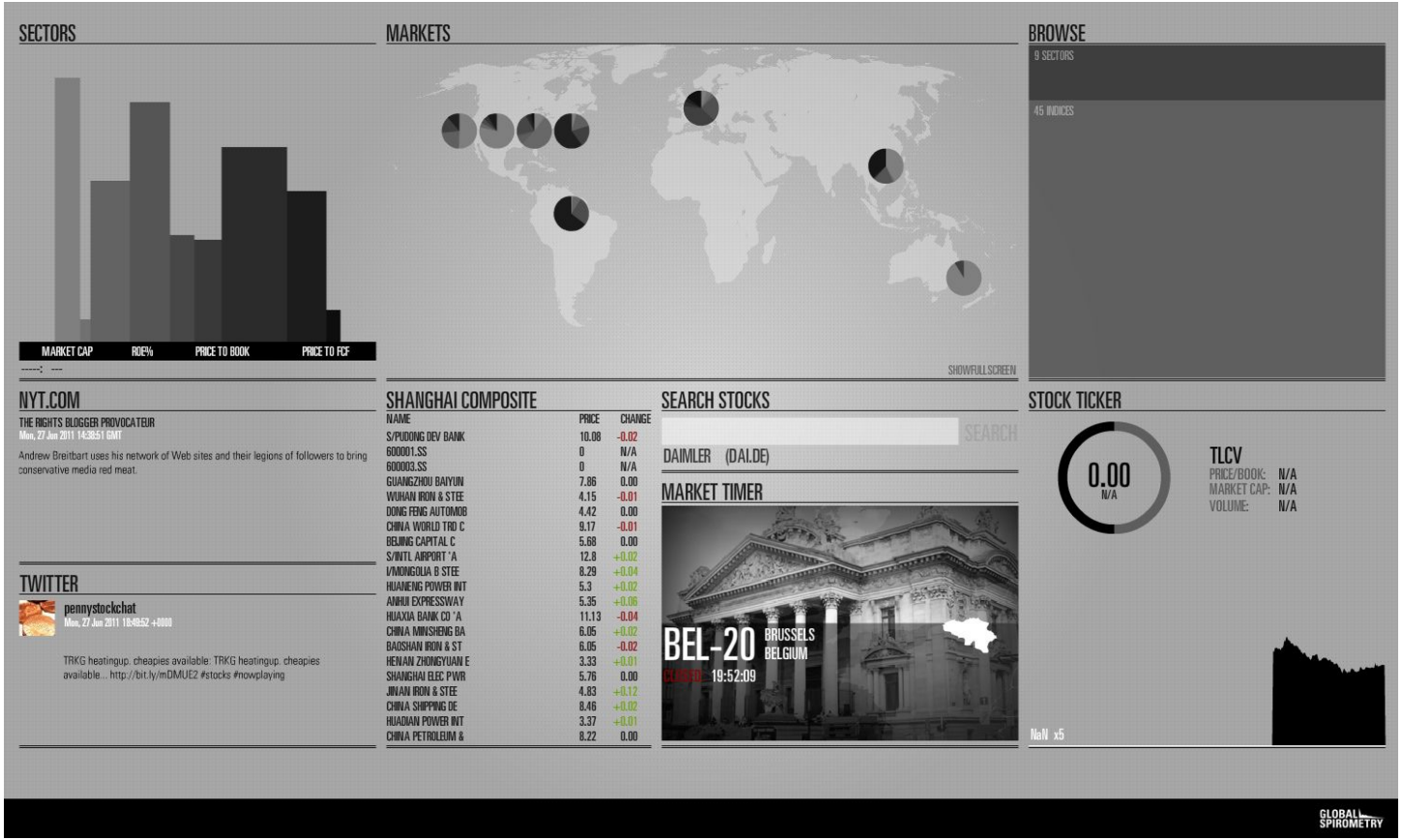
Weather

http://weatherspark.com/

https://weatherspark.com/



Data Dashboards



http://globalspirometry.com

Resources for more examples

- https://sites.google.com/view/visres/
- Visualization conferences
- Blogs
  - http://infosthetics.com/
  - http://fellinlovewithdata.com/
  - http://eagereyes.org/
  - http://flowingdata.com/
  - http://www.informationisbeautiful.net/
- Examples
  - Beautiful Data (McCandless)
  - Now You See it (Few)
  - Tufte Books: Visual Display of Quantitative Information (and others)
  - … (many more, ask me for details)

- Textbooks
  - **Tamara Munzner. Visualization Analysis and Design. A K Peters Visualization Series, CRC Press, 2014.**
  - Readings in Information Visualization: Using Vision to Think (a bit old now but good intro)
  - Information Visualization (Robert Spence – a light intro, I recommend as a start)
  - Information Visualization Perception for Design (Colin Ware, focused on perception and cognition)
  - Interactive Data Visualization: Foundations, Techniques, and Applications (Ward et al. – most recent)

It is difficult to create

CREATE VISUALIZATIONS

Why have a human in the loop?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

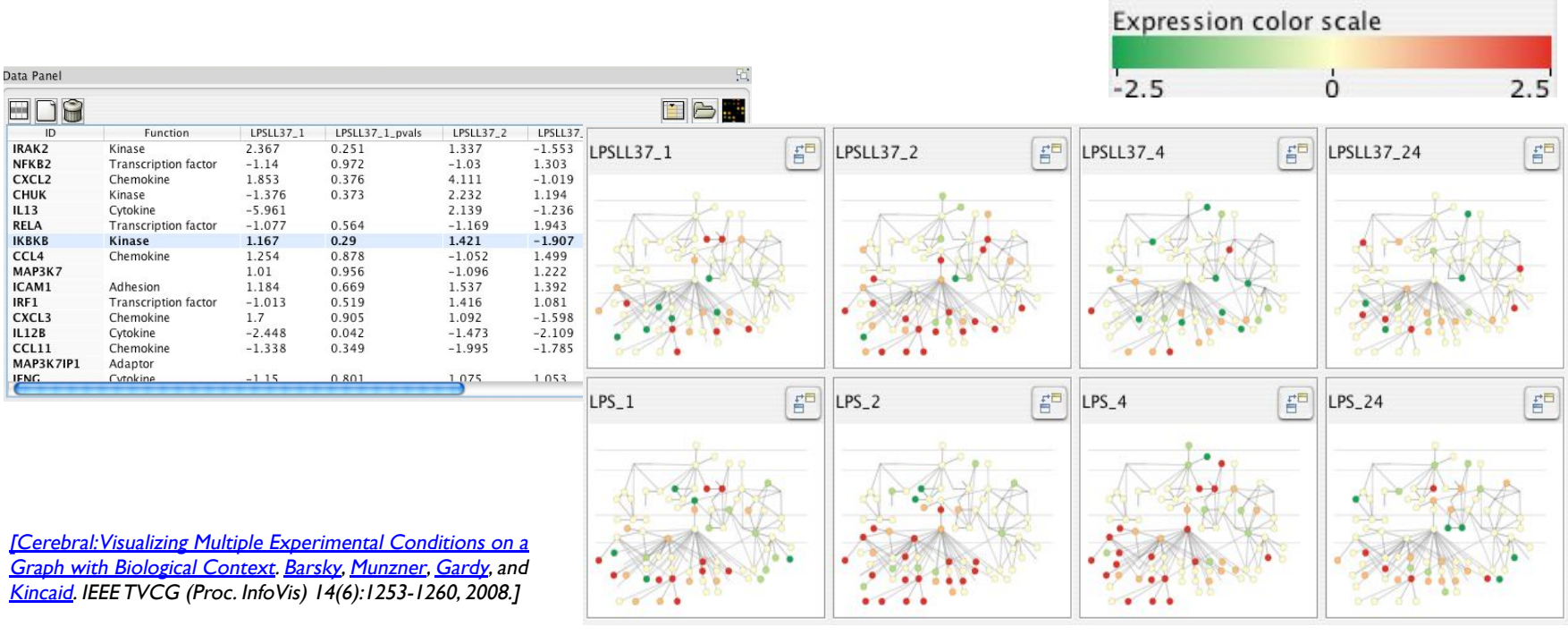
Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- don't need vis when fully automatic solution exists and is trusted
- many analysis problems ill-specified
  - don't know exactly what questions to ask in advance
- possibilities
  - long-term use for end users (ex: exploratory analysis of scientific data)
  - presentation of known results (ex: New York Times Upshot)
  - stepping stone to better understanding of requirements before developing models
  - help developers of automatic solution refine/debug, determine parameters
  - help end users of automatic solutions verify, build trust

Why use an external representation?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- external representation: replace cognition with perception



[Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context. Barsky, Munzner, Gardy, and Kincaid. IEEE TVCG (Proc. InfoVis) 14(6):1253-1260, 2008.]

Why represent all the data?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, details matter
  - confirm expected and find unexpected patterns
  - assess validity of statistical model

Example

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

Raw Data from Anscombe's Quartet

[Source: Anscombe's quartet, Wikipedia]

Statistical Analysis

For all four columns, the statistics are identical

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

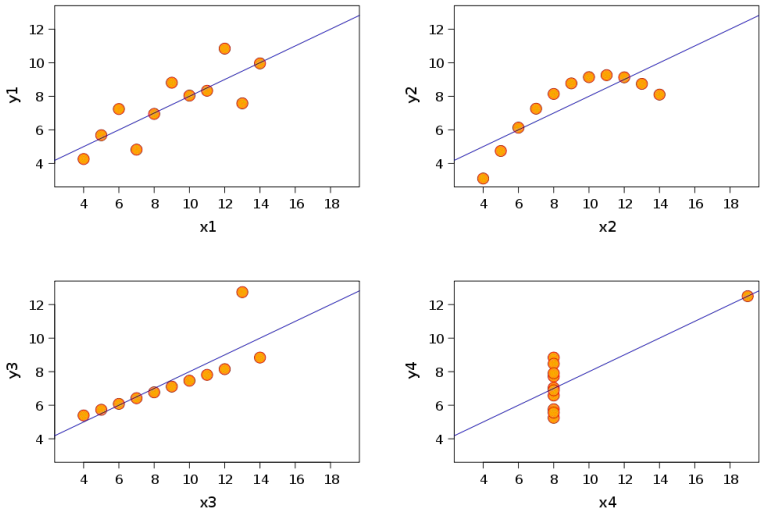
Mean of x	9.0
Variance of x	11.0
Mean of y	7.5
Variance of y	4.12
Correlation between x and y	0.816
Linear regression line	$y = 3 + 0.5x$

[Source: Anscombe's quartet, Wikipedia]

Visual Representation of the Data

Visual representation reveals a different story

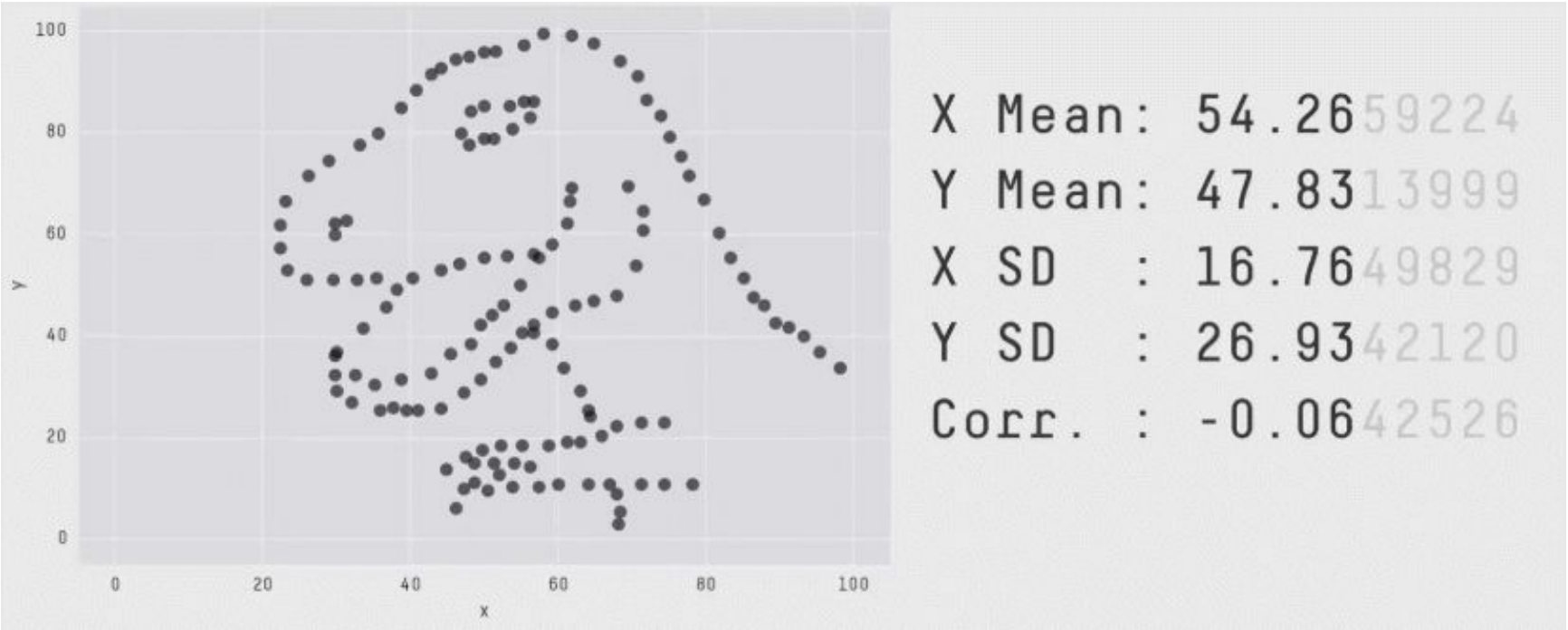
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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



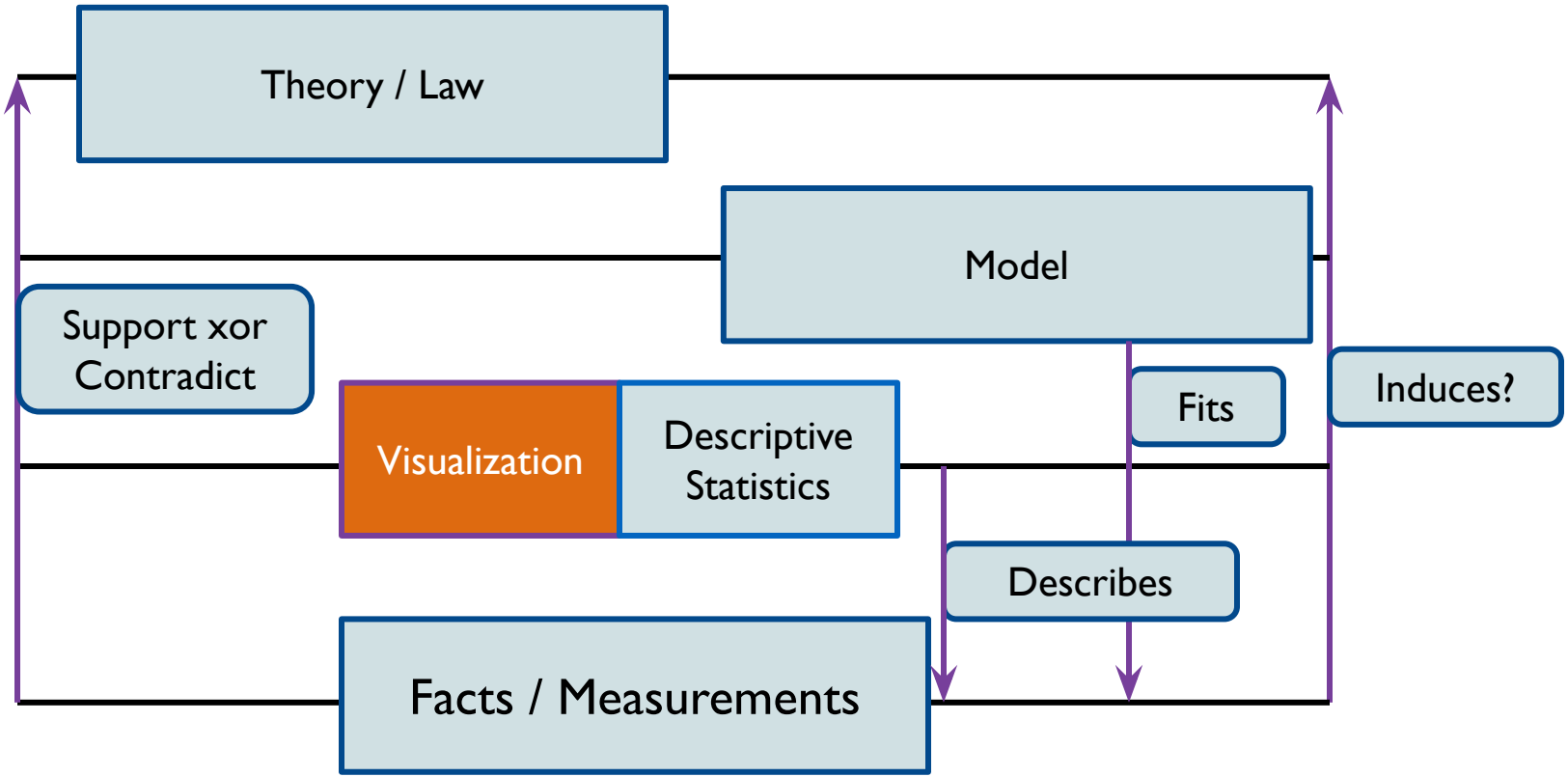
[Source: Anscombe's quartet, Wikipedia]

Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing [CHI17]

https://www.autodeskresearch.com/publications/samestats



Where does Visualization Stands?

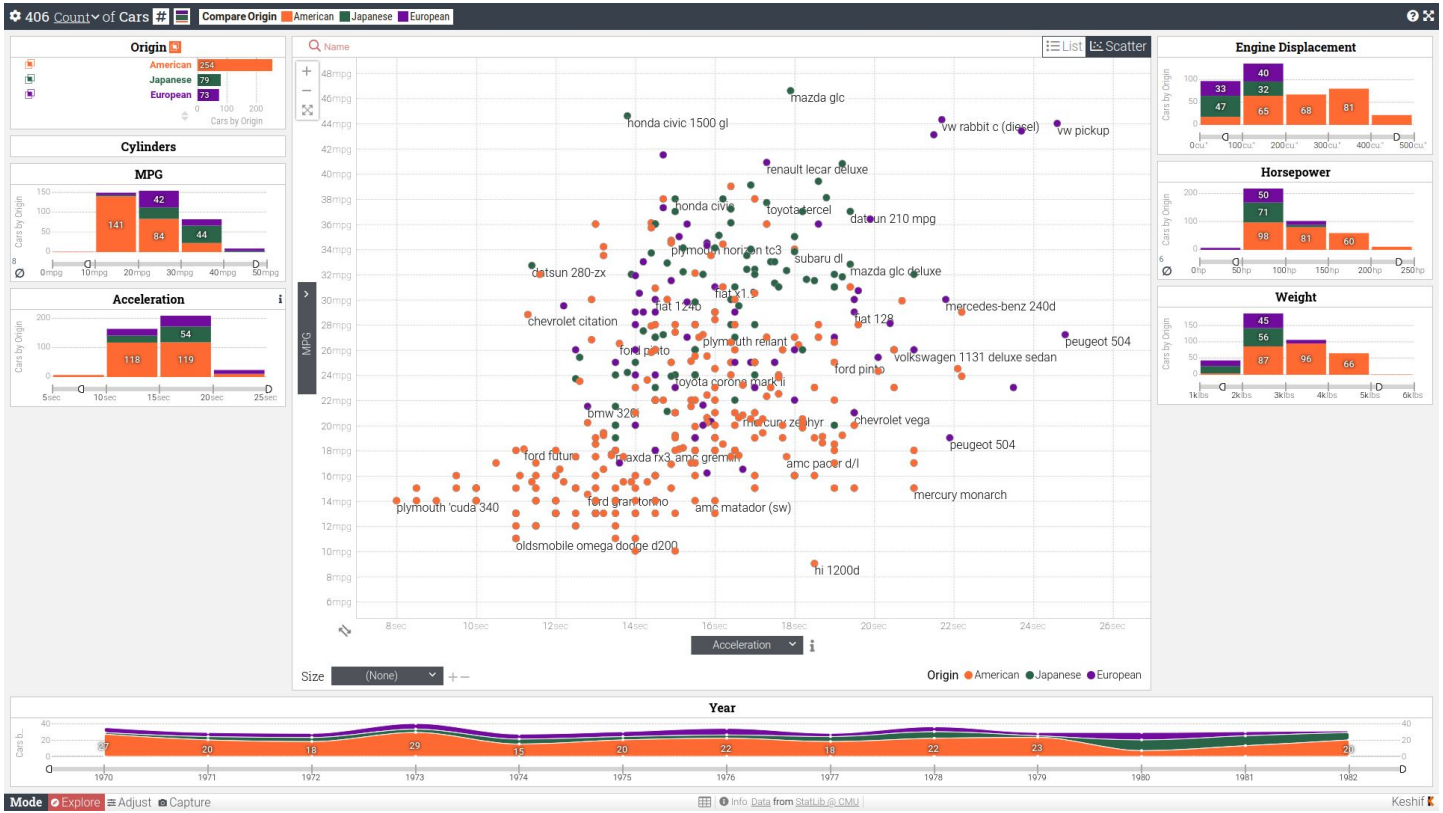


Recap

- So far you
  - learned what information visualization is
  - learned about the advantages of visualization
  - saw a number of examples (historical and new)

- Next
  - you will get to know your data
  - you will learn about the basic components of visualization

Visualization is Interactive



https://gallery.keshif.me/cars

Cars data structure

	A	B	C	D	E	F	G	H	I
1	Car	MPG	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model	Origin
2	Chevrolet Chevelle Malibu	18	8	307	130	3504	12	70	US
3	Buick Skylark 320	15	8	350	165	3693	11,5	70	US
4	Plymouth Satellite	18	8	318	150	3436	11	70	US
5	AMC Rebel SST	16	8	304	150	3433	12	70	US
6	Ford Torino	17	8	302	140	3449	10,5	70	US
7	Ford Galaxie 500	15	8	429	198	4341	10	70	US
8	Chevrolet Impala	14	8	454	220	4354	9	70	US
9	Plymouth Fury iii	14	8	440	215	4312	8,5	70	US
10	Pontiac Catalina	14	8	455	225	4425	10	70	US
11	AMC Ambassador DPL	15	8	390	190	3850	8,5	70	US
12	Citroen DS-21 Pallas	0	4	133	115	3090	17,5	70	Europe
13	Chevrolet Chevelle Concours (sw)	0	8	350	165	4142	11,5	70	US
14	Ford Torino (sw)	0	8	351	153	4034	11	70	US
15	Plymouth Satellite (sw)	0	8	383	175	4166	10,5	70	US
16	AMC Rebel SST (sw)	0	8	360	175	3850	11	70	US
17	Dodge Challenger SE	15	8	383	170	3563	10	70	US
18	Plymouth 'Cuda 340	14	8	340	160	3609	8	70	US
19	Ford Mustang Boss 302	0	8	302	140	3353	8	70	US
20	Chevrolet Monte Carlo	15	8	400	150	3761	9,5	70	US
21	Buick Estate Wagon (sw)	14	8	455	225	3086	10	70	US
22	Toyota Corolla Mark ii	24	4	113	95	2372	15	70	Japan
23	Plymouth Duster	22	6	198	95	2833	15,5	70	US
24	AMC Hornet	18	6	199	97	2774	15,5	70	US
25	Ford Maverick	21	6	200	85	2587	16	70	US
26	Datsun PL510	27	4	97	88	2130	14,5	70	Japan
27	Volkswagen 1131 Deluxe Sedan	26	4	97	46	1835	20,5	70	Europe
28	Peugeot 504	25	4	110	87	2672	17,5	70	Europe

Data

- Data is the foundation of any visualization
- The visualization designer needs to understand
  - the data properties
  - know what meta-data is available
  - know what people want from the data

Datasets

Data Types

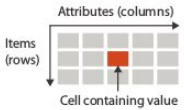
→ Items → Attributes → Links → Positions → Grids

Data and Dataset Types

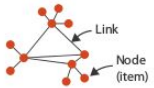
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

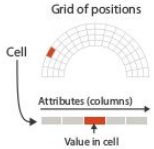
→ Tables



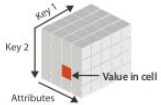
→ Networks



→ Fields (Continuous)



→ *Multidimensional Table*



→ *Trees*

→ Geometry (Spatial)

Dataset Availability

→ Static

→ Dynamic

Attributes

Attribute Types

→ Categorical

→ Ordered

→ *Ordinal*

→ *Quantitative*

Ordering Direction

→ Sequential

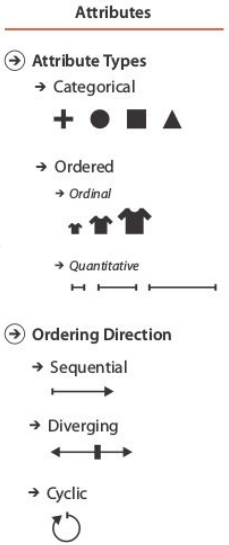
→ Diverging

→ Cyclic

Nominal, Categorical, Ordinal,and Quantitative

- Nominal (labels)
  - Fruits: apples, oranges (open ended)
- Categorical
  - Male, Female (closed)
- Ordered
  - Quality of meat: grade A, AA, AAA
  - Can be counted and ordered, but not measured

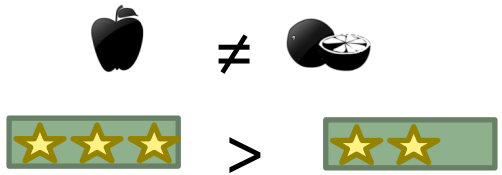
- Quantitative: Interval
  - no clear zero (or arbitrary)
  - e.g. dates, longitude, latitude
  - usually compare differences (intervals)
- Quantitative: Ratio
  - meaningful origin (zero)
  - physical measurements (temperature, mass, length)
  - counts and amounts

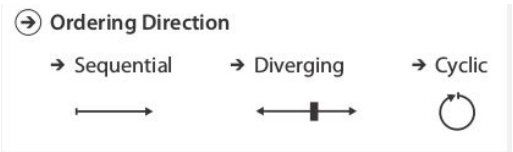


S.S. Stevens, On the theory of scales of measurements, 1946

Nominal, Ordinal and Quantitative

- Nominal (labels)
  - Operations: =, ≠
- Ordered
  - Operations: =, ≠, <, >
- Quantitative: Interval
  - Operations: =, ≠, <, >, -, +
  - Can measure distances or spans
- Quantitative: Ratio
  - Operations: =, ≠, <, >, - , +, ×, ÷
  - Can measure ratios or proportions





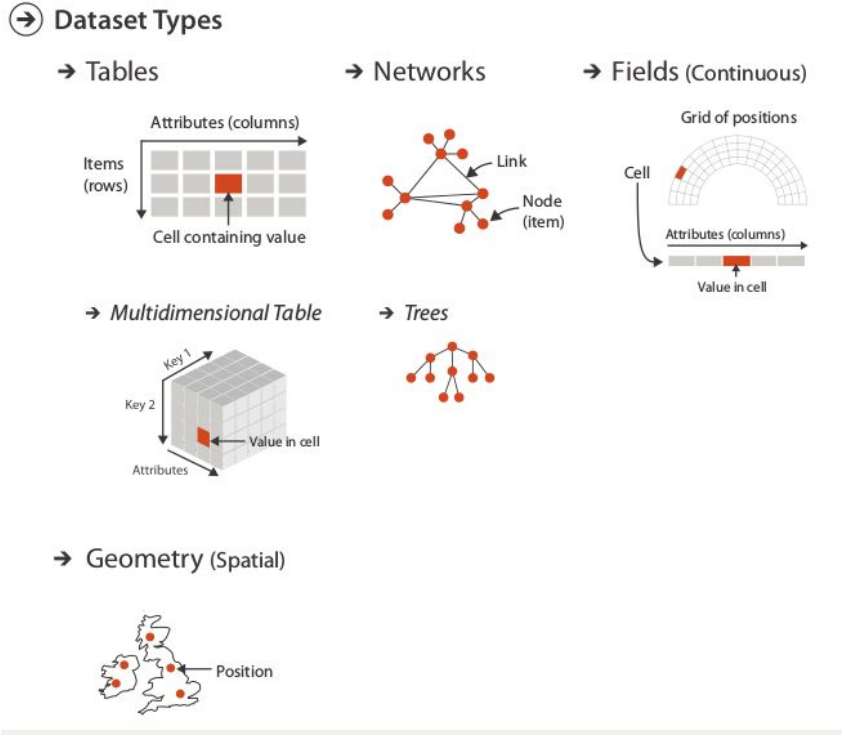
[1989 – 1999] + [2002 – 2012]

10kg / 5kg

S.S. Stevens, On the theory of scales of measurements, 1946

Why is this important?

- Nominal, ordinal, and quantitative data are best expressed in different ways visually
- They can be combined according to their Dataset Types
- But they can also be mapped visually through visual channels



Mapping attributes

•

Map
from categorical and ordered attributes

→ Color

→ *Hue* → *Saturation* → *Luminance*

→ Size, Angle, Curvature, ...

→ Shape

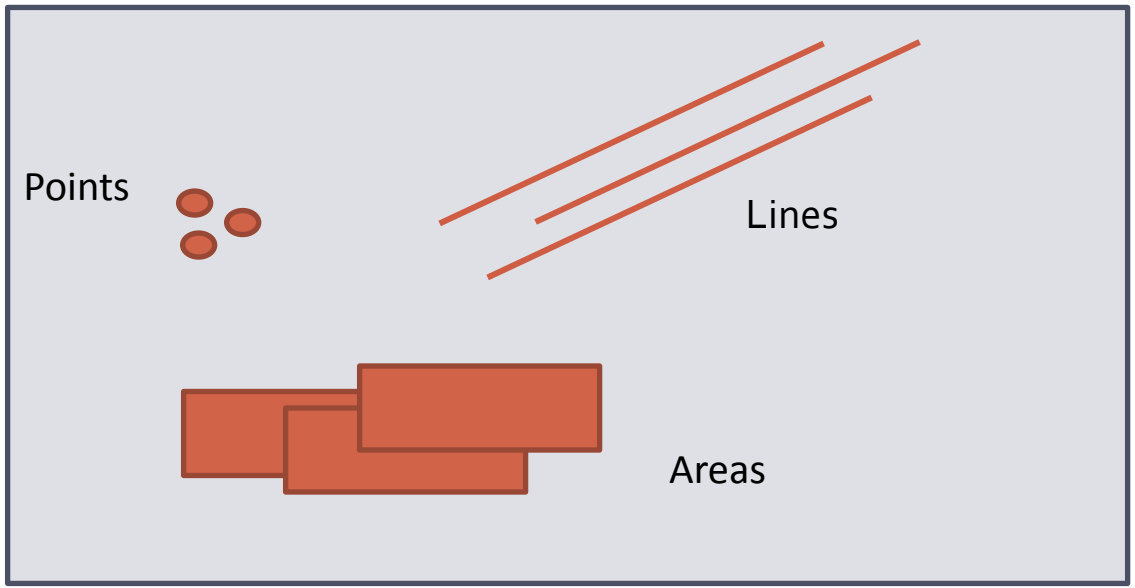
→ Motion
Direction, Rate, Frequency, ...

Why is this important?

- Nominal, ordinal, and quantitative data are best expressed in different ways visually
- Data types often have inherent tasks
  - temporal data (comparison of events)
  - trees (understand parent-child relationships)
  - …
- But:
  - any data type (1D, 2D,…) can be expressed in a multitude of ways!

Visualization's Main Building Blocks

Marks which represent:

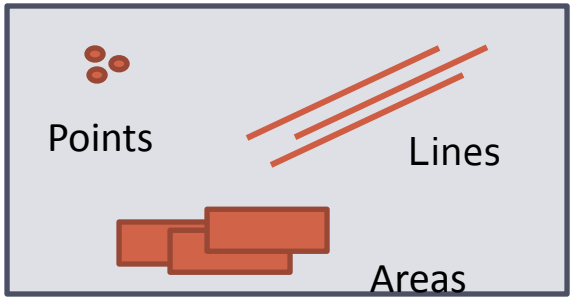


From Semiology of Graphics (Bertin)

The following slides on the topic adapted from Sheelagh Carpendale

Points

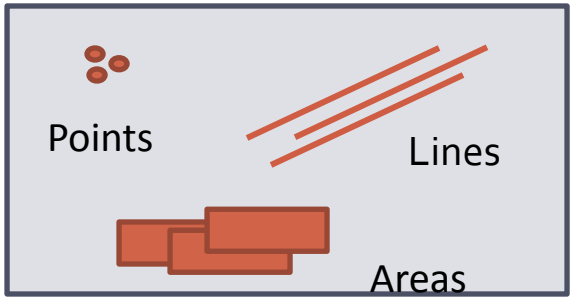
- "A point represents a location on the plane that has **no theoretical length or area**. This signification is independent of the size and character of the mark which renders it visible."
- a location
- marks that indicate points can vary in all visual variables



From Semiology of Graphics (Bertin)

Lines

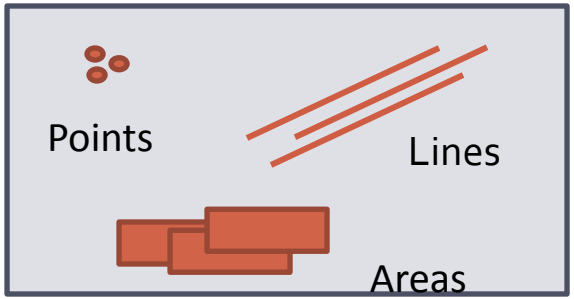
- "A line signifies a phenomenon on the plane which has **measurable length but no area**. This signification is independent of the width and characteristics of the mark which renders it visible."
- a boundary, a route, a connection



From Semiology of Graphics (Bertin)

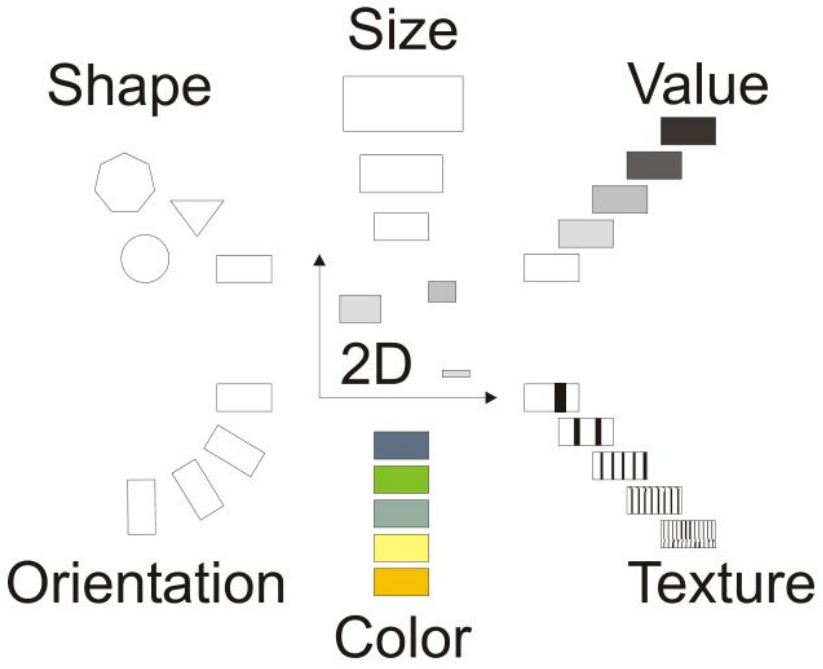
Areas

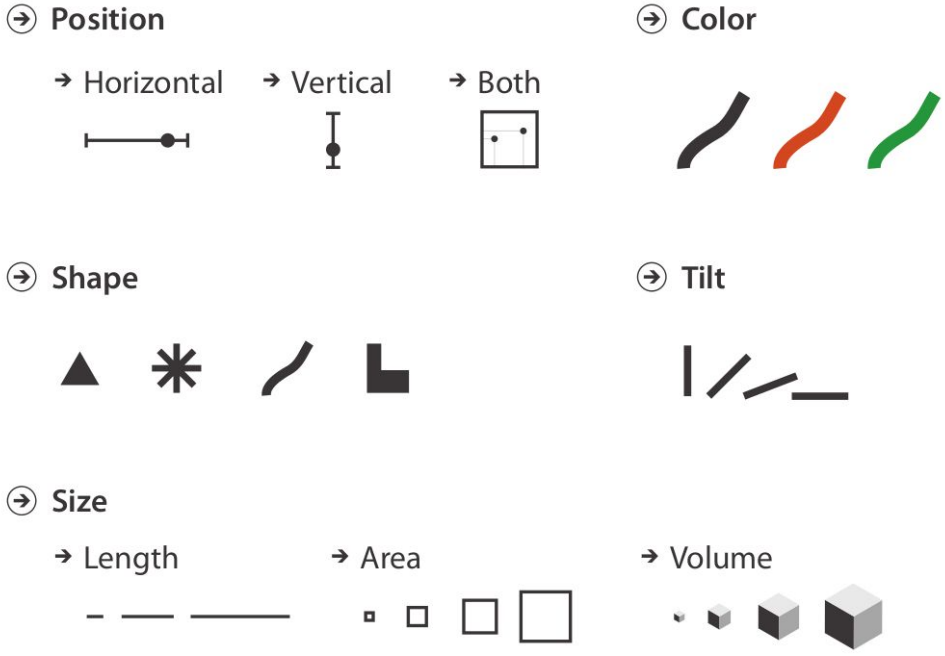
- "An area signifies something on the plane that **has measurable size**.
  This signification applies to the entire area covered by the visible mark."
- an area can change in position but not in size, shape or orientation without making the area itself have a different meaning



From Semiology of Graphics (Bertin)

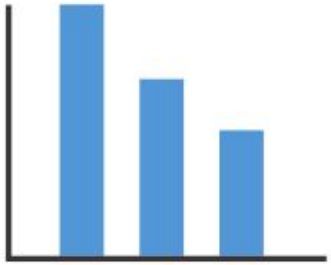
Visual Variables Applicable to Marks

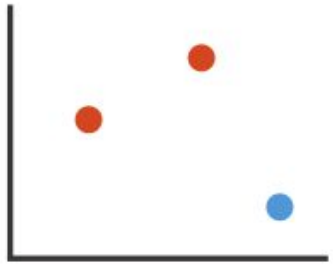


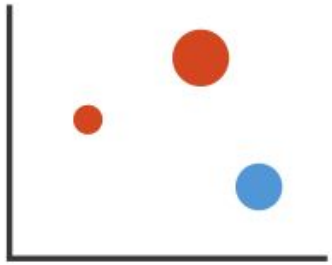


From Semiology of Graphics (Bertin)

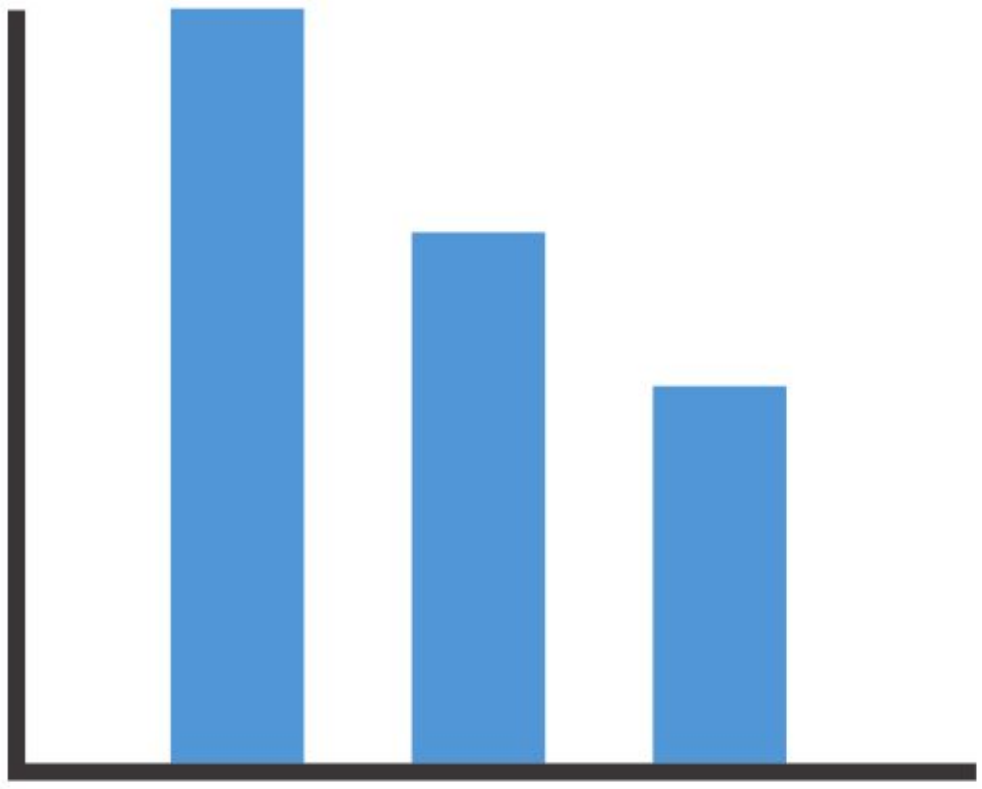
Practice...



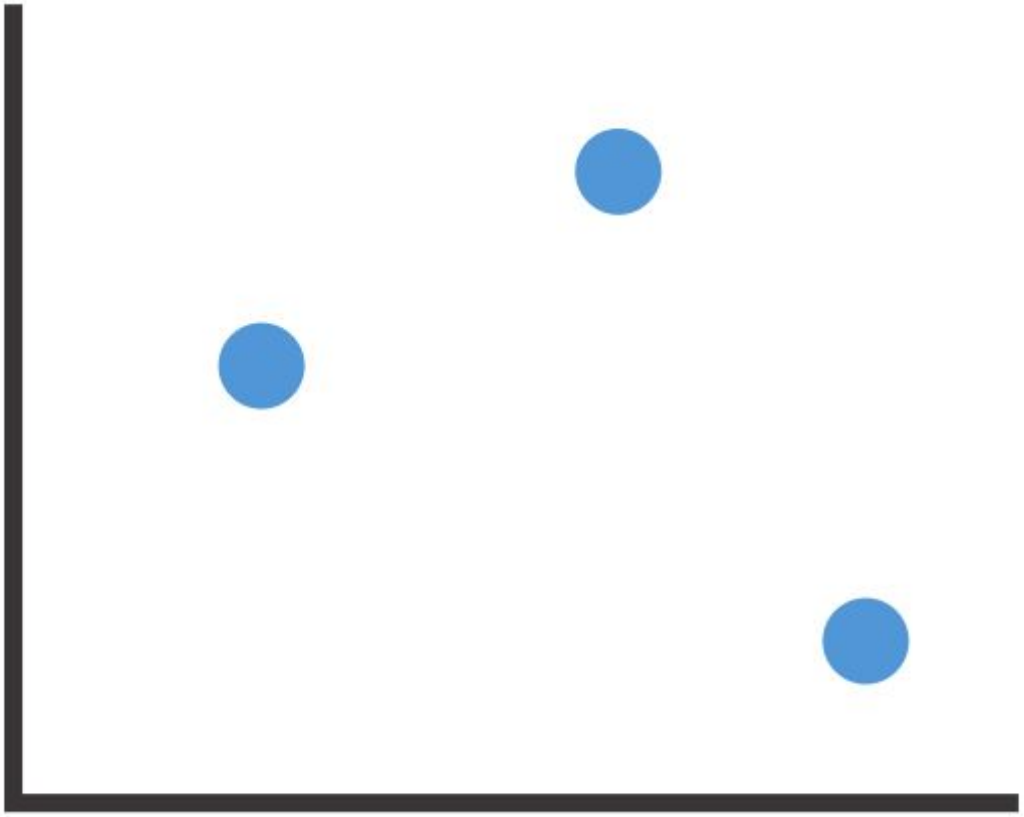




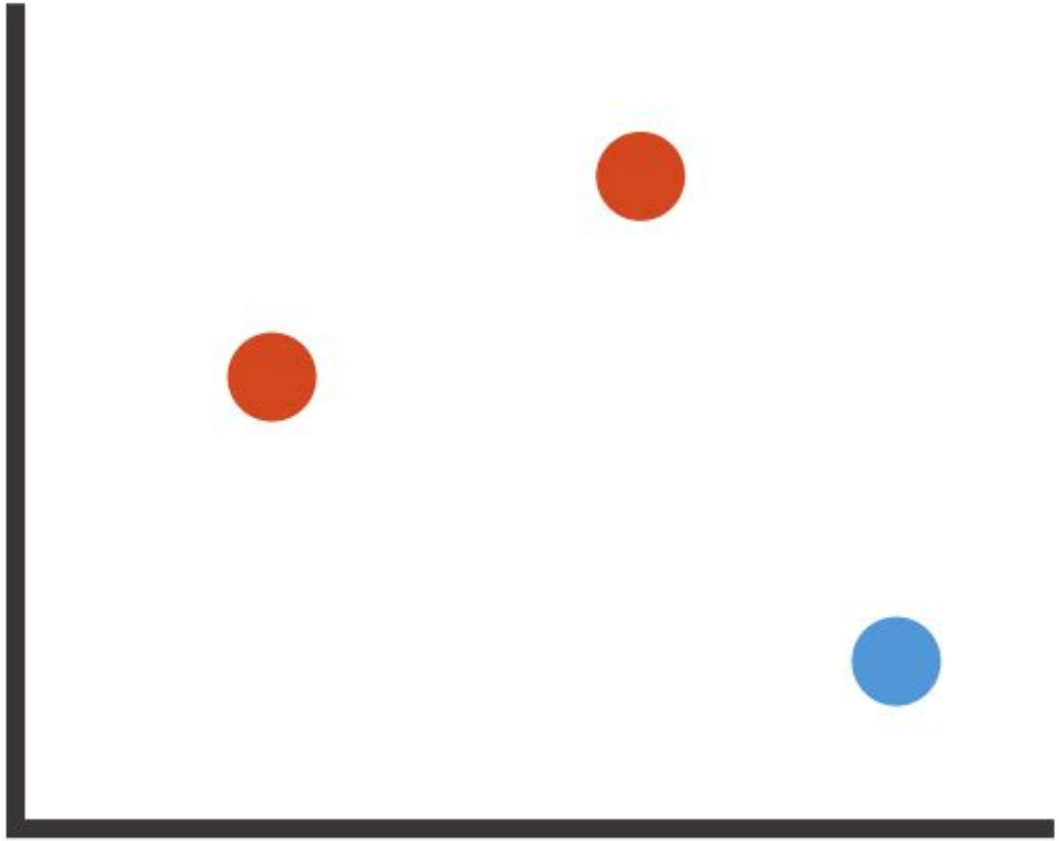
What channels are used?



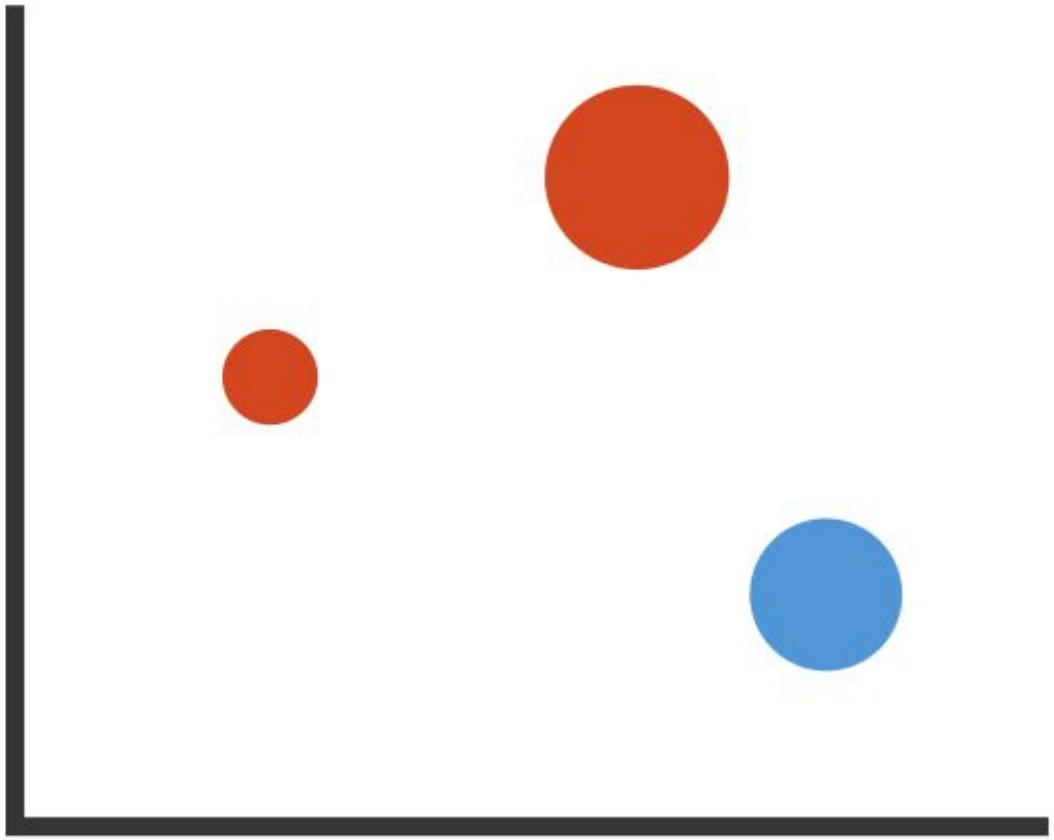
What channels are used?



What channels are used?



What channels are used?



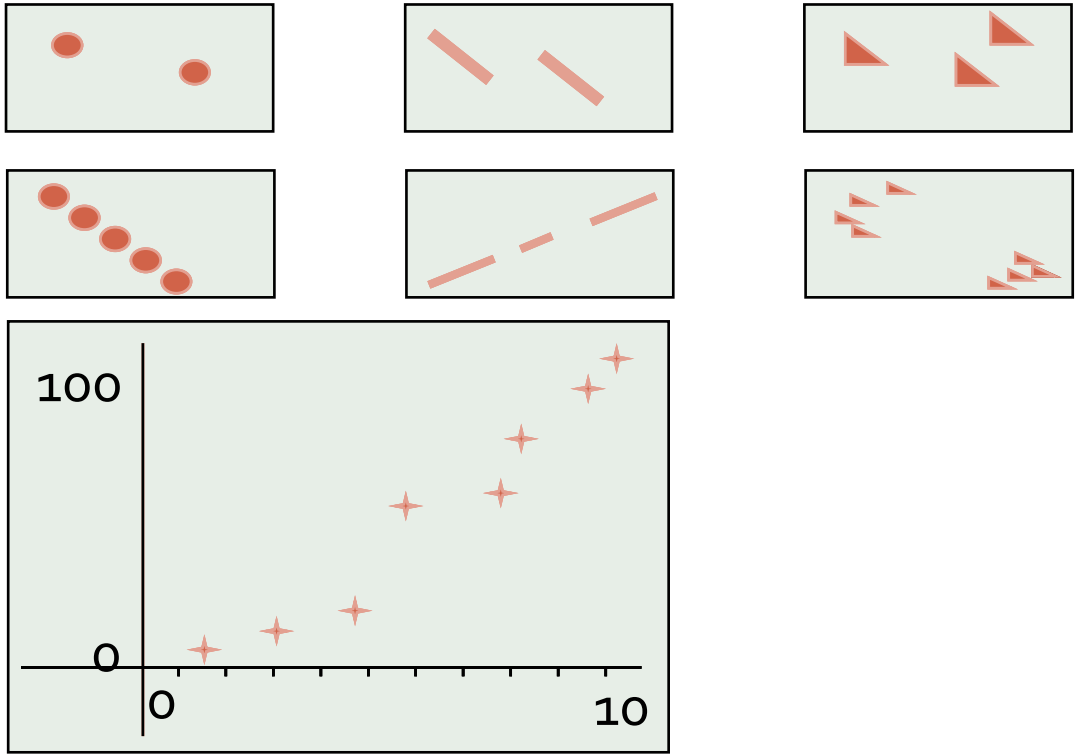
Characteristics of Visual Variables

- Selective:
  Can this variable allow us to spontaneously differentiate/isolate items from groups?
- Associative:
  Can this variable allow us to spontaneously group items in a group?
- Ordered:
  Can this variable allow us to spontaneously perceive an order?
- Quantitative:
  Can the difference between two marks in this variable be interpreted numerically ?
- Length (resolution):
  Across how many changes in this variable are distinctions possible?

From Semiology of Graphics (Bertin)

Visual Variable: Position

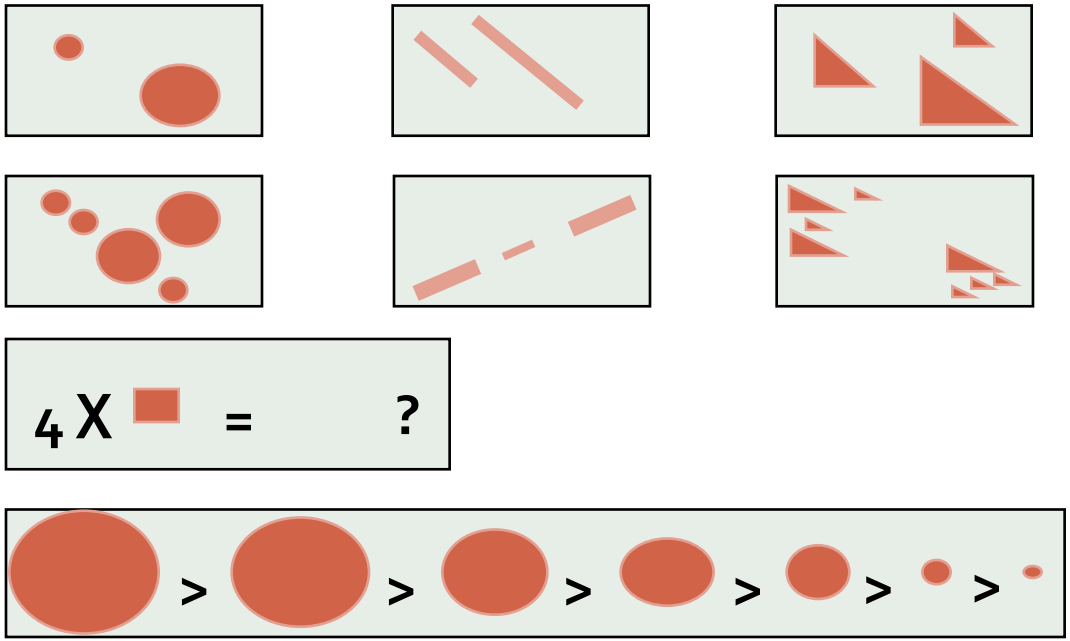
- selective
- associative
- quantitative
- order
- length (resolution)



From Semiology of Graphics (Bertin)

Visual Variable: Size

- selective
- associative
- quantitative
- order
- length (resolution)



Size

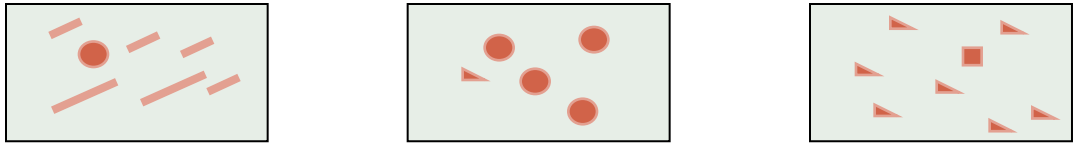
points

lines

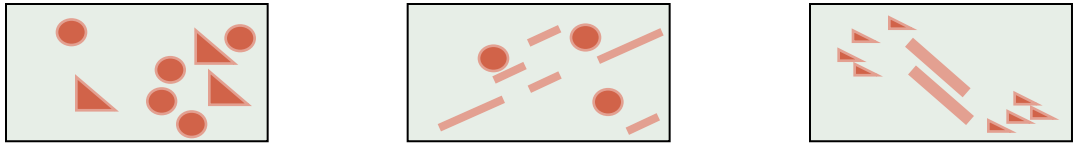
areas

Visual Variable: Shape

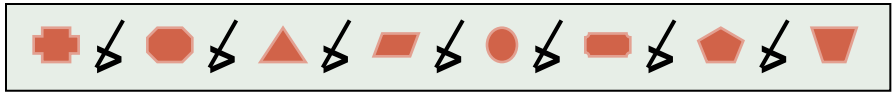
- selective



- associative



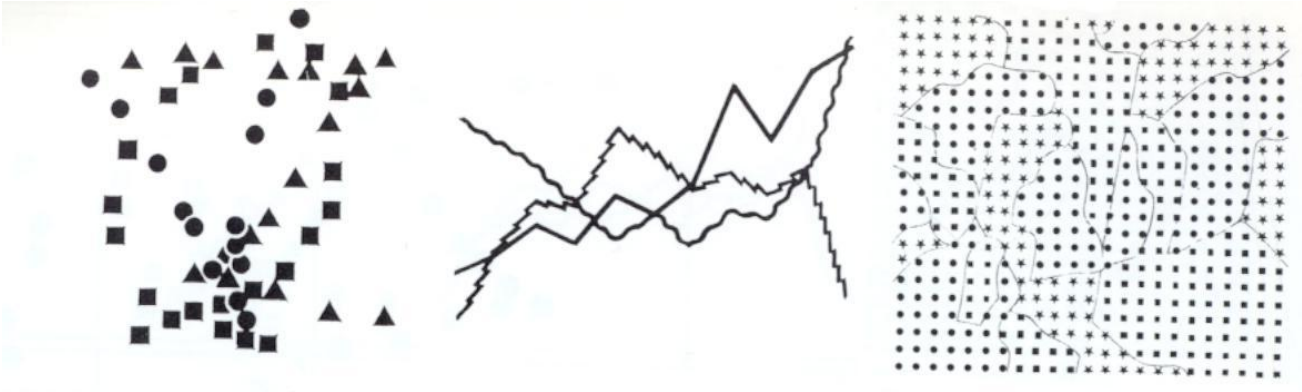
- ordered



- quantitative

- length (resolution)
  - infinite

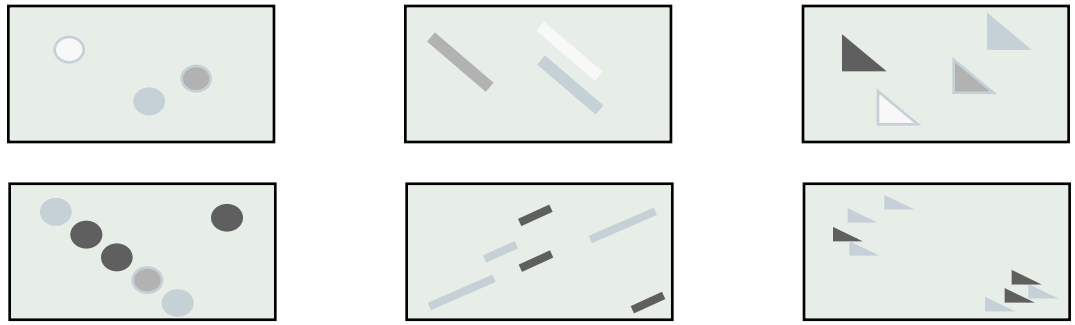
Shape



points lines areas

Visual Variable: Value

- selective
- associative
- quantitative
- order
- length (resolution)
  - theoretically infinite but practically limited
  - association and selection ~ < 7 and distinction ~ 10

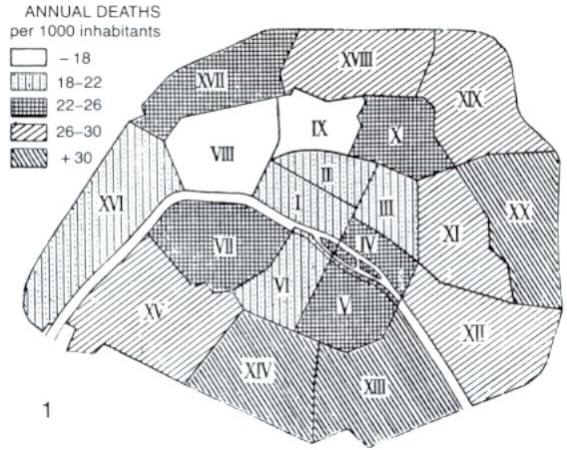


Value

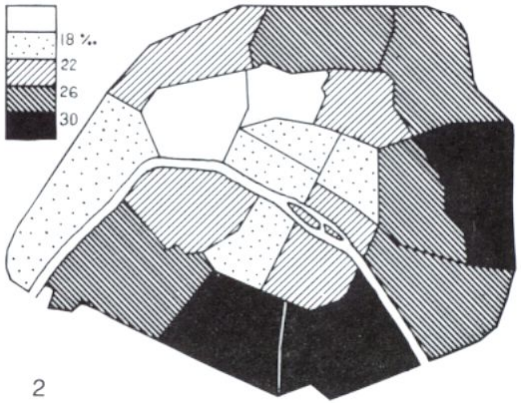
points lines areas

Value

ordered, cannot be reordered



Values not ordered correctly according to scale
Information has to be read point by point

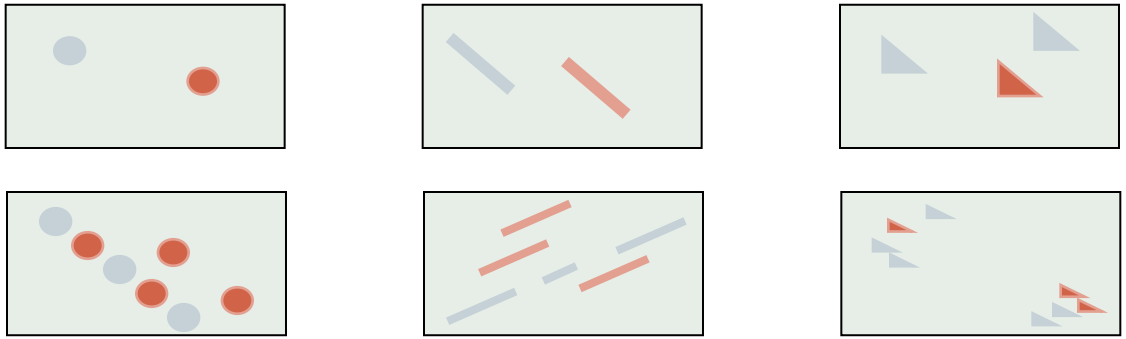


Values ordered correctly
Image much more useful

annual deaths per 1000 inhabitants, Paris

Visual Variable: Color

- selective
- associative
- quantitative
- order

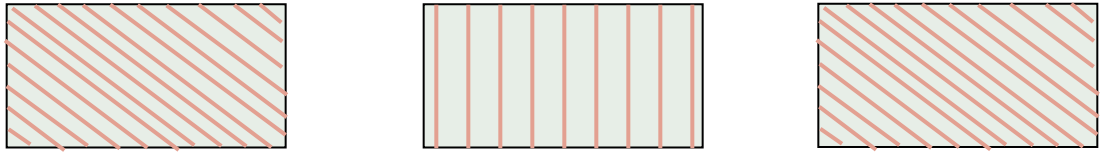


- length (resolution)
  - theoretically infinite but practically limited
  - association and selection ~ < 7 and distinction ~ 10

Visual Variable: Orientation

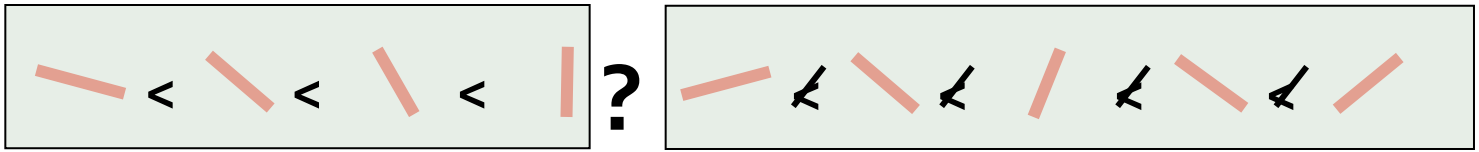
- selective

- associative



- quantitative

- order



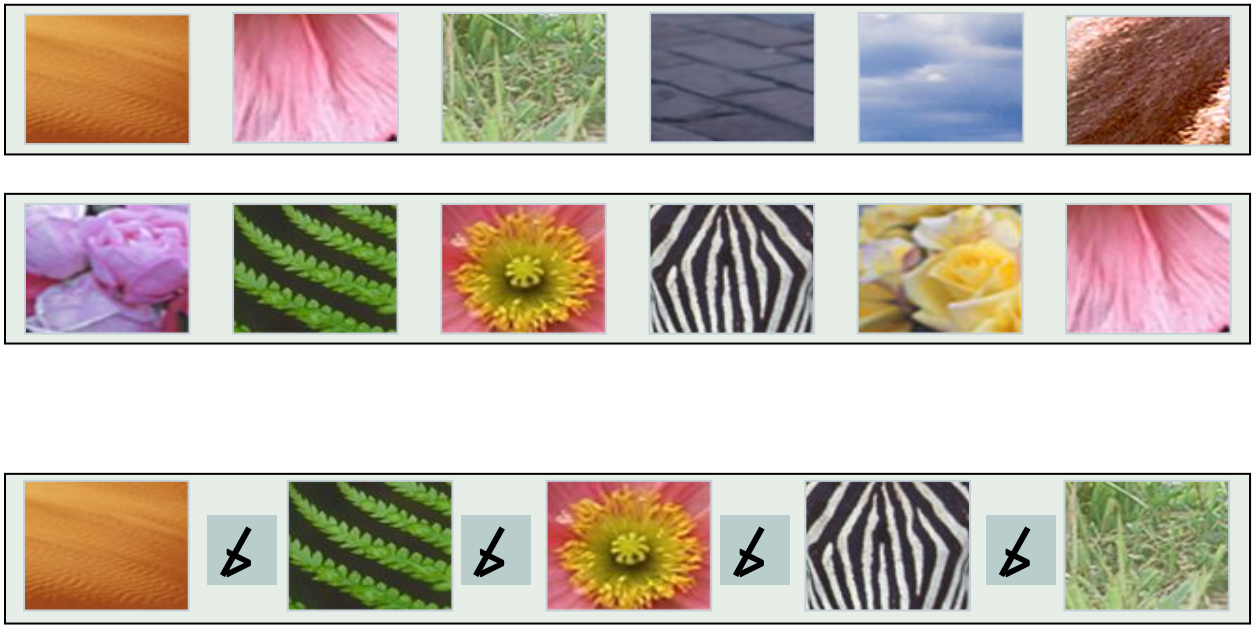
- length (resolution)
  - ~5 in 2D; ? in 3D

Orientation

points lines areas

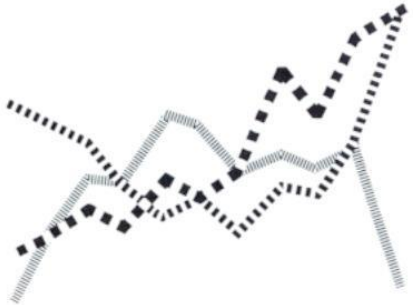
Visual Variable: Texture

- selective
- associative
- quantitative
- order
- length (resolution)
  - theoretically infinite

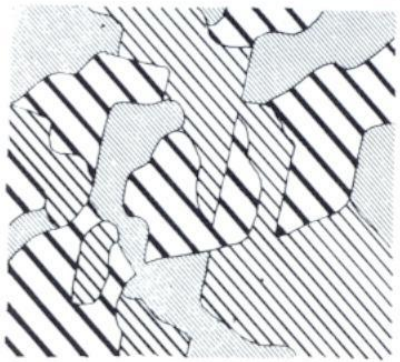


Texture

points



lines

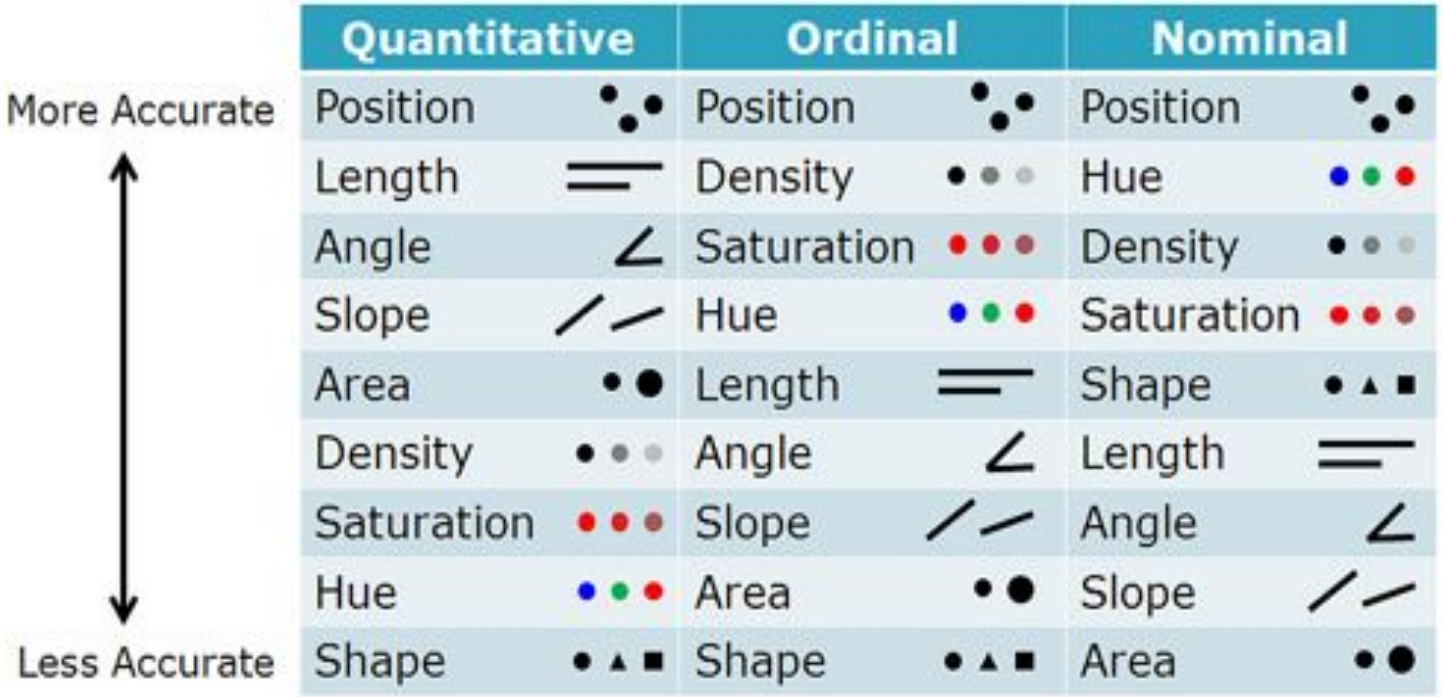


areas

Visual Variables

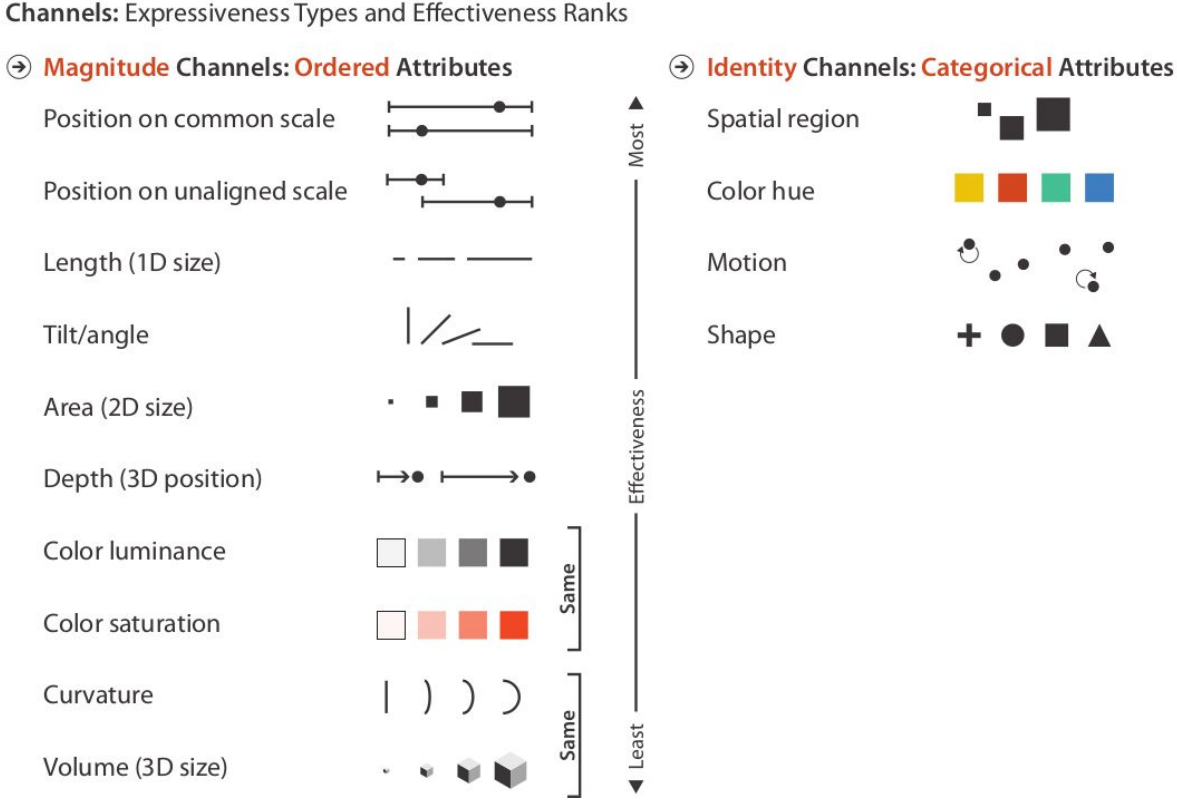
Visual Variable	Selective	Associative	Quantitative	Order	Length
Position	Yes	Yes	Yes	Yes	Dependant on resolution
Size	Yes	Yes	Approximate	Yes	Association: 5; Distinction: 20
Shape	With Effort	With Effort	No	No	Infinite
Value	Yes	Yes	No	Yes	Association: 7; Distinction: 10
Hue	Yes	Yes	No	No	Association: 7; Distinction: 10
Orientation	Yes	Yes	No	No	4
Grain	Yes	Yes	No	No	5
Texture	Yes	Yes	No	No	Infinite
Motion	Yes	Yes	No	Yes	Unknown

Summary



Jacques Bertin refined by Cleveland&McGill then by Card&Mackinlay

Summary


Channels: Expressiveness Types and Effectiveness Ranks
Magnitude Channels: Ordered Attributes
Position on common scale
Position on unaligned scale
Length (1D size)
Tilt/angle
Area (2D size)
Depth (3D position)
Color luminance
Color saturation
Curvature
Volume (3D size)
Same
Same
Most
Effectiveness
Least
Identity Channels: Categorical Attributes
Spatial region
Color hue
Motion
Shape

Summary

- Now you know the main building blocks are **marks**
- Marks are modified by **visual variables**
- Visual variables have **specific characteristics**
- These characteristics influence how the data will be perceived

Visualization Tools

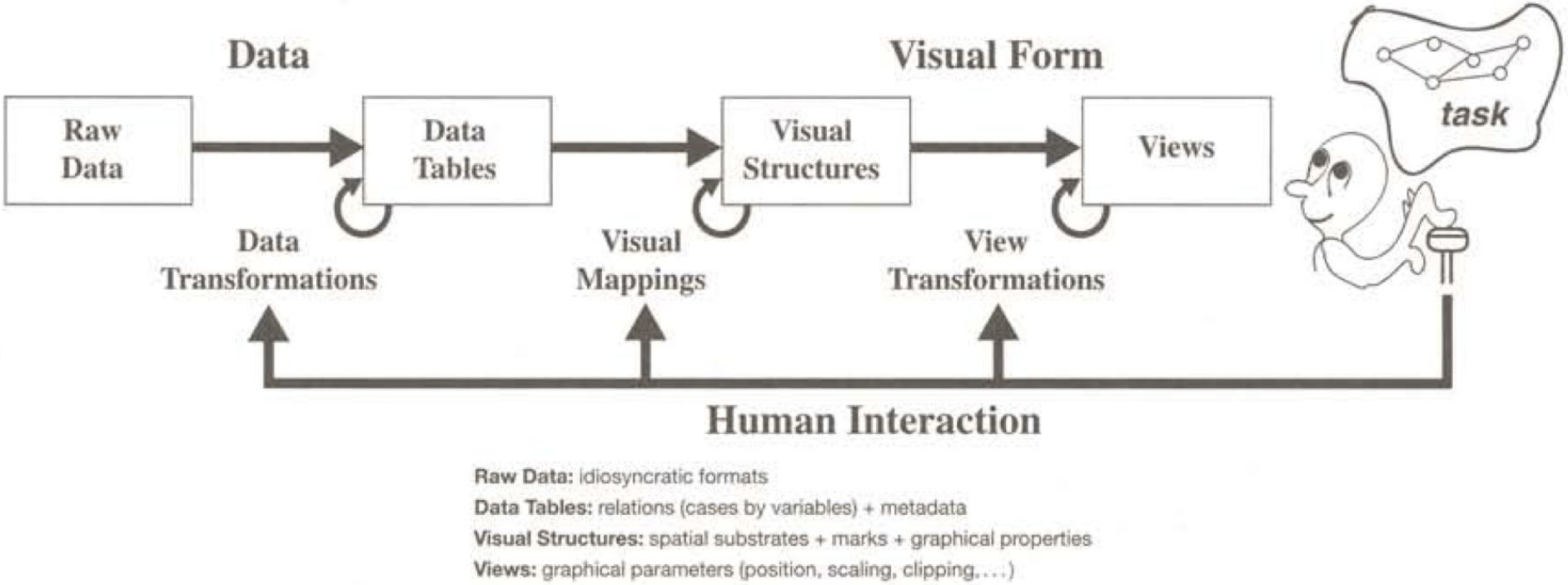
Outline

Don't do it yourself!

If you are asked to do it yourself: don't do it yourself!

- Architecture
- High-level systems
- Low-level systems
- Toolkits
- Specialized environments

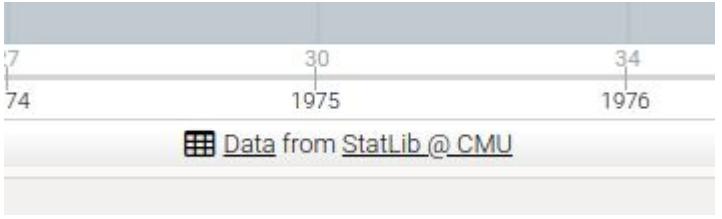
Conceptual Pipeline



[Card, Mackinlay, Shneiderman, Readings in Information Visualization: Using Vision to Think, 1999]

Example: Raw Data

Bottom of https://keshif.me/demo/cars



lib.stat.cmu.edu/datasets/cars.data

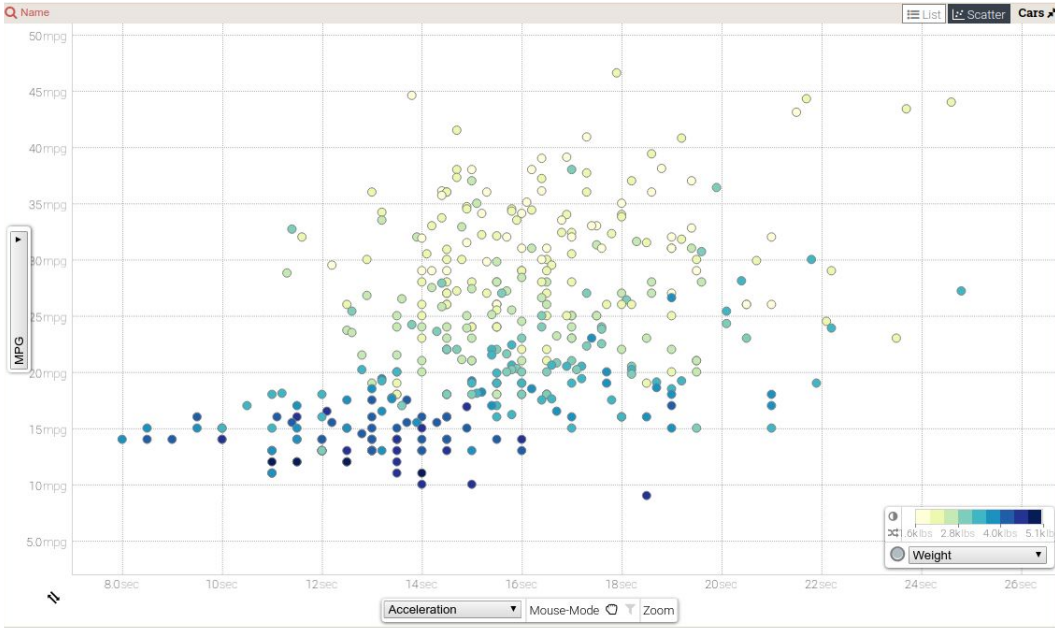
```
# To unbundle, sh this file
echo cars.data.des 1>&2
sed 's/.//' >cars.data.des <<'//GO.SYSIN DD cars.data.des'
--f
-mpg REAL 1
-cylinders REAL 2
-displacement REAL 3
-horsepower REAL 4
-weight REAL 5
-acceleration REAL 6
-model.year REAL 7
-origin REAL 8
//GO.SYSIN DD cars.data.des
echo cars.names 1>&2
sed 's/.//' >cars.names <<'//GO.SYSIN DD cars.names'
-"chevrolet chevelle malibu"
-"buick skylark 320"
-"plymouth satellite"
```

Data Tables

	A	B	C	D	E	F	G	H	I
1	Car	MPG	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model	Origin
2	Chevrolet Chevelle Malibu	18	8	307	130	3504	12	70	US
3	Buick Skylark 320	15	8	350	165	3693	11,5	70	US
4	Plymouth Satellite	18	8	318	150	3436	11	70	US
5	AMC Rebel SST	16	8	304	150	3433	12	70	US
6	Ford Torino	17	8	302	140	3449	10,5	70	US
7	Ford Galaxie 500	15	8	429	198	4341	10	70	US
8	Chevrolet Impala	14	8	454	220	4354	9	70	US
9	Plymouth Fury iii	14	8	440	215	4312	8,5	70	US
10	Pontiac Catalina	14	8	455	225	4425	10	70	US
11	AMC Ambassador DPL	15	8	390	190	3850	8,5	70	US
12	Citroen DS-21 Pallas	0	4	133	115	3090	17,5	70	Europe
13	Chevrolet Chevelle Concours (sw)	0	8	350	165	4142	11,5	70	US
14	Ford Torino (sw)	0	8	351	153	4034	11	70	US
15	Plymouth Satellite (sw)	0	8	383	175	4166	10,5	70	US
16	AMC Rebel SST (sw)	0	8	360	175	3850	11	70	US
17	Dodge Challenger SE	15	8	383	170	3563	10	70	US
18	Plymouth 'Cuda 340	14	8	340	160	3609	8	70	US
19	Ford Mustang Boss 302	0	8	302	140	3353	8	70	US
20	Chevrolet Monte Carlo	15	8	400	150	3761	9,5	70	US
21	Buick Estate Wagon (sw)	14	8	455	225	3086	10	70	US
22	Toyota Corolla Mark ii	24	4	113	95	2372	15	70	Japan
23	Plymouth Duster	22	6	198	95	2833	15,5	70	US
24	AMC Hornet	18	6	199	97	2774	15,5	70	US
25	Ford Maverick	21	6	200	85	2587	16	70	US
26	Datsun PL510	27	4	97	88	2130	14,5	70	Japan
27	Volkswagen 1131 Deluxe Sedan	26	4	97	46	1835	20,5	70	Europe
28	Peugeot 504	25	4	110	87	2672	17,5	70	Europe

Visual Structure

items = [{x: 10, y: 20, intensity: 0.3}, …];



Views


Created with Keshif
406 Cars
Origin
American 254
Japanese 79
European 73
Cylinders
3 4
4 207
5 3
6 84
8 108
5 Rows
MPG
Acceleration
Name
List
Scatter
Cars
MPG
Weight
Acceleration
Mouse-Mode
Zoom
Engine Displacement
Horsepower
Weight
Year
Data from StatLib @ CMU
Keshif

High-level systems

- On desktops
  - **Commercial:** Tableau, Spotfire, Qlik, Power BI
  - **For networks:** Cytoscape, Tulip, Gephi
- On the Web
  - **Commercial:** The same with their web versions
  - **Free**: Polestar, Voyager https://vega.github.io/

Always start to explore data with those to avoid wasting time.

Low-level Systems

Proliferation

- Many examples using Processing
  - https://processing.org/
- Web-based D3 very rich but **extremely difficult to harness**
  - D3.js

Toolkits

Easy to use but need some coding

- Java (becoming a bit obsolete)
  - Cytoscape, Gephi
- Python
  - Many toolkits, e.g. Tulip (C++) can be used as a toolkit
  - Matplotlib, Seabord, ggpy, Bokeh, Altair
- JavaScript easy to deploy for web-based visualizations
  - D3 (be careful of the learning curve)
  - Vega-lite
- VTK and ParaView
  - Specialized for 3D but also work in 2D
  - https://www.paraview.org/

Specialized environments

- R uses ggplot2 and extensions
- Python comes with (too) many toolkits
  - Matplotlib, Seaborn, Bokeh, Plotly, Altair
- Difficult to deploy as full applications for end-users
- Easy to use with notebooks for replicating analyses

Grammar of Graphics

Book of Leland Wilkinson, 1999

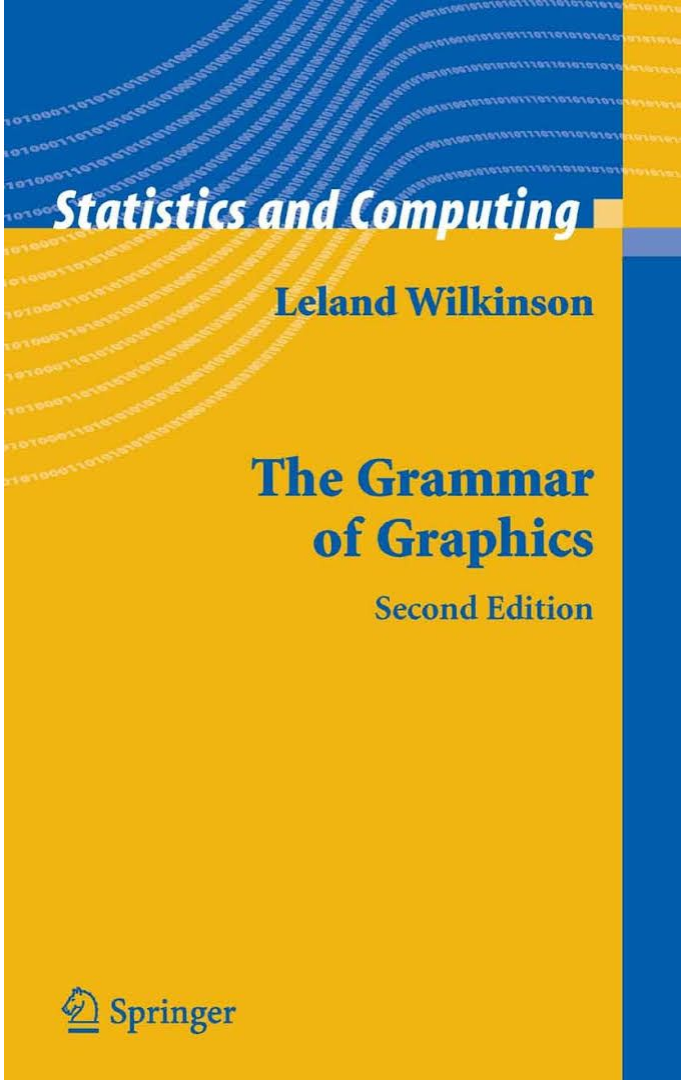
Syntax to specify the construction of visualizations.

Implemented by:

- R with Ggplot2
- JavaScript with Vega lite

Use Vega lite if you can

Limited in scalability (50k items)



Interaction with Visualization

Not an afterthought

Demos...

Conclusion

- Visualization is now mature
- Use it for debugging, use it for showing many results to users
- Stick to well-established rules or collaborate with practitioners
- Not meant to embellish but to to convey meaning effectively
- Use high-level toolkits rather than hand-crafted systems if you can
- Beware of fashionable unreadable visualizations
  - For graphs, avoid hairballs!
- Visualizations allow to increase recall and decrease precision
  - Humans can find what they want effectively
- Don't hesitate to ask around you for methods, don't reinvent the wheel!

The Grammar of Graphics

https://uwdata.github.io/visualization-curriculum/intro.html

Assignment

Reproduce figures from the IPCC report 2023 with legends and captions

- https://www.ipcc.ch/report/ar6/syr/figures
- https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_LongerReport.pdf
- https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_07_Supplementary_Material.pdf (and others)
- https://www.ipcc.ch/site/assets/uploads/2019/04/IPCC-visual-style-guide.pdf
- http://guidance.climatesciencecognition.com/

Using either:

- A Python notebook (using Altair or another system)
- An Observable notebook (using Vega-lite or Plot)

First, produce a figure as similar as possible from the original
then improve it with hyperlinks and interactions, and document the rationales.