DATA VISUALIZATION

Contents

Contents

1.1 Definitions and Objectives

"Transformation of the symbolic into the geometric"

[McCormick et al. 1987]

"Finding the artificial memory that best supports our natural means of perception."

[Bertin 1967]

"The greatest value of a picture is when it forces us to notice what we never expected to see."

- **[John Tukey, 1977]**

"The use of computer-generated, interactive, visual representations of data to amplify cognition."

[Card et al. 1999]

1.1 Definitions and Objectives

Objective of Data Visualization:

Study techniques and algorithms for creating visualizations based on the principles of graphics, visual art, perceptual psychology and cognitive science.

Transform a large amount of raw data into a graphical representation by exploiting the superiority of human visual processing to detect patterns and their ability to infer. The objective is to improve the interpretation and communication of data.

[Alexandru 2008]

Population:

- Develop applications for big data visualization,
- Develop big data visualization tools and systems

1.2 Need for data visualization

- **Growth of data production**

The volume of data generated worldwide is expected to exceed 180 zettabytes 10^{21}) by 2025, representing an average annual growth of nearly 40% over five years [w1].

- **The ability to take data and be able to:**
  - understand them,
  - treat them,
  - extract useful information,
  - visualize them,
  - communicate them

- **This will be an important skill to acquire for the future due to the free and omnipresence of data.**

1.2 Need for data visualization

Why create visualizations?

- Answer or discover questions
- Make a decision
- See data in specific context
- Expand memory
- Finding Patterns
- Present arguments or tell a story
- Get inspired
- Communicate information (share, collaborate)

1.2 Need for data visualization

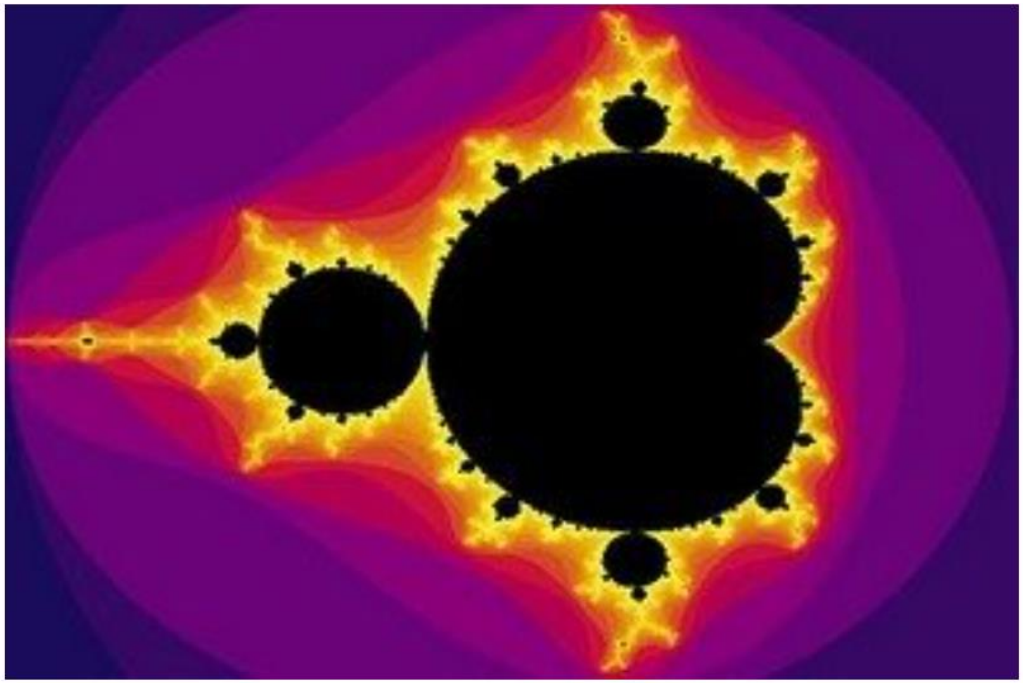
Areas of Visualization

- Health
- Policy
- Scientist (Engineering, Mathematics, etc.)
- Finance (Investment performance monitoring)
- Logistics (Determine the best routes.)
- Geographic Information Systems
- Others

1.2 Need for data visualization

Areas of Visualization
Finance (Investment performance monitoring)
Source: [W3]

1.2 Need for data visualization



Areas of Visualization
Scientist (Engineering, Mathematics, etc.)
Source: [W4]

1.2 Need for data visualization

- **Some challenges:**

- Increase of data in the network
- Text and documents
- Images
- Geospatial Data
- Heterogeneous Data

1.2 Need for data visualization

Technics Related to Visualization

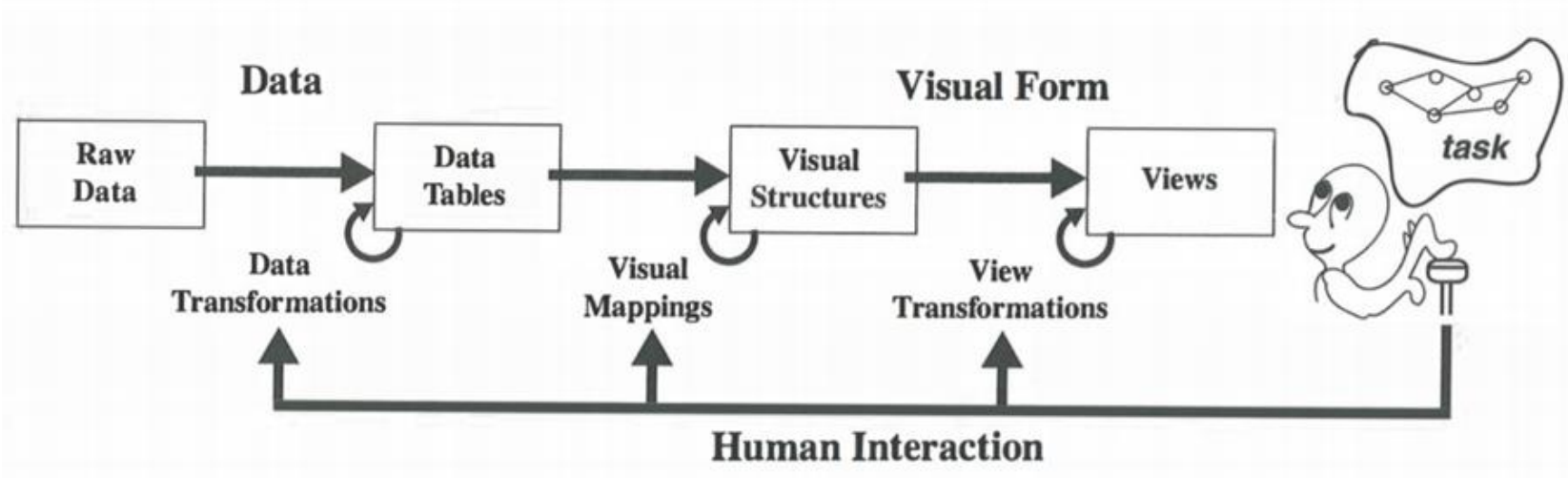
- Data Analysis
- Data mining
- Data Bases
- HCI (Human-Computer Interaction)
- Computer Graphics

1.2 Need for data visualization

Modes of Visualization

- Interactive Visualization (web)
- Presentation
- Interactive (Storytelling)

1.3 Process of Visualization



Source: [Card et al. 1999]

Visual Structures: the space in which we're going to create the visualization

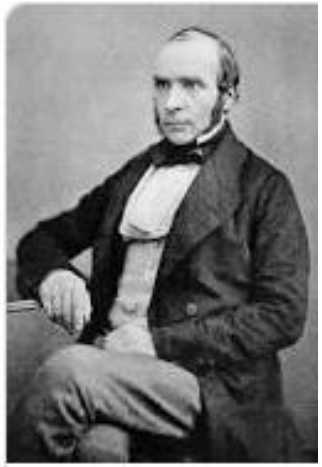
Views: Graphical parameters

1.4 Examples of Visualization

Dr. John Snow (1813–58). London practicing obstetrician/anesthesiologist who conducted a detailed epidemiologic investigation of London cholera epidemic adjacent to the now famous Broad St. pump [w2].

1.4 Examples of Visualization

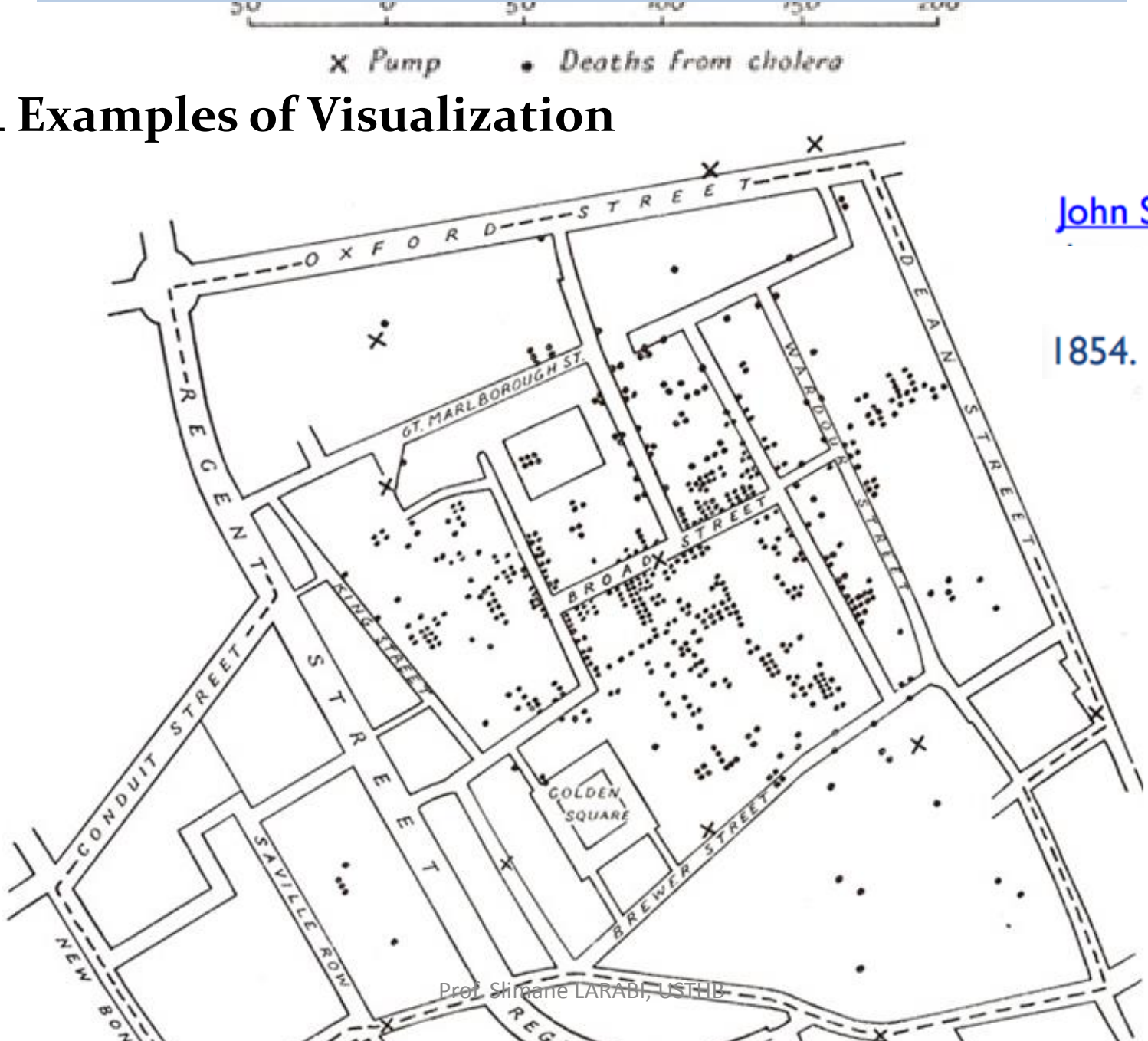
1.4 Examples of Visualization



Memorial (the pump without handle), in London, road Broad street (Actually Broadwick).

The pump handle was removed on September 8, 1854. Removal of the handle prevented additional cholera deaths, supporting Snow's theory that cholera was a waterborne, contagious disease

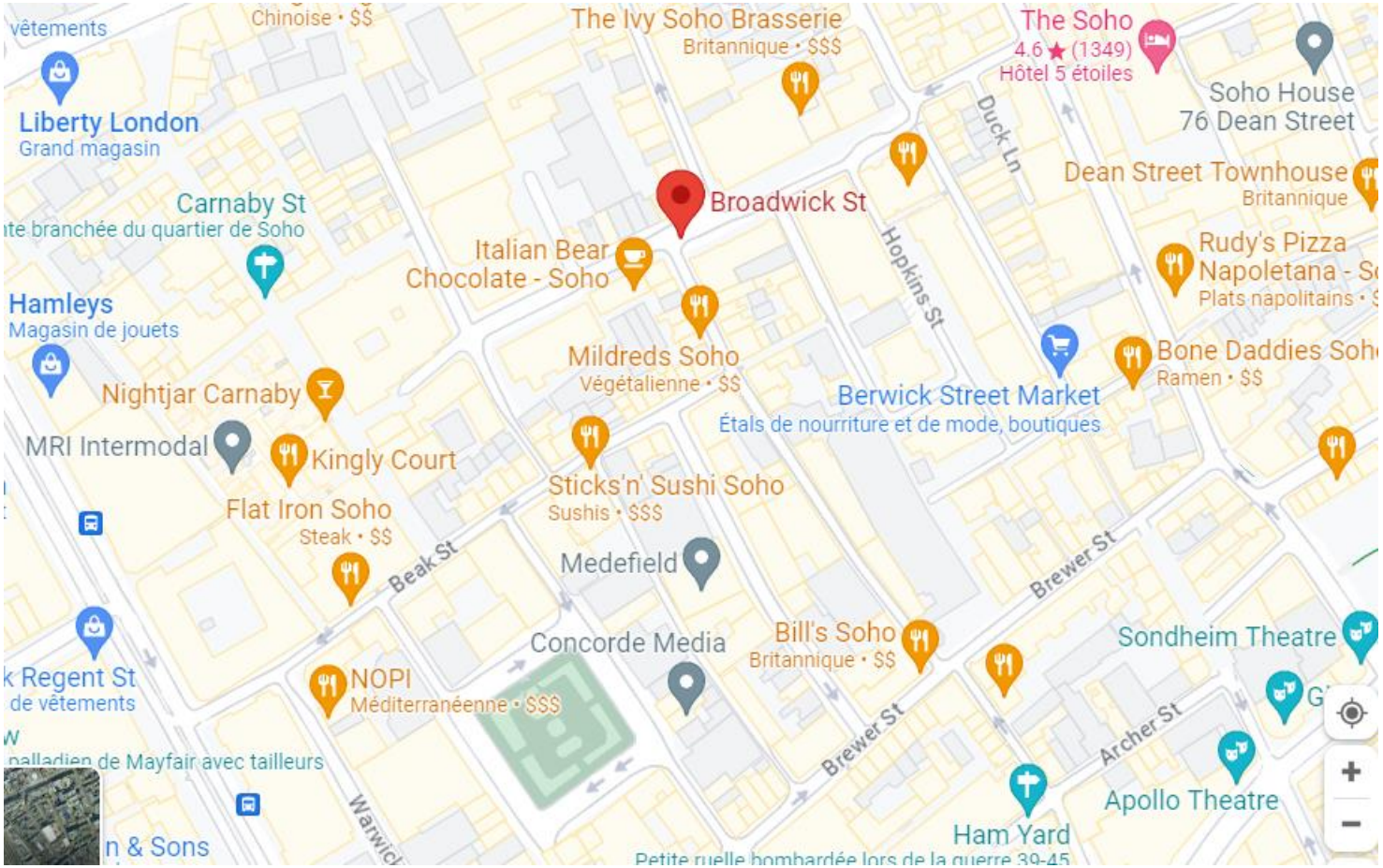
1.4 Examples of Visualization



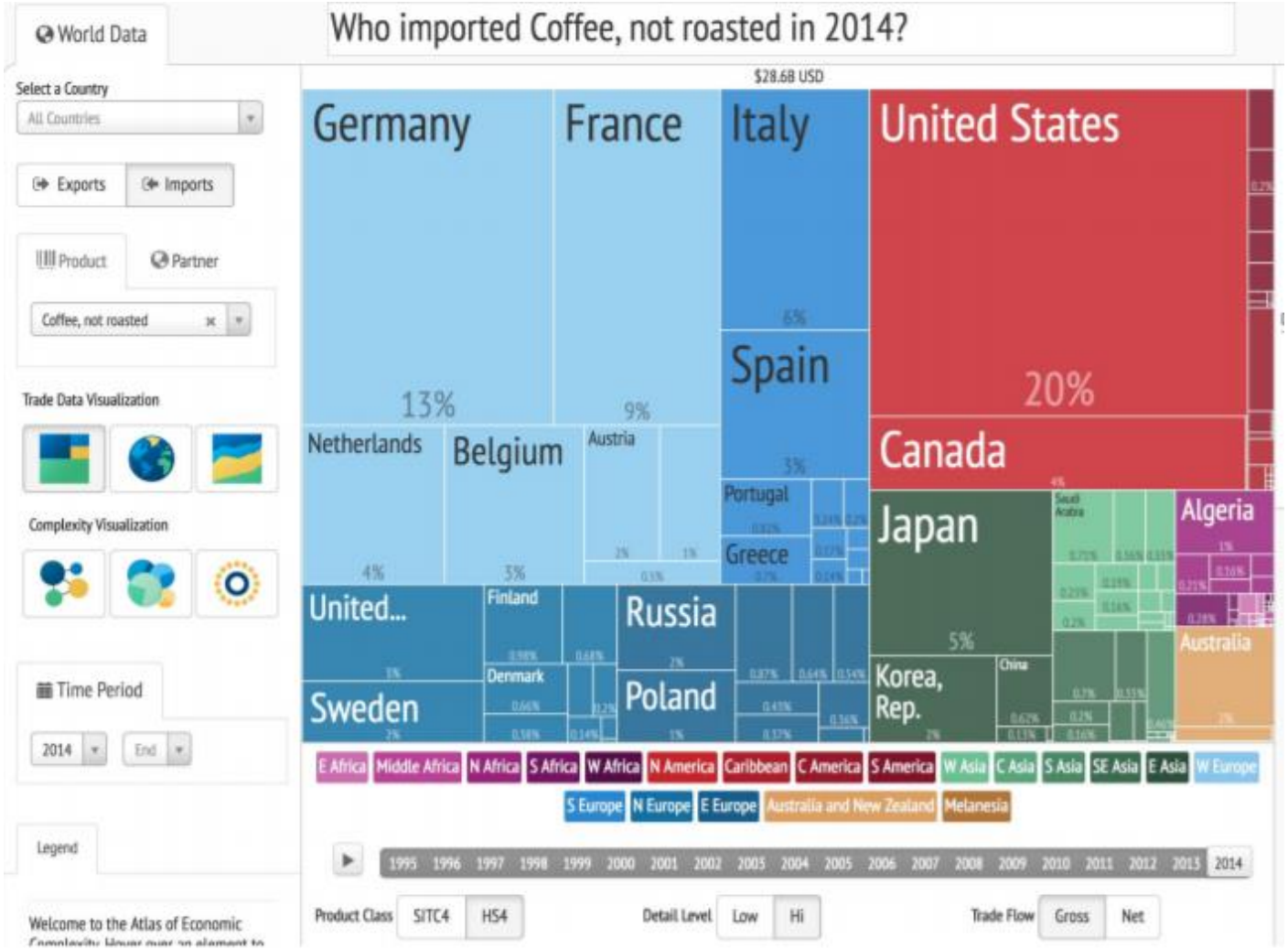
John Snow,

1854. London.

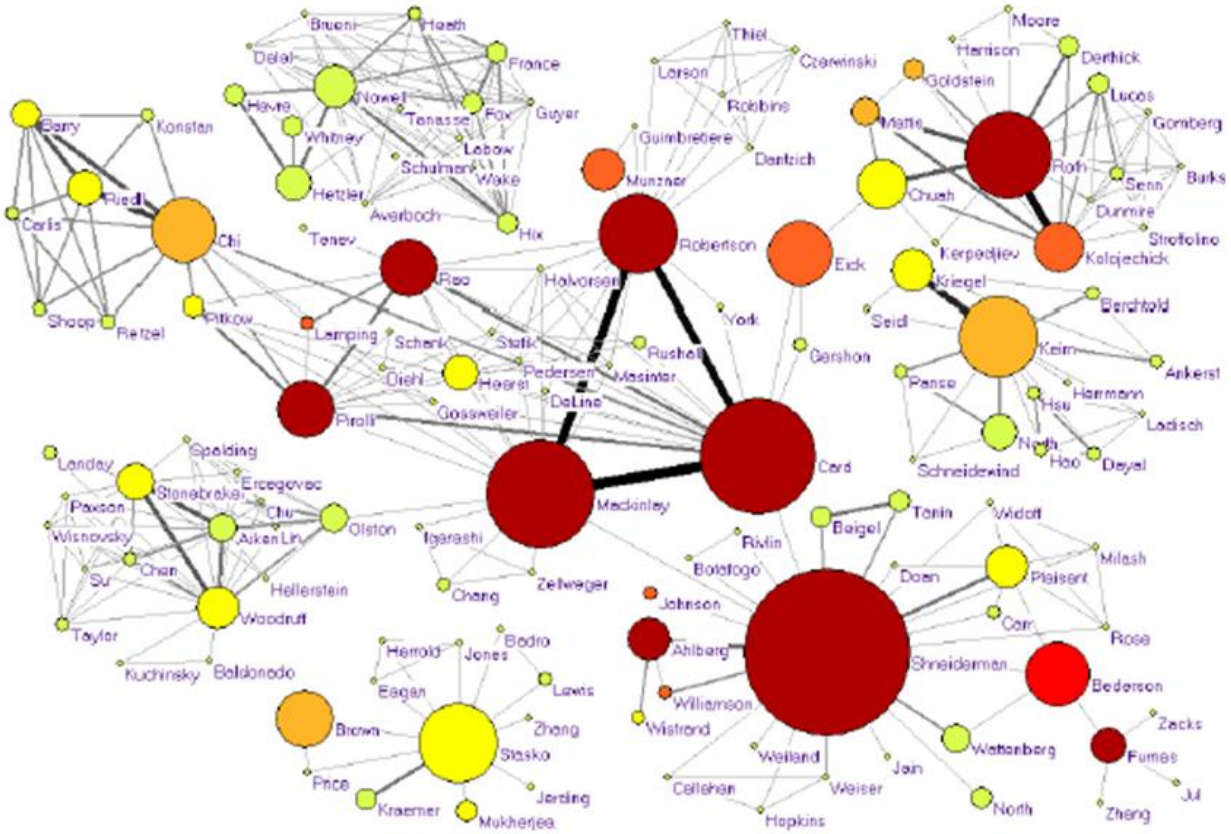
1.4 Examples of Visualization



1.4 Examples of Visualization



1.4 Examples of Visualization

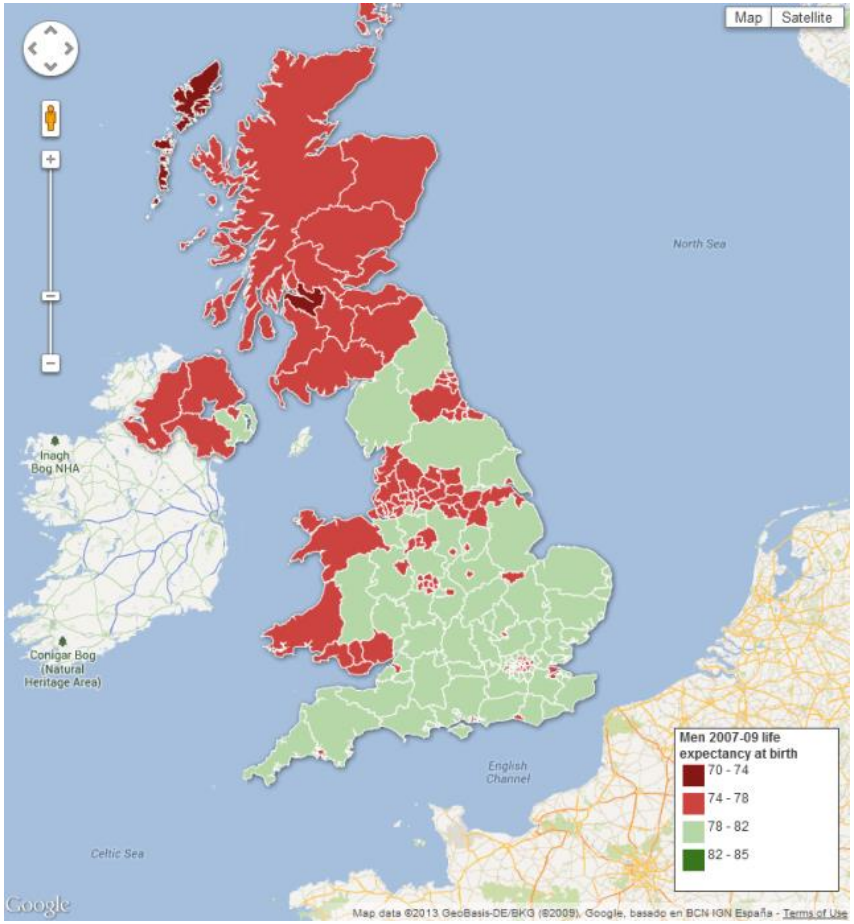


1.4 Examples of Visualization

1.4 Examples of Visualization

THE YELLOW LINES REPRESENT BUSES IN REAL TIME AND THE RED CORRESPONDS TO DENSITY OF PEOPLE. SEP 2006.

1.4 Examples of Visualization



1.4 Examples of Visualization

Hans Rosling

2007

1.5 Basic Principle of Visualization

Attributes to use in visualization production [3]

Preattentive Visual Properties:
What are the visual elements that attract the attention of an observer for a data visualization (attributes of pre-attentive processing) ?

1.5 Basic Principle of Visualization

Attributes to use in visualization production [Card et al. 1999]

What are the visual elements that attract the attention of an observer for a data visualization (attributes of pre-attentive processing) ?

Color, Hue, Intensity, **Shape**, **Motion**, **Spatial position**.

1.5 Basic Principle of Visualization

•**Color**:

How many digits 9 are written on the following list ?

1 2 4 6 5 3 4 7 8 3 4 5 6 2 3 5 6 9 0 7 6 4 3 1 2 4 5 6 3 4 5
3 5 6 7 3 9 5 4 6 1 3 2 9 0 7 5 4 6 3 2 1 2 6 9 0 5 4 2 3 1 4
2 4 3 5 4 3 1 6 9 6 4 7 8 6 9 0 3 2 1 4 6 7 3 4 1 2 5 4 6 2 6
6 5 3 4 5 2 9 8 7 5 4 6 5 6 2 3 4 5 8 9 0 8 9 0 3 2 4 5 6 1 2
3 4 5 2 3 1 6 7 8 3 4 6 5 1 2 7 9 8 6 5 6 4 3 5 7 8 1 0 2 1 4
1 3 2 5 4 6 4 5 3 2 7 0 5 6 7 8 3 2 4 5 6 7 5 4 2 3 6 7 8 4 3
5 6 3 4 5 6 7 2 3 4 1 2 6 7 8 3 4 6 7 2 3 4 1 5 6 4 2 3 4 1 4

1.5 Basic Principle of Visualization

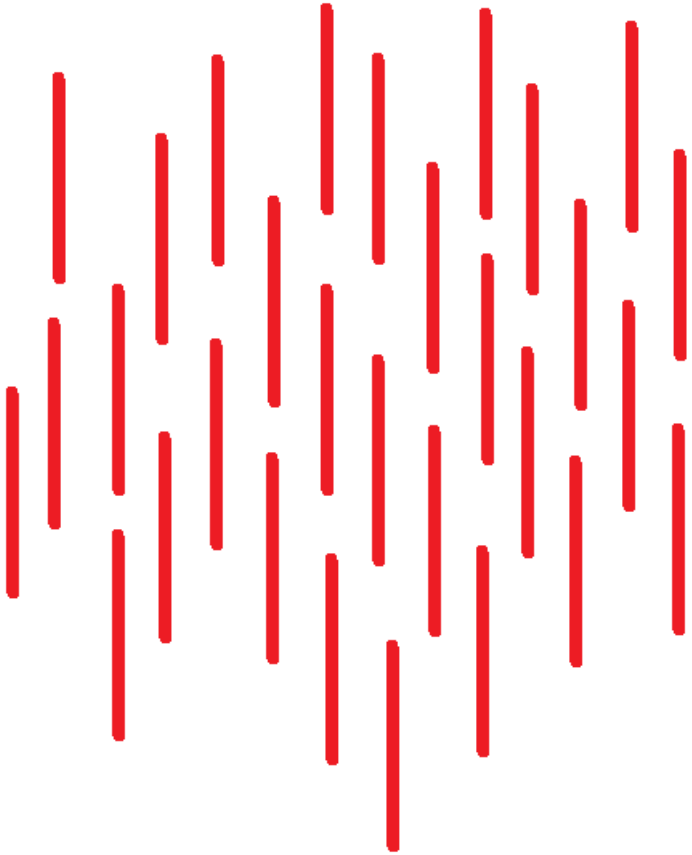
•**Color**:

How many digits 9 are written on the following list ?

1 2 4 6 5 3 4 7 8 3 4 5 6 2 3 5 6 9 0 7 6 4 3 1 2 4 5 6 3 4 5
3 5 6 7 3 9 5 4 6 1 3 2 9 0 7 5 4 6 3 2 1 2 6 9 0 5 4 2 3 1 4
2 4 3 5 4 3 1 6 9 6 4 7 8 6 9 0 3 2 1 4 6 7 3 4 1 2 5 4 6 2 6
6 5 3 4 5 2 9 8 7 5 4 6 5 6 2 3 4 5 8 9 0 8 9 0 3 2 4 5 6 1 2
3 4 5 2 3 1 6 7 8 3 4 6 5 1 2 7 9 8 6 5 6 4 3 5 7 8 1 0 2 1 4
1 3 2 5 4 6 4 5 3 2 7 0 5 6 7 8 3 2 4 5 6 7 5 4 2 3 6 7 8 4 3
5 6 3 4 5 6 7 2 3 4 1 2 6 7 8 3 4 6 7 2 3 4 1 5 6 4 2 3 4 1 4

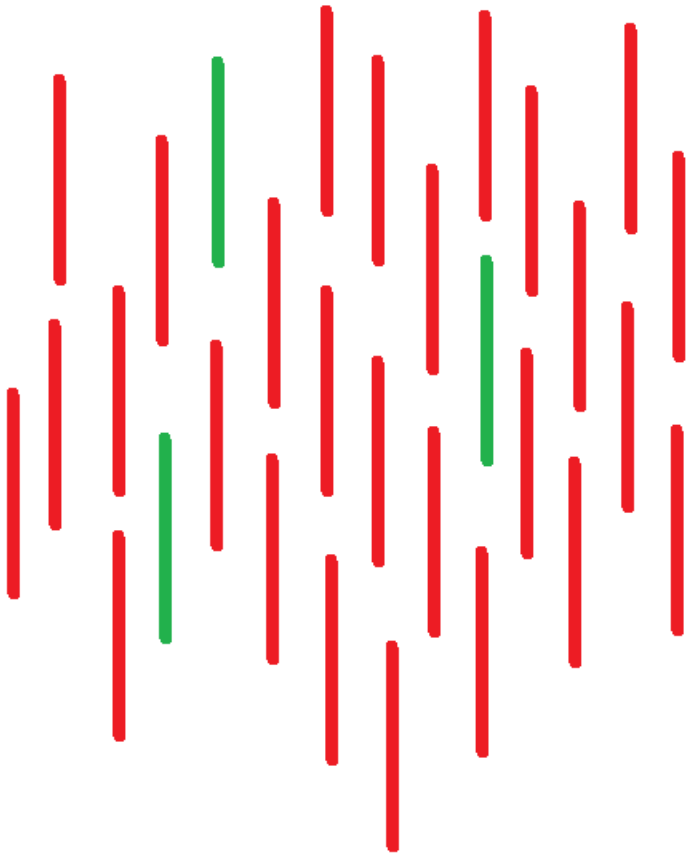
1.5 Basic Principle of Visualization

The Hue



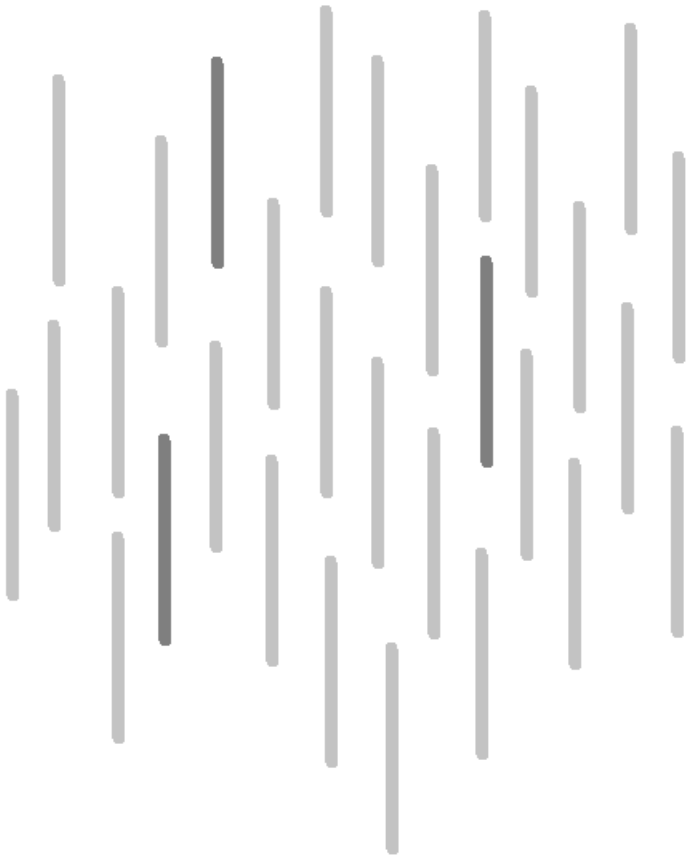
1.5 Basic Principle of Visualization

The Hue



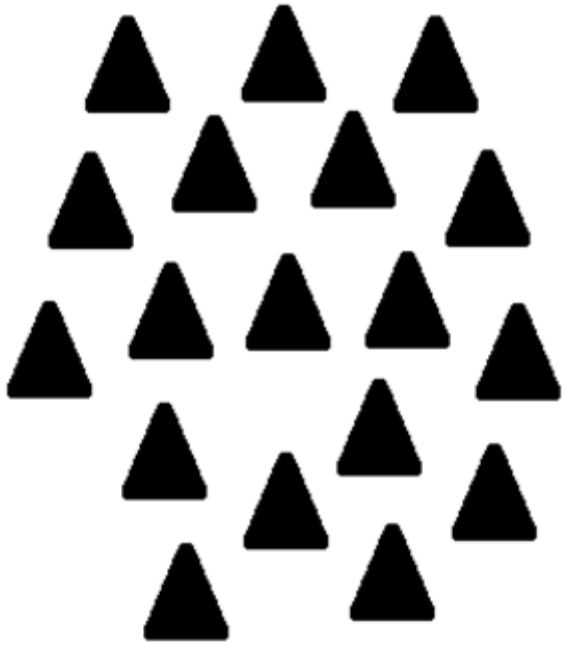
1.5 Basic Principle of Visualization

The Intensity



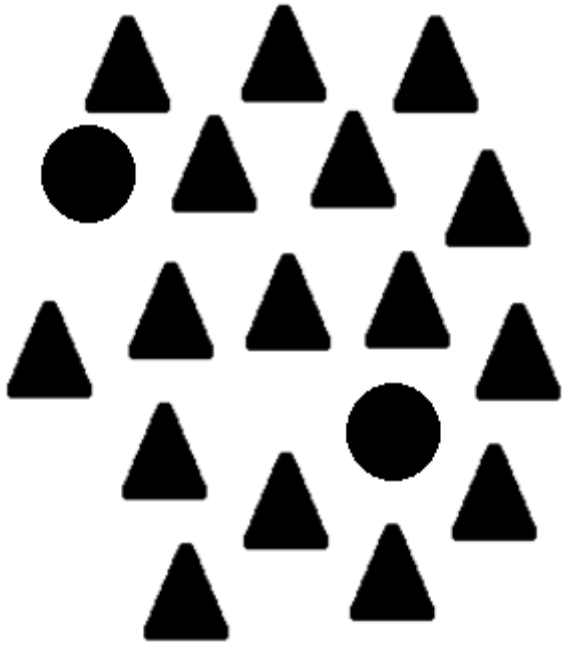
1.5 Basic Principle of Visualization

The Shape



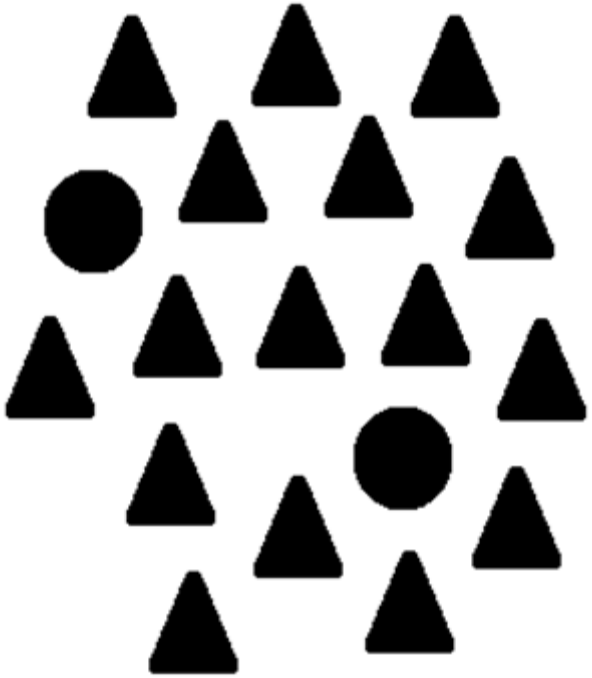
1.5 Basic Principle of Visualization

The Shape



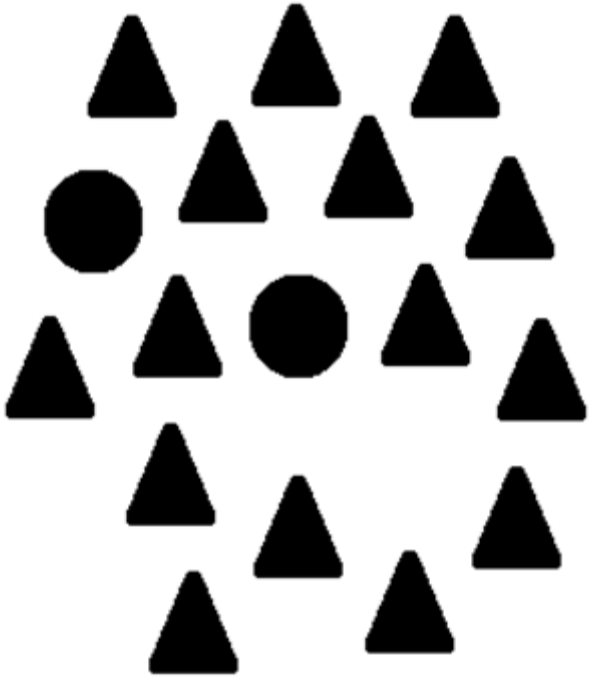
1.5 Basic Principle of Visualization

The Motion



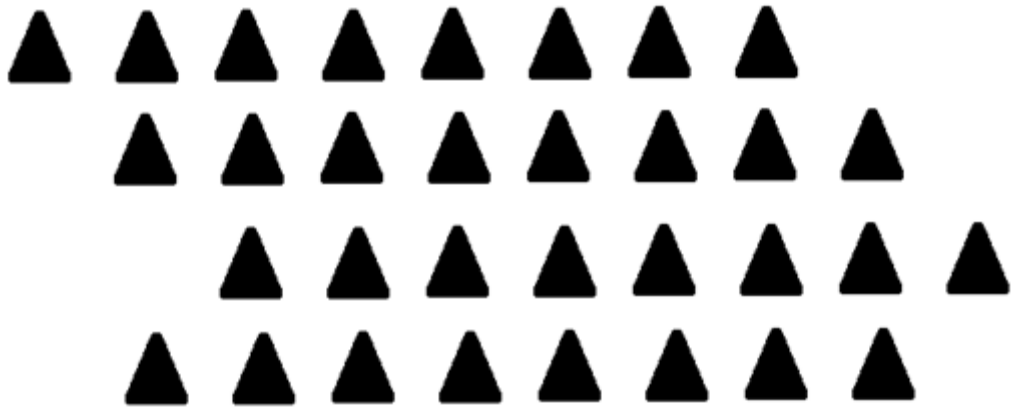
1.5 Basic Principle of Visualization

The Motion



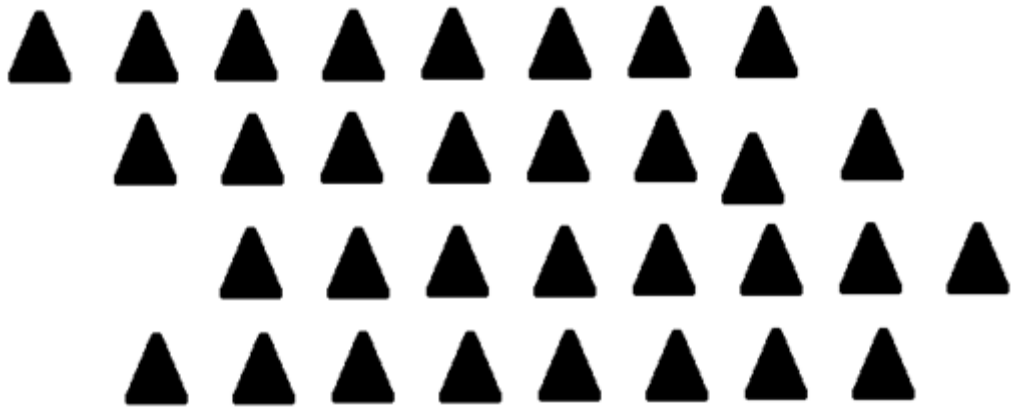
1.5 Basic Principle of Visualization

The Spatial Position



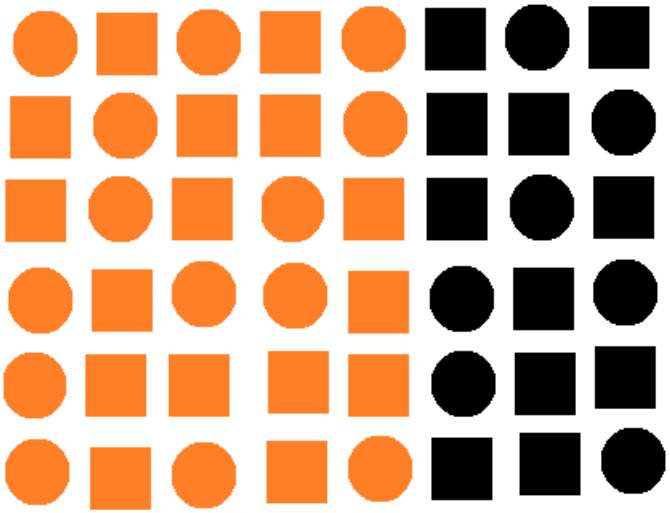
1.5 Basic Principle of Visualization

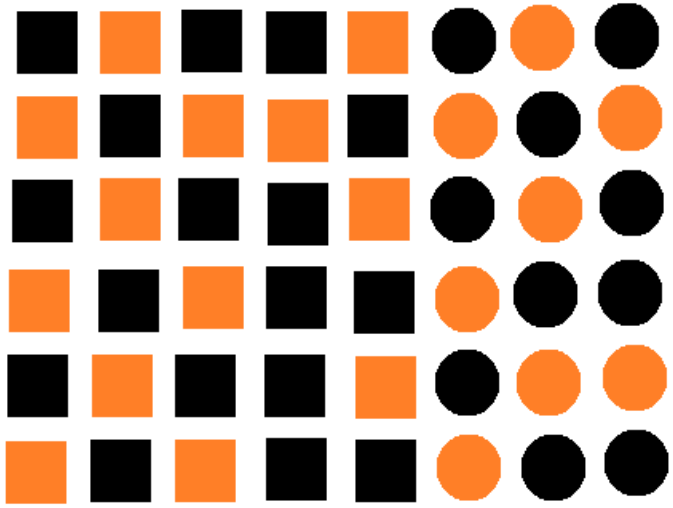
The Spatial Position



1.5 Basic Principle of Visualization

Shape or color are pre-attentive attributes that allow
for example to delimit the border present in the presentation.





1.5 Basic Principle of Visualization

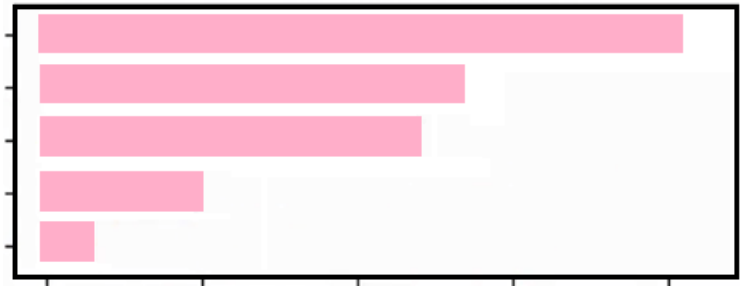
Precautions to consider when using color:

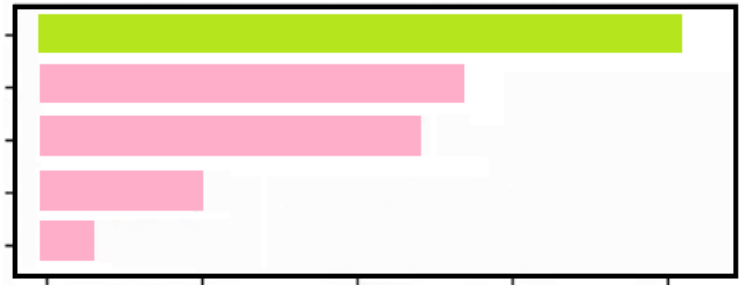
- Black and white (more than enough)

1.5 Basic Principle of Visualization

Precautions to consider when using color:

Use medium shades or pastel colors
Use color to enhance information





1.6 The Visual Language

Images are perceived as a set of signs.

The emettor code the information with signs.

The receptor receives and decodes the information from theses signs

« J. Bertin, Sémiologie Graphique. Les diagrammes, les réseaux, les cartes, Paris, La Haye, Mouton, Gauthier-Villars, 1967 »

- **Definition of Graphical semiology:**
- « The set of rules that allow the use of a graphical system of signs in order to transmit the information ».

1.6 The Visual Language

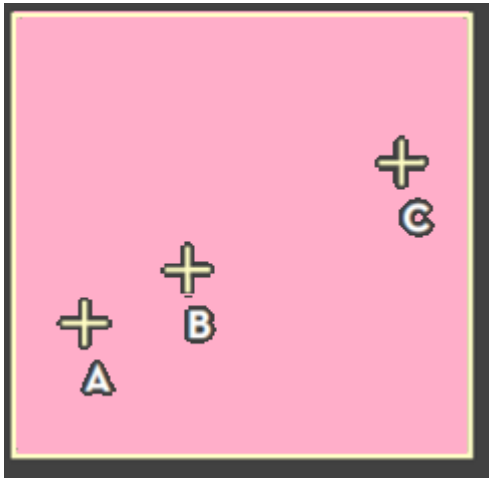
Les variables de Jacques Bertin

Cartographe français,
auteur de la sémiologie graphique

1.6 The Visual Language

Semiology is a science of signs

Example of graphical semiology (Jacques Bertin, 1967)

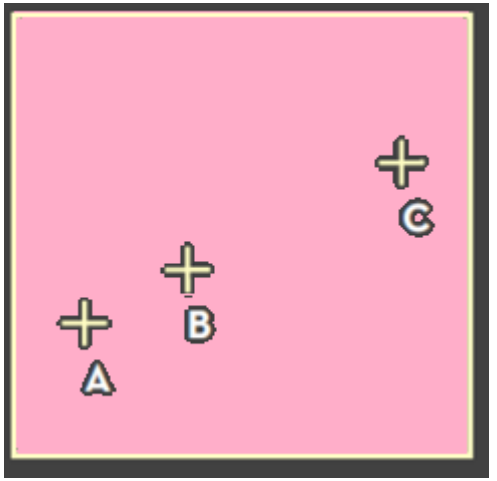


1.6 The Visual Language

Semiology is a science of signs

Example of graphical semiology (Jacques Bertin, 1967)

A, B, C are distinct
B is between A and C
BC is two times more length than AB

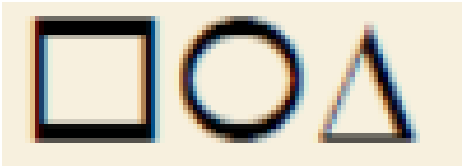


1.6 The Visual Language

Types de variables:

- Nominal,
- Ordinal,
- Quantitative

N – Nominal
Fruits: Apple, oranges, …
Shape: square, triangle, …
Of transport: truck, boat



1.6 The Visual Language

Types de variables:

- Nominal,
- Ordinal,
- Quantitative

N – Nominal

Category: sex (M/F)
Quality (quality of chocolate: b, n,..)
Color: r, v, b, j (whatever the permutation)

1.6 The Visual Language

Types de variables:

- Nominal,
- Ordinal,
- Quantitative

O – Ordinal:
Quality of meal: A, AA, AAA
Mention: P, AB, B, TB
Size: S, M, L, XL
Rang, Satisfaction, Fanciness

1.6 The Visual Language

Types de variables:

- Nominal,
- Ordinal,
- Quantitative

Q – Quantitative:

Interval: age
Ratio: measure: (discrete or continue)

1.6 The Visual Language

Types de variables:

- Nominal,
- Ordinal,
- Quantitative

Operations:

N –Operations: =, ≠
O –Operations: =, ≠, <, >
Q - Operations: =, ≠, <, > - %

1.6 The Visual Language

- **Example: N, O, Q?**

• Population	• Q-Ratio
• Year	• Q-Interval
• Age	• Q-Ratio
• Sex (M/F)	• N
• Familial Situation	• N

1.6 The Visual Language

	Point	Region
Space	(Latitude, Longitude) e.g. “My current location” **Categorical**	Geographic Identifier e.g. Countries, States, Counties **Categorical**
Time	Instant e.g. “Right now” **Quantitative**	Interval e.g. “This year” **Ordinal**
Quantity	Value e.g. 5.2 **Quantitative**	Interval e.g. “5 - 10 years old” **Ordinal**

1.6 The Visual Language

Visual variables:

- Size,
- Shape,
- Color,
- Value,
- Texture,
- Orientation.

1.6 The Visual Language

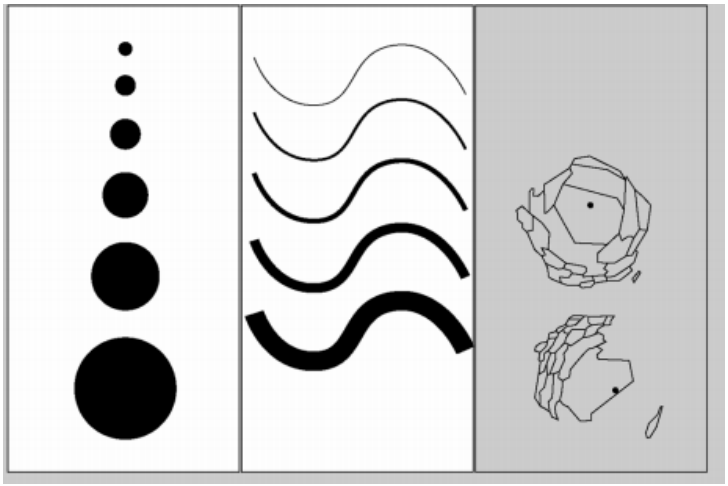
Size :

- The representation using a symbol of a quantitative variable imply a modification proportional of its size related to the visualized value.
- The change of the size is performed on the width, the length, or the superficies of the symbol.

1.6 The Visual Language

Size :

- The representation using a symbol of a quantitative variable imply a modification proportional of its size related to the visualized value.
- The change of the size is performed on the width, the length, or the superficies of the symbol.



Points, lines, superficies

1.6 The Visual Language

Shape:

A change in form more often implies a qualitative change in the variable. Depending on the class to which the observation belongs, the visual form of the symbol will vary.

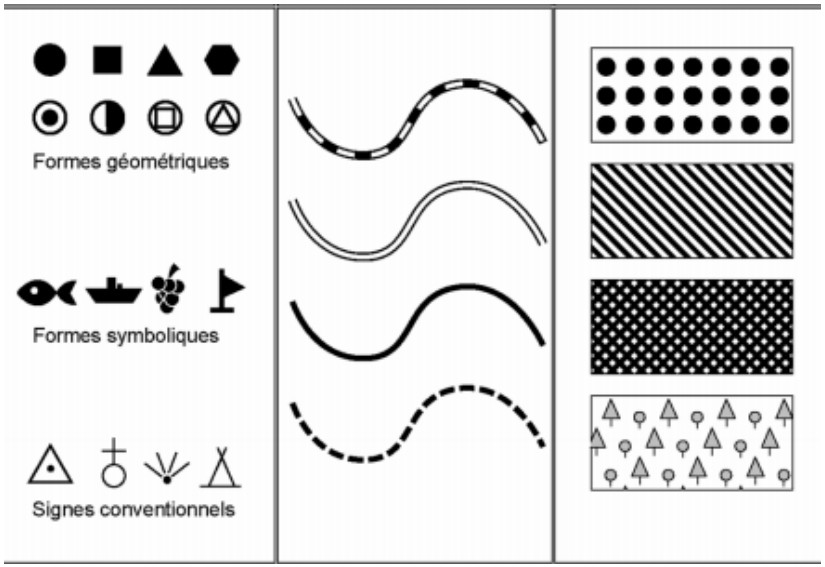
Types of establishment (emergency station, clinic, CHR, CHU)

Transportation

1.6 The Visual Language

Shape:

A change in form more often implies a qualitative change in the variable. Depending on the class to which the observation belongs, the visual form of the symbol will vary.

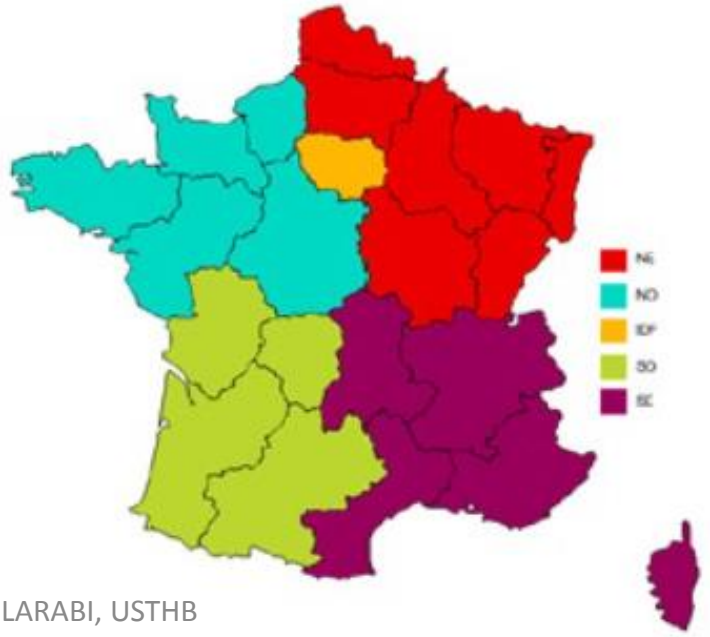


Points, lines, superficies

1.6 The Visual Language

Colour:

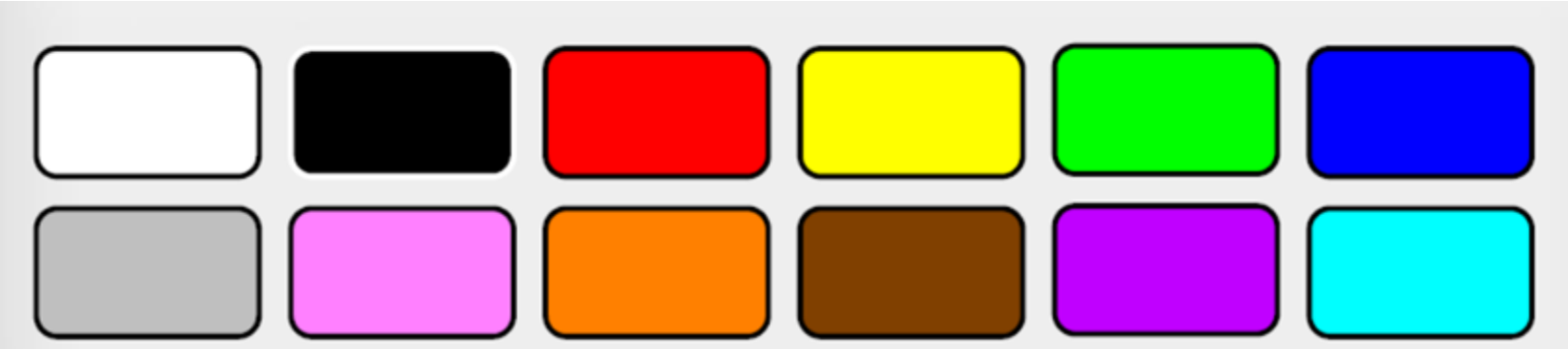
- The variation of symbol by color benefits from excellent perception by the human eye.
- Color can vary according to three shades: tone, saturation and intensity



1.6 The Visual Language

Colour:

12 recommended colours



[Healy, "Choosing effective colors for data visualization" Proc. Visualization, 1996]

1.6 The Visual Language

Colour:

The Hue:

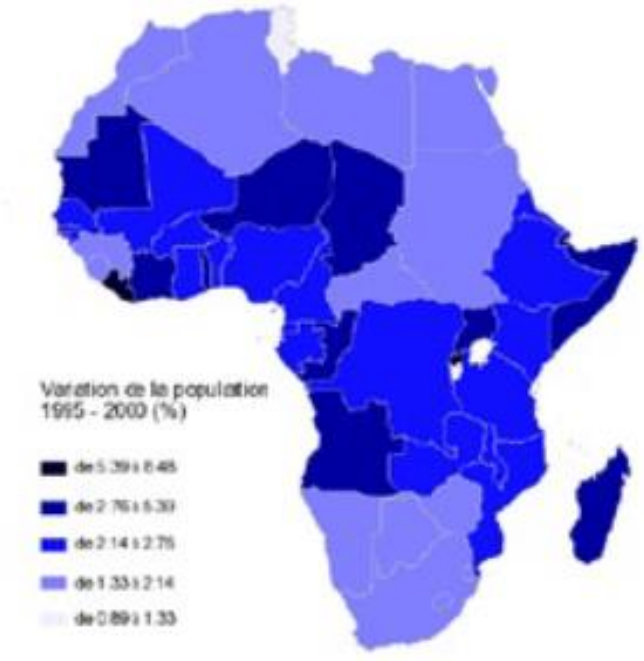
- Variation in tone (dominant color) is more frequently used for a qualitative variable.

1.6 The Visual Language

Colour:

The Saturation:

The variation of intensity and/or saturation makes it possible to construct palettes of gradients. They illustrate changes in quantitative data.



1.6 The Visual Language

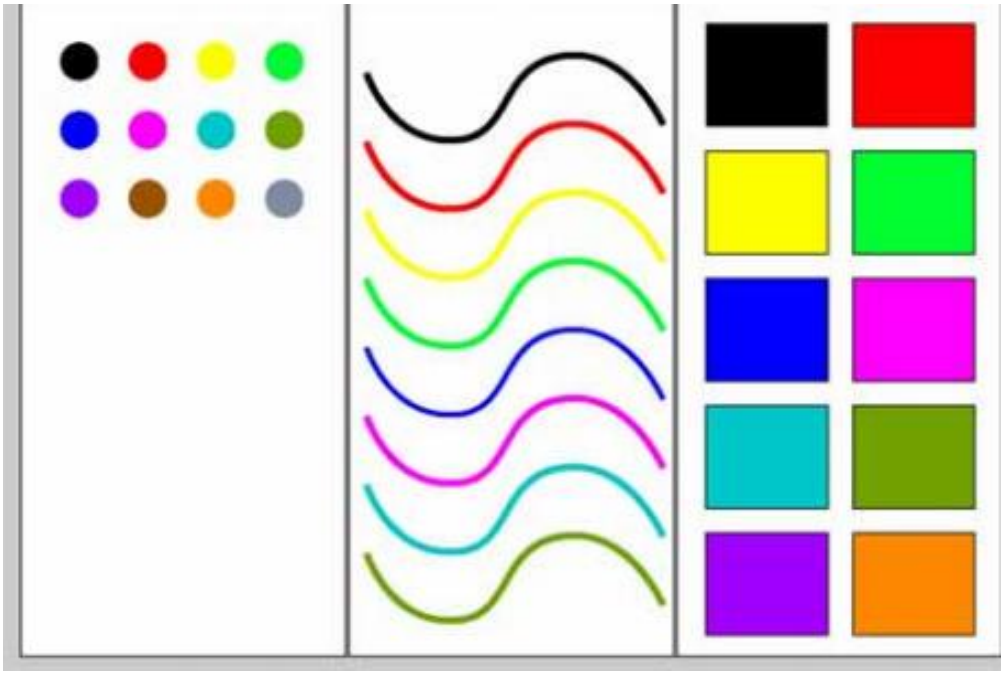
Colour:

The Intensity:

Intensity also corresponds to the visual variable of “value”

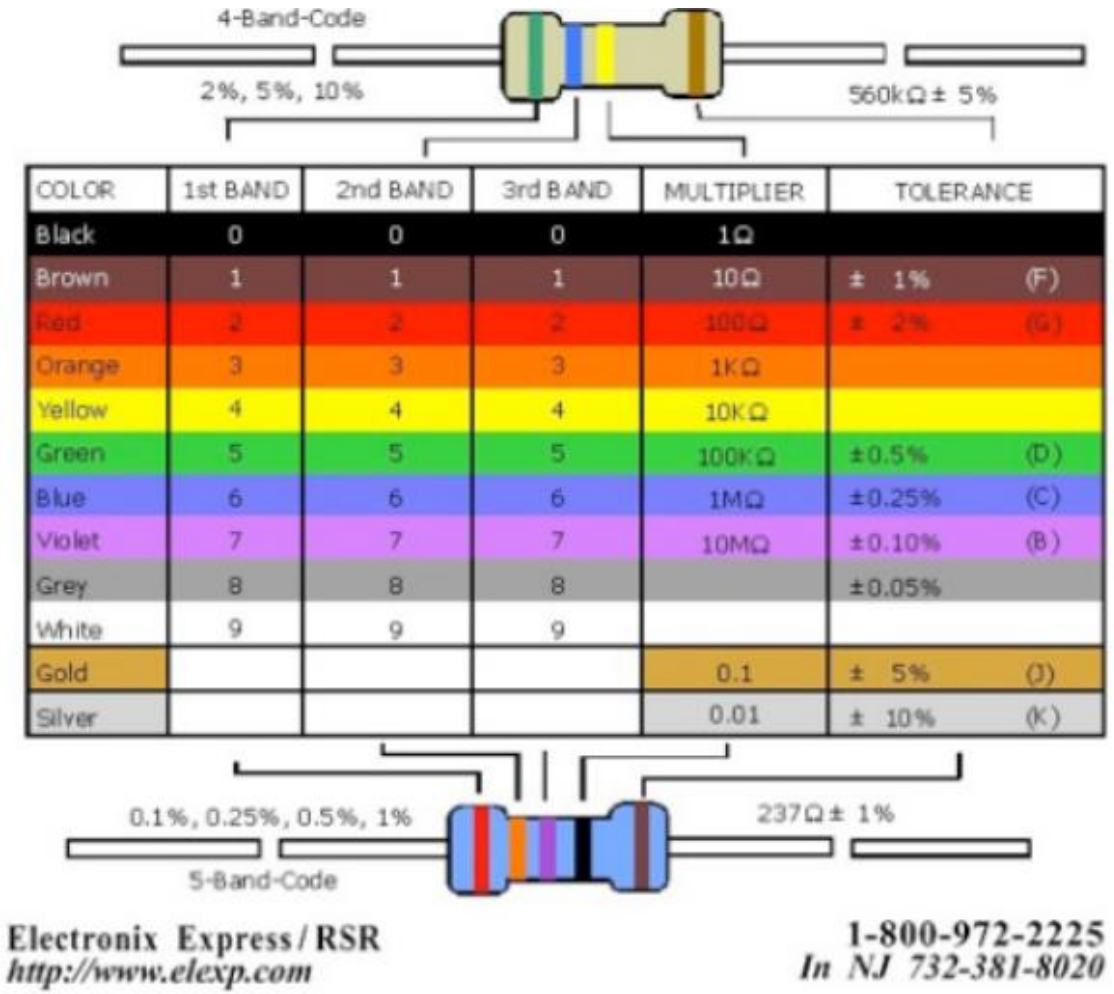
1.6 The Visual Language

Colour:



1.6 The Visual Language

Colour:

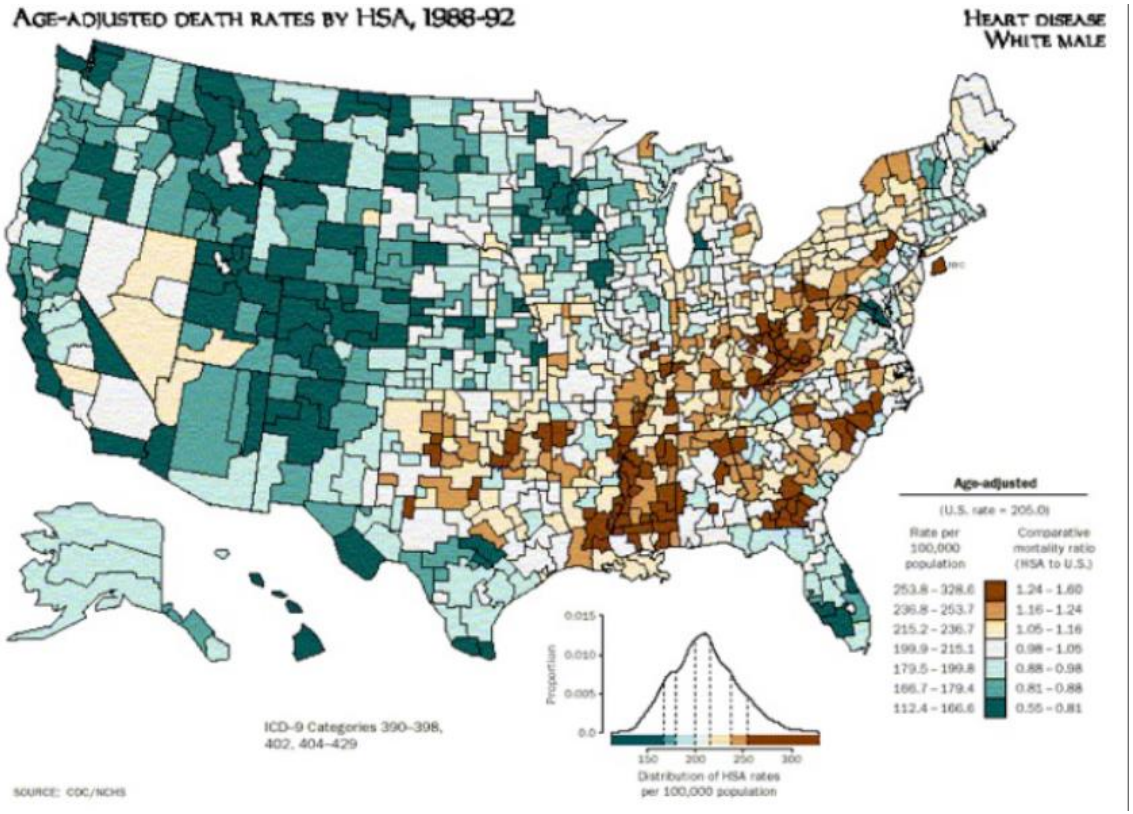


COLOR	1st BAND	2nd BAND	3rd BAND	MULTIPLIER	TOLERANCE	
Black	0	0	0	1Ω		
Brown	1	1	1	10Ω	± 1%	(F)
Red	2	2	2	100Ω	± 2%	(G)
Orange	3	3	3	1KΩ		
Yellow	4	4	4	10KΩ		
Green	5	5	5	100KΩ	±0.5%	(D)
Blue	6	6	6	1MΩ	±0.25%	(C)
Violet	7	7	7	10MΩ	±0.10%	(B)
Grey	8	8	8		±0.05%	
White	9	9	9			
Gold				0.1	± 5%	(J)
Silver				0.01	± 10%	(K)

1.6 The Visual Language

Colour:

Quantitative Colour



1.6 The Visual Language

The value

- The value is “the ratio between the quantity of black and white on a given surface” (D. Poidevin, “The map, Means of action”).
- Black is considered to have greater value than white. Thus, a high value will be represented with a greater proportion of black than white. Value variations will therefore be used for quantitative data.

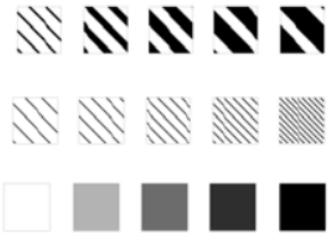
1.6 The Visual Language

The value

In a variation by color, this representation will be influenced by the "intensity" of the color.

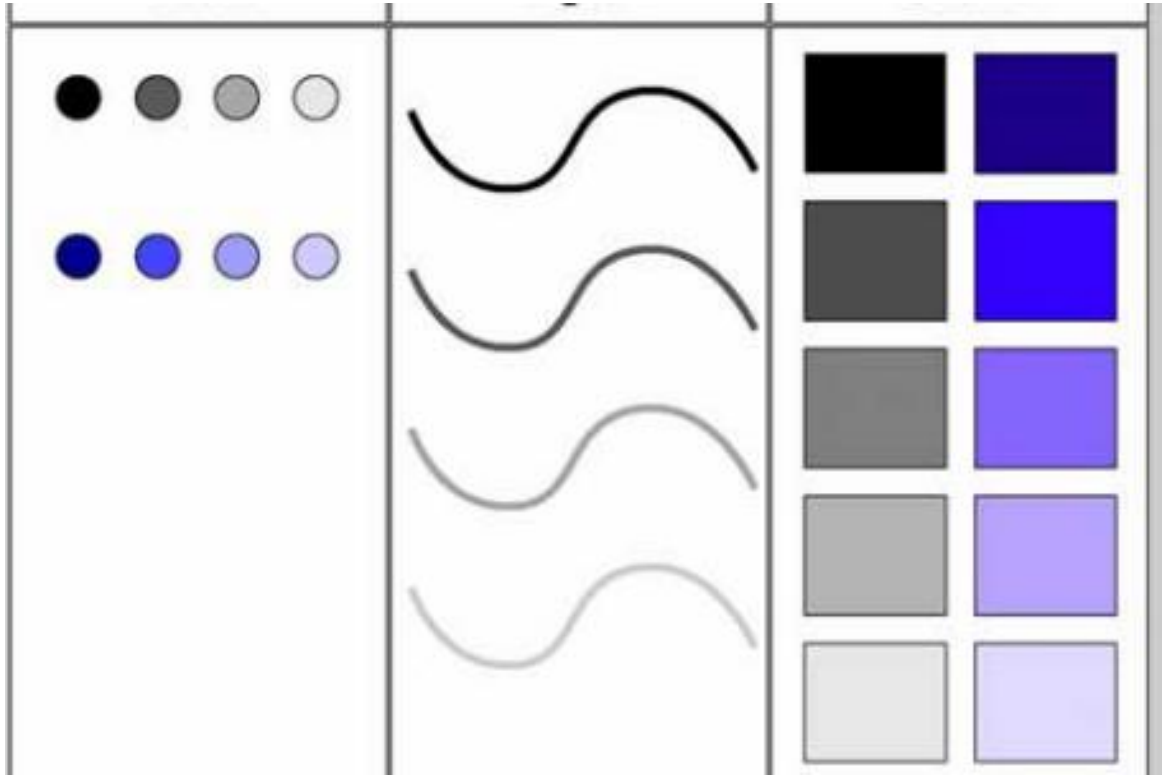
But the value can also be brought into play in other types of representation (hatching, dotted lines, grains, etc.).

- Modification of the value by the thickness of the line:
- Modification of the value by the spacing of the line:
- Modification of the value by the hue:



1.6 The Visual Language

The value



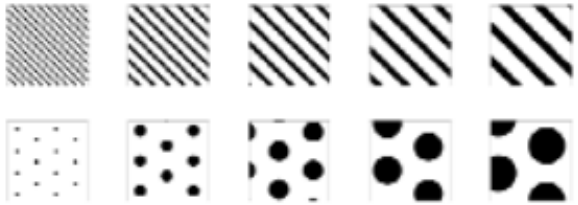
1.6 The Visual Language

The texture

Unlike variation by value, which involves a white-black evolution, variation by grain aims to maintain this balance. Here, it is the number of elements in the frame that varies, and not the value.

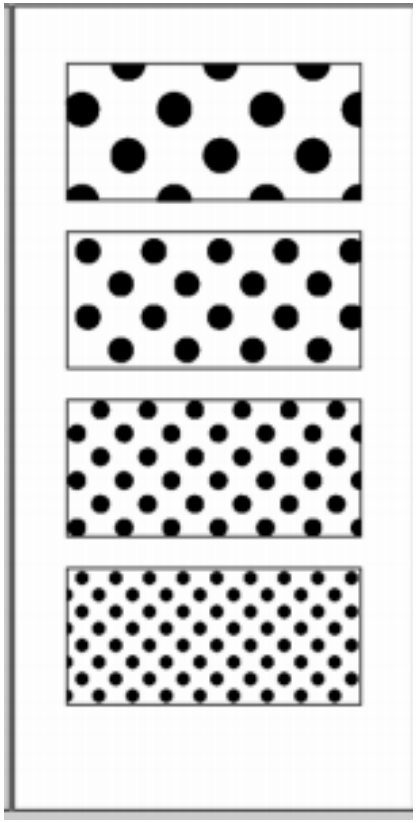
To maintain this balance, a grain variation begins with a very tight pattern of very small spots, and ends with a small number of thick spots spaced more widely apart.

This variation can be identified by a zoom on a specific train (the size/spacing ratio of the elements is preserved).



1.6 The Visual Language

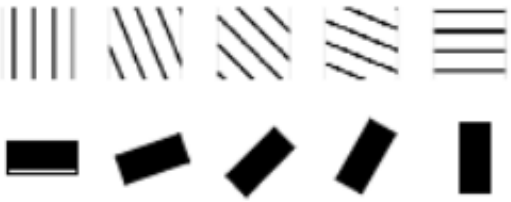
The texture



1.6 The Visual Language

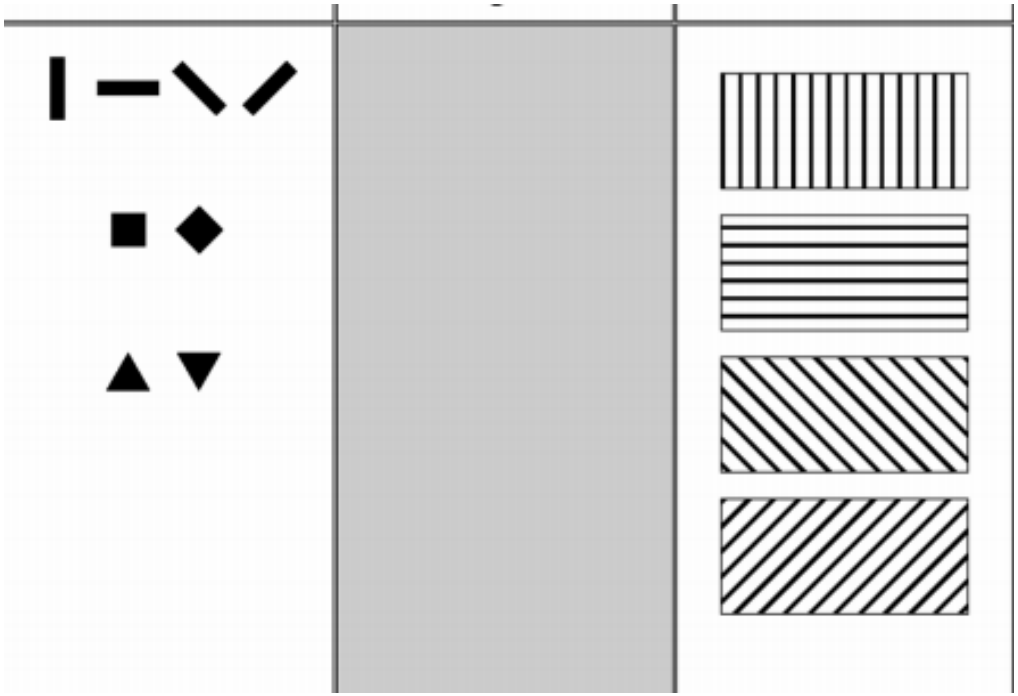
The Orientation

The sixth visual variable, orientation, involves changing the angle of the hatching in the representation. This change in orientation makes it possible to distinguish certain classes of observation (qualitative data). It concerns point or surface objects.
In surface representation, the distinction is sometimes difficult to make when too many classes are recorded. This variation by orientation should therefore be used with caution.



1.6 The Visual Language

The Orientation



References

[McCormick et al. 1987]
McCormick, B. H., DeFanti, T. A., & Brown, M. D. (1987). Visualization in scientific compu-ting: Report of The NSF Advisory Panel on Graphics, Image Processing and Workstations.
http://www.sci.utah.edu/vrc2005/McCormick-1987-VSC.pdf
[Bertin 1967]
Jacques Bertin (1967), Sémiologie graphique. Les diagrammes. Les réseaux. Les cartes, Paris/La Haye, Mouton ; Paris, Gauthier-Villars.
[John Tukey, 1977]
John W. Tukey Exploratory Data Analysis, 1977
[Card et al. 1999]
Stuart Card, Xerox PARC, Jock Mackinlay, Xerox PARC, Ben Schneiderman,
Readings in Information Visualization: **Using Vision to Think**,, U. Maryland
[**Alexandru 2008**]
Alexandru C. Telea , "**Data Visualization** -- **Principles and Practice**", 2008. A K Peters/CRC Press,
[W1]
(https://fr.statista.com/)
[W2]
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7150208/
[W3]
https://www.andlil.com/mise-en-place-station-trading-6273.html
[W4]
https://www.01net.com/actualites/deces-de-benoit-mandelbrot-le-papa-des-objets-fractals-522424.html