# 02-2 Data representations

Monday, March 13, 2023 17:34

2-2 Data representa...

1

# DATA REPRESENTATION

Sungahn Ko
sako@unist.ac.kr

1

## Disclaimer

2

- The slides cannot be distributed, posted or used outside of this class
- Slides in this course courtesy of
  - Dr. Abish Malik (Purdue)
  - Dr. Yun Jang (Sejong Univ.)
  - Dr. Ross Maciejewski (ASU)
  - Dr. Niklas Elmqvist (UMD)
  - Dr. David Ebert (Purdue)

2

## Data Models

3

- Data models/representations are structured forms suitable for computer-based transformations
- These structures exist in the original data or are derivable from the original data
- Structures retain the information and knowledge content and the related context within the original data
- These structures are transformable into lower-dimensional representations for visualization and analysis

3

## Data Models vs. Conceptual Models

4

- Data models are mathematical abstractions
  - We can perform numerical operations, addition, subtraction, etc.
- Conceptual models are our mental constructs
  - These contain semantic structure and support reasoning
  - Think of giving directions to someone using landmarks

4

## What Types of Data Models Do We Have?

5

Statistical

- ?? **Data Models**
  - These are data records that you deal with in excel, or typical database records
  - Data records are of fixed length and well-defined
  - Data is in rows and columns where each column has a domain type
  - There may be a relationship defined in a schema between tables
  - A database is a collection of these relations

| | A | B | C |
|---|---|---|---|
| 1 | Animal | Brain Weight (kg) | Body Weight (kg) |
| 2 | Mountain beav | 1.35 | 465 |
| 3 | Cow | 465 | 423 |
| 4 | Grey wolf | 36.33 | 119.5 |
| 5 | Goat | 27.66 | 115 |
| 6 | Guinea pig | 1.04 | 5.5 |
| 7 | Dipliodocus | 11700 | 50 |
| 8 | Asian elephant | 2547 | 4603 |

, E.F. (1970). “A Relational Model of Data for Large Shared Data Banks”. Communications of the ACM 13(6).

5

## What Types of Data Models Do We Have?

6

Table

- ?? **Data Models**
  - This consists of variables or measurements (think of things like the census)
  - Categories are factors relating to the measurements
  - Observations or cases

| Patient ID | Date | Chief Complaints | Location |
|---|---|---|---|
| 9398 | 4/16/09 | Ear pain | Karachi |
| 10816 | 4/16/09 | Stuffy nose | Lebanon |
| 1491 | 4/16/09 | Fever | Allepo |
| 16237 | 4/16/09 | Head bleed | Yemen |

6

## Data Types

7

- Nominal ??
  - Data whose categories have no implied ordering
  - Examples include political affiliations of a population
- Ordinal ??
  - Data that has a specified order, but no specified distance metric
  - Examples include beverage sizes at McDonalds (Small, medium, large)
- Interval ??
  - Data that has measurable distances
  - Examples include periods of time (second, minute, etc.) – the zero point is arbitrary
- Ratio ??
  - Same as interval, but include a zero point
  - Example include Celsius scale, height above sea level

SS Stevens, "On the Theory of Scales of Measurement,"

7

## Mapping Data

8

- We need to know how to assign quantitative dimensions of our data to *aesthetic attributes*[1] of the data

| ?? Form | ?? Surface | ?? Motion | ?? Sound | ?? Test |
|---|---|---|---|---|
| Position<br>Size<br>Shape<br>polygon<br>glyph<br>image<br>Rotation<br>Resolution | Color<br>hue<br>brightness<br>saturation<br>Texture<br>pattern<br>granularity<br>orientation<br>Blur<br>Transparency | Direction<br>Speed<br>Acceleration | Tone<br>Volume<br>Rhythm<br>Voice | Label |

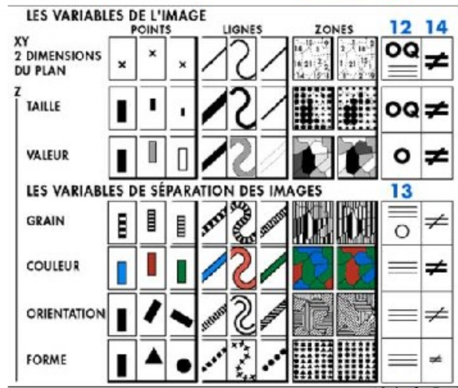
8

Aesthetic Attributes < represent both cont. & categorical vari

# 8 Visual Variables

10

- Visualization is concerned primarily with a mapping to visual form
- [x,y]
  - Position
- [z]
  - Size (Taille)
  - Value (Valeur)
  - Color (Couleur)
  - Texture (Grain)
  - Orientation
  - Shape (Forme)




10

# Position

11

- **Position** refers to a location in a multi-dimensional space
- **Continuous variables** map to densely distributed locations
- **Categorical variables** map to a lattice
- Positions are ordered, but the ordering may or may not have meaning in terms of what is being measured
- Sometimes, position is just a way to keep things from overlapping

11

## ?? Position[1]

12

- Cleveland[2] rates position ?? on a common scale as the **best way to represent** a **quantitative dimension visually**
- This reflects research findings that points or line lengths placed adjacent to a common axis enable judgments with the least bias or error
- However, this recommendation has a caveat, as it depends on how far the graphic primitive (point, line, etc.) is from a reference axis[3]
- If a graphic is far from an axis, the multiple steps needed to store and decode the variation can impair judgment

12

## Size[1]

13

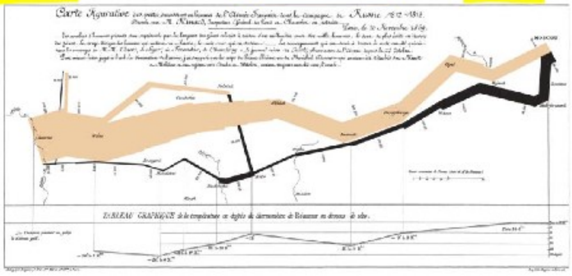
- Bertin[2] defines size variation in terms of length or area
- For three dimensions we have volume
- Cleveland[3] ranks area ?? and volume ?? representations among the **worst attributes** to use for graphing data
- Some designers assign size to only one dimension of an object
- Think of the bar chart where the width of the bar is typically constant, but the height is varied

13

## Size

14

- Size for lines is usually equivalent to thickness
- This is less likely to induce perceptual distortion
- Size can be used to great effect with path



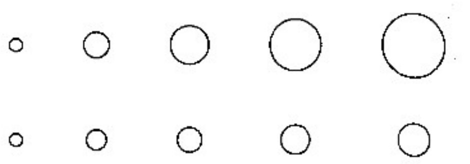
– : Mapping Napoleon's

14

## Size

15

- For objects with rotational symmetry, we can map size to the diameter rather than area
- Representing data through area or volume should probably be confined to positively skewed data that can benefit from the perceptual equivalent of the square root transformation



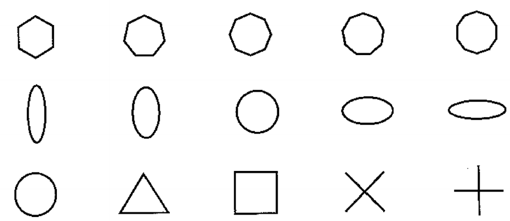
Top row changes diameter from 1-5
Bottom row changes area from 1-5

15

## Shape

16

- Shape refers to the shape or boundary of an object
- Examples would include map symbols
- Shape must vary without affecting size, rotation and other attributes

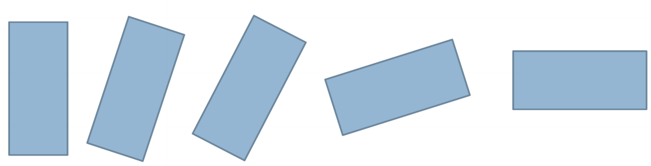


16

## Rotation

17

- This is the rotational angle of the graphic primitive
- Lines, areas and surfaces can only rotate if they are positionally unconstrained

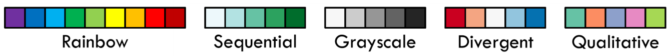


17

# Color

18

- Color is really a psychological phenomenon
- The physical stimulus for color is light
- We see color because of the photoreceptors in our retina
- We will discuss color in-depth in a future lecture
- Use of Rainbow?




18

# Texture

20

- Texture includes pattern, granularity, and orientation
- Granularity is the repetition of a pattern per unit of area
- Orientation is the angle of the pattern
- Textures can be described in a variety of ways
  - **Fourier transform** – decomposes a grid of brightness values into sums of trigonometric components
  - This decomposition is orientation dependent
  - **Auto-correlogram** – characterize the spatial moments of a texture

20

## Texture

21

- Texture alone can be a basis for perception
- Two gray areas that have the same overall level of brightness can be discriminated if their texture is different

21

## Texture

22

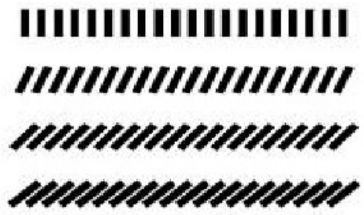
- **Granularity** – Changing the size or resolution of patterns in a texture changes the granularity
  - Less grainy patterns (those with low-frequency spatial components) are more difficult to resolve
- **Pattern** – Patterns make use of increasing degrees of randomness to encode data

22

# Texture

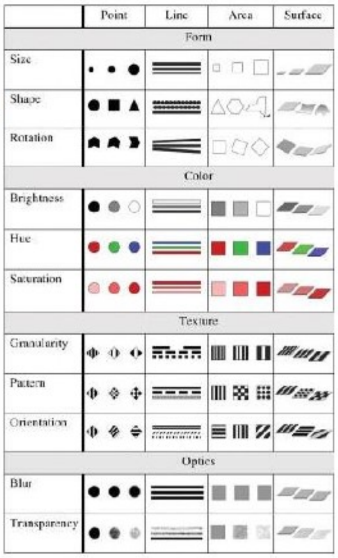
23

- **Orientation** – Variation in texture orientation can introduce visual illusions (making lines not seem parallel) and is typically avoided as an encoding in textures



23

24




24

## Groups of Representation

25

- Bertin describes four groups of representations
- *Diagrams* – correspondences are between a single component (think of things like a bar chart or a line graph)
- *Networks* – correspondences between different components are linked together
- *Maps* – correspondences between components are arranged according to geography
- *Symbols* – correspondences between components is linked to a symbol or a shape

25

## Taxonomy by Data Type[1]

26

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)
- Text and Documents[2]

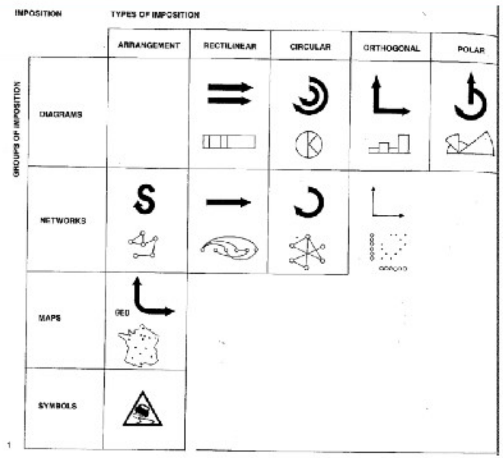
–

– Hanrahan's "From Data to Image" lecture: http://graphics.stanford.edu/courses/cs448

26

# Types of Graphic Representation

27




27

# Cleveland's Hierarchy[1]

28

- Cleveland evaluated elements when isolated
- Tasks were restricted to magnitude and ratio comparisons
- Research indicates this hierarchy may be best in pre-attentive stages or when focusing only on portions of a graphic[2,3,4]

Best

1 - Position along a common scale
2 – Position along nonaligned scales
3 – Length
4 – Angle/Slope
5 – Area
6 – Volume
7 - Color

Worst

Not high ranked in some senses

28