Introduction to Data Vizualisation

Smart Analytics for Big Data

Iragaël Joly
Grenoble INP - UMR GAEL

iragael.joly@grenoble-inp.fr

"2020-10-07"

Introduction

Preliminary to data analysis and vizu

- Raw, unprocessed data is often not ready for visualisation.
- Techniques to turn data into information including :

Derivation and feature extraction

- Derivation and feature extraction : to add to existing data without requiring external datasets.
- additional columns *or features* based upon the existing data

Derived data element : using a mathematical, logical or other type of transformation, e.g. arithmetic formula, composition, aggregation.

Feature extraction : no math function, but a rule.

- a person in a family / a city from list of adresses

Combining datasets

- **Merging or Combining** two datasets with 10 columns each will be one dataset with 19 columns.
- Merging data can only be done with a shared value (jointure key)

Enriching geographic data

- Combine geographic data based upon location : spatial join.
- Essential stage in preparing geographic data for analysis is geocoding
- Geocoding is the process of taking any reference or description of a physical location and adding the actual physical location coordinates to the data.

Qualitative and quantitative data analysis

- qualitative research deals with open-ended, often text-based data
  - Barplot, surface plot, ...
- quantitative research tries to focus on objective, measurable data in the form of numbers or other structured data.
  - boxplot, histogram, scatter plot

Data Visualisation

The goals of visualising data

Communicate information clearly and efficiently to users through graphical means

- Aesthetic form and functionality to communicate its key-aspects of a dataset
- Communication is only one goal of a data visualisation
- Data visualisations as a support during the data analysis stage

Visualization and perception

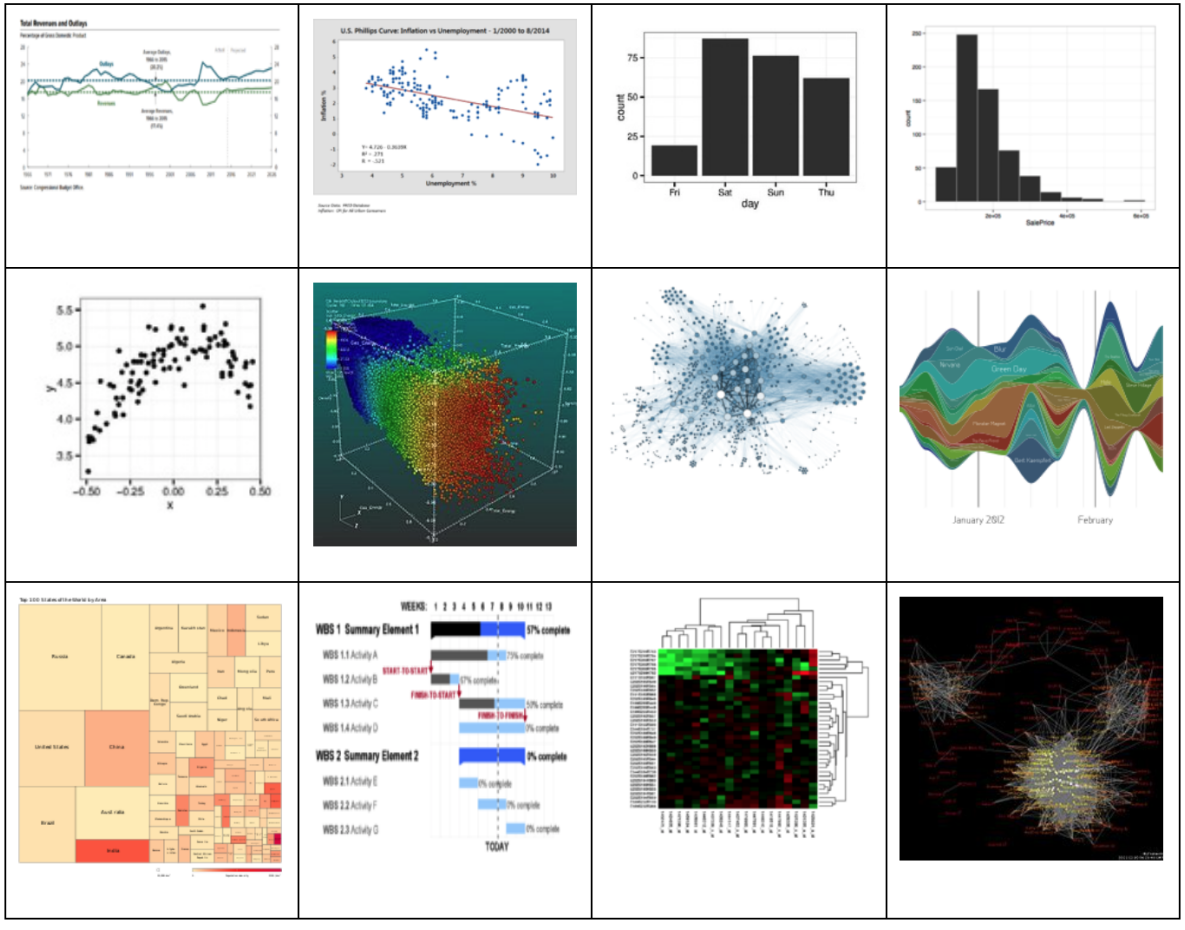


Figure 1 – Examples of data visualisation *from Wikipedia*

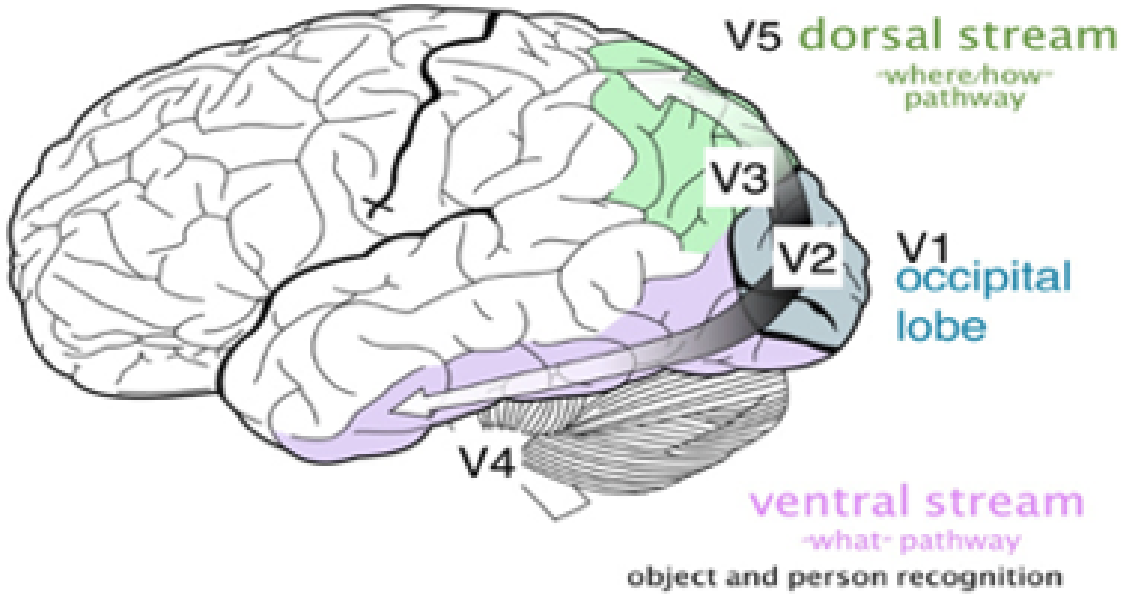


Figure 2 – Pathways in visual cortex

Human eye is drawn towards

- brighter colours
- larger items
- things which stand out through difference.

The pop-out effect - Inside the visual cortex

- **dorsal** stream : *where/how*
  - information about our surroundings in real time
  - where things are and how they relate to other things
- **ventral** stream : *what*
  - 'what' the thing is
  - much slower process

Best data visualisations appeal to the dorsal stream and make information pop

Visualization and colours

Colour makes information ‘pop out’ also convey meaning

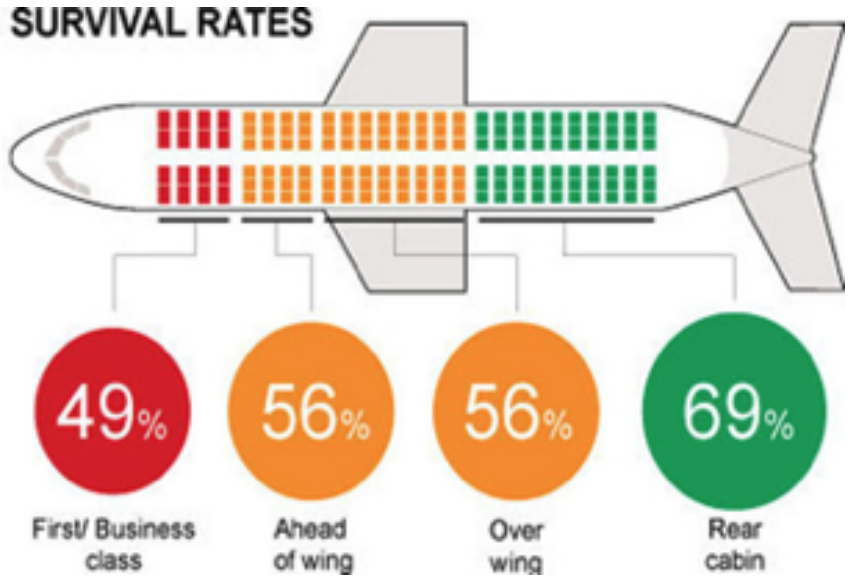


Figure 3 – Traffic-light colours indicating safety

keep the number of colours used to a minimum

Choosing the correct visualisation for data
Choosing the correct visualisation for data depends on two key aspects :

- the type of data
- the message to be conveyed.

Data type	Description	Example visualisation
Time series	Observations of the same objects over time	Line chart, motion chart, polar area diagram, Gantt chart, bar chart
Population	Observation of different objects at a single point in time	Bar chart, map, treemap, pie chart
Multivariate	Observations of different objects at different points in time	Multidimensional motion chart, bar chart, treemap

Example with ultivariate data

Yields of cereal in the East African Community of Tanzania, Rwanda, Uganda, Kenya, Burundi and South Sudan

(available as open data from the World Bank.

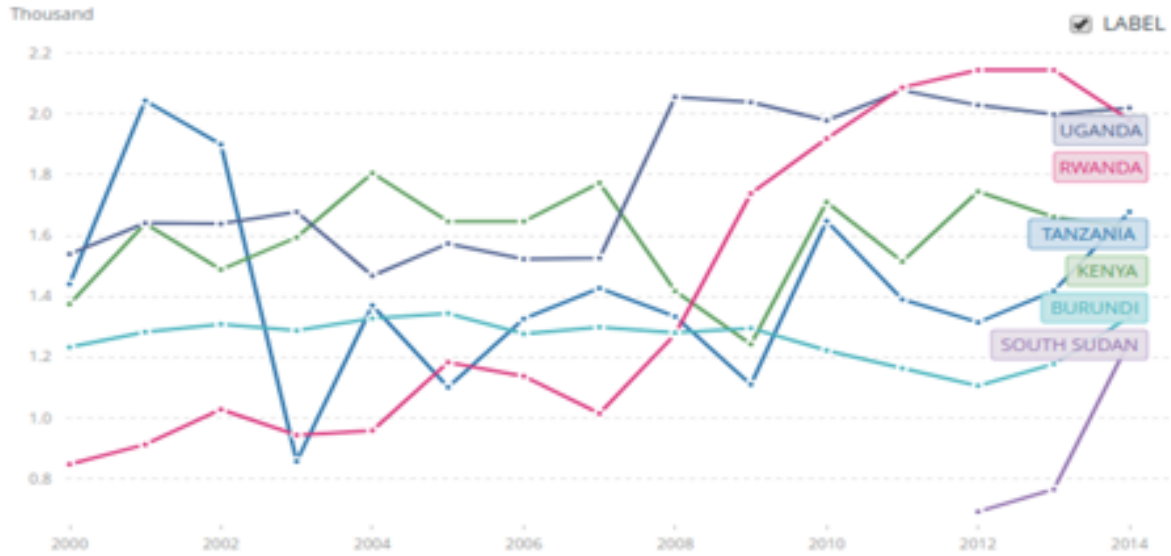


Figure 4 – Time series dataset visualisation of cereal yield in the East African Community since 2000

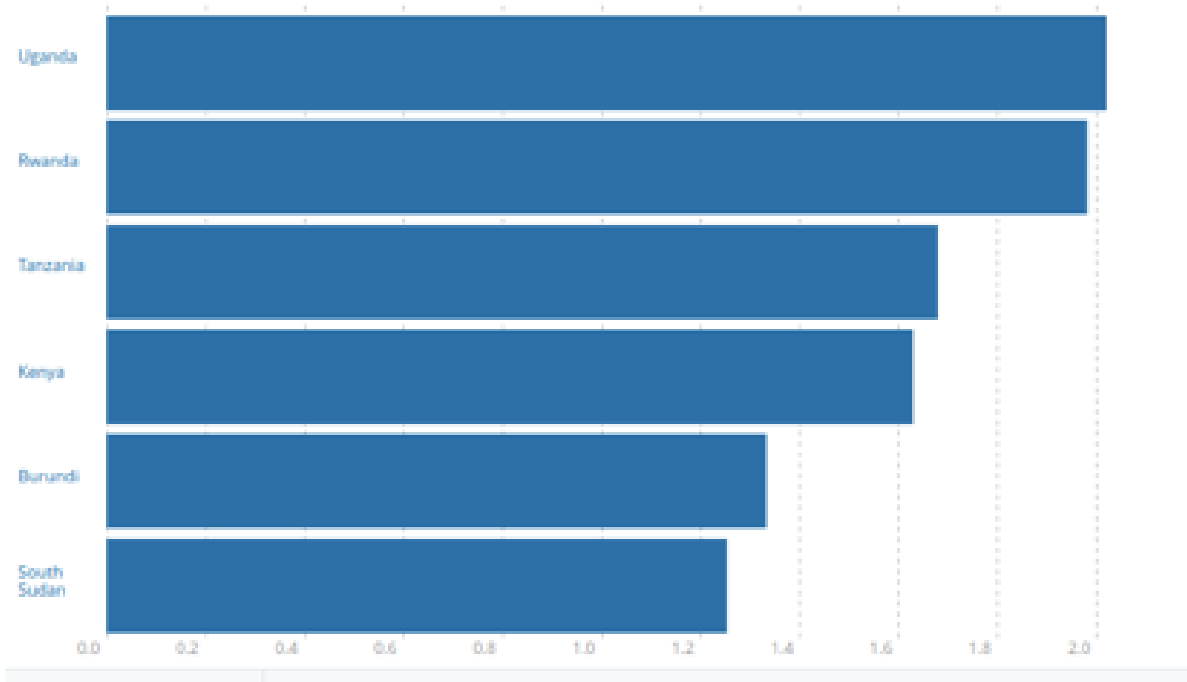


Figure 5 – Population bar chart visualisation of cereal yield in the East African Community in 2014

Figure 6 – EAC countries in alternative mapping

Visualization

Aid to analysis

Summaries lose information, details matter

– confirm expected and find unexpected patterns – assess validity of statistical model

Anscombe's Quartet

Identical statistics

X : mean $=9$; variance$=10$

Y : mean $=8$; variance$=4$

$Cor(X, Y) = 1$

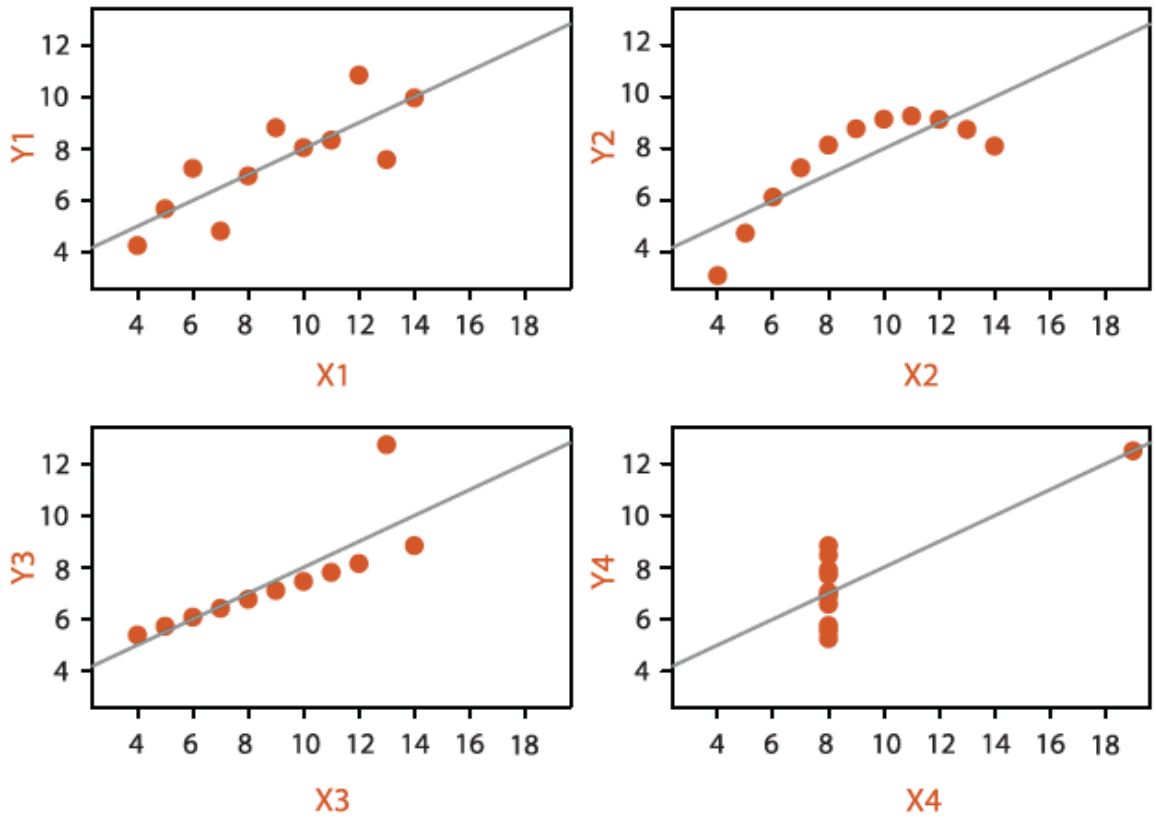


Figure 7 – Anscombe's Quartet

Framework

Munzner (2015) proposes a framework

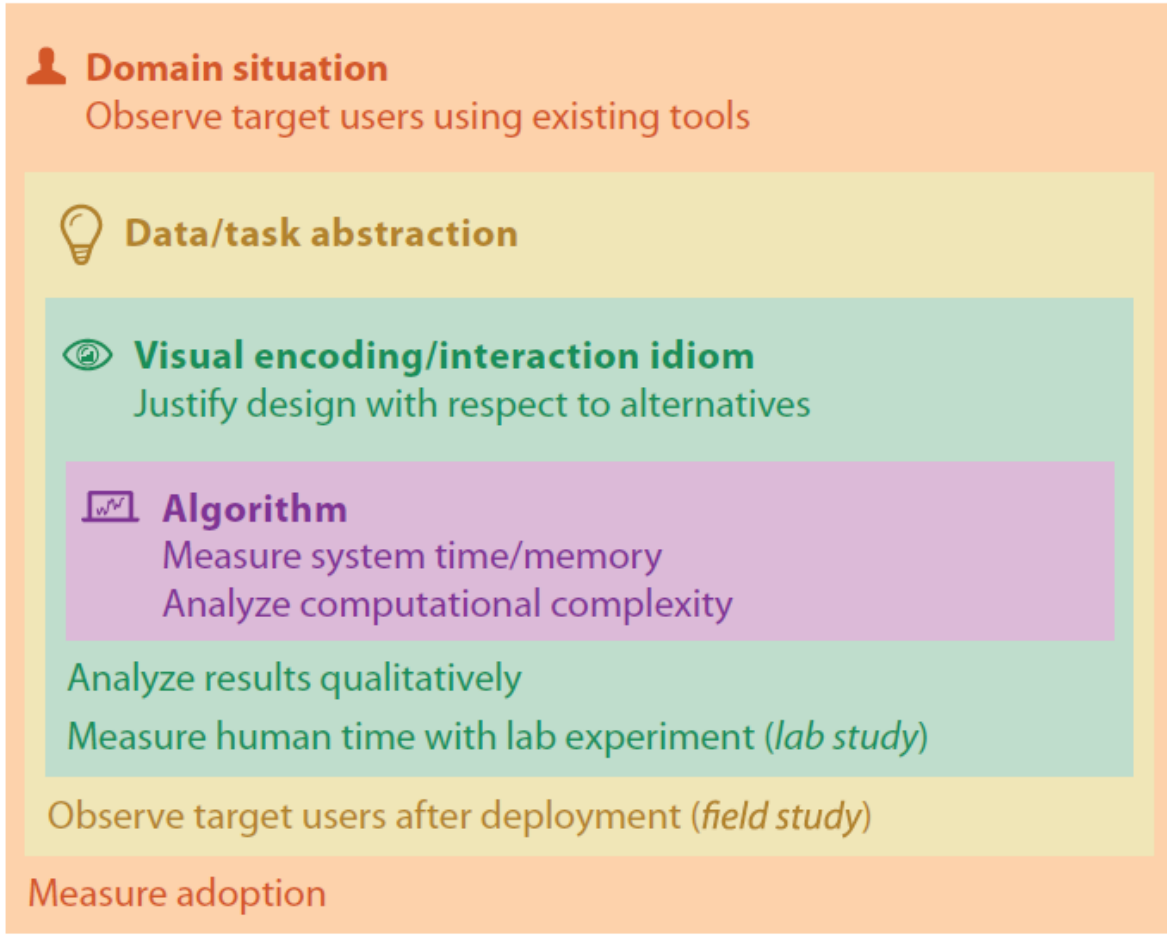


Figure 8 – Munzner Framework

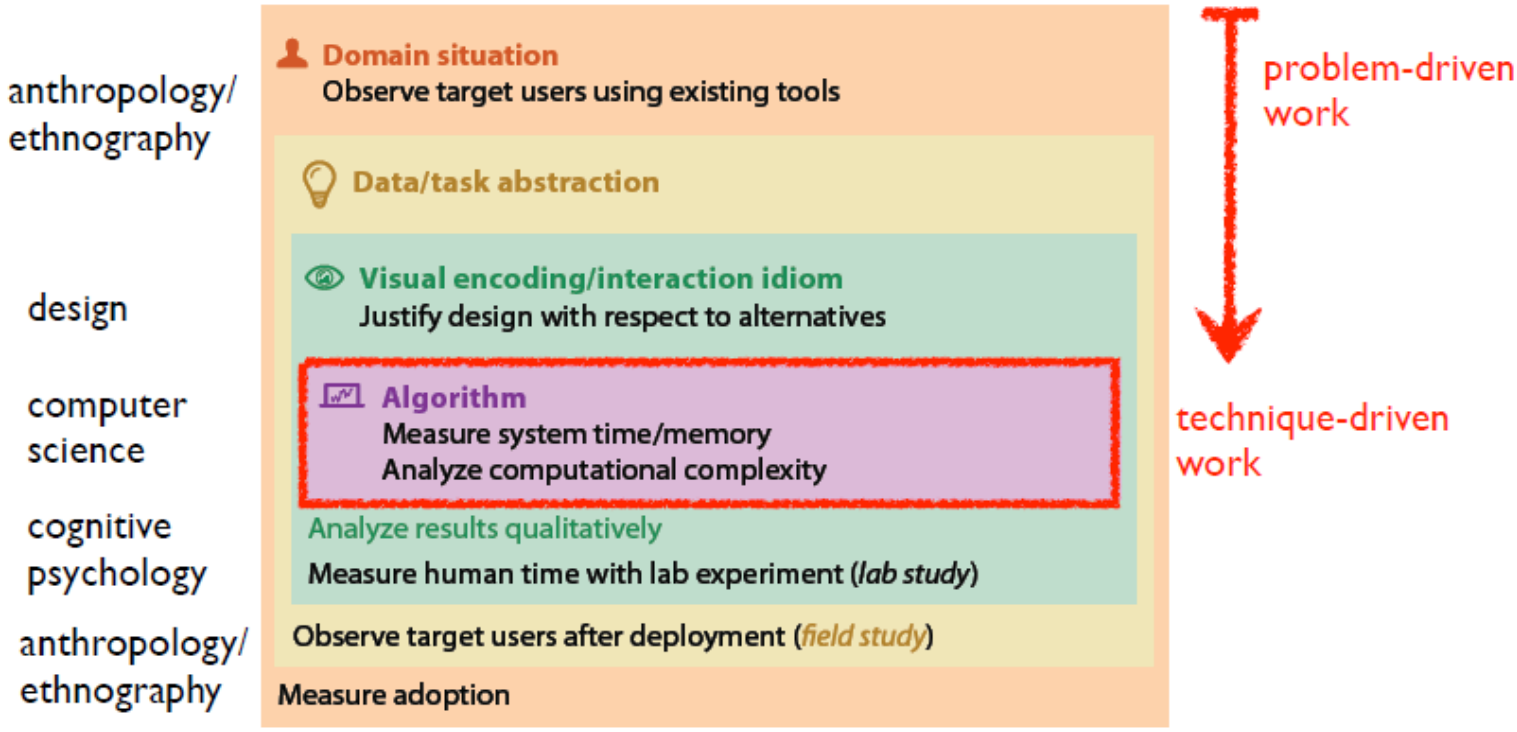


Figure 9 – Munzner Framework

What ? Why ? How ?

What can be visualized : data, datasets, and attributes

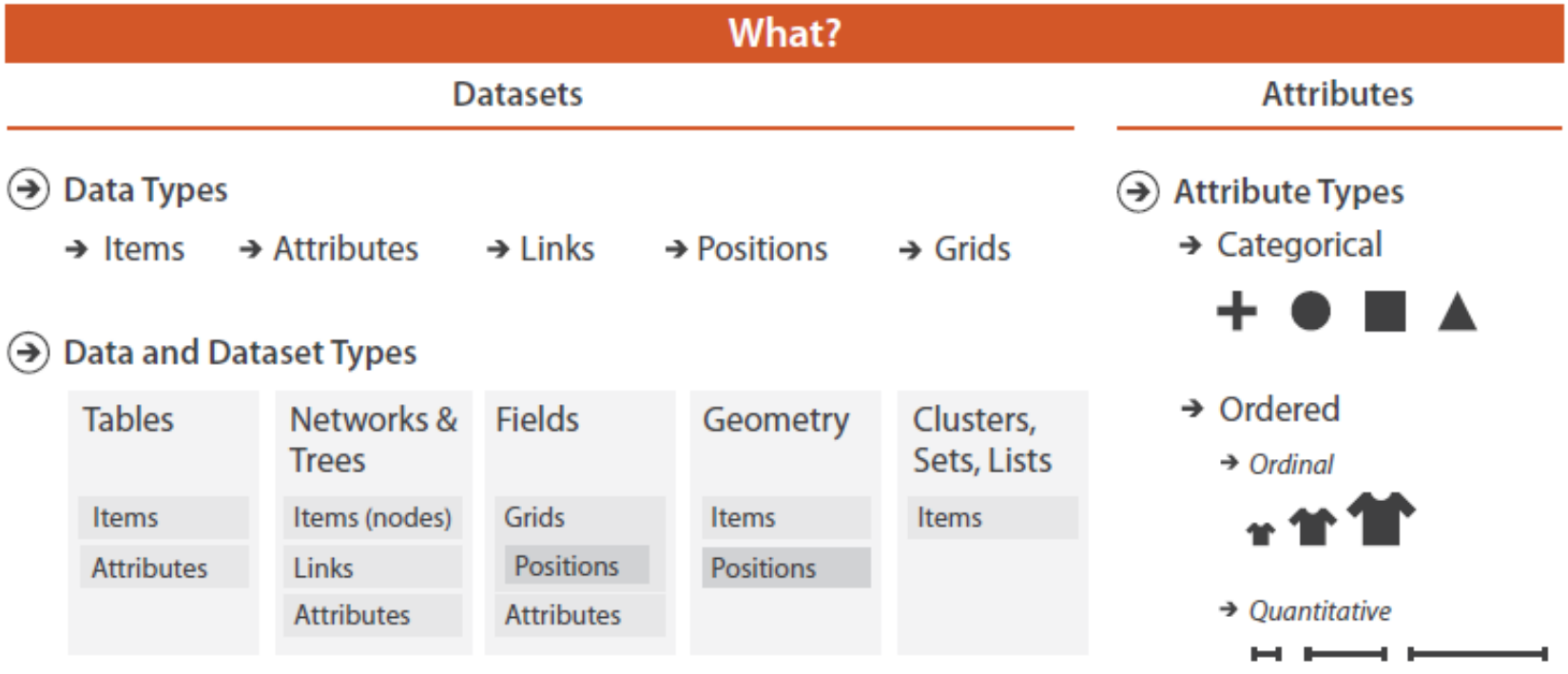
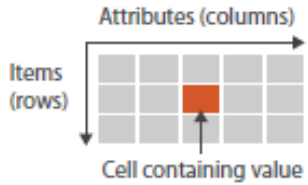


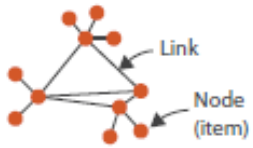
Figure 10 – What - 1

Dataset Types

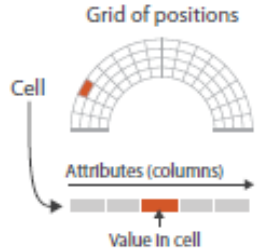
Tables



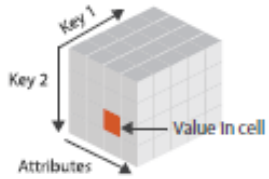
Networks



Fields (Continuous)



Multidimensional Table



Trees

Geometry (Spatial)

Ordering Direction

Sequential

Diverging

Cyclic

Dataset Availability

Static

Dynamic

Figure 11 – What - 2

Why people are using vis in terms of actions and targets

{action, target} pairs :

- discover distribution
- compare trends
- locate outliers
- browse topology

Figure 12 – Why

How to design vis idioms : encode, manipulate, facet, and reduce.

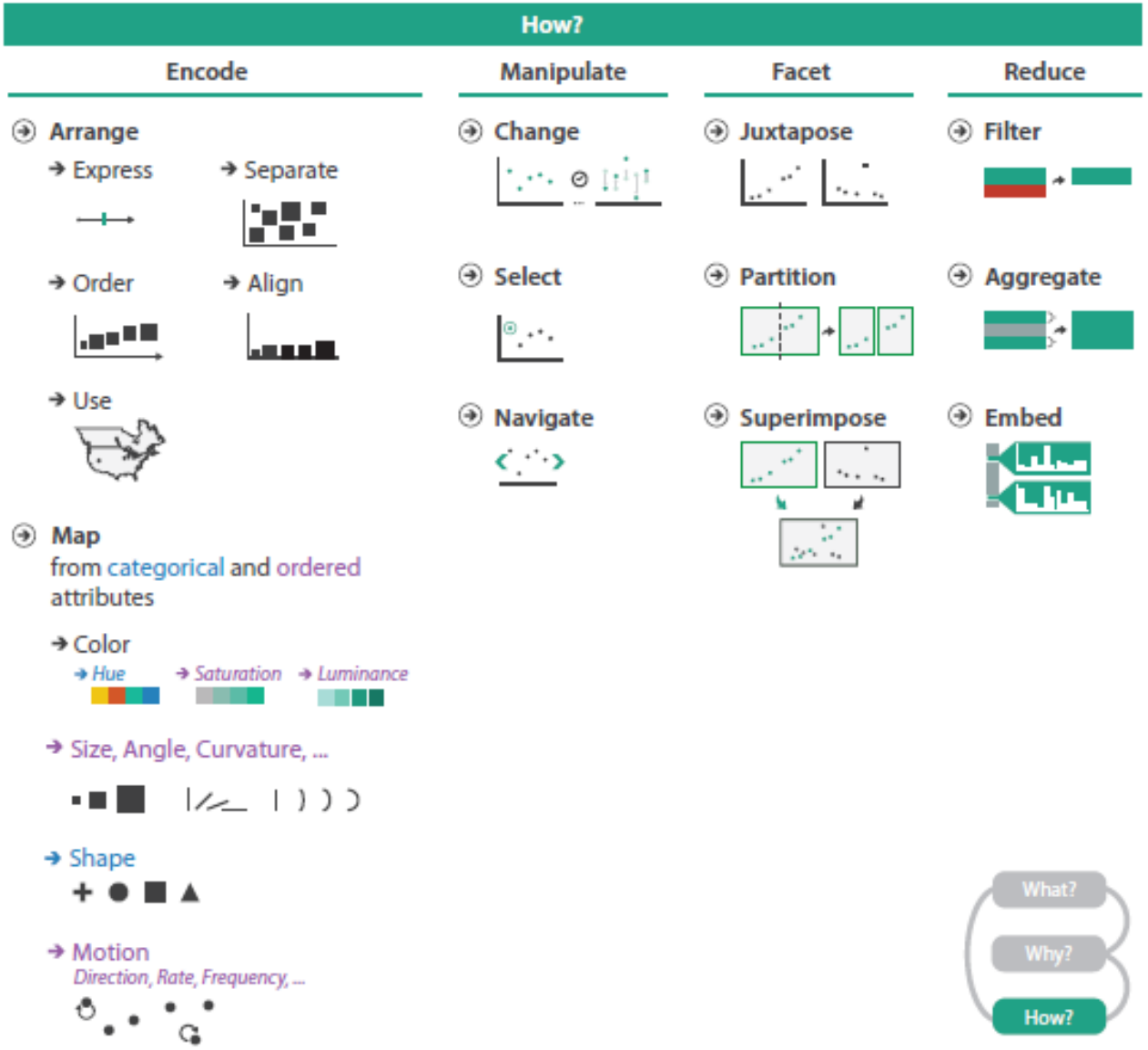


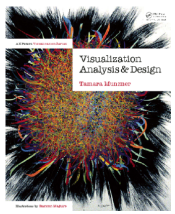
Figure 13 How

Eight rules of thumb.

- No Unjustified 3D
  - The Power of the Plane
  - The Disparity of Depth
  - Occlusion Hides Information
  - Perspective Distortion Dangers
  - Tilted Text Isn't Legible
- No Unjustified 2D
- Eyes Beat Memory
- Resolution over Immersion
- Overview First, Zoom and Filter, Detail on Demand
- Responsiveness Is Required
- Get It Right in Black and White
- Function First, Form Next

Bibliography

- Stephanie Glen. “Misleading Graphs : Real Life Examples” From StatisticsHowTo.com : Elementary Statistics for the rest of us ! - (https://www.statisticshowto.com/misleading-graphs/)
- https://datavizcatalogue.com/resources.html
- Tamara Munzner : Talks



Munzner, Tamara. 2015. *Visualization Analysis and Design*. CRC Press, Routledge.