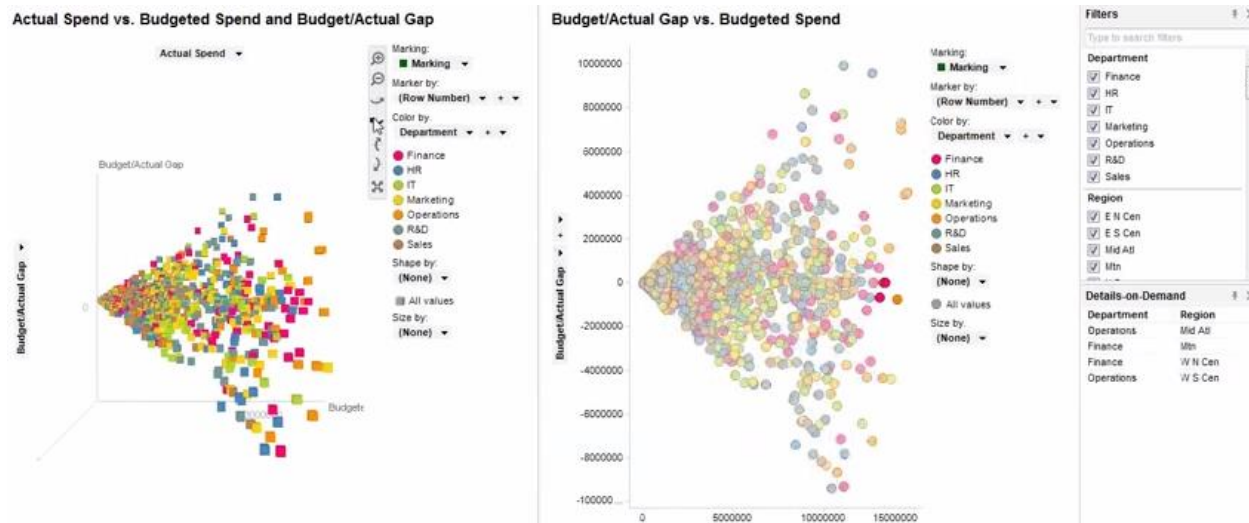




Issues in Visualization: Data and Design Process



Interesting questions:

Why have a human in the decision-making loop?

Why have a computer in the loop?

Why use an external representation?

Why depend on vision?

Why show the data in detail?

Why use interactivity?

What is the design space of visualization idioms?

Why focus on tasks?

Why are most designs ineffective?

What resource limitations matter?

How can better be measured? (Munzner, 2014, chap. 1)

Why and how to represent data visually?

- The human visual system is a most powerful pattern seeker

“seeing is understanding...”

- We easily see patterns displayed in certain ways

but not in others ...



An exercise in preattentive processing: how many “3”?

69704259347493
58728294954642
44396854634235
6658789376

(Nussbaumer Knaflic, 2015)

What if the task is counting the “7”?

69704259**3**47493
58728294954642
44**3**968546**3**42**3**5
6658789**3**76

C. Nussbaumer Knaflic, Storytelling with Data ,Talks at Google, 2015
[Storytelling with Data | Cole Nussbaumer Knaflic | Talks at Google \(youtube.com\)](#)

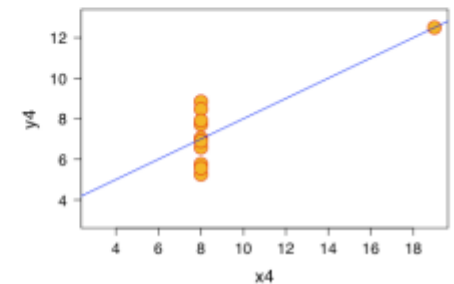
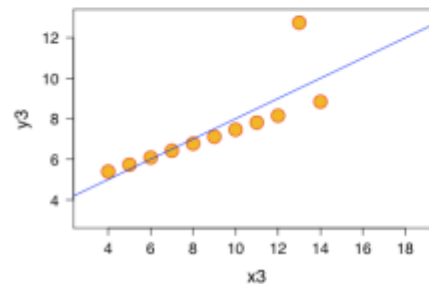
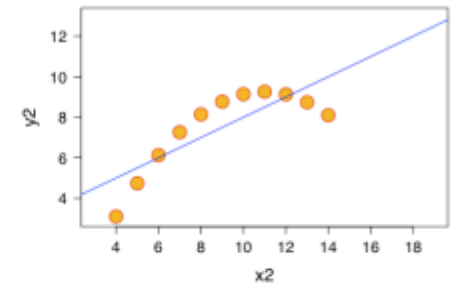
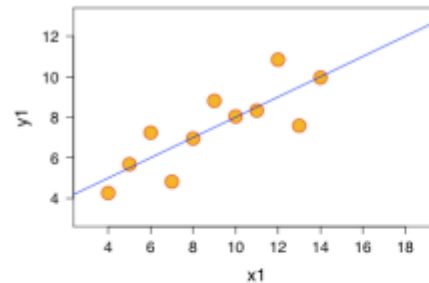
Why show the data in detail? and not only aggregated parameters

Visualization helps in situations where seeing the dataset structure in detail is better than seeing only a brief summary of it (losing information). (Munzner, 2014)

Ascombe quartet: data sets with same simple statistical properties (Tufté, 1983)

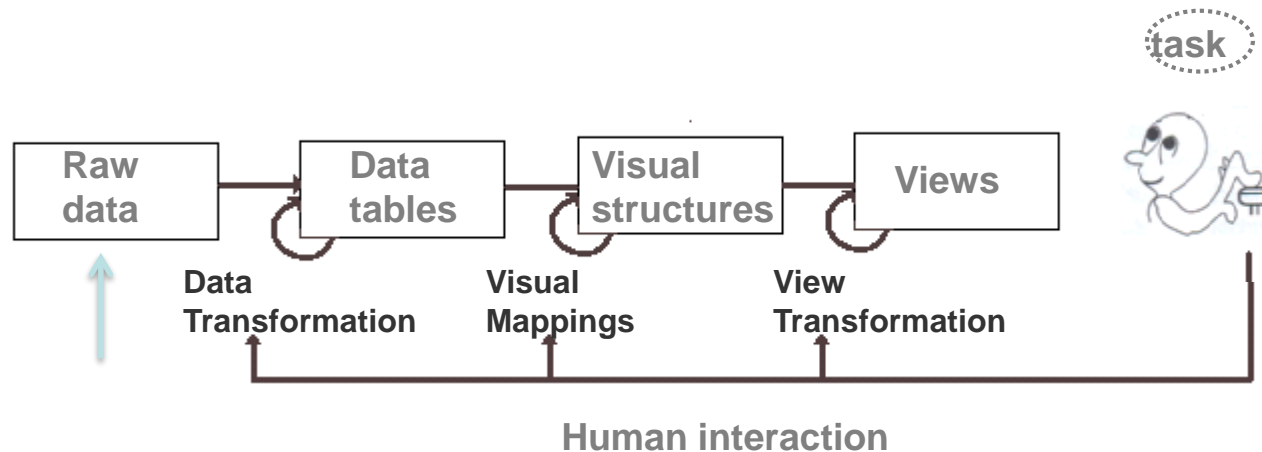
Anscombe's Quartet: Raw Data

	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	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
var.	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
corr.		0.816		0.816		0.816		0.816



Information Visualization Reference Model

This course focus on Information Visualization

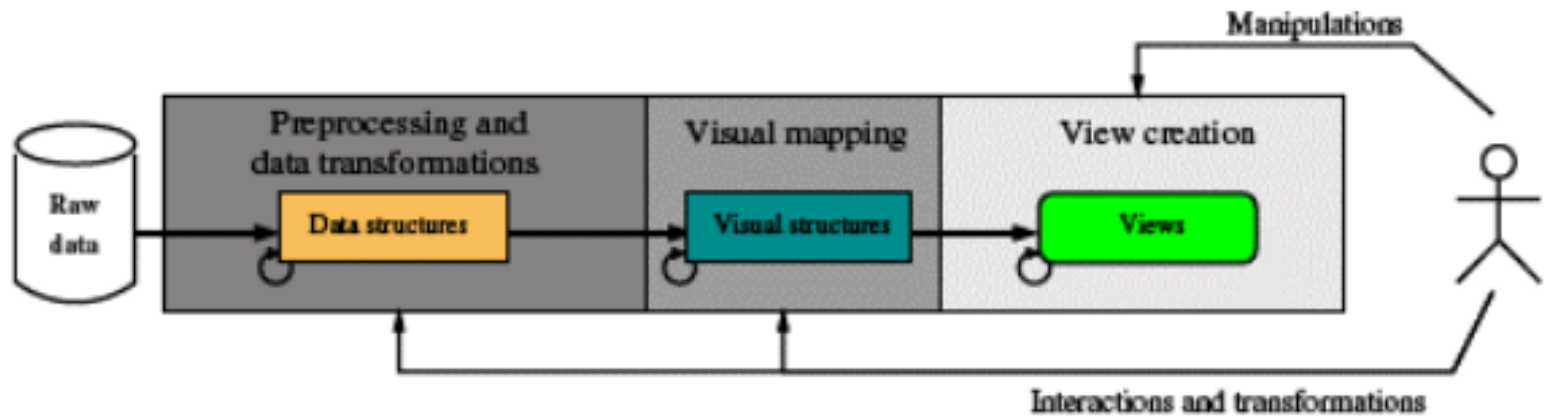


Visualization can be described as the **mapping of data to visual** form **supporting human** interaction for visual sense making (Card et al., 1999)

Visualization is a **Human in the loop process!** -> which calls for specific **methods**

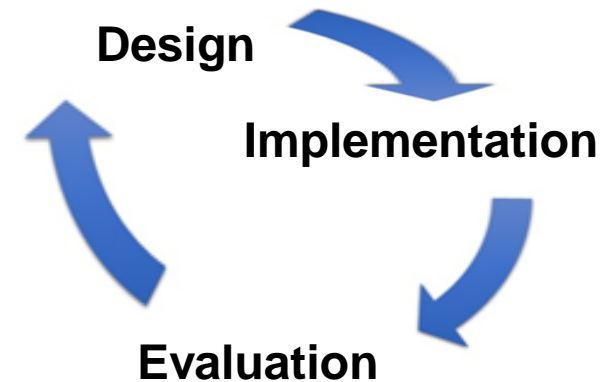


The process of creating a Visualization



How can we produce a Visualization?

- There are **many possible** Visualization solutions, but how to select the more adequate to a specific case?



- To obtain **efficacy** it is fundamental:
 - a correct definition of **goal** and **user tasks**
 - apply adequate methods and **evaluate**in **several iterations** until the goals are ~satisfied ...
- I.e. using a **human-centered approach**

Framework for analyzing Visualization use

- It is a “wicked problem”

<https://www.interaction-design.org/literature/article/wicked-problems-5-steps-to-help-you-tackle-wicked-problems-by-combining-systems-thinking-with-agile-methodology>

Visualization usage can be analyzed in terms of:

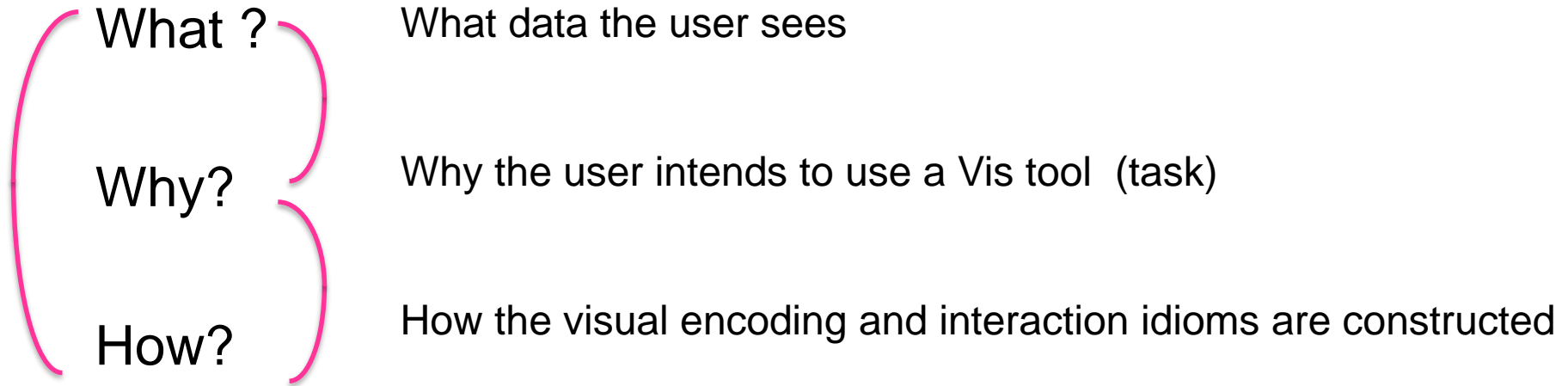
- Why the user needs it
- How to design the visualization
- What data is shown



Visualization Analysis & Design

(Munzner, 2014)

Framework for analyzing Visualization use

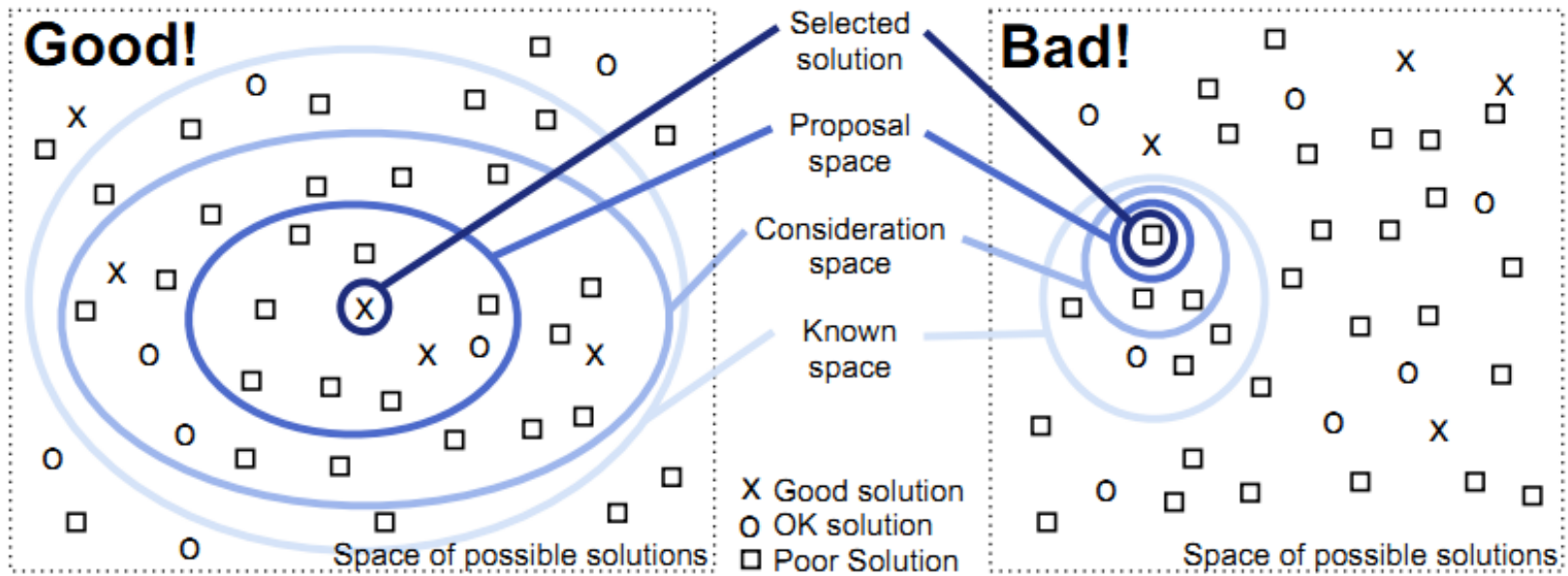


(Munzner, 2014)

Simple Vis tools may be analyzed as an instance;

Complex tools may require analysis in terms of a sequence of instances

The problem of Visualization design



(Munzner, 2014)

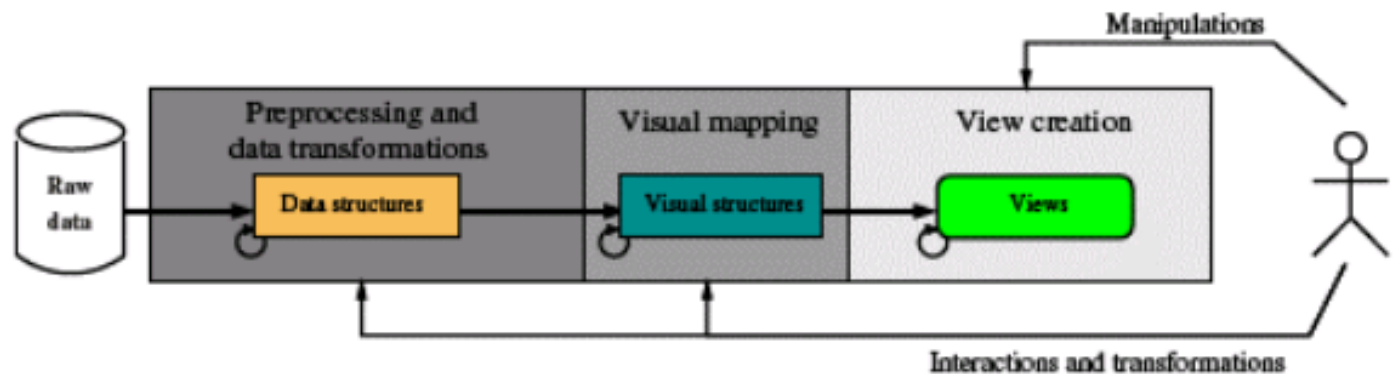
- Only a very small number of possibilities are reasonable ...
most are ineffective

Consider multiple alternatives that seem interesting and then select the best!

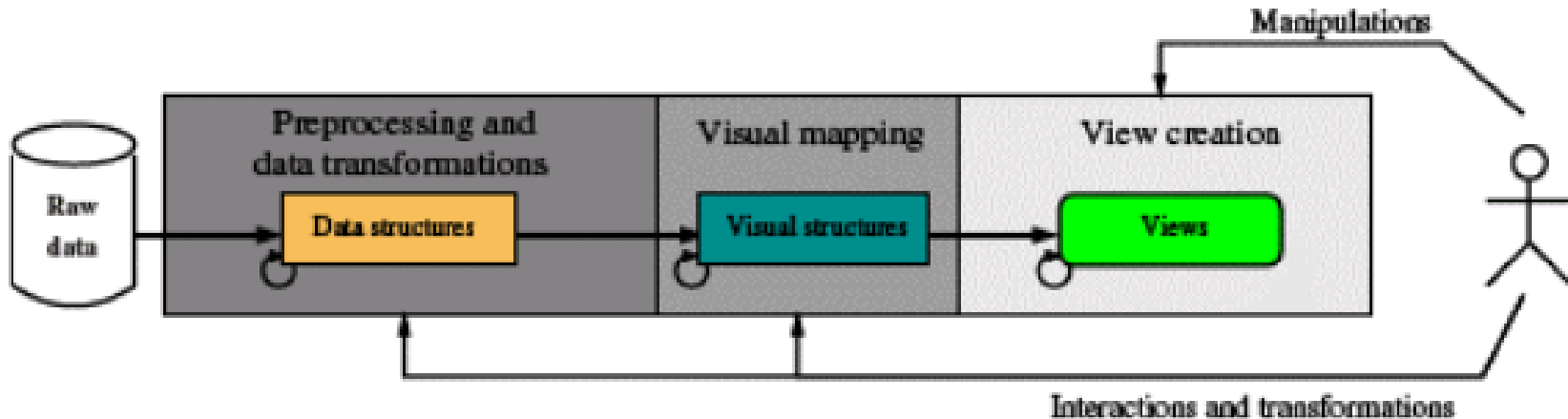
(based on evaluation ...)

Creating Visualizations

- **Good design and evaluation** is the key to success in producing a Visualization
- Visualization S/W can provide many visual templates;
- In spite of variation, all S/W packages follow the same generation process



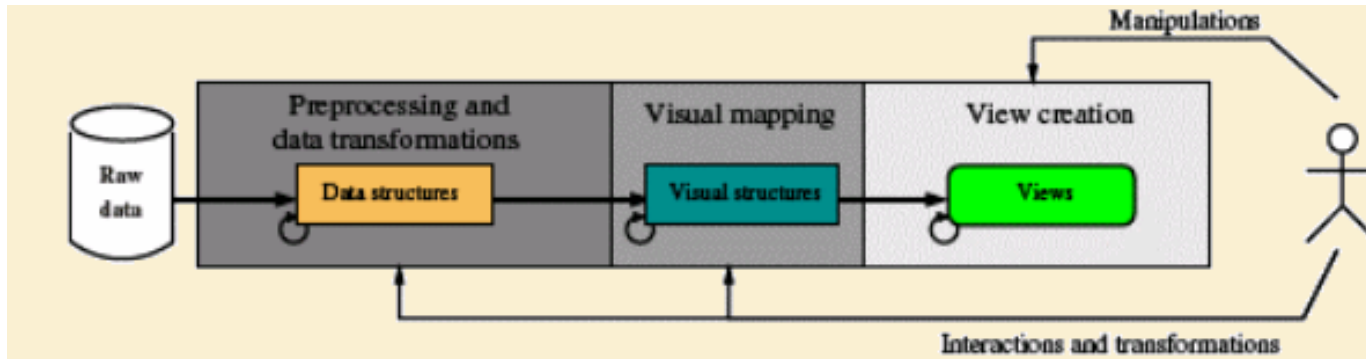
(Mazza, 2009)



- Creating a visualization can be modeled as a process including several stages:
 - preprocessing and transformation
 - visual mapping **not forgetting evaluation in several iterations!**
 - creation of views

(Mazza, 2009)

Note: this is similar to the visualization reference model

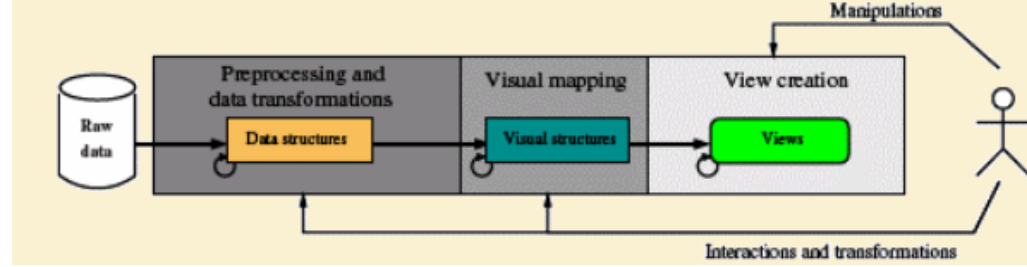


1. Preprocessing (Data transformation):

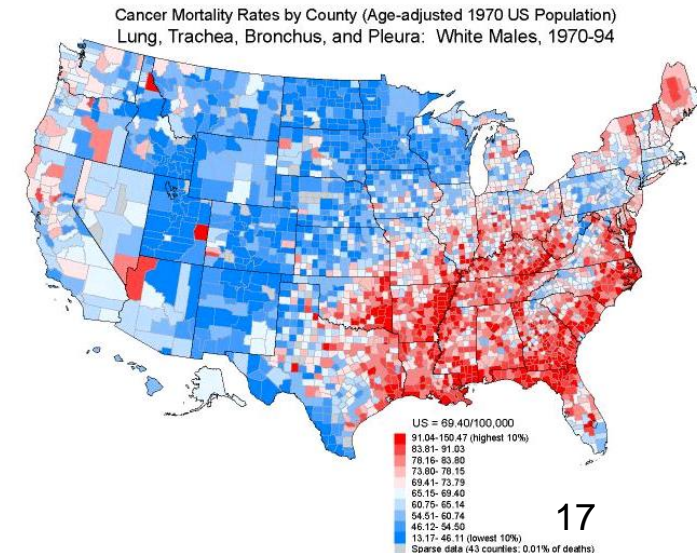
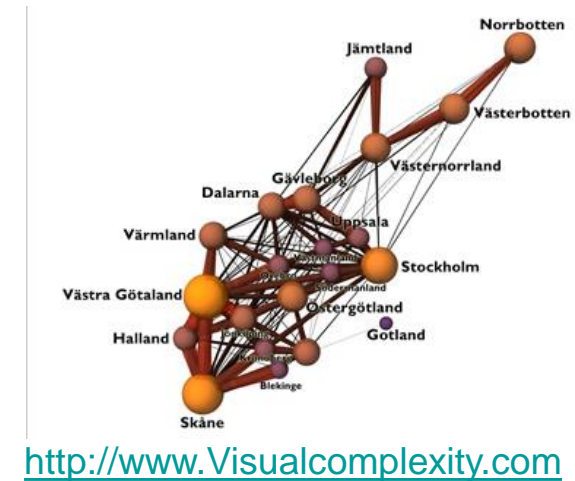
- Abstract data (which don't have a specific connection with physical space) are rarely in a suitable format for automatic treatment and visualization
- Raw data (data supplied by the world around us, a.k.a. data sets) have to be given an organized logical structure to be processed by the Visualization S/W

This is a fundamental step; it is assumed to be addressed in another course!

2. Visual mapping:



- It is necessary to decide:
 - which visual structures to use to **represent** the data
 - their location in the display
- Some types of abstract data can be easily mapped to a spatial location
- Examples:
 - . data with a topological or geographical structure
- Many types of data don't have an easy correspondence with the dimensions of the physical space around us



Three **structures** must be defined in the **visual mapping**:

- Spatial substrate
- graphical elements
- graphical properties

- **Spatial substrate** - dimensions in physical space where the visual representation is created (can be defined in terms of axes and type of data)
- **Graphical elements** - anything visible appearing in the space
points, lines, surfaces, volumes
- **Graphical properties** – properties of the graphical elements to which the human retina is very sensitive - **retinal variables**:
size, orientation, color, texture, and shape

- **Spatial substrate** axes (x, y, ...)
type of data (quantitative, ordinal, categorical)

- **Graphical elements** points
lines
surfaces
volumes

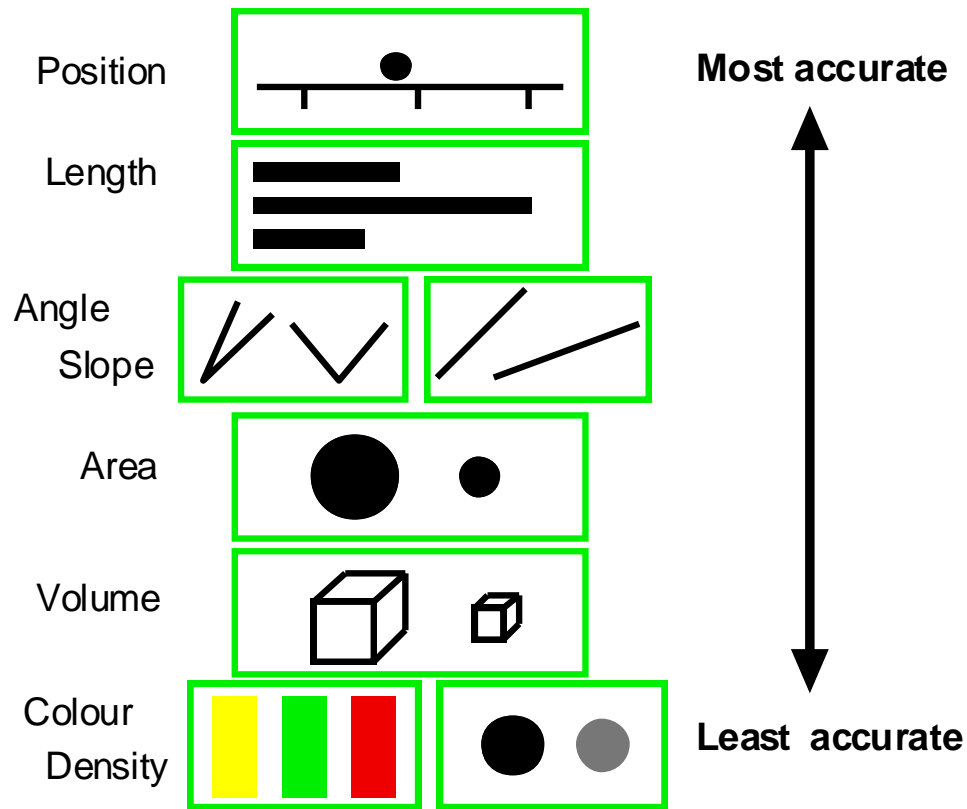
- **Graphical properties** retinal variables:
size,
orientation
color (depends on physiology and culture)
texture
shape

	Association The marks can be perceived as SIMILAR	Selection The marks are perceived as DIFFERENT, forming families	Order The marks are perceived as ORDERED	Quantity The marks are perceived as PROPORTIONAL to each other
Size				
Value				
Texture				
Colour				
Orientation				
Shape				

Bertin's guidance regarding the suitability of various visual **encoding methods to support common tasks**

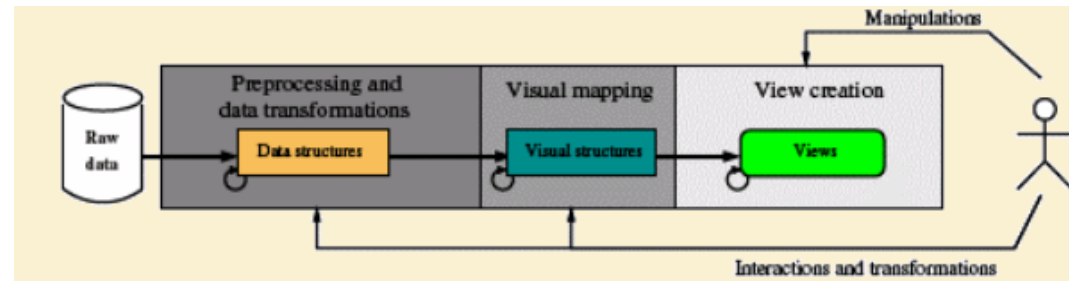
(Spence, 2007)

Note that **only size is adequate to represent quantity accurately**



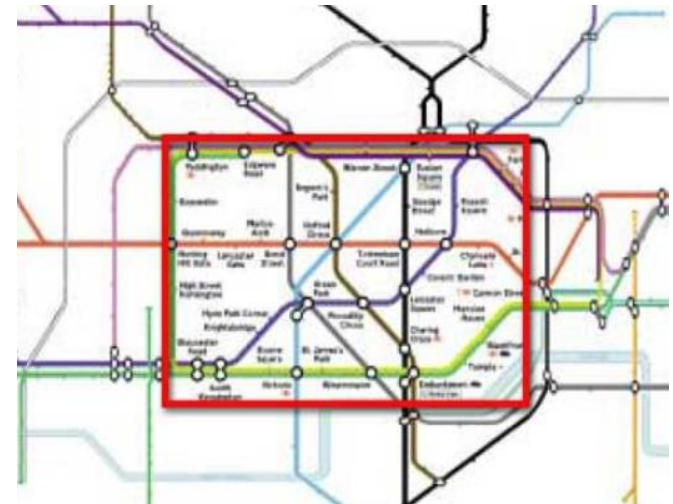
The relative difficulty of **assessing quantitative value** as a function of encoding method, as established by Cleveland and McGill (Spence, 2007)

3. Creation of views:



- Views are the final result of the generation process
- Producing them corresponds to the computer graphics phase:
- Often the quantity of data to represent is too large for the available space
- To overcome this problem there are **presentation and interaction** techniques as:

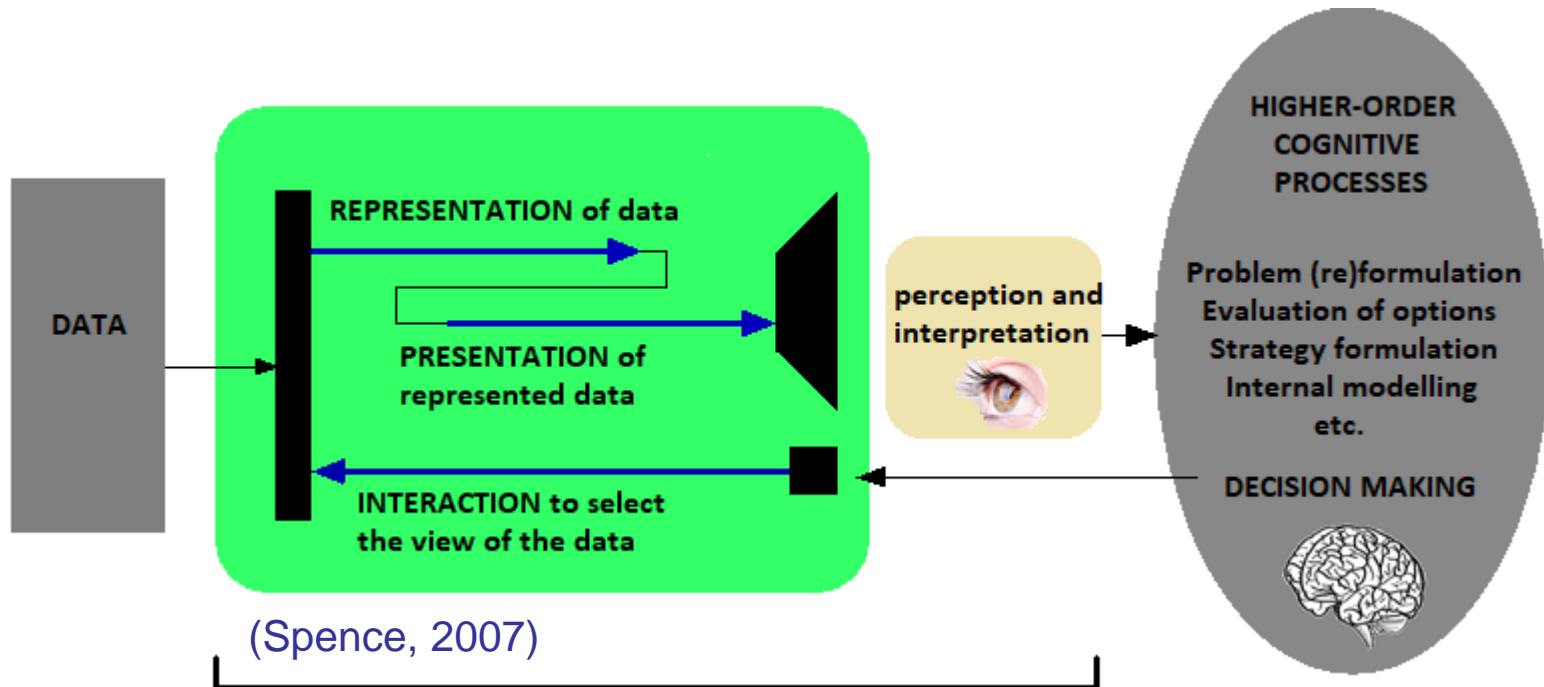
- Zooming
- Panning
- Scrolling
- Focus + context
- Magic lenses
- ...



<https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/bifocal-display>

The process of visualization

Let us increase the known solution space organizing the methods!



Interaction with data governed by high-order cognitive processes involves:

- Representation
- Presentation
- Interaction

- Data may have a lot of different forms and there are many techniques and systems to visualize them

- A data classification is important to:
 - predict what visualization techniques are adequate
 - make easier the communication about the data
 - allow a more systematic approach to Visualization
 -

Data Abstraction

name	rank	gender	year
Jacob	1	boy	2010
Isabella	1	girl	2010
Ethan	2	boy	2010
Sophia	2	girl	2010
Michael	3	boy	2010

Item/object

Attribute/variable

- Four basic dataset types:

InfoVis

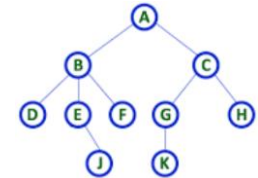
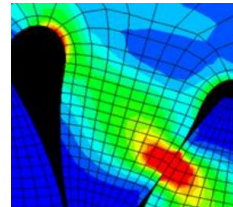
- Tables

- Networks

SciVis

- Fields (continuous)

- Geometry (spatial)



- Five basic datatypes

- Items

- Attributes

- Links

- Positions

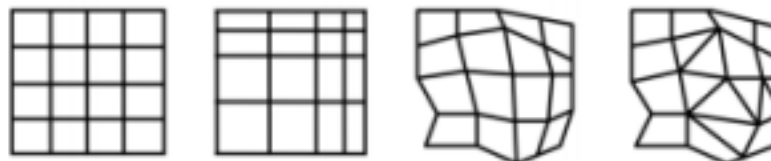
- Grids

Categorical

Ordered

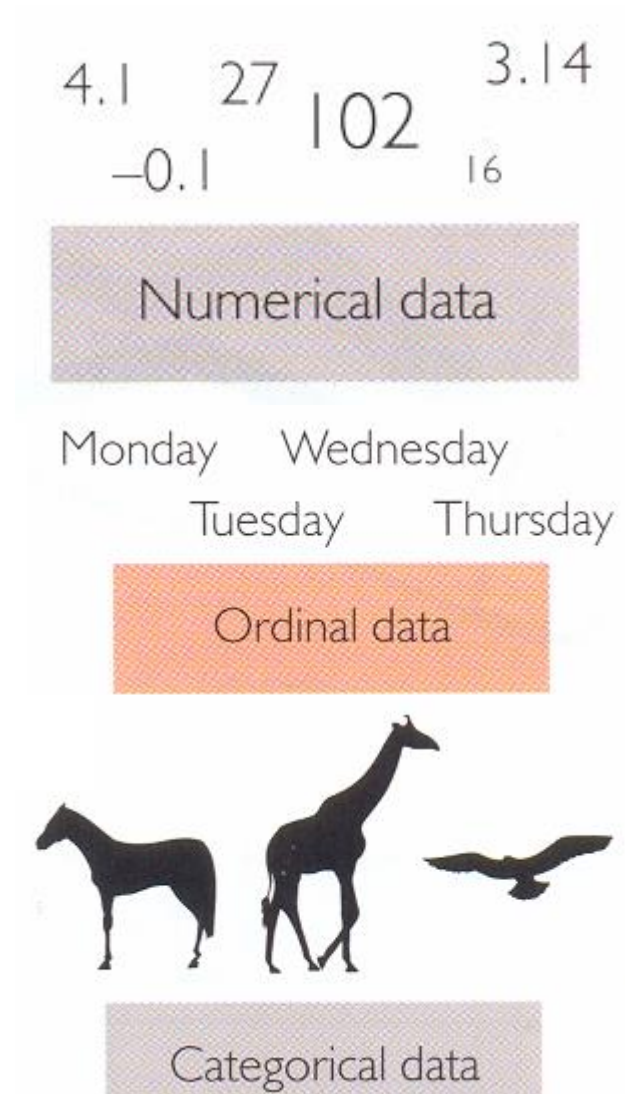
Ordinal

Quantitative



(Munzner, 2014)

- **Data representation level:**
 - Qualitative (or categorical)
 - Quantitative (or numeric)
- **Data nature:**
 - Continuous
 - Discrete
- **Measuring scale:**
 - Nominal
 - Ordinal
 - Interval
 - Ratio



(Spence, 2007)

- Examples of measuring scales and types of data:
 - **nominal** --> car brands, gender, animal species...
 - **ordinal** --> week days, preferences, levels measured in a Likert-type scale
 - **Interval** --> date, IQ, temperatures in °C
 - **Ratio** --> temperatures in °K, weight, height
- The ratio scale represents the **highest level of representation**, has a absolute, non-arbitrary zero (unlike the interval scale)
- This is a general classification and might be used to select the statistical methods to use with the data

Model, structure and format of the data to Visualize

- Consider a data set with three columns:

latitude *longitude* *d*

- Which is the most adequate way to visualize these data?
- If *d* is depth, the selected visualization technique may involve interpolation (ex: contours)
- If data represent location and the number victims of traffic accidents, interpolation and contours do not make sense

Know the data structure is not enough

It is necessary to know the phenomenon behind the data!

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

latitude *longitude* *d*

- Which is the most adequate way to visualize these data?
- If *d* is depth or altitude?

the selected visualization
technique may involve
interpolation

(e.g. isocontours,
isosurfaces, 3D surface)

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

latitude

longitude

d

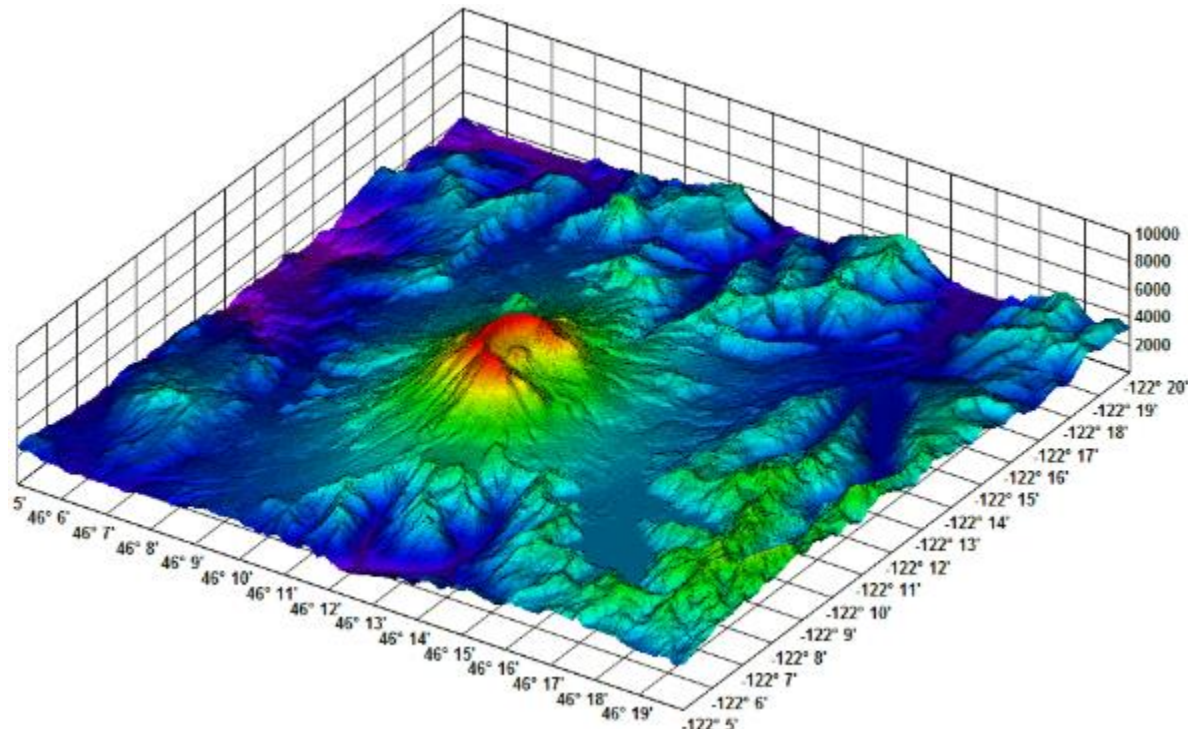


- Which is the most adequate way to visualize these data?

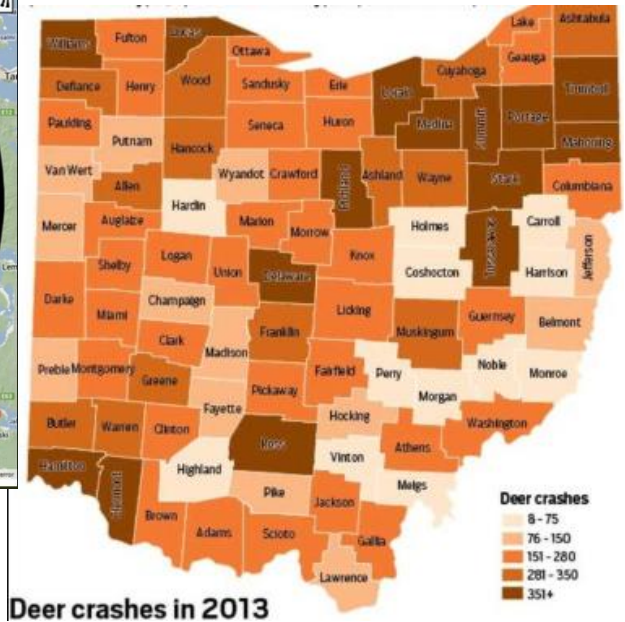
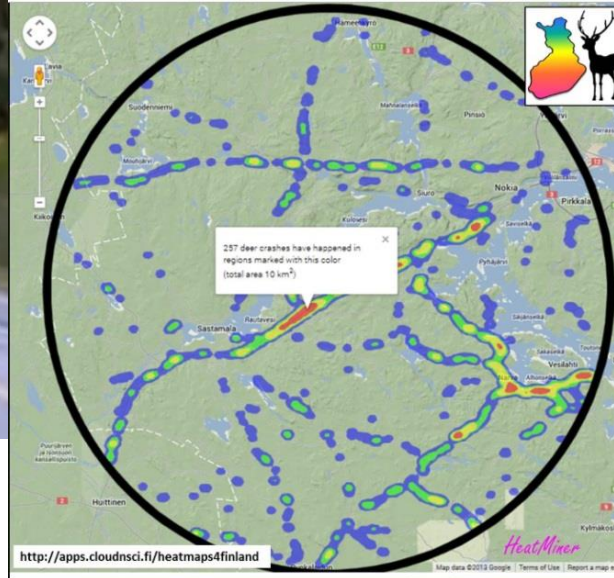
- If *d* is depth or altitude?

the selected visualization technique may involve interpolation

(e.g. isocontours, isosurfaces, 3D surface)



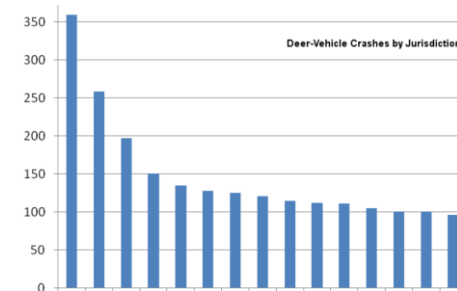
- What if the data represent location and the number of “deer crashes”?



- Interpolation and contours don't make sense!

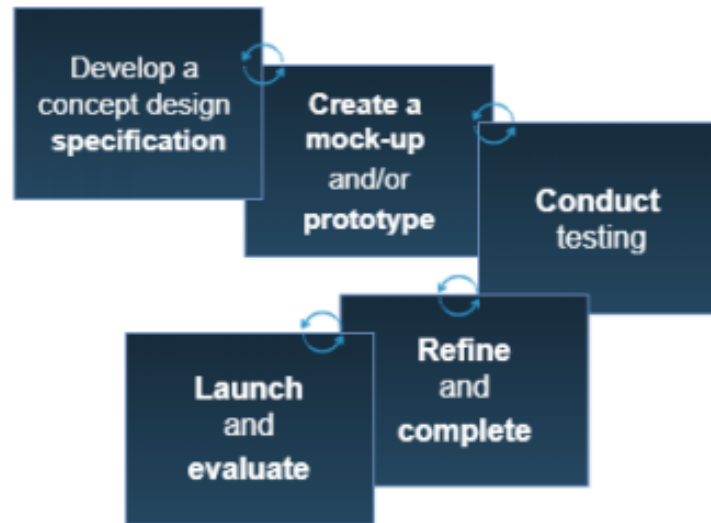
Know the data structure is not enough

It is necessary to **know the phenomenon behind the data and the users' tasks**



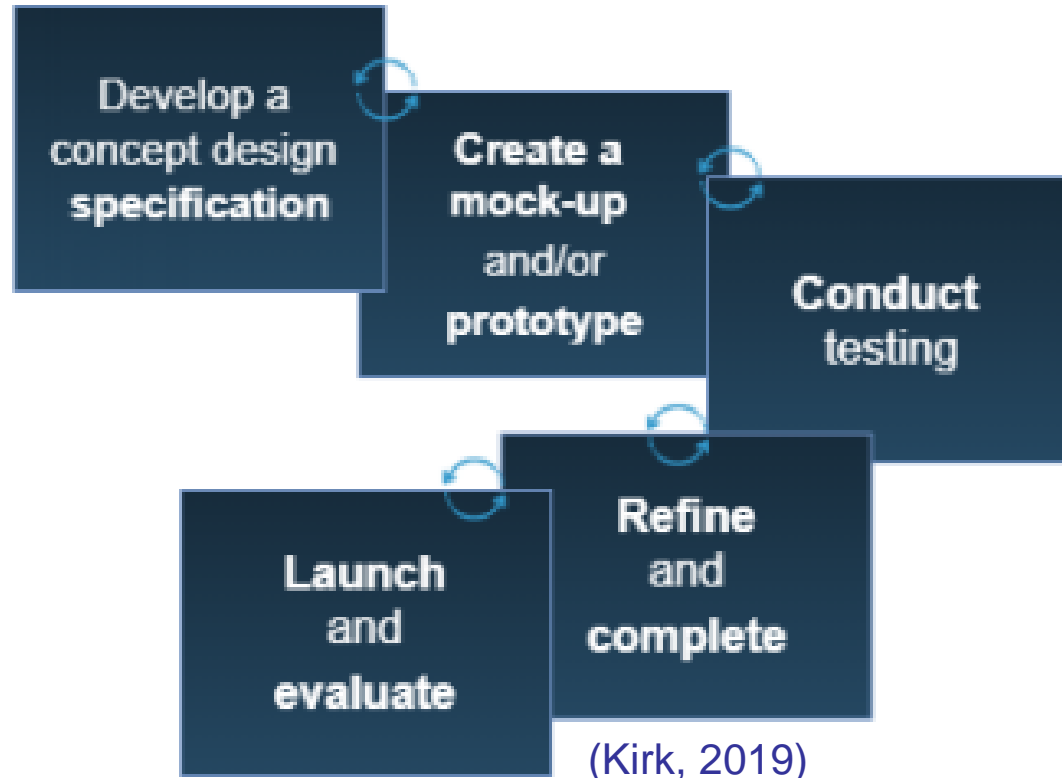
Developing a Visual data exploration application

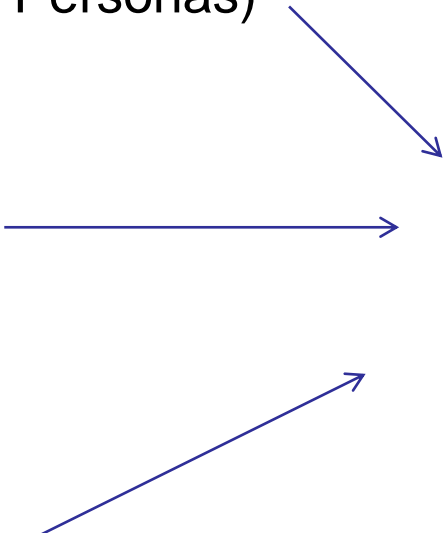
a very brief introduction



Phases of the development cycle of a visual data exploration application

after understanding the users and their questions to elicit requirements:



- There are several methods that can be used to help requirement elicitation of an interactive application, e.g.
 - User models (such as Personas)
 - Scenarios
 - Task analysis
 - Requirements
 - Low fidelity prototype
- Typical in user-centered design approaches (UCD)
- To apply in the Visualization mini-project
- 
- ```
graph LR; A[User models (such as Personas)] --> B[Typical in user-centered design approaches (UCD)]; C[Scenarios] --> B; D[Task analysis] --> B; E[Requirements] --> B;
```

# Personas

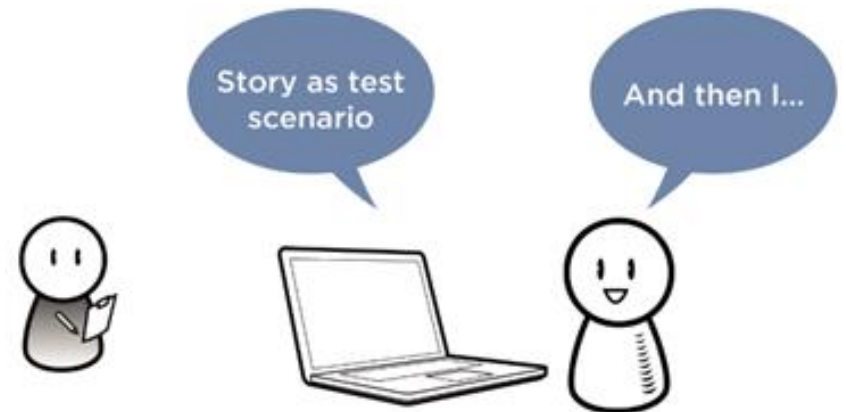
- Personas are fictional characters based on user research to help understand:
  - users' needs,
  - experiences,
  - behaviors
  - goals.
- Make the design task at hand less complex
- Guide the ideation processes, and help to achieve the goal of creating a good user experience for the target user group

<https://www.interaction-design.org/literature/article/personas-why-and-how-you-should-use-them>

# Scenarios

- Stories and contexts about how the user groups use a future product/service
- Note the goals and questions to be achieved and sometimes define the possibilities of how the user(s) can achieve them on the application
- Scenarios are critical for
  - designing
  - UX testing

(Note: Different from User stories and Use cases)

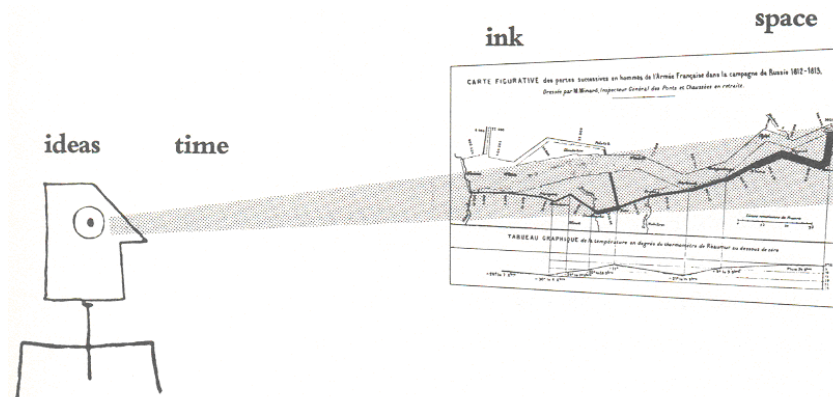


<https://www.usability.gov/how-to-and-tools/methods/scenarios.html>

<https://www.interaction-design.org/literature/topics/user-scenarios>



# Effective Visualization

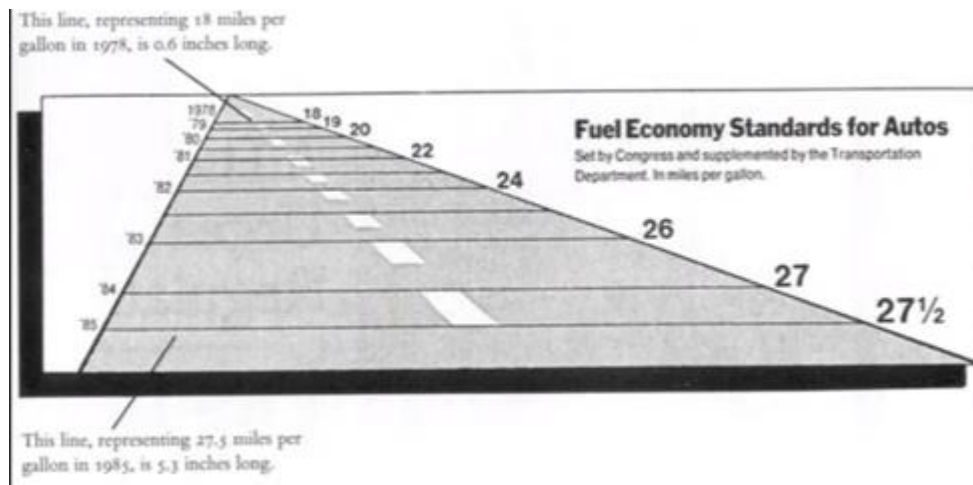
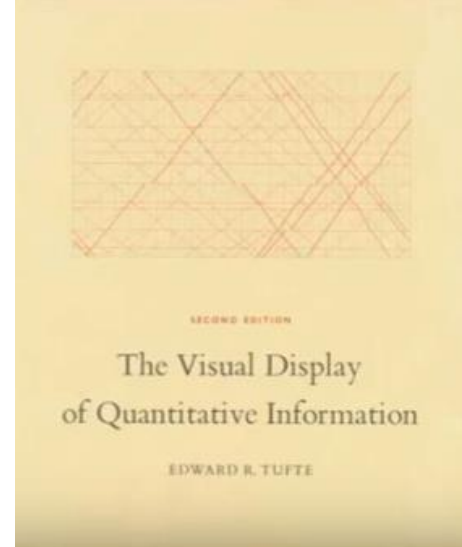


# Effective visualization

Implies saying the **truth** about the data

Tufte presents a lot of commented examples in his book:

Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983



**There are methods to evaluate visualizations that should be used along the process of creating a visualization**

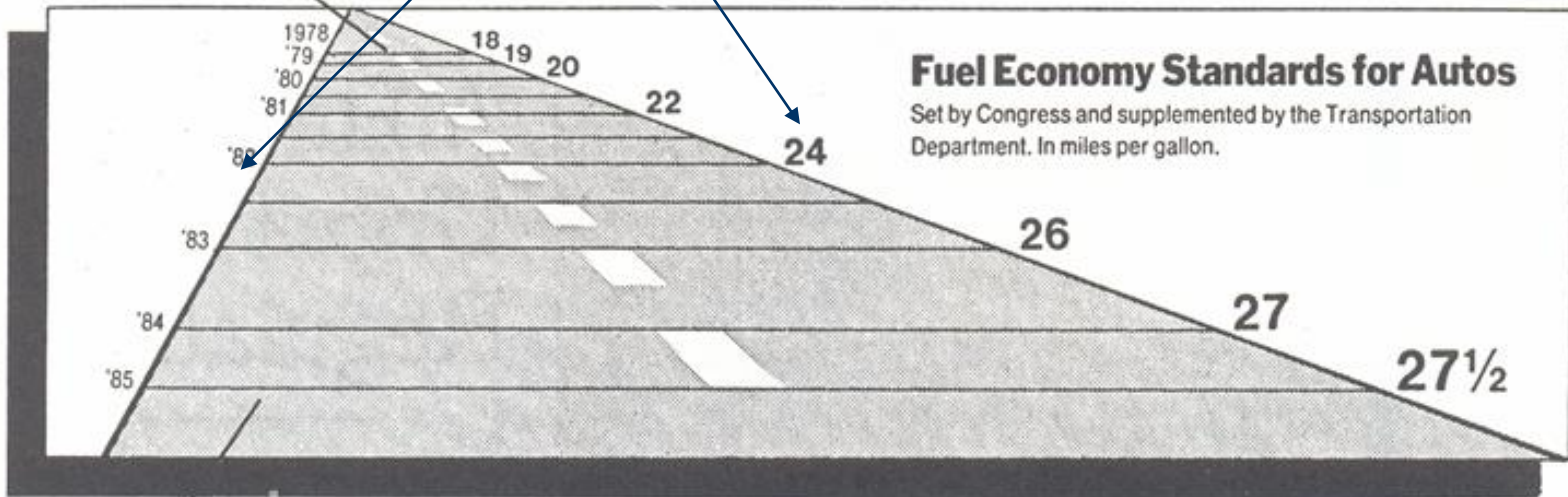
[https://infovis-wiki.net/wiki/Lie\\_Factor](https://infovis-wiki.net/wiki/Lie_Factor)

this example has several problems:

Lie Factor = 14.8

Legends have a constant size in one side and variable in the other

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

In roads, future usually lies in front, not behind

- Perception varies with

- context

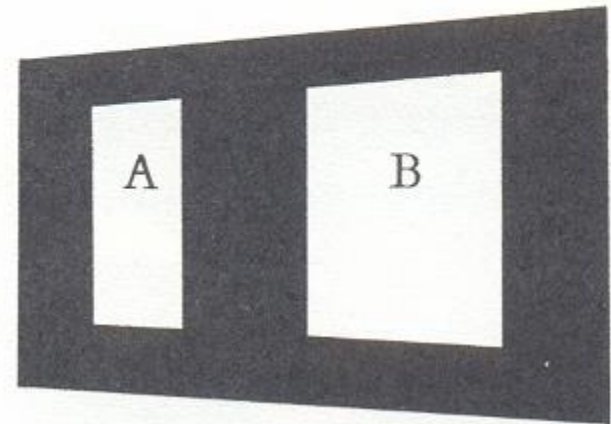
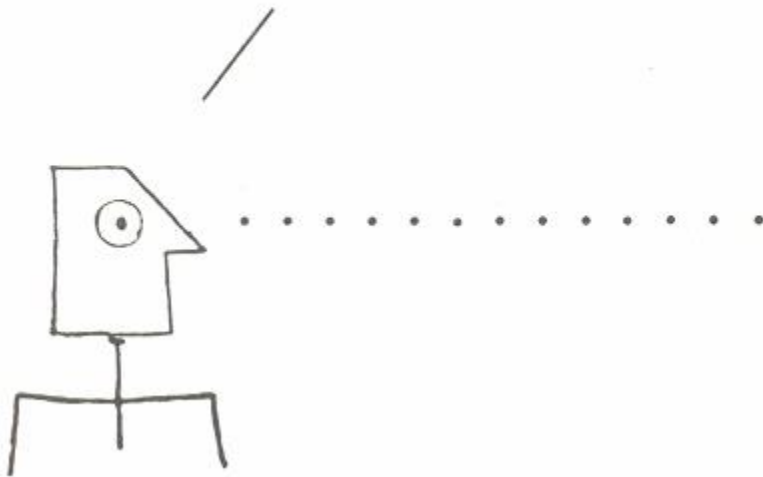


(Tufte, 1983)



- How do we know that the visual image represents the underlying numbers?
- One way to try to answer these questions is to conduct experiments on the visual perception of graphics

**I think I see that area B  
is 3.14 times bigger than  
area A. Is that correct?**



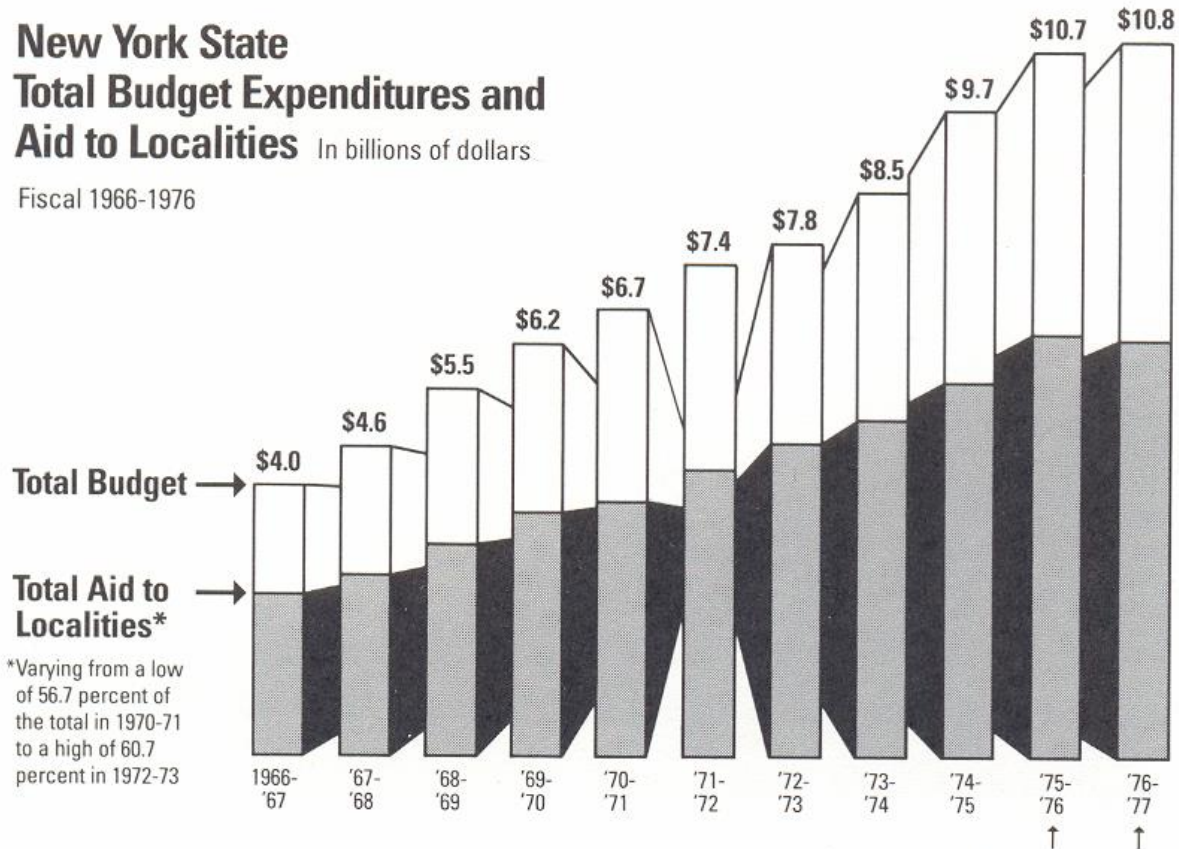
*(Tufte, 1983)*

Another example:

## New York State Total Budget Expenditures and Aid to Localities

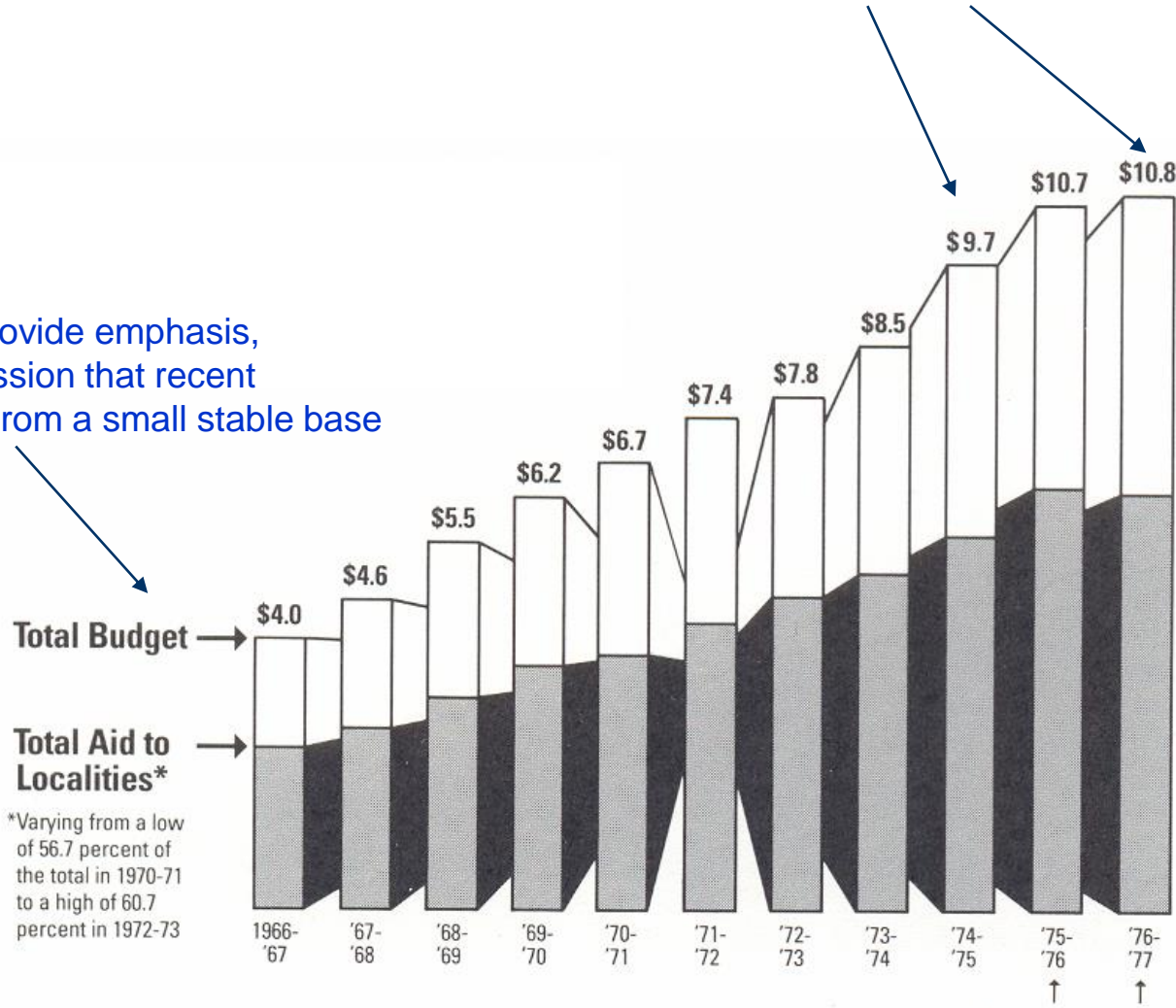
In billions of dollars

Fiscal 1966-1976



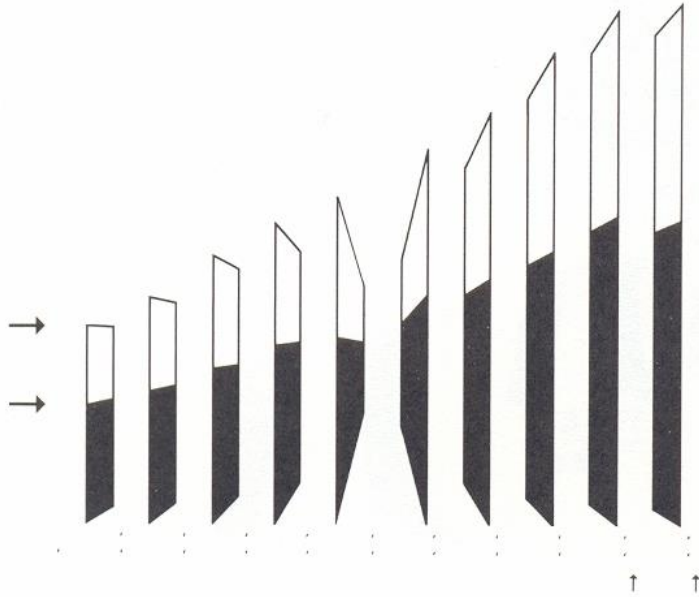
These three parallelipeds have been placed in na optical plane in front of the other eight, creating the image that the newer budgets tower over the older ones

Horizontal arrows provide emphasis, Encourage the impression that recent years have shot up from a small stable base

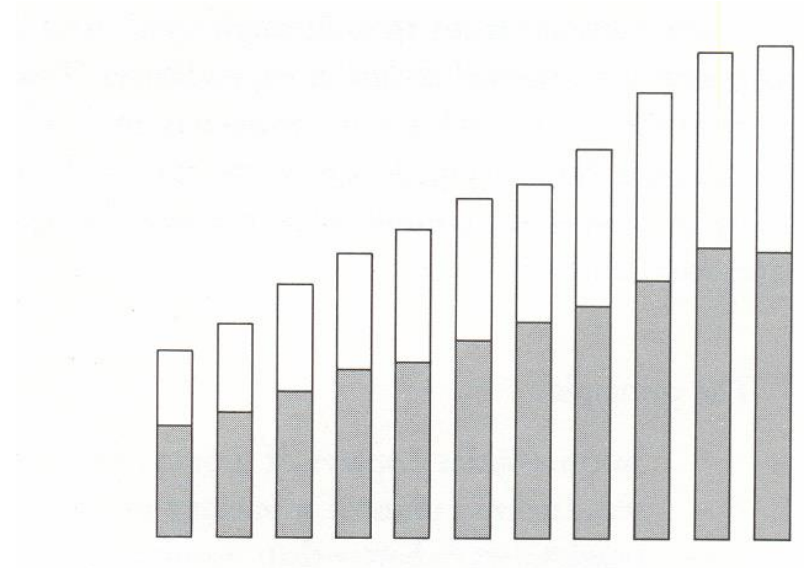


Arrows pointing straight up emphasize recent growth

# Leaving behind the distortion



# we have a calmer view:



Two statistical lapses also bias the chart :

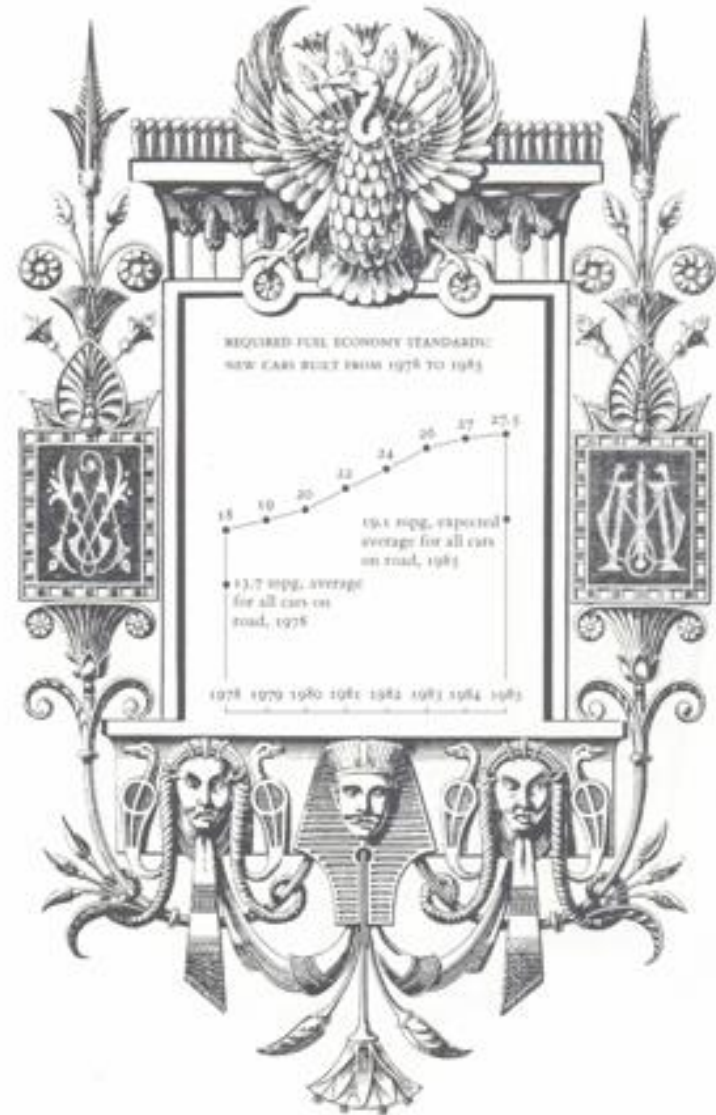
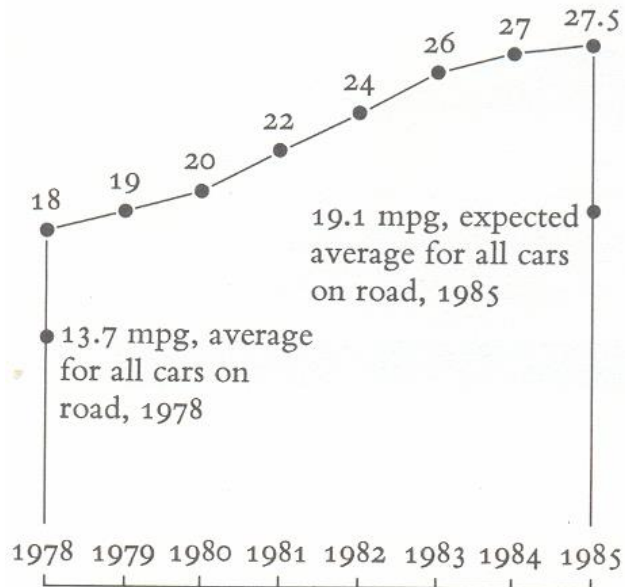
- Population increased 10%
- there was substantial inflation

Final result ➔

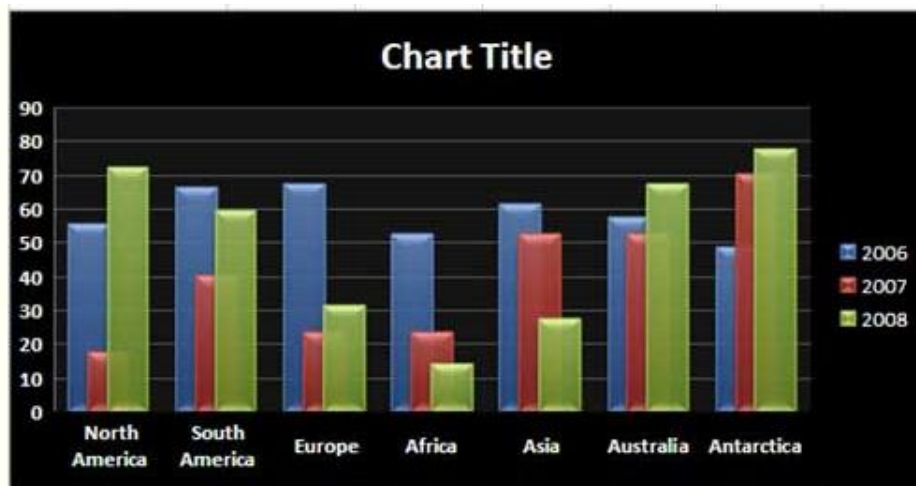
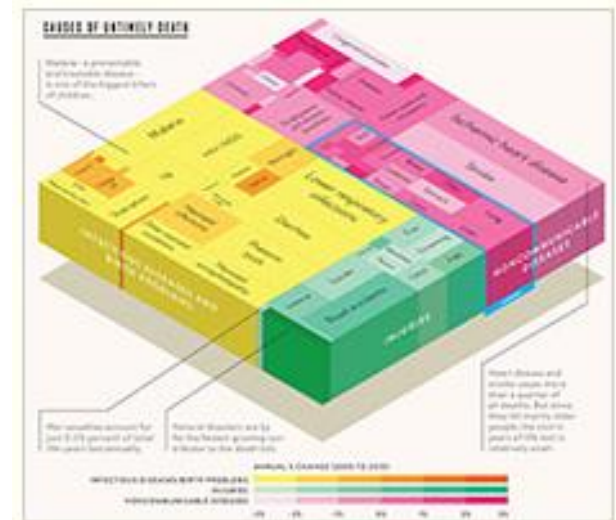
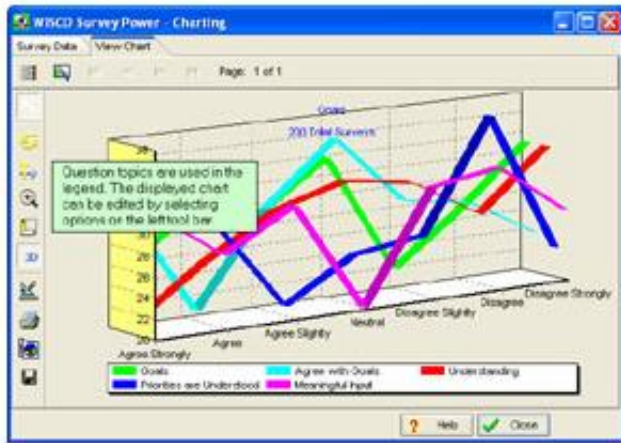


# Decorations without lies:

REQUIRED FUEL ECONOMY STANDARDS:  
NEW CARS BUILT FROM 1978 TO 1985



# Effective visualization: more poor examples analyzed



**There are methods to evaluate visualizations that should be used along the process of creating a visualization**

## Main bibliography

- Spence, R., *Information Visualization, An Introduction*, Springer, 2014
- Munzner, T., *Visualization Analysis and Design \**, A K Peters, 2014
- Kirk, A., *Data Visualisation- A Handbook for Data Driven Design*, 2nd. Ed., Sage, 2019
- Fisher, D., Meyer, M., *Making Data Visual. A practical Guide to using Visualization for insight \**, O'Reilly, 2017
- Ware, C., *Information Visualization, Perception to Design \**, 3rd ed.,Morgan Kaufmann, 2012
- Alberto Cairo, *How Charts Lie: Getting Smarter about Visual Information*, W. W. Norton & Company; 1st edition, 2020
- Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983
- Explore books marked \* and other books at the playlist:  
<https://learning.oreilly.com/playlists/74bfec5e-4346-48ff-82b4-657fda6922b6>