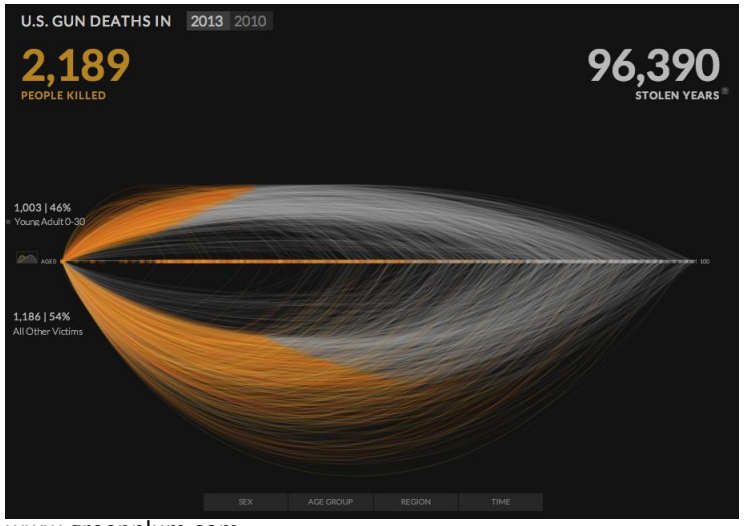
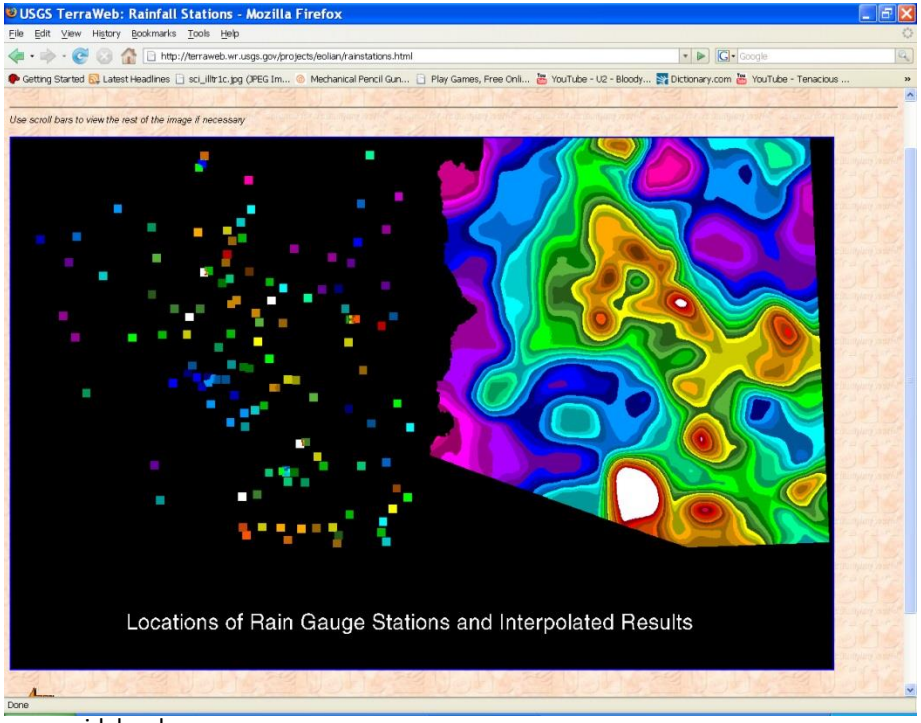


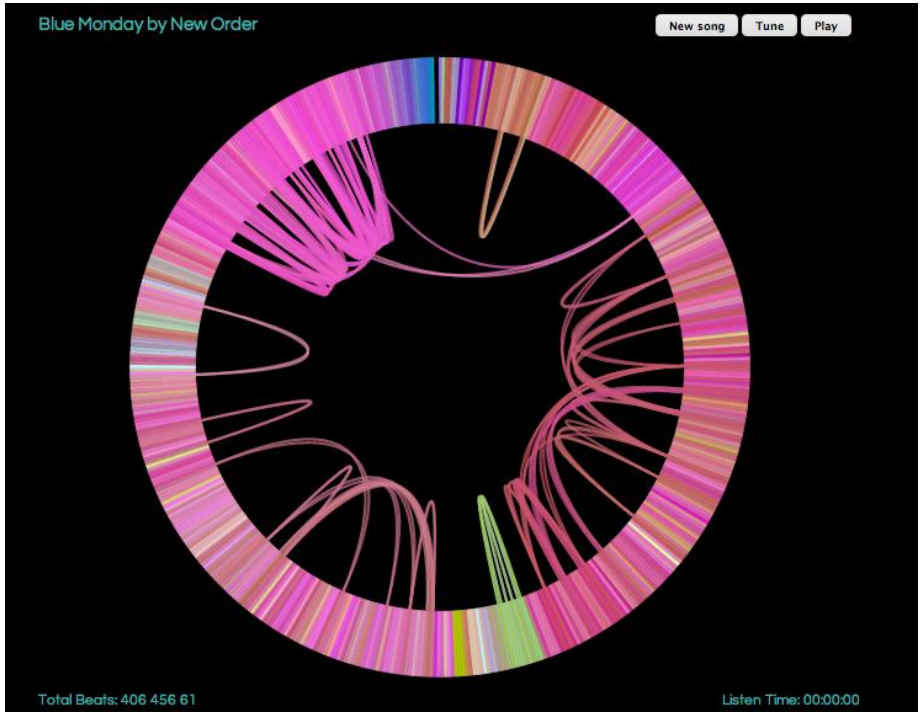
www.greenplum.com

www.cmu.edu

8. Designing efficient visualizations



www.idvbook.com



www.business2community.com

Efficient visualization

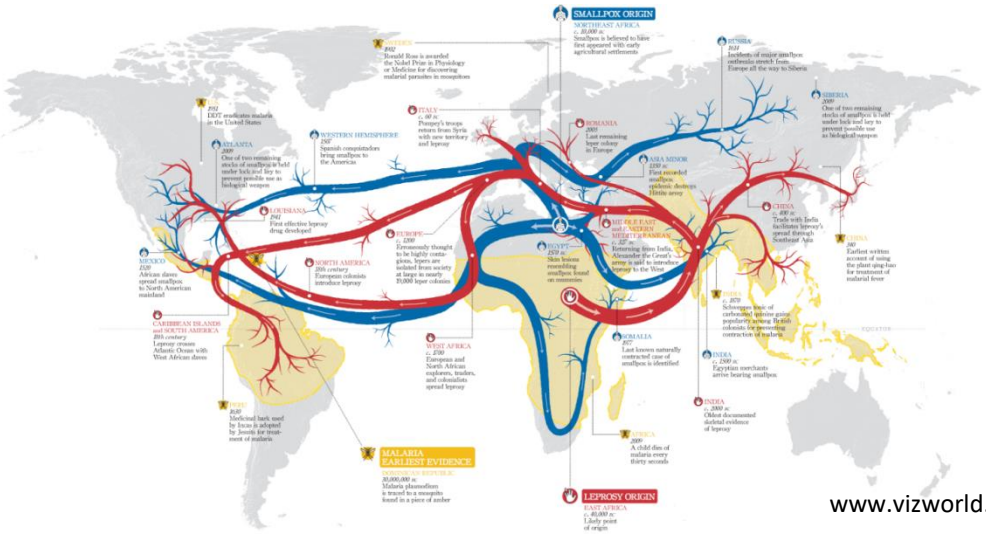
- Successful visualization = efficient and correctly conveys the desired information to the audience + fulfills the initial requirements (exploration, confirmation of hypothesis, presentation, …)
- We can choose from unlimited number of options for
  - Mapping the data to graphic entities and their attributes
  - Interactive manipulation with the visualization
- Searching for the best combination is not an easy task…

Non-efficient visualization

- Too complex and confusing for the end user
- Undesirable deformation, clipping or data loss within the mapping process
- Insufficient support for interaction – changes in views, changing the color palette, …
- Poor aesthetics

Efficient visualization

- It is much easier to design a bad visualization than the good one ☺
- Our goal is to show
  - possible ways to design efficient visualizations
  - common problems



www.vizworld.com

A PERIODIC TABLE OF VISUALIZATION METHODS

Data Visualization
Visual representations of quantitative data in schematic form (either with or without axes)

Information Visualization
The use of interactive visual representations of data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space. The image can be changed by users as they proceed working with it

Concept Visualization
Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses.

Strategy Visualization
The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.

Metaphor Visualization
Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed

Compound Visualization
The complementary use of different graphic representation formats in one single schema or frame

C continuum																G graphic facilitation
Tb table	Ca cartesian coordinates											Me meeting trace	Mm metro map	Tm temple	St story template	Tr tree	Ct cartoon
Pi pie chart	L line chart											Co communication diagram	Fp flight plan	Cs concept sceleton	Br bridge	Fu funnel	Ri rich picture
B bar chart	Ac area chart	R radar chart cobweb	Pa parallel coordinates	Hy hyperbolic tree	Cy cycle diagram	T timeline	Ve venn diagram	Mi mindmap	Sq square of oppositions	Cc concentric circles	Ar argument slide	Sw swim lane diagram	Gc gantt chart	Pm perspectives diagram	D dilemma diagram	Pr parameter ruler	Kn knowledge map
Hi histogram	Sc scatterplot	Sa sankey diagram	In information lense	E entity relationship diagram	Pt petri net	Fl flow chart	Cl clustering	Lc layer chart	Py minto pyramid technique	Ce cause-effect chains	Tl toulmin map	Dt decision tree	Cp cpm critical path method	Cf concept fan	Co concept map	Ic iceberg	Lm learning map
Tk tukey box plot	Sp spectogram	Da data map	Tp treemap	Cn cone tree	Sy system dyn./ simulation	Df data flow diagram	Se semantic network	So soft system modeling	Sn synergy map	Fo force field diagram	Ib ibis argumentation map	Pr process event chains	Pe pert chart	Ev evocative knowledge map	V Vee diagram	Hh heaven 'n' hell chart	I infomural

Cy **Process Visualization**

Hy **Structure Visualization**

☼ **Overview**

▫ **Detail**

◎ **Detail AND Overview**

< > **Divergent thinking**

> < **Convergent thinking**

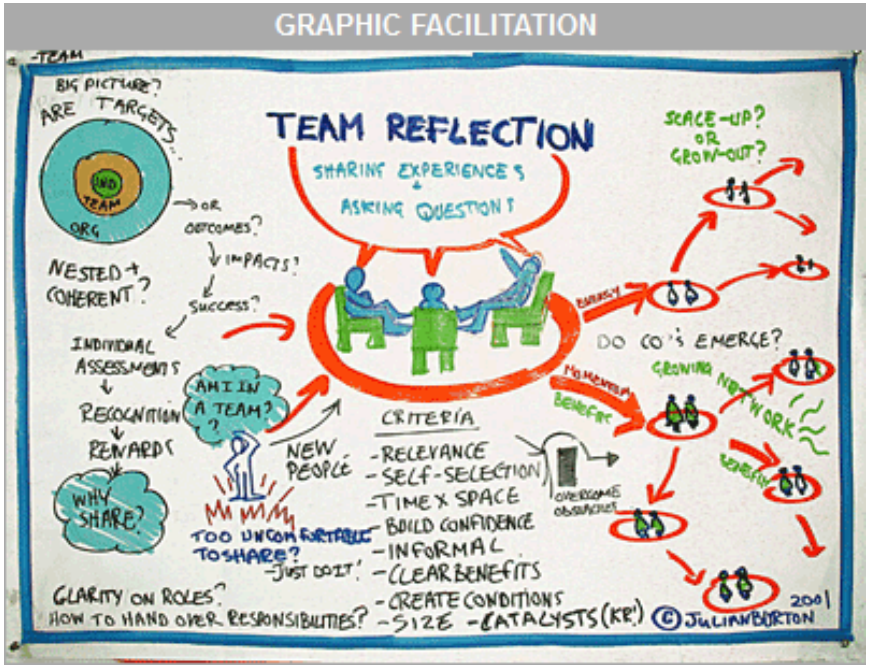
Su supply demand curve	Pc performance charting	St strategy map	Oc organisation chart	Ho house of quality	Fd feedback diagram	Ft failure tree	Mq magic quadrant	Ld life-cycle diagram	Po porter's five forces	S s-cycle	Sm stakeholder map	Is ishikawa diagram	Tc technology roadmap
Ed edgeworth box	Pf portfolio diagram	Sg strategic game board	Mz mintzberg's organigraph	Z zwicky's morphological box	Ad affinity diagram	De decision discovery diagram	Bm bcg matrix	Stc strategy canvas	Vc value chain	Hy hype-cycle	Sr stakeholder rating map	Ta taps	Sd spray diagram

Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

version 1.5

Periodic table of visualization methods

- http://www.visual-literacy.org/periodic_table/periodic_table.html



Main steps when designing visualization

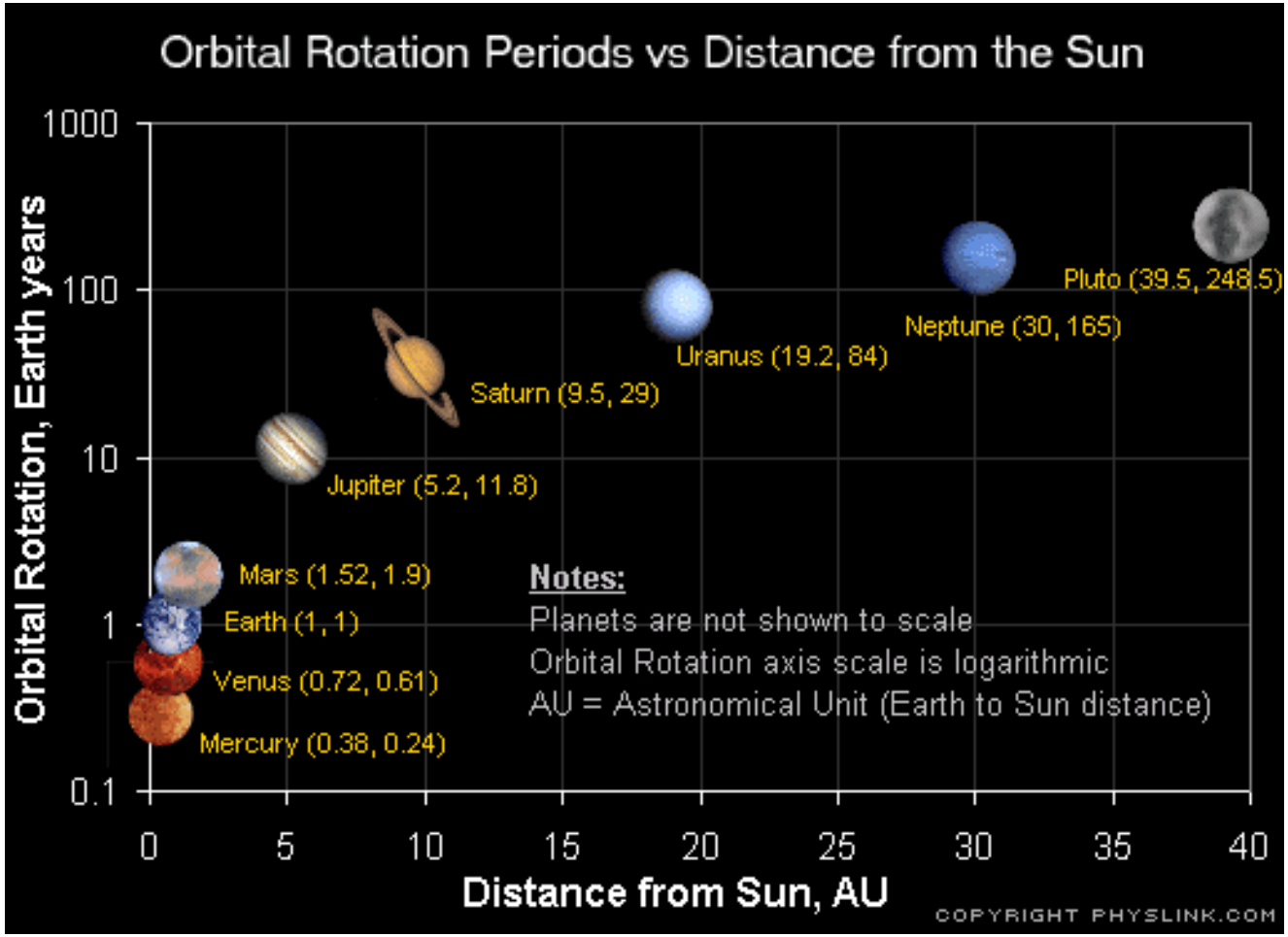
1. Decision about the data mapping to graphic attributes
2. Selection and implementation of methods for changing the view
3. Selection of the amount of data for visualization
4. Displaying additional information – labels, legends, …
5. Overall aesthetics of the resulting visualization

Intuitive mapping of data to visualization

- We need to consider the data semantics and the typical user
- Suitable data mapping to graphic attributes has to correspond to the mental model of the user
- The higher consistency between the design and user expectations, the higher chance of correct interpretation

Intuitive mapping of data to visualization

- Intuitive mapping = faster interpretation



www.physlink.com

Intuitive mapping of data to visualization

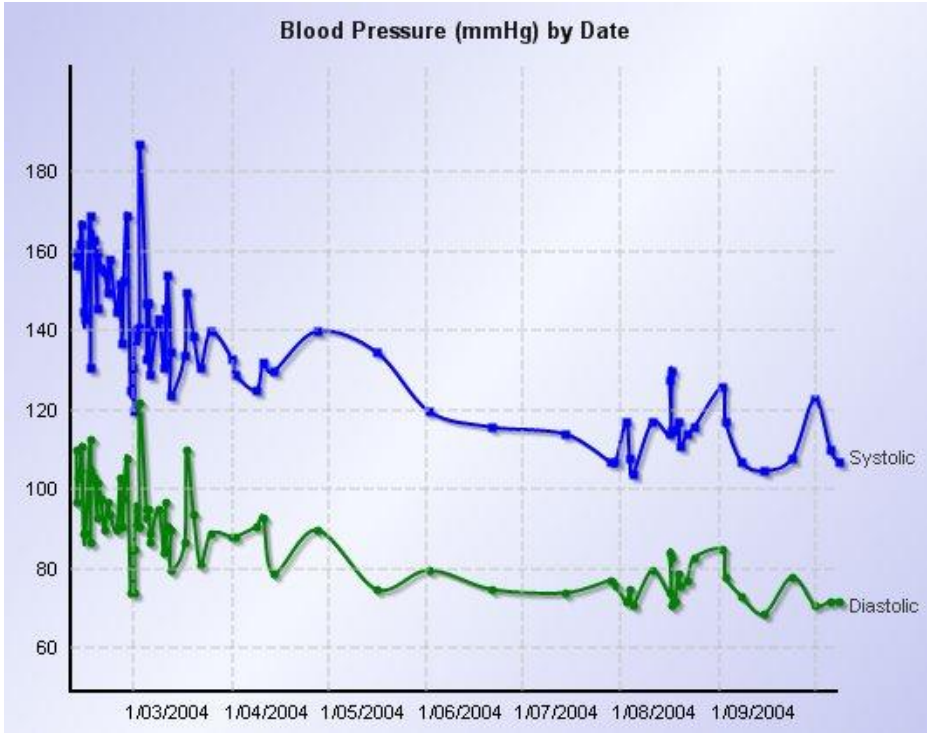
- The most common is the mapping of spatial data attributes (longitude, latitude) to positions on screen
- First visualizations used the human capability of understanding the relationship between the 2D position on screen and the corresponding 3D position in the real world
- Animations enabled intuitive and natural representation of time-dependent data

Types of mapping

- Some mappings are intuitive only in a corresponding context
  - e.g., mapping temperature to color
- Color has a specific interpretation in many fields – cartography, geology, …
- The application domain highly influences the color used

Types of mapping

- Other possible representation of temperature (when having the values from thermometer) – height, length of line
- Length often used in medicine – visualizing blood pressure, …



www.my-health-software.com

Selecting the type of mapping

- We have to keep the compatibility between data scaling and scaling of the corresponding graphic entities or attributes
  - For sorted data attributes (e.g., age) we shouldn't use unsorted graphic attributes (shape)
  - Unsorted data attributes (place of birth) shouldn't be mapped onto sorted graphic attributes (length)

Selecting the type of mapping

- Sometimes it is interesting to explore data using an alternative mapping – may reveal interesting data properties
- **Rule of thumb for mapping = the most intuitive default mapping, appropriate to the common user + enabling the user to apply different settings**

Selection and modification of views

- One view onto data is often insufficient
- We have to assess the most appropriate views and changes of views which are typically used by the users, and to provide the users with intuitive manipulation with views

Selection and modification of views

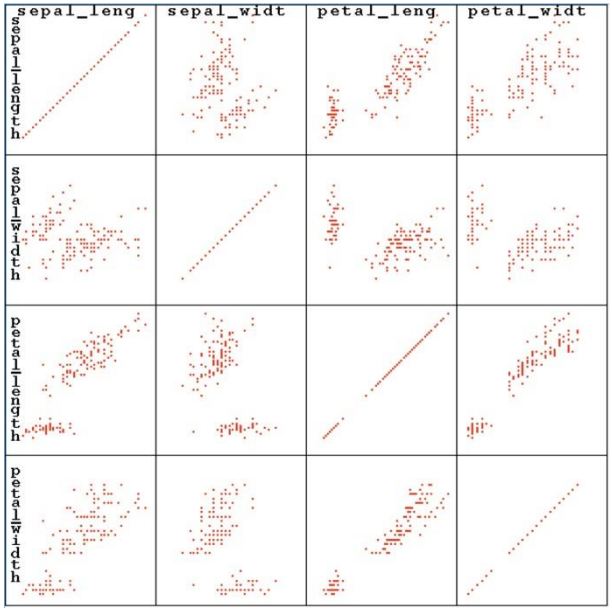
- Each view should be clearly marked and selection of a new view should require minimal amount of interaction from the user
- View changes can be categorized according to the user priorities
- These categories are:

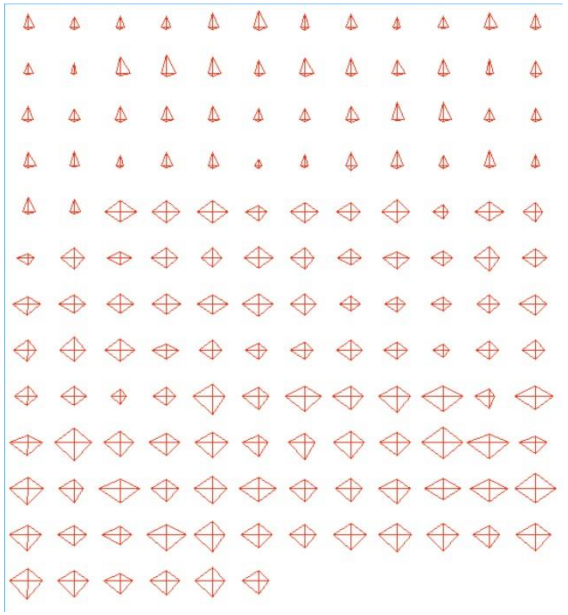
Selection and modification of views

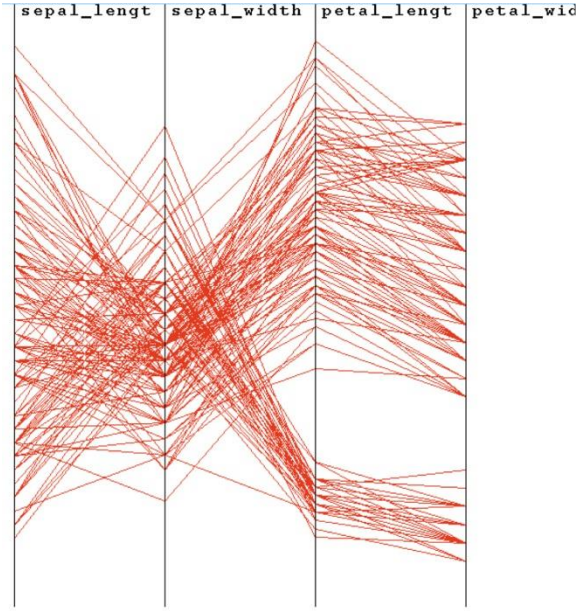
- **Shifting and zooming** – required when the whole dataset cannot be viewed at once in the required resolution.
- **Color palette modification** – almost always required. At least the user should be able to choose between several color palettes. Better to have the possibility to change individual colors or whole palettes.

Selection and modification of views

- **Modification of mapping** – the user can switch between different types of visualization of the same dataset. It helps to reveal different data properties.

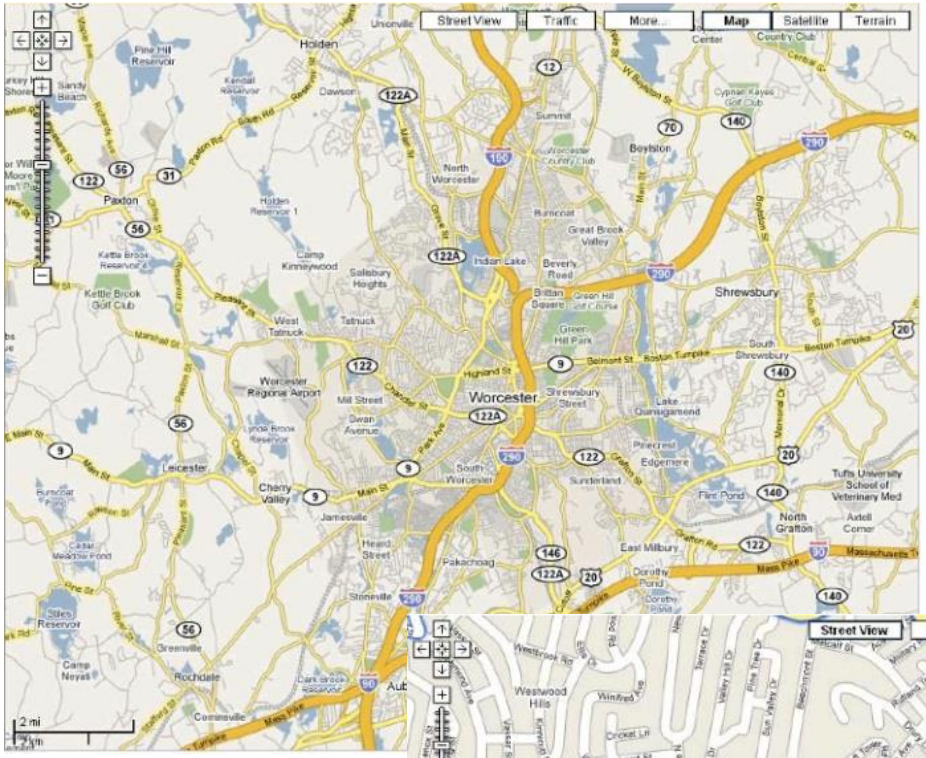






Selection and modification of views

- **Controlling the scale** – modification of range and distribution of the data before mapping itself.
- **Controlling the level-of-detail** – enables to stress/suppress details and support views on different levels of abstraction. Possibility to switch between levels.


Street View
Traffic
More...
Map
Satellite
Terrain
Holden
Paxton
Worcester
Shrewsbury
Leicester
Cherry Valley
Jamesville
Westwood Hills
Elm Park
Beaver Brook
Worcester
2 mi
1000 ft
200 m


N
Somerset St
S
Somerset St

Selection and modification of views

- In all cases
  - the manipulation has to be intuitive and easy to understand and remember
  - it has to have correct preciseness
- If possible, it is preferable to enable direct manipulation (changes are specified directly in the visualization)

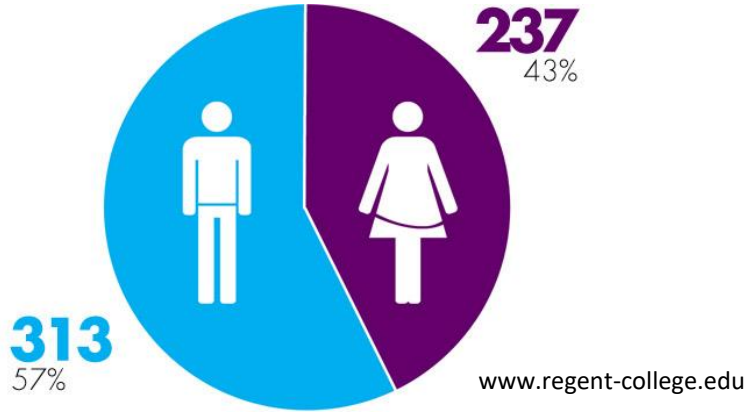
Density of information

- One of the key aspects when designing visualizations
- Two extreme cases:
  1. Very small amount of information to show
  2. We are trying to convey too much information

1. Small density

- Often we determine between two or three different values
  - Male/female ratio
  - Deriving other values from the basic ones -

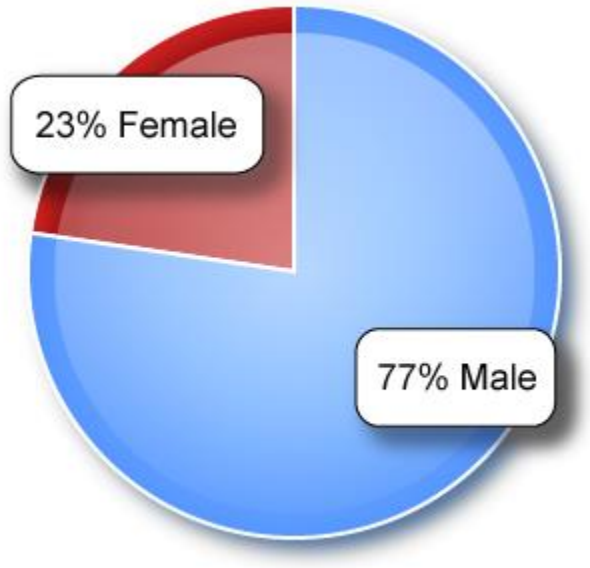
    e.g., two numbers and their sum or difference



- In these cases, it is more efficient to show these values directly as text – saving screen space

1. Small density

Only because we are capable of creating visualizations, we don't have to create them!



vs.

23% Female
77% Male

2. High density

- Too much information can lead to confusion and wrong interpretation by the user
- Important information in data can be easily scattered and the user won’t know where to focus attention

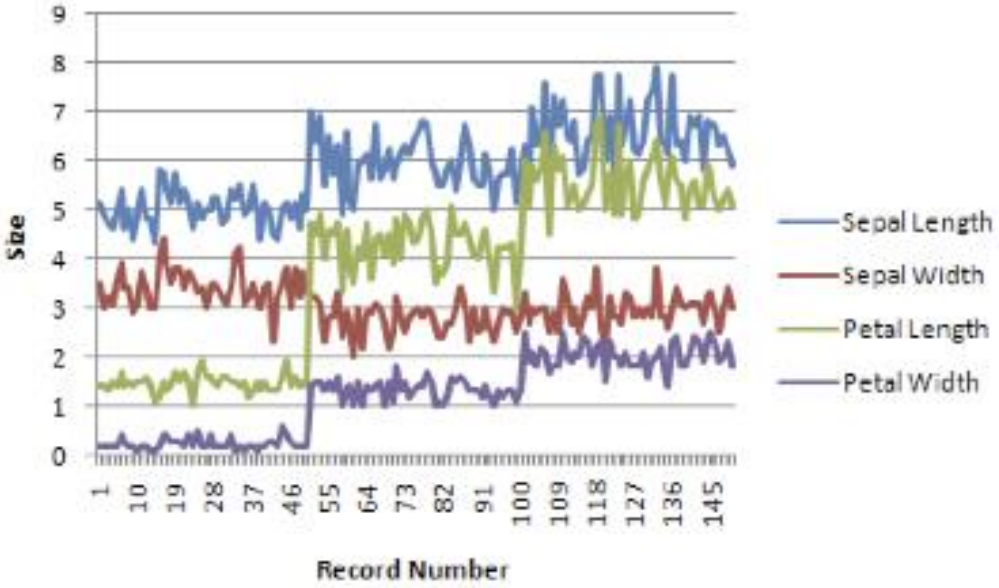
2. High density

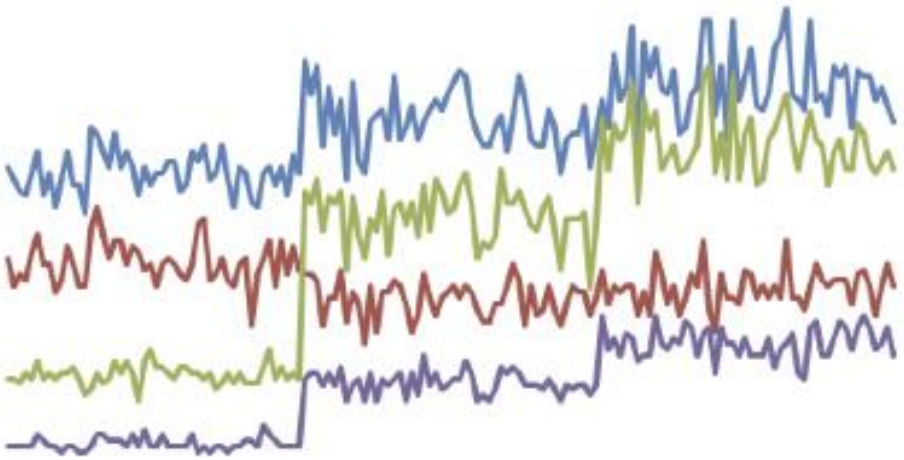
- Many efficient solutions:
  - e.g., the user can show and hide visualization components on demand – the user often decides which parts are the most important ones and the rest is displayed on demand
  - using multiple screens – disjoint panels or we can allow partial overlaps

Keys, labels, legends

- Common problem is insufficient support for additional information (legend) helping to prevent misinterpretation
- This additional information should contain detailed description of the selected mapping properties
- Symbols need key to their encoding
- Color mapping has to be explained

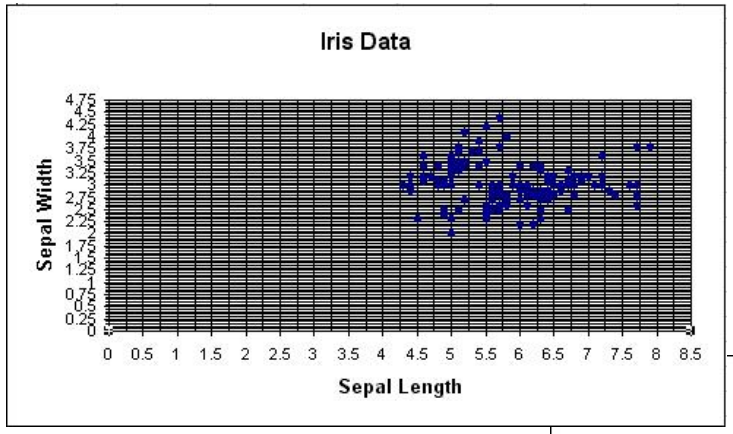
Keys, labels, legends

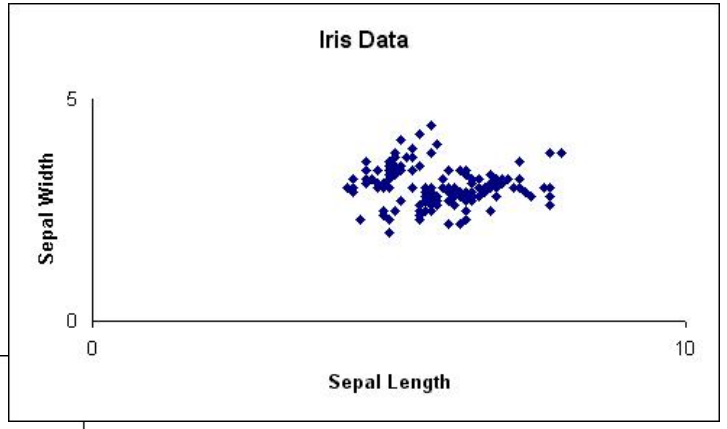


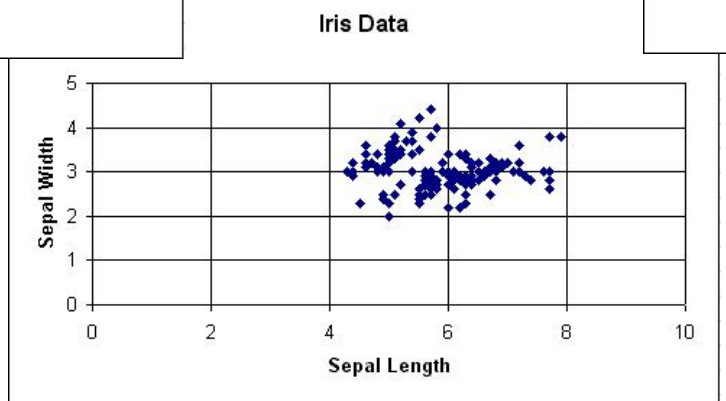


Keys, labels, legends

- It can have positive or negative impact
- Wrong selection of labels and their density can hide the data values themselves

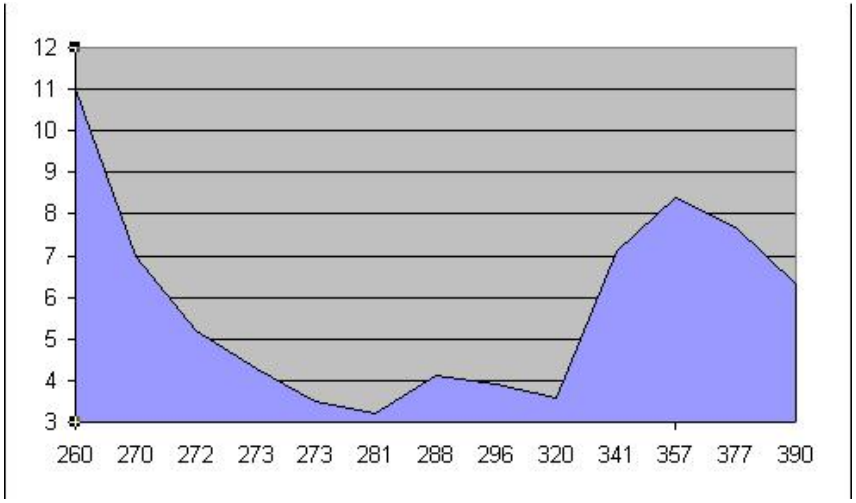


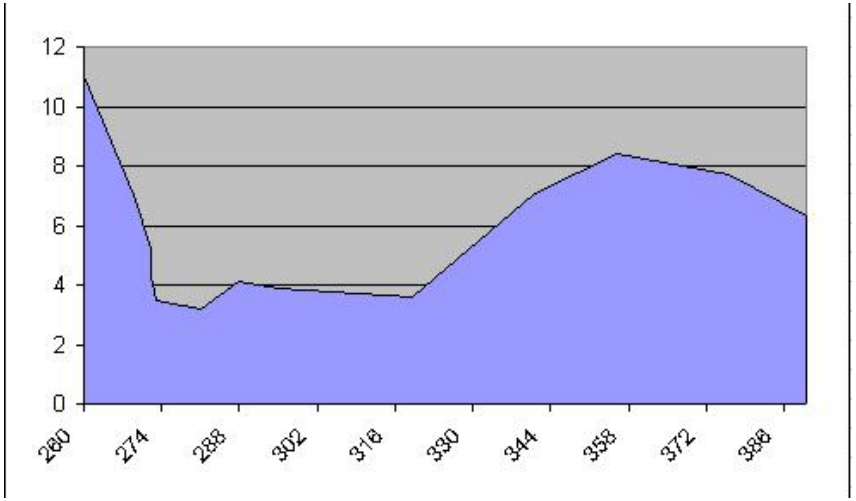




Keys, labels, legends

- Positions of labels determine the readability

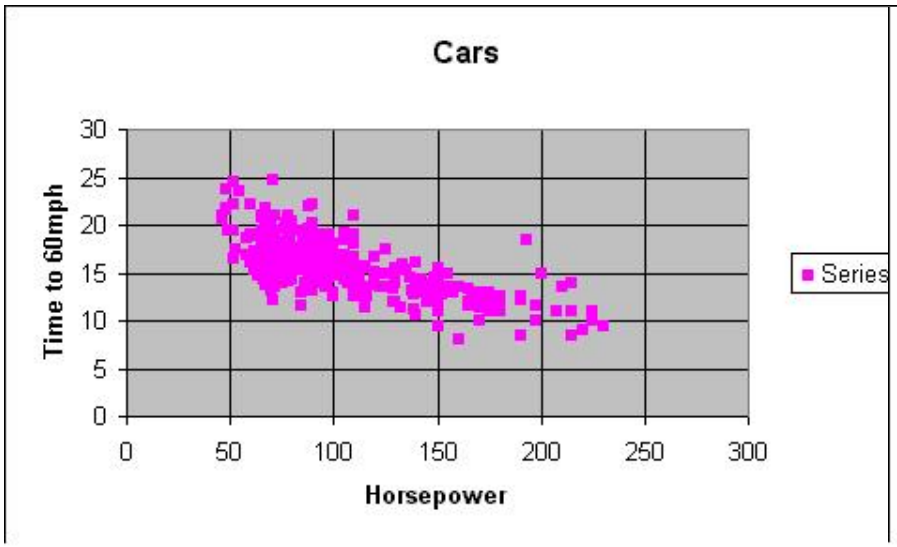


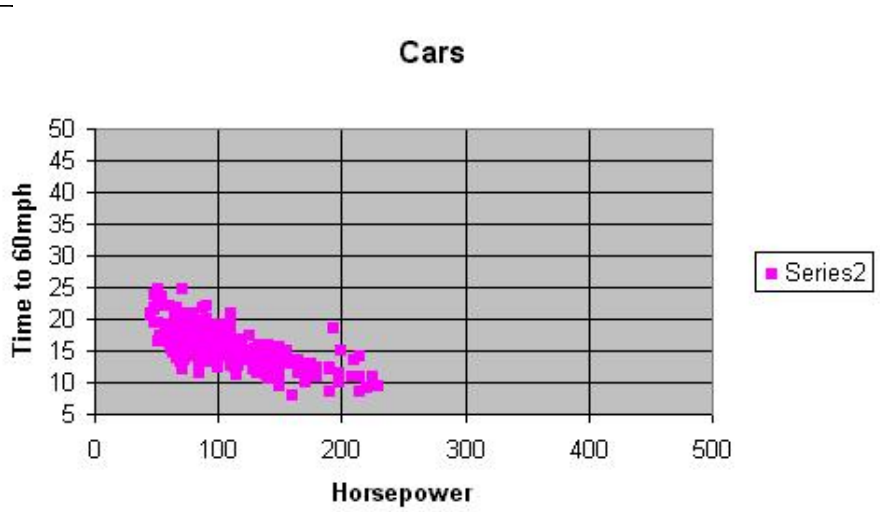


Keys, labels, legends

- We need to determine the data range to be displayed
- We have to display this range in order to avoid misinterpretation
- E.g., when the data represents percentage, we expect the range between 0 – 100. But this can lead to screen space wasting:

Keys, labels, legends





Keys, labels, legends

- When using **multiple views**, we need to keep the consistency in labeling
- Changes in position of labels or range for the same data can lead to misinterpretation
- When the changes are necessary, the labeling should reflect this and convey it to the user

Using color

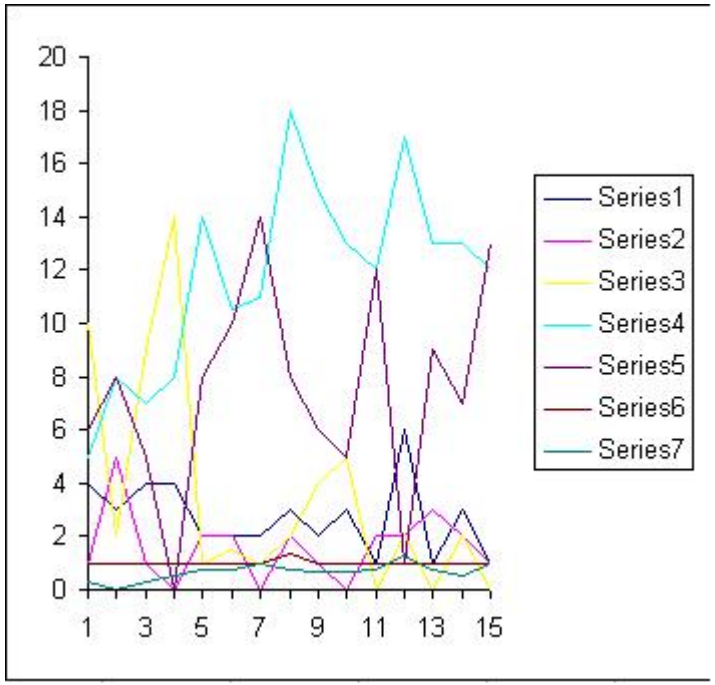
- Color is one of the most commonly used parameters in visualization but also most often in a wrong way ☺
- Wrong color palette selection or trying to convey too much information using color leads to inefficient and misleading visualization
- Color perception depends on the context

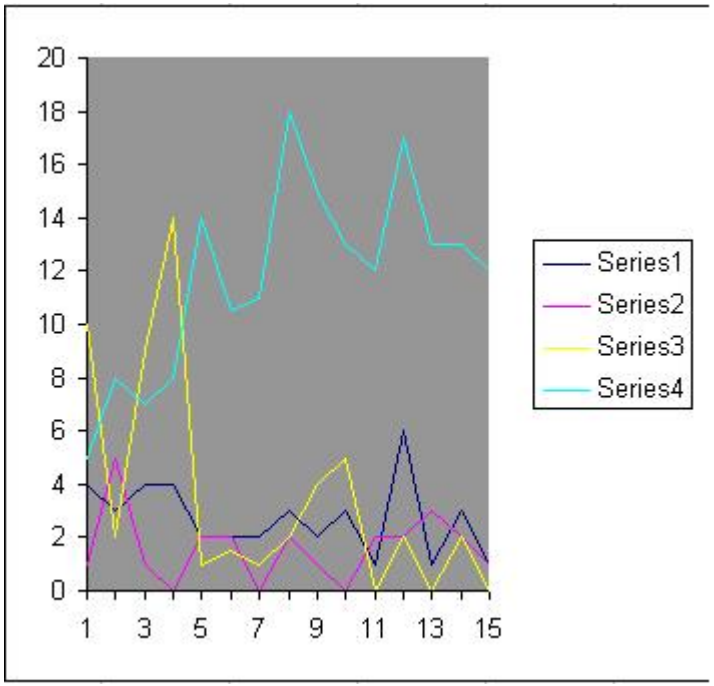
Using color

- Keep in mind that many people have problems with color perception (up to 10% of male population)
- The following rules can help when designing colors for visualization:

Using color

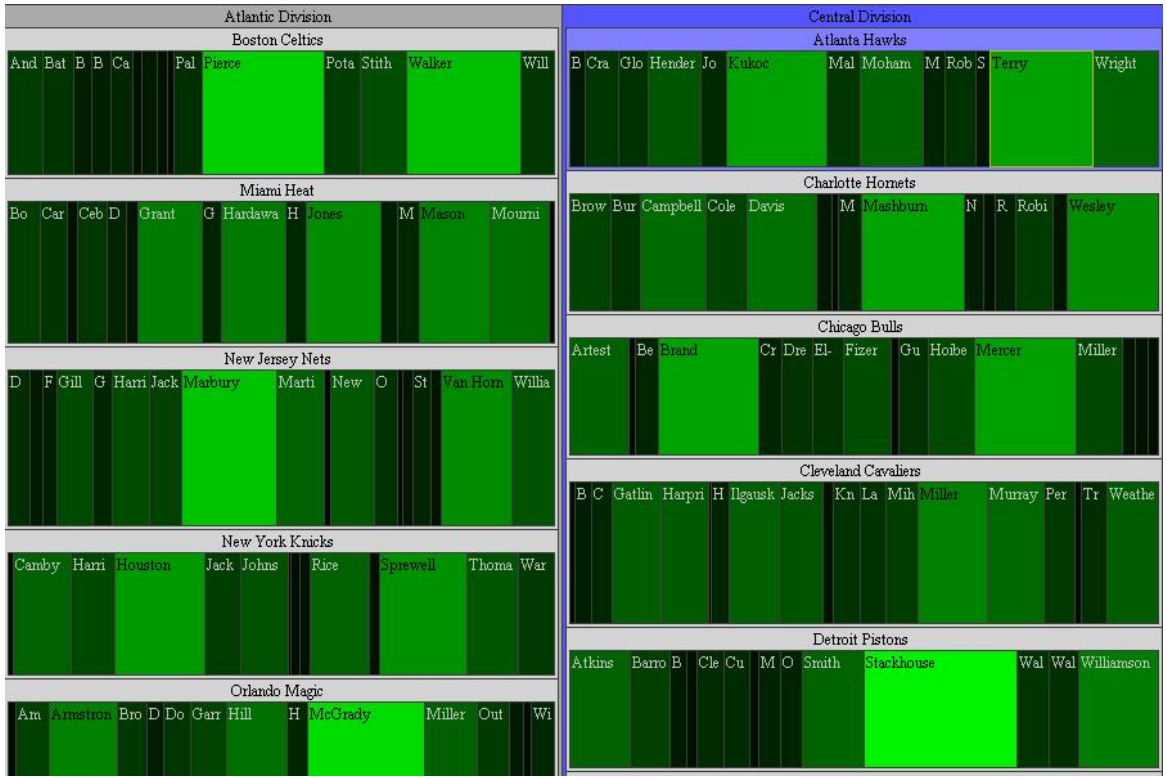
1. When the user needs to clearly distinguish between values, we can use only a limited number of colors





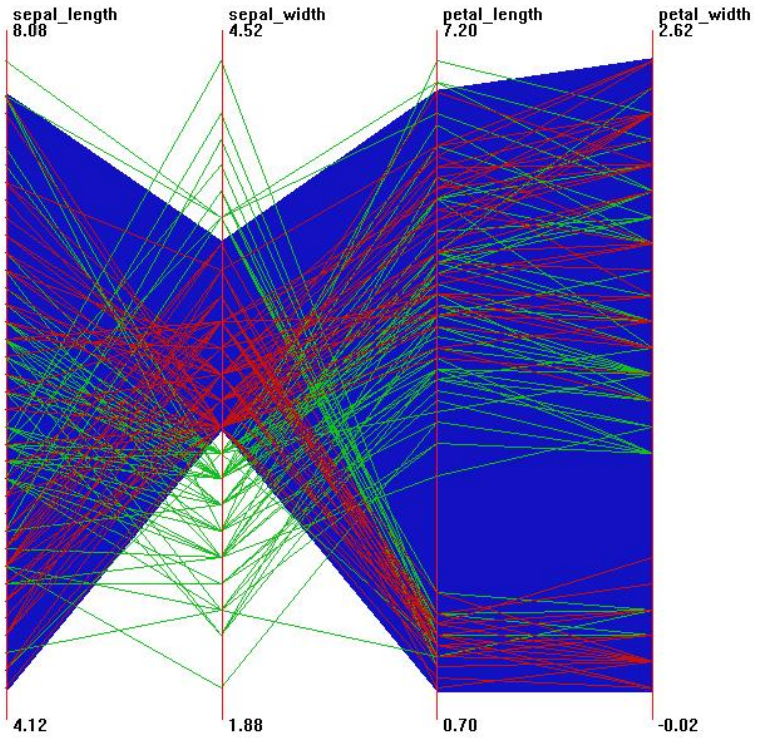
Using color

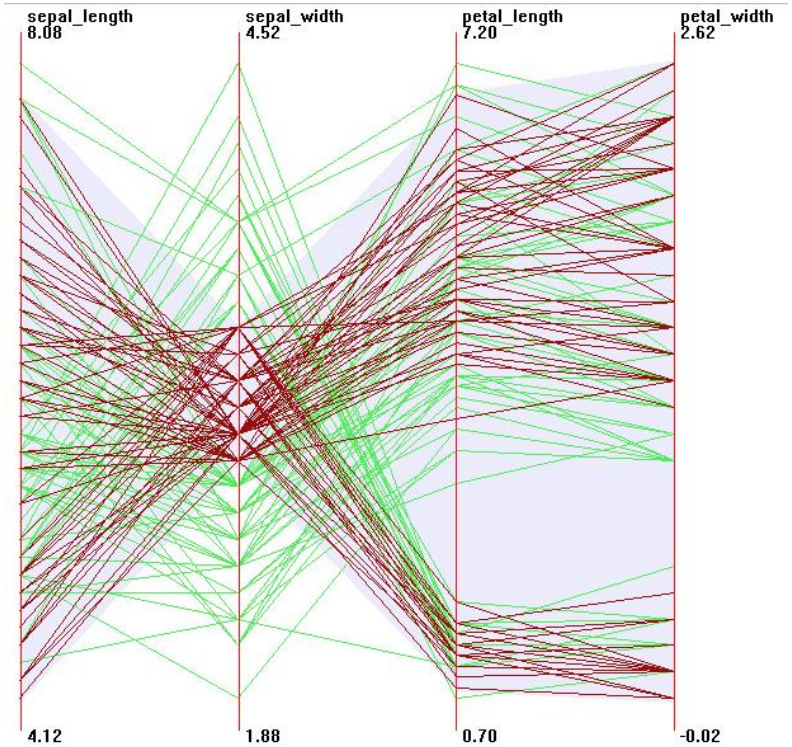
2. If possible, we should use redundant mapping of the same property, e.g., mapping to color and size at once



Using color

3. When creating color palette, we need to change not only the color hue but also its saturation





Using color

4. Don’t forget about the legend explaining the color mapping

- Color can add visual plausibility but can also significantly decrease the efficiency of the visualization
- Some designers prefer to make the initial design in greyscale
- Adding color after testing the first prototype can be more efficient

Importance of aesthetics

- Finally, we should assess the aesthetic part of the visualization
- The best visualizations are both informative and visually appealing
- On the other hand, “ugly” visualizations can discourage the user from the communication process
- Visually appealing visualizations can attract user’s attention to perform more deep data exploration

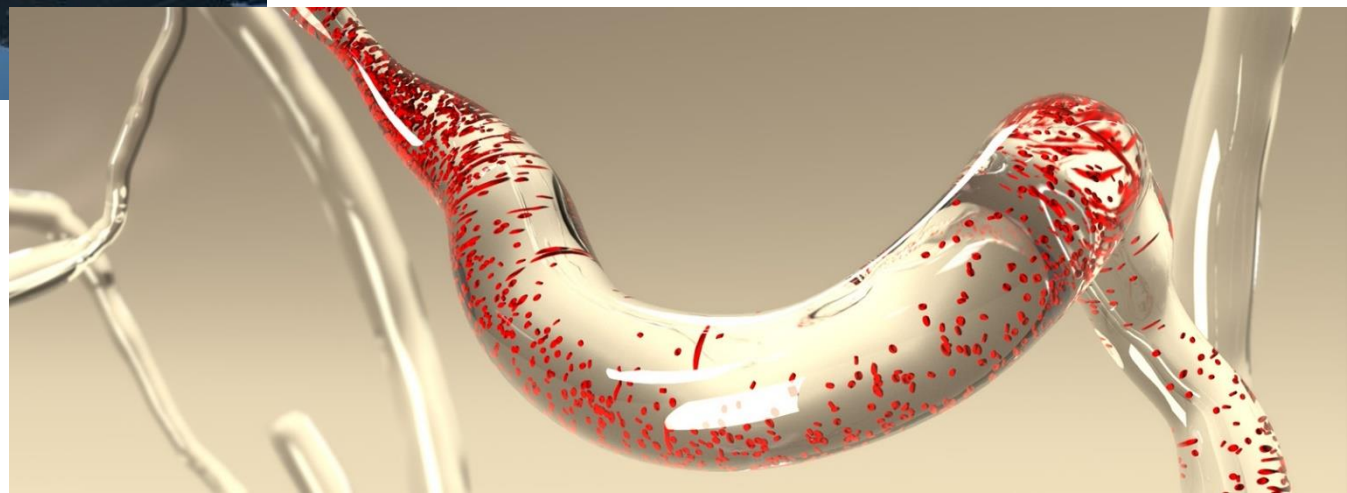
Importance of aesthetics

- Several rules for design of attractive visualizations:
  - Focus
  - Balance
  - Simplicity

www.tableausoftware.com

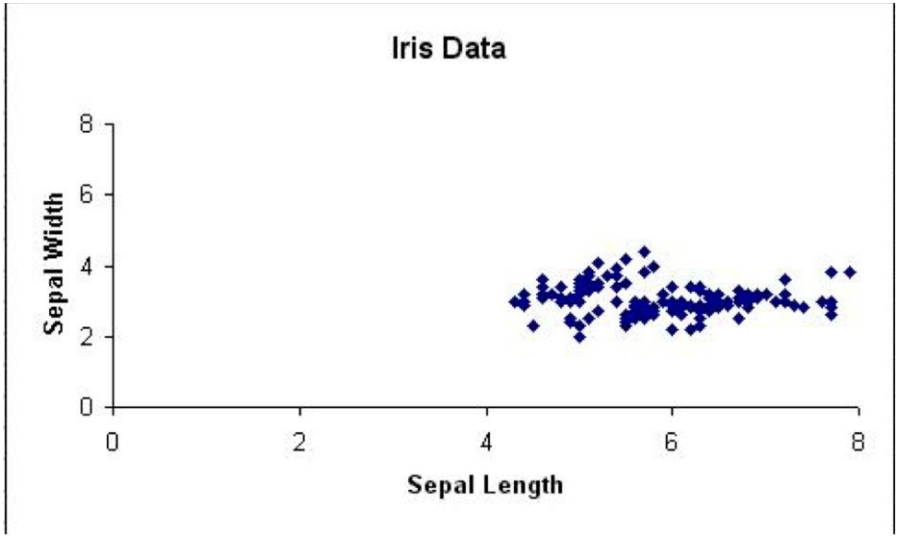
Focus

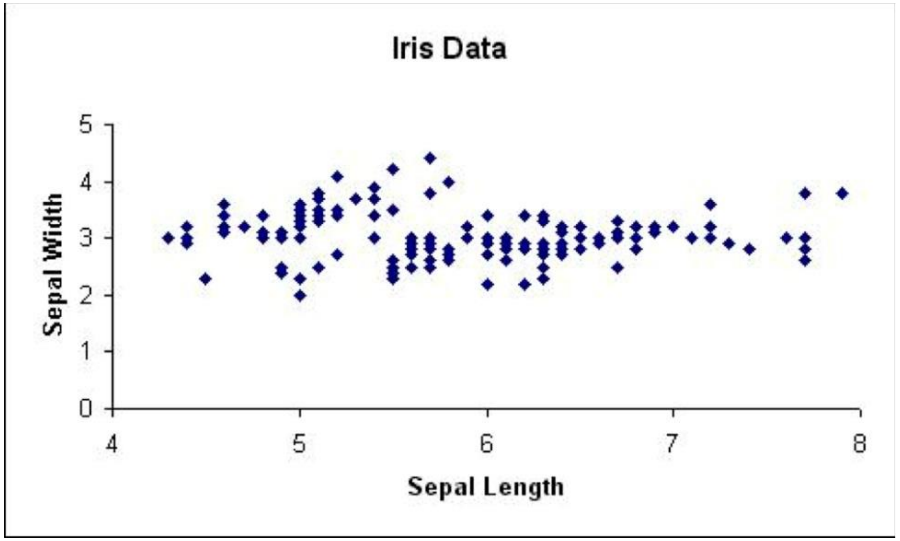
- The user should be navigated to the most important parts of visualization



Balance

- The screen space should be used efficiently, the most important component should be positioned in the center

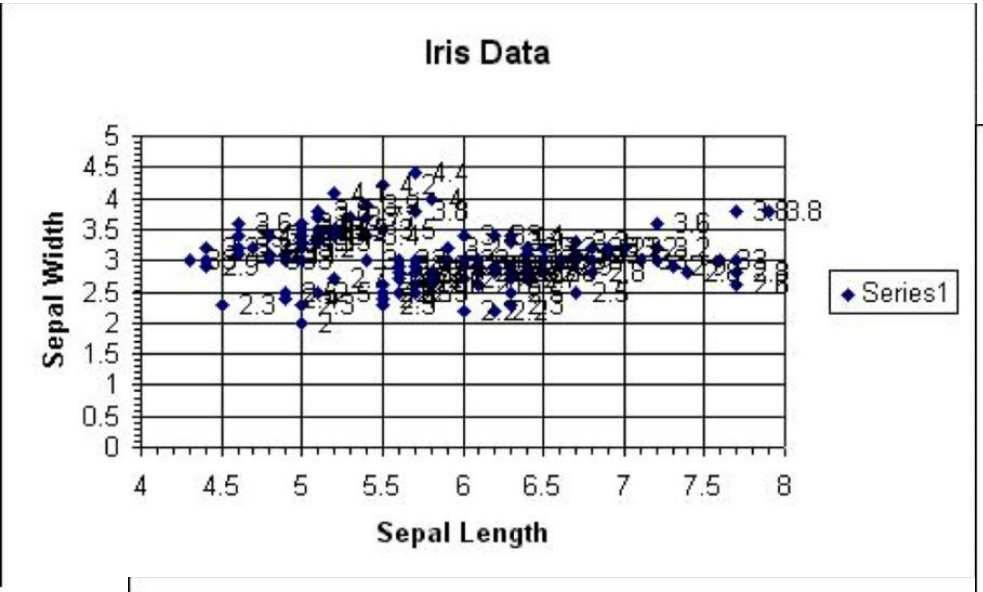


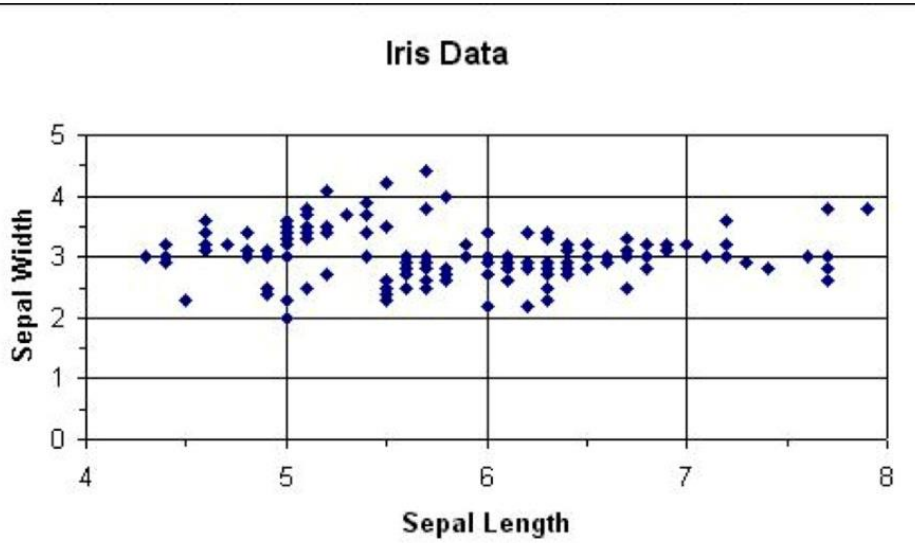


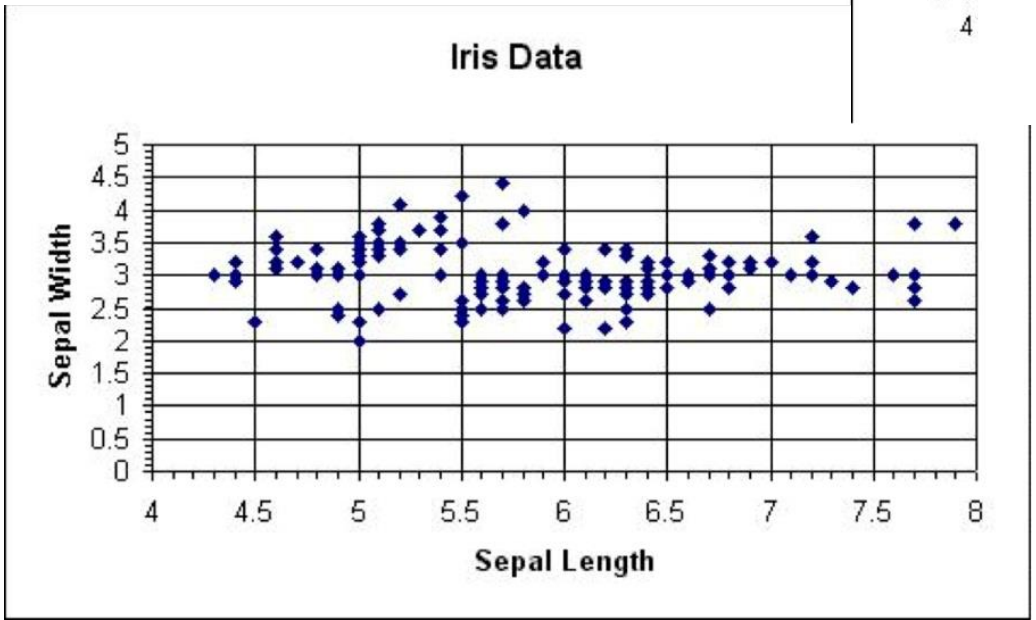
Simplicity

- Rule of thumb – not to squeeze as much information as possible to the screen and not to use different graphic tricks only because they are available
  - do not use histograms in combination with 3D Phong shading when the same information can be conveyed using scatterplot or line chart
- Common procedure for removing these errors = iterative removal of visual properties used and measuring the loss of conveyed information

Simplicity

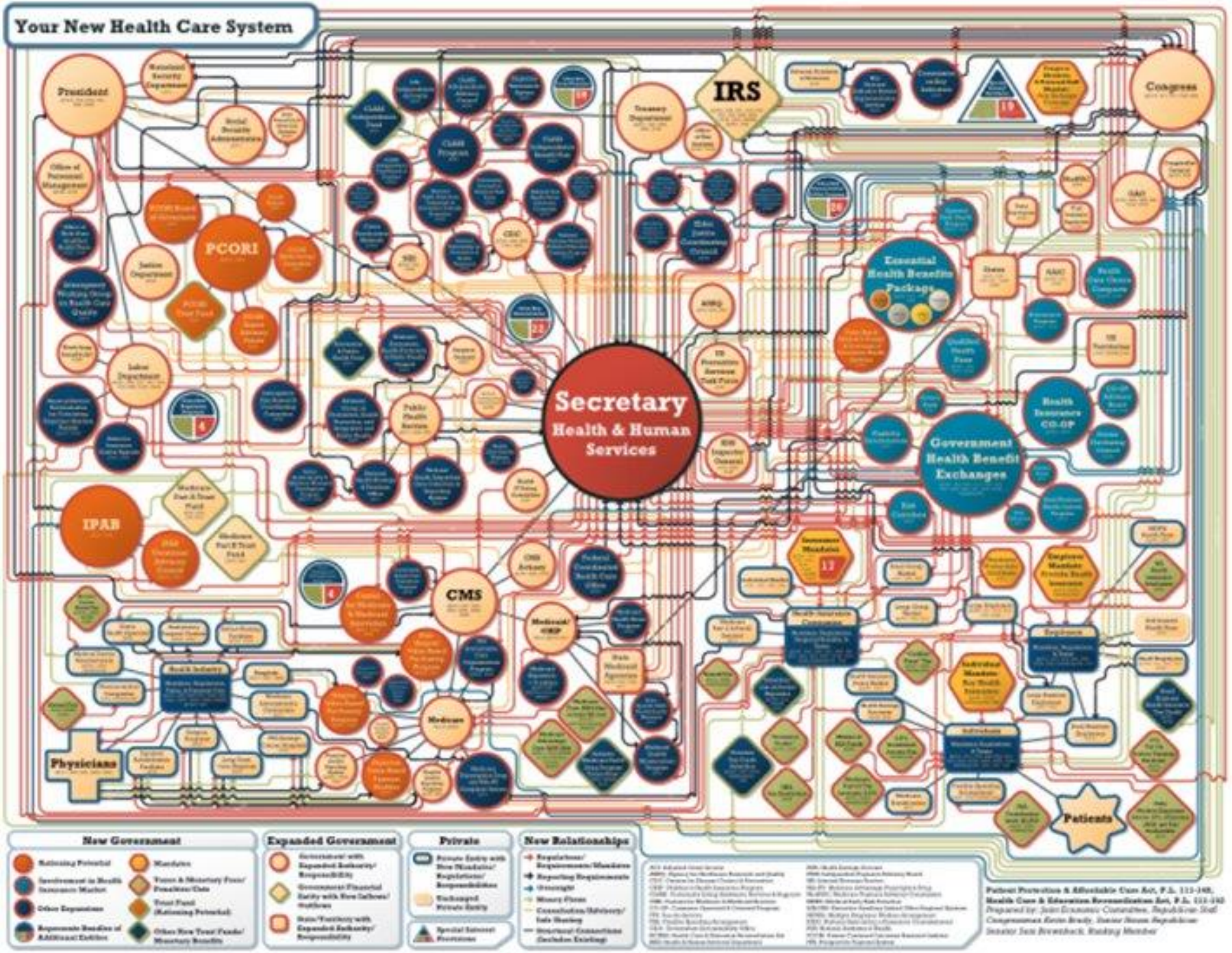

Iris Data
Sepal Width
Sepal Length
Series1


Iris Data
Sepal Width
Sepal Length

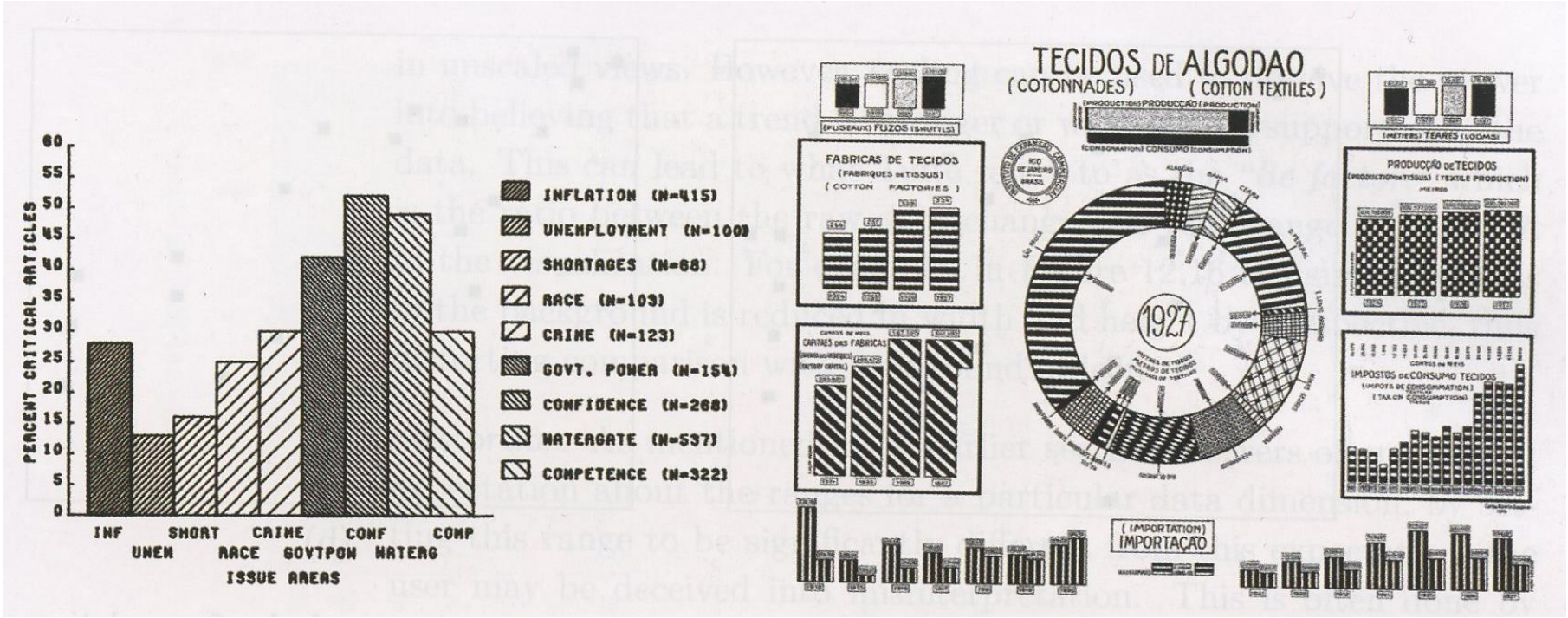

Iris Data
Sepal Width
Sepal Length

The importance of aesthetics

- Literature contains a lot of “ugly” visualization examples
- https://gizmodo.com/8-horrible-data-visualizations-that-make-no-sense-1228022038
- It is desirable to let to assess the aesthetics of the proposed visualization before showing it to the customer


Your New Health Care System
President
IRS
Congress
PCORI
Secretary
Health & Human Services
Essential Health Benefits Package
Government Health Benefit Exchanges
Health Insurance CO-OP
IPAB
CMS
Medicare
Physicians
Patients
New Government
Expanded Government
Private
New Relationships
Patient Protection & Affordable Care Act, P.L. 111-148,
Health Care & Education Reconciliation Act, P.L. 111-152
Prepared by: Joint Economic Committee, Republican Staff
Congressman Kevin Brady, Senior House Republican
Senator Jim Bunning, Ranking Member

The importance of aesthetics



Problems when designing efficient visualizations

- We will focus on common problems which can appear even when following the previously mentioned rules
- These problems have deeper roots and are related to the decision what and how to visualize

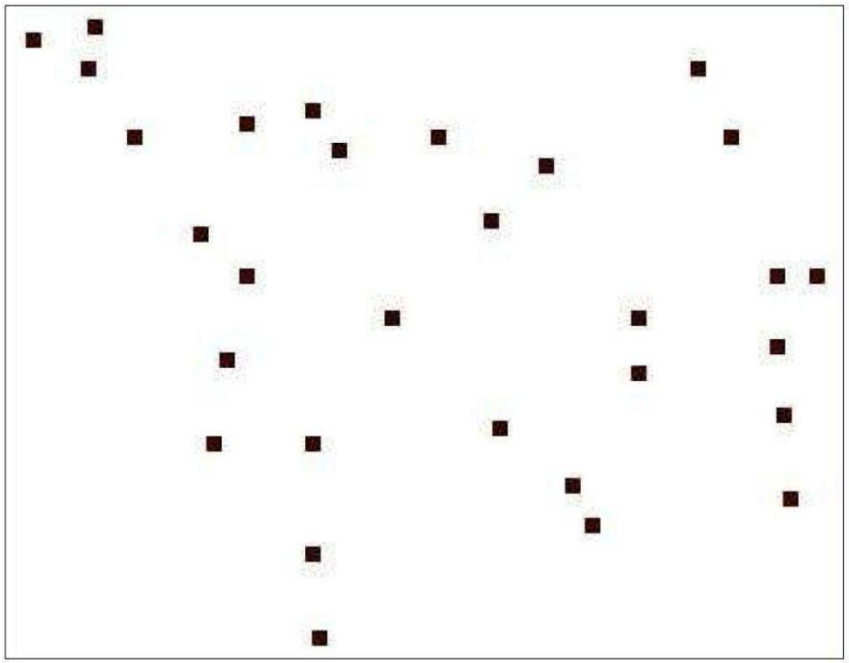
Misleading visualizations

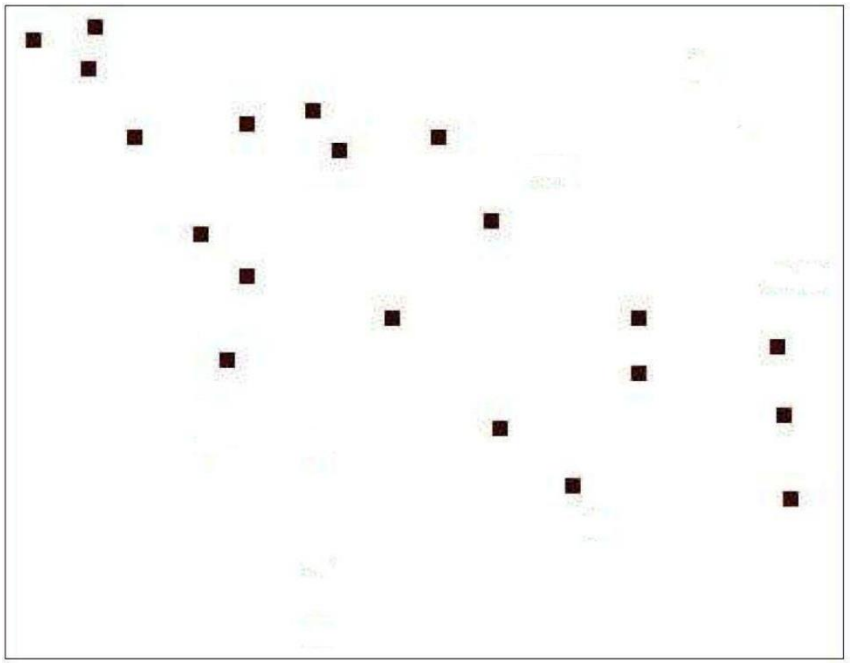
- One of the basic rules is that the visualization should display exactly the required data, nothing more or less
- History showed several examples of visualizations leading to complete misinterpretation – **„viz lies“** (http://www.vislies.org/)
- Lets have a look on several strategies for creation of viz lies – to be able to avoid them in the future ☺

Misleading visualizations

- **Data scrubbing**
  - Raw data can be very rough and we are tempted to remove this roughness
  - The selection can be performed in that way that we can reveal relationships which are not present in the original data

Misleading visualizations





Misleading visualizations

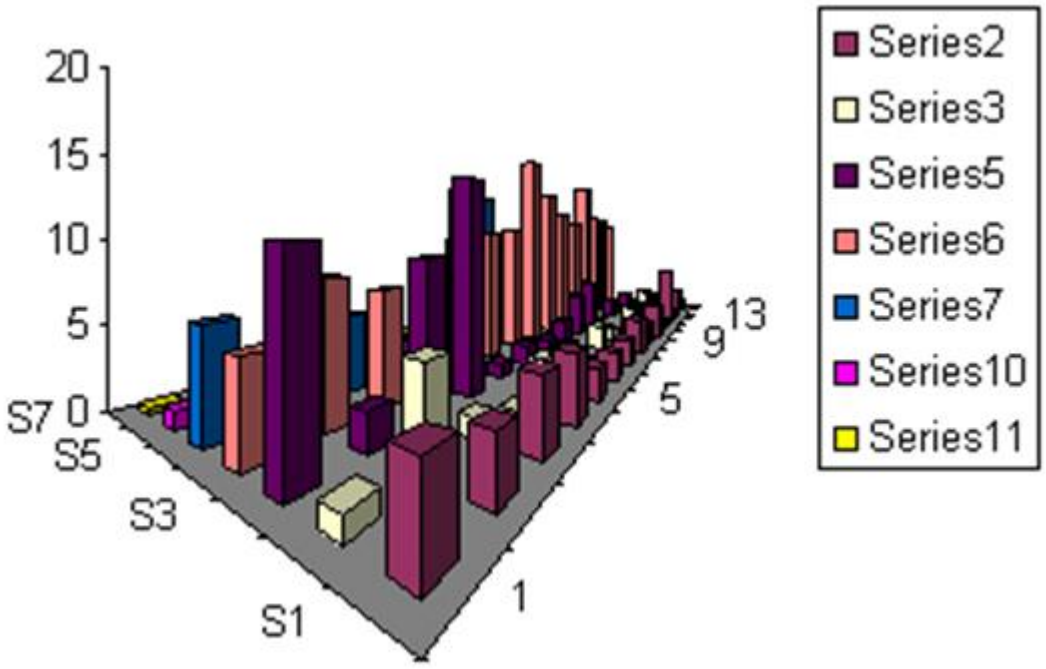
- Commonly used practice is removal of outliers from the dataset
- It is believed that they appeared in the dataset because of the errors in the data acquisition, but we shouldn’t remove them without informing the user

Misleading visualizations

- **Unbalanced scaling**
  - Scaling is very powerful, it helps to reveal patterns and structures barely visible without scaling
  - Scaling can lead to confusion of the user, believing that the trends in data are much stronger or weaker than they are in fact = **„lie factor“** (ratio between the change in the raw data and change displayed in visualization)

Misleading visualizations

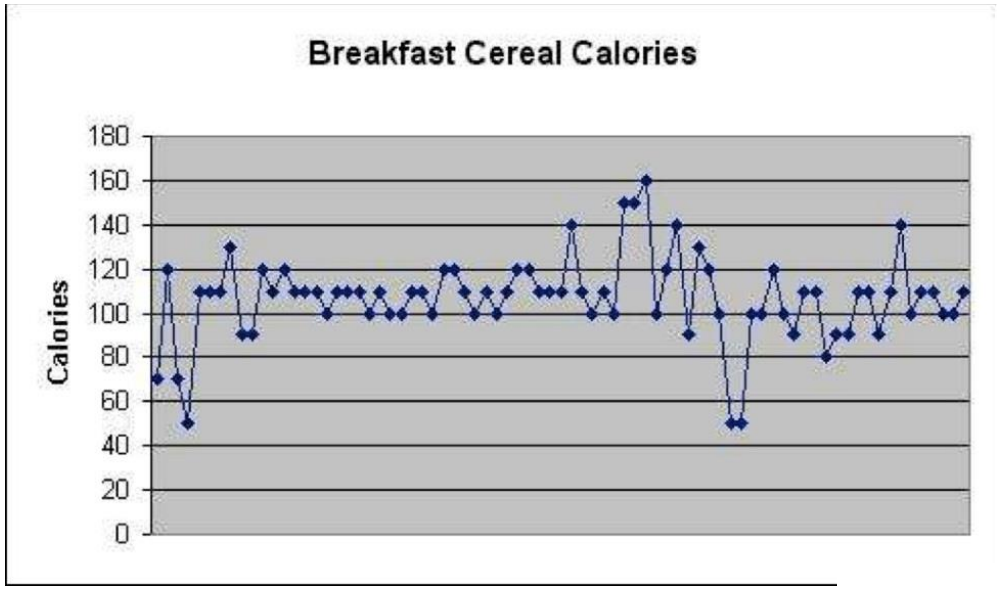
- Size (width and height) of objects further from camera is reduced with perspective projection – we cannot compare values in front and back

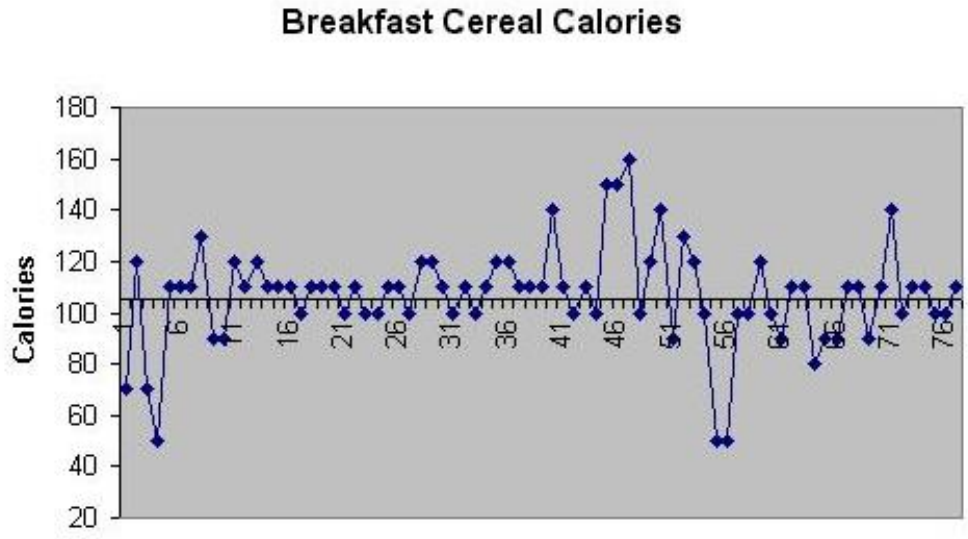


Misleading visualizations

- **Range distortion**
  - Users have often expectations regarding the data range
  - Setting a different range can lead to confusion
  - Typical example is shifting the axes in graph that it does not correspond to the expected zero value

Misleading visualizations





Misleading visualizations

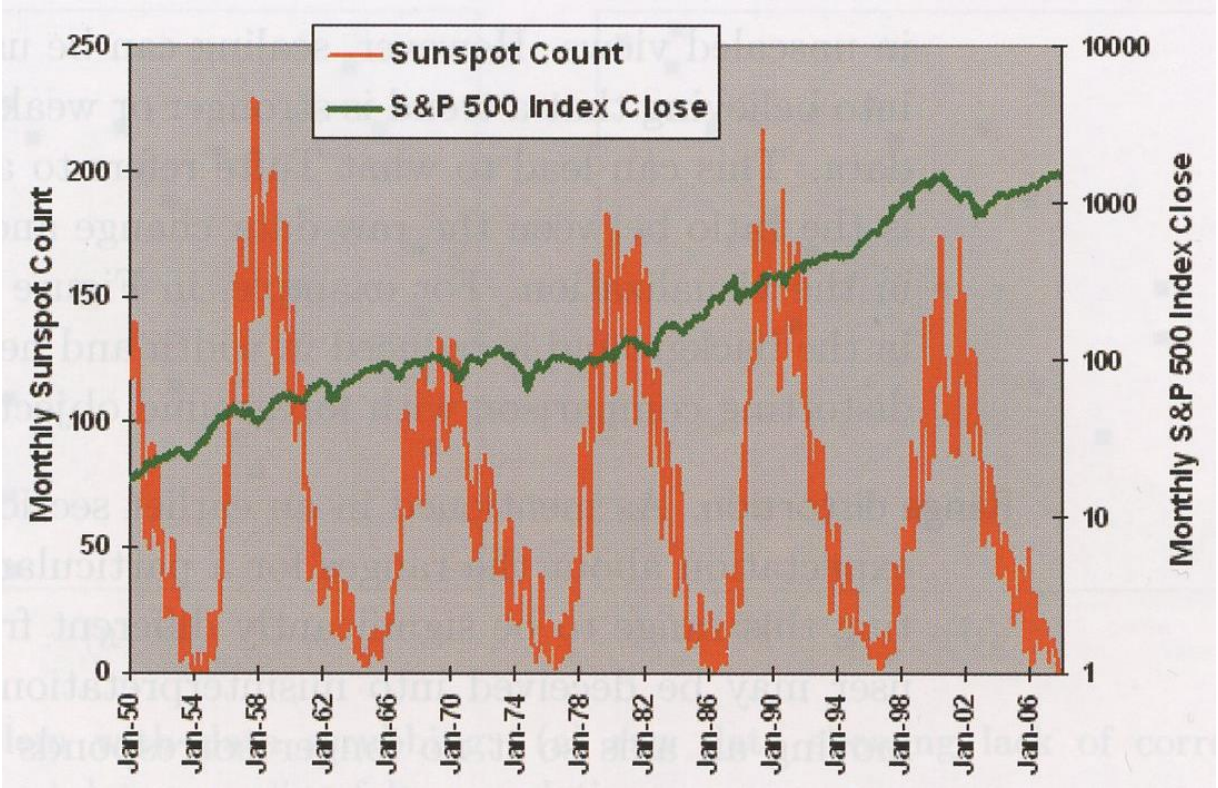
- **Abusing dimensionality**
  - The number of errors in interpretation is proportional to the power of the number of dimensions of the displayed data
  - Error when assessing volume is much higher than when assessing area and that is higher than when assessing length
  - Mapping scalar values to volume highly increases the chance of misinterpretation
  - The rule: **the simpler the better**

Visual nonsenses

- Visualizations are created in order to convey information and this information should make sense
- We often combine datasets from different sources
- It is easy to combine these datasets to one visualization and try to identify some structure in them …

Visual nonsenses

- Relationship between the stock market and the number of sunspots

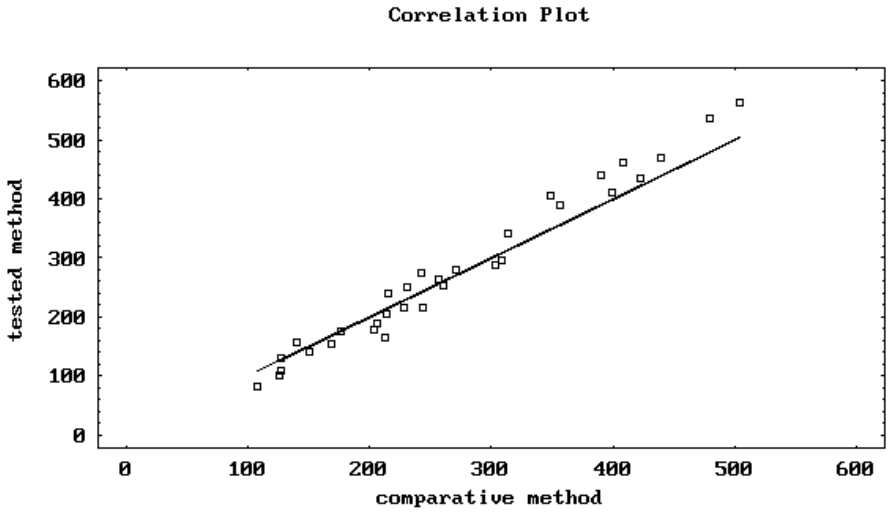


Visual nonsenses

- When deciding if to combine the datasets or not, we need to find some logic in their combination
- Another factor is the compatibility between the temporal and spatial ranges in the datasets
- Also the compatibility of units plays a role

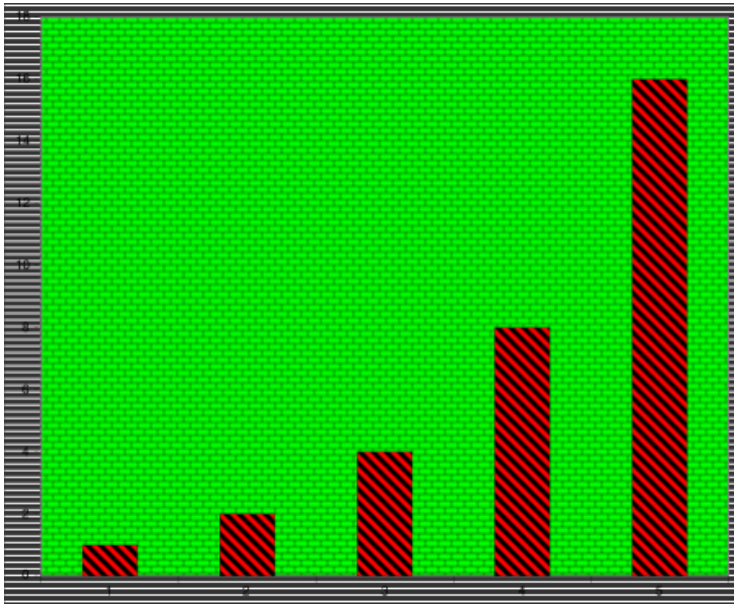
Visual nonsenses

- We often try to apply operations suitable for sorted or continuous data to unsorted, categorical data – simply cause the visual result is a sorted graphic representation



Data loss thanks to Chart Junk

- **Chart Junk** = all visual parts of graphs and diagrams which are not necessary for correct interpretation, they distract the user



http://en.wikipedia.org/wiki/Chartjunk

Data loss thanks to Chart Junk

- This additional information can lead to very complex visualizations, occlusion, and suppressing the visualization itself
- We need to decide the amount of additional graphics elements which contribute to aesthetics and do not distract the user – complex task
- The user can influence the amount as well

Data loss thanks to Chart Junk

- **Rule of thumb:**

  To provide support for additional tools and enable the user to remove them on demand

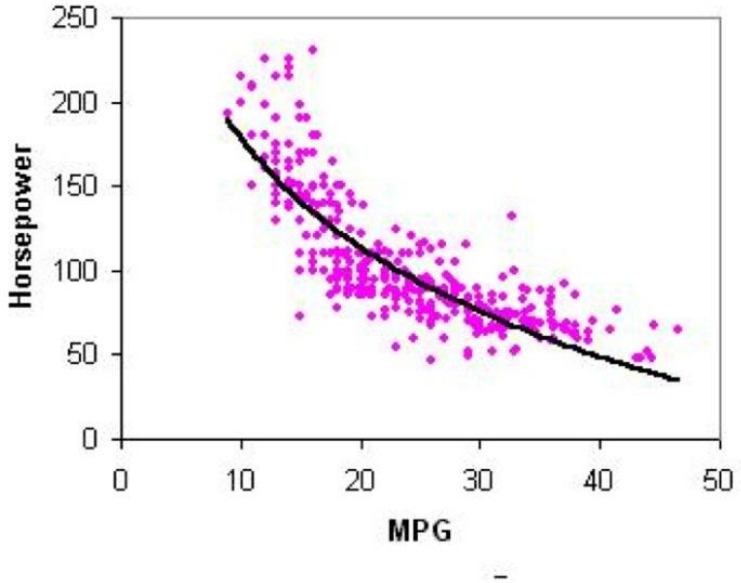
Raw vs. derived data

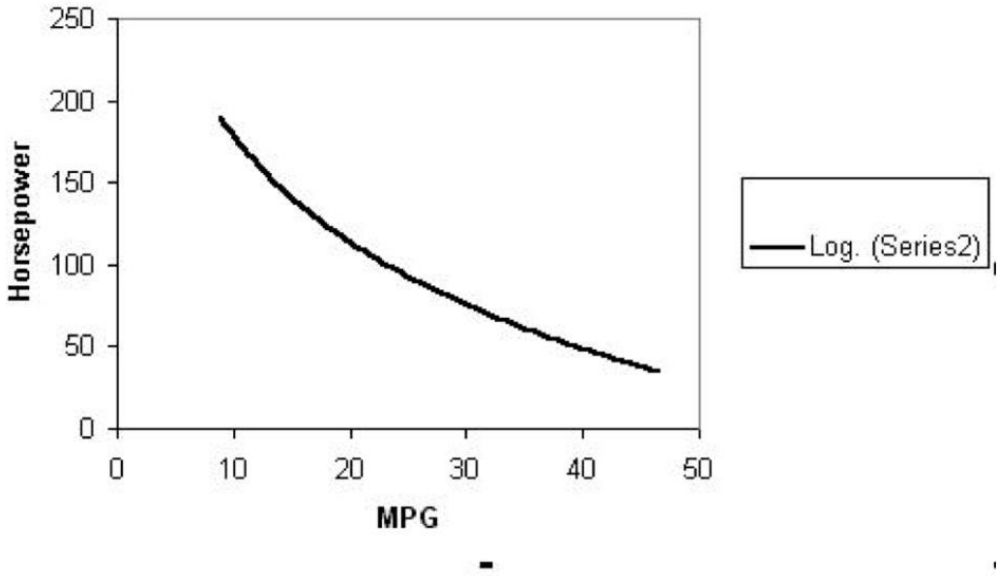
- Common practice is to calculate an analytical model of data, represented by curves or surface
- This can again lead to misleading interpretation

Raw vs. derived data

- In some visualizations the common practice is to remove all raw data and replace them by a smooth approximation derived from the data
- The user then has to believe that the representation is correct which often is not true
- The best way is to show the raw data as well as its approximation and enable the user to filter these on demand

Raw vs. derived data

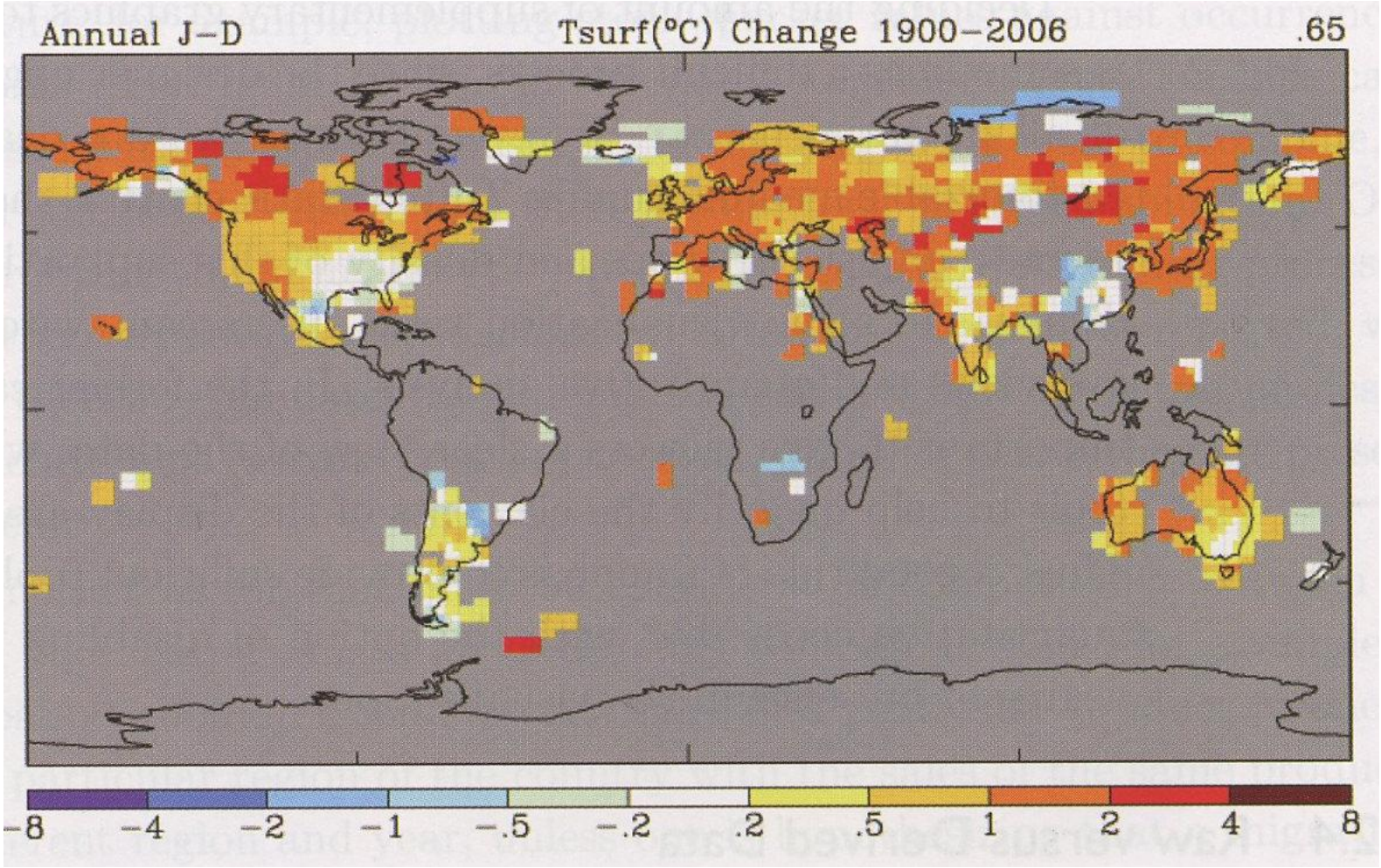




Raw vs. derived data

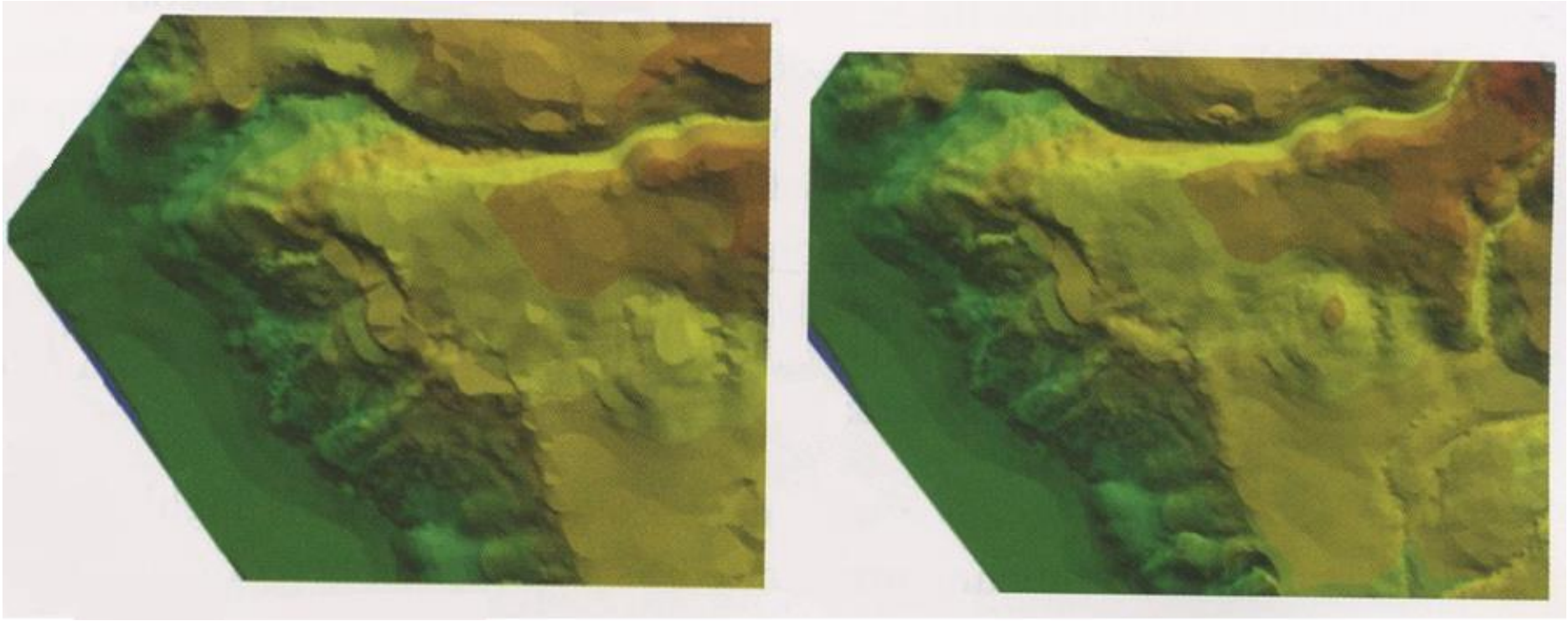
- Another commonly used technique is **resampling** when the raw data originally positioned randomly or in a sparse grid are used for creating denser grid
- In this way we create more “rich” visualizations which are misleading the user
- The denser resampling, the higher probability of wrong data interpretation

Raw vs. derived data



Raw vs. derived data

- Another problem can be insufficient sampling:



Raw vs. derived data

- The user has to have access to the raw data and has to be informed about every data preprocessing
- In some fields (e.g., radiology) the experts are strictly against any data preprocessing

Interesting links

- Four Pillars of Effective Visualizations

http://www.youtube.com/watch?v=nrsdgvauqKg

- Designing Data Visualizations

http://www.youtube.com/watch?v=lTAeMU2XI4U

- Designing Data Visualizations with Noah Iliinsky

http://www.youtube.com/watch?v=R-oiKt7bUU8