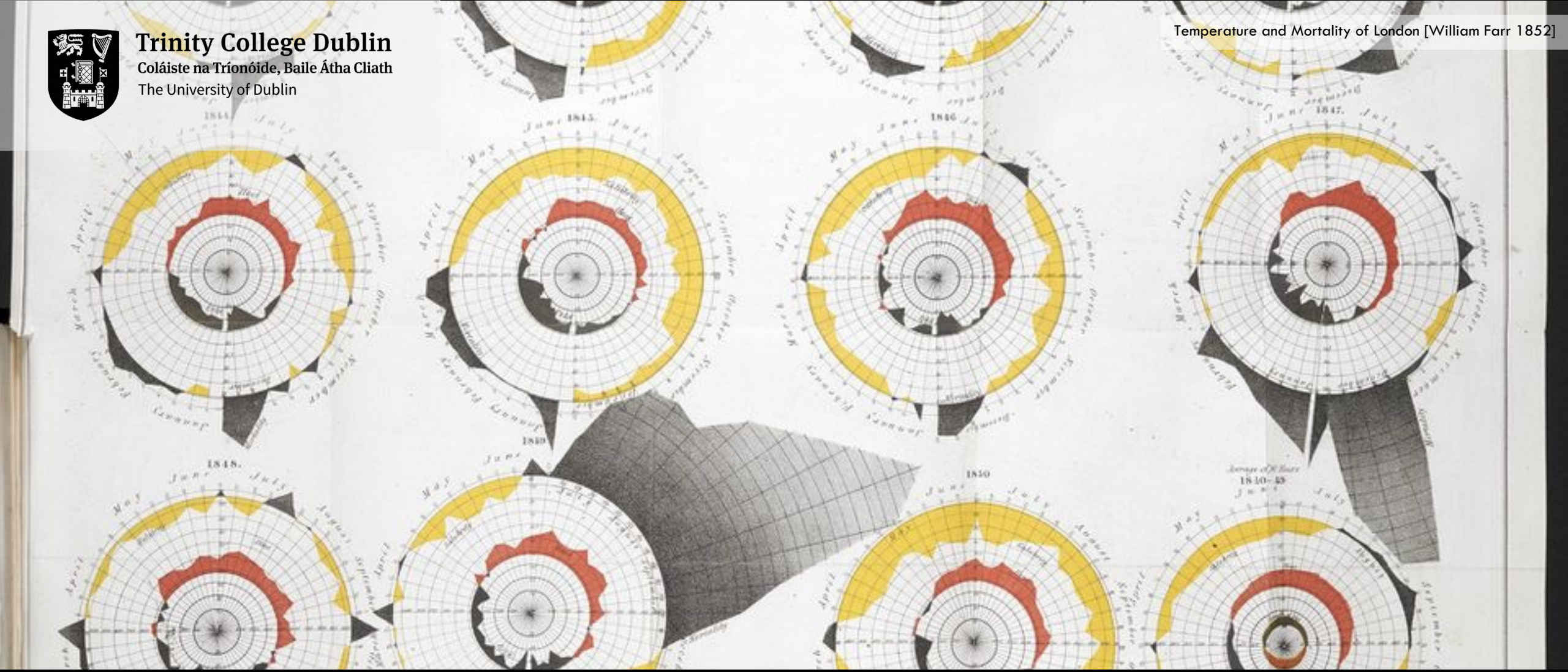
Graphical Integrity

Avoiding Distortions in Visualisation

Temperature and Mortality of London [William Farr 1852]



Aside: **Vis. Analysis Example**

Farr's analysis of the cholera epidemic

Temperature and Mortality of London [Farr 1852]

The following is a simple example of an analysis of a classic visualization, similar to what you are being asked to provide for assignment 2.1

The objective is to critically analyze a visualization in order to inform how we design our own visualizations effectively.

About the visualization [https://goo.gl/eN2TjS]

- William Farr investigated cholera in the same era as John Snow
- This historic visualization investigates one aspect of his hypothesis of a miasma link to the disease
- N.B. This background detail is not required in your analysis. We're only interested in what you can read from the visualization itself.


TEMPERATURE AND MORTALITY OF LONDON For every week of 11 years (1840-50.)
Registrar General's Report on Cholera in England. 1848.
Plate 4.
1840.
1841.
1842.
1843.
1844.
1845.
1846.
1847.
1848.
1849.
1850.
Average of 10 Years 1840-49
Circular Diagrams showing the relative Mortality and the mean temperature in each week of the 11 Years 1840-50.
The distance between each circle represents either 100 deaths or 10° of temperature. the radii represents the weeks of the year the numbers of which are arranged round the outside circle from 1 to 52. Besides these circles which serve for a scale there are two others. The two extra circles, the one running outside the yellow colour and inside the black; the other outside the blue and inside the red colour) represent respectively the average weekly deaths of the ten Years 1840-49 corrected for increase of Population, and the mean temperature (at 25) of the 29 years 1721-1849.
The outer irregular line represents the Deaths, and the inner irregular line the mean temperature in each week of the year.
The black colour denotes the extent by which the weekly deaths exceed the average (Excessive Mortality) and the yellow colour denotes the extent by which the weekly deaths are below the average (Salubrity). The red colour denotes the extent by which the mean temperature of the week exceeds the mean temperature of the 29 years, (Heat) and the blue (Cold) the extent by which the mean weekly temperature is below the mean temperature of the 29 years. Thus in the year 1848 the number of deaths registered in the 25th week was 516; the average weekly number for the 10 years was 1026 the deaths were therefore 212 below the average, which difference is shown by the width of the yellow colour. The mean temperature for the same week was 68°3, while the average mean for the 29 years was only 48°4, so that the mean temperature of that week was 20° above the average mean; which excess is shown by the width of the red colour.

[Farr 1852]

Overview

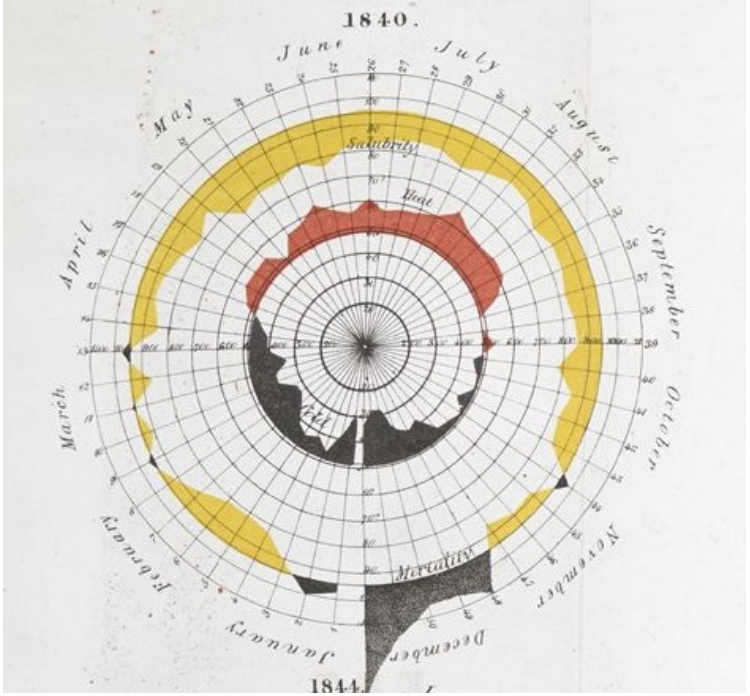
Based on the title the chart depicts temperature and mortality across several weeks over an 11-year period

Based on a coarse visual scan

- The visualization appears to use a **multi-faceted idiom**; sometimes called "**small multiples**" visualizations, comprised of several instances of some form of **radial area chart**
- Each facet is also comprised of two charts **elided** (or overlapped) into one, noting from the two different axis scales for the radial axes, namely temperature and deaths

Based on the captions and labels..

- Each facet (subplot) represents a year of data, the 12th chart depicts the average of the 11 years
- Weeks numbered 1-52 are arranged by angular position (clock-wise from bottom); Month labels are also used, grouping relevant weeks of the year
- There are 2 area plots in each facet. One of the plots (the outer plot) is labeled mortality (for positive values) and salubrity (negative); and similarly heat/cold for the inner plot.



Data

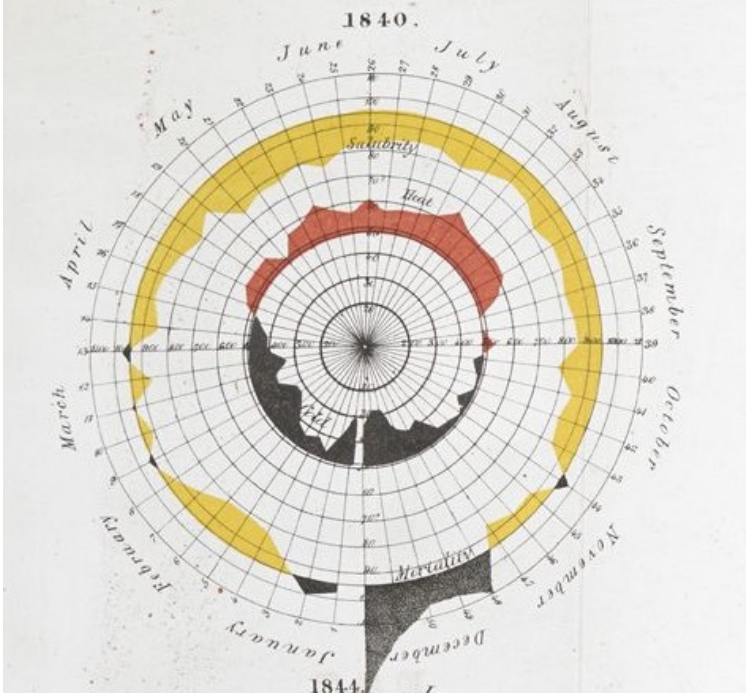
Most likely a Table dataset with..,

- 2 **quantitative, measurement** attributes:
  - Deaths: **quantitative, sequential, measurement**
  - Temperature (in degrees): **quantitative, diverging, interval**

- 2 **derived quantitative** attributes (same types corresponding to above):
  - Average deaths for the year:
  - Average temperature for the year

- 2 **derived categorical** attributes:
  - deaths being above (mortality) or below average (salubrity)
  - temperature being hot/cold relative to average

- **Ordinal cyclic attribute** of date over 12 years decomposed into weeks (key attribute), month and year)

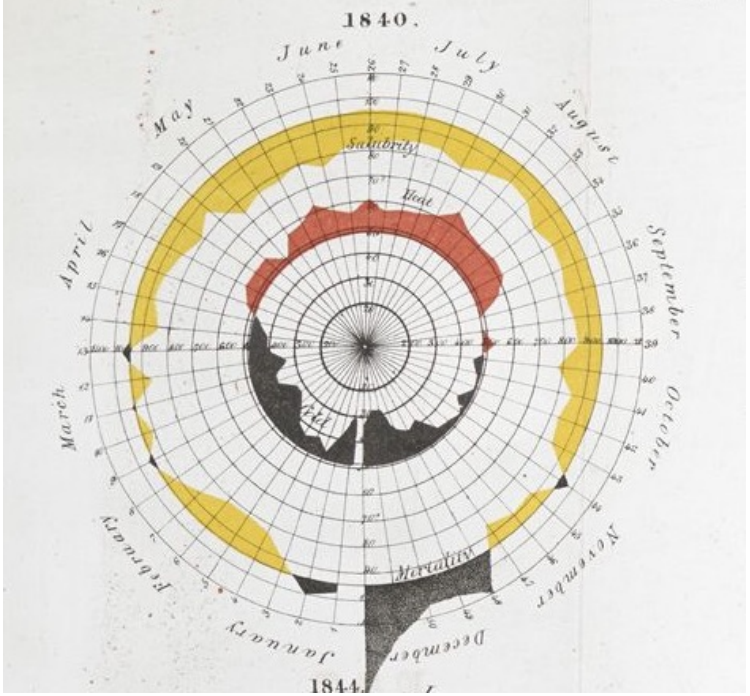


Encoding(s)

- **Position** (left-to right then top-bottom) is used to arrange the individual facets by **year**
- **Angular position** in the individual facets is used to **arrange** the categorical attribute week/month
- **Radial position** in the individual facets is used to encode the average and actual values of deaths and temperature
- At the same time, **radial size** relative to the average of each circular area chart appears to depict quantitative attributes of mortality and temperature
- **Colour** is used for categorical encoding of positive and negative values for deaths and temperature (into hot/cold, mortality/salubrity). This could be said to be a redundant encoding as we can already tell from the line plot if a value is positive/negative.

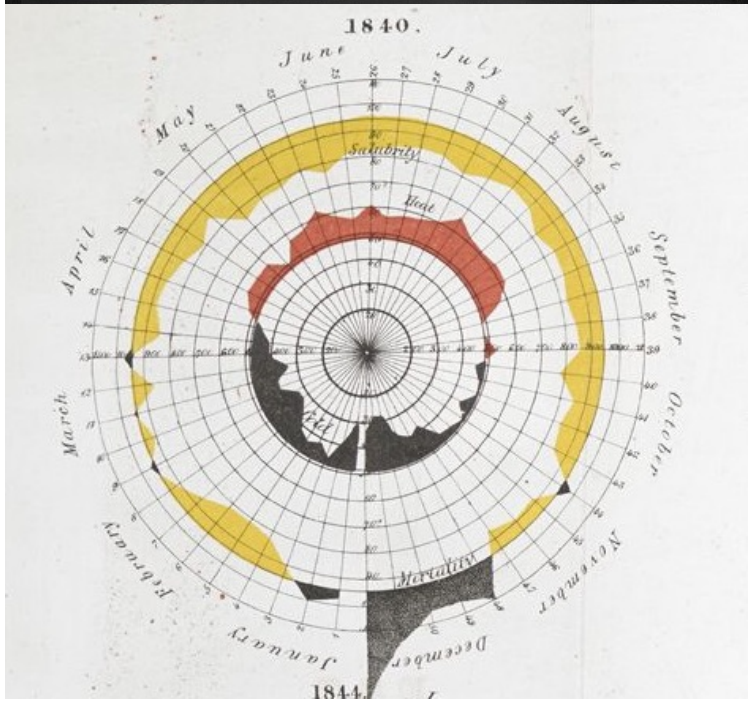
  SIDE NOTE: in Farr's caption the "Cold" values are Blue. The colour has apparently been lost over time and this now appears black.

- At a high level, the facets could be seen as Glyphs (using shape encoding) to characterize each year and compare similarity and extreme outlier amongst the years



Task(S)

- Query or Compare values of mortality and temperature at particular dates.
- Identify trends in mortality and temperature
- Search, or more specifically browse for the existence of (or explore the degree of) Correlation between values of the 2 quantitative attributes temperature and mortality
- Compare patterns of mortality and temperature across 11 years



Qualitative Analysis

Circular plots such as this are problematic due to relative difficulty of comparing areas and angles

- Areas increase outwards
- Unaligned axes lead to slight difficulty in comparison of values across different months

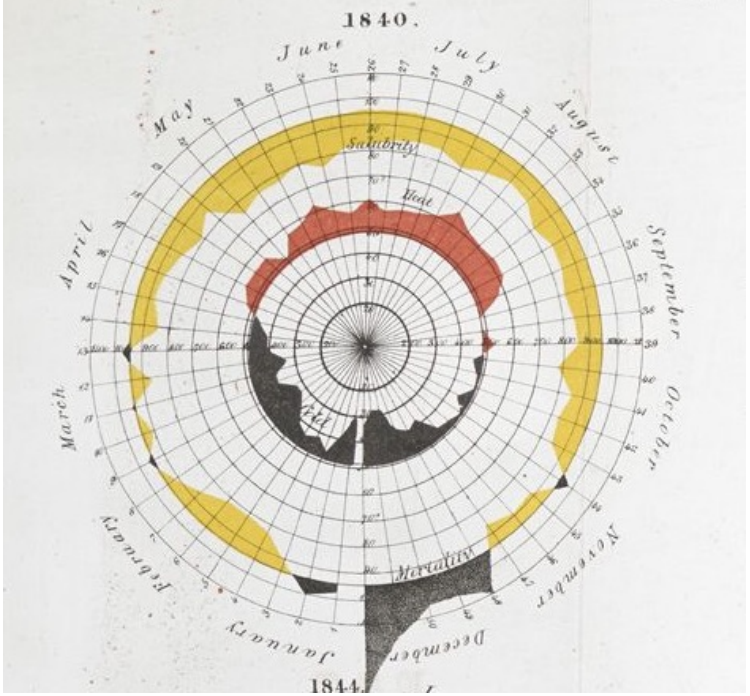
Some of these issues are ameliorated here because…

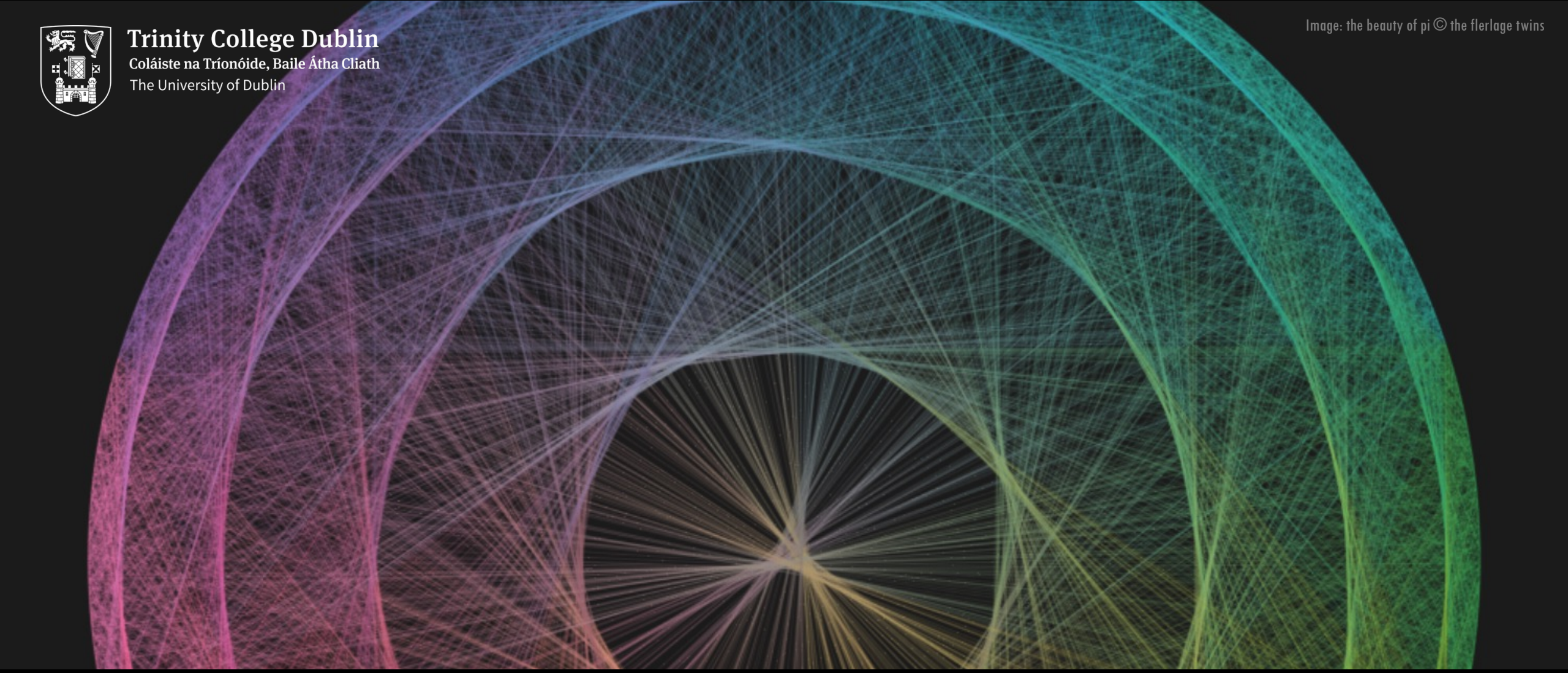
- main attributes are encoded radially i.e. linear positional encoding along each axes, which makes querying individual weeks fairly effective
- redundant encoding by colour aids high level comparison across weeks
- reference lines aid lookup of values, but cause some clutter.

The visualization is reasonably effective for detecting trends and correlations,

The visualization is probably less effective than a simple table for querying individual values. It is likely that a simple rectilinear line/area chart would be more effective on a per year basis;

Radial facets for each year make sense due to the cyclic nature of the key attribute (week/month). Also, at a high level we can quickly see similarities, patterns (or the lack thereof) across years





Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Image: the beauty of pi © the flerlage twins

Visualization Design

Recap of key concepts

Stages of Visualizing Data

Below, based on Nussbaumer [2015] are suggested steps for **explanatory** visualization:

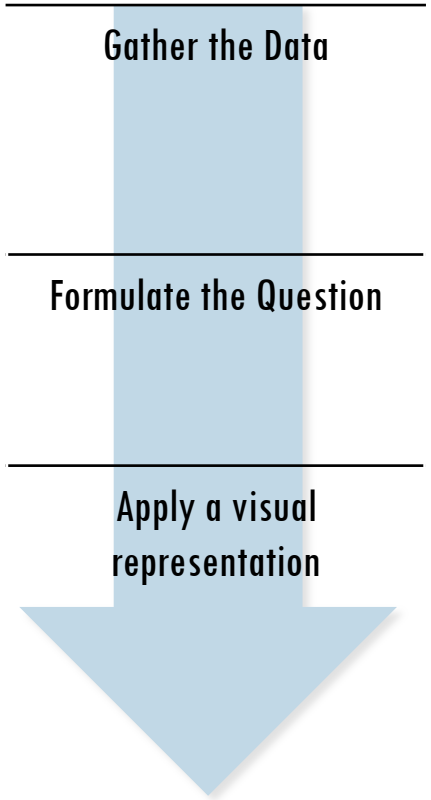
Understand the context	• Identify audience, what they need to know or do and why
Choose an effective visual	• decide how best to encode data
Eliminate clutter	• Remove anything not adding informative value
Focus attention	• Leverage pre-attentive perception to draw users attention to important parts
Tell a story	• Use the visual as a foundation to walk the audience through the analysis

[Nussbaumer 2015] Cole Nussbaumer, Storytelling with Data, John Wiley & Sons, 2015

Stages of Visualizing Data

Based loosely on [Fry 2008] an approach that might be followed for **exploratory** visualization:

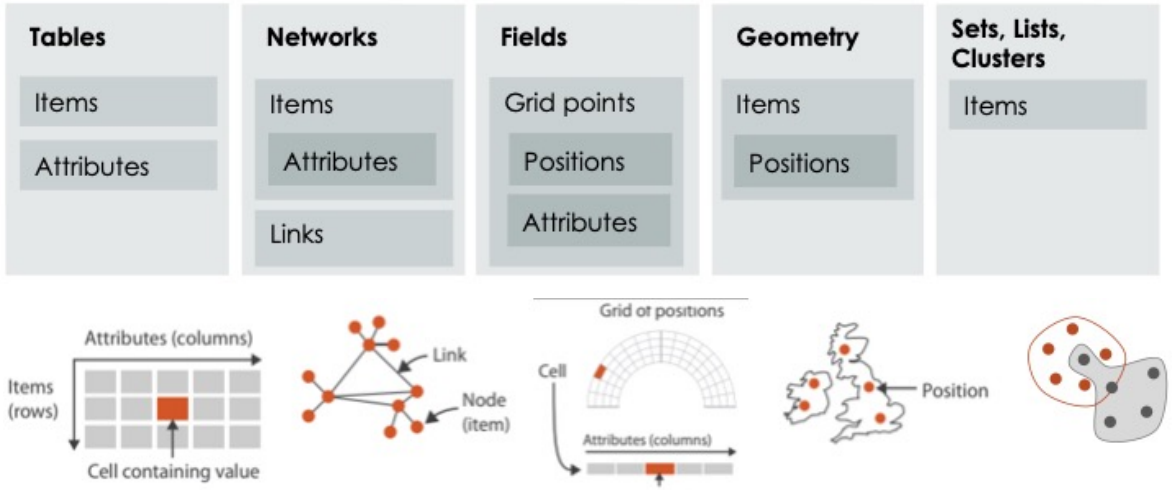
Acquire	Obtain the data
Parse	Provide some structure (attributes+data types)
Filter	Remove extraneous data
Mine	Apply math, stats & data mining to discern patterns
Represent	Choose idioms + encoding channels
Refine	Improve efficacy of representation; transform
Interact	Add methods for user to manipulate/control visuals



[Fry 2008] Ben Fry, Visualizing Data, O'Reilly Press, 2008

1: Data

Data/Dataset Types



Attribute Types

Categorical

Ordinal

Quantiative

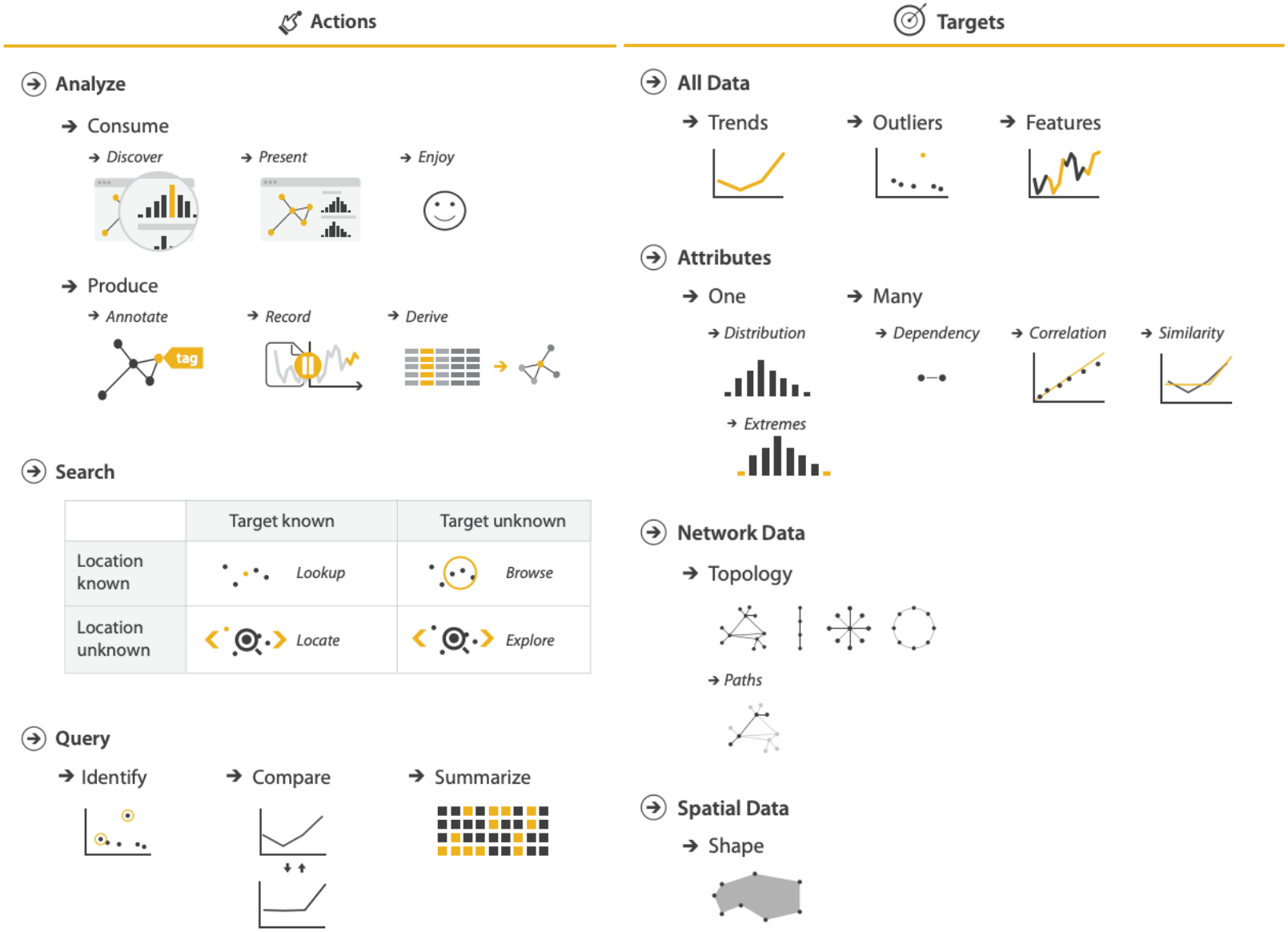
RECOMMENDED READING

- CS7DS4 Lecture 5 Data and Task Abstractions
- Chapter 2 – "Data Abstraction" in Visualization Analysis and Design, T. Munzner 2014
- [Free Alternative]: Visualization Design and Analysis: Abstractions, Principles, and Methods (DRAFT). T. Munzner, 2012. https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=60

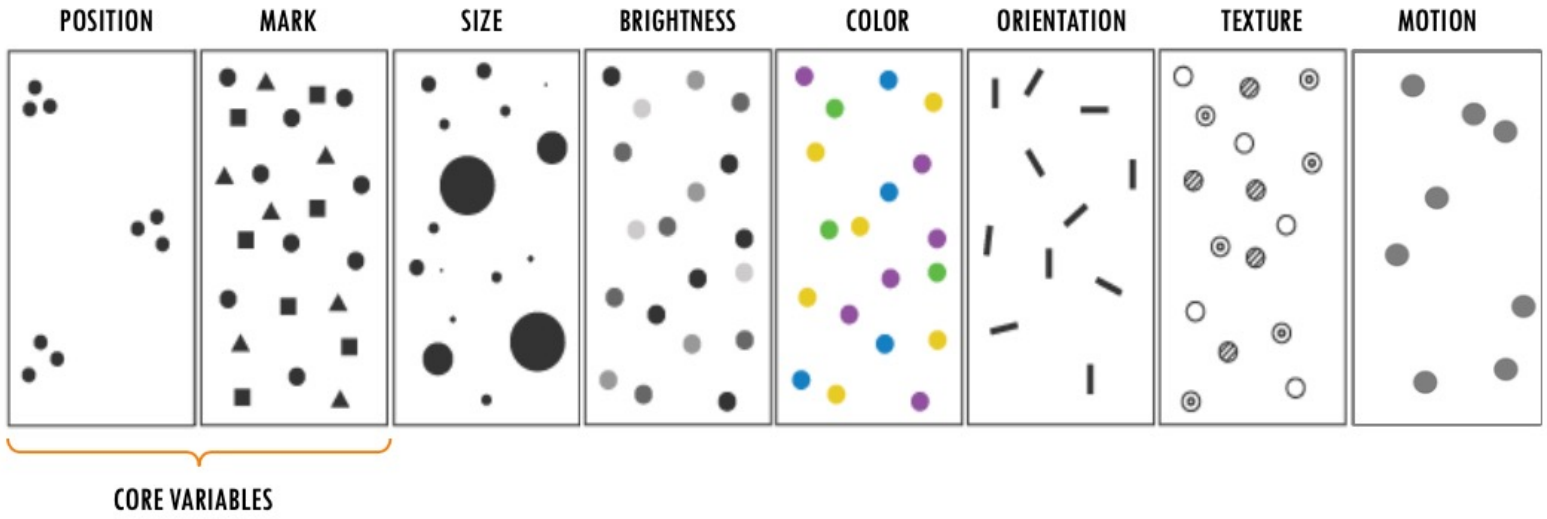
2: task

RECOMMENDED READING

- CS7DS4 Lecture 5 : Data and Task Abstractions
- Chapter 3 – "Task Abstraction" in **Visualization Analysis and Design**, Tamara Munzner 2014 [Available as e-book in Library Reading Rooms]
- [Free Alternative]: **A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner**, Infoviz, 2013. https://www.cs.ubc.ca/labs/imager/tr/2013/MultiLevelTaskTypology/brehmer_infovis13.pdf



3: Encoding Channels



Arrange

→ Express → Separate

→ Order → Align

→ Use

Map
from categorical and ordered attributes

→ Color

→ *Hue* → *Saturation* → *Luminance*

→ Size, Angle, Curvature, ...

→ Shape

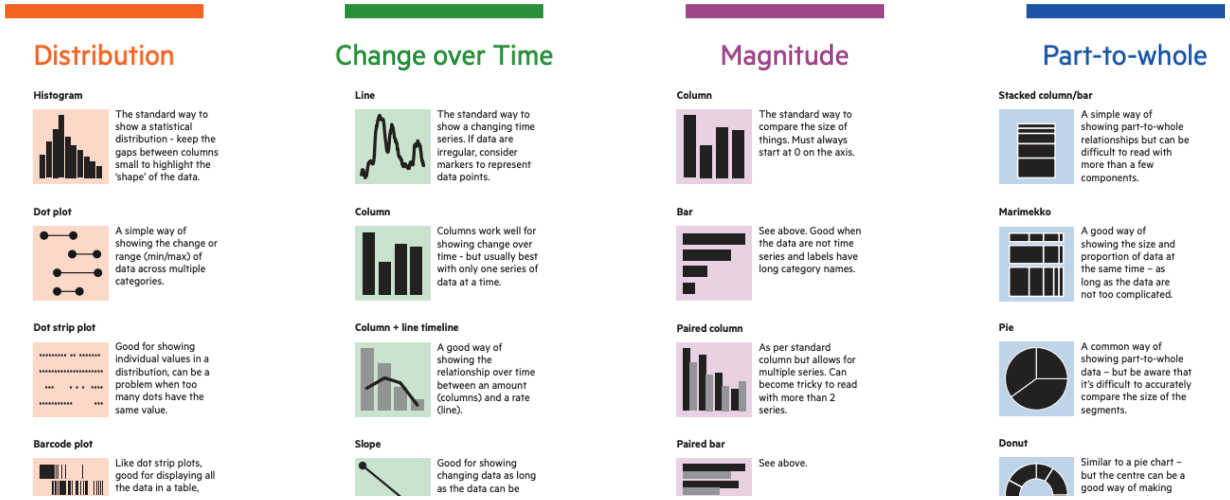
→ Motion
Direction, Rate, Frequency, ...

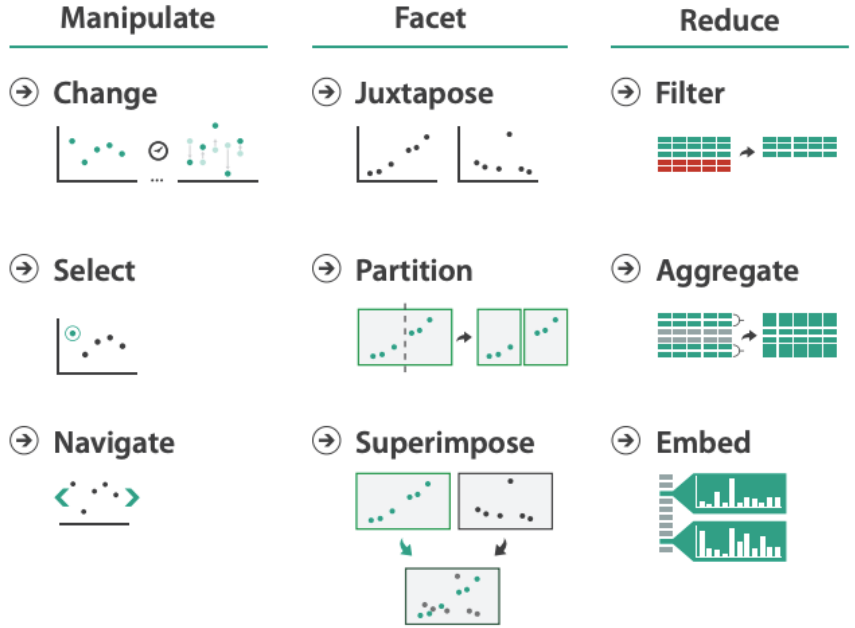
RECOMMMENDED READING

- CS7DS4 Lecture 4 : Visual Encoding
- Chapter 4 – "Visualization Foundations" in Interactive Data Visualization, Ward et al 2010
- [Free Alternative] Chapter 4 – "Visual Encoding Principles" in Information Visualization: Principles, Techniques and Practice, [Draft] Munzner 2011

  https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=56

4: Idioms

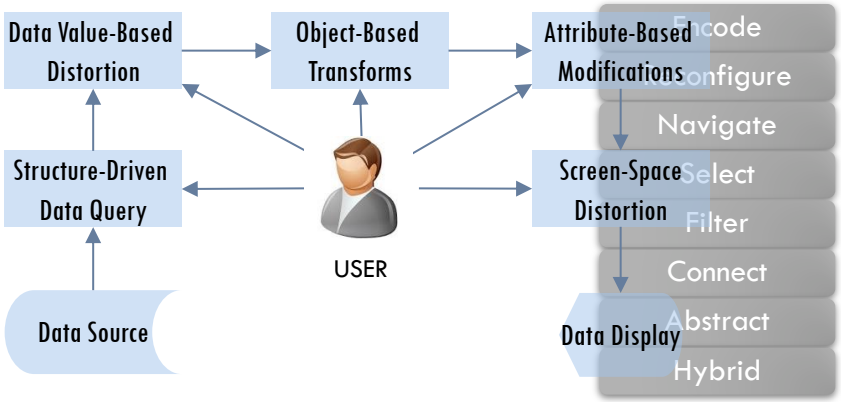




Too many to list. See something like the above cheat sheet by the Financial Times:
https://github.com/ft-interactive/chart-doctor/raw/master/visual-vocabulary/Visual-vocabulary.pdf

RECOMMMENDED READING

- CS7DS4 Lecture 7,8 9,10
- [Idioms] Chapter 8 – "Making Views" in Visualization Design and Analysis: Abstractions, Principles, and Methods (DRAFT). Tamara Munzner, 2012.
  https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=142
- [Transformations] Sec 8.4-8.6 and Ch 9. in Visualization Design and Analysis: Abstractions, Principles, and Methods [Draft]
  https://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/ClassLectures/PDF-old/book.120803.pdf#page=187
- [Interaction] Interactive Data Visualization. Ward et al (2012). Ch 10, Ch 11. https://goo.gl/FyioEK



Data Integrity (Continued)

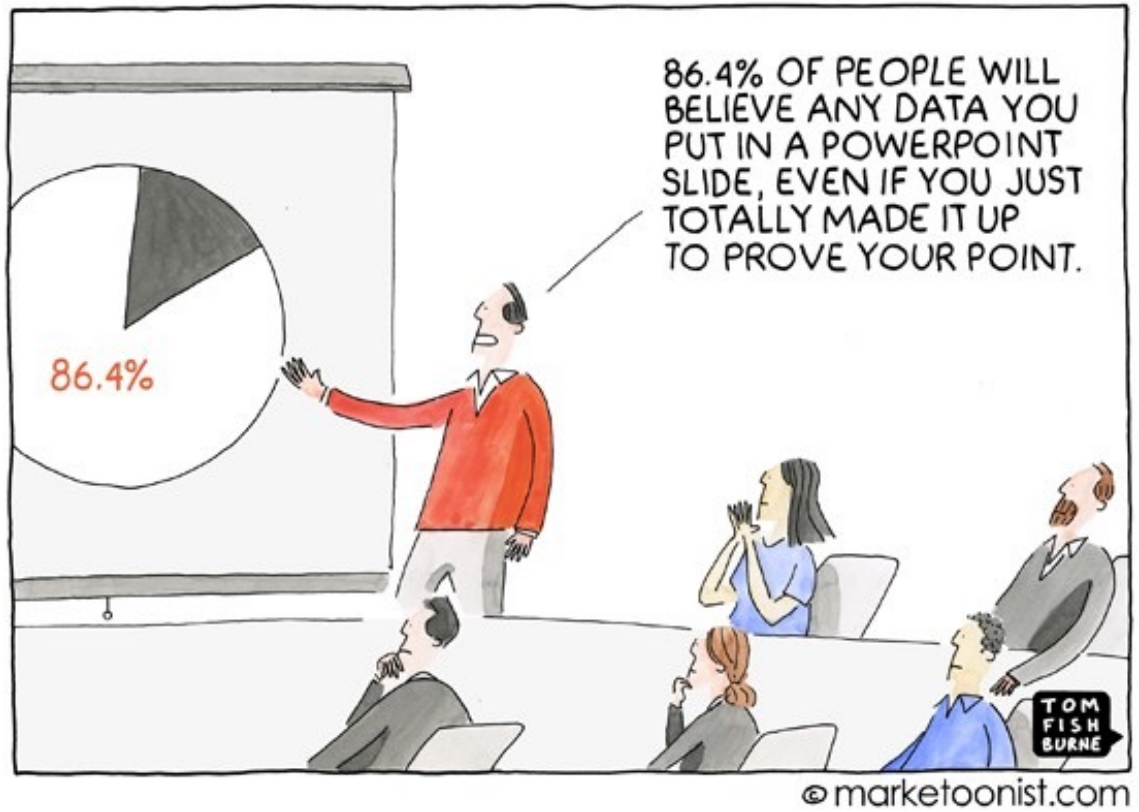
Principles for Effective Visualization

Address a clear objective	function over form; show the data; expressiveness/effective
Avoid distortion	lie factor; dangers of scrubbing, deriving, sampling, high dimensionality
Address Complexity	maximize data density; parallelism; reduce clutter focus and attention, change blindness

REFERENCES: the Visual Display of Quantitative Information. Ch1 & 2. Edward Tufte. 1983.
[Externally hosted PDF Excerpt here: https://tinyurl.com/uhaxq7k]

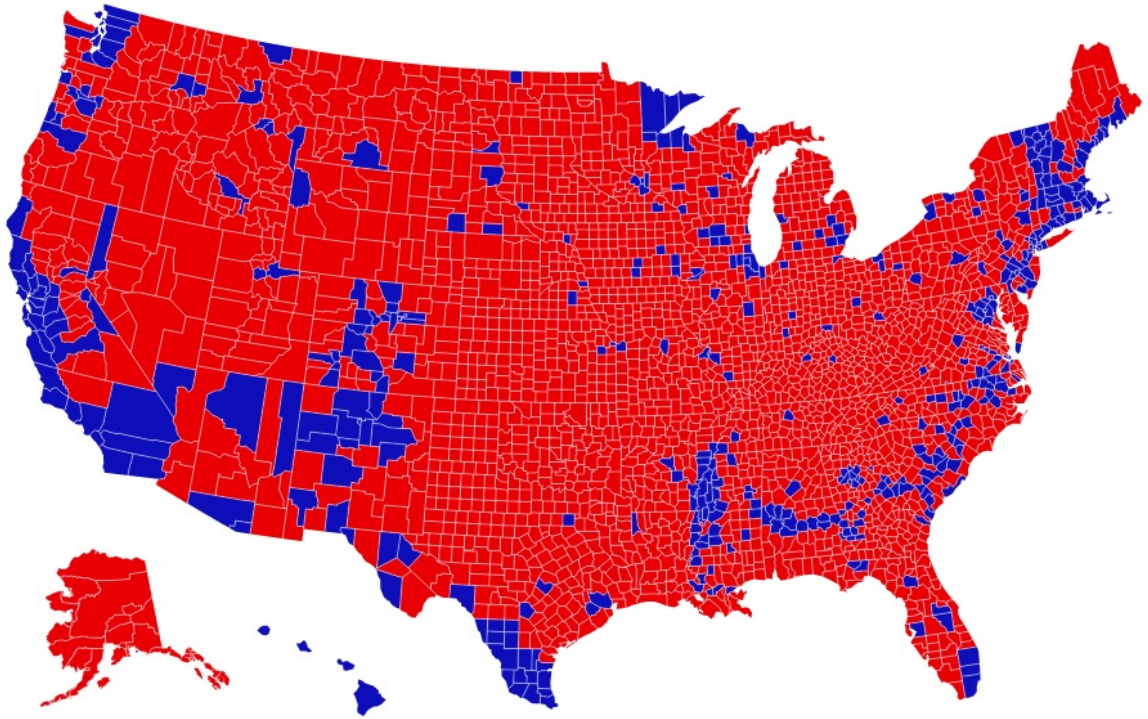
Principles of Data integrity [Tufte '83]

- Representation of numbers should **match the true proportions**.
- **Labelling** should be clear and detailed.
- Show **data variation, not design variation**.
- The number of information-carrying **dimensions depicted should not exceed the number of dimensions in the data**
- Graphics must not quote data out of context



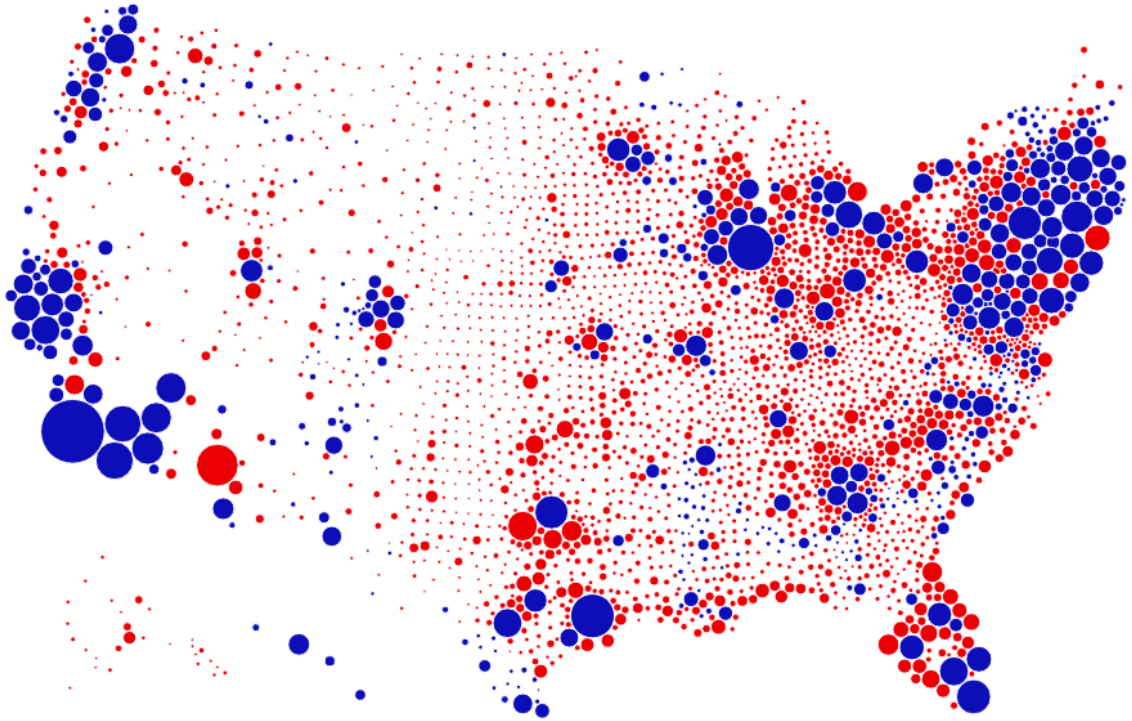
From: E. Tufte. The Visual Display of Quantitative Information (1983)

Distortions in Perception



Binary Choropleth Map

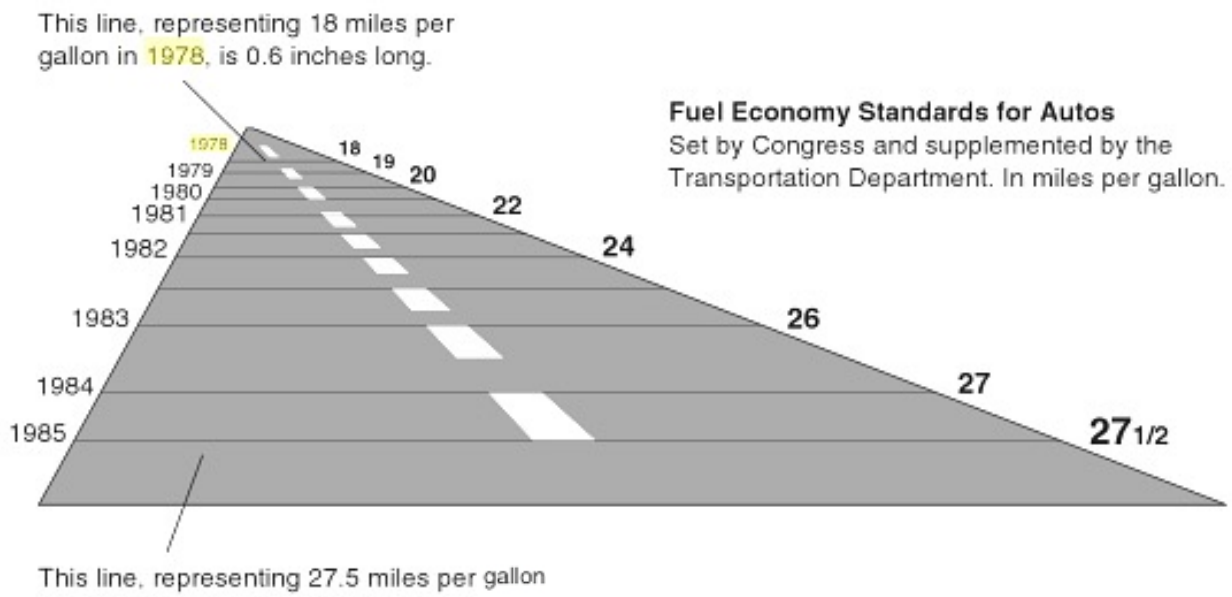
2016 US election result by county © HuffPost

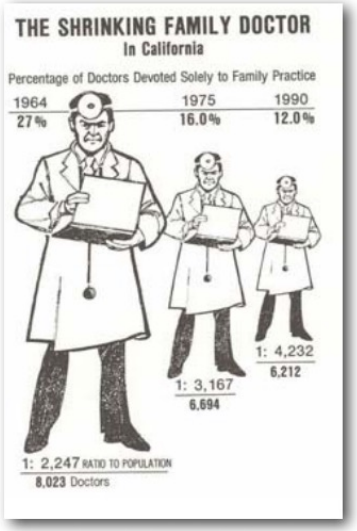


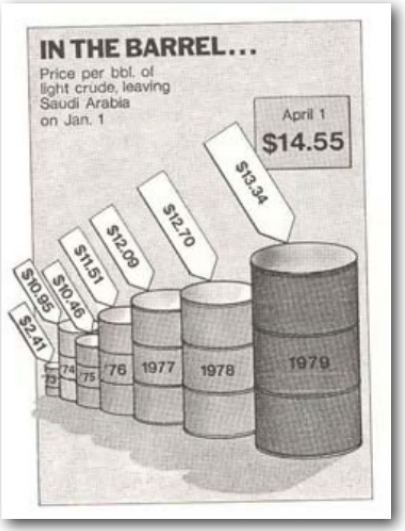
Cartogram/bubble chart

with area encoding of population
© karim douieb

Assessing Distortion : Lie Factor [Tufte 83]







$$Lie\ factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data}$$

Lie factor: 2.8

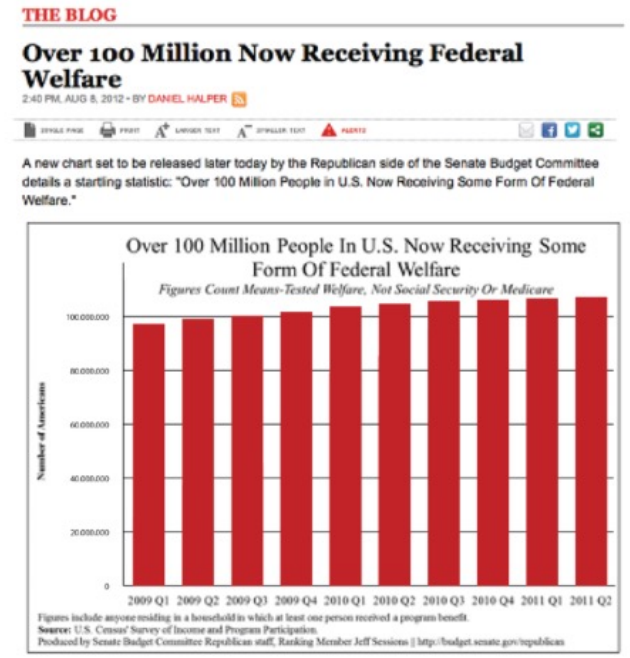
Lie factor: 9.5

For the above example: $Lie\ factor = \frac{5.3-0.6}{0.6} / \frac{27.5-18}{18} = 14.8$

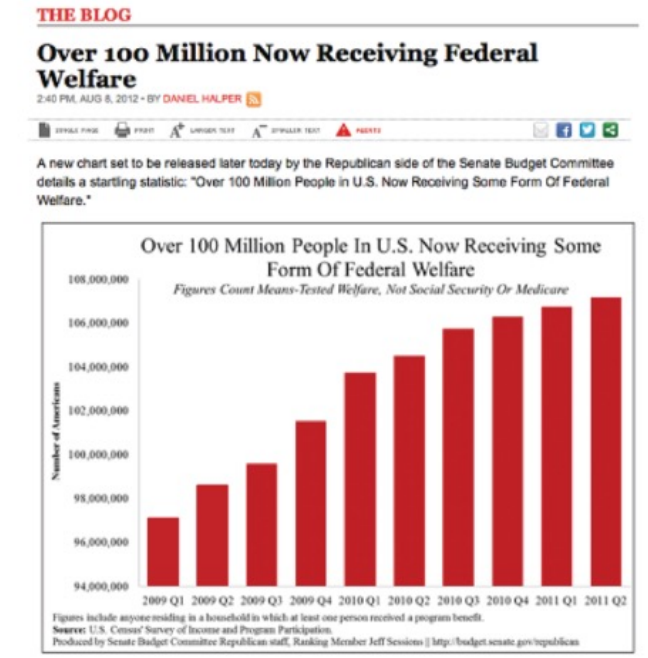
Range Distortion

Using unexpected data ranges may mislead readers.

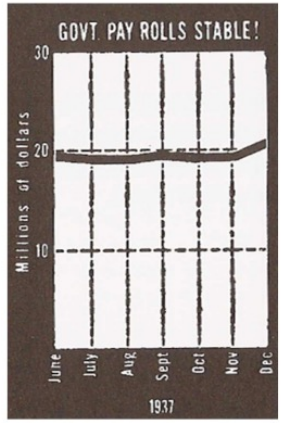
- Relative judgement is perceptually strong component of visual system.
- Need to balance with reducing screen space usage.

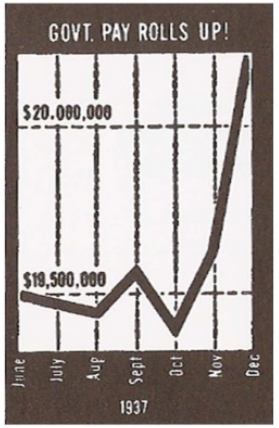


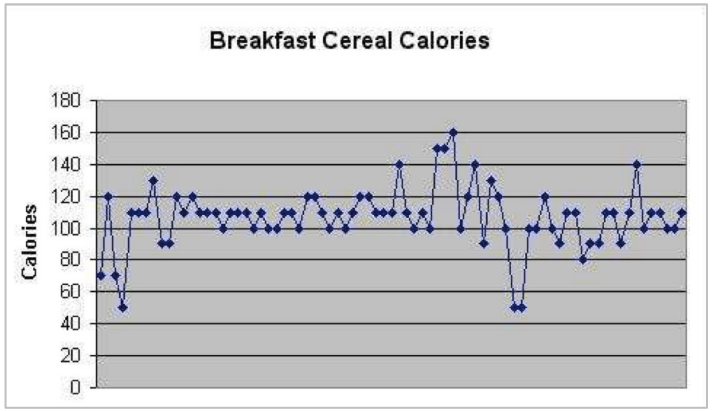
lie factor: 1

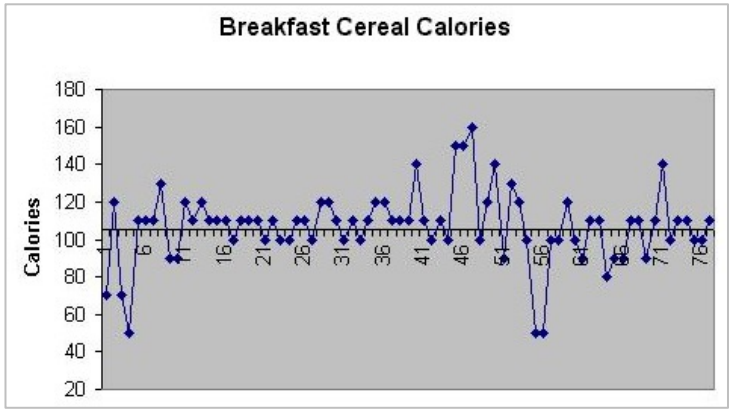


lie factor: 16,08



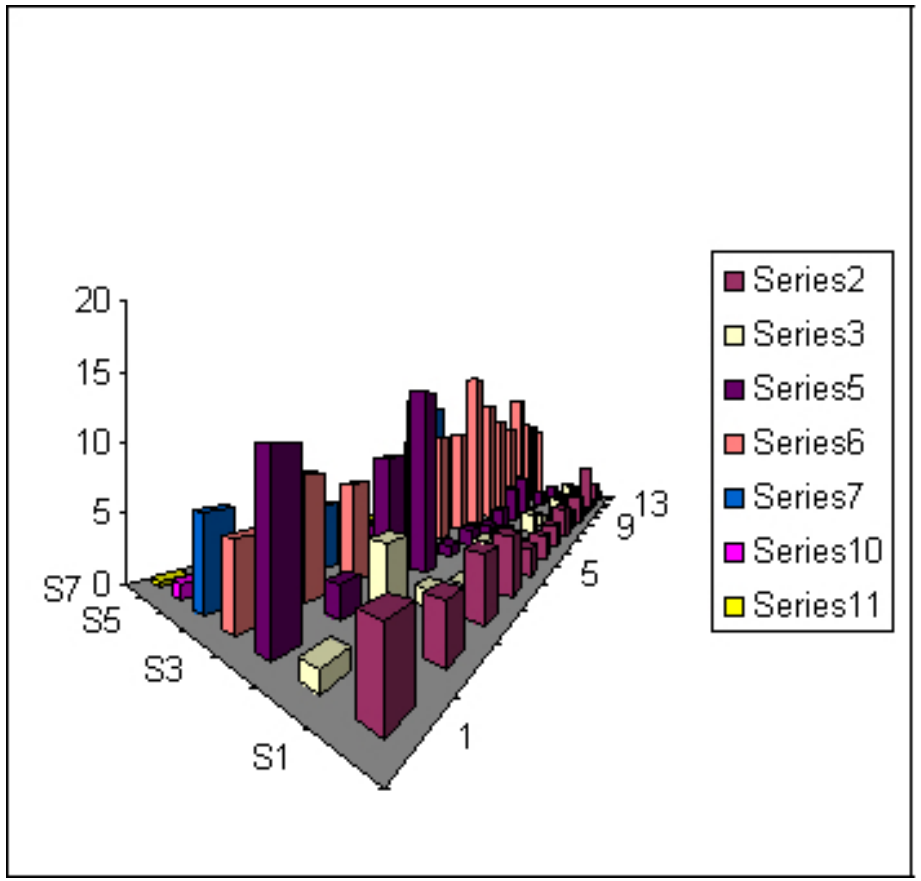


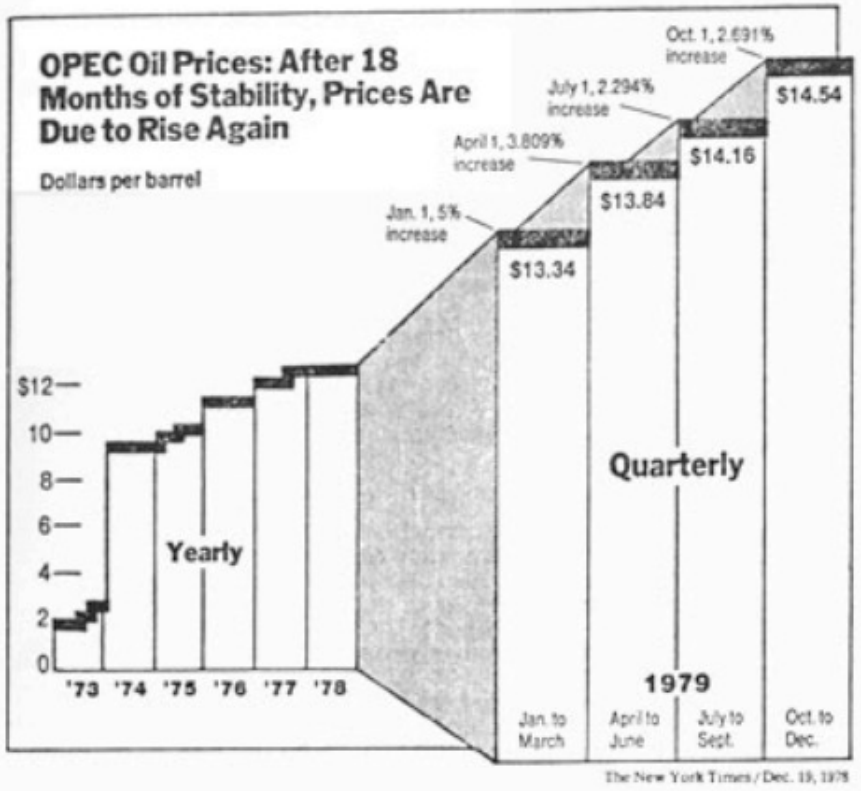




Unbalanced Scaling

Avoid distortion of numbers by graphic devices and design variations





Five different vertical scales show the price:

During this time	one vertical inch equals
1973–1978	$8.00
January–March 1979	$4.73
April–June 1979	$4.37
July–September 1979	$4.16
October–December 1979	$3.92

And two different horizontal scales show the passage of time:

During this time	one horizontal inch equals
1973–1978	3.8 years
1979	0.57 years

Data Scrubbing [ward 2014]

Raw data can be rough

But removing data potentially creates bias

e.g. outlier removal

- Avoid unless proven that data results from flaws in acquisition process
- Always inform reader of any scrubbing/smoothing

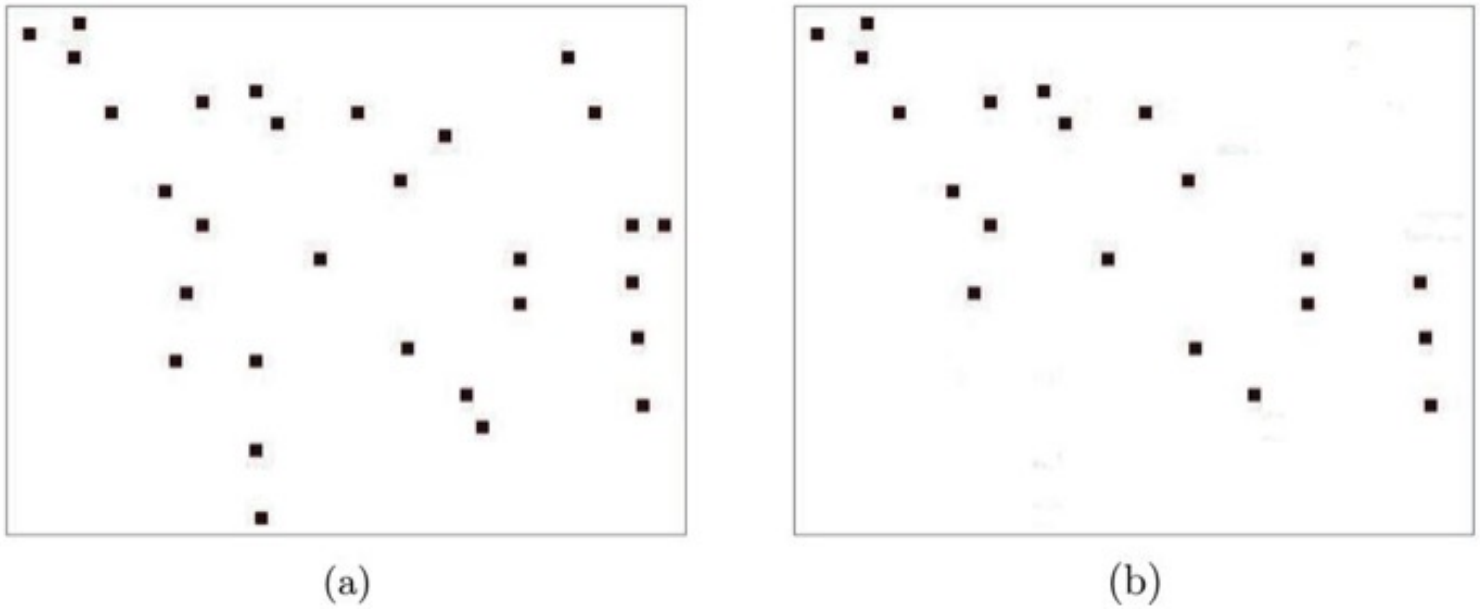
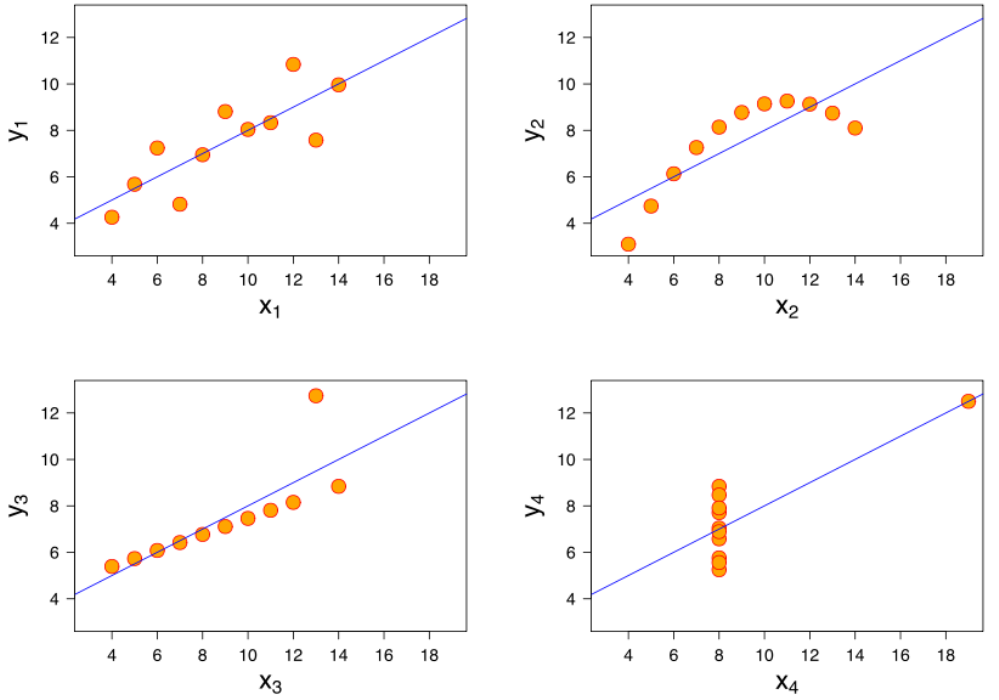


Figure 13.15. The problem with data scrubbing: (a) raw data showing lack of correlation; (b) scrubbed data revealing false correlation.

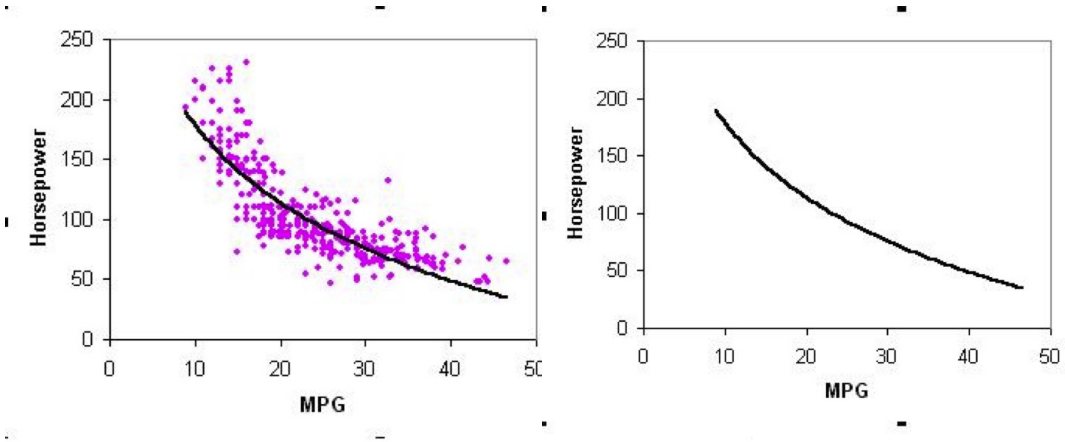
Raw vs. Derived Data

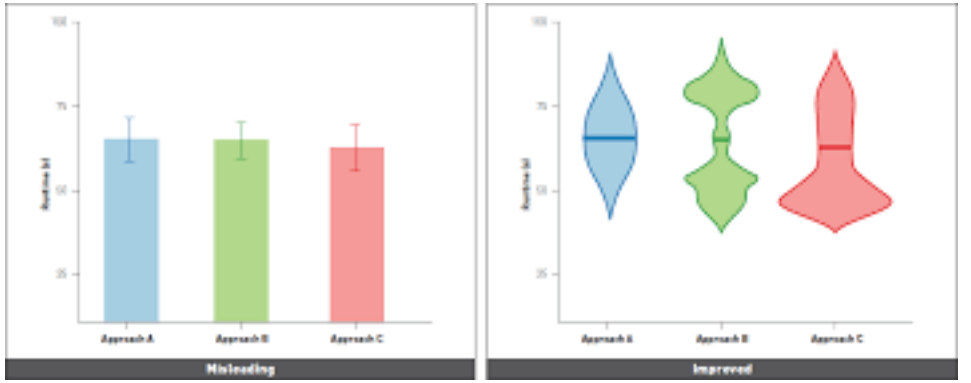
Anscombe (1973)



Showing only derived attributes can mislead readers

- Derived data is important in analysis e.g. fitting data to curve/surface
- Ideally show both raw and fitted model and allow filtering by the user
- Or encoding channels that indicate the presence/nature of the abstracted data





Insufficient Sampling

Uniform sampling

Contour info used to add sample points where significant changes occur

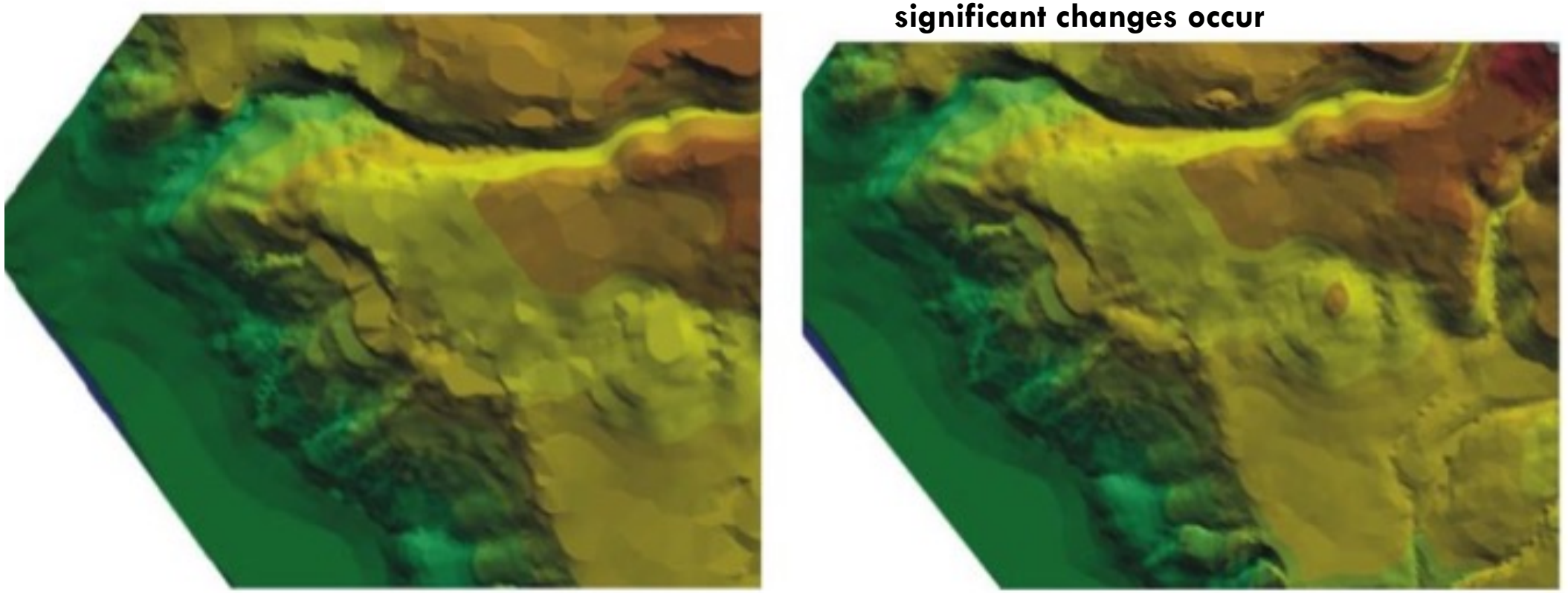


Figure 13.21. Different sampling and interpolation of the same data set. Some of the details in the right image are not seen in the left image.

Figure & caption from Ward 2014. Image © Hemphill 2008.

Non-uniform Sampling and Interpolation

Interpolation is common process to address sparsity of data

- However Readers may believe dataset is larger than it is
  - Potential solution: use cues to distinguish real sample points
- If data has high variability, interpolated values may lead to erroneous analysis
  - Avoid interpolation if localized judgements will be made

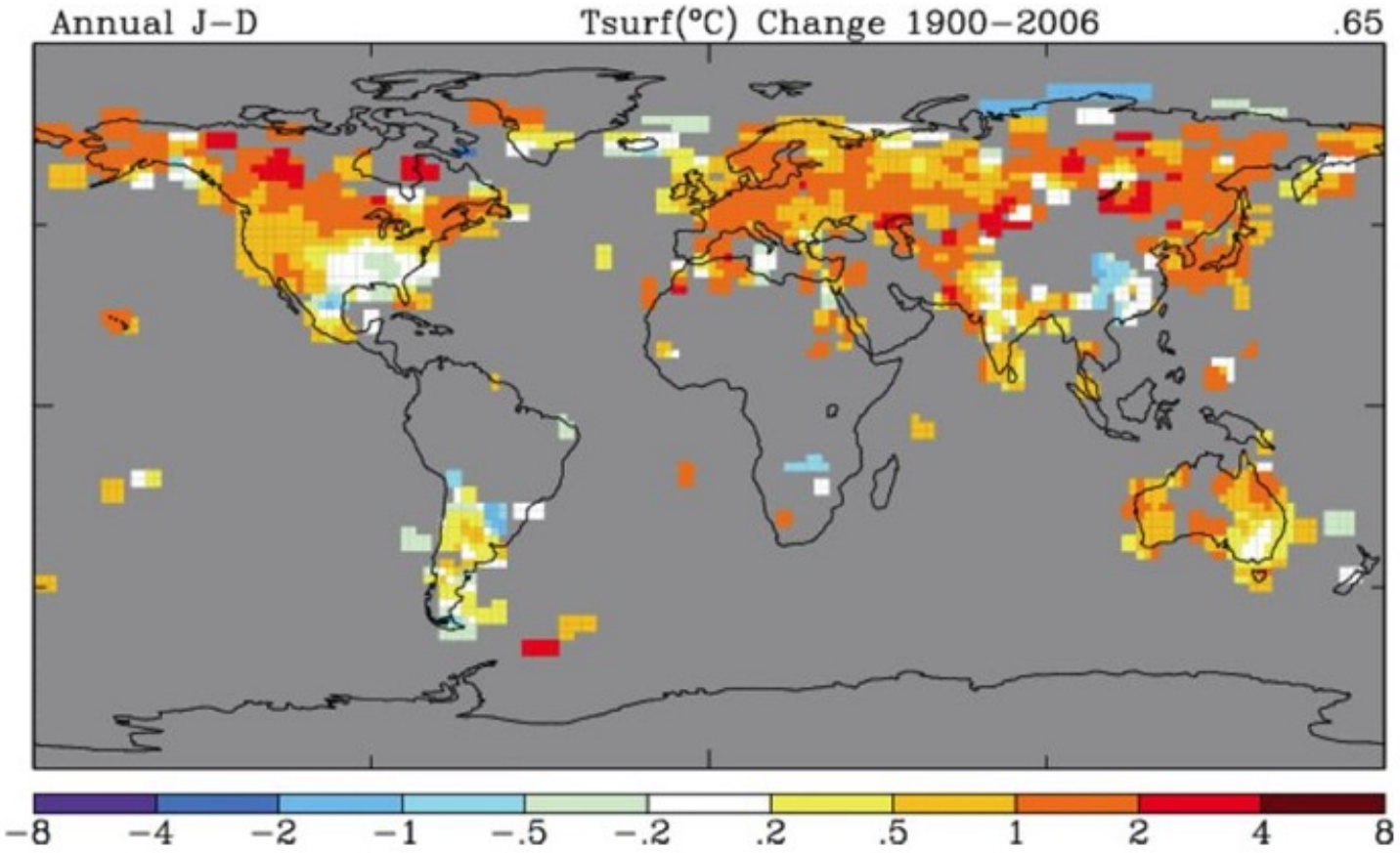


Figure 13.20. Sparse global temperature change data would give erroneous values for most of the planet if interpolated.

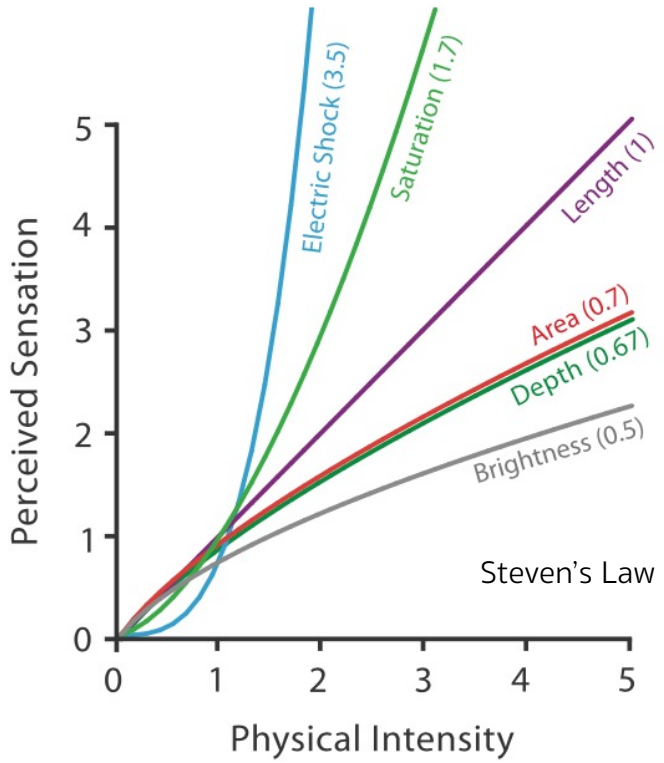
Figure & caption from Ward 2014. Image © Goddard Institute for Space Studies [http://data.giss.nasa.gov/gistemp/maps/]

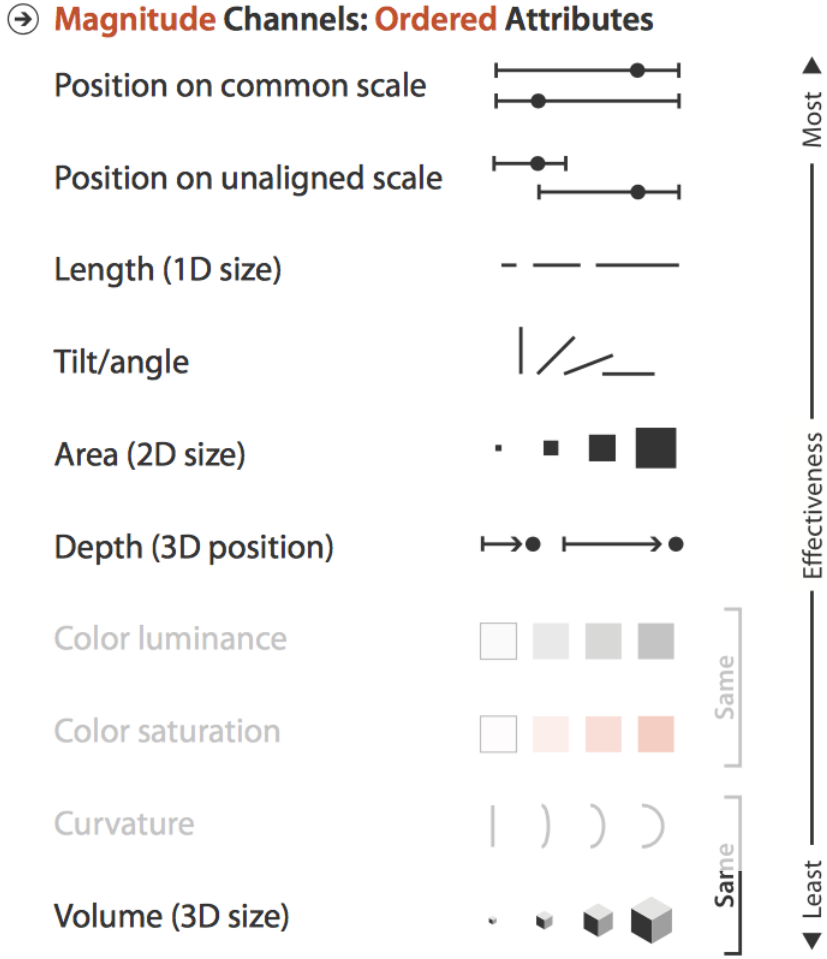
Avoid unjustified 3D

Figures from Munzner [2014]

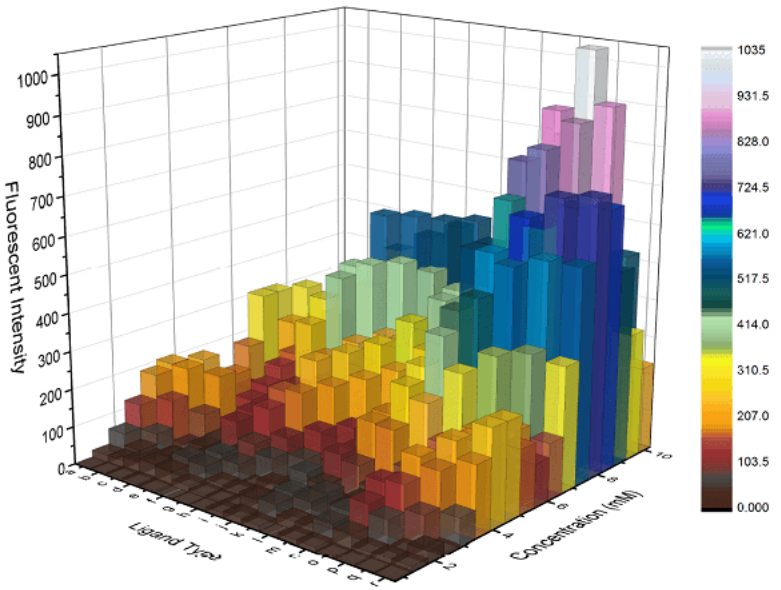
Spatial position/size channels are highly ranked:

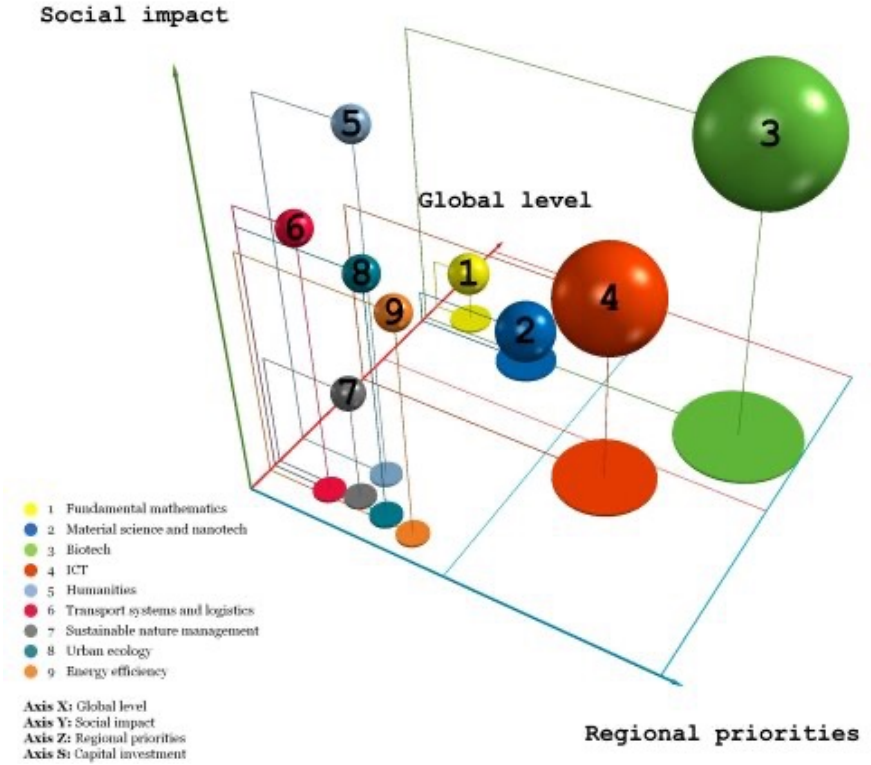
- But this pertains to planar (2D) spatial position
- Area, depth and volume perception is much less precise





Volume Encoding (3D Size)

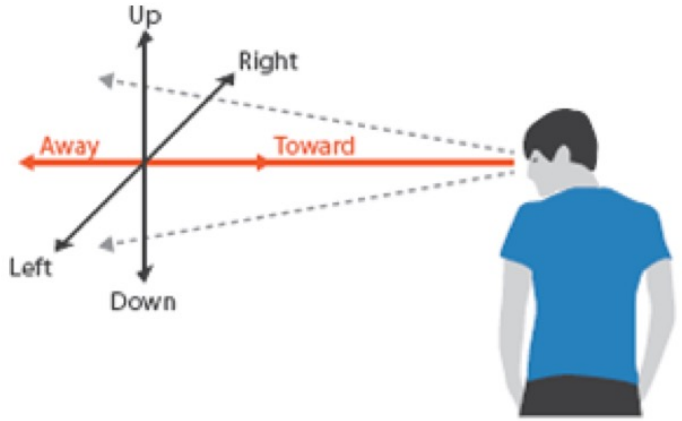




Idiom	3D Bar chart
Data	Table three quantitative attributes, one categorical attribute
Encode	3D columns marks positioned along X-Z plane based on 2 key attributes, third attribute encoded by Y-height, here also encoded redundantly by colour
Task	Compare/look up values, trends, find outliers/maxima

Idiom	3D Bubble chart
Data	Table, four quantitative attributes and one categorical attribute
Encode	Sphererical marks positioned in 3D according to3 key attributes, fourth attribute encoded by size, categorical attribute encode by colour
Task	Trends, distribution, compare values

3D Depth Ambiguity



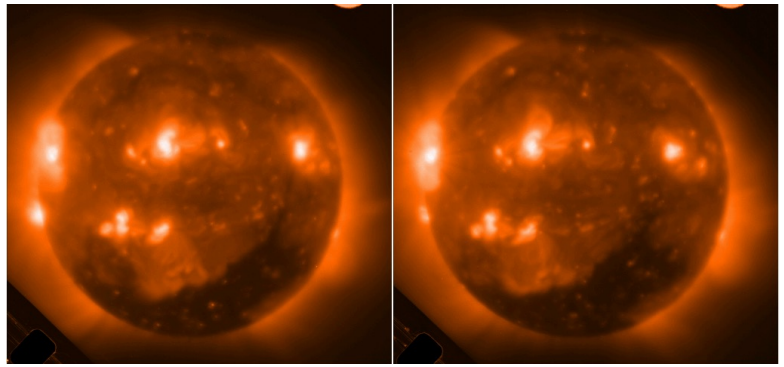
We don’t perceive in 3D: but in 2.05D [Ware 08]

In the real world we get depth information from various depth cues including parallax (movement) and binocular disparity (stereo)

But in conventional digital displays this sense is missing

In addition, stereo vision is

- Very effective at ~ 2m
- Moderately effective up to ~10m
- Minimally effective up to ~20 m

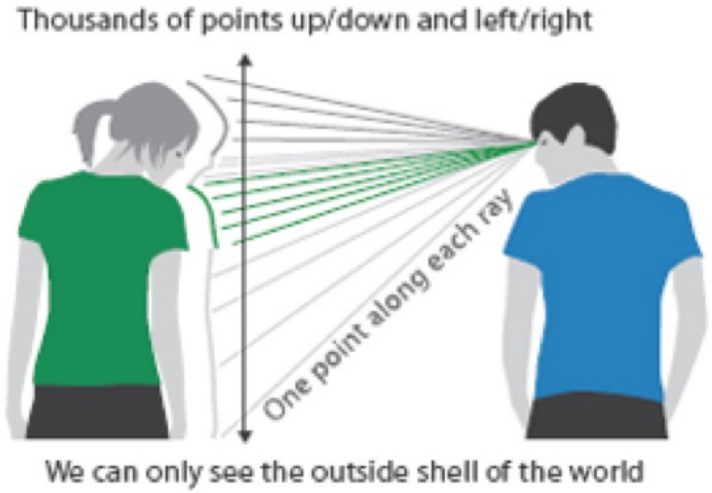


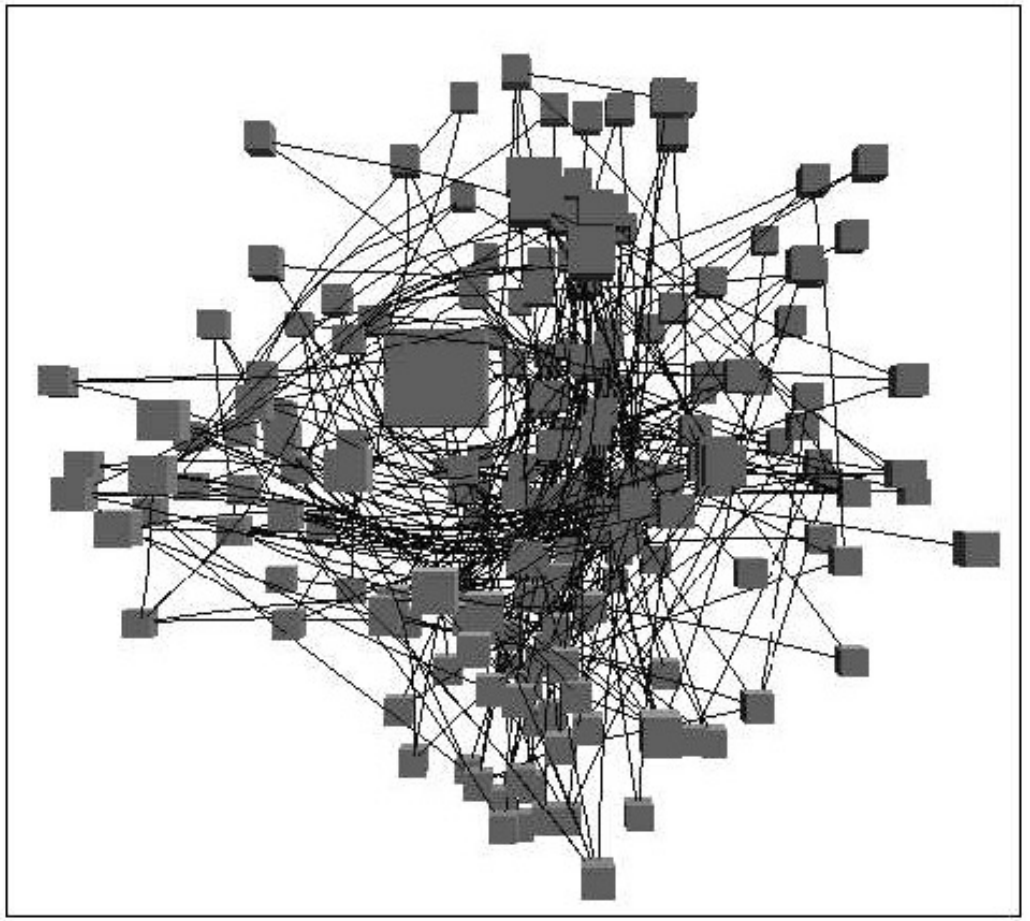
3D Occlusion

Occlusion is an important depth cue

However

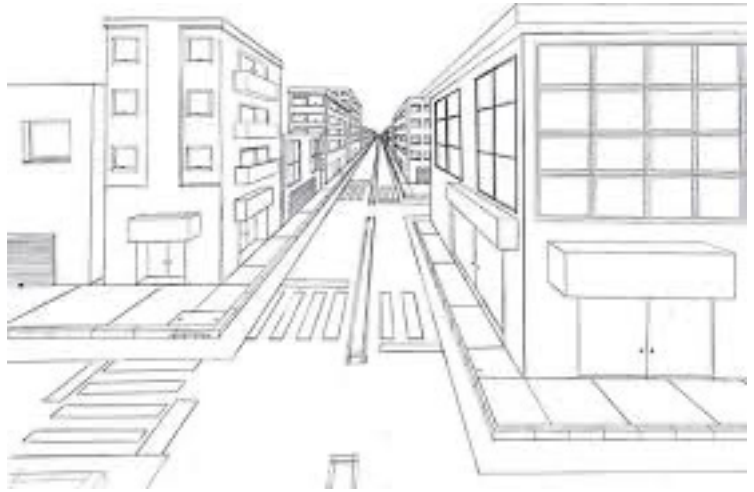
- It hides information in the background
- Leads to interaction complexity
  - e.g. the picking/selection problem in 3D





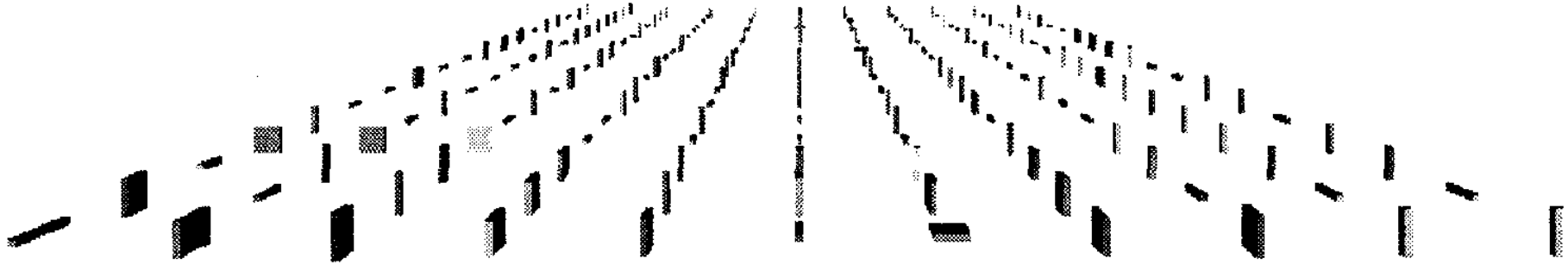
Carpendale 1996

3D Perspective distortion



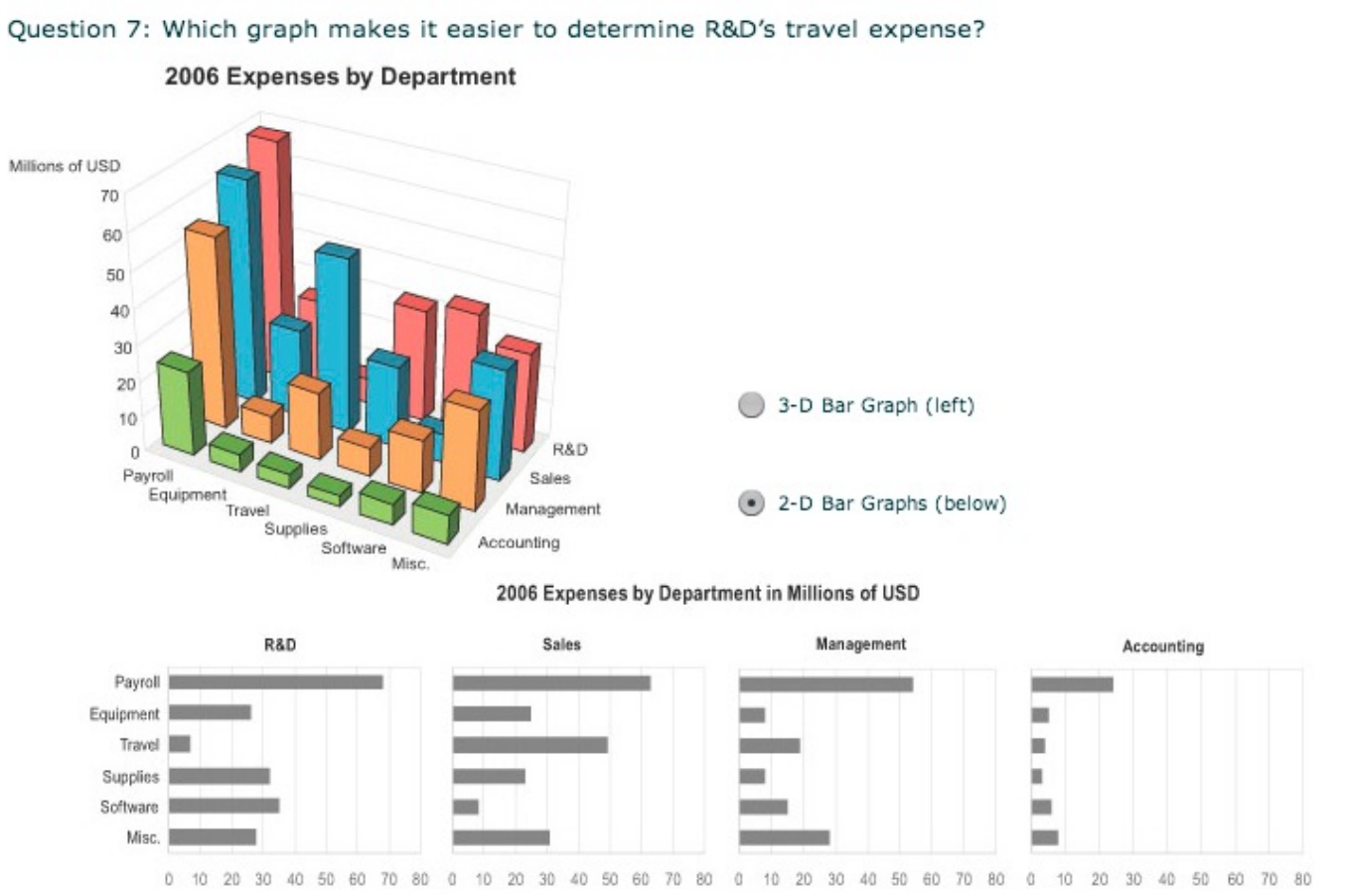
Perspective distortion

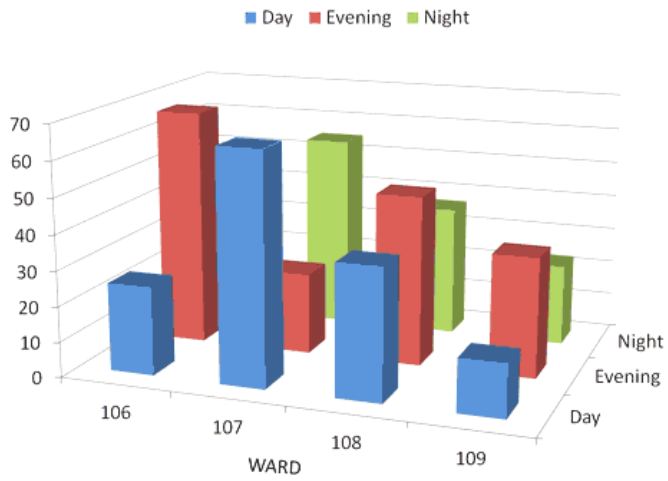
- interferes with all size channel encodings
- power of spatial position / size is lost



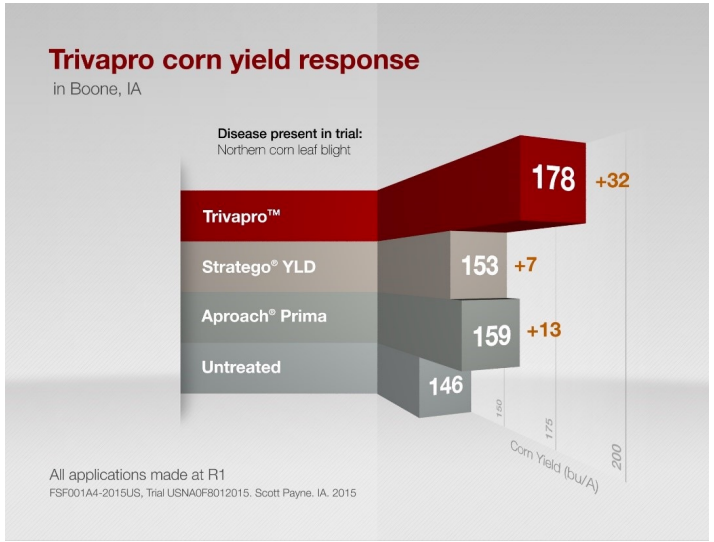
[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

Ineffective 3D Bar Charts

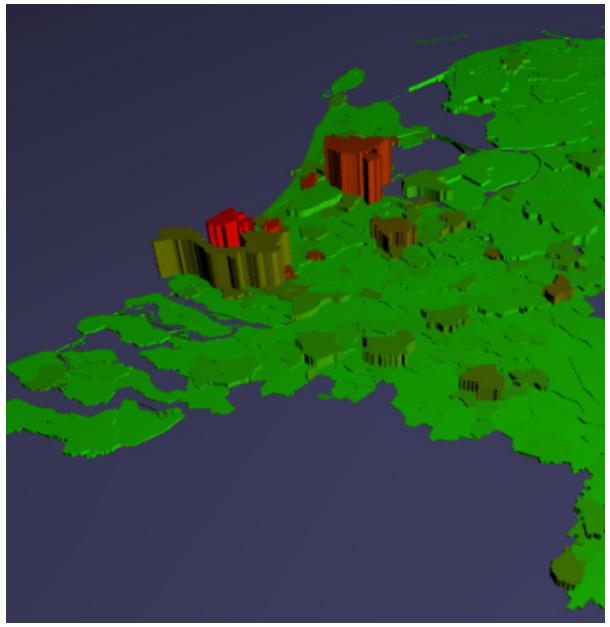




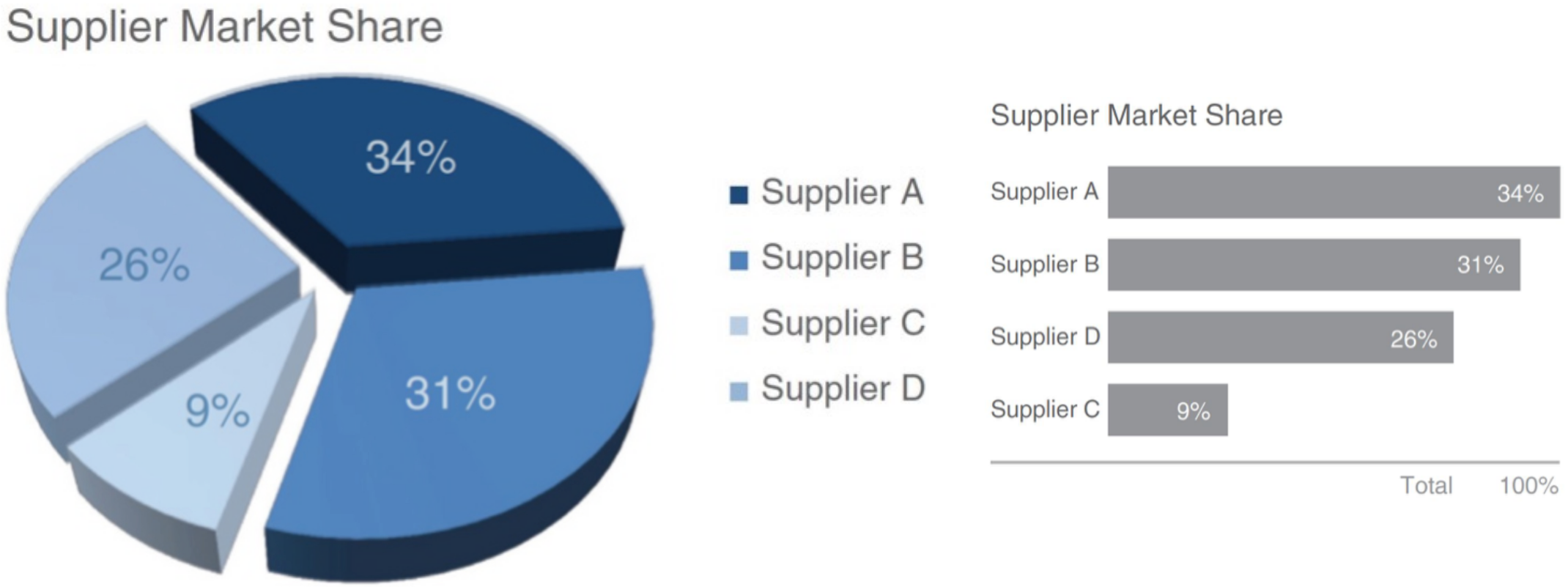
Is the second blue bar touching the 50 line?



Imprecise 3D volume perception



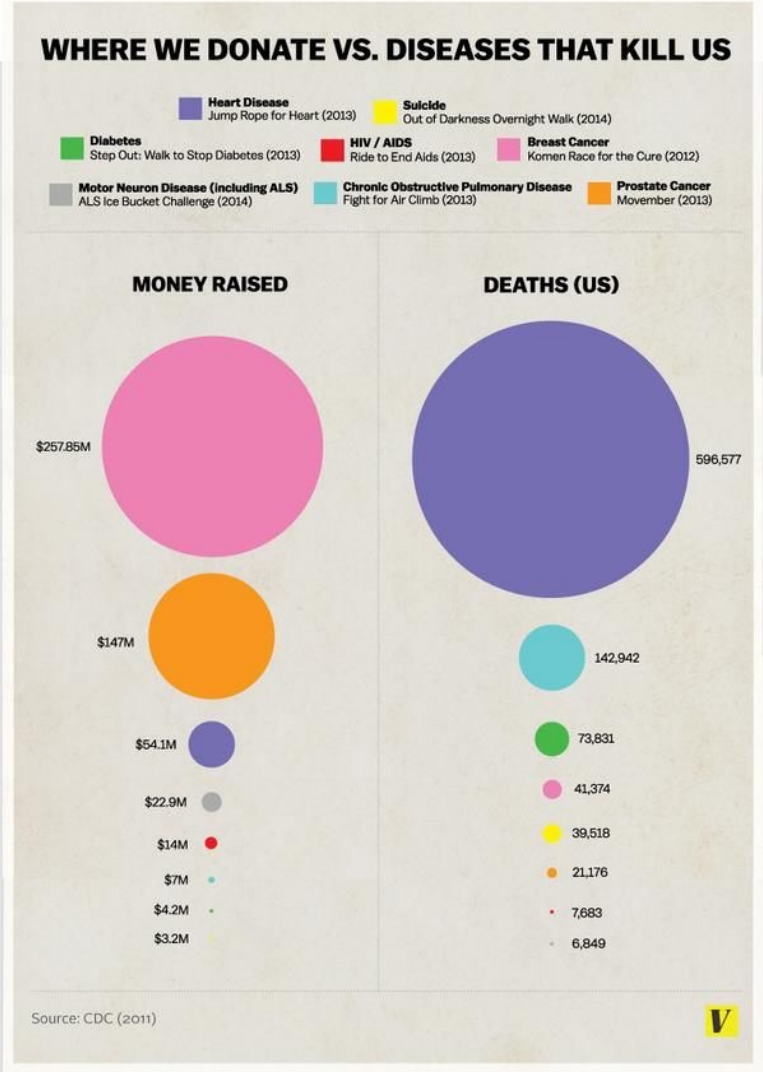
Ineffective 3D Pie Charts

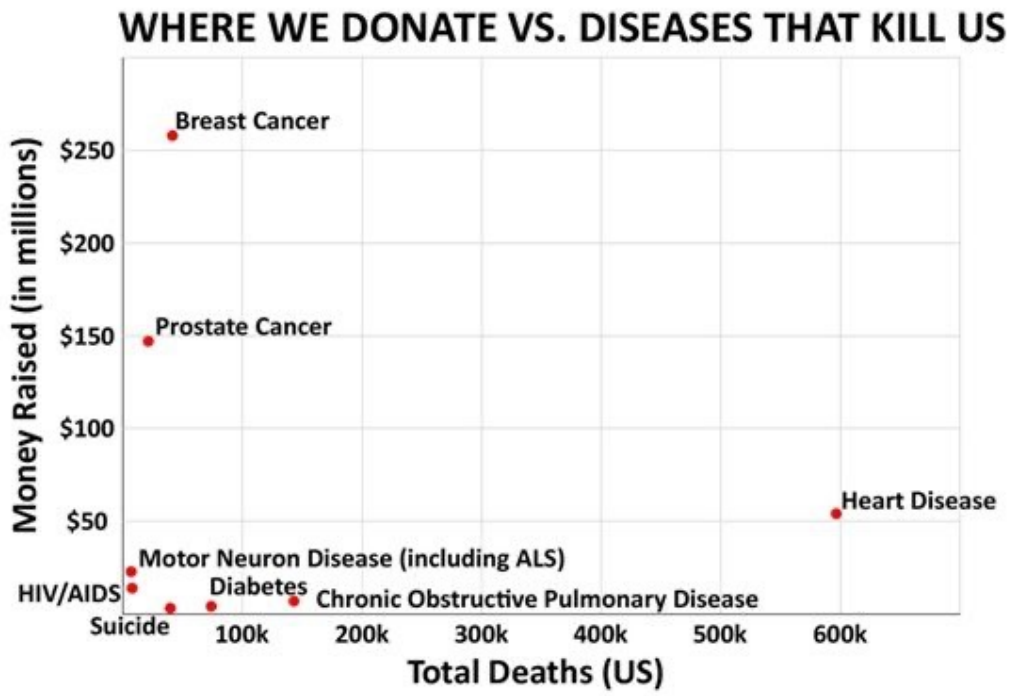


Supplier A looks misleadingly bigger than Supplier B

Avoid Unjustified 2D

Don't use two-dimensional idioms when one will do

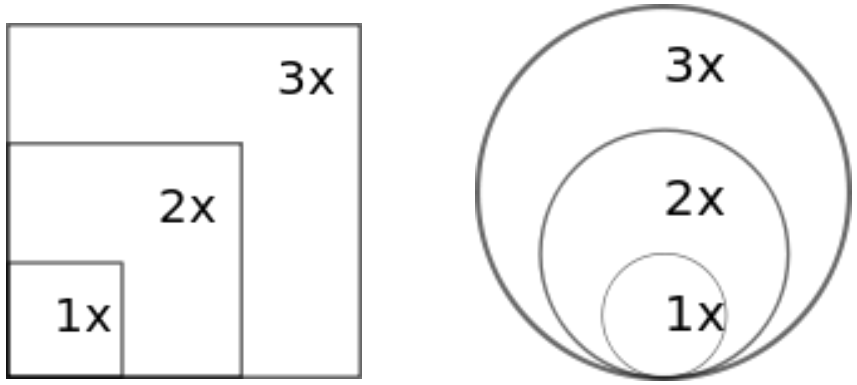




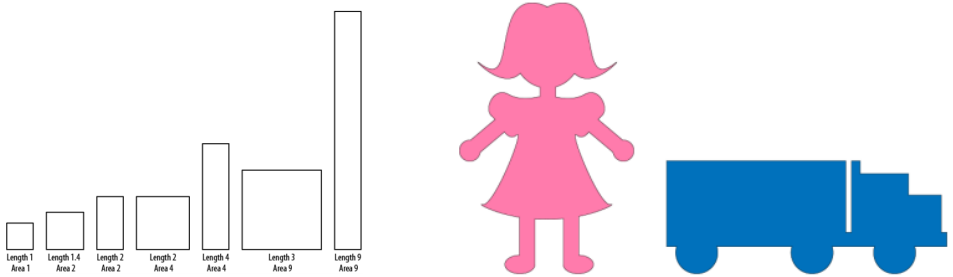
Easy to see which is bigger in the bubble chart (left)
but "how much bigger" is clearer in the simple scatter plot to (right)

Avoid Unjustified 2D

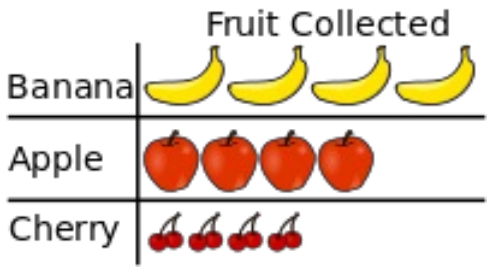
Avoid 2D area if 1D length will do

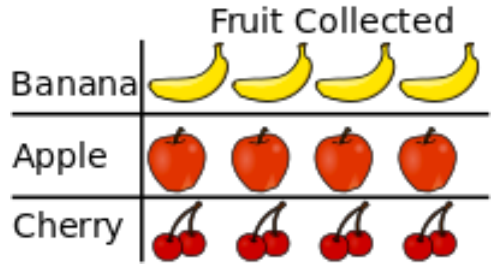


Non-uniformity of alignment, orientation and shape affect users perception of area



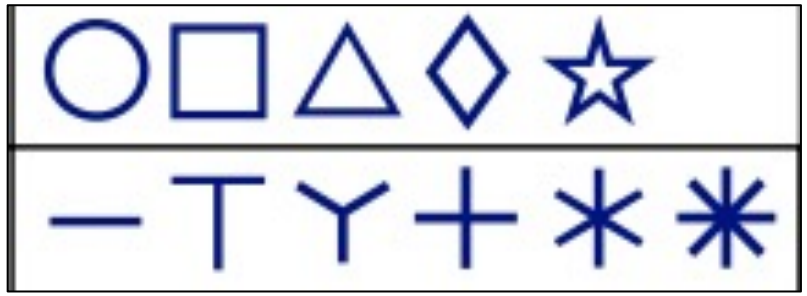
Avoid 2D glyphs if area is not constant



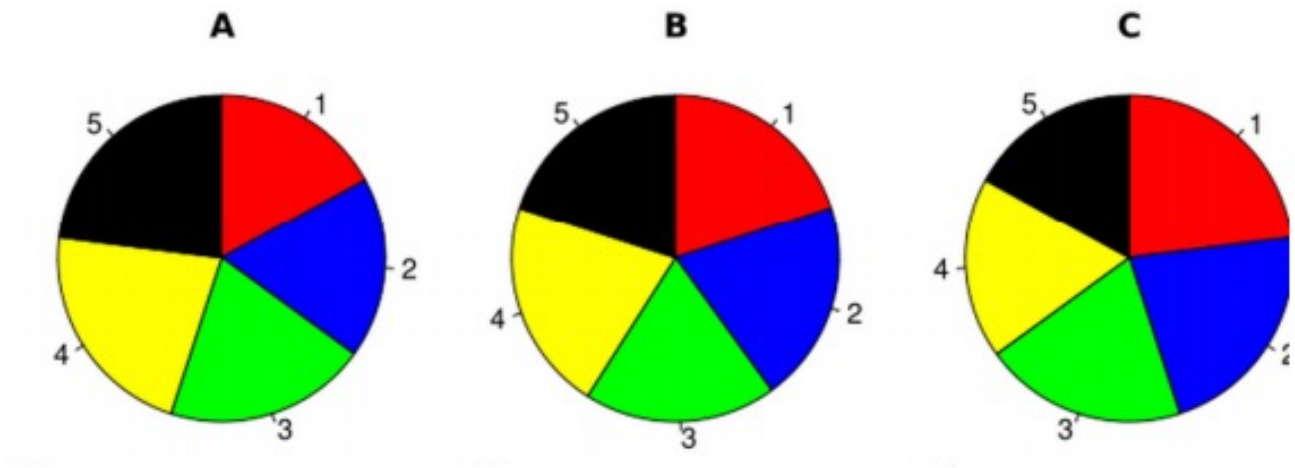


Shapes affect size perception

[Smart and Szafir 2019]



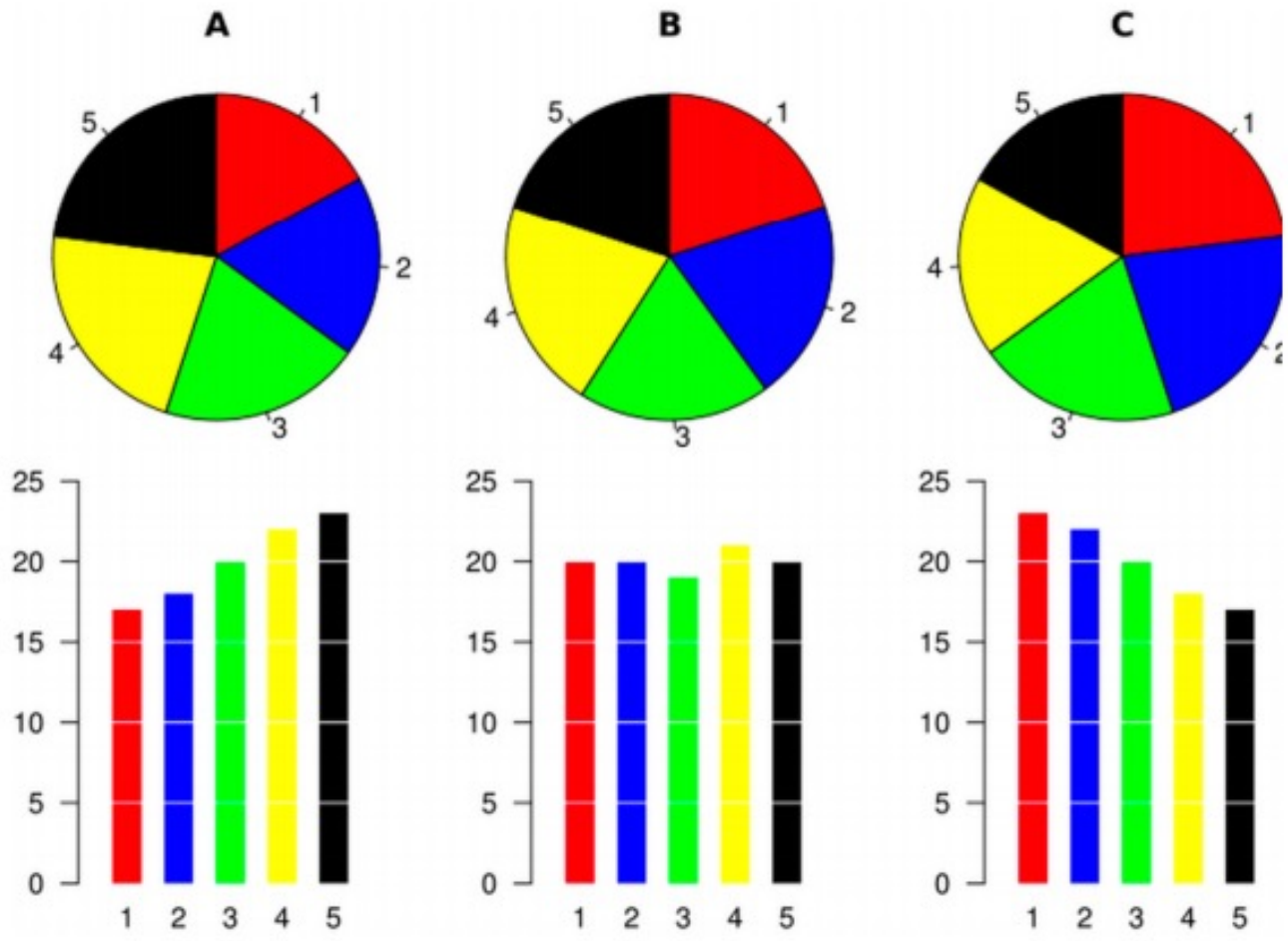
Aside: **Pie Charts Efficacy**



Can you spot differences? (Alexander Lex 2014)

Aside: **Pie Charts Efficacey**

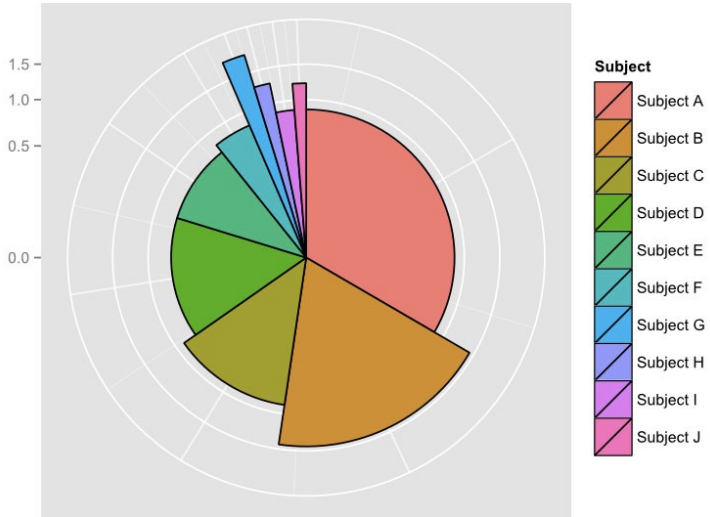
This example by Lex (2014) shows disadvantages of pie charts (comparing areas subtended by angles is not as effective as one-dimensional length judgements)

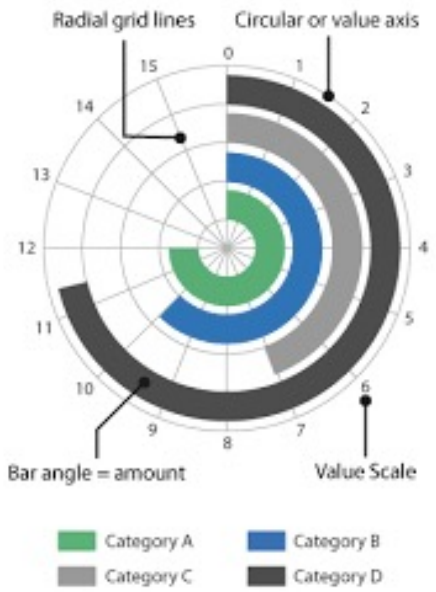


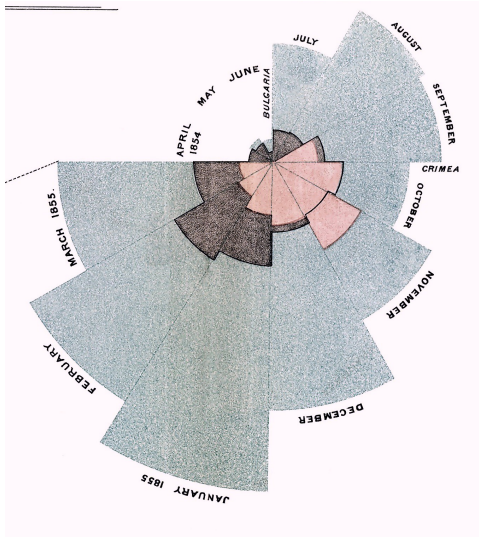
Avoid Unjustified 2D

Avoid pie charts and radial layouts if possible

- Angular size difficult to judge accurately
- Indiscriminable as segments become small







Particularly difficult when using different **radii:** more detail on the outer circumference while data is squashed in the center

However, some valid uses

- Part-to-whole relationships
- Cyclic data
- Hierarchical space-filling idioms

Resolution vs. immersion

Immersive Virtual Reality (VR) highly popular recently and increasingly proposed for visualization.

Some advantages e.g., scalability, intuitive exploration, etc.

However, Immersion typically not helpful for abstract data

- Viz doesn't really require what VR provides: sense of presence or stereoscopic 3D [Munxner 2014]
- Virtual reality for abstract (non-spatial) data difficult to justify

Resolution is much more important

- pixels are the scarcest resource
- desktop also better for workflow integration

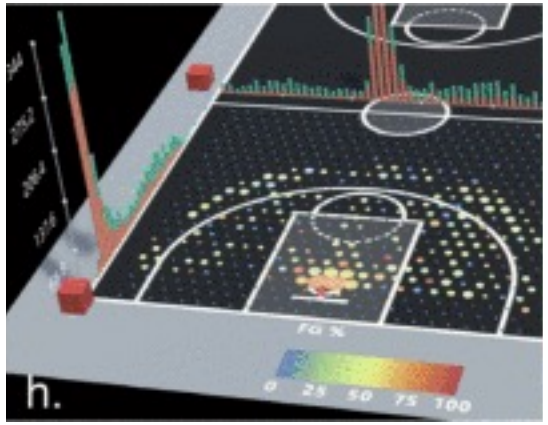


Image from Sicat et al (2019) DXR: a toolkit for immersive data visualization

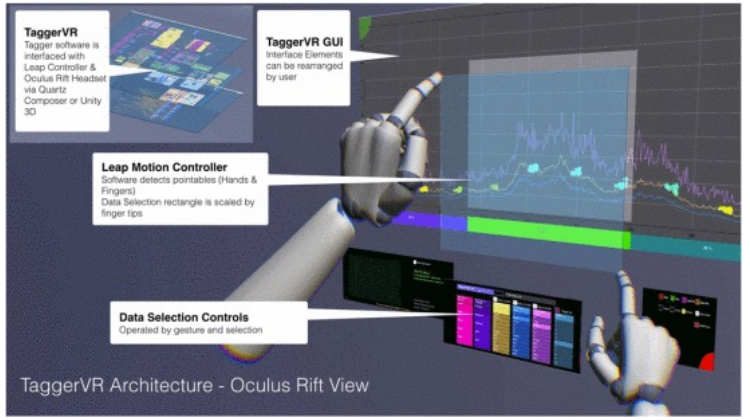
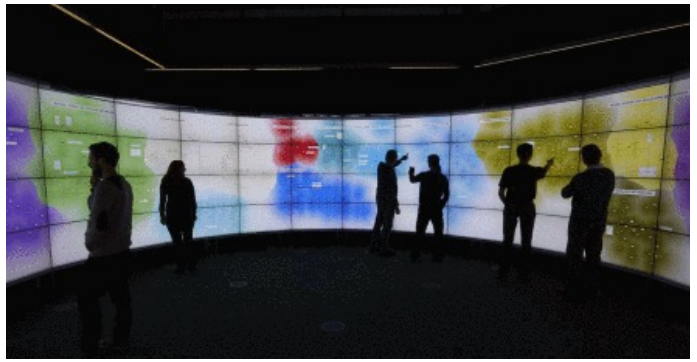


Image from Morse et al (2015) TaggerVR system for interactive Viusal Analytics

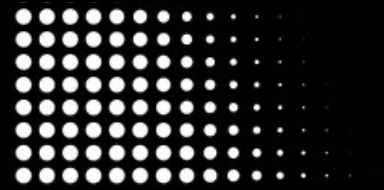


Big Data Immersive Analytics by Chandler et al (2015)

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The End

For now



Recommended reading

Most of this lecture was based on the book “Designing Data Visualizations” by N. Iliinsky & J. Steele. O’Reilly Press. 2011.

- A PDF excerpt of relevant chapters is available here: http://courses.ischool.utexas.edu/unmil/files/Designing_Data_Visualizations.pdf

Other references for this lecture include:

- Chapter 1: “Graphical Excellence” and Chapter 2: “Graphical Integrity” in The Visual Display of Quantitative Information. Edward Tufte. 1983. [Externally hosted PDF Excerpt here: https://tinyurl.com/uhaxq7k]

- Chapter 6 ”Rules of Thumb” in Visualization Analysis and Design. Munzner (2014)
- Chapter 13 “Designing Effective Visualizations” in Interactive Data Visualization. Ward et al (2015)

Other References

[Nussbaumer 15] Cole Nussbaumer Knaflic. “Storytelling with Data: A Data Visualization Guide for Business Professionals”. Wiley & Sons. 2015.

[Ware 08] Colin Ware. Visual Thinking For Design. Morgan Kaufmann. 2008

[Hanrahan 2011] Pat Hanrahan. “CS448B Data Visualization – Course Notes”, Stanford University. 2011

[Stone 2005] D. Stone, C. Jarrett, M. Woodroffe and S. Minocha. User Interface Design and Visualization. Morgan Kuafmann. 2005.

(Card et al 1991)

References

http://www.tbray.org/ongoing/data-ink/di6

http://www2.cs.uh.edu/~gnawali/courses/cosc6397-f13/intro-visualization.pdf

https://en.wikipedia.org/wiki/Misleading_graph

http://people.stat.sfu.ca/~cschwarz/Stat-650/Notes/PDF/ChapterBadgraphs.pdf

https://study.com/academy/lesson/edward-tufte-6-principles-of-graphical-integrity.html

http://courses.ischool.utexas.edu/unmil/files/Designing_Data_Visualizations.pdf

http://www-personal.umich.edu/~mmmc/516/notes/TuftePrinciples.pdf

Evaluating Visualizations [Ward]

FORMS OF EVALUATION

Usability Tests: Effective, Efficient, Engaging, Error Tolerant and Easy to Learn [Stone 2005]; through user observation

Expert Reviews: heuristic evaluation in controlled environment

Field Tests (beta): extended testing in normal user environment

Case studies and use cases: real, contrived examples of solving a particular problem/task, neds to be realistic & generalizable

SOME COMMON MEASURES

Computational performance

Memory performance

Data limitations: bounds of data that can be visualized

Occlusion: likelihood of one subset obscuring others; how many views needed to view full data set

Complexity: learning curve, parameters

Usability: ease of task, intuitiveness

Accuracy: frequency of success at performing task

Qlik

# of Incidents
203.1k

Same Day Compliance %
55%

Person Type
Employee
Customer

Treatment
First Aid Injury
Medical Treatment Injury

# of Incidents and Severity by Loc
OCEANIA
© Qlik, OpenStreetMap contributors

80k
60k
40k
20k
0
# of Incidents
63.25k
59.34k
52.42k
16.57k
9.43k
2.86k
12
FAI - First Aid Inj...
<Undefined>
No Treatment - Report Only
MTI - Medical Treatment In...
LTI - Lost Time Injury
RWI - Restricted Work Injury
Fatality
Classification

Claim Staus
Open
<Undefined>
Closed
11.8%
84.2%
Incident

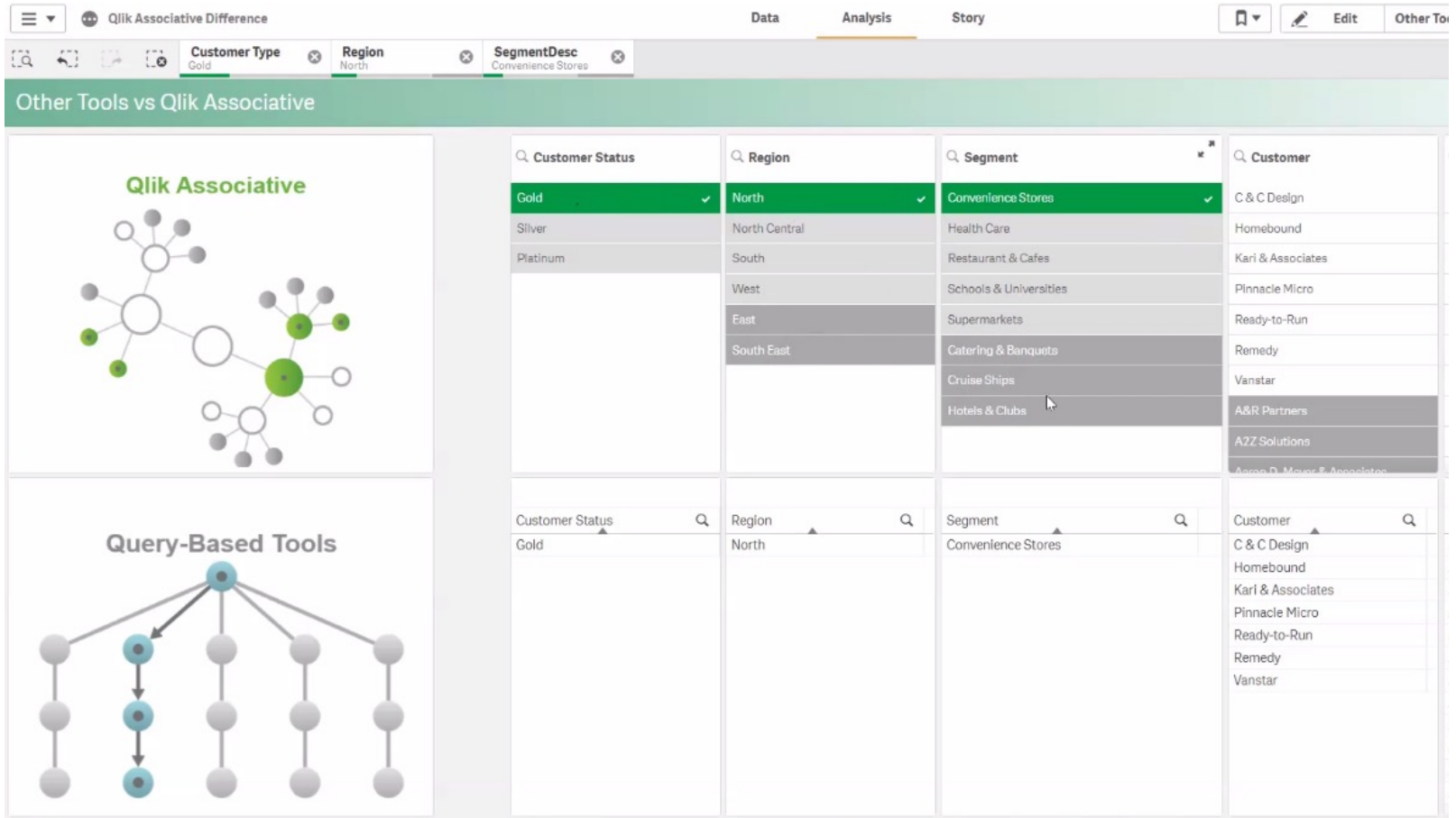
SpotLight Tool: Qlik

GUI-based Viz Tool

- Often cited amongst top-10 Visualization/BI Tools

Unique Features

- Dashboard style vis similar to Tableau
- Data Associations
- Natural Language Search
- Relatively Easy to learn – tonnes of resources
- 1 year Free Academic License (market value of EUR 3000)
- Industry recognized certification (after brief 16.5hr free training course)



Demos:
demos.qlik.com/qliksense

Examples:
www.qlik.com/us/solutions

Register:
qlikid.qlik.com/register

Apply for Academic Programme:
qlik.com/academicprogram

Function over form

"It's like a finger pointing away to the moon. Don't concentrate on the finger or you will miss all that heavenly glory"

- Bruce Lee

Focus on objective, not method

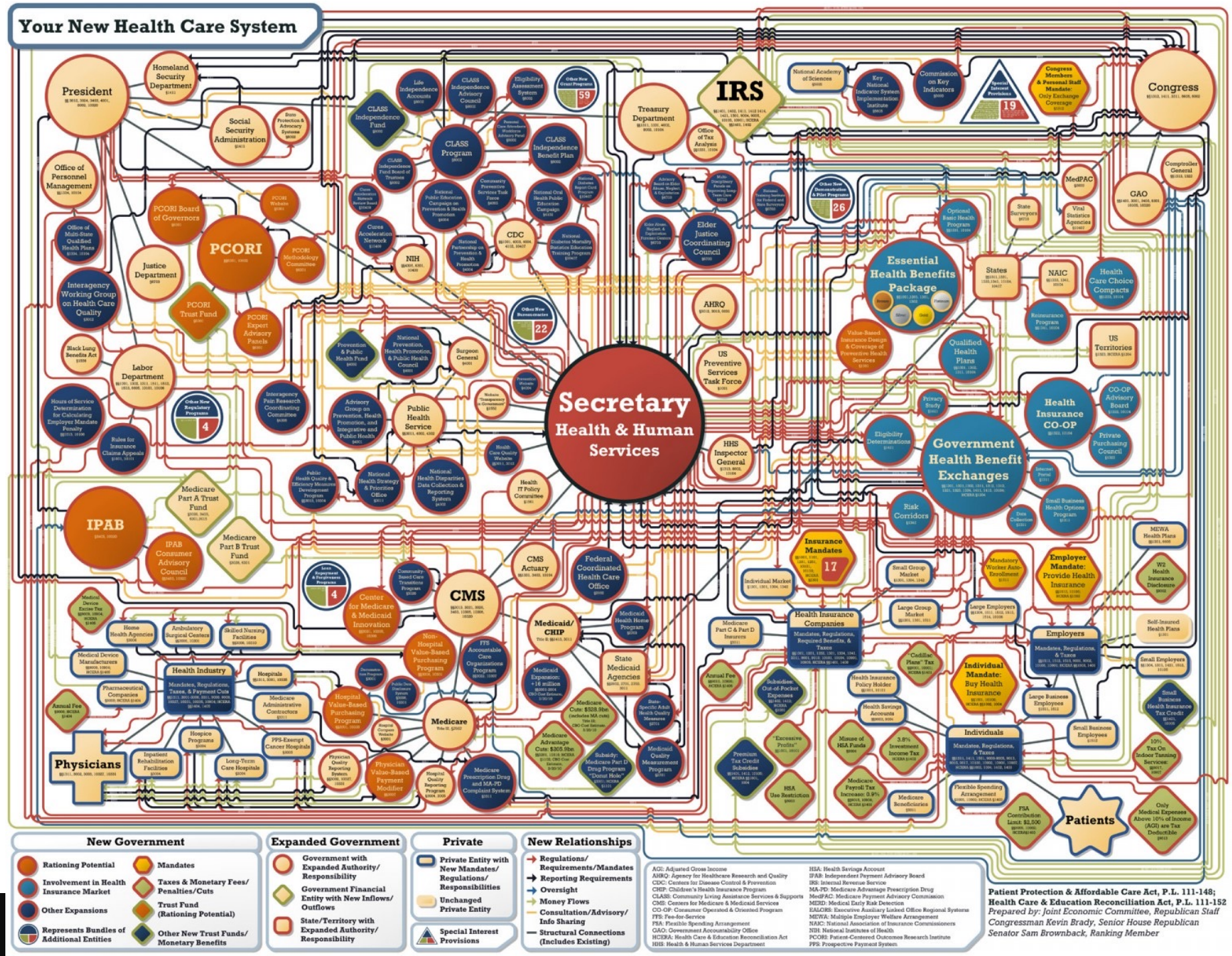
- Choose encoding's for a specific purpose not merely because you can - Don't distract
- Iterate/evolve but don't lose sight of main objective

Understand and define objective clearly before designing the visualization

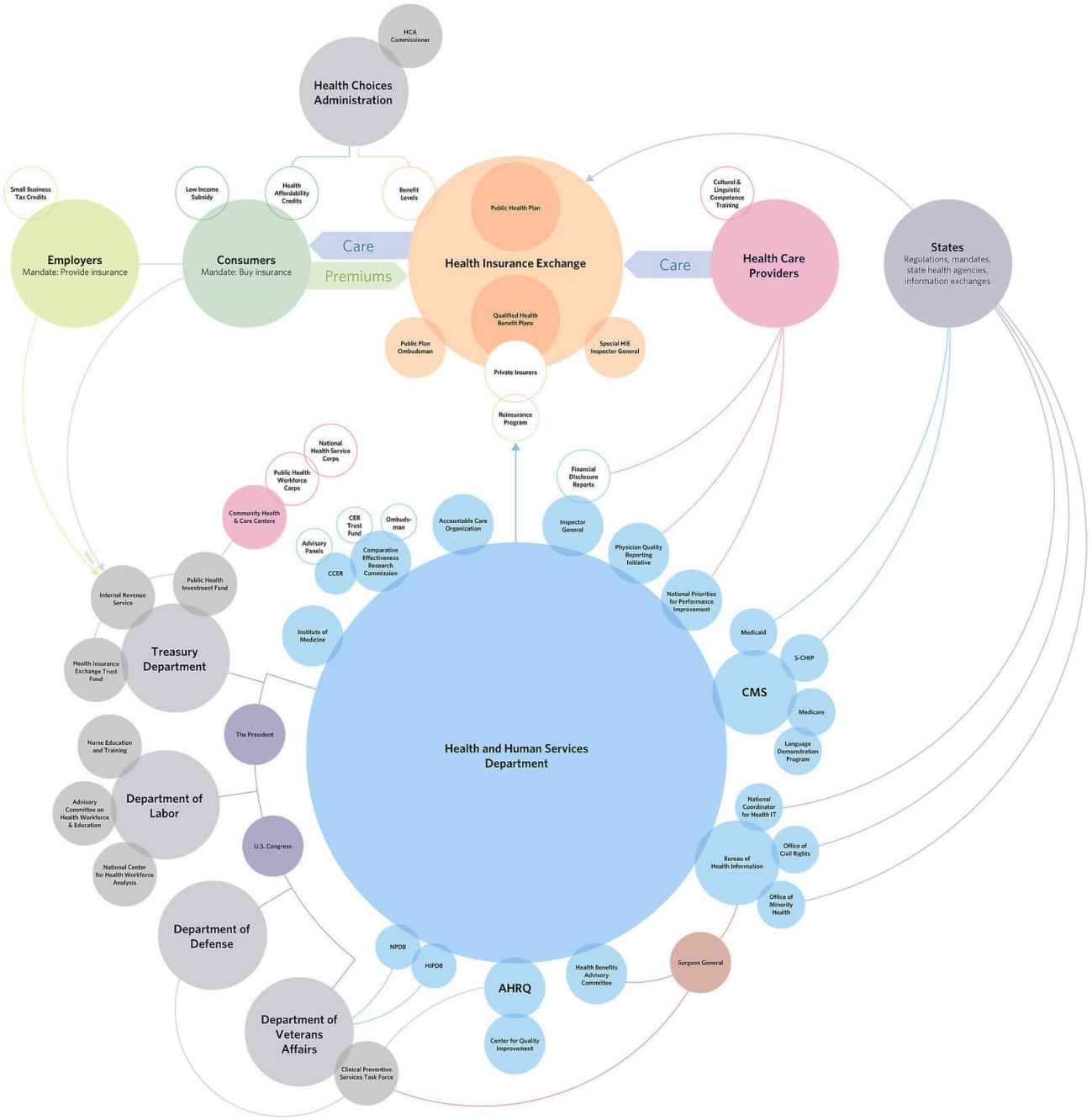
- Have you understood the task yourself?
- Is the objective exploratory or explanatory
- What is the Action + Target?
- Can you justify why you have chosen a particular way to present the data?

Image © salesforce

Keep it Simple (KISS)



KEEP IT SIMPLE (KISS)



Visualization Principles (Tufte 1983)

Graphical displays should:

- show the data
- induce viewer to think about **substance rather than about the methodology** of the visualization design
- **avoid distorting** what the data have to say
- present **many numbers in a small space**
- make **large data sets coherent**
- **encourage the eye to compare** different pieces of data
- reveal the data at **several levels of detail**, from a broad overview to the fine structure.
- **serve a reasonably clear purpose**: description, exploration, tabulation or decoration.
- be **closely integrated with the statistical and verbal descriptions** of a data set.

Principles of Data integrity:

- Representation of numbers should match the true proportions.
- Labeling should be clear and detailed.
- Show data variation, not design variation.
- The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data
- Graphics must not quote data out of context

From: E. Tufte. The Visual Display of Quantitative Information (1983)

Aside: Rules of Thumb (Munzner 2014)

No-redundant dimensionality: 3D vs. 2D vs. 1D

Static vs Dynamic

Resolution vs Immersion

Aside: Rules of Thumb (Munzner 2014)

Static vs Dynamic

Resolution vs Immersion

Detail on Demand

Responsiveness and interaction

Placement

Consider position first

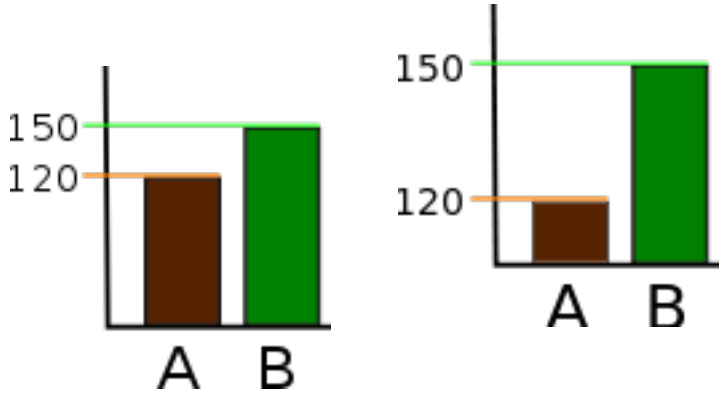
Layout and Axes

Proximity / Relative Placement

Absolute Placement

Patterns

From Illinski & Steele
Is this really needed here?
Perhaps move to Encoding lecture next year



Guidelines for Effective Visualizations (Ward)

Selecting and modifying views

- Mapping to encodings
- Scaling
- Level-of-Detail

What about these?

Visualization Pitfalls [ward 2014]

Misleading Visualizations

- Data scrubbing
- Range Distortion
- Unbalanced Scaling
- Abusing Dimensionality

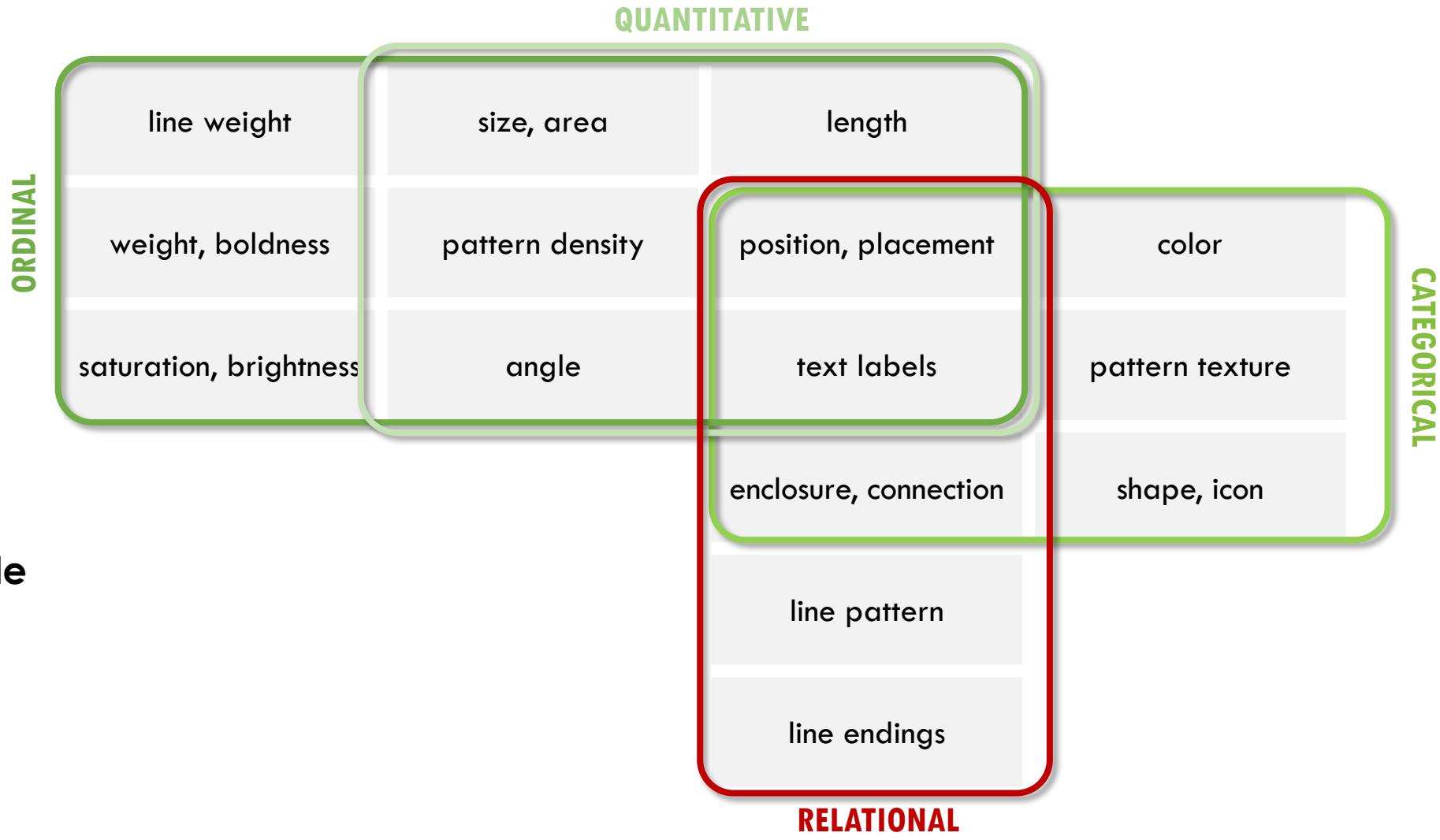
Visual Nonsense

Losing Data in Chart-Junk

Raw vs. Derived Data

Absolute vs Relative Judgement

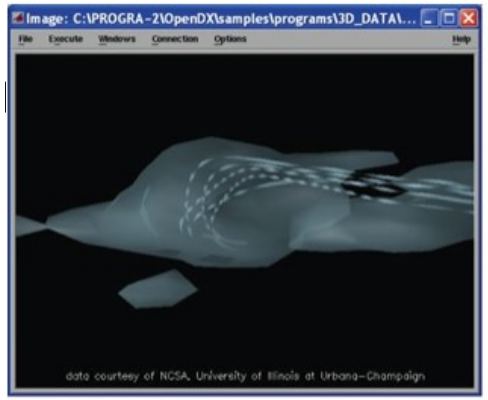
Applicability of Encodings

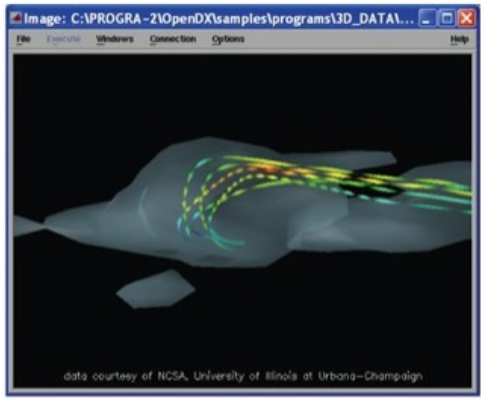


Not all encodings are suitable to all types of data

From Ilinsky and Steele 2011

Visualization Aesthetics [Wa

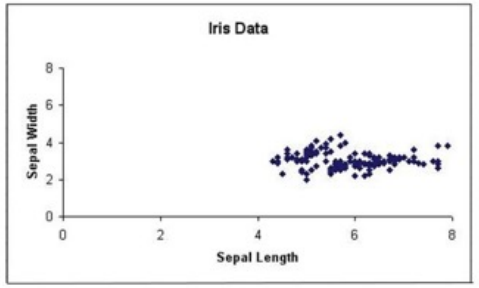


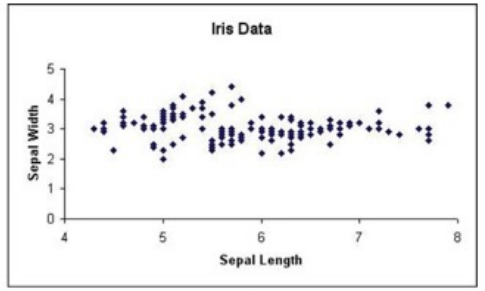


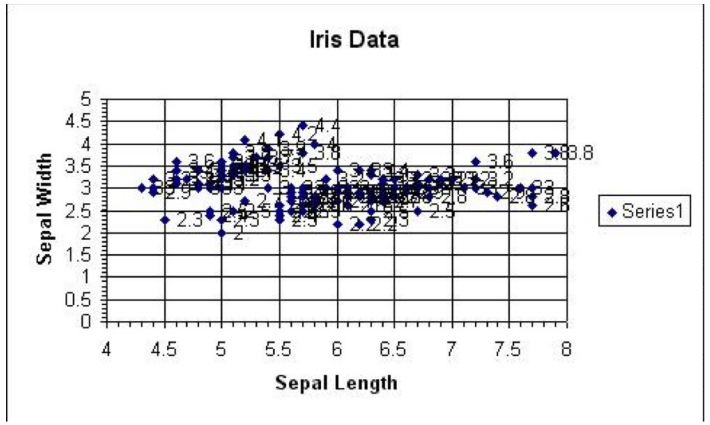
Focus: draw attention to the parts that are important

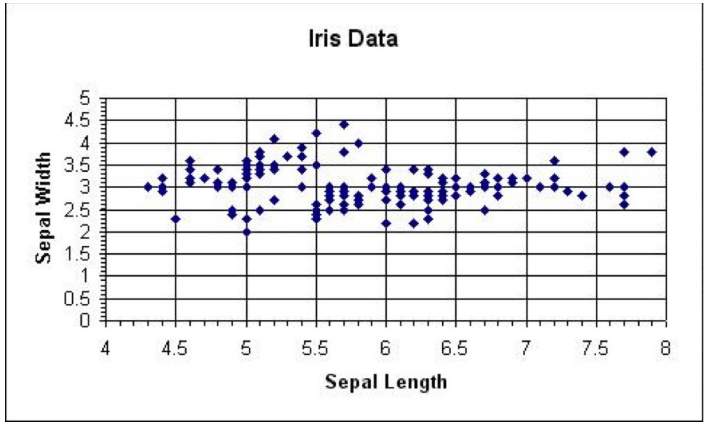
Balance: use screen space, with most important components in the centre. Emphasis should not be given to any particular border

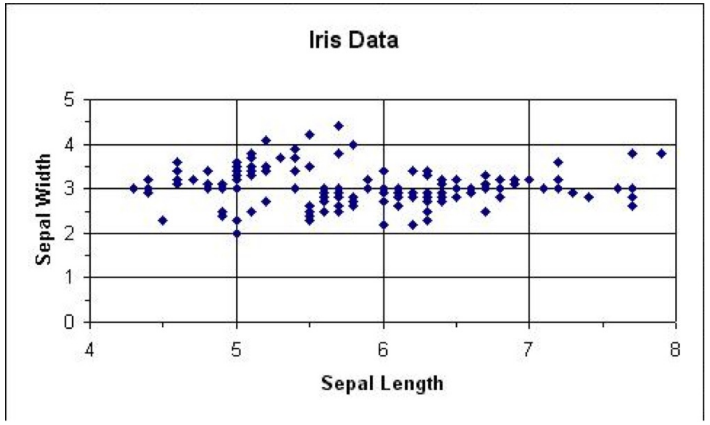
Simplicity: avoid information overload; don't overuse graphics gimmicks just because they are available











No unjustified 2D

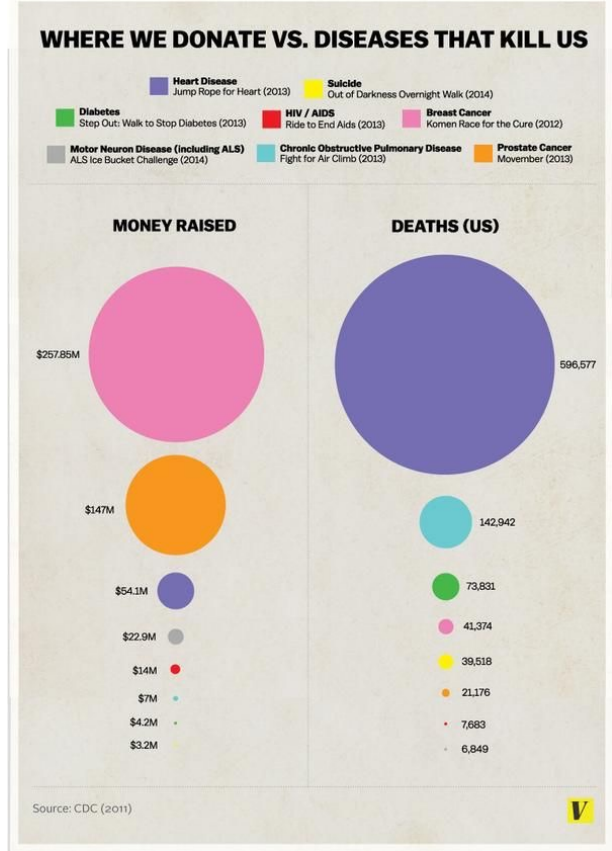
Don’t use two dimensions when one will do

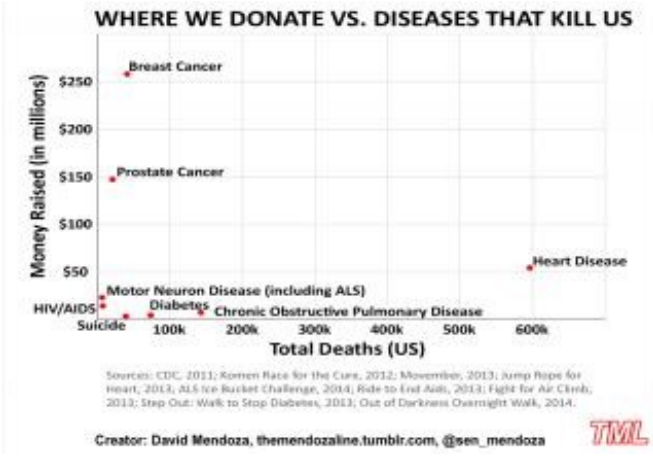
E.g. node-link tree not always useful

- Especially if reading text is central to task!
- Arranging as network means lower information density and harder label lookup compared to text lists

But benefits outweigh costs when topological structure is important for task

- e.g. Number of jumps, Path through network





More style than Substance?

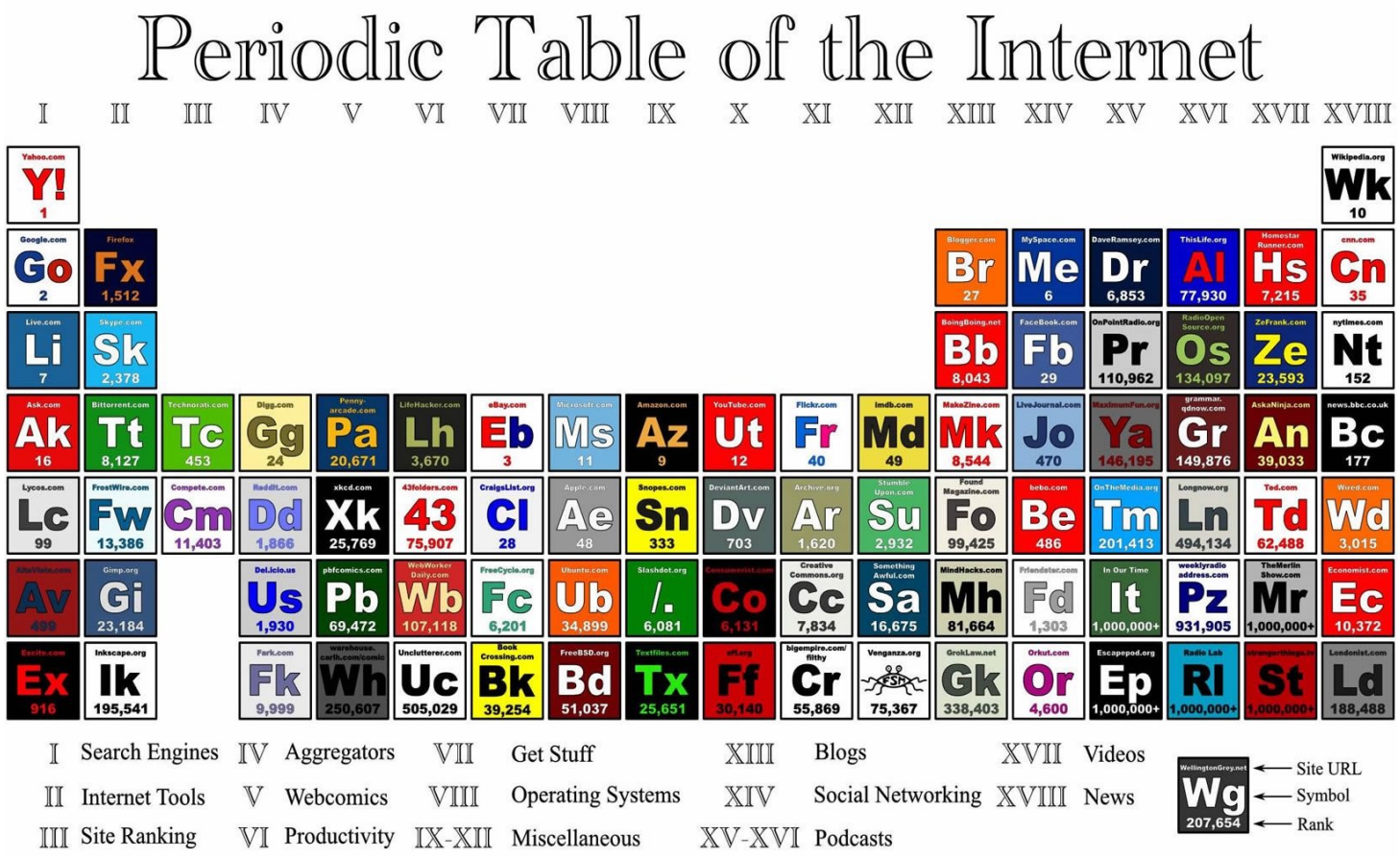


Chart by Wellington Grey (2007)

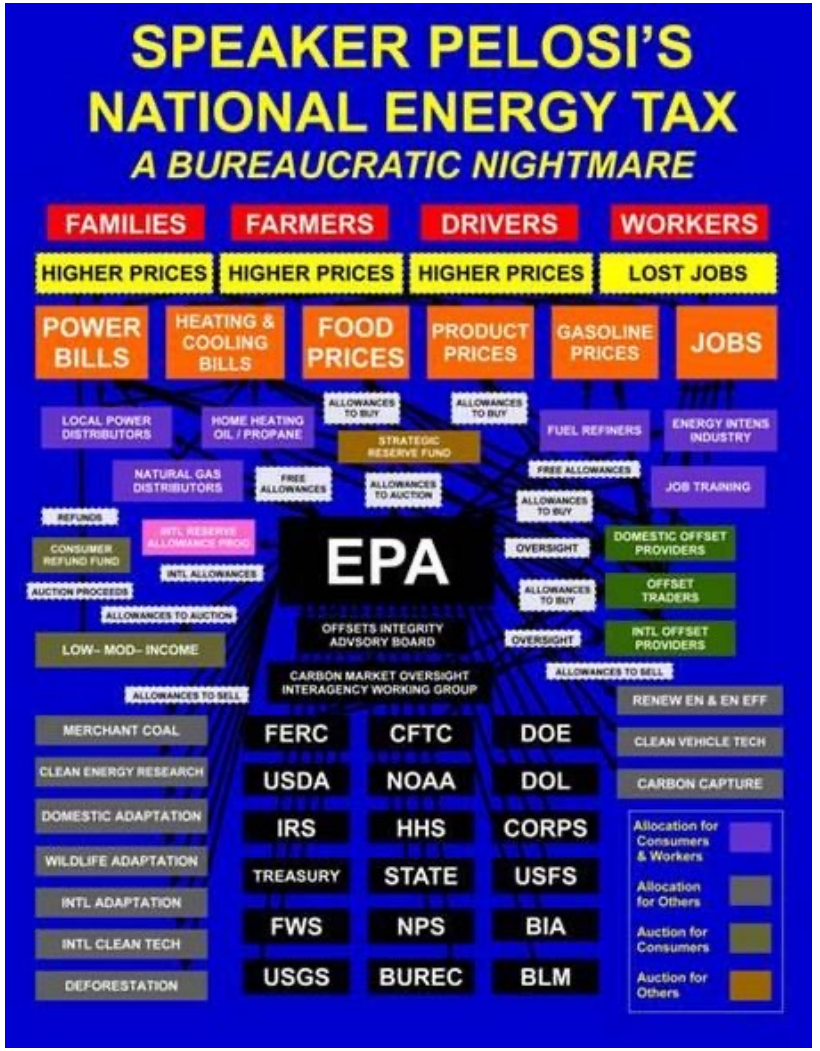


Chart by Brian Boehner [2008]

Masquerading as visualizations: above visualizations have a lot of graphical elements, but the encodings and idioms contribute little functionally beyond decorative value (perhaps to engage)

Function over form

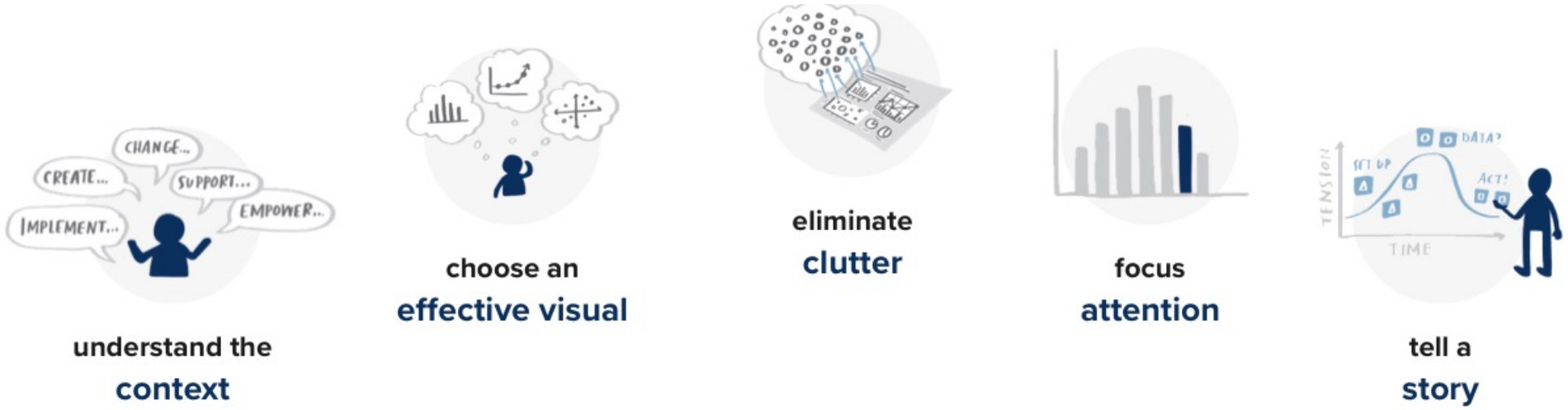
"It's like a finger pointing away to the moon. Don't concentrate on the finger or you will miss all that heavenly glory" - Bruce Lee

Understand and define objective clearly before designing the visualization

- Have you understood the task yourself?
- What is the Action + Target?
- Can you justify why you have chosen a particular way to present the data?

Focus on objective, not method

- Choose encodings for a specific purpose not merely because you can - Don't distract
- Iterate/evolve but don't lose sight of main objective



Good Visualization Guidelines

Prologue

Video Examples for final Assignme

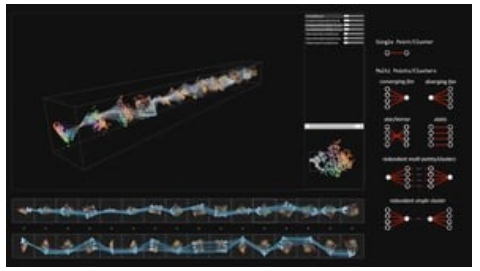
Note: below images are links

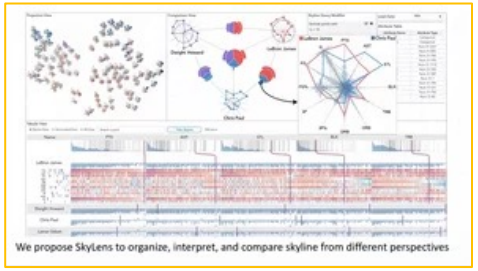
Vis 2017 : 30 Second Videos [LINK]

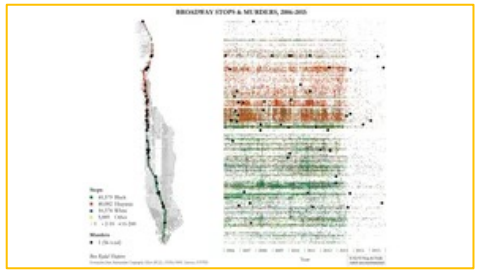
- Shorter than required but shows how much detail can be expressed in a short video
- These are technical papers : you are not really expected to provide scientific novelty as they are

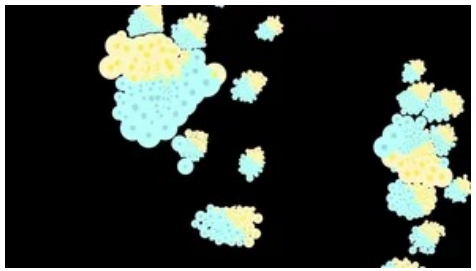
VAST Challenge Videos

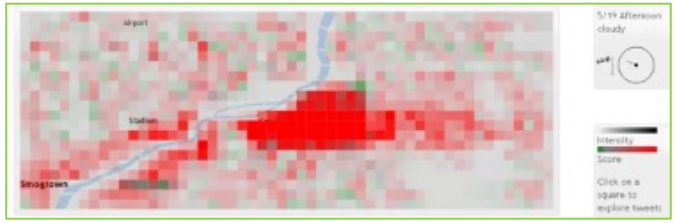
- Closer to the brief of this project i.e. novel implementation of visualization of some data
- Mostly longer than 2 mins
- A lot detail of also the user process of analyzing the data (not really expected in this assignment)

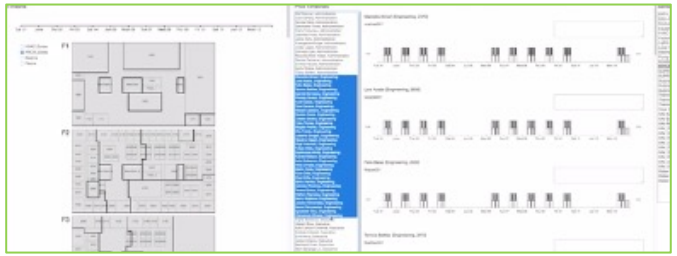


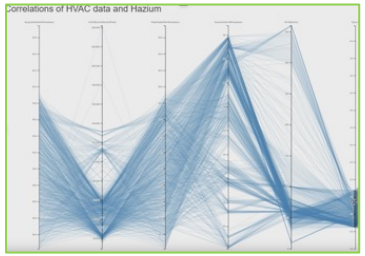












Video Suggestions

You may just use captions / subtitles to describe the video [voice over not really expected]

Get a screen capture program. Free version with watermark is fine. E.g.

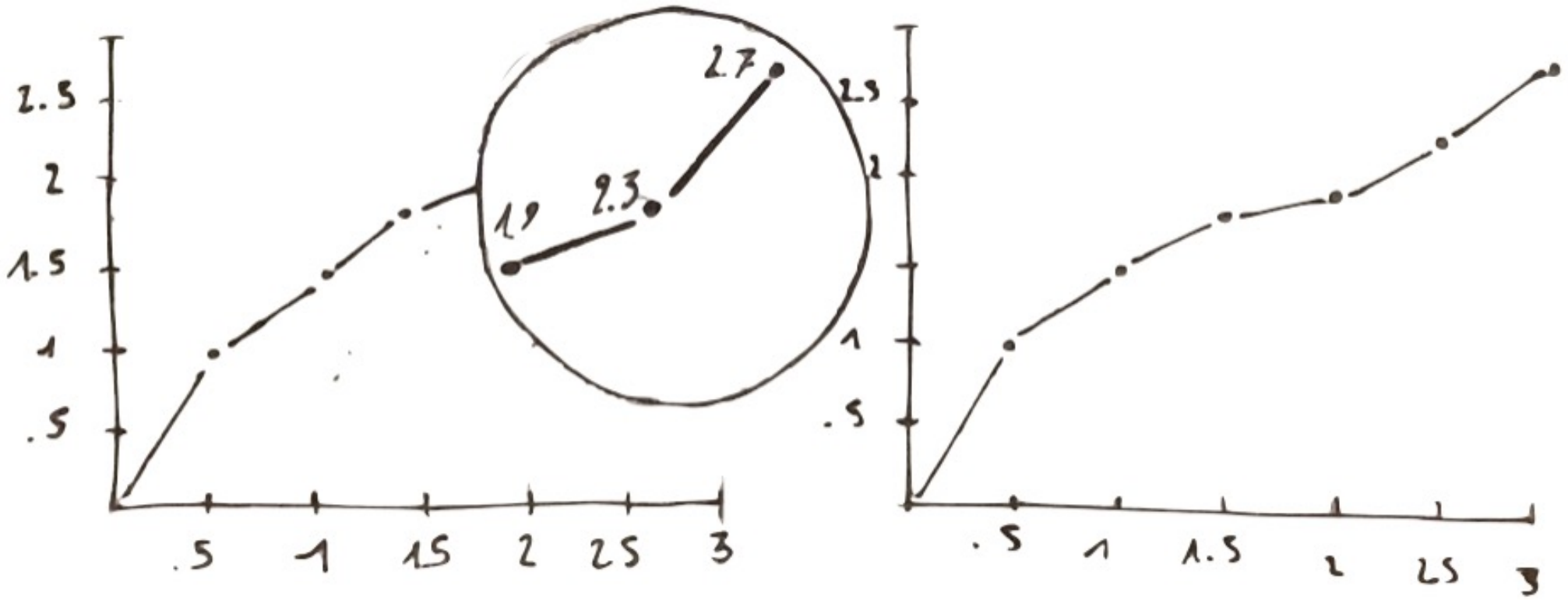
- https://www.techsmith.com/download/camtasia/ [mac/windows]
- https://www.bandicam.com/ [windows]
- http://www.fraps.com/download.php [windows]

Some free video editing tools out there BUT just keep it functional I'm not really expecting a really fancy video

- Windows Movie Maker as part of Windows Essentials Pack (really easy to use. Sadly now discontinued but you can still find downloads on the web and works in win 10) [windows]
- iMovie [mac]
- VirtualDub [windows]

Create a PowerPoint presentation + switch through slides automatically

Unintended Distortion



Resolution Vs. immersion

Immersive Virtual Reality (VR) highly popular recently and often proposed for visualization. Some advantages e.g. scalability, intuitive exploration, etc.)

However Immersion typically not helpful for abstract data

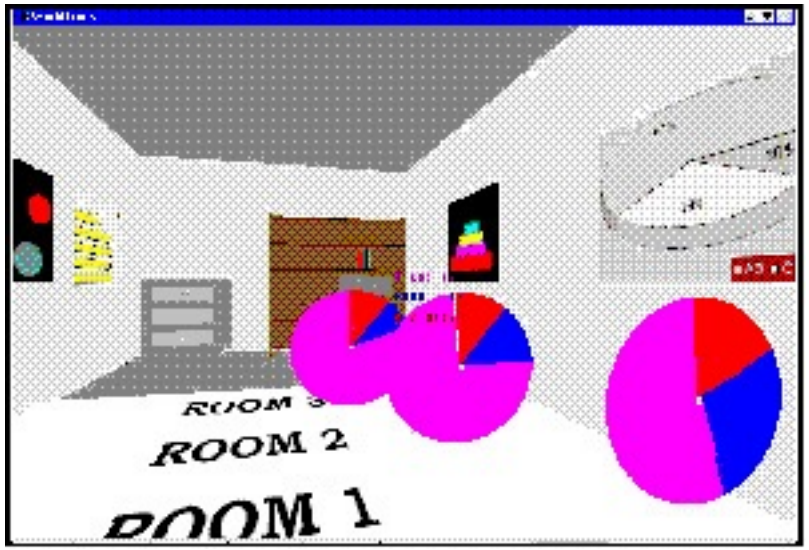
- Viz doesn't really require what VR provides: sense of presence or stereoscopic 3D

Resolution is much more important

- pixels are the scarcest resource
- desktop also better for workflow integration

Virtual reality for abstract data difficult to justify

Munzner (2014)



[An information visualization tool using VR. Kirner and Martins. Proc. Symp. Applied Computing 2000]

Visualization Principles (Cleveland 85)

Clear vision

- Make the data stand out. Avoid **superfluidity.**
- Use **visually prominent graphical elements** to show the data.
- Do not **clutter** the data region.
- Use a reference line when there is an important value that must be seen across the entire graph, but do not let the line interfere with the data.
- Do not allow data labels in the data region to interfere with the quantitative data or to clutter the graph.
- Avoid putting notes, keys, and markers in the data region. Put keys and markers outside the data region and put notes in the legend or in the text.
- **Overlapping** plotting symbols must be visually distinguishable.
- Superposed data sets must be readily visually discriminated.
- Visual clarity must be preserved under reduction and reproduction.

Clear understanding

- Put major conclusions into graphical form. Make legends comprehensive and informative.
- Error bars should be clearly explained.
- Proofread graphs.
- Strive for clarity.

Scales

- Choose the range of tick marks to include/nearly include the range of data.
- Subject to the constraints that scales have, choose the scales so that the data fill up as much of the region as possible.
- It is sometimes helpful to use the pair of scale lines for a variable to show two different scales.
- Choose appropriate scales when graphs are compared.
- Do not insist that zero always be included on a scale showing magnitude.
- Use a logarithmic scale when it is important to understand percent change or multiplicative factors.
- Showing data on a logarithmic scale can improve resolution.

General strategy

- A large amount of quantitative information can be packed into a small region.
- Graphing data should be an interactive, experimental process.
- Graph data two or more times when it is needed.
- Many useful graphs require careful, detailed study.

From William S. Cleveland, The Elements of Graphing Data (1985)