ASSIGNMENT 3: NOVEL VISUALIZATION

Design, Implementation and Presentation

ASSIGNMENT 3 : DEALING WITH COMPLEXITY

[WORTH 40% DESIGN, IMPLEMENTATION, AND PRESENTATION]

Motivations :

- apply concepts discussed in lectures to visualize a data set of some complexity
- design and implement a visualization comprised of more than a single static idiom (or chart type) OR that requires a something more than a well-established chart type

Objectives [see details in following slides]

1. Choose an interesting dataset, that necessitates visualization with more than a single established chart
2. Design and implement a *novel* visualization for this data set
3. Choose any tool (you may even use a combination of several)
4. Write a report, discussing your design in the context of the issues discussed in the course
5. Create a 2-minute presentation video

1. DATA

You are free to choose any interesting data set. Some suggested data sources are provided on blackboard. It is up to you to set the scope.

There is no explicit bottom-limit to how complex the data-set should be, but there should be some rationale for visualizing it with something more elaborate than a single established chart/idiom.

- As a rough approximate, try to use Minard's Data set (Assignment 1.2) as a baseline. Try to find something at least as complex as this in some way.
- The higher the complexity the more likely that the novelty of your visualization will be high, but try to balance this with feasibility (for a time-budget of about 45 hours of work).

Complexity can be based on various criteria, such as...

- *Volume*: Large items (high-dimensional), many items, overall size
- *Variety*: Multi-variate/many attributes, heterogenous, non-uniformly distributed
- *Structure*: task is related to complex relationships between elements
- *Visual complexity* needed for the task: e.g. prone to over-plotting, many/complex graphical elements

2. NOVELTY

Novelty, here, merely means that you are NOT simply replicating an existing visualization of your chosen data. i.e. either the dataset has not yet been visualized OR it has not been visualized in the way that you are going to visualize it. In addition, there are two essential requirements:

Requirement #1 : at least one of the following should apply to your result

- The overall visualization uses more than one idiom, combined in a new way, **OR**
- The visualization uses an idiom that is unique in some way e.g. combining several encoding channels or a completely new idiom

Requirement #2: your visualization must feature at least one of the following technical elements:

- Multiple facets (see Lecture 10) e.g. small multiples, or a dashboard of different idioms, **OR**
- Dynamic/animated view (animation may be linear and passive i.e. without any input from user), **OR**
- Interactive manipulation of the view (some form of real-time changes can be made by the user to what is show on screen)

Some additional guidelines (this may impact your grade):

- The full visualization should be visible on one page/screen without excessive scrolling or turning the page (e.g. separate tabs should be avoided). You may allow the user to change the view interactively (e.g. modify some part of the view by clicking), but you should not submit several unconnected visualizations.
- The submitted visualization should be a 'finished product'; the user should be able view/use it without looking at your code or having to change your code/script
- Choice of (and number of) idioms used should be justified and not merely to 'show off' ("function over form").

3. CHOICE OF TOOL

You may choose any generalized* visualization tool.

* Essentially, the only thing to AVOID is using a third-party black-box solution that has been specifically created to visualize something closely matching your dataset with little/no design input from you (e.g. using an pre-compiled off-the-shelf "MRI Visualization tool" for an MRI scan OR plugging new data into a dashboard someone else created). Avoid this as you will score poorly in terms of novelty.

This is not an exhaustive list but examples of **tools you may certainly use include**: processing, d3.js, Tableau, qlikview, PowerBI, HighCharts, GoogleCharts / Google Data Studio, plotly, three.js, p5.js, jupyter notebook, VtK, OpenGL, R, matplotlib, shiny, matlab, visme, datawrapper, rawgraphs, bokeh, altair, Mathematica, …

4. THE REPORT

Write a report about your implementation of max 3 pages (not including references) + 1 full page image (or screenshot) of your visualization

- Template provided on blackboard. You may simply replicate this as best as you can according to the sample PDF that describes the spec.

The report should discuss

- The tools/technologies used and (in brief) any steps taken to pre-process the data
- The dataset : general description, more detailed analysis of dataset/data types, attribute types
- The tasks supported by the visualization
- Encoding channels and idioms (including interactive elements, if any) AND a justification for why particular encodings are used
- A discussion about the novelty of the visualization; if/why it is particularly complex
- A critical analysis of the strengths and weaknesses of the visualization

You *may* support this with additional images but these must be included within the 3-page limit

Do not assume the reader has seen Assignment A2, if you are submitting a related implementation

5. THE VIDEO

A presentation video of max 2 minutes outlining the technical contributions of the paper

- This should essentially summarise the main points of the technical paper with visuals of your visualization result

- use captions or voice over to present a summary of both the artefact and the paper

- Please note that these may be made available to students in the class (and future classes) to view as an example. If you wish to remain anonymous, do not include your name in the video.

DELIVERABLES

1. **Technical paper**
   - This should be in PDF format of upto 4-pages (+ references)
   - For marking reasons, **you should not add the report insidea a zip file; and should not submit it as a link.**

2. **A LINK to a presentation video of max 2 minutes outlining the technical contributions of the paper**
   - Please note that these may be made available to students in the class (and future classes) for educational purposes, to view anonymously as an example.
   - DO NOT SUBMIT THE ACTUAL VIDEO ON BLACKBOARD (ONLY A LINK)

3. **Visualization "artefact" i.e. the actual visual output or programme that you have created:**
   - This could be ...
     - a link to your interactive visualization if it is online;
     - a locally executable binary (you must provide instructions on how to run it);
     - a video if the visualization is animated but non-interactive;
     - or a high resolution image/PDF if it is static
   - You MUST submit a link to download the project source files and data set
   - DO NOT SUBMIT THE ACTUAL VIDEO ON BLACKBOARD (ONLY A LINK)

NB. If your artefact itself is a video, you may submit two videos: an artefact video and a presentation video (there is no limit on the length of the artefact video but if larger than 10Mb, submit a link)

IMPORTANT NOTES: PERMISSIONS TO SHARED VIDEOS AND FILES

Numerous students in the past have shared files that are protected and inadvertently can't be accessed by the examiners. This affects marking progress. Also note that I am not the only one that needs to see these. Therefore, please ensure the following:

1. You <u>MUST</u> either:

a) Make the file viewable by anyone with the link OR

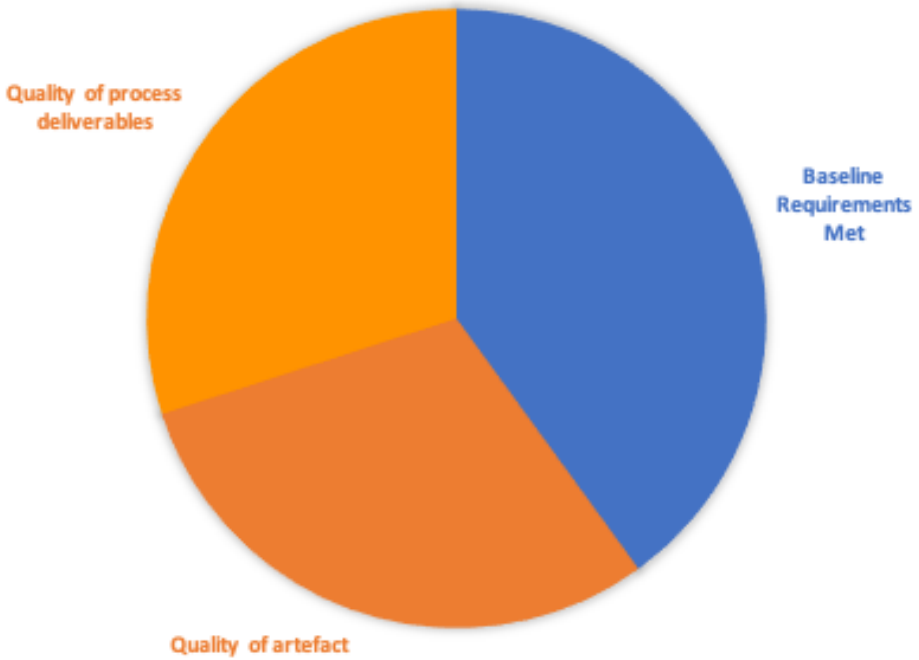
b) Make the file viewable to cs7ds4.staff@gmail.com OR

c) Password protect the file and submit the password and link in the text of your report

2. You MUST not change the file after submission; the file MUST remain online (unchanged) until August 30th, 2023

REMINDER: you may not submit a link to the report. This must be in PDF format uploaded to blackboard

MARKING SCHEME



This project is worth 40% of the module, broken down as follows:

0 to 16/40 for meeting the baseline requirements : all deliverables submitted on time, artefact successfully visualizes the data, the process paper and video together reasonably describe the data, task, idioms and encodings channels employed.

+ 0 to 12/40 marks for quality of the artefact beyond the baseline including but not limited to: Novelty, Complexity and How well it achieves the intended data visualization task(s) using the idioms employed

+ 0 to 12/40 marks for the quality of process deliverables beyond the baseline

- Paper is well written (clear and concise with minimal errors and suitable references); video is well put together and summarizes the paper and artefact to someone not previously familiar with it
- Paper expresses clearly the complexity/novelty of the visualization; explains/justifies the design choices made; makes a good attempt at discussing strengths and weaknesses

Late Submissions will incur a 10% penalty for each day late up to a maximum of 6 days. After that all submissions are capped at 40% of the project mark until 31st December, 2022, after which a zero mark will be given.

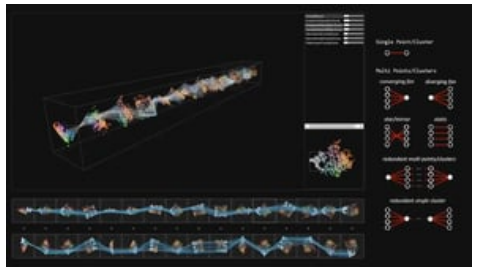
EXAMPLES OF SHORT PRESENTATIONS

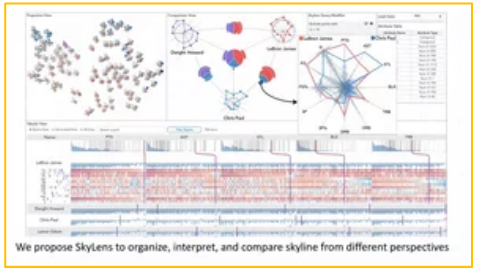
Vis 2017 : 30 Second Videos [LINK]

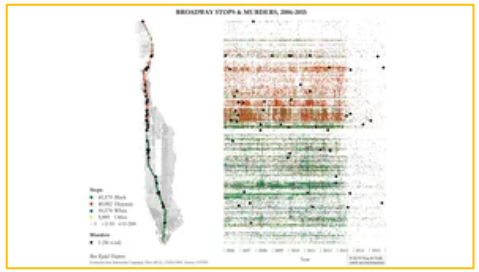
- Shorter than required but shows how much detail can be expressed in a short video
- These are technical papers : you are not really expected to provide scientific novelty as they are providing in these conference videos

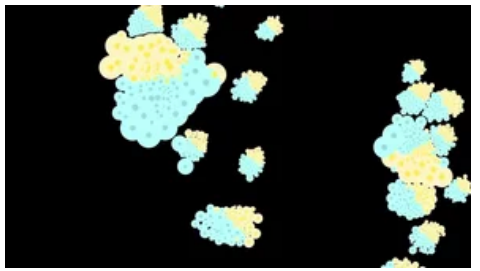
VAST Challenge Videos

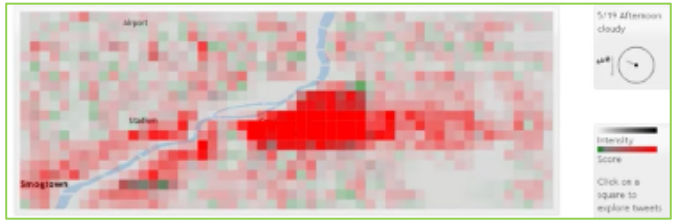
- Closer to the brief of this project i.e. novel implementation of visualization of some specific data
- Mostly longer than 2 mins
- A lot detail of also the user process of analyzing the data (not really expected in this assignment)

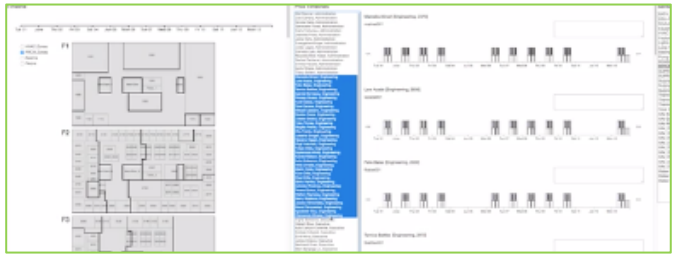












TIPS ON CREATING THE VIDEO [THESE ARE JUST SUGGESTIONS. ALL OPTIONAL]

One of the simplest options for this projects is to create a PowerPoint presentation and save this as a video.

- Here's one of many tutorials on the web that show you how to do this (ignore the second half transitions are really not needed)
  - https://helpdeskgeek.com/office-tips/how-to-turn-a-powerpoint-presentation-into-a-video/
- You should add a voice-over or captions / subtitles to describe the video

Some free screen capture options:

- Microsoft Stream [through browser, any platform] – free with TCD id at https://stream.microsoft.com
- Quicktime [Mac – bundled by default] see; https://libguides.rowan.edu/c.php?g=248114&p=4711659
- Xbox Game bar [free on windows]
- The following are other suggested 3rd party software that can be used free (with watermark, this is fine)
  - Camtasia: https://www.techsmith.com/download/camtasia/ [mac/windows] – free version with watermark
  - Bandicam: https://www.bandicam.com/ [windows] – free version with watermark
  - Fraps: http://www.fraps.com/download.php [windows] – free version with watermark

For more advanced editing, there are loads of free tools out there BUT just keep it functional. I'm not really expecting a fancy video

- iMovie [mac]
- VirtualDub [windows]

DESIGNING EFFECTIVE VISUALIZATIONS | 21/11/2022

Part 2 : Avoiding Distortions | Graphical Integrity

PRINCIPLES OF GRAPHICAL EXCELLENCE

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

Graphical displays should:

- **show the data**
- induce viewer to think about **substance rather than the methodology**
- serve reasonably **clear purpose**: description, exploration, tabulation or decoration.

- **avoid distorting** data

- present **many numbers in a small space**
- **make large data sets coherent**
- reveal the data at several **levels of detail**

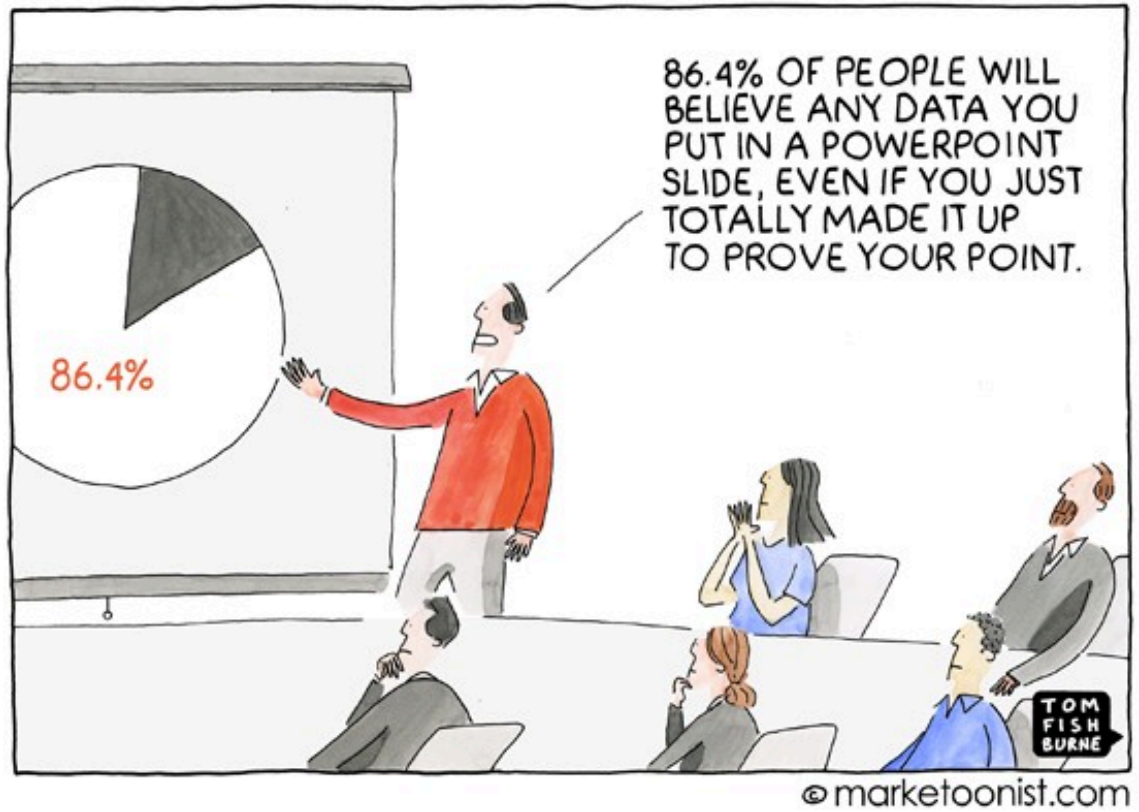
Address Objective

Avoid Distortion

Handle Complexity

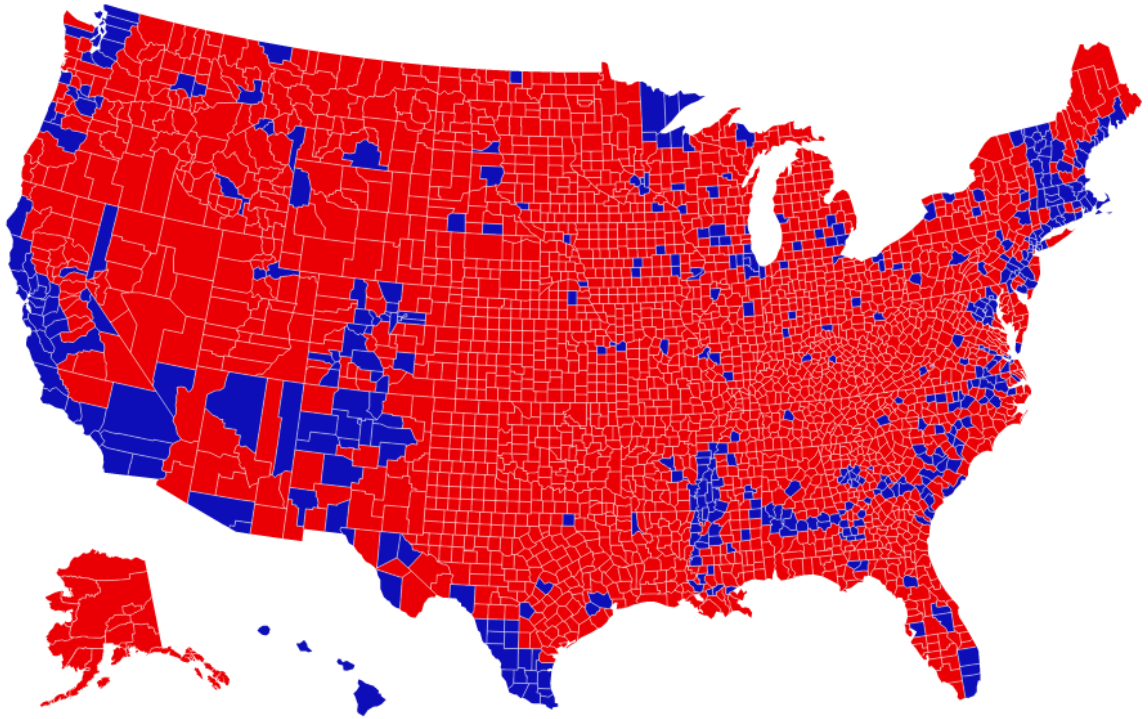
PRINCIPLES OF DATA INTEGRITY [TUFTE '83]

- 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



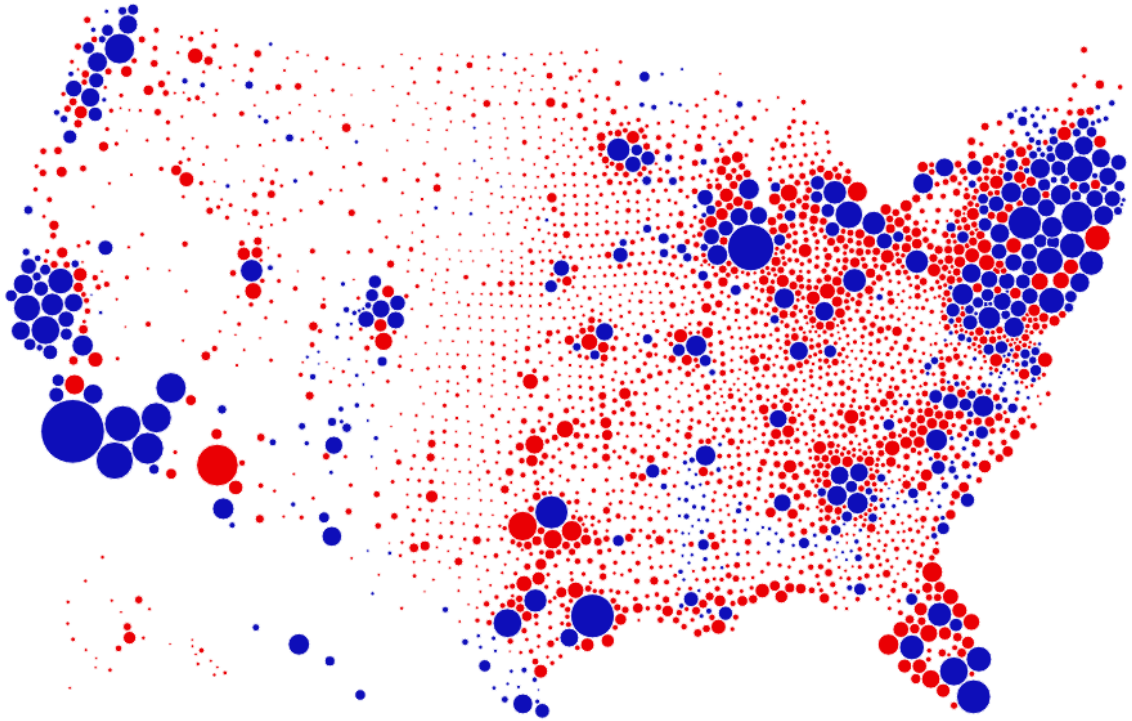
From: E. Tufte. The Visual Display of Quantitative Information (1983)

DISTORTIONS IN PERCEPTION



Binary Choropleth Map

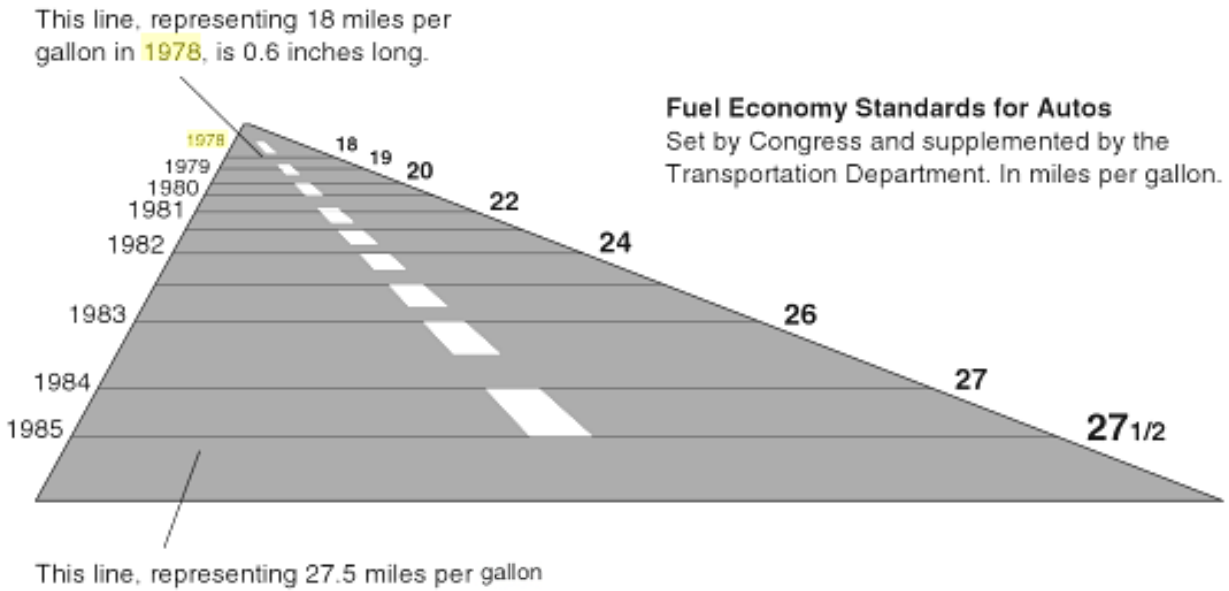
2016 US election result by county © HuffPost

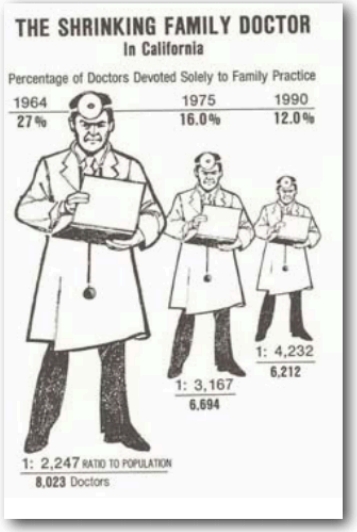


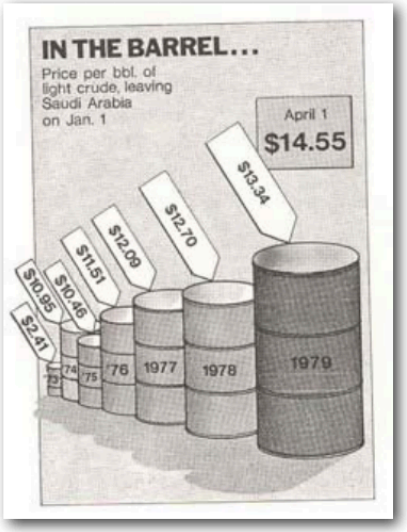
Cartogram/bubble chart

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

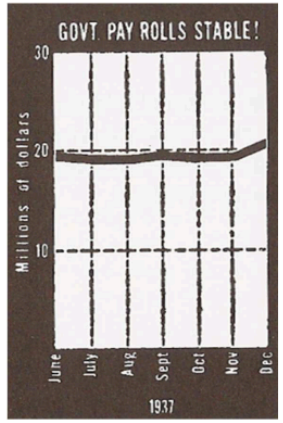
Lie factor: 9.5

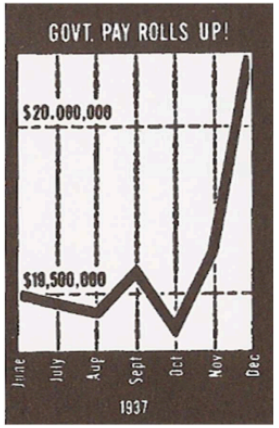
For the above example: $Lie\ factor = \frac{5.3-0.6}{0.6} / \frac{27.5-18}{18} = 14.8$

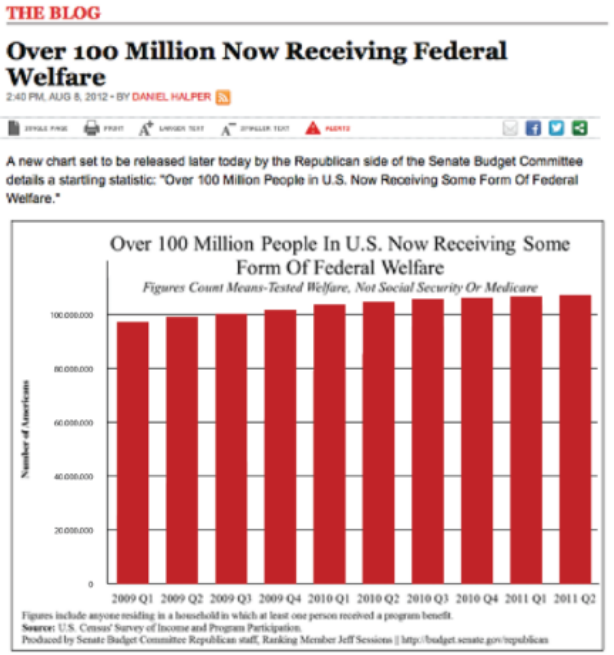
RANGE DISTORTION

Using unexpected data ranges may mislead readers.

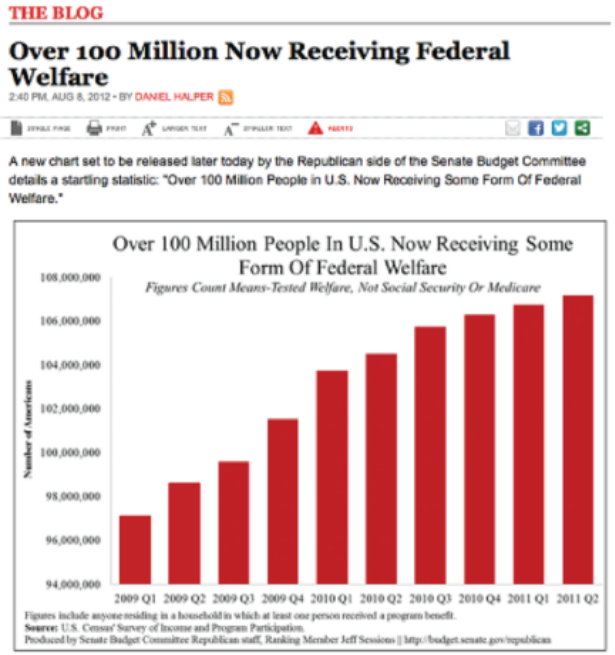
- Relative judgement is perceptually strong component of visual system.
- Need to balance with reducing screen space usage.



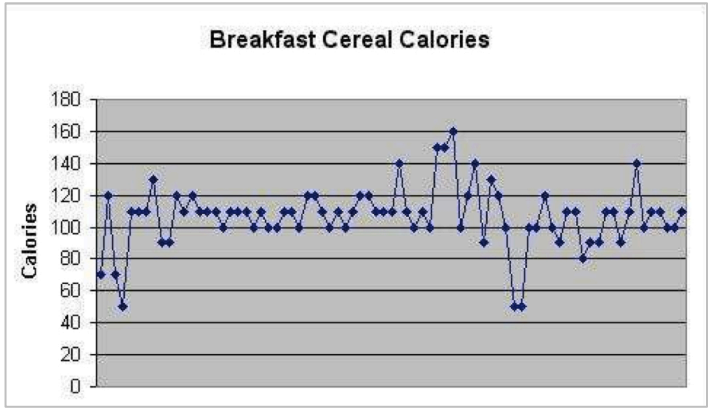


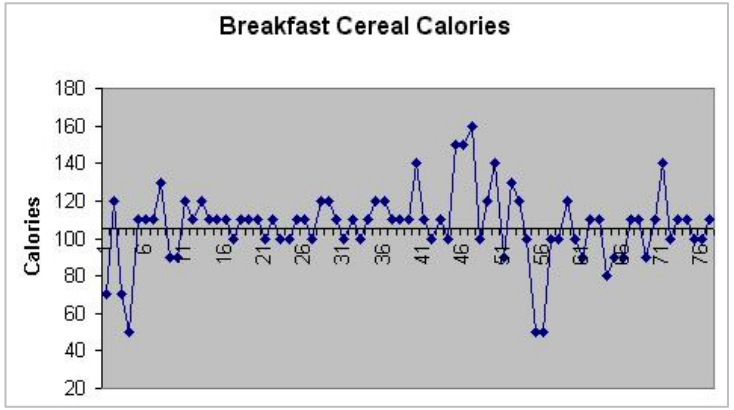


lie factor: 1



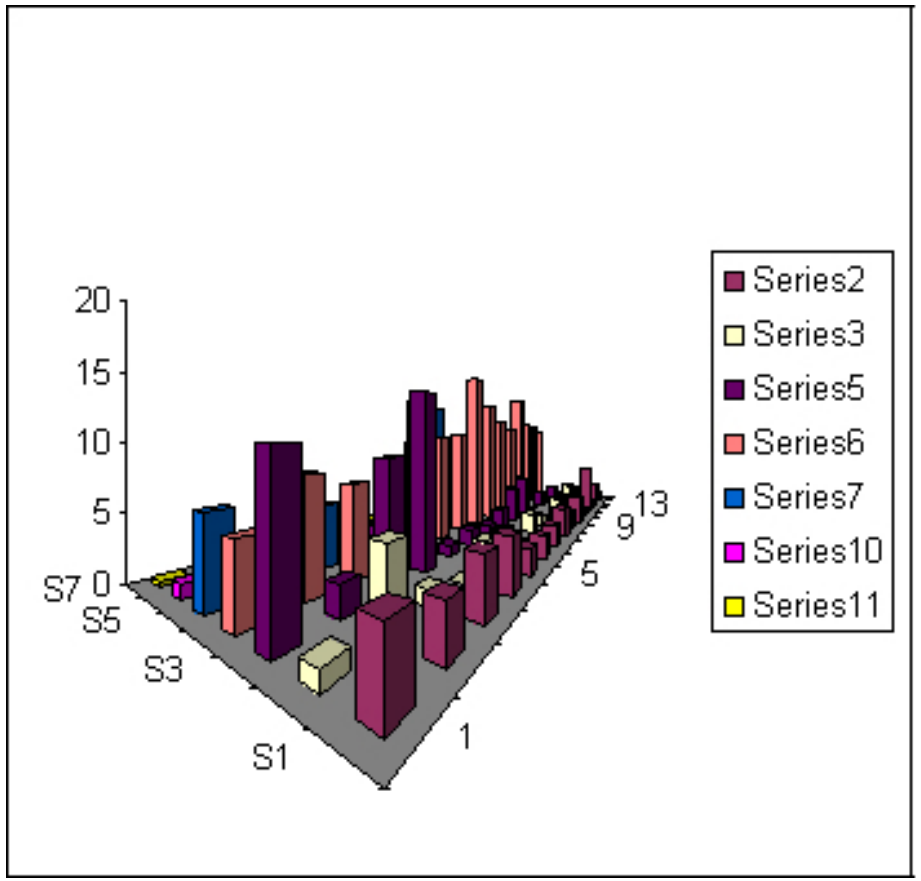
lie factor: 16,08

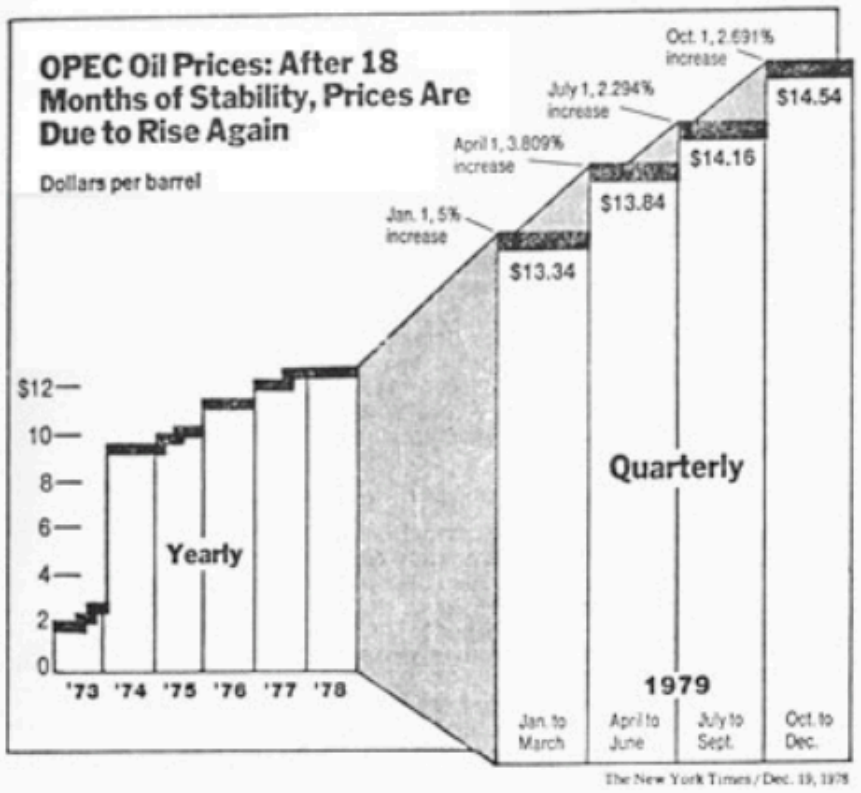




UNBALANCED SCALING

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**

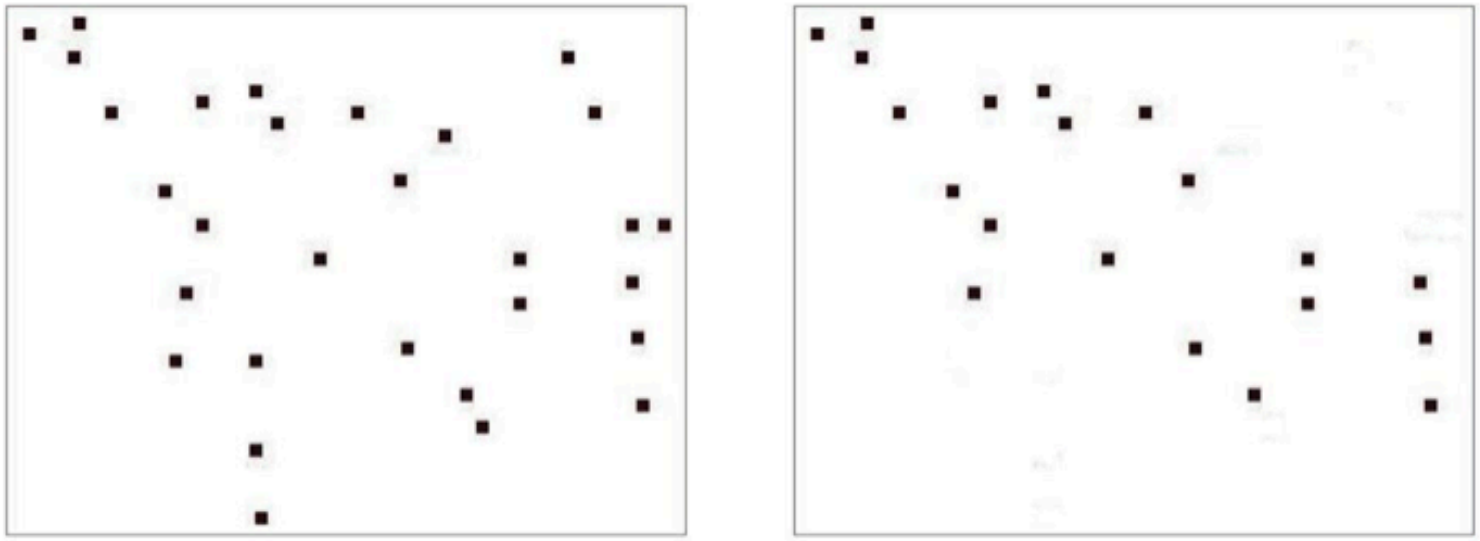


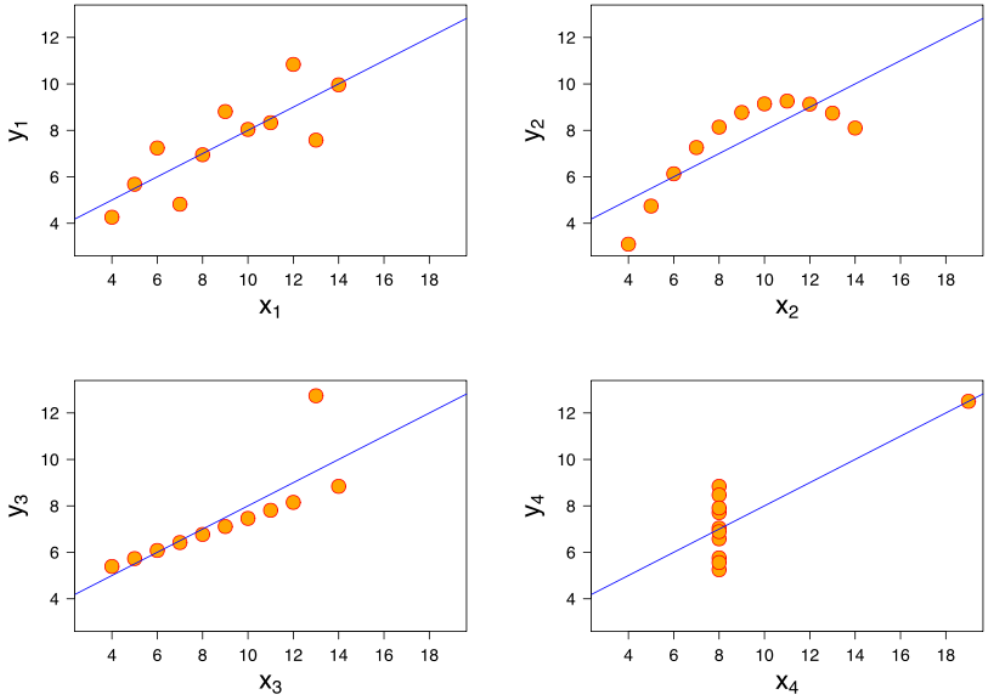
Figure 13.15. The problem with data scrubbing: (a) raw data showing lack of correlation; (b) scrubbed data revealing false correlation.

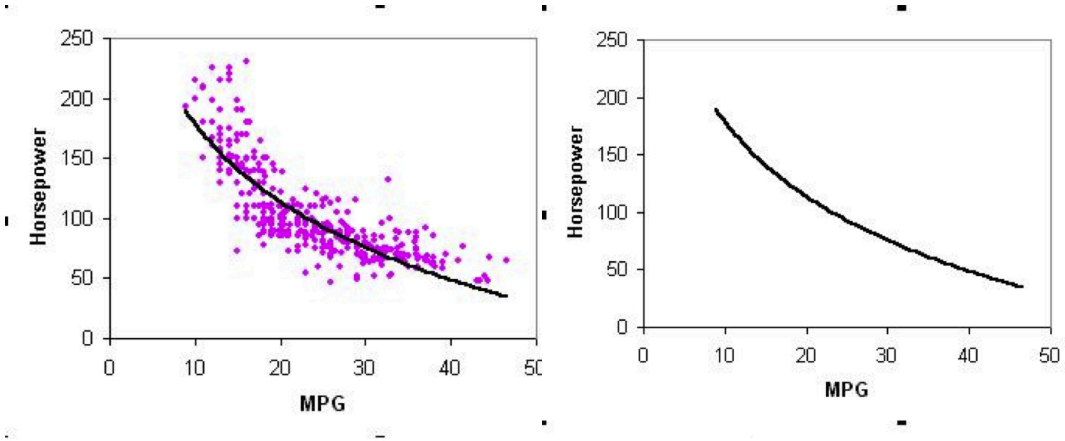
RAW VS. DERIVED DATA

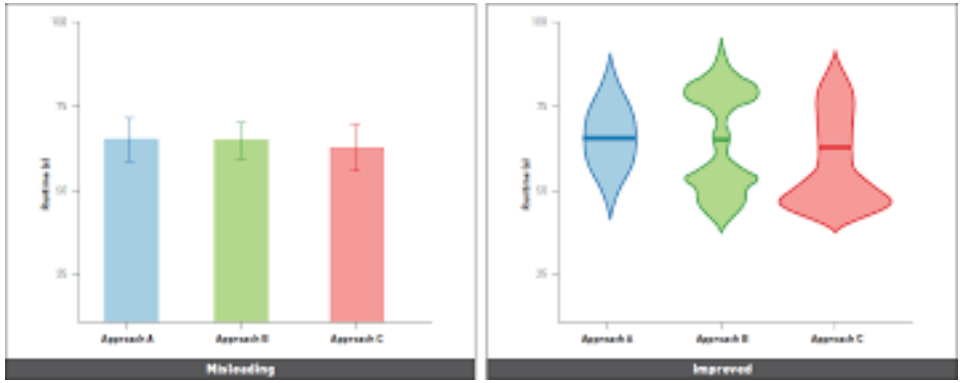
Showing only derived attributes can mislead readers

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

Anscombe (1973)







INSUFFICIENT SAMPLING

Uniform sampling

Contour info used to add sample points where significant changes occur

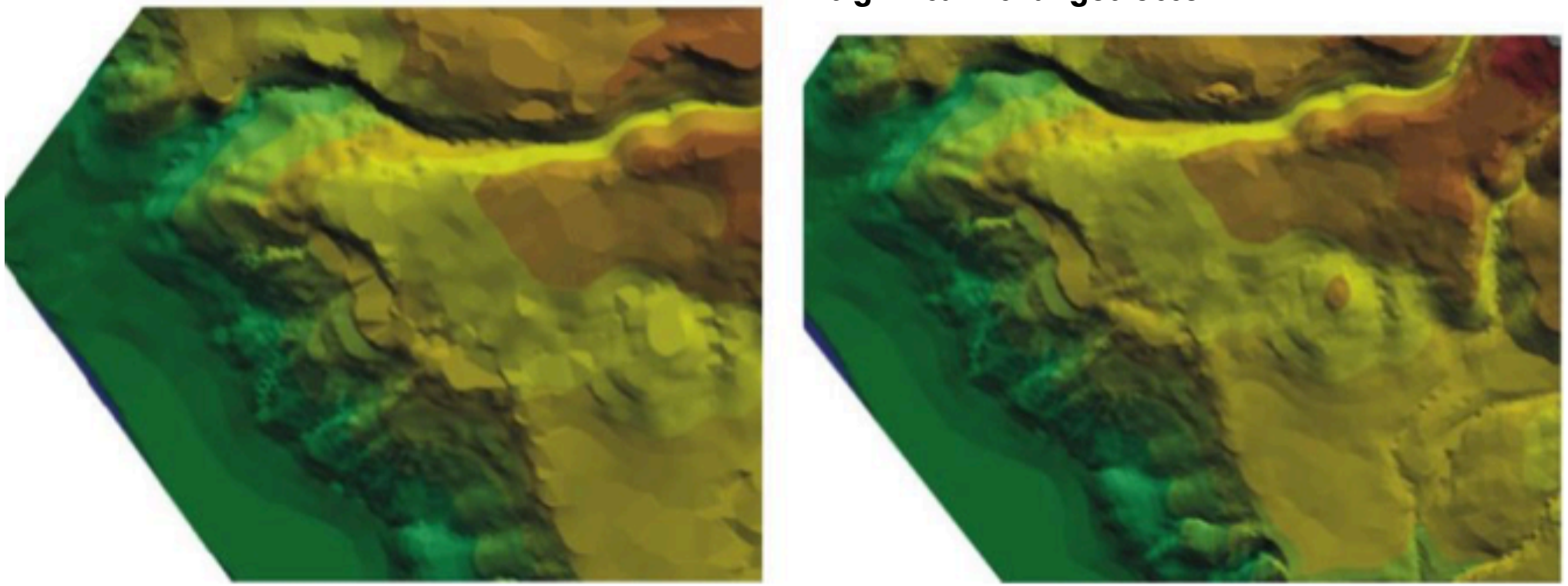


Figure 13.21. Different sampling and interpolation of the same data set. Some of the details in the right image are not seen in the left image.

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

In signal processing, the Nyquist rate, twice the highest frequency of a given function is required to avoid under-sampling (aka aliasing)

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

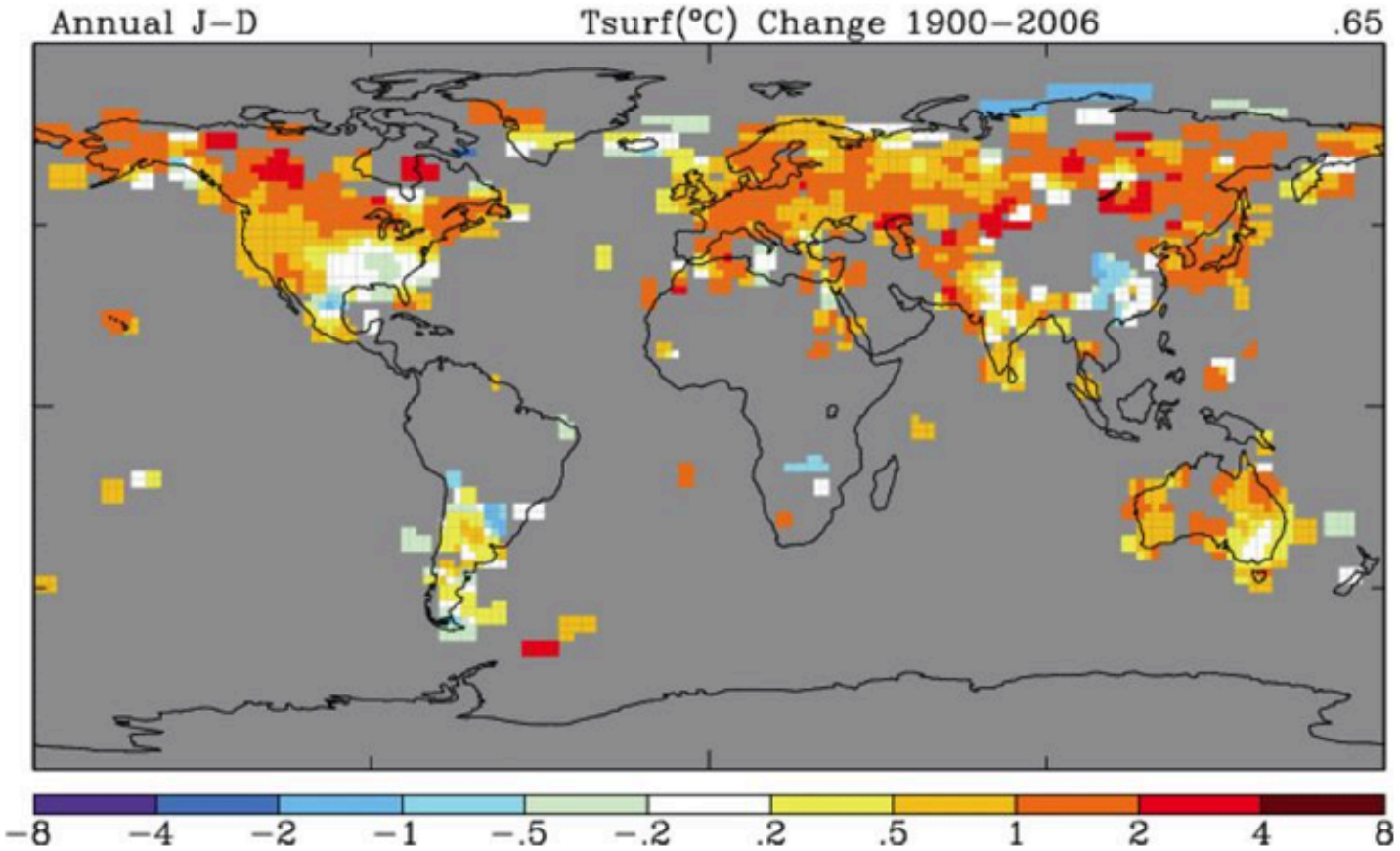


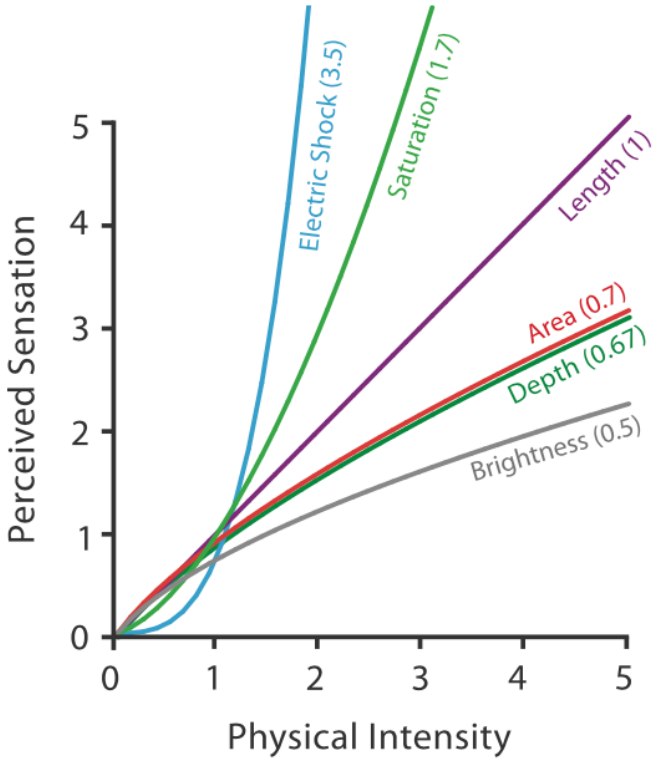
Figure 13.20. Sparse global temperature change data would give erroneous values for most of the planet if interpolated.

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

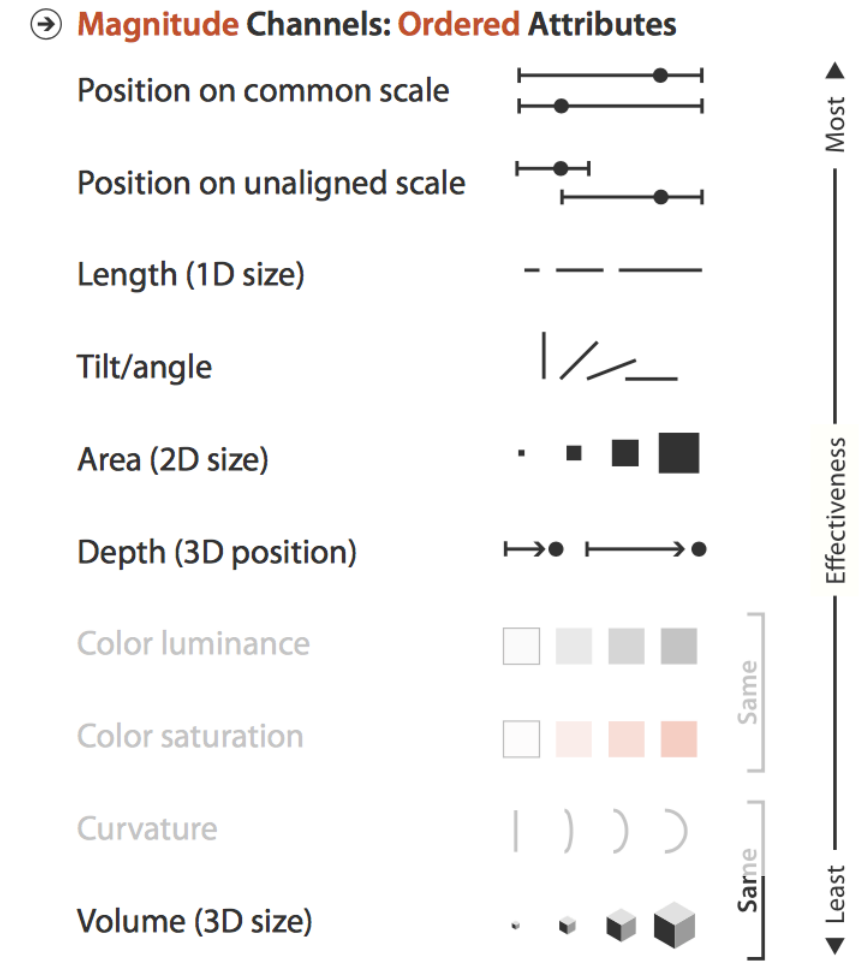
AVOID UNJUSTIFIED 3D

Spatial position/size channels are highly ranked:

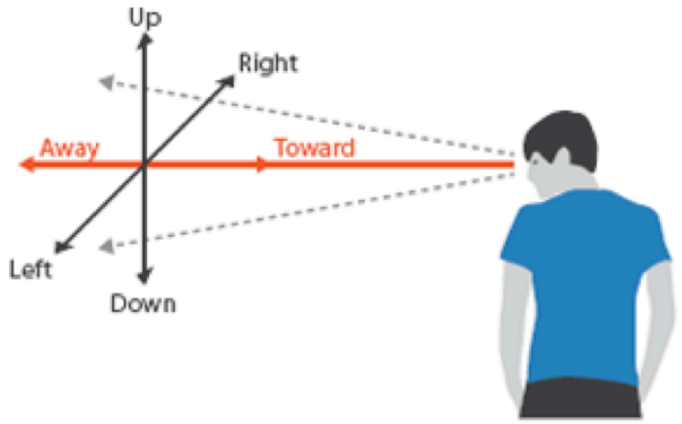
- But this pertains to planar (2D) spatial position
- Area, depth and volume perception is much less precise



Steven's Law



3D DEPTH AMBIGUITY



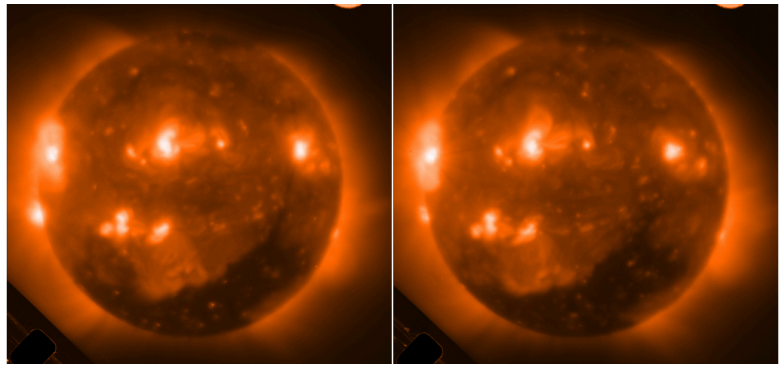
We don't perceive in 3D: but in 2.05D [Ware 08]

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

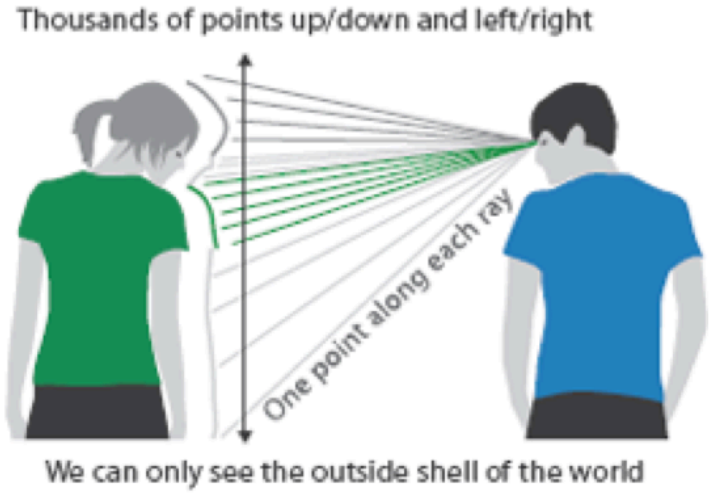


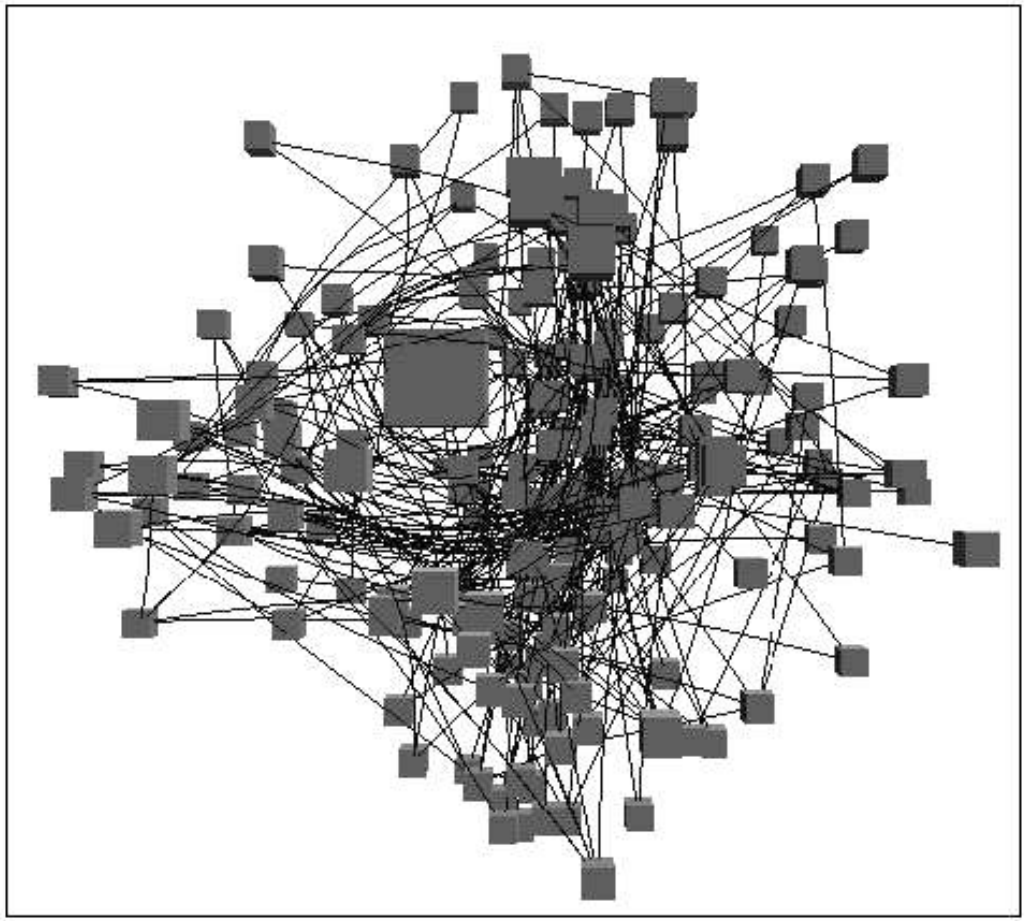
3D OCCLUSION

Occlusion is an important depth cue

However

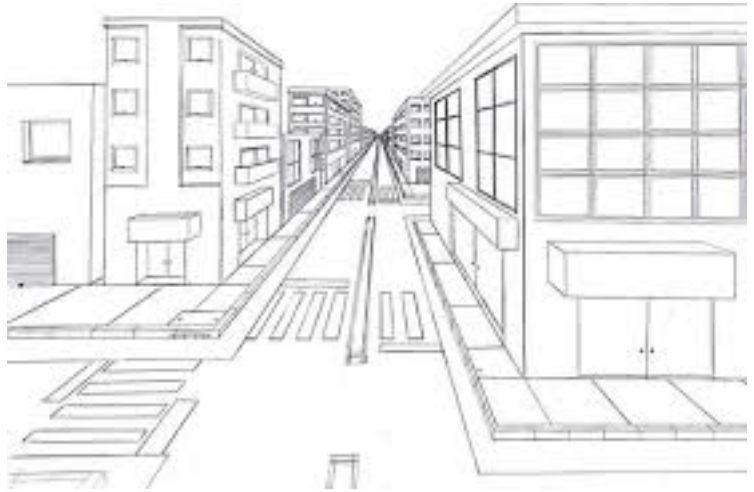
- It hides information in the background
- Leads to interaction complexity
  - e.g. the picking/selection problem in 3D





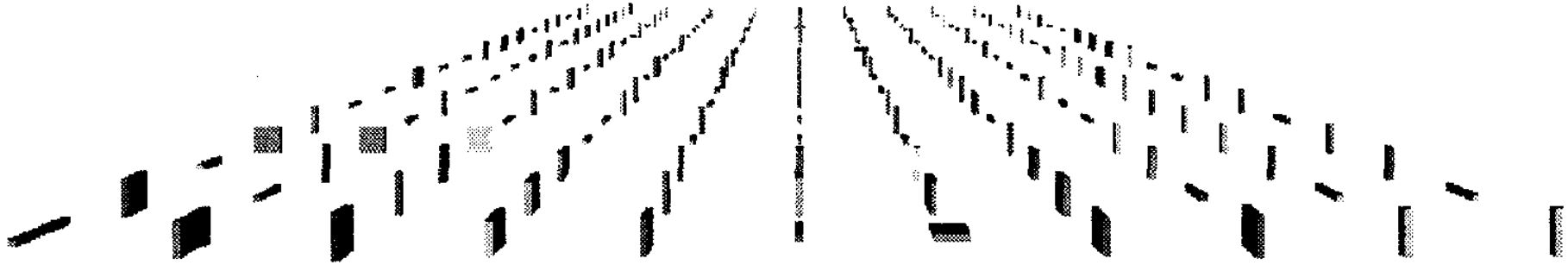
Carpendale 1996

3D PERSPECTIVE DISTORTION



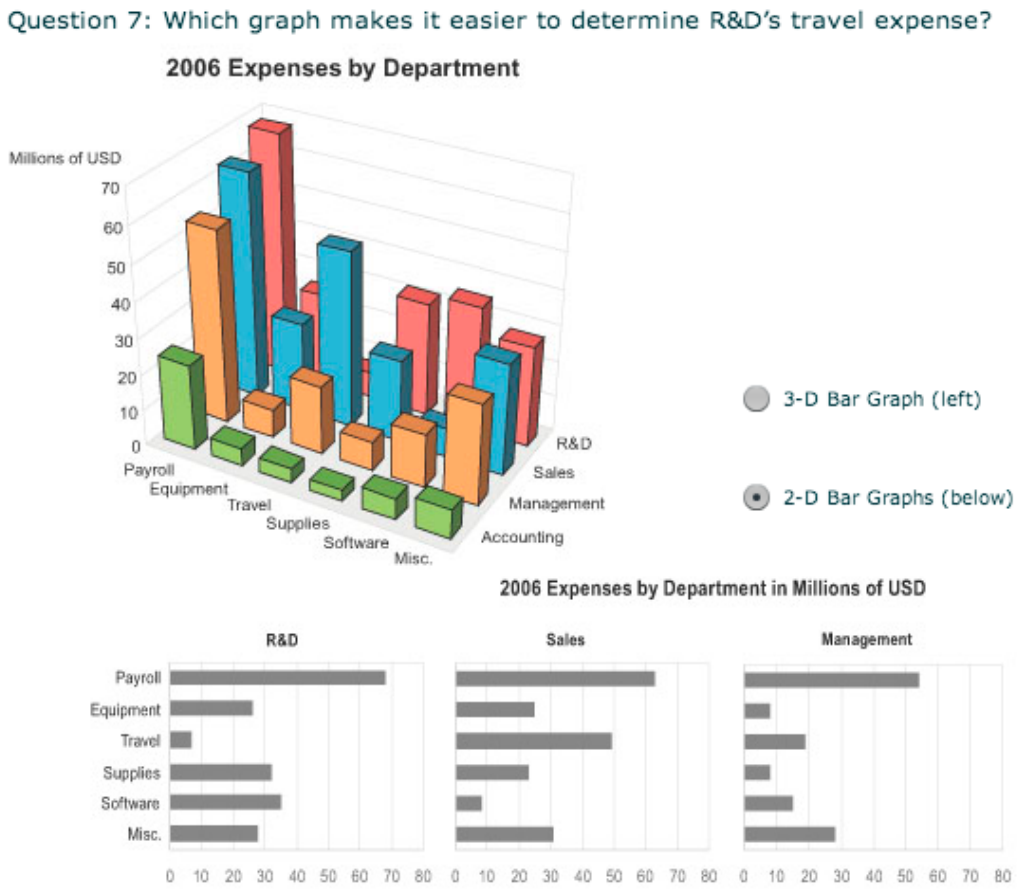
Perspective distortion

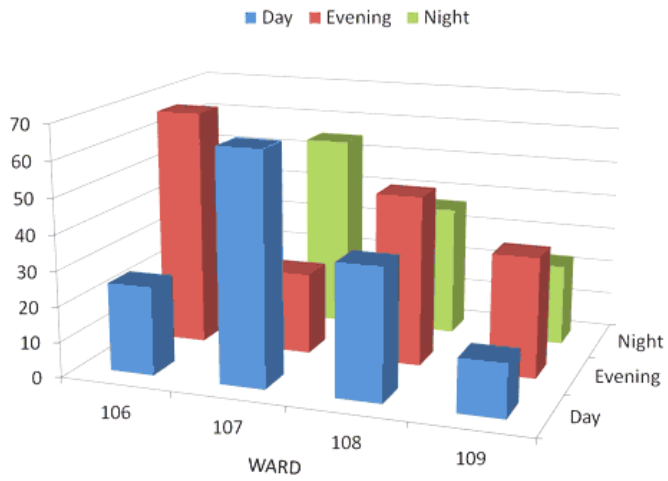
- interferes with all size channel encodings
- power of spatial position / size is lost



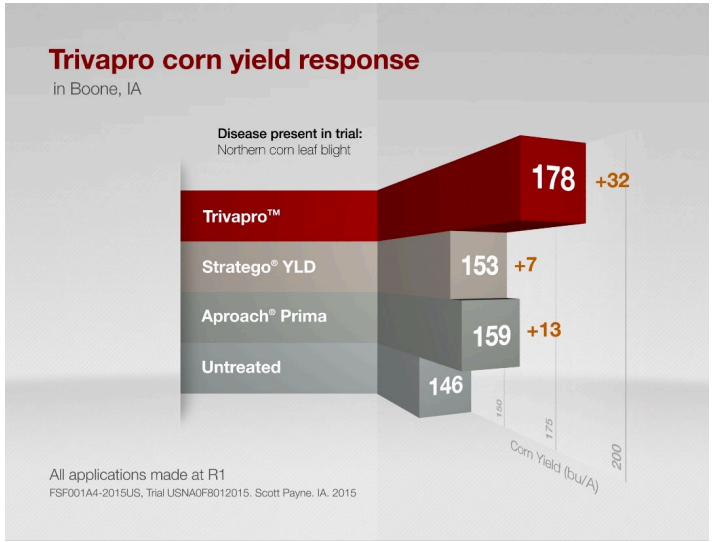
[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

INEFFECTIVE 3D BAR CHARTS

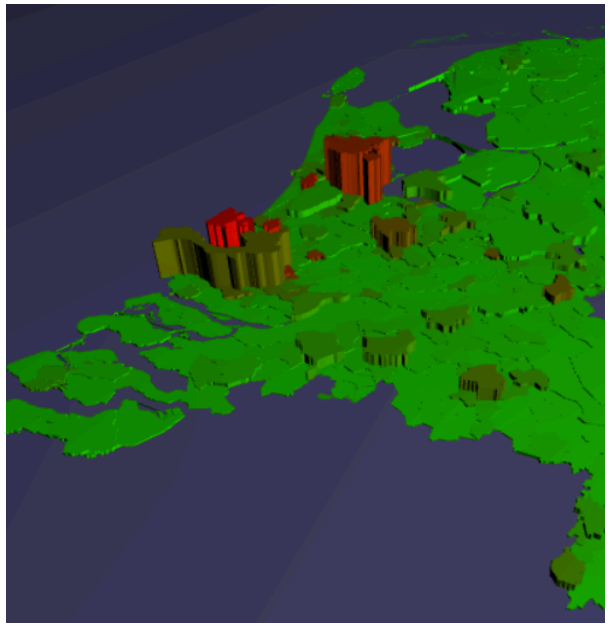


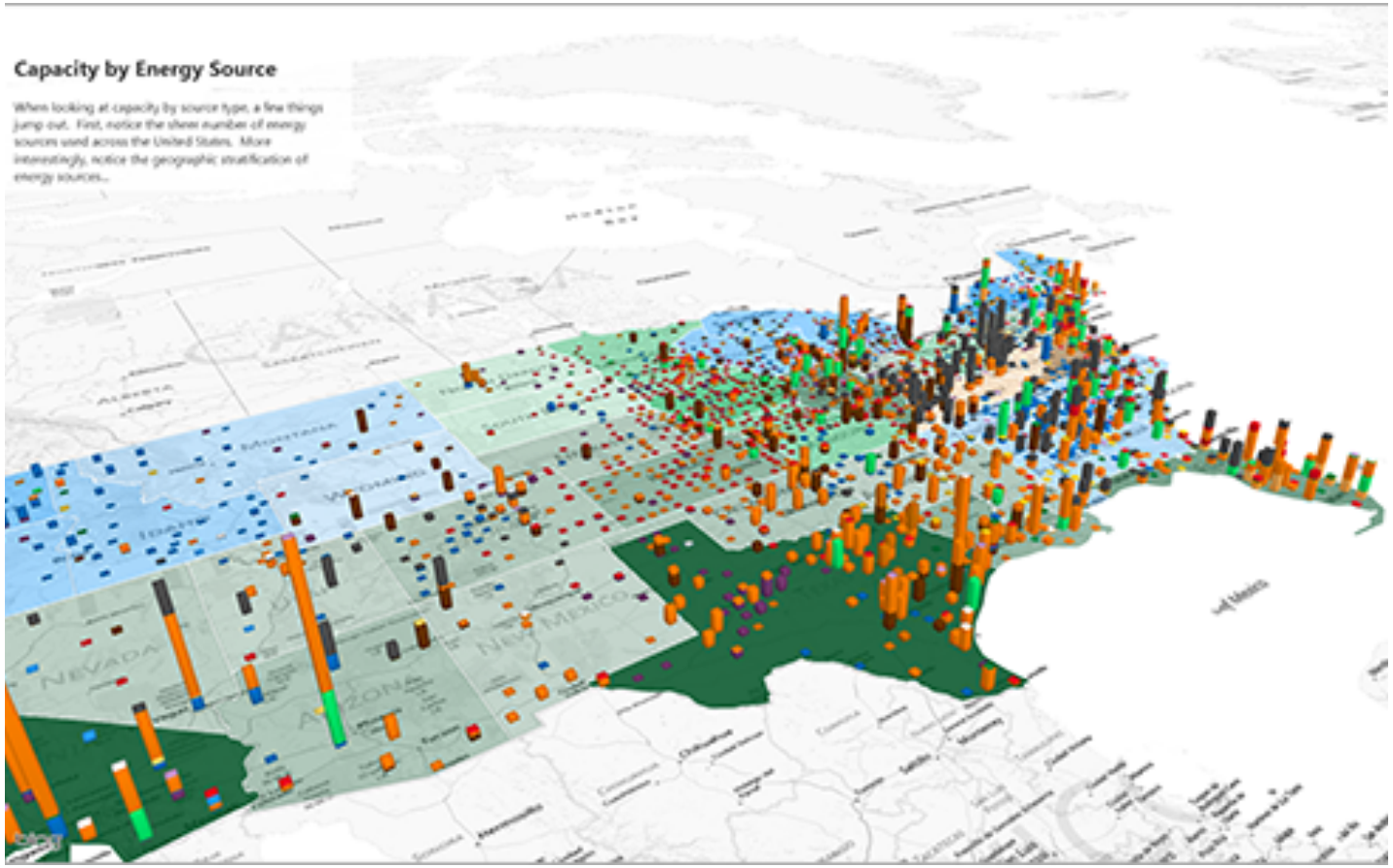


Is the second blue bar touching the 50 line?

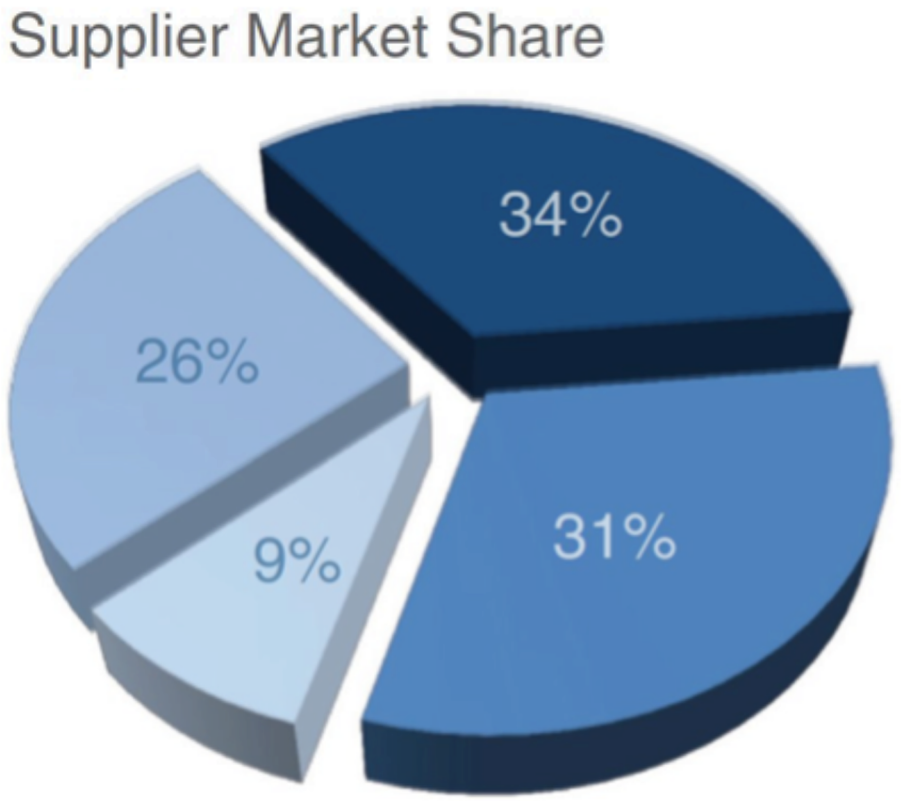


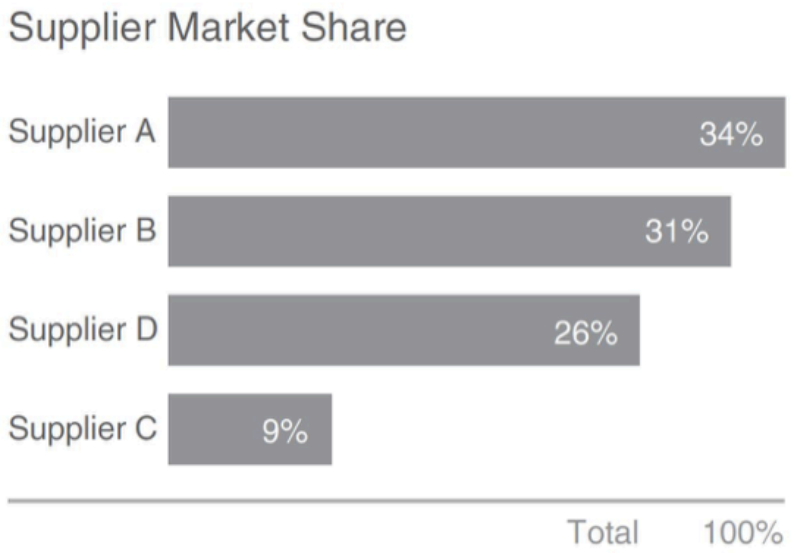
IMPRECISE 3D VOLUME PERCEPTION





INEFFECTIVE 3D PIE CHARTS

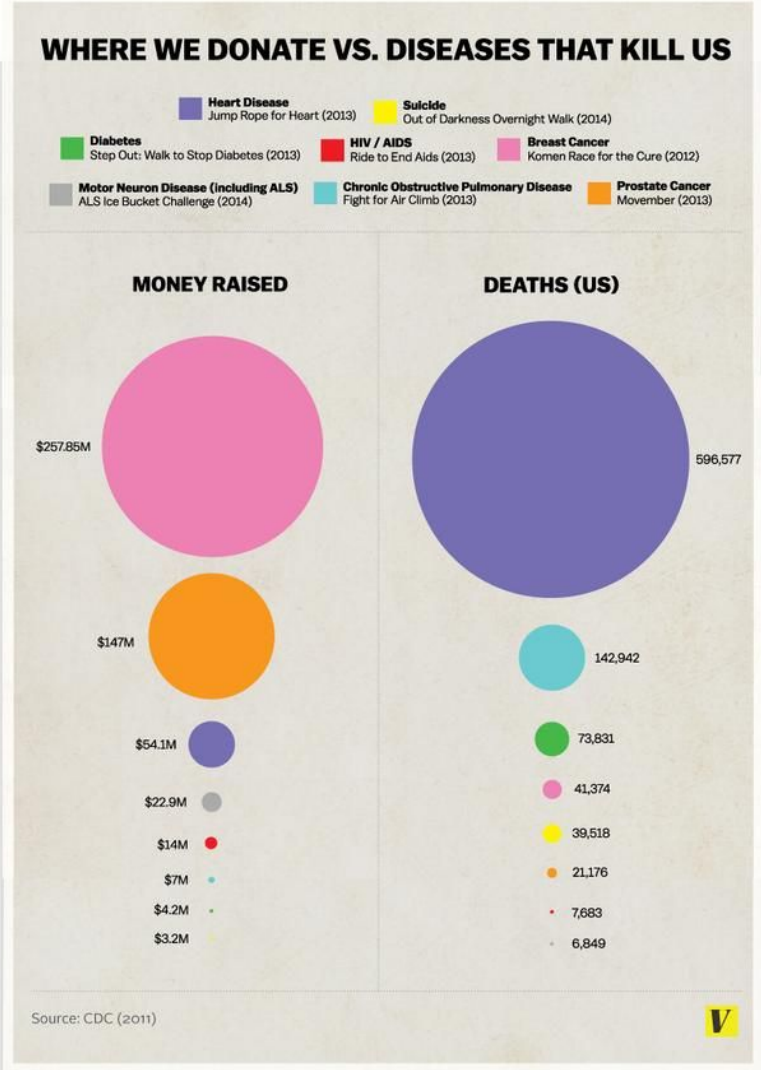


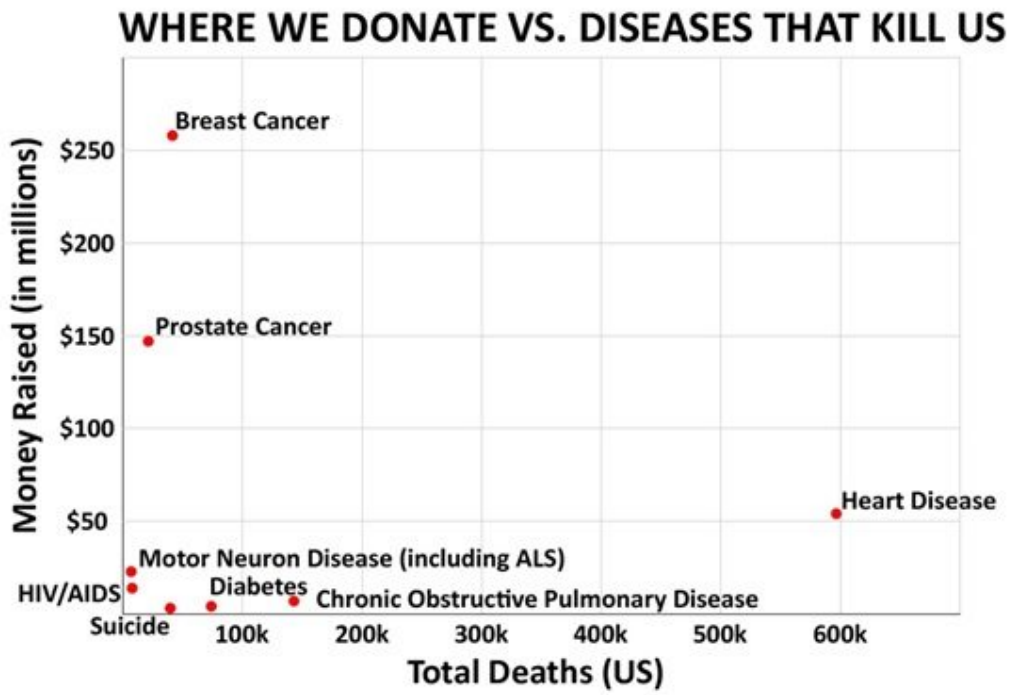


Supplier A looks misleadingly bigger than Supplier B

AVOID UNJUSTIFIED 2D

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

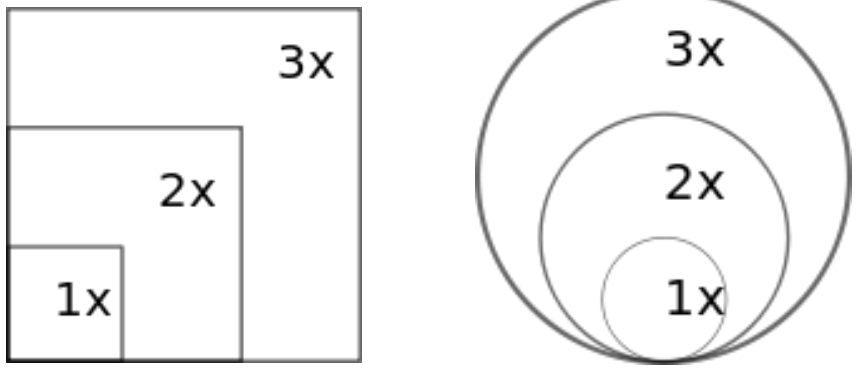




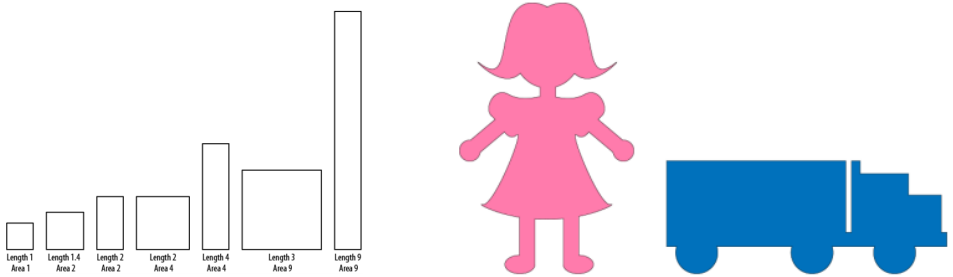
Easy to see which is bigger in the bubble chart (left)
but "how much bigger" is clearer in the simple scatter plot to (right)

AVOID UNJUSTIFIED 2D

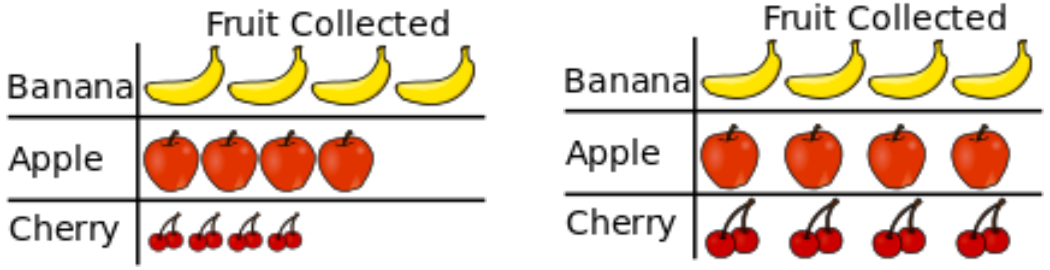
Avoid 2D area if 1D length will do



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

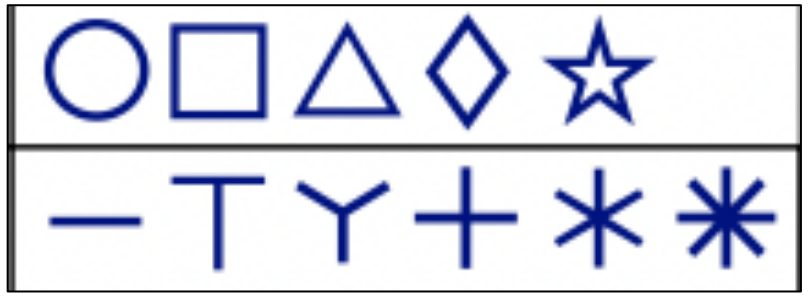


Avoid 2D glyphs if area is not constant



Shapes affect size perception

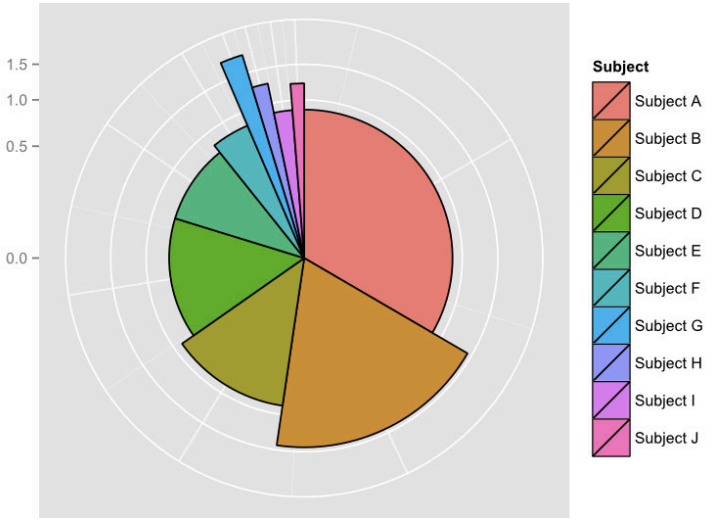
[Smart and Szafir 2019]

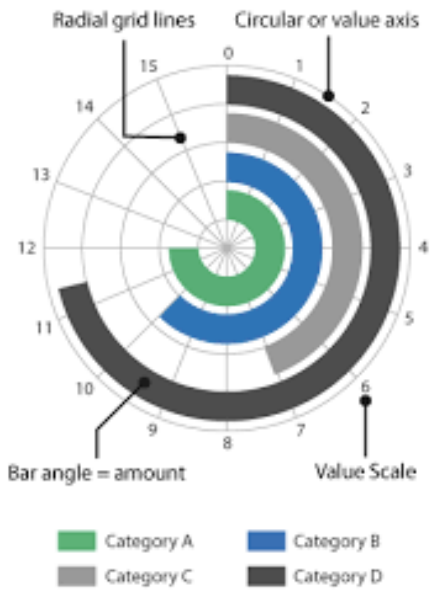


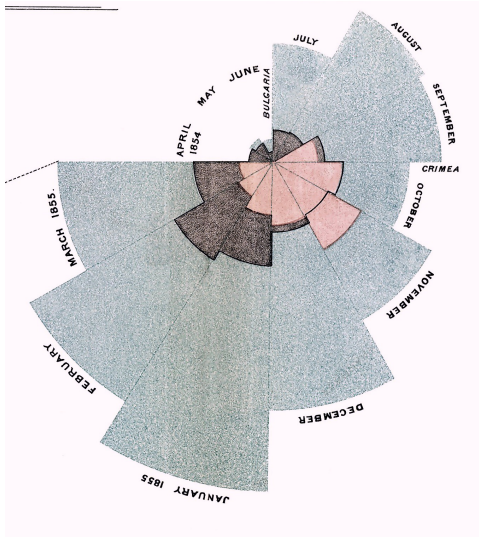
AVOID UNJUSTIFIED 2D

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

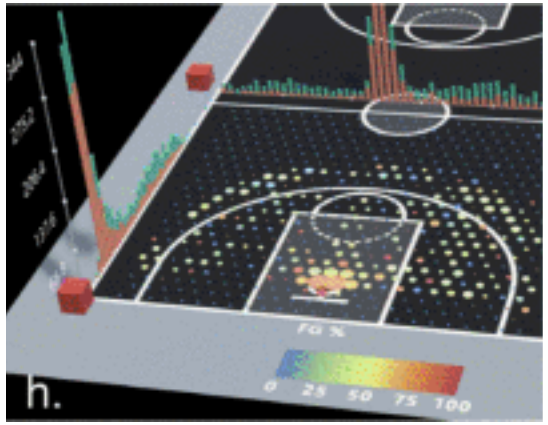


Image from Sicat et al (2019) DXR: a toolkit for immersive data visualization

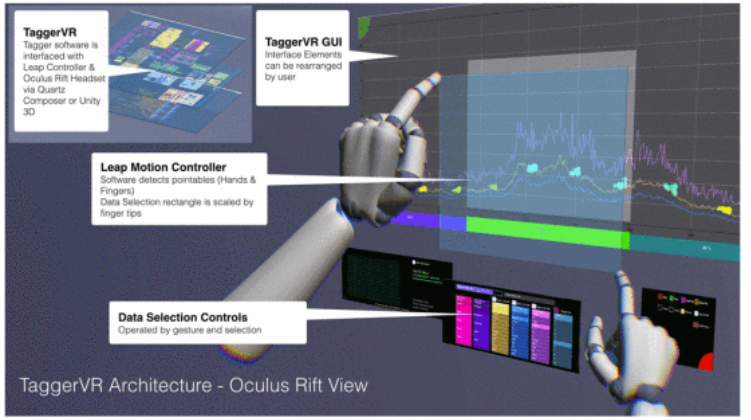
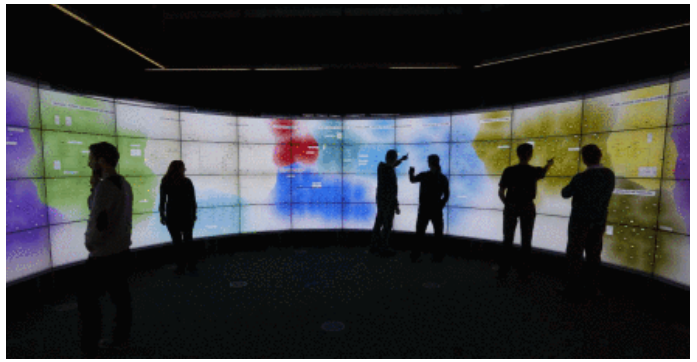


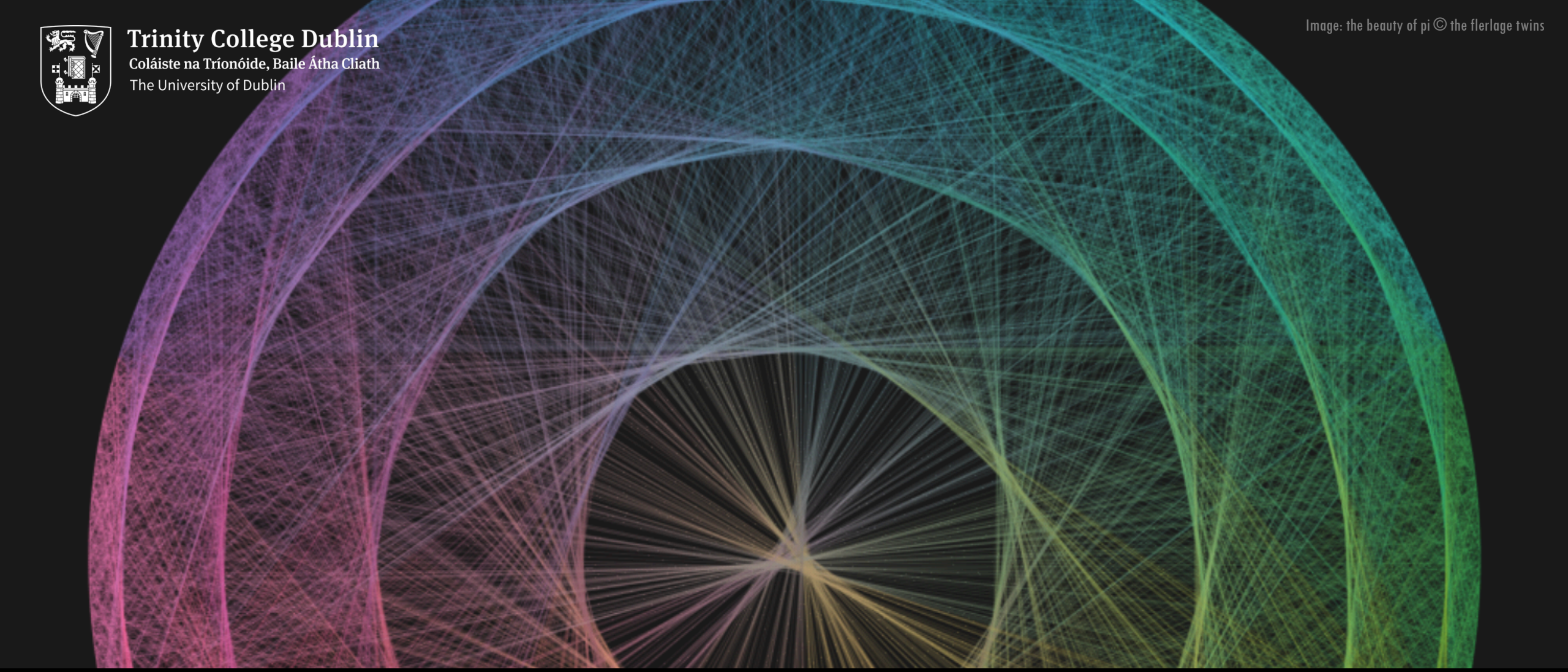
Image from Morse et al (2015) TaggerVR system for interactive Viusal Analytics



Big Data Immersive Analytics by Chandler et al (2015)

THE END

For now



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:

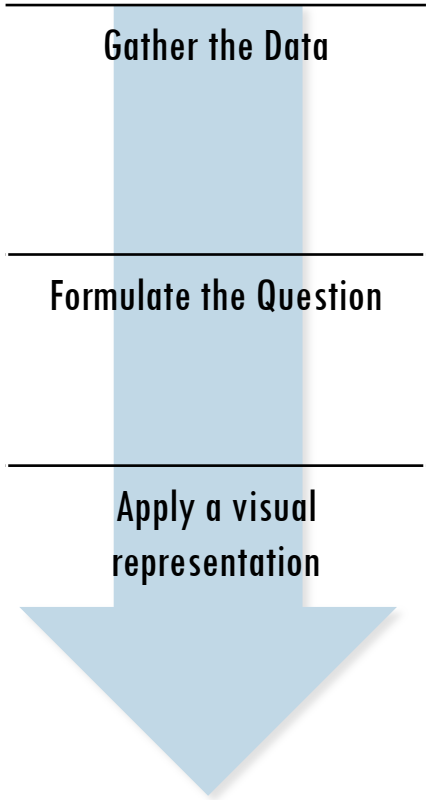
Understand the context	• Identify audience, what they need to know or do and why
Choose an effective visual	• decide how best to encode data
Eliminate clutter	• Remove anything not adding informative value
Focus attention	• Leverage pre-attentive perception to draw users attention to important parts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

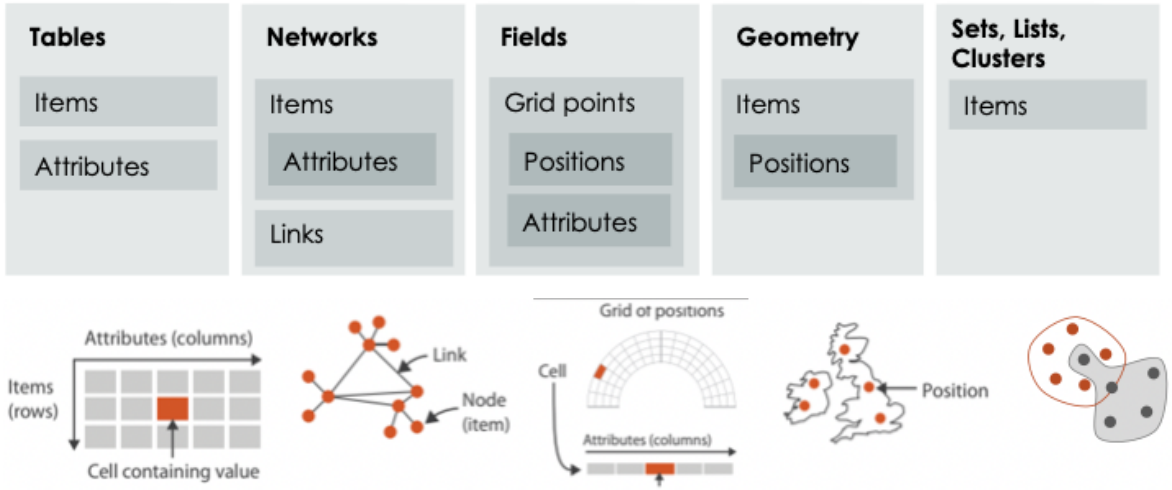
Acquire	Obtain the data
Parse	Provide some structure (attributes+data types)
Filter	Remove extraneous data
Mine	Apply math, stats & data mining to discern patterns
Represent	Choose idioms + encoding channels
Refine	Improve efficacy of representation; transform
Interact	Add methods for user to manipulate/control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



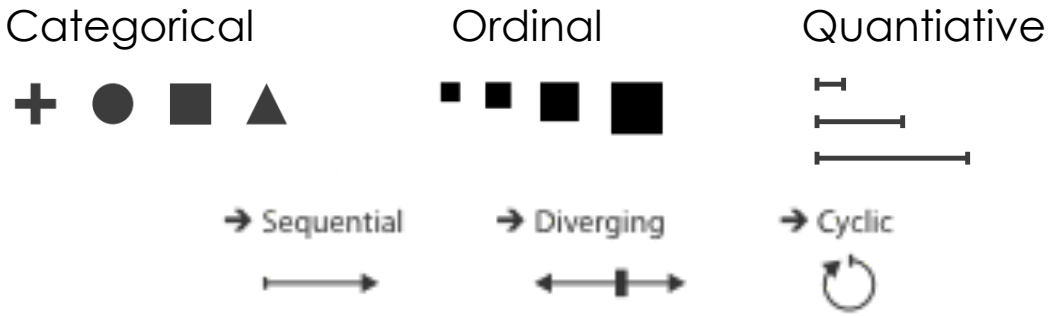
[Fry 2008] Ben Fry, Visualizing Data, O'Reilly Press, 2008

1: DATA

Data/Dataset Types



Attribute Types



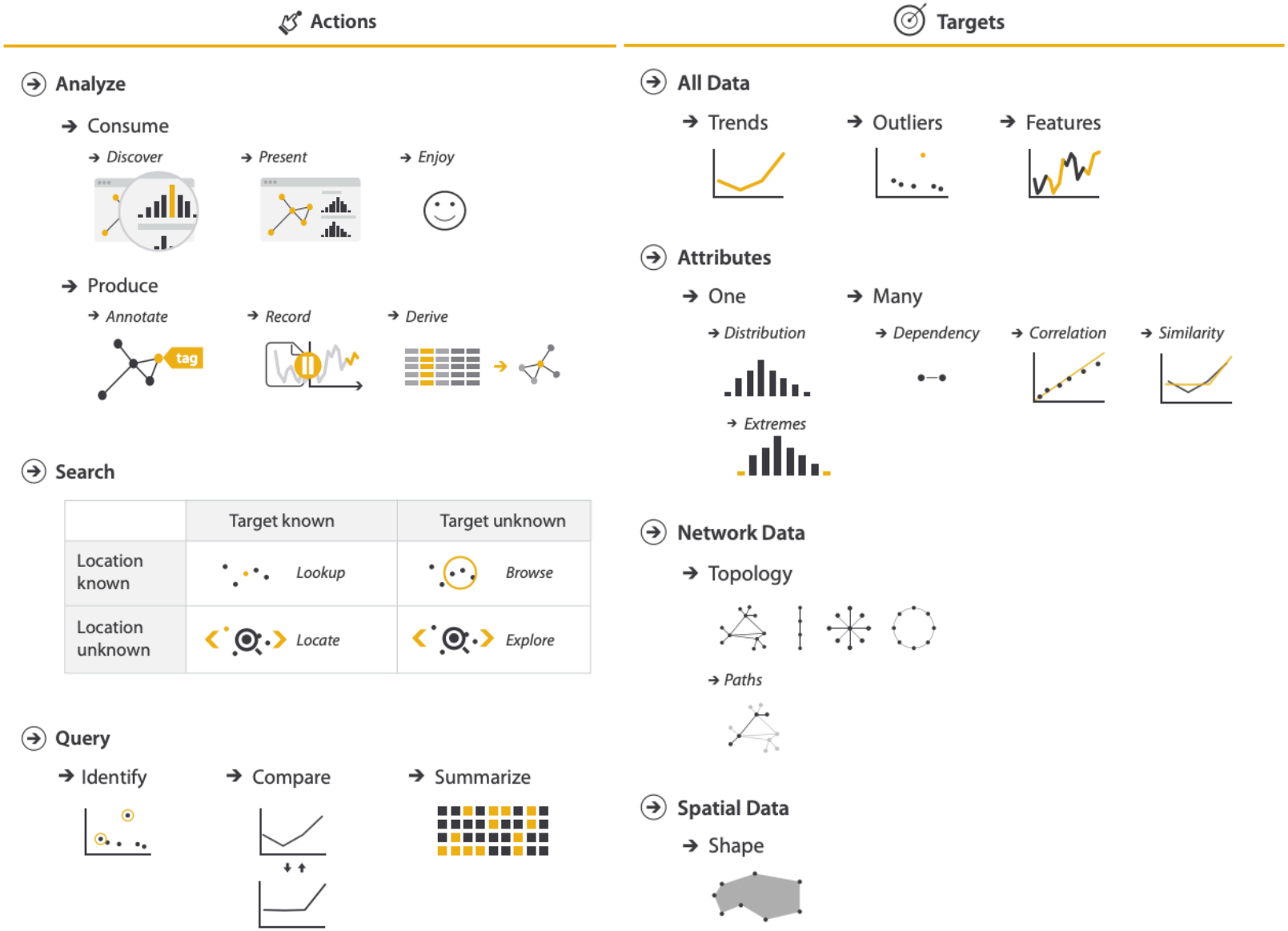
RECOMMENDED READING

- 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

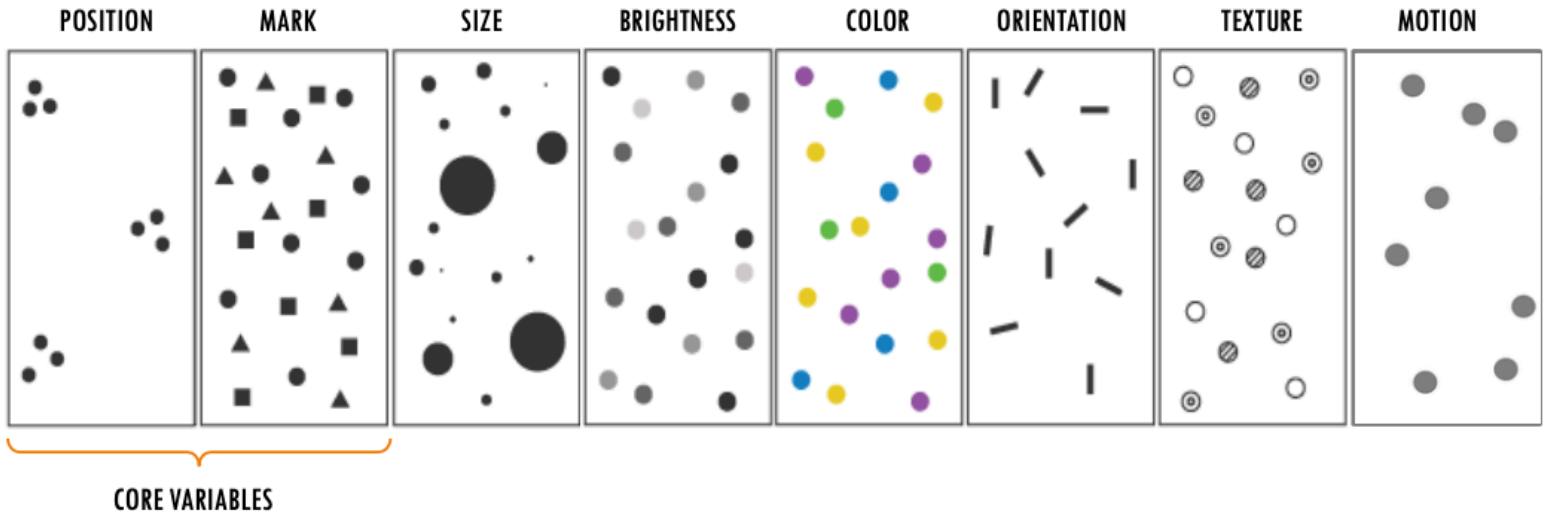
2: TASK

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



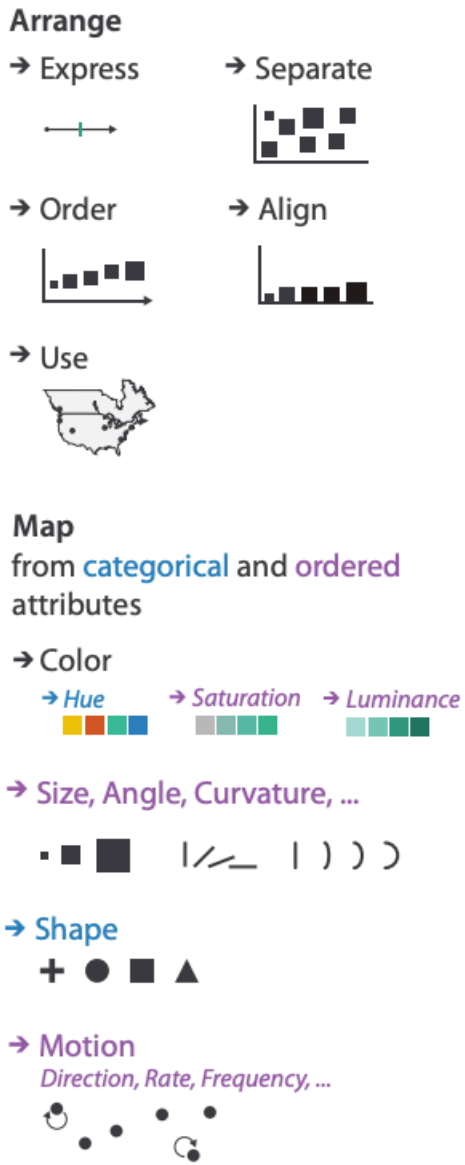
3: ENCODING CHANNELS



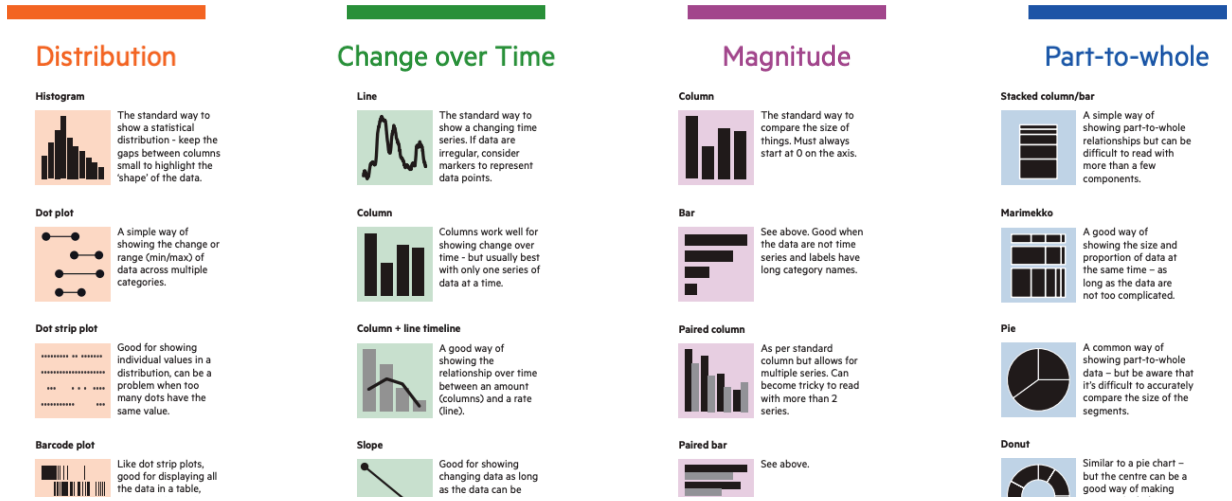
RECOMMMENDED READING

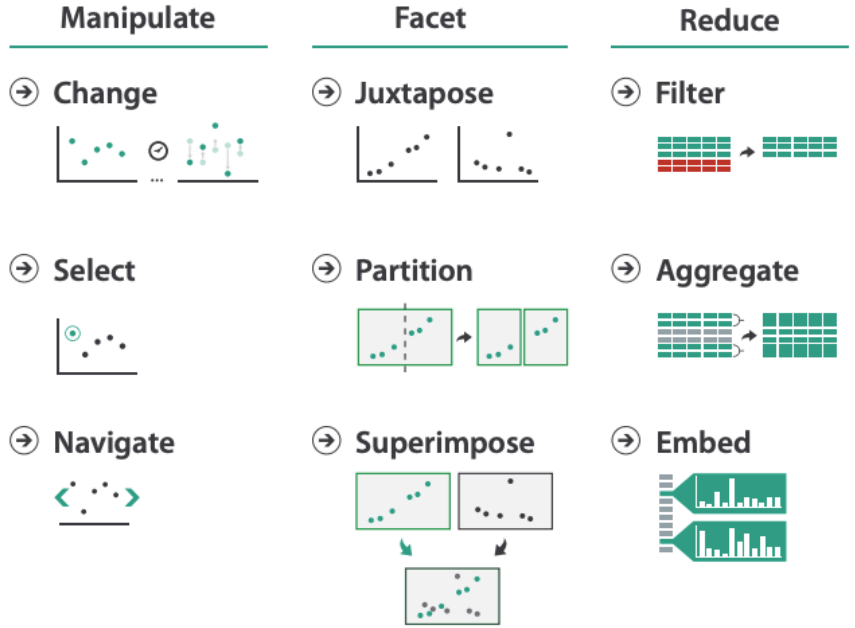
- CS7DS4 Lecture 4 : Visual Encoding
- Chapter 4 – "Visualization Foundations" in Interactive Data Visualization, Ward et al 2010
- [Free Alternative] Chapter 4 – "Visual Encoding Principles" in Information Visualization: Principles, Techniques and Practice, [Draft] Munzner 2011

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



4: IDIOMS

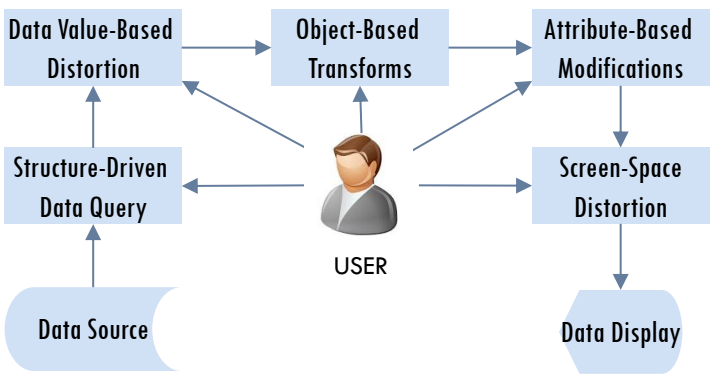




Too many to list. See something like the above cheat sheet by the Financial Times:
https://github.com/ft-interactive/chart-doctor/raw/master/visual-vocabulary/Visual-vocabulary.pdf

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



THE END

For now