Data Visualization for Biomedical Applications

Lecture 2

BMI706 - 29 March 2018

Nils Gehlenborg, PhD

Administrative

Piazza

https://piazza.com/harvard/spring2018/bmi706/home

Is everyone getting notifications?

Administrative

- Who worked on the homework assignment?

- Homework assignment 2 will be posted by the end of the day

    - Design Exercise: longitudinal cancer genomics

        - Implementation: create linked interactive views in R using `crosstalk`

- Start thinking about your project

    - Identify data sets

Review of Session 1

What is data visualization?

Computer-based visualization systems provide visual representations of datasets intended to help **people** carry out some task **more effectively**.

— *Tamara Munzner*

Why do we do data visualization?

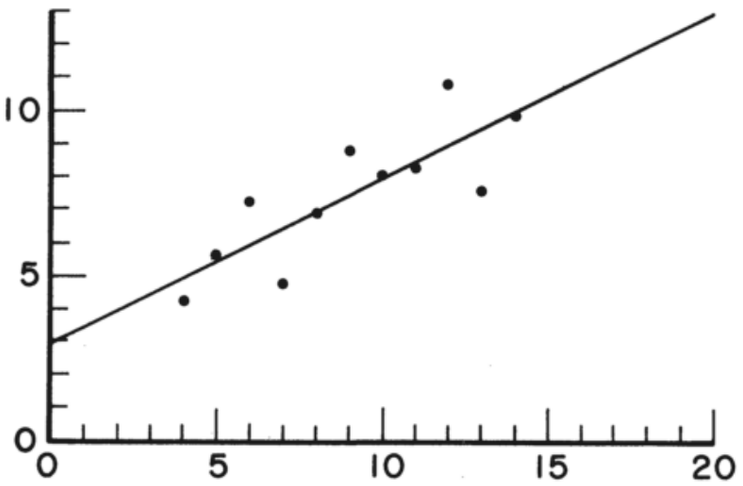


Figure 1

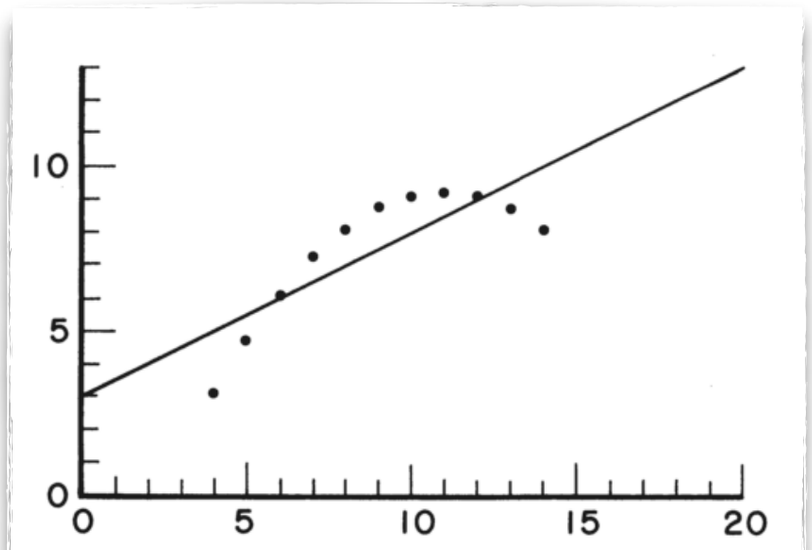


Figure 2

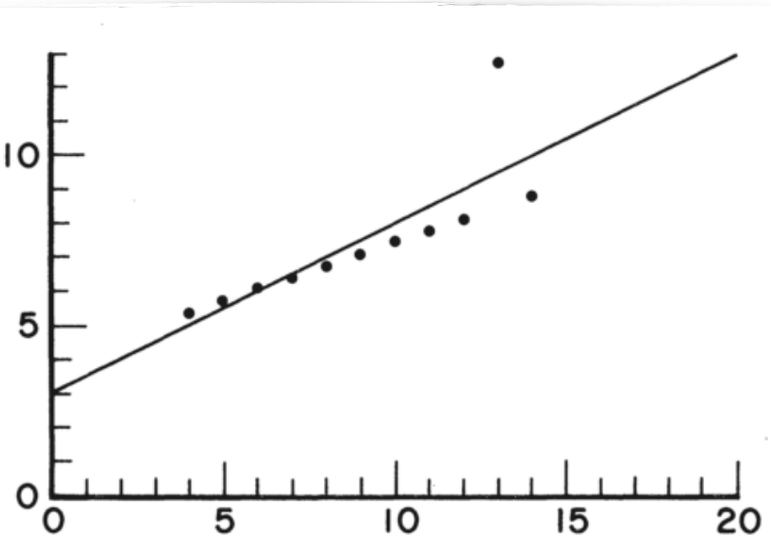


Figure 3

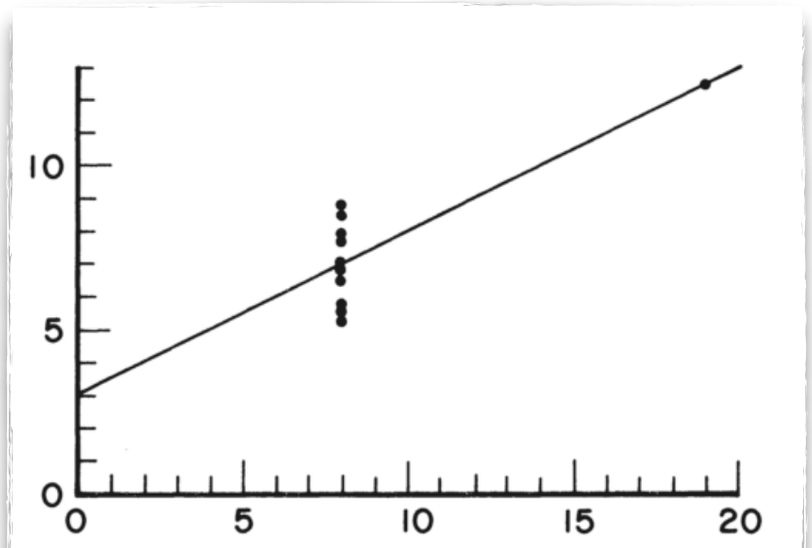


Figure 4

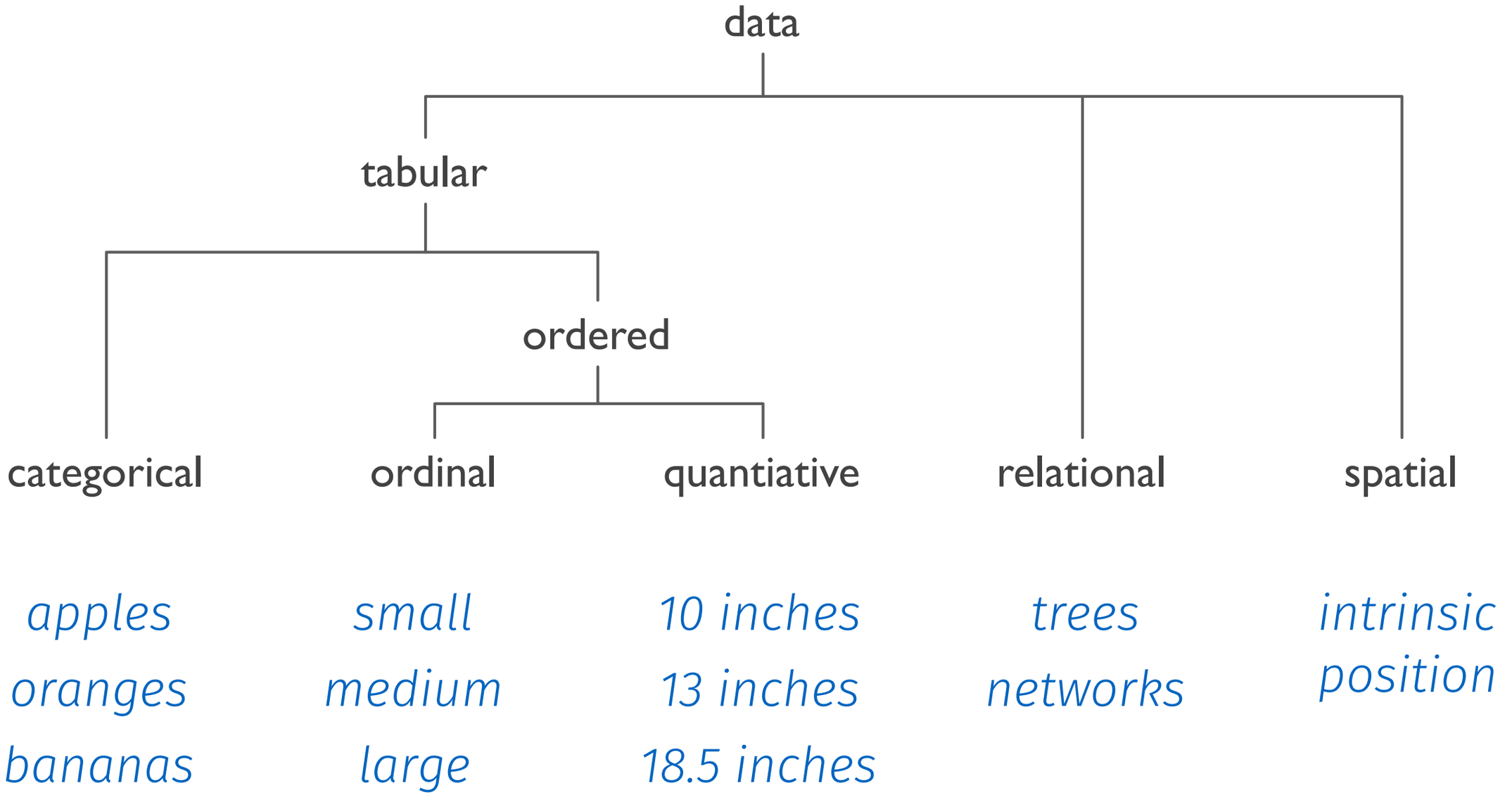
Why do we need a human in the loop?

How does visualization work?

453 x 862 = ?

```
   906
+ 27,180
+ 362,400
---------
  390,486
```

Visual Encoding of Data



slide adapted from Munzner 2011, Visualization Principles

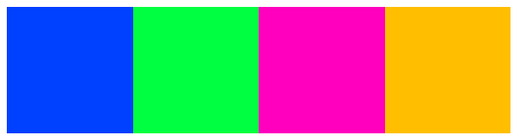
Lecture 2: Learning Goals

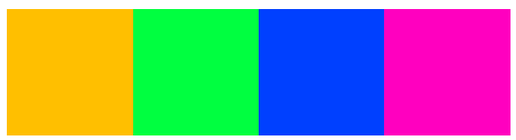
- How can visualizations fail?
- What are some common interaction techniques?
- What is the typical design process for a data visualization tool?
- How can we evaluate a visualization method or tool?

Visualization Pitfalls

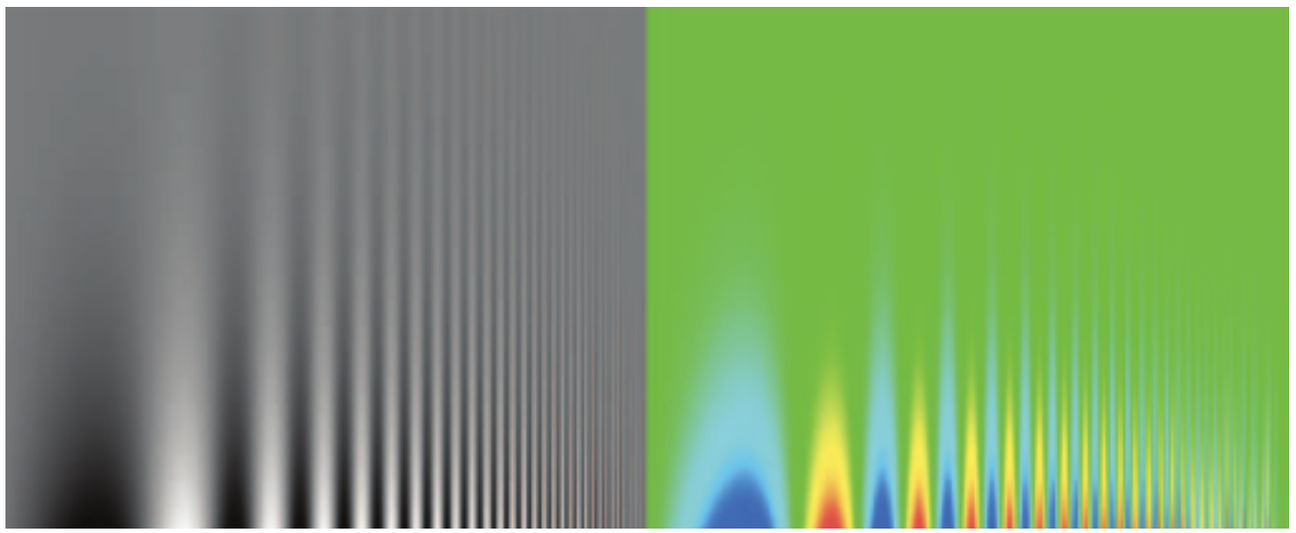
Color Pitfalls: Rainbow Color Map

hard to order



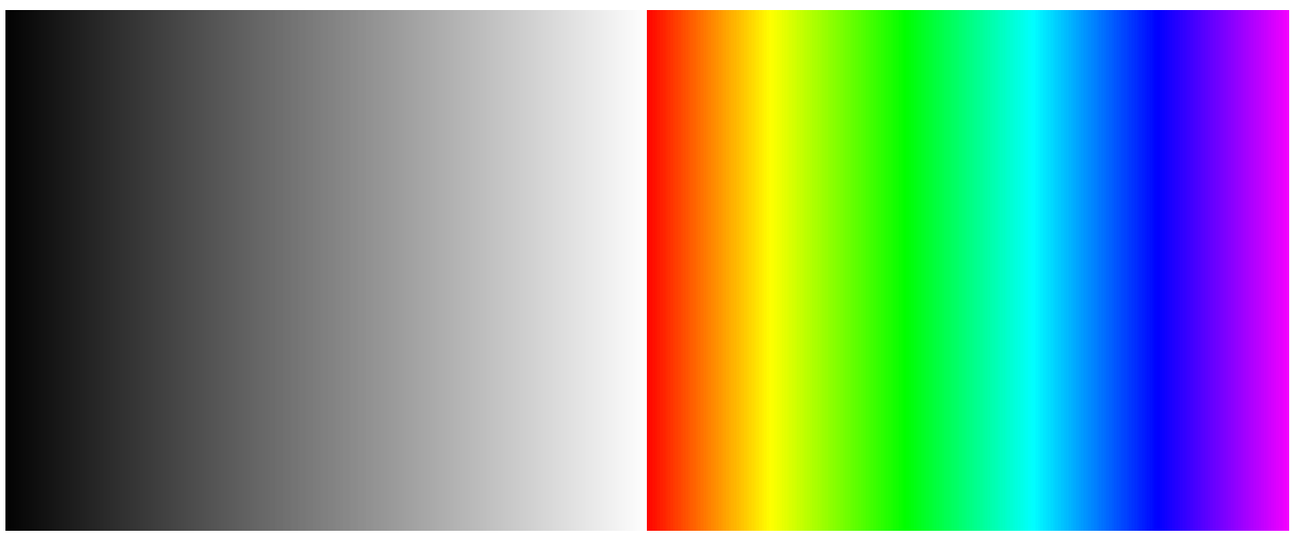


lower resolution



easy to order

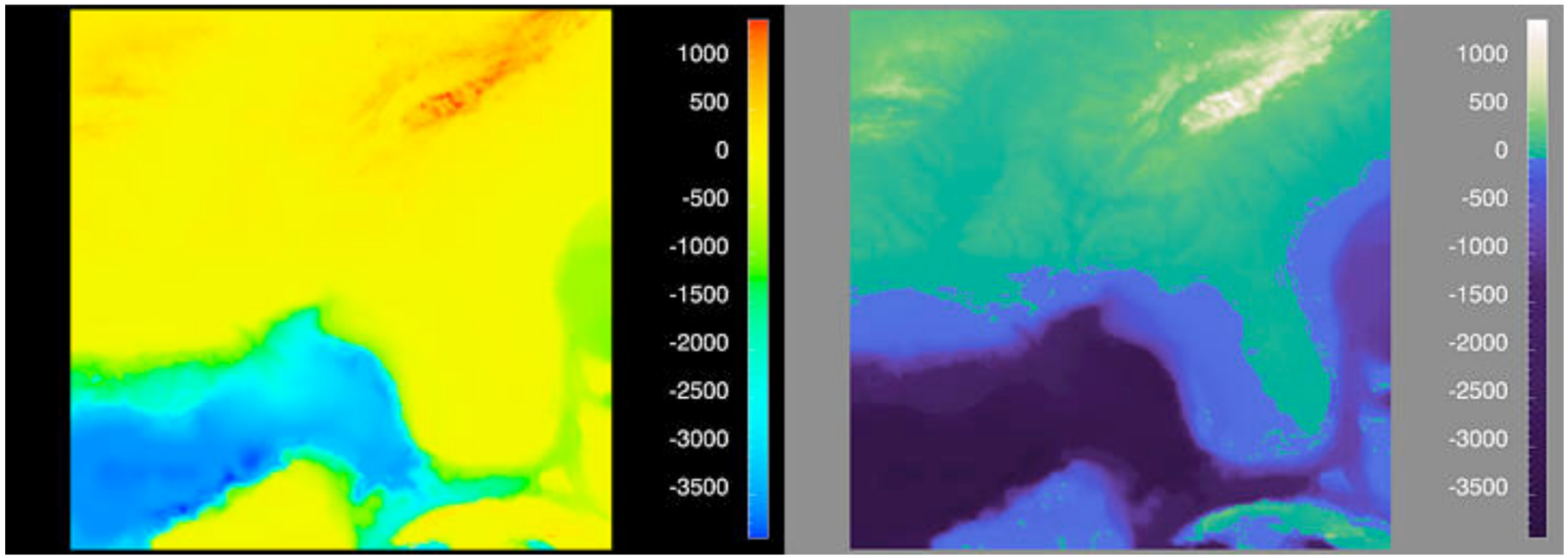
creates artifacts



after Borland & Taylor 2007, IEEE CG&A

Color Pitfalls: Rainbow Color Map

Southeastern United States and Gulf of Mexico



Problems

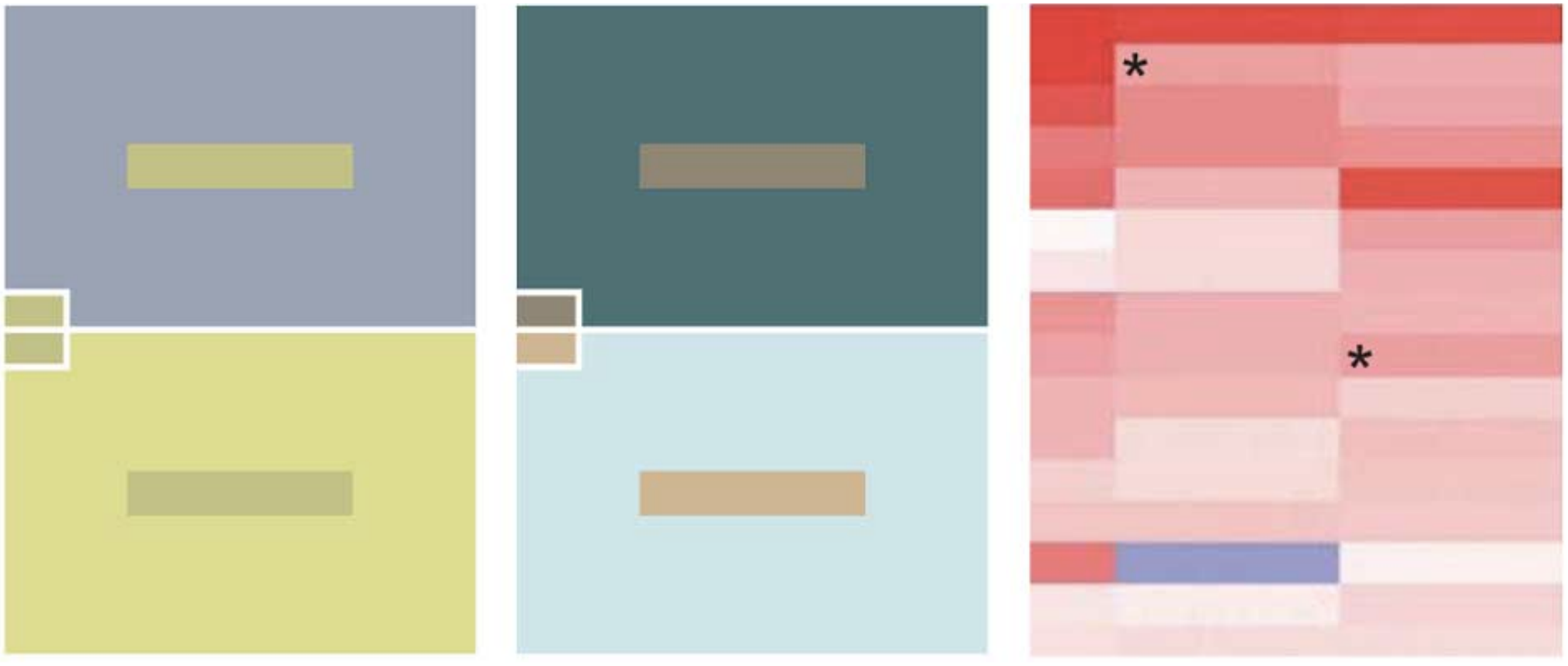
Zero crossing not explicit.

Lack of ordering of colors makes it hard to interpret the map.

Rogowitz & Treinish 1996, http://researchweb.watson.ibm.com/people/l/lloydt/color/color.HTM

Color Pitfalls: Relativity

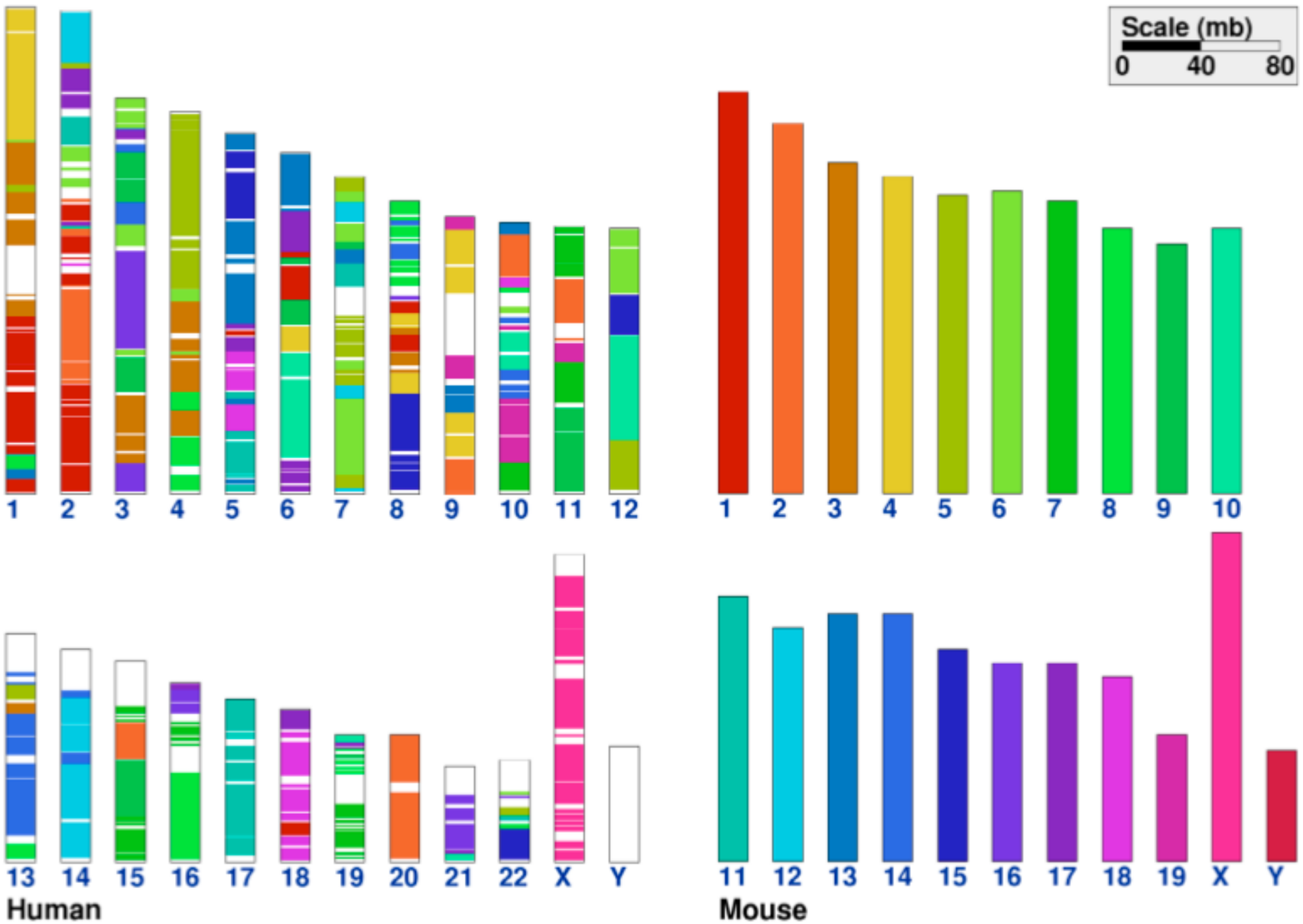
Color is a relative medium and context matters



Wong 2010, Nature Methods

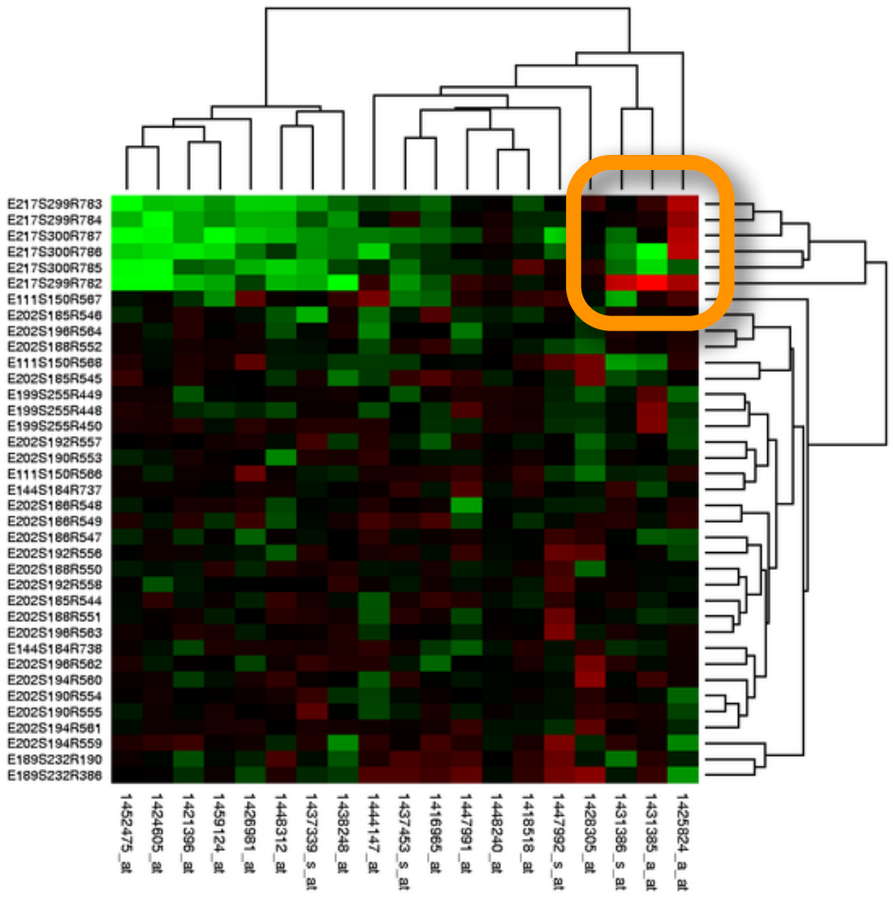
Color Pitfalls: Discriminability

Only 6-12 colors are visually discernible!

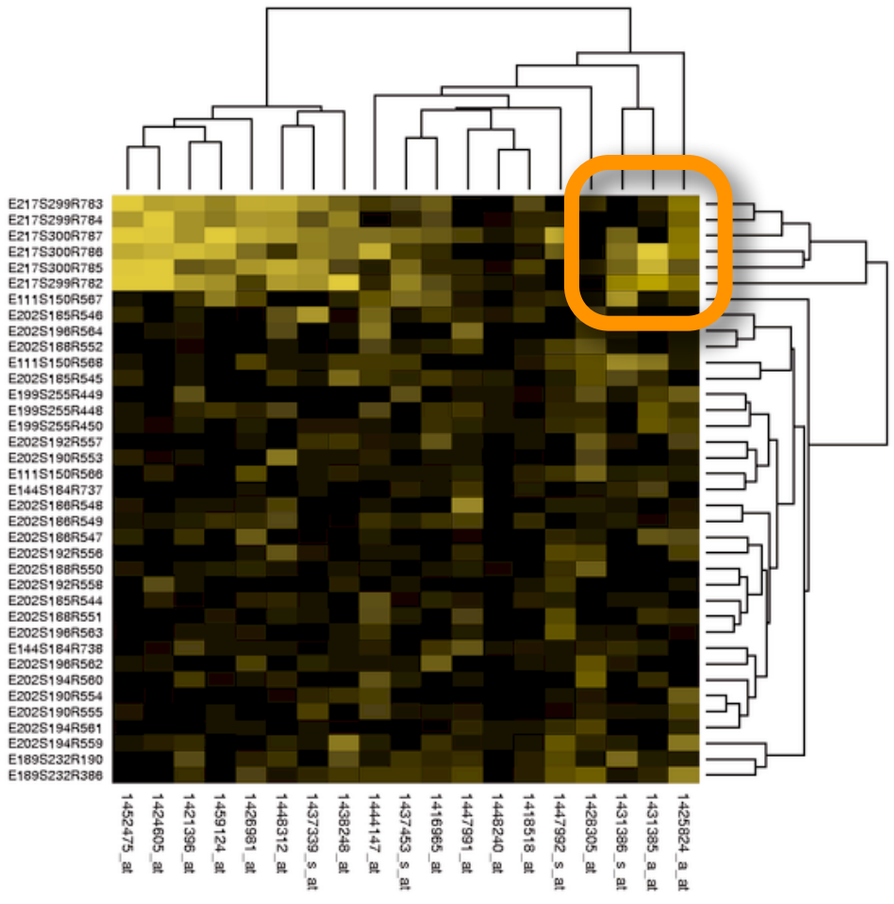


Sinha and Meller, Bioinformatics 2007

Color Pitfalls: Color Blindness



Normal Vision



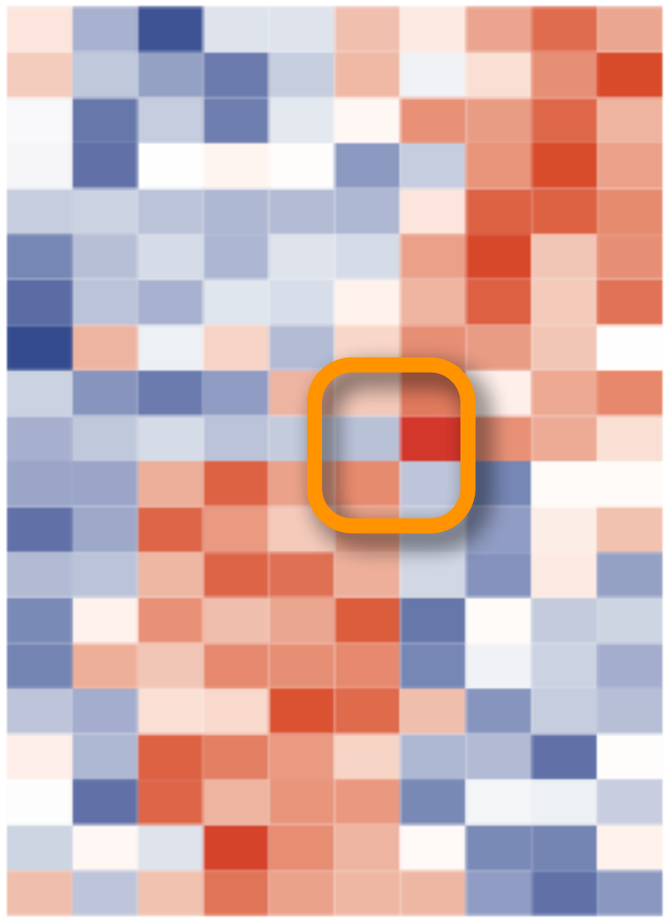
Deuteranope Vision

(“Red-Green Blindness”)

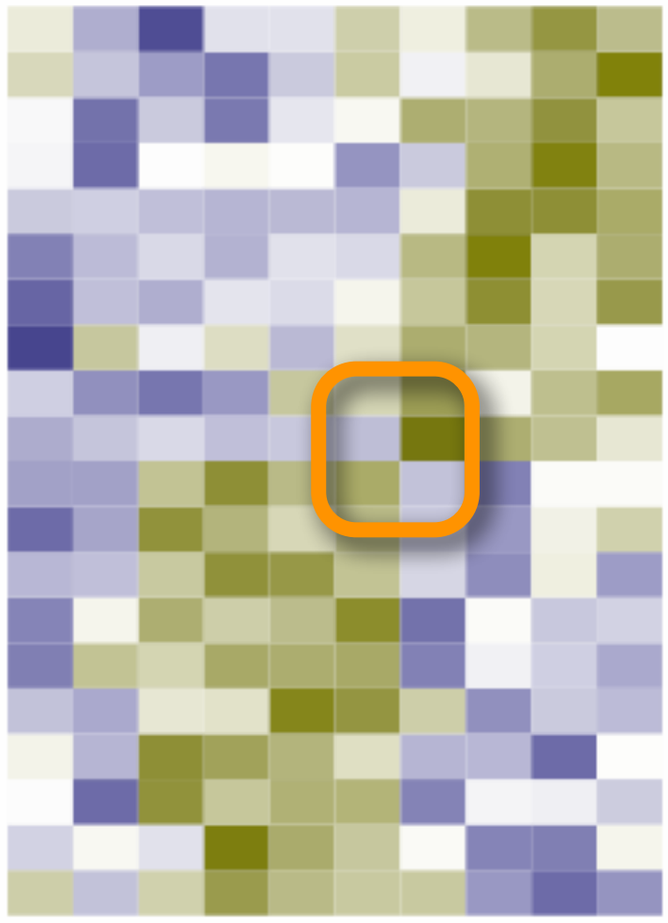
~ 7% of male population affected

estimate: Howard Hughes Medical Institute, http://www.hhmi.org/senses/b130.html

Color Pitfalls: Color Blindness



Normal Vision



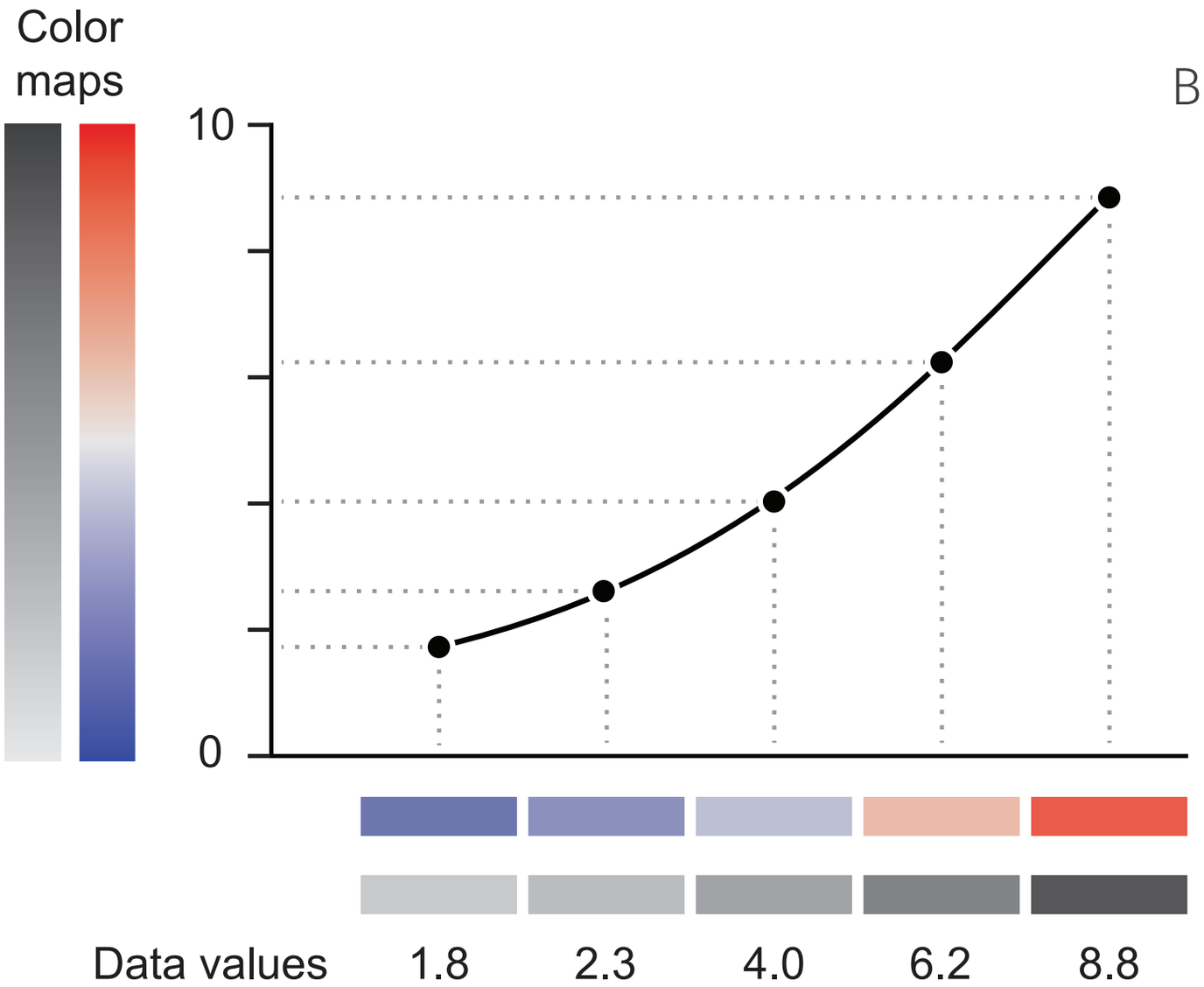
Deuteranope Vision

(“Red-Green Blindness”)

~ 7% of male population affected

estimate: Howard Hughes Medical Institute, http://www.hhmi.org/senses/b130.html

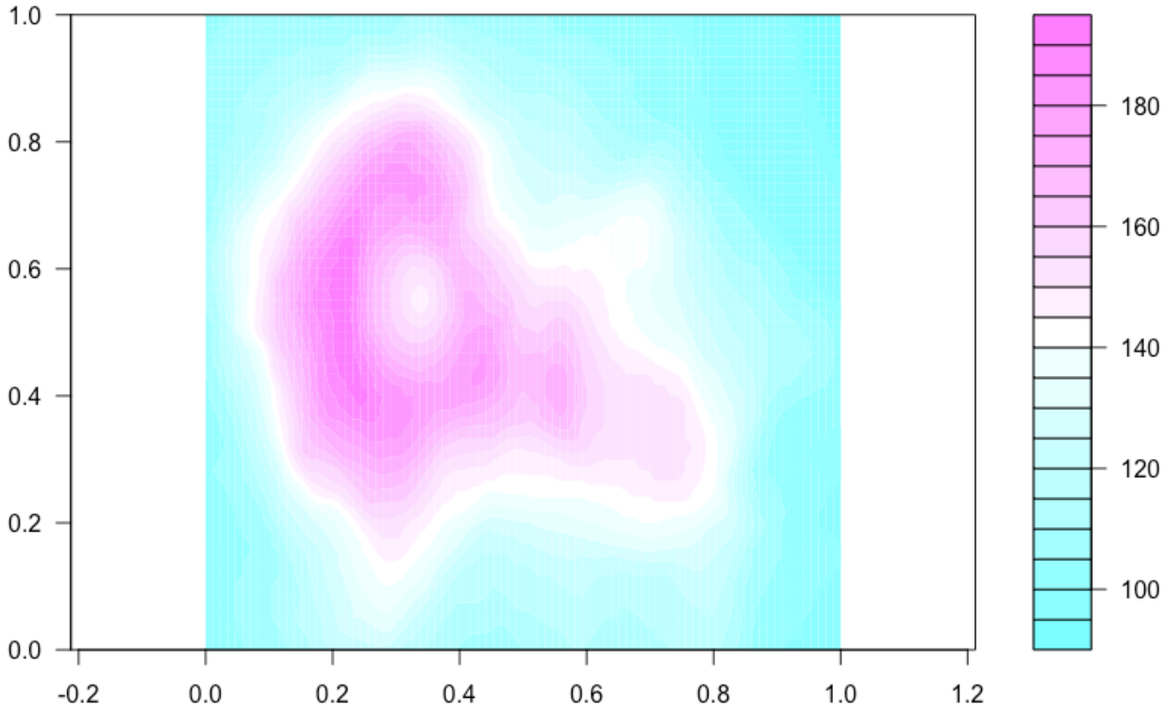
Color Pitfalls: Color Mapping

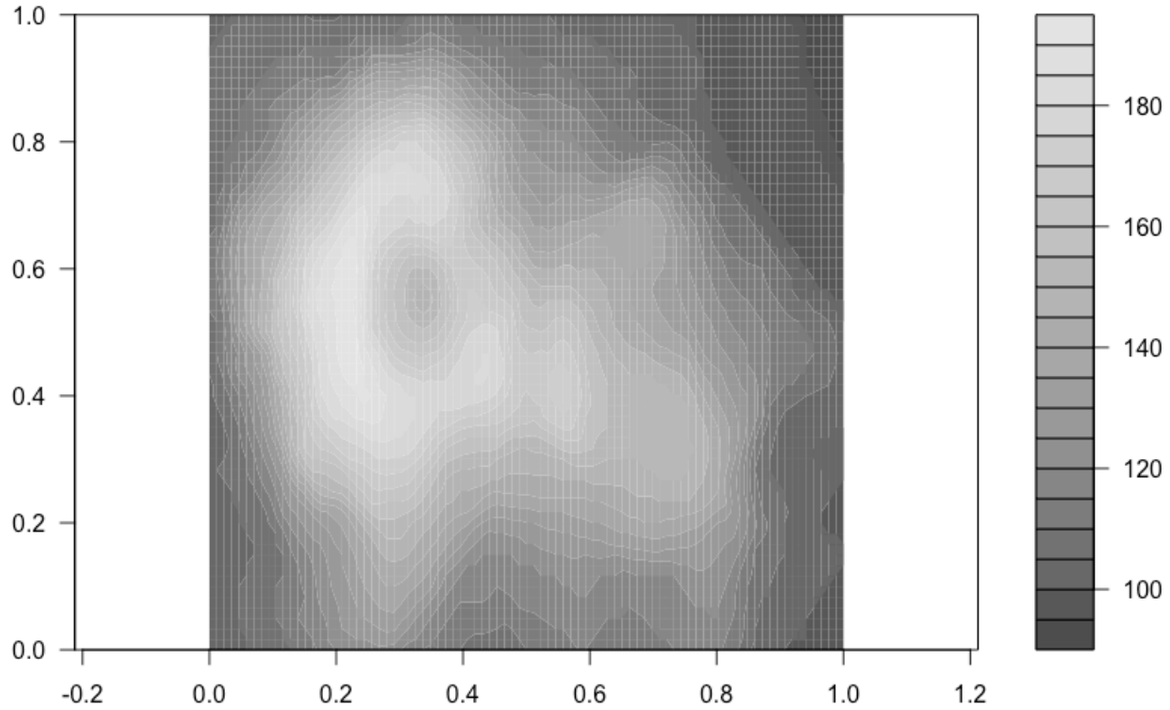


Bad color mapping!

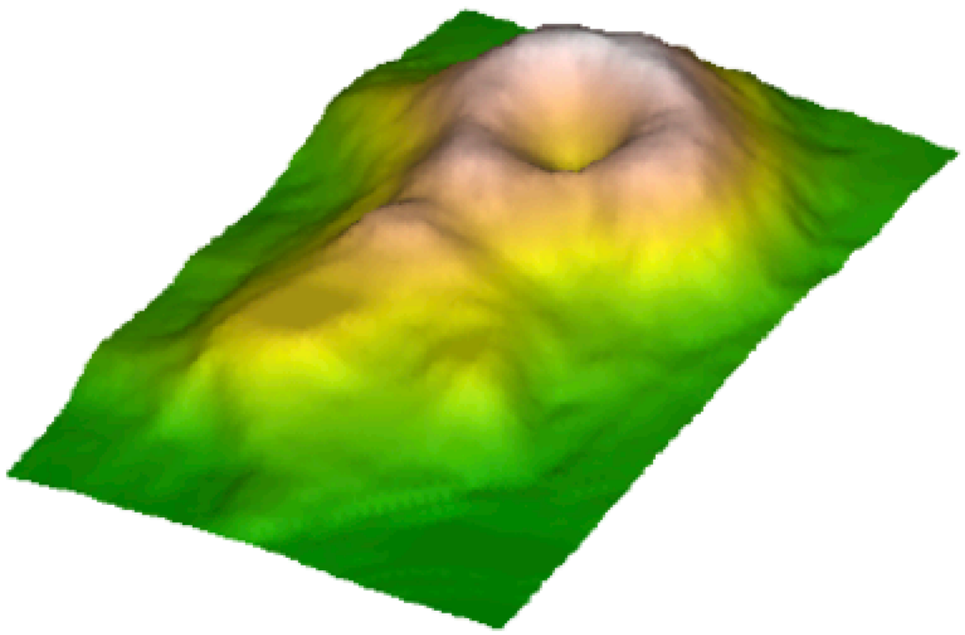
Gehlenborg and Wong, Nature Methods, 2012

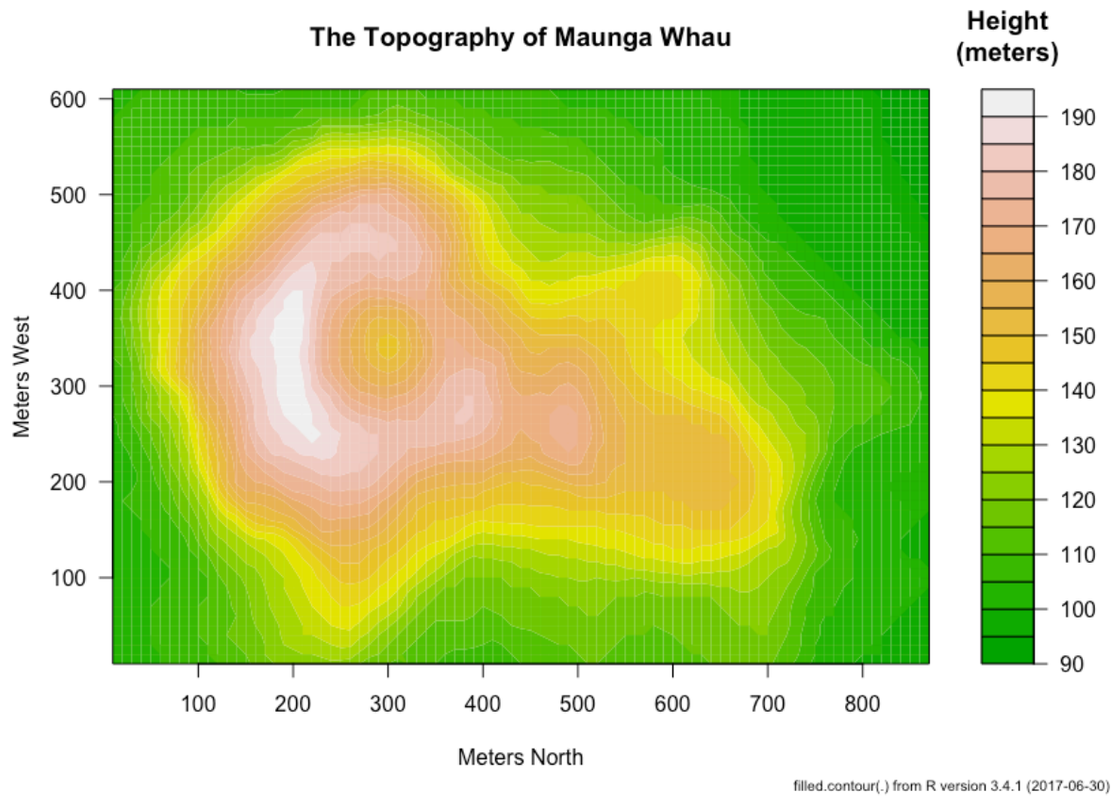
Color Pitfalls: Color Mapping





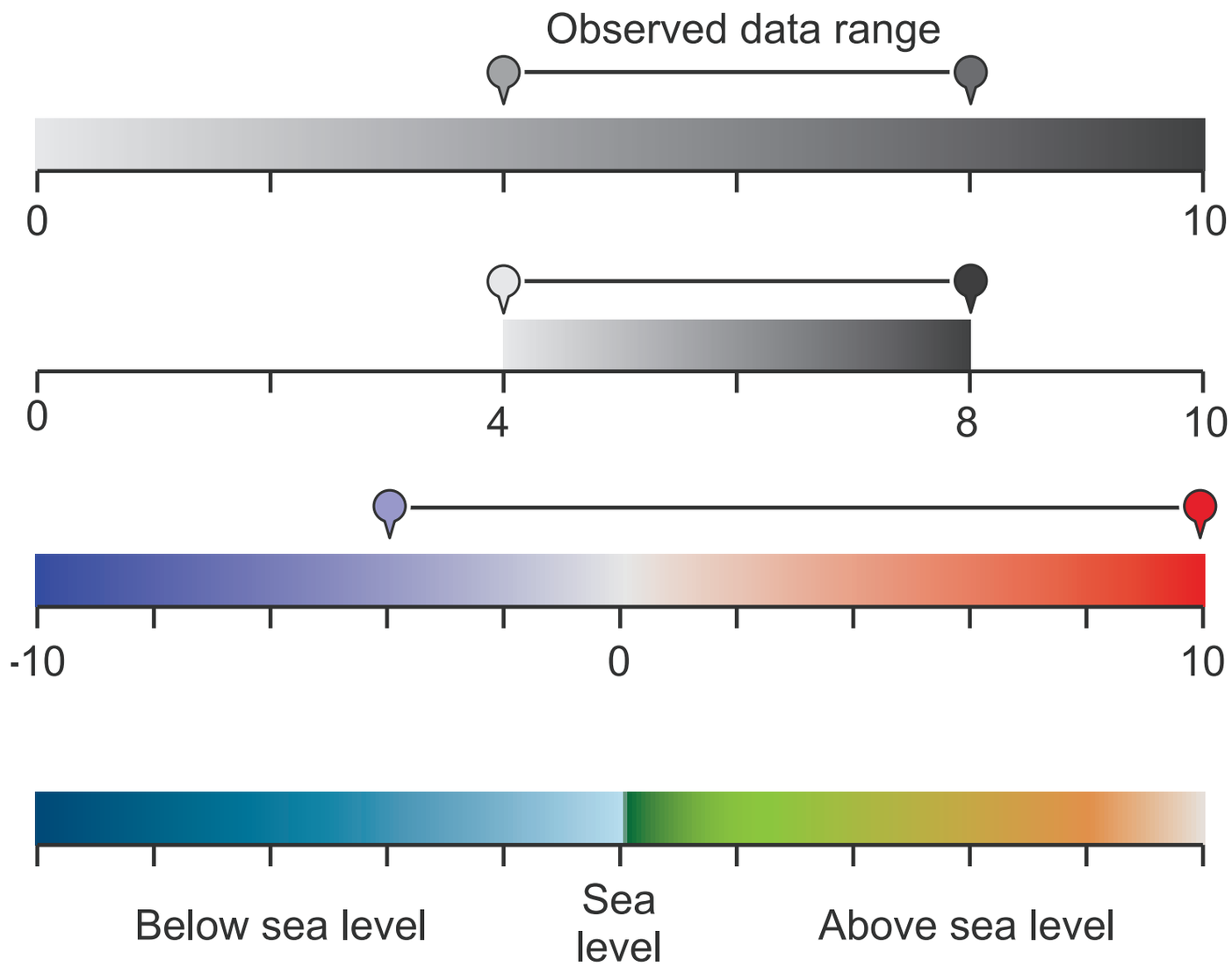
Color Pitfalls: Color Mapping





Color Pitfalls: Color Mapping

Good color mapping!



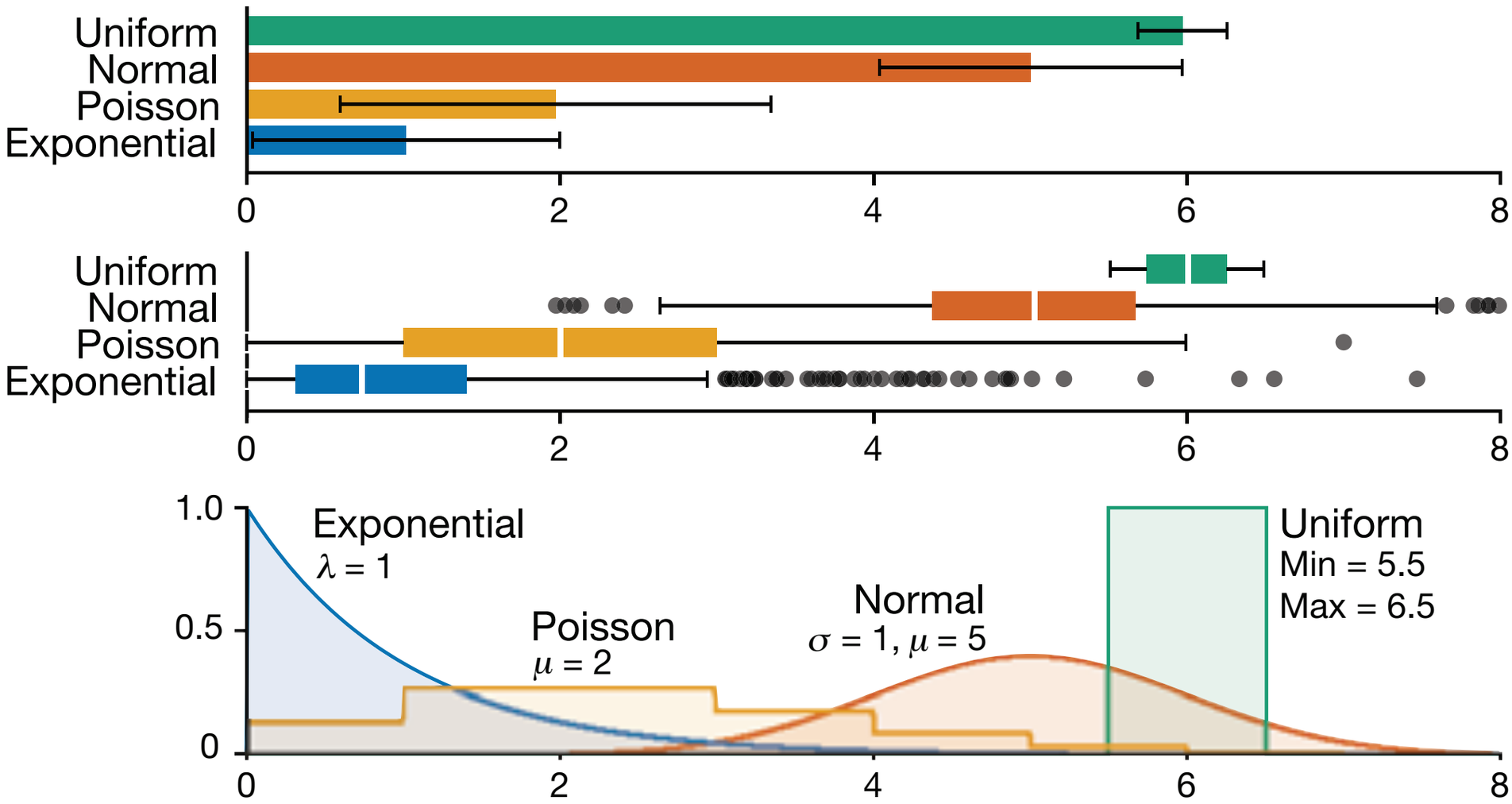
Gehlenborg and Wong, Nature Methods, 2012

Remember!

Color used poorly is worse than no color at all.

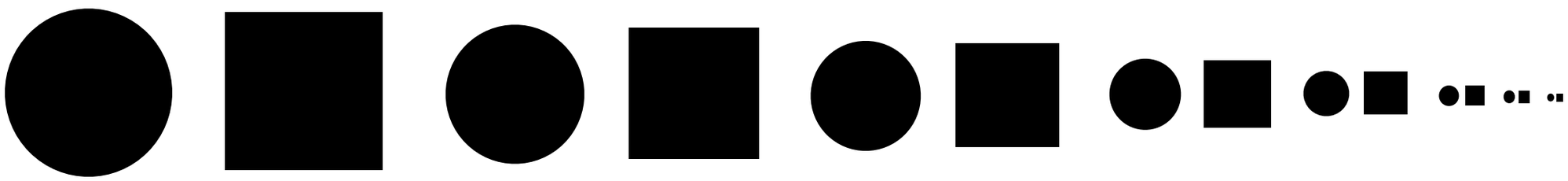
— *Edward Tufte*

Location Pitfalls: Data Location



Streit and Gehlenborg, Nature Methods, 2014

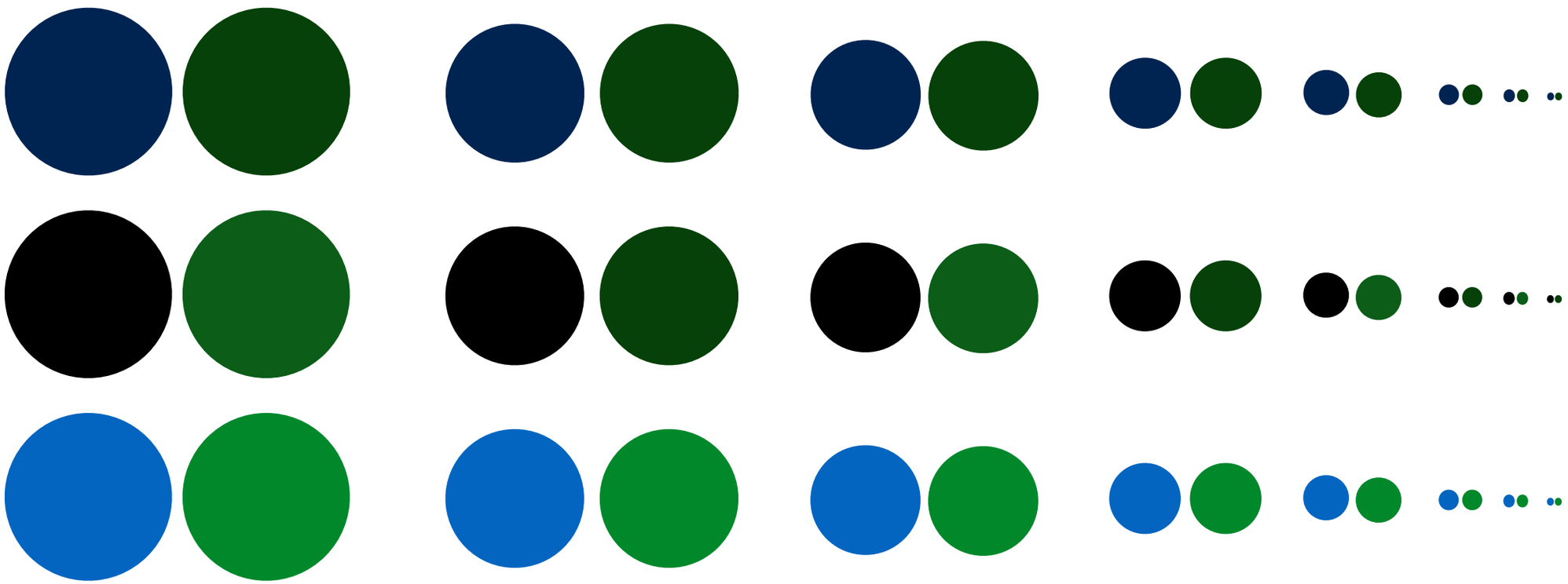
Encoding Pitfalls: Interference



shape and size

after Mackinlay, ACM Transactions on Graphics 5, 110-141, 1986

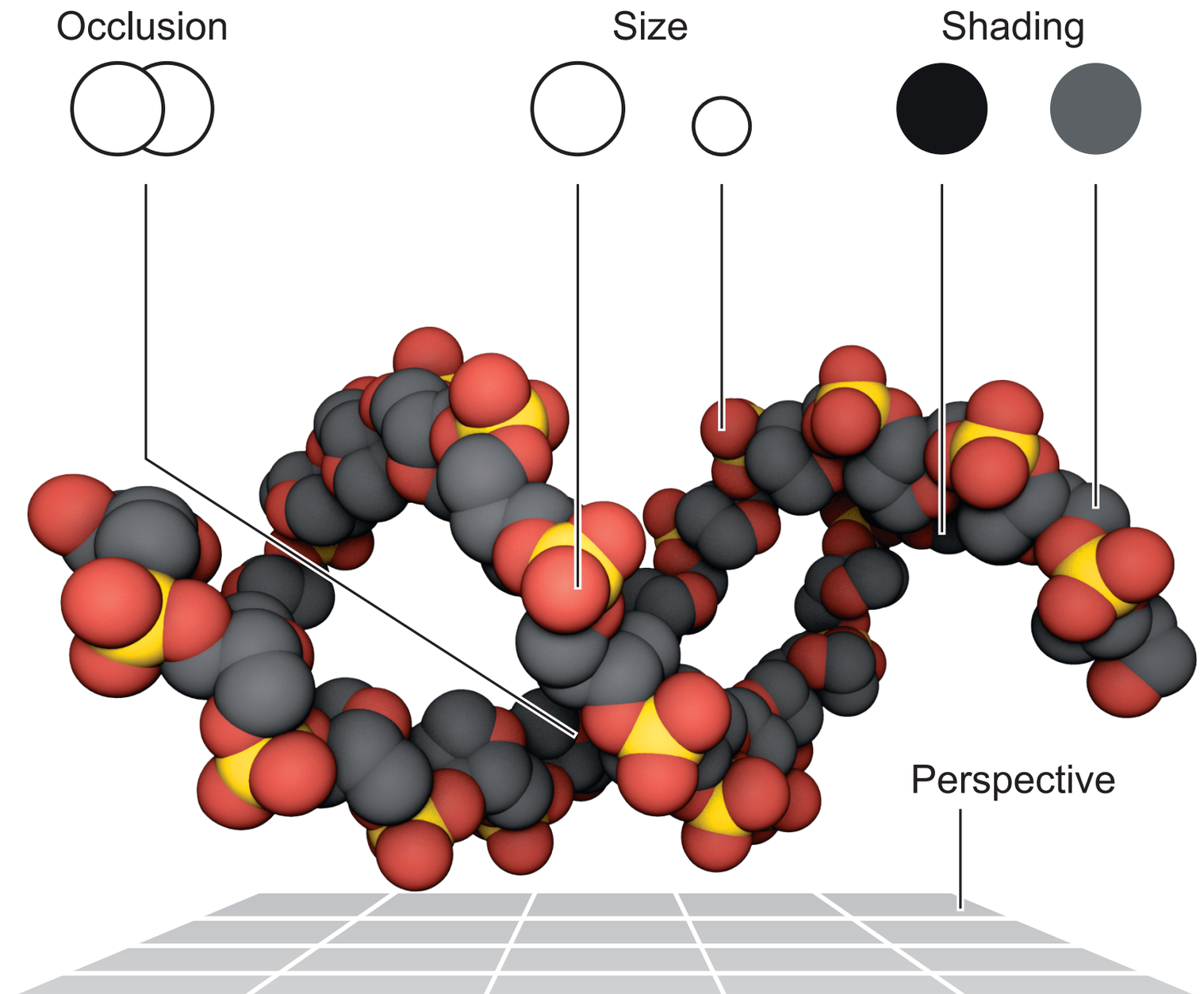
Encoding Pitfalls: Interference



color and size

after Mackinlay, ACM Transactions on Graphics 5, 110-141, 1986

3D: Depth Cues

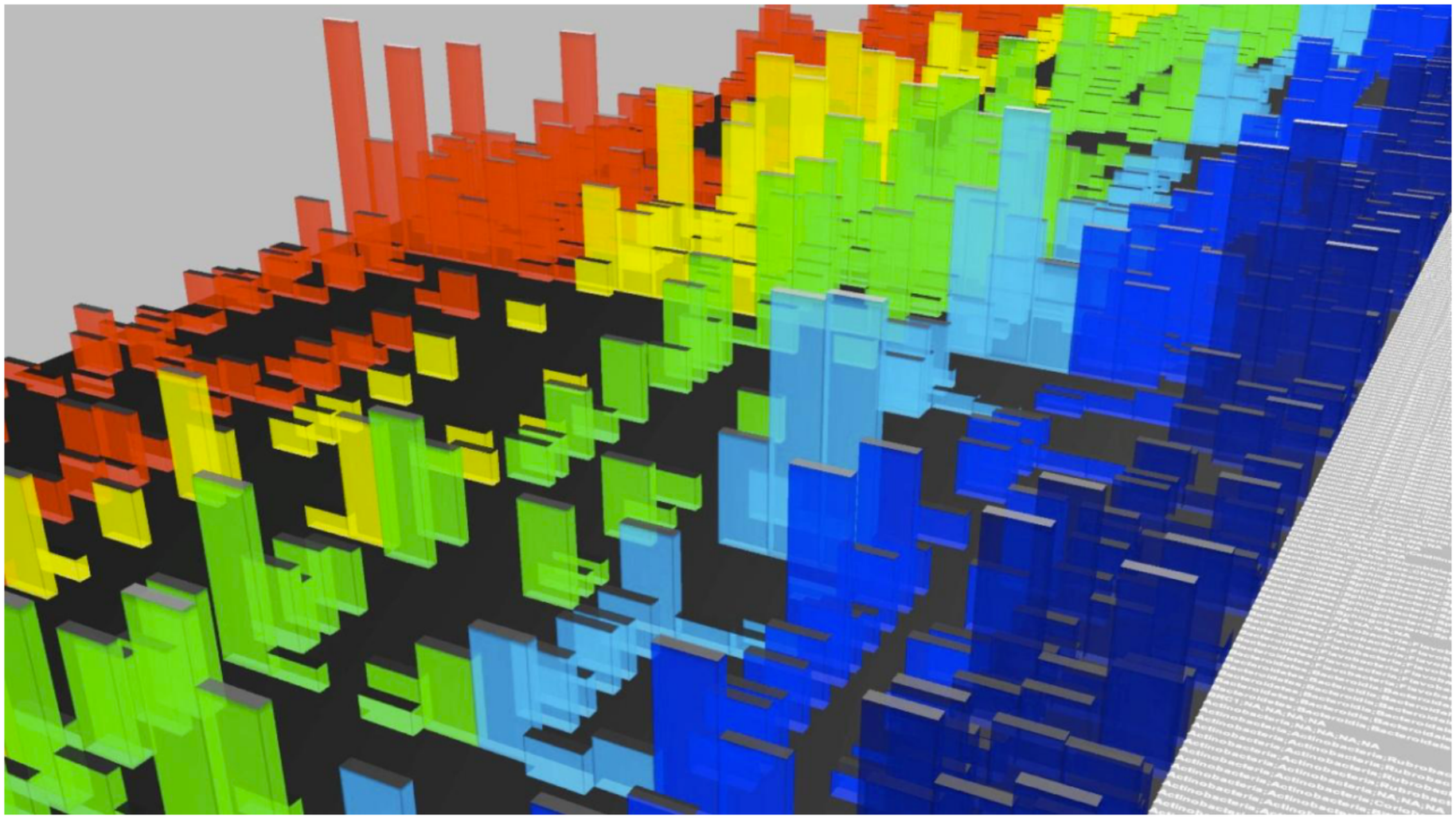


Gehlenborg and Wong, Nature Methods, 2012

3D Pitfalls: Perspective

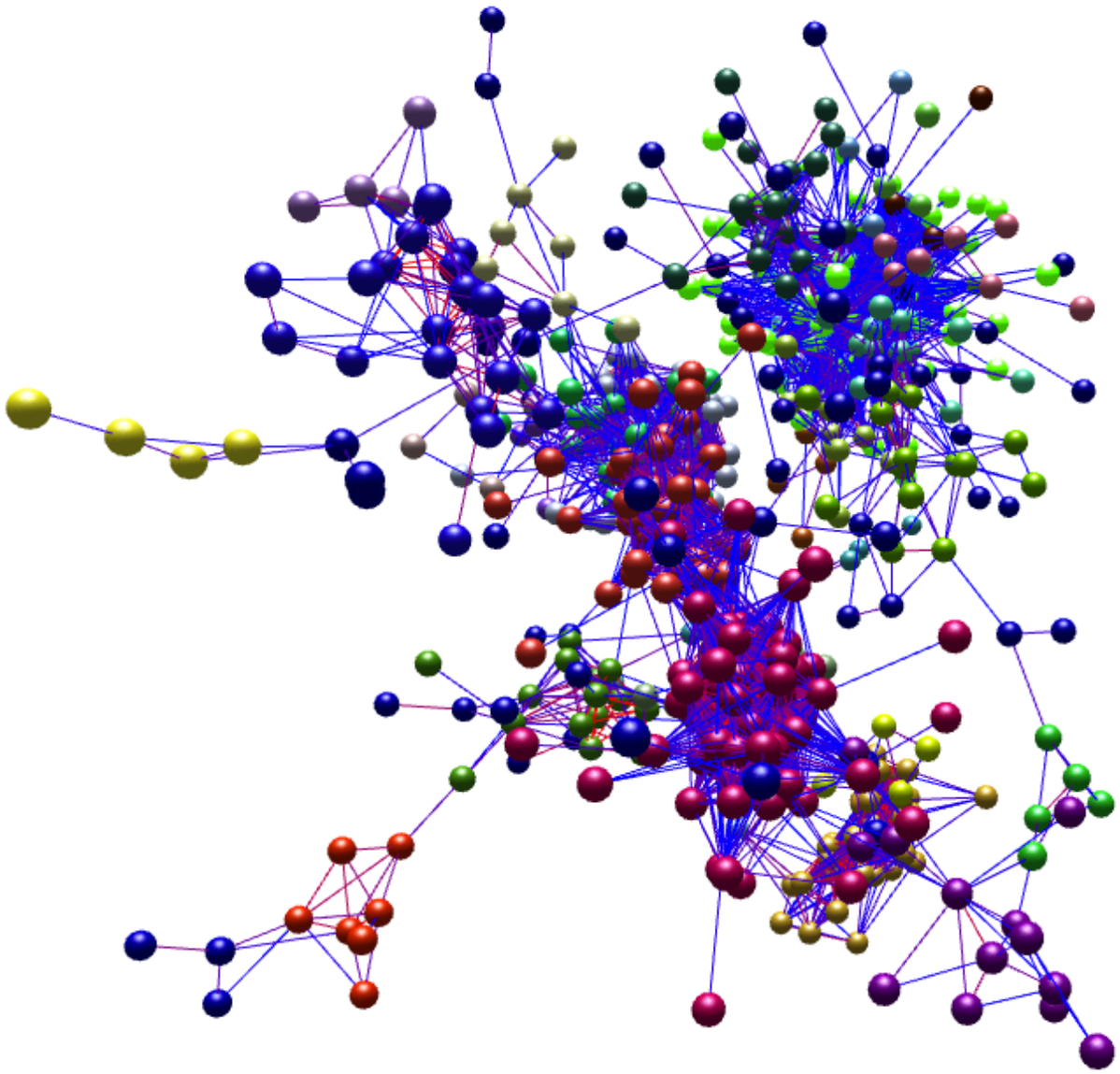
Perspective distortion: interferes with size channel encoding

Shading: interferes with color, lightness, and saturation channel encodings



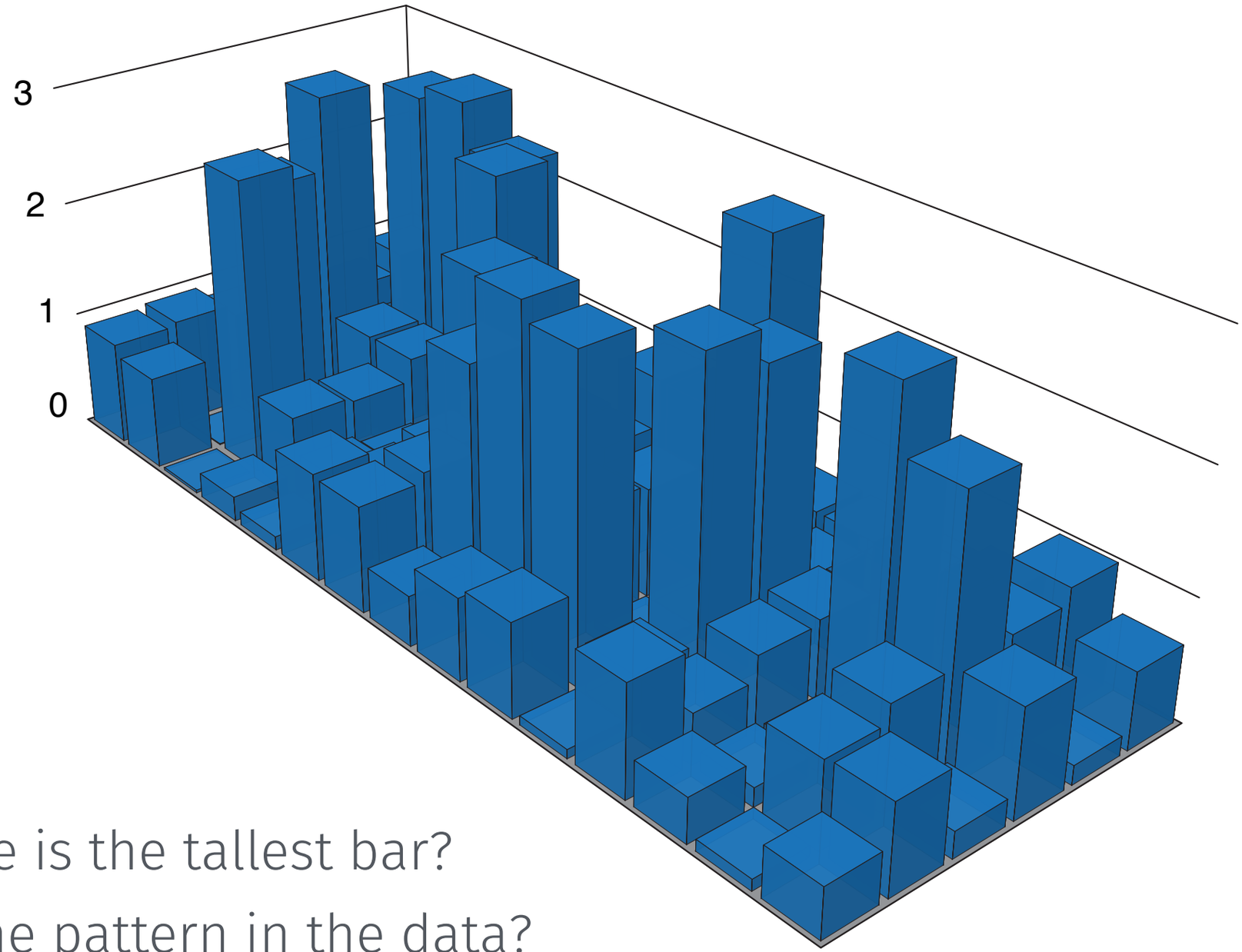
Moore et al. 2011, Proceedings of Pacific Symposium on Biocomputing 2011

3D Pitfalls: Occlusion



BioLayout 3D 2.0 sample dataset, http://www.biolayout.org

3D Pitfalls: Occlusion & Perspective

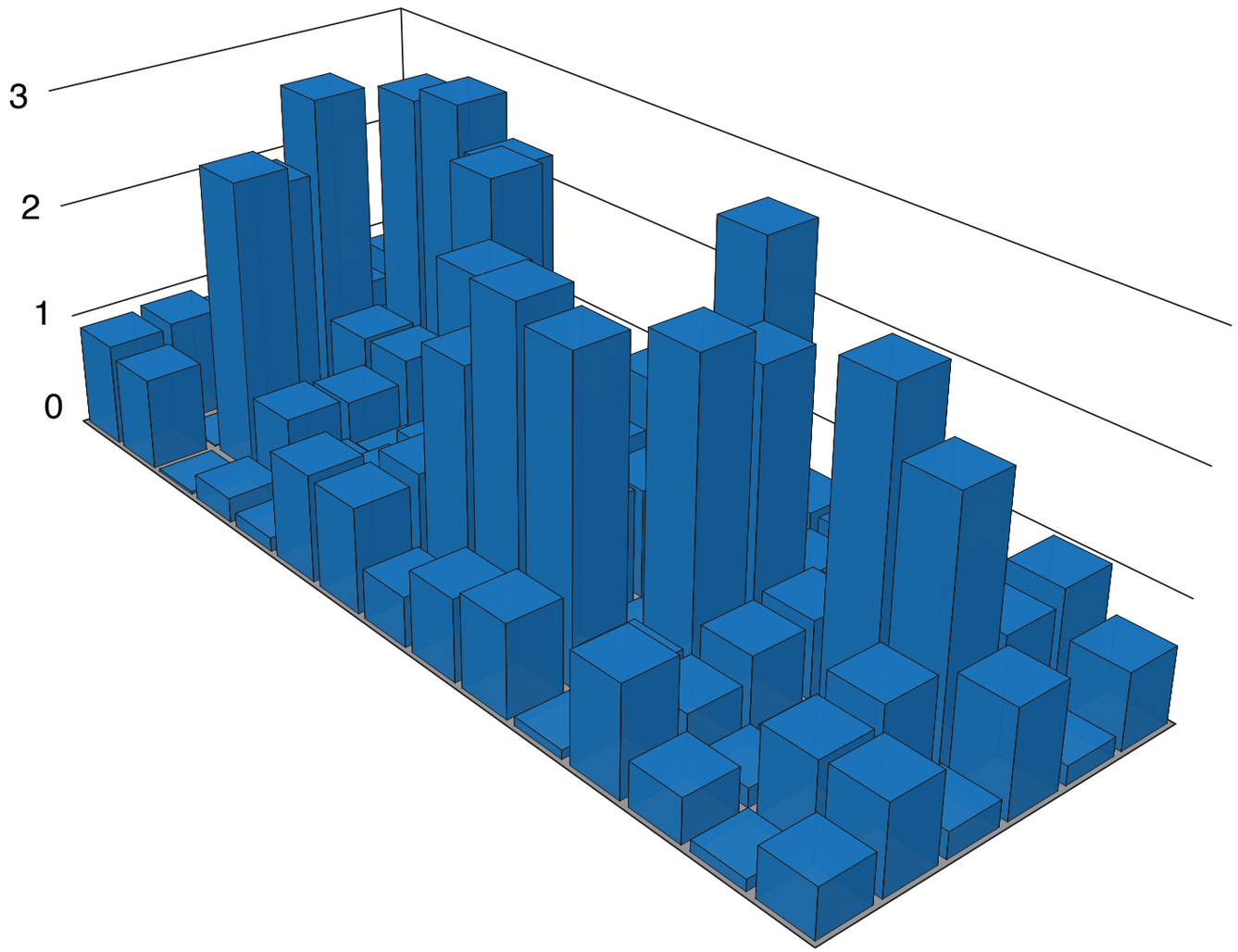


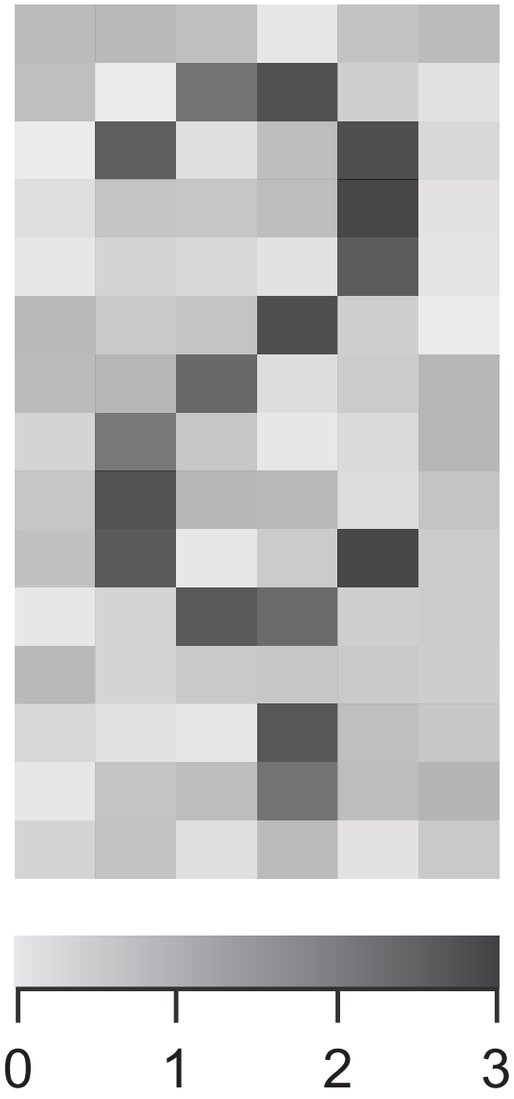
Which one is the tallest bar?

What is the pattern in the data?

Gehlenborg and Wong, Nature Methods, 2012

3D Pitfalls: Occlusion & Perspective



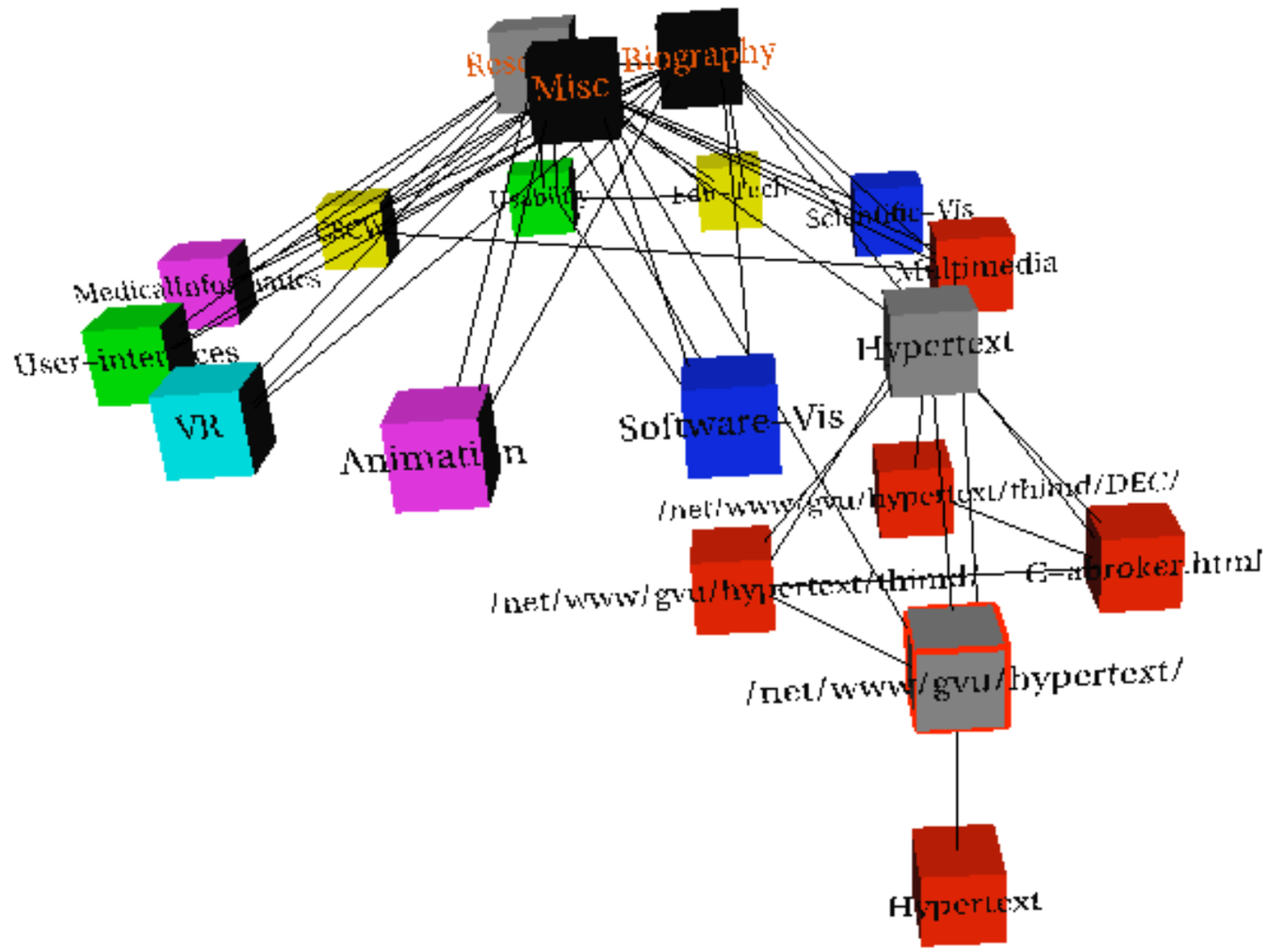


Which one is the tallest bar?

What is the pattern in the data?

Gehlenborg and Wong, Nature Methods, 2012

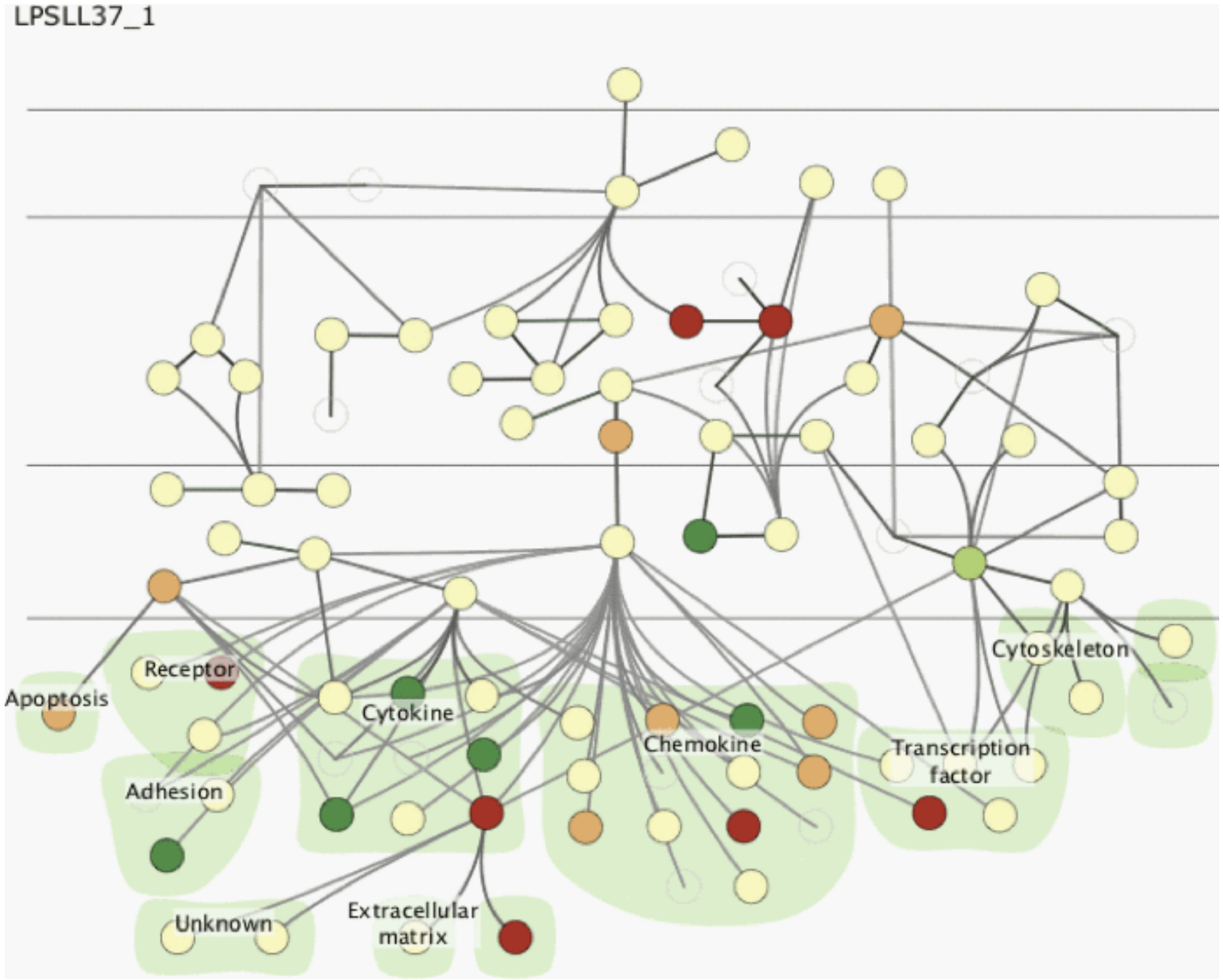
3D Pitfalls: Text Legibility



Mukherjea and Foley, 1995

Animation Pitfall

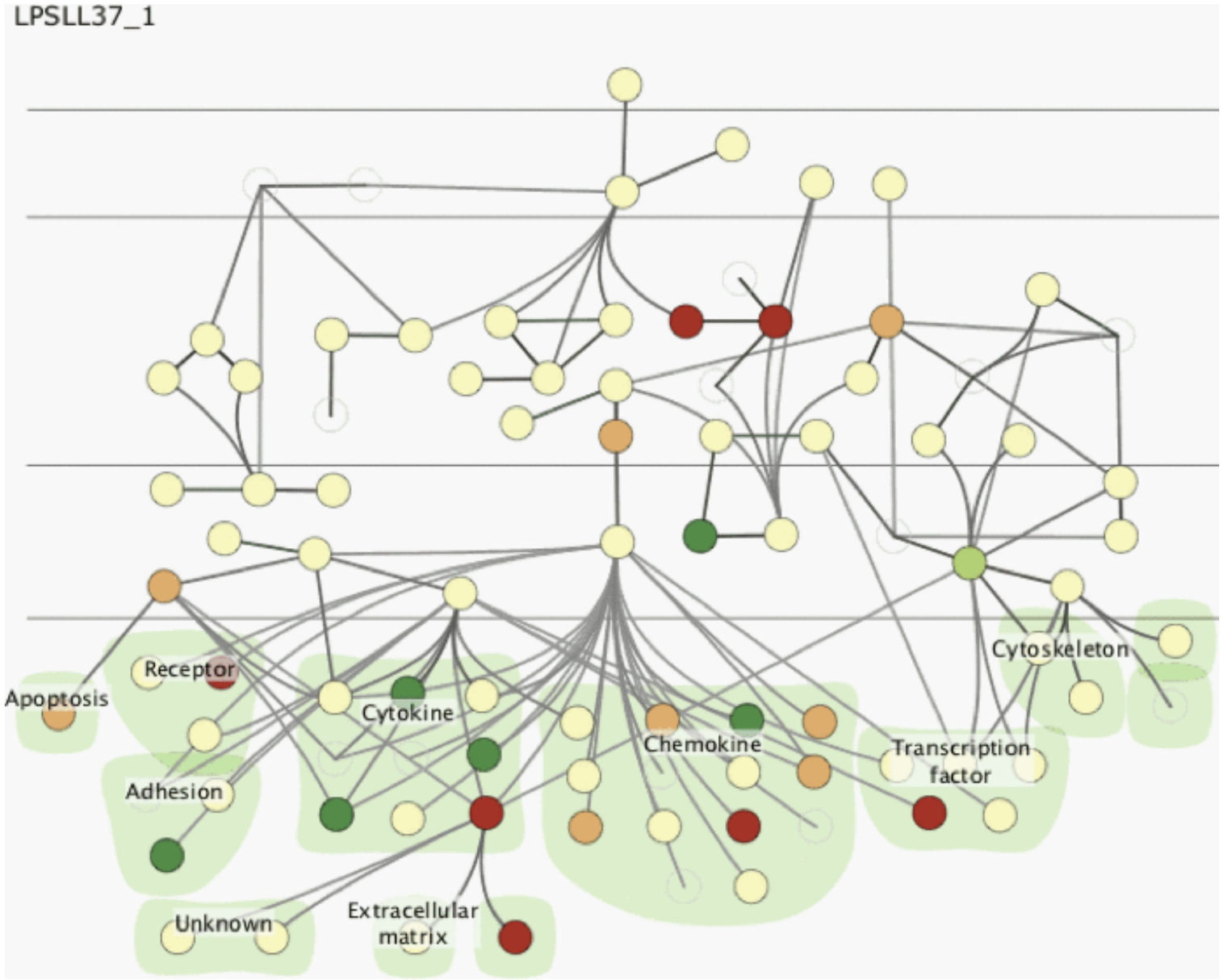
Global comparisons are difficult



Barsky et al. 2008, Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context

Animation Pitfall

Global comparisons are difficult



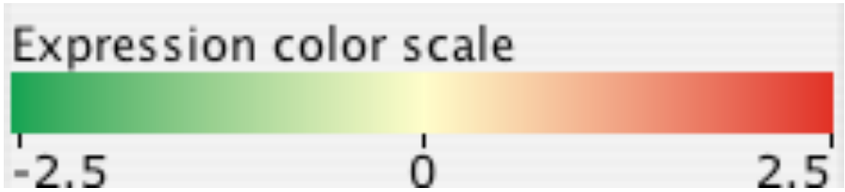
Barsky et al. 2008, Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context

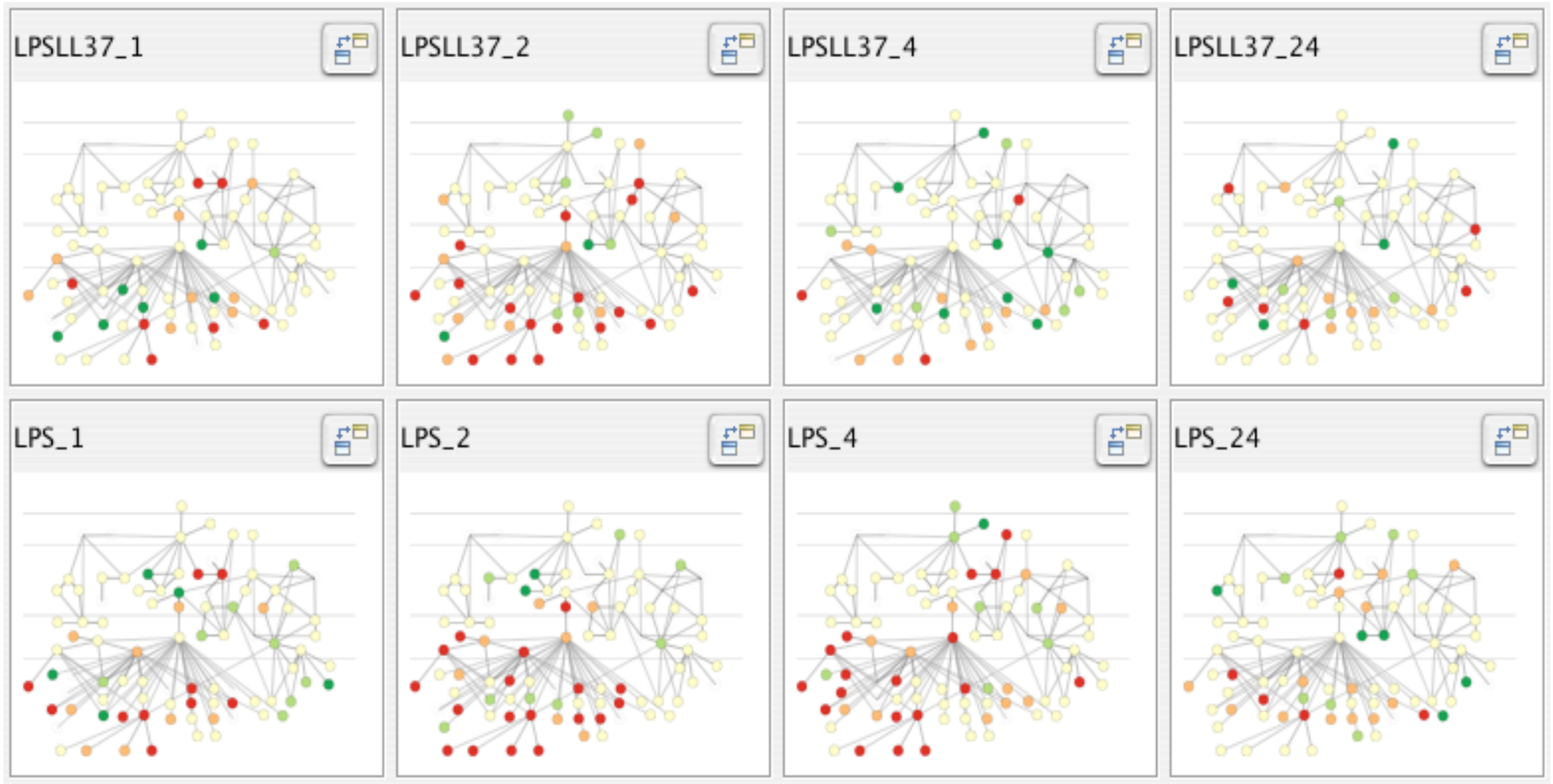
Animation Pitfall

Small Multiples

one view per state

show time with space



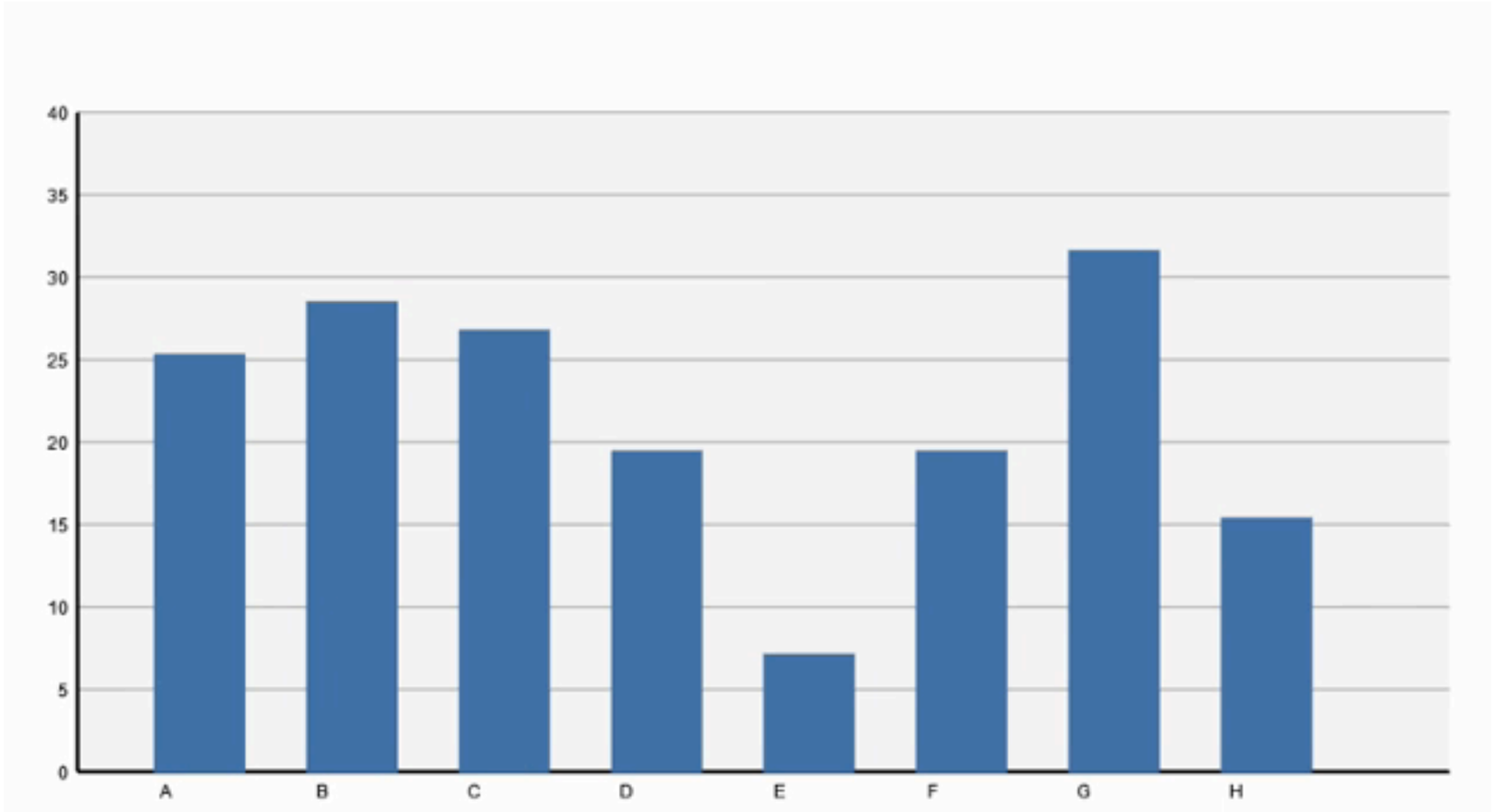


Barsky et al. 2008, Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context

Animation

- external versus internal memory
  - easy to compare by moving eyes between views
  - hard to compare view to memory of what you saw
- when to use animation?
  - **good:** chronological storytelling
  - **good:** transition between states
  - **poor:** multiple states with multiple changes

Animation



Heer and Robertson, Animated Transitions in Statistical Data Graphics, 2007

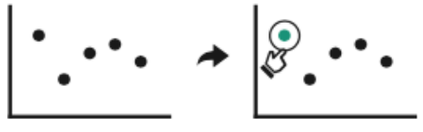
Interactions

Single View Interactions

Manipulate

Change over Time

Select

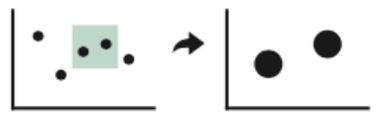


Navigate

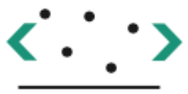
→ Item Reduction

→ *Zoom*

Geometric or *Semantic*



→ *Pan/Translate*

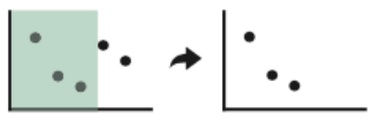


→ *Constrained*

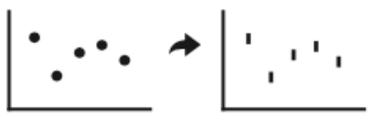
→ Attribute Reduction

→ *Slice*

→ *Cut*



→ *Project*



Munzner, Visualization Analysis and Design, 2014 (Chapter 11)

Exploration: Hypothesis Generation

- Visualization for exploration is an “Exploratory Data Analysis” technique (Tukey 1977). Statistical graphics such as box plots and scatter plots are early examples.

- When there is a specific question that can easily be determined algorithmically (“What is the highest value?”), then visualization is usually not the right tool.

- When it is not clear what should be asked or when the answer can not be summarized easily (“What is the distribution of the values?”), then visualization is an excellent choice.

- **Visualization for exploration is challenging because the data sets are getting bigger and more heterogeneous.**

Exploration: Information Seeking Mantra

- In explorative settings the user is normally dealing with large amounts of data.

- Impossible to grasp everything at once.

- Solution: Make visualizations **interactive** to support the user in exploring subsets of the data at different resolutions.

- Ben Shneiderman's **Information Seeking Mantra**:
    - Overview first
    - Zoom and filter
    - Then details on demand.

Shneiderman 1996, in Proceedings IEEE Symposium on Visual Languages

Exploration: Visual Analytics Mantra

- Keim-Mansmann-Thomas **Visual Analytics Mantra**:
    - Analyze first
    - Show the important
    - Zoom, filter and analyze further
    - Then Details on demand

Keim et al., 2009, SIG KDD Explorations

Interaction Techniques: Basic Methods

- **Selection**
    - click on elements
    - lasso/drag over elements
    - hover over elements
- **Manipulation**
    - move elements
    - sort elements
    - add or delete elements (or filter elements)

Interaction Techniques: Advanced Methods

- Focus + Context (*single* view)

- Overview + Detail (*multiple* views)

- Brushing + Linking (*multiple* views)

- Zooming + Panning

    - geometric zooming (modifies graphical representation)

    - semantic zooming (modifies selection of data)

    - often combined (Google Maps)

Pan and Zoom in D3

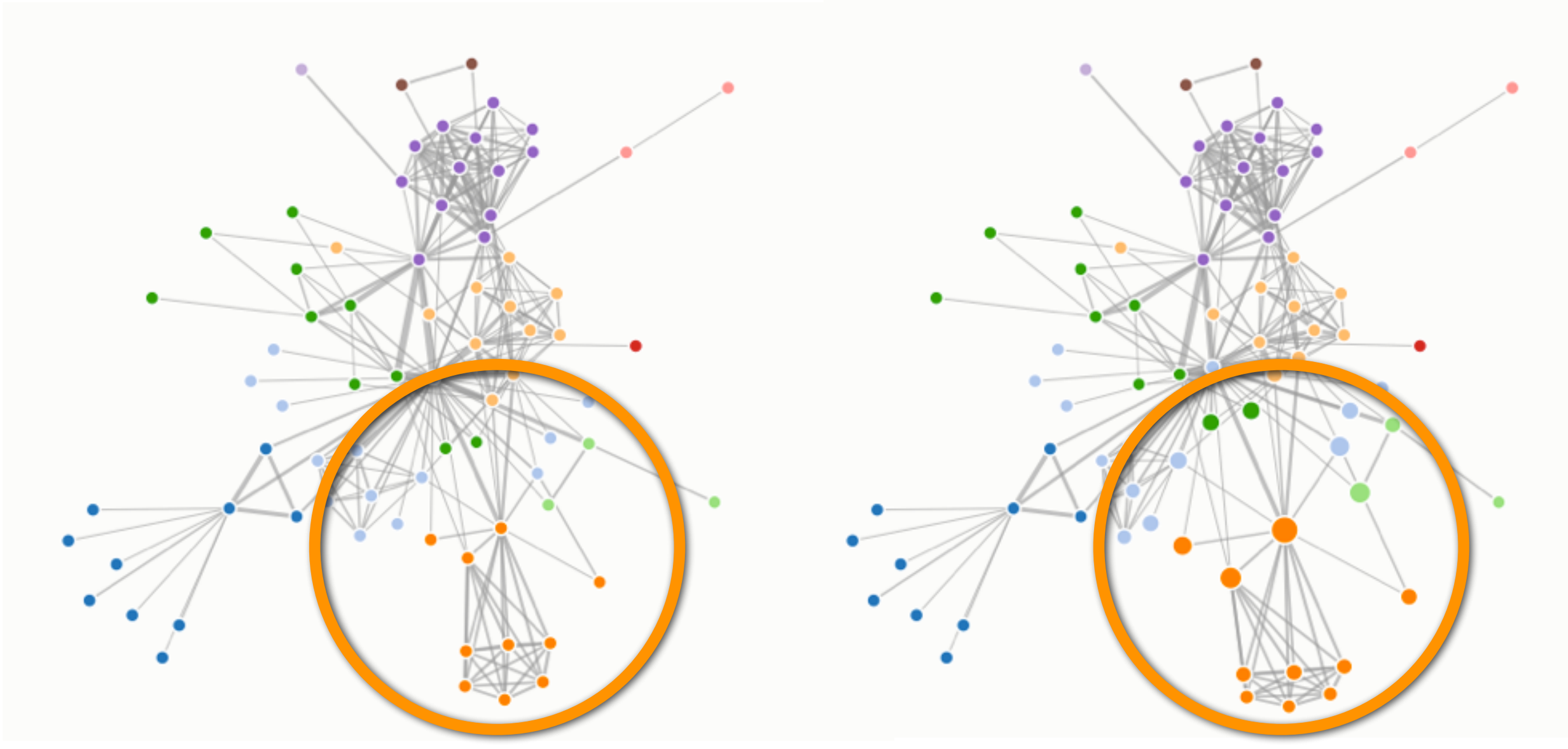
https://bl.ocks.org/mbostock/db6b4335bf1662b413e7968910104f0f

Semantic Zoom

https://bl.ocks.org/mbostock/3680957

Interaction Techniques: Focus + Context

Fisheye Distortion



Fisheye and Rectilinear Cartesian Distortion in D3

https://bost.ocks.org/mike/fisheye/

Mike Bostock, http://www.d3.js

Multi-View Interactions

Facet

Juxtapose and Coordinate Multiple Side-by-Side Views

→ Share Encoding: Same/Different

→ *Linked Highlighting*

→ Share Data: All/Subset/None

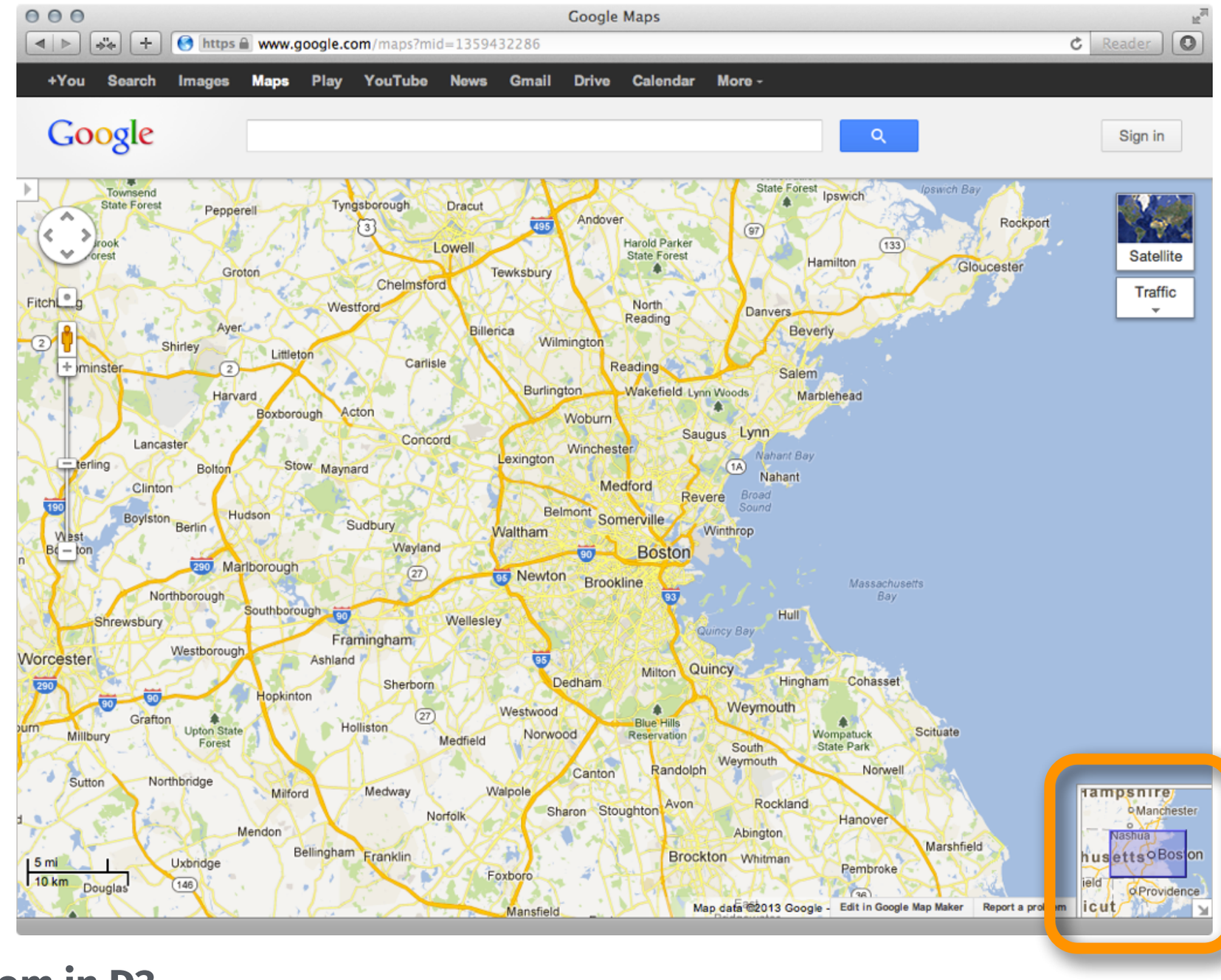
→ Share Navigation

		Data		
		All	Subset	None
Encoding	Same	Redundant	Overview/ Detail	Small Multiples
	Different	Multiform	Multiform, Overview/ Detail	No Linkage

Partition into Side-by-Side Views

Superimpose Layers

Interaction Techniques: Overview + Detail



Then …

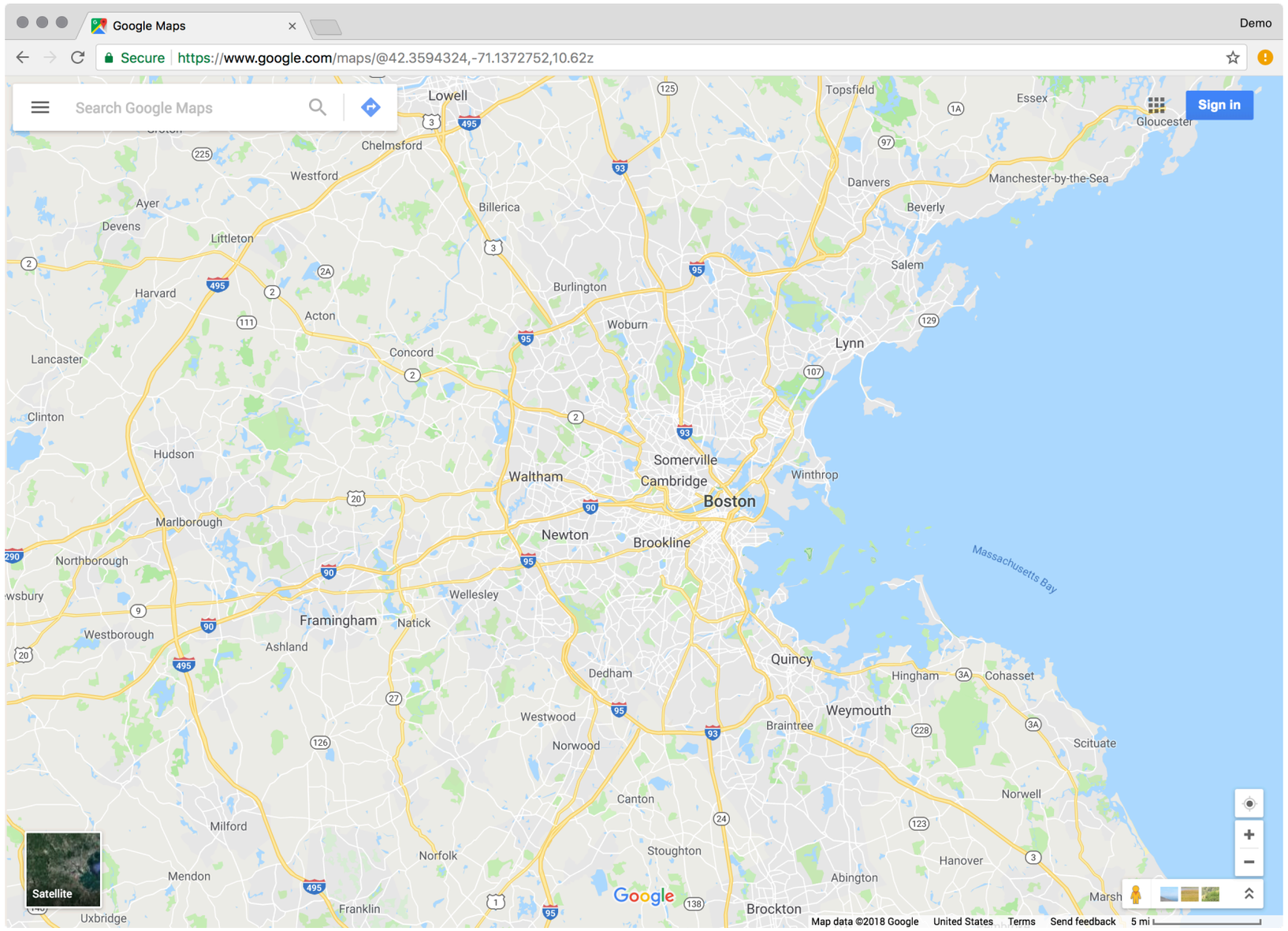
Brush and zoom in D3

https://bl.ocks.org/mbostock/f48fcdb929a620ed97877e4678ab15e6

https://bl.ocks.org/mbostock/34f08d5e11952a80609169b7917d4172 (also overview + detail)

https://maps.google.com

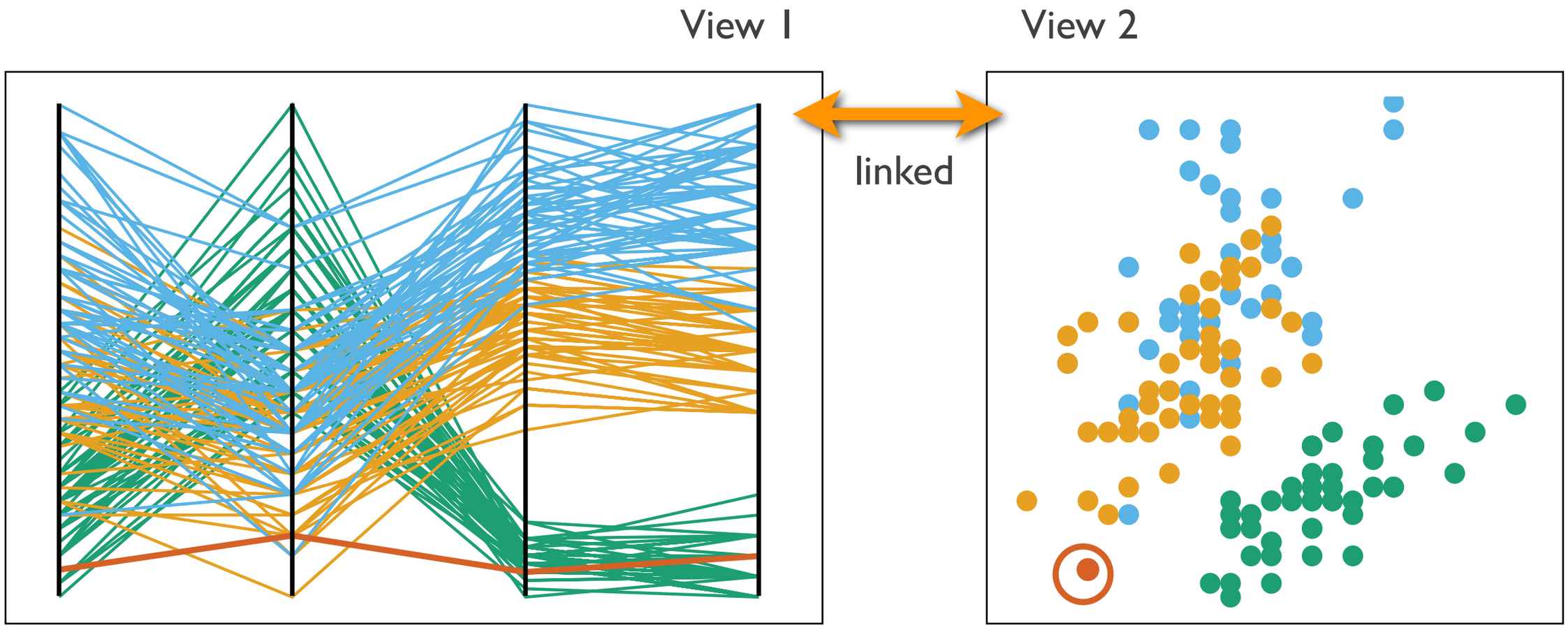
Interaction Techniques: Overview + Detail



… & today!

https://maps.google.com

Interaction Techniques: Brushing + Linking



Brushing and Linking

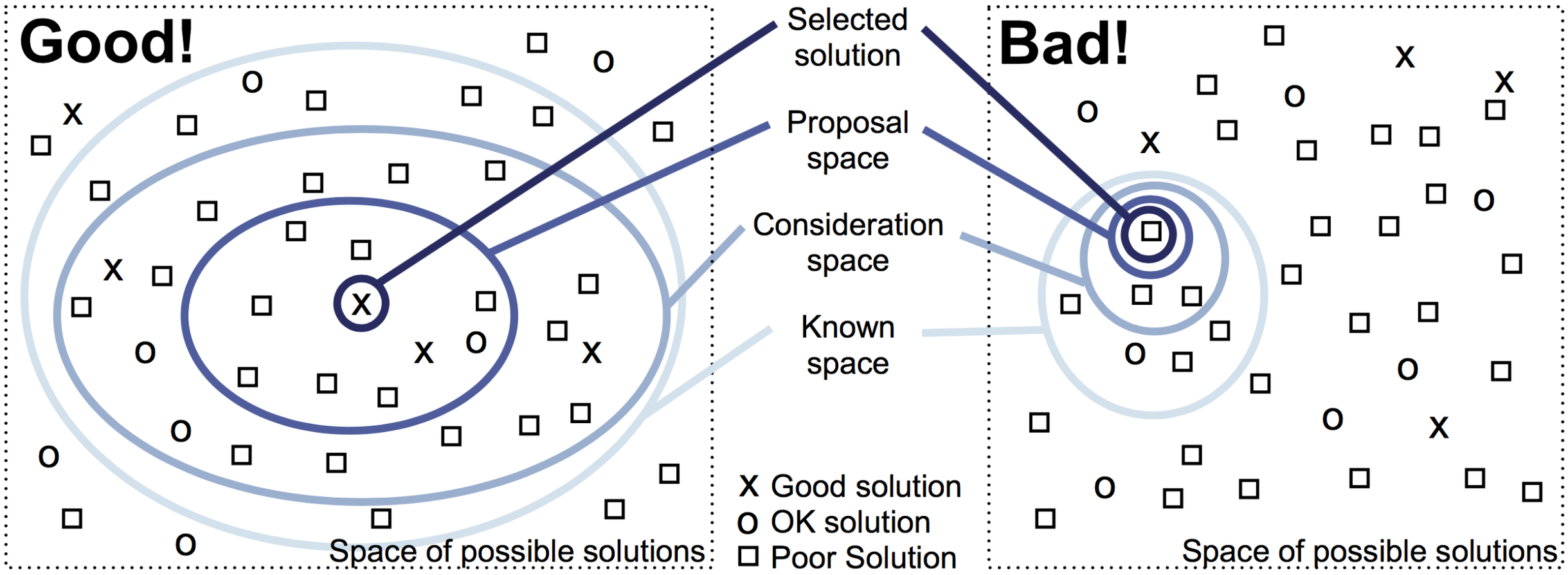
https://bl.ocks.org/mbostock/4063663

https://stateofobesity.org/adult-obesity/

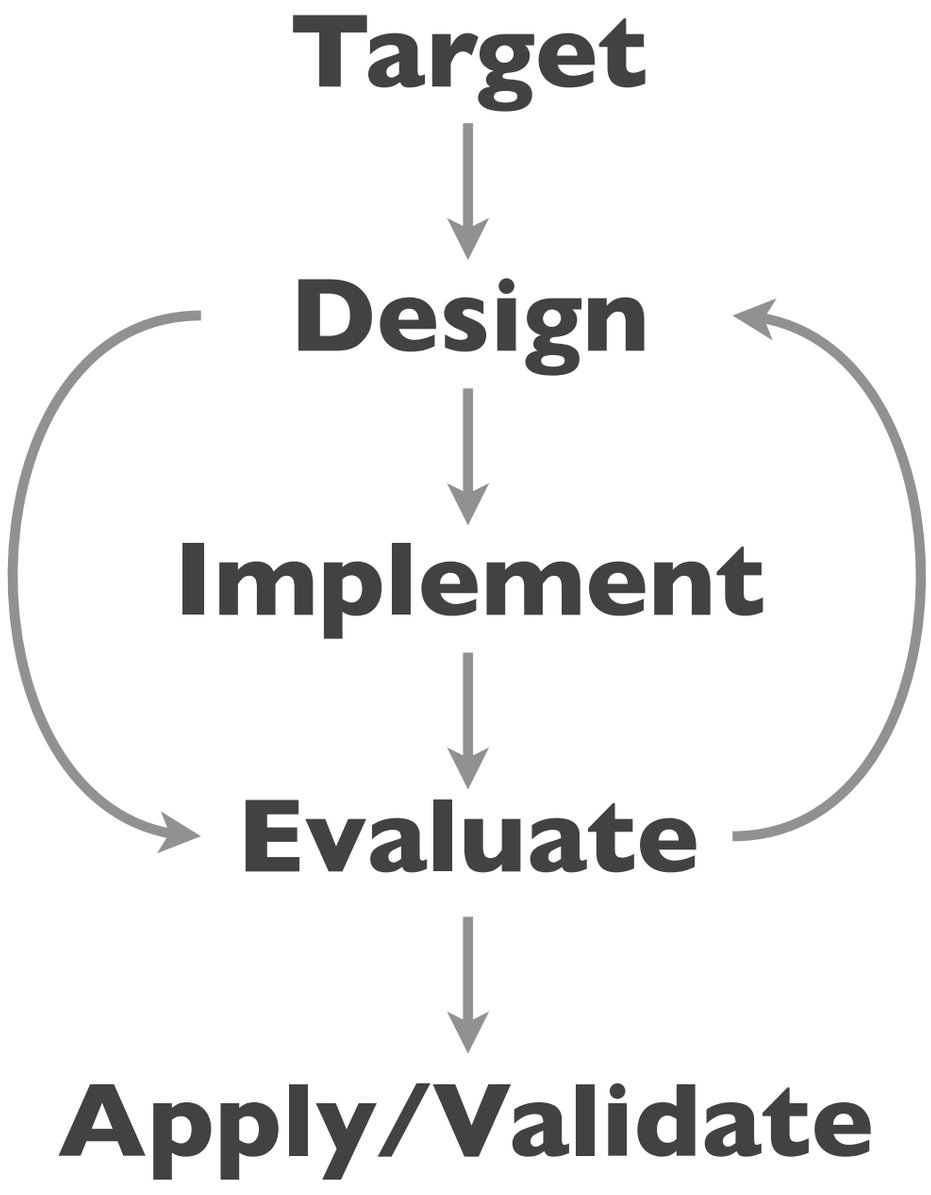
Gehlenborg and Wong, Nature Methods, 2012

Design Process

Goal of the Design Process



User-Centered Participatory Design



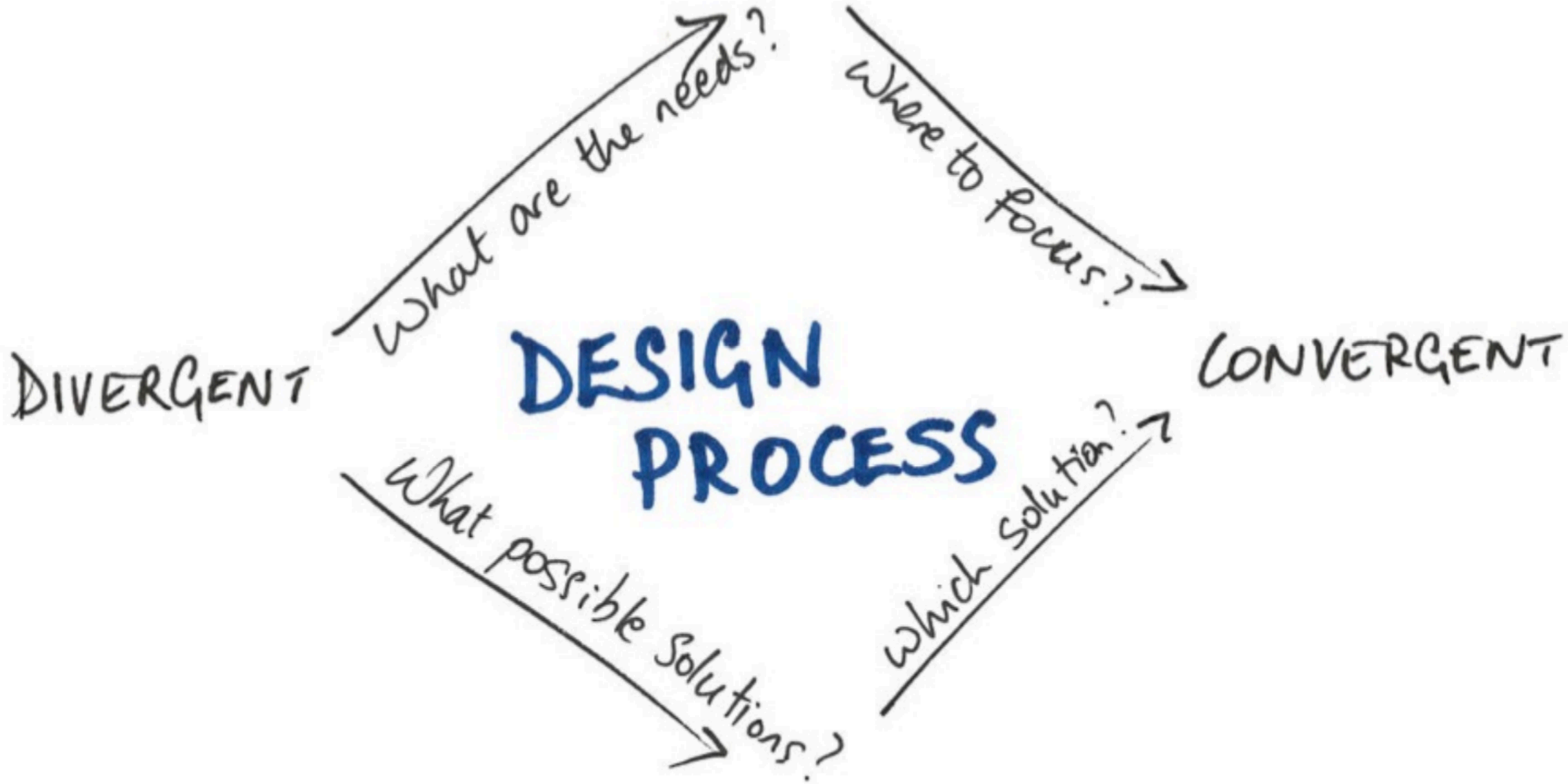
user-centered design

usability engineering

participatory design

adapted from Miriah Meyer, University of Utah and Cydney Nielsen, University of British Columbia

Five Design-Sheet Methodology



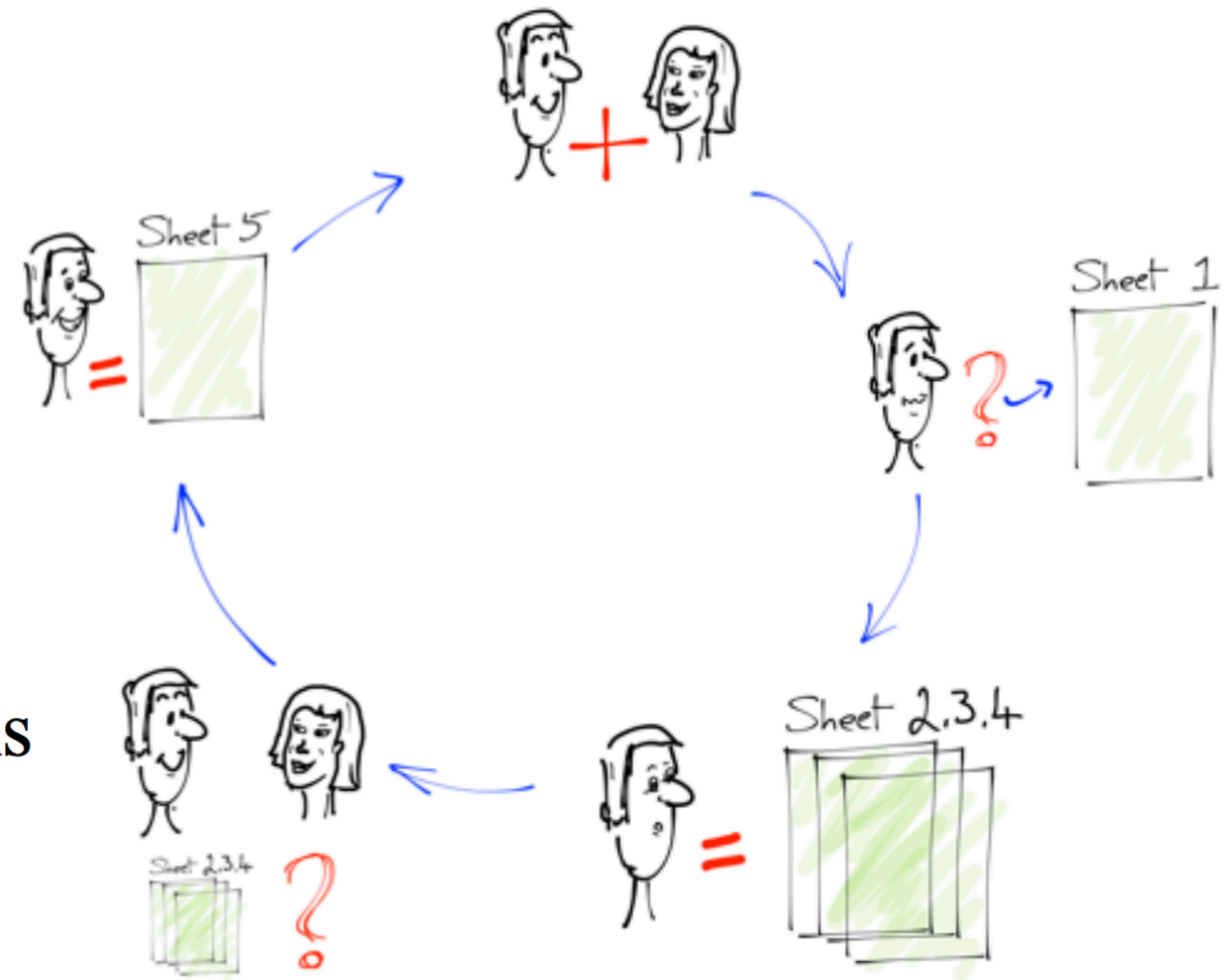
http://fds.design

Five Design-Sheet Methodology

Five sheets

Five parts per sheet

Five parts to the process



Sheet 1 = Ideas

Sheets 2,3,4 = Main designs

Sheet 5 = Realization