

DeCART Summer School Data Visualization

Day 1

2 August 2019

Nils Gehlenborg, PhD

Design Exercise

<https://github.com/hms-dbmi/decart-2019-data-visualization>

Get *decart-2019_fludata.pdf* and follow instructions.

Laboratory-Confirmed Influenza Cases in the first ten weeks of 2010

	Asia	S. America	Australia	N. America	Asia	S. America	Africa	Europe	Europe	Africa	N. America
Week	Afghanistan	Argentina	Australia	Canada	China	Colombia	Egypt	Germany	Ireland	South Africa	USA
1	5	4	2	41	2179	36	739	26	23	0	366
2	13	21	1	15	2213	36	396	24	8	1	396
3	4	6	1	8	2228	14	192	18	4	0	447
4	0	1	0	14	2027	11	80	NA	8	0	402
5	0	4	1	12	1813	8	56	NA	4	0	404
6	0	0	1	6	1353	9	47	NA	0	0	361
7	1	3	0	6	799	7	32	NA	0	0	380
8	1	1	4	7	1218	5	16	NA	1	1	424
9	NA	0	3	3	1333	7	8	3	0	0	445
10	1	3	1	7	1614	5	8	7	0	0	475

Source: World Health Organization FluNet database (<http://who.int/flunet>)

Tasks (15 minutes for discussion and sketching, 3 minutes per team for reporting back)

1. Form teams of 5 to 8 students.
2. Discuss how you would visualize this data set. *Do not ask the instructor if any questions come up during your discussion, instead note the question and decision taken.*
3. Create sketches on provided paper using the provided markers (black plus two colors).
4. Report results back to class, including the questions that have come up in your discussion.

Design Exercise

- Interaction? Static?
- Multiple Views?
- Encoding? # of classes?
- Audience?
- Preprocessing and transformations? Normalization? Imputation?
- Use case and application? Explanation? Exploration?

Learning Goals

- What is data visualization, how does it work, and what are the principles behind it? (Day 1)
- How can we design and evaluate visualizations? How can we visualize tabular and network data? (Day 2)
- How can we visualize time-series and genomic data? (Day 3)
- How to use Altair <https://altair-viz.github.io>) to make interactive visualizations in Python.

What is data visualization?

What is data visualization?

The use of computer-supported, interactive, visual representations of data to amplify cognition.

– Stu Card, Jock Mackinlay & Ben Shneiderman

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

– Tamara Munzner

What is data visualization?

The use of computer-supported, interactive, visual representations of data to amplify cognition.

— Stu Card, Jock Mackinlay & Ben Shneiderman

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

— Tamara Munzner

What is data visualization?

Human

Data

Visualization

What is data visualization?

The purpose of computing is insight, not numbers.

— Richard Hamming

The purpose of **visualization** is insight, **not pictures**.

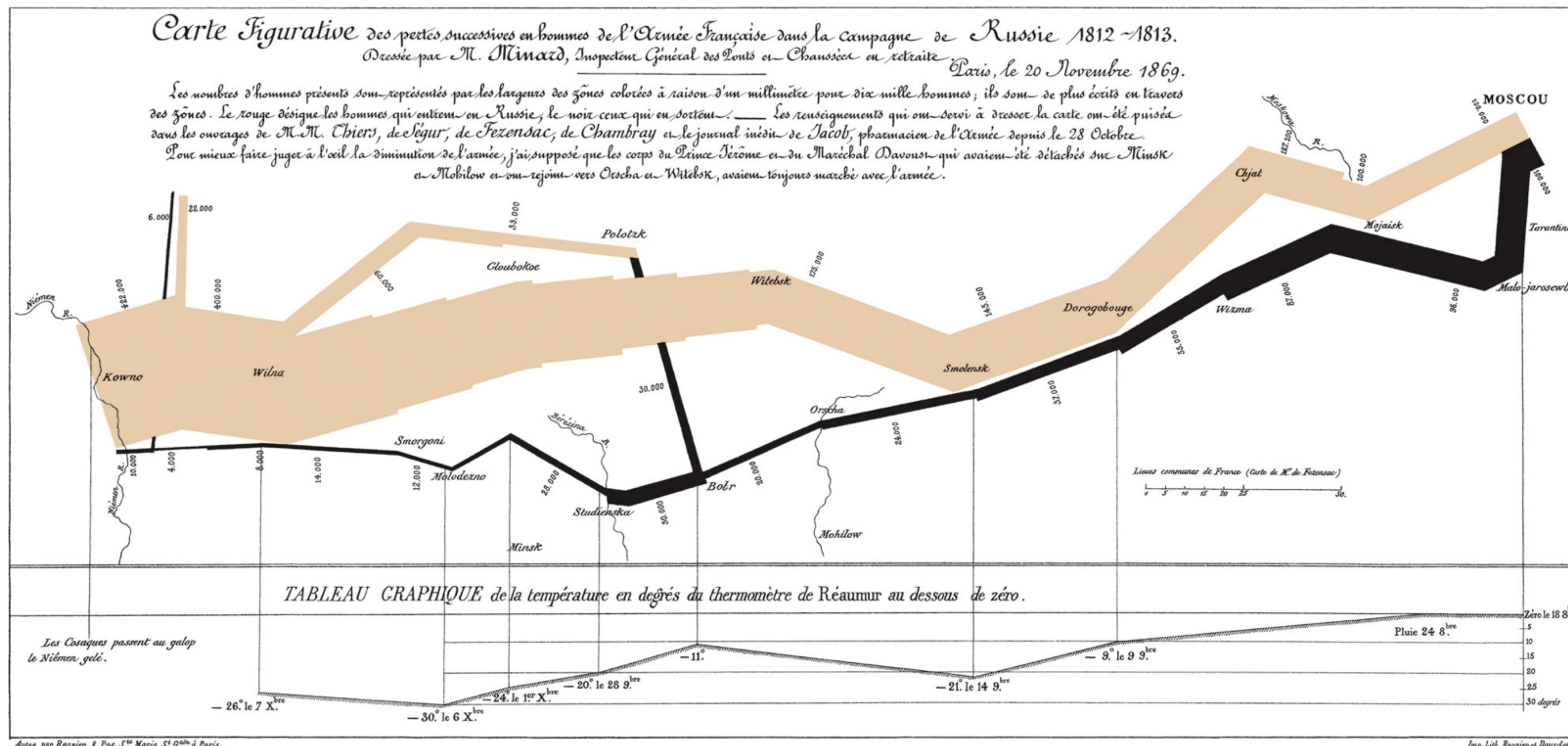
— Stu Card, Jock Mackinlay & Ben Shneiderman

Why do we need it?

A good sketch is better than a long speech.

—Napoleon Bonaparte

Charles Minard: Napoleon's March on Moscow



Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent Orscha et Wilebsk, avaient toujours marché avec l'armée.

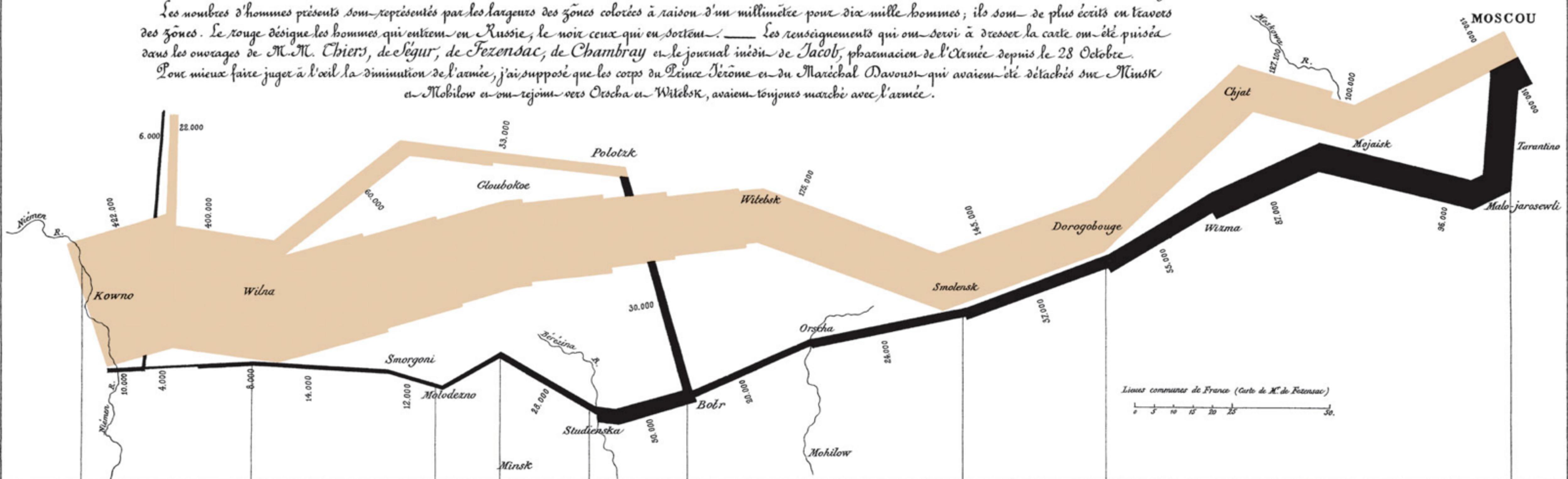


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Niémen gelé.



Crossing the Berezina River, Peter von Hess



Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent Orscha et Wilebsk, avaient toujours marché avec l'armée.

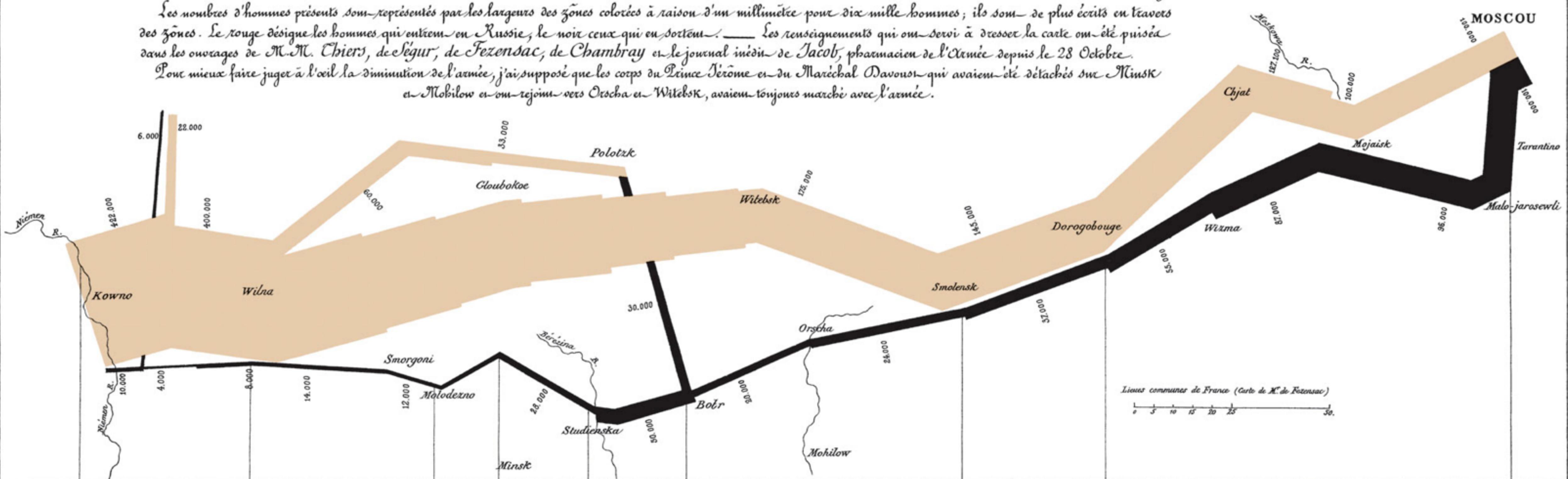
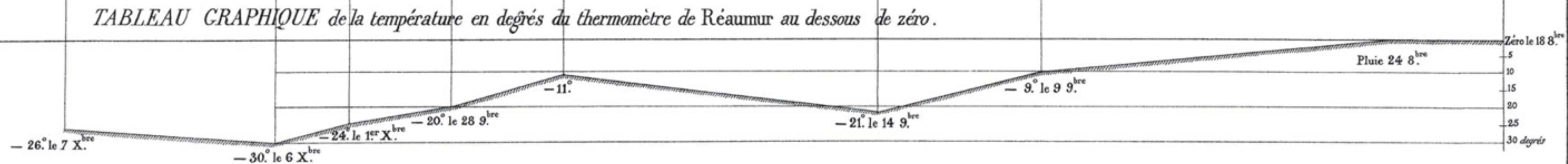
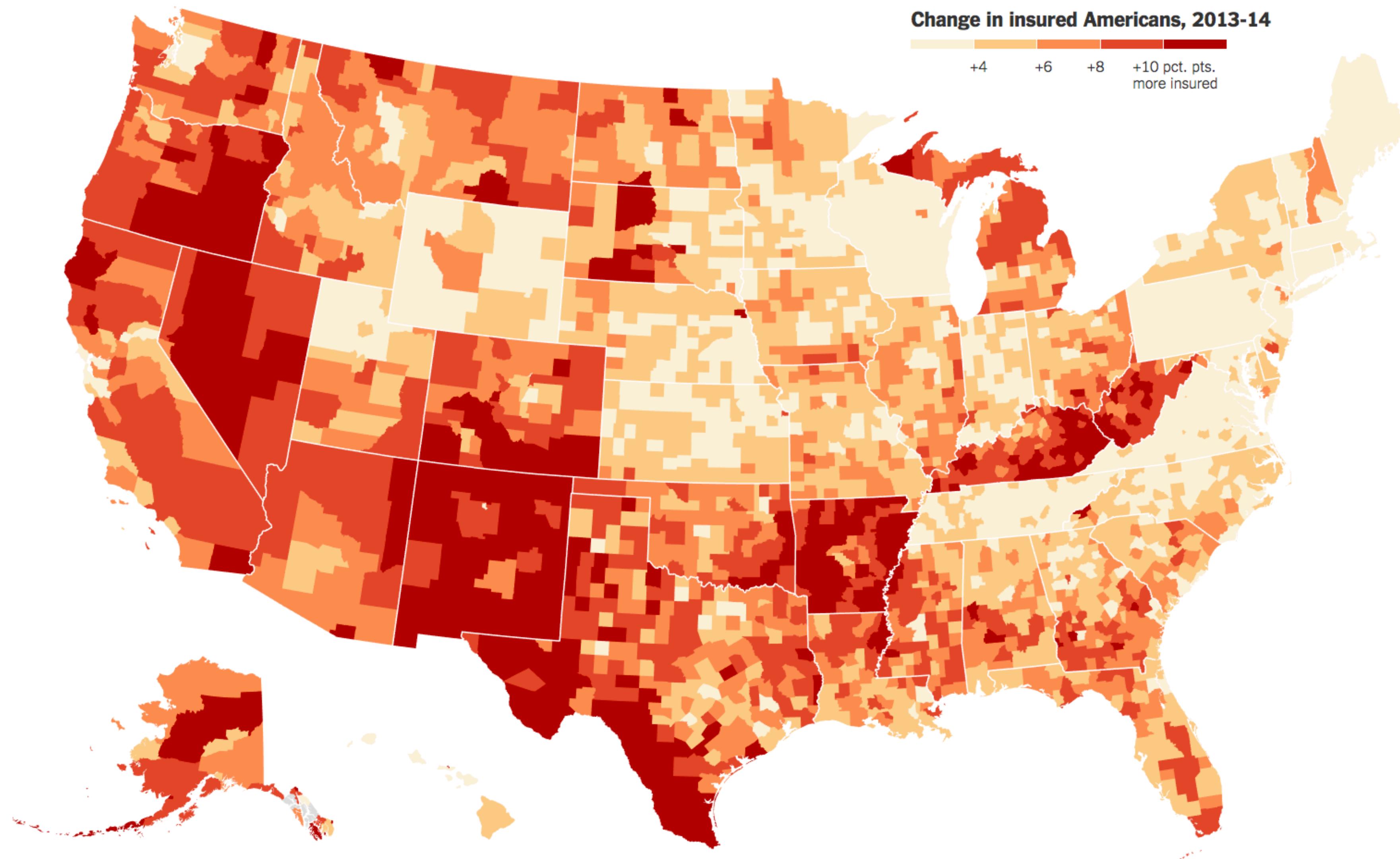


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

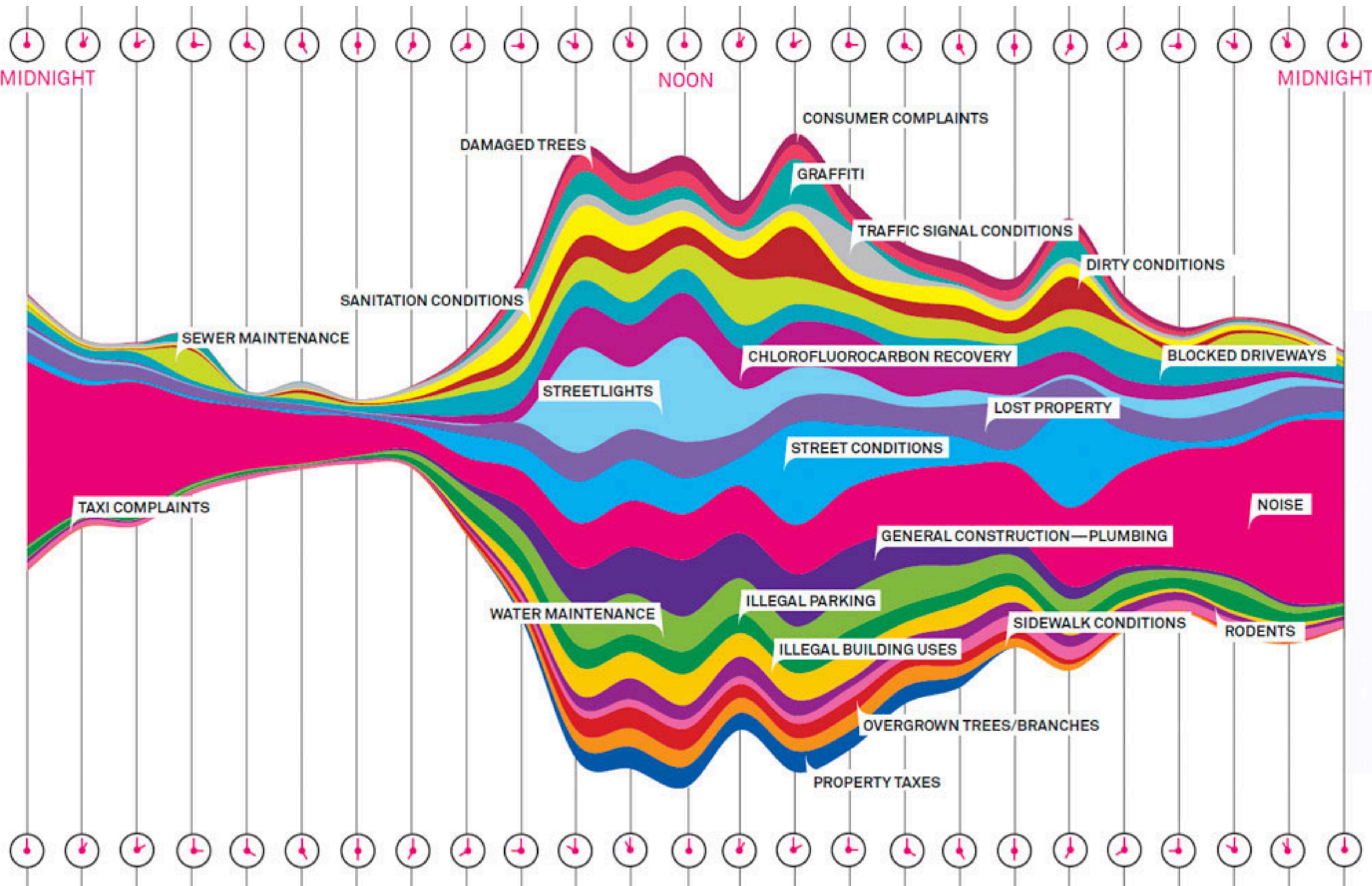
Les Cosaques passent au galop
le Niémen gelé.



New York Times: ACA Health Insurance Gains 2013 - 2014



Wired: 34,522 311 calls in New York City between 9/8/10 and 9/15/10



Why do we need it?

I believe it when I see it.

— Unknown

Table 1.1: Anscombe's Quartet (Anscombe, 1973). In each of the four data sets mean $\mu_{X_i} = 9.0$, variance $\sigma_{X_i}^2 = 11.0$, $\mu_{Y_i} = 7.5$, $\sigma_{Y_i}^2 = 4.12$, correlation $\text{cor}(X_i, Y_i) = 0.816$ and the linear regression line is $Y_i = 3 + 0.5X_i$ for $i \in \{1, 2, 3, 4\}$.

X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

$\text{mean}(X) = 9$, $\text{var}(X) = 11$, $\text{mean}(Y) = 7.5$, $\text{var}(Y) = 4.12$,
 $\text{cor}(X,Y) = 0.816$, linear regression line $Y = 3 + 0.5*X$

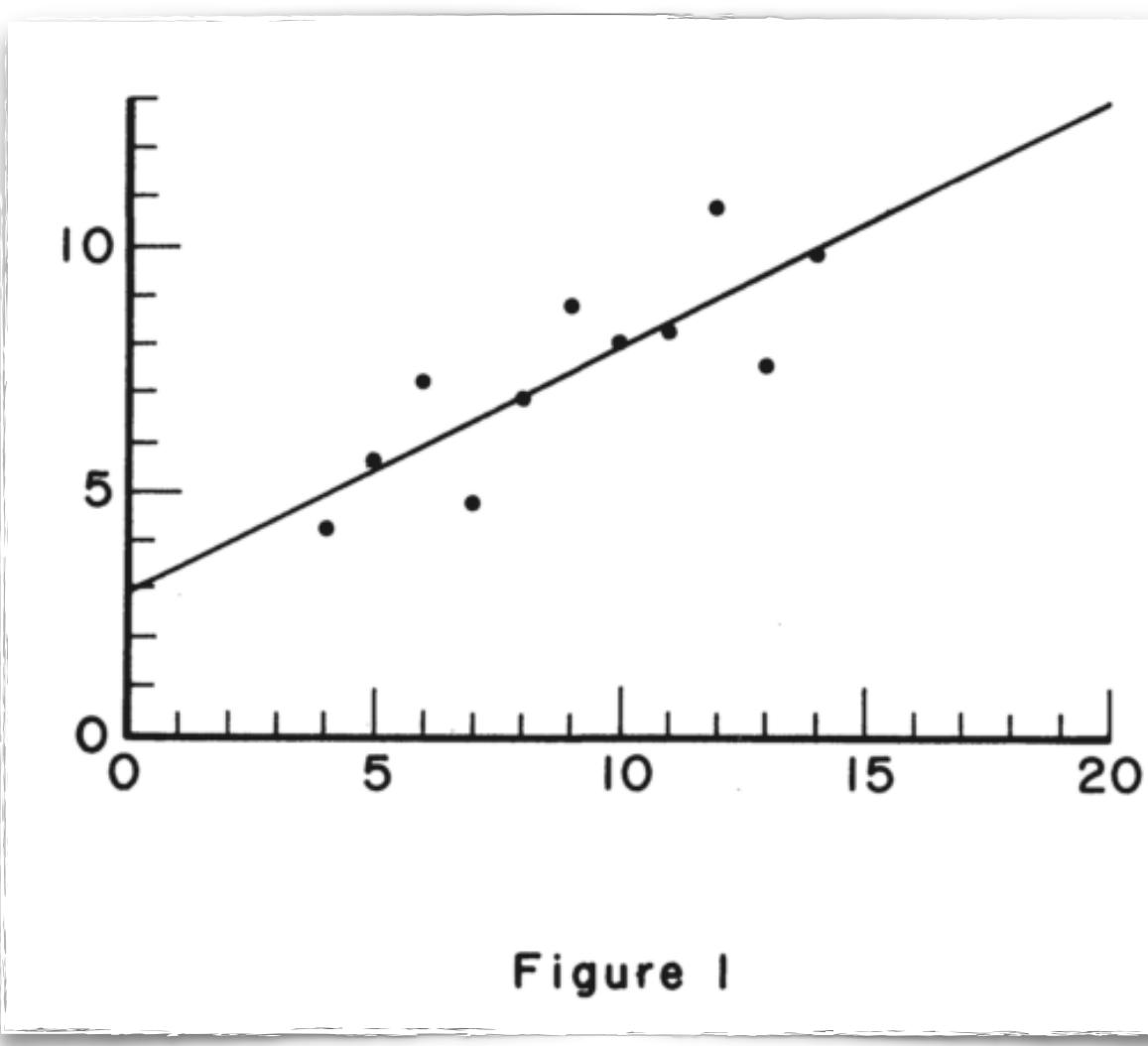


Figure 1

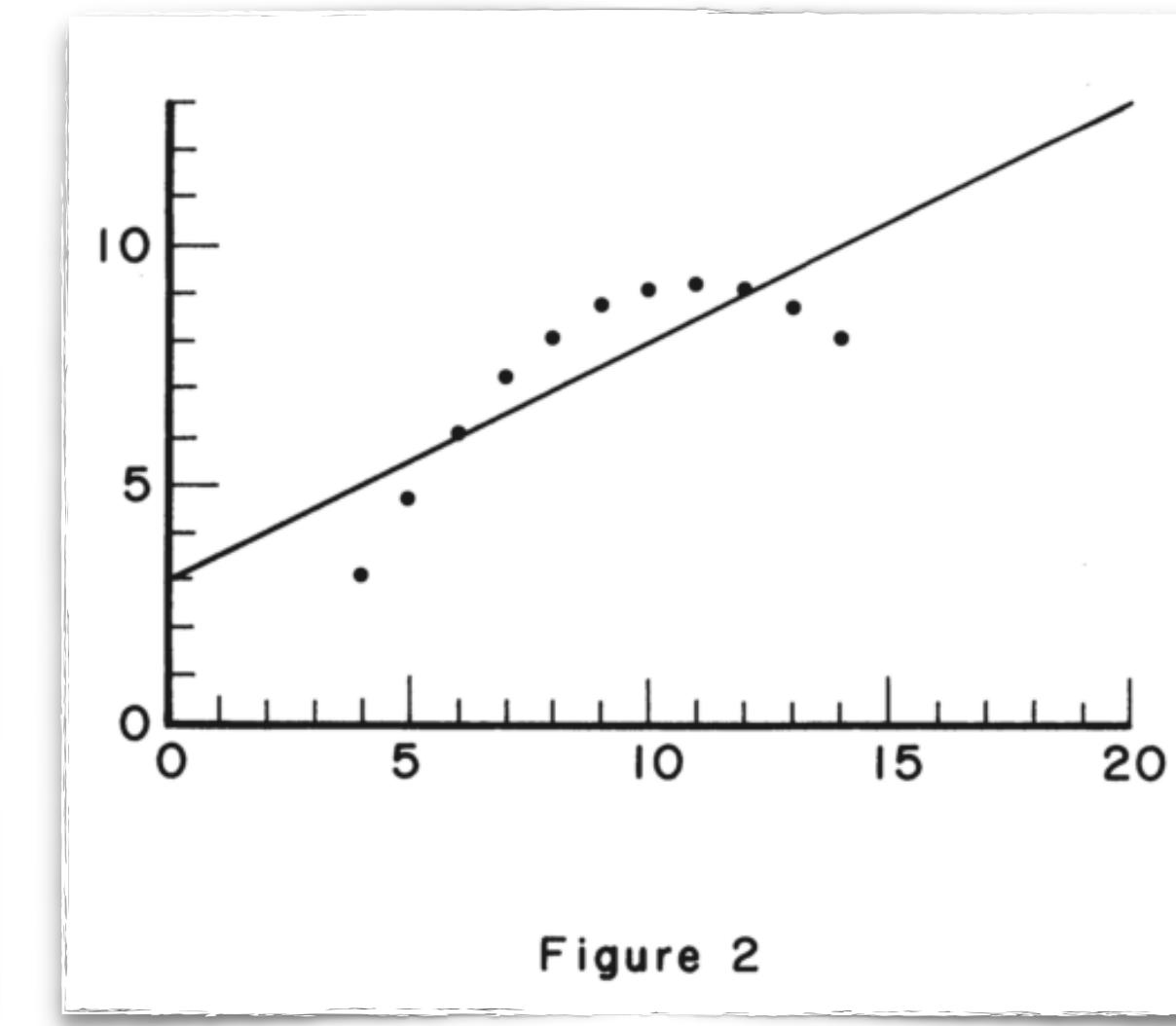


Figure 2

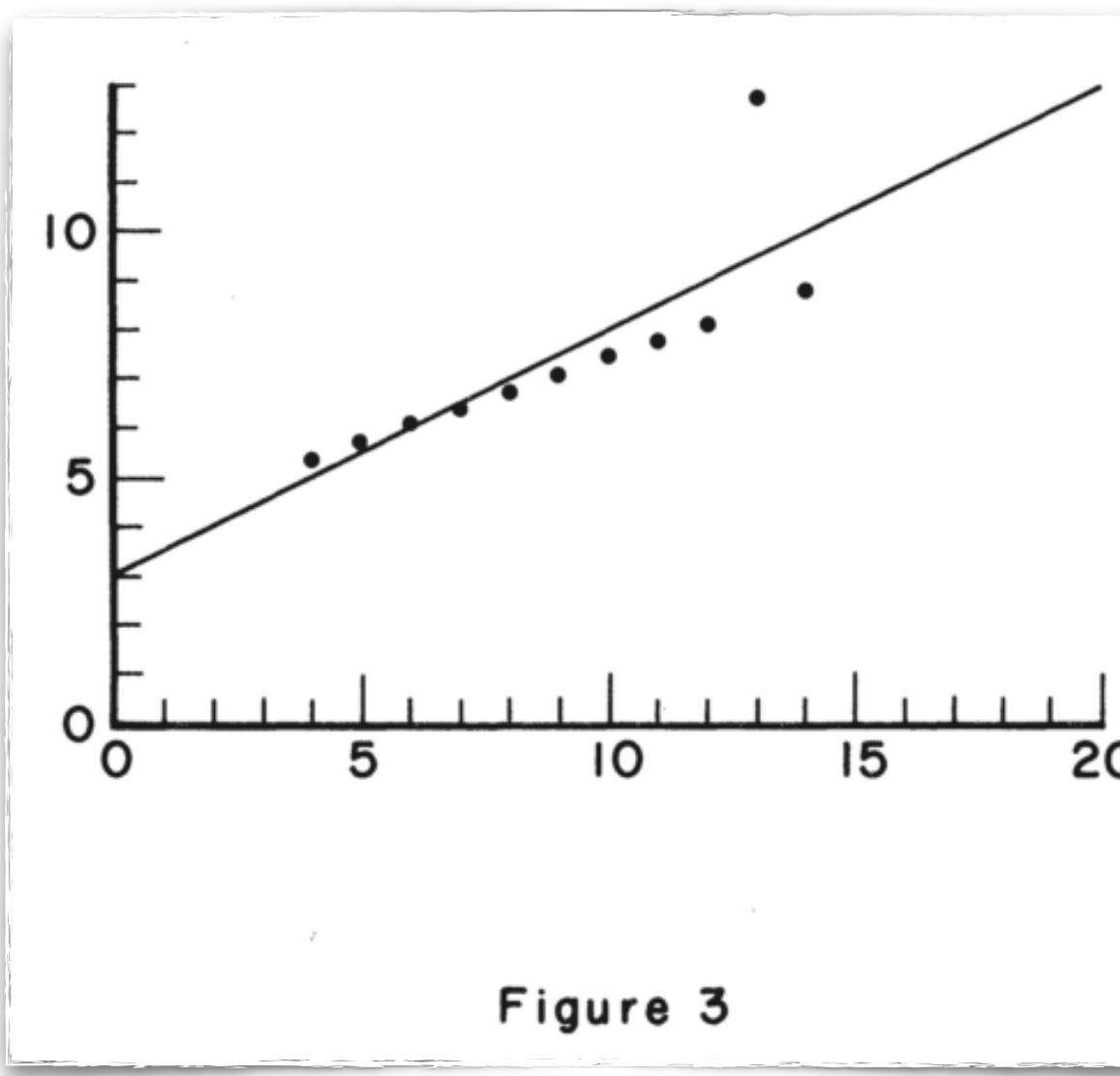


Figure 3

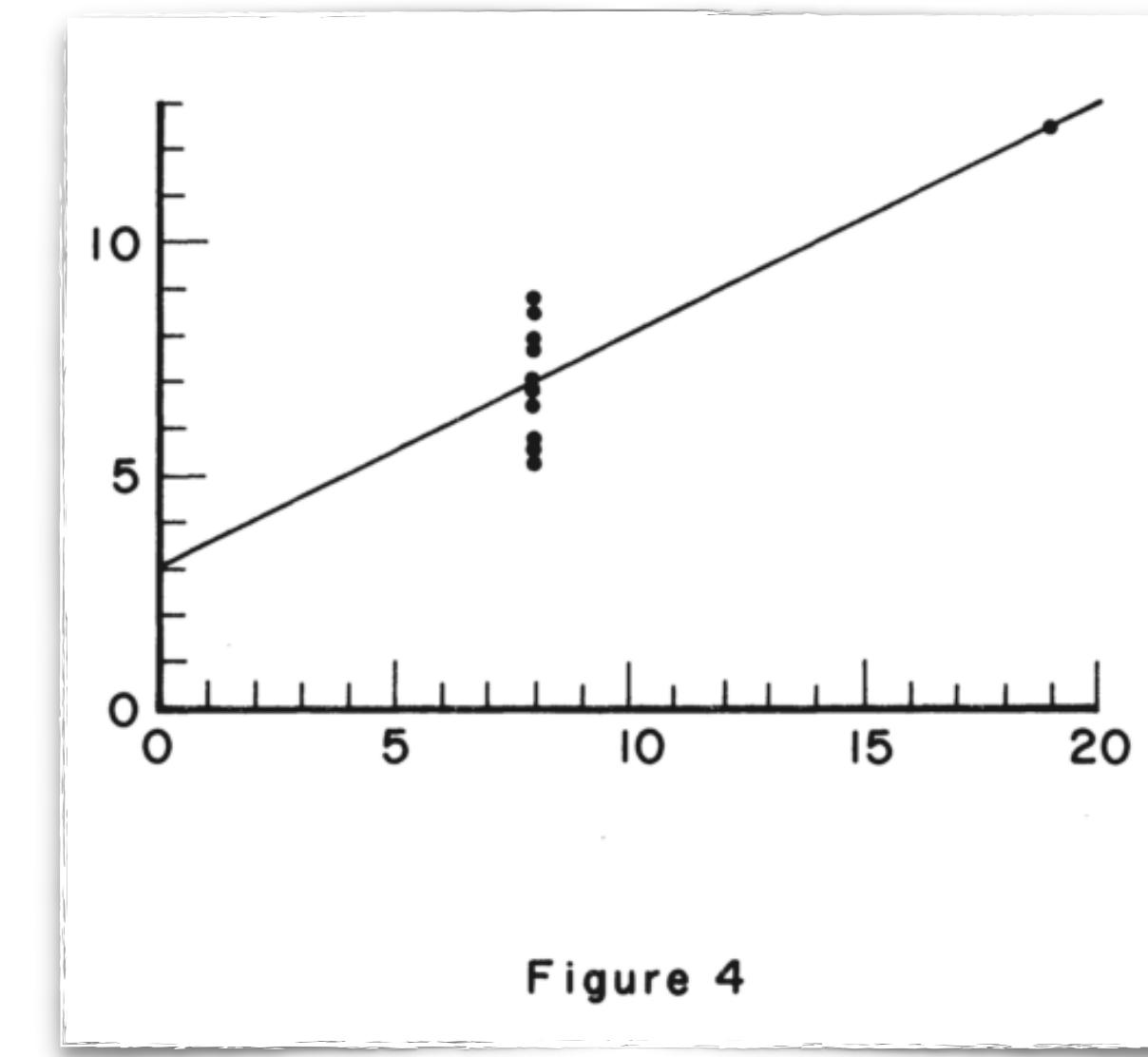
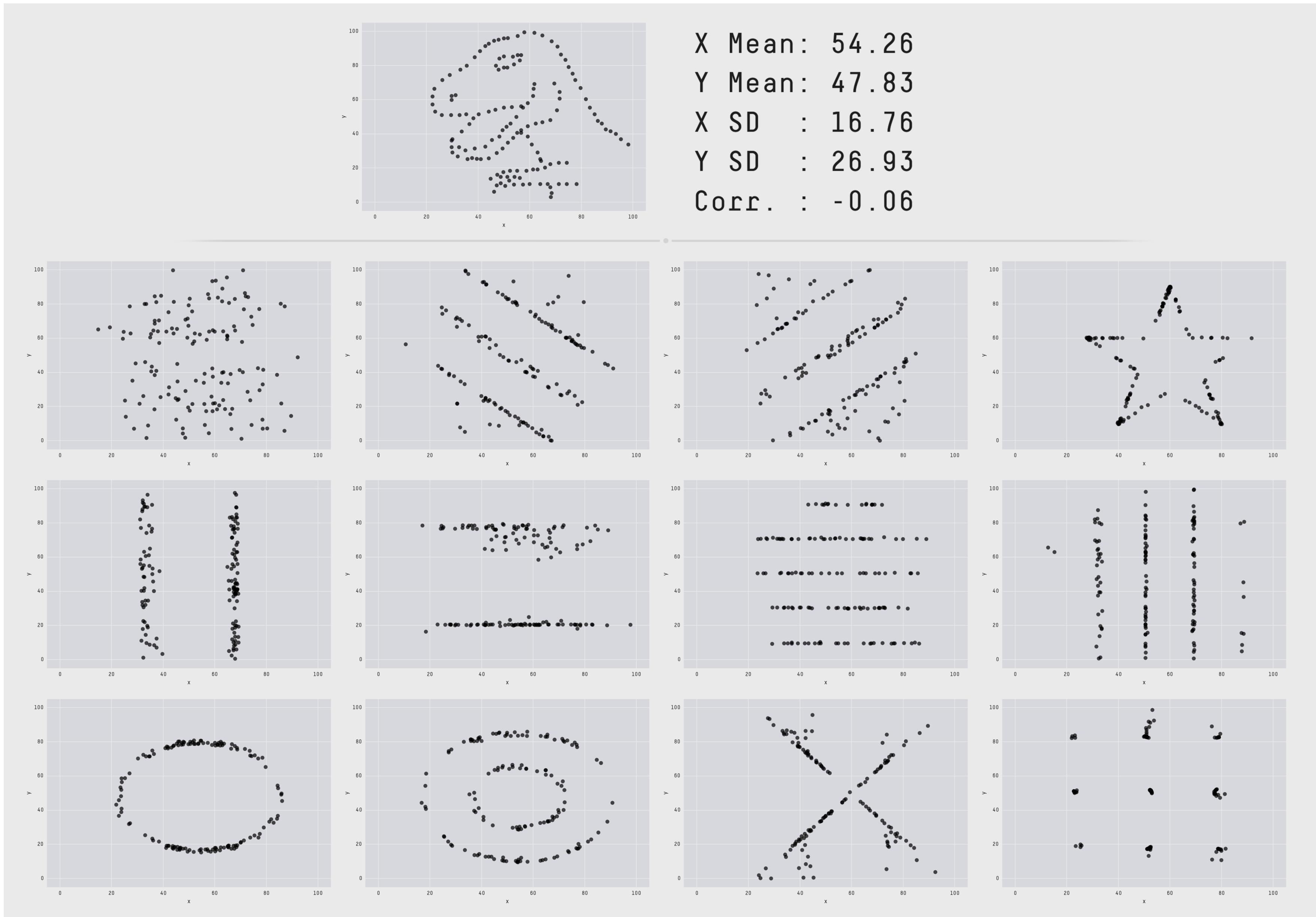


Figure 4

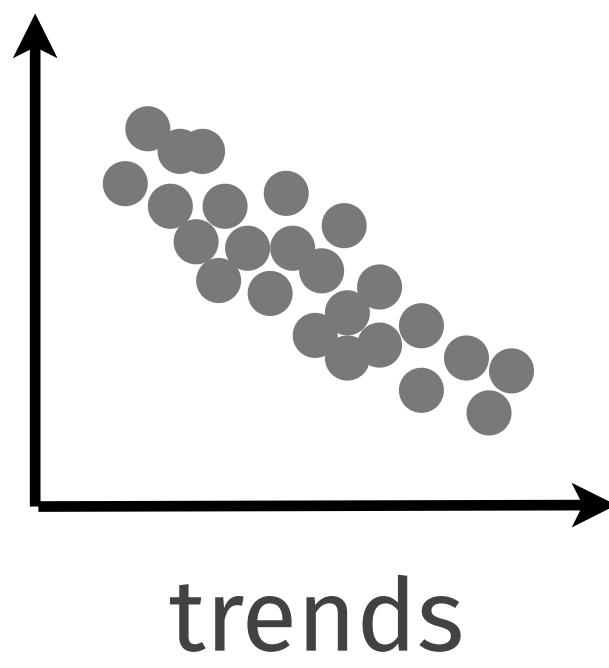


Why do we need it?

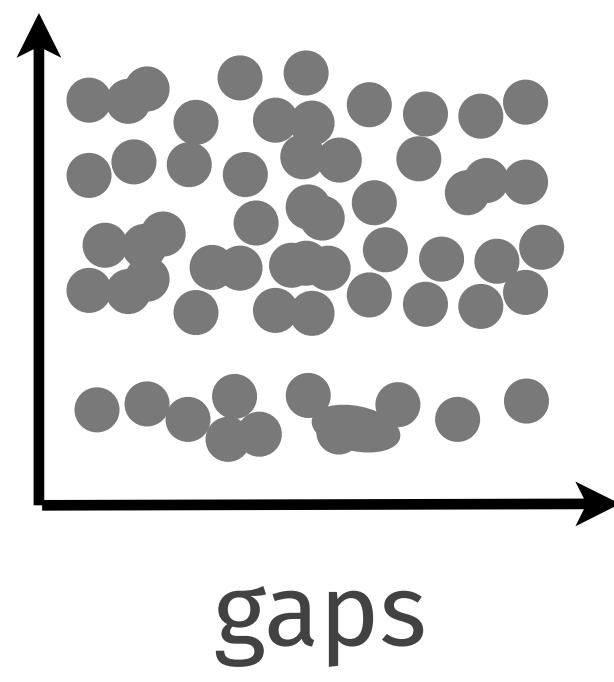
I'm wondering if there are any interesting patterns in my data.

—Almost Everyone

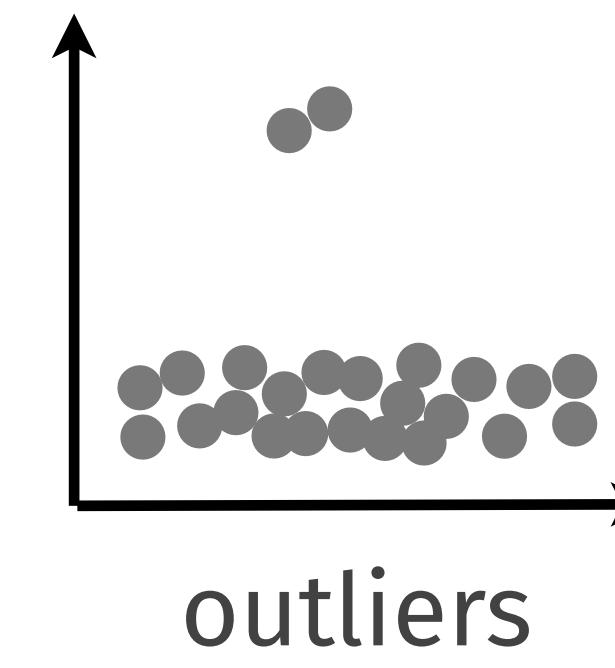
Exploration: Hypothesis Generation



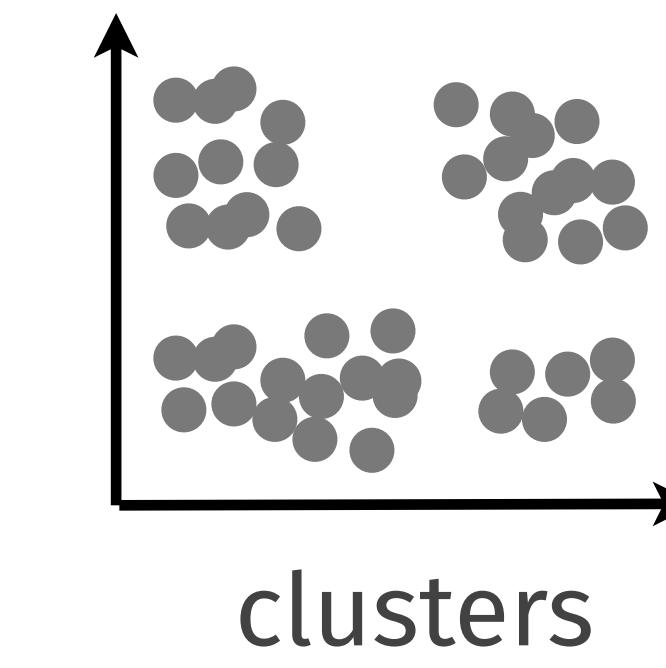
trends



gaps

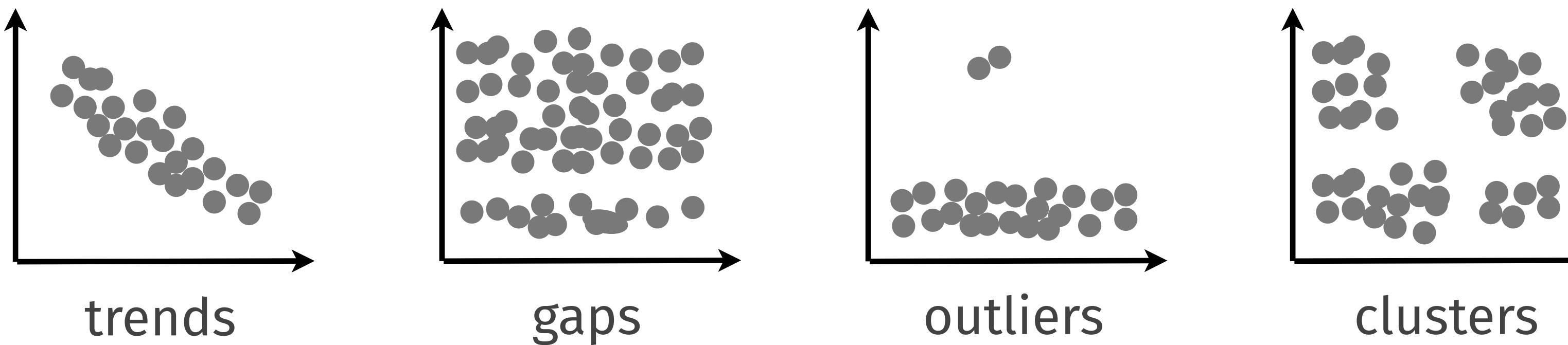


outliers



clusters

Exploration: Hypothesis Generation



Why? Generate hypotheses that can be tested with statistical methods or follow-up experiments.

How? Visualization is employed to perform pattern detection using the human visual system.

Visualization Use Cases

Exploration
Confirmation
Explanation

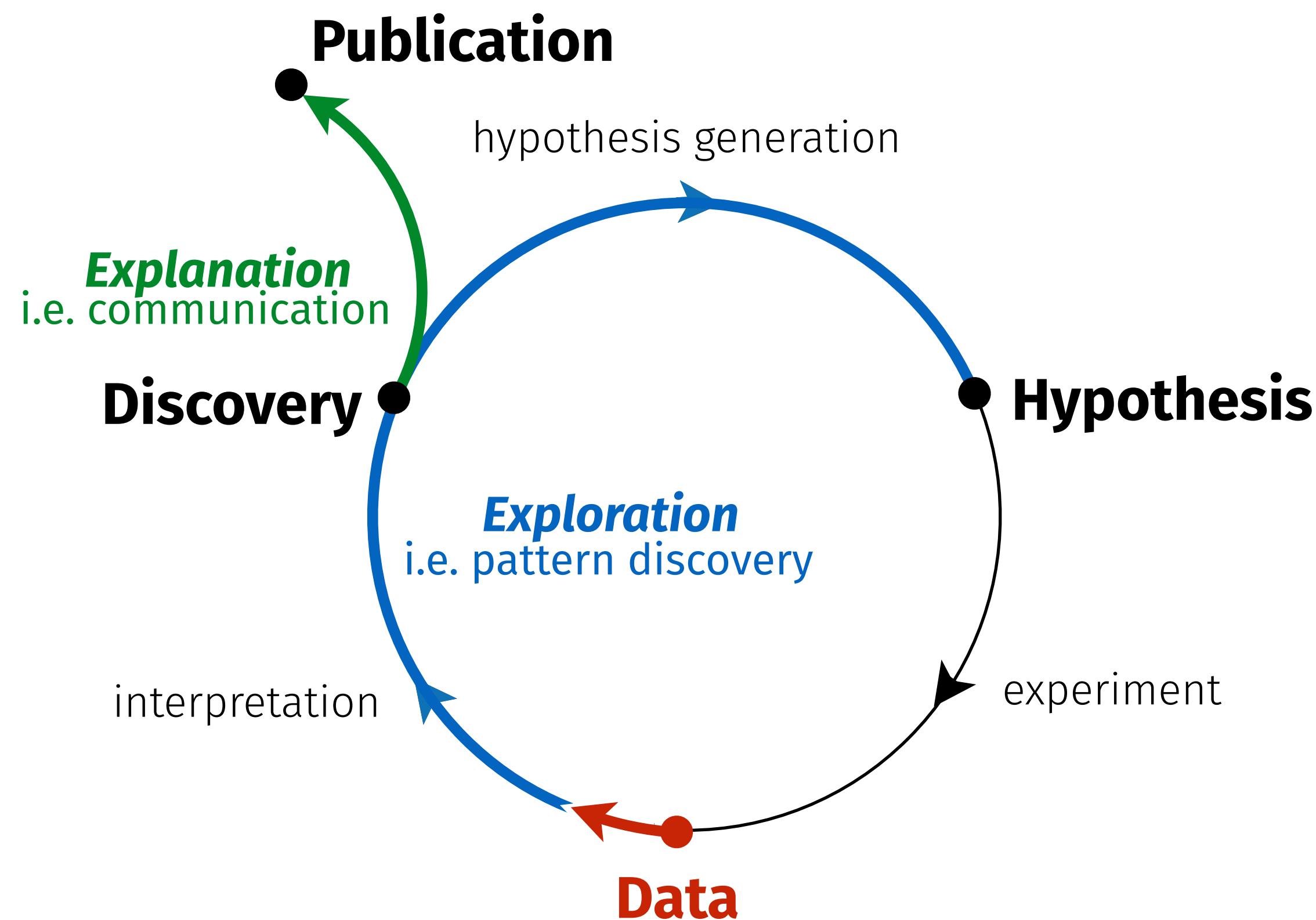
Visualization Use Cases

Exploration

Confirmation

Explanation

Discovery Process

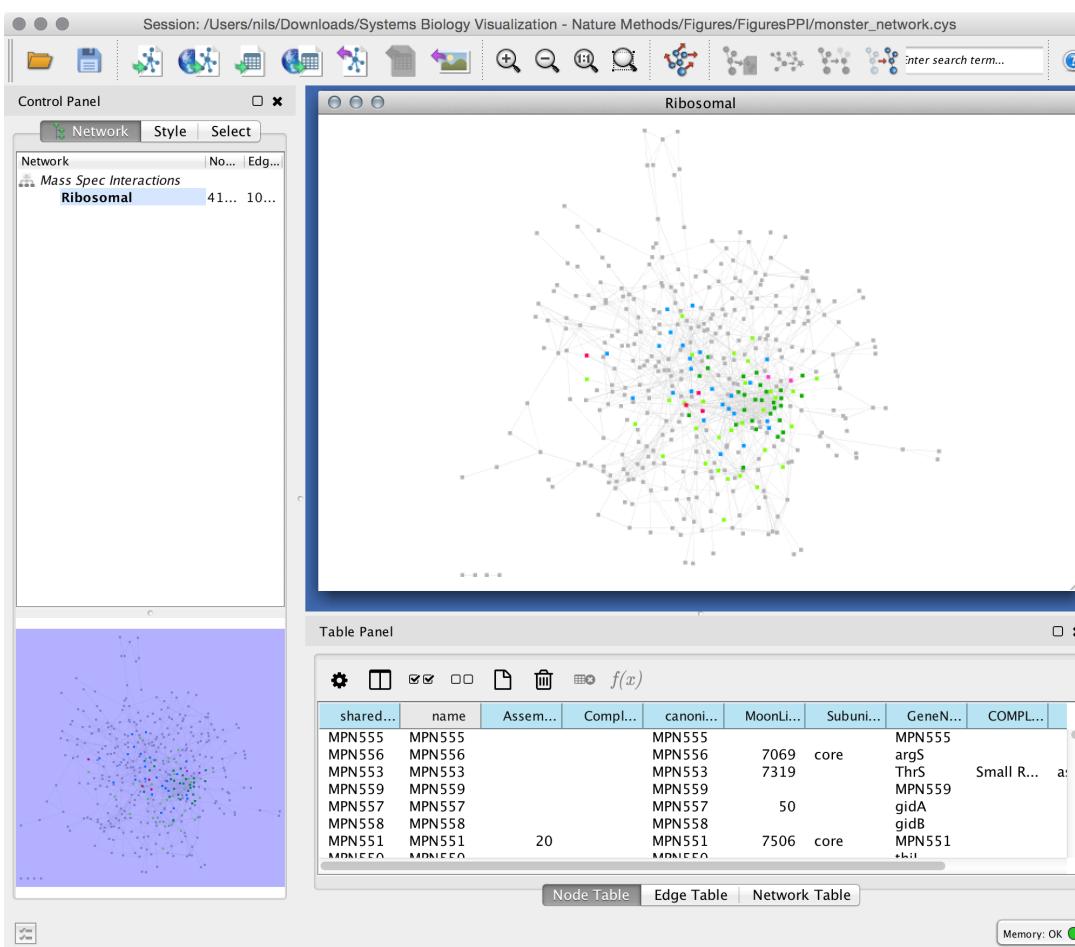


Discovery Process



Discovery Process

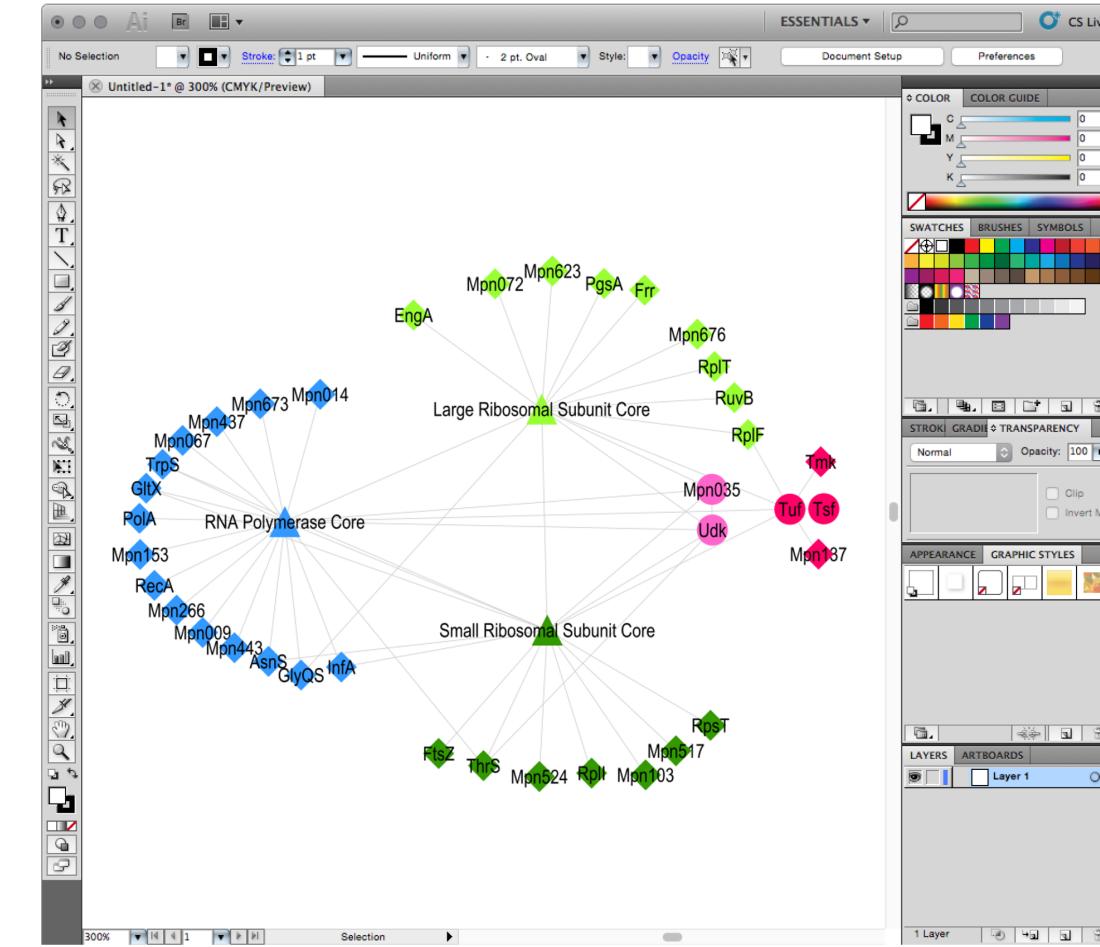
Exploration



Cytoscape

Insight

Explanation



Discovery Process



DeCART 2019 Visualization Course



Syllabus

- Session 1: Introduction to Data Visualization
- Session 2: Design Process, Evaluation, and Interaction
- Session 3: High-Dimensional Data
- Session 4: Networks
- Session 5: Genomes and Epigenomes
- Session 6: Time and Event Sequences
- Session 7: *Project Presentations* and Review

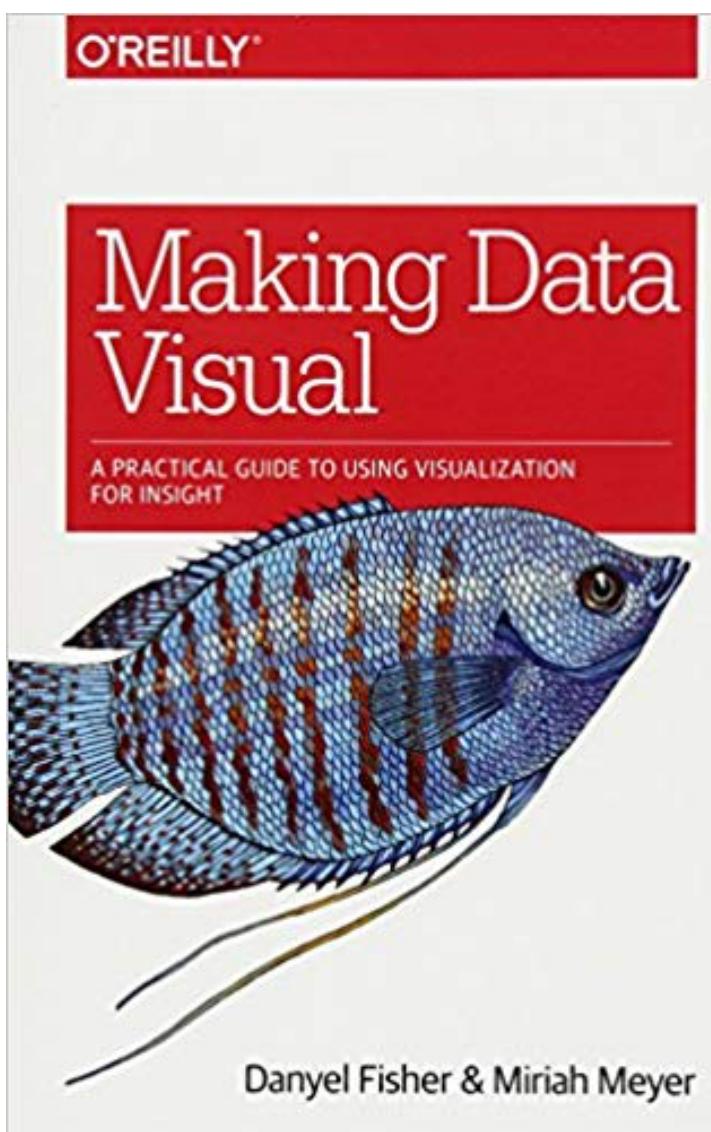
Syllabus

Canvas

<https://canvas.harvard.edu/courses/55733>

Reading

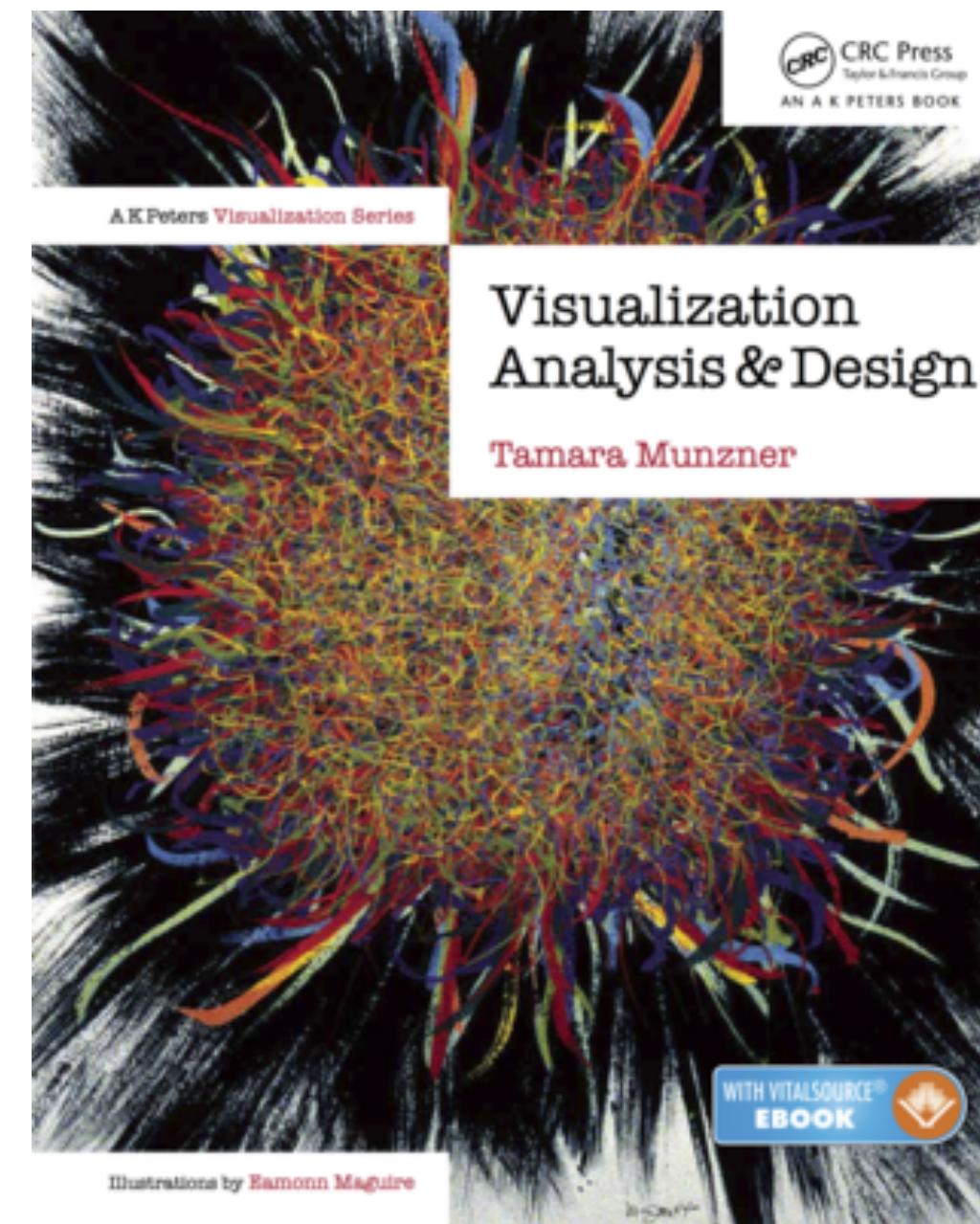
Casual Reading



Danyel Fisher & Mariah Meyer, Making Data Visual, O'Reilly, 2018

Reading

Textbook with In-Depth Background



Tamara Munzer, Visualization Analysis & Design, CRC Press, 2014

Gehlenborg Lab @ DBMI

Methodology

Data Visualization

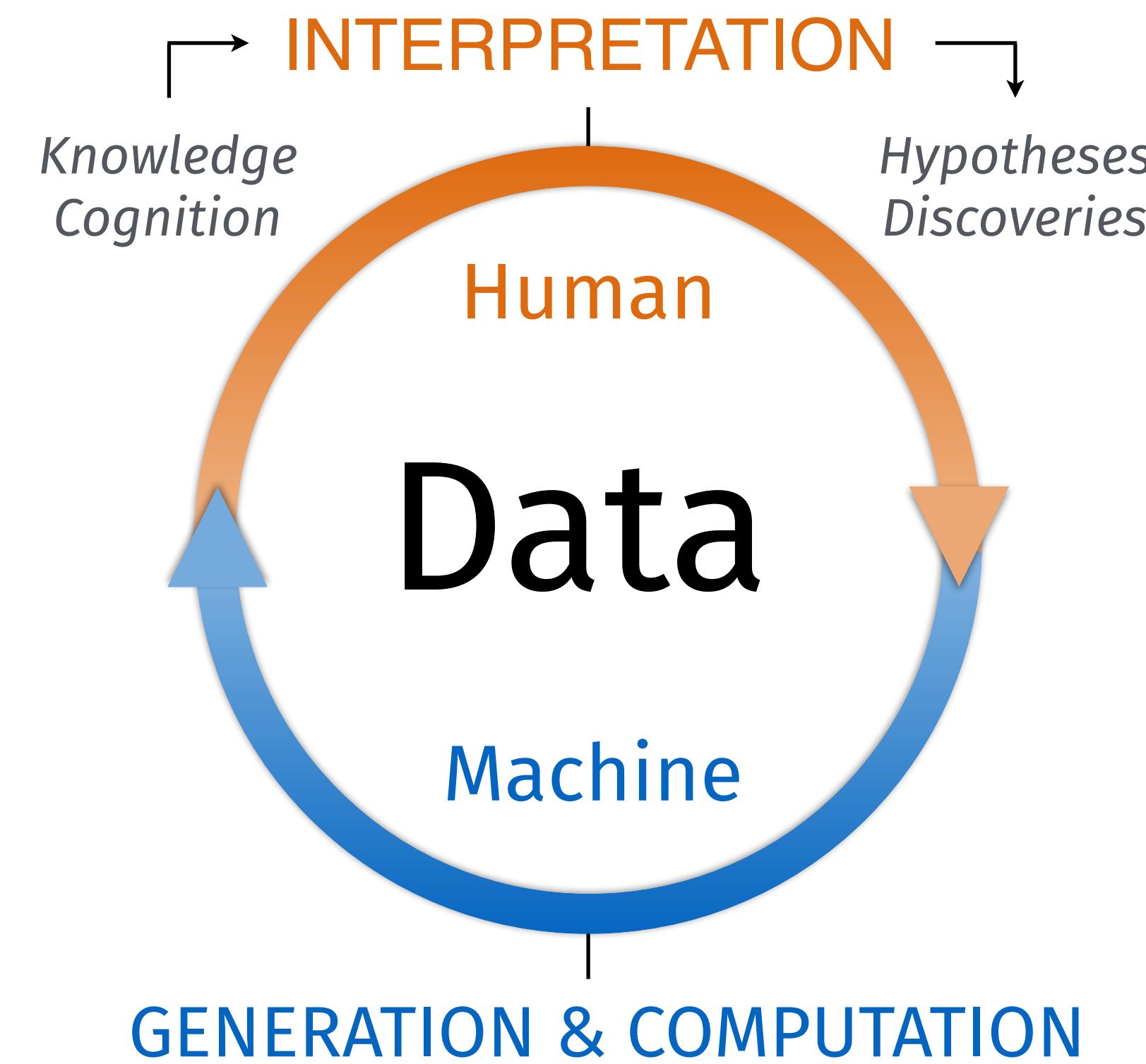
Reproducible Research

Applications

Cancer Genomics

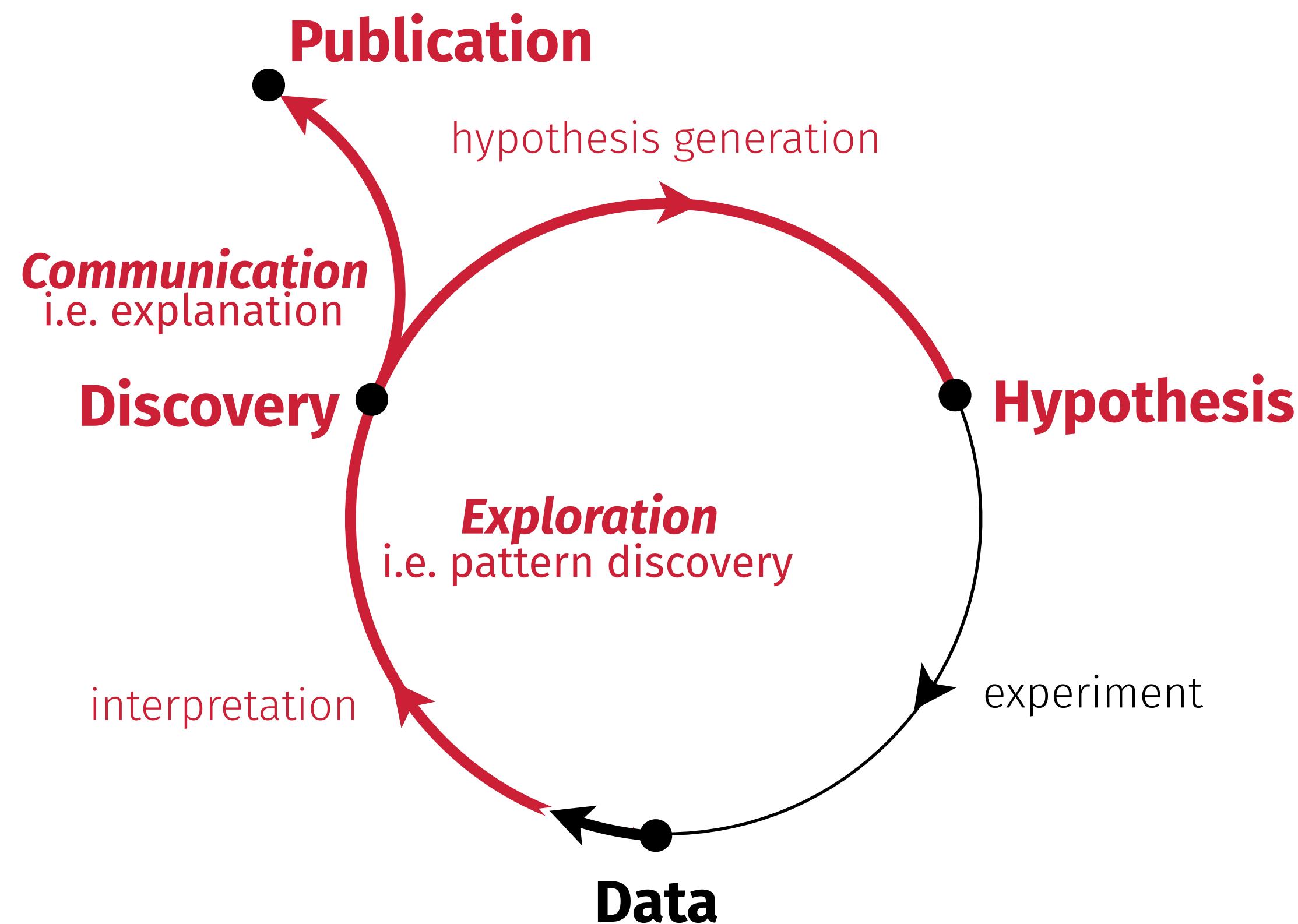
3D Genome Structure

Methods for Data Visualization and Exploration

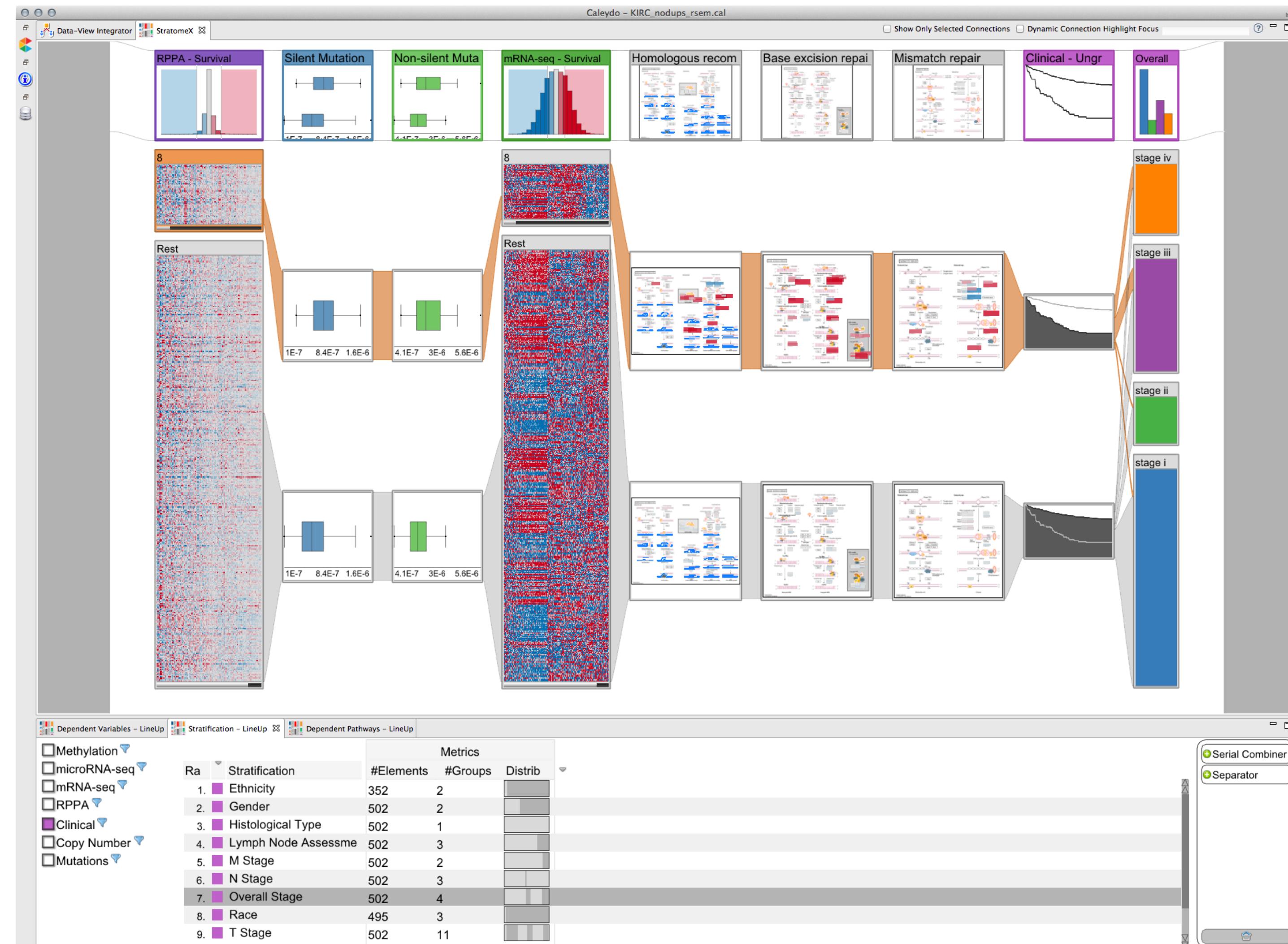


Tools for Reproducible Research

Spectrum of Projects

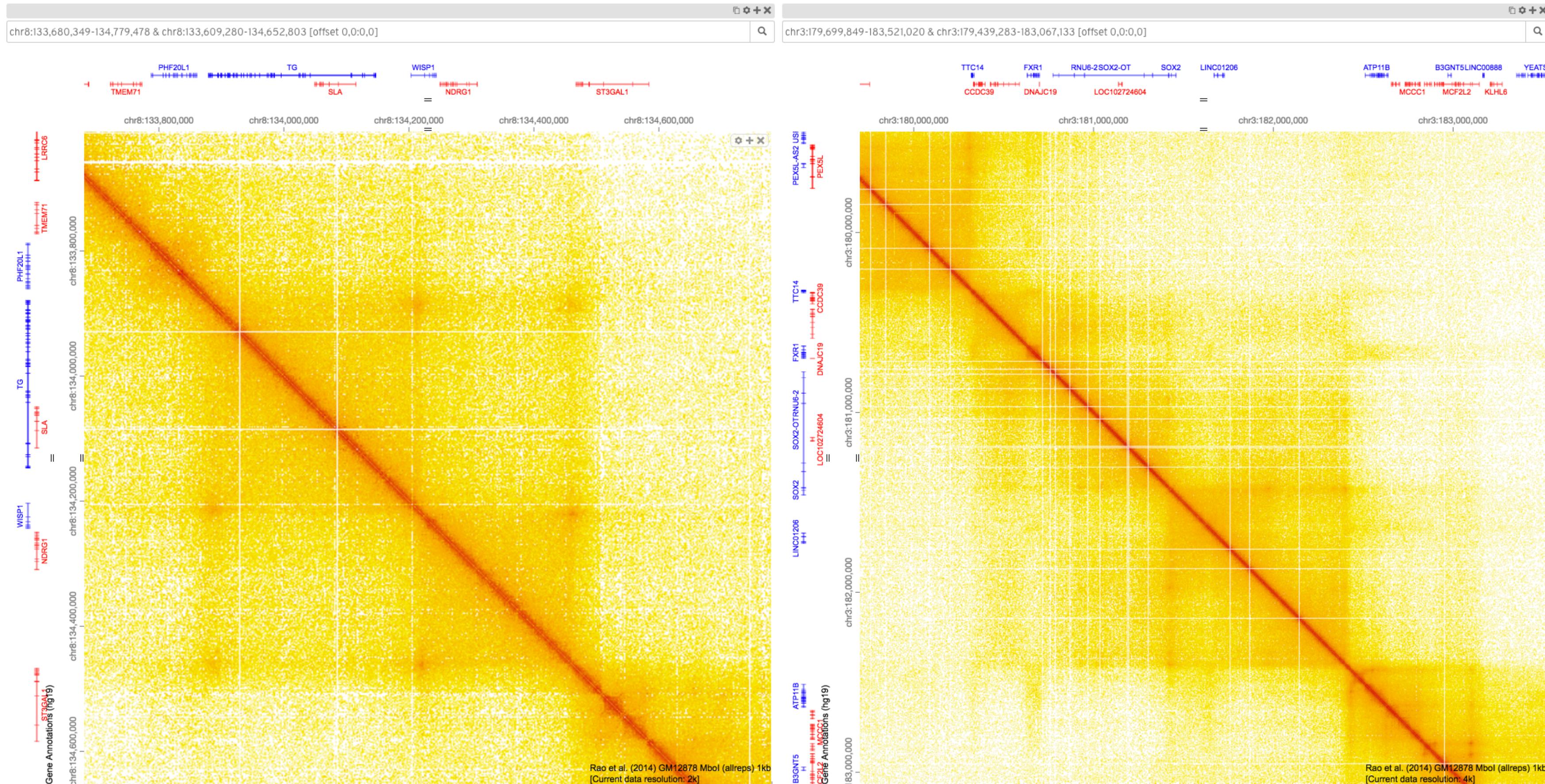


Methods for Data Visualization and Exploration



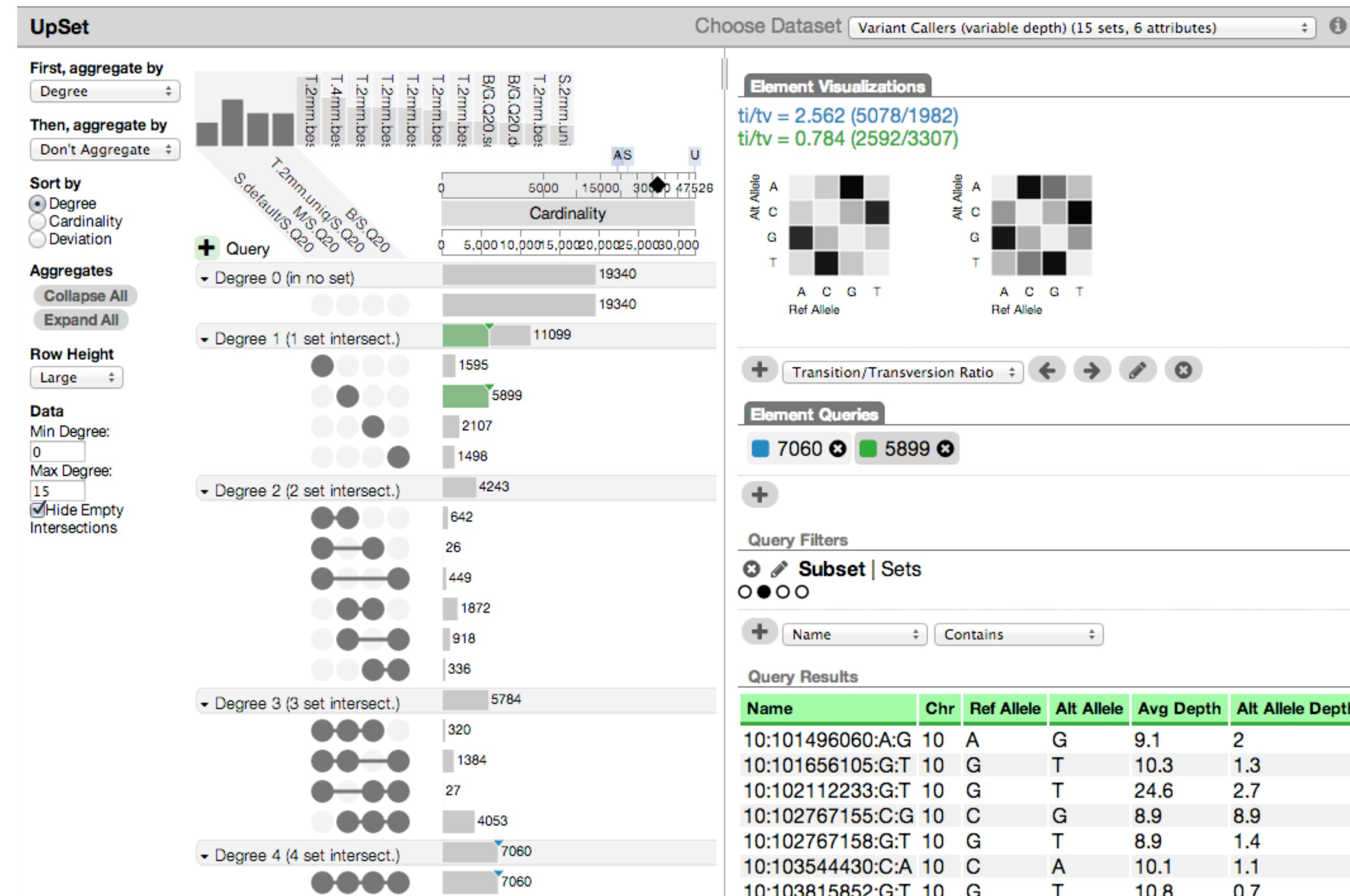
StratomeX - Cancer Subtypes: <http://stratomex.caleydo.org>

Methods for Data Visualization and Exploration



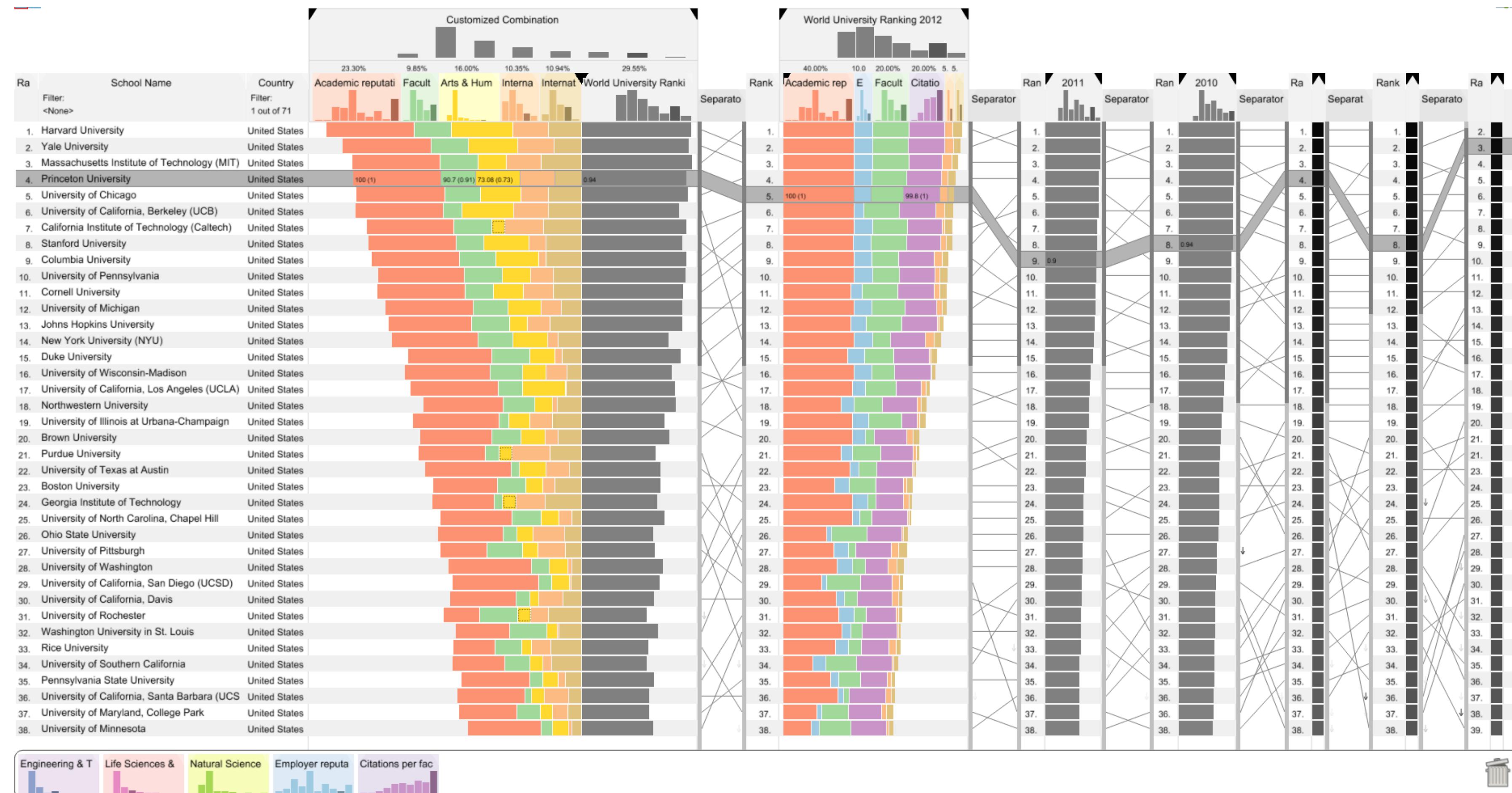
HiGlass - Visualization of Chromatin Interaction Maps: <http://higlass.io>

Methods for Data Visualization and Exploration



Upset & UpSetR - Visualization of Set Intersections: <http://upset.caleydo.org>

Methods for Data Visualization and Exploration



LineUp - Visualization of Rankings: <http://lineup.caleydo.org>

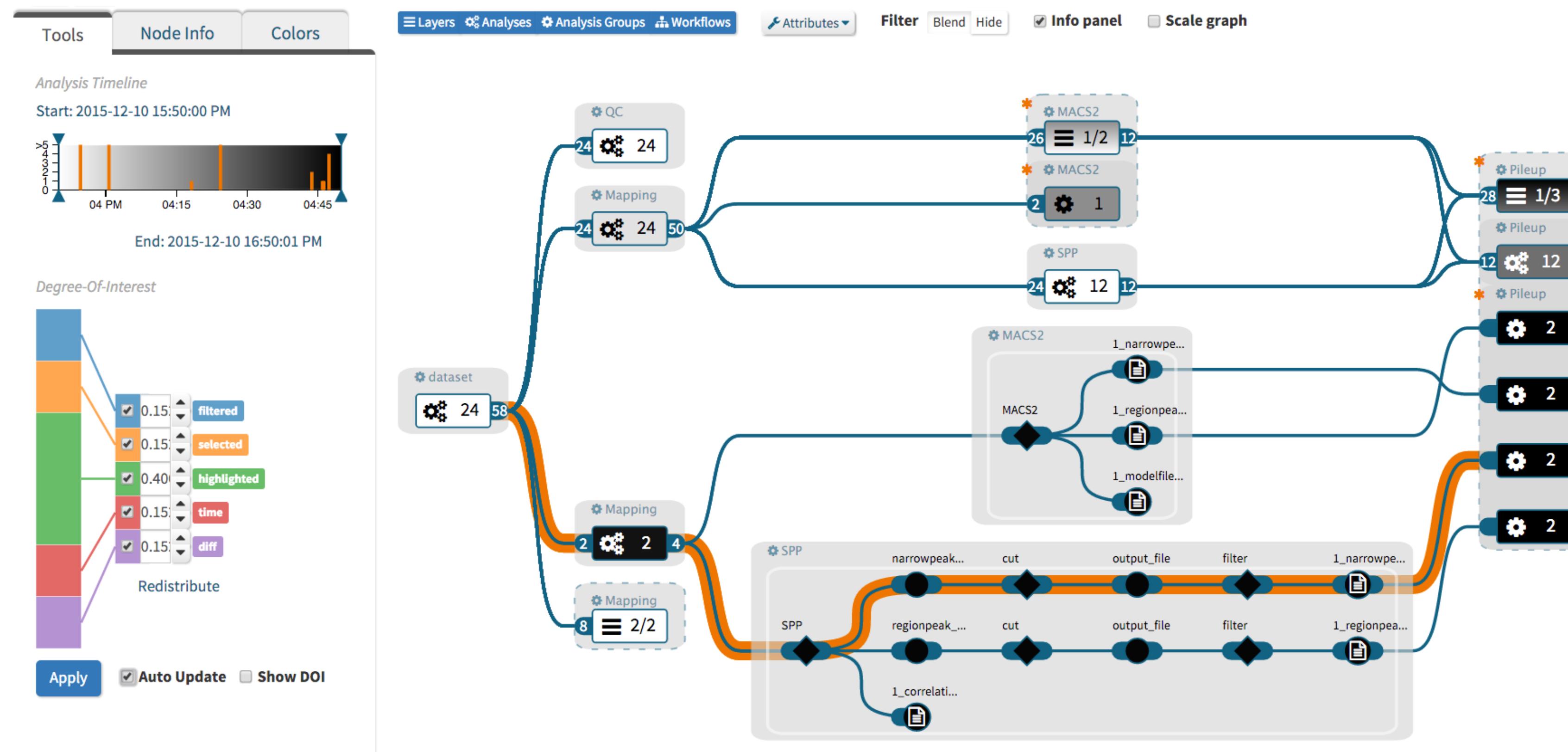
Tools for Reproducible Research

The screenshot shows the Stem Cell Commons Refinery Platform interface. The top navigation bar includes links for Stem Cell Commons, Collaboration, Statistics, About, and user Nils Gehlenborg. The main content area is divided into three sections: Data Sets, Analyses, and Workflows.

- Data Sets:** Displays 206 data sets. Examples include:
 - An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming [RNA-Seq]
 - Droplet barcoding for single cell transcriptomics applied to embryonic stem cells
 - NKX2-1 occupancy in human lung adenocarcinoma cell lines
 - Transcription factor ChIP-seq in expanded human hematopoietic stem and progenitor cells
 - Cardiac transcription factors in HL-1 cells: genome binding profiling
 - Gene expression analysis of cdx4, sall4, and cdx4+sall4 morpholino(s) injected embryos at 3-somite stage
 - Epigenetic profiling of WT and Ezh2-null MLL-AF9 murine leukemic cells
 - Genome wide uH2A localization analysis highlights Bmi1-dependent deposition of the mark at repressed genes
 - Cdx2 transcription factor binding in intestinal villus and gene expression profiling in Cdx mutant mice
 - Cell-Type-Specific TGF-beta Signaling is Targeted to Genes that Control Cell Identity: ChIP-Seq (mouse)
 - Mapping polycomb complexes in human and mouse embryonic stem cells (mouse)
 - Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and factor binding (ChIP-seq)
- Analyses:** Displays 60 analyses. Examples include:
 - Test workflow: 5 steps no delay 2016-1-15@9:39:07
 - FastQC 2016-1-15@9:17:49
 - FastQC 2016-1-14@11:33:31
 - FastQC 2016-1-13@21:10:53
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-23@10:53:37 - 3 pairs
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-22@17:39:12
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-18@14:20:51 - 3 pairs
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-18@12:29:15 - 13880/3 pairs
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-17@15:50:05
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-16@17:11:52 - new bowtie
 - TF ChIP-Seq analysis using MACS2: danRer7 2015-9-16@13:01:33
 - FastQC 2015-9-16@12:39:55
 - FastQC 2015-9-16@12:20:18
 - FastQC 2015-9-16@11:06:31
 - MACS2 Demo
 - Oct4 FastQC
 - My release test!
- Workflows:** Displays 5 workflows. Examples include:
 - FastQC
 - TF ChIP-Seq analysis using MACS2: danRer7
 - TF ChIP-Seq analysis using MACS2: hg19
 - TF ChIP-Seq analysis using MACS2: mm10
 - Test workflow: 5 steps no delay

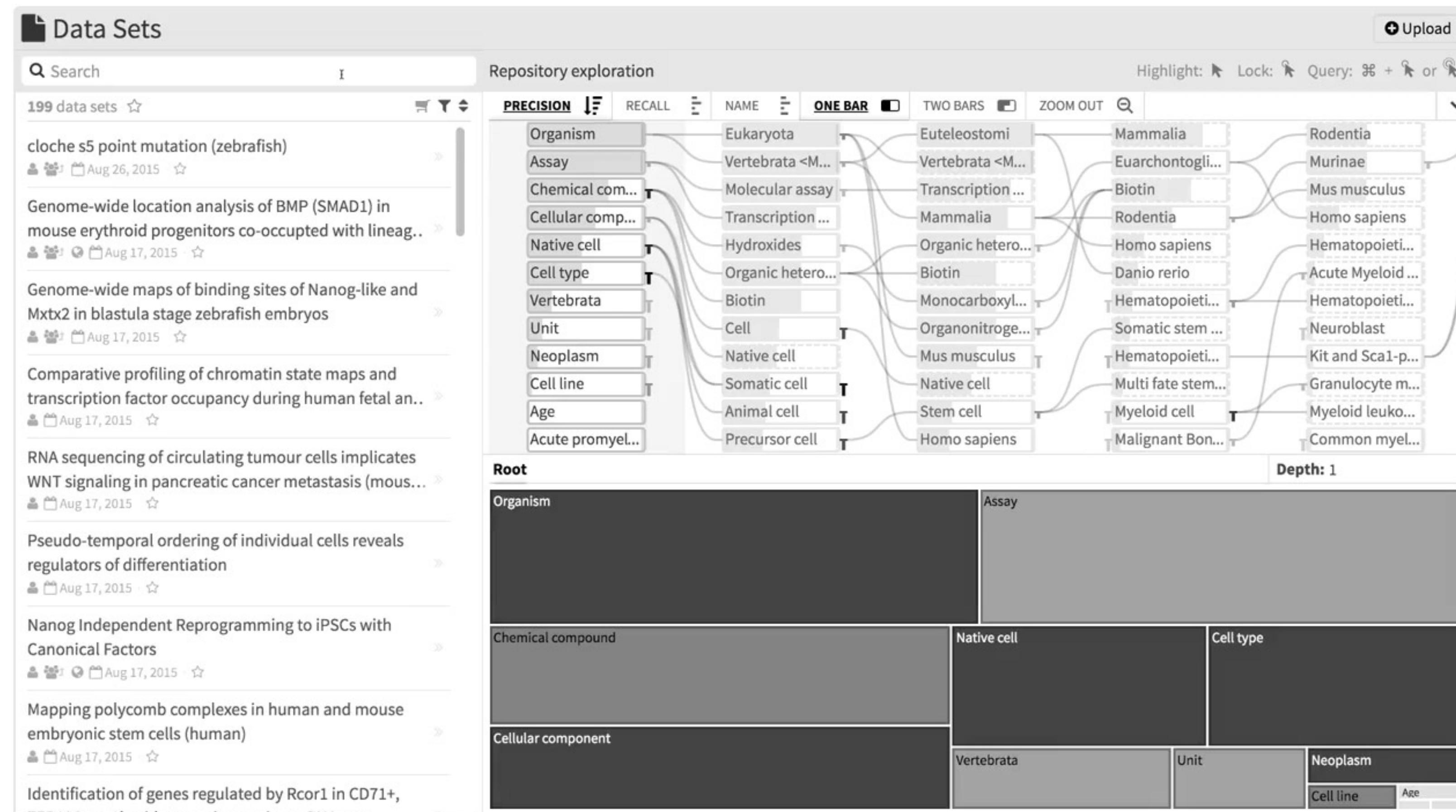
Refinery Platform: <http://refinery-platform.org>

Tools for Reproducible Research



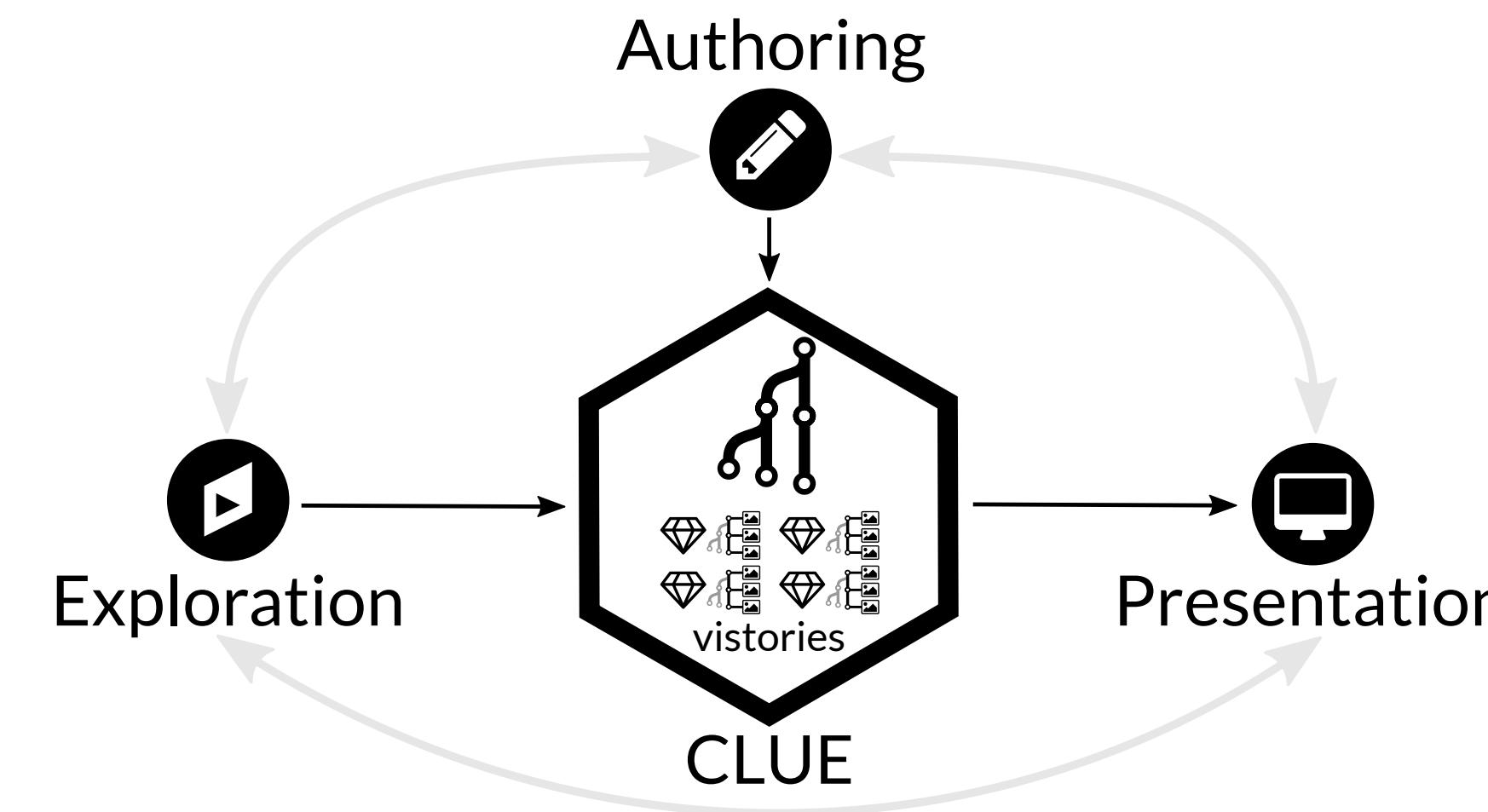
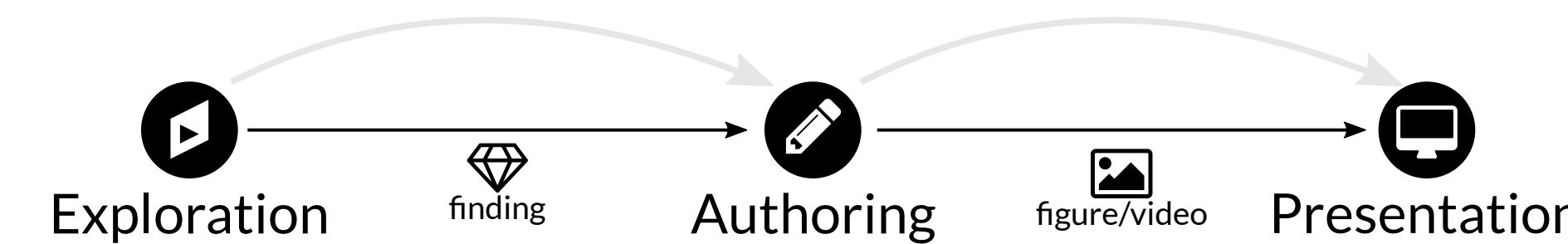
AVOCADO Provenance Visualization
Refinery Platform: <http://refinery-platform.org>

Tools for Reproducible Research



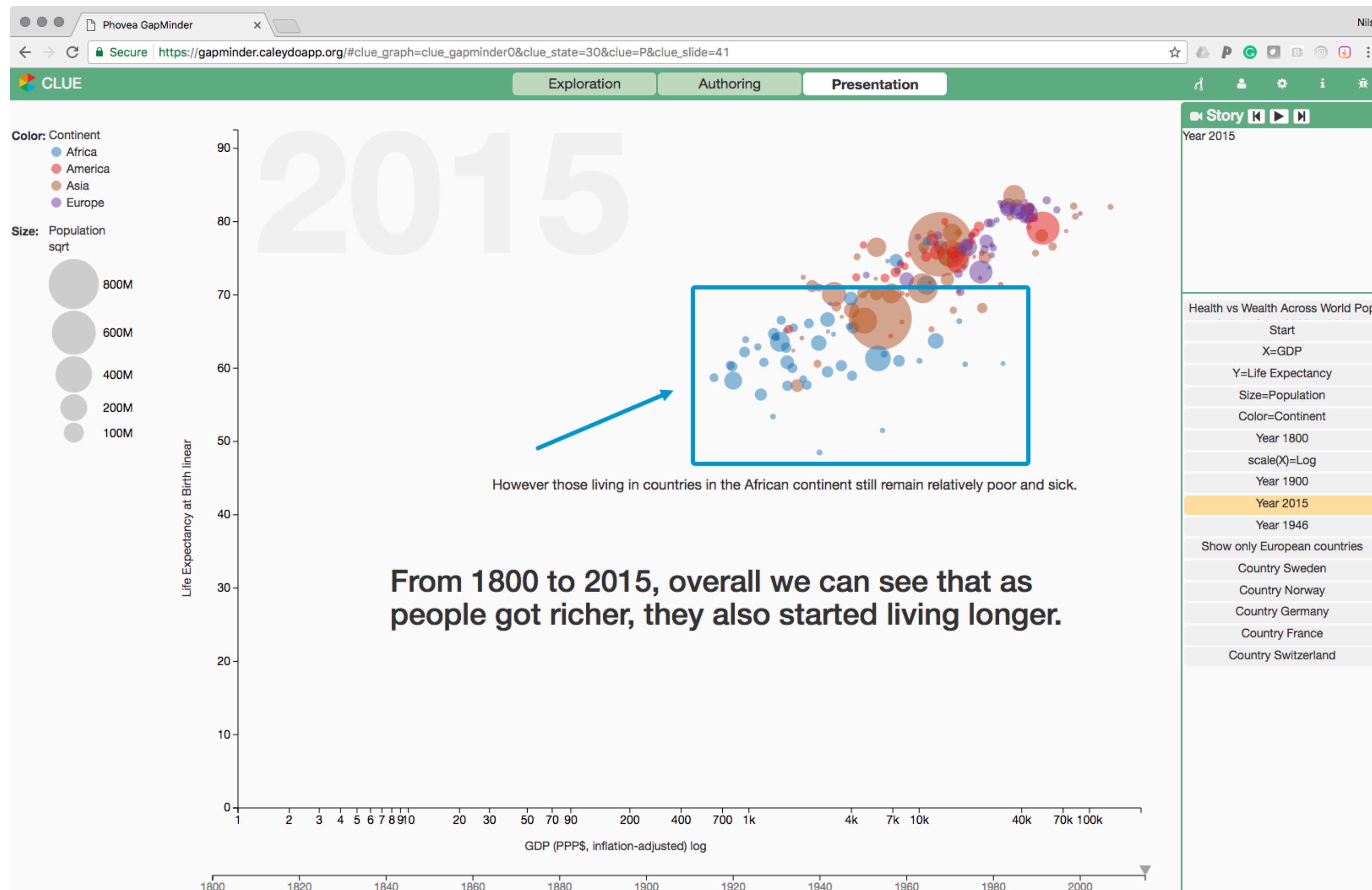
SATORI Ontology-Guided Repository Exploration
Refinery Platform: <http://satori.refinery-platform.org>

Tools for Reproducible Research



Vistories & CLUE: <http://vistories.org>

Tools for Reproducible Research



Vistories in GapMinder: <http://vistories.org/v/gapminder>

How does it work?

Visualization is really about external cognition, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind.

— Stu Card

How does it work?

Why do we use the visual system and not other sensory systems?

Information bandwidth of visual system is much higher than of all other sensory system.

How does it work?

Visualization uses perception to free up cognition.

How does visualization work?

MALWMRLLPLALLALWGPDPA
AAFVNQHLCGSHLVEALYLVCG
ERGFFYTPKTRREAEDLQVGQV
ELGGPGAGSLQPLALE GSLQK
RGIVEQCCTSICSLYQLENYCN

How does visualization work?

MALWMRLLPLALLALWGPDPA
AAFVNQHLCGSHLVEALYLVCG
ERGFFYTPKTRREAEDLQVGQV
ELGGGPGAASLQPLALEGSLQK
RGIVEQCCTSICSLYQLENYCN

How does visualization work?

Visualization uses perception to free up cognition.

Visualization is an external cognitive aid
and augments working memory.

How does visualization work?

$$453 \times 862 = ?$$

How does visualization work?

$$\underline{453 \times 862 = ?}$$

906

+ 27,180

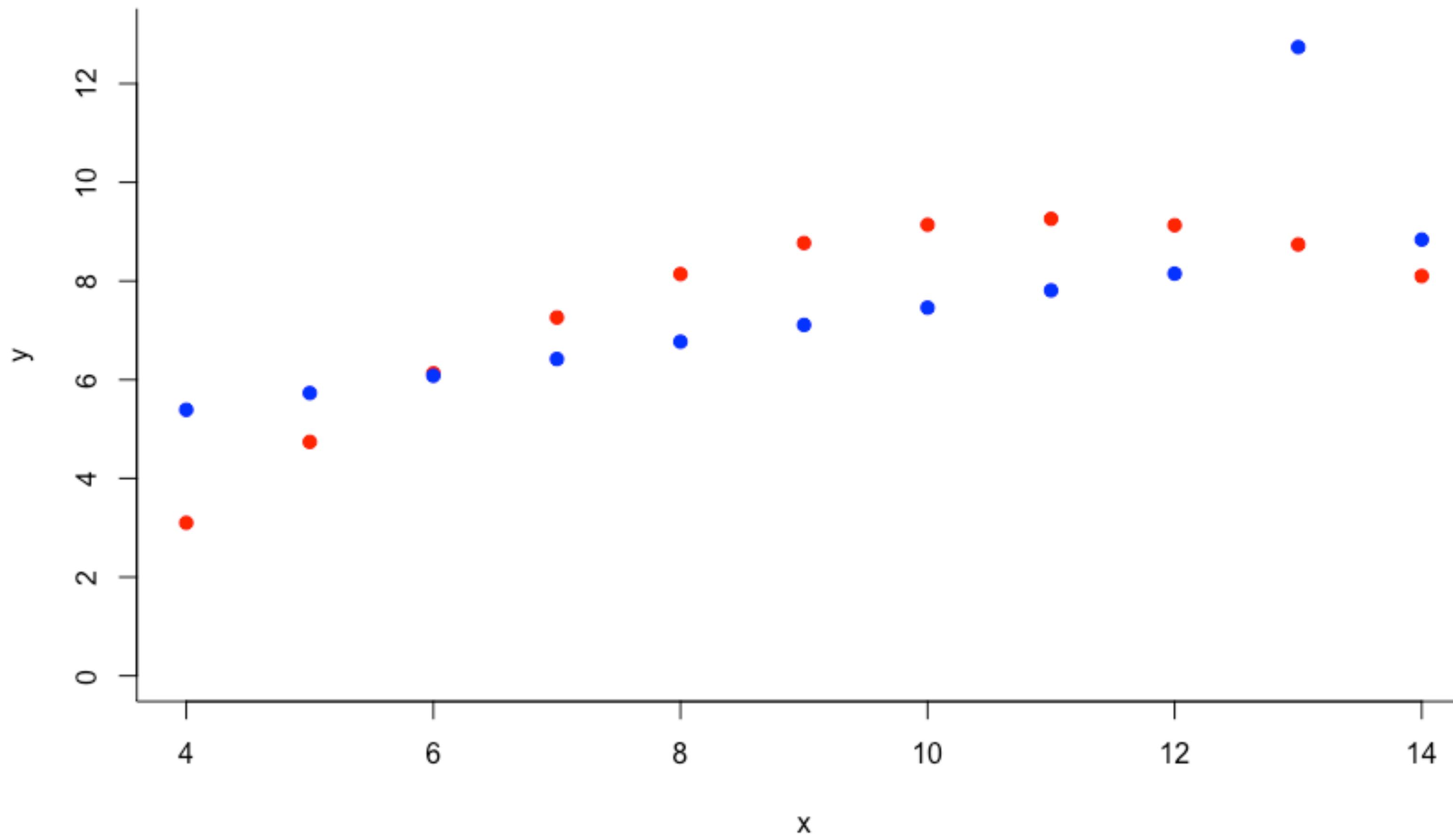
+ 362,400

390,486

How does visualization work?

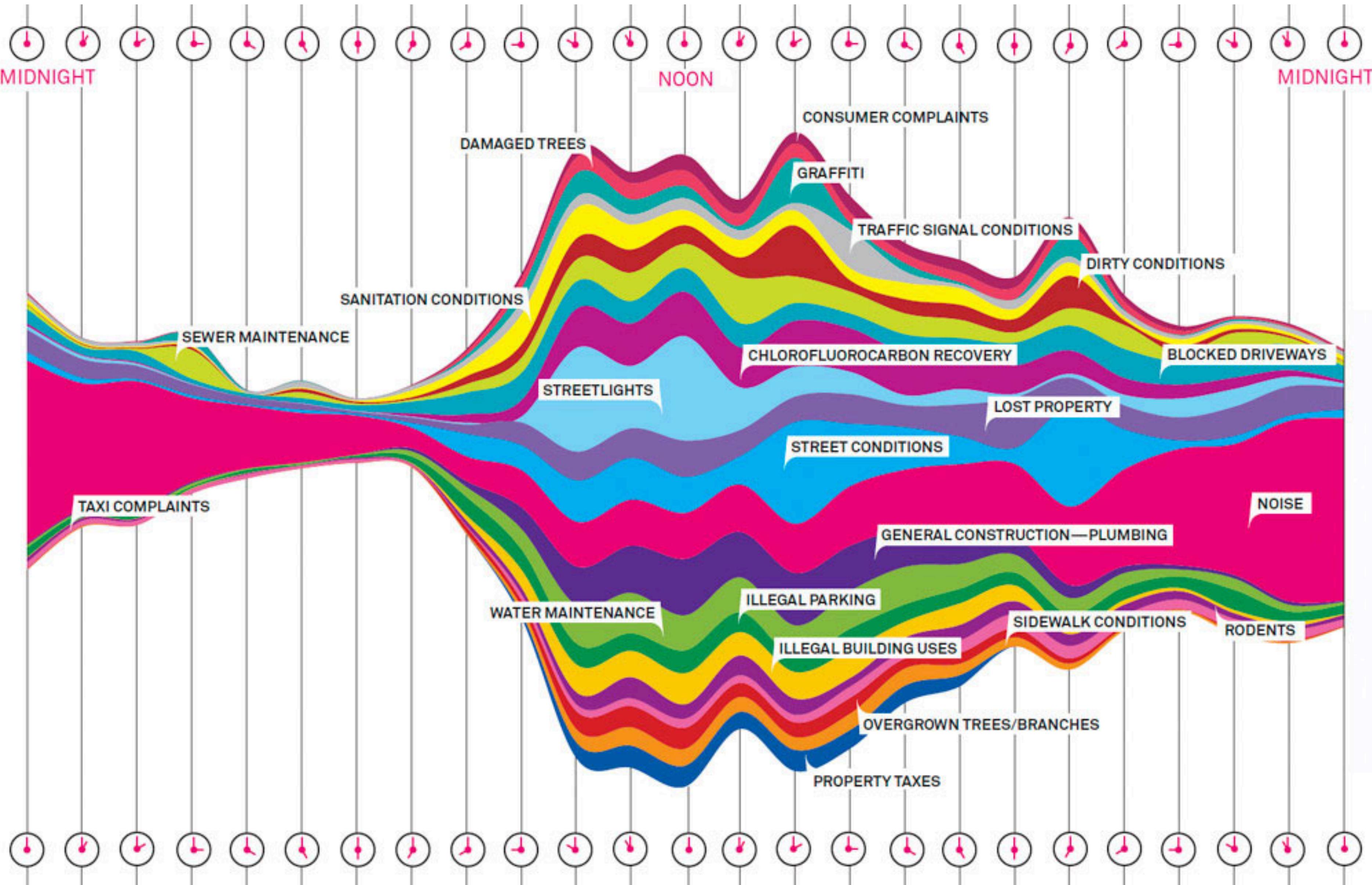
	f_1		f_2		Tasks
	x	y	x	y	
1	10	9.14	10	7.46	What is the shape of f_1 ? Of f_2 ?
2	8	8.14	8	6.77	
3	13	8.74	13	12.74	How many times do f_1 and f_2 intersect?
4	9	8.77	9	7.11	
5	11	9.26	11	7.81	
6	14	8.10	14	8.84	Do they cross 0?
7	6	6.13	6	6.08	
8	4	3.10	4	5.39	...
9	12	9.13	12	8.15	
10	7	7.26	7	6.42	
11	5	4.74	5	5.73	

How does visualization work?



Design Critique

Wired: 34,522 311 calls in New York City between 9/8/10 and 9/15/10



Visual Encoding of Data

data

apples
oranges
bananas

small
medium
large

10 inches
13 inches
18.5 inches

trees
networks

intrinsic
position

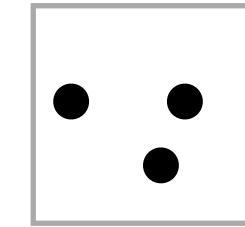
Visual Channels: Rankings

Categorical
What? Where?

Relational
With whom?



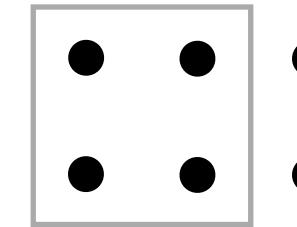
position*
planar



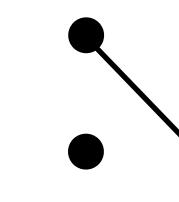
color hue
shape



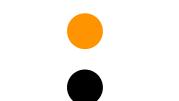
containment



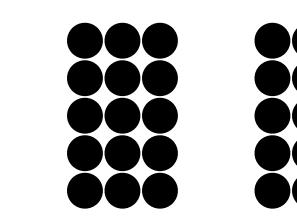
connection



similarity



position*
proximity

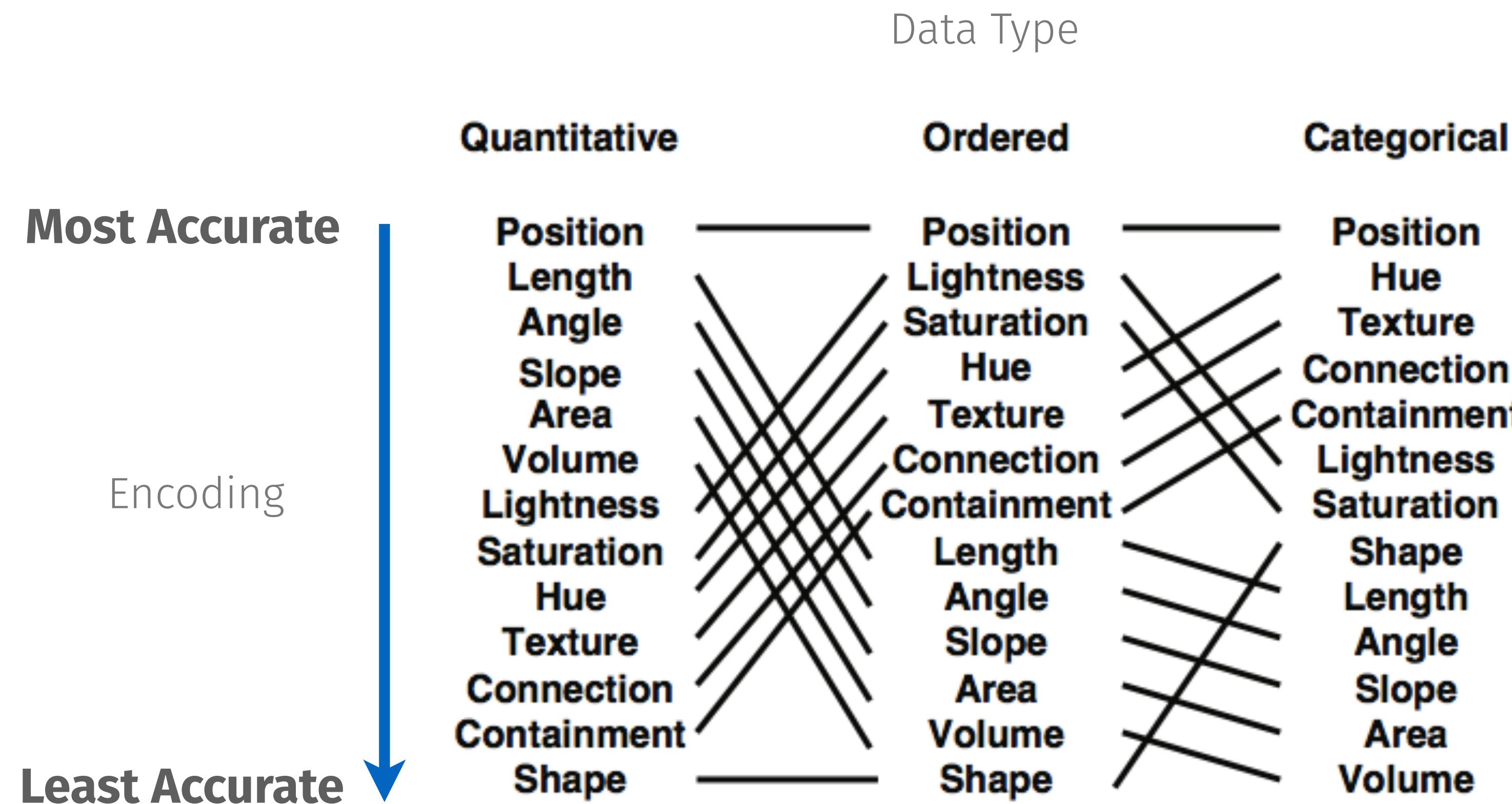


Visual Channels: Rankings

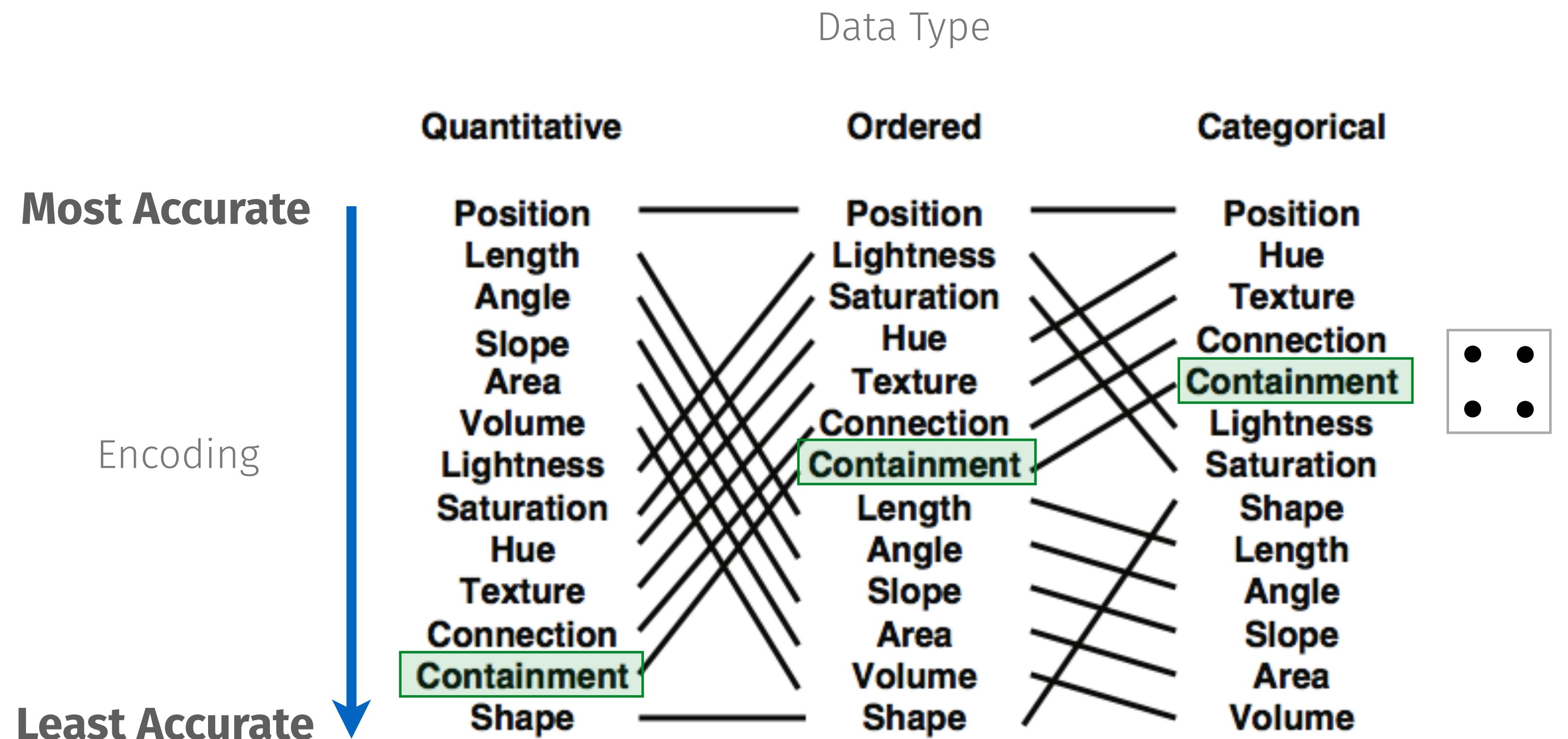
Ordinal &
Quantitative
How much?

position*	common scale	
position*	unaligned scale	
length (1D)		
angle/tilt		
area (2D)		
curvature		
volume (3D)		
lightness		
color saturation		

Ranking of Encodings

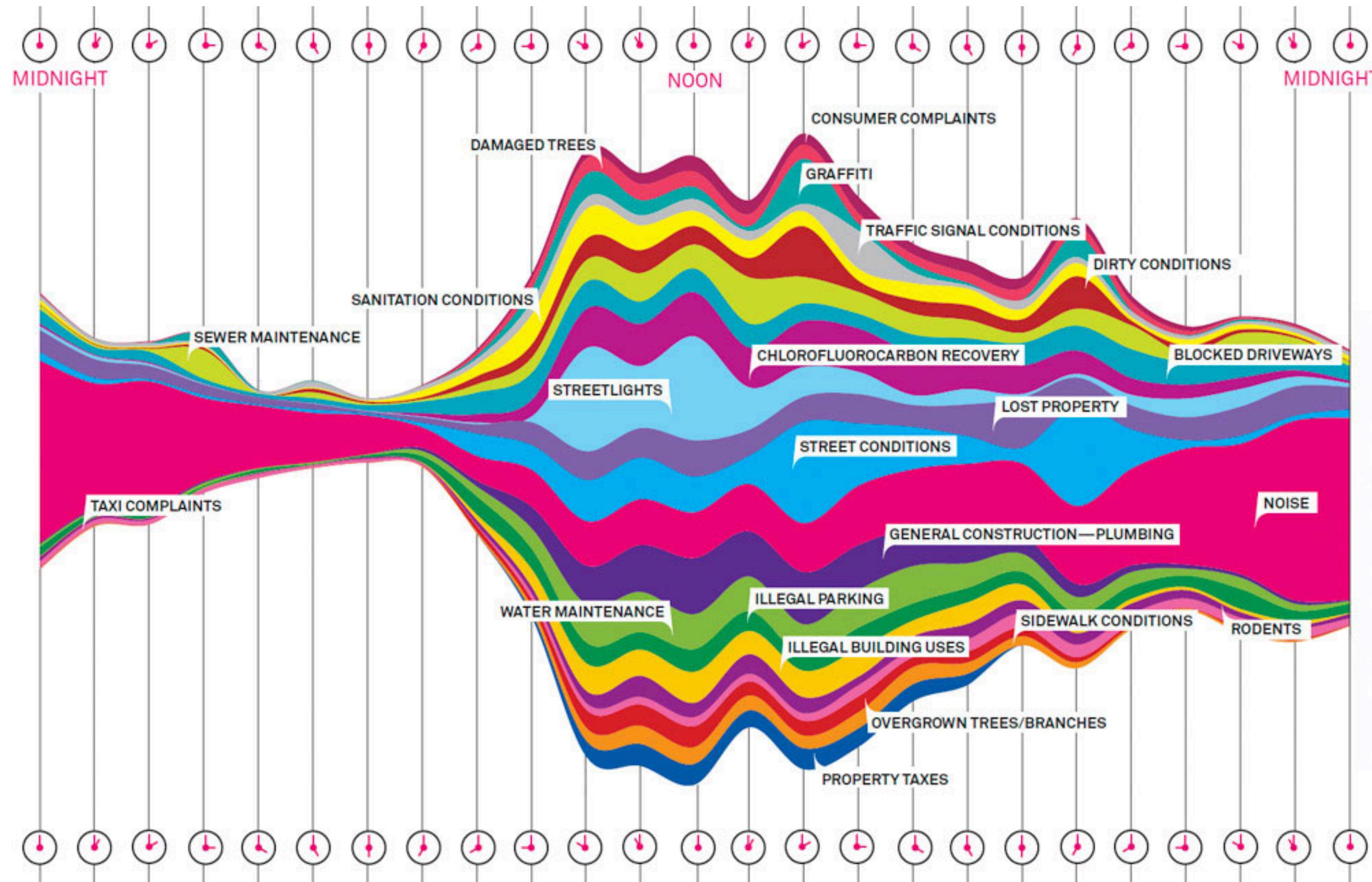


Ranking of Encodings



Back to our Design Critique ...

Exercise: Design Critique



Ranking of Encodings

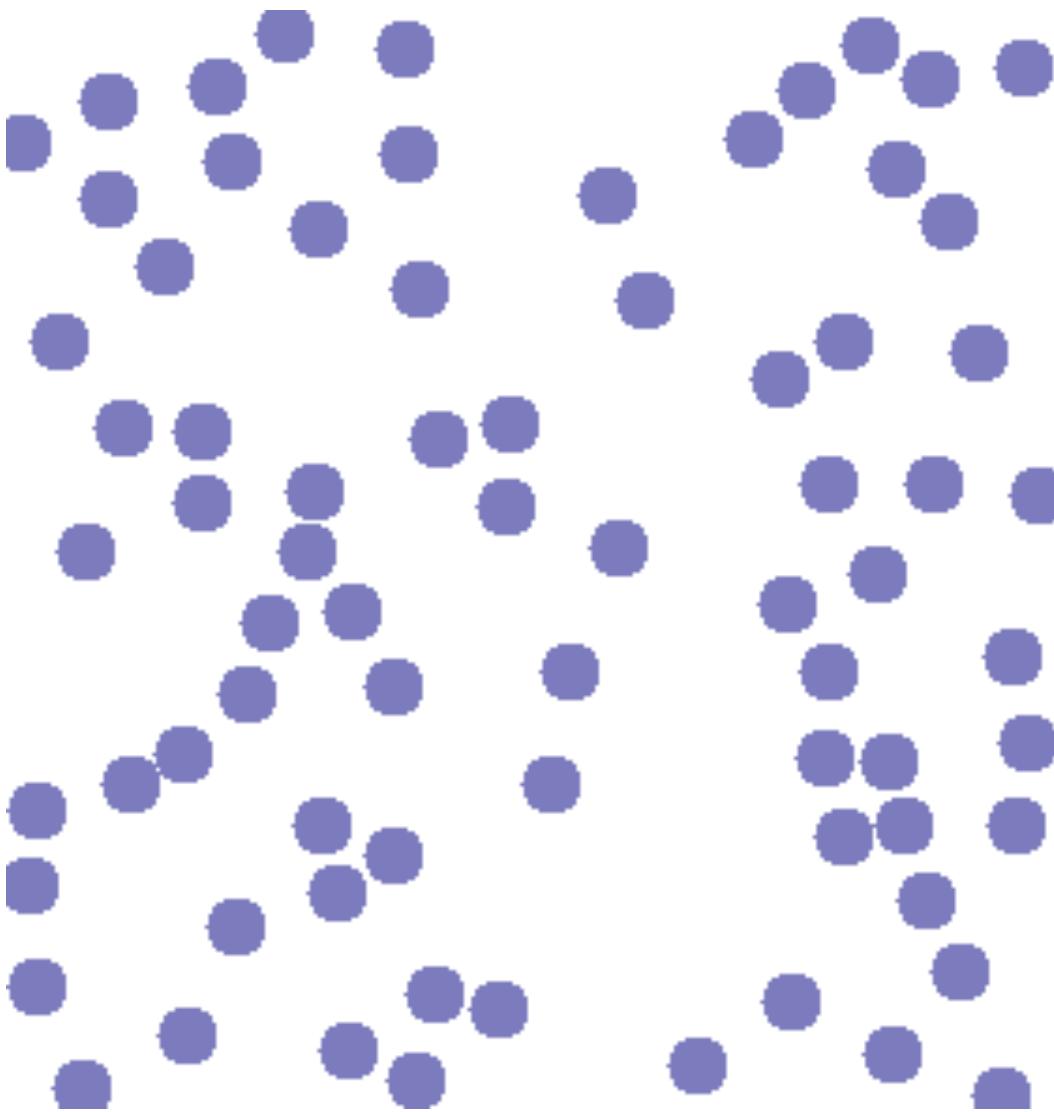
Principle of Importance Ordering (Mackinlay 1986):

Encode more important information more effectively.

- How accurately can the data be read from the visualization?
- How many classes can be distinguished?
- Can the channels be separated from each other?
- Which channels are processed preattentively?

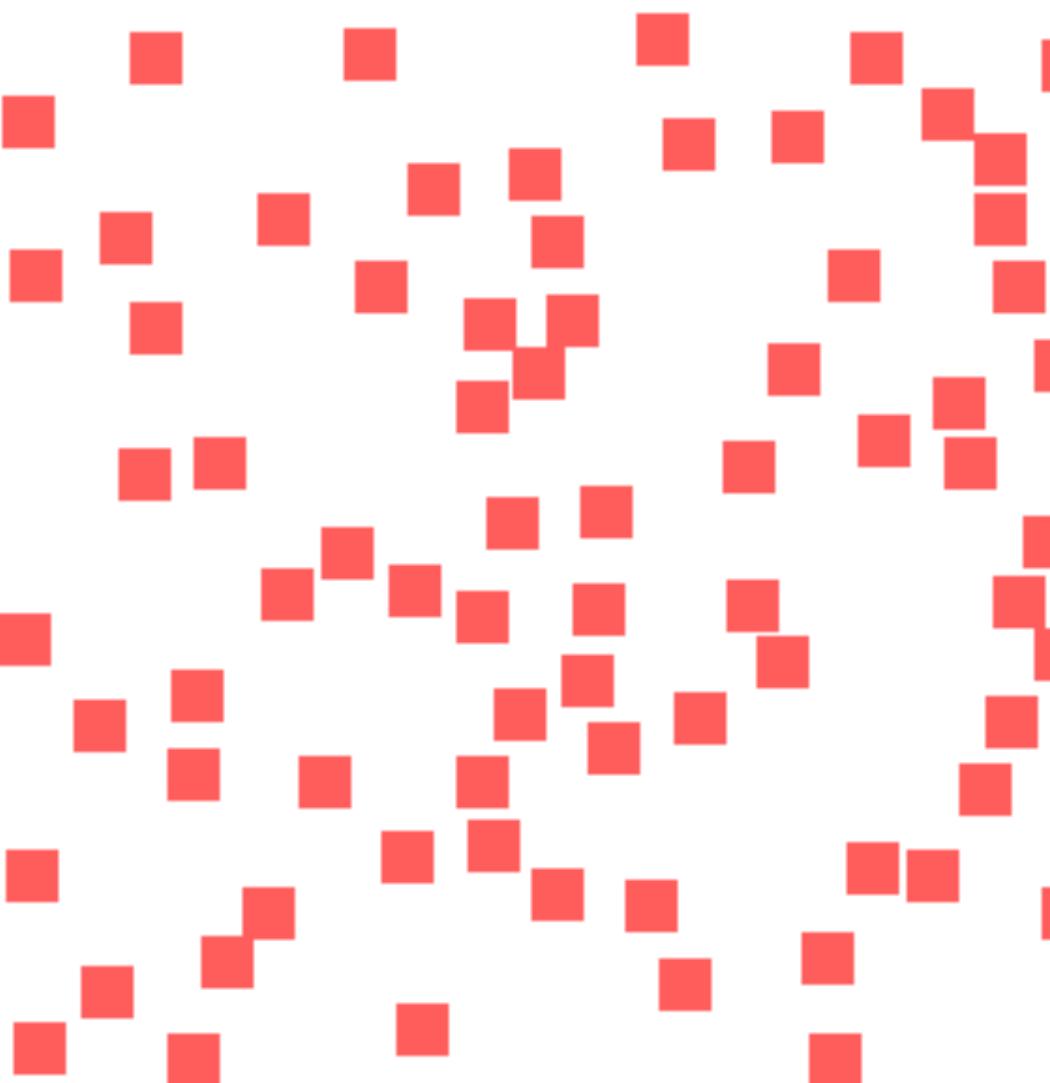
Preattentive Processing: Color

Can you spot this: ?



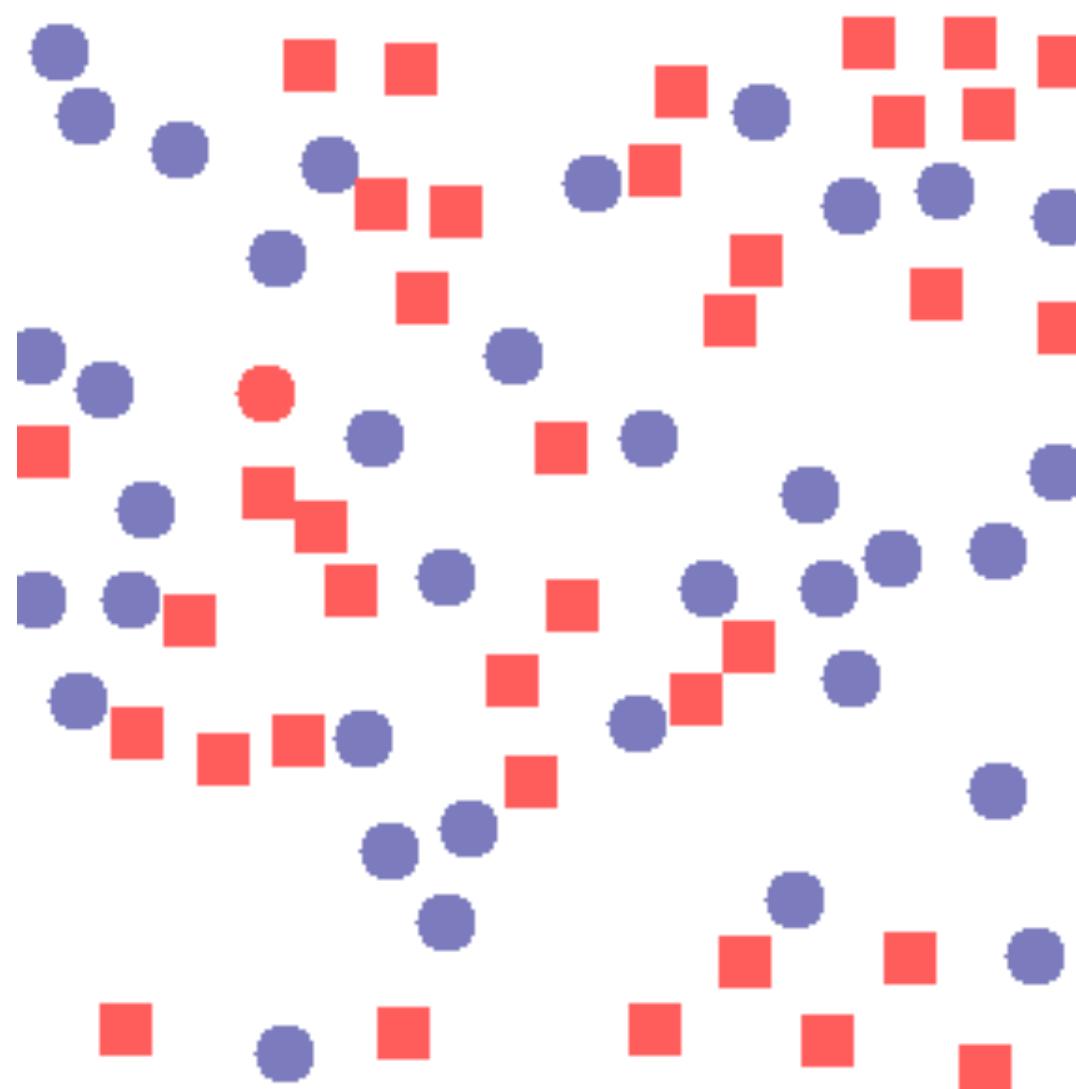
Preattentive Processing: Shape

Can you spot this: 

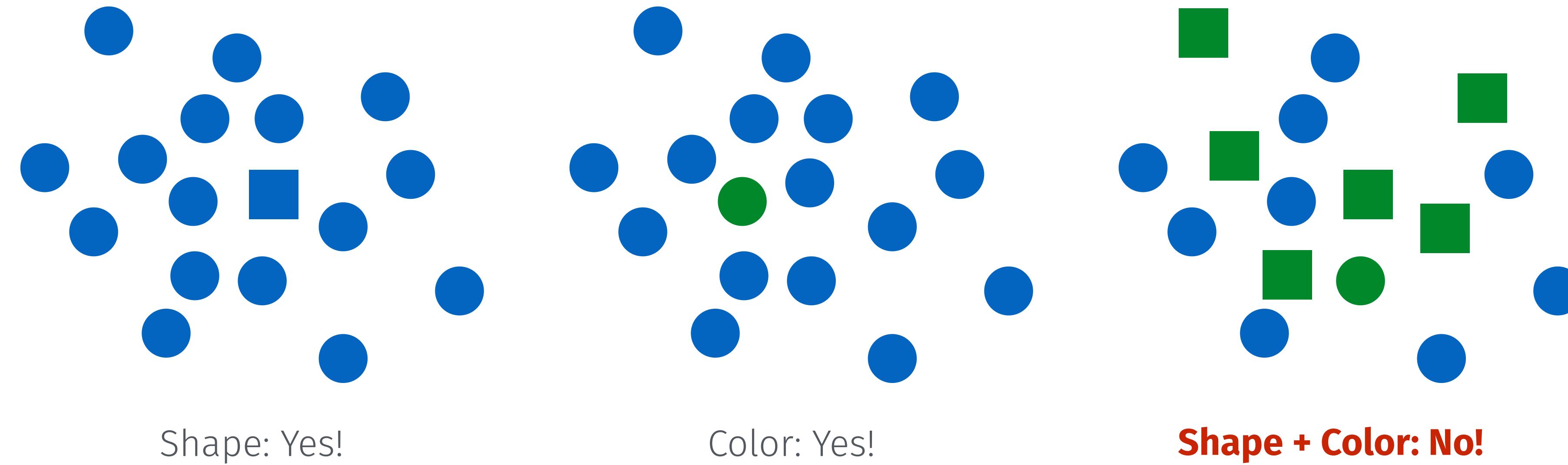


Preattentive Processing: Shape & Color

Can you spot this: ?

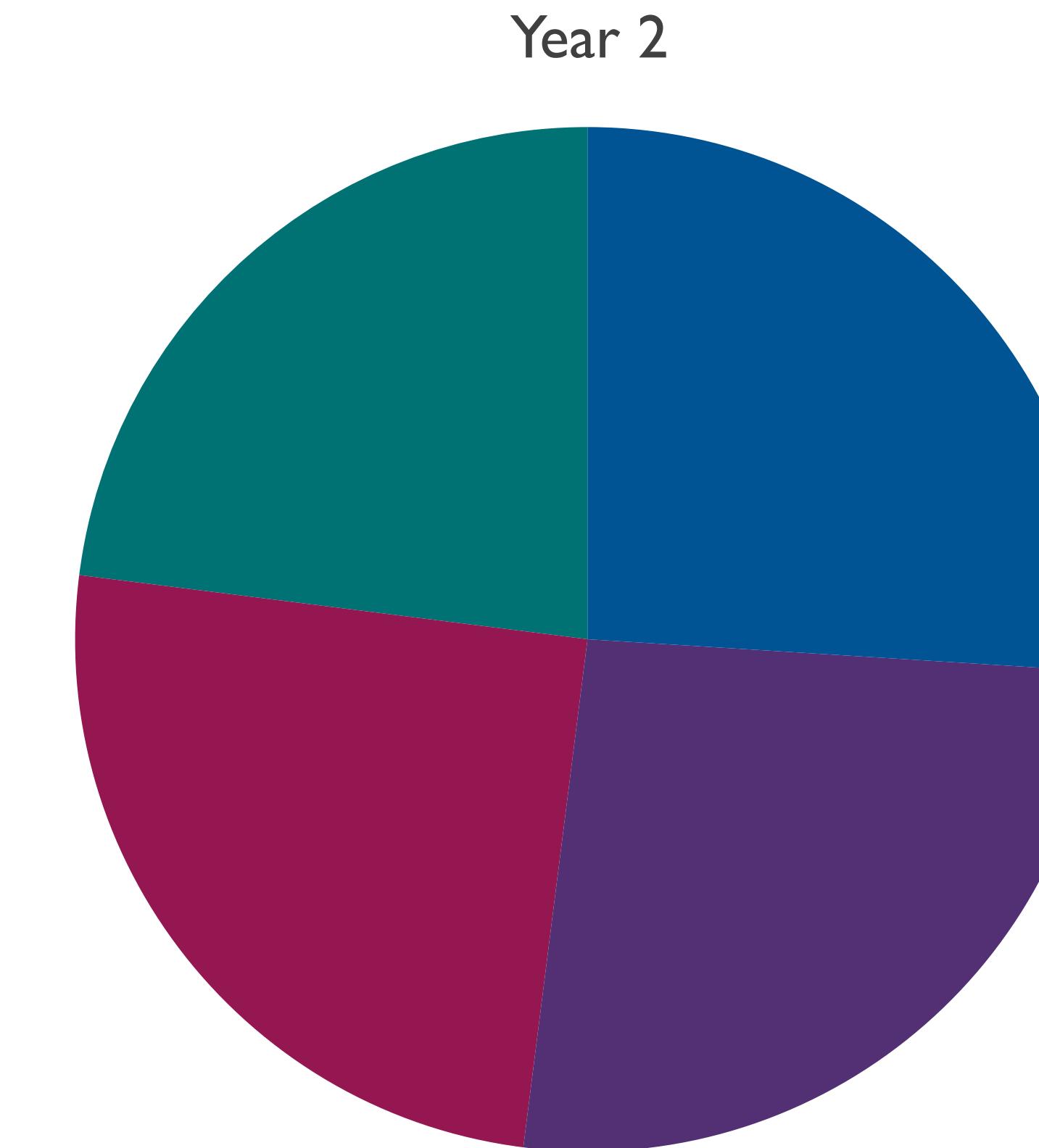
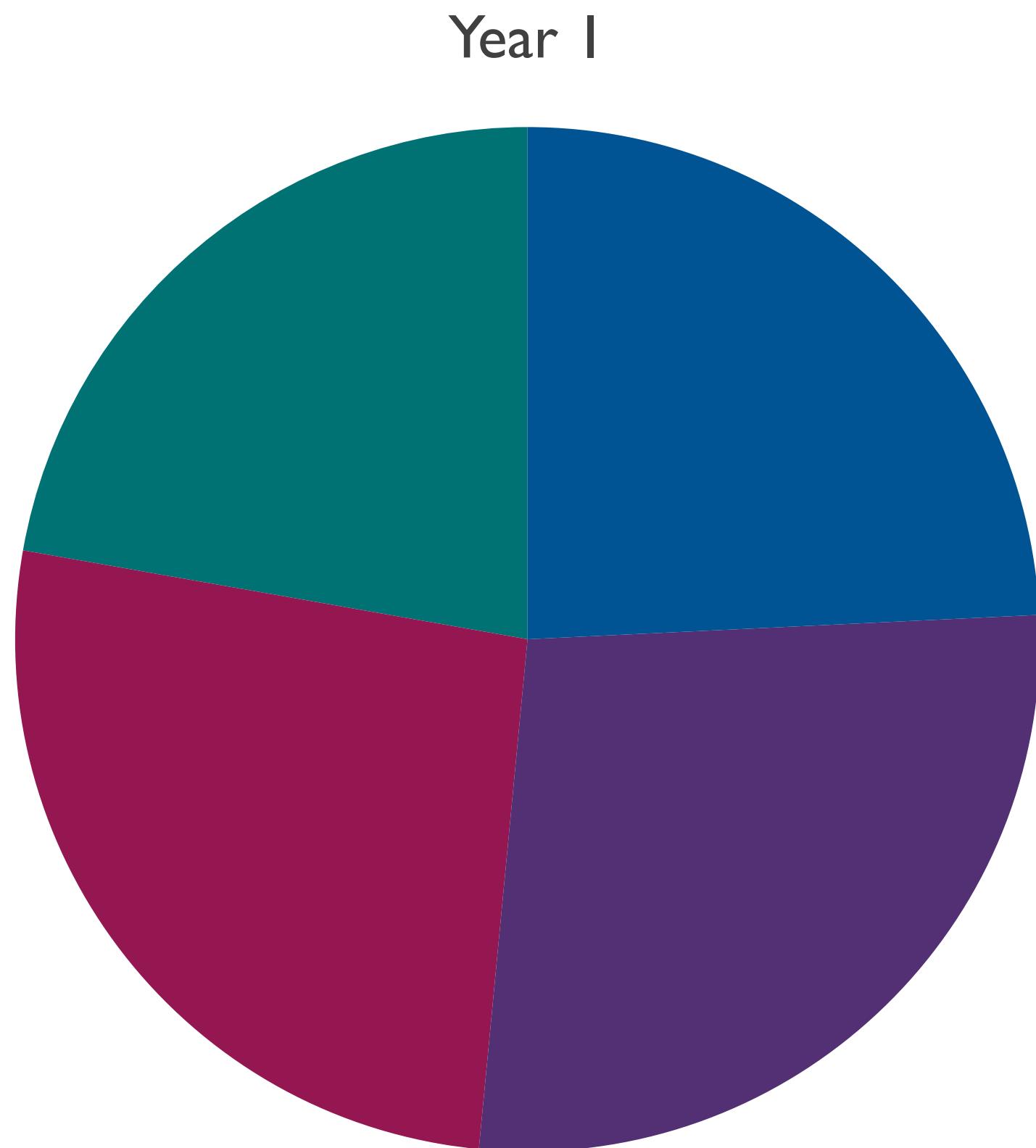


Preattentive Processing



- visual properties that can be perceived in less than 250 ms
- no sequential scanning of the image required, unlike text or numbers
- examples for other visual properties that can be processed preattentively: orientation, curvature, direction of motion, size and others

Using Rankings



A

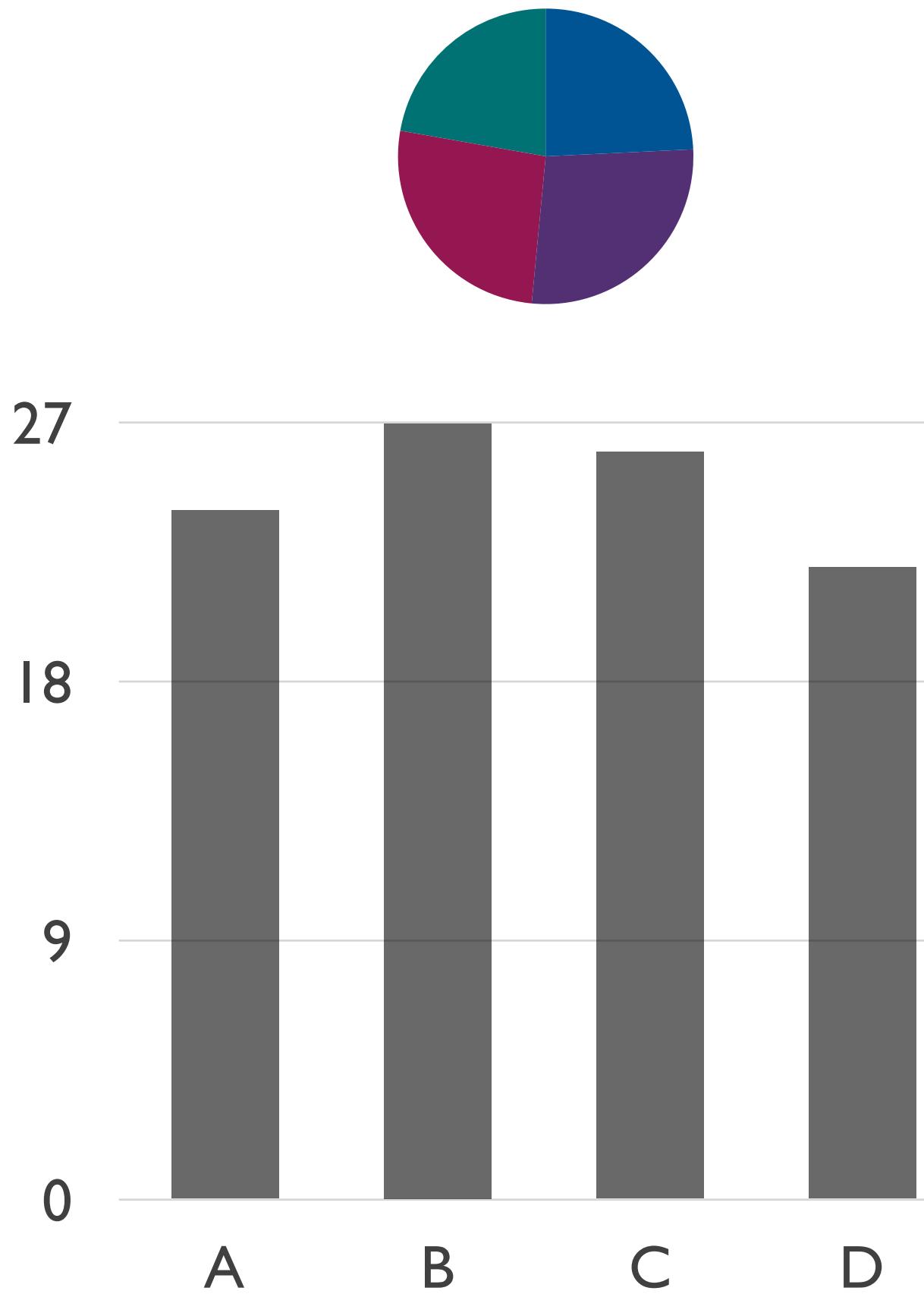
B

C

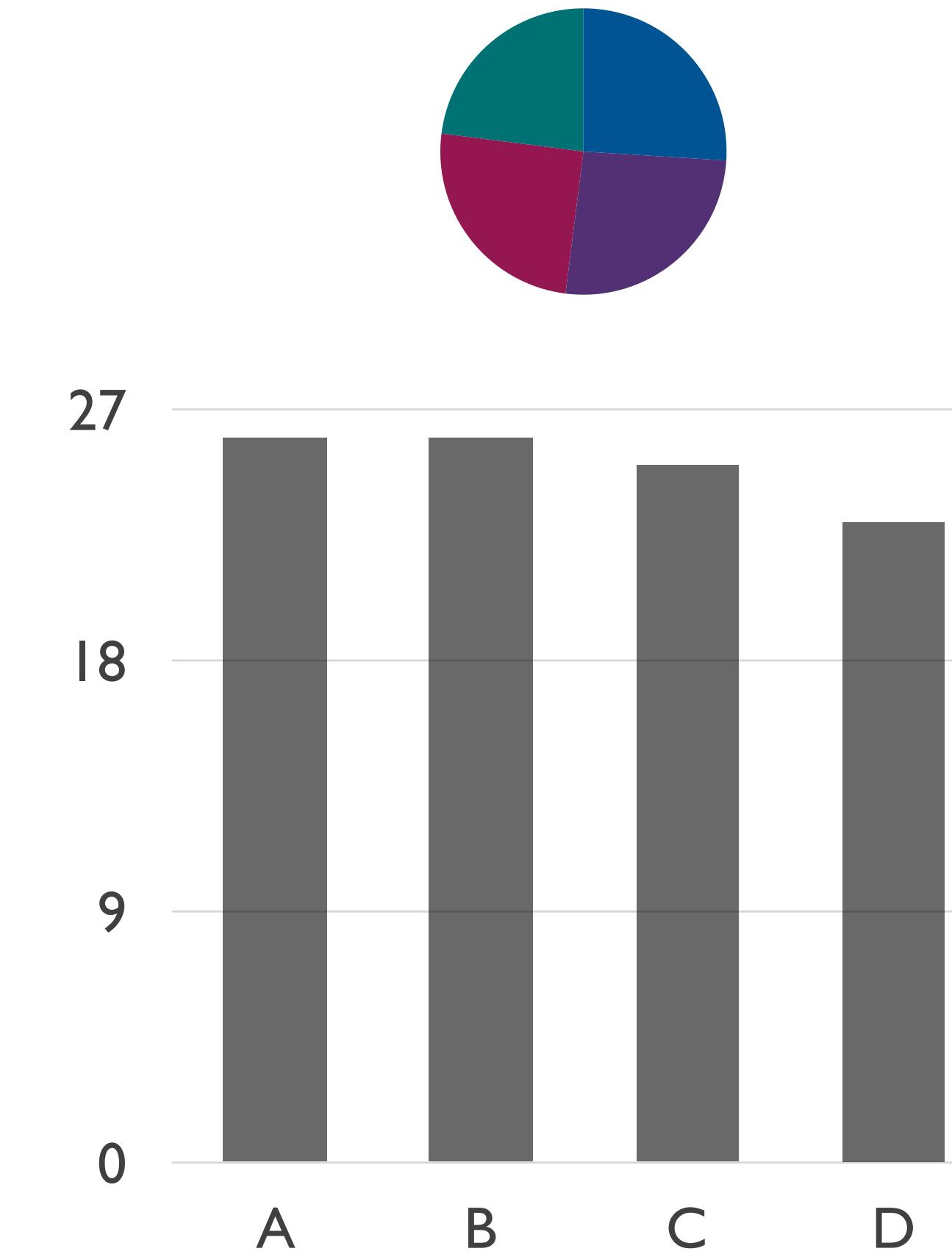
D

Using Rankings

Year 1



Year 2



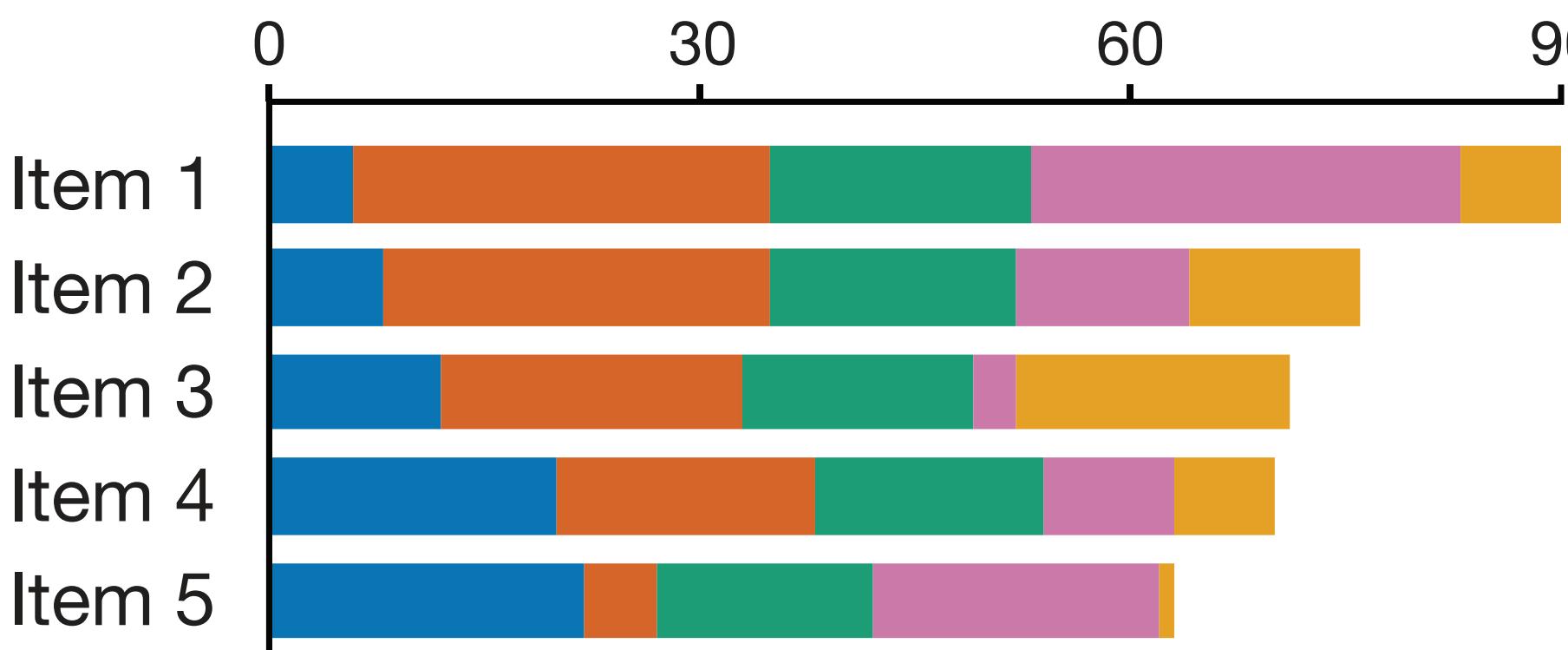
Bar Charts for Items & Categories

	1	2	3	4	5
Item 1	6	29	18	30	7
Item 2	8	27	17	12	12
Item 3	12	21	16	3	19
Item 4	20	18	16	9	7
Item 5	22	5	15	20	1

Bar Charts for Items & Categories

	1	2	3	4	5
Item 1	6	29	18	30	7
Item 2	8	27	17	12	12
Item 3	12	21	16	3	19
Item 4	20	18	16	9	7
Item 5	22	5	15	20	1

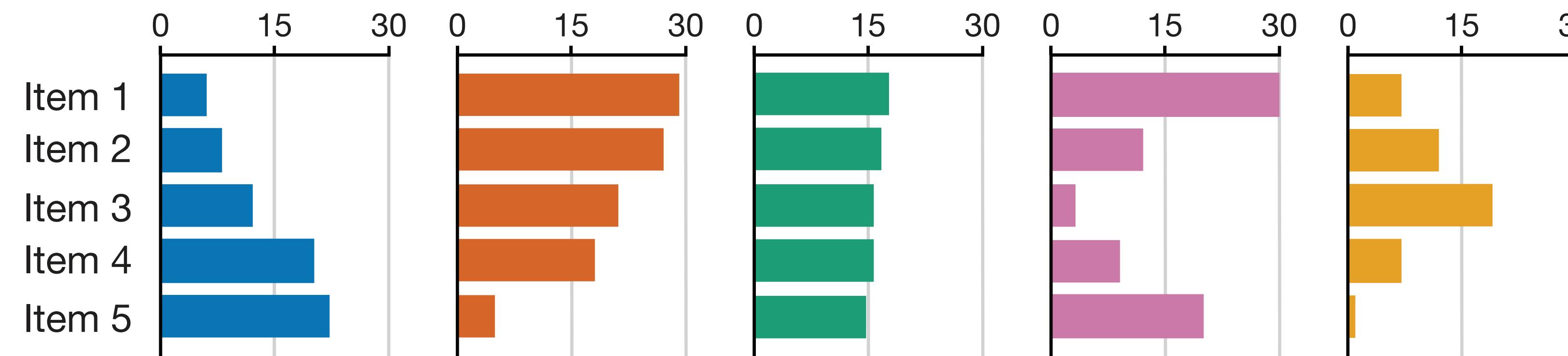
Stacked Bar Chart



Bar Charts for Items & Categories

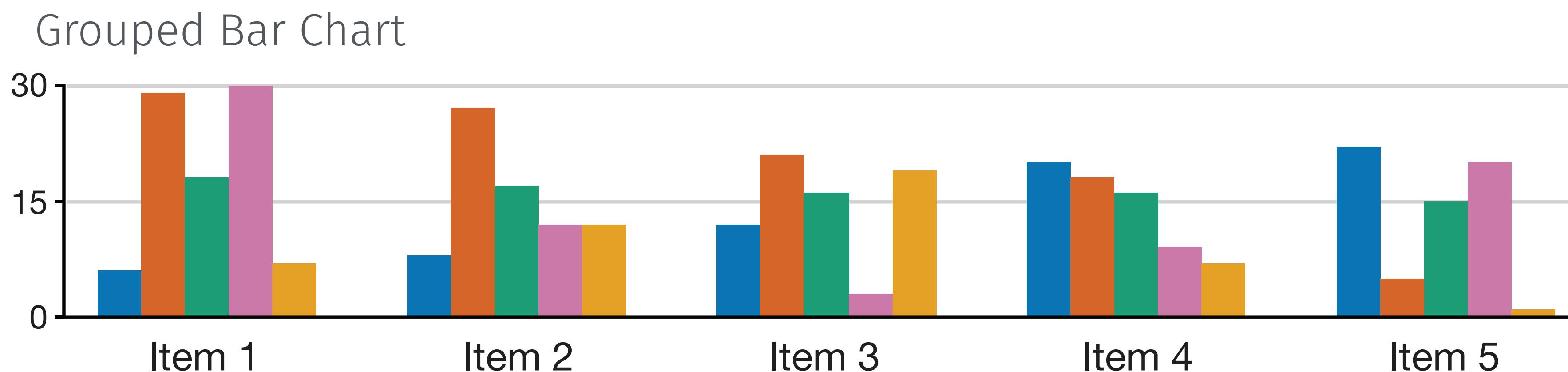
	1	2	3	4	5
Item 1	6	29	18	30	7
Item 2	8	27	17	12	12
Item 3	12	21	16	3	19
Item 4	20	18	16	9	7
Item 5	22	5	15	20	1

Layered Bar Chart



Bar Charts for Items & Categories

	1	2	3	4	5
Item 1	6	29	18	30	7
Item 2	8	27	17	12	12
Item 3	12	21	16	3	19
Item 4	20	18	16	9	7
Item 5	22	5	15	20	1



Bar Charts for Items & Categories

- **Stacked Bar Chart**

- if focus is on comparing the overall quantities across items but also need to illustrate contributions of each category to the total

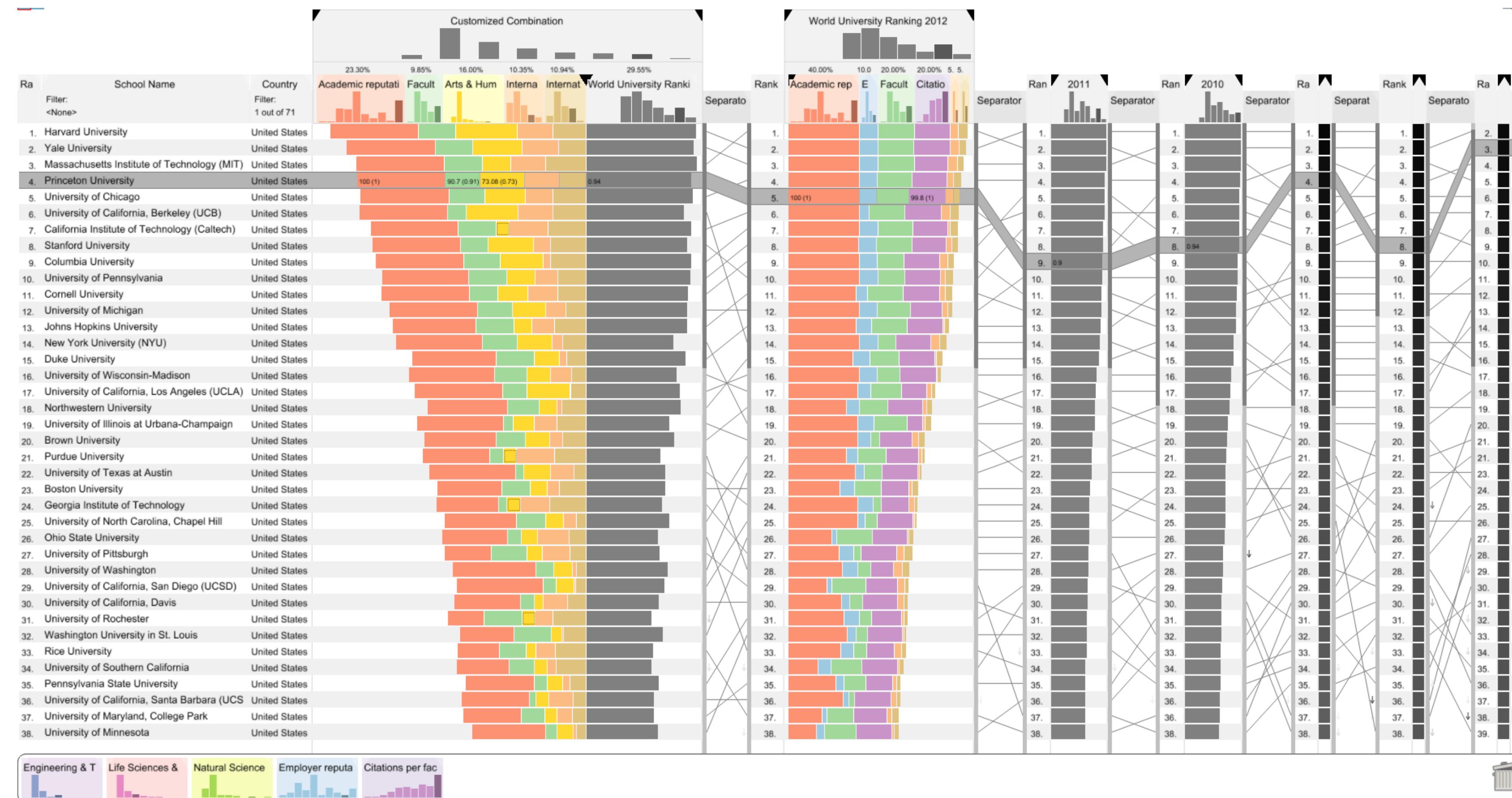
- **Layered Bar Chart**

- if focus is on distribution of values in each category across all items
- comparisons within each category are more accurate than in stacked bar charts due to common baseline for the values in each category

- **Grouped Bar Chart**

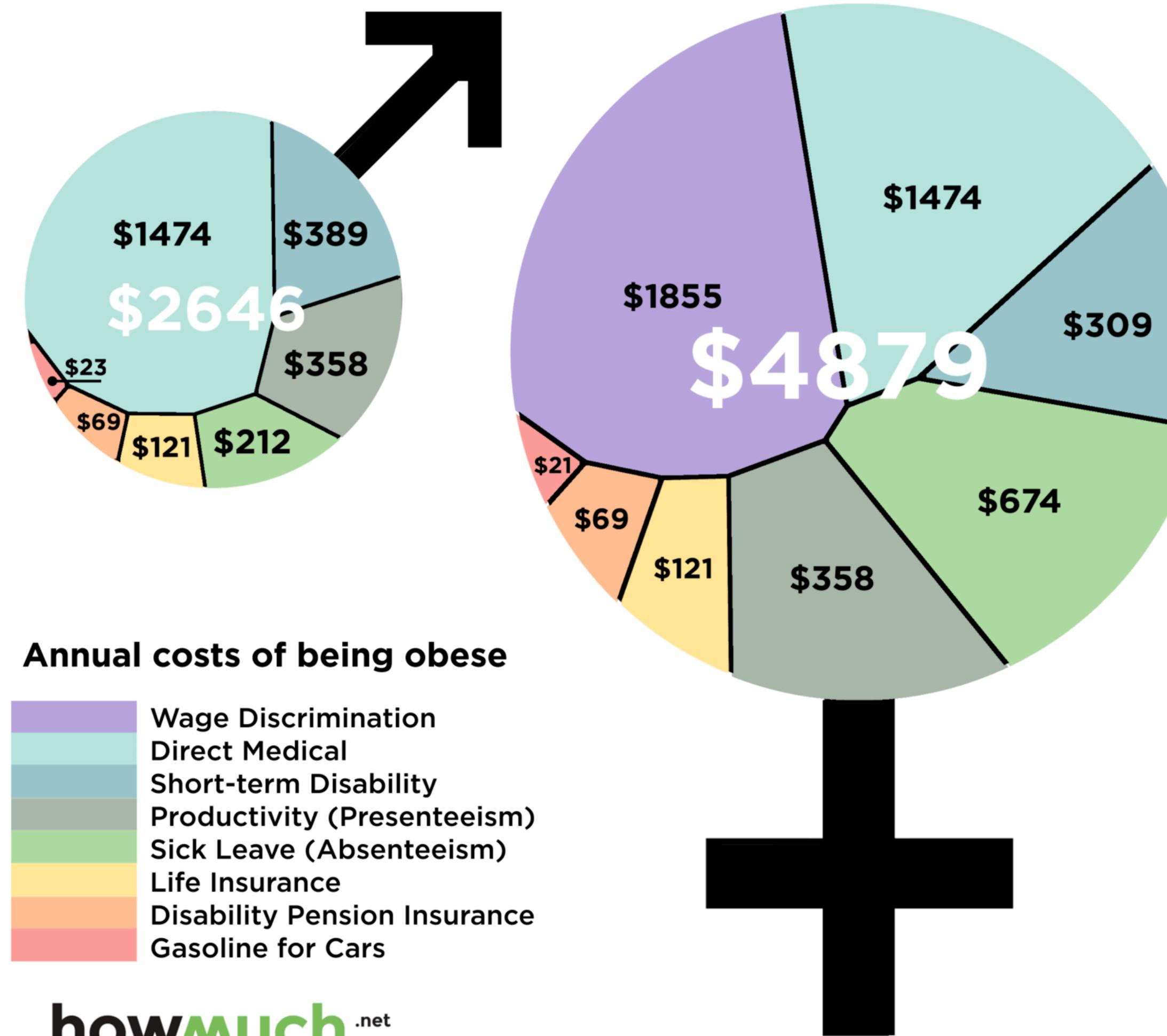
- if focus is on comparison of values across categories within each item while still enabling comparisons across items
- if quantities add up to the same total for each item, then a grouped bar chart is equivalent to multiple pie charts, yet a grouped bar chart affords more accurate readings of values and comparisons

LineUp: Ranking Visualization



Redesign Exercise

The Individual Costs of Being Obese in the United States



Redesign Exercise

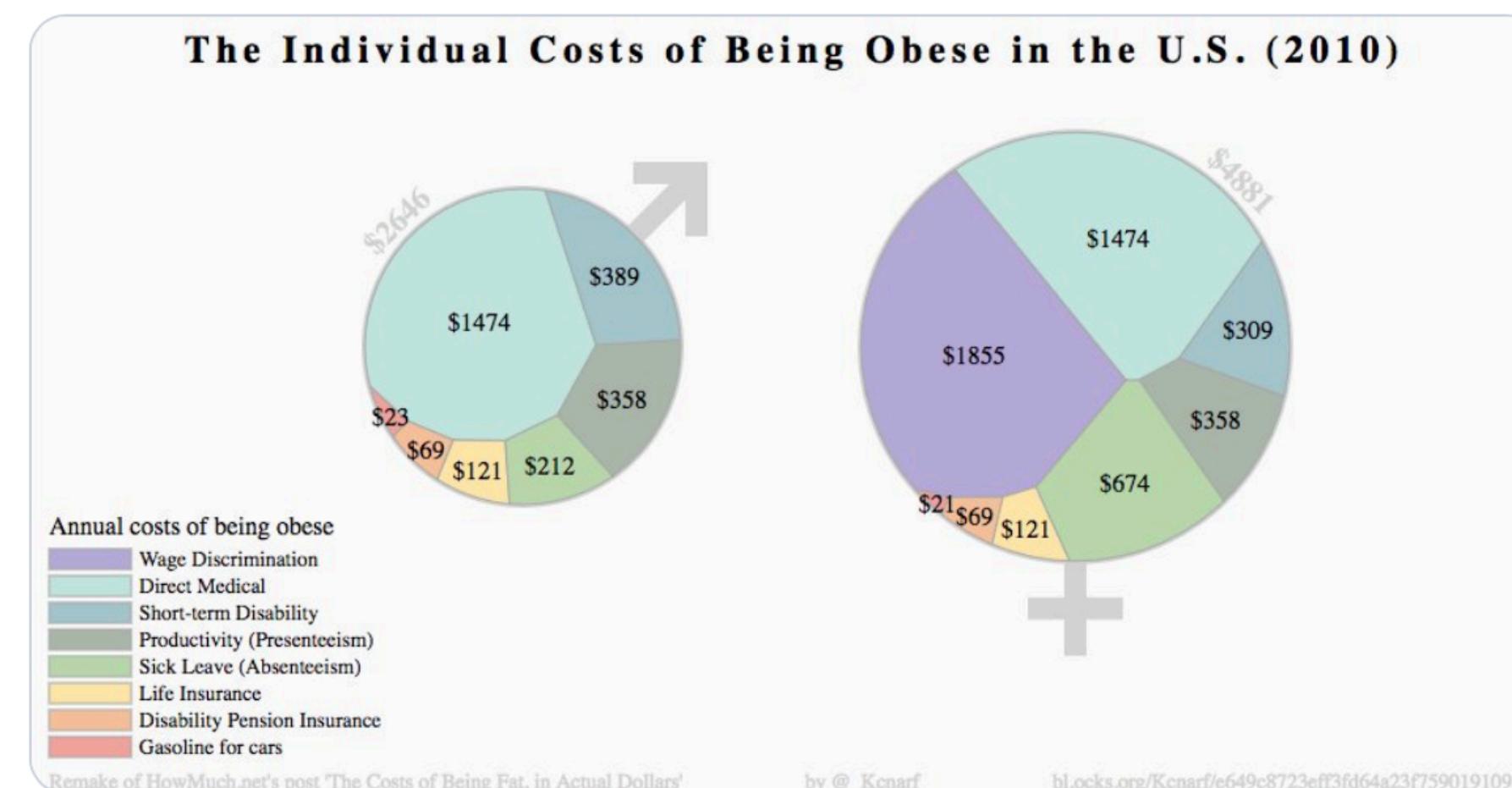


Franck Lebeau
 @_Kcnarf



🎉 Updated API for d3-voronoi-map (github.com/Kcnarf/d3-voro...), which now allows to control the final tessellation.

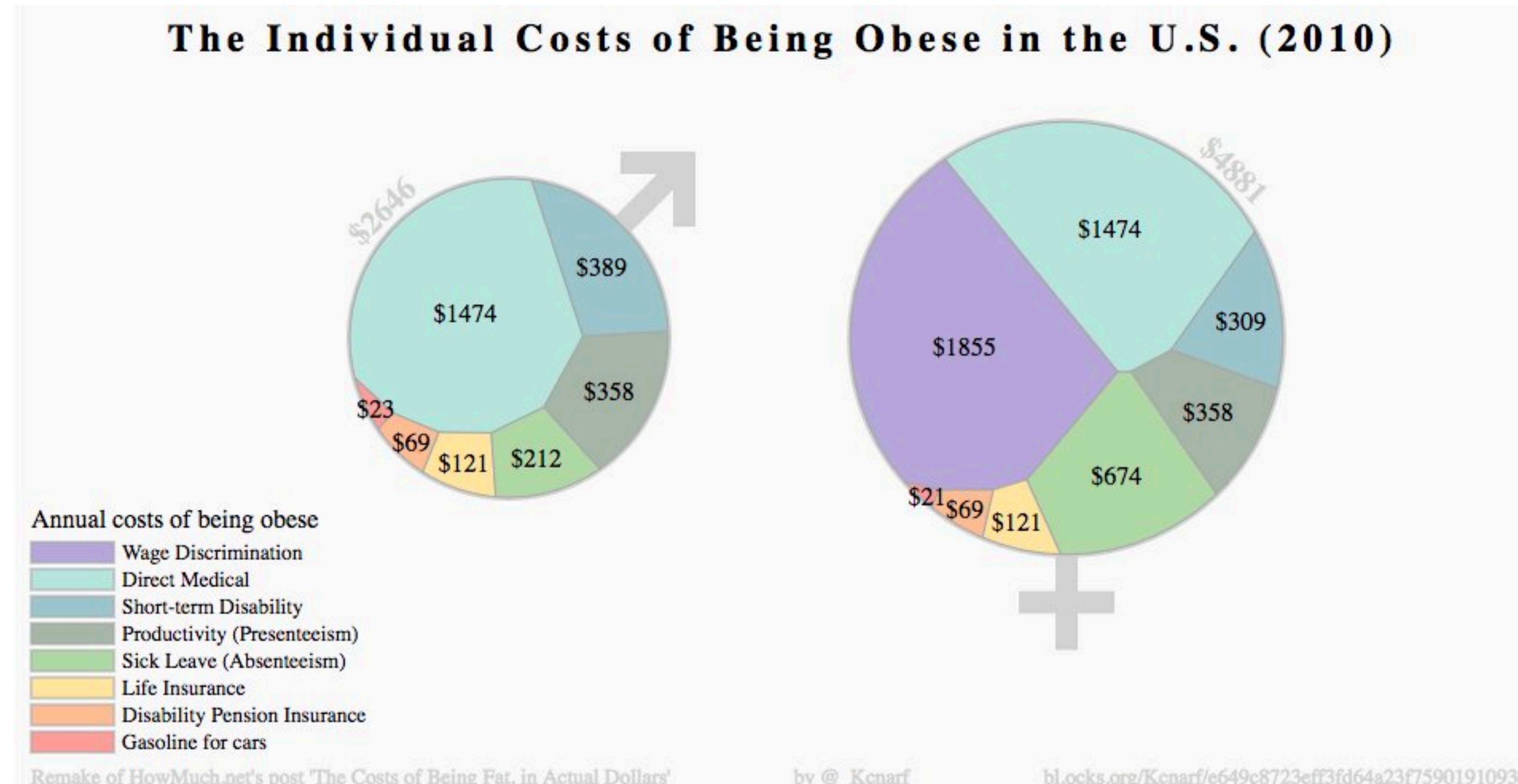
Here ([#D3js](https://bl.ocks.org/Kcnarf/e649c87...) remake of a [@howmuch_net's viz](#) (howmuch.net/articles/obesi...)), placing same cell types at the same positions eases comparison.



8:44 AM · Apr 25, 2018 · Twitter Web Client

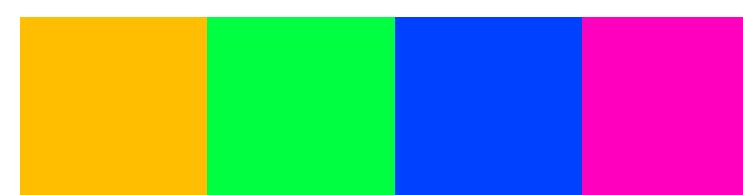
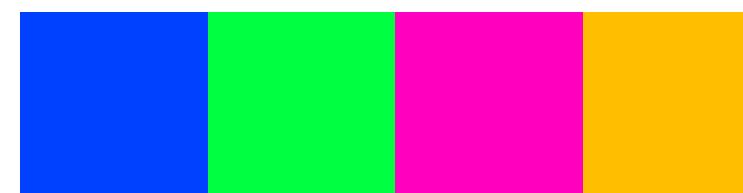
https://twitter.com/_Kcnarf/status/989153096055050240

Redesign Exercise

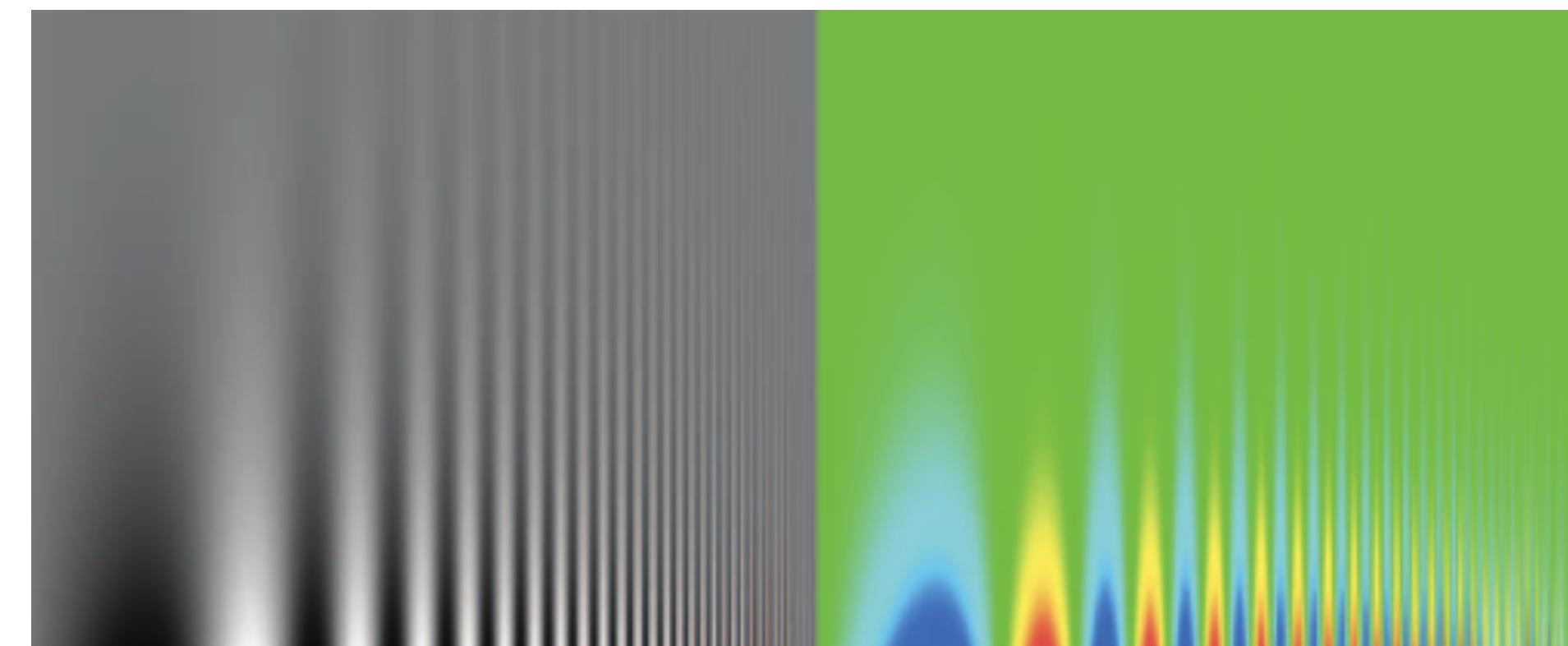


Color Pitfalls: Rainbow Color Map

hard to order



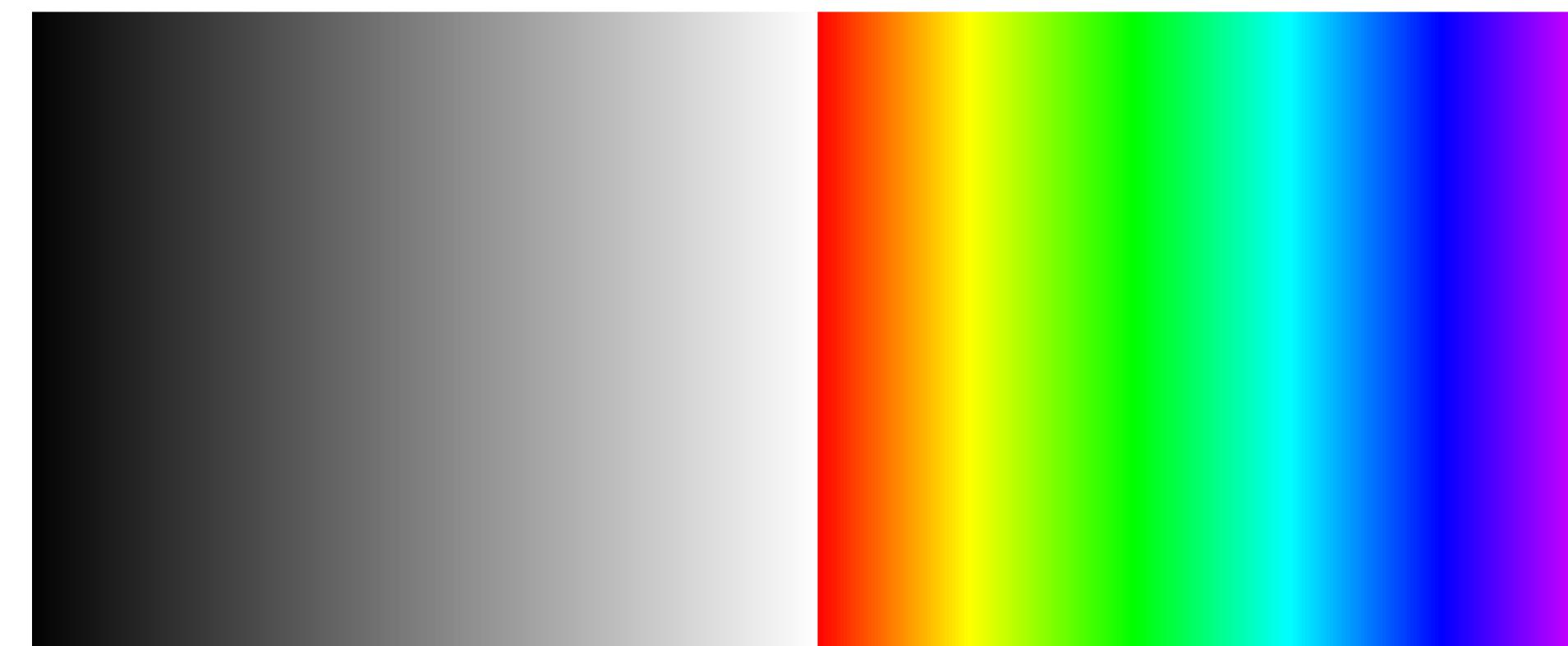
lower resolution



easy to order



creates artifacts



Alternative Color Maps

Normal Vision



Available in the
viridis
R package!

Alternative Color Maps

Deutanopia

(Red-Green
Blindness)



Available in the
viridis
R package!

Alternative Color Maps

Desaturated



Available in the
viridis
R package!

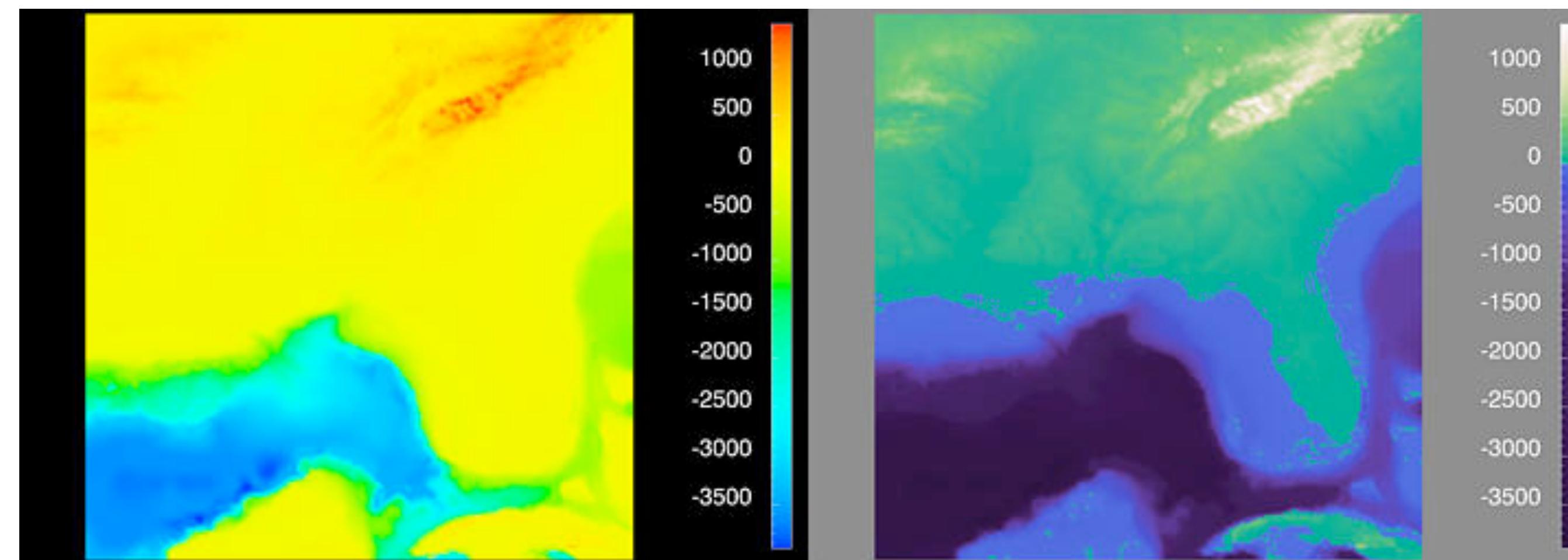
Alternative Color Maps



viridis and **magma** are good choices for a wide range of scenarios.

Color Pitfalls: Rainbow Color Map

Southeastern United States and Gulf of Mexico



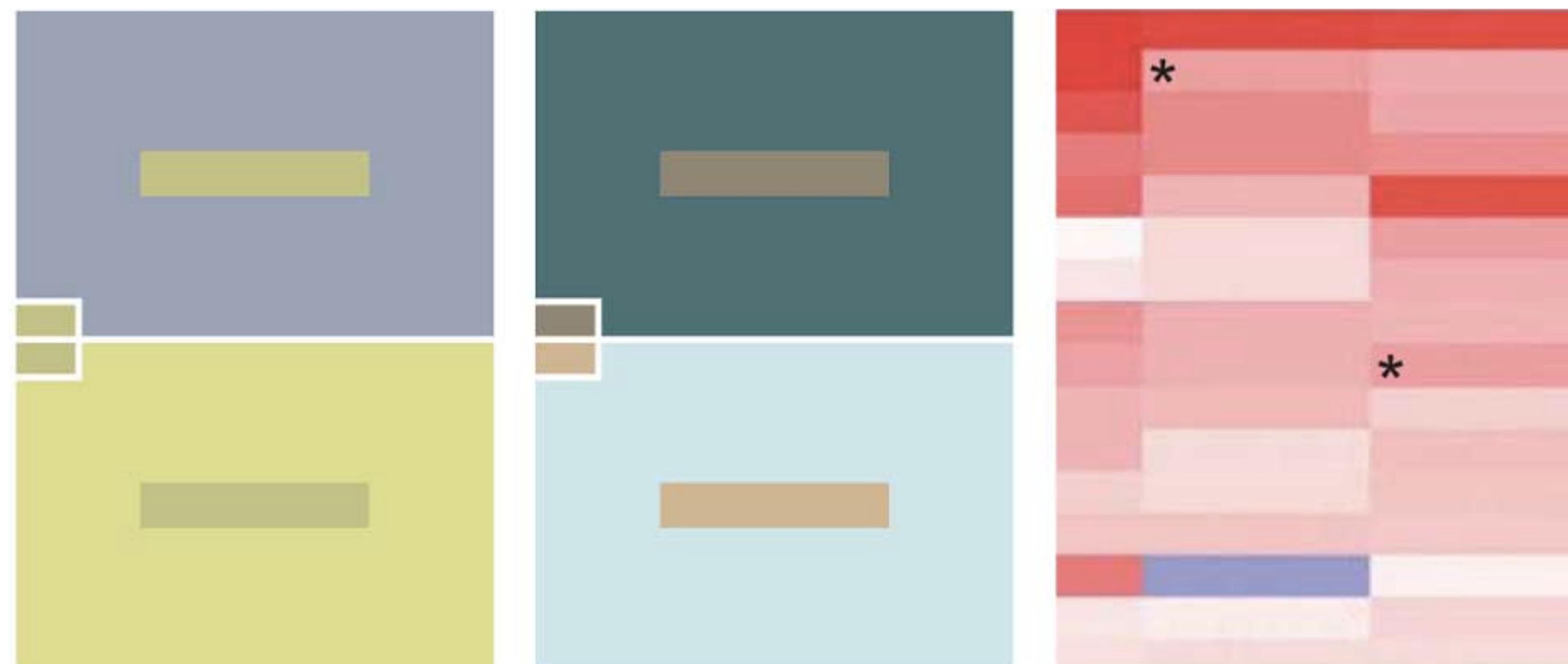
Problems

Zero crossing not explicit.

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

Color Pitfalls: Relativity

Color is a relative medium and context matters



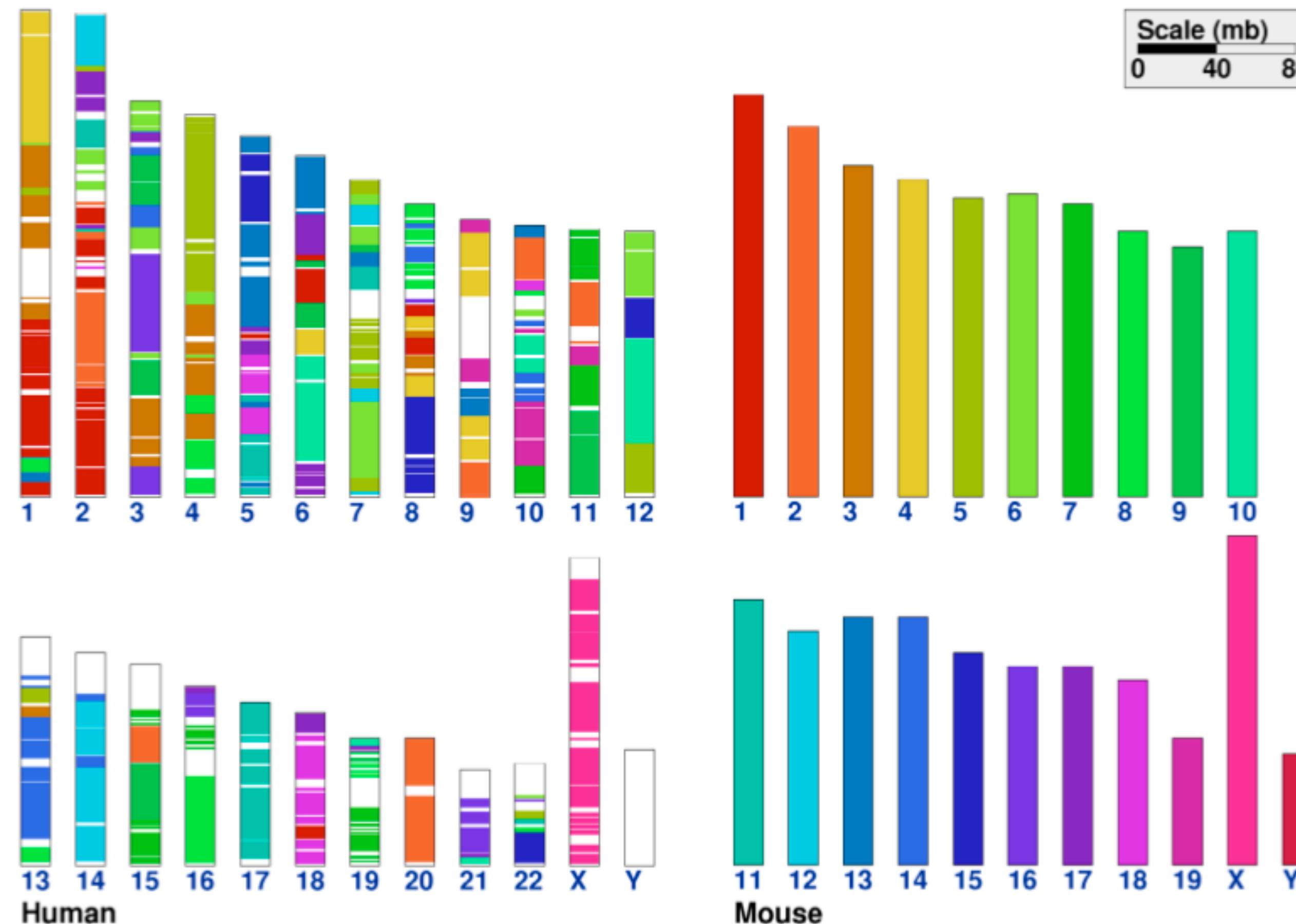
Color Pitfalls: Relativity

Color is a relative medium and context matters

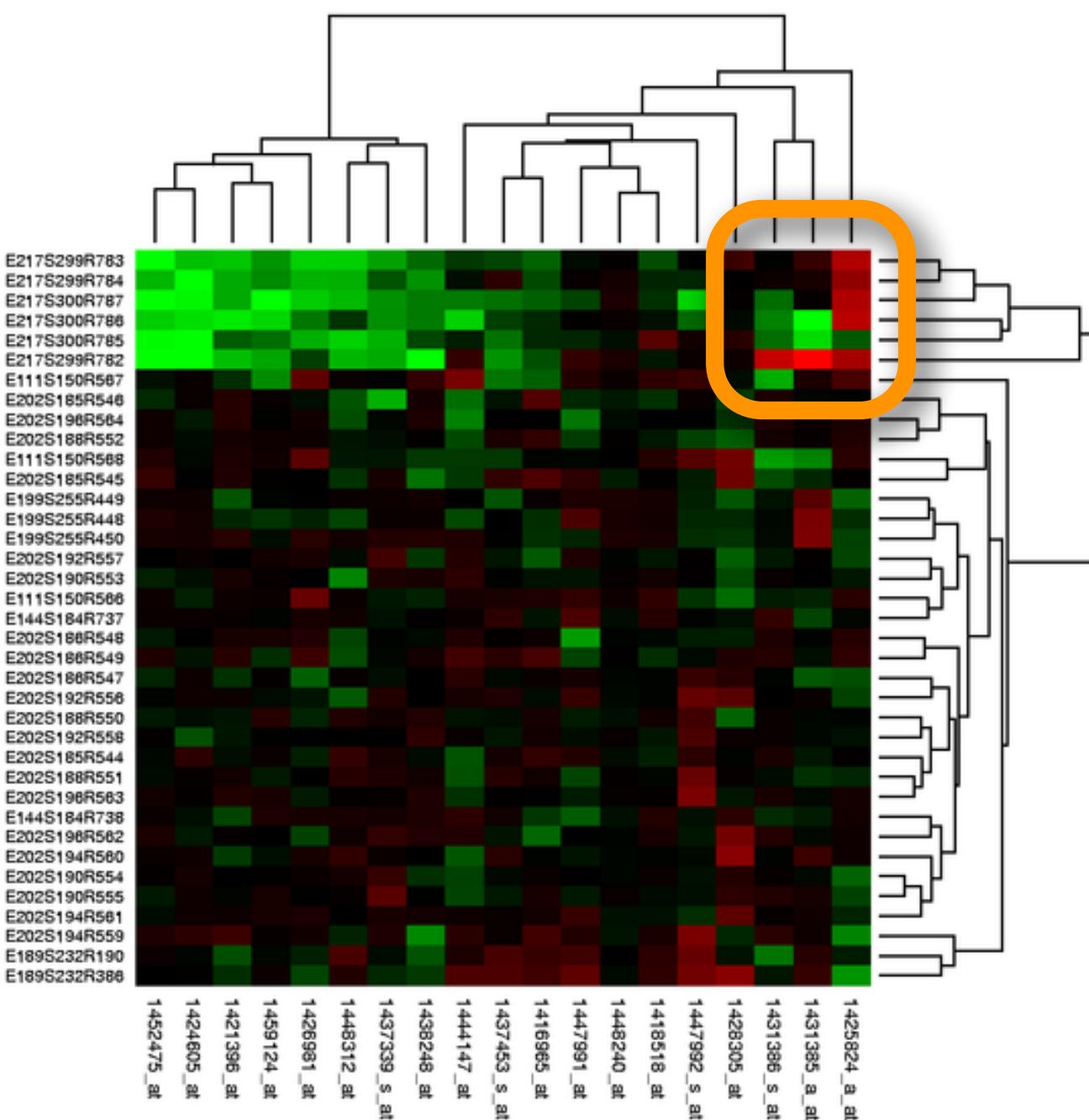


Color Pitfalls: Discriminability

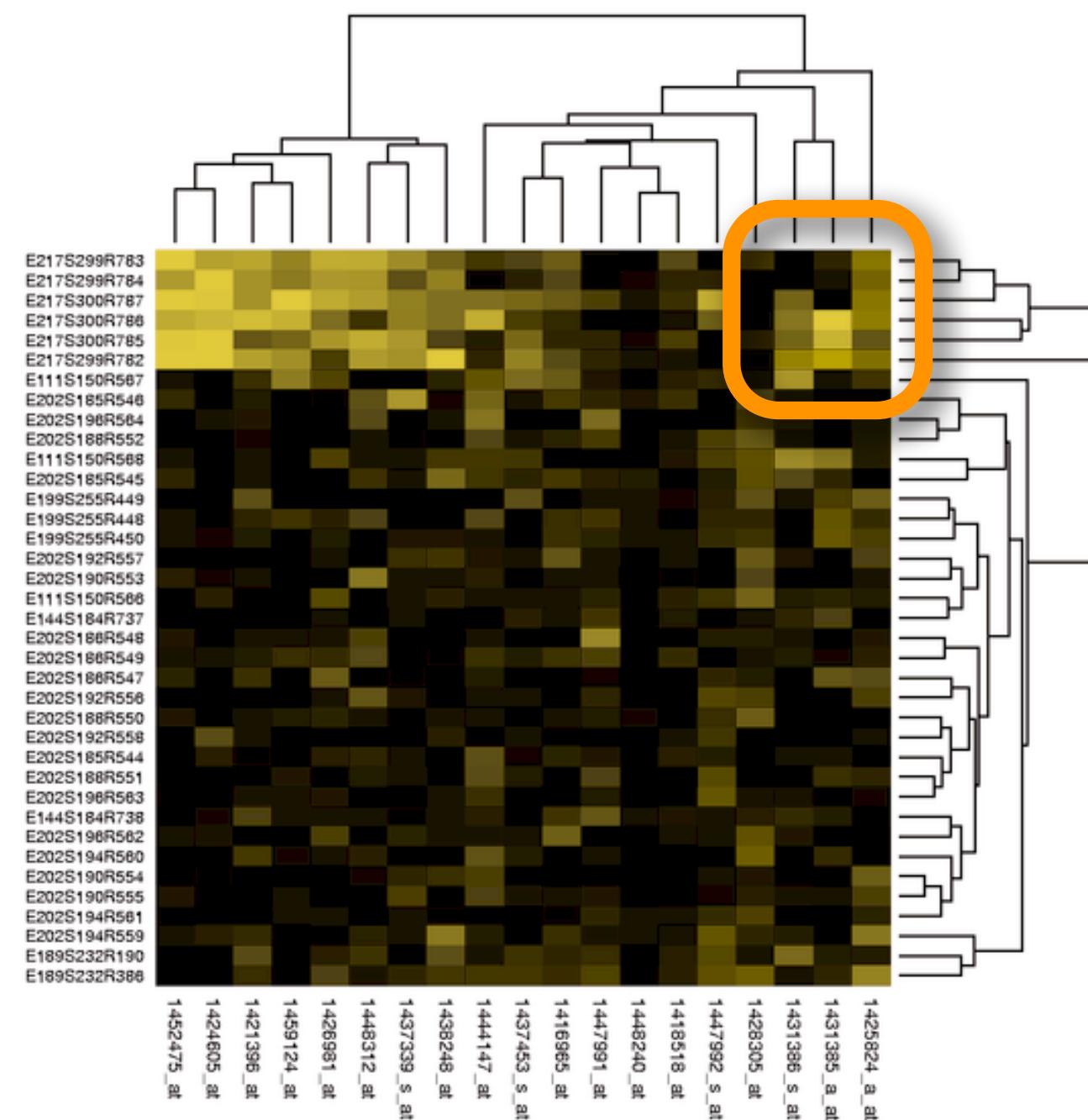
Only 6-12 colors are visually discernible!



Color Pitfalls: Color Blindness

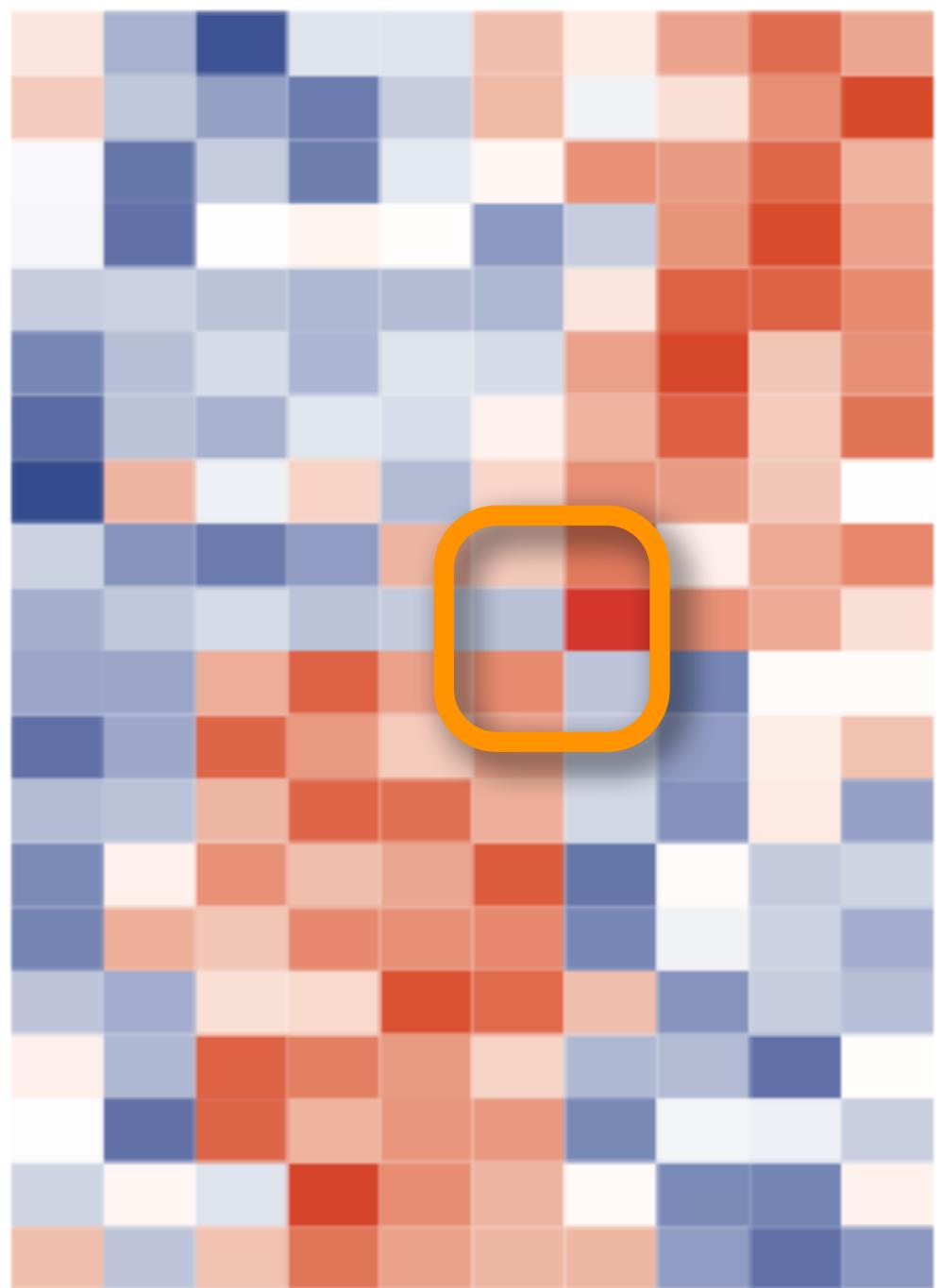


Normal Vision

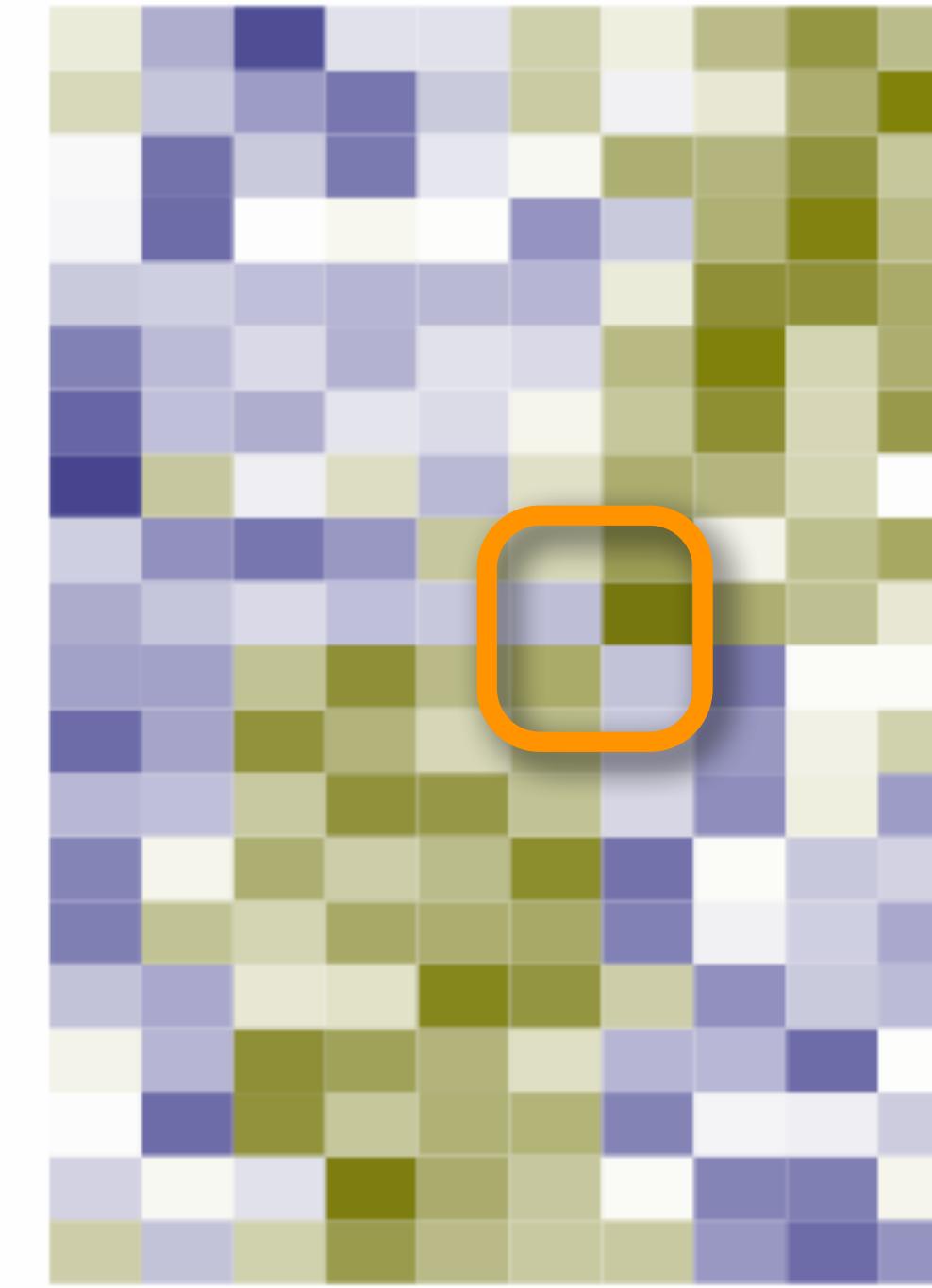


Deuteranope Vision
("Red-Green Blindness")
~ 7% of male population affected

Color Pitfalls: Color Blindness

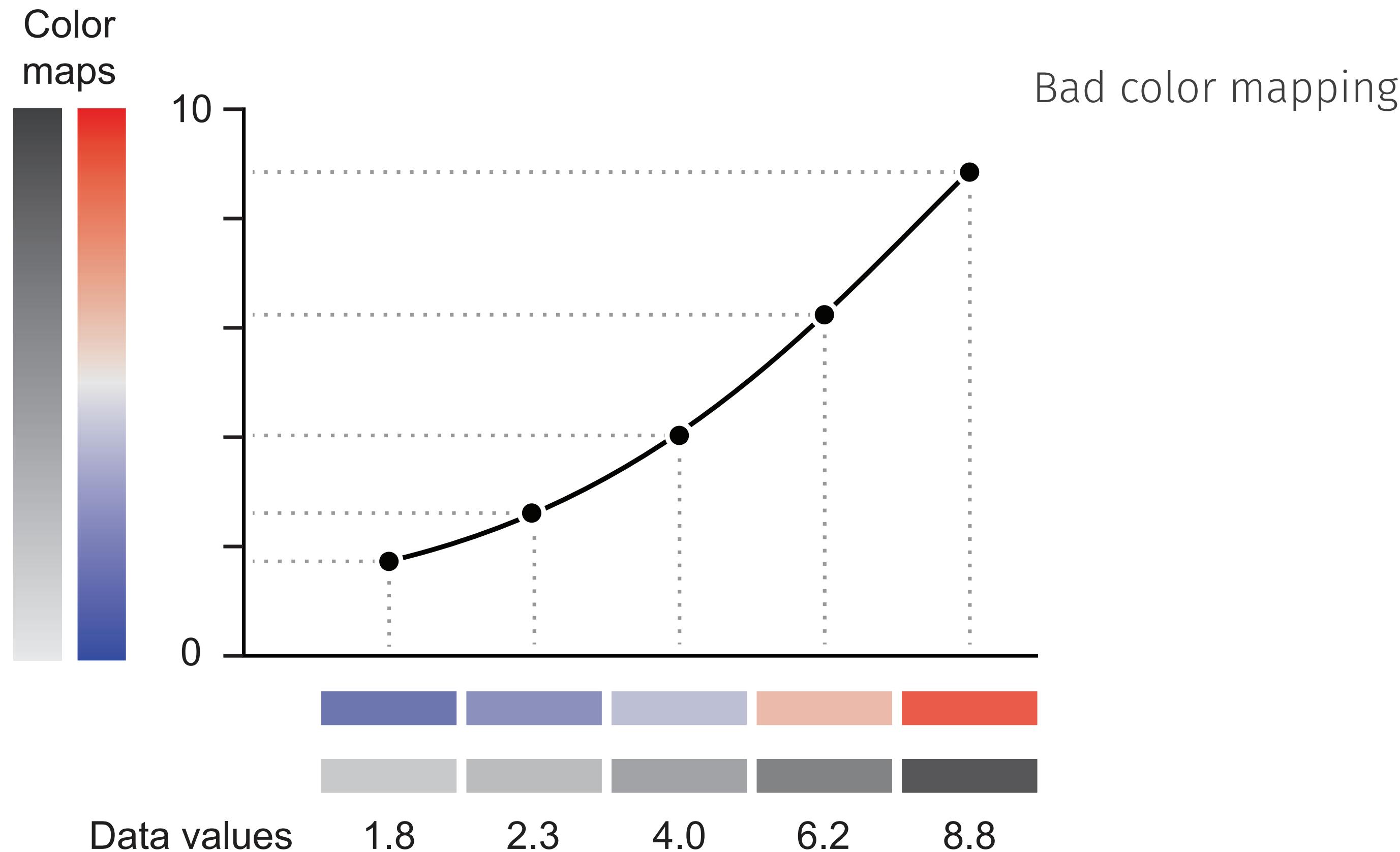


Normal Vision

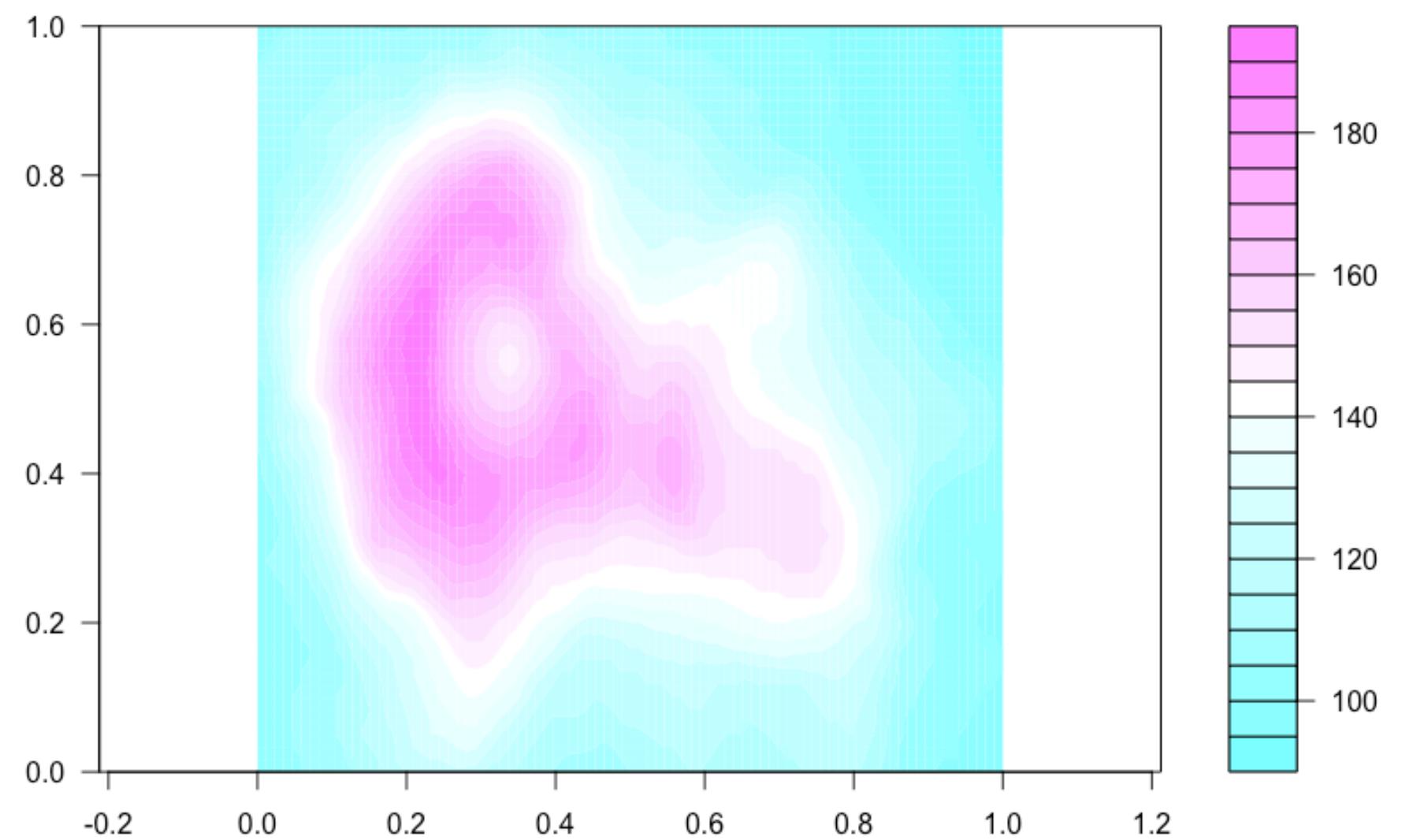


Deutanope Vision
("Red-Green Blindness")
~ 7% of male population affected

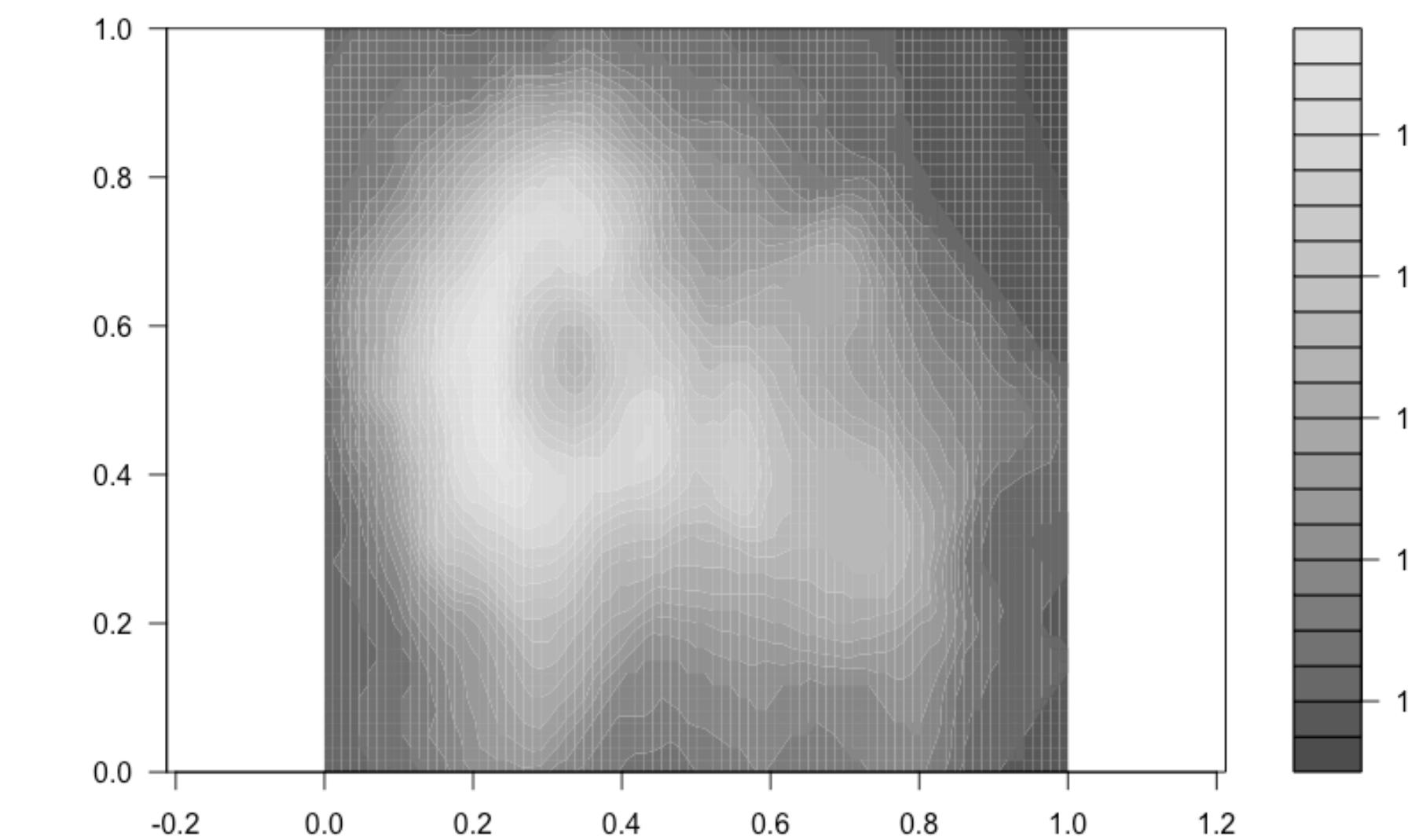
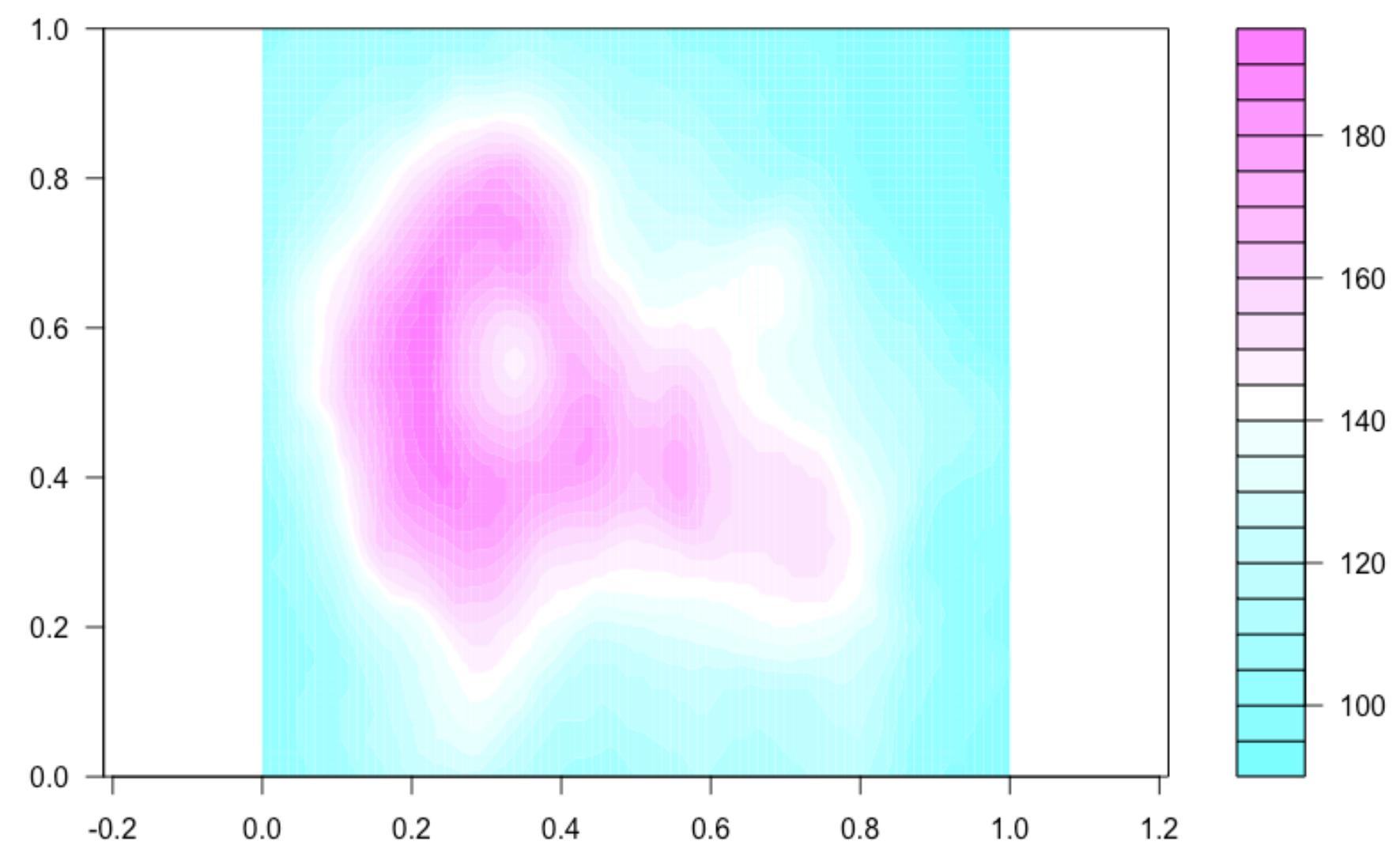
Color Pitfalls: Color Mapping



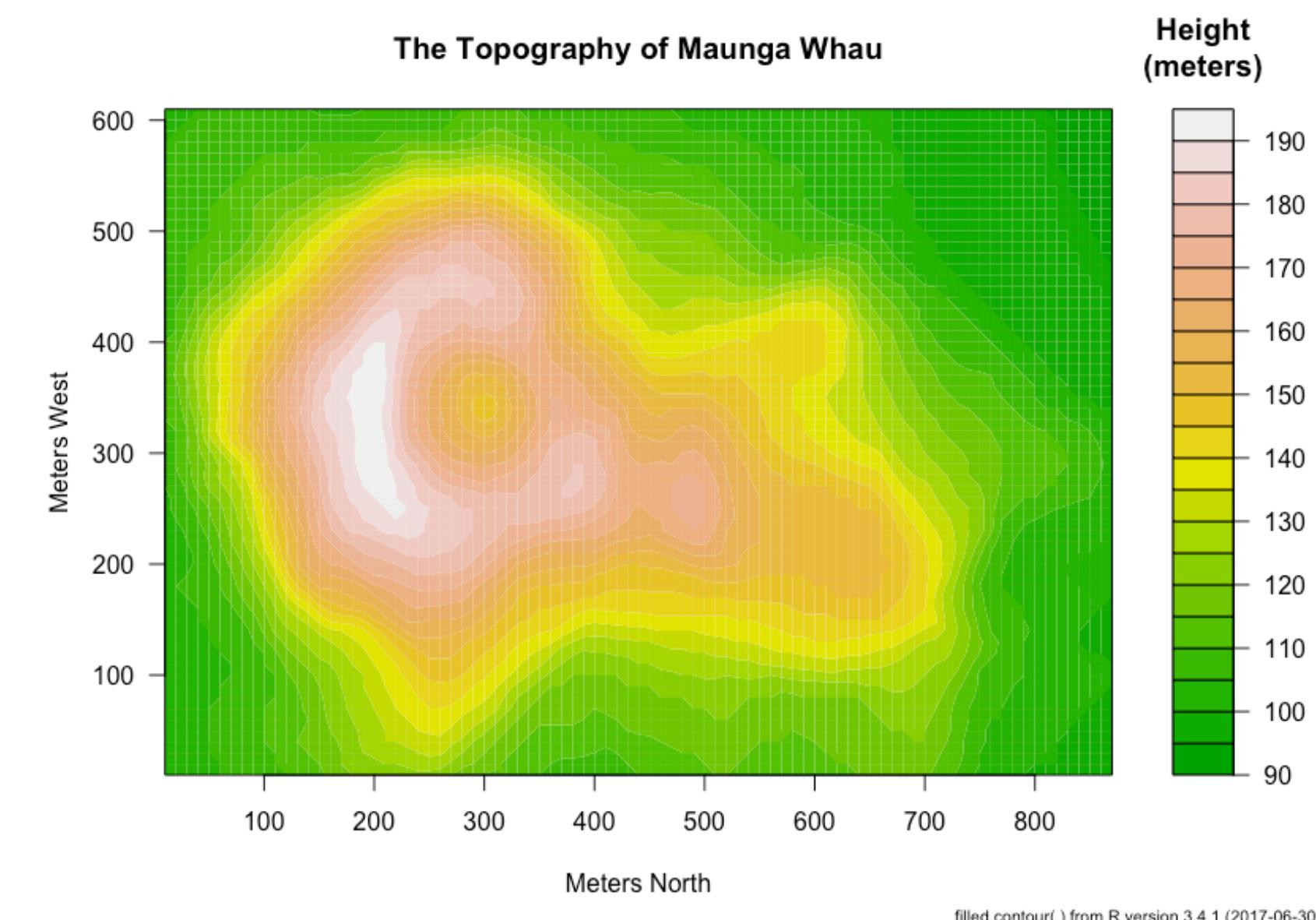
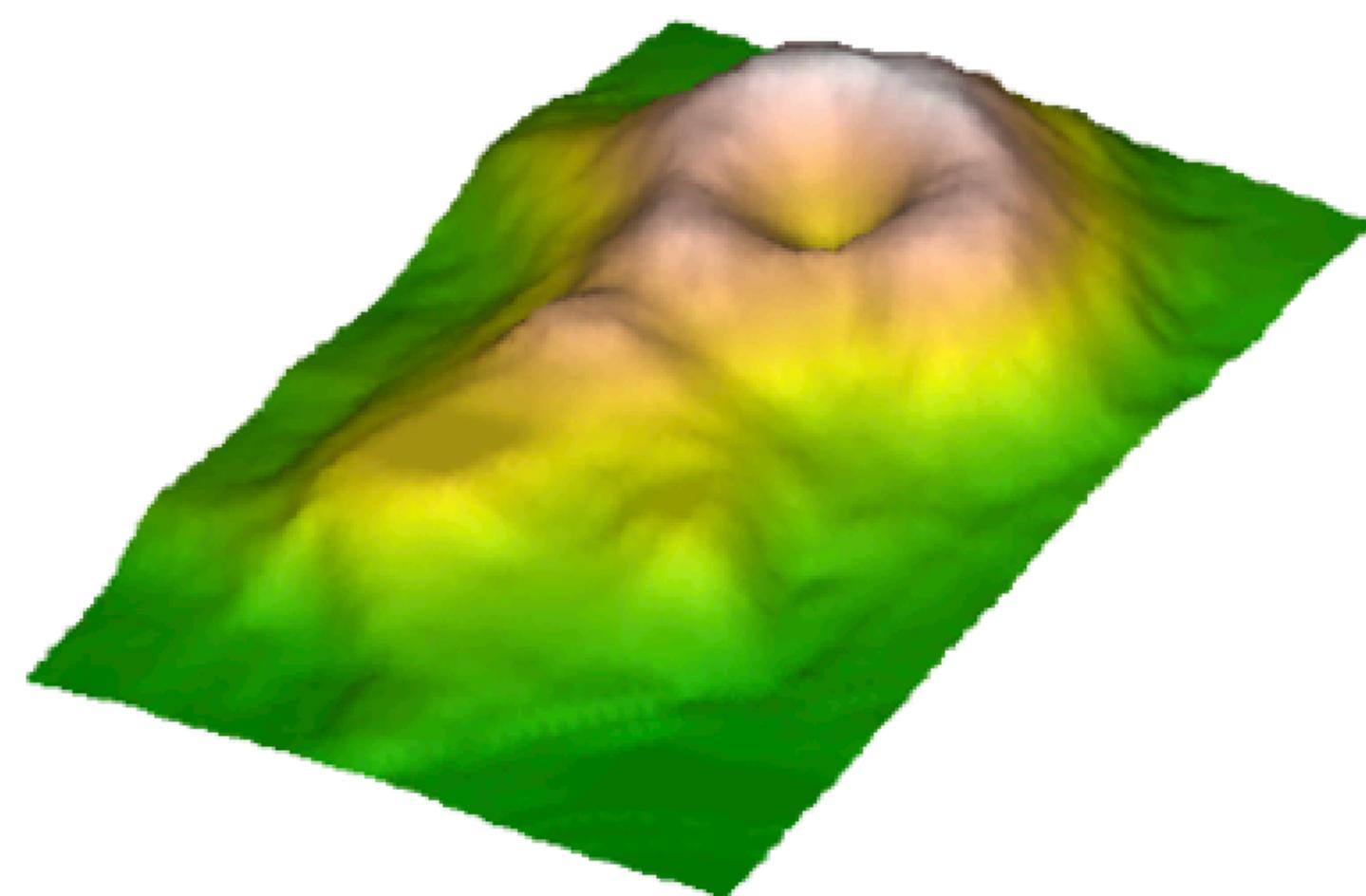
Color Pitfalls: Color Mapping



Color Pitfalls: Color Mapping

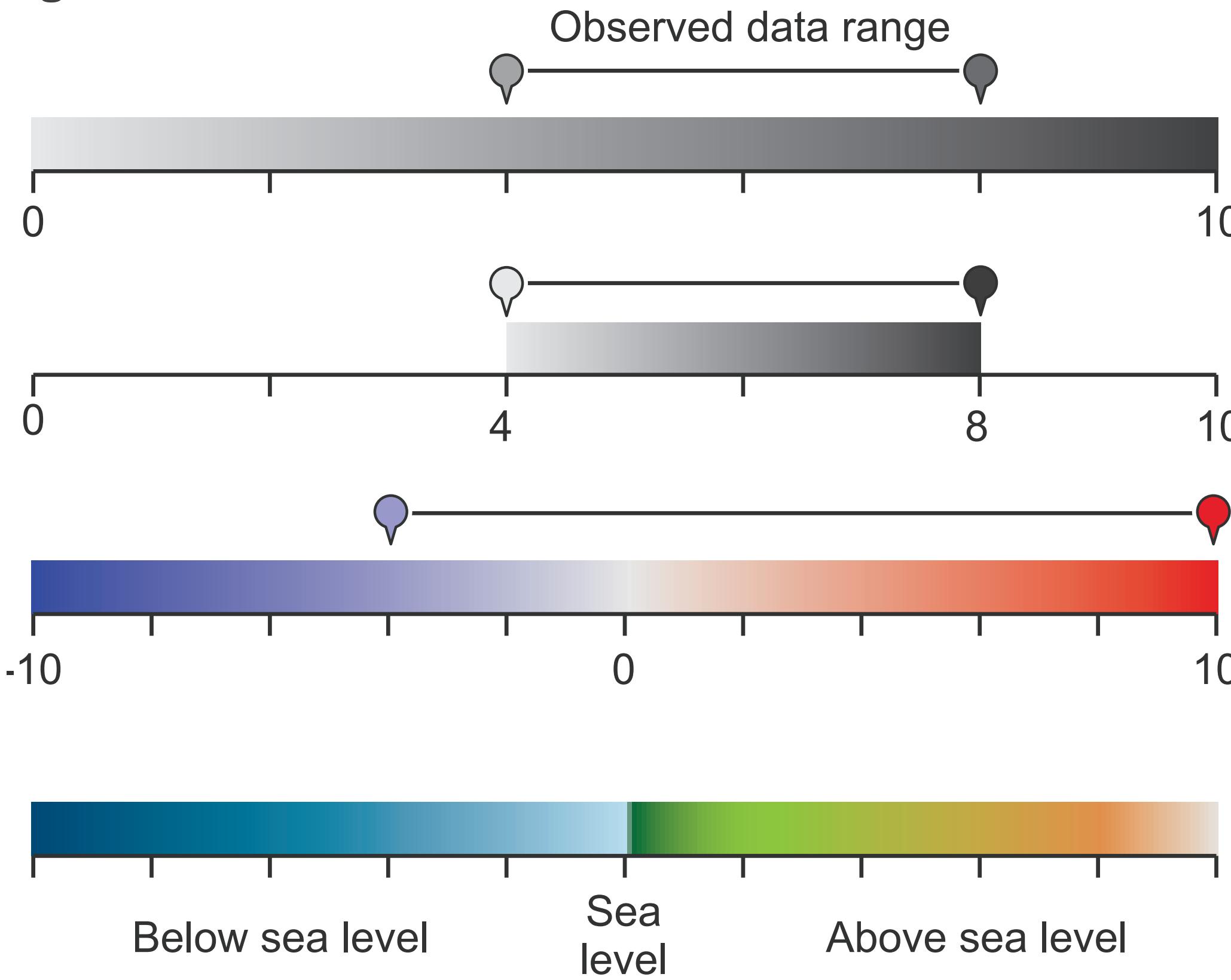


Color Pitfalls: Color Mapping



Color Pitfalls: Color Mapping

Good color mapping!

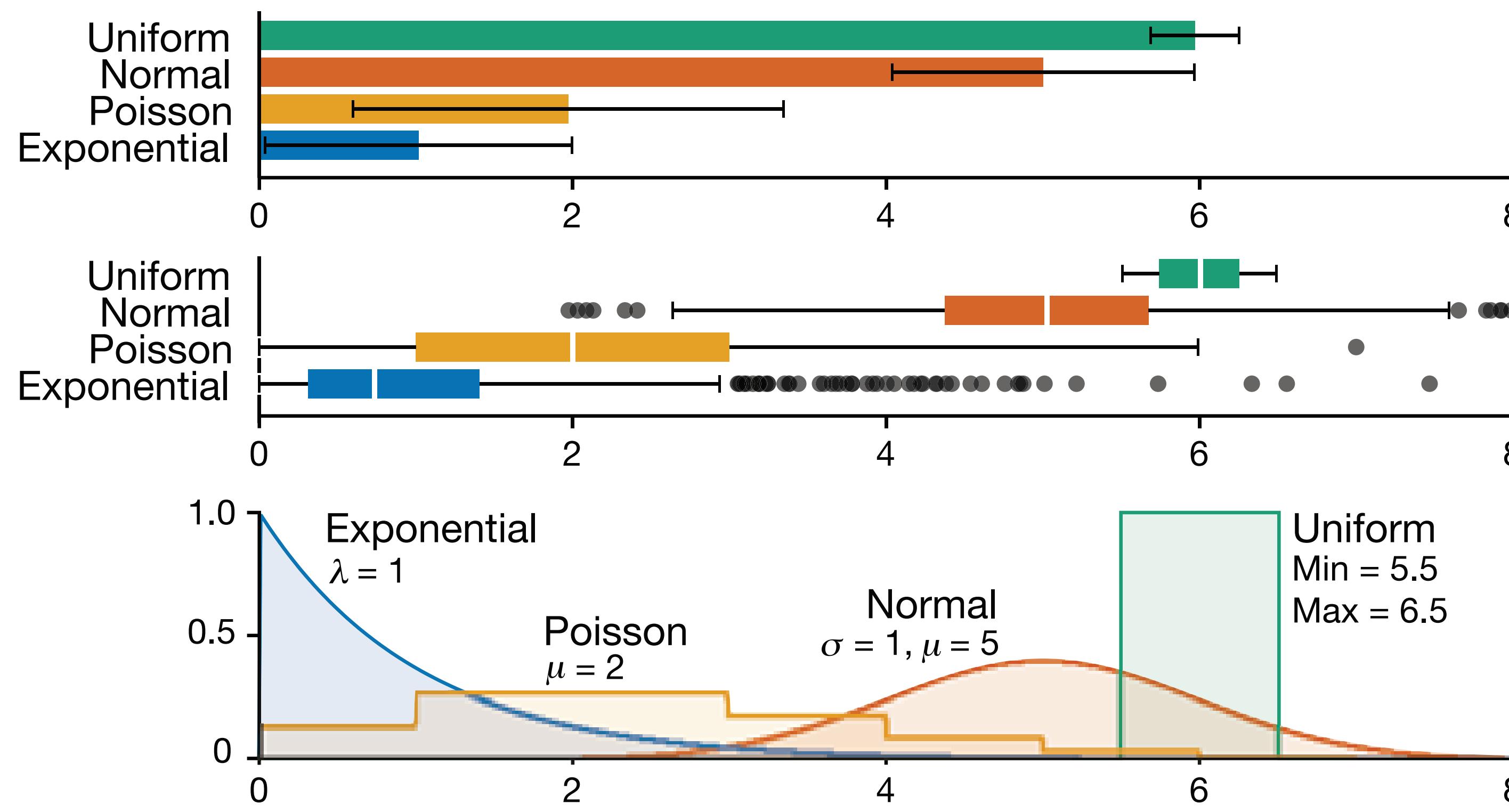


Remember!

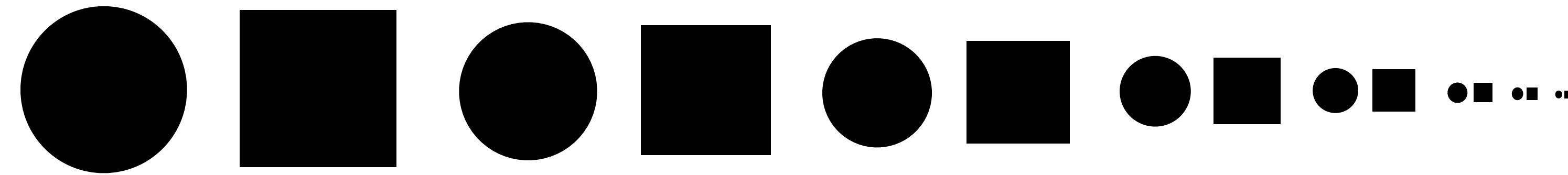
Color used poorly is worse than no color at all.

— Edward Tufte

Location Pitfalls: Data Location

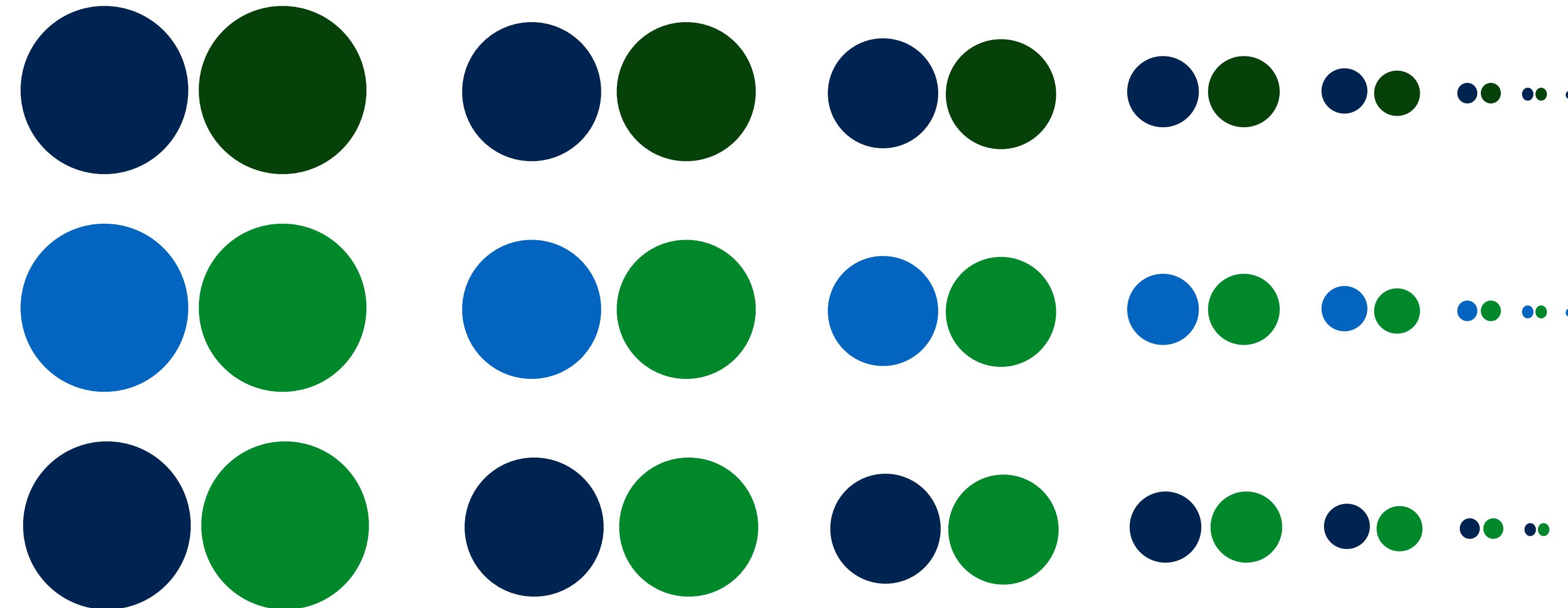


Encoding Pitfalls: Interference



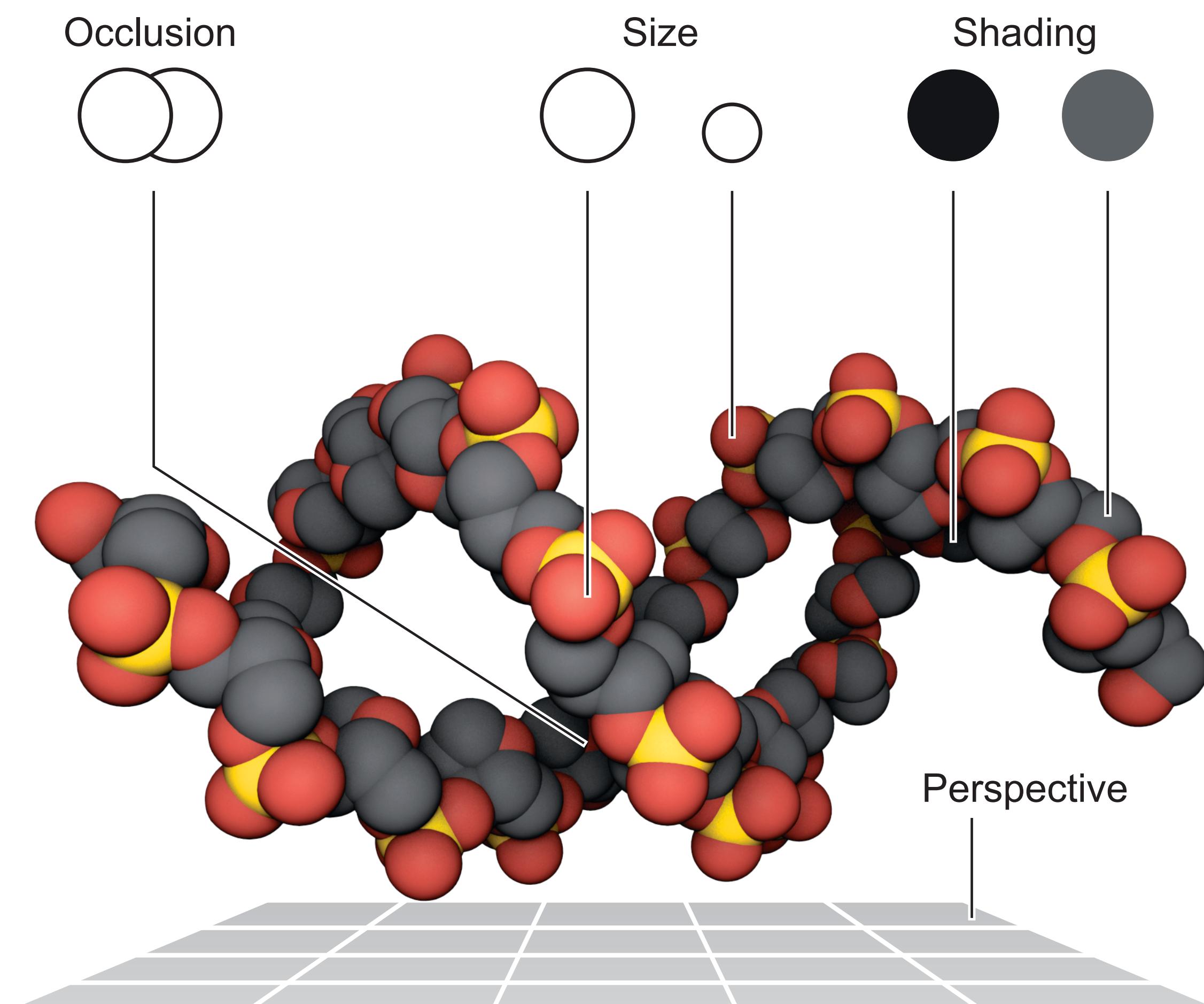
shape and size

Encoding Pitfalls: Interference



color and size

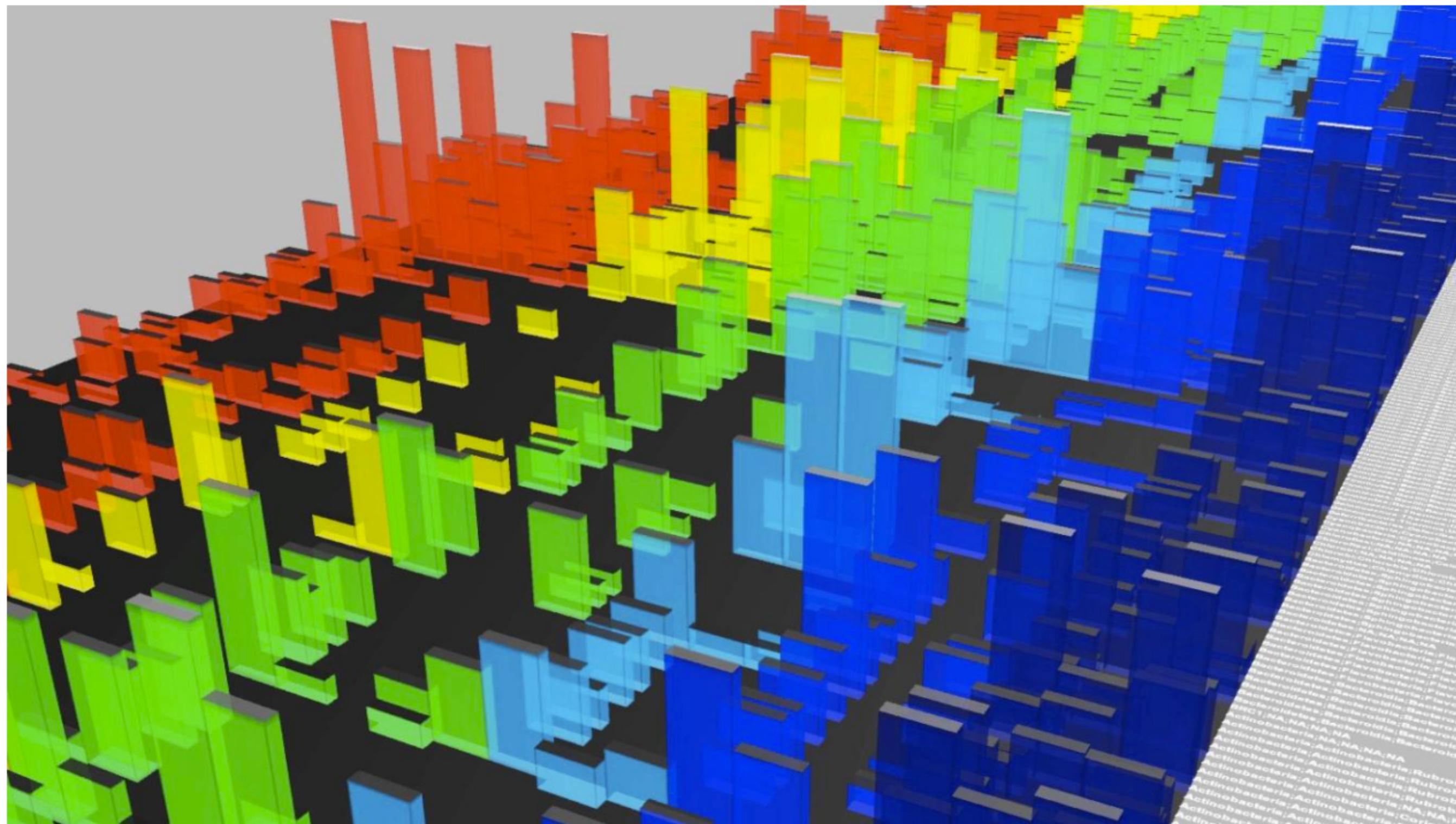
3D: Depth Cues



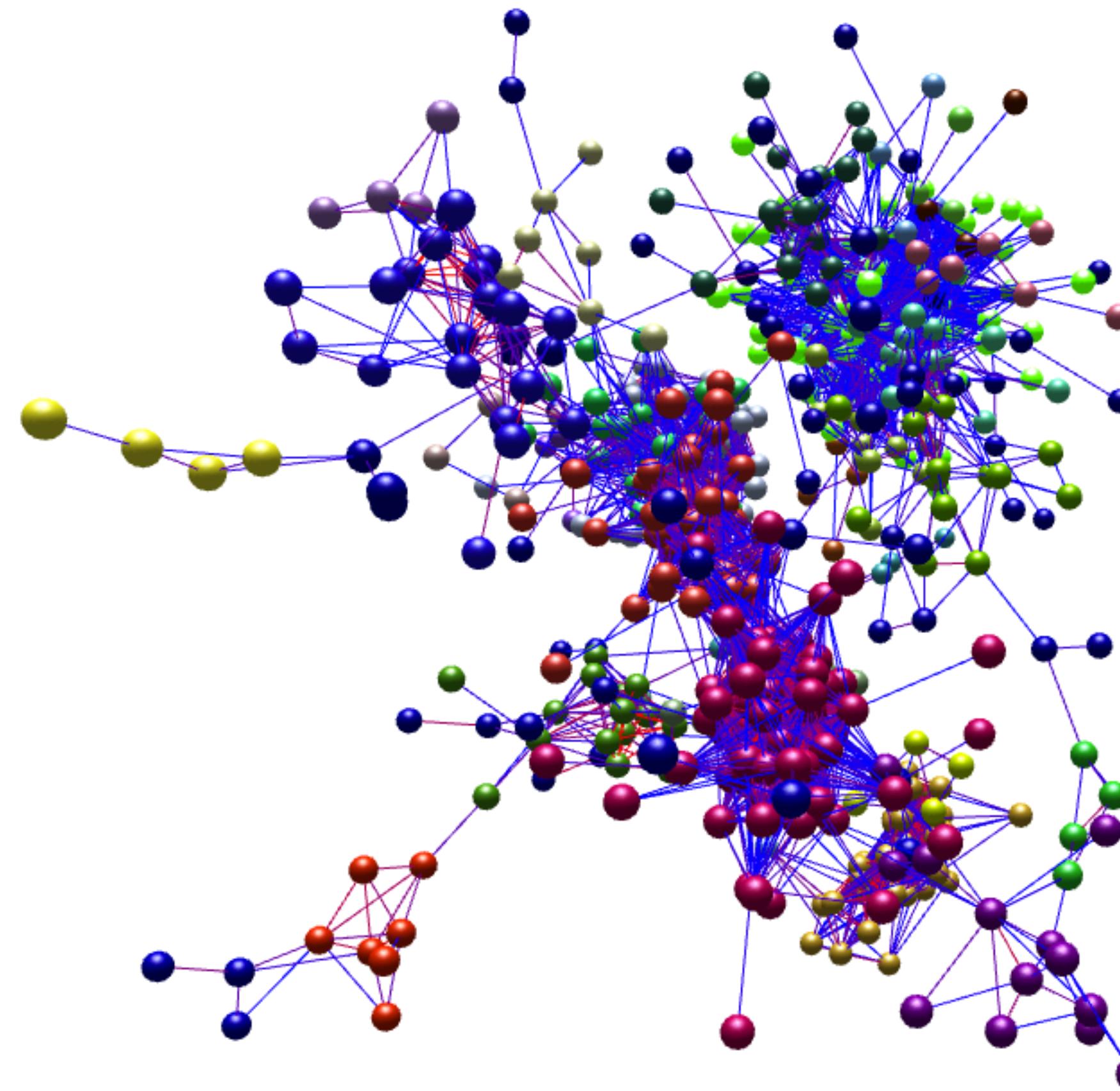
3D Pitfalls: Perspective

Perspective distortion: interferes with size channel encoding

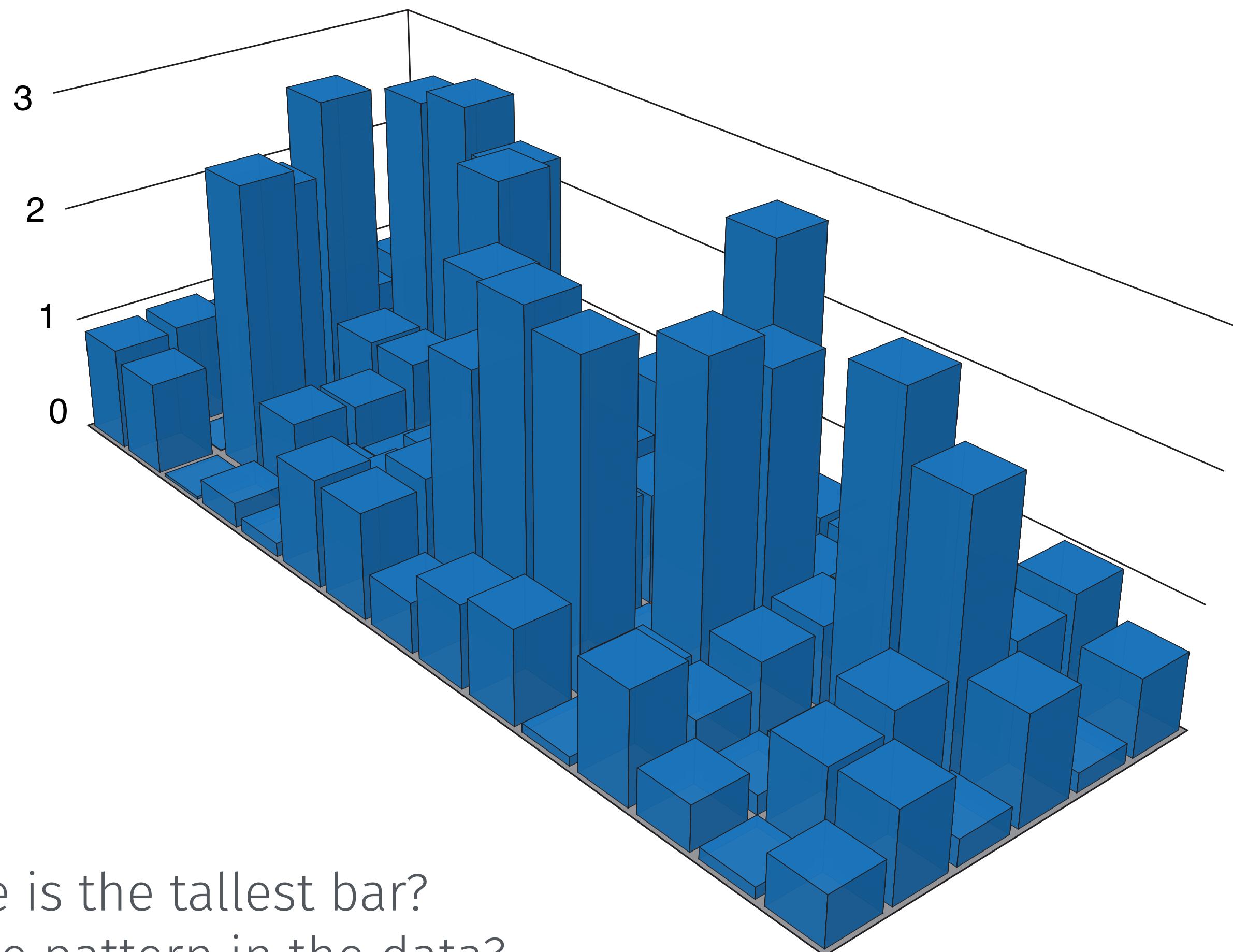
Shading: interferes with color, lightness, and saturation channel encodings



3D Pitfalls: Occlusion

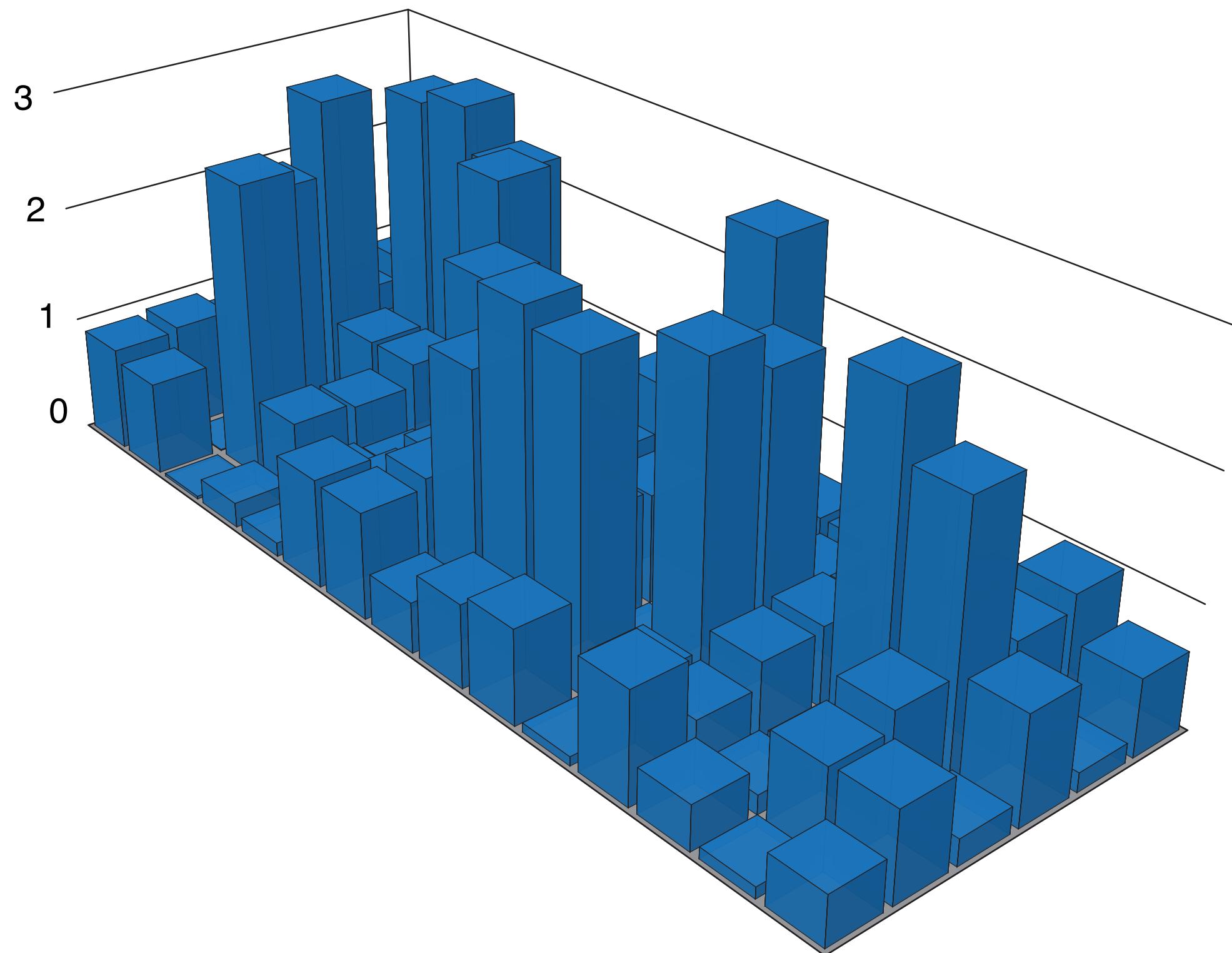


3D Pitfalls: Occlusion & Perspective

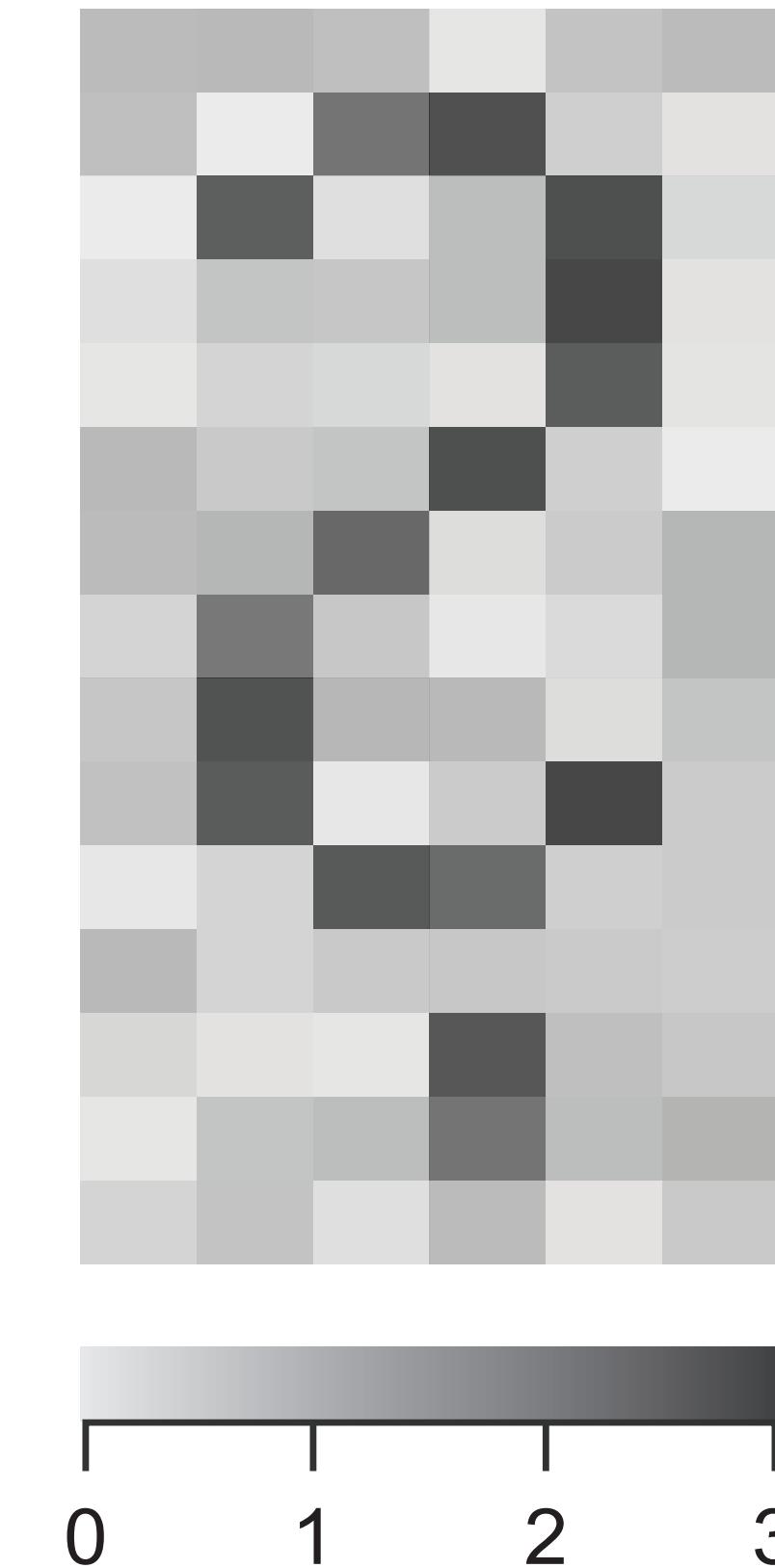


Which one is the tallest bar?
What is the pattern in the data?

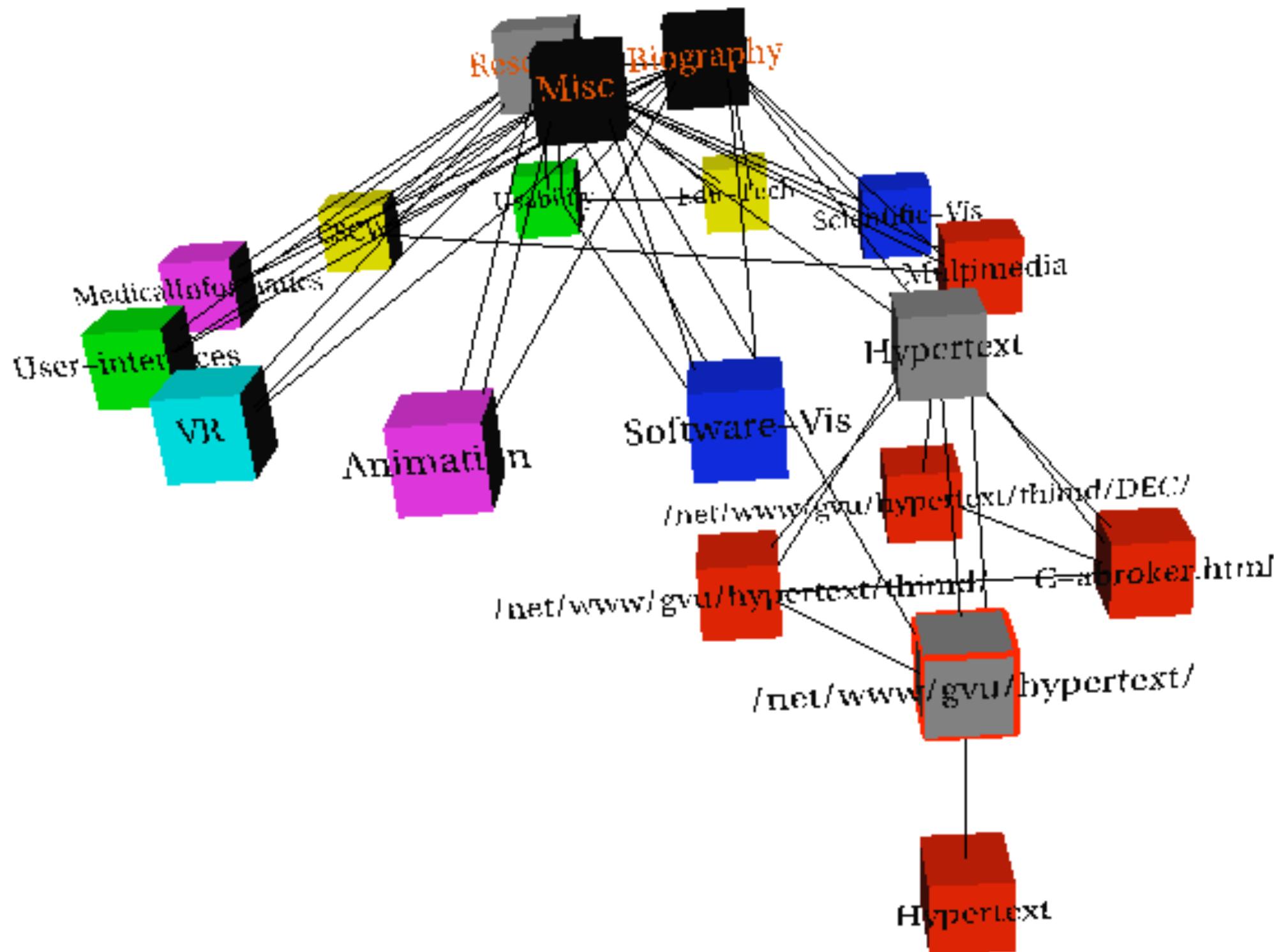
3D Pitfalls: Occlusion & Perspective



Which one is the tallest bar?
What is the pattern in the data?



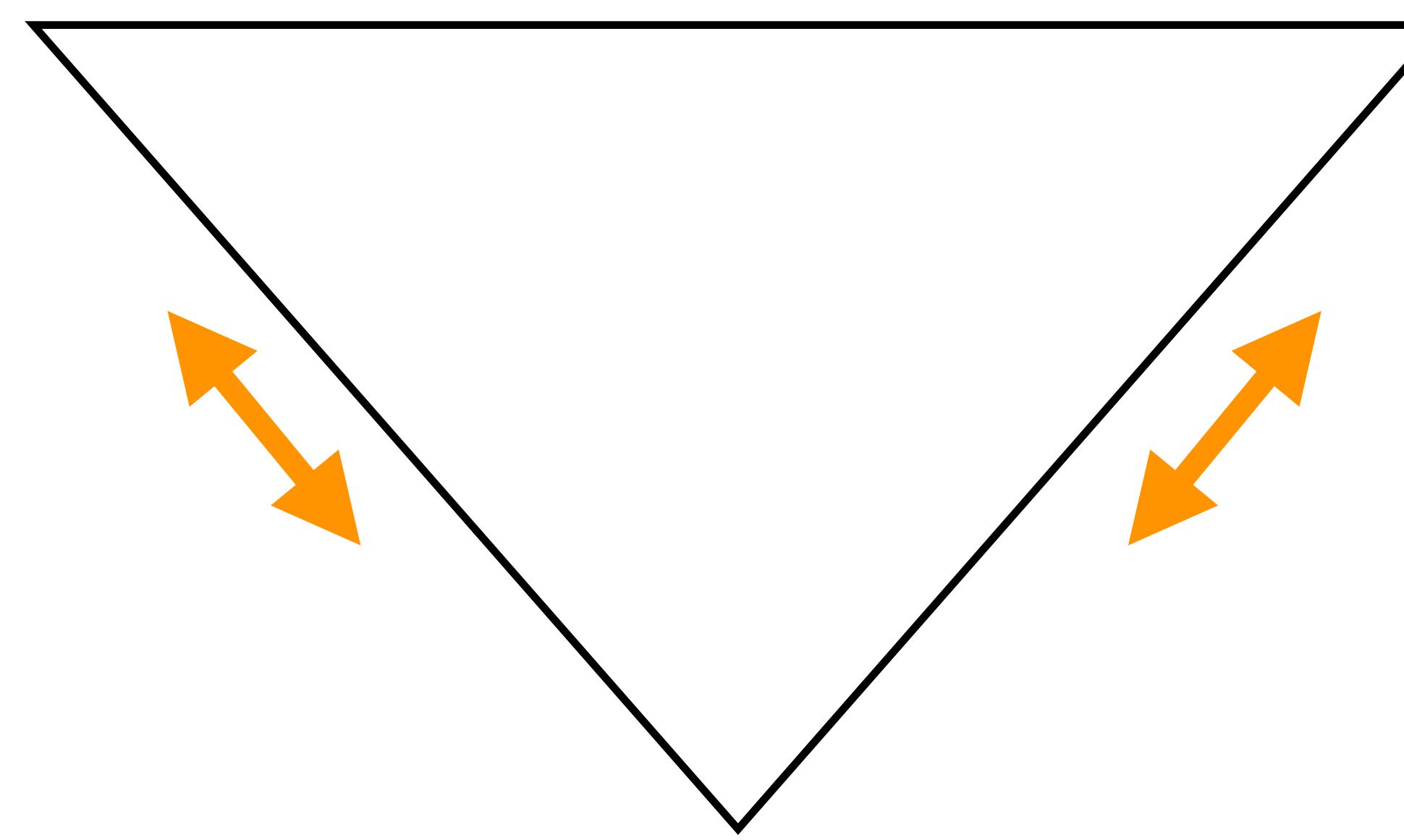
3D Pitfalls: Text Legibility



Visualization for Communication

dialogue between computer & analyst

Confirmation **Exploration**



Presentation

dialogue between analyst & audience

Insights into the Paper, Data into the Database

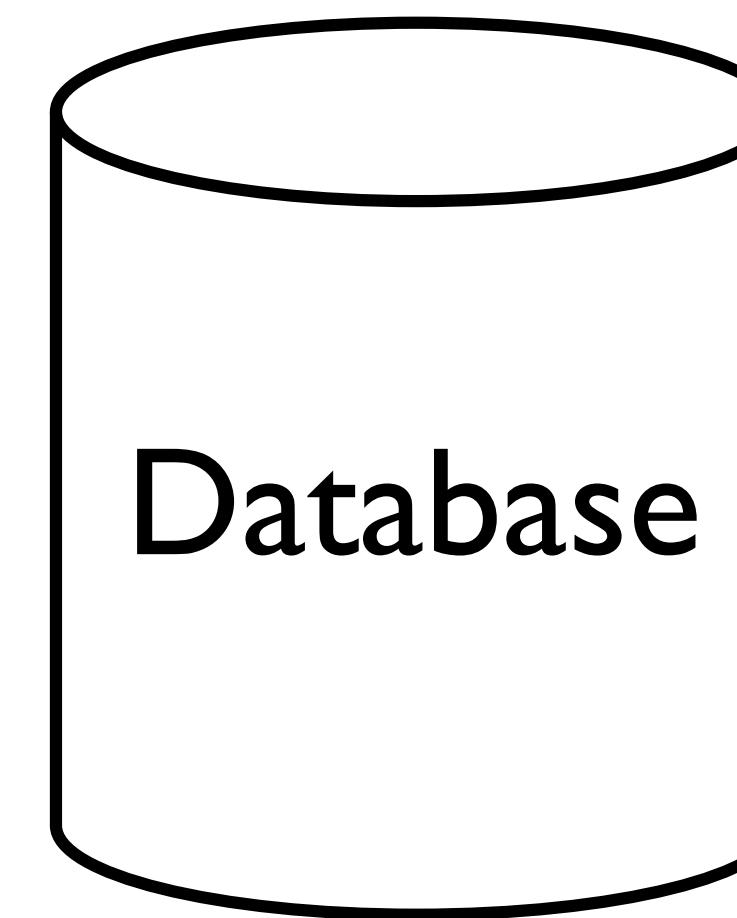
ARTICLE

doi:10.1038/nature11412

Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network*

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.

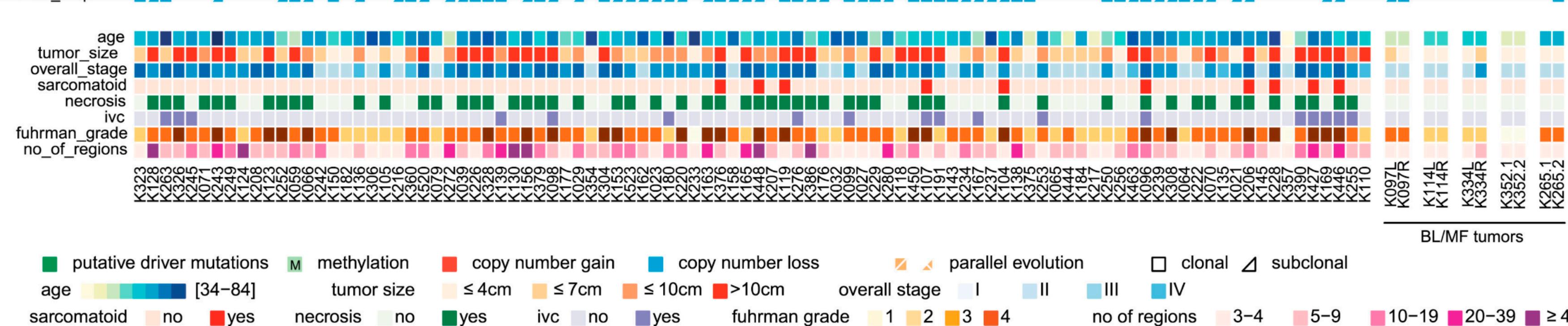
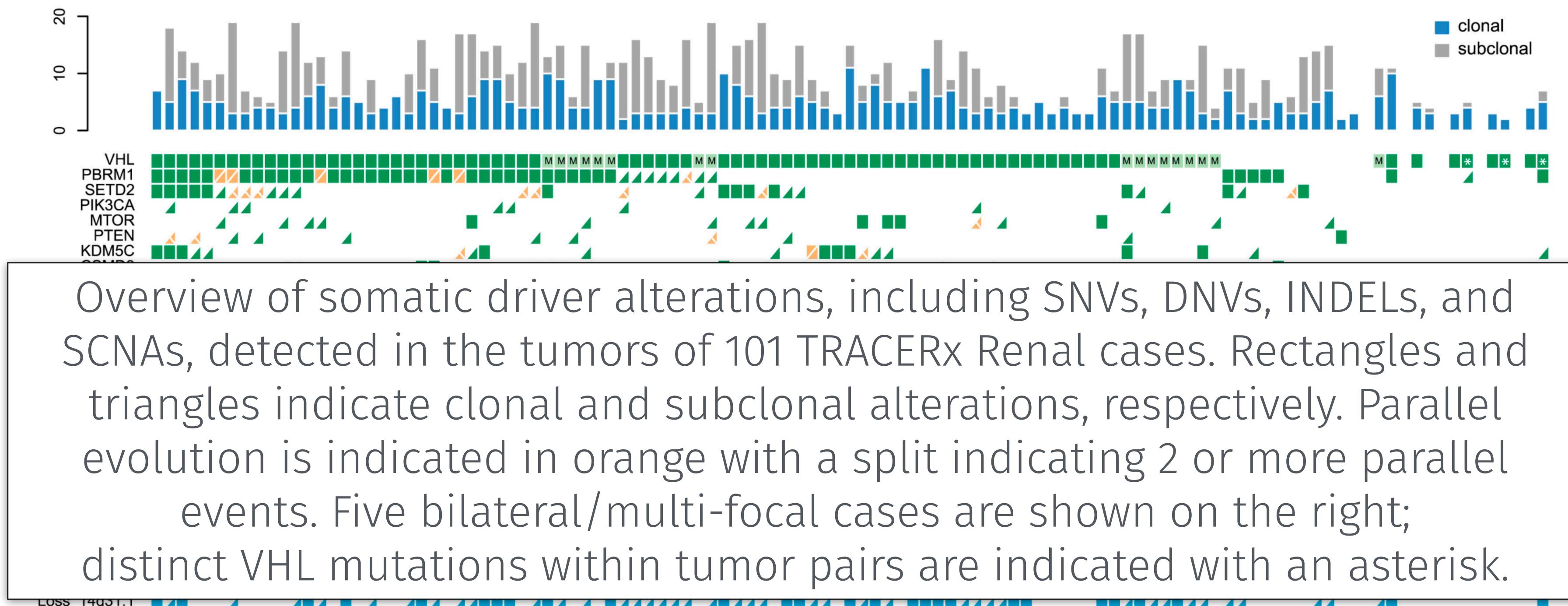


Insights into the Paper, Data into the Database

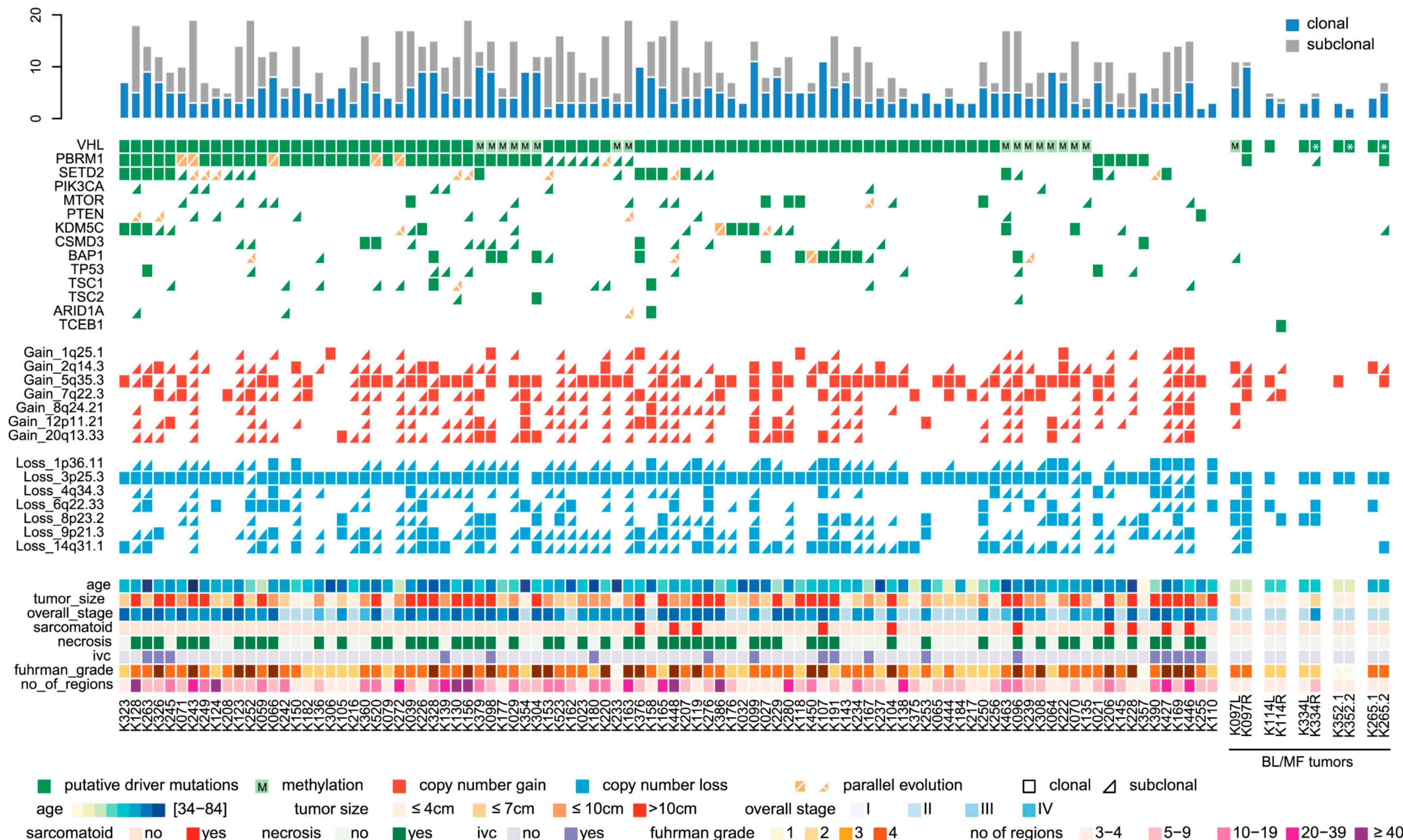
DEls, and
ngles and
y. Parallel
e parallel
ight;
asterisk.



Insights into the Paper, Data into the Database



Insights into the Paper, Data into the Database

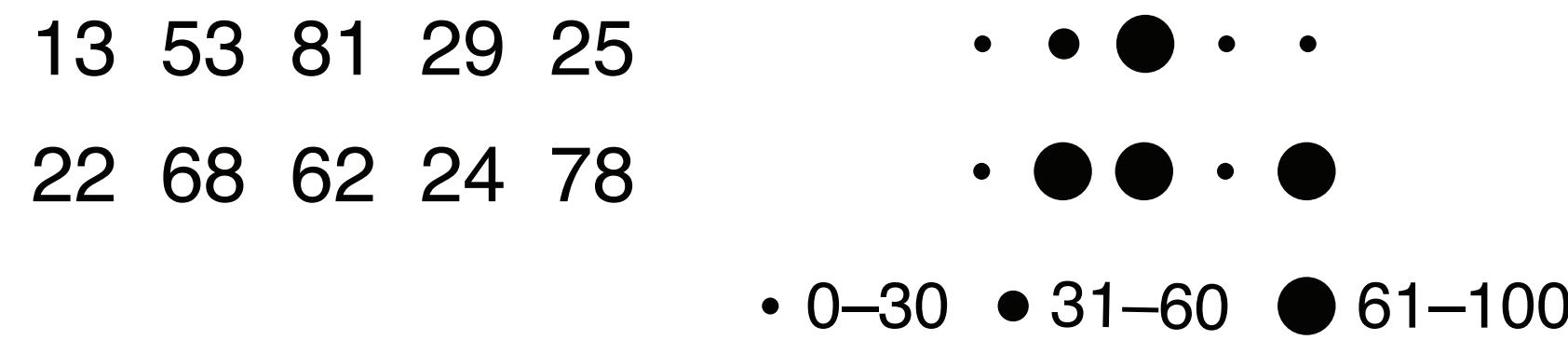


Overview
SCNAs, det
triangles
evolutio
even
distinct V

Focus on the Message: Reduce Data Detail

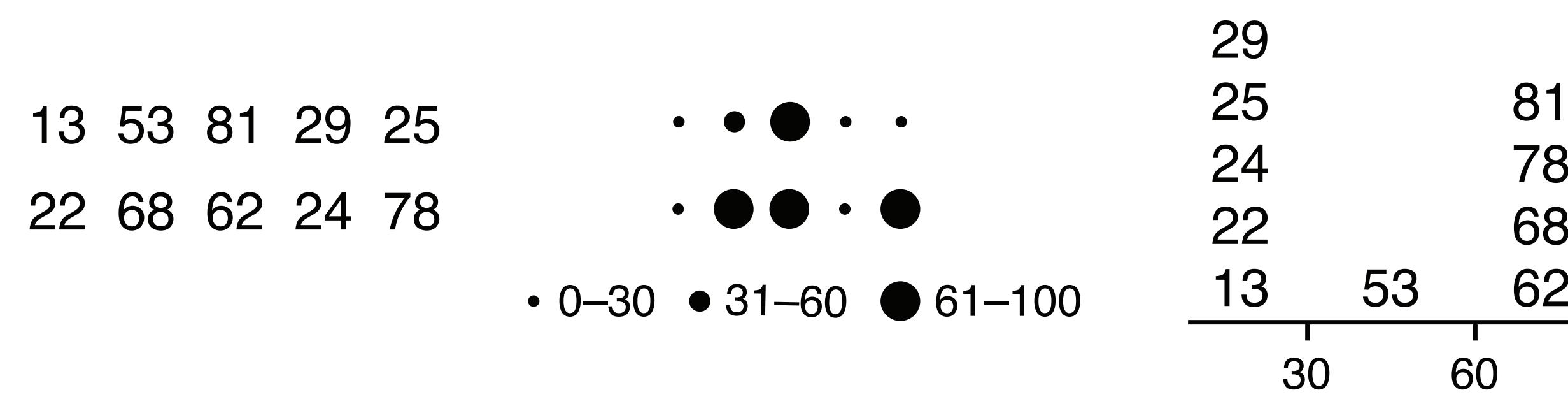
13 53 81 29 25
22 68 62 24 78

Focus on the Message: Reduce Data Detail

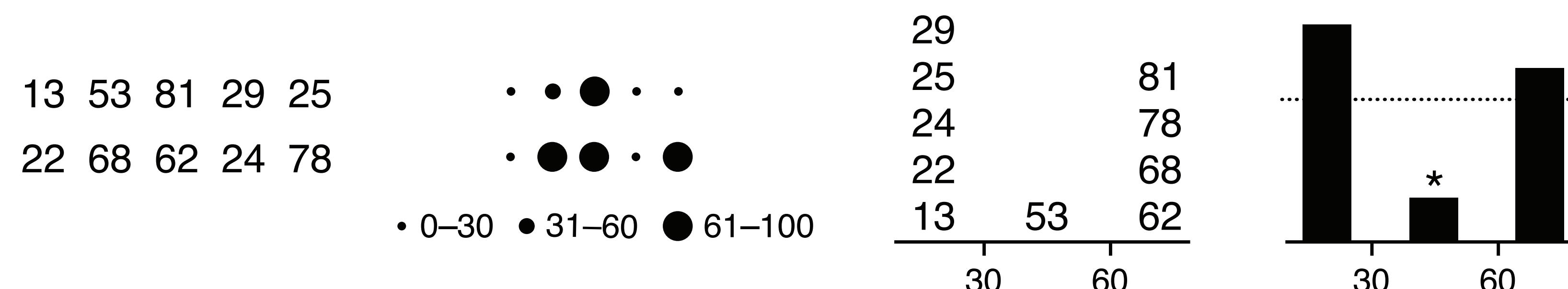


Focus

Focus on the Message: Reduce Data Detail



Focus on the Message: Reduce Data Detail



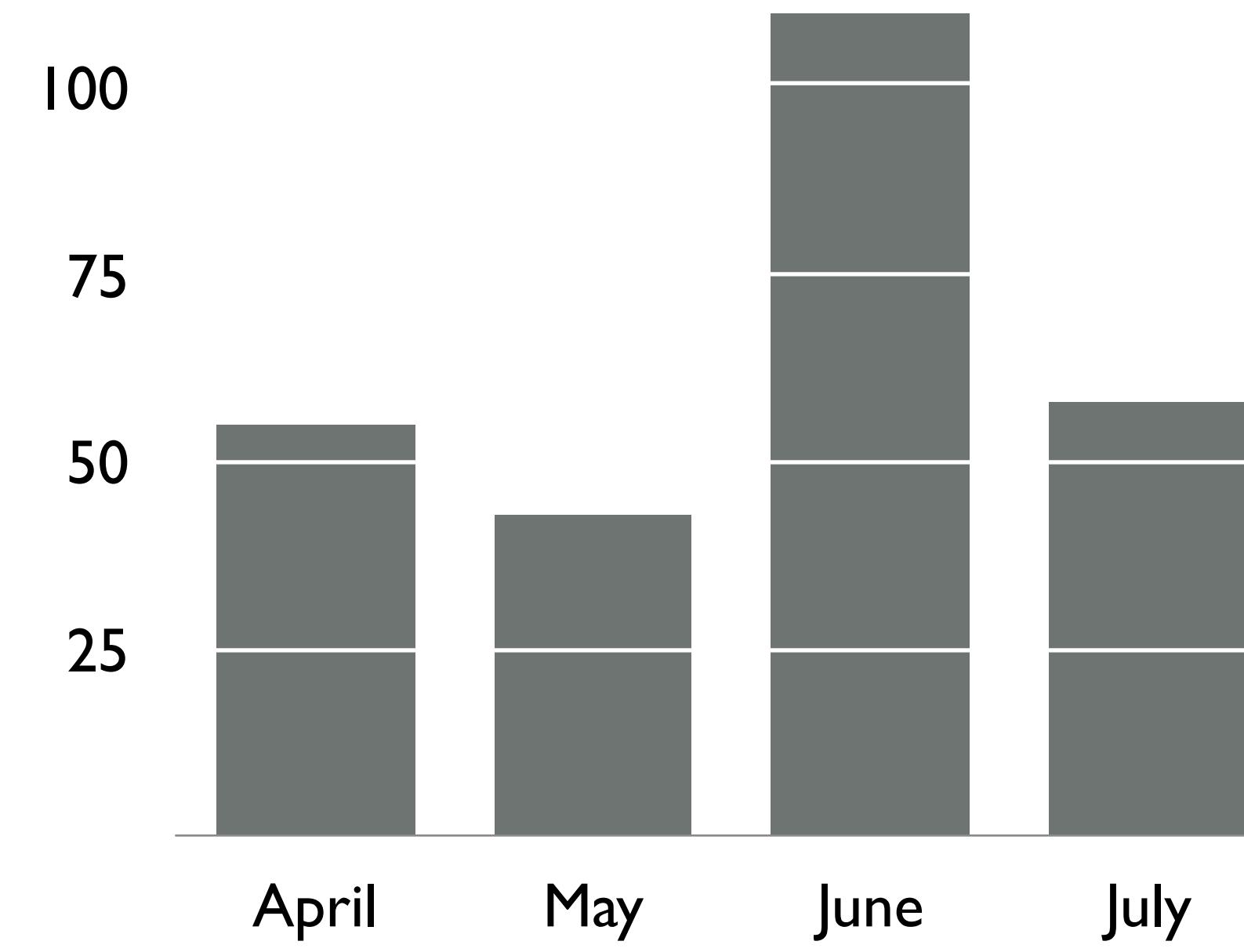
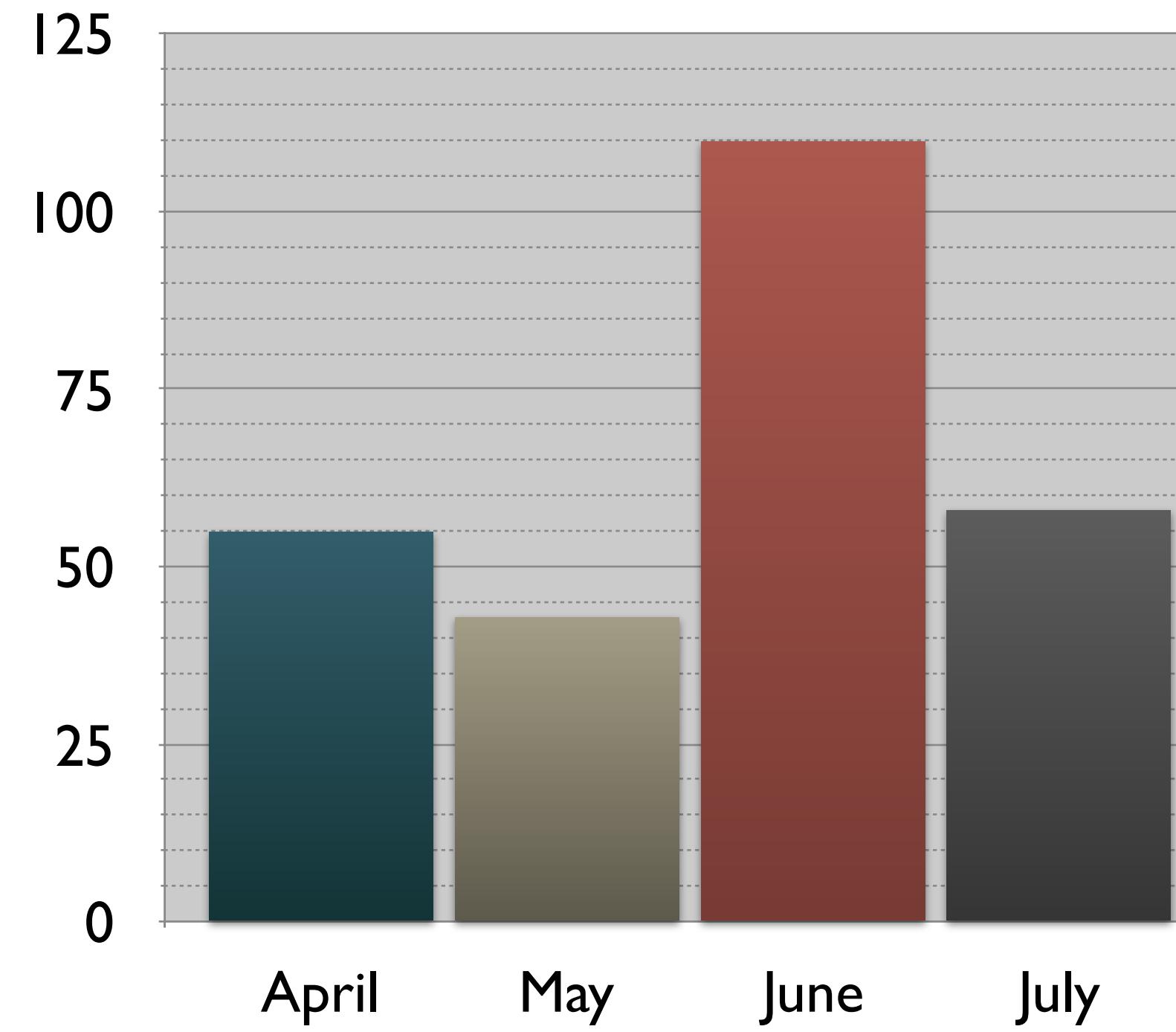
Focus

Use Layout to Convey Meaning



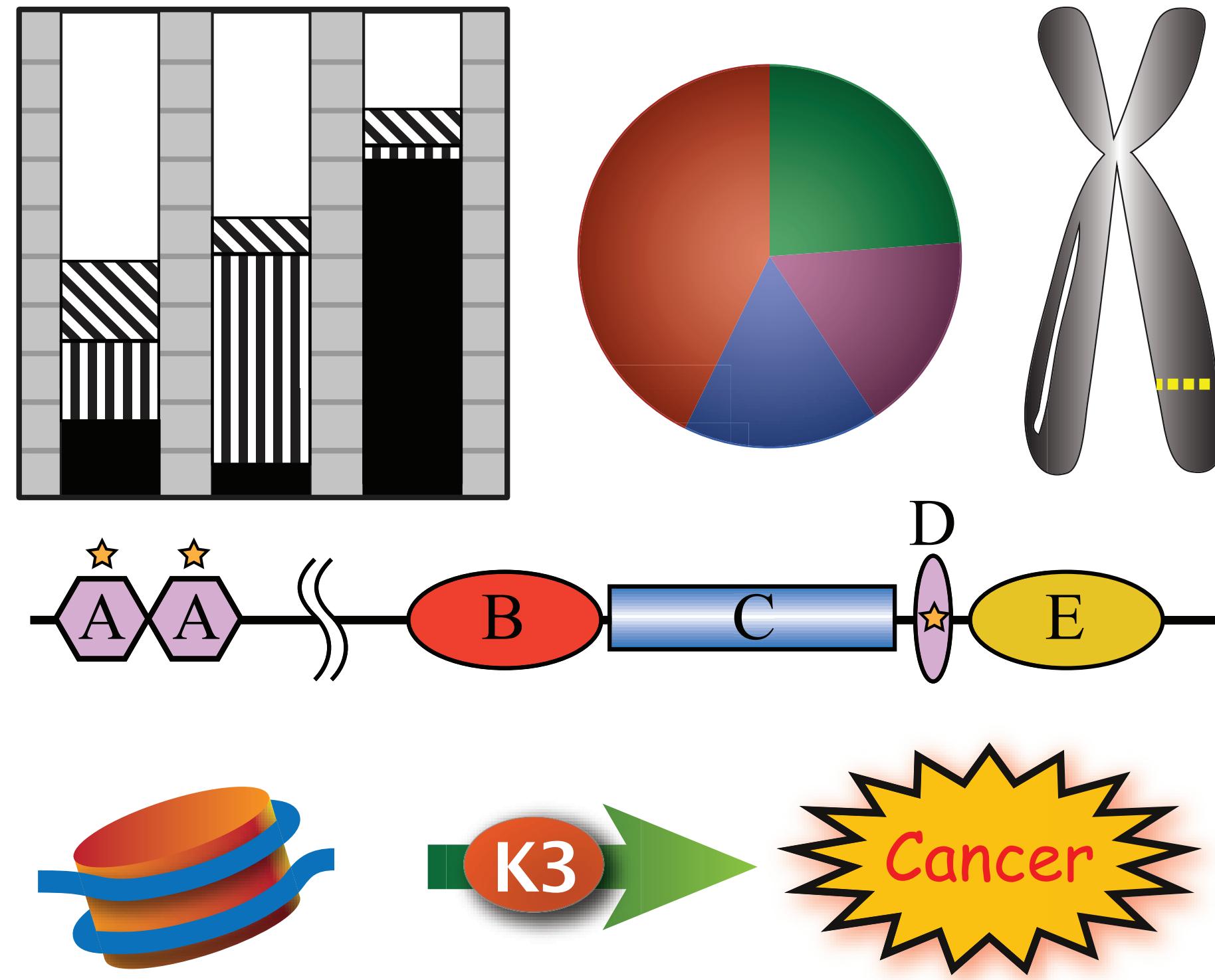
after Wong, Nature Methods, 2011

Focus on the Data: Maximize Data/Ink Ratio

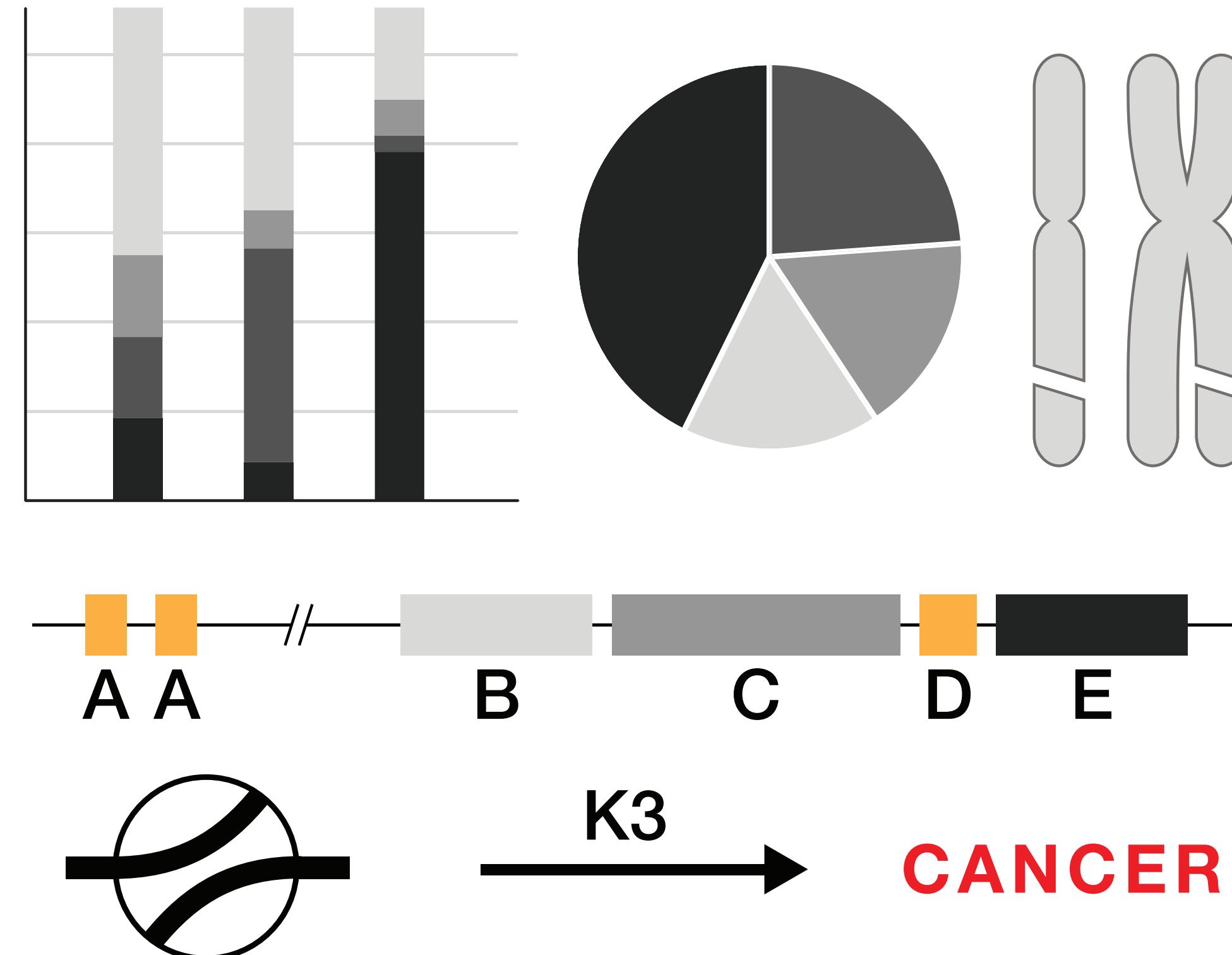


after Tufte, *The Visual Display of Quantitative Information*, 2nd Ed., 2001

Don't shout at your audience!

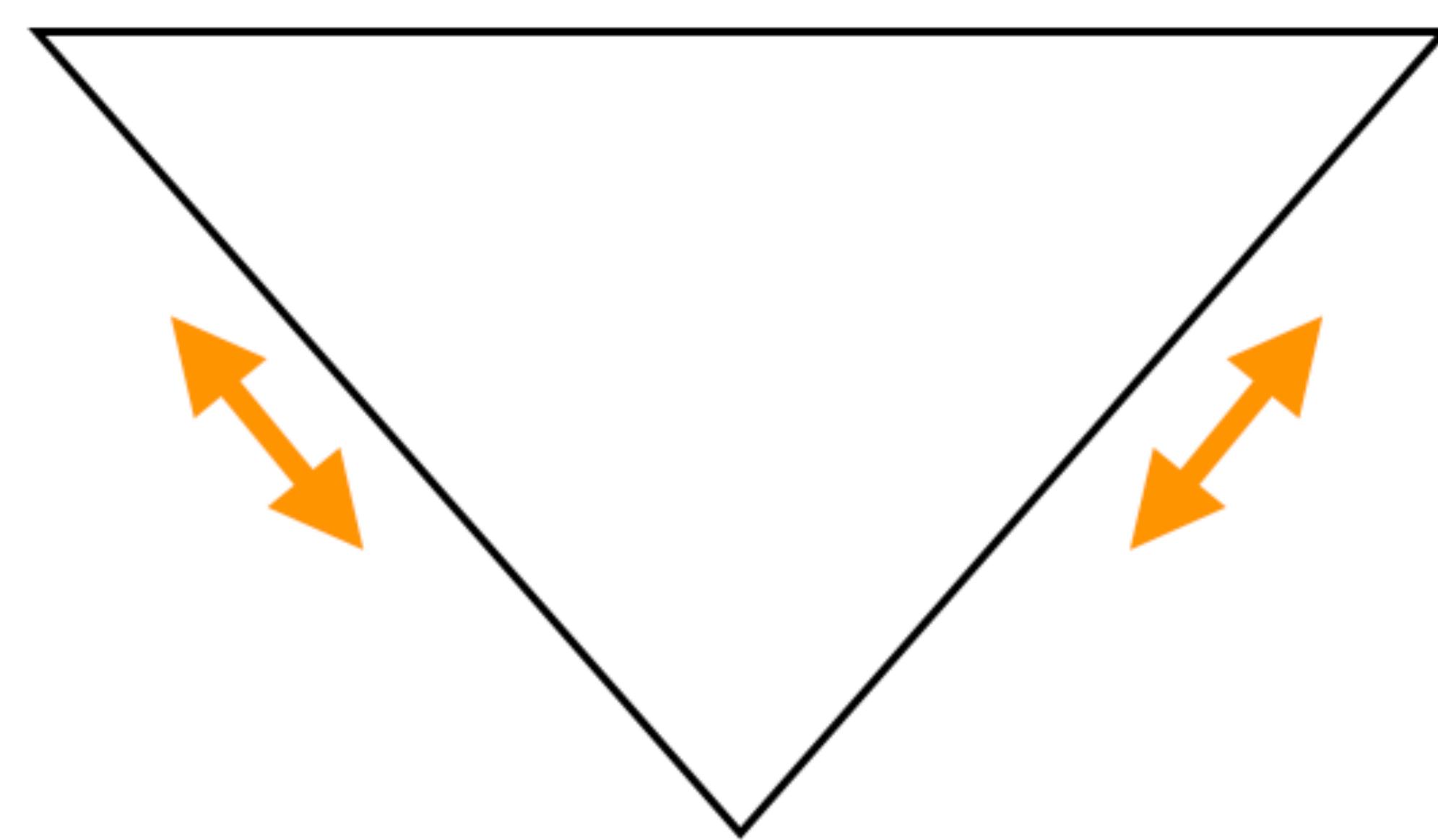


Don't shout at your audience!



dialogue between computer & analyst

Confirmation **Exploration**



Presentation

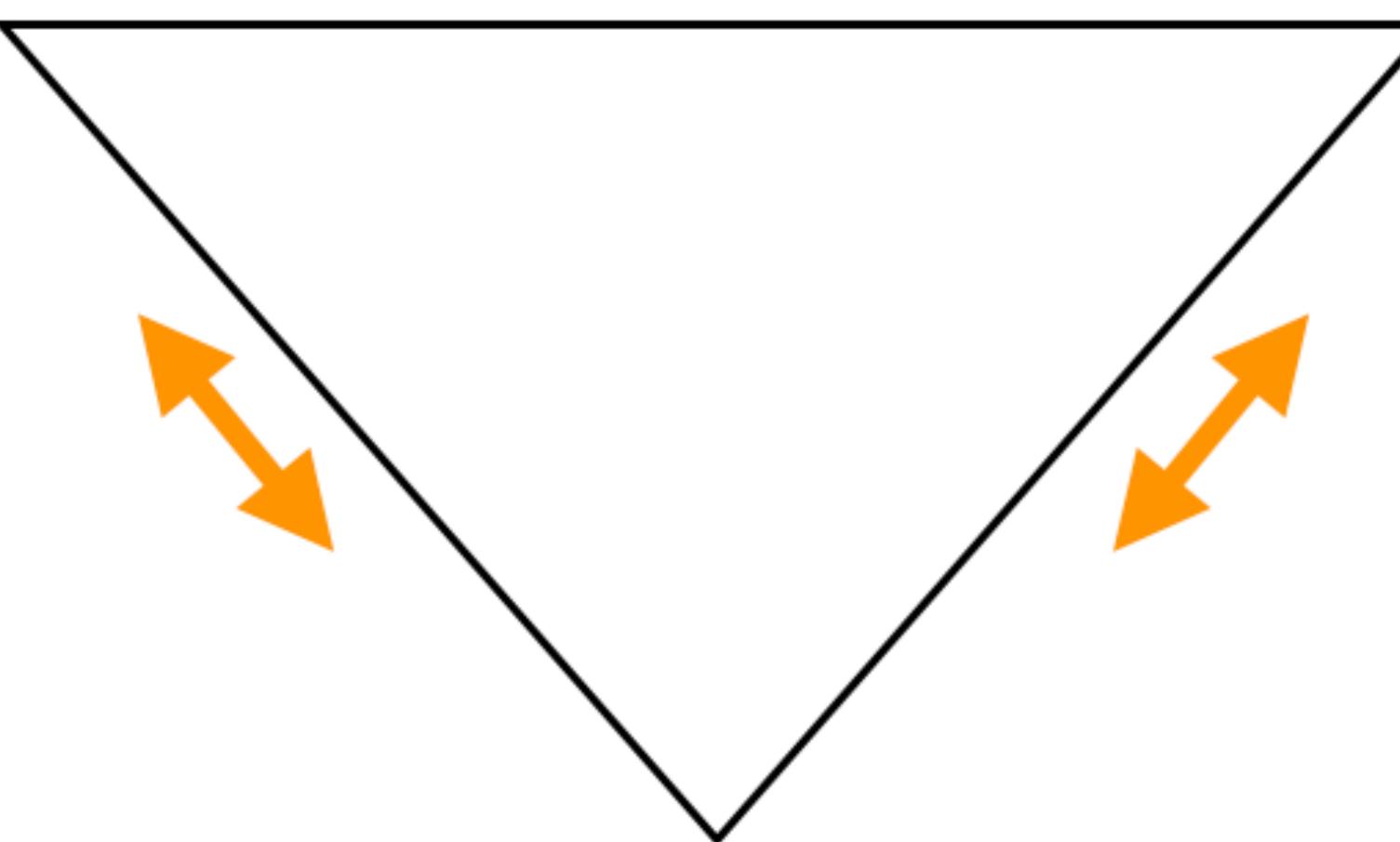
dialogue between analyst & audience

as much data as possible

dialogue between computer & analyst

Confirmation

Exploration

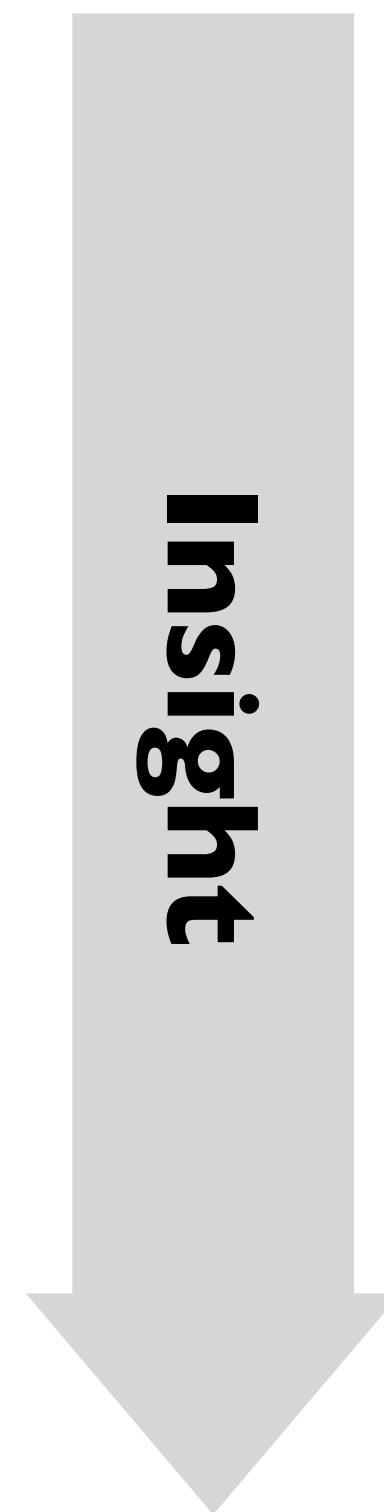


Presentation

dialogue between analyst & audience

as much data as necessary

Insight



Single View Interactions

Manipulate

⌚ Change over Time



🔍 Select



🧭 Navigate

→ Item Reduction

→ Zoom
Geometric or Semantic



→ Pan/Translate



→ Constrained



→ Attribute Reduction

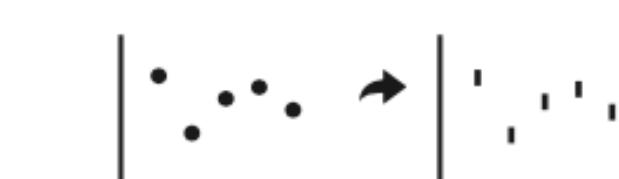
→ Slice



→ Cut

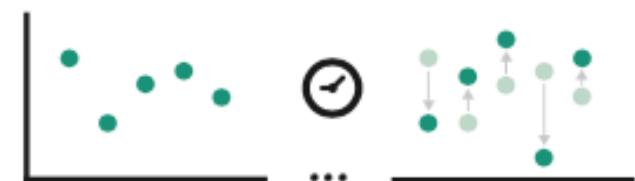


→ Project



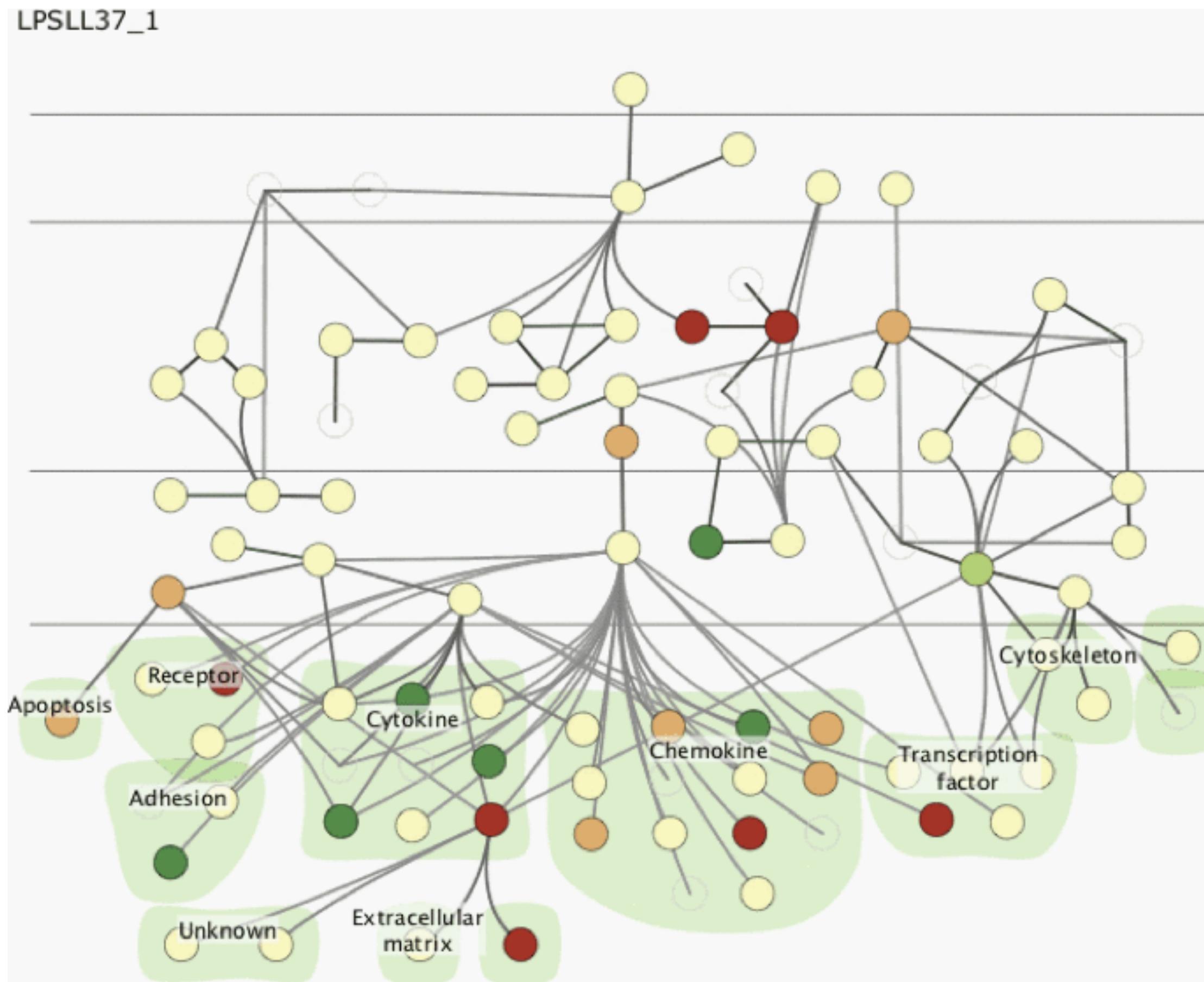
Animation

⌚ Change over Time



Animation Pitfall

Global comparisons are difficult

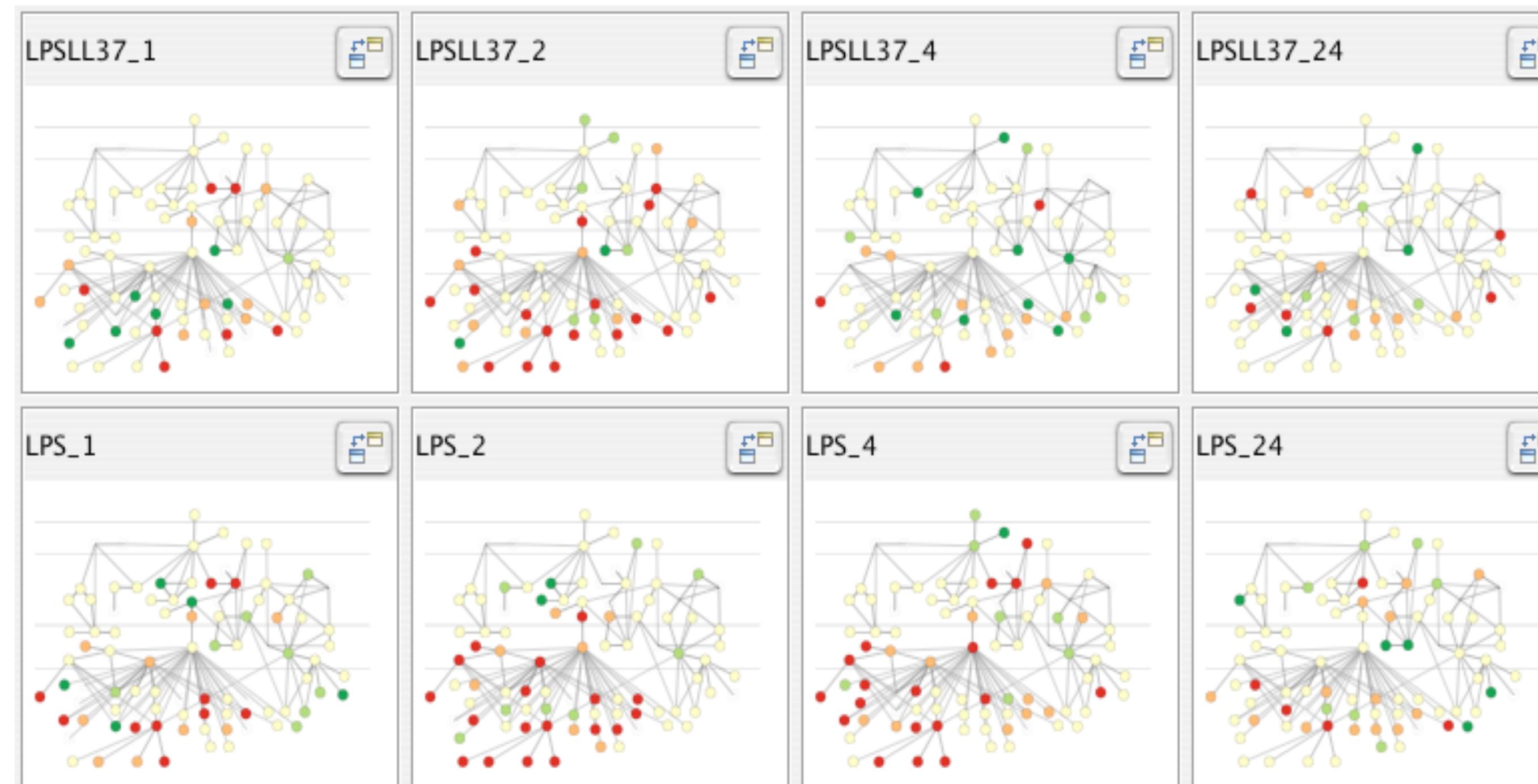
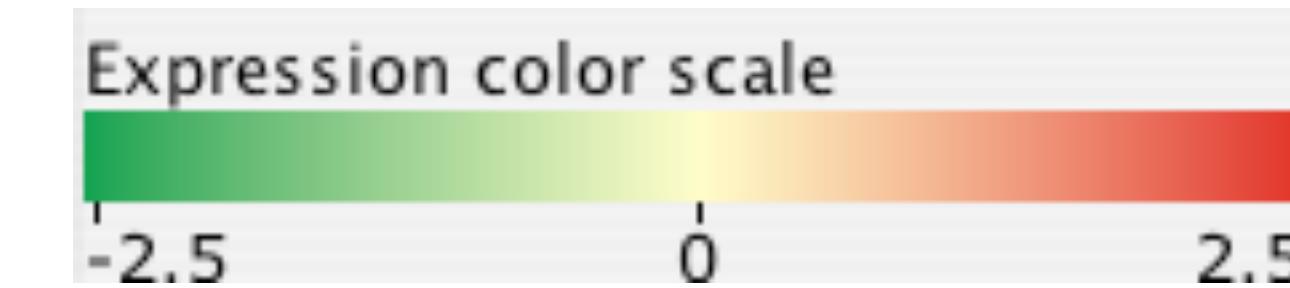


Animation Pitfall

Small Multiples

one view per state

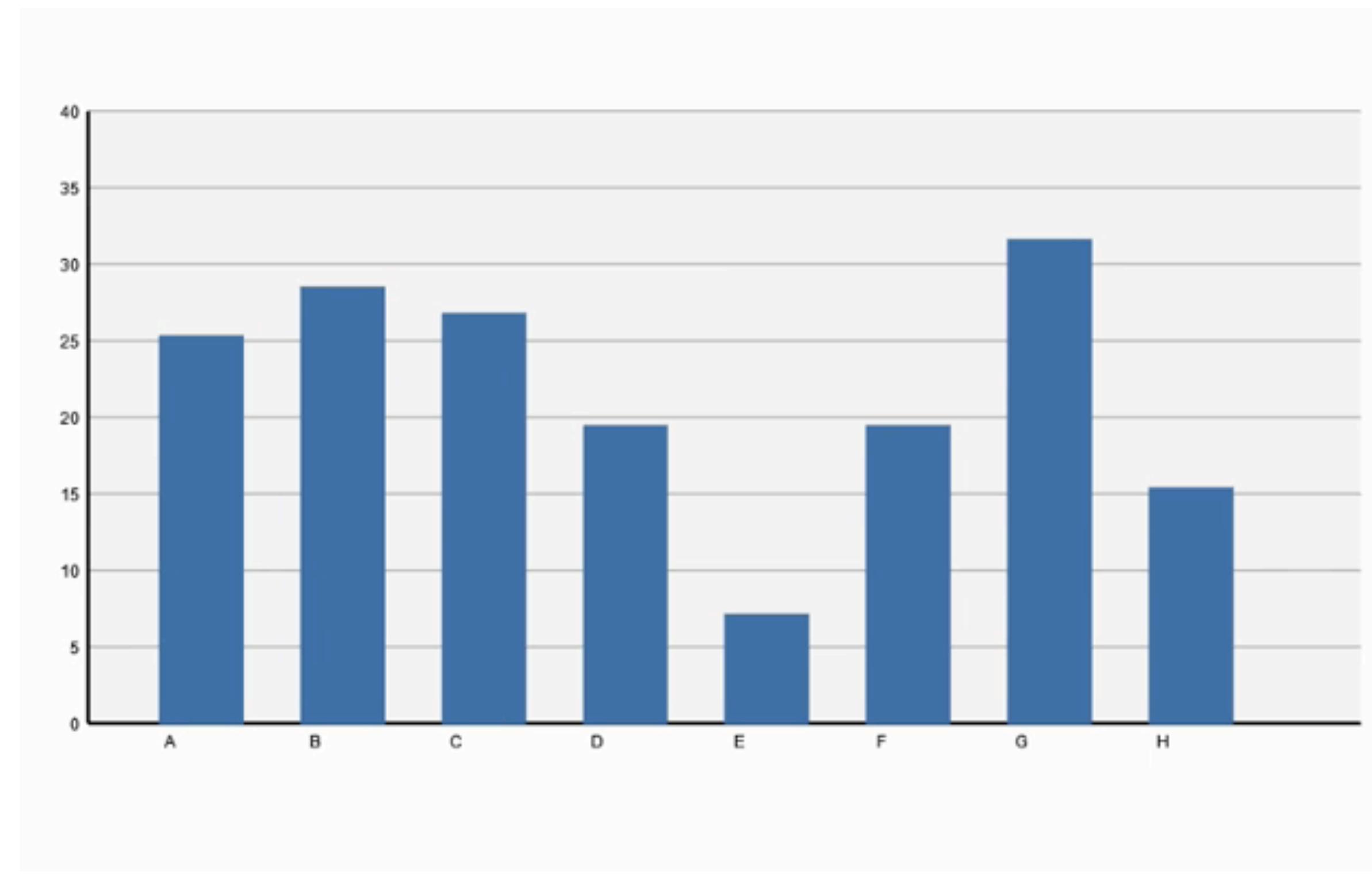
show time with space



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



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

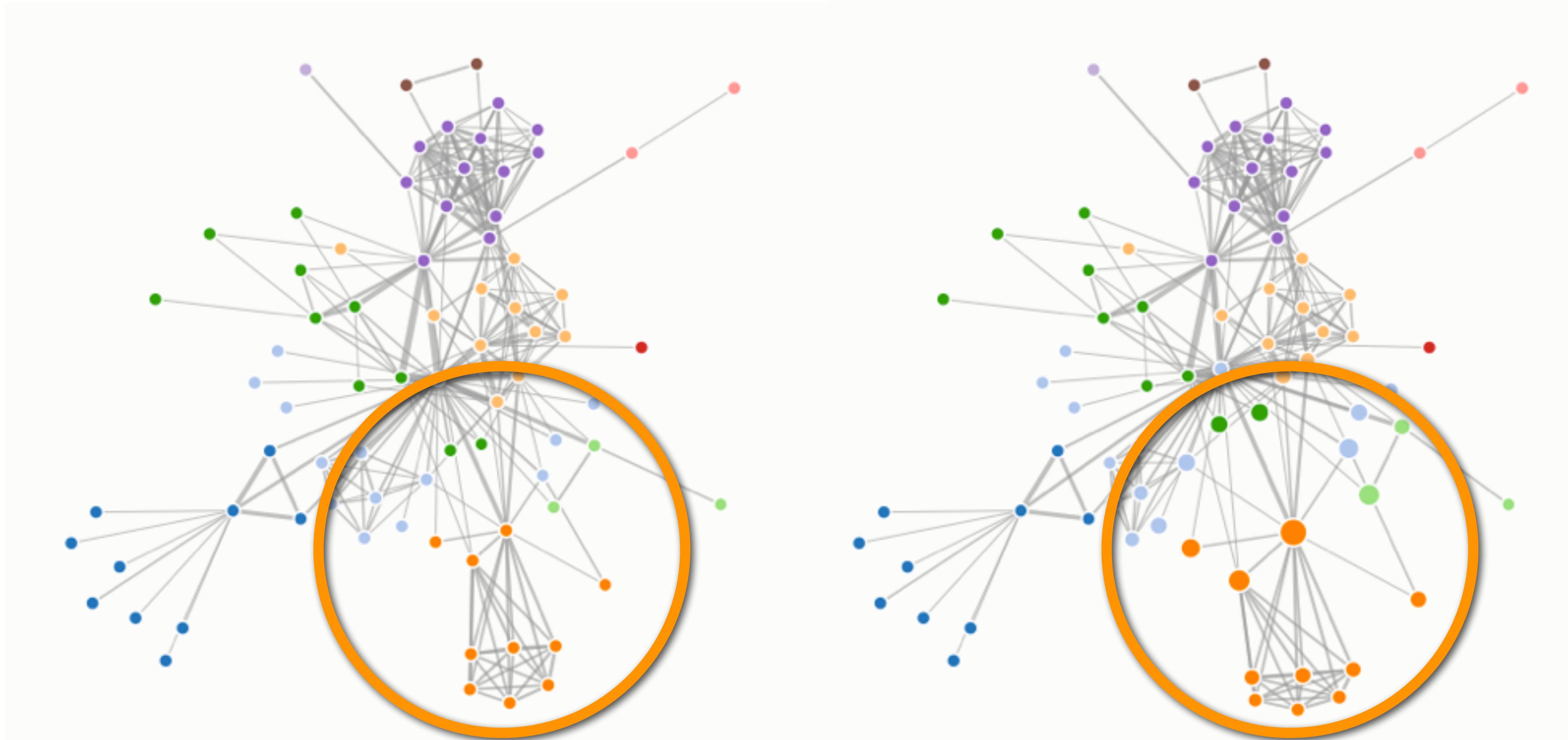
<https://bl.ocks.org/mbostock/db6b4335bf1662b413e7968910104f0f>

Semantic Zoom

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

Interaction Techniques: Focus + Context

Fisheye Distortion



Fisheye and Rectilinear Cartesian Distortion in D3

<https://bostocks.org/mike/fisheye/>

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

Multi-View Interactions

Facet

④ Juxtapose and Coordinate Multiple Side-by-Side Views

→ Share Encoding: Same/Different

→ *Linked Highlighting*



→ Share Data: All/Subset/None



→ Share Navigation

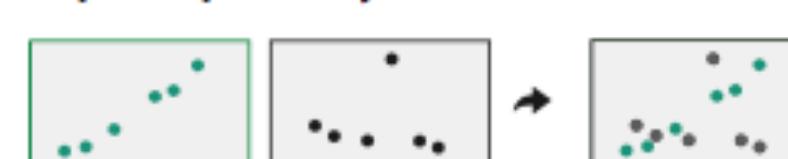


		Data		
		All	Subset	None
Encoding	Same	Redundant		Overview/Detail
	Different			Multiform, Overview/Detail

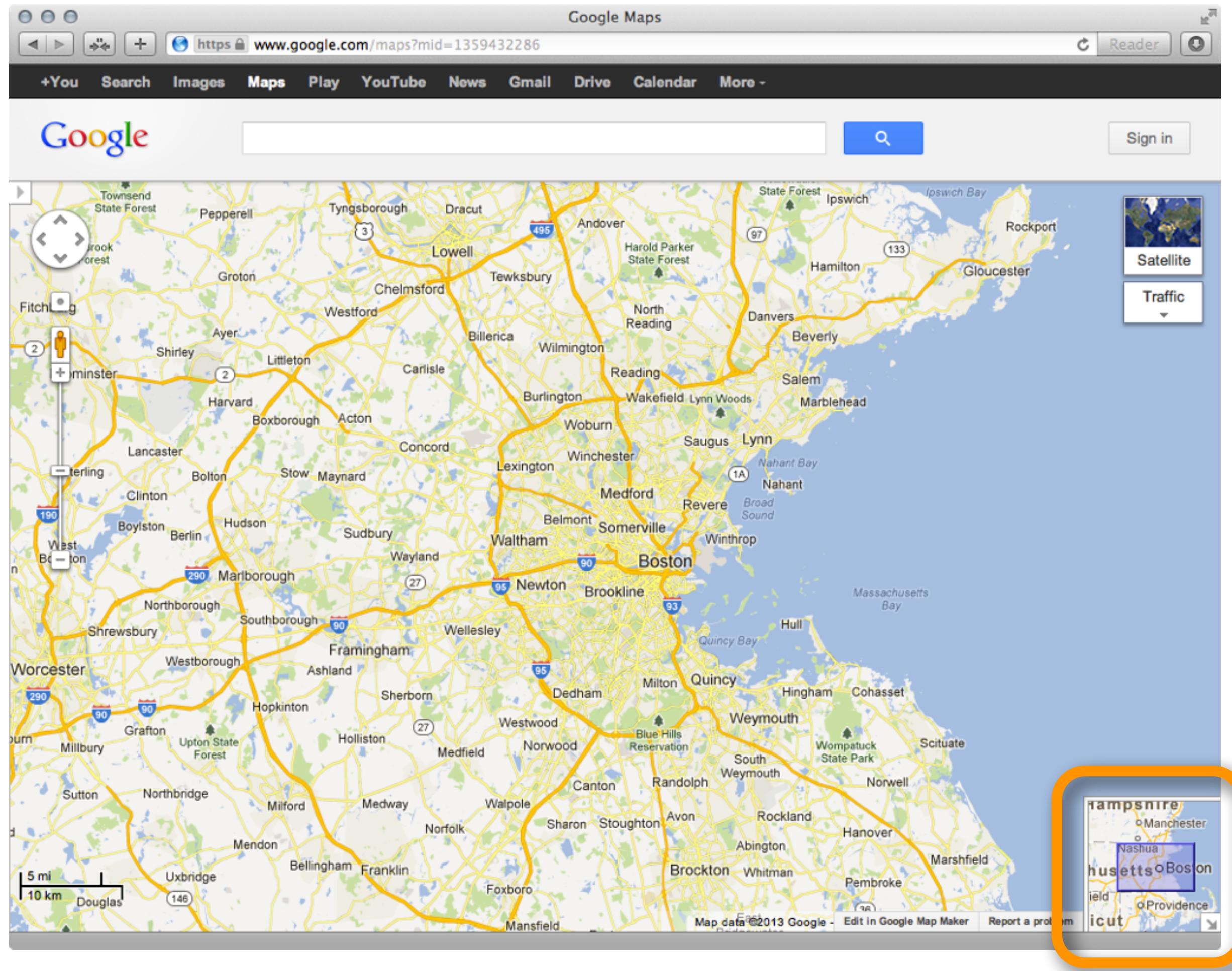
④ Partition into Side-by-Side Views



④ Superimpose Layers



Interaction Techniques: Overview + Detail



Then ...

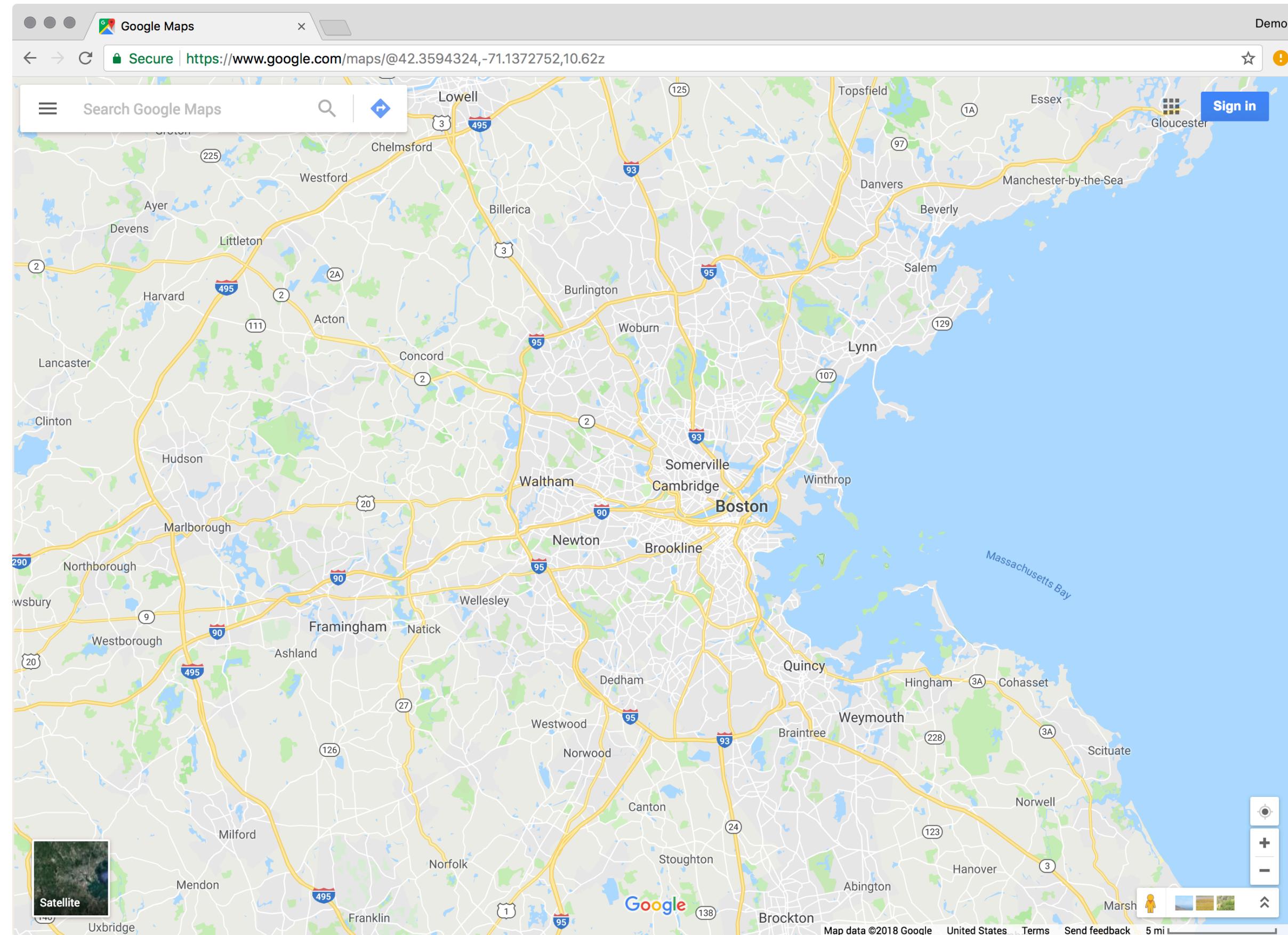
Brush and zoom in D3

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

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

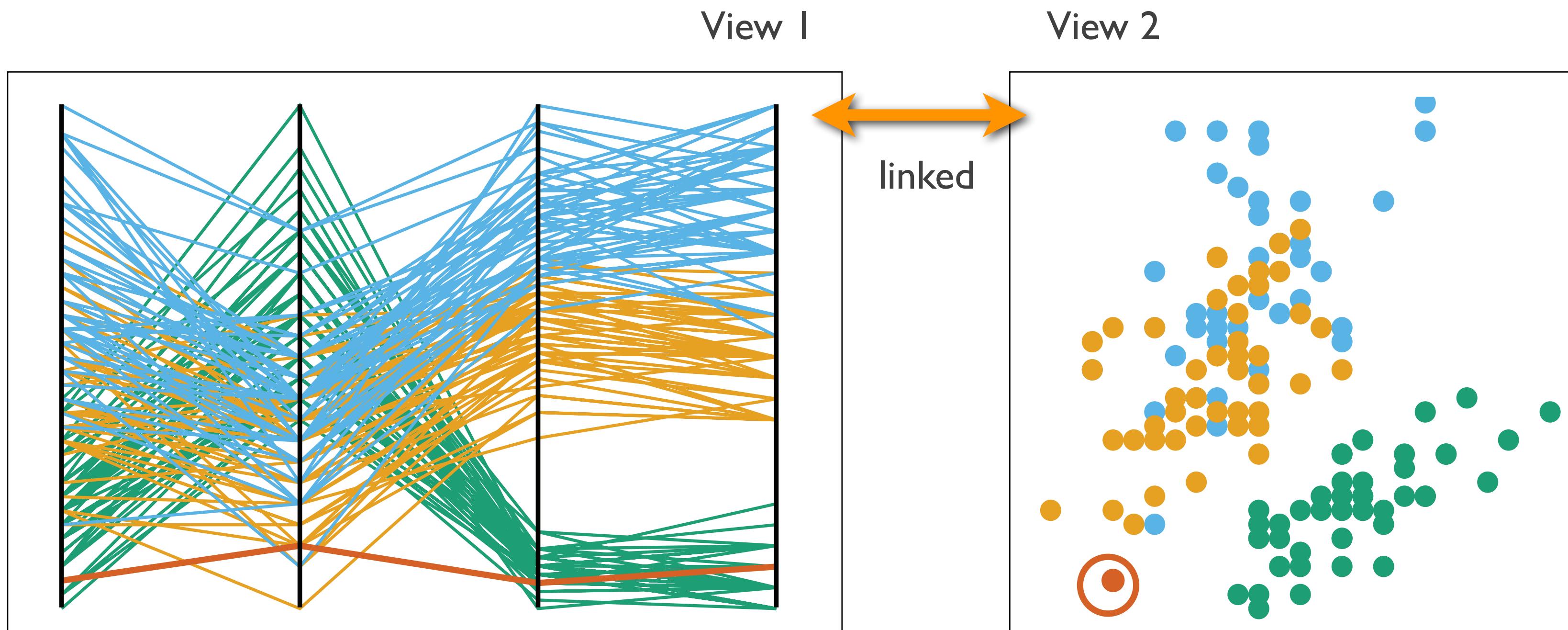
<https://maps.google.com>

Interaction Techniques: Overview + Detail



... & today!

Interaction Techniques: Brushing + Linking



Brushing and Linking

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

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

Altair

- **Website**

<https://altair-viz.github.io>

- **Notebooks**

<https://github.com/hms-dbmi/decart-2019-data-visualization/>