

# Storytelling with Data

## Module 8: Effective visuals – encoding, grammar, and layers

**Scott Spencer**

Faculty and Lecturer  
Columbia University



# **Unanswered, or new, questions from discussion?**

# Agenda

Timeline for upcoming deliverables

Today's objectives

Communicating data with color

The grammar of graphics

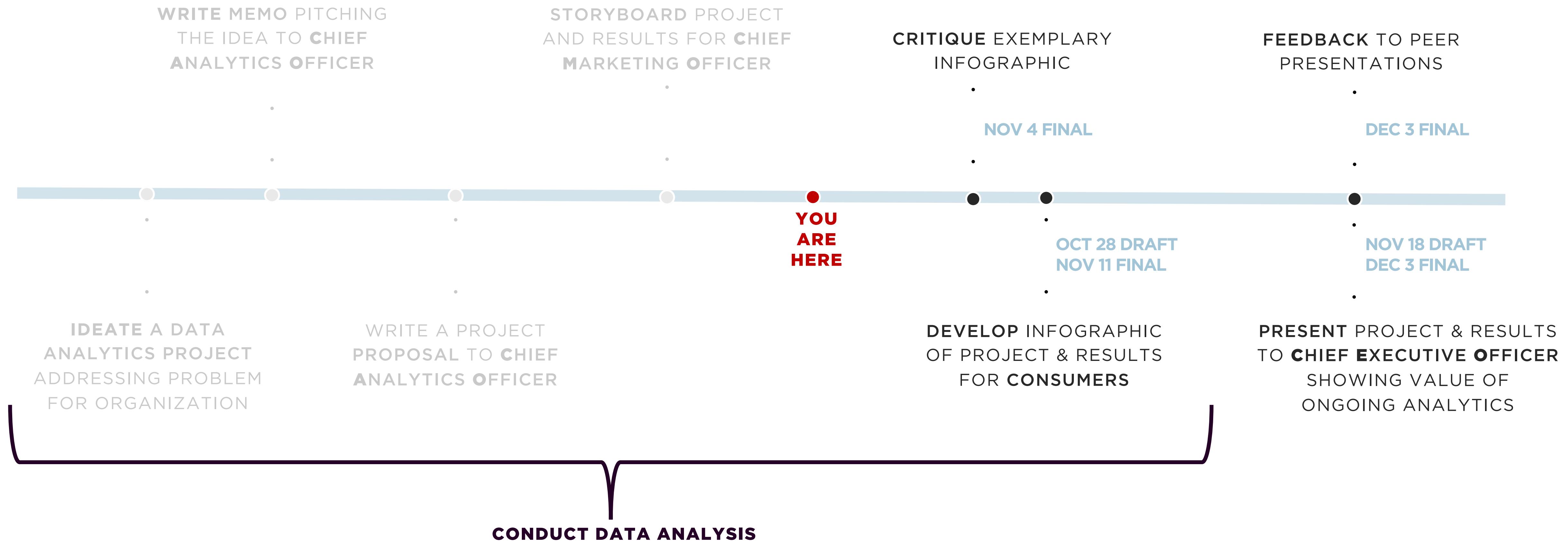
From graphics back to narrative

Complexity from common graphics

# Our project timeline

# Upcoming deliverables

**Information graphic** – reframe your story, this time building off the messages you built for the marketing team in order to craft an infographic that displays the results of the analytic work in a way that is accessible, engaging, and exciting for a **general or consumer audience**.



# Today's Objectives

# Objectives

1

Creating a visual narrative  
from common graphics

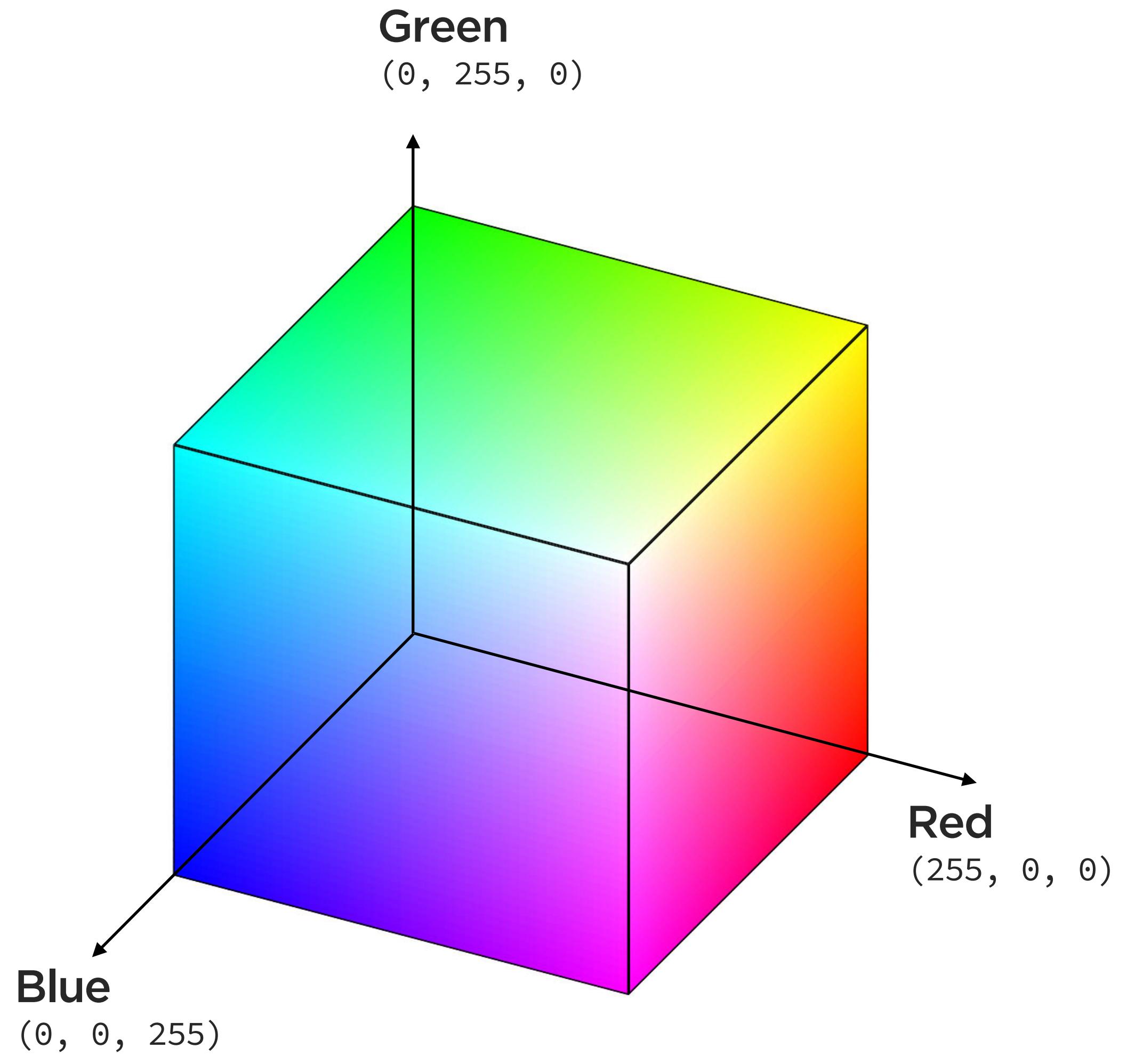
2

Consider the role of  
layering and hierarchy in  
graphic narrative

3

Building complexity from  
common graphics, examples

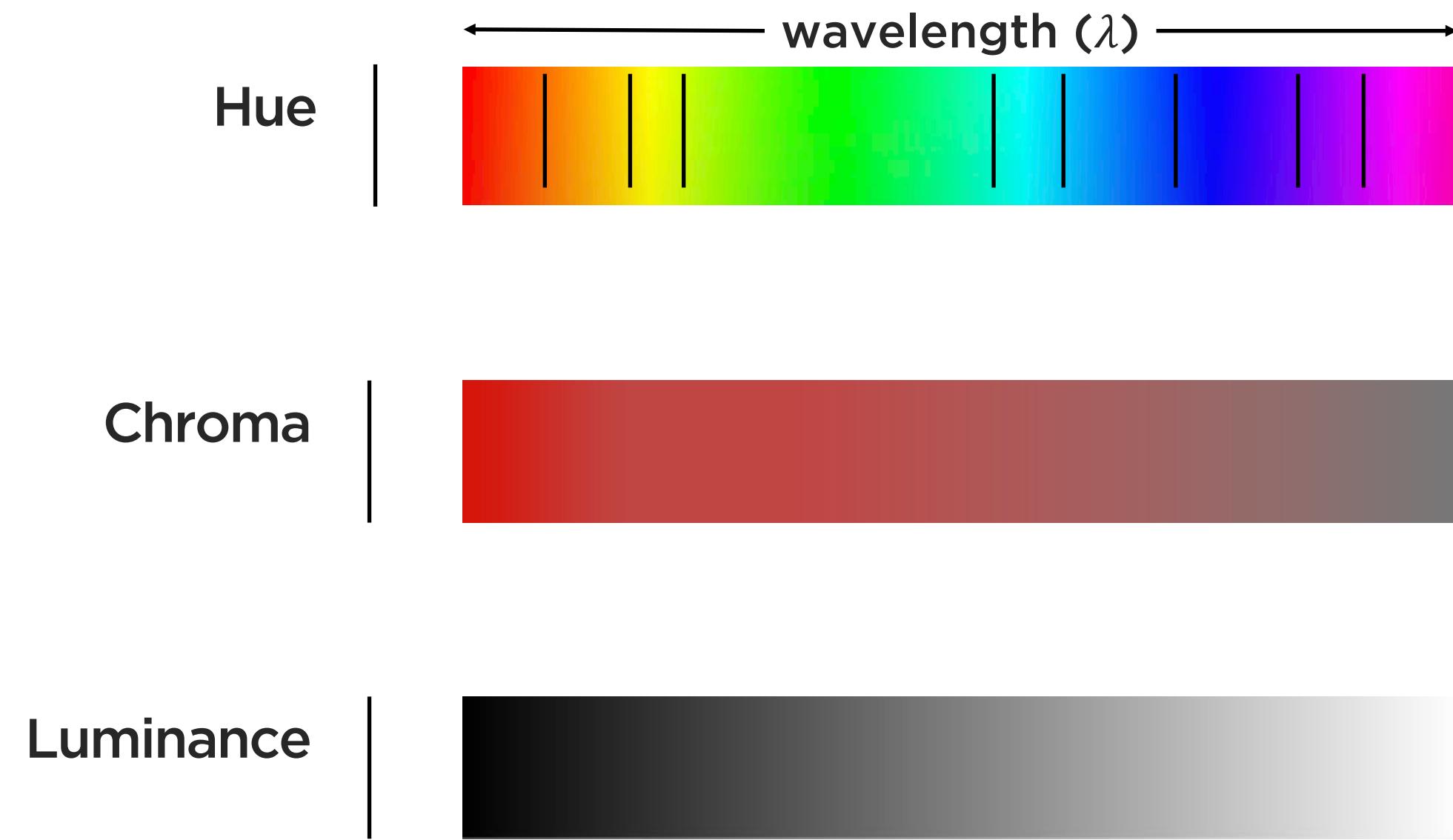
# Communicating data with color: perception of values



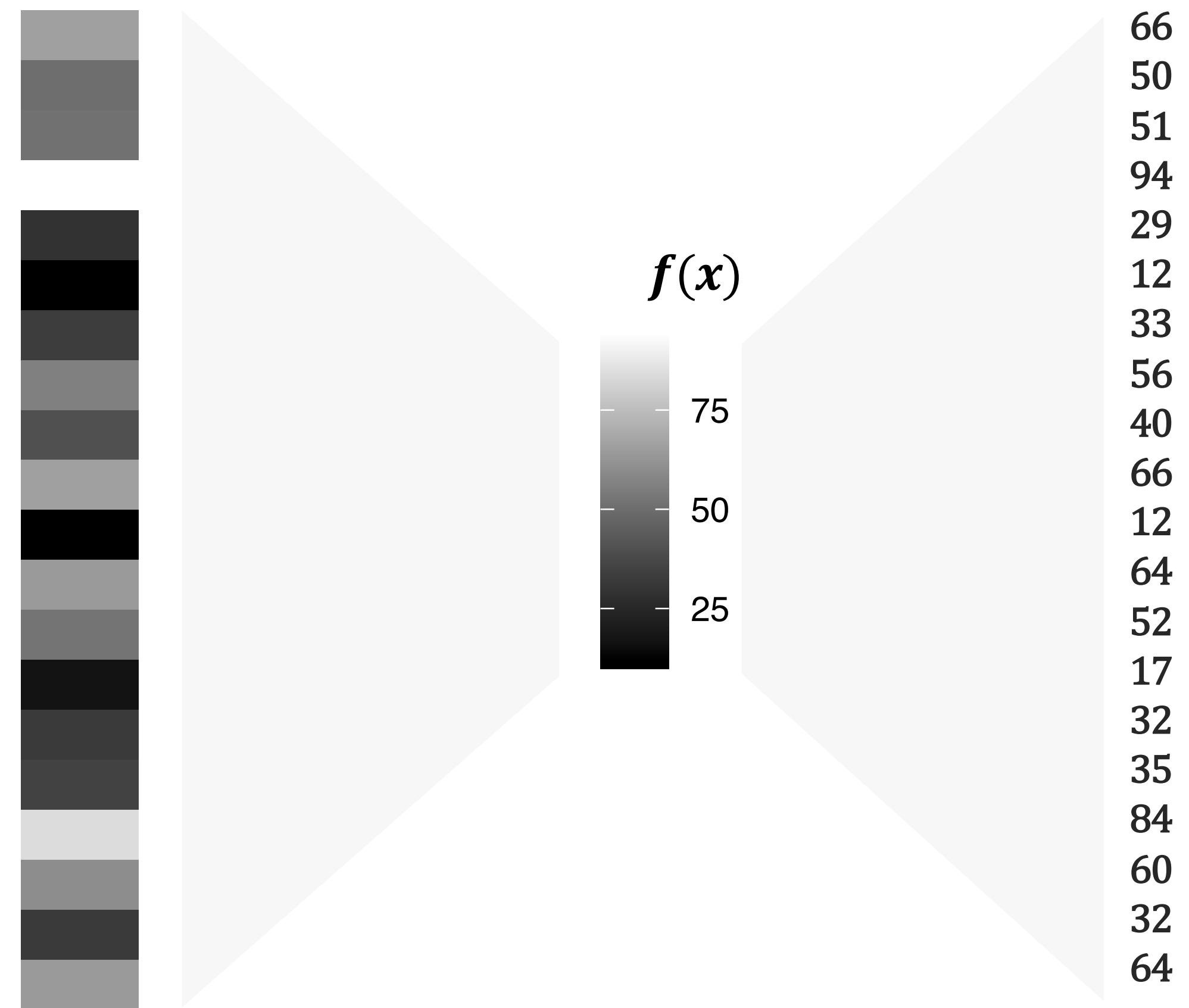
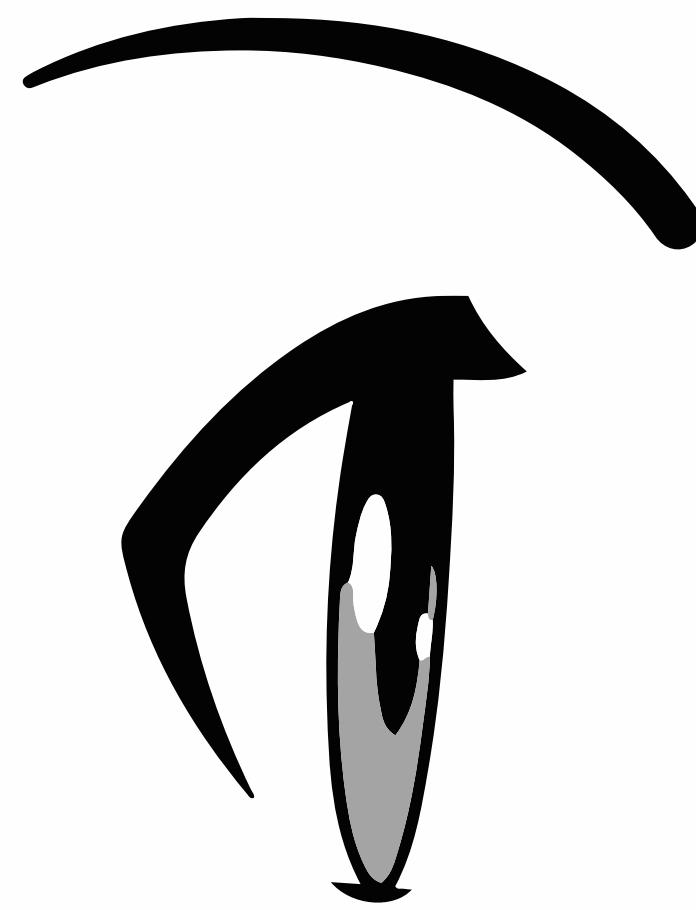
We can encode data using *color spaces*, which are mathematical models.

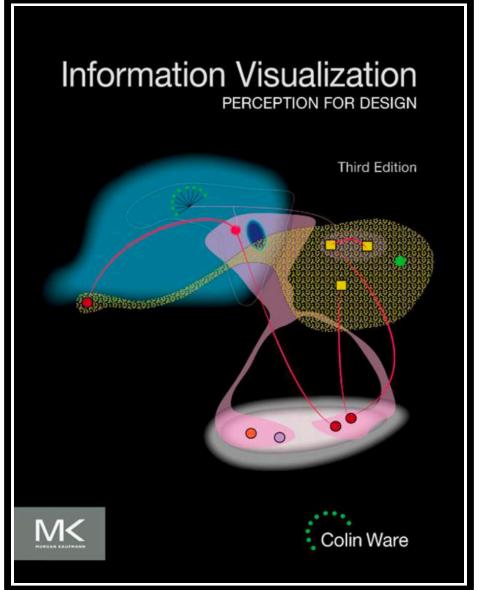
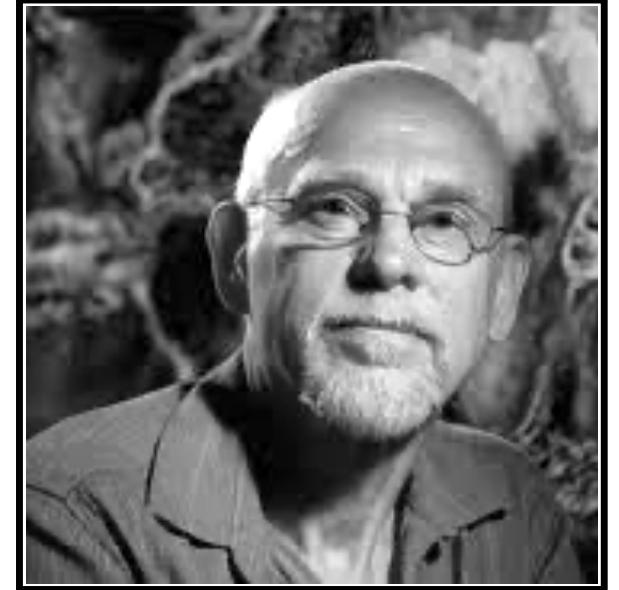
The common color model **RGB** has three dimensions — **red**, **green**, and **blue** — where those hues are mixed to produce a specific color.

Notice this color space, described below as hue, chroma, and luminance, has **uneven distances** and brightness between perceived hues.



How can we map data to light, whether using its hue, saturation, or luminance?





# Information Visualization perception for design

## Ware

He is the Director of the Data Visualization Research Lab which is part of the Center for Coastal and Ocean Mapping at the University of New Hampshire. He is cross appointed between the Departments of Ocean Engineering and Computer Science.

### Luminance and brightness

**Luminance** is the measured amount of light coming from some region of space.

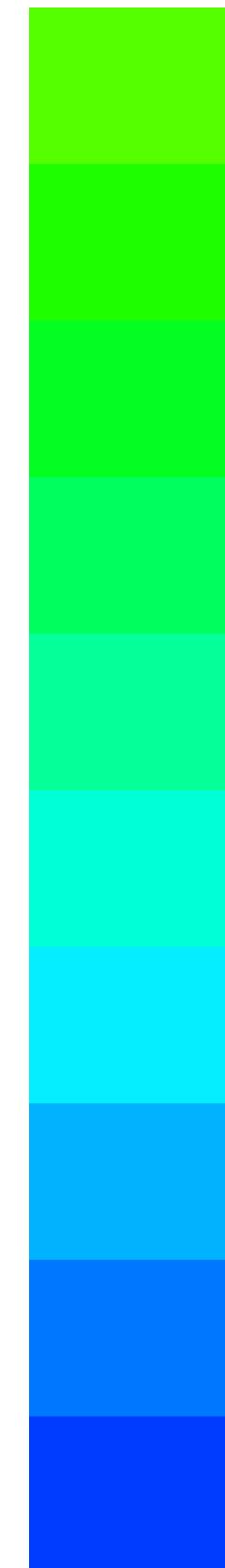
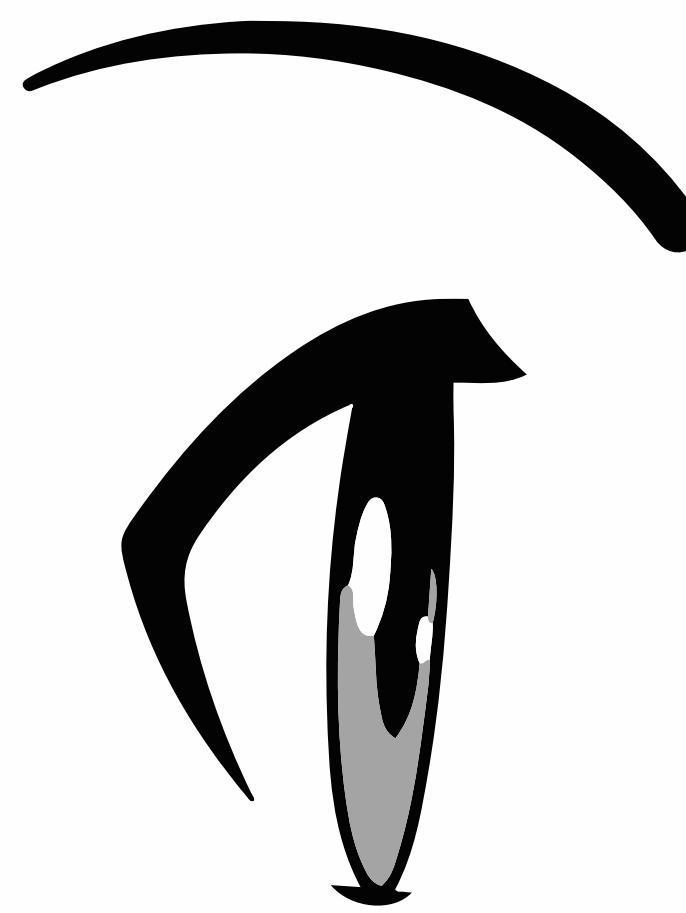
**Brightness** is the perceived amount of light coming from that region of space.

**Brightness perception is a nonlinear function of luminance.**

Perceived brightness is a very nonlinear function of the amount of light emitted. That function follows the power law:

$$\text{Perceived brightness} = \text{Luminance}^n$$

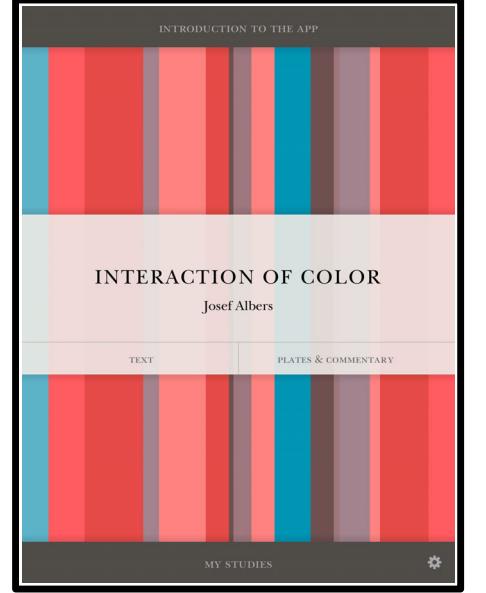
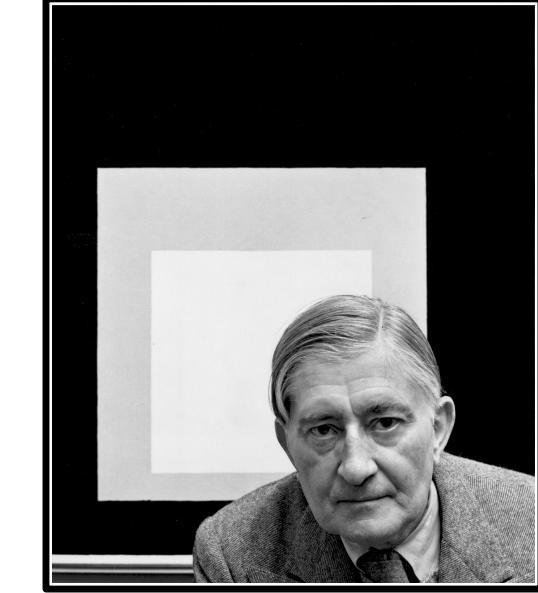
Where the **value of  $n$  depends on the size of the patch of light**. For circular patches of light subtending 5 degrees of visual angle,  $n$  is 0.333, whereas for point sources of light  $n$  is close to 0.5.



**As with luminance, hue values in the RGB color space fail to uniformly scale across values.**

**Uneven wavelengths** between what we perceive as colors, as we saw in the RGB color space, results in, for example, almost identical hues of green across a range of its values while our perception of blues change more rapidly across the same change in values.

We also perceive a lot of variation in the lightness of the colors here, with the cyan colors in the middle looking brighter than the blue colors.



# Interaction of Color

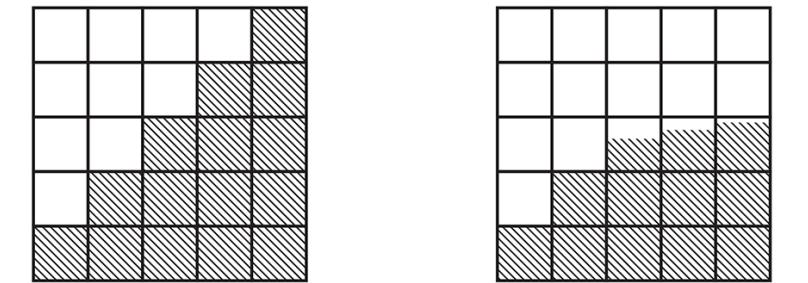
## Albers

Josef Albers was a German-born American artist and educator whose work, both in Europe and in the United States, formed the basis of modern art education programs of the twentieth century.

### Color intervals

Color intervals are the distance in light intensity between one color and another, analogous to musical intervals (the relationship between notes of different pitches).

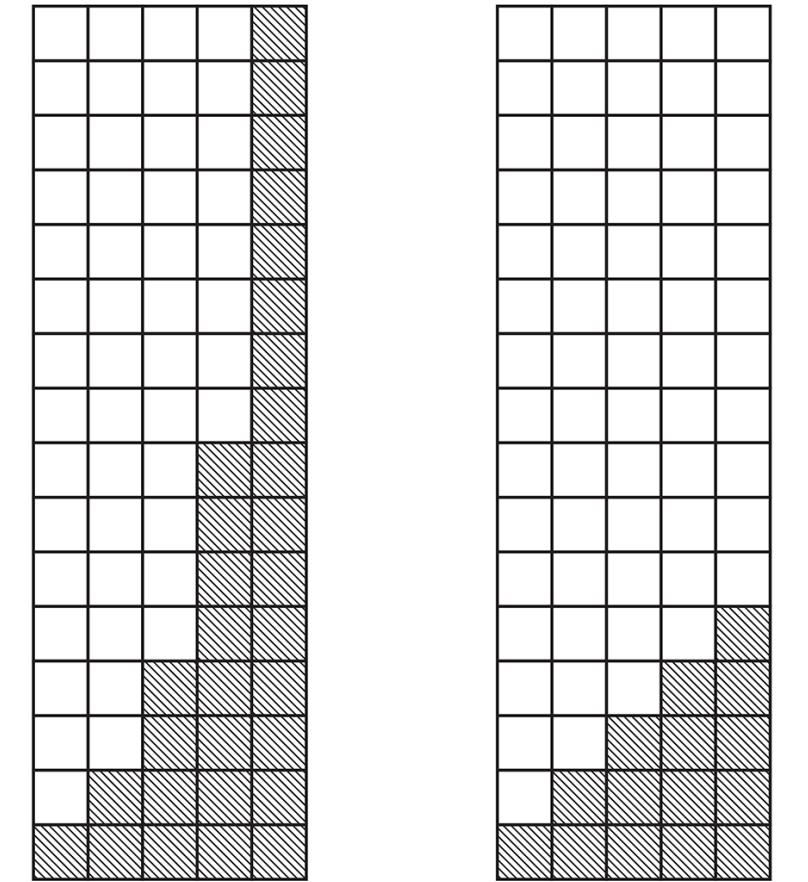
Weber's law,  
applied to creating  
steps we perceive  
as evenly spaced



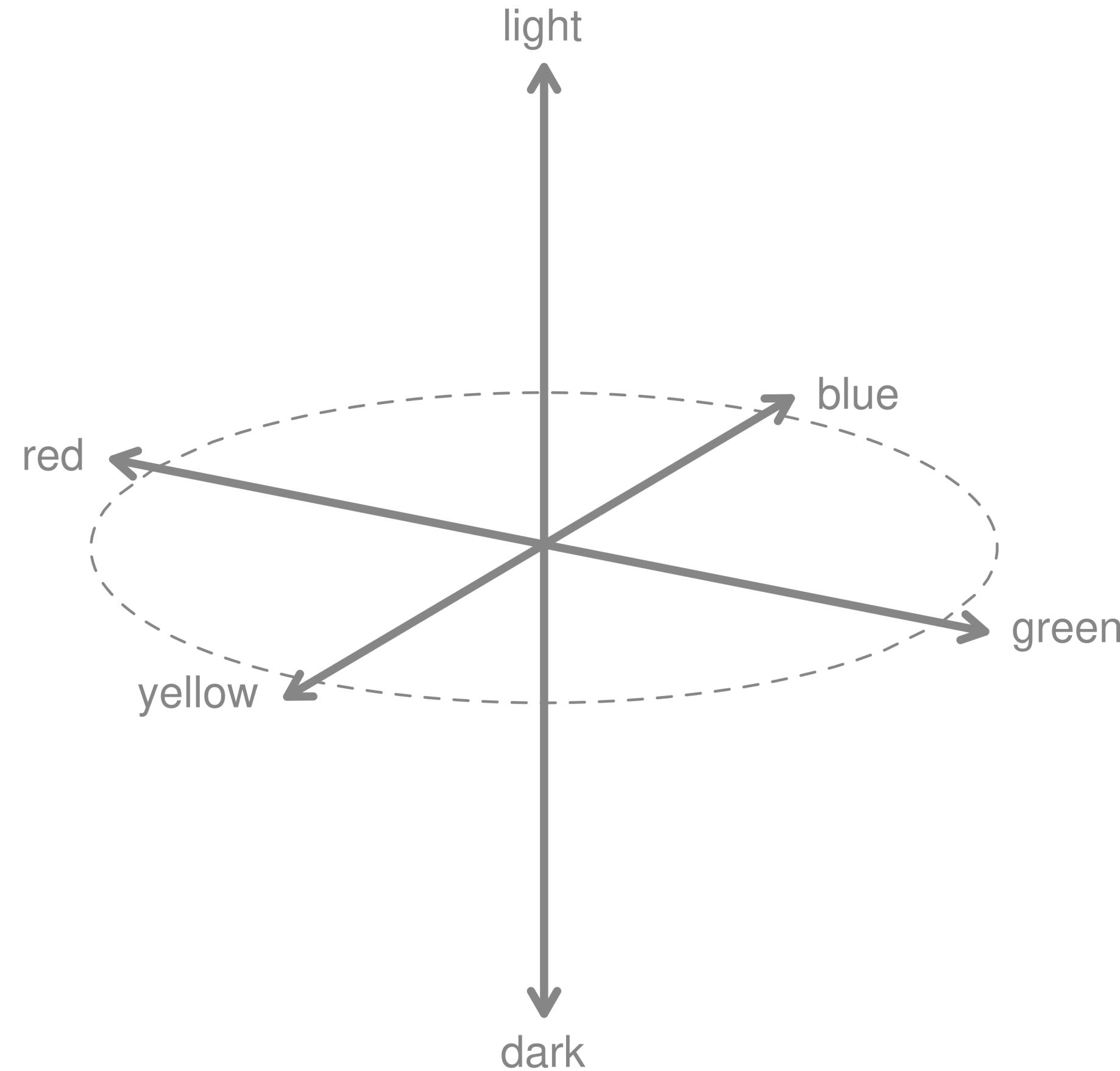
THIS PHYSICAL FACT      REDUCES TO      THIS PSYCHOLOGICAL EFFECT

The visual perception of an arithmetical progression depends upon a physical geometric progression.

In a **simplification** shown to the right, this means: if the first 2 steps measure 1 and 2 units in rise, then step 3 is not only 1 unit more (that is, 3 in an arithmetical proportion), but is twice as much (that is, 4 in a geometric proportion). The successive steps then measure 8, 16, 32, 64 units.



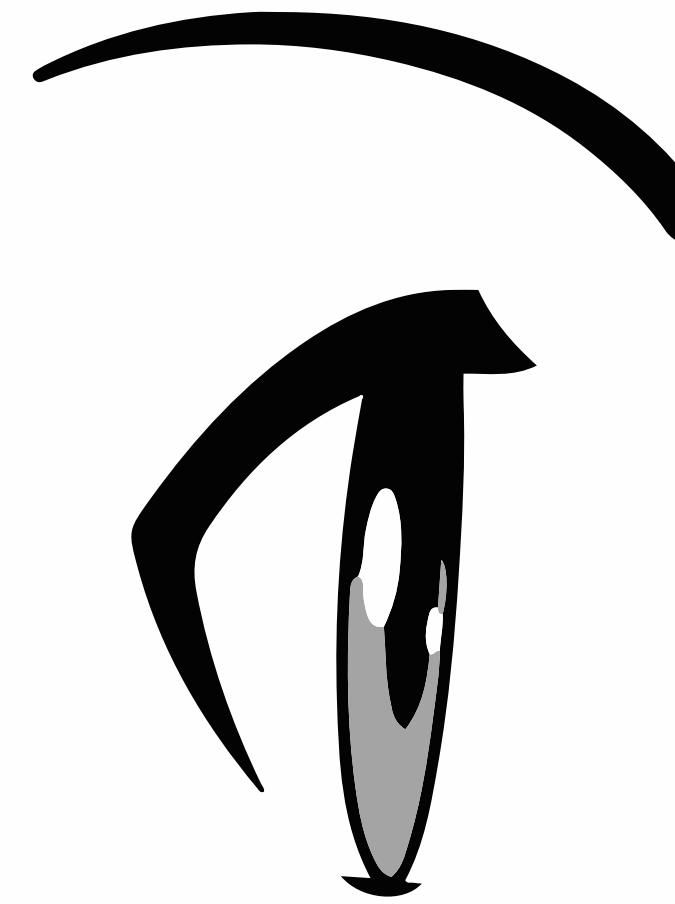
THIS PHYSICAL FACT      PRODUCES      THIS PSYCHOLOGICAL EFFECT



## Other color spaces describe color as we see it.

**Humans** compute color signals from our retina cones via an opponent process model, which makes it impossible to see reddish-green or yellowish-blue colors.

The International Commission on Illumination (CIE) **studied human perception and re-mapped color so that we perceive color changes uniformly**. Their **CIELUV** color model has two dimensions —  $u$  and  $v$  — that represent color scales from red to green and yellow to blue.

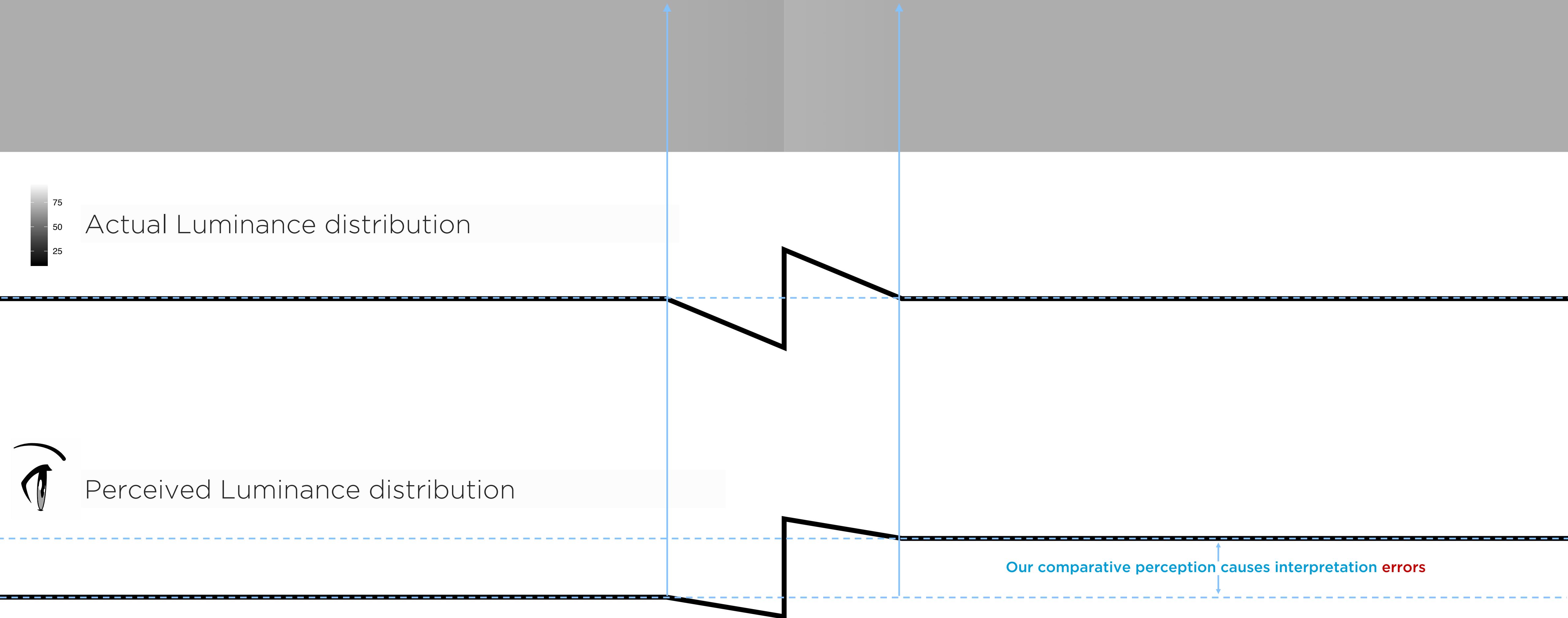


***Perceptually uniform* color spaces better represent quantity.**

More modern color spaces improve upon CIELUV by mapping colors as perceived into the familiar and intuitive **Hue-Chroma-Luminance** dimensions. To the left, our green-to-blue hues in 10 equal steps using the HCL model are now perceptually uniform. For an R implementation, see package **colorspace**.

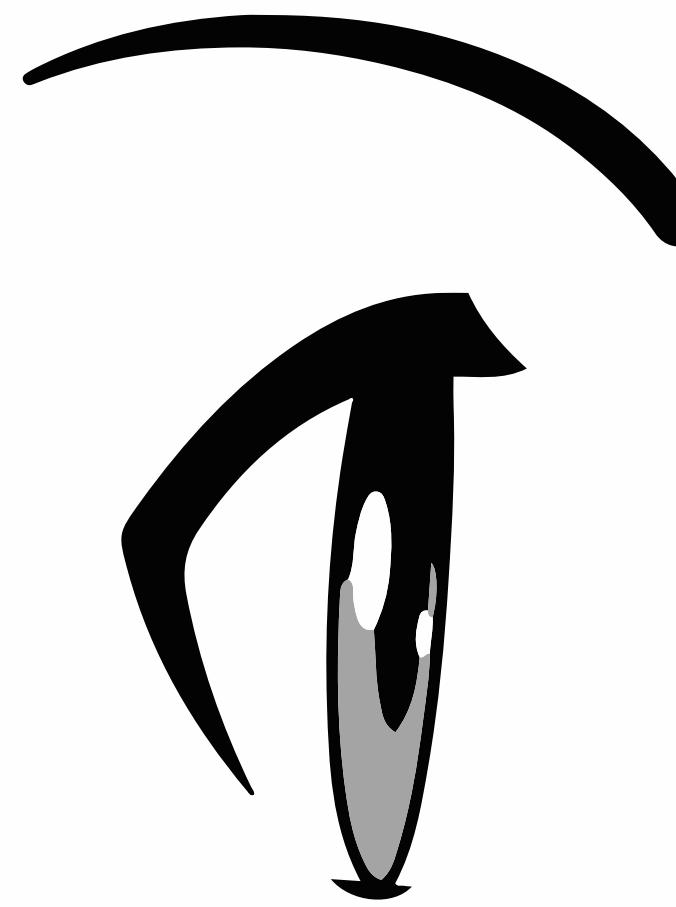
Humans have evolved to see edge contrasts, like here:

We see comparative — not absolute — luminance value.



Our comparative perception has implications for how to accurately represent data using luminance. **Background or adjacent** luminance — or hue or saturation — can influence how our audience perceives our data's encoded luminance value.

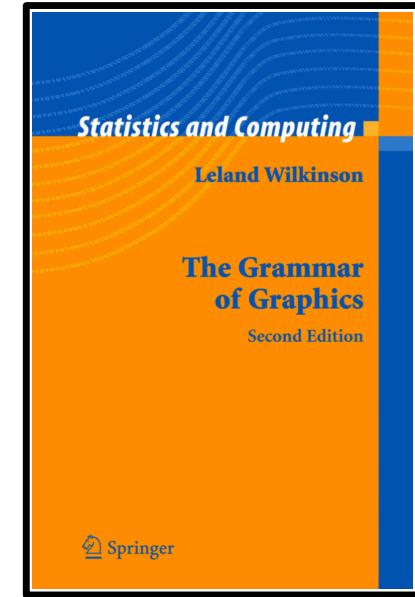




## Perceptually uniform color spaces also help in distinguishing categorical data.

With categorical data, we do not want one color value to appear brighter than another. Instead, we want to choose colors that both separate categories while holding their brightness level equal.

# Graphics have a grammar



# The grammar of graphics

## Wilkinson

Leland, formerly VP of Statistics at Tableau, is a statistician and computer scientist at H2O.ai, and Adjunct Professor of Computer Science at University of Illinois at Chicago. His research focuses on scientific visualization and statistical graphics.

A grammar links visual perception to graphics

Theory of graphics, not chart typology

Charts are just instances of more general objects

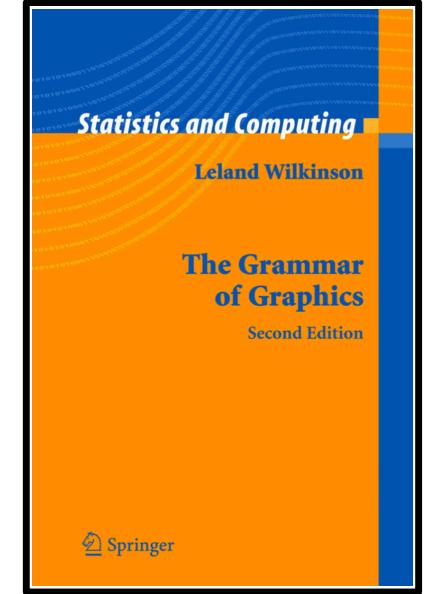
Thinking in terms of charts limit re-use for new, different cases

Once we understand how our audience can best perceive our data comparisons, a grammar helps us build graphics to accomplish that goal.

We often call graphics charts ... pie charts, bar charts, line charts, and so on. But **Customary usage and standards can blind us to the diversity of the graphics domain**; a formal system can liberate us from conventional restrictions.

Charts are usually instances of much more general objects. Once we understand that a pie is a divided bar in polar coordinates, we can construct other polar graphics that are less well known. We will also come to realize why a histogram is not a bar chart and why many other **graphics that look similar nevertheless have different grammars**.

If we ... develop charting instead of a graphing: **we inevitably will offer fewer charts than people want**. And we will have no way to add new charts ... without generating complex new code.



# The grammar of graphics

## Wilkinson

Leland, formerly VP of Statistics at Tableau, is a statistician and computer scientist at H2O.ai, and Adjunct Professor of Computer Science at University of Illinois at Chicago. His research focuses on scientific visualization and statistical graphics.

### So what's a graph, and a graphic?

A **graph** is a set of points. A mathematical graph cannot be seen. It is an abstraction. A **graphic, however, is a physical representation of a graph**. This representation is accomplished by realizing graphs with aesthetic attributes such as size or color.

### Six components to statistical graphs

**Algebra** the operations that allow us to combine variables, specify graph dimensions

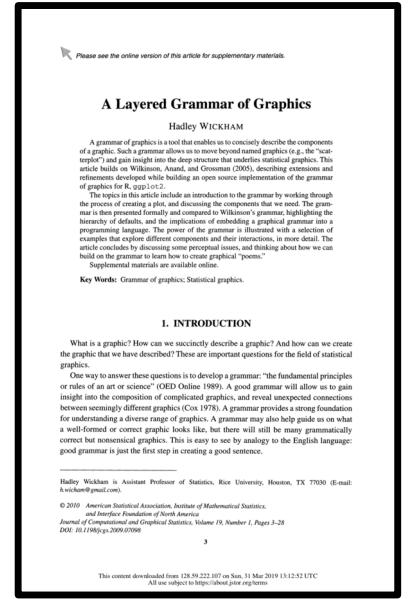
**Scales** the representation of variables on measured dimensions (e.g., linear, log)

**Statistics** the functions that allow graphs to change their appearance and representation

**Geometry** the creation of geometric graphs from variables

**Coordinates** coordinate systems, from polar coordinates to more complex map projections and general transformations

**Aesthetics** covers the sensory attributes used to represent graphics.



# A layered grammar of graphics

*Wickham*

Hadley, an adjunct professor at Rice University, is the creator of the R package ggplot2, which is a limited implementation of Wilkinson's seminal work. gg is short for grammar of graphics.

GPL

ggplot2

DATA → Defaults

TRANS  
Data  
Mapping

ELEMENT → Layer

Data  
Mapping  
Geom  
Stat  
Position

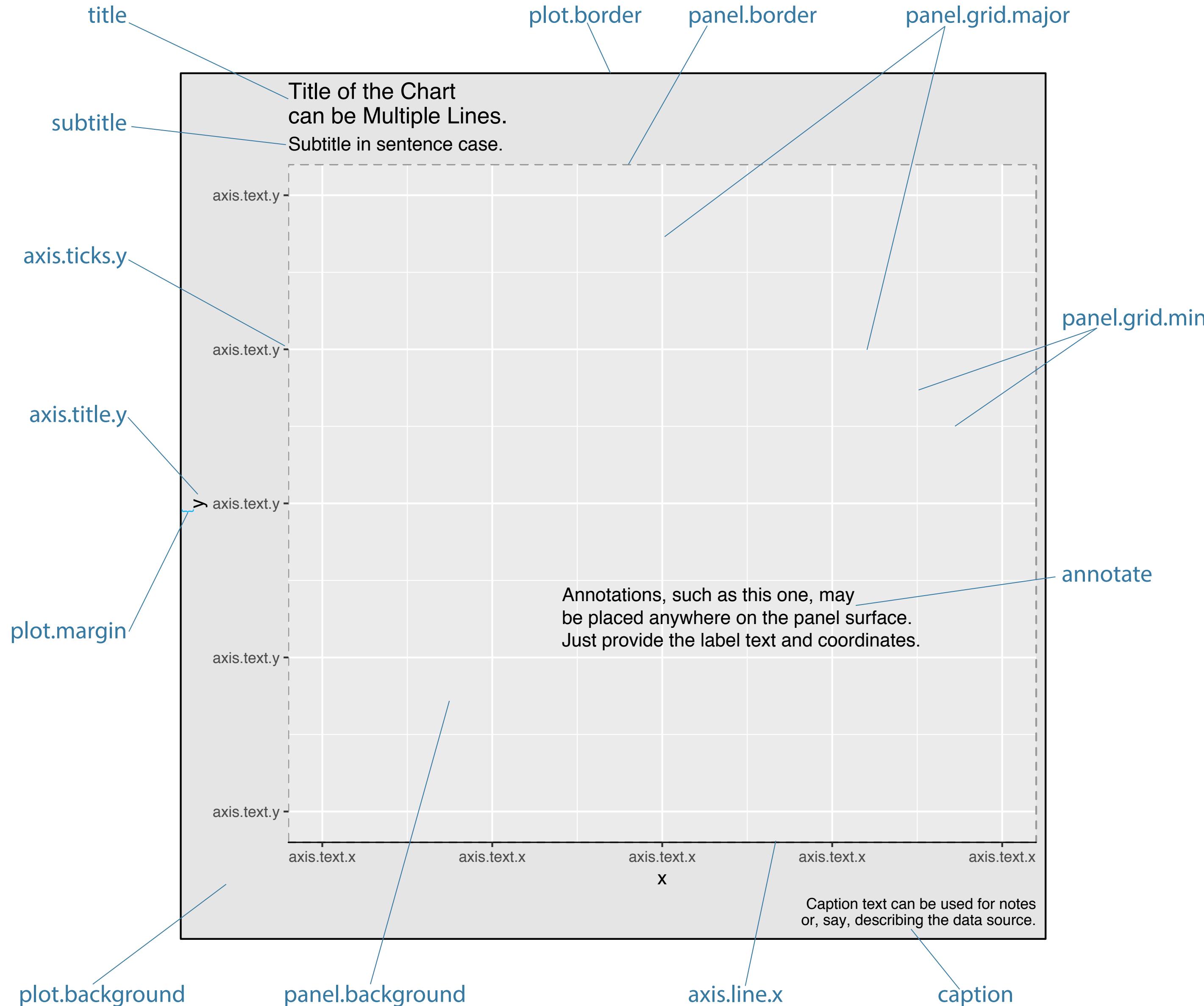
SCALE → Scale

GUIDE

COORD → Coord

Facet

# ggplot2, an R library implements this grammar of graphics.



```
# load grammar of graphics  
library(ggplot2)
```

```
p <-
```

```
# functions for data ink
```

```
ggplot(data = <data>,  
       mapping = aes(<aesthetic> = <variable>,  
                     <aesthetic> = <variable>,  
                     <...> = <...>)) +  
  geom_<type>(<...>) +  
  scale_<mapping>_<type>(<...>) +  
  coord_<type>(<...>) +  
  facet_<type>(<...>) +  
  <...> +
```

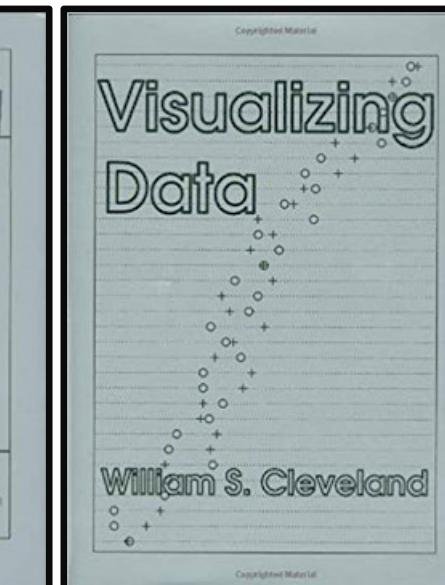
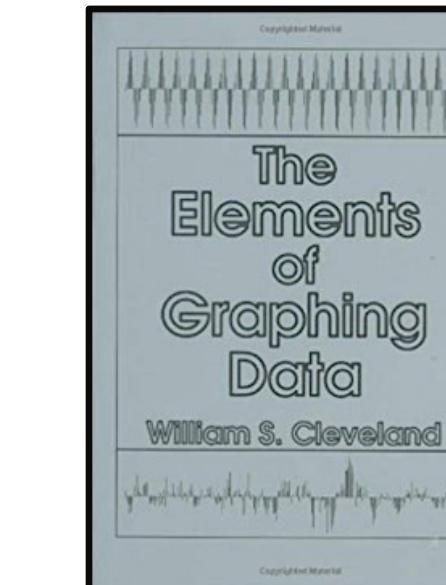
```
# functions for non-data ink
```

```
labs(<...>) +  
theme(<...> = <...>) +  
annotate(<...>) +  
<...>
```

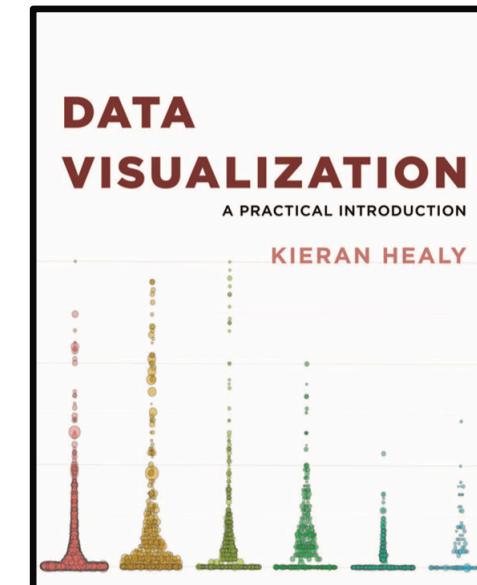
```
element_blank()  
element_line(<...> = <...>)  
element_rect(<...> = <...>)  
element_text(<...> = <...>)
```

## Learning & References

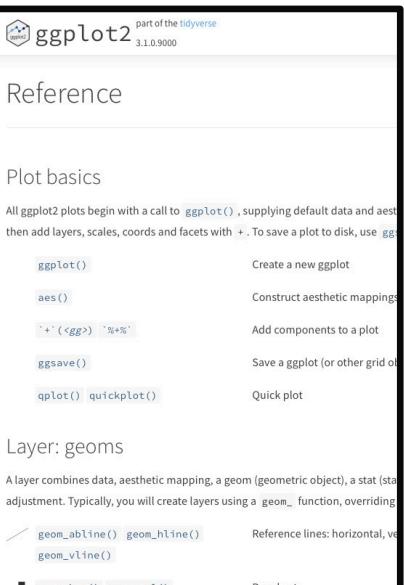
---- visual relationships in data ----



Implementation in R



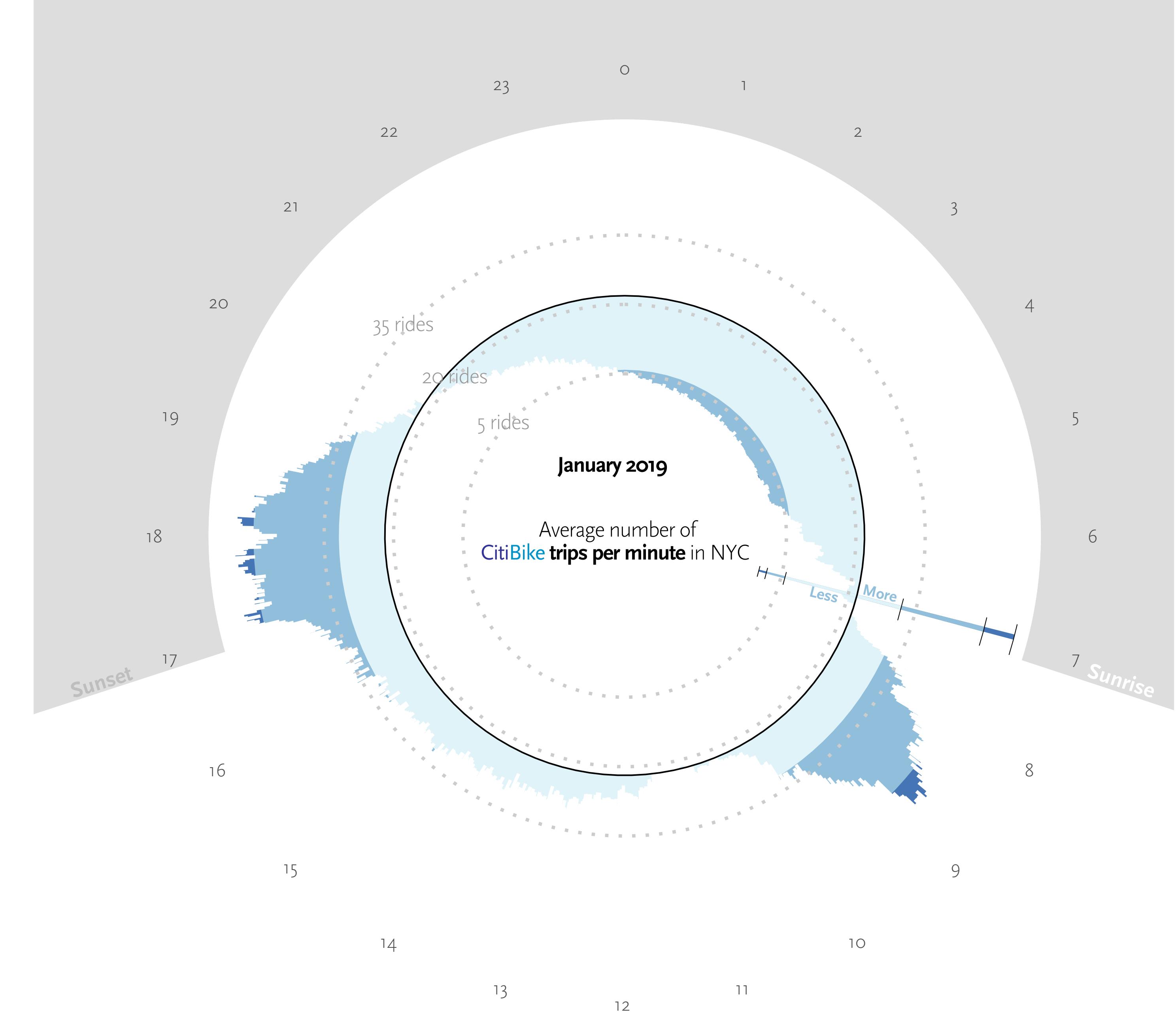
ggplot reference



socviz.co

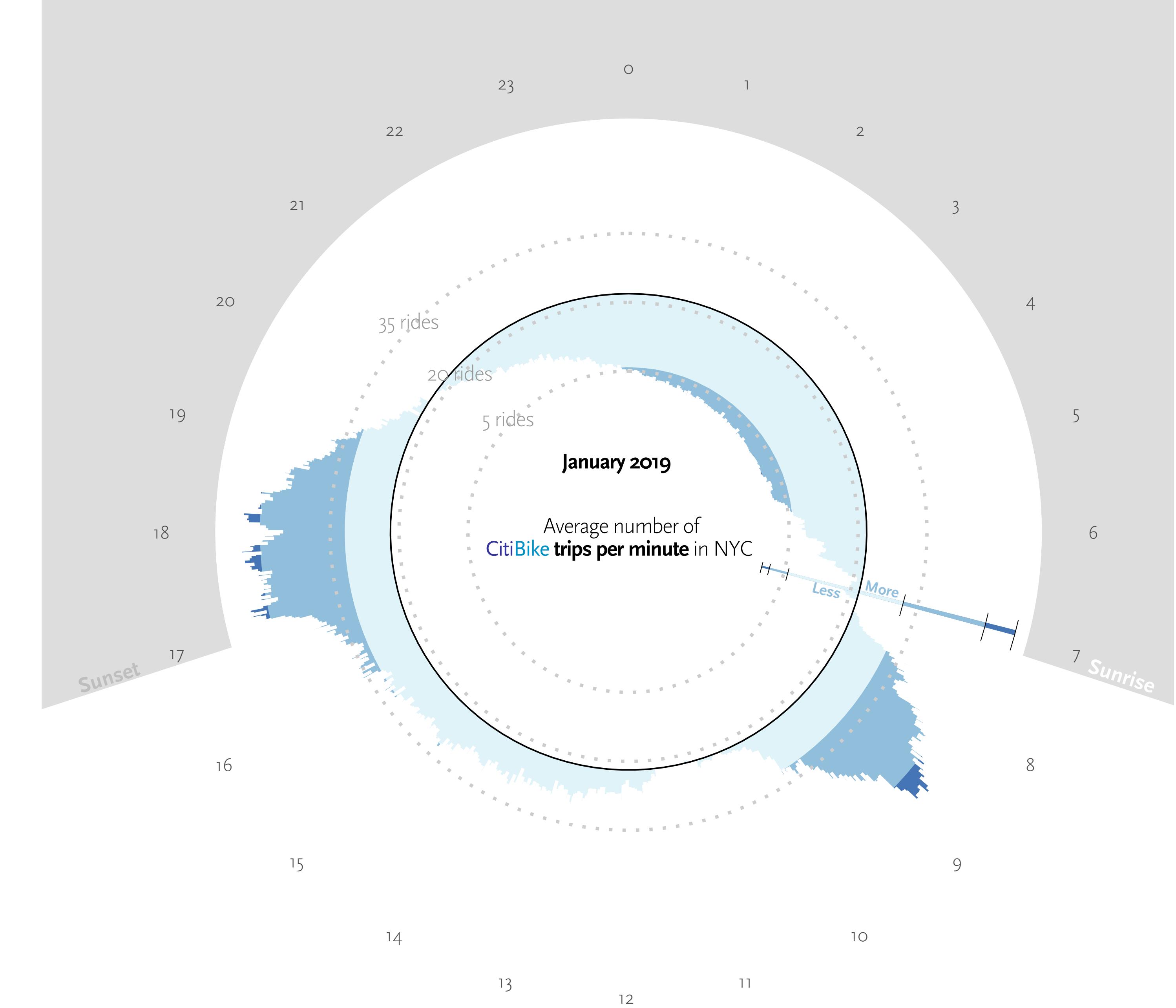
ggplot2.tidyverse.org

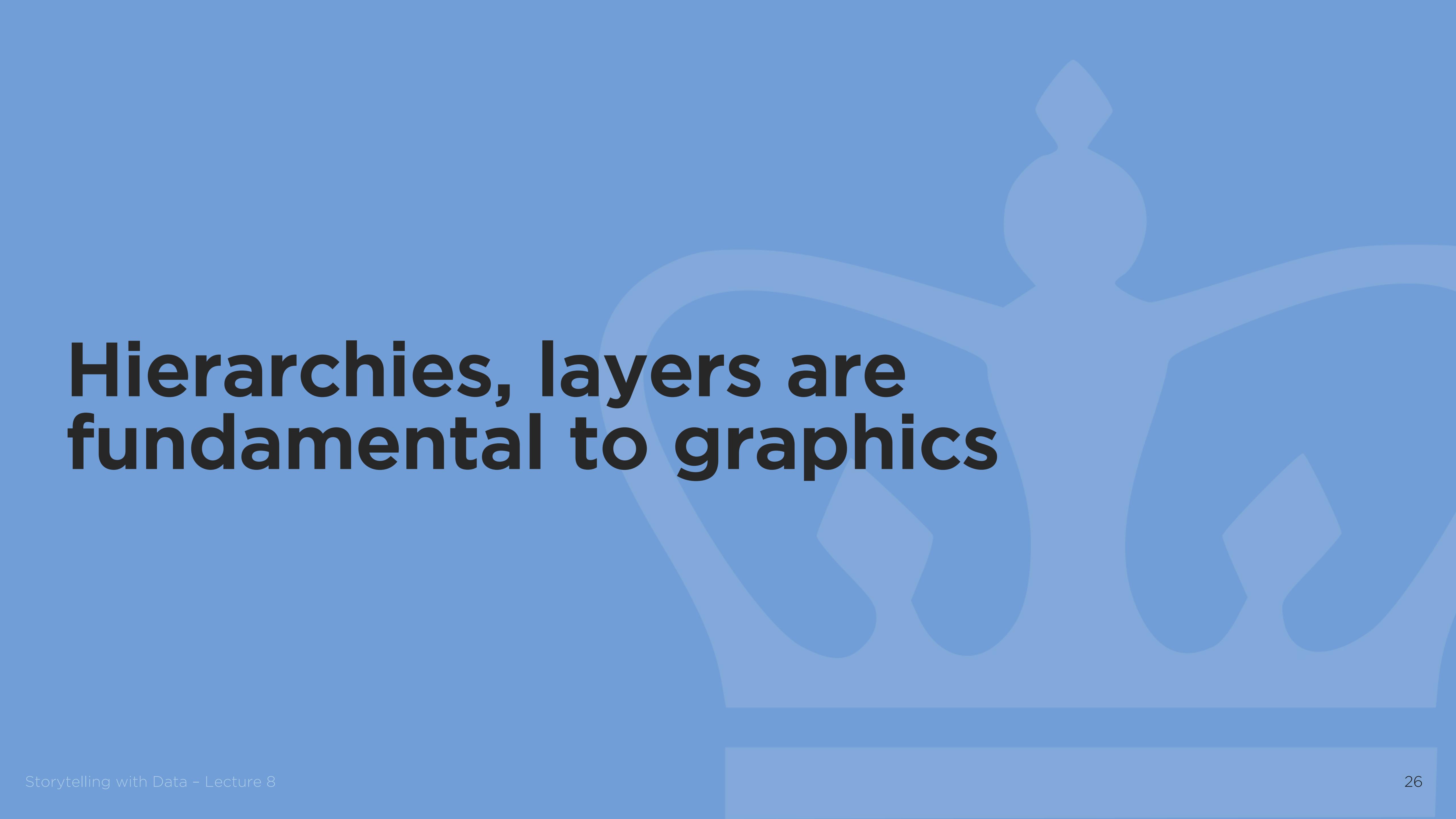
# Think about graphics as l/a/y/e/r/s:



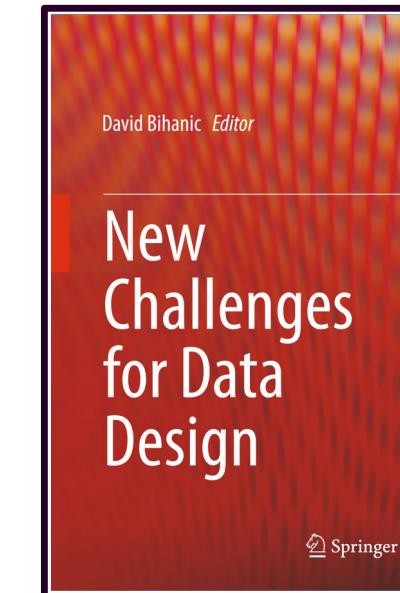
Think about  
graphics as l/a/y/e/r/s:

Here, a white layer  
partly masks a  
rectangular band of blues





# **Hierarchies, layers are fundamental to graphics**



# The New Aesthetic of Data Narrative

*Lupi*

Her work is part of the permanent collection of the Museum of Modern Art. She is co-founder and Design Director of Accurat, a data-driven design firm. After receiving her master's degree in Architecture, she earned her PhD in Design at Politecnico di Milano.

Create displays with layers, hierarchies

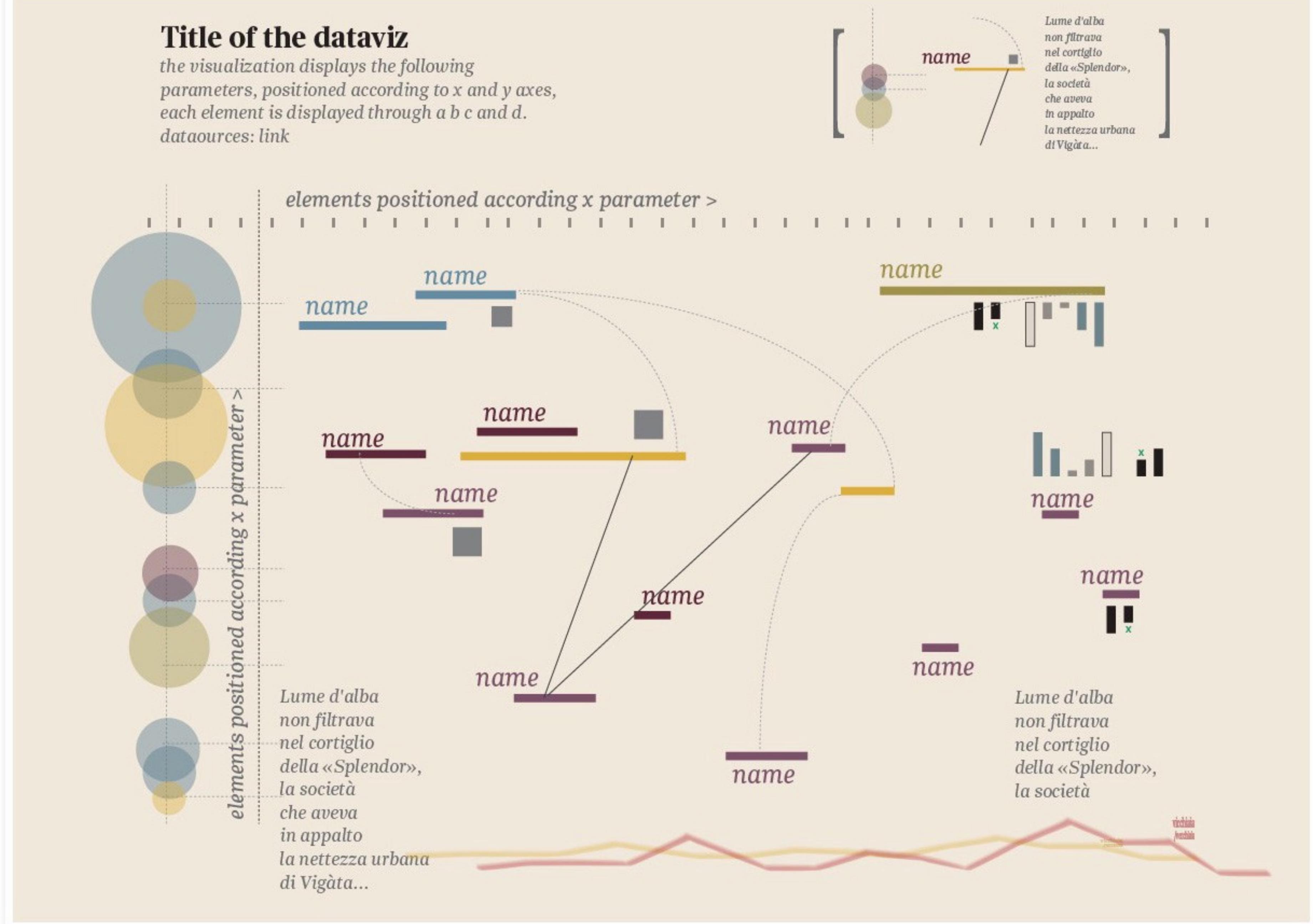
## DATA HUMANISM

small big data  
data bandwidth ~~quality~~  
  
imperfect infallible data  
subjective impartial data  
inspiring descriptive data  
serendipitous predictive data  
  
data conventions ~~possibilities~~  
data to simplify complexity / depict  
data processing ~~drawing~~  
data driven design  
data is numbers ~~people~~  
data will make us more efficient ~~human~~.

Her process involves layering multiple sub-narratives over a main construct. To achieve this multilayered storytelling with data ... everything depends on the concept of layering, establishing hierarchies, and making them clear.

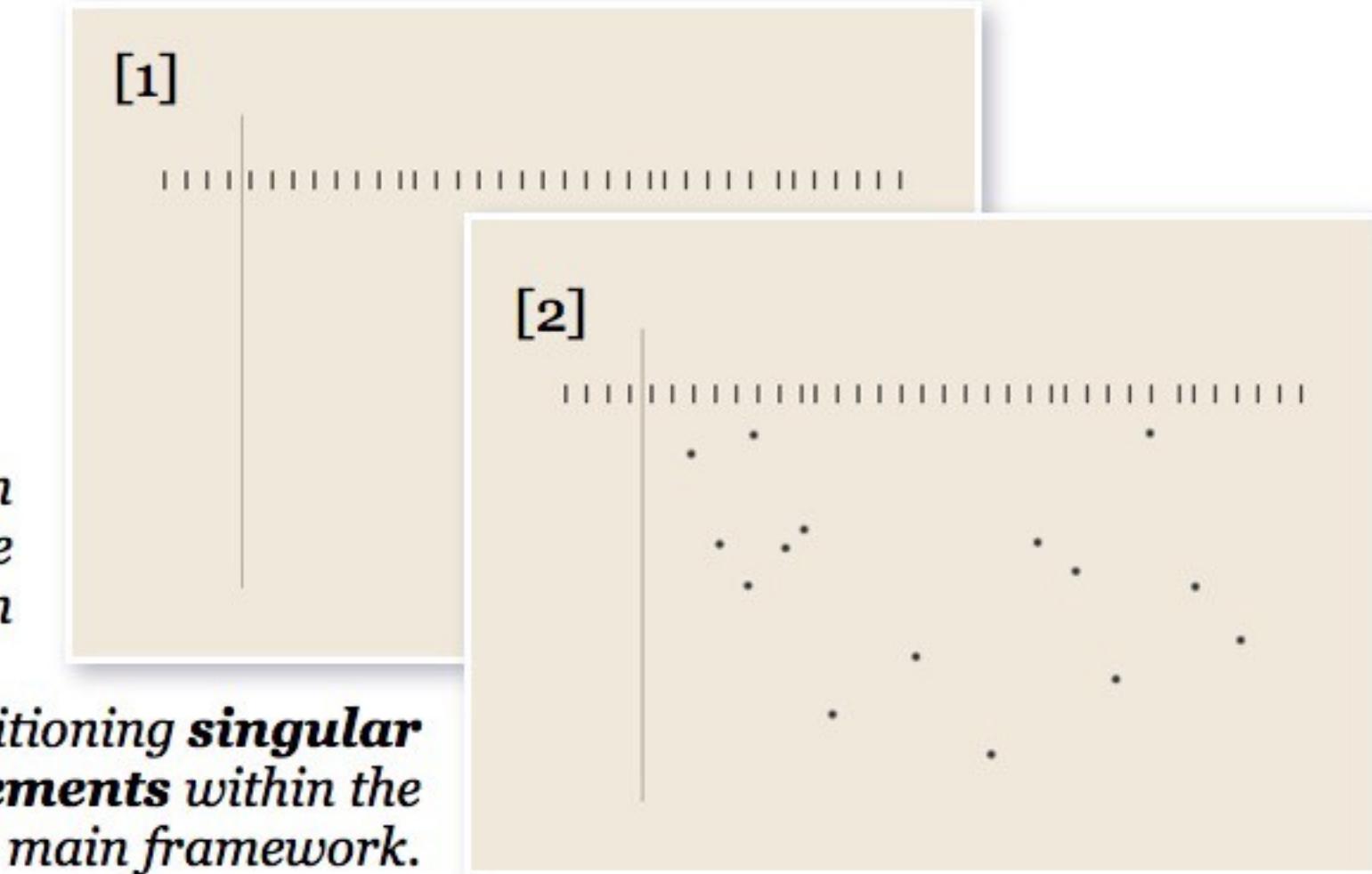
## Title of the dataviz

the visualization displays the following parameters, positioned according to x and y axes, each element is displayed through a b c and d. datasources: link

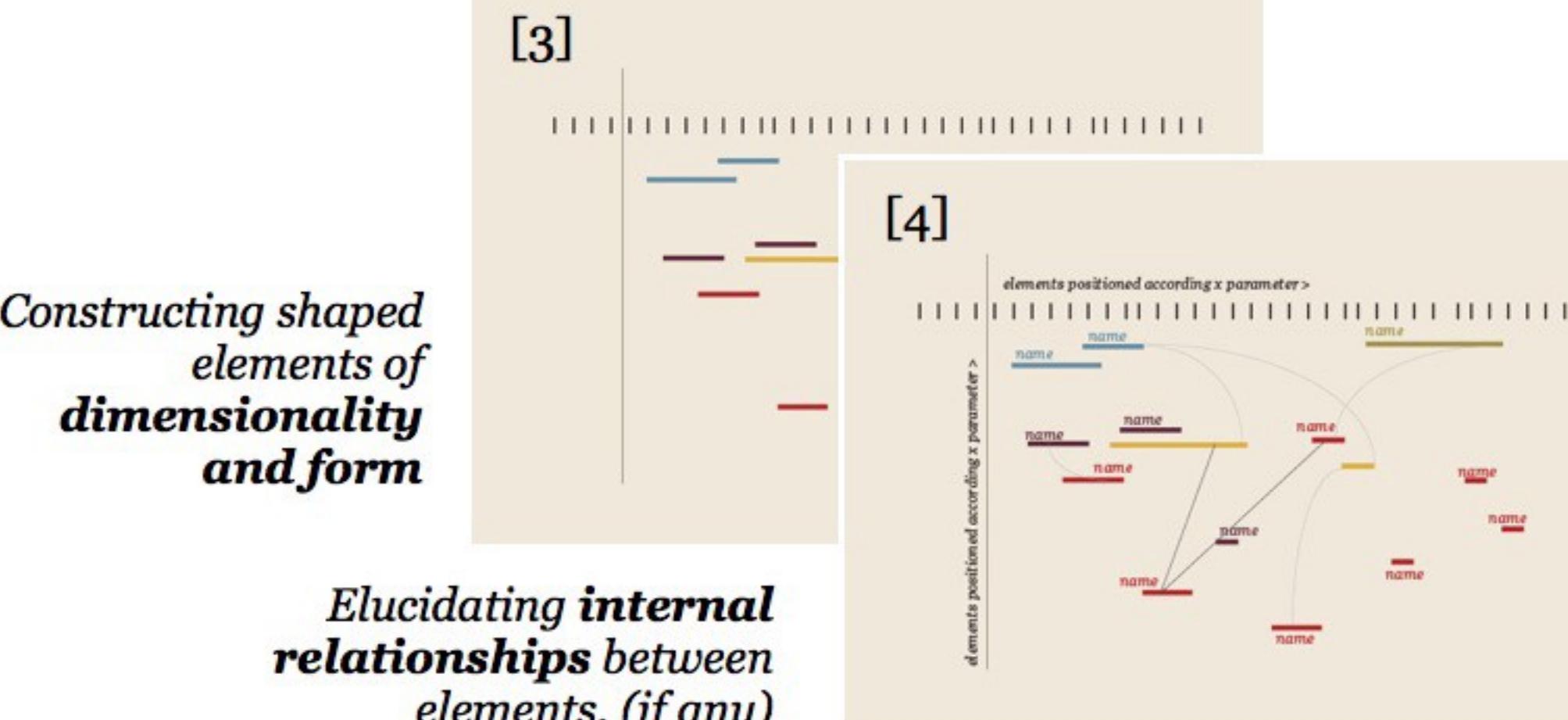


# Layering, and making hierarchies clear

*Composing the main architecture of the visualization*

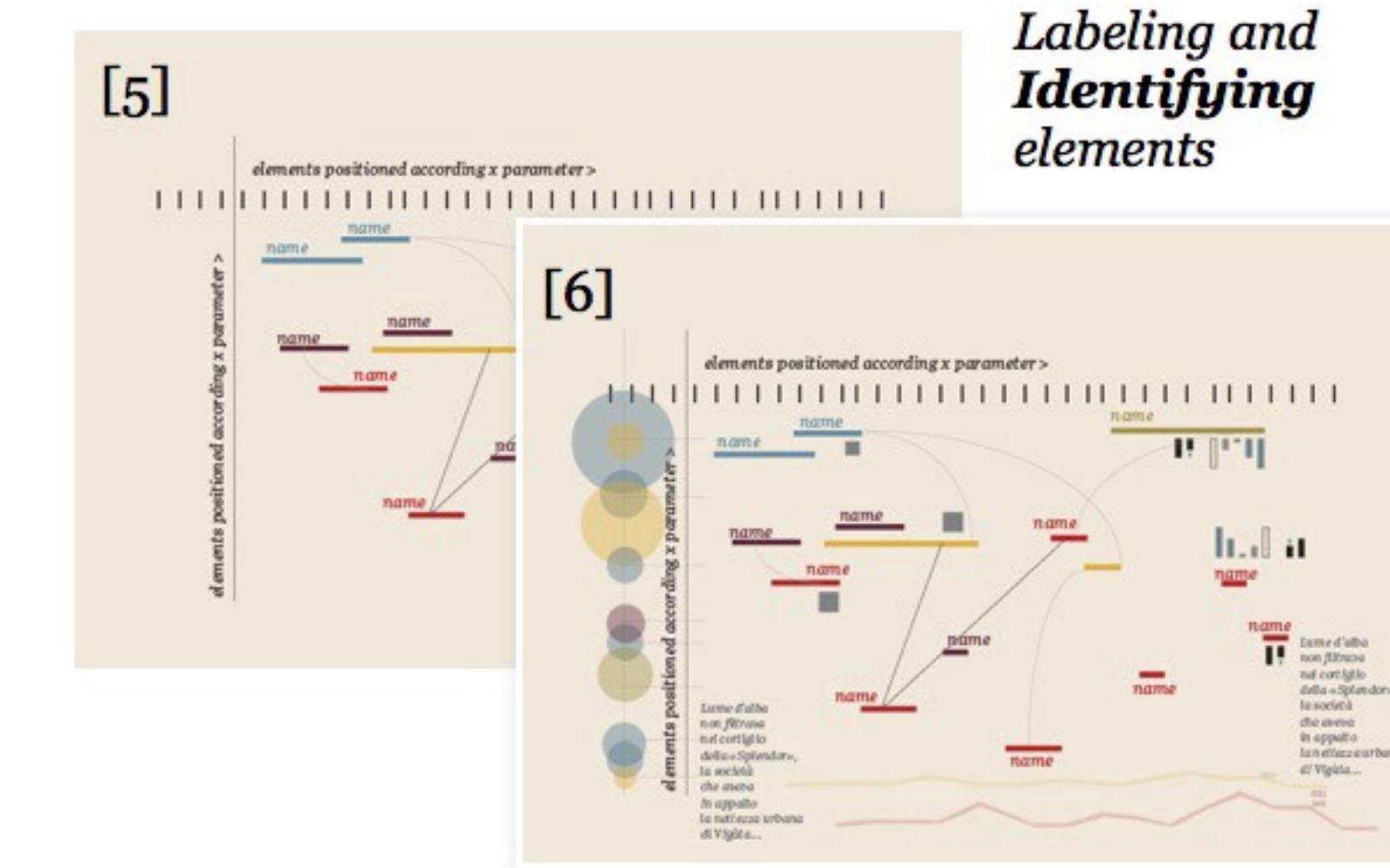


*Positioning singular elements within the main framework.*

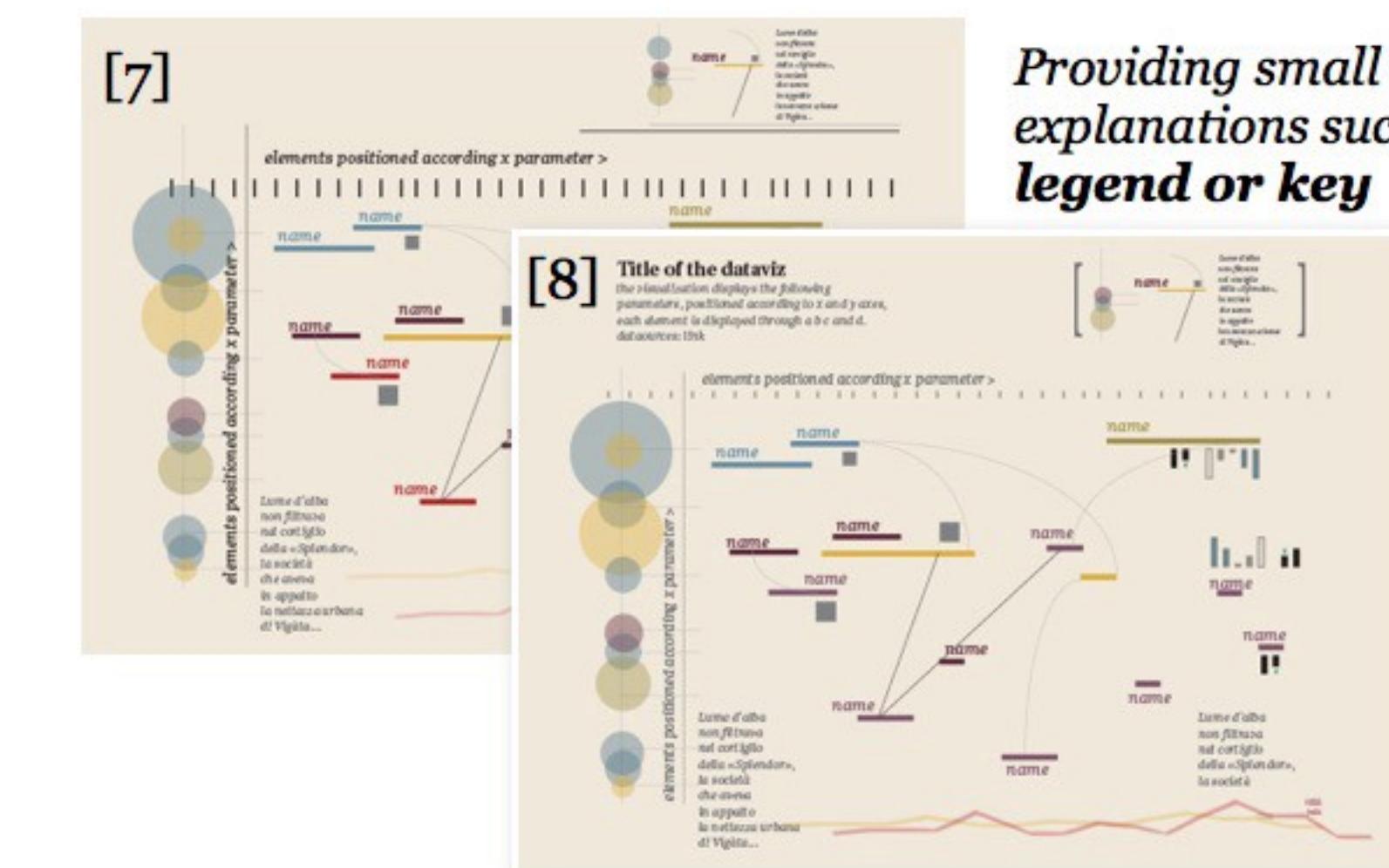


*Constructing shaped elements of dimensionality and form*

*Elucidating internal relationships between elements. (if any)*



*Labeling and Identifying elements*



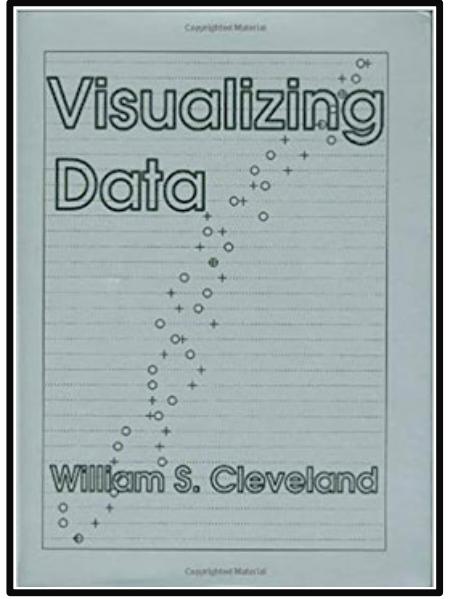
*Providing small visual explanations such as a legend or key*

*Fine-tuning and stylizing of elements shapes, colors, and weights to make hierarchies pop out.*

*Supplementing the greater story through the addition of “minor or tangential tales” elements*

# Comparing data

TODO: DEVELOP SECTION



# Visualizing Data

*Cleveland*

A graduate of Princeton, William is a computer scientist and Professor of Statistics and Professor of Computer Science at Purdue University, known for his work on data visualization, particularly on nonparametric regression and local regression.

**Univariate data**

Details.

# **Building information complexity from common business graphics**

**TODO: DEVELOP SECTION**



# Ensemble Graphics

*Unwin & Valero-Mora*

Anthony is a professor of computer-oriented statistics and data analysis at the University of Augsburg. He is a fellow of the American Statistical Society, and his research focuses on data visualization, especially in interactive graphics.

## Ensemble graphics

Graphics grouped together without text chunks between them are considered ensembles, when all the graphics can be seen at once.

The point is to assemble a range of different views of data in a coherent whole to discover or convey information that would otherwise be spread over different places, making it difficult and awkward to evaluate them simultaneously.

## Ensembles versus infographics

Infographics are not included, although these are also closely related. Our definition excludes narratives. It does **allow ensembles to be components of narratives**.

## Coherence

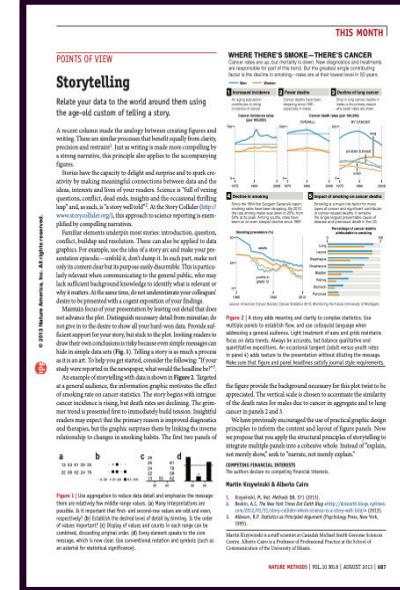
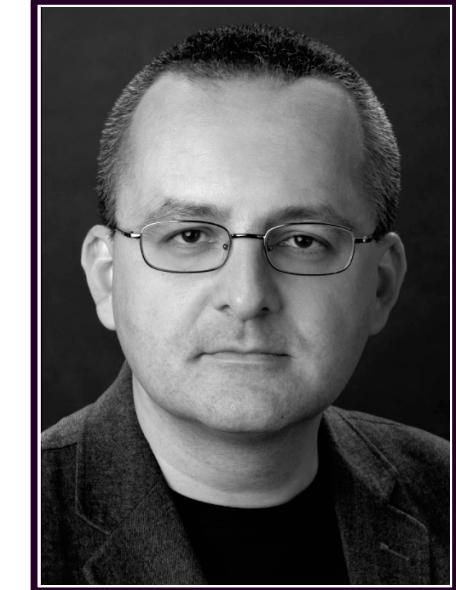
The term “coherent” covers many different aspects: a coherent **theme**, a coherent **look**, consistent **scales**, **formatting**, and **alignment**. Coherence facilitates understanding.

## Apply best practices in design

Requirements for coherence are more design than statistics, but they are driven by the statistical information to be conveyed, and it is therefore essential that statisticians concern themselves with them.

# Storytelling in graphics

TODO: DEVELOP SECTION



# Storytelling

*Cairo & co-author*

Cairo is a journalist and designer, and the Knight Chair in Visual Journalism at the School of Communication of the University of Miami (UM), and director of the visualization program its Center for Computational Science.

## WHERE THERE'S SMOKE—THERE'S CANCER

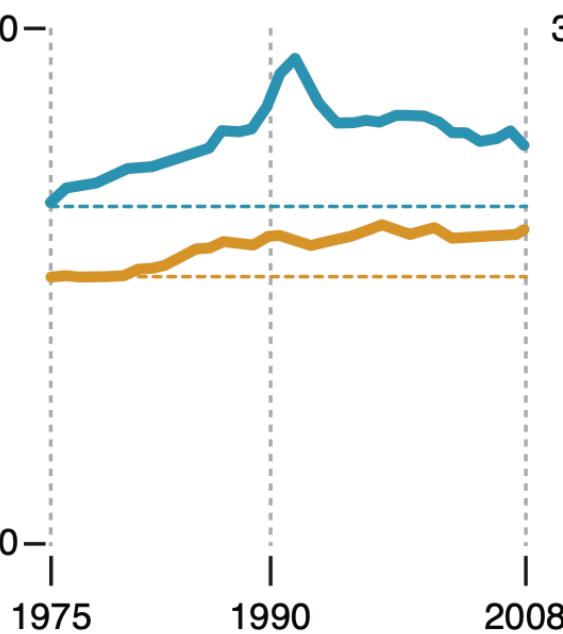
Cancer rates are up, but mortality is down. New diagnostics and treatments are responsible for part of this trend. But the greatest single contributing factor is the decline in smoking—rates are at their lowest level in 50 years.

Men      Women

### 1 Increased incidence

An aging population contributes to rising incidence of cancer.

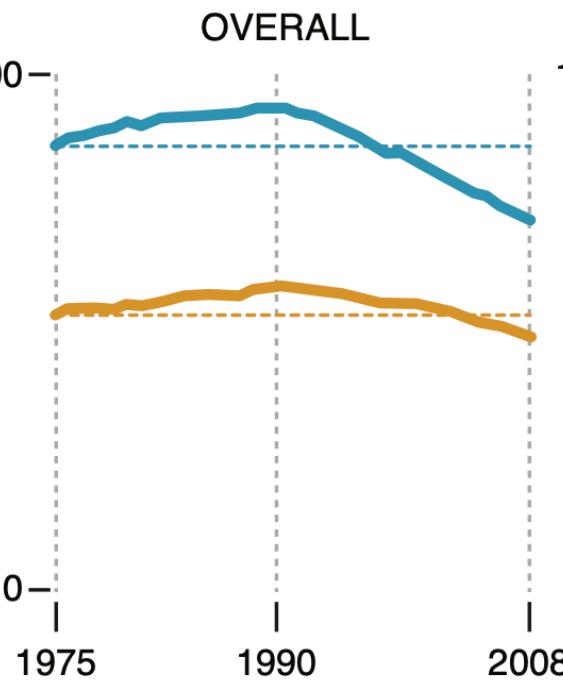
**Cancer incidence rates (per 100,000)**



### 2 Fewer deaths

Cancer deaths have been dropping since 1991, especially in males.

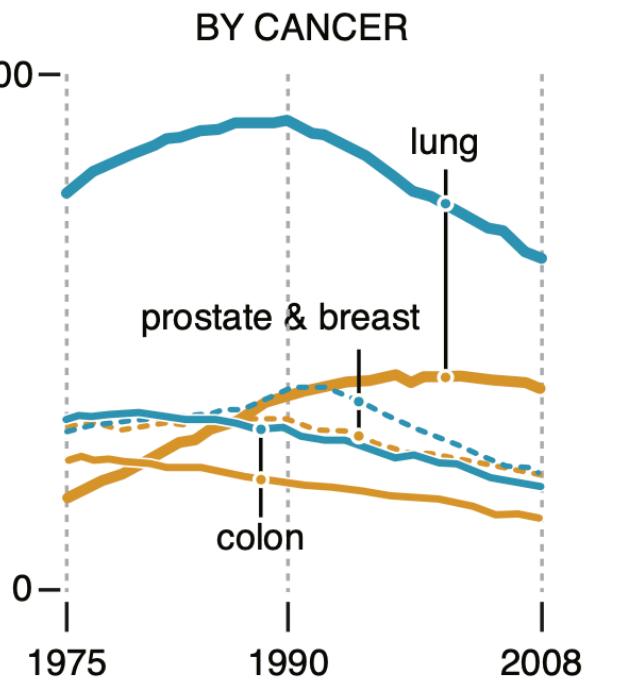
**Cancer death rates (per 100,000)**



### 3 Decline of lung cancer

Drop in lung cancer deaths in males is the primary reason why death rates are down.

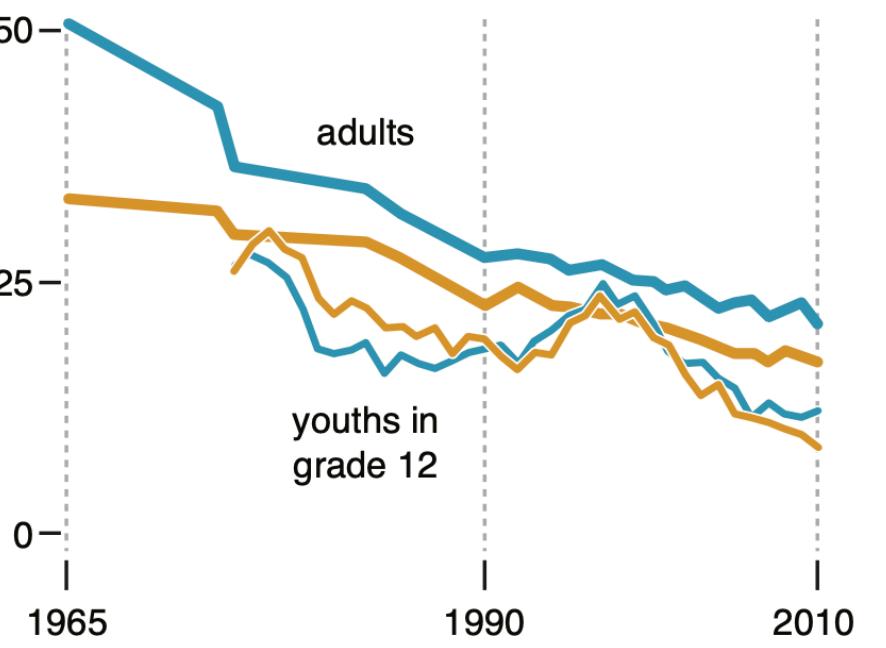
**BY CANCER**



### 4 Decline in smoking

Since the 1964 first Surgeon General's report, smoking rates have been dropping. By 2010, the rate among males was down to 20%, from 50% at its peak. Among youths, rates have been on an even steeper decline since 1997.

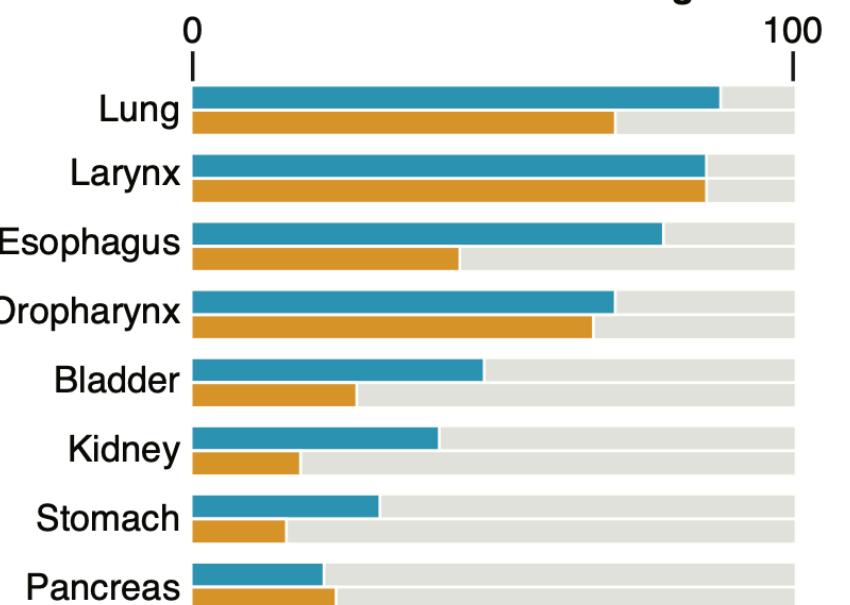
**Smoking prevalence (%)**



### 5 Impact of smoking on cancer deaths

Smoking is a major risk factor for many types of cancer and significant contributor to cancer-related deaths. It remains the single largest preventable cause of disease and premature death in the US.

**Percentage of cancer deaths attributable to smoking**

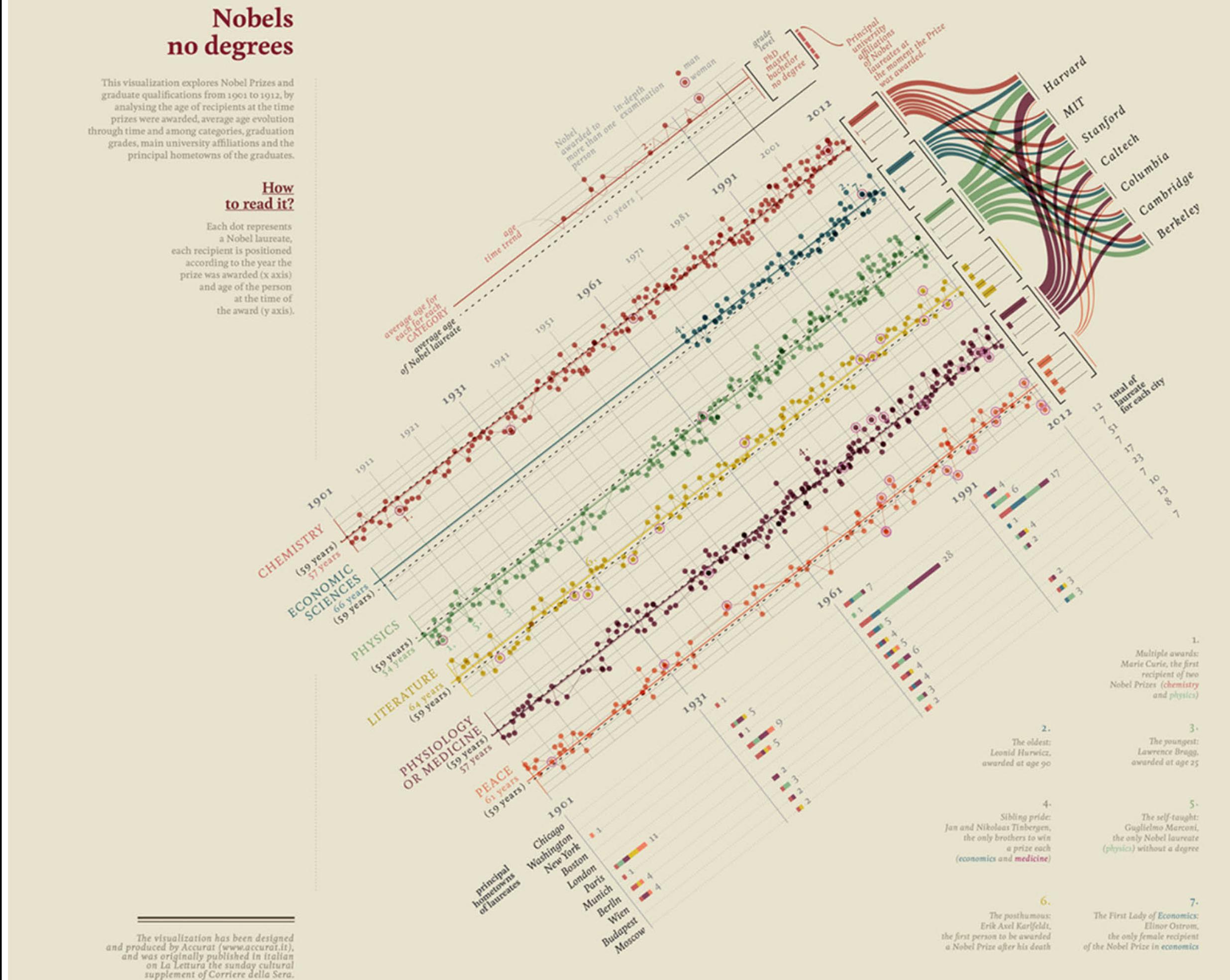


source: American Cancer Society Cancer Statistics 2012; Monitoring the Future (University of Michigan).

# **Example infographic as an ensemble with narrative**

# Winner, Information is Beautiful Award

Lupi heads a design firm and this visualization is a representative example from her series of exploratory data-visualizations originally published for *La Lettura*, their Sunday cultural supplement, spanning two years.



Storytelling with Data – Lecture 8

# Components are integrated with color, font—placement—lines, and annotation.

All four displays share data information. These relationships are integrated together using consistent color, in the case of prize categories,



and integrated using the relative placement of each display for both categories and other shared data. Placement of the histograms and Sankey diagram alongside the same categories in the central display reinforces the connection.

And notice, for example, that the per-era stacked bars share dates by aligning a common baseline with the x-axis of the central display.

## Nobels no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 1912, by analysing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

### How to read it?

Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded (x axis) and age of the person at the time of the award (y axis).



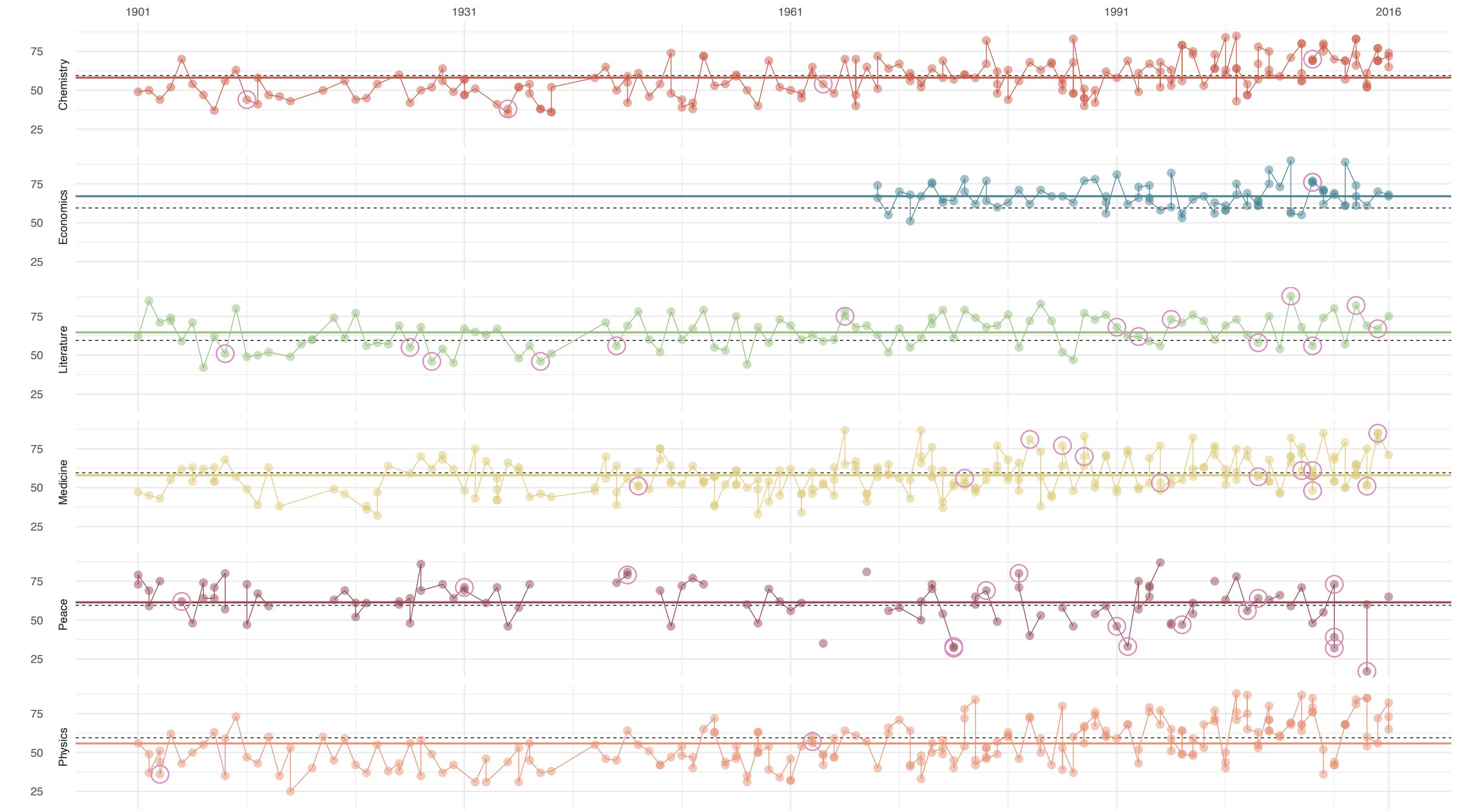
# Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



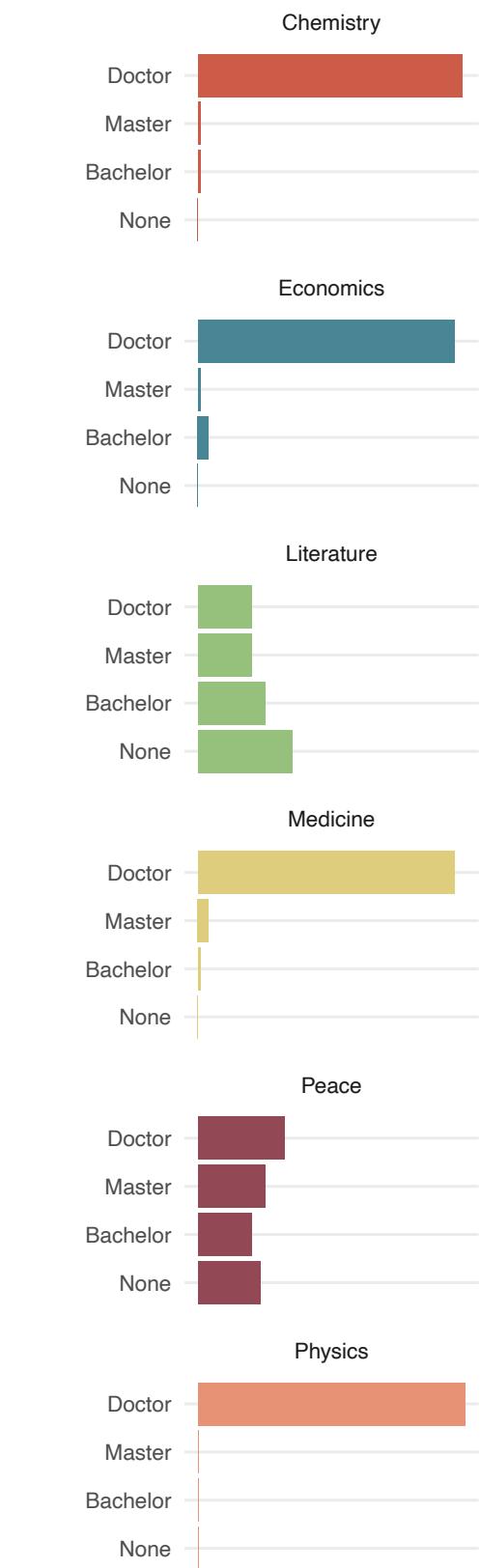
# Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



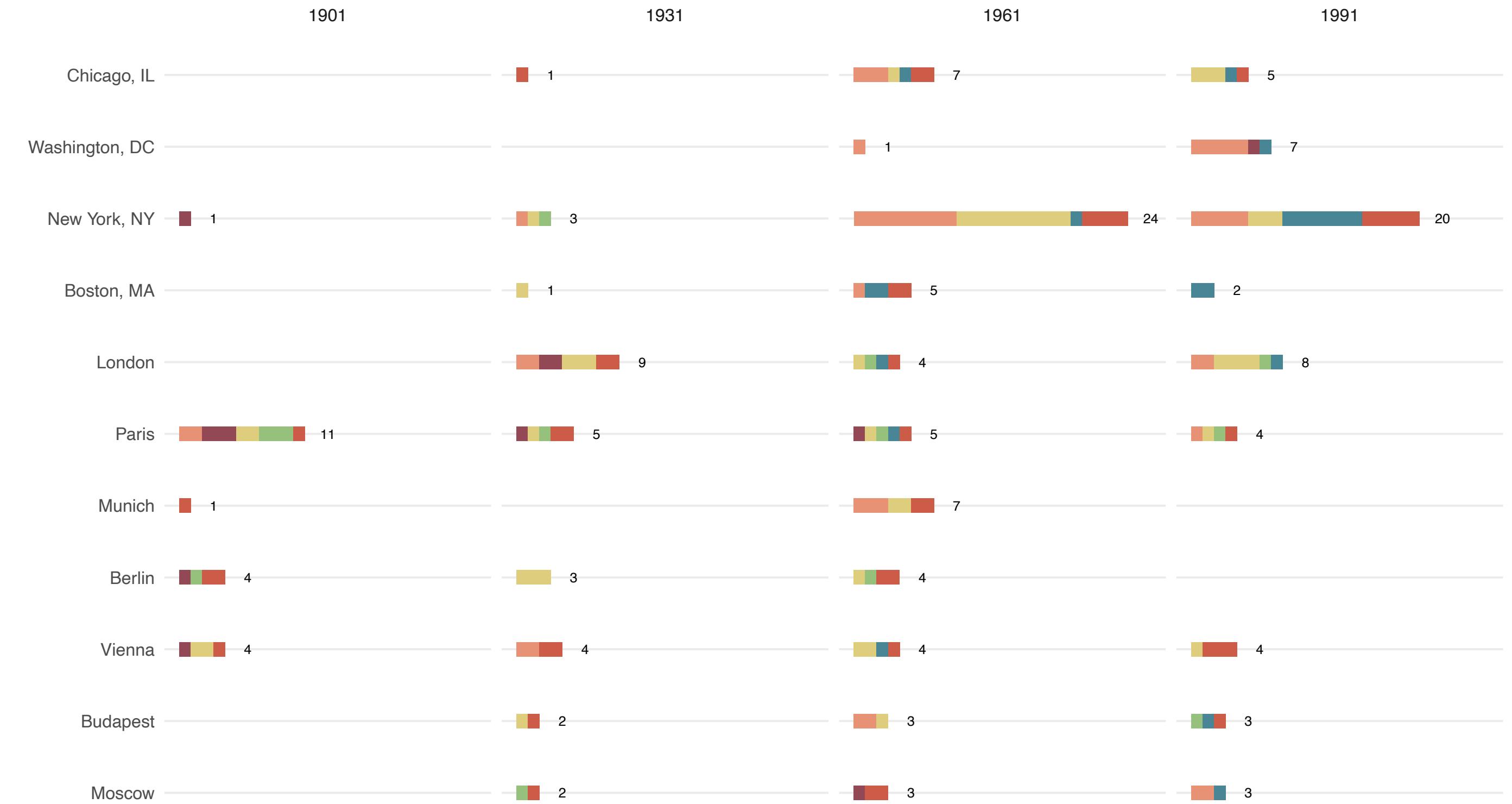
# Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



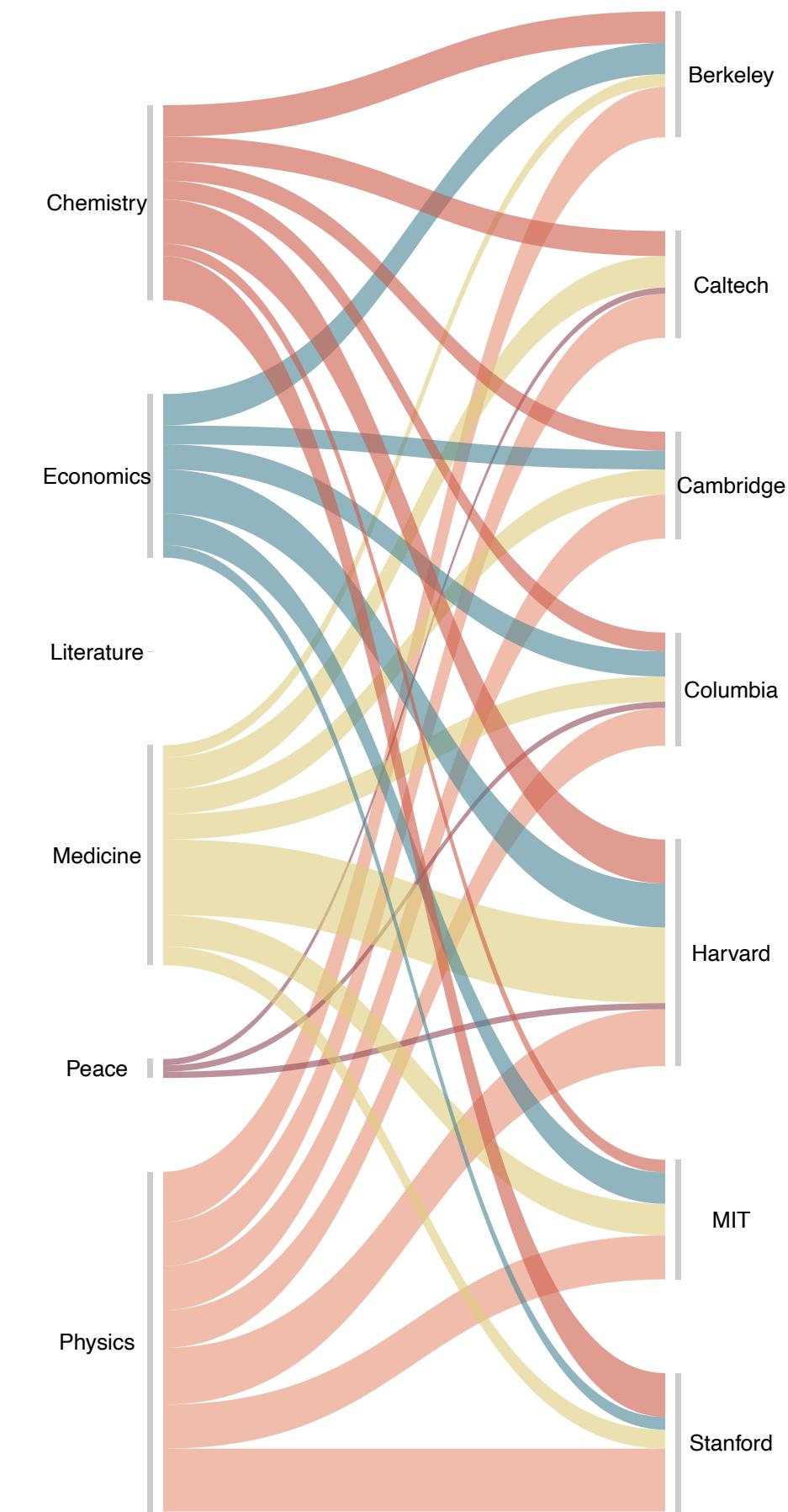
# Approximating the components

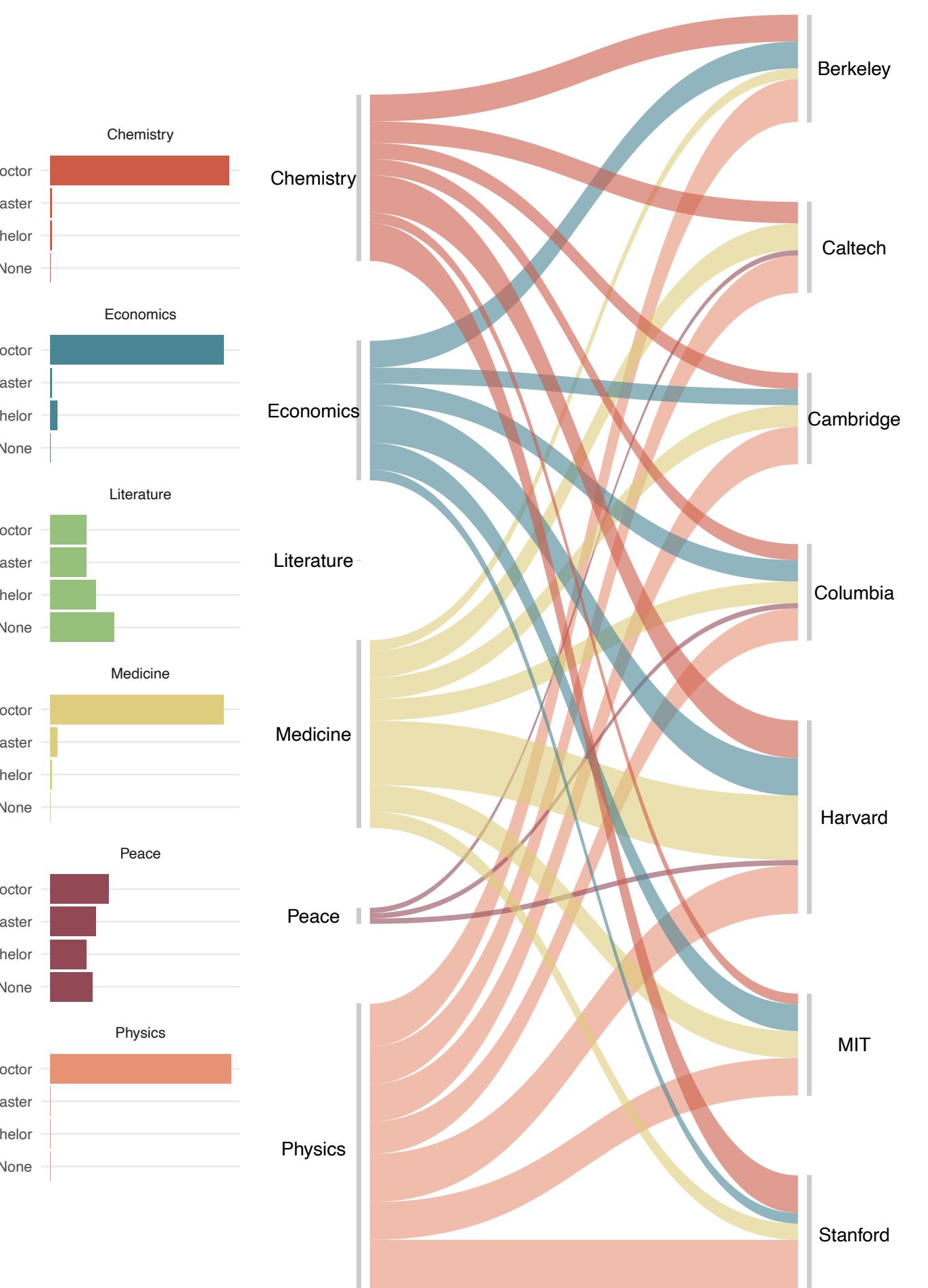
Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



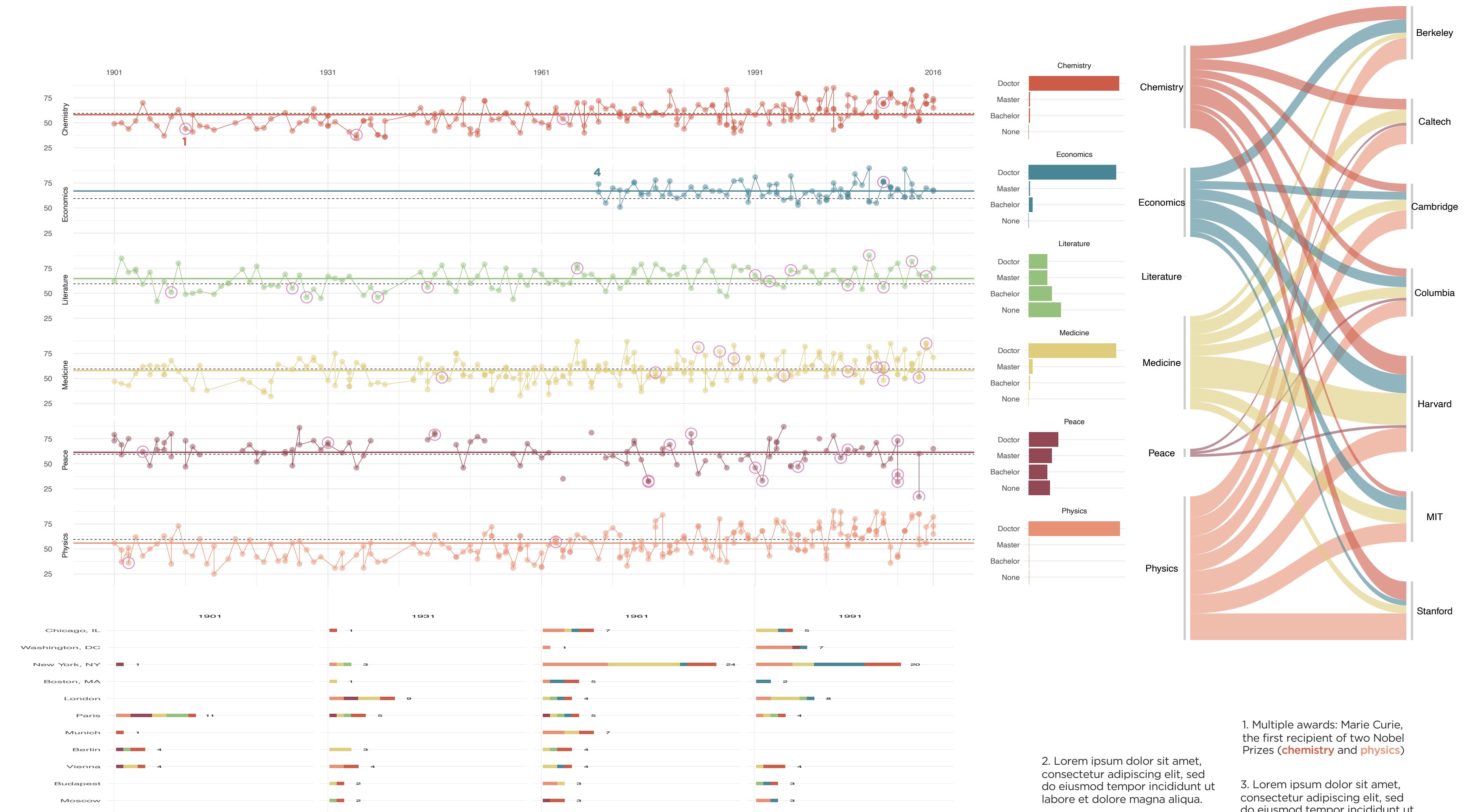


# Nobels, no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 2012, by analyzing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

## How to read it?

Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded (x axis) and age of the person at the time of the award (y axis).



1. Multiple awards: Marie Curie, the first recipient of two Nobel Prizes (**chemistry** and **physics**)

2. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

3.

3. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

4. Sibling pride: Jan and Nikoalaas Tinbergen, the only brothers to win a prize each (**economics** and **medicine**)

6. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

7. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

## Nobel Laureates and Prizes

Chart type: *connected scatter plot, alluvial diagram, small multiples of stacked bar charts*

This example is based on the [award-winning design](#) by Accurat.

[Download Dataset in CSV Format](#)



Follow us on Twitter 

Ask questions, report issues and request features 

© All Rights Reserved.

# On the horizon

# For Next Week, Module 9:

## Agenda next week

### The minimum

Next deliverable, *draft* information graphic

Perceptions of uncertainty, missing data

Information graphics

**Andrews, R J.** *Info We Trust: How to Inspire the World with Data*. Wiley, 2019. Print. Chp 17-18

While the whole book is an excellent read of plain language descriptions of what we've been covering, focus on chapters 17-18 for class discussion.

**Hullman, Jessica et al.** *Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty*. IEEE Transactions on Visualization and Computer Graphics 24.1 (2017): 446–456. Web.

How do audiences perceive uncertainty? Consider approaches to communicating uncertainty important to your messages about your projects.

**Song, Hayeong, and Danielle Albers Szafir.** *Where's My Data? Evaluating Visualizations with Missing Data*. IEEE Transactions on Visualization and Computer Graphics 25.1 (2018): 914–924. Web.

How should we reason, explore, visualize and communicate the implications of missing data?

# For online discussion

## Your turn

Ask me a question and I'll try to answer.

## Your project

In creating your project, did you find data missing from the information you gathered? If so, how would you describe or visualize it? How did that affect your analysis?

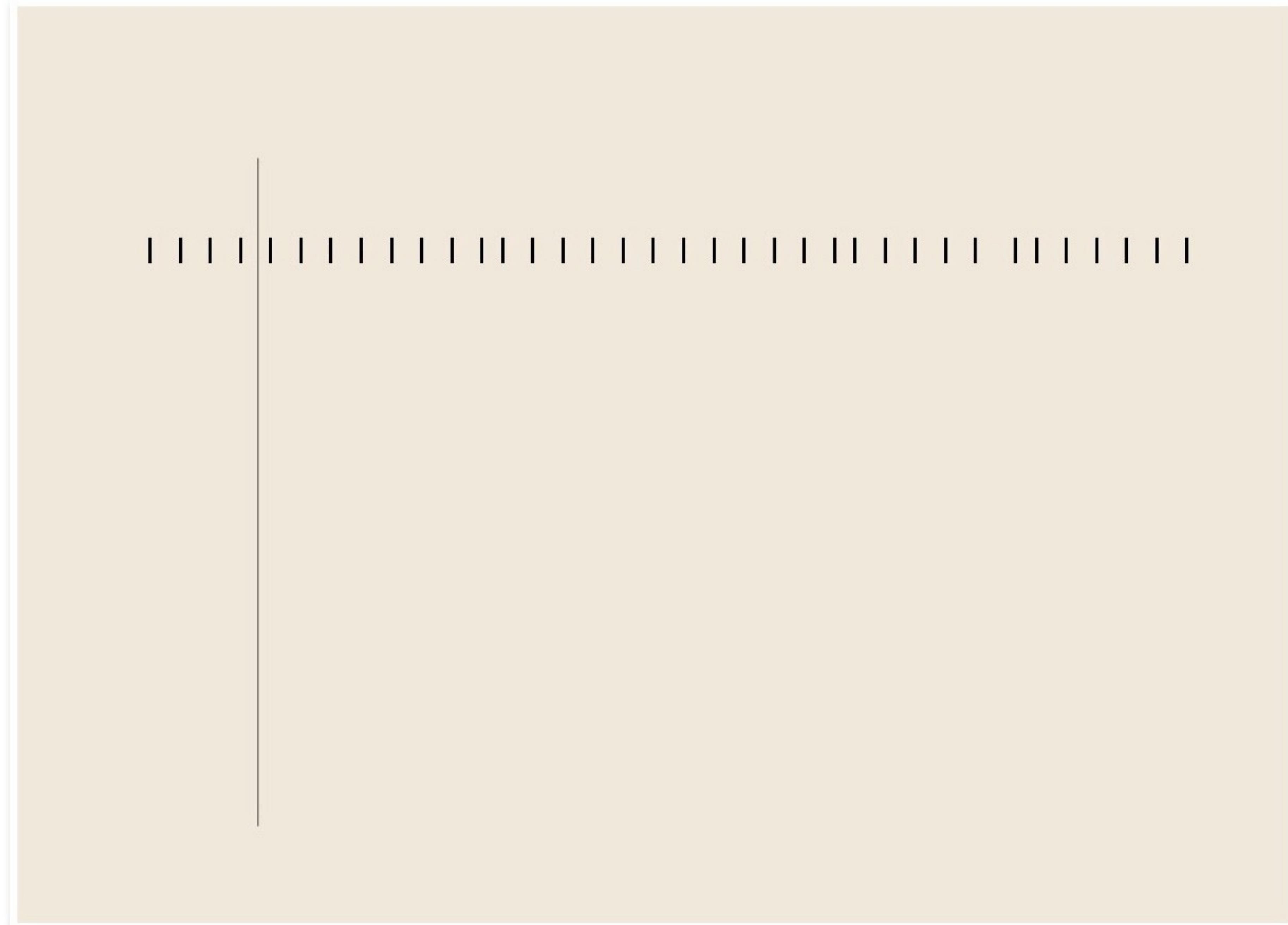


See you  
next week!

# SAVE AREA

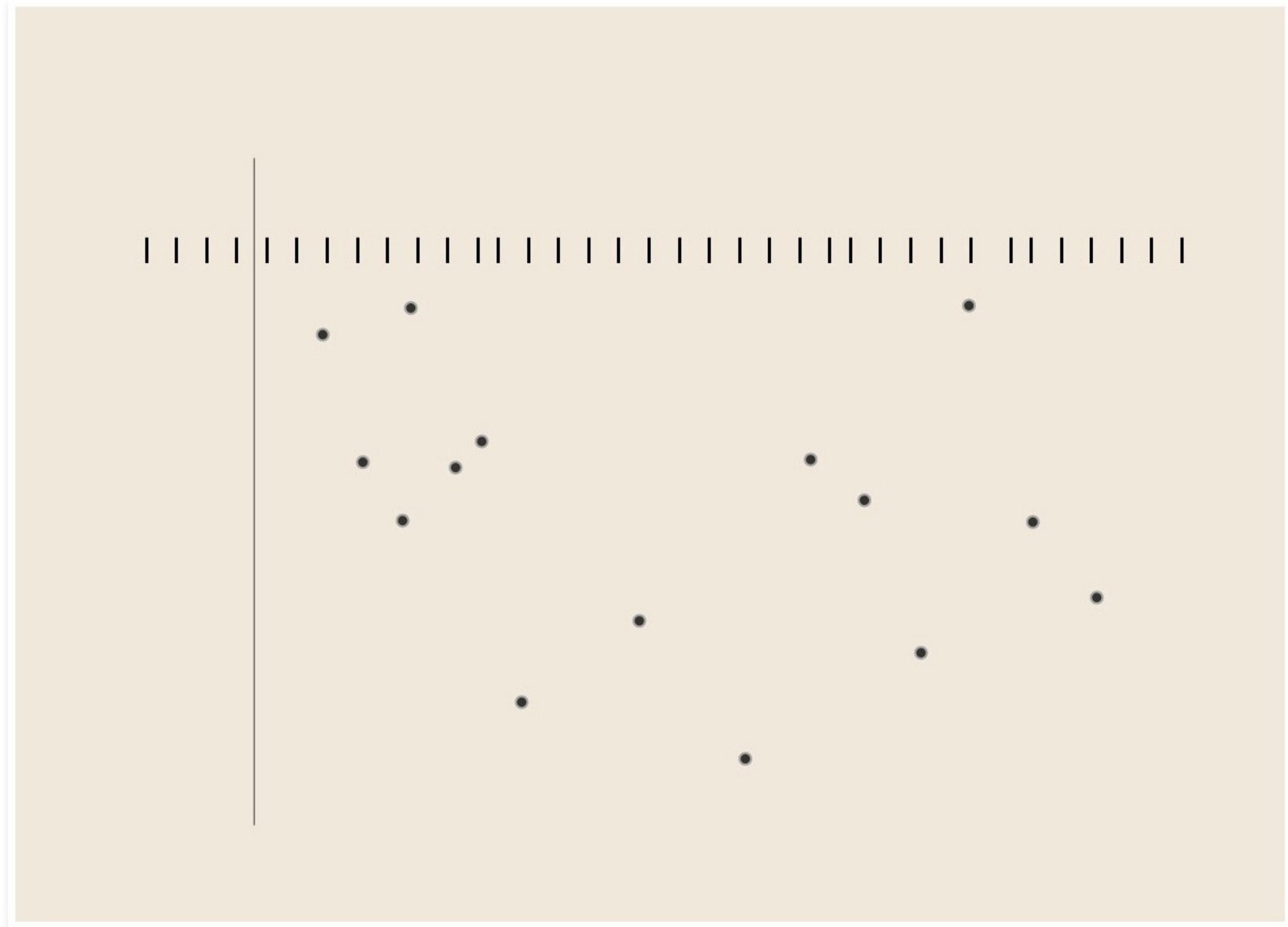
## [1] Composing the main architecture of the visualization

Composing the main architecture: this acts as the formalized base through which the main story will be mapped and displayed, upon this, one will see the most relevant patterns emerging from the story: the essential “map” that conceptually identifies where we are. This base is essentially a matrix or pattern that will serve as our organizer. It may be composed of cells, or distances, or other interrelated multiples.



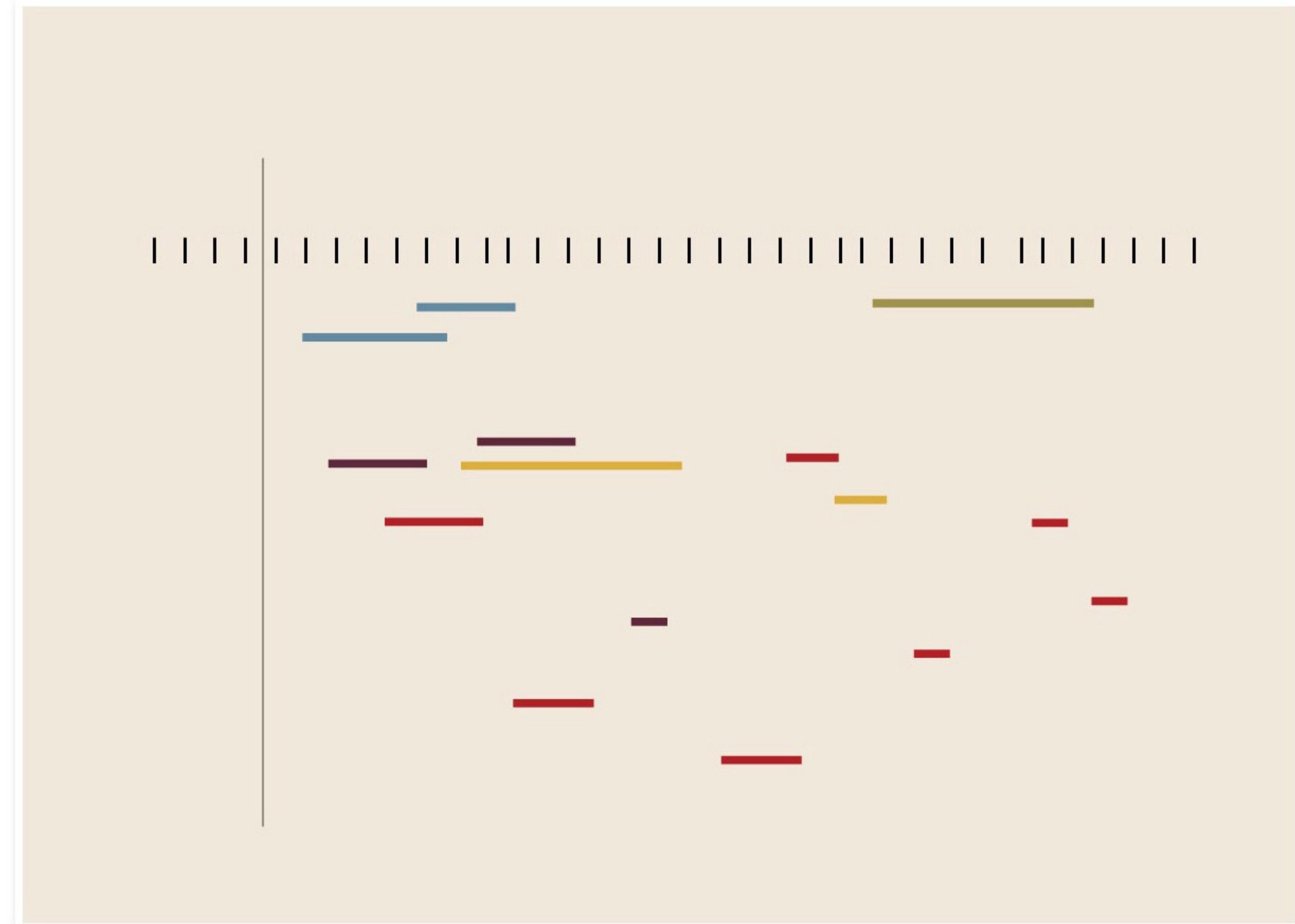
## [2] Positioning singular elements within the main framework.

This process will test the effectiveness of the main architecture; the placement of elements reveals or confirms weaknesses and strengths, which may lead to modification of the main architecture.



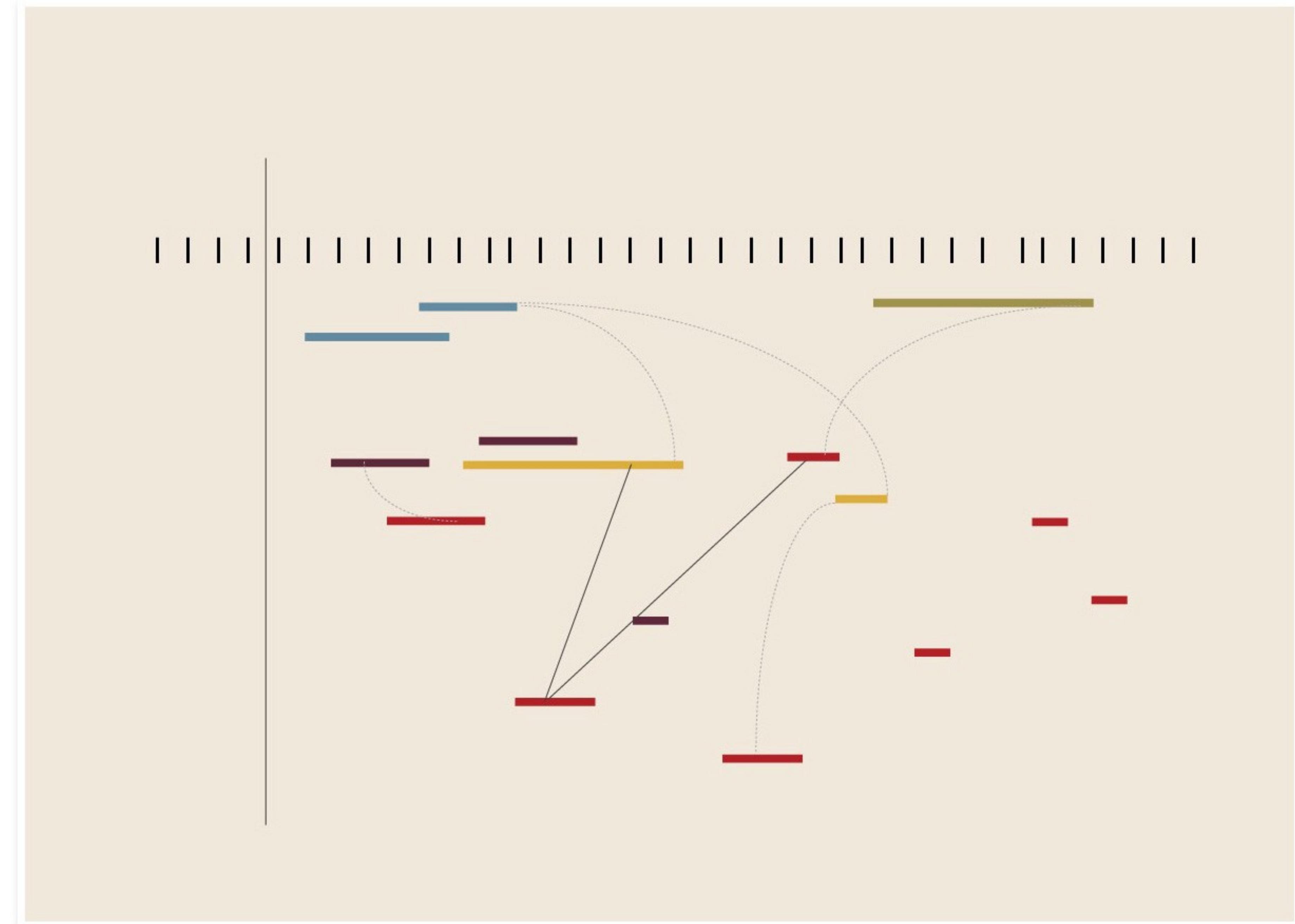
### [3] Constructing shaped elements of dimensionality and form

Constructing shaped elements of dimensionality and form (essentially polygons) with quantitative and qualitative parameters and positioning these within the main architecture. As these elements have form they must also be identified through colors according to opportunities to establish categorizations, thus advancing clarity and relationships that serve to enhance the story.



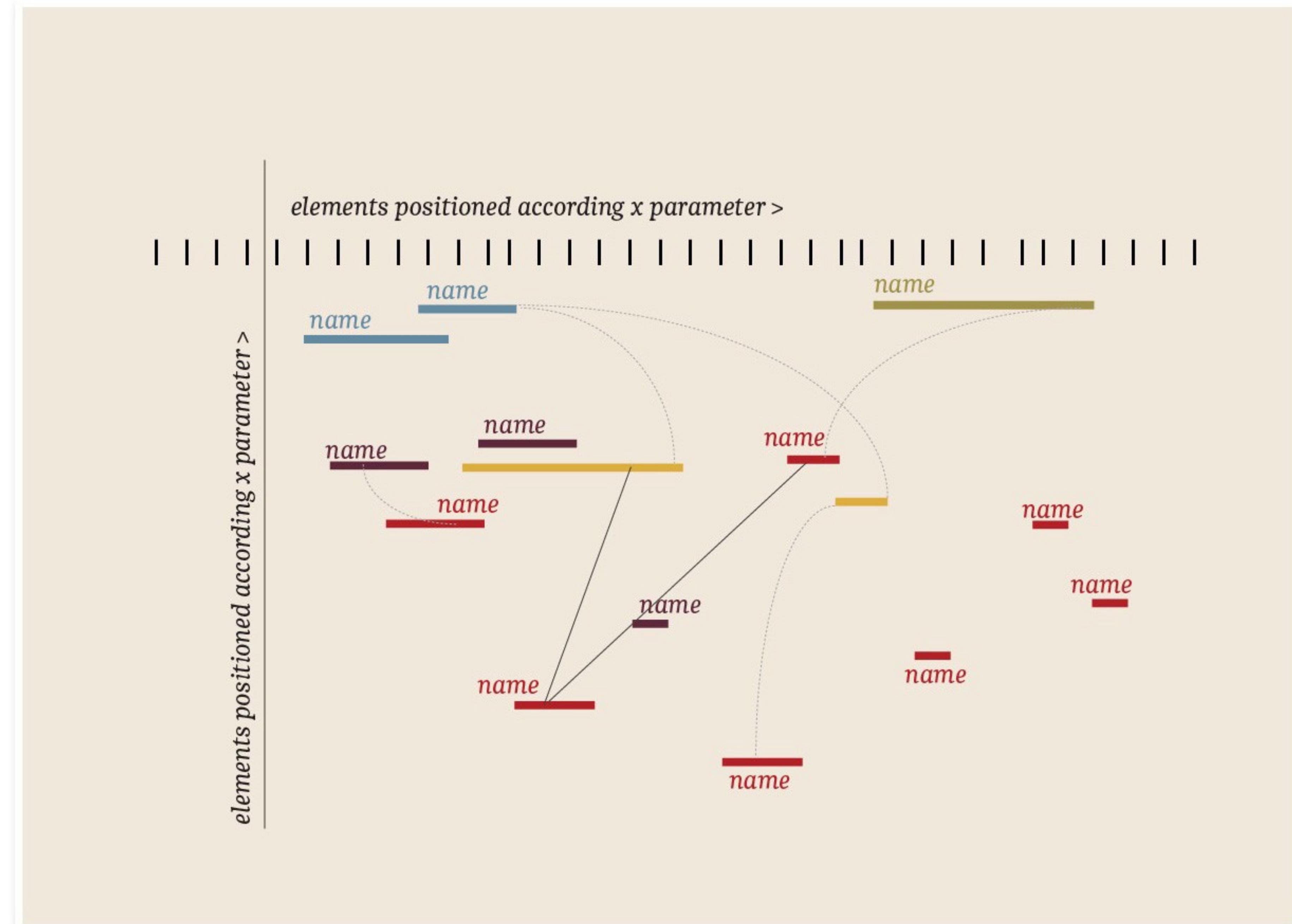
## [4] Elucidating internal relationships between elements (if any)

These links, directives, and qualifiers serve to give the story a comprehensive texture and correlate dependencies within the story.



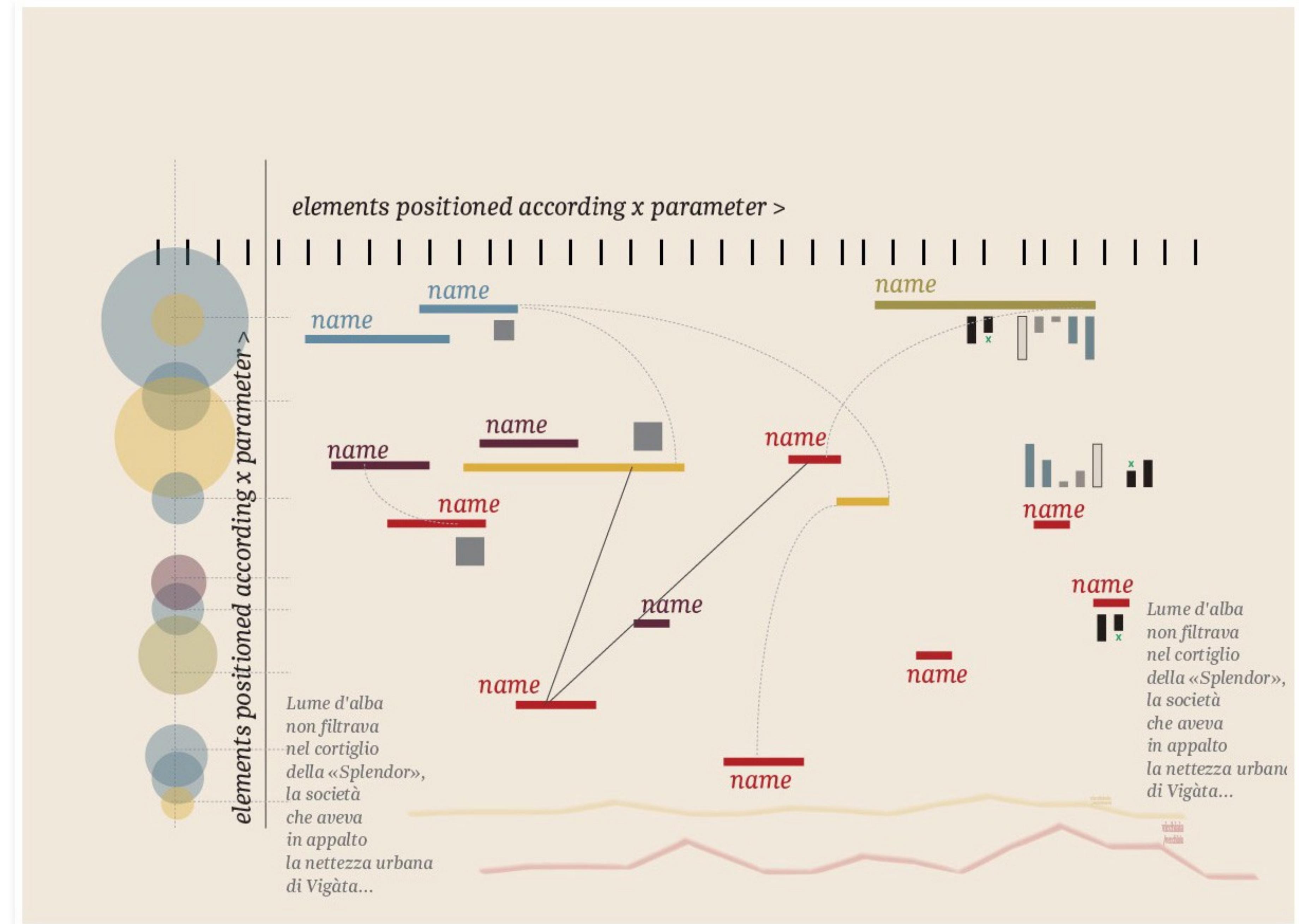
# [5] Labeling and identifying

Through the addition of explanatory labels and short texts we provide requisite last mile clarity throughout the presentation.



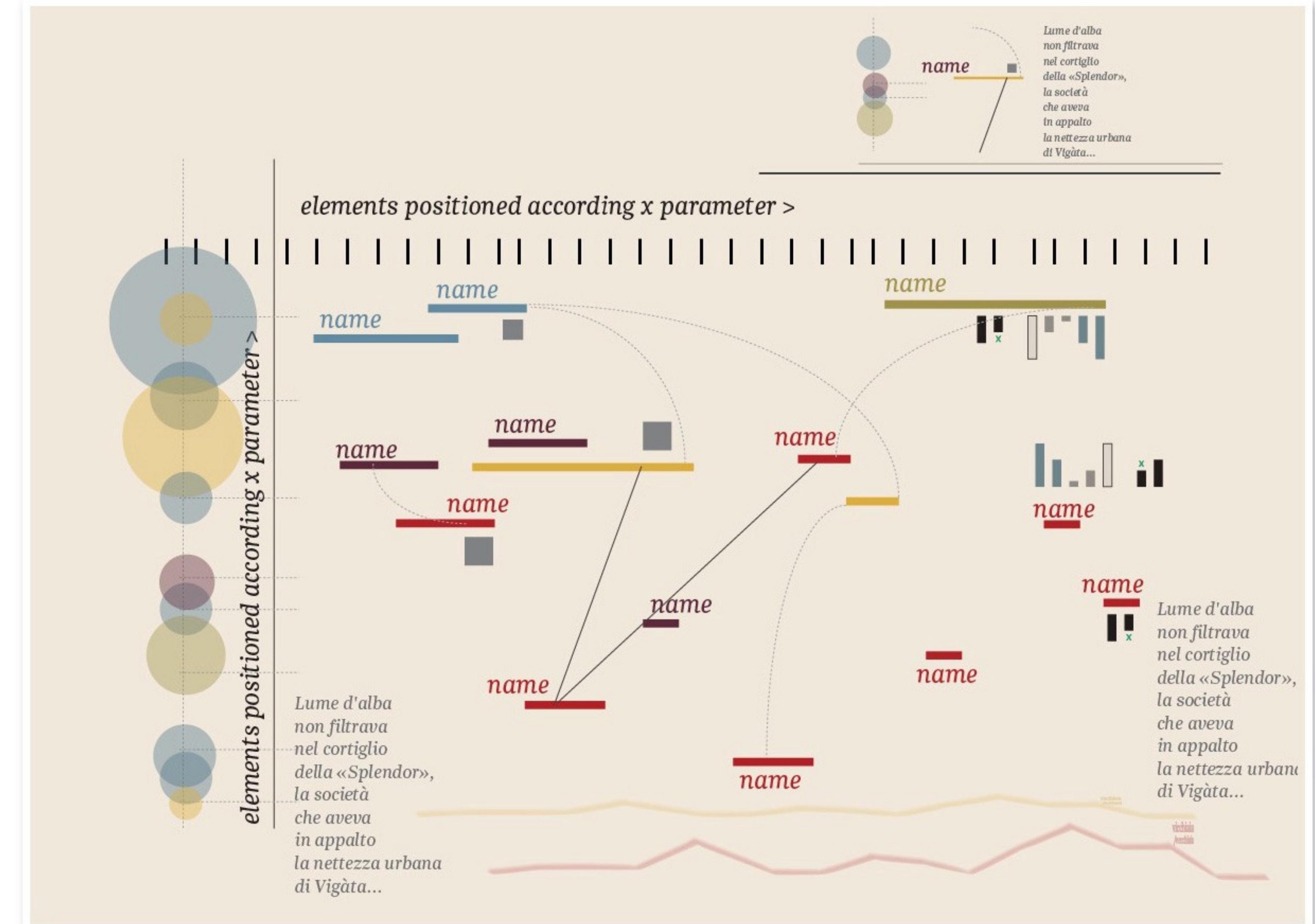
## [6] Supplementing the greater story through the addition of minor or tangential tales elements.

We consider this a very important step to contextualize the phenomena in a wider world. These components link the story to external ideas, other times, or other places. Elements that are rendered here may come from very diverse sources—analysis that is undertaken once we have strongly established the core story. These elements, which may take the form of small images, textual components, graphic symbols, etc., are to be located where they best help to enrich the overall comprehension: they must not distract from the main story.



**[7] Providing small visual explanations such as a legend or key that assists readers and the general public who may not be familiar with norms of data visualization.**

These are composed to enlighten the layered idea of the visualization, often constructed as miniatures of the layers themselves. The process usually involves simplification of the general architecture (e.g. the x and y axes, base timelines, or map components) as well as minimal explicit shapes, colors, and dimensions of singular elements. These explanations also provide units of measurement for distances and volumes.



# [8] Fine-tuning and stylizing of elements' shapes, colors, and weights to make hierarchies pop out.

By visually highlighting the most relevant elements and lightening the other background layers of information, we should be able to allow information to be selectively and sequentially revealed, helping readers discover stories by themselves and recognizing the patterns or interrelationships from one element within the story to another.

