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

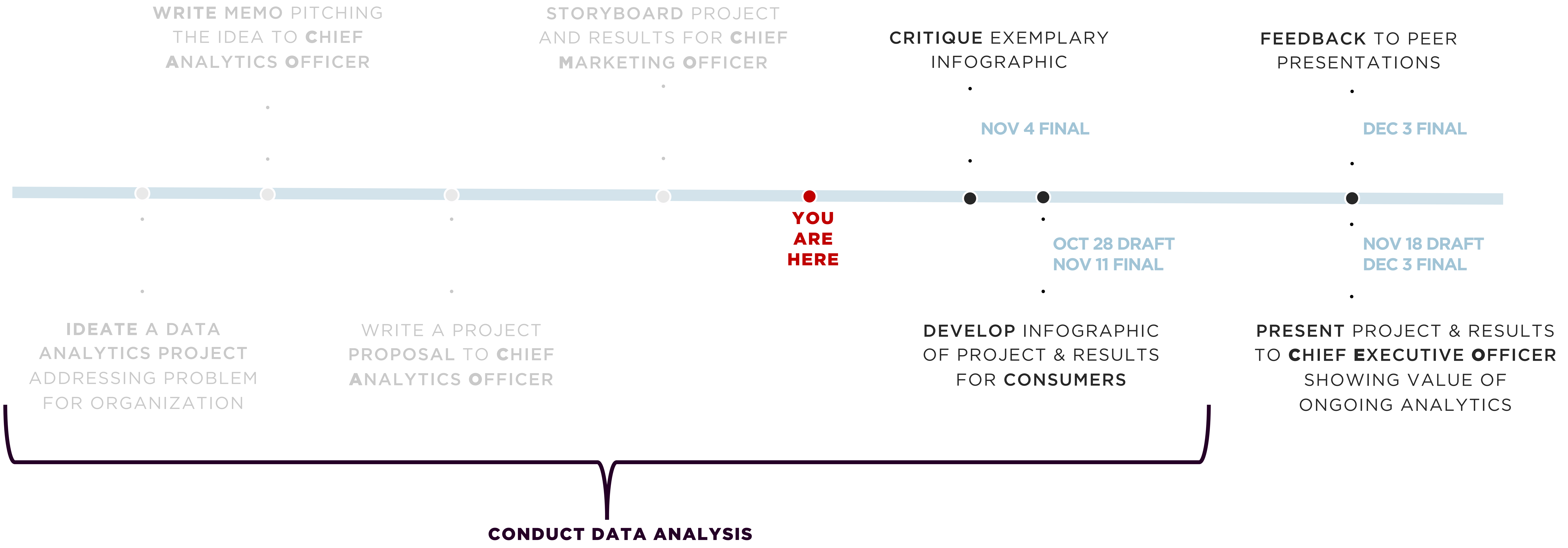
From graphics back to narrative

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

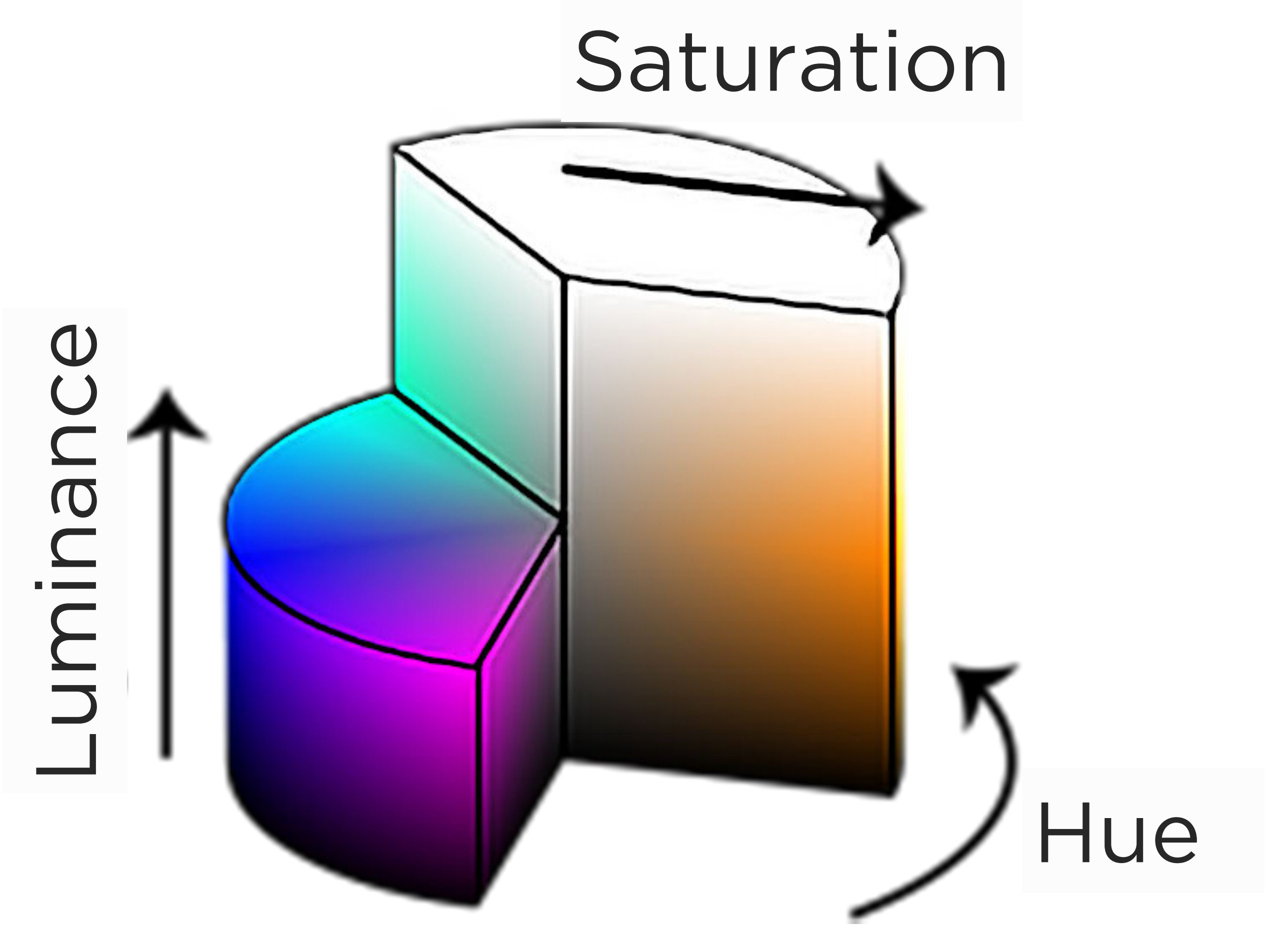
Communicating data with color: perception of values

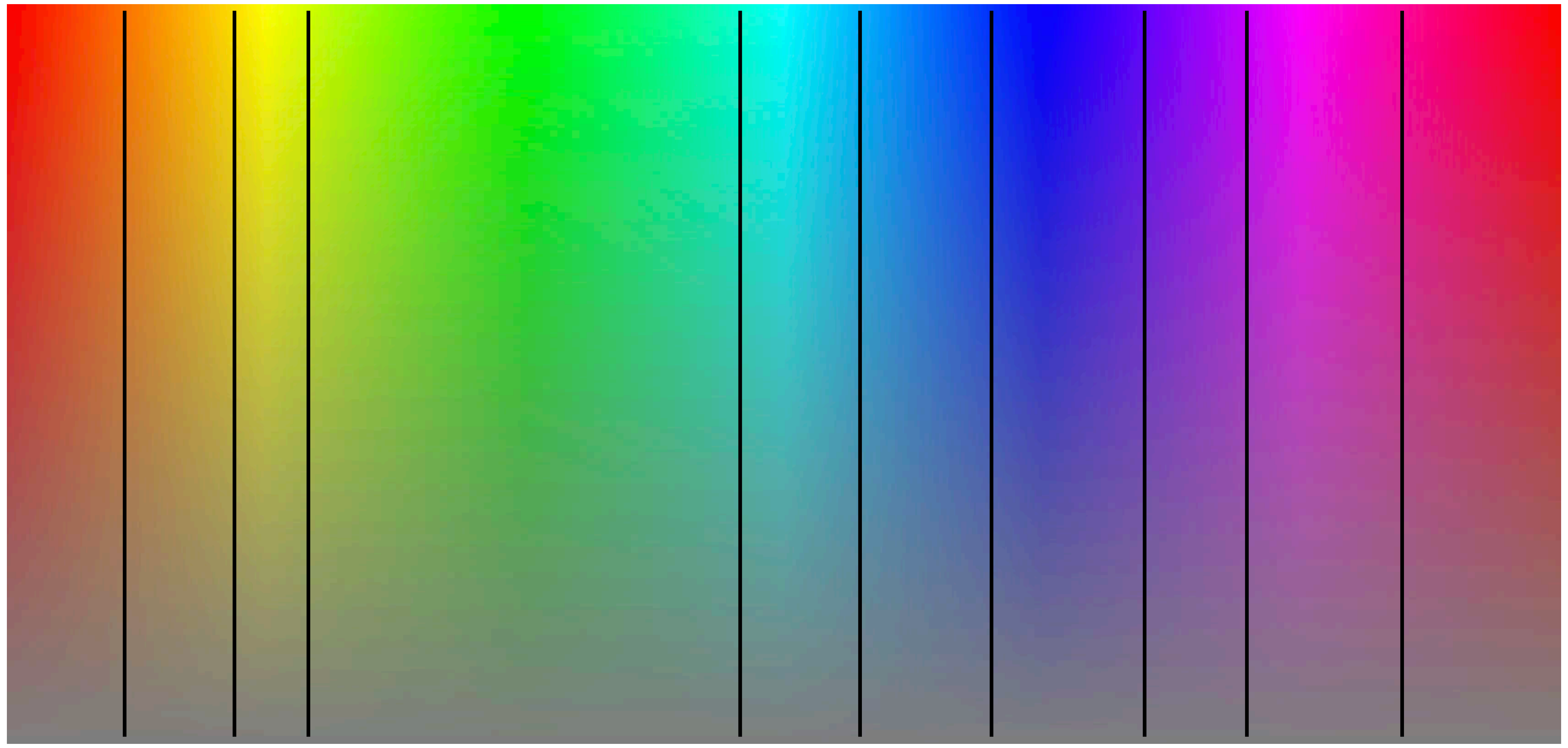
We can represent data using *color spaces*, which are mathematical models.

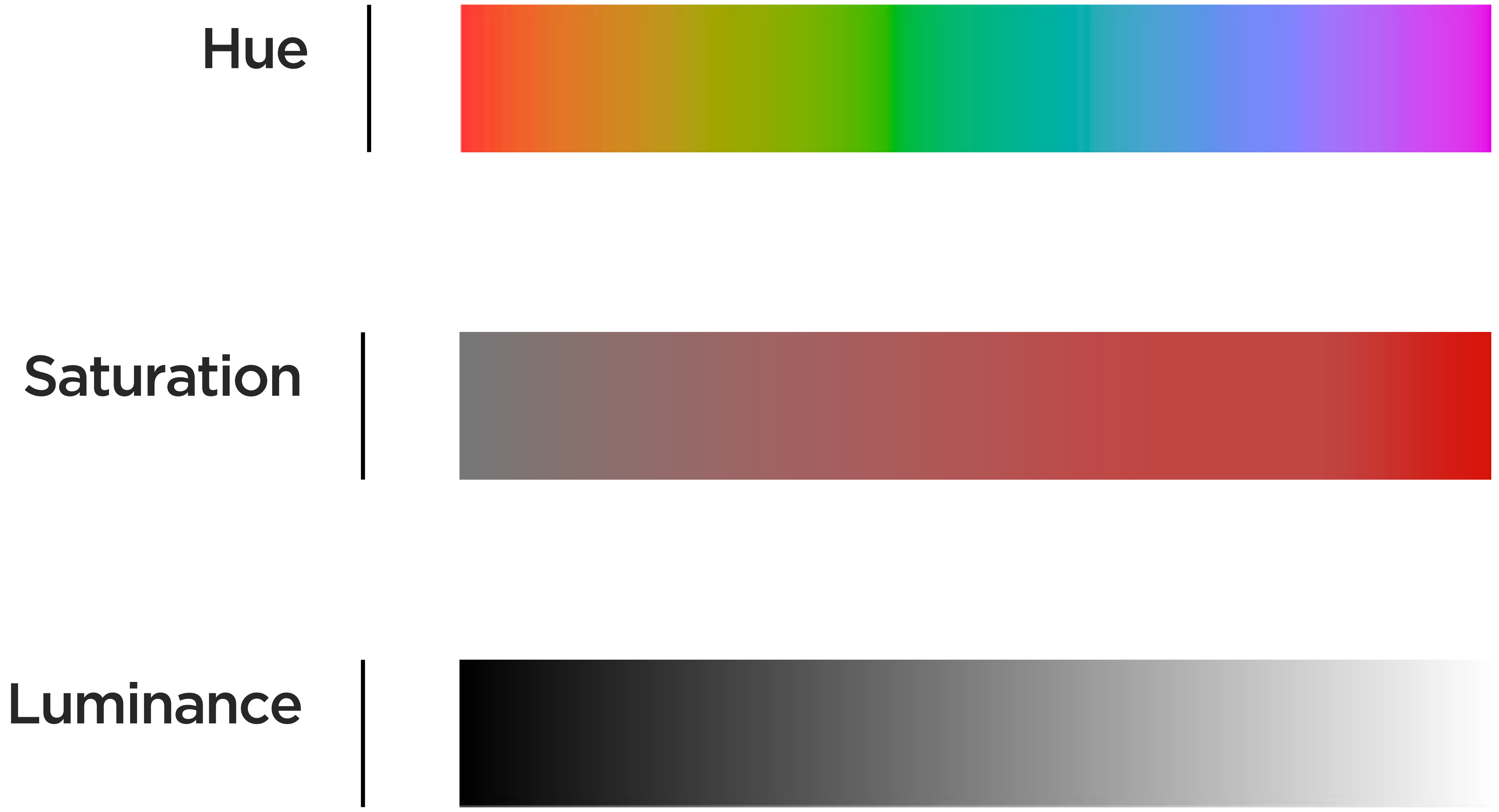
The common color model **RGB** has three dimensions — red, green, and blue — where those hues are mixed to produce a specific color. When defining colors in these dimensions, one has to know the sequence of colors in the color spectrum, e.g. mixing pure red and green makes yellow.

HSL is another color space that has mapped the same RGB color space into the alternate dimensions of hue, saturation, and luminance.

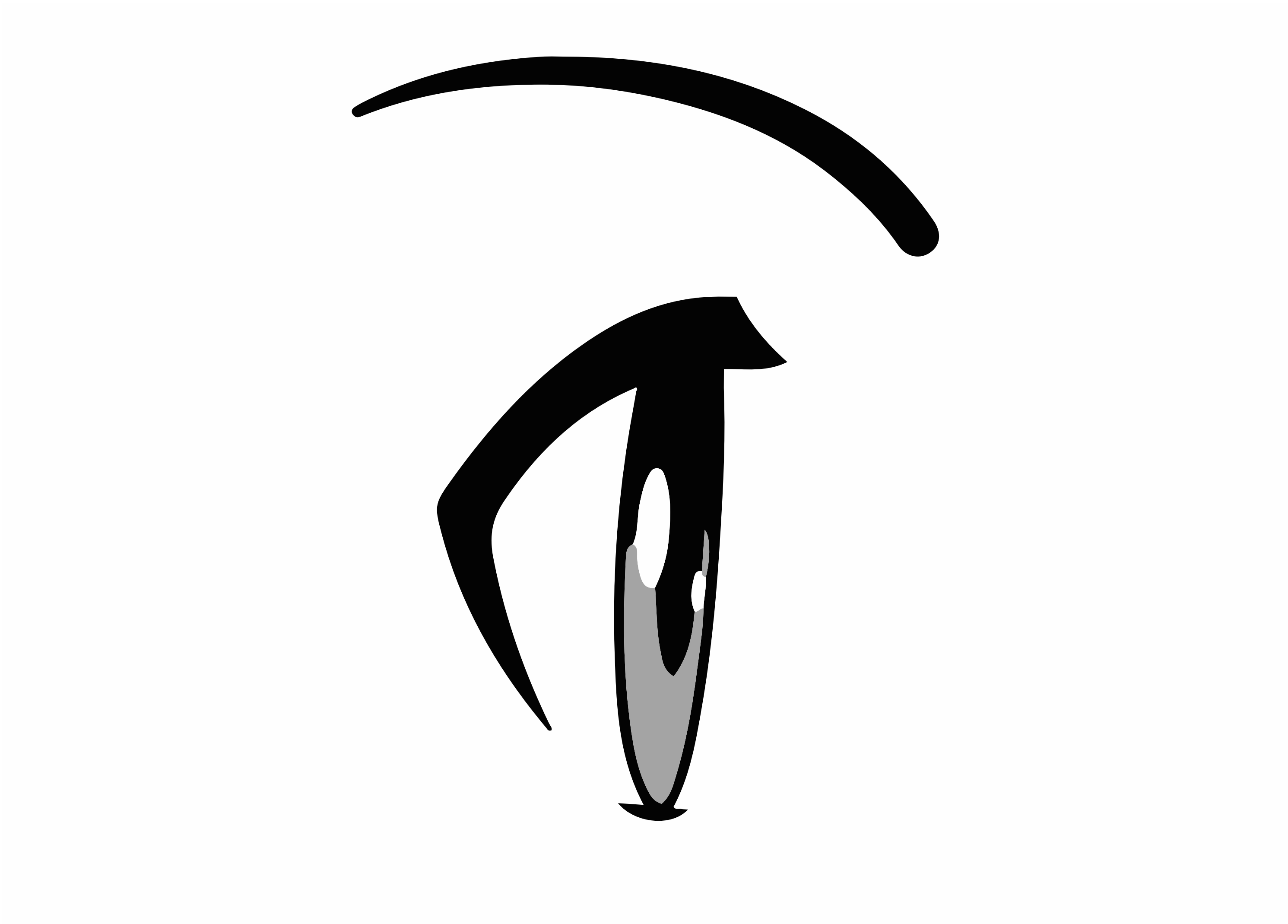
Notice this color space — shown below — has **uneven distances** between colors, **and** what we perceive as **brightness** is also uneven from left to right.

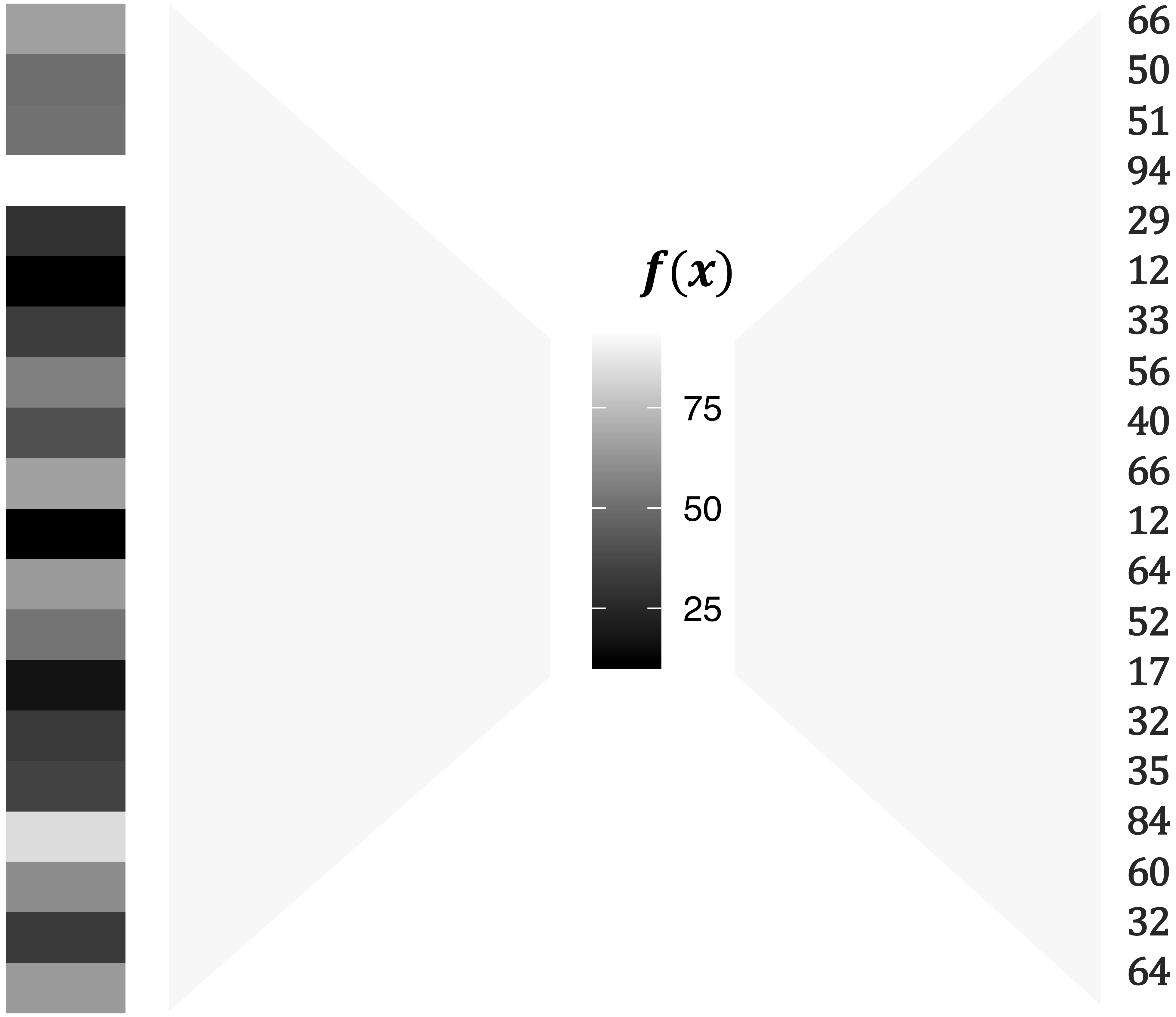


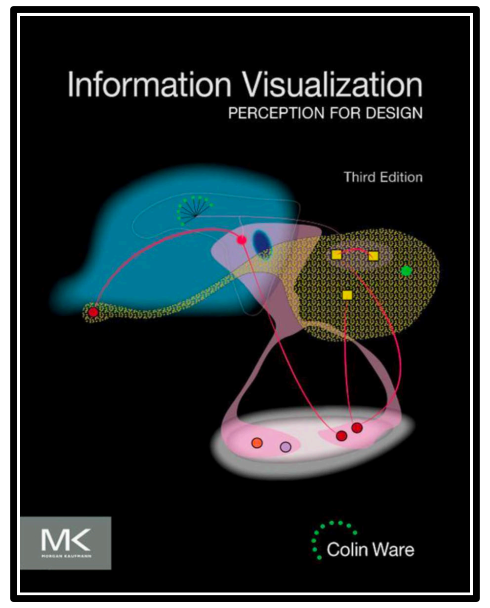




How can we map data to light, whether using its **hue**, saturation, or luminance?







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.

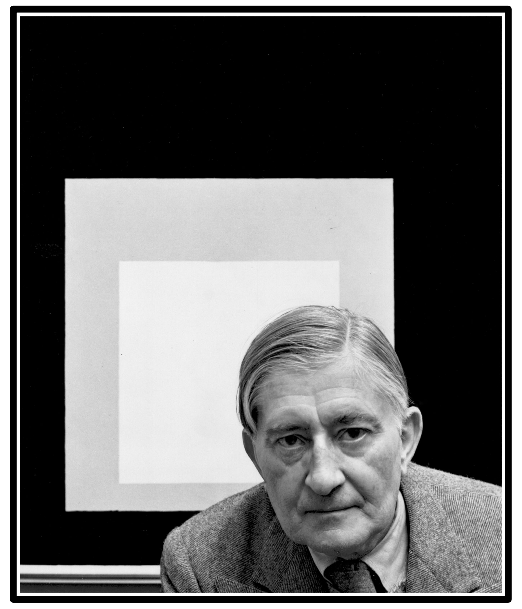
Brightness perception is a nonlinear function of luminance.

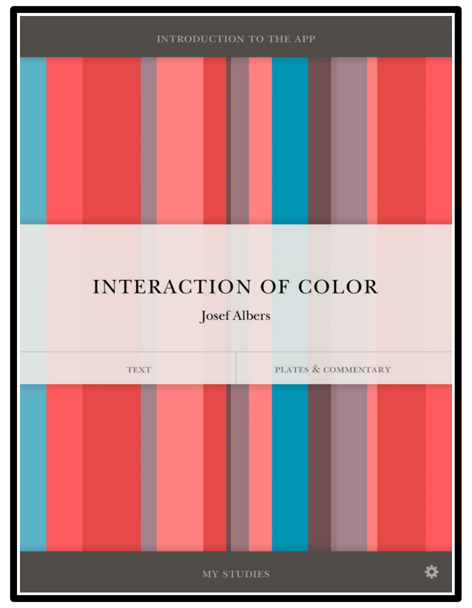
Perceived brightness is a very nonlinear function of the amount of light emitted. That function follows the power law:

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

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.

Ware, Colin. *Information Visualization*. Third. Elsevier, 2013. Print.





Interaction of Color

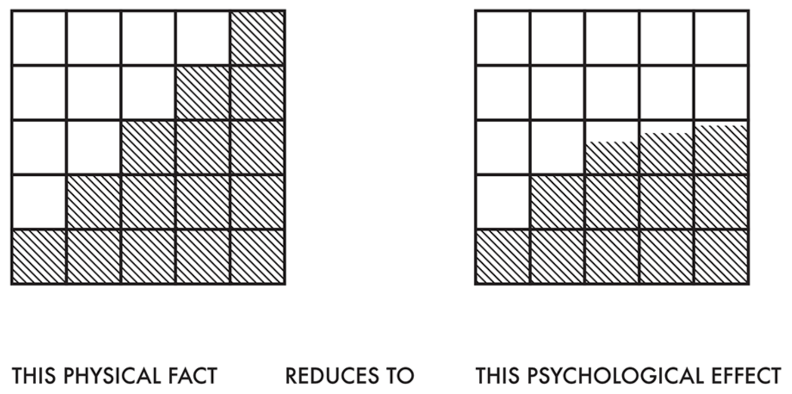
Albers

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

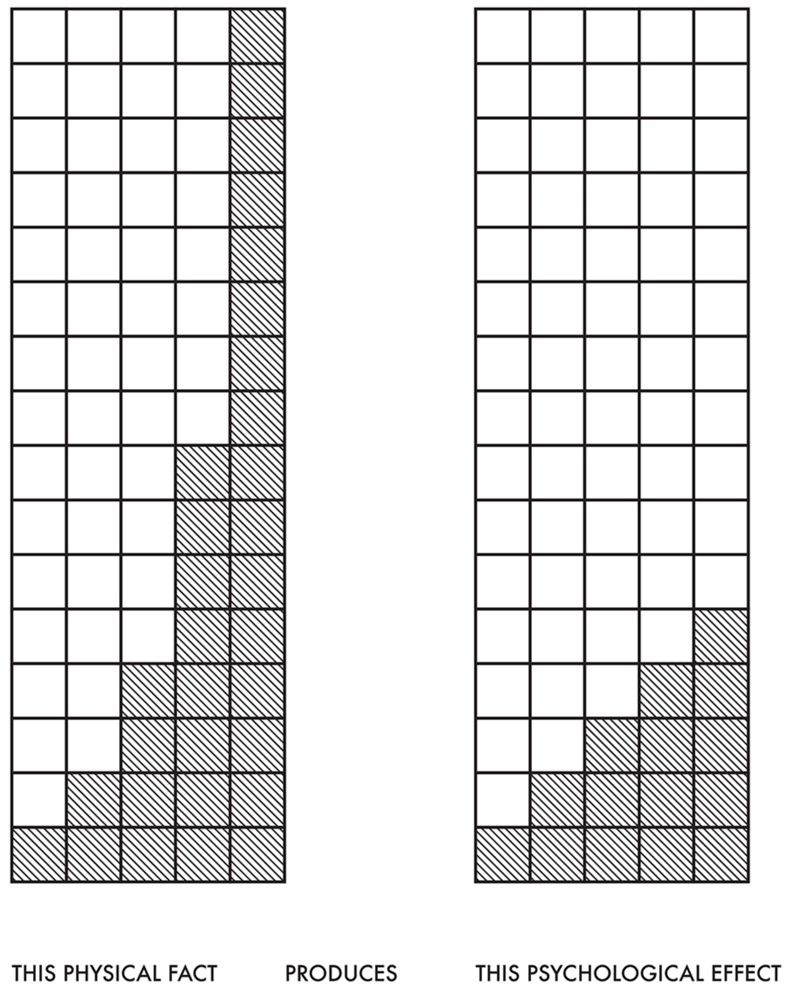
Color intervals are the distance in light intensity between one color and another, analogous to musical intervals (the relationship between notes of different pitches).

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

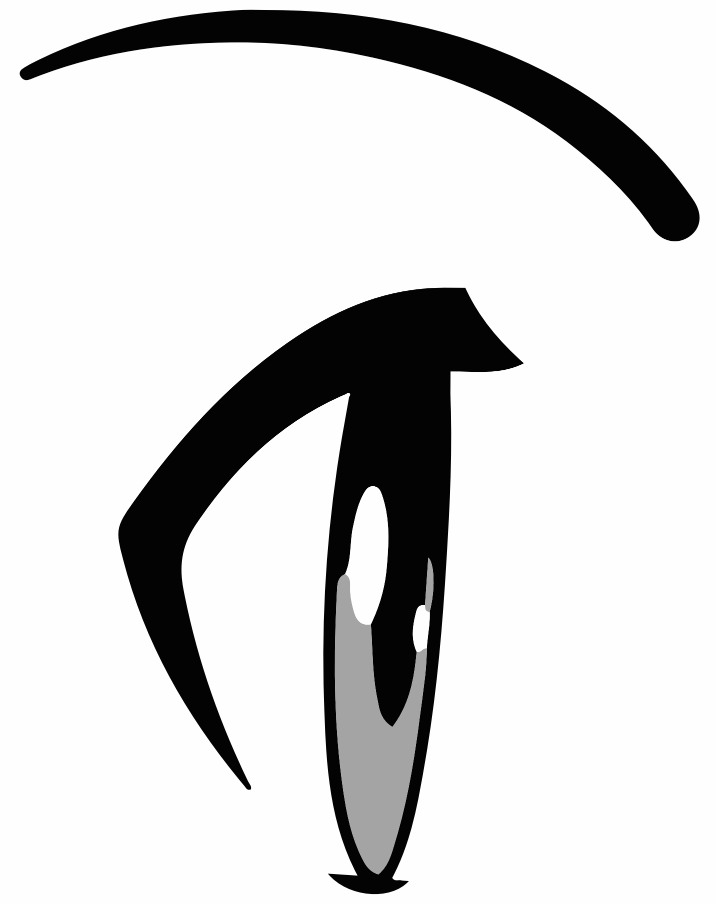


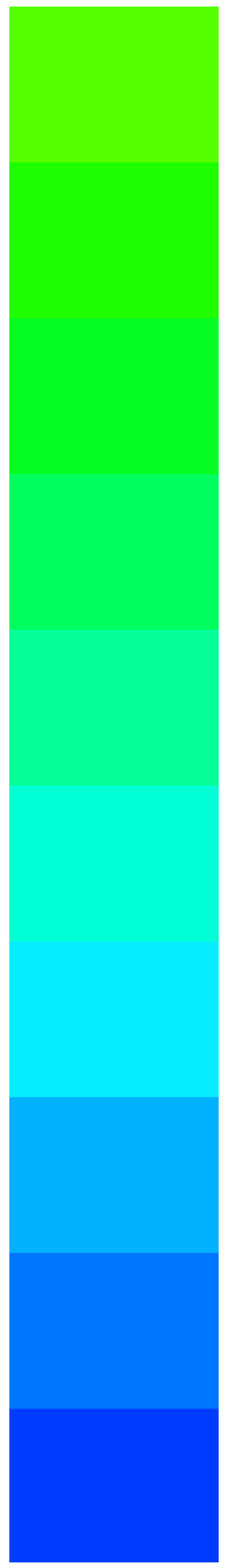
The visual perception of an arithmetical progression depends upon a physical geometric progression.

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.



Albers, Josef. *Interaction of Color*. Yale University Press, 2006. Print.

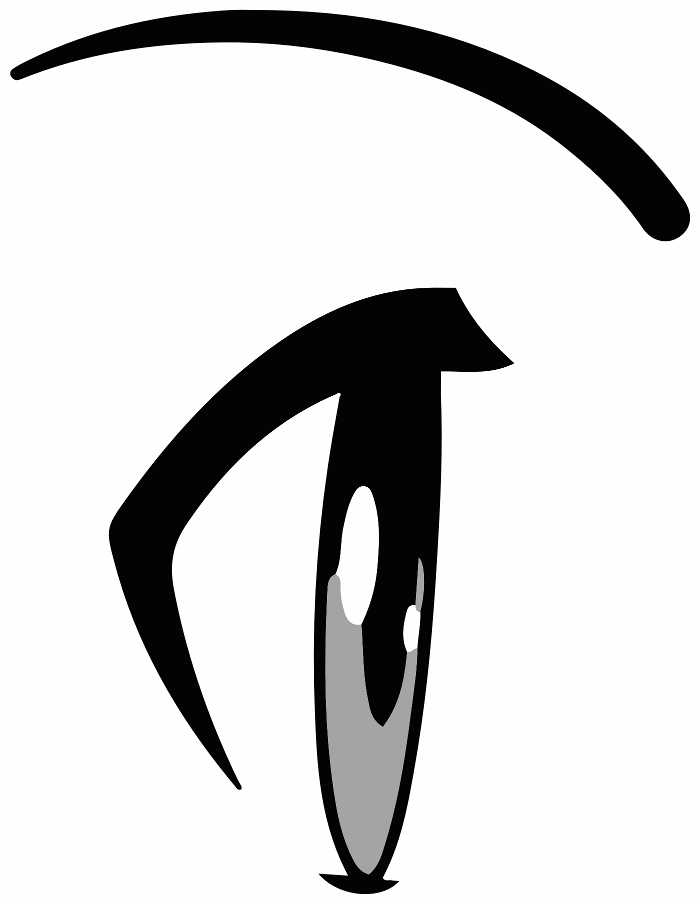


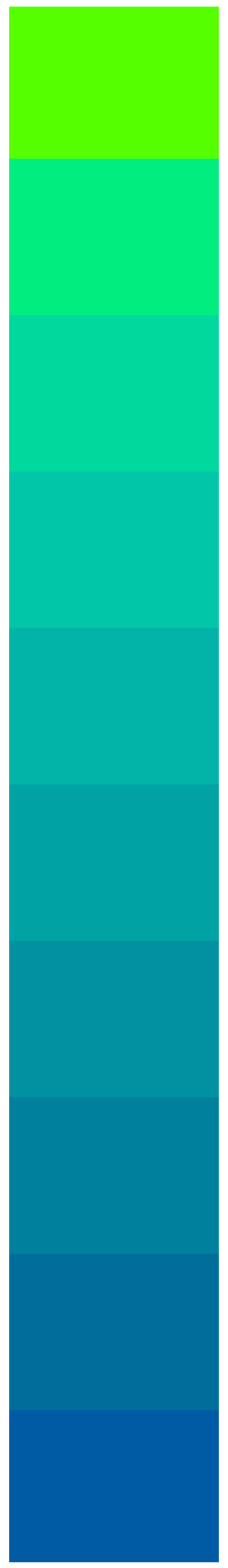


Changing the hue values in the sRGB color space from green to blue in 10 equal steps fails to uniformly scale.

The relatively wide band of green in this color space results hues we perceive as almost identical, while the blue colors are more diverse. 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.

This happens because the default sRGB color space (and any color model built on it like HSV and HSL) is irregular, which means that **even though the rectangles have evenly spaced hue values, the corresponding effect is not linear to the human eye**.





But other color spaces are *perceptually uniform*, better for representing data.

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. To create a color in the CIELUV color space, one has to define the lightness of the color (l), whether it is reddish or greenish (u), and whether it is yellowish or bluish (v).

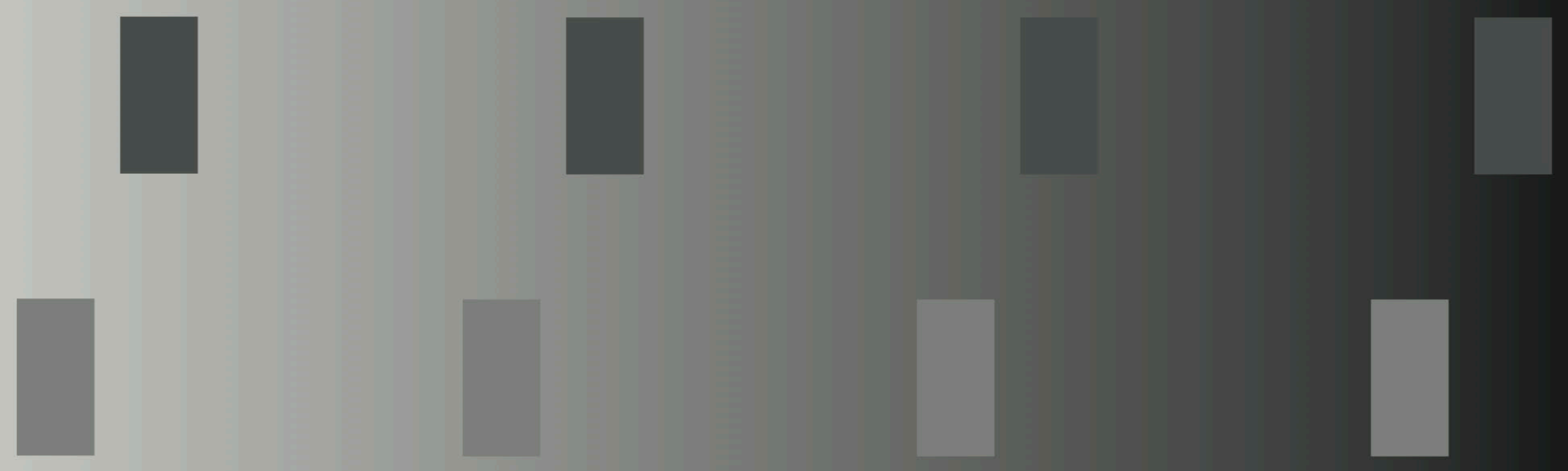
More modern implementations, like **HSLuv** improve upon this work by mapping colors as perceived into the familiar and intuitive HSL dimensions. **To the left**, our green-to-blue hues in 10 equal steps using the HSLuv model are perceptually uniform.

You can learn more about CIELUV at https://en.wikipedia.org/wiki/CIELUV and HSLuv at http://www.hsluv.org

Humans have evolved to see edge contrasts, like here: We see comparative — not absolute — luminance value.

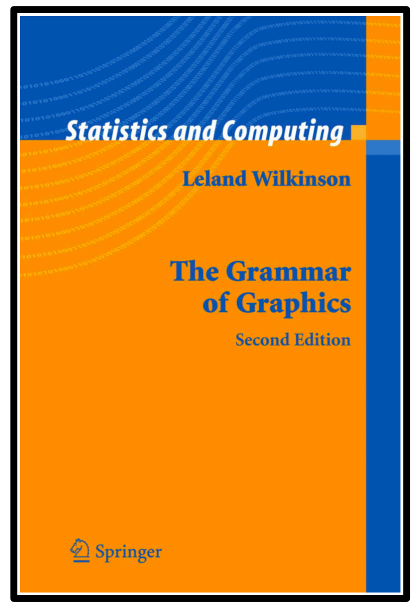
Ware, Colin. *Information Visualization*. Third. Elsevier, 2013. Print.

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.



Ware, Colin. *Information Visualization*. Third. Elsevier, 2013. Print.

Graphics have a grammar



The grammar of graphics

Wilkinson

Leland, formerly VP of 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

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

Theory of graphics, not chart typology

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.

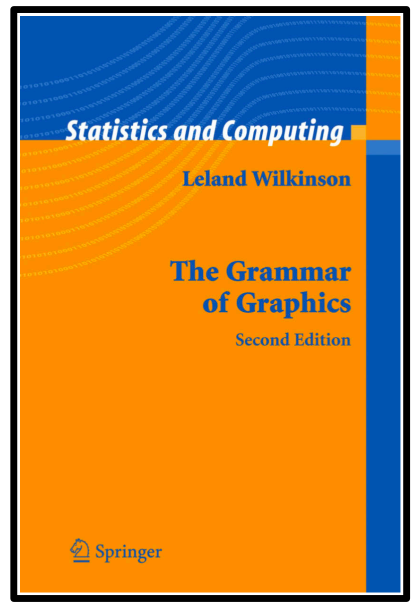
Charts are just instances of more general objects

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**.

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

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**.

Wilkinson, Leland. *The Grammar of Graphics*. Second. Springer, 2005. Print.



The grammar of graphics

Wilkinson

Leland, formerly VP of 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)

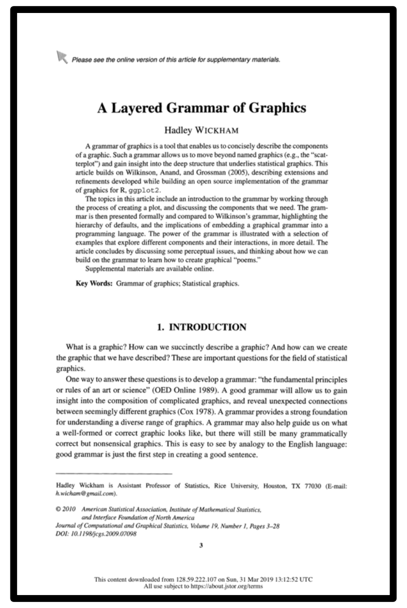
Statistics the functions that allow graphs to change their appearance and representation

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.

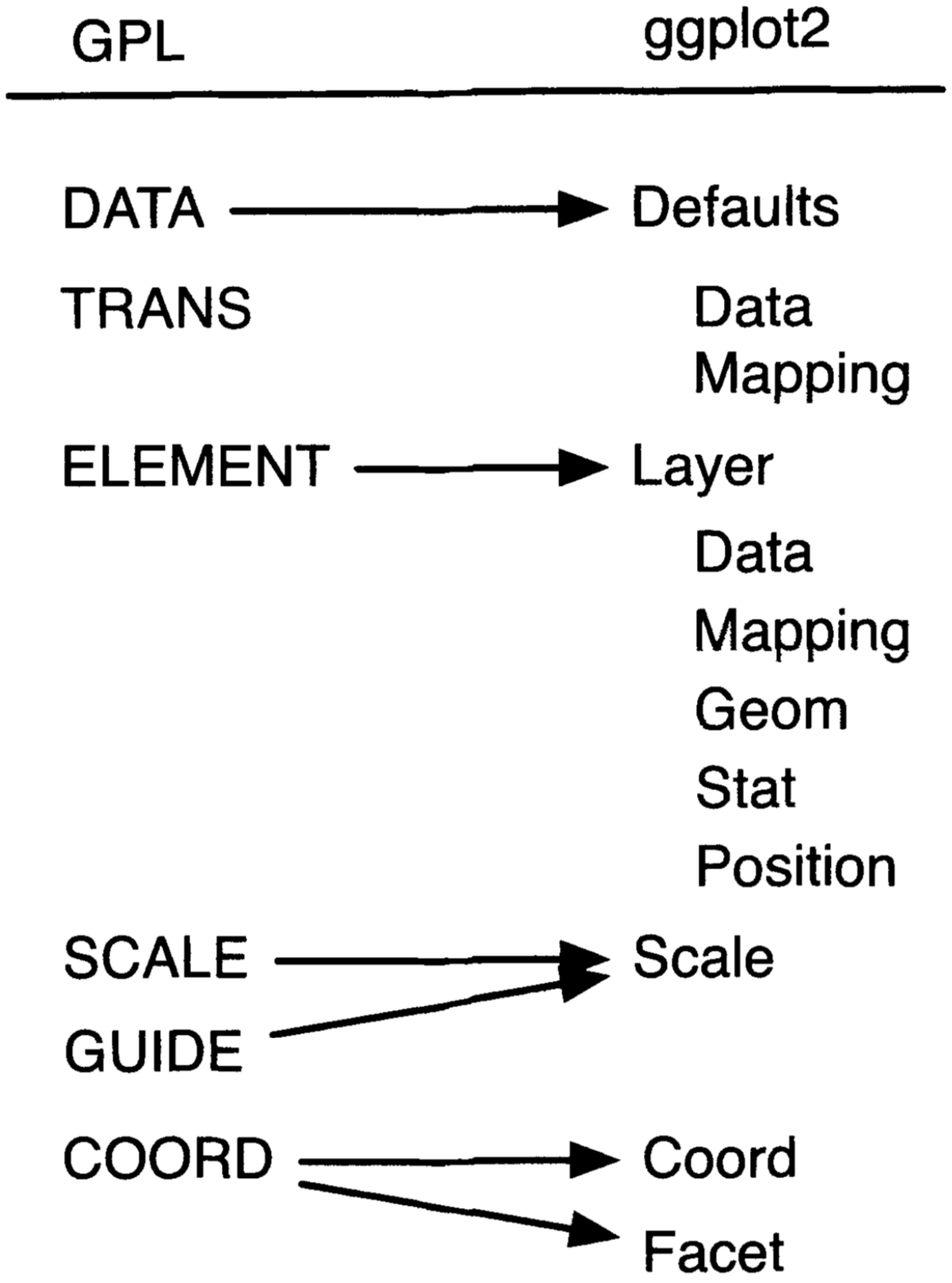
Wilkinson, Leland. *The Grammar of Graphics*. Second. Springer, 2005. Print.



A layered grammar of graphics

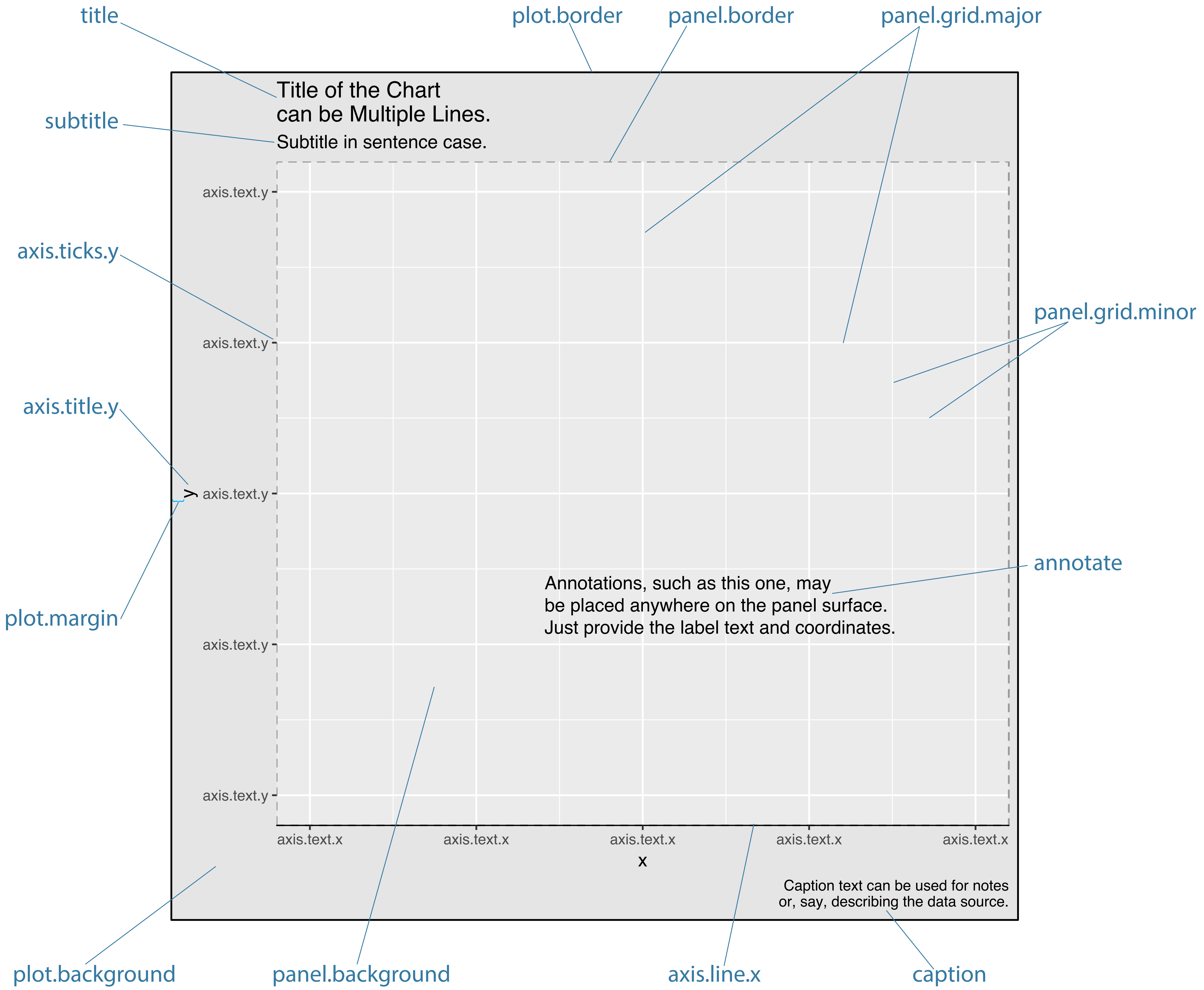
Wickham

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.



Wickham, Hadley. "A Layered Grammar of Graphics." Journal of Computational and Graphical Statistics 19.1 (2010): 3–28. Web.

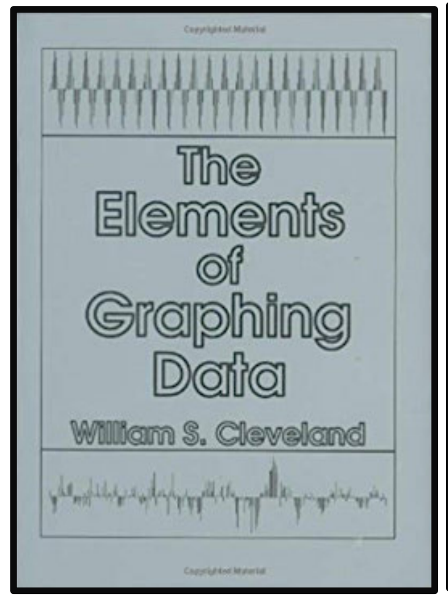
ggplot2, an R library implements this grammar of graphics.

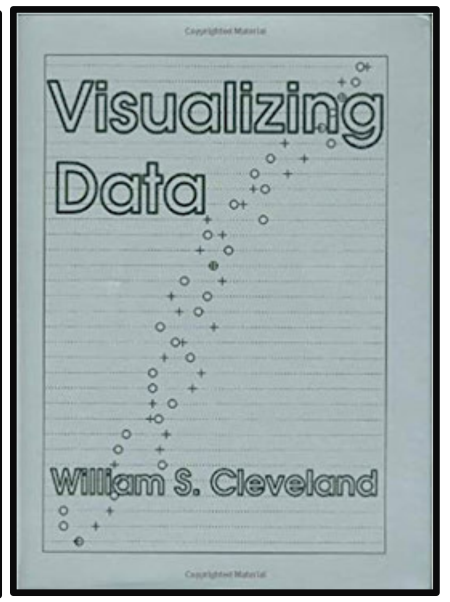



```
# load grammar of graphics
library(ggplot2)

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(<...> = <...>)
```

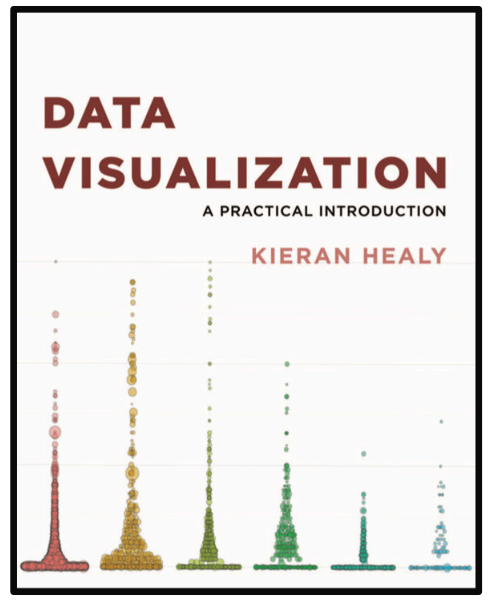
Learning & References

——— visual relationships in data ———



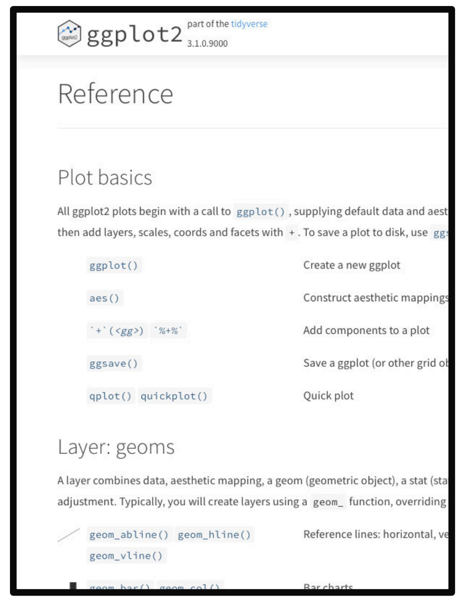


Implementation in R



socviz.co

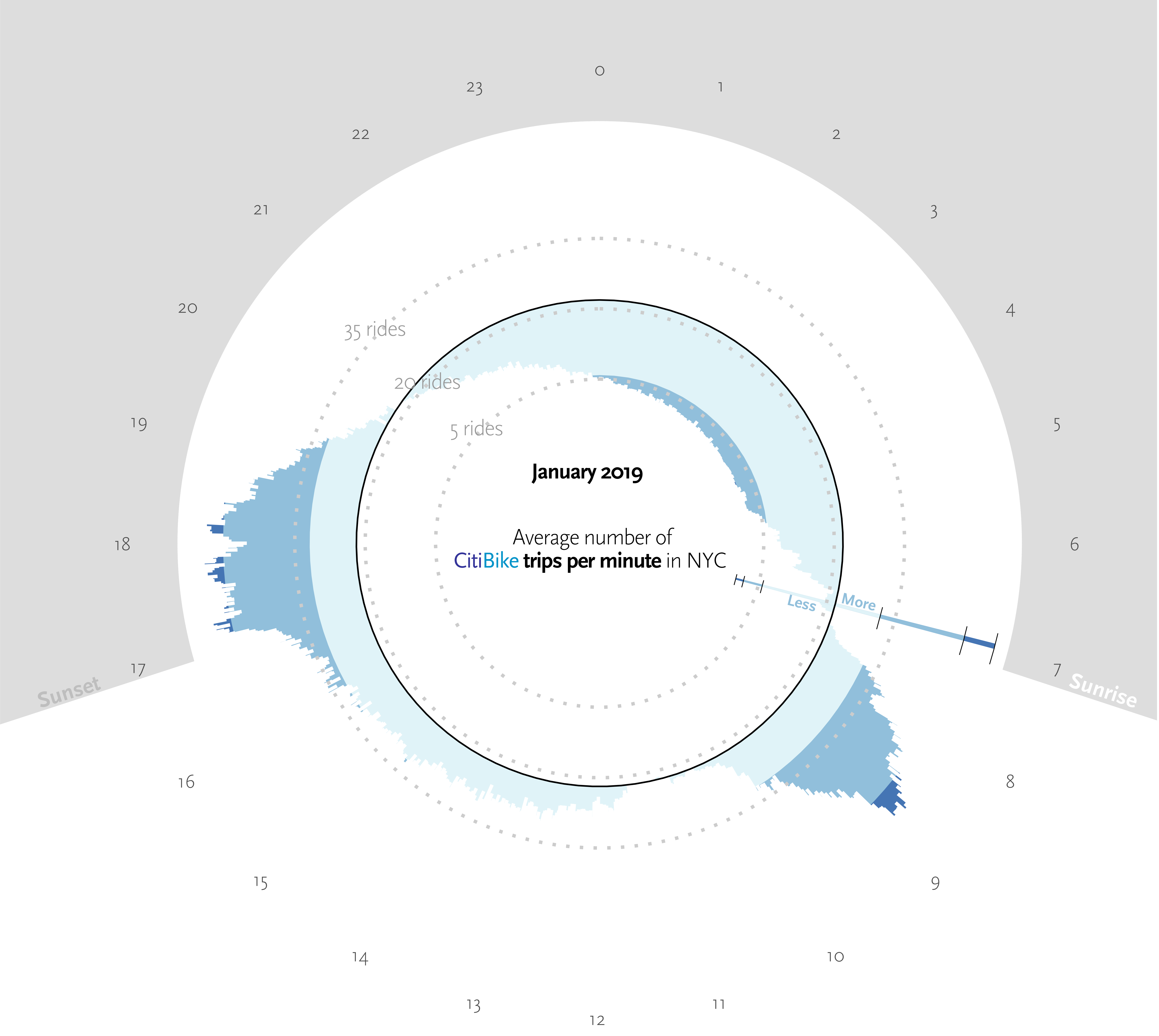
ggplot reference



ggplot2.tidyverse.org

Spencer, Scott. *Quantitative Persuasion Amid Uncertainty*. Forthcoming. Print.

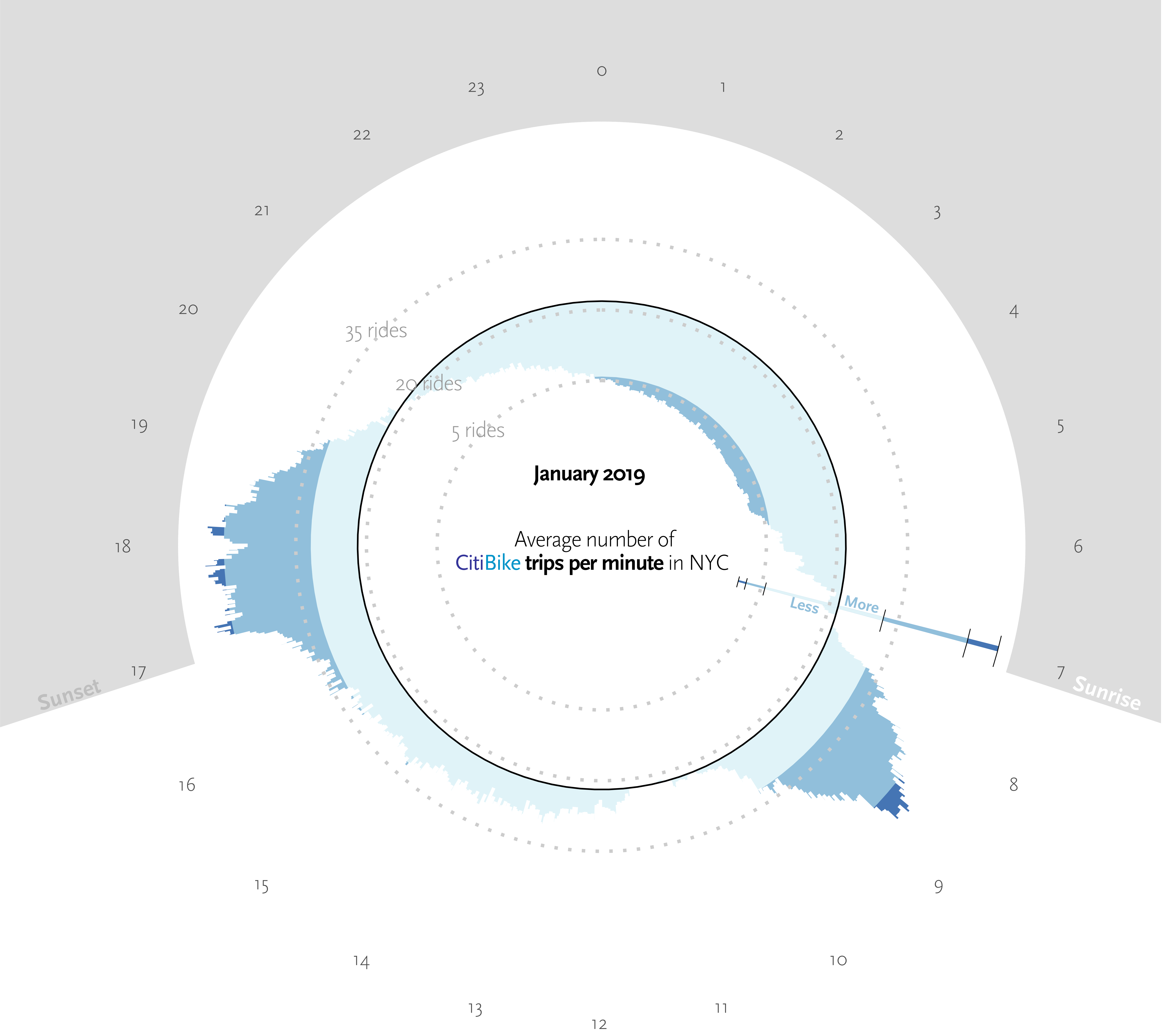
Think about graphics as l/a/y/e/r/s:



Spencer, Scott. *Exploring CitiBike Ride Data*. 28 Feb 2019. Web. https://ssp3nc3r.github.io/post/exploring-citibike-ride-data/

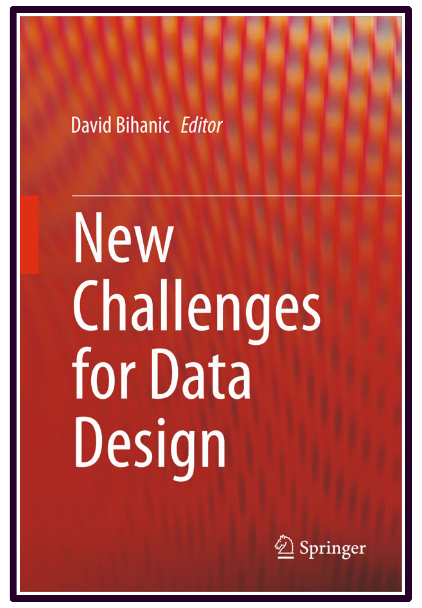
Think about graphics as l/a/y/e/r/s:

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



Spencer, Scott. *Exploring CitiBike Ride Data*. 28 Feb 2019. Web. https://ssp3nc3r.github.io/post/exploring-citibike-ride-data/

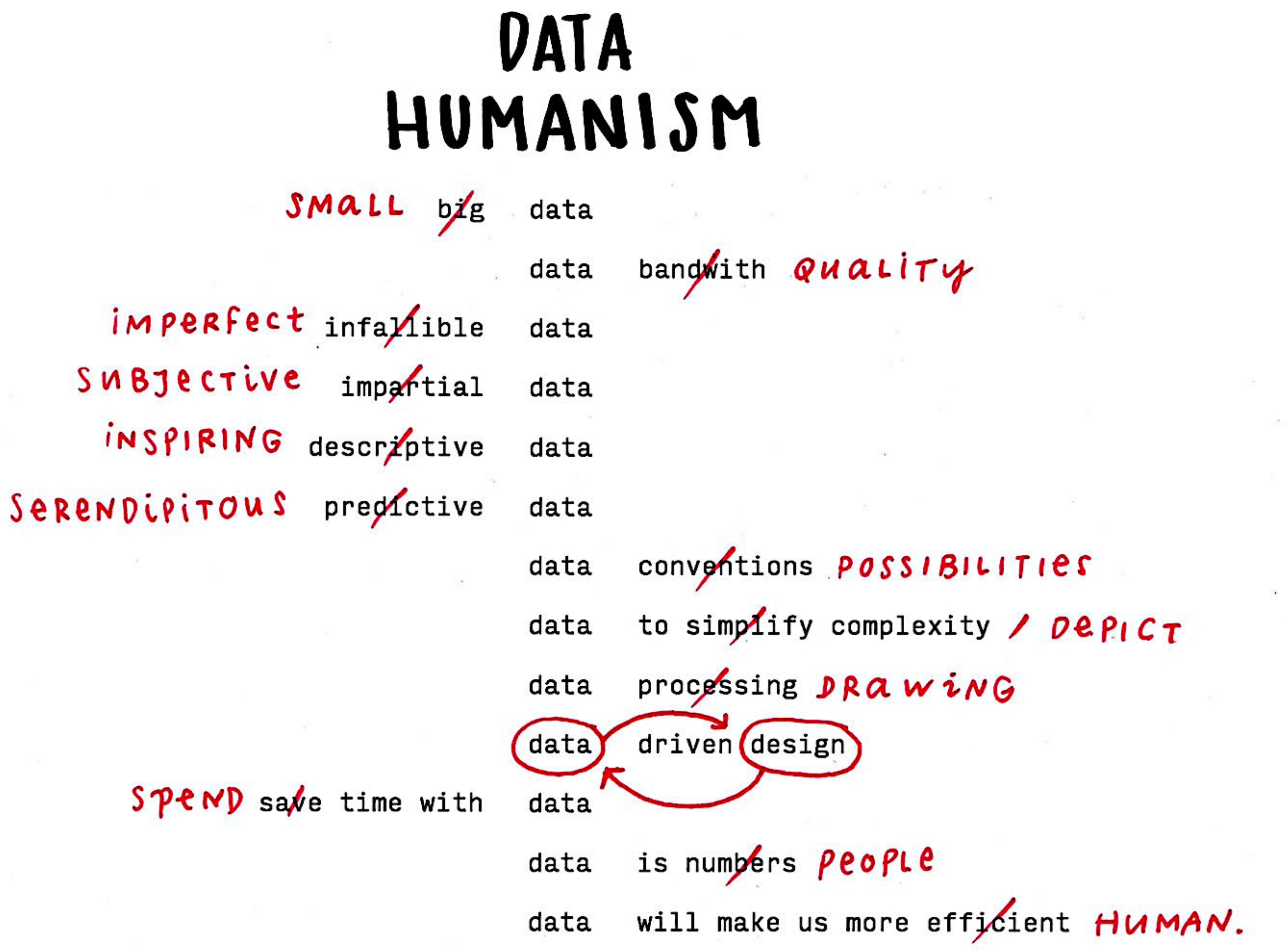
Hierarchies, layers are fundamental to graphics



The New Aesthetic of Data Narrative

Lupi

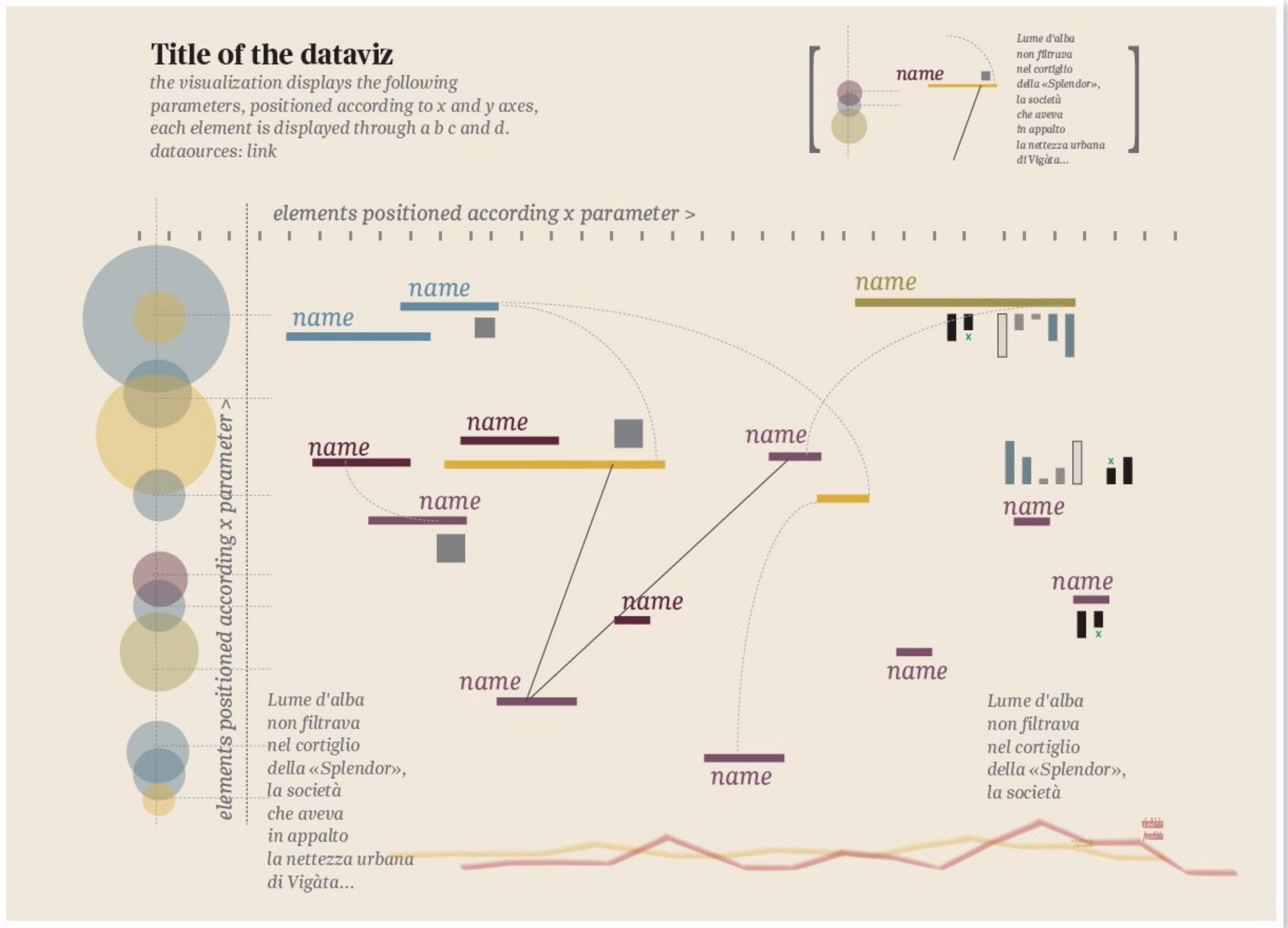
Her work is part of the permanent collection of the Museum of Modern Art. She is co-founder and Design Director of Accurat, a data-driven design firm. After receiving her master's degree in Architecture, she earned her PhD in Design at Politecnico di Milano.



Create displays with layers, hierarchies

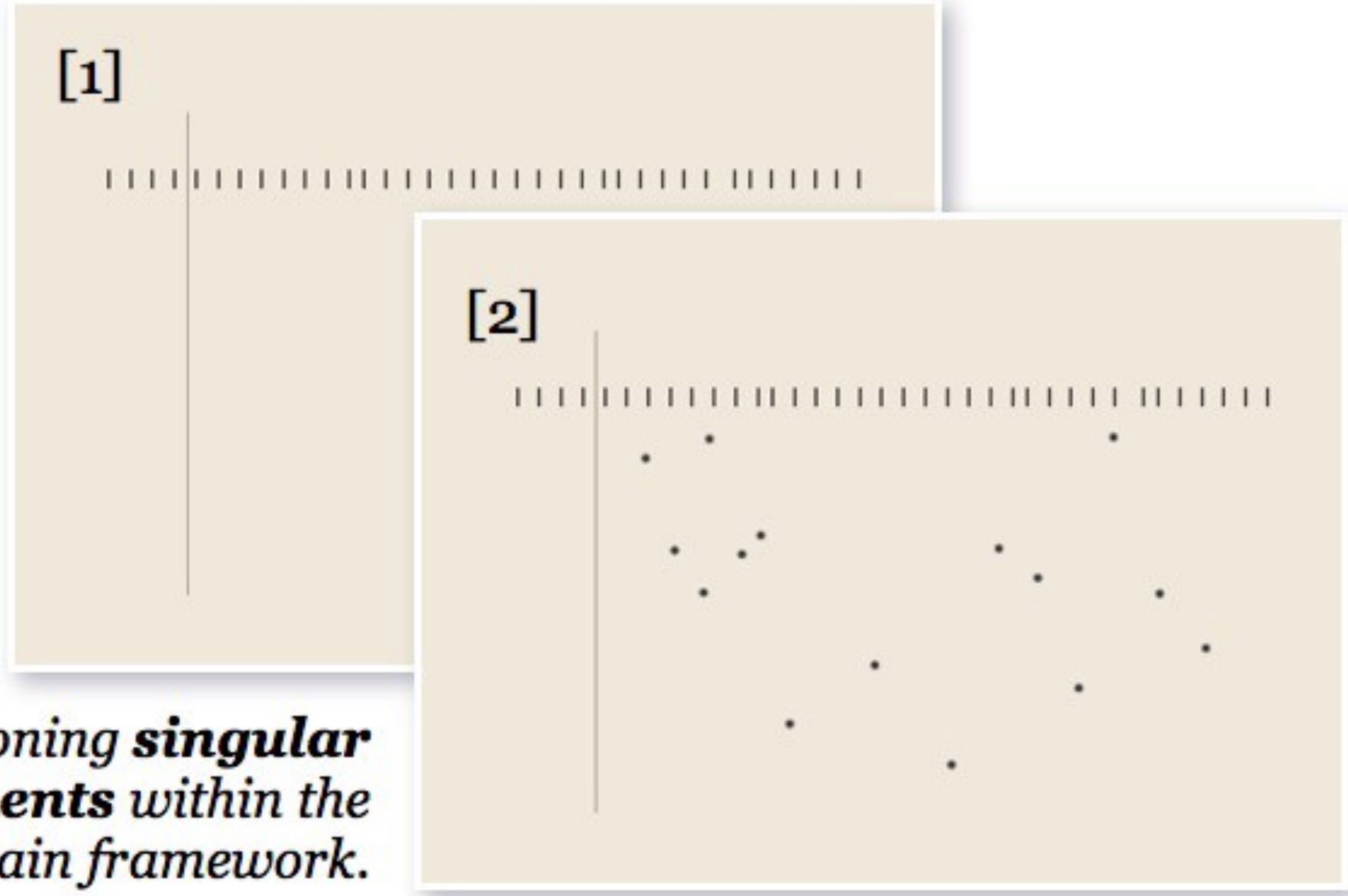
Her process involves layering multiple sub-narratives over a main construct. To achieve this multilayered storytelling with data … everything depends on the concept of layering, establishing hierarchies, and making them clear.

Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.



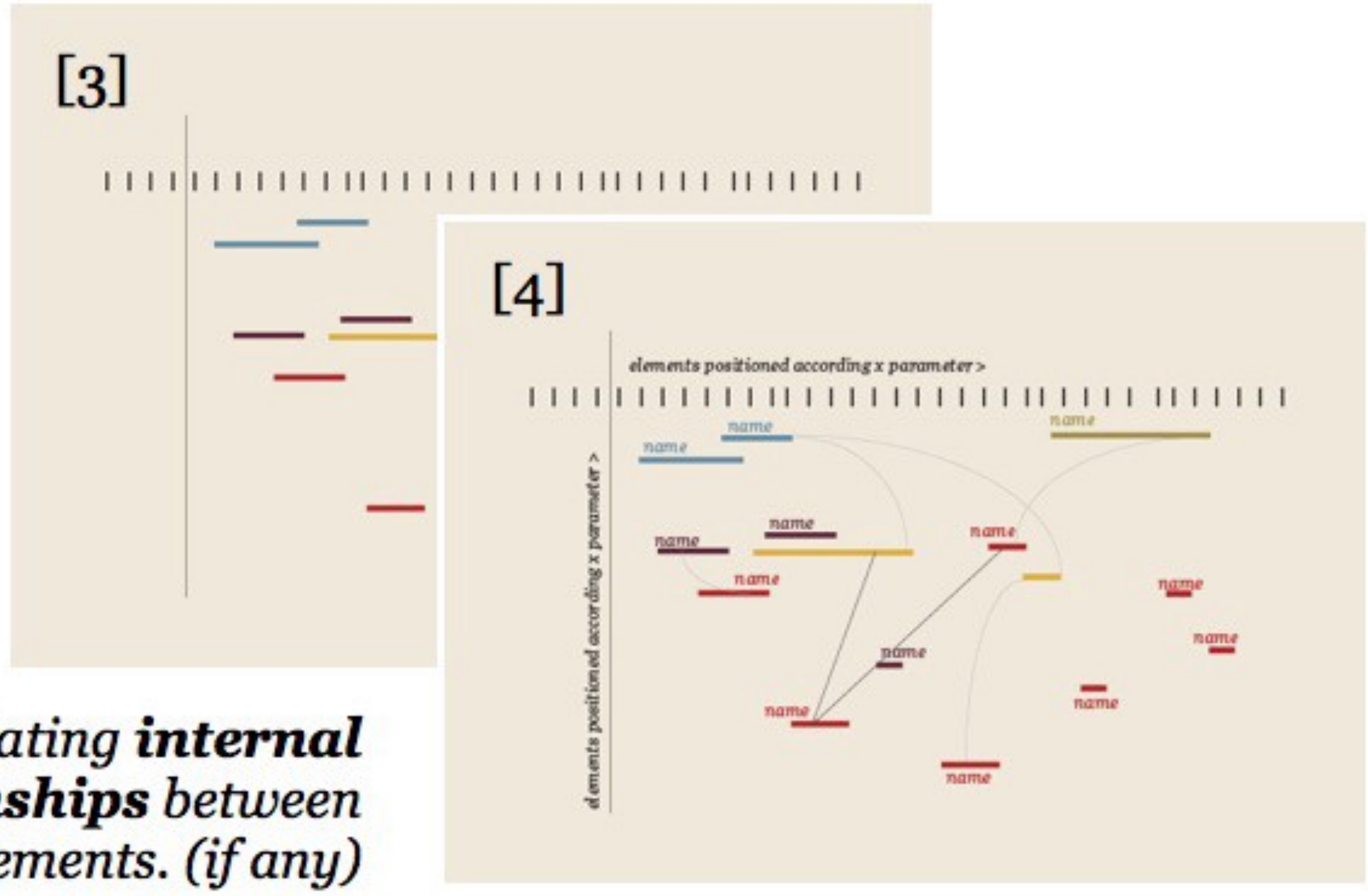
Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

Layering, and making **hierarchies clear**



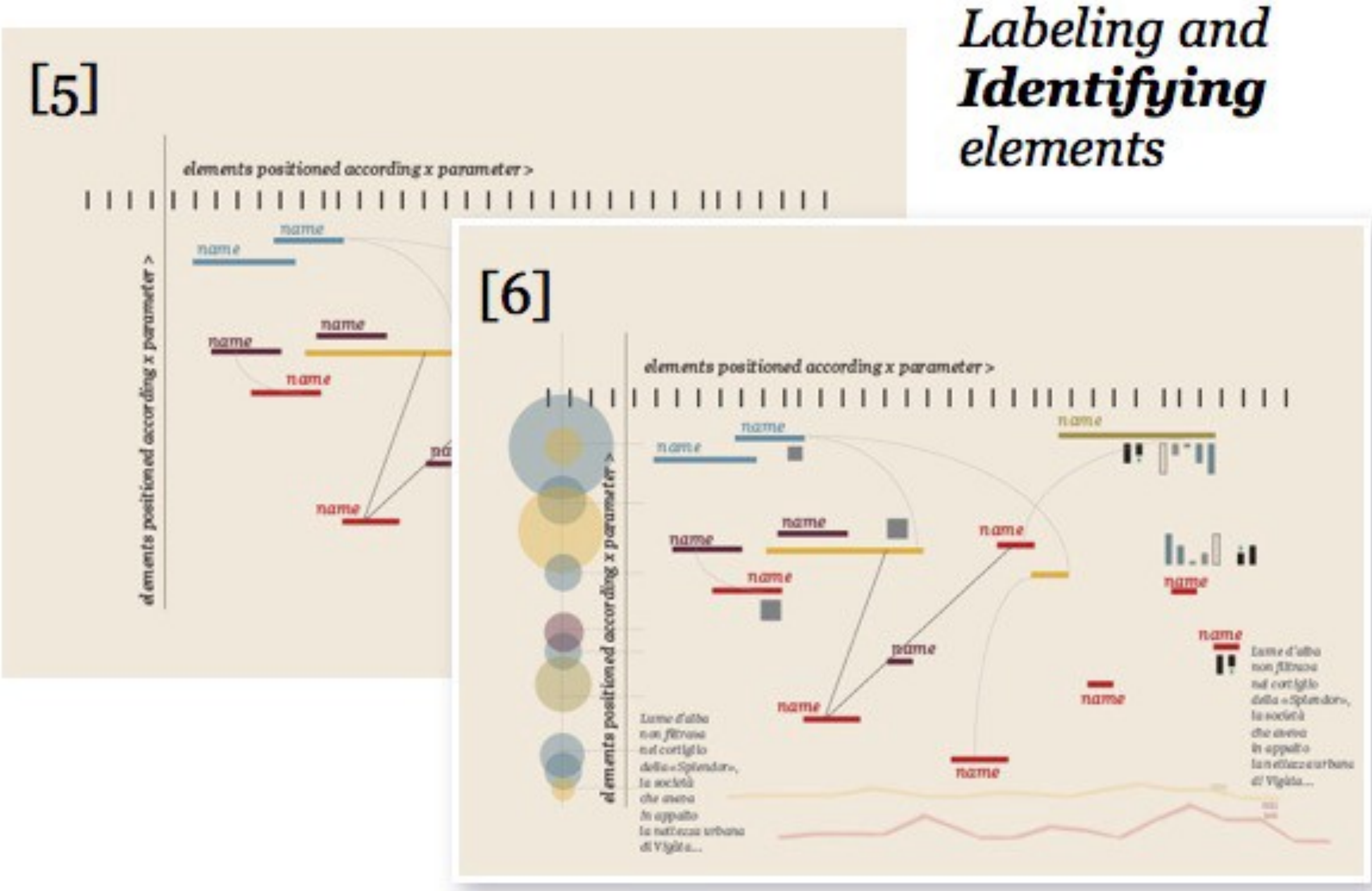
*Composing the main **architecture** of the visualization*

*Positioning **singular elements** within the main framework.*



*Constructing shaped elements of **dimensionality and form***

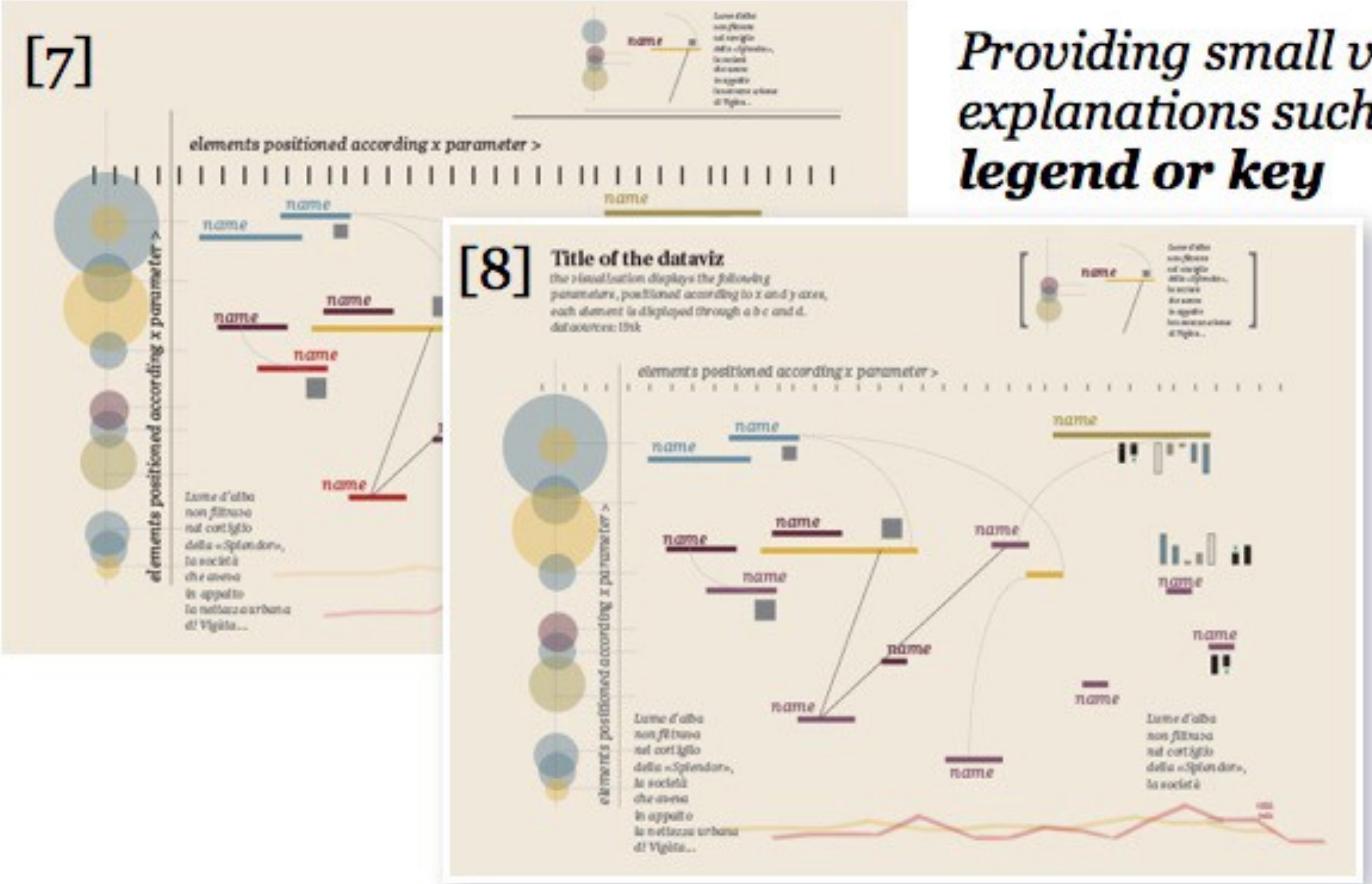
*Elucidating **internal relationships** between elements. (if any)*



*Labeling and **Identifying** elements*

Supplementing the greater story through the addition

*of **"minor or tangental tales"** elements*



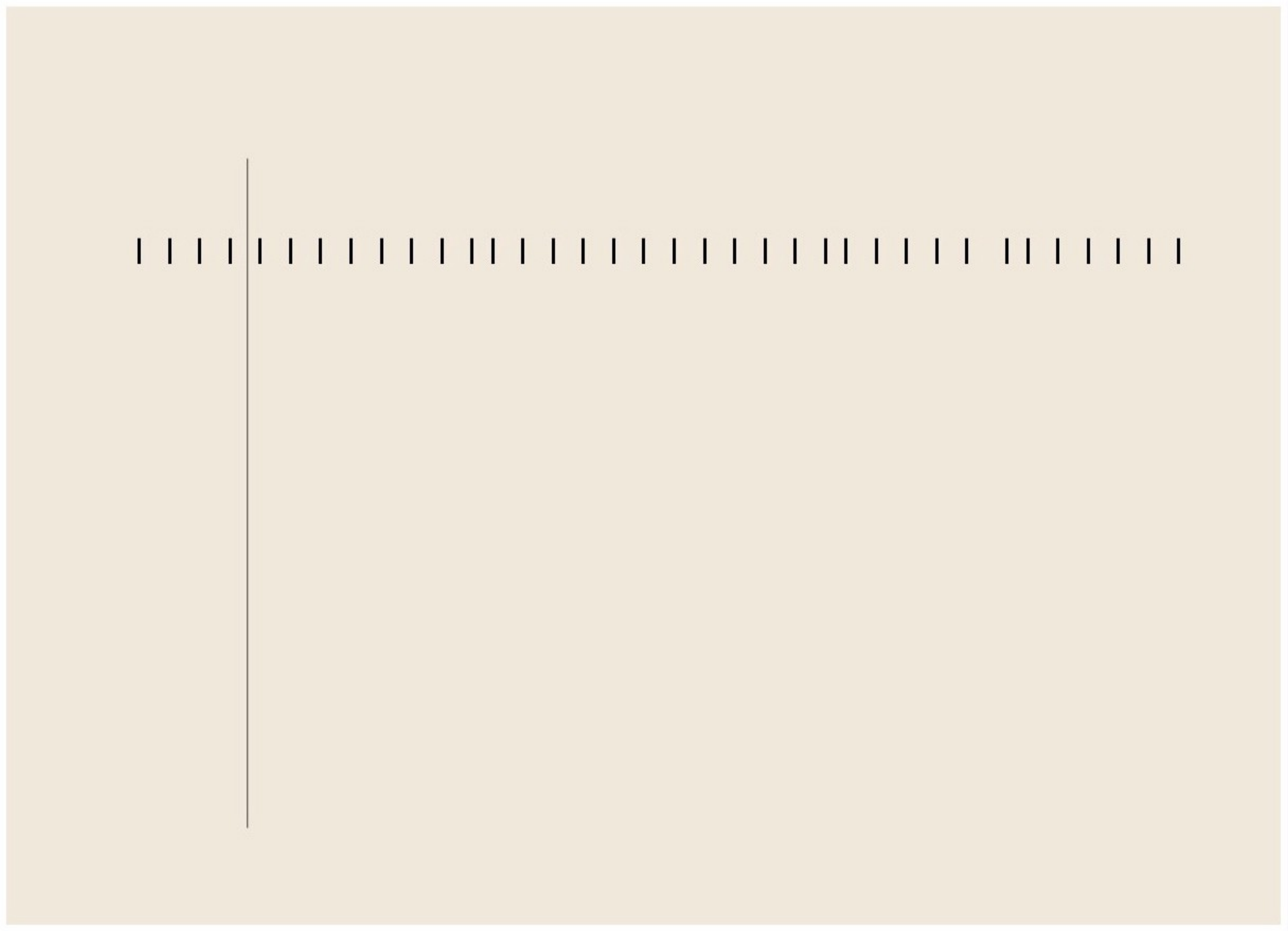
*Providing small visual explanations such as a **legend or key***

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

Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

[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.

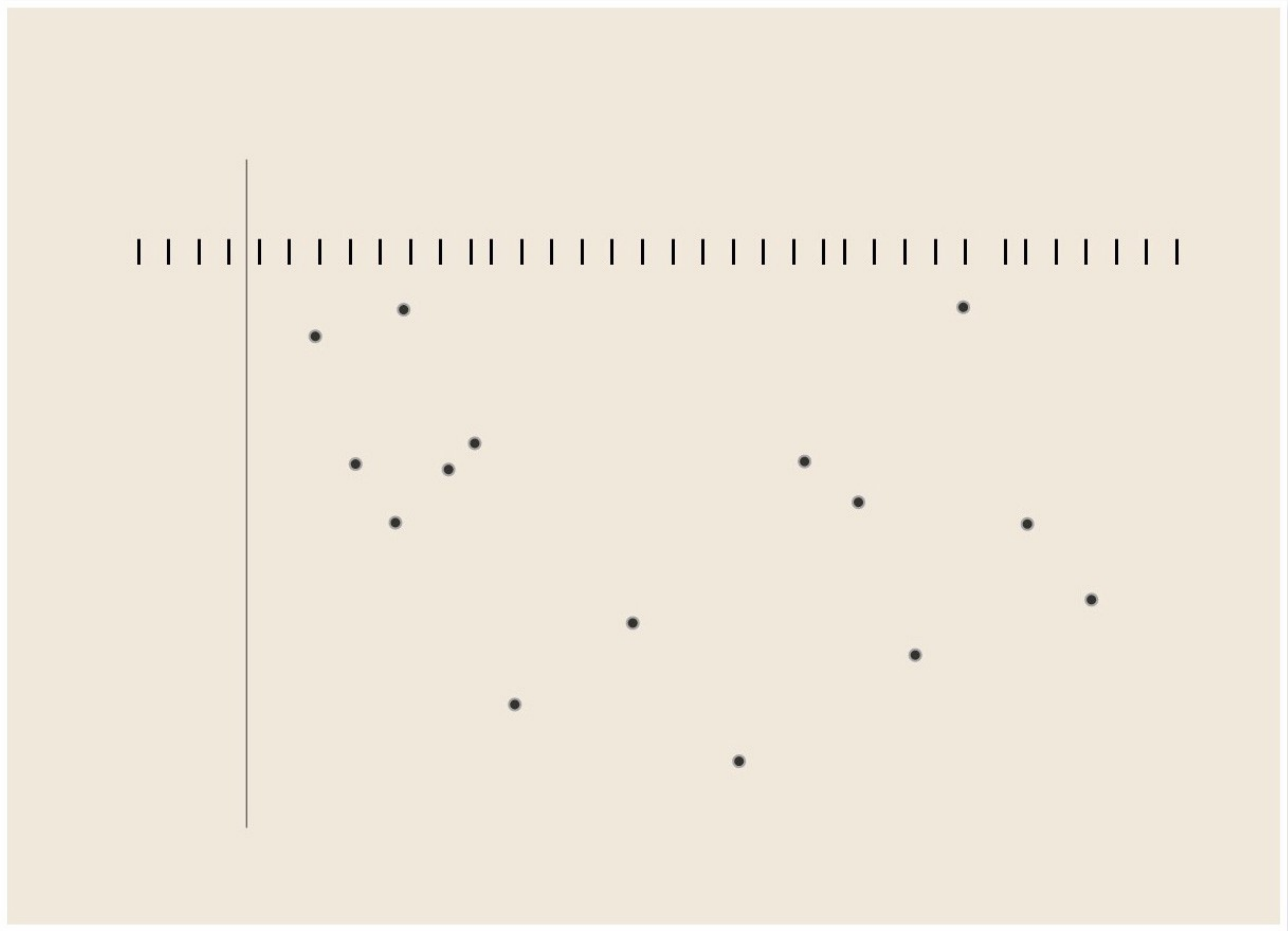


Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.

Lupi, Giorgia. “The Architecture of a Data Visualization: Multilayered Storytelling Through ‘Info-Spatial’ Compositions.” medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

[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.

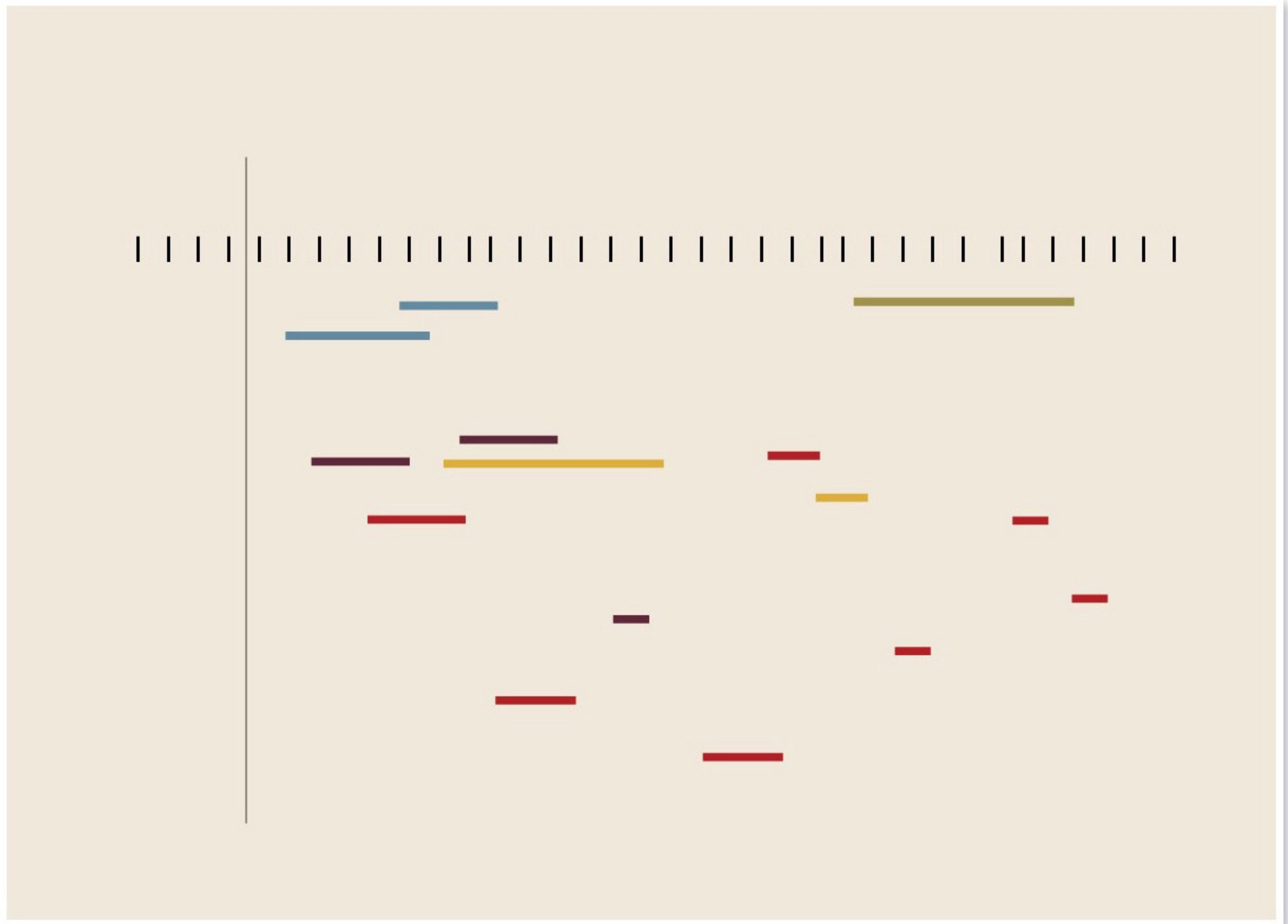


Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.

Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

[3] Constructing shaped elements of dimensionality and form

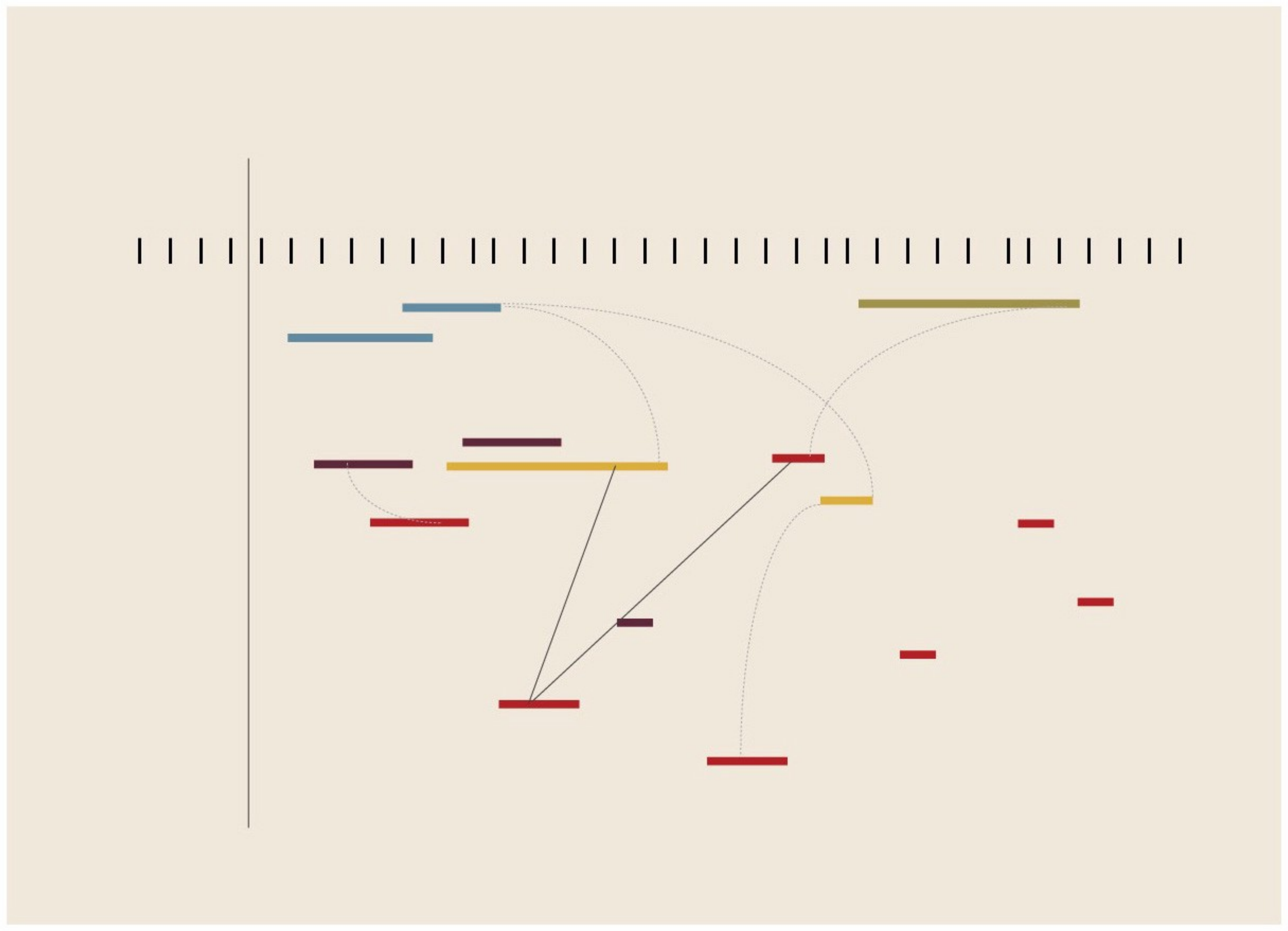
Constructing shaped elements of dimensionality and form (essentially polygons) with quantitative and qualitative parameters and positioning these within the main architecture. As these elements have form they must also be identified through colors according to opportunities to establish categorizations,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



Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

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

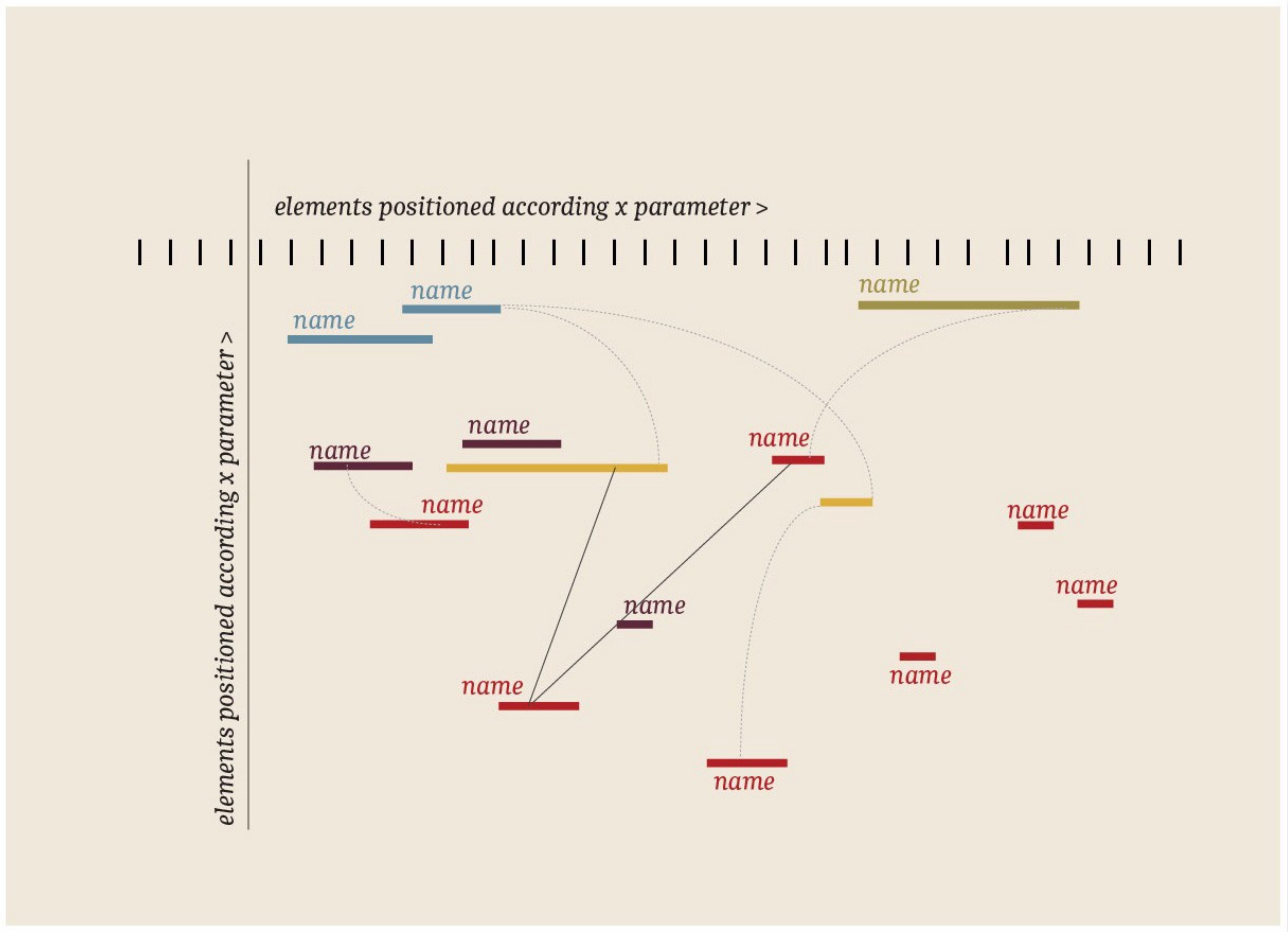
These links, directives, and qualifiers serve to give the story a comprehensive texture and correlate dependencies within the story.



Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

[5] Labeling and identifying

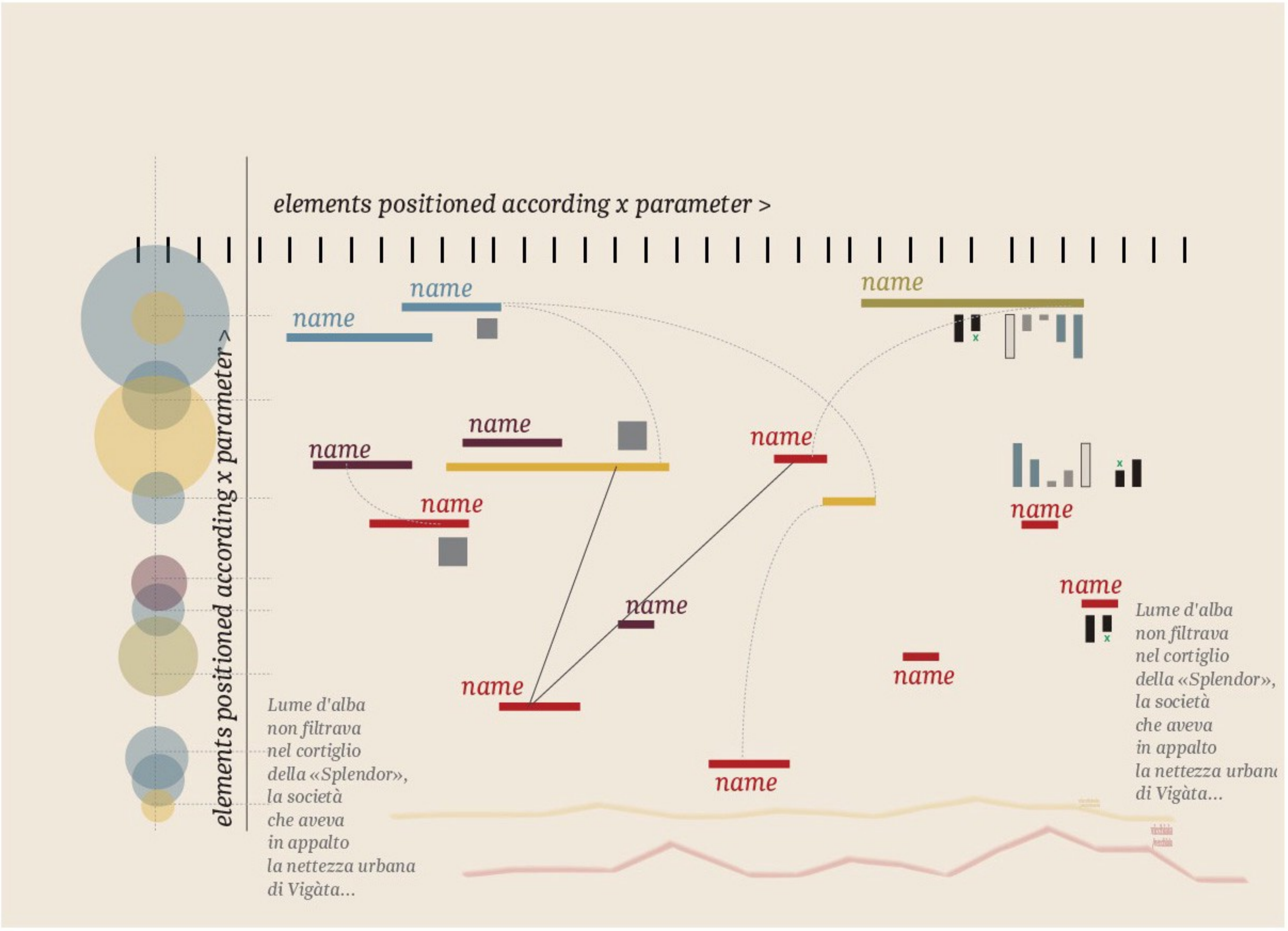
Through the addition of explanatory labels and short texts we provide requisite last mile clarity throughout the presentation.



Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

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

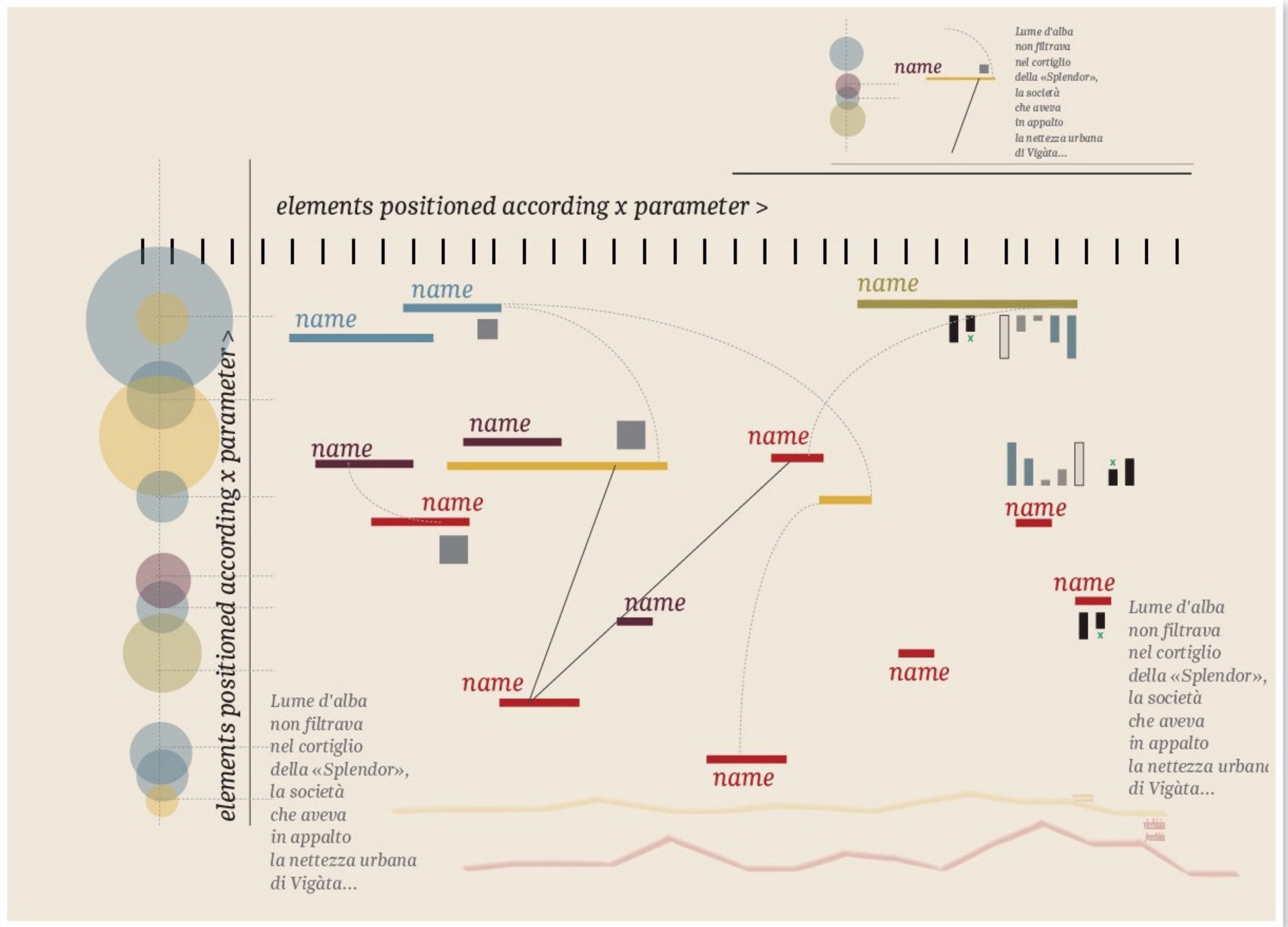
We consider this a very important step to contextualize the phenomena in a wider world. These components link the story to external ideas, other times, or other places. Elements that are rendered here may come from very diverse sources—analysis that is undertaken once we have strongly established the 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.



Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

[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.

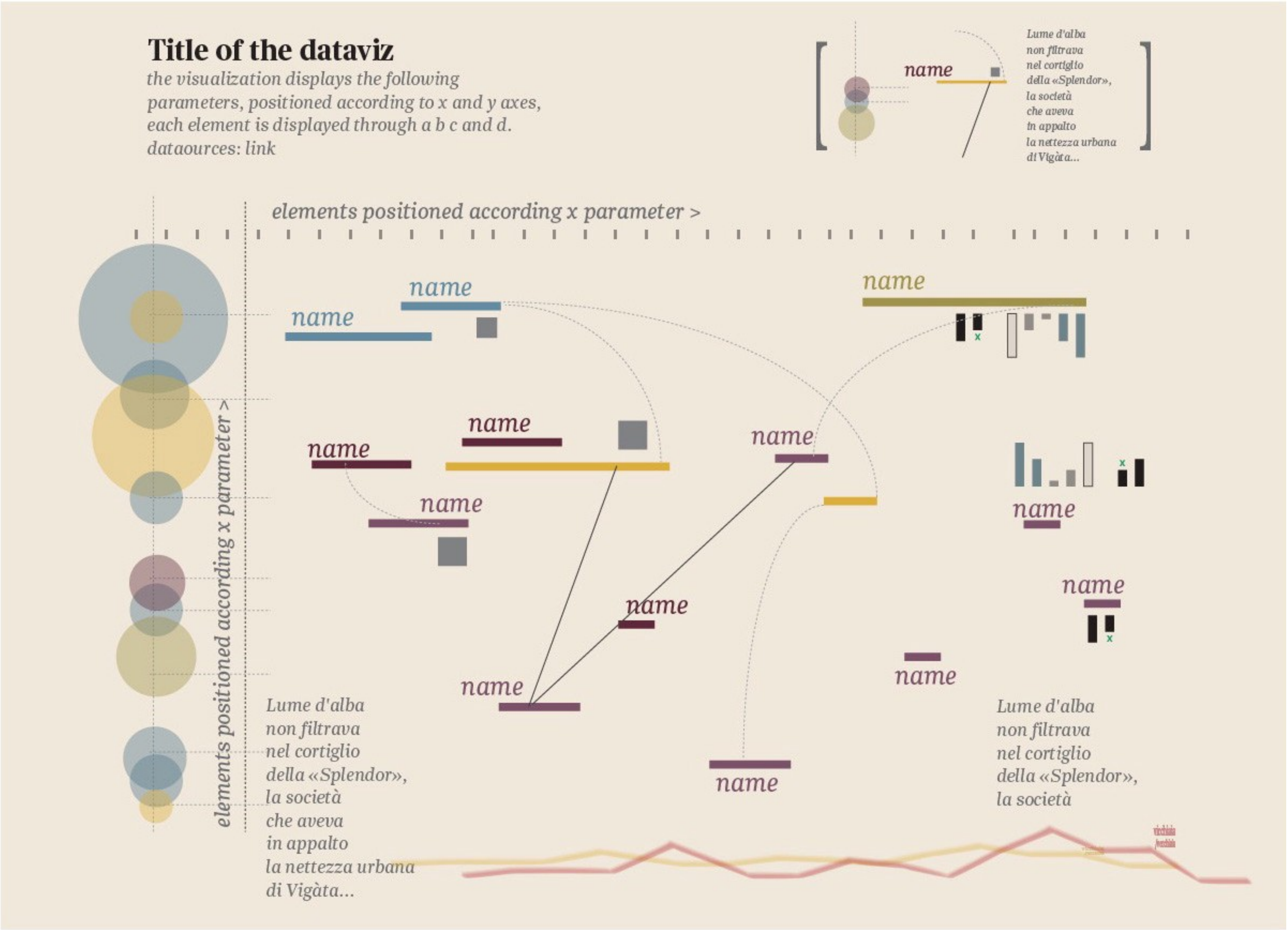


Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.

Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

[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.



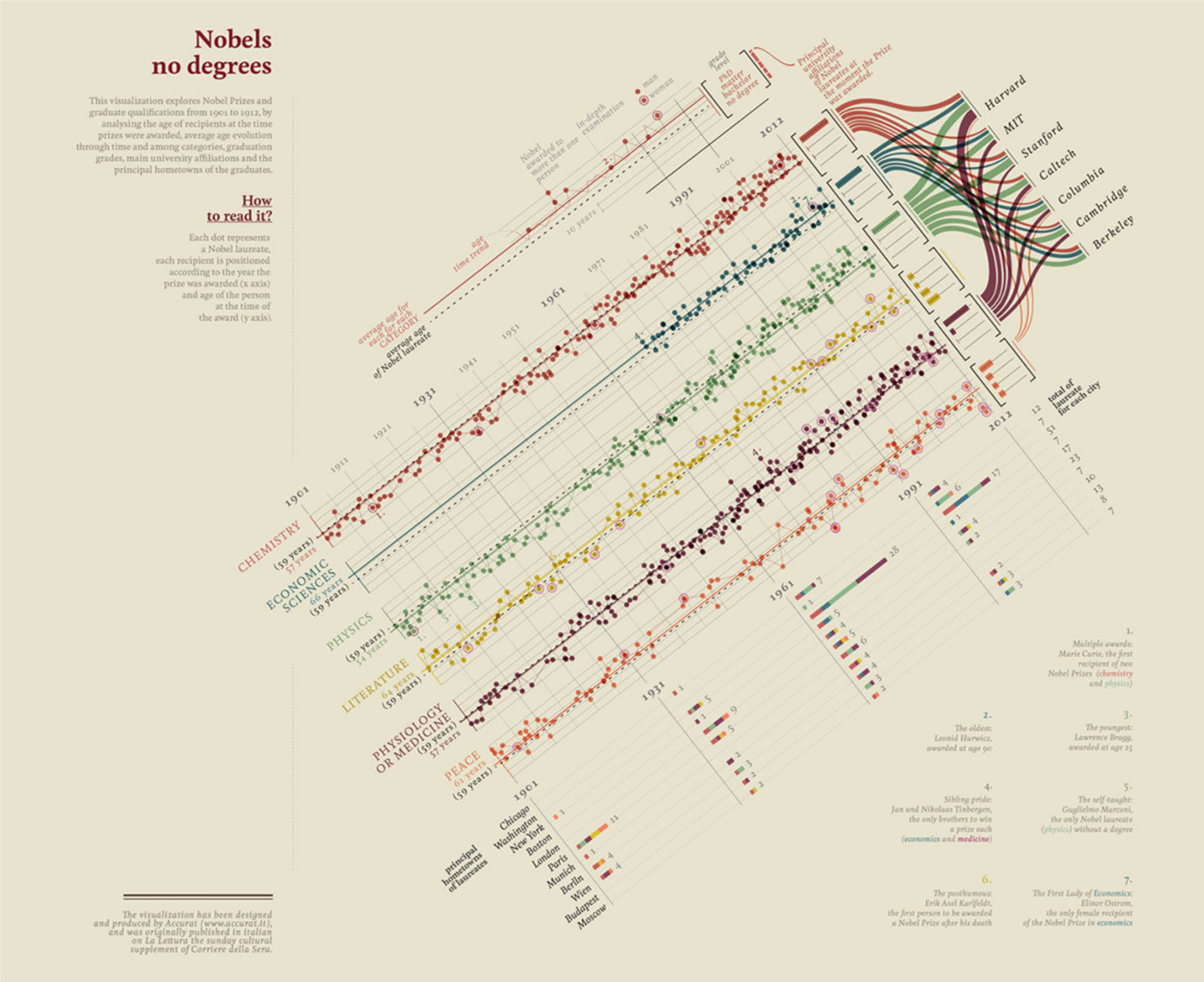
Lupi, Giorgia. The New Aesthetic of Data Narrative in Chapter 3 of Bihanic, David. New Challenges for Data Design. Springer, 2015. Print.
Lupi, Giorgia. "The Architecture of a Data Visualization: Multilayered Storytelling Through 'Info-Spatial' Compositions." medium.com. N.p., 26 Feb. 2015. Web. 10 Mar. 2019.

Building information complexity from common business graphics

Winner, Information is Beautiful Award

Lupi & Fragapane, et al

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.



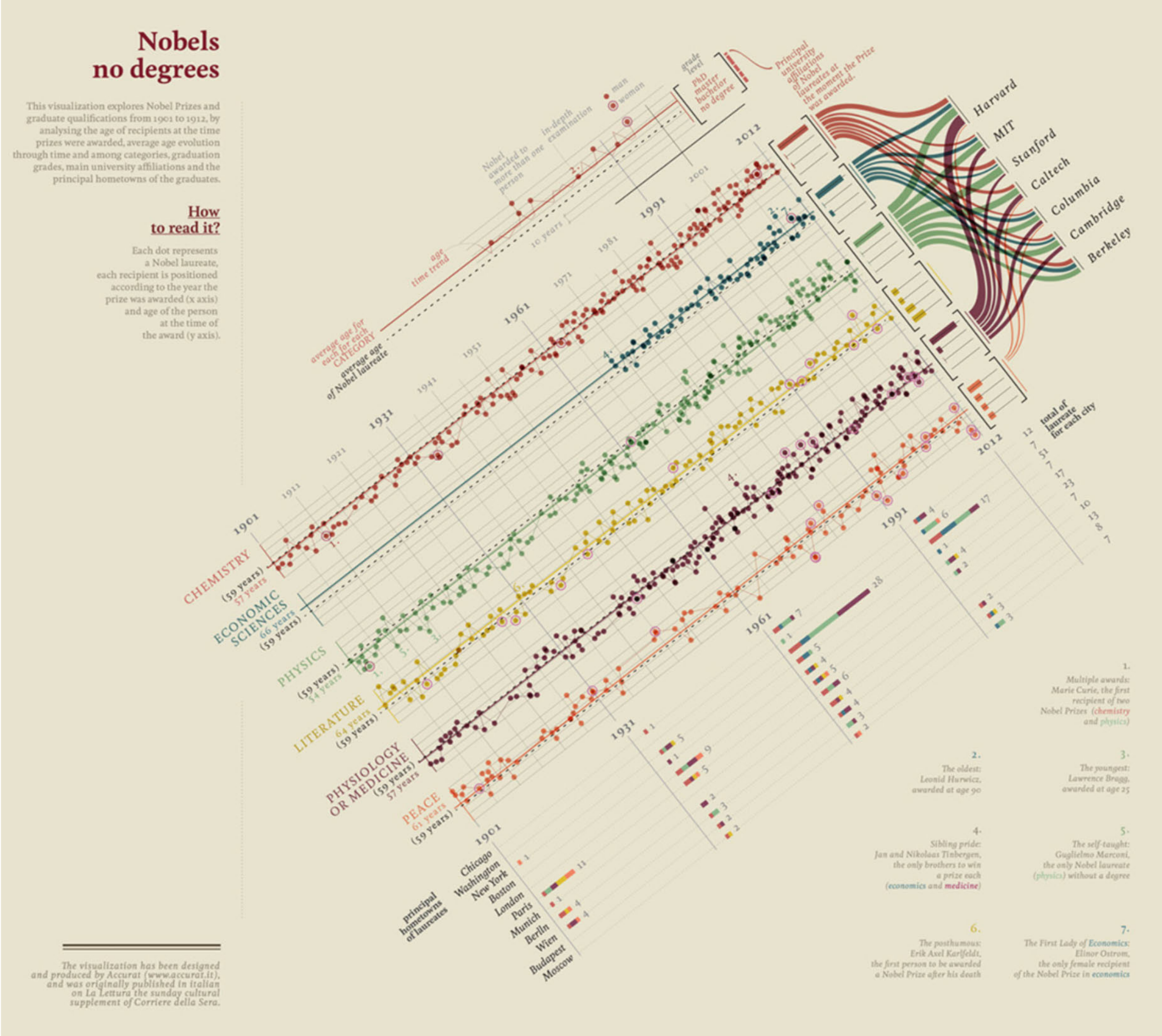
Lupi, Giorgia. Visual Data - La Lettura. Web. http://giorgialupi.com/lalettura

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.



Lupi, Giorgia. Visual Data - La Lettura. Web. http://giorgialupi.com/lalettura

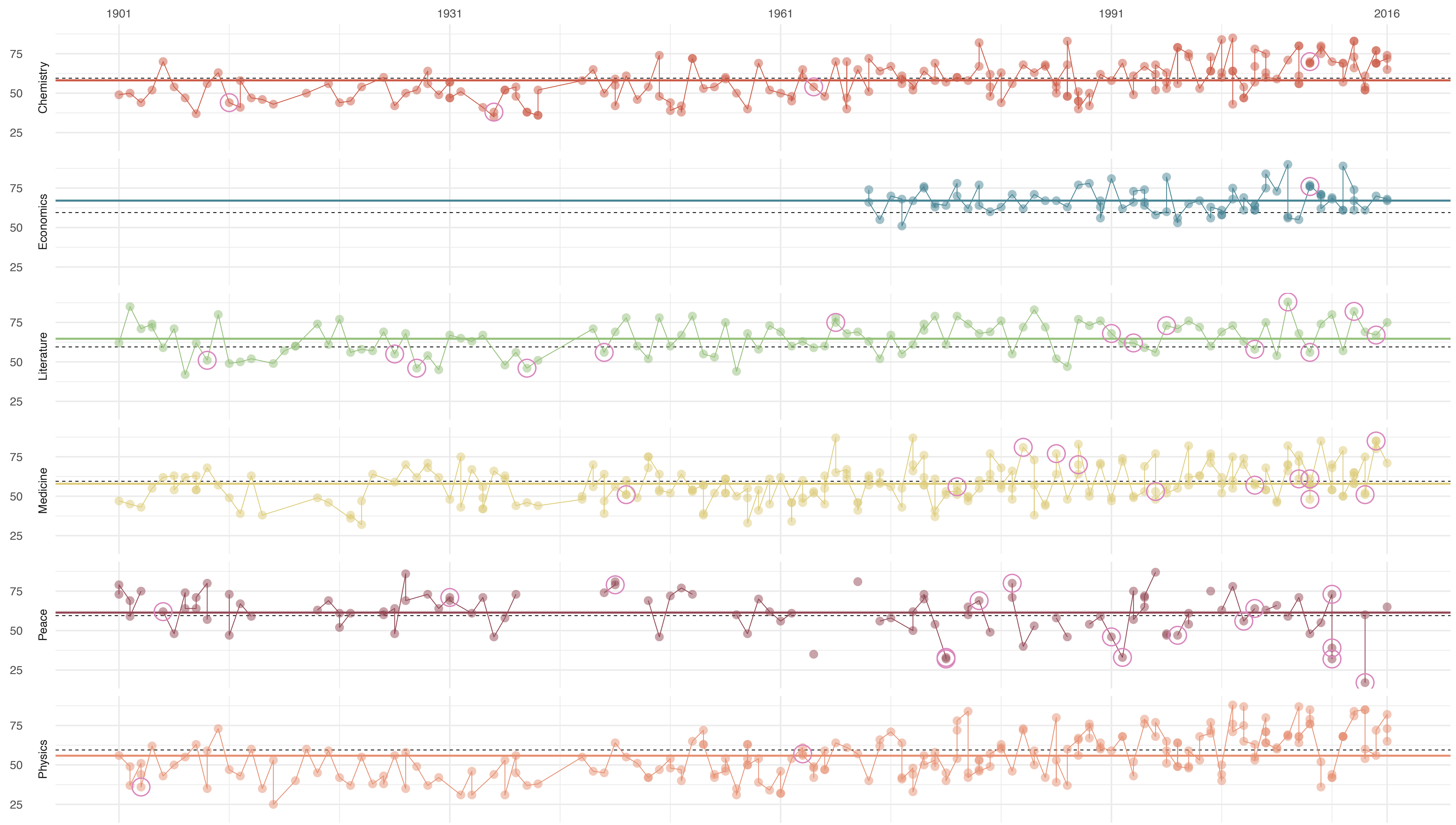
Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



Spencer, Scott. *Approximating the components of Lupi's Nobel no degrees*. 25 Mar 2019. Blog. https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/

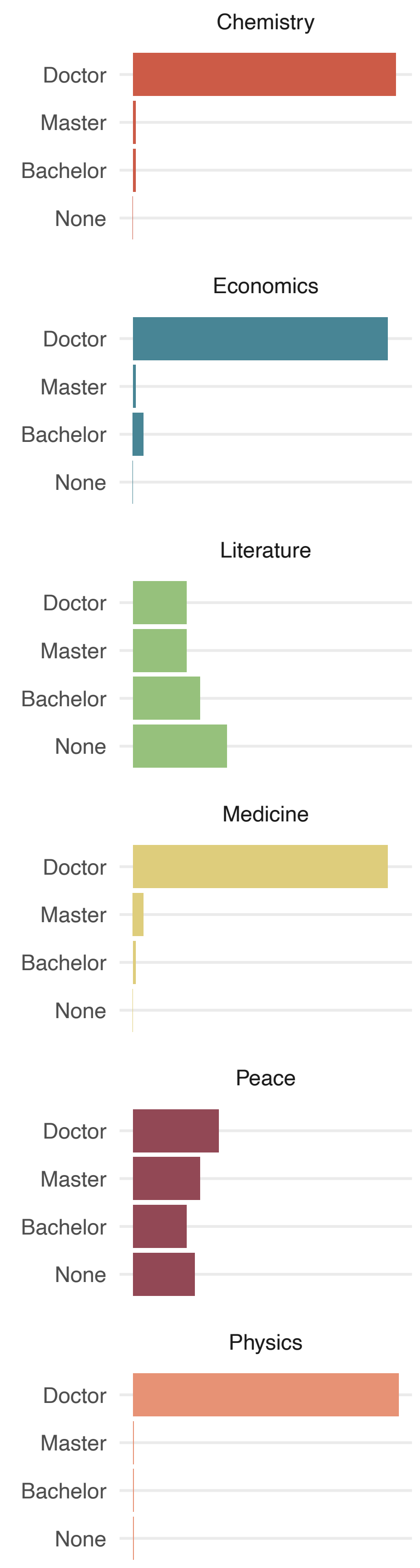
Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



Spencer, Scott. *Approximating the components of Lupi's Nobel no degrees*. 25 Mar 2019. Blog. https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/

Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram

Spencer, Scott. *Approximating the components of Lupi's Nobel no degrees*. 25 Mar 2019. Blog. https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/

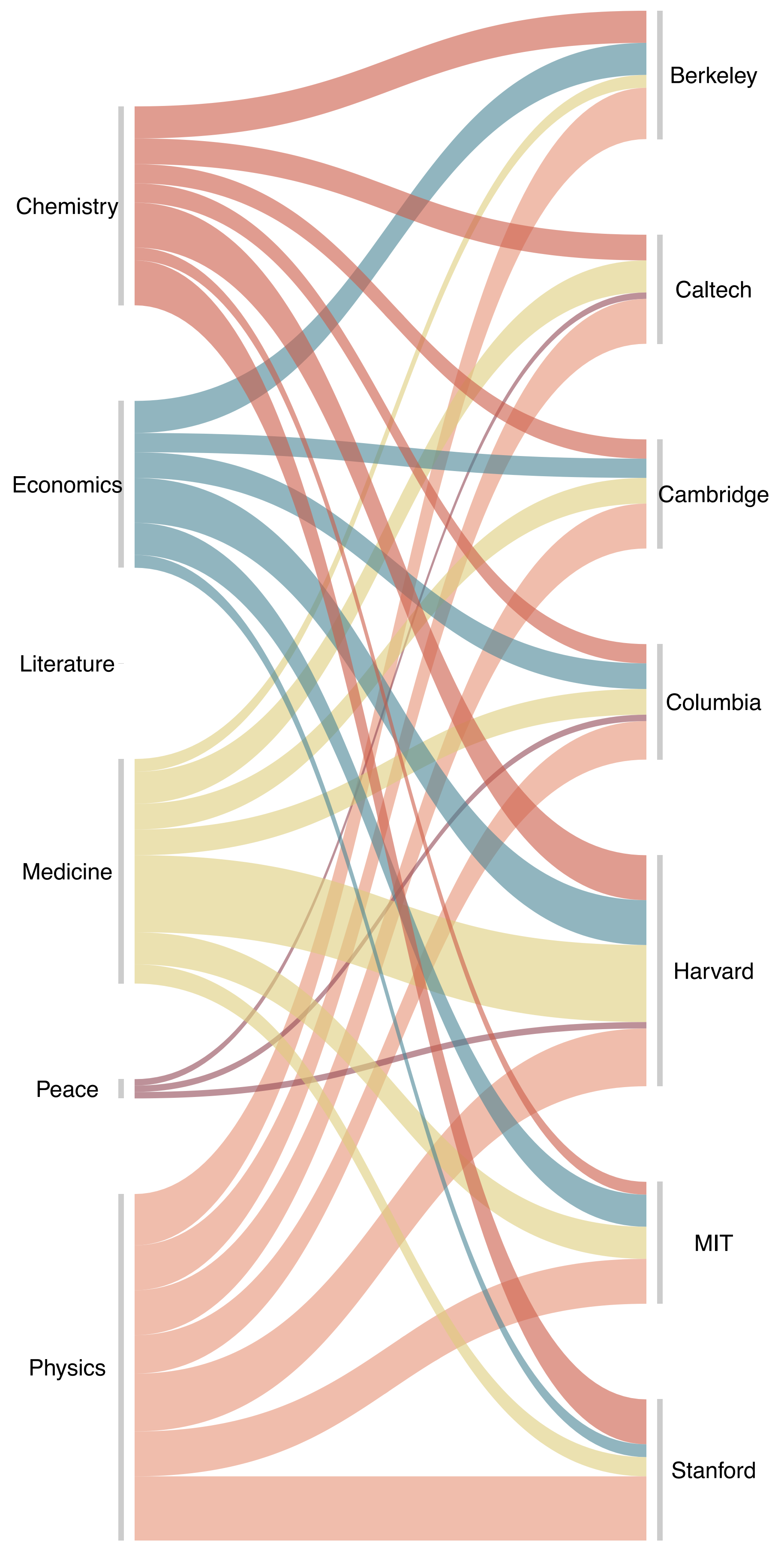
Approximating the components

Central display layering points, lines

Per-category histograms

Per-era stacked bar charts

Sankey diagram



Spencer, Scott. *Approximating the components of Lupi's Nobel no degrees*. 25 Mar 2019. Blog. https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/

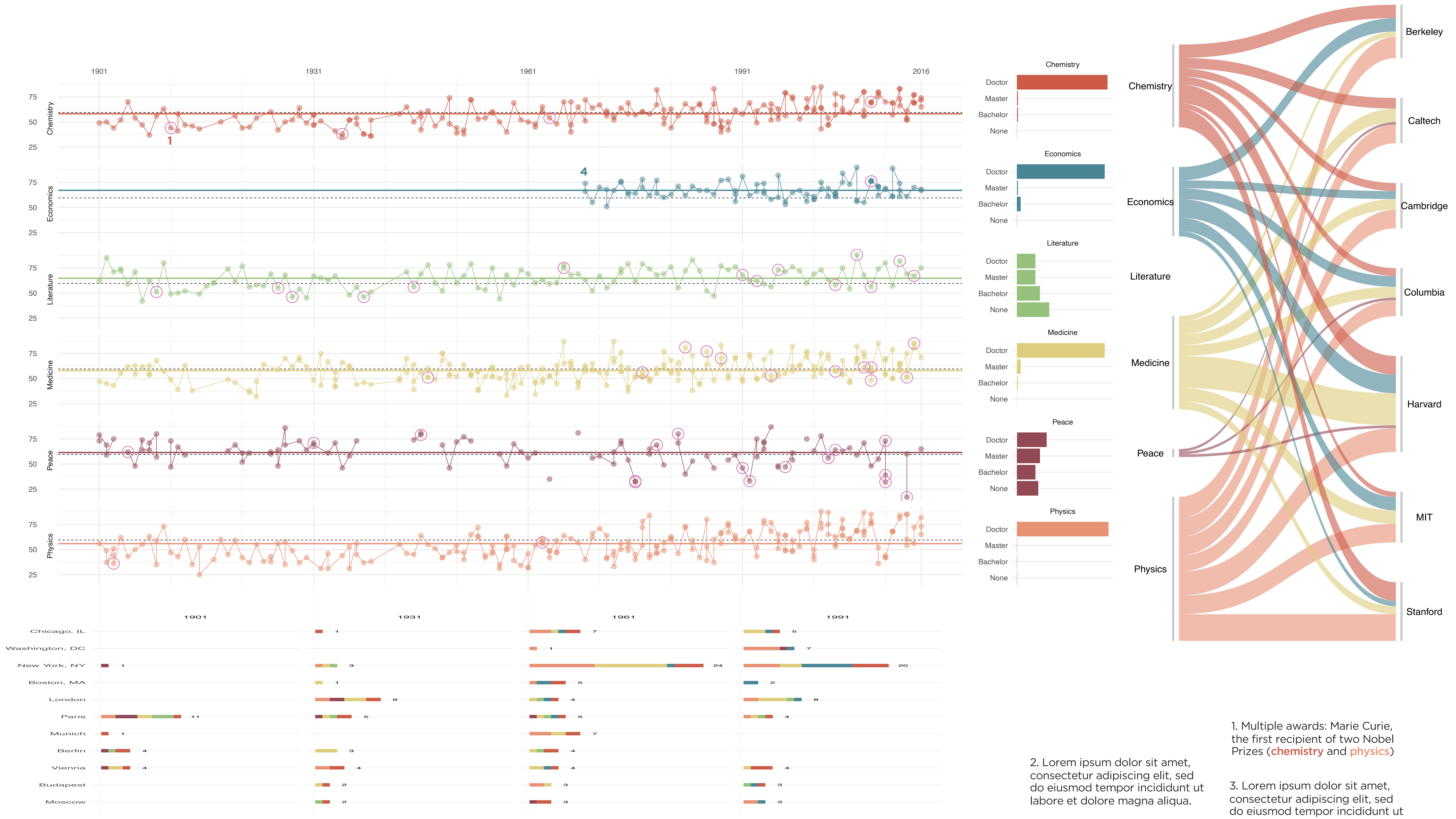
Spencer, Scott. *Approximating the components of Lupi's Nobel no degrees*. 25 Mar 2019. Blog.
https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/

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

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

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

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.

Spencer, Scott. *Approximating the components of Lupi's Nobel no degrees*. 25 Mar 2019. Blog. https://ssp3nc3r.github.io/post/approximating-the-components-of-lupi-s-nobel-no-degrees/

Data Illustrator

Getting Started   Gallery   About   Release Notes   Launch

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

Data Illustrator. Nobel Laureates and Prizes Demo. 2019. Web. http://data-illustrator.com/example.php?v=nobel-no-org
Liu, Zhicheng et al. “Data Illustrator.” New York, New York, USA: ACM Press, 2018. 1–13. Web.

Examples?

TODO: develop this section

On the horizon

For Next Week, Module 9:

Agenda next week

Next deliverable, *draft* information graphic

Perceptions of uncertainty, missing data

Information graphics

The minimum

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 might 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!

Spencer, Scott. Illustration of Columbia University Baseball. Columbia University. 3 Apr 2019.