Storytelling With Data

The storytelling process, ***with images and data graphics***

Scott Spencer | Columbia University

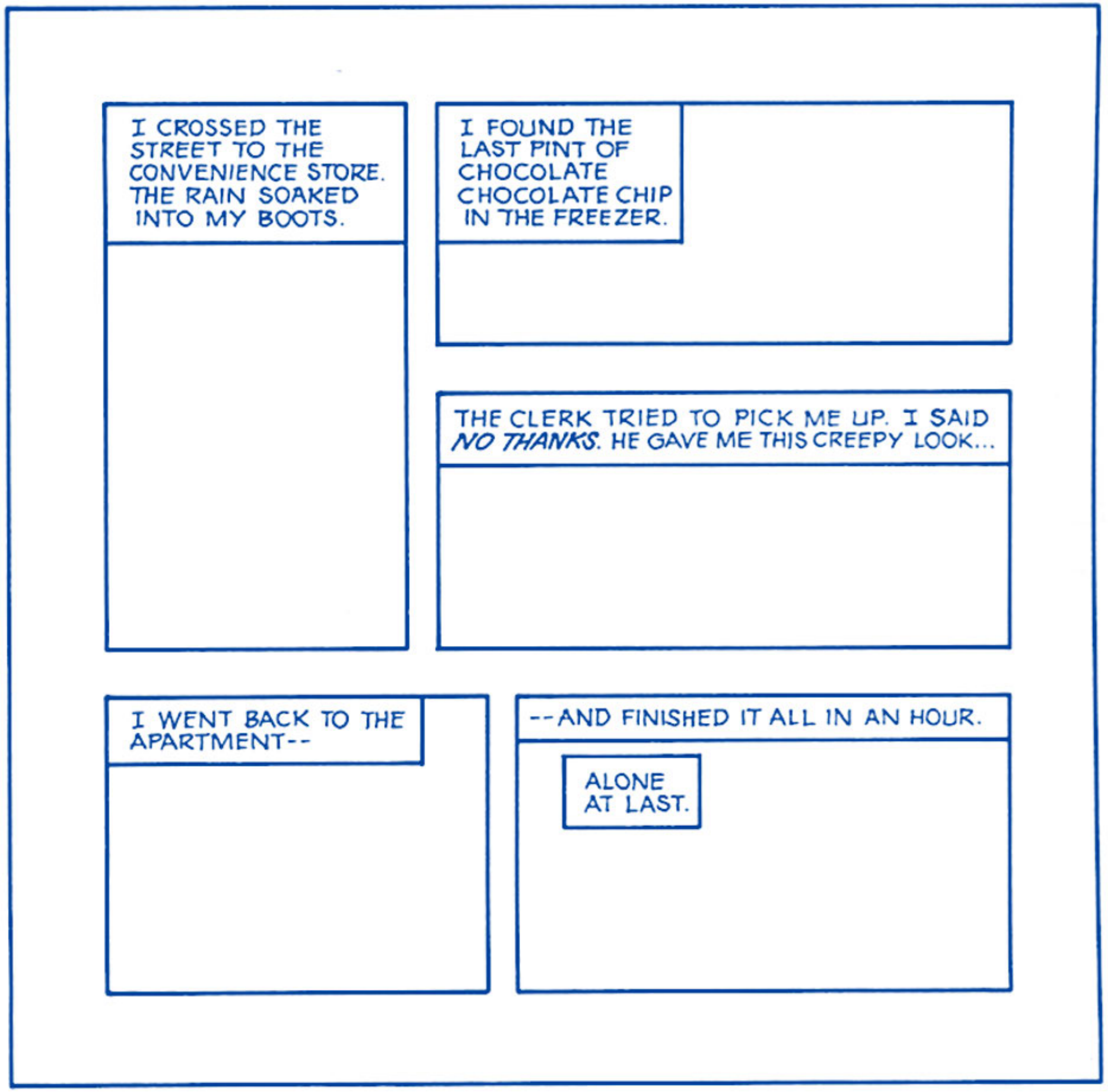
Conceptual project timeline

integrating images and words in narrative to create shared meaning

Edward Tufte, in *Visual Explanations* — his book about pictures of verbs to show causes and effects, explanations and narratives — cites to **comics** for understanding the idea of "**visual narrative**".

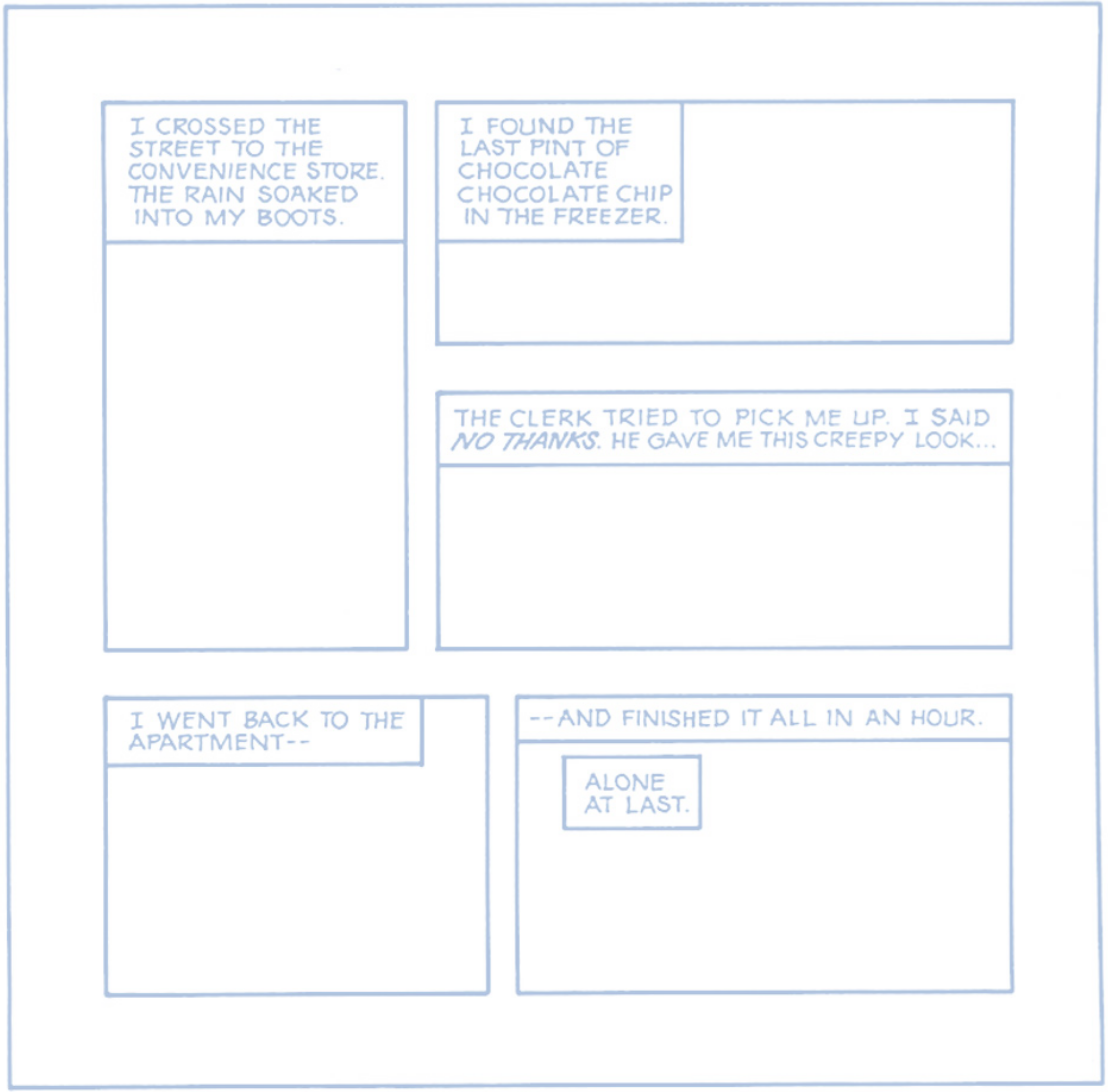
Tufte, Edward

A very simple story using only words

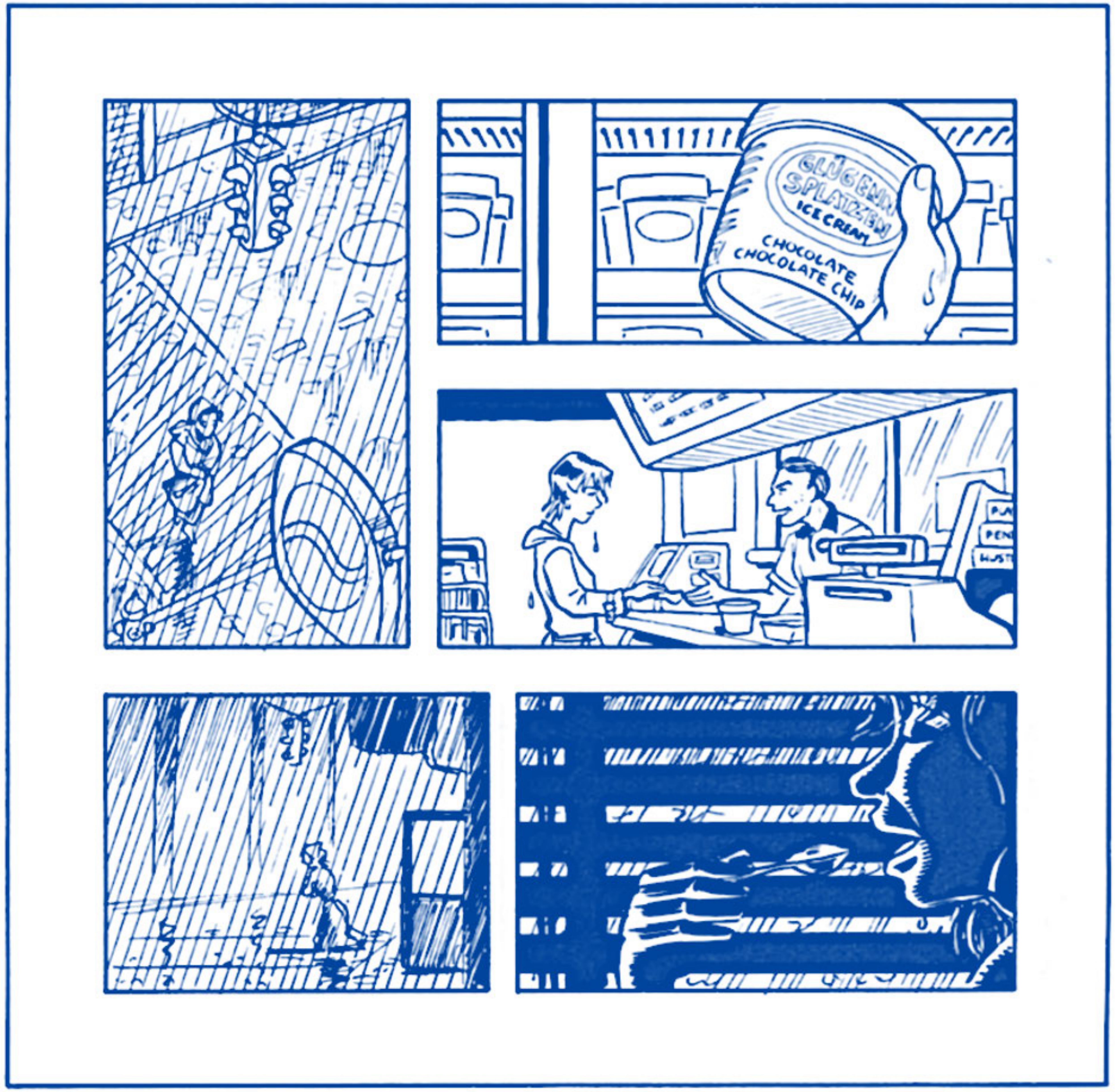


McCloud

A very simple story using words

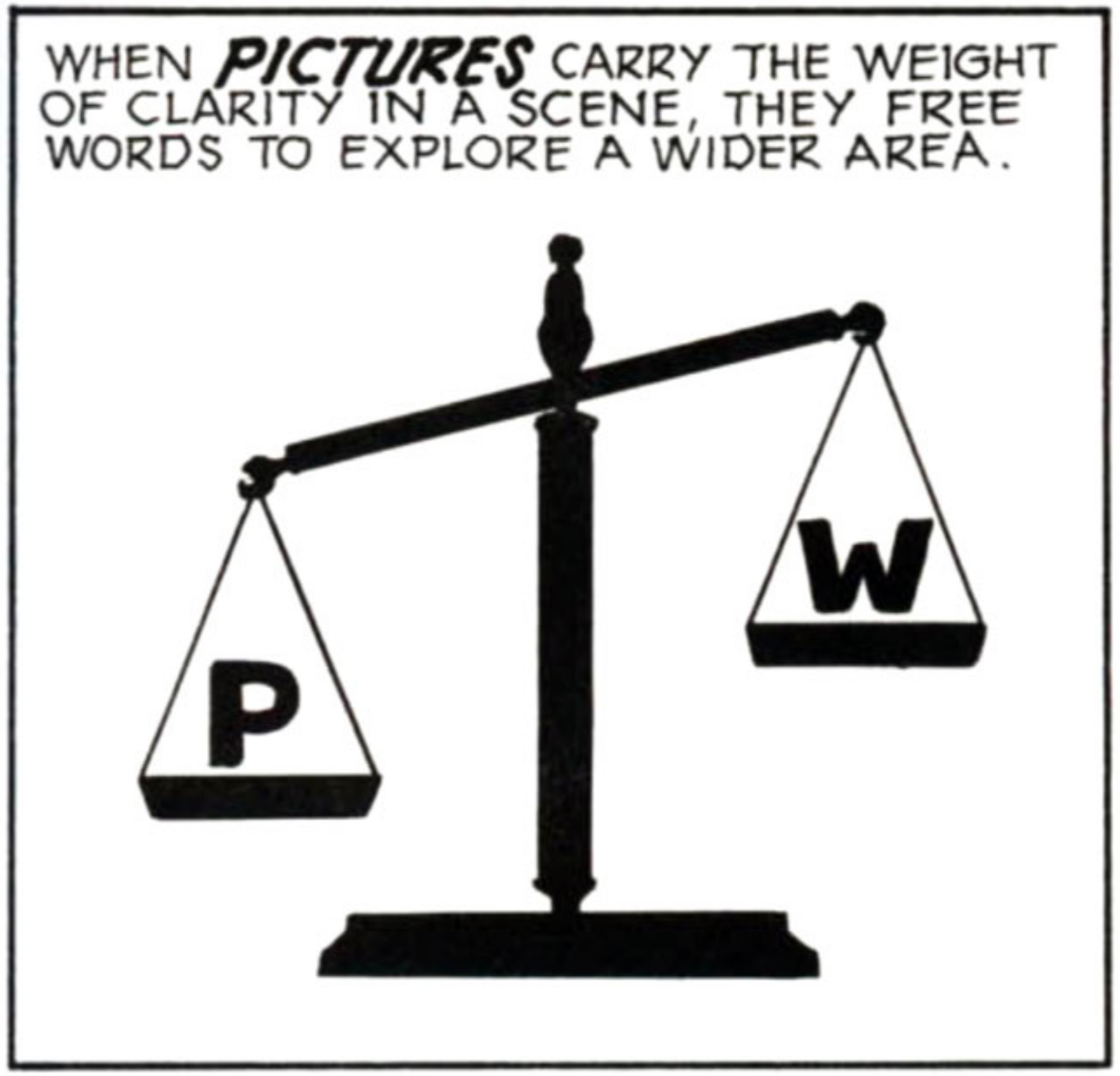


The same story using only images



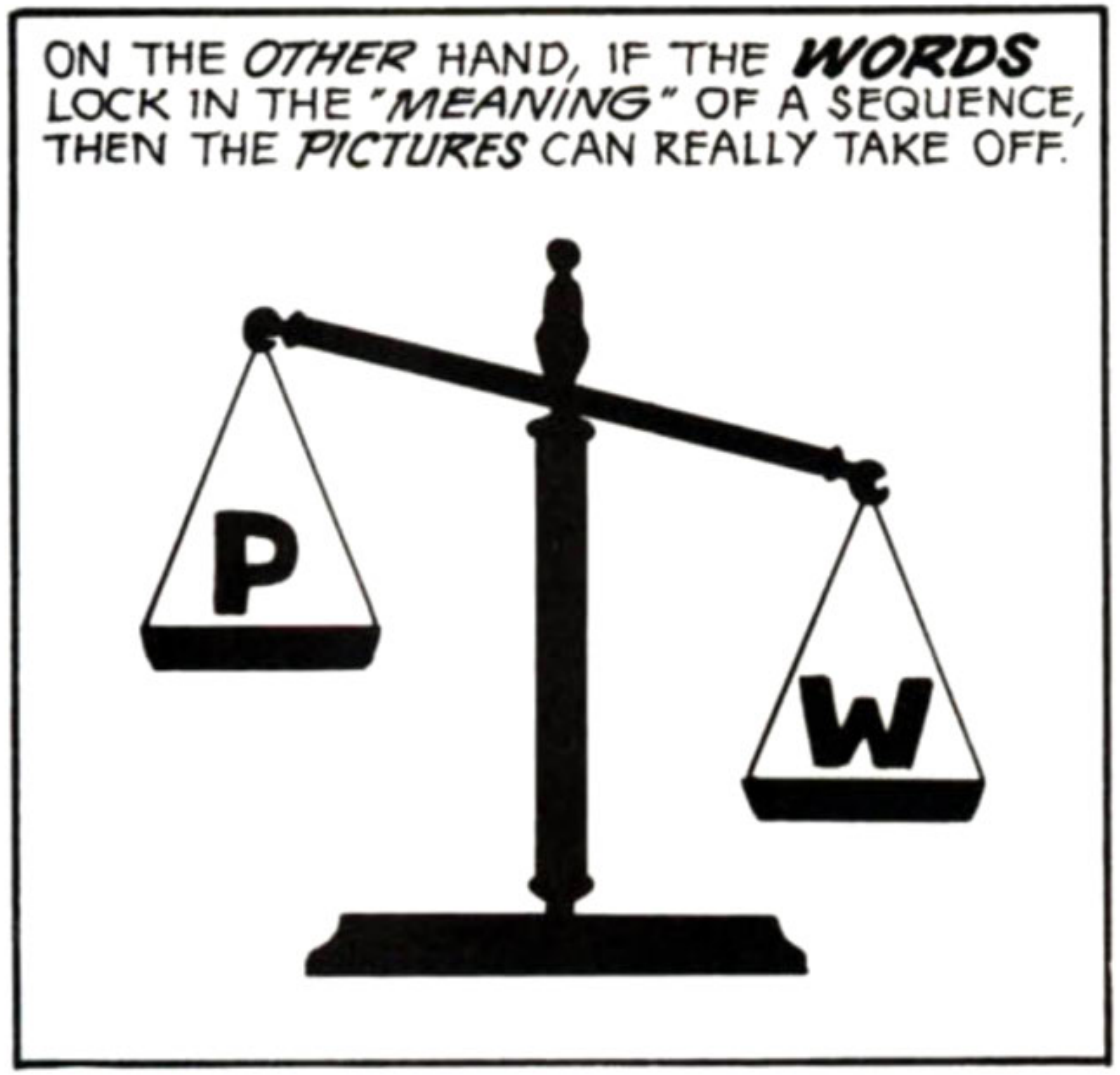
McCloud

Shared meaning of words and images



McCloud

Shared meaning of words and images



McCloud


INTERDEPENDENT COMBINATIONS AREN'T ALWAYS AN *EQUAL BALANCE* THOUGH AND MAY FALL *ANYWHERE* ON A SCALE BETWEEN TYPES ONE AND TWO.
GENERALLY SPEAKING, THE MORE IS SAID WITH *WORDS*, THE MORE THE PICTURES CAN BE FREED TO GO EXPLORING AND *VICE VERSA*.
P
W
P
W
W
P

McCloud

The **balance between text and visualization** becomes an issue when too much text takes away from the data but too little text leaves the viewer confused and unable to see the connections.

Kosara

Storyboards are a tool to test narrative containing mixed media types

Storyboard | | Date: | To: | From:

Visual narrative		Written narrative
	1	
	2	
	3	

Visual narrative		Written narrative
	4	
	5	
	6	

Storyboard | *The Next Rembrandt*, demonstrating our data science skills we can apply to client needs. | Date: 2020 February 29 | To: Ing CMO | From: Scott Spencer

Visual narrative | Written narrative

1

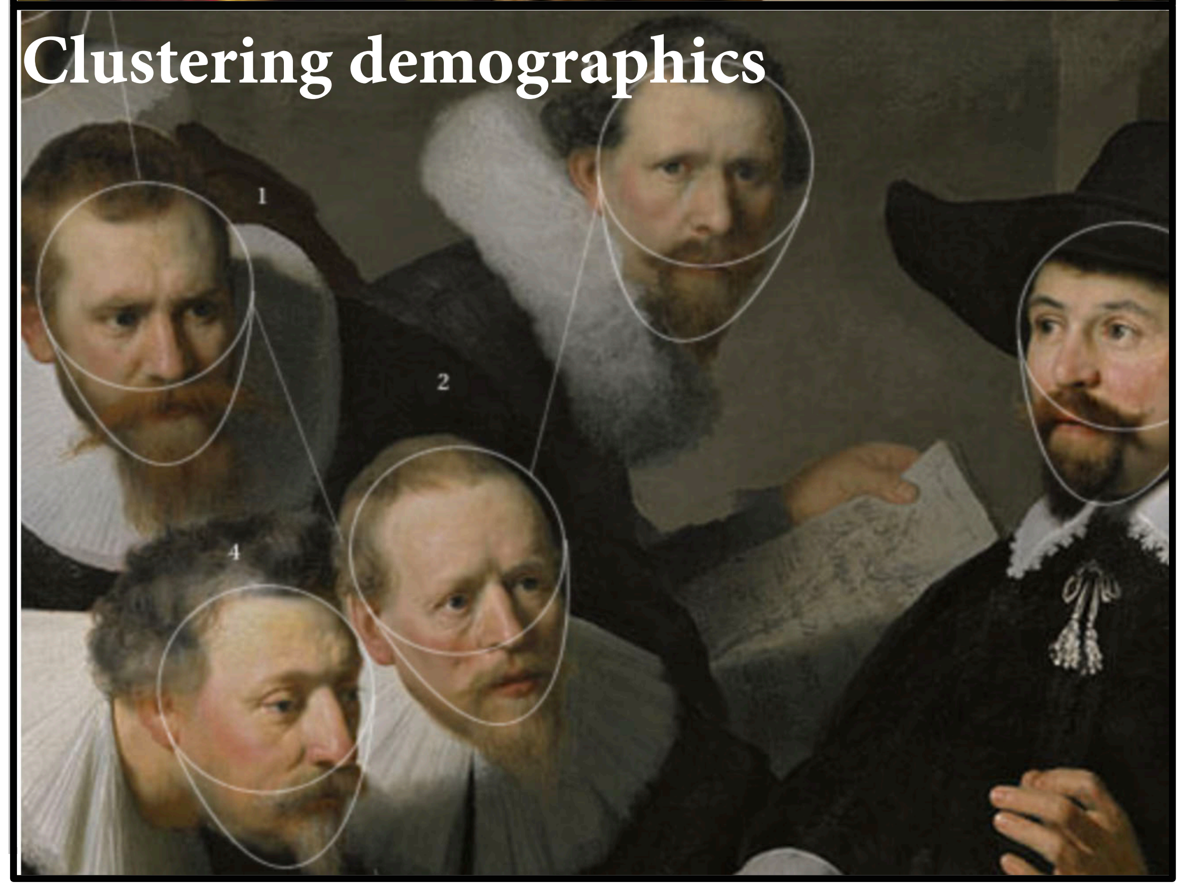
We should market ourselves as data-driven, forward-thinking, and creative.

To demonstrate our creativity, we attempted to turn paintings into data, analyze it, and create an entirely new painting perhaps indistinguishable from the master, Rembrandt. Did Rembrandt paint this? Or did *we*?

2

To answer the challenge, we gathered a extensive pool of data. Perhaps to some, a large collection of paintings are not data. But we partnered with museums to collect many of Rembrandt's works.

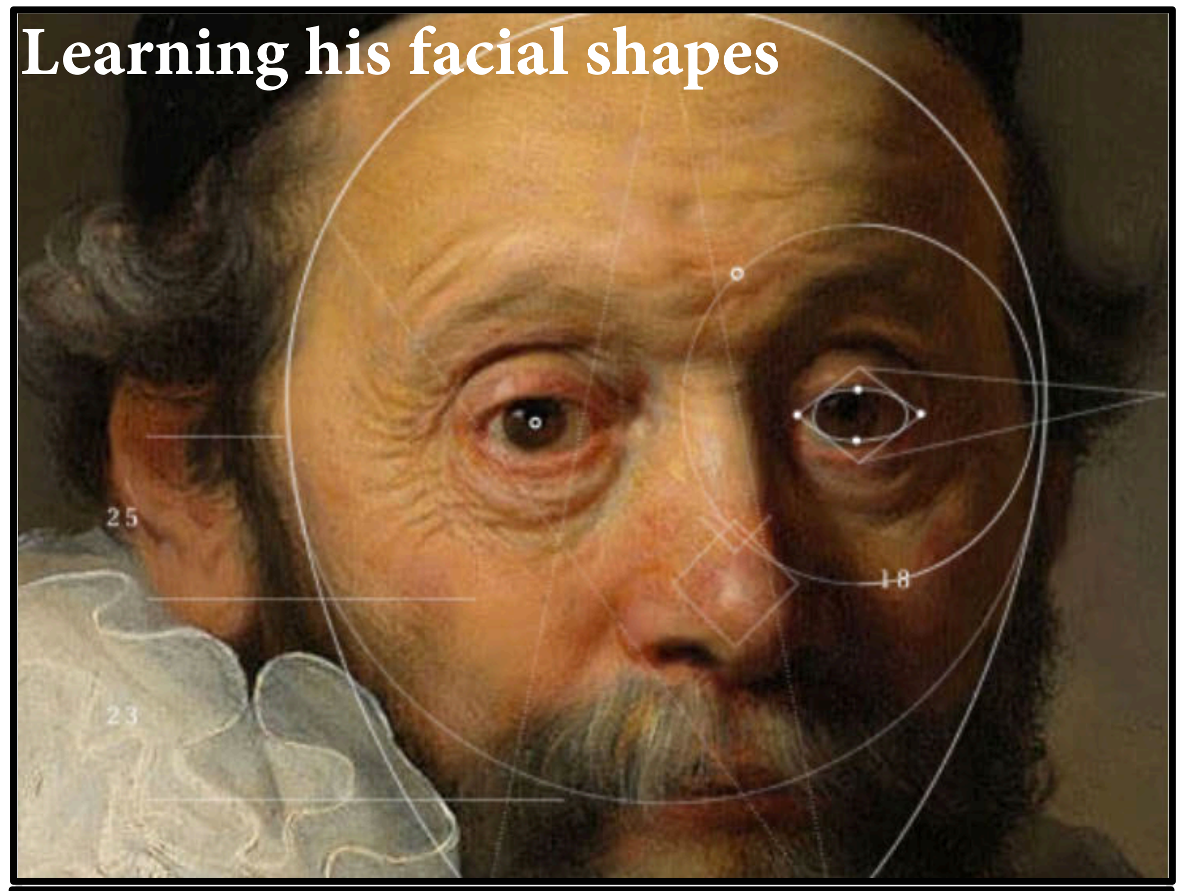
Our analysis of these works — these data — helped us understand how to paint like Rembrandt.



3

With these paintings, we coded and trained an algorithm to see what we can see through our own eyes, to identify the subject of each painting, to learn the demographics of this master painter's focus.

Visual narrative | Written narrative



4

That wasn't enough. Then, our algorithm identified the features of his subjects, like the shape of the face and eyes.

5

To make a painting like Rembrandt, though, we would need it to look and feel like his originals. Paintings have texture and depth. We measured the actual depth of his paint strokes on each of the works we collected, turning messy, real-world information into structured data for our algorithm.

6

After transforming the data generated by our algorithm, we were able to create the new painting by feeding it back into a 3D printer.

We apply this same dedication using technology to inform our customers. We can do even more with data for our clients. Let's make them curious about how we can use data to create solutions for them.

Do we think this is a story? Do the visuals add to the written narrative? Would it engage the audience? Explain.

Visual narrative | Written narrative

1

We should market ourselves as data-driven, forward-thinking, and creative.

To demonstrate our creativity, we attempted to turn paintings into data, analyze it, and create an entirely new painting perhaps indistinguishable from the master, Rembrandt. Did Rembrandt paint this? Or did *we*?

2

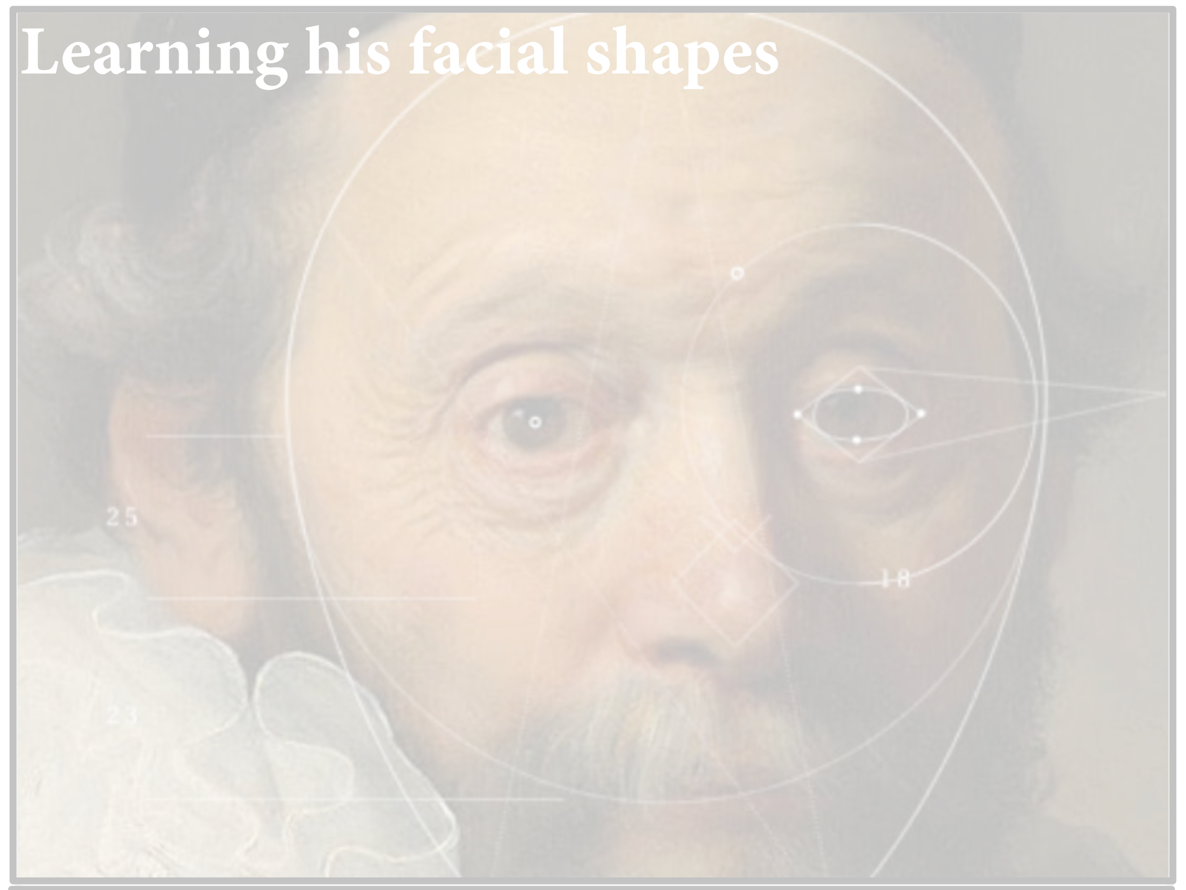
To answer the challenge, we gathered a extensive pool of data. Perhaps to some, a large collection of paintings are not data. But we partnered with museums to collect many of Rembrandt's works.

Our analysis of these works — these data — helped us understand how to paint like Rembrandt.

3

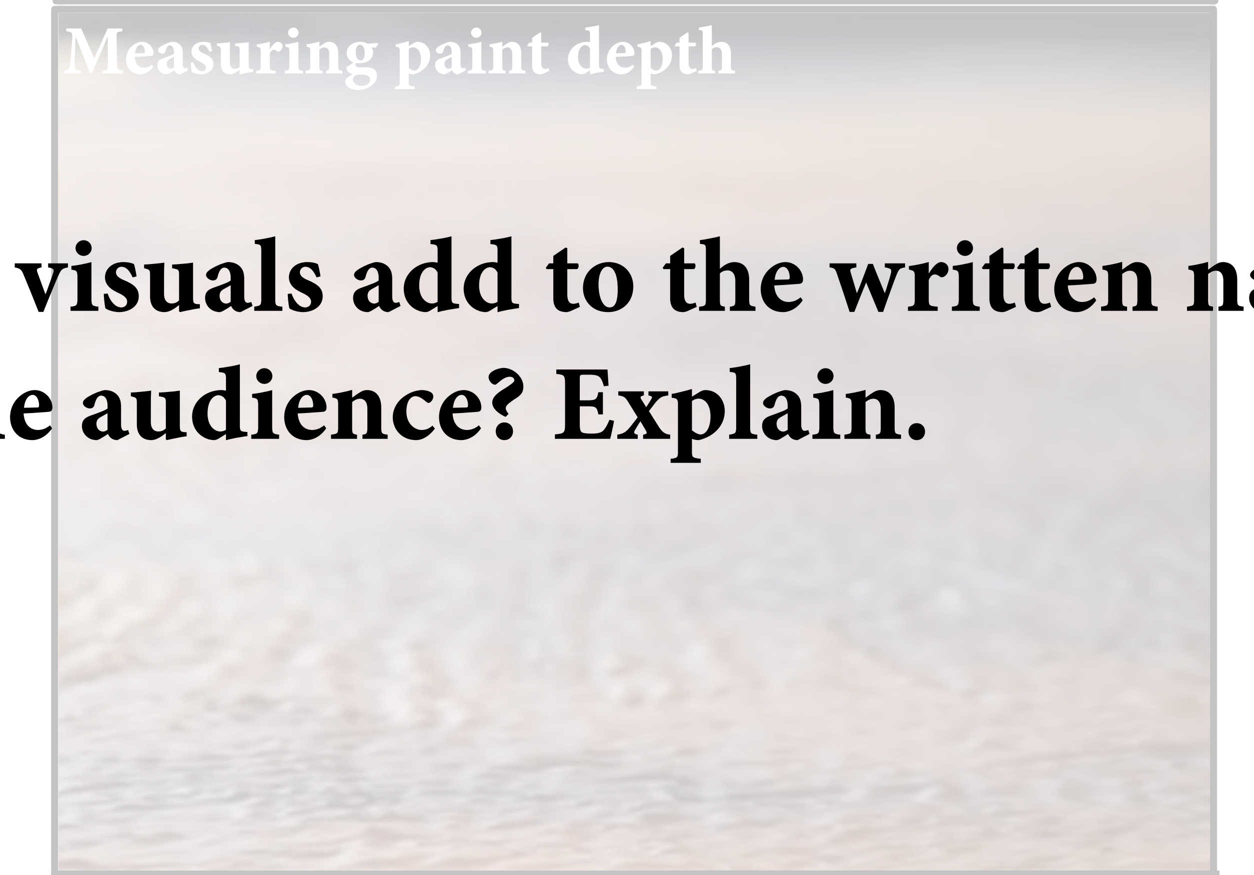
With these paintings, we coded and trained an algorithm to see what we can see through our own eyes, to identify the subject of each painting, to learn the demographics of this master painter's focus.

Visual narrative | Written narrative



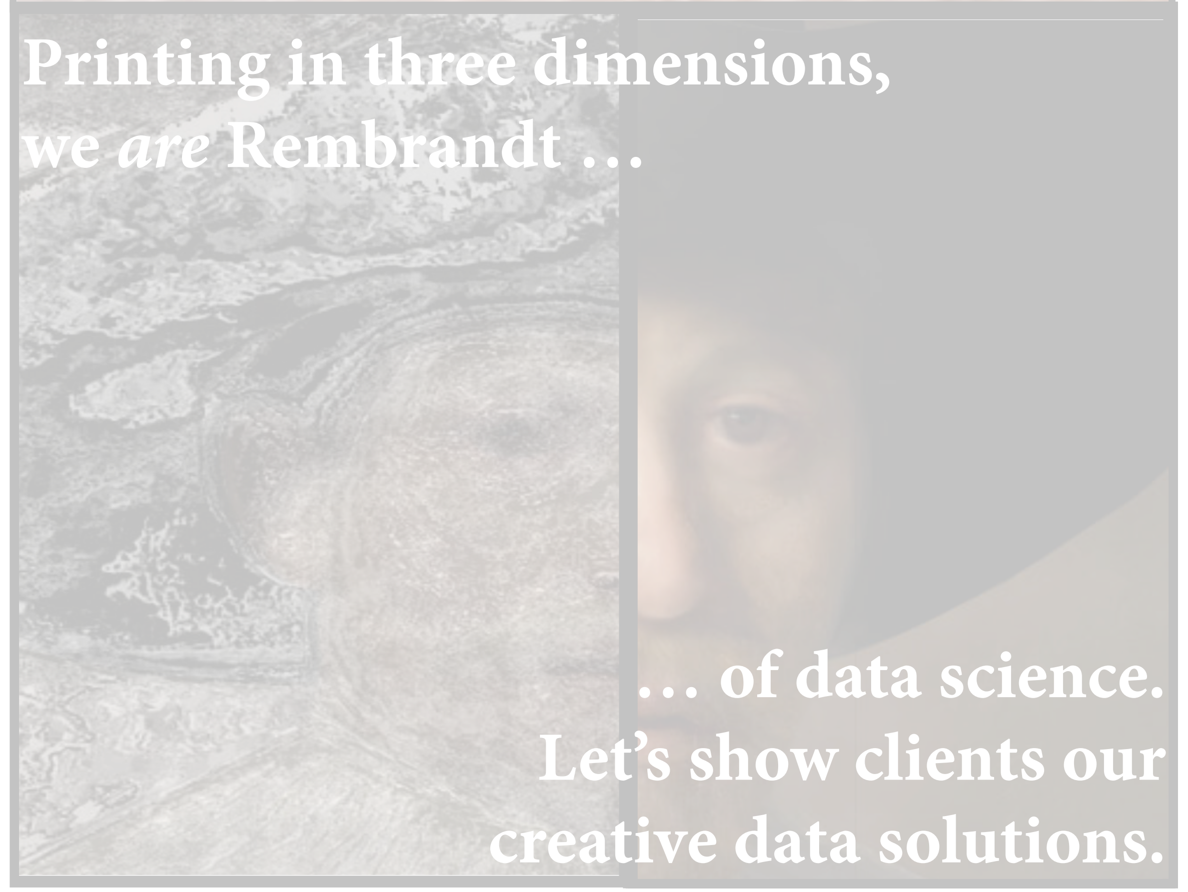
4

That wasn't enough. Then, our algorithm identified the features of his subjects, like the shape of the face and eyes.



5

To make a painting like Rembrandt, though, we would need it to look and feel like his originals. Paintings have texture and depth. We measured the actual depth of his paint strokes on each of the works we collected, turning messy, real-world information into structured data for our algorithm.



6

After transforming the data generated by our algorithm, we were able to create the new painting by feeding it back into a 3D printer.

We apply this same dedication using technology to inform our customers. We can do even more with data for our clients. Let's make them curious about how we can use data to create solutions for them.

use **data graphics** to support and amplify your narrative

Why show data graphically?

Classic example, datasets from Anscombe

1		2		3		4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Anscombe

Classic example, datasets from Anscombe

1		2		3		4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary statistics of data: *are they the same*?

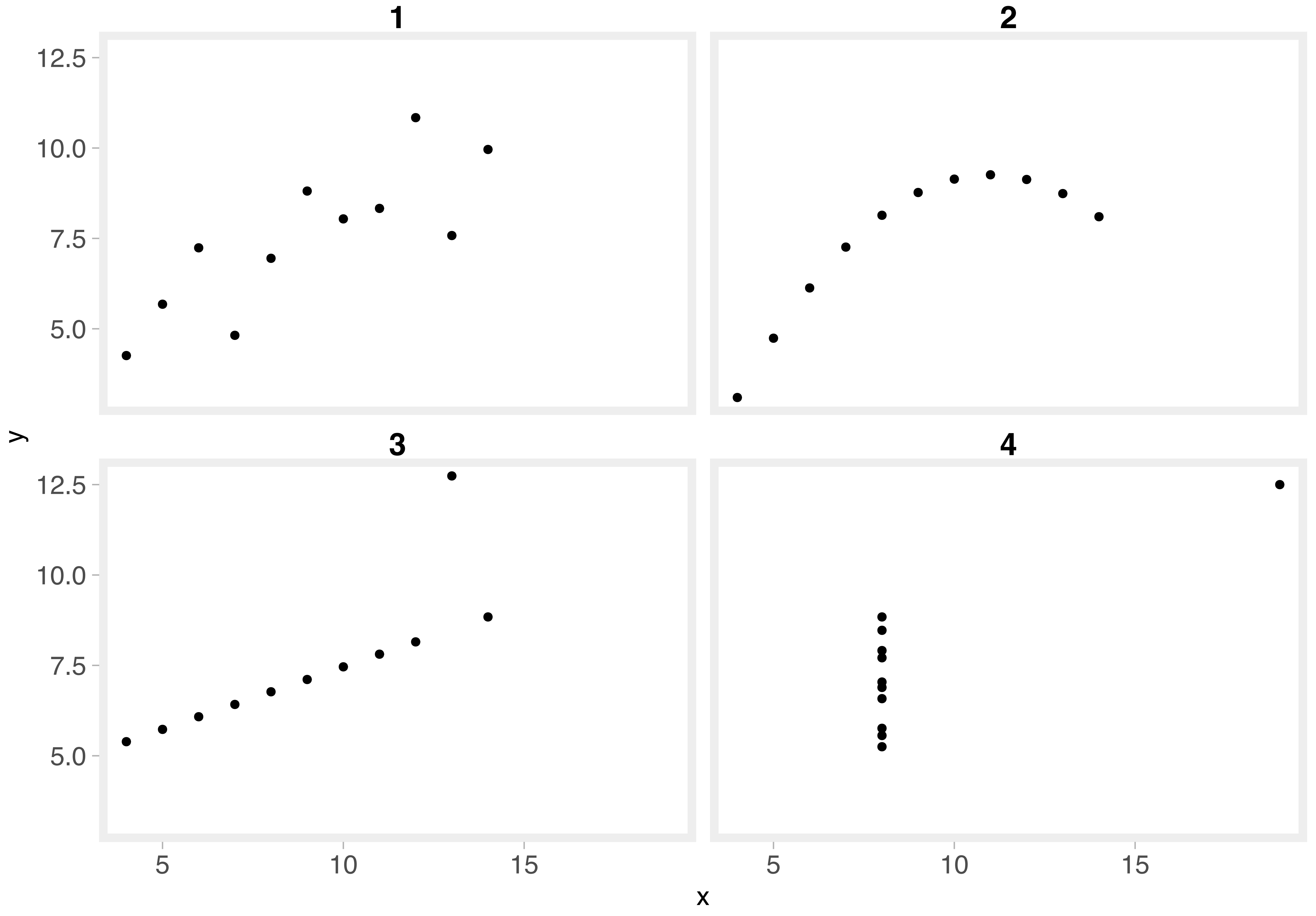
	1		2		3		4	
	x	y	x	y	x	y	x	y
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
sd	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03

Parameter	Mean	Std Err	t-val	p-val
Dataset 1				
(Intercept)	3.000	1.125	2.667	0.026
x	0.500	0.118	4.241	0.002
Dataset 2				
(Intercept)	3.001	1.125	2.667	0.026
x	0.500	0.118	4.239	0.002
Dataset 3				
(Intercept)	3.002	1.124	2.670	0.026
x	0.500	0.118	4.239	0.002
Dataset 4				
(Intercept)	3.002	1.124	2.671	0.026
x	0.500	0.118	4.243	0.002

Classic example, datasets from Anscombe

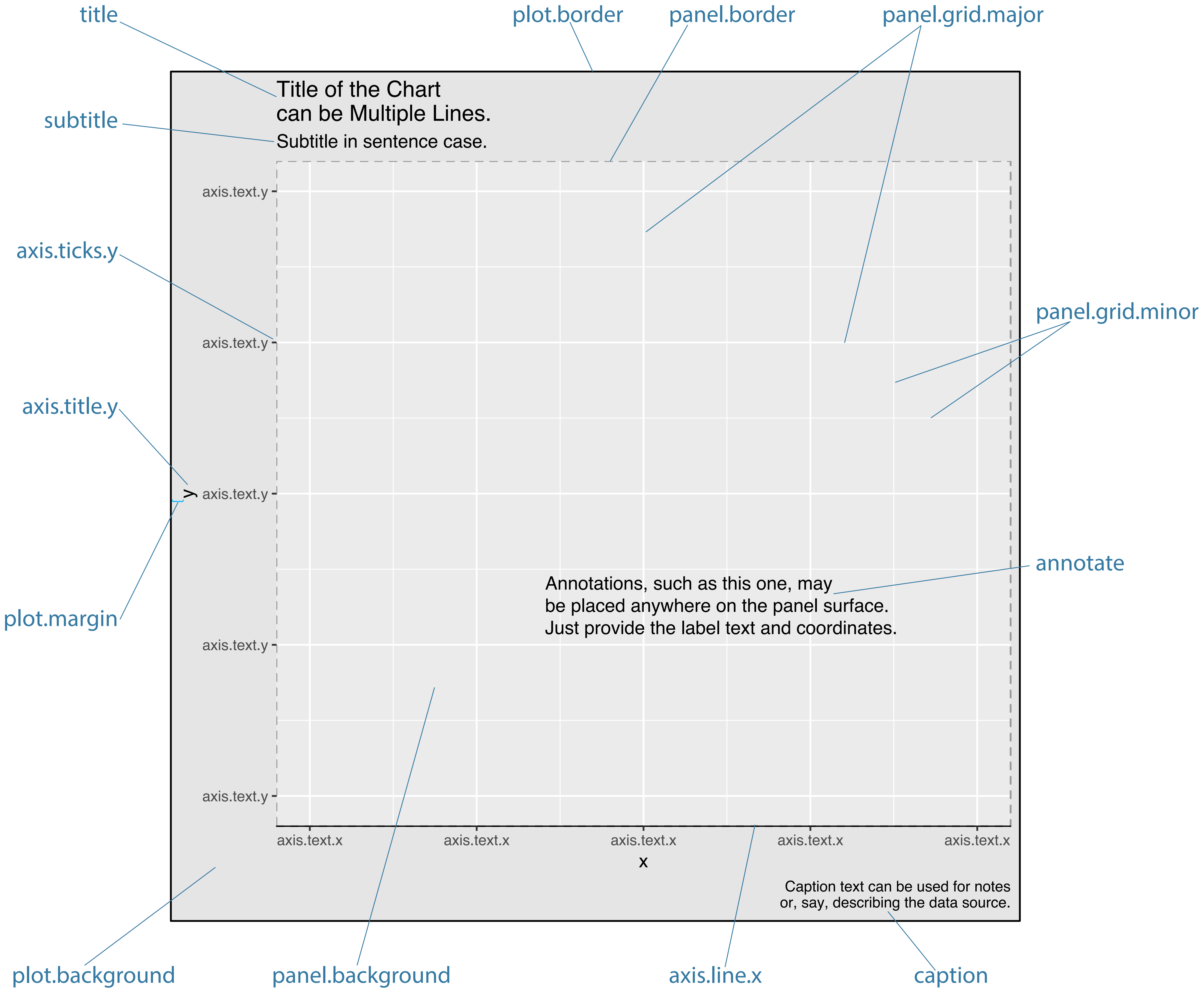
1		2		3		4	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Graphics show how the datasets are different

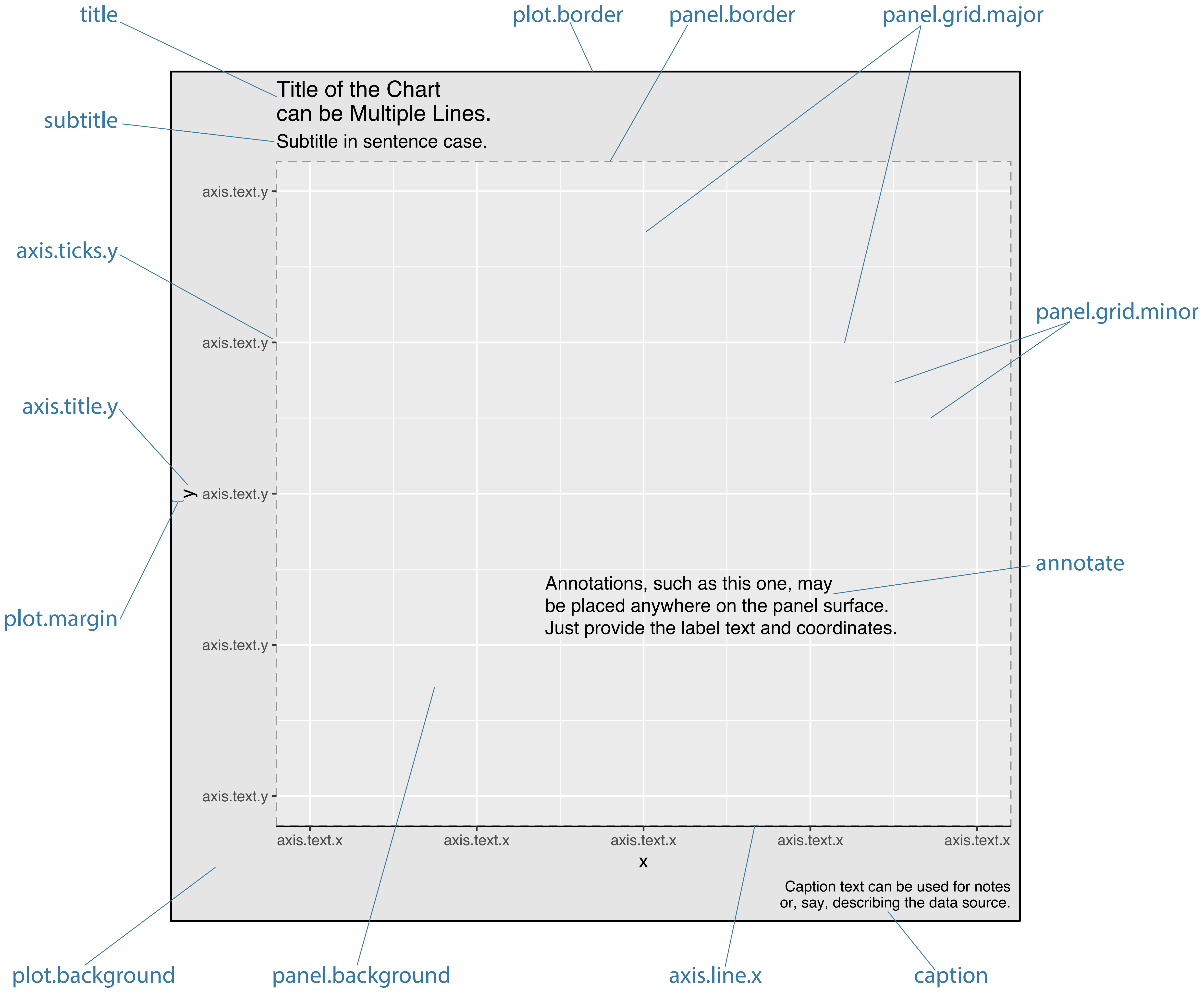


graphics, the *non-data-ink*

graphics, the *non-data-ink*



graphics, the *non-data-ink*



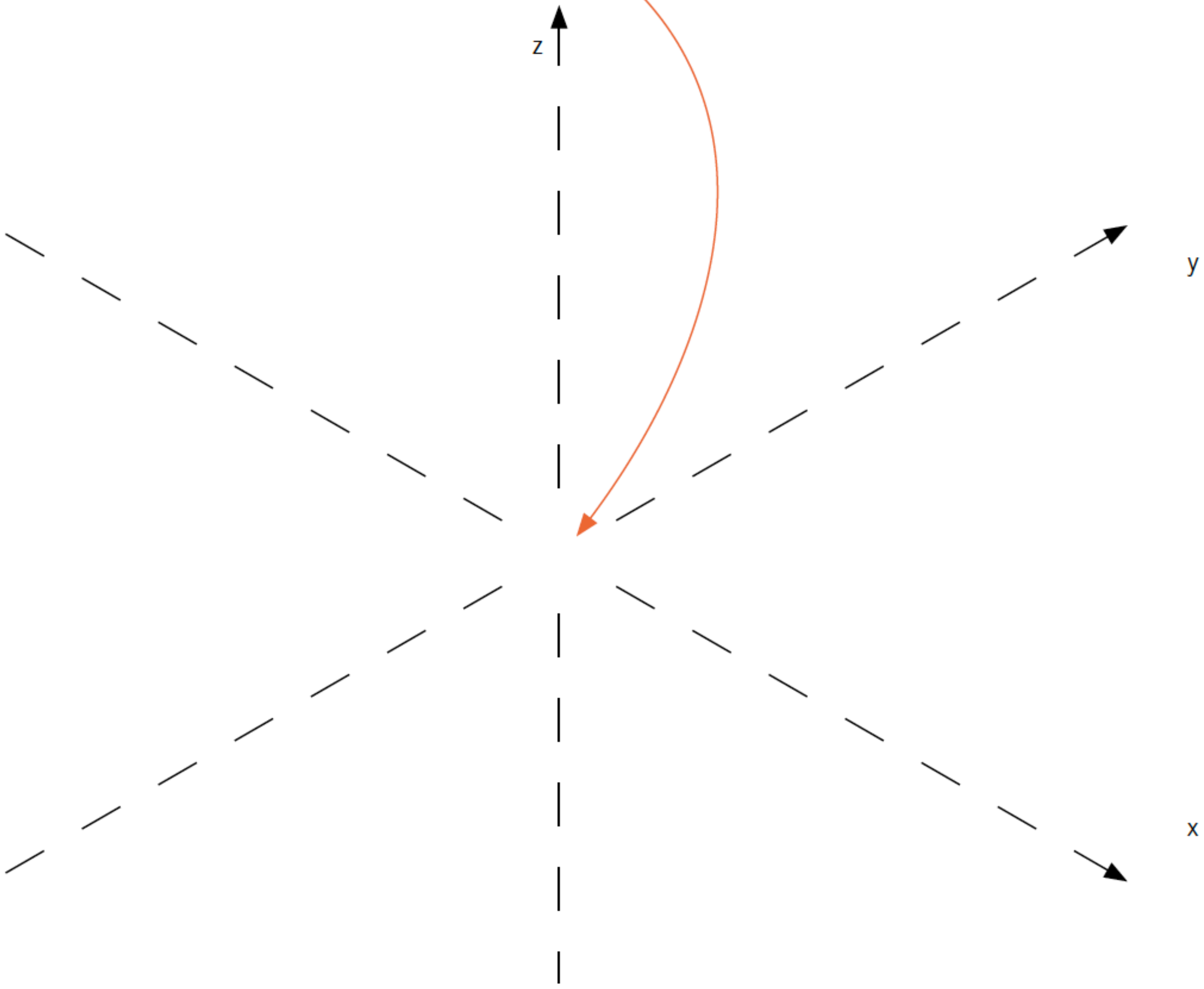
Coding graphic elements, *example in R/GGplot2*

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

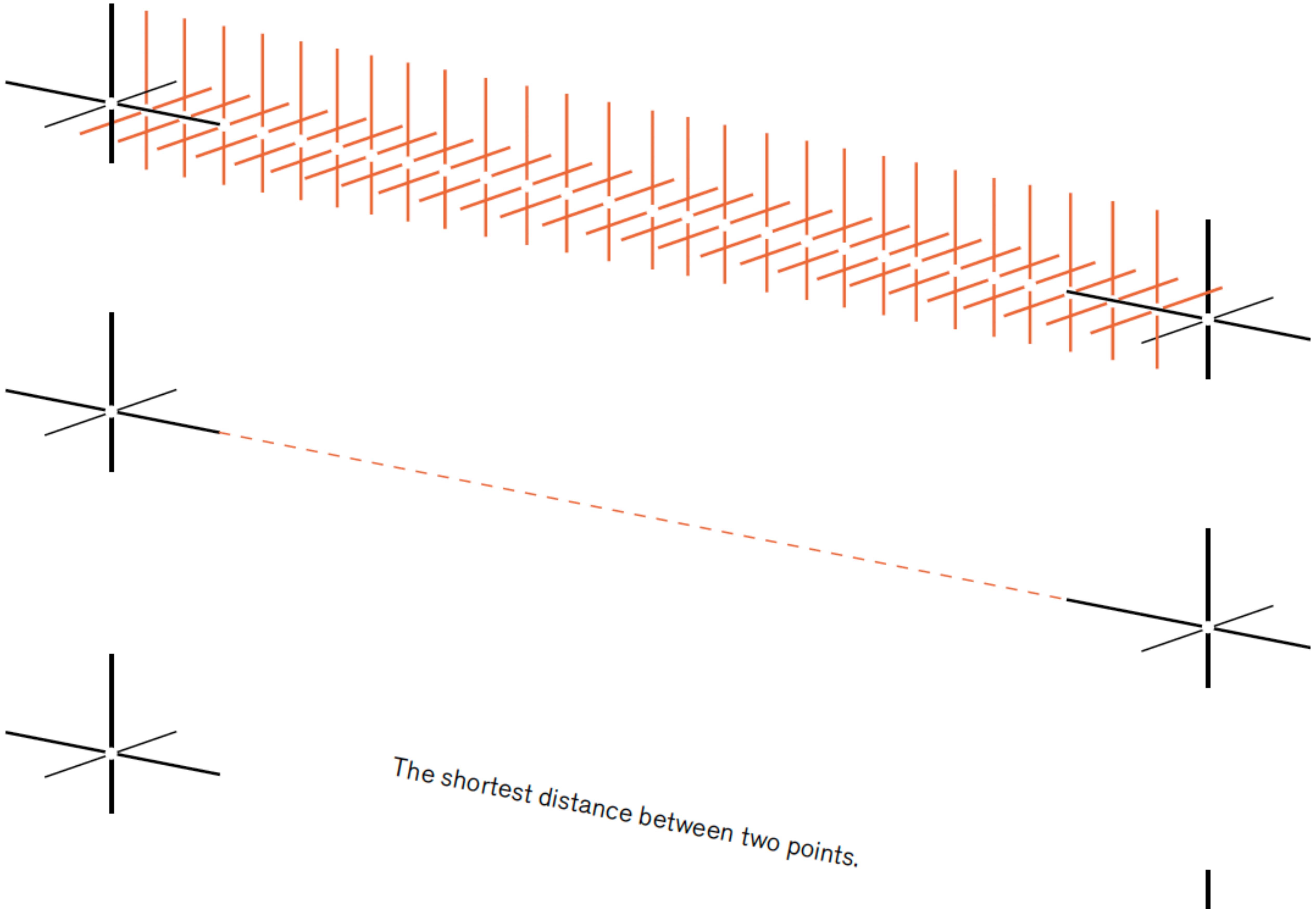
graphics, *points, lines, surfaces, volumes*

Point. You cannot see or feel a point; it is a place without area. The point has a position that can be defined by coordinates (numbers on one, two, or three axes).

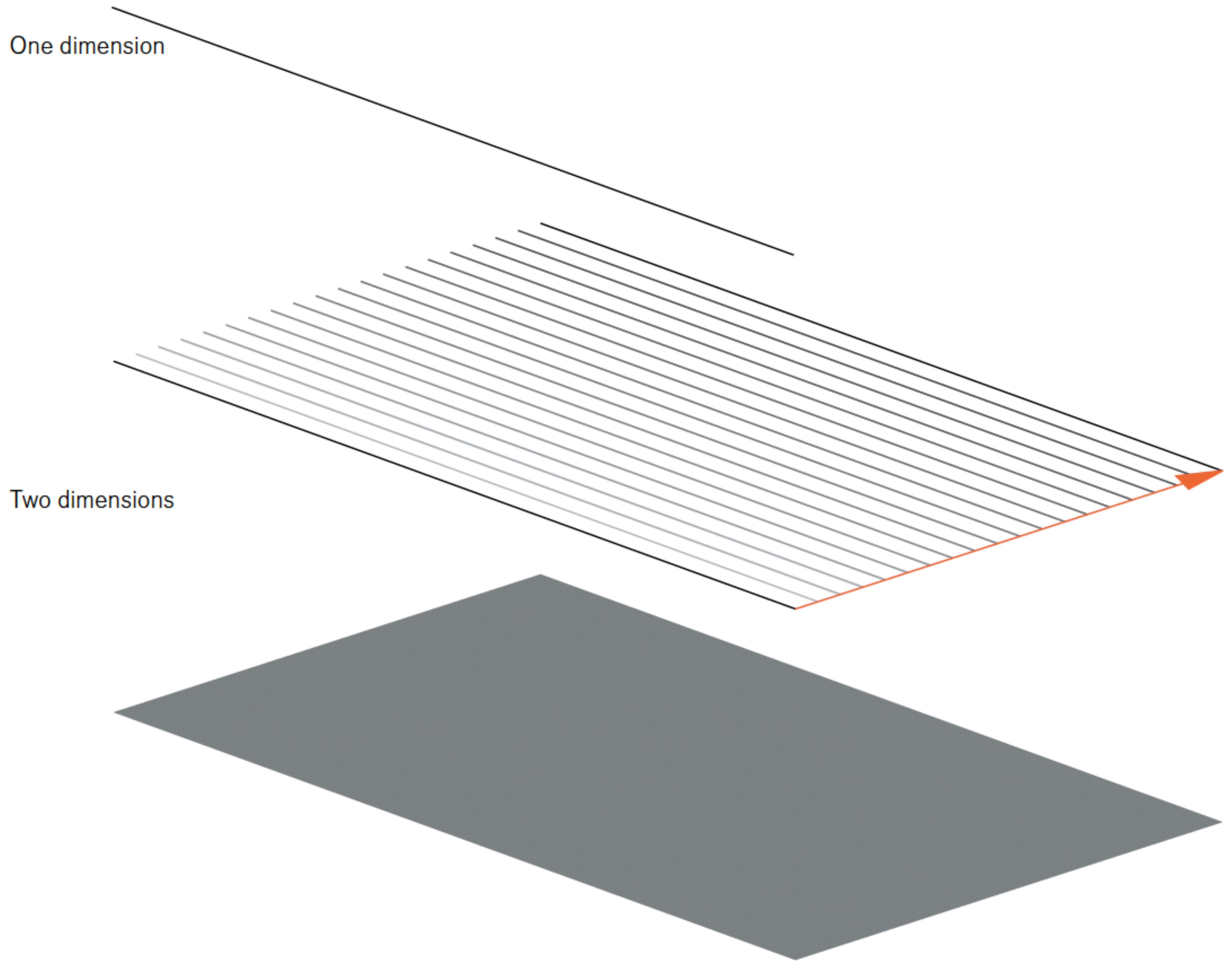


Leborg

Line. A line can be understood as a number of points that are adjacent to one another. A line can be infinite or have two endpoints. The shortest distance between two points is a straight line.

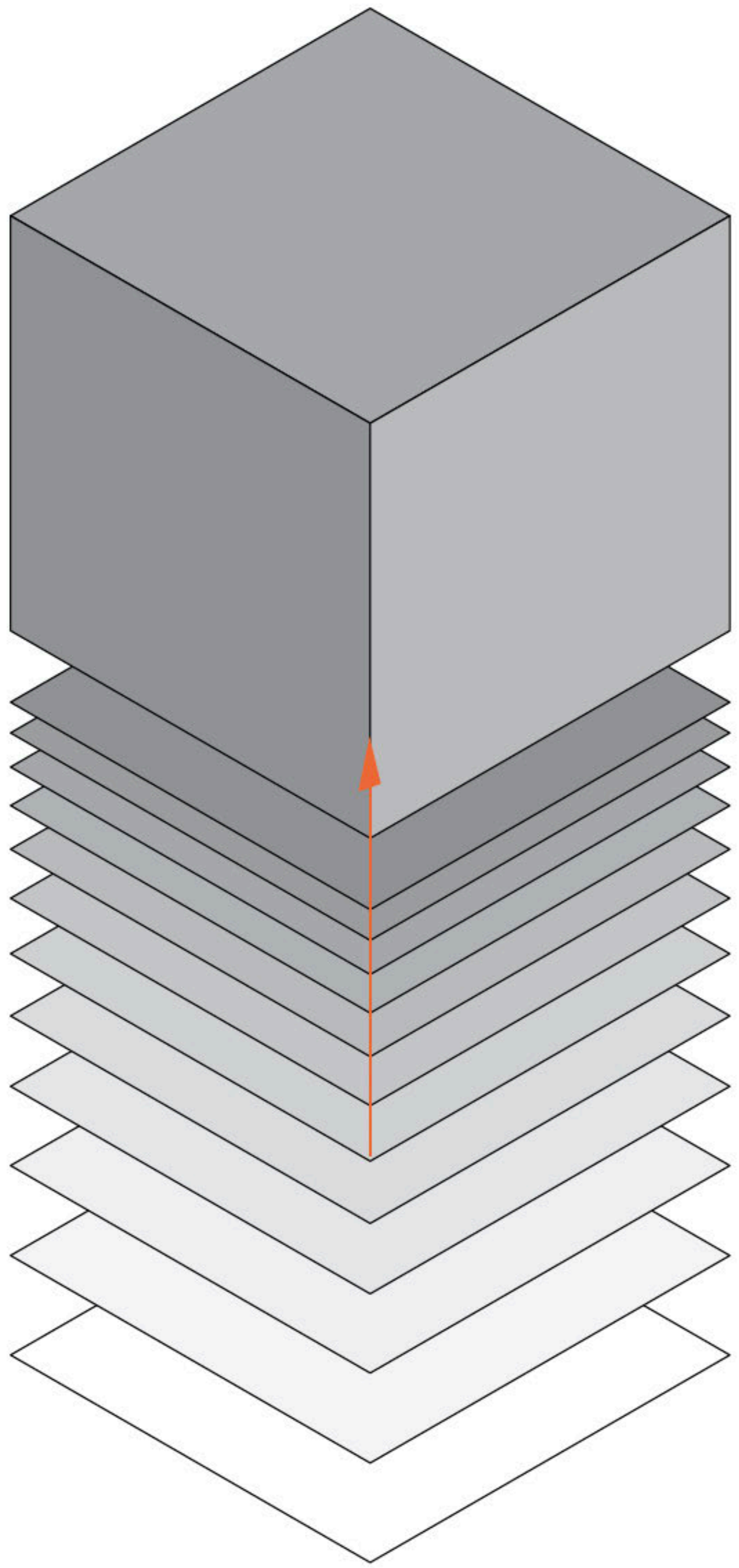


Leborg



Surface. A surface is defined by two lines that do not coincide or by a minimum of three points that are not located on a line. If the two lines have one coinciding point, the surface will be a plane.

Leborg

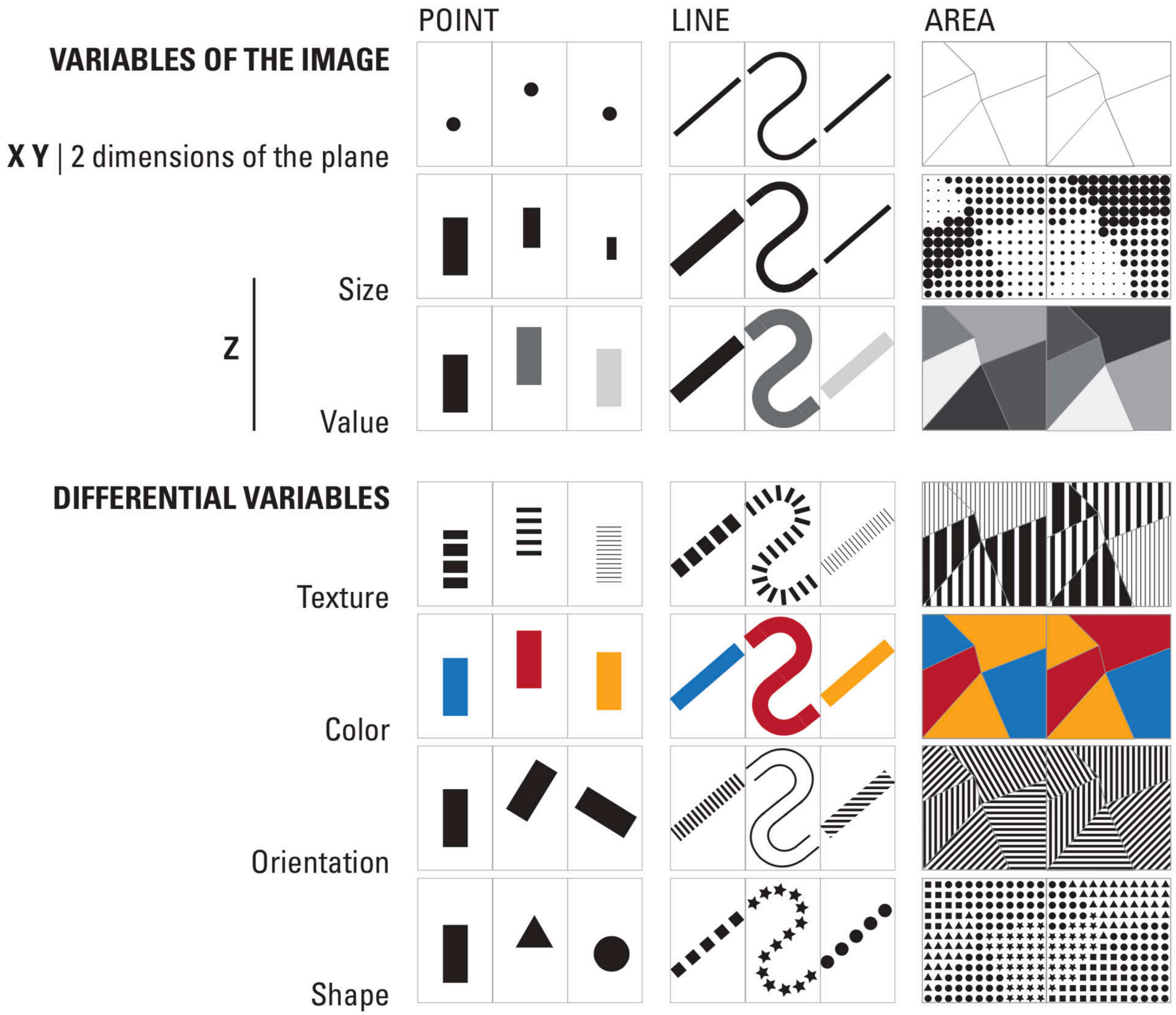


Volume. A volume is an empty space defined by surfaces, lines, and points.

Leborg

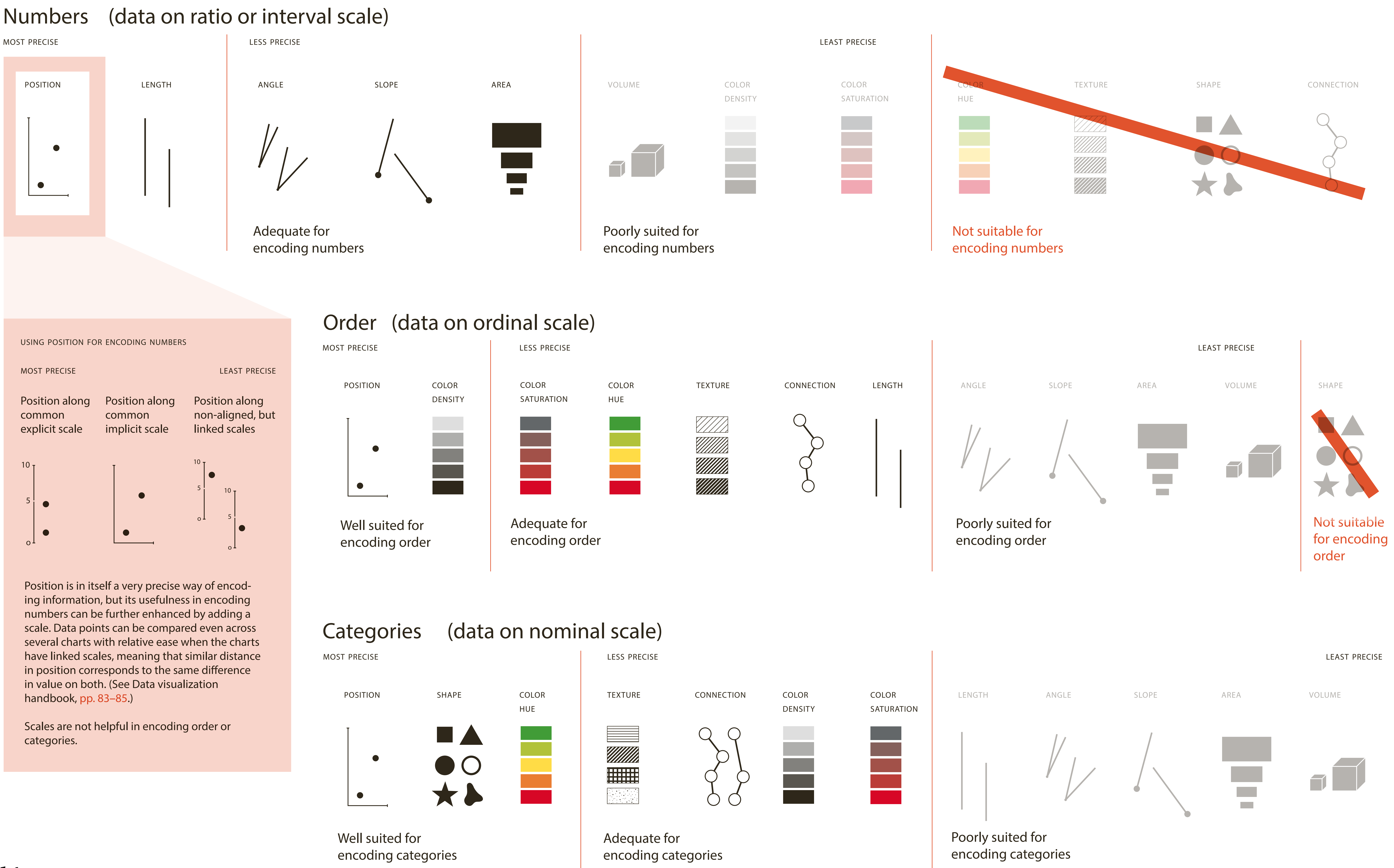
graphics, *data-ink encodings*

graphics, options for *data-ink encodings*



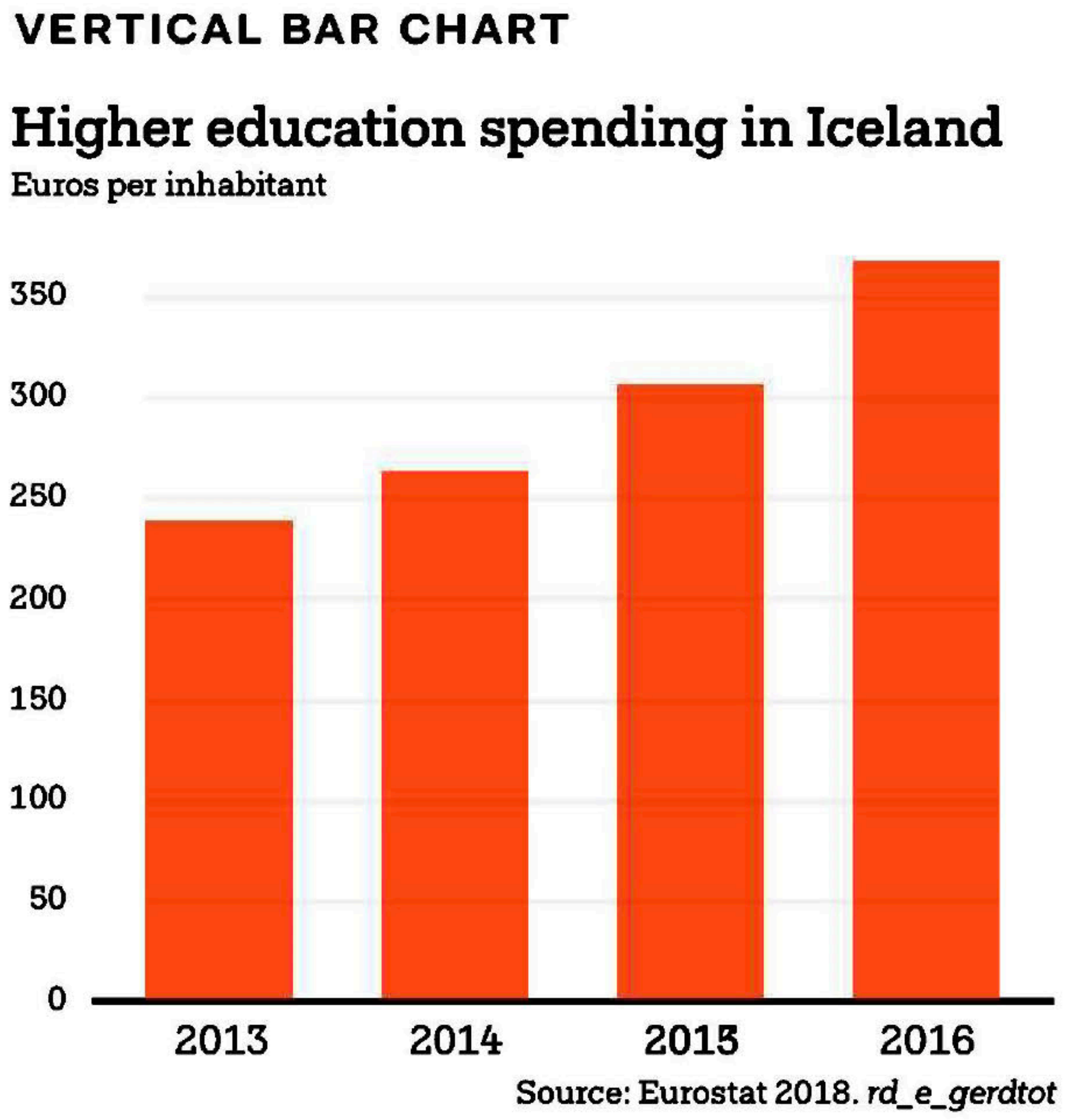
Bertin

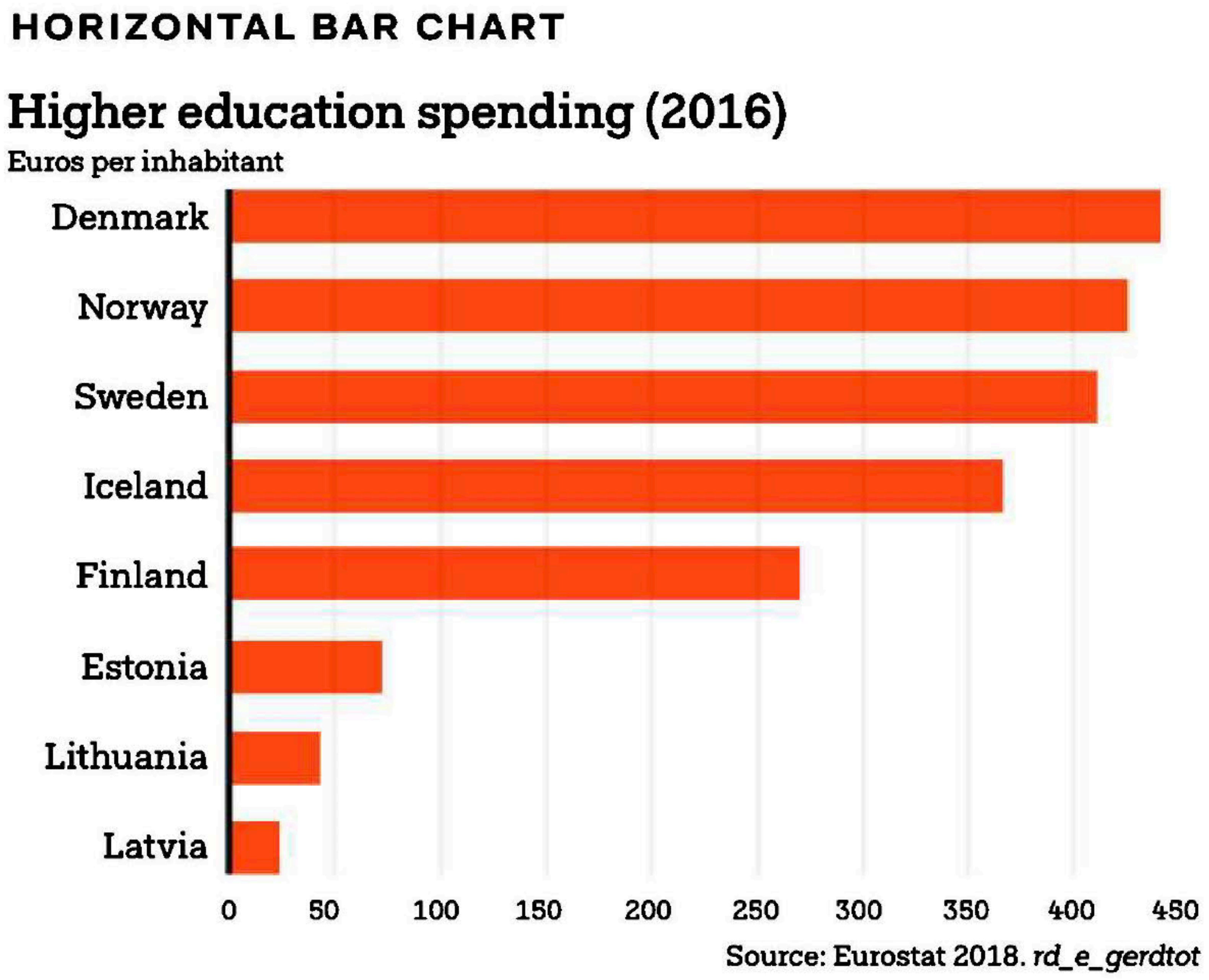
graphics, accuracy of *data-ink decodings*



Koponen & Hildén

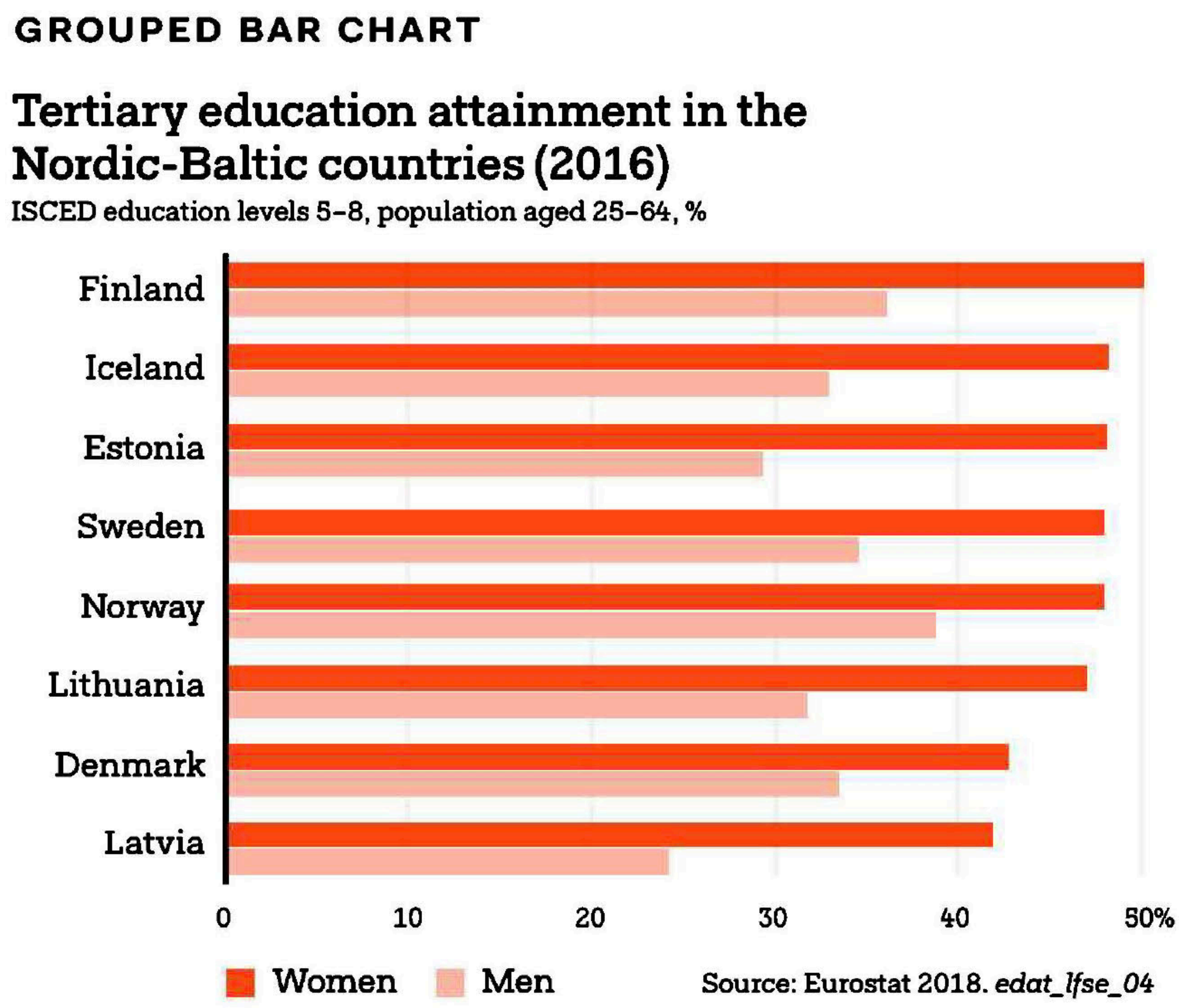
What data encodings do we find in the following named charts?

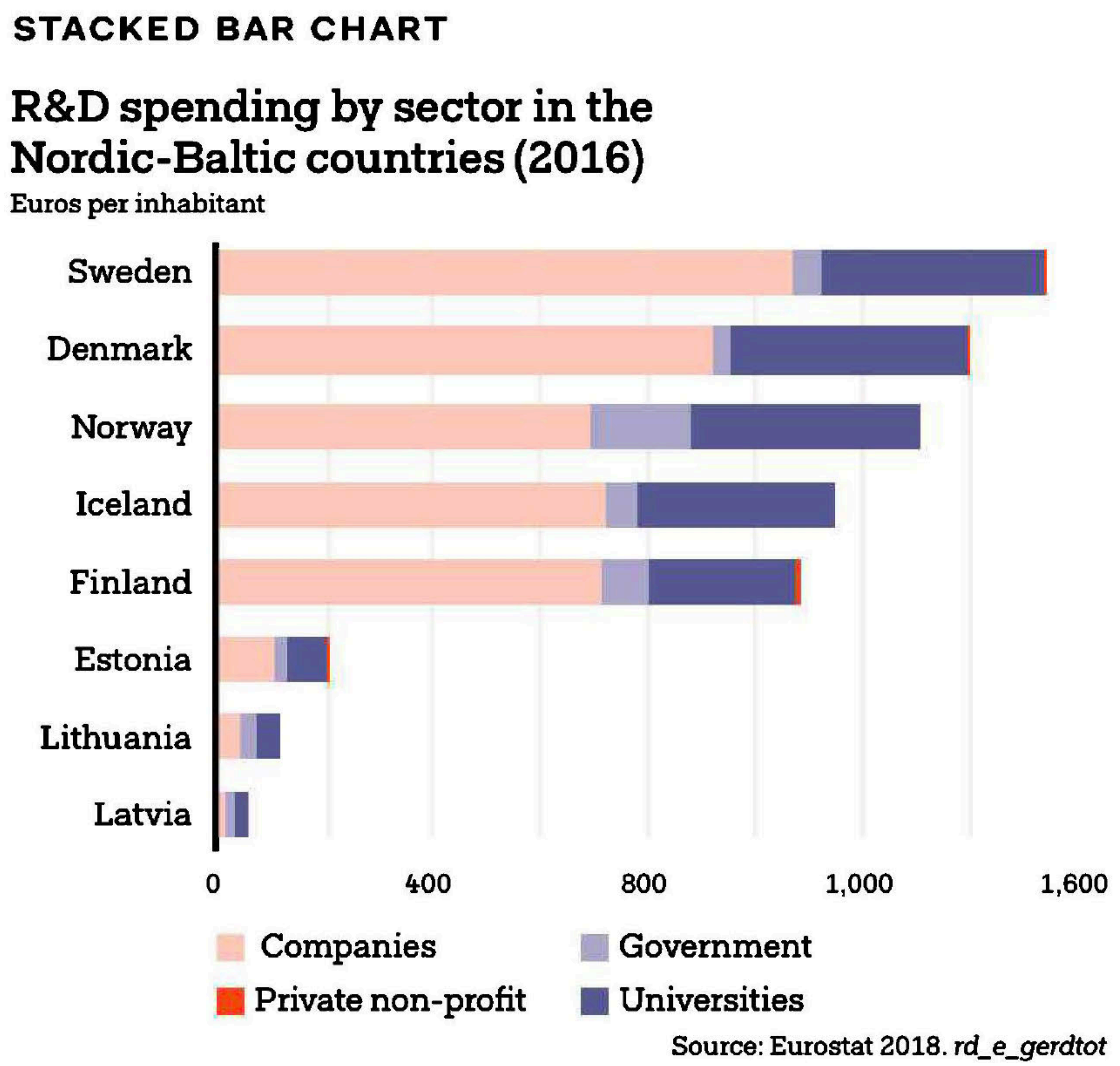


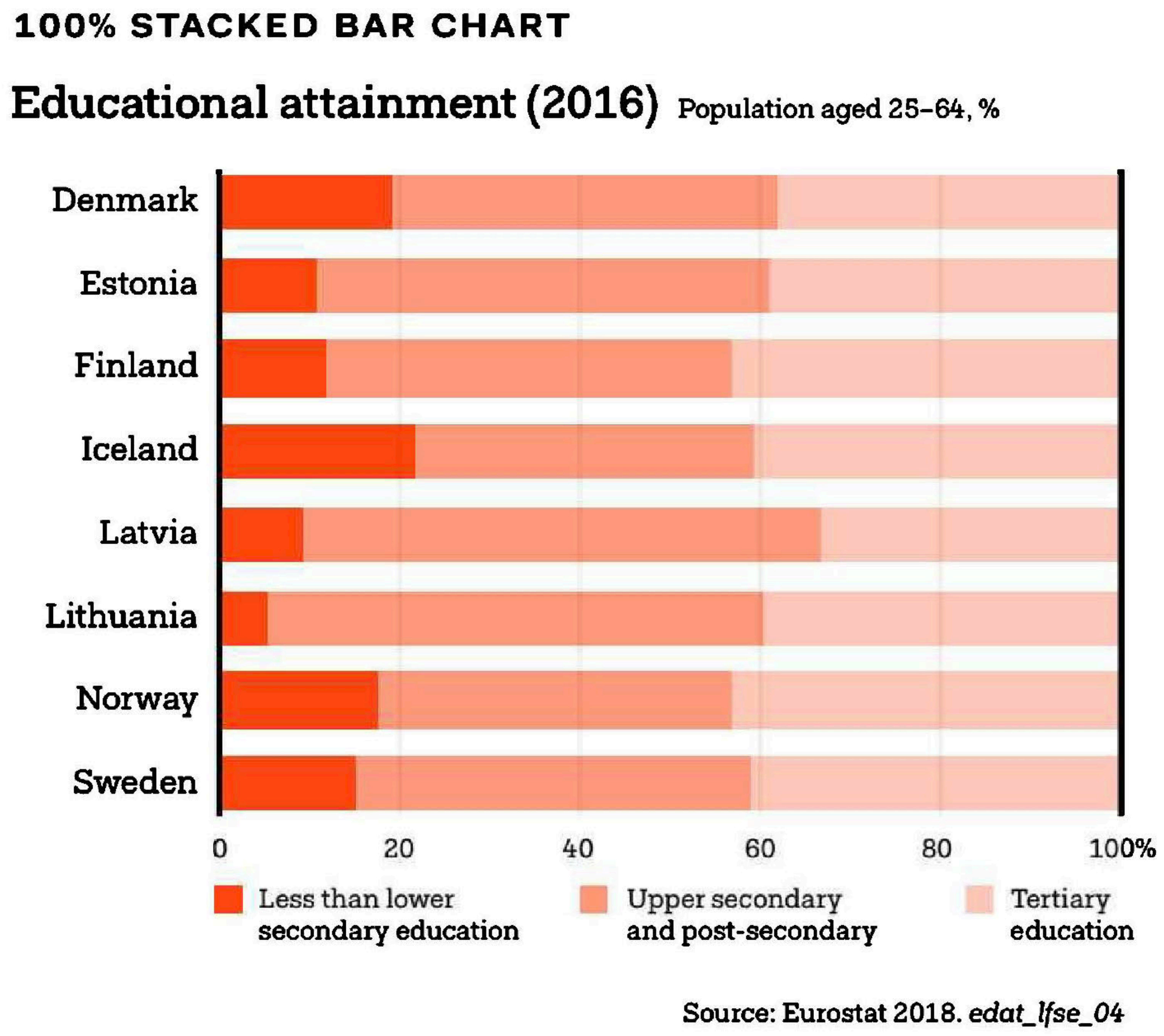


Koponen & Hildén

Common graphics, identify the *data-ink encodings*

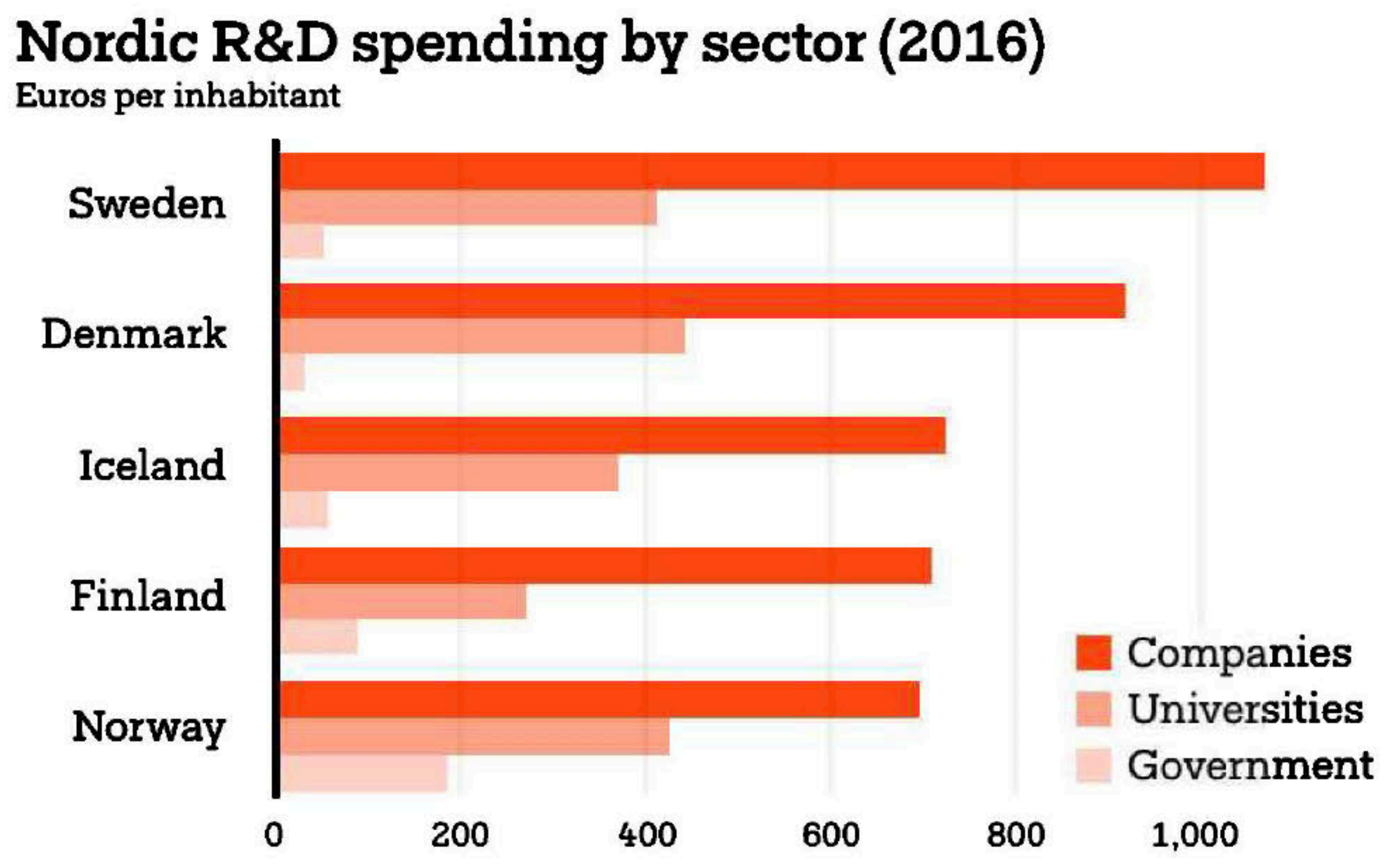


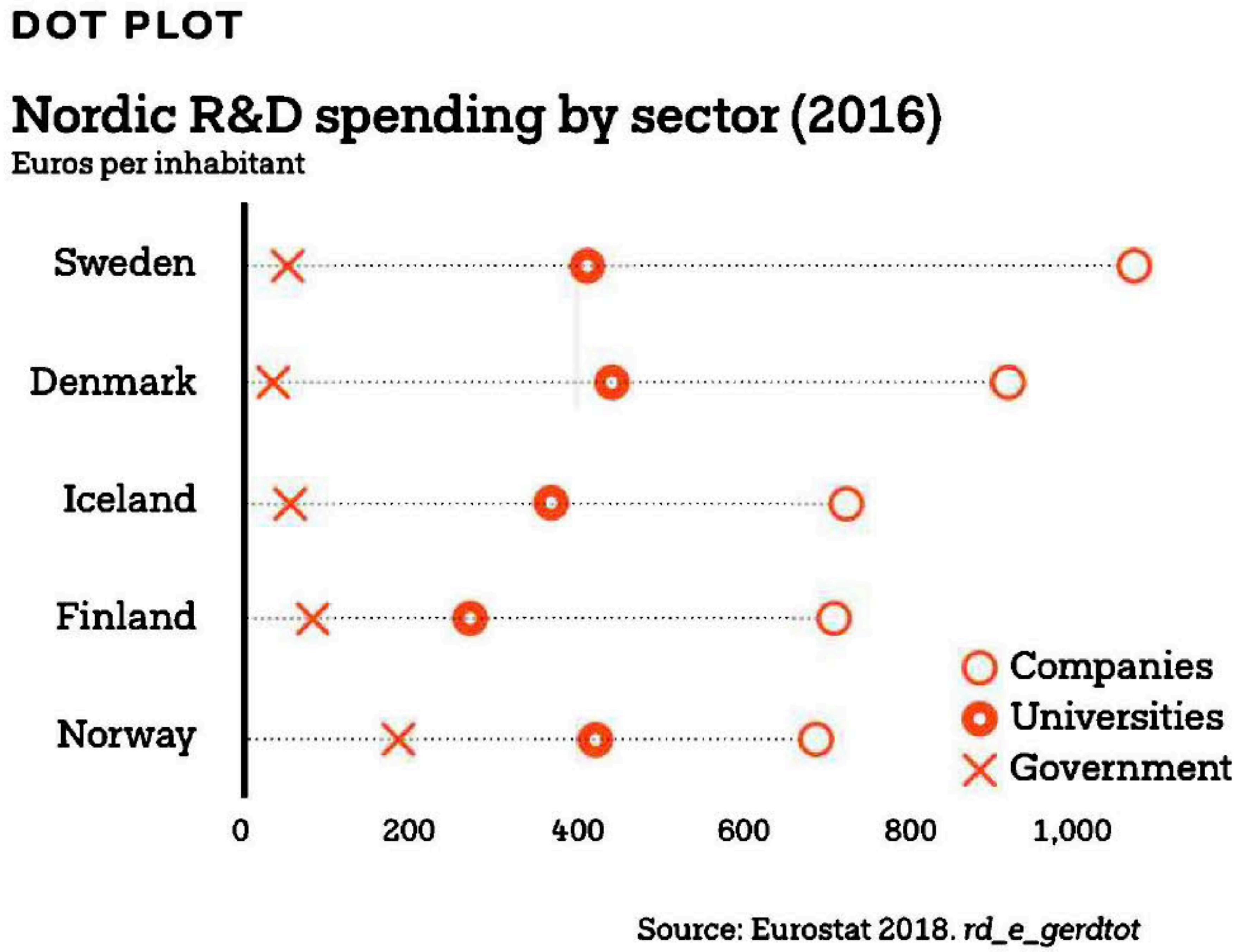




Koponen & Hildén

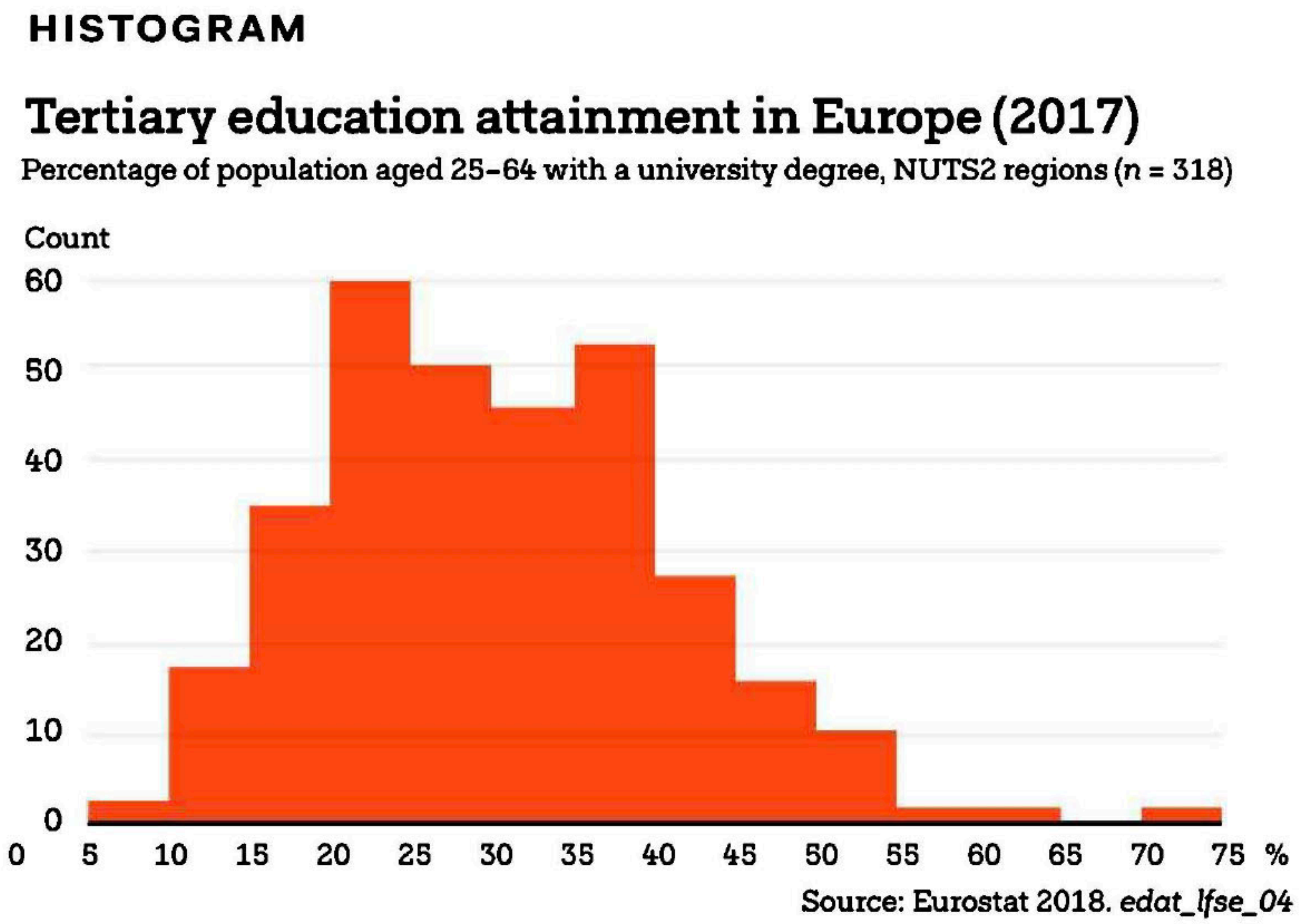
Common graphics, identify the *data-ink encodings*

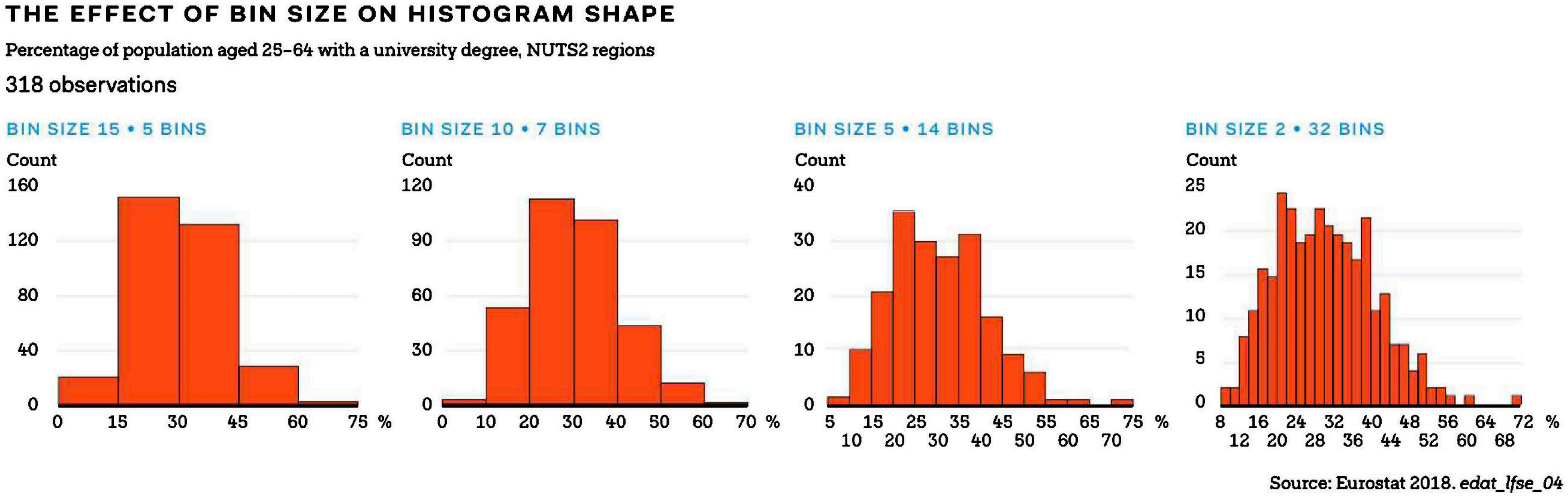




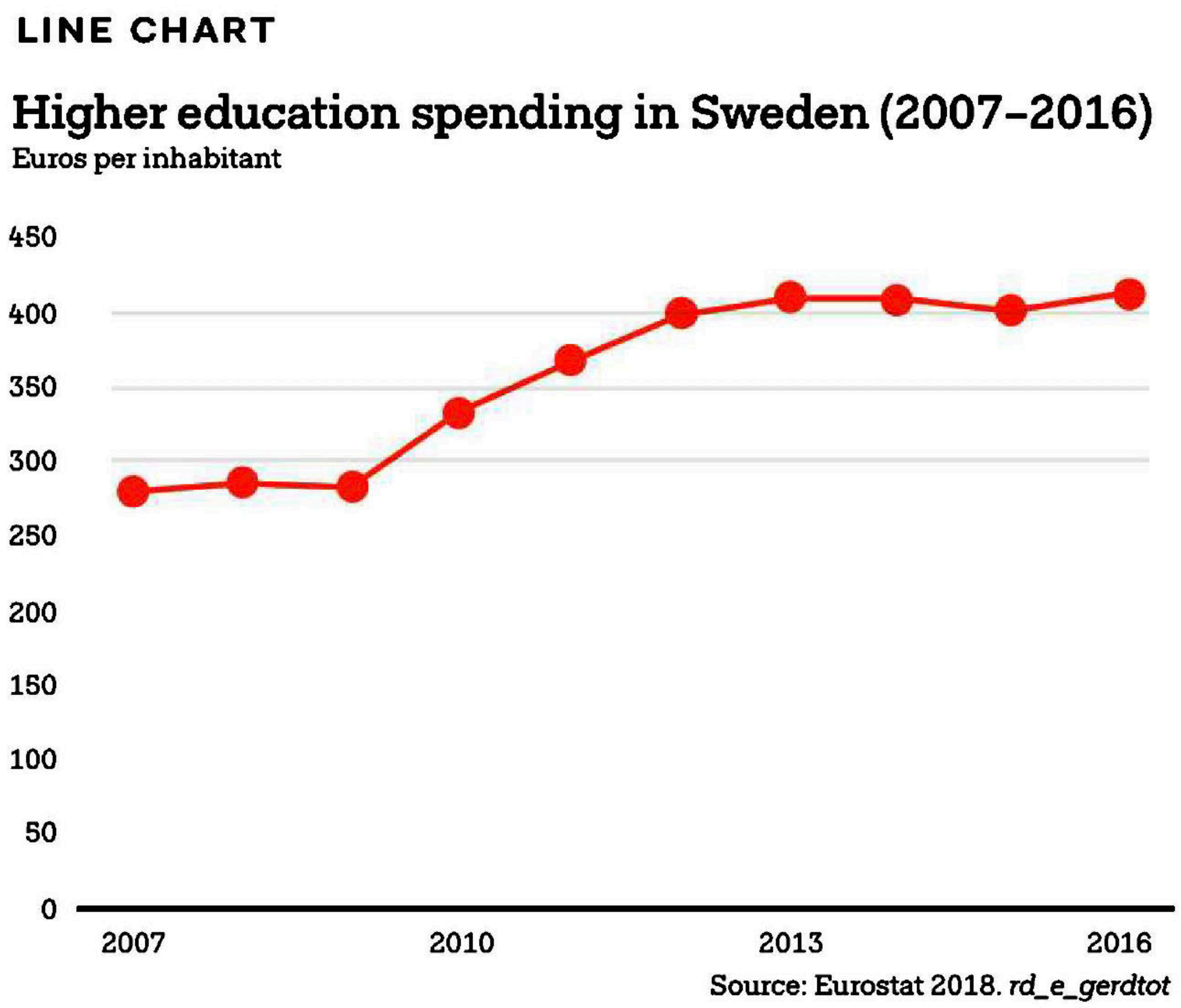
Koponen & Hildén

Common graphics, identify the *data-ink encodings*



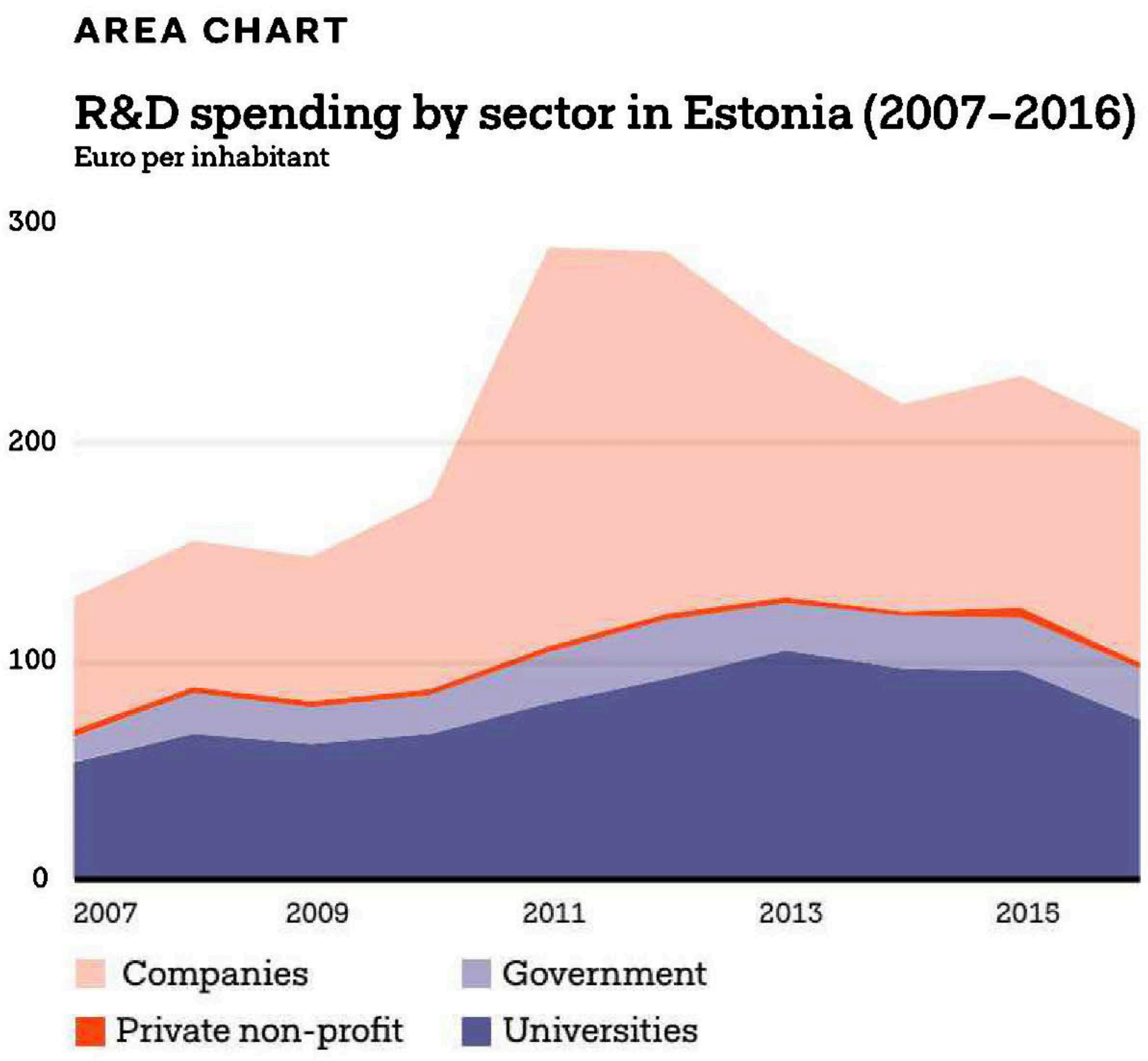


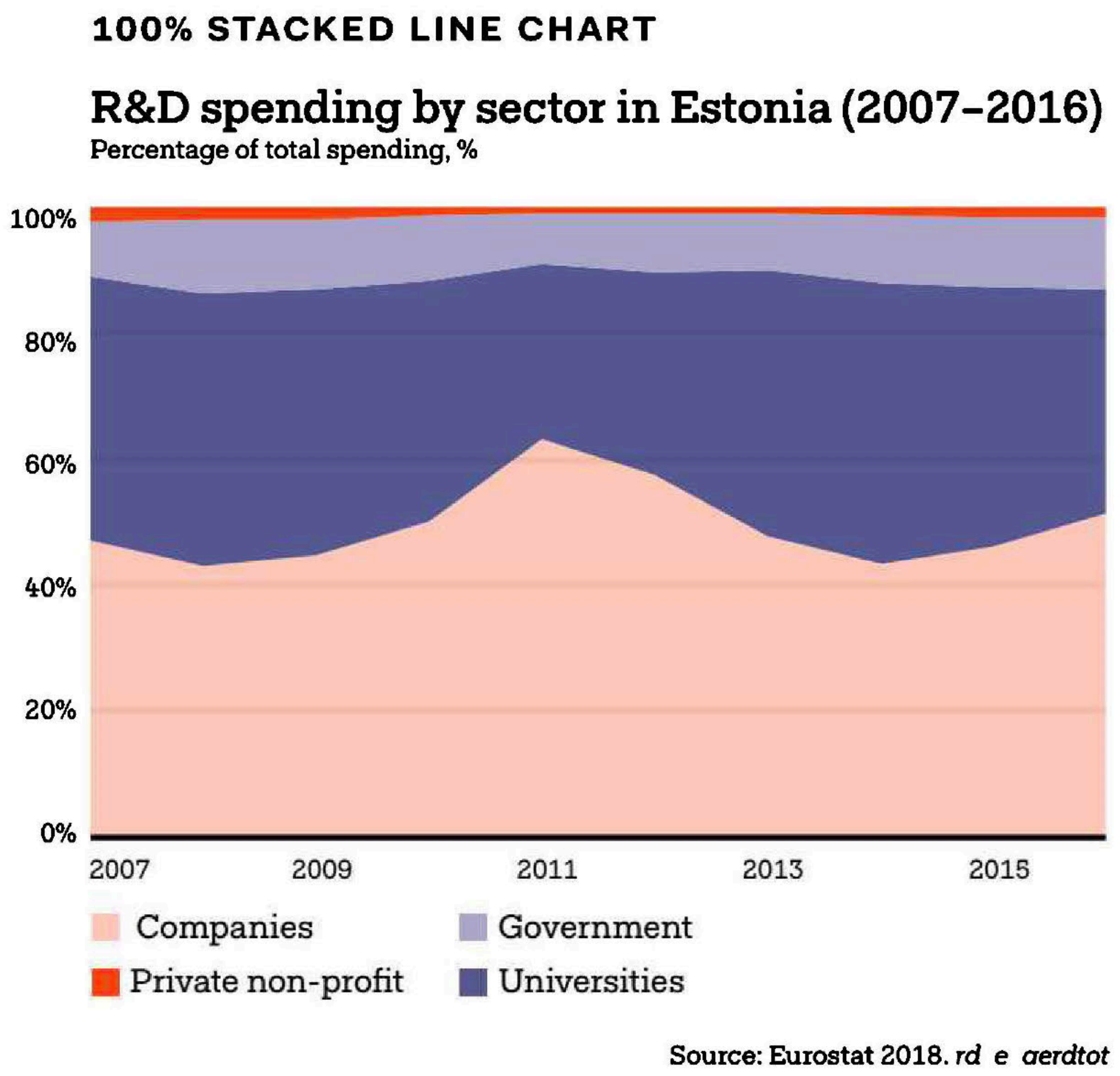
Koponen & Hildén



Koponen & Hildén

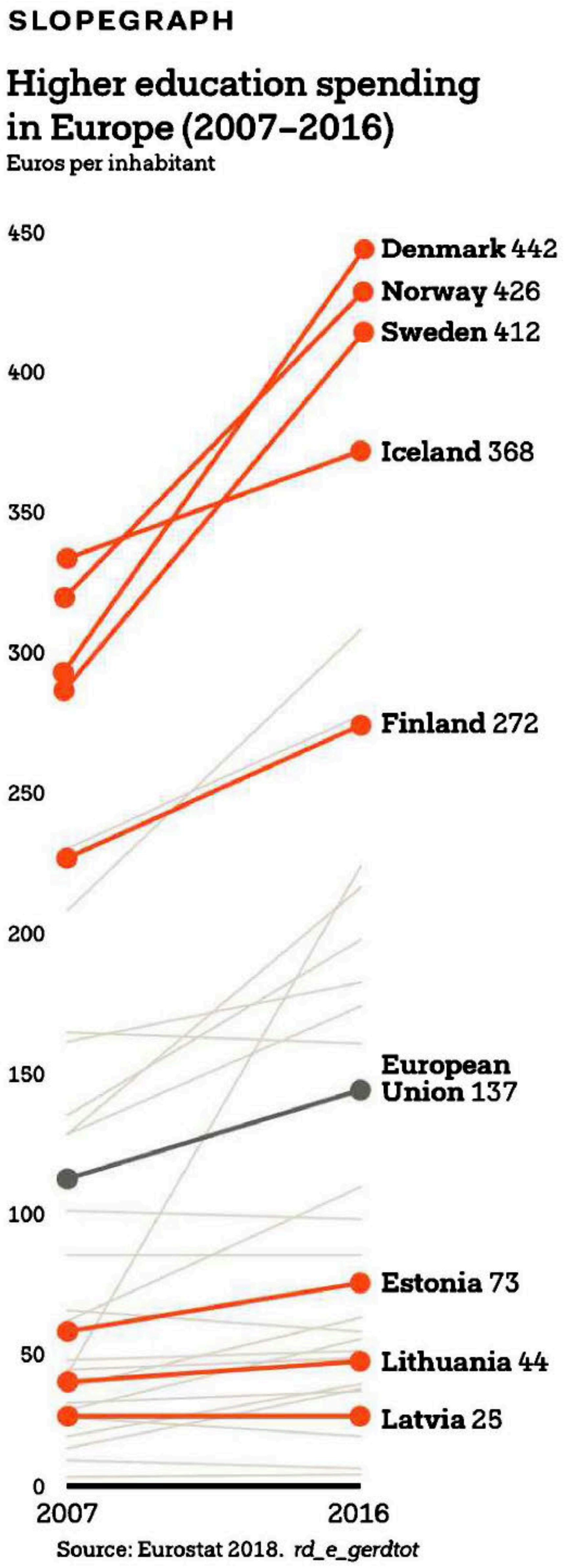
Common graphics, identify the *data-ink encodings*



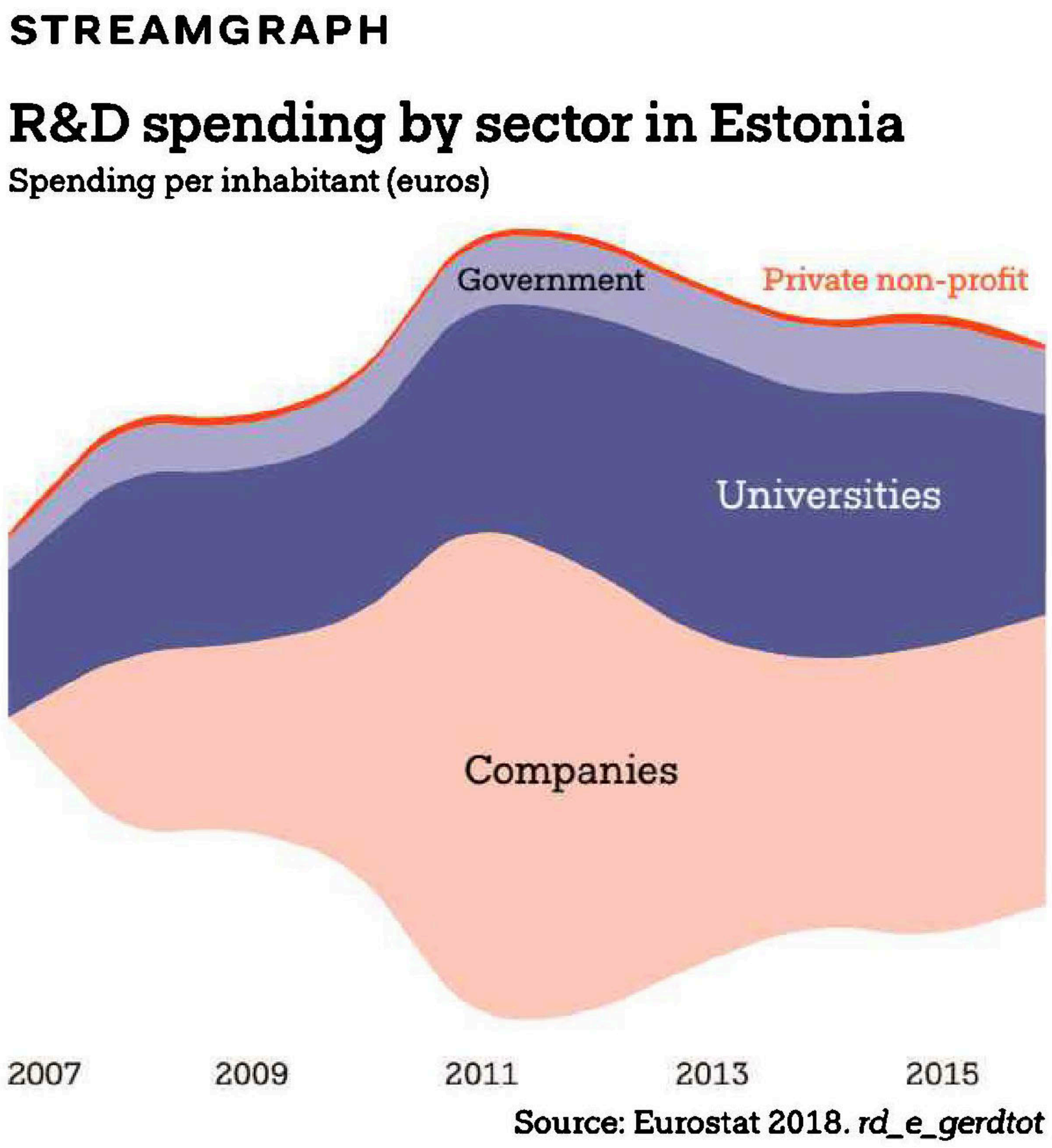


Koponen & Hildén

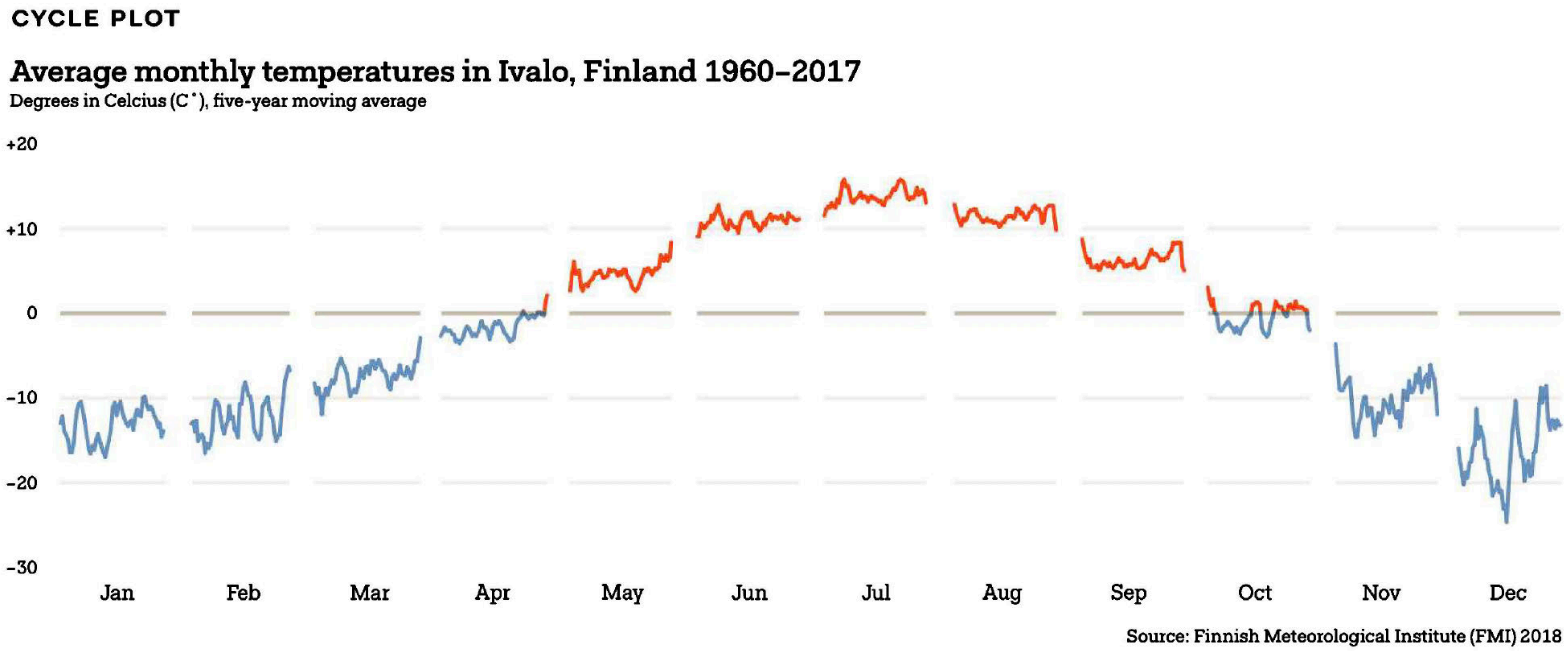
Common graphics, identify the *data-ink encodings*



Koponen & Hildén

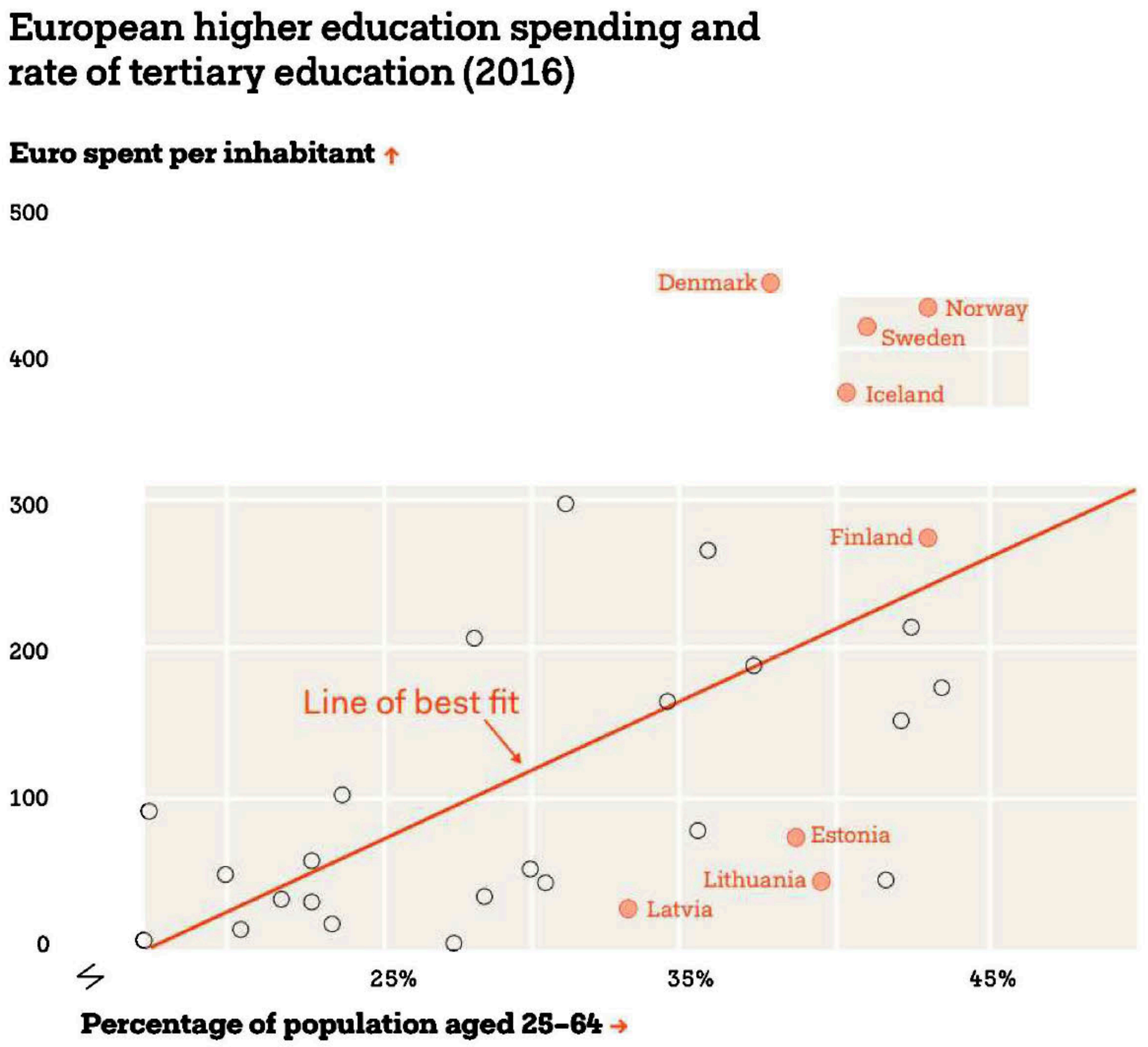


Koponen & Hildén

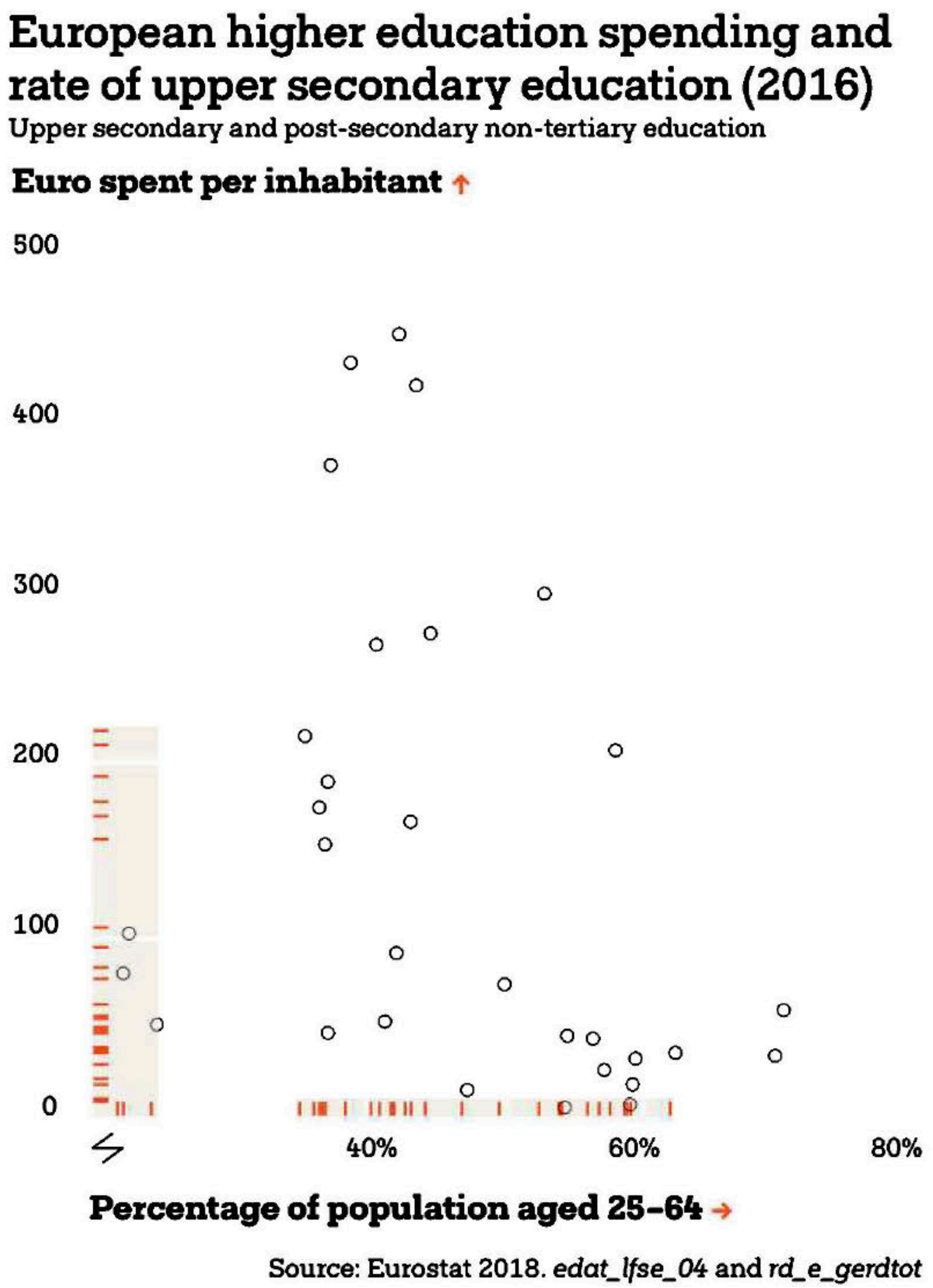


Koponen & Hildén

SCATTERPLOT

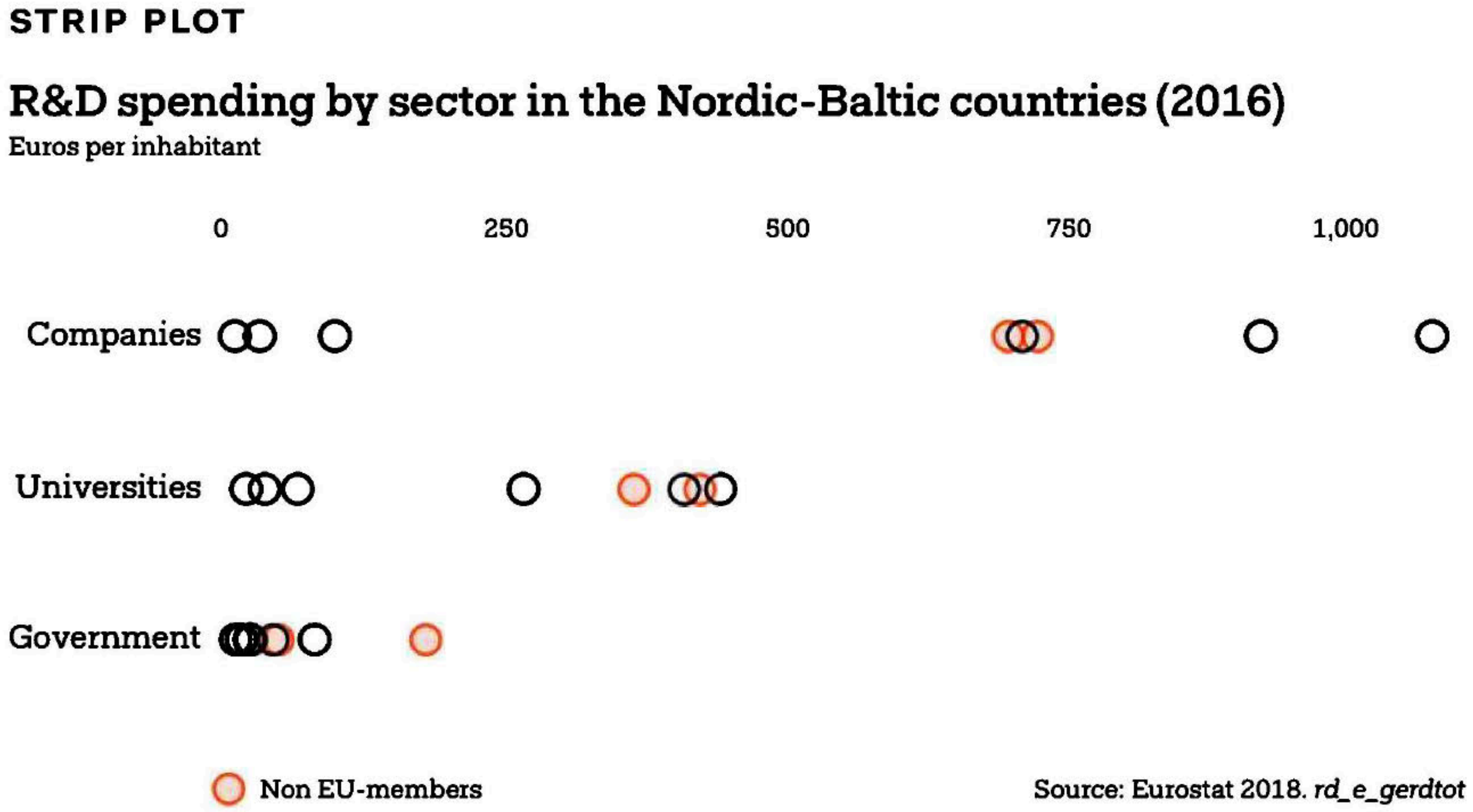


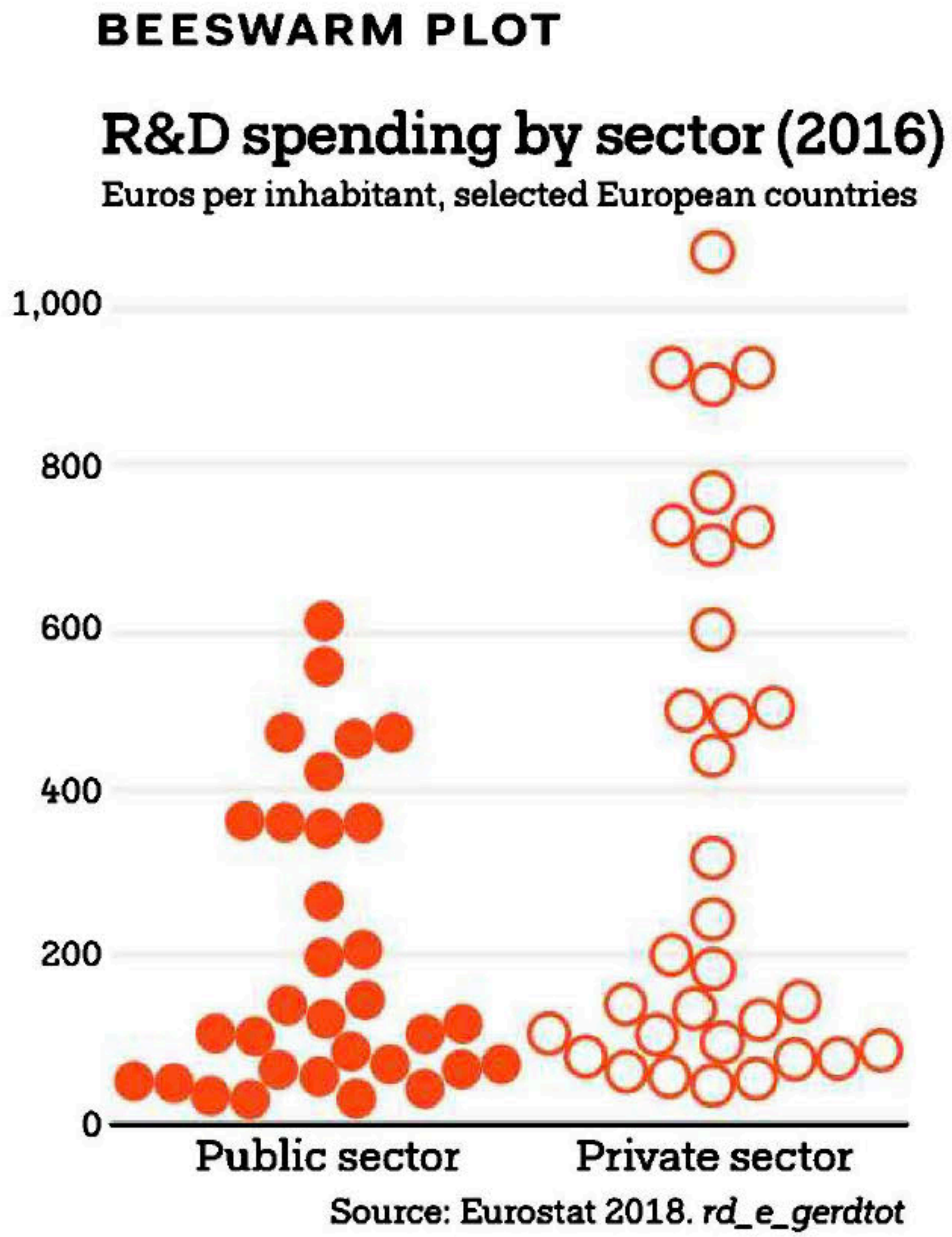
DOT-DASH PLOT



Koponen & Hildén

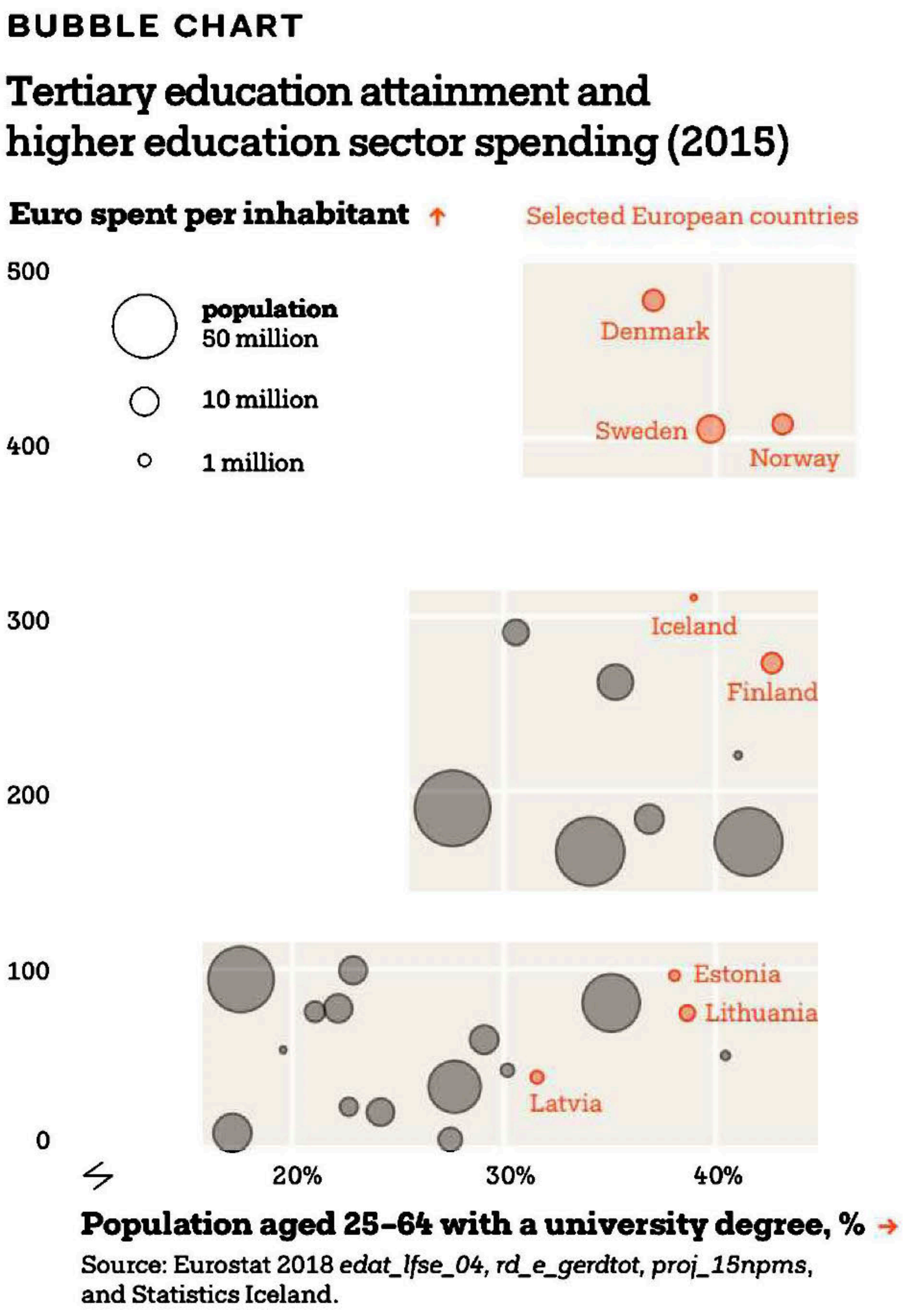
Common graphics, identify the *data-ink encodings*

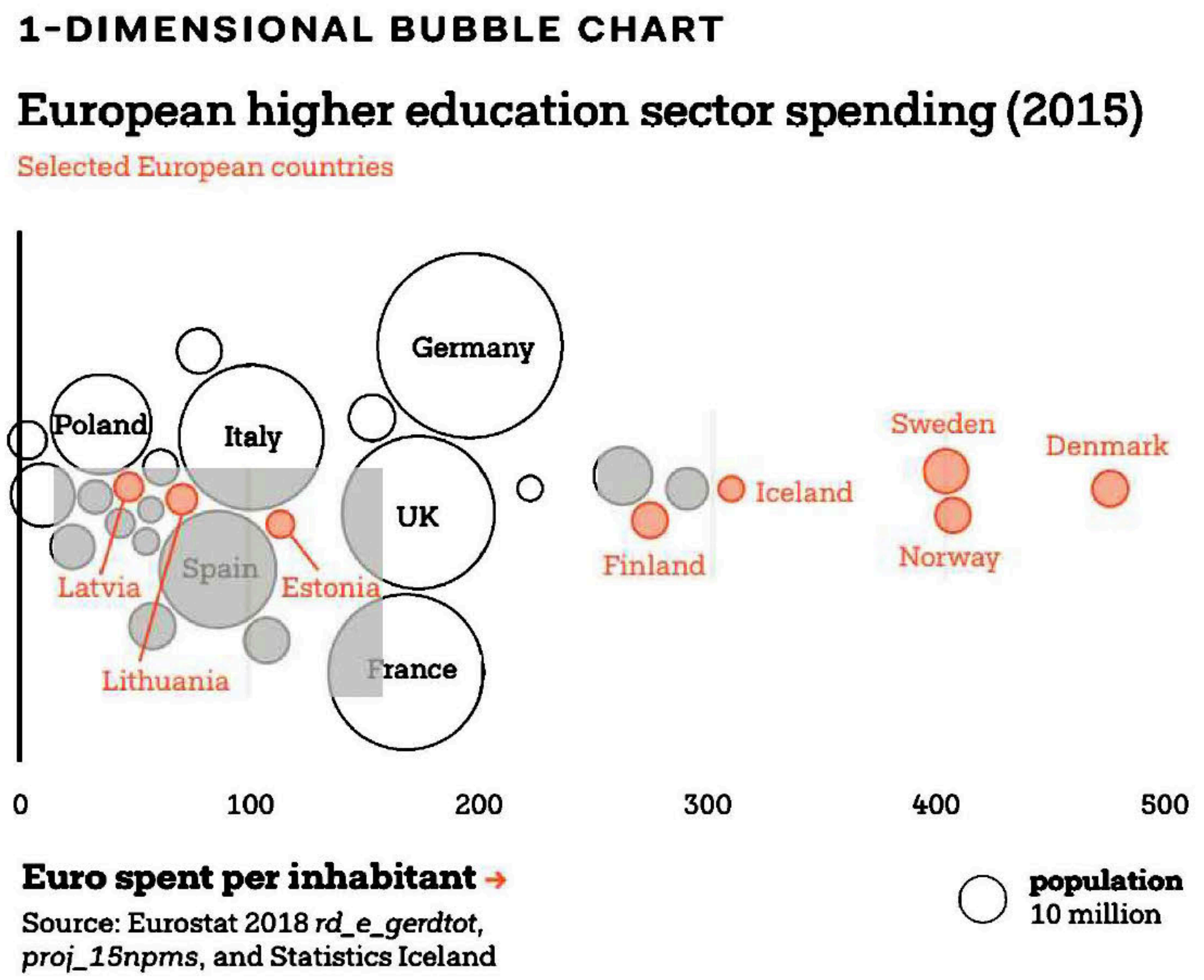




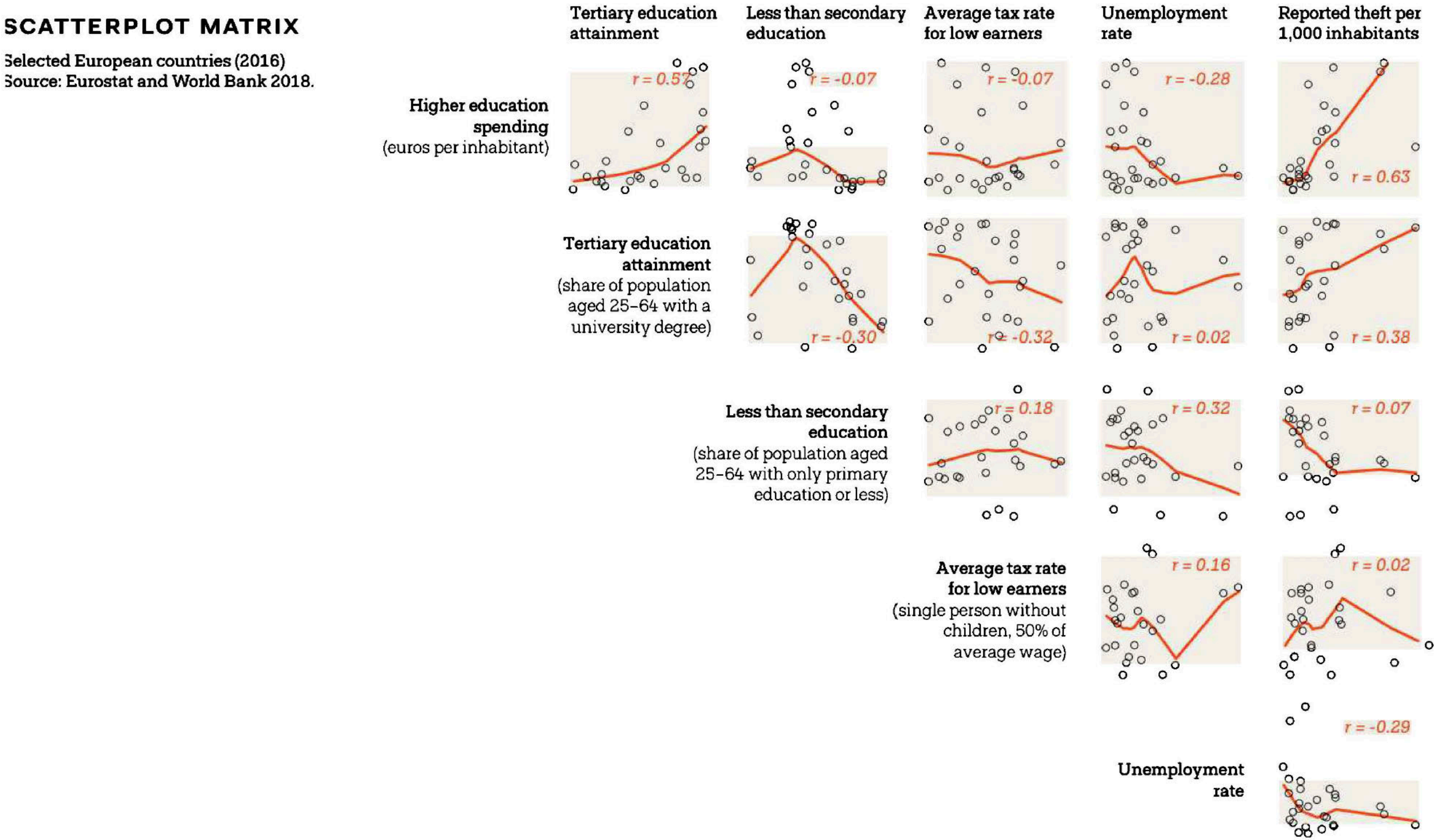
Koponen & Hildén

Common graphics, identify the *data-ink encodings*



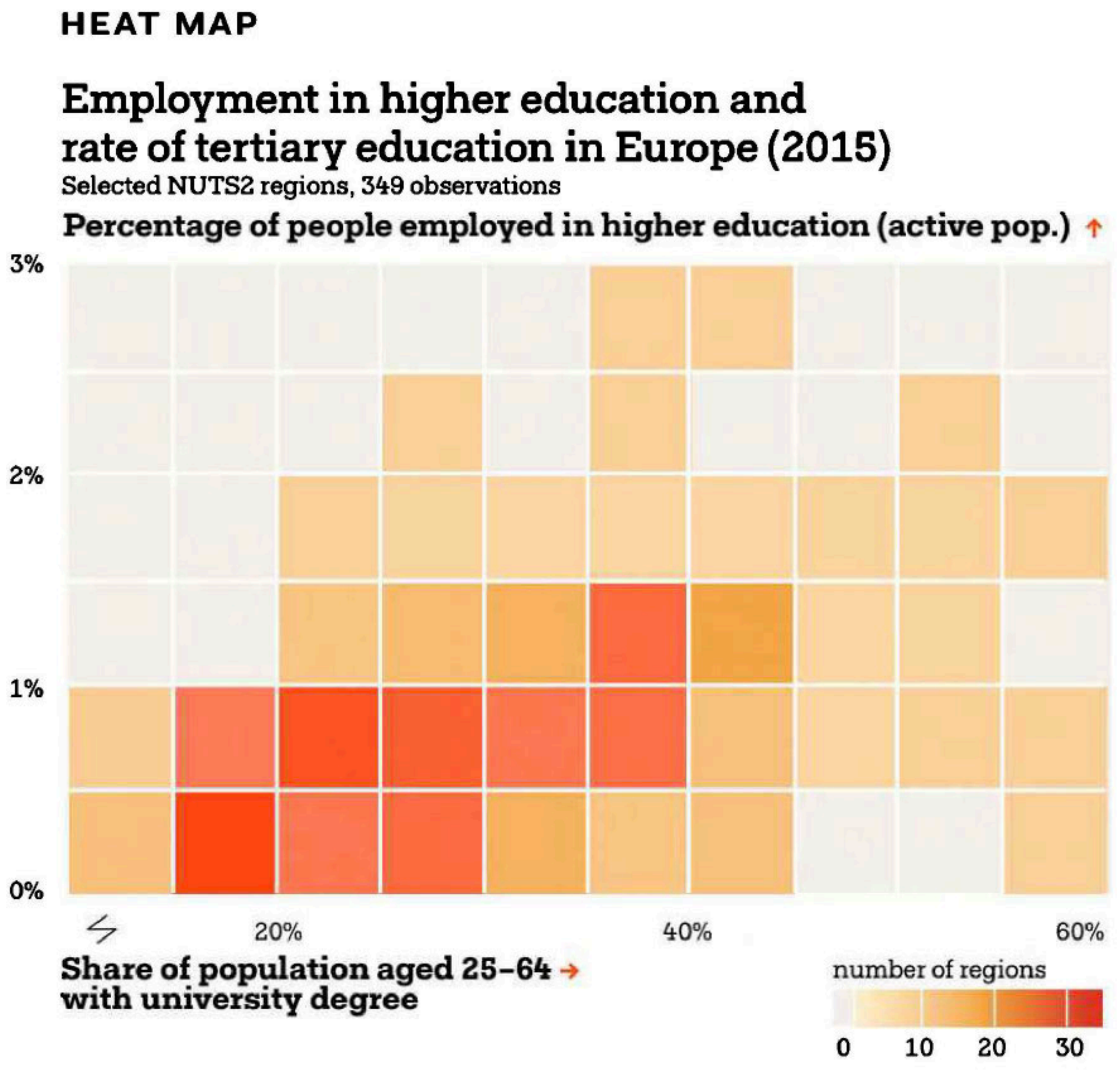


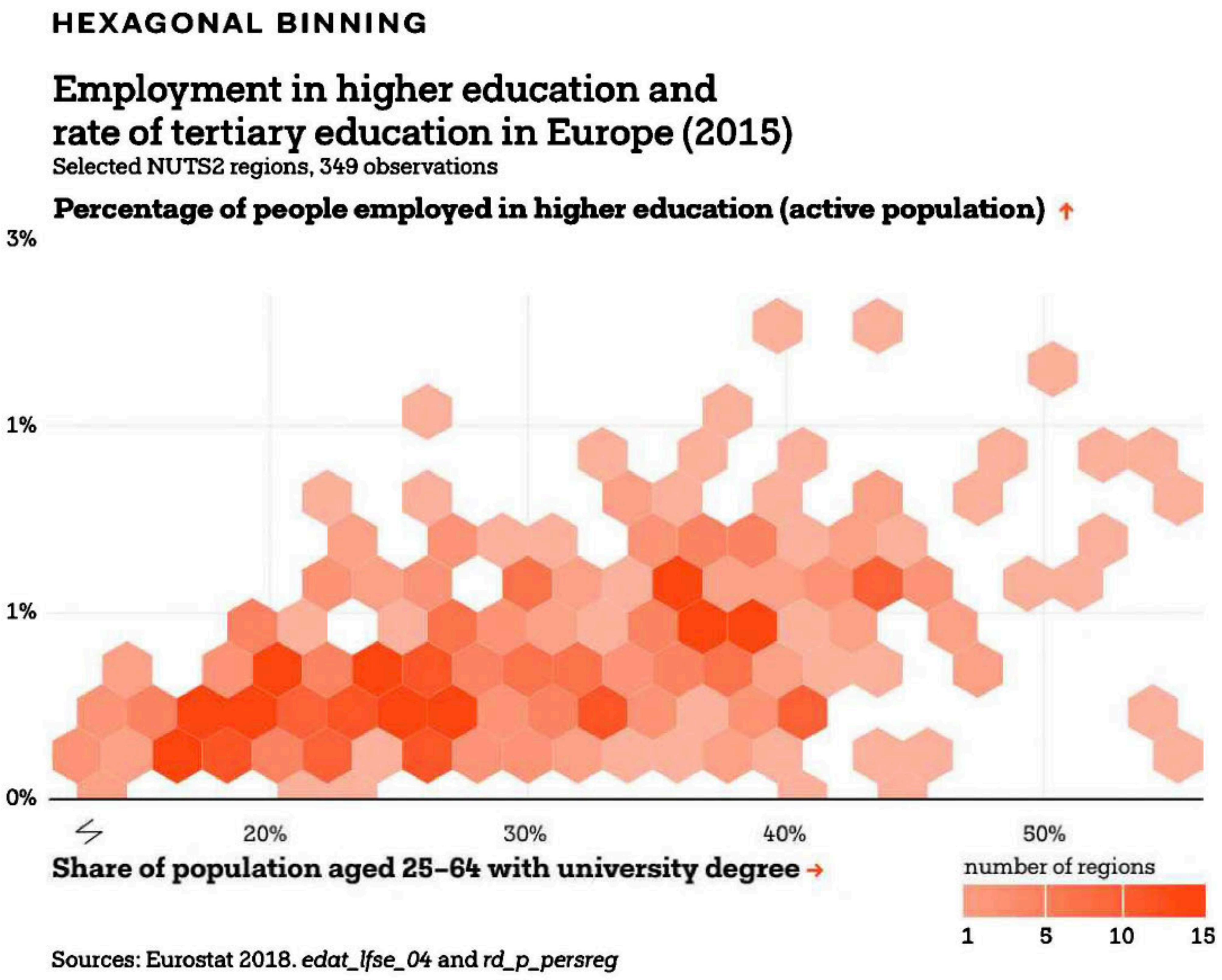
Koponen & Hildén



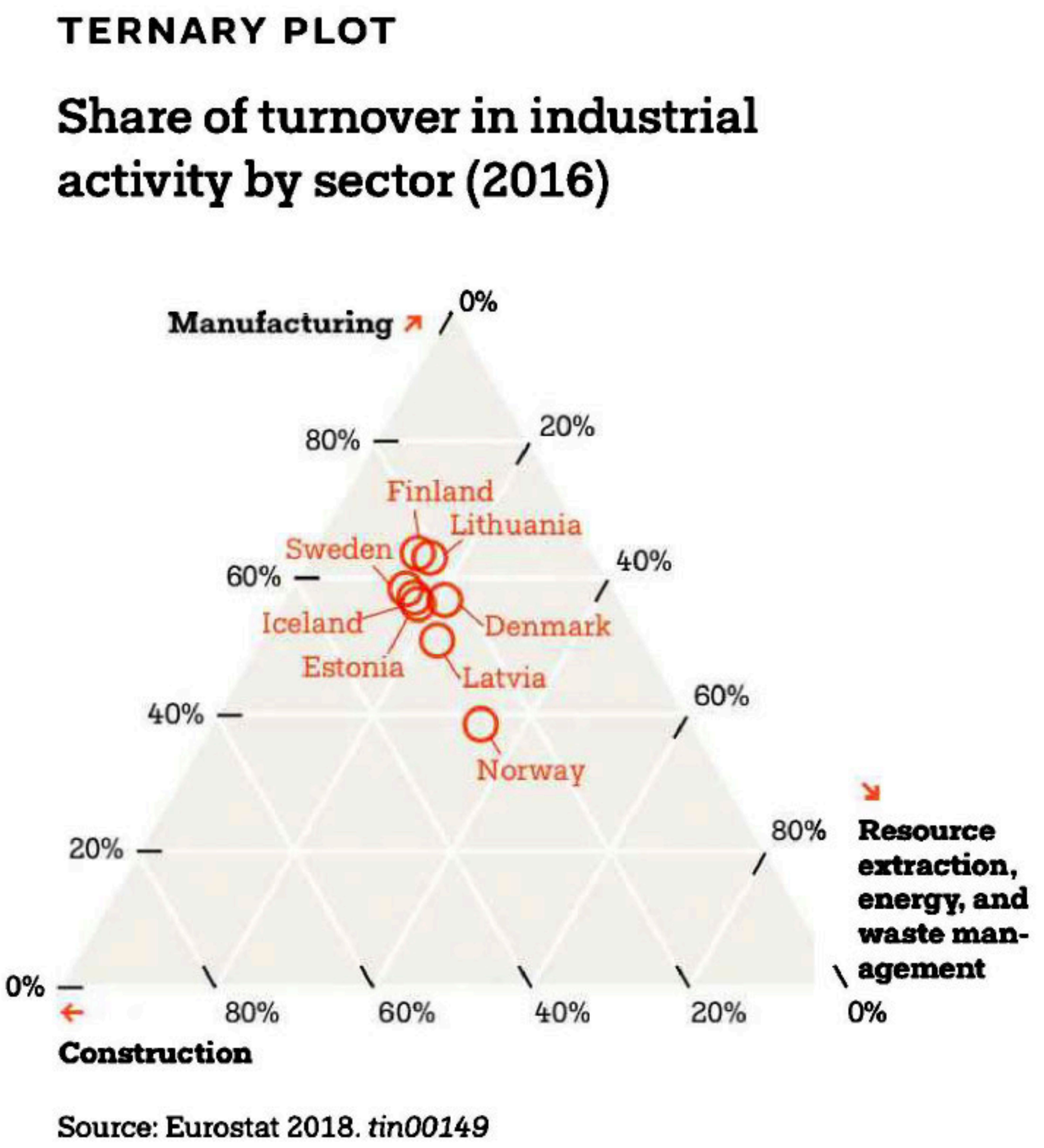
Koponen & Hildén

Common graphics, identify the *data-ink encodings*





Koponen & Hildén



Koponen & Hildén

Studio work — exploratory data analysis

Koponen & Hildén

References

Anscombe, F J. "*Graphs in Statistical Analysis*." The American Statistician 27, no. 1 (February 1973): 17–21.

Bertin, Jacques. *Semiology of Graphics: Diagrams Networks Maps*. Redlands: ESRI Press, 2010.

Holtz, Yan, and Conor Healy. "*From Data to Viz*," 2018. www.data-to-viz.com.

Ing. *The Next Rembrandt*, April 2016. https://www.nextrembrandt.com.

Koponen, Juuso, and Jonatan Hildén. *Data Visualization Handbook*. First. Finland: Aalto Art Books, 2019.

Kosara, Robert, and Jock Mackinlay. "*Storytelling: The Next Step for Visualization*." Computer 46, no. 5 (May 2013): 44–50.

Leborg, Christian. *Visual Grammar*. Princeton Architectural Press, 2004.

Lee, Bongshin, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. "*More Than Telling a Story: Transforming Data into Visually Shared Stories*." IEEE Computer Graphics and Applications 35, no. 5 (September 2015): 84–90.

McCloud, Scott. *Understanding Comics: The Invisible Art*. Kitchen Sink Press, 1993.

Riche, Nathalie Henry, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. *Data-Driven Storytelling*. CRC Press, 2018.

Tufte, Edward R. *Visual Explanations. Images and Quantities, Evidence and Narrative*. Graphics Press, 1997.

Wickham, Hadley. "*A Layered Grammar of Graphics*." Journal of Computational and Graphical Statistics 19, no. 1 (January 2010): 3–28.

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