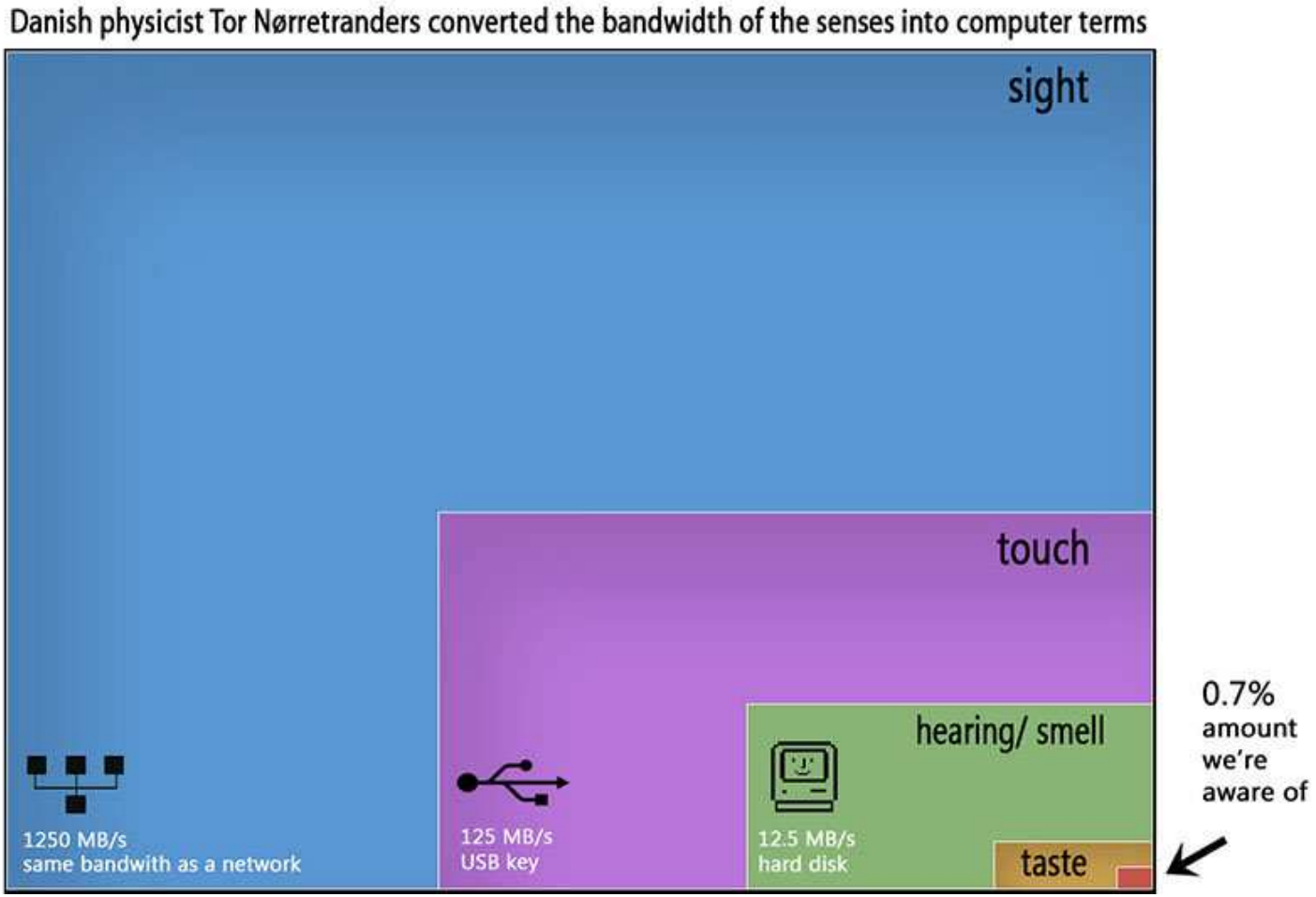
Digitise, Optimise, Visualise: Data Visualization

Peter H. Gruber

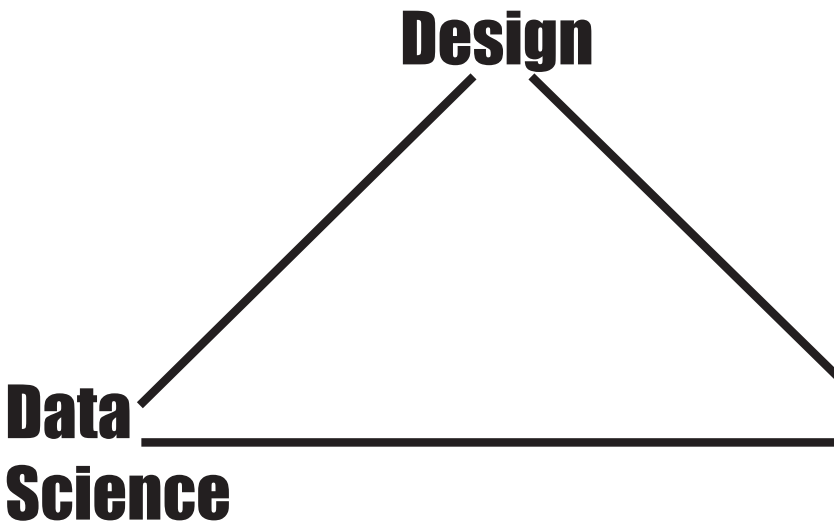
July 1-5, 2019

Why data viz?



- Data Viz is the highway to the brain.
- Brain has incredible graphics processor. → Viz as cognitive aid.

Unique Skillset



Dataviz

What is data viz?

Translation of data into graphic language to enable ...

- Exploratory data analysis
  (i.e. formualte hypothesis, John Tukey, 1977)
- Confirmation
- Presentation
- Story telling

How is it done?

- Mapping of data onto graphic objects

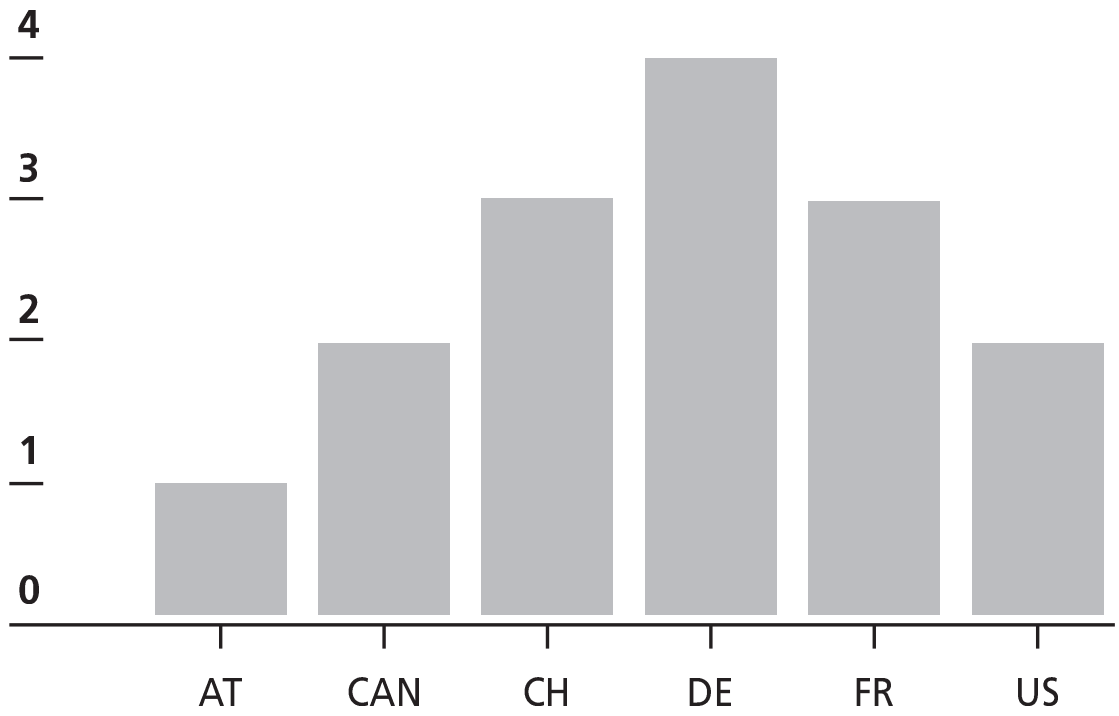
Context is important

- Application, scientific discipline,
- Audience: background, data literatcy
- Culture (e.g. colors)

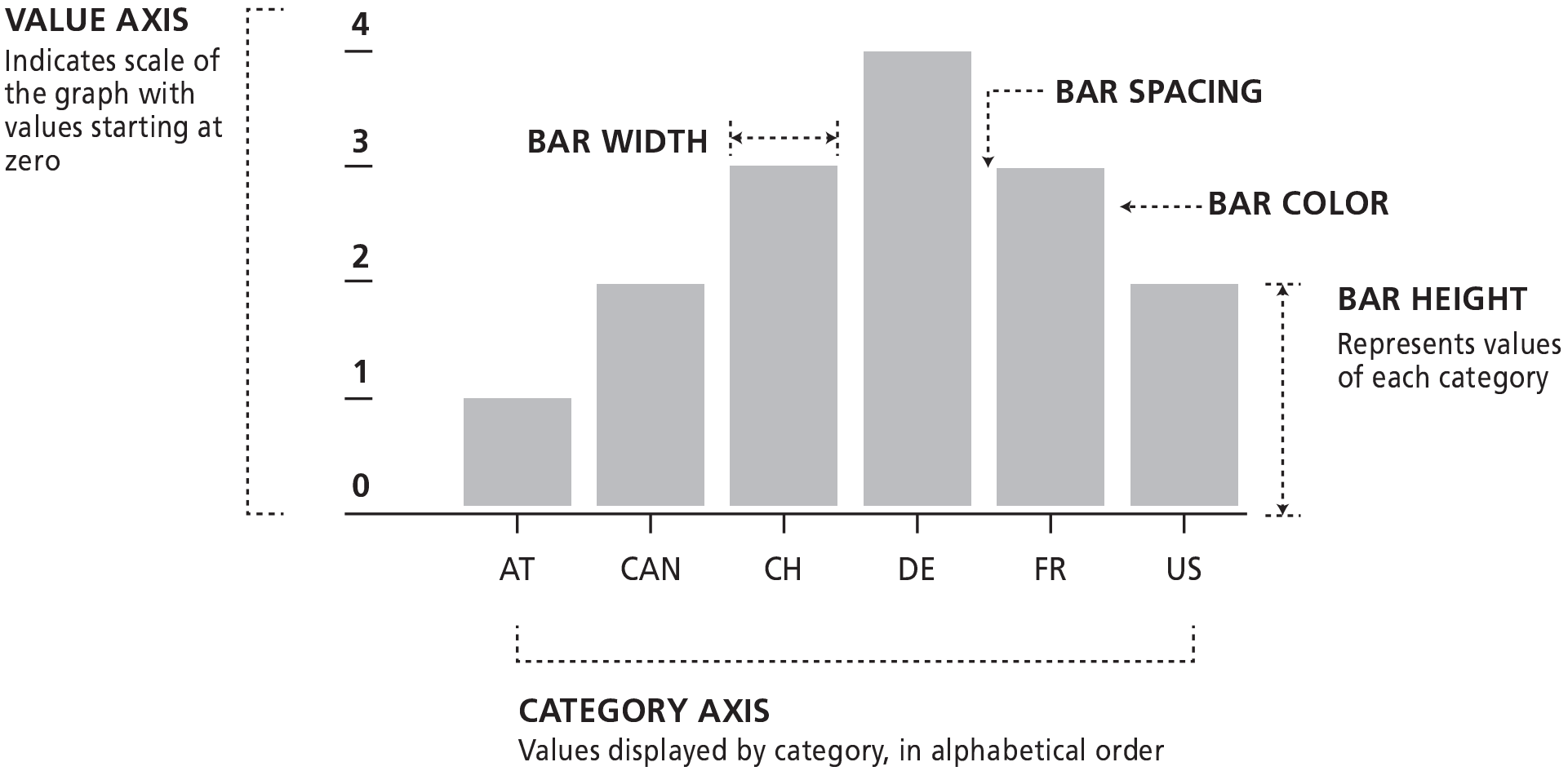
Distinctions

- **Infographics**
  Explain a mechanism
- **Information vizualization**
  Translate facts into images
- **Data visualization** ⟵ *this is what we do*
  Translate data into graphics
  Goals: exploration – confirmation – presentation
- **Data art**
  Translate data into graphics with some artistic freedom
  Goal: beauty
- **Data decoration**
  Add graphical elements to data that have no information content

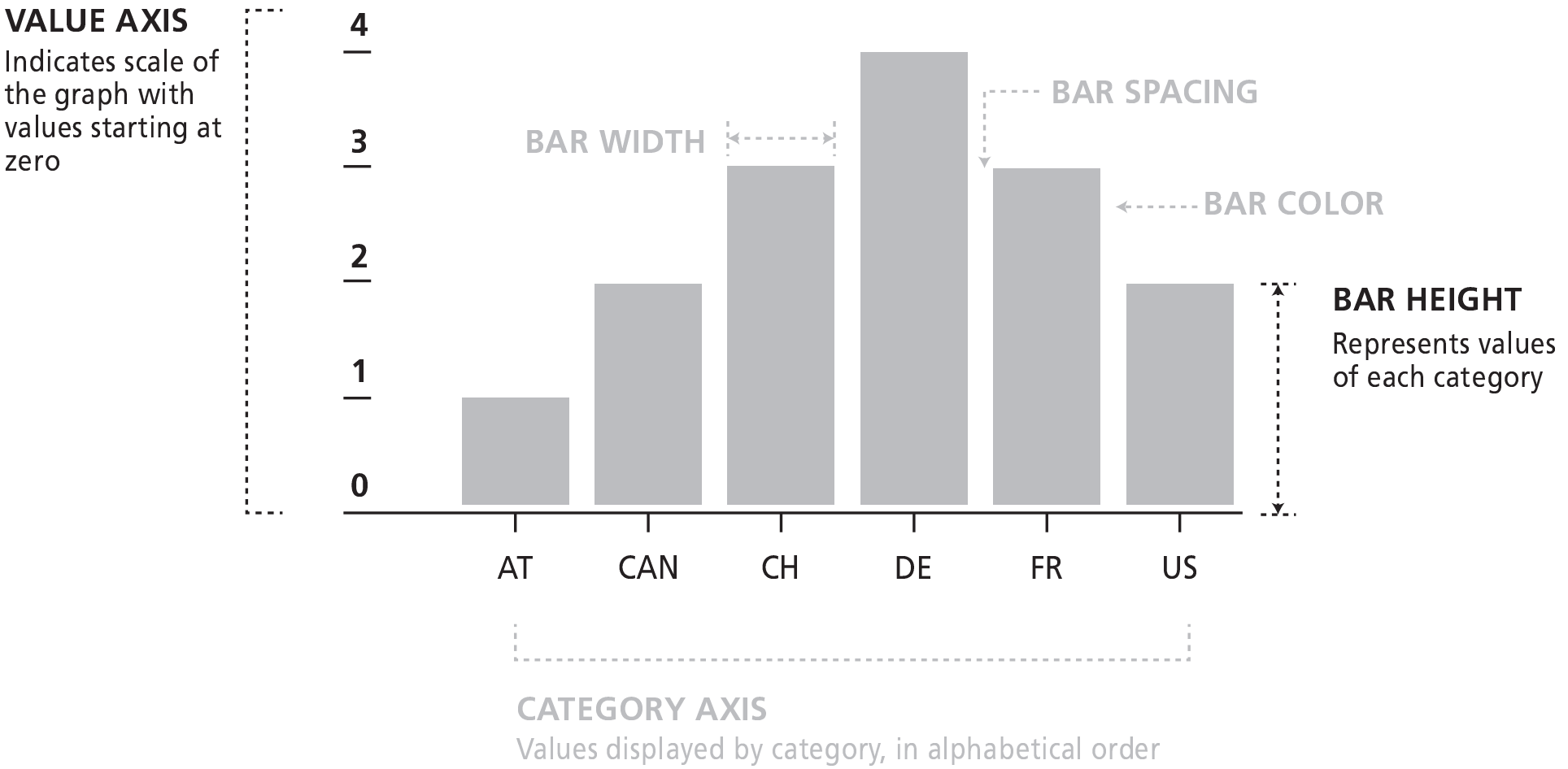
Example



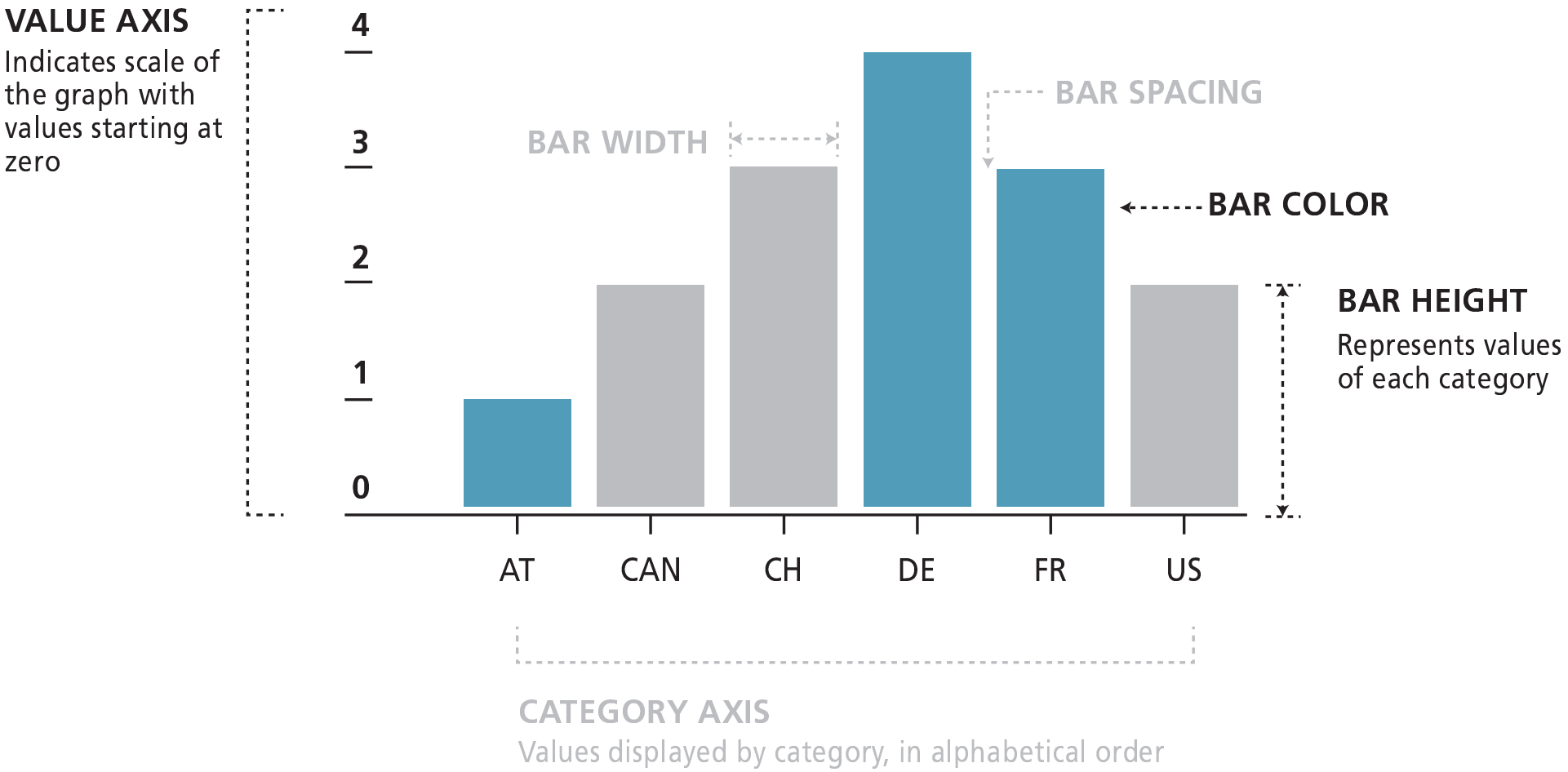
Example with annotations



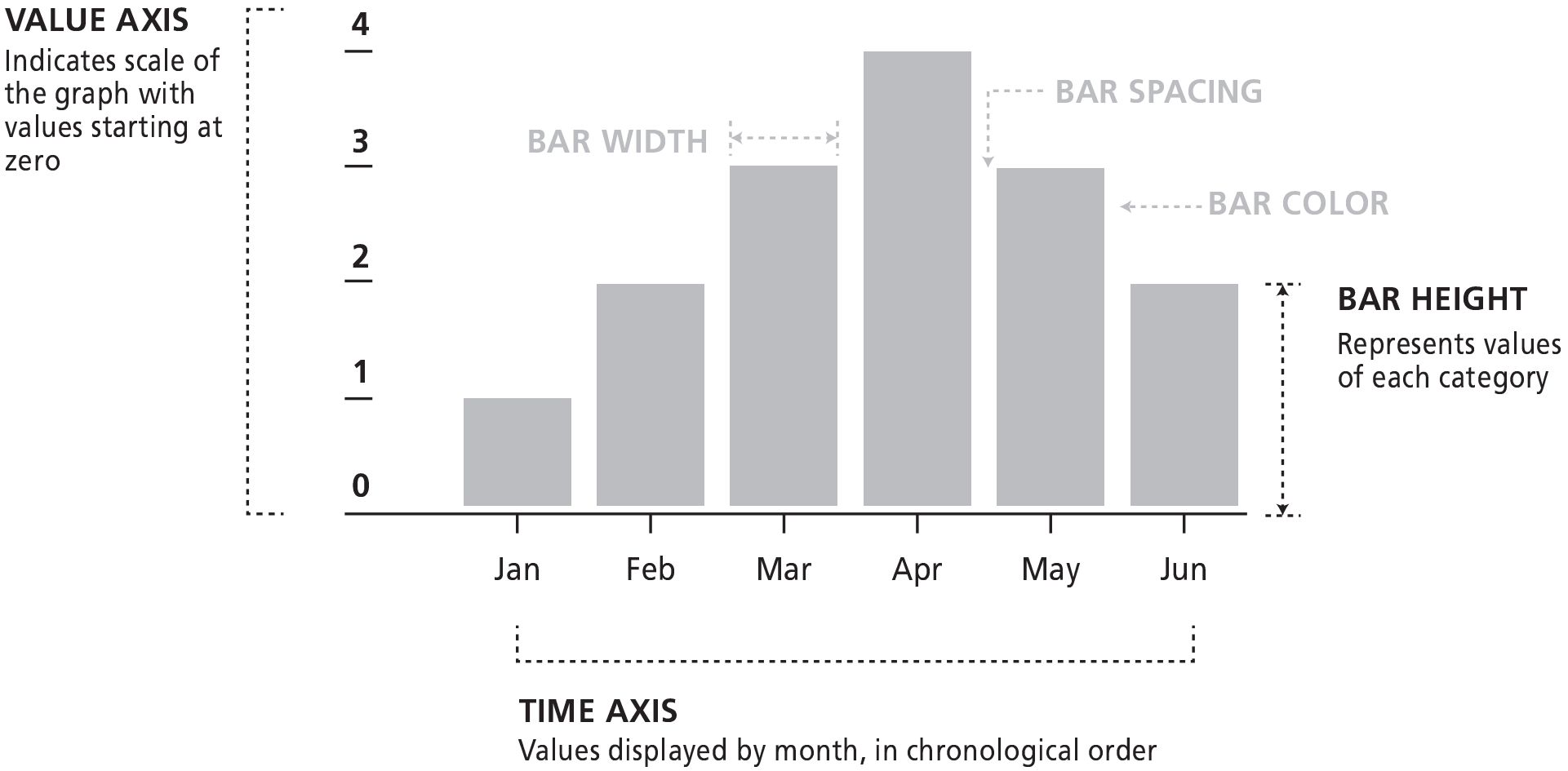
Information content (1)



Information content (2)



Information content (3)



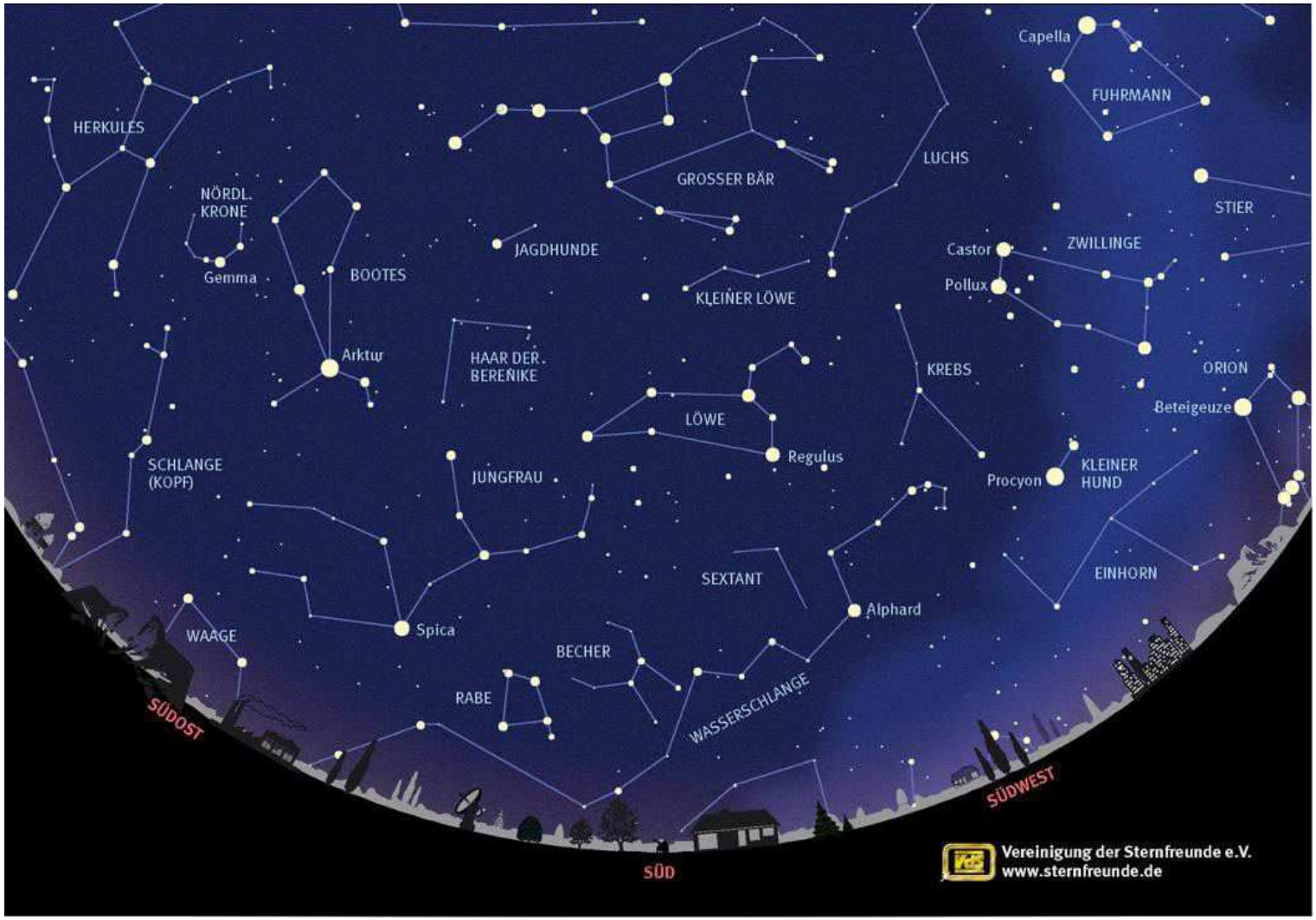
Dataviz = science?

- *Def:* A statistic is a calculable function of the data
  $\rightarrow$ every visualization is a statistic.
- Data visualizations are falisifiable.
- Scientific foundation in cognitive science + statistics.
- Good viz. leads to statistically superior inference.
- Design = finding the best possible (most efficient) solution for a given problem.

Perception

The night sky ...

... and how we see it



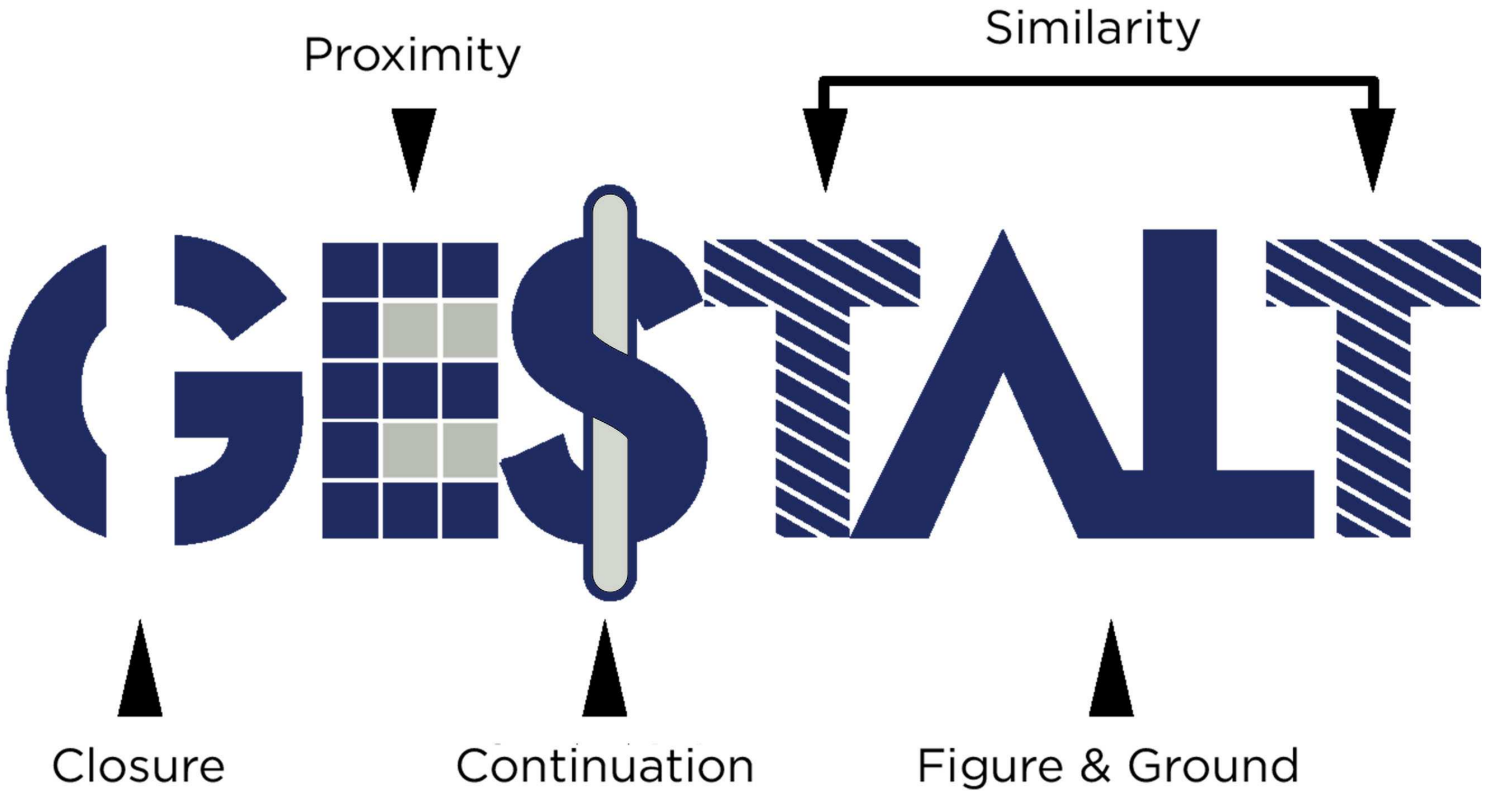
Gestalt Principles

The brain *wants* to find sense in our environment

- Christian von Ehrenfels (1890)
- Law of Prägnanz:
  "We order our experience in a manner that is . . .
  - regular,
  - orderly,
  - symmetrical,
  - and simple."
- Pre-attentive processing

→ Exercise: seeing plots

Gestalt Principles



More: symmetry, common fate, past experience, common region, connectedness

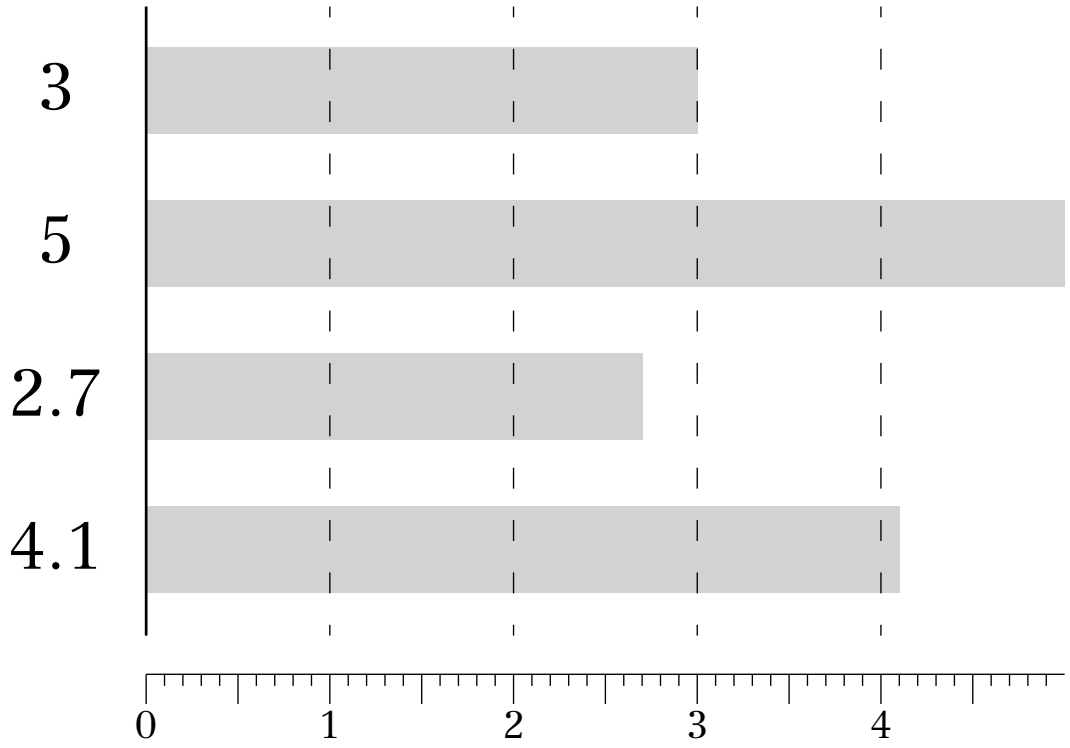
→ Exercise: how can we make use of the gestalt principles?

The translation problem

Q: How to translate numbers into graphics?

- Jacques Bertin (cartographer, 1967), Semiology of Graphics
  List of "les variables de l'image"
- Leyland Wilkinson (2005), Grammar of graphics
  Aesthetics = mappings of the data

The translation process: simple case



Data $\longrightarrow$ Length

The translation process: advanced case

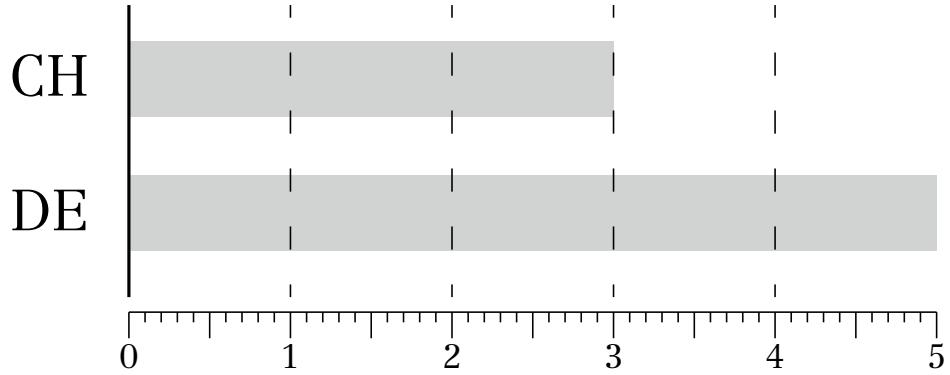
Data — DE, DE, CH, CH, DE, CH, DE, DE

↓

Statistics — 3xCH, 5xDE (count)

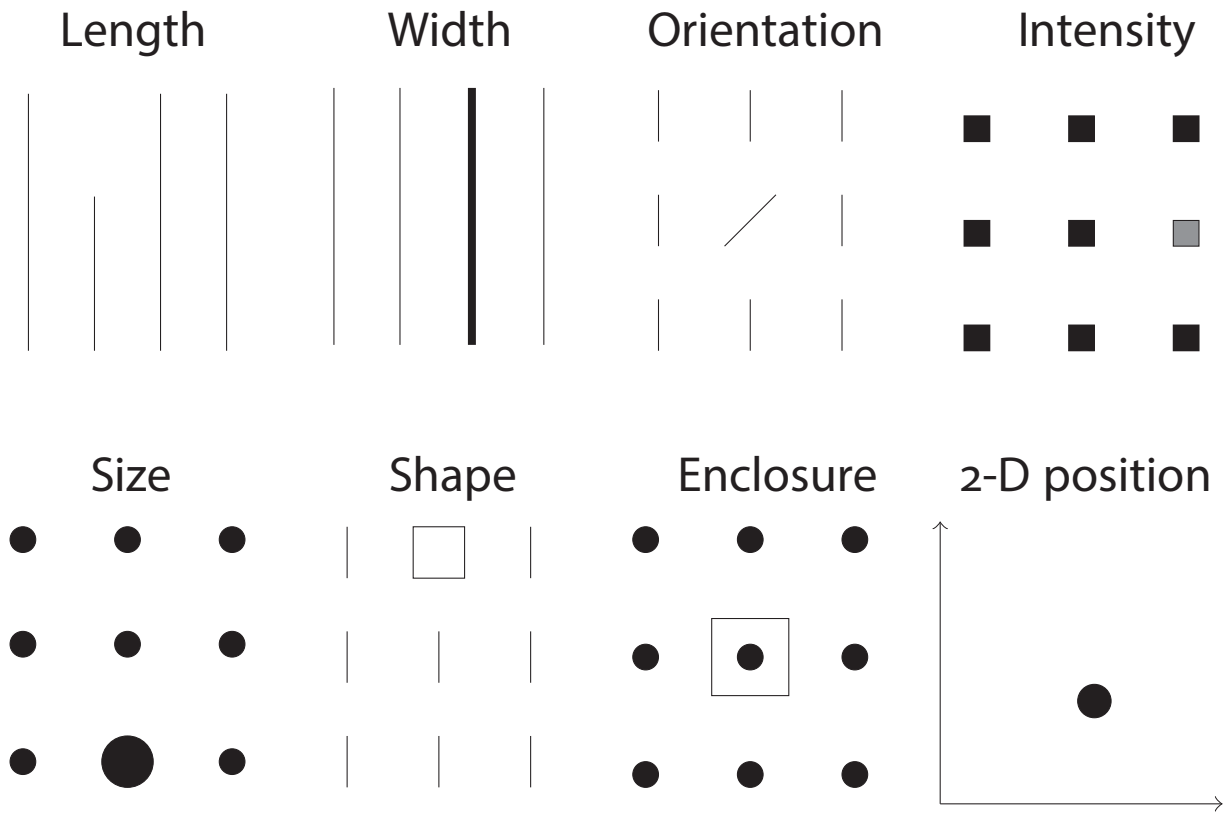
↓

Aesthetics — length



Data $\longrightarrow$ Statistics $\longrightarrow$ Length

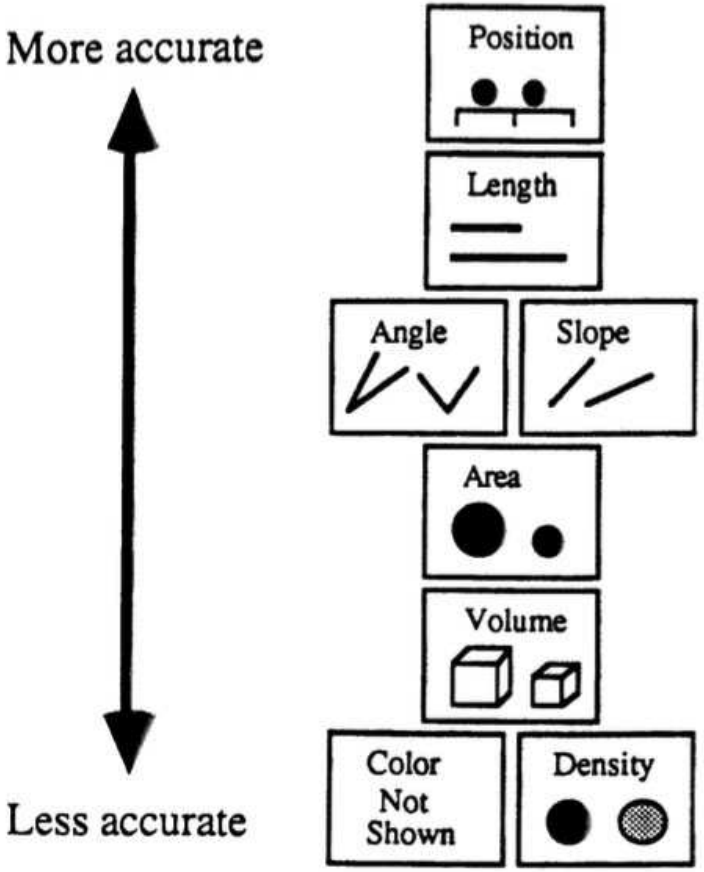
Aesthetics



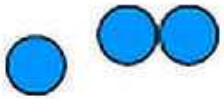
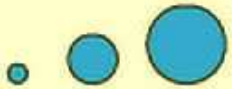
Visual Encodings

Which *encodings* work best?

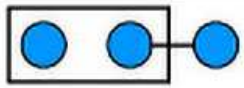
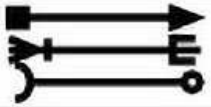
- Jock D. Mackinlay (1986): Automating the design of graphical presentations of relational information



Visual Encodings (2a)

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal
	position, placement	yes	infinite	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good
	length	yes	many	Good	Good
	size, area	yes	many	Good	Good
	angle	yes	medium	Good	Good
	pattern density	yes	few	Good	Good
	weight, boldness	yes	few		Good
	saturation, brightness	yes	few		Good

Visual Encodings (2b)

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical
	color	no	few (<20)			Good
	shape, icon	no	medium			Good
	pattern texture	no	medium			Good
	enclosure, connection	no	infinite			Good
	line pattern	no	few			
	line endings	no	few			
	line weight	yes	few		Good	

From: Julie Steele, Noah Ilinsky: Designing Data Visualizations, Safari books

Color

Humans and color

Perception

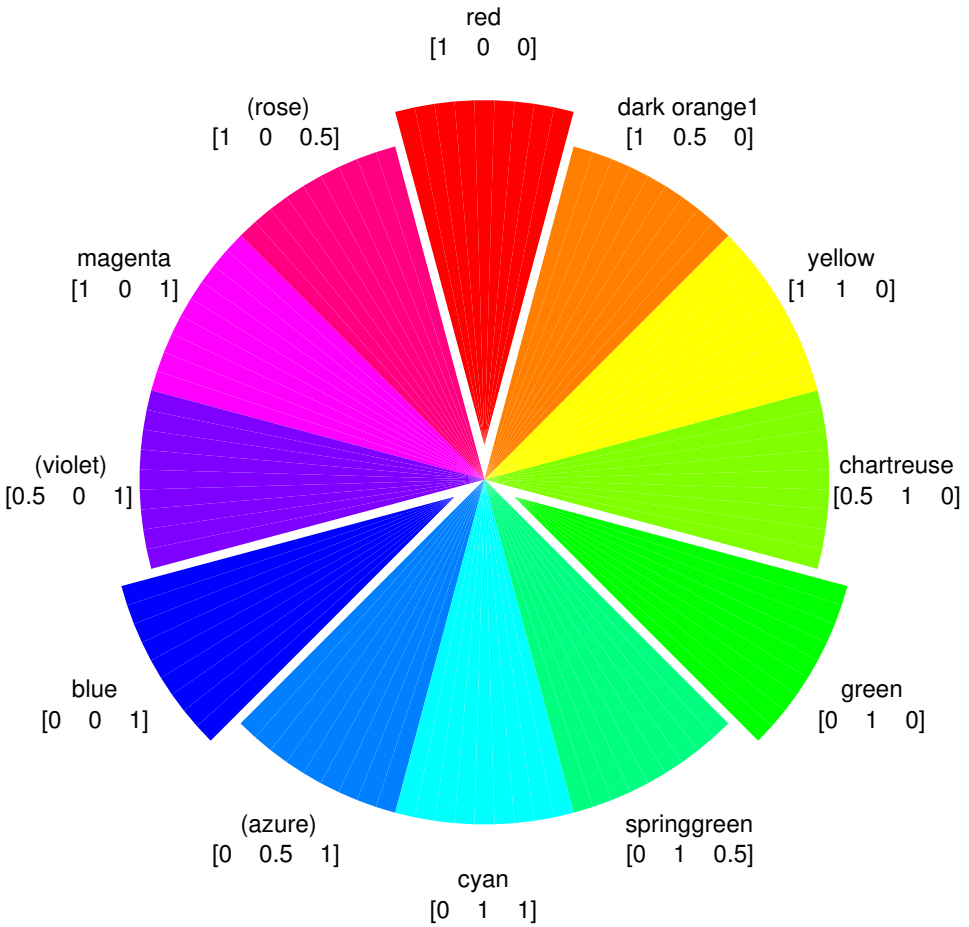
- Human retina has 100-150 mio light-sensitive cells
  Only 7 mio are color-sensitive
- Approx. 8% of men and 0.5% of women are color-blind
  Online test: `enchroma.com/test/instructions`
- Continuous: Can distinguish $>$ 1mio colors *side-by-side*
- Discrete: Can distinguish, recognize and name 100s of colors

Communication

- **Names:** "blue", "sky blue", "steel blue", "navy blue"
- **Formula:** how much **R**ed, **G**reed and **B**lue light is in a color?
- RGB is color mixing by light: (0,0,0) = black, (1,1,1)= white

RGB Color Wheel

- Most computers: 256 shades of R/G/B each
  $\rightarrow 256^3 \approx 16$ mio colors



Alternatives to RGB

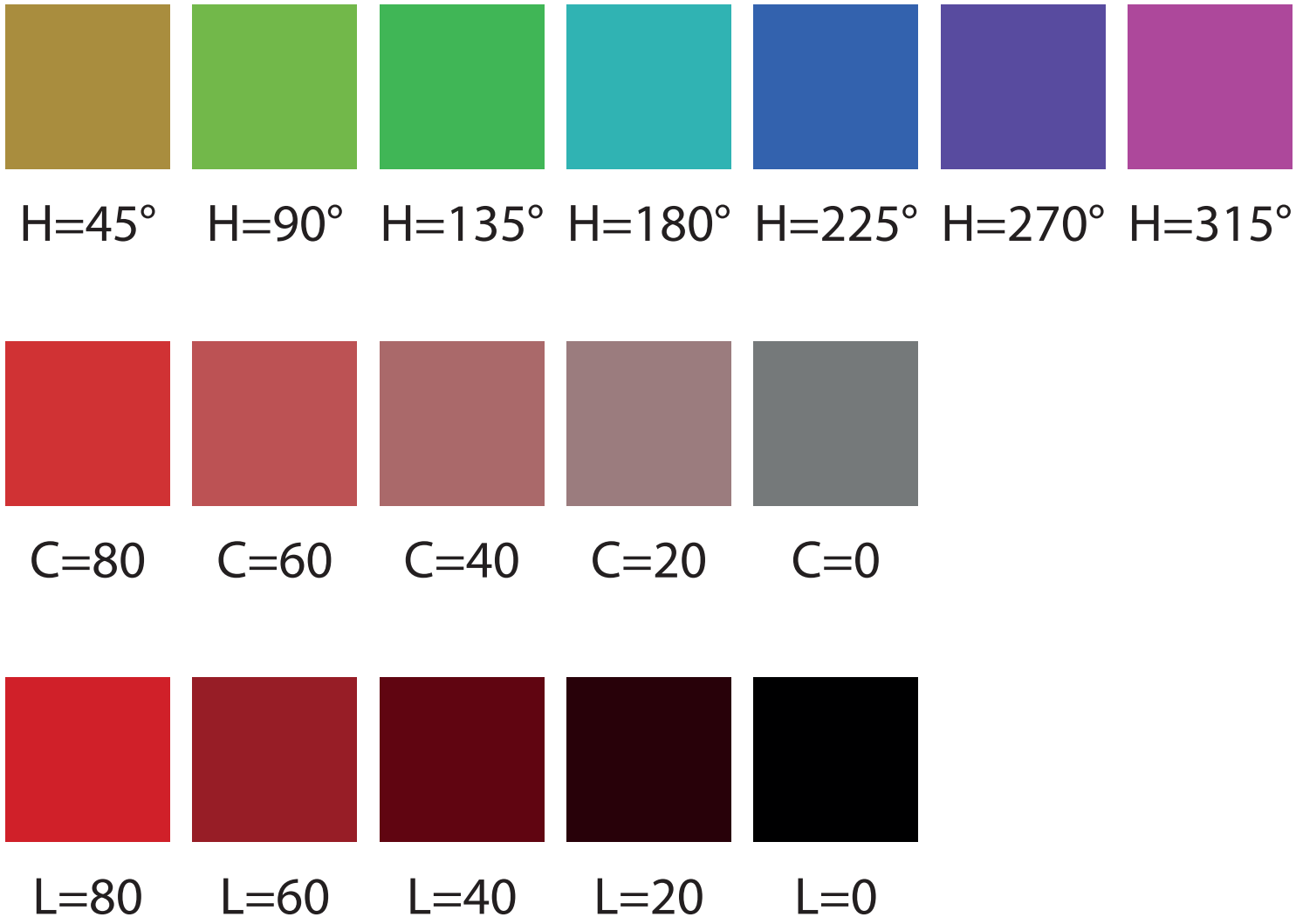
Is the RGB system the best?

- Technically motivated
- Color Space of the eye is
  (1) black-white, (2) blue-yellow, (3) red-green
- Luminance (1): $L = 0.31R + 0.59G + 0.1B$
  $\rightarrow$ RGB coloes with equal color sums have different brightness

HCL Model

- Perception-based color space **HCL**
  - **H**ue [0-360] = angle in the color wheel
  - **C**roma [0-100] = color intensity between grey and intense
  - **L**uminance [0-100] = *perceived* brightness

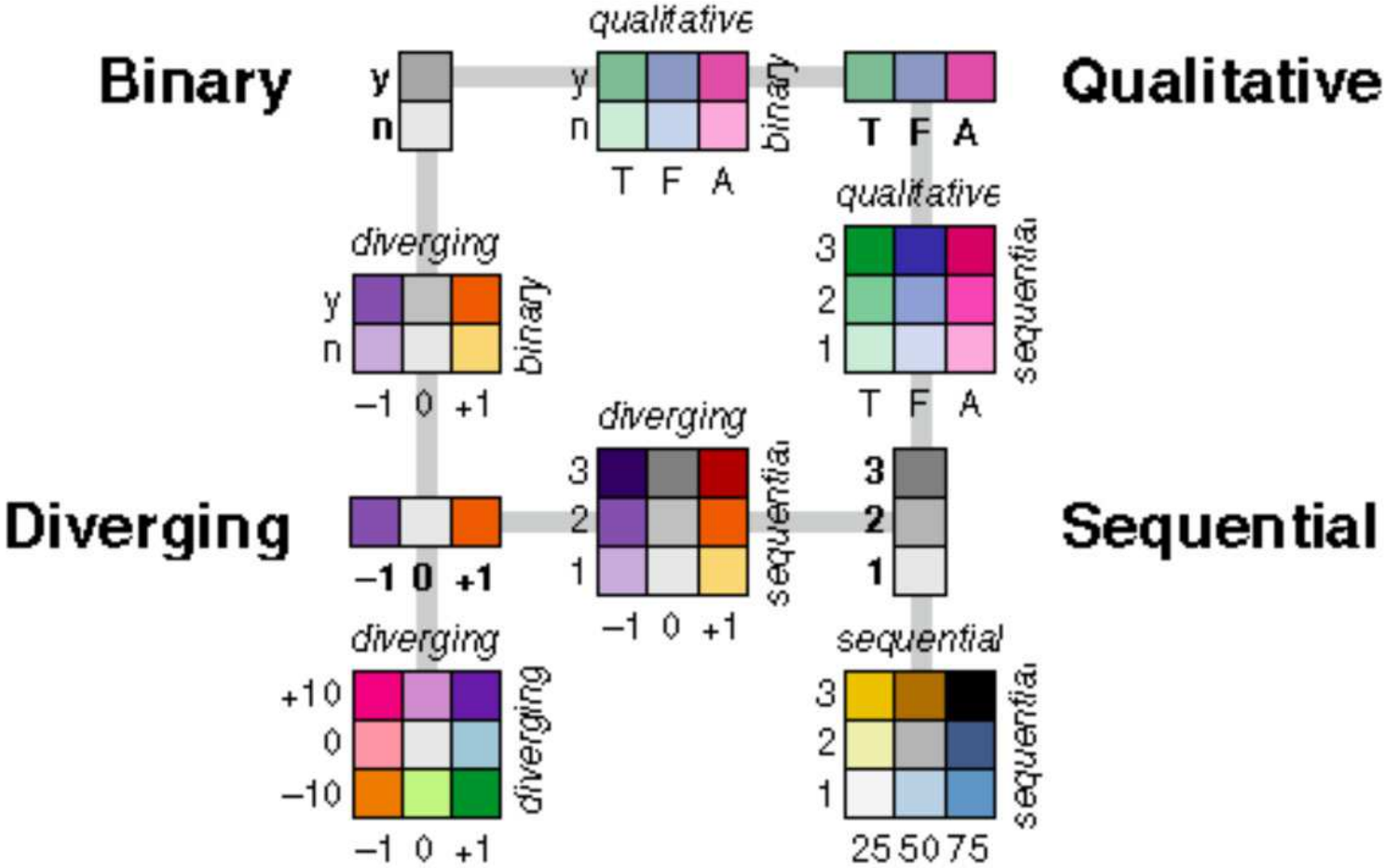
HCL Model



How to choose a color scheme for visualization?

- Default colors not a good idea
- Type of variable → color scheme
  - Binary – contrast
  - Qualitative – distinction
  - Diverging – continuous scale with neutral center
  - Sequential – continuous scale
- Possible to combine 2 schemes
- Cultural connotations
  - Hot - cold
  - Countries, companies, continents
- Color blindness, printing process, limitations of beamers
- Predefined schemes: `colorbrewer.py`
  See `www.colorbrewer2.org`

Color Brewer Combinations



See `colorbrewer2.org`

Source `www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html`

The graphical method has considerable superiority for the exposition of statistical facts over the tabular.

A heavy bank of figures is grievously wearisome to the eye, and the popular mind is as incapable of drawing any useful lessons from it as of extracting sunbeams from cucumbers.

Arthur & Henry Farquhar
in Economic and Industrial Delusions (1891)