University of Zurich[UZH]

URPP Evolution in Action

URPP tutorial

Principles of data visualization

Dr. Heidi E.L. Tschanz-Lischer
University of Zurich
Switzerland

23 June, 2017

Outline

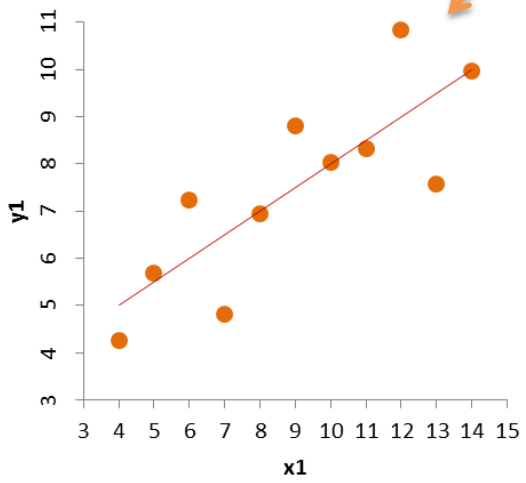
1. The properties of the data or information (HTL)
2. Use of salience, colors, consistency and layout (HTL)
3. The rules mapping data to images (SW)
4. Examples of effective visualizations in biological sciences (SW)
5. Presentation and discussion of “good” and “bad” graphics (HTL & SW)

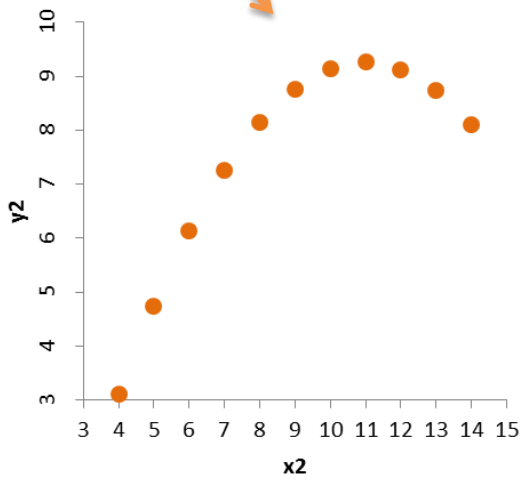
Information visualization

2 main objectives:

- Data analysis
  - Understand the data
    → "a picture is worth a thousand numbers"
  - Derive information from them
  - The goal is to generate hypotheses that can be tested
- Communication
  - Of information
  - Involves simplification
    → don't distort what the data has to say
    → What information does the reader need to be successful?

x1	y1	x2	y2
10	8.04	10	9.14
8	6.95	8	8.14
13	7.58	13	8.74
9	8.81	9	8.77
11	8.33	11	9.26
14	9.96	14	8.1
6	7.24	6	6.13
4	4.26	4	3.1
12	10.84	12	9.13
7	4.82	7	7.26
5	5.68	5	4.74

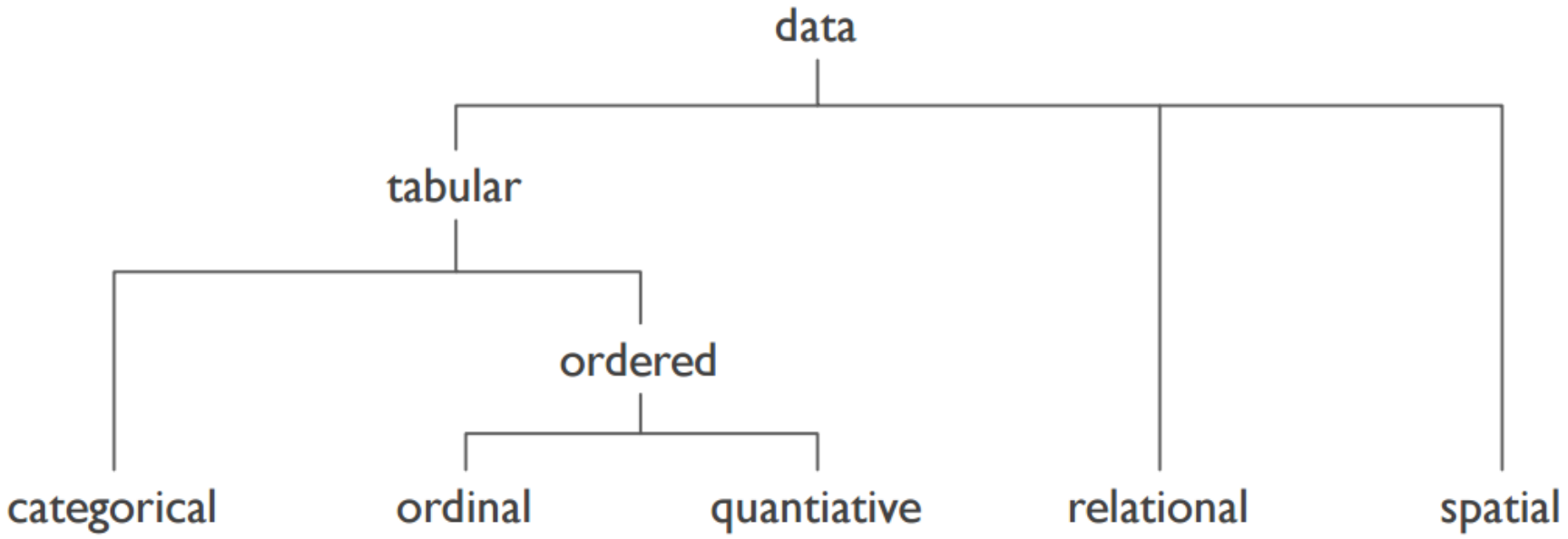




Information visualization

- Common data visualization issues:
  - Inappropriate display choices that distort reality (i.e. pie charts, 3D charts)
  - Too much information
  - Poorly designed display choices that use noisy fill patterns, line styles or saturated/bright colours
  - Encoding quantitative data inaccurately
  - Inconsistent ordering and placement
  - Inconsistent or reversed scales
  - Proportional axis scaling
  - Using counts (not percentages) when comparing different totals

Types of data



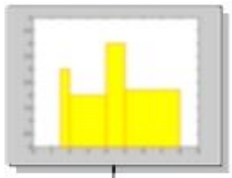
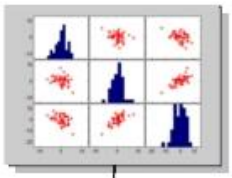
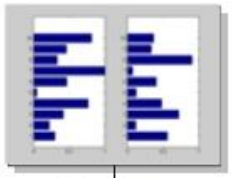
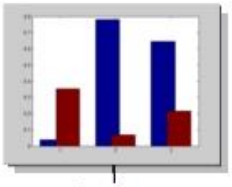
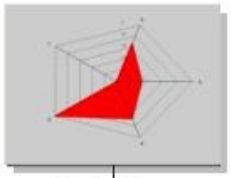
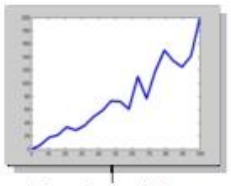
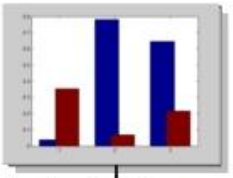
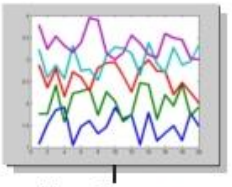
- Sex: male, female
- Genotype: AA, AT, AG, …

- Days: Mon, Tue, …
- Abundance: common, rare, …

- Length: 7, 15, 23 mm, …
- Expression level

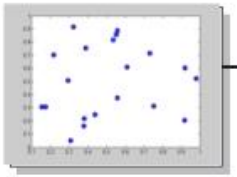
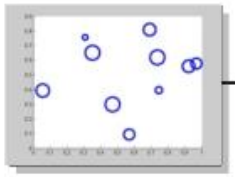
- Trees
- Networks

- Maps

Two Variables per Item
Many Categories
Many Items
Few Items
Few Categories
One Variable per Item
Among Items
Cyclical Data
Non-Cyclical Data
Many Periods
Single or Few Categories
Many Categories
Few Periods
Over Time

Comparison

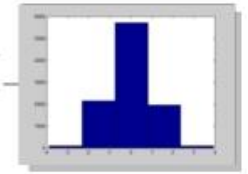
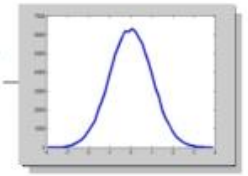
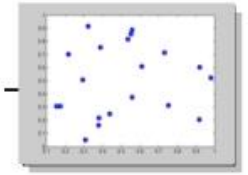
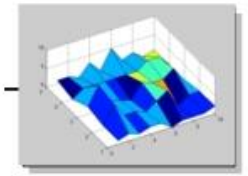
Two Variables
Three Variables

Relationship

What would you like to show?

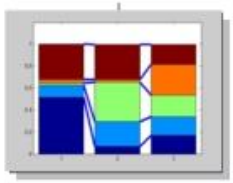
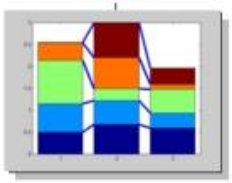
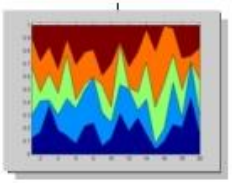
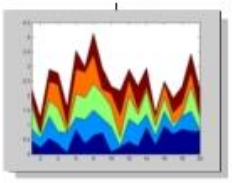
Distribution

Single Variable
Few Data Points
Many Data Points
Two Variables
Three Variables

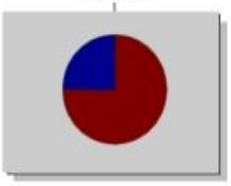
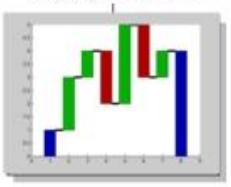
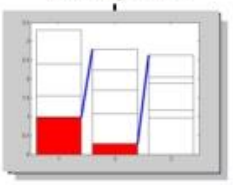





Composition

Changing Over Time
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter

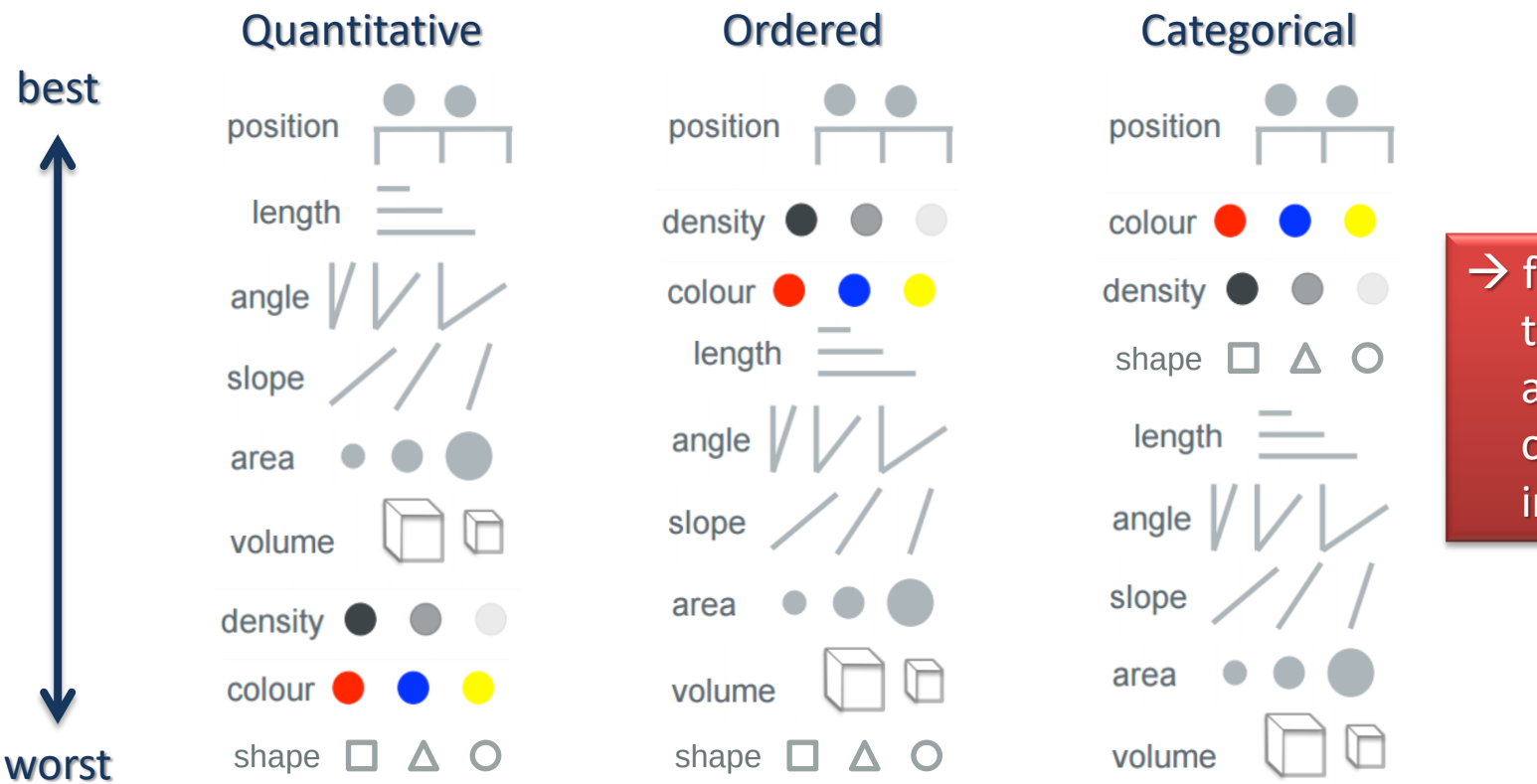





Static
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components

Language of graphics

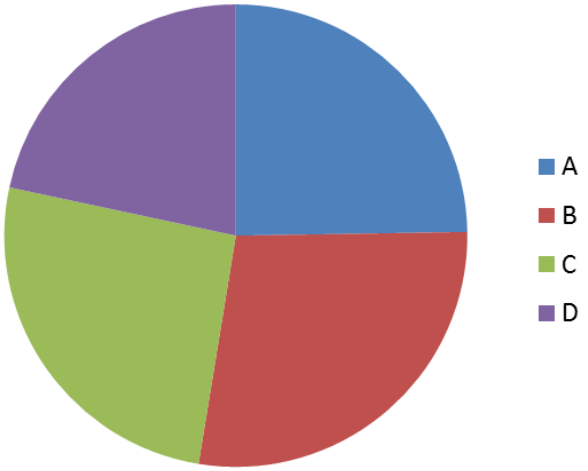
- Graphics can be thought of as forming a sign system:
  - Each mark (point, line, or area) represents a data element
  - Choose visual variables to encode relationships between data elements



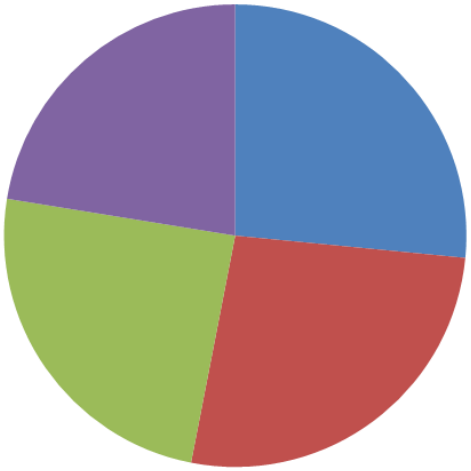
→ find images that express and effectively convey the information

Ranking example

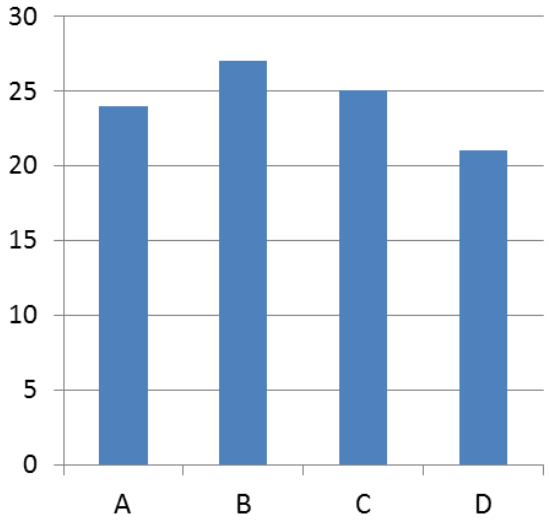
Year 1

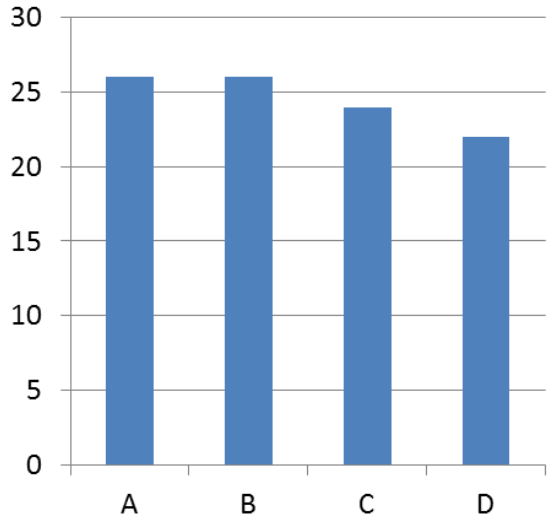


Year 2



→ Difficult to see differences





→ Much better!

Sorting

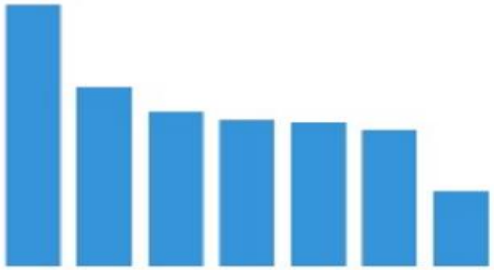
Alphabetical

Helps people to find what they are looking for

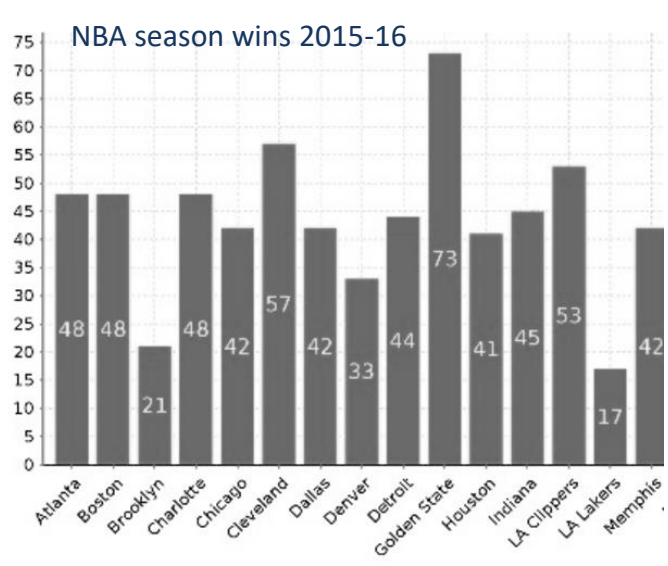
Ascending

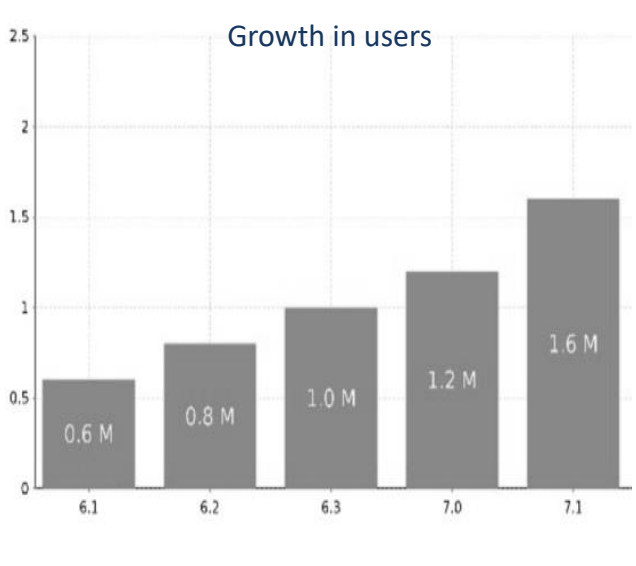
Tells a story of growth

Descending



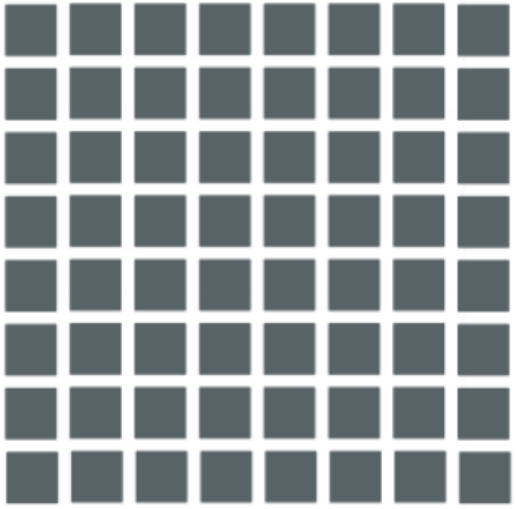
Helps compare largest to smallest

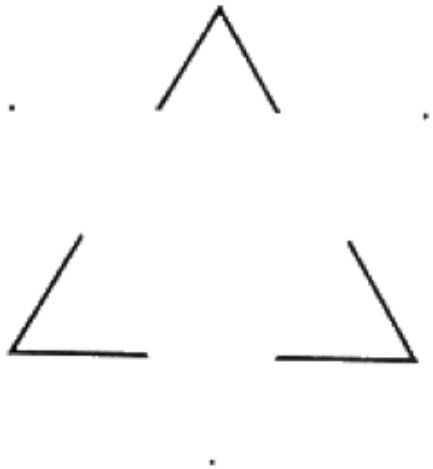




Gestalt principles

- Gestalt theory: principles of pattern recognition
  - "The whole is greater than the sum of its parts"
  - The mind visually perceives objects in a certain way collectively
  - Tries to structure/organize what we see into patterns

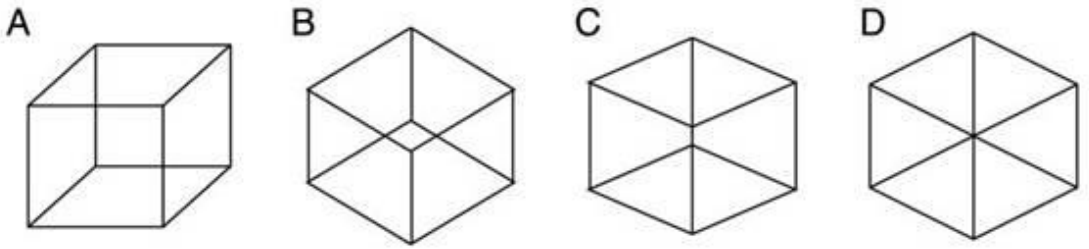




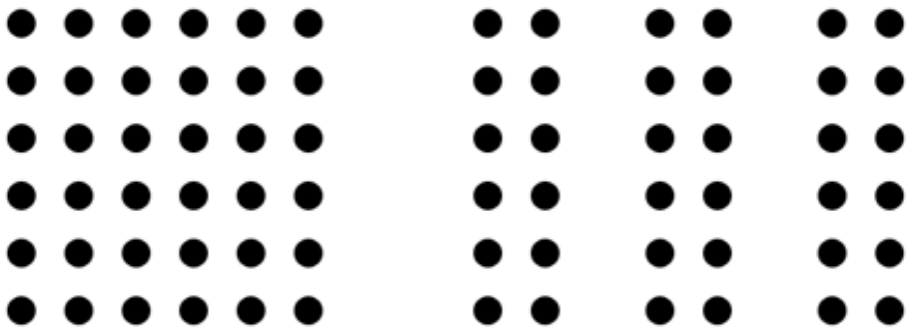
Gestalt effects

Main principles:

- **Simplicity:**
  Every pattern is seen such that the resulting structure is as simple as possible
  - Different projections perceived as 2 or 3D
  - depending on the simpler interpretation

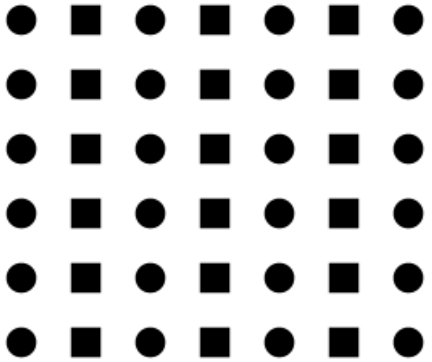


- **Proximity:**
  Things that are near to each other appear to be grouped together

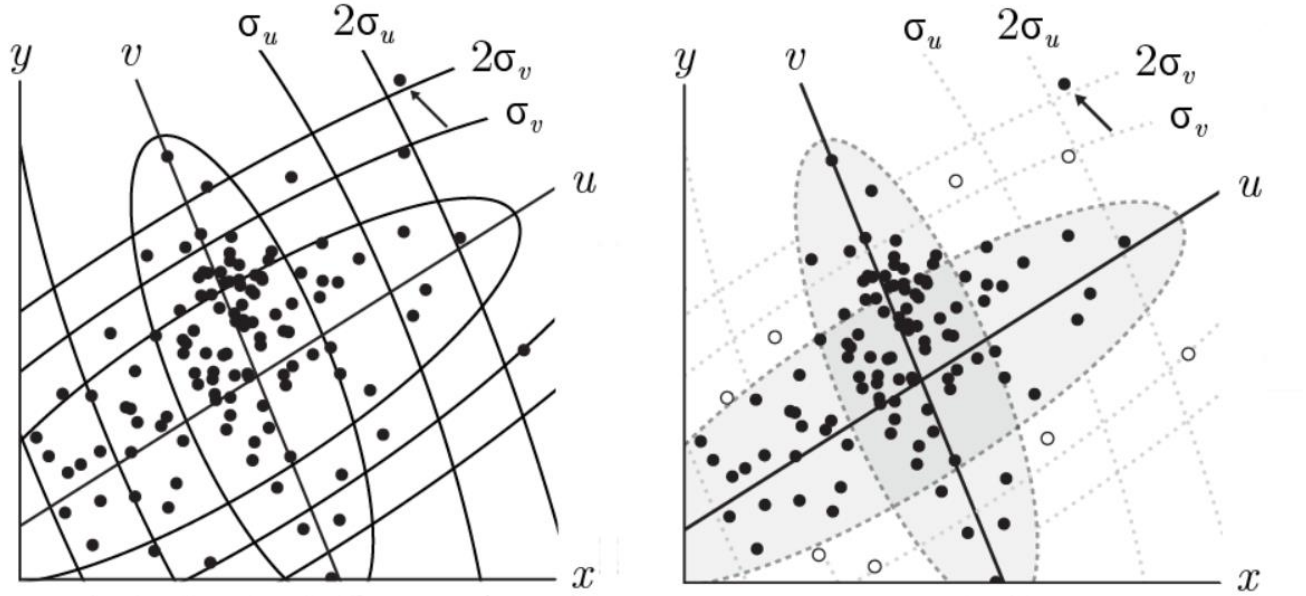


Gestalt effects

- **Similarity:**
  Similar things appear to be grouped together



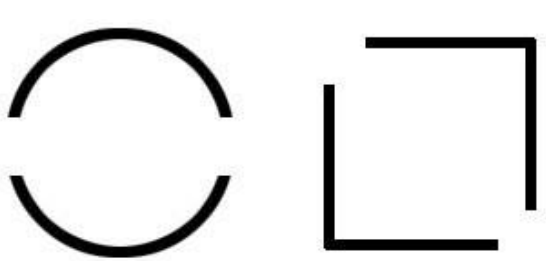
  - Use of similarity of lines and transparency to clarify images:

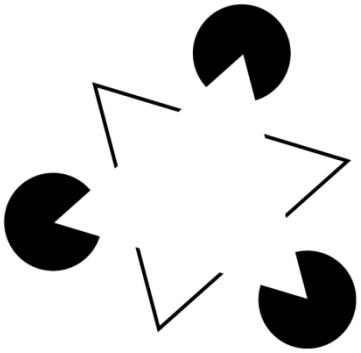


M. Krzwinski, behind every great visualization is a design principle, 2012

Gestalt effects

- **Closure:**
  We perceptually close up or complete, objects that are not complete



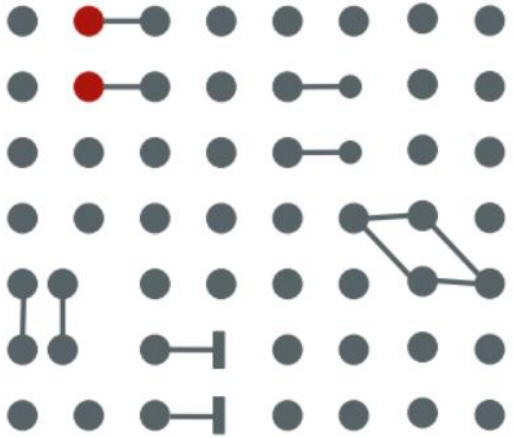


→ We automatically connect graphics, although they may be independent!

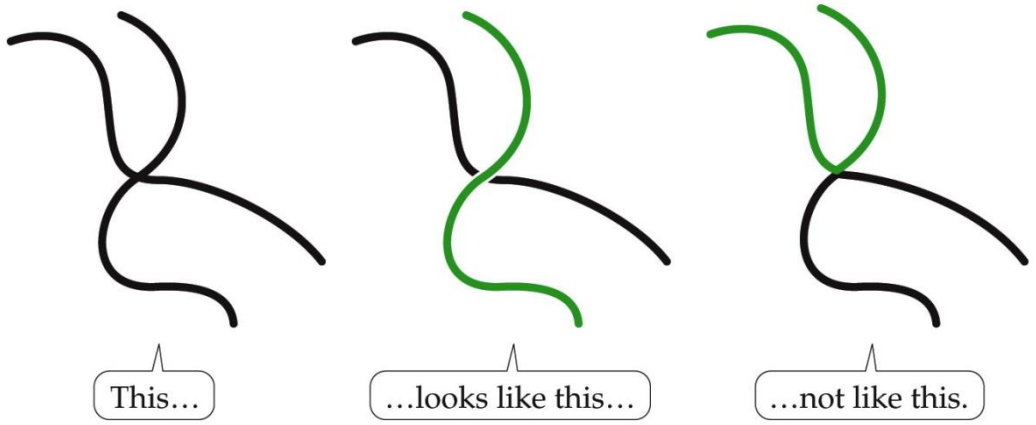
- **Symmetry:**
  Symmetrical objects are grouped together [ ]{ }[ ]

Gestalt effects

- **Connectedness:**
  Things that are physically connected are perceived as a unit
  → stronger than colour, shape, proximity or size

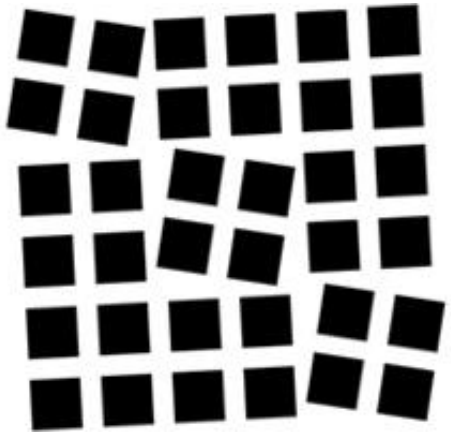


- **Continuity:**
  Points connected in a straight or smoothly curving lines are seen as belonging together
  → lines tend to be seen as to follow the smoothest path



Gestalt effects

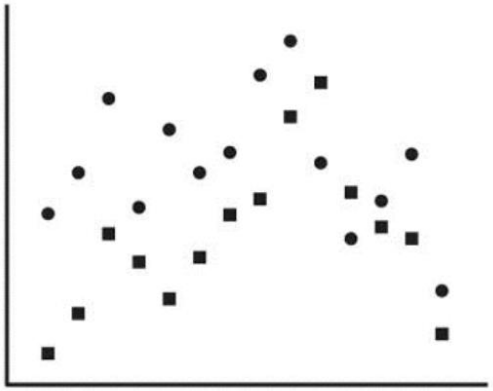
- **Common fate:**
  Things that are moving in the same direction appear to be grouped together

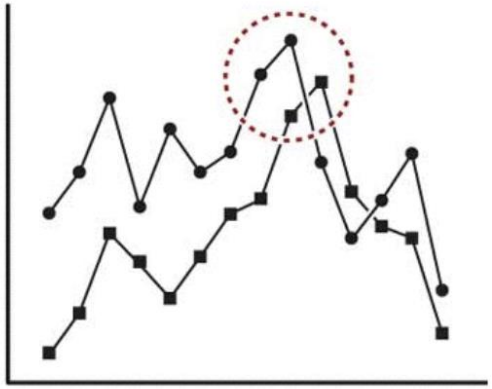


- **Background & smallness:**
  Smaller areas are seen as figures against larger background

Gestalt effects – Examples

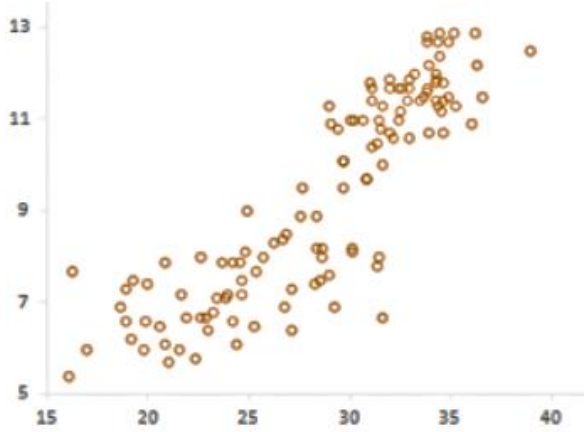
Use connectedness to make it clearer:



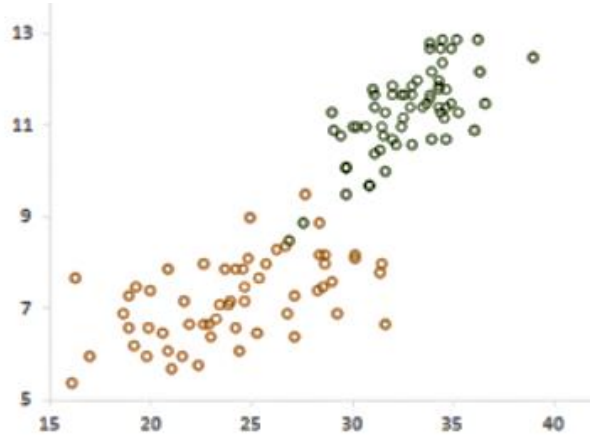


Grouping:

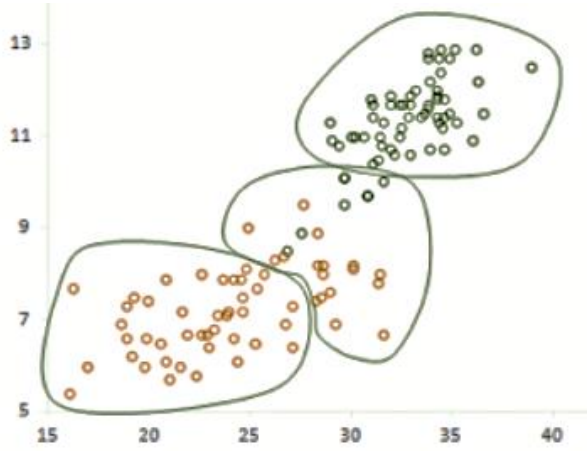
Proximity:



Similarity:

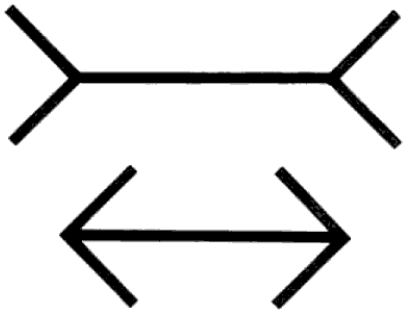


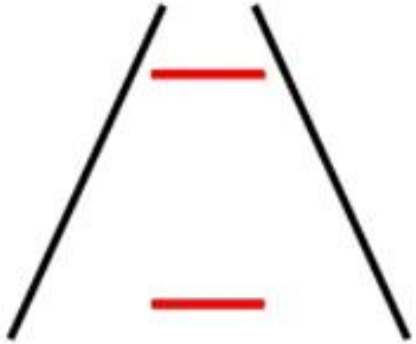
Enclosure:



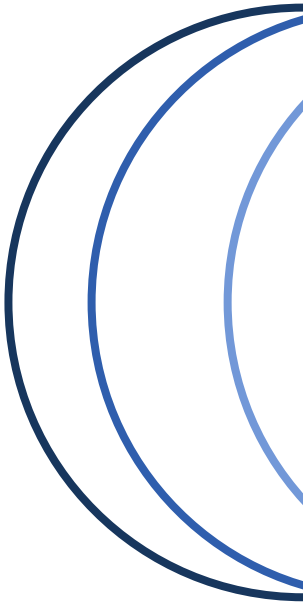
Context affects perceptual tasks

- Comparing values:
  - Length





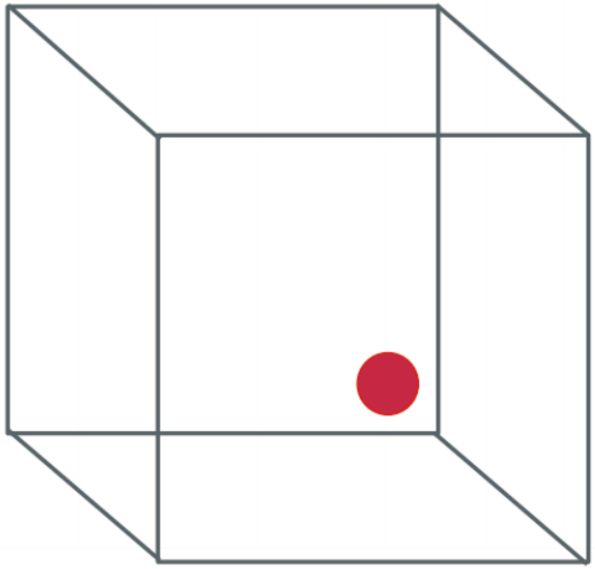
  - Curvature



Context affects perceptual tasks

- Area

- Position in 3D



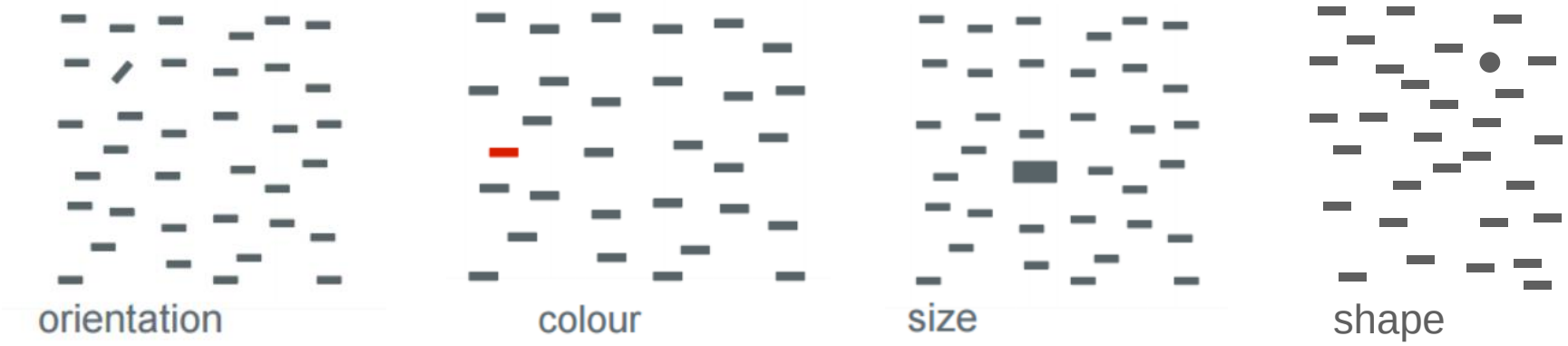
→ 3D effects reduce comprehension

Preattentive visual features

Preattentive Visual Features

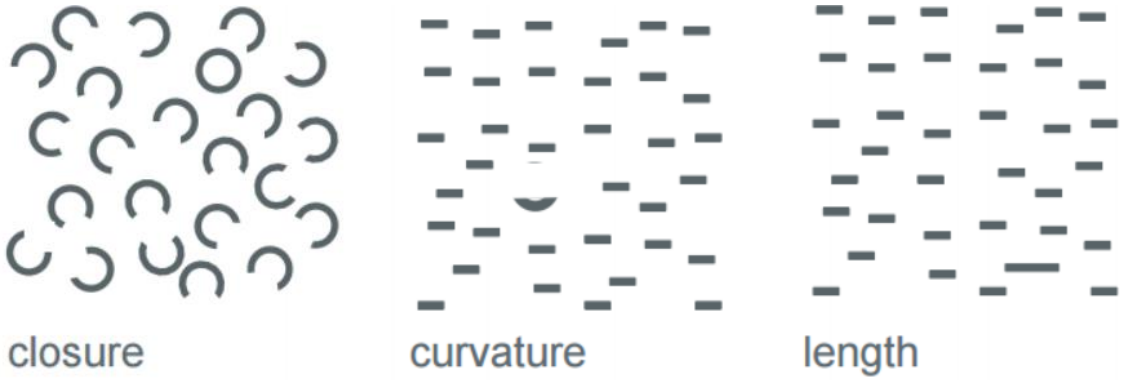
- Ability of human visual system to rapidly identify certain basic visual properties (< 200 -250 ms)
- A unique visual property is processed preattentively (e.g.: color red)

- This is important for design of visualizations
  - What can be perceived immediately?
  - Which properties are good discriminators?
  - What can mislead viewers?



Preattentive visual features

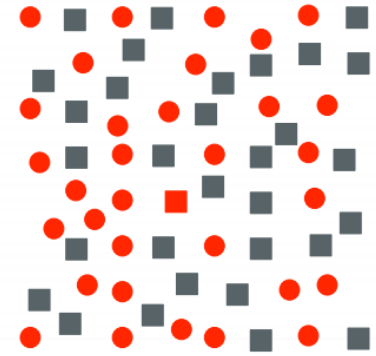
- Some are less effective:



- Some properties are asymmetric
  - Sloped line among vertical lines is preattentive
  - Vertical line among sloped ones is not

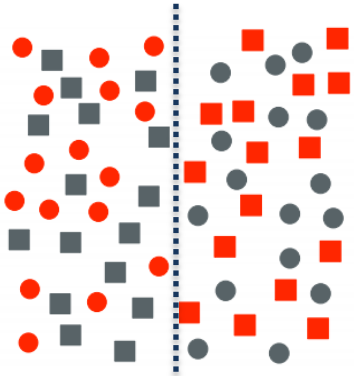
More than 2 preattentive visual features

- Combination of 2 features→ cannot be detected preattentively



→ Difficult to detect the red square
→ Serial search required

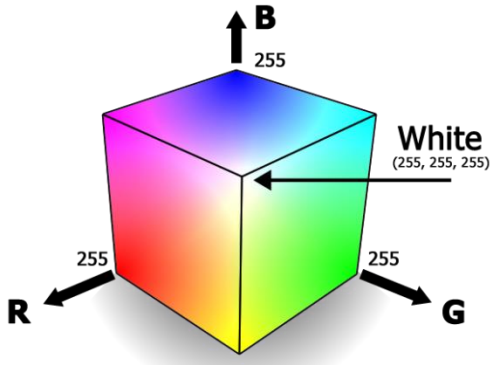
- Boundary detection
  - Users rapidly and accurately detect texture boundary between two groups of elements



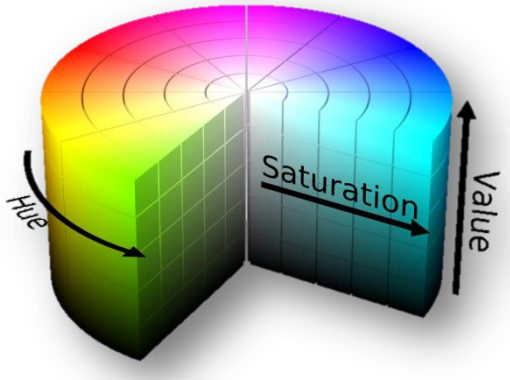
→ all elements in each group need a common visual property

Colour

- "Colour used poorly is worse than no colour at all" *Edward Tufte*
- Colour can cause the wrong information to stand out
  → make meaningful information difficult to see
- Colour space:
  - RGB: computer use



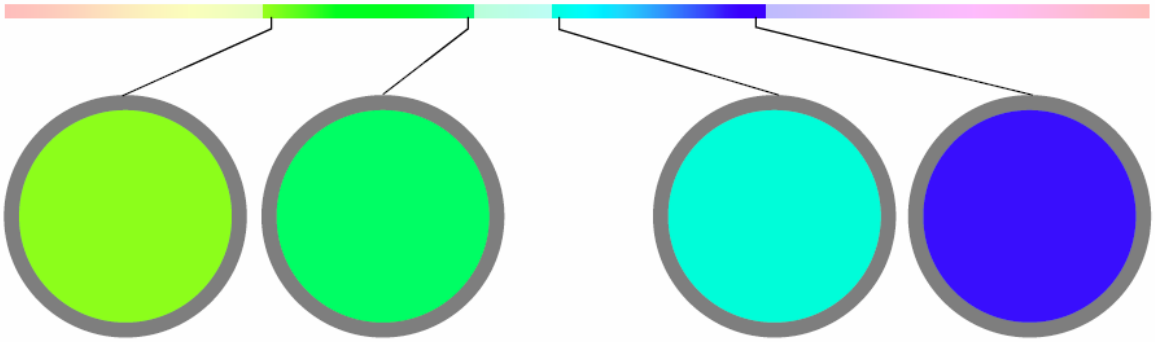
  - HSV: describe colour in terms of
    - Hue (colour)
    - Saturation
    - Lightness (value)



Perceptual differences

- Our eyes do not response linear to colours

  → play attention in quantitative encoding!



HSB COLOR SPACE	ΔH = 60		ΔH = 60	
HUE	83	143	171	231
SATURATION	1	1	1	1
BRIGHTNESS	1	1	1	1

Mimics nonlinear eye perception ←

Lab COLOR SPACE	$\Delta E_{ab} = 35$		$\Delta E_{ab} = 176$	
L	91	88	90	35
a	-59	-81	-58	70
b	87	60	4	-102

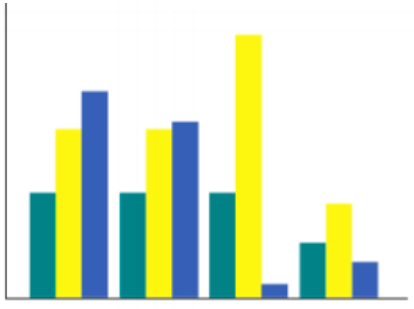
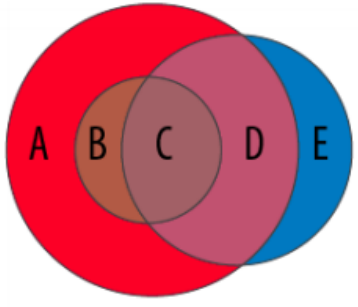
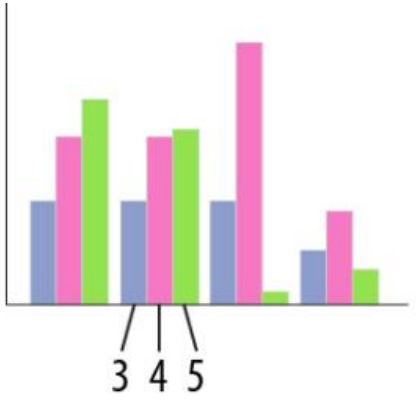
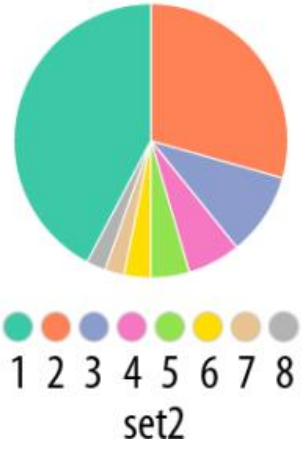
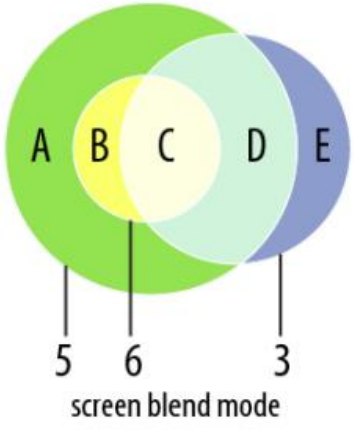
Brewer palettes

- Brewer palettes provide a range of palettes based on HSV model, which make life easier for us

Avoid to use hue to encode quantitative variables!

Fig. Courtesy of M. Krzwinski,

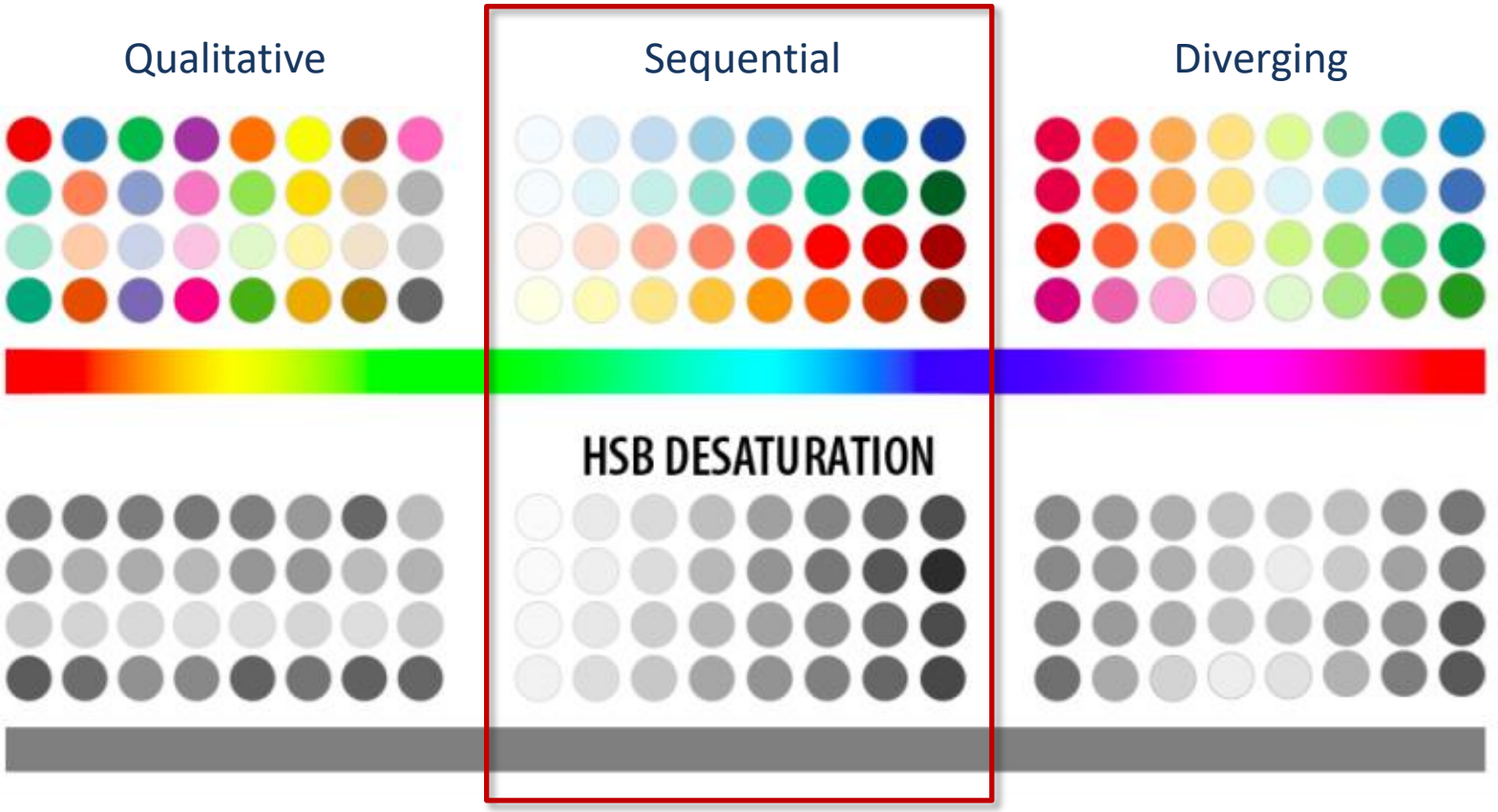
Examples

	One color dominant	Difficult to distinguish	Muddy and dark
Poor use of colour			
Brewer colours			

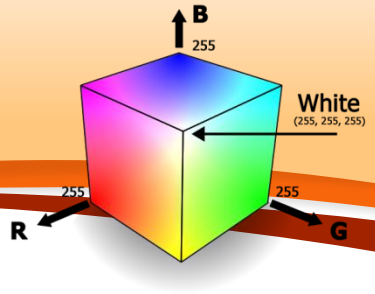
M. Krzwinski, behind every great visualization is a design principle, 2012

Conversion to grey scale

- Ensure that chosen colour set work well in grey scale (for printing)
  → Sequential palette works well here

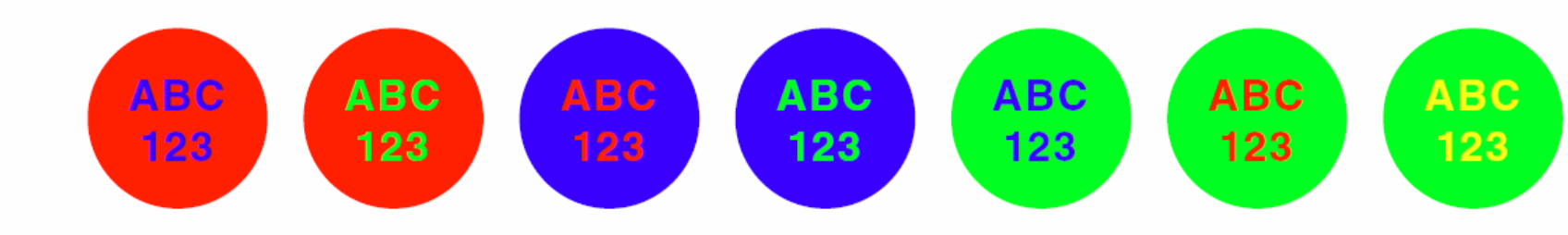


Contrast



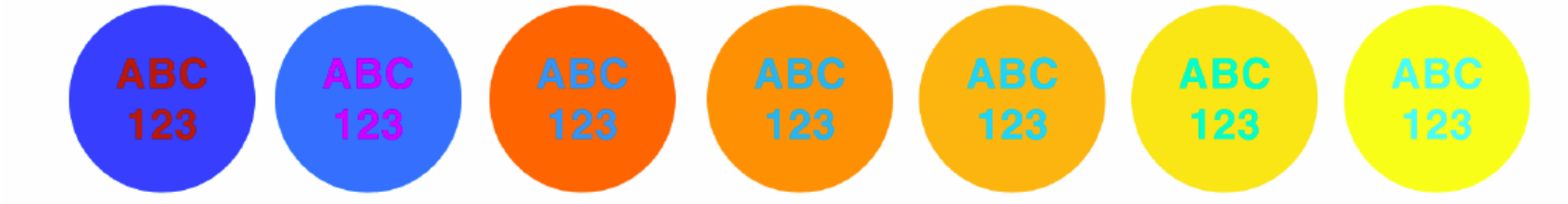
Contrast is determined by the difference in the color and brightness

- Avoid pure colors next to each other:



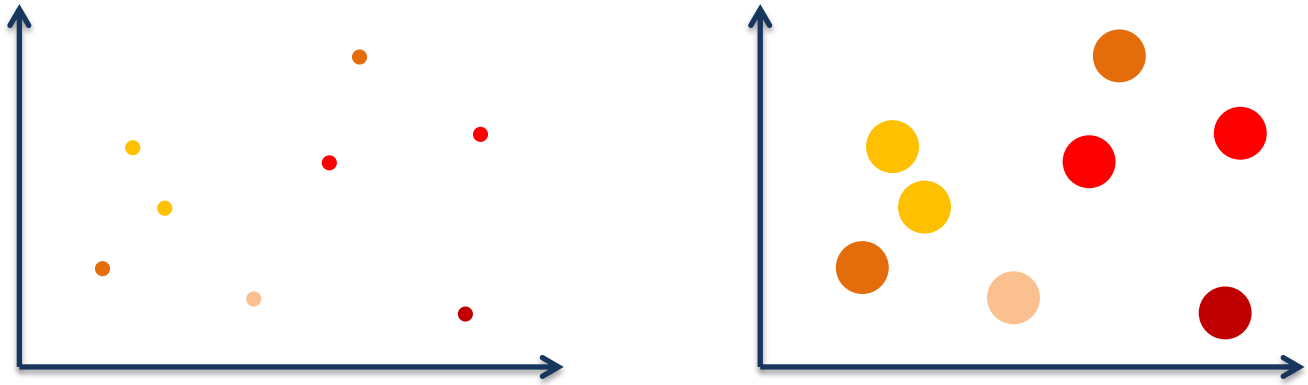
- Avoid adjacent colors with similar luminance (perceived brightness)

$$\Delta L = 0$$



Guidelines

- Saturation and colour are not separable in small regions



→ in small regions use bright, highly saturated colors

- Saturation interacts strongly with size
  - → more difficult to perceive in small regions
  - → for points and lines use just two saturation levels

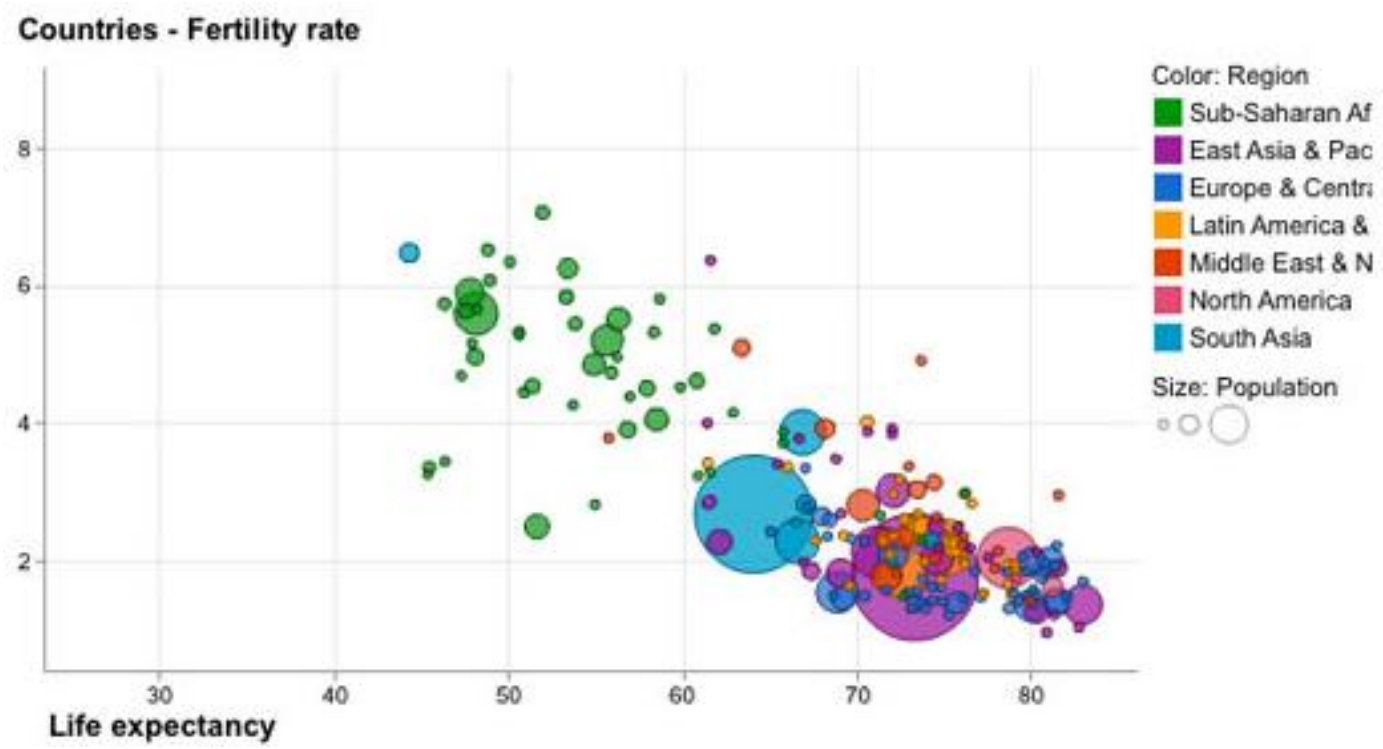
Guidelines

- Higher saturation makes large areas look bigger
  → use low saturation pastel colors for large regions and backgrounds

- Luminance and saturation are most effective for ordinal data
  → because they have an inherent ordering

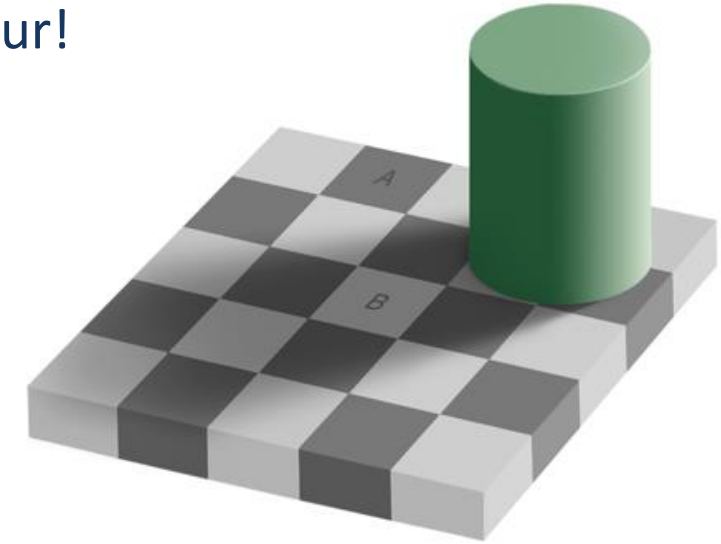
Guidelines

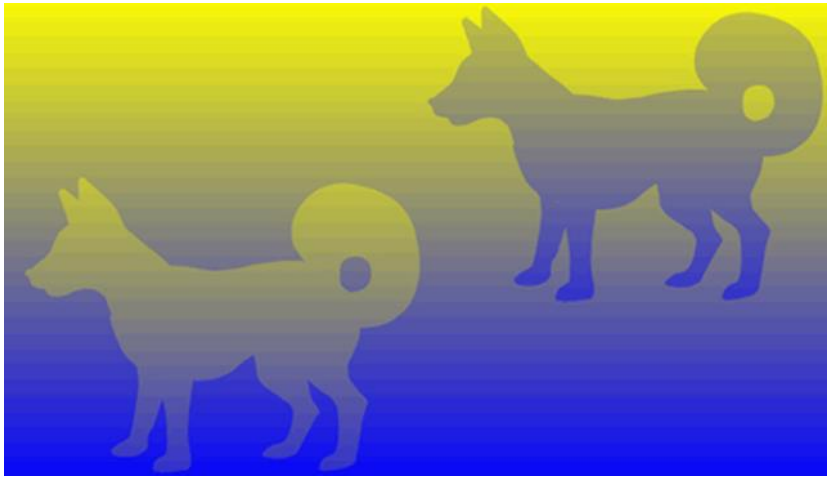
- Colour is great for categorical data
  - → because there is no inherent ordering
  - → but limit number of colour to 6-12 for distinguishability



Perceptual colour

- Be aware: context affects perceived colour!





→ Especially heat maps may be affected!

100	99	90
90	45	70
70	95	65
40	3	30
50	20	10
40	5	45
30	10	0

Differences in human colour perception

White and gold?

→ Discount the blue side (like ambient light of night time)

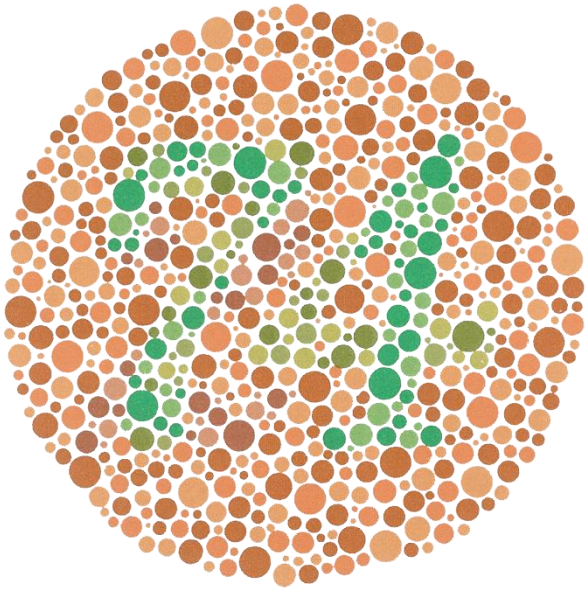
What colour is this dress?

Depends on how the brain processes the various hues of a daylight sky (“white balance”)

Blue and black?

→ Discount the gold side (like a sunny day)

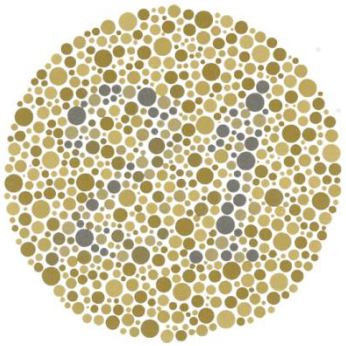
Accessibility of colours

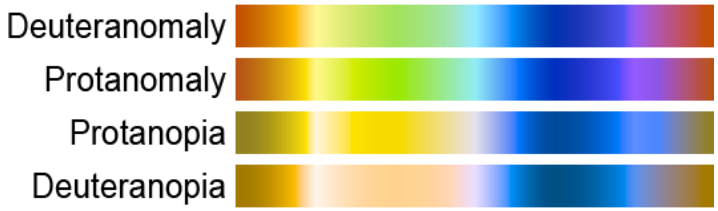


What do you see?

- 74: normal vision vision

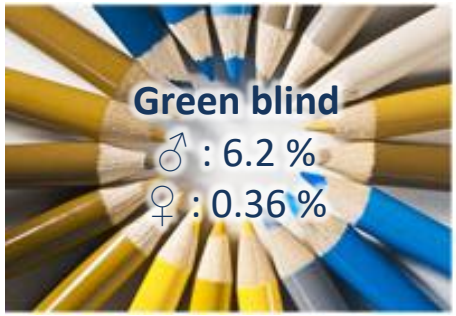
- 21: red/green blind

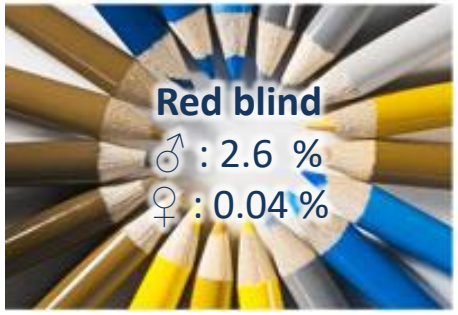




- Nothing: color blind

→ Be aware of (red/green) colour blind people!





Colour blindness

problematic!

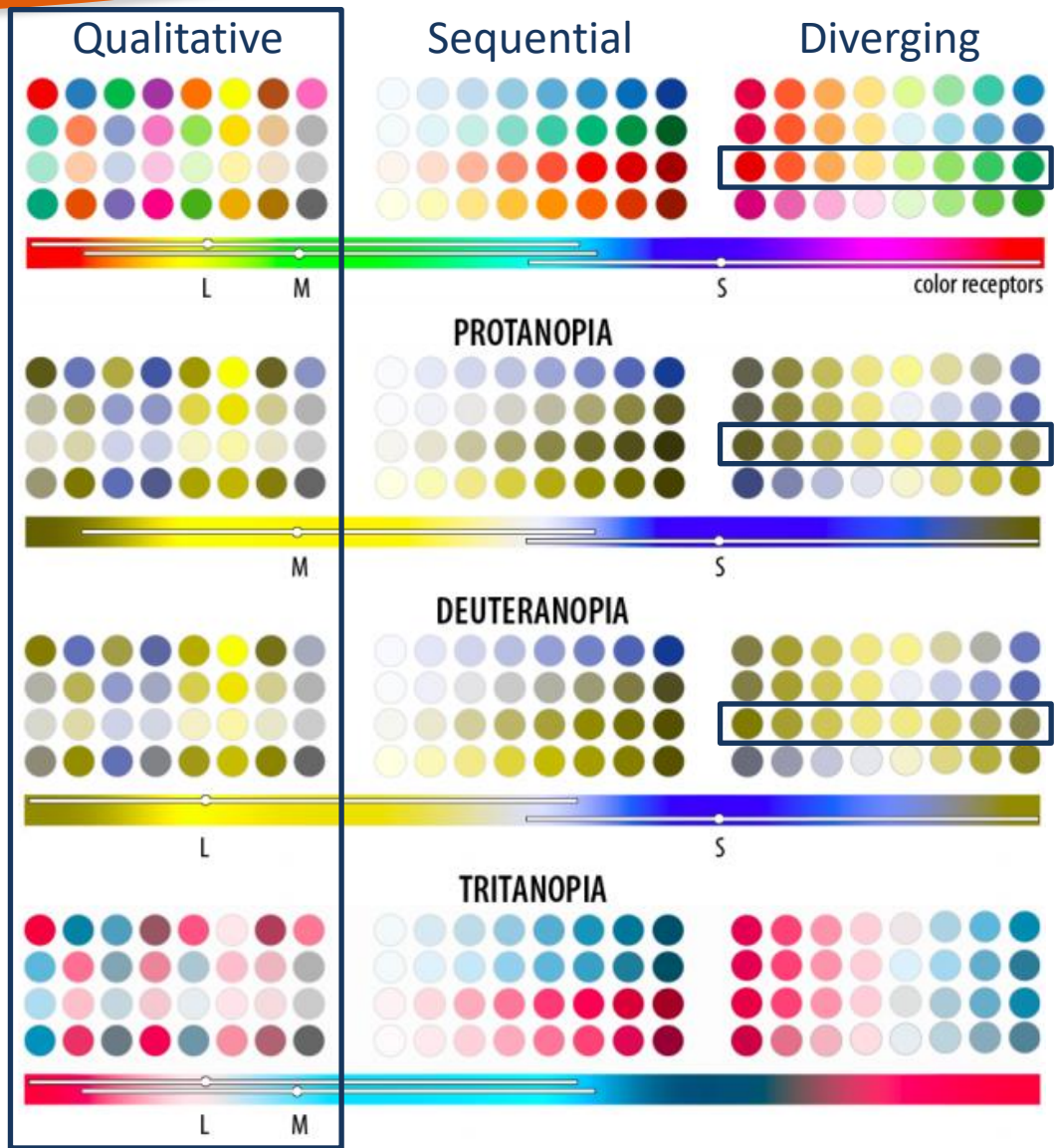


Fig. Courtesy of M Krzywinski

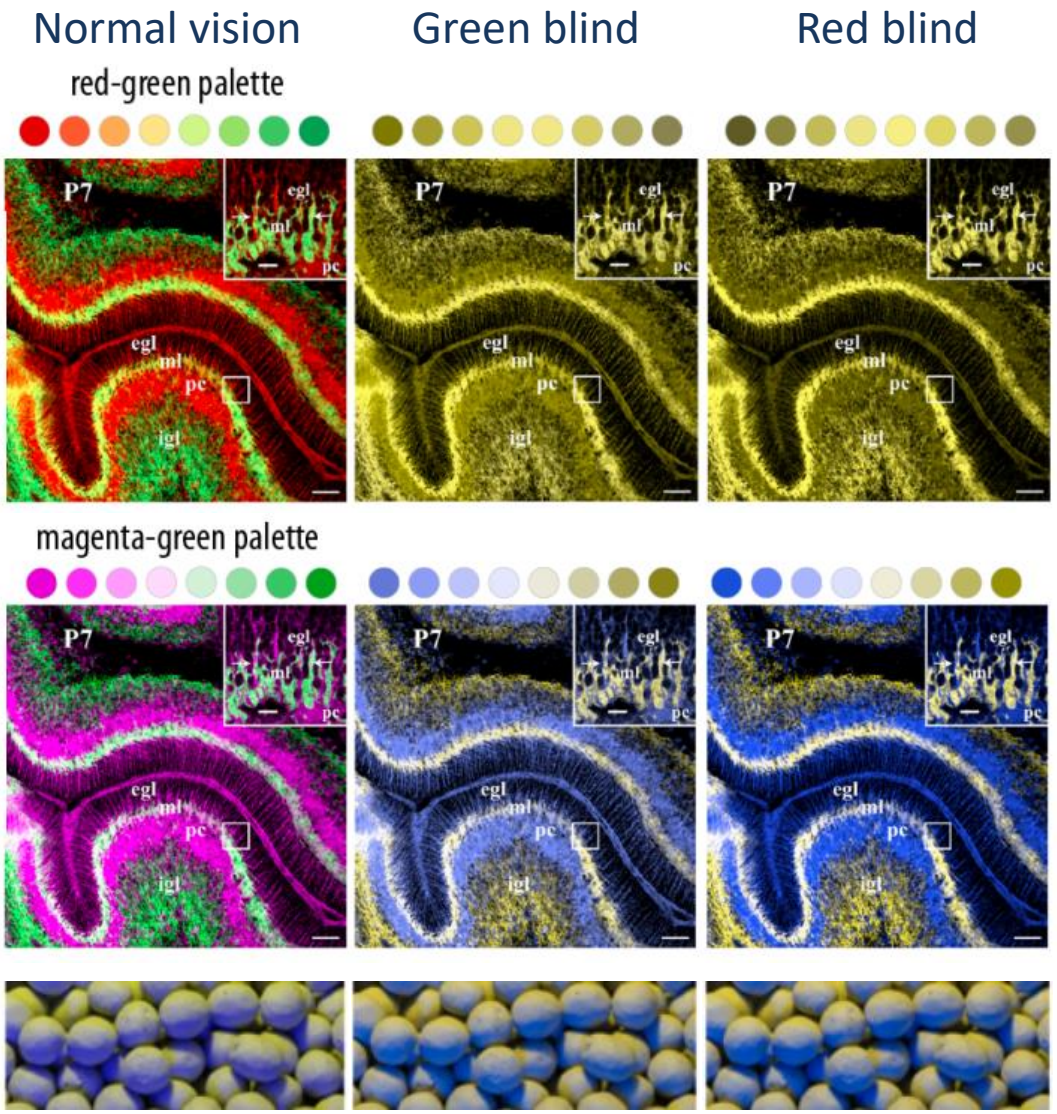
Example

Immunofluorescence images

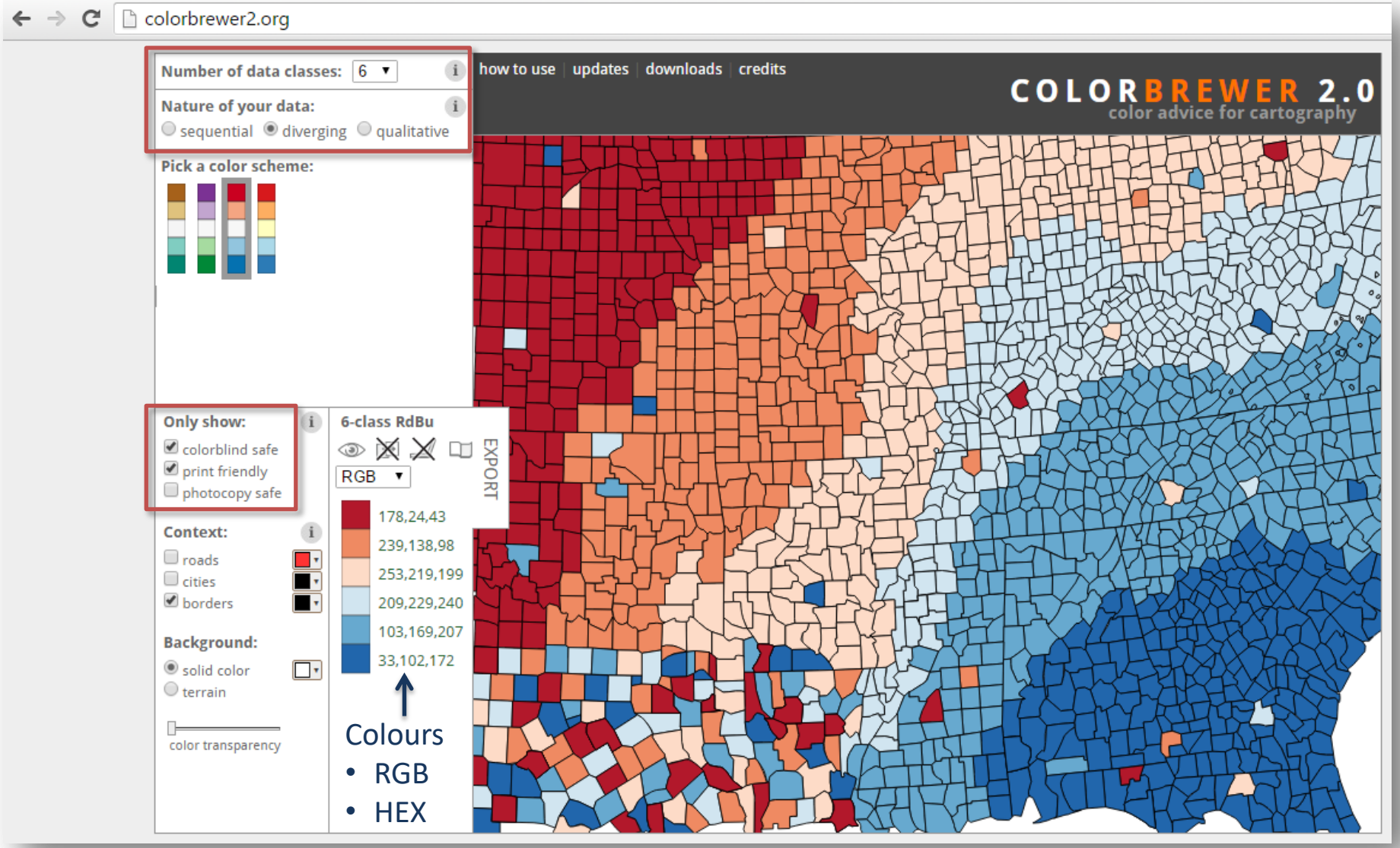
- Red-green image:

- Remapped to magenta-green:
  → suitable for red/green blindness

- Blue-yellow even better
  → talk about same colours



colorbrewer.org



Acknowledgment

- http://mkweb.bcgsc.ca/vizbi/2012/principles.pdf
- https://vizbi.org/2011/Presentations/Data_visualization_Nils_Gehlenborg.pdf