University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Data analysis and visualization (DAV)

Lecture 07

Łukasz P. Kozłowski

l.kozlowski@mimuw.edu.pl

Warsaw, 2025

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Data analysis and visualization (DAV)

Lecture 07
Perception

Łukasz P. Kozłowski

Warsaw, 2025

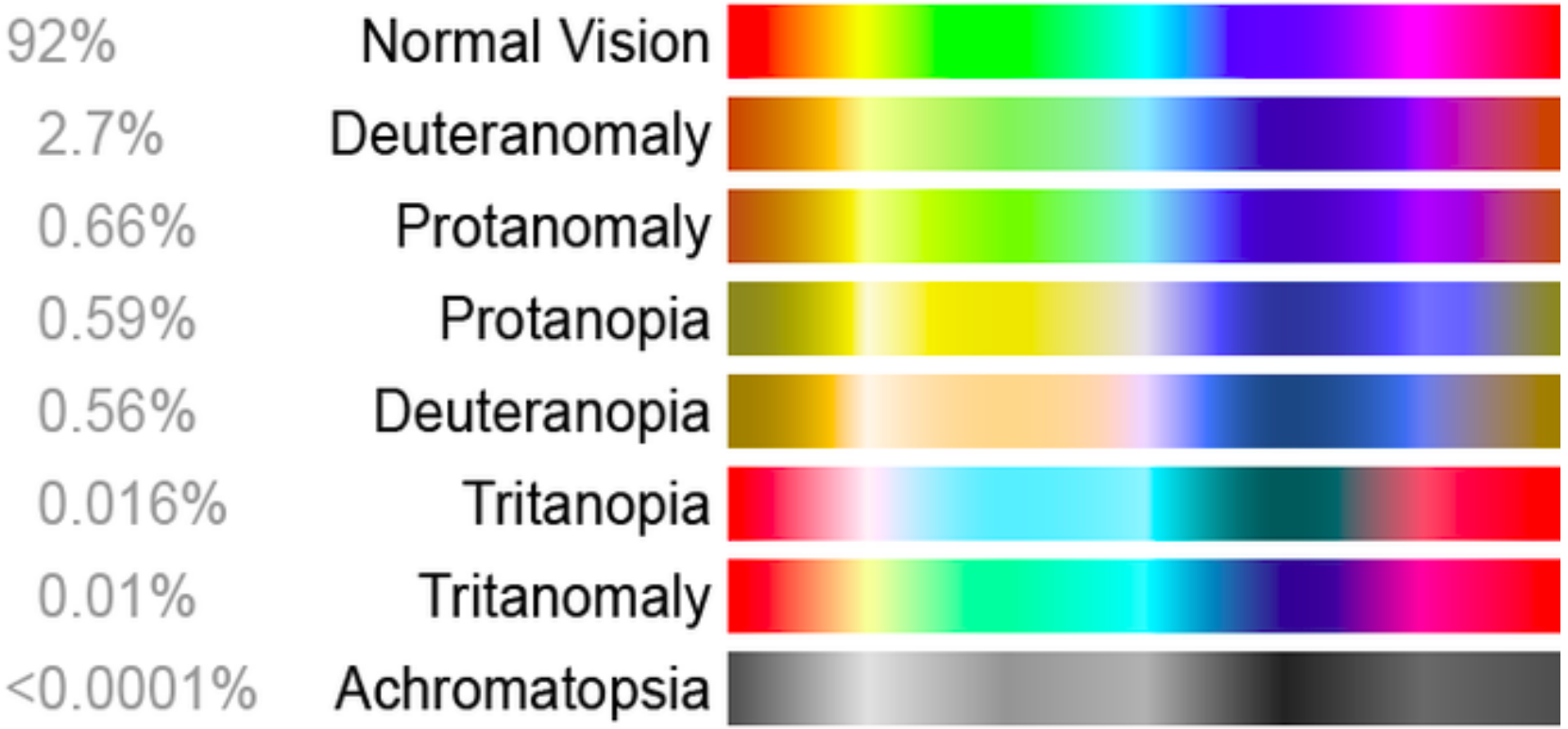
l.kozlowski@mimuw.edu.pl

Image perception

We, as the species, evolved to use the colors (in order to spot the predator or food)

Humans are more sensitive for color differences than for gray tones

Beware: Color blindness



Normal

Deuteranopia

Tritanopia

Monochromacy

Remeber:

Any plot is highly reduced/distilled form of information

Colors can distract and misinterpret

Frequently, black and white plot is more readable than color

Color interpretation depends also on external factors like:

- equipment (TV, computer screen parameters)

- sunshine

the same picture in different light conditions

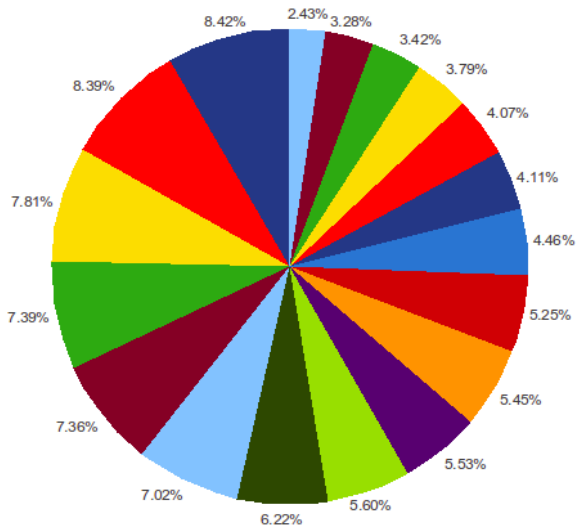
Bad font, poor light,
and you cannot see anything

During the color design for plots:

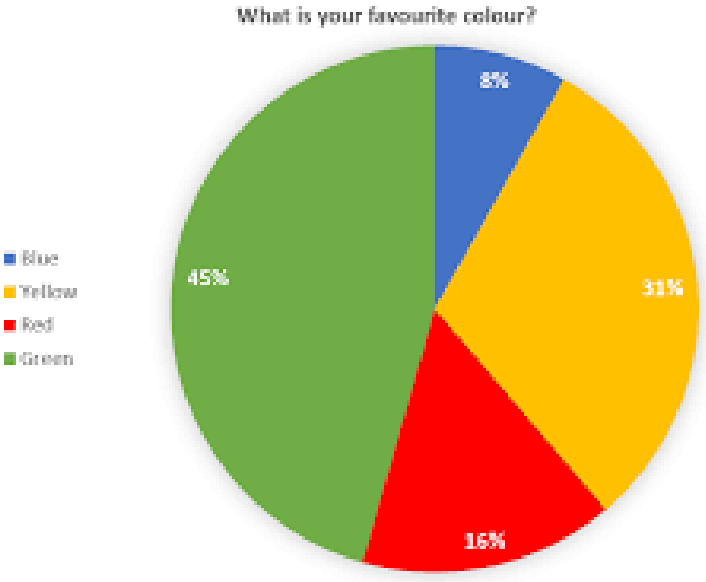
- limit to relatively small number of colors (3-4 on one plot, rarely more)

During the color design for plots:

- limit to relatively small number of colors (3-4 on one plot, rarely more)



vs.

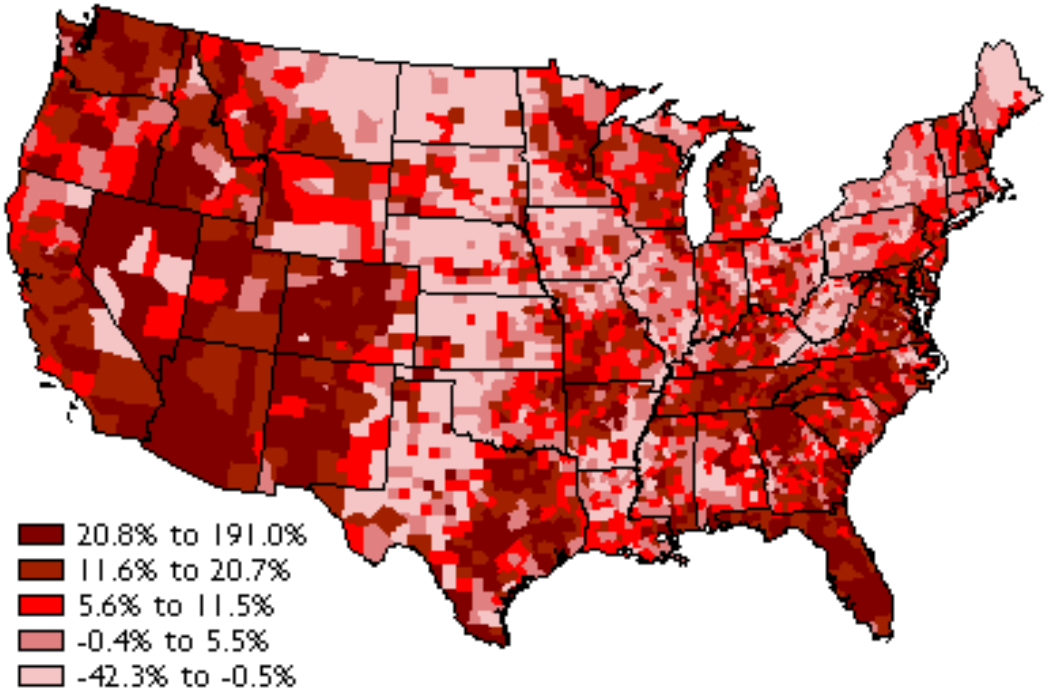


During the color design for plots:

- remember that color saturation is the worst choice for showing differences (e.g. maps)

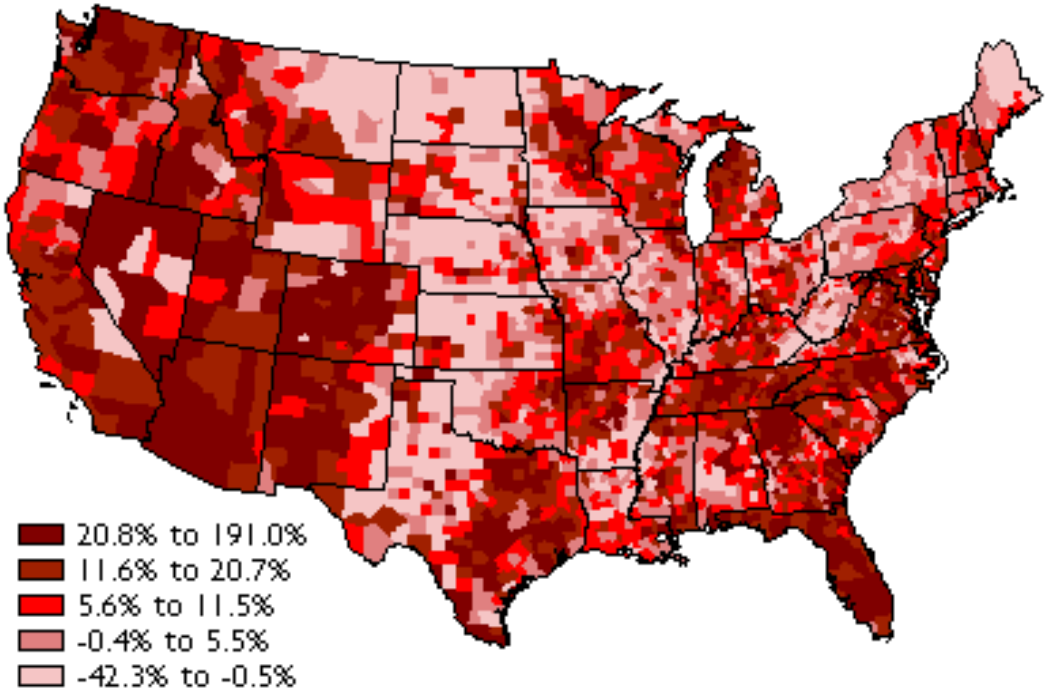
During the color design for plots:

- remember that color saturation is the worst choice for showing differences (e.g. maps)



During the color design for plots:

- remember that color saturation is the worst choice for showing differences (e.g. maps)

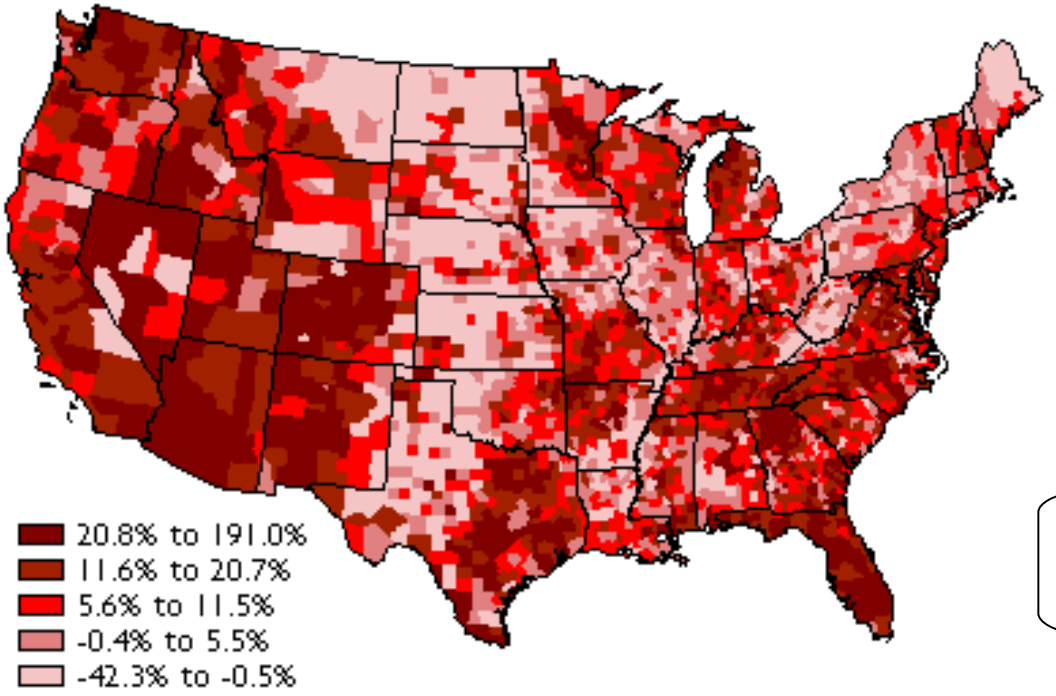


ColorBrewer for the rescue

Color Advice for Maps http://colorbrewer2.org/

During the color design for plots:

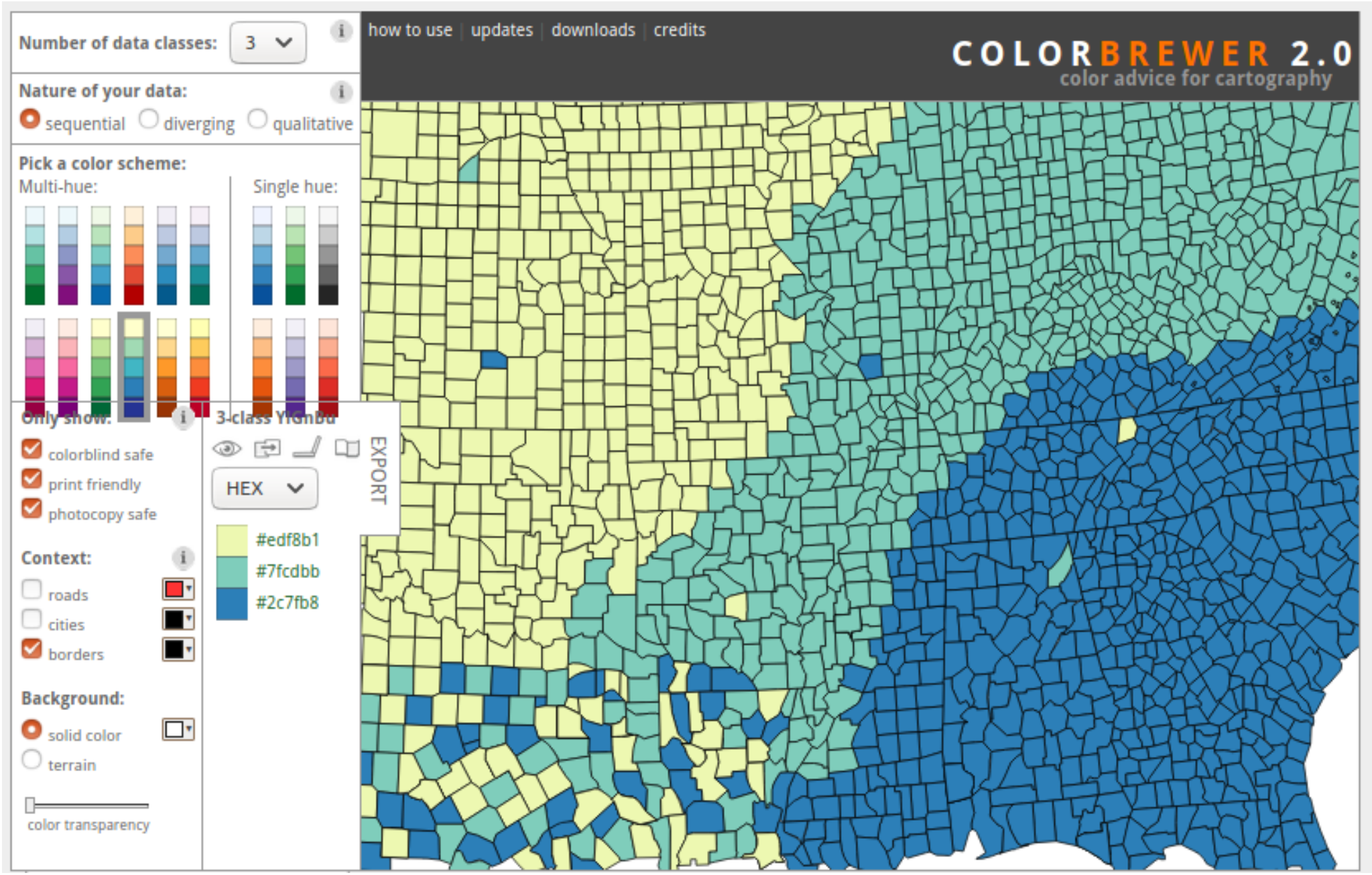
- remember that color saturation is the worst choice for showing differences (e.g. maps)



Play with it for 5 min

ColorBrewer for the rescue

Color Advice for Maps http://colorbrewer2.org/



ColorBrewer for the rescue

Color Advice for Maps http://colorbrewer2.org/

	Southern Italy	Northern Italy	Spain	Bulgaria	Switzerland	France	Southern Germany	Northern Germany	Austria	Hungary	CEU	Sweden	Czech Republic	Poland	Russia	Estonia	Lithuania	Latvia	Finland (Helsinki)	Finland (Kuusamo)
Southern Italy	0.000	0.005	0.005	0.004	0.004	0.005	0.006	0.008	0.006	0.006	0.008	0.009	0.007	0.010	0.012	0.013	0.014	0.015	0.016	0.023
Northern Italy	0.005	0.000	0.003	0.003	0.003	0.003	0.004	0.005	0.004	0.004	0.005	0.007	0.005	0.007	0.009	0.010	0.011	0.012	0.013	0.020
Spain	0.005	0.003	0.000	0.002	0.001	0.001	0.002	0.003	0.002	0.002	0.002	0.004	0.003	0.005	0.007	0.008	0.009	0.010	0.011	0.017
Bulgaria	0.004	0.003	0.002	0.000	0.001	0.002	0.001	0.002	0.001	0.001	0.003	0.004	0.002	0.003	0.004	0.005	0.006	0.007	0.009	0.015
Switzerland	0.004	0.003	0.001	0.001	0.000	0.000	0.001	0.002	0.001	0.001	0.002	0.003	0.002	0.004	0.005	0.006	0.007	0.008	0.009	0.015
France	0.005	0.003	0.001	0.002	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.003	0.005	0.005	0.006	0.008	0.008	0.015
Southern Germany	0.006	0.004	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.003	0.004	0.005	0.006	0.013
Northern Germany	0.008	0.005	0.003	0.002	0.002	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.002	0.003	0.003	0.004	0.006	0.012
Austria	0.006	0.004	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.003	0.003	0.004	0.005	0.006	0.013
Hungary	0.006	0.004	0.002	0.001	0.001	0.001	0.000	0.001	0.000	0.000	0.001	0.002	0.000	0.001	0.002	0.003	0.003	0.004	0.006	0.013
CEU	0.008	0.005	0.002	0.003	0.002	0.001	0.000	0.000	0.001	0.001	0.000	0.001	0.001	0.003	0.004	0.004	0.005	0.006	0.006	0.013
Sweden	0.009	0.007	0.004	0.004	0.003	0.002	0.001	0.001	0.001	0.002	0.001	0.000	0.002	0.002	0.003	0.003	0.004	0.005	0.005	0.011
Czech Republic	0.007	0.005	0.003	0.002	0.002	0.002	0.001	0.001	0.000	0.000	0.001	0.002	0.000	0.001	0.001	0.002	0.002	0.003	0.006	0.012
Poland	0.010	0.007	0.005	0.003	0.004	0.003	0.002	0.001	0.001	0.001	0.003	0.002	0.001	0.000	0.001	0.001	0.001	0.002	0.006	0.012
Russia	0.012	0.009	0.007	0.004	0.005	0.005	0.003	0.002	0.003	0.002	0.004	0.003	0.001	0.001	0.000	0.001	0.001	0.002	0.006	0.012
Estonia	0.013	0.010	0.008	0.005	0.006	0.005	0.003	0.003	0.003	0.003	0.004	0.003	0.002	0.001	0.001	0.000	0.001	0.001	0.004	0.009
Lithuania	0.014	0.011	0.009	0.006	0.007	0.006	0.004	0.003	0.004	0.003	0.005	0.004	0.002	0.001	0.001	0.001	0.000	0.001	0.007	0.013
Latvia	0.015	0.012	0.010	0.007	0.008	0.008	0.005	0.004	0.005	0.004	0.006	0.005	0.003	0.002	0.002	0.001	0.001	0.000	0.007	0.013
Finland (Helsinki)	0.016	0.013	0.011	0.009	0.009	0.008	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.006	0.006	0.004	0.007	0.007	0.000	0.005
Finland (Kuusamo)	0.023	0.020	0.017	0.015	0.015	0.015	0.013	0.012	0.013	0.013	0.013	0.011	0.012	0.012	0.012	0.009	0.013	0.013	0.005	0.000

CEU - Utah residents with ancestry from Northern and Western Europe

Heat maps is still better than too big table

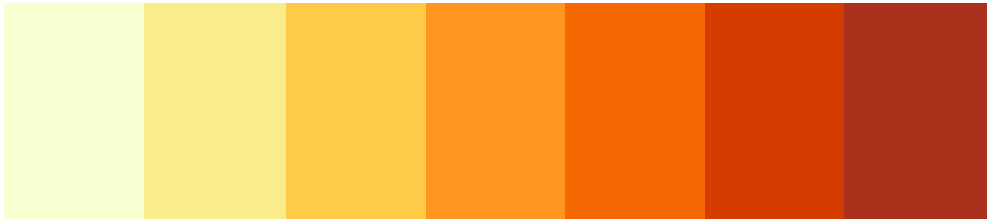
During the color design for plots:

- use HTML safe colors charts

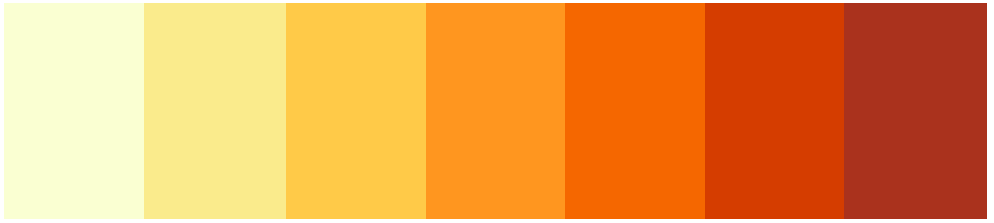
During the color design for plots:

- use HTML safe colors charts

  e.g. https://www.w3schools.com/colors/colors_picker.asp

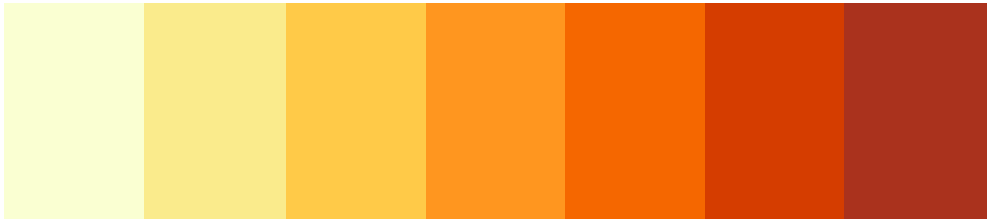


Ordered color list



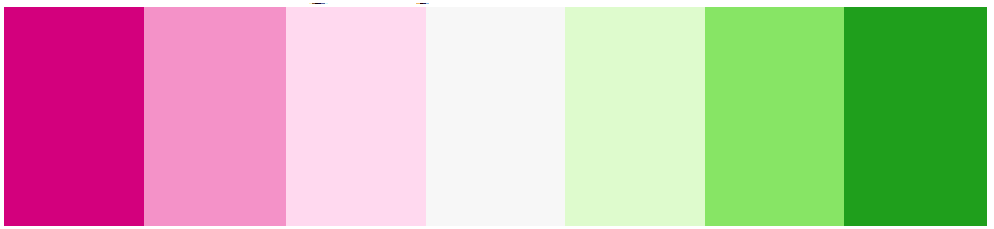
Ordered color list

Values: 0-100

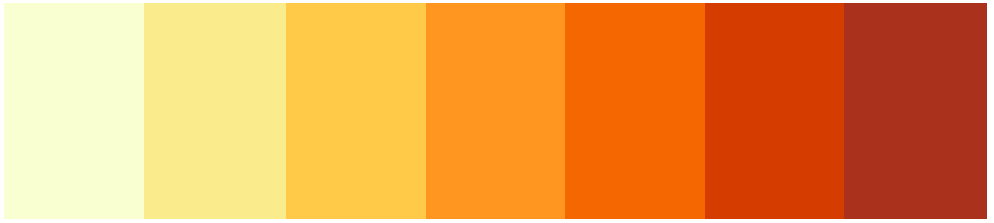


Ordered color list

Values: 0-100

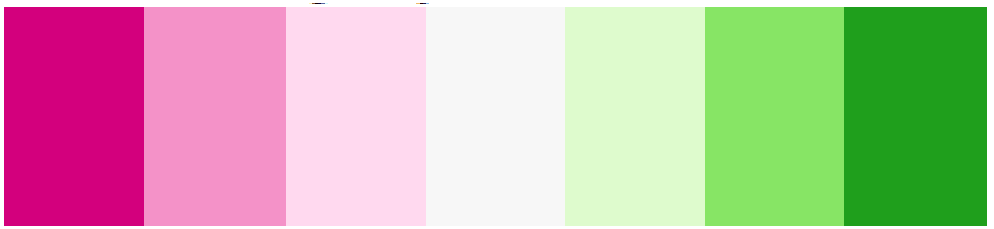


Ordered color list (rainbow)



Ordered color list

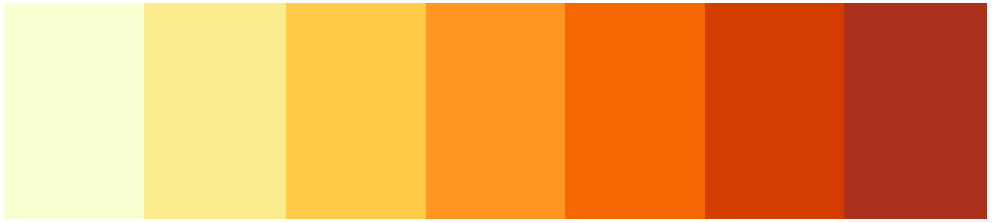
Values: 0-100



Ordered color list (rainbow)

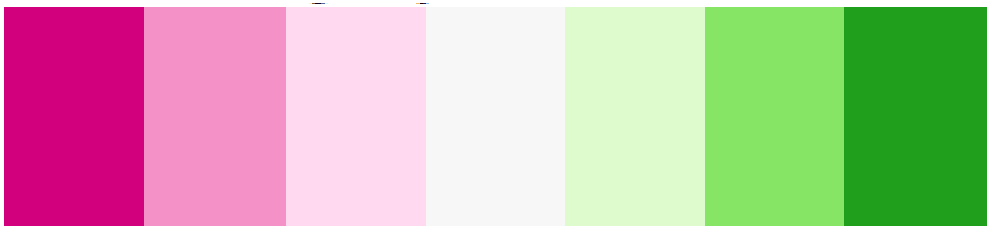
Values:

from -100 to 100



Ordered color list

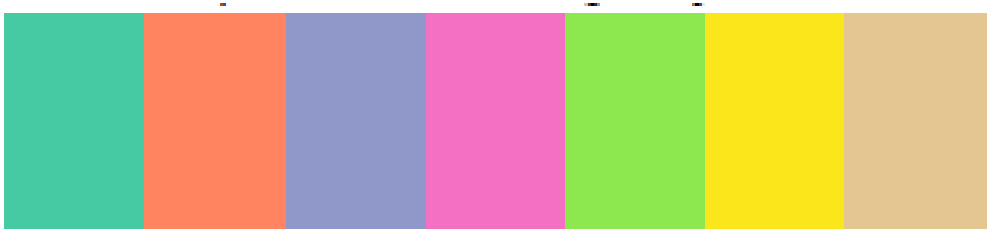
Values: 0-100



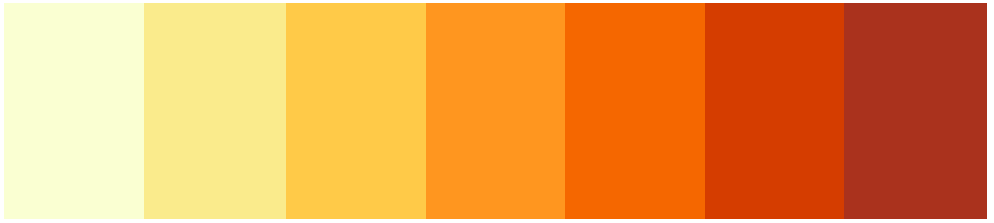
Ordered color list (rainbow)

Values:

from -100 to 100

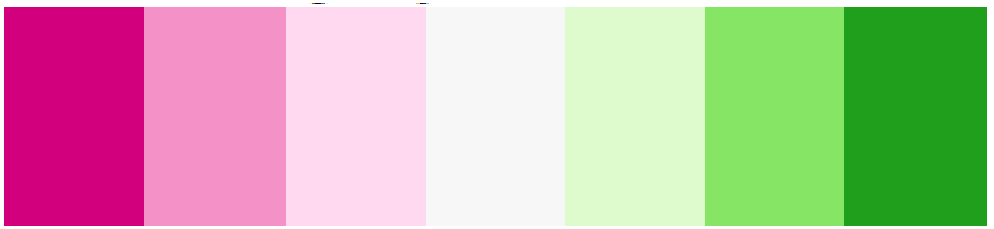


Unordered color list



Ordered color list

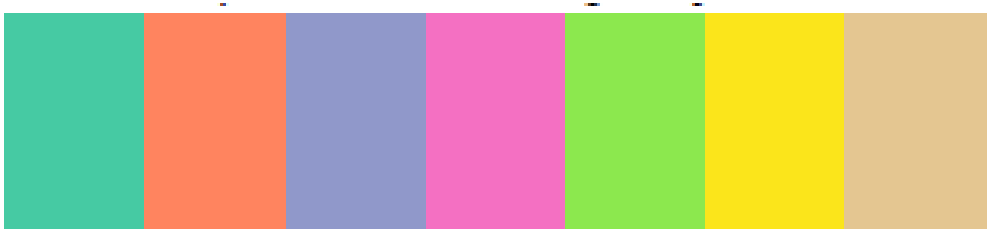
Values: 0-100



Ordered color list (rainbow)

Values:

from -100 to 100



Unordered color list

Values: classes

Thus as color vision is so frequently wrong or disrupted you need to remember about this in your plots

Scales

interval scale

ratio scale

ordinal scale

nominal scale

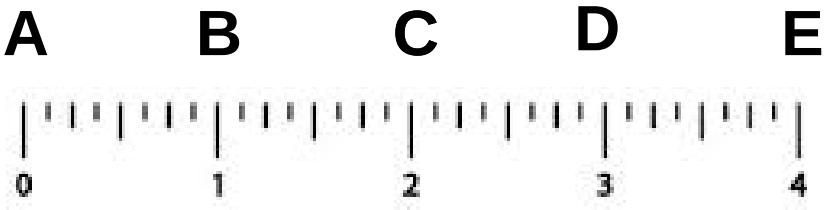
Scales

Interval scale

- the most common (default)

- each unit has the same span/length

e.g. temperature, time, length



A - B = D – E = 1 (unit)

Scales

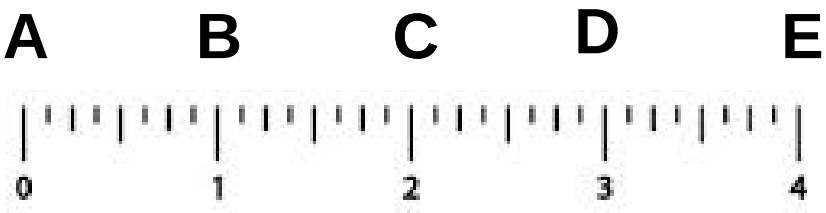
Interval scale

- the most common (default)

- each unit has the same span/length

e.g. temperature, time, length

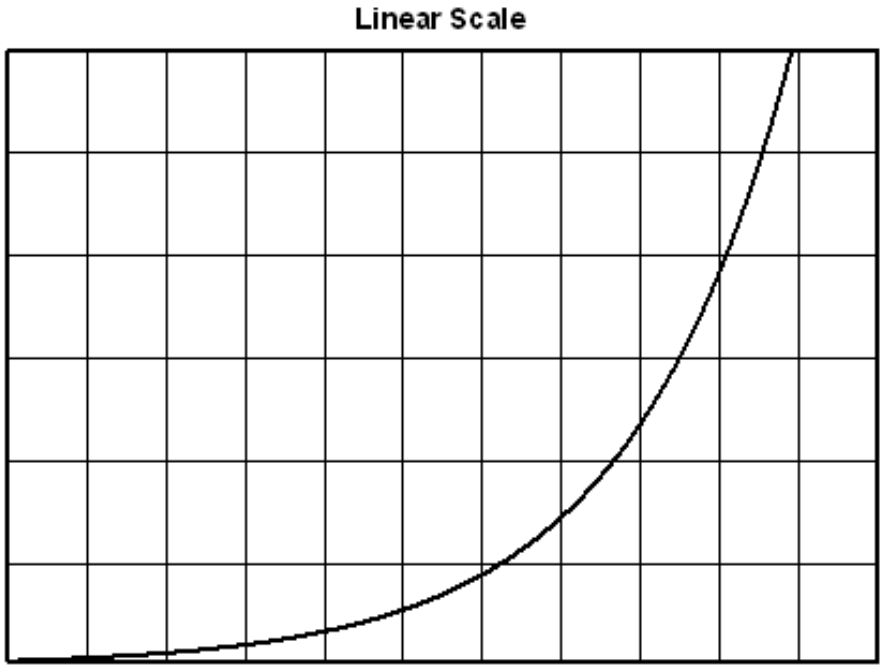
When there is no indication on the plot, it probably is a linear (interval) scale



A - B = D – E = 1 (unit)

Scales

Ratio scale – why to use it?

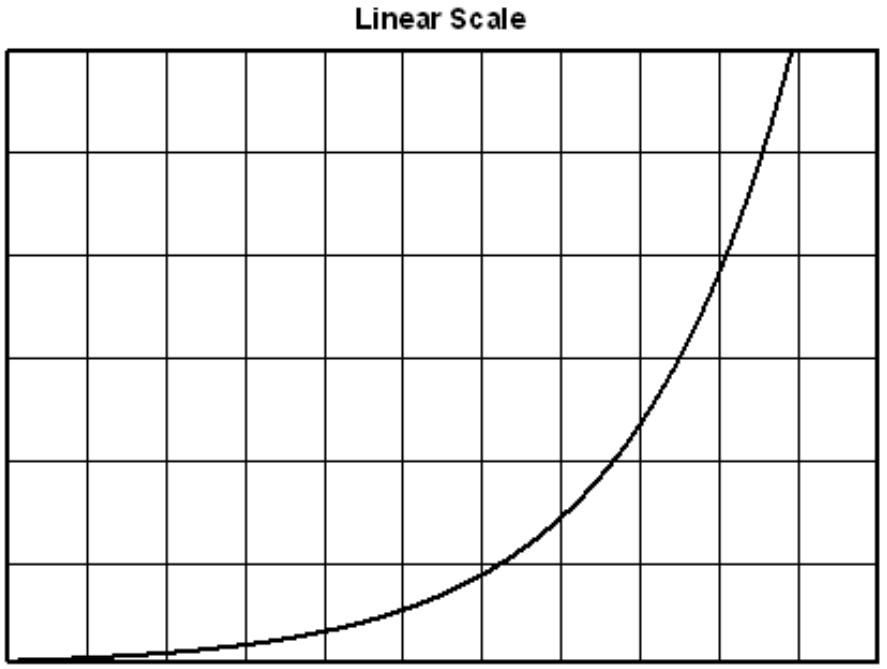


a trend that appears out of control

Conclusions: the sky is the limit, the trend is exploding

Scales

Ratio scale – why to use it?

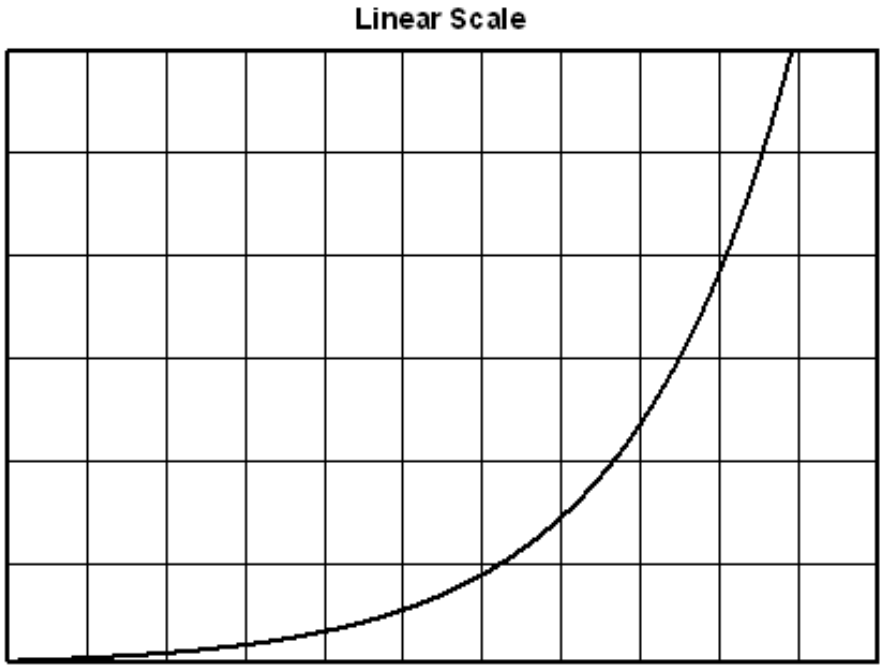


a trend that appears out of control

Conclusions: the sky is the limit, the trend is exploding

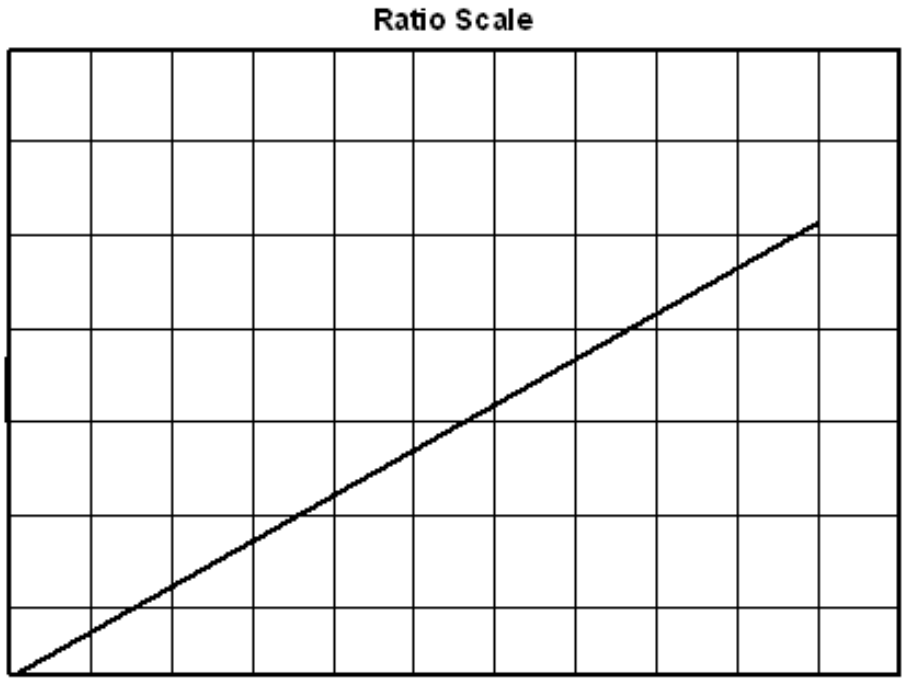
Scales

Ratio scale – why to use it?



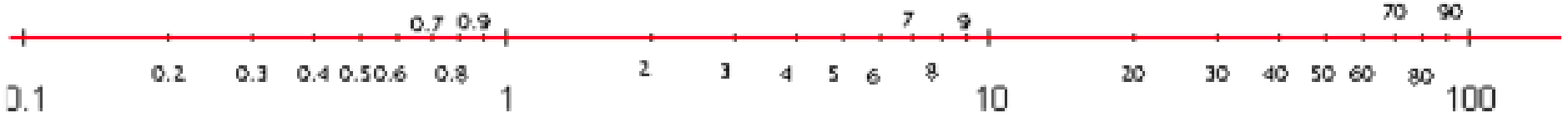
a trend that appears out of control

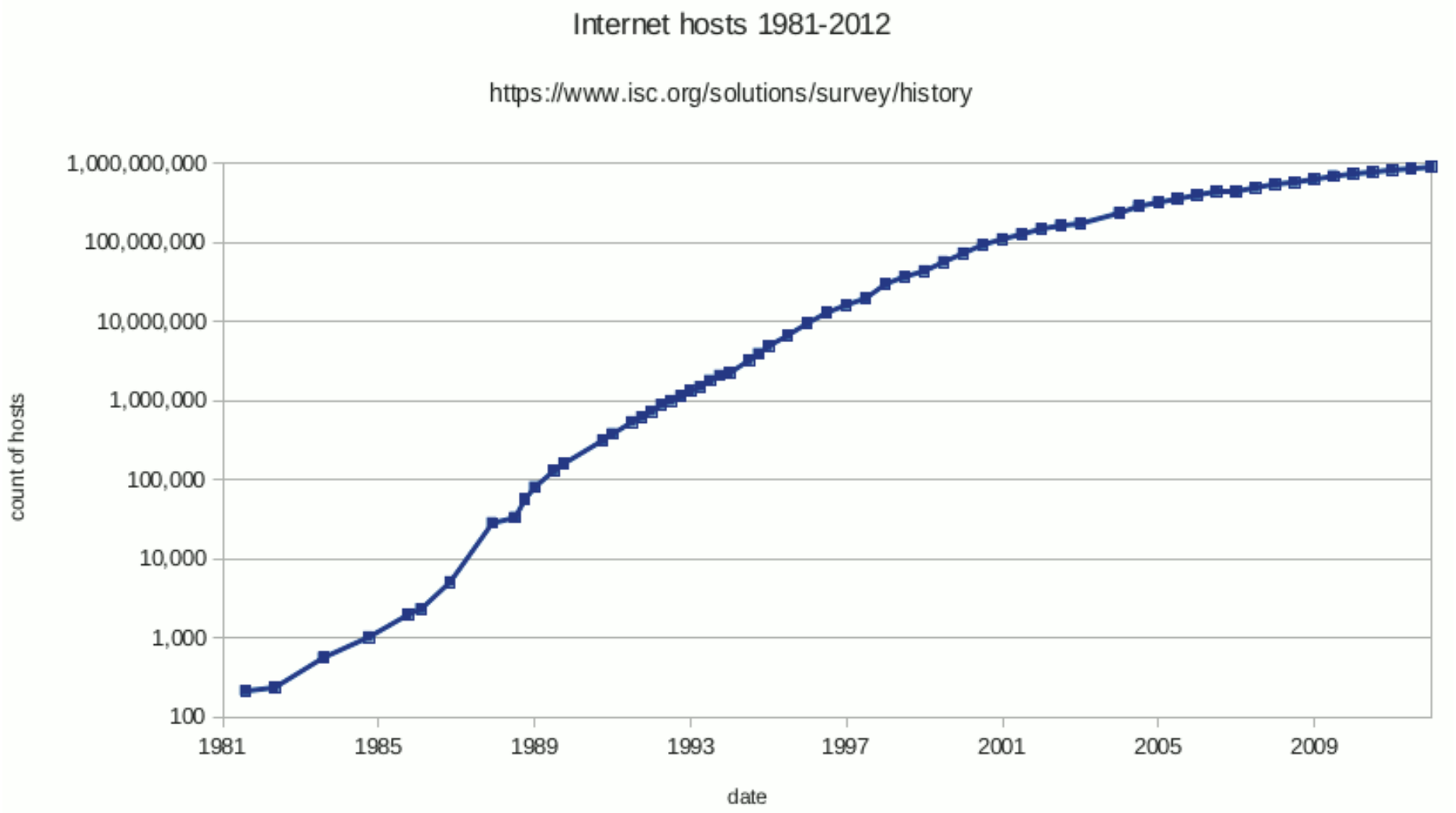
Conclusions: the sky is the limit, the trend is exploding



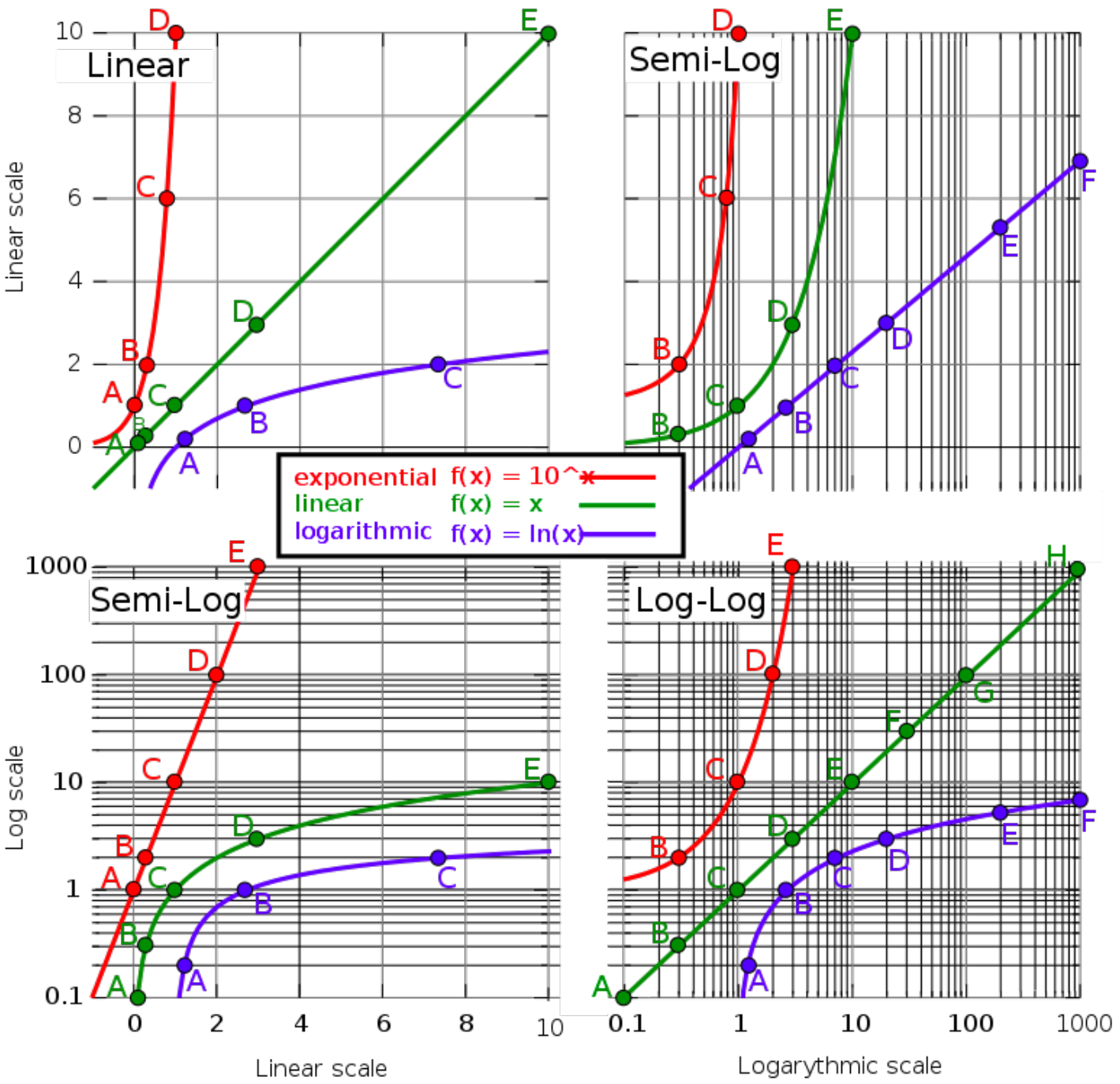
The same on ratio scale
(5% growth rate)

A logarithmic scale from 0.1 to 100





Graph on a logarithmic scale


Linear
Semi-Log
Semi-Log
Log-Log
exponential f(x) = 10^x
linear f(x) = x
logarithmic f(x) = ln(x)
Linear scale
Log scale
Linear scale
Logarythmic scale

Dirrect Applications

- **Entropy (information theory)**

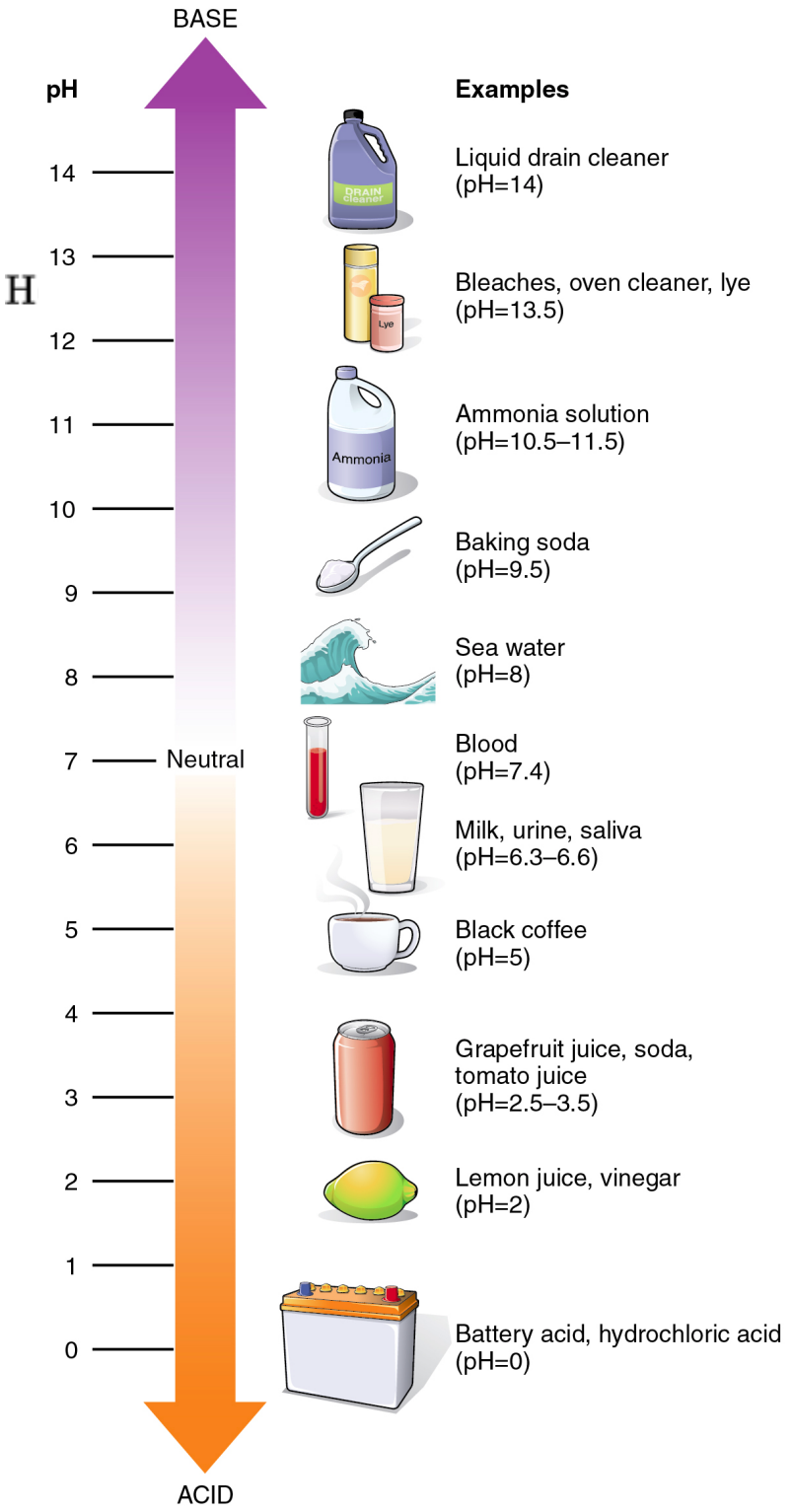
$$S = \log_2 2^N = N = 2$$

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log_b \mathrm{P}(x_i)$$

Dirrect Applications

- Entropy (information theory)

- pH

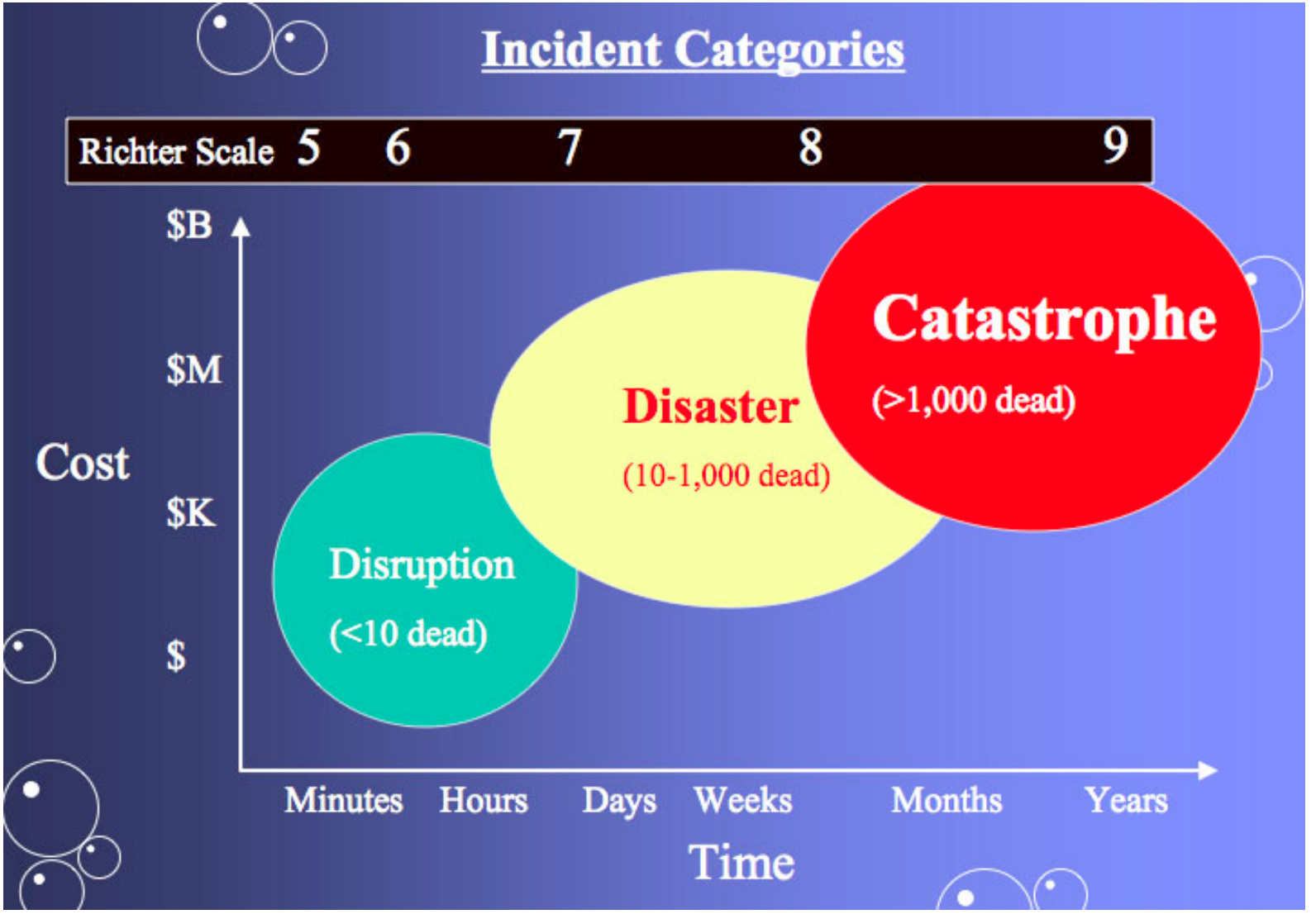


Dirrect Applications

- Entropy (information theory)

- pH

- earthquakes



Scales

Ratio scale – when (not) to use it?

- appropriate for investment prices, sales figures, income, or any other absolute amounts being plotted over a period of time

- should not be used to plot anything in which a relationship is already inherent in the amounts — such as percentages (like the inflation rate), ratios between two items (such as a gold-silver ratio or price-earnings ratio) — because the benefit provided by a ratio scale is already built into the figures being plotted

Ratio scales tell us:
- about the order
- the exact value between units
- have an absolute zero (important for the statistics*)

* variables can be meaningfully added, subtracted, multiplied, divided (ratios). Central tendency can be measured by mode, median, or mean; measures of dispersion, such as standard deviation and coefficient of variation can also be calculated from ratio scales

Scales

Ordinal scale

We care only about the order, the difference has no meaning

1st place is better than second, but how much is worth golden medal in comparison to silver?

Is having 2 silver medals is like having 1 golden?
Is placing ten times at 10th position is equal to race only once and win?

Scales

Ordinal scale – other examples

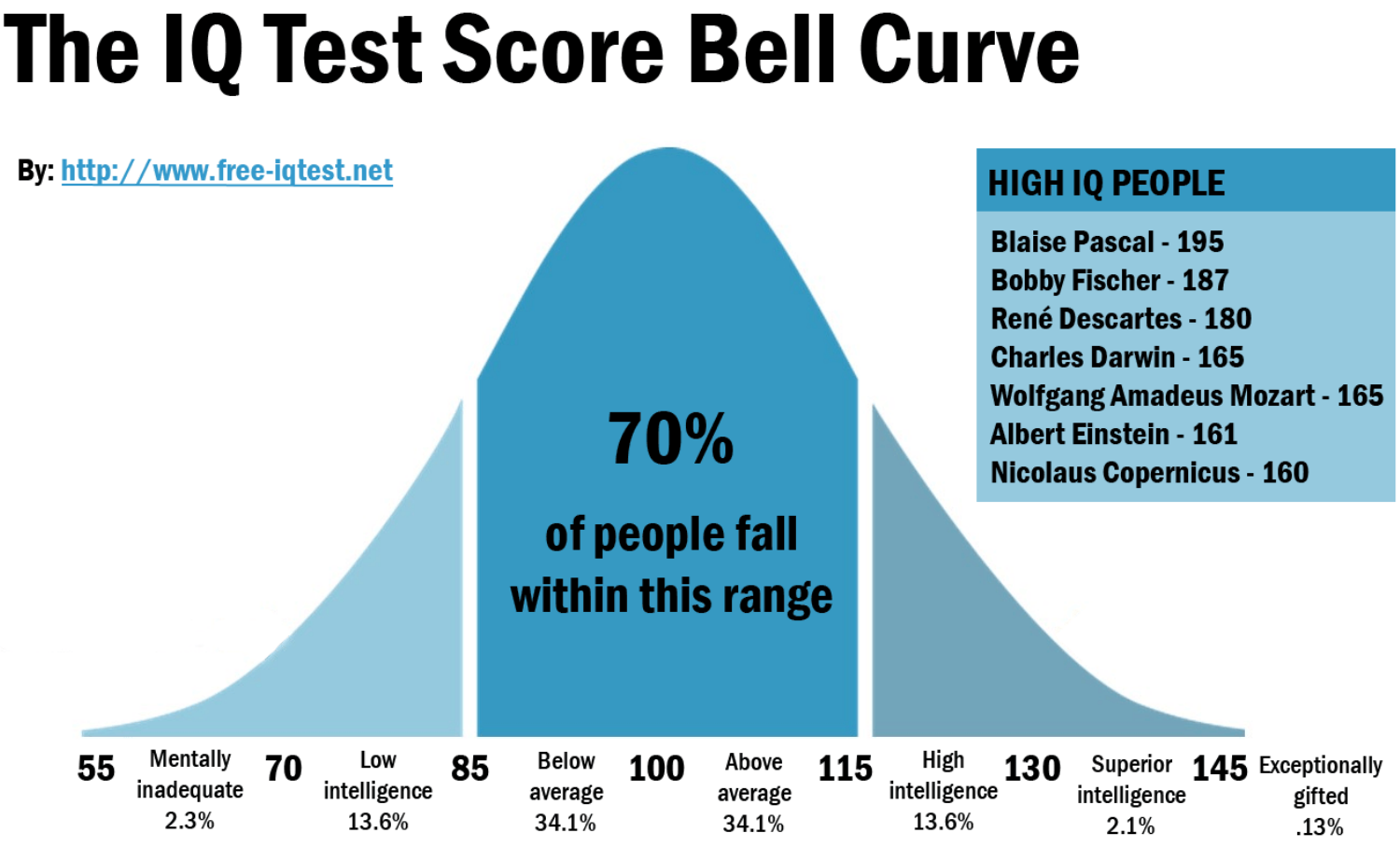
Percentage points	Letter grade	Quality points
97.0–100.00	A+	4.33
93.0–96.99	A	4.00
90.0–92.99	A–	3.67
87.0–89.99	B+	3.33
83.0–86.99	B	3.00
80.0–82.99	B–	2.67
77.0–79.99	C+	2.33
73.0–76.99	C	2.00
70.0–72.99	C–	1.67
67.0–69.99	D	1.00
0.0–66.99	F	0.00

Grades in Polish schools are 1-6

Is getting 6 equal getting twice 3?
How many 5 you need to get to compensate one 4?

Scales

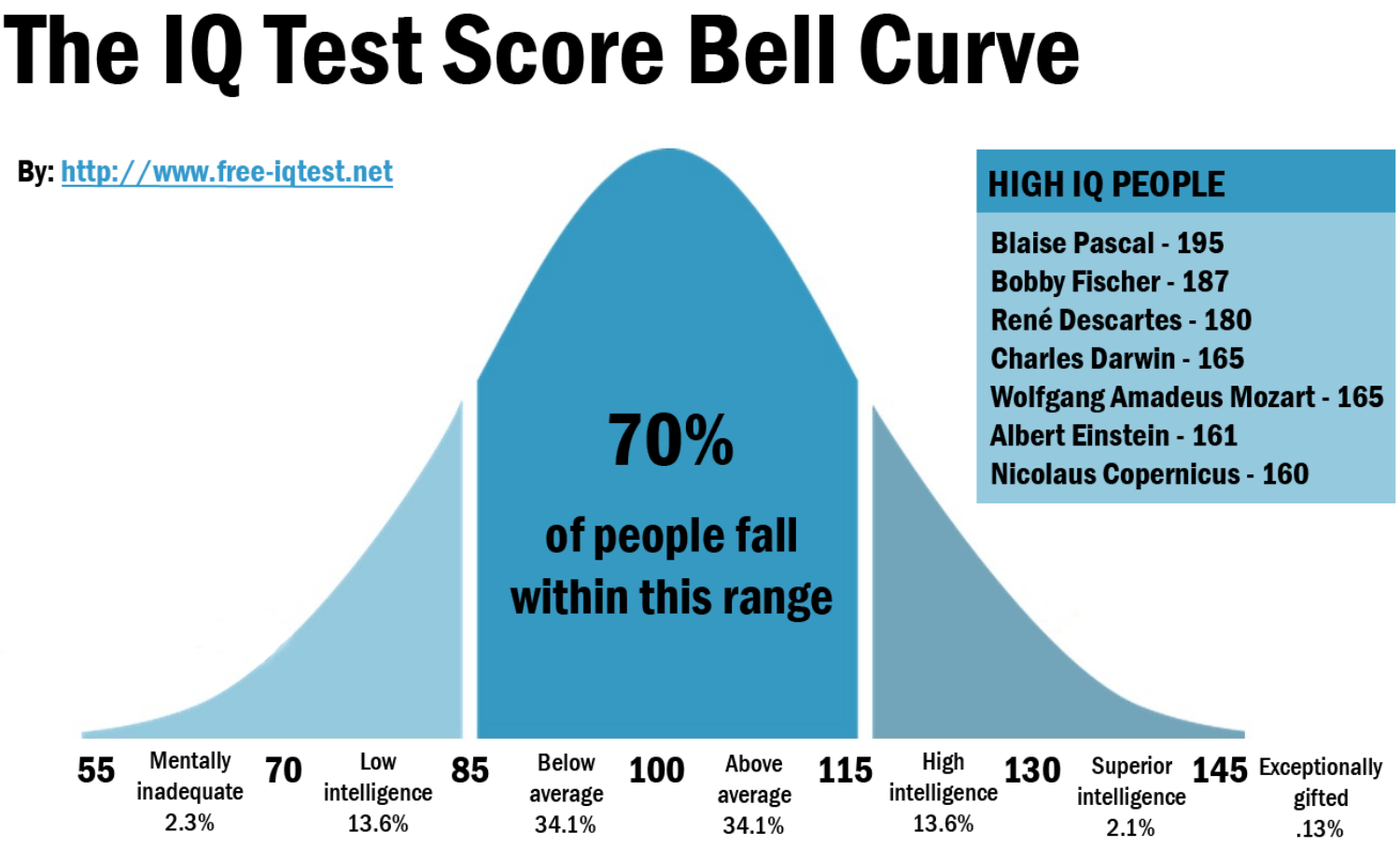
Ordinal scale – other examples



The scale is normal with sigma 15

Scales

Ordinal scale – other examples



The scale is normal with sigma 15

BUT

IQ should not be interpreted as interval scale

delta IQ 100 and IQ 130 is smaller than delta IQ 130 and IQ160

Scales

Nominal scale

- used for labeling variables, without any quantitative value

- you can call them names, labels, classes, etc.

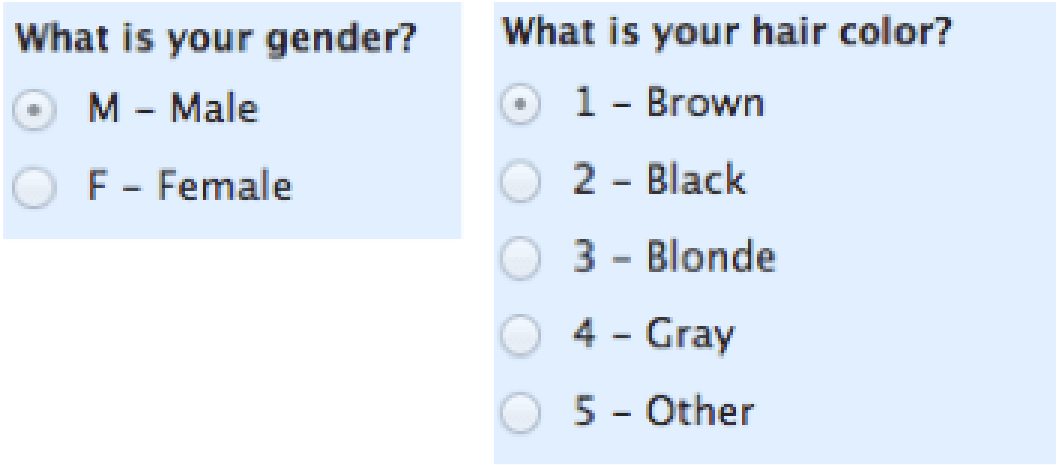
Scales

Nominal scale

- used for labeling variables, without any quantitative value

- you can call them names, labels, classes, etc.

Examples: PESEL, post codes, sex, phone numbers



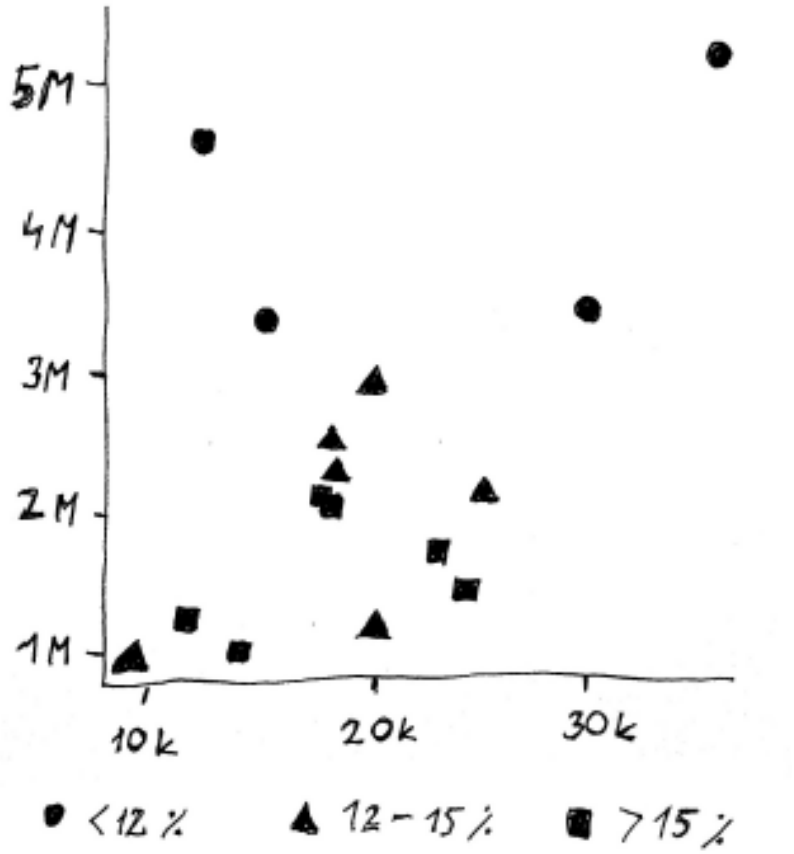
Scales - summary

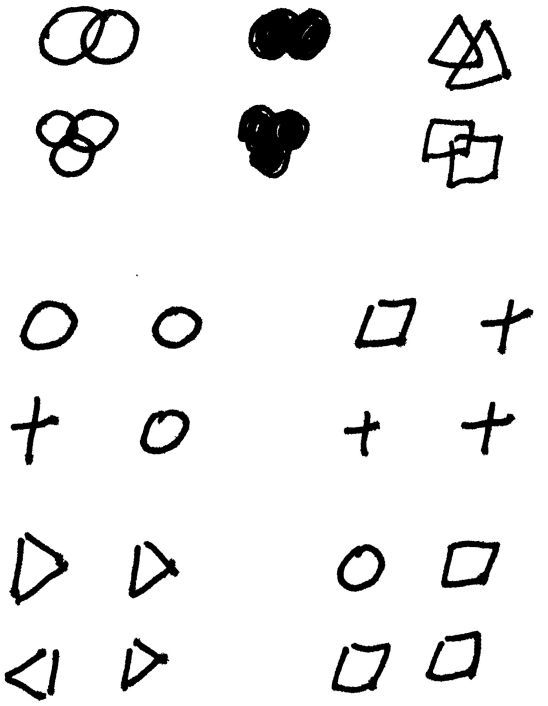
Level of Measurement	Summarize	Display	Compare
Nominal/Ordinal (Categorical Data)	Mode	Bar Chart Pie Chart	Percent distribution
Interval/Ratio (Numerical Data)	Mode Median Mean	Stem-and-Leaf Relative Frequency Table Histogram Box and Whisker Dot plot Data shape	Percent distribution Range Minimums and maximums Percentiles Standard deviation Variance

Scales - summary

Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✔	✔	✔
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Mean			✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

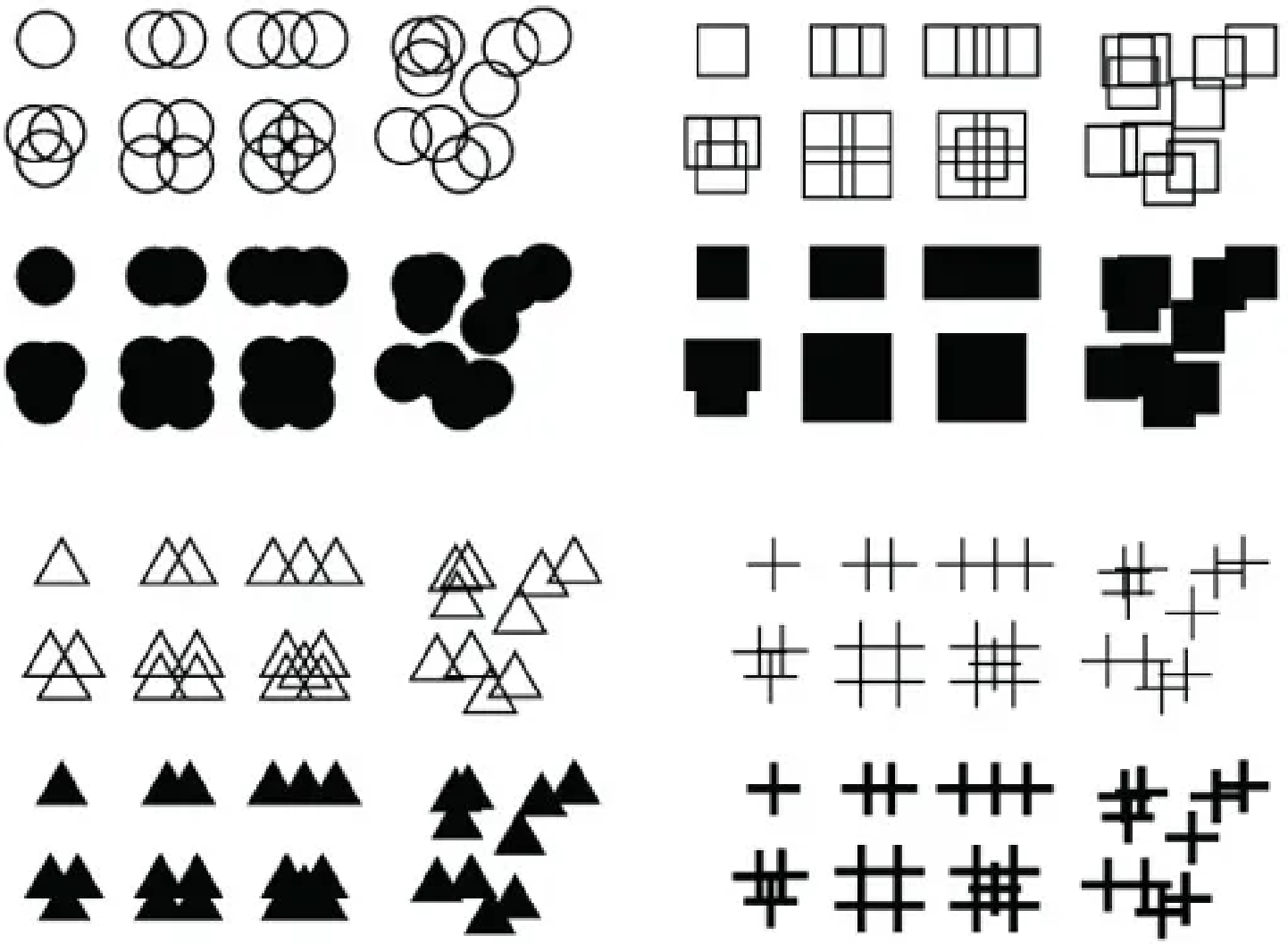
Plotting symbols





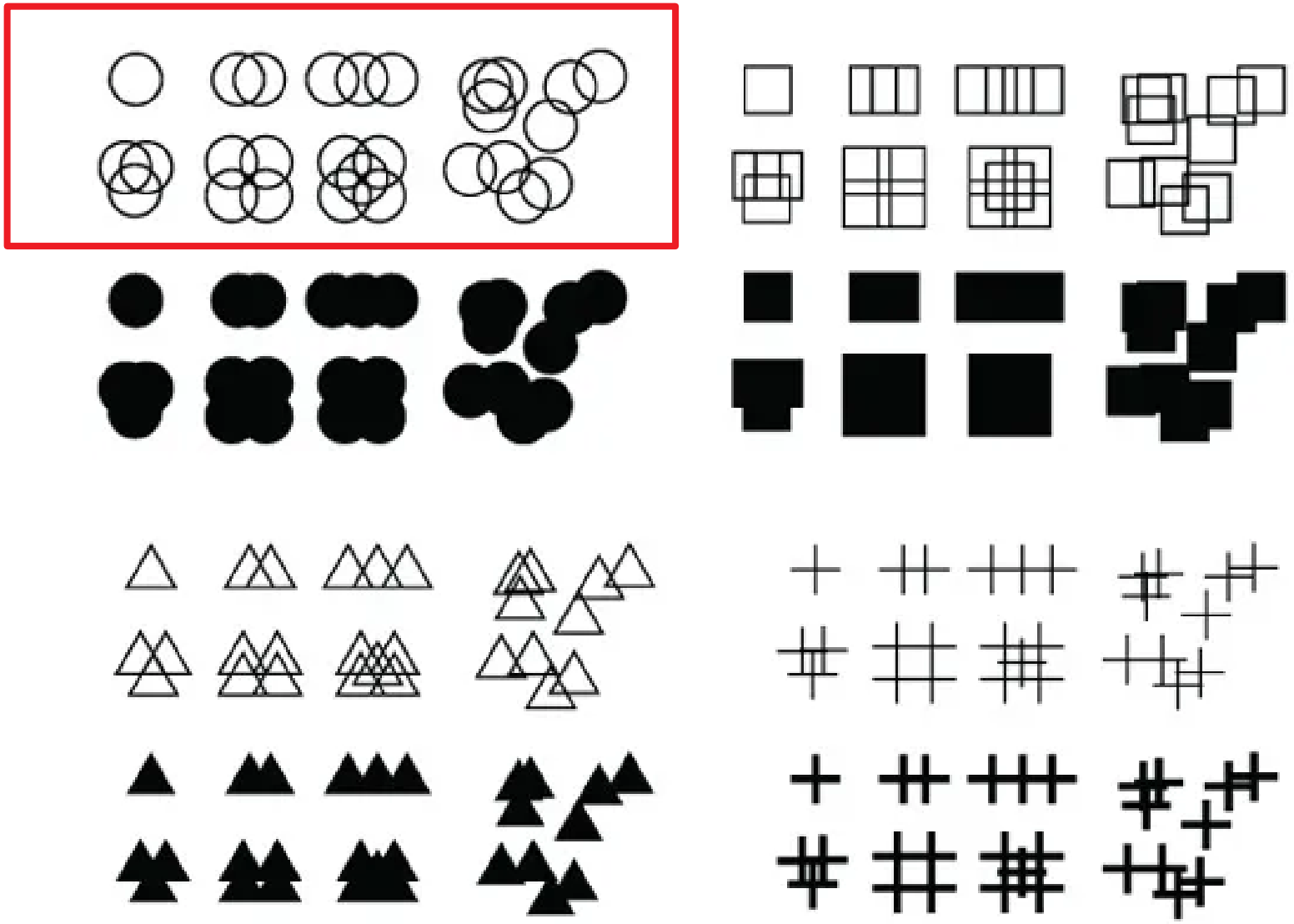
Which one is better?

Plotting symbols



10.1038/nmeth.2490

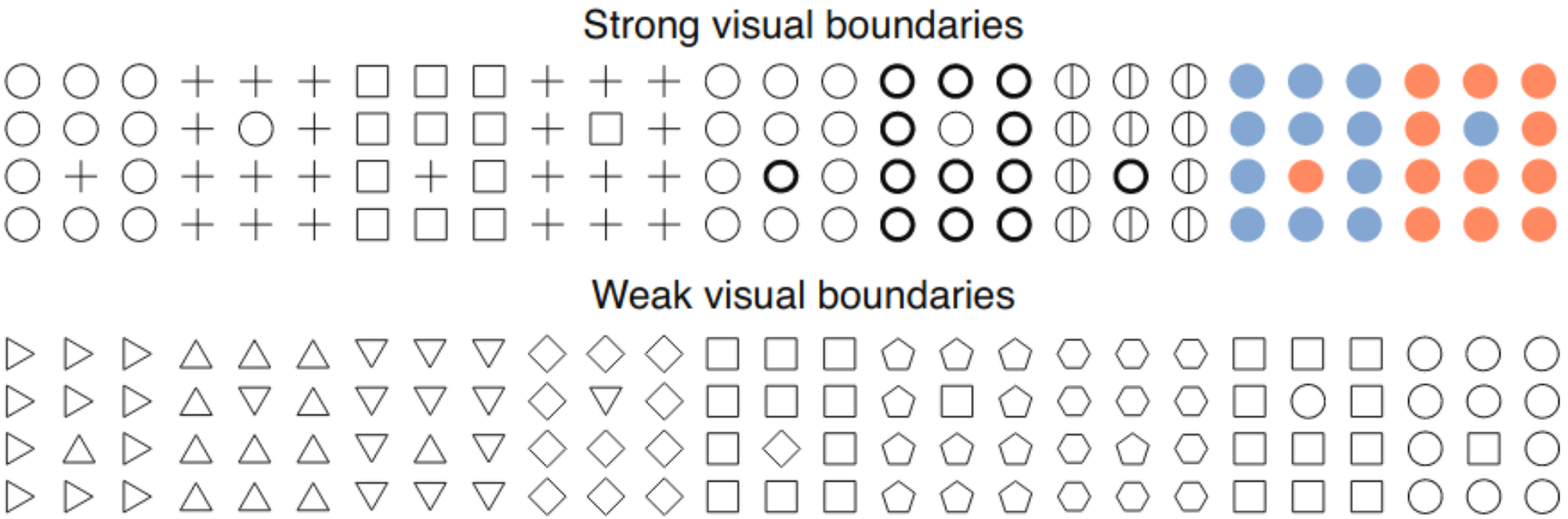
Plotting symbols



The hollow circle is a flexible and robust plotting symbol

10.1038/nmeth.2490

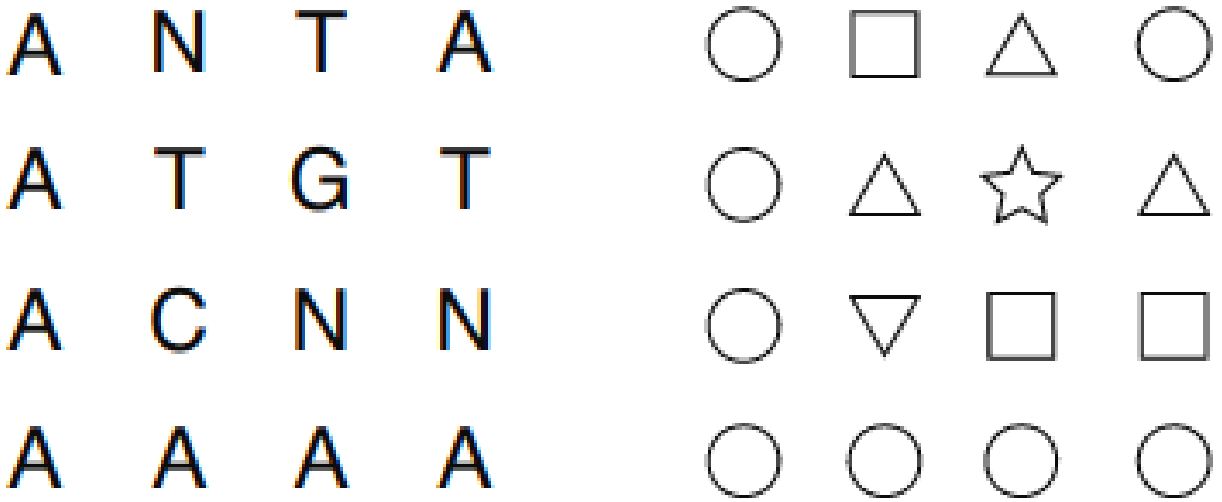
Plotting symbols



Symbols that contrast with one another make good combinations

10.1038/nmeth.2490

Plotting symbols



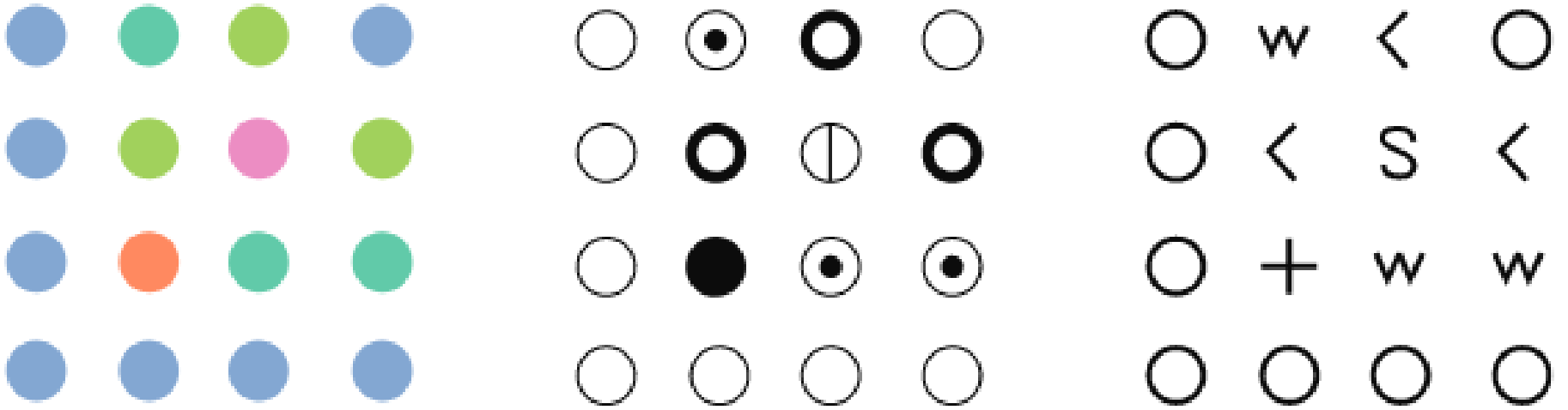
Letters simplify legend lookups,
but many appear the same
(such as C/G, B/R/P and E/F/H)

- use proper font
- do not mix upper and lower case

Shapes are powerful discriminator
but beware that, for a given width, they may appear to have different sizes owing to differences in areas

10.1038/nmeth.2490

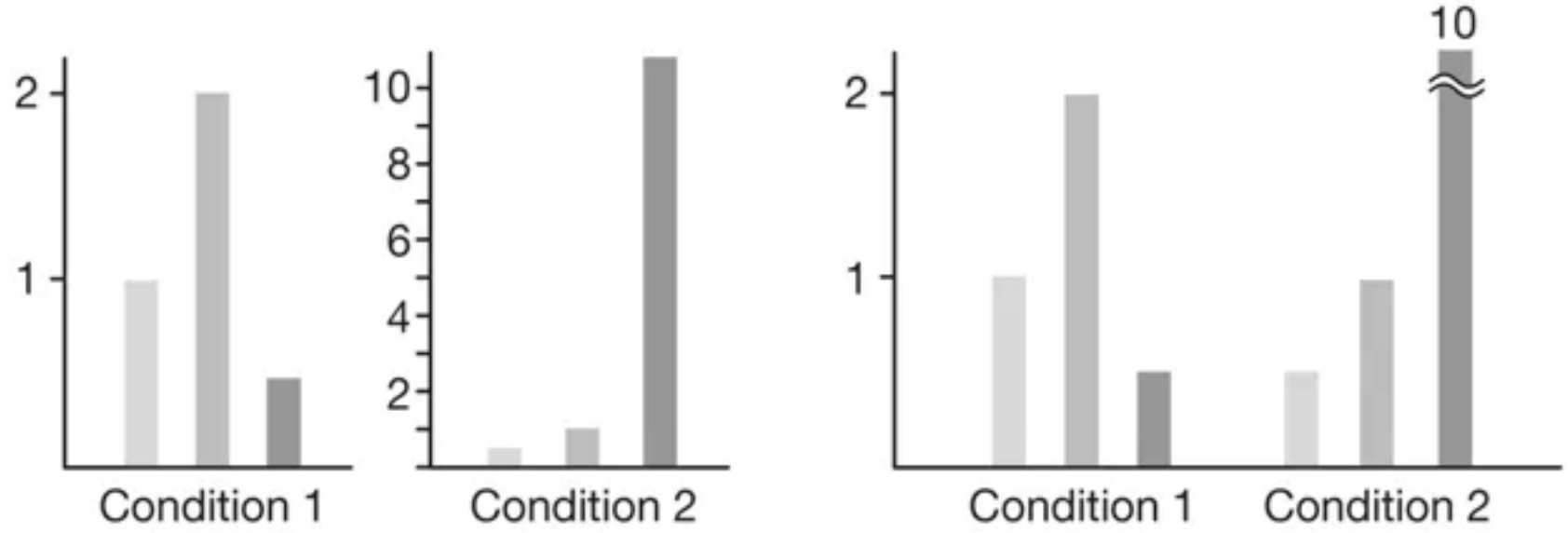
Plotting symbols



You can use colors (left) for differentiation or for black-and-white applications, vary the fills (middle) for low data densities and use texture symbols (right) when overlap is high

10.1038/nmeth.2490

Axes, ticks and grids



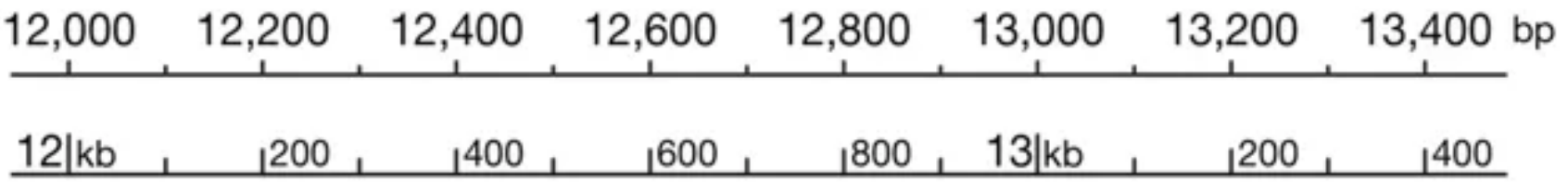
- if absolute differences are important, maintain axis scaling across panels

- draw a single y axis to emphasize that the scale is fixed

- in bar plots, use breaks to shorten outlier elements that would otherwise compress the dynamic range of the data

10.1038/nmeth.2337

Axes, ticks and grids

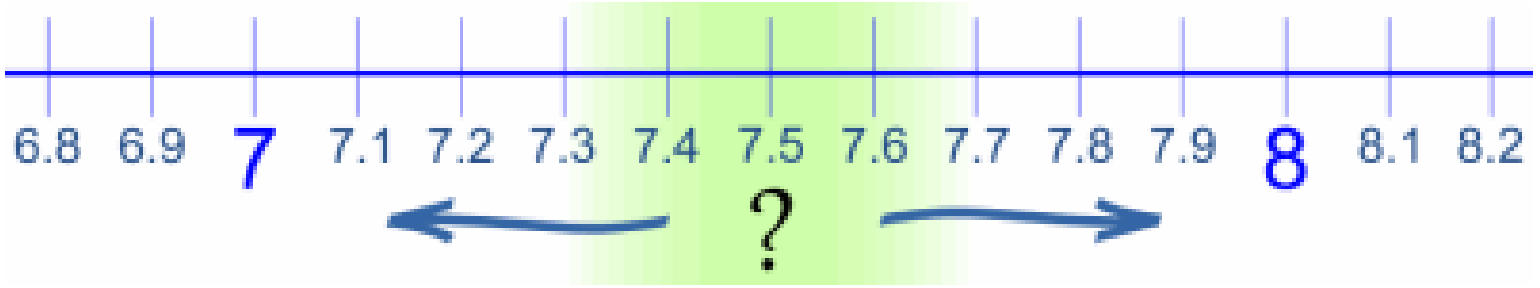


- avoid duplication of nonsignificant digits in tick marks (reduce or remove them if you can by adjusting the units)

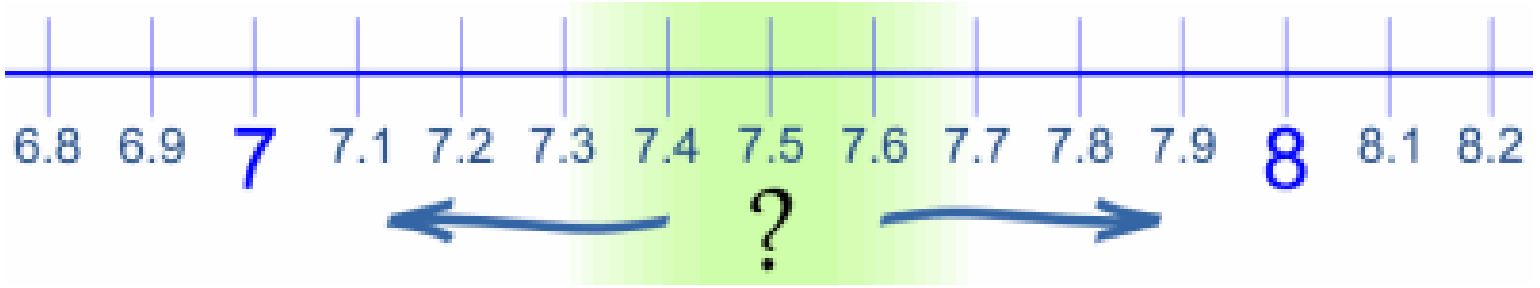
- move from to shorten e.g. 12,000 to 12kb or 1,000,000 to 1M

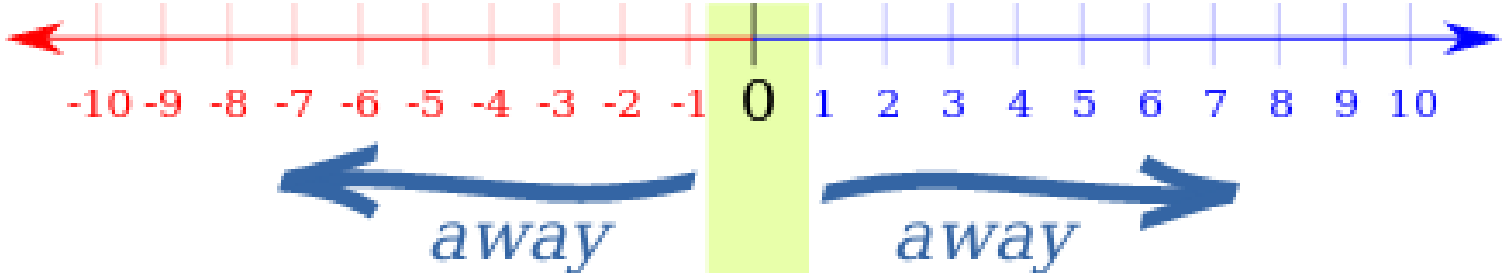
10.1038/nmeth.2337

Rounding

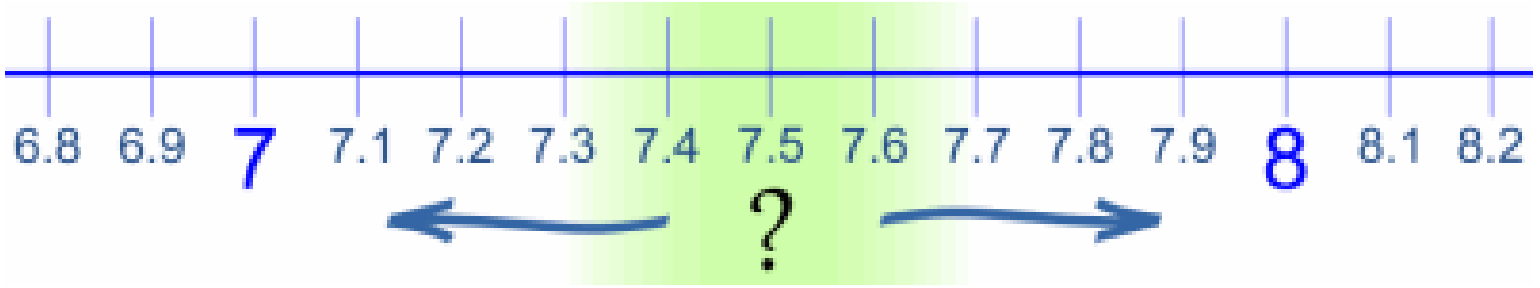


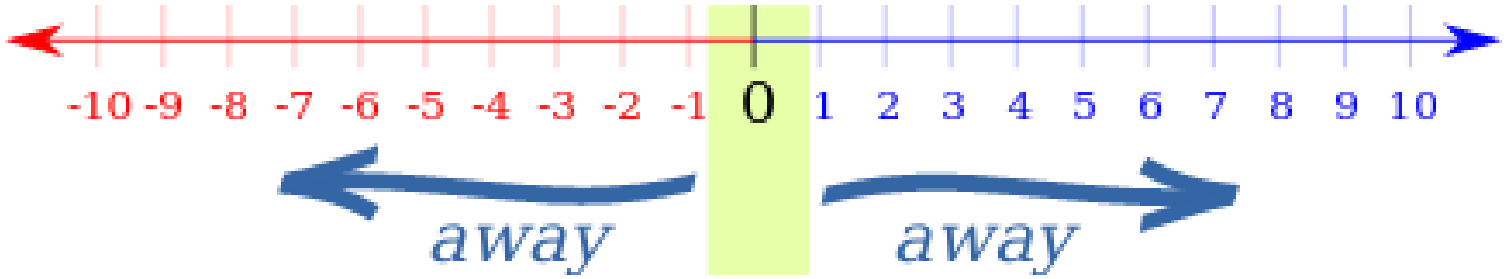
Rounding

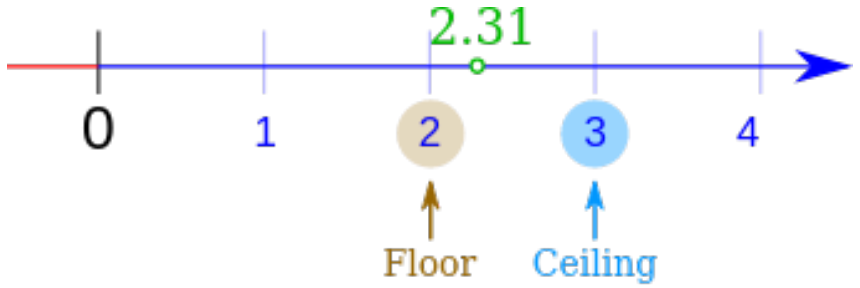

6.8 6.9 7 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 8 8.1 8.2
?


-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
away
away

Rounding


6.8 6.9 7 7.1 7.2 7.3 7.4 7.5 7.6 7.7 7.8 7.9 8 8.1 8.2
?


-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
away
away


2.31
0 1 2 3 4
Floor
Ceiling

Rounding

1443736684.0 **Population of China in Year 2021**

Rounding

1443736684.0 **Population of China in Year 2021**

1443736684 **Population of China in Year 2021**

Rounding

1443736684.0	**Population of China in Year 2021**
1443736684	**Population of China in Year 2021**
1,443,736,684	**Population of China in Year 2021**

Rounding

1443736684.0	**Population of China in Year 2021**
1443736684	**Population of China in Year 2021**
1,443,736,684	**Population of China in Year 2021**
1 443 736 684	**Population of China in Year 2021**

Rounding

1443736684.0 **Population of China in Year 2021**

1443736684 **Population of China in Year 2021**

1,443,736,684 **Population of China in Year 2021**

1 443 736 684 **Population of China in Year 2021**

Country specific
(comas, dots, spaces, ...)

Rounding

1443736684.0	**Population of China in Year 2021**
1443736684	**Population of China in Year 2021**
1,443,736,684	**Population of China in Year 2021**
1 443 736 684	**Population of China in Year 2021**
1,443 Million	**Population of China in Year 2021**

Rounding

1443736684.0	Population of China in Year 2021
1443736684	Population of China in Year 2021
1,443,736,684	Population of China in Year 2021
1 443 736 684	Population of China in Year 2021
1,443 Million	Population of China in Year 2021
1,44 Billion	Population of China in Year 2021

Rounding

1443736684.0	Population of China in Year 2021
1443736684	Population of China in Year 2021
1,443,736,684	Population of China in Year 2021
1 443 736 684	Population of China in Year 2021
1,443 Million	Population of China in Year 2021
1,44 Billion	Population of China in Year 2021
1,44 Bi	Population of China in Year 2021

Rounding

1.57845454545348412211111

Rounding

1.57845454545348412211111

2.344847398437943894794243

2.784353534543

3

1.432328948593543

Rounding

1.57845454545348412211111	1.58
2.34484739843794389479424 3	2.34
2.784353534543	2.78
3	3.00
1.432328948593543	1.43

Rounding

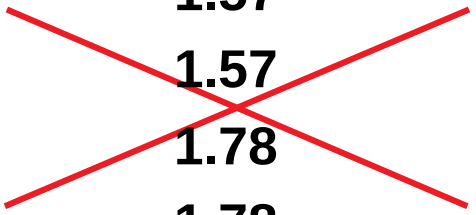
1.57845454545348412211111	1.58	1.6
2.34484739843794389479424 3	2.34	2.3
2.784353534543	2.78	2.8
3	3.00	3.0
1.432328948593543	1.43	1.4

Rounding

1.57845454545348412211111	1.57
1.57484739843794389479424 3	1.57
1.784353534543	1.78
1.7842328948593543	1.78

Rounding

1.57845454545348412211111	1.57
1.57484739843794389479424 3	1.57
1.784353534543	1.78
1.7842328948593543	1.78



Rounding

1.57845454545348412211111

1.57484739843794389479424 3

1.784353534543

1.784232894859 3543

1.578 45454545348412211111

1.574 84739843794389479424 3

1.784353534543

1.784232894859 3543

Rounding

1.57845454545348412211111	1.578 45454545348412211111
1.57484739843794389479424 3	1.574 84739843794389479424 3
1.784353534543	1.7843 53534543
1.784232894859 3543	1.7842 32894859 3543

Rounding

1.57845454545348412211111	1.57845454545348412211111
1.57484739843794389479424 3	1.57484739843794389479424 3
1.784353534543	1.784353534543
1.784232894859 3543	1.784232894859 3543

Rounding

1.57845454545348412211111

1.57484739843794389479424 3

1.784353534543

1.784232894859 3543

1.5784 5454545434841221111 1

1.5748 4739843794389479424 3

1.7843 53534543

1.7842 328948593543

Key question: does this make sense to use the given precision?

Rounding

1.57845454545348412211111 ± 0.2334322324323

1.57484739843794389479424 3 ± 0.2734322324323

1.784353534543 ± 0.1934322324323

1.7842328948593543 ± 0.4134322324323

Rounding

1.5784 ± 0.2334322324323

1.5748473984379438947942 43 ± 0.2734322324323

1.784353534543 ± 0.1934322324323

1.7842328948593543 ± 0.4134322324323

1.5784 ± 0.2334322324323

1.5748 ± 0.2734322324323

1.7843 ± 0.1934322324323

1.7842 ± 0.4134322324323

Rounding

1.5784 ± 0.2334322324323

1.57484739843794389479424 3 ± 0.2734322324323

1.784353534543 ± 0.1934322324323

1.7842328948593543 ± 0.4134322324323

1.5784 ± 0.2334

1.5748 ± 0.2734

1.7843 ± 0.1934

1.7842 ± 0.4134

Rounding

1.5784 ± 0.2334322324323

1.57484739843794389479424 3 ± 0.2734322324323

1.784353534543 ± 0.1934322324323

1.7842328948593543 ± 0.4134322324323

1.5784 ± 0.2334

1.5748 ± 0.2734

1.7843 ± 0.1934

1.7842 ± 0.4134

1.57 ± 0.23

1.57 ± 0.27

1.78 ± 0.19

1.78 ± 0.41

Rounding

1.5784 ± 0.2334322324323

1.57484739843794389479424 3 ± 0.2734322324323

1.784353534543 ± 0.1934322324323

1.7842328948593543 ± 0.4134322324323

1.5784 ± 0.2334

1.5748 ± 0.2734

1.7843 ± 0.1934

1.7842 ± 0.4134

1.57 ± 0.23

1.57 ± 0.27

1.78 ± 0.19

1.78 ± 0.41

1.6 ± 0.2

1.6 ± 0.3

1.8 ± 0.2

1.8 ± 0.4

Rounding

1.5784 ± 0.2334322324323

1.57484739843794389479424 3 ± 0.2734322324323

1.784353534543 ± 0.1934322324323

1.7842328948593543 ± 0.4134322324323

1.5784 ± 0.2334

1.5748 ± 0.2734

1.7843 ± 0.1934

1.7842 ± 0.4134

1.57 ± 0.23

1.57 ± 0.27

1.78 ± 0.19

1.78 ± 0.41

1.6 ± 0.2

1.6 ± 0.3

1.8 ± 0.2

1.8 ± 0.4

Rounding

1.57845454545348412211111

1.57484739843794389479424 3

1.784353534543

1.784232894859 3543

1.5784 5454545434841221111 1

1.5748 4739843794389479424 3

1.7843 53534543

1.7842 328948593543

Key question: does this make sense to use the given precision?

Rounding

1.57845454545348412211111

1.57484739843794389479424 3

1.784353534543

1.784232894859 3543

1.5784 54545454348412211111

1.5748 47398437943894794243

1.7843 53534543

1.7842 328948593543

Key question: does this make sense to use the given precision?

Consider example: 1.57845454545348412211111 (that should be presented as 1.5784) is theoretical calculation (e.g. prediction based deep learning model), but … this is only prediction of some natural phenomen that due to the technique we use we can measure with 0.2 precision.

Rounding

1.57845454545348412211111

1.57484739843794389479424 3

1.784353534543

1.784232894859 3543

1.5784 5454545434841221111 1

1.5748 4739843794389479424 3

1.7843 53534543

1.7842 328948593543

Key question: does this make sense to use the given precision?

1.6 ± 0.2

1.6 ± 0.2

1.8 ± 0.2

1.8 ± 0.2

Consider example: 1.57845454545348412211111 (that should be presented as 1.5784) is theoretical calculation (e.g. prediction based deep learning model), but … this is only prediction of some natural phenomen that due to the technique we use we can measure with 0.2 precision.

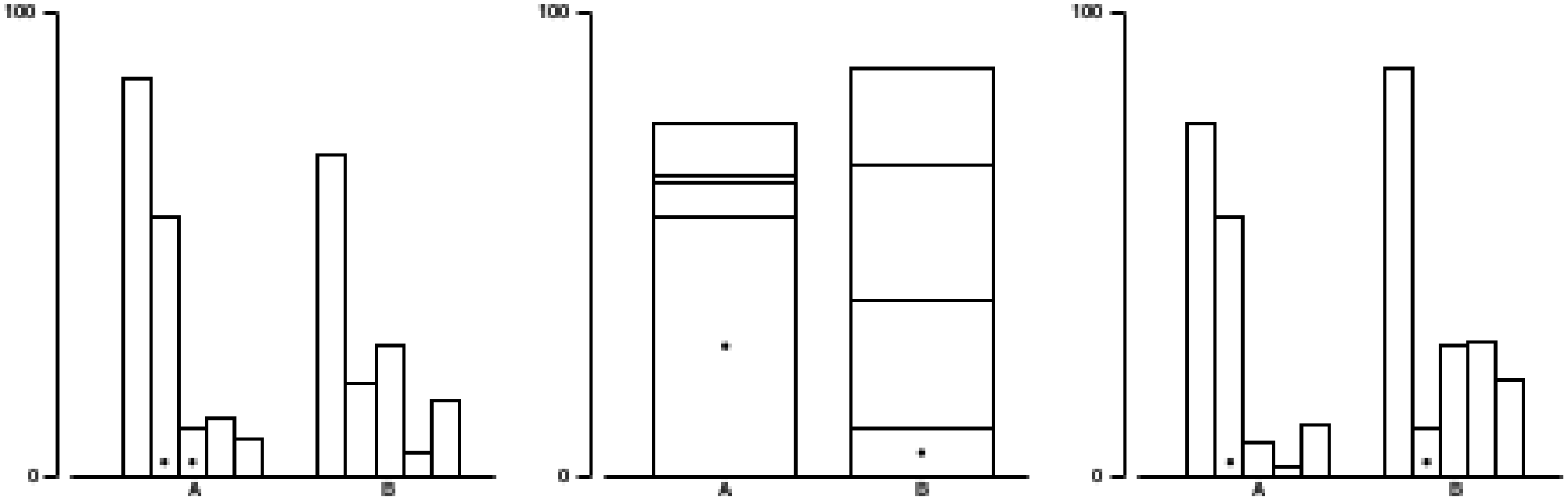


Figure 1: Stimuli for judgment tasks T1, T2 & T3. Subjects estimated percent differences between elements.

10.1145/1753326.1753357

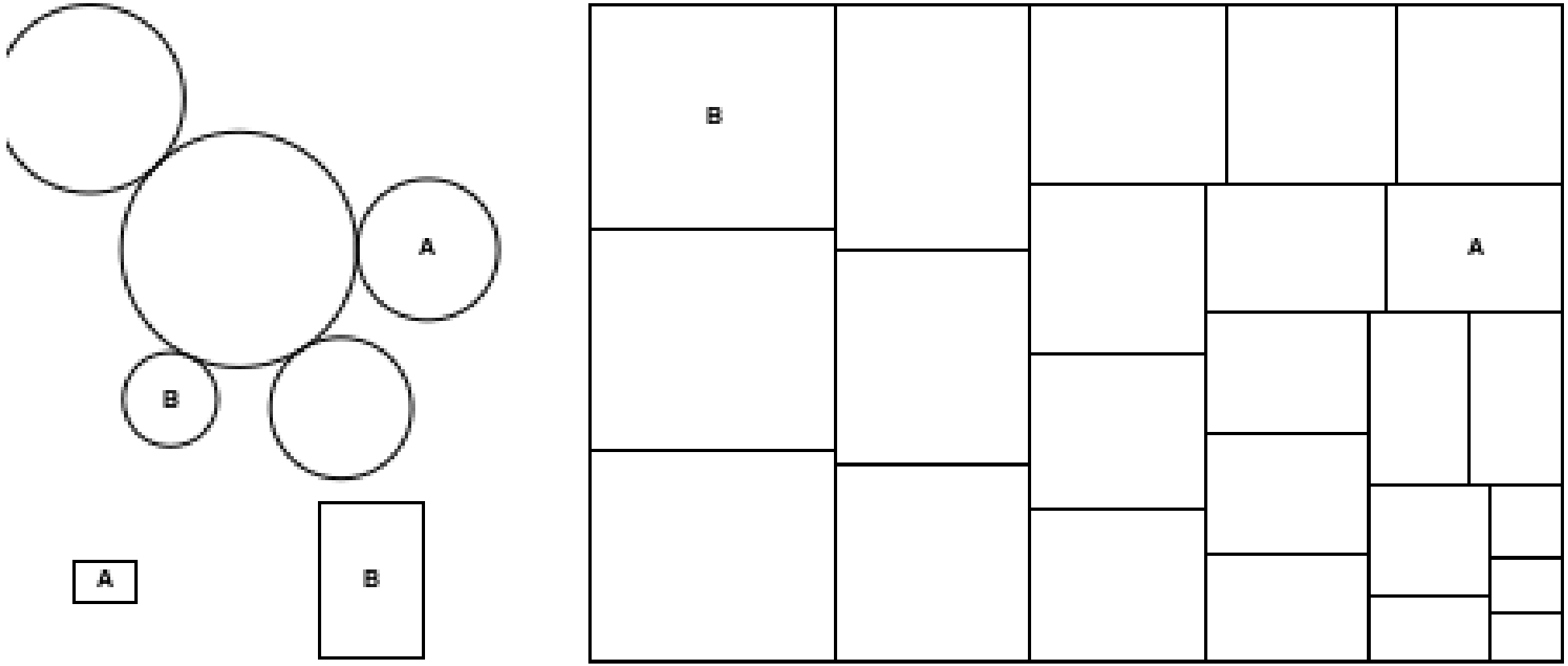
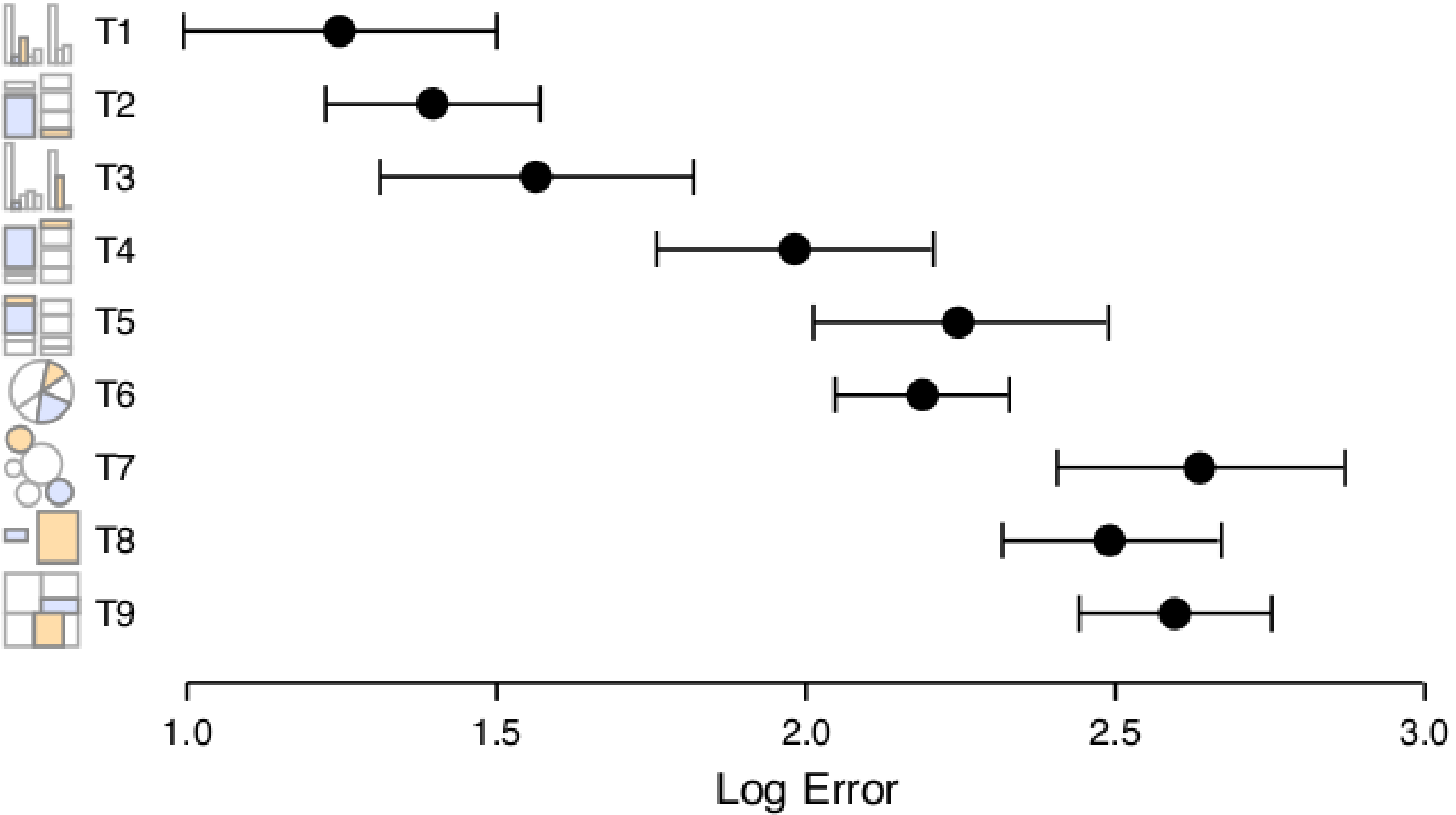


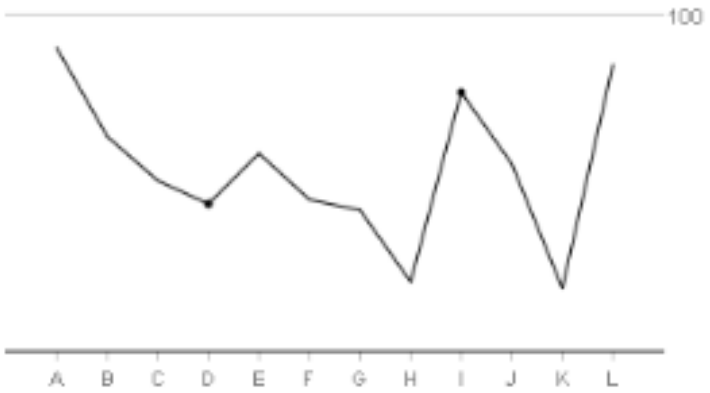
Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).

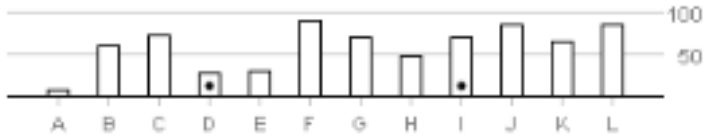
10.1145/1753326.1753357

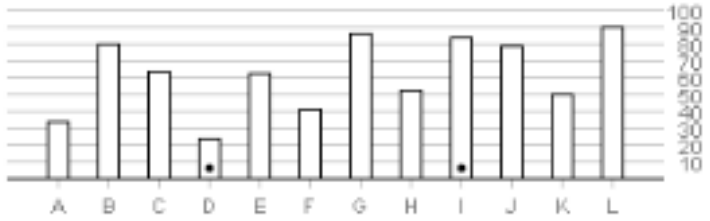


10.1145/1753326.1753357

Stimuli varying chart type, chart height, and gridline spacing

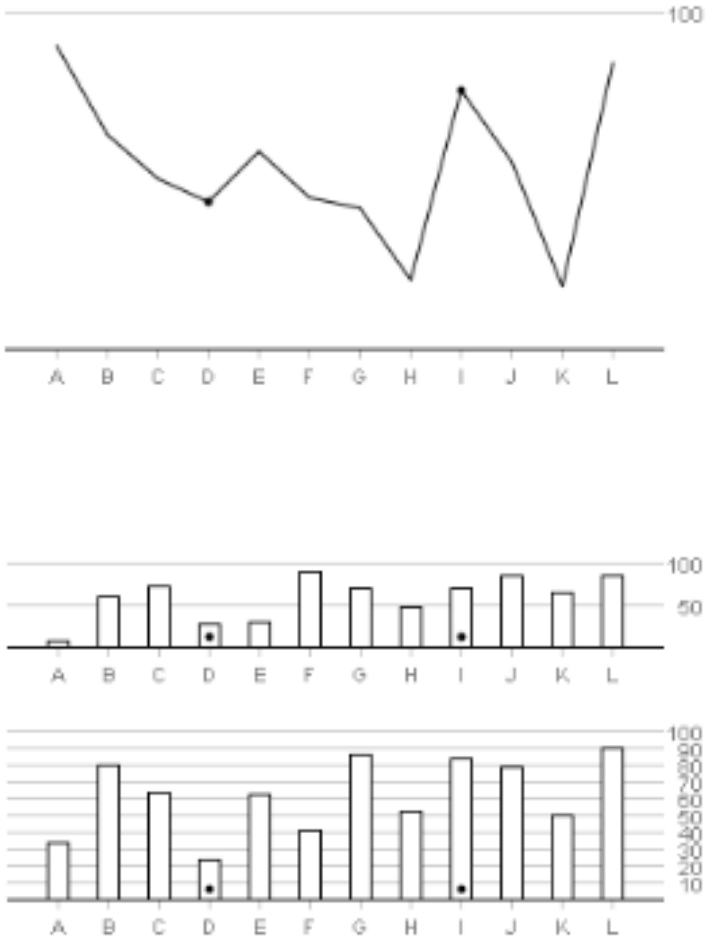


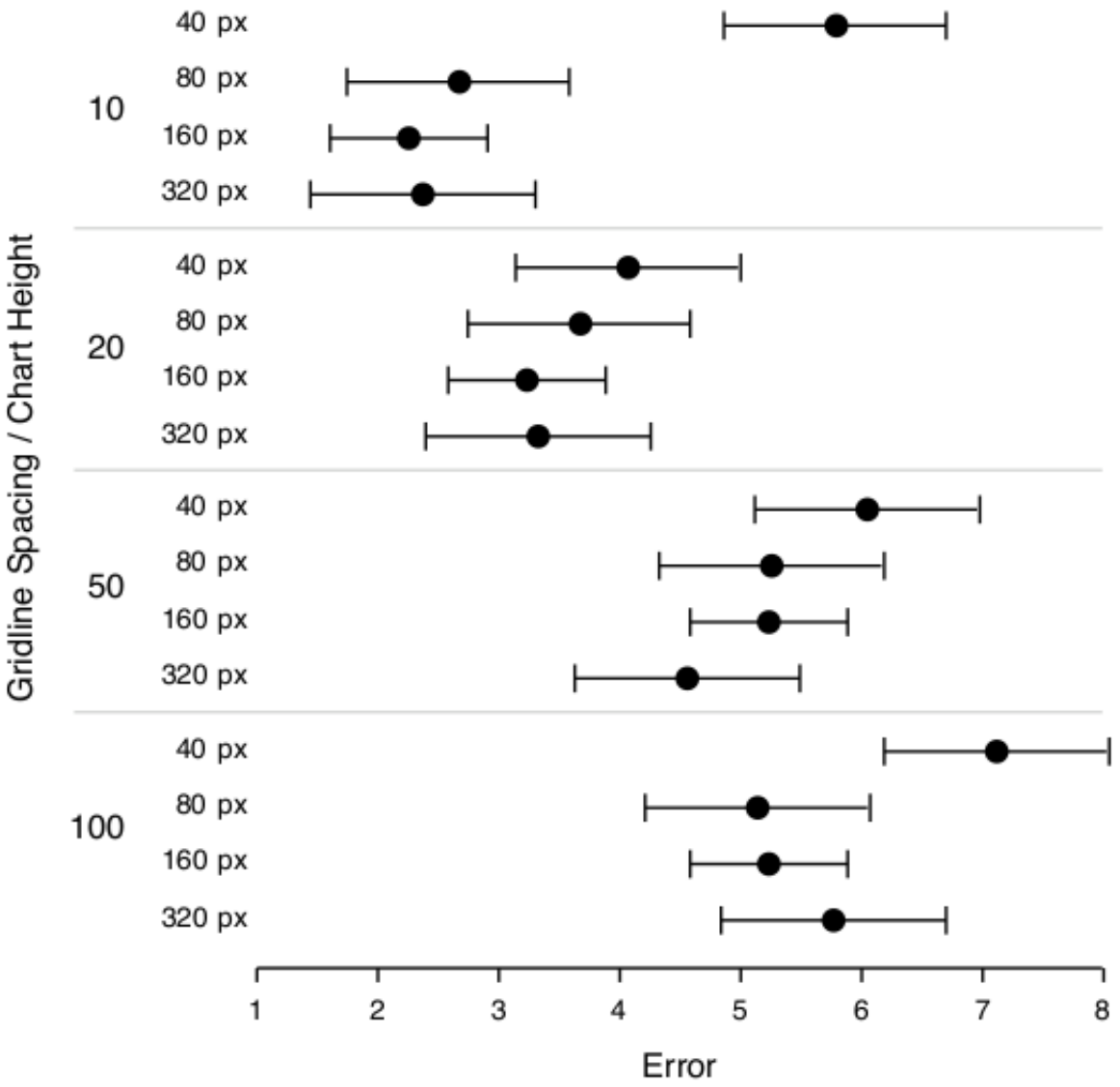




10.1145/1753326.1753357

Stimuli varying chart type, chart height, and gridline spacing





10.1145/1753326.1753357

Thank you for your time
and
See you at the next lecture

Any other
questions & comments

l.kozlowski@mimuw.edu.pl