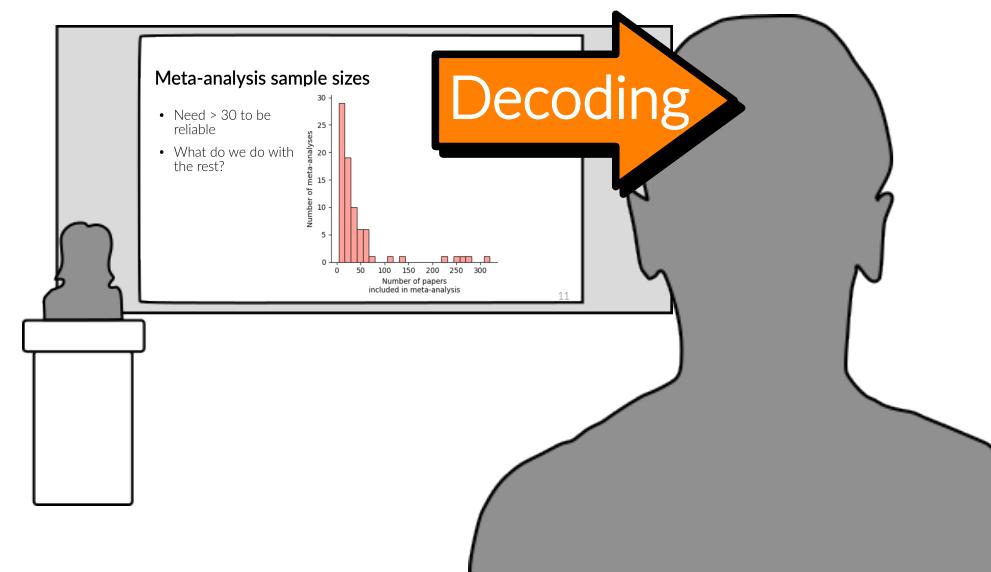


Introduction to Data Visualization

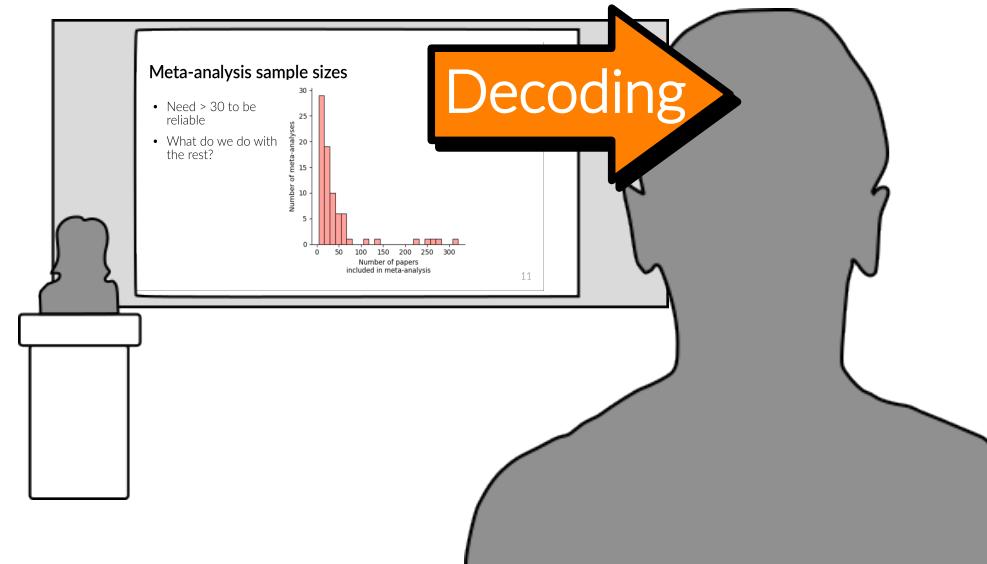
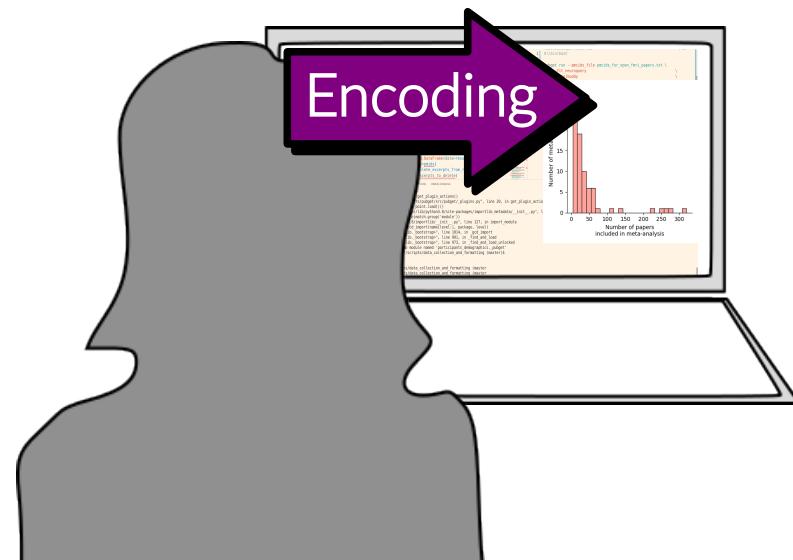
Part 1: Decoding

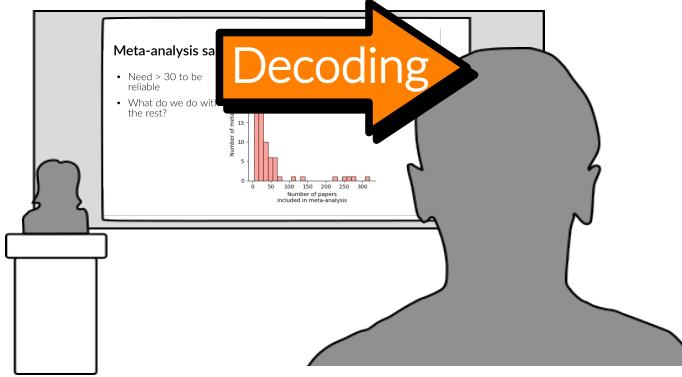
Kendra Oudyk



Goal

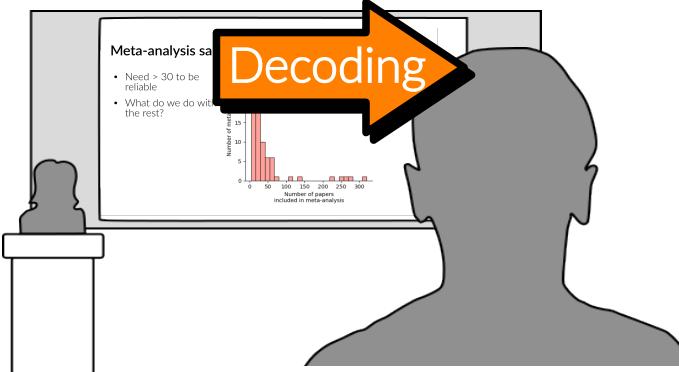
Use principles of visual **encoding** and **decoding**
to **efficiently** create visualizations
that are **effective** and **reproducible**





To plan an **effective visualization**, we need to think about

- **Message**
 - **What** we want to communicate
- **Perception**
 - **How best** to communicate it
- **Conventions**
 - **How** it's usually communicated
- **Context**
 - **Where** it will be seen

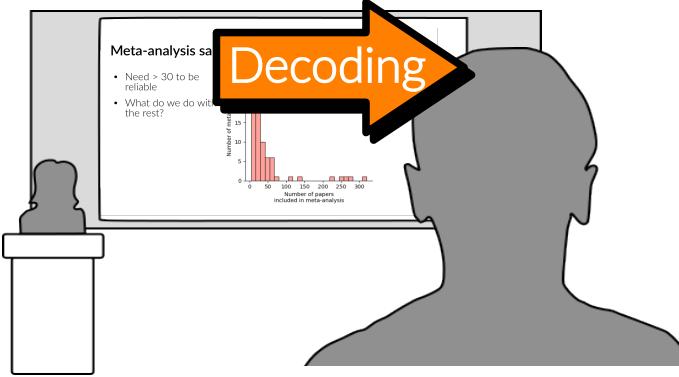


To plan an **effective visualization**, we need to think about

- **Message**
 - What we want to communicate
- **Perception**
 - How best to communicate it
- **Conventions**
 - How it's usually communicated
- **Context**
 - Where it will be seen

Message

- Raw data has no message
- Abstract it
 - “There are more males than females in science”
--> A difference between magnitudes
 - “These brain areas activate together”
--> A grouping / pattern

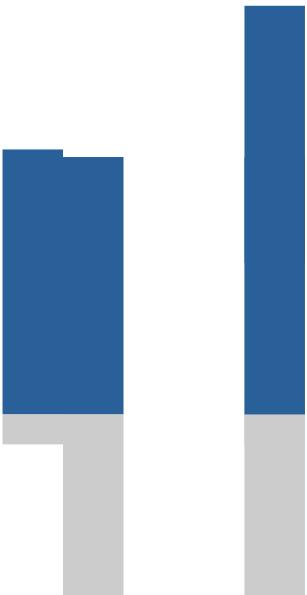


Decoding

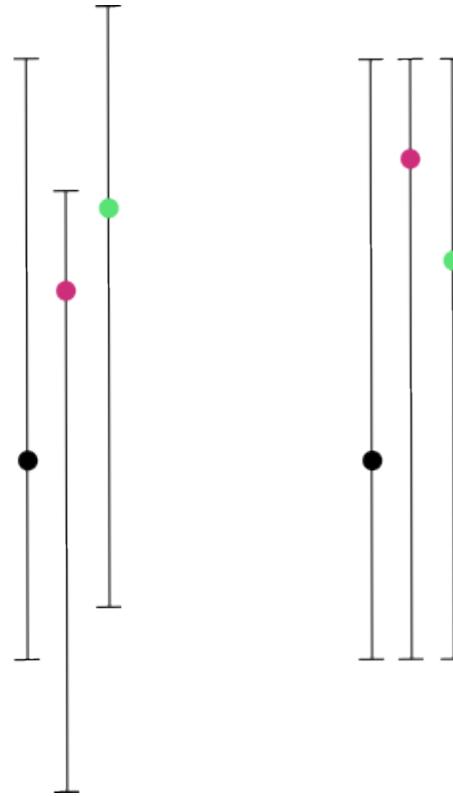
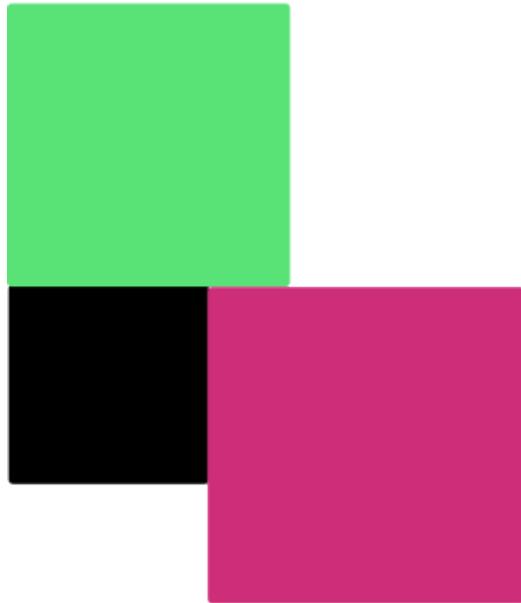
To plan an **effective visualization**, we need to think about

- **Message**
 - What we want to communicate
- **Perception**
 - How best to communicate it
- **Conventions**
 - How it's usually communicated
- **Context**
 - Where it will be seen

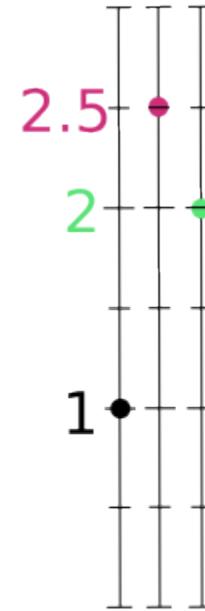
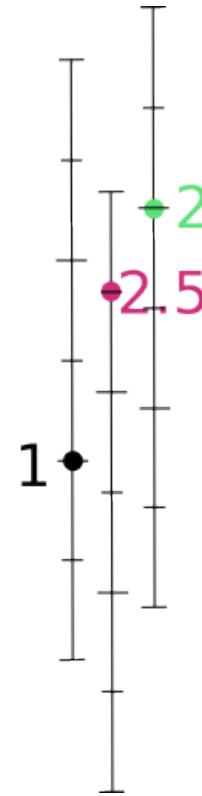
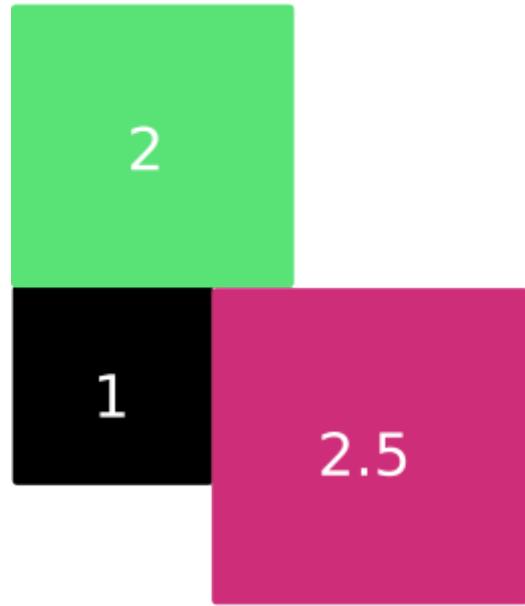
Which blue bar is bigger?



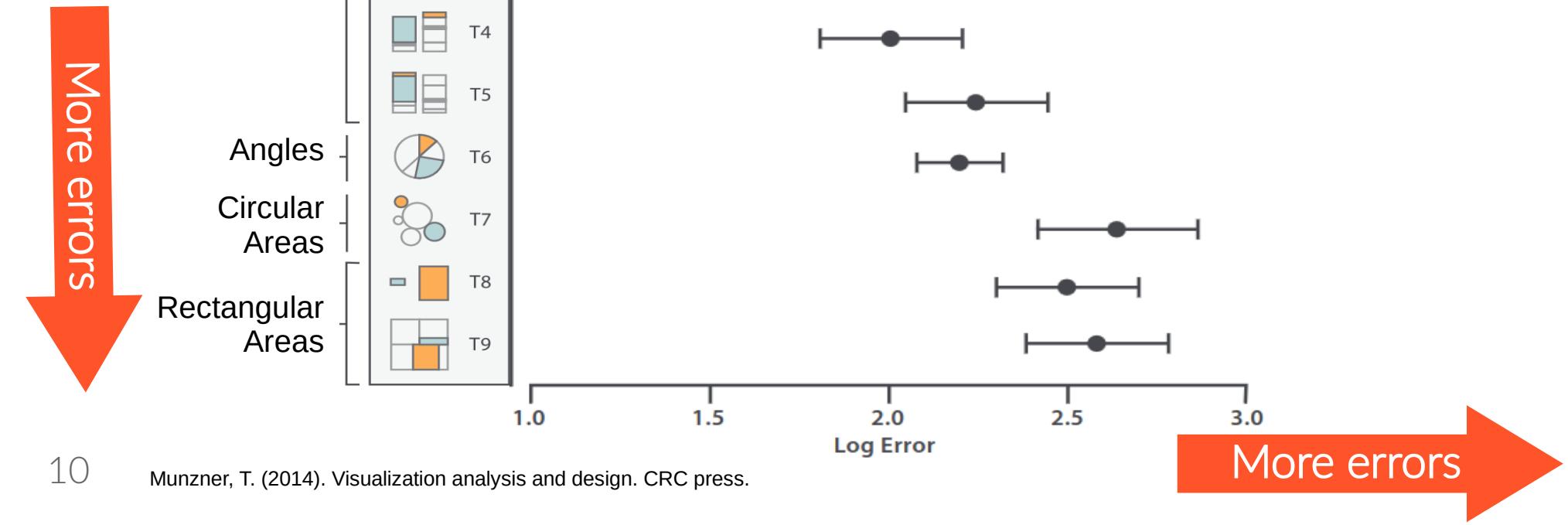
Which is 2x black, green or pink?



Which is 2x black, green or pink?



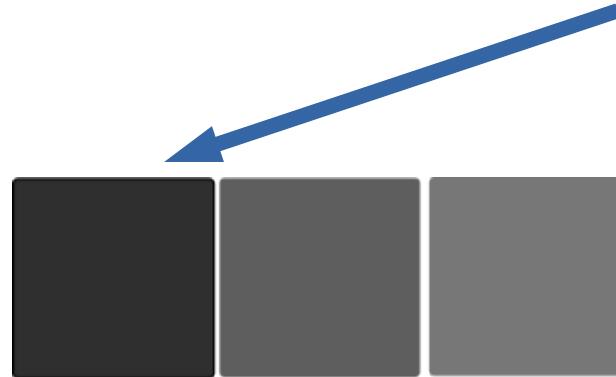
We're better at judging aligned positions



Decoding values vs. patterns

- “There are more males than females in science”
--> A difference between **magnitudes**
- “These brain areas activate together”
--> A grouping / **pattern**

Which is 2x lighter than the leftmost box?

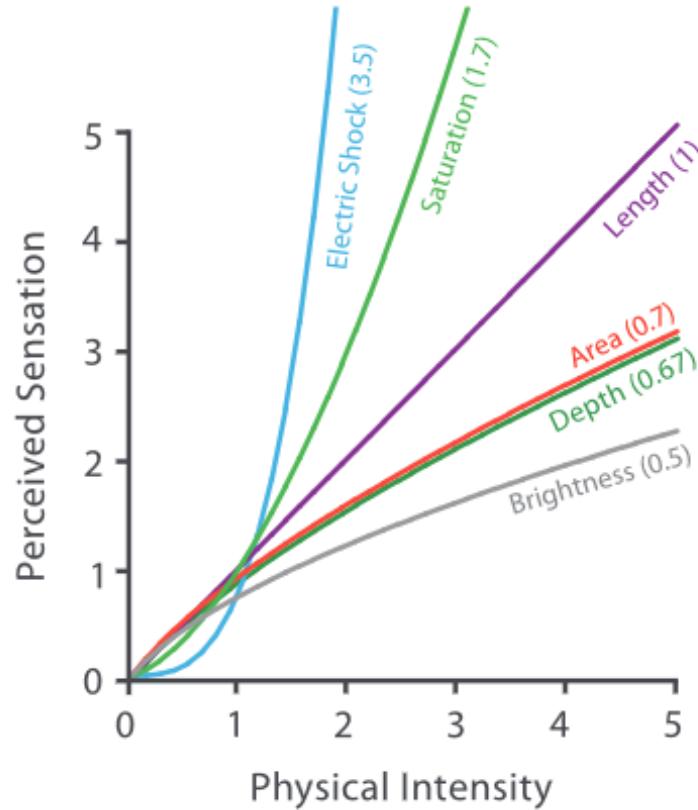


Which is 2x lighter than the leftmost box?



Physical intensity vs perceived intensity

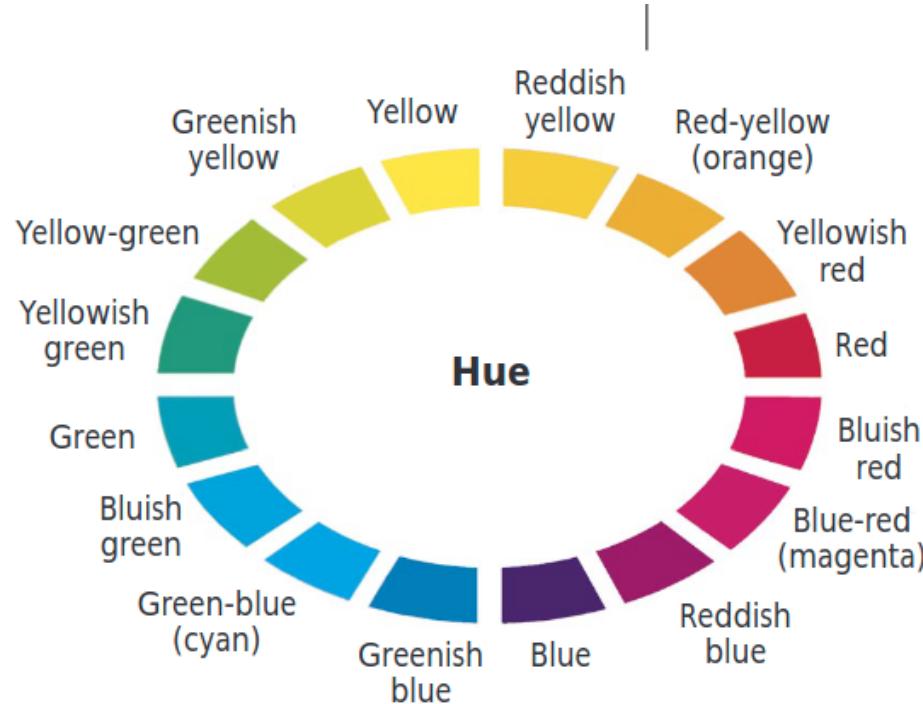
Steven's Psychophysical Power Law: $S = I^N$



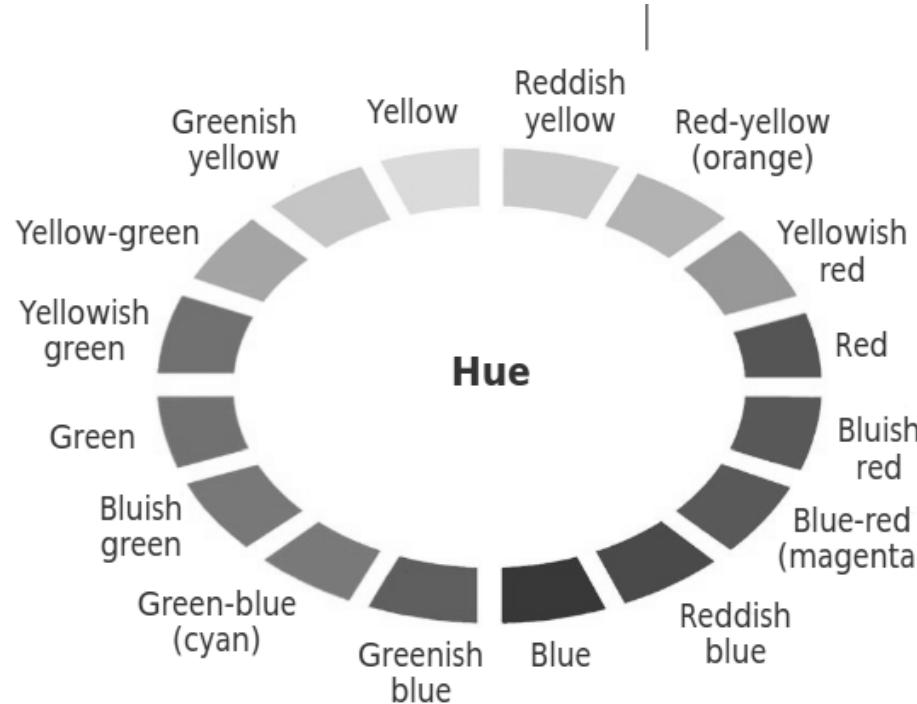
Color: salient but complicated



Hue: how we talk about color



Luminance: an important subconscious cue



Which is
lightest?



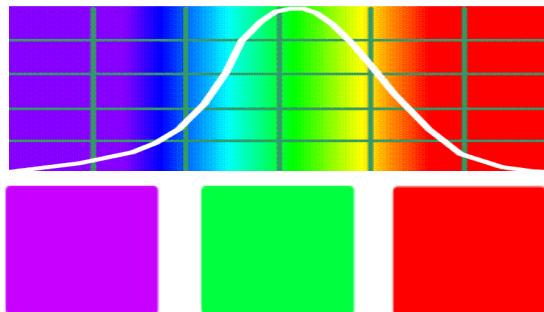


Physical
Lightness



			
Physical Lightness	50	50	50
Perceptual Lightness	51	88	53

Spectral sensitivity to luminance (lightness)



Physical
Lightness

50	50	50
----	----	----

Perceptual
Lightness

51	88	53
----	----	----

Spectral sensitivity to luminance (lightness)

HSL

(physically accurate)



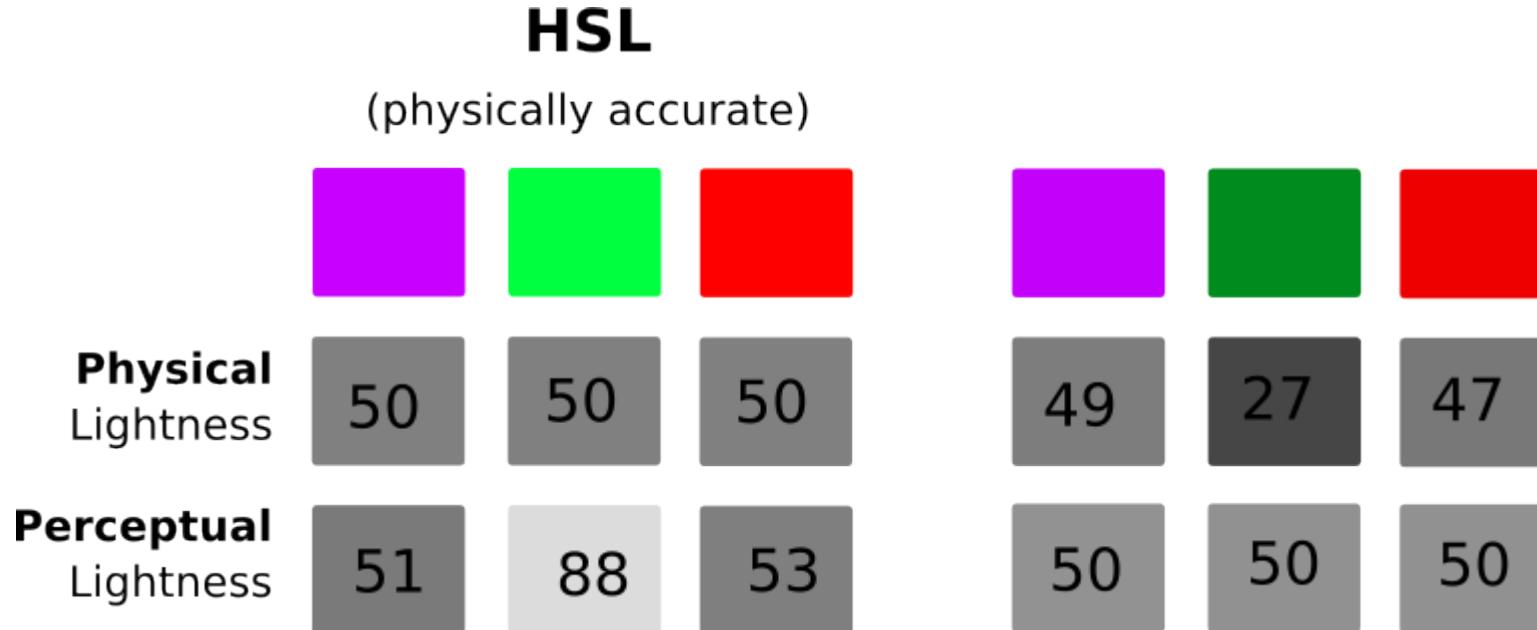
Physical
Lightness

50	50	50
----	----	----

Perceptual
Lightness

51	88	53
----	----	----

Spectral sensitivity to luminance (lightness)

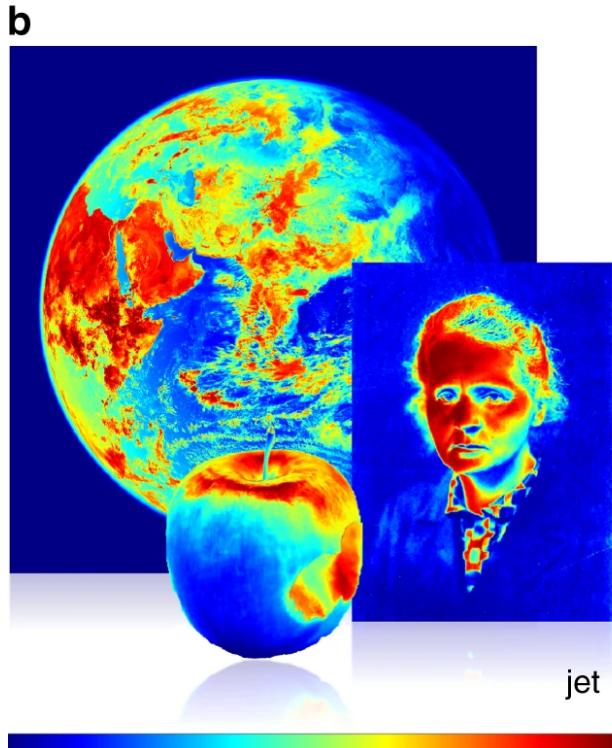


Spectral sensitivity to luminance (lightness)

	HSL (physically accurate)			H*S*L* (psycho physically accurate)		
						
Physical Lightness	50	50	50	49	27	47
Perceptual Lightness	51	88	53	50	50	50



Named colors don't work well for ordered values

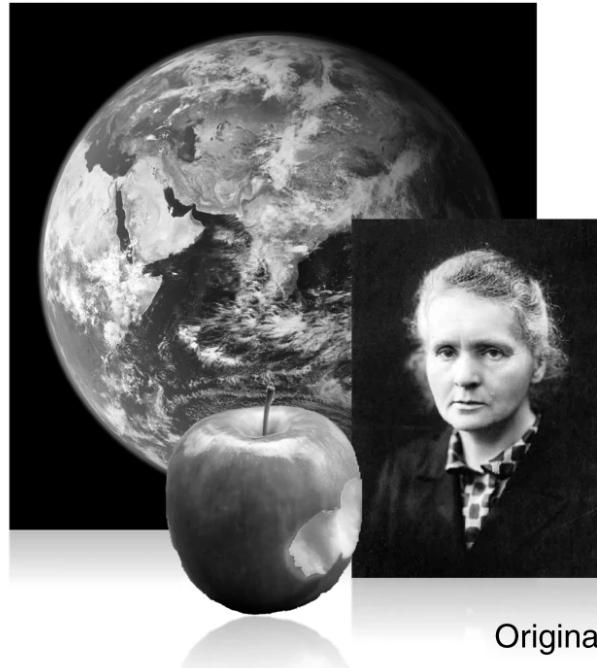
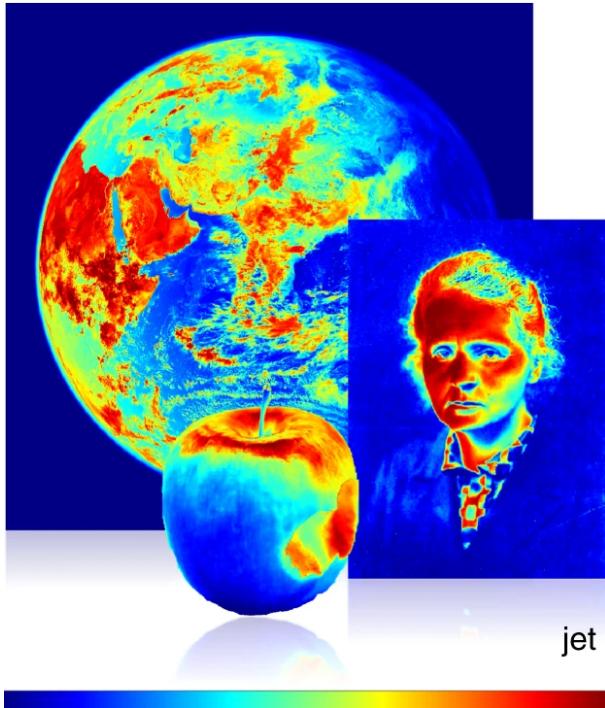


Named colors don't work well for ordered values



Certain colormaps do work

b



c



Certain colormaps do work

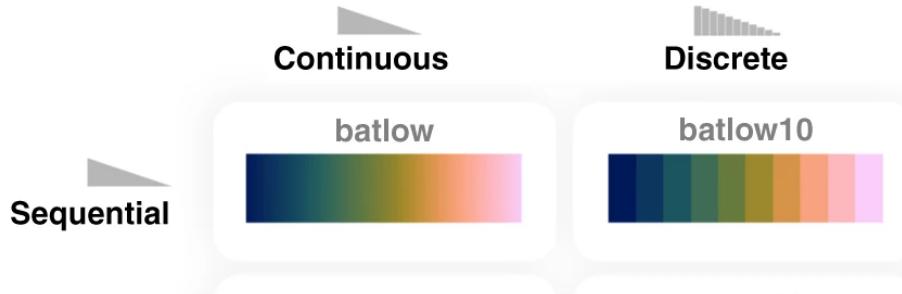
b



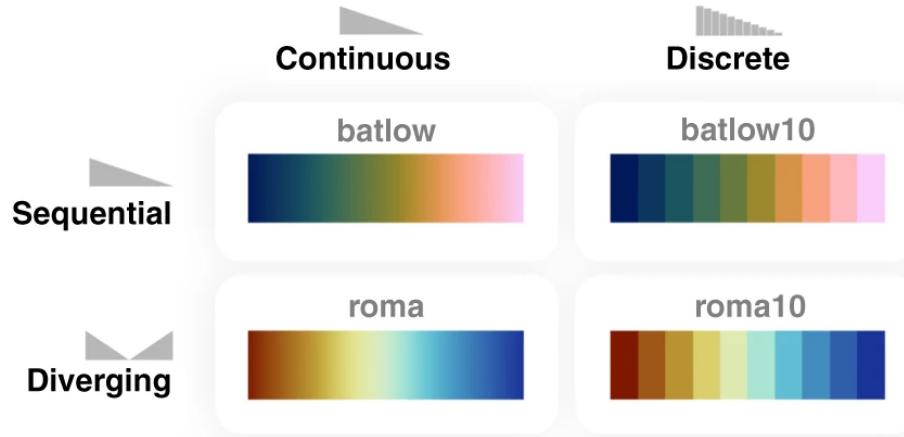
c



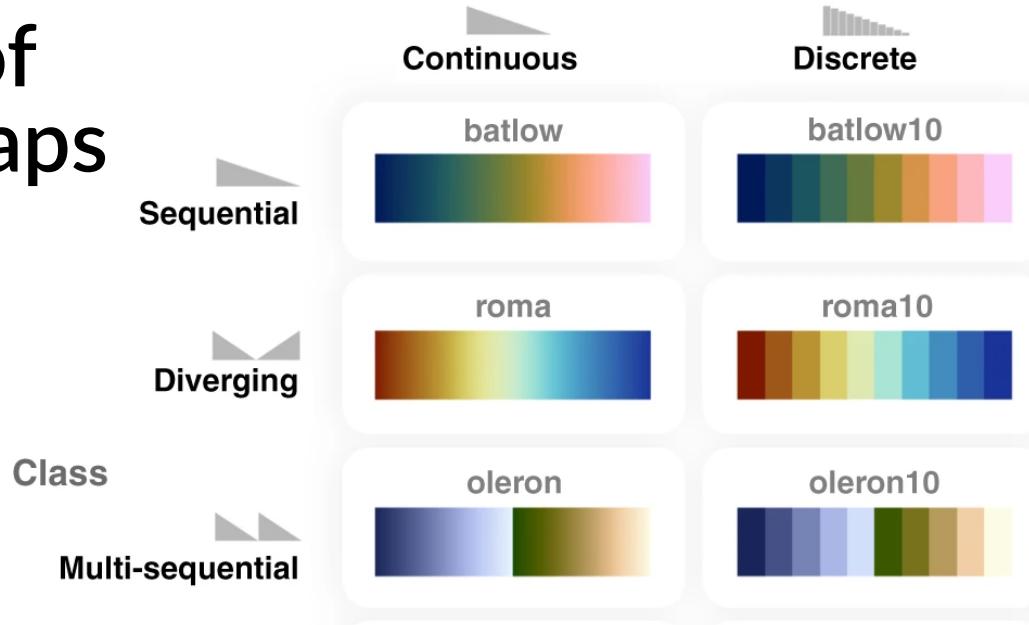
Types of colormaps



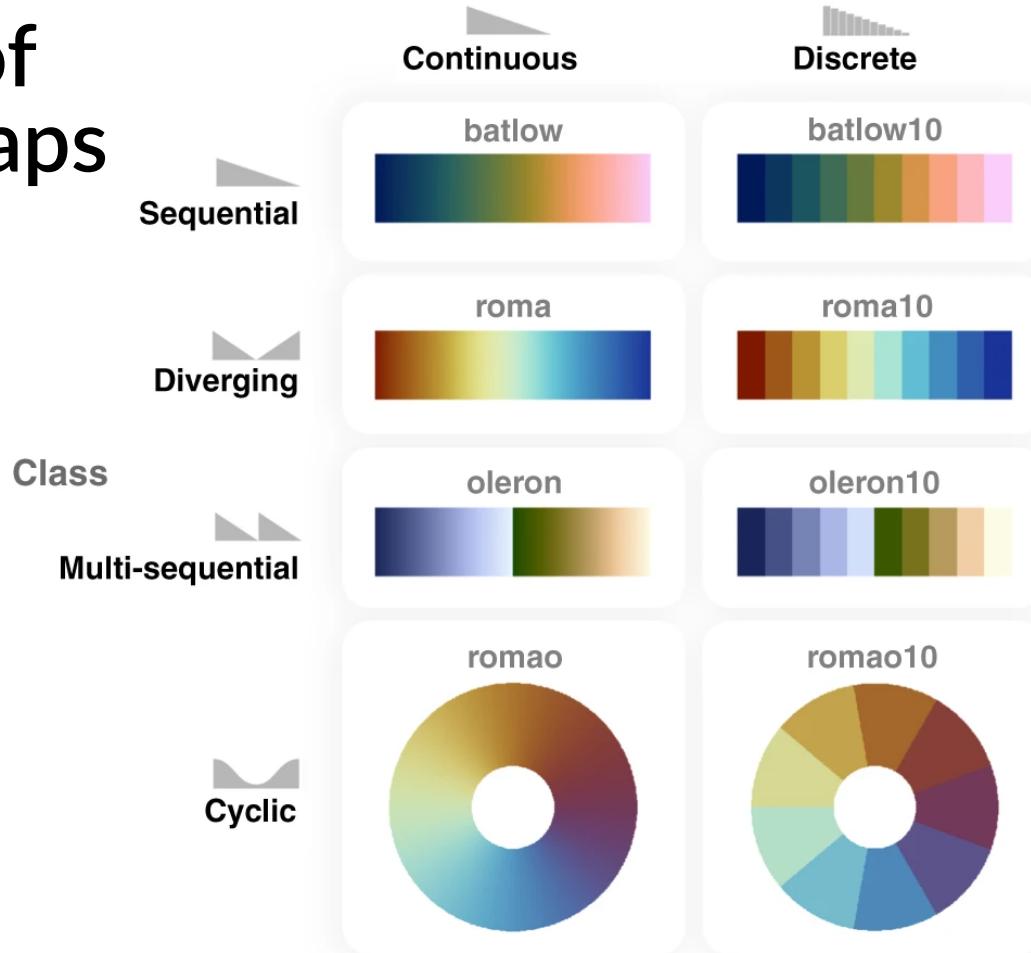
Types of colormaps



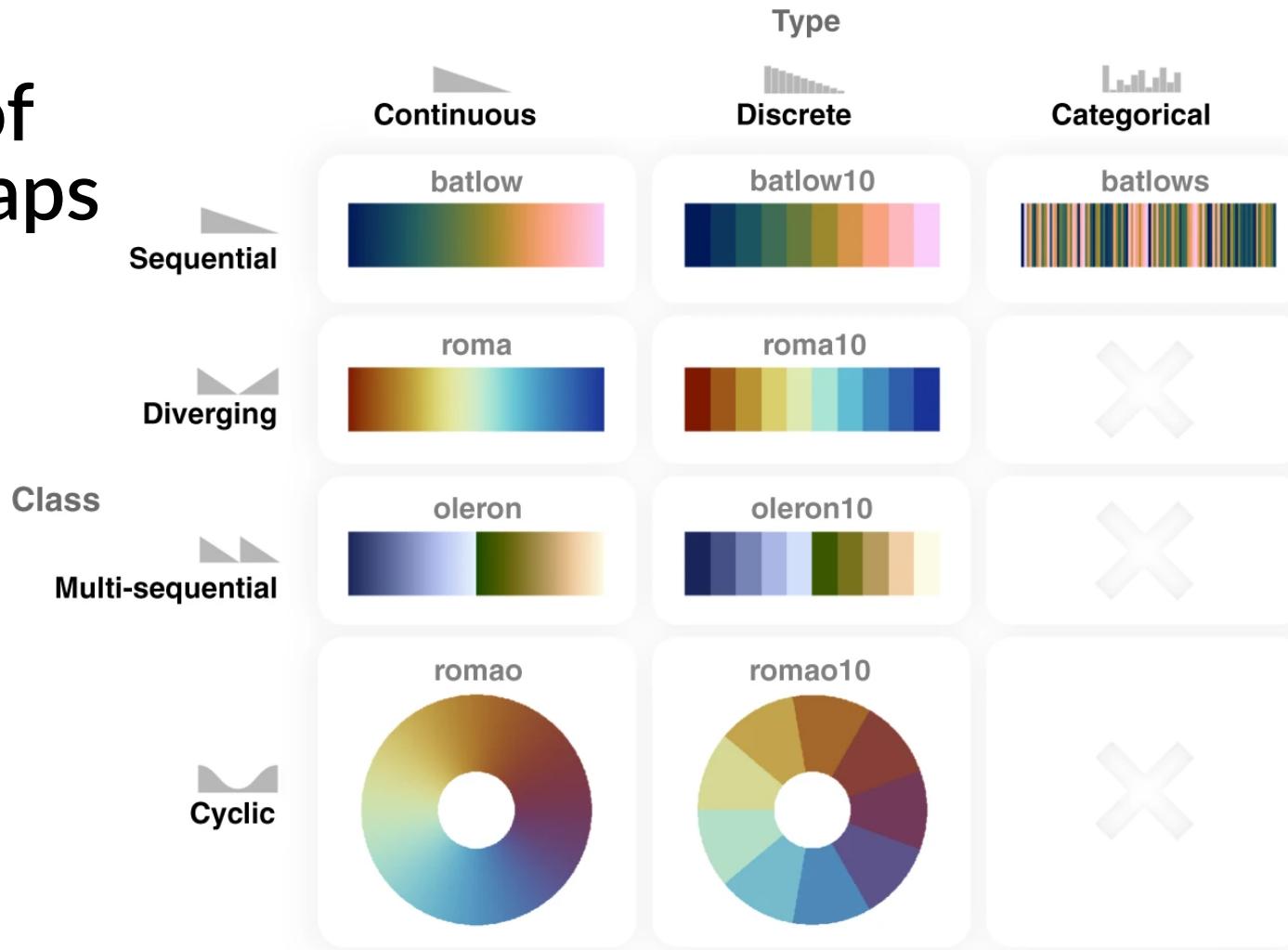
Types of colormaps



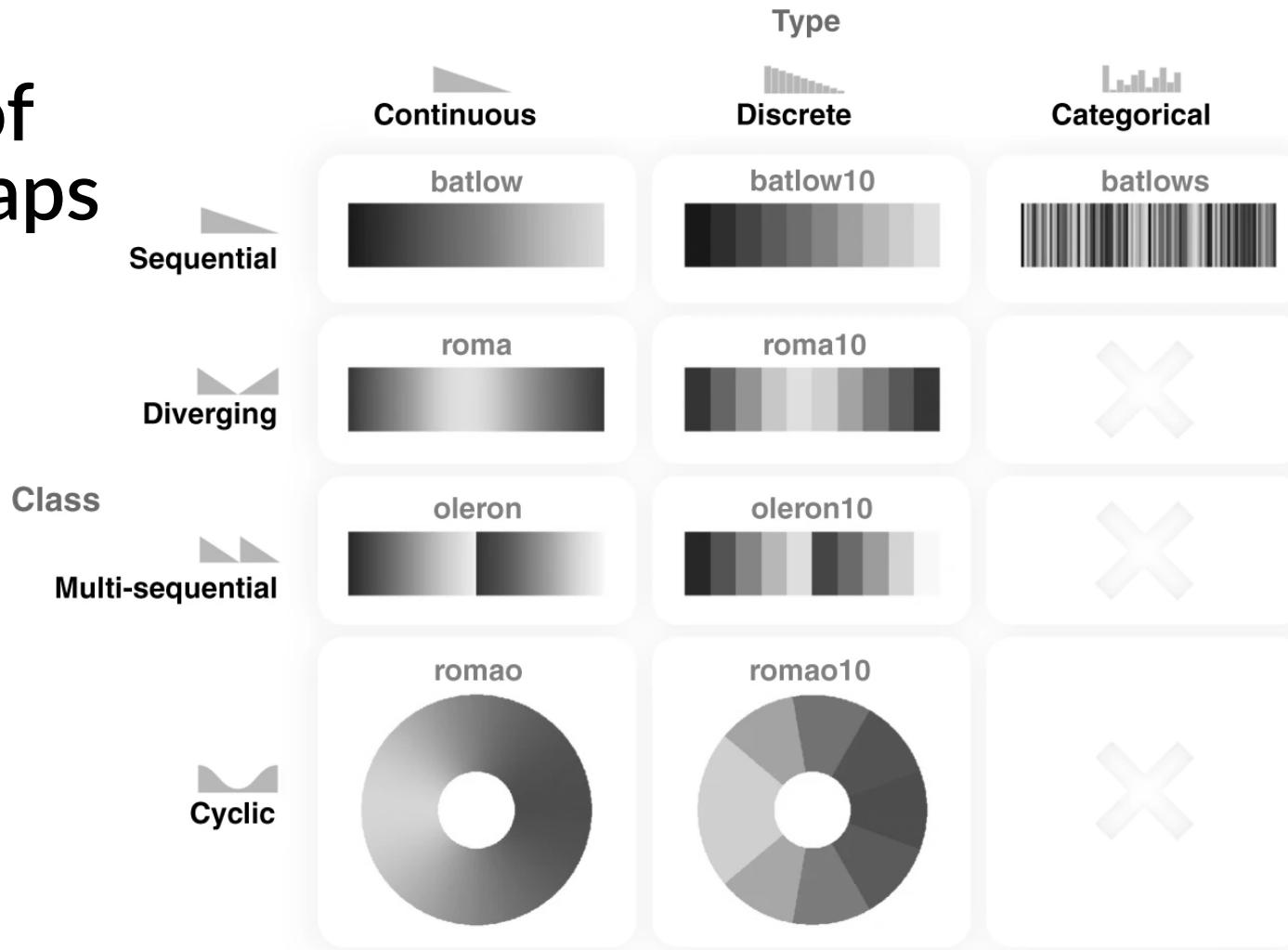
Types of colormaps



Types of colormaps



Types of colormaps



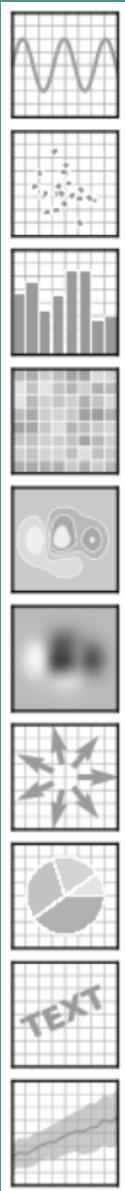
How do I remember all that??

- Test yourself, your lab mate, your friend
“Which bar is higher?”
- Look at the image in grayscale



(For future reference)

Choosing effective charts



Channels: Expressiveness Types and Effectiveness Ranks

⇒ **Magnitude Channels: Ordered Attributes**

Position on common scale



Position on unaligned scale



Length (1D size)



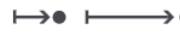
Tilt/angle



Area (2D size)



Depth (3D position)



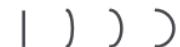
Color luminance



Color saturation



Curvature



Volume (3D size)



⇒ **Identity Channels: Categorical Attributes**

Spatial region



Color hue



Motion



Shape



Text

Fr
Ca

Canada France Canada France

https://matplotlib.org/cheatsheets/_images/cheatsheets-1.png

Munzner, T. (2014). Visualization analysis and design. CRC press.

Quick start

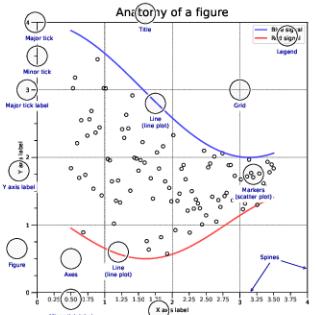
```
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
```

```
X = np.linspace(0, 2*np.pi, 100)
Y = np.cos(X)
```

```
fig, ax = plt.subplots()
ax.plot(X, Y, color='green')

fig.savefig("figure.pdf")
fig.show()
```

Anatomy of a figure



Subplots layout

```
subplot[s](rows,cols,...) API
fig, axs = plt.subplots(3, 3)

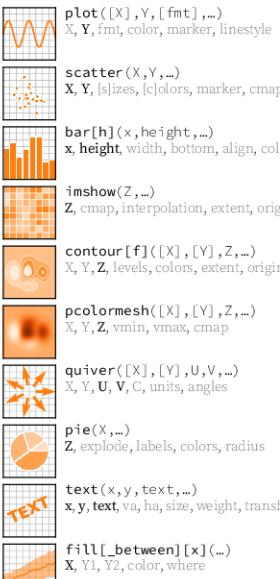
G = gridspec(rows,cols,...) API
ax = G[0,:]

ax.inset_axes(extent) API
d=make_axes_locatable(ax)
ax = d.new_horizontal('10%')
```

Getting help

- matplotlib.org
- github.com/matplotlib/matplotlib/issues
- discourse.matplotlib.org
- stackoverflow.com/questions/tagged/matplotlib
- https://gitter.im/matplotlib/matplotlib
- twitter.com/matplotlib
- Matplotlib users mailing list

Basic plots



Scales



Tick locators

```
from matplotlib import ticker
ax.[x|y]axis.set_[minor|major]_locator(locator)

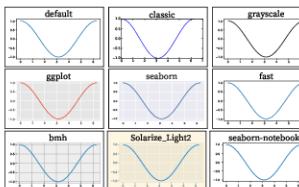
ticker.NullLocator()
ticker.MultipleLocator(base=0.5)
ticker.FixedLocator([0, 1, 5])
ticker.IndexLocator(base=0.5, offset=0.25)
ticker.LinearLocator(numticks=3)
ticker.MaxNLocator(n=4)
ticker.LogLocator(base=10, numticks=15)
```

Animation

```
T = np.linspace(0, 2*np.pi, 100)
S = np.sin(T)
line, = plt.plot(T, S)
def animate(i):
    line.set_ydata(np.sin(T+i/50))
anim = mpl.animation.FuncAnimation(
    plt.gcf(), animate, interval=5)
plt.show()
```

Styles

```
plt.style.use(style)
```



Quick reminder

```
ax.grid()
ax.set_[xy]lim(vmin, vmax)
ax.set_[xy]label(label)
ax.set_[xy]ticks(ticks, [labels])
ax.set_[xy]ticklabels(labels)
ax.set_title(title)
ax.tick_params(width=10, ...)
ax.set_axis_[on|off]()
```

```
fig.suptitle(title)
fig.tight_layout()
plt.gcf(), plt.gca()
mpl.rcParams['axes', linewidth=1, ...]
[fig|ax].patch.set_alpha(0)
text=r'$\frac{e^i}{\pi^j}$'
```

Keyboard shortcuts

ctrl + S	Save	ctrl + W	Close plot
R	Reset view	F	Fullscreen 0/1
V	View forward	B	View back
P	Pan view	Z	Zoom to rect
X	X pan/zoom	Y	Y pan/zoom
G	Minor grid 0/1	G	Major grid 0/1
L	X axis log/linear	L	Y axis log/linear

Ten simple rules

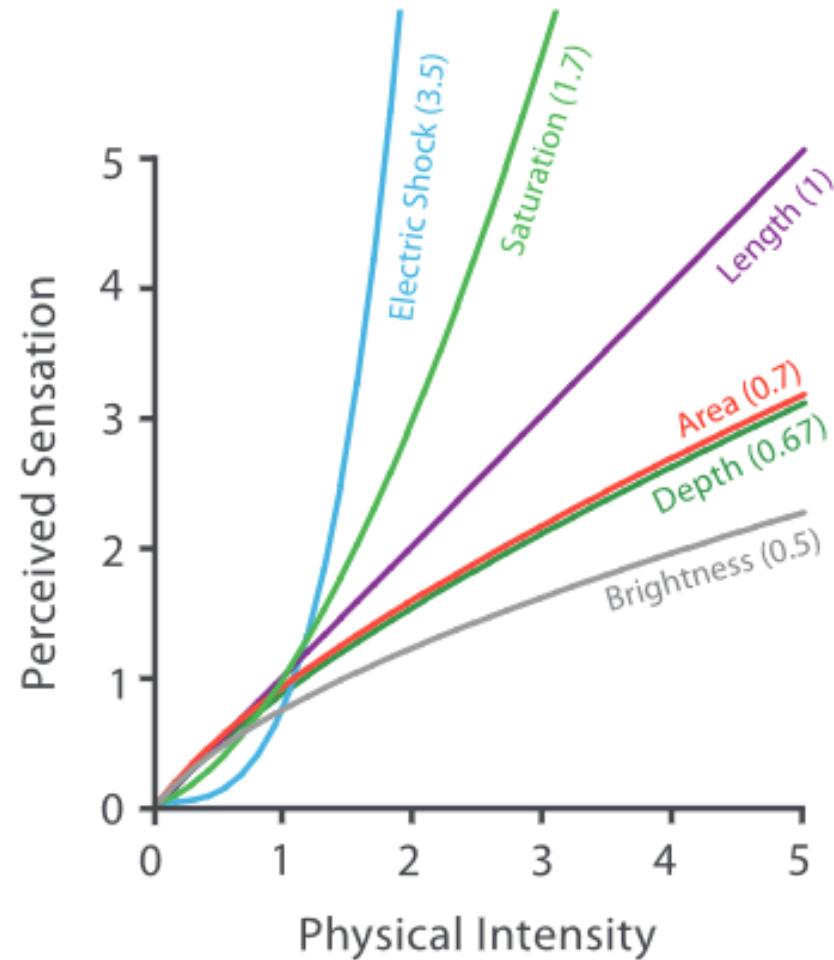
1. Know your audience
2. Identify your message
3. Adapt the figure
4. Captions are not optional
5. Do not trust the defaults
6. Use color effectively
7. Do not mislead the reader
8. Avoid "charjunk"
9. Message trumps beauty
10. Get the right tool

Lineplot

Ordered

Categorical

Steven's Psychophysical Power Law: $S = I^N$



Lineplot

Ordered

Position on a common scale

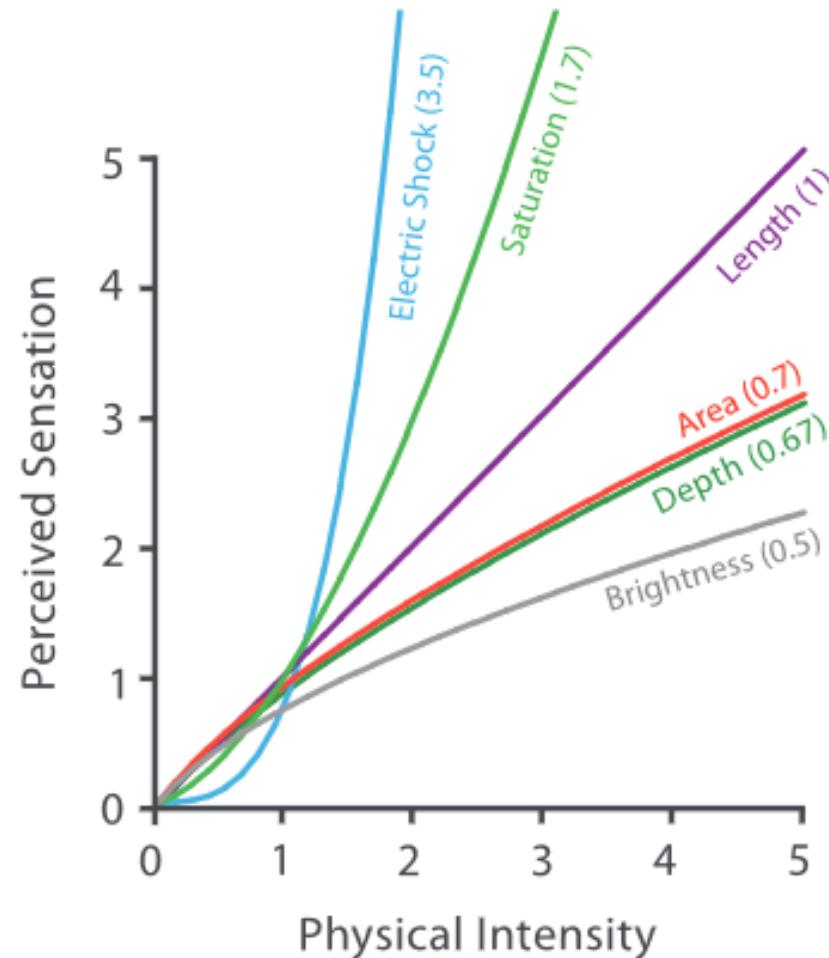


Text (equation)

Canada Canada Canada
France France France

Categorical

Steven's Psychophysical Power Law: $S = I^N$



Lineplot

Ordered

Position on a common scale



Canada Canada Canada
France France France

Text (equation)

Categorical

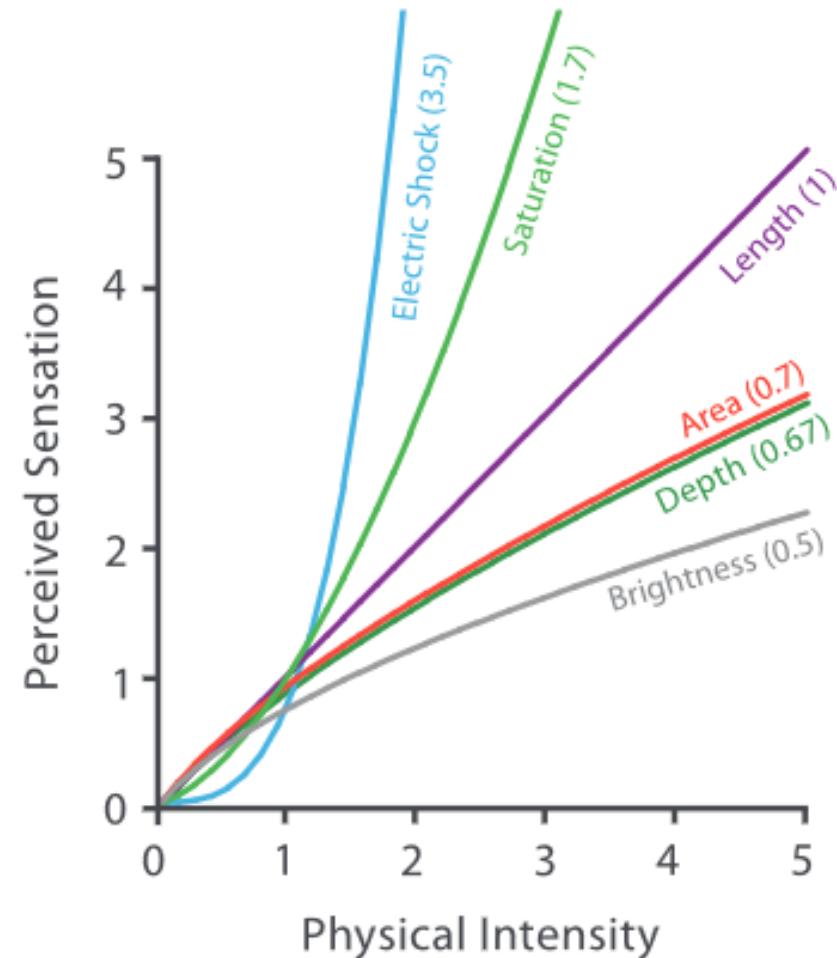
Hue



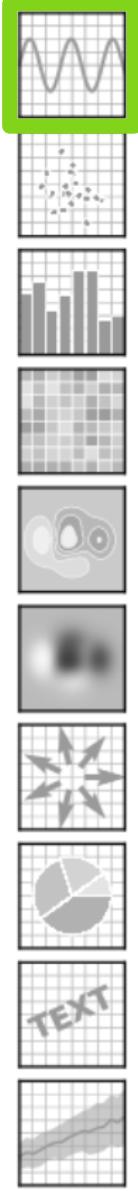
Text (identity)

Canada Canada Canada
France France France

Steven's Psychophysical Power Law: $S = I^N$



Lineplot



Channels: Expressiveness Types and Effectiveness Ranks

⇒ **Magnitude** Channels: **Ordered Attributes**

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



⇒ **Identity** Channels: **Categorical Attributes**

Spatial region



Color hue



Motion



Shape



Text



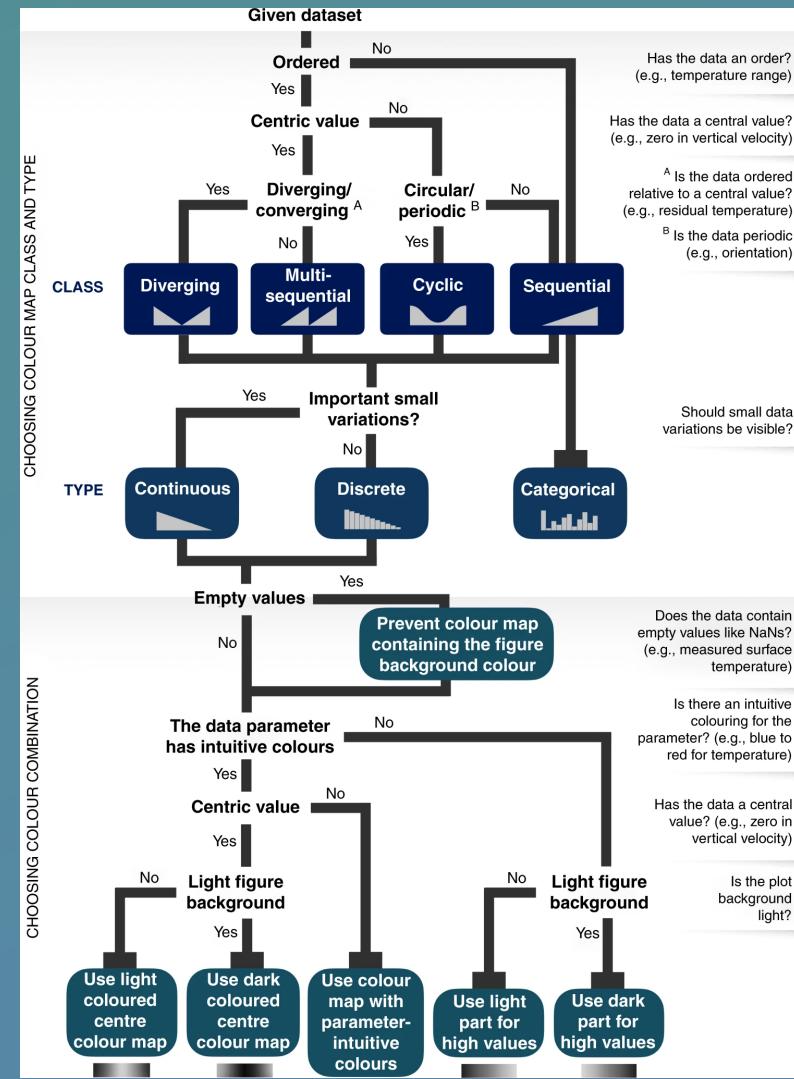
Effectiveness Rank (Y-axis): Most Effective ↑ | ↓ Least Effective

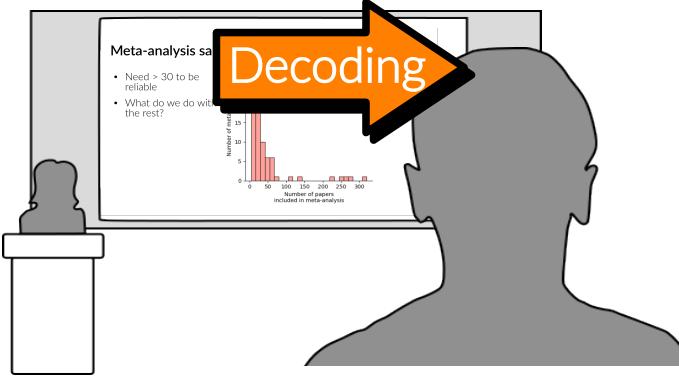
https://matplotlib.org/cheatsheets/_images/cheatsheets-1.png

Munzner, T. (2014). Visualization analysis and design. CRC press.

(For future reference)

Choosing effective colormaps



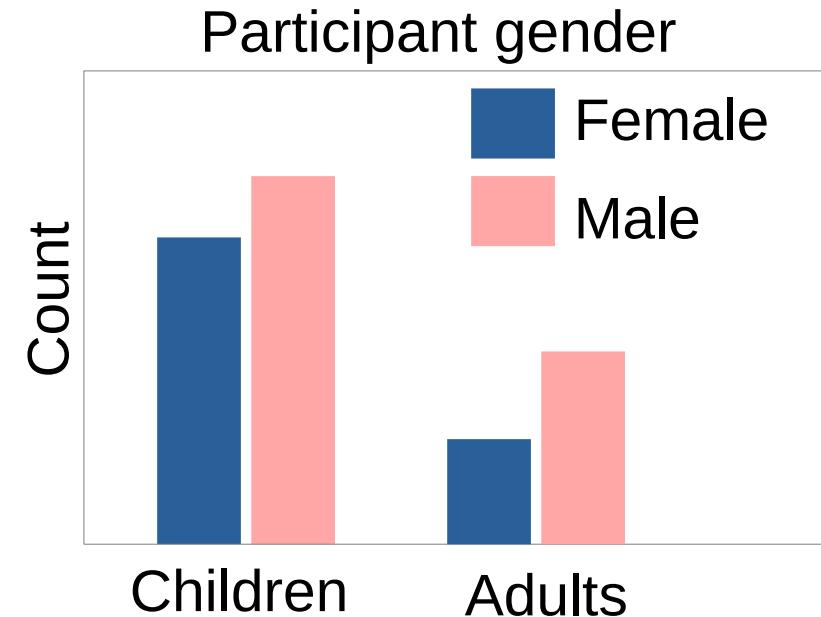


Decoding

To plan an **effective visualization**, we need to think about

- **Message**
 - What we want to communicate
- **Perception**
 - How best to communicate it
- **Conventions**
 - How it's usually communicated
- **Context**
 - Where it will be seen

Culture: Color associations



Science: Rainbow colormaps are disliked

← **Tweet**



Patrick Mineault 
@patrickmineault

Does anybody else have a visceral negative reaction when they see the jet colormap?

Science: Rainbow colormaps are disliked

Chris Holdgraf 🐘 @choldgraf@hachyderm.io
@choldgraf

Academia twitter: when you review papers, do you instruct authors to change their colormap from jet to either viridis/parula?

Option	Percentage
yes	24.5%
no	36.2%
just wanna see answers	35.1%
I require plots in ASCII	4.3%

94 votes · Final results

1:48 PM · Apr 7, 2018

Science: Rainbow colormaps are disliked?

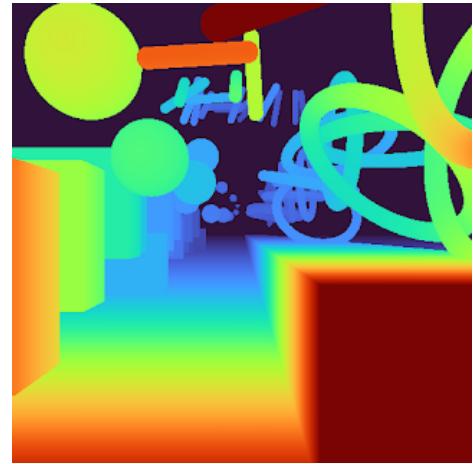


Simon Eickhoff
@INM7_ISN

Turbo, a colormap that looks like jet without its downsides

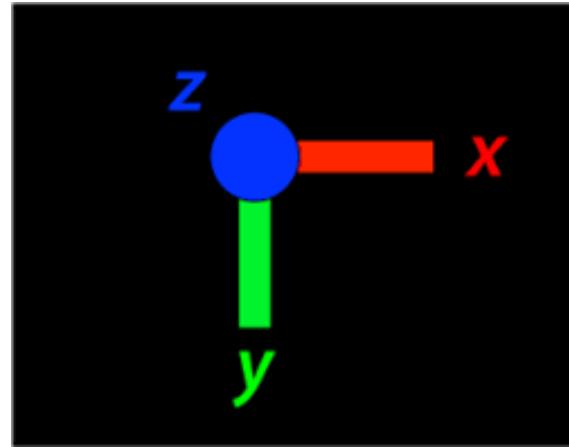
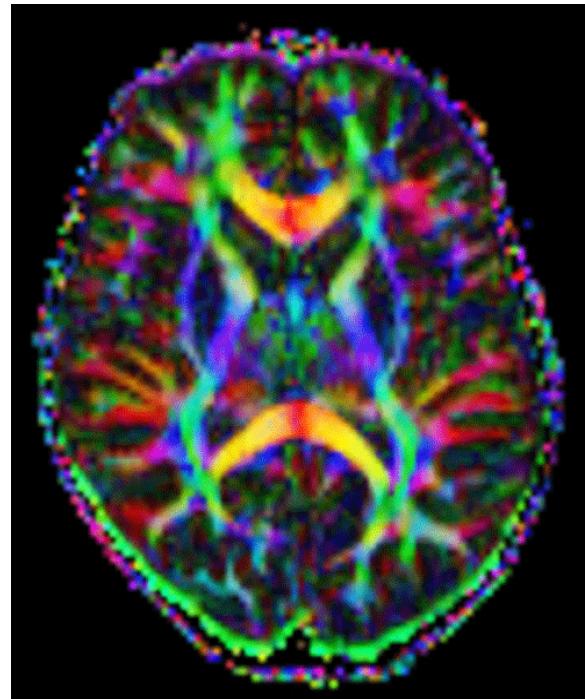
Great news for somebody like me, who likes jet as nameable colors are indispensable in many applications: Try explaining anybody, which shade of parula actually denotes the interesting finding

buff.ly/30h9N8O



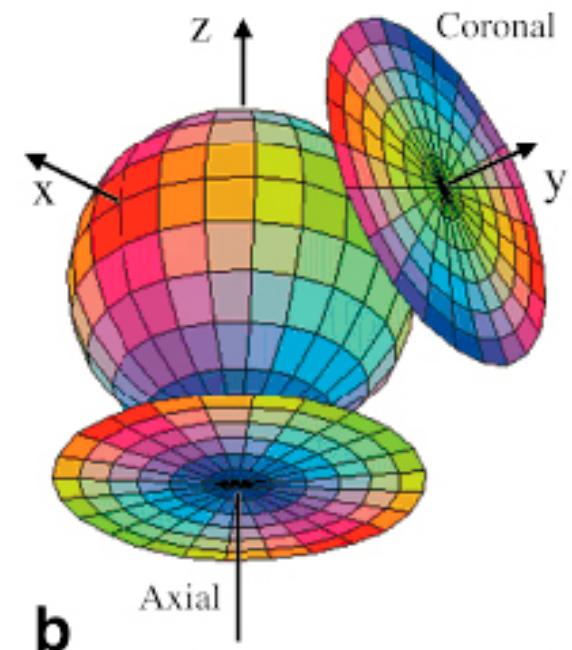
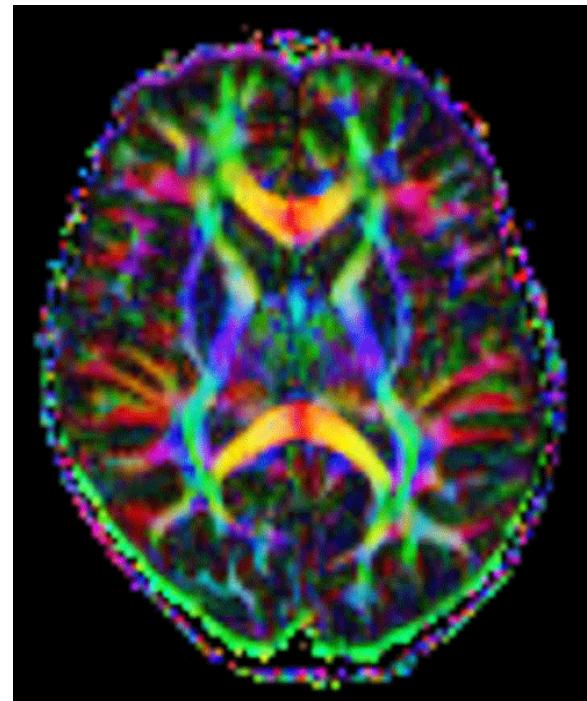
Neuroscience: Colors = 3D direction

Principal Diffusion Direction



Neuroscience: Colors = 3D direction

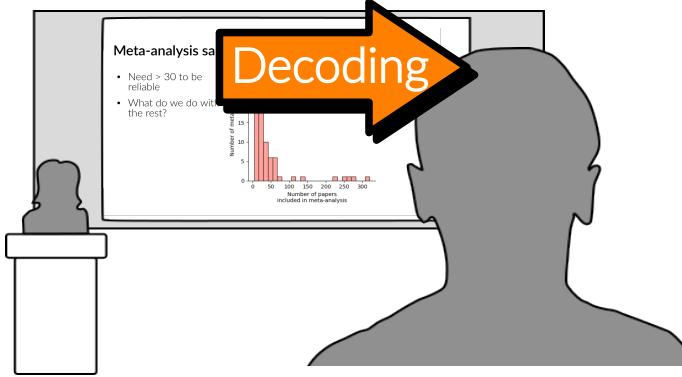
Principal Diffusion Direction



(For future reference)

Examples of visualization conventions

- Culture
 - Pink is female
 - Time goes left to right
- Science
 - Rainbow colormaps are disliked
- Neuroscience
 - MRI in grayscale
 - DWI colors = directions

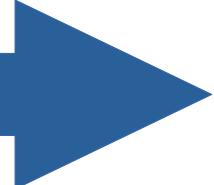


Decoding

To plan an **effective visualization**, we need to think about

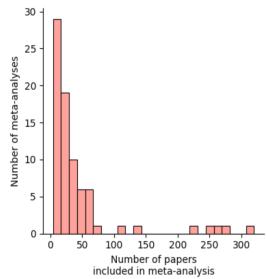
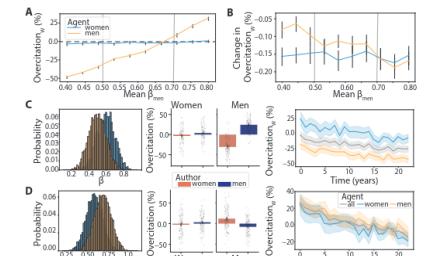
- **Message**
 - What we want to communicate
- **Perception**
 - How best to communicate it
- **Conventions**
 - How it's usually communicated
- **Context**
 - Where it will be seen

More self-explanatory



Meta-analysis sample sizes

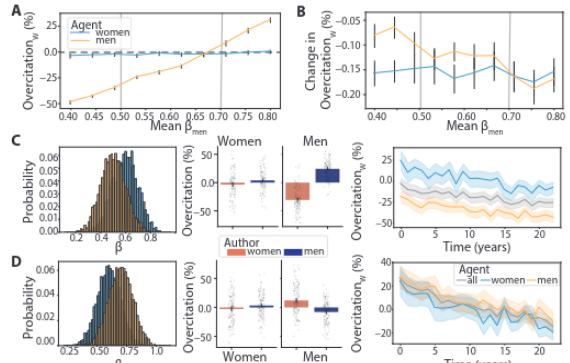
- Need > 30 to be reliable
- What do we do with the rest?

representation in the field leads to more equitable citation practices, but does not reverse temporal trends.

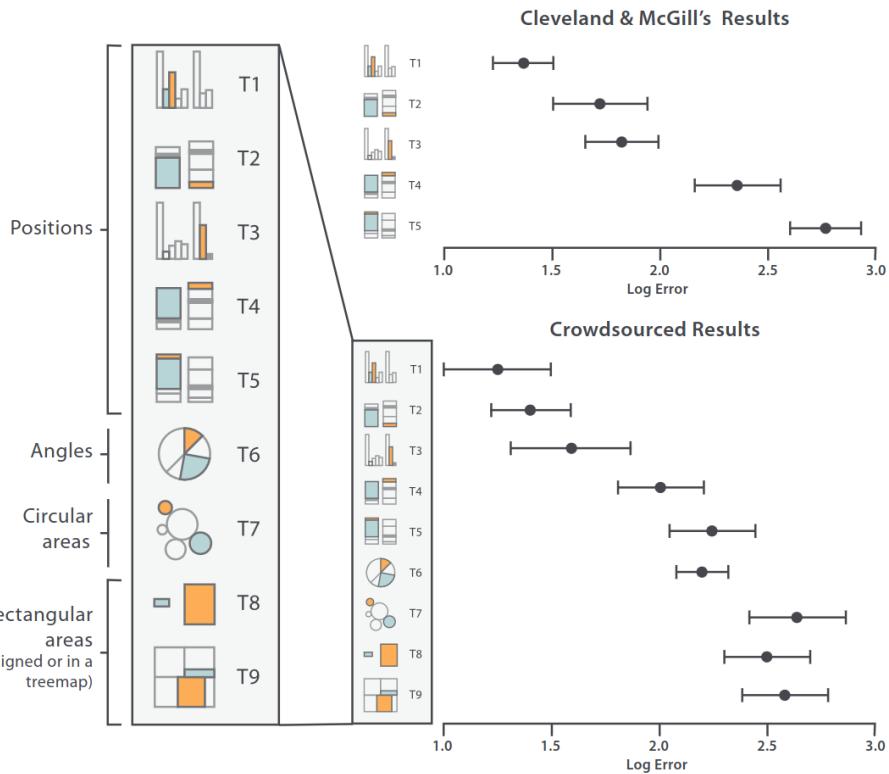
Interventions to biased meeting preferences

We next simulated the effects of intervention to the meeting preferences of men. Here we select 10 values for γ that range between 0.01 and 0.06 (the original value used). The value of 0.001 effectively means agents ignore gender when selecting meeting partners, while authors in their citation histories. Broadly, we demonstrate that increasing the selectivity of agents that men have for citing women leads to higher overcitation of women (Fig. 3A), and leads to sharper decreases in the overcitation of women over time (Fig. 3B). We next examine the effect of interventions to more equitable citation practices (Fig. 3C, 0.01). We find that for γ values of 0.01, man agents still significantly under-



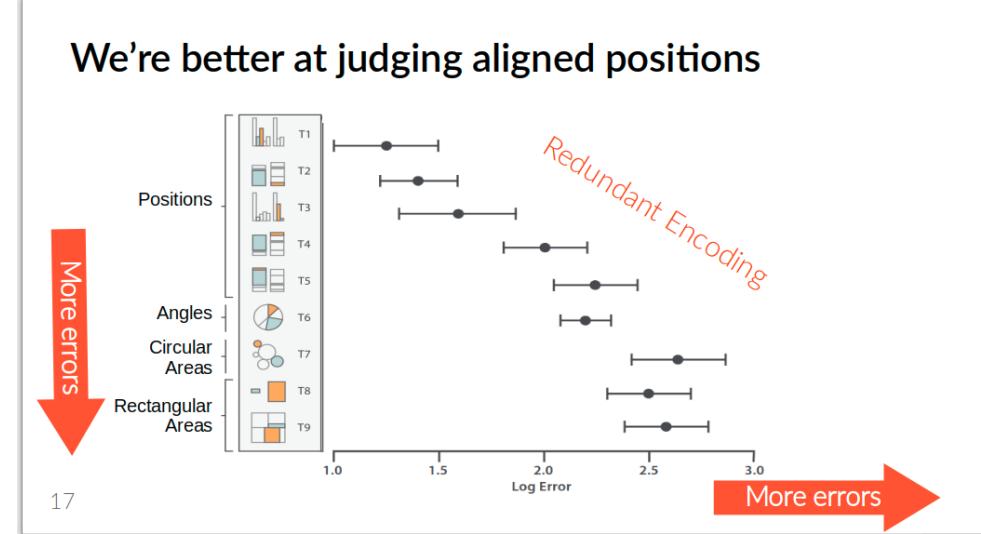
- How long would it take them to understand this slide?
- How long will they see the slide?

E.g., Figure in textbook



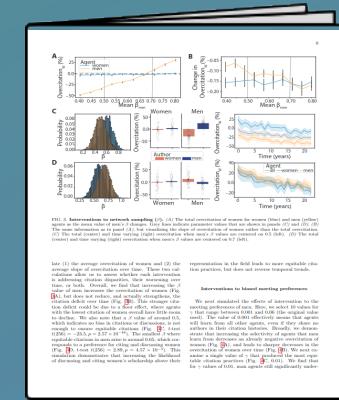
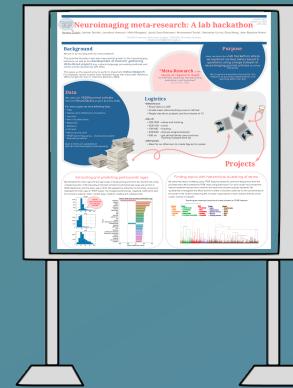
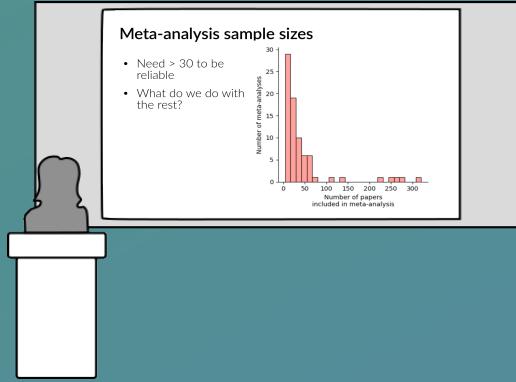
My slide

We're better at judging aligned positions



(For future reference)

Contextual adjustments



Time to understand

Proximity to viz

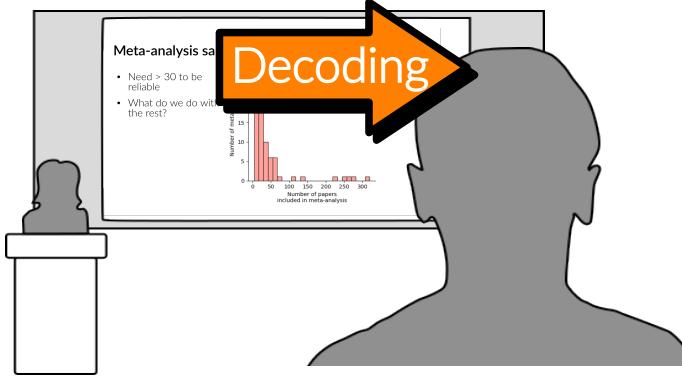
Less

Text explanations

More

Self-explanatory

Visual complexity



To plan an **effective visualization**, we need to think about

- **Message**
 - **What** we want to communicate
- **Perception**
 - **How best** to communicate it
- **Conventions**
 - **How** it's usually communicated
- **Context**
 - **Where** it will be seen

Example 1 – Simple data and figure

- Original paper on the ABIDE dataset
 - 964 subjects
 - 396 male
 - 51 female

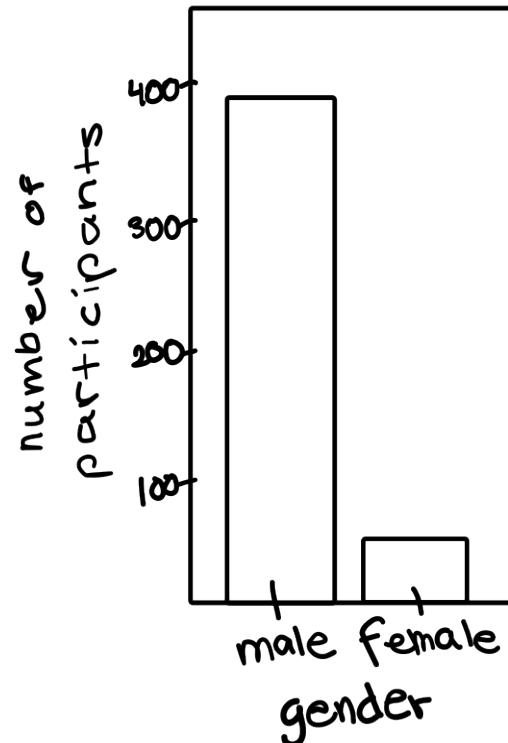
← Lets visualize this

What's our message?

- More male than female participants
 - 2 categories, 2 values

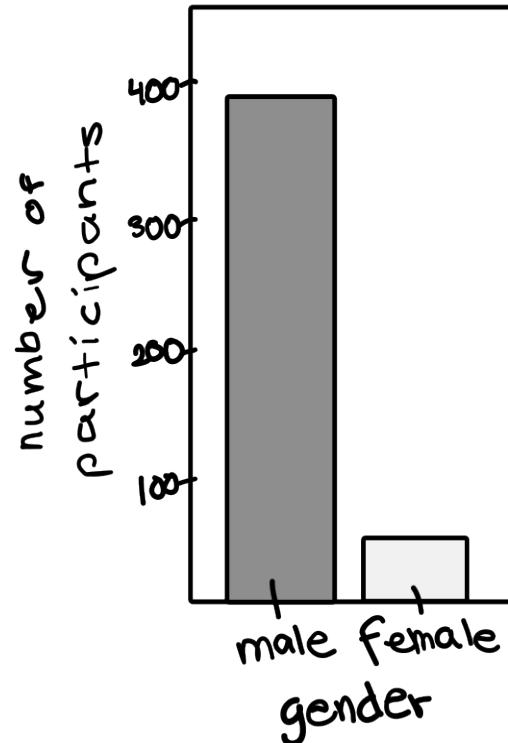
How should we communicate this message?

- Clearly show 2 categories and 2 different values



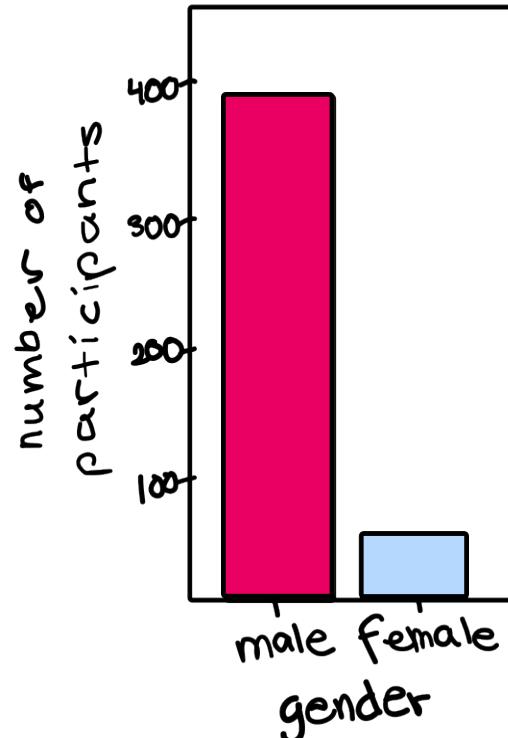
How should we communicate this message?

- Clearly show 2 categories and 2 different values



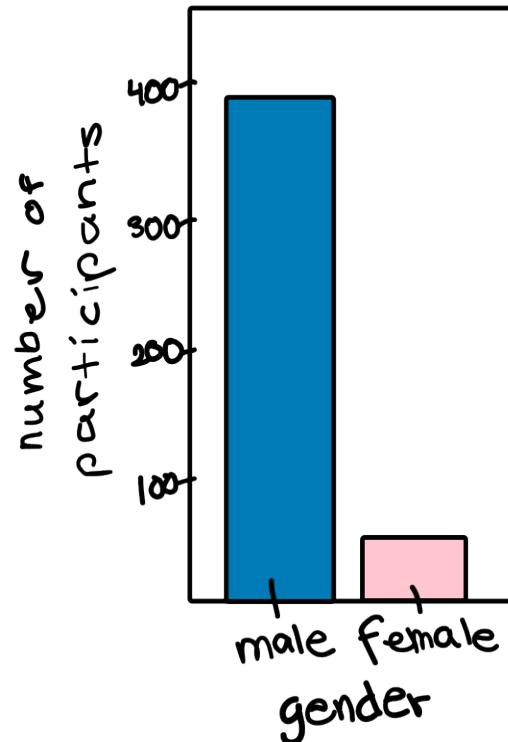
How should we communicate this message?

- Clearly show 2 categories and 2 different values



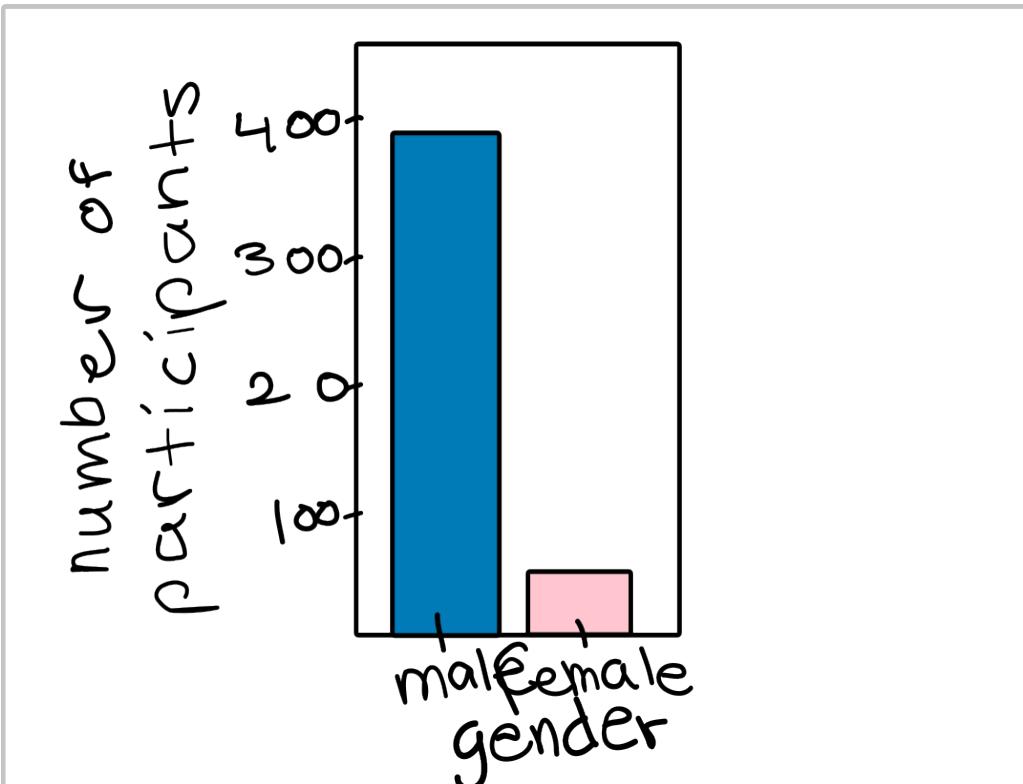
Are there any conventions we should think about?

- Gender & color



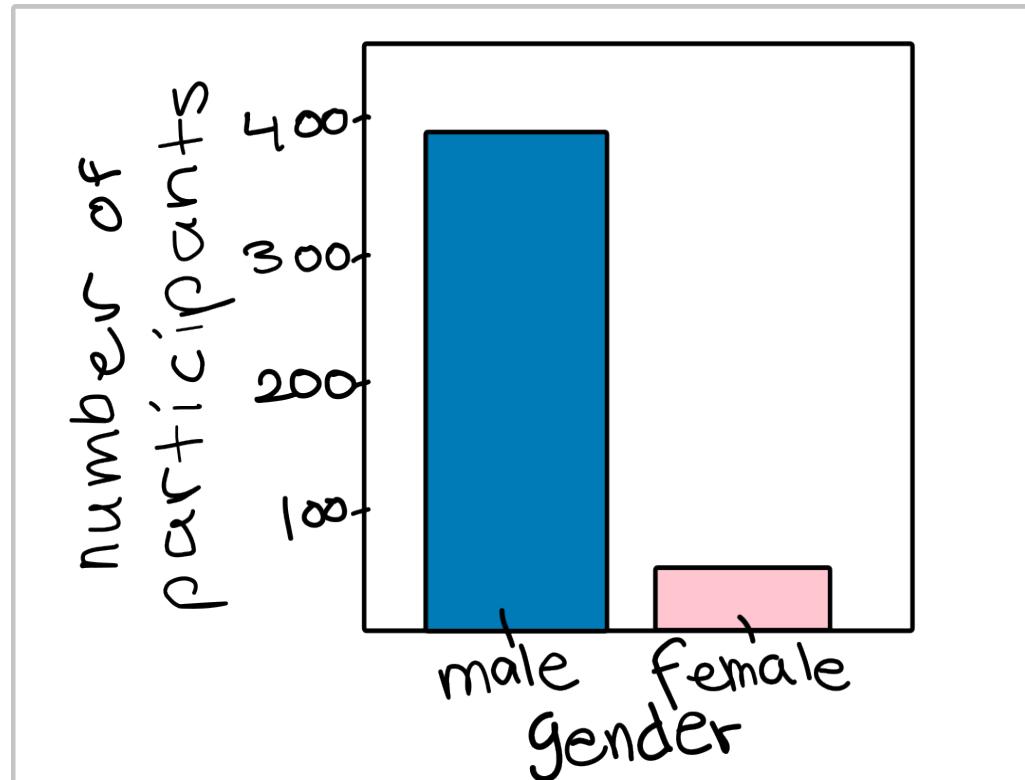
Do we need to adapt it to the context?

- Academic presentation



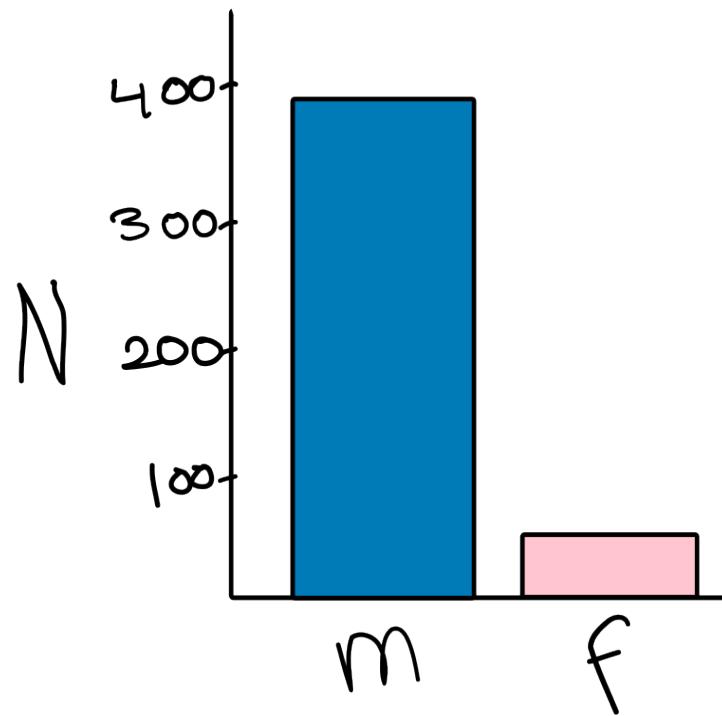
Do we need to adapt it to the context?

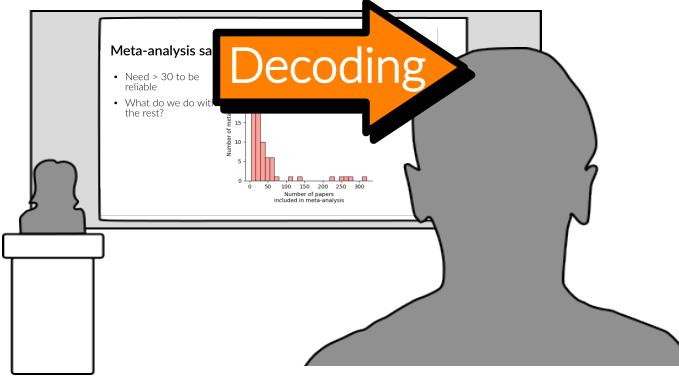
- Academic presentation



Do we need to adapt it to the context?

- Academic presentation





To plan an **effective visualization**, we need to think about

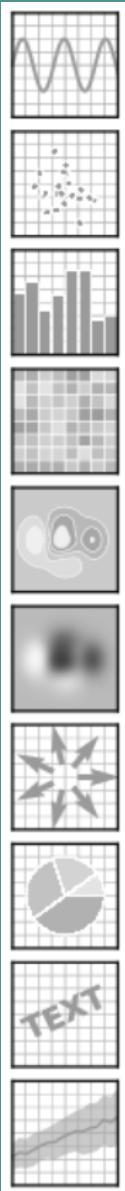
- **Message**
 - **What** we want to communicate
- **Perception**
 - **How best** to communicate it
- **Conventions**
 - **How** it's usually communicated
- **Context**
 - **Where** it will be seen

The end of part 1

Reference slides...

(For future reference)

Choosing effective charts



Channels: Expressiveness Types and Effectiveness Ranks

⇒ **Magnitude Channels: Ordered Attributes**

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



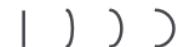
Color luminance



Color saturation



Curvature



Volume (3D size)



⇒ **Identity Channels: Categorical Attributes**

Spatial region



Color hue



Motion



Shape



Text

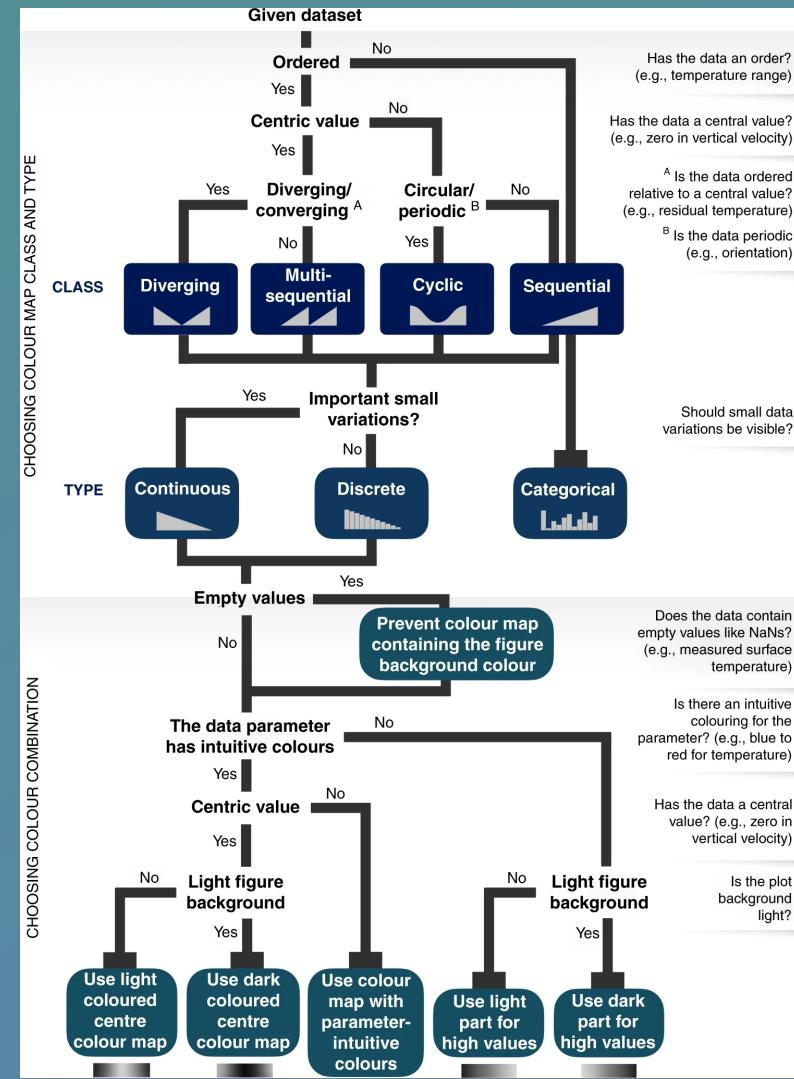


https://matplotlib.org/cheatsheets/_images/cheatsheets-1.png

Munzner, T. (2014). Visualization analysis and design. CRC press.

(For future reference)

Choosing effective colormaps



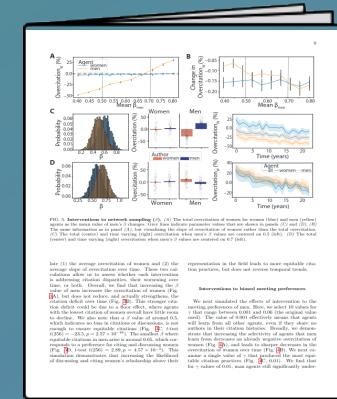
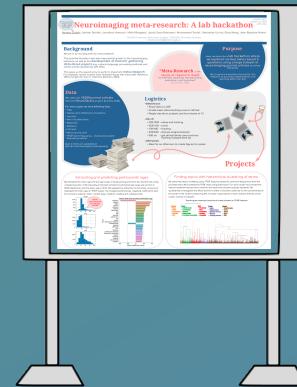
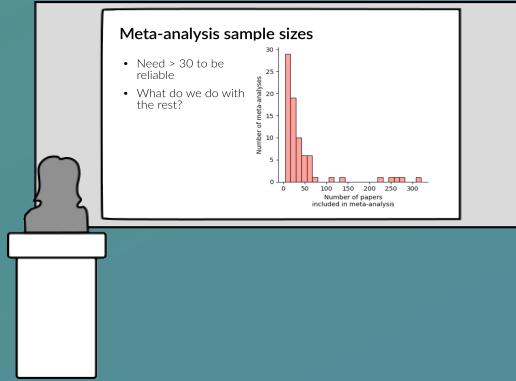
(For future reference)

Examples of visualization conventions

- Culture
 - Pink is female
 - Time goes left to right
- Science
 - Rainbow colormaps are disliked
- Neuroscience
 - MRI in grayscale
 - DWI colors = directions

(For future reference)

Contextual adjustments



Time to understand

Proximity to viz

Less

Text explanations

More

Self-explanatory

Visual complexity