

# Introduction to Data Visualization

Aaron Geller

(with help from Christina Maimone, Frank Elavsky, Steve Franconeri  
and Adam Miller)



# Initial Considerations

First, think carefully about these:

1. Choose exploratory vs. explanatory visualization.
2. Use storytelling to share your message.
3. Use your tools wisely (salience, symbols, colors, dimensions, axes/grids).
4. Carry these important suggestions in your back pocket.

Improve Data-to-Ink Ratio

Increase Efficiency

Ensure Visual Quality

Consider Accessibility

Reduce Clutter

Organize and Guide

Two great resources: [Nature Methods points of view](#), [Google material design principles](#)



# Important Suggestions

**Improve Data-to-Ink Ratio**

Is this [shape] necessary?

**Ensure Visual Quality**

Is this geometry telling the truth?

**Reduce Clutter**

Is this color choice or layout necessary?

**Increase Efficiency**

Is it too hard or time consuming to read?

**Consider Accessibility**

Is this colorblind safe? Is the font size large enough?

**Organize and Guide**

Should I regroup my data? Can I add helpful text?



# Edward Tufte's “Data to Ink Ratio”

The golden rule of visualization

**Remove**  
to improve  
(the **data-ink** ratio)



# Edward Tufte's “Data to Ink Ratio”

The golden rule of visualization (even for tables)

---

**Remove**  
to improve  
the **data tables** edition

# Exploration

---

*Help YOU learn about your data*

**Audience**

You and your collaborators

**Number of Visualizations**

Many

**Visualization Message**

Unknown

**Formatting**

Not important

# Exploration

*Help YOU learn about your data*

## Audience

You and your collaborators

## Number of Visualizations

Many

## Visualization Message

Unknown

## Formatting

Not important

# Explanation

*Help OTHERS learn about your data*

Pick one

Probably 1 per dataset

Why include the visualization?

Important, possibly restricted



# Storytelling

**Communication:** (noun) a process by which info is exchanged between individuals through a common system of symbols, signs, behavior.

Humans communicate via **stories**



# Storytelling

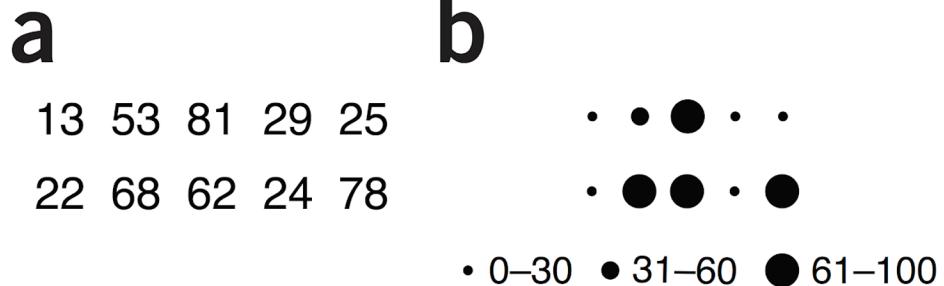
a

13 53 81 29 25

22 68 62 24 78



# Storytelling





# Storytelling

a

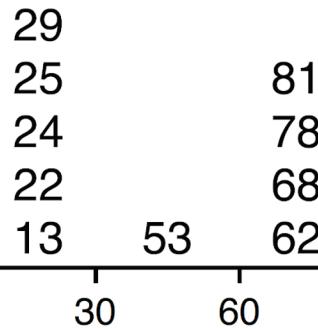
13 53 81 29 25  
22 68 62 24 78

b

• • ● • •  
• ● ● • ●

• 0–30 • 31–60 ● 61–100

c





# Storytelling

a

13 53 81 29 25  
22 68 62 24 78

b

• • ● • •  
• ● ● • ●

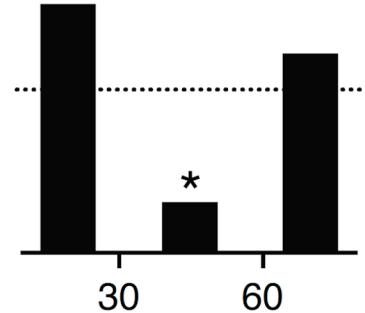
• 0–30 • 31–60 ● 61–100

c

29  
25  
24  
22  
13  
53  
62  
81  
78  
68

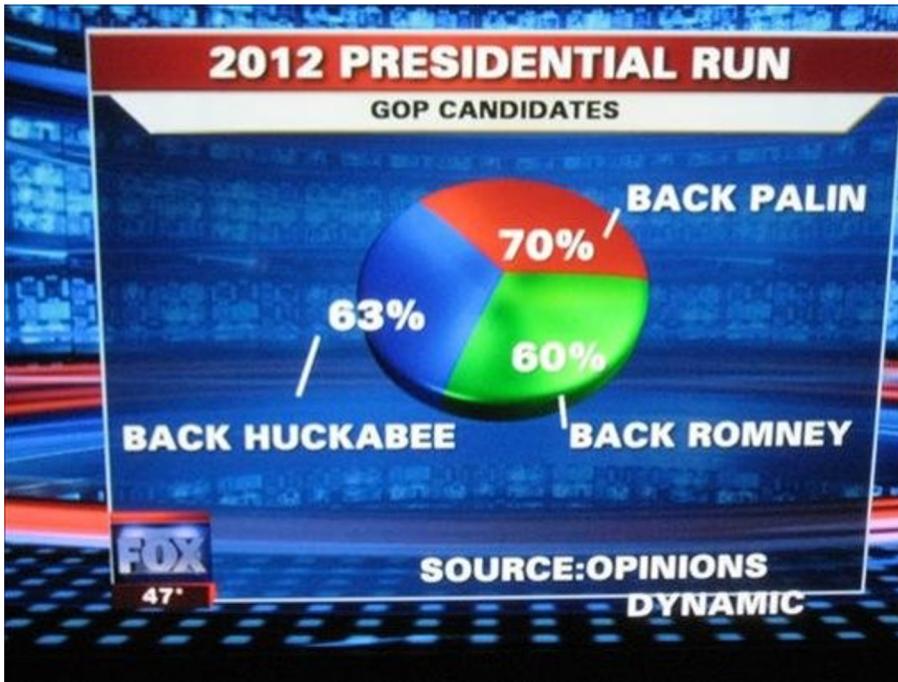
30 60

d



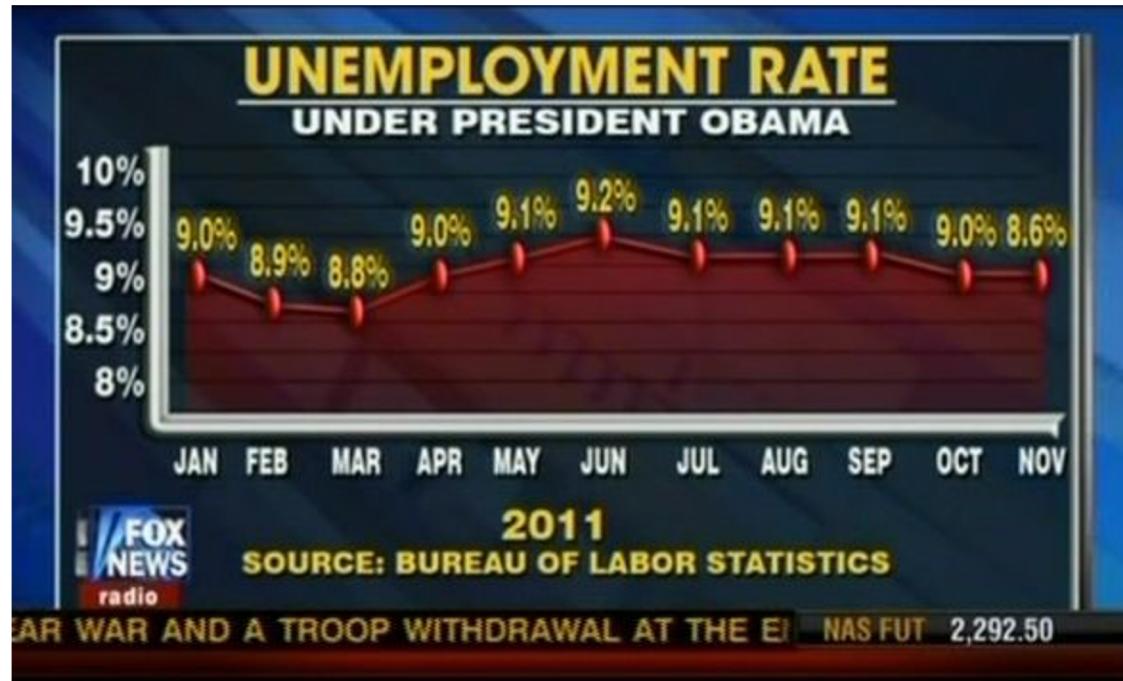


# Use Your Tools Wisely!





# Use Your Tools Wisely!





# Your Tools: Salience

Guide the viewer to your result.

**a**

Easy

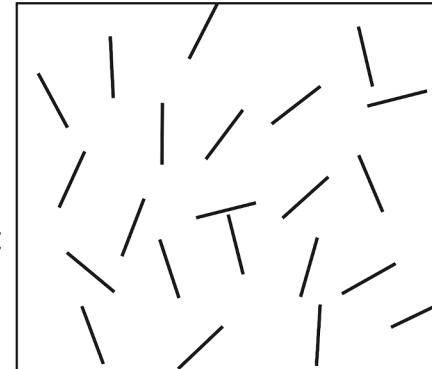
**A**

Difficult

**P**

MSVTLHTVFCERTPKTC  
EMESRCVPQEGVQWRDL  
GS**A**LQPGFGGFKQVFCL  
SLPRTGRGGNSIWWGKK  
FEDEYSEYSEYLKH**A**VR  
GVVSMSNNGPNTNGSQF  
FITYGKQPHLDMKYTGF  
GKVIDGLEK**A**PVNEKTY  
RPLNDVHIKDITIHNPF

Easy  
  
Difficult



**b**

Color

Size

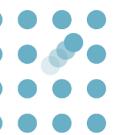
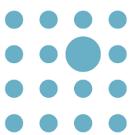
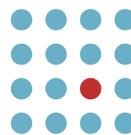
Orientation

Shape

Added mark

Motion

Grouping



Wong (2010)

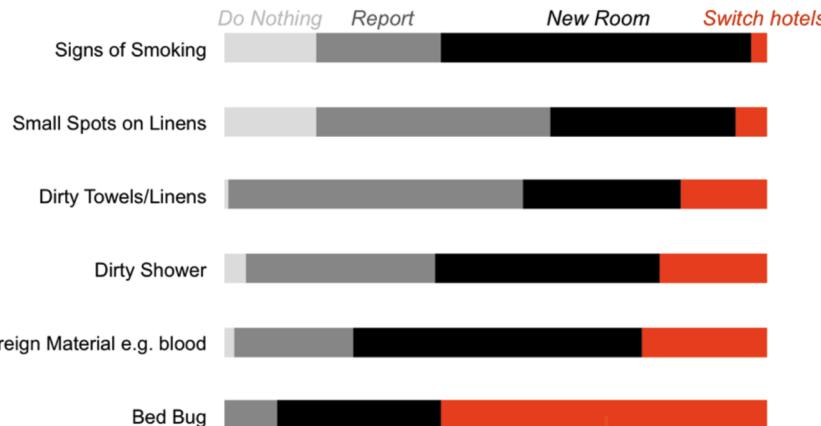


# Your Tools: Salience

Guide the viewer to your result.

**Bedbugs are the #1 cause of lost guests**

Clean design



Guide attention

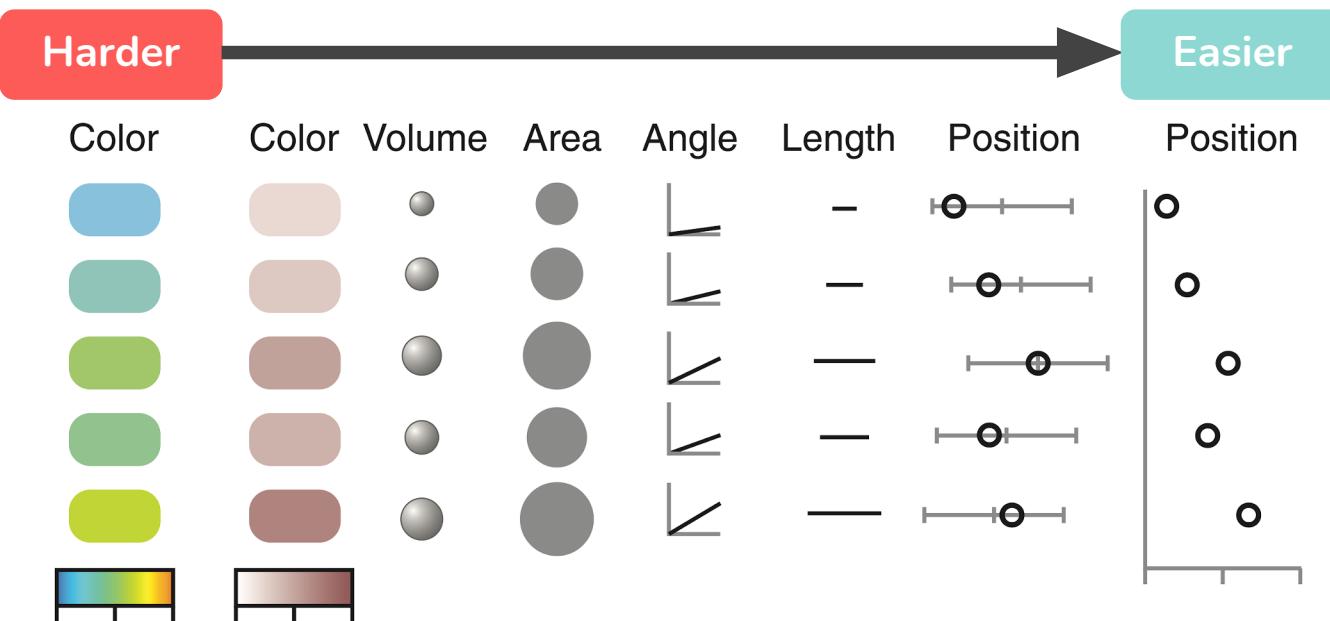


*60% of those surveyed said they would switch to a new hotel after finding a bedbug - and would not return to that chain*



# Your Tools: Salience

Use easy-to-estimate visual representations

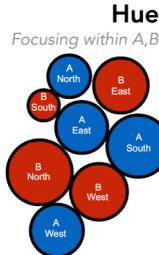
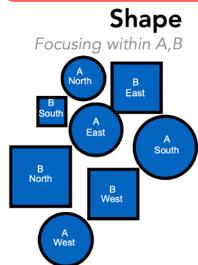




# Your Tools: Salience

Organize for comparison.

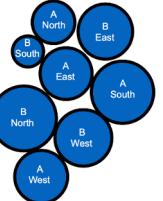
Harder



**Position V or H**  
Focusing within A,B

	North	South	East	West
A	●			
B		●		
A			●	
B				●

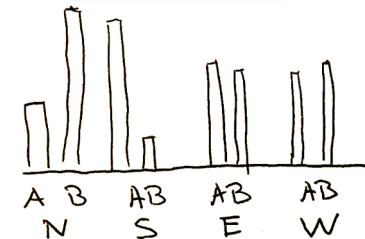
**Just labels**  
Focuses on nothing



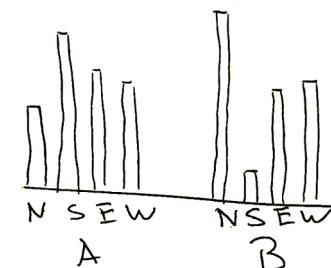
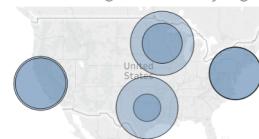
Focusing within N,S,E,W

	North	South	East	West
North	A			
South	B	A		
East			A	
West				B

Easier



**Maps are Position V+H**  
For focusing within nearby regions



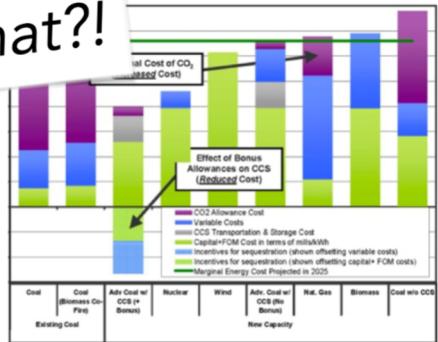
Slide adapted from Steve Franconeri (Northwestern)



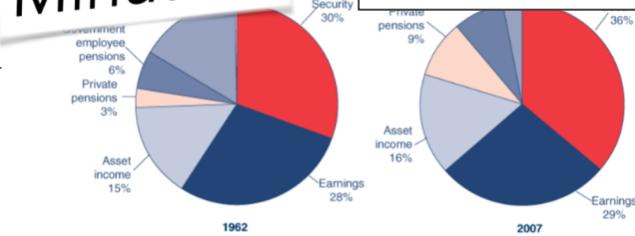
# Your Tools: Salience

Technique matters.

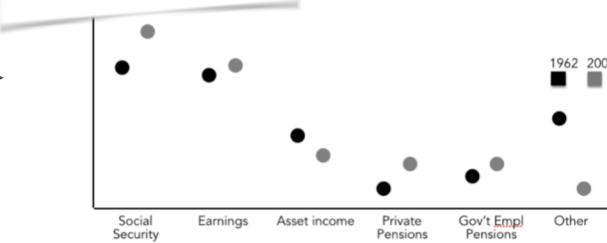
What?!



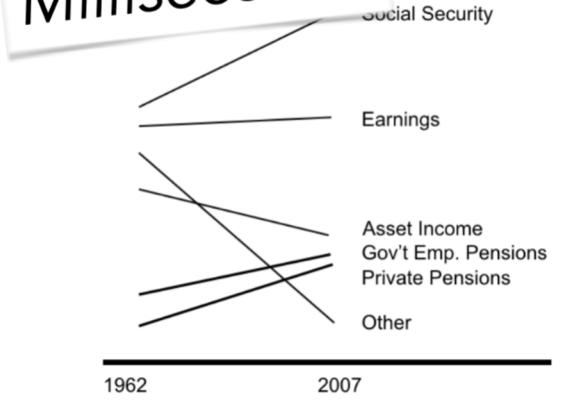
Minutes



Seconds



Milliseconds



Slide adapted from Steve Franconeri (Northwestern)



# Your Tools: Salience

Consider the time it takes your viewer to digest.

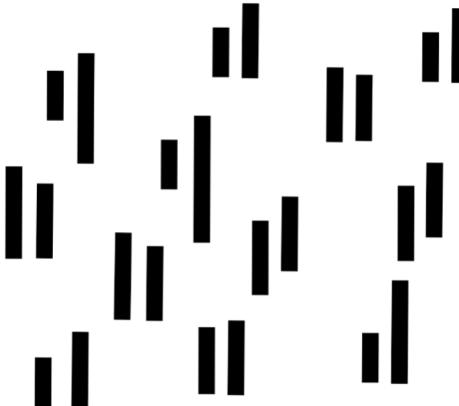
Seeing distributions is easy

*Who's the biggest?*



Comparisons are expensive

*Who's decreasing?*

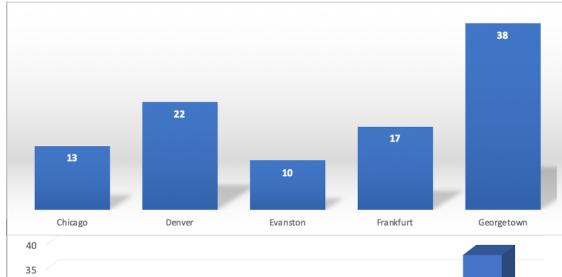


Slide adapted from Steve Franconeri (Northwestern)

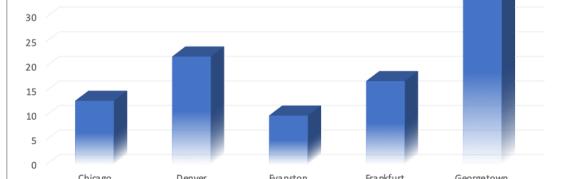


# Your Tools: Salience

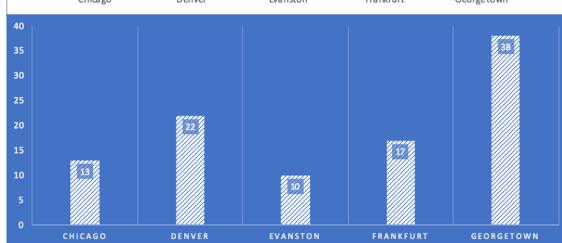
Keep it simple.



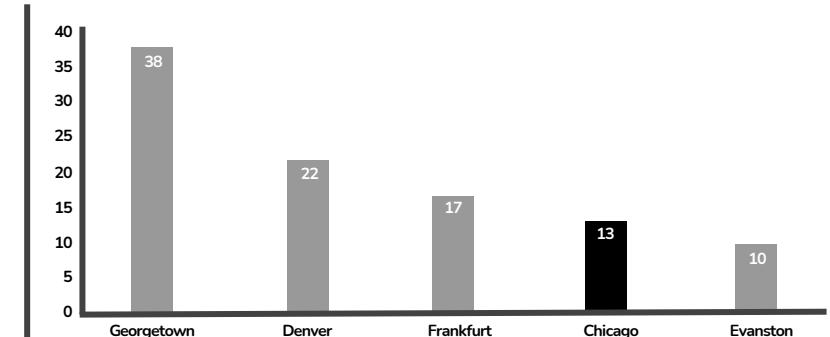
No shadows



No gradients



No patterns



Yes

*Expert tips: Order bars by value (if categories aren't ordered), and guide the reader to important value(s)*



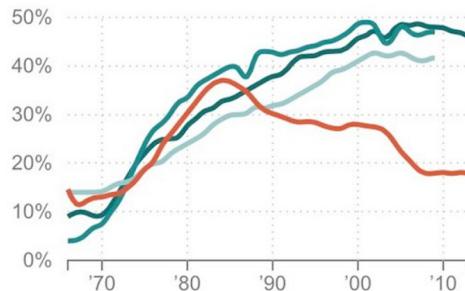
# Your Tools: Salience

Try direct labels instead of legends.



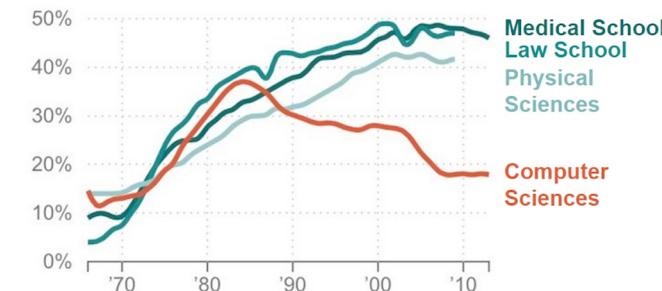
## What Happened To Women In Computer Science?

% Of Women Majors, By Field



## What Happened To Women In Computer Science?

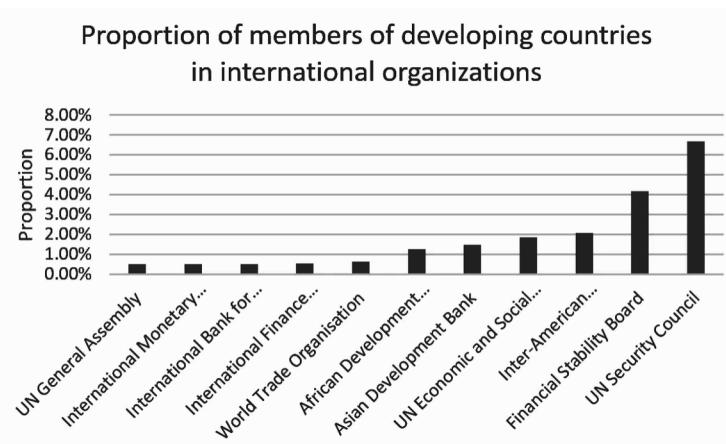
% Of Women Majors, By Field





# Your Tools: Salience

Make sure labels are legible (orientation, font size, etc.).



No

Developing countries make up less than 7% of international organization membership.

Proportion of members of developing countries in international organizations, most current year of data (2015, 2016)

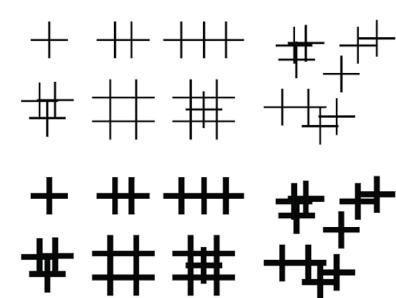
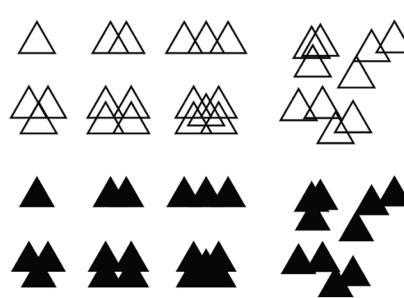
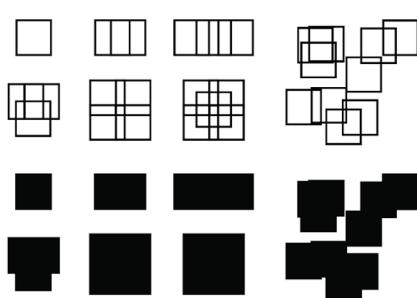
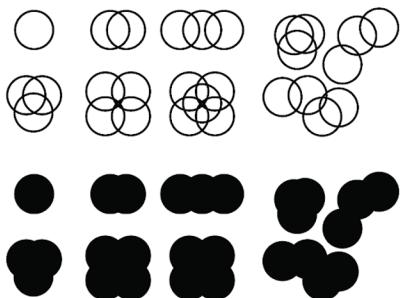


Yes



# Your Tools: Symbols

Open circles are the most flexible.

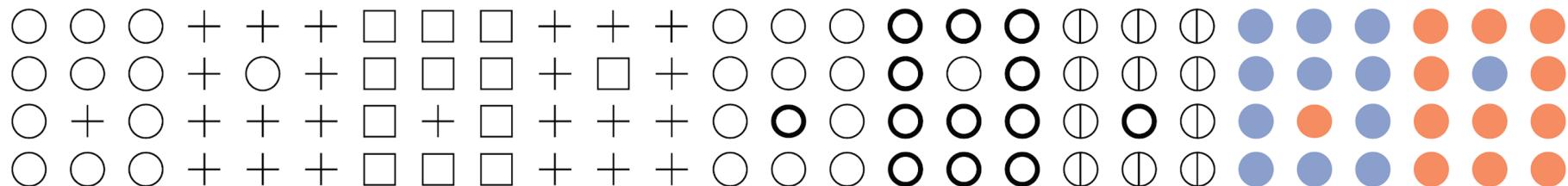




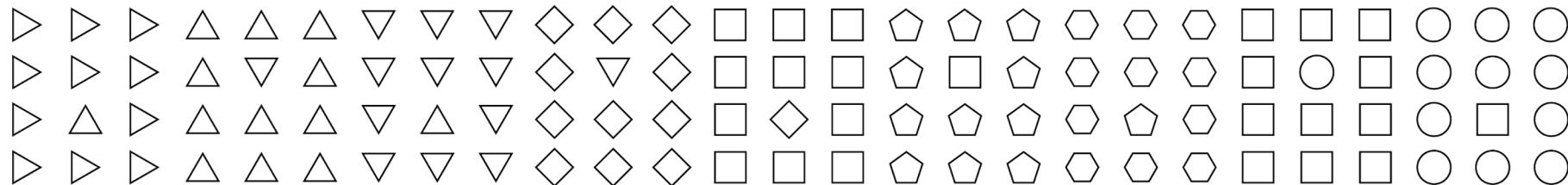
# Your Tools: Symbols

Form strong visual boundaries.

Strong visual boundaries



Weak visual boundaries





# Your Tools: Symbols

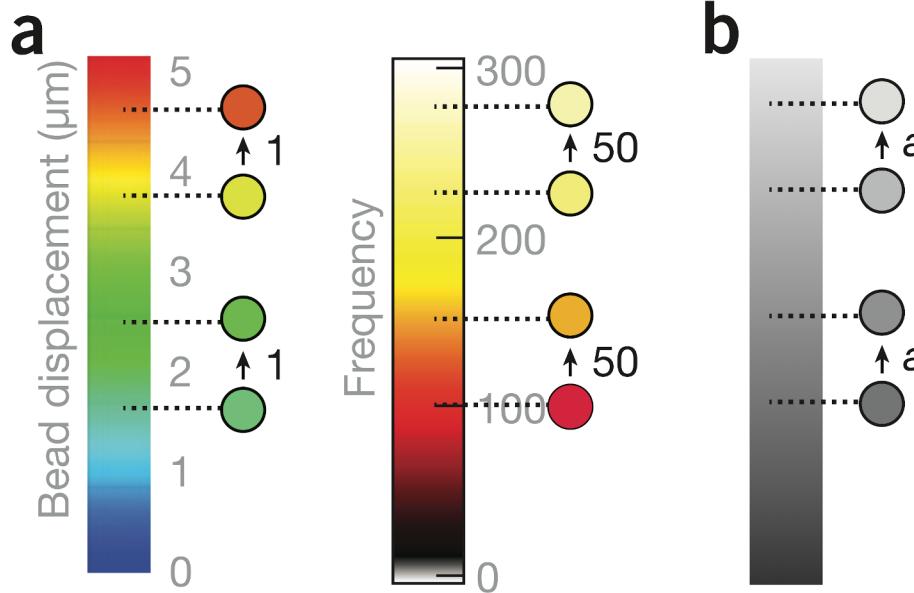
Form strong visual boundaries.

A	N	T	A	○	□	△	○	●	●	●	●	○	●	○	○	○	w	<	○
A	T	G	T	○	△	☆	△	●	●	●	●	○	○	○	○	○	<	S	<
A	C	N	N	○	▽	□	□	●	●	●	●	○	●	●	●	○	+	w	w
A	A	A	A	○	○	○	○	●	●	●	●	○	○	○	○	○	○	○	○



# Your Tools: Colors

Choose colormaps wisely, but note that color is not ideal for representing quantitative data.





# Your Tools: Colors

Colors can have meaning.

Good

Bad

Banana

Sky



# Your Tools: Colors

Colors can have meaning.

Good

Bad

Banana

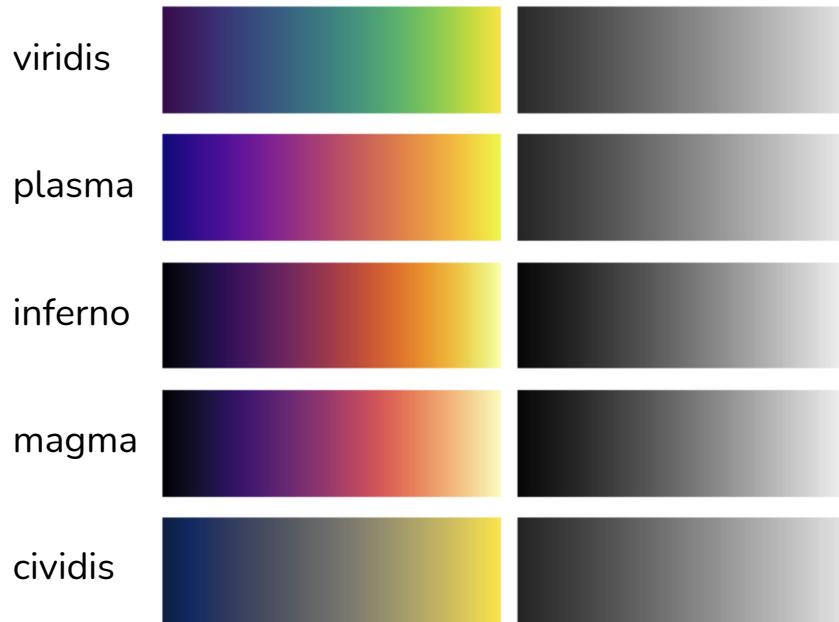
Sky



# Your Tools: Colors

What does it look like in greyscale? Is it colorblind safe?

Perceptually uniform sequential colormaps



<https://www.color-blindness.com/coblis-color-blindness-simulator/>

<https://colorbrewer2.org/>

<https://coolors.co/>

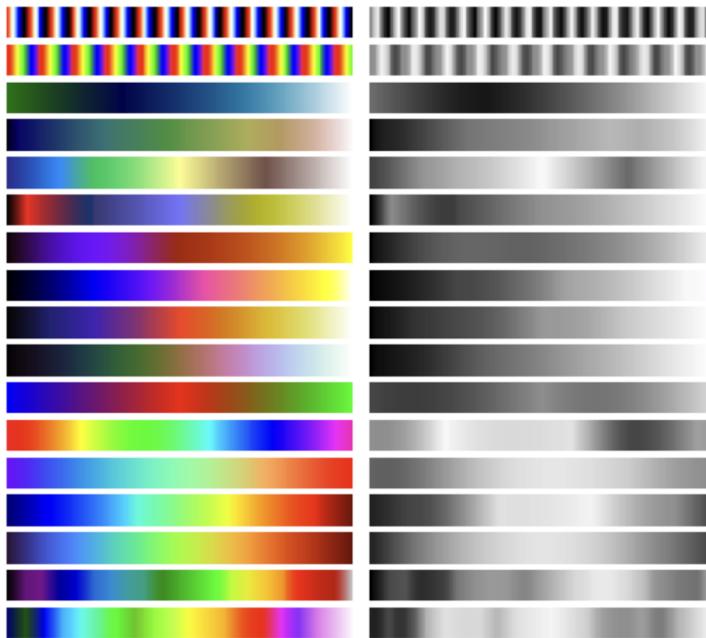
[matplotlib](#)



# Your Tools: Colors

What does it look like in greyscale? Is it colorblind safe?

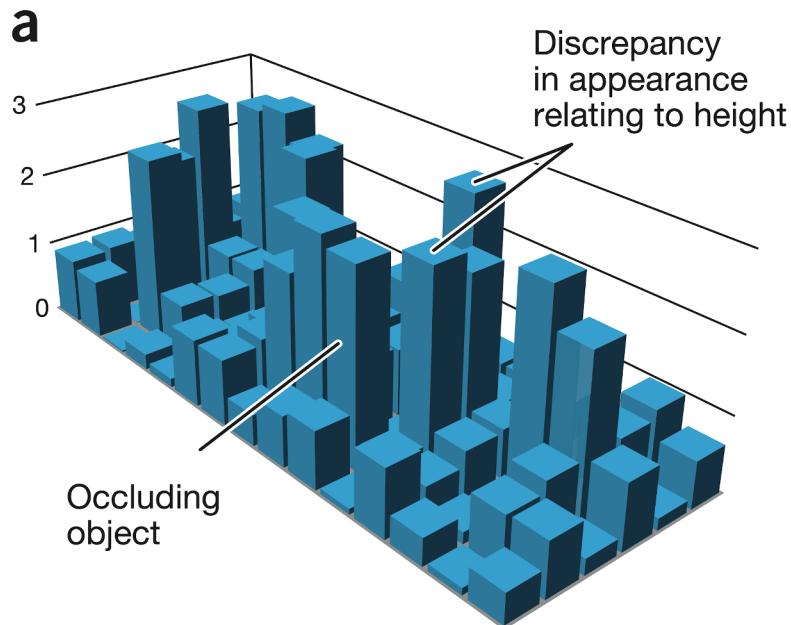
There are LOTS of (often very bad) colormaps out there. Be careful!





# Your Tools: Dimensions

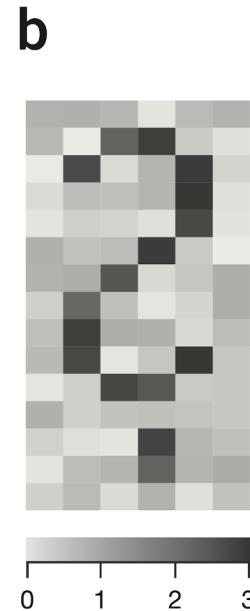
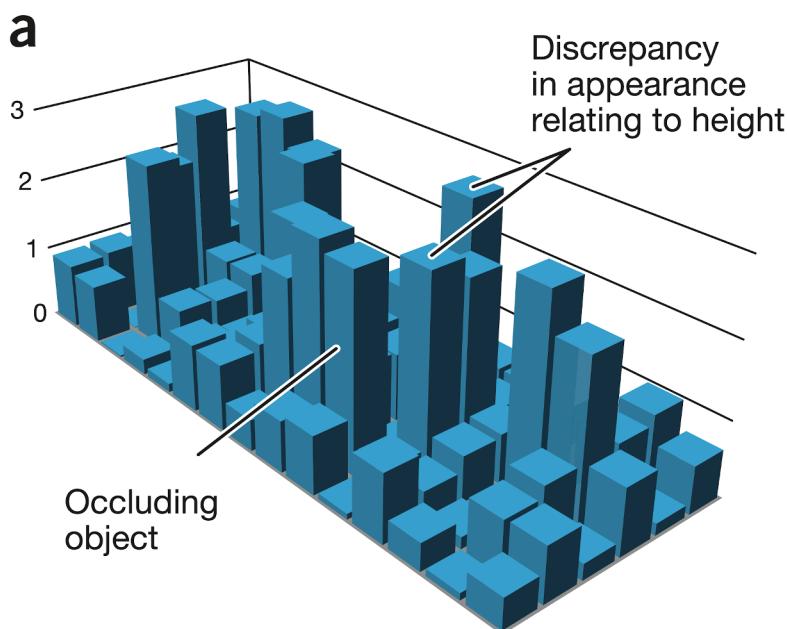
Stick to 2D whenever possible.





# Your Tools: Dimensions

Stick to 2D whenever possible.

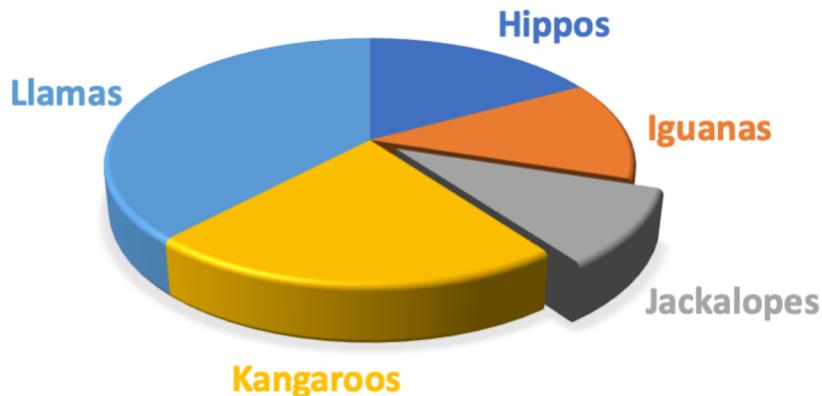


Gehlenborg & Wong (2012)

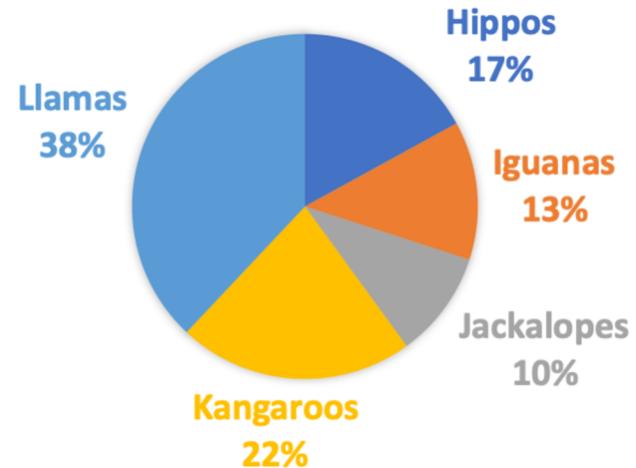


# Your Tools: Dimensions

Stick to 2D whenever possible.



No



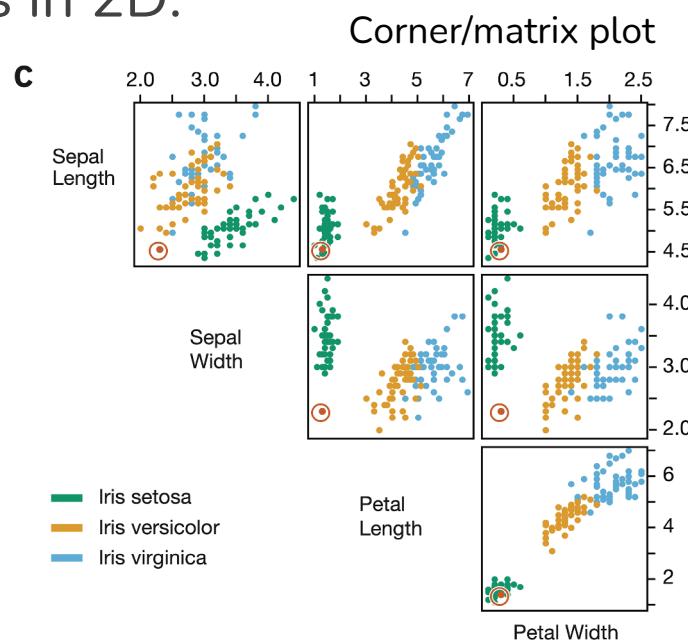
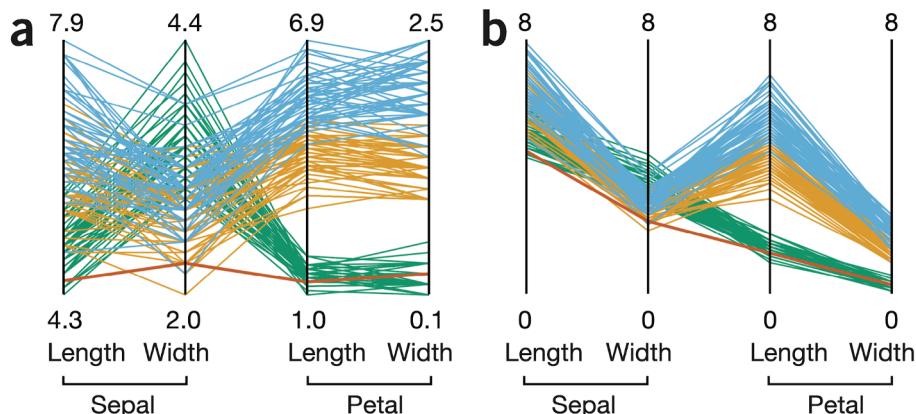
Yes



# Your Tools: Dimensions

Combine multiple dimensions in 2D.

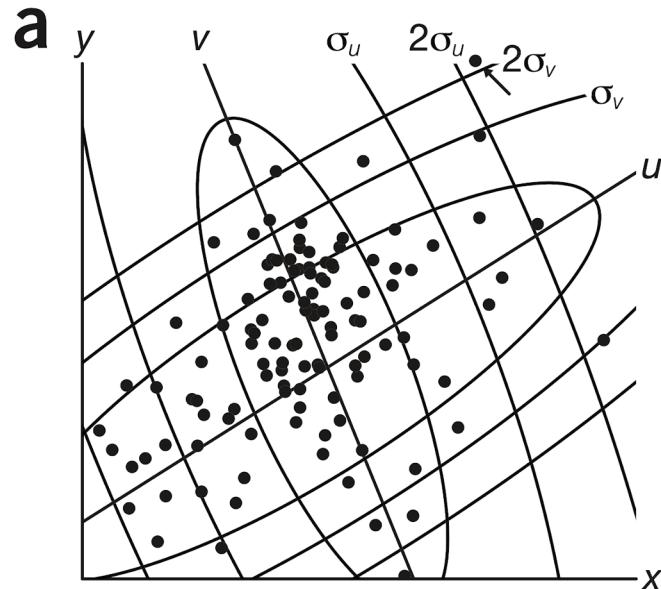
Parallel coordinate plots





# Your Tools: Axes & Grids

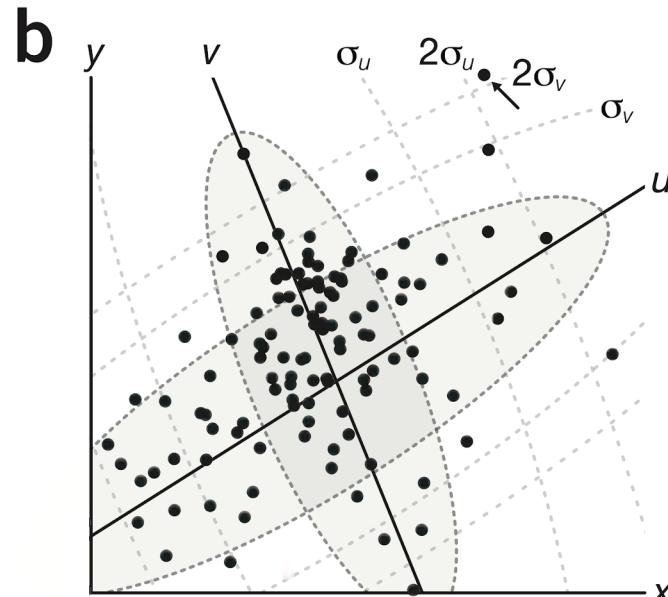
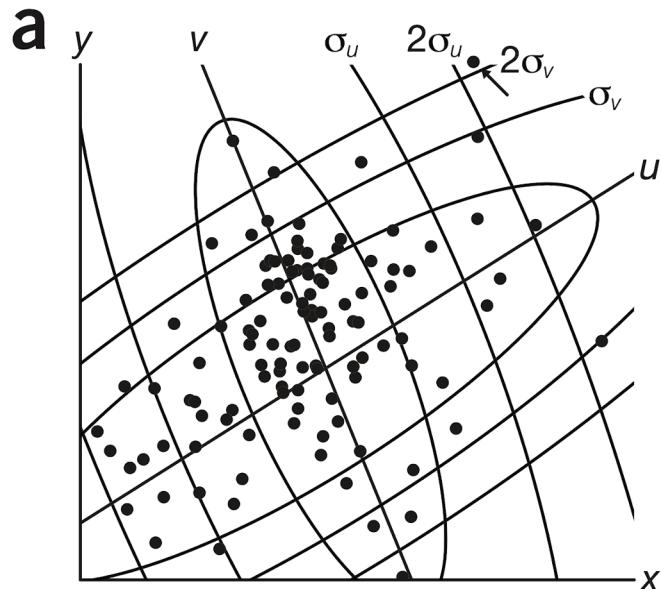
Make navigational elements visually distinct.





# Your Tools: Axes & Grids

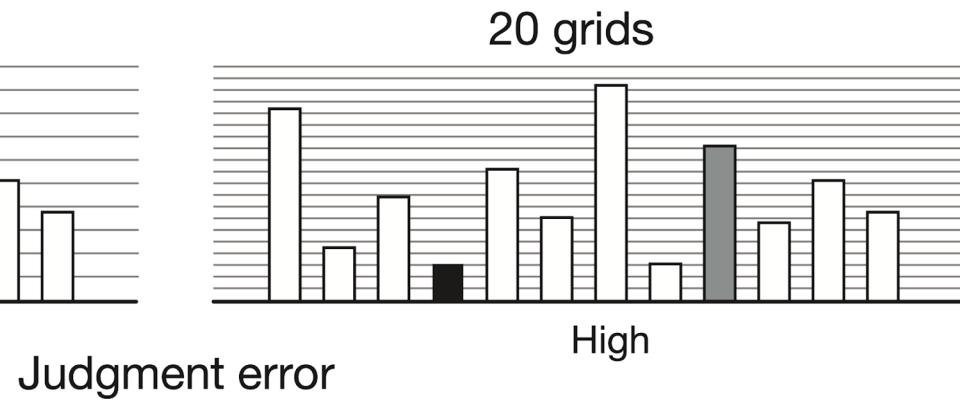
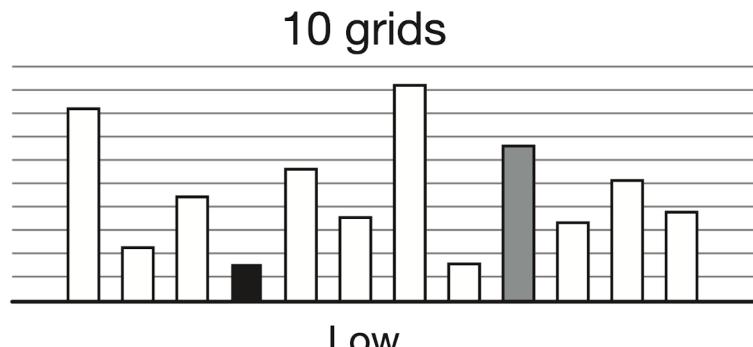
Make navigational elements visually distinct.





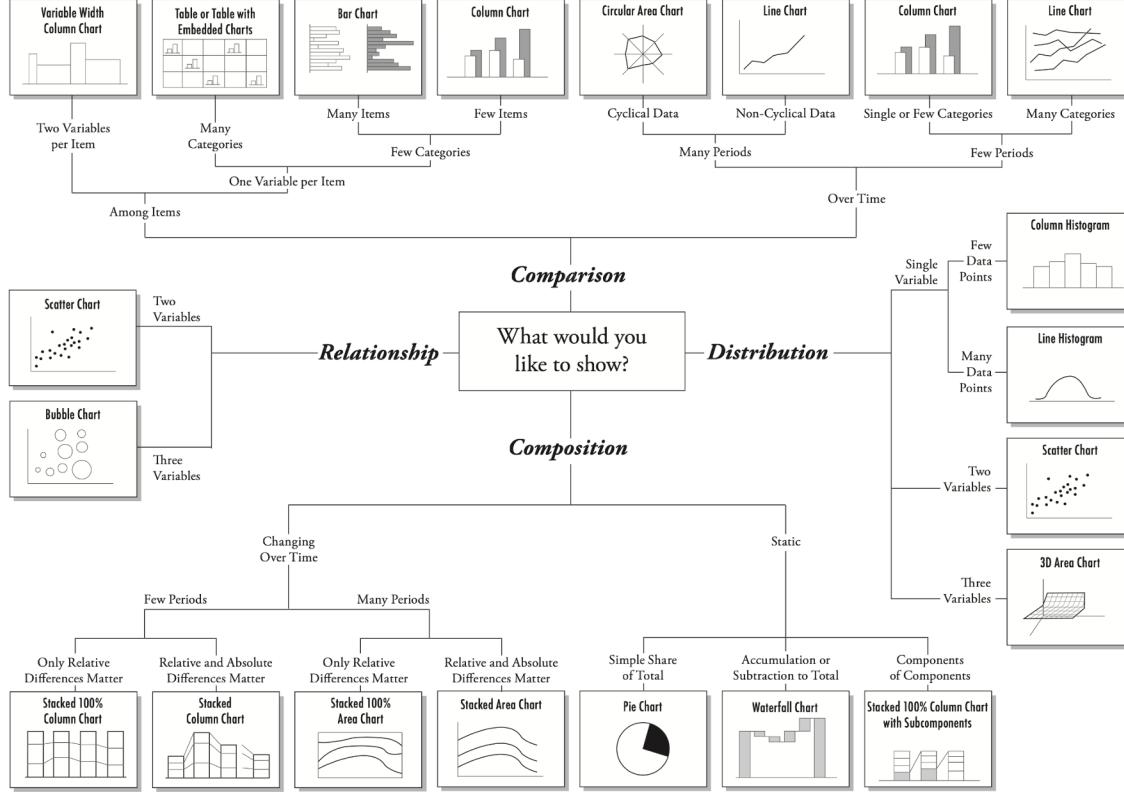
# Your Tools: Axes & Grids

Use grid lines judiciously.





# Chart Suggestions





# Summary

First, think carefully about these:

1. Choose exploratory vs. explanatory visualization.
2. Use storytelling to share your message.
3. Use your tools wisely (salience, symbols, colors, dimensions, axes/grids).
4. Carry these important suggestions in your back pocket.

Improve Data-to-Ink Ratio

Increase Efficiency

Ensure Visual Quality

Consider Accessibility

Reduce Clutter

Organize and Guide

Two great resources: [Nature Methods points of view](#), [Google material design principles](#)



# Any questions?

(I have a hands-on Python walkthrough next.)



# Hands on with Python

IntroToDataVis.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Comment Share

RAM Disk Editing

What is wrong with this figure?

Poor Usage of Characteristics in Graphs

Poor Usage of Characteristics in Graphs

Efficiency of Cognition  
Truncating Axes  
More Data per Dimension  
More Dimension per Axis  
Color Choice  
Breaking, Subverting, Soping

Let's agree that this is a monstrosity. Now, how do we improve it?

```
[1] import numpy as np
import pandas as pd
import matplotlib
from matplotlib import pyplot as plt
%matplotlib inline
```

1. Read in the data

I'm using pandas.

```
[4] url = 'https://raw.githubusercontent.com/ageller/IDEAS_FSS-Vis/master/matplotlib/bar/bar.csv'
data = pd.read_csv(url)
data
```

Label	Value
Encoding Visual Meaning	10.0
Truncating Axes	3.0
Data to Ink Ratio	9.5
Efficiency of Cognition	9.0

[Click here to view notebook.](#)