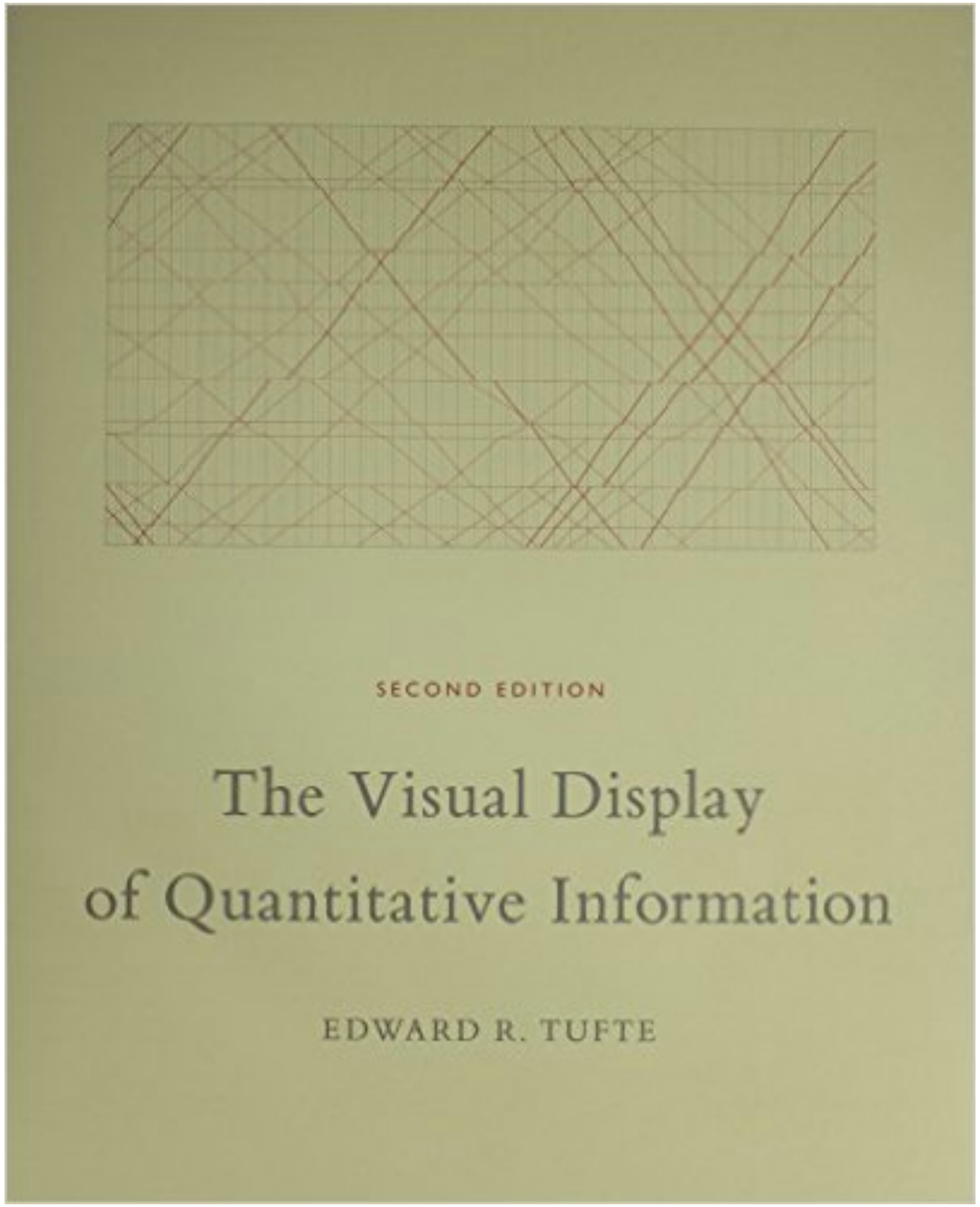
Improving Figure Clarity

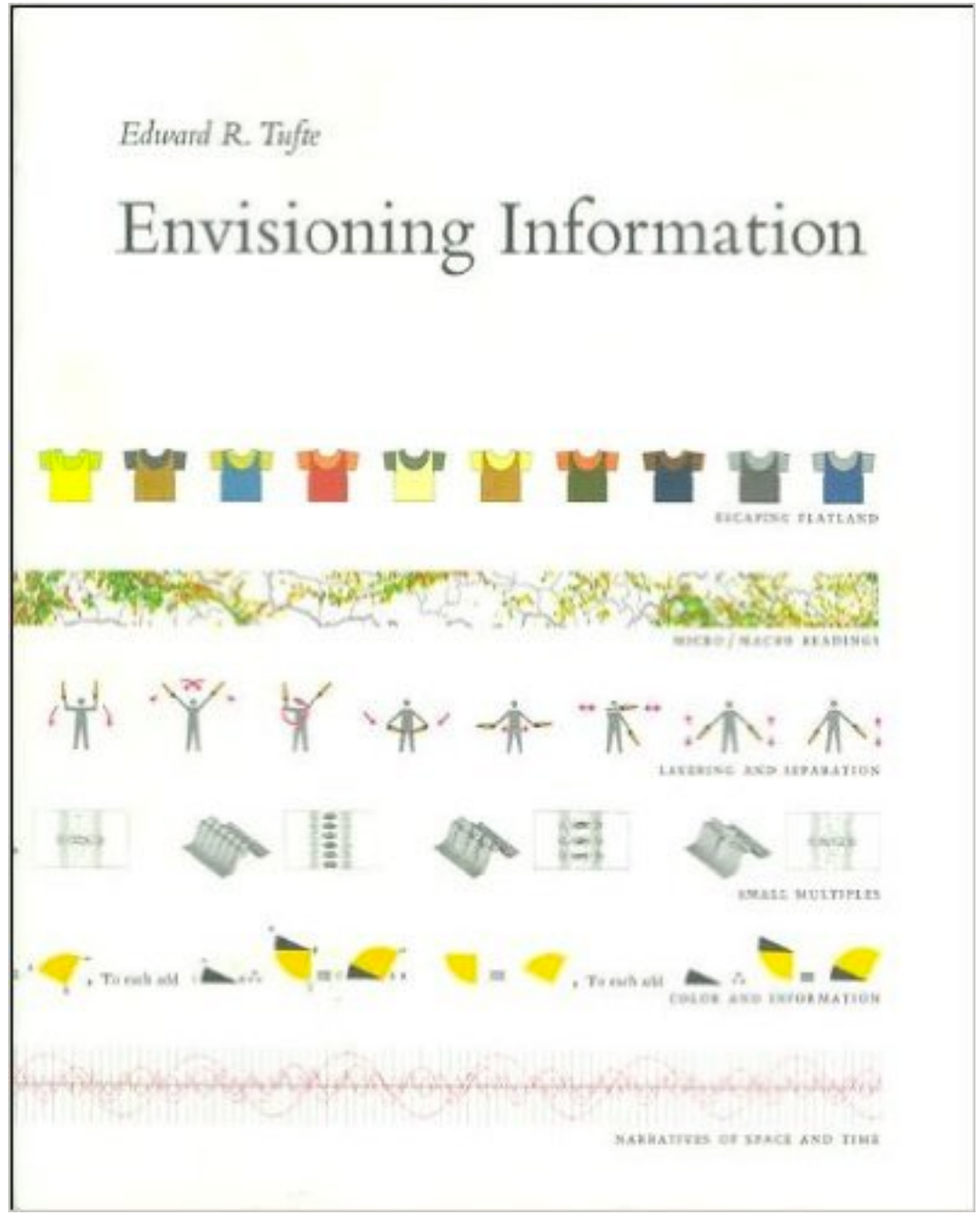
Steven Braun

Data Analytics and Visualization Specialist

May 9, 2017

Simplify charts and graphs
to aid communication


SECOND EDITION
The Visual Display
of Quantitative Information
EDWARD R. TUFTE


Edward R. Tufte
Envisioning Information
ESCAPING FLATLAND
MICRO/MACRO READINGS
LAYERING AND SEPARATION
SMALL MULTIPLES
COLOR AND INFORMATION
NARRATIVES OF SPACE AND TIME

GRAPHICAL INTEGRITY

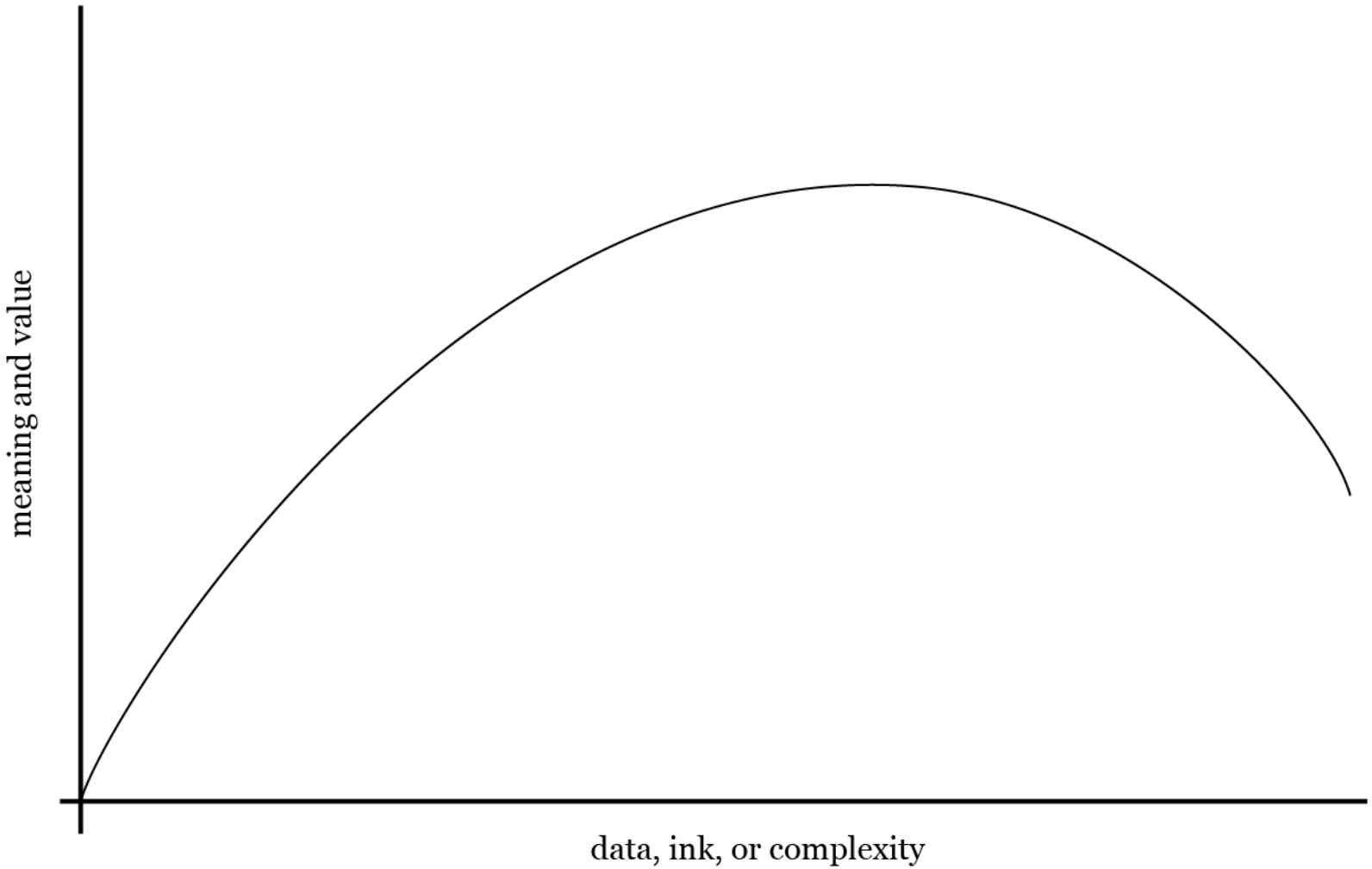
"Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data."

Edward Tufte, "Visual Display of Quantitative Information"

TUFTE'S ROLES OF GRAPHICAL DISPLAYS

1. Show the data
2. Avoid distorting what the data have to say
3. Encourage comparisons
4. Reveal the data at several levels of detail
5. Serve a reasonably clear purpose
6. Be closely integrated with the statistical and verbal descriptions

$$\text{“data-ink ratio”} = \frac{\text{data ink}}{\text{total ink used to print the graphic}}$$


meaning and value
data, ink, or complexity

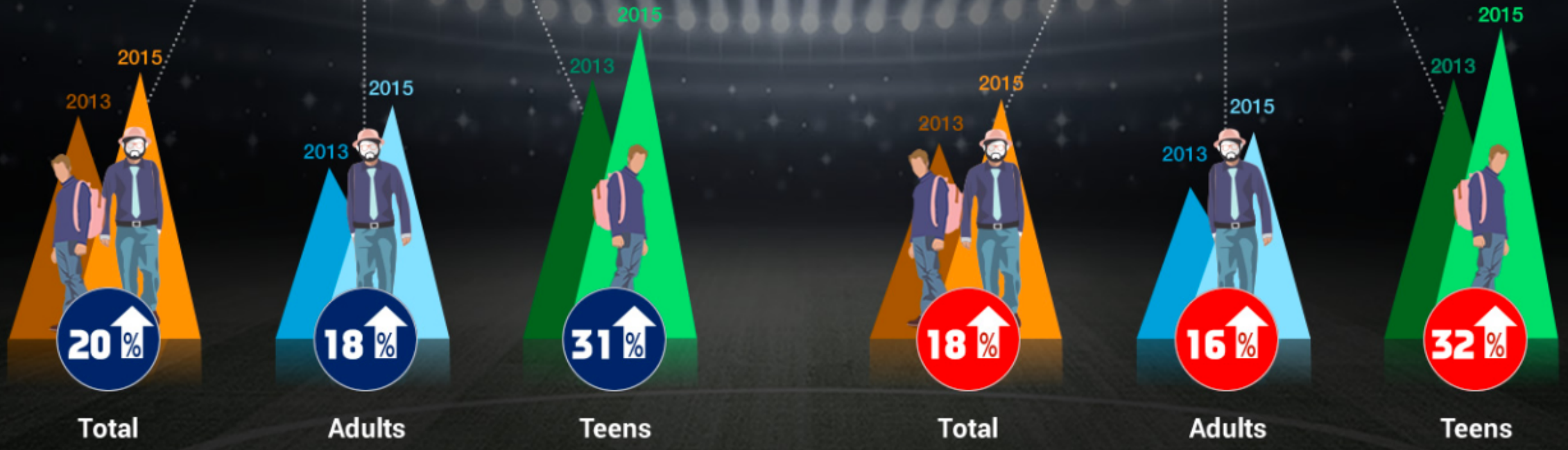
56.8 Million Fantasy Players

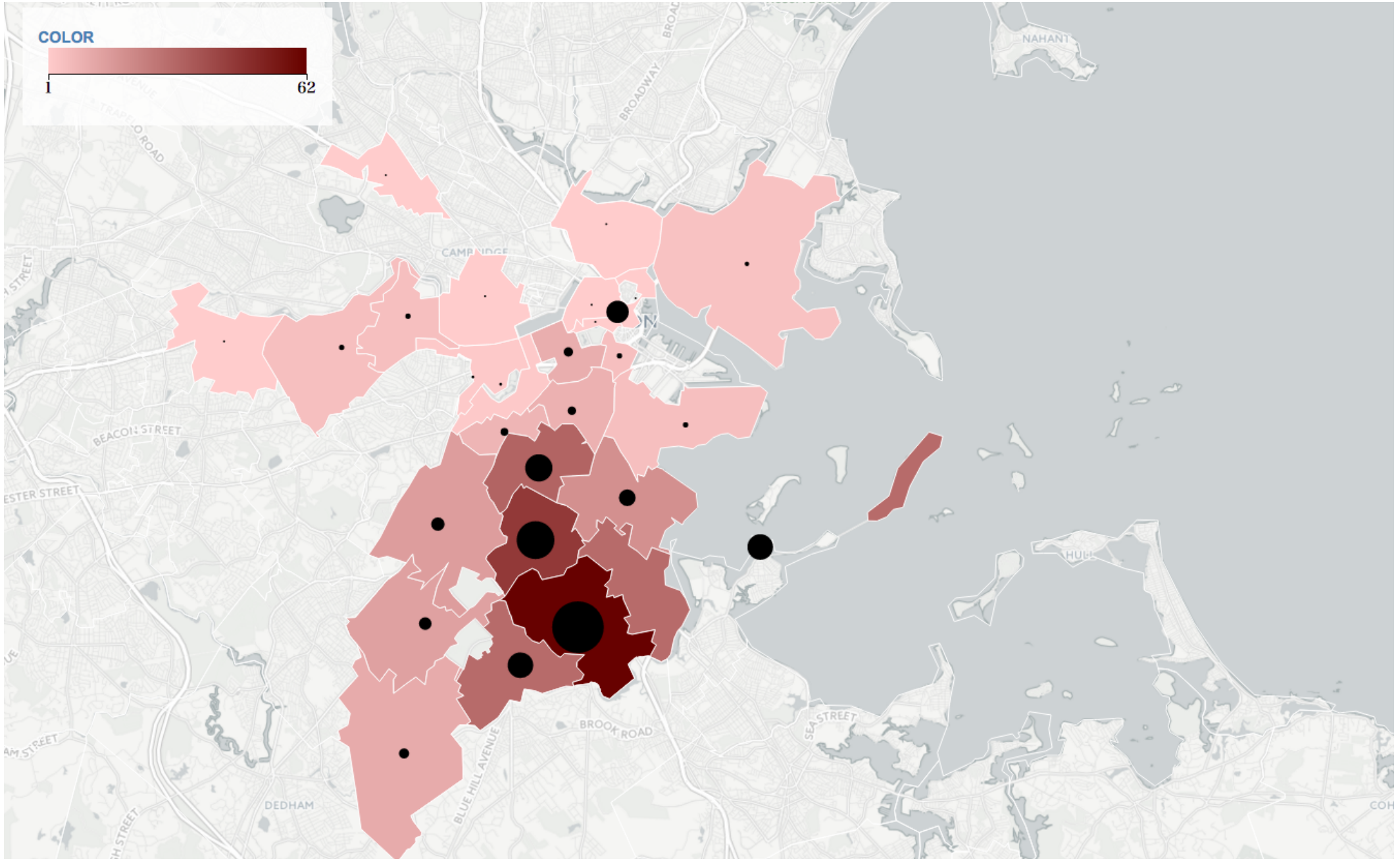
Aged 12+ in North America

51.6 million
42 million adults18+

5.2 million
4.2 million adults18+

Participation Growth Among Adults & Teens




COLOR
1
62
NAHANT
BROADWAY
CAMBRIDGE
BEACON STREET
BROOK ROAD
DEDHAM
BLUE HILL AVENUE
SEA STREET

Is data-ink ratio a good measure of effectiveness?

Sometimes.

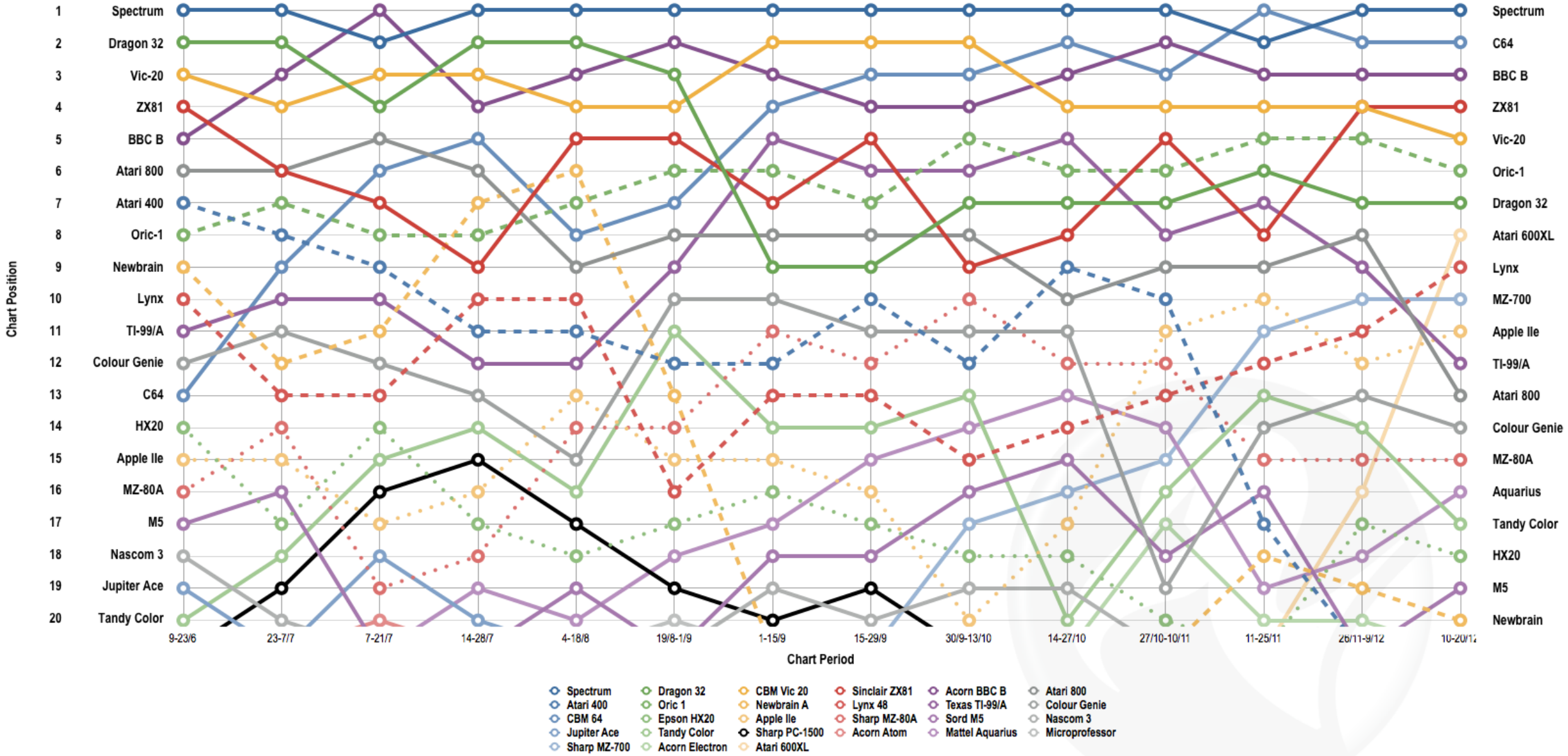
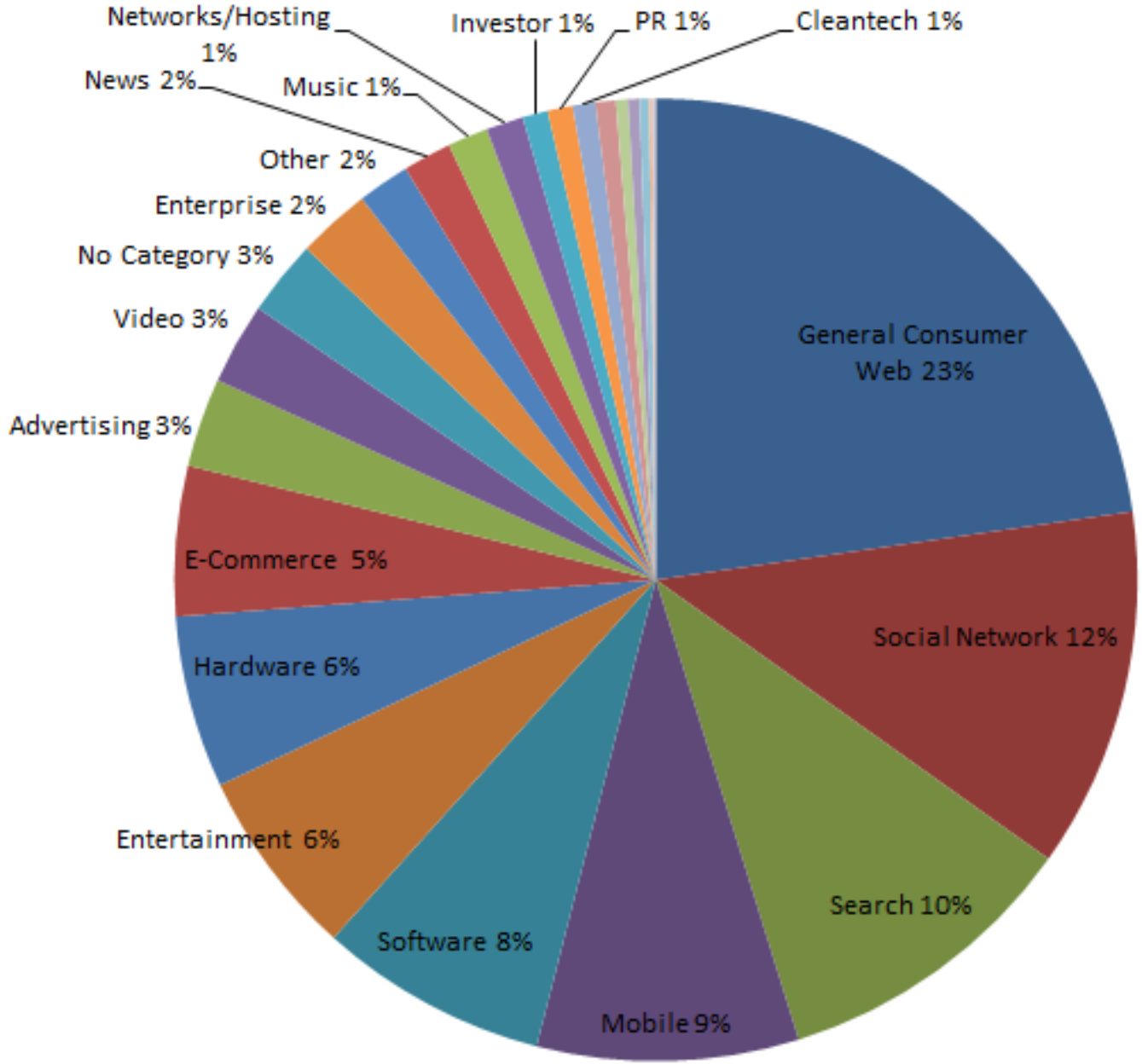
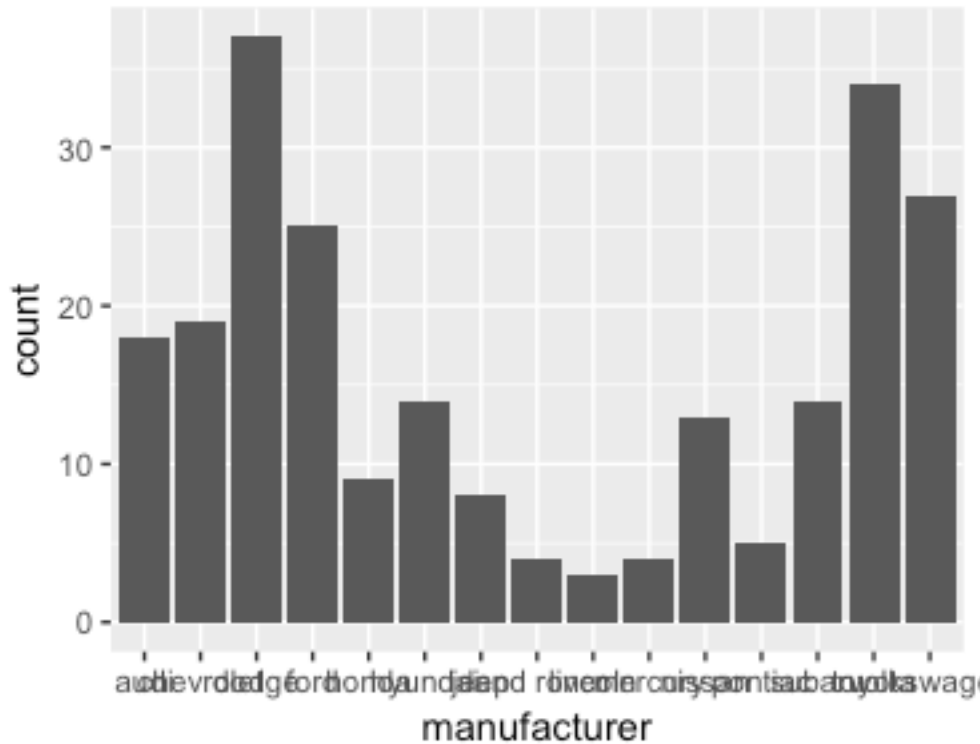
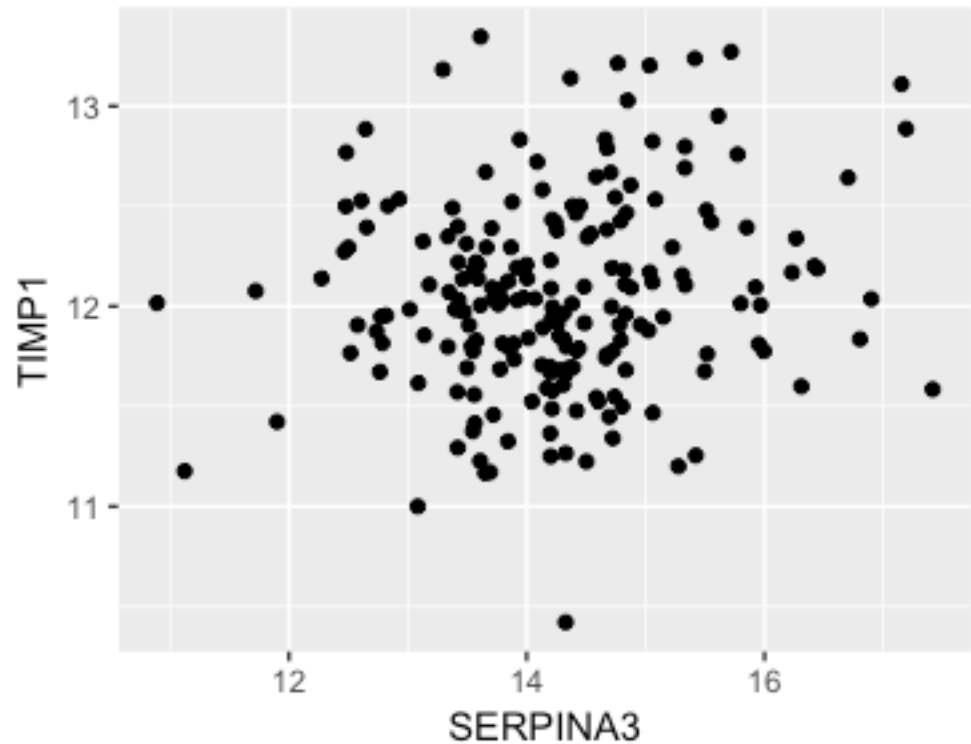


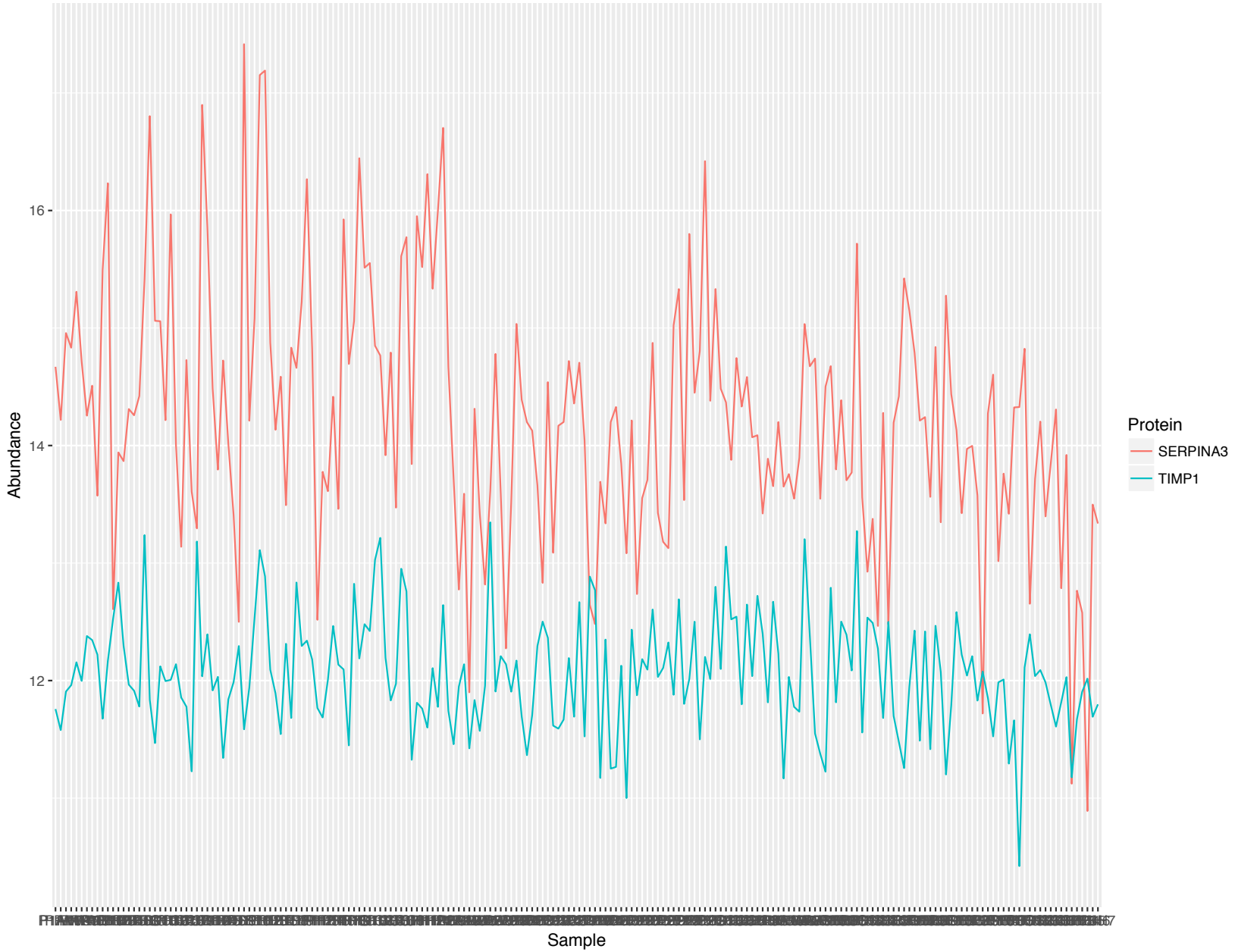
Chart Position
1 Spectrum
2 Dragon 32
3 Vic-20
4 ZX81
5 BBC B
6 Atari 800
7 Atari 400
8 Oric-1
9 Newbrain
10 Lynx
11 TI-99/A
12 Colour Genie
13 C64
14 HX20
15 Apple IIe
16 MZ-80A
17 M5
18 Nascom 3
19 Jupiter Ace
20 Tandy Color
Spectrum
C64
BBC B
ZX81
Vic-20
Oric-1
Dragon 32
Atari 600XL
Lynx
MZ-700
Apple IIe
TI-99/A
Atari 800
Colour Genie
MZ-80A
Aquarius
Tandy Color
HX20
M5
Newbrain
9-23/6
23-7/7
7-21/7
14-28/7
4-18/8
19/8-1/9
1-15/9
15-29/9
30/9-13/10
14-27/10
27/10-10/11
11-25/11
26/11-9/12
10-20/12
Chart Period
Spectrum
Atari 400
CBM 64
Jupiter Ace
Sharp MZ-700
Dragon 32
Oric 1
Epson HX20
Tandy Color
Acorn Electron
CBM Vic 20
Newbrain A
Apple IIe
Sharp PC-1500
Atari 600XL
Sinclair ZX81
Lynx 48
Sharp MZ-80A
Acorn Atom
Acorn BBC B
Texas TI-99/A
Sord M5
Mattel Aquarius
Atari 800
Colour Genie
Nascom 3
Microprofessor


General Consumer Web 23%
Social Network 12%
Search 10%
Mobile 9%
Software 8%
Entertainment 6%
Hardware 6%
E-Commerce 5%
Advertising 3%
Video 3%
No Category 3%
Enterprise 2%
Other 2%
News 2%
Networks/Hosting 1%
Music 1%
Investor 1%
PR 1%
Cleantech 1%

Often, we need to work within and against **software defaults** that may impede chart simplification


30
20
10
0
count
manufacturer


13
12
11
TIMP1
12
14
16
SERPINA3


Abundance
16
14
12
Sample
Protein
SERPINA3
TIMP1

Simple Tips for Improving Basic Charts

SIMPLIFYING CHARTS AND GRAPHS

Reduce number of elements

Fewer elements means fewer bits to process

Use Gestalt principles

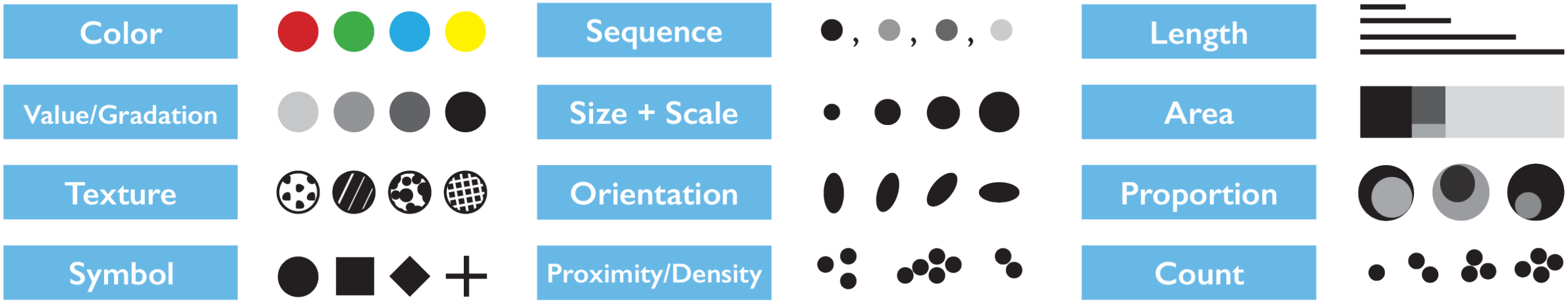
Alignment, spacing, grouping, and hierarchy

Use strong visual encodings

Position, length, proportion, hue, and more

Use contrasting visual encodings to
enhance the salience
of particular data elements

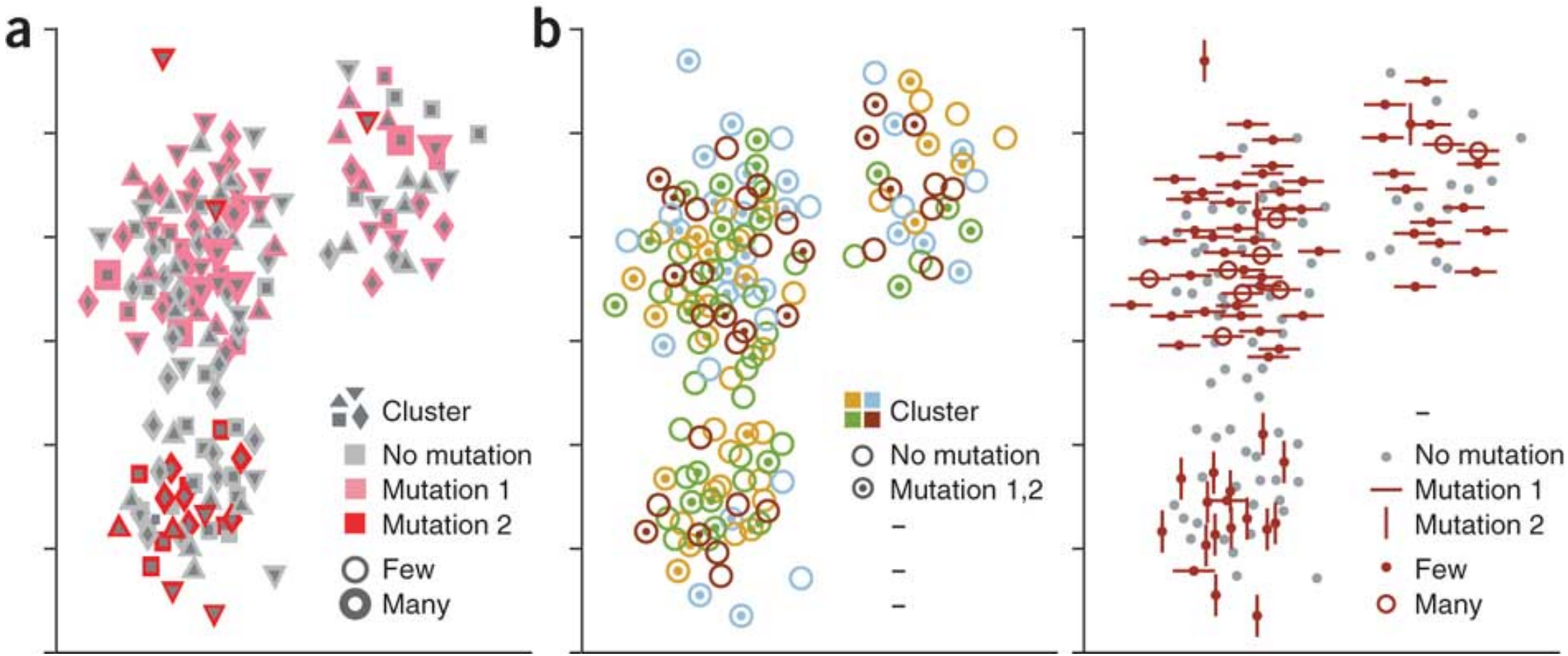
VISUAL ENCODINGS



Instances of **A** are easy to find, but instances of **P** are not

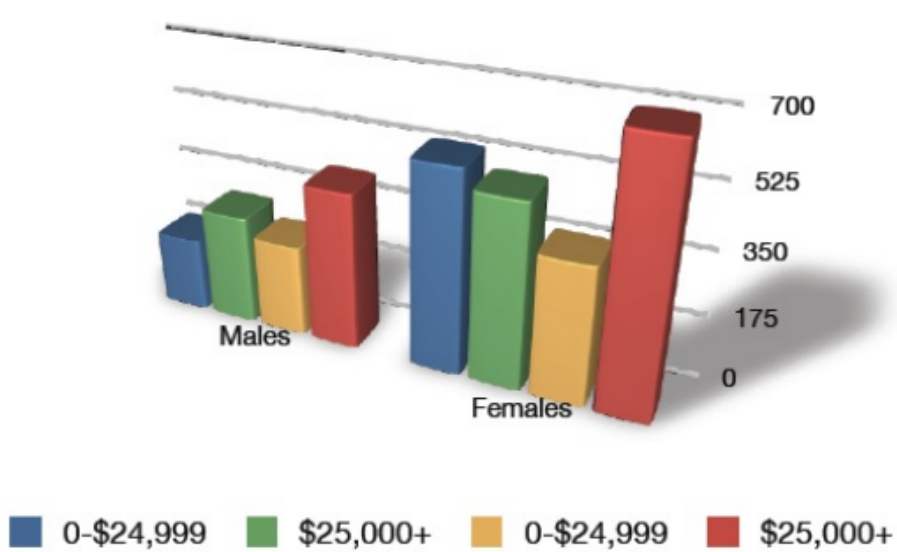
Wong, Bang. Points of View: Salience. *Nature Methods* 7, 773 (2010)

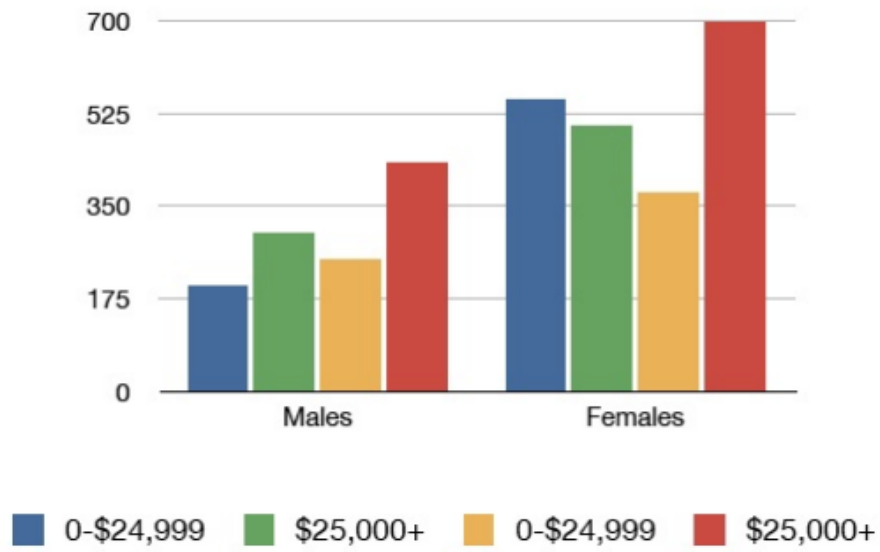
Redundant encodings can aid perception, but visual conjunctive searches come at a cost



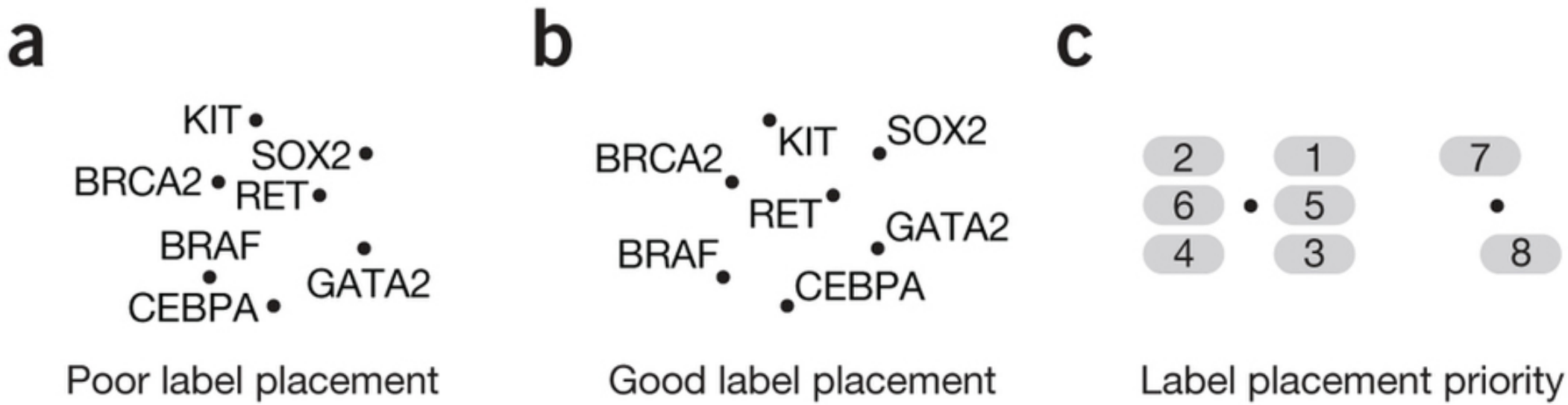
Wong, Bang. Points of View: Salience. *Nature Methods* 7, 773 (2010)

Avoid any unjustified use of
3D representation

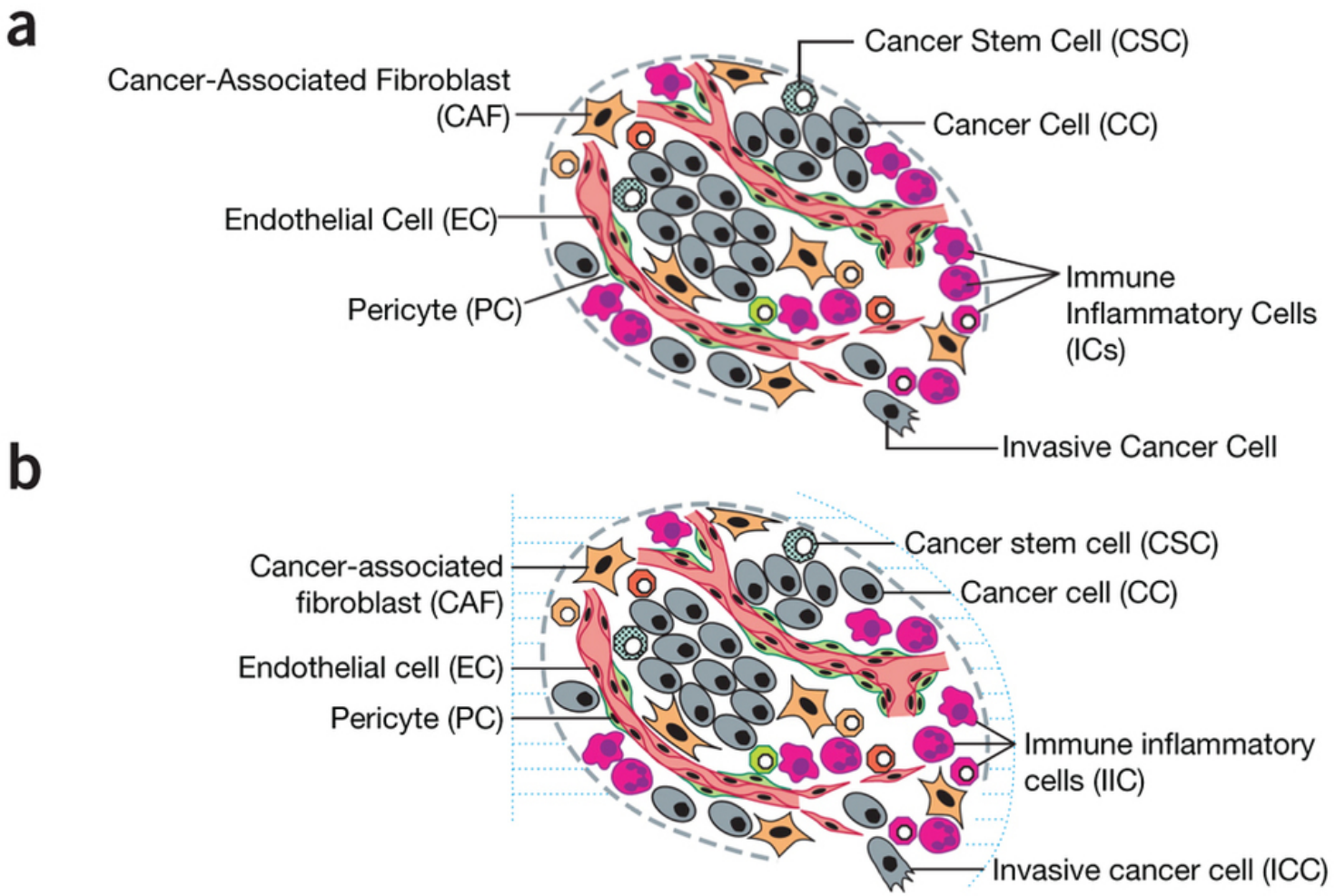

700
525
350
175
0
Males
Females
0-$24,999
$25,000+
0-$24,999
$25,000+


700
525
350
175
0
Males
Females
0-$24,999
$25,000+
0-$24,999
$25,000+

Use data labels judiciously
and position them without obscuring
important parts of the visualization

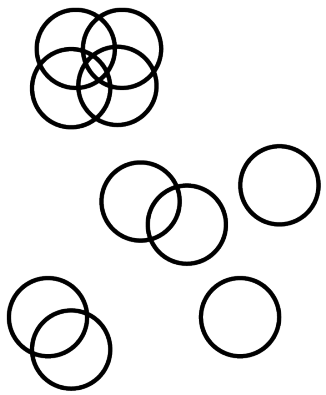


Krzywinski, Martin. Points of View: Labels and callouts. *Nature Methods* 10, 275 (2013)

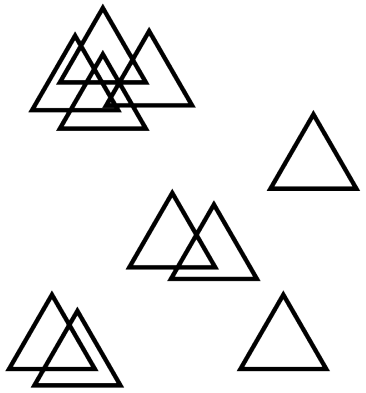


Krzywinski, Martin. Points of View: Labels and callouts. *Nature Methods* 10, 275 (2013)

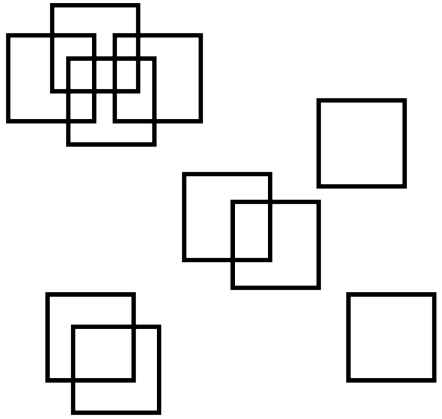
For dense plots, use open markers
instead of closed



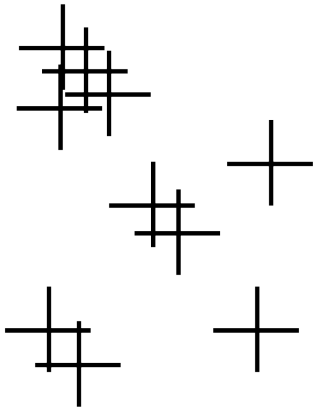
Circle



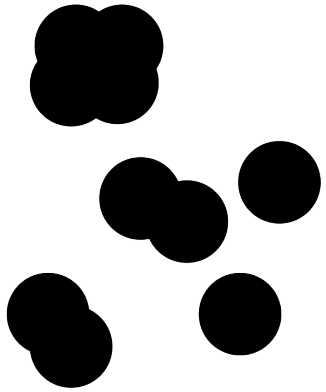
Triangle

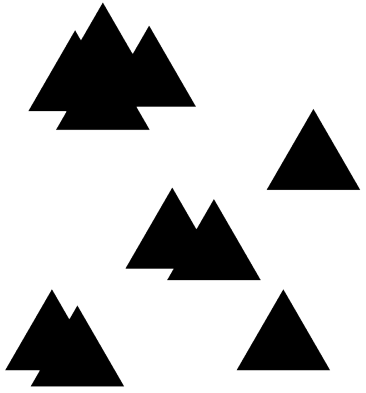


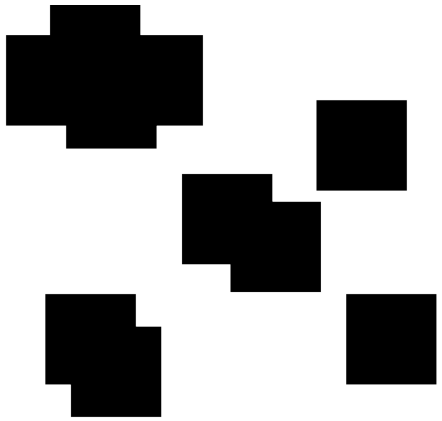
Square

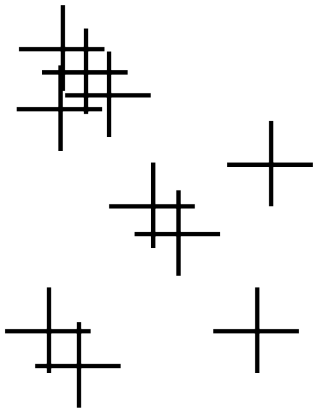


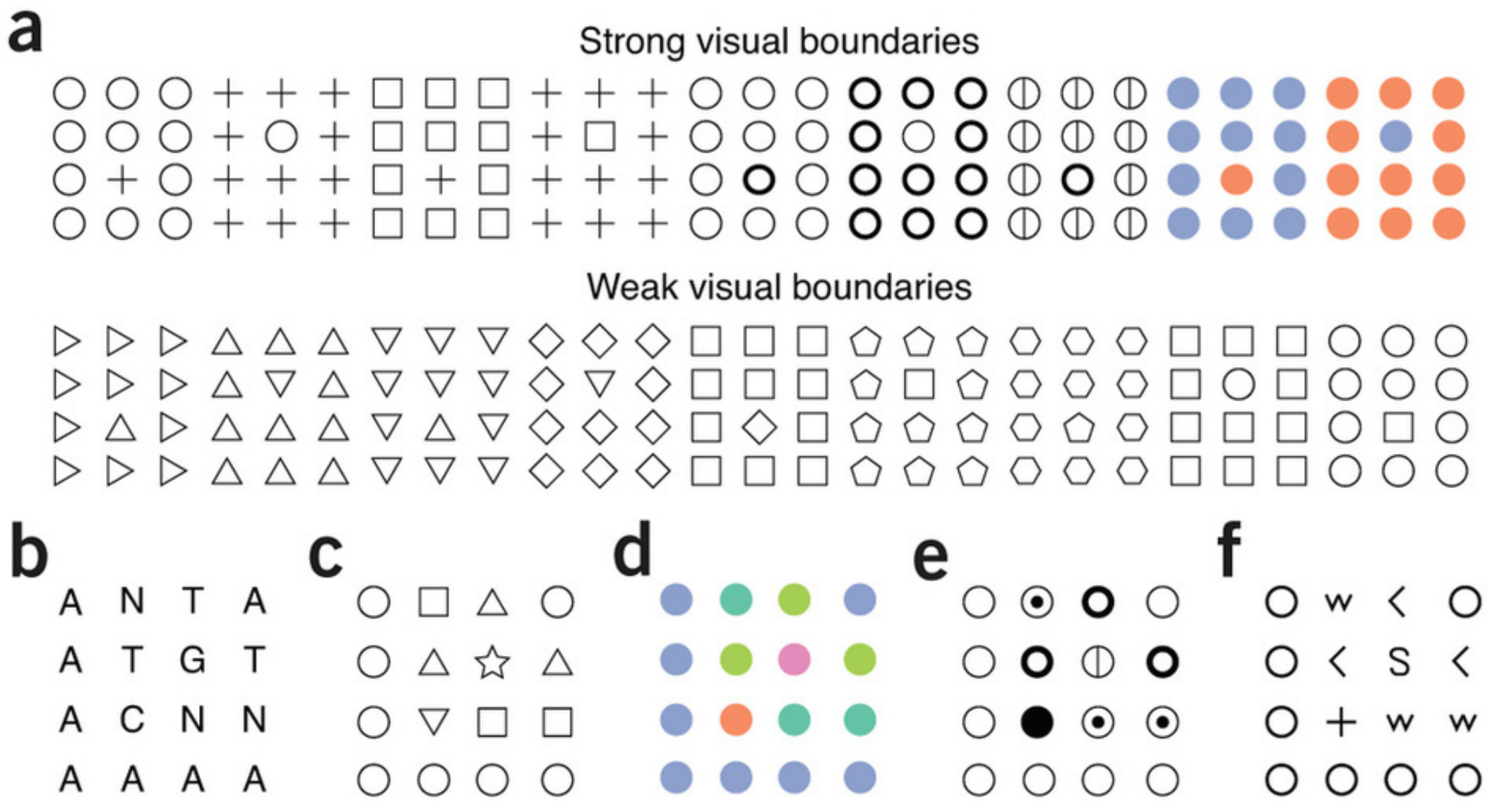
Crosshair



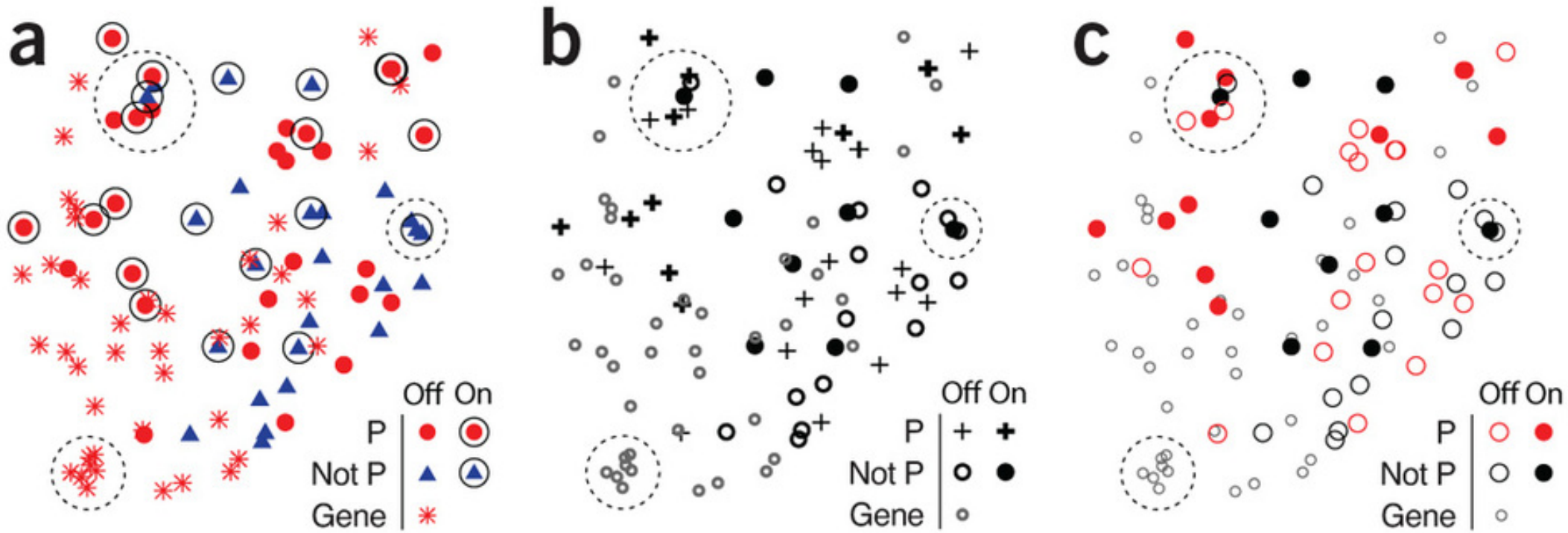






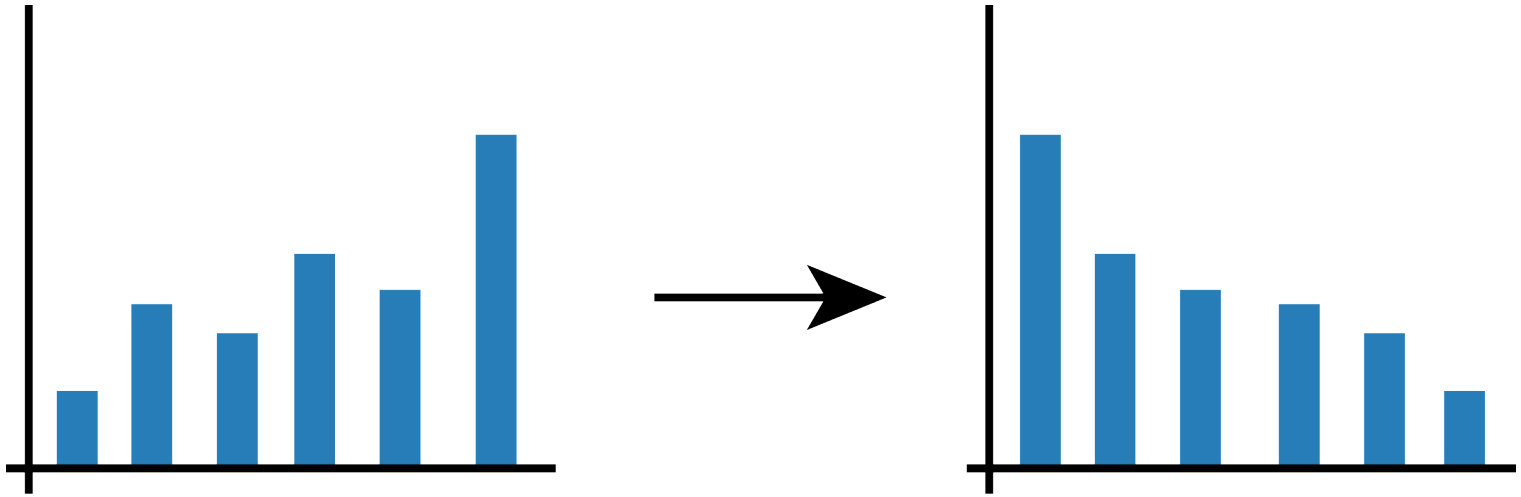


Krzywinski, Martin and Wong, Bang. Points of View: Plotting symbols. *Nature Methods* 10, 451 (2013)

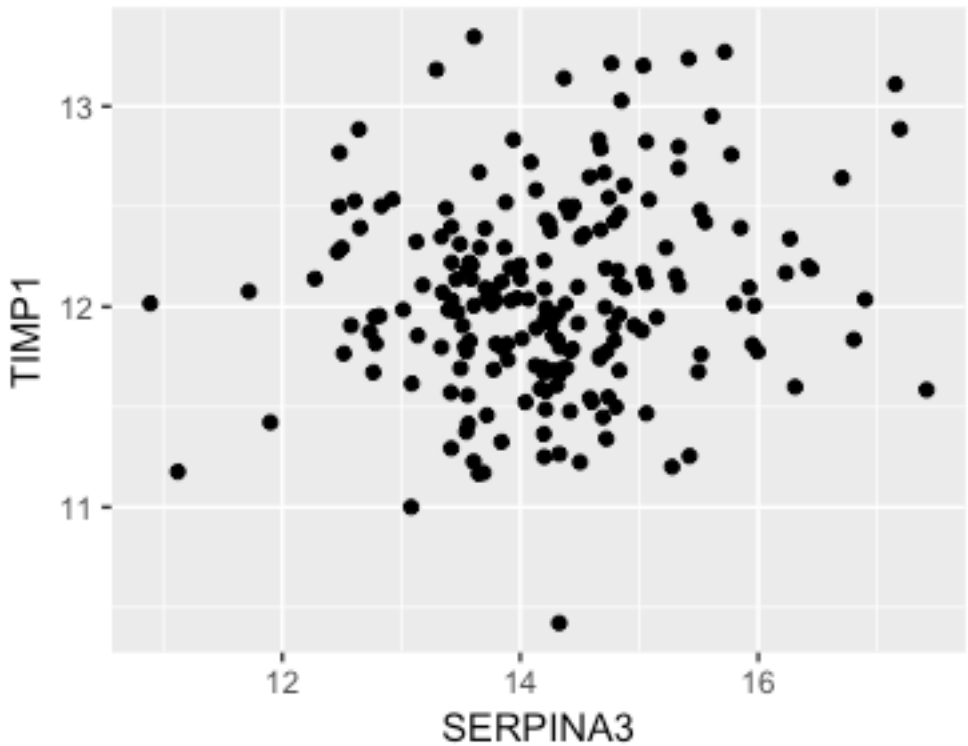


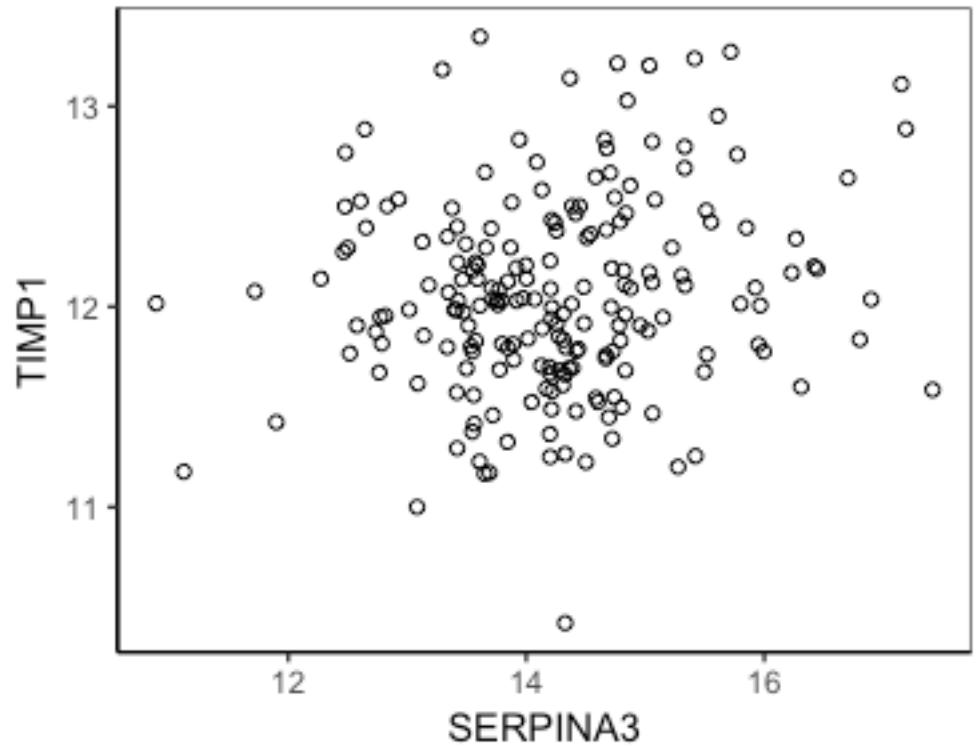
Krzywinski, Martin and Wong, Bang. Points of View: Plotting symbols. *Nature Methods* 10, 451 (2013)

Order bar charts by descending magnitude

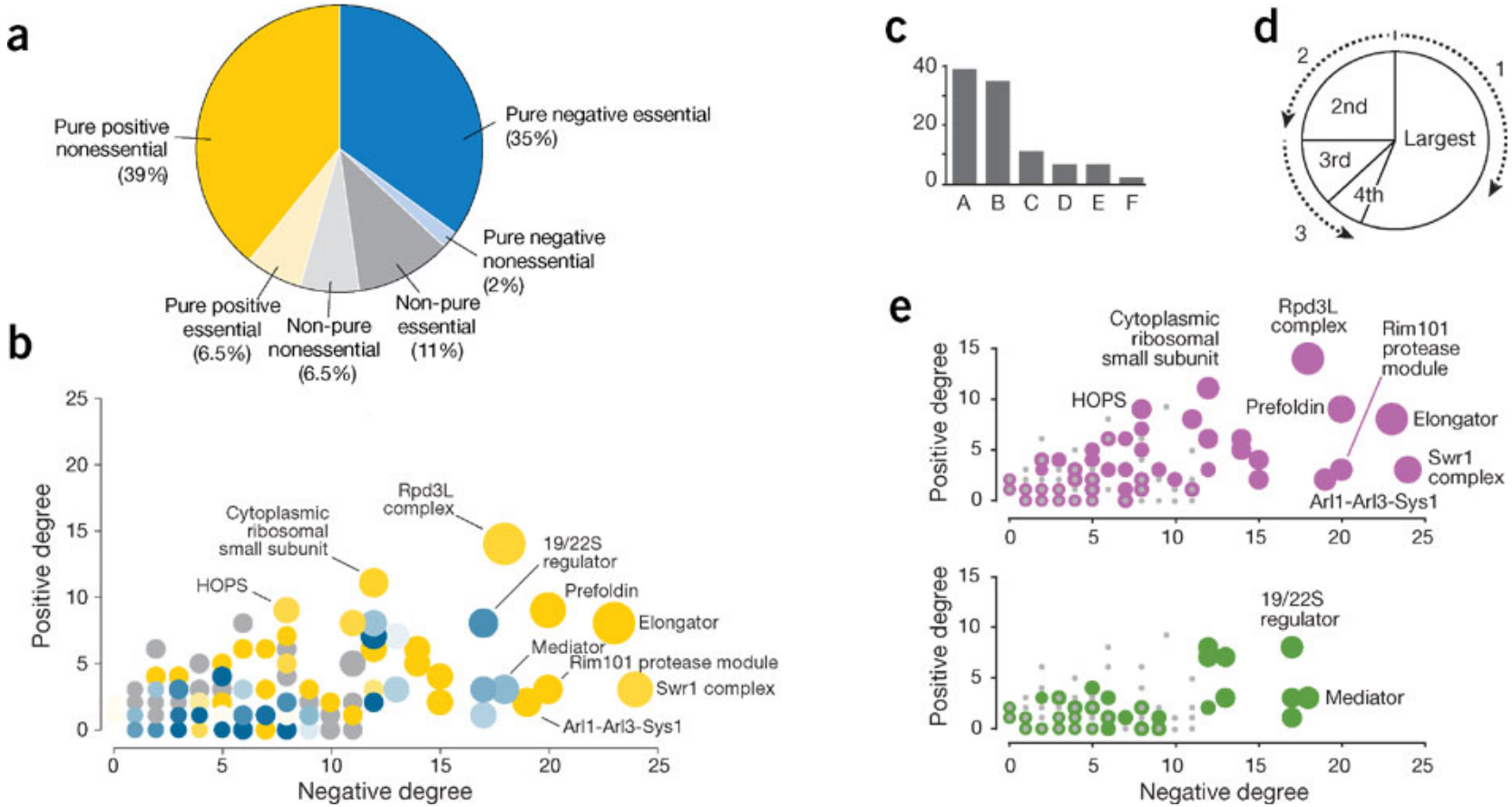


Eliminate or reduce unnecessary default elements, such as gridlines and excessive tick marks


TIMP1
13
12
11
12
14
16
SERPINA3


TIMP1
13
12
11
12
14
16
SERPINA3

Arrange the two largest wedges of a pie chart at the 12:00 position and order the remaining wedges counterclockwise in descending size



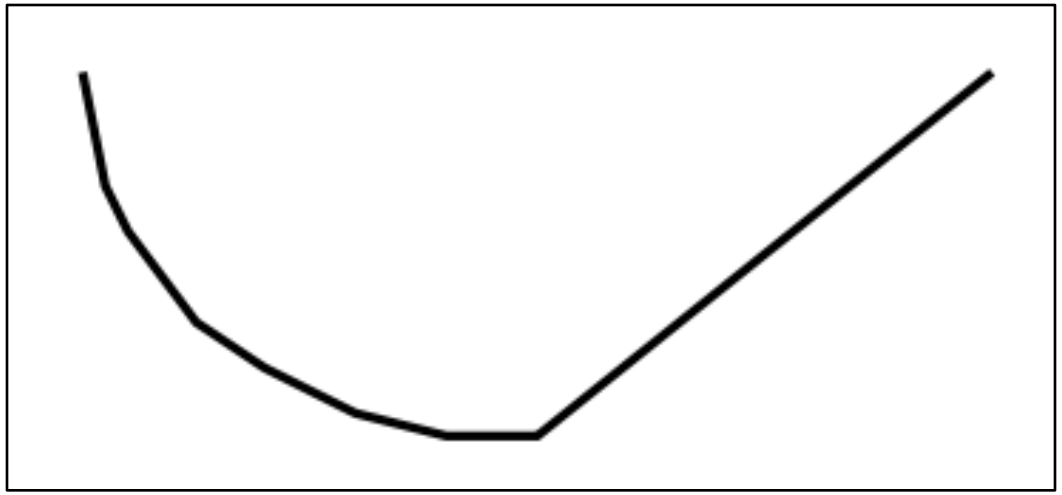
Wong, Bang. Points of View: Points of review (part 2). *Nature Methods* 8, 189 (2011)

For two-dimensional representations, the aspect ratio should be chosen to bank line segments to 45° angles

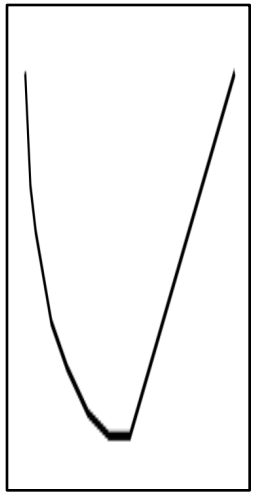
ASPECT RATIO

height / width

1:2



2:1

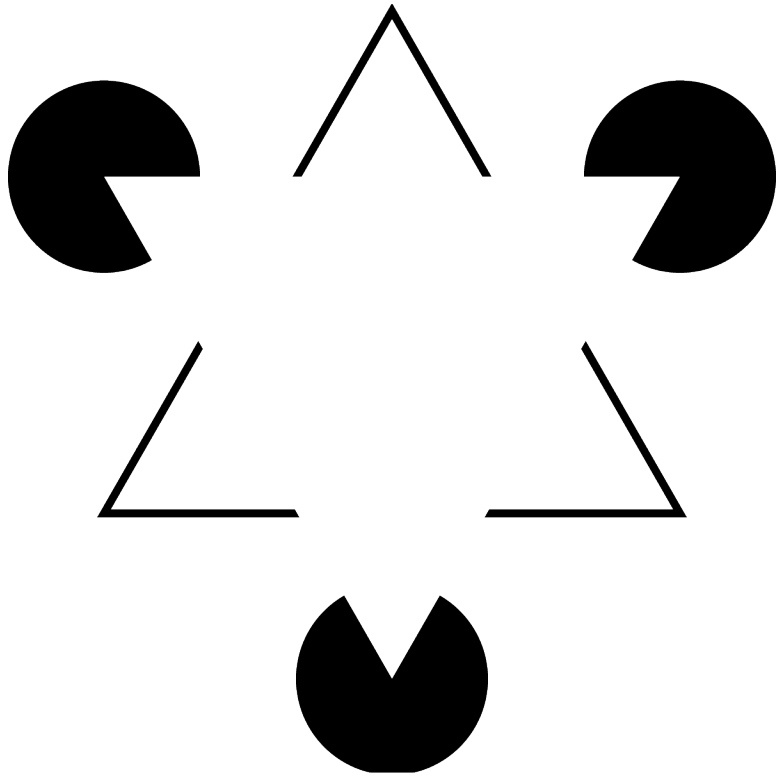


1:4

Avoid Gestalt illusions that can engender incorrect readings of a visualization

GESTALT ILLUSIONS

Closure

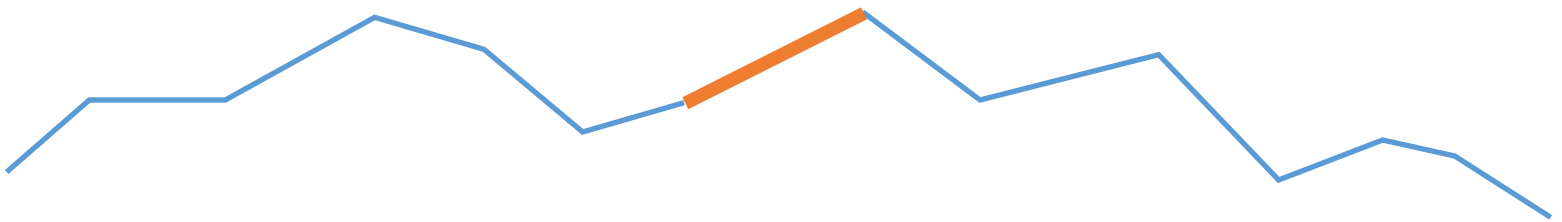


GESTALT ILLUSIONS

Closure

GESTALT ILLUSIONS

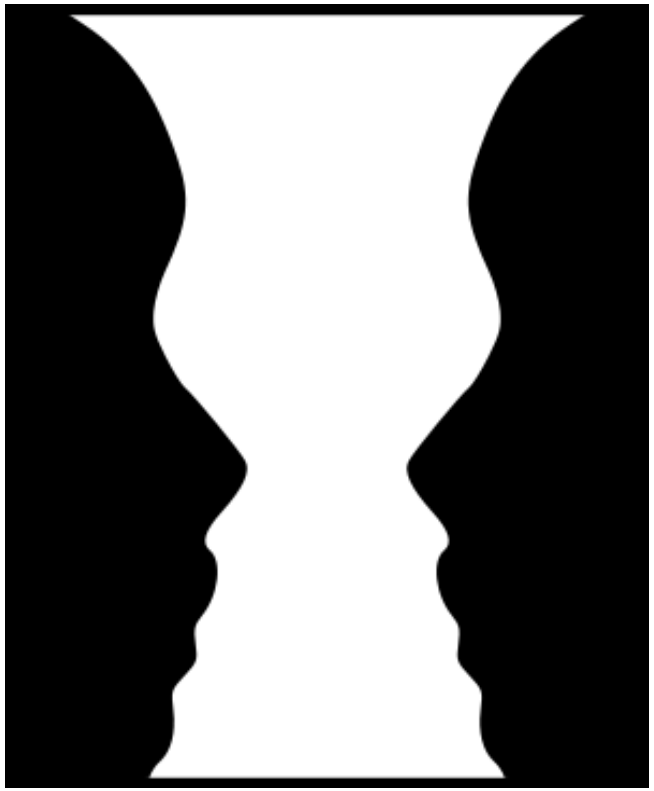
Closure

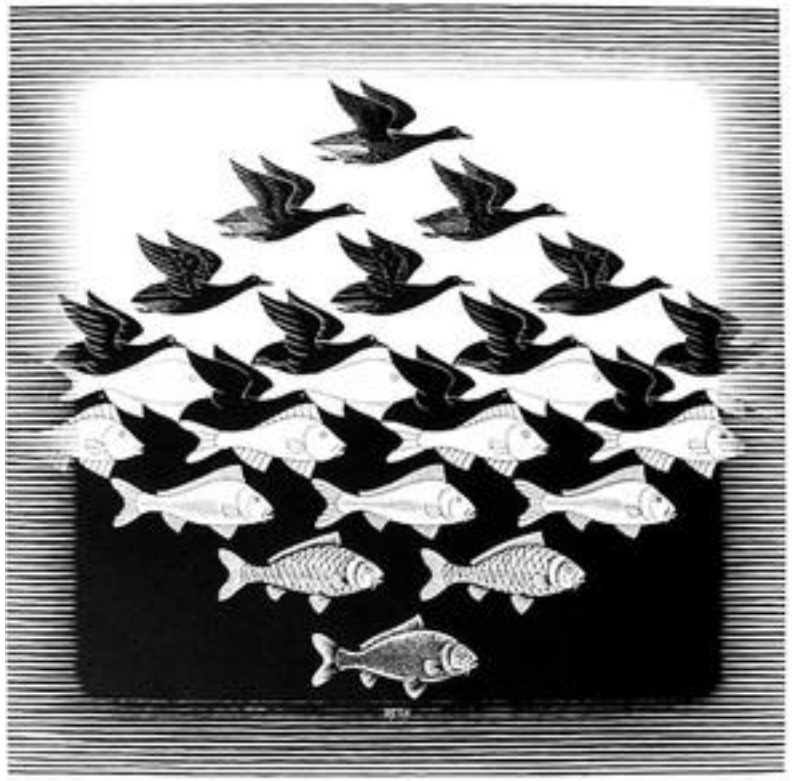


It is generally better to explicitly encode null values than leave them implied in your representation

GESTALT ILLUSIONS

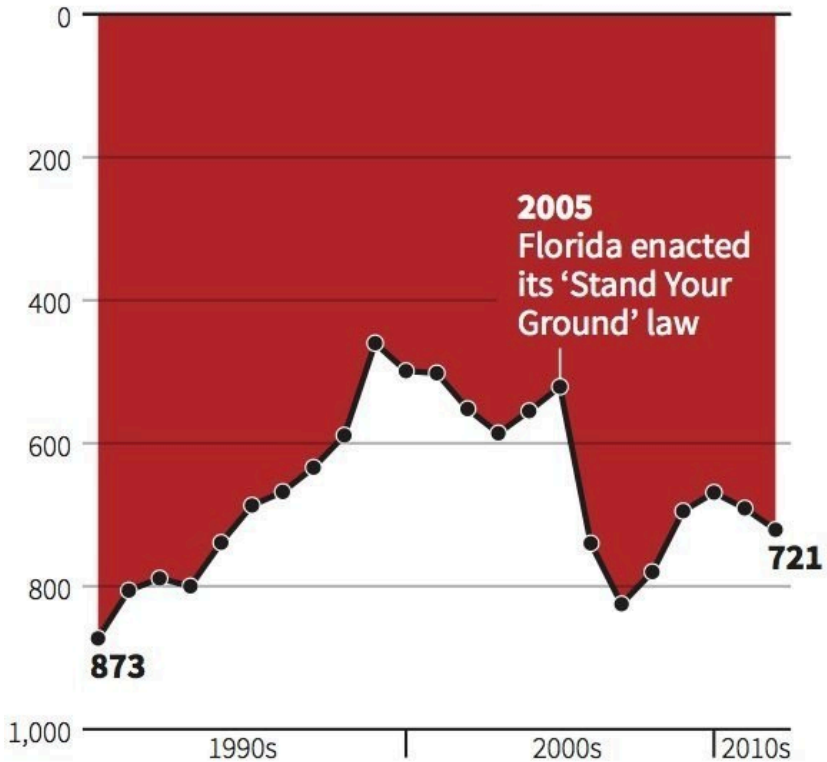
Figure/ground





Gun deaths in Florida

Number of murders committed using firearms



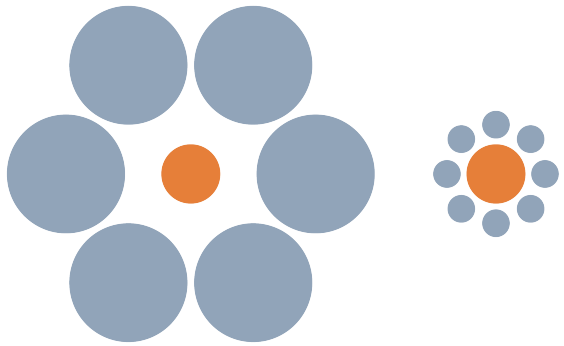
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

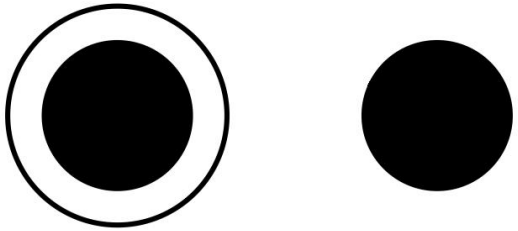
REUTERS

ILLUSIONS OF SCALE

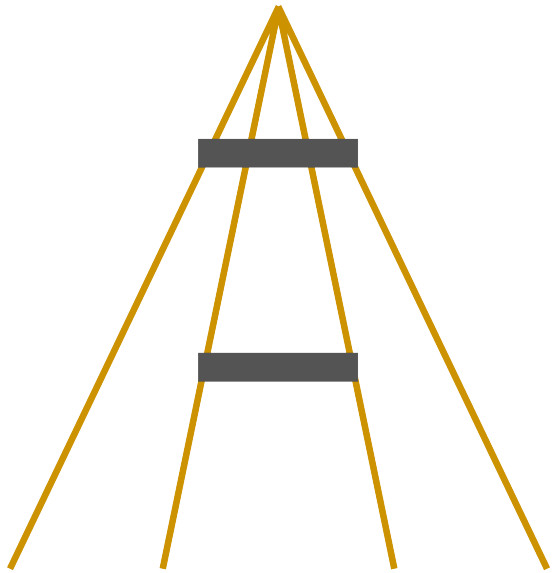
Ebbinghaus

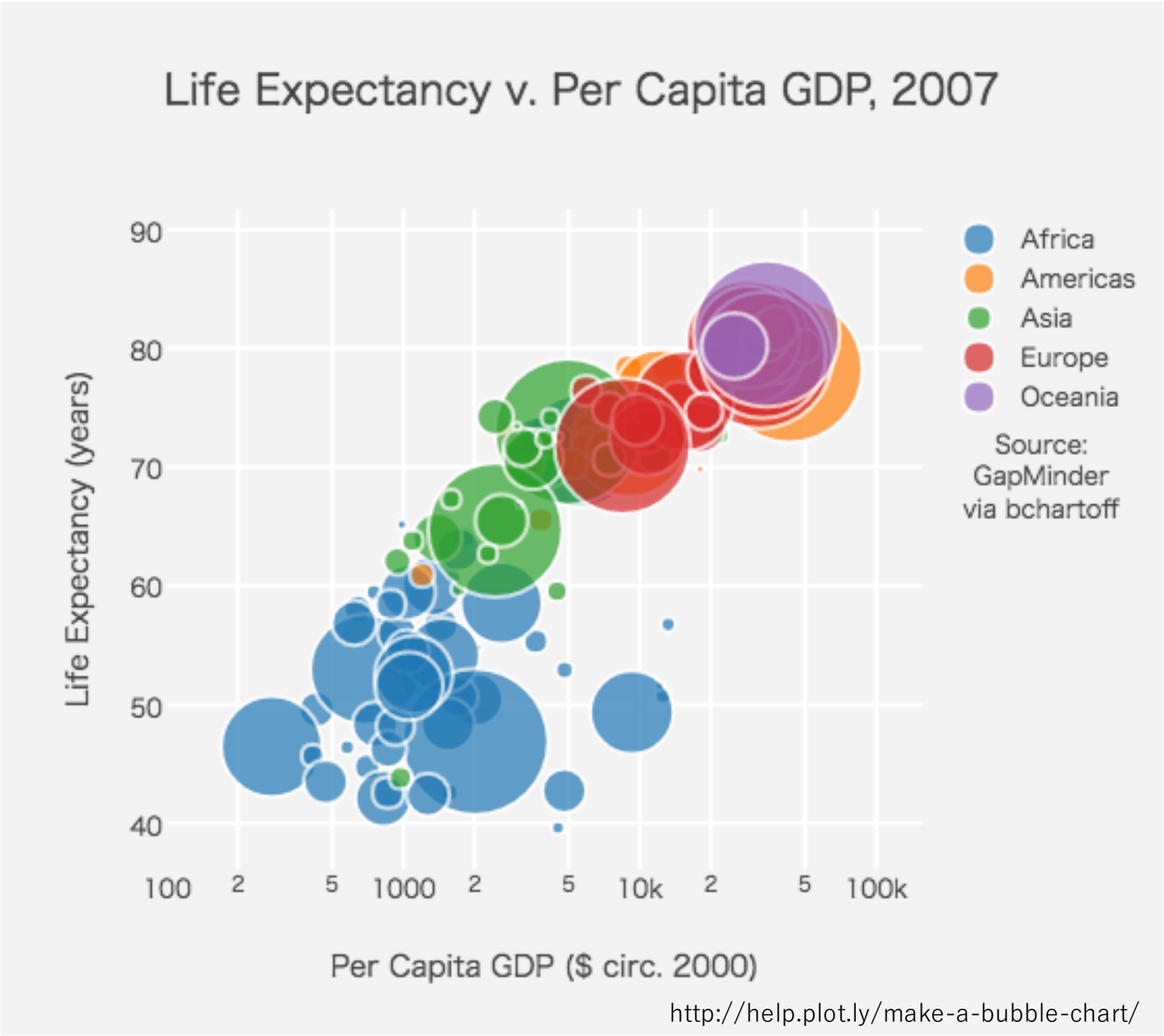


Delboeuf



Ponzo





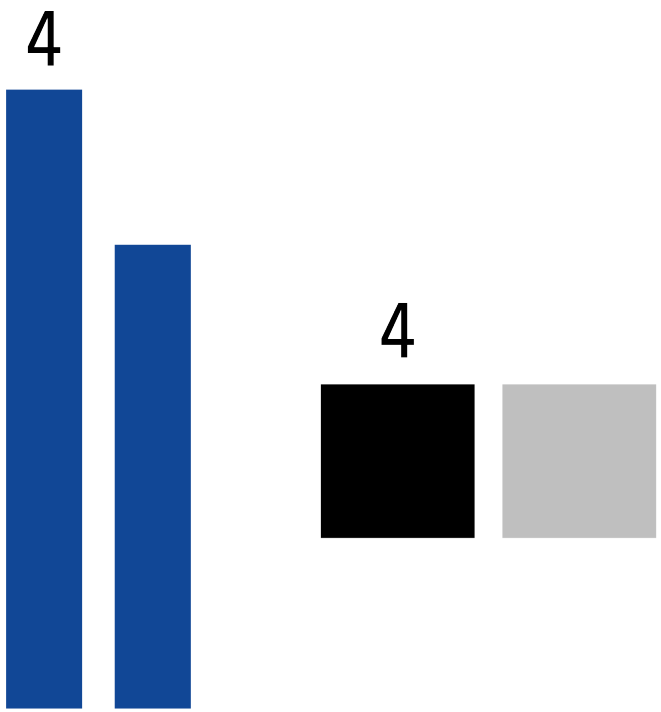
http://help.plot.ly/make-a-bubble-chart/

Use visual encodings that can be interpreted accurately

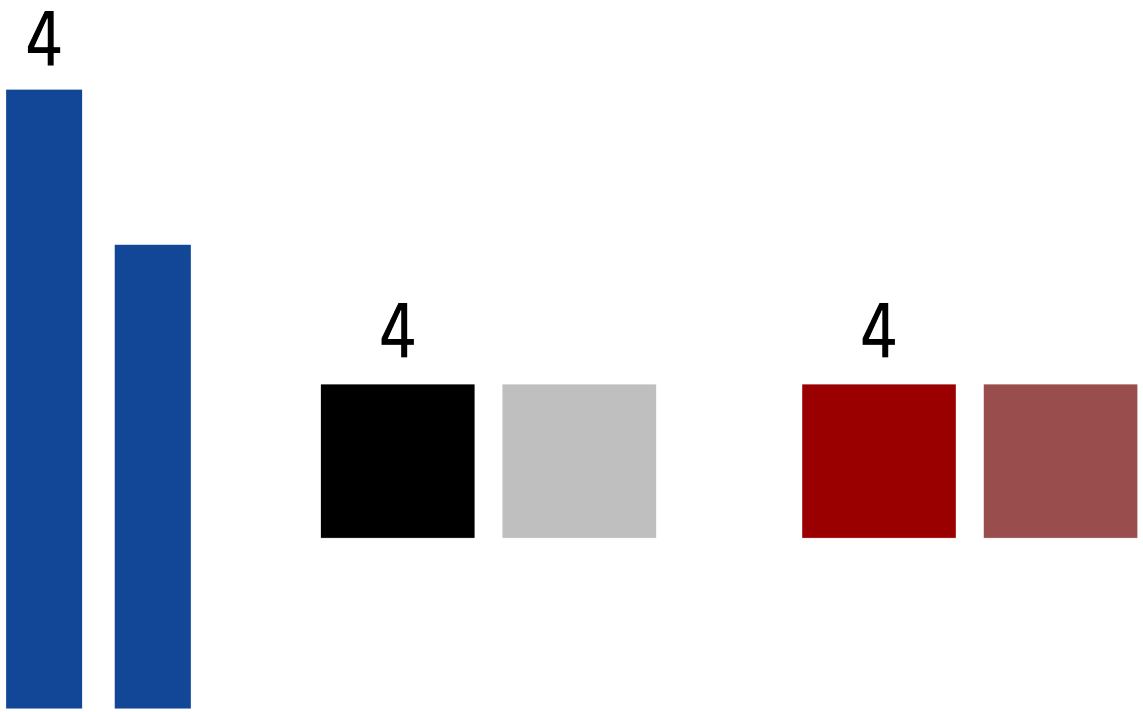
Unequal Encodings

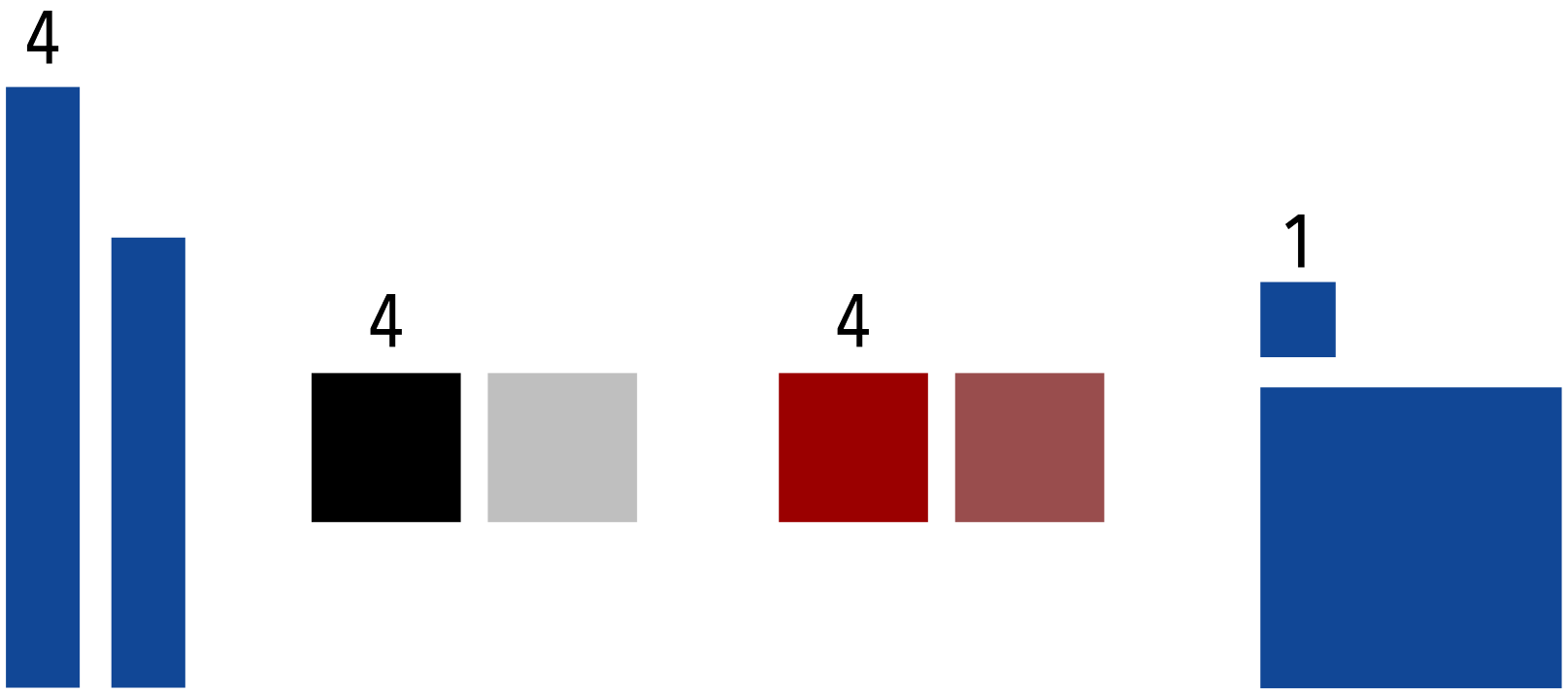
Our visual acuity for distinguishing differences varies across encodings like length, area, brightness, and saturation

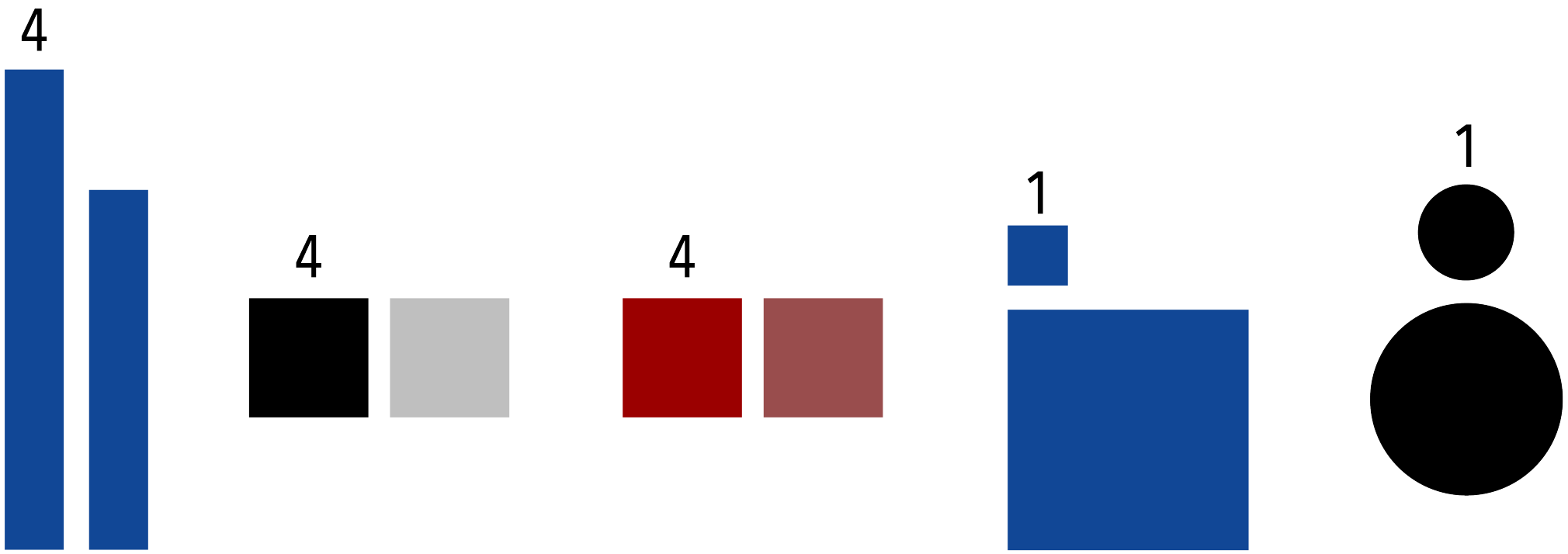

4

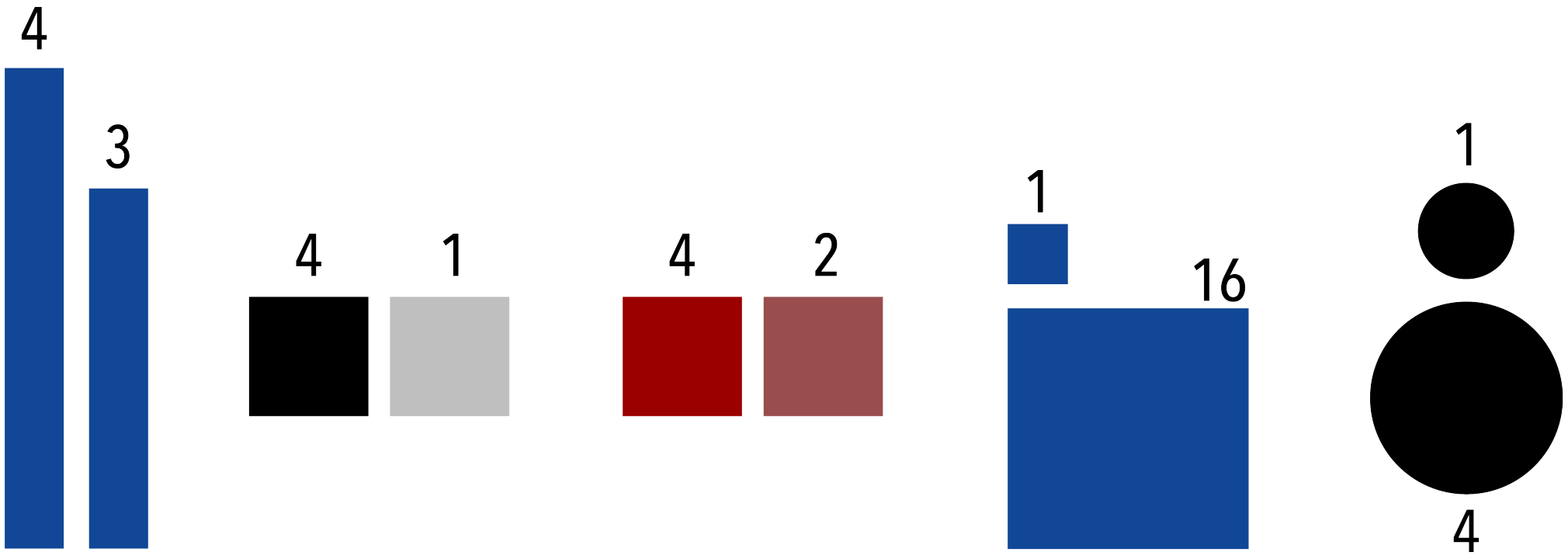

4

4


4
4
4


4
4
4
1


4
4
4
1
1


4
3
4
1
4
2
1
16
1
4

WEBER'S LAW

The perceived change in a stimulus is proportional to the initial stimulus

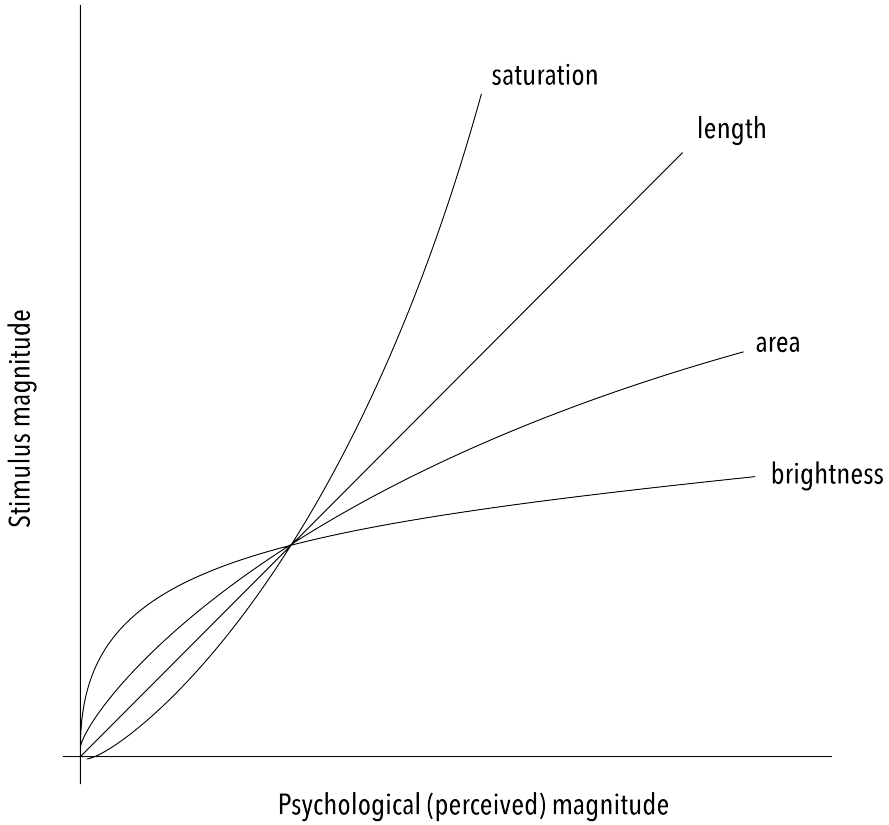
Also known as the Just Noticeable Difference, the amount something must be changed in order for a difference to be detectable

STEVEN'S POWER LAW

The perceived magnitude or intensity of a stimulus

is related to the actual stimulus intensity

by a power law



Improving Figure Clarity

http://www.random.org/integers/

1. In small groups, determine an interpretation for the data set provided
2. Work together to produce a visualization of the data that clearly communicates that interpretation
3. Pass your visualization on to another group
4. In your group, create a new derivative visualization that improves upon the one you have received, considering the concepts discussed during lecture
5. Repeat this process 3 times

Improving Figure Clarity

http://www.random.org/integers/

24	54	71	75	34
9	76	74	34	43
71	17	4	81	60
9	89	67	80	37

1. In small groups, determine an interpretation for the data set provided
2. Work together to produce a visualization of the data that clearly communicates that interpretation
3. Pass your visualization on to another group
4. In your group, create a new derivative visualization that improves upon the one you have received, considering the concepts discussed during lecture
5. Repeat this process 3 times

Discussion

1. In what ways was your original visualization improved by other groups?

2. Did you find improvements on your visualization that were surprising?

3. Are there specific ways in which you could have communicated your data more effectively from the beginning?