Bringing Data to Life Through Visualisation

Dr Sarah Marshall sarah.marshall@aut.ac.nz Dept Math. Sciences

Today

Why Visualise Data?

Elements of Data Visualisation

Encouraging Curiosity through Data Visualisation

Why Visualise Data?

Exploratory

- Familiarisation with data
- Discover hidden patterns
- Explore large datasets

Explanatory

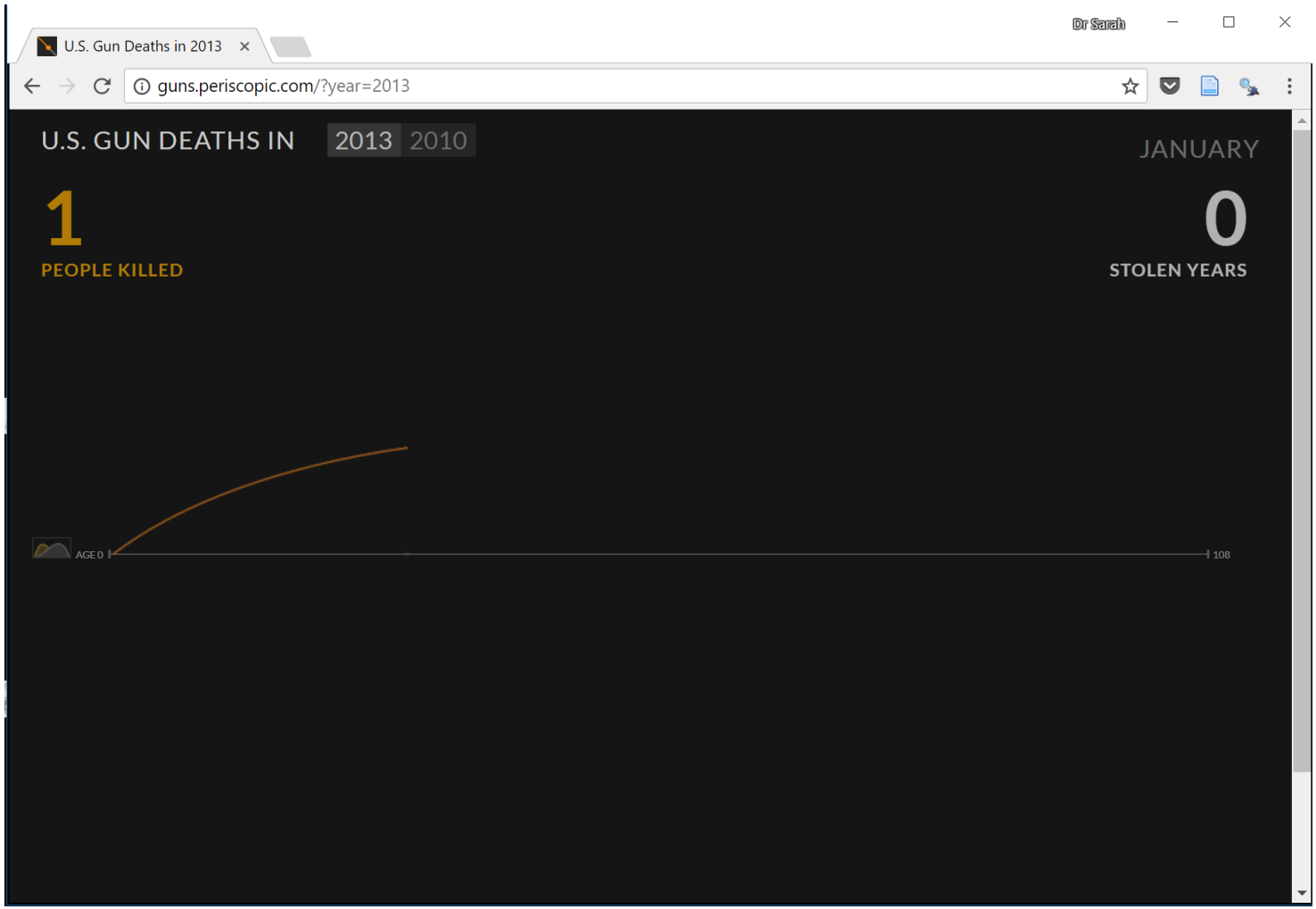
- Communicate insight
- Highlight presence of phenomenon
- Convey meaning


About 440,000 Americans die each year from diseases related to smoking.
90% of them started as teen smokers.
Smoking kills.
truth
THETRUTH.COM

Picture Superiority Effect

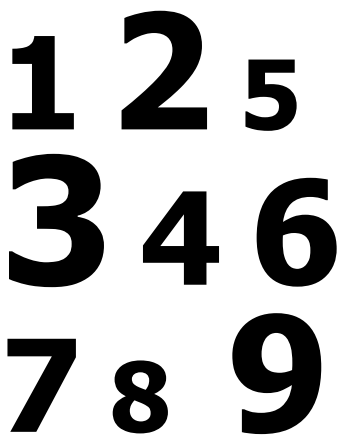
Picture Superiority Effect

Gun Deaths in the USA

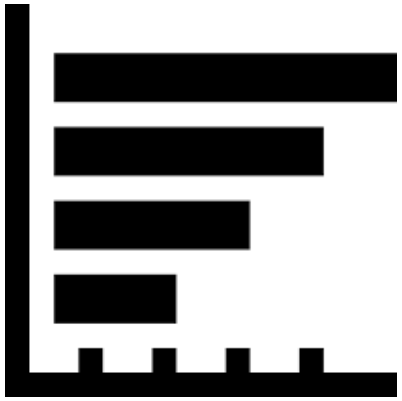


http://guns.periscopic.com/

Elements of Data Visualisation



Data



Visualisation Tools

Design Principles

DATA

1 2 5
3 4 6
7 8 9

Data Types

Numeric

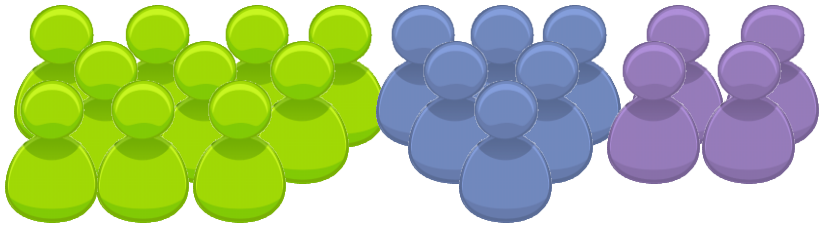
- Values are numbers
- Continuous or discrete

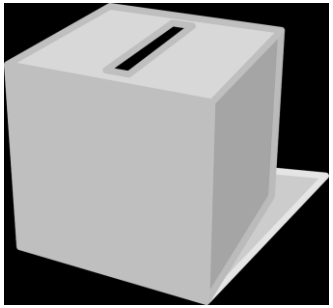
Categorical

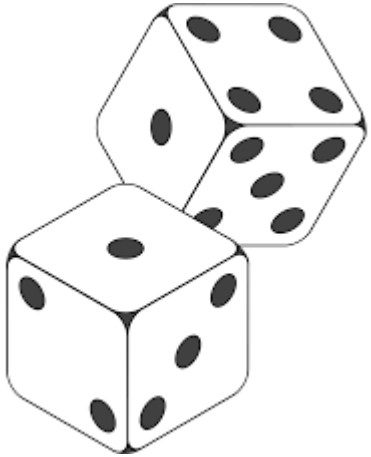
- Values are selected from a number of categories.
- Nominal or ordinal

Numerical Data

Discrete







Continuous

Activity: Type of Data

Discrete or Continuous?

- Number of students in this class
- Result from rolling a dice
- Height of All Blacks
- Time taken to get to AUT this morning
- Number of cats at home
- Age of America's Cup sailors

Categorical Data

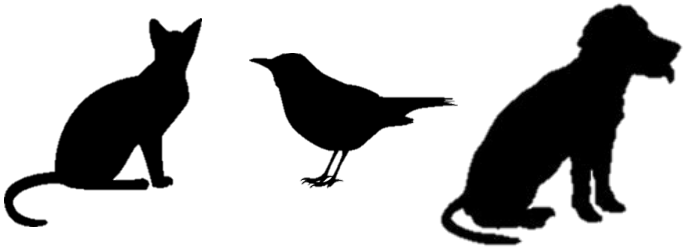
Nominal

No order

Hair Colour?

Gender?

Favourite?



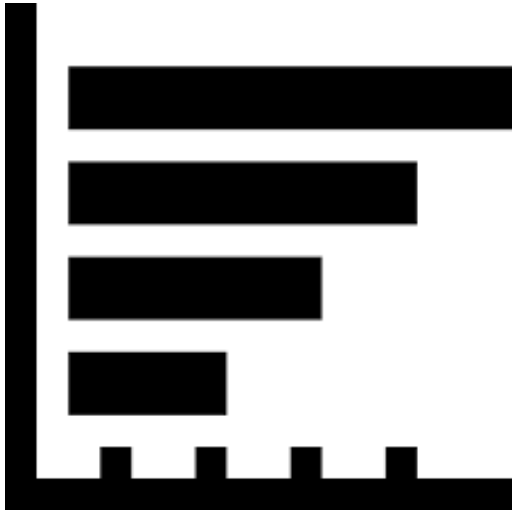
Ordinal

Meaningful order

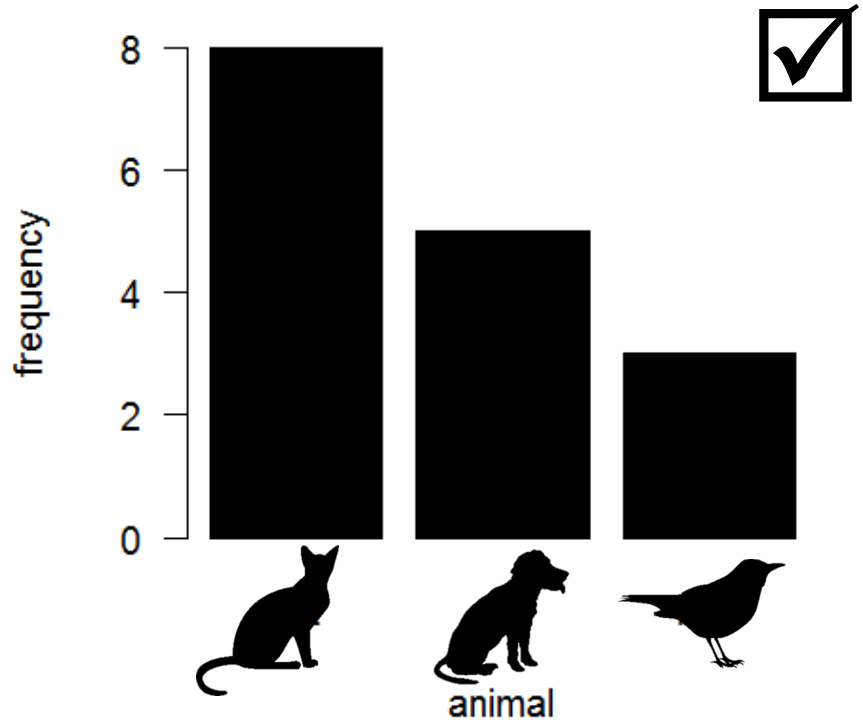
A+ A- A
B+ B B- …

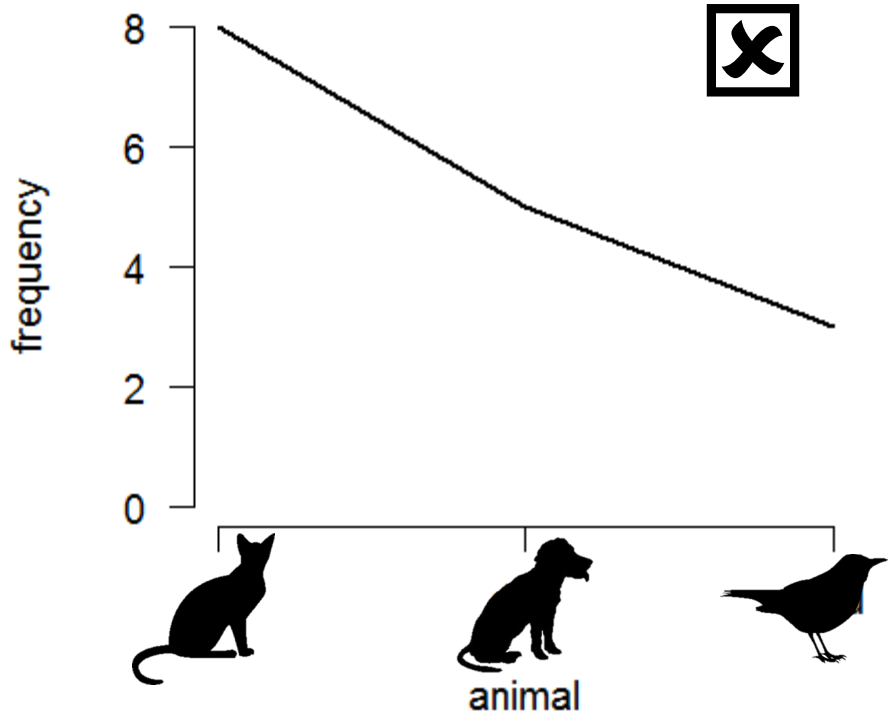
Strongly disagree
Disagree
Neither agree nor disagree
Agree
Strongly agree

Visualisation Tools

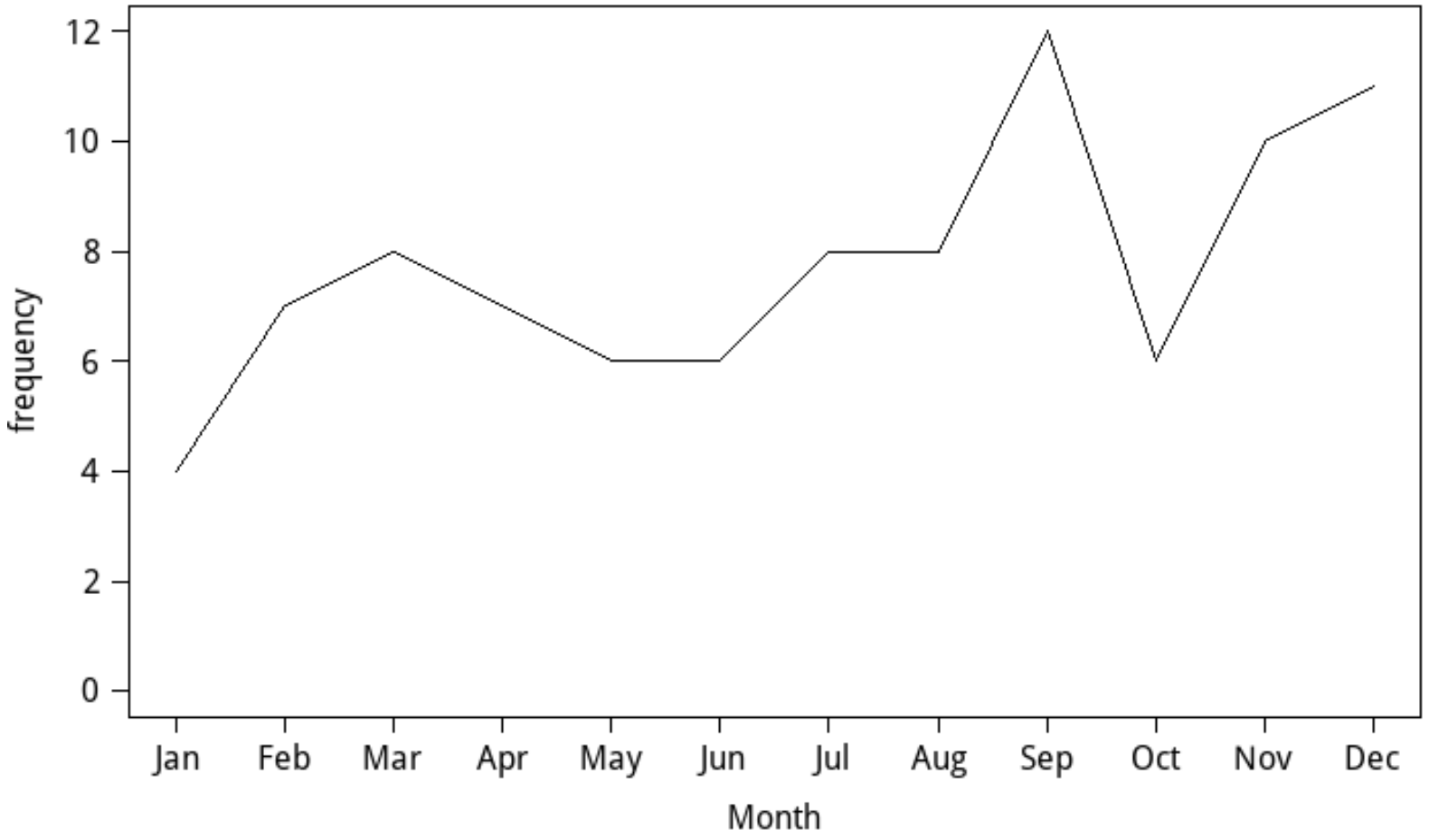


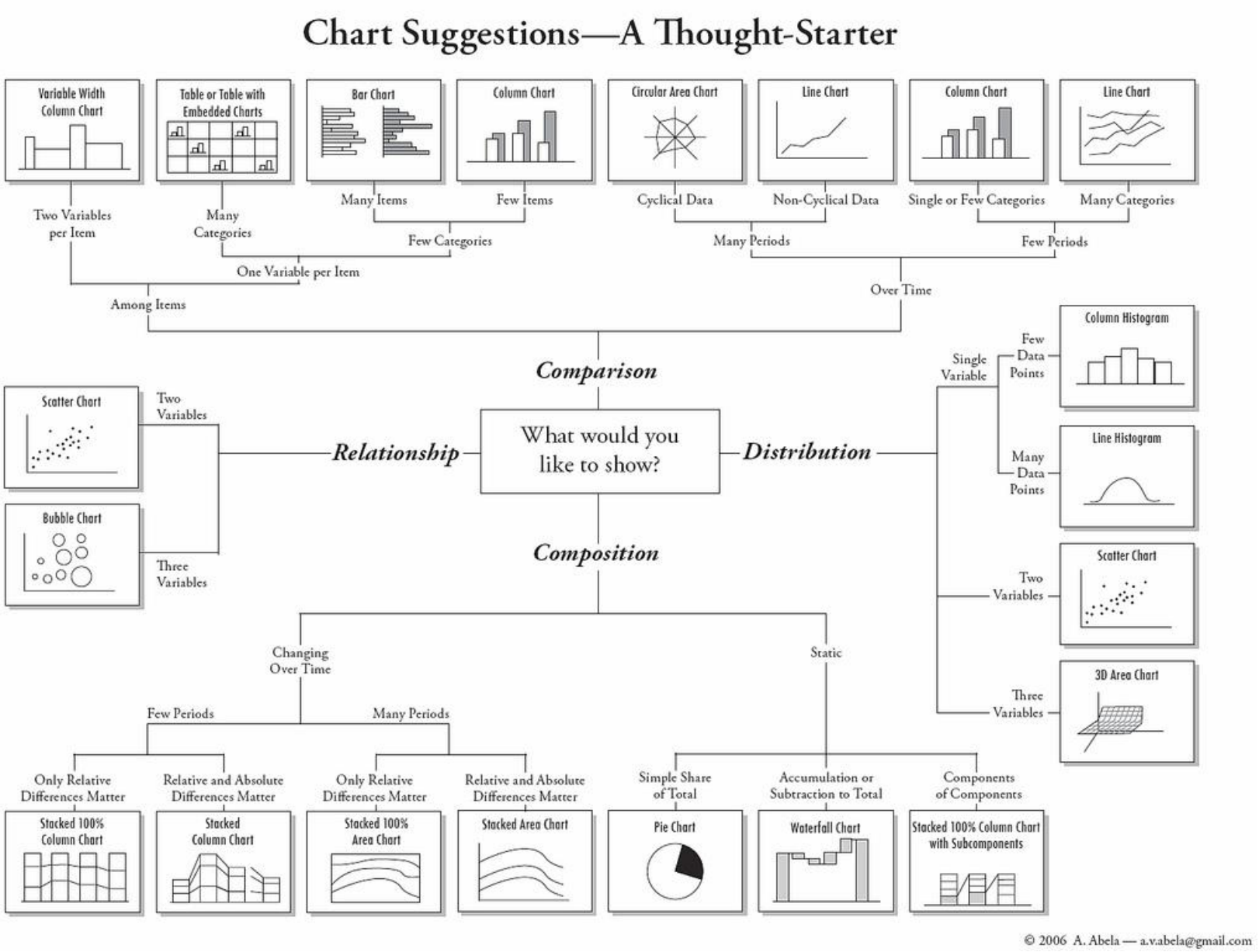
Discrete





Continuous (Trends Over Time)





http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html

Visualisation Software

Excel
Google Sheets
R (including Shiny)
Power BI
Tableau
Google Trends

Caution: default options are not necessarily the best options

Design Principles

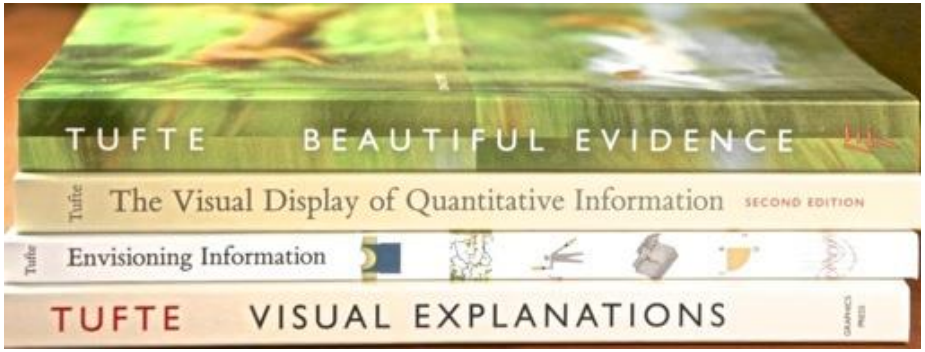
Design Principles

Graphical Integrity

Display data accurately

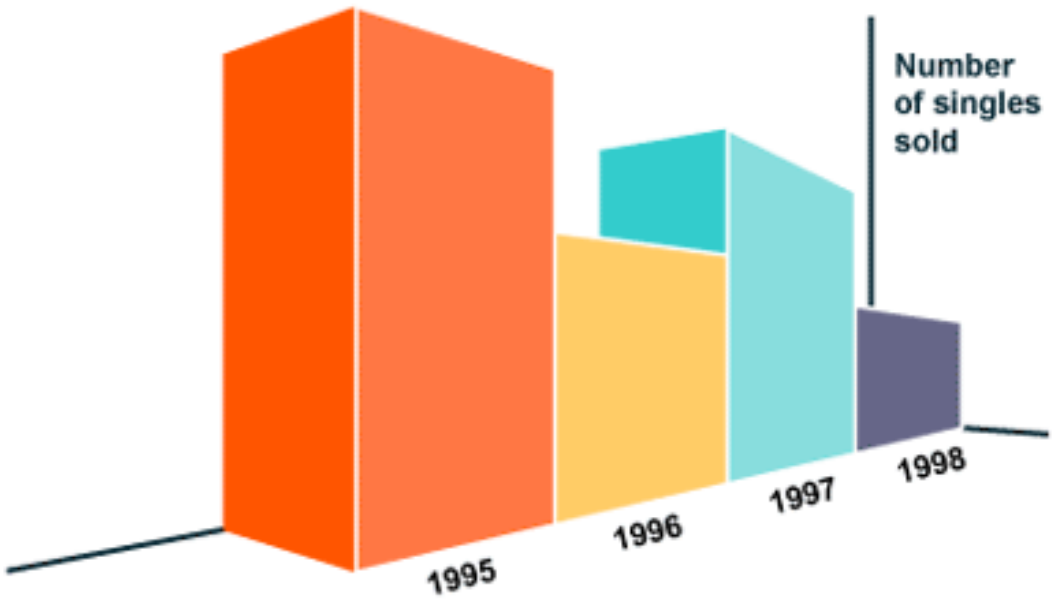
Graphical Excellence

Display data effectively and clearly



Edward Tufte

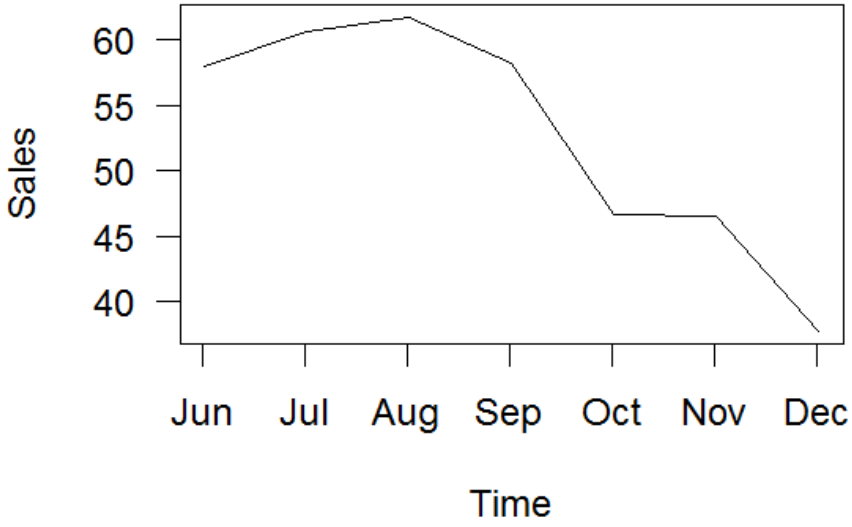
Proportion

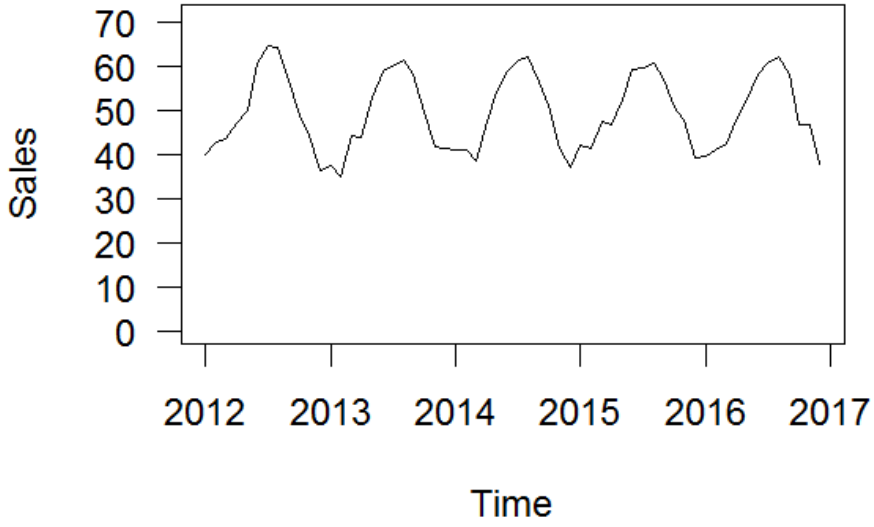


Sales 1995 > 1997?

http://www.bbc.co.uk/schools/gcsebitesize/maths/statistics/representingdata2rev5.shtml

Context & Numerical Scale





Woollen Blanket Sales

Everything should be made as

SIMPLE

as possible, but not

SIMPLER.

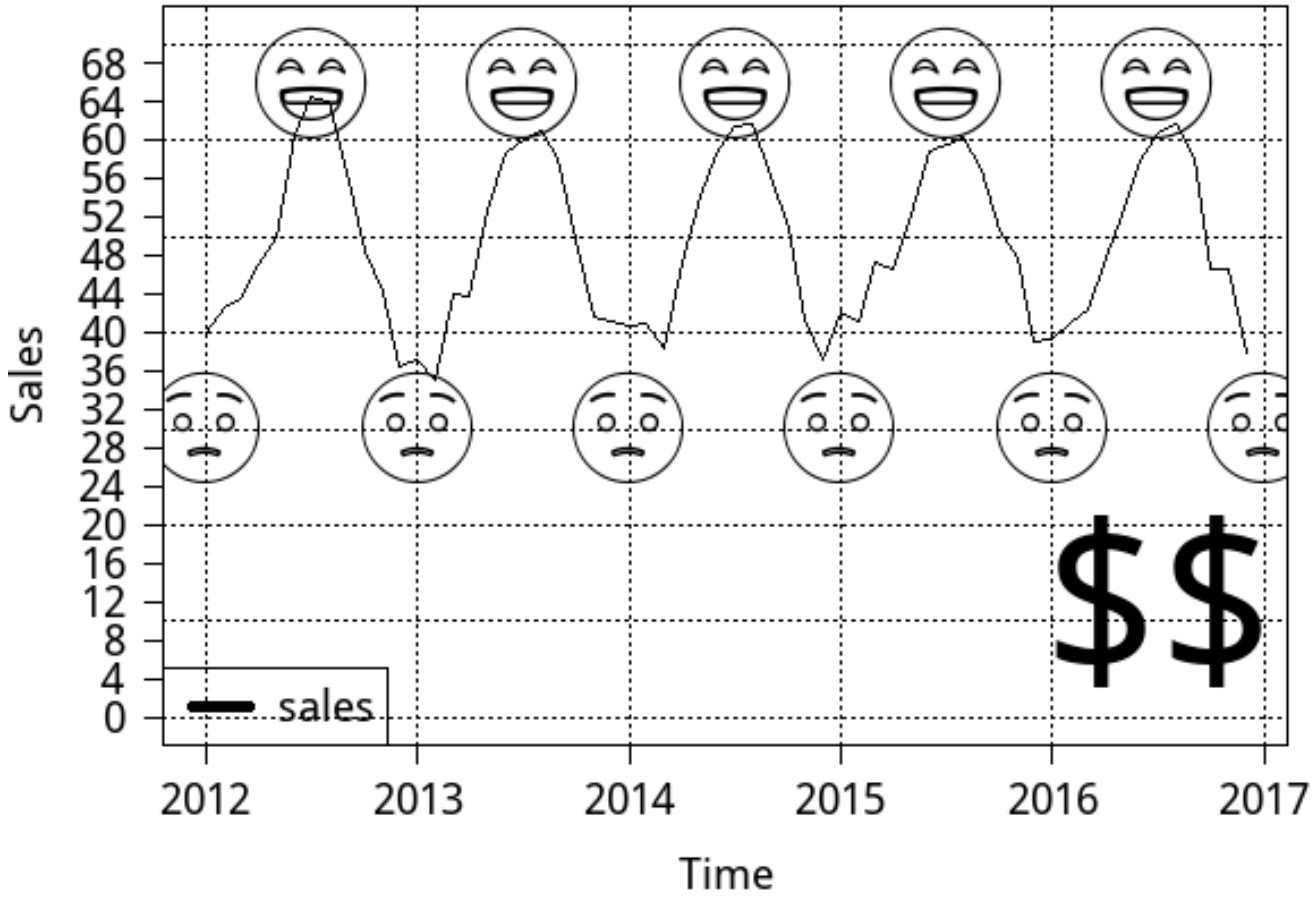
Albert Einstein

Maximise Data Ink Ratio

$$\text{Data Ink Ratio} = \frac{\text{Data Ink}}{\text{Total Ink Used in Graphic}}$$

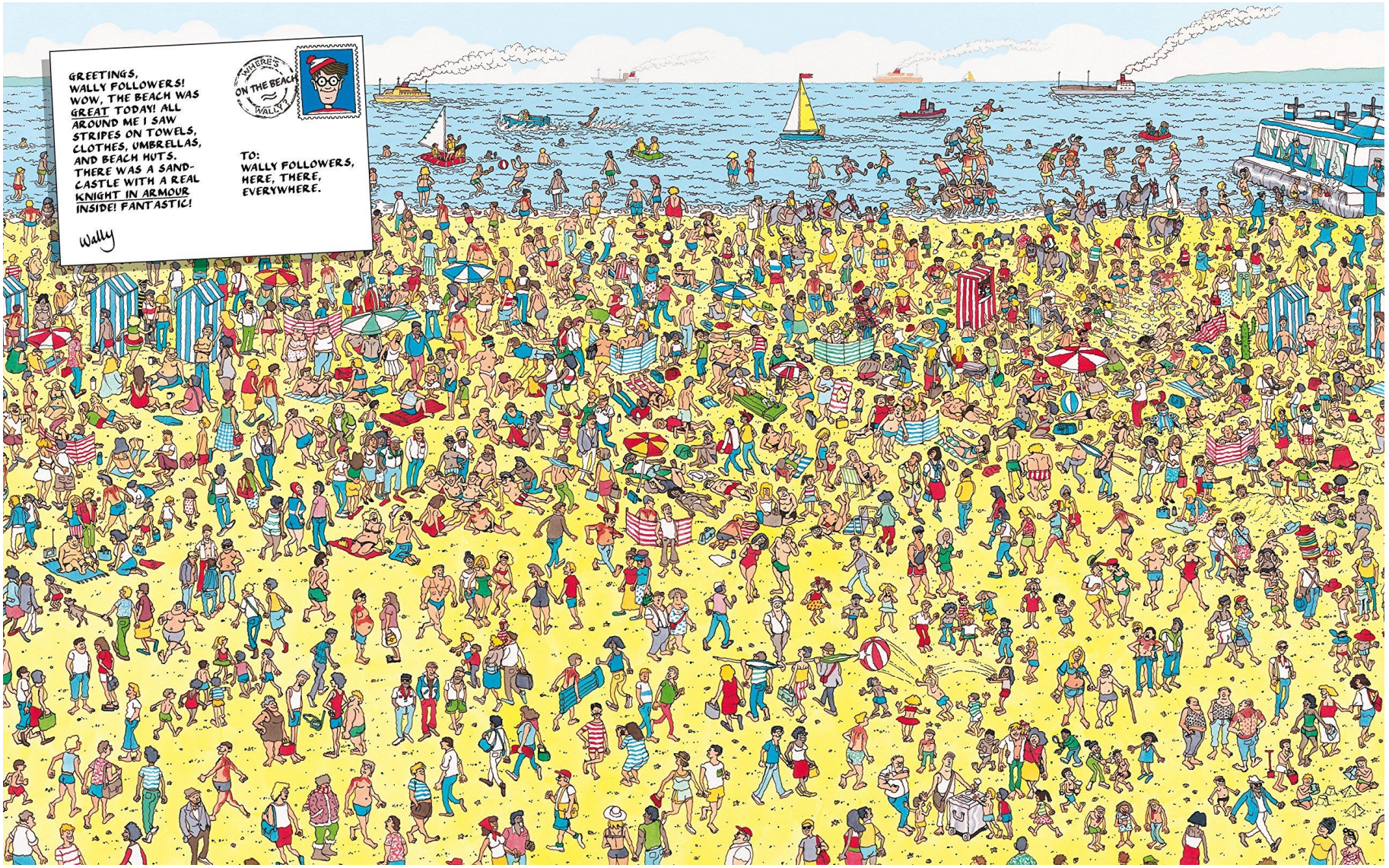
Minimise “Chart Junk”

Chart Junk

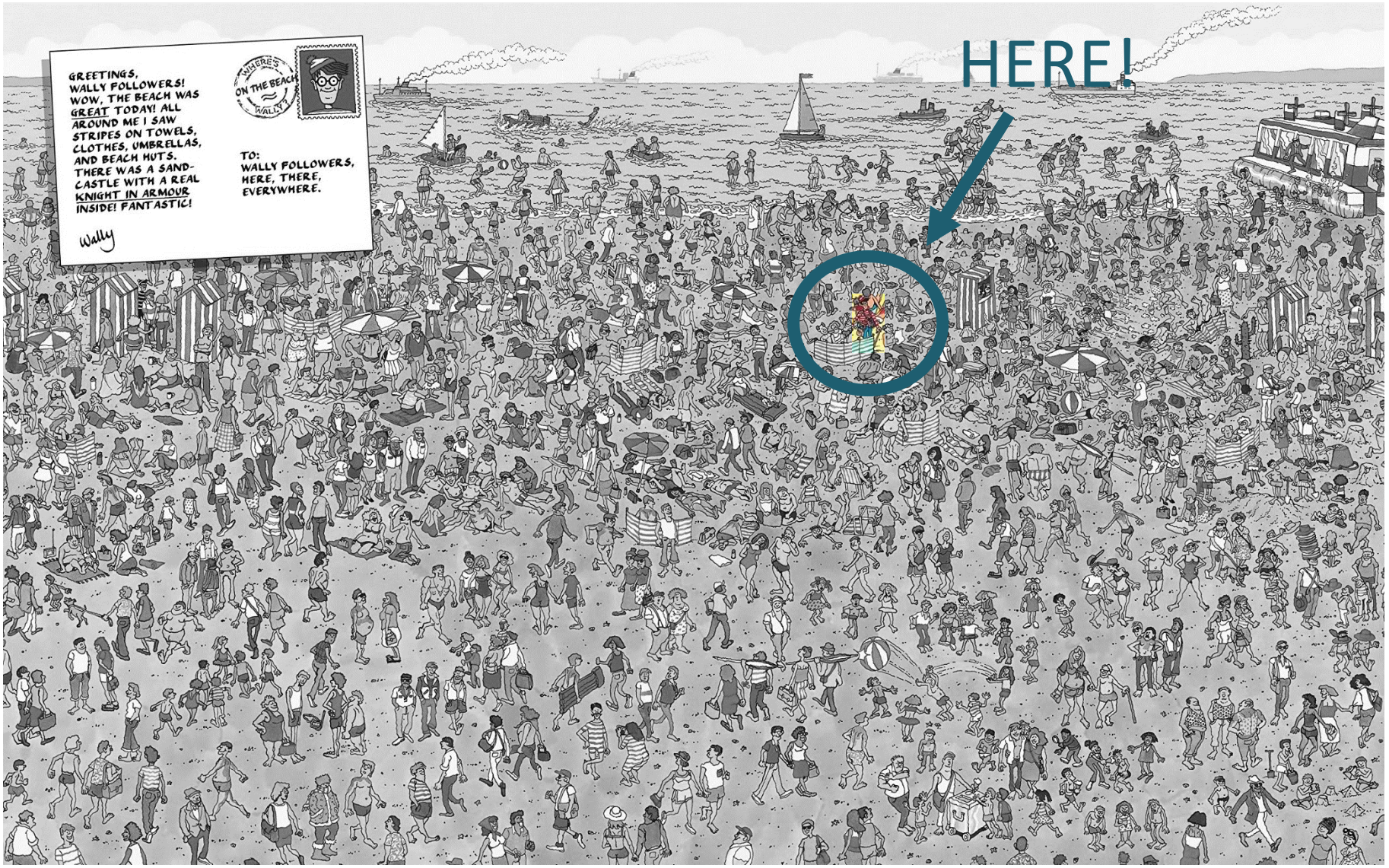

Sales
68
64
60
56
52
48
44
40
36
32
28
24
20
16
12
8
4
0
sales
$$
2012
2013
2014
2015
2016
2017
Time

Where's Wally?

Where's Wally?



WHERE'S WALLY?



How Many Threes?

How Many Threes?

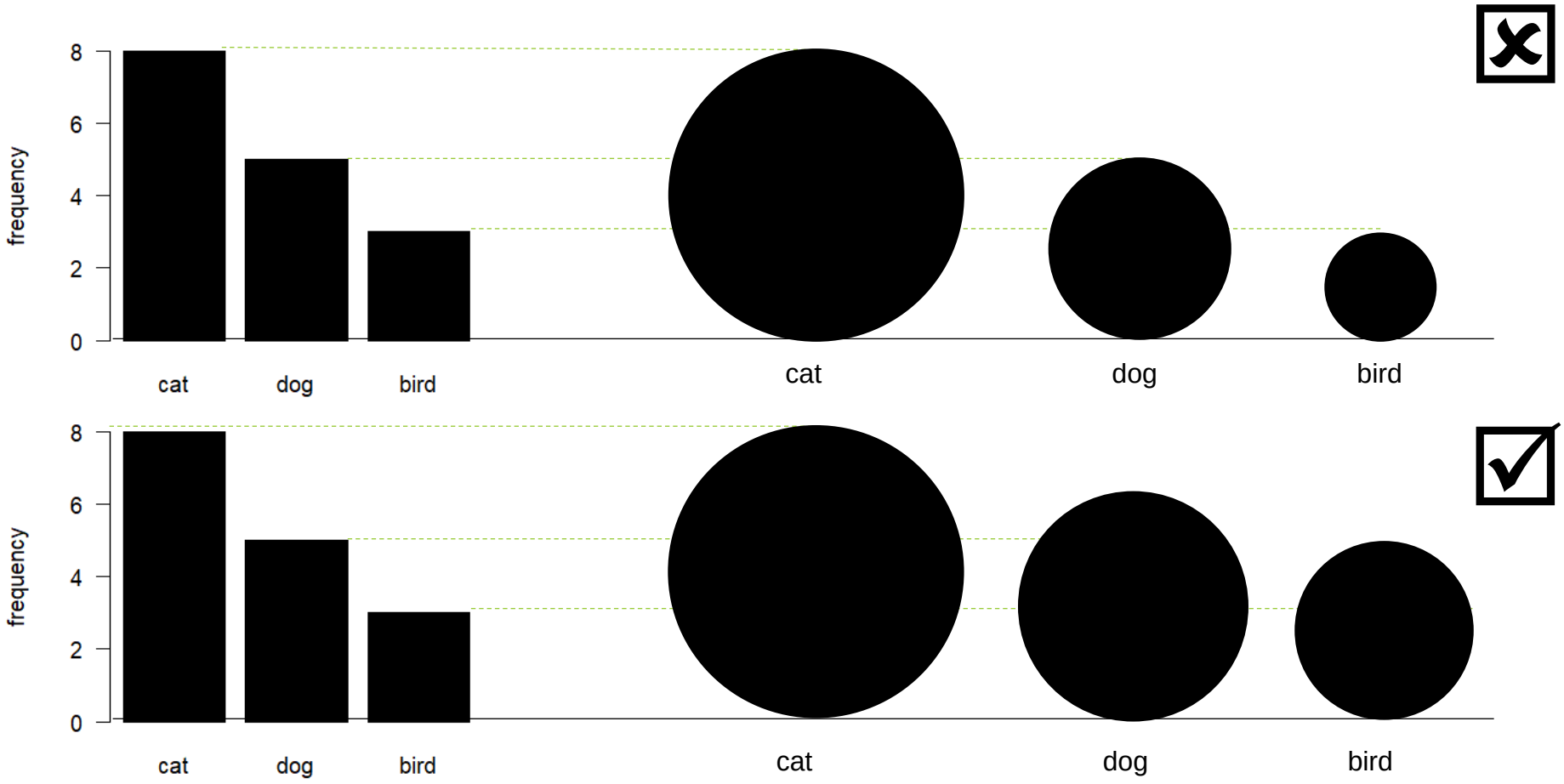
1913186243924148522327154403504531688853193131975961424215384486

How Many Threes?

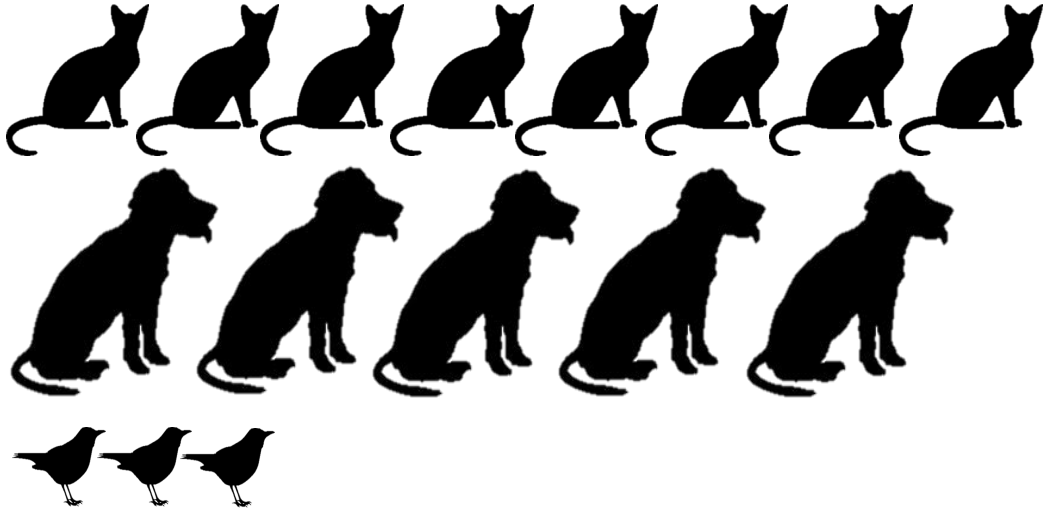
1913186243924148522327154403504531
6888531931319759614242153844486

Scale

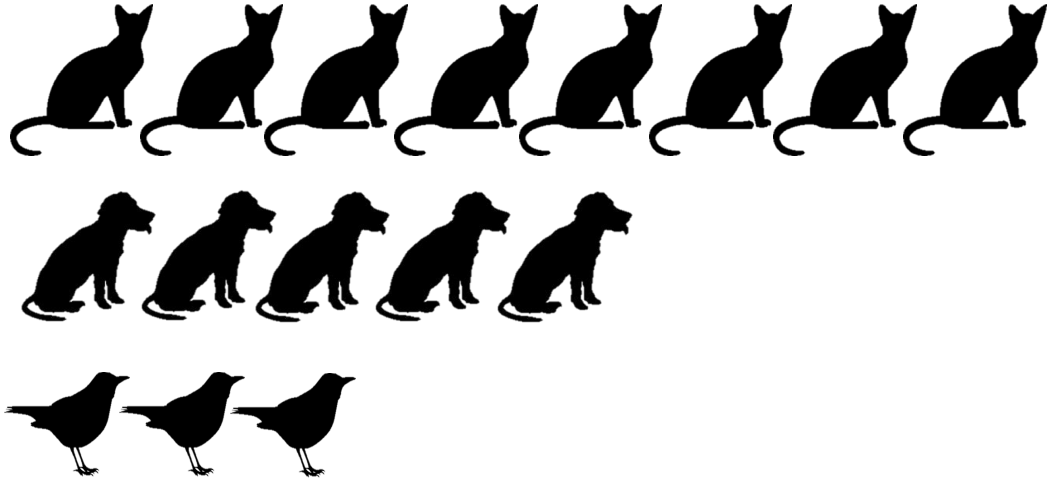
Circle area proportional to frequency not bar height



Scale



☒



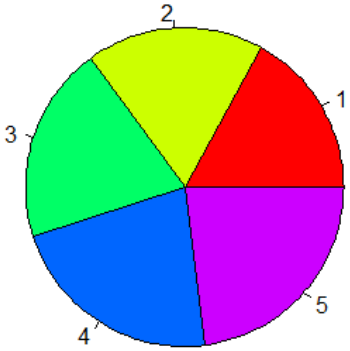
☑

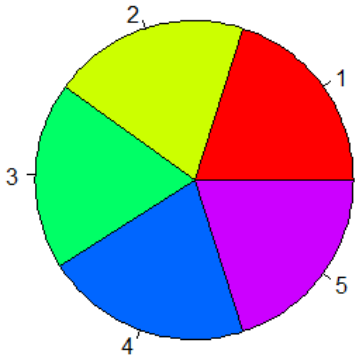
Elements of Design

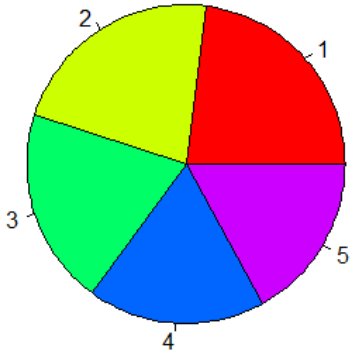
Colour
Contrast
Scale

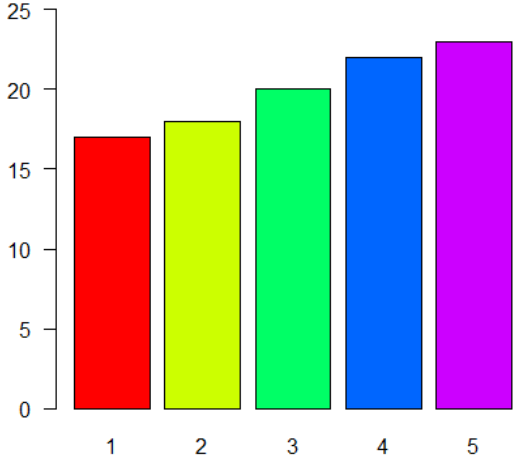
- Impact effectiveness of visualisations
- Use thoughtfully to convey meaning and highlight key points

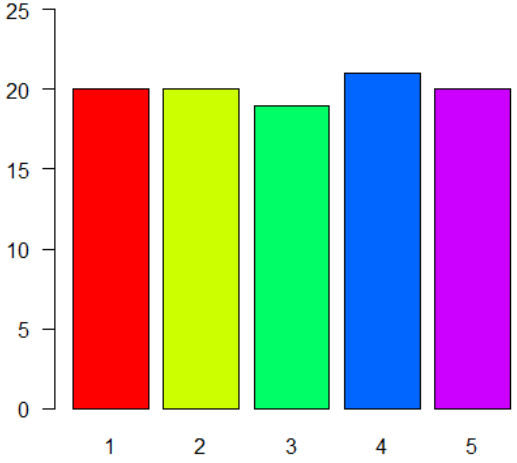
Bars > Pies

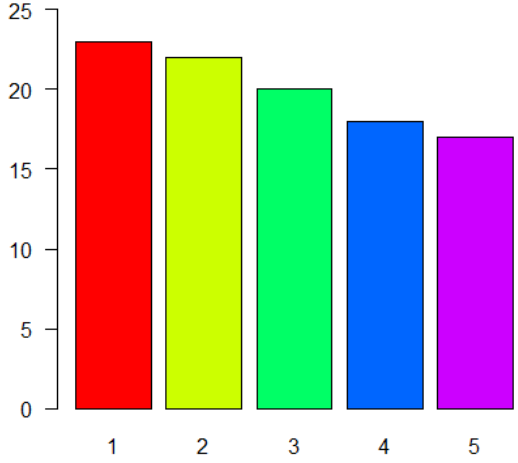

1
2
3
4
5


1
2
3
4
5

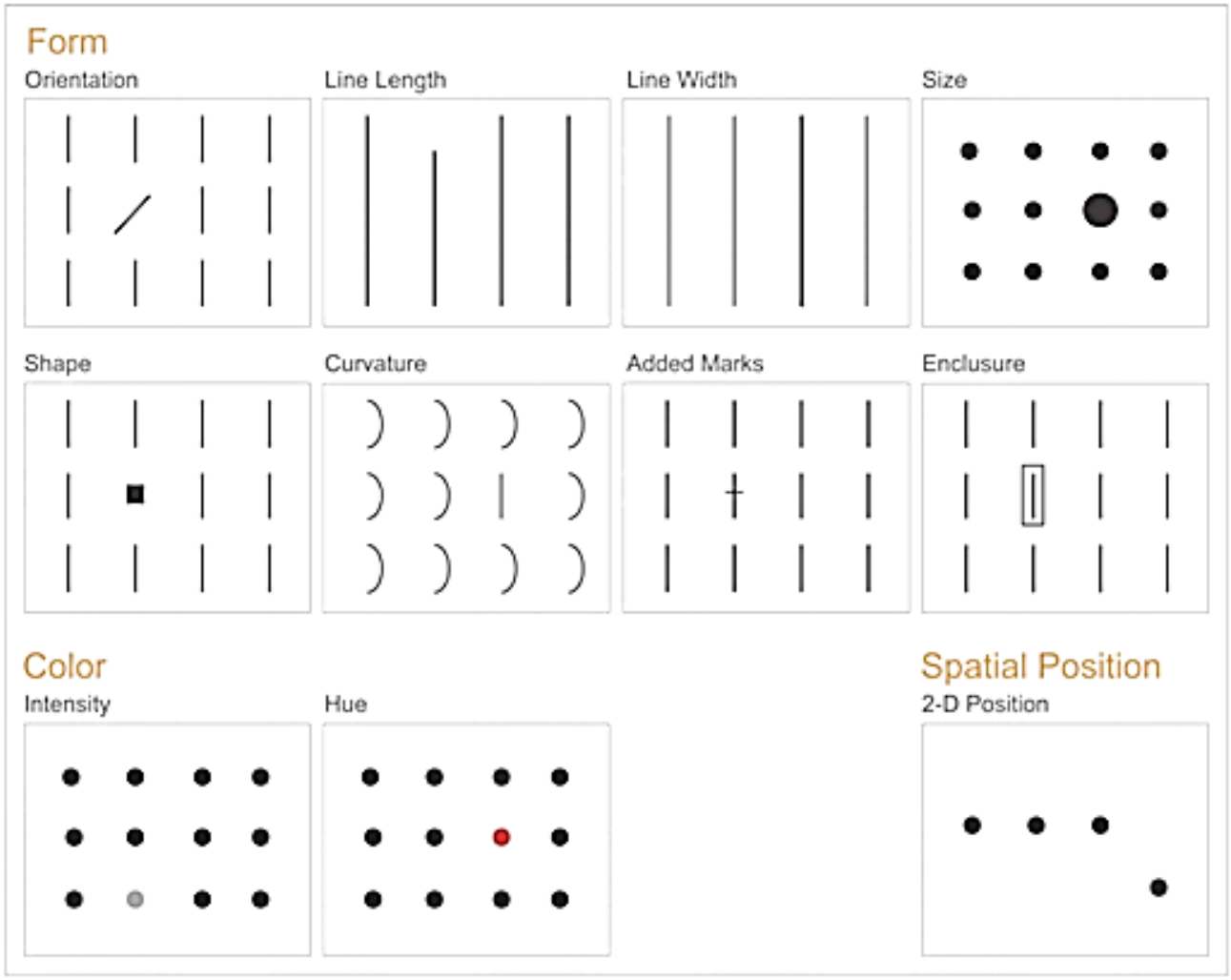

1
2
3
4
5


0
5
10
15
20
25
1
2
3
4
5


0
5
10
15
20
25
1
2
3
4
5


0
5
10
15
20
25
1
2
3
4
5

Pre-attentive Attributes



https://www.perceptualedge.com/articles/ie/visual_perception.pdf

Stephen Few