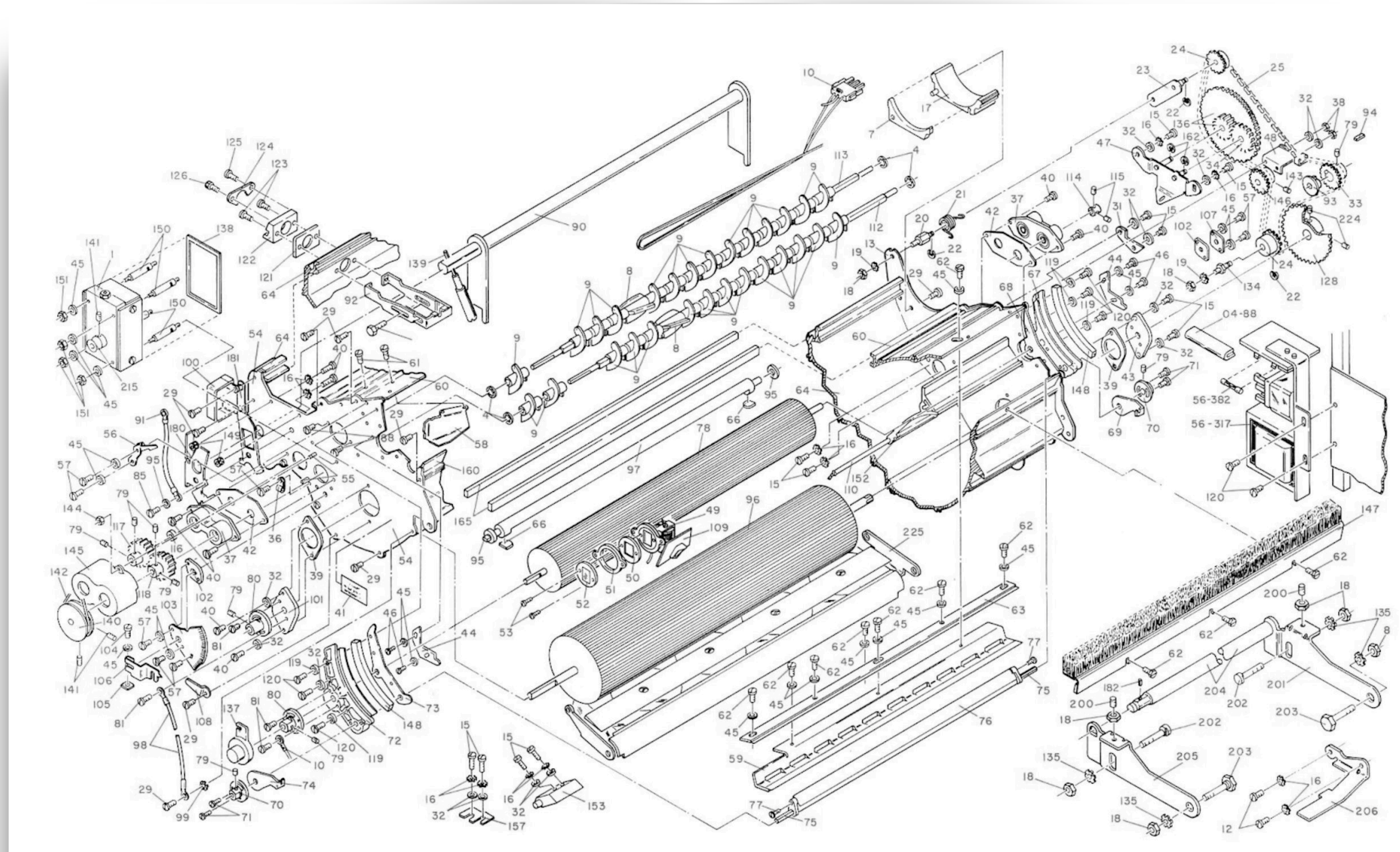
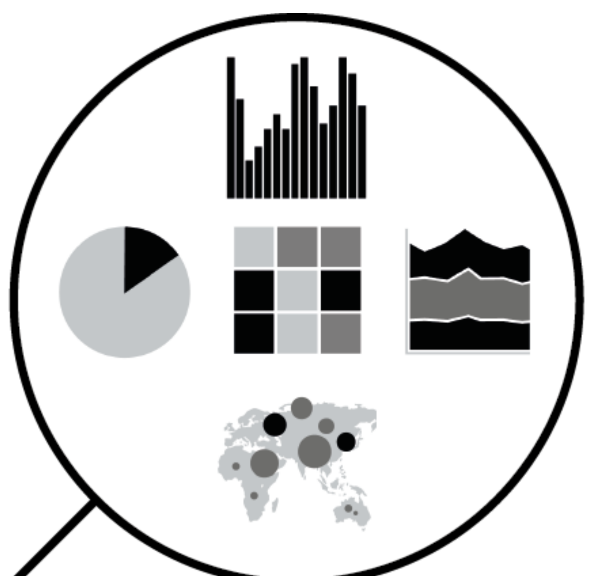
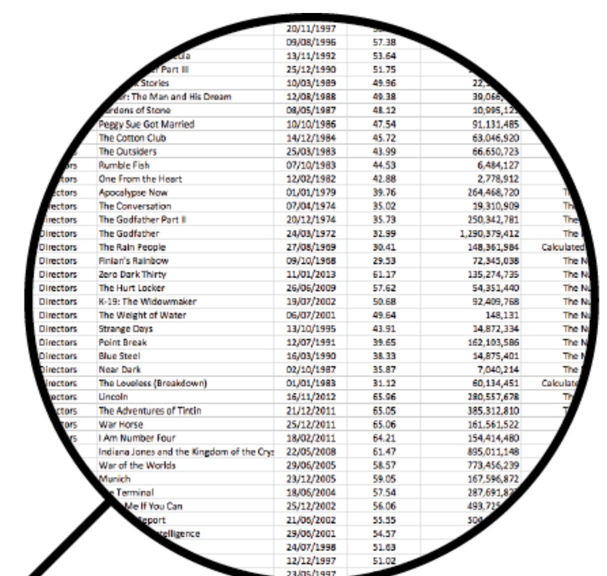


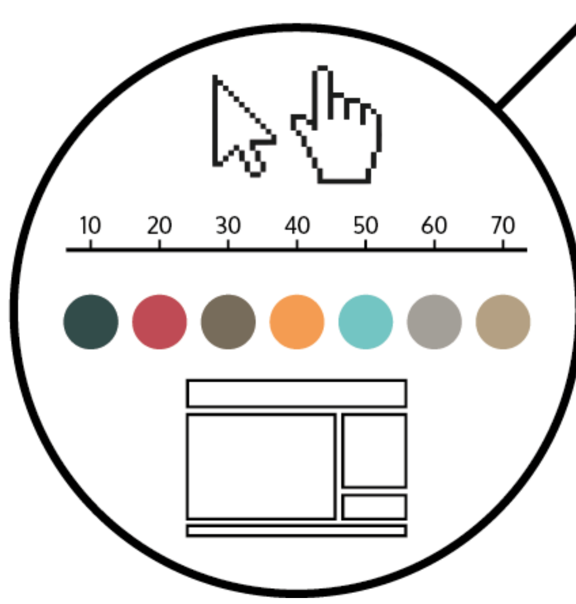
foundations of data visualisation

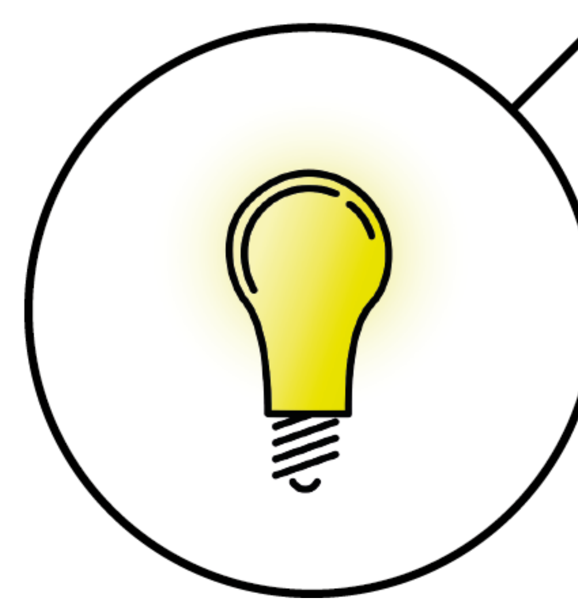
visualisation





The representation and presentation of data to facilitate understanding





Kirk, 2016, Data Visualisation

ingredients

data is the key actor in the dataviz process

tabular form is what we consider the raw form of data

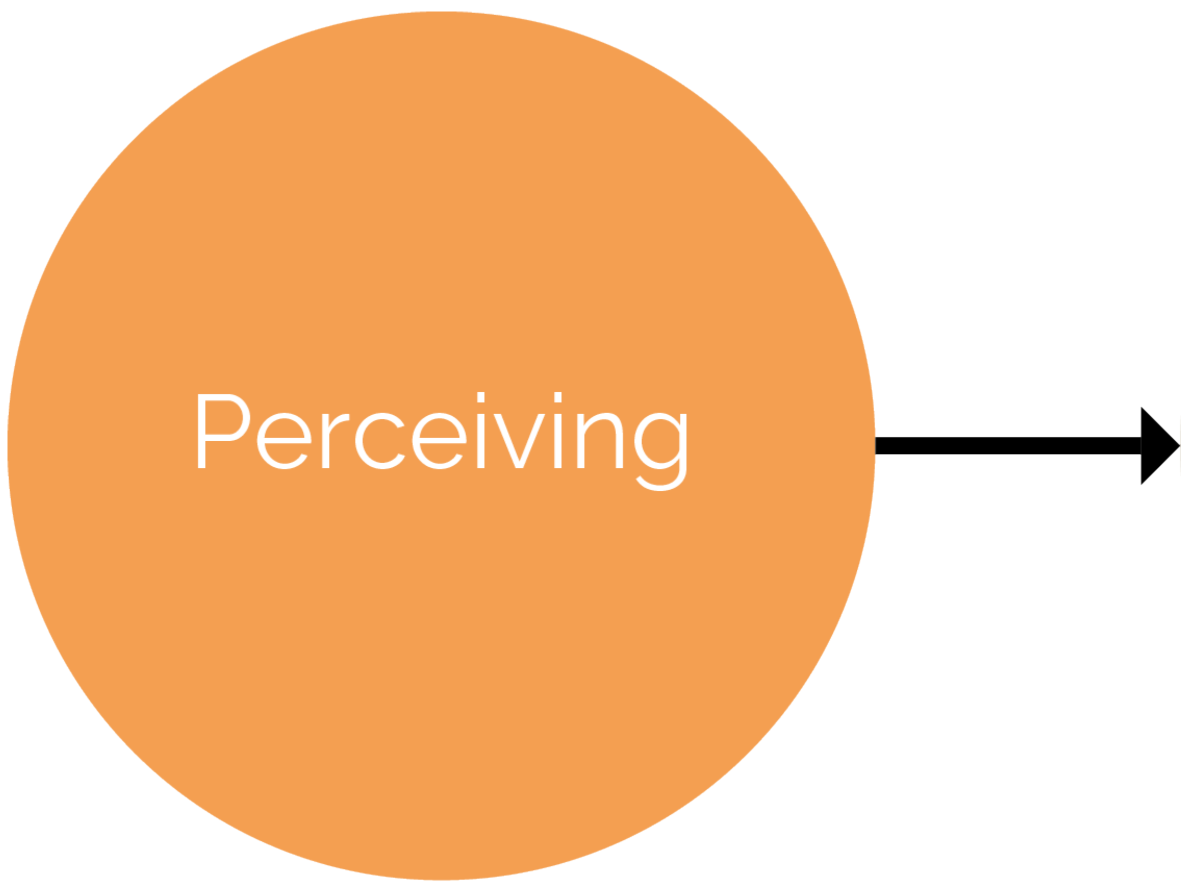
raw data cannot be compared and contrasted effectively

data representation is the visualization of data through the combination of marks and attributes

data presentation concerns all other visible design decision beyond the representation: interactivity, annotations, colours

facilitating **understanding** should be the goal

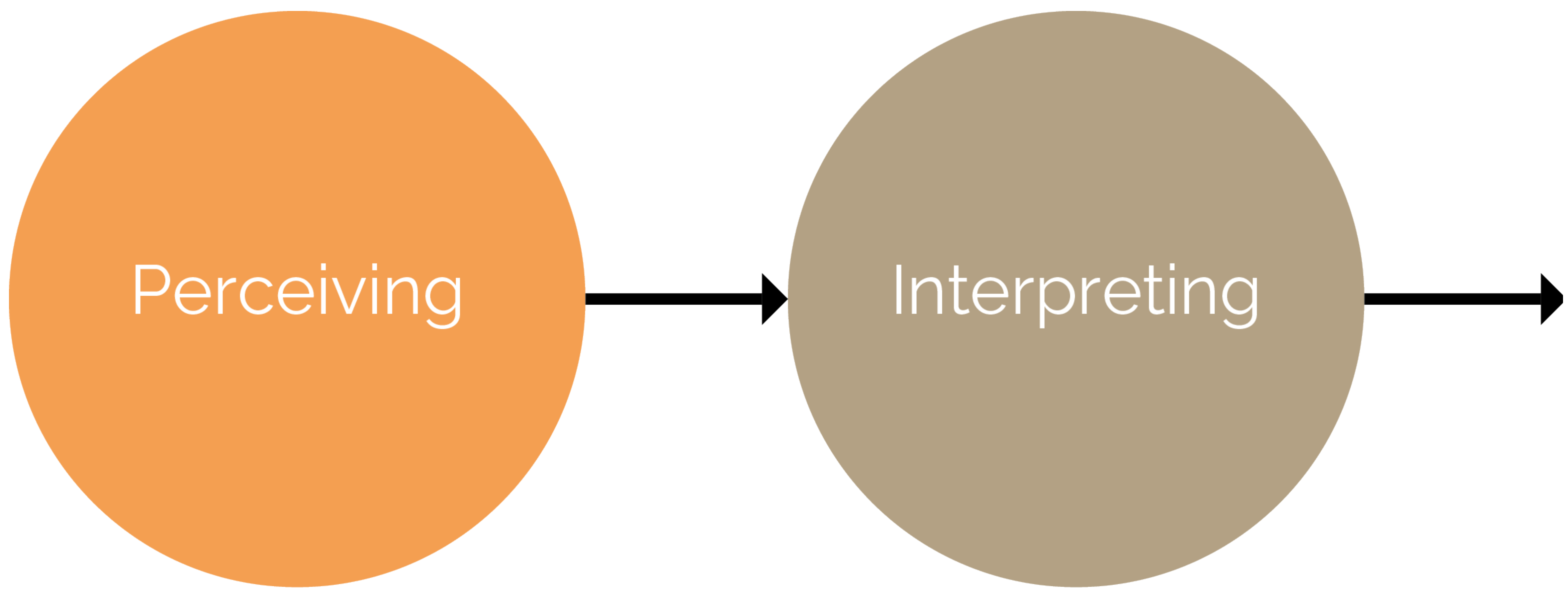
understanding



What does it show?

Where is big, medium, small?
How do things compare?
What relationships exist?

understanding



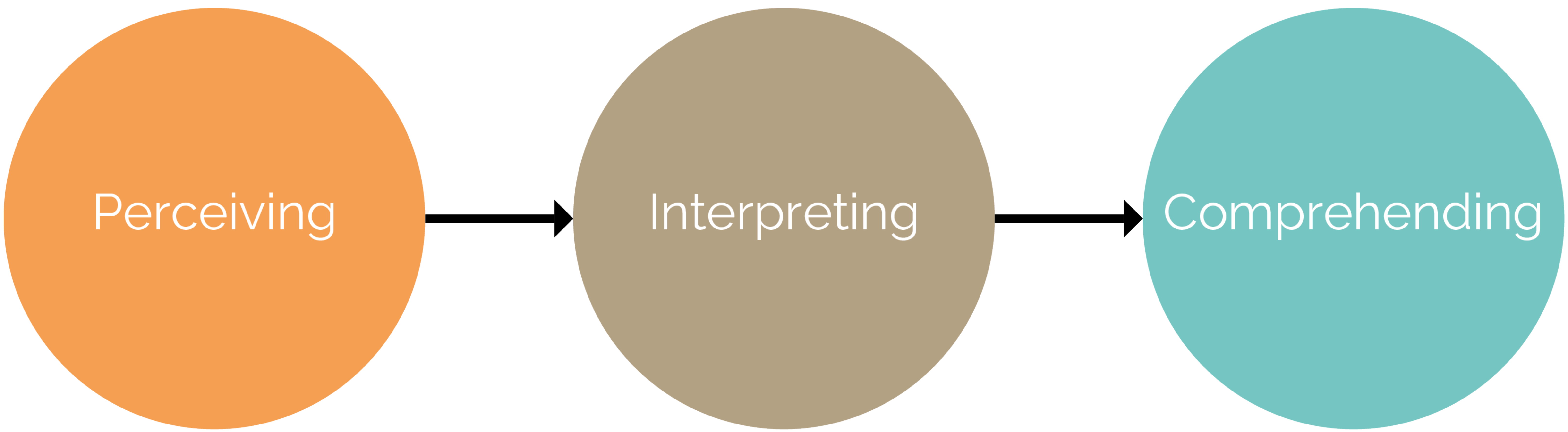
What does it **show**?

Where is big, medium, small?
How do things compare?
What relationships exist?

What does it **mean**?

What is good and bad?
Is it meaningful or insignificant?
Unusual or expected?

understanding



What does it **show**?
Where is big, medium, small?
How do things compare?
What relationships exist?

What does it **mean**?
What is good and bad?
Is it meaningful or insignificant?
Unusual or expected?

What does it **mean to me**?
What are the main messages?
What have I learnt?
Any actions to take?

perceiving

Perceiving

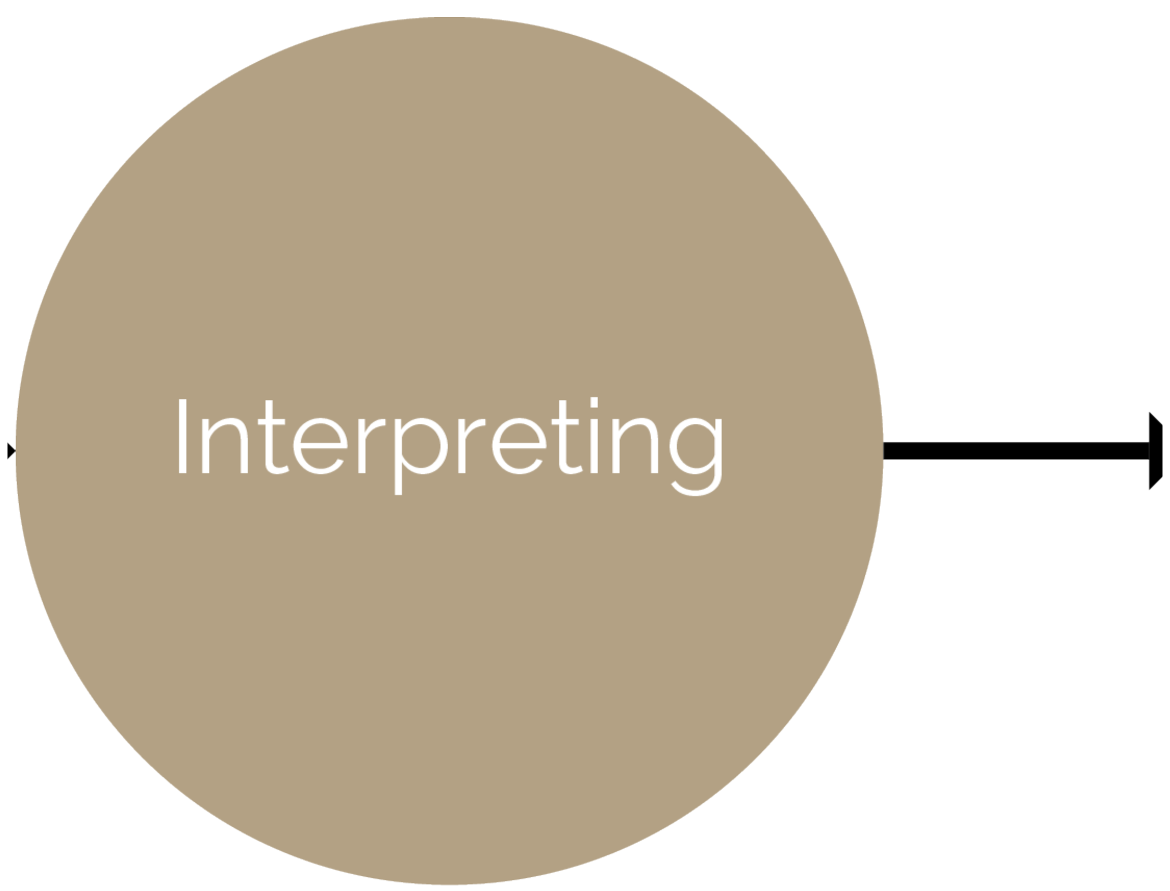
What does it **show**?

Where is big, medium, small?
How do things compare?
What relationships exist?

- where are the largest/middle-sized/smallest values?
- what proportion of the total does that value hold?
- how do these values compare in ranking terms?
- to which other values does this have a connected relationship?

the art of simply being able to read a chart
efficiently decode the representation of the data (shapes/sizes/colors)

interpreting

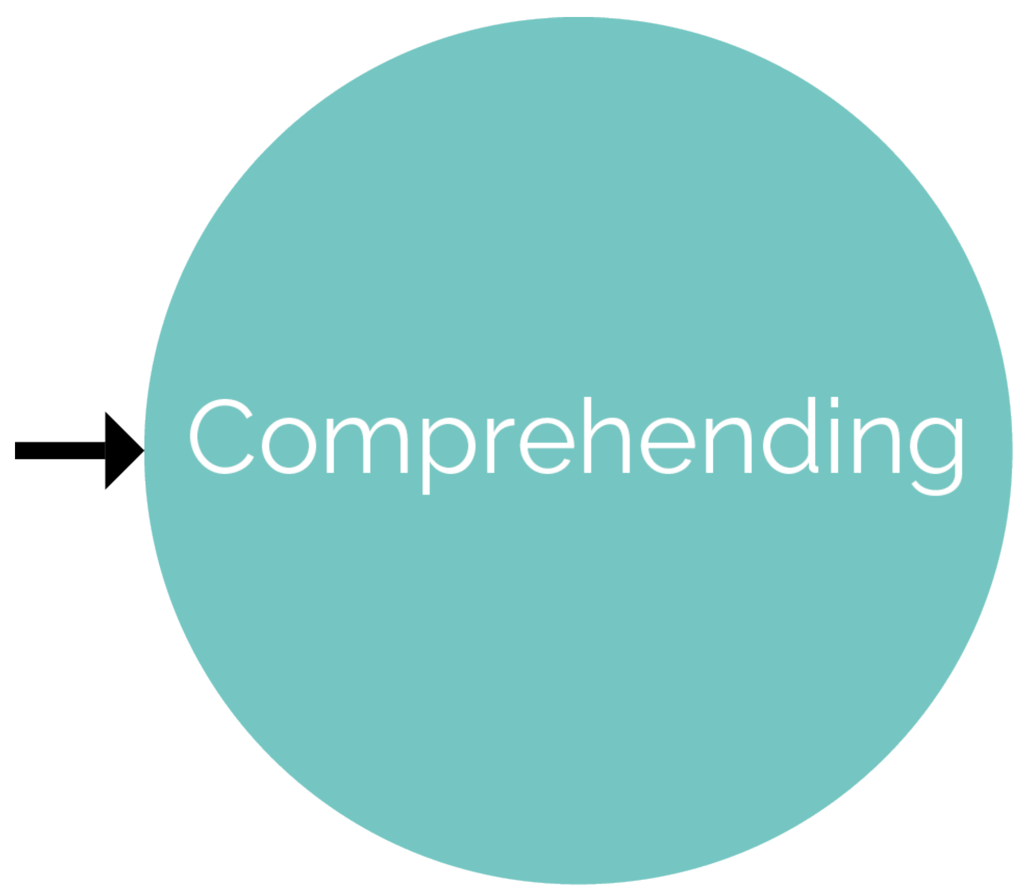


What does it **mean**?
What is good and bad?
Is it meaningful or insignificant?
Unusual or expected?

- is it good to be big or it is better to be small?
- what does it mean to go up or to go down?
- is that relationship meaningful or insignificant?
- is the decline/increase of that category especially surprising?

the art of converting the perception into meaning
use pre-existing knowledge to frame the implication of the viz

comprehending

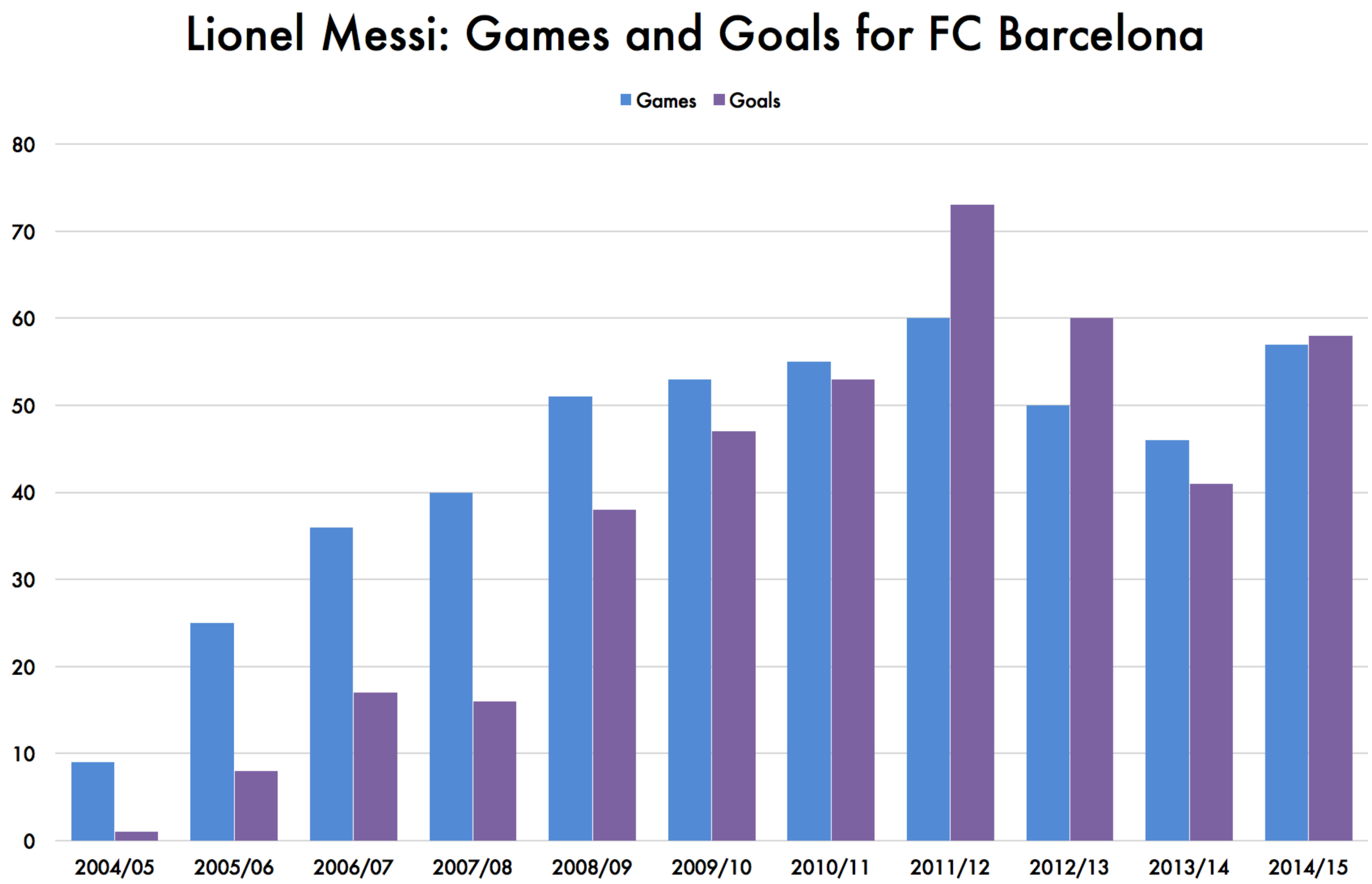


What does it **mean to me**?

What are the main messages?
What have I learnt?
Any actions to take?

- why is this relevant? to whom?
- has it confirmed what I suspected or enlightened me with new knowledge?
- has this impacted me emotionally or left me indifferent?
- does this new understanding force me to take action on the subject?

the act of reasoning about the consequences of interpretation
what is the novelty carried by this dataviz about the subject?

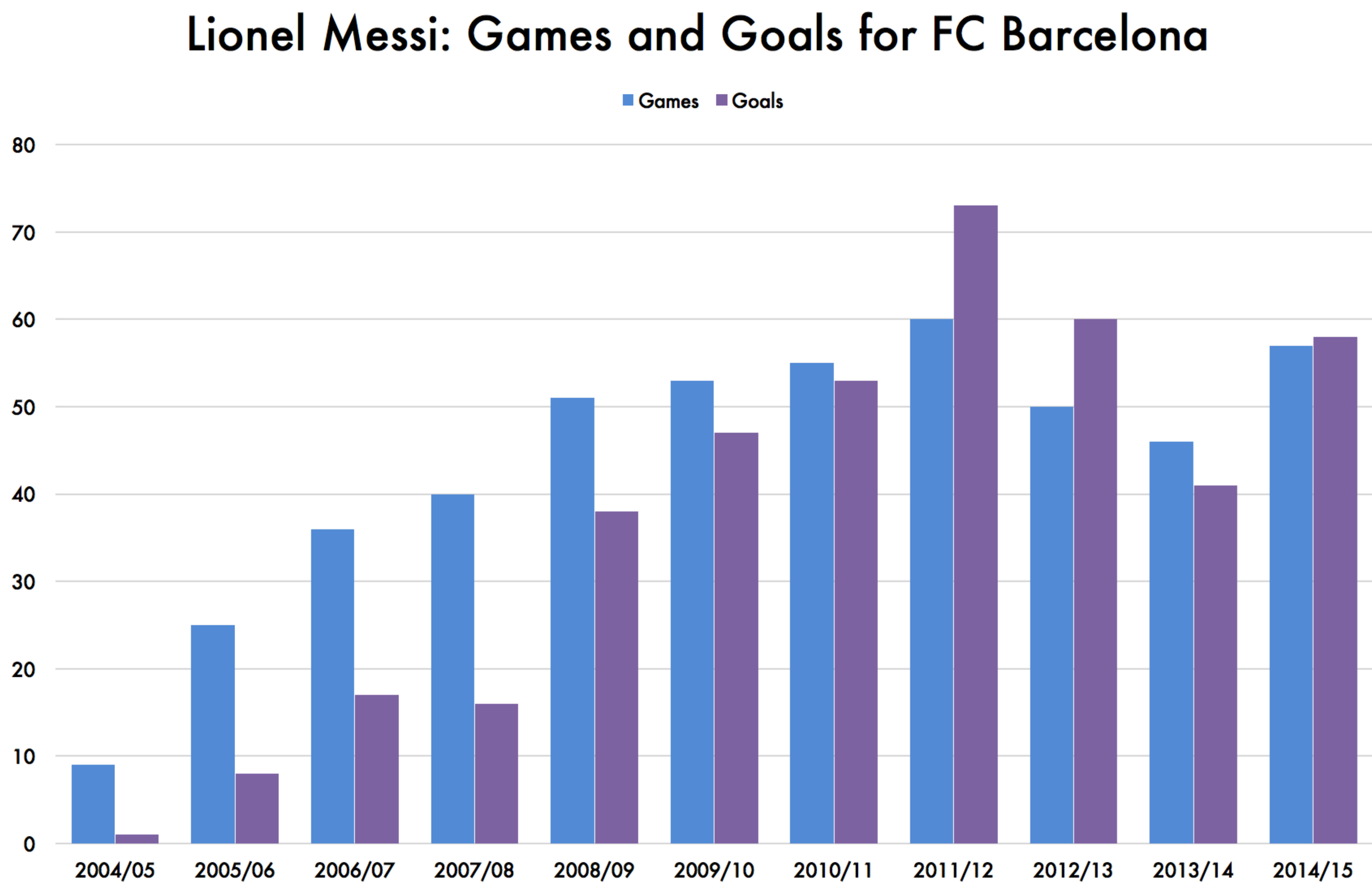


example

perceiving

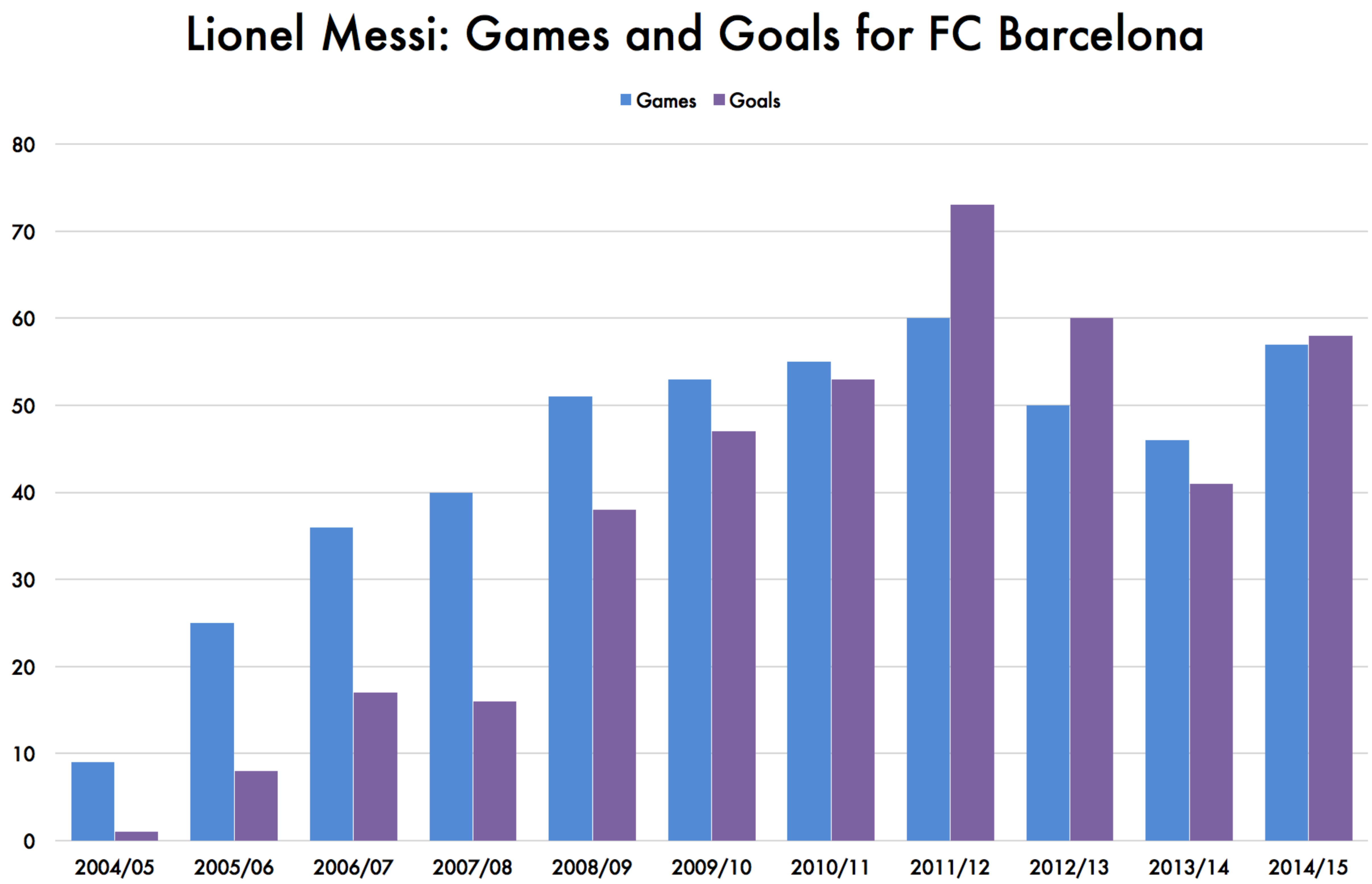
- Identifying chart type, axes and labels
- identifying big, small and medium values
- identifying trends for both columns
- making higher/lower comparison

example



interpreting

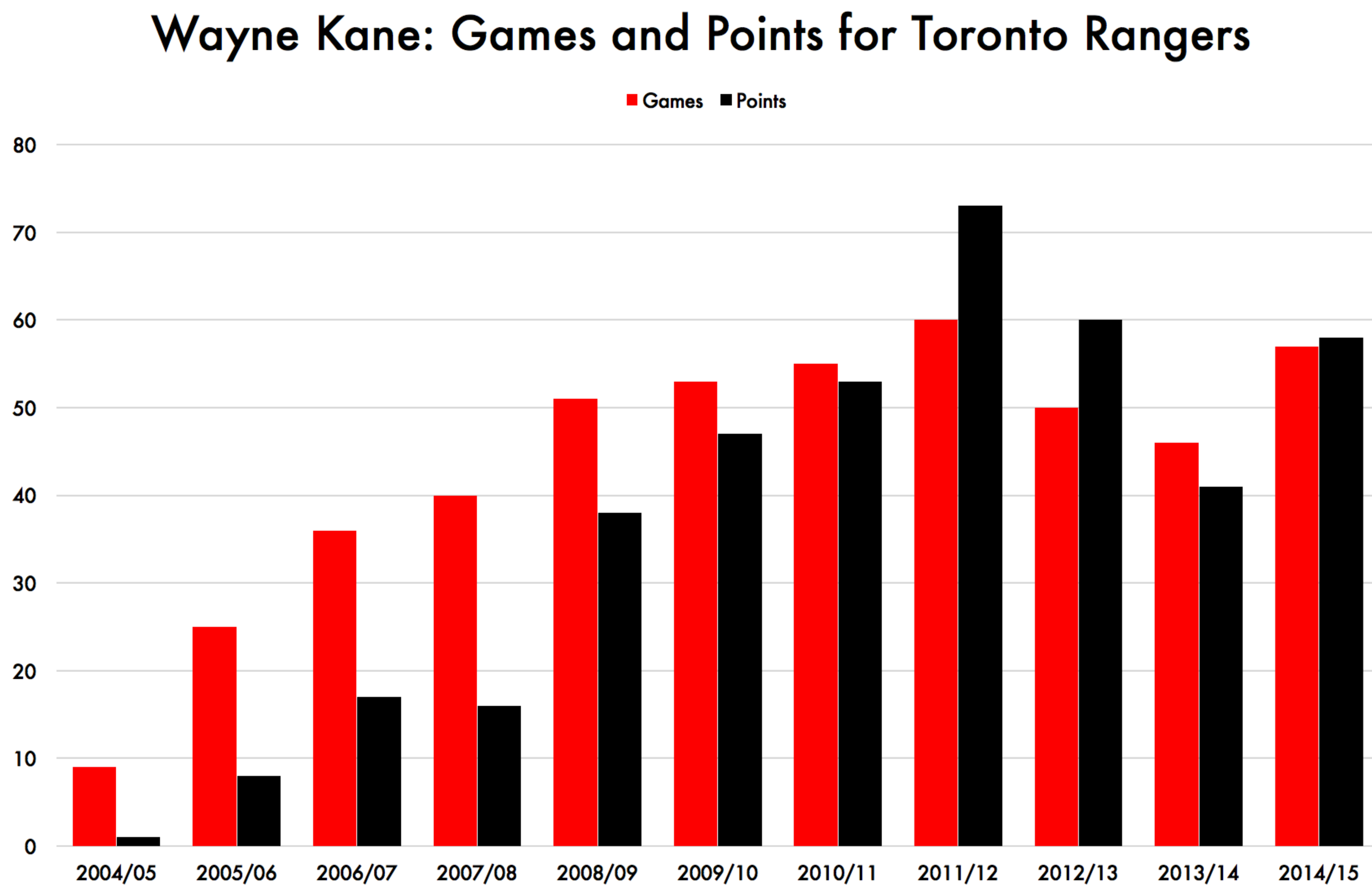
- previous knowledge: scoring 25 goals is good, scoring 70 is amazing
- scoring them in la lira / ucl is even more amazing
- a ratio of ~1 goal/game is very rare in football
- even more remarkable if coupled with messi's age



example

comprehending

- enforce my belief of rating messi one of the top world player
- what is related to the decreasing performances in 2012/2013 & 2013/2014?



example

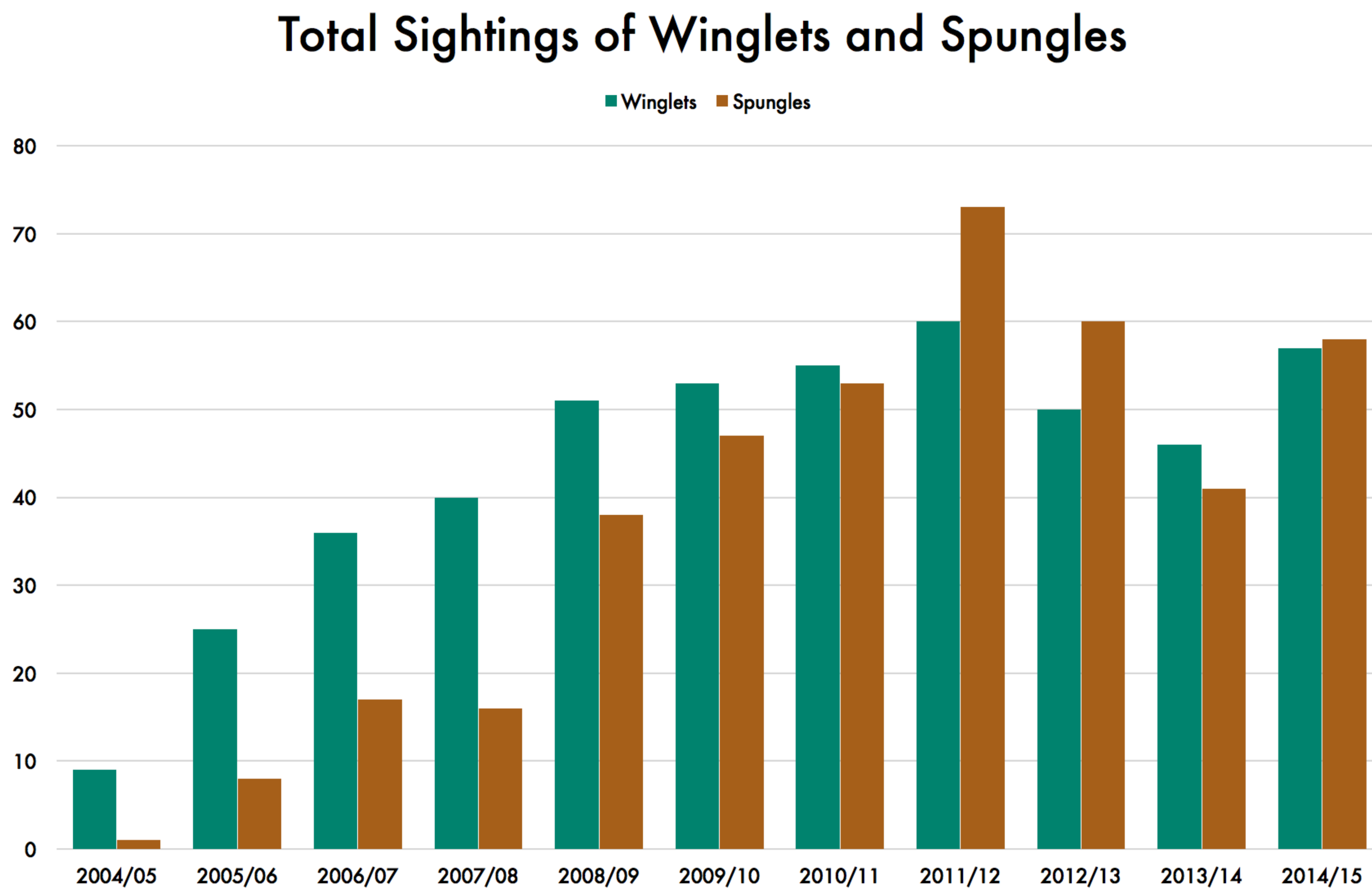
perceiving 

interpreting 

if I am no hockey expert

comprehending 

example



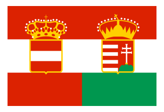
perceiving

interpreting

comprehending

wwii example

abraham wald

 1902  1950

ph.d. maths, uni vienna, 1928

member of columbia us
statistical research group
during wii

how to protect
bombers to get shot by
german planes?

How to optimise position of
additional protecting armor?

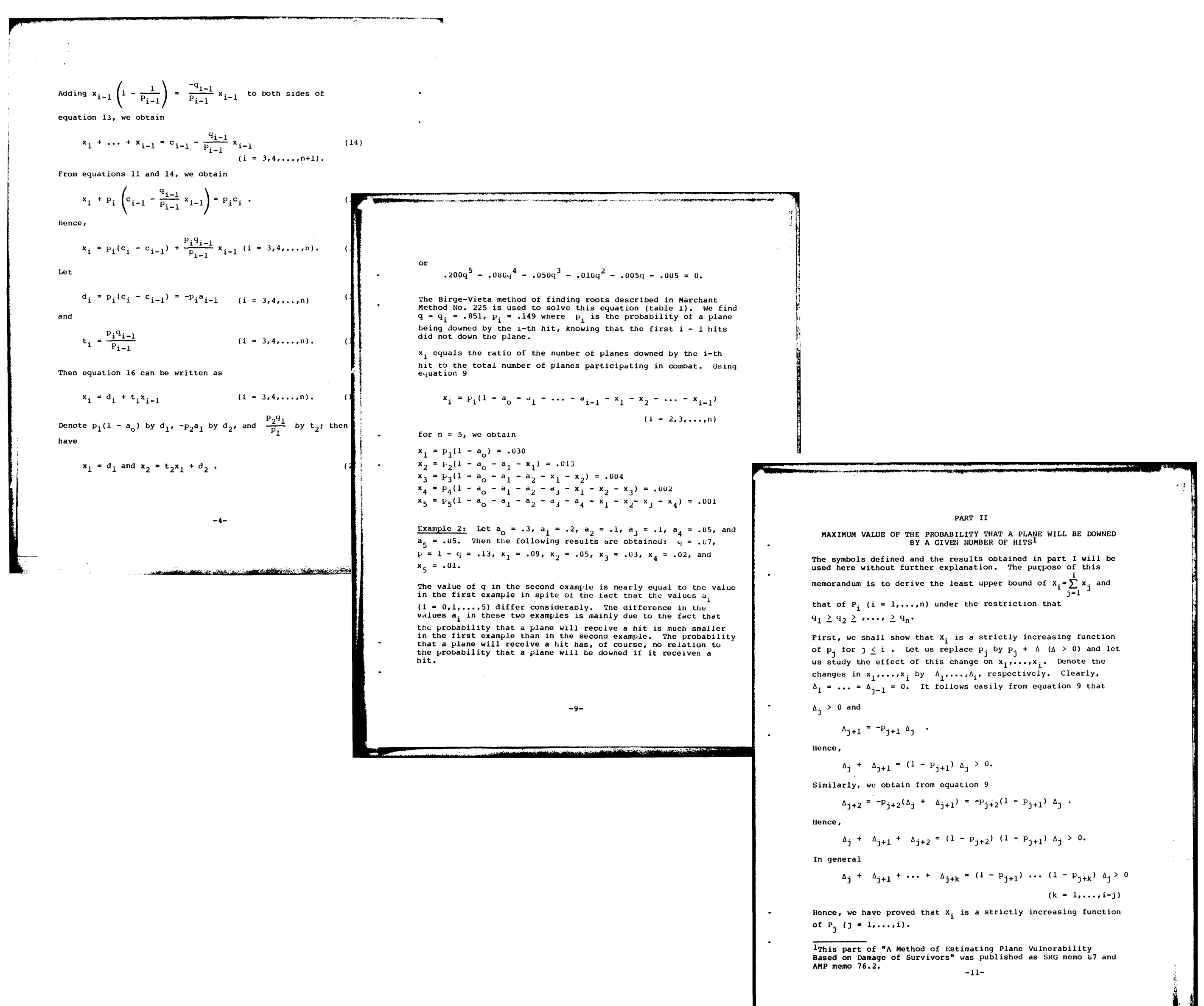
wwii example

data…

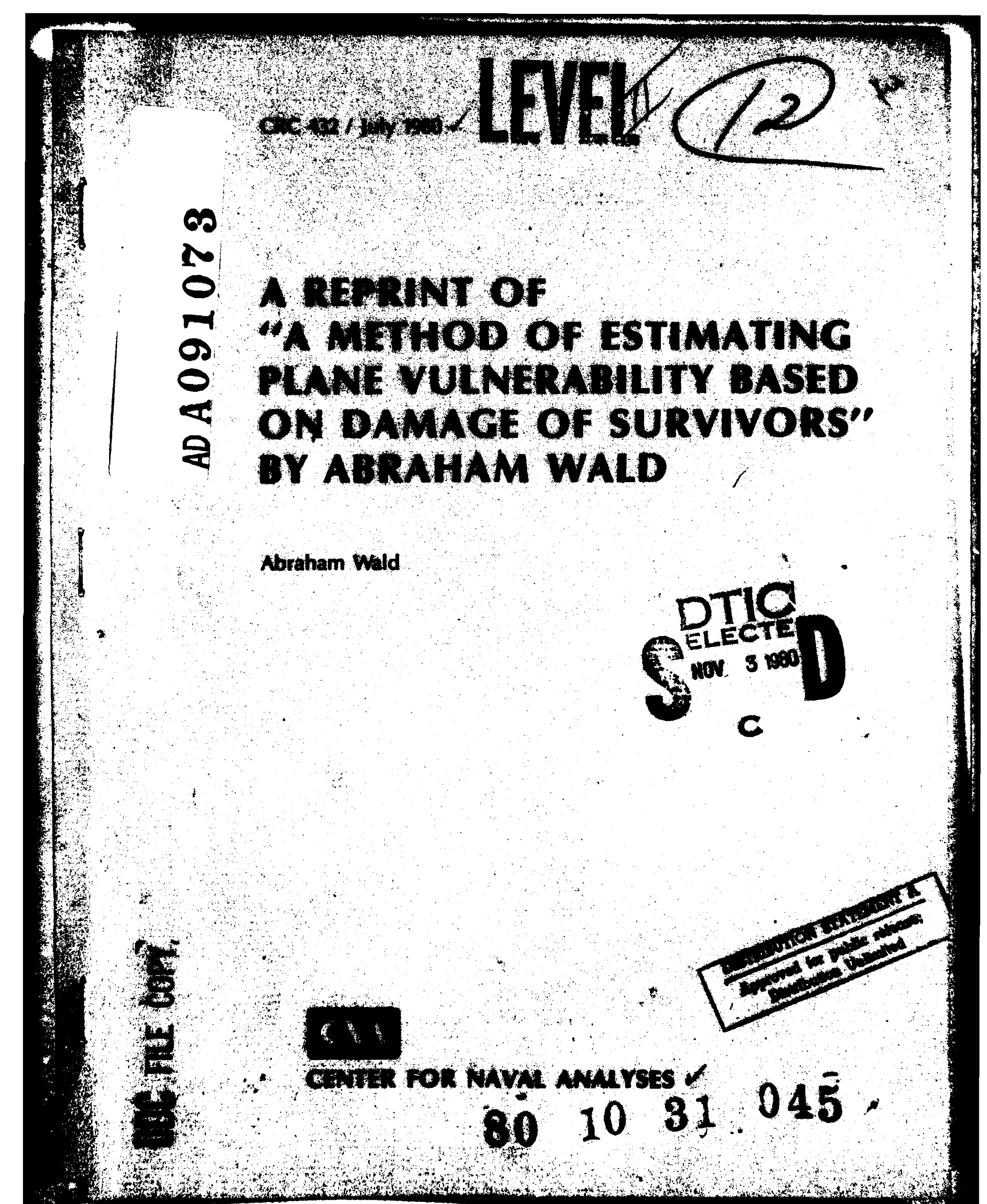
The hits on the returning planes were distributed as follows:

Part number	Number of hits observed on part	Ratio of number of hits observed on part to total number of observed hits ($\delta(i)$)
1	39	.193
2	78	.386
3	31	.154
4	54	.267
Total number of hits	202	

…model…



…solution



LEVEL

AD A091073

A REPRINT OF "A METHOD OF ESTIMATING PLANE VULNERABILITY BASED ON DAMAGE OF SURVIVORS" BY ABRAHAM WALD

Abraham Wald

DTIC ELECTE NOV 3 1980

CENTER FOR NAVAL ANALYSES

80 10 31 045

DDC FILE COPY

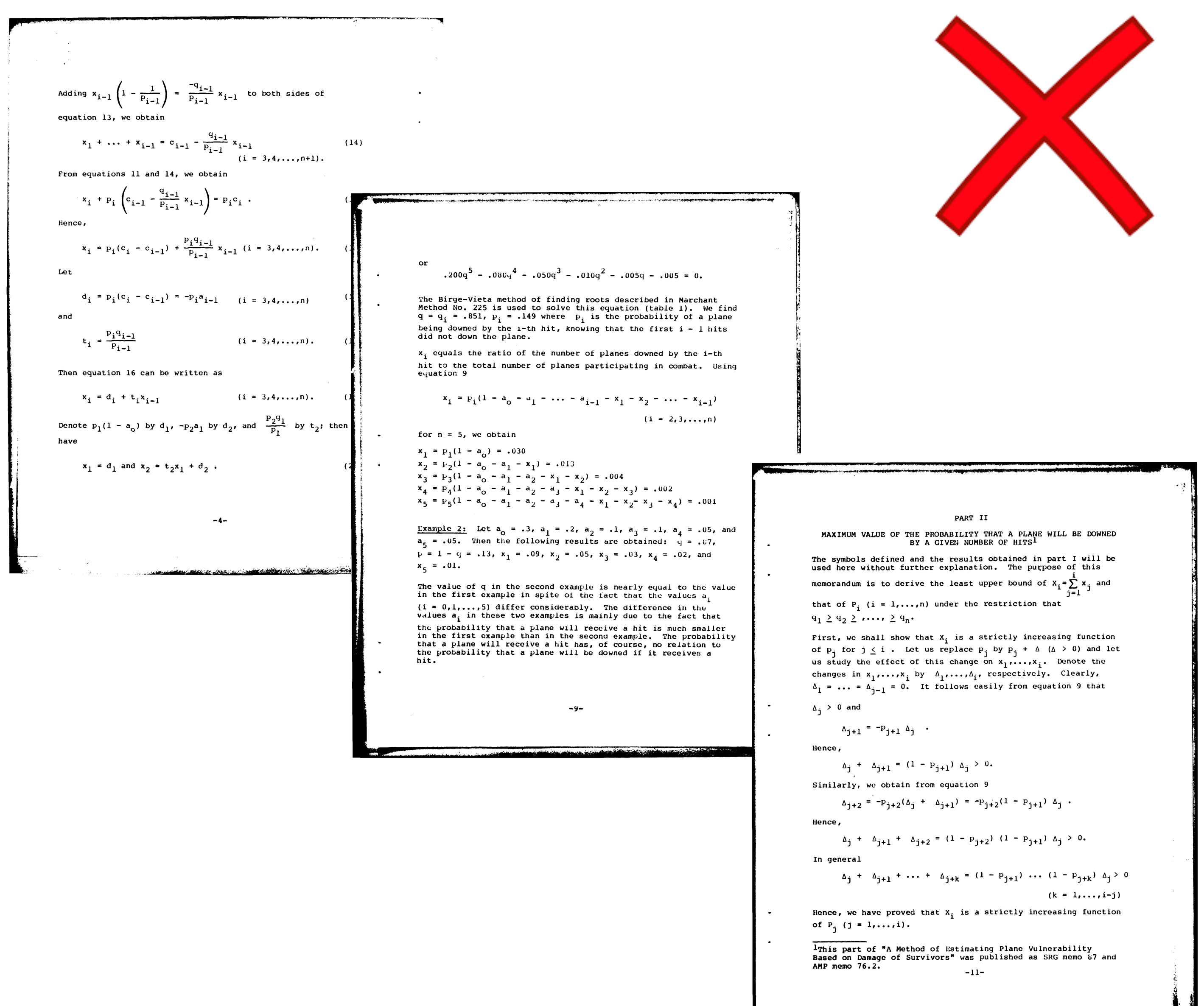
wwii example

data…

The hits on the returning planes were distributed as follows:

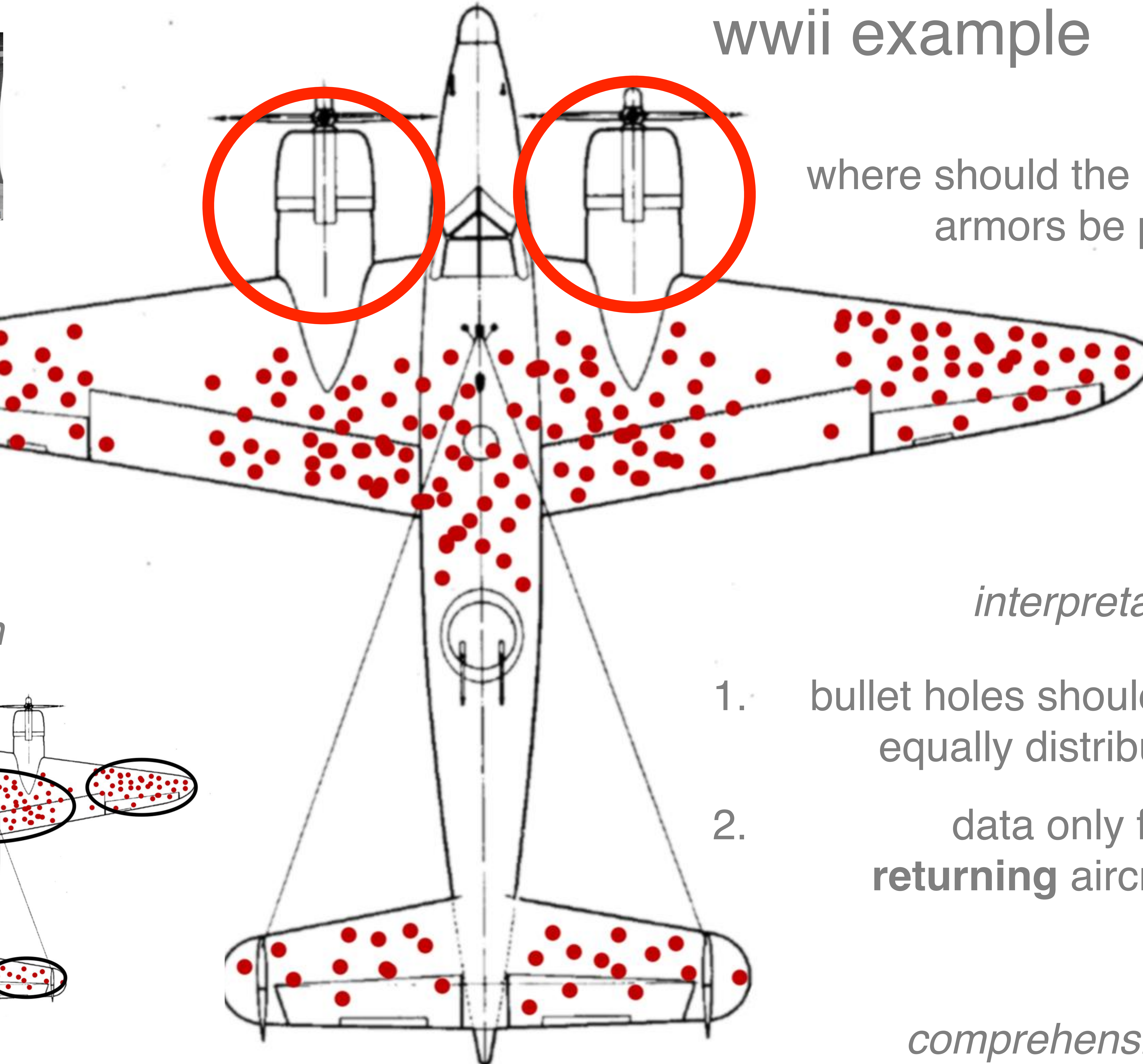
Part number	Number of hits observed on part	Ratio of number of hits observed on part to total number of observed hits ($\delta(i)$)
1	39	.193
2	78	.386
3	31	.154
4	54	.267
Total number of hits	202	

…model…

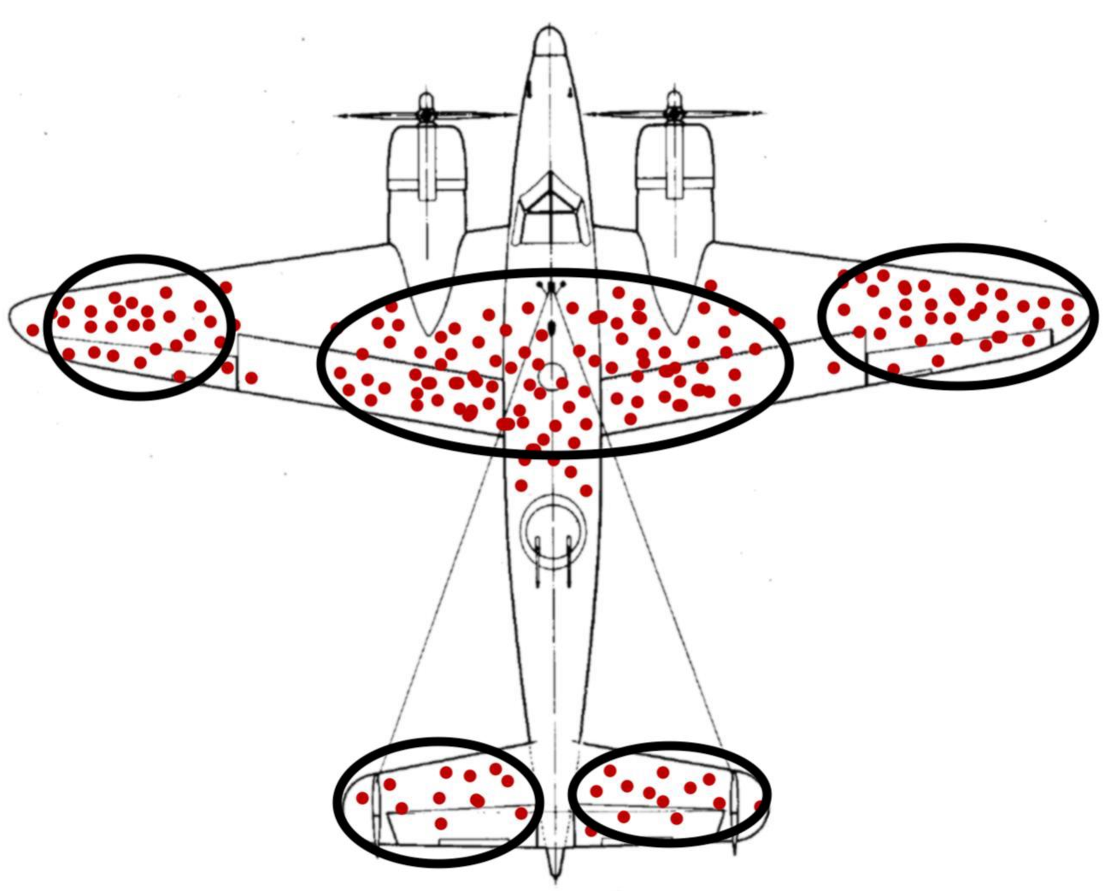


wwii example

where should the new armors be put?



perception

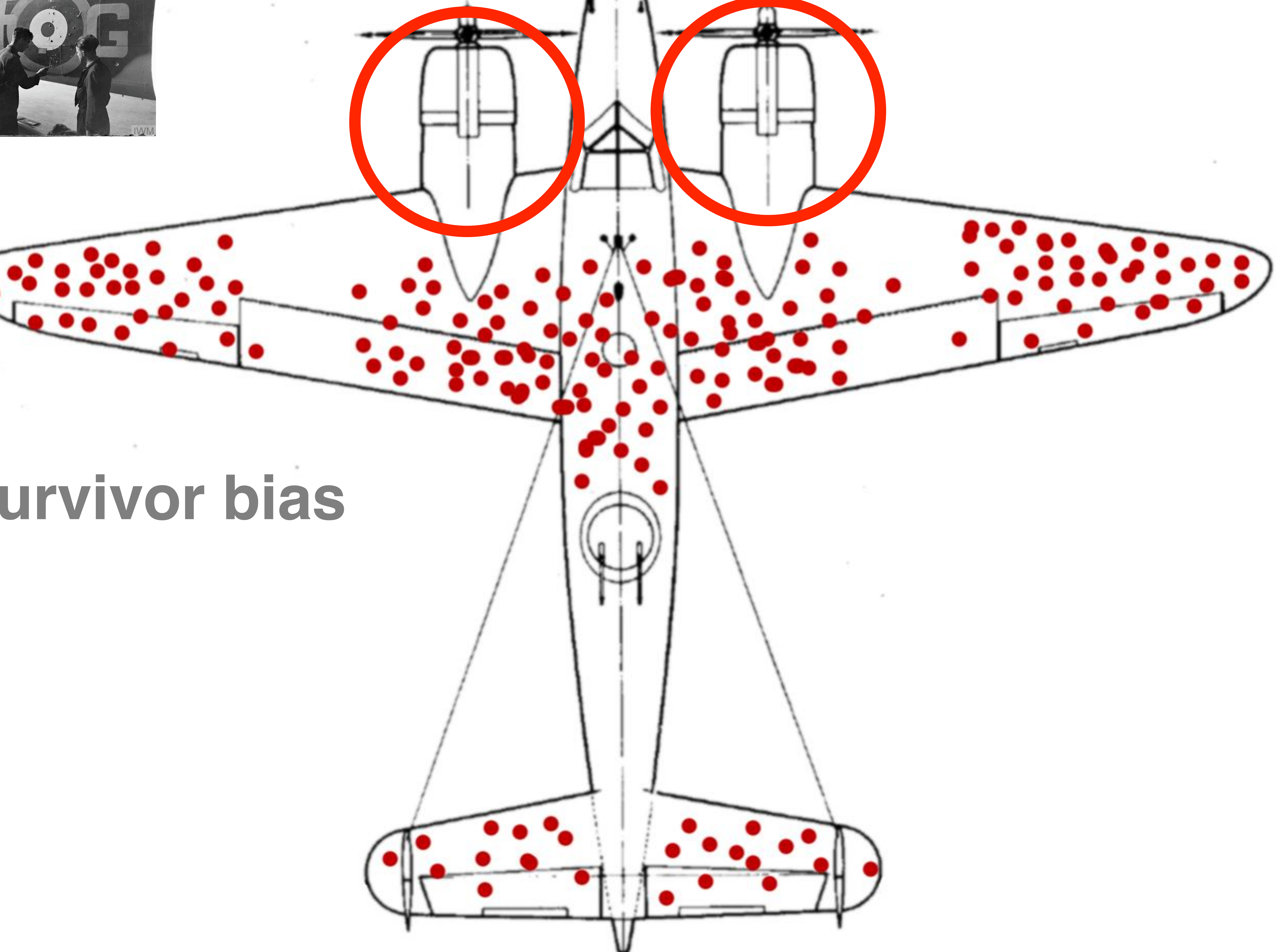


interpretation

1. bullet holes should be equally distributed
2. data only from **returning** aircrafts

comprehension?

wwii example



survivor bias

principles

inspired by d. rams

principles of good design

good design…

… is innovative

… makes a product useful

… is aestehetic

… makes a product understandable

… is unobtrusive

… is honest

… is long lasting

… is thorough down to the last detail

… is environmentally friendly

… is as little design as possible

principles

inspired by d. rams

principles of good dataviz

Principle 1	Principle 2	Principle 3
Good data visualisation is **TRUSTWORTHY**	Good data visualisation is **ACCESSIBLE**	Good data visualisation is **ELEGANT**

principles

Principle 1

Good data visualisation is **TRUSTWORTHY**

integrity + accuracy + legitimacy

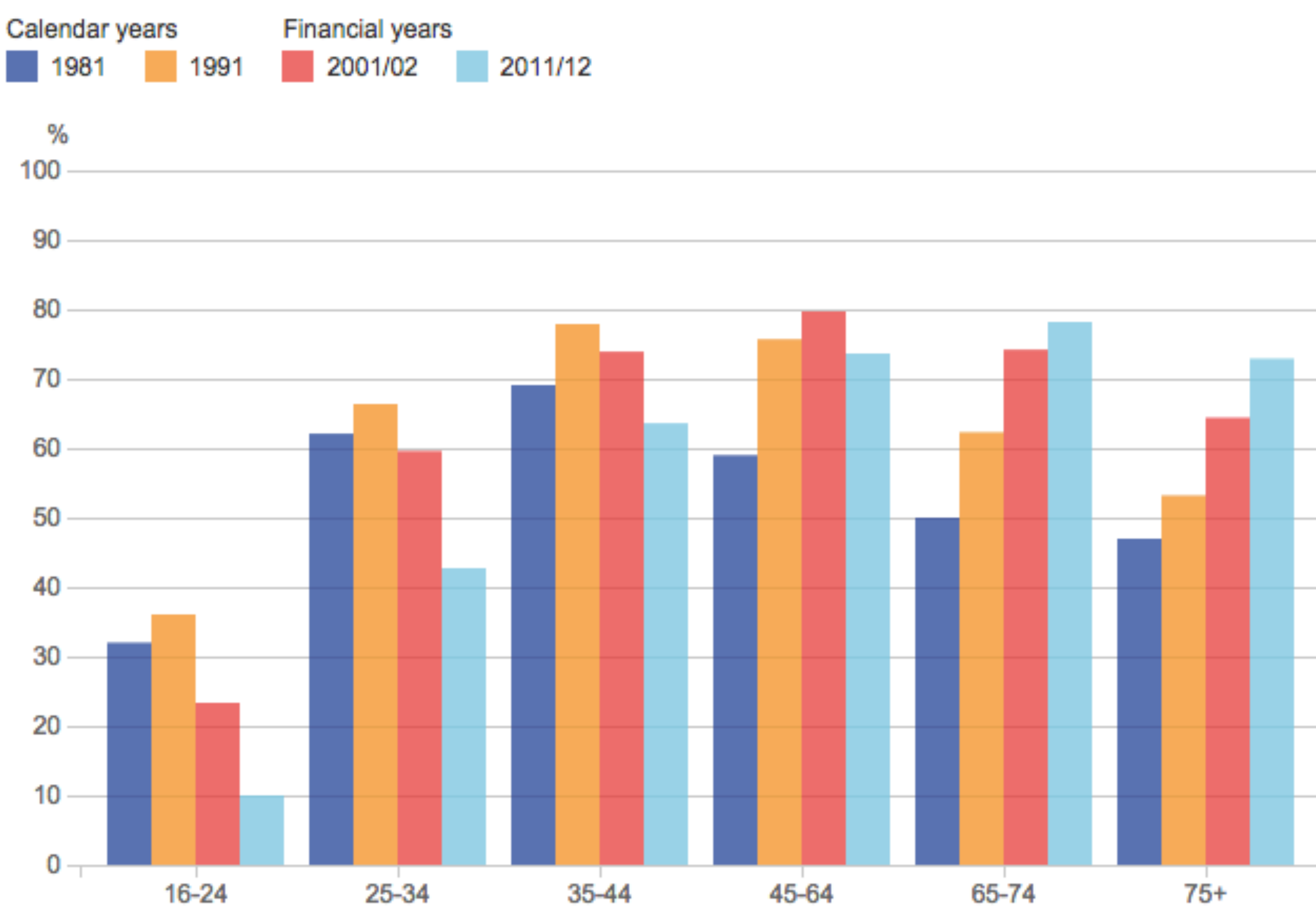
- **truth** is an obligation
- different legitimate versions of truth can exists in dataviz, being the outcome of different pathways
- pure objectivity is rarely possible in dataviz
- there is a need to show that the shown truth is **trustable**
- even a true graph may not be viewed as trustworthy

Principle 1

Good data visualisation is **TRUSTWORTHY**

principles

- colors
- fonts
- background
- data description
- data source



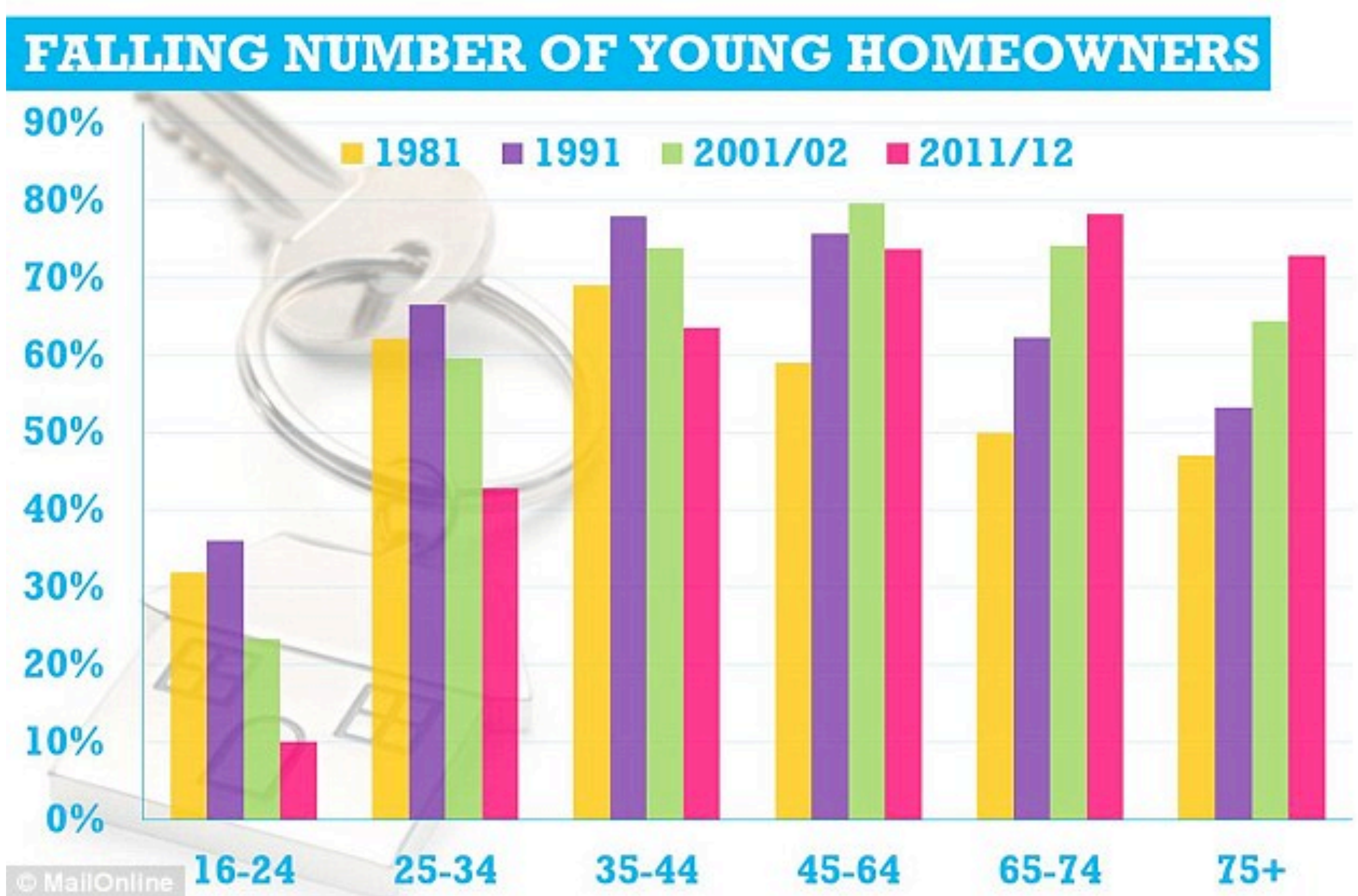
Percentage of each age group that are home owners[5], England, 1981 to 2012

Source: English Housing Survey (EHS) 2012 to 2013, Table FC2101, DCLG ; EHS 2001/02, Table S106, DCLG

uk office for national statistics

same true data,
different trustworthiness

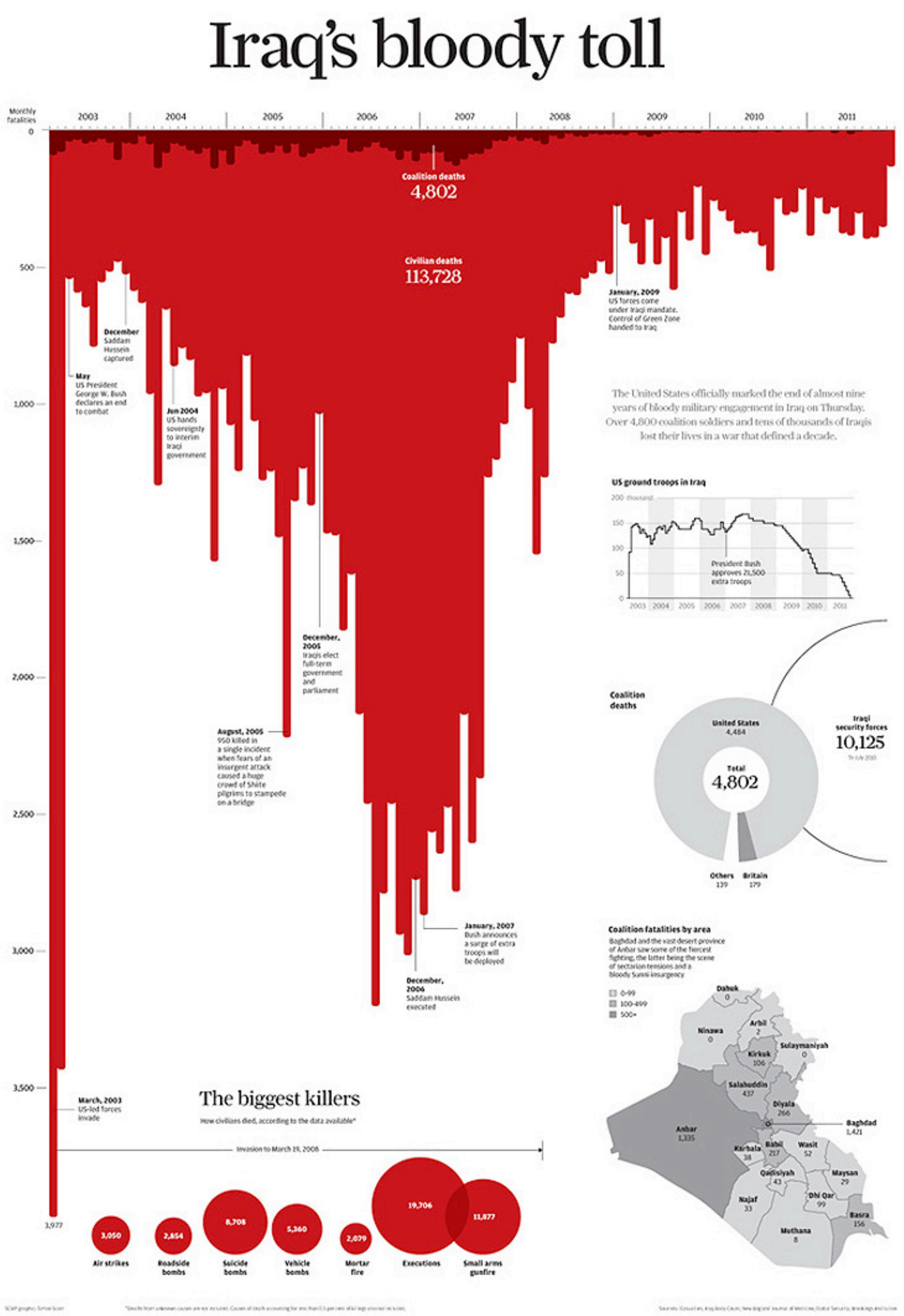
daily mail

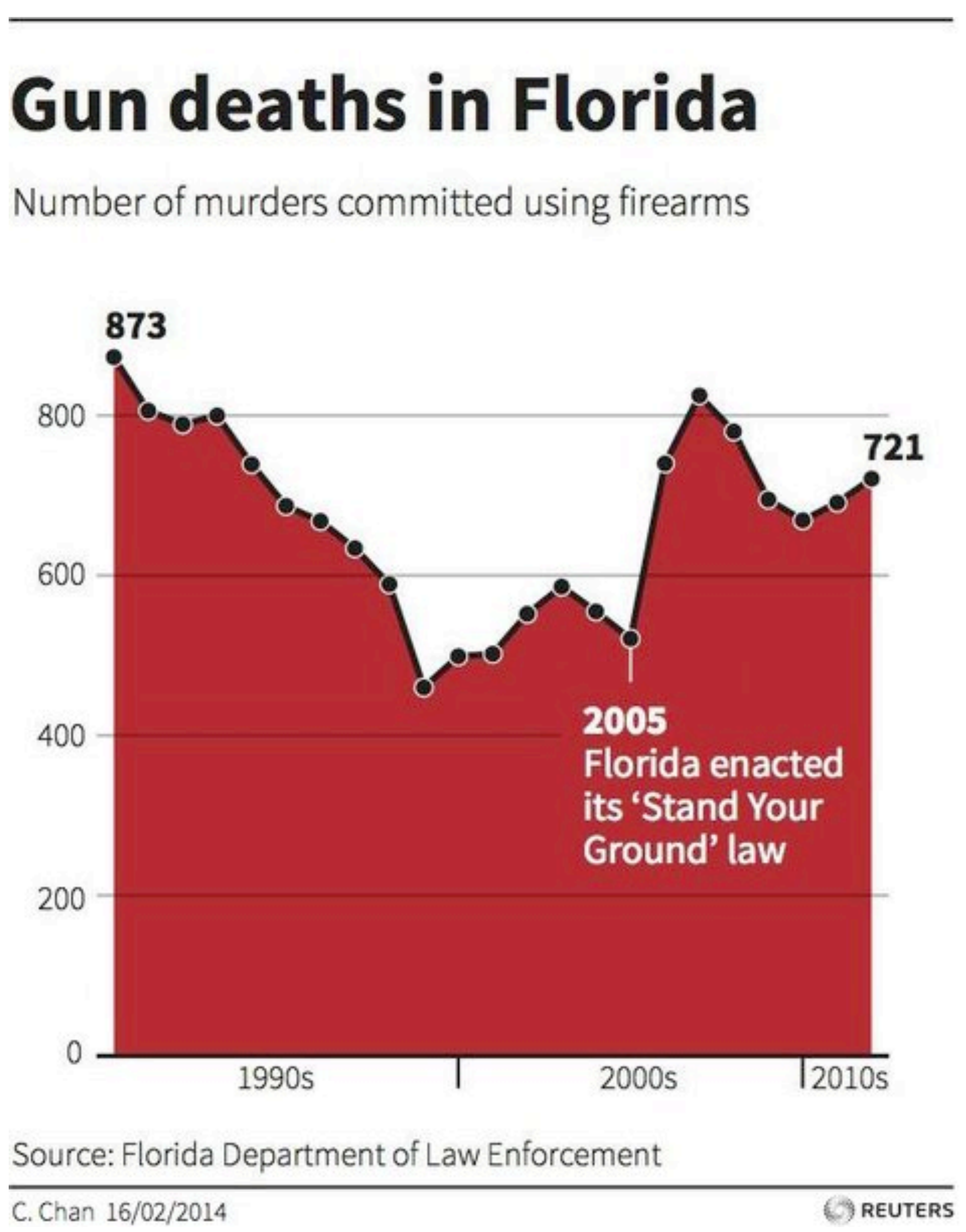


Principle 1

Good data visualisation is **TRUSTWORTHY**

principles





Principle 1

Good data visualisation is **TRUSTWORTHY**

principles

pursuing trustworthiness in **data processing**

- how was data collected: from where & using what criteria?
- what calculation of modification have you applied to it?
- explain the approach in details
- have you made any significant assumption or observed any special counting rules that may not be common?
- have you **removed** or **excluded** any data?
- how representative it is?
- what biases may exist that could distort interpretation?

Principle 1

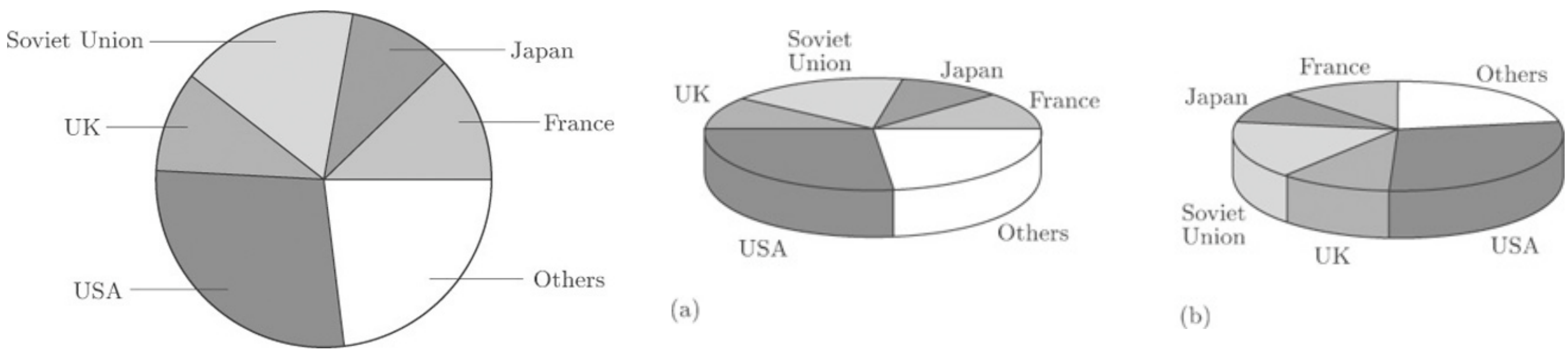
Good data visualisation is **TRUSTWORTHY**

principles

pursuing trustworthiness in **data representation**

- never **deceive** the receiver

- avoid misunderstandings, inaccuracies, confusions and distortions
  - ★ quantitative values represented by areas can be disproportionally perceived



[Open University]

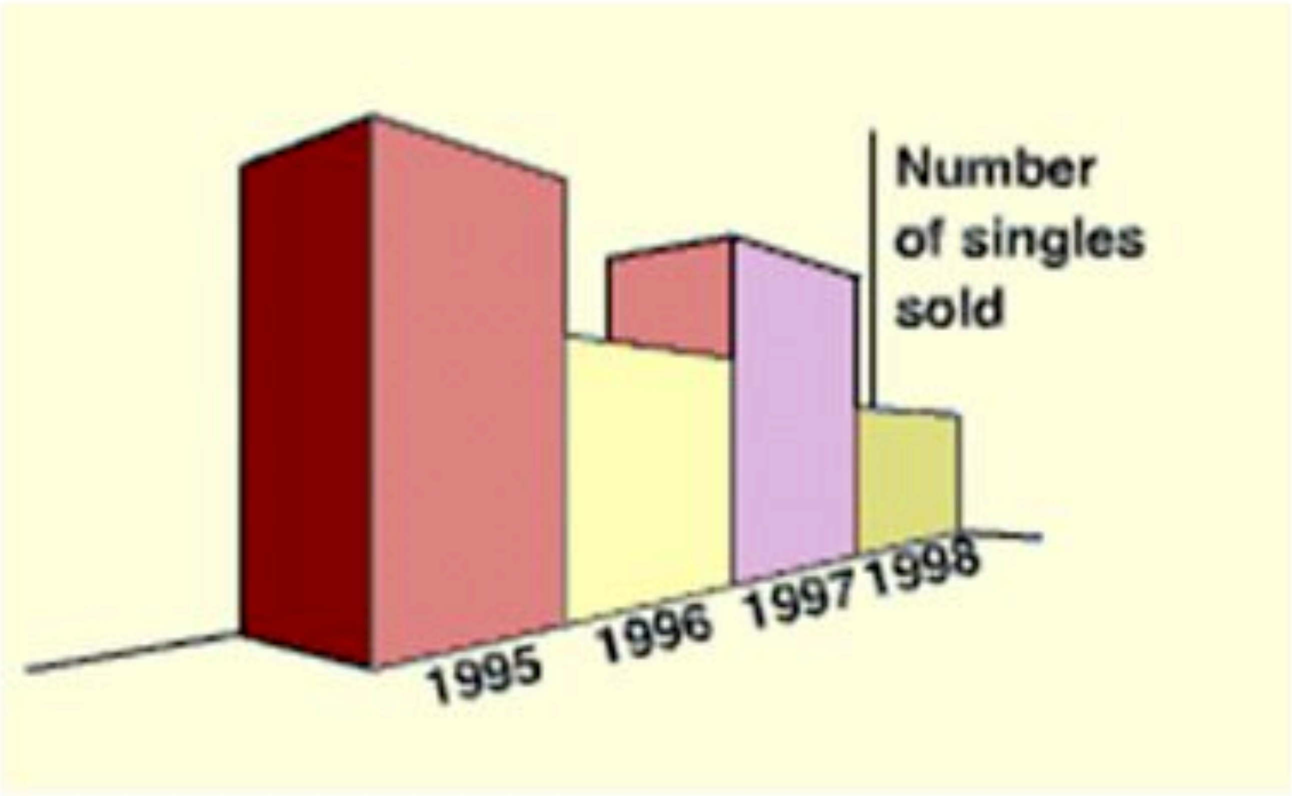
Principle 1

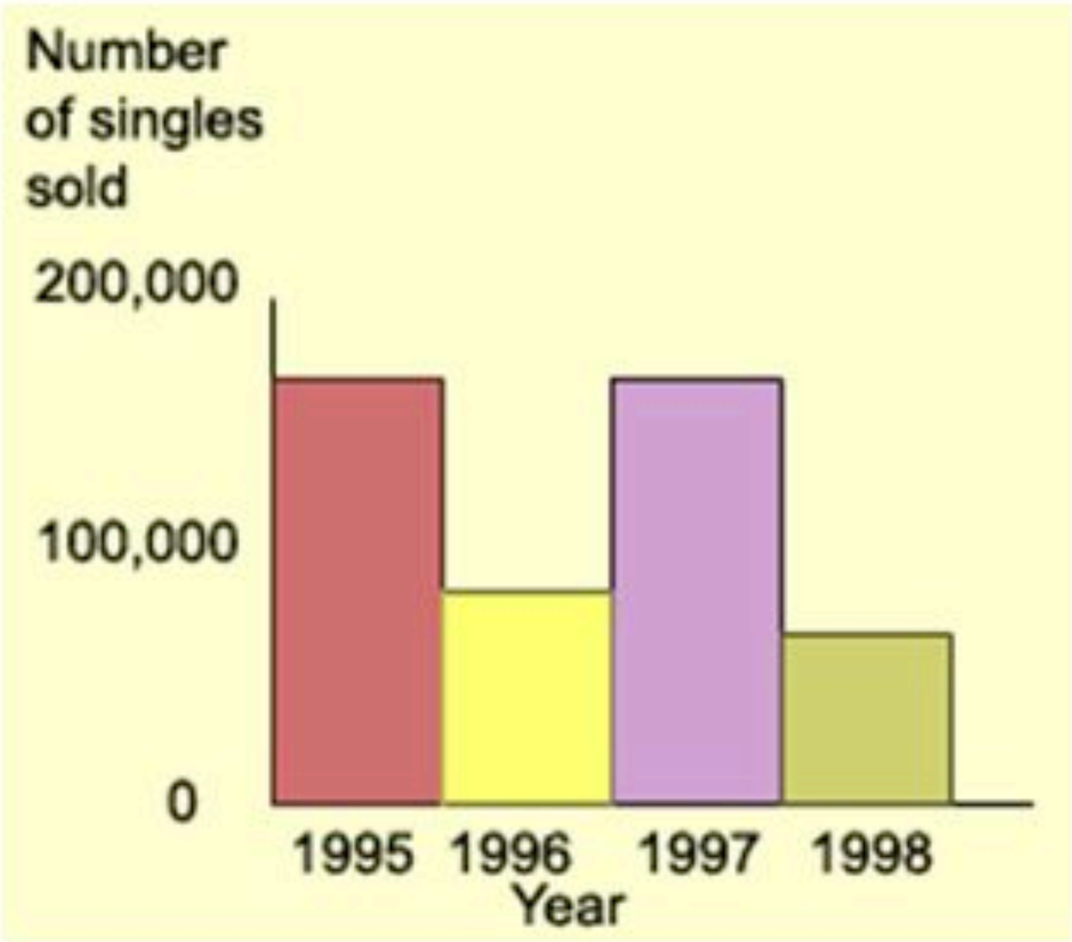
Good data visualisation is **TRUSTWORTHY**

principles

pursuing trustworthiness in **data representation**

- never **deceive** the receiver
- avoid misunderstandings, inaccuracies, confusions and distortions
  - ⋆ 3d representations are often nothing more than distraction — distortion — decoration; use only if there are 3 dimensions in data and the viewer can change point of view to see different 2d perspectives





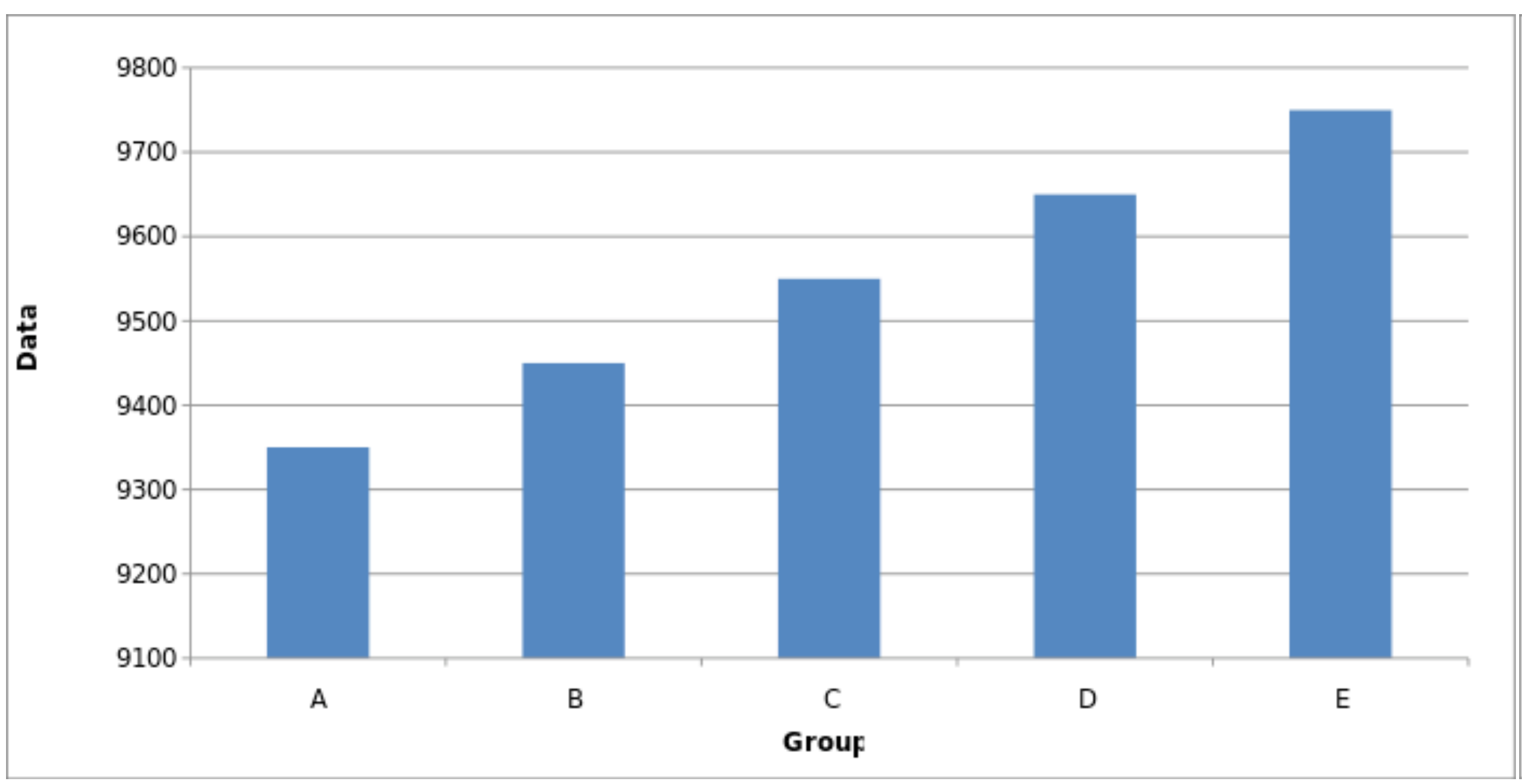
Principle 1

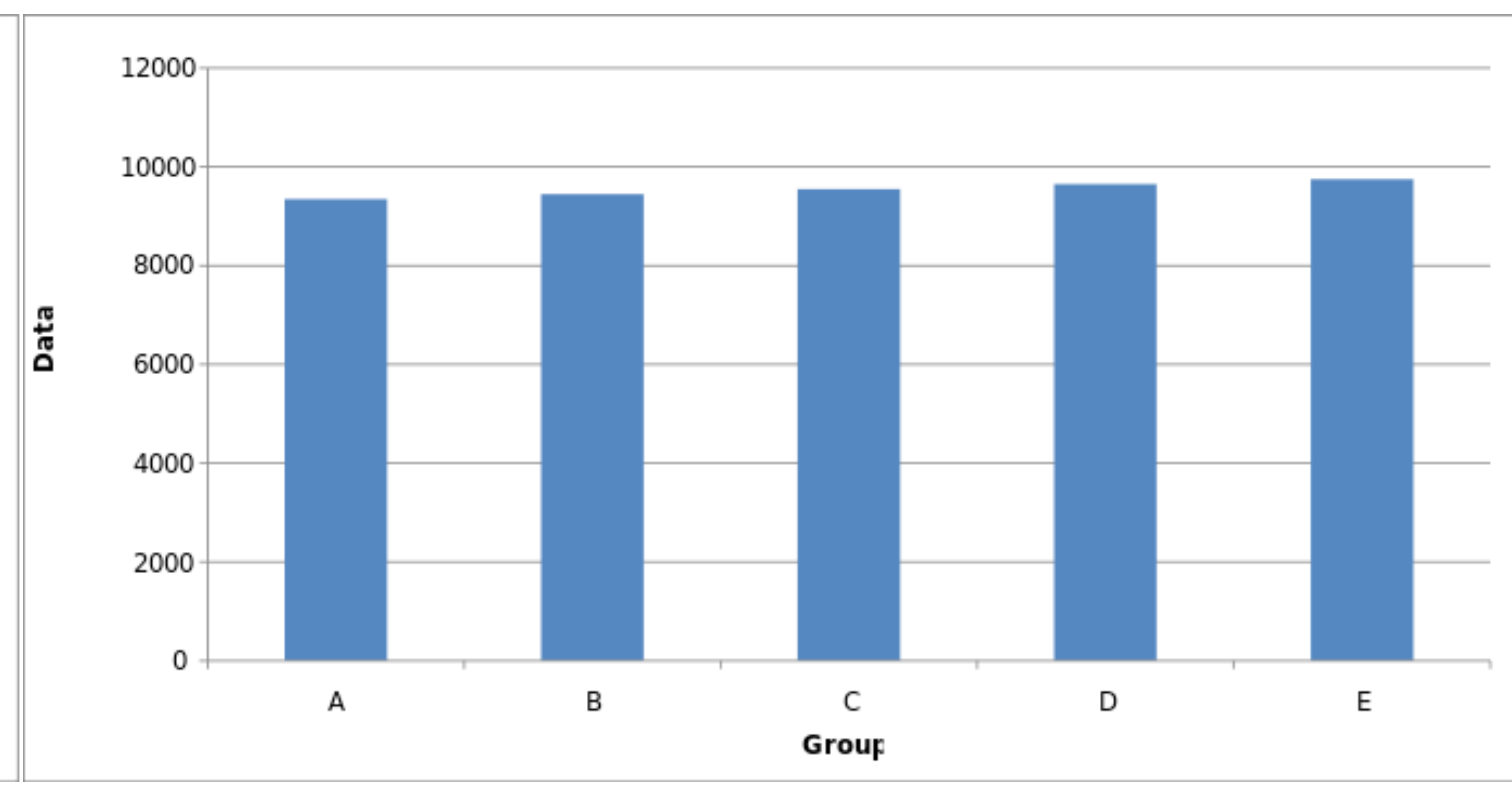
Good data visualisation is **TRUSTWORTHY**

principles

pursuing trustworthiness in **data representation**

- never **deceive** the receiver
- avoid misunderstandings, inaccuracies, confusions and distortions
  - ★ axes (in bar charts) should never be truncated, origin must be 0





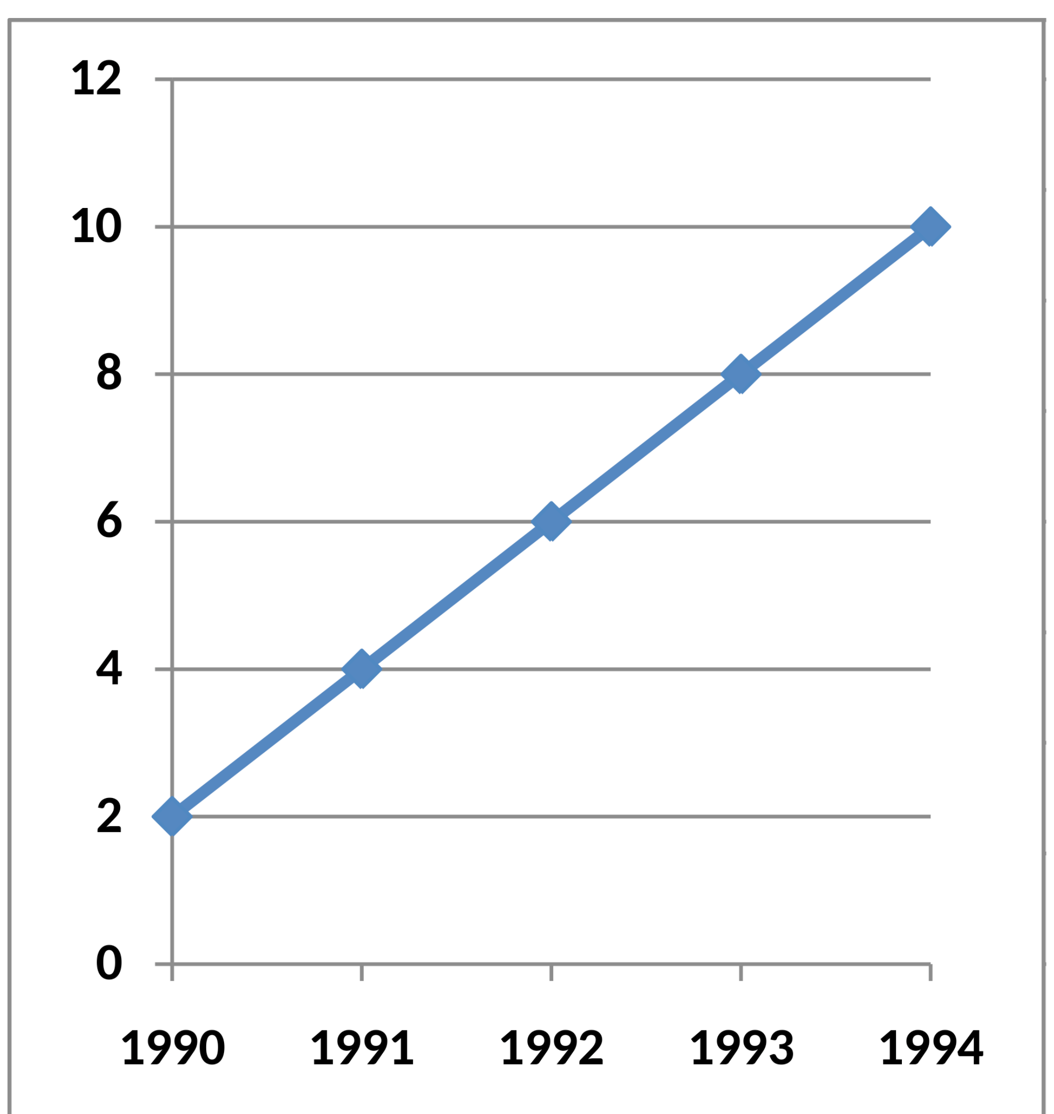
Principle 1

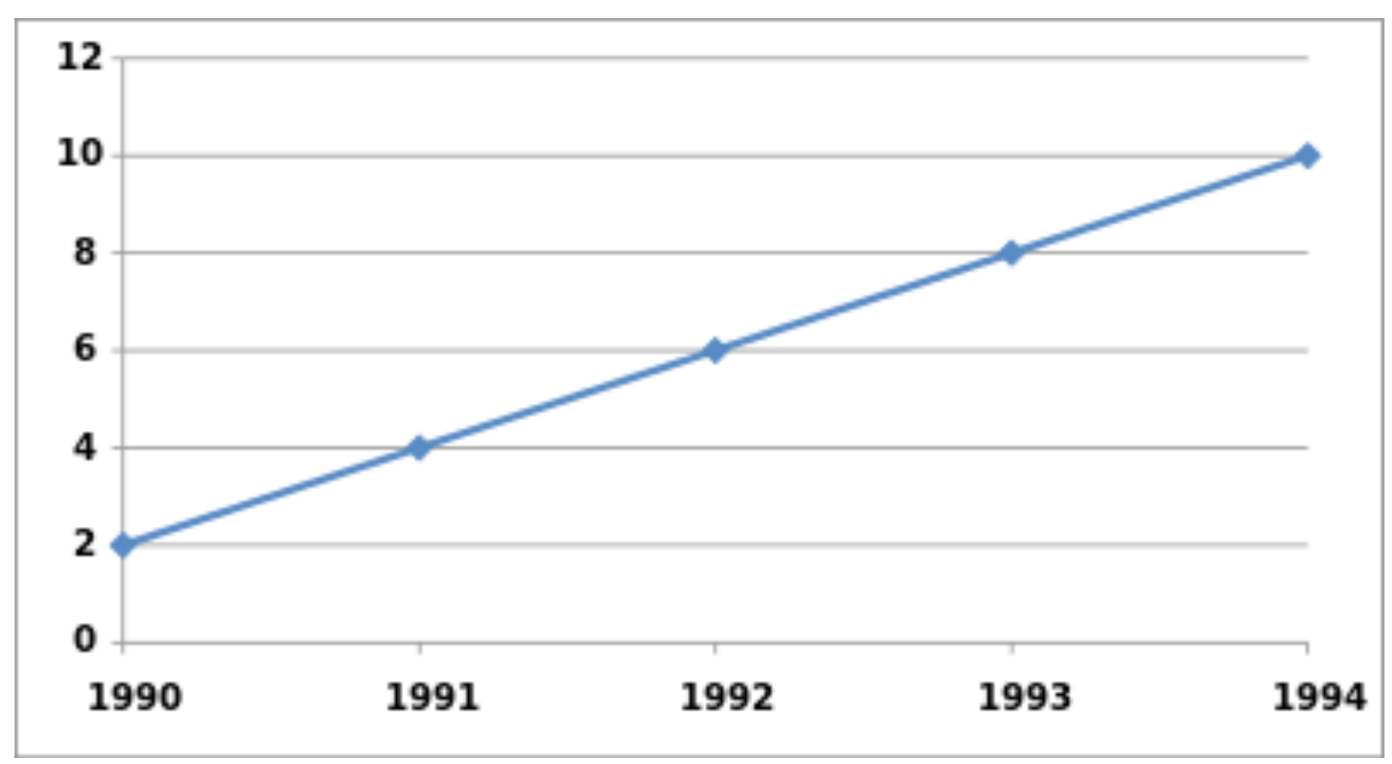
Good data visualisation is **TRUSTWORTHY**

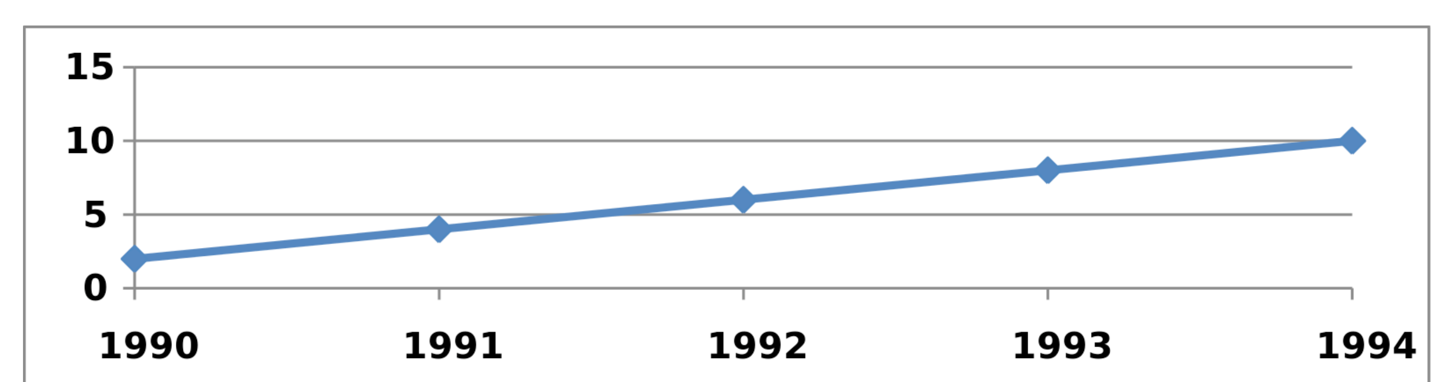
principles

pursuing trustworthiness in **data representation**

- never **deceive** the receiver
- avoid misunderstandings, inaccuracies, confusions and distortions
  - ★ aspect ratio in line chart is influential, since it modifies the perception of the steepness of connecting lines, either embellishing it or dampening it







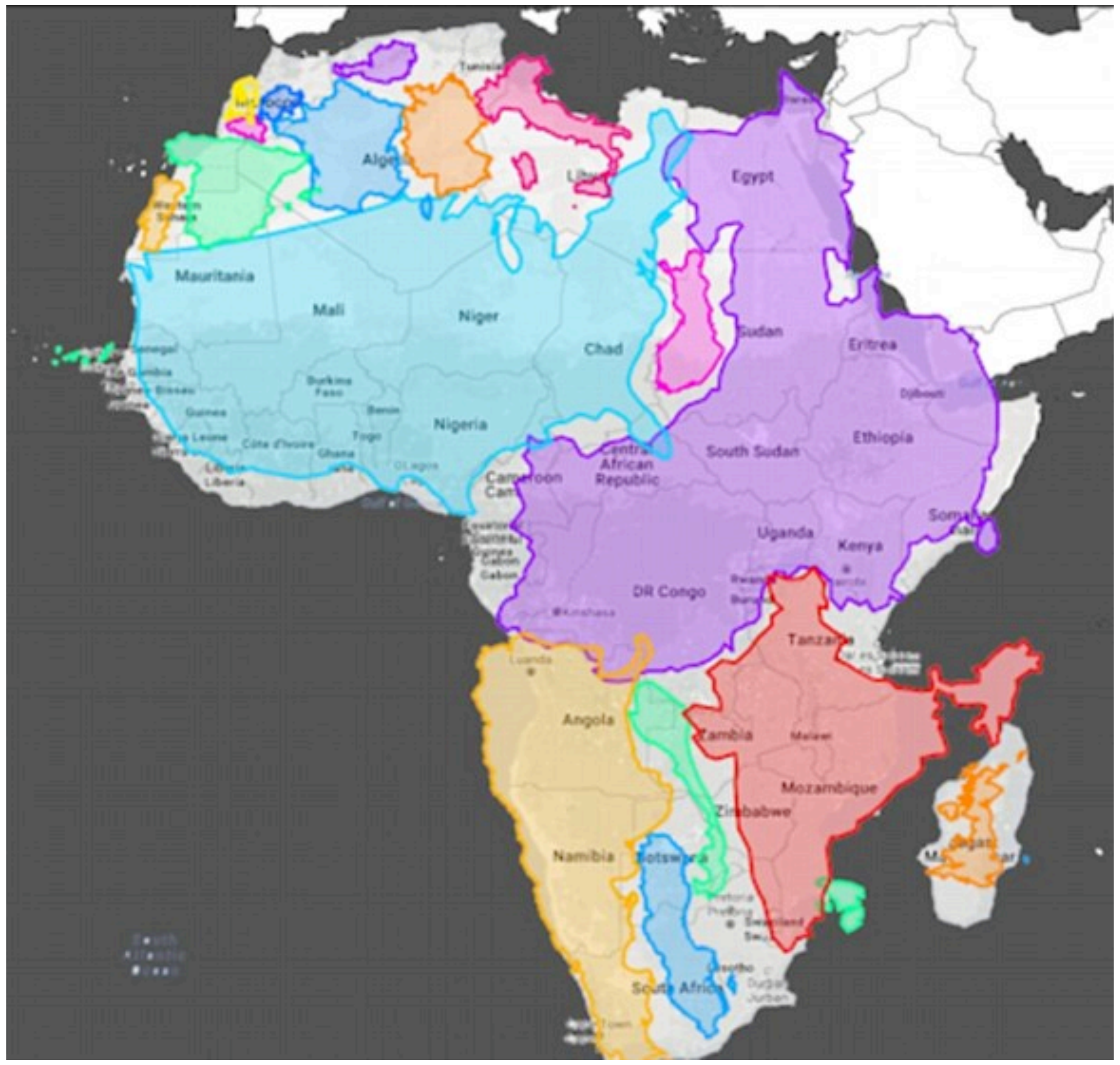
Principle 1

Good data visualisation is **TRUSTWORTHY**

principles

pursuing trustworthiness in **data representation**

- never **deceive** the receiver
- avoid misunderstandings, inaccuracies, confusions and distortions
  - ⋆ different projections (mercator, etc) may alter the perception of a thematic map, distorting size/shape of regions.



Principle 1

Good data visualisation is **TRUSTWORTHY**

principles

pursuing trustworthiness in **data presentation**

- if it looks significant, it should be
- absent annotations such as introduction/guides, axis title and labels, footnotes, data sources fail to inform the reader
- inconsistent or inappropriate color schemas
- confusing layouts
- does the chosen solution work, and, specifically, does it work in the way it promises to do?

Principle 2

Good data visualisation is **ACCESSIBLE**

principles

the viewer should experience minimum friction between the act of understanding (**effort**) and the achieving of understanding (**reward**)

reader's side:

- subject/matter appeal
- dynamic of need
- subject/matter knowledge
- what they need to know
- issue of unfamiliar representation
- time
- format
- personal taste
- attitude & emotion

author's side

- lack of focus
- not deep enough
- oversimplified representation
- unfit for the setting
- visually inaccessible
- misjudged format, details get lost
- too many interactive options
- too complex representation
- absent annotations

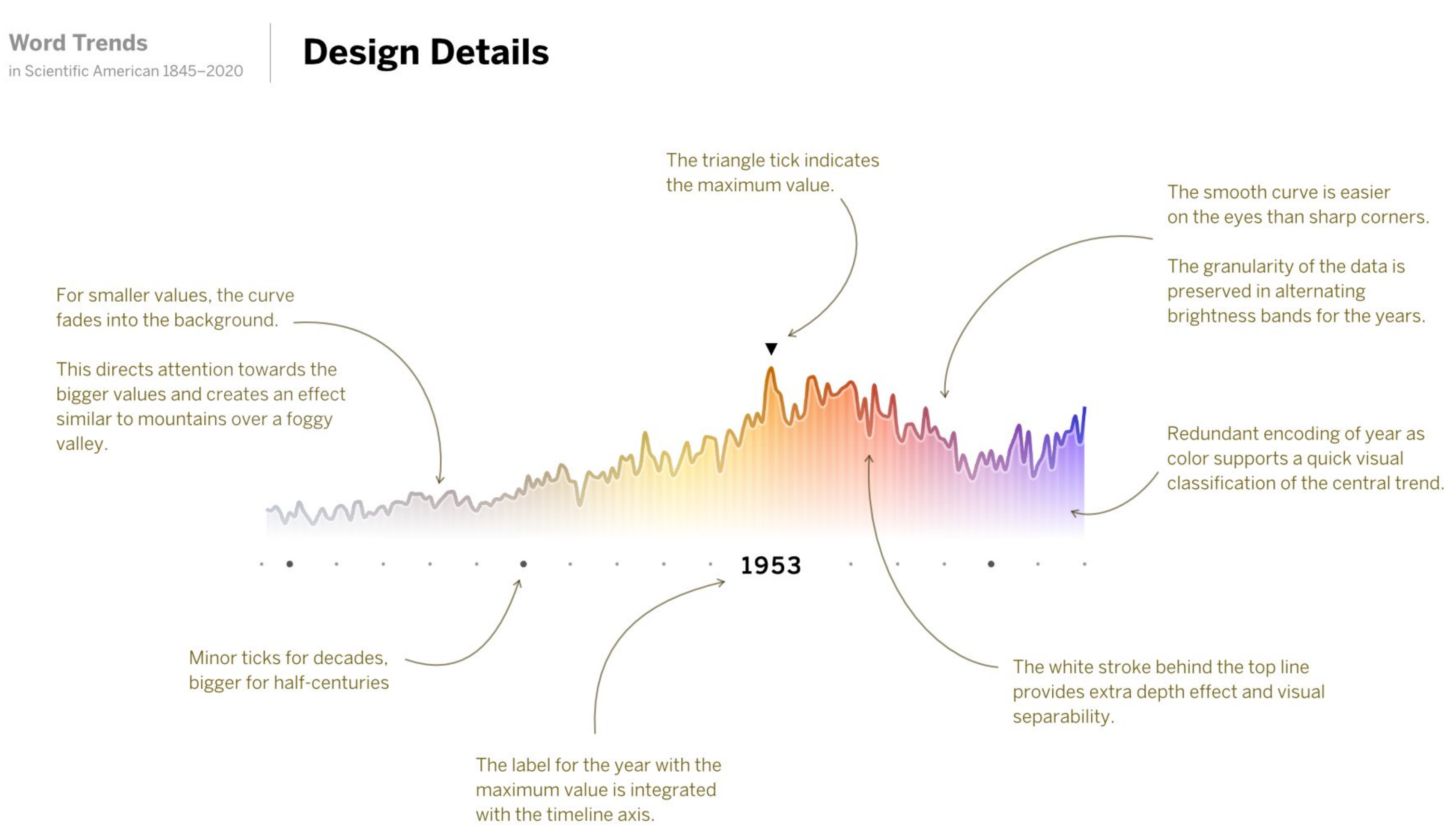
Principle 3

Good data visualisation is **ELEGANT**

principles

achieving a visual quality to attract your audience, without letting the style overcome the substance

- eliminate the arbitrary: ‘’remove to improve”
- thoroughness
- develop a style
- decoration should be additive, not negative
- offer elegant & appealing presentation congruent with the subject
- do not pursue minimalism at all costs



Source: https://truth-and-beauty.net/

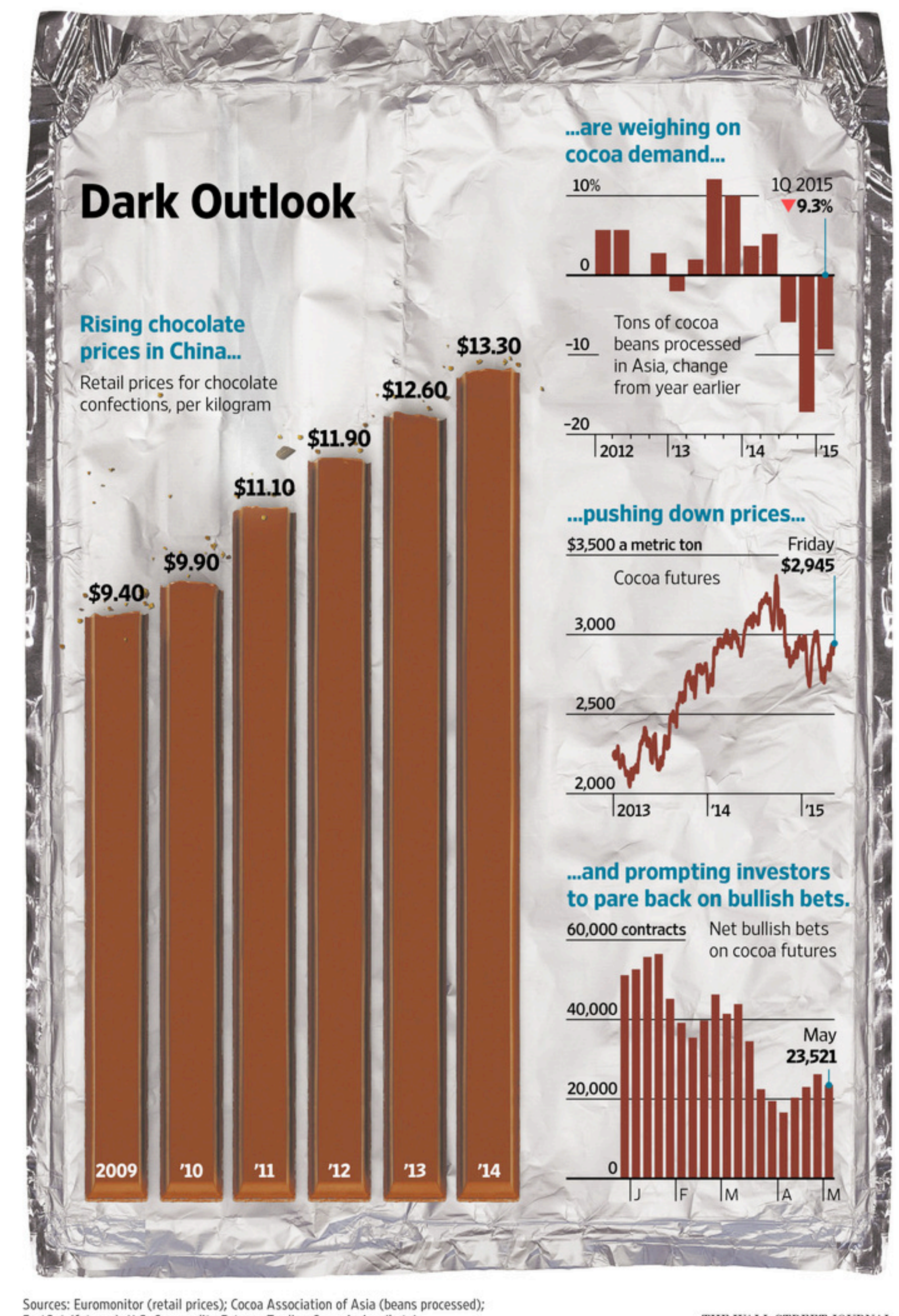
Principle 3

Good data visualisation is **ELEGANT**

principles

achieving a visual quality to attract your audience, without letting the style overcome the substance

- eliminate the arbitrary: ‘’remove to improve”
- thoroughness
- develop a style
- decoration should be additive, not negative
- offer elegant & appealing presentation congruent with the subject
- do not pursue minimalism at all costs



Principle 3

Good data visualisation is **ELEGANT**

principles

achieving a visual quality to attract your audience, without letting the style overcome the substance

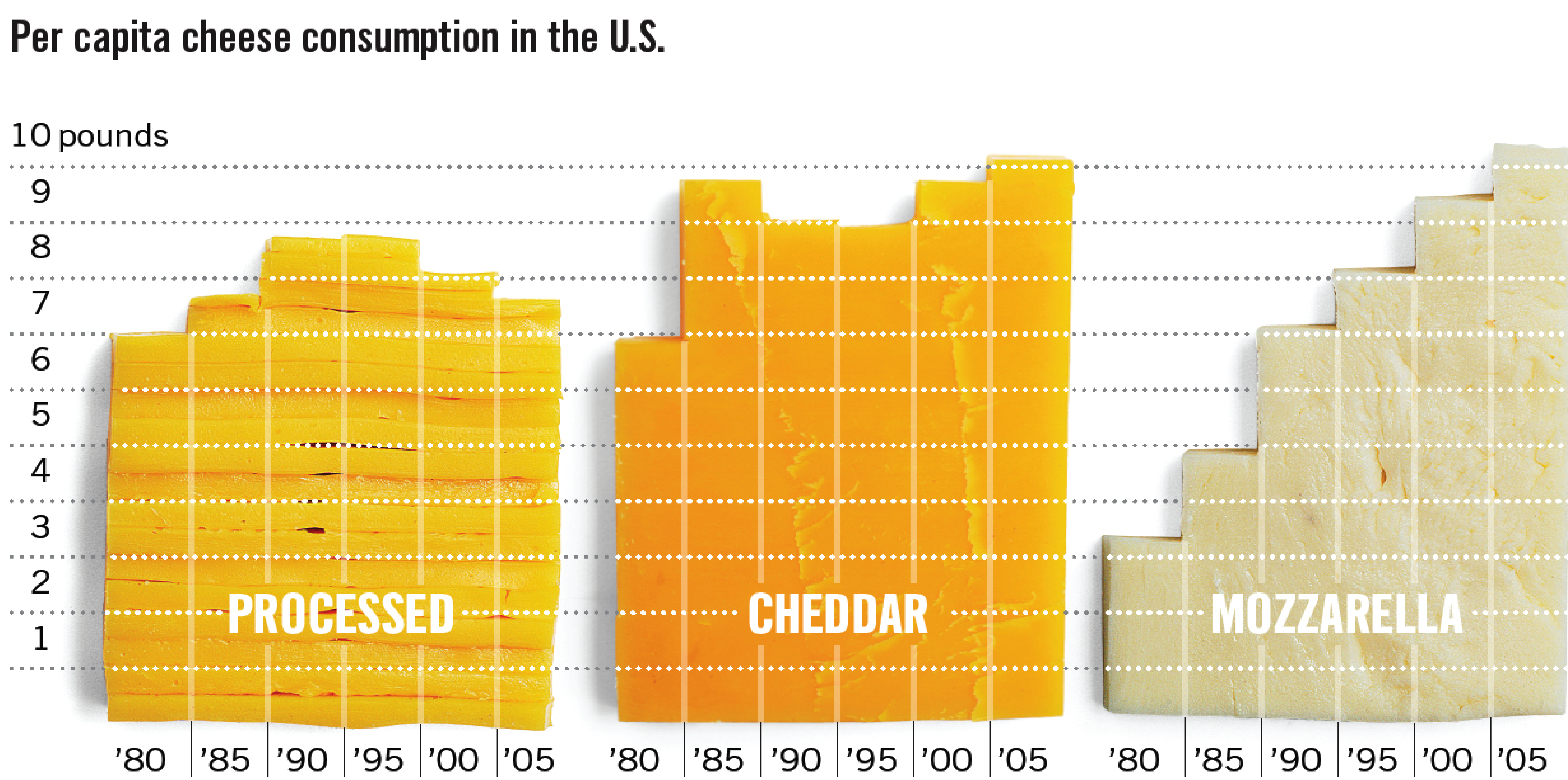
- eliminate the arbitrary: ''remove to improve''
- thoroughness
- develop a style
- decoration should be additive, not negative
- offer elegant & appealing presentation congruent with the subject
- do not pursue minimalism at all costs

Principle 3

Good data visualisation is **ELEGANT**

principles

- decoration should be additive, not negative

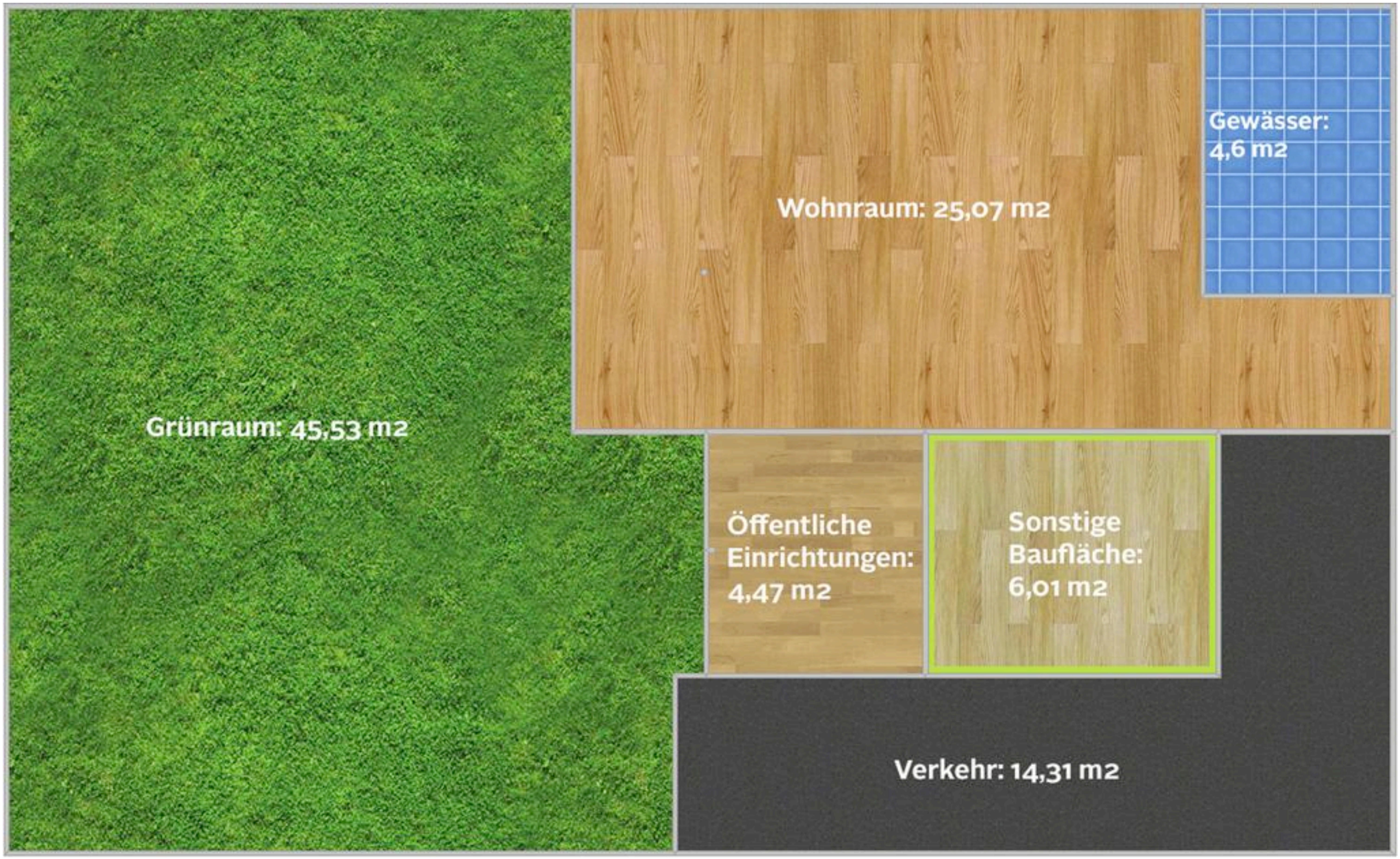


Principle 3

Good data visualisation is **ELEGANT**

principles

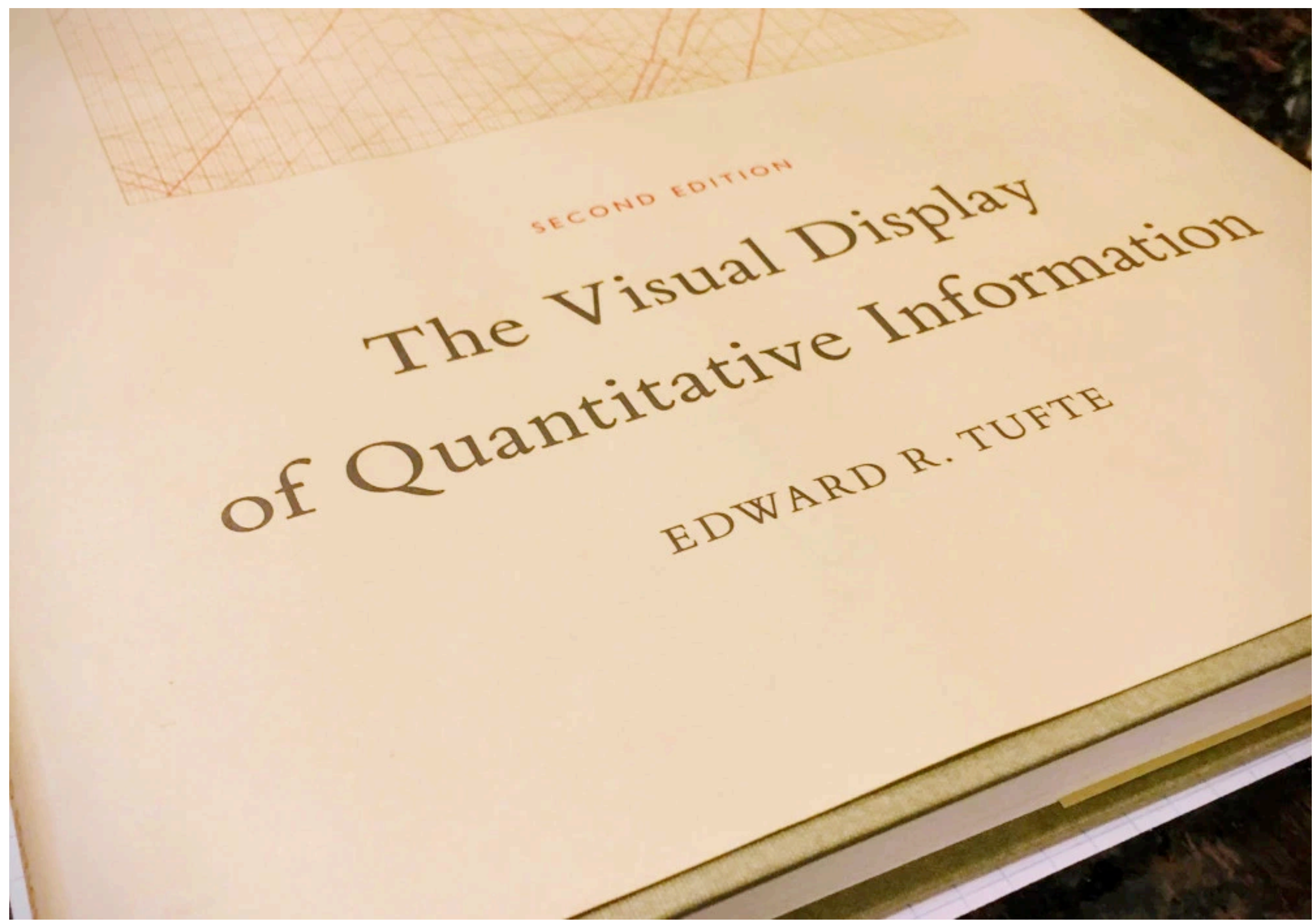
- decoration should be additive, not negative



“if vienna would be an apartment”

practice guidelines

by e. r. tufte



practice guidelines

by e. r. tufte

excellence in dataviz consists of complex ideas communicated with clarity, precision and efficiency; graphical displays should

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, technology of production
- avoid distorting what data have to say
- present many numbers in a small space
- make large dataset coherent
- encourage the comparison of different pieces of data
- reveal the data at several levels of detail
- serve a reasonably clear purpose: description, exploration, tabulation or decoration
- be closely integrated with the statistical and verbal description of the dataset

Edward Tufte