Factfullness

Inform but don’t persuade

Bias in data and data products

A common concern for everyone using data is bias.

Technically: systems that deviate from a design requirement.
Robust technical testing.

Humanely: systems that deviate from people's expectations.
Extensive social testing.

Example: 2020 A-level result estimation algorithm.
Example: automated face detection and tracking algorithms.

Most systems are designed to be biased (to make choices) the question to answer is are those choices the ones we want?

Who decides on what is the unbiased position?

Ethics and Data Science

What should we do in Data Science to be more ethical?

One person who thought about this in detail was Hans Rosling

Factfullness is his response to this question, as a set of ten recommendations for Data Scientists (and indeed everyone) about using and presenting facts.

The Gap

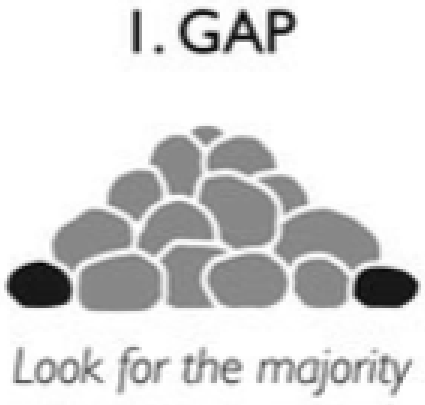
Where is the gap.

Beware of averages.

Look at spreads (distributions)

(even a Tukey boxplot is better than single numbers)

Beware extremes – rarely is world divided into a binary groups.



Negativity

Negative news is news.

Good news is rarely news

Gradual improvements are not news

Expect to mainly see bad news.

Are things bad but getting better?

More bad news may just be better records.

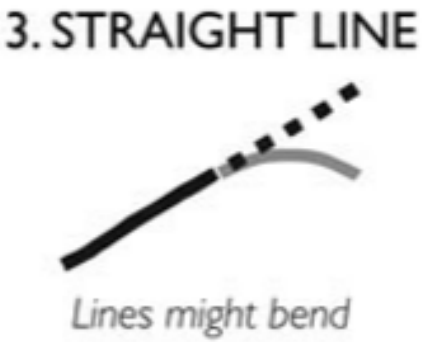
Be very wary of rosy pasts - rarely true.

Don’t assume straight lines

Straight lines are rare

Never assume that things will continue in straight lines

S-Bends are much more common



Manage fears

Understand real risks

Risk = danger * exposure

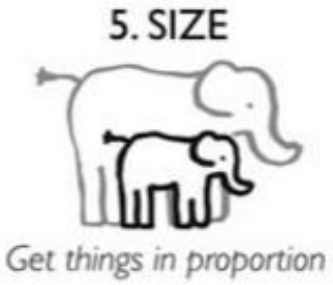
Size

Recognize impressive numbers and check them

Compare the context e.g. Globally

80/20 check what makes up 80% of the effect

Divide to get a rate especially when comparing different size groups



Generalization

Generalisations are useful, question categories relentlessly.

Look for difference within groups.

Look for similarities across groups.

Look for differences across groups.

Check how big the majority is.

Beware of vivid newsworthy examples.

Destiny

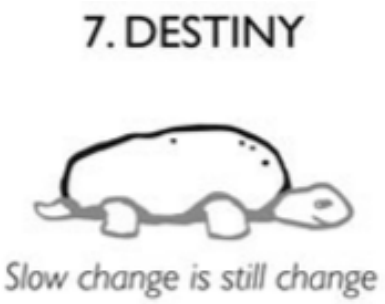
Things appear constant because change happens slowly.

Slow change is still change.

Track gradual improvements.

Update your knowledge

Collect examples of cultural change



Have more than one perspective

Test ideas

Limited expertise

Hammers and nails

Numbers but not only numbers

Beware simple ideas and simple solutions

Beware blame

Identify scapegoats, resist blaming an individual

Look for causes, not for villains

Look for systems not for heroes

Resist Urgency

Decisions are rarely urgent.

Take a breath, look for time to reflect.

It can be now and ever

Insist on data that is relevant & accurate

Beware predictions without uncertainty, look for options

Be wary of drastic action,
Kaizen!

Trust and Trustworthiness

Trust in data visualization

Trying to generate more trust may be counterproductive.

Baroness Onora O'Neill

Emeritus Professor of Philosophy
Cambridge University

But one aim might be to be trustworthy when we present data.
Transparency alone != Trustworthiness.

There are three properties that could help generate trust:

1) Accessibility – can we get the data we need.
2) Usability – is it in a form, with the tools needed, to be used.
3) Assessibility – can we test the data, reanalyse & compare it.