Data-based decision making

Jānis Lazovskis
L-EARN Bootcamp, 15 March 2021

No decision is based on just the data

Every decision must consider data and its context

Overview

- What is data?
  - Technically
  - Non-technically
- Examples
  - Big data
  - Social context
- Exploring data
  - The scientific method
  - Observation and randomness
- Checklist for decisions
  - Check before you wreck

What is data?

How it feels: A number, a symbol something that has no emotions, that is objectively true

Balance: Raw data is best, but it is huge and hard to understand. Extrapolated data is easy to understand, but far removed from the actual source.

Knowledge: You may only have to deal with particular types of data, but knowing all types gives you more perspective. Your decisions will be better reasoned.

Raw data

Measurements, observations, numbers

Processed data

Sums, averages, combinations

Extrapolated data

Predictions, estimates, new data

Theoretical data

Definitions, theorems, proofs

Raw data: *tables*

30000 x 10 table with numbers between 0 and 1 (value table)
30000 x 30000 table with with either 0 or 1 (neighbor table)

Find the 10 most important rows, based on values of neighbors.

Processed data: *dictionaries*

Four different vaccines, similar measurements, different trials, same goal.
How many people, what ages, what measurements, what doses, what date, ...

Decide which vaccines to distribute in what way to whom.

Theoretical data: *equivalences*

S = { -2, -1, 0, 1, 2, 3, ..., 11}
[s] = {t in S with t^2-s^2=5k for some integer k} for every s in S

Identify which sets $[s_1]$, $[s_2]$ are the same and which are different.

Examples

Everywhere: Data exists in every decision at every level: personal, governmental, educational, social, etc.

Perspective: If you can see the data, you can better understand the reasons for the decisions.

Theory: Expertise with theoretical data (mathematics) gives you powerful tools to describe and process data. Understanding the tools is key.

Exploring data

A day in the life: What does a data scientist do? How do they interpret and present data?

Not non-scientific: Exploring is part of learning, and learning is necessary for expertise.

Power and perspective: Be careful of your biases and how your experience shapes the results you see. Others may see different things in the same data - you must find common ground.

The scientific method

Pose a hypothesis, test the hypothesis against the data, conclude yes / no. Classically effective and watertight.

The non-scientific method

Look at the data in different ways, use your intuition, what result are you trying to find?

Randomness

- What is random? Dangerous to consider random data as non-random
- No single number is random. Randomness is defined on infinitely large groups of numbers
- Example: COVID vaccination data

➔ Is it understandable?

Find someone else with a different mindset and see if they can honestly understand your work.

➔ Is it consistent?

If others with the same data made different conclusions, you need to justify why you are correct and where they made mistakes.

➔ Is it sensational?

Sensational results are often based on selective / small data. Impressive results = impressive justification.

➔ Is it important?

Why should anyone be interested in your decision? Be aware of your public and context.

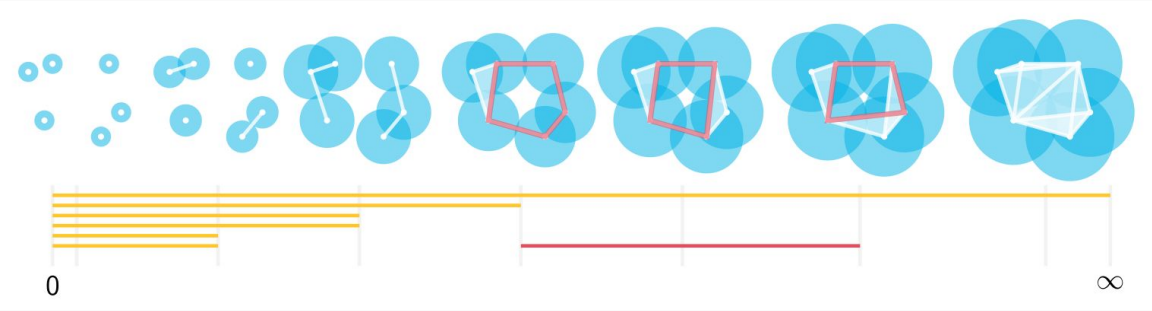
Checklist for a good data-based decision

This is not a universal list. To succeed, personally recommended:

- A critical and honest mindset
- A toolbox for analysis

Thank you for your attention

1. What is data + Examples
2. Exploring data + Checklist



Resources:

- FiveThirtyEight fivethirtyeight.com
- The Economist Graphic Detail blog economist.com/graphic-detail
- Bart de Langhe, Harvard Business Review “Covid-19 Vaccine trials are a case study on the challenges of data literacy”
- James Gleick “The Information: A History, A Theory, A Flood”
- RAND corporation “A million random digits with 100 000 normal deviates”