

Jeroen van de Nieuwenhof

data journalism & open source intelligence (OSINT)

✉ jeroen@vandenieuwenhof.nl

🐦 @KenSentMe



Paul Bradshaw

@paulbradshaw

...



Data journalists! How do you explain your job to others?

(I tend to say "Just like a normal journalist, but using data too - and the stories might include visualisation or interactivity") #ddj



Cecília do Lago ✓ @ceciliadolago · 2 u

Als antwoord op @paulbradshaw

I do journalism with number. Its easy explain that way because in my country journalists hate dealing with numbers



1



3



Winnie Kamau @WinnieKamau254 · 4 u

Als antwoord op @paulbradshaw

Hmm.... I explain it as using facts and figures to tell a story and give examples... The example part is the eye opener for many 🤔



1



Alex Homer ✓ @alexhomer · 55 m

Als antwoord op @paulbradshaw

I'd say first that I'm an investigative journalist interested in public interest reporting and secondly that a lot of my work involves working with data.



Adrian Bridgwater ✓ @ABridgwater · 4 u

Als antwoord op @paulbradshaw

I just say, well, you know the Daily Mail? Yeah? Well, not like that.



2



DATA (DRIVEN) JOURNALISM

DATA (DRIVEN) JOURNALISM

Journalism is the activity of gathering, assessing, creating, and presenting news and information.

*Purpose: to provide people with **verified information** they can use to make better decisions*

*A systematic process – a discipline of **verification** – that journalists use to find not just the facts, but also the **“truth about the facts.”***

DATA (DRIVEN) JOURNALISM

a form of investigative journalism: research, analyse, present

based on large amounts of data

DATA AS A SOURCE

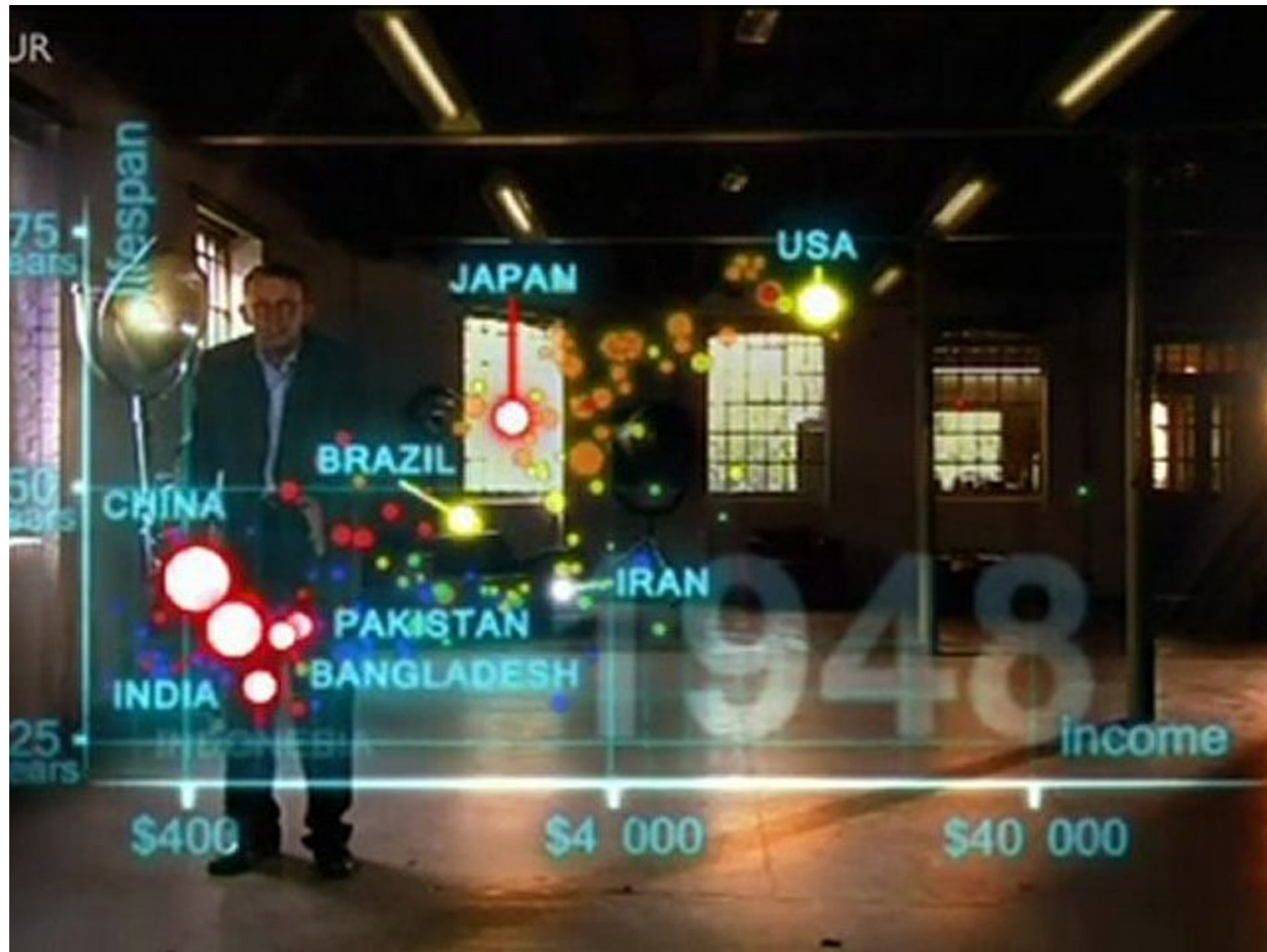
telling stories that originate from data



*the data alone doesn't suffice
use it as a **starting point** for a
conversation/investigation*

DATA AS A TOOL

use data to bring the story you want to tell



MIND THE BIAS

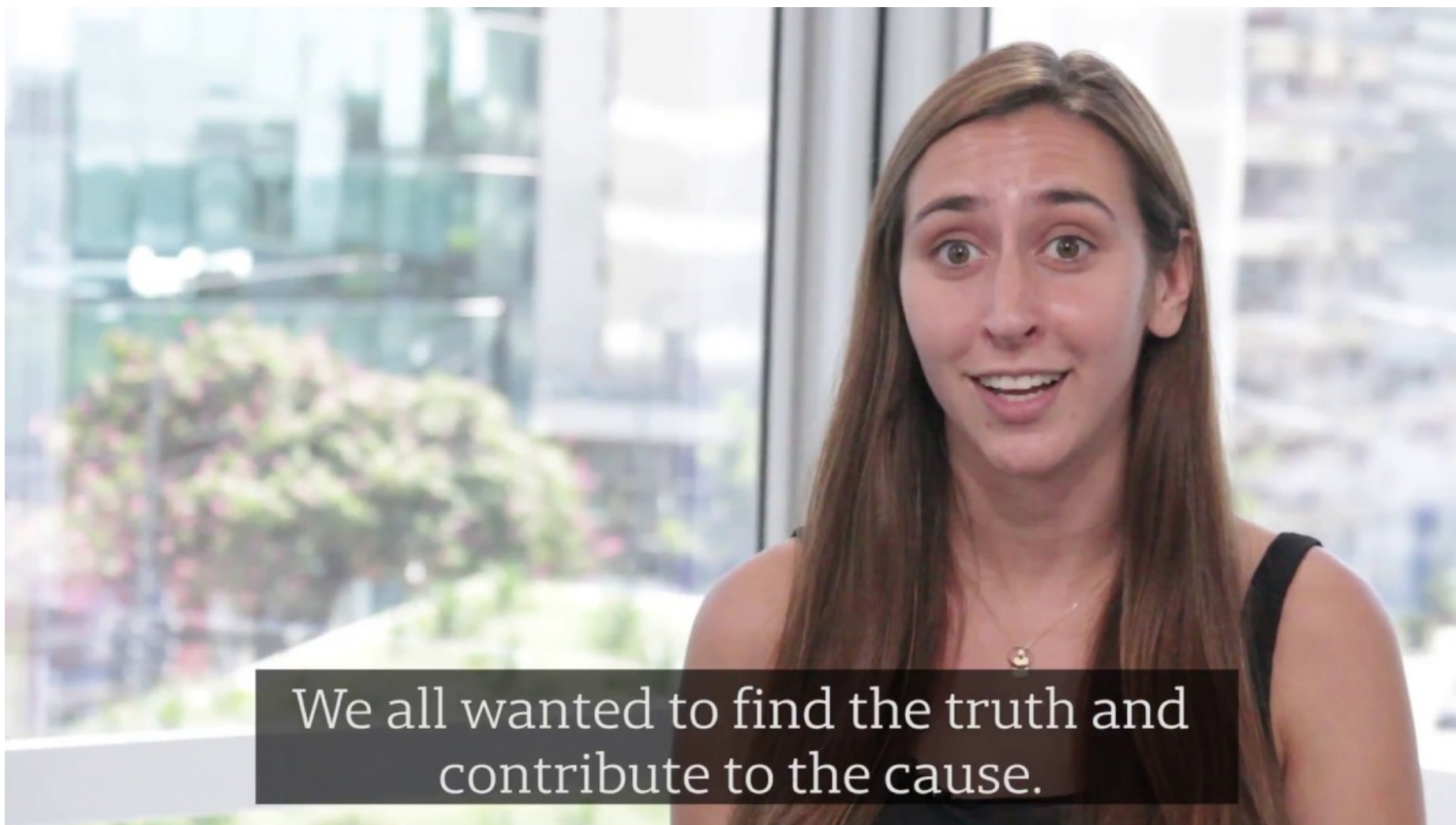
we tend to pick statistics that confirm our worldview

MIND THE BIAS

whoever creates the database, determines how we view the world

WHAT IS DATA?

*data is not only about numbers and statistics
data can be text, photos, audio, etc.*



We all wanted to find the truth and
contribute to the cause.

5/6 STEPS TO BRING STORIES WITH DATA

(1. RESEARCH QUESTION OR HYPOTHESIS)

*what about **topic X**... ?*

*i suspect **A** does **B** to **C** through **D** (with help from **E**) because of **F** ...*

2. FINDING DATA

words in text - Trump vs Biden speeches,
compare words used (word cloud)

ready to go - search for 'database'
or 'data set'. E.g. IMDB, Last.fm

statistics – (almost) every country has a statistics
bureau (like CBS or US Census Bureau)

COUNTING



2. FINDING DATA

open data - national (data.overheid.nl) or local government (data.eindhoven.nl) etc.

monitor tools - companies that collect and sell data. for sports/entertainment: Gracenote

scraping - run a script/program that automatically collects data from a website. E.g. dresses from Zalando or houses on Funda.nl

wob/foia – for the die hards. Request government data that isn't already publicly available.

ALWAYS QUESTION

who provides the data?

why are they sharing it?

how was it collected?

what's *missing*?

3. CLEANING THE DATA

*learn **spreadsheets** (like Excel), **Numpy/Pandas** (python libraries), **R**, etc.*

3. CLEANING THE DATA

in order to:

remove clutter

check if it's in the right format

are there empty fields (and why)?

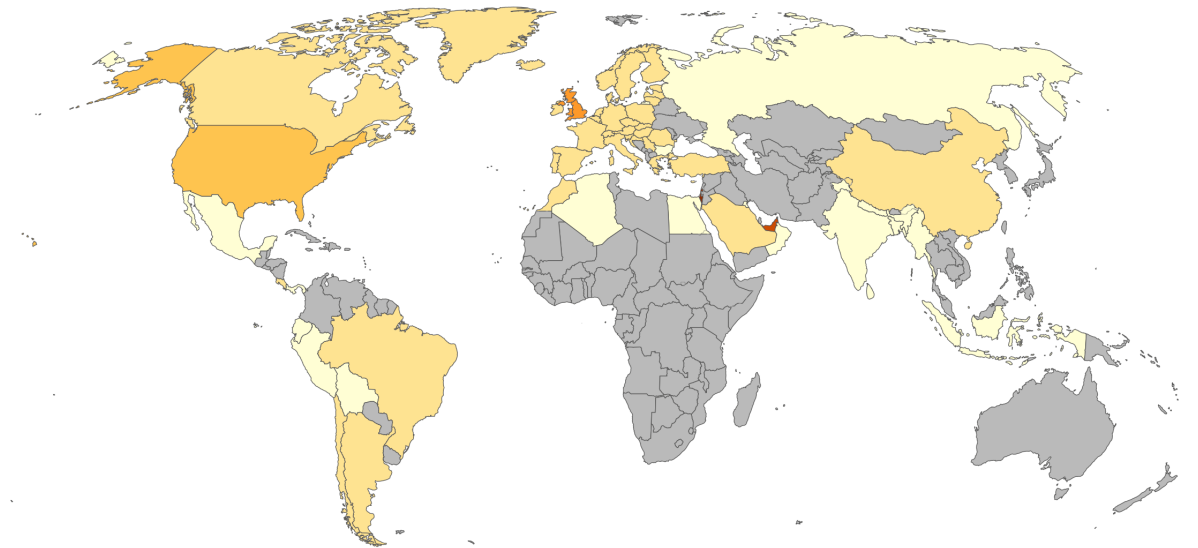
is it complete?

4. ANALYSING THE DATA

*look for **patterns** – averages or geographical*

Where is the vaccine being given?

Total reported vaccine doses administered per 100 people



Note: Total vaccinations refers to the number of doses given, not necessarily the number of people vaccinated

Source: OWID, gov.uk dashboard, ONS, updated 1130 GMT on 11 Feb



4. ANALYSING THE DATA

*look for **extremes** – best, highest, lowest, most frequently*

4. ANALYSING THE DATA

compare – then/now, here/there, groups

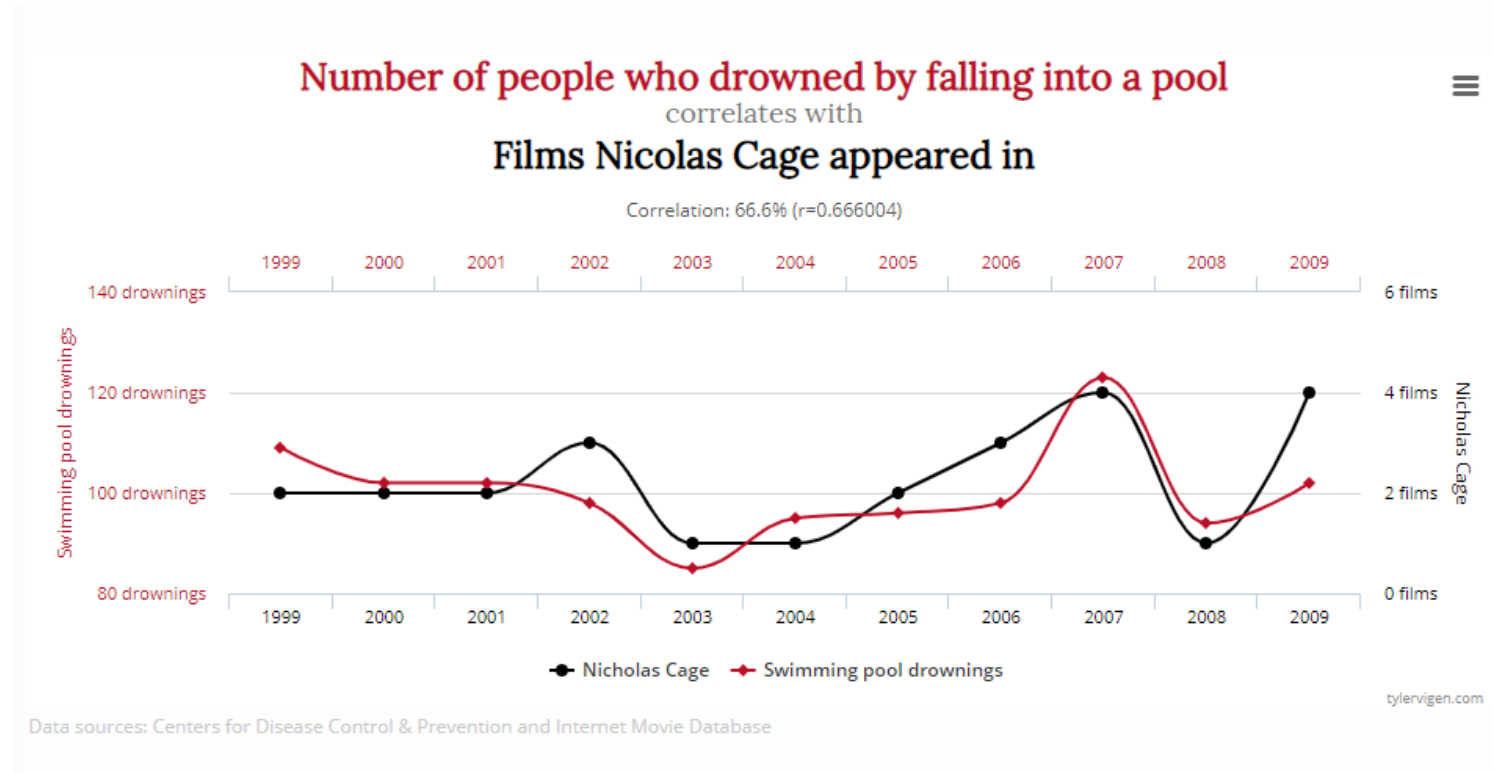
4. ANALYSING THE DATA

*look for **correlations***

does one thing lead to another?

which factors are at play?

CORRELATION \neq CAUSATION



5. DATA EXPLAINED

starting point of your story:

numbers often don't answer the 'why'-question

so start researching

6. BRING YOUR STORY

QUESTIONS?

