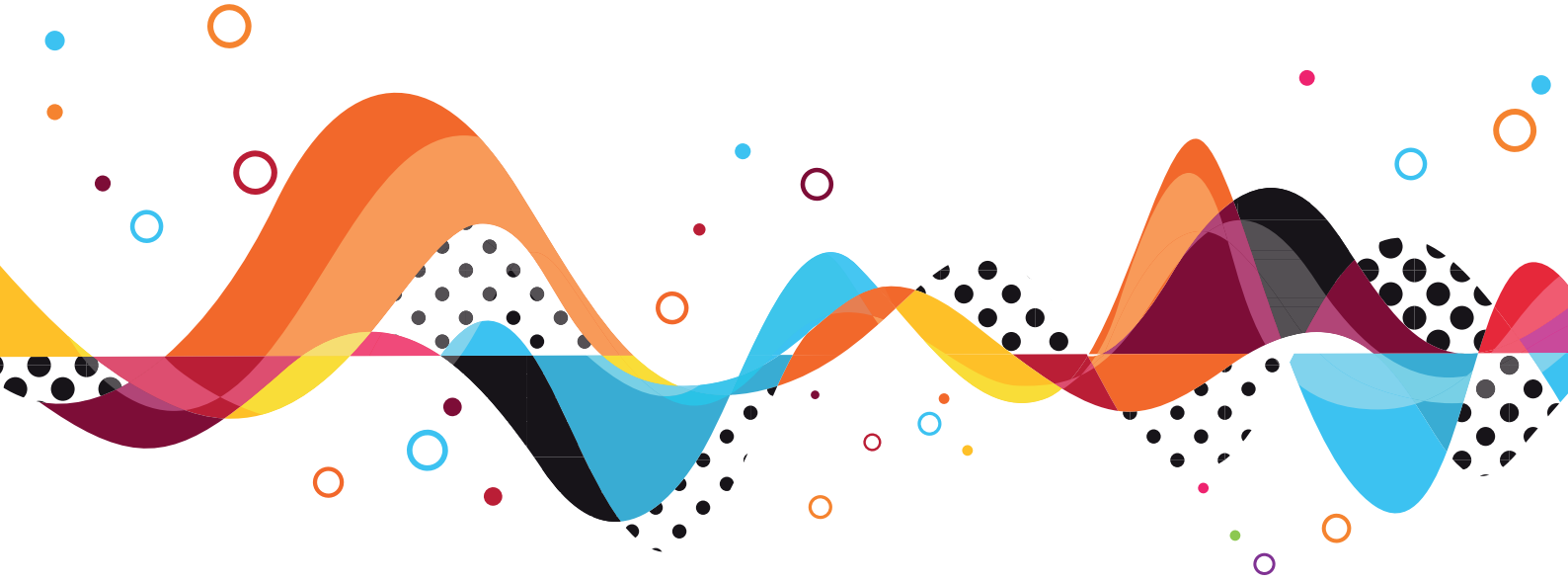
ART/DATA/HEALTH
data as creative material
for health and wellbeing

University of Brighton

DATA SKILLS TOOLKIT:

LEARNING TO WORK WITH DATA



An ART/DATA/HEALTH resource

UKRI Arts and Humanities Research Council

Published by *ART/DATA/HEALTH: data as creative material for health and wellbeing*, University of Brighton, Brighton, UK. July 2020.

A full-text PDF of this resource is available from: **https://artdatahealth.org/**

Suggested citation: Aristea Fotopoulou, Harriett Barratt and Elodie Marandet.2020. *Learning to Work with Data: A Data Skills Toolkit*. Brighton, UK: ART/DATA/HEALTH: data as creative material for health and wellbeing.

CONTENTS

1

ABOUT THIS TOOLKIT

The ART/DATA/HEALTH Data Skills Toolkit has been developed by researchers at the University of Brighton. The project 'ART/DATA/HEALTH: data as creative material for health and wellbeing' is funded by the Arts and Humanities Research Council (UKRI-AHRC Innovation Fellowship) and led by Dr Aristea Fotopoulou at the University of Brighton.

This Data Skills Toolkit is a free resource available for anyone interested in helping others to understand health and wellbeing data. We hope that third-sector and charity organisations will find it useful for telling stories with health and wellbeing data, and that they will make use of the toolkit in their advocacy work.

The resources included in this Toolkit are the result of a series of workshops that were carried out in Brighton, UK, in early 2020. The ART/DATA/HEALTH project worked with two groups of participants: one made up of employees at the domestic abuse charity RISE (**https://www.riseuk.org.uk/**), and another group which brought together people from a range of organisations in the health, wellbeing and therapeutic sectors. Each group of participants met for two types of workshop: **datahubs**, in which they learnt about how to work with data and create stories with them, and **creative art workshops**, led by the commissioned artist (see Section 2: About the ART/DATA/HEALTH project).

HOW TO USE THIS TOOLKIT

This Toolkit provides you with a step-by-step guide to the training provided in ART/DATA/HEALTH datahub sessions. Each section is designed to help you with a specific aspect of working with data: an introduction to data for advocacy, a guide to working with data, an overview of visualising data for storytelling, and a guide to telling data stories. You can do this training alone or as a group with your colleagues. You will find a range of information, exercises and further resources in this Toolkit. The exercises are collated at the end of Toolkit as handouts that you can print out, photocopy and share, as well as a list of resources, links and further reading. You will also find several examples of data arts projects spread throughout this booklet, to show how broad and engaging this practice can be.

UKRI Arts and Humanities Research Council

The Arts and Humanities investigate the values and beliefs which underpin both who we are as individuals and how we undertake our responsibilities to our society and to humanity globally.

2

ABOUT THE ART/DATA/ HEALTH PROJECT

The academic research project 'ART/DATA/ HEALTH: data as creative material for health and wellbeing' creates an innovative and interdisciplinary process that offers new tools, at the intersections of data science with art practice, to approach two key issues: digital skills and health literacy.

The project uses health and wellbeing data as the source of experiential stories and as the source material for creative expression. In a series of exploratory workshops, a community of artists, service workers, academics and people living in Brighton and Hove used a combination of creative media, storytelling and data analytics to explore evidence around health and wellbeing.

They are co-producing creative work that takes various forms, using both anonymised personal and open statistical health and wellbeing data. The project was adapted mid-way through to reflect the extraordinary circumstances of the Corona virus pandemic.

The artists involved in the project are: Anna Dumitriu, working with staff from the domestic abuse charity RISE on the rise of domestic abuse during C19; Kate Genevievve; Hydrocracker; and Caroline Beavon, focusing specifically on C19 data.

Key partners are:

Brighton & Hove City Council

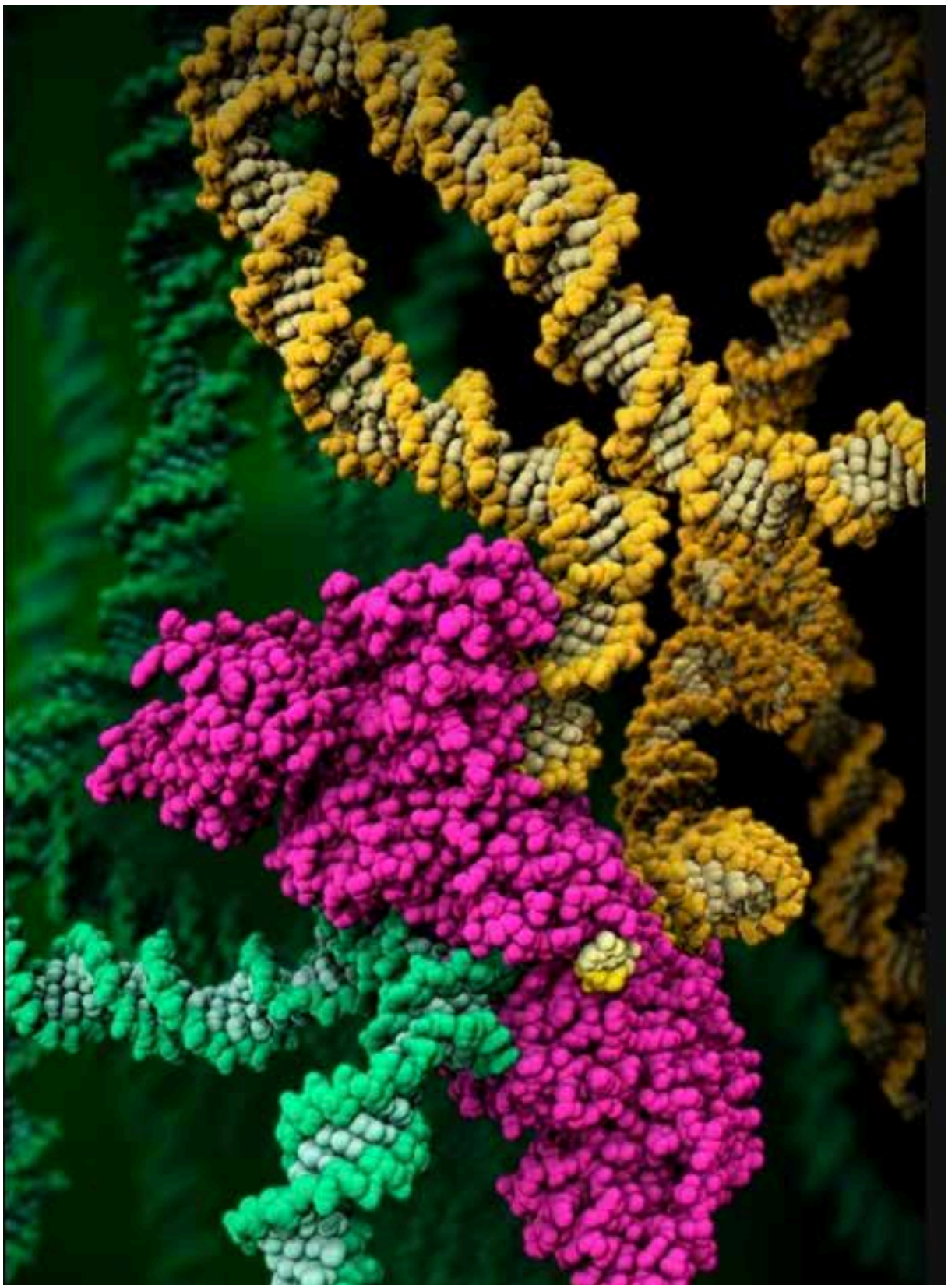
UKRI Arts and Humanities Research Council

BRIGHTON DIGITALFESTIVAL

NHS Brighton and Hove Clinical Commissioning Group

BRIGHTON FRINGE

Freedom from Abuse and Violence Rise



3

HEALTH AND WELLBEING DATA

TYPES OF HEALTH AND WELLBEING DATA

Health and wellbeing data take lots of different forms, including:

- **Patient data:** information on someone's own health and wellbeing collected by a health body such as a GP surgery or hospital. This might include medication records, or blood pressure readings, for example. Some health insurers also collect this information.
- **Individually collected data,** for example, through wearable devices and mobile phone apps that track heart rate or exercise patterns.
- **Data collected by third parties,** for example by a third sector or community group, by employers, and by companies.

WHY WORK WITH HEALTH AND WELLBEING DATA?

Working with health and wellbeing data can give individuals, researchers, and organisations important insights. For example, data might allow governments to identify links between health and poverty, or help individuals to notice what they could change to live more healthily.

For third sector organisations, data are the materials that can fuel strategic planning and advocacy work. They not only highlight problems, but also act as the evidence to underpin clear, engaging stories.

THINGS TO BE AWARE OF WHEN WORKING WITH HEALTH AND WELLBEING DATA

Anyone working with data needs to understand how to treat them so that they can be used and shared safely and ethically. There are several ways you can anonymise data to protect the individuals concerned:

- **Anonymised data** are data from which it is not possible to identify individuals either directly or indirectly. Anonymised data can be **aggregated**. This means that statistical data about several individuals has been combined to show general trends or values, without identifying individuals.
- **Pseudonymisation** means using a unique identifier to distinguish individuals within a data set. This means that their 'real world' – most commonly their name – is not evident, but that their records can still be tracked and linked together.
- Data can also be **de-personalised**. This means that the information does not identify an individual, because identifiers or identifiable data have been scrambled or removed from that single record. However, the information is still about an individual person, so it needs to be protected.

Some individuals have special protection under law. **Special category personal data** are data defined under the Data Protection Act (1998) as data that identify a living individual regarding his or her:

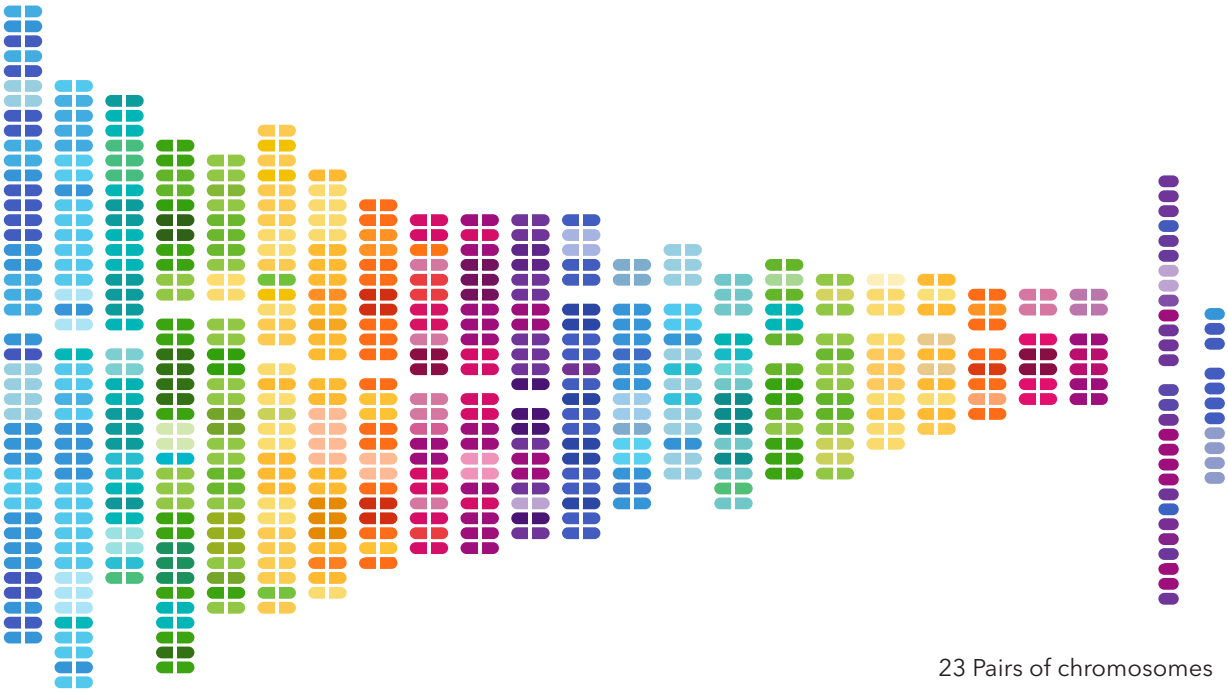
- racial or ethnic origin,
- political opinions,
- religious beliefs or other beliefs of a similar nature,
- membership of a trade union,
- physical or mental health or condition,
- sexual life,
- generic data and biometric data convictions,
- legal proceedings against the individual or allegations of offences committed by the individual.

HOW TO WORK RESPONSIBLY WITH HEALTH AND WELLBEING DATA

When working with data, it's vital to think about what problem is being explored, who the stakeholders are, and who holds the power to decide what is collected and how it is used.

The ART/DATA/HEALTH project puts emphasis on working responsibly with partners and data, and communicating clearly and openly. Organizations can work with health and wellbeing data in ways that improve public trust and understanding. Some key ways in to do this are to:

- Engage community organisations in research design and data analysis
- Design your consent processes with the user at the center: make sure they are properly informed about how their health and wellbeing data may be used
- Focus on complex social issues, particularly the factors which contribute to unequal access to healthcare – some of which are hugely stigmatised
- Explore how individual stories can bring shared experiences and themes to light.
- Develop skills in data analysis and security.
- Be conscious about implicit bias or structural inequalities that your work may be perpetuating, particularly in relation to protected characteristics.



23 Pairs of chromosomes

PATIENT DATA

Patient data are data collected about a patient whenever they go to a doctor or receive social care. This can include details about the patient's health, medication record, health and social care needs, and other information. These data are recorded and stored in a care record.

Patient data use has allowed researchers and health professionals to provide **better care and improve health,** by:Improving individual care, for example by

- improving newborn screening;
- Making more accurate diagnoses, for instance by researching the earliest changes in the development of dementia;
- Enhancing treatment and prevention, e.g. by learning more about blood cancer;
- Increasing patient safety, for example by demonstrating the safety of a treatment for heart failure;
- Improving the planning of NHS services, e.g. by analysing the local variations in stroke prevention;
- Evaluating policy e.g. extending the flu vaccination programme to children; and
- Better understanding diseases, for instance by following up people discharged from hospital after a heart attack. (Understanding Patient Data, 2018)

Digital data approaches and tools have changed many **treatment practices**. For example, NHS.uk's e-prescribing system is widely used, as is their online advice repository, which had 550 million visits in 2016 (NHS, 2017). In some countries and areas of practice, the use of digital tools is now mandatory: for example, the US provided financial incentives for pharmacies switching to e-prescribing for chemotherapy drugs in the late 2000s, given how severe the impact of human error in dosage measurements would be (Zadeh and Tremblay, 2016).

In the UK, The Department for Health and Social Care (2018) is trying to make the use of health and wellbeing data more secure, open and inclusive. For example, the NHS Long Term Plan (2019) lays out an ambition for developers to create new software to improve how local health and care records (LHCRs) are recorded. Five LHCR regions, together covering 40% of the population of England, have been awarded £7.5 million each to carry out this work (Castle-Clarke and Hutchings, 2019).

Many health sector research funders (such as the Wellcome Trust, the Medical Research Council, the Biotechnology and Biological Sciences Research Council, and Cancer Research UK) either encourage or require their funded researchers to share data (UK Data Services, n.d.).

Meanwhile, technical policy initiatives such as FHIR standards (Fast Healthcare Interoperability Resources, an international standard for the exchange of electronic health records) have emerged to enable and promote data sharing.

However, there are some big challenges for both individuals and organisations around how patient data is collected, used and shared. For many patients, privacy is a huge concern. Organisations, meanwhile, need to make sure that they are fully adhering to legal, technical and ethical standards across the whole process of working with health and wellbeing data, from collection to use to storage.

There are **strict controls around how and why companies can access and use patient data**. National bodies play a key role in raising awareness of these controls. However, a 2019 report from the Nuffield Trust found that there was often a lack of **national-level strategic thinking and leadership** in regard to digital standards and data sharing (Castle-Clarke and Hutchings, 2019). As a result, there are examples across the UK of providers refusing to engage with local data sharing efforts (ibid.).

A 2019 study found that the majority of studies into data privacy reported that people are **generally worried about the security of data** and fear 'data leakage' (for example, through hacking or by companies selling data) (Skovgaard, Wadmann, & Hoeyer, 2019). The sharing of **sensitive data** (including information on mental health, sexual health, sexual preference, and religion) holds greater concern than other types of data, as does the use of data by commercial companies (Understanding Patient Data, 2018). More transparent information and communication, and space for dialogue around data use and data sharing, are key to securing users' trust.

There are also **demographic differences** in relation to attitude to data sharing. Young people are more supportive of electronic patient records and more likely to think of the benefits of patient data sharing (Understanding Patient Data, 2018), while ethnic minorities may be less confident about their data being secure if used for health research (Hunn, 2017). There may also be a risk of the emergence of a two-tier model of access, under which only those with financial means are able to afford to protect their own data (Sharon and Lucivero, 2019).

The UK's controversial **care.data NHS project** encapsulates many of these issues. Announced in 2013, the programme sought to extract data from GP surgeries into a central database. It was positioned as a new chance to make real inroads into harnessing the potential benefits of sharing health data, from longitudinal research to the monitoring of public health populations. However, fears around risk management were publicly voiced by ethicists and clinicians from the start (Sheather and Brannan, 2013). These concerns ranged from the lack of clarity around what patients were told about consent and potential risks (McCartney, 2014), to the deliverability of the programme's scope, budget and schedule, to the lack of sufficient ethical protocols and safeguarding (Mann, 2014). The programme was postponed, suspended and restarted several times in different locations between 2014 and 2016, when it was eventually abandoned. The care.data project has been described as an example of the failure to consider the need for a **'social licence'** (Carter et al, 2015). In the end, it failed to secure public trust, created a rupture in traditional understandings of the role of the general practitioner, and failed to articulate itself as a 'public good'.

DATA COLLECTED BY INDIVIDUALS, USING MOBILE PHONE APPS, WEARABLES

Individuals collect data by using mobile phone apps and wearable devices such as Fitbit. The number of connected wearable devices worldwide has more than doubled over the last three years, increasing from 325 million in 2016 to 722 million in 2019, and is forecast to reach more than one billion by 2022 (Statista, 2020). Using wearables to collect health and wellbeing data can help individuals to understand and make changes to their behaviour patterns, from diet and exercise to sleep. This is why self-tracking using wearables and other digital technologies is often understood to empower people.

However, there are many privacy and security issues associated with self-tracking devices, as well as questions such as: who retains rights of the data, how do companies use them, where do they end up, and who profits from selling these data?

Wearables and apps collect a broad range of data, including biological processes such as heart rate and sleep patterns, lifestyle factors such as diet, distance and speed travelled, behavioural and postural data such as walking gait (to prevent falls), and much more besides. They are collected through a combination of automated data collection (for example, a pedometer built into a phone or a wristband) and self-generated data (such as those data provided to health apps on mood, symptoms and physical condition). The format of the data is primarily quantitative, allowing easy exporting and comparison, though some apps also use qualitative input such as free text. Many offer customisable features – for example, the 'Clue' menstruation and fertility tracking app allows users to create 'non-standard' symptoms to be tracked on top of those provided (see Fotopoulou 2017). Although many commercially available wearables seek to position themselves as fashion items, some seek to be as unobtrusive as possible. For example, the 'Hospital Without Walls' project in Australia piloted a portable radio which included a discreet accelerometer to detect falls amongst an elderly care population (Wu and Luo, 2019). Some biosensors, act at the level of cells, enzymes and biological material. For example, blood glucose monitors used at home by diabetics account for 85% of the global market (Mehrotra, 2016).

This has also led to patient-led, 'DIY' movements. The 'We Are Not Waiting' campaign (**https://openaps.org/**) is a group of patients with diabetes who have developed their own glucose monitoring system more quickly than commercial companies have been able to. The explosion in the global use of wearables has led to a new sense of how we see ourselves, known as the 'quantified self'. This term refers both to the phenomenon of self-tracking, and to a community of users – known as Quantified Self or QS – who swap tips and resources.

In this way, self-tracking might embed a sense of an 'ideal' self – a fit, healthy and busy self whose every move is logged. This ideal is a normative one (Fotopoulou and O'Riordan, 2016) – it does not allow for widespread differences in bodies, health conditions or lifestyles. Self-tracking also shifts responsibility onto the individual for their own care. In this way, it can be seen as a symptom of a more general phenomenon: a shift towards healthcare systems where the state plays a smaller part than the commercial sector.

As well as apps and wearables which upload personal data to centralised databases, users are increasingly encouraged to share their data with other users, organisations, and reasearchers. The drive towards 'data-sharing' – embodied by phrases such as 'sharing is caring' – even takes on moral undertones (Fotopoulou, 2018).

Some employers are also now offering wearables as part of their employee benefits scheme. However, questions around the use of data generated by wearable devices are complex in a workplace setting. Research carried out by PwC in the UK in 2015 found that, while two thirds of workers (65%) wanted their employer to take an active role in their health and wellbeing, only 46% would accept a free piece of wearable technology if their employers had access to the data recorded (PwC, 2016). Four in 10 said that they didn't fully trust their employer to use it for their benefit, while 37% didn't trust their employer not to use the data against them in some way. Workers who would be happy to use a wearable device at work are most likely to want to trade their personal data in exchange for flexible working hours, free health screening and health and fitness incentives.

While users are offered a summary of their data, it is often difficult to find and extract the data sets themselves (most data standards recommend that data are exportable in a csv format to allow easy analysis – see the **Five-star deployment scheme** for open data licensing, built by Tim Berners-Lee, and the Open Data Institute's **Data Certificates**). In conjunction with the difficulty of user access, many apps retain the rights to sell on the data they collect in an anonymised form. It may also often be the case that the full data sets are only legible to health professionals – for example, devices which measure the haemoglobin concentration in the blood, or the severity of a hand tremor, produce data which are not easily read by patients, and need to be interpreted.

Self-tracking has also led to concerns about over-diagnosis (Li et al, 2017): is it always the best thing to have access to lots of quantitative data? How are they 'read' or prioritised alongside hands-on examinations, or people's everyday experiences of their own health?

DATA COLLECTED BY THIRD PARTIES

Many kinds of health and wellbeing data are collected by organisations whose main activity lies outside of traditional healthcare. This includes charities, local government, employers, community groups, commercial bodies and many more.

The research fieldwork of the ART/DATA/HEALTH project involved working with community groups that collected health and wellbeing data of both qualitative and quantitative nature. These related to experiences of domestic abuse; data which tracked someone's journey through a support service, including rates of re-engagement; details of volunteers supporting a 'befriending' service with isolated people, and many more. Each set of data is used by these organisations in specific ways – for example, for communications, lobbying, and bid-writing purposes, as well as to inform patients' treatment options. Some of these data are entered into a national, anonymous database, while others are kept locally by a small number of staff members.

New technological practices – such as the use of data storage in the 'cloud' – mean that the health and wellbeing sectors start to adapt to a **decentralised, interconnected structure**. This gives an opportunity for decision-making to be spread across a wide range of individuals, researchers and organisations, and for health research to more fully take into account a wider range of people, from participants to third sector organisations. But enhanced connectivity also means that there is a need for clear, shared approaches to ethics, consent and governance of health and wellbeing data, which inevitably differ from those that apply to commercial data (Bot, Wilbanks, and Mangravite, 2019).

There is a **global market in healthcare data** beyond the local community level. Third parties gather de-identified data from healthcare systems, pharmacies and other sources, and sell it on to buyers interested in analysing large data sets. These are known as **'data brokers' or 'information brokers'**; their customers include government bodies, marketing agencies, and research organisations. For example, the US health information technology firm Iqvia (formerly IMS Health), one of the biggest companies in the health data market, generated $8 billion in revenue in 2017 (Arndt, 2018).

These companies' access to health and wellbeing data differs from country to country. In the UK, there are strict controls around how and why companies can access and use national patient data, as outlined above (Understanding Patient Data, 2018). However, models of data ownership and sharing differ hugely across the private companies offering internet solutions, wearables and app technologies. For example, a 2018 study (Binns et al, 2018) found that the majority of 959,000 apps from the US and UK Google Play stores transferred data to third parties, and that many of these operated on a transnational basis (and therefore not necessarily in adherence to the legal system of the country of use).

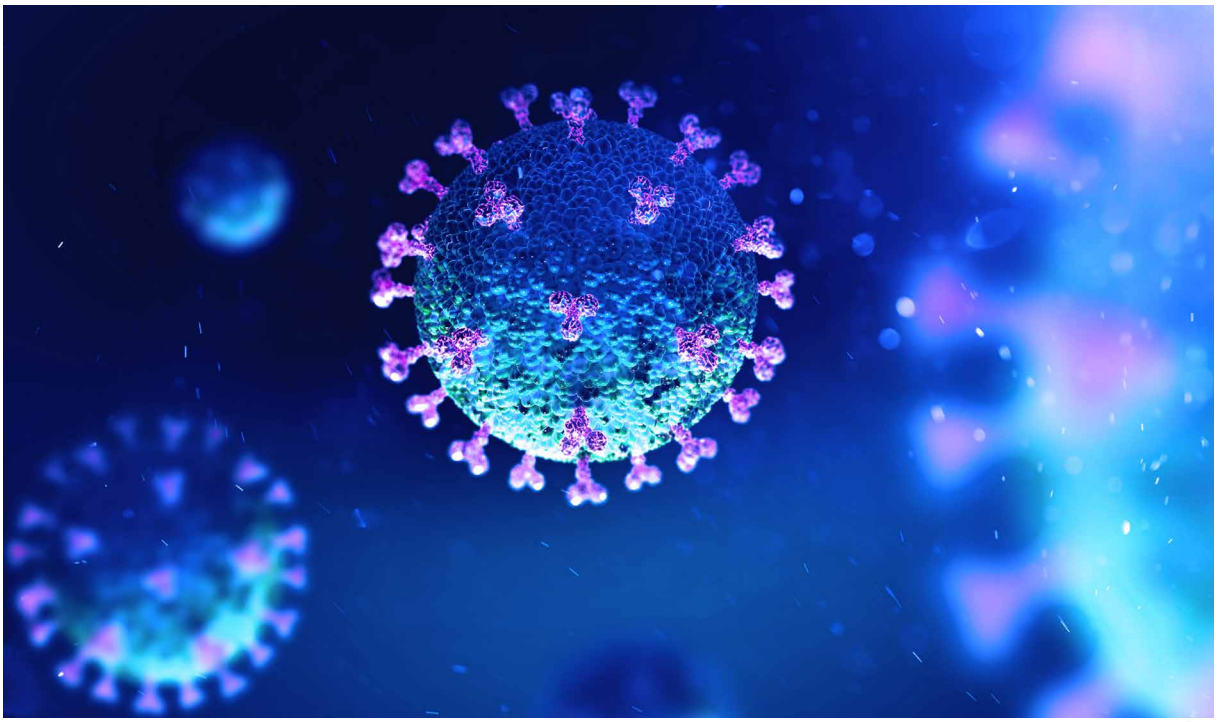
In the EU, the General Data Protection Regulation (GDPR) legislation established in 2018 seeks to answer some of these concerns. However, while consent forms may be technically compliant with GDPR, this may not mean that they adhere to high standards of **informed consent**. When dealing with health and wellbeing data, it is important to place the user at the centre of your consent processes and to make sure that they know exactly how their information will be used (Schairer, Rubanovich, and Bloss, 2018).

HEALTH AND WELLBEING DATA DURING THE COVID-19 PANDEMIC

The COVID-19 pandemic has turned public attention to the crucial role played by data in enabling us to understand rapid developments worldwide and to take appropriate measures. However, the crisis has also highlighted that these data do not speak for themselves; their collection, use and presentation to the public are complicated.

Take for example the number of deaths from COVID-19. Controversies about these numbers have shown that counting the victims is more complex than expected. In the UK, the daily announcements in the first weeks of the pandemic only included deaths in hospital of those who tested positive for COVID-19. Even then, those who had recently passed away may not be counted as there was generally a delay of a few days in hospital reports, potentially leading to considerable discrepancies. For example, while on 27 March 2020 the government announced that 926 COVID-19 deaths had so far taken place in English hospitals, NHS England now reports that the true figure was 1,649 (Richardson and Spiegelhalter, 2020).

A more reliable number is the one collated by the Office for National Statistics from death certificates issued by local authorities. This would include deaths occurring in the community and care homes. However, deaths can be reported for up to thirteen days after a person passed away, so numbers are likely to lag even further behind. In addition, in the absence of systematic testing, it is likely that some deaths caused by COVID-19 may not be registered as such, leading to an underestimation of mortality rates.
A more accurate number still – but suffering even further delays – is arrived at by looking at excess of deaths i.e. the number of extra deaths recorded during a certain period (compared to a similar period in previous years). This excess can therefore be attributed to either COVID-19 or to the new context. But new issues crop up: how to ascertain how many deaths are down to new policies such as lockdowns, and their effects (e.g. reduced medical cover, increase in domestic violence, poverty, mental health issues)? Or how many lives have been saved by these changes (e.g. reduced travel, increased air quality)? What is more, it is likely that some of those who lost their lives to COVID-19 would have passed away later that year, effectively bringing their loss earlier rather than adding to the overall numbers. However, this figure may provide a better approximation of the truth. For example, on 22 April 2020, the Financial Times published extrapolations showing that the likely number of "excess deaths" since the start of the pandemic in the UK could be in the region of 41,000, rather than the official 17,337 fatalities officially recorded (Giles, 2020).



Without mass testing, and given the wide range and variable severity of the symptoms, ascertaining how many people have had COVID-19 is also far from simple and is best done in retrospect. For example, on 21 April 2020, a study by researchers at Hong Kong University's School of Public Health published in The Lancet found that more than 232,000 people might have been infected in the first wave of COVID-19 in mainland China (Tsang et al, 2020). In an article in March 2020, Tomas Pueyo showed how 22 COVID-related deaths in Washington state probably translated into 3,000 cases state-wide, using approximate mortality rate, average length of the illness before death, predicted transmission rate, and looking for any large clusters than would skew results (Pueyo, 2020).

Reliable data is key in ensuring that any responses to the pandemic addresses inequalities. Data journalism has been crucial in shedding light on the fact that Black, Asian and minority ethnic (BAME) populations may be suffering disproportionately from the pandemic, with deaths of BAME healthcare workers especially high. See this **simple but striking** data visualisation by Bristol-based artist Niki Groom, using statistics cited in Sathnam Sanghera's article, **'Coronavirus and ethnicity: black and Asian NHS medics on the front line'.**

This issue has also exposed **gaps in data gathering**. For example, in the UK, ethnicity is not registered on death certificates (Barr et al., 2020). In the US, neither the Centers for Disease Control and Prevention's information site, nor the Johns Hopkins University database and the COVID Tracking Project offer data desegregated by ethnicity (Kendi, 2020a). But data from hospital deaths suggests that, up to 19 April 2020, 19% of those who died in hospital in England were from BAME backgrounds when BAME residents make up 15% of the country's population (Barr et al, 2020). The release by the Office for National Statistics of data for COVID-19-related deaths sorted by local authority was key in enabling analysts to take this further. It allowed them to establish correlations between COVID-19 deaths and local authority data such as population and environmental characteristics and revealed that a high proportion of BAME residents was the strongest predictor of a high COVID-19 death rate. For every 10% increase in ethnic minority residents there were 2.9 more COVID-19 deaths per 100,000 people (Barr et al, 2020). These correlations were echoed in the US after the New York Times released data by zip code (Kendi, 2020b).

4

WHAT IS DATA ART?

Data-based art uses creativity to help audiences understand complex data, and to engage with the issues that these data explore. Data art can help people learn; it can also be entertaining, joyful and thought-provoking.

Data-based art includes what is known as data visualisation (or dataviz) and data storytelling (for example interactive design and infographics), but as any other art form, it can move beyond representation to become more critical, experimental, and explorative. Data arts also move beyond the visual, to create artwork using sound, dance, movement, or sculpture.

a. Greenpeace poster campaign: 'How to starve to death on a full stomach'
Take a look at these two examples. Both of them are trying to raise awareness of different forms of pollution in very striking, visually clear ways. Each of these campaigns is backed up by complex sets of data, but it is these images that tell a story and hopefully, lead to changes in behaviour.

This example uses two linked images to hone in on one example of global sea pollution: the stomach contents of a young albatross.

The caption reads: 'How to starve to death on a full stomach. The 272 pieces of rubbish pictured above were fed to this fledging albatross along with fish caught by its mother. The plastic accumulated in its stomach until it was literally 'too full to eat'. Careless and unregulated dumping is just one of the ways we're killing our oceans. Become an ocean defender at oceans.greenpeace.org.'

Photography for data visualization here has been used to grab the viewer's attention in order to highlight one aspect of a bigger issue. The caption then offers a way for people to get involved to help address this issue.

Campaign poster by Greenpeace with Publicis Mojo Auckland. Photograph of Albatross chick, © David Liittschwager, 2005.

Photograph of stomach contents, © Susan Middleton, 2005.

b. WWF community campaign: 'Black Cloud'
Similarly, this campaign uses one very striking piece of art to highlight a key issue at a single glance. The text on the balloon reads: 'Drive one day less and look how much carbon monoxide you'll keep out of the air we breathe.'Along with an increase in new volunteers, WWF received widespread news coverage of the campaign.

HOW CAN DATA ART HELP IMPROVE HEALTH AND WELLBEING?

Data art can help health and wellbeing stakeholders spread key messages about health and wellbeing, and can provoke debate and discussion. The ART/DATA/HEALTH project explores how creative approaches based on a range of data can help people to explore health and wellbeing in an engaging and meaningful way.

GET INVOLVED WITH DATA ART

There are lots of online platforms available for you to explore more examples of data art – take a look at these, and at 'Further Reading' in Section 6:

- Watch **TEDtalks on data art**
- The blog **Information is Beautiful** offers a range of artistic and infographic pieces to explore all kinds of data, including the short film **'Making Data into Art'**.
- The free **Flowing Data** blog offers resources for exploring data visualisation yourself as well as examples of data art.

Many cities and regions also run festivals including data artworks (for example, Brighton, Bristol, Leeds, Lincoln, Keele and Halifax all have Digital Festivals), while the V&A runs an annual Digital Design Weekend as part of the London Design Festival.

The Open Data Institute runs a data art programme called 'Data as Culture' (**https://theodi.org/service/data-as-culture/**). They commission international artists to produce work that explores how data can be positive, engaging forces for good. Artworks have included a semi-sentient vending machine, data collection performances, photographs, networked artworks, pneumatic machines, live-coding performances and 'stitch-hacked' jumpers.

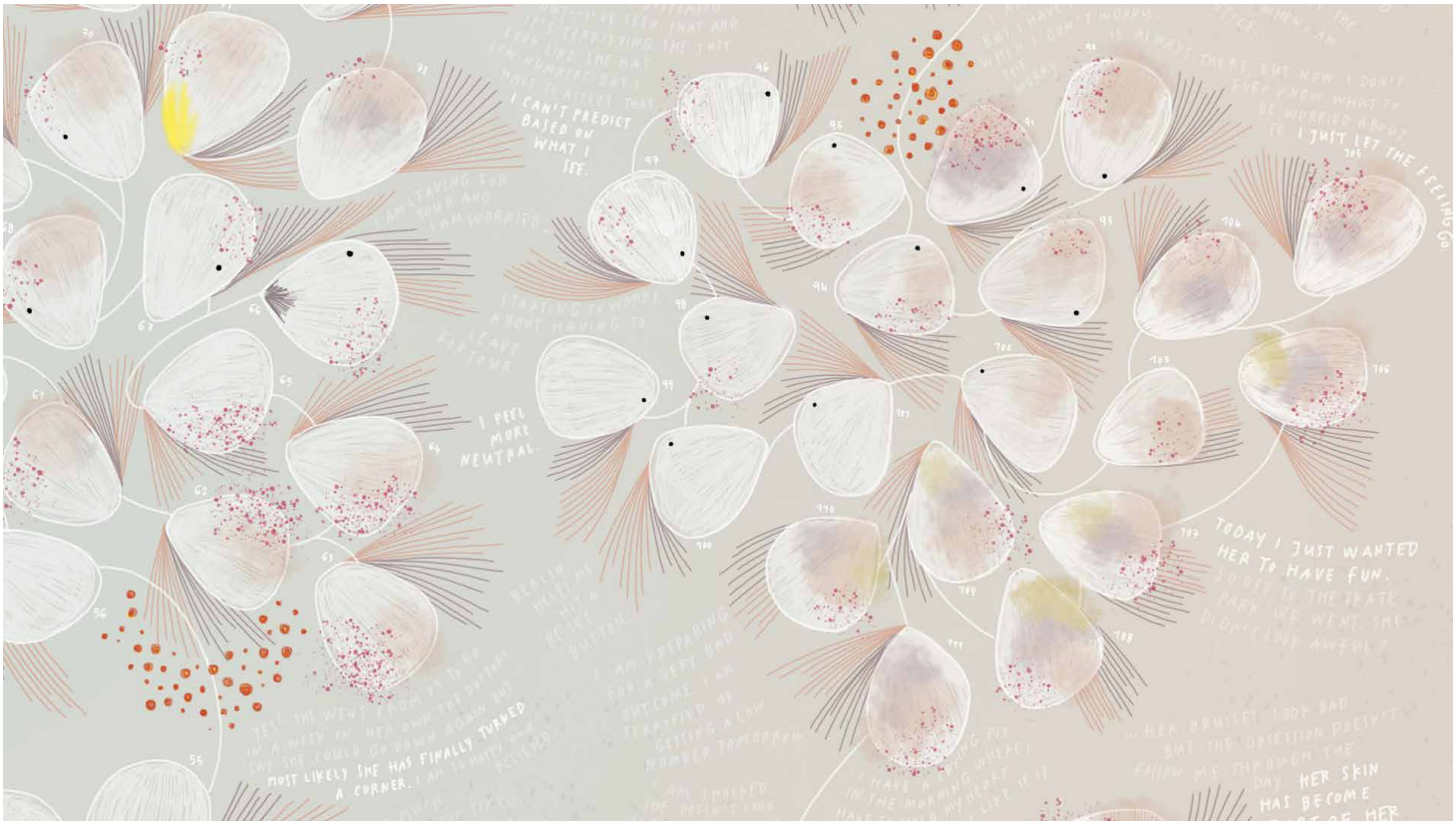
Data art showcase: Giorgia Lupi

Giorgia Lupi's project *Bruises—The Data We Don't See* uses data art to capture the impact that the illness of a child has on a family. Lupi's collaborator, the musician Kaki King, tracked her three-year-old daughter's auto-immune disease symptoms over a four-month period to provide the raw data for the artwork.

Each day was represented by a white petal, and the petals were grouped to make a flower. Each time there was a lab test, a new flower would be started. Lab results were expressed by red bubbles at the base of each flower. Bruises and spots on the petals reflected marks on the child's skin caused by her condition.

The work also visually registered medications, days where the mother was away for work, days of hope, Kaki's level of fear, and notes for each day. The work of Giorgia Lupi seeks to humanise data, to re-connect it with people and their experiences, and to engage the viewer at an emotional level rather than solely a cognitive one. She coins this process 'data humanism': an approach that re-personalises the individuals at the heart of big data. As Lupi puts it, "We are ready to question the impersonality of a merely technical approach to data and to begin designing ways to connect numbers to what they really stand for: knowledge, behaviors, people".

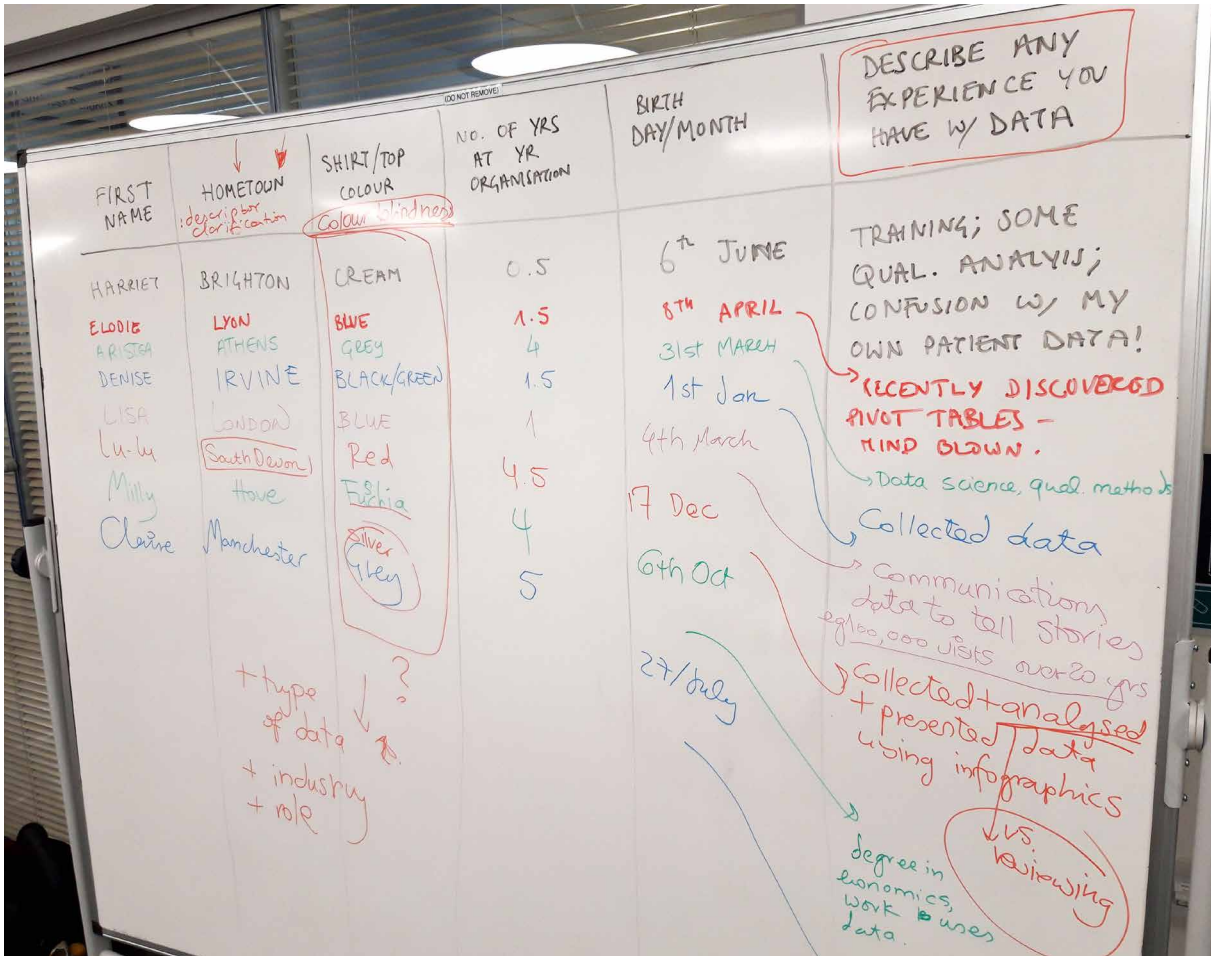
See here to read more about the project, and watch Giorgia Lupi's TED talk on 'data humanism' here.



© Giorgia Lupi, Bruises – The Data We Don't See, reproduced with permission from the artist

5

LEARNING TO WORK WITH DATA: A STEP-BY-STEP GUIDE



An example of the analogue spreadsheet exercise form an ART/DATA/HEALTH workshop (Photograph: ART/DATA/HEALTH project)

5.1 Warm-ups

THE ANALOGUE SPREADSHEET: AN ICEBREAKER FOR 10-30 PEOPLE

To get in the mindset of thinking about and analysing data, try this 'analogue spreadsheet' exercise. This is an introductory activity – perfect for non-technical newcomers to familiarise themselves with basic concepts like 'data', 'datasets', 'data types' and 'cleaning data'.

MATERIALS REQUIRED:

- Giant gridded piece of paper
- Large, thick markers
- Tape to hang Analogue Spreadsheet
- Flat wall to hang Analogue Spreadsheet

PREPARATION:

Create a grid on your giant sheet of paper with at least five columns of personal information. Try to cover most of the major data types with the columns, e.g.:

- First name – qualitative data
- Hometown – geographic data
- Colour of your top – categorical data
- Number of siblings you Have – quantitative data
- Day & month of your birthday – temporal data
- Describe any experience with data – open text as data

Feel free to customize the columns for your audience, but make sure you are not collecting sensitive personal information or anything that would be embarrassing.

RUNNING THE ACTIVITY:

Filling in the Analoque Spreadsheet:

Ask your participants to fill out a row about themselves. You can invite a couple of volunteers to introduce themselves using the spreadsheet.

10 minutes – 'What are Data?' Discussion

- Tell participants that they have just made a 'dataset', basically a table of systematically collected observations about the world. Each row is an observation.
- Ask the group: 'What's missing from this data set? Does your row capture everything about yourself as a person?' Data are a reduction of the world: while they are systematic, they will not cover every aspect of a topic or a person.
- Another example: 'What about top colour?' People invariably have on tops that consist of more than one colour, patterns, and so on. You can use this as an illustration of how the data often reduce the complexity of the world into something manageable.

10 minutes – 'Data Types' Discussion

- Ask participants: 'Often, we think of data as consisting of only numbers. But what types of data other than numbers do you see represented here?'
- As people volunteer types of data, then follow up and ask them, 'What could we do with these types of data? How could we start to organize it?' For example, if people see temporal data, we could plot it on a timeline. If people see geographic information, we might want to make a map. If people see text data, we could look for the most common words.
- The point: Different types of data lend themselves to different ways of exploring, grouping and sorting. Noting what data type a column is can lead you to look at it in a certain way.
- Ask participants: 'What challenges would you face if you were going to make a map of these data?' Typically, people specify the 'hometown' column very differently. Sometimes it has a town and state. Sometimes a country. Same with the 'top colour' and 'birthday' columns. Data are often collected informally and in a non-standardised way, but to plot them systematically we need to clean and standardise them.
- Ask participants: 'What kinds of patterns can you start to see from these data?' For example, in the 'number of siblings' columns – who has the most? The least? What's the average? Walk through the other columns and start to notice their patterns. Ask participants: 'Anything surprising that you see here?'
- Return to the people who introduced themselves at the beginning – ask the group 'How typical is Sandra of the rest of the group? Would you say she's a typical participant at this workshop?'

5 minutes – Review.

Remind participants that they learned that:

- Data are systematic observations of the world
- A dataset is the collection of those observations
- Data are a helpful reduction of the world – it's important to keep in mind they do not capture everything
- The types of data collected (temporal, geographic, quantitative, etc.) affects the kinds of exploration and pattern making that you can do
- Data often need to be cleaned and standardized
- Data storytelling and analysis involves looking for patterns, making comparisons, finding outliers, and then testing that knowledge in dialogue with others.

RESOURCE:

See the full version of this exercise, created by Catherine D'Ignazio, at **https://datatherapy.org/activity-analog-spreadsheet/**.

COLLECTING YOUR ORGANISATIONAL DATA

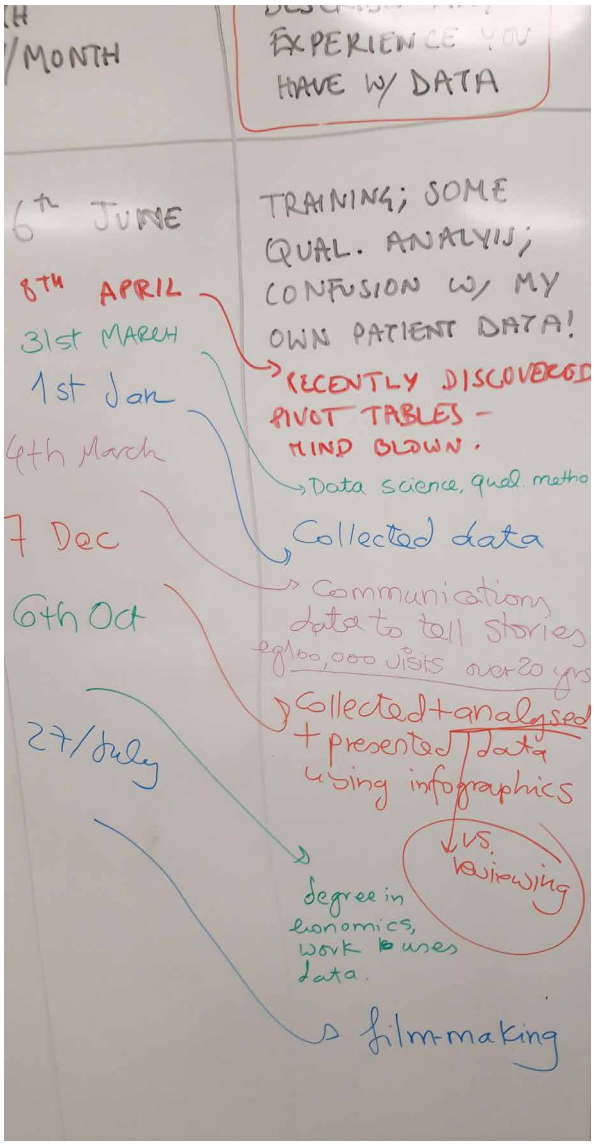
Most third-sector and charity organisations collect data of some sort.[1] This might range from simple contact details of clients, customers and partners, through to a complex range of outcomes related to your work.

For example, an organisation that supports people with health and wellbeing issues will probably track how people are referred into the service, a range of quantitative and qualitative datasets that arise from an initial client assessment, how each person progresses through the service, and what the outcomes are. These data sets may or may not be easy to compare across different teams, platforms and formats. This can be because people collect slightly different data that are difficult to compare directly, or because file formats cannot easily be merged.

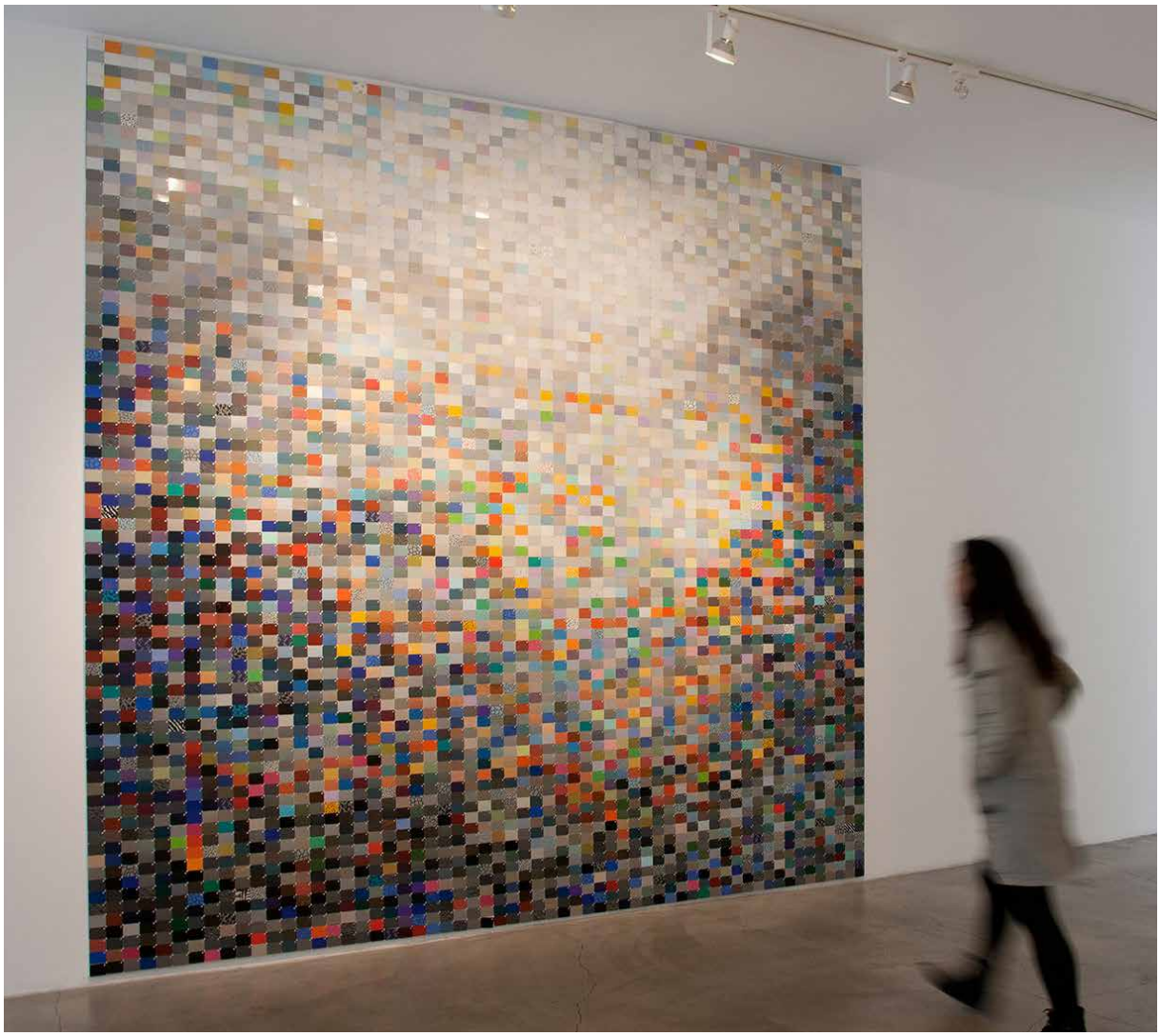
FEARS, HOPES AND OPPORTUNITIES

In order to work with data, organisations need to have a shared vision of exactly what data they are talking about and how they would like to use them. Different members of staff might also have particular fears and hopes in relation to working with data. It is useful to discuss these in detail before a team embarks upon an organisational data project. For example, teams could discuss together:

- What does this organisation understand by 'data'? Does everyone share an understanding of what data means for your organisation/ team?
- What are people's hopes for working with data? What does your team/ organisation want to achieve or change?
- What does your team fear most when it comes to working with data? How will these fears affect the organisation's work?
- What are the opportunities within the organisation's wider sector for using data analysis to work differently?
- Look at your existing data sets being: Are they complete enough? What data are missing? Or are there so many that it is hard to filter out the most important pieces of information?
- Does the team have the right skills and resources in-house to collect and analyse its data? If yes, who is best placed to do so? If no, what training or external support is needed?



[1] In the UK, third-sector, community and charity organisations often deliver services to communities that in other countries might come under the remit of the state. For example, many therapeutic or behavioural treatments in the health and wellbeing sector – such as counselling or addiction services – are run by third sector organisations, even if they may hold contracts with the NHS and other government bodies.



© Laurie Frick, 'Moodjam', reproduced with permission from the artist

Data art showcase: Laurie Frick

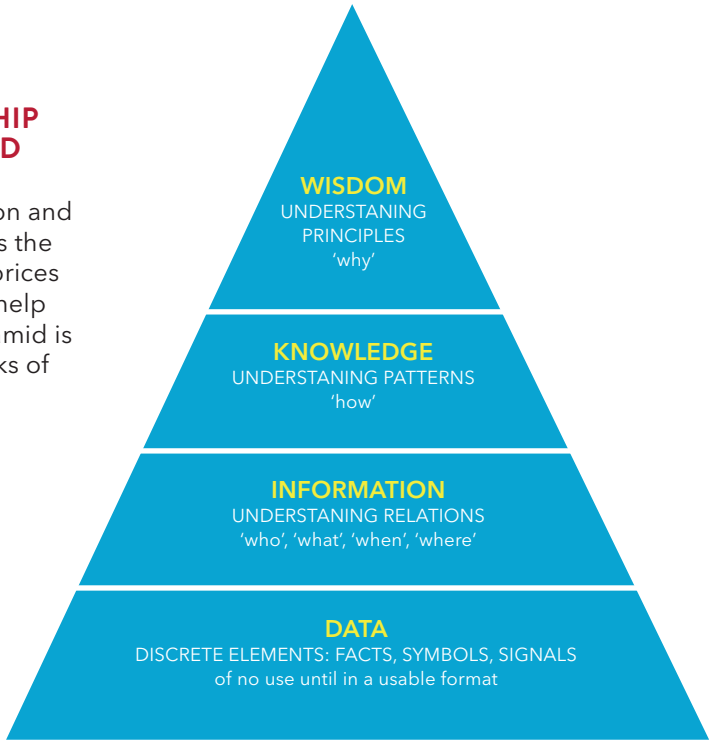
In "Floating Data", Laurie Frick used her own Fitbit walking records and combined them with location data to create a two-storey tall structure made from 60 anodized aluminum panels. She aimed to capture 'the experience of walking speed, and the feel of walking through a busy neighborhood'.

In another work (above), Frick used the online diary Moodjam to track her feelings. She then translated them into canvases and billboard-sized murals made of mosaic-like pieces of Italian laminate countertop samples, with different shades reflecting her changing emotions.

5.2 Introduction to data for advocacy

UNDERSTANDING THE RELATIONSHIP BETWEEN DATA, INFORMATION AND KNOWLEDGE

Data are the material from which information and knowledge can be derived. Think of data as the images, videos, reviews, descriptions and prices that form the basis of information that can help you purchase a bike, for example. This pyramid is a helpful way to visualise the building blocks of how data can be used to build knowledge:



Open Data Institute, CC BY 4.0, based on G Bellinger, D Castro, A Mills, 'Data, information, knowledge, and wisdom', 2004

DEFINITIONS AND THINGS TO CONSIDER

Data are collected everywhere – from mobile GPS and accelerometers to traffic movements and AI. When working with data, you need to think about lots of different aspects: how you collect it, who controls it, and issues of privacy and security. You have probably heard about lots of different terms related to data – big data, open data, or just 'data'. What are they? What are they for? Where are they from? How are they used, and by whom?

'Big data' are extremely large data sets that reveal all kinds of different patterns. The sheer amount of data is usually much, much more than a household computer can cope with in one go. As such, it is often multinational companies with the infrastructure who end up collating and analysing this data. This is why licensing and access is so important.

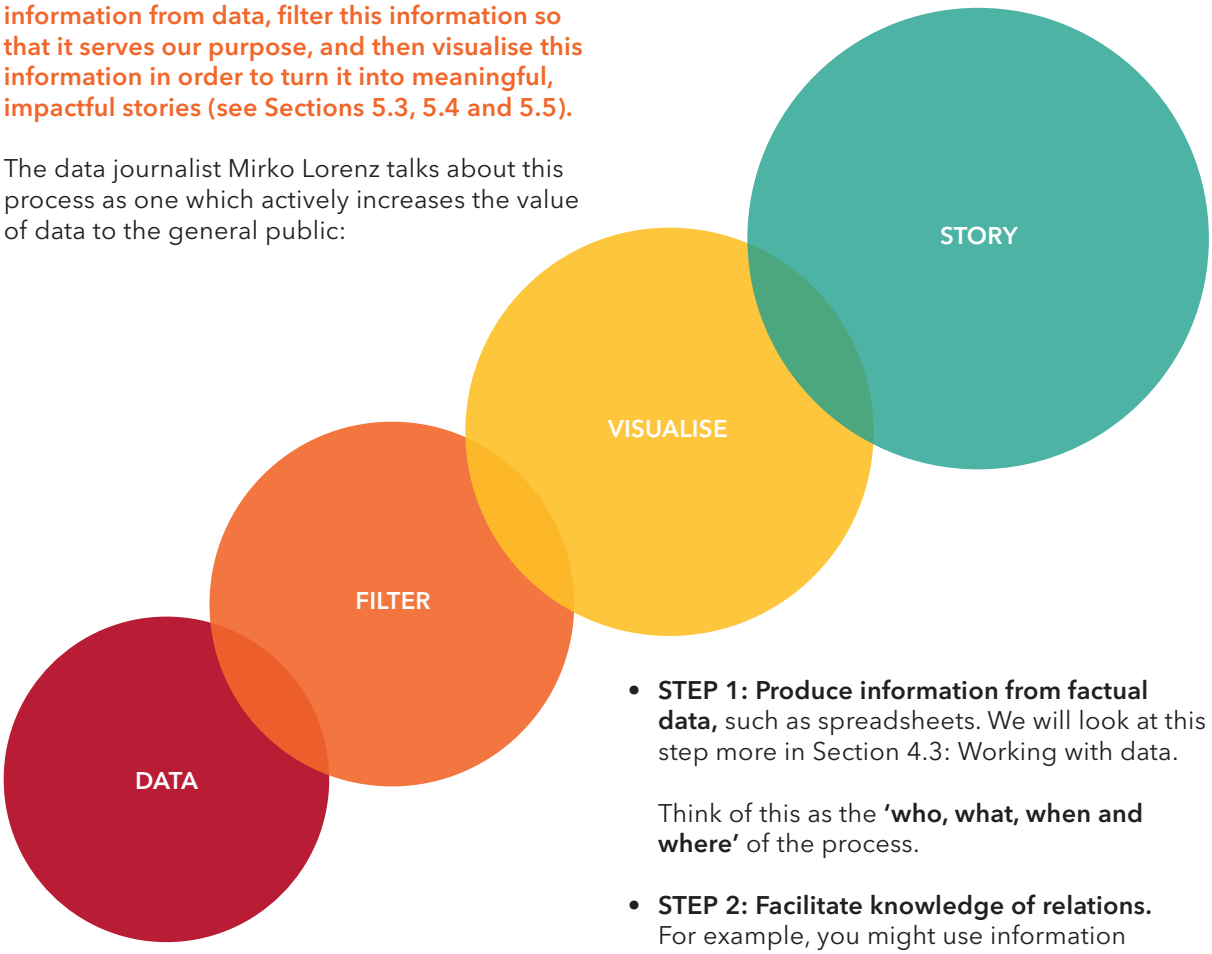
'Open data' are data that can be freely used, re-used and redistributed by anyone, though there are different levels of access and attribution required (See Section 5.3: Working with Data for more details).Data also includes the smaller, local level data sets that you collect every day in your organisations (see 'Collecting your organisational data' in Section 5.1: Warm-ups).

UNDERSTANDING HOW DATA IS USED TO ADVOCATE FOR CHANGE

How can we use data for advocacy?

Data by themselves are not particularly useful. To change behaviour, we need to extract information from data, filter this information so that it serves our purpose, and then visualise this information in order to turn it into meaningful, impactful stories (see Sections 5.3, 5.4 and 5.5).

The data journalist Mirko Lorenz talks about this process as one which actively increases the value of data to the general public:



'Data-Driven Journalism = Process' by Mirko Lorenz, 2010, licensed under CC BY 2.0

- **STEP 1: Produce information from factual data,** such as spreadsheets. We will look at this step more in Section 4.3: Working with data.

  Think of this as the **'who, what, when and where'** of the process.

- **STEP 2: Facilitate knowledge of relations.** For example, you might use information on unemployment statistics to examine the relationship between gender and unemployment

  Think of this work as the **'how'** of your process.

- **STEP 3: Help the public recognise patterns.** For example, we might use open data to compare rates of unemployment between different regions of the UK, so that we can start to think about the factors that influence these figures.

  Think of this step as another **'how'** of your process.

- **STEP 4: Advocacy** – Finally, you need to produce interpretations of a phenomenon, as well as suggestions for improvement. You can then communicate findings to senior level decision makers.

  This step is vital, because it tells people **why** they should pay attention to the story you are telling them, based on the information you have gathered from your raw data.

WHO CAN TELL DATA STORIES?

Anyone can use data to tell a story. There are particular groups, however, who can find this method especially useful.

Communities – from local community groups to international charities – can also use data to tell stories which challenge the stories being told in the media, shape how particular groups are represented, campaign for change, and add different kinds of knowledge to public debate.

It's important for data scientists to be able to make their research into more widely applicable and accessible narratives (see **https://theodi.org/elearning** for more on this, and check out Section 6: Resources). Data journalists also use data as the raw materials for the stories they tell, and these often result in beautiful, engaging graphics.

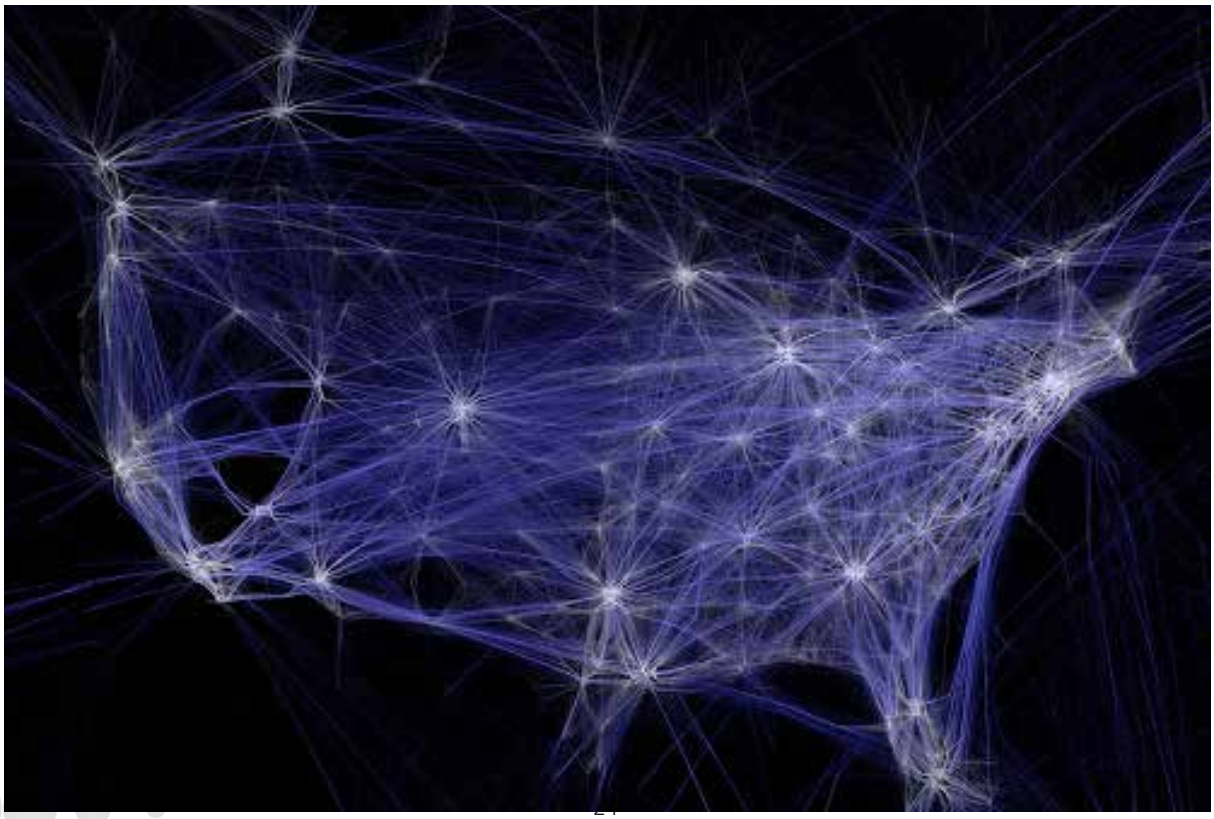
(for example, see the 'Information is Beautiful' blog at **https://informationisbeautiful.net/**).

It's important to think about whose data you prioritise, whether you value some forms of knowledge over others, and who may be over- or under-represented in the stories that are usually in the media. Be alert and responsible in how you interpret your own data to avoid falling into these practices.

Hans Rosling's Gapminder Foundation is a good example of an organisation that uses data to reshape public narratives about key social issues (see **https://www.gapminder.org/**).

CREATIVE MEDIA FOR VISUALISATION

In comparison with conventional ways of presenting evidence, a visual executive summary can help audiences understand and relate to our issue quickly. However, using information for advocacy does not only mean making graphs! Graphs can be impersonal; they cannot capture complex experience; the audience reads data, not stories; and the voices of marginalised communities are often missing. Here are some different approaches to visualising data (In Section 5.4: Visualising data for storytelling you learn more about how to translate data into a meaningful, well-evidenced story)



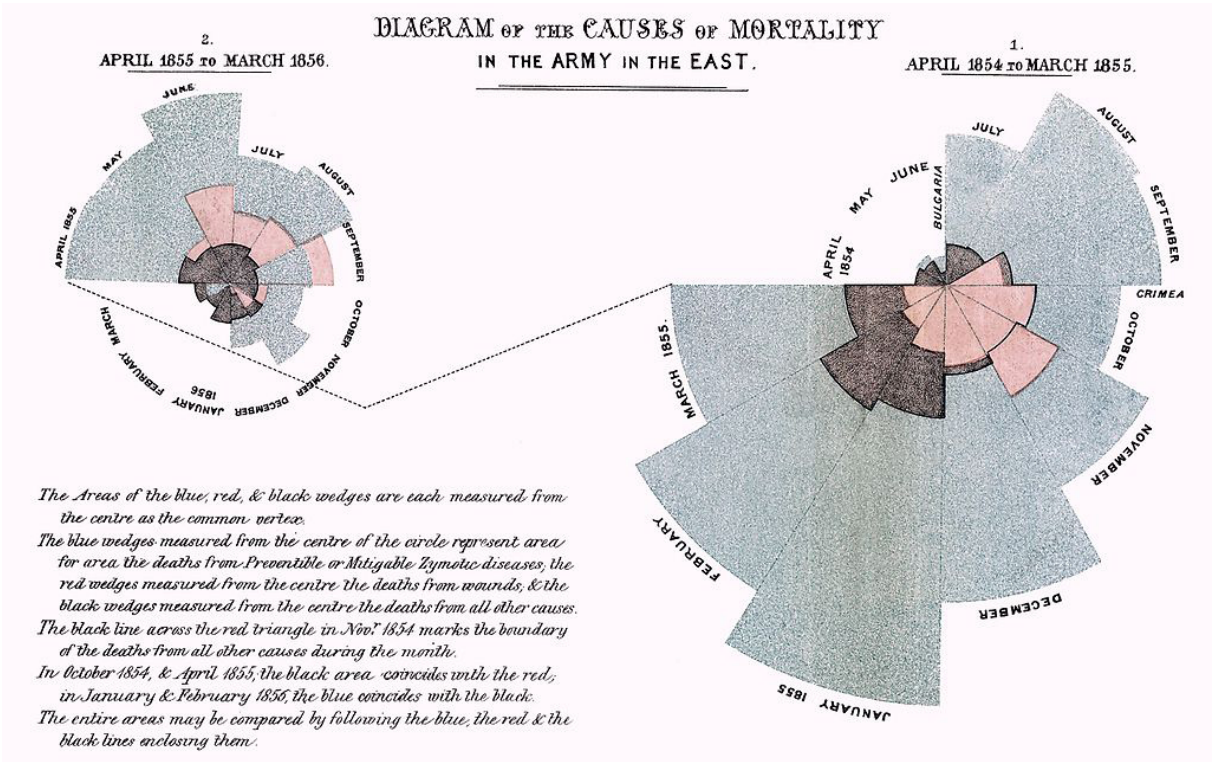
5.2a Ways of representing data

One of the most important benefits of visualisation is that it allows us visual access to huge amounts of data in easily digestible visuals. Well-designed data graphics are usually the simplest and, at the same time, the most powerful.

One of the pioneers in this area was **Florence Nightingale**. While serving as a manager and trainer of nurses during the Crimean War, where she organised the care of the wounded, she uncovered a scandal: more soldiers were killed by preventable diseases caused by unsanitary healthcare than as a result of battlefield wounds.

Nightingale implemented a number of changes, which lead to improved survival rates. But she also meticulously recorded soldiers' mortality data over two years. When the war was finished, she convinced the government to set up a Royal Commission of Enquiry and presented her data in a manner that was revolutionary at the time – and impossible to ignore.

The 'polar area graph' she created shows soldiers' cause of deaths over two years, with the blue wedges representing soldiers dying from diseases that could have been prevented. Her data collection, analysis and visualisation led to a revolution in hygiene worldwide and saved countless lives.



'Diagram of the Causes of Mortality in the Army in the East' by Florence Nightingale, Wikimedia Commons, CC0 1.0

TYPES OF DATA VISUALISATION

- Bar charts and pie charts – to communicate strong messages quickly
- Histograms – much like a bar chart, but it groups numbers into ranges per bar
- Line charts – for displaying time series only
- Maps – though take care to choose the right type of map (see here for more details)
- Infographics – a collection of eye-catching imagery, charts and minimal text to give a clear overview of a particular topic
- Creative media – artistic representations using a range of media

Think about what the best organising principle for the information might be, rather than what is the most convenient. You will need to:

- Balance aesthetics and design with getting the right narrative and foregrounding the most important information
- Use colour to make information 'pop out' and convey meaning (see the example below)
- Make things immediately clear and accessible without dumbing down the issue
- Remember that the success of advocacy and campaign communications can only be judged by how they affect an audience
- Give your audience one clear task or message: for example, a call to change one small piece of behaviour, a volunteer sign-up option, or a donation link

EXAMPLES OF EFFECTIVE DATA VISUALISATION

Tracking overhead surveillance
Journalist Peter Aldhous made a map showing flights over Los Angeles made by the FBI (red) and the Department of Homeland Security (blue) from August to December 2015 for a BuzzFeed article about aerial surveillance of U.S. cities. The circles show when a plane circled over one area for prolonged periods of surveillance.

The U.S. government claimed that these planes were following up on specific leads related to drug smuggling, human trafficking, and other criminal activities, though Buzzfeed reported critics' views that the data showed a tendency to target particular communities by ethnicity and other characteristics. This example shows how effective visualisation can help to add value and meaning to a campaign designed to effect policy and practice.

See the work of digital cartographers Axis Maps for lots more beautiful map visualisations.

Aeroplane seating and COVID-19 infections
This graphic from the *National Geographic* presents the likelihood of catching coronavirus on a plane, using denser colours to represent higher percentages of risk. The use of colour here makes information 'pop out'.

Evidencing potential NHS savings
Mastadon C, in collaboration with the Open Data Institute, identified that if generic medication could be prescribed over branded medication, the NHS could potentially save £200m in a year. To demonstrate this, they analysed 37 million rows of statin prescription data in different regions across England. They visualised the percentage of proprietary statin prescription as a map, using a clear labelling system of colour density per region.

5.2b Data visualisation during Covid-19

As the COVID-19 pandemic progressed in 2019 and 2020, effective visualisation has been particularly important to communicate ideas and facts about the situation.

Good data visualisation can be incredibly useful in conveying information in a format that is easy to understand and help shape policy debates. For example, this data visualisation by Financial Times journalist Bob Haslett compared the US and China in terms of how effectively the spread of the virus was contained. The graphics in this New York Times article contrast confirmed infection with unconfirmed infections, using a subtly creative approach to mimic the spread of the virus through the air. It shows clearly the different between the the very small number of confirmed cases of COVID19 in five major US cities on 1st March 2020 (23 cases) with what researchers now suspect to have been 30,000 unconfirmed infections.

COMPUTER MODELLING

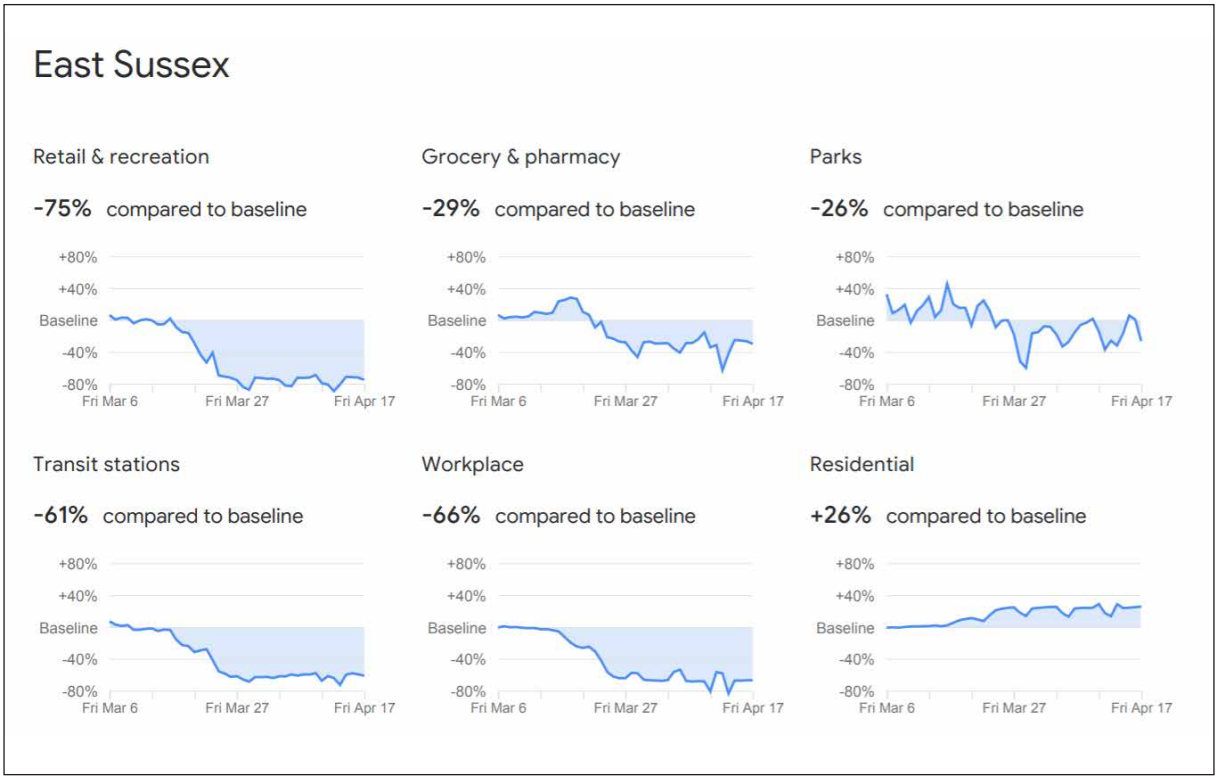
Computer modelling of the COVID-19 pandemic has also been key in discussing and evaluation ways forward. Caitlin Rivers, an infectious diseases modeller with the Johns Hopkins Center for Health Security argues: 'modeling plays a really important role in understanding how an outbreak is unfolding, where it might be going, and what we should be thinking through'.

A Washington Post model visualisation (Stevens 14 March 2020) – shared extensively in social media as the key to understanding social distancing – shows a simulation of people depicted as dots, which move around and thus 'infect' one another. It shows changes of count of the recovered, healthy and sick over time, but interestingly it does not depict deaths.

However, it's worth remembering that modelling is only as good as the data it is based on. Models provide indications, not accurate predictions. The quality of available COVID-19 data, in particular, is poor: "Right now the quality of the data is so uncertain that we don't know how good the models are going to be in projecting this kind of outbreak," says Marc Lipsitch, an epidemiologist at the Harvard T.H. Chan School of Public Health (Greenfieldboyce 2020).

Reponses to the COVID19 pandemic also provide a good example of the many and various ways you might choose to approach a particular topic, rather than focusing on the most obvious types of data. While many visualisations have portrayed the numbers of cases per country, visualisations showing the decrease of air pollution or the risk factors for different jobs have also been widely shared. You can read more about this here.

Google offers Community Mobility Reports which allow you to search by region to see how social distancing measures are working locally. For example, these graphs show how Brighton and Hove is performing compared to its usual baseline:



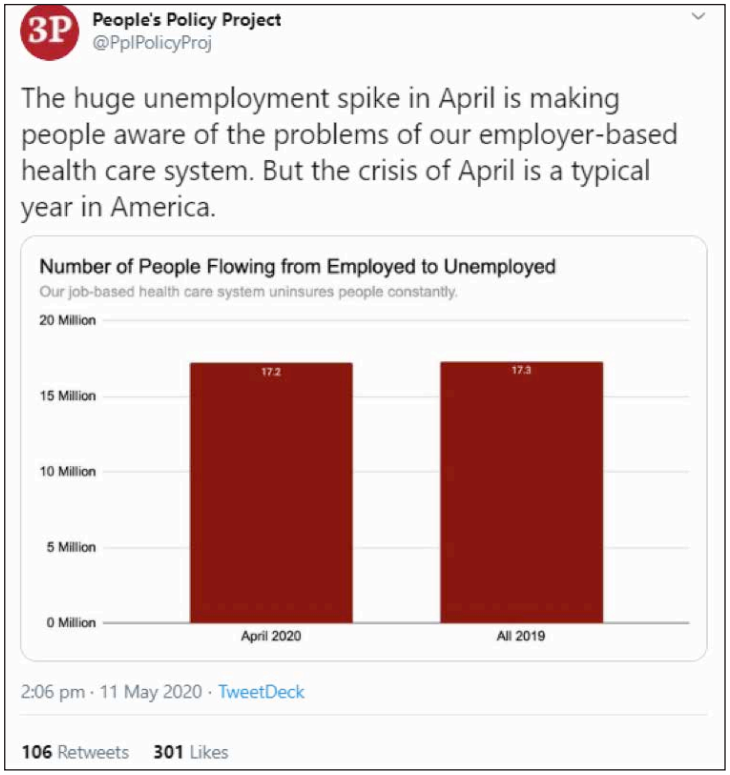
© Google Community Mobility Reports – figures accurate to 14th May 2020

You can read more about data visualisation approaches to the COVID19 pandemic here and here (resource links also appear in full in the Resources section).

TRAINING YOUR CRITICAL EYE

Be aware that many data visualisations are misleading – always take a critical view. Do the segments of a pie chart actually add up to 100%? Are the segments proportionate? Do the horizontal and vertical axes of a bar chart make sense?

For example, take a look at this supposedly comparative bar chart comparing the rate of people moving from employment to unemployment in two periods of time, tweeted by the US think tank People's Policy Project:



People's Policy Project, *The huge unemployment spike in April is making people aware of the problems of our employer-based health care system. But the crisis of April is a typical year in America*, Tweet, 11 May 2020 **https://twitter.com/PplPolicyProj/status/1259832193117020162**

WHAT DO YOU NOTICE?

At first glance, the two bars on this chart show comparable data. But the data labels within each red bar are extremely hard to read. The title 'huge unemployment spike' is not visible at all! Notice how the y-axis (running horizontally) is not an even scale: it gives the same weight to one month ('April 2020') as it does to twelve months ('All 2019'). And why does 2020 appear first, given that English speakers read left to right?

It would be more effective to use colour to make a clear case for the story it is trying to tell. A completely different type of visual graphic might have done the job. And there is a real lack of further information – who made this graphic? Where are the figures collected from? Even if these are publically available, it's good practice to clearly reference your data. You can follow the link to the original tweet (**https://twitter.com/PplPolicyProj/status/1259832193117020162**) to see how commentators have asked for a description of the sources used by the think tank, as well as clearer labelling of the graph.

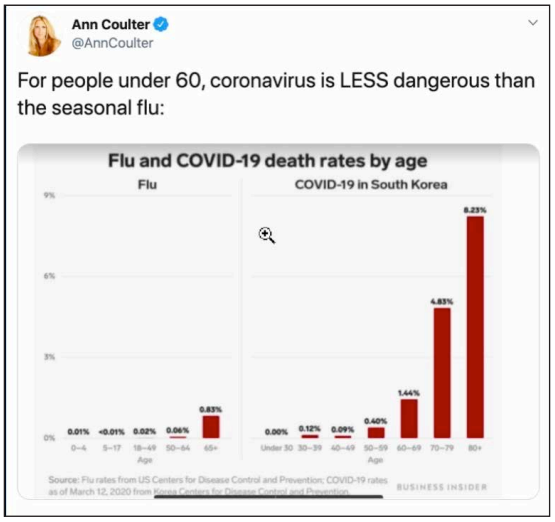
MISLEADING DATA VISUALISATIONS

Some data visualisations may be unintentionally misleading, but be alert to graphics that lead you to a conclusion that support specific agendas.

For example, this graph from business insider was tweeted by Ann Coulter, a far-right US media pundit, to minimise the impact of COVID-19. Have a look at the two graphs.

The message is *For people under 60, coronavirus is LESS dangerous than the seasonal flu*. If you notice the visualisation however, you will see how the two graphs (Flu on the left and C19 on the right) show different age categories and different countries. The two graphs compare death rates associated with seasonal flu in the US with deaths associated with COVID-19 in South Korea. The death rates due to C19 in South Korea have been one of the lowest rates recorded worldwide. The two countries have demonstrated very different ways of dealing with the spread of the virus, so comparing the two data sets without additional context is misleading, and is only used here for political purposes.

What is most striking however is that the figures don't support the assertion made in the tweet! The graph on the left shows 0.08% death rate for flu for Americans under the age of 65. The graph on the right shows 0.61% death rate for South Koreans under 60 for COVID-19. This is over 7 times higher!



Ann Coulter, *For people under 60, coronavirus is LESS dangerous than the seasonal flu*, Tweet, 24 March 2020, **https://twitter.com/anncoulter/status/1242484117373100037?lang=en**

This example illustrates how bad visualisations are not only easily misread but also misrepresented. Read the article about Coulter's misuse of the original figures by journalist Andy Kiersz.

RESOURCE:
see **Viz.wtf** for more examples of data visualisation gone wrong.

DATA VISUALISATION: POINTS TO REMEMBER

- **Correlation does not mean causation –** just because two data sets appear to show a trend over time, this doesn't mean that one causes the other.
- **Complement, don't supplement –** make sure your use of data complements other facts, background information and contextual research. The data analysis and visualization cannot be used to substitute common sense or facts that we know from experience.
- **Make sure you're working with the right data to begin with!** One example of working with the wrong sample relates to the UK 2008 Norovirus outbreak. Two thousand confirmed cases were reported, and the Health Protected Agency calculated that the 'reported to unreported' ratio was 1:1500. This would have meant that 3 million people had it in total; however, there was only a 5% confidence rate in this ratio, which meant that there was a 95% chance that this number didn't in fact have Norovirus, making the statistic almost meaningless.

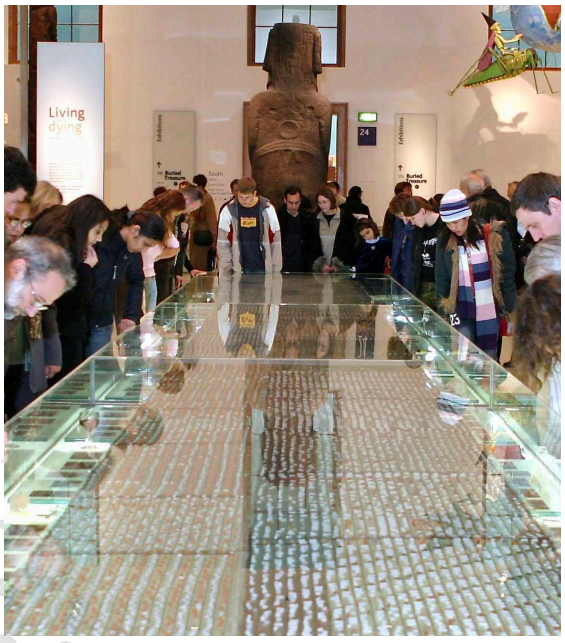
RESOURCE:

Have a look at the list of videos for the course Calling Bullshit (INFO 198 / BIOL 106B. University of Washington) available on the UW Information School's YouTube channel. **https://callingbullshit.org/videos.html**

- **Avoid 'overfitting' your data to a concept.** This means that data are used with too much confidence and not enough evidence. For example, the GoogleFluTrends case in 2013 incorrectly predicted the spread of the flu virus based on Google searches that were later shown to be related to other seasonal events, such as the high school basketball season.

Data art showcase: Pharmacopoeia

Pharmacopoeia is an art collective formed of Liz Lee and Susie Freeman. Their project, 'Cradle to Grave', has been on public display at the British Museum since 2003. It illustrates the medical histories of one man and one woman across two 13-metre lengths of fabric populated with personal objects, images, and documents, and – most strikingly – over 14,000 drugs (the estimated average prescribed to each British citizen across their lifetime, not including over-the-counter remedies). See here to read more about the 'Cradle to Grave' commission.



© Pharmacopoeia, Cradle to Grave, 2003, reproduced with permission from the artists

5.3 Working with data

WHAT ARE OPEN DATA?

Open data are data that anyone can access, use or share. For data to be open, they should have no limitations that prevent them from being used in any particular way.

Open data must be free to use, but this does not mean that they must be free to access – a reasonable cost is one that reflects how much it costs the holders of the data to reproduce them more widely. Once the user has the data, they are free to use, reuse and redistribute them – even commercially.

OPEN DATA PLATFORMS AND DATA AGGREGATORS

A platform is a major piece of software on which smaller pieces of software and content can be run. Open data platforms are pieces of software that make it simpler to publish and manage open data on the Web.

Data Aggregators and portals provide access to a number of datasets in one place. The first examples of governmental portals included data.gov and data.gov.uk.

Aggregators may be run by national governments, municipal authorities, international organisations or domain-specific interest groups such as an academic community. There are over 500 open data portals listed on the directory **dataportals.org.**

OPEN DATA: WHY DO WE NEED THEM?

Open data have the potential to help grow economies, transform societies and protect the environment, for example:

- Innovation and growth in data-driven businesses
- Opportunities for governments from growing data economies
- Impact on society and public policy through improved mobility, new ways of working, and data-driven policymaking
- Benefits for the environment, through research collaborations that increase agricultural productivity

Open data can help make governments more transparent. They can provide the evidence that public money is being well spent and policies are being implemented. For example, according to the campaign group **'Follow the Money'** in Nigeria, open data ensures that public funds are spent implementing the policies promised to the people.

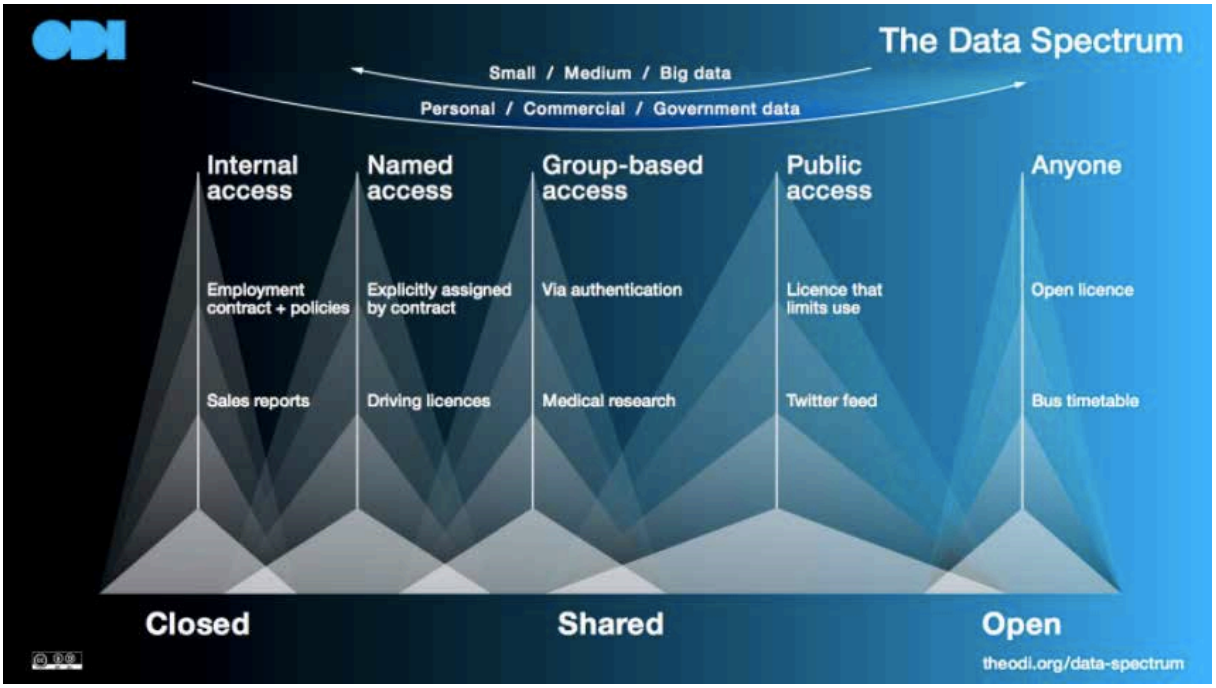
However, it is vital for data to be clearly and thoroughly licensed to make sure that they are fully usable as open data. In the case of COVID19, for example, gov.uk hosts regularly updated open data on infection rates, death rates and testing on a **dedicated platform**.

OPEN DATA LICENSES

Why license open data?

Without a licence, data are not truly open. A licence tells anyone that they can access, use and share your data. Unless you have a licence, data may be 'publicly available', but users will not have permission to access, use and share them under copyright or database laws.

The Open Data Institute provides a lot of information on the different kinds of licence. For example, this image shows what the ODI calls the 'Data Spectrum'. This outlines the range of approaches to data from the most closed (e.g. internal organisational data which is restricted to a small group of people) to the most open (open licence data accessible by anyone).



The Data Spectrum' by the Open Data Institute, licensed under CC-BY v4.0

CREATIVE COMMONS

Creative Commons licences are widely used for open content. Version 4.0 (**https://creativecommons.org/licenses/by/4.0/**) explicitly considers data licensing. There are three Creative Commons types of open licence, with different levels of attribution required:

- **Public domain (CC0):** CC0 is the 'no copyright reserved' option in the Creative Commons toolkit. It effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain, so that anyone may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.
- **Attribution (CC-BY v4.0):** Under this licence you can share data by copying and redistributing them in any medium or format. You can also adapt it by remixing, transforming and building upon materials for any purpose, even commercially. The licensor cannot revoke these freedoms, as long as you follow the license terms.
- **Attribution & Share-alike (CC-BY-SA v4.0):** Under this option, you must give appropriate credit, provide a link to the license, and indicate if changes were made (this is the 'attribution' part of the licence). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original ('Share-alike'). There is a 'No additional restrictions' clause, meaning that you may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

OPEN GOVERNMENT

Some government publishers have chosen to develop their own licences. For example, the UK and French governments use custom open government licenses. The best examples are short, compatible with widely used licences, and easy to comply with.

Under these, you are free to:

- Copy, publish, distribute and transmit the information;
- Adapt the information;
- Exploit the information commercially and non-commercially for example, by combining it with other information, or by including in your own product or application.

For more on data licences, see the Open Data Institute's Data Certificates (**https://certificates.theodi.org/en/about**) and Tim Berners-Lee's 5 star deployment scheme for open data (**https://5stardata.info/en/**).

Data art showcase: Jer Thorp

Jer Thorp is a data artist who was involved in a project entitled "**Cloudy with a Chance of Pain**". The project drew on citizen science to determine a possible link between weather and arthritis symptoms.

Although a popular belief, the link between arthritis and the weather has never been scientifically proven. The research involved the design of a 'quantified self' app, and invited thousands of patients to take part by logging their symptoms. The reporting involved 10 different metrics and involved the collection of geolocation data which were then matched with weather data. Infographics were then developed to explore aspects of the data.

The project helped to confirm a link between rheumatism and the weather. Some of the data visualisations were then turned into physical representations for a celebratory event.

© Jer Thorp / *The Office for Creative Research with Will Dixon and uMotif*, reproduced with permission from the artist

5.4 Telling data stories

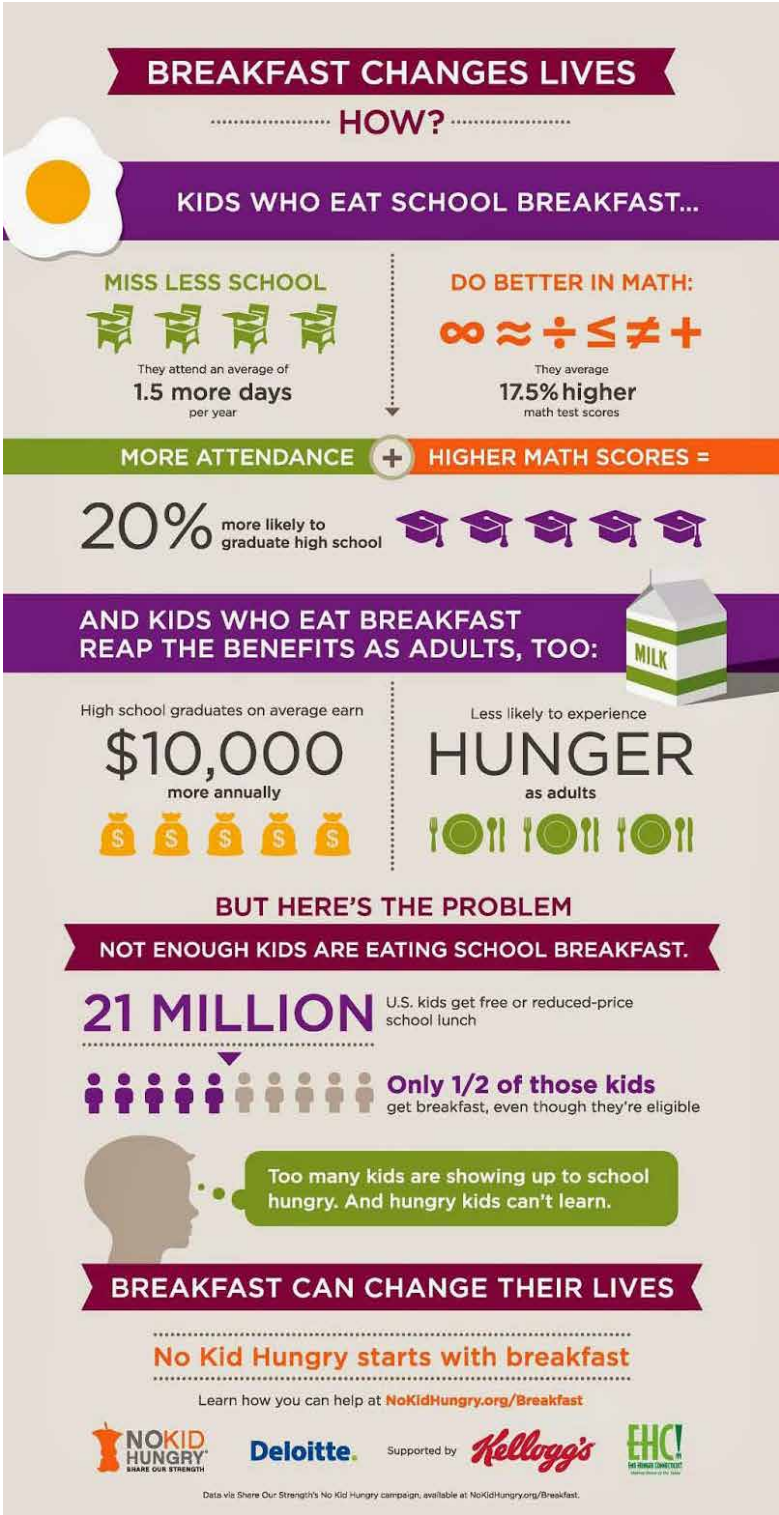
In this section, you'll first look at an example of a data visualisation to examine the data the visualisation uses, explore the ways it shows those data, understand the narrative it is trying to tell, and evaluate whether it combines those together to tell the data story well.

This is followed by a series of questions and resources to help you to define your own goals when telling data stories using creative techniques, including planning your time, selecting your data and format, and structuring your story.

A) EXPLORING THE DIFFERENT WAYS IN WHICH DATA-DRIVEN STORIES CAN HAVE A REAL IMPACT IN THE WORLD.

Take a look at this example of a data story, and ask yourself:

- What datasets have been used in this example?
- How are the data presented?
- What is the story this graphic is trying to tell?
- What are the data and visual representations trying to say? Summarise it in just one sentence, like you would in a tweet.
- Is the story told well in your view?
- What elements best support the story?
- Are there extraneous pieces that don't help the story along?
- Are there pieces that actively distract from the main point or contradict it?



© No Kid Hungry (https://www.nokidhungry.org/)

RESOURCES

You can find a larger version of this image on **Handout 2**. See also the 'Training your critical eye' section under **5.2: Introduction to data for advocacy**.

This visualisation presents a story with a beginning, middle and an end – or, to put it another way, an observation, a problem, and a solution. A good data story needs to be understandable in simple terms, and to end with a message, which may be call to action. Here, it might be: a) school breakfast is good for children who might not otherwise be able to access it; b) not enough children are taking up the offer of free school breakfasts; c) you can help change this by supporting this organisation's work.

The use of different font sizes, a subtle but striking colour palette, and a clearly linear structure help to support this message. A simple visualisation is used to provide statistical evidence; it shows the percentage of children taking up free school breakfasts that they are eligible for. In general the graphic is not overloaded with information.

What do you notice about the data sets used here? A tiny note at the bottom of the visualisation says that the data is from the 'Share Our Strength' campaign. What kind of assumptions and interpretations might have been made in the use of this data? For example, can we be sure that children accessing school lunches also find it hard to fund breakfast? It may well be the case – just notice that certain selections from, and interpretations of, the larger data sets have been made in order to build this story. The story guides you to think that there is causation between eating school breakfast and higher earning in adult life but it gives no data about who these kids are (do they belong is BAME communities for example?) or other barriers to graduating school and entering employment. It also does not tell us WHY kids eligible for breakfast do not get it.

WHAT ARE YOUR OWN OBSERVATIONS OF THIS DATA STORY? DO YOU FIND IT EFFECTIVE? WHY/ WHY NOT?

Now that you can pick apart the meaning behind a data visualisation, you're ready to start building your own.

B) TELLING DATA STORIES USING CREATIVE NARRATIVES

Where to start?

It can seem overwhelming to build a data story if you haven't done it before. The questions in this section will help you to define your goals and consider your options. This work is best done as a team.

Planning your time

Plan your time carefully in any data project. Follow these steps to help you plan efficiently.

The 80:20 rule: Gathering, organising, filtering and cleaning of data will take 80% of your time. Done right, this can significantly reduce the time you spend on the analysis and visualisation of your data.

Develop your narrative at the beginning: Spend the first stages of a project thinking about your message and your audience. This will help keep the data collection focused and time to a minimum.

Selecting your data

First of all, ask yourself the following questions:

- What's the most interesting question you want to ask the dataset you are using?
- Do you need any other datasets to answer this question?
- How could you get the other data you need to answer this question?

RESOURCE:

You can also print out these questions in handout form: see **Handout 1: Asking Good Questions** in the Resources section.See also Handout 5: Finding data

Once you're confident that you have the right data, look in more detail at what you are going to do with it, using the following questions.

CREATIVE NARRATIVE CHECKLIST

What and why:

- What is the aim of your story? What is the change you are trying to make or the problem you are trying to solve, and why?
- What is the benchmark you are trying to reach? For example, a change in the discourse, or a shift in the understanding of a particular event or public attitude to an issue.

Who:

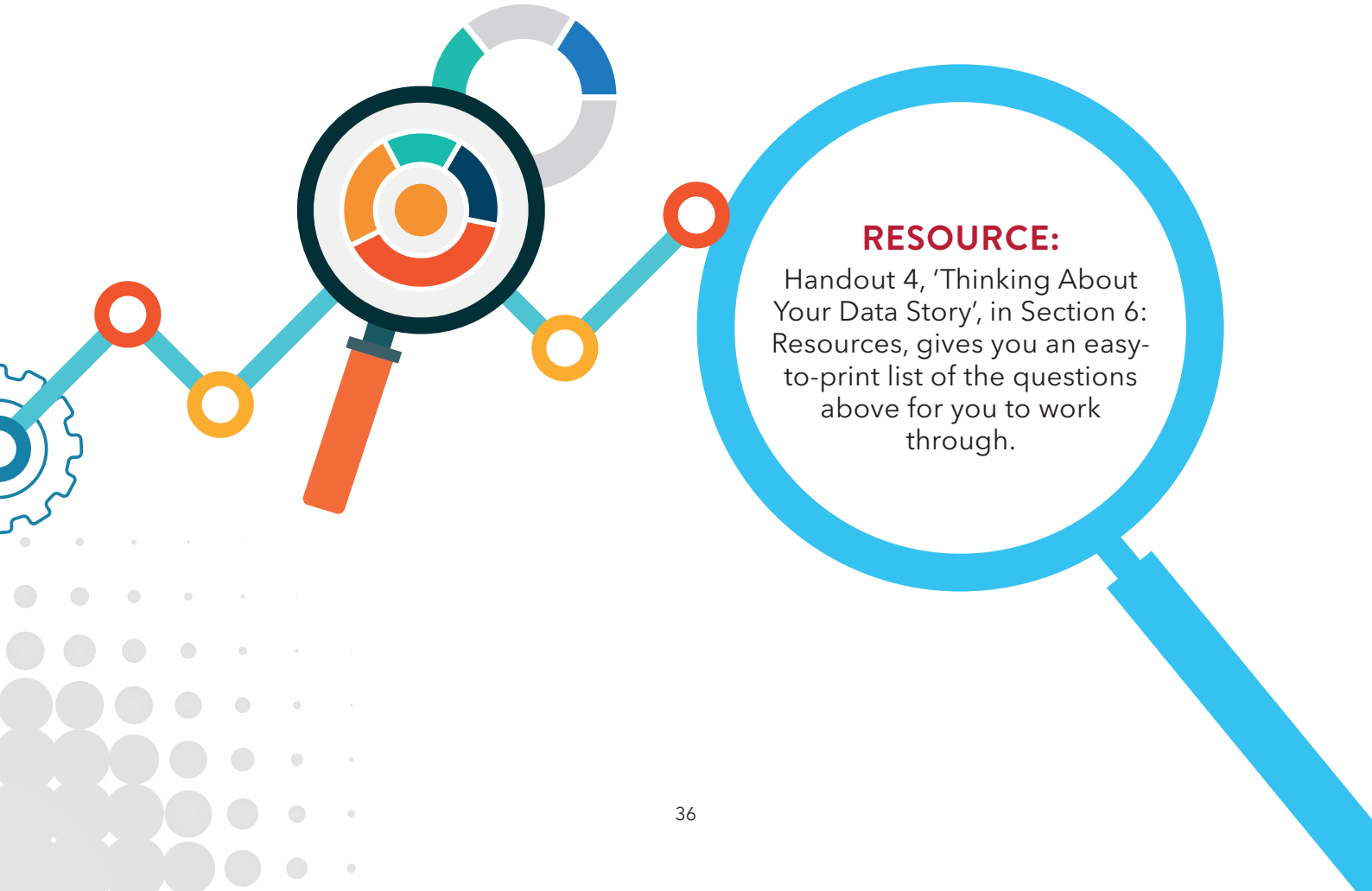
- Who is the audience and why?
- What is their current position and what will nudge or change this?
- What is their current level of understanding of the issue?
- Why should they care?
- What do you want them to do? Why aren't they doing it already?
- What may change their mind and why? What can we tell them that will make them do what we want them to do?
- What influences them? Who do they listen to?
- What is their communication environment?

How:

- How are you going to do it? What will get the audience to change their position?
- What is the role of data in our story?
- What data will we use to tell it?
- What form of intervention will you choose: for example, do you want to interrupt the audience's thinking, or to appeal to their emotions?
- What would change if you were successful?
- How will you use emotional, rational or moral appeals to their values and how will you get them to care?
- How will the information you use get them to care?

Implementation:

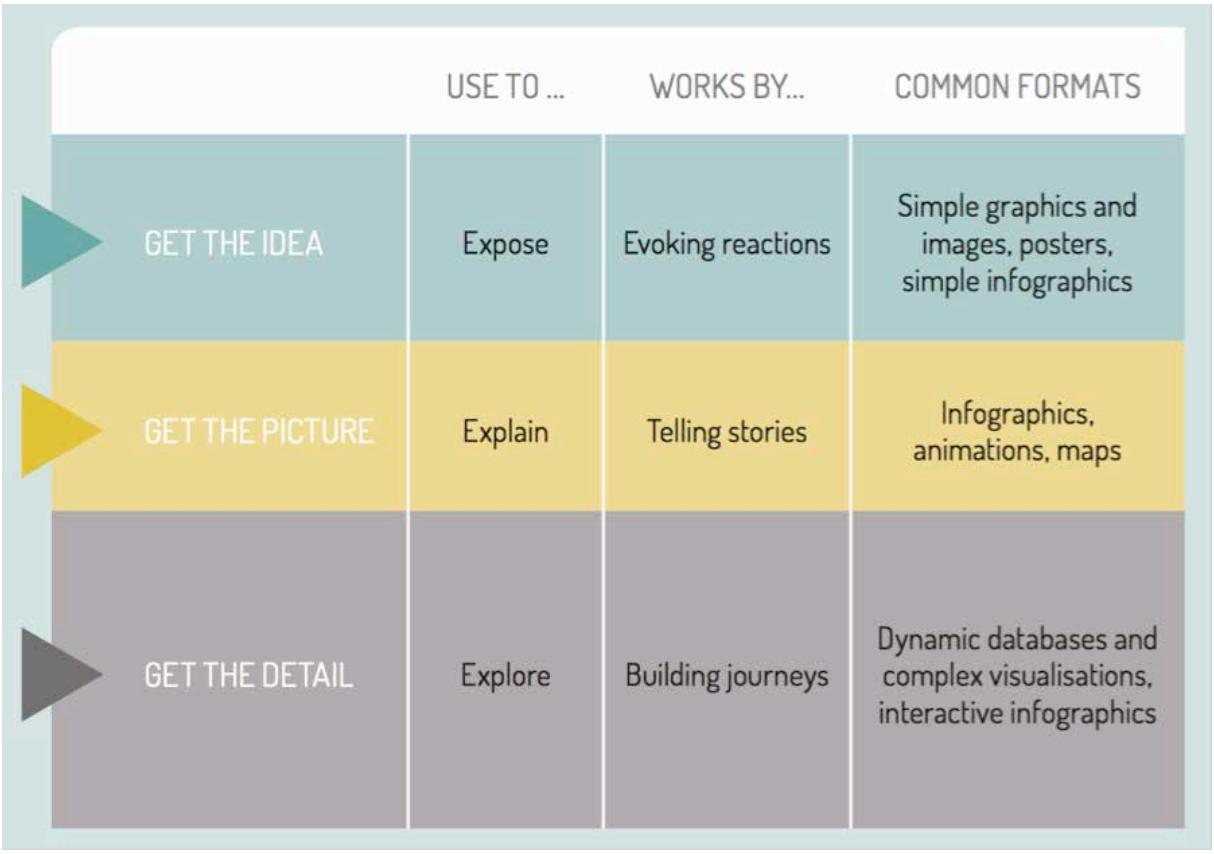
- How will you implement the visual information campaign?
- To which networks will you present the campaign?
- What technologies can you use to deliver the campaign and how will technologies help distribution of or interaction with the campaign?
- How will the design of the information you choose to present support the content?
- What will be the campaign's communicative power?



RESOURCE:

Handout 4, 'Thinking About Your Data Story', in Section 6: Resources, gives you an easy-to-print list of the questions above for you to work through.

What format is most suitable?
To help you think about the most appropriate format to use, think about your ultimate aim:



	USE TO ...	WORKS BY...	COMMON FORMATS
GET THE IDEA	Expose	Evoking reactions	Simple graphics and images, posters, simple infographics
GET THE PICTURE	Explain	Telling stories	Infographics, animations, maps
GET THE DETAIL	Explore	Building journeys	Dynamic databases and complex visualisations, interactive infographics

(Source: Visualising Information for Advocacy three-step approach)

How to tell stories: Planning and framing
Creating a narrative for a campaign from data is a careful balancing act that involves working continually on four fronts at once:

1. **What is the point?** What is it that we want the audience to understand and why? Write a summary of your story with the main points that you want it to cover, using bullet points.

2. **Working outwards from the data:** be clear about what the data tell you. Consider whether the data need to be simplified, contextualised or completed with other data to make our point.

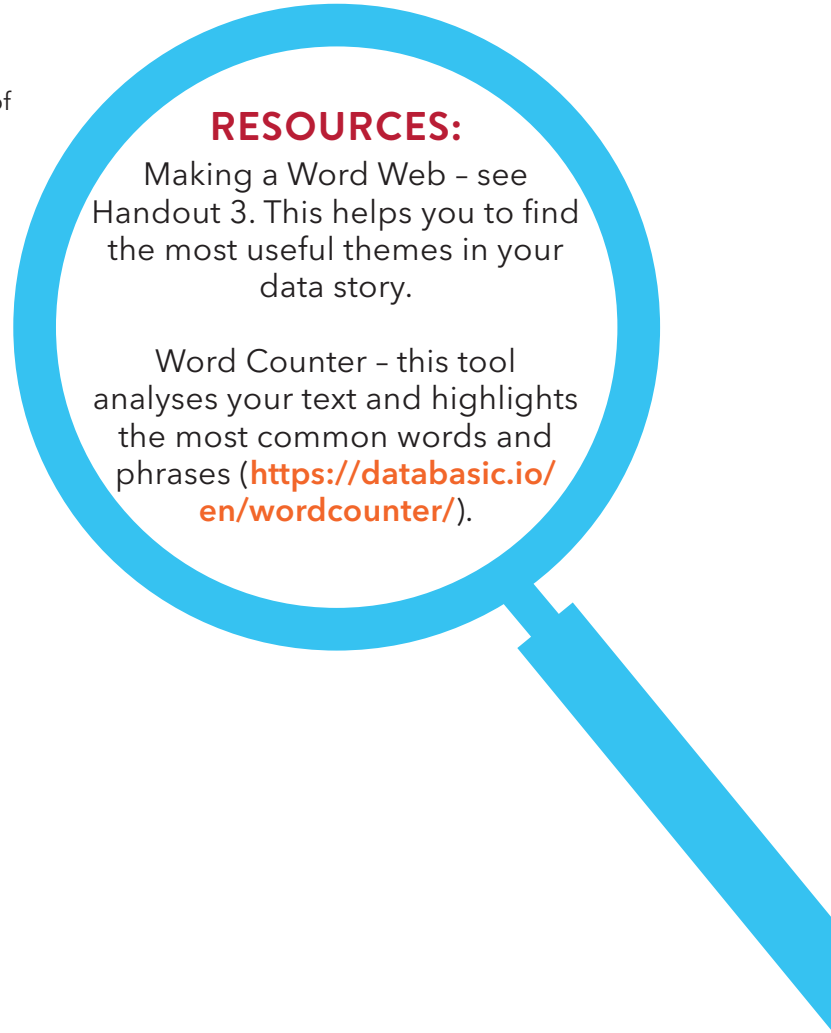
You will need evidence to draw on in your story. Evidence may come from information or quotes you find from another news story, a report, a book or an interview.

3. **Designing your information:** how will you bring your story together in rough data? How can you frame it in compelling ways without misleading or over-fitting data?

4. **Finding visual stories:** what visual devices can be used to present the information in an engaging way? How can the visual design help organise and give meaning to the data?

Evaluating tools for organising and visualising information: points to bear in mind

- When using external infographics or dataviz platforms, their services may be available for free but you may end up losing control of your data. In some cases the data and the resulting visualisation are stored on the program's website/ server and are made public and downloadable by everyone. This may be an issue if you are working with sensitive or politically controversial data.
- You will need the help of a data designer with coding experience to create impressive, interactive visualisations with a "wow" factor.
- Some of these tools come with a mind-boggling array of options and customisable features. Getting your head round the complexity of the tool and what it is capable of can take a long time, but if you plan to do this kind of work more than once, the training is worth it.
- Take note of what the final format of the visual will be. For example, it may only be embeddable in a website and not downloadable. It may be the case that it can only be viewed on the web as the data are stored online. This will not be of much use if you need to print out a poster.
- A lot of the free, open source tools do not have comprehensive support documentation, forums or intuitive interfaces and are quite simply not user-friendly.



Data art showcase: Stefanie Posavec and Miriam Quick

In the series 'Air Transformed': Better with Data Society Commission' (**http://www.stefanieposavec.com/airtransformed**), data designer Stefanie Posavec worked with data researcher Miriam Quick to visualise the burden that air pollution places on bodies.

Drawing on open air quality data from Sheffield, UK, they created 'wearable data objects' such as necklaces and glasses in order to inspire conversations about air pollution. In 'Touching Air', the pair made three necklaces from Perspex segments of different textures, based on data from sensors measuring the health damaging large particulate (PM10) levels. The larger and spikier the segment, the more particulates in the air at that time.

Posavec explains: 'By running their fingers over each necklace, the wearer can literally feel how the air quality in Sheffield went up and down over the course of each week. Dangerous particulate levels have the potential to hurt/prick the finger of the wearer.'

6

RESOURCES

6.1 Online resources

TOOLS AND TECHNIQUES

- **'Analogue Spreadsheet' exercise** to introduce concepts related to data.
- **Bournemouth University Datalabs:** a comprehensive and innovative project that brought together NGOs, designers and journalists to develop data aggregation, digital storytelling and visualisation workshops for addressing civic and humanitarian issues.
- **Data Carpentry:** workshops designed to teach basic concepts, skills, and tools for working with data, with an explicit focus on the skills needed to conduct research in different disciplines.'
- **Data Literacy and data visualization'**: ITunes-based University course from Ohio State University, introducing students to data visualisation and its use in political science.
- **Databasic.io**: a set of tools for students, journalists, non-profit organisations and community activists to help build their data literacy.
- **Deconstructing Data viz** (Data Culture Project): an activity to help you take apart the meaning behind a data visualisation.
- **Flowing Data**: resources for exploring data visualisation yourself.
- **Gapminder Foundation**: offers free training and teaching resources to gather an evidence-based view of world issues.
- **Infogr.am**: a visualisation tool to help you create beautiful graphics from your raw data.
- **Information is Beautiful**: a blog offering a range of artistic and infographic pieces to explore all kinds of data.
- **Introduction to cleaning data** – free online course
- **Introduction to Open Refine data cleaning tool** from the School of Data
- **LinkedIn Excel tutorials**
- **Mirador**: a tool for visually exploring complex datasets
- **Open Data Institute resources**: A set of introductory courses to using and working with data.
- **The Data Journalism Handbook**: Designed for journalist and targeted at people with no prior knowledge of data science, it offers a text-based introduction to interrogating and using data to tell stories.
- **Visual Capitalist**: a wealth of data visualisations related to global economics and politics
- **Visualising Information for Advocacy**: online learning resources that accompany a companion book, exploring how to influence issues using information, design, technologies and networks.
- **Word Counter**: visualise the most commonly used words and phrases in your sample text.
- **World Bank Open Data Essentials**: The World Bank open data toolkit aims to give learners from governments and the World Bank an introduction to how to start and manage an open data initiative.

DATA LICENSING

- **Creative Commons** licences
- **Five-star deployment scheme** for open data licensing, built by Tim Berners-Lee
- The Open Data Institute's **Data Certificates**

6.2 Further reading

ARTICLES, BLOGS AND TALKS:

- 'The digital age of data art': **https://techcrunch.com/2016/05/08/the-digital-age-of-data-art/**
- 'The rise of the data artist': **https://www.theatlantic.com/entertainment/archive/2015/05/the-rise-of-the-data-artist/392399/**
- Watch TEDtalks on data art at **https://www.ted.com/playlists/201/art_from_data**
- The blog **https://informationisbeautiful.net/** offers a range of artistic and infographic pieces to explore all kinds of data.
- *Washington Post* visualisation showing the effects of social distancing on the spread of the Coronavirus: **https://www.washingtonpost.com/graphics/2020/world/corona-simulator/?fbclid=IwAR2hC8DXtddmohRgJijWfAuq9Xzh7XTo30FNspZ69Rv-3oTKbeodrsJ7lYY**
- 'Ten Considerations Before You Create Another Chart About COVID-19': this piece, while specific to the pandemic, is a good set of rules for assessing how appropriate any data visualisation project is: **https://medium.com/nightingale/ten-considerations-before-you-create-another-chart-about-covid-19-27d3bd691be8**
- 'Five questions to ask when you see a coronavirus map': **https://medium.com/@garrett.dash.nelson/five-questions-to-ask-when-you-see-a-coronavirus-map-8e7ec56feeac**
- 'Thirteen Things to Visualize About COVID-19 Besides Case Loads': **https://medium.com/nightingale/thirteen-things-to-visualize-about-covid-19-besides-case-loads-581fa90348dd**

EXAMPLES OF DATA ART PROJECTS:

- Stefanie Posavec and Giorgia Lupi, Dear Data (**http://www.dear-data.com/**)
- Giorgia Lupi, The Bruises We Don't See (**https://medium.com/@giorgialupi/bruises-the-data-we-dont-see-1fdec00d0036**): a data art project that explores the impact of childhood illness on a family (see Section 4)
- Laurie Frick, Floating Data and Moodjam (**https://lauriefrick.com**): data art using self-tracking records (see Section 5.2)
- Pharmacopoeia, Cradle to Grave (**http://www.pharmacopoeia-art.net/exhibitions/collections/**): (see Section 5.3)
- Jer Thorp, Cloudy with a Chance of Pain (**https://www.jerthorp.com/cloudywithachanceofpain**): (see Section 5.4)
- Stefanie Posavec and Miriam Quick, Air Transformed (**http://www.stefanieposavec.com/airtransformed**): a visualization of the burden that air pollution places on bodies (see Section 5.4)

6.3 References to works cited

Unless otherwise specified, all online links were accessed on 13th May 2020.

Arndt, R. Z. (2018). How third parties harvest health data from providers, payers and pharmacies. *Modern Healthcare*, 48(15), p. 18. Available at: **https://search.proquest.com/docview/2024121781**.

Barr, C., Kommenda, N., McIntyre, N. and Voce, A. (2020). Ethnic minorities dying of Covid-19 at higher rate, analysis shows. *The Guardian*. Available at: **https://www.theguardian.com/world/2020/apr/22/racial-inequality-in-britain-found-a-risk-factor-for-covid-19**.

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., and Shadbolt, N. (2018). Third Party Tracking in the Mobile Ecosystem. *WebSci '18: Proceedings of the 10th ACM Conference on Web Science*, pp. 23–31. Available at: **https://doi.org/10.1145/3201064.3201089**.

Bot, B. M., Wilbanks, J. T., and Mangravite, L. M. (2019). Assessing the consequences of decentralizing biomedical research. *Big Data and Society*, [online] 6(1). Available at: **https://doi.org/10.1177/2053951719853858**.

Carter, P., Laurie, G. T., Dixon-Woods, M., Carter, P., Laurie, G. T., & Dixon-woods, M. (2015). The social licence for research: why care.data ran into trouble. *Journal of Medical Ethics*, 41(5), pp. 404–409. Available at: **jstor.org/stable/44014080**.

Castle-Clarke, S., & Hutchings, R. (2019). *Achieving a digital NHS: lessons for national policy from the acute sector*. Available at: **https://www.nuffieldtrust.org.uk/research/achieving-a-digital-nhs-lessons-for-national-policy-from-the-acute-sector%0Ahttps://www.nhsx.nhs.uk/**.

Department of Health and Social Care. (2018). *The future of healthcare: our vision for digital data and technology in health and care*. Available at: **https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care**.

Devlin, H. (2020). Why are people from BAME groups dying disproportionately of Covid-19? *The Guardian*. Available at: **https://www.theguardian.com/world/2020/apr/22/why-are-people-from-bame-groups-dying-disproportionately-of-covid-19**.

Fdez-Arroyabe, P., Fernandez, D. S., & Andres, J. B. (2019). Work environment and healthcare: a biometeorological approach based on wearables. In N. Dey, A. S. Ashour, S. J. Fong, and C. Bhatt (Eds.), *Wearable and Implantable Medical Devices: Applications and Challenges*, pp. 141–161. Available at: **https://doi.org/10.1016/B978-0-12-815369-7.00006-9**.

Ford, E., Boyd, A., Bowles, J. K. F., Havard, A., Aldridge, R. W., Curcin, V., and Sperrin, M. (2019). Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond. *Learning Health Systems*, [online] 3(3). Available at: **https://doi.org/10.1002/lrh2.10191**.

Fotopoulou, A (2020) Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility, Information Communication and Society. Available at: **https://www.tandfonline.com/eprint/MZPZRPPG4FPPQBGGGNAS/full?target=10.1080/1369118X.2020.1716041**

Fotopoulou, A. (2018). From networked to quantified self: Self-tracking and the moral economy of data. In Z. Papacharissi (Ed.), *A Networked Self: Platforms, Stories, Connections*. New York, USA: Routledge.

Fotopoulou, A., and O'Riordan, K. (2016). Training to self-care: fitness tracking, biopedagogy and the healthy consumer. *Health Sociology Review*, 26(1), pp. 54–68. Available at: **https://doi.org/10.1080/14461242.2016.1184582**

Giles, C. (2020). UK coronavirus deaths more than double official figure, according to FT study. *Financial Times*. Available at: **https://www.ft.com/content/67e6a4ee-3d05-43bc-ba03-e239799fa6ab**.

Hunn, A. (2017). *Survey of the general public: Attitudes towards health research*. Available at: **https://www.hra.nhs.uk/documents/1422/HRA_NIHR_general_public_omnibus_survey_2017_FINAL.pdf**.

IPSOS MORI for the Wellcome Trust. (2016). *The One-Way Mirror: Public attitudes to commercial access to health data*. Available at: **https://www.ipsos.com/sites/default/files/publication/5200-03/sri-wellcome-trust-commercial-access-to-health-data.pdf**.

Kendi, I. X. (2020a). Why Don't We Know Who the Coronavirus Victims Are? *The Atlantic*. Available at: **https://www.theatlantic.com/ideas/archive/2020/04/stop-looking-away-race-covid-19-victims/609250/?utm_source=twitter&utm_medium=social&utm_campaign=share**.

Kendi, I. X. (2020b). What the Racial Data Show. *The Atlantic*. Available at: **https://www.theatlantic.com/ideas/archive/2020/04/coronavirus-exposing-our-racial-divides/609526/**.

Li, X., Dunn, J., Salins, D., Zhou, G., Zhou, W., Schüssler-Fiorenza Rose, S. M., and Snyder, M. P. (2017). Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. *PLOS Biology*, [online] 15(1). Available at: **https://doi.org/10.1371/journal.pbio.2001402**.

Mann, N. (2014). The care.data programme: explicit detail on protocols and safeguards is needed. *BMJ: British Medical Journal*, Vol. 348. Available at: **https://doi.org/10.1136/bmj.g2201**.

McCartney, M. (2014). Care.data doesn't care enough about consent. *BMJ: British Medical Journal*, Vol. 348. Available at: **https://doi.org/10.2307/26514579**.

NHS. (2017). *Harnessing technology and innovation*. Available at: **https://www.england.nhs.uk/five-year-forward-view/next-steps-on-the-nhs-five-year-forward-view/harnessing-technology-and-innovation/**.

NHS. (2019). *The NHS Long Term Plan*. Available at: **https://www.longtermplan.nhs.uk/**.

Pateraki, M., Fysarakis, K., Sakkalis, V., Spanoudakis, G., Varlamis, I., Maniadakis, M., and Koutsouris, D. (2020). Biosensors and Internet of Things in smart healthcare applications: challenges and opportunities. In In N. Dey, A. S. Ashour, S. J. Fong, and C. Bhatt (Eds.), *Wearable and Implantable Medical Devices: Applications and Challenges*, pp. 25–53. Available at: **https://doi.org/10.1016/B978-0-12-815369-7.00002-1**.

Pueyo, Tomas. (2020). Coronavirus: Why You Must Act Now. *Medium*. Available at: **https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca**.

PwC. (2016). *Use of wearables in the workplace is halted by lack of trust*. Available at: **https://pwc.blogs.com/press_room/2016/06/use-of-wearables-in-the-workplace-is-halted-by-lack-of-trust-pwc-research.html**.

Richardson, S. and Spiegelhalter, D. (2020). Coronavirus statistics: what can we trust and what should we ignore? *The Guardian*. Available at: **https://www.theguardian.com/world/2020/apr/12/coronavirus-statistics-what-can-we-trust-and-what-should-we-ignore**.

Sanghera, S. (2020). 'Coronavirus and ethnicity: black and Asian NHS medics on the front line'. *The Times*. Available at: **https://www.thetimes.co.uk/edition/news/coronavirus-and-ethnicity-black-and-asian-nhs-medics-on-the-front-line-dhznrc3kc**.

Schairer, C. E., Rubanovich, C. K., & Bloss, C. S. (2018). How Could Commercial Terms of Use and Privacy Policies Undermine Informed Consent in the Age of Mobile Health? *AMA Journal of Ethics*, 20(9), pp. 864–872. Available at: **www.amajournalofethics.org**.

Sharon, T., & Lucivero, F. (2019). Introduction to the Special Theme: The expansion of the health data ecosystem – Rethinking data ethics and governance. *Big Data and Society*, Vol. 6. Available at: **https://doi.org/10.1177/2053951719852969**.

Sheather, J., & Brannan, S. (2019). Patient confidentiality in a time of care.data: The benefits of unlocking patient data are enormous; hopefully the risks can be managed. *BMJ: British Medical Journal*, 347(7935). Available at: **https://www.jstor.org/stable/43513265**.

Skovgaard, L. L., Wadmann, S., & Hoeyer, K. (2019). A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good. *Health Policy*, Vol. 123, pp. 564–571. Available at: **https://doi.org/10.1016/j.healthpol.2019.03.012**.

Statista. (2020). *Connected wearable devices worldwide 2016-2022*. Available at: **https://www.statista.com/statistics/487291/global-connected-wearable-devices/**.

Stockdale, J., Cassell, J., & Ford, E. (2019). "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland [version 2]. *Wellcome Open Research*, Vol. 3, p. 6. Available at: **https://doi.org/10.12688/wellcomeopenres.13531.2**.

Tsang, T. K., Wu, P., Lin, Y., Lau E. H. Y., Leung, G. M. and Cowling, B. J. (2020). Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *The Lancet*, Open Access. Available at: **https://doi.org/10.1016/S2468-2667(20)30089-X, https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(20)30089-X/fulltext**.

UK Data Service. (n.d.). *Why share data*? Available at: **https://www.ukdataservice.ac.uk/manage-data/plan/why-share.aspx**.

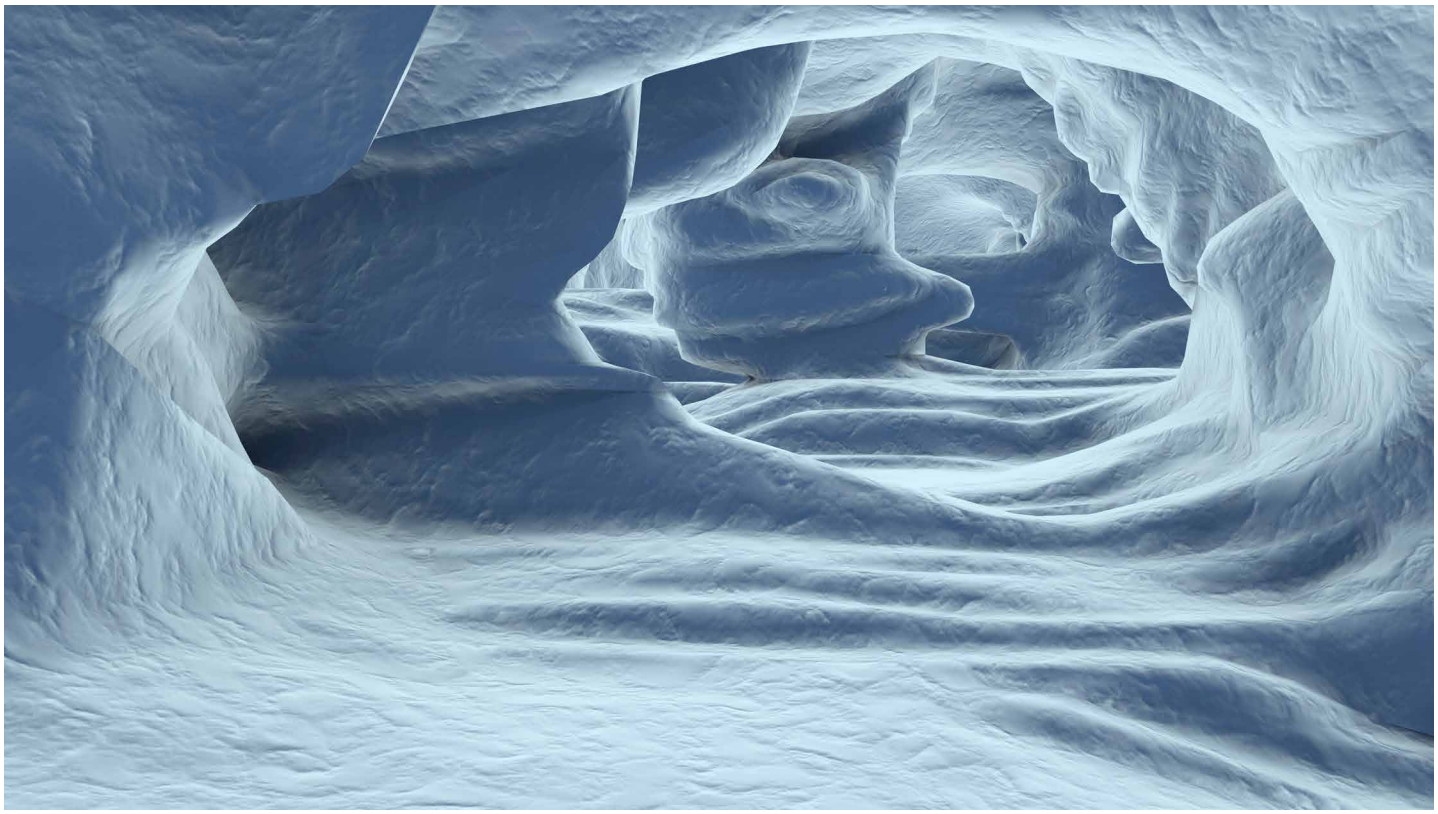
Understanding Patient Data. (2018). *Public attitudes to patient data use: A summary of existing research*. Available at: **https://understandingpatientdata.org.uk/sites/default/files/2019-05/Public attitudes key themes 0.pdf**.

Wu, M., & Luo, J. (2019). Wearable Technology Applications in Healthcare: A Literature Review. *Online Journal of Nursing Informatics (OJNI)*, [online] 23(3). Available at: **https://www.himss.org/resources/wearable-technology-applications-healthcare-literature-review#_ENREF_22**.

Zadeh, P. E. & Tremblay, M. C. (2016). A review of the literature and proposed classification on e-prescribing: Functions, assimilation stages, benefits, concerns, and risks. *Research in Social and Administrative Pharmacy*, 12:1, pp. 1-19.

6.4 Handouts

In the following pages you will find the handouts for four of the exercises featured or mentioned in this booklet:
Handout 1: Asking Good Questions (Section 5.3: Working with data)
Handout 2: Analysing data visualisations (Section 5.4: Telling data stories)
Handout 3: Making a Word Web (Section 5.4: Telling data stories)
Handout 4: Thinking About Your Data Story (Section 5.4: Telling data stories)



HANDOUT 1:
ASKING GOOD QUESTIONS

Our Question

What's the most interesting question you want to ask the dataset you are looking at?

Other Data We Need

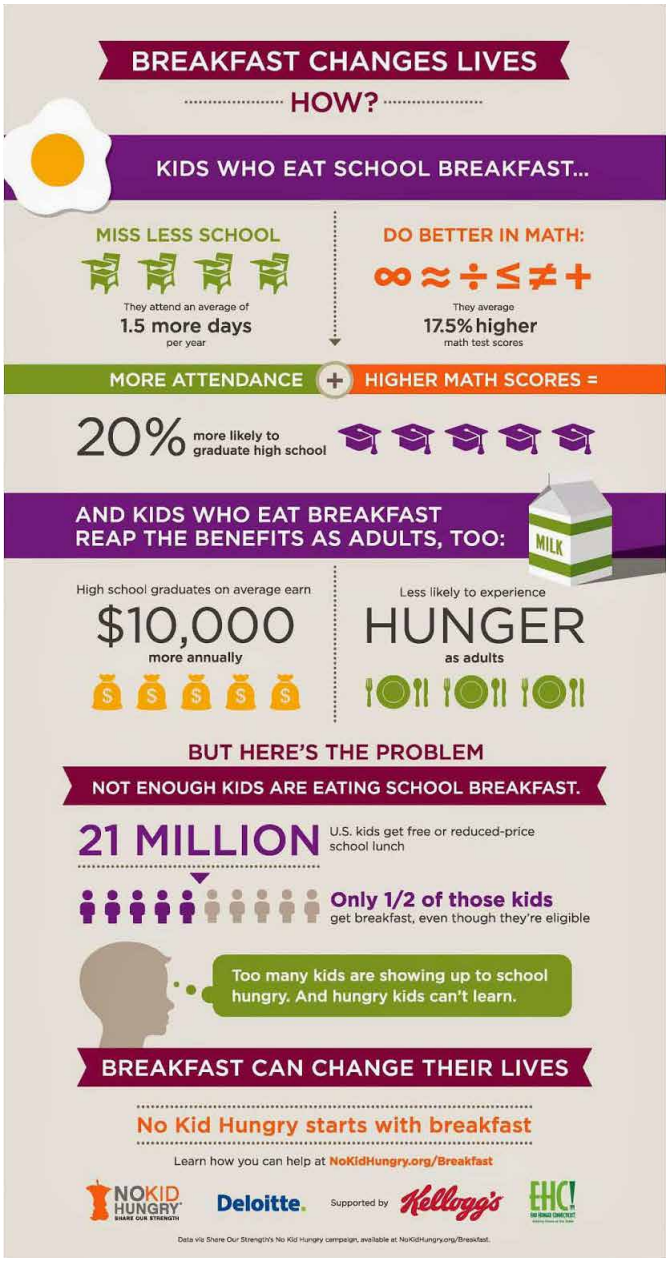
Do you need any other datasets to answer this question?

Sources for that Other Data

How could you get the other data you need to answer this question?

UKRI Arts and Humanities Research Council

HANDOUT 2:
ANALYSING DATA VISUALISATION



© https://www.nokidhungry.org/

Questions to help you analyse a data story

- What datasets have been used in this example?

- How are the data presented?

- What is the story this graphic is trying to tell?

- What are the data and visual representations trying to say? Summarise it in just one sentence, like you would in a tweet.

- Is the story told well in your view?

- What elements best support the story?

- Are there extraneous pieces that don't help the story along?

- Are there pieces that actively distract from the main point or contradict it?

HANDOUT 3: MAKING A WORD WEB

Set-up

This activity will take 15 to 20 minutes. You will need big sheets of paper, lots of pens of different colors, and sticky notes.

Background

Sometimes your data story is centered around an abstract idea, such as "climate", "injustice", or "rights". Abstract ideas are hard to picture, and even harder to draw. This activity helps you brainstorm more concrete ideas that are related to the abstract concepts. It also helps you come up with visual symbols you could later incorporate into a visual design that tells your data story. Making word webs encourages collaborative teamwork and collective brainstorming.[2]

Kick off the activity

Begin by spreading out large pieces of paper, each with an abstract concept written in the middle. The concepts should be from a data-driven story you are working with. Words like "poverty", "injustice", and "happiness" are all good examples of abstract concepts that come from data-driven stories.

Give each participant a pen and break them up into groups of 5 or 6, with each group assigned to one of the pieces of paper you just showed. Tell the participants they should start by drawing a line from the central word and writing another word that they associate with the first one. Keep adding words connected to the first word or to the ones that other people add. Give the groups 6 minutes to brainstorm and write words. Each paper should end up looking like a web of words, connected by lines.

Once the time is up, hand out the sticky notes. Give folks another five minutes to identify any words that can easily be sketched out, and then have them draw those on a note and stick it next to the word.

Have everyone share back

Bring everyone back together, and have each group hang the sheets of paper on the wall. Give participants a few minutes to walk around looking at what other groups created. Ask the full group about the connections they saw, the unique or unconventional ideas conveyed, and the pictures that stood out to them as iconic or effective. Highlight any drawings that carry a particular tone, or ones that may only make sense in certain contexts or cultures. Discuss which pictures are most effective for conveying the concept of the data-driven story you started with.

Note that there are online resources which can help you brainstorm, such as the Noun Project (**https://thenounproject.com**).

HANDOUT 4: THINKING ABOUT YOUR DATA STORY

These questions will help you to plan the what, why who, and how of your data story.

The what and the why:

- What is the aim? What is the change you are trying to make or the problem you are trying to solve, and why?
- What is the benchmark you are trying to reach? For example, a change in the discourse, or a shift in the understanding of a particular event or issue.

The who:

- Who is the audience and why?
- What is their current position and what will nudge or change this?
- What is their current level of understanding of the issue?
- Why do they care?
- What do we want them to do?
- Why aren't they doing anything already?
- What may change their mind and why?
- What can we tell them that will make them do what we want them to do? What influences them?
- Who do they listen to?
- What is their communication environment?

The how:

- How are you going to do it? What will get the audience to reconsider or change their position?
- What is the role of data in our story?
- What data will we use to tell it?
- What form of intervention will you choose? For example: interruption, education or coercion?
- What would change if you were successful?
- How will you use emotional, rational or moral appeals to their values and how will you get them to care?
- What role will the information you use play in this?

Implementation:

- How will you implement the visual information campaign?
- To which networks will you present the campaign?
- What technologies can you use to deliver the campaign and how will technologies help distribution of or interaction with the campaign?
- How will the design of the information you choose to present support the content?
- What will be the campaign's communicative power?

HANDOUT 5: FINDING DATA

Now that you know what open data are, what they're used for, and how they are licenced, you're ready to explore an open dataset. You can apply this to your own data once you're more familiar with how to do it.

Sourcing data

First, let's find some data. In this exercise, we'll look at homelessness data, but you can search for whatever is relevant to your organisation.

- Google 'data.gov.uk'
- Go to the search bar and type 'homelessness'
- Choose the first one, 'Homelessness statistics'
- Choose the 'Live tables' option
- We want the data in .csv or .xlsx format. Click on Download for the last three files.

Cleaning and filtering data

Let's take a look at the file and see if there's anything we can get rid of.

- Import the file into Google Sheets (or copy this file).
- Select the columns that aren't useful, right click and delete them.
- Download a clean .CSV via 'File > 'Download as' > 'CSV'

RESOURCE:
see an introduction to cleaning data here and to Open Refine here.

UKRI Arts and Humanities Research Council

Asking questions of the data

In groups, brainstorm some answers to the following questions, and pick the best question to report back:

- Does this tell us anything about causes of homelessness?
- How does homelessness affect women and men differently?
- What can these data tell us about the history of a homeless person?

Questions to ask of any data set include:

- What are your observations?
- Do you observe something out of the ordinary?
- Explore the available rows and make a note of anything that stands out or is interesting in the data.
- What's the most interesting question you want to ask the dataset you are looking at?
- Do you need any other datasets to answer this question?
- How could you get the other data you need to answer this question?

Combining data sets

If you would like to work through another exercise, follow these steps:

- Google "data.gov.uk"
- Go to the search bar and type "homelessness"
- Choose from the right hand lists "Homelessness prevention and relief (10 May 2014)"
- Choose from the right hand side "Homelessness from April 2018"
- Choose from the right hand list "show more"
- Download "Assessments - table A9 & Assessments table A10"
- We want the data in .csv format.

First things first, we want to make a nice clean dataset. Let's merge the files first.

- Go to Google, and ask for "combine csv online"
- Drag /select the files and place it in the appropriate boxes.
- Click on 'Merge files'
- Download the result

Let's Dig

- Open https://databasic.io/en/wtfcsv/
- Upload your file and click on 'Analyze'
- Click on the panels to see what data are included.

RESOURCES:

Find out more about working with pivot tables to analyse data here.For something a little more in depth, check out this visual data exploration tool.