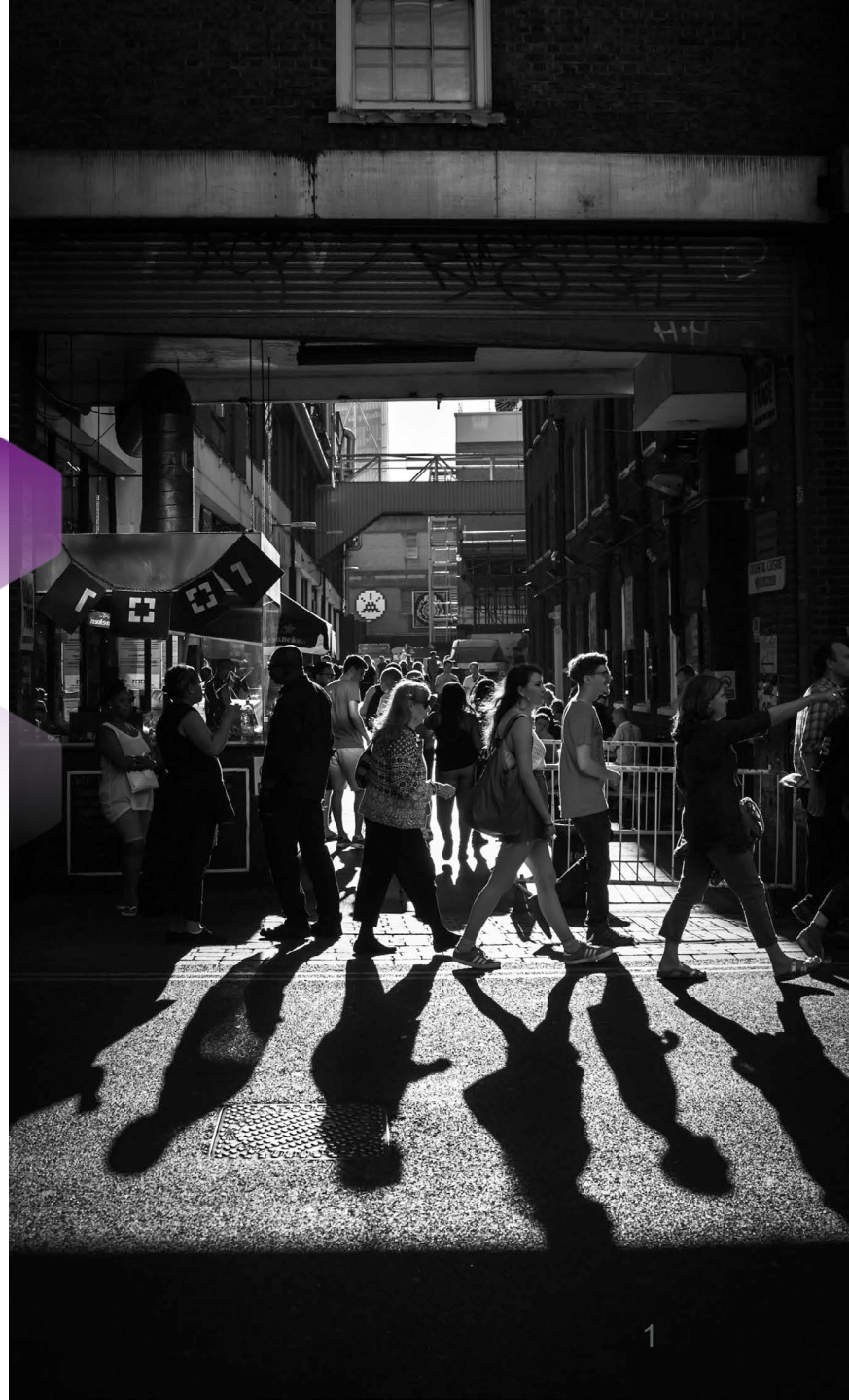
UK Data Service

Computational Social Science: An Introductory workshop

Dr. J. Kasmire

Research Fellow at Cathie Marsh Institute and UK Data Service



UKRI Economic and Social Research Council

Table of Contents for this workshop

Computational social science is…

The use of computational and empirical methods to address social science questions.

This requires:

- Human-thinking to identify important research questions,
- Computer-thinking to turn questions into computational/empirical methods,
- Human-thinking to effectively communicate the results.

Computational Social Science is NOT just…

- using computers within a social science research project,
- using digital versions of purely traditional social science methods, or
- using digital but purely non-empirical methods.

Let’s have some examples of CSS projects:

Collect, process, and analyse millions of online news articles to show changing political attitudes

Another CSS example project:

Use real-time weather and traffic data to show how travellers react

Yet another CSS example project:

Combine data from novel wearables/apps to establish correlation between social media activity and heart rate

And the final CSS example project:

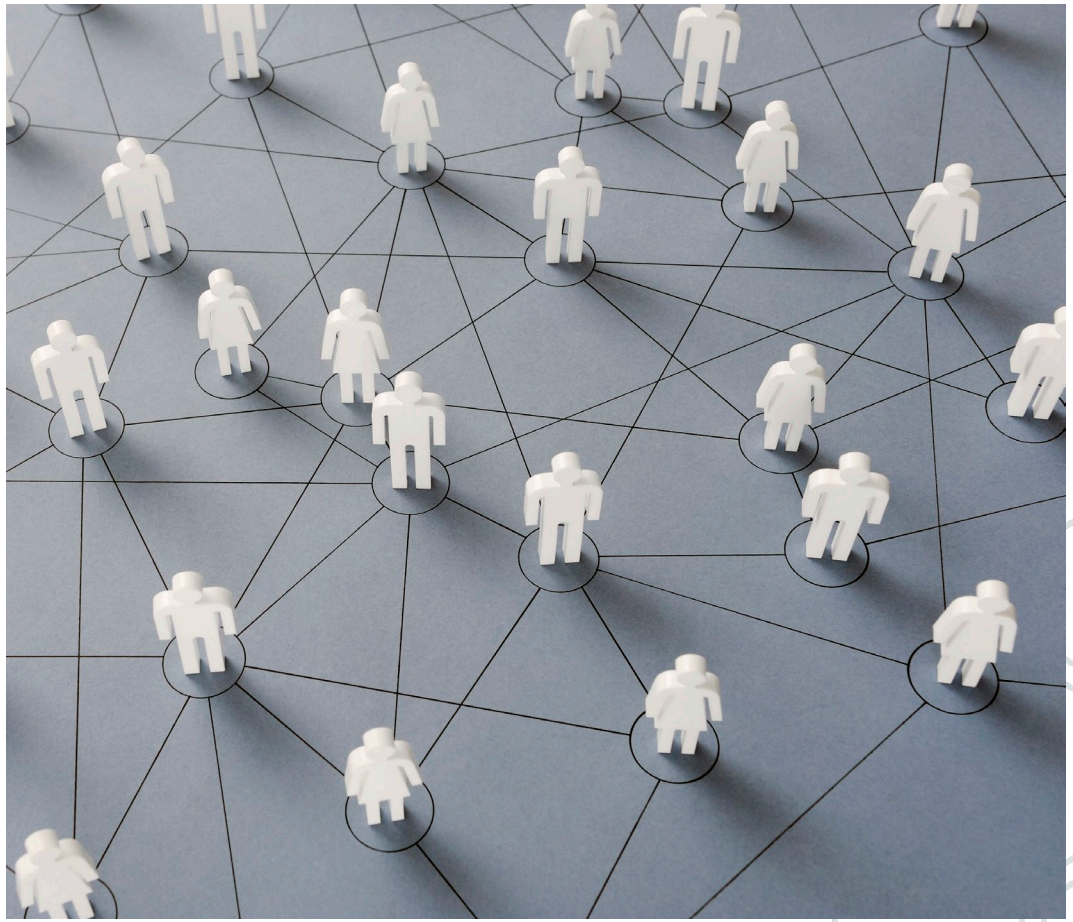
Import, process and format centuries of parish records to map family names over time

Key factors in CSS:

- Data volume, complexity, speed, difficulty or novelty is more important than exact data source/type.

- Data must pertain to people, actions, behaviours, choices, statements, etc.

- Exact research question is not important BUT must be a social science question.



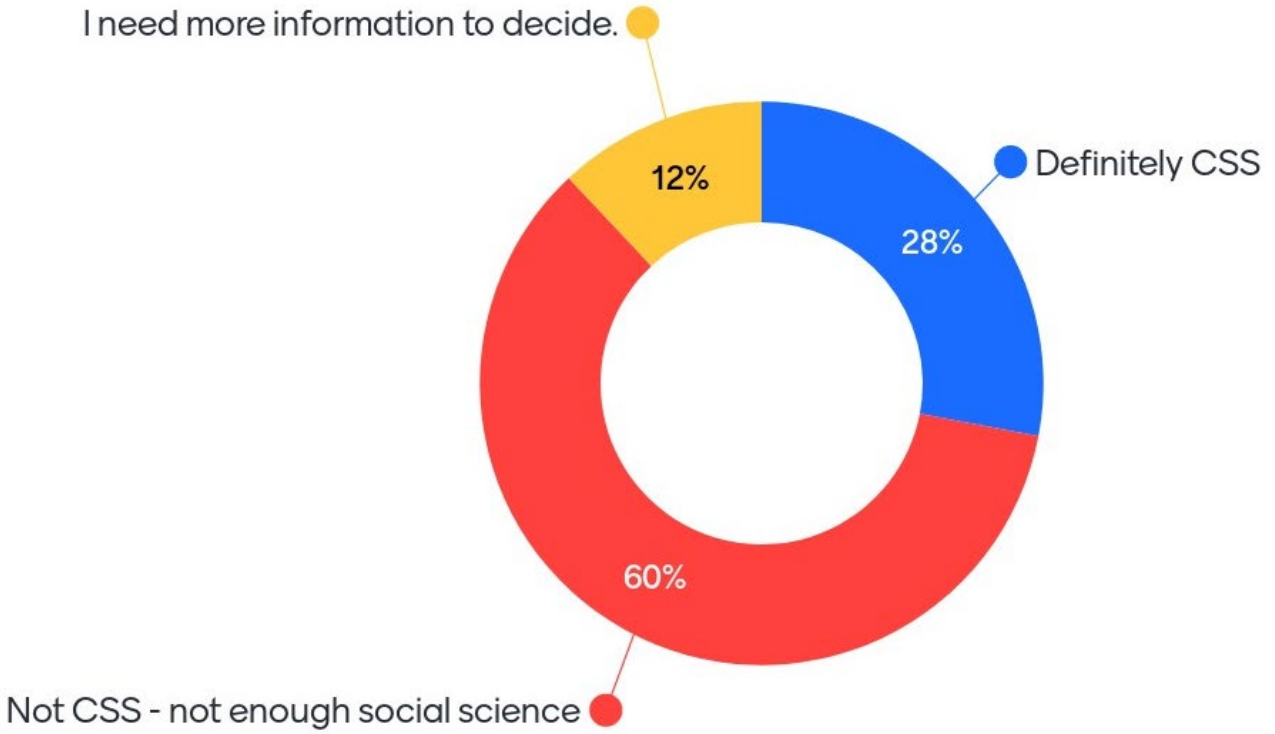
In essence, CSS is:

“an opportunity to do socially valuable research that would not be possible without computational methods and tools”
(Halford & Savage 2017)

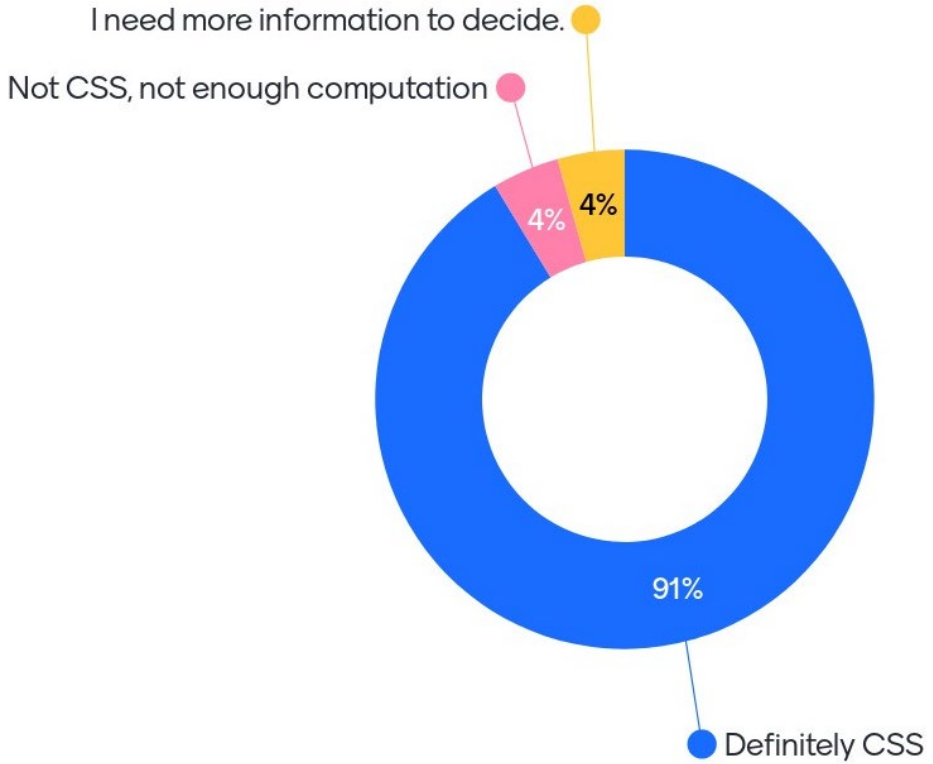
Interaction Time

The following slides give you a chance to vote on whether you think the described project is or is not an example of CSS!

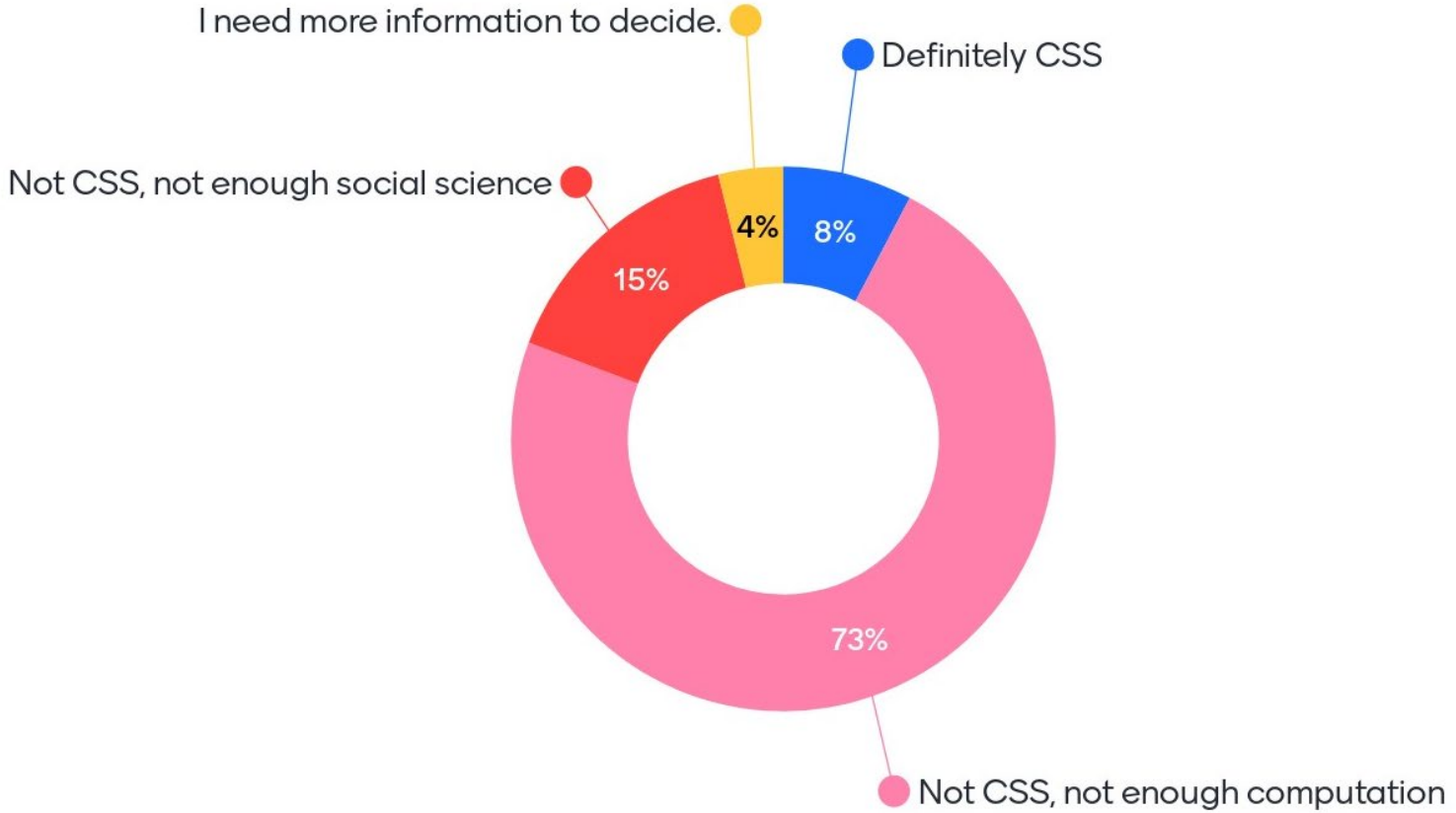
CSS or not?
Scan historic recipes and use AI algorithms to recognise text to identify ingredients and measures used over time.



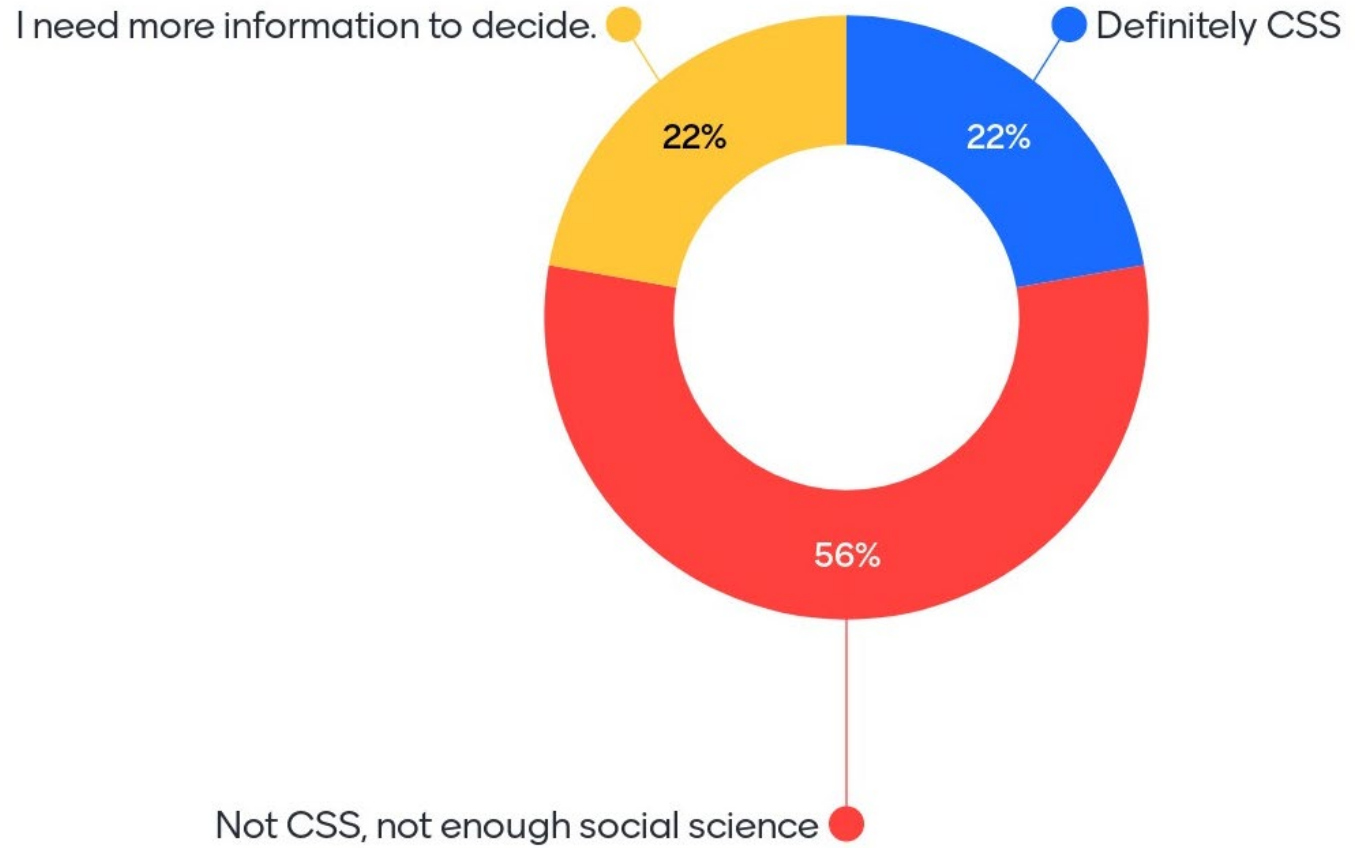
CSS or not?
Use gamified smart home displays to understand how people interact with energy saving technologies.



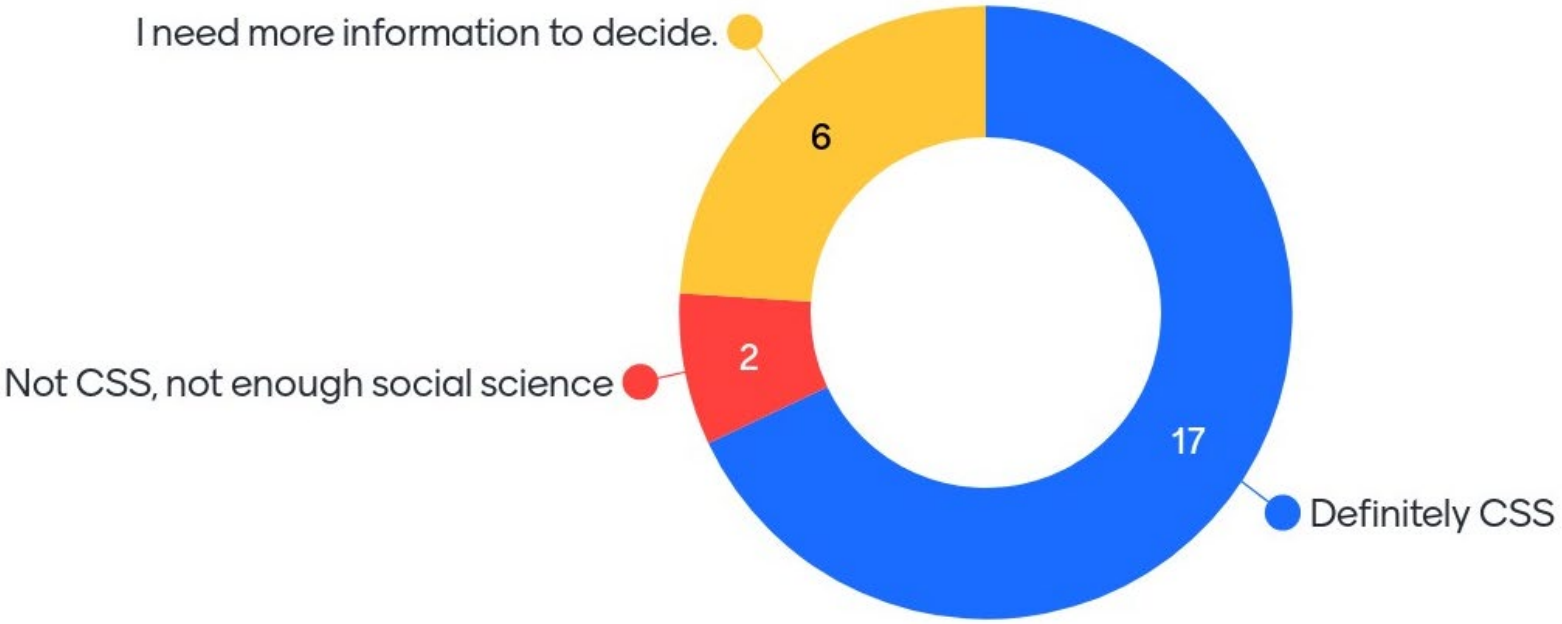
CSS or not?
Advertise for survey participation on social media and store the responses in a data base.



CSS or not?
Read in real time weather and air pollution data to create complex models of hyperlocal air quality.



CSS or not?
Train neural net on social media data to create a believable ‘chat bot’ that counteracts online radicalisation.



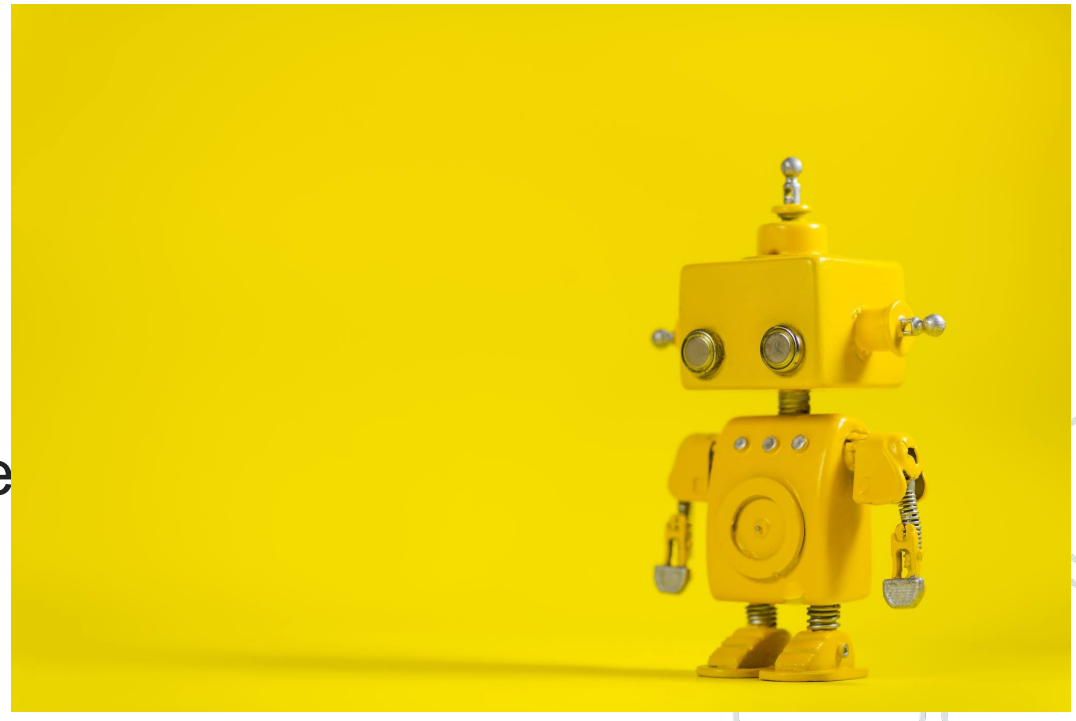
What have you learnt about CSS? a word cloud.

Social Scientists …

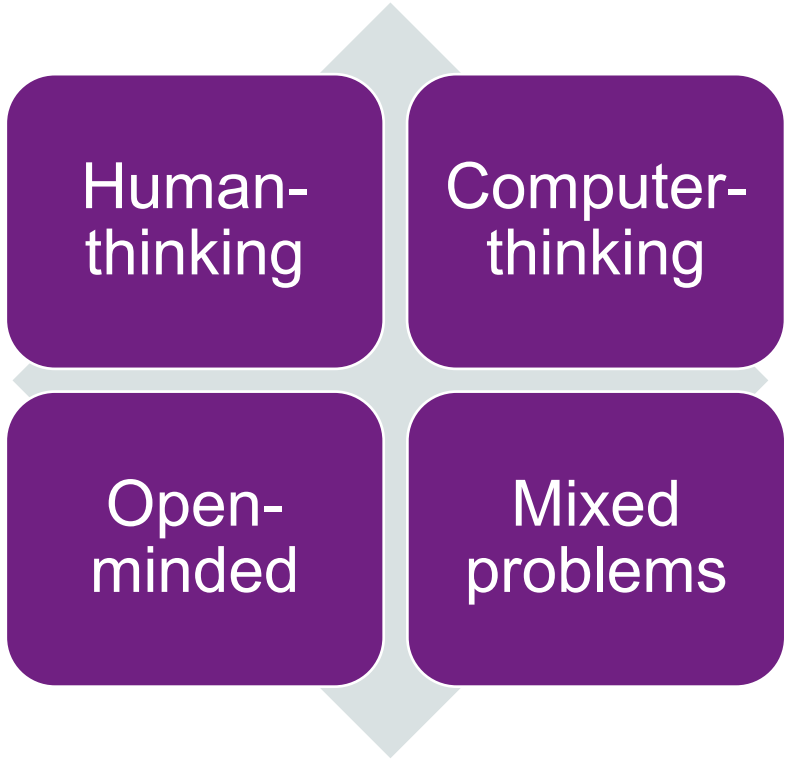
- Social scientists “think like people”
- Study people, interactions, behaviours, etc.
- Thinking skills = abstraction, inference, fuzzy categories, background knowledge, etc.
- Data skills = response categorisation/coding, quality evaluation, pattern detection, etc.
- Use computers, but do not usually write computer code

… and Computer Scientists

- Computer scientists “think like computers”
- Solve information/processing problems
- Thinking skills = concrete definitions, absolutes, strict hierarchies/categories, clearly defined and scoped variables/rules, etc.
- Data skills = Collect/analyse/manipulate data through code/tech/computational methods
- Not usually taught to identify/motivate research projects with societal impact/value



How to do CSS?


Human-thinking
Computer-thinking
Open-minded
Mixed problems

Human-thinking skills include:

- Identifying important problems or knowledge gaps,
- Considering possible solutions,
- Connecting problems to relevant theories or perspectives, and
- Collecting relevant information and research to frame approach.

Human-thinking is:

Easy(ish) for social scientists trained in abstraction, communication, subtle context, and shared societal knowledge.

Harder for computer/data scientists not trained in ill-defined, overlapping, context-dependent concepts or using assumptions/background knowledge for interpretation

Computer-thinking skills include:

Skills like:

- accessing, organising, processing and vast and/or complex data,
- writing (collaborative) code, and
- documenting workflows.

Computer-thinking is:

Easy(ish) for computer/data scientists trained in computational methods, strict rules, exclusive definitions, and extremely formal and structured processes (Jewett and Kling 1991)

Harder for social scientists, but they can build on training to code responses, format surveys, and draw statistical analyses from complex data (among others)

Open-minded (and eager to learn)

No one starts out with all of the skills they need.

No one knows all the skills they might need to acquire

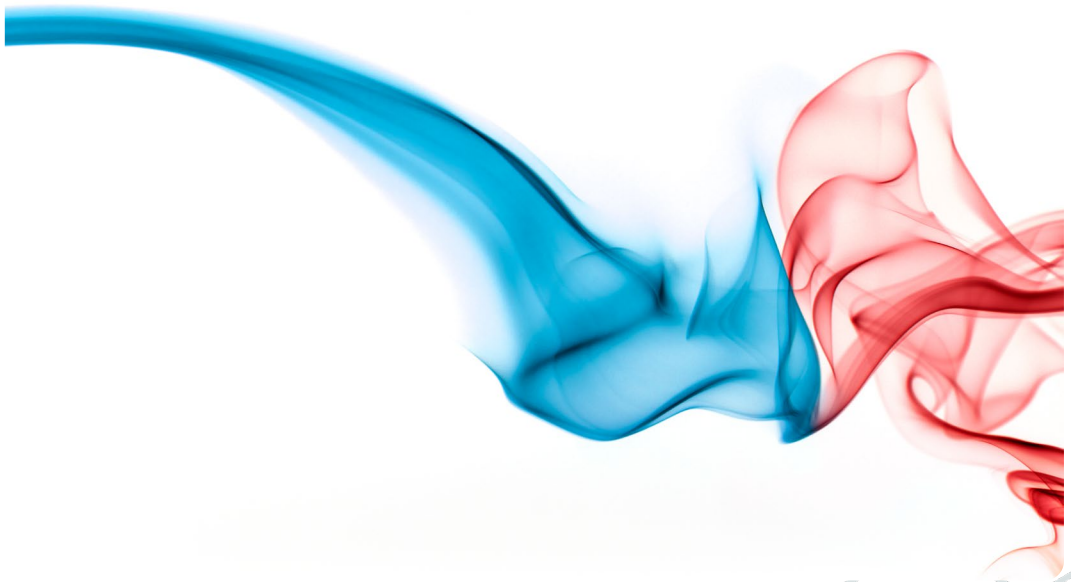
Approach with an open mind, curiosity, and a willingness to learn.

Some skills will be easier to pick up or use than others.

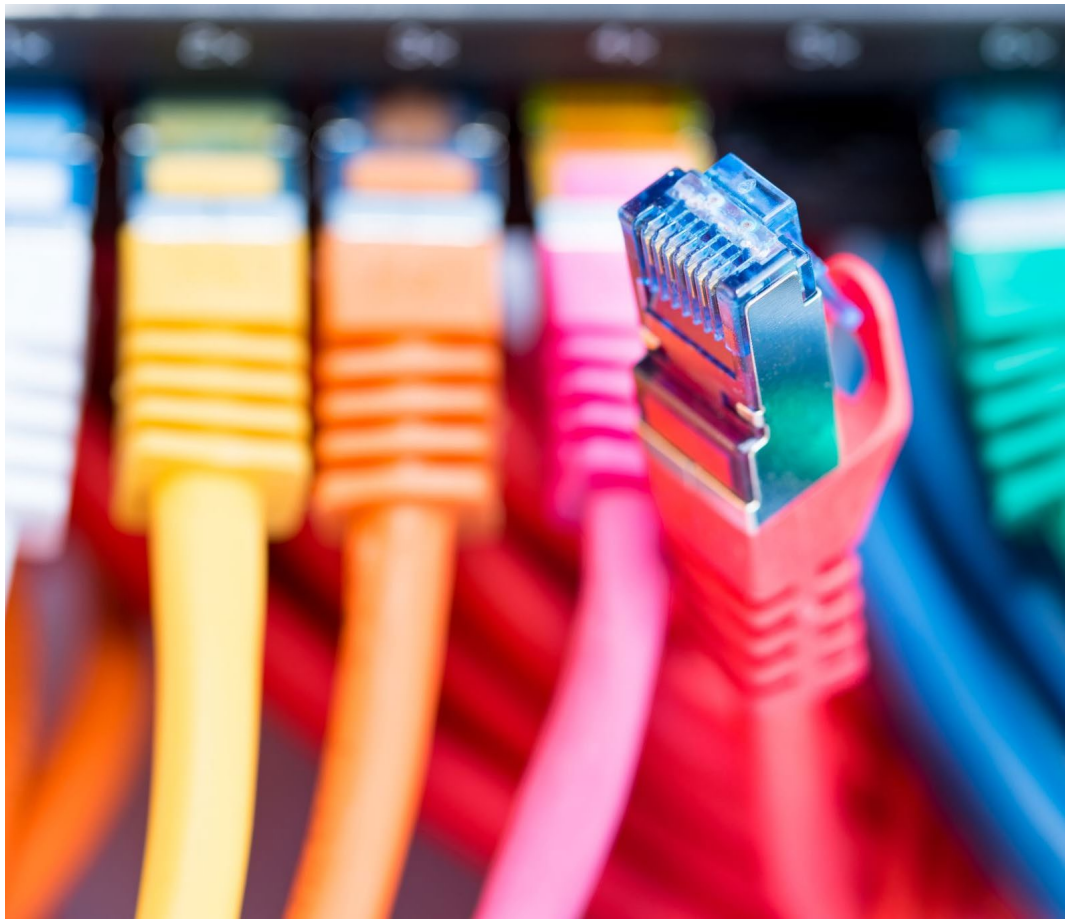
You can’t do it all yourself - be prepared to collaborate.

Mixed problems

Need human-thinking AND computer-thinking



Mixed problems (and the need to pick ‘em)



- Will become more important as:
- Resources are digitised
- Interactions, objects and processes become 'smart' or networked
- Large volumes of data are made available/are updated faster
- Other changes in the future

Practically, you can follow my 8-step process!

- Identify the problem
- Explore the problem
- Formalise the concepts
- Collect data, implement software, verify
- Experiment and analyse data
- Discussions and conclusions
- Communicate, publish, present
- Share, document and validate

1. Identify the problem

- Be as clear and specific as possible about the pattern, problem, lack of insight.

- Also identify who is involved, where it is, etc.

Essentially, brainstorming with reflection and editing.

Ideally with others.

WRITE IT DOWN! TRACK VERSIONS!

2. Explore the problem

- Gather information and perspectives in multiple ways (surveys, observations, secondary data analysis, app creation, web-scraping, API's, expert interviews, etc.).

- Spell out sub-problems, processes, relationships, simplicifations, assumptions, related issues, existing specialties, etc.

This edits/revises your brainstorm output.
WRITE DOWN WHAT YOU DID AND WHERE YOU GOT IT!

Participants on their (theoretical) step 1 & 2!

Literature review

I've scribbled a few ideas down and spoken to a retired professor in that subject area who recommended a couple of books. Ive started to read them.

Interested in Changing Labour Voteshares between the 2017 and 2019 General Elections. Extensive literature review allows identification of gaps in research, and what factors are already linked to changing voteshre.

asking a bunch of questions I had. literature review Need to do a systematic review with specific key words

spotting the gap via literature review; problematisation via 'engaged scholarship'; own career experience

Lit review- gap/contradictions

review all existing research on area of interest.

Step 1. I was interested in the relationship between people's awareness of a range of cultural objects (books, music, films, games) and their taste. I explored data initially to seek patterns and discussed it with the group I was working with.

Met with an academic advisors to develop a better understanding of the topic at hand. Built upon previous ideas and information available.

I did your step 1 without realising it would become research. I was chatting with my friend about racism in football + digital platofrms and thought this could be important research - now its my phd

scanning policy documents and speak to policymakers

Read recent articles to identify problems. Look into past research papers that already tackle related topics

3. Formalise the concepts

- Make all the concepts and processes explicit, formal and both computer and human understandable.

- Often known as ‘pseudo-code’.

- Example:
- “trust” is defined as a variable between 0 and 100.
- “trust” between two parties increases following mutually beneficial interactions.
- Existing levels of “trust” decrease to zero if an interaction is judged to be deceitful.
- Etc.

4. Collect data, implement software, verify

- Select and implement one or more methods.

- The choice of method will be highly dependent on the research topic.

- Thoroughly check that the selected method has been implemented correctly – essentially answering the question “Did we do the thing right?”

Participants on their (theoretical) step 3 & 4!

Used mind maps, flow charts, piloting survey and interview questions

I did a mini-project in a masters module - very brief exploration of content on Twitter at a specific moment (2020 football Euros)

Collect both qualitative and quantitative data to gain a deeper insight for my problem

Sounds like pilot study for step 3 and 4Pilot study ~ to see if operational definition of a variable is valid Could Also link to software: pilot study to see if software, does what it does

Rendering the data as a network graph to see how certain entities or concepts interact

Pencil n paper...not that far in current project.

choosing, reordering and manipulating sets of variables

5. Experiment and analyse data

- Run the experiments! Build the models! Analyse the data! Or otherwise use the methods selected in previous step!

- Identify and explain the results within the context of the experiments/model/method.

6. Discussions and conclusions

- Going beyond the experiment/model/method, draw some conclusions about what the results mean.

- Do you support policy recommendations?

- Who or what do these results affect? Why does it matter?

- What should change? Who benefits from that proposed change?

Participants on their (theoretical) step 5 & 6!

Used NVivo to organise and synthesise primary qualitative data; used Reflexive Thematic Analysis to analyse the pqd. Sadly, no machine learning!

longitudinal studies to make conclusion about a your problem

I suppose my main conclusion from the mini/pilot study was that more exhaustive research needed to be done online surrounding different types of content shared on Twitter during football tournaments

Mostly spent time modelling the data using a very long decision tree, analysed using descriptive statistics

Present relevant results and interpretation to study participants.

7. Communicate, publish, present

- All of the previous steps must be communicated to multiple audiences in multiple ways.

- Short term and long term engagement.

- Public, academic, political, students, etc.

- Consider conferences, journals, blogs, white papers, academic societies, workshops or university classes, etc.

8. Share, document, validate

- Help make sure the ‘right thing was done’ by allowing your work to be studied, reproduced and/or modified as needed through openly available:
- Workflows (methodologies/steps taken)
- Code
- Data

- As transparent, well documented and openly as.
- This comes from the writing down and version tracking.

Important to Note

- These steps are NOT LINEAR!
- Most (or all) will require many ITERATIONS.
- Documentation (step 8) actually applies THROUGHOUT all the other steps – don’t wait until the end to start!

Participants on their (theoretical) step 7 & 9!

open science framework

Reports and synthesis of work undertaken. Not personally active in research, but report on the research of others

all documents change controlled by version and date

Present relevant results and interpretation to study participants.

Using GitHub for step 8

Papaja + Github

Sharing your code or even just ideas can come at a cost if you are an ethnic minority, female or immigrant researcher working at a less p university - people with more authority sometimes 'borrow' your code, ideas without any attribution.

ReShare

I like poster presentations - I did a couple at undergrad and like visual resources. I'm also writing a blog post right now, which is actually relevant to current events (world cup in Qatar) so good timing!

Conferences are useful for novice researchers like me. Helps me condense my paper into a concise presentation and gain insights from academicians I otherwise would not have access to

Website (blog), contributing to academic papers, GitHub, social media, posters, research group discussions, reading groups, WhatsApp group

References

Arnold, B., L. Bowler, S. Gibson, P. Herterich, R. Higman, A. Krystalli, A. Morley, M. O'Reilly, K. Whitaker and others (2019). The Turing Way: a handbook for reproducible data science (Version v0.0.4).

Brooker, P. D. (2019). Programming with Python for Social Scientists, SAGE Publications Limited.

Halford, S. and M. Savage (2017). "Speaking Sociologically with Big Data: Symphonic. Social Science and the Future for Big Data Research." Sociology **51**(6): 1132--1148.

Heiberger, R. H. and J. R. Riebling (2016). "Installing computational social science: Facing the challenges of new information and communication technologies in social science." Methodological Innovations **9**.

Jewett, T. and R. Kling (1991). "The dynamics of computerization in a social science research team: A case study of infrastructure, strategies, and skills." Social Science Computer Review. **9**(2): 246—275.

Kasmire, J. and D. McDonnell (2020). Thinking like a computational social Scientist: Organising thoughts and organising data. Big Surv '20 Big Data meets Survey Science.

Take away points.

Another word cloud!

UK Data Service

Thank you.

Dr. J. Kasmire

Email j.kasmire@manchester.ac.uk

@JKasmireComplex on Twitter

@JKasmireComplex@mastodonapp.uk

@UKDataService on Twitter