Being a Computational Social Scientist

Dr. J. Kasmire
UK Data Service

Dr Diarmuid McDonnell
UK Data Service

UK Data Service Webinar
12 May 2020

New Forms of Data Training Series

Upcoming webinars:

- Introduction to Text-Mining (02 June 2020)
- Text-Mining: Basic Processes (16 June 2020)
- Text-Mining: Advanced Options (29 June 2020)

Upcoming coding demonstrations:

- Collecting data I: web-scraping (13 May 2020)
- Collecting data II: APIs (20 May 2020)

Past webinars:

- Web-scraping for Social Science Research: APIs as a Source of Data
- Web-scraping for Social Science Research: Websites as a Source of Data

UK Data Service

Table of Contents

UK Data Service

What is computational social science?

Isn't most social science already dependent on/supported by computers?

Computational social science takes us beyond traditional activities due to:

- Big data
- Writing code
- Applying analytical methods from other disciplines

More formally, *computational social science* is an interdisciplinary branch of research, defined more by its methods and data than its substantive topics (Heiberger & Riebling, 2016).

UK Data Service

Why be a computational social scientist?

Computational social science is an opportunity.

Provides methodologies and tools for working with data (Halford & Savage, 2017):

- Process and analyse large-scale, unstructured data (e.g., text corpora)
- Capture data generated/published in real-time (e.g., smart home sensors)
- Access information on new/previously unmeasured activities (e.g., online-only behaviours)
- Access information on familiar/currently measured activities at an unprecedented scale, dynamism or complexity (e.g., social networks)
- Empirical social science is having its moment in the sun (move over social theorists…)

How do you be a computational social scientist?

Firstly, be a social scientist!

Social scientists possess knowledge - theoretical and empirical - of social systems and phenomena, and already have advanced data skills, especially around:

- Categorising and coding responses (qualitative and/or quantitative)
- Evaluating data quality (e.g., why is this survey response missing?)
- Making inferences from data (e.g., how representative is this pattern?)

It’s just the computational side you need to develop!

UK Data Service

How do you be a computational social scientist?

There is a lot to it, but *computational social science* typically involves the following practices:

- Writing programming scripts to collect and manipulate data.
- Employing analytical techniques - many derived from computer/information sciences - to reveal patterns in data.
- Using technological tools and e-Research best practice to structure and document your research workflow.

Focus on the skills and knowledge needed, or our *Big Five for Computational Social Science.*

Big Five for Computational Social Science

1. **Thinking computationally**
2. **Writing code**
3. **Knowing your computational environment**
4. **Understanding and manipulating data**
5. **Documenting and enhancing your workflow**

UK Data Service

Big 5 of CSS: #1

Thinking computationally

UK Data Service

What are human problems?

- Staying alive
  - Responding to danger
  - Curiosity about the unknown
  - Recognising patterns

Sub-rational thinking is just thinking...

What else are human problems?

- Other people
  - Understanding intentions
  - Predicting likely responses
  - Collaboration and/or competition

UK Data Service

Human thinking about human problems?

- Working memory capacity according to Kimberg, 1997 :
  - Limited focus (perhaps only one item)
  - Complexity matters
  - Different capacity for structured vs. unstructured
  - Chunking (and others) as a work-around strategy
- Bounded rationality according to Simon, 1997:
  - Satisficing
  - Heuristics

Human thinking about human problems?

- Communication assumptions according to Pinker, 2003 :
  - Speaker and listener interpret the same way
  - Objects and actions as distinct
  - Mid-level categories, but with knowledge of hierarchies
  - Transitive properties
- 2 types of thinking according to Kahneman, 2011:
  - Fast, error-prone, intuition-based
  - Slow, (more) accurate, rationality-based

Computer thinking for human problems…

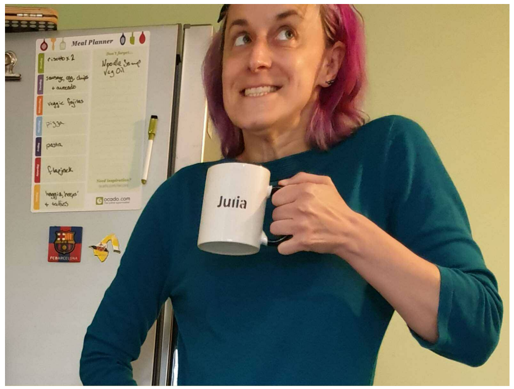
- Especially human problems that humans are not good at:
  - More working memory needed (high volume or complexity)
  - Best option more important than speed
  - Reciprocal communication not assumed

Example problem =

UK Data Service

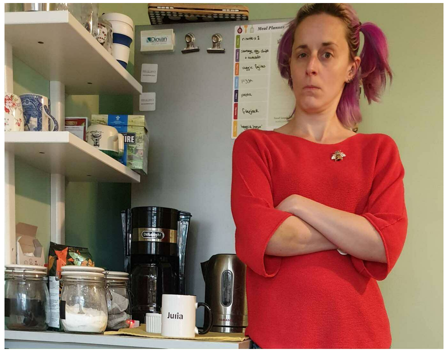
A human tries to tell a human how to make a cup of tea.

- Human 1 says:
  - Put the kettle on, put a tea bag in a mug.
- Human 1 says:
  - Pour just boiled water into the mug.
- Human 1 says:
  - Remove tea bag, add a splash of milk
- Human 1 says:
  - Enjoy!

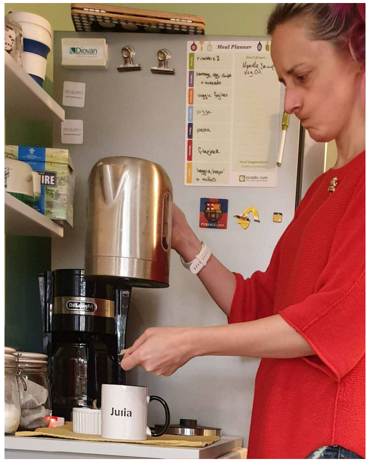


UK Data Service

A human tries to tell a computer how to make a cup of tea.



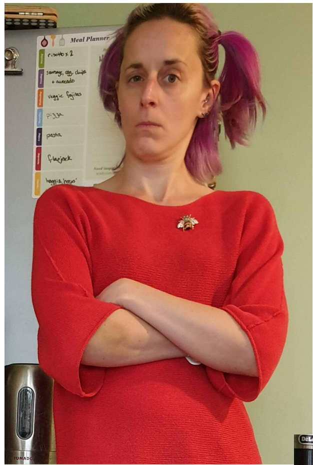
- Human 1 says:
  - Put the kettle on, put a tea bag in a mug.
- Human 1 says:
  - Right. There is an electric kettle right there. That is the kettle. And that, right there, is the mug. And that, right there, is the tea bag. OK?!?
- Human 1 says:
  - Put the kettle on, put the tea bag in the mug.
- Human 1 says:
  - Gah! Put the kettle on its base, turn it on.

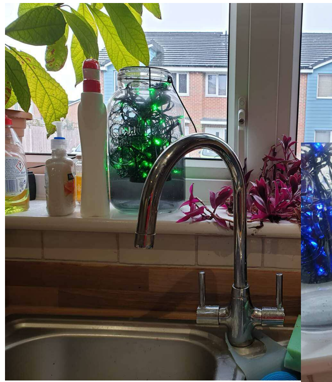


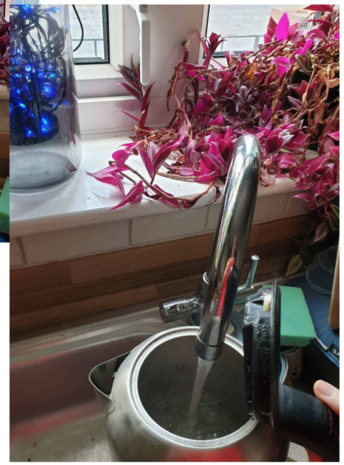
UK Data Service

A human tries to tell a computer how to make a cup of tea.

- Human 1 says:
  - Put 250 ml of water in the kettle. Return it to its base. Turn it on.
- Human 1 says:
  - There is a tap. Put 250 ml of water from the tap into the kettle. Return kettle to its base. Turn it on.
- Human 1 says:
  - Put one tea bag into a mug.
- Human 1 says:
  - When the kettle indicates the boil cycle has finished, pour water into the mug.
- Human 1 says:
  - Gah! Start again, but pour water from the kettle into the mug.







UK Data Service

Humans and computers... Different.

Humans	Vs.	Computers
Abstract concepts		Concrete definitions
Inference		Defined terms/rules only
Shared/background knowledge		Nothing carried over from other scopes
Fuzzy categories		Strict categories
Context-dependence		Absolutes
Expertise accumulated over time		Equations, formuae, mathematical models
Citation lists		Charts, graphs, visualisations
Mental models		Detailed methods sections
Debates and discussions		Data sets, code repositories, etc.

UK Data Service

Some problems need both…

- Human thinking needed to identify the problem, possible solutions, relevant info, etc.
- Computer thinking needed to work accurately and reproducibly with large volumes of (complex) data.

- More mixed problems arise as
  - Resources become digital
  - Social media/Internet of Things/etc. grows more prevalent
  - Real-time data becomes available
  - Etc.

Computational Social Science example

- Travel to Work – before and after pandemic lockdown
  - Research questions like “How does lockdown change work commutes?
    - Right now?
    - In 3 months? 6 months?
    - A year? 3 years?
  - Knowledge of resources
    - UK Data Service census data (including 2011 travel to work data, household composition, etc.)
    - GIS data (maps, transport systems, etc.)
    - Apple and/or google maps open data
    - Upcoming 2021 census
  - Acquire data
  - Work with data
  - Compare complex patterns of behaviour derived from large volumes of data with spatial and temporal structure
  - Communicate results

UK Data Service

Big 5 of CSS: #2, #3, and #4

Writing code

Knowing your computational environment

Understanding and manipulating data

Big 5 of CSS: #5

Documenting and enhancing your workflow

UK Data Service

Documenting your workflow

Increasingly, people in the scientific community are demanding greater transparency and reproducibility of research. Put simply:

"*Reproducible research is necessary to ensure that scientific work can be trusted*." (The Turing Way Community, 2019)

Reproducibility can be summarised as the availability of data and code to fully rerun an analysis.

UK Data Service

Documenting your workflow

Reproducible: the same analysis steps performed on the same dataset consistently produces the same answer.

Replicable: the same analysis performed on different datasets produces qualitatively similar answers.

Robust: the same dataset is subjected to different analysis workflows to answer the same research question and a qualitatively similar or identical answer is produced.

Generalisable: Combining replicable and robust findings allow us to form generalisable results. (The Turing Way Community, 2019)

UK Data Service

Documenting your workflow

Five Simple New Rules of the Sociological Method (Gayle, 2017):

1. **Tell us about your software.**
2. **Tells us about your data.**
3. **Show us how you got your data ready.**
4. **Show us all the analysis you did.**
5. **Save all of this work openly.**

UK Data Service

Enhancing your workflow

There are a growing number of technological tools and approaches for supporting your workflow:

- **Writing code** – Jupyter notebooks + a programming language (Python, R, Julia)
- **Knowing your computational environment** – Anaconda distribution of Python; `conda` and `pip` package management tools; Binder
- **Understanding and manipulating data** – open data, open-source packages (`pandas` in Python, `tidyverse` in R)
- **Documenting and enhancing your workflow** – Jupyter notebooks, Github

Summary and links to further reading

- Bibliography

  - Kimberg, D. Y., et al. (1997). "Effects of bromocriptine on human subjects depend on working memory capacity." Neuroreport **8**(16): 3581-3585.
  - Simon, H. A. (1997). Models of bounded rationality: Empirically grounded economic reason, MIT press.
  - Pinker, S. (2003). The language instinct: How the mind creates language, Penguin UK.
  - Kahneman, D. (2011). Thinking, fast and slow, Macmillan.

UK Data Service

Questions

Dr. J. Kasmire

julia.kasmire@manchester.ac.uk

@JKasmireComplex

twitch.tv/jkasmirecomplex

Dr. Diarmuid McDonnell

diarmuid.mcdonnell@manchester.ac.uk

@DiarmuidMc

Further resources and help

Repository: https://github.com/UKDataServiceOpen/new-forms-of-data

Youtube: https://www.youtube.com/user/UKDATASERVICE

Help: ukdataservice.ac.uk/help/

Subscribe to UK Data Service news at https://www.jiscmail.ac.uk

@UKDataService

UKDataService

UK Data Service