Doing 'what works' in teaching with data

Jen Buckley, Vanessa Higgins and Alle Bloom

UK Data Service and University of Manchester

Some context

- Recognised need for data skills leading to teaching initiatives.
- Growing discussions on pedagogy.
- Data as a hook, important for engaging students (Nind, 2020).
- UK Data Service has a long history supporting the use of data in teaching and learning, …
- but can we better support the use of data in teaching?

Higher Education Teaching Consultation 2021

Survey

- Non-probability sample with 109 valid responses.
- Varied disciplines and career stages but most were lecturers or above and sociology and politics most common disciplines.
- Nearly half had substantial experience of quantitative data and methods and the teaching of quantitative methods.
- Topics: use of data in teaching, teaching materials, statistics software, dissertation students, and reproducibility and CSS.
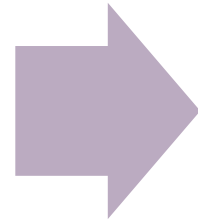
Follow-up interviews

- Three interviewers carried out 16 interviews online via Zoom.
- Open questions on their experience and teaching practices and evaluation of existing teaching resources produced by the UK Data Service.

What data is being used in teaching and why?

- Lecturers use a wide variety of data
  - 60 percent reported using data from the UK Data Service, but 50 preferred datasets named.
- Survey data most popular.
- But also, data…
  - from the Police ('students love that')
  - a RCT (which the lecturer had to "beg and borrow" from a colleague
  - on pupil preferences for Justin Bieber (may have been fake data)
  - that was synthesized.

Varied preferences and experiences of selecting data
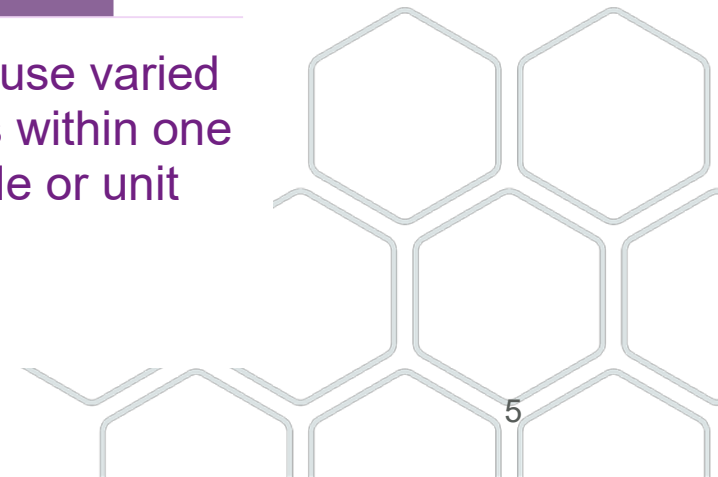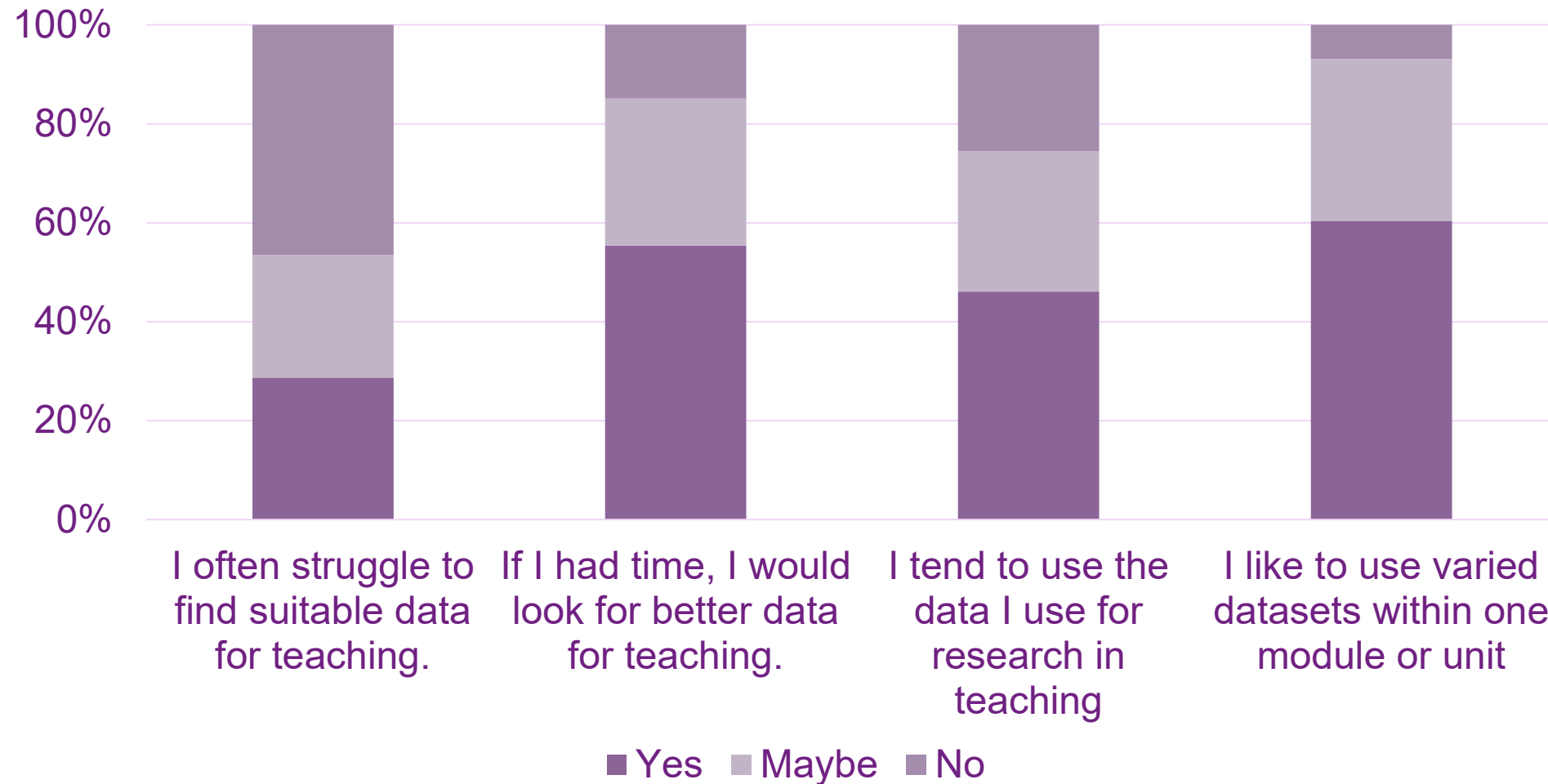
Six common factors influencing the choice of data

Real world data and research

The (perceived) interests of students

Size and complexity of the dataset

Suitability for teaching methods

Familiarity

Data access

Progression from simple to more complicated

Something that is small enough so that it's easily understood. Size relates to number of variables not observations. Datasets cut down to 10-20.

'Consistently trying to use datasets that people find manageable and intuitive… a little bit of data manipulation but not too much. Not as easy as pressing a button and getting the answer'.

Students are taught the 'how to' and then 'set loose' on real data.

For first years, data is uber clean, it does exactly what they need it to do. Second year is somewhere in between – longer list, clean it a bit. Third year – this is how horrible it can be to access data, this is how you would go about cleaning.

Six common factors influencing the choice of data (2)

Real world data and research

The (perceived) interests of students

Size and complexity of the dataset

Suitability for teaching methods

Familiarity

Data access

When using data for teaching, how often have the following applied?

|	Never	Sometimes	Usually
I make no prior changes to the data, students see the full data	28%	56%	16%
I prepare data to make datasets more suitable (e.g. such as reducing the number of variables, setting missing values and recoding variables.	12%	43%	46%
I use teaching datasets from the UK Data Service	47%	34%	20%

n=101
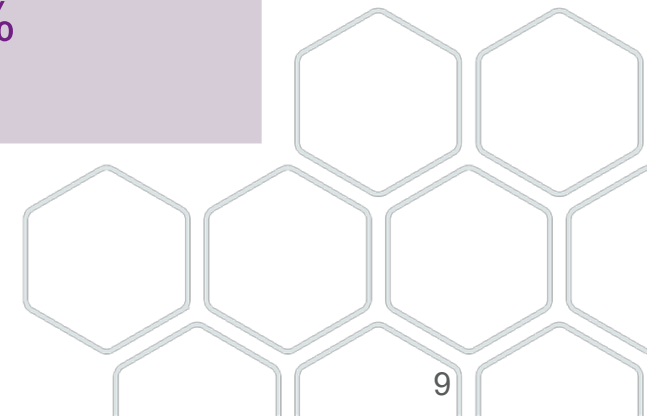
Suggested principles for teaching datasets

- Should use real data with manageable issues.

- Can be very small and simple (for example up to 20 (or 30 variables) or something larger and more complex (up to 100 variables). Due to variations in lecturer preferences, tend towards including some additional variables to offer choice.

- Be intuitive and understandable (through easy-to-understand variables and accessible documentation that provides well-formatted information about variables including level of measurement).

- Include multiple continuous variables that can be dependent and independent variables, even if they need to be derived using additive scores or factor analysis.

- Be current and connect to student interests (topical or linked to core social science themes).

How much do you use the following software for teaching?
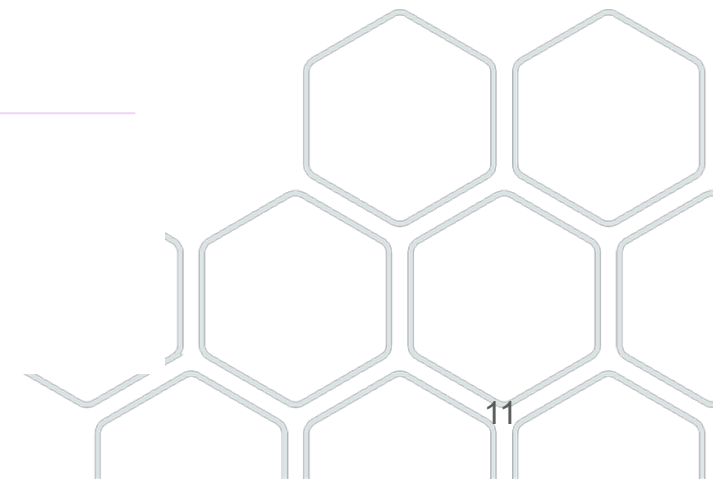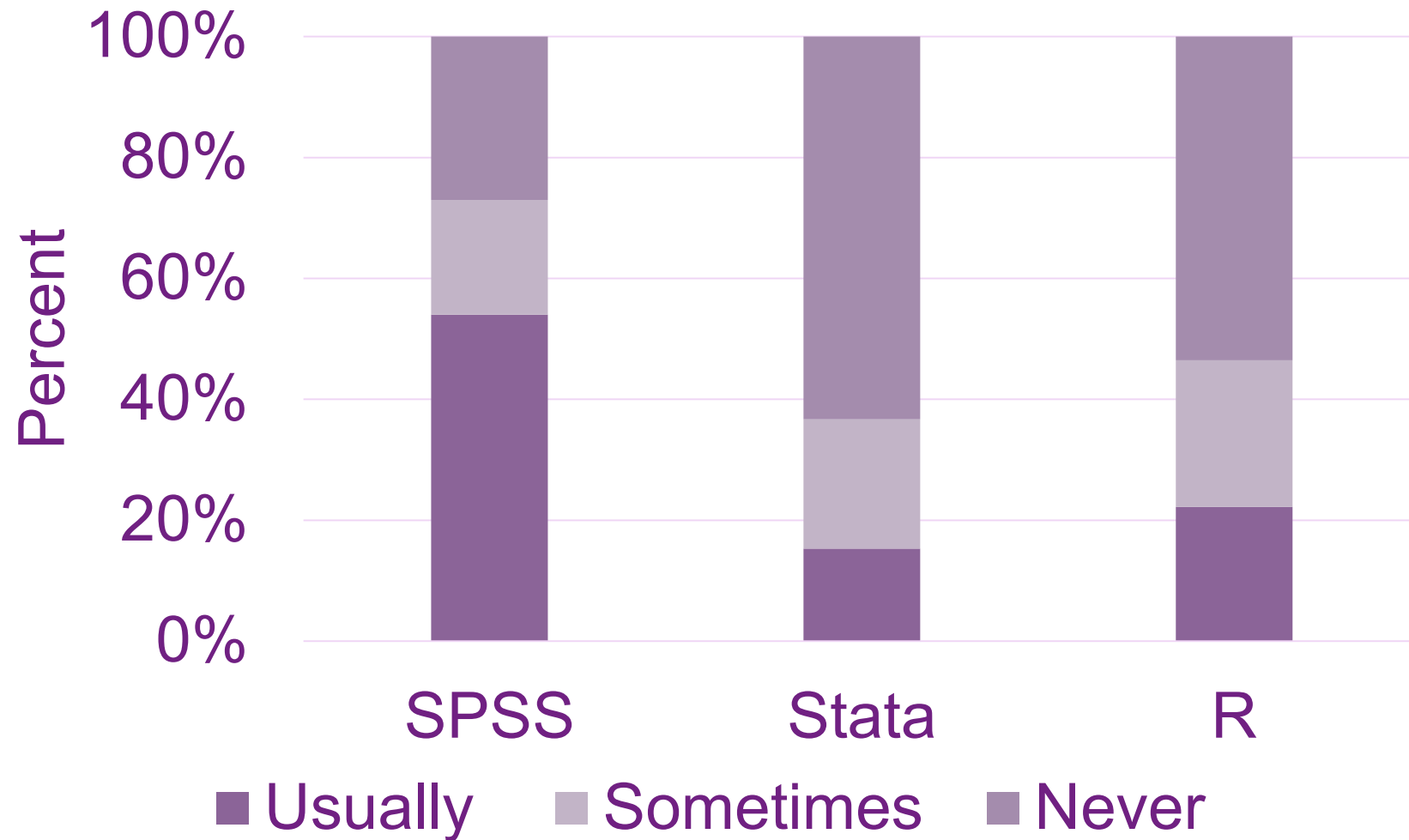
Programming: a fork in the road

Alternative perspectives

- SPSS more intuitive and programming potentially alienating.
- Not enough time to teach programming.
- 'A fetishization of programming languages and a focus on advanced techniques at the expense of basic statistical literacy'.

Advocates for R and programming

- Open source better.
- Pedagogic benefits of programming.
- Employability.
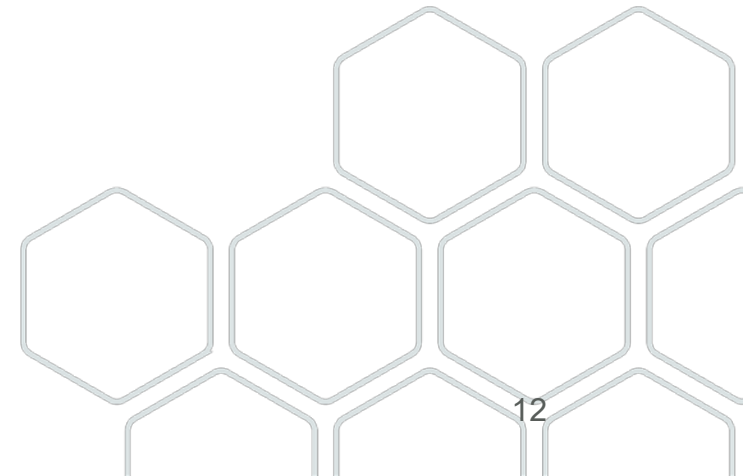
Conclusion

- Varied data is used but data is not simply a hook ready to be picked up.
- Time, knowledge and skills involved.
- Some common factors influencing choices.
- Potential benefits of 'teaching data' and sharing knowledge.
- But context (teaching time, time to prepare) influences views and practices.
- And bigger questions around curriculum.

Thank you.

Jen Buckley

jennifer.buckley@manchester.ac.uk