Exploring sex, sexual orientation and gender identity through interactive visualisations

Louise Capener, Nadia Kennar, J. Kasmire, Alle Bloom
UK Data Service, University of Manchester
Presentation outline

• Background
• Visualisations
  • scatterplots
  • maps
• Summary.
Previous research: Sexual orientation

• Sexual identity data is commonly collected across surveys e.g. Natsal, Understanding Society and Health Survey.

• However, these lack detail at a lower geographical level and often have limited samples.


• However, these statistics are considered ‘unofficial’ and were created to assess methods more than produce results.
Previous research: Gender

• Gender identity data much less widely collected (ONS 2020):
  • Trial of questions in Crime Survey for England and Wales, Active Lives Children and Young People survey, Taking Part Survey.
  • More widely collected in admin data, but still not extensive.

• International research includes the Canadian Census 2021 and Census’ from Nepal (2011), India (2011) and Pakistan (2017), which all included a ‘third gender’ option.
Census 2021

• Which of the following best describes your sexual orientation?
  • Straight/heterosexual.
  • Gay or Lesbian.
  • Bisexual.
  • Other sexual orientation, write in.

• Is the gender you identify with the same as your sex registered at birth?
  • “Yes” and “No, write in gender identity”.

• Questions were voluntary, varied slightly between England and Wales.

• Allowed respondents to request a confidential individual form, which overrode the information on a household form.
Why make interactive visualisations?

• The 2021 census data provides the information at a level of geography that allows local authority areas to be visualised. This is a level of detail not available with other surveys.

• The data therefore provides a unique opportunity to display this data at this detail.

• ONS created maps to display the initial data, but these are single-variate.

• The aim with our visualisations is to display the data in a way that allows users more control over what they can explore, allowing them to examine their own assumptions about their local area.
What features influence non-response rates?

- Datasets: Gender identity and sexual orientation.
- Area type: Lower tier Local Authorities.
- Scatterplots explored 2 variables:
  1. Main language.
  2. Religion.

- Programming language: Python.
- Integrated development environment (IDE): Jupyter Notebook
- Hosted on: GitHub.
- Packages: Bokeh, pandas.

- Platform for user interaction: Bokeh server.
- Hosted on: Heroku.
Interactive mapping

Datasets: Gender identity and sexual orientation.

Area Type: Local Authority.

Visualisation: Using two R packages to highlight the distribution of these categories in urban/rural areas, providing an interactive tool for exploratory analysis.

- Programming language: R.
- Hosted on: GitHub.
- Packages: Leaflet, tmap, ggplot2.
- Platform for user interaction: Shiny.
- Hosted on: Shiny.io.

- Let’s dive in!
References


Thank you.

Louise Capener and Nadia Kennar

Email:
Louise.capener@manchester.ac.uk
Nadia.kennar@manchester.ac.uk

Twitter:
@CapenerLouise
@NadiaKennen

Want to check out our code?

GitHub:
UKDataServiceOpen/InteractiveGender