University of Essex

Data Visualisation Overview

Akitaka Matsuo
Essex IADS

Data Visualisation

University of Essex

- Why we visualise data?
  - Overall, we have a story to tell from the data
    - Strength of the association
      - Movement in time-series
      - Relation between two variables
    - Existence of outliers

Things to consider

University of Essex

- Types of variables to visualize?
- How many variables to be involved?

- You can find a lot of galleries online
  - https://datavizcatalogue.com/
    - basic but very helpful to make you familiarise the types of visualisation
  - https://www.informationisbeautifulawards.com/showcase
    - Award winning visualisation. Good to get some inspirations

Types of data

University of Essex

- Numeric variables
  - Continuous
  - Discrete
- Ordinal
- Categorical variables

- Other
  - Network data

Number of variables involved

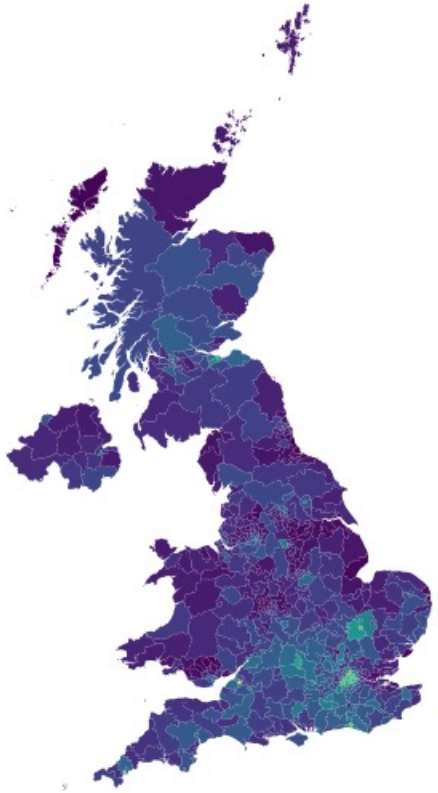
University of Essex

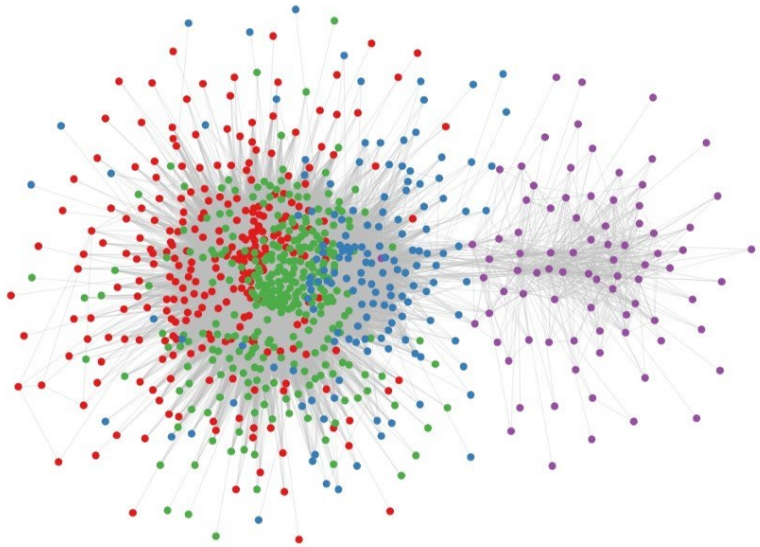
- One variable
  - Histogram
  - Density
  - Proportion
    - Bar, pie
- Two variables
  - Scatter plot
  - Heat map
- More than two
  - Scatter plot with additional information (colors, shapes)
  - Bubble chart

Other types of plots (examples)

University of Essex

- Geospace plot
  - Choleopleth map
  - Dot map
- Network plot
- Word cloud





Python Graphic Packages

University of Essex

- Matplotlib
  - Provide basic API for graphics
  - Long history
  - Not super well suited for data science pipeline
- Seaborn
  - Work really well with Pandas DataFrames
  - Reasonably beautiful figures without customisation
  - Calling Matplotlib functions behind the scene
- Pandas

- And more
  - e.g. plotly

Matplotlib

University of Essex

- Started as a project for bringing MATLAB-ish graphic capability in Python
  - Predates Pandas
- Provides flexible/powerful API
  - A number of add-on toolkits for visualization (e.g. seaborn)
- In Jupyter
  - Use option of
    - `%matplotlib inline`
  - It's recommended to run entire graphic in one coding cell

Seaborn

University of Essex

- A very popular visualisation package
- Working smoothly with Pandas DataFrame
- A number of visualisation options that does that require less coding

Documentation/gallery

University of Essex

- Both packages have excellent gallery
  - https://matplotlib.org/gallery.html
  - https://seaborn.pydata.org/examples/index.html