Lab 2: Data Processing and Visualisation

Week 3

January 2026

1 Introduction

Over the next two labs, we will work with a dataset based on social media posts (tweets) by members of UK parliament on Twitter (now X). These data correspond to the years 2017-2019, a period of Brexit negotiation and leading up to the 2019 general election. The goal of this lab is to visualise patterns and trends in these data.

1.1 Inspecting data files

- Download the data file (mp_tweet_counts_month.csv) from GitHub
- Open the file in Excel or other spreadsheet editor. What columns/variables are present in the dataset? What kind of distribution do you expect (e.g. categorical, continuous)? What units of analysis are available in these data?
- Import it into Power BI and create a new report:
  - Click 'Create', then 'CSV', then 'Upload file' and upload the file (mp_tweet_counts_month.csv), then click 'Next'
  - Inspect the data in preview; familiarise yourself with the format
  - Click 'Next', then 'Create a report'. Choose any file name e.g. 'css-lab2' and click 'Create'

1.2 Explore the data Power BI and interpret results

Creat the following figures. Refer back to the Lab 1 instructions and consult the Power BI tutorials if you are unsure.

- A bar chart showing the total number of tweets by MPs from each party. Which party's MPs tweeted the most? Does this correspond to the party share in parliament at the time?

- A line plot showing the number of tweets per party for each month. Are there are many interesting trends in tweet frequency?
- A line plot showing the number of Brexit-related tweets among the two main parties (Conservative and Labour) over time. Which party tweeted about Brexit more?
- A line plot showing the average number of tweets by MPs of each party over time. Is the average the same or different to the total?
- A scatterplot showing the relation between the number of retweets and the number of Brexit-related tweets.
  - Were MPs who tweeted about Brexit more likely to be retweeted often?
  - Why might this pattern by spurious?

2 Extensions

Plots showing distributions require tools that are not available in all versions of Power BI. If you are using Power BI in your browser, these will be more challenging.

- A histogram showing the distribution of the total number of monthly tweets per MP. This requires first binning the data, which can be done easily on the desktop version of Power BI.
- Now add party to make 'small multiples' of the distribution of tweet counts by party. Do you see any differences between the distributions (beyond the fact that there are fewer tweets by MPs from smaller parties)?
- Try using a log transformation on the Y axis to clarify any differences in the distribution.

3 Compare with Excel

Try making the same plots as above with Microsoft Excel (or LibreOffice). What differences do you observe compared with Power BI? Which tool do you find easier to use, or more powerful?