Problem Set 1

Data Visualisation for Social Scientists

Due: January 28, 2026

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on GitHub.

- This problem set is due before 23:59 on Wednesday January 28, 2026. No late assignments will be accepted.

Roll Call Votes in the European Parliament

Data Manipulation

First, you need to download data from the first European Parliament, including information on each MEP from EP1 and how they voted in each recorded roll-call vote during EP1.

1. Load these datasets into your global environment:

    - `mep_info_26Jul11.xls` (includes MEP characteristics from EP1–EP5, you need "Sheet = EP1")
    - `rcv_ep1.txt` (EP1 roll-call votes)

2. Briefly describe (2–3 sentences each) the unit of analysis and key variables in each of these two datasets.

3. The `rcv_ep1` data are in a wide format, with V1, V2, ..., Vn as separate vote columns.

- Identify which columns are ID/metadata (*MEPID, MEPNAME, MS, NP, EPG*) and which columns are vote decisions ($V_1 \ldots V_n$). Tidy the voting data such that each row/observation is a single vote for a single MEP.

- Create a summary table of counts of decision categories (e.g. Yes/No/Abstain/Present but did not vote/Absent) across all votes.

4. Construct a new dataset that combines MEP-level information with their vote decisions from EP1 in long format (from part 3). Check for missingness.

5. Compute, for each EP group in EP1:

   - The mean rate of Yes votes (Yes over Yes+No+Abstain) across all roll calls.

   - The mean abstention rate.

   - The mean vote preferences along the two contested dimensions (NOM-D1 and NOM-D2).

Data Visualization

1. Plot the distribution of the first NOMINATE dimension by EP group, and explain any trends you see.

2. Make a scatterplot of *nomdim1* (x-axis) and *nomdim2* (y-axis), with one point per MEP and color by EP group.

3. Produce a boxplot of the proportion voting *Yes* for each vote by EP group to visualize cohesion.

4. Display the proportion voting *Yes* across votes by national party using a bar plot.