Group Project for Computational Social Science

Deadline: 27 March 12:00 **SHARP** Late submissions will not be accepted

Use a real-world dataset to answer a clear question about society. Turn a research question grounded in social science into a measurable outcome, apply appropriate computational methods on the dataset, communicate results clearly with evidence, collaborate responsibly and effectively.

Methods: You may use any methodology or tool, but we can only provide support for the tools introduced in the labs. Explain your methods clearly, provide sufficient detail, and justify your choices. Acknowledge the limitations and ethical implications of your approach.

Outcome: Be specific on what you test and measure. Provide enough evidence on what you measure verifies your claims. You may add visualizations, tables, infographics as evidence. Report your results based on what you claim and what you observe.

Deliverables:

- **Abstract:** a one-page abstract that states your research question(s), describes the data, the planned methods, summarises the expected outcomes, and states the division of labor. This is for instructor feedback; it is required but not marked. **Deadline: 27 February**

- **Report:** A report of up to 8 pages (figures and tables included; references excluded), structured as follows: Introduction (with Background and Related Work either as a separate section or integrated), Data, Methods, Results, Discussion (including reflection on results, decisions, and implications), Limitations, Ethics Statement, GenAI Disclosure, and a Contribution statement that clearly states the contribution of each student. **The provided LaTeX template must be used.**

  A detailed literature review is not required. An appendix may be included for supporting material that cannot fit in the main report (e.g., oversized figures or implementation details).

- **Supplementary materials:** a short set of materials showing evidence of work and reproducibility, e.g., KNIME workflow screenshots, Power BI dashboard/report screenshots, and (if used) shareable AI conversation links or exported chat transcripts.

Data: You **must** use one instructor-provided dataset shared in the following Github repo: `https://github.com/TomBidewell/Understanding_Society_with_Big_Data_CSS_Course_Data`

- **Climate Change:** global/regional temperature anomalies (NASA GISTEMP v4); global sea-level change (1993–present); natural disasters (2004–2025); Google Trends search interest; TV news snippets mentioning climate change (BBC 2017–2020; CNN/MSNBC/Fox 2009–2020); multilingual climate-related headlines (sampled, 2015–2020); and anonymised Twitter retweet networks (selected weeks in 2019) plus tweet metadata with sentiment/topic/stance (sampled).

- **Housing (Scotland):** cleaned Reddit discussions on rent from r/Edinburgh (posts/comments plus node/edge lists for discussion networks, collected Summer 2025); private-sector rental price statistics (2010–2024) by property size and Broad Rental Market Areas; and transcripts of Scottish Parliament discussions related to rental prices (Jan 2010–May 2025).

- **Political Polarisation:** UK MP Twitter interaction/retweet networks with early vs. late one-month windows and a tweet corpus sample; posts by American politicians on Truth Social with post-level attributes and emotion/hate/fear-related indicators; and a cross-national dataset of ideological polarisation in Western Europe (Dalton index; 20 countries; 1945–2023).

Instructions: Use this checklist to review your group project report before submission. Ensure all items are addressed to meet the assessment criteria.

1. Research Question & CSS Framing

☐ Do you state your research question clearly and concretely?

☐ Do you motivate the research question?

☐ Do you specify your population?

☐ Do you link the question to a social-science theory or a concept?

☐ Do you define how each abstract concept is measured and justify your choice?

☐ Do you specify the outcomes you will measure?

2. Data

☐ Do you name the dataset(s), the source, and the time window?

☐ Do you describe the data quantitatively and qualitatively?

☐ Do you report the sample, sampling method, and size?

☐ Do you explain and justify the preprocessing steps?

3. Methods & Results & Discussion

☐ Do you specify and justify your methods and what outcomes they estimate *before* showing the results?

☐ Are your methods appropriate for the data you employ?

☐ Do you validate or evaluate your methods and results?

☐ Do you communicate your results with appropriate visualizations showing the evidence (plots, tables, infographics)?

☐ Does each figure/table have a clear caption that explains what the reader should learn?

☐ Do you interpret the reported outcomes clearly and tie them back to the research question?

☐ Are your claims consistent with your methods and results?

☐ Do you discuss and reflect on your work, your decisions, the implications of your results and future work?

4. Limitations

☐ Do you acknowledge your limitations and state what you cannot conclude?

☐ Do you address the representativeness of your sample?

☐ Do you discuss your methodological assumptions and measurement error?

5. Ethics & Responsible Practice

☐ How could your analysis violate privacy? How do you mitigate it?

☐ How could your analysis cause harm? How do you mitigate it?

☐ Are there other ethical risks and implications?

☐ Did you include the **mandatory** GenAI disclosure stating how you used ChatGPT/GenAI tools?

6. Writing & Formatting

☐ Did you cite your sources for motivation, social science framing, data, methods, and discussion?

☐ Is the report structured with clear headings (Introduction, Data, Methods, Results, Discussion, Limitations, Ethics Statement)?

☐ Are figures readable (axis labels, units, legends) and referenced in the text?

☐ Is the report within the page limit?

Criterion (Weight)	Fail (0–39%)	Pass (40–49%)	Good (50–59%)	Very Good (60–69%)	Excellent (70–79%)	Outstanding (80–100%)
Research Question & Social Science Framing (15%)	Missing or unclear question.	Question is vague, weakly motivated, or not grounded in social science.	Clear question but partial theoretical grounding or unclear outcome	Clear question & outcome explicitly linked to a social-science theory or concept.	Clear, well-defined, focused, justified, and theoretically grounded question & outcome	Exceptionally precise, original, and theoretically integrated question & outcome.
Data Understanding & Preparation (10%)	Inappropriate dataset or not described.	Appropriate and described dataset, description is poor.	Dataset is described with basic preprocessing but not justified.	Dataset is clearly described, preprocessing steps are explained and justified.	Well-justified data description & preprocessing with attention to quality & representativeness.	Deep, critical understanding of data with extensive, quality preprocessing steps.
Methods (15%)	Methods inappropriate or incorrect.	Basic methods with weak justification.	Appropriate methods that are described but not justified; no validation or evaluation.	Appropriate & justified methods with some form of validation.	Well-justified methods tied back to the research questions with thorough validation	Sophisticated, well-validated, insightful and creative methodological choices.
Evidence, Results, Verification, Figures, Tables & Supplementary Material (35%)	Claims unsupported, inconsistent with results, no supp. material, no GenAI disclosure with respect to results	Results weakly supported by evidence, partial supp. material. Figures & Tables unreadable, without captions, not referenced	Results are presented with figures or tables, but evidence is thin or not clearly tied to the research question. Minor issues with figures & tables.	Results are clearly presented with well-chosen evidence, figures & tables that directly addresses the research question. Some verification of findings is attempted.	Strong results, clearly communicated with clean figures & tables, supported by sufficient evidence and appropriate verification.	Exceptional, original, and rigorously verified results, communicated with clarity and depth, and supported by comprehensive and persuasive evidence, creative figures & tables.
Limitations & Ethics (15%)	Absent or superficial discussion.	Some issues noted with little depth.	Key issues discussed only briefly, the impacts are not considered.	Key issues are discussed with reference to how they affect the conclusions & some mitigations.	Thoughtful discussion of limitations and ethical risks with clear mitigation strategies.	Nuanced, reflective, creative and responsible limitation & ethical analysis.
Reflection (10%)	No meaningful reflection on decisions, trade-offs, or learning, no GenAI disclosure on writing	Superficial or generic reflection that lacks engagement with the research process.	Some reflection on challenges or possible improvements, but discussion is descriptive rather than analytical.	Reflection engages with specific methodological decisions and some description of trade-offs.	Thoughtful reflection on trade-offs, improvements, limitations of choices, and lessons learned from the research process.	Deep & critical reflection demonstrating intellectual growth, methodological maturity, & strong awareness of research decisions & implications.