Assignments for GEI1002

<u>Individual Assignment (40%)</u>

Write a 1000-word essay (excluding references) where you analyze an article that uses data to examine a social or cultural phenomenon.

Which article to use?

Choose an article that was published after January 2022, either from the news or academia. Describe the article's source and ensure its credibility. If you use a blog, make sure it is reputable, not a tutorial, and the blog author is also the data analyst. Include the article's reference in your essay. If you found the article in an aggregator, like ResearchGate, cite the original source instead.

How to analyze the article?

To analyze the article, explain how the data´s context and definitions influenced the study. Describe how the data was obtained and processed, and give your personal evaluation on whether the data was adequately used. In this analysis, you must reference at least one of these readings (some of which will be discussed in tutorials):

- "Sheared sheep" from *Living in Data* by Jer Thorp.
- *The functional art* by Alberto Cairo
- *Weapons of math destruction* by Cathy O'Neil
- "Ways of seeing data: Toward a Critical Literacy for Data Visualizations as Research Objects and Research Devices" by Jonathan Gray et al.

Discuss the chosen reading's ideas and how they apply to your analysis. Include at least one quote from the reading with a proper citation in any familiar format. Describe the quote's context, showing your understanding of the reading, and avoiding generic claims.

What format to use?

Submit your essay in PDF format with your name and matrix number at the beginning. Do not use headers and footers to repeat your name on each page.

Rubrics for individual essay

Criteria	Pts
Relevance of the chosen article The chosen article should effectively demonstrate the use of data to study a social or cultural phenomenon, and you should provide a clear description of the article's source and credibility.	4 pts
Understanding and application of readings You should show a thorough understanding of the chosen reading(s) and effectively apply the ideas to the context of their article analysis.	4 pts
Use of evidence and citation You should use at least one quote from the chosen reading, properly cited, and describe the quote's context while demonstrating their understanding of the reading. The selected article should also be properly referenced.	4 pts
Analysis of data´s context and definitions	10 pts

You should provide a well-reasoned explanation of how the data context and definitions have shaped the analysis in the chosen article.	
Evaluation and critical thinking You should effectively assess the rationale behind the definitions used by the authors of the article, the usage of the data, and the persuasiveness of the analysis. You should demonstrate your ability to think critically about the chosen article and its findings.	10 pts
Structure and organization The essay should be well-organized, with clear transitions between paragraphs and a logical flow of ideas. The introduction should provide an overview of the essay's purpose, and the conclusion should summarize your analysis and evaluation.	4 pts
Grammar, spelling, and formatting The essay should be written in clear and concise language, free of grammatical and spelling errors. It should adhere to the formatting guidelines provided (PDF format, no headers or footers, and including the student's name and matrix number). The essay should be within the specified word limit (max 1000 words, excluding references).	4 pts

Group Project (35%)

Group members

Regarding team formation, choose your own members (4-5 people per team), all of whom must be from the same tutorial group as you.

Description

For this project, you will need to provide a quantitative description of a social or cultural phenomenon, preferably from Singapore or Southeast Asia, using data visualizations. You can use an existing dataset, which you'll need to clean and organize, or create your own dataset from scratch.

You can approach your analysis in one of two ways:

1. Text classification of individual sentences
2. Network analysis

If you choose text classification, make sure to manually check some results and provide inter-coder reliability metrics, as discussed in class.

In both cases, your analysis can focus on either:

1. Change over time
2. Comparison between groups

You must explain how you defined the phenomenon you are studying. For this, please describe alternative definitions you could have used, and justify your choice. Remember, your main goal is to provide a quantitative description. Please avoid giving causal explanations (why things are happening) in your write-up, as this is beyond the scope of the class.

Submit your project as a ZIP file (one per team) containing:

1. The data you used, along with the manual categories provided by each coder (if applicable).

2. The Jupyter notebooks you've used.
3. A project report (in PDF format) including the literature review and your analysis (2000 to 3000 words, excluding references). For the literature review, summarize at least six existing research papers. State if there's any quantitative research in the field and how your work is different. If there isn't any, explain the unique aspects of your data-based approach. If only a few papers use data, the rest can be from traditional research. Provide a detailed explanation of how you obtained, cleaned, or created your data. Describe what insights can be gained from your quantitative analysis, but avoid causal explanations.
4. A separate PDF file with the names and matrix numbers of all group members, along with a description of each person's contribution. Normally, all students will receive the same grade, unless there is a significant difference in the workload distribution.
5. Grading criteria for group project

Criteria	Pts
Clarity and Definition of the Phenomenon Clearly explain and define the social or cultural phenomenon you have chosen for your project. Discuss alternative definitions and justify why you chose the one you did. This helps to set the context for your analysis and to ensure that you have a well-defined research area.	7 pts
Literature Review Provide a comprehensive and well-structured literature review that summarizes at least six existing research papers relevant to your project. Discuss any quantitative research in the field, how your work is different, and clarify the unique aspects of your approach. This demonstrates your understanding of the current research landscape and how your project fits into it.	7 pts
Data Collection and Processing Explain in detail the process of obtaining, cleaning, or creating your dataset. This includes documenting any challenges faced, justifying decisions made, and describing the meticulousness of your data processing. Proper data management is important for accurate and reliable analysis. *If you don't include the dataset and Jupyter Notebooks, you will be given 0 points for this rubric.*	7 pts
Quantitative Analysis and Visualizations Show your ability to conduct a quantitative description using either text classification or network analysis approaches. This should include well-crafted data visualizations to support your findings and demonstrate insights gained from your quantitative analysis. Your analysis should be coherent, clear, and accurate.	7 pts
Structure and organization of the report The report should be well-organized, with clear transitions between paragraphs and a logical flow of ideas. The introduction should provide an overview of the research purpose, and the conclusion should summarize your analysis.	4 pts
Format of the submission Make sure you submit the files in the correct formats (PDF for the project report and the group member´s contribution, Jupyter notebooks for the code and spreadsheets for the data). Don´t exceed the word limit.	3 pts