Lab 1: Introduction to Computational Social Science Tools

January 2026

1 Introduction

Welcome to your first lab! In Computational Social Science (CSS), we rely on a diverse toolkit to collect, analyze, and visualize social data. Today, we will set up the core tools you will use throughout the semester:

- **PowerBI:** For rapid analysis and visualization.

- **KNIME:** For building reproducible data processing workflows that include advanced analysis and machine learning.

- **Gephi:** For Social Network Analysis.

- **ELM:** For having unlimited ChatGPT access with privacy.

- **NotebookLM:** For basic text analysis and any other tasks you need assistance with.

2 Setting Up

2.1 Data

To learn these tools, we need data. While you can use any dataset, we recommend using the **Climate Change News** dataset to follow some of the specific interpretation steps in this guide.

Download the file `climate_related_news_in_english.csv` from this GitHub repository.

2.2 PowerBI

PowerBI is an industry-standard tool for "Business Intelligence". For us, it is an excellent way to quick way to analyze a dataset by understanding its *descriptive statistics* (e.g., "What is the distribution of news sources?") or do visualizations.
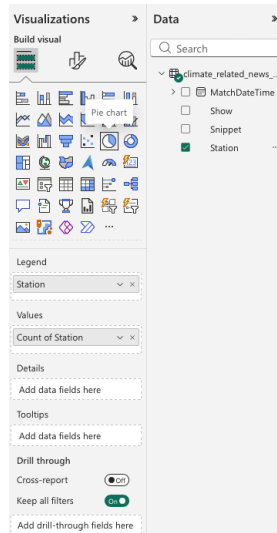
Figure 1: Caption

1. **Access:** Go to `https://app.powerbi.com`. Sign in with your school email (no installation or registration required).

2. **Import:** You should see a prompt: "Add data to start building a report."
   Click **Get Data** → **Upload file** → choose `climate_related_news_in_english.csv`.
   Note: If this fails for any reason, try again.

3. **Preview:** Once uploaded, you will get a preview of the data. PowerBI detects data types and delimiters automatically (you may need to manually configure this for "dirty" data, but the defaults work here). Click **Next**.

4. **Create:** Click **Create Report**. You can name it "Lab 1 Test".

5. **Visualize:** You will be greeted by a blank screen prompting you to "Build Visuals with your data."

   - On the right, under the **Visualizations** tab, click the **Pie Chart** icon.
   - In the **Data** tab (rightmost side), expand `climate_related_news.csv` and check the box for **Station**.
   - Drag and drop "Station" into the **Values** field in the visualization pane.

   Result: You have created a visualization showing the proportion of news coverage by station, a key metric in media analysis.

2.3 KNIME

In CSS, *reproducibility* is critical. If you clean your data in Excel by hand, no one can replicate your steps. KNIME allows you to build a visual "workflow" where every step (filtering, cleaning, analyzing) is a visible "node." For our course, it is a coding-free approach to do advanced data analysis. Thus, it is good both for teaching and professional use.

1. **Download:** `https://www.knime.com/downloads`. You do not need to enter your information; just tick "I have read and agree to the terms..." and click Download.

2. **Install:** Follow the standard installation prompts (by clicking "Next" every time...). (Windows tutorial: `https://www.youtube.com/watch?v=bBO8oJnow7U`).

3. **Workflow:** Open KNIME and create a new workflow (name it "Lab 1").

4. **Read Data:** Drag and drop your CSV file onto the blank canvas. A **CSV Reader** node will appear.

5. **Execute:** Click the node, then click the **Execute** button (the yellow play arrow below).

6. **Inspect:** Inspect the table below. Look at the Snippet column. What do you see?

7. **Explore:** This is it for today's exercise but you can try creating a Pie Chart using this tool to explore further. See how KNIME handles visualization differently from PowerBI.

2.4 Gephi

Gephi is the leading open-source software for visualizing networks and doing simple analysis. In this class, we use it to study relationships (edges) between social actors (nodes).

There are two versions of Gephi, the standard desktop version and the browser version named Gephi Lite. We recommend **Gephi Lite** as it does not require installation.

1. Go to `https://lite.gephi.org/`.

2. You will be prompted to open a file. Select the sample dataset **Les Misérables**.

3. This graph represents the characters in the famous novel Les Misérables. Characters are *nodes*, and their interactions are *edges*.

4. Play with the network as you like. We will cover more in the upcoming labs.

Gephi Lite currently only supports GEXF, GRAPHML or Graphology JSON. To convert network datasets in csv format to GEXF, you can use `https://medialab.github.io/table2net/`. This is an open-source tool that works locally in your browser.

3 Data Exploration with AI

A good first step for any data analysis to get a summary of the data to understand what it contains. This can be a quantitative analysis (e.g., looking at the means of values it contain) or qualitative (e.g., looking at the themes the text data cover). We will do the latter in this lab using AI (specifically the "Large Language Models"). We will generate mind maps to extract prominent themes hidden within large volumes of plain text.

3.1 Prerequisites

We will use the `climate_related_news_in_english.csv` dataset again. However, Generative AI tools have strict word count limits. ELM accepts approximately 240,000 words, while Google NotebookLM allows 500,000. Thus, you must split large data files into smaller segments before uploading.

You can split files using a tool like ConvertCSV. This website processes files locally in your browser and states that it does not retain data; however, we recommend against uploading sensitive or private data to any third-party tools.

1. Enter https://convertcsv.com/text-split.htm

2. Click "Choose file" tab

3. Click "Choose file" button and select "`climate_related_news_in_english.csv`"

4. Choose "# of Files" (instead of # of Lines Per File). Set this 20 for this dataset.

5. Click "split file and download"

6. Wait for the processing to finish; you will receive a ZIP file containing 20 parts of your document.

7. You will get a zip file that includes 20 parts of your document with .txt extension.

Note: The resulting files will have a .txt extension. This is intentional and beneficial, as Google NotebookLM does not natively support .csv files efficiently at this moment. The extension change does not alter the actual content of your data.

3.2 ELM (Edinburgh Access to Language Models)

ELM provides free and unlimited access to GPT Models under zero retention agreement, meaning that 1) You can use GPT-5 with no rate limits 2) Your data will not be used by OpenAI for model training.

Privacy & Ethics FAQ

Q: Does my instructor see my chats?
> No. Your interactions are private. Instructors cannot access your chat history.

Q: Who can see my data?
> Only you, unless developers are legally compelled to access it (e.g., in cases of felony/illegal activity). See: `https://elm.edina.ac.uk/elm/help/privacy`.

Instructions

1. Go to `https://elm.edina.ac.uk`, login, and go to settings (top right)

2. Select **GPT-5.2**, the largest GPT model to date.

3. *Note:* You may see options for **GPT-5 Mini** and **GPT-5 Nano**. They are less intelligent but faster and cheaper, will be useful when analyzing thousands of documents later. For chatting, it is always better to speak to a larger more intelligent model.

4. Click on the **Documents** tab in the left-hand menu.

5. Create a new folder. You can name it "data".

6. Click the upload icon and select *documents*. For this exercise, you do not need to upload all the files; just upload the first one.

7. Select the uploaded file by checking the box next to it.

8. Ask ELM "Create a mind map of this document". Interpret the results. What is the data about?

9. Experiment with other questions such as "What are the prominent themes in this document?" The answer should be similar, but not exactly the same!

While ELM is great for chatting and simple document analysis, its document processing capability is currently limited. We will use a tool designed for processing multiple documents named **Notebook LM**.

3.3   Google NotebookLM

NotebookLM is a tool designed for **Grounded Research** by Google. It is specifically designed to answer questions by referring to the sources you designate. If you upload a cooking book and ask "What is the capital of Scotland?" it will reply with "The source you provided does not contain such info." This is great for reducing "hallucinations", i.e., AI making stuff up.

Meanwhile, since it is primary function is to receive and search among sources, it excels at finding patterns across large datasets that are too big for a standard chat assistants like ChatGPT or ELM that do not have this function. Therefore, it is great for data analysis!

Interestingly, NotebookLM does not support .csv yet. However you can just rename your .csv as .txt (e.g., climate_related_news_in_english_part1.csv.txt) and upload.

1. **Create:** Go to `https://notebooklm.google.com` and create a new notebook.

2. **Source:** Click the **+ Add Source** and upload the 20 parts of now splitted `climate_related_news_in_english.csv`.

3. **Summary:** Once uploading is complete, click on the Mind Map on the Studio menu (right).

4. **Interpret:** Once created, look at the Mind Map. Click on the individual concepts in the map to learn more.

5. **Explore:** Chat with the AI to learn more about the dataset. Also try creating a video overview or an infographic.

For reference: this is the notebook you should be able to get after following these steps: `https://notebooklm.google.com/notebook/03579c46-207e-42ea-9e6f-531dddbf006c`

Google NotebookLM cannot perform quantitative analysis (e.g., "How many articles mention rising temperatures?") because it goes through chunks of texts to identify most relevant information and use them only, rather than reading the entire dataset at once. Consequently, it cannot see the "full picture" required to count frequencies or aggregate statistics across a whole file. It is like asking a research assistant to answer a question by only reading the top 10 most relevant pages of a 1,000-page book; they can explain the themes, but they cannot accurately count exactly how many times a specific theme appears in the whole book.

Optional: Sign up to Google Student offer to have elevated access to Gemini 3 and extra NotebookLM features. `https://gemini.google/students/`