Instructions for the hands-on session "Harvesting, Processing and Visualising Geo-Encoded Data from Social Media"

1. Data harvesting

- Go to https://github.com/clarinsi/tweetcat and "Download ZIP"
- Extract the tool on your desktop
- Open a command line tool (cmd.exe, Terminal)
- Navigate to the working directory of the tool
- Go to the `1-harvest` directory

1.1. First, we will run data harvesting on a existing project (`hbs` project) - collect some data written in Croatian / Bosnian / Serbian!
- Open with a text editor the `hbs.py` file
- Enter your credentials (four tokens) into the file
- Run the harvesting process
  ```
  python harvesting.py hbs
  ```
- You can monitor the process in the `hbs.log` file. Observe how tweets are written in the `hbs` directory.
- Stop the collection process by modifying the content of the `hbs.busy` file to "stop".

1.2. Create your own project! Define your own list of seed terms (specific for a low-frequency language) and start the collection process.
- Define a file with seed terms, one per line.
- Make a copy of the `hbs.py` file and rename it after your project (`[project].py`).
- Modify the content of the project definition file with your list of seed terms and language(s) of interest.
- Run the harvesting process like in case of the `hbs` project.

1.3. Filter the data collected in the `bi_geo` project with the `filter_geo.py` script to contain only tweets published in Great Britain or Ireland.
- Analyse the `bi_geo.py` file.
- Run the filtering process on the `bi_geo` project (removing tweets not published in Great Britain or Ireland).
  ```
  python filter_geo.py bi_geo
  ```
- Observe that the filtered tweets were written to the `bi_geo.filter` directory.

2. Variable extraction

- Move to the `2-extract` directory

2.1. Extract variables for the `hbs` project.

- Run the variable extraction process on the `hbs` project

  `python extraction.py hbs`

- Analyse the output in `hbs.csv`.

2.2. Extract variables for the `bi_geo` project.

- Run the already defined variable extraction process on the `bi_geo` project.
- Add a variable to the `bi_geo` project containing the number of friends of a specific user.

2.3. If there is time left, extract variables of interest from your own project!