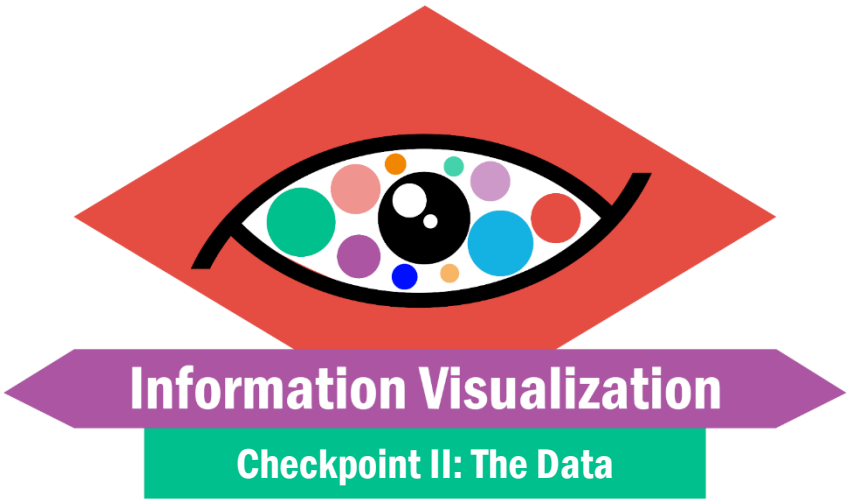

Objective

Show that your data has now been processed and is ready to use:

- You have integrated data sources (made sure country codes match across files, etc.) and cleaned the data (outliers, missing values, etc.).
- Data is processed and ready to visualize. This includes the calculation of derived measures, from the simplest (averages, sums of values of countries for a continent, etc.) to the more complex, based on your domain and questions.
- An appropriate data abstraction has been selected for your visualization.
- You have produced a file (or set of files) in .json or .csv for use in your visualization.

ULTIMATE GOAL: at the end of this Checkpoint, you have the datafiles you need for your visualization.

Requirements

Look at the materials to understand the tasks you need to do:

- Decide which information from the original dataset (or datasets) you will use.
- Parse the original data into .json or .csv format to use with D3.
- Compute derived measures if needed.
- Decide an appropriate data abstraction.

Deliverables

Create a **2-page document using the provided template** and submit it online, until two days before your class (ex: classes on Monday must submit until Friday end of day) which states:

- The initial dataset (size, format, …).
- Which data were selected and/or derived.
- Highlight the derived measures you calculated or carefully justify why you didn't need any.
- The data abstraction you have selected (and why), including:
    - Description of the dataset type (spatial, table, field, etc.).
    - Description of each item and attribute (nominal/ordinal/etc., diverging/sequential scale, etc.) - be clever about this. If you have five attributes which are all the same type, etc. you can describe them together, instead of producing a large and redundant table. DO NOT FORGET TO INCLUDE DERIVED ATTRIBUTES.
    - Semantics (what does each attribute and item stand for.

- How the dataset was processed (cleaned, problems found, how did you fix missing values, cross-referenced different tables/datasets, which tool did you use to do it, etc.).
- How the dataset matches the questions you presented in Checkpoint I (idea: show you can answer them with the data at hand).
- Your final dataset that you are going to use in the project.

Penalties

- Documents over 2 pages long: **1 grade point penalty per extra page**.
- Document uploaded after the deadline: **0.5 grade points penalty per hour of delay.**
- Document template altered (wider margins, smaller font, etc.): **1 grade point penalty.**

Tasks to perform during the lab

The professor will provide feedback. The grade will be made known one week later (see below).

Grading

Your work will be graded according to the following parameters:

- Data parsing (is it parsed into .json or .csv format?).
- Data complexity (information is complex / interesting enough?).
- Derived measures (Are they considered? And, if so, are they relevant?).
- Data abstraction (Correct? Adequate? Relevant?).
- Data Complete (Have enough items and attributes for the goals set in Checkpoint I).
- Match between data and tasks/questions from Checkpoint I.

Additional Notes

After you deliver your document, your work will be graded. HOWEVER, this grade **can be improved by up to two grade points** if you correct any faults pointed out by the professor and submit a revised version of the document HIGHLIGHTING THOSE CHANGES up the beginning of the class taking place 7 days after you receive feedback in class.