Analysing Structured Data with Pandas and ElementTree

Prework

Python: https://programminghistorian.org/en/lessons/introduction-and-installation
Jupyter Notebooks: https://glam-workbench.github.io/getting-started/
Noteable: https://www.ed.ac.uk/information-services/learning-technology/noteable/accessing-noteable
Noteable User Guide: https://noteable.edina.ac.uk/user_guide/#hide_ge_7
Using Jupyter Notebooks and Noteable:
https://github.com/edina/Exemplars2020/blob/master/TeachingDocs/Tutorials/UsingNoteableBeginner.ipynb

Week 1: Pandas

Pandas Documentation

- User Guide: https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html
- Pandas in 10 Minutes: https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html
- Tutorials: https://pandas.pydata.org/pandas-docs/stable/getting_started/tutorials.html

Assignment

Watch the videos in sections 6. Introduction to Pandas and 7. Baby Names with Pandas:
https://www.linkedin.com/learning/python-data-analysis-2015/dataframes-in-pandas?u=50251009
Follow along in your own Jupyter Notebook!

Find your own CSV file to load and turn into a DataFrame (or create your own DataFrame from a lists or a dictionary). What questions can you ask about it using the methods and functions in Pandas? As a starting point, consider calculating, per column of your DataFrame, the minimum value, maximum value, the total number of unique values, and the most common value.
For example, the UN publishes CSV data and the UK Office for National Statistics publishes datasets in xls format, which you can "Save as" a CSV file when you open the data in Excel:
https://data.un.org/
https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/articles/traveltrends/2019/relateddata

Go Further

Example: Creating a DataFrame from Tweets:
https://github.com/sul-cidr/python_workshops/blob/master/data_manipulation.ipynb

Example: Transforming a string into a DataFrame:
https://chrisalbon.com/python/data_wrangling/pandas_regex_to_create_columns/

Learning about NumPy: 4. Introduction to NumPy, 5. Weather Data with NumPy:
https://www.linkedin.com/learning/python-data-analysis-2015/numpy-overview?u=50251009

Learning about Pandas: 2. Series and DataFrames, 3. Data Input and Validation, 4. Basic Analysis, 6. Indexing, 7. Groupby, 8. Reshaping:
https://www.linkedin.com/learning/pandas-essential-training/dataframes?u=50251009