POLS 7387 Global Governance
Professor Denise Garcia
Retrieving and Preparing Mapping Data in Excel

Important Vocabulary

- **Excel**: software that reads .xlsx and .csv files (tabular data)
- **Workbook**: the entire Excel file
- **Sheet:** individual sheets within an Excel workbook file
- **Columns:** vertical data (letters on top of Excel workbook)
- **Rows**: horizontal data (numbers on side of Excel workbook)
- **Metadata**: information about your data, such as what different column variables represent
- **Geospatial Mapping:** digital mapping that uses data, especially latitude and longitude coordinates, to map information

Dataset

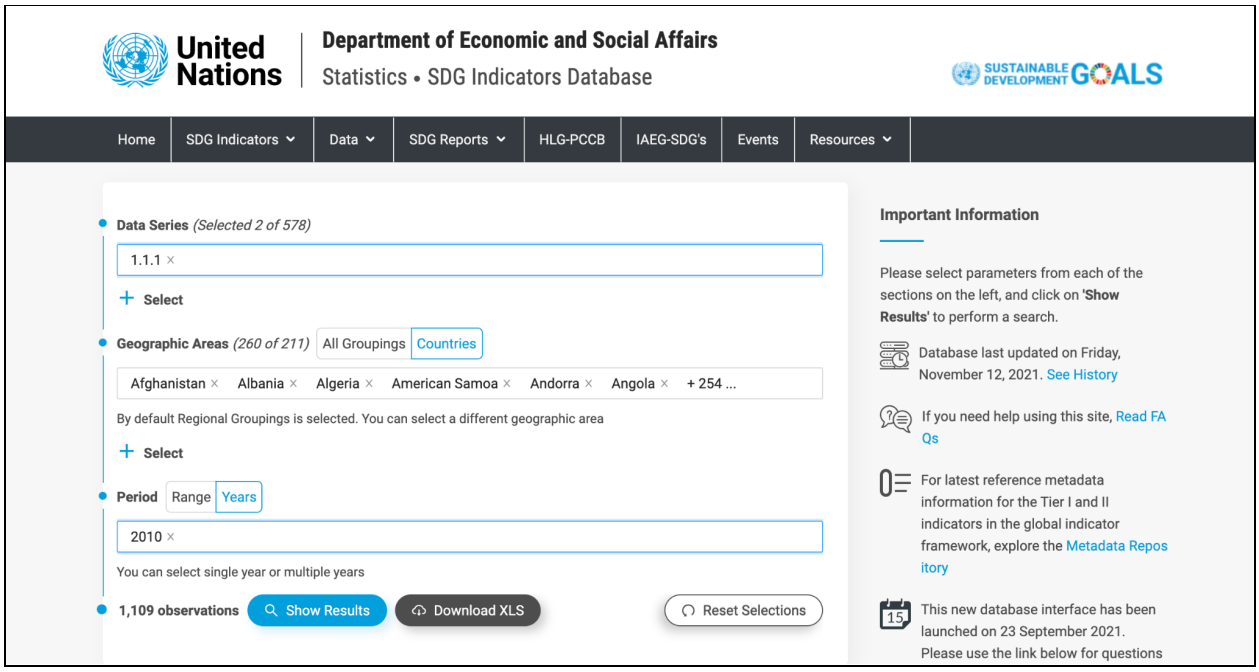
- UN SDG Database: https://unstats.un.org/sdgs/UNSDG/IndDatabasePage > data that needs to be processed before geospatial mapping
- MetaData of UN SDG Database: https://unstats.un.org/sdgs/metadata > Metadata is essential in order to understand what the dataset does show, and what it does not show.

File Types

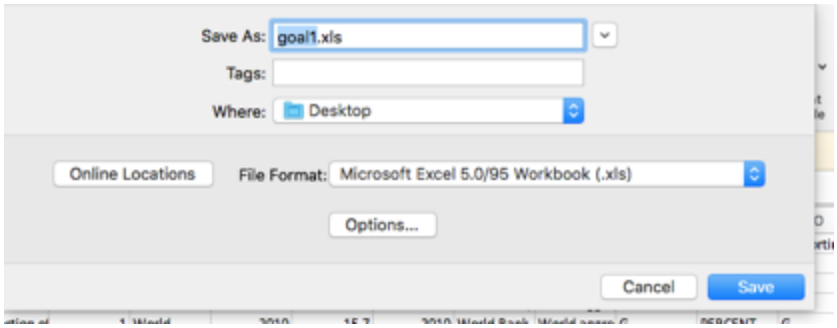
- .xlsx = Excel file; and can usually only be opened in Excel
- .csv = comma-separated values, an open format for delimited data organized into rows and columns. Can be read by Excel & mapping software programs like Google My Maps and ArcGIS
- .kml = mapping file that can be fed into mapping software such as Google My Maps

Data Cleaning in Excel for the Global Sustainable Development Goals Indicators Database

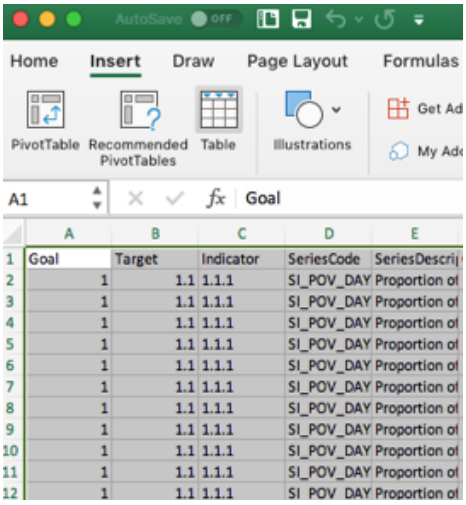
1. Go on the UN Sustainable Goals Website where the data is held: https://unstats.un.org/sdgs/UNSDG/IndDatabasePage
2. From here, look through to determine which goal(s) you would like to focus on. Remember to explore the Metadata repository of the goal you chose so you understand what is in the data
3. Pick the data you want to download – **make sure to be as specific as possible**



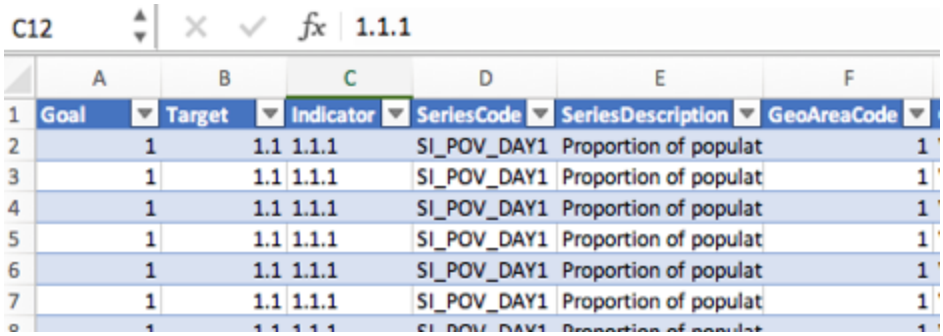
4. Download the data by clicking the "download" button on the bottom of the webpage
5. Once your data is downloaded, it should appear on the bottom of your web browser → click on it to open (it should then open in Excel as a workbook)
6. The data needs to be saved as an Excel file, but right now it is saved as a CSV. To change the format, go to File → Save as. Name your file something you will remember and select .xlsx file where it says File Format



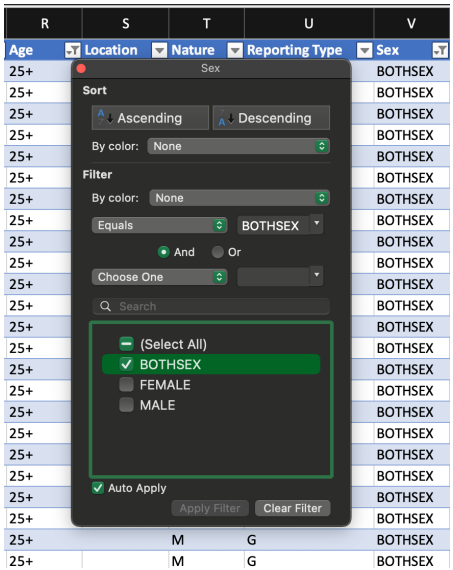
7. Now we want the data to be in Table format so that we can clean it. First, select all the data [control + a/command + a]
8. Next, click Insert → Table → Okay



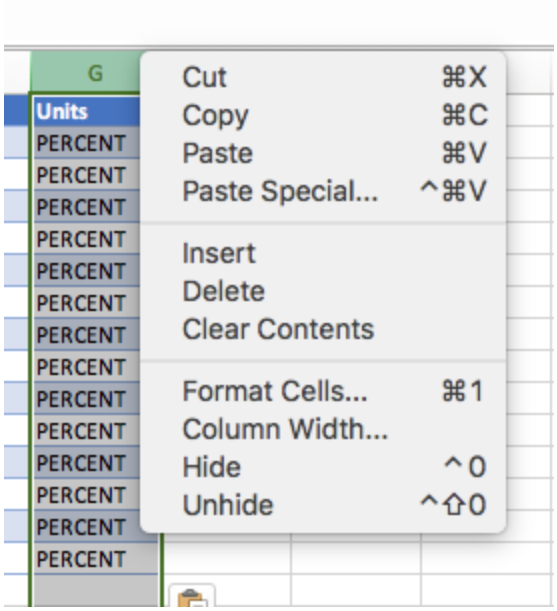
9. The file should then look similar to this in terms of format:



10. Now you can begin cleaning the data using the different columns. Filter the data to what you would like to include by clicking the column arrows next to each variable name. Then select the data you would like to keep in your file.



11. Once the data is filtered, the next step is to copy all of the data (use control + a/command + a again) to put in a new blank Excel file
    a. After copying the data [control/command +c] → open a new blank Excel file (File → New)
    b. Paste the data [control/command + v] into the new file and
    c. Save your new file as something you will remember [File → Save as]
12. You may also want to delete variables that you do not need for mapping. To do so highlight the column that you want to delete → right/control click → delete



13. Once you are happy with the way your file looks and it is ready for GIS, the last step is to save the file as a CSV (File → Save as → click "CSV UTF-8 (comma delimited) (.csv)" under file format → Save)

Things to Note

- Remember to always save your work as you go [command/control + s] or File → Save
- Relatedly, save your data using file names you will remember and store them in a folder you will remember later on
- When choosing how to filter your data, make sure that the data is complete
    - For example, if choosing two years (2001, 2002) make sure there is data for both dates for the countries you choose
- To realign your format – select all [control + a] → Wrap Text

Example Steps for Preparing Data - Class Demonstration

Goal 1, Target 1.1, indicator 1.1.1

1. Go to https://unstats.un.org/sdgs/UNSDG/IndDatabasePage
2. Select Data
    a. Goal 1, Target 1.1, indicator 1.1.1, called "Employed population below international poverty line, by sex and age (%)"
    b. Countries: Bolivia, Chile, Colombia, Costa Rica, Ecuador, Panama, and other specific countries (not regions like "West Asia" or "World")
    c. Year 2010
3. Download this file
4. Open in Excel → Save as Excel document (.xlsx)
5. Select all cases (control + a/command + a)
6. Insert → Table
7. Clean/Filter the data:
    a. Select series code SI_POV_EMP1
    b. Select both sexes
    c. Select age 25+
8. Delete all columns except: value, country name, country code, year indicator
9. If there are formatting issues with the columns → select all again (control + a/command + a) and "wrap text"
10. Copy and paste your data into a new spreadsheet and save it as a .csv