# Collecting and Preparing Data for Mapping in Excel

Cara Marta Messina and Jeff Sternberg

POLS 3409 Global Governance

Dr. Denise Garcia

Fall 2019

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Workshop Agenda

- Objectives
- Collecting UN Sustainable Goals data
- Preparing data for mapping using Excel
- Examples for retrieving and prepping data

Slides, handouts, and data available at
**http://bit.ly/dti-fall2019-garcia**

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# **Workshop Objectives**

- Know places from where to collect data
- Understand how to collect, store, and clean data in Excel
- Apply this method to the UN Sustainable Development Goals data
- Prepare the UN SDG data for GIS mapping

Northeastern University
*NULab for Texts, Maps, and Networks*

# Collecting Data

- A dataset is a collection of several pieces of information called variables (usually arranged by columns).
- A variable can have one or several values (information for one or several cases).
- Qualitative data, quantitative data
- String versus numeric

| Country | Year | GDP |
|---------|------|-----|
| USA | 2015 | 19.39 |

Northeastern University
NULab for Texts, Maps, and Networks

# Where Can You Get Data to Map?

https://data.humdata.org > Clean data with different files types that are mostly prepared to be mapped

https://unstats.un.org/sdgs/indicators/database > messier data that needs to be processed before geospatial mapping

*Feel free to ask questions at any point during the presentation!*

# Understanding Your Data

Where does this data come from? Who collected it? For what purpose?

Metadata for your data – use the metadata to understand what the particular variables represent.

For example, you might see unclear column names. Some data will have metadata attached, which may explain what the columns mean.

Important data for mapping: *geocoordinates* (latitude and longitude) are necessary to map, not just names of countries and cities.

Northeastern University
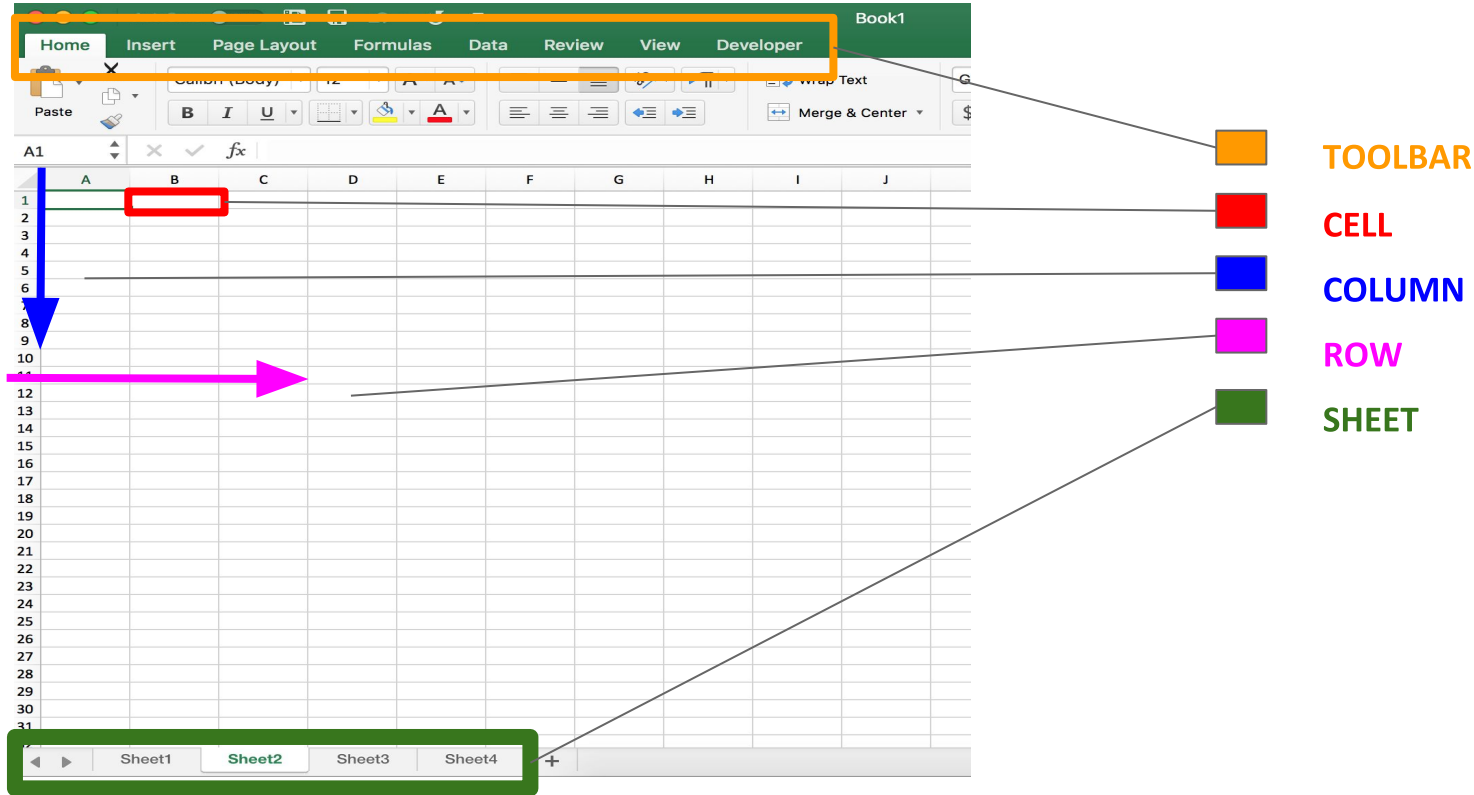*NULab for Texts, Maps, and Networks*

# Excel

Excel is a program that is used to create and edit tabular data (spreadsheets). In Excel, data are organized into rows and columns; data can be presented and analyzed using Excel's functions, such as pivot tables, charts, formulas, and more.



Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point during the presentation!*

# Anatomy of Excel



TOOLBAR

CELL

COLUMN

ROW

SHEET

*Feel free to ask questions at any point during the presentation!*

# Collecting & Prepping Data

Follow the instructions on page 4 of the handout:

- Go to [https://unstats.un.org/sdgs/indicators/database](https://unstats.un.org/sdgs/indicators/database)
- Select the first goal: **Goal 1, Target 1.1, indicator 1.1.1** "Employed population below international poverty line, by sex and age (%)"
- Create a table from your Excel data
- Clean your data: choose the correct columns and variables; make sure there are only **one** set of variables for each country
  - Select "2010" in year
  - Select only *countries* in the country variable (ignore "World," "North America", etc). We want *countries* because they have geocoordinates
- Copy and paste

Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point during the presentation!*

# Collecting Data from HumData

- Go to [https://data.humdata.org](https://data.humdata.org) > DATA
- Under the filters, click "Geodata" (featured) and formats (.XSLX or .KML).
  - .XSLX are Excel files
  - .KML are mapping files that can be used on Google Earth – Bahare will show you the process for mapping with .KML files
- If the data is in an .XSLX file, it will mostly be prepared for you already and in a table (you will still need to choose data)

Northeastern University
*NULab for Texts, Maps, and Networks*

# FOR WEDNESDAY: Download, prep, and bring this data to class

**Follow the instructions from the handout (all the instructions and the instructions on page 3-4)**

**Collect data from this goal:**

Select Goal 16, Target 16.2, indicator 16.2.2, called "Age and sex distribution of detected victims of trafficking in persons (%)"

**Prepare this data (if you can) and save it as a .CSV file:**

Females only, and only those over the age of 15

Northeastern University
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point during the presentation!*

# Thank you!

If you have any questions, contact us at:

**Cara Marta Messina**
Digital Teaching Integration
Assistant Director
messina.c@husky.neu.edu

**Jeff Sternberg**
Digital Teaching Integration
Research Fellow
sternberg.je@husky.neu.edu

Slides, handouts, and data available at **http://bit.ly/dti-fall2019-garcia**

Northeastern University
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*