Using Eurostat and downloading data

Contents

- Browsing Eurostat
  - Configuring datasets
- Data formats
- Downloading data from Eurostat

Co-funded by the
Erasmus+ Programme
of the European Union

Browsing Eurostat

The Eurostat ways

- Via the Eurostat website
  - http://ec.europa.eu/eurostat
- Through the browser:
  - Complete database (tree form)
  - Browse by Theme
  - Browse A-Z
  - Bulk download
- Automated / scripted
  - JSON or UNICODE requests
  - SDMX requests

Other ways

- Excel addins
  - FRED
  - Quandl
  - (…)

- Via SPARQL endpoints
  - eurostat.linked-statistics.org

Data downloading pros and cons

Data via Browsing

- Useful for *ad hoc* work

- Basic customization

- Time consuming

- Unsuitable for:

  – Repeated analysis on changing data

  – Structured reports

... vs scripted

- Unsuitable for *ad hoc* work

- Full customization

- Time efficient

  – Timely to setup the 1st time

- Suitable for:

  – Repeated analyses

  – Structured reports

Bulk downloading

Visit http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing

- What is this;
  - "Manual" process.
  - Complete datasets **at once**
  - No customization
  - **Additional data available:**
    - European Commission
    - International Trade
- When to use?
  - We know the dataset we want.
  - We want the "complete" dataset
  - Massive analysis

- Important to know:
  - Supplementary info
    - Indexes
    - Listings of other data
  - Each dataset contains others
  - All the data is included
  - Takes time to customize
  - Limited formats (tsv, sdmx)
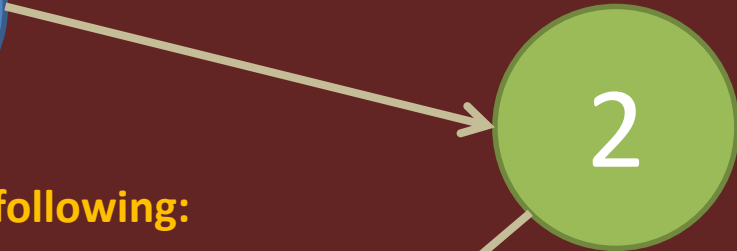 - Need to repeat when data is updated.

Let's browse ...

http://ec.europa.eu/eurostat

1

Visit the following:
- Database
- A-Z
- By theme

2

Try the following:
- Open a table
- Let's customize it
- Write its code name

3

Visit the following:
- Bulk download
- Look for the table

4

Try the following:
- Web services →
- JSON & Unicode →
- Query builder

Data format(s)

Basic premises:
- data always organized in table form [*rows and columns*]
- Columns → data *features*
- Rows→ data *points*

- Plain text delimited files (txt)
  - Columns separated by a delimiter (or more)
  - Rows separated by a delimiter
  - "freedom" to choose delimiters

- **C**omma-**s**eparated **v**alues (csv)
  - Comma " , " used as the delimiter
  - The most common format for data
  - May require some manipulation for text data

- **T**ab-**s**eparated **v**alues (tsv)
  - The tab [TAB⇆] is used as the delimiter
  - Tab → tabular → table form
  - Easier to use than csv, but slightly less common than csv

- Spreadsheet files
  - EXCEL (xls, xlsx) more common than others.
  - "Inherent" tabular format
  - Inefficient storage (compared to the other formats.

- Other types.
  - Most common: json, xml, …

Data types.

Data type is defined *per column*

- Text
  - Character ('C') or String ('S')

- Number / Numeric
  - Integer
  - With decimal points
    - Single or Double precision

- Special types
  - Time or Date
  - Geo-data
  - List [collection of]
  - Video
  - Audio
  - Image

Important:

Always make sure that data is read properly – i.e. the proper type

Why? *Because* …

- E.g. You can add or subtract numbers or do other calculations.
  - 2+2=4 the average of (3, 4, 5) is 4.

- E.g. you can manipulate text
  - *Concatenate*: Donald Duck → DonaldDuck
  - *Change case*: ABC12345 → abc12345

- … you can perform *data-aware* operations:
  - Month '2' ↔ February
  - 2+1 hrs = 180' = 10.800"
  - (37.940880, 22.944860) ↔ Corinth, Corinthia, Greece

Let's download!!!

- Browse by theme →
- Population and social conditions →
- Population (Demography, Migration, and Projections)

- *Left navigation* :: Population data → Main tables
  - Choose a table, then
  - More data
    - [+] to customize
    - Pivoting
- (…)
- **Download options**
  - [# data tables]
  - Flags & Footnotes
  - Separators

Scripted downloads – things to consider

SDMX

- **S**tatistical **D**ata and **M**etadata E**x**change
- The full monty
- Full customization
- English speakers only
- Accessible via REST and SOAP
- Some software tools
- Requires SDMX compatibility

JSON & Unicode

- All datasets except international trade
- Size restrictions
- Customizable
- English, French, German
- Accessible via REST requests
  – REpresentation State Transfer
  – Widely used.

A stepwise process to using Eurostat data

1. Locate the dataset of interest
   – *Table name*

2. Create download query
   – *Select and arrange data*
   – *Verify size limit*
   – *Choose download format*

3. Get the data
   – *Prepare for analysis*

4. Analyze
   – *Imaginatively, creatively, and common sense …* 😊

Summary

- Browsing Eurostat
  - Configuring datasets
- Data formats
- Downloading data from Eurostat