United Nations Food and Agriculture Statistics User Guide

Spring 2025

INST490 - 0102

Kofi Boateng, Christian Janssen, Queenie Ma, Joelle Mbella Banim, Adrian Strasser-King

Table of Contents

Introduction

This user guide serves to provide an overview of the datasets provided in the Food and Agriculture Organization of the United Nations' FAOSTAT website.

Overview of the Data Collection

FAOSTAT is a comprehensive database managed by the Food and Agriculture Organization of the United Nations. It provides free access to global food and agricultural data, covering a wide range of topics essential to analysis, research, and policymaking. The dataset holds multiple decades of information and includes data on the production, trade, food security, emissions, and other key agricultural indicators.

The intended audience for this guide includes policymakers, researchers, economists, and agricultural professionals looking to explore global food and agricultural statistics for decision-making and analysis.

The data is separated into 17 domains or sections, each containing between 1 and 10 datasets. When downloaded, each dataset contains between 3 and 8 files. As of May 2025, these domains as a whole contain a total of 295 files.

Domains include, but are not limited to:

- Production: Crop and livestock production across various countries and years.
- Trade: Imports and exports of agricultural commodities.
- Food Balance Sheets: Consumption patterns, including calorie and nutrient availability.
- Prices and Value Chains: Agricultural producer prices and market trends.
- Emissions and Environment: Greenhouse gas emissions from agriculture, forestry, and land use
- Investment and Government Expenditure: Public Funding and private investments in agriculture

Each dataset is structured with country-level and in some cases regional statistics, categorized by year, commodity, and the measurement unit (i.e. tons, hectares, USD).

With this data source, users are able to:

- Browse and download the datasets in XLSX, CSV, and JSON format
- Generate custom queries using the FAOSTAT Data API
- View data visualizations and analytical reports

Data Source and Collection Methods

FAOSTAT aggregates data from:

- National governments and statistical offices

- International organizations such as the World Bank, International Monetary Fund (IMF), and the World Trade Organization (WTO)
- Surveys, censuses, and remote sensing technologies
    - Remote Sensing technologies covers aspects such as:
        - Satellite Imagery from NASA, European Space Agency (ESA), and the Group on Earth Observations (GEO) to obtain satellite imagery that tracks crop growth, deforestation, soil moisture levels, and land use changes
        - Geographic Information Systems (GIS) process satellite data to create models of agricultural productivity, drought impact, and irrigation patterns
        - Drones and UAVs are used in some cases for high-precision monitoring of farmlands, typically inr regions where satellite data isn't sufficient or clouds affect visibility.
        - Microwave (Radar) and Thermal Sensors (Infrared) are used to estimate vegetation health, soil conditions, and climate factors such as temperature and precipitation
- FAO research and field studies
    - Agricultural surveys as described above with national and regional offices on crop yield, livestock population, and farming practices.
    - Soil and land assessments using programs like the Global Soil Partnership to evaluate soil quality, degradation, and fertility levels. The collected soil samples are also tested for their nutrient content and erosion risk
    - Climate and environmental monitoring on site studies to analyze the impact of climate change on food production and biodiversity loss.
    - Household and market surveys to collect data on food consumption patterns, market prices, and food security indicators which are then used with the FAO's Food Insecurity Experience Scale (FIES) to assess hunger levels in vulnerable populations
    - Pest and disease surveillance field teams monitor the spread of various agricultural pests and diseases such as Desert Locust outbreaks or Wheat Rust disease

Dataset Description

Data Fields and Structure

The FAOSTAT datasets are structured to provide detailed food and agricultural statistics across various domains. Below is a list of the common fields found in most FAOSTAT datasets, along with their data types and descriptions:

- **Area Code (M49)**: *Integer*
  The United Nations M49 code, a standard numerical identifier for countries or regions. For example, 8 represents Albania, and 76 represents Brazil.
- **Area**: *String*
  The name of the country or geographical area (e.g., "Albania", "Brazil").
- **Item Code**: *Integer*
  A unique code assigned to each agricultural item, such as crops or livestock products. These codes are specific to FAOSTAT and may map to international classifications like the Central Product Classification (CPC). For instance, wheat has item code 15 and CPC code ‘0111.
- **Item**: *String*
  The name of the agricultural commodity or product (e.g., "Wheat", "Maize").
- **Element Code**: *Integer*
  A code representing the specific aspect of the item being measured, such as production quantity or area harvested. For example, 5510 might denote "Production".
- **Element**: *String*
  A description of what is being measured (e.g., "Production", "Area Harvested", "Yield").
- **Year**: *Integer*
  The year the data pertains to (e.g., 2010, 2011). Shown as YXXX where the numbers following the Y are the year
- **Value**: *Float*
  The numerical measurement for the given area, item, element, and year. This could represent quantities like tonnes of production or hectares of land, with decimal precision as needed. Doesn’t have a distinct column, specified per year
- **Unit**: *String*
  The unit of measurement for the value (e.g., "tonnes" for production, "hectares" for area harvested, "USD" for monetary values).
- **Flag**: *String*
  A code indicating the status or source of the data, such as "E" for estimated or blank for official data. Here is a brief example of some flags you may come across:

    - A: Official figure
    - E: Estimated value
    - I: Imputed Value
    - M: Missing Value (data cannot exist; not applicable)
    - X: Figure from international organizations
- **Flag Description**: *String*
  An optional field providing a textual explanation of the flag (e.g., "Estimated"). This field may not appear in all datasets.

Depending on the dataset, additional fields may be present. For example, trade datasets might include "Partner Code" and "Partner" to denote trading partner countries, while food balance sheets might feature fields for specific consumption components like "Food Supply" or "Feed". These common fields ensure consistency across the 17 domains, such as Production, Trade, and Emissions, while allowing flexibility for domain-specific data.

File Formats

FAOSTAT datasets are available for download in the following formats:

- **CSV (Comma-Separated Values)**: A lightweight, text-based format suitable for most data analysis tools.
- **XLSX (Microsoft Excel)**: A spreadsheet format for users preferring Excel-based workflows.
- **JSON (JavaScript Object Notation)**: A structured format ideal for developers and API integrations.

These options cater to a wide range of users, from researchers using statistical software to developers building applications via the FAOSTAT Data API.

Size and Volume

The size and volume of FAOSTAT data sets vary significantly depending on the domain and the filters applied during download. A single dataset, such as one covering "Crops and Livestock Products," may contain millions of records when including all countries, items, elements, and years. The data spans multiple decades—often from the 1960s to the present—and covers most countries and regions globally. File sizes can range from a few megabytes for small, filtered datasets to several gigabytes for comprehensive datasets encompassing entire domains. Users can customize their downloads to manage size, selecting specific years, countries, or commodities as needed.

Detailed information about the file size and type of each dataset, alongside other metadata fields, can be viewed upon accessing the “FAOSTAT Domain Metadata Table” hosted on the linked website: https://opendatasets1.github.io/UMD-OpenDataset/

Data Sample

Below is a small, representative sample of a FAOSTAT dataset for the *Crop and livestocks product* domain, illustrating the structure and typical content:

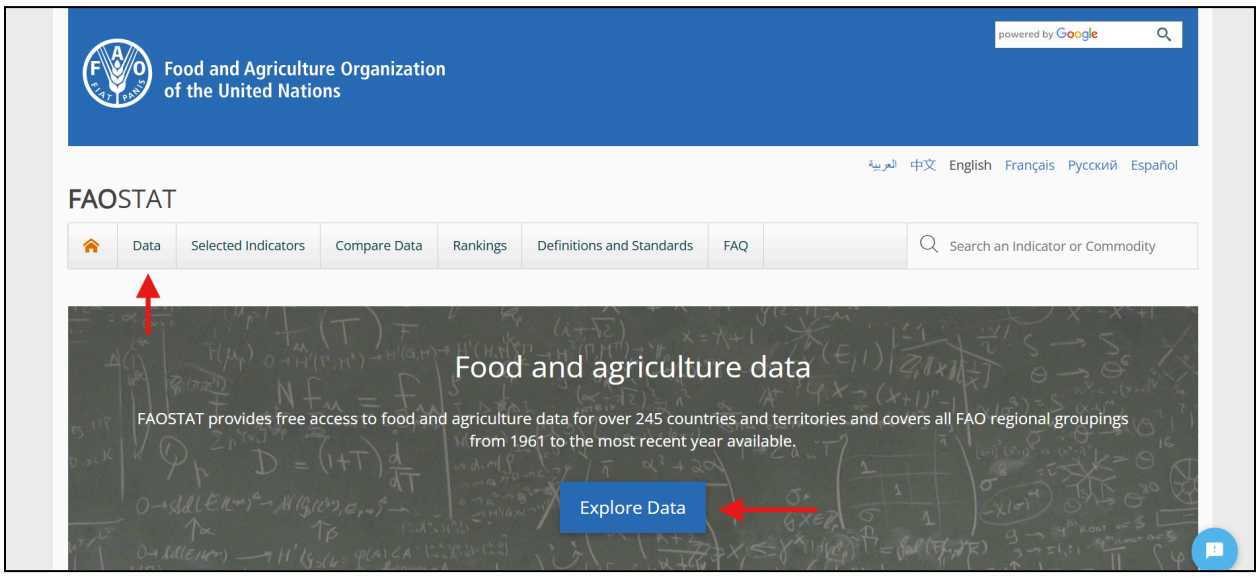
Domain Code	Domain	Area Code (M49)	Area	Element Code	Element	Item Code (CPC)	Item	Year Code	Year	Unit	Value	Flag	Flag Description
QCL	Crops and livestock products	2	Afghanistan	5312	Area harvested	1371	Almonds, in shell	2023	2023	ha	37000	A	Official figure
QCL	Crops and livestock products	2	Afghanistan	5312	Yield	1371	Almonds, in shell	2023	2023	kg/ha	1810.8	A	Official figure
QCL	Crops and livestock products	2	Afghanistan	5312	Production	1371	Almonds, in shell	2023	2023	t	67000	A	Official figure
QCL	Crops and livestock products	2	Afghanistan	5312	Area harvested	1654	Anise, badian, coriander, etc.	2023	2023	ha	25439	E	Estimated value

This sample demonstrates how each row combines “Domain,” "Area," “Element,” “Item,” and “Year” with their unique corresponding FAOSTAT codes to create an entry. The "Flag" field provides additional context to the “Value” field, such as whether the data entered is official (A) or estimated (E). This structured format allows users to analyze trends, compare regions, and integrate the data into broader research or policy frameworks, aligning with FAOSTAT's comprehensive data scope.

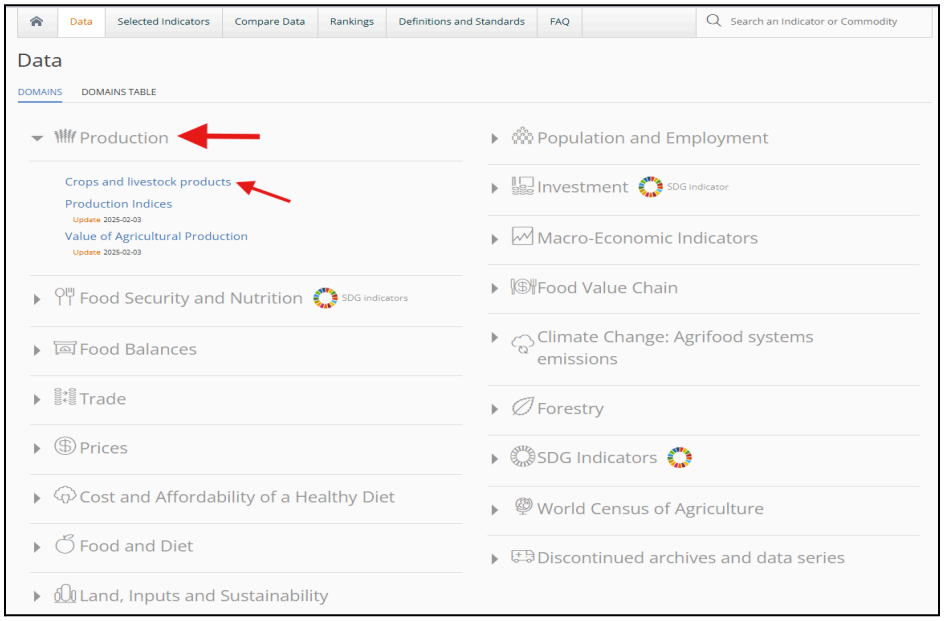
Accessing the Dataset

The data is hosted at the Food and Agriculture Organization of the United Nations' website, which can be accessed through the following link: FAOSTAT.

Upon launching the website in the user's preferred web browser, users will be led to the homepage. From here, users can either click the "Data" tab in the navigation bar or "Explore Data" in the middle of the page to access the datasets.



Upon entering the "Data" tab, a list of domains is shown. Each domain title is a clickable dropdown banner, containing subsections for that domain. For the rest of this guide, we will be using the "Production" domain and the "Crops and livestock products" subsection as an example.



Download Instructions

1. Through the FAOSTAT Website

- Navigate to the FAOSTAT home page.
- Click on the "Data" tab or "Explore Data" to browse available datasets.
- Select a domain (e.g., "Production") and a subsection (e.g., "Crops and livestock products").
- Apply filters for specific countries, years, or commodities.
- Click "View" to display the data, then select "Download" to save it in formats like CSV or Excel.

Note: All FAOSTAT data can be accessed and downloaded directly from the FAOSTAT home page by navigating to the desired dataset and using the download options. No account is required, making it accessible to all users.

2. Through the FAOSTAT API

- **Overview:** For users hoping to access the data programmatically, FAOSTAT themselves do not offer an API, however the 3rd party projects listed below act as an access point.
- **Python Package:**
  - **URL:** https://pypi.org/project/faostat/
  - **Description:** This is a Python package designed to simplify interaction with the FAOSTAT API. It allows users to list datasets, retrieve parameters, and download data directly into Python environments.
  - **Installation:** Install via pip with the command: pip install faostat
  - **Usage Example:** To list available datasets, use:
    import faostat; faostat.list_datasets()
  - **Status:** The package is actively maintained, with the latest version (1.1.2) released on June 1, 2024, indicating that the API remains functional.
  - **Terms:** Data usage is subject to FAO's general terms and conditions.
- **GitHub Repository:**
  - **URL:** https://github.com/FAOSTAT/faostat-api
  - **Description:** This repository contains source code and documentation for the FAOSTAT API. The specific module linked enables users to access FAOSTAT data through REST web services. However, it has not been updated since 2015, suggesting it may not reflect the latest API developments. The Python package is a more current resource for API access.

Additional FAO Resources

- **FAOLEX Open Data**
  - **URL:** https://www.fao.org/faolex/opendata/en/

    - **Description:** FAOLEX is a database of national legislation, policies, and bilateral agreements related to food, agriculture, and natural resources. It provides open data access to legal texts, which could be relevant for studies involving regulatory frameworks.
    - **Conclusion:** This resource is not directly tied to FAOSTAT's statistical data but is valuable for legal or policy analysis.
- **AGROVOC Thesaurus**
    - **URLs:**
        - Access: https://www.fao.org/agrovoc/access
        - Browse: https://agrovoc.fao.org/browse/agrovoc/en/
    - **Description:** AGROVOC is a multilingual thesaurus covering terminology in agriculture, fisheries, forestry, and related fields. It supports both browsing via a web interface and programmatic access via their REST API for integrating standardized terms into applications or research.
    - **Conclusion:** While separate from FAOSTAT's numerical data, AGROVOC is useful for semantic consistency or terminology standardization in reports.

Permissions and Licenses

The FAOSTAT dataset is provided under specific permissions and licenses that govern its use, sharing, and adaptation. The rules are as follows:

- **License:** Data is available under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits users to:
    - Share, adapt, and use the data for any purpose including commercial use as long as proper attribution is made.
- **Attribution:** Users must credit the Food and Agriculture Organization of the United Nations using this format:
    - "FAO [YYYY (year of last update)]. [Name of database: Name of dataset OR Name of database]. [Accessed on [DD Month YYYY]]. [URL] Licence: CC-BY-4.0."
    - **Example:** "FAO. 2025. FAOSTAT: Crops and livestock products. [Accessed on 06 April 2025]. https://www.fao.org/faostat/. License: CC BY 4.0."
- **Restrictions:**
    - FAO states that datasets contained within their corporate statistical databases "shall not be used for or in conjunction with the promotion of a commercial enterprise and/or its product(s) or services (s), and/or in any way that suggests that FAO endorses any specific company, products or services."(*Statistical database terms of use: FAO: Food and Agriculture Organization of the United Nations*).
    - Misrepresenting the data or suggesting FAO's approval is prohibited.

- **Warranty and Liability**:
    - FAO offers no warranty on the data's accuracy, completeness, or suitability for any purpose.
    - Users take on all risks related to using the data, and FAO is not responsible for any resulting issues or damages.
- **Full Terms of Use**:
    - For the complete details, see FAO's Statistical Database Terms of Use.
    - Terms may change over time, review them regularly.

Data Preprocessing and Preparation

Being such a large dataset, FAOSTAT may require cleaning such as handling missing values and converting data types, depending on your analysis needs. Bulk downloads are zipped, requiring decompression, whereas specific dataset downloads from filtering are uncompressed as CSV or Excel files.

File Decompression

As mentioned above, if you bulk download data from FAOSTAT, it often comes as a zip file. To access the data:

- Locate the downloaded zip file on your computer.
- Use your native system unzip settings(Archive Utility for macOS, Windows File Explorer for Windows), or the “unzip” command in Linux to extract the contents.
- Open the extracted files for further processing.

Setting Up the Environment

To process FAOSTAT data, you’ll need to use various software or libraries based on your use case:

- **Spreadsheet Software:** Microsoft Excel or Google Sheets for basic analysis
- **Python Libraries:** You can install different libraries to tailor your data processing. Pandas for data manipulation, NumPy for numerical computations, and Matplotlib/Seaborn for visualization. For integrating FAOSTAT databases right into your codebase, use the faostat package discussed in the previous section, “Accessing the Dataset”.

Data Cleaning Steps

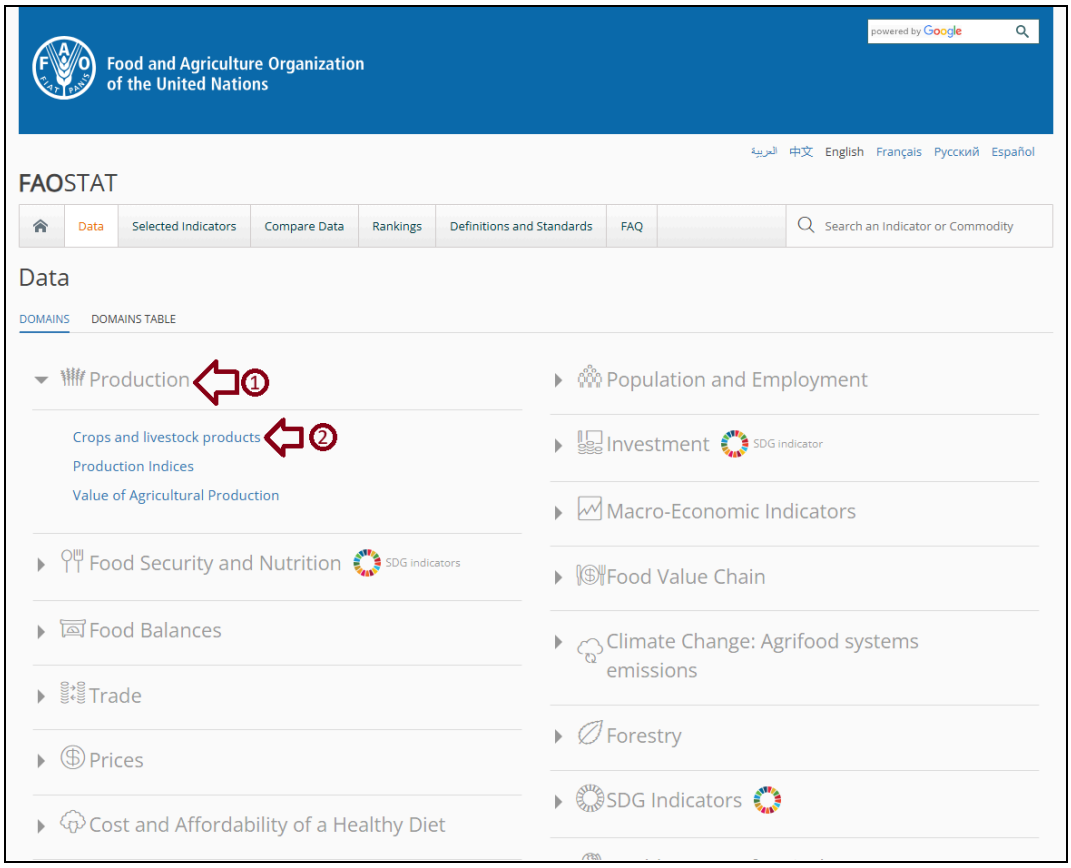
After downloading, clean the data appropriately for your use case to ensure accuracy. This guide wont walk you through programming this and instead provide suggestions:

- **Handling Missing Values:** Check for missing data using pandas “isnull()”
- **Data Type Conversions:** Ensure numerical columns are integers and/or floats, and that categorical data is properly formatted
- **Removing Duplicates:** You can utilize “drop_duplicates()” from the pandas library.
- **Normalization:** Use min-max scaling or standardizations based on what's needed for your analysis

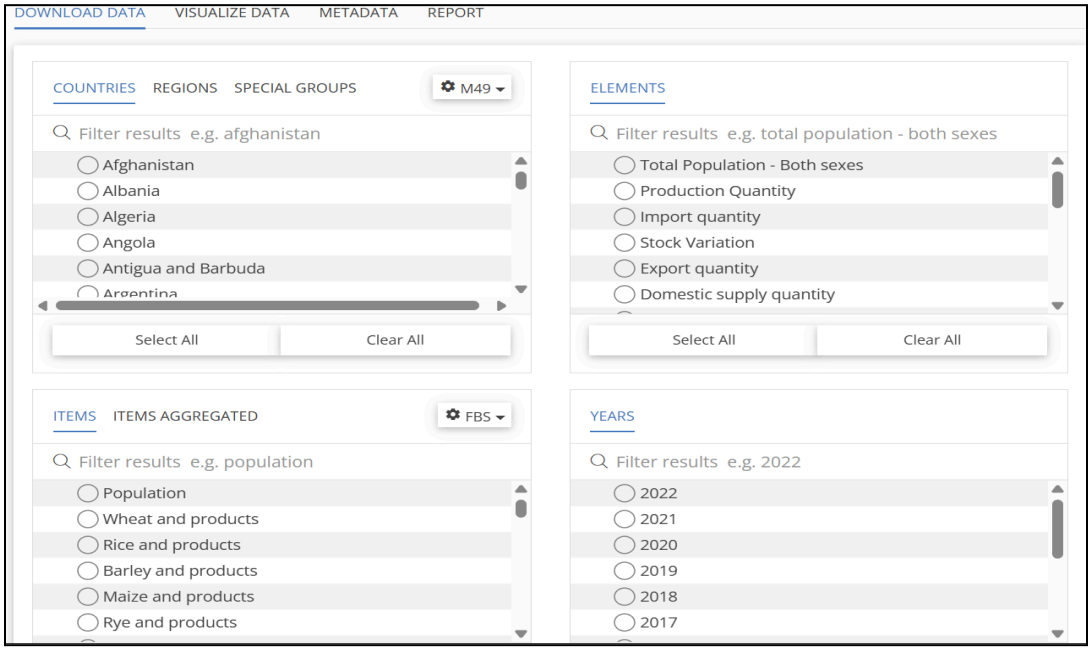
Using Data and Filters on the Website

You can alternatively explore and manipulate the FAOSTAT data without downloading:

1. Visit the FAOSTAT Data Table
2. Select a domain and a dataset (i.e. Production, Crops)



3. Apply filters for countries, years, or commodities, then click “Show Data” to see the table.
4. Sort, search, or copy data from the table for quick use.



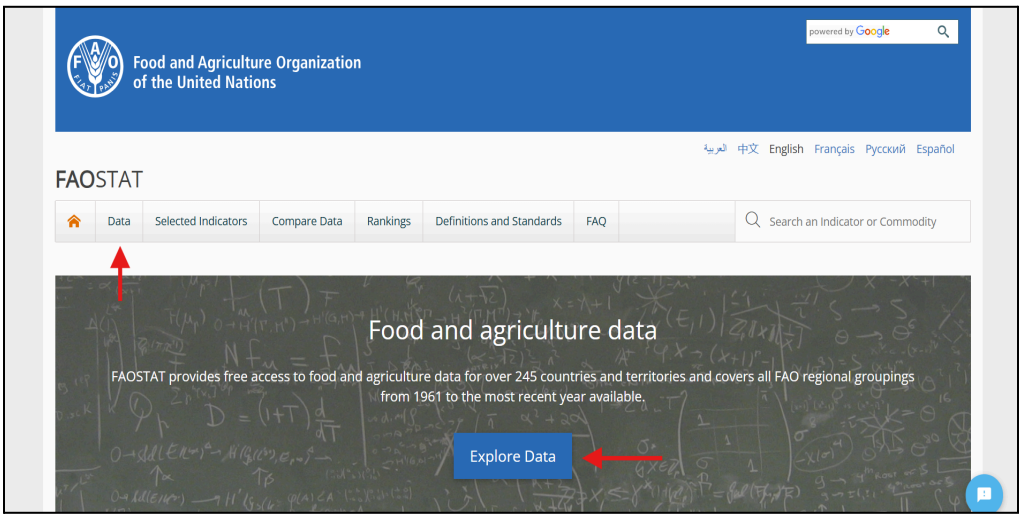
Usage Examples

In-System Visualizations:

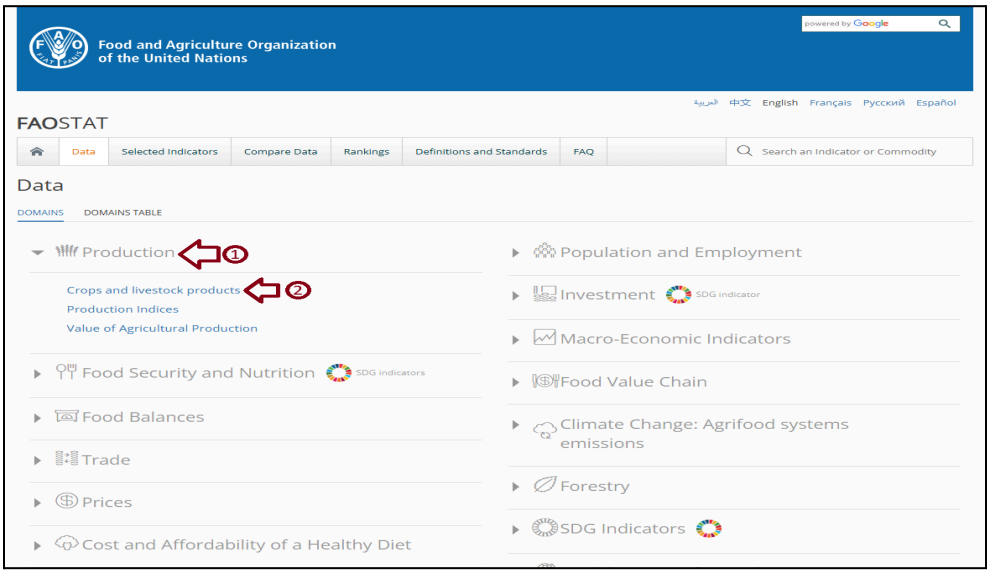
FAOSTAT's built-in visualization tools allow users to create quick, interactive charts directly from the database interface. These include line graphs, maps, pie charts, and bar charts generated using selected variables like country, item, year, and aggregation type. The purpose of these visualizations is to help users explore trends and compare values across countries or time periods without needing to download or process data externally.

Using FAOSTAT Visualize Data:

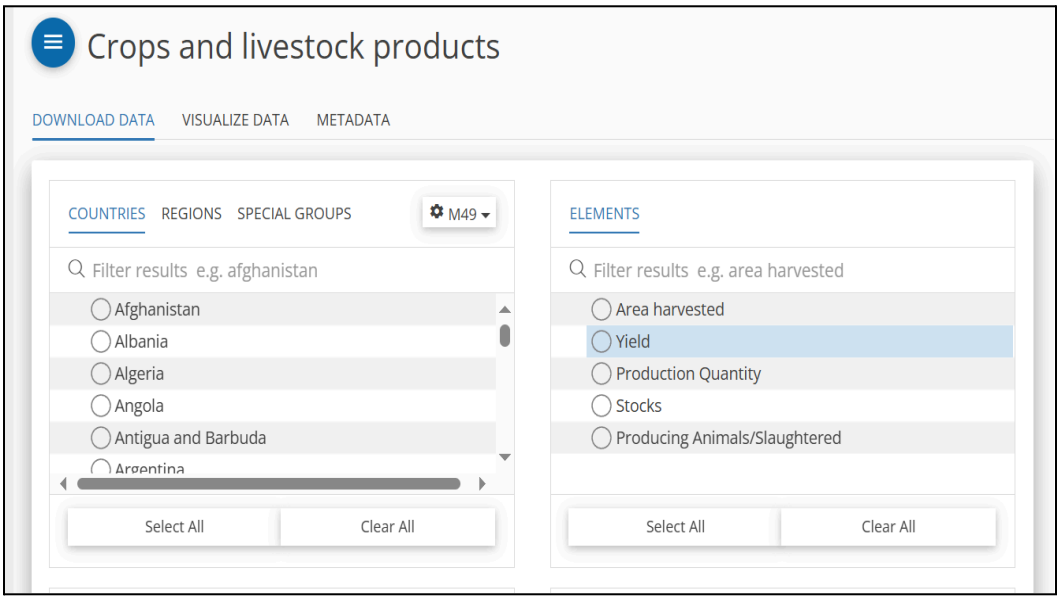
Step 1: Visit the [FAOSTAT Data Table]



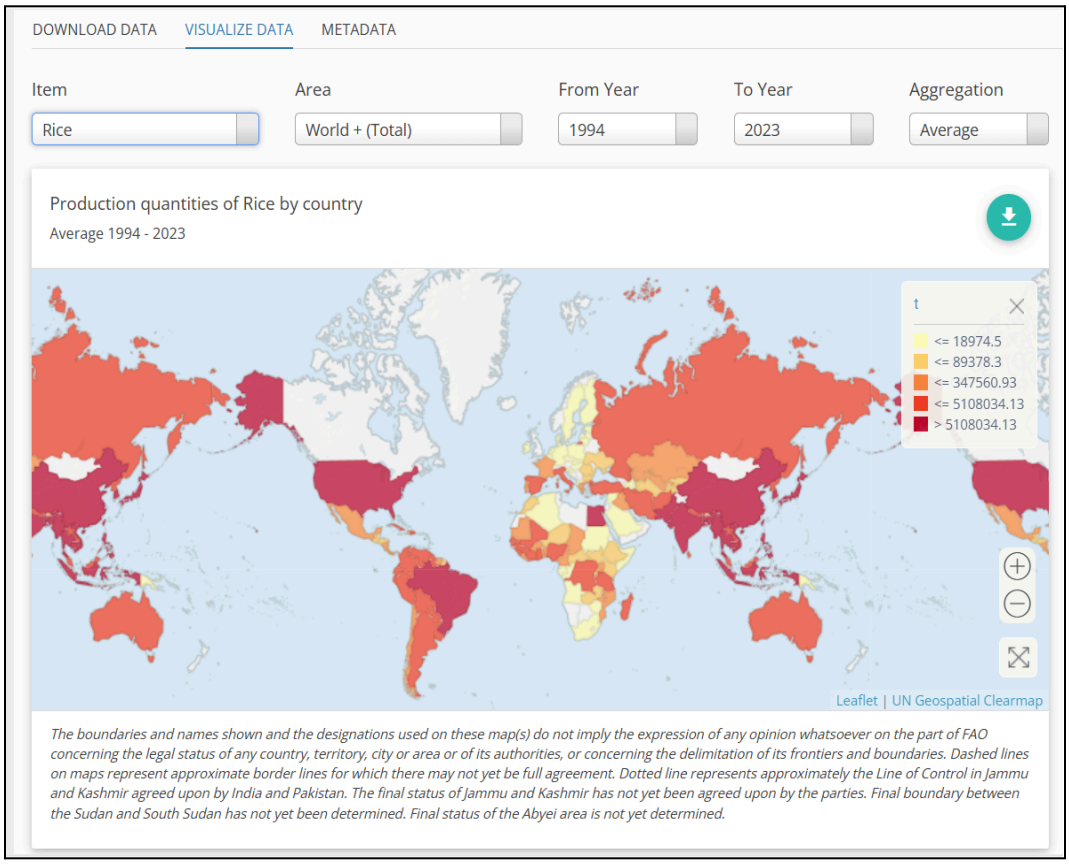
Step 2: Select a domain and a dataset (i.e. Production, Crops)

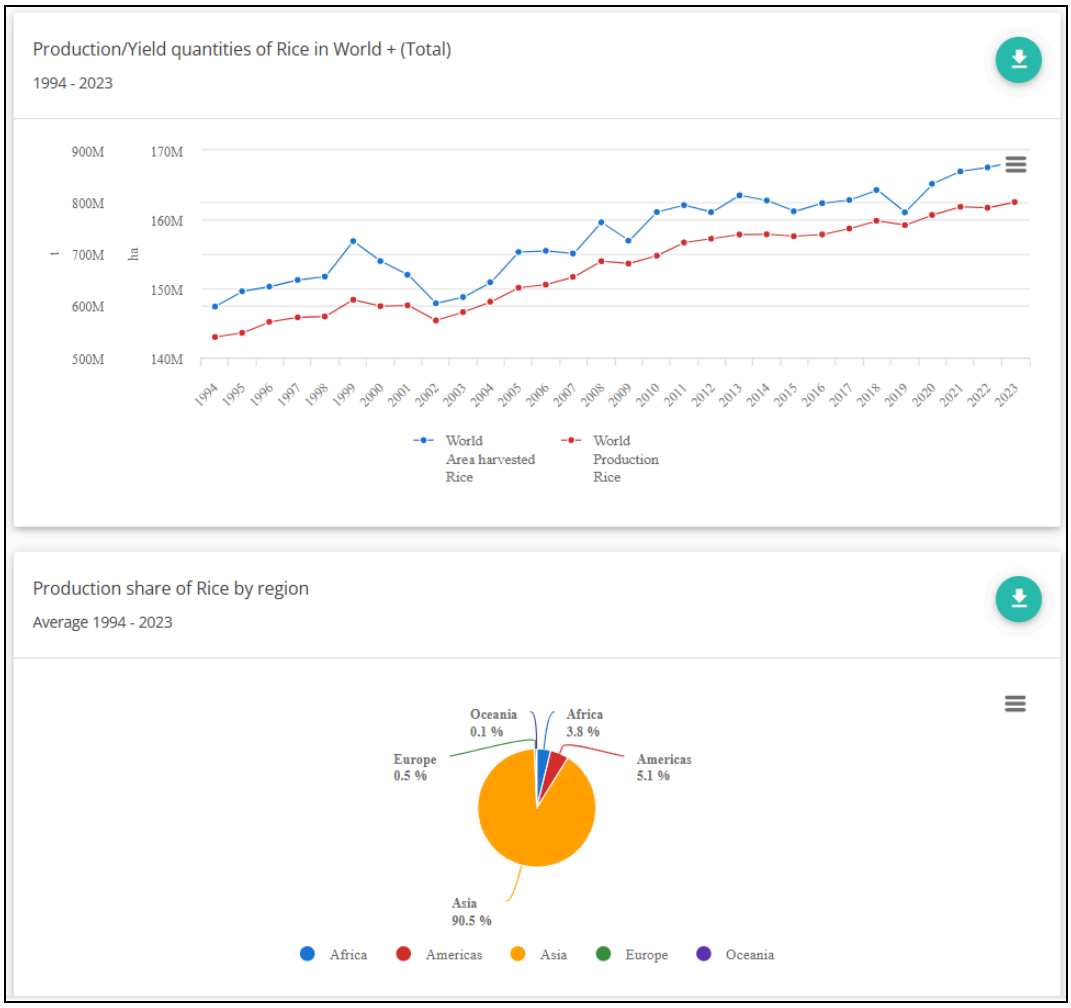


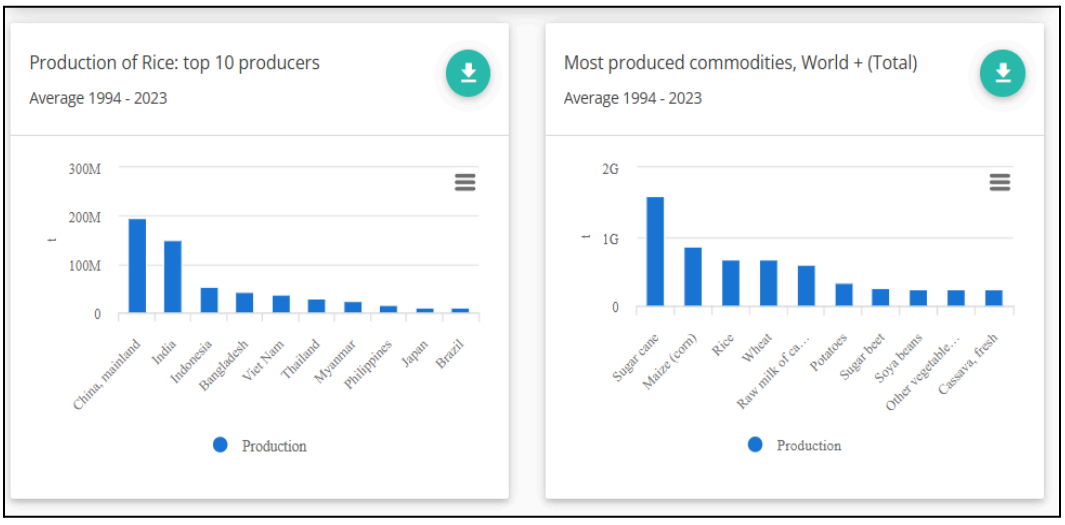
Step 3: Navigate to “VISUALIZE DATA” Section



Step 4: Apply filters for Item, Area, From Year To Year, and Aggregation to see the visualizations. Scroll down to see more visualizations.

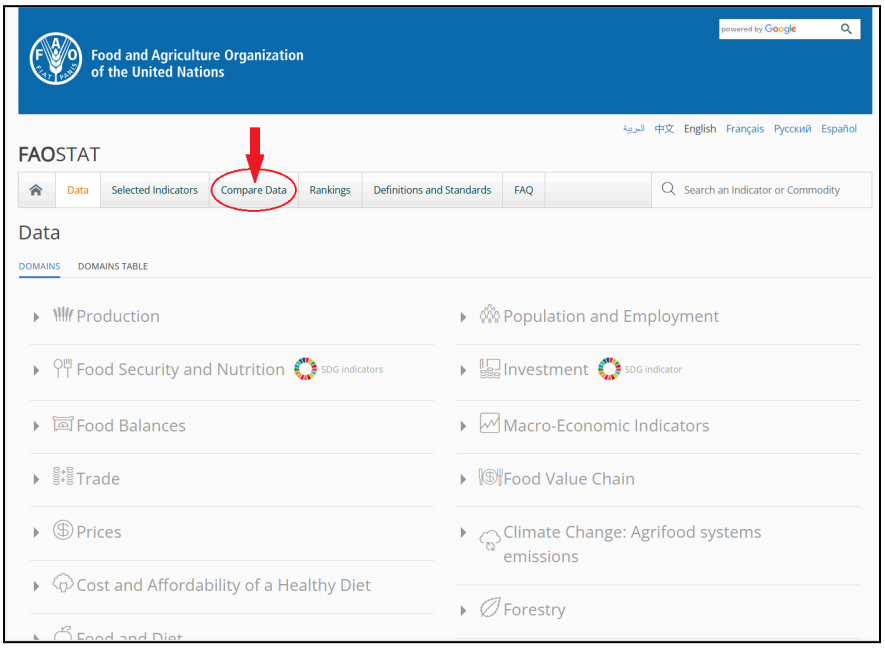



Production/Yield quantities of Rice in World + (Total)
1994 - 2023
900M
800M
700M
600M
500M
t
170M
160M
150M
140M
ha
1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023
World Area harvested Rice
World Production Rice
Production share of Rice by region
Average 1994 - 2023
Oceania 0.1 %
Africa 3.8 %
Europe 0.5 %
Americas 5.1 %
Asia 90.5 %
Africa
Americas
Asia
Europe
Oceania

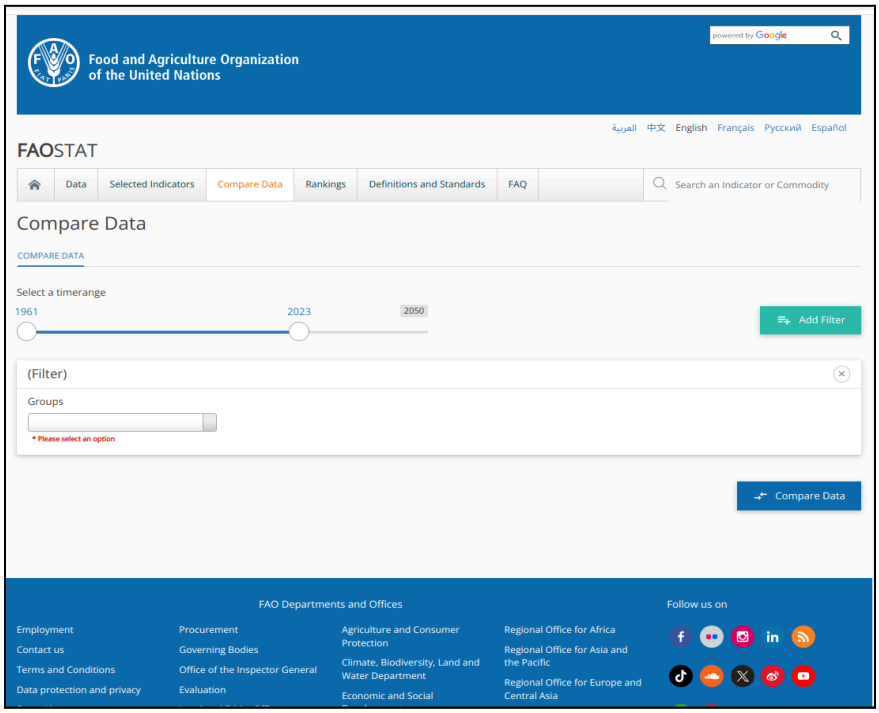

Production of Rice: top 10 producers
Average 1994 - 2023
300M
200M
100M
0
t
China, mainland
India
Indonesia
Bangladesh
Viet Nam
Thailand
Myanmar
Philippines
Japan
Brazil
Production
Most produced commodities, World + (Total)
Average 1994 - 2023
2G
1G
0
t
Sugar cane
Maize (corn)
Rice
Wheat
Raw milk of ca…
Potatoes
Sugar beet
Soya beans
Other vegetable…
Cassava, fresh
Production

Using FAOSTAT Compare Data

Step 1: Navigate to the **"Compare Data"** tab

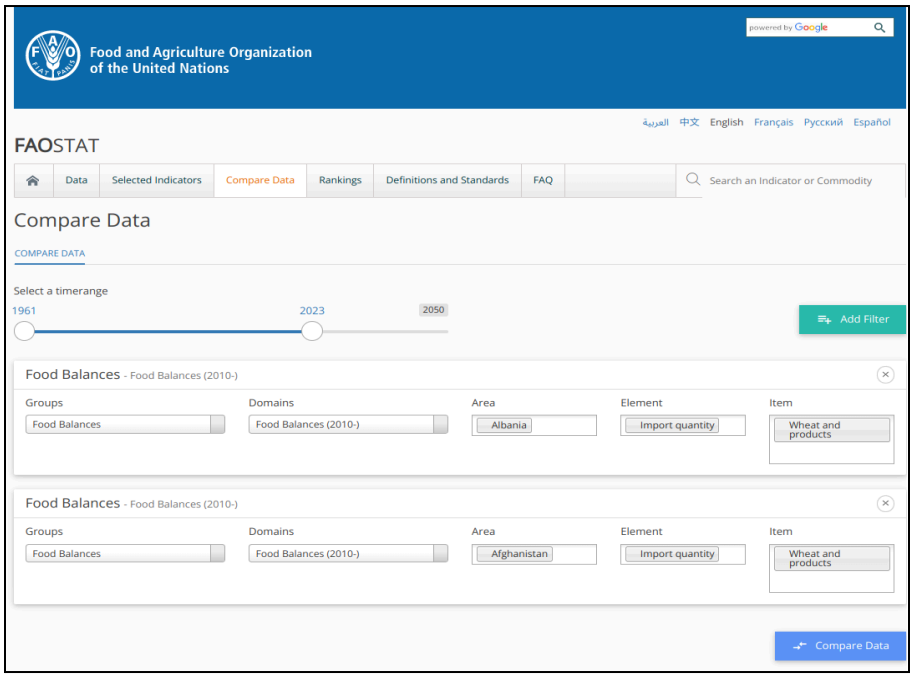


Step 2: Select a time range by dragging the two circles to desired year interval with the headline called **"Select a timerange"**

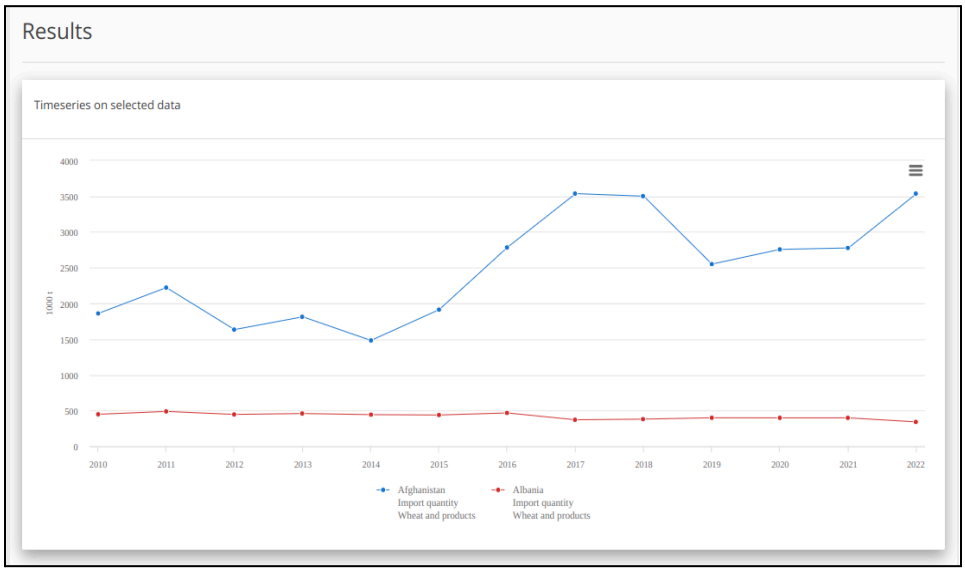


Step 3: Select and enter in all the dropdowns to select a specific variable.

Select "**Add Filters**" to compare multiple sets of data



Step 4: Click "**Compare Data**" and scroll down to see the comparison visualization.

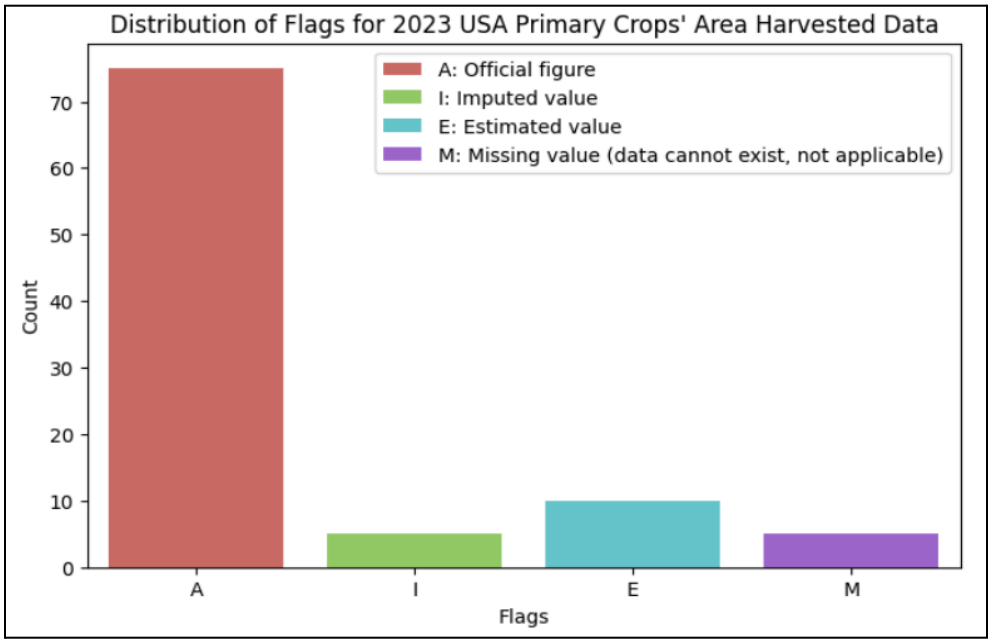


Out-of-System Visualizations:

Out-of-system visualizations are created by downloading FAOSTAT data and analyzing it using external tools such as Excel, R, Tableau, or Python. These charts can highlight patterns not easily captured or recognized by FAOSTAT's interface. The purpose is to demonstrate more advanced use cases, offering users greater flexibility and analytical depth beyond the platform's built-in capabilities.

Visualization 1:

This visualization was generated in Python with Matplotlib, a bar graph that displays the distribution of data quality flags for "Primary Crops" items and "Area harvested" elements in the United States for 2023 in the "Crops and livestock products" domain. Each bar represents the total count of data entries marked with a specific flag:

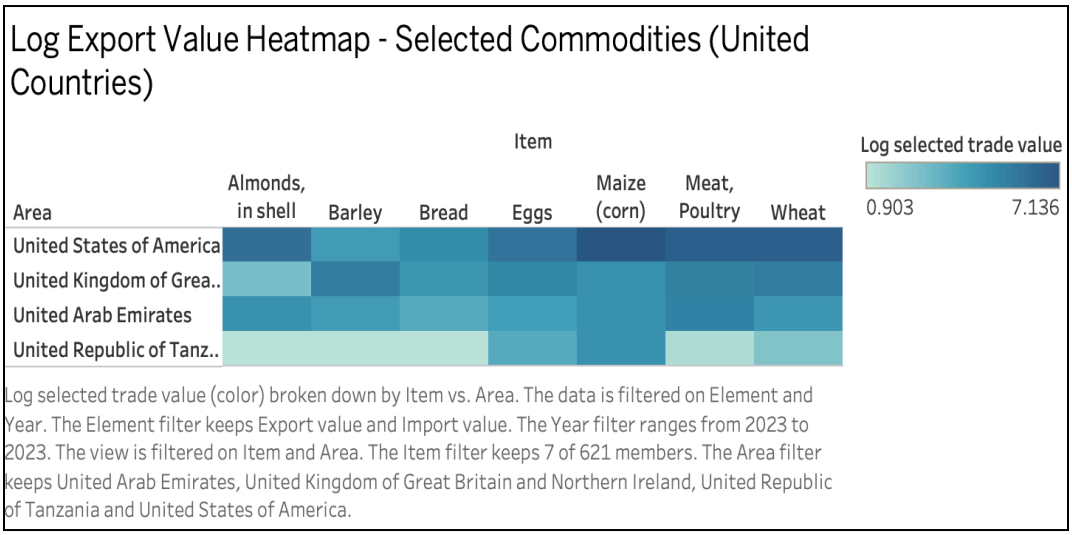


Flag Description	
Official figure	75
Estimated value	10
Imputed value	5
Missing value (data cannot exist, not applicable)	5

Visualization 2:

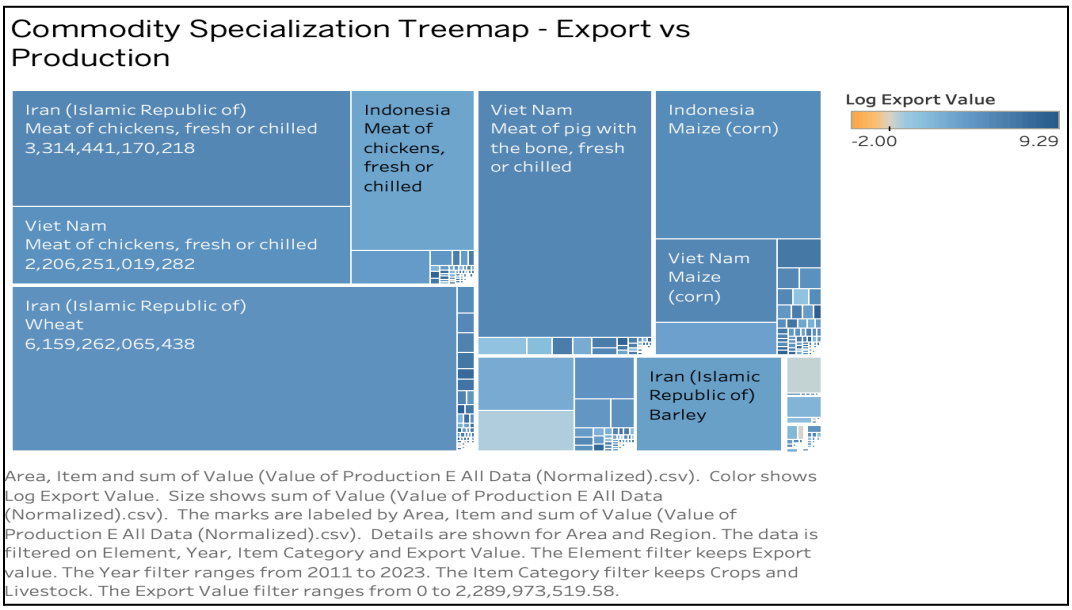
This visualization was generated in Tableau, a view created to compare the export value of all "United" countries for specific items such as Almonds, Barley, Eggs, and more. Due to the values reaching to the billions for some countries or just the thousands for others, a logarithmic scale was applied to balance it with base 10, along with only using data from 2023. For comparison, the United States of America had an export value of 14,116,169,000

in comparison to the United Republic of Tanzania's 74,263,000 for Maize, in units of 1000 USD.



Visualization 3:

This visualization was generated in Tableau, a treemap designed to highlight commodity specialization across countries by comparing export and production values for various items like Wheat, Maize, Meat, and more. To manage the wide range of export values, which span from millions to billions, a logarithmic scale was applied using base 10, alongside data filtered to the year 2023. For instance, Iran (Islamic Republic of) showed a production value of 6,613,943,643,829 and an export value of 254,926,000 for Wheat, while Viet Nam had an export value of 2,435,125,223,243 for Meat of chickens, in units of 1000 USD.



Troubleshooting and FAQs

Below are some problems you may encounter when working with FAOSTAT data and their solutions or general support topics:

1. File Formatting Issues
    a. CSV files may not display correctly in your software if the delimiter settings aren't correct (FAOSTAT typically uses commas) and ensure the encoding is set to UTF-8.
    b. Some Excel files may be too large or crash your software, using a data library recommended in the "Setting Up the Environment" to process the data can mitigate this.
2. Data Corruption
    a. If a downloaded file is incomplete or unreadable, attempt to download it again.
3. Having Trouble Starting?
    a. View a brief demo created by FAO itself: FAOSTAT DEMO

Contact Information

For additional help or to address issues not covered here:

- For general FAOSTAT issues, visit the help section or contact the FAOSTAT Statistics Division at "faostat@fao.org". Alternatively, their full contact information is below:

  Food and Agriculture Organization of the United Nations
  Viale delle Terme di Caracalla
  00153 Rome, Italy
  T: (+39) 06 570 55303

- Looking for like-minded individuals? Check out the Global Forum on Food Security and Nutrition (FSN Forum)! Their goal is to foster awareness, knowledge sharing, debate and mutual learning on a wide range of issues that affect food security and nutrition, and food systems. Access it below:
    - https://www.fao.org/fsnforum/