# Dataset Nutrition Label:
# Impact of Climate Change on Global Food Security

Thuy Nguyen, Nola Fung, Nguyen Le

# Motivation

- **Why Was the dataset created?**

USDA's Food Security data (1980–2013) tracks global commodities, aiding research. NOAA's Climate Time Series notes temperature anomalies, supporting climate studies. Integration aligns with human rights, stressing ethical use for global challenges.

- **Who created the dataset?**

The International Food Security dataset was curated by the USDA (U.S. Department of Agriculture), while the Global Climate Time Series dataset is provided by the NOAA (National Oceanic and Atmospheric Administration).

- **Who funded the creation of the dataset?**

The USDA likely funded the International Food Security dataset, while the NOAA probably funded the Global Climate Time Series dataset. Government funding likely played a crucial role in supporting the creation of these datasets.

# Collection Process
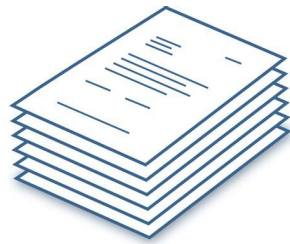
- **How was the data acquired?**

  The USDA obtained the International Food Security dataset by tracking global commodity data. The NOAA collected the Global Climate Time Series dataset by combining sea and land temperature measurements over time.

- **If the dataset is a sample, what was the sampling strategy?**

  The International Food Security dataset spans globally over three decades, indicating broad inclusion, while the Global Climate Time Series dataset represents continuous temperature anomaly measurements, not following a traditional sampling approach.

- **Over what timeframe was the data collected?**

  The International Food Security dataset spans 1980 to 2013, while the Global Climate Time Series dataset timeline is not explicitly mentioned but likely corresponds to the observational period.

# Mechanisms/procedure used to collect the data

The International Food Security dataset likely utilized international trade records, agricultural reports, and collaborations to collect information on commodity import, export, and supply availability. While specific procedures are not explicitly mentioned, statistical methods were likely employed for data processing.

The Global Climate Time Series dataset, focusing on temperature anomalies, used satellite observations, ground-based monitoring stations, and scientific research for data collection. The creation involved combining sea surface temperature with land surface air temperature data. Specific procedural details are not provided, but these practices are common in climate science for monitoring temperature trends.
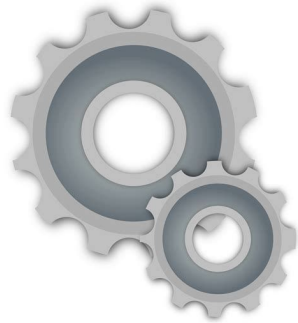
# Joined Dataset: About

Data Creation Range: 2003-2013

We created a joined dataset out of the food security dataset and the climate change dataset. These are both important worldwide issues, and we wanted to see if they are connected to each other in some way.
By joining the datasets, it is easier to see the correlation between climate anomalies and changes in food availability. We added a couple variables, as well:

- **kg_food_per_person** is the total grains/cereals divided by the population for a country in a certain year.
- **Amount Category** is a simple measure of whether the listed amount is relatively high or low. It is on a scale of low, medium-low, medium-high, and high.
- **new_numerical** multiplies the climate anomalies by 2 to highlight/exaggerate the trends.

# Joined Dataset: Composition

- **Content**
Each of the 26088 instances in the dataset is a record of information related to food availability for different countries. The climate dataset is a sample of 10 years from a larger time period, but the food security dataset includes all instances. The instances consist of reported statistics across 10 years related to food production, food availability, and population for different countries. Additionally, the dataset contains the global climate anomalies for each year and the calculated amount of food per person (in kilograms, based on total grains/cereals) for each country.

- **Handling**
The instances are labeled with the category of data that they contain (type of crop vs population vs other, etc.), but don't have IDs. Each instance is complete (i.e., there is no missing information). The instances are organized by country, so related information is grouped together. It makes most sense to look at data by country, as outside of that, any one part of the data doesn't rely on other parts of the dataset.

- **Sensitivity**
The dataset does not contain potentially confidential information. The food security dataset consists of public records, and the climate dataset is for global climate anomalies, so neither are confidential. The instances are for entire countries, so it is not possible to identify individuals or subpopulations from the dataset. The dataset is not related to controversial topics and is based purely on numbers, so it doesn't contain offensive or sensitive information.

# Joined Dataset: Uses

- **What tasks could the dataset be used for?**
  The dataset could be used to see how food availability is affected by climate change, and it could be used to compare food availability between countries and over time.

- **Is there anything about the collection/preprocessing methods that might impact future uses?**
  It is important to note that the kg_food_per_person variable is based on total grains/cereals and does not account for other types of food. The date range should also be kept in mind.

- **Are there tasks the dataset shouldn't be used for?**
  There is not much room for misinterpretation of the dataset that might lead to harms, so any task should be fine as long as it's relevant.
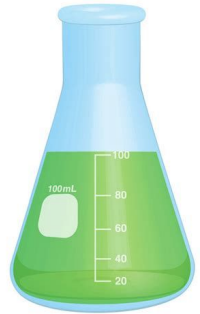
# Joined Dataset: Preprocessing/Cleaning/Labeling

- **Was any preprocessing/cleaning/labeling of the data done?**

  First, we considered what information is related to the question we tried to answer and kept only those. We then remove other less irrelevant information and summarize them as simple infographics.

- **If so, was the "raw" data saved?**

  Most of the raw data was saved, the unnecessary information was removed. In addition, additional information composed from the raw data was added.

- **Is the software used to preprocess/clean/label the data available?**

  The software used to preprocess and clean the data is RStudio which is widely available for free. For labeling, we used Google Slide to document and display collected data.

# Joined Dataset: 8 Ethics Principles

- **Privacy**

  The data does not reveal any private information of any individual, only information that was made available to the public.

- **Accountability**

  The source for the data was provided. Any new information and conclusions are our effort to provide better understanding and meaning to the raw data so people can be better informed.

- **Safety and Security**

  This information did not composed by AI and we did not use any AI-related tools.

- **Transparency and Explainability**

  The data is raw and open source. We aim to deliver our idea in a way that is straightforward and easy to understand.

# Joined Dataset: 8 Ethics Principles (Cont.)

- **Fairness and Non-discrimination**

  We judged information with fairness and no bias. Many countries are included in the data to show inclusiveness.

- **Human Control of Technology**

  There is no AI that impacts the processing of raw data.

- **Professional Responsibility**

  We are responsible for the new information, we follow the guidelines and keep the end goal in mind.

- **Promotion of Human Values**

  We hope this new data can help society be better informed.

# Joined Dataset: Shiny App

Designed to create an interactive web application for exploring and visualizing the joined dataset, which contains information related to "Country," "Commodity," "Year," "kg_food_per_person," and "Amount_Category." Here's a breakdown of how the Shiny app would be used with this dataset:

**Interactive Data Exploration:**
- Year Range Slider: Users can select a range of years using the slider, allowing them to focus on specific periods of interest.
- kg_food_per_person Range Slider: Users can filter the dataset based on the range of kg_food_per_person values, helping them explore how different levels of food availability per person relate to other variables.
- Amount Category Dropdown: Users can filter the data based on different levels of "Amount_Category," providing the ability to investigate specific categories or ranges of amounts.

**Dynamic Plotting:**
- The scatter plot in the main panel dynamically updates based on user input. Users can observe how changes in the selected variables affect the distribution of data points in the scatter plot.

**Faceted Display:**
- The scatter plot is faceted by the combination of "Country" and "Commodity," providing separate panels for each unique combination. This helps users explore how the relationships between variables differ across different countries and commodities.

**Enhanced Readability:**
- The app includes enhancements for readability, such as rotated x-axis labels to prevent overlap and improved font size for facet labels. These adjustments make it easier for users to interpret the visualizations.