Technical Report

Nick Ochsner, Jenaya Modeste, Kai Lee, Rama Yakkala, ITCS 4122/5122 Fall 2025

Data & Collection

The dataset used in this project was obtained from Kaggle:
Climate Changing Indicator Worldwide(1961 to 2022)
Source:

www.kaggle.com/datasets/zubairamuti/climate-forcaste-of-all-countries1961-to-2022/data

The dataset includes **approximately 225 rows**, each representing a unique country or territory. One row summarizes global temperature change. The main columns used in our analysis were:

- **Country identifiers:** Country name and ISO3 code
- **Temperature anomaly values:** ***F1961*** through ***F2022***, representing the annual temperature anomaly above the baseline for each year

Why This Dataset

This dataset provides a broad, global perspective of temperature anomaly trends across many countries over six decades. It is directly relevant to our topic of climate change because it allows year-over-year comparison and net change calculations.

Known Limitations

- Many countries have **incomplete data**, with missing values spanning several years.
- Countries with fewer recorded years have **limited ranges for calculating net change**, which may affect comparisons.
- The dataset ends in **2022**, so it does not incorporate temperature anomalies from 2023–2025.

These limitations were considered when interpreting final results.

Data Cleaning & Preparation

To prepare the dataset for analysis, several preprocessing steps were performed. The goal was to isolate useful variables, handle missing data appropriately, and create derived metrics for comparing temperature trends between countries.

1. Column Selection

- Removed irrelevant or labeling-related columns.
- Isolated the year-based columns (F1961–F2022), which served as the primary variables.

2. Handling Missing Values

- Missing values were left as NaN and skipped during calculations rather than imputed.
  This decision ensured no artificial inflation or distortion of the temperature trends.

3. Derived Feature Creation

Several functions were implemented to generate additional analytical features:

- **get_first_last**
  Extracts the *first* and *last* valid temperature anomaly for each country.
  Output: two new columns: *First Valid Value* and *Last Valid Value*
- **get_min_max**
  Identifies each country's minimum and maximum anomaly values.
  Output: *Min Temp Anomaly* and *Max Temp Anomaly*
- **get_first_last_avg**
  Computes the average of the first 5 and last 5 years of valid data.
  This provides a more robust measure for countries with inconsistent yearly reporting.
- **get_net_change**
  Calculates temperature change using several methods:
    - First → Last

- Min → Max
- Avg of first 5 years → Avg of last 5 years

These variations help determine the most reliable representation of long-term change.

4. Removing Aggregate Row

The dataset included a row labeled **"World"**, representing the global average.
This was removed to avoid skewing country-level comparisons while still acknowledging its importance as a reference point for global climate trends.

Processing & Methods

Tools & Libraries Used

- **Pandas** (e.g., read_csv, drop, dropna, isna, duplicated, bfill, ffill)
- **NumPy** (e.g., np.polyfit)
- **Matplotlib** (e.g., plt.figure, plt.plot, plt.title)
- **Seaborn**
- **Plotly Express** (px.choropleth)

Visual Mapping Choices

- **Line Charts (Matplotlib):**
  Used to show long-term temperature anomaly trends for selected countries.
  Line charts are ideal for illustrating patterns over time, such as warming rates and year-over-year variability.

- **Choropleth Maps (Plotly Express):**
  Used to visualize net temperature change across countries.
  This enables intuitive geographic comparison and highlights regional warming patterns or anomalies.

No interactivity beyond the default Plotly hover features was implemented.

Validation & Quality Checks

Several measures were taken to verify the accuracy and reliability of the analysis:

Data Checks

- Confirmed temperature anomaly ranges were consistent with known global climate datasets (~−1°C to +2°C).
- Ensured countries with insufficient data were not falsely treated as low-change regions.

External Validation

We cross-checked general global warming trends using credible scientific sources:

- NASA Global Climate Change
- United Nations climate reports
- Intergovernmental Panel on Climate Change (IPCC)
- Paris Agreement documentation

These sources supported the dataset's overall trends despite missing values.

Alternative Analyses

Although we considered exploring other narratives (e.g., no warming, cooling claims), the overwhelming consistency of evidence from both the dataset and scientific literature provided no credible basis for such alternatives. Instead, we focused on objectively evaluating the dataset to avoid confirmation bias.

Reproducibility & Environment

Reproducing the Analysis

The full code and workflow are available here:
GitHub Repository: https://github.com/NinerNay/ITCS-4122-5122-Fall-2025

Required Environment

The analysis can be run in:

- Jupyter Notebook
- Google Colab
- Any Python environment with the following libraries installed:

Required Libraries

- Pandas
- Matplotlib
- Seaborn
- NumPy
- Plotly Express

Basic Python knowledge is sufficient to reproduce all steps.

Ethics & Transparency

Ethical Considerations

- **Data completeness:** Missing values and inconsistent reporting may cause bias in comparisons between countries.

- **Interpretation:** Temperature anomaly trends are complex; visualizations simplify large-scale climate dynamics and may not capture localized climate phenomena.

- **Responsible communication:** Care was taken to avoid overstating findings or implying causation where only correlation was observed.

AI Transparency Statement

AI tools (ChatGPT) were used in:

- Generating certain plot ideas

- Improving clarity of explanations

- Assisting with some feature-engineering logic

- Editing and refining report wording

All code logic was either written manually or reviewed extensively to ensure understanding and correctness.