Project for CSPB 4502: Data Mining

Jonathan C. Trull

World Climate Trends

Team Members – Group 11

Jonathan Trull (sole member)

Project Description

In this project, I will analyze historical climate data to identify regional temperature trends and assess the impact of population growth and associated urbanization on rising temperatures. The study will explore correlations between urbanization and climate change. The findings will help uncover patterns in temperature increases and provide insights into how human activities contribute to global warming. Conversely, I will analyze if regional changes in temperature lead to people migrations associated with the changing climate.

- How have historical temperature trends varied by region, and what are the most significantly warming areas?

- Is there a correlation between urbanization (e.g., populations growth, development) with rising temperatures?

- How do temperature changes align with economic growth indicators across different countries?

- How do temperature increases influence population change and migration patterns?

Prior Work

- Large weather and conflict effects on internal displacement in Somalia with little evidence of feedback onto conflict - https://www.sciencedirect.com/science/article/pii/S0959378023000079

- Climate change and migration: A review and new framework for analysis - https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.886

- Evaluating contributions of urbanization and global climate change to urban land surface temperature change: a case study in Lagos, Nigeria

- Urbanization contributes little to global warming but substantially intensifies local and regional land surface warming - https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021EF002401?utm_source=chatgpt.com

Data Sets

- NOAA Global Historical Climatology Network (GHCN) - https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily
  - Source: U.S. National Centers for Environmental Information
  - Downloaded

- NASA Global Surface Temperature Analysis (GISTEMP) - https://data.giss.nasa.gov/gistemp/
  - Source: U.S. National Aeronautics and Space Administration
  - Downloaded

- World Bank Open Data on Population Estimates and Projections - https://datacatalog.worldbank.org/search/dataset/0037655/Population-Estimates-and-Projections
  - Source: World Bank Group
  - Downloaded

- NASA Nighttime Lights Data - https://earthobservatory.nasa.gov/features/NightLights
  - Source: U.S. National Aeronautics and Space Administration
  - Downloaded

Proposed Work

- Data Collection & Extraction
  - Download datasets
  - Extract variables needed to answer research questions (Climate, Urbanization, & Socioeconomic Data)

- Data Cleaning
  - Handle missing data
  - Remove duplicates and inconsistent data if applicable
  - Standardize on measurement units and data format

- Data Transformation
  - Aggregate daily climate data into monthly and yearly averages
  - Create derived data for research questions – temperature anomaly, urbanization index, urban heat island index

- Data Integration
  - Merge datasets and align fields
  - Align datasets based on common timeframes

- Exploratory Data Analysis
  - Check for trends and outliers
  - Look for changes over time
  - Perform statistical regression on dependent variables
  - Visualize results

Analysis Toolkit

- **Interactive Development Project Environment**: Jupyter Notebook

- **Data cleaning, pe-processing and analysis toolkit**: Python3, Pandas, Numpy, Scikit-learn

- **Data visualization toolkit**: seaborn, matplotlib

Evaluation of Results

- **Validity of Records –** 90% or greater of records are valid and complete

- **Root Mean Squared Error** – as low as possible

- **Correlation Strength (Pearson's Correlation Coefficient)** – as close to 1 as possible

Subjective Evaluation: I will evaluate if the findings make sense with prior work and with the real world. For example, do known phenomena explain the correlation between the variables? Are there anomalies that would call into question the overall conclusions? For example, are there regions with increasing temperatures that do not have decreasing populations? Are there increasing levels of urbanization that do not have increasing levels of temperature?