Climate Change and Mortality

Natasha Soldin
Auston Balwinski

Motivation

Climate change is a complex global crisis. Understanding the nature of its effects is important to communicate if solutions are to be sought (Lee, et al.). This report is an effort to explore the nature of the relationship of climate change driven mortality.

The wide-ranging and complicated measurements of climate change puts limits on the scope of this project. As a representative measurement, we use annual averages of temperature, maximum temperature, and precipitation. Mortality is our dependent variable, detailed in number of deaths for numerous categories of causes, ages, and sex.

Extensive research has been conducted on the diverse impacts of climate change. Temperature is known to have an effect on mortality in unequal ways across demographics (Calleja-Agius, et al.). This report affirms the existence of this relationship and details the effect on demographic and mortality causes. While we find this relationship is significant, temperature variance associated with climate change is a small factor when compared to absolute temperature. Further, we detail the magnitude of each climate variable's effect on age groups and causes of death.

Data Sources

Climate

Climate data is gathered from the Climatic Research Unit Gridded Time Series (CRU TS) from the University of East Anglia. This dataset is composed of spatially aggregated climate data. For our interests we have extracted the variables as yearly averages by country for precipitation in millimeters, surface air temperature in degrees celsius, and maximum surface air temperature in degrees celsius. Each variable is collected in its own dataset, which contains a row entry for each 246 countries and a column for each year: 1901 to 2023.

Data is accessed through The World Bank Group's (WBG) Climate Change Knowledge Portal, a hub for climate related data. The website is a GUI catalog where the source, variables, and aggregation can be requested by the user. Data is then downloaded in CSV format.

Mortality

The secondary dataset in this project is the World Health Organization (WHO) Mortality Database containing mortality statistics collected from national civil registration systems worldwide. The database spans from 1950 to present, covering 227 countries and territories worldwide.

It documents causes of death using International Classification of Disease (ICD) coding frameworks, with ICD-10 being the most recent classification system. Deaths are also categorized by detailed age groups (infant, childhood and adult age bands at regular intervals up to 95+ years) and sex (male, female or unspecified).

Population

Population data is from WBG's World Development Indicators database housed on the WBG's DataBank website. The dataset contains yearly populations for 1989-2023 by 217 countries. The website is a GUI where the source, variables, and aggregation can be requested by the user. Data is then downloaded in CSV format.

Data Limitations

The mortality dataset is limited by coverage, completeness, and quality issues. Data availability varies substantially across countries and time periods, with some nations not reporting to the WHO at all. The database depends entirely on national civil registration systems, creating significant gaps in regions with underdeveloped health infrastructure. Many countries experience underreporting due to inadequate recording systems, shortage of medical certifiers for death certificates, and politically motivated concerns.

Additionally, the WHO publishes data exactly as submitted by countries without adjusting for incompleteness or quality variations, meaning the database reflects the varying capabilities of national surveillance systems. There are typically significant reporting delays, as shown in the missing data rates from recent years, as countries compile and verify their records before submission.

Data Manipulation Methods

Climate Dataset

The climate datasets were processed (*Figure 3.1*) from three separate CSV files containing mean temperatures, maximum temperatures, and precipitation data from the CRU time series.

Methods:

- **Data Transformation and Restructuring:** The datasets required restructuring from wide format, where years were represented as column headers, to long format suitable for analysis (i.e. where each row was a country-year combination). This transformation involved cleaning column names to remove formatting inconsistencies, converting each dataset from wide to long format and merging all three climate variables into a comprehensive dataset with mean temperature, maximum temperature, and precipitation measurements for each country-year observation.
- **Handling Duplicates:** During validation checks, a critical data quality issue emerged in that duplicate country-year combinations existed due to overseas territories sharing names with their parent countries. For example, France appeared multiple times for the same year because French territories such as Guadeloupe and French Guiana were listed under the same country name as metropolitan France. To resolve this issue, a comprehensive territory mapping system was implemented to distinguish between mainland countries and their territories. Countries with overseas territories were renamed to clearly identify the specific geographic

entity, such as "France (Metropolitan)" versus "France (Guadeloupe)". This approach preserved the distinct climate characteristics of each territory while eliminating data duplication.

- **Climate Metric Calculations:** Historical baseline averages were calculated for the period 1950-1980 for temperature, maximum temperature and precipitation by country. Climate differentials were then computed as the differences between actual measurements and these baseline averages, enabling analysis of climate change patterns relative to historical norms.
- **Population Data Integration:** The WBG population dataset was chosen over the WHO population dataset due to coverage regarding the country-year requirements. It required preprocessing including removal of year indicator suffixes from column headers, elimination of metadata rows and transformation from wide to long formatting. This cleaned population dataset was then merged with the climate dataset to enable the calculation of population-adjusted metrics.

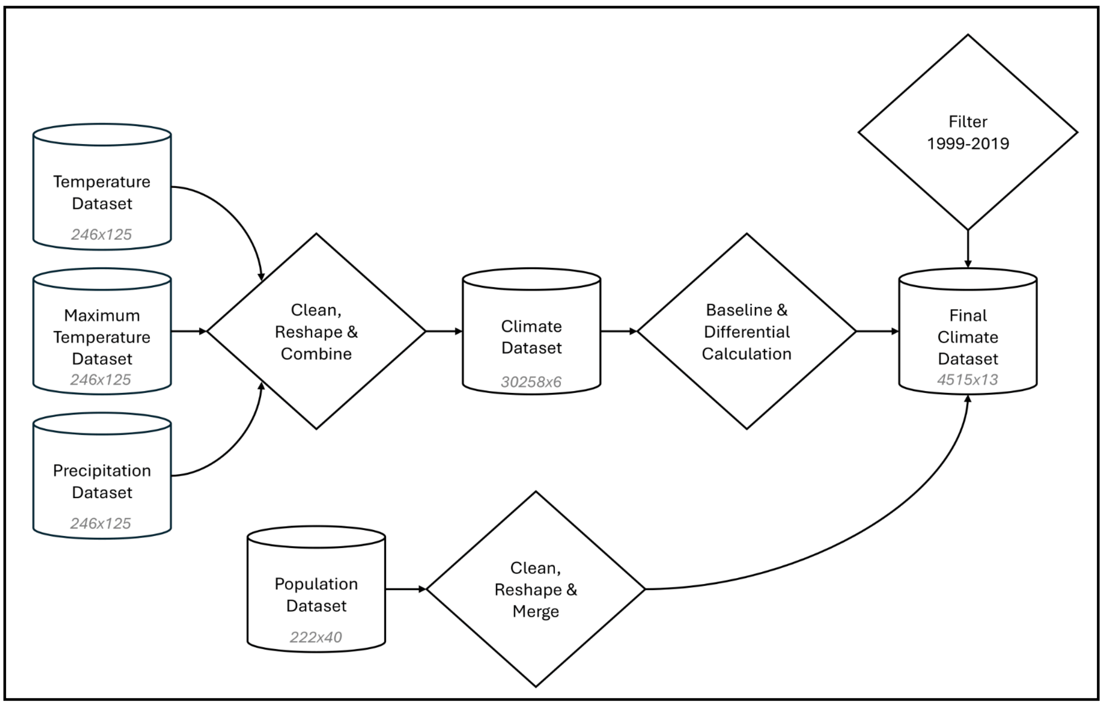

Figure 3.1: Climate Dataset Processing Flow Diagram

The final processed climate dataset contains unique country-year records with mean temperature, maximum temperature, precipitation measurements, baseline averages, differential measurements and population statistics.

Mortality Dataset

The mortality datasets were processed (*Figure 3.2*) from six separate WHO mortality files containing detailed death records by country, year, sex, age groups, and cause of death.

Methods:

- **File Combination:** The six mortality files were systematically loaded and combined into a single dataset. This process included file validation with case-insensitive matching, column structure verification across all files, and duplicate removal after concatenation to ensure data integrity.
- **Cause of Death Re-categorization:** The original detailed ICD-10 cause codes were mapped to broader aggregated categories using a comprehensive dictionary system. This involved converting granular cause codes (e.g. specific disease subcategories) to higher-level groupings (e.g. all cancers, all cardiovascular diseases) to enable meaningful analysis across cause categories.
- **Age Group Binning:** Death records across multiple narrow age bands were consolidated into broader, analytically useful age groups (e.g. Ages_0, Ages_1_4, etc.). This aggregation simplified the dataset structure while preserving age-related mortality patterns for analysis.
- **Data Aggregation and Restructuring:** The detailed records were aggregated to create a unique country-year row format suitable for analysis. This involved grouping deaths by country, year, sex, and cause, then creating separate breakdowns for sex-specific mortality, cause-specific mortality, and age-specific mortality. These breakdowns were then merged into a comprehensive dataset with each row representing a unique country-year combination. This process was carefully validated to ensure the same death figures per category that existed in the raw datasets remained intact.
- **Country Name integration:** Country codes were mapped to readable country names by merging with a country reference dataset.

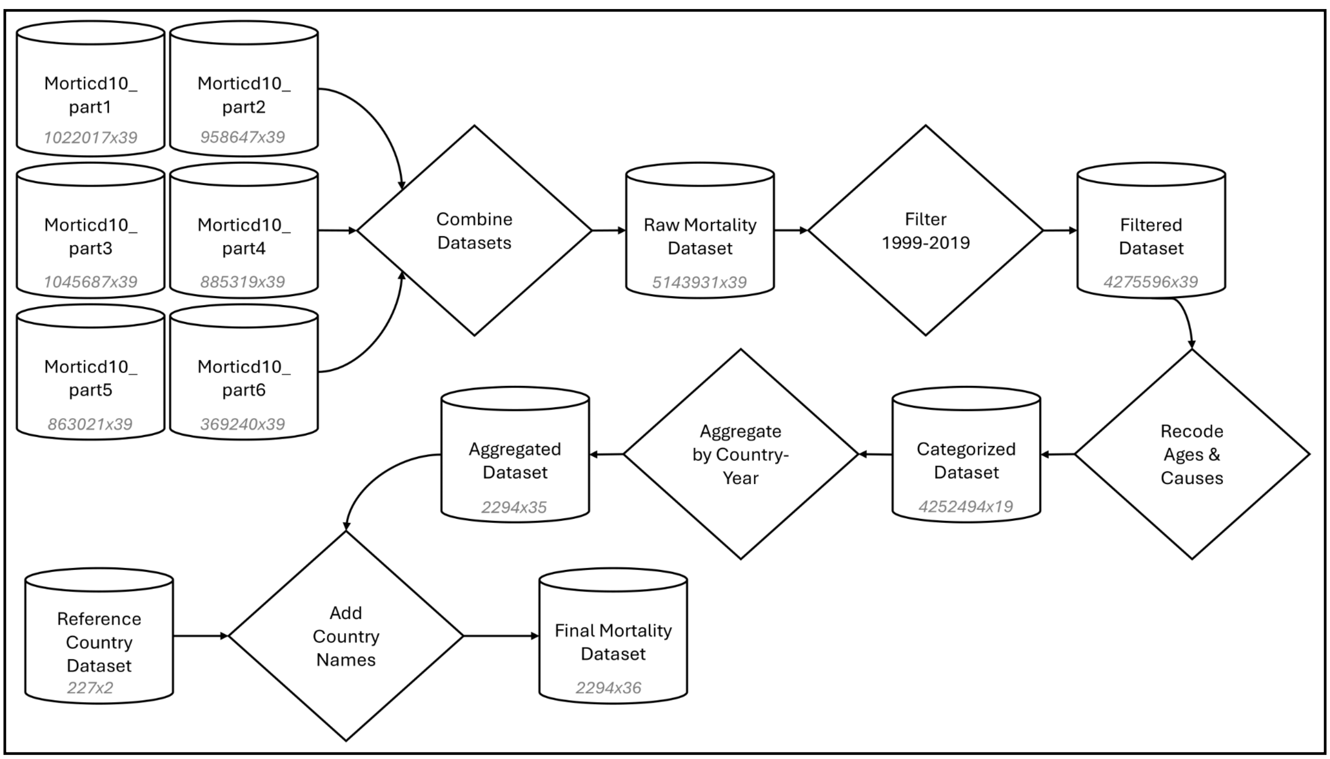

Figure 3.2: Mortality Dataset Processing Flow Diagram

The final processed mortality dataset contained unique country-year records with comprehensive mortality breakdowns by age, sex and cause of death along with country population data, ready for integration with climate data.

Combined Dataset

These two datasets were then combined (*Figure 3.3*) based on unique country-year combinations to create a final combined dataset.

Methods:

- **Country Name Harmonization:** Initial analysis revealed that 146 countries existed in the mortality dataset and 239 in the climate dataset, only 111 countries had exact name matches between the datasets. This left 35 countries in the mortality dataset, the limiting dataset, requiring manual mapping to their corresponding climate dataset names. A comprehensive mapping dictionary was created to resolve naming differences. Additionally, territories with parenthetical notations in the climate dataset (e.g., "Anguilla (U.K.)") were systematically mapped to their mortality dataset equivalents.
- **Filtering and Merging:** To ensure data quality, both datasets were filtered to retain only countries that could be successfully matched through either exact matches or manual mapping. This filtering process reduced the mortality dataset from 2,294 to 2,217 observations and the climate dataset from 4,515 to 2,730 observations. A left join was then performed on the standardized country names and years, preserving all mortality data while adding climate variables where available. The final merged dataset contained 2,086 country-year observations spanning 130 unique countries.
- **Death Rate Calculation:** A death rate variable was calculated as the ratio of total deaths to total population, enabling population-adjusted mortality analysis.
- **Logarithmic Transformation:** Distributions of death counts for all age, cause, and sex variables are highly right-skewed. Log transformations were taken of the variables to allow for more accurate and statistically sound comparison. These transformations help conform the variables to a more normal distribution.

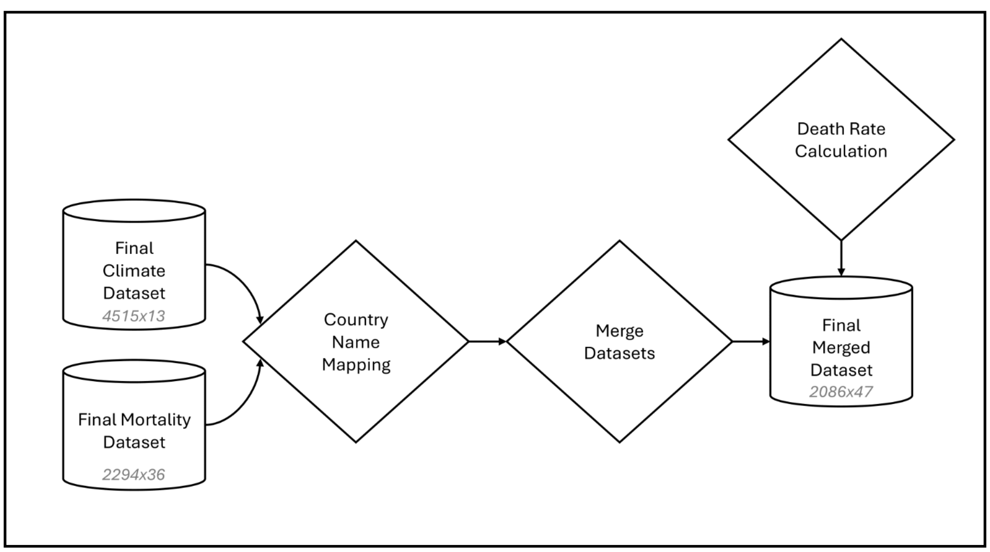

Figure 3.3: Combined Dataset Processing Flow Diagram

- **Data integrity validation:** Procedures were implemented to ensure completeness of the final dataset. A missing data assessment was performed to find that the final dataset achieved complete data coverage with zero missing values across all 47 variables (i.e. although there did exist missing data in the individual datasets, once combined, those entries fell away due to their lack of corresponding data in the other dataset), thus elimination the need for imputation. A duplicate investigation was also conducted to ensure the absence of duplicate country-year combinations in the final dataset. Finally, data integrity validation was performed by tracking specific country-year combinations (e.g., Andorra 2019 with 301 total deaths) through each stage of the data processing pipeline. This validation confirmed that mortality figures were preserved accurately from the raw WHO datasets through all transformation steps to the final analytical dataset.

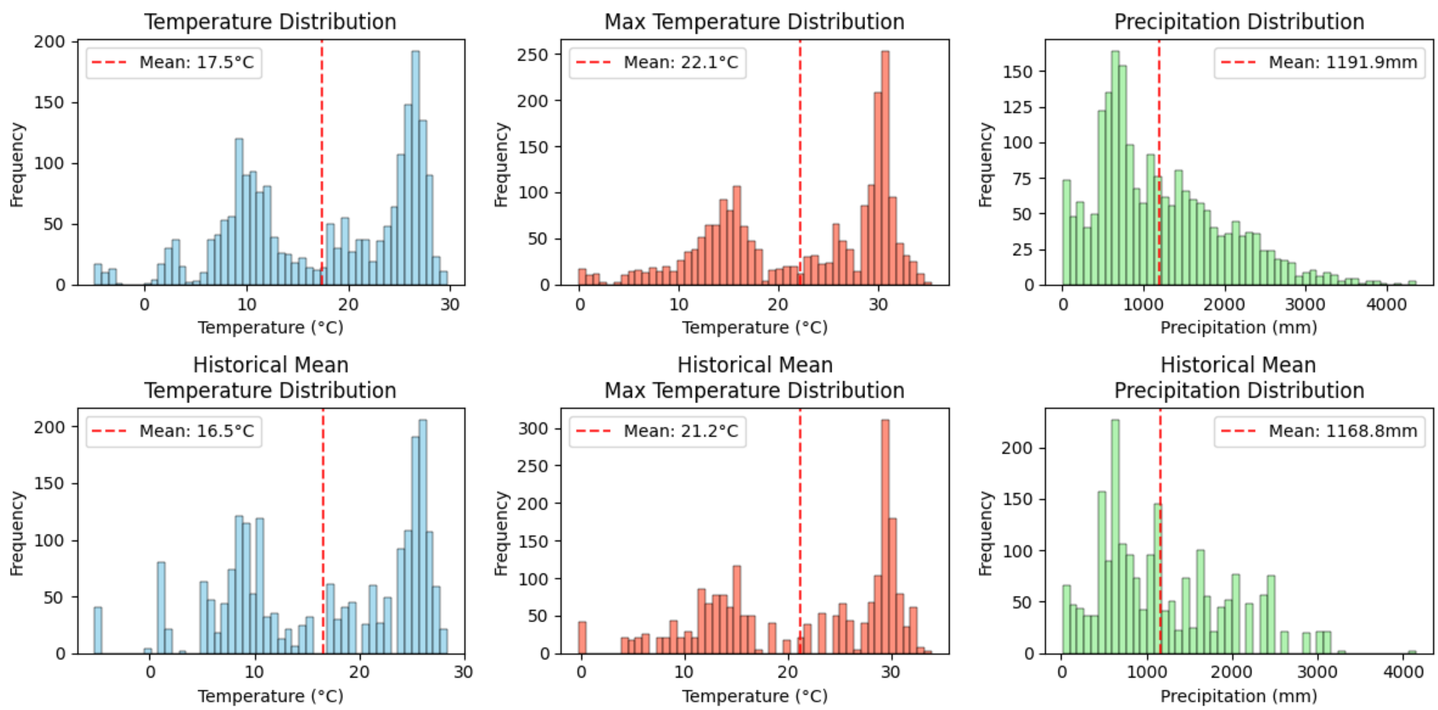

Figure 4.1: Histogram Distributions of Climate Variables

Analysis

Exploratory

Initial exploration involved examining the distribution of key variables through descriptive statistics to understand the central tendencies and variability of climate measures. Histograms (*Figure 4.1, top row*) were generated for temperature, maximum temperature, and precipitation variables to assess distribution characteristics. Compared to historical baselines (*Figure 4.1, bottom row*) for these variables we can see similarities in distribution. To confirm the

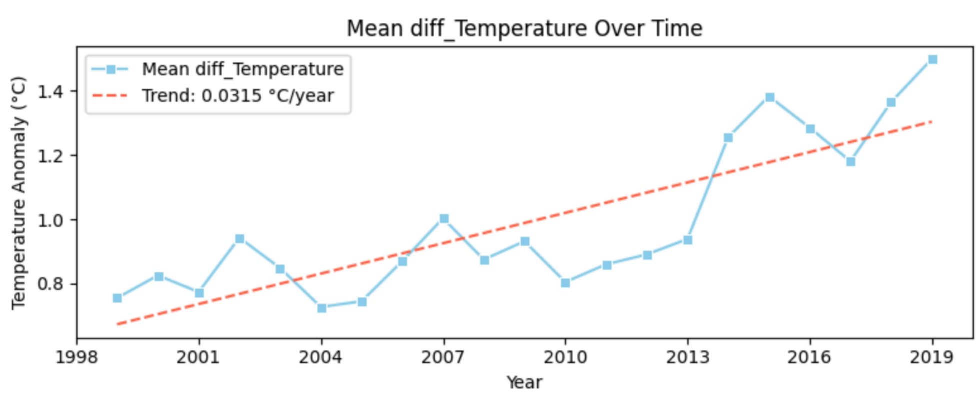

Figure 4.2: Means of Temperature Differential from Historical Baseline Average

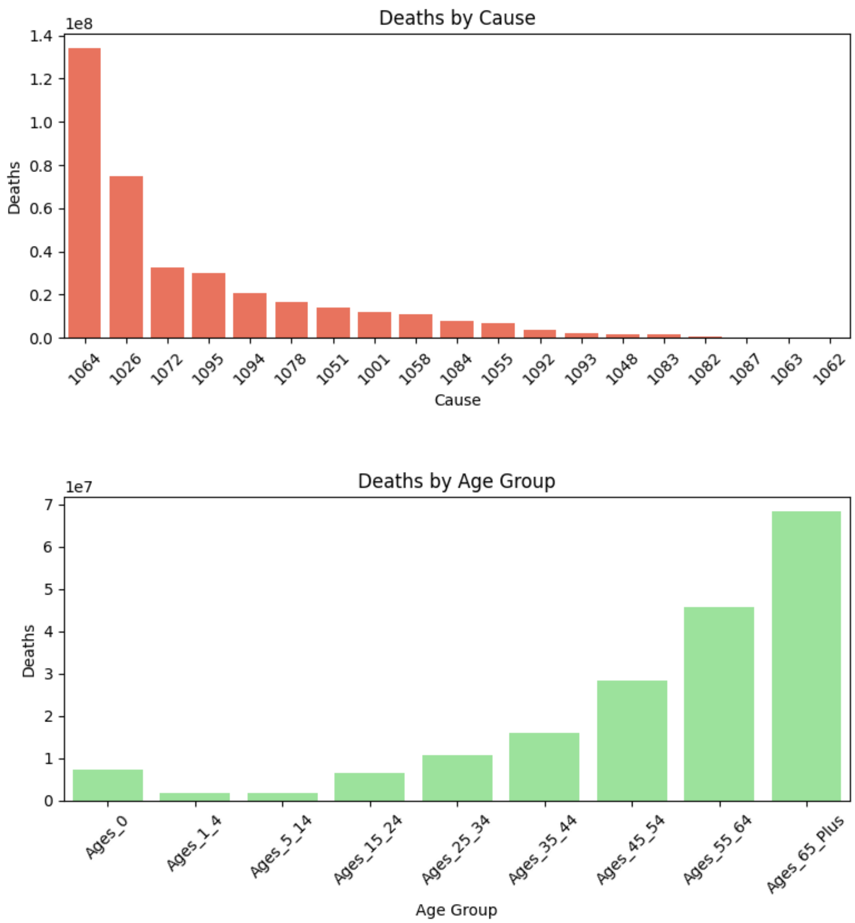

Figure 4.3 and 4.4: Total Deaths by Cause and Age Group, respectively

existence of increasing temperatures, a plot and trend line (*Figure 4.2*) was generated using the difference between temperature and historical baseline temperature (temperature means of years 1950-1980). The differential temperature data shows an positive yearly trend of 0.0315 degrees celsius. This trend was missing from the original temperature data but that could be attributed to missing data interfering with progressive trend.

Mortality-related variables were then examined, focusing on age and cause of death. The categorical plots (*Figure 4.3 & 4.4*) reveal the expected pattern of higher mortality in older age groups, with the 65+ category accounting for the largest proportion of deaths globally. A notable deviation from this trend was observed in the <1 year age group, reflecting elevated infant mortality rates,

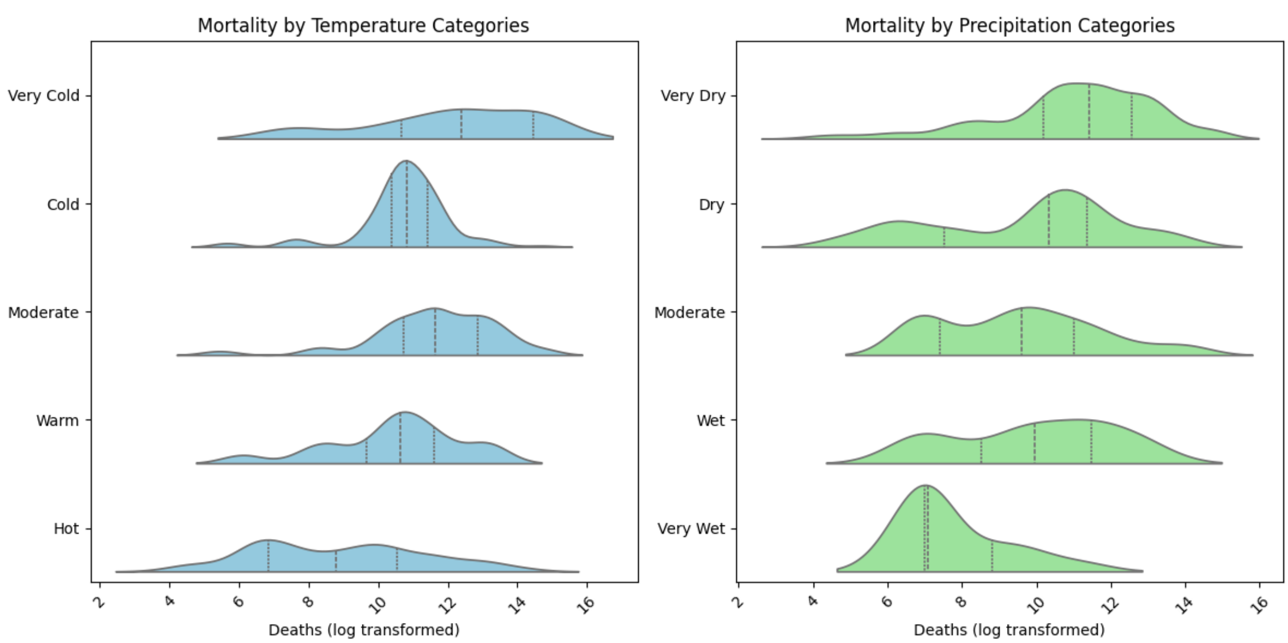

Figure 4.5 and 4.6: Mortality by Temperature and Precipitation Categories, respectively

highlighting two distinct at-risk populations. Analysis of cause-specific mortality identified circulatory system diseases (Cause_1046) and neoplasms (Cause_1026)(e.g., cancer) as the leading causes of death in the dataset.

To further examine patterns across climate extremes, the climate variables temperature and precipitation were binned into categorical ranges to create interpretable categories. Split violin plots (*Figure 4.5 & 4.6*) were generated to examine mortality distributions across these climate categories, revealing systematic patterns in how extreme climate conditions relate to mortality outcomes.

Variable Relationships

Correlation analysis examining climate-mortality relationships across age groups and causes of death categories reveal distinct vulnerability patterns that provide insight into which populations and health conditions are most climate-sensitive.

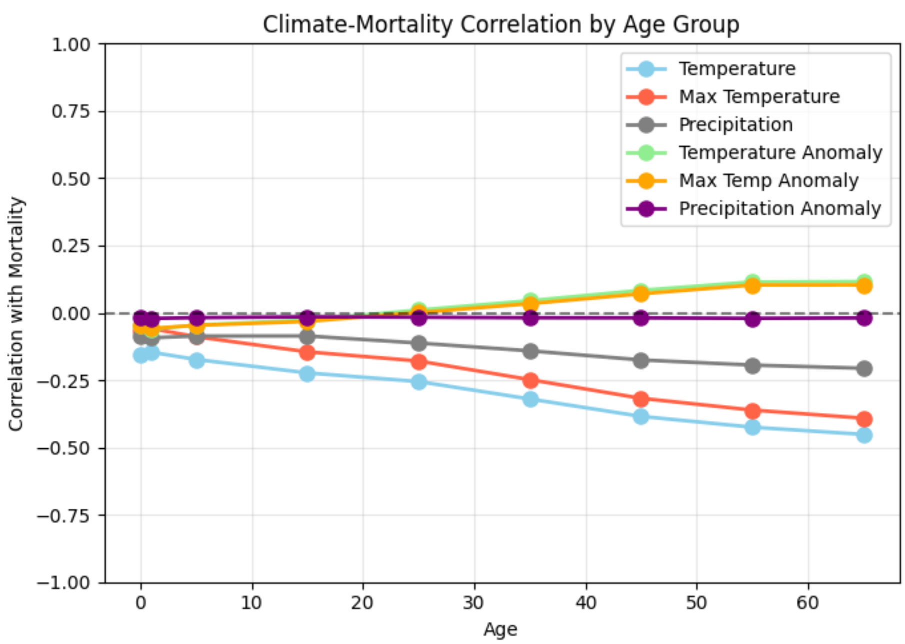

Figure 4.7: Climate-Mortality Correlation by Age Groups

Age-specific analysis (*Figure 4.7*) demonstrates clear vulnerability gradients, with elderly populations showing the strongest negative temperature correlations (-0.45), indicating higher mortality corresponding with colder temperatures. While small, there is a notable positive correlation of temperature differential with mortality in groups older than 25. These effects are statistically significant (p-values <0.05), but marginally influential with an overall correlation score of +0.09 for all ages. Though it does indicate a slight climate change related effect on mortality by age.

Cause-specific analysis (*Figure 4.8*) reveals substantial heterogeneity in climate sensitivity across disease categories. Overall correlations suggest colder temperatures have a bigger impact on mortality than does the impact from temperature differentials. As with age, precipitation differential is the least meaningful to mortality causes.

Categorically, neoplasms (Cause_1026), mental and behavioral disorders (Cause_1055), and circulatory system diseases (Cause_1064) exhibit the strongest climate sensitivities with temperature correlations around -.50, indicating higher mortality from these conditions in colder temperatures. When looking at differential temperatures we see an opposite effect, though of much smaller magnitude with an overall score of +0.13.

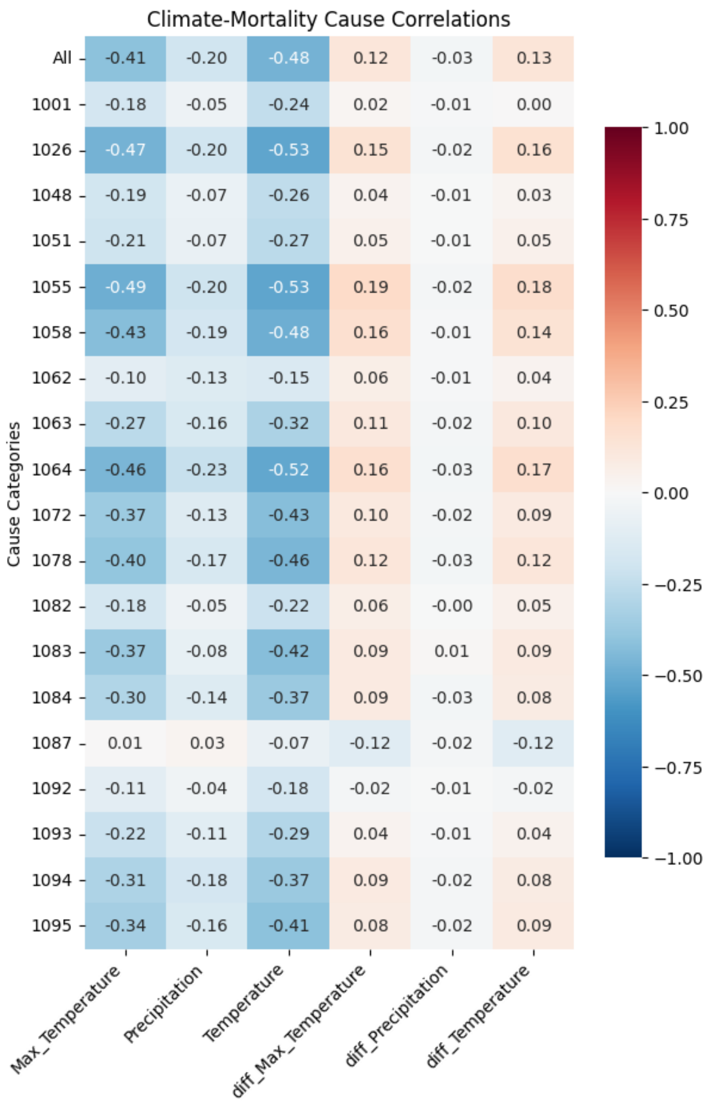

Figure 4.8: Climate-Mortality Cause Correlation Heatmap

Most categories have correlations that indicate temperature increases do have a small relation to mortality across most causes. Notably, maternal mortality (Cause_1087) shows essentially no correlation with temperature and a negative differential temperature correlation of -0.12. Confusingly, this indicates a maternal mortality reduction with increasing temperature.

Conclusion

All age groups and most mortality cause categories show stronger correlations with absolute temperature than with climate anomalies or precipitation, suggesting that absolute temperature patterns are more important for health outcomes than the gradual year-to-year temperature increase from climate change. Cold temperatures have an outsize impact on mortality. However, it is difficult to conclude that warming temperatures will improve mortality rates. Mortality correlations with temperate differentials could indicate a vulnerability to changing climate (Lee & Dessler).

This analysis should serve as direction for further study into the nature of climate change vulnerability. It has been suggested that a population's climate adaptability is more important to mortality rates than absolute temperatures (Lee & Dessler). There are many possible confounding variables to examine such as economic and developmental factors (Calleja-Agius, et al.).

Concerningly, the trend of increasing temperatures shows no evidence of abating (Lee, et al.). Every country in our dataset has experienced rising temperatures. It may be the case that the small, but significant mortality-differential temperature relationship grows over time as the temperature differential widens across the globe.

Analysis Limitations

Data Aggregation Challenges

Multiple levels of data aggregation introduced significant limitations. Annual climate averages masked critical seasonal variations and extreme weather events that likely drive climate-mortality relationships, potentially obscuring heat waves, cold snaps, and other short-term climate impacts on health. The data integration process required substantial filtering, excluding countries without matching climate and mortality records, which reduced datasets by 22-40% (mortality: 2,294→2,217; climate: 4,515→2,730 observations). This selective retention created potential bias toward developed countries with robust data collection systems while excluding regions that may exhibit different climate-health relationships. Additionally, country-level aggregation eliminated important sub-national heterogeneity including urban-rural disparities, altitude effects, and regional climate variations that could reveal more nuanced vulnerability patterns.

Confounding Variable Concerns

The analysis could not adequately control for critical socioeconomic factors, healthcare system quality, or climate adaptation measures that likely mediate climate-mortality relationships. Countries with advanced healthcare infrastructure, air conditioning availability, or effective early warning systems may show weaker climate-mortality associations due to protective factors rather than genuine climate insensitivity. Economic development levels, urbanization patterns, and public health interventions vary substantially across countries and time periods, potentially confounding observed climate-health relationships. Additionally, changing population demographics over time, including aging populations and shifting disease burdens, could confound temporal trend analysis as countries may show mortality changes independent of climate factors.

Statistical Methodological Approach

The analysis relies primarily on correlation-based approaches and could not establish causal relationships between climate variables and mortality outcomes. The focus on linear relationships may have missed important nonlinear climate-health relationships, and the analysis did not account for potential time lag effects where climate conditions might influence mortality in subsequent years.

Statement of Work

This project was conceptualized, designed, and outlined in mutual development through collaborative discussions. Division of work felt equal for both partners and communication were clear and consistent. Both partners contributed to coding notebook organization, editing, and commenting, including on each others' work. Visualizations were a mix of both individual and collaborative conception and formatting. Final report responsibility was divided into sections except for jointly writing the *Analysis* sections. Suggestions for improvement in future collaborations include planning more detailed project outline, timelines, expectations, and responsibilities. Researching and discussing more about the data sources to understand the capabilities and limitations of the data would be productive in future projects to hopefully better avoid complications later in the project.

- **Natasha Soldin:** Organized and created collaborative project tools. Data intake and cleaning, including aligning naming conventions. Joined raw data into combined DataFrames and transformed to aggregate values. Prepared tests for ensuring data uniformity and quality throughout the intake and manipulation processes. Created visualizations for data coverage, interactive exploratory comparison graphs, interactive maps for country-level metrics, and data flow diagrams detailing data intake and processing. Primary author of report sections: *Data Sources: Mortality*, *Data Limitations, Data Manipulation Methods*, and *Analysis Limitations*.
- **Auston Balwinski:** Acquired climate and population datasets. Data manipulation efforts to create climate baselines and differentials, death rate, and logarithmic transformations, and categorizations for mortality causes and binning age groups. Created visualizations for logarithmic distributions, age and cause mortality distributions, temperature trend line graphs, and age correlation graph. Primary author of report sections: *Motivation*, *Data Sources: Climate*, *Conclusion,* and *Statement of Work*. Formatted final report document, fit visualizations, and created references.

Data Source Links

WBG Population Dataset
WBG Precipitation Dataset
WBG Maximum Temperature Dataset
WBG Maximum Temperature Dataset
WHO Mortality Datasets

References

Calleja-Agius, Jean et al. "The effect of global warming on mortality." Early human development vol. 155 (2021). https://doi.org/10.1016/j.earlhumdev.2020.105222

Lee, Hoesung, et al. "Climate Change 2023 Synthesis Report," 2023, Intergovernmental Panel on Climate Change. iicc.ch. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf

Lee, J., & Dessler, A. E. "Future temperature-related deaths in the U.S.: The impact of climate change, demographics, and adaptation." GeoHealth, 7, e2023GH000799. (2023). https://doi.org/10.1029/2023GH000799