A Statistical Analysis of the Relationship Between Global Warming and Flood Frequency (1985-2021)

BIRMINGHAM CITY UNIVERSITY
Faculty of Computing, Engineering and the Built Environment

Asfand Yar
25110740

Word Count: 2307

August 25, 2025

Contents

1 Executive Summary

This report presents a statistical investigation into the relationship between global climate change and flood patterns. Using techniques in applied statistics this project analyzes two key datasets: NASA's global temperature anomaly records and the Dartmouth Flood Observatory's global flood archive, from 1985 to 2021.

The core questions from this analysis were: Has there been a significant increase in global temperatures and flood frequency? Is there a statistically significant correlation between them? And does a warming year predict more floods in subsequent years?

The analysis revealed a significant warming trend where global temperatures are increasing by approximately 2.07°C per century. A significant, positive correlation was found between temperature anomalies and flood frequency in the same year. However, time series analysis determined this relationship to be simultaneous. A warm year does not predict floods in the following years.

A counter intuitive finding was also discovered regarding flood severity. While flood frequency correlated with temperature, the number of people displaced showed a significant negative correlation. This suggests that improved global reporting, disaster preparedness, and resilience may be mitigating the human impact of floods over time. The report concludes that while a warming planet is statistically linked to more frequent flooding, the story is complex, influenced by both physical climate mechanisms and evolving socio-economic factors.

2 Introduction

The global climate is changing and not for the better and there are well documented increases in global temperature. Mean surface temperature being a primary indicator(IPCC 2021). One major effect of a warming climate is that it can change how water moves through the environment which may cause more frequent and intense heavy rainfall and flooding events.(Trenberth 2011). Such changes pose a direct threat to communities worldwide through an increased risk of flooding, which is among the most devastating and costly of natural disasters(Committee 2024).

The Dartmouth Flood Observatory(Observatory n.d.) and other databases have compiled extensive records of flood events globally, providing a valuable resource for empirical analysis. It is difficult to clearly prove a strong statistical link between global warming and flood trends.

The purpose of this report is to utilize applied statistical methods to empirically investigate the relationship between global temperature anomalies and flood activity between 1985 and 2021. It moves beyond theoretical models to quantify trends, test correlations, and explore the nature of this relationship using real world data.

2.0.1 Statistical Questions

This report is guided by the following key statistical questions

Q1: Is there a statistically significant increasing trend in global temperature anomalies from 1985 to 2021?

Q2: Is there a statistically significant increasing trend in the annual frequency of reported floods over the same period?

Q3:Is there a significant correlation between the annual global temperature anomaly and the number of floods in that same year?

Q4: Does a relationship exist between temperature anomalies and the severity of flood impacts, measured by fatalities and population displacement?

Q5: Does a warming year predict an increase in flood frequency in subsequent years, or is the relationship contemporaneous?

Q6: Does global temperature anomaly correlate with the severity of floods (fatalities, people displaced)?

Q7: How are the flood impact metrics (fatalities, displacements) distributed, and are there outlier years.

3 Methodology

3.1 Exploratory Data Analysis

Descriptive Statistics was done such as the mean, median, and standard deviation were generated to provide an overview of the central tendency and variability within the dataset.

The **distributions** of the variables were further examined using histograms and boxplots, which helped visualize the spread, identify potential outliers, and assess symmetry.

To quantify the shape of the distributions more formally, **skewness** was calculated using boxplots, providing insight into whether the data were normally distributed or exhibited positive or negative skew.

3.2 Linear Regression

Simple Linear regression was used to quantify trend over time. it tests the null hypothesis that the slope of relationship is zero. using p-value null hypothesis was put to test

3.3 Correlation Analysis

Pearson's product-moment correlation and **Spearman's rank-order correlation** were used.

Pearson correlation is appropriate for assessing the strength and direction of a linear relationship between two continuous, approximately normally distributed variables (e.g., Annual_Anomaly and Flood_Count). Spearman's correlation, a non-parametric method, was employed for variables with high skewness and outliers (e.g., Total_Deaths), because it looks at whether values consistently increase or decrease, not whether they follow a straight-line pattern.

3.4 Time Series Analysis

Augmented Dickey-Fuller test for stationarity and the **Cross Correlation** Function.

Standard correlation assumes that data points are independent but in time series this is often not true because values are related over time. The ADF test helps check whether a time series has a trend. The CCF is used to see if one time series leads, lags, or moves together with another showing whether changes in one variable happen before or after or at the same time as changes in the other.

4 Dataset Description

This study uses two trusted, publicly available datasets that provide long-term global records on climate and disaster events. Combining these datasets is essential for studying how climate change is linked to extreme flood events. A summary of the key information is given below.

4.0.1 NASA GISS Surface Temperature Analysis (GISTEMP)

Source: NASA Goddard Institute for Space Studies (GISTEMP, 2023)
The dataset provides a globally averaged estimate of surface temperature change, expressed as an anomaly relative to a 1951-1980 baseline period. It is a critical benchmark for monitoring global climate change.

This dataset gives the independent variable for the analysis: a reliable measure of global warming. It covers a long period (1880 to 2025) with yearly data, which makes it useful for studying trends. For this study the main variable used was the annual mean anomaly (J.D), as it shows how global temperatures change from year to year and how this may affect weather patterns(GISTEMP Team 2025)(Lenssen et al. 2024).

This dataset is crucial to answer the questions posed by this report such as analyzing global temperature trend and link it to annual flood counts.

Table 1: Summary of the Global Temperature Dataset (GISTEMP v4)

Variable	Type	Units	Description
Year	Integer	Year	Calendar year (1881–2024).
Jan–Dec	Numeric	°C anomaly	Monthly mean temperature anomalies relative to 1951–1980 baseline.
J.D	Character (numeric values)	°C anomaly	Annual mean temperature anomaly (Jan–Dec average).
D.N	Character (numeric values)	°C anomaly	Dec–Nov mean anomaly.
DJF	Character (numeric values)	°C anomaly	Winter mean anomaly (Dec–Feb).
MAM	Numeric	°C anomaly	Spring mean anomaly (Mar–May).
JJA	Character (numeric values)	°C anomaly	Summer mean anomaly (Jun–Aug).
SON	Character (numeric values)	°C anomaly	Autumn mean anomaly (Sep–Nov).
Dataset dimensions: 144 rows (years), 19 variables.			

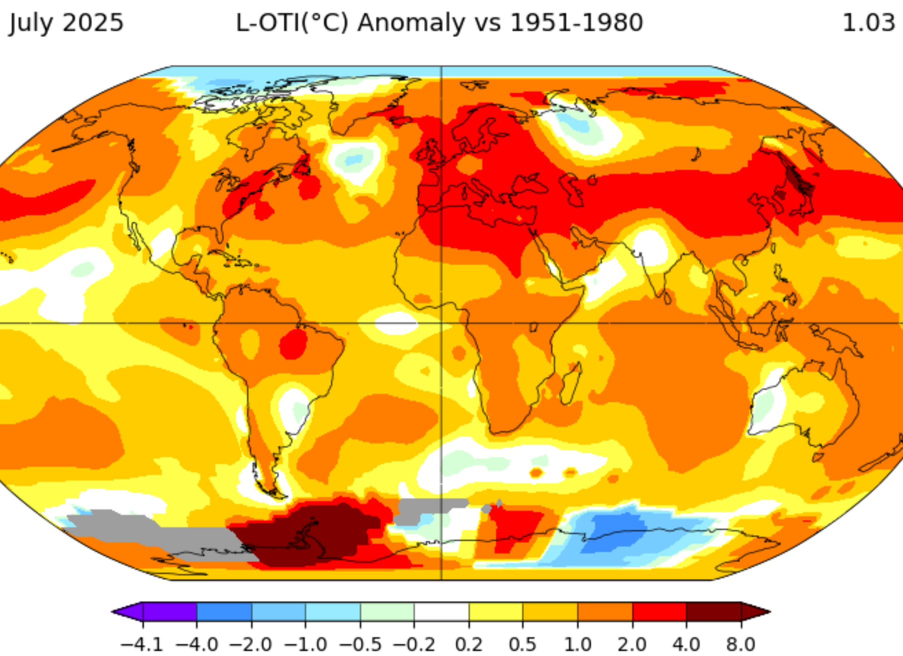

Figure 1: Global surface temperature anomaly map (°C) compared to the 1952–1980 average, based on NASA GISS GISTEMP data (1981–2024)(for Space Studies GISS).

4.1 Dartmouth Flood Observatory (DFO) Global Flood Archive

Source: Dartmouth Flood Observatory (2023)
The DFO database is a global inventory of significant flood events, compiled from news, governmental, instrumental, and remote sensing sources. It records events that meet specific criteria for significant social or economic disruption

This dataset provides the dependent variables for the analysis: standardized metrics of global flood activity. While subject to biases inherent in event reporting (e.g., improved media coverage over time, under-reporting in remote regions), it represents one of the most comprehensive and long-running archives of its kind. For this study, the data was aggregated annually to create consistent count and impact variables (Flood_Count, Total_Deaths, Total_Displaced) compatible with the annual temperature data(Observatory n.d.).

Table 2: Summary of the flood dataset (1985–present).

Variable	**Type**	**Description**
ID	Numeric	Unique event identifier
GlideNumber	Numeric	Global Disaster Identifier (if available)
Country	Character	Primary country affected
OtherCountry	Character	Secondary affected countries (if any)
long	Numeric	Longitude of event
lat	Numeric	Latitude of event
Area	Numeric	Affected area size (km^2)
Began	Date (POSIXct)	Start date of the flood event
Ended	Date (POSIXct)	End date of the flood event
Validation	Character	Source of validation (e.g., News, Official)
Dead	Numeric	Number of deaths reported
Displaced	Numeric	Number of displaced people
MainCause	Character	Reported main cause (e.g., heavy rain)
Severity	Numeric	Event severity (categorical scale)
Year	Numeric	Calendar year of the event

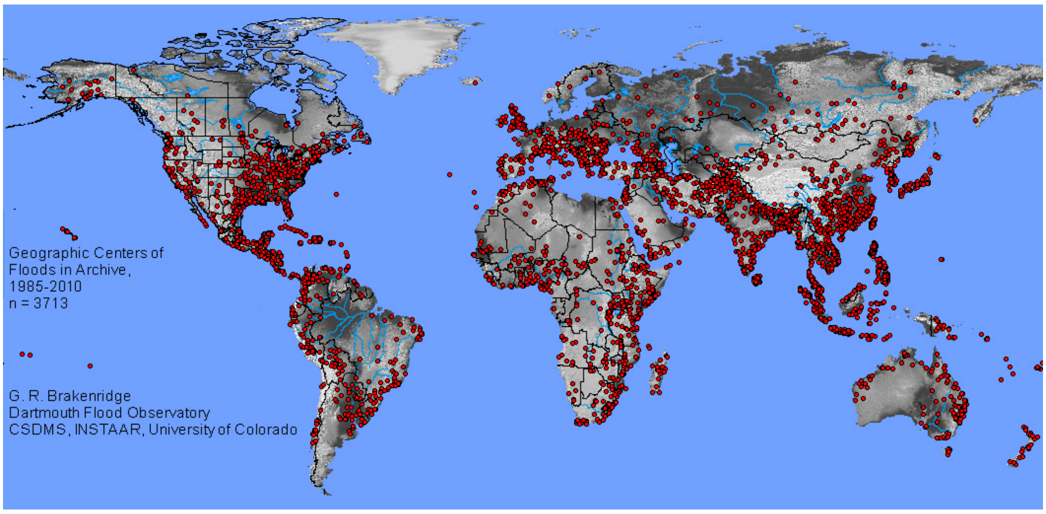

Figure 2: Geographic Centers of floods in the FloodArchive GIS file, 1985-2010(Observatory n.d.)

4.2 Data Pre-processing

Before doing statistical analysis on the dataset it went through cleaning.

4.2.1 NASA GISTEMP Data Processing

Data Loading the csv file was loaded, skipping intial metadata lines to correctly import the column headers.

Column Cleaning The critical J.D column (annual mean anomaly) was stored as a character type due to the presence of "***" placeholders for missing data in early years. Missing values were cleaned and the datatype was changed to numeric and the column name was changed to annual anomaly.

Variable Selection only the Year and the cleaned annual anomaly variable (renamed to Annual_Anomaly for clarity) were selected for analysis. All monthly and seasonal anomaly columns were excluded to focus on the annual trend.

4.2.2 Dartmouth Flood Observatory Data Processing

Temporal Aggregation a new Year variable was created by extracting the year component from the Began timestamp using R's date-time functions.

Annual Summary Calculation the event-level data was aggregated to the annual level using group_by(Year) and summarise(). This created three new key variables for each year

1. Flood_Count: The total number of flood events.
2. Total_Deaths: The sum of the Dead column.
3. Total_Displaced: The sum of the Displaced column.

4.2.3 Data Integration

The final step involved merging the processed temperature and flood datasets.

Merging Procedure: The NASA and DFO datasets were merged on the Year field using an inner_join(). Resulting in a final analytical dataset spanning the period 1985 to 2021.

Table 3: Summary of the Combined Dataset (1985–2021)

Variable	**Type**	**Description**
Year	Numeric	Observation year (1985–2021).
Annual_Anomaly	Numeric	Annual mean global temperature anomaly (°C).
FloodCount	Integer	Total number of flood events recorded in that year.
AvgSeverity	Numeric	Average severity index of flood events in that year.
TotalDeaths	Numeric	Total number of deaths caused by flood events in that year.
TotalDisplaced	Numeric	Total number of people displaced due to floods in that year.

Table(3) shows the final dataset after cleaning both dataset and combining them together from the year 1980 to 2021.

4.3 Time Series Analysis Dataset

For the time series analysis another step was taken for the dataset to work. Data had to be converted to time series objects, Annual_Anomaly and FloodCount had to be converted to time series objects. Using the command ts and frequency set to 1.

5 Results and Discussion

5.1 Exploratory Data Analysis

5.1.1 Distribution of Global Temperature Anomalies

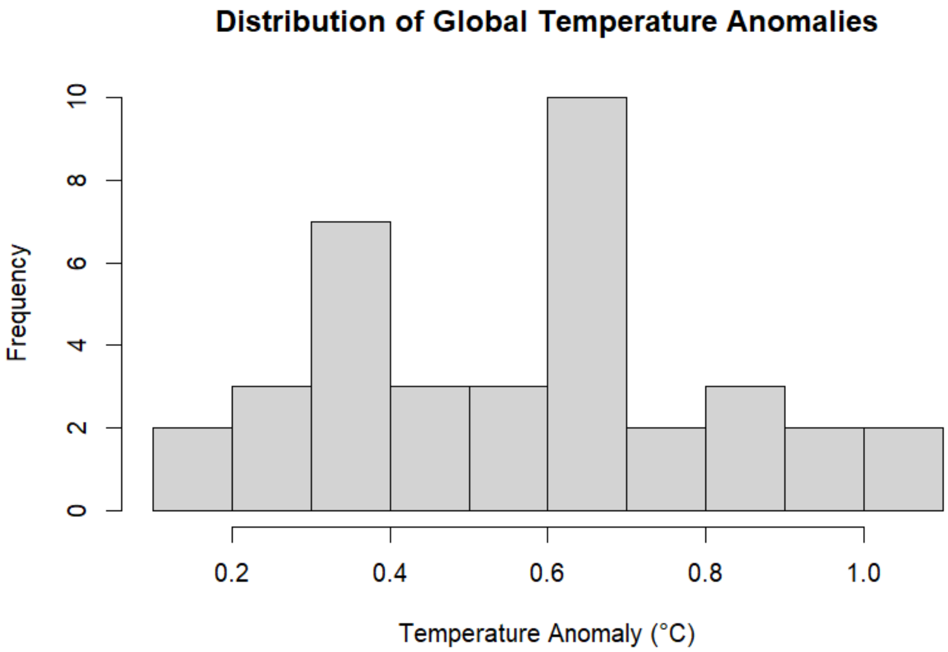

Figure 3: Distribution of global temperature anomalies

Figure(3) shows the distribution of the column Annual_Anomaly from the combined dataset as seen in Table(3). Skewness was calculated and was **0.1402481** this emperically shows that the data is skewed to the right.

5.1.2 Distribution of Annual Flood Count

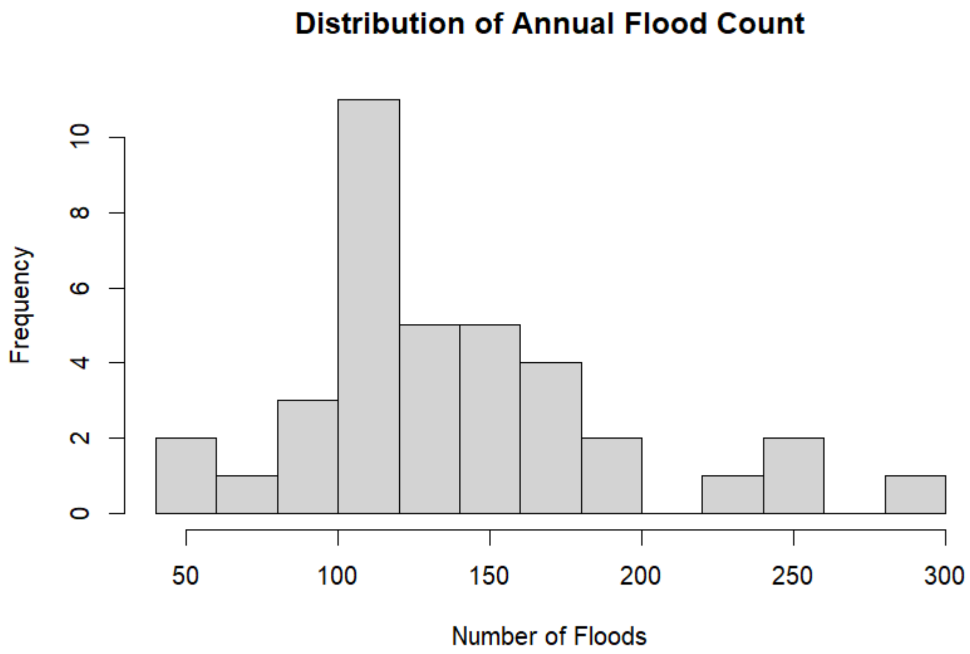

Figure 4: Distribution of Annual Flood Count

Figure(4) skewness was calculated was obtained to be **0.9043698** and is skewed to right.

5.1.3 Distribution of Annual Flood Deaths

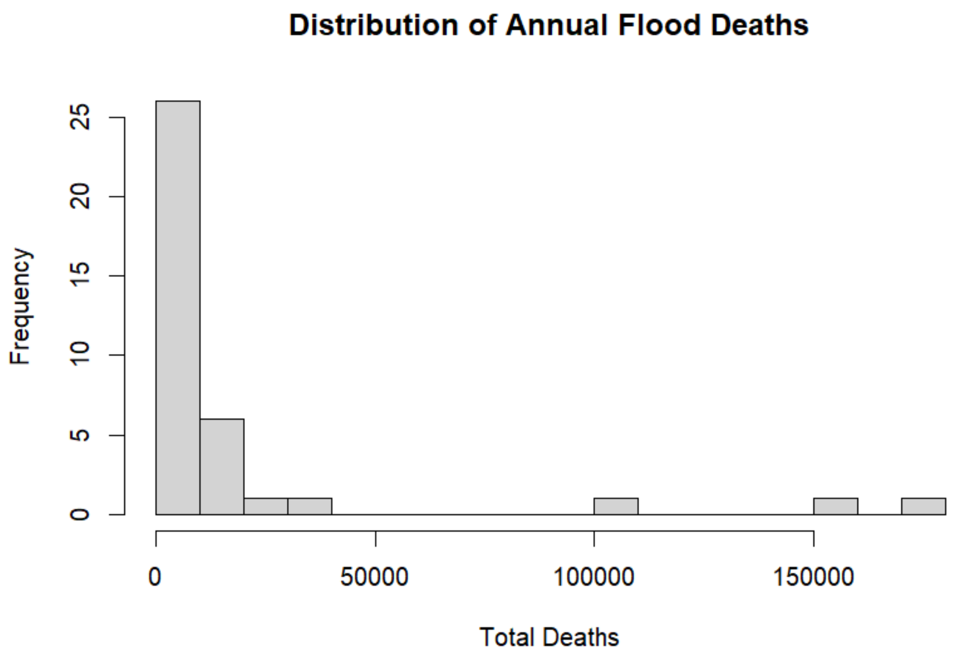

Figure 5: Distribution of Annual Flood Deaths

Figure(5) is highly right skewed and its skewness is calculated to be **3.019159**.

5.1.4 Distribution of Annual Total Displaced

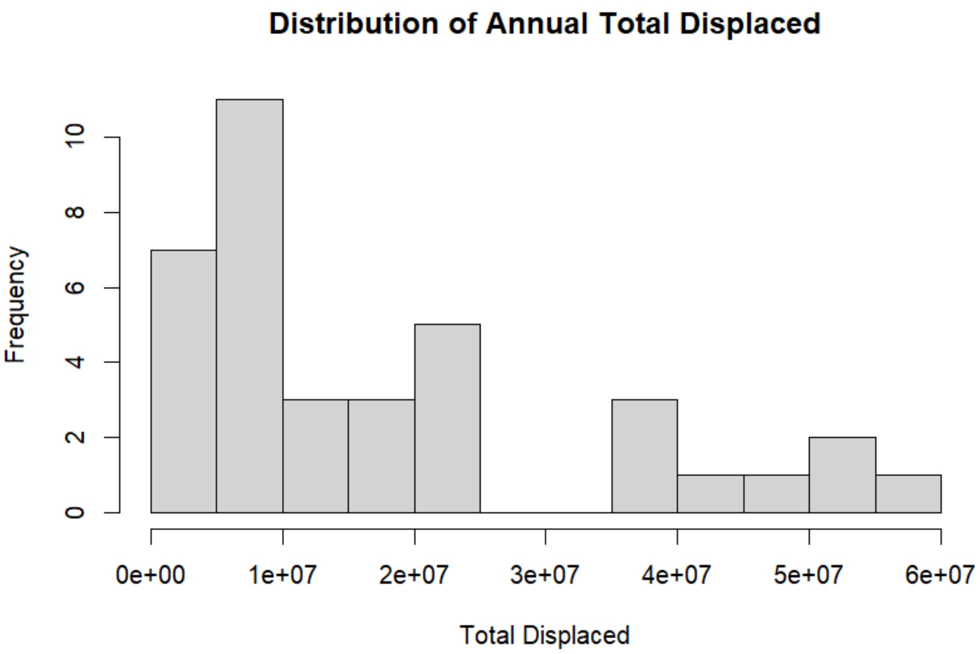

Figure 6: Distribution of Annual Total Displaced

Figure(6) is skewed to the right and its skewness is calculated to be **1.033155**

5.1.5 Outlier Check

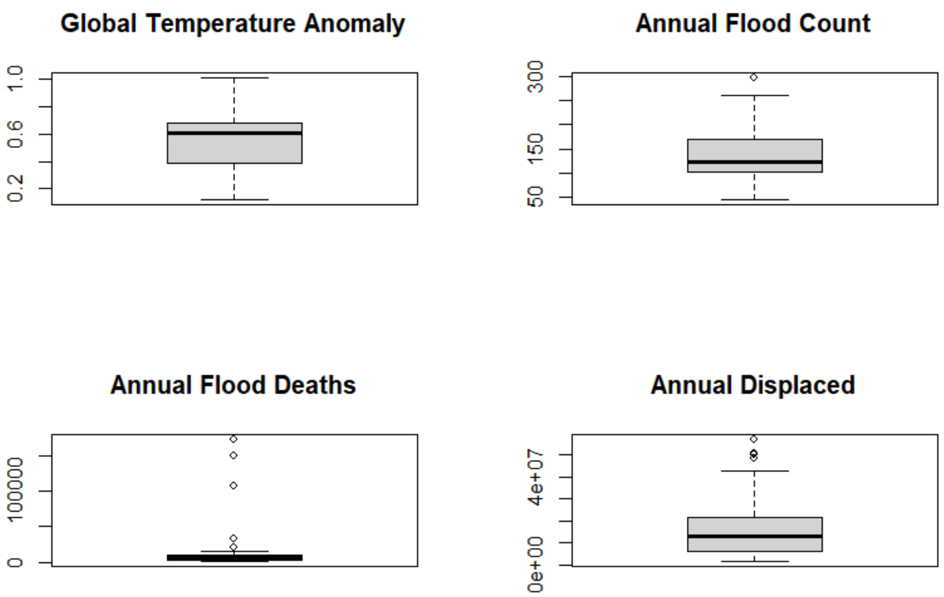

Figure 7: Boxplot for 4 columns to check for outliers

Figure(7) shows box plot of the columns. And Annual Flood Deaths showed the highest distribution of its data.

Table 4: Summary Statistics of Selected Variables in the Combined Dataset (1985–2021)

Statistic	Annual_Anomaly (°C)	FloodCount	TotalDeaths	TotalDisplaced
Min.	0.1200	45	1,177	1,452,016
1st Qu.	0.3900	102	3,218	5,831,709
Median	0.6100	123	6,570	12,814,753
Mean	0.5649	138.6	18,607	17,857,721
3rd Qu.	0.6800	171	10,361	21,678,074
Max.	1.0100	297	173,123	57,099,221

Table(4) empirically shows the distribution of the columns.

5.1.6 Trend in Global Temperature Anomalies

Further diving down into EDA process we were able to identify the trends.

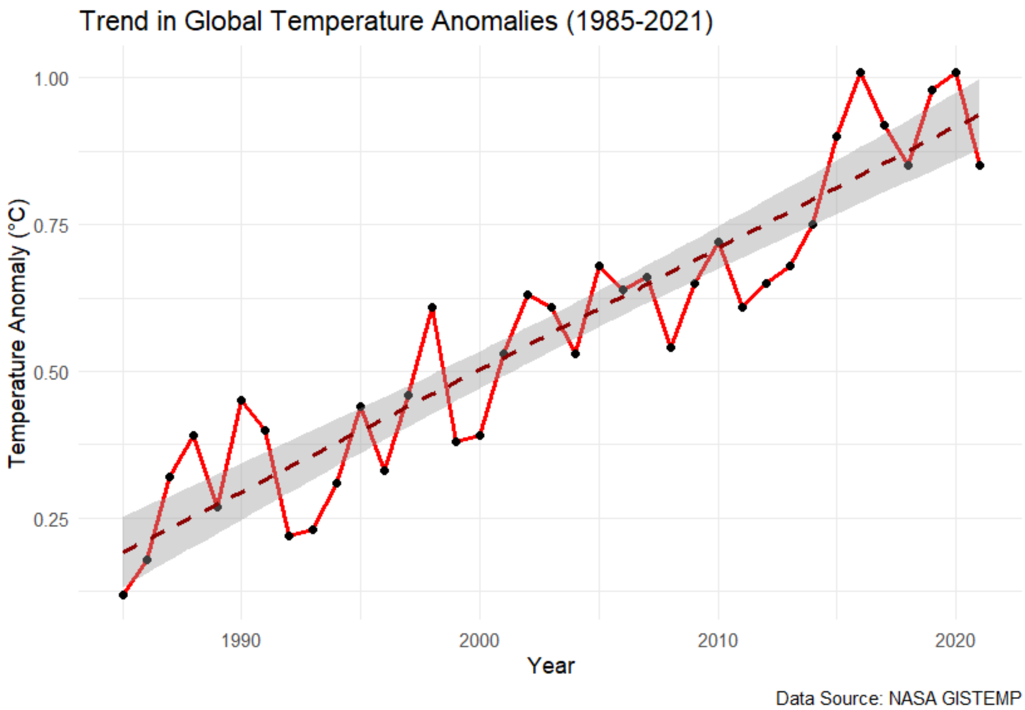

Figure 8: Trend for Global Temperature Anomalies

Figure(8) clearly shows the trend for global temperature anomalies is moving upwards.

5.1.7 Trend in Flood Metrics

This section discuses the trends seen by flood metric including annual number of flood counts, annually displaced and annual flood deaths.

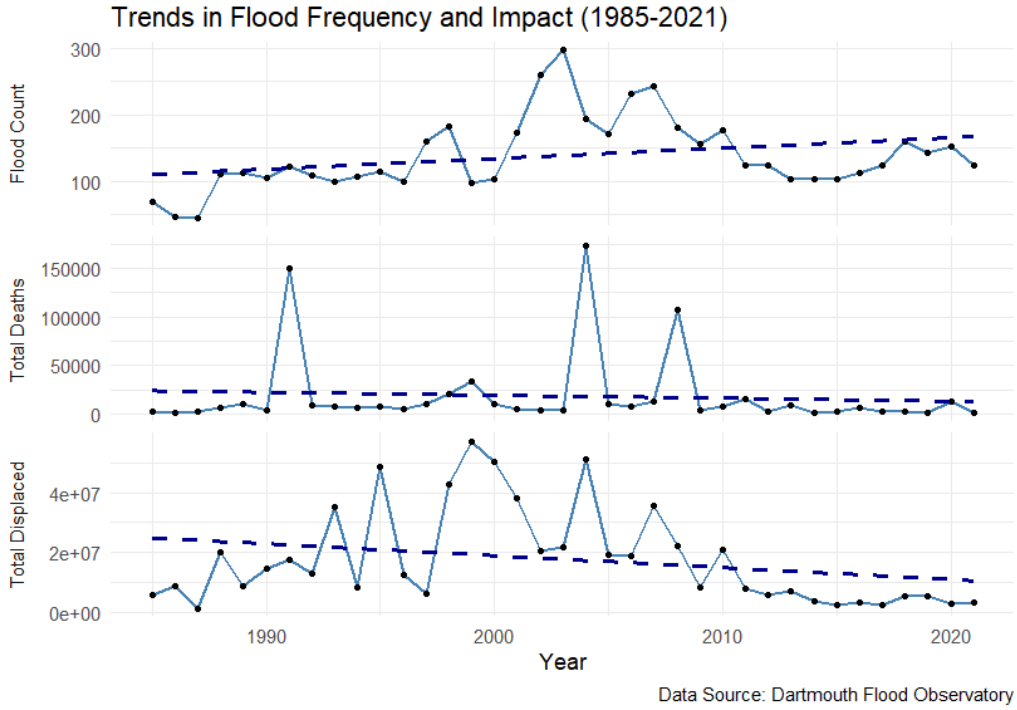

Figure 9: Trend for Flood Metrics

From the figure(9) some insights can be gathered that the flood count is increasing but also counter intuitively the trend in annual death and displaced is going downwards.

5.2 Correlation Analysis

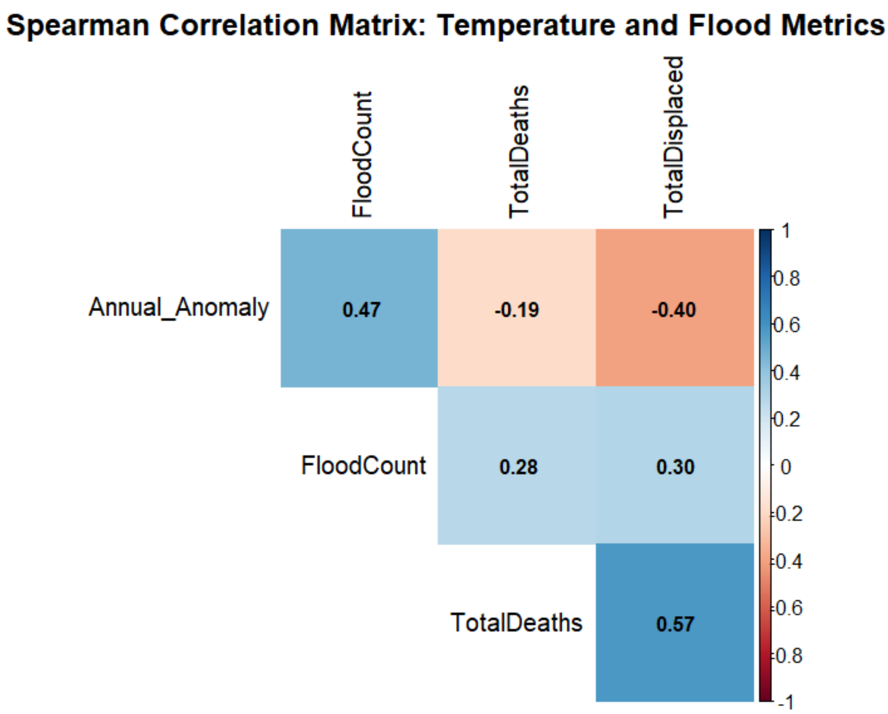

Figure 10: Heatmap of correlations between temperature anomaly, flood count, deaths, and displaced using Spearman method (1985–2021).

5.2.1 Correlation between Temperature and Flood Count

correlation analysis was done using Peason method

Table 5: Correlation between Annual Temperature Anomaly and Flood Count (1985–2021)

Statistic	**Value**
Correlation Coefficient (Pearson's r)	0.3574
t-value	2.2642
Degrees of Freedom (df)	35
p-value	0.0299
95% Confidence Interval	[0.0378, 0.6107]
Alternative Hypothesis	True correlation is not equal to 0

The table(5) shows that Pearson correlation coefficient between annual temperature anomaly and flood count is 0.357, indicating a moderate positive relationship. The correlation is statistically significant ($p = 0.0299$), suggesting that years with higher global temperature anomalies tend to have more flood events. The 95% confidence interval [0.038, 0.611] confirms that the correlation is unlikely to be zero.

5.2.2 Correlation between Temperature and Total Deaths

To find out the correlation spearman method was used as the columns were not normally distributed.

Table 6: Spearman Correlation between Annual Temperature Anomaly and Total Flood Deaths (1985–2021)

Statistic	**Value**
Correlation Coefficient (Spearman's ρ)	-0.1880
S-value	10,022
p-value	0.2652
Alternative Hypothesis	True correlation is not equal to 0

The table(6) shows that Spearman correlation coefficient between annual temperature anomaly and total flood deaths is -0.188, indicating a weak negative relationship. The correlation is not statistically significant ($p = 0.265$), suggesting that there is no strong evidence that higher global temperatures are associated with changes in total flood deaths during 1985–2021."

5.2.3 Correlation between Temperature and Total Displaced

Table 7: Spearman Correlation between Annual Temperature Anomaly and Total Flood Displaced (1985–2021)

Statistic	**Value**
Correlation Coefficient (Spearman's ρ)	-0.404
S-value	11,841
p-value	0.0132
Alternative Hypothesis	True correlation is not equal to 0

The Table(7) shows Spearman correlation coefficient between annual temperature anomaly and total flood-displaced people is -0.404, indicating a moderate negative relationship. The correlation is statistically significant ($p = 0.013$), suggesting that in years with higher global temperature anomalies, the total number of people displaced by floods tends to decrease. This may reflect the influence of other confounding factors or the uneven distribution of flood events globally.

5.3 Linear Regression

5.3.1 Linear Regression on Annual Temperature Anomaly and Year

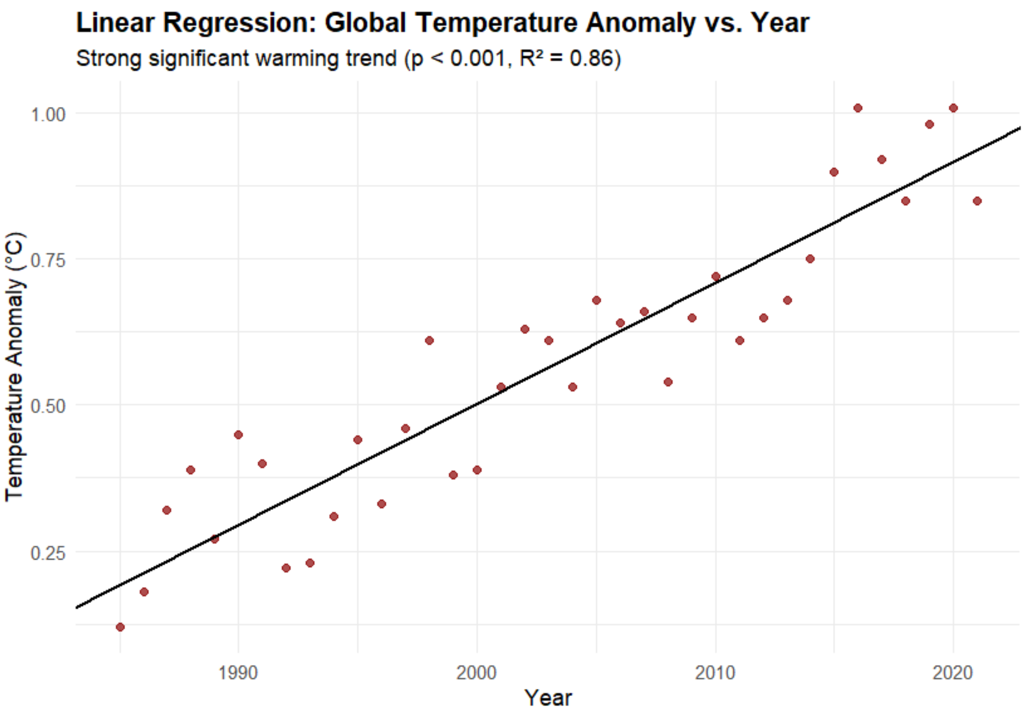

Figure 11: Linear regression of annual global temperature anomaly versus year (1985–2021), showing a significant warming trend over time.

Table 8: Linear Regression Results: Annual Temperature Anomaly vs. Year (1985–2021)

Coefficient	**Estimate**	**Std. Error**	**t value**	**p-value**
Intercept	-40.953	2.814	-14.55	<2e-16 ***
Year	0.02073	0.00141	14.76	<2e-16 ***

Statistic	**Value**
Residual Std. Error	0.09123 on 35 DF
Multiple R^2	0.8615
Adjusted R^2	0.8576
F-statistic	217.7 on 1 and 35 DF
p-value (F-test)	$<$ 2.2e-16

As shown in the table(8) that the linear regression shows a statistically significant positive trend in annual global temperature anomalies from 1985 to 2021. The slope coefficient (0.0207 °C/year, p ¡ 0.001) indicates that the global mean temperature anomaly has increased on average by approximately 0.021 °C per year. The model explains 86.2% of the variance in annual anomalies (Adjusted $R^2 = 0.8576$), indicating a strong fit. Residuals are small and evenly distributed, suggesting the model assumptions are reasonably met.

5.3.2 Linear Regression on Annual Flood Count vs Year

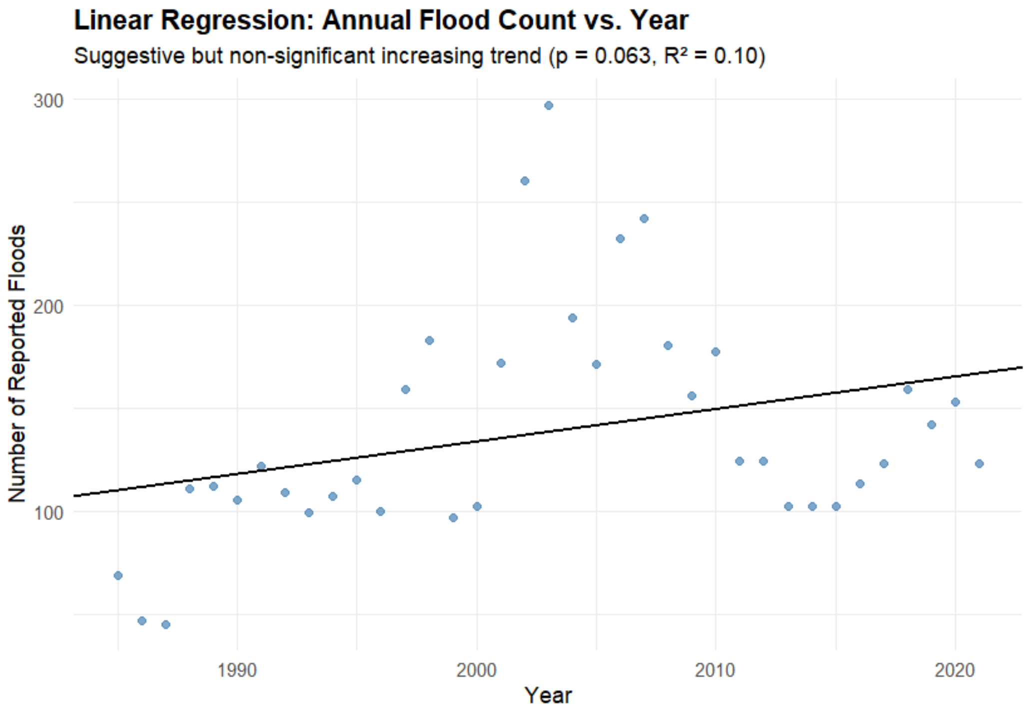

Figure 12: Linear regression of annual flood count versus year (1985–2021), showing a weak positive trend that is not statistically significant.

Table 9: Linear Regression Results: Annual Flood Count vs. Year (1985–2021)

Coefficient	**Estimate**	**Std. Error**	**t value**	**p-value**
Intercept	-3019.23	1644.61	-1.836	0.0749 .
Year	1.5766	0.8211	1.920	0.0630 .

Statistic	**Value**
Residual Std. Error	53.32 on 35 DF
Multiple R^2	0.0953
Adjusted R^2	0.0695
F-statistic	3.687 on 1 and 35 DF
p-value (F-test)	0.0630

The table(9)linear regression of annual flood count versus year suggests a slight positive trend also supported by figure(12), with an estimated increase of 1.576 floods per year. However, the relationship is not statistically significant at the 5% level (p = 0.063), and the model explains only 9.5% of the variance in flood counts (Adjusted R^2 = 0.0695). This indicates that while there may be a weak upward trend, other factors likely dominate the year-to-year variability in flood events.

5.3.3 Linear Regression Annual Flood Count vs Temperature Anomaly

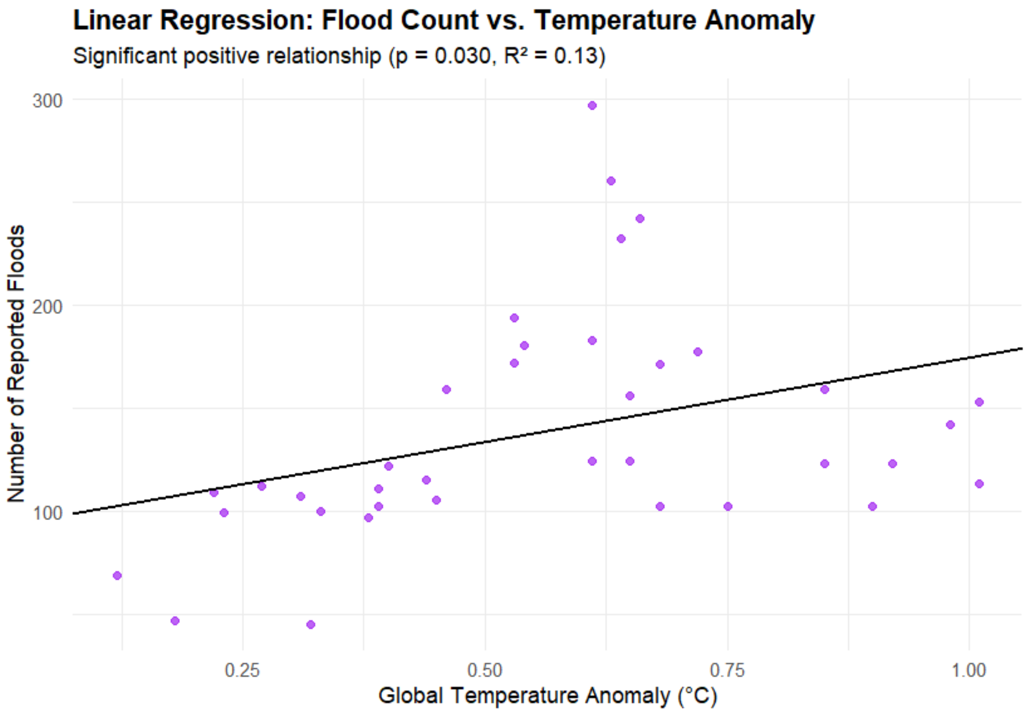

Figure 13: Linear regression of annual flood count versus global temperature anomaly (1985–2021), showing a significant positive relationship

Table 10: Linear Regression Results: Annual Flood Count vs. Temperature Anomaly (1985–2021)

Coefficient	Estimate	Std. Error	t-value	p-value
Intercept	92.48	22.13	4.178	0.000186 ***
Annual Anomaly	81.74	36.10	2.264	0.0299 *

Statistic	Value
Residual Std. Error	52.36 on 35 DF
Multiple R^2	0.1278
Adjusted R^2	0.1028
F-statistic	5.127 on 1 and 35 DF
p-value (F-test)	0.02986

As shown, the linear regression shows a statistically significant positive relationship between annual temperature anomaly and flood count. The slope coefficient (81.74, $p = 0.0299$) indicates that for each 1 °C increase in global temperature anomaly, the annual number of floods increases by approximately 82 events. The model explains about 12.8% of the variance in flood counts (Adjusted $R^2 = 0.1028$), suggesting that while temperature is a contributing factor, other variables also influence flood frequency.

5.4 Time Series Analysis

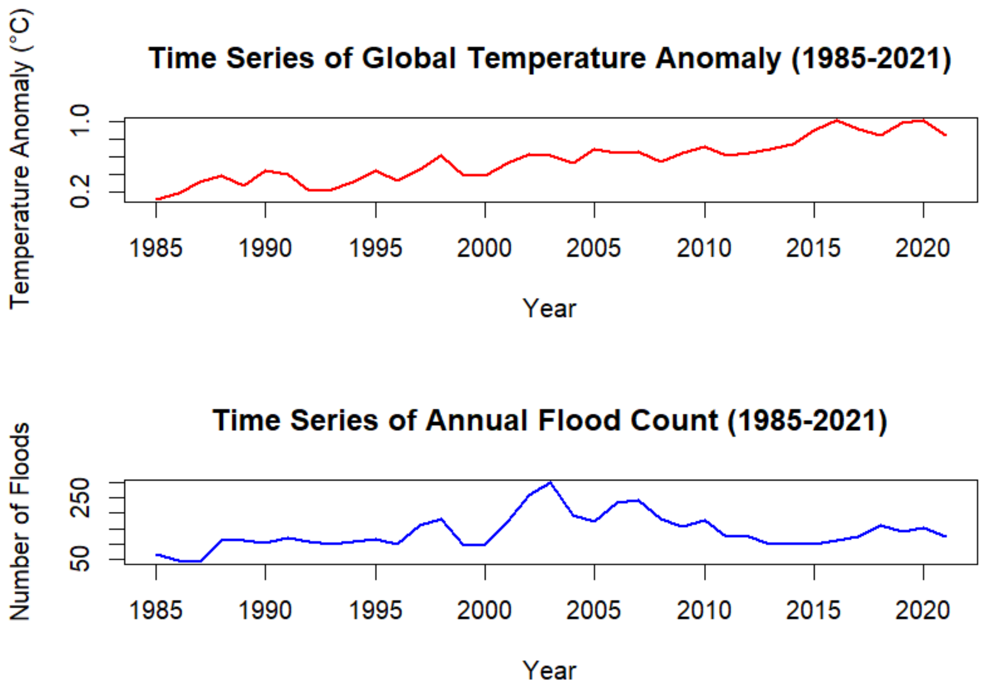

Figure 14: Time series plots of global temperature anomaly (top) and annual flood count (bottom) from 1985 to 2021.

5.4.1 Stationary Test

Table 11: Results of Augmented Dickey-Fuller (ADF) Test for Stationarity

Variable	Dickey-Fuller Statistic	Lag Order	p-value	Stationarity Conclusion
Temperature Anomaly	-2.7615	3	0.2763	Not stationary
Flood Count	-1.4449	3	0.7897	Not stationary

The Augmented Dickey-Fuller (ADF) test was applied to examine whether the time series data for global temperature anomaly and annual flood count are stationary. The results show that both series have high p-values (0.2763 for temperature anomaly and 0.7897 for flood count), which are above the 0.05 significance level. This indicates that we fail to reject the null hypothesis of a unit root, meaning both series are non-stationary in their original form. Differencing is required to make them stationary for further time series modeling.

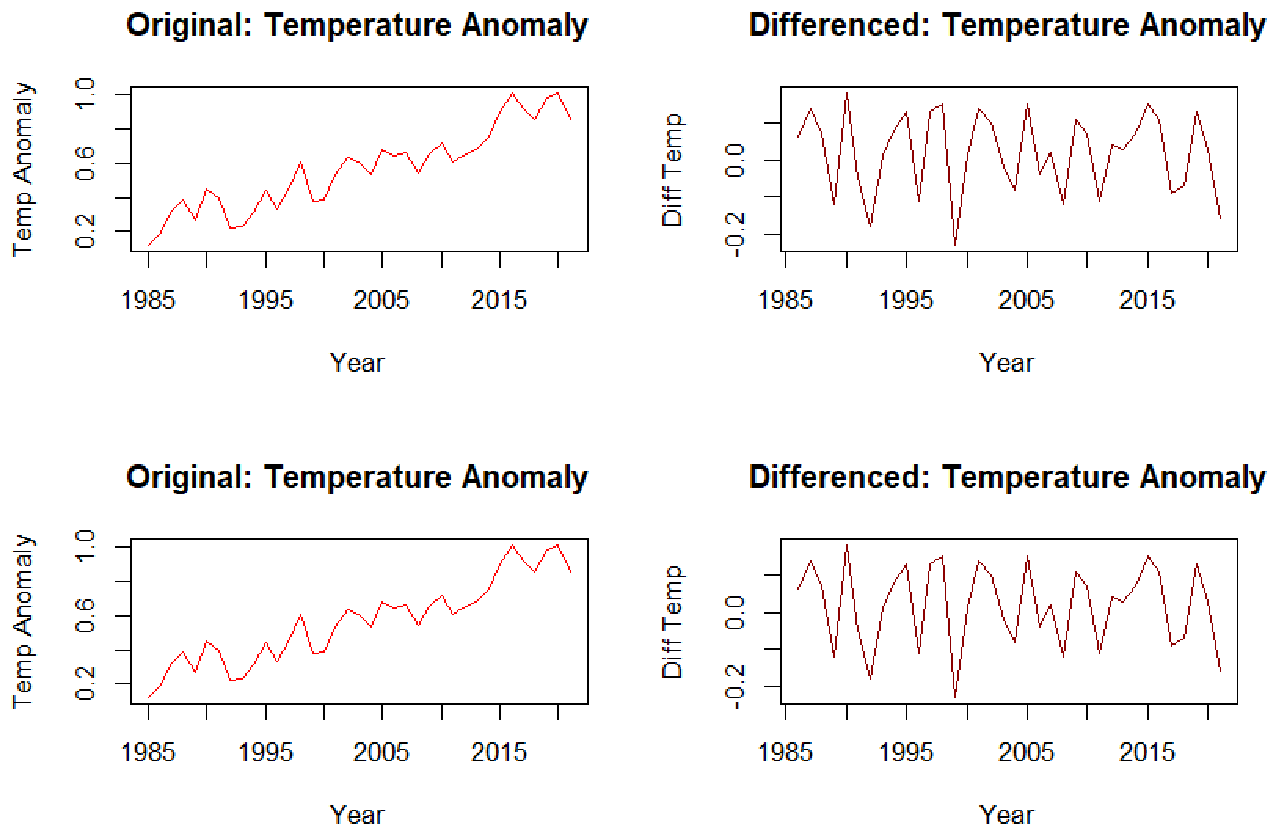

Figure 15: Time series of the original global temperature anomaly and its differenced version, showing how differencing removes trends and helps achieve stationarity

5.4.2 Cross Correlation

Table 12: Cross-Correlation between Temperature Anomaly and Flood Count (Lags ±5 Years)

Lag (Years)	Correlation Coefficient
-5	0.031
-4	0.060
-3	0.062
-2	0.141
-1	0.285
0	0.357
1	0.229
2	0.174
3	0.195
4	0.194
5	0.178

The table(12) depicts the cross-correlation analysis and shows that the strongest positive relationship at lag 0 (correlation = 0.357), suggesting that increases in global temperature anomalies are at the same time associated with higher flood counts. Smaller correlations at other lags indicate weaker delayed effects.

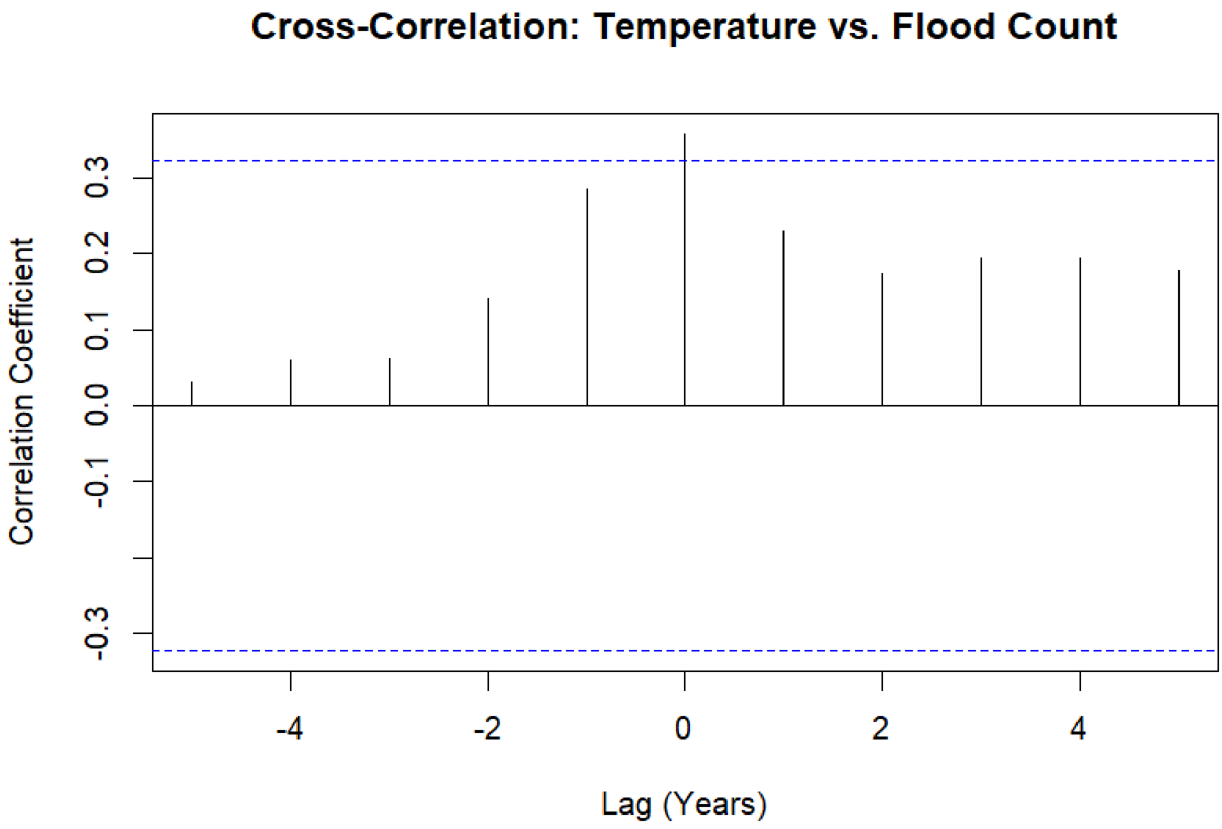

Figure 16: Cross-correlation function (CCF) between global temperature anomaly and annual flood count, showing the strength and direction of correlations across time lags (±5 years)

6 Conclusions

This report used several statistical methods to study the link between global temperature changes and floods from 1985 to 2021. The results showed clear patterns that highlight both the effects of climate change and the challenges of measuring its impact.

The main findings are that the global temperatures have risen at a strong and significant rate, about 2.07°C per century. There is also a positive and statistically significant link between higher temperatures and the number of floods, meaning warmer years tend to have more flood events. Time series analysis showed that this link happens in the same year, with no strong evidence that hotter years directly cause more floods in the years after.

Moreover, the study also found a negative relationship between temperature and the number of people displaced by floods. This may be because better forecasting, disaster response, and resilience measures are reducing the human impact of floods, even as they happen more often.

Finally, the evidence suggests that climate change is increasing the number of floods worldwide. However, the overall effects are shaped by other important factors, such as improvements in disaster reporting and risk management. While global warming raises the risk of flooding, society has also become better at adapting and reducing vulnerability. This shows why continued investment in climate resilience and adaptation is crucial. Future studies could provide deeper insight by looking at specific regions and including more detailed models that account for social and economic differences.

7 Appendix

Link to code: GitHub Repository

References

Committee, U. C. J. E. (2024), 'Flooding costs the u.s. between $179.8 and $496.0 billion each year'. Estimate in 2023 dollars for total annual U.S. flood costs.

for Space Studies (GISS), N. G. I. (2025), 'Gistemp surface temperature analysis maps', Online tool, NASA GISS. Sources and parameters: GHCNv4_ERSSTv5_1200km_Anom_7_1981_2024_1952_1980_100_180_90_0_2_ROB. Accessed: 2025-08-20.
URL: *https://data.giss.nasa.gov/gistemp/maps/*

GISTEMP Team (2025), 'Giss surface temperature analysis (gistemp), version 4', https://data.giss.nasa.gov/gistemp/. Dataset accessed 2025-08-20.

IPCC (2021), 'Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change'.
URL: *https://www.ipcc.ch/report/ar6/wg1/*

Lenssen, N., Schmidt, G. A., Hendrickson, M., Jacobs, P., Menne, M. & Ruedy, R. (2024), 'A gistempv4 observational uncertainty ensemble', *Journal of Geophysical Research: Atmospheres* **129**(17), e2023JD040179.

Observatory, D. F. (n.d.), 'Flood archive', http://floodobservatory.colorado.edu/Archives/index.html. Accessed: 2023-10-27.

Trenberth, K. E. (2011), 'Changes in precipitation with climate change', *Climate Research* **47**(1), 123–138.