A Multi-source United States Flood Database

Group Members: Ashmita Dhakal, Baichuan Ni, Clara Tu, Songxi Yang

- **Abstract**

This research paper presents a comprehensive analysis of the multifaceted impacts of flooding, delving into its economic repercussions, demographic effects, geographical patterns, and seasonal variations across different states. Central to this study is the development of a spatial database, a pivotal tool in dissecting the nuances of flood impacts. The paper addresses several critical questions: the economic impact of floods, the influence of floods on populations, the correlation between geographic features and flooding, and the seasonality of floods in various states.

Through this spatial database, we gain insights into how floods affect economies, highlighting the crucial role of flood prediction and intervention in mitigating financial losses. The study also explores the immediate and long-term effects on populations, including displacement, health issues, and demographic shifts. An examination of geographic features most prone to flooding is conducted, aiming to enhance future predictions and preparedness measures. Additionally, the paper identifies seasonal patterns of flooding in different states, a key factor in efficient resource allocation and emergency planning.

This research is pivotal in understanding the broad spectrum of flood impacts and serves as a guide for policy-makers, urban planners, and disaster management authorities in formulating effective strategies to manage and mitigate the consequences of flooding.

- **Problem statement**

Flooding represents one of the most common and devastating natural disasters, impacting economies, populations, and environments globally. Despite its prevalence, there remains a significant gap in understanding the multifaceted effects of flooding, particularly in relation to economic damage, demographic changes, geographical vulnerability, and seasonal trends. The lack of a comprehensive spatial database to analyze these aspects hinders effective policy-making and disaster-management strategies. This gap in knowledge and resources poses a critical challenge in predicting, preparing for, and mitigating the wide-ranging impacts of flood events. Consequently, communities, economies, and ecosystems continue to suffer from the inadequately managed consequences of flooding, highlighting the urgent need for an integrated and detailed analysis to inform more resilient and effective responses.

- **Datasets**

USFD1 (flood events): damage and basic information about the flood events

USFD2 (geographic features): geographic features related to the flood events

Weather_NOAA_1981_2010: U.S. Climate Normals from 1981 to 2010

Station: Stations that record the weather information in the weather dataset

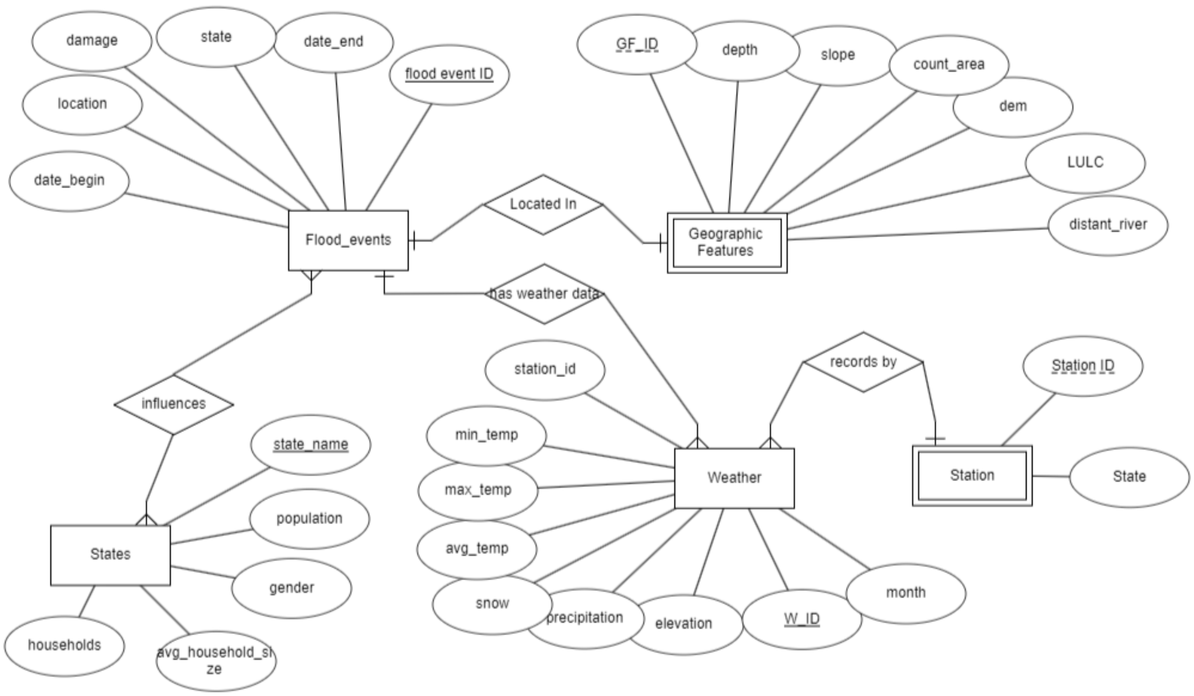
- **Database design and manipulations**

Entities:

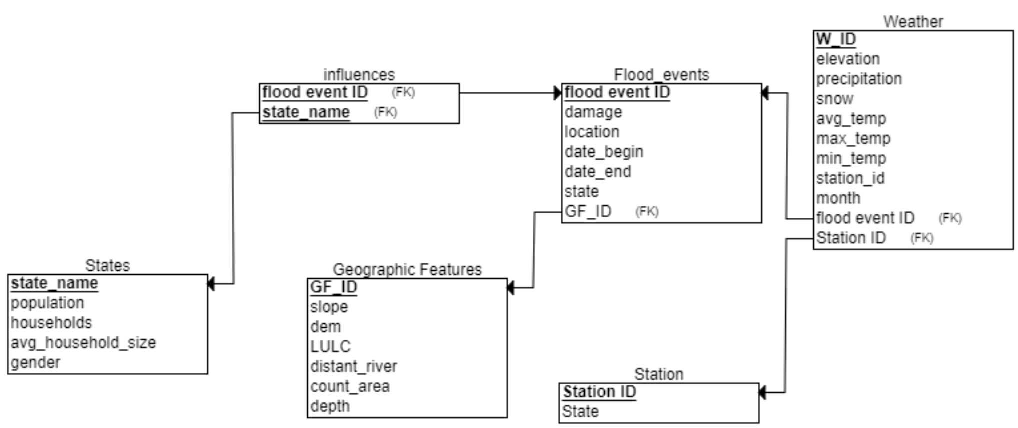
1. Flood events
   - Attributes: flood event ID, damage, location, date_begin, date_end, state
2. Geographic Features
   - Attributes: GF_ID, depth, slope, count_Area, dem, LULC, distant_river
3. States
   - Attributes: state_name, population, avg_household_size, households, gender
4. Weather
   - Attributes: W_ID, elevation, precipitation, snow, min_temp, max_temp, avg_temp, station_id, month
5. Station
   - Attributes: Station ID, State

Relationships, both spatial and nonspatial, between objects:
Geographic features are inherently connected to flood events, often determining or being influenced by their occurrence. Flood events, in turn, are closely associated with specific weather data, such as rainfall intensity, duration, and patterns, which are crucial for understanding and predicting flood dynamics. This weather data is recorded by stations strategically placed to capture meteorological conditions. Furthermore, flood events have a significant impact on demography, influencing population distributions, movements, and demographic changes due to displacement, economic shifts, and health impacts. These relationships showcase a complex interplay between natural phenomena and human dynamics, where spatial aspects and nonspatial elements are intrinsically linked through the occurrence and consequences of flood events.

ER Diagram



Logical schema:



- **Case studies and results**

1. Economic impact of floods

Top ten states with the greatest damage:

SELECT state, SUM(damage) AS total_damage

FROM flood_events

GROUP BY state

ORDER BY total_damage DESC

LIMIT 10;

Query   Query History

```sql
SELECT state, SUM(damage) AS total_damage
FROM flood_events
GROUP BY state
ORDER BY total_damage DESC
LIMIT 10;
```

Data Output   Messages   Notifications

	state character varying (100)	total_damage numeric
1	TEXAS	49692232550
2	NEW JERSEY	22448169440
3	PUERTO RICO	19393590000
4	LOUISIANA	9914312100
5	TENNESSEE	4856714400
6	NORTH DAKOTA	3919837300
7	CALIFORNIA	3582752490
8	IOWA	3493201490.0007
9	NEW YORK	3437619890
10	MICHIGAN	2967849130

From the statistics of the top ten states with the greatest damage, the research observes that the data on flood damages in various U.S. states and territories reveal a strong geographical correlation. Texas, with the highest damage of approximately $49.69 billion, illustrates the impact of its vast geographical area and exposure to diverse weather phenomena. Coastal states like New Jersey and Louisiana, with damages of $22.45 billion and $9.91 billion, respectively, face significant flood risks due to their proximity to the ocean, making them prone to storm surges and hurricane impacts. This is similarly seen in Puerto Rico, a Caribbean territory subjected to tropical storms, resulting in around $19.39 billion in damages.

Inland states, while not directly impacted by oceanic storms, also report substantial flood damages due to other factors. Tennessee and North Dakota, with damages of $4.86 billion and $3.92 billion, respectively, demonstrate the risks associated with river flooding. Meanwhile, states like California, Iowa, New York, and Michigan, despite their diverse climatic and geographic profiles, show significant flood-related costs ranging from $2.97 to $3.58 billion. These figures underscore the varied nature of flood risks across the United States, influenced by a combination of topographical features, river systems, and weather patterns.

Create a year column:

```
ALTER TABLE flood_events

ADD COLUMN year INT;


UPDATE flood_events

SET year = CASE

        WHEN date_begin ~ '^\d{4}' THEN CAST(SUBSTRING(date_begin, 1, 4) AS INT)

        ELSE NULL

    END;
```

Ten years that have the greatest damage:

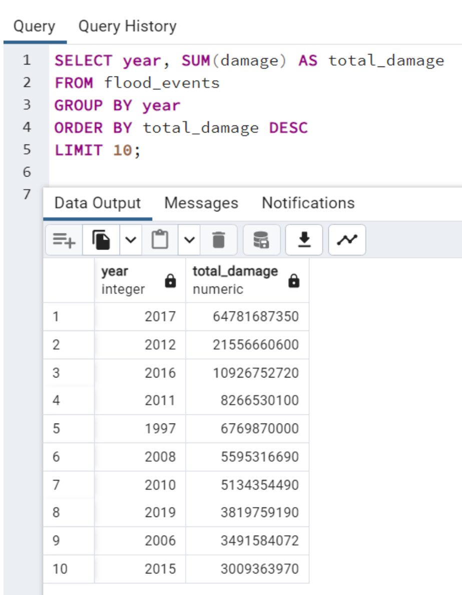
```
SELECT year, SUM(damage) AS total_damage

FROM flood_events

GROUP BY year

ORDER BY total_damage DESC

LIMIT 10;
```




```
SELECT year, SUM(damage) AS total_damage
FROM flood_events
GROUP BY year
ORDER BY total_damage DESC
LIMIT 10;
```

	year integer	total_damage numeric
1	2017	64781687350
2	2012	21556660600
3	2016	10926752720
4	2011	8266530100
5	1997	6769870000
6	2008	5595316690
7	2010	5134354490
8	2019	3819759190
9	2006	3491584072
10	2015	3009363970

The data on the greatest flood damages over various years reveals a noteworthy trend, particularly highlighting the recent increase in flood-related damages. The year 2017 stands out significantly with an exceptionally high damage cost of approximately $64.78 billion, which is notably higher than any other year listed. This spike could be attributed to specific high-impact flood events in that year, possibly linked to severe weather phenomena or climatic changes.

Another observation is the concentration of high-damage years in the 2010s, with 2012, 2016, 2011, 2019, and 2015 all featuring in the top ten. This trend suggests an increasing frequency and/or intensity of damaging flood events in the recent decade. The presence of earlier years like 1997 and 2006, while significant, indicates that high-cost flood events are not exclusively a recent phenomenon, but their escalation in the latter half of the 2010s is notable. This pattern might point towards broader environmental changes or increased vulnerability due to factors such as urbanization, climate change, or changes in land use, necessitating a reevaluation of flood management and mitigation strategies.

2. Influence of floods on populations

Top ten states with the highest flood count over population ratio in 2000:

```
SELECT fp.state,
    fp.flood_count,
    p.pop,
    CASE
        WHEN p.pop > 0 THEN fp.flood_count / CAST(p.pop AS FLOAT)
        ELSE 0
    END AS ratio
FROM (
    SELECT state, COUNT(*) as flood_count
    FROM flood_events
    WHERE year = 2000
```

```
    GROUP BY state
) fp
JOIN states p ON fp.state = p.state_name
ORDER BY ratio DESC
LIMIT 10;
```

	state character varying (100)	flood_count bigint	pop double precision	ratio double precision
1	NORTH DAKOTA	56	642200	8.720024914356898e-05
2	VERMONT	42	608827	6.898511399790089e-05
3	IOWA	173	2926324	5.911853916381098e-05
4	KENTUCKY	217	4041769	5.3689362256972135e-05
5	WEST VIRGINIA	88	1808344	4.8663307423808745e-05
6	OKLAHOMA	137	3450654	3.970261869199288e-05
7	HAWAII	48	1211537	3.9619095413511925e-05
8	MAINE	44	1274923	3.4511888168932554e-05
9	ALASKA	18	626932	2.871124779082899e-05
10	NEBRASKA	49	1711263	2.863382191983348e-05

The data reveals significant insights into the impact of floods on different states' populations in the year 2000. North Dakota, despite a smaller population, experienced a disproportionately high impact from flooding, evidenced by the highest ratio of flood events to population (approximately 8.72e-05). This suggests that a significant percentage of its residents were affected by floods. Similarly, Vermont, with a ratio of around 6.90e-05, also faced considerable flood-related challenges relative to its population size.

On the other hand, states like Iowa and Kentucky, despite having a higher absolute number of flood events (173 and 217 respectively), show a lower per capita impact when compared to North Dakota and Vermont. This is reflected in their lower ratios (approximately 5.91e-05 for Iowa and 5.37e-05 for Kentucky). These figures indicate that while floods were frequent, their impact was spread across a larger population, slightly reducing the relative burden per individual. Overall, the data highlights how the impact of natural disasters like floods varies significantly across states, not just in terms of frequency but also in terms of per capita effect.

3. The correlation between geographic features and flooding

To find the correlation between geographic features and flood events, we need to use tables having information about geography and flood events.

a. Influence of geographic features on damage due to flood events

The correlation between damage by the flood events and different geographic features such as distance to river, depth, slope and contributing by using statistical function called corr() as follows:

```
SELECT
   corr(fe.damage, gf.depth) AS damage_depth_corr,
   corr(fe.damage, gf.slope) AS damage_slope_corr,
   corr(fe.damage, gf.count_area) AS damage_contrib_area_corr,
   corr(fe.damage, gf.distant_river) AS damage_distant_river_corr
FROM flood_events fe, geographic_features gf
Where fe.flood_event_id = gf.gf_ID
```

The result of this query is:

damage_depth_corr	damage_slope_corr	damage_contrib_area_corr	damage_distant_river_corr
0.002536068	-0.003282914	-0.000431828	-0.000589668

Based on this data, it appears that there is an extremely weak positive correlation between damage due to flood events and 500-yr flood depth. However, there seems to be no linear relationship between these two attributes. Similarly, there seems to be a weak negative correlation between damage and slope, contribution area, and distance to the nearest river, inferring there is an almost negligible relationship between these attributes and the resulting damage from flood events.

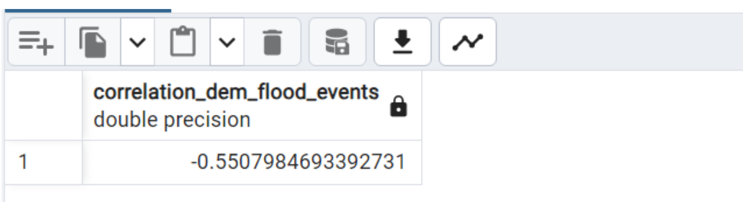
b. Correlation between geographic features and total flood events.

I. Total flood count and DEM

```
SELECT
```

```
  corr(total_flood_events, dem) AS correlation_dem_flood_events
FROM (
   SELECT
      gf.dem,
      COUNT(fe.flood_event_ID) AS total_flood_events
   FROM geographic_features gf
   LEFT JOIN flood_events fe ON gf.gf_ID = fe.gf_ID
   GROUP BY gf.dem
) AS subquery
```

Result:



	correlation_dem_flood_events double precision
1	-0.5507984693392731

The value for correlation between DEM and total number of flood events suggest that as the DEM values change (increase or decrease), there is a moderate tendency for the total count of flood events to change in the opposite direction. This correlation suggests that as elevation changes, the number of flood events might change too, but there are other factors for floods such as distance to the river, contributing area to the river, and so on.

II. Total flood count and distance to major river

```
 SELECT
   corr(total_flood_events, distant_river) AS correlation_distant_river_flood_events
FROM (
   SELECT
      gf.distant_river,
      COUNT(fe.flood_event_ID) AS total_flood_events
```

```
  FROM geographic_features gf
  LEFT JOIN flood_events fe ON gf.gf_ID = fe.gf_ID
  GROUP BY gf.distant_river
) AS subquery
```

Result:

	correlation_distant_river_flood_events double precision
1	-0.0034383925388485814

The correlation coefficient value between total count of flood events and distance to nearest river is extremely weak, negative and almost negligible. There is almost no observable linear relationship between the total count of flood events and the distance to the nearest major river. This means that changes in the total number of flood events do not significantly correlate with changes in the distance to the nearest major river.

III. Total flood count and contributing area

```
 SELECT
   corr(total_flood_events, count_area) AS correlation_distant_river_flood_events
FROM (
   SELECT
     gf.count_area,
      COUNT(fe.flood_event_ID) AS total_flood_events
   FROM geographic_features gf
   LEFT JOIN flood_events fe ON gf.gf_ID = fe.flood_event_ID
  GROUP BY gf.count_area
) AS subquery
```

Result:

	correlation_distant_river_flood_events double precision
1	-0.004449671260056443

This correlation coefficient value between the total count of flood events and the contributing area indicates a negative, extremely weak, almost negligible correlation between these two variables. There is almost no observable linear relationship between them.

These correlation values don't mean DEM value, distance to the nearest river, and contributing area cause the flood events to happen. It just shows there is a connection between flood events and elevation changes, distance to the river, and contributing area. This suggests that as these aspects change, the occurrence of flood events might also change, but it is most likely that factors beyond these measures also contribute to the occurrence of floods.

4. Flood Seasonality

```
SELECT
  STATE,
  CASE
    WHEN date_begin ~ '^\d{12}$' AND SUBSTRING(date_begin, 1, 4)::INT BETWEEN 1990
AND 1999 THEN '1990-1999'
    WHEN date_begin ~ '^\d{12}$' AND SUBSTRING(date_begin, 1, 4)::INT BETWEEN 2000
AND 2009 THEN '2000-2009'
    WHEN date_begin ~ '^\d{12}$' AND SUBSTRING(date_begin, 1, 4)::INT BETWEEN 2010
AND 2019 THEN '2010-2019'
  END AS TimePeriod,
  EXTRACT(MONTH FROM TO_TIMESTAMP(SUBSTRING(date_begin, 1, 8),
'YYYYMMDD')) AS FloodMonth,
  COUNT(*) AS FloodCount
FROM
  flood_events, states
WHERE
  STATE IS NOT NULL
```

```
  AND state = states.state_name
  AND date_begin ~ '^\d{12}$'
  AND SUBSTRING(date_begin, 1, 4)::INT BETWEEN 1990 AND 2019
GROUP BY
  STATE,
  TimePeriod,
  FloodMonth
ORDER BY
  STATE,
  TimePeriod,
  FloodMonth;
```

The analysis of flood seasonality in the United States across three decades (1990s, 2000s, and 2010s) in Section 4 presents a dynamic and evolving pattern of flood occurrence. In the 1990s, the flood seasonality varied across different regions. The Midwest, particularly the central states, predominantly experienced spring floods. This pattern can be attributed to melting snow and spring rains, which are common in these areas. In contrast, coastal states in the Northeast exhibited a mix of winter and autumn flooding, possibly due to nor'easters and late-season hurricanes impacting these regions.

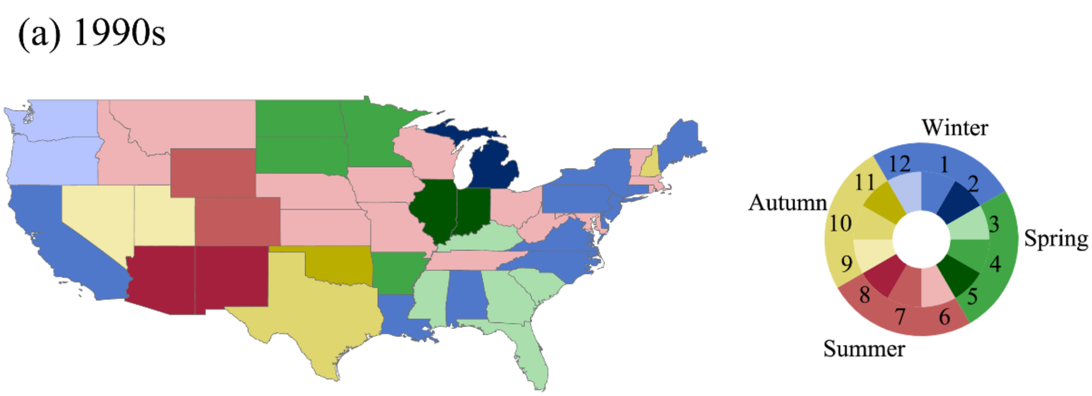


Figure 1. US flood seasonality in the 1990s.

In the 2000s, there was a noticeable shift towards autumn and summer flooding in many states, particularly in the South and Southeast.

(b) 2000s

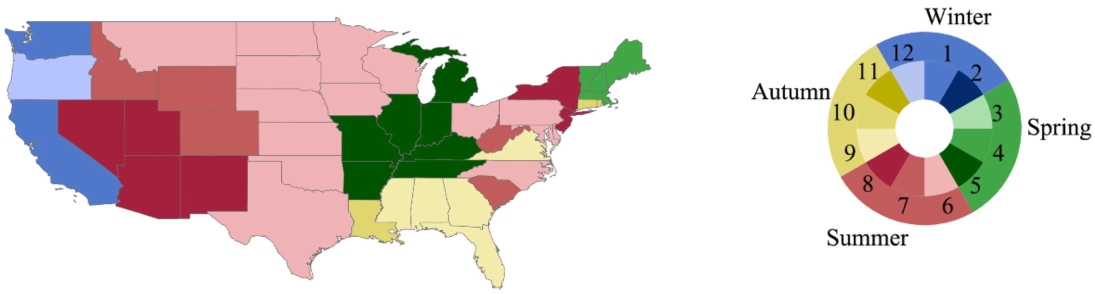


Figure 2. US flood seasonality in the 2000s.

The 2010s continued the trend of summer flooding in the South and Southeast but also showed an emergence of winter flooding in the Northwest.

(c) 2010s

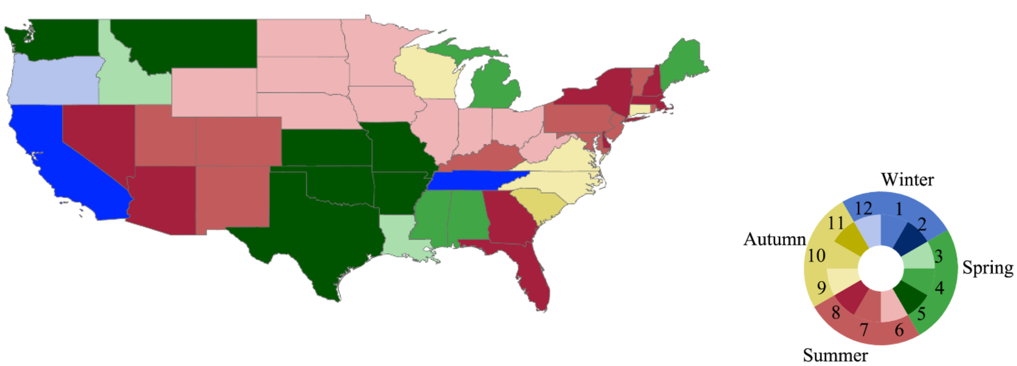


Figure 3. US flood seasonality in the 2010s.

The observed seasonality patterns suggest a complex interplay of climatic factors influencing flood risks across different regions. The analysis also implies a need for region-specific flood management strategies. For example, the central U.S. may benefit from enhanced flood mitigation measures in place for the spring season. In the Northwest, increased monitoring during winter months could be crucial for early flood warning systems.

- **Conclusion**

This research project offers a comprehensive analysis of the multifaceted impacts of flooding across various states in the United States, revealing the complex and varied effects on economic, demographic, and geographical fronts. Key findings include:

1. **Economic Impact**: The study highlights significant economic damages caused by floods, with total losses amounting to billions of dollars. The greatest damages are observed in states with diverse geographical features and weather patterns, including coastal areas and inland regions susceptible to river floods.
2. **Demographic Effects**: Floods profoundly impact population dynamics, leading to displacement and affecting health and social structures. The analysis of flood event counts ratio to population in different states emphasizes the necessity of effective disaster management and relief measures to address these population shifts.
3. **Geographical Correlation**: The research identifies a weak correlation between flood damage and geographical features such as flood depth, slope, and proximity to rivers. This finding indicates the complexity of flooding, where geographical features play a role but are not the sole determinants of flood impacts.
4. **Seasonality Trends**: The analysis of flood seasonality over three decades reveals evolving patterns, with different regions experiencing distinct seasonal trends. This variability underscores the need for region-specific flood preparedness and mitigation strategies, adapting to changing climatic conditions and flood patterns.

In conclusion, this research underscores the importance of an integrated approach to understanding and managing the impacts of flooding. The development of the spatial database has been instrumental in providing valuable insights into the economic, demographic, and geographical aspects of flooding. These findings serve as a guide for policymakers, urban planners, and disaster management authorities, emphasizing the need for tailored strategies that take into account the unique characteristics of each region. As climate change continues to influence weather patterns, this research provides a foundation for developing more resilient and effective responses to the challenges posed by flooding.

References

[1] Li, Zhi, Mengye Chen, Shang Gao, Jonathan J. Gourley, Tiantian Yang, Xinyi Shen, Randall Kolar, and Yang Hong. "A multi-source 120-year US flood database with a unified common format and public access." Earth System Science Data 13, no. 8 (2021): 3755-3766.

[2] Zenodo, "United States Flood Database," 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7545697.

[3] NOAA, "U.S. Climate Normals," 2023. [Online]. Available: https://www.ncei.noaa.gov/products/land-based-station/us-climate-normals.