Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1877–1887
https://doi.org/10.54941/ahfe1008416

The FLOODPERCEP Dataset: Flood Hazard Maps and Imagery for Studying Human Perception of Flood Risk

Amir Behzadan[1,2], Subashree Dinesh[1,3], Armita Dabiri[1,2], and Arunima Maitra[1,3]

[1]Connected Informatics and Built Environment Research (CIBER) Lab, Institute of Behavioral Science, University of Colorado, Boulder, CO 80309, USA
[2]Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, CO 80309, USA
[3]Department of Computer Science, University of Colorado, Boulder, CO 80309, USA

ABSTRACT

Understanding public risk perception is essential as climate-related hazards, including floods, become more intense, frequent, and widespread. This perception varies across population groups, and is influenced by factors, such as sociodemographic characteristics, previous hazard exposure, and the format of risk information. Individuals process visual hazard information, such as flood hazard maps and photographic imagery, differently. We present FLOODPERCEP, a new dataset designed to examine public interpretation and perception of flood risk. The dataset integrates geospatial risk products from governmental and non-governmental entities, as well as web-sourced, geolocated photographic evidence of historical inundation events from 260 U.S. locations, complete with metadata, alongside contextual Google Maps views. The dataset also contains quantitative and qualitative responses from $n = 796$ study participants who viewed this visual data. This extensive set of information enables detailed analyses of the level of comprehension of multi-format flood risk information, as well as public risk perception and behavioral intentions. Ultimately, this dataset aims to support the design, development, and deployment of more effective, inclusive, and evidence-based hazard communication strategies.

Keywords: Flood risk perception, Multimodal dataset, User study, Hazard communication, Human-computer interaction, Disaster informatics

INTRODUCTION

Climate disasters are increasing in magnitude, frequency, and geographic extent (Summers et al., 2022), with floods ranking among the most pervasive and damaging worldwide (Jonkman, 2005; Hirabayashi et al., 2013). According to the International Disaster Database by the Centre for Research on the Epidemiology of Disasters (CRED), floods are the most reported type of disaster globally, accounting for approximately 37%

Received August 26, 2026; Revised September 28, 2026; Accepted October 8, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

of all recorded disasters from 2000–2026 (EM-DAT, 2026). In the U.S., flooding is estimated to cause hundreds of billions of dollars in annual economic losses, amounting to about 1–2% of the nation's total GDP (U.S. Joint Economic Committee, 2025; NOAA, 2025). Globally, between 1990 and 2022, more than 4,700 major flood events affected billions of people, caused over 200,000 deaths, and led to trillions of dollars in losses, with mortality rates declining over time but remaining disproportionately high in low-income countries (Liu et al., 2024).

Effective communication during flood events is key to disseminating timely facts about hazard forecasts, protective actions, and evacuation instructions (Rahman et al., 2024). The increased engagement of the general public in the process of information gathering and sharing, coupled with the growing number of information modalities (text, video, interactive tools) and sources (government alerts, social media, peer-to-peer exchange) can contribute to earlier evacuations, coordinated community response, as well as improved understanding of flood severity, timing, and safe routes (Salley, Fox and Schubert, 2025). However, the mere communication of disaster risk, while necessary, may not be sufficient to instill the desired level of preparedness among residents. Research shows that an individual's own perception of risk may also influence their level of preparedness when faced with a natural hazard (Scovell et al., 2021). The correlation between communication efficacy and risk perception is further driven by various sociodemographic and contextual factors (Perlstein, 2024). Past studies have found that disaster risk perception varies across sociodemographic groups by gender, age, education, and prior disaster experience, as well as time constraints, access to information, and emotions such as fear and anxiety (Cuesta et al., 2022; Lechowska, 2018; Perić and Cvetković, 2019). Flood risk perception affects not only disaster management but also decisions about buying or renting property and acquiring flood insurance (Seenath et al., 2025; Dinesh, Dabiri and Behzadan, 2025). Furthermore, visual depictions of flood scenes using photographs, street-view simulations, or conventional flood maps enhance situational awareness and emotional engagement during real events (Gao et al., 2025). Dynamic, street-level images, for example, produce more accurate judgements of the water depth and stronger affective responses (Bakhtiari et al., 2024).

To date, flood hazard maps produced by government agencies, such as the Federal Emergency Management Agency (FEMA), National Oceanic and Atmospheric Administration (NOAA), and United States Geological Survey (USGS) remain the dominant long-term tool for communicating flood risk, with their design influencing how the public interprets the magnitude of potential hazards (Lim, Brandt and Seipel, 2024). Meanwhile, the rise of crowdsourcing platforms and social media channels has created vast amounts of volunteered geographic information (VGI), allowing affected communities to share images, videos, and narratives in near real-time

(Dujardin et al., 2025; Navalkar et al., 2025). Flood data has grown from a few hundred maps and reports per year to tens of thousands of geotagged posts daily, including roughly 88 million flood-related tweets collected worldwide between 2014 and 2018 (de Bruijn et al., 2019). Researchers have integrated these user-generated content with satellite and sensor data to produce rapid flood maps and to support crisis mapping (Destefanis et al., 2025). Despite the abundance of multimodal data, the literature lacks systematic evidence on whether visual imagery or official hazard maps can effectively shape flood risk perception. A recent South Tyrol, Italy case study found that exposure to hazard maps altered public awareness, risk perception, and willingness to act (Pedoth et al., 2025), motivating a dataset that juxtaposes historical flood images and official maps to assess their differential impact on risk perception, essential for designing communication strategies that translate crowdsourced inputs into actionable public awareness.

In this paper, we present FLOODPERCEP, a dataset capturing public perception of flood risk using flood hazard maps and historical flood imagery sourced from the internet. Flood hazard maps come from three sources: FEMA, NOAA, and Climate Central. Historical images are sourced from searching content on Google Images and social media sites, and consequently stored with detailed metadata. The dataset also includes Google Maps views for each flood location, as well as quantitative and qualitative participant responses on how they interpreted and reacted to the presented flood risk information.

METHODS

Ethics Statement

Research design, protocol, and data collection instruments were reviewed and approved under the 'exempt' category by the Institutional Review Board (IRB) of the University of Colorado Boulder on July 22, 2024 (protocol #24-0438). Participants were at least 18 years old, self-reported U.S. residents, and consented to take part in this study. Study information, timeline, and potential risks and benefits were provided to the participants before asking for consent. To increase participation from historically underrepresented groups in disasters, the consent form and all survey questions were made available in both English and Spanish, allowing participants to take part in the study using their preferred language.

Image Set Curation

Figure 1 shows a sample entry from the FLOODPERCEP dataset, corresponding to Location 114 (Huntington Beach, California).



Figure 1: Sample flood visualization of Location 114 (Huntington Beach, California) across five categories: FEMA, NOAA, Climate Central, Google Maps, and Historical Image.

In total, 260 flood locations in the U.S. are represented in the FLOODPERCEP dataset by five visual sources: historical flood images (from online/social media searches), flood-map screenshots from FEMA's National Flood Hazard Layer (NFHL) Viewer (FEMA, 2024), NOAA's Sea Level Rise Viewer (NOAA, 2024), and Climate Central's Coastal Risk Screening Tool (Climate Central, 2024), plus Google Maps views of each location. All extracted maps also include a legend (overlaid or side-attached) as well as an additional plain U.S. map with a pin marking the exact coordinates.

Participant and Survey Design

The survey was administered over a one-month period between September and October 2024, in both English and Spanish, using Qualtrics and distributed to eligible U.S. participants through Amazon Mechanical Turk (MTurk). Each participant was randomly assigned to one of the five visualization conditions source (FEMA map, NOAA map, Climate Central map, Google Maps, or Historical Image) and viewed two locations, yielding a total of 1,592 visuals reviewed in the study. They also responded to a series of questions evaluating comprehension, perceptions, and behavioral responses to presented information. The distribution of the reviewed visuals across the five visualization conditions was relatively balanced: FEMA (21.36%),

NOAA (17.84%), Climate Central (20.35%), Google Maps (19.60%), and Historical Image (20.85%). Figure 2 illustrates the distribution of 796 participants across U.S. states, as well as the source locations of the 260 flood visualizations in the dataset.

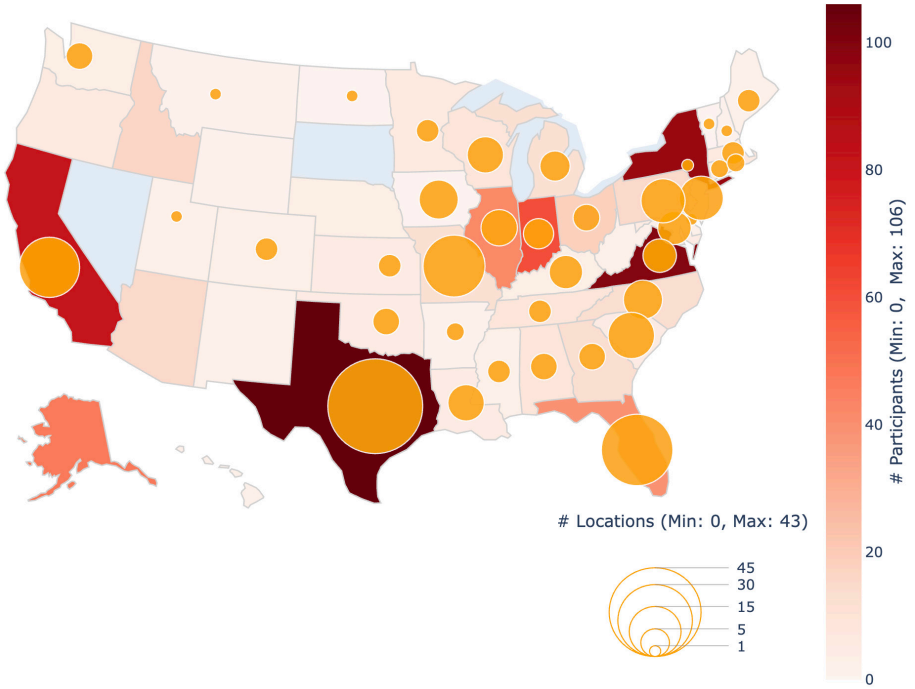


Figure 2: Geographical distribution of study participants and flood visualizations.

The Qualtrics survey followed a structured sequence, beginning with informed consent and eligibility verification, followed by questions collecting participants' MTurk Worker ID, age, and self-reported U.S. Zip code. Next, participants completed background questions capturing sociodemographic factors and information relevant to flood preparedness and perception, including gender, race, ethnicity, education level, primary spoken language, prior flood experience, previous use of flood maps, and actions taken to prepare for or respond to flooding events. Most of these background questions were multiple-choice, with an 'Other' option provided when appropriate. Table 1 describes the sociodemographic characteristics of study participants and compares the distribution of our sample with the general U.S. population (U.S. Census Bureau, 2021, 2022, 2025). In the survey question about race, one response option was 'Middle Eastern or North African'; since the U.S. Census classifies this group as 'White,' we added these participants to the 'White' category to ensure a fair and consistent comparison with Census data (Ennis et al., 2024).

Table 1: Sociodemographic distribution of the study sample in comparison to the U.S. population. Percentages for the U.S. population are based on adults aged 18 years and older.

Variable	Category	Sample (*n*)	Sample (%)	Population (%)
Gender	Male	562	70.6	48.8
	Female	234	29.4	51.2
Race	White alone	601	75.5	75.3
	Black alone	21	2.5	13.7
	American Indian and Alaska Native	26	3.3	1.3
	Asian alone	106	13.3	6.4
	Native Hawaiian or Pacific Islander	2	0.3	0.3
	Two or more races	38	4.8	3.1
	Prefer not to answer	2	0.3	-
Ethnicity	Hispanic or Latino	364	45.7	18.2
	Not Hispanic or Latino	432	54.3	81.8
Education	High school graduate or less	20	2.5	38.7
	Some college	30	3.8	16.5
	Associate degree	19	2.4	9.9
	Bachelor's degree	530	66.6	22.1
	Master's degree	179	22.5	9.5
	Doctoral or other professional degree	16	2.0	3.3
	Prefer not to answer	2	0.2	-
Age	18 to 24	65	8.2	16.2
	25 to 34	462	58.0	16.6
	35 to 44	163	20.5	16.0
	45 to 54	73	9.2	14.7
	55 to 64	25	3.1	15.2
	65 or over	8	1.0	21.3

Each visualization is further accompanied by a set of assessment questions rating perceived flood risk severity, estimated flood frequency in years, understanding of the risk information, and confidence in the accuracy and clarity of the material. Additional questions measured behavioral intentions such as the likelihood of purchasing or renting property in the area, obtaining flood insurance, or implementing preventive measures such as installing barriers or relocating as well as frequency of flood risk map use in decision-making and ease of interpreting the information. Most items used five-point Likert scales, with exceptions for numerical estimates and open-ended responses, and each visualization block concluded with an open-ended

question inviting the respondent to add any comment or concern about the image/map they just viewed. Table 2 summarizes the survey questions.

Table 2: Summary of survey questions on Qualtrics for MTurk participants.

Category	Question Contents
Qualification Questions	• MTurk Worker ID, age, and valid U.S. Zip code
Background Questions	• Gender • Race • Ethnicity • Education • Language • Prior flood preparation actions taken • Prior flood response actions taken • Prior experience with flood maps
Assessment Questions (for each visualization)	• How severe do you think the flood risk, if any, is in the area shown on the map? • In your opinion, how frequently (in years), if at all, will the area shown on the map be flooded in coming years (e.g., once every 3 years)? • How well do you understand the flood risk information, if at all, presented on the map? • How confident are you in the information content and presentation of the map? • Based on the information shown on the map, how likely are you to buy property in the area shown? • Based on the information shown on the map, how likely are you to rent property in the area shown? • Based on the information shown on the map, how likely would you be to purchase flood insurance if you lived in this area? • How likely are you to take preventive measures (e.g., building barriers, relocating) after viewing this map? • How often do you use flood risk maps in making decisions? • How easy was it for you to interpret the information presented on this map? • Do you have any additional comments or concerns about the above map you want to share?

Data Records

The FLOODPERCEP dataset is provided in CSV format, along with historical flood images and maps in JPG format, organized into five folders by source: NOAA, FEMA, Climate Central, Historical Images, and Google Maps. Each historical flood image is accompanied by associated metadata in a separate CSV file. The dataset and accompanying tutorials can be found on CIBER Lab's GitHub page: https://github.com/ciber-lab/floodpercep.

Technical Validation

Multiple validation steps ensured data quality during collection and processing. MTurk and Qualtrics verified valid U.S. Zip codes. Flood image metadata (longitude, latitude, city, state) was programmatically linked to its

corresponding dataset entry to prevent mismatches. Data integrity checks helped exclude records missing key geographic or demographic fields, including participants who abandoned the survey after the demographic section. Since all survey items were mandatory, no further missing values occurred once incomplete surveys were dropped. Initial data collection occurred in two batches (500 and 1,000 responses) totaling approximately 1,500 responses in English and Spanish. After removing duplicate MTurk worker IDs, 830 unique IDs remained. Next, responses completed in under 100 seconds, indicative of insufficient engagement, were excluded, yielding a final set of 796 valid responses. Since each participant evaluated two visualizations, the cleaned dataset expands to 1,592 rows, each representing ratings for a single flood visualization. This structured approach ensures every entry is complete and consistent for subsequent analysis.

USAGE NOTES

The FLOODPERCEP dataset integrates sociodemographic, spatial, and behavioral information to study how individuals perceive and respond to flood risks. The study collects participant characteristics, prior flood experiences alongside GPS-tagged flood visualizations and Likert-scale ratings of perceived risk, confidence, and preventive intentions. These variables enable analyses of how factors such as age, education, or map literacy influence risk perception and preparedness, the extent to which visual cues and distance affect judgment, and how communication materials shape trust in flood information, while supporting spatial mapping and policy design. To facilitate adoption in future research, the dataset includes detailed descriptions, tutorials, and technical notes on accessing, preprocessing, and integrating the raw files into workflows.

As with many large-scale datasets of human data, the FLOODPERCEP dataset also has limitations. The five-point Likert scores are self-reported, which may introduce bias, and the sample may not accurately reflect the broader U.S. population, limiting generalizability. Compared to national figures, the sample skews toward male, Hispanic or Latino, and highly educated. The racial representation is also slightly skewed, with fewer Black participants and more Asian participants. Furthermore, age is concentrated among 25- to 34-year-olds.

CONCLUSION

FLOODPERCEP is a multimodal dataset designed to support research on how individuals interpret and respond to flood risk information presented through different visualization modalities. By integrating flood hazard maps from multiple government sources, in addition to historical flood imagery and Google Maps context, as well as participant responses from 796 individuals, the dataset provides a rich resource for investigating relationships between visual representations, risk perception, comprehension, and behavioral intentions. The accompanying metadata, documentation, and technical validation support reproducibility and facilitate reuse across disciplines. While

the participant sample may not be fully representative of the U.S. population and the responses are based on self-reported measures, the presented dataset provides a foundation for future research in hazard communication, human-computer interaction, disaster informatics, and risk perception. In the long term, we anticipate that this dataset will support the development of more effective, evidence-based flood risk communication.

ACKNOWLEDGMENT

This work was supported by a seed grant from the Institute of Behavioral Science (IBS) at the University of Colorado Boulder. Initial collection and geotagging of historical flood images were carried out by Dr. Bahareh Alizadeh (former CIBER Lab researcher) as part of a project funded by the Texas Sea Grant program of the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce [award #NA18OAR4170088].

REFERENCES

Bakhtiari, V., Piadeh, F., Chen, A.S. and Behzadian, K. (2024) 'Stakeholder analysis in the application of cutting-edge digital visualisation technologies for urban flood risk management: A critical review', *Expert Systems with Applications*, 236(1), p.121426. http://doi.org/10.1016/j.eswa.2023.121426.

Climate Central (2024) *Coastal Risk Screening Tool* [online]. Available at: (Accessed: 12 June 2024).

Cuesta, A., Alvear, D., Carnevale, A. and Amon, F. (2022) 'Gender and public perception of disasters: A multiple hazards exploratory study of EU citizens', *Safety*, 8(3), p. 59. https://doi.org/10.3390/safety8030059.

de Bruijn, J.A., de Moel, H., Jongman, B., de Ruiter, M.C., Wagemaker, J. and Aerts, J.C.J.H. (2019) 'A global database of historic and real-time flood events based on social media', *Scientific Data*, 6(1), p.311. https://doi.org/10.1038/s41597-019-0326-9.

Destefanis, T., Guliyeva, S., Boccardo, P. and Fissore, V. (2025) 'Advancing flood detection and mapping: A review of earth observation services, 3D data integration, and AI-Based techniques', *Remote Sensing*, 17(17), p.2943. https://doi.org/10.3390/rs17172943.

Dinesh, S., Dabiri, A. and Behzadan, A.H. (2025) 'The influence of information source and modality on flood risk perception and preparedness'. *Winter Simulation Conference (WSC),* Seattle, WA, 7–10 December. Piscataway, NJ: IEEE, pp. 2264–2275. http://doi.org/10.1109/wsc68292.2025.11339065.

Dujardin, S., Arifi, D., Schmidt, S., Linard, C. and Resch, B. (2025) 'Tracing online flood conversations across borders: A watershed-level analysis of geo-social media topics during the 2021 European flood', *Natural Hazards and Earth System Sciences*, 25(7), pp. 2351–2369. https://doi.org/10.5194/nhess-25-2351-2025.

Emergency Events Database (EM-DAT) (2026) *EM-DAT | The International Disaster Database* [online] Available at: https://www.emdat.be/ (Accessed: 18 August 2026).

Ennis, S., Tiv, M., Fernandez, L. Bhaskar, R. and Porter, S. (2024) 'Examining Racial Identity Responses Among People with Middle Eastern and North African Ancestry in the American Community Survey' [online] Available at: https://www.census.gov/library/working-papers/2024/adrm/CES-WP-24-14.html (Accessed: 22 August 2026).

Federal Emergency Management Agency (FEMA) (2024) *FEMA Hazard and Risk Information Platform* [online]. Available at: https://hazards-fema.maps.arcgis.com/ (Accessed: 12 June 2024).

Gao, S., Yang, T., Xu, Y., Mou, N., Wang, X. and Huang, H. (2025) 'Enhancing disaster situation awareness through multimodal social media data: Evidence from typhoon Haikui', *Applied Sciences*, 15(1), pp. 465–465. https://doi.org/10.3390/app15010465.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H. and Kanae, S. (2013) 'Global flood risk under climate change', *Nature Climate Change*, 3(9), pp. 816–821. https://doi.org/10.1038/nclimate1911.

Jonkman, S.N. (2005) 'Global Perspectives on Loss of Human Life Caused by Floods', *Natural Hazards*, 34(2), pp. 151–175. https://doi.org/10.1007/s11069-004-8891-3.

Lechowska, E. (2018) 'What determines flood risk perception? A review of factors of flood risk perception and relations between its basic elements', *Natural Hazards*, 94(3), pp. 1341–1366.

Lim, N.J., Brandt, S.A. and Seipel, S. (2024) 'Assessment of how uncertainty representation in flood maps can affect geographic-based decisions', *Discover Water*, 4(1), p. 120.

Liu, Q., Du, M., Wang, Y., Deng, J., Yan, W., Qin, C., Liu, M. and Liu, J. (2024) 'Global, regional and national trends and impacts of natural floods', 1990–2022', *Bulletin of the World Health Organization*, 102(6), pp. 410–420. https://doi.org/10.2471/BLT.23.290243.

Navalkar, A., Tripathy, S.S., Gupta, M., Basu, S., Tripathy, P., Banerjee, A., Sekharan, S., Murtugudde, R. and Ghosh, S. (2025) 'NLP-driven crowdsourcing for urban flood monitoring: Insights from Mumbai', *Sustainable Cities and Society*, 132, p. 106795. https://doi.org/10.1016/j.scs.2025.106795.

NOAA (2024) *Sea Level Rise Viewer* [Online]. Available at: https://coast.noaa.gov/slr/ (Accessed: 12 June 2024).

NOAA (2025) *Billion-dollar weather and climate disasters* [Online]. Available at: https://www.ncei.noaa.gov/access/billions/ (Accessed: 10 November 2025).

Nominatim.org. (2026) *Overview - Nominatim API 3.6.0* [Online]. Available at: https://nominatim.org/release-docs/3.6/api/Overview/ (Accessed: 10 November 2025).

Pedoth, L., Stawinoga, A.E., Thaler, T., Schneiderbauer, S., Carnelli, F. and Damyanovic, D. (2025) 'Impacts of hazard maps on individual reaction? Results from a case study in South Tyrol, Italy', *International Journal of Disaster Risk Reduction*, 130, p. 105803. https://doi.org/10.1016/j.ijdrr.2025.105803.

Perić, J. and Cvetković, V.M. (2019) 'Demographic, socio-economic and phycological perspective of risk perception from disasters caused by floods: Case study Belgrade', *International Journal of Disaster Risk Management*, 1(2), pp. 31–43. https://doi.org/10.18485/ijdrm.2019.1.2.3.

Perlstein, S. (2024) 'Risk perception and interpersonal discussion on risk: A systematic literature review', *Risk Analysis*, 44(7), pp. 1666–1680. https://doi.org/10.1111/risa.14264.

Rahman, M.M., Shobuj, I.A., Islam, Md.R., Islam, Md.R., Hossain, Md.T., Alam, E., Hattawi, K.S. Al and Islam, M.K. (2024) 'Flood preparedness in rural flood-prone area: A holistic assessment approach in Bangladesh', *Geomatics, Natural Hazards and Risk*, 15(1), p. 2379599. https://doi.org/10.1080/19475705.2024.2379599.

Salley, C., Fox, N. and Schubert, A. (2025) 'Bridging the gap in flood risk communication: A comparative study of community and organizational social media posts using natural language processing', *Frontiers in Communication*, 10, p. 1553746. https://doi.org/10.3389/fcomm.2025.1553746.

Scovell, M., McShane, C., Swinbourne, A. and Smith, D. (2021) 'Rethinking risk perception and its importance for explaining natural hazard preparedness behavior', *Risk Analysis*, 42(3), pp. 450–469. https://doi.org/10.1111/risa.13780.

Seenath, A., Mahadeo, S.M.R. and Blackett, M. (2025) 'Decision-making under flood predictions: A risk perception study of coastal real estate', *Risk Analysis*, 45(7), pp. 1899–1925. https://doi.org/10.1111/risa.17706.

Summers, J. K., Lamper, A., McMillion, C. and Harwell, L. C. (2022) 'Observed Changes in the Frequency, Intensity, and Spatial Patterns of Nine Natural Hazards in the United States from 2000 to 2019', *Sustainability*, 14(7), pp. 1–23. https://doi.org/10.3390/su14074158.

U.S. Census Bureau (2021) *AGE AND SEX, American Community Survey, ACS 1-Year Estimates Subject Tables, Table S0101* [Online]. Available at: https://data.census.gov/table/ACSST1Y2021.S0101/ (Accessed: 22 August 2026).

U.S. Census Bureau (2022) *Educational Attainment, American Community Survey, ACS 1-Year Estimates Subject Tables, Table S1501* [Online]. Available at: https://data.census.gov/table/ACSST1Y2022.S1501/ (Accessed: 22 August 2026).

U.S. Census Bureau (2025), *Annual Estimates of the Resident Population by Sex, Age, Race, and Hispanic Origin for the United States, April 1, 2020 to July 1, 2025 (NC-EST2025-ASR6H)* [Online]. Available at: https://www2.census.gov/programs-surveys/popest/tables/2020-2025/national/asrh/nc-est2025-asr6h.xlsx/ (Accessed: 22 August 2026).

U.S. Joint Economic Committee. (2025) *Flooding Costs the U.S. Between $179.8 and $496.0 Billion Each Year* [online]. Available at: https://www.jec.senate.gov/public/index.cfm/democrats/2024/6/flooding-costs-the-u-s-between-179-8-and-496-0-billion-each-year (Accessed: 22 August 2026).