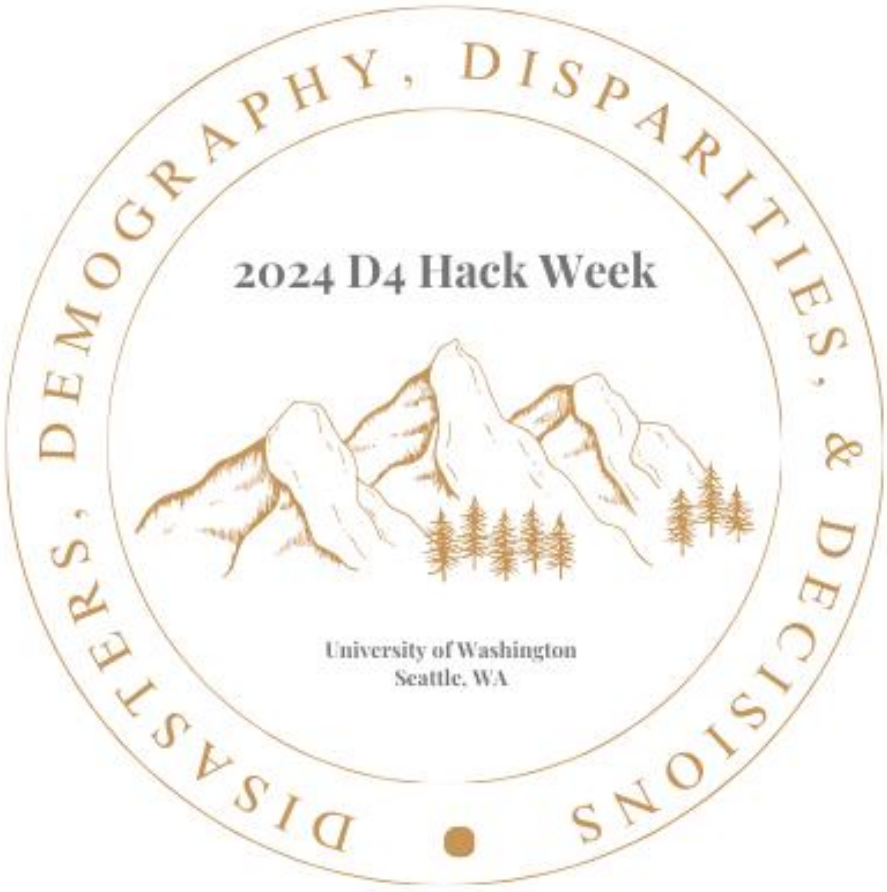


Report Published: 02/20/2025

Hack Week: September 9th – September 13th, 2024

GitHub Site: D4 Hackweek

Acknowledgements

This workshop is supported by a partnership between the University of Washington (UW)'s Center for Studies in Demography and Ecology (CSDE), the National Science Foundation AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), and the University of Washington eScience Institute.

Funding for the workshop derives from a grant from NOAA to AI2ES (Award NA23OAR40505031) and a center grant to CSDE from the Eunice Kennedy Shriver National Institutes of Child Health and Human Development via the P2C HD042828 mechanism.

In his role as a *John C. Garcia Term Professor*, Dr. Shah recognizes the generous financial support made possible by Carole Garcia.

Support for the workshop organization is also provided through the Weyerhaeuser Endowed Professorship in Environmental Policy at the University of Washington Evans School of Public Policy & Governance.

Special thank you to the Organizing Team, including: Ann Bostrom, Sara Curran, Sameer Shah, David John Gagne II, Amy McGovern, John Williams, Andrea Schumacher, Julie Demuth, Jonathon Mote, Adam Tasca, Jorge Celis, Masha Vernik, Maddie Farris, Courtney Allen, Desiree Salais, and August Wettstein.

Thank you to all the participants who actively engaged and produced important research products.

WORKSHOP REPORT
The 2024 D4 Hack Week:
Disasters, Demography, Disparities, and Decisions

CONTENTS

1. Goals of workshop

The University of Washington (UW) hosted a 5-day collaborative workshop from September 9-13, 2024, to advance research products and methods for improving observations, assessments, and forecasts across appropriate temporal and spatial scales to accomplish three goals:

1. Investigate the human behavior and societal adaptive responses to, and impacts of, severe weather and climate-related events, particularly flooding associated with atmospheric rivers, hurricanes, and severe storms, but also including other extreme events such as heat or fire.
2. Address the research gaps linking mitigation to adaptation and resilience in relation to severe weather. This was intended to involve exploring co-benefits for human well-being from climate adaptation strategies that will further contribute to resilience to extreme weather events and climate mitigation.
3. Explore pathways to better understand the dynamics of decisions and population disparities in responses to and impacts of past extreme climate / weather events.

Funded by the National Institutes of Health through the UW Center for Studies in Demography and Ecology (CSDE) and the National Oceanic and Atmospheric Administration (NOAA) through the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), and supported by UW eSciences and the National Center for Atmospheric Research (NSF NCAR), the workshop aimed to create improved data products and methods (data integration, data assimilation, analytic tools, new approaches to analyses) that integrate social and weather and/or climate data across space and time, through interdisciplinary collaborations. Data products were envisioned as informing decision models that can guide decision making to address the needs of individuals, households, neighborhoods, and communities, with projections of impacts on the scales of minutes to hours, days, weeks, or years. Data and their analyses were intended to be capable of informing impact analysis and risk reduction planning.

The workshop focused on the data and analytic challenges of linking climate- and weather-related impacts and mitigation efforts to human behavior, health, and well-being by:

- facilitating all participants' access to data sources and tools in a secure computing environment, when required, and provide open source tools and resources otherwise;
- exploring and innovating new approaches, including the use of machine learning (ML), to assimilate, curate, and analyze the relevant multiplicity of data required to assess and predict impacts of extreme weather events across contexts and scales.
- incorporating vital, mediating societal, institutional, organizational, and relational factors into understandings of the environment. The workshop organizers assume that societal and natural landscape factors are crucial mechanisms in managing mitigation and adaptation, but until now these factors have been incompletely incorporated into models that link climate change or weather to human and societal well-being;

- catalyzing research that more completely models weather- and climate-human linkages and accelerates knowledge and provision of tools for policymakers, emergency managers, businesses, and the public.

Prior to the workshop, participants were expected to participate in several activities related to preparing for the workshop, which included an online session to introduce the participants and their projects or workshop roles to one another. In that same session, the organizers introduced the goals of the workshop, and participants described their workshop-related expertise, topical focus, and their methodological focus.

Workshop materials were shared and developed with and by participants and organizers using the Jupyter Notebook platform: https://d4hackweek.github.io/d4book/

2. Lessons for the NOAA Societal Data Insights Initiative

NOAA liaison Jonathon Mote kicked off the closing session of the D4 hack week by asking attendees to consider the manifest functions of the hack week, which he identified as: exploring AI/ML methods, identifying new or core datasets, exploring the usefulness of this type of event, and exploring the challenges posed by integrating social, behavioral, and demographic data with weather hazard data. He also suggested considering the latent functions of the hack week, specifically, gaining user experience insights in a collaborative setting. Lessons for NOAA SDII emerged across all of these, the characteristics of which are described in more detail in the remainder of this report, and are implicit in the products proposed in section 5.

Key takeaways from the D4 hack week include:

- The **extensive advance preparation** for the D4 hack week paid off. Teams arrived ready to work, and made progress on their projects during the week. They were also able to learn about data sources, tools, and research methods from interactions with one another and the Floaters (both informal and in the 'team speed-dating' sessions at the beginning of the workshop, and office hours with Floaters), and to contribute to active plenary, panel-led discussions of *AI for data integration*, *Data integration opportunities and challenges*, and *Uncertainty, focusing on the analysis and representation of uncertainties that surface with integration of diverse datasets*.

- The organization and operation of the D4 hack week could be characterized as a **team-of-teams** approach. As noted in section 5, the teams' and floaters' breadth of expertise was key to the success of the workshop. Although team projects were distinct and unique, by design the projects and teams included in the hack week shared common questions, problems, and goals that were central to the workshop. Striking this balance seems important for the success of such events. Bringing people in for an in-person meeting was deemed crucial by many; the social life of information at the workshop was enhanced by hack week community meals and happy hours.

- While the types of methodological advances teams demonstrated at the hack week appear extremely promising, a variety of **obstacles currently hinder data access and integration to study societal impact**, such as:

  (a) data challenges – access (lack of awareness of existing data; legal, administrative, fiscal, and technical challenges to accessing existing data); privacy (data walls, differences in privacy of individually identifiable data across the public and private sectors); aggregation (Simpson's paradox, mismatches in data collection and aggregation approaches across datasets).

  (b) data integration challenges - uncertainty, discrepancies between units of analysis in different datasets, and technical and computational challenges of analyzing, storing, and representing (e.g., visually) integrated datasets due to their size and diverse characteristics.

- **Roles and limitations of AI for SDII**: Most of the teams at the workshop were led by seasoned demographic and environmental researchers, but few of the teams had experience with machine learning or were accustomed to considering machine learning as an approach to data integration and analysis. With hands-on help from Floaters, some teams gained insight into the potential value of machine learning for data assimilation and analysis during the hack week. In hindsight the tutorial that Floater Jason Stock walked through during the hack week may have been too advanced or have required more time and support for hack week attendees to benefit fully from it. Organizers suspect this is indicative of the state of the field, and that social and demographic researchers may be lagging in the uptake of recent rapid advances in ML.

- **Hack week-specific lessons**: On balance this type of event offers a way of learning and advancing knowledge relevant to SDII that differs from what other types of workshops or individually funded research projects might contribute. The issues raised and discussed at the hack week are highly relevant to the eventual shape and success of SDII.

  There was agreement that even more NOAA participation in the D4 hack week would have been valuable, but participating team sizes should be kept small (3-5), and the workshop size as a whole should ideally be limited to under 30 participants.

  With regard to the integration of ML and other methods advancements into SDII small-group in-person tutorials tailored for participants' level of familiarity with methods/data might be more effective than the online-only or plenary tutorials the D4 hack week provided.

  While at the end of the week teams wished they'd had more time to work on their projects, most of the scheduled activities were assessed as valuable (with the exception of plenary discussions, which some thought could be shortened; see section 6 and Appendix 1). Furthermore, it was difficult to schedule even the few days we asked participants to carve out for the hack week.

3. D4 Hack Week Development

The CSDE 2022 Population Dynamics Center renewal included funding for a new effort to "intentionally and deliberately spark new collaborations and launch new research initiatives that effectively leverage the capacities of demographers, data scientists, health scientists and scholars of inequality" in annual 2-day events each autumn, organized by each Primary Research Area (PRA) around presentations on the theme of Demography, Data Innovation, and Population Disparities. A theme of studying disparate impacts (e.g., housing, and migration) on vulnerable populations from climate change and environmental hazards and disasters emerged from an initial CSDE brainstorming session, with the Environments and Populations PRA identified as lead for an initial event. In February 2023 Sara Curran (CSDE Director), Sameer Shah, and Ann Bostrom (PRA chair for Environments and Populations) met to start drafting the goals and logistics for this event, the emerging focus of which became a data hacking event with invited teams, convened to develop new research collaborations and advance novel data integration and linkages for insights on changes in climate, demography, and vulnerability. Specifically, this event would tackle the challenges of identifying and accessing temporally and spatially overlapping data (e.g., census, social, flood, ecological, and novel types of data) at high enough resolution to understand what might happen to people who are particularly vulnerable and inform interventions to increase their resilience and adaptive capacity.

In spring 2023 the opportunity also arose for risk communication researchers in AI2ES to propose research to NOAA that would help inform NOAA's Societal Data Insights Initiative. The opportunity to collaborate with UW CSDE on this effort promised important synergies, and was developed as the second thread of the proposal to NOAA. This thread proposed to "convene a handful of small, interdisciplinary expert research teams for a week-long active, hackathon-type workshop. This workshop will explore and innovate new approaches, including the use of machine learning (ML), to integrate meteorological and other environmental data with demographic and other types of social data to study changes in flood events, such as how the evolution of a flood threat and flooding, and institutional responses to these, can affect short-term mobility, information searches, or displacement. Specifically, the workshop will explore pathways to better understand the dynamics of decisions and population disparities in responses to and impacts of past atmospheric river event(s) along the U.S. West Coast." The proposal *[BIL: Integrating social and meteorological data to assess the dynamics of flood hazards and impacts: An interdisciplinary approach leveraging AI, risk communication, and data sciences]* was funded by NOAA in late summer 2023.

The initial organizing group expanded over the next few months to include two graduate research assistants and representatives from NOAA and from the other threads in the BIL grant, as well as consultations with eSciences at the UW. With an initial meeting in December, 2023, we set a regular cadence of virtual meetings to organize the hack week, inspired by UW

eSciences hackathon experience (https://2024.hackweek.io/), Shah et al. (2023), Voorheis et al. (2023), NOAA SAB (2021), and McGovern et al., 2023.[1]

The organizing group established a google drive and folders (AI2ES), developed a webpage about the event (CSDE), and developed other infrastructure and materials to support the hack week, including a Jupyter notebook. We identified participants in April, then surveyed them to learn their plans, needs, and expectations, and held a virtual orientation meeting for all invited participants. This was followed by other virtual meetings and the preparation and distribution of a Jupyter notebook, which contained several tutorials, in advance of the workshop to help participants prepare. Critical to the entire effort was the continuity afforded by superb support from graduate research assistants, CSDE staff, NOAA liaisons, and members of the organizing group (see section 8.2).

The organizing group distributed the call for applications to participate in the hack week (travel and per diem fully funded) widely, but also sent targeted invitations to known research groups, and developed a collaborating across threads (CAT) demonstration project based on the third thread in the NOAA BIL grant (new longitudinal research on an atmospheric river event), led by Julie Demuth and Andrea Schumacher, which had successfully collected longitudinal data from an atmospheric event at the end of March. The demonstration project involved a wider set of the workshop organizers, including also new AI2ES postdoctoral scholar Jorge Celis advised by PI Amy McGovern. Individual applicants invited to participate in the workshop were either added to an existing team, invited to form a new team (composition suggested by the organizing group), or invited to act as "Floaters" to help teams troubleshoot a variety of data and methodological issues, before and at the hack week. One new team was formed this way and three individuals elected to be Floaters. Floaters also joined some of the hack week planning meetings. In addition to CAT demonstration project meetings (led by Andrea Schumacher and Julie Demuth), separate virtual meetings were also held by subsets of the organizing committee on infrastructure planning (led by David John Gagne) and on travel logistics and budgeting (led by CSDE Program Coordinator Madeleine Farris and Graduate Research Assistant Masha Vernik).

[1] McGovern, A., Gagne, D.J., Wirz, C.D., Ebert-Uphoff, I., Bostrom, A., Rao, Y., Schumacher, A., Flora, M., Chase, R., Mamalakis, A. and McGraw, M., 2023. Trustworthy Artificial Intelligence for Environmental Sciences: An Innovative Approach for Summer School. Bulletin of the American Meteorological Society, 104(6), pp.E1222-E1231.

NOAA Science Advisory Board (2021). A Report on Priorities for Weather Research. NOAA Science Advisory Board Report, 119 pp. https://sab.noaa.gov/wpcontent/uploads/2021/12/PWR-Report_Final_12-9-21.pdf

Shah, S.H., O'Lenick, C.R., Wan, J.S., Ramos-Valle, A., Ash, K.D., Wilhelmi, O.V., Edgeley, C.M., Molina, M.J., Moulite, J., Pizarro, C.A.C. and Emard, K. (2023). Connecting physical and social science datasets: challenges and pathways forward. Environmental Research Communications, 5(9), p.095007. .DOI 10.1088/2515-7620/acf6b4

Voorheis, J. L., Colmer, J. M., Houghton, K. A., Lyubich, E., Munro, M., Scalera, C., & Withrow, J. R. (2023). Building the prototype Census environmental impacts Frame Working Papers 23-20, Center for Economic Studies, U.S. Census Bureau, National Bureau of Economic Research.NBER Working Paper No. 31189, https://www.nber.org/system/files/chapters/c14833/revisions/c14833.rev0.pdf

4. Teams, project descriptions, and progress

The eight teams who participated in the hack week developed their research ideas and goals before they arrived for the hack week, and made progress on these during the week, as described in this section.

4.1 Claim to Flame team

***Team members**: Joan Casey, Lauren Wilner, Vivian Do, Heather McBrien, and David Coomes*
***Institutions**: University of Washington, Columbia University*

Project description: This project plans to integrate data on wildfires and wildfire boundaries, FEMA household claims, and demographics, with CalEnviroScreen data to ask several questions about FEMA assistance for wildfire disasters: What are the population characteristics of those actively applying for FEMA assistance for wildfire disasters? Within this group, what are the individual (e.g., owner vs. rental status, level of disaster preparedness, reliance on electrical medical equipment) and area-level (e.g., neighborhood poverty) factors associated with successfully receiving FEMA aid or the amount of FEMA aid? Do the factors associated with successfully receiving FEMA aid vary by disaster type (i.e., wildfire disaster vs. wildfire disaster co-occurring with extreme heat)? The overarching goals of the project are to:

1. Describe FEMA claims related to wildfire in California from 2010-2019
2. Among similarly-exposed ZCTAs, compare individual and ZCTA-level characteristics of those making a FEMA claims post-wildfire
3. Evaluate sociodemographic inequities in (a) successful claims and (b) amount of claim paid out

Progress: During the hack week, the team conducted multiple analyses addressing their three research questions, with initial findings leading to the following questions and conclusions:

- What does missing data mean for variables across the FEMA dataset?
- Complexity of assigning ZCTA-level wildfire exposure.
- Four out of ten ZCTAs only have one claim. Low exposure to fire or low access to FEMA?
- What are the reasons someone may not file a FEMA claim, or may have a FEMA claim rejected, that are outside the scope of the claims dataset?
    - For example, homeowner insurance provider or individual-level internet access.
- Who is the population applying to FEMA claims?
    - How do they differ from populations affected by fires but did not apply to FEMA claims? What are ways to accurately characterize this population?

Caption: Claim to Flame – Joan Casey (Univ. of Wash), Karen Chen (Univ. of Wash), David Coomes (Univ. of Wash), Vivian Do (Columbia Univ.), Heather McBrien (Columbia Univ.), and Lauren Wilner (Univ. of Wash)

4.2 DEMUS: Demography of Environmental Migration in the United States

***Team members**: Elizabeth Fussell, Jack DeWaard, Katherine Curtis, James Done, and Sara Ronnkvist*
***Institutions**: Brown University, Population Council, University of Wisconsin – Madison, NSF National Center for Atmospheric Research*

***Project Description**:* This project plans to integrate IRS county-level migration data with data on tropical cyclones, wildfire events, flood events, wet bulb temperatures and air pollution, county health rankings and roadmaps data, and national neighborhood data archive data, to ask the broad question: How do tropical cyclones affect county-level migration systems, where migration systems are the counties connected through in- and out-migration flows, and do these dynamics differ for age, sex, race/ethnicity, and nativity groups?

Progress: This team has been working with secure data in Federal Statistical Research Data Centers (FSRDCs) for years, and were able during the D4 hack week to actually visit the NW FSRDC in person, for the first time, to look at the data and report together. The team is

examining spatio-temporal dynamics of migration flows ("currents") using both established and emerging tools, with guiding principles of parsimony and interpretability for the computational elements necessary to handle the magnitude of the data. They reported having created a matrix of county-to-county migration flows in the RDC for all counties and county equivalents from 2000 to 2020, comprehensive of the social security population, decomposable by age, sec, race/ethnicity and nativity, and linkable to environmental hazards as well as other place-level characteristics, measured at the county level. The team is looking at environmental justice through the lens of interactive intersectionality,[2] including spatial and other interconnections and interdependencies, disaggregated by populations and places. Currently their aims to contribute to open, reproducible science contravene RDC restrictions on research; while the team would like for the dataset they have created to become publicly available, under the same restrictions that the IRS data are available, at this point they don't see a path forward on this. Anything that is only knowable with the data within the RDC are not discussable outside the RDC until they have been vetted by the Census. Census differential-privacy algorithms might potentially be applicable to take the data out of the RDC, but the RDCs have decided they are not in the business of disclosing datasets.

Caption: DEMUS - Elizabeth Fussell (Brown Univ.), Katherine Curtis (Univ. of Wisconsin), Sara Ronnkvist (Univ. of Wisconsin). Not Pictured: Jack DeWaard (Population Council), James Done (NSF National Center for Atmospheric Research)

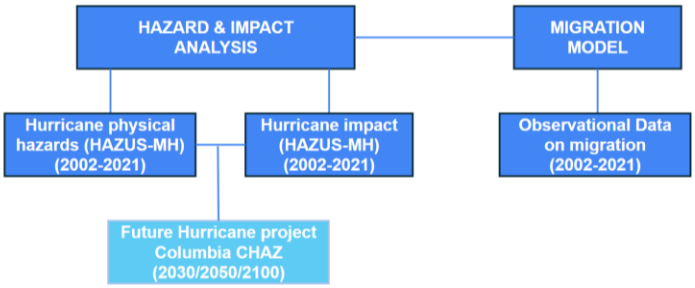
[2] Maharjan, A., del Valle, A., Erulkar, A., Mishra, A., Steidl, C., Singh, C., Sharma, D., Riosmena, F., Pinillos, G., Abel, G., DeWaard, J., Ha, J. T., Donato, K. M., Madise, N., Nawrotzki, R., Nevarez, R., McLeman, R., & Hussein, S. A. (2024). The Migration Intersections Grid: An Organizing Framework for Migration Research in and through the Twenty-first Century. International Migration Review, 58(4), 1937-1973. https://doi.org/10.1177/01979183241275469

4.3 Columbia Hurricane Migration

***Team Members**: Fabien Cottier, Mona Hemmati, Andrew Kruczkiewicz, and Kytt MacManus*
***Institution**: Columbia University*

Project Description: This project is integrating data on tropical cyclones, migration, and vulnerability to examine to what extent flooding and hurricanes contribute to shape migration flows in the US. The project also asks how early warning systems can be appropriately leveraged to inform risk reduction with a view towards decreasing disproportionate impact from different types of floods / hurricanes.

Progress: The team accomplished all of its hack week objectives, though they noted they could have used an additional day. Their objectives included building an initial predictive model of migration as a result of hurricane (peak) wind gusts, for demonstration and evaluation purposes, validating that model against external data sources on migration; and understanding how a predictive model of migration might support decision-makers, to produce an output/tool whose effectiveness could be evaluated by decision makers.



The team considered a variety of models, including random forest, boosting, and neural nets, and encountered challenges from data paucity and scale mismatches. They emphasized that tropical cyclones are inherently multi-hazard events, with each hazard driving different spatiotemporal impact patterns. The team recognized the need for a robust co-production process, in order to define migration in this context, and to assess what it would take for a model to be sufficient or appropriate to influence policy making. The team concluded that forecasts of socioeconomic impacts from a specific hurricane may lead to migration, or might inform regional and global efforts related to extreme events and mobility. Promising prospect included the possibilities of applying AI/ML, and incorporating risk perception data into their analyses, which might inform risk communications related to migration.

4.4 Disaster Demography

***Team members**: Deborah Balk, Dylan Connor, Melanie Gall, Lori Hunter, Jenna Tipaldo, and Helen Wilson Burns*
Institutions: City University of New York, Arizona State University, University of Colorado Boulder

Project Description: This team plans to integrate county-level Spatial Hazard Events and Losses Database for the United States (SHELDUS) data with individual- and household-level American Community Survey (ACS) Public Use Micro-data (PUMA) to examine how the burden

of hazards on the U.S. population has changed over the past 15-years and what implications this has for risk mitigation and emergency preparedness efforts. SHELDUS data include property damage (in dollars) and fatality counts from natural hazard events.

Progress: The team were excited to be able to work in person as a team and interact with like-minded research teams. The team started on one supercomputer and migrated their work to the NCAR supercomputer during the hack week (supported by the D4 hack week organizing committee and floaters). The team noted that aggregated data on vulnerability and vulnerability metrics (e.g., SOVI) are susceptible to ecological fallacy, whereas individual data are much sparser in their representations. The team illustrated how ACS PUMA level data (100K persons) can align or mismatch the county-level SHELDUS data, and noted that data are constrained to state boundaries, all of which complicated data integration. During the hack week the team linked a 1% ACS sample to SHELDUS from 2012-2021, with probabilistic allocation of ACS persons to counties[3] (resulting in about 10 million person-county observations for each ACS year). While the preliminary results they produced linking SHELDUS data to ACS PUMA were interesting, they were not able during the hack week to disaggregate by hazards in order to examine how exposures differed by hazard. Insights of relevance for decision makers included that micro-data will provide much more nuanced information about who is at risk, and thus will help focus resiliency efforts, adaptation planning, and early warning. To advance the project further, the team noted a need for continuing supercomputer support, engagement with others, and support for visualization of the multiple strata of data.

4.5 The Floodsters

Team members*: ChangHoon (Chang) Hahn, Sharif Islam, and Lidia Cano*
Institutions: Princeton University, Massachusetts Institute of Technology

Project Description: To address the broad questions of “How will existing flood management tools deal with future climate scenarios?” and “Which communities will be the most affected/most protected as a result?”, this team is planning to integrate Federal Emergency Management Agency (FEMA) data from the National Flood Insurance Program, American Community Survey data, data from First Street, metadata on perceptions of future events/risks, and possibly international data (EM-Dat and the Extreme Weather dataset).

Progress: In their larger project this team is looking at how flood adaptation methods have worked historically, but also how they will work in the future, for urban communities most at risk. During the hack week this team built on their recently published research[4] to examine how flood adaptation might be effective and for whom in a future with climate change. With support from a

[3] Ethan Sharygin suggested the team consider using maximum entropy for imputation of the county/tract associated with a PUMA record, as suggested by Ruther, M., Maclaurin, G., Leyk, S., Buttenfield, B., & Nagle, N. (2013). Validation of spatially allocated small area estimates for 1880 Census demography. *Demographic research*, *29*, 579-616.

[4] Cano Pecharroman, L., & Hahn, C. (2024). Exposing disparities in flood adaptation for equitable future interventions in the USA. *Nature Communications*, *15*(1), 8333. https://doi.org/10.1038/s41467-024-52111-0

floater, they were able to compare FirstStreet flood forecasts to CMIP6 forecasts (the median of the max precipitation forecasted per year over the 36 models), to estimate median flood losses and FEMA Community Rating System (CRS) savings for each zip code in Houston, New York, and Chicago up to 2052.

Caption: Floodsters – ChangHoon Hahn (Princeton Univ.), Lidia Cano (MIT), Sharif Islam (MIT).

4.6 DemograFires

***Team members**: Mathew Hauer, Alexis Santos, and Sunshine Jacobs*
***Institutions**: Florida State University and Pennsylvania State University*

Project Description: This project plans to integrate ACS/Census Data and IRS migration data with novel demographic change modeling and data on environmental hazard events to ask: How do populations change in association with environmental events? How do different race/ethnic groups change in association with environmental events? And what are the long-range impacts of environmental change on demographic change?

Progress: This team considered both theory-driven and data-driven approaches to variable selection, and ultimately went with a leading conceptual theory of vulnerability as the intersection of exposure, sensitivity, and adaptive capacity. Not all key theory-driven variables were available through the American Community Survey, however. The team also encountered issues of aggregation, examined both census tracts and counties as the unit of analysis, and

wrestled with the modifiable areal unit problem. A key takeaway was that the level of aggregation really matters because vulnerability is impactful.

Caption: DemograFires – Mathew Hauer (Florida State Univ.), Sunshine Jacobs (Florida State Univ.), and Alexis Santos (Penn State Univ.)

4.7 Migration Mavericks

***Team members**: Ethan Sharygin, Justin Stoler, Mary Angelica Painter, Sameer Shah*
Institutions: Portland State University, University of Miami, University of Colorado Boulder

Project description: This team is considering addressing the following questions:

1) How are people displaced after a wildfire and what are their characteristics? How are water insecurity and distrust in water utilities and services related and shaped by hazard experiences? What is the relationship between government capacity, community resiliency, and weather events? Data under consideration for integration to address these questions include 2024 nationally representative survey data with modules on water insecurity and institutional trust, migration destinations post-wildfire, rural capacity index, local hazard mitigation plan status, and community resiliency estimates.

2) Estimate population displacement effects from sudden onset disaster and compare network of migration destinations after sudden onset disaster to regular migration pathways, commute patterns, and parametric migration models.

Progress: The team found a dearth of scholarship on wildfire-induced migration within the U.S., and mixed evidence regarding the potential influence of wildfires on migrations. The team envisions developing a simulated model of migration flows to help decision-makers locate places likely to experience such flows. Among data considered: IRS income statistics, longitudinal employer household dynamics, integrated parcel data and mail forwarding, consumer credit panel data, and data on rural capacity and local hazard mitigation status. The team aims to integrate gravity and radius models, which have not traditionally been used to investigate such questions, and to make advances regarding spatial and temporal granularity (e.g., where do people get immediately displaced). Among questions raised in the discussion of this project were how to think about the temporality of vulnerability in a place, and what regulatory and rate environments may be influencing homeowners' insurance.

Caption: Migration Mavericks – Mary Angelica Painter (Univ. of Colorado Boulder), Ethan Sharygin (Portland State Univ.), and Sameer Shah (Univ. of Washington)

4.8 CAT Demonstration Project

Team members*: Andrea Schumacher, Julie Demuth, DJ Gagne, Jorge Celis, Amy McGovern, Sara Curran, Sameer Shah, Masha Vernik, Ann Bostrom*

***Institutions**: NSF National Center for Atmospheric Research, University of Oklahoma, University of Washington*

Project Description: This team aims to integrate multidisciplinary, time evolving data to investigate changes in driving behaviors before and during a southern California atmospheric river flooding event. The motivation for this project is to investigate whether people change their driving behaviors as an atmospheric river (AR) flooding event is threatening and occurring and, if so, 1) who changes their driving behaviors, 2) how driving behaviors change, and 3) when those changes occur. This case study focuses on an AR event that occurred in Southern California from 29-31 March 2024, and integrates data from multiple disciplines such as longitudinal panel survey data, American Community Survey (ACS), weather, and mobility data to examine changes in driving behavior surrounding this event through various disciplinary lenses. This project also seeks to address questions related to the process of integrating data from different disciplines, including: How do different datasets, individually and together, answer these questions? Where are there similarities and differences? How can we have a focused examination on population disparities? What types of data are needed to meaningfully explore these groups? What can we learn in doing so?

Progress: The CAT Demonstration Project benefited from numerous discussions and insights at the workshop, including a tip about scikit-mobility from eSciences researcher Spencer Wood, and numerous insights from Floaters. During the week the team made progress on accessing mobility data, including obtaining a sample dataset from Veraset. The team also began examining the representativeness of their March 2024 longitudinal survey data by comparison with the ACS, and crystalized their research questions and analytical strategies.

4.9 FLOATERS

***Team members**: Jason Stock, Tyler Fricker, Patrick Greiner*
***Institutions**: Colorado State University, University of Louisiana Monroe, Vanderbilt University*

***Task description*:** Floaters are scientists who bring their expertise to tackle data integration challenges, model development, AI, machine learning solutions, and more. They will hold office hours and dedicate time to answering team questions.

Jason Stock is a Computer Science PhD candidate at Colorado State University (CSU) advised by Professor Chuck Anderson. His current research interests are in neuro-inspired attention methods, generative diffusion models, creating interpretable-by-design machine learning algorithms, and modeling weather and climate change.

Tyler Fricker is an Assistant Professor of Geography in the School of Sciences at the University of Louisiana Monroe. His research interests are, broadly, in the environmental impacts of natural hazards on society and the connection between weather and climate. Much of his work bridges the gap between climate change science and climate-society interaction.

Patrick Greiner is an environmental sociologist at Vanderbilt University (at the University of Washington as of fall 2024). His research and teaching address questions at the intersection of structural inequality, development processes, and environmental change.

5. Proposed products and next steps

The Center for Studies in Demography and Ecology (CSDE) is organizing a follow-on virtual half-day workshop, scheduled for February 7, 2025. The aim of this workshop will be to update on the progress teams have made on their D4 projects, and to coordinate on additional next steps stemming from the workshop, including the following proposed products:

Papers

- **"Team of Teams" Paper**
    - Why: One factor that contributed to the hack week's success was the teams' breadth of expertise. While teams' projects were distinct, they shared a set of questions, problems, and goals.
    - What: We propose a paper discussing the 'space between teams' to achieve cross-pollination between adjacent projects. This paper would emphasize the importance of achieving the right balance of difference and similarity between teams – enough similarity in order to talk to each other, but not so similar that we're in an echo chamber. Such a paper may include a network-like conceptualization of knowledge and people in our hack week. It may also examine how teams' ideas changed over time after exposure to other teams' projects.

- **Data Walls**
    - Why: To various extents, teams ran into problems related to data access, making it difficult to answer research questions and discuss findings. Teams ran into different kinds of challenges depending on if the data was from a public or private source.
    - What: We propose a paper discussing data walls – why they exist, how much they cost, and how and when to overcome them. This paper may also discuss challenges to public-private partnerships to overcome datawalls. We would draw a distinction between the inaccessibility of micro-data from government sources, as well as the over accessibility of private data (if you have the money). Ethical considerations around privacy would figure prominently in the paper, and we would address whose data should be protected, by whom, and how.

- **Aggregation problem:**
    - Why: Virtually every team carefully considered aggregation in some, often multiple, forms. Teams made decisions about the appropriate level of spatial

aggregation, and considered the appropriateness of including other indices that aggregate data, like social vulnerability and risk indices.
    - What: We propose a paper that would address questions such as: What gets lost when we over-aggregate? How do we move from specific to general without losing important details? How do we emphasize important details without becoming anecdotal and lose broader scientific relevance? Here, we may also discuss the 'story of the data' and how it's important to know where data comes from in order to understand how to aggregate it and at what level. Incorporating qualitative data throughout the research process may also be a way to buffer against the aggregation problem.

Boundary object:

- Why: Throughout the week, we grappled with similar questions around data integration for different problems, contexts, and disciplines. It would be useful to create a tool for thinking through data integration that would be malleable enough for use in other contexts.
- What: We propose creating a boundary object, which would be a conceptual framework for integrating social and environmental datasets. The boundary object would include big questions we grappled with throughout the week, including data integration challenges, data walls, spatial and conceptual data aggregation, use of Machine Learning and AI for data integration. It may also discuss the benefits of team science and integration of ideas, and when and where it would especially be useful. We hope that such a framework would be useful for researchers and practitioners encountering similar challenges around integration of various kinds and modes of data. [Program with a methodologically oriented core (POR)]

Jupyter Cookbooks:

- Why: A practical tool to teach others about data integration and alternative format for thinking through the questions probed throughout the hack week.
- What: Jupyter Cookbooks would be a series of tutorials on topics related to the hack week, such as: accessing datasets, using AI for data integration and analysis, and deciding the appropriate level(s) of spatial aggregation. We could model this set of cookbooks on Project Pythia, and we'd need to make sure to update dependencies so the cookbooks would run even with updates. We may also be able to recycle the existing tutorials into this.

Data sources database:

- Why: A theme that emerged throughout the hack week was the importance of knowing where your data comes from – without a solid understanding of your data sources, results are hard to interpret. Furthermore, teams seemed to benefit from knowing about

others' datasets, and it might be useful for people doing similar work to compile relevant datasets.

- What: We could compile all the sources we used. This could take the format of a paper or a searchable online database (so a database of databases). We would want people across disciplines to be able to use the database. Such a database would emphasize the story of the data – how was the data collected? Who and how were key decisions made about its collection? We would also focus on the human side of the data – the data are not detached from people's lives, but represent a small point from their lives.

6. Evaluation survey

The evaluation survey was informed by the workshop goals and the closing questions drafted in advance of the workshop by NOAA D4 liaison Jonathon Mote. After review and revisions by the organizing team the survey was implemented in Qualtrics through Slack and email online, in the late afternoon of the final full day of the workshop, Thursday, September 12th, during a long break before the closing discussion. Responses were anonymous. Reminders were sent immediately after the workshop and a few days later.

Eight teams attended, including the Collaborating Across Threads (CAT) Demonstration project team, led by members of the D4 organizing group. From the other seven teams 26 attended the workshop in person, and three remotely. Three in-person attendees acted as Floaters, working across teams by appointment and in ad hoc interactions to provide research support throughout the workshop.

What was your role in this hack week? *(check all that apply) N=26*		
Organizer	3	12%
Team lead	9	35%
Team member	15	58%
Floater	3	12%

Table 1. Evaluation survey respondent characteristics

Nominally, the organizing group ultimately included 25 people (see Appendix 2), of whom one was also a Floater, two were NOAA liaisons, one was an eScience liaison, who did not attend the workshop, two attended the workshop remotely (although not in its entirety), and six were CSDE staff or affiliates who contributed in several ways to the workshop, but did not participate in teams. The survey was sent to the entire participant list (partial, remote, and full-workshop in-person attendees). All floaters and over half of team workshop participants who were not organizers responded.

6.1 Evaluations of how well the workshop met participants' expectations

Participants were enthusiastic about their overall workshop experience, with over 60% reporting that it exceeded their expectations. On average participants reported that the workshop met most or all of their expectations on every dimension we explored, including networking, computational infrastructure and support, making progress on their research projects, developing new collaborations, and learning about new methods and tools for integrating demographic and other social data with environmental or earth system science data (Figure 1).

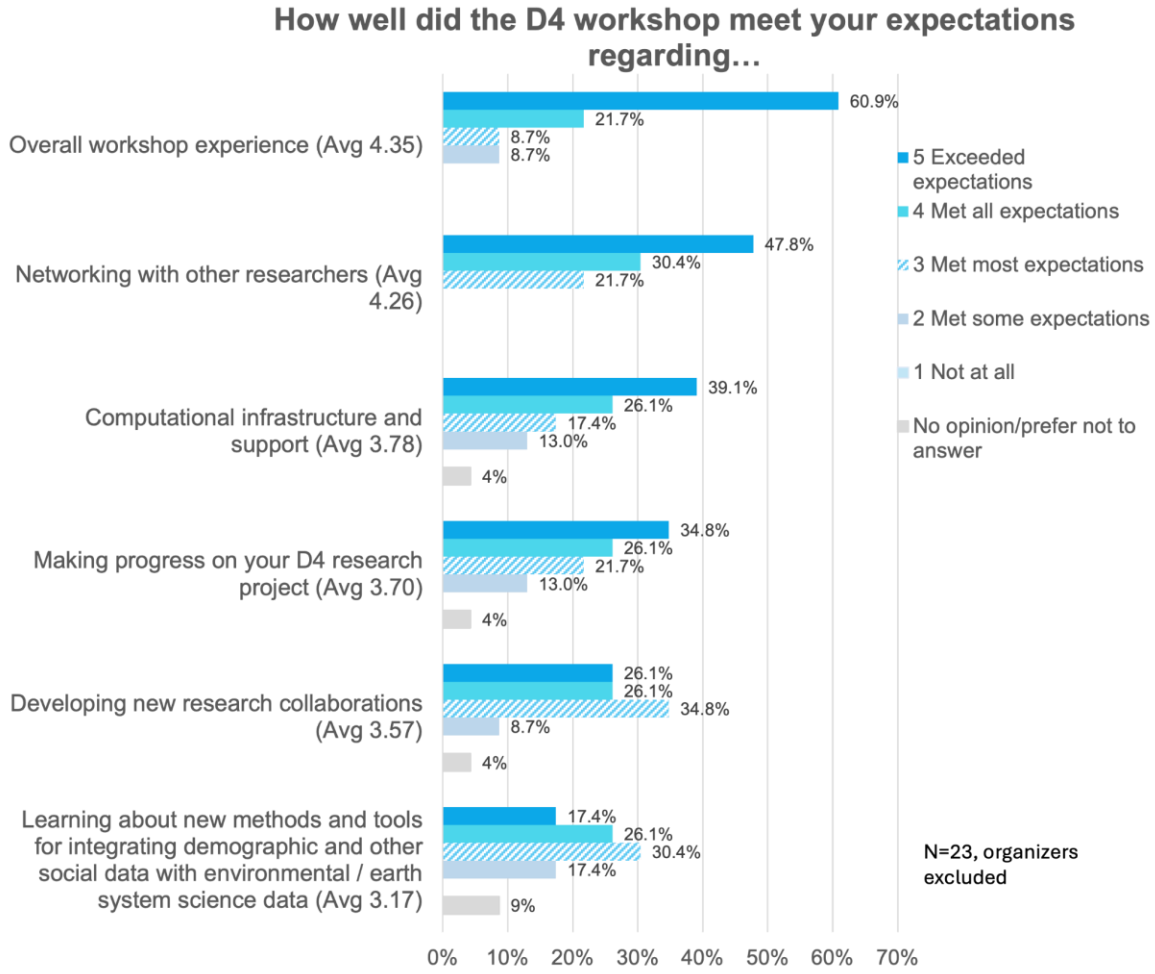


Figure 1. How well the D4 workshop met participants' expectations, on a scale from 1=Not at all, to 5=Exceeded expectations, with an explicit No opinion/prefer not to answer response category. Responses from organizers are excluded.

6.2 Evaluations of how well the workshop met organizers' expectations

Only a small proportion of the organizing group responded to the survey. Comparisons between organizers' ratings and those who did not self-identify as organizers show that organizers'

ratings were on average a little higher than those of other participants. However, a majority of the core organizers attended the debriefing meeting Friday September 13th, the day after the workshop. The general consensus at the debriefing meeting was that the workshop succeeded in achieving its primary goals, and exceeded expectations in several respects.

Evaluations of helpfulness and productiveness of elements of the workshop

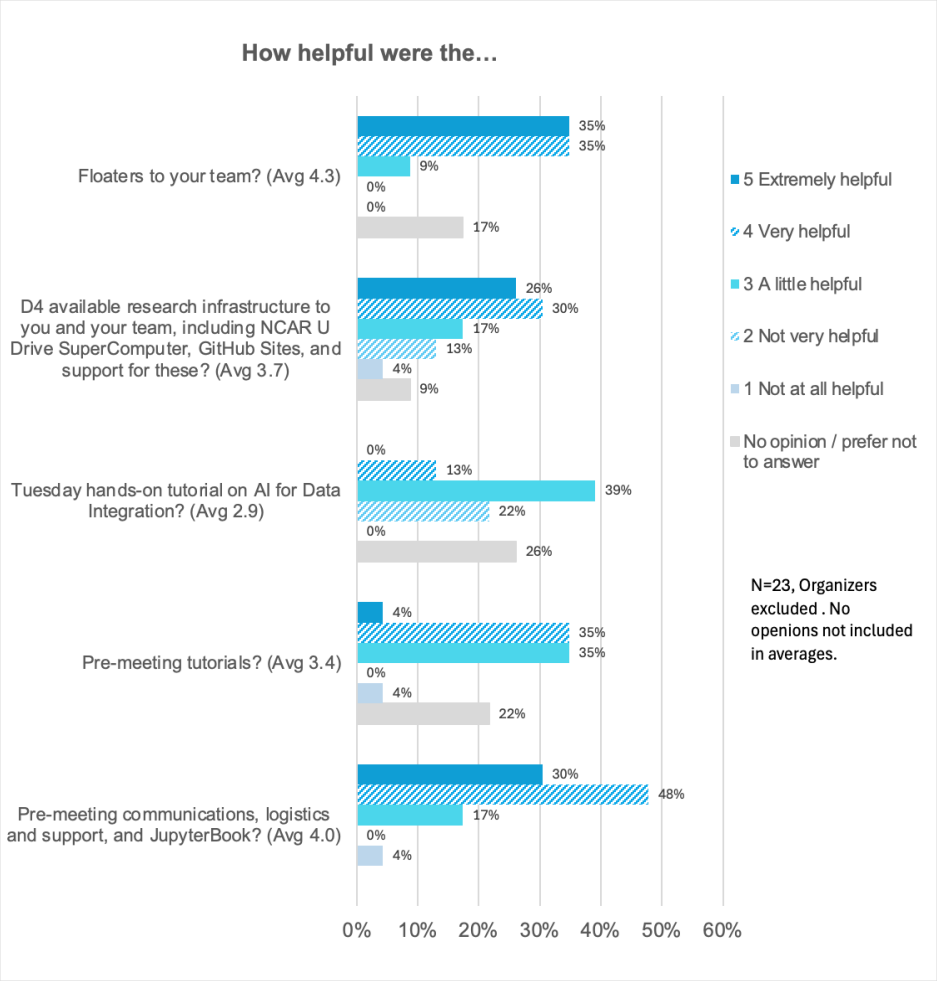


Figure 2a. Assessments of the helpfulness of workshop elements, excluding organizing committee members.

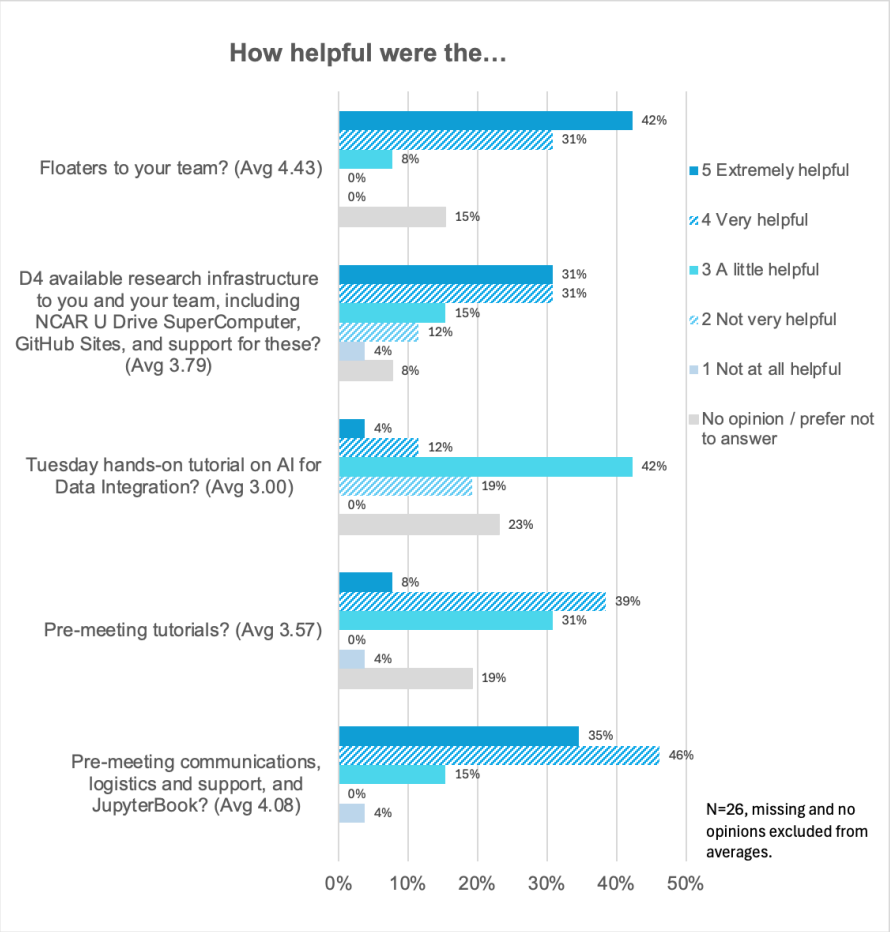


Figure 2b. Assessments of workshop element effectiveness, including respondents who were members of the organizing committee.

7. Acknowledgments

We gratefully acknowledge the many sources of funding and support for the D4 Hack Week. This workshop was supported by a partnership between the University of Washington (UW)'s Center for Studies in Demography and Ecology (CSDE), the National Science Foundation AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), the NSF National Center for Atmospheric Research, and the UW eScience Institute. Funding for the workshop derived from a grant from NOAA to AI2ES (Award NA23OAR40505031) and a center grant to CSDE from the Eunice Kennedy Shriver National Institutes of Child Health and Human Development via the P2C HD042828 mechanism. In his role as a *John C. Garcia Term Professor*, Dr. Shah acknowledges the generous financial support made possible by Carole Garcia. Support for the workshop organization was also provided through the Weyerhaeuser Endowed Professorship in Environmental Policy held by Dr. Bostrom in the UW Daniel J. Evans School of Public Policy & Governance.

8. Appendices

8.1 Open-ended responses from the D4 workshop evaluation survey

Comments on what did or did not meet your expectations are most welcome:
- more time for hacking - flexibility in teams. Most teams seemed pre-formed and existing collaborations. - discussion sessions were interesting but overall too long and did not introduce new insights. Some felt like therapy sessions for researchers. Perhaps a single discussion session near the end with participant-submitted questions would've been more productive. - Instead of discussion sessions, interactive breakout tutorials by organizers or participants would've been more useful. - some "hacks" felt like a pre-existing collaboration meeting rather than a hack/sprint using a new dataset or new methodology.
Expected more discussion on disparity and decision-making.
Hospitality and facilities are especially excellent. Got less accomplished than hoped but being in person very much strengthened the collaboration and confidence in ability to complete the research.
I know the workshop was constrained by time, given that it is only Monday-Thursday, but I think the workshop would have benefited from both more time to network and maybe a bit more time to work on our projects. Even if it was one more day, that would have helped with adding more to the "speed dating" event and time for more project development. I think there would be more room for more interactive trainings, especially taking into account people's expertise and competency with different methodological approaches. Maybe hosting small group trainings would help make sure that people are getting most from learning these approaches.
I was truly impressed with how successfully the hackweek brought together a diverse group of scholars into a research-driven environment. I have had several bad experiences with more discipline- or domain-based workshops, but this interdisciplinary or convergence space sparked quite a bit of collaboration. Ultimately, I am more energized to ask and answer questions that require data integration than I was prior to the event.
It would have been helpful to have a lot more time for us to actually work with each other. Ultimately, we ended up with only 1.5 days to do all the actual work. I think having the event be the full week would help, in addition to having fewer sessions and more work time.
Lots of excitement to learn more about AI/ML. Excellent networking opportunities.
Masha was an amazing asset!
Overall good, needed more focus on the decisions and disparities elements
Team DEMUS already has most of our tech needs met because of the conditions working in the RDC. While interesting to learn about, our main goals were advancing our own project and networking. I truly enjoyed networking and especially connecting with researchers who may join the DEMUS team (Jess) or another project. The organizers set a great collaborative tone for the meeting, most panels were engaging, and the meeting spaces were great for providing collective and break out spaces. In the end, it struck me that the emphasis on statistical analysis and computing was not well balanced with the

importance of research questions, theory, qualitative data, as well as research ethics and applications. That's okay, but sometimes it felt as teams were exploring their data without a clear sense of direction or goals.
The tone and organization of the event was great, especially for providing focused time and an in-person setting to push forward on our team's research.
the training talks were too superficial to learn much; i would have preferred more speed-dating and workshopping our projects and actually working on projects
This has been a great experience! It has by far exceeded a lot of the workshops I have participated in the past.

How did you consider vulnerable populations in your D4 work this week?	**What insights might NOAA take away from your approach to better serve these communities?**
Conceptual modeling and use of SVI data	Co-production is key!
we are including them in our analysis	more hazard data that can be easily connected with population data (floods, fire, etc.)
Using ACS data to examine the representativeness of our data.	Work with demographers! NOAA should incentivize interdisciplinary work that includes demographers to help ensure demographic data is used properly. Like all datasets, demographic data has A LOT of nuance and can lead you to different conclusions depending on choices like aggregation level, data source, etc.
Focus of our research project, and discussion point re developing our research team composition	Emphasize deep engagement with social dimension of the social and ecological interface (as opposed to superficial or "tacked on" treatment of social dimension). Invest in development of accessible social data.
Consideration of many aspects of vulnerability in the population (focus of analysis).	Vulnerability is complex. Should explore approaches that do not necessarily rely on SOVI.
We tried to think about the underlying systems that lead people to be vulnerable more than descriptions of these populations. We often resort to demographics and proxies, but focusing on systems can help to develop policy solutions that lead to reductions in vulnerability.	Understand the underlying systems and try to incorporate them into hazard risk!
I grounded my theoretical approach based on a c critical concern based on vulnerability	I would encourage first a practice of considering how many of these communities' perceptions are inherently foreign to my privileged perspective
Ethics considerations	Focus approach to better serve underrepresented communities.

in our work we consider the effectiveness of flood adaptation policy considering community characteristics that include income, race, educational attainment, housing status (renter vs owner)	Our approach has potential to be used to answer other causal inference questions in the same realm helping make decisions that consider a climate justice angle
We specifically examined exposure by demographics/vulnerabilities.	TBD
N/A	Thinking about vulnerability as more than a quantitative or countable variable.
census variables, other indices, disparities in exposure / effect	identifying disparities in exposure and outcome and information and resources
We studied the demographic and socioeconomic profile of High-risk and Medium-risk wildfire areas.	Need for better data to understand NOAA. Particularly, event analysis or measures in county or subcounty level measures of actual exposure. Maybe as a processed dataset?
Quantified the problems in representativeness of the data with respect to socioeconomic status	Highly granular spatial demographic data will be helpful to study risk and disparities. Existing summary indexes of risk are weakly specified and weakly validated.
Theory;?Indicators of social vulnerability in modeling	Ideally, use it to better support vulnerable communities
We are concerned about the impact of wildfire disasters and FEMA assistance on marginalized populations, including by race/ethnicity and wealth. I appreciated the question from one participant about looking closely at fires that affect those on indigenous land.	
I was pleased to learn that EJ was being considered by many teams, but I would like to hear about how statistical methods could be applied to improve statistical power for small subgroups in the population. For example, what can be learned about small N race and ethnic groups that are dispersed across geographic areas unequally and therefore difficult to study with quantitative methods. Is there a role for differential privacy, bootstrapping, or something else?	I would be happy to share our migration system's approach with NOAA. Our focus on coastal counties and their migration systems are helpful for considering where residents of those counties might go in an evacuation or planned relocation.
We focused the analysis on demographics associated with vulnerability in the literature	Honestly I don't think this is the best approach to understand vulnerability, at least not by itself. I would encourage interviews in tandem to better understand vulnerability in one specific location/context.

We looked at exposure to disasters by demographic characteristics	It should be kept in mind during study design and data collection how data might be integrated with social science data such as from the Census to better answer questions and serve vulnerable communities.

What was the most promising or exciting method or tool you learned about this week for integrating demographic or other social data with environmental data, if any?
AI/ML implementation
Certainly, the aspect of AI role in helping make more efficient use of time for analysis is the most promising tool.
Demographic data available at RDC
I had never worked with a supercomputer before, which is very helpful for large datasets
It was interesting to compare demographic methods with existing demographic methods I know
It was interesting to think about different demographic datasets that we heard about and the ways in which people are matching things that are spatiotemporally different.
movement data in census, bayesian methods
Neural Network Analysis seems promising for our project (thanks Jason!).
Neural networks
Neural networks
Neutral ml blows my mind
Not tool so much as considerations -- need to consider the metadata/story of all the datasets you're using. Need to pay attention to unit of analysis.
Possibilities of machine learning and Bayesian approaches
Seeing other research projects and their myriad approaches was very informative. More so that AI sessions for data integration, IMHO
The PUMA-county data integration method
This wasn’t where my focus was.
Use of AI
XGboost

Please share any specific constructive comments you have on pre-meeting communication, logistics and support, the JupyterBook, or the tutorials:

It was a lot and set expectations quite high. Nonetheless incredibly useful.
I don't use python and really struggled with the tutorials. I wish the tutorials were in R! I also would have appreciated more focus on the conceptual motivation for certain approaches
I think these should have been developed further in advance and have a meeting to discuss them there. I feel not everyone looked at them.
Love JupyterBook - nice organizational feature.
JupyterBook and links to content and other platforms including slack were really helpful. Well done!
I think it would be helpful to host virtual meetings to help people get set up with tools that they do not use regularly in their work (e.g., Jupyter Notebook, Python).
The tutorial on AI for Data Integration was a wonderful recap for those that already know these AI tools. I would have opted for something a bit less mathy and more focused on explaining the overarching functions and advantages of certain AI approaches to data integration, as well as cautions for those using them for the first time
There was a lot of great information but there were also a lot of links and websites and sources of that information which made it a bit hard to follow. I really appreciate everything being so thorough but i think there was a little bit of information overload. If we could streamline to have things be a bit more concise and consolidated to one site, I think that would be helpful! I didn't have a chance to do all tutorials, but they seem like they are useful!
The JupyterBook was a great resource that was clearly made with end users in mind. I really enjoyed its presence during the hackweek.
The hands-on ML tutorial could have been more inclusive and hands on. There should've been time dedicated to participants actually setting up the jupyter notebooks and the instructor should've allowed the participants to run the notebook on the fly.
Appreciated that there were many channels of communication (email, slack, jupyter, drive, etc). It was easy to find info when I needed it.
I think the AI/ML tutorials are a useful resource, but I don't think going over them as a group was helpful. Maybe some time to go through them as a small group with some floater help would be more useful so we can tailor them to our own project.
It was a bit overwhelming to adopt all those tools. We already have a system for keeping records of our progress and it felt odd to take up a new system that would be public.
It was unclear which tutorials to prioritize / which would be really useful and I ran into issues with linking Github to the Jupyter workspace which went unresolved (not needed on my end for the project)

Please share any specific constructive comments you have on the Discussion panel sessions:
Some were more useful than others. No finger pointing.
The discussion was wonderful! It was not all relevant, but it was very insightful to me.
Uncertainty!! Data integration!!

Very thought provoking in a bunch of ways (storms, different data and methods, uncertainty).
It was great to have a conversation while in the early stages of a project, helping focus the direction of our research.
I found the AI sessions to be a bit buzzword heavy but not very actionable. The panel about uncertainty was helpful and generally I think it would have been more helpful to have more work time than more panels!
The panels were excellent topics that produced really good questions and feedback. I thought they were a high point during the week.
- single discussion session near the end with prepared questions submitted by the participants and organizers beforehand.
helpful to have early reactions to work with new ideas
I enjoyed the integration panel with DJ and Jess. Both super smart and Jess was really helpful thinking about issues that come up in integration.
I felt that the AI/ML one should have been created with the audience in mind -- it seemed many in the room had not used ML techniques before and it would've been useful to have more of a primer (Jorge's slide but more detailed).
We wanted more time to meet with our team and work earlier in the week

Please share any specific constructive comments you have on the team speed dating:
3 sessions was too much for 1 day. Would have liked to have met with every team over the week.
It was really nice to meet the other teams, but less trouble shooting happened for our project
It was extremely helpful to get a sense of what each team was up to and where there were potential spaces for synthesis or collaboration.
I did not think these would be as productive as they were.
As a floater, I was able to learn more about what the floaters' expertise were!
More time.
Speed dating was so much fun and really help to make us feel connected to the work of the other groups.
I wish we had time to meet with all the teams! I think it would also be nice to have some helpful sources for the speed dating, such as guiding questions and how to best use that time.
The team speed dating was excellent! It was great to meet people who are doing similar work right off the bat and have a full 30 minutes to do so.
The team speed dating was another great aspect of the week, but I would recommend expanding the time given moving forward.
Rather than a team-on-team speed dating, the hackathon could've started with a brief "pitch meeting" by all of the teams. Then participants could join or approach whichever team of interest.
this was my favorite part and felt the most productive for both the science and building relationships
Make them longer?

It was great to meet the other teams in small groups - 100% recommend doing that instead of large group discussions.
Really it was just fun to meet and hear about other projects.
I think it was good! Some guiding questions would've been helpful.
We wish we could have met all the teams!

Please share any specific constructive comments you have on the research infrastructure or support, including Floaters:
Both of these elements were exceptional!
I used my own university resources to work. The floaters were really nice to talk with and provided interesting insights.
I was a part of setting it up and a floater. I did not use them myself.
Patrick, Jason, and RDC were instrumental for our team. :)
I only wish we came more prepared (and we came with an already linked dataset) in order to make progress to make more use of the floaters.
Jason rocked it
It was great to have so many resources available! I think it is much better for teams to make their own repositories so that they have ownership over them moving forward, though. That is essential for the future of these projects!
As a floater, I enjoyed the ability to move between teams and provide general and specific feedback without being connected to any one problem.
the infrastructure could've been useful if, e.g., the AI tutorial was interactive. Floaters were very helpful overall. Future hackathons should have more participants with diverse expertise (ML, climate modeling, migration modeling) that are not pre-designated teams.
Floaters are an amazing resource. I would recommend making the Floater part of teams.
We couldn't use the GitHub site off the bat because of the licensing that it would put our code and data under. We came to the event without plans to use NCAR cluster but have been thinking about introducing a component to the project to try it.
Floaters assistance a was critical in helping us implementing our model
I appreciate the floaters' expertise, but I think we weren't far enough along in our project to make much use of them.
Jason was helpful although we kind of iterated on the value of tools for analyzing observed data versus projections. More familiarity with formal demography might help AI/ML folks understand what benefits of those approaches are visa vis what we already do.
Floaters were great!
This was awesome to have!!

Considering the relative allocation of time to different activities (including both work and social activities) during this workshop, for future workshops like this one: Which elements should receive more time, if any? Why? Which should receive less time, if any? Why?
There should be more time to work!
Another day of team work would have been helpful.
I think the time balance was overall good, but more breaks would have been good. Tuesday morning felt really long.
The allocation of time was quite good in my opinion. It might be beneficial to dedicate slightly more time to working and a little less time to presentation of the work (e.g. 15 to 20 minutes per presentation)
I would have liked to see more time devoted on the first day to introducing the different groups and their rough project ideas (in addition to speed dating exercise). The panels could have been a bit shorter - maybe an hour instead of 1.5 hours.
More work time +1 or +2 days. No less.
Seems like we need one more day?
A little time to start at beginning for groups to get work underway. (Saving some of the panels for later days or afternoon sessions, as a break from longer blocks of group work.)
More time for networking and working on projects, less time (or more constructive time) for trainings. The panel discussions were great and I think the time allotted for them was perfect.
We should get another day
Thursday was very pack, maybe having allocated at least one of the sessions on Wednesday.
more team speed dating time would be helpful, more time to work by teams since day 1, as well as more time with the floaters if needed
I think there needs to be significantly more time spent on the actual work for the hack week. I think this should have been 5 full days with at least 4 days' worth of time dedicated to doing the actual work. While workshops and sessions are helpful, the point of this was to talk with each other across teams and work on our projects within our team. And I think having more space for both of those things would have been a lot more helpful. That means spending a lot less time on sessions like workshops and panels. Regarding social time, I think having more unstructured social time could be nice but the happy hour and dinner are both great!
I would suggest increasing the time spent during speed dating and for project research and decreasing the time spent presenting and providing group updates.
more time dedicated to hacking. In hackathons I found more productive in the past the clear goal was to use new data, new methods, or just a new project to hack and produce concrete preliminary results that lead to papers. There was not enough time to produce concrete results for most teams.
less time for talks more time for work more time for connecting in small groups
Less time to Data Sciences ML/AI aspects. It is a great thing, however, sometimes the language and conversation was difficult to follow. If kept, please simplify the language.

More time in networking and work.
One more day would be really useful. The panel discussions were very interesting, but they could be potentially spread out more across the week
A little more time for group work on our project would have been nice.
More team work time would be great. More time for floaters. More data visualization instruction.
More - work time Less - discussions
more time for team speed dating, team work sessions especially earlier in the week less time for panels / talks (or spread them out more rather than most of Monday & Tuesday being panels/talks)

We welcome any additional comments or suggestions you have, especially guidance to NOAA for their societal data insights initiative.
Appreciated the interdisciplinary balance and seniority mix and size of the event (number of participants). Felt that one more working day might have enabled more concrete progress but given schedule constraints and start of semester, that could also have been a hardship.
Echoing, more inputs from those with boots on the ground.
Maybe allow NOAA guy, who is very sweet and a nice person, to talk a little bit about their priority areas and what is NOAA looking for.
One more day would be great!!!
thank you for a great week!
Thanks for making this necessary space possible.
Thanks for this awesome opportunity!
Thanks! This was a really positive experience
There are a lot of teams with similar challenges and aggregating these challenges into a whitepaper can be beneficial.
This was a great event and one that has reinforced my positive view of interdisciplinary or convergence science. Thanks to the organizing committee!
Weather impacts humans through its impacts on the built and natural environment. There is a shortage of data on what’s on the ground and how it exacerbates or mitigates hazard impacts. A focus on these modifications or adaptations it’s important for investigating population impacts.
Would love to have had the NOAA team come meet all the groups, as if on a "listening tour".

8.2 Agenda

Prior to D4 Hack Week				
When	**What**	**Details**	**Location**	**Remote option**
1-2 weeks before	Participate in asynchronous tutorials	We will provide tutorials on methods and accessing datasets that you are welcome to complete prior to the hack week.	See the D4 JupyterBook.	Fully remote
1-2 weeks before	Project scoping	Teams revisit their goals for the D4 hack week and scope out what they hope to achieve during the week.	See the D4 JupyterBook.	Fully remote
Monday, September 9th				
1:00-1:30 PM	Welcome & catered lunch	Please arrive at the UW eScience Institute by 1PM if possible.	eScience Institute - Common area	In person
1:30-2:30 PM	Team planning & floaters huddle	Teams revisit their goals for the week. Floaters connect on how to best support teams.	Physics 520: DEMUS, Claim to Flame, Disaster Demography eScience Seminar room: Floodsters, Migration Mavericks eScience Common area: Floaters, CAT eScience Meeting room: Columbia Hurricane Migration, FSU & Penn State	In person (teams welcome to incorporate remote members)
2:30-2:40	Break			
2:40-3:45 PM	Discussion: AI for Data Integration	Panel (panelists Amy McGovern, Jorge Celis, David John Gagne, June Yang) and plenary discussion of AI tools for data integration, facilitated by Ann Bostrom. AI/ML can be used to help overcome many of the challenges that are posed by the integration of social, environmental, and other data (e.g., mismatched temporal or spatial	Physics 520	Hybrid

		scales). The panel will queue up a discussion about the state of the art regarding AI methods for data integration, and opportunities for methodological advancements using AI.		
3:45-5:15 PM	Team speed dating	Teams connect with other teams during three ~half-hour sessions: 3:45 Intro to session – Sara Curran **3:50 to 4:15** (1) DEMUS with Columbia Hurricane Migration (2) FSU & Penn state with Claim to Flame (3) The Floodsters with Disaster Demography (4) Migration Mavericks with CAT **4:20 to 4:45** (2) DEMUS with FSU & Penn State (3) Columbia Hurricane Migration with Disaster Demography (4) Claim to Flame with Migration Mavericks (1) The Floodsters with CAT **4:50 to 5:15** (1) DEMUS with CAT (2) Columbia Hurricane Migration with FSU & Penn state (3) The Floodsters with Migration Mavericks (4) Claim to Flame with Disaster Demography	(1) Physics 520 (2) eScience Institute – Seminar Room (3) eScience Institute - Common area or outside (4) eScience Institute - Meeting Room	In person
5:15-6:00 PM	Hosted Happy Hour	Big Time Brewery, 4133 University Way NE	Big Time Brewery	In person
6:00 PM Adjourn	Dinner and evening on your own	The evening is free for teams to use as they see fit. You can see a list of recommended restaurants and rules about reimbursement for meals on your own on the JupyterBook – logistics page.	Up to you	

Tuesday, September 10th				
8:30-9:00 AM	Coffee Connect with breakfast / Standup	Teams provide 2-minute informal updates on their progress so far and goals for the day over coffee and breakfast (provided).	eScience Institute - Common area	In person
9:00-9:30 AM	CAT Demonstration project	The Collaborating Across Threads (CAT) team will share their methods and preliminary results from their analyses.	Physics 520	Hybrid
9:30-9:40 AM	Break			
9:40-10:45 AM	Hands-on AI for data integration exercise	Tutorial-associated exercise led by Jason Stock	Physics 520	Hybrid
10:45-11:00 AM	Break			
11:00 AM-12:30 PM	Discussion: Data integration	Panel (panelists Matt Dunbar, Tyler Fricker, and Andrea Schumacher) and plenary discussion facilitated by Sameer Shah. A key goal of this hack week is to generate data products that integrate social and weather and/or climate data across space and time, through interdisciplinary collaborations, and develop methods for such data integration. This panel will highlight and kick off a discussion of key data integration challenges and approaches to addressing them.	Physics 520	Hybrid
12:30-1:30 PM	Catered lunch			In person
1:30-5:00 PM	Team work time		See room assignments at end of this document	In person (teams welcome to incorporate remote members)
6:00PM Adjourn	Dinner and evening on your own	The evening is free for teams to use as they see fit. Restaurant suggestions available on the D4 JupyterBook – logistics page.	Up to you	

Wednesday, September 11th				
8:30-9:30 AM	Coffee Connect with breakfast / Standup	Teams and floaters provide 5-minute updates on their progress so far and goals for the day over coffee and breakfast	eScience Institute - Common area	In person
9:30 AM-5:00 PM	Team work time (whole day after Standup)	Wednesday will be a dedicated team work day. Lunch and dinner will be on your own.	See room assignments at end of this document	In person (teams welcome to incorporate remote members)
9:30-10:30 AM	Floater drop-in hours	Floaters will be available for drop-in questions; teams can sign up for floater support during other times (see D4 Jupyter book - floaters)	eScience Institute - Common area	In person
1:30-2:30 PM	Floater drop-in hours	Floaters will be available for drop-in questions; teams can sign up for floater support during other times. (see D4 JupyterBook – floaters)	eScience Institute - Common area	In person
Evening	Dinner on your own	Restaurant recommendations and reimbursement rules for meals on your own are available in the D4 JupyterBook – logistics page. The rest of the evening is free for teams to use as they see fit.	Up to you	
Thursday, September 12th				
8:30-9:00 AM	Coffee Connect with breakfast	We will open the space and provide coffee and breakfast in the mornings for those who would like to join early.	eScience Institute - Common area	In person
9:00-10:30 AM	Presentations	Three teams will have 30 minutes each to present (informally) their methods, challenges, and findings, including time for questions/discussion.	Physics 520	Hybrid
10:30-10:45 AM	Break			
10:45-11:00 AM	CAT Update	Update from the Collaborating Across Threads (CAT) team.	Physics 520	Hybrid
11:00-12:30 PM	Discussion: Uncertainty	Panel (panelists David John Gagne, Julie Demuth, Jessica Godwin, facilitator Sara Curran) and plenary discussion of how to analyze and	Physics 520	Hybrid

		represent different kinds of uncertainties that surface with integration of diverse datasets.		
12:30 -1:30 PM	Catered lunch	Catered lunch will be provided		In person
1:30-2:30 PM	Presentations	Two teams will have 30 minutes each to present (informally) their methods, challenges, and findings, including time for questions/discussion.	Physics 520	Hybrid
2:30-2:45 PM	Break			
2:45-3:00 PM	Evaluation survey	Fill out evaluation survey	Physics 520	In person
3:00-4:00 PM	Presentations	Two teams will have 30 minutes each to present their methods, challenges, and findings, including time for questions/discussion.	Physics 520	Hybrid
4:00-4:45 PM	Closing discussion facilitated by Jon Mote	Floaters kickoff, followed by open discussion around observations and key takeaways from the hack week	Physics 520	In person
6:00-8:00 PM	Community Dinner at Mamma Melina Ristorante & Pizzeria	All D4 Hack Week participants are invited to dinner at Mamma Melina, 5101 25th Ave NE, Seattle, WA 98105.	Mamma Melina	In person
Friday, September 13th				
10:00 AM - 12:00 PM	Discussion: Debrief & next steps (optional for participants)	Hack Week organizers, as well as any participants who would like to join, will debrief the D4 Hack Week and discuss next steps for producing outputs and outcomes.	eScience Meeting Room	Hybrid

LOCATIONS: *A searchable UW Campus map is available here.*

- ***The eScience Institute*** *(including the eScience Seminar room, eScience Meeting room, and eScience Common area) is located on the 6th Floor of the Physics / Astronomy Tower, UW Seattle.*
- ***Physics 520*** *is on the 5th floor of the Physics / Astronomy Tower, which is located on the North-East corner of the intersection of 15th Ave NE and NE Pacific Street*
- *Winkenwerder Forest Sciences Laboratory (room 021) and Bloedel Hall (room 292) are just south of the Drumheller fountain on the UW Seattle campus.*
- *Raitt Hall (rooms 221, 223 & 229) is located north of Drumheller Fountain on The Quad.*
- *The University Inn is located at 4140 Roosevelt Way NE, Seattle, WA 98105. The Inn is within easy walking distance of the Physics/Astronomy Tower. The phone number of the inn is (206) 632-5055 or (866) 866-7977. Check in is at 3 PM and check out is at noon (https://phys.washington.edu/visitor-information).*

TEAMS: Details available in the D4 Jupyter book

Team Name	Project Leads and Members	Room for Team work on Tuesday & Wednesday
Claim to Flame	Joan Casey, Lauren Wilner, Vivian Do, Heather McBrien, David Coomes, and Karen Chen	Physics 520
DEMUS (Demography of Environmental Migration in the United States)	Elizabeth Fussell, Jack DeWaard, Katherine Curtis, James Done, Katie McConnell and Sara Ronnkvist	eScience Meeting room
Columbia Hurricane Migration	Fabien Cottier, Mona Hemmati, Andrew Kruczkiewicz, and Kytt MacManus	Raitt 223
Disaster Demography	Deborah Balk, Dylan Connor, Melanie Gall, Lori Hunter, Jenna Tipaldo, and Helen Wilson Burns	Bloedel 292
The Floodsters	ChangHoon (Chang) Hahn, Sharif Islam, and Lidia Cano	eScience Seminar room
FSU & Penn State	Mathew Hauer, Alexis Santos, and Sunshine Jacobs	Winkenwerder 021
Migration Mavericks	Ethan Sharygin, Mary Angelica Painter, and Justin Stoler	Raitt 221
Floaters	Tyler Fricker, Patrick Greiner, and Jason Stock	
CAT Demonstration Project	Andrea Schumacher, Julie Demuth, DJ Gagne, Jorge Celis, Amy McGovern, Sara Curran, Sameer Shah, Masha Vernik, Ann Bostrom	eScience Common area

Tuesday backup room: Raitt 229

8.3 Organizing Committee

Ann Bostrom – University of Washington Evans School of Public Policy & Governance and the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)
Sara Curran – University of Washington Center for Studies in Demography & Ecology
Sameer Shah – University of Washington School of Environmental and Forest Sciences
David John Gagne II – NSF National Center for Atmospheric Research and the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)
Amy McGovern – University of Oklahoma and the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography
John Williams – The Weather Company; NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)
Andrea Schumacher – NSF National Center for Atmospheric Research
Julie Demuth – NSF National Center for Atmospheric Research and the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)
Jonathon Mote – National Oceanic and Atmospheric Administration
Adam Tasca – National Oceanic and Atmospheric Administration
Masha Vernik – University of Washington School of Environmental and Forest Sciences
Courtney Allen – University of Washington Center for Studies in Demography & Ecology
Jorge Celis – University of Oklahoma and the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)

8.4 Glossary and Acronyms

D4

Stands for Disasters, Demography, Disparities, and Decisions.

Disasters

Serious disruptions to the functioning of a community that exceed its capacity to cope using its own resources. Disasters can be caused by natural, man-made and technological hazards, as well as various factors that influence the exposure and vulnerability of a community.

Git

A popular version control system that is used in many open source software projects to manage their software code base.

GitHub

A service platform that allows developers to create, store, manage and share their code using the git command.

GitHub Actions

Continuous integration and continuous delivery (CI/CD) GitHub feature that allows you to automate computational workflows for a GitHub repository.

GitHub Pages

GitHub feature that allows you to host a website connected to a repository or organization

Hackweek (or hack week)

Participant-driven events that strive to create welcoming spaces to learn new things, build community and gain hands-on experience with collaboration and team science.

Project Jupyter

Project Jupyter (name derived from “JUlia PYThon and R”) exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.

Jupyter Book

Jupyter Book is an open source project for building beautiful, publication-quality books and documents from computational material.

JupyterHub

JupyterHub allows you to deploy an application that provides remote data science environments (typically Jupyter Lab) to multiple users. It can be deployed with a cloud service provider, or on your own hardware.

JupyterLab

JupyterLab is the next-generation web-based user interface for Project Jupyter intended to replace the JupyterNotebook interface.

Jupyter Notebook

Open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

Machine Learning Models

A machine learning model is a program that can find patterns or make decisions from a previously unseen dataset.

AI

Artificial intelligence, or AI, is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities.

Acronyms:

D4 - Disasters, Demography, Disparities, and Decisions.

CSDE - Center for studies in Demography and Ecology.

NOAA - National Oceanic and Atmospheric Administration

AI2ES - NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography

SHELDUS - county-level hazard data set for the U.S. and covers natural hazards such thunderstorms, hurricanes, floods, wildfires, and tornadoes as well as perils such as flash floods, heavy rainfall, etc.

FEMA - Federal Emergency Management Agency