

# SUPPLEMENTARY MATERIALS FOR

# OceanHackWeek

## AN INCLUSIVE, COLLABORATIVE APPROACH TO DEVELOPING OCEANOGRAPHY DATA SCIENCE SKILLS

By Catherine Mitchell, Wu-Jung Lee, Filipe Fernandes, Joseph Gum, Alex Kerney, Emilio Mayorga, Thomas Moore, Nick Mortimer, Natalia Ribeiro, and Valentina Staneva.

**TABLE S1.** OceanHackWeek in numbers. Participants self-identify in the application materials as considering themselves a minority in their research field in terms of ethnic/racial identity and gender identity. n/a indicates data are not available.

	2018	2019	2020	2021	2022	2023	2024
<b>Modality</b>	In-person	In-person	Virtual	Hybrid	Hybrid	Hybrid	In-person
<b>Location(s) for in-person</b>	USA	USA	–	USA	3 USA locations, Brazil, Peru, Uruguay, Australia	USA, Australia	USA, Australia
<b>Time zone(s) for virtual</b>	–	–	UTC-7 (USA PDT)	UTC-7, UTC+10 (Australia EST)	UTC-7 (2 events: one English, one Spanish)	UTC-7	–
<b>Number of participants</b>	52	54	46	70	161	61	41
<b>% Gender minority</b>	31	37	64	58	44	n/a	35
<b>% Ethnic/racial minority</b>	14	26	36	45	37	n/a	18
<b>% Non-US institutions</b>	8	20	52	55	45	n/a	37
<b>Programming languages</b>	Python	Python	Python + R	Python + R	Python + R	Python + R	Python + R
<b>Number of tutorials</b>	13	8	10	7	21	8	9
<b>Number of projects</b>	11	11	8	11	17	11	11

TABLE S2. OceanHackWeek Tutorials

TUTORIAL	INSTRUCTOR(S)	YEAR	DOCUMENTS	VIDEO
<b>DATA ACCESS</b>				
Integrated Ocean Observing System (IOOS) data access	Emilio Mayorga	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Ocean Observing Institute (OOI) data access	Friedrich Knuth	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Access and clean-up ocean observation data	Emilio Mayorga	2019		<a href="#">YouTube</a>
Searching and obtaining data from multiple servers	Filipe Fernandes Alison Gray Ben Tupper	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Satellite data from the cloud (Python)	Chelle Gentemann	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
OBIS and raster data analysis	Camille Ross Nick Record Ben Tupper	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data access	Filipe Fernandes	2021	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data access in Python	Filipe Fernandes	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data access in R	Johnathan Evanilla	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Cloud-based data access and exploration	Don Setiawan	2022		<a href="#">YouTube</a>
Extracting OBIS data by Large Marine Ecosystem (LME) regions in R — <i>In Spanish</i>	Enrique Montes	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data access methods	Marty Hidas	2023	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data access in Python	Callum Rollo	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>
<b>DATA VISUALIZATION</b>				
Cutting-edge data visualization tools	Jake VanderPlas	2018		<a href="#">YouTube</a>
Data visualization part 1: Basics and geospatial visualization	Filipe Fernandes	2019	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Advanced data visualization	Filipe Fernandes Amanda Tan	2019	<a href="#">GitHub</a>	<a href="#">YouTube</a>
HoloViz	Jim Bednar	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data visualization in R	Maria Pachiadaki	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data visualization in Python	Aimee Barciauskas	2021	<a href="#">GitHub</a>	<a href="#">YouTube</a>
R oce package	Clark Richards	2021	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data visualization in R	Ben Tupper	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data visualization in Python	Derya Gumustel	2022	<a href="#">GitHub</a>	
Visualizing tidy spatial data in R — <i>In Spanish</i>	Paola Corrales Elio Campitelli	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
<b>PYTHON TOOLS</b>				
Geospatial and mapping tools	Filipe Fernandes	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Xarray	Joe Hamman	2018		<a href="#">YouTube</a>
Xarray	Deepak Cherian	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Jupyter, NumPy, Pandas, and Matplotlib	Leticia Portella	2020		<a href="#">YouTube</a>
Xarray	Julius Busecke	2021	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Xarray	Deepak Cherian	2022		<a href="#">YouTube</a>
Dask	Deepak Cherian	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Xarray	Danilo Silva	2023	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Xarray	Richard Signell	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>

Table continues on next page...

TUTORIAL	INSTRUCTOR(S)	YEAR	DOCUMENTS	VIDEO
<b>R TOOLS</b>				
Analysis of spatial networks with sfnetworks — <i>In Spanish</i>	Lorena Abad	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
<b>MACHINE LEARNING</b>				
Machine learning	Valentina Staneva Joseph Gum	2019	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Machine learning overview	Valentina Staneva	2020	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Neural networks in R	Johnathan Evanilla	2021	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Machine learning introduction	Valentina Staneva	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Machine learning applications	Filipe Fernandes	2022	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Machine learning with Python	Filipe Fernandes	2023	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Introduction to transformers	Jiarui Yu	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>
<b>DATA ANALYSIS WORKFLOWS</b>				
Synoptic validation	Rob Fatland	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Data mining	Valentina Staneva	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
AI-assisted programming with Python and ChatGPT	Myranda Shirk	2023	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Ocean 'omics data	Julia Brown	2023	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Sofar Spotter comparison to model data	Ciara Dorsay	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>
AI assisted programming	Myranda Shirk	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Species distribution modeling in R	Camille Ross	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>
<b>WORKING WITH BIG DATA</b>				
Cloud computing	Amanda Tan	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Handle “big” larger-than-memory data	Joe Hamman Ryan Abernathey	2019	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Cloud computing 101	Amanda Tan	2019	<a href="#">GitHub</a>	<a href="#">YouTube</a>
<b>GIT AND GITHUB</b>				
Git/Github	Valentina Staneva	2018		<a href="#">YouTube</a>
Git, Github, and project collaboration	Valentina Staneva Tim Crone	2019	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Git/Github	Filipe Fernandes	2020		<a href="#">YouTube</a>
GitHub workflow	Wu-Jung Lee	2022		
Version control with Git and GitHub — <i>In Spanish</i>	Linda Cabrera Denisse Fierro-Arco	2022		<a href="#">YouTube</a>
Git and Github	Derya Gumustel	2023		
GitHub primer	Derya Gumustel	2024	<ul style="list-style-type: none"> <li>• <a href="#">About Git and Github</a></li> <li>• <a href="#">Git Setup and Basics</a></li> </ul>	<a href="#">YouTube</a>

Table continues on next page...

TUTORIAL	INSTRUCTOR(S)	YEAR	DOCUMENTS	VIDEO
<b>REPRODUCIBLE RESEARCH &amp; TOOLS</b>				
Data sharing and collaboration tools	Amanda Tan Don Setiawan	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Reproducible research and tools	Filipe Fernandes	2018	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Reproducible science	Joseph Gum	2019	<a href="#">GitHub</a>	
Reproducible research	Joseph Gum	2020		<a href="#">YouTube</a>
Reproducible science and data management	Mathew Biddle	2021	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Managing conda environments	Filipe Fernandes	2022	<a href="#">GitHub</a>	
JupyterHub	Wu-Jung Lee	2022		
JupyterHub — <i>In Spanish</i>	Emilio Mayorga	2022		
Best practices for reproducible research	Nick Record	2022		<a href="#">YouTube</a>
Reproducible research	Joseph Gum	2023	<a href="#">GitHub</a>	<a href="#">YouTube</a>
Conda and environments	Filipe Fernandes	2023		<a href="#">YouTube</a>
JupyterHub	Emilio Mayorga	2023		<a href="#">YouTube</a>
Reproducible research and project management	Eli Holmes	2024	<a href="#">Slideshow</a>	<a href="#">YouTube</a>
JupyterHub and Pixi environments	Christian Sanderson	2024	<a href="#">GitHub</a>	<a href="#">YouTube</a>

## OceanHackWeek Participant Testimonials

“I had no intention to pitch a project and planned on joining another team... During the project pitching session, I felt encouraged by the inviting and friendly atmosphere and decided to mention a project that I have been trying to start for a few months... The OceanHackWeek (OHW) participants I interacted with were super encouraging and by the end of the day I was leading a team of five to develop a Python package!... The positive atmosphere created by the OHW organizers was contagious: everyone was supportive, kind, and encouraging.”

– Kyla Drushka, 2021, USA, virtual participant

“OceanHackWeek (OHW) has an important message, especially to young scientists: ‘It is impossible to do science alone.’ Therefore, we were very encouraged to work in groups... As each participant had a different background, everyone had something to teach and learn, which provided a democratic and inclusive environment.”

– Mauricio Reboucas Rocha, 2021, Brazil, virtual participant

“I also enjoyed the community aspect of the hackathon. The organisers made a real effort to include participants, from suggesting additions to the code of conduct on Monday through to presenting our hackathon projects at the end of the week”

– Callum Rollo, 2020, UK, virtual participant

“My experience at the OceanHackWeek 2019 was... a once in a lifetime learning experience. The people involved were amazing, from the organization group to all the participants. The project group I joined was extremely democratic and involved. I learned not only from more advanced researchers but also from telling what I knew to my group partners.”

– Ágata Piffer Braga, Brazil, in-person participant

## OceanHackWeek Alumni Stories

“I dove into learning Python in late 2018, eager but overwhelmed by the vast array of libraries available for analyzing Earth data. Attending OceanHackWeek (OHW) in 2019 was a game-changer—it didn’t just help me learn Python, it transformed the way I approach problem-solving. By 2020 and 2021, I was thrilled to step into an organizing role, helping others discover the same magic I’d experienced.

“As a hands-on learner, OHW was perfect for me. I could tackle real-world challenges alongside experts and peers, collaboratively building solutions while learning in a supportive, dynamic environment. What once felt frustrating suddenly became fun! Spending an entire week immersed in shared learning—absorbing knowledge from others and contributing what I knew—propelled my research in ways I hadn’t imagined. And the best part? I gained an incredible community I could turn to whenever I hit roadblocks.

“This transformative experience was made possible by OHW’s commitment to fostering an inclusive and open atmosphere, where everyone felt encouraged to learn, share, and grow. It wasn’t just a workshop—it was a turning point that continues to inspire me.”

– *Chelle Gentemann, NASA SMD Office of the Chief Science Data Officer, Open Science Program Scientist*

“Attending OceanHackWeek (OHW) in summer 2021 has had a lasting impact on my career as an ocean scientist. It strengthened my coding skills, expanded my professional network, and fostered collaborations that have continued well beyond the event. In the months following OHW, my project group continued to meet and develop our python-based product. Because our project aligned with many of our ongoing research interests, our group continued to exchange feedback, discuss results, and support one another through conference presentations. In the years since OHW, the confidence I gained has empowered me to teach ocean-focused coding courses at my home institution. I’ve also reconnected with OHW alumni at conferences and even co-organized a town hall with a colleague I met during the event.”

– *Jackie Vetch, Postdoctoral Fellow, Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution*

## OceanHackWeek Organizational Structure

OceanHackWeek has two organizational arms: the Steering Committee and the Organizing Committee. Both committees are composed of ocean-focused individuals across all career stages, and with a mix of professions (academia, nonprofit research institutes, federal agencies, and occasionally industry partners). Committee members volunteer their time and expertise, and a diverse skillset among them is important; typically, the committees consist of at least two or three researchers/academics and at least two or three cyberinfrastructure experts (including data management and standards experts, as well as experts in computational platforms, e.g., JupyterHub).

The eight to twelve Steering Committee members chart the continued evolution of OceanHackWeek, conduct strategic planning, and focus on longer term issues like funding. The Organizing Committee forms each year to plan all aspects of the annual event: deciding tutorial topics and inviting instructors, planning the project brainstorming session, creating the schedule and in-person logistics, setting up the JupyterHub, and managing applications, participant selection, and other tasks.

The 10 to 20 member Organizing Committee is composed of a subset of Steering Committee members, previous attendees, and/or interested individuals, often with a number of individuals from that year’s hosting institution. The Organizing Committee typically starts meeting approximately six months before the annual event, with one-hour meetings every two weeks until two months before the event, when meetings begin to take place weekly. The Organizing Committee uses a distributed leadership model, rotating the chair of each meeting among the committee members. The main duties of the Organizing Committee are separated into task groups focused on: event coordination, participant application/selection, communication, tutorials, projects, and cyberinfrastructure. Each committee member joins one or two task groups, with each task group working asynchronously and/or meeting as necessary to complete their responsibilities. In years with a hybrid event, the task groups have taken the integration of virtual and in-person participants into consideration. For the actual event, we typically have an additional five to ten helpers who act as project facilitators and mentors, assist with in-person logistics, and serve as tutorial instructors. These helpers join the Organizing Committee meetings approximately two to four weeks before the event. In the years with a virtual option, we typically had more helpers (~10), in years with only an in-person option, fewer (~5).