The Earth Data Science Corps: A Model for Teaching & Learning **Environmental Data Science Skills (Award #1924337)** Nathan Quarderer¹ (current co-PI); Jennifer Balch¹ (current PI, former co-PI); Leah Wasser (former PI)²

¹ University of Colorado Boulder/CIRES/Earth Lab/ESIIL; ² pyOpenSci

The Earth & Environmental Sciences (EES) produce vast amounts of data at a pace and on a scale that precipitate a need for EES researchers who are equipped with the technical data analytic skills required to work with large EES data sets. There are currently limited opportunities to learn these critical earth and environmental data science (EDS) skills leading to a gap between the demand for and supply of well trained data analysts, and contributes to a lack of diversity in the workforce. One model for meeting these demands is the NSF-supported Harnessing the Data Revolution (HDR) Earth Data Science Corps (EDSC) which has engaged with 60 students and 8 faculty partners from Minority Serving Institutions (MSIs) and Tribal Colleges and Universities (TCUs) in the 3 years of the program. Through online instruction and a 12-week paid internship that includes training in fundamental Python programming and geospatial science, we have demonstrated significant growth across different aspects of participants' technical Python and data science skills, as well as their science identity and sense of belonging to a larger population of data scientists. These findings will be discussed along with implications for teaching EDS to members from historically underrepresented communities.

Partner Institutions & Faculty Partners

Oglala Lakota College (Kyle, SD)

• Jim Sanovia, Jason Tinant

United Tribes Technical College (Bismarck, ND) • Emily Biggane, Jessica Logan, Sajjad Akam

Metropolitan State University Denver (Denver, CO) • David Parr, Sylvia Brady

Front Range Community College (Westminster, CO) Jennifer Muha

- CU Boulder (Boulder, CO)
- William Travis, Imtiaz Rangwala

Student Demographics (*n* = 53)



www.PosterPresentations.co

Course Content

- Week 0 Software Carpentry Training Workshop
- Shell, bash, git, GitHub, version control
- Week 1 Tool Introduction 5 #slack (CO (
- Spatial Chat, Slack, Google Colab, Discourse
- discrete matplextlib Week 2 – Python Fundamentals • Variables, lists, operators, plotting with matplotlib
- pandas Week 3 – Tabular & Time-Series Data
- Pandas, parse_dates, 2013 Boulder flood, precipitation/discharge
- Week 4 Vector & Raster Spatial Data 🕞 GeoPandos 📈 xarray 🕵
- Geopandas, earthpy, rioxarray, CRS reproj., .clip, .crop, .shp, .tif
- Week 5 Clean & Reproducible Code
- Science communication
- Weeks 6 11 Group Project Work Time
- Practice/ final presentations ; peer + professional mentoring

Project Topics





Year 1 (2020)

- •COVID-19 South Dakota Tribal Data Infrastructure (OLC, UTTC) •Plotting the Platte: Measuring Vernal Change over Multiple Seasons
- Using LandSat Imagery (MSU)
- •Flooding on Tribal Lands (UTTC; FRCC)
- Emerald Ash Borer Detections in Minnesota Cities (FRCC)
- •NEON Characterizing Vegetation Using NEON Remote Sensing Data (UTTC; CU)
- •USGS NC CASC Climate Projections for Habitat Sustainability (CU)

Year 2 (2021)

- •Air Quality Data & Transportation Emissions Changes over time & COVID-related trends in Denver, CO (MSU)
- •Climate Scenarios for Managing Grassland Ecosystems in the Great Plains (CU)
- •Envisioning Resilient Campuses in a Changing Climate (OLC)
- •Methane Flux Dynamics in the Prairie Pothole Wetlands (UTTC)

Year 3 (2022)

- Predicting Streamflow Data from Snowpack: An Exploration (CU) •Correlation between Snowpack Levels and Reservoir Levels in Denver Water Supply (MSU)
- Pine Ridge Reservation Old Church Discovery (**OLC**)
- •Bird Behaviors within the Prairie Pothole Region (**UTTC**)

Research Outline

- Questions How does involvement in an immersive, online, projectbased data science learning environment contribute to the development of participants' :
- Self-confidence in their Earth data science technical skills?
- Science and data science communication skills?
- Science identity and sense of belonging?

Methods

- 5 Likert-like survey instruments
- Retrospective post-then-pre design
- Rasch measurement
- ANOVA, post-hoc *t*-testing

Survey Responses (%)

Python Skills (n = 46)

	Not Comfortable Slightly C	omfortable	Moderate	ely Comfortable	Ve	ry Comfortable							
	BEFORE		AFTER										
Use Python to work with time series data	87	11 2	29	30		41							
Work with spatial data in Python	87	94	2 30	35	j	33							
Import text files into Python	85	13 2	11	37		52							
Summarize data in Python	85	13 2	22	46		32							
Plot data in Python with matplotlib	85	13 2	11	39		50							
	Not Comfortable Slightly	Comfortable	Modera	tely Comfortabl	e 📕 Ve	ery Comfortable							
	BEFORE		AFTER										
Work with tabular data in Python	85	11 22	22	37		41							
Manipulate data in Python	80	16 4	31	3	9	30							
Document Python code	79	15 42	13	46		41							
Use loops and conditional statements to create efficient workflows	76	13 9 2	9	41	33	3 17							

Data Science Skills (n = 46)

Not Confident Slightly Confident Moderately Confident Very Confident

	BEFORE	Ε				
I can use Jupyter Notebooks	78	13 7 2	2 20	30		48
I can publish my code	76	13 7 4	7	30	35	28
I can write efficient and modular code	67	24 7 2	2 24	4	57	17
can document code so that others can easily understand and use it	61	20 17 2	47	57		32
l can use scientific programming to work with data	57	30 11 2	13	57	,	30

Not Confident Slightly Confident Moderately Confident Very Confident

	BEFORE							AFTER									
l can use remote sensing data	!	55		24	17	4	9	:	28	43			20				
can create reproducible workflows	44			37		4	2	20		50		2	28				
I can find and access data	28	3	3	32		7	9		30		61						
I can draw analytical conclusions from data	24	37		30		9	9		50			41					
l can visualize scientific data	24	37		32		7	15	5	37		48						
							L										

Data Science Communication (n = 48)

		Stron	igly Disgree	e 🗾 Dis	gree	Neithe	er A	Agree	nor Dis	agree	Agree	St	rongly	Agree		
			BE	FOR	Ε			AFTER								
I can create plots and map visualize data using Pytl	s to hon	Z	14	31	19	6			81	81			19			
l am comfortable writing code analyze d	e to lata	25	25			13 6	6	6	6 18		63					
I know how to document formulas/ca so others can understand and us	ode se it	13	49		19	9 13 6		18		69						
With enough time I feel I can le new programming langua	arn ges	25	38 31		31	6		44	Ļ	56						
I am confident I can communic about science by writing bl	ate ogs	6 13	31 50			6	6 13 56			25		5				
	L	Strongly	/ Disgree	Disgre	e	Neither A	٩gr	ree no	or Disag	ree 🗾 A	gree	Stror	ngly A	gree		
			BEF	ORE				AFTER								
l am comfortable analyzing data	6	44		31		19		19		50	50		31			
When I write about science, I can present my ideas clearly		31	56			13		19		50	50		31			
I am confident in my ability to explain science to non-scientists	2	25	56			19		6	5 56		6		38			
I am confident in my ability to collaborate with others on projects	2	25	56			19		12		44		44				
I feel comfortable using data in spreadsheets	19		31 50					38				62				

Data Science Practices (n = 48)

Science Identity (n = 46)

Ordinal responses from each Likert instrument converted to ratio scale through Rasch measurement [item/person reliability, fit statistics, item difficulty, person ability] Significant growth observed across each survey dimension

Pytho Data S Data S Data S Scien







Survey Responses cont. (%)



		Stro	ongly	/ Disgree		Disgre	e	Neit	her A	gre	e nor Di	sagree		Agree	S	trongly A	gree		
				BEF	OF	RE				AFTER									
I have come to think of myself as a 'scientist'	11	1	7	30		22	2	20)	2	9	28		24		37			
Being a scientist is an important reflection of who I am	11	11		32		26		20)	7	26	;	:	26		41			
In general, being a scientist is an important part of my self-image	9	9		28		30		24		24	20		33	3		41			
I have a strong sense of belonging to the community of scientists	7	22	2	38	3		2	24	9	9		35			32	24	4		
I feel like I belong in the field of science	7	17		30		24		22		7	15	:	30			48			

Rasch Measurement; ANOVA; *t*-testing



Post-hoc *t*-testing

	t	Cohen's <i>d</i>
n Skills	15.43 ***	3.22
Science Skills	10.45 ***	2.18
Science Communication	10.27 ***	2.09
Science Practices	7.50 ***	1.53
ce Identity	3.32 **	0.69