The Earth Data Science Corps: A Model for Teaching & Learning Environmental Data Science Skills (Award #1924337)

Nathan Quarderer[1] (current co-PI); Jennifer Balch[1] (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

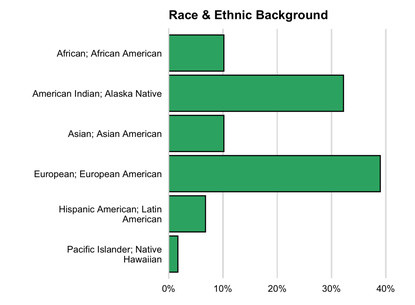
United Tribes Technical College (Bismarck, ND)
- Emily Biggane, Jessica Logan, Sajjad Akam

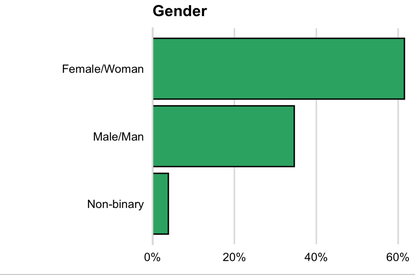
Metropolitan State University Denver (Denver, CO)
- David Parr, Sylvia Brady

Front Range Community College (Westminster, CO)
- Jennifer Muha

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

Student Demographics (*n* = 53)





Course Content

Week 0 – Software Carpentry Training Workshop
- Shell, bash, git, GitHub, version control

Week 1 – Tool Introduction
- Spatial Chat, Slack, Google Colab, Discourse

Week 2 – Python Fundamentals
- Variables, lists, operators, plotting with matplotlib

Week 3 – Tabular & Time-Series Data
- Pandas, parse_dates, 2013 Boulder flood, precipitation/discharge

Week 4 – Vector & Raster Spatial Data
- 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

EDSC Student Project Blogs & Presentations!!!

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

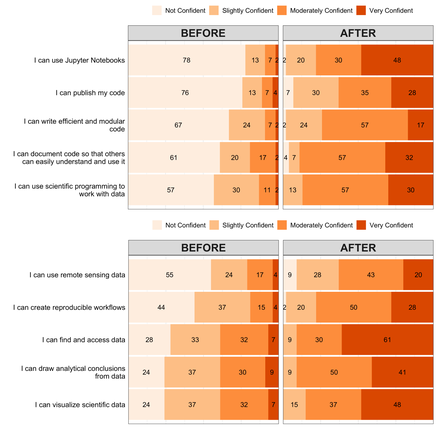
Questions - How does involvement in an immersive, online, project-based data science learning environment contribute to the development of participants' :
- Self-confidence in their Earth data science technical skills?
- Science and data science communication skills?
- Science identity and sense of belonging?

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

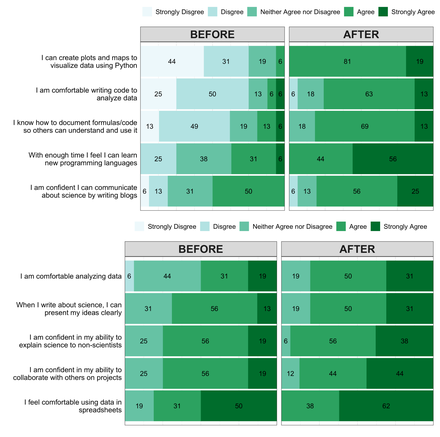
Survey Responses (%)

Python Skills (*n* = 46)

Data Science Skills (*n* = 46)

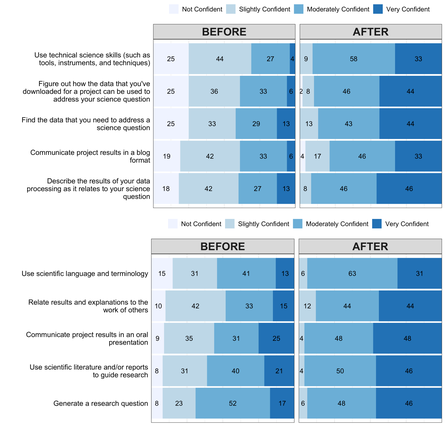


Data Science Communication (*n* = 48)

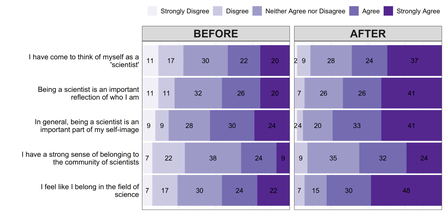


Survey Responses cont. (%)

Data Science Practices (*n* = 48)

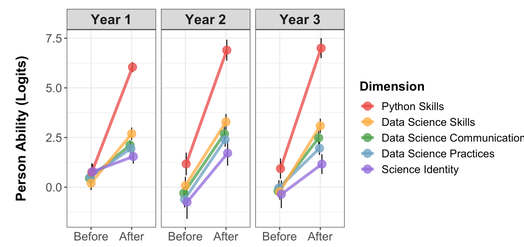


Science Identity (*n* = 46)



Rasch Measurement; ANOVA; *t*-testing

- 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



ANOVA

	Df	Sum Sq	Mean Sq	*F* value	*p* value	*Eta sq. (part)*
Trial	1	1004.44	1004.44	324.75	< 2e-16 ***	0.41
Dimension	4	563.71	140.93	45.56	< 2e-16 ***	0.28
Cohort	2	1.34	0.67	0.22	0.81	0.0009
Residuals	469	1422.77	3.09			

Post-hoc *t*-testing

	t	Cohen's *d*
Python Skills	15.43 ***	3.22
Data Science Skills	10.45 ***	2.18
Data Science Communication	10.27 ***	2.09
Data Science Practices	7.50 ***	1.53
Science Identity	3.32 **	0.69