

San Carlos Apache College - Python Workshop

Nate Quarerer
Earth Analytics Education (Earth Lab/CIRES)

Monday, November 6th, 2023



Agenda:

- Create GitHub accounts (<https://github.com>)
- Introduce [Earth Lab](#) & [ESIIL](#)
- Introduce GitHub & Codespaces
- Break
- Introduce EDS and GIS in Python

What is Earth Data Science?

In-Demand Skills at Intersection of Science and Data Science: Earth Data Science

Earth & Environmental Science



- Domain knowledge
- Scientific inquiry

Communication & Collaboration “Super Skills”



- Communicate with non-technical audiences
- Collaboration best practices

Core Computation Skills



- Scientific programming
- Command line (bash)
- Git / Github

Earth & Environmental Data Science (EDS)

Reproducible Workflows



- Shared, collaborative workflows
- Version control
- Reproducible reports

Data Structures

- Raster
- Vector
- Hierarchical
- Json
- Tabular / Time Series

Use Diverse Data Types



- Learn data structures
- Find & get data
- Clean data

Data Sources

- Remote Sensing
- Social Media
- *In situ* Measured
- Long-term Sensor Networks

Why EDS Education?

Why EDS Education?

1. Volume of earth & environmental data



Why EDS Education?

1. **Volume of earth & environmental data**
2. **Gap between supply of & demand for Earth Data Scientists**

THE DATA SCIENCE / ANALYTICS LANDSCAPE



2,350,000

DSA job listings in 2015

By 2020, DSA job openings
are projected to grow

15%

364,000

Additional job listings
projected in 2020

Demand for both Data
Scientists and Data Engineers
is projected to grow

39%

DSA jobs remain open

5 days

longer than average

DSA jobs advertise average salaries of

\$80,265

With a premium over all BA+ jobs of

\$8,736

81%

Of DSA jobs require workers with
3-5 years of experience or more

IBM, 2017

- The McKinsey Global Institute predicts that the U.S will experience a shortage of 250,000 employees trained in data skills by 2024. (Henke et al, 2016)
- #3 on the Best Jobs for 2020 list on Glassdoor

Why EDS Education?

1. **Volume of earth & environmental data**
2. **Gap between supply of & demand for Earth Data Scientists**
3. **Limited opportunities to learn Earth Data Science**



Data Science
UNIVERSITY OF COLORADO BOULDER



PennState
Institute for Computational
and Data Sciences

coursera
UDACITY

edX

ESF

State University of New York College of
Environmental Science and Forestry

MIT

xPRO  UNIVERSITY OF WISCONSIN
DATA SCIENCE

Duke MIDS
Master in Interdisciplinary Data Science

BERKELEY
Institute for
Data Science

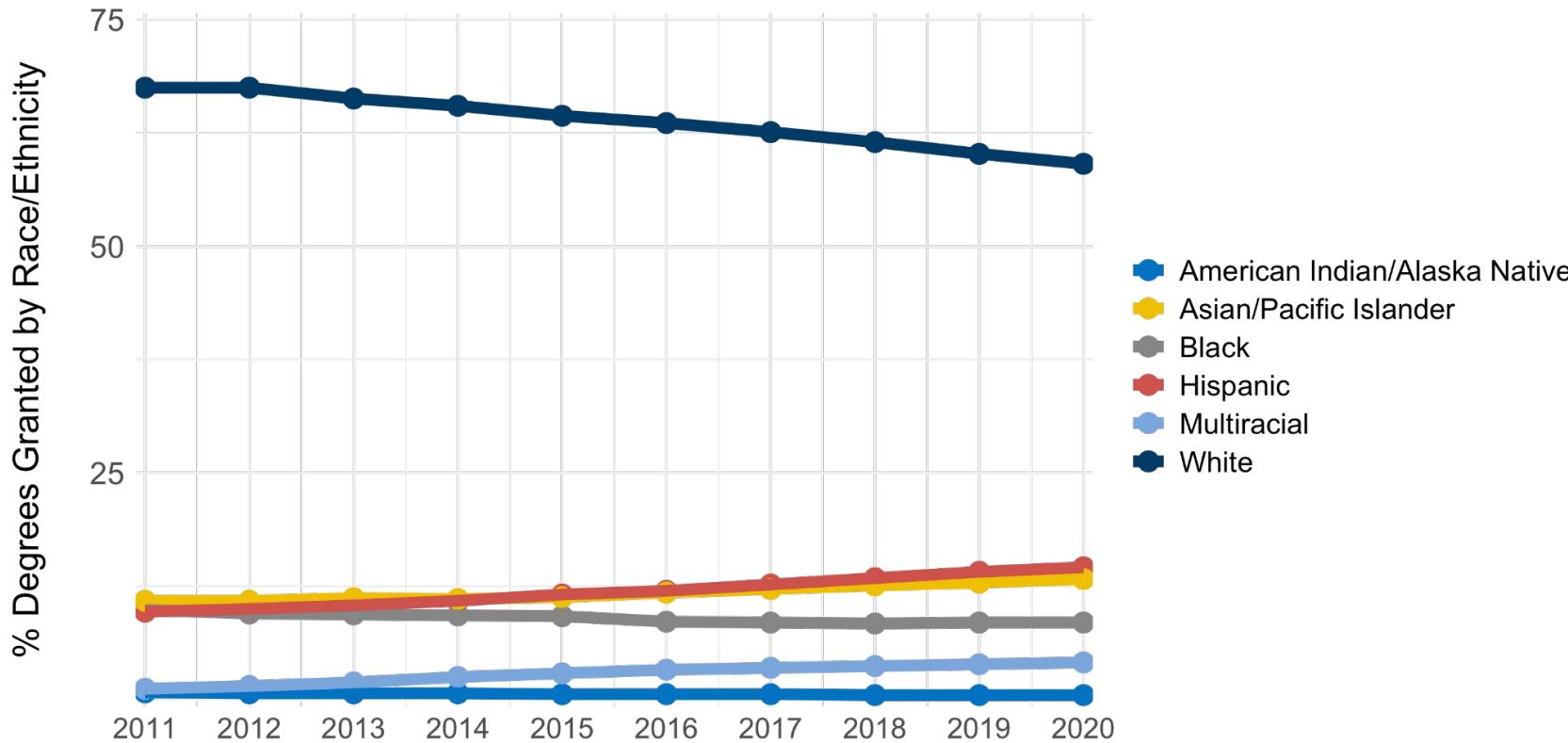


BRENSCHOOL
UC SANTA BARBARA

Why EDS Education?

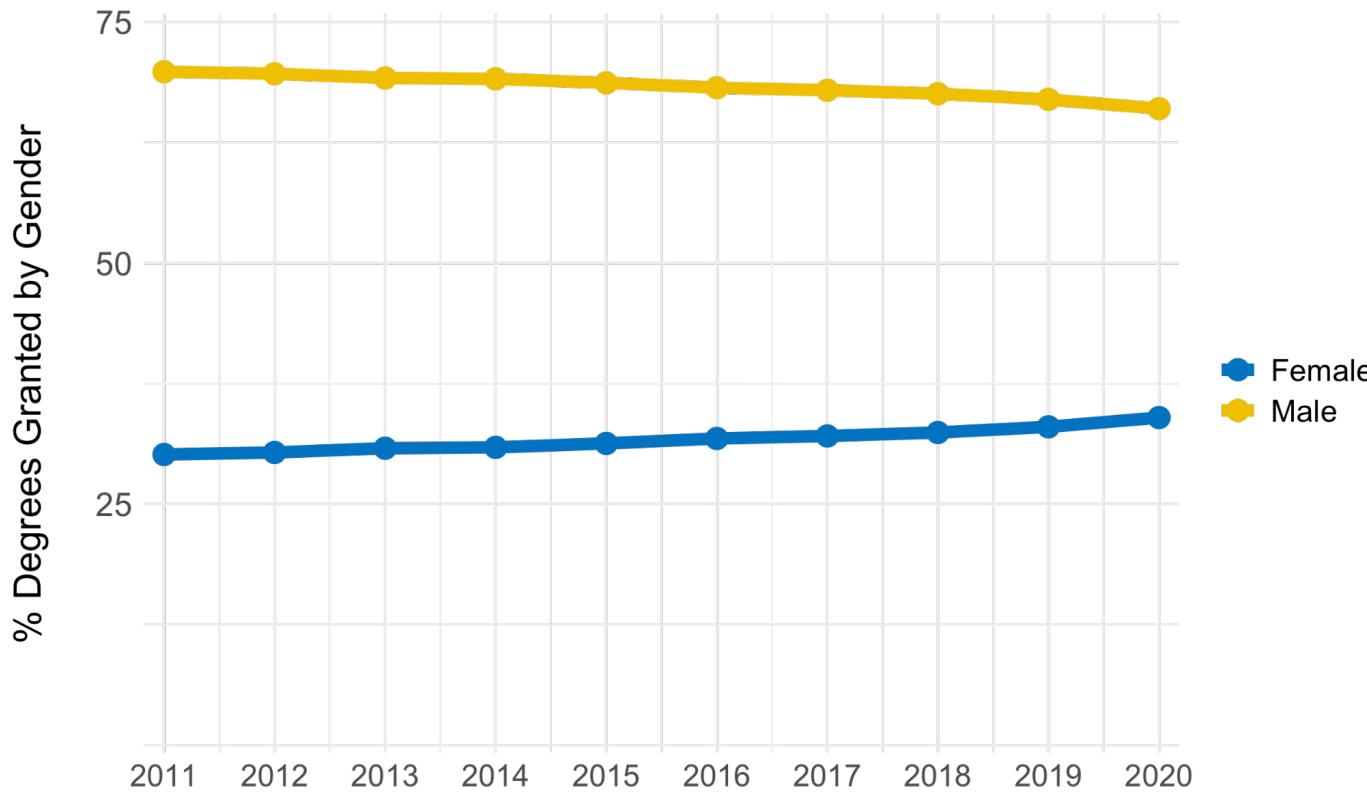
1. **Volume of earth & environmental data**
2. **Gap between supply of & demand for Earth Data Scientists**
3. **Limited opportunities to learn Earth Data Science**
4. **Lack of diversity in STEM (esp. Computer Science & Engineering)**

STEM Degrees & Certificates Granted by Race/Ethnicity in the US: 2011-2020



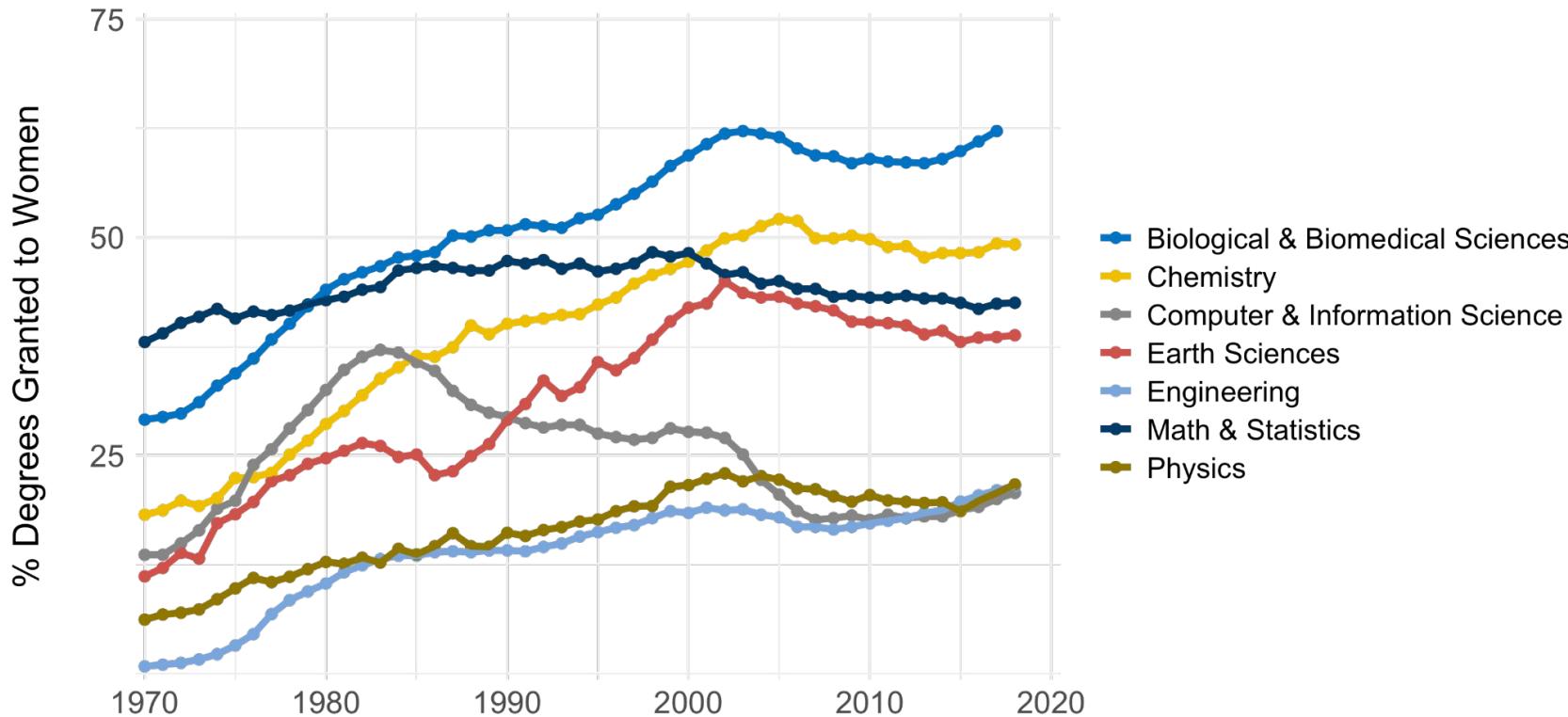
Source: National Center for Education Statistics,
Table 318.45 - Trends in STEM Degrees by Race/Ethnicity
https://nces.ed.gov/programs/digest/d21/tables/dt21_318.45.asp

STEM Degrees & Certificates Granted by Gender in the US: 2011-2020



Source: National Center for Education Statistics,
Tables 318.45 - Trends in STEM Degrees by Gender
https://nces.ed.gov/programs/digest/d21/tables/dt21_318.45.asp

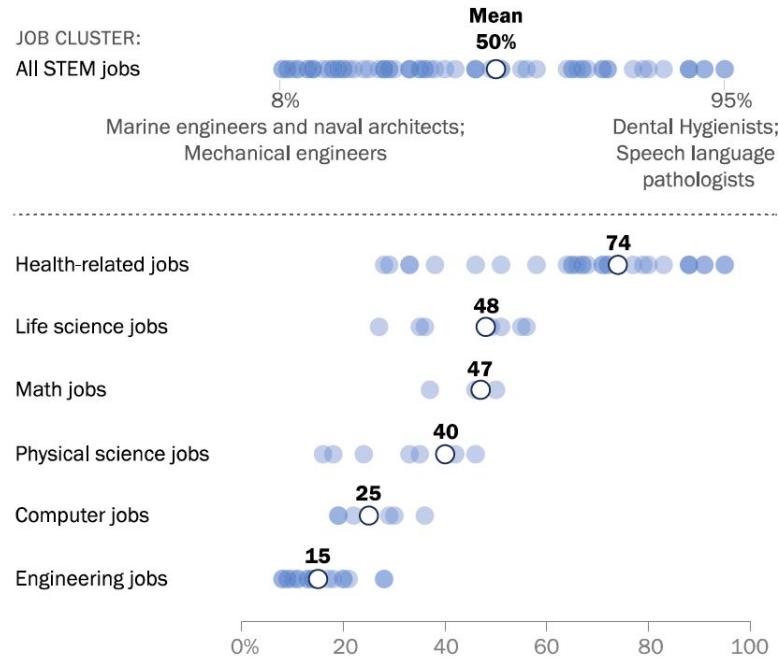
STEM Bachelor's Degrees Granted to Women in the US by Field: 1970-2018



Source: National Center for Education Statistics,
Tables 325 - Trends in Degrees by Field and Gender
https://nces.ed.gov/programs/digest/d19/tables_3.asp#Ch3Sub25
<https://www.aps.org/programs/education/statistics/womenmajors.cfm>

Representation of women in STEM varies across job clusters

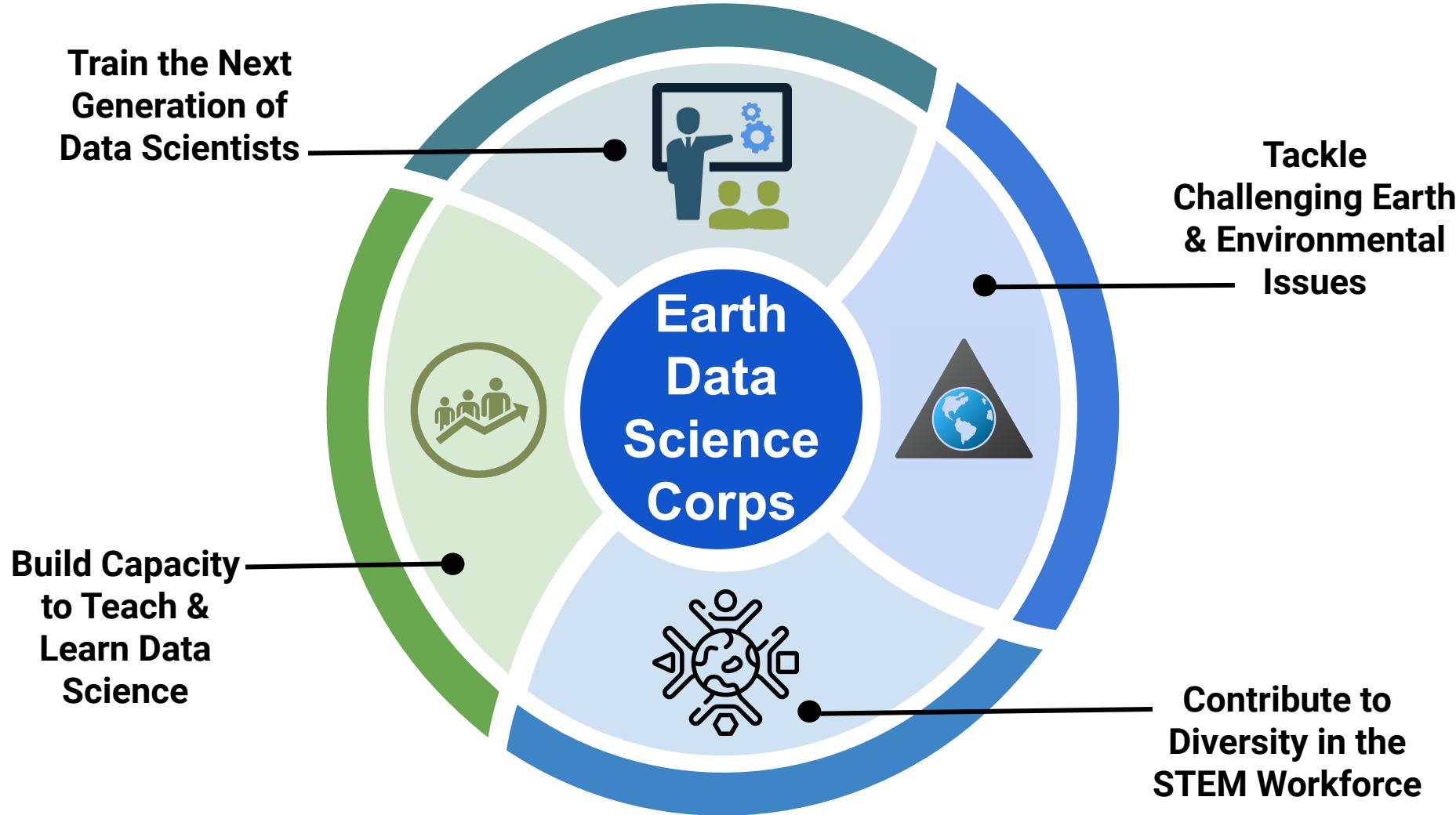
Share of women in each of the following job clusters



Note: Based on employed adults ages 25 and older. Each circle represents a single occupation (e.g., mechanical engineer, registered nurse). STEM stands for science, technology, engineering and math. Engineering includes architects.

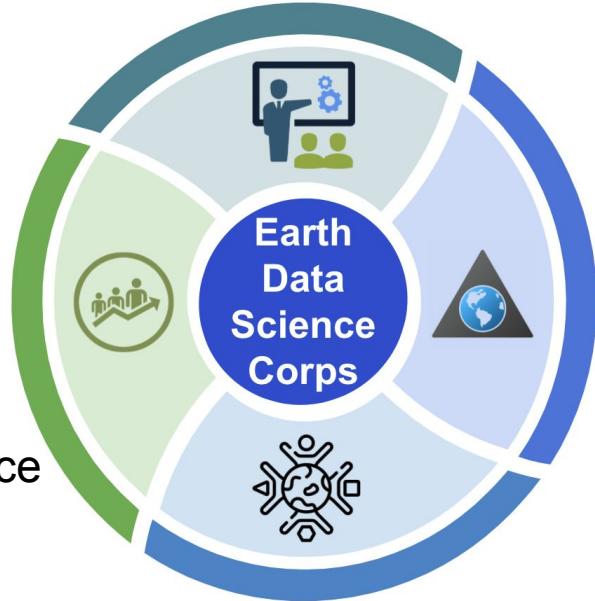
Source: Pew Research Center analysis of 2017-19 American Community Survey (IPUMS) “STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity”.

**What is the
EDSC?**



The Earth Data Science Corps

- Earth Lab Education team & 4 partner institutions
- **12-week** paid summer internship (\$5,000) for undergraduates; **45 total students** in Yr 1 & 2
- Learn **Python**; relevant **GIS, Earth/Environmental** science
- **Online** (Zoom/Spatial Chat); Modeled after GEOG 4463/5463 (**EDS Bootcamp**)
- **Faculty mentors** develop critical EDS teaching skills
- Supported through **NSF Award #1924337**; \$1.3M; 3 yrs (Wasser, Balch, Quardeerer)
- Harnessing the Data Revolution (**HDR**) (10 Big Ideas)



**Who is the
EDSC?**

Partner Institutions



UNITED TRIBES
TECHNICAL COLLEGE®



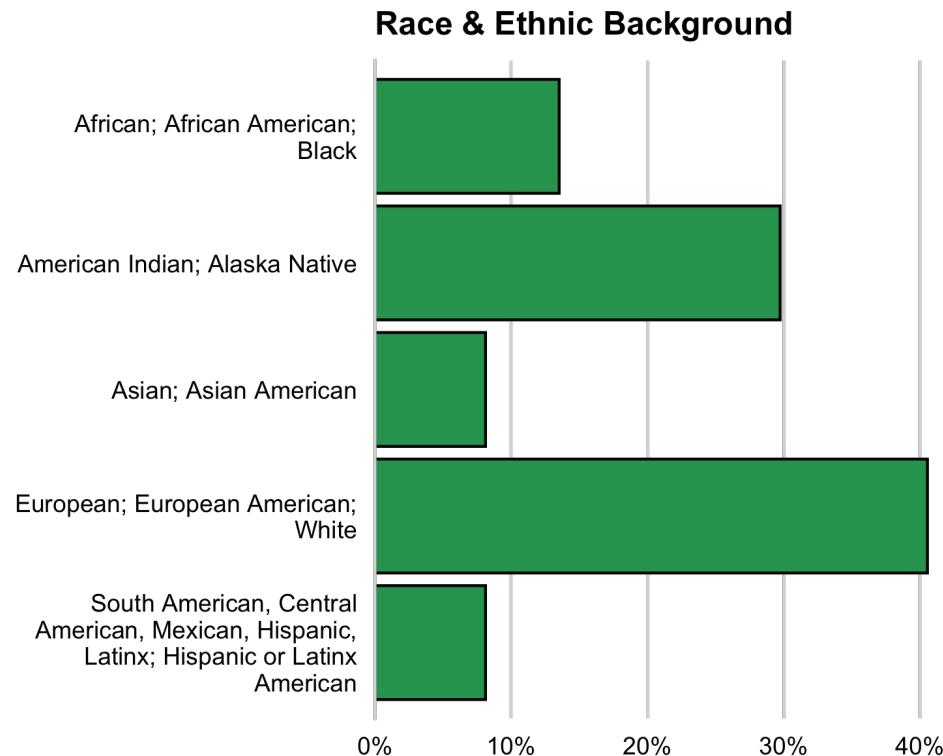
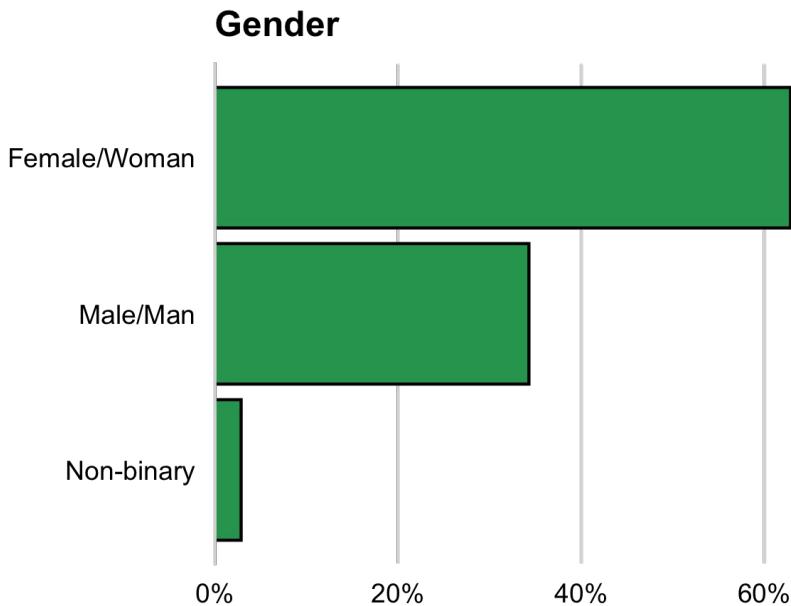
METROPOLITAN
STATE UNIVERSITYSM
OF DENVER



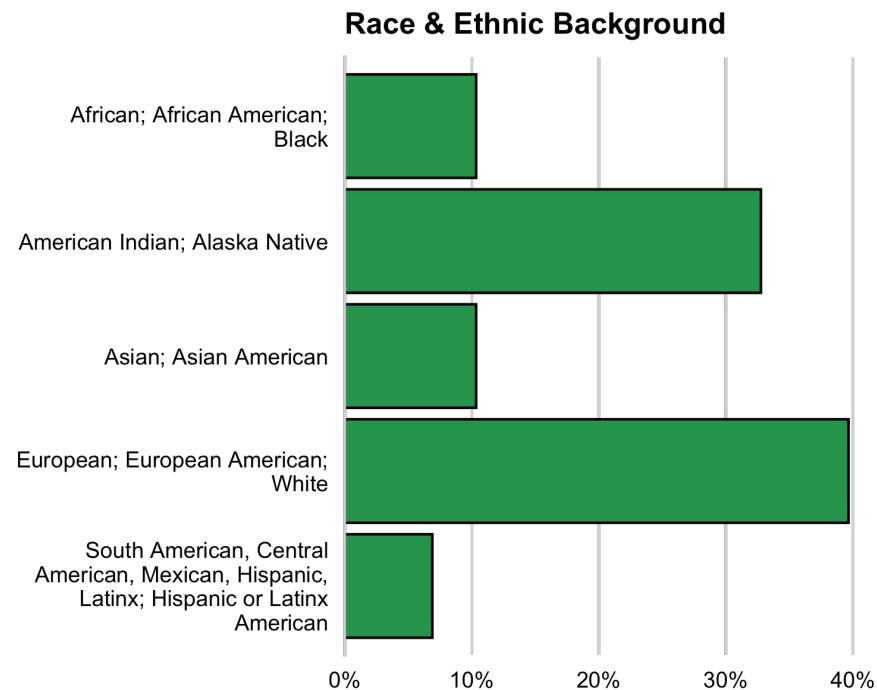
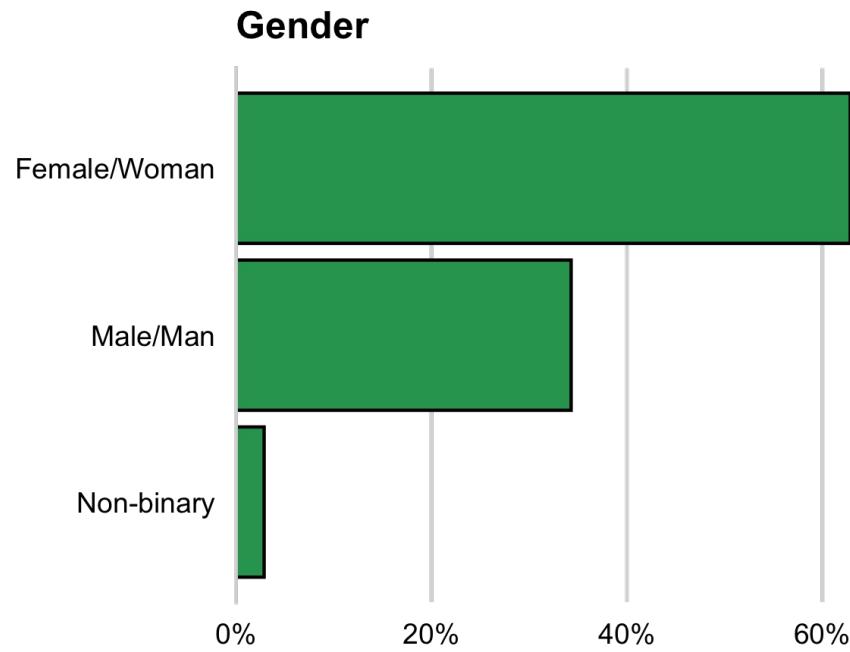
University of Colorado
Boulder

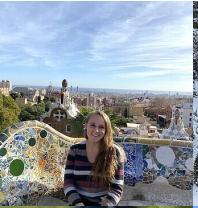


Student Demographics



EDSC Student Demographics (2020-2022)





**What does the
EDSC do?**

wk0

Software Carpentry
Training Workshop



- Shell
- Bash
- Intro. to Python
- Version Control w/
Git & Github

wk1

Kickoff,
Icebreakers,
Introduce Tools



- Spatial Chat
- Slack
- Discourse
- Google Colab

wk2

Python
Fundamentals



- Variables
- Lists
- Operators
- Plotting w/
matplotlib

wk3

Tabular &
Time-Series
Data



- Pandas
- `parse_dates`
- 2013 Boulder flood
- Precipitation
- Discharge

wk4

Vector &
Raster
Spatial Data



- Geopandas,
EarthPy, Rioxarray
- CRS reprojections
- Clip, crop
- .shp, .tif

wk5

Clean &
Reproducible Code

Science
Communication

- What the Fork?!
- Data Visualization
- Workflow Diagrams
- Message Box

wk6-10

Group Project
Work Time

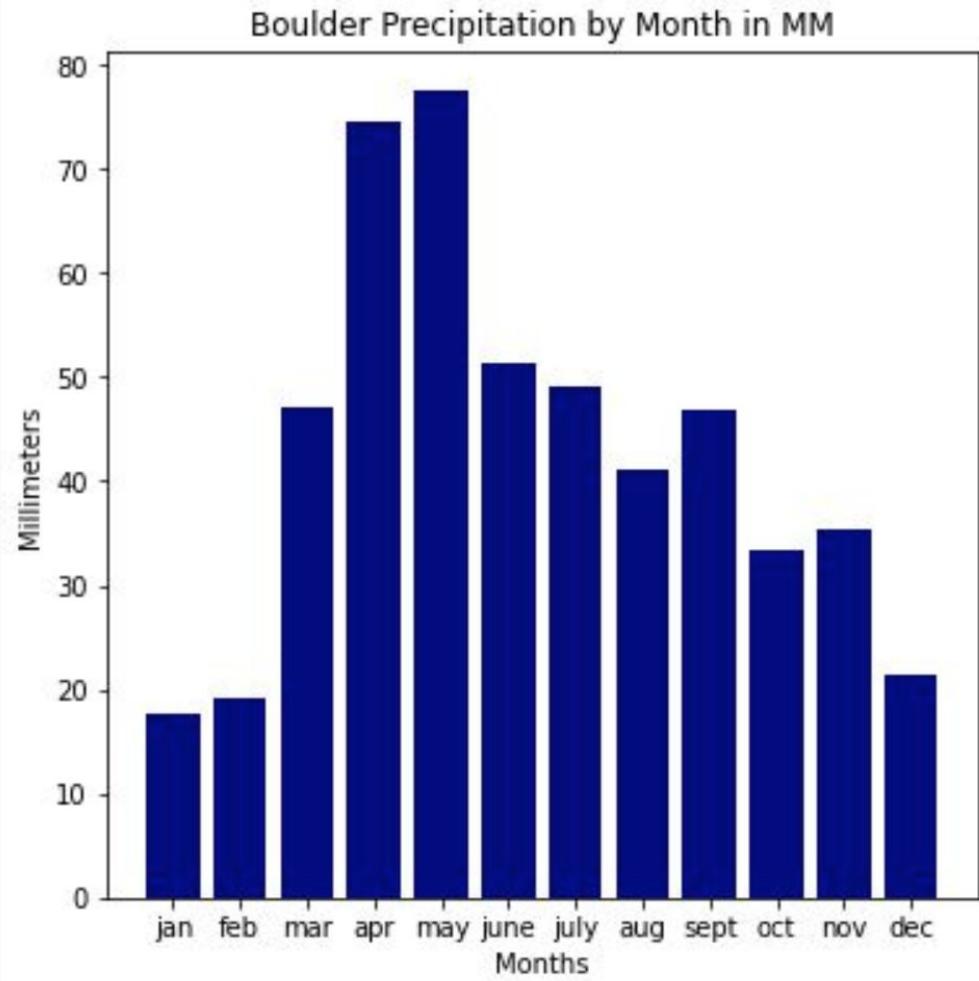
- Work with peers at home institution
- Led by Faculty Mentors
- Earth/ENVS Topics with Local & Cultural Relevance

wk11

Final
Presentations

- 10 min talk; 5 min questions
- Recorded for CIRES' YouTube
- Blogs published on EL website

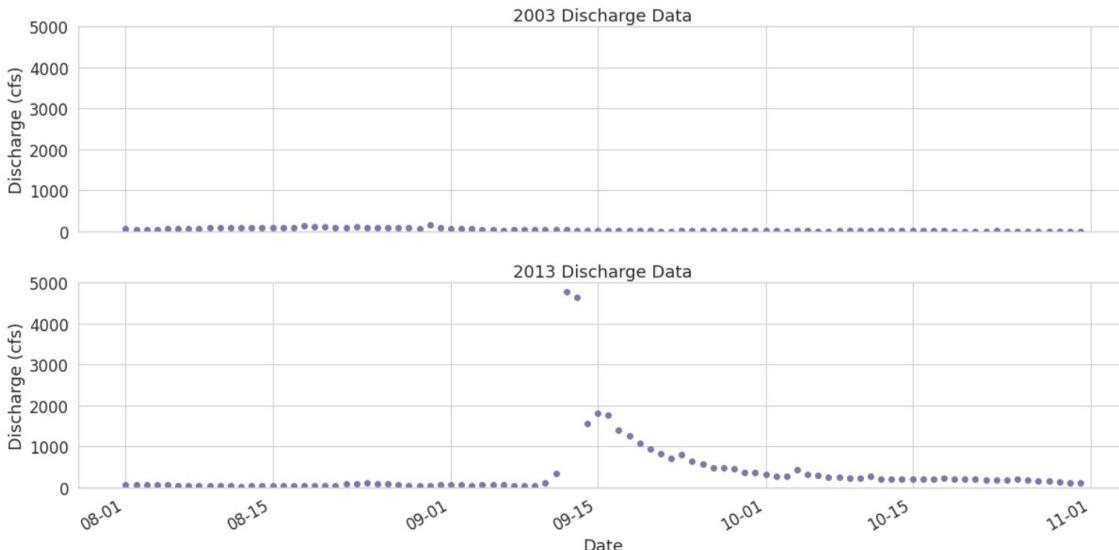
```
# Copy/paste the code from the cell above to this cell  
# Modify values as instructed  
  
import matplotlib.pyplot as plt  
fig, ax = plt.subplots(figsize=(6, 6))  
ax.bar(boulder_precip_months,  
       boulder_precip_mm,  
       color="navy")  
ax.set(title="Boulder Precipitation by Month in MM",  
      xlabel="Months",  
      ylabel="Millimeters")  
plt.show()
```



```
discharge_2003= stream_discharge_boulder['2003-08-01':'2003-10-31']
discharge_2003
```

```
discharge_2013= stream_discharge_boulder['2013-08-01' : '2013-10-31']
discharge_2013
```

Average Stream Discharge 2003 vs. 2013 - Boulder, CO



```
f, (ax1, ax2) = plt.subplots(2, 1, figsize=(20, 10))
plt.suptitle("Average Stream Discharge 2003 vs. 2013 - Boulder, CO", fontsize = 25)

ax1.scatter(x = discharge_2003.index.values,
            y = discharge_2003["disValue"],
            color = "m")
ax1.set(title = "2003 Discharge Data",
        xlabel = "Date",
        ylabel = "Discharge (cfs)",
        ylim = [0, 5000])

ax2.scatter(x = discharge_2013.index.values,
            y = discharge_2013["disValue"],
            color = "m")
ax2.set(title = "2013 Discharge Data",
        xlabel = "Date",
        ylabel = "Discharge (cfs)",
        ylim = [0, 5000])

plt.subplots_adjust(hspace=0.25)

date_form= DateFormatter("%m-%d")
ax1.xaxis.set_major_formatter(date_form)
ax2.xaxis.set_major_formatter(date_form)
```

```
# Add your code to plot the data here
f, ax1 = plt.subplots(figsize=(10, 10))

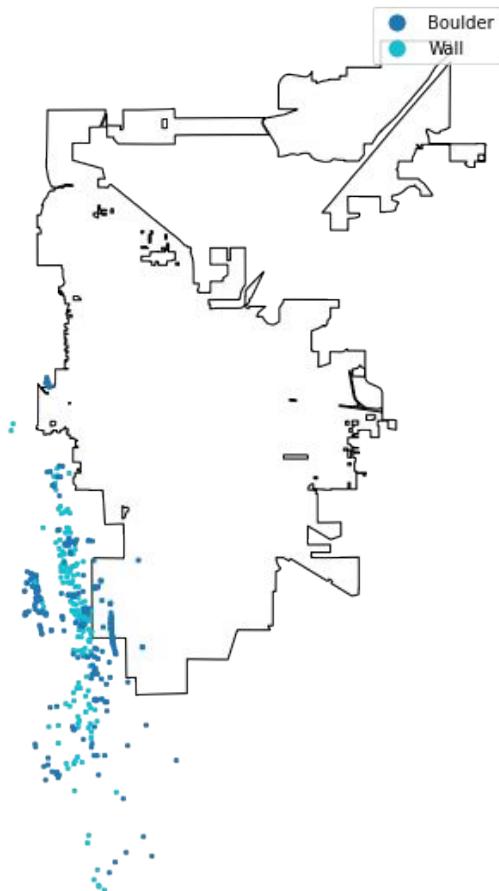
city_limits.plot(ax=ax1,
                  color="white",
                  edgecolor="black")

boulder_climbing_gdf.plot(ax=ax1,
                           markersize = 5,
                           marker = "X",
                           column='FormationType',
                           legend=True)

ax1.set_axis_off()

# Add a title
# I used this so I can change fontsize
f.suptitle("Climbing Locations near Boulder, CO", fontsize=15, color="Black")
#ax1.set(title="Climbing Locations in Boulder, CO")

plt.show()
```



```

fig, ax = plt.subplots(figsize=(15, 10))
ep.plot_bands(ned_30m_path_clip,
               ax=ax,
               cmap="hot",
               extent = ned_cl_extent)

# Plot the hillshade
ep.plot_bands(hillshade,
               ax=ax,
               alpha=0.5,
               extent = ned_cl_extent,
               cbar=False)

# Plot the S Zone trails
rmnp_trails_4326_clip.plot(ax=ax,
                            color="limegreen",
                            linestyle="dashdot",
                            linewidth=3,
                            label= "Trails")

# Plot the S Zone boundary
southzone_boundary.plot(color = "none",
                         edgecolor = 'black',
                         linewidth=2,
                         ax=ax)

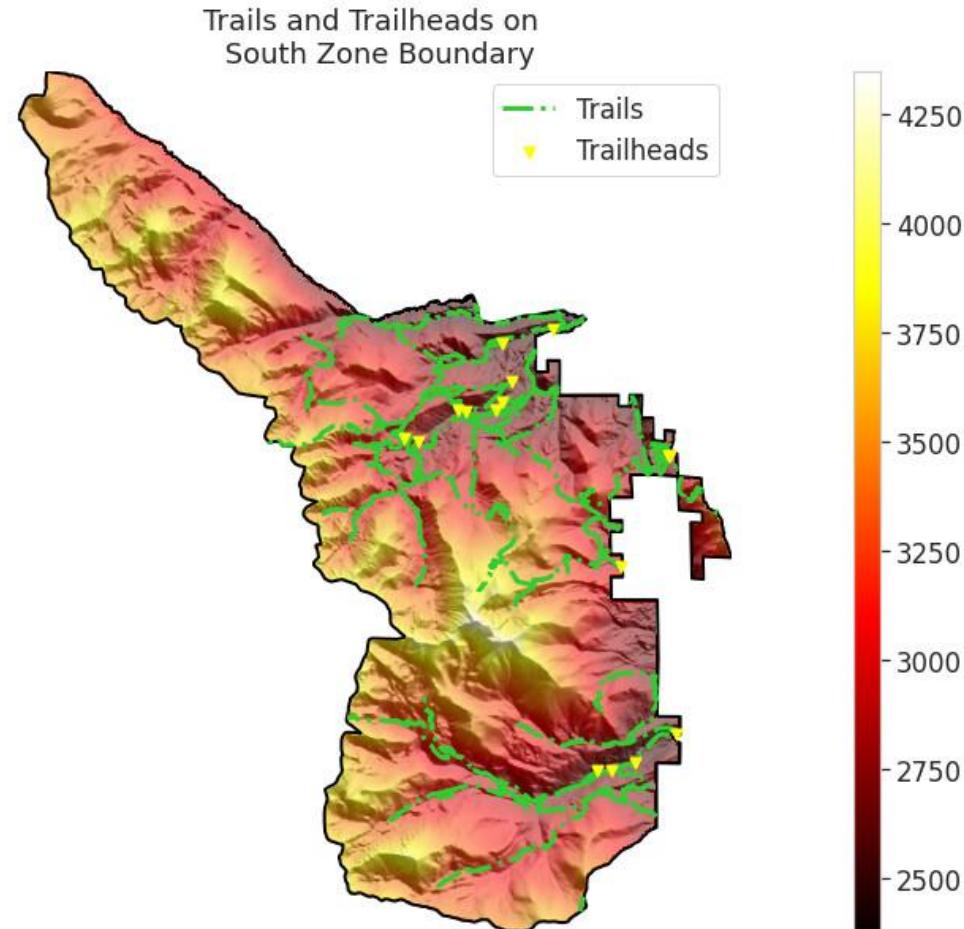
# Plot the S Zone trailheads
trailheads_clip.plot(ax=ax,
                      color="yellow",
                      marker="v",
                      markersize=50,
                      zorder=10,
                      label= "Trailheads")

ax.set(title = "Trails and Trailheads on \n South Zone Boundary")

# Turn off the outline or axis border on your plot
handles, labels = ax.get_legend_handles_labels()
ax.legend(handles, labels)

ax.axis('off')
plt.show()

```



EDSC Student Projects

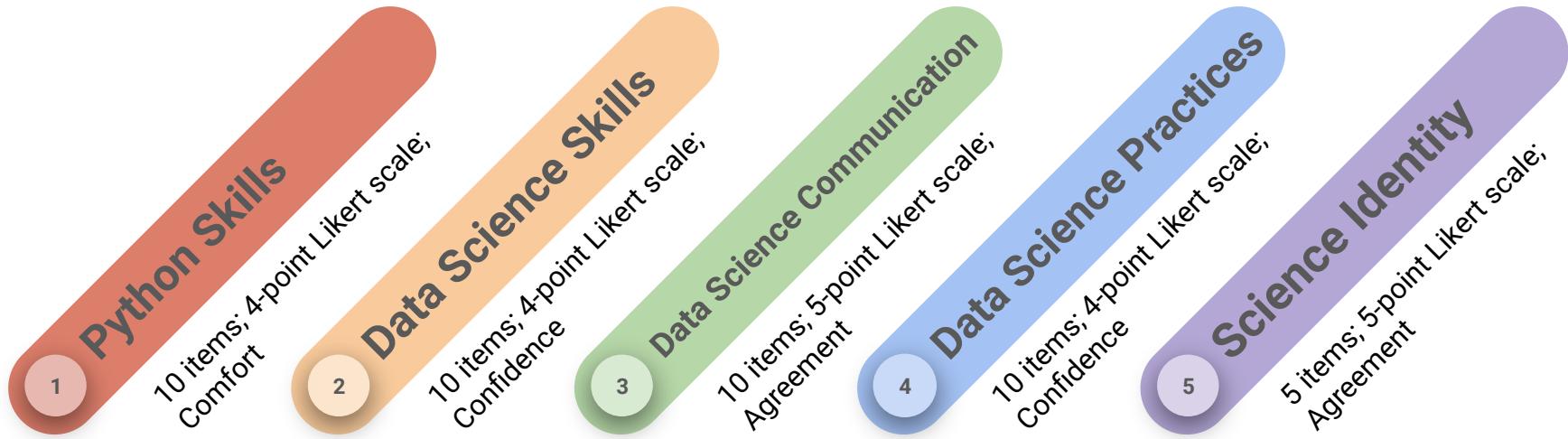
EDSC Student
Project Blogs &
Presentations!!!



Research Questions

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

Survey Dimensions



Post-then-pre retrospective approach

- Retrospective pretest methodology (Pratt et al 2000)
 - At end of program students report on their current knowledge (post-test)
 - Then, complete same task but perceive where they think they were when program began (pre-test)
 - Cuts down on time spent taking surveys (test fatigue)
 - Only requires students to be present once
 - Helps to avoid *response shift bias* (change in frame of reference following instruction)
 - Reduces likelihood of *pretest overestimation*
 - Has been shown to more accurately detect effect of intervention

Please provide two responses for each statement to tell us **how confident you feel** in doing these tasks **AFTER you finished participating** in the Earth Data Science Corps program **compared to BEFORE** participating in the program.

Python Skills

Not Comfortable Slightly Comfortable Moderately Comfortable Very Comfortable



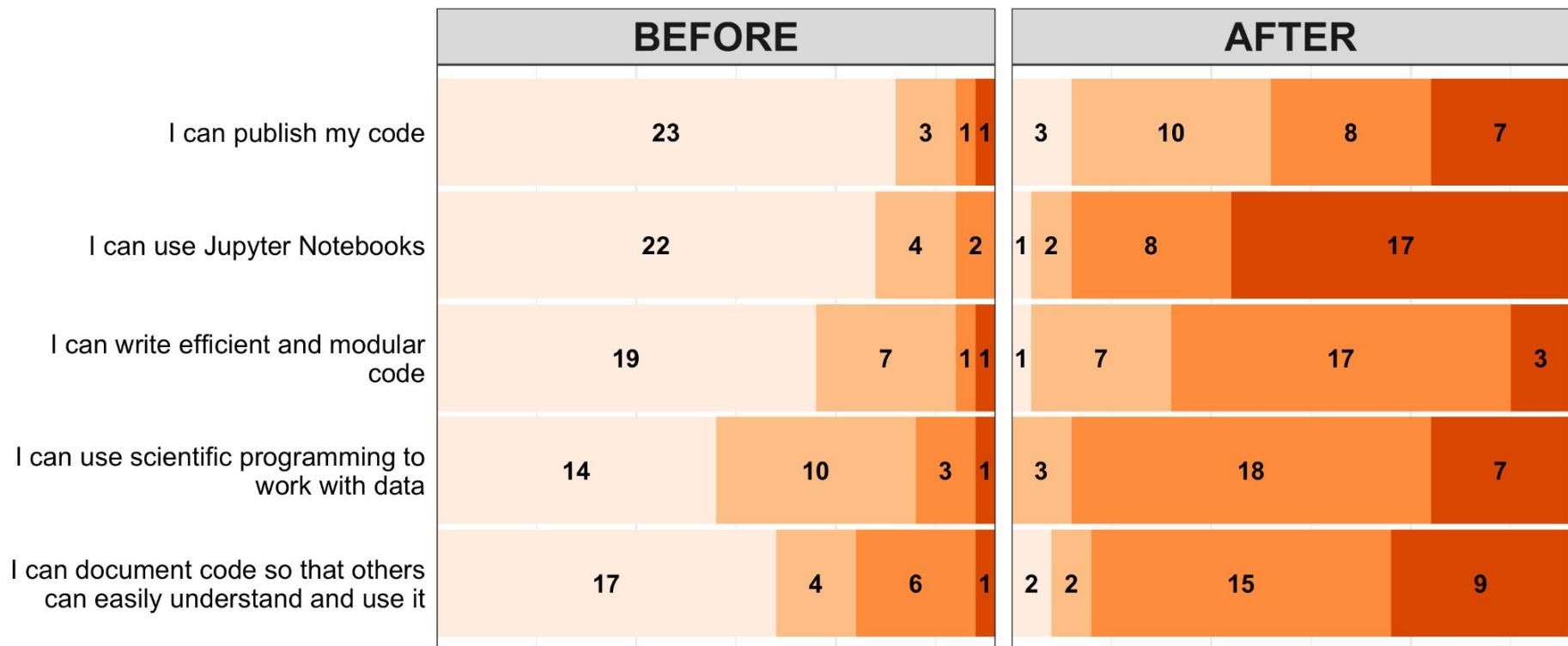
Python Skills cont.

Not Comfortable Slightly Comfortable Moderately Comfortable Very Comfortable



Data Science Skills

Not Confident Slightly Confident Moderately Confident Very Confident

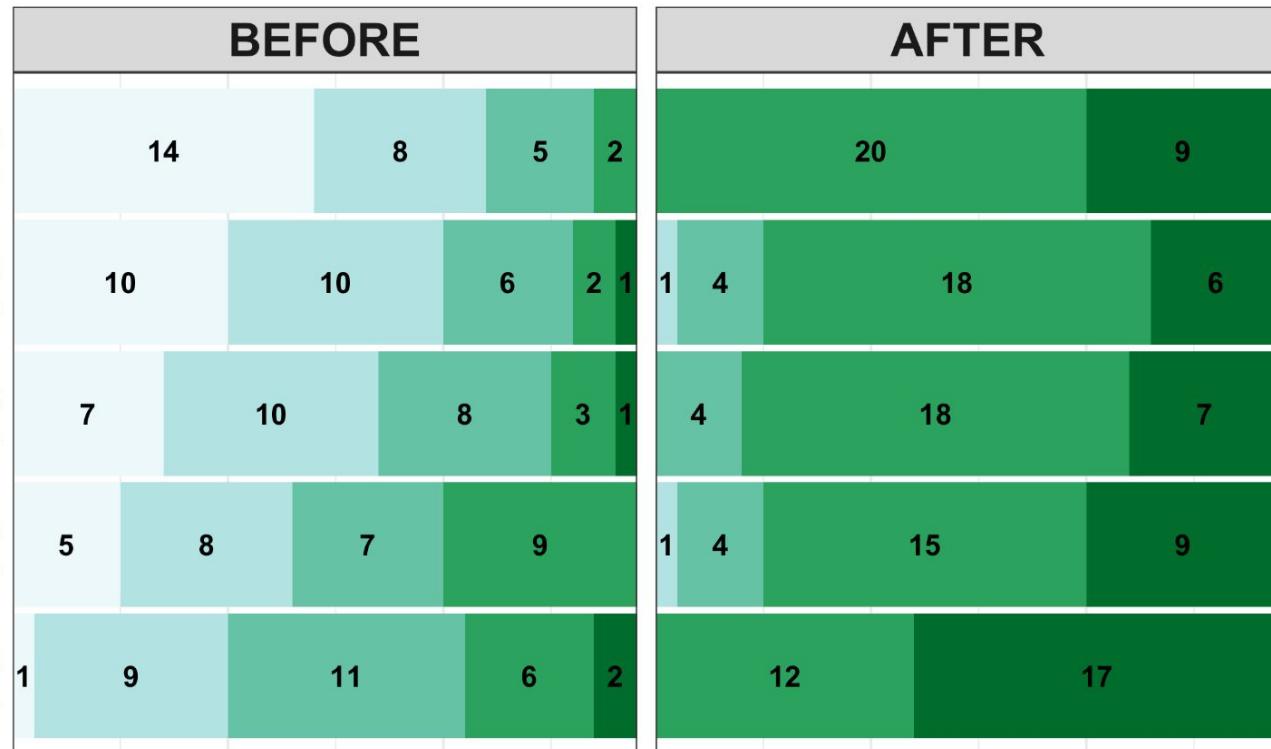


Data Science Skills cont.



Data Science Communication

Strongly Disagree Disagree Neither Agree nor Disagree Agree Strongly Agree

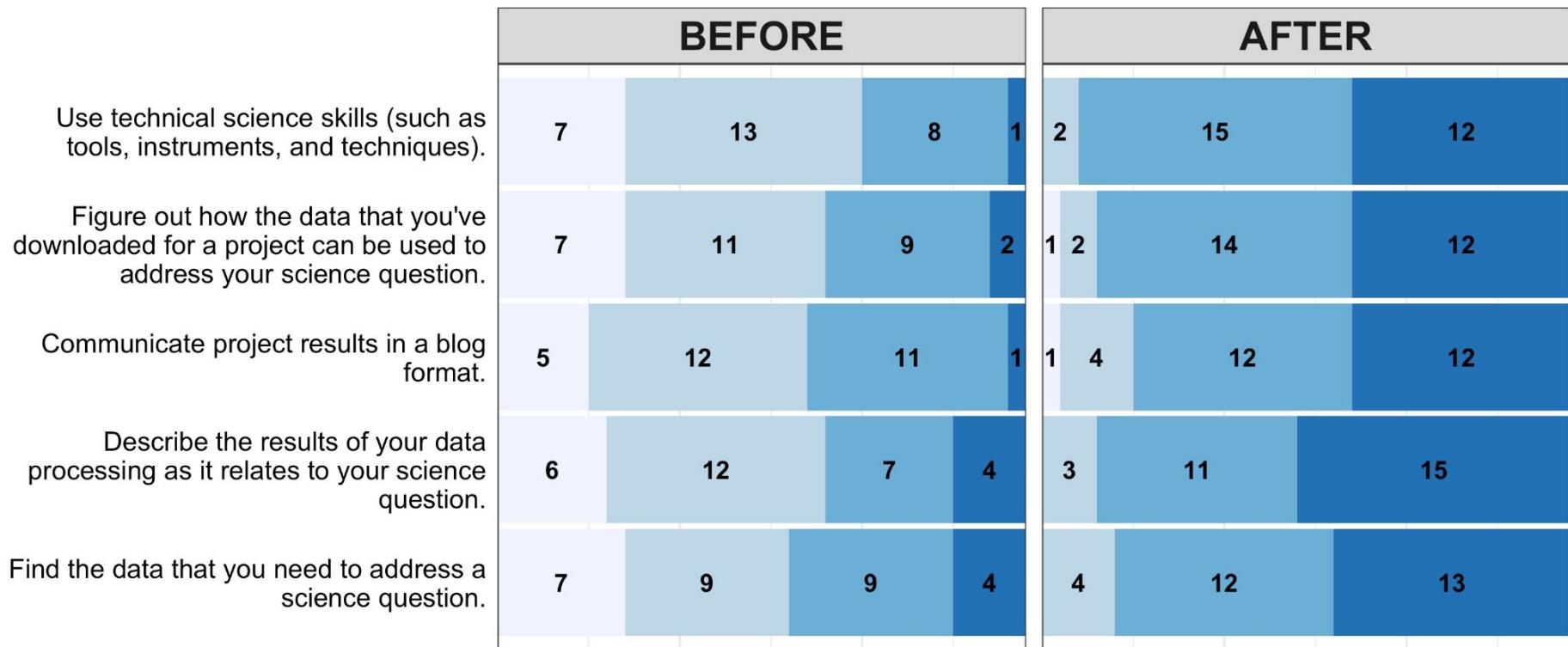


Data Science Communication cont.

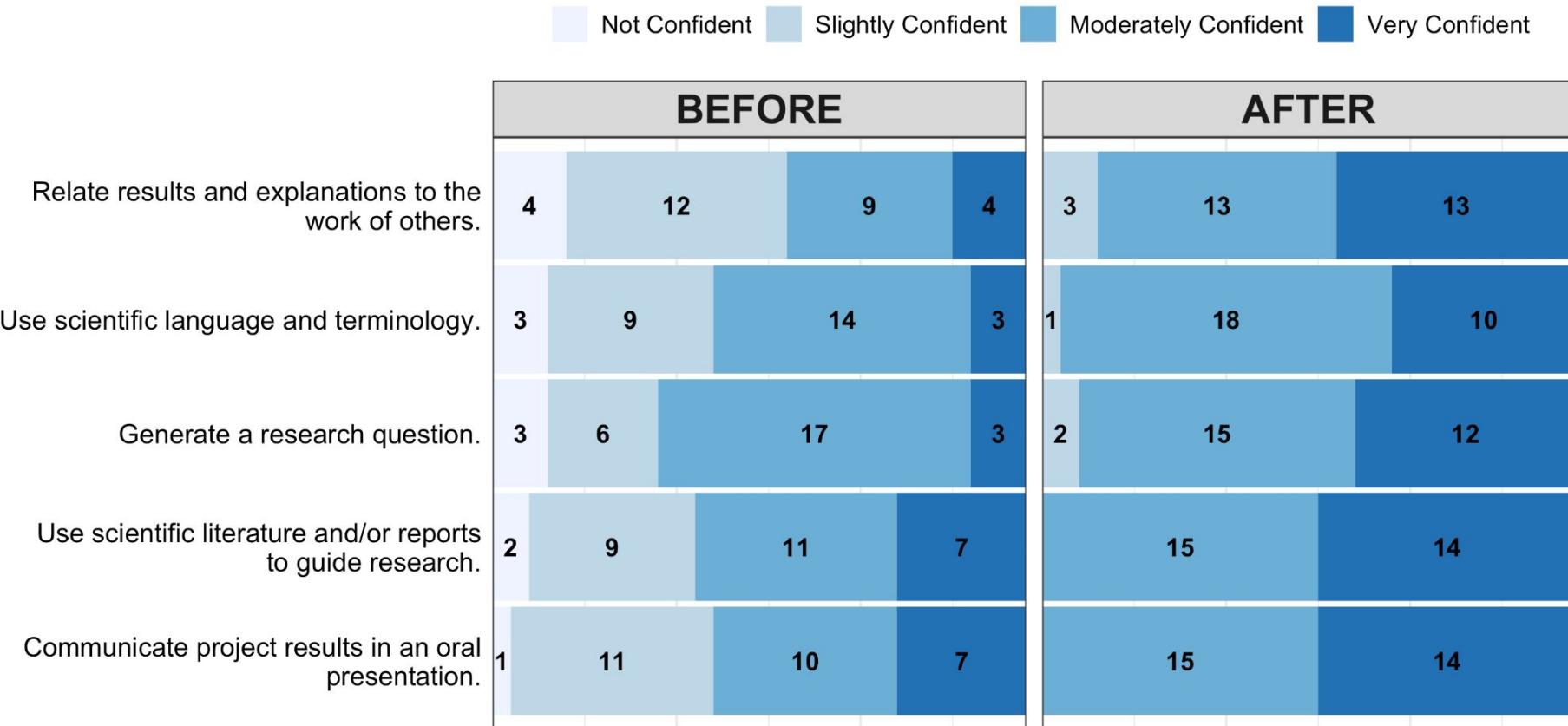


Data Science Practices

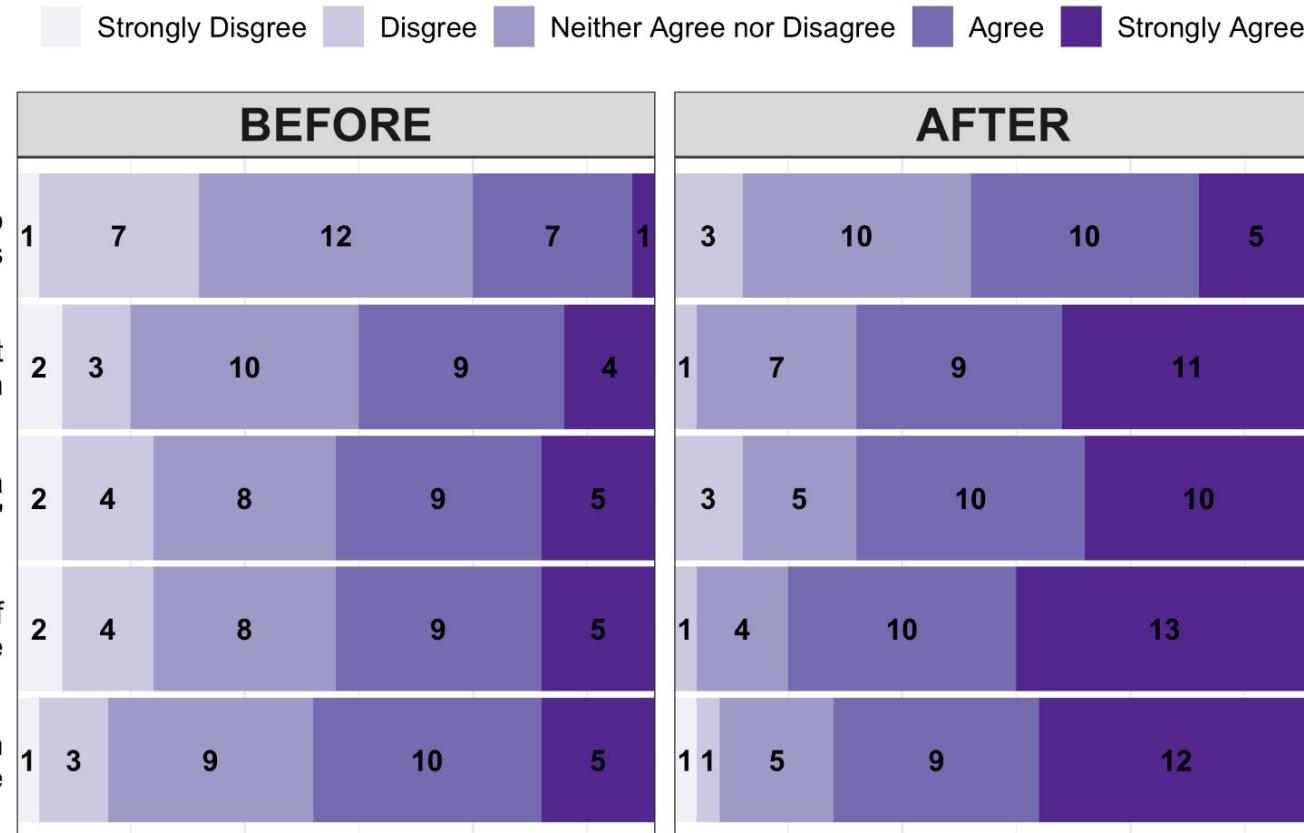
Not Confident Slightly Confident Moderately Confident Very Confident



Data Science Practices cont.



Science Identity





**Building capacity to teach
and learn technical data
skills for diverse student
populations**

Modeled after Earth
Lab's EDSC Program



PARTNER INSTITUTIONS



METROPOLITAN
STATE UNIVERSITY
OF DENVER



WHO?

**Undergraduate
students who
are....**

- New to data science
- Interested in earth and environmental science
- Previously limited to earth data science educational resources



- Attending one of our partner institutions

WHY?

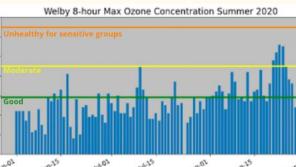
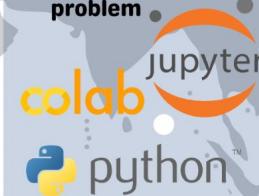
Meet demand for earth data scientists in workforce + Increase diversity in STEM + tech sector



- Make technical skills and educational resources available to students and faculty from historically underrepresented groups

HOW?

- Students are recruited for a paid summer internship and core data science skills
- Students solidify skills by completing a real-world data problem



2020 Days above Good (0.059 ppm): 37
2020 Days above Moderate (0.075 ppm): 4
Average: 0.057 ppm

A group from our 2021 EDSC cohort studied air quality data and transportation-related emissions in Denver, CO. ESIL STARS students can expect to participate in similar, student-designed projects.

ESIIL Stars



- Modeled after NSF-supported **HDR Earth Data Science Corps**
<https://earthlab.colorado.edu/edsc>
- **5-month** paid internship (March - August)
- **Build capacity** to teach and learn EDS using **Python**
- Tribal Colleges & Universities; Hispanic Serving Institutions
- **13** students; **6** faculty; **9** advanced interns (EDSC)



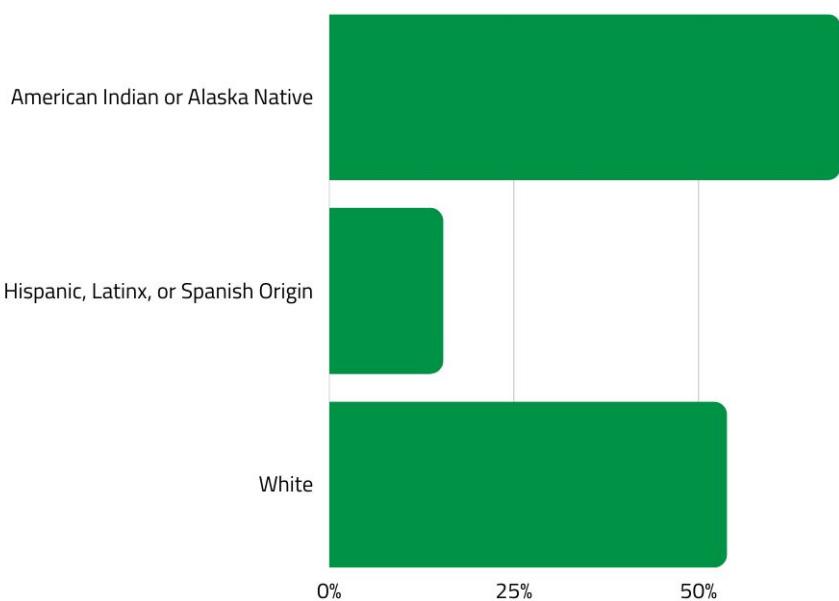
UNITED TRIBES
TECHNICAL COLLEGE



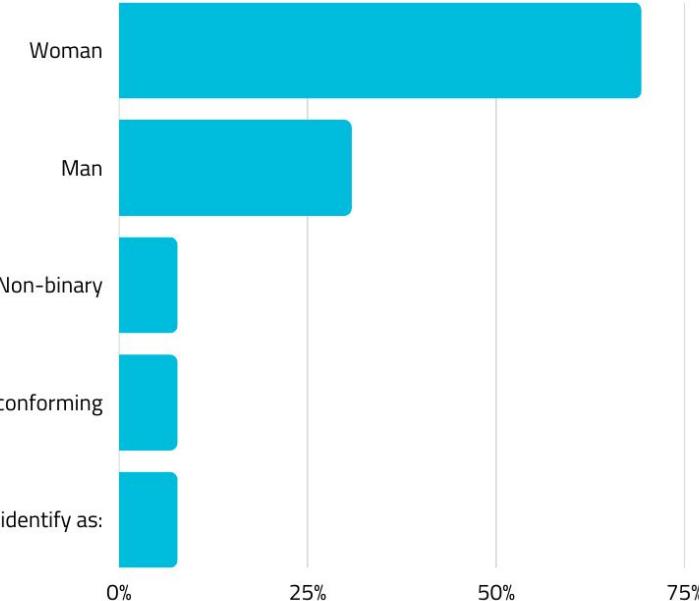
METROPOLITAN
STATE UNIVERSITY
OF DENVER

ESIIL Stars: Student Demographics (n = 13)

Racial and Ethnic Identity



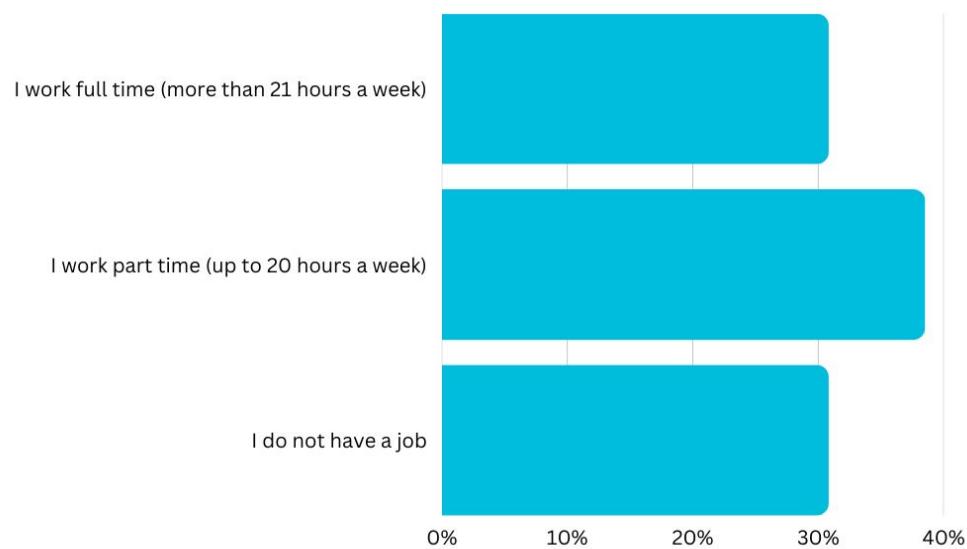
Gender Identity



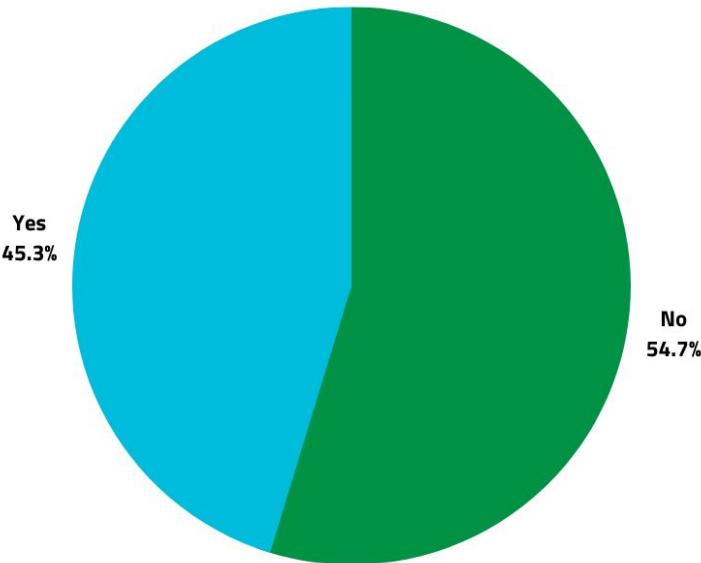
ESIIL Stars: Student Demographics (n = 13)



Employment Status



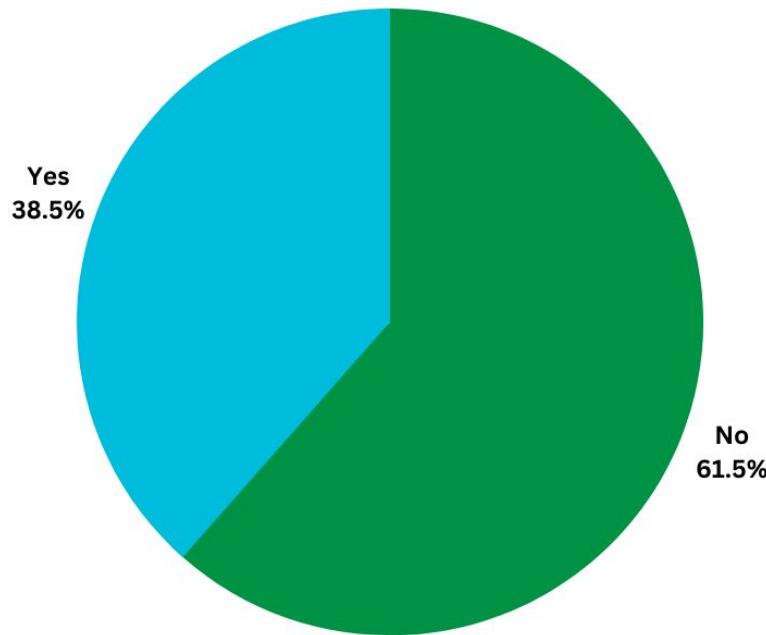
Are you a first generation college student (meaning neither of your parents or guardians completed a Bachelors degree)?



ESIIL Stars: Student Demographics (n = 13)



Do you identify as LGBTQ+?



LGBTQ+ includes (but isn't limited to) Agender, Aromantic, Asexual, Bisexual, Gay, Gender fluid, Gender non-conforming, Genderqueer, Intersex, Lesbian, Non-binary, Pansexual, Queer, Questioning, Trans, and Two spirit.

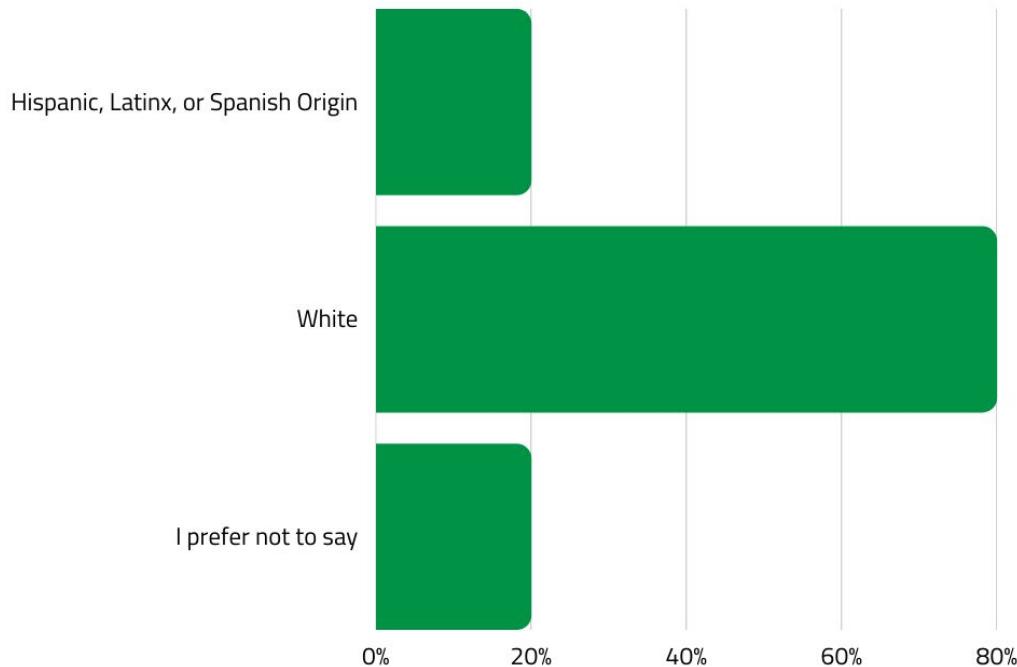
ESIIL Stars: Student Demographics

- 100% of students are not a member of the Armed Forces
- 23.1% of students identify as having a disability or being neurodiverse
- 76.9% of students are full-time, 23.1% are part-time
- 61.5% of students have other commitments outside of a job and school but do not take a reduced number of course credits, 15.4% of students have other commitments that require them to take a lower number of credits than they would ideally take

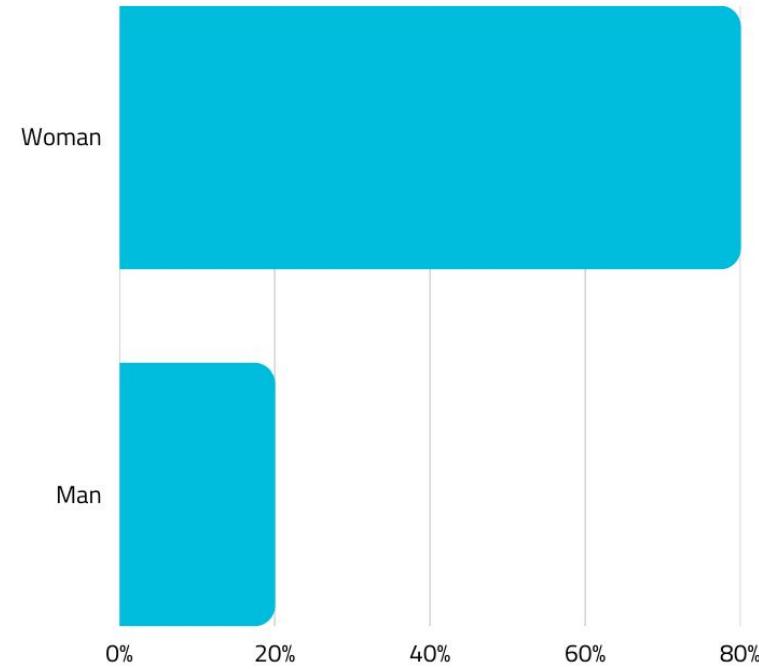
ESIIL Stars: Faculty Demographics (n = 5)



Racial and Ethnic Identity



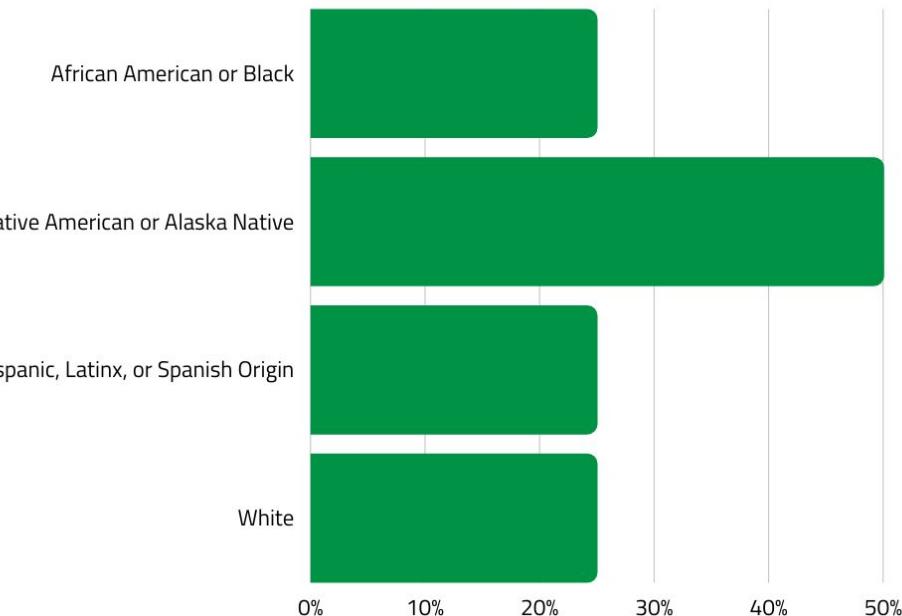
Gender Identity



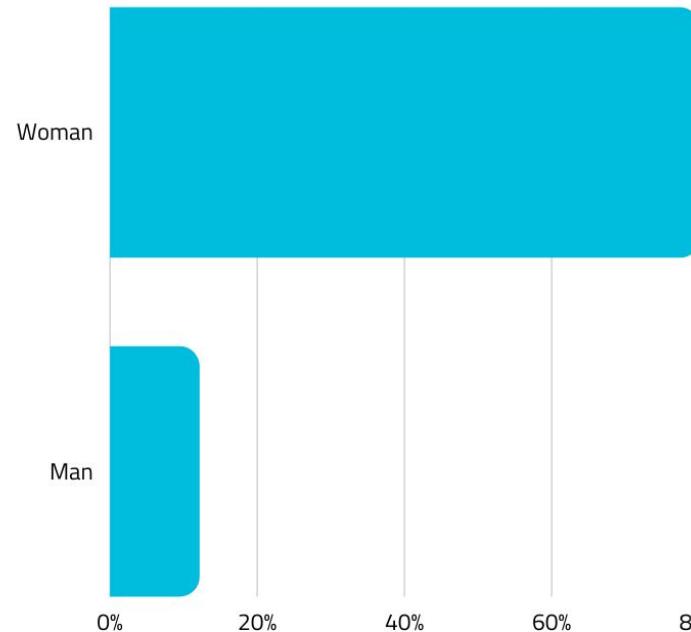
ESIIL Stars: Advanced Intern Demographics (n = 8)



Racial and Ethnic Identity



Gender Identity

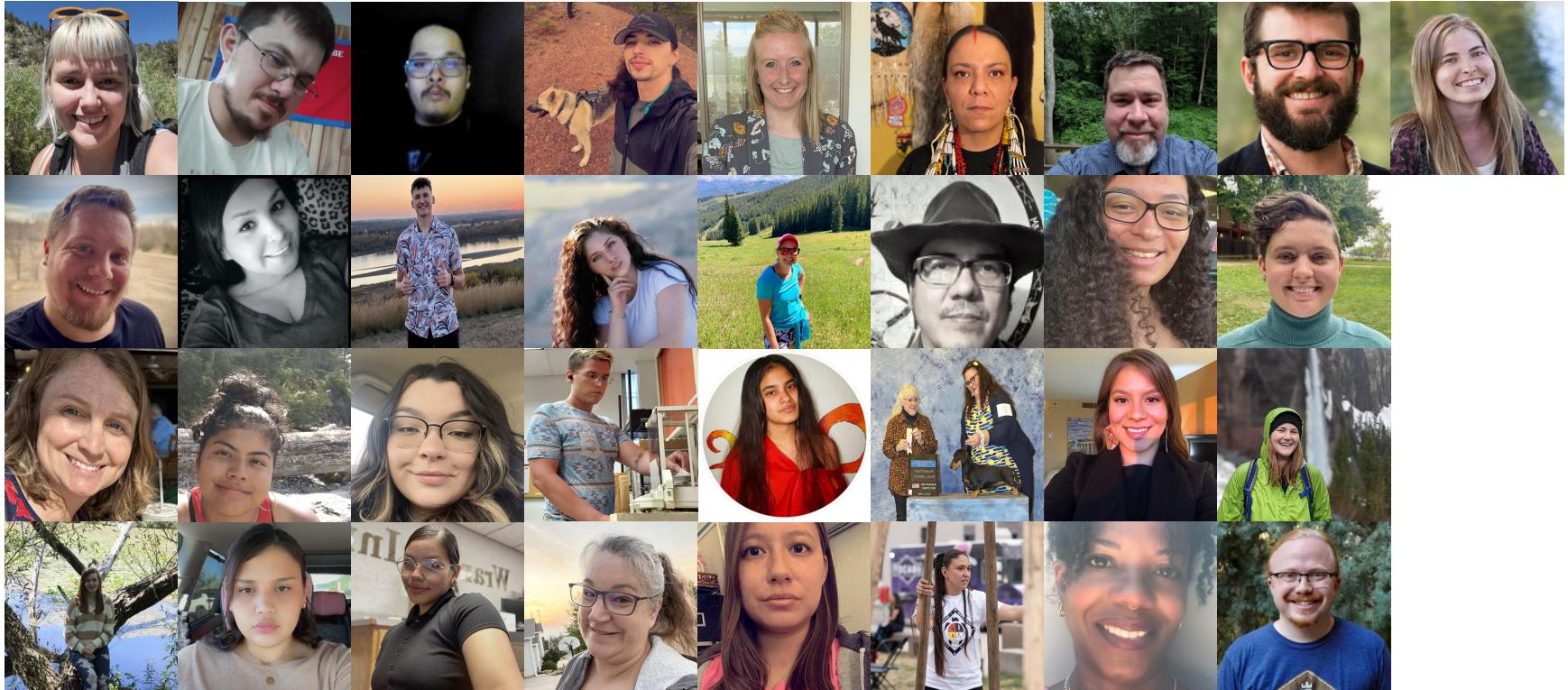


ESIIL Stars: Faculty Demographics



- 0% of faculty identify as LGBTQ+
- **80%** of faculty are not members of the Armed Forces, 20% preferred not to answer the question.
- 0% of faculty identify having a disability or neurodiverse.

ESIIL Stars



ESIIL Stars: Curriculum



GitHub: Classroom, Codespaces, Pages



Syllabus:

<https://cu-esiil-edu.github.io/esiil-stars-syllabus-2023/>

Notebooks:

<https://github.com/cu-esiil-edu/stars-curriculum>

Textbook: <https://www.earthdatascience.org/>

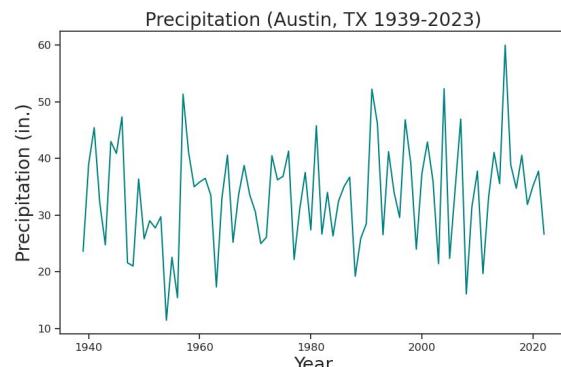
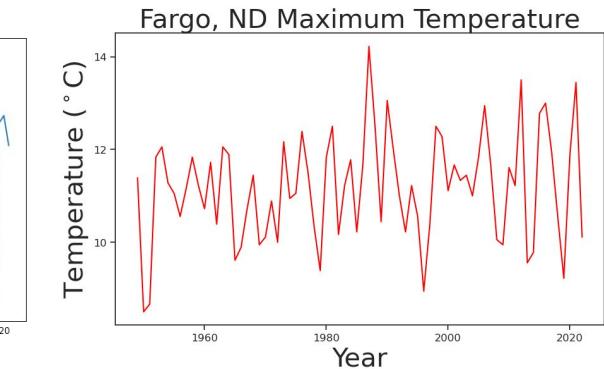
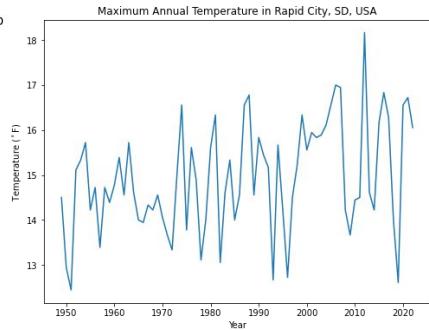
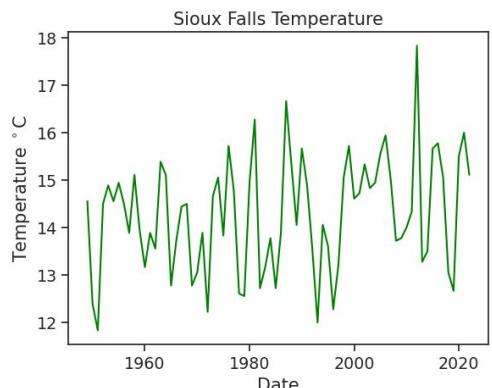
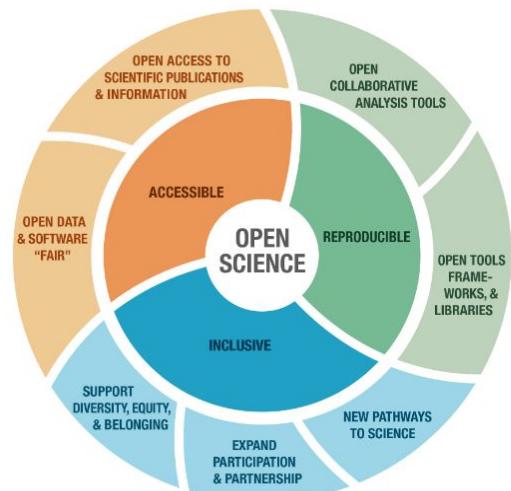


ESIIL Stars: Curriculum



Get started with open reproducible science!

Open reproducible science makes scientific methods, data and outcomes available to everyone. That means that everyone who wants should be able to **find, read, understand, and run** your workflows for themselves.



<https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/>

ESIIL Stars: Curriculum



In March of 2019 there were floods in South Dakota, USA

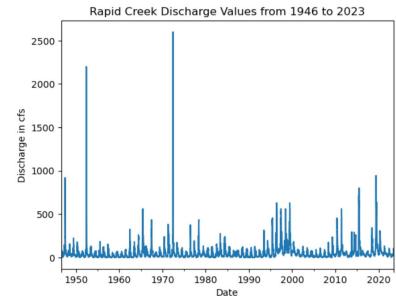
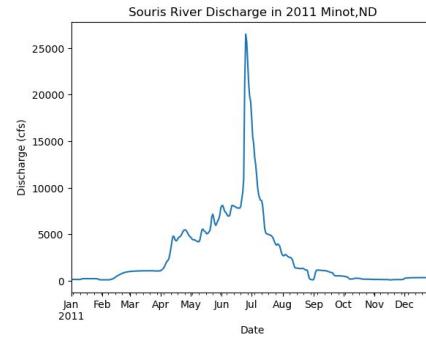
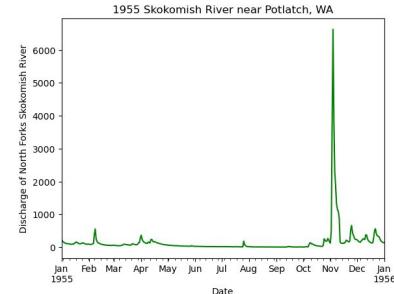
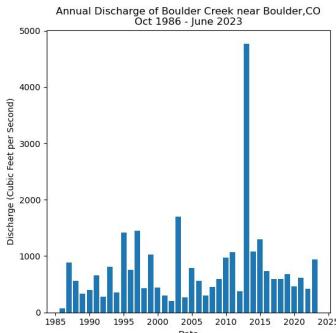


Image source: The Intercept April 5, 2019

In March 2019, large parts of South Dakota were flooded for weeks. What happened to cause this flooding? What impacts did the flooding have? Before we look at data about the flooding, we need to check out what other sources are saying about it.

Here are some resources from different sources to get you started:

<https://waterdata.usgs.gov/nwis/>



ESIIL Stars: Curriculum

Reclaiming Water Rights on the Gila River

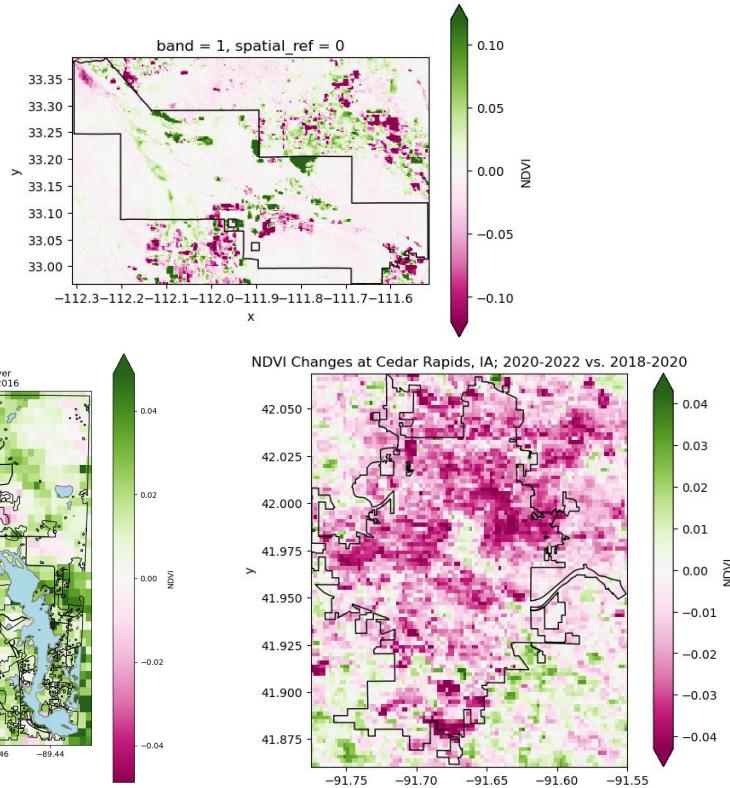
The Gila River Reservation south of Phoenix, AZ is the ancestral home of the [Akimel O'otham and Tohono O'odham tribes](#). The Gila River area was known for its agriculture, with miles of canals providing irrigation. However, in the 1800s, European colonizers upstream installed dams which cut off water supply. This resulted in the collapse of Gila River agriculture and sky-rocketing rates of diabetes and heart disease in the community as they were forced to subsist only on US government surplus rations.

In 2004, the Gila River community won back much of its water rights in court. The settlement granted senior water rights nearly matching pre-colonial water use. Work has begun to rebuild the agriculture in the Gila River Reservation. According to the Gila River Indian Community, "It will take years to complete but in the end the community members will once again hear the sweet music of rushing water."



Image source: New York Times

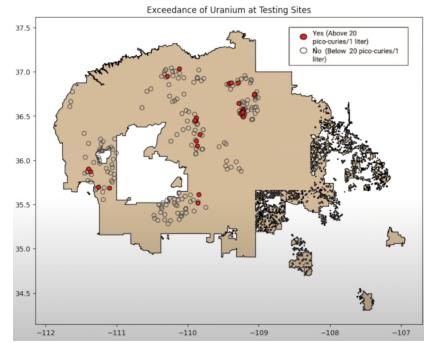
<https://modis.gsfc.nasa.gov/data/dataproducts/mod13.php>



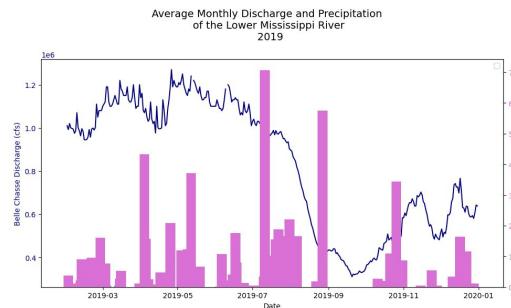
ESIIL Stars - Final Projects



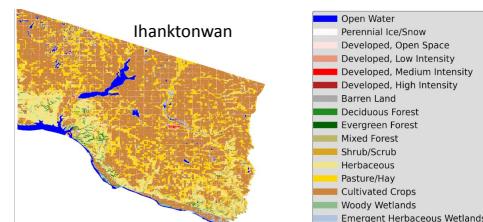
- a.) Water Quality, Quantity, and Accessibility on the Diné (Navajo) Reservation - **MSU Denver**
i.) <https://youtu.be/l8ltH6REyzo>



- b.) Relationships Between Precipitation and Discharge on Mississippi River Tributaries - **United Tribes Technical College**
i.) <https://youtu.be/U-p7johQTu8>



- c.) Habitat Suitability Modeling of Thíŋpsinla (timpsula, tinspila, prairie turnip) - **Oglala Lakota College**
i.) https://youtu.be/4R3H_Ump0aq



ESIIL Stars - Final Projects



- a.) Leedle, C., Sheldon, S., Patrick, R., Smedes, M., Jacquez, N., Johnson, M., & Rohlehr, K. (2023). Water Quality, Quantity, and Accessibility on the Diné (Navajo) Reservation (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.24162888.v1>

- b.) Davis, T., Sanchez, S., Stroh, K., Alberts, B., Biggane, E., & Logan, J. (2023). Flowing America (Version 2). figshare. <https://doi.org/10.6084/m9.figshare.24143079.v2>

- c.) Weddell, E., Landreaux, A., Dupree, S., DuBray, M., Yellow Thunder, E., Bagola, L., Phelps, M., White Eyes, J., Phelps, T., Griffith, C., & Gehring, D. (2023). Habitat Prediction Modeling of Tinpsila on South Dakota Reservations (Version 2). figshare. <https://doi.org/10.6084/m9.figshare.24143073.v2>