

# The DataSquad Experiment: Some Lessons for Building RSE Capacity

Paula Lackie - DataSquad founder &  
Academic Technologist @ Carleton College

Elliot Pickens - DataSquad Alum &  
PhD student @ the University of Pennsylvania



# A CREATIVE SOLUTION TO AN INTERCONNECTED PROBLEM - THE DATASQUAD ORIGIN STORY

*Cleaning, Wrangling, Sculpting,  
& Visualizing data since 1993*



## DataSquad

**Carleton College**

*for the love of the pipe*

*“Our aim is to make your research and  
work life easier!”*

- Started in the 1990's at Carleton College
- 2014 DataSquad was born
- Gives students valuable work & life experience

Carleton ITS DataSquad ([carletondatasquad.bitbucket.io](http://carletondatasquad.bitbucket.io))

*With Thanks to O'RIELLY® for a generous copyright allowance on their book motif.*

### Assistant Data Scientist

- Entry level - basic code
- Task oriented
- Seeking any experience that will help them get an internship

### Lead Data Scientist

- Aspiring professional
- Encouraged to parse projects into tasks
- Manage communications
- Verify project *completion*



### Technical Writer

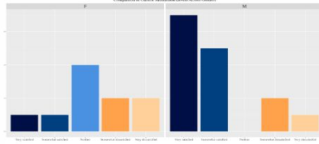
- Entry Level - no code required
- Interested in programming
- Often looking to move into other roles
- Begin work by documenting ongoing projects
- Seeking work experience

### Project Management Intern

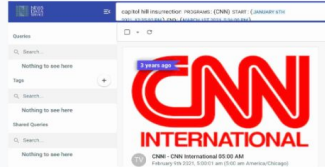
- Aspiring professional
- Interested in managerial experience
- Helps keep expectations realistic



# Example Projects from Carleton College:



**Data Visualization**  
*Sociology and  
Anthropology Department*



**Data Scraping**  
*Political Science  
Department*



**File Management for ITS**  
*ITS*



**Labeling and Sorting Alumni Images**  
*Office of Alumni Relations*

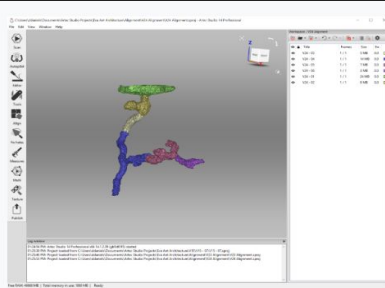


**Converting Stata to R**  
*Economics Department*



**Project Workflow**  
*DataSquad*

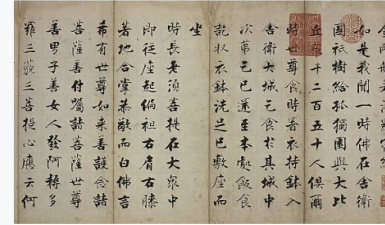
# Example Projects from UCLA:



**Digitalizing Ant Nests  
to Study Tunneling  
Trends**



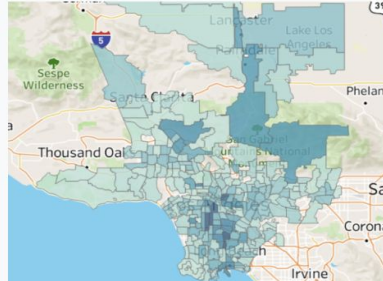
**Squirrels in Motion**



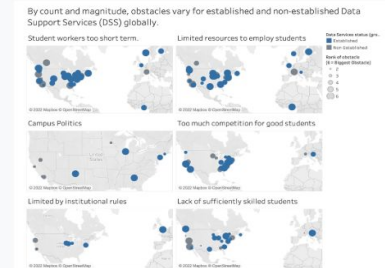
**Analyzing Ancient  
Chinese Buddhist Text  
with Python**



**Como-Q - Your Skincare  
Assistant**



**Civilian Deaths at Hand  
of Police**



**DataSquad  
International**

# Why does this matter?

- Two main criteria for employers seeking to fill data-related roles – wide range of skills AND prior experience
- Lack of diversity in data-related work –
  - 10-15% of data scientists and machine learning researchers were women
  - 80% of Artificial Intelligence professors are male
- New solutions are needed that can offer opportunities to a more diverse pool of students to learn and practice the hard and soft skills needed in their future employment

# Who are DataSquad members?

- Over 95 students have joined and participated in the DataSquad (typical size is ~10-13 students)
- Diverse cohorts –
  - 68% non-white (vs. 30% in college population)
  - 30% non-male (including non-binary) (vs. 50% in college population)
  - 59% come from outside the US and many of the US-students are first generation Americans. (vs. 11% in college population)
- Most come in as computer science majors
- Other disciplines include: Cognitive psychology, Linguistics, Mathematics and from less expected disciplines including Music, Dance, History

# Some key learnings...

## 1. Students are young versions of us

- *without the benefit of life experience - be direct - be kind -be honest*

## 2. Hire carefully - hire creatively

- *use referrals from your successful students.*

## 3. Provide a “promotion track”

- *Something that shows a progression in responsibility*

## 4. Have them sign a confidentiality agreement

- *& teach them (by example) to follow it.*

- *Regularly remind them what “Confidentiality” means in practice*



# Some key learnings...

## 5. Have a mission statement

- & teach them (by example) to follow it.

## 6. Have a job manual

- perpetually call it a “draft” and engage with them to keep it real.

## 7. Follow a prescribed onboarding process

- because the work is so freeform, this helps provide structure

## 8. Hire a Student PROJECT MANAGER !!!

- mentor this student to help you keep the group on track



## **Dexter Kong** '26 - Project Management Intern

"Being a part of the DataSquad has been an invaluable part of my college experience and leveled up many of my skills, with emphasis on the ones I was lacking in. For example, having this environment, mentorship, and team showed me how to communicate both as an individual through meticulous documentation and as a team player through meaningful updates."

"The same way reading about how to swim won't make you float, no amount of classroom 'assignments and reflections' would have ever helped me the way that DataSquad has."



**Auiannce Euwing '26** Assistant Data Scientist

**“Through this experience, I’ve learned to think more critically and gained a tremendous boost in confidence. As a Black woman in computer science, I’ve sometimes doubted my abilities, but DataSquad has shown me that I truly belong in this field and that I have the technical skills to back up my worth.”**



## Mustafe Osman '26

“In my internship last summer at Franklin Templeton, I saw firsthand how the ability to turn complex data into actionable insights; something I first practiced at DataSquad can directly influence high-stakes decisions in global markets.”





**Oscar Smith '20**

“The DataSquad requires planning for future maintenance, documentation, and communication: some of the most important aspects of programming.”

**Jon Bisila '18**

“Working on the DataSquad gave me **real-world experience** that I couldn't get in class. With projects that extend beyond the usual term limits, I really got to dig in, this gave me both the satisfaction *and realistic disappointments* of working as a programmer.”





## Dashiell Coyier '29

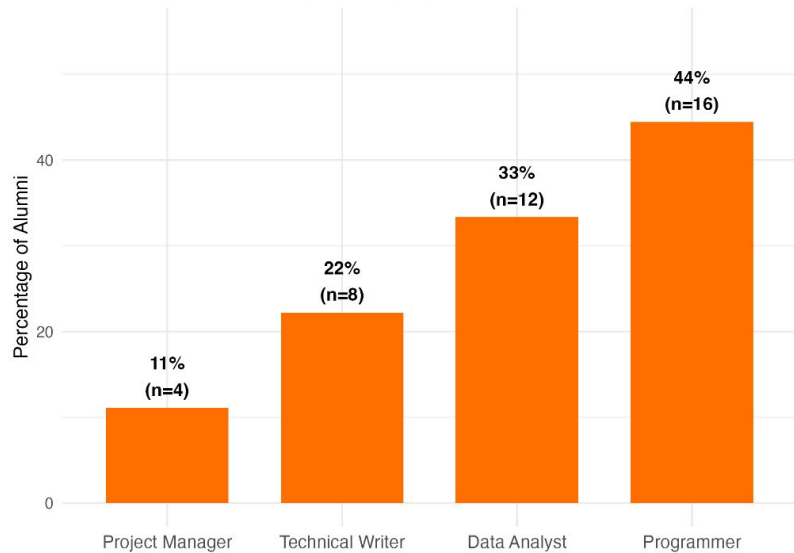
"I have been on the DataSquad for all of three weeks! I joined the team as a technical writer, and in this fast-paced environment, I have already learned the importance of both careful documentation and consistent communication. Asking questions on the Slack channel has been a lifeline when navigating this new environment."



## The DataSquad Experience: Diverse Roles and Responsibilities

### DataSquad Role Participation

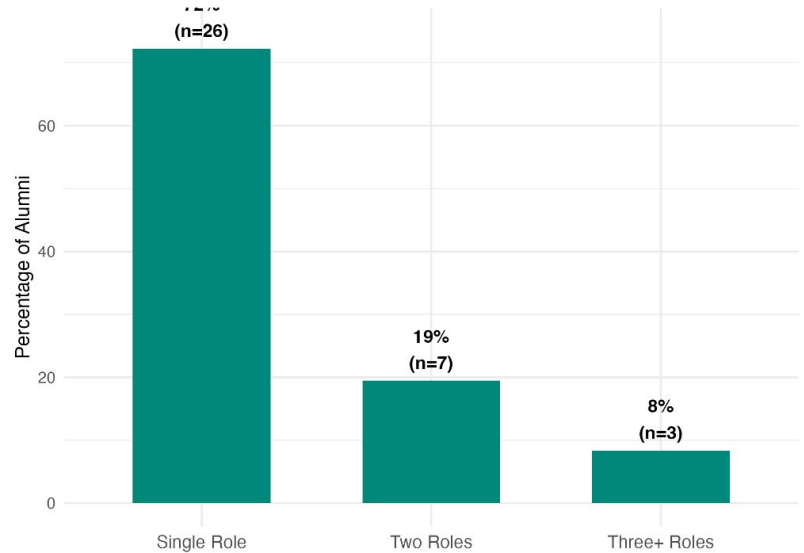
Students can hold multiple roles simultaneously



Percentages sum to >100% because students can have multiple roles

### Role Diversity: Many Students Hold Multiple Roles

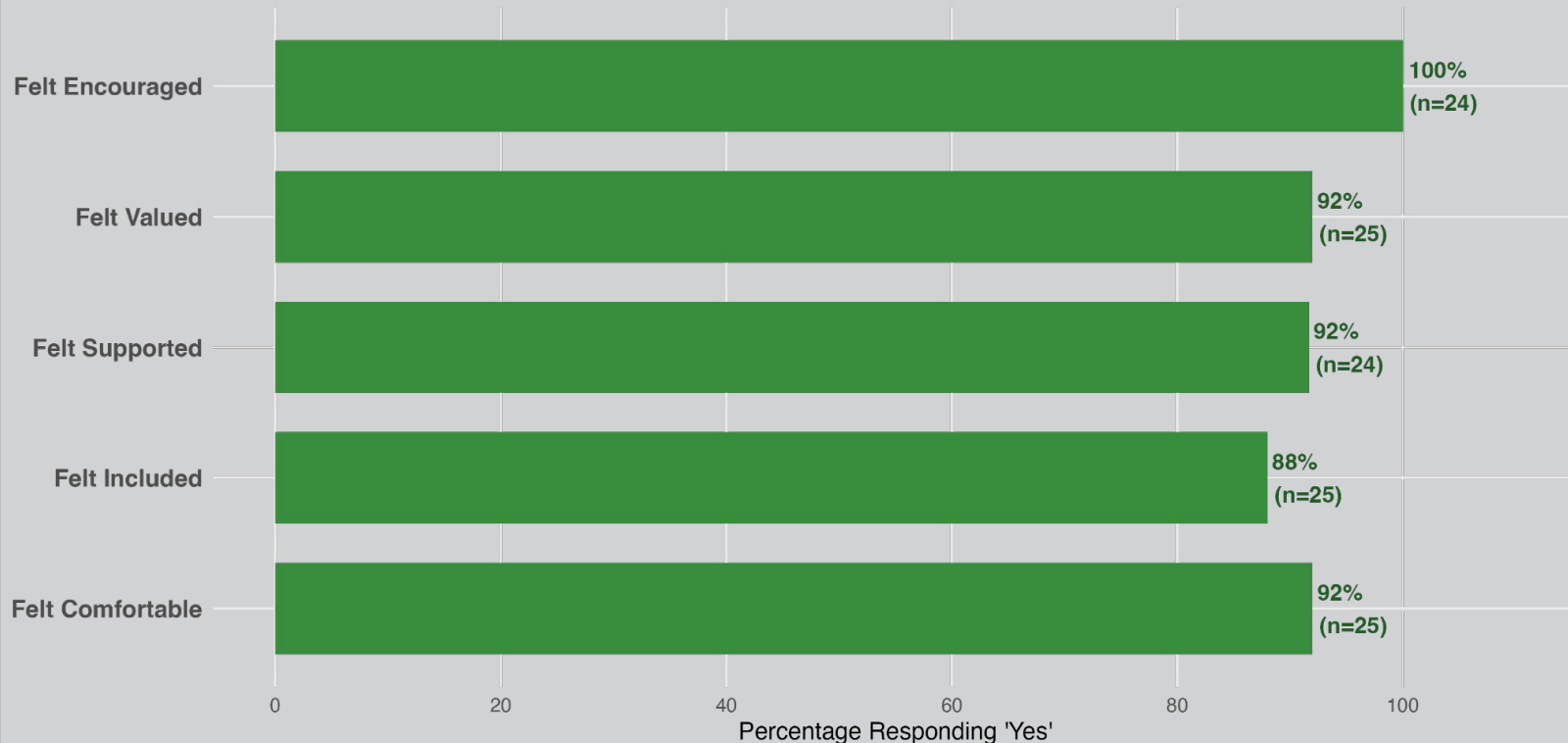
Distribution of role complexity among alumni



Multi-role experience may enhance breadth of skill development

## DataSquad Learning Environment: Exceptionally Positive

Percentage of alumni responding 'Yes' to each dimension

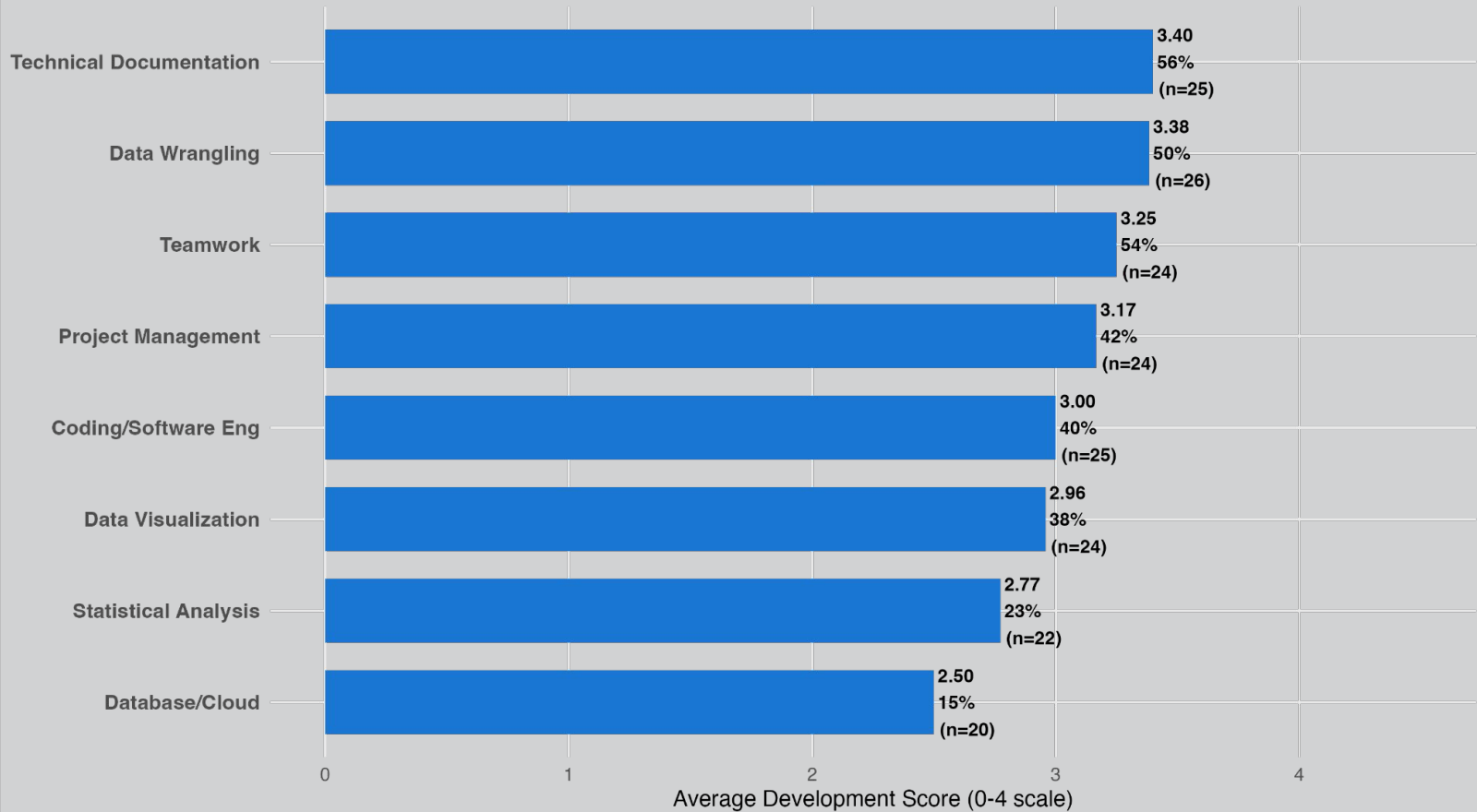


All dimensions show >88% 'Yes' responses, with 100% reporting feeling encouraged



## DataSquad Skills Development: Overall Alumni Ratings

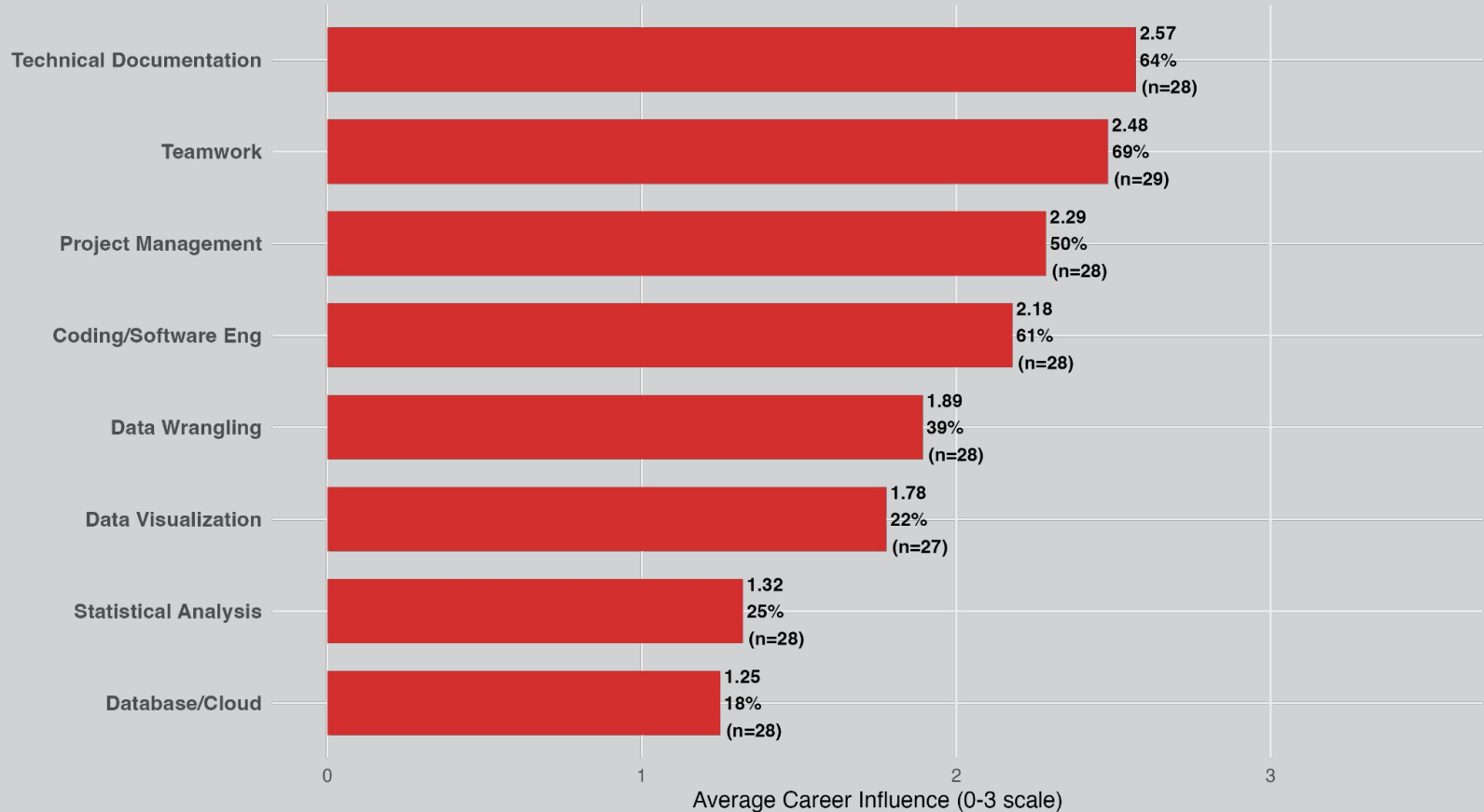
Which skills do alumni report developing most through DataSquad?



Numbers show: Average score, % reporting 'Yes, a Lot!', and sample size  
Higher scores indicate greater skill development

## Career Impact: Which DataSquad Skills Matter Most?

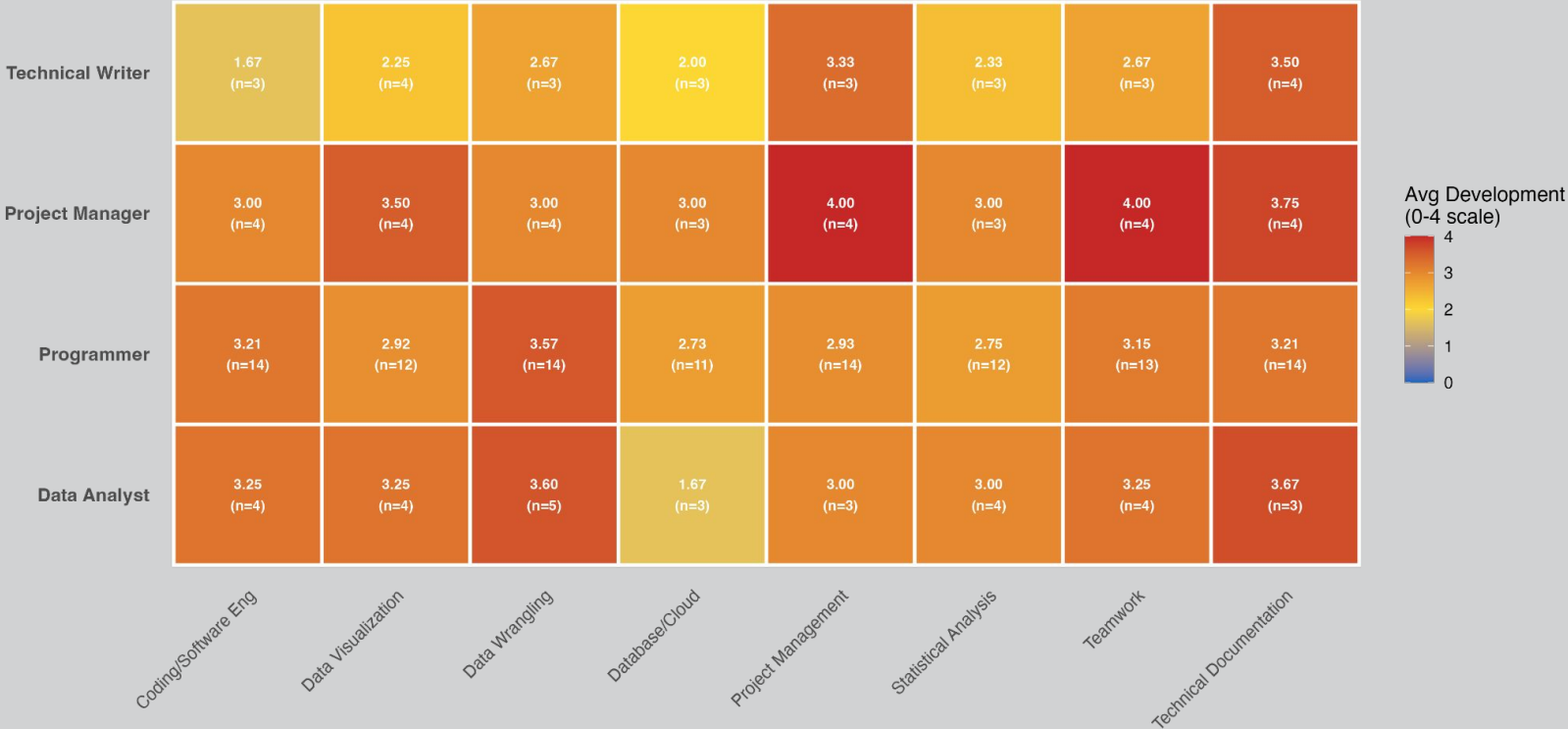
Alumni ratings of how influential each skill has been in their careers



Numbers show: Average influence, % reporting 'Very Influential', and sample size  
0=Not Applicable, 1=Not Influential, 2=Tangentially Influential, 3=Very Influential

# Skills Development by Role: Different Paths to Success

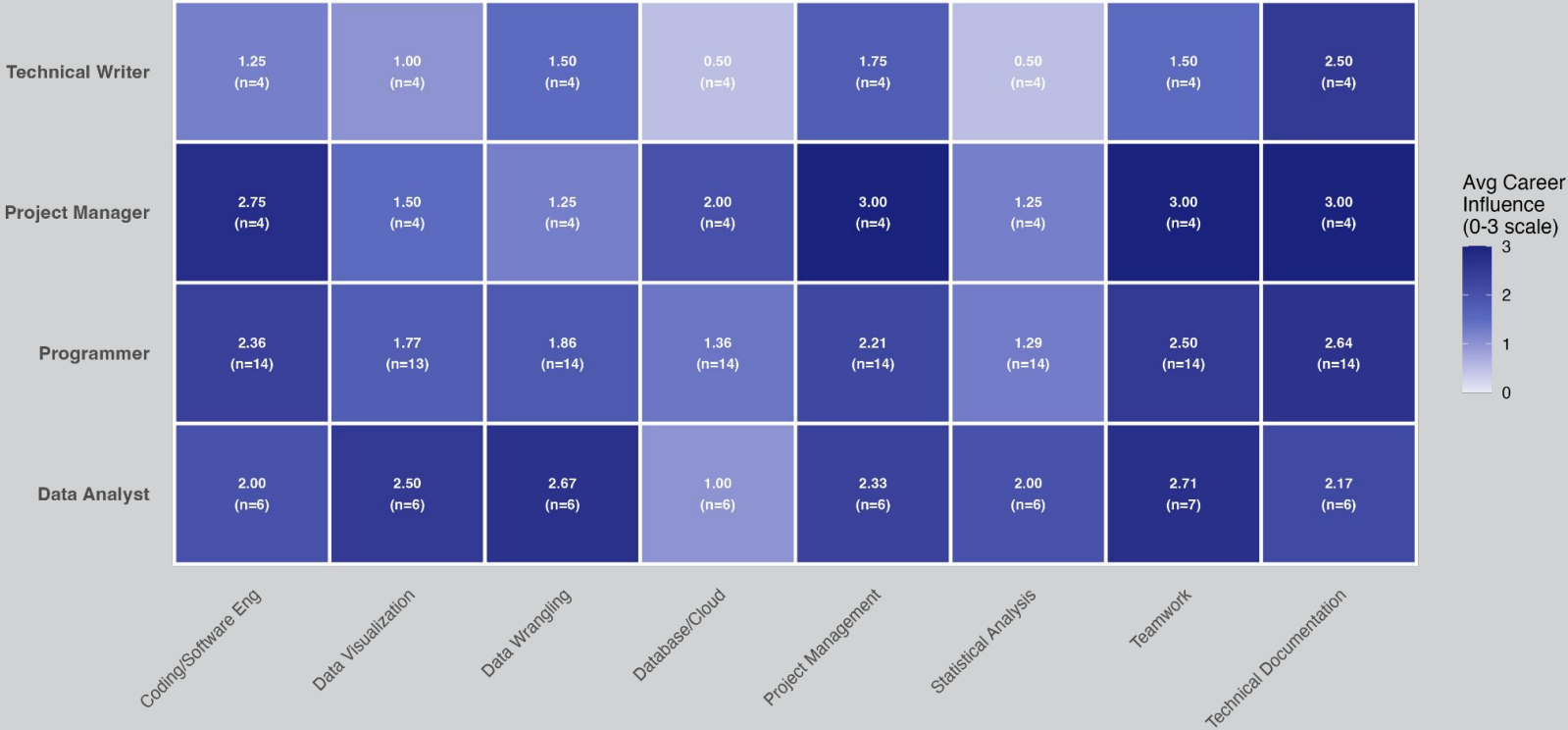
How different DataSquad roles develop different skill sets



Darker red = higher development. Numbers show average score and sample size.

# Career Impact by Role: Which Skills Matter for Whom?

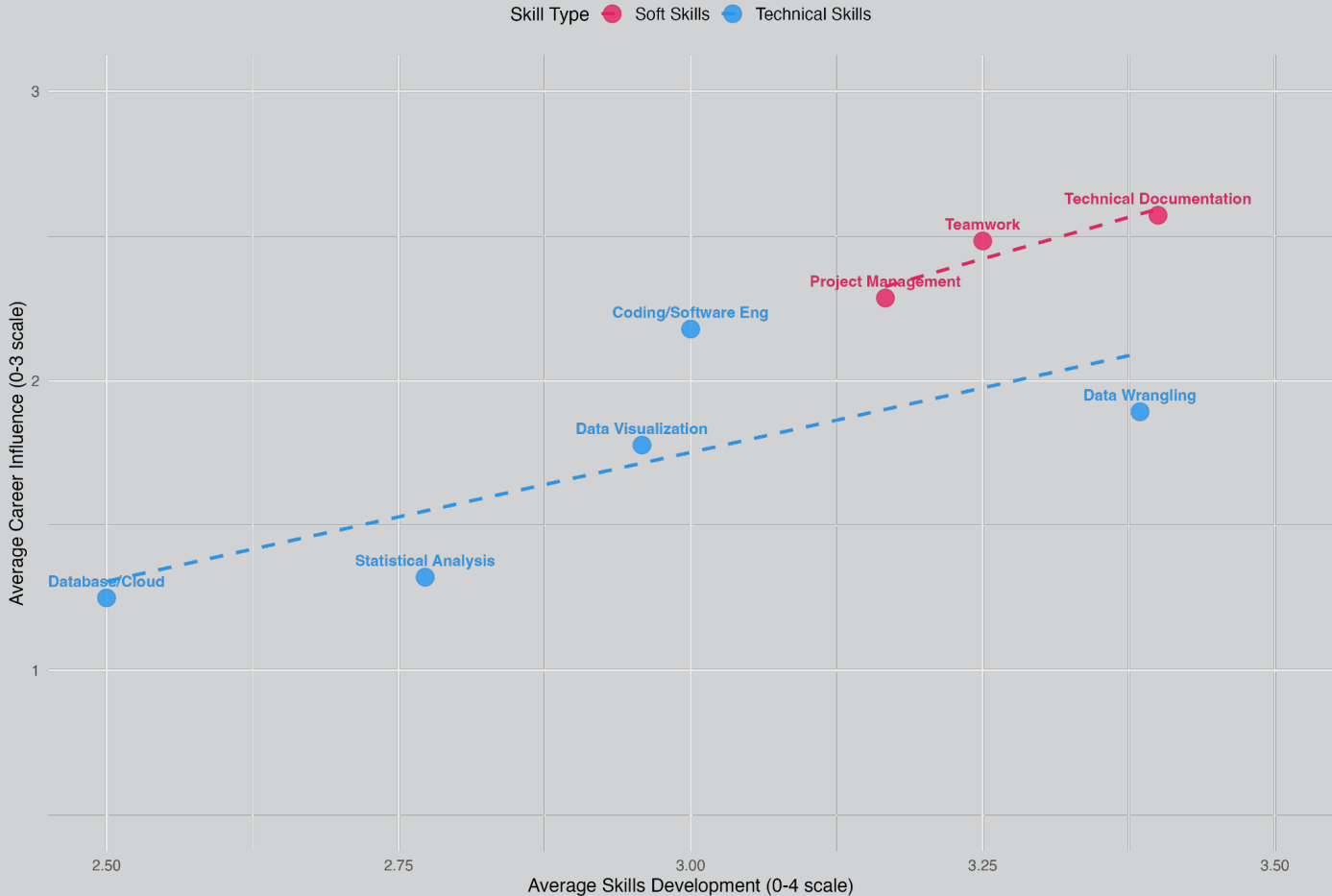
Long-term career influence of DataSquad skills across different roles



Darker blue = higher career influence. Numbers show average score and sample size.

# DataSquad ROI: Skills Learned vs Career Impact

Do skills developed in DataSquad translate to career influence?



## CAN THE DATASQUAD MODEL BE EXPANDED TO MEET THESE CHALLENGES ELSEWHERE?

- 2021 – launch of a new [DataSquad at UCLA](#)
- Looking at developing blueprint for developing and supporting the distribution of the highly flexible DataSquad model (*aka: DataSquad International*)
- Must recognise the potential barriers to this!
  - Managing a DataSquad takes time & effort!
  - Takes a cultural shift - *what is the mandate?*

# Thank you!

Here are links to some previous presentations:

- [DataSquad-Testimonials-Lackie-IASSIST19](#) (for more from students and clients)
- [LIBER-DataSquad International-2022](#) (Europe's research library community)
- Sample documents: [PublicFolder-Managing Student DataSquad-IASSIST19](#)
- DataSquad related offerings from the [IASSISTdata Youtube channel](#)

