California Education Learning Lab

2022 INSPIRE CONVENCING

THE NEXT GRAND CHALLENGE: DATA SCIENCE WORKSHOP

October 14, 2022
Introduction to Data Science (IDS) for High School Students

Suyen Machado
Director, Data Science Education Project
Co-author, Introduction to Data Science
University of California, Los Angeles
Collaborating Organizations

IDS

Introduction to Data Science

NSF

mobilize

transforming public schools ucla center x

los angeles unified school district

ucla department of statistics
2013: 1st Data Science A-G Approved Course
2014: Pilot at Los Angeles Unified
2017: Adoption in Other Districts
2020: Program Status
IDS Components

Introduction to Data Science

Robert Gould
Suyen Machado
Terri Anna Johnson
James Molyneux

Web Tools

Year-long Curriculum + Technology Suite + PD
Retrieved from:
https://nasaeclips.arc.nasa.gov/teachertoolbox/the5e
The Data Science Test

- Awash in Data
- Data Moves
- Data Properties

https://bestcase.wordpress.com
LESSONS

PROJECTS

LABS
Introduction to Data Science

IDS Curriculum

Unit 1 - Data and Visualizations
Unit 2 - Distributions, Probability, and Simulations
Unit 3 - Data Collection Methods: Traditional and Modern
Unit 4 - Predictions and Models

Guidelines for Assessment and Instruction in Statistics Education (GAISE)

Computational Thinking Standards
The Data Cycle

1. Ask questions
2. Consider data
3. Analyze data
4. Interpret data

Dr. Rob Gould
Lead IDS Author & UCLA Statistics Professor
IDS Technology

Web Tools

R Studio

Participatory Sensing
PARTICIPATORY SENSING

An approach to data collection and interpretation in which individuals, acting alone or in groups, use their personal mobile devices and web services to systematically explore interesting aspects of their worlds, ranging from health to culture.
SCALE Immersion Model for Professional Learning (SIMPL™)

Teacher’s Lens - intro
Student Lesson - intro
Student Lesson - body
Student Lesson - reflect
Teacher’s Lens - reflect

IDS Professional Development

Year 1
9 days
IDS Implementation
Deeper dive into: R Skills, Concepts, Pedagogy

Year 2
4 days
Welcome to the IDS Teachers' Community of Practice!
A repository for sharing resources, best practices, information,

IDS Professional Learning Community

Google Group
Bi-monthly Office Hours
Email ticket support

Introduction to Data Science - Best Practices Video Series
By IDS Admin - 1 post - 1 view
I teach AP Statistics and IDS. I’ve noticed that IDS students get a better conceptual understanding of the p-value by performing shuffling in RStudio than my AP students do using the traditional method.

Brandon Hawks
IDS Teacher
Blue Valley Southwest High School
Kansas
I teach AP Statistics and IDS. I've noticed that IDS students get a better conceptual understanding of the p-value by performing shuffling in RStudio than my AP students do using the traditional method.

Brandon Hawks
IDS Teacher
Blue Valley Southwest High School
Kansas

To: support@idsucla.org

I got an email from a former student who is in college now (I'll include it below because y'all deserve credit for making such a great IDS program).

Former IDS Student's Email:

I'm just emailing you to let you know that your class really helped me in college. I chose to do stats for my math GE credit and thanks to your class, I'm ahead of the curve. I found that I am actually very good at statistics and data science and I miss learning R script. If you know any free websites where I can continue to learn it please let me know. I also wanted to let you know that you are an amazing teacher. Learning was fun and exciting in your class, you inspired me to continue learning data science and R script even if it's not my career goal or major. Keep on doing what you're doing.
Want to learn more?

http://www.introdatascience.org

@uclaIDS

info@idsucla.org
DATA SCIENCE AT UC BERKELEY
INSTITUTIONAL TRANSFORMATION

Anthony Suen
Director of Data Science Discovery
UC BERKELEY DATA SCIENCE EDUCATION STRATEGY

Leverage Deep Multidisciplinary Strengths

Scalability with No Barriers to Entry

Student Powered Social Impact
We incubate and accelerate data science research with academic, government, non-profit, and industrial partners.

Over 500 UC Berkeley student researchers participate in the program each year.

The end results have led to advancements in domains as diverse as medicine, disaster readiness, astrophysics, criminal justice reform, and archaeology.
My journey to data science at UC Berkeley

Interdisciplinary Focus

Cross Domain Policy Advocacy & Data

Startup Incubation
2013
BIDS Launched
2015
Data 8 Pilot of 100 students
(Now 2,000 students)
2016-2017
Interim Division of Data Science

- Data Science Discovery
- Connector Support
- Data Science Modules
- Data Peer Consulting
- Data Scholars
- Student Teams Model
2018
National Workshop on Data Science Education

Data Science Major
2019
UC Data Science Leadership Meeting
2020

Shared interest and momentum across workshop participants from California to build an alliance focused on statewide collaboration opportunities
Where are we now in 2022

Majority of UCB Students take a DS course, 3rd biggest major with 1000+ students

More than 500 students participate in over 150 impactful Discovery Projects

Building Education consortiums across California and US
Building on Student Energy

Student energy driving research and education innovation in the space across campus and beyond.
Thank you!

anthony.suen@berkeley.edu
CourseKata's Approach to Data Science: Using Data to Improve Data Science Education

Ji Y. Son, PhD, Cal State LA
# The Practice of Data Science

## Data Literacy
The goal: Informed citizen (someone who can understand a news article with data and summary statistics in it and think about it critically)

## Data Analytics
The goal: Uses data and statistics in service of a content area (e.g., biologist, economist, public health official, entrepreneur)

## Data Engineering
The goal: Leads in data science (e.g., data leader in an org, developing new statistical techniques) or specializes in some area (e.g., machine learning, data infrastructure)
The Practice of Data Science

DATA LITERACY
The goal: Informed citizen (someone who can understand a news article with data and summary statistics in it and think about it critically)

DATA ANALYTICS
The goal: Uses data and statistics in service of a content area (e.g., biologist, economist, public health official, entrepreneur)

DATA ENGINEERING
The goal: Leads in data science (e.g., data leader in an org, developing new statistical techniques) or specializes in some area (e.g., machine learning, data infrastructure)
After CA community colleges reduced remedial math courses (AB705), 50% of math enrollment is now in statistics courses.

OPPORTUNITY TO MODERNIZE STATS
OPPORTUNITY TO MODERNIZE STATS

A Modeling Approach

DATA = MODEL + ERROR

Connecting stats to modeling with algebraic functions

Interactive Textbook (R)

In-Class Jupyter Notebooks
OPPORTUNITY TO MODERNIZE
HOW WE STUDY
STUDENT LEARNING
The “Better Book” Model

Using student data to improve data science education.

Data are surprising!
Swayed by pictures!

Excerpt:

It is important to note that what is equal about the four quartiles is the number of data points included in each...
Swayed by pictures!

The quartiles are equally sized. What is “equal” about the quartiles?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>They each have the same range on the variable (i.e., 1-10, 11-20, 21-30, 31-40).</td>
<td>28%</td>
</tr>
<tr>
<td>B</td>
<td>They each have the same data points.</td>
<td>1%</td>
</tr>
<tr>
<td>C</td>
<td>They each have the same number of data points.</td>
<td>46%</td>
</tr>
<tr>
<td>D</td>
<td>They each have the same interval.</td>
<td>25%</td>
</tr>
</tbody>
</table>
Swayed by pictures!

Excerpt:

It is important to note that what is equal about the four quartiles is the number of data points included in each...
Swayed by pictures!

The quartiles are equally sized. What is "equal" about the quartiles?

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>They each have the same range on the variable (i.e., 1-10, 11-20, 21-30, 31-40).</td>
<td>8%</td>
</tr>
<tr>
<td>B</td>
<td>They each have the same data points.</td>
<td>5%</td>
</tr>
<tr>
<td>C</td>
<td>They each have the same number of data points.</td>
<td>84%</td>
</tr>
<tr>
<td>D</td>
<td>They each have the same interval.</td>
<td>3%</td>
</tr>
</tbody>
</table>
The "Better Book" Model

The genius of collaborating with other researchers!
1. Video Watching

control group: 9 videos
experimental group: 9+20

Dr. Laura Fries
Students watch smaller proportion of long videos
Students watch more minutes of long videos
2. Motivation
2. Motivation
3. Clickstream & Student Achievement
KEEP STUDENTS LIKE SARAH IN MIND

Video taken after Sarah took a face-to-face statistics class that used CourseKata in 2021
THANK YOU!

@cogscimom
@coursekata
Teaching Hard Things to All Students

If you want to find out more, join our CourseKata.org community!
Nathalie Guebels
Assistant Professor, Computer Science
Santa Barbara City College
What are we trying to do with DATA SCIENCE?

- Provide easy transfer pathway into the Data Science major at 4-year institutions
- Prepare ALL students across disciplines with Data Science skills applicable to research in their disciplines/areas of interests
- Increase Data Science awareness at SBCC
- Increase participation in our first DS course by recruiting across disciplines
- Offer an Associate’s Degree and a Certificate of Achievement in Data Science

=> Demystify Data Science for all Community College Students.
Our Data Science journey so far....

2019:
SBCC collaborates with UCSB on NSF grant proposal

January 2020:
SBCC joins the Central Coast Data Science Partnership (CCDSP)

Fall 2020:
SBCC joins the California Alliance for Data Science Education

Summer 2021:
SBCC students participate in Summer Research Fellowships at UCSB

Fall 2021:
CS 118/MATH 118: new “Data Science for All” course launch

Summer 2022:
SBCC students participate in Summer Research
& present at Undergraduate Research Symposium at UCSB

Fall 2022:
Associate’s Degree and Certificate in Data Science proposal submitted
& in review with SBCC Curriculum and Articulation Committee
First SBCC Course: Data Science for All (SBCC's version of Data 8) Launched Fall 2021

- **CS 118/MATH 118**: Cross-listed Computer Science/Mathematics (co-taught by CS/MATH)
- Developed as part of the Central Coast Data Science Partnership (CCDSP) effort
- Designed in parallel with the new DS series at UCSB CS 5A/B and with the amazing support of the Data 8 Adoption Team at UCB
- CSU/UC Transferable,
- satisfies CSUGE Area B4, IGETC Area 2A (Mathematics/Quantitative Reasoning)
Our experience with the new course so far…

- **Fall 2021 & Spring 2022:**
  1 section (~24 students).
  Taught synchronously online.

- **Fall 2022:**
  2 sections (~40 students).
  Lecture online, with option for in-person lab.

- High female participation compared to Computer Science courses!!

- Students with **different career goals**

*Pictures taken by SBCC office of Communications in CS104: Intro to programming (lab), Fall 2019*
Surprises, observations and challenges along the way...

Collaboration is KEY:
- CCDSP, the Data 8 Adoption Team and the California Alliance
  => all helped us with bringing quality curriculum and infrastructure support
- Collaboration between Computer Science and Mathematics
  => *Special thanks to my colleague from Mathematics, Andrea Cullinen*, who helped with the DS course proposal and co-teaches the course with me

**Students are excited about the course & DS! Student quotes:**

  "I love this course because it feels very applicable to real-life analysis of data using Python."

  "Relevance of topic in today’s job market."

**Challenges:**
- New course approval process takes a long time
- When a course is approved, it does not mean students will sign up
  => *the course needs to count towards a transfer requirement or degree*
- A community college student is not necessarily a UC Berkeley student
  ⇒ *Drop rate still seems a little high*
- Limited funding and not enough faculty trained to teach Data Science
What we hope for going forward....

- Launch new **AS Degree and Certificate in Data Science**
- Continue to work on **course articulation** with 4-year institutions to facilitate **smooth transfer pathways**
- Advertise more **paid summer internship opportunities for CC students** at 4-year institutions
- Increasing **Data Science awareness** at the community college level and recruit **across disciplines**!
- Work with our local **High School** on a **dual enrollment course**
Advice for others....

- **Collaborate** and apply for grants/funding
- Don’t work alone, reach out to other departments, disciplines and/or institutions
- Get started on curriculum proposals early and work closely with your articulation officer

=> Collaboration across institutions is imperative to create a well-prepared and diverse Data Science workforce.
THANK YOU!

contact info:
Nathalie Guebels,
Assistant Professor, Computer Science
SBCC lead, Central Coast Data Science Partnership (CCDSP)
Santa Barbara City College

nfguebels@pipeline.sbcc.edu
Innovating Data Science Curriculum

Eric Van Dusen, Data Science Undergraduate Studies, UC Berkeley
ericvd@berkeley.edu
What was my journey to Data Science

Had worked on campus as researcher and lecturer

Came to faculty workshop on Data 8 and got recruited to teach a connector course

Found a niche to do outreach to more faculty on the Data Science approach to pedagogy
Teaching

Economic Development Connector Course

Open Science Connector Course

Economic Models Connector Course - Data 88E

Honors Thesis Seminar

Staff job

Outreach - both on and off campus

Technical Liaison for teaching Jupyterhub

Modules and Connectors - two interdisciplinary initiatives
Data Science Modules & Connectors Program

Exposure to data science tools that give students the opportunity to work hands-on with a dataset relevant to their course.

- Lab assignments, homeworks, or projects

Principles of data analysis, statistics, and computing, in context of their domain.

- Over 5 years creating curriculum:
  - Students served: 13,000 +
  - Courses served: 80 +
  - Notebooks created: 250 +
  - Departments served: 40 +
    - Economics, Political Science, Sociology
    - Ethnic Studies, Social Justice, Environmental Justice
    - Biology, Physics, Earth Science

*If these Jupyter notebooks can succeed, Data Science is truly interdisciplinary*
Teaching Other Teachers

- Innovation in Pedagogy
- Coding for Professors
- Instructional Technology

Pedagogy workshop for UC Berkeley Professors

National Workshop for Data Science Education
California Panel at National Workshop 2020–22

- 2020 - First Panel
- 2021 - Summary of efforts at UC CSU CCC
- 2022 - Learning Lab and CID Chair
An advisory group from a diverse coalition of academic institutions.

- **16 California Institutions**
  - 6 UCs, 5 CSUs, 4 CCCs, 1 Private Institution
- **Quarterly Board Meetings**
- **Key Topics Discussed**
  - Articulation efforts
  - Key Classes for Transfers
  - DS Major Requirements and Pathways
Collaborative Vision

- **Coordination - DS Efforts across institutions**
  - Creating the first California-wide public survey of UC, CSUs and CCCs DS Programs
  - Gaining support from UCOP CSUCO and CCCO
  - Linkages for NSF (HDR) proposals

- **Process - Aiding the Transfer and Articulation process for CCCs**
  - Organized a Community of Practice Workshop (Nov 5)
  - Facilitating articulation for pilot institutions
  - Fundamentals of Data Science requirements
  - Mathematics background requirements
  - UC Office of the President - articulation coordination

- **Support - Infrastructure and Curriculum**
  - Creating resources/training for Intro to Data Science adoption
  - Cloud pilot partnership with 2i2c.org and Cloudbank
2021-2022 Progress Summary:

- **CADSE Website**
- **UC, CSU, CCC** Summary of Data Science Programs
- **Mathematics background** requirements for Intro. to Data Science
- **Fundamentals of Data Science** requirements
- **Intro To Stats Project**: Beginning Database of CSU Intro To Stats Courses
- A few of our 2021 milestones:
  - February 2021 meeting 149 attendees from 65 institutions across California
  - Participated in the CMC^3 and ADSA Conferences (December)
  - Microsoft and CADSE’s Faculty Residency Program Completed
  - Held a CCC Instructors Community of Practice Workshop (October)
  - February 2022 meeting (February)
  - CSU-Wide Workshop (April)
Let’s work together

- [data.berkeley.edu/californiaalliance](http://data.berkeley.edu/californiaalliance)
- [data.berkeley.edu/2022workshop](http://data.berkeley.edu/2022workshop)
- [ds-help@berkeley.edu](mailto:ds-help@berkeley.edu)
- [ericvd@berkeley.edu](mailto:ericvd@berkeley.edu)