

Data Science and Society: Learn It and Lead It

Data Science for Future Leaders Workshop
Iona University, New Rochelle, New York USA
October 23, 2024

Florence D. Hudson, Executive Director, Northeast Big Data Innovation Hub,
COVID Information Commons, and Prototype-Open Knowledge Network Education Gateway,
The Data Science Institute at Columbia University in the City of New York
Founder & CEO, FDHint, LLC; Editor and Author; IEEE Engineering in Medicine and Biology Society

Agenda

- The Northeast Big Data Innovation Hub
- The National Student Data Corps
- Critical Success Factors for Work, Life and Leadership

Northeast Big Data Innovation Hub



Funded by NSF grants #1550284,
1748395, 1916585, 2028999,
2139391, 2333532;
NIH AIM-AHEAD;
US Department of Transportation.

- The Northeast Hub is a community convener, collaboration hub, and catalyst for data science awareness, education and innovation.
- Our mission is to build a diverse, equitable, and inclusive community for data science learners, educators, researchers and professionals with open, free, online valuable content and accessible resources.
- We strive to increase data science capacity and talent, emphasizing underserved communities.



Education +
Data Literacy



Health



Urban to Rural
Communities



Responsible Data Science:
Security + Privacy + Ethics

Northeast Big Data Innovation Hub Focus Areas



Education & Data Literacy

National Student Data Corps
(NSDC)

NSDC Global Chapter
Community

Data Science Resource
Repository (DSRR)

Data Science
Flashcard Series



Health

COVID Information
Commons (CIC)

CIC Student Paper Challenge

CIC Working Groups

AIM-AHEAD Data Science
Training Core Portal



Urban to Rural Communities

Transportation Data Science
Project (TDSP)

Climate and Nature

Smart Cities Data Exchange

Driver Video
Privacy Challenge



Responsible Data Science

Security, Privacy & Ethics

Responsible Data Science
Comic Books

IEEE/UL 2933 TIPPSS
Standard Working Group

IEEE Cybersecurity TIPPSS for
Industry 2024 Workshop Series

IEEE Connected Healthcare
Cybersecurity 2021 Workshops

Seed Fund Awards and Success Stories

25 Seed Fund grants - \$600,000+ in funds to awardees in Northeast



Education +
Data Literacy
12 awards



Health
10 awards



Urban to Rural
Communities
5 awards



Responsible
Data Science
7 awards

Data Science Related Seed Fund Projects included:

- Curricular Structures to Blend Data Science & the Digital Humanities
- Development of a Data Analytics Learning Community
- Algorithmic Justice and Responsible AI
- Using Data Science to Study Environmental Racism, Justice, and Policy

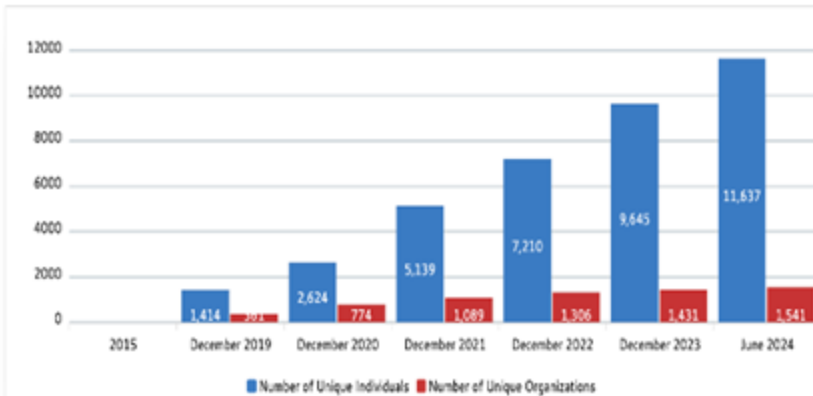


<https://nebigdatahub.org/seed-fund-about/>

NEBDHub includes 11,637 individuals from 1,541 organizations across the U.S. and 62 other countries

An Overview of our Community

Community Growth of the Northeast Big Data Innovation Hub



One Page NEBDHub June 2024 Quarterly Analytics Report

Growth Highlights

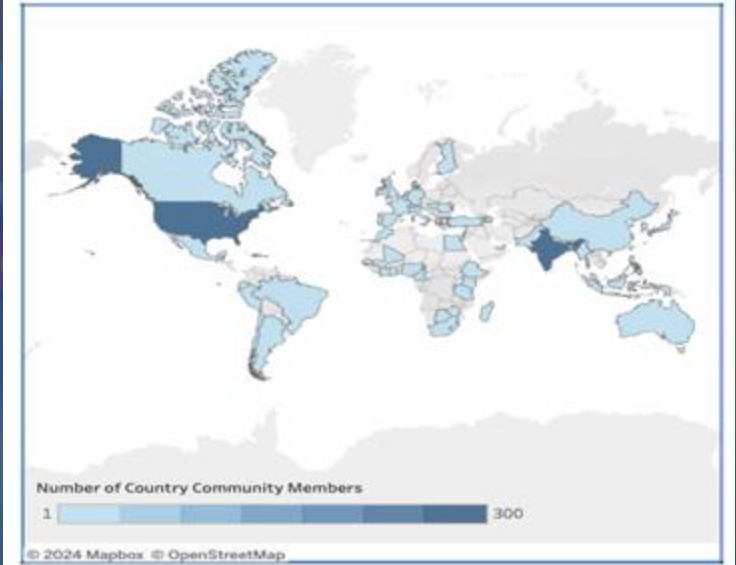
- The Hub community grew 39% Year over Year, the NSDC 61%, and the CIC 43%.
- The Hub community spans all 50 U.S. states, plus Puerto Rico, Washington, D.C., and 62 countries outside of the U.S.
- More than 24% of the Hub's academic community is affiliated with MSIs (2,169 individuals, up 76% YoY).
- The Hub now reaches 1,541 organizations globally, 229 of which are MSIs.

Hub Individual Growth



Community	Metric	Count	Quarterly Growth (%)	Annual Growth (%)
Northeast Big Data Innovation Hub	Total Individuals	11,637	13%	39%
	MSI-Affiliated Individuals	2,169	17%	76%
National Student Data Corps	Total Individuals	7,162	21%	61%
	MSI-Affiliated Individuals	1,452	26%	103%
COVID Information Commons	Total Individuals	4,193	14%	43%
	MSI-Affiliated Individuals	828	19%	92%

NEBDHub Community by Country



8X growth since 2019!

National Student Data Corps



The National Student Data Corps (NSDC) is a community-developed initiative that teaches data science fundamentals to students across the United States and around the world, with a special focus on underserved institutions and students.

We are grateful to the NSDC Founding Committee, which includes 24 members from academia, industry and nonprofits, who have come together to cultivate a community of diverse insights and perspectives to envision, develop, launch, and grow the NSDC portal and community, a groundbreaking and inclusive new program in data science.

NSDC Provides Free Online Data Science Resources for Learners and Educators

National Student Data Corps – Learner Central

Jump right in! Learn data science at your own pace by leveraging this collection of free resources for learners of all backgrounds. Whether you're just getting started or seek to learn advanced machine learning techniques, you'll be able to find something for your data science journey.



Data Science Ethics

Learn about the most important ethical issues facing data scientists today and how you can obtain and use data responsibly.
Prerequisites: None!

Getting Started

Get acquainted with the basics. Review the essential arithmetic and algebraic skills you'll need to begin working with data.
Prerequisites: None!



Introductory to Intermediate

Dive into data collection and analysis. Learn about statistical procedures, draw conclusions from data, and develop an intuition for common machine learning methods.
Prerequisites: Arithmetic, Pre-Algebra.



Intermediate to Advanced

Want to expand your data science knowledge? Learn about advanced statistical tools, state-of-the-art machine learning methods, and mechanics of optimization.



National Student Data Corps Video Library

Check out these videos on data science topics based on IBM's OpenDSA4AI curriculum and presented by Columbia University Research students.

Introduction to Data Science

With [Yuxuan Wang](#)

Introduction to Data Science – Part 1 (Yuxuan Wang)



Introduction to Data Science – Part 2 (Yuxuan Wang)



Data Acquisition and Wrangling

With [Ranyin Zhang](#)

Data Acquisition and Wrangling – Part 1 (Ranyin Zhang)



Data Acquisition and Wrangling – Part 2 (Ranyin Zhang)



Data Acquisition and Wrangling – Part 3 (Ranyin Zhang)



Supervised Machine Learning

With [Tomislav Gajanic](#)

Introduction (Tomislav Gajanic)



Introduction to Decision Trees (Tomislav Gajanic)



Decision Trees (Tomislav Gajanic)











National Student Data Corps – Educator Central

Welcome Educators!

Data science is more important than ever in today's digital world and has applications in all disciplines. The resources below can help you whether you are building a new data science program at your school, running a lab and looking for datasets, or just interested in how you can incorporate data science topics in your classroom.

Open Courses

Build a data science curriculum that works for your students using these materials from different universities and organizations.

 <p>Data 8: Foundations of Data Science</p> <p>This course from UC Berkeley covers statistics, data visualization, machine learning, and basic programming in Python.</p>	 <p>Data: Past, Present, Future</p> <p>This Columbia course integrates the teaching of algorithms and data manipulation with the cultural, philosophical, and ethical controversies from which these techniques emerged.</p>	 <p>Data Science Ethics</p> <p>Visit NSDC – Learner Central for resources on data science ethics issues including privacy, informed consent, and bias.</p>	 <p>NSDC 101: Introduction to Data Science</p> <p>Created and taught by NSDC Founding Committee member Ruy Anand, this introductory course was adapted from the Data 8 course. Assignments and corresponding code workbooks can be accessed at bit.ly/101datacourse.</p>
 <p>International Data Science in Schools Project</p>	 <p>OpenDSA4AI</p> <p>Explore lecture slides, assignments, and more.</p>	 <p>Open P-TECH</p> <p>This curriculum from IBM covers machine learning, AI, and more.</p>	 <p>The Missing Semester</p> <p>This MIT course covers the last 10 years of computer science.</p>

Learn Data Science in 3 minutes or less with the NSDC Flashcards

Flashcard Topics

- Data Science Ethics
- Data Science Pipeline
- Data Visualizations
- Data Types
- Data Storytelling
- Ontologies
- ... and more!



Learn and practice data science and earn a certificate with the NSDC Data Science Projects!



Advanced SQL Queries



Analyzing Pandemic Response



Climate & Permafrost-fed Lakes



COVID-19 Data Dashboard



COVID-19 Data Visualization



Data Cleaning



Data Essentials



Image Classification



Intermediate SQL Queries



Intro to SQL Queries



Movie Recommendations



Movie Reviews Sentiment Analysis



Transportation Projects



Yahoo Finance & Stocks



Fold Data Science Projects into your Curriculum!

- **Transportation Data Science Project**

- City College of NY, Columbia University, Bluefield State University, University of South Carolina, New York University, Rensselaer Polytechnic Institute, Purdue University are!

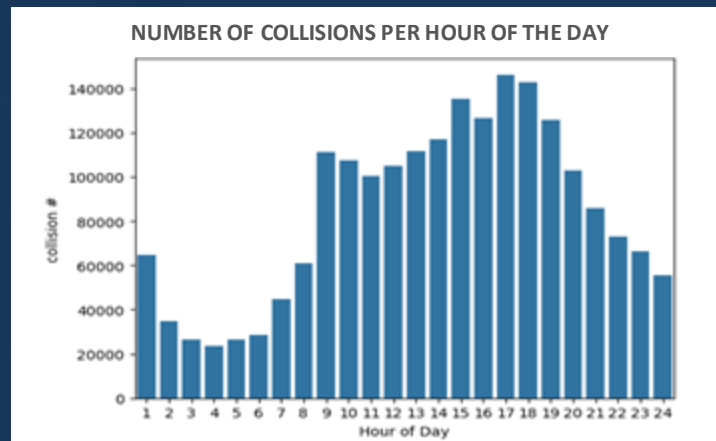
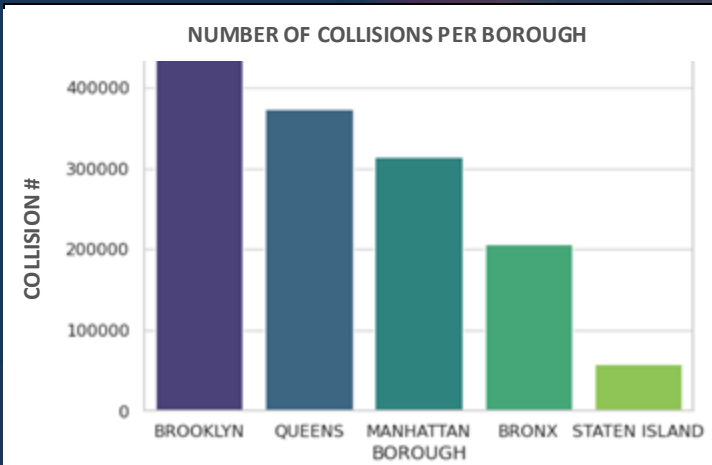
- **CIC Analyzing Pandemic Responses Project & CIC Data Essentials Project**

- University of Washington, Tacoma is too!

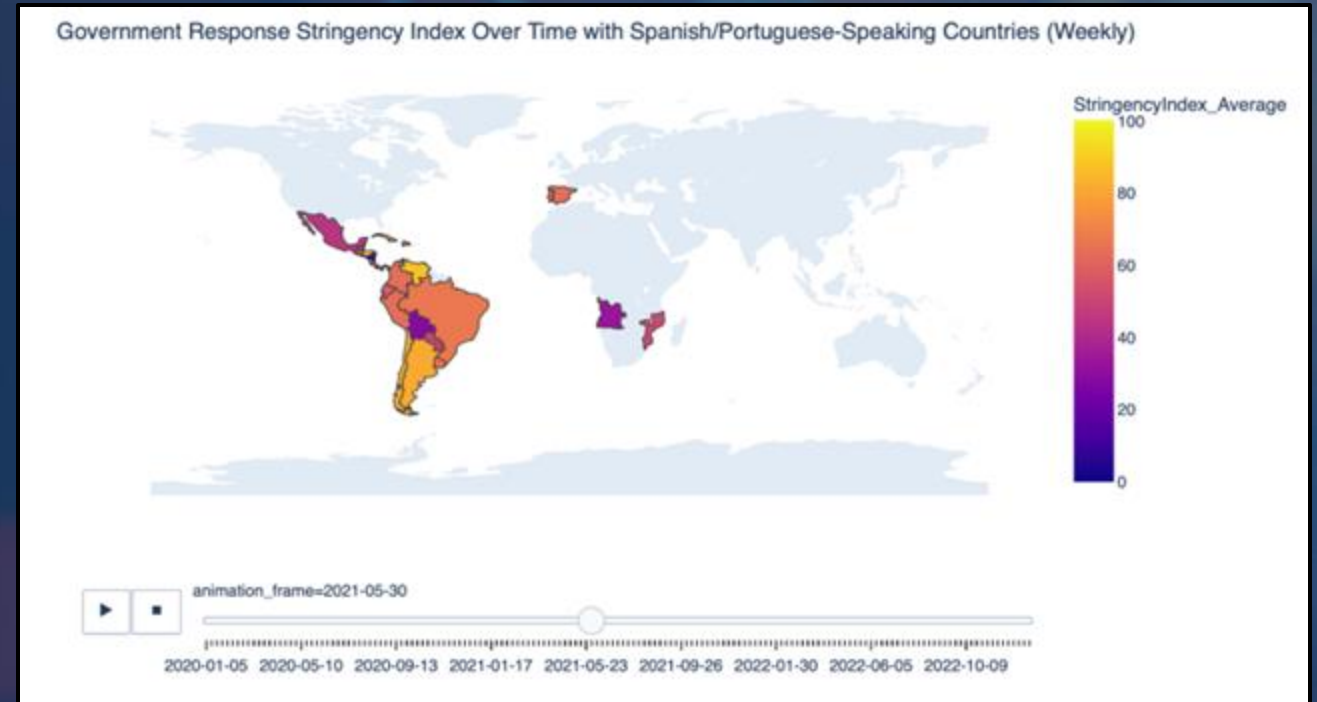
 CCNY COLUMBIA
UNIVERSITY Uof
SC PURDUE
UNIVERSITY W
TACOMA

Create Real-World Data Visualizations!

Transportation Data Science Project



CIC Analyzing Pandemic Response Project



Data Science Resource Repository

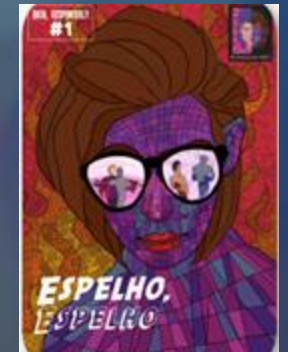
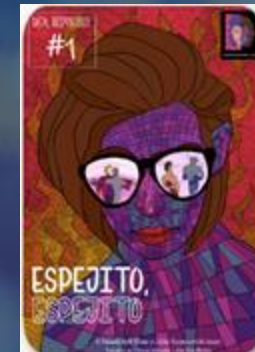
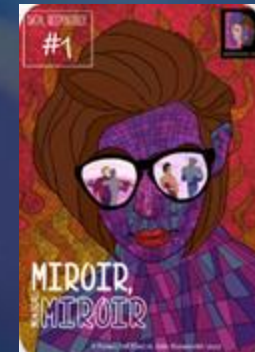
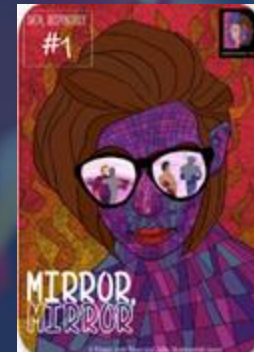
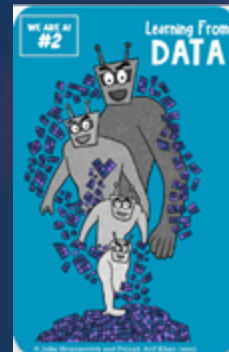
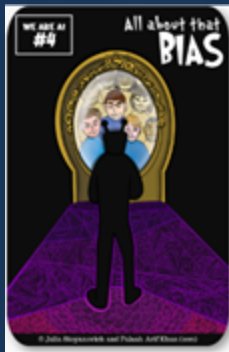
The Data Science Resource Repository (DSRR) is a curated set of 1,100+ resources for learners, educators, researchers, career explorers, and professionals that promotes data science literacy and learning, including 200+ Spanish language materials.

Resources for Educators

- Data Science Curriculum
- Data Science Programs
- Books, blogs
- Open Datasets

Resources for Students

- Self-Guided Lessons
- Data Science Challenges
- Open Projects & Use Cases
- Data science comic books



NSDC MasterClass Video Series

The NSDC MasterClass Video Series showcases experts who share how data science tools and techniques are being leveraged in various domains. Future episodes may highlight the intersection of data science and healthcare, finance, athletics, entertainment, technology, education, public policy, and more.

An Introduction to AI in Precision Oncology with Ian Maurer



Welcome to the
National Student Data Corps
MasterClass Series
August 2023

Episode #1: An Introduction to AI in Precision Oncology

TODAY'S SPEAKER:

Ian Maurer
CTO of Genentech

MODERATORS:
 
Florence Holcomb
NSDC Program Director and NSDC Founder
Emily Rubenberg
NSDC Program Manager

Generative Artificial Intelligence with Svitlana Volkova



Welcome to the
National Student Data Corps
MasterClass Series
September 2023

Episode #2: Generative Artificial Intelligence

TODAY'S SPEAKER:

Svitlana Volkova
Chief of Scientific Office of Science and Technology at Agilent

MODERATORS:
 
Florence Holcomb
NSDC Program Director and NSDC Founder
Emily Rubenberg
NSDC Program Manager

Data for Fame, Fortune & Championships with John Daniels



Welcome to the
National Student Data Corps
MasterClass Series
October 2023

Episode #3: Looking for an Edge: Mastering Data Science for Fame, Fortune, and Championships

TODAY'S SPEAKER:

John Daniels
Professor of Statistics at Central Michigan University

MODERATORS:
 
Florence Holcomb
NSDC Program Director and NSDC Founder
Emily Rubenberg
NSDC Program Manager

When Not to Use Machine Learning with Matthew Carbone



Welcome to the
National Student Data Corps
MasterClass Series
January 2024

Episode #4: When Not to Use Machine Learning

TODAY'S SPEAKER:

Matthew Carbone
Computational Scientist at Brookhaven National Laboratory

MODERATORS:
 
Florence Holcomb
NSDC Program Director and NSDC Founder
Emily Rubenberg
NSDC Program Manager

Interactive Massive-Scale Graph Analytics with David Bader



Welcome to the
National Student Data Corps
MasterClass Series
February 2024

Episode #5: Open-Source Framework for Interactive Massive-Scale Graph Analytics

TODAY'S SPEAKER:

David A. Bader, PhD
Distinguished Professor & Director, Institute for Data Science, New Jersey Institute of Technology

MODERATORS:
 
Florence Holcomb
NSDC Program Director and NSDC Founder
Emily Rubenberg
NSDC Program Manager

AI and DEIA: Wrong Problems, Wrong Solutions with Matthew Salzano



Welcome to the
National Student Data Corps
MasterClass Series
August 2024

Episode #6: AI and DEIA: Wrong Problems, Wrong Solutions

TODAY'S SPEAKER:

Matthew Salzano, Ph.D.
Institute, Diversity, Equity, and Access (IDEA) Fellow, Stony Brook University

MODERATORS:
 
Florence Holcomb
NSDC Program Director and NSDC Founder
Emily Rubenberg
NSDC Program Manager

NSDC Data Science Career Panels

Inspirational Panelists!!



Haleigh Stewart
Undergraduate Senior
at Columbia University



Benjy Sango
Undergraduate Senior
at Columbia University



**Edwood
Brice**



**Julia
Stoyanovich**



**Yoni
Malchi**



**Shane
Chang**



**Brandon
Sepulvado**



**Asya
Spears**



**Amen Ra
Mashariki**



**Daniel
Balouek-Thomert**



**Zoran
Obradovic**



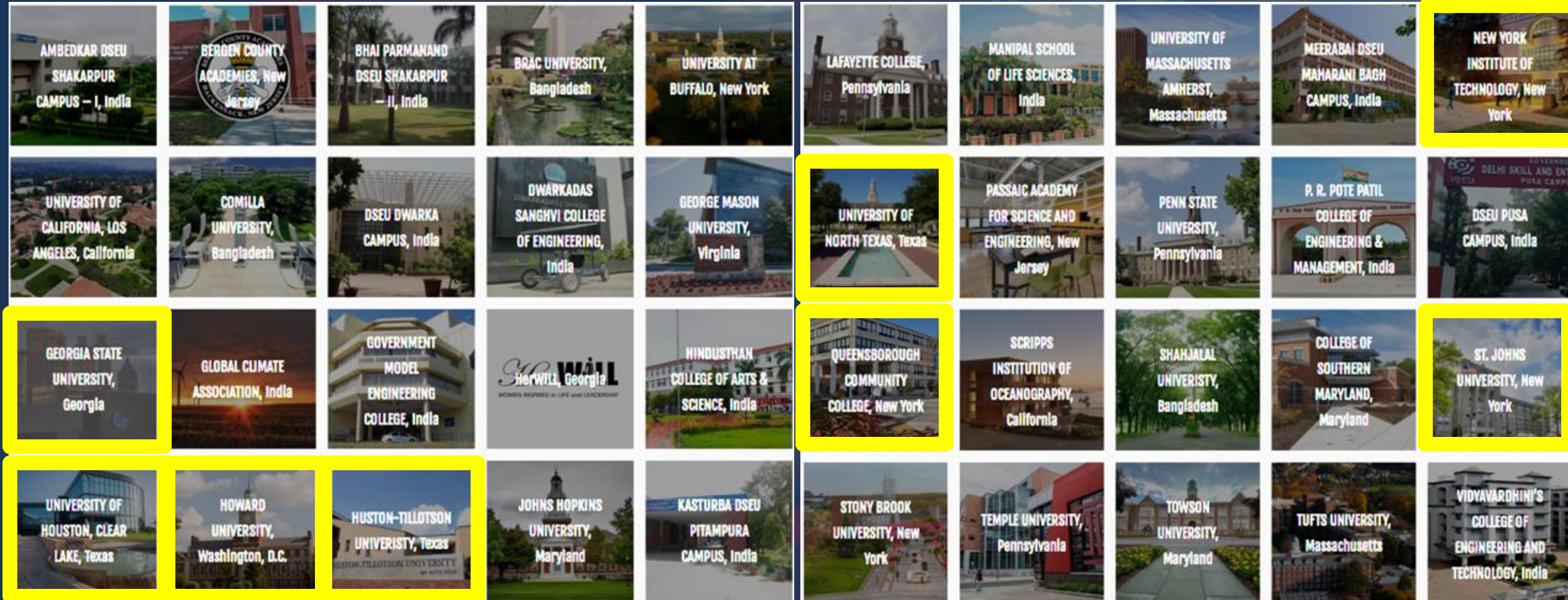
**Marija
Stanojevic**

Watch past panels on demand here:
bit.ly/nsdc-panels
or read summaries of the panels here:
bit.ly/nsdc-summaries

Gain Leadership Experience and start an NSDC Chapter!



MSIs and HBCUs in Yellow



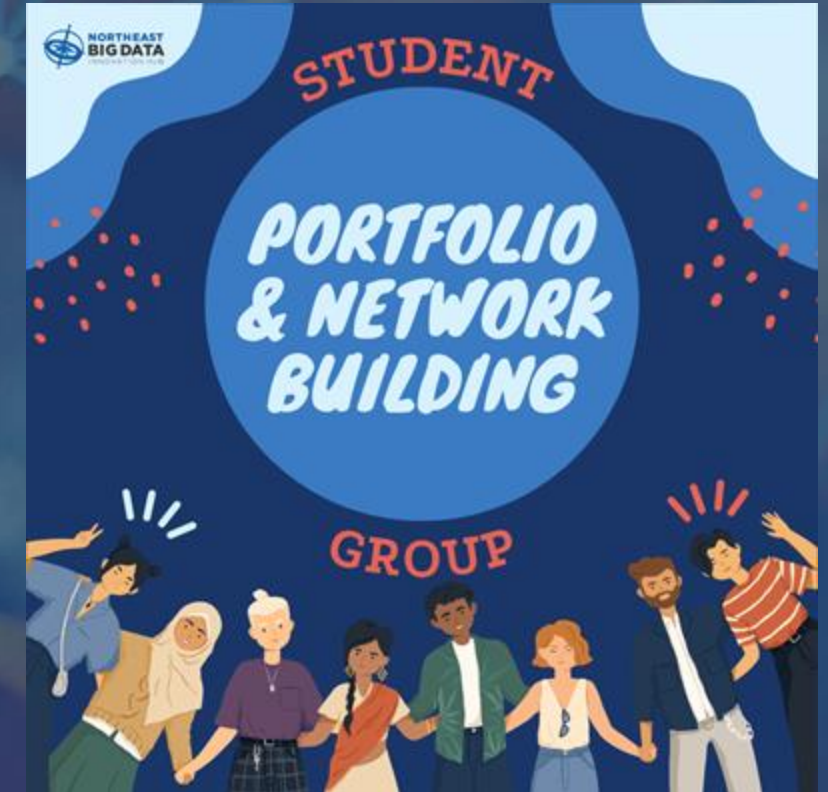
Student Portfolio & Network Building (PNB) Group

Four types of sessions and sets of online materials:

- A: Peer networking - download the toolkit!
- B: Resume and interview prep- download the toolkit!
- C: Professional communications - October 25, 2024
- D: Build your digital portfolio - January 10, 2025

Join sessions for free at any point in the cycle to build up your professional networking and interview skills.

Use the PNB LinkedIn Group to stay in touch between sessions!



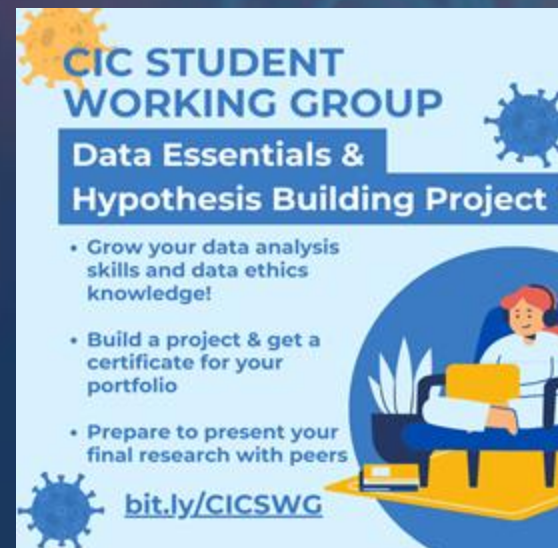
COVID Information Commons (CIC) Student Working Group

Four open-source, free COVID + data science projects for student learners of all backgrounds.

Next Meeting: Friday, November 1st

- Join the Working Group and attend monthly meetings with COVID researchers and students
- Strengthen your data analysis skills and gain practical experience with scientific research
- Receive a certificate for your resume and LinkedIn

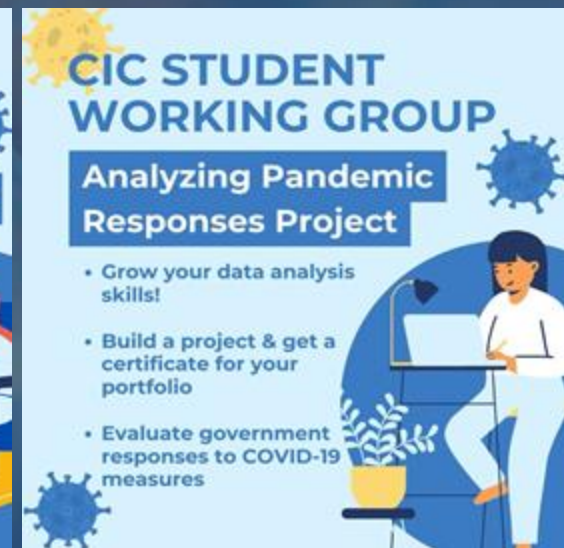
bit.ly/CICSWG



CIC STUDENT WORKING GROUP
Data Essentials & Hypothesis Building Project

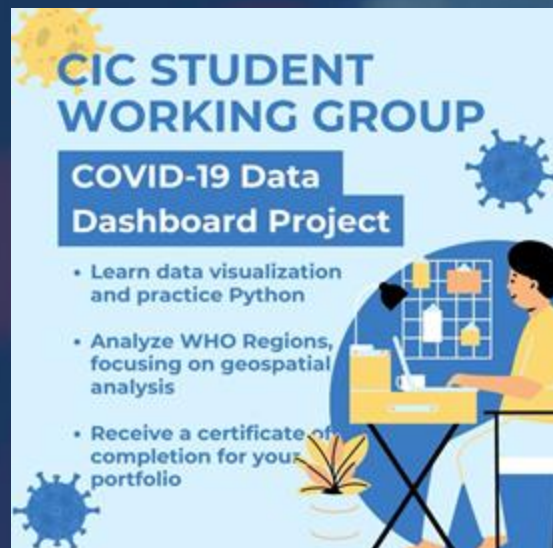
- Grow your data analysis skills and data ethics knowledge!
- Build a project & get a certificate for your portfolio
- Prepare to present your final research with peers

bit.ly/CICSWG



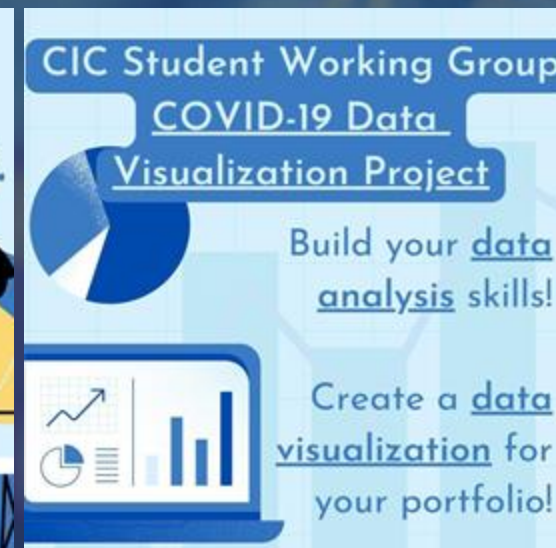
CIC STUDENT WORKING GROUP
Analyzing Pandemic Responses Project

- Grow your data analysis skills!
- Build a project & get a certificate for your portfolio
- Evaluate government responses to COVID-19 measures



CIC STUDENT WORKING GROUP
COVID-19 Data Dashboard Project

- Learn data visualization and practice Python
- Analyze WHO Regions, focusing on geospatial analysis
- Receive a certificate of completion for your portfolio



CIC Student Working Group
COVID-19 Data Visualization Project

Build your data analysis skills!

Create a data visualization for your portfolio!

Critical Success Factors for Work, Life & Leadership

- ***You can do it! Plan your life to achieve your purpose***
 - Determine your work and life goals and grow into your successful future
- ***Believe in Yourself, especially when others don't believe in you***
 - Mantra ... "I Believe in Me"
 - After a speed bump ... Be on YOUR side
- ***Build your support network*** – Mentors, Networking, Personal "cheerleaders"
- ***Build your skills***
 - Collaboration skills – teamwork, teambuilding, build an ark
 - Leadership skills
 - Communications skills – express your ideas – written, verbal, online
 - Technical skills – continually become an expert in your field
 - Business and financial skills – for work and life
- ***Follow your instincts, including living your values***



Florence DiStefano Hudson – My Professional Journey

Education:

- Princeton University BSE in Aerospace and Mechanical Engineering
- Executive Education Certificates Harvard and Columbia Universities

Professional Industry, Government and Academia Experience:

- NASA and Grumman Aerospace Engineer
- Hewlett Packard
- IBM Vice President Strategy and Marketing and Chief Technology Officer
- Internet2 Senior Vice President and Chief Innovation Officer,
- NSF Cybersecurity Center of Excellence Special Advisor
- FDHint, LLC, Founder and CEO
- IEEE P2933 Standards Working Group for Clinical Internet of Things and TIPSS, Chair
- NSF Northeast Big Data Innovation Hub, Columbia University, Executive Director and Steering Committee
- Author and Editor with Elsevier, Springer Nature, IEEE, NACUBO, Wiley, Taylor & Francis, etc.

Board Experience:

- Nonprofits: JLEP, SHORE, SWE, IEEE, NeurON, ERN, etc. ; Industry - IEC Electronics (NASDAQ)
- Academia/Govt: Columbia, Princeton, Stony Brook Universities, Cal Poly, UC San Diego, NCI

Florence DiStefano Hudson – My Personal Journey



Family and Faith



Hillary Clinton



Mae Jemison
First US African American
Female Astronaut



Sally Ride, First US Female Astronaut



Bonnie Dunbar, Astronaut



Alan Bean, Astronaut



IBM VP and CTO - Mentor and BFFs



Princeton Mentees



U.S. Supreme Court
Justice Sonia Sotomayor



Nobel Laureate
Daniel Kahneman



My Princeton and SWE mentor, Yvonne Brill,
1st US female Rocket Scientist and President Obama



Key Learnings from My Journey

- **There are many speed bumps in life ... we all have them**
 - Keep rolling over your own speed bumps ... until you get where you're going
- **People who you think are supposed to support you, are not always on your side**
 - Professors, bosses, co-workers, family ... don't be shocked
- **You can be a threat, just by being yourself - intelligent, educated, accomplished**
 - When you perceive it, it's there ... build your wisdom
 - Find a role and place where you are respected for being yourself
- **Be confident enough to follow your instincts**
 - If it doesn't feel right, it's not ... don't be afraid to walk away
- **Forgive, but don't forget ...**
- **Find your happy place, your comfort zone, re-charge**
- **Never give up on your dreams ...**

Create a vision and make it a reality: TIPPSS - IEEE Standard for Trust, Identity, Privacy, Protection, Safety, Security

IEEE TIPPSS Working Group is a 2024 IEEE Emerging Technology Awardee

Publications to learn more about TIPPSS

<https://fdhint.com/writing>



Sources: "Enabling Trust and Security: TIPPSS for IoT", IEEE IT Professional 2018; <https://www.computer.org/csdl/magazine/it/2018/02/mi2018020015/13rBUxDiIT>
"Wearables and Medical Interoperability: The Evolving Frontier", IEEE Computer 2018, ©2018 IEEE, <https://ieeexplore.ieee.org/document/8481273>

Create your path ... follow your dreams ... be a star

*“Don’t go where the path may lead,
go instead where there is no path and leave a trail.”*



- Ralph Waldo Emerson (1803-1882)



A star is not something that flashes through the sky.

It stays in place and gives off a strong, steady glow.

A star always works at being a star.

Stars never take themselves for granted.

That's why they're stars.



You can do it!

Thank you

Florence Hudson

florence.hudson@columbia.edu

Executive Director, Northeast Big Data Innovation Hub
at Columbia University - <https://nebigdatahub.org>

Founder & CEO, FDHint - <https://fdhint.com/about>

<https://www.linkedin.com/in/florencehudson/>



Connect with us! Join us!



Join the Hub/NSDC
Mailing List:
bit.ly/HubNewsSignup



LinkedIn
Northeast Big Data
Innovation Hub



Slack
bit.ly/nsdc-slack-channel
bit.ly/cic-slack-channel



Twitter/X
[@NEbigdatahub](https://twitter.com/NEbigdatahub)



Instagram
[@nebigdatahub](https://www.instagram.com/nebigdatahub)



YouTube
bit.ly/NEastYT



Columbia University
Data Science Institute

Questions? Email us at contact@nebigdatahub.org!