Data Science for Social Justice (DSSJ):

A Tool for Broadening Participation on Dual Fronts

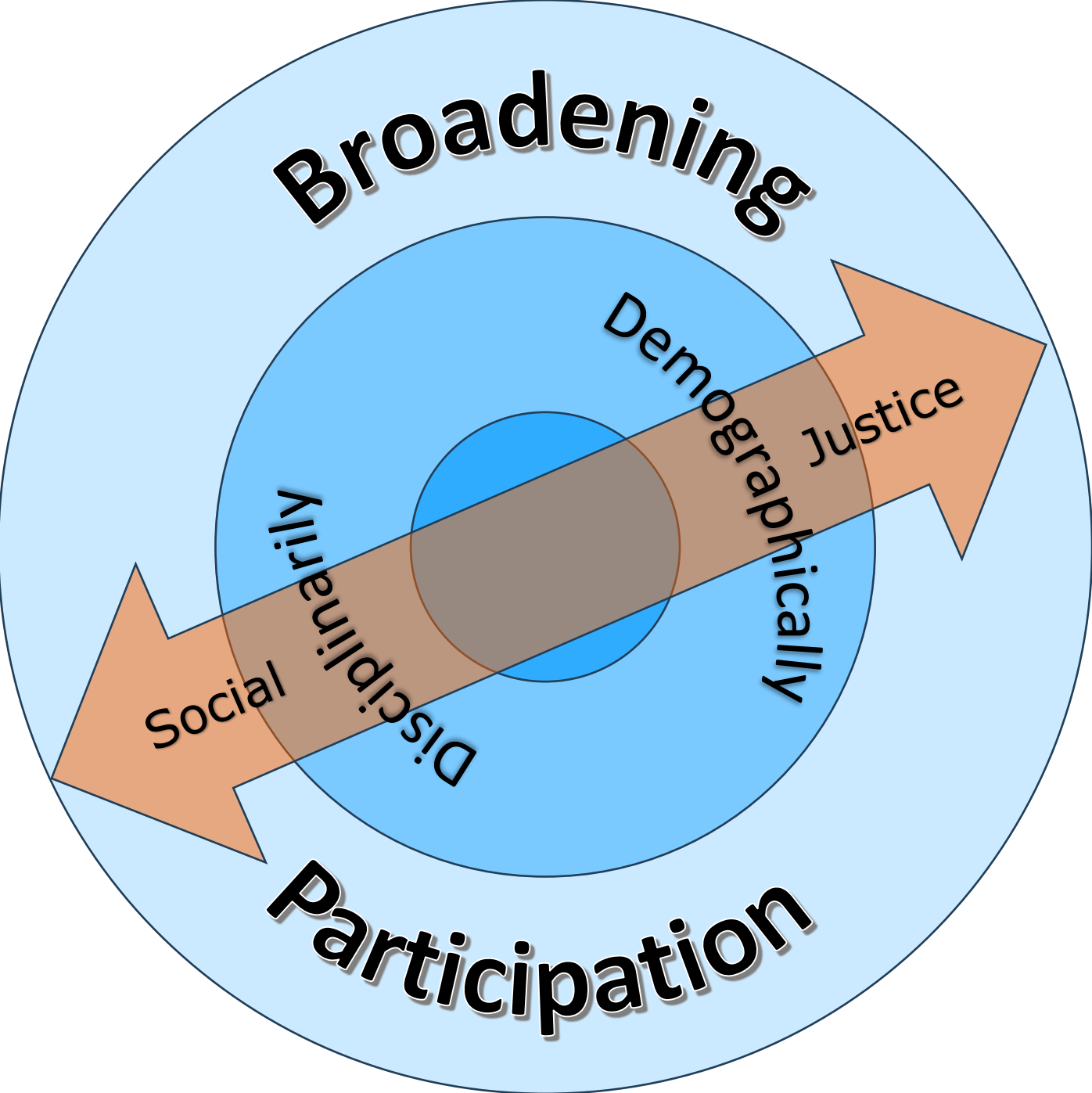**R. N. Uma**
NC Central University
Durham, NC

DSSJ PROJECT WEBSITE
https://sites.google.com/view/dssj

Dual Fronts to

Broaden Participation

in Data Science/STEM

using Social Justice

Demographically...

NC Central University

- Constituent member of the University of North Carolina System

- Historically Black College/University (HBCU)

- Liberal Arts University

- ~8,200 students (~6,400 UG)

- Student Profile
  - 75% African American, 10% White, 6% Hispanic
  - Over 80% of students are from historically marginalized communities (for STEM)
  - < 10% of UG degrees awarded are in STEM

- Research Triangle Park
  - NC State University (~35,000 students; ~25,000 UG)
  - UNC Chapel Hill (~30,000 students; ~19,000 UG)
 - Duke University (~16,000 students; ~7,000 UG)

BS Mathematics

CORE COURSES					
GENERAL EDUCATION COURSES					
MATHEMATICS CORE					
SCIENCE CORE					
COMPUTING CORE					
GENERAL MATHEMATICS	SECONDARY EDUCATION LICENSURE	COMPUTATIONAL AND ENGINEERING MATHEMATICS CONCENTRATION			
		DATA SCIENCE	S/W ENGG	COMPUTATIONAL SCI	NCSU 3+2

Rationale for using Social Justice

- Persistent underrepresentation of minorities and women in STEM despite several efforts.

Number of Bachelor's Degrees Awarded in Math/Stats/CS (MCS) (1966-2018) [NSF]
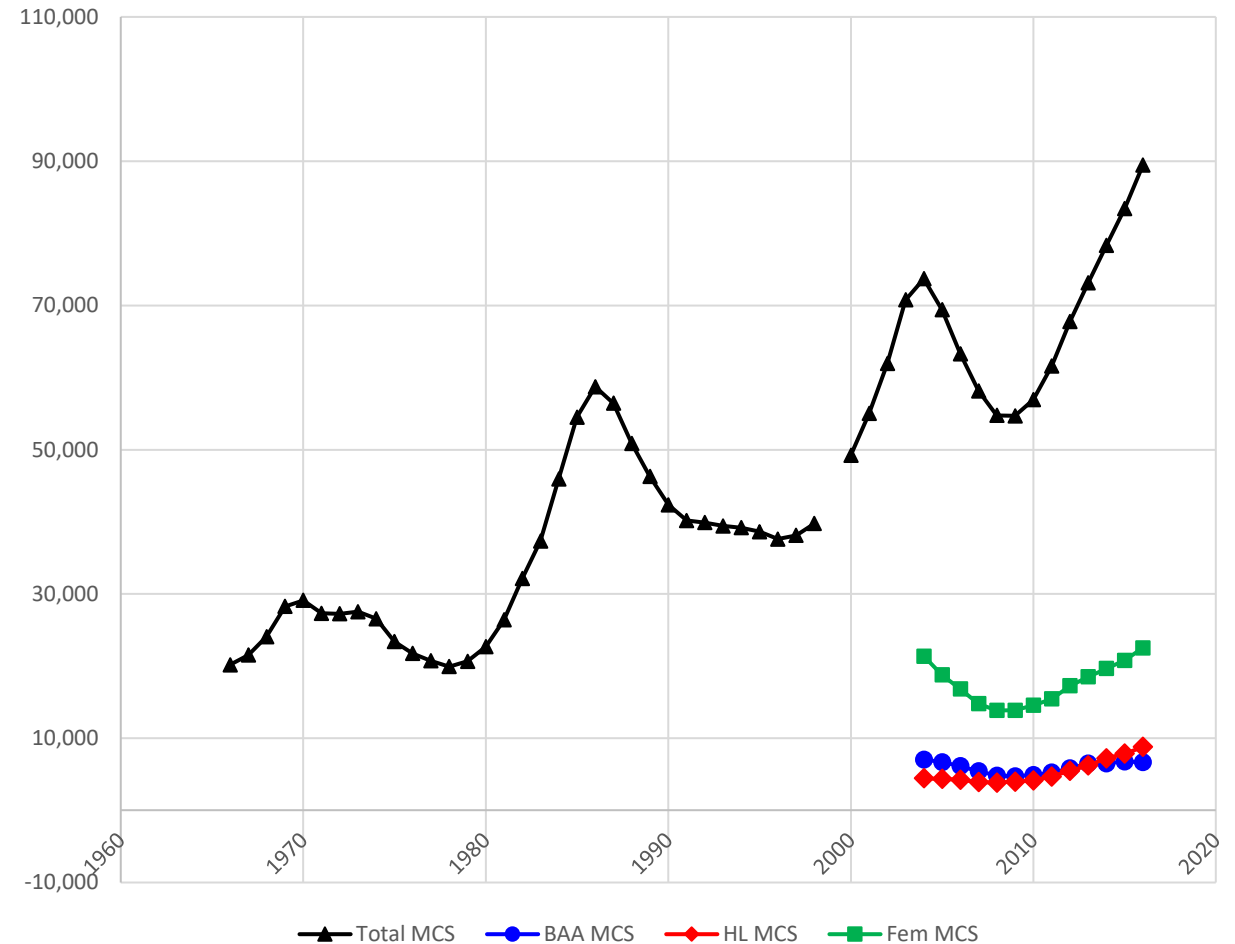
Rationale for using Social Justice

- North-East Scientific Training (NEST) Retreat 2014 [Campbell, Skvirsky, Wortis, Thomas, Kawachi, & Hohmann 2014]
  - Sought input from across the academic spectrum, to diversify the STEM workforce

  - Actionable items
    - "*adapt STEM training to include or make room for a social justice component*".
    - provide assistance to communicate science broadly
    - enable STEM to STEAM – interdisciplinary cross-talk and training
    - provide research career information prior to grad school
    - provide guidance and assistance for work-life balance
    - value other skills
    - provide access to invested mentors
    - provide opportunities for ancillary training

Rationale for using Social Justice

- Minority Students more motivated by "*Equity Ethic*" than a big paycheck. [McGee & Bentley 2017]
  - Seek STEM careers that integrates STEM expertise with "social justice, empathy, and equity matters".

- To attract and retain a diverse STEM student body
  - connect STEM content to real-world problems [Gorman 2010, Kim & Song 2013, Larios-Sanz, Simmons, Bagnall & Rosell 2011]

  - link STEM to "personal and culturally valued outcomes" [Estrada, et al. JWG on Improving URMs Persistence in STEM 2016]

Data Science for Social Justice (DSSJ)

- Incorporated in a *Freshman Seminar* course.

  - Group project

    - Identify and explain a social problem/issue
    - Describe the issue
    - Why is it important to address the issue
    - (Your solutions or recommendations to address the issue)

Pre-DSSJ

- **Freshman Seminar**

  - Prior Project – verbatim from a student-submitted report:

    "In 2010, there was an increase in the polices brutality cases. The increased of death was from 397 to 426 deaths. Black people were twenty four percent of people kill by cops. There are major racial disparities in how police officers use force on people."

Pre-DSSJ

- **Freshman Seminar**

  - Prior Project – verbatim from a student-submitted report:

    "In 2010, there was an increase in the polices brutality cases. The increased of death was from 397 to 426 deaths. Black people were twenty four percent of people kill by cops. There are major racial disparities in how police officers use force on people."

DSSJ Example Data Set: Fatal shootings by police officers
(downloaded on July 23, 2019 from https://github.com/washingtonpost/data-police-shootings)

name	date	manner_of_death	armed	age	gender	race	city	state	signs_of_mental_illness	threat_level	flee	body_camera
Patrick Wetter	1/6/2015	shot and Tasered	knife	25	M	W	Stockton	CA	FALSE	attack	Not fleeing	FALSE
Ron Sneed	1/7/2015	shot	gun	31	M	B	Freeport	TX	FALSE	attack	Not fleeing	FALSE
Hashim Hanif Ibn Abdul-Rasheed	1/7/2015	shot	knife	41	M	B	Columbus	OH	TRUE	other	Not fleeing	FALSE
Nicholas Ryan Brickman	1/7/2015	shot	gun	30	M	W	Des Moines	IA	FALSE	attack	Car	FALSE
Omarr Julian Maximillian Jackson	1/7/2015	shot	gun	37	M	B	New Orleans	LA	FALSE	attack	Foot	TRUE
Loren Simpson	1/8/2015	shot		28	M	W	Huntley	MT	FALSE	undetermined	Not fleeing	FALSE
James Dudley Barker	1/8/2015	shot	shovel	42	M	W	Salt Lake City	UT	FALSE	attack	Not fleeing	TRUE
Artago Damon Howard	1/8/2015	shot	unarmed	36	M	B	Strong	AR	FALSE	attack	Not fleeing	FALSE

DSSJ Example Data Set: Fatal shootings by police officers

- **Research Question (Sample):**

  What has been the impact of the *Black Lives Matter movement on police killings*? Have police departments been using the data that is being collected?

DSSJ Example Data Set: Fatal shootings by police officers

- **Guiding Questions – Easy (Sample):**
  - Create a summary statistics of the data to get a snapshot of the data - mean, median, standard deviation.
  - Create graphs of number of fatalities, by each race, gender, year, etc.
  - Create graphs to show distribution by race and by year within each race.
  - Repeat the previous item but with two other relevant variables of your choice.
  - Are whites or blacks more prone to dying as a result of police shooting because of mental health issues?

- **Guiding Questions – Intermediate (Sample):**
  - Explore other visualizations that may be better suited to represent this data.
  - How do the demographics of the state impact the race of the victim?

# DSSJ Example Data Set:
Fatal shootings by police officers

- **Guiding Questions – Advanced (S**

  - Can you predict the ou—en the characteristics of—al victim? Justify by building a—model to support your answer.

  - —igns of mental illness affect the rates of p—ce violence? When combining mental illness and race, are there correlations?

Dispatch mental health teams instead of police (Durham, NC)

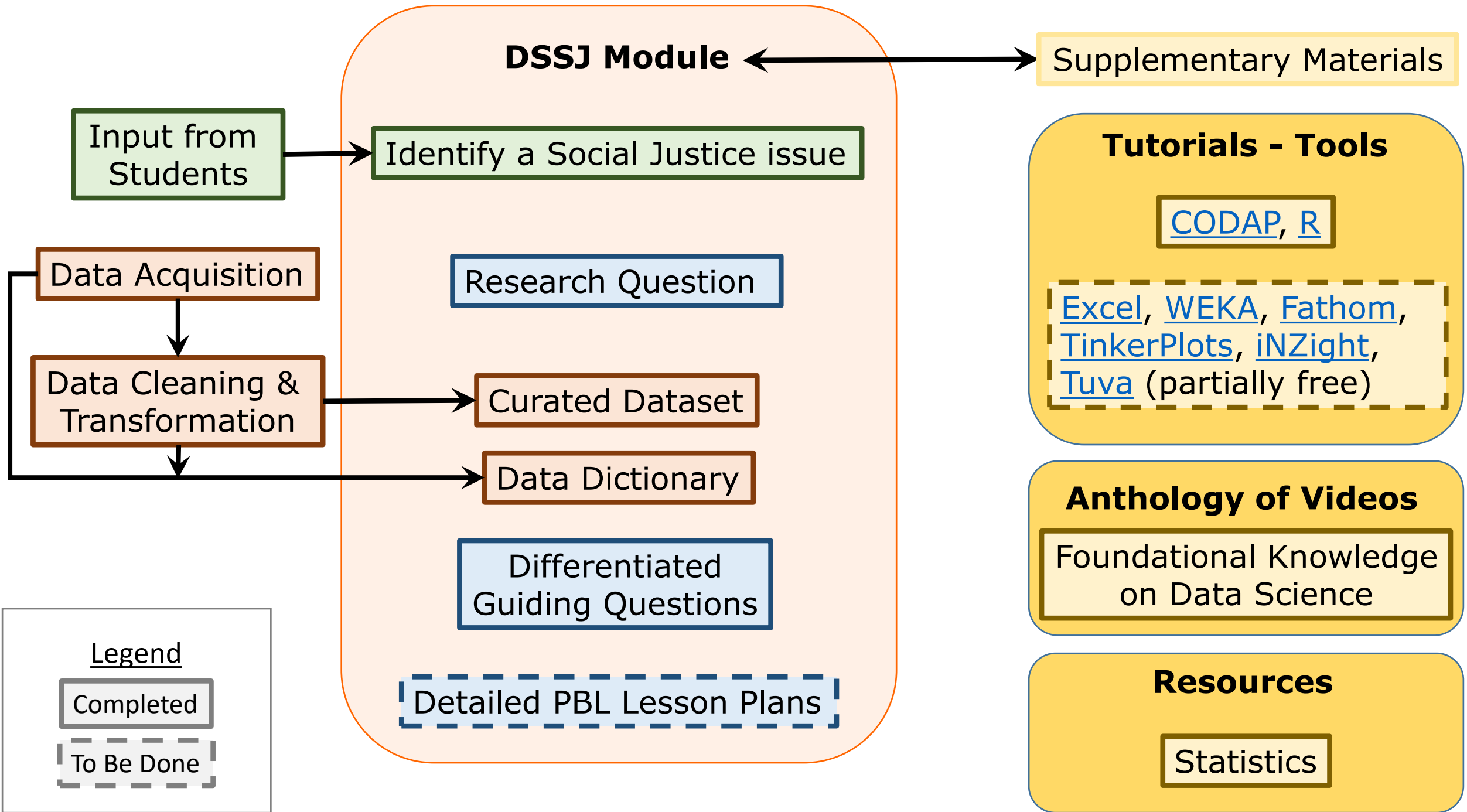# Freshman Seminar - DSSJ-Infused
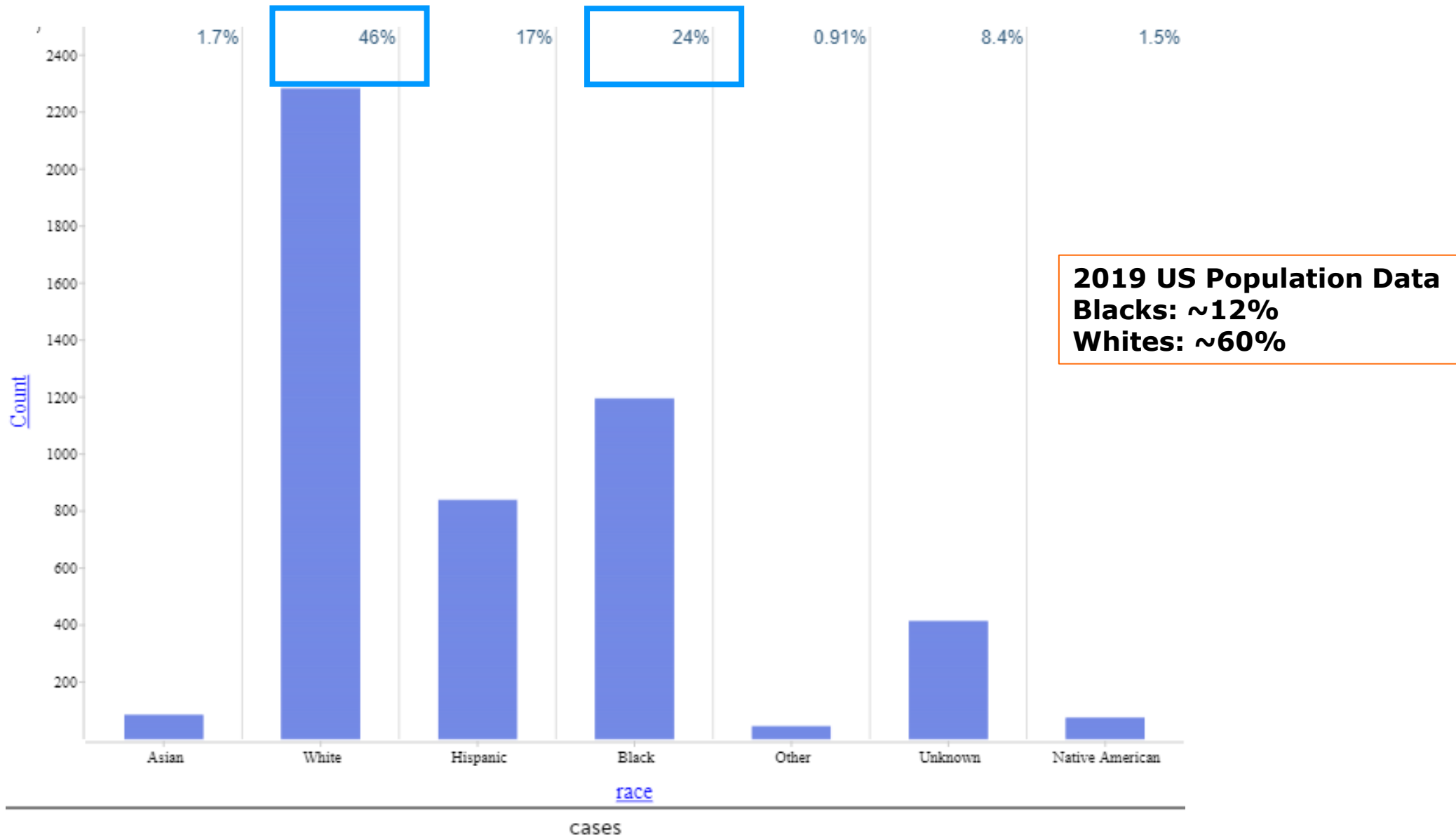
Challenges with this Approach

- **Instructors**
  - Access to data sets
  - Ability to clean data sets as needed
  - Familiarity with necessary tools: CODAP, R, Python, github
  - Ability to involve students from STEM and non-STEM background

- **Students**
  - Ability to pick up programming concepts in R/Python
  - Basic knowledge of statistics
  - Logical and critical thinking

DSSJ Module ⟷ Supplementary Materials

Input from Students → Identify a Social Justice issue

Data Acquisition

Data Cleaning & Transformation → Curated Dataset

Data Dictionary

Research Question

Differentiated Guiding Questions

Detailed PBL Lesson Plans

Tutorials - Tools

CODAP, R

Excel, WEKA, Fathom, TinkerPlots, iNZight, Tuva (partially free)

Anthology of Videos

Foundational Knowledge on Data Science

Resources

Statistics

Legend

Completed

To Be Done

Curated Social Justice Data Collection(https://sites.google.com/view/dssj/projects)

1. Police killings (2 data sets: shootings & all types)
2. Domestic Violence (NC)
3. Gun Violence
4. COVID-19
5. Sexual Assault (NC)
6. Environmental Pollution Data (2 data sets: by year & NC)
7. Cost of College Education
8. Child Abuse
9. LGBTQ+ Data

curated datasets - completed

10. Mental Health Issues
11. Social Media Effects
12. Immigration
13. Economic Justice
14. Energy Justice
15. ...

under development

Usage Metrics from figshare
(as of July 16, 2023)

Dataset	#Views (#V)	#Downloads (#D)	Conversion Rate (Percentage of ...hs)
Fatal Police Violence	1131	342	
Fatal Police Shootings	417		
COVID-19	244		
Domestic Violence (NC)			73.75
Sexual Assault (NC)			85.79
Environmental ... (2019)		103	41.87
Environm...		56	34.78
Gun Viole...	...05	124	40.66
Cost of Coll...	255	83	32.55
Child Abuse	Available with an End User Agreement		
TOTAL	**3182**	**1267**	**39.82**

Our conversion rate: 39%
Google Ads (all industries): 4.40%
Google Ads (jobs/education): 4.15%

Study Results
[Zulli Lowe, et al., 2022a, 2022b]

- To what extent does this project influence STEM enrollment and retention outcomes at an HBCU?

  - Increased desire to become more knowledgeable about data science

  - Anecdotally, a couple of students indicated they would consider changing their major to pursue data science (Math/CS).

Study Results
[Zulli Lowe, et al., 2022a, 2022b]

- How does this project impact the socio-cognitive factors (self-efficacy, identity, etc.) known to mediate and moderate STEM enrollment, persistence, and success?

  - Increase in students' data science knowledge

  - Greater interest in learning about data science as a career option

 - Greater awareness of big data use across employment sectors

Study Results
[Zulli Lowe, et al., 2022a, 2022b]

- What are the key features of this pathway model that will guide replication efforts by other STEM programs at this HBCU and other institutions? …

  - Students valued the use of CODAP as an analysis tool for data exploration

  - Students described the exposure to data explorations as "powerful"

  - Students and faculty found the topics to be relevant and interesting

  - Selection of social justice topics with input from students at a HBCU boosted student engagement

Study Results
[Zulli Lowe, et al., 2022a, 2022b]

- … What are the key features of this pathway model that will guide replication efforts by other STEM programs at this HBCU and other institutions? …

  - An interdisciplinary approach that incorporates the perspectives of other disciplines is important

  - Identifying the right tool for the data exploration required careful attention so that the data science introduction would not put up the well-known barriers for HMC students in entering STEM and computing

Study Results
[Zulli Lowe, et al., 2022a, 2022b]

- … What are the key features of this pathway model that will guide replication efforts by other STEM programs at this HBCU and other institutions?

    - For many students the word "data" is equated with "math" and triggers the same beliefs students hold about themselves as math students

    - Beyond developing future data scientists, HBCU faculty from the humanities and social sciences raised an equally loud call for the need to develop social scientists and scholars from a wide array of fields to be engaged in data science

Disciplinarily...
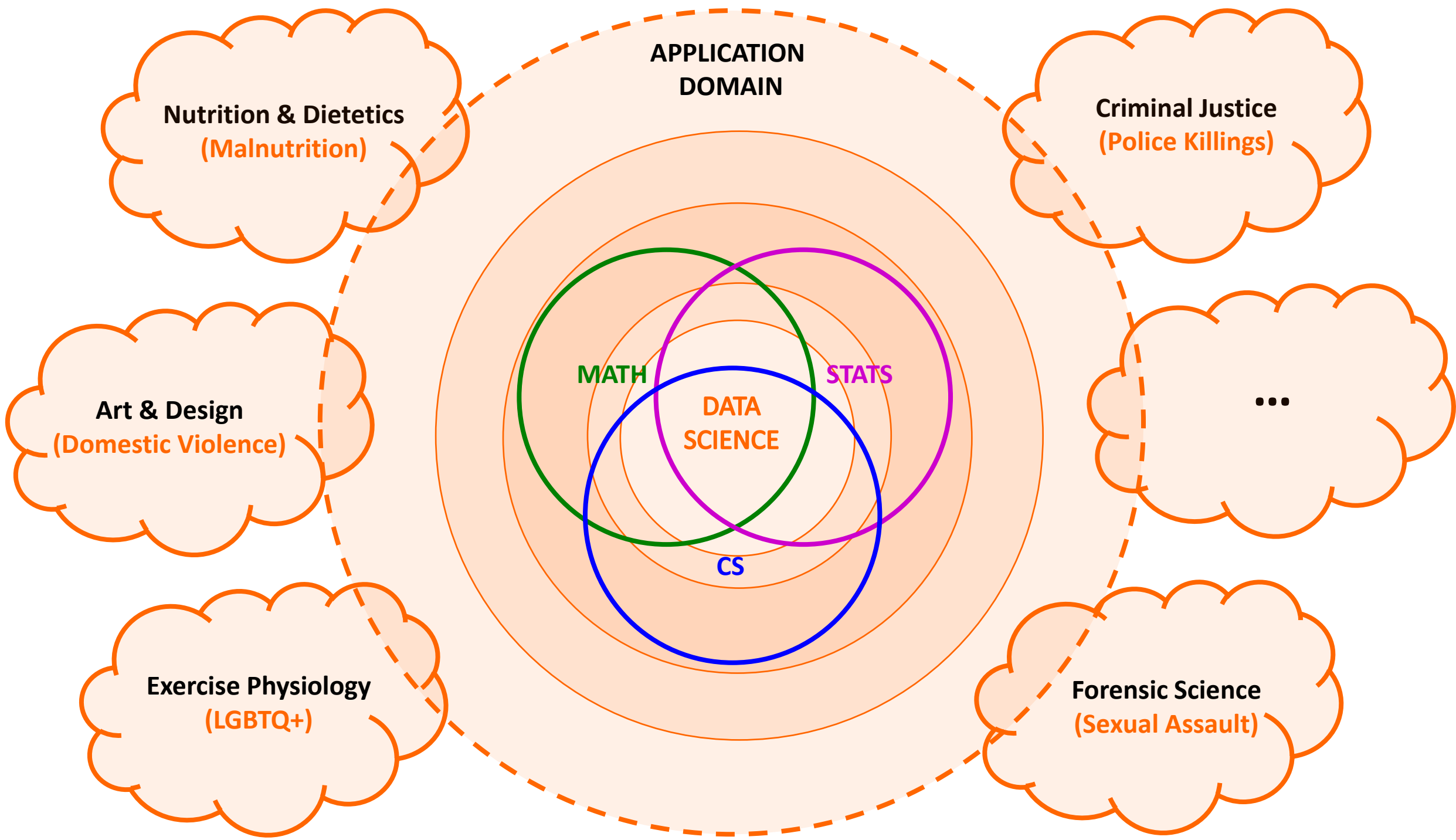
APPLICATION DOMAIN

Nutrition & Dietetics
(Malnutrition)

Criminal Justice
(Police Killings)

Art & Design
(Domestic Violence)

Exercise Physiology
(LGBTQ+)

Forensic Science
(Sexual Assault)

MATH

STATS

DATA SCIENCE

CS

• • •

Research Questions

- How do practitioners value this data science certification pathway?

- Does certification increase students' foundational knowledge of data science?

- How does the experience of the certification pathway (including entry, course-taking, post-certification) shape participants' identities, self-efficacy, perceptions of barriers, and career trajectory?

- To what extent does the DSSJ focus of the pathway resonate and interact with their identity as a member of a traditionally marginalized group?

- How does the innovation proposed transform undergraduate computing education?

Environmental Science

Demographics, as a percentage, of those receiving a BS degree in each of the listed CIP STEM disciplines (NCES, 2022). Data is averaged over five years 2016-2021. The last column represents the total percentage of Historically Marginalized Communities (HMC) that includes the following: B/AA-Blacks/African American, Hisp/Lat-Hispanic/Latino, AIAN-American Indian/Alaska Native, and NHPI-Native Hawaiian or Other Pacific Islander.

Discipline	B/AA	Hisp/Lat	AIAN	NHPI	Asian	White	HMC
NRC (includes Env Sci)	**2.56**	**10.36**	**0.73**	**0.20**	**4.35**	**71.84**	**13.85**
CIS	8.26	10.24	0.30	0.21	15.70	48.34	19.00
Engineering	3.93	11.00	0.25	0.13	11.23	56.85	15.30
Biological Sciences	7.97	13.39	0.35	0.18	14.64	53.53	21.89
Math/Stat	3.93	9.88	0.19	0.12	11.81	49.16	14.12
Phy Sci	5.17	11.08	0.31	0.13	9.21	60.77	16.68
US Population (USCensus, 2022)	**13.6**	**18.9**	**1.3**	**0.3**	**6.1**	**75.8**	**34.1**

Fall Enrollment of Lower Division Undergraduate Students seeking Bachelor's Degree in Environmental Science (UNC-Data, 2022).
(Note: The data for UNC Chapel Hill seems to be off in UNC Data Dashboard (UNC-Data, 2022).)

UNC System	2017	2018	2019	2020	2021
NCCU	12	15	14	15	13
App State	74	89	96	88	101
NC State	115	144	144	117	160
UNC-CH	2	2	3	-	-
UNC Pem	20	17	16	12	7
UNC Wilm	53	31	41	42	64
WCU	52	48	66	55	40

Improve Diversity in Environmental Science Program

- Innovate a gateway environmental science course at NCCU taken by majors and non-majors

  - innovate within an existing gateway course (ENSC 1000:Introduction to Sustainable Planet) by introducing data science for environmental justice project-based learning (PBL) modules.

Detailed-PBL Module

Cultural
Sensitivity

Instructor
Materials

Student
Worksheets

Mini-PBL Module

Need to knows

Voice and Choice

Entry
Event

Basic-PBL Module

Curated
Authentic Datasets

Research/Driving Question

Guiding Questions
(differentiated to the
interest level
of the students)

Sample
Activities

Debrief and
Reflection

Critique and Revision

Research Questions

- Does the incorporation of DSEJ-PBL modules within a gateway environmental science course result in increases in environmental science course-taking and in the selection of an environmental science major at an HBCU?

- To what extent do constructs such as student knowledge, interest, career aspirations, and attitudes, mediate the impact of the innovation?

Evangelization ...

Education & Workforce
➤ Institutional Engagement
➤ Faculty Development
➤ Broadening Participation
➤ Student Training / Experiences

Infrastructure
➤ Facilitate Sharing of Data, and Resources Across the Region
➤ **Social Impact Areas:** Geosciences, Criminal Justice, Food and Water Science

Community Engagement
➤ Connecting and Mobilizing the Regional HBCU and HSI Data Ecosystem
➤ Supporting the National NSF INCLUDES Ecosystem

Partnership & Collaboration
➤ Working with existing project-based programs on a shared evaluative frameworks
➤ Addressing Societal Challenges in partnerships with government, non-profits, industry, and academic partners

EDUCATION

INFRASTRUCTURE

ENGAGEMENT

PARTNERSHIP

Acknowledgement & Collaborators

QUESTIONS ?

Study Observations
[Zulli Lowe, et al., 2022a]

- **Effectiveness of the materials piloted (statistically significant findings)**

  - Increase in students' data science knowledge

  - Greater interest in learning about data science as a career option

  - Greater awareness of big data use across employment sectors

  - Increased desire to become more knowledgeable about data science

Study Observations (contd)

- **Analysis of qualitative data
  (student focus groups, course observations, instructor reviews, course artifact reviews):**

  - Students valued the use of CODAP as an analysis tool for data exploration.

  - Students described the exposure to data explorations as "powerful".

  - Students and faculty found the topics to be relevant and interesting.

Additional Key Observations
[Zulli Lowe, et al., 2022b]

- selection of social justice topics with input from students at a HBCU boosted student engagement

- an interdisciplinary approach that incorporates the perspectives of other disciplines is important

- identifying the right tool for the data exploration required careful attention so that the data science introduction would not put up the well-known barriers for HMC students in entering STEM and computing

- for many students the word "data" is equated with "math" and triggers the same beliefs students hold about themselves as math students

- beyond developing future data scientists, HBCU faculty from the humanities and social sciences raised an equally loud call for the need to develop social scientists and scholars from a wide array of fields to be engaged in data science