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

<table>
<thead>
<tr>
<th>Core Courses</th>
<th>General Education Courses</th>
<th>Mathematics Core</th>
<th>Science Core</th>
<th>Computing Core</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Mathematics</strong></td>
<td><strong>Secondary Education Licensure</strong></td>
<td><strong>Computational and Engineering Mathematics Concentration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Data Science</strong></td>
<td>S/W Engg</td>
<td>Computational Sci</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table above provides an overview of the different courses and concentrations within the BS Mathematics program.
Rationale for using Social Justice

• Persistent underrepresentation of minorities and women in STEM despite several efforts.
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)
Prior Project – verbatim from a student-submitted report:

“In 2010, there was an increase in the police 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.”
“In 2010, there was an increase in the police 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

<table>
<thead>
<tr>
<th>name</th>
<th>date</th>
<th>manner_of_death</th>
<th>armed</th>
<th>age</th>
<th>gender</th>
<th>race</th>
<th>city</th>
<th>state</th>
<th>signs_of_mental_illness</th>
<th>threat_level</th>
<th>flee</th>
<th>body_camera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patrick Wetter</td>
<td>1/6/2015</td>
<td>shot and Tasered</td>
<td>knife</td>
<td>25</td>
<td>M</td>
<td>W</td>
<td>Stockton</td>
<td>CA</td>
<td>FALSE</td>
<td>attack</td>
<td>Not fleeing</td>
<td>FALSE</td>
</tr>
<tr>
<td>Ron Sneed</td>
<td>1/7/2015</td>
<td>shot</td>
<td>gun</td>
<td>31</td>
<td>M</td>
<td>B</td>
<td>Freeport</td>
<td>TX</td>
<td>FALSE</td>
<td>attack</td>
<td>Not fleeing</td>
<td>FALSE</td>
</tr>
<tr>
<td>Hashim Hanif Ibn Abdul-Rasheed</td>
<td>1/7/2015</td>
<td>shot</td>
<td>knife</td>
<td>41</td>
<td>M</td>
<td>B</td>
<td>Columbus</td>
<td>OH</td>
<td>TRUE</td>
<td>other</td>
<td>Not fleeing</td>
<td>FALSE</td>
</tr>
<tr>
<td>Nicholas Ryan Brickman</td>
<td>1/7/2015</td>
<td>shot</td>
<td>gun</td>
<td>30</td>
<td>M</td>
<td>W</td>
<td>Des Moines</td>
<td>IA</td>
<td>FALSE</td>
<td>attack</td>
<td>Car</td>
<td>FALSE</td>
</tr>
<tr>
<td>Omarr Julian Maximilian Jackson</td>
<td>1/7/2015</td>
<td>shot</td>
<td>gun</td>
<td>37</td>
<td>M</td>
<td>B</td>
<td>New Orleans</td>
<td>LA</td>
<td>FALSE</td>
<td>attack</td>
<td>Foot</td>
<td>TRUE</td>
</tr>
<tr>
<td>Loren Simpson</td>
<td>1/8/2015</td>
<td>shot</td>
<td></td>
<td>28</td>
<td>M</td>
<td>W</td>
<td>Huntley</td>
<td>MT</td>
<td>FALSE</td>
<td>undetermined</td>
<td>Not fleeing</td>
<td>FALSE</td>
</tr>
<tr>
<td>James Dudley Barker</td>
<td>1/8/2015</td>
<td>shot</td>
<td>shovel</td>
<td>42</td>
<td>M</td>
<td>W</td>
<td>Salt Lake City</td>
<td>UT</td>
<td>FALSE</td>
<td>attack</td>
<td>Not fleeing</td>
<td>TRUE</td>
</tr>
<tr>
<td>Artago Damon Howard</td>
<td>1/8/2015</td>
<td>shot</td>
<td>unarmed</td>
<td>36</td>
<td>M</td>
<td>B</td>
<td>Strong</td>
<td>AR</td>
<td>FALSE</td>
<td>attack</td>
<td>Not fleeing</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
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 (Sample):
  • Can you predict the outcome given the characteristics of a potential victim? Justify by building a predictive model to support your answer.
  • How do signs of mental illness affect the rates of police violence? When combining mental illness and race, are there correlations?
Freshman Seminar - DSSJ-Infused

2019 US Population Data
Blacks: ~12%
Whites: ~60%
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
Curated Social Justice Data Collection

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

curated datasets - completed

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Views (#V)</th>
<th>#Downloads (#D)</th>
<th>Conversion Rate (Percentage of Clickthroughs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal Police Violence</td>
<td>1131</td>
<td>342</td>
<td>30.24</td>
</tr>
<tr>
<td>Fatal Police Shootings</td>
<td>417</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19</td>
<td>244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic Violence (NC)</td>
<td>244</td>
<td>103</td>
<td>45.90</td>
</tr>
<tr>
<td>Sexual Assault (NC)</td>
<td>240</td>
<td>56</td>
<td>73.75</td>
</tr>
<tr>
<td>Environmental Justice (2019)</td>
<td>305</td>
<td>124</td>
<td>41.87</td>
</tr>
<tr>
<td>Environmental Justice (NC)</td>
<td>161</td>
<td>56</td>
<td>34.78</td>
</tr>
<tr>
<td>Gun Violence</td>
<td>183</td>
<td>157</td>
<td>85.79</td>
</tr>
<tr>
<td>Cost of College</td>
<td>305</td>
<td>124</td>
<td>32.55</td>
</tr>
<tr>
<td>Child Abuse</td>
<td>255</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>3182</strong></td>
<td><strong>1267</strong></td>
<td><strong>39.82</strong></td>
</tr>
</tbody>
</table>

**Our conversion rate:**

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...
Exercise Physiology (LGBTQ+)
Nutrition & Dietetics (Malnutrition)
Criminal Justice (Police Killings)
Art & Design (Domestic Violence)
Forensic Science (Sexual Assault)
COURSE #1: Exploratory Data Analysis of Social Justice Issues

COURSE #2: Foundations of Data Science (Python) (UC Berkeley’s Data8)

COURSE #3: Machine Learning for All

COURSE #4: Responsible Data Science (NYU)

COURSE #5: Discipline-Specific Capstone Project (PBL)

The DATA Certificate program could potentially be expanded through offering optional *Special Topics in Data Science* in:

- Data Mining
- Text Mining
- Time-Series Modeling
- …
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

RECOMMENDED
NSF HBCU-UP BPRP (EES)
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.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>B/AA</th>
<th>Hisp/Lat</th>
<th>AIAN</th>
<th>NHPI</th>
<th>Asian</th>
<th>White</th>
<th>HMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRC (includes Env Sci)</td>
<td>2.56</td>
<td>10.36</td>
<td>0.73</td>
<td>0.20</td>
<td>4.35</td>
<td>71.84</td>
<td>13.85</td>
</tr>
<tr>
<td>CIS</td>
<td>8.26</td>
<td>10.24</td>
<td>0.30</td>
<td>0.21</td>
<td>15.70</td>
<td>48.34</td>
<td>19.00</td>
</tr>
<tr>
<td>Engineering</td>
<td>3.93</td>
<td>11.00</td>
<td>0.25</td>
<td>0.13</td>
<td>11.23</td>
<td>56.85</td>
<td>15.30</td>
</tr>
<tr>
<td>Biological Sciences</td>
<td>7.97</td>
<td>13.39</td>
<td>0.35</td>
<td>0.18</td>
<td>14.64</td>
<td>53.53</td>
<td>21.89</td>
</tr>
<tr>
<td>Math/Stat</td>
<td>3.93</td>
<td>9.88</td>
<td>0.19</td>
<td>0.12</td>
<td>11.81</td>
<td>49.16</td>
<td>14.12</td>
</tr>
<tr>
<td>Phy Sci</td>
<td>5.17</td>
<td>11.08</td>
<td>0.31</td>
<td>0.13</td>
<td>9.21</td>
<td>60.77</td>
<td>16.68</td>
</tr>
<tr>
<td>US Population (USCensus, 2022)</td>
<td><strong>13.6</strong></td>
<td><strong>18.9</strong></td>
<td><strong>1.3</strong></td>
<td><strong>0.3</strong></td>
<td><strong>6.1</strong></td>
<td><strong>75.8</strong></td>
<td><strong>34.1</strong></td>
</tr>
</tbody>
</table>
The table shows the fall enrollment of lower division undergraduate students seeking Bachelor’s Degree in Environmental Science for various institutions over the years 2017 to 2021. The data is provided by UNC-Data, 2022.

(Note: The data for UNC Chapel Hill seems to be off in UNC Data Dashboard (UNC-Data, 2022).)

<table>
<thead>
<tr>
<th>UNC System</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCCU</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>App State</td>
<td>74</td>
<td>89</td>
<td>96</td>
<td>88</td>
<td>101</td>
</tr>
<tr>
<td>NC State</td>
<td>115</td>
<td>144</td>
<td>144</td>
<td>117</td>
<td>160</td>
</tr>
<tr>
<td>UNC-CH</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UNC Pem</td>
<td>20</td>
<td>17</td>
<td>16</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>UNC Wilm</td>
<td>53</td>
<td>31</td>
<td>41</td>
<td>42</td>
<td>64</td>
</tr>
<tr>
<td>WCU</td>
<td>52</td>
<td>48</td>
<td>66</td>
<td>55</td>
<td>40</td>
</tr>
</tbody>
</table>
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.
Basic-PBL Module

Curated Authentic Datasets

Research/Driving Question

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

Critique and Revision

Debrief and Reflection

Mini-PBL Module

Need to knows

Voice and Choice

Sample Activities

Detailed-PBL Module

Cultural Sensitivity

Student Worksheets

Instructor Materials

Entry Event

Sample Activities
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
Acknowledgement & Collaborators

**DATA SCIENCE FOR SOCIAL JUSTICE**

**NSF HBCU-UP BPRP (HRD#1912408)**
Alade Tokuta (NCCU)
Rebecca Zulli Lowe (Cynosure)
Adrienne Smith (Cynosure)

**DATA SCIENCE FOR ENVIRONMENTAL JUSTICE**

**NSF HBCU-UP BPRP(EES - recommended)**
Alade Tokuta (NCCU)
Rebecca Zulli Lowe (Cynosure)
Adrienne Smith (Cynosure)
Elisabeth Stoddard (WPI)
Marja Bakermans (WPI)
Rakesh Malhotra (NCCU)

**DATA ANALYTICS CERTIFICATE FOR NON-COMPUTING MAJORS**

**NSF IUSE CUE (CNS#2245958)**
Alade Tokuta (NCCU)
Rebecca Zulli Lowe (Cynosure)
Adrienne Smith (Cynosure)
Debzani Deb (WSSU)
Sambit Bhattacharya (FSU)

**BROADER FACULTY/STUDENT NETWORK**

**NSF INCLUDES (EES - pending)**
Renata Rawlings-Goss (PI – GaTech)
Marc Boumedine (UVI)
Earvin Balderama (CalState Fresno)
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