June 23, 2022

Advanced Computing for Social Change Curriculum Workshop

ACSC Curriculum Workshop

Kate Cahill (kcahill@osc.edu)

XSEDE

Extreme Science and Engineering Discovery Environment

NSF

Workshop Outline

- 1 – 1:30 Intro to ACSC Program
- 1:30 – 2:30 Curriculum Examples
  Prof. Matthew Platt, Morehouse College
  Prof. Sarah Donovan, Wagner College
- Break
- 2:45 – 3:30 Resource Review
- 3:30 – 4:00 General Discussion

https://www.osc.edu/ACSCFaculty

ACSC/C4C History

Data Science as an Enabler

SPICE 2019
Summer Immersion Program

Research4Change 2019
REU SITE

Research4Change 2020
REU SITE

SPICE 2020
Summer Immersion Program

NSF REU Site

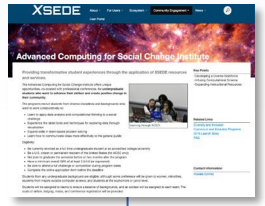
NSF CyberTraining

NSF Includes (SPICE)

PEARC19
Maternal and Infant Health

NASA Direct-STEM
Cal State Los Angeles

ACSC20

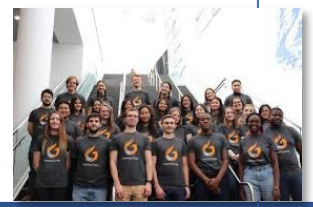


SC16
Black Lives Matter

SC17
Immigration

SC18
Violence

SC19
Maternal and Infant Health



2016 2017 2018 2019 2020

Funding for Chaminade's Data Science Program

Chaminade's First Data Science Students

NSF

XSEDE

Advanced Computing for Social Change • Computing4Change (C4C)

Founded in 2016 and co-located with SC

Engages undergraduates from diverse backgrounds and disciplines

Participants learn to apply data analysis and computational thinking to a social challenge

Students work in groups and:

- Construct a non-biased question for exploration
- Use computational resources to create visualization to confirm or debunk hypothesis
- Present results in teams using evidence- based analysis and visualization

Social Challenges

- Black Lives Matter (2016)
- Immigration (2017)
- Violence (2018)
- Maternal & infant morbidity and mortality in the U.S. (2019)
- COVID-19, violence, health, environment (2020)

How does the language that you speak impact your birthing experience, specifically with birth hysterectomies?

Alejandra Garcia Orosco, Biology and Microbiology, Cal State LA

IMPACT: 2016-2020 C4C Student Tracking

Participant Gender Distribution (N=126)	Count	Percent
Female	80	63%
Male	43	34%
Not Listed	3	2%

Participant Racial/Ethnic Distribution (N=118)	Count	Percent
Asian	20	17%
African American/Black	25	21%
Hispanic/Latinx	36	31%
Native Hawaiian Pacific Islander	13	11%
White	18	15%
Not Listed	6	5%

- Evaluation data show effective engagement of students from non-STEM disciplines
- Majority participants recruited from MSIs and institutions with limited research opportunities
- Increase in # of underrepresented XSEDE student users
- Participants demonstrate increased participation in XSEDE sponsored research opportunities

The Process

STEP 1: Identify your question

STEP 2: Ask: “Is the answer informed by data?”

STEP 3: What data do you need?

STEP 4: Does that data exist? If not, what are your proxies?

STEP 5: Does your data need cleaning?

STEP 6: Ask your question analyzing your data?

STEP 7: Are you using the right tools to do your analysis?

STEP 8: Have you looked at your data. If no, why not?

STEP 9: Are you surprised by your answers? If yes, did you verify that your answers are correct?

STEP 10: Do you need more data? If yes, go to step 3.

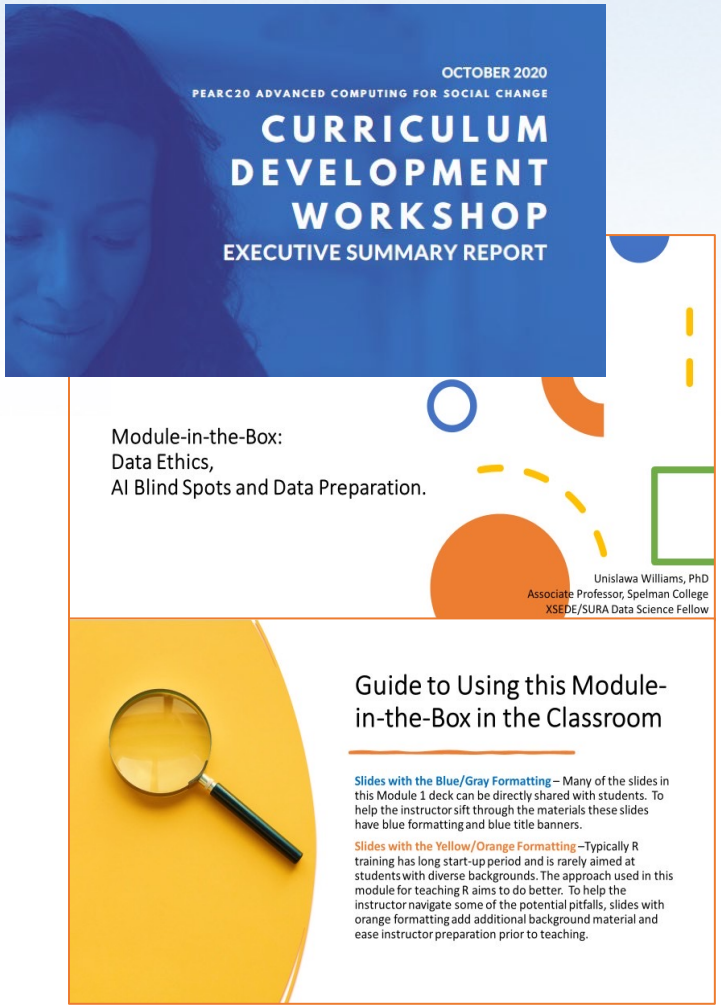
STEP 11: What did you find? What is your message or story?

STEP 12: Put your story together.

STEP 13: Do you have any proposed solutions?

STEP 14: Tell your story (presentation)

ACSC Curriculum Development (2020 – 2021)



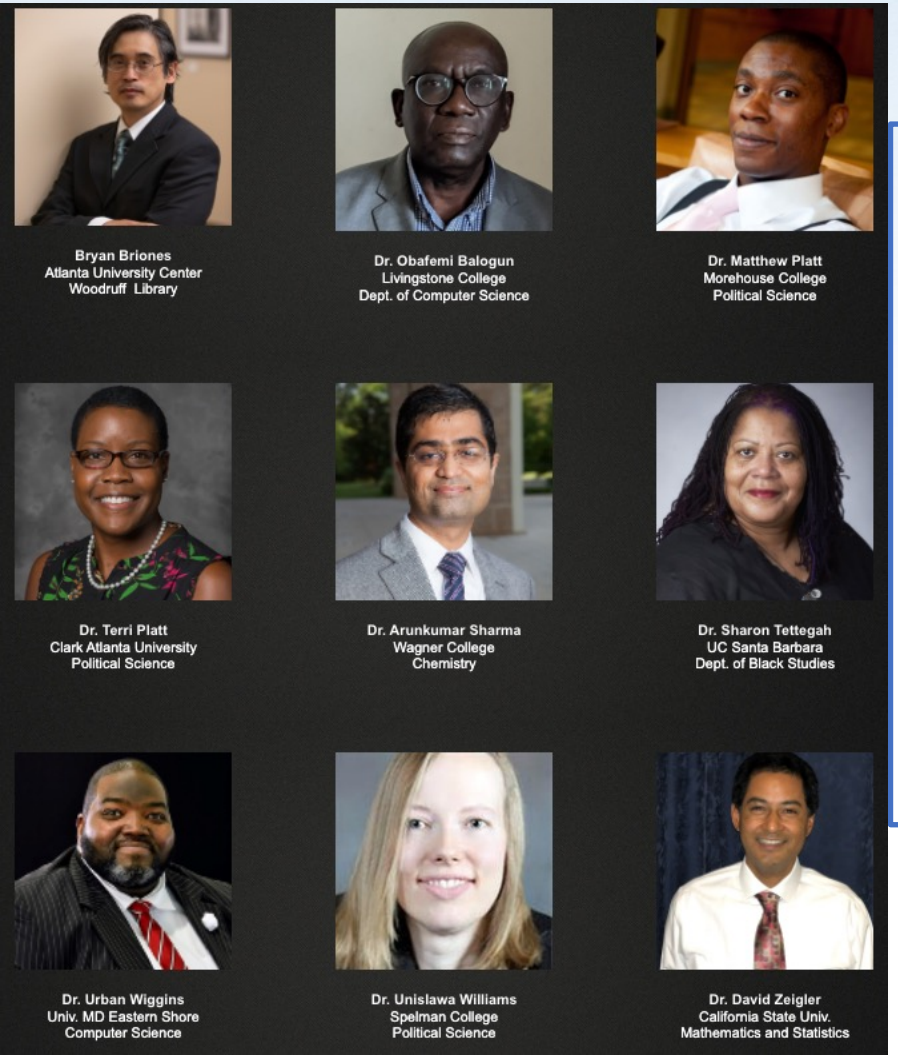
ACSC Faculty Curriculum Development Workshop, July 2020

- 34 applicants; 20 accepted; 18 participated from 10 institutions including Albany State University, AUCC, Livingstone College, UMES, and UC Santa Barbara
- Disciplines – Biology, Computer Science, Economics, Philosophy, Political Science and World Languages,
- Evaluated via focus group, observation, and survey
- Overall, participants valued their experience, enjoyed the opportunity to network with like-minded faculty, and appreciated being exposed to new tools.

Advanced Computing for Social Change (ACSC) Modules-in-a-Box Curriculum Materials

- Data Ethics, AI Blind Spots, and Data Preparation using maternal and infant mortality and morbidity problems and datasets – implemented at UPRM
- Data Ethics, Privacy, Modeling/Analysis using immigration problems and datasets
- Data Ethics and Scientific Visualization using COVID-19 and health disparities problems and datasets

ACSC Curriculum Development 2022



ACSC Data Science Community Development Fellows 2022

- 9 Fellows, meeting monthly to discuss how to infuse social justice into curriculum.
- Disciplines include Chemistry, Political Science, Computer Science, Mathematics, Black Studies
- Fellows are working together to develop new curriculum modules
- Summer Faculty Workshop coming up in June

Modules-in-a-Box

Module 1

- **Data Ethics,**
  AI Blindspots and Data Preparation

Module 2

- **Data Ethics,**
  Privacy and Data Modeling/Analysis

Module 3

- **The Pandemic, Data Ethics and Visualization**

Module features

Modules were created by Prof. Unislawa Williams of Spelman College as an XSEDE Data Science Fellow

- Use R as data analysis platform
- No previous R experience assumed
- Separate topics but data science skills are sequential
- Slides include student material and instructor information
- Hands on practice included as well as homework and project ideas including grading rubrics and suggestions for further study

Module 1

Student Learning Outcomes (SLO): Understand what are AI Blindspots, how they arise, and how we may mitigate or even avert them.

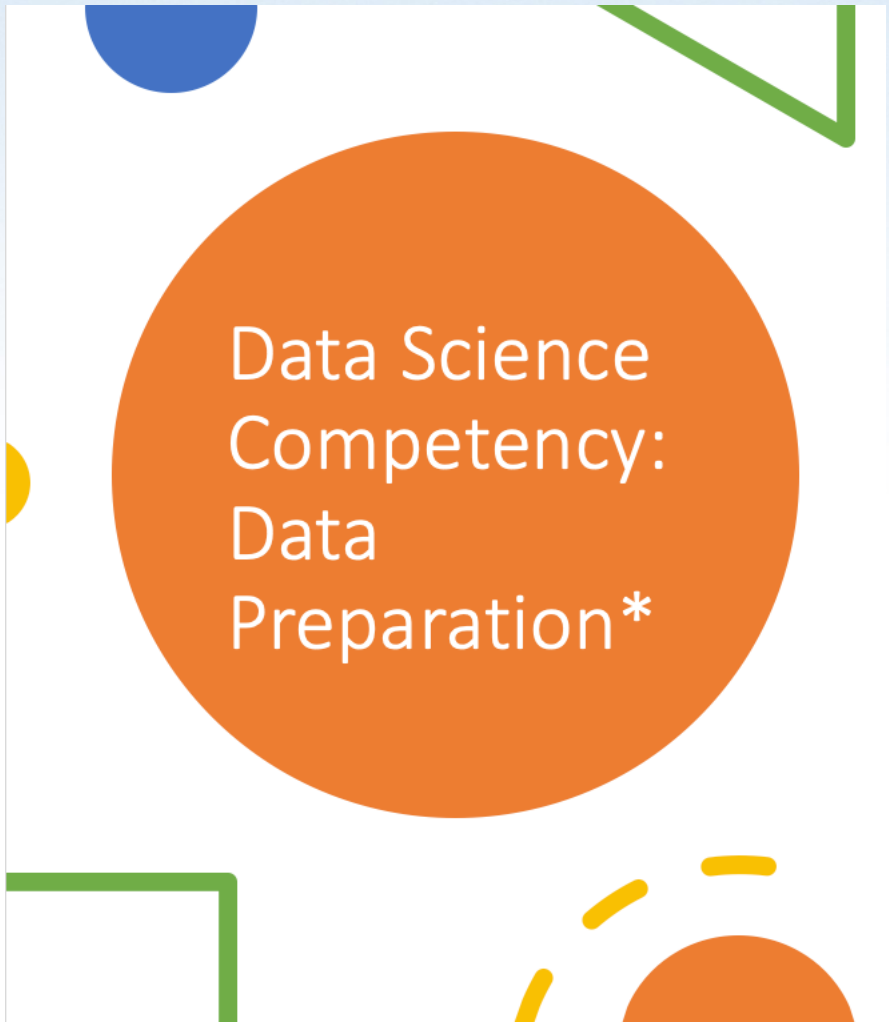
Presentations: Data's Ethical Component

Discussion Point: When should we start talking about data ethics when teaching data science?

The advantages of teaching data ethics along side data science is that the two subjects become well integrated.

***MIT Media Lab:**
https://aiblindspot.media.mit.edu/

Module 1



Student Learning Outcomes:

- "Demonstrate ***[some]*** ability to identify and collect data—multiple formats (IBM Analytics, 6)." ***Exercise 1.2***
- "Demonstrate ***[some]*** ability to manipulate, transform, and clean data (IBM Analytics, 6)." ***Presentations: R Skills***
- "Demonstrate ***[some]*** expertise with techniques to deal with missing values, outliers, unbalanced data, as well as data normalization (IBM Analytics, 6)." ***Exercise 1.2***
- "Demonstrate through a project ***[some]*** ability to construct usable data sets (IBM Analytics, 6)." ***Exercise 1.1***

Discussion Point: Is data science just about data analysis?

***IBM Analytics.** "The Data Science Skills Competency Model: A Blueprint for the Growing Data Scientist Profession." International Business Machines (IBM) Corporation. www.ibm.com/downloads/cas/7109RLQM

Module 1

Students will [begin to] understand how data is originated from diverse sources

- Students will [begin to] understand the relationships between objects and their representation in a digital data repository, such as a database or a group of files, by exploring a number of example datasets. ***Exercise 1.1 & 1.2***
- Students will [begin to] actively undertake a data acquisition task and the design of a relevant dataset. ***Exercise 1.1 & 1.2***

Ability to recognize factors affecting and techniques employed to cope with the quality of data

Students will [begin to] understand potential sources of data errors including measurement, encoding, derivation, and missing value problems. ***Presentations.***

- Students will [begin to] understand how bias (i.e. systematic data acquisition errors introduced by instruments , and numerical errors - introduced in data encoding and processing), become part of the dataset. ***Presentations.***

Ability to [begin to] organize, describe and manage data

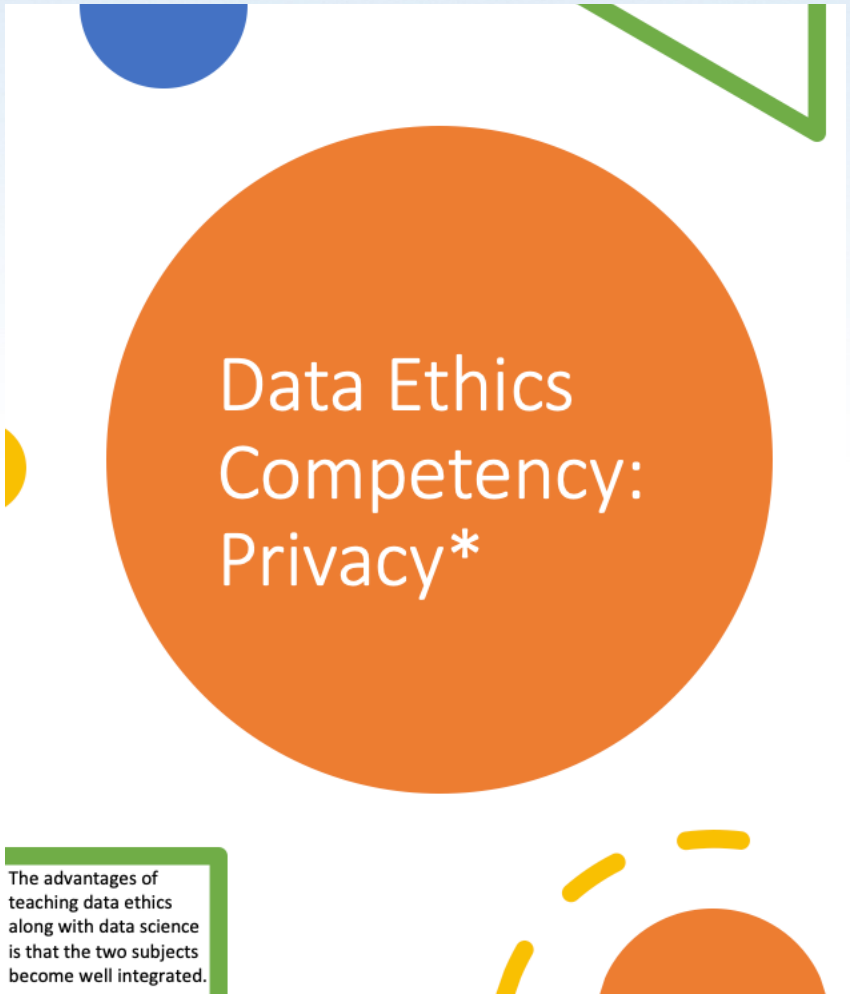
- Students will [begin to] understand that some data are made private, or are licensed and require attribution, while other datasets are open to the public domain. They will also be introduced to questions of data use related to intellectual property rights, privacy policies, and legal protections of some data such as the HIPPA requirements. ***Exercise 1.2***

Understanding of Data Preservation and Sharing

- Students will [begin to] understand the diverse motivations and barriers (technological , cultural, legal, ethical) associated with data sharing. ***Exercise 1.2***

*Quoted from **HPC University Educators.** "Computational and Data Science Education Competencies."
http://hpcuniversity.org/educators/competencies/

Module 2



Student Learning Outcomes (SLO): Demonstrate awareness of data privacy issues.*

Homework: Data's Ethical Component

Discussion Point: Should we take data privacy seriously? Do you personally take data privacy seriously?

- ***Organisation for Economic Cooperation and Development (OECD) Privacy Framework:** https://www.oecd.org/sti/ieconomy/privacy-guidelines.htm

Module 2

Student Learning Outcomes (SLO):

- "Demonstrate ***[some]*** competency with data modeling techniques (IBM Analytics, 6)." ***Homework***

Discussion Point: Is data science just about data analysis?

***IBM Analytics.** "The Data Science Skills Competency Model: A Blueprint for the Growing Data Scientist Profession." International Business Machines (IBM) Corporation. www.ibm.com/downloads/cas/7109RLQM

Module 2

Student Learning Outcomes (SLO):

Students will [begin to] create a conceptual model: ***Homework***

- Illustrate a conceptual modeling process through examples
- Identify the key parameters of the model
- Estimate model outcomes
- Utilize modeling software and/or spreadsheets to implement model algebraic equations (e.g. Vensim, Excel, MATLAB, Mathematica)
- Construct a simple computer visualization of the model results (e.g. infectious disease model, traffic flow, etc.)
- Validate the model with data
- Discuss model quality and the sources of errors*

*Quoted from **HPC University Educators.** "Computational and Data Science Education Competencies."
http://hpcuniversity.org/educators/competencies/

Module 2

Class Exercise:
A/B Testing

To get started, let's go to https://rdrr.io/snippets/ and copy/type in the first bulleted line below. Imagine the following situation: The numbers are the results from testing a new version of a website. The version A of the website received 600 views and 32 conversions (or clicks on the ad), while version B received 400 views and 44 conversions. Here is the test for Chi-squared that can tell us if the difference is statistically significant. What do the results tell us about which website version is better A or B (Cite: https://rpubs.com/JanpuHou/280223)?

You can copy the bullets into R directly: YOU DO NOT HAVE TO UNDERSTAND EVERY SYMBOL IN THE COMMAND – FOCUS ON UNDERSTANDING COLOR CODED ELMENTs!

- prop.test(c(**32**,**44**), c(**600**,**400**)) #this command performs a Chi-squared test
- #Note the p-value. In many social sciences, a p-value less than 0.05 is considered significant and the difference unlikely to have been generated by chance. What p-value do you see and what can you say about the difference between the websites?
- **#We can also use what we've learned to visualize the results.**
- Websites <- as.table(rbind(c(**32**,**44**), c(**600**,**400**))) #create a data table
- dimnames(Websites) <- list(Success = c("Conversions", "Views"),
- Versions = c("Website A","Website B")) #provide labels
- mosaicplot(~Versions + Success, data = Websites)
- **#What would you communicate to the website team – which website is better? Consider model quality and potential sources of error in your evaluation.**

Note: If your screen shows + press escape (this happens when you missed a symbol).

Module 3

Student Learning Outcomes (SLO): Demonstrate awareness and begin to evaluate Big Tech AI Principles.

Homework: Data's Ethical Component

Discussion Point: What is the role of Big Tech AI Principles?

- AI at Google: Our Principles https://www.blog.google/technology/ai/ai-principles/
- Microsoft AI: Responsible AI https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6
- Facebook AI: Facebook's Five Pillars of Responsible AI https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/

Module 3

Student Learning Outcomes (SLO):

- "Demonstrate ***[some]*** ability to visualize data and extract insights (IBM Analytics, 7)." ***Homework***

Discussion Point: Is data visualization just about insight sharing? Can data visualization help with analysis?

***IBM Analytics.** "The Data Science Skills Competency Model: A Blueprint for the Growing Data Scientist Profession." International Business Machines (IBM) Corporation. www.ibm.com/downloads/cas/7109RLQM

Module 3

Good Visualizations…

tell a good story	
with help of a narrative	Watch the following video to help you build a good narrative https://hbr.org/video/6069960867001/storytelling-with-data-a-good-charts-workbook-tool
awareness of the audience	The following link discusses the audience https://hbr.org/2013/04/the-three-elements-of-successf
and the use of clear labels	The following video shows how clear labels (titles, legends, units) can clarify and highlight the story https://www.youtube.com/watch?v=5Zg-C8AAIGg

Following are some resources to help you answer your questions about visualizations:

- Table or a Figure? https://writingcenter.unc.edu/tips-and-tools/figures-and-charts/
- What's considered good practice? https://porg/doi/10.1021/acsenergylett.1c00415,ubs.acs. https://pubs.acs.org/doi/10.1021/jz500997e
- Do you have any amazing examples? https://visme.co/blog/best-data-visualizations/.

Questions?