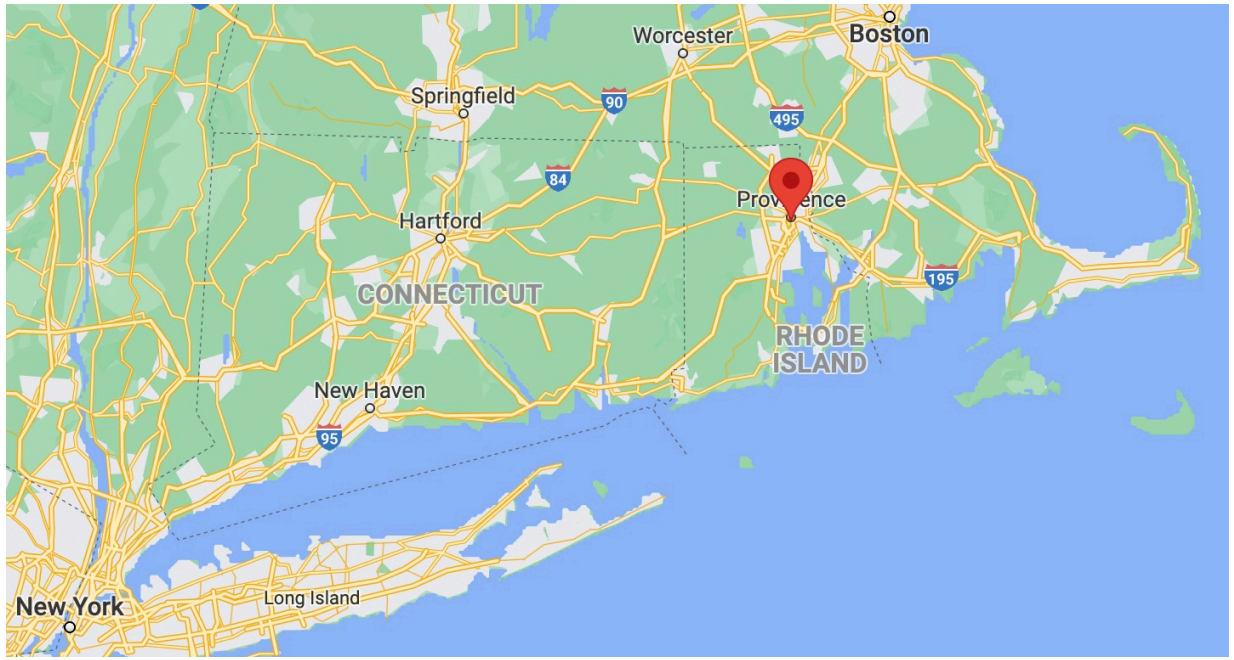
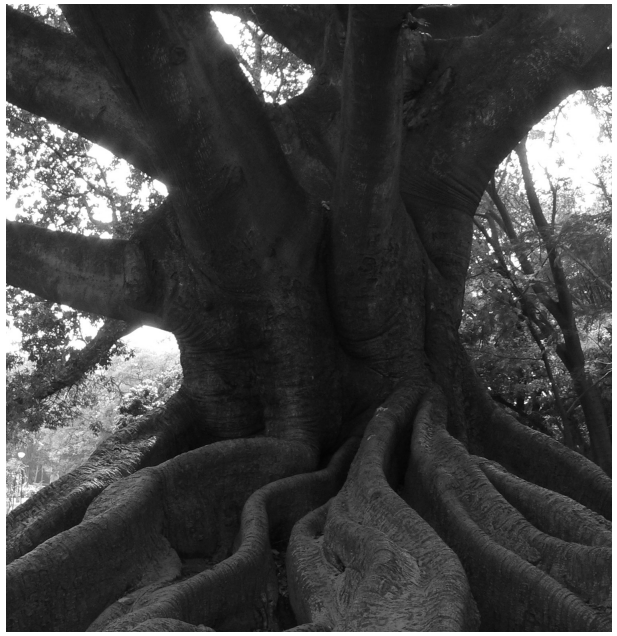
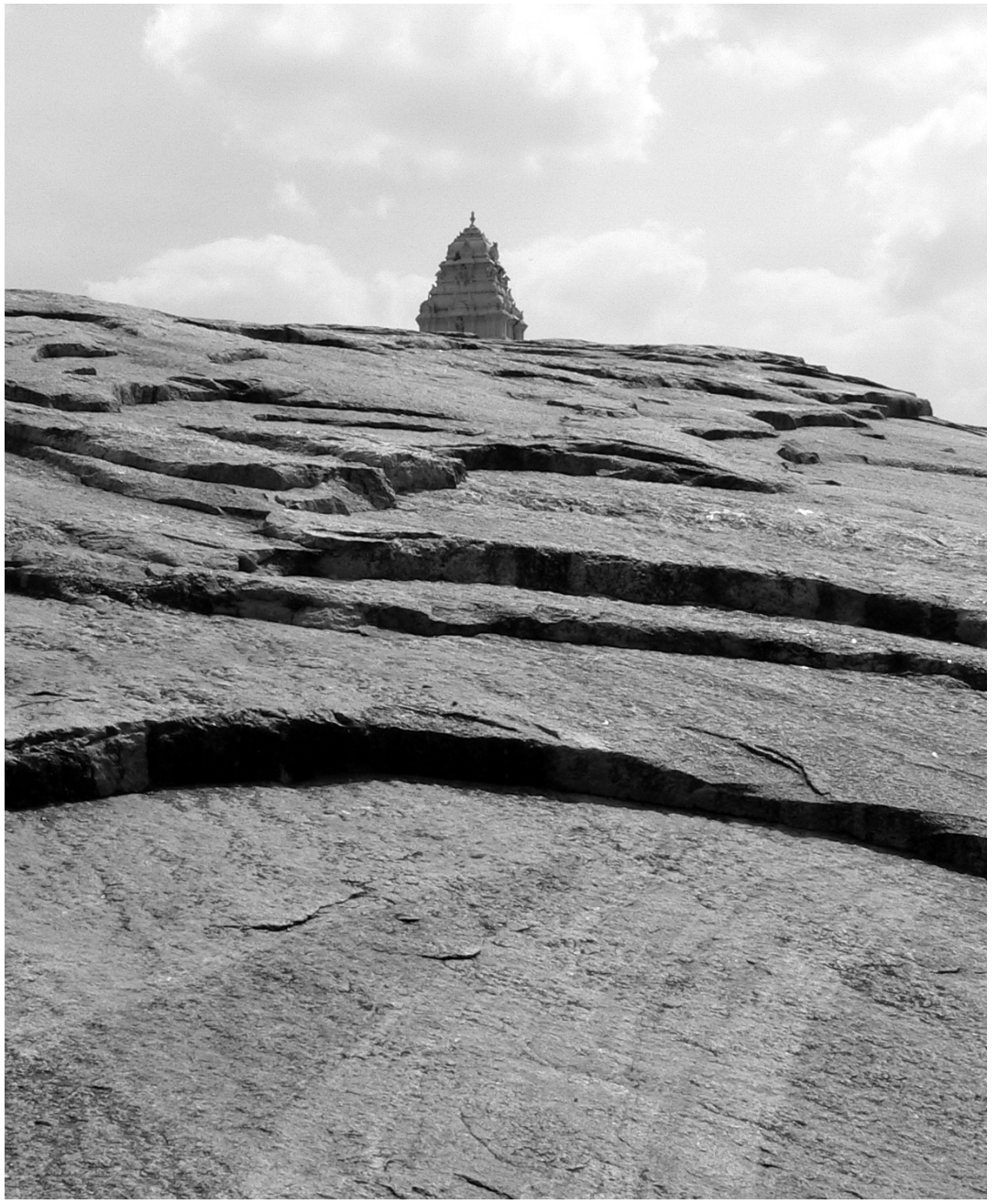
Data-Centricity: A Challenge and Opportunity for CS Education

Shriram Krishnamurthi, *Brown University*

Joint work with Kathi Fisler


Worcester
Boston
Springfield
90
495
84
Hartford
CONNECTICUT
RHODE
ISLAND
195
New Haven
95
New York
Long Island





What is in (Collegiate) CS1?

Numbers / Strings

Variables

Assignment statements

Conditionals

Loops

Functions

Arrays

Data Structure Progress Since the 1970s

1970s	Pascal	arrays
1980s	C	arrays
1990s	C++	arrays
2000-2015	Java	ArrayList
2015-now	Python	associative arrays

Pressure From Above:
“The Unreasonable Effectiveness of Data”

Pressure From Below:
Code Synthesis from Text

Pressure From Below:
“LowCode/NoCode” Ssytems

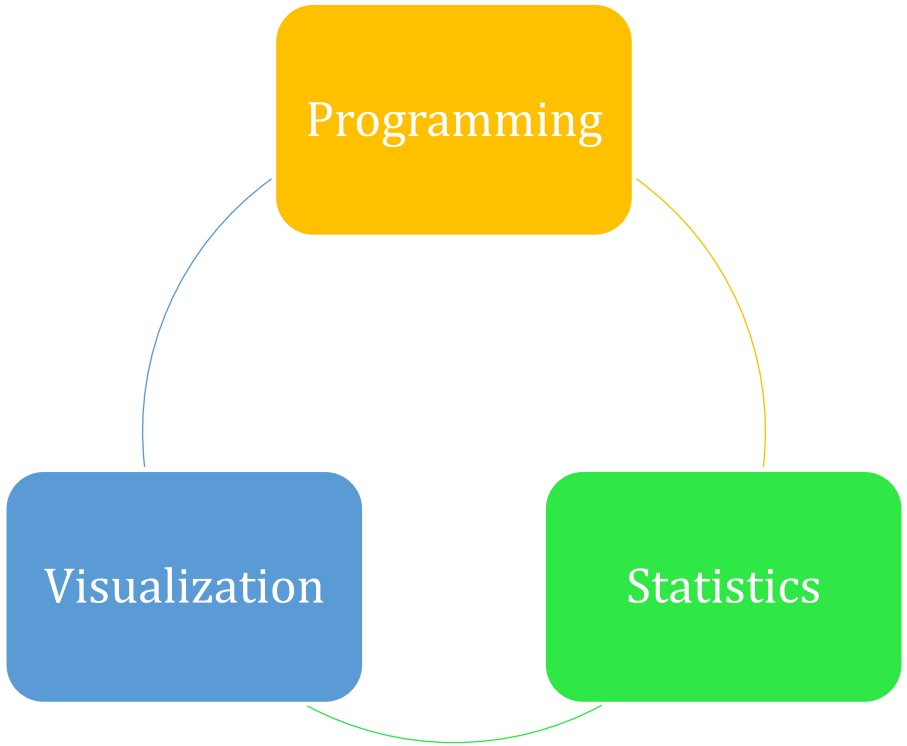
Needs from many more disciplines

Student diversity concerns

Negative societal impacts

Major new curricular trend:
Data Science

Courses, bootcamps, even degrees



How Should We View This?

“They don’t teach even the basics!”

“This is a threat to our degree programs!”

“This is actually a (valid) criticism of current CS Ed”

What Are the Valid Criticisms of CS 1?

Curricula do not engage students

Curricula do not connect to society/real-world phenomena

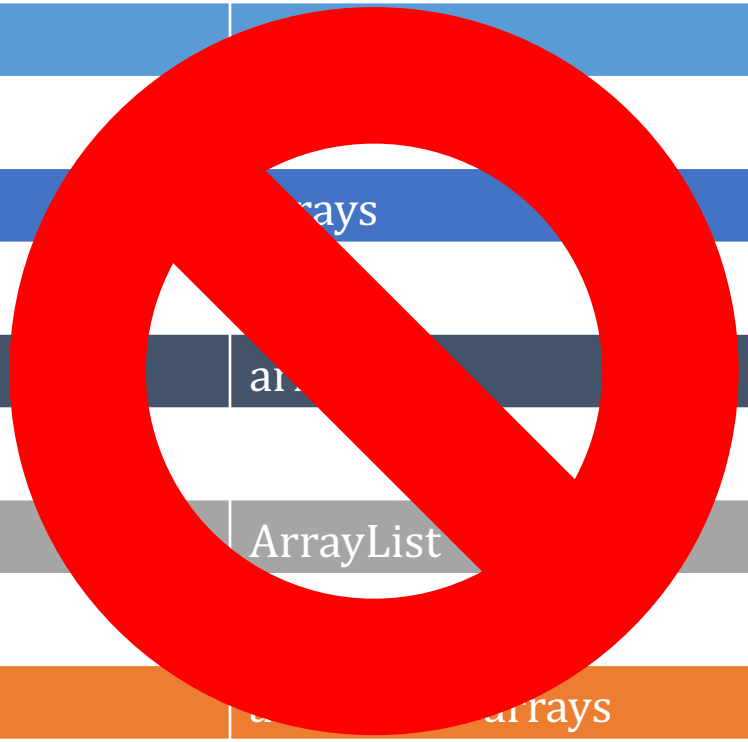
You can’t do anything useful after a course

Computing is used everywhere; students aren’t prepared

Reforming CS for the Data Era

Data Structure Progress Since the 1970s

1970s	Pascal	
1980s	C	rays
1990s	C++	ar
2000-2015	Java	ArrayList
2015-now	Python	arrays



First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

Rich structure

Already parsed

Familiar even to kids

MODELING AS A CORE COMPONENT OF STRUCTURING DATA

CLIFFORD KONOLD
University of Massachusetts Amherst
konold@srri.umass.edu

WILLIAM FINZER
Concord Consortium
wfinzer@concord.org

KOSOOM KREETONG
University of Massachusetts Amherst
kkreeton@educ.umass.edu

Time & Date	Segment Number	Direction of traffic	Width of Lanes	Type of vehicle	Speed (mph)	Distance to the vehicle in front (in ft)
8:00am 8/12/2010	03	westbound	15'	SUV	75	50'
				car	64	75'
8:00 am 8/12/2010	03	Eastbound	15'	car	24	15'
				SUV	12	12'
				car	12	8'
				car	12	8'
				truck	12	15'
4:00 pm 8/15/2010	08	westbound	20'	car	73	25'
				car	75	35'
				truck	50	50'
4:00 pm 8/15/2010	08	Eastbound	20'	car	85	60'

(763)

Graphs in order from front to back

8:00AM on 8/12/10
Segment 03 (W/E, 15' wide per side)

West

car type	speed (mph)	ft to next car
C	64	75
S	75	50

East

T	12	15
C	12	8
C	12	8
S	12	12
C	24	15

4:00PM on 8/15/10
Segment 08 (W/E, 20' wide per side)

West

car type	speed (mph)	ft to next car
T	50	50
C	75	35
C	73	25

East

C	85	60

First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

Data Commons

Data Commons aggregates data from a wide range of sources into a unified database to make it more accessible and useful. More on why we are building Data Commons.

Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum

Emmanuel Schanzer
Bootstrap
USA

Nancy Pfenning
University of Pittsburgh
USA

Flannery Denny
Bootstrap
USA

Sam Dooman
Brown University
USA

Joe Gibbs Politz
University of California, San Diego
USA

Benjamin S. Lerner
Northeastern University
USA

Kathi Fisler
Brown University
USA

Shriram Krishnamurthi
Brown University
USA

"I chose my dataset because…" Students could choose more than one option. Here, 42.6% said they chose it because it was something they already knew something about, while just as many (40.2%) chose it because "I don't know much about it, and I was curious". In addition, 16% said "It affects me personally". Numerous students used free-response to express specific interests such as representing a favorite hobby or activity, or "I am interested in the medical field". Still, small numbers contained a potentially negative response: 11.7% said they had no reason, 4.3% said it was their partner's choice, and 3.2% said they did not have a choice (presumably meaning the dataset was chosen by the teacher).

Overall, then, we see that the personalization is accompanied in general by a high degree of engagement and, most importantly, low degrees of *dis*-engagement. This suggests that the technique of providing a (limited) range of curated datasets still gives many students room for personal expression.

First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

"How many tickets sold with a student discount?"

Interesting problems

Concrete, easy to envision and manipulate

Lend themselves to decomposition and planning

Computing Education Research Blog

A Challenge to Computing Education Research: Make Measurable Progress

August 16, 2010 at 9:50 pm | 41 comments

In computing education research, we repeatedly conduct studies showing just how little students are learning in our classes. We've been doing it for almost 30 years. I have a challenge to our community: ***Let's show that we can do better!***

After 30 years, why hasn't somebody *beaten* the Rainfall Problem? Why can't someone teach a course with the ***explicit*** goal of their students doing ***much*** better on the Rainfall Problem — then publish how they did it? *We ought to make measurable progress.*

I don't think that this is an impossible goal. In fact, I bet that some of the existing research projects in computing education could "beat" (generate published reports with better results) these current studies.

- The TeachScheme approach focuses on design based on data. I bet that their students could beat the Rainfall Problem or the McCracken working group problem.

First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

Variety of programming forms (functions, methods, loops, …)

```
t1 = sieve tickets using discount:
  discount == "student"
end

sum(extract tickets from t1 end)
```

Pyret

First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

Data quality, normalization, cleansing

Consequences of bad data

First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

Falsehoods Programmers Believe About Names

First Name	Last Name	Email	Tickets	Discount	Birthday
Josie	Zhao	josie	2	birthday	Jan 1, 1970
Guy Lewis	Steele Jr	guy@steele.com	4		
Bart	Simple	bart@simpsons.com	1	student	
Robby		robby	2		
Shweta	Mysore	shweta@np.org	1		
Alvina	Velasquez	avlina@school.edu	three	student	
Sam	Ochibe	s@sweb.com	10		

What gets collected about you?

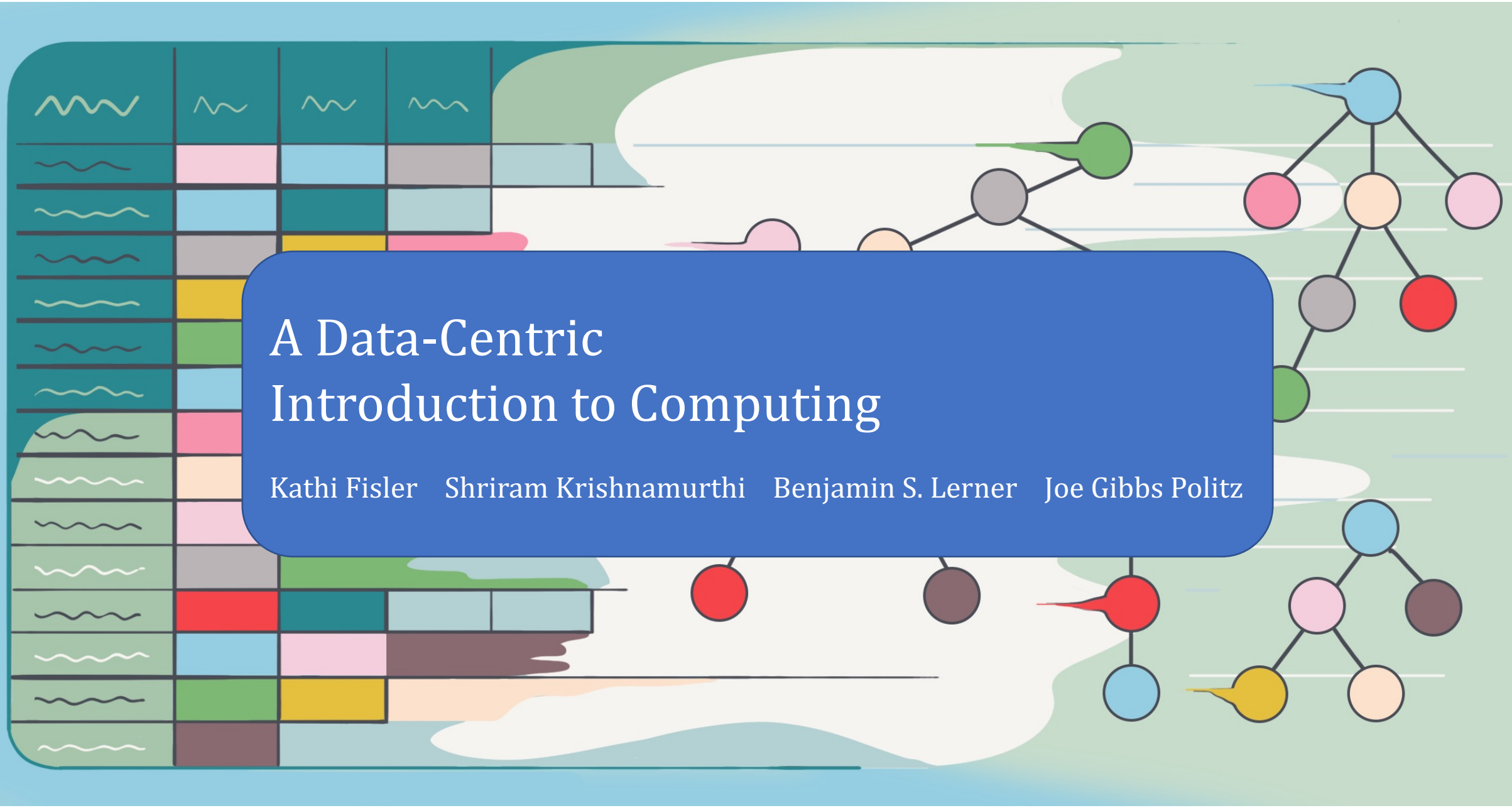
What gets *inferred* about you?

What *decisions* are made on those inferences?

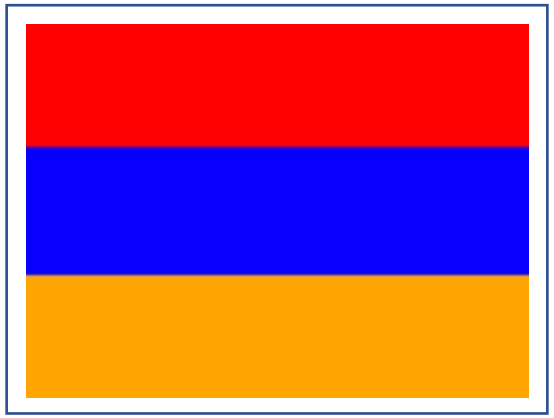
Limitations

```
ancestry = table: name, birthyear, eyecolor, fempar, malepar
    row: "Anna", 1997, "blue", "Susan", "Charlie"
    row: "Susan", 1971, "blue", "Ellen", "Bill"
    row: "Charlie", 1972, "green", "NoInfo", "NoInfo"
    row: "Ellen", 1945, "brown", "Laura", "John"
    ...
end
```

Motivate new data structures (and rest of CS)


A Data-Centric
Introduction to Computing
Kathi Fisler Shriram Krishnamurthi Benjamin S. Lerner Joe Gibbs Politz

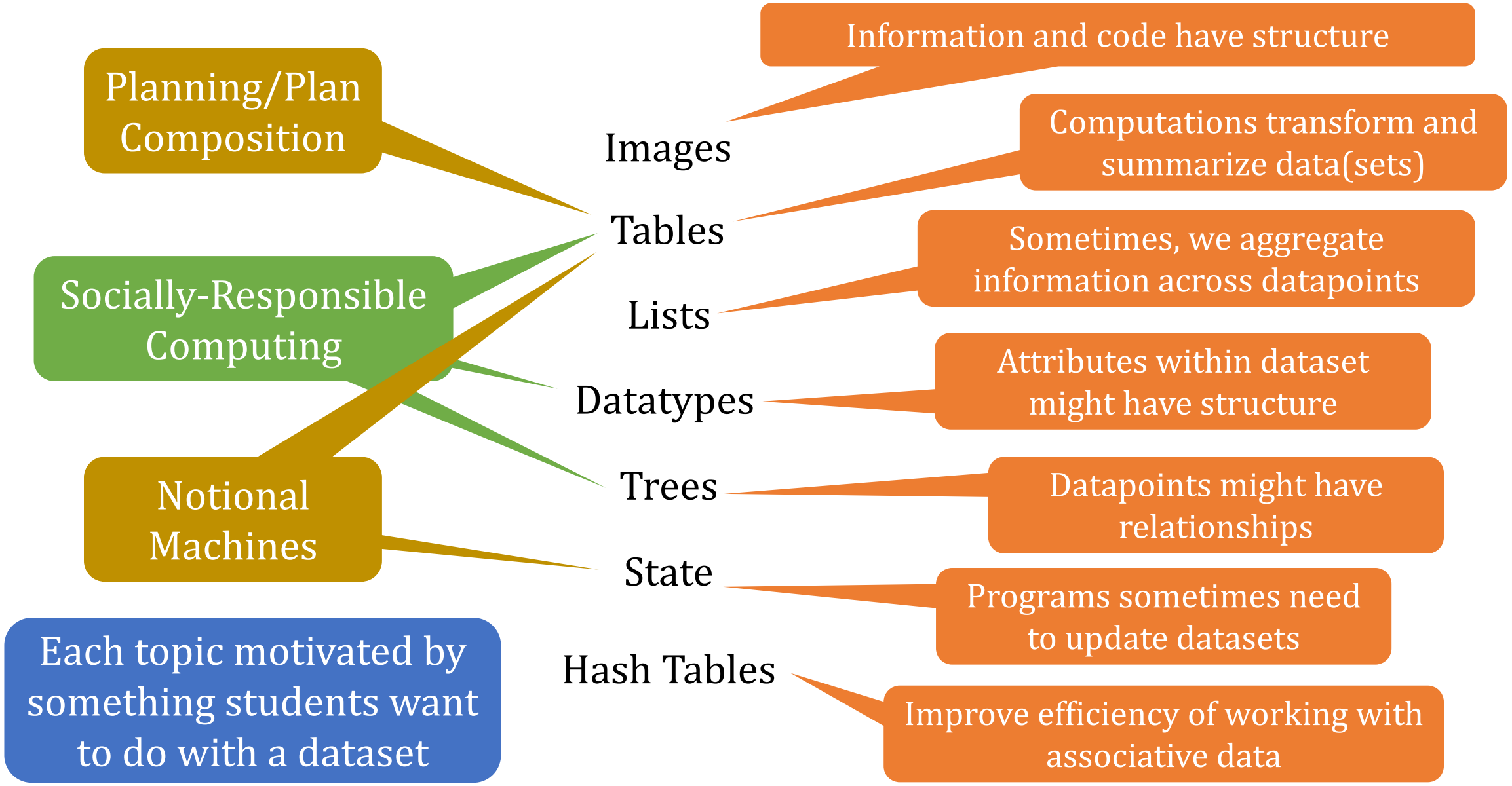
Data-Centric Intro to Computing (DCIC)



```
frame(
  above(rectangle(200, 50, "solid", "red"),
    above(rectangle(200, 50, "solid", "blue"),
      rectangle(200, 50, "solid", "orange"))))
```

Structure of code follows structure of image

Data-Centric Intro to Computing (DCIC)



Research-Driven Curriculum!

Using Design Alternatives to Learn About Data Organizations

Xingjian Gu Max A. Heller Stella Li Yanyan Ren Kathi Fisler Shriram Krishnamurthi
Contact:sk@cs.brown.edu
Computer Science Department
Brown University
Providence, RI, USA

Developing Behavioral Concepts of Higher-Order Functions

Shriram Krishnamurthi
Brown University
Providence, RI, USA

Kathi Fisler
Brown University
Providence, RI, USA

What Help Do Students Seek in TA Office Hours?

Yanyan Ren
Computer Science Department
Brown University
Providence, Rhode Island, USA

Shriram Krishnamurthi
Computer Science Department
Brown University
Providence, Rhode Island, USA

Kathi Fisler
Computer Science Department
Brown University
Providence, Rhode Island, USA

Adapting Student IDEs for Blind Programmers

Emmanuel Schanzer
schanzer@bootstrapworld.org
Brown University / Bootstrap
Providence, RI, USA

Sina Bahram
sina@sinabahram.com
Prime Access Consulting
Cary, NC, USA

Shriram Krishnamurthi
schanzer@bootstrapworld.org
Brown University / Bootstrap
Providence, RI, USA

Early Post-Secondary Student Performance of Adversarial Thinking

Nick Young
Brown University
Providence, RI, USA

Shriram Krishnamurthi
Brown University
Providence, RI, USA

Article


COLUMN

Data-centricity: a challenge and opportunity for computing education

Authors: Shriram Krishnamurthi, Kathi Fisler Authors Info & Affiliations

Publication: Communications of the ACM • July 2020 • https://doi.org/10.1145/3408056

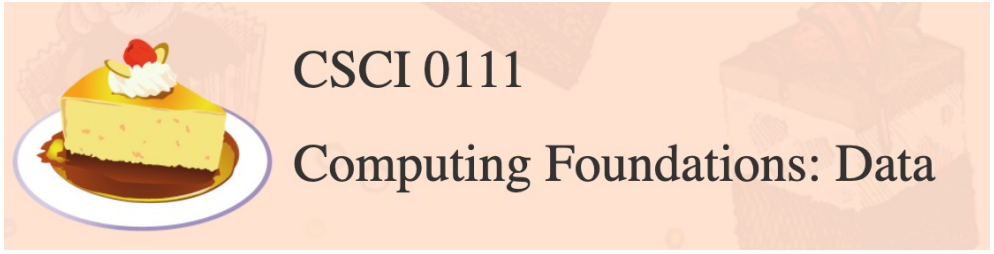
CACM July 2020

Book

All materials are free and on-line!

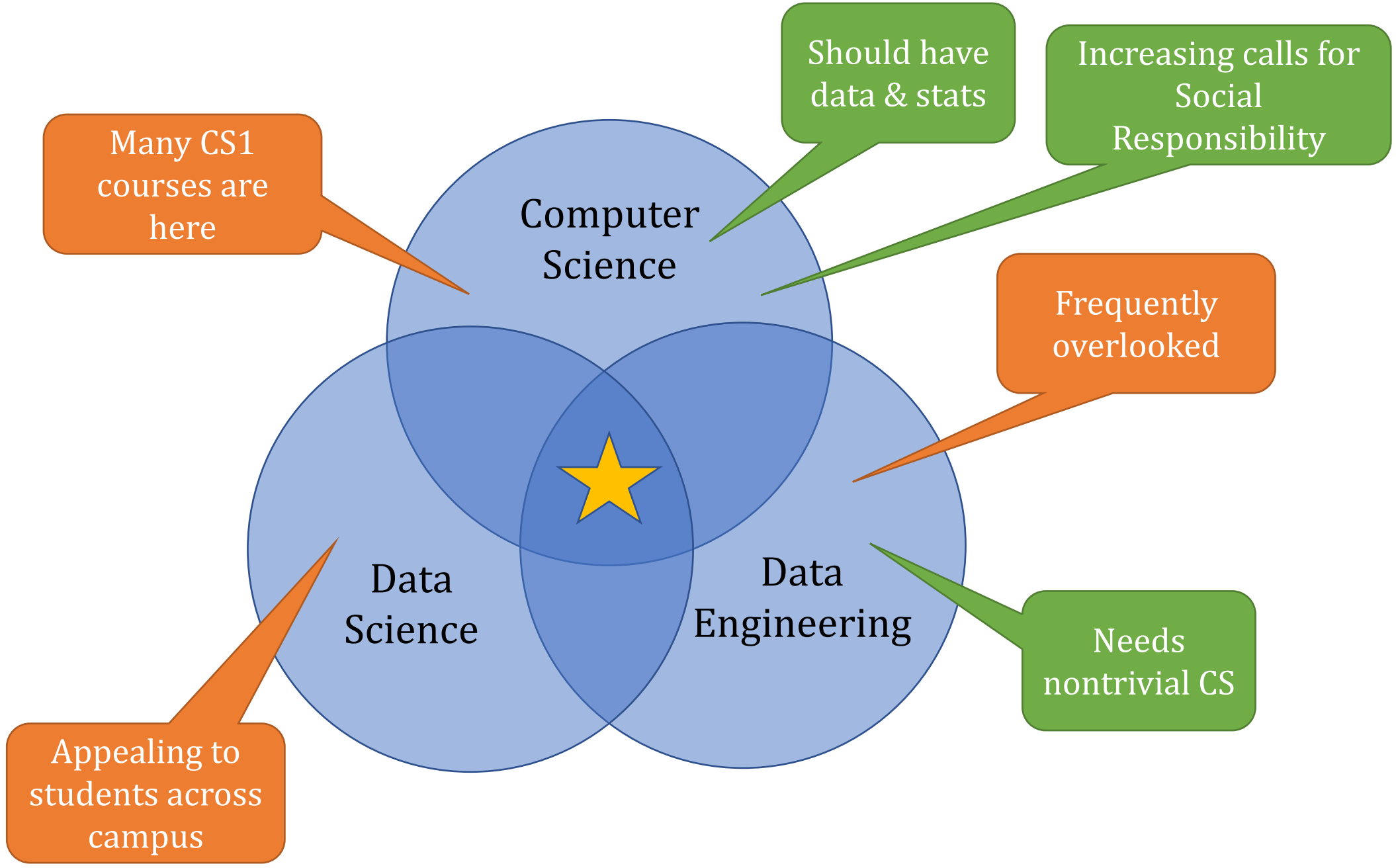
`dcic-world.org`

Course



(at Brown & others)

Stepping Back Out


Should have data & stats
Increasing calls for Social Responsibility
Many CS1 courses are here
Computer Science
Frequently overlooked
Data Science
Data Engineering
Needs nontrivial CS
Appealing to students across campus

What About Primary/Secondary?

Ports all these ideas to students as young as 10yo

State of CS 1

The Rise of Data Science

Redesigning CS 1 with Data