Functional Programming: the Glue for Introducing Computing through Data Science

Kathi Fisler
Brown University and Bootstrap
(Pyreteer and Racketeer)
joint work with Shriram Krishnamurthi

“Language Wars” have been around for some time

tooling

syntax

problem styles

real-world relevance

University Computer Science/Industry Programmers

Some pre-college education
(C++, Visual Basic, Java)

Expanded pre-college education
(Scratch, AppInventor, Javascript, ...)

Data Science!

What's the best language for teaching data science?

What problem are you trying to solve?

A challenge for programming education

Not just the “natural” programmers

Regardless of intended major

Provide computing and data science education to all students

Pedagogies that support students with different skills

in ways that support equity and diversity in computing

Problems that resonate across cultures

Without enough CS/Informatics teachers to go around

while working within the constraints of schools and university departments

Acknowledge impacts of computing on people and society

A Data-Centric Approach to Computing via Functional Programming

I come at this from …

BOOTSTRAP
www.bootstrapworld.org

BROWN

Computing Education, grades 6-16

author of K-12 CS standards in multiple US states

CS department administration of a large undergrad program

Researcher in computing education, formal methods, and verification

Data Centric?

Not just the “natural” programmers

Regardless of intended major

Without enough CS/Informatics teachers to go around

lead with data,
not control operators

lead with data that students recognize and care about

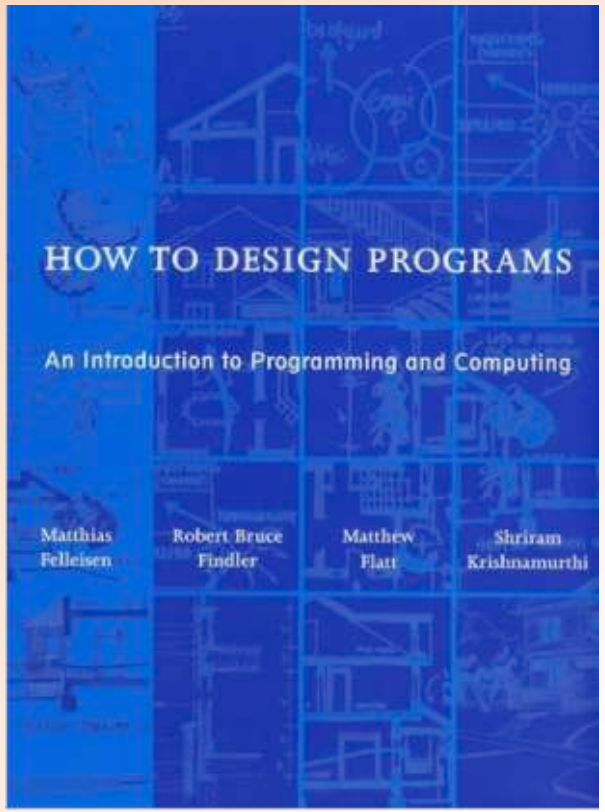
lead with questions that teachers from other disciplines care about

Pedagogies that support students with different skills

Problems that resonate across cultures

Acknowledge impacts of computing on people and society

Not just any old functional programming!



Atomics (nums, string, images)

Structs

Lists

List[Structs]

Trees

Mutual Recursion

Generative Recursion

State

Ooh! Datasets!

Could we do this earlier?

Questions like
"how many tickets sold with a student discount"
let students explore problem decomposition in a concrete, physical format

Task Planning!

Rich, structured data, in a familiar format!

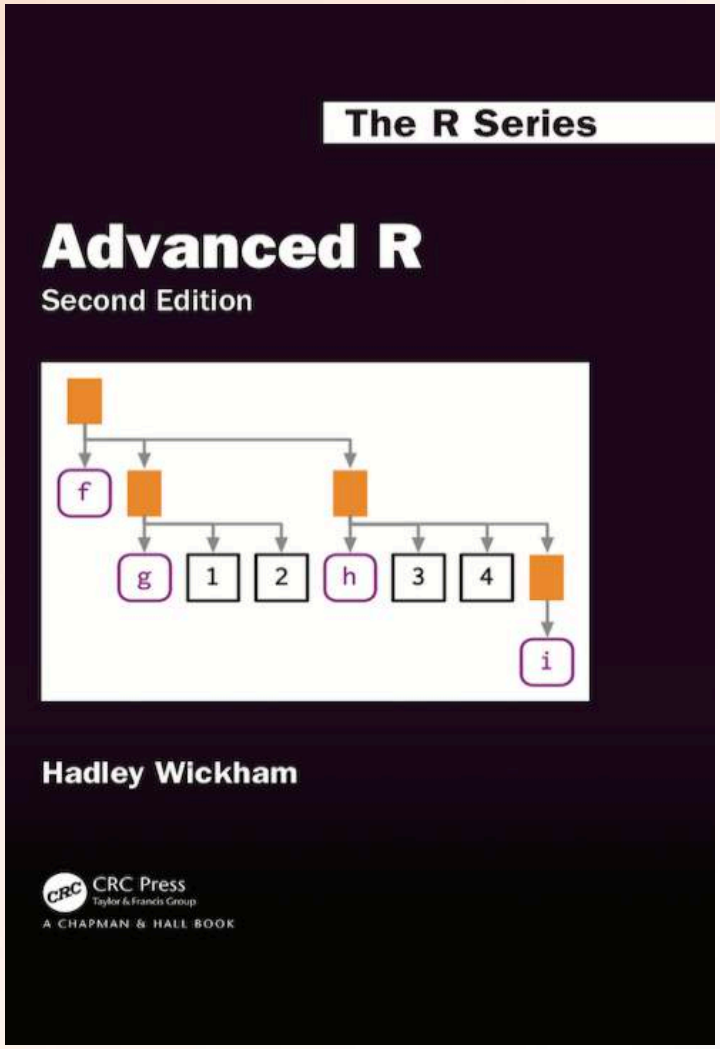
	A	B	C	D	E
1	**Name**	**Email**	**Num Tickets**	**Discount Code**	**Delivery**
2	Josie Zhao	jo@mail.com	2	BIRTHDAY	email
3	Sam Ochibe	s@sweb.com	1		pickup
4	Bart Simple	bart@simpson.org	5	STUDENT	yes
5	Ernie O'Malley	ernie.mail.com	0	none	email
6	Alvina Velasquez	alvie@schooledu	3	student	emall
7	Zander	zandaman	10		email
8	Shweta Chowpatti	snc@this.org	three		pickup

Many authentic tasks that can raise impact issues

As much data engineering as data science

Recipe for preparing data for analysis: normalize, locate suspicious data, use visualization to sanity check, analyze

Functional programming underlies tools for processing tabular-data



Introduction

R, at its heart, is a **functional** language. This means that it has certain technical properties, but more importantly that it lends itself to a style of problem solving centred on functions. Below I'll give a brief overview of the technical definition of a functional *language*, but in this book I will primarily focus on the functional *style* of programming, because I think it is an extremely good fit to the types of problem you commonly encounter when doing data analysis.

Recently, functional techniques have experienced a surge in interest because they can produce efficient and elegant solutions to many modern problems. A functional style tends to create functions that can easily be analysed in isolation (i.e. using only local information), and hence is often much easier to automatically optimise or parallelise. The traditional weaknesses of functional languages, poorer performance and sometimes unpredictable memory usage, have been much reduced in recent years. Functional programming is complementary to object-oriented programming, which has been the dominant programming paradigm for the last several decades.

Data-Centric Intro to Computing (DCIC)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

Task Planning

Socially-Responsible Computing

Atomics (nums, string, images)

Structs

Lists

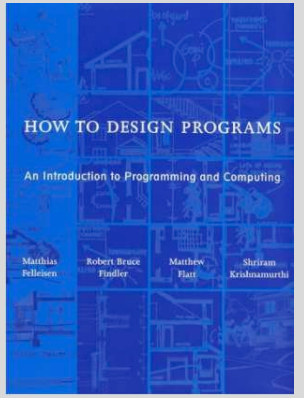
List[Structs]

Trees

Mutual Recursion

Generative Recursion

State



Data-Centric Intro to Computing (DCIC)

Images

Tables

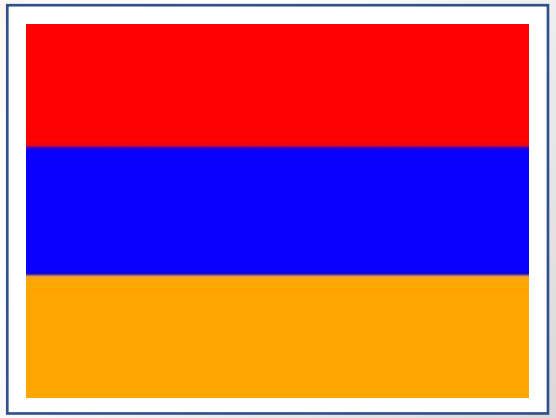
Lists

Datatypes

Trees

State

Hashtables



```
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)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

	A	B	C	D	E
1	**Name**	**Email**	**Num Tickets**	**Discount Code**	**Delivery**
2	Josie Zhao	jo@mail.com	2	BIRTHDAY	email
3	Sam Ochibe	s@sweb.com	1		pickup
4	Bart Simple	bart@simpson.org	5	STUDENT	yes
5	Ernie O'Malley	ernie.mail.com	0	none	email
6	Alvina Velasquez	alvie@schooledu	3	student	emall
7	Zander	zandaman	10		email
8	Shweta Chowpatti	snc@this.org	three		pickup

```
tickets = table: name, email, num, discount, delivery
  row: “Josie Zhao”, “jo@email.com”, 2, “BIRTHDAY”, “email”
  row: “Sam Ochibe”, “s@web.com”, 1, “”, “pickup”
  ...
end

build-column(tickets, “fee”, lam(r :: Row): r[“num”] * 10 end)
```

Or import from Google Drive

Data-Centric Intro to Computing (DCIC)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

	A	B	C	D	E
1	**Name**	**Email**	**Num Tickets**	**Discount Code**	**Delivery**
2	Josie Zhao	jo@mail.com	2	BIRTHDAY	email
3	Sam Ochibe	s@sweb.com	1		pickup
4	Bart Simple	bart@simpson.org	5	STUDENT	yes
5	Ernie O'Malley	ernie.mail.com	0	none	email
6	Alvina Velasquez	alvie@schooledu	3	student	emall
7	Zander	zandaman	10		email
8	Shweta Chowpatti	snc@this.org	three		pickup

Extract a column

Higher-order functions (resembles those on tables)
then introduce (only) structural recursion

Data-Centric Intro to Computing (DCIC)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

	A	B	C	D	E
1	**Name**	**Email**	**Num Tickets**	**Discount Code**	**Delivery**
2	Josie Zhao	jo@mail.com	2	BIRTHDAY	email
3	Sam Ochibe	s@sweb.com	1		pickup
4	Bart Simple	bart@simpson.org	5	STUDENT	yes
5	Ernie O'Malley	ernie.mail.com	0	none	email
6	Alvina Velasquez	alvie@schooledu	3	student	emall
7	Zander	zandaman	10		email
8	Shweta Chowpatti	snc@this.org	three		pickup

How to represent timestamps?
string? number?

Many opportunities to discuss data-design tradeoffs and connect to real-world issues

(e.g., storing lists and structs in CSV files in systematic ways, "falsehoods programmers believe about names/dates")

Data-Centric Intro to Computing (DCIC)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

```
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
```

Tables aren’t always a useful data structure

Challenge of searching for ancestors highlights the need for data structures beyond tables (here comes CS …)

Data-Centric Intro to Computing (DCIC)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

Back to tables! (pandas)

Positive!

Get it?

Tables (could easily build into Racket)

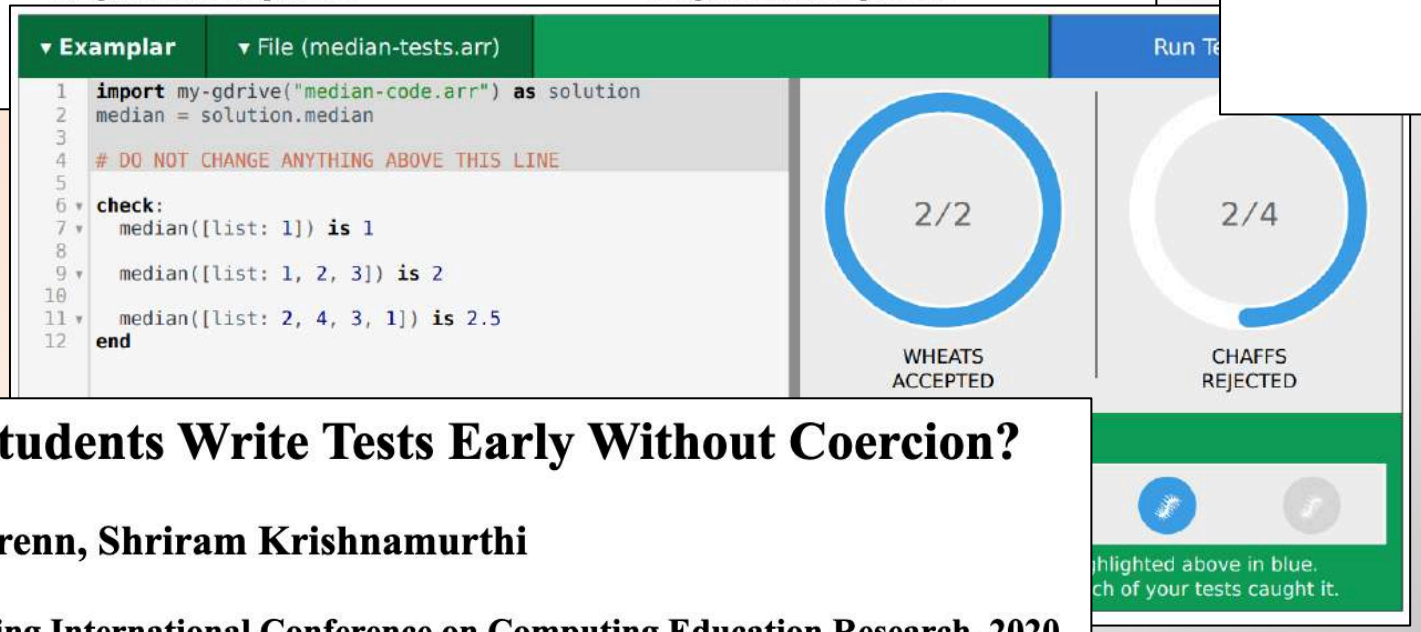
Examples part of function definitions, separate from tests

Python-esque syntax (we know, but ...)

Executable Examples for Programming Problem Comprehension

John Wrenn
Computer Science Department

Shriram Krishnamurthi
Computer Science Department



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

Data Design Druid

Will Students Write Tests Early Without Coercion?

John Wrenn, Shriram Krishnamurthi

Koli Calling International Conference on Computing Education Research, 2020

Driven by CS Education Research

Data-Centric Intro to Computing (DCIC)

Images

Tables

Lists

Datatypes

Trees

State

Hashtables

Back to tables! (pandas)

CSCI 0111

Computing Foundations: Data

Textbook in progress (Mar 2021)

550 students so far

non-CS majors LOVE it

many have become/added CS

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

Checkpoint!

Not just the “natural” programmers

Regardless of intended major

Without enough CS/Informatics teachers to go around

lead with data,
not control operators

lead with data that students recognize and care about

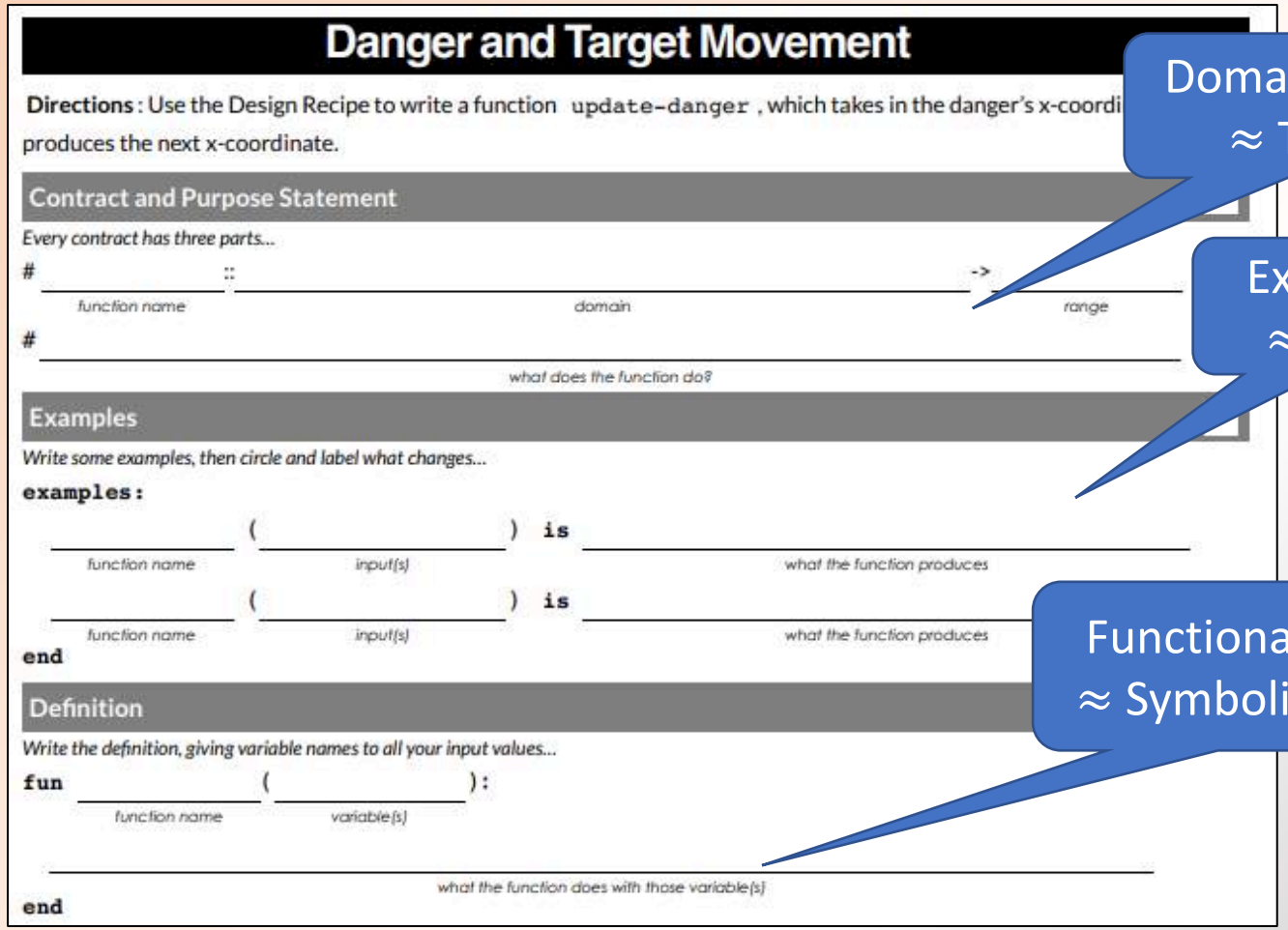
lead with questions that teachers from other disciplines care about

Pedagogies that support students with different skills

Problems that resonate across cultures

Acknowledge impacts of computing on people and society

Computing and Data-Science in K-12



Danger and Target Movement

Directions : Use the Design Recipe to write a function `update-danger` , which takes in the danger's x-coordi
produces the next x-coordinate.

Contract and Purpose Statement

Every contract has three parts...

\# _____ :: _____ -> _____
function name *domain* *range*

\# _____
what does the function do?

Examples

Write some examples, then circle and label what changes...

examples:

_____ (_____) **is** _____
function name *input(s)* *what the function produces*

_____ (_____) **is** _____
function name *input(s)* *what the function produces*

end

Definition

Write the definition, giving variable names to all your input values...

fun _____ (_____):
function name *variable(s)*

_____
what the function does with those variable(s)

end

Domain/Range ≈ Types!

Examples ≈ Tests!

Functional Code ≈ Symbolic Form

BOOTSTRAP
www.bootstrapworld.org

(thanks, Philip Wadler)

Integrate intro computing/data science into existing pre-college classes (algebra, science, social studies)

Questions and projects that matter in the host discipline

Assess learning in the host discipline

With teachers new to computing

Also backed by research ...

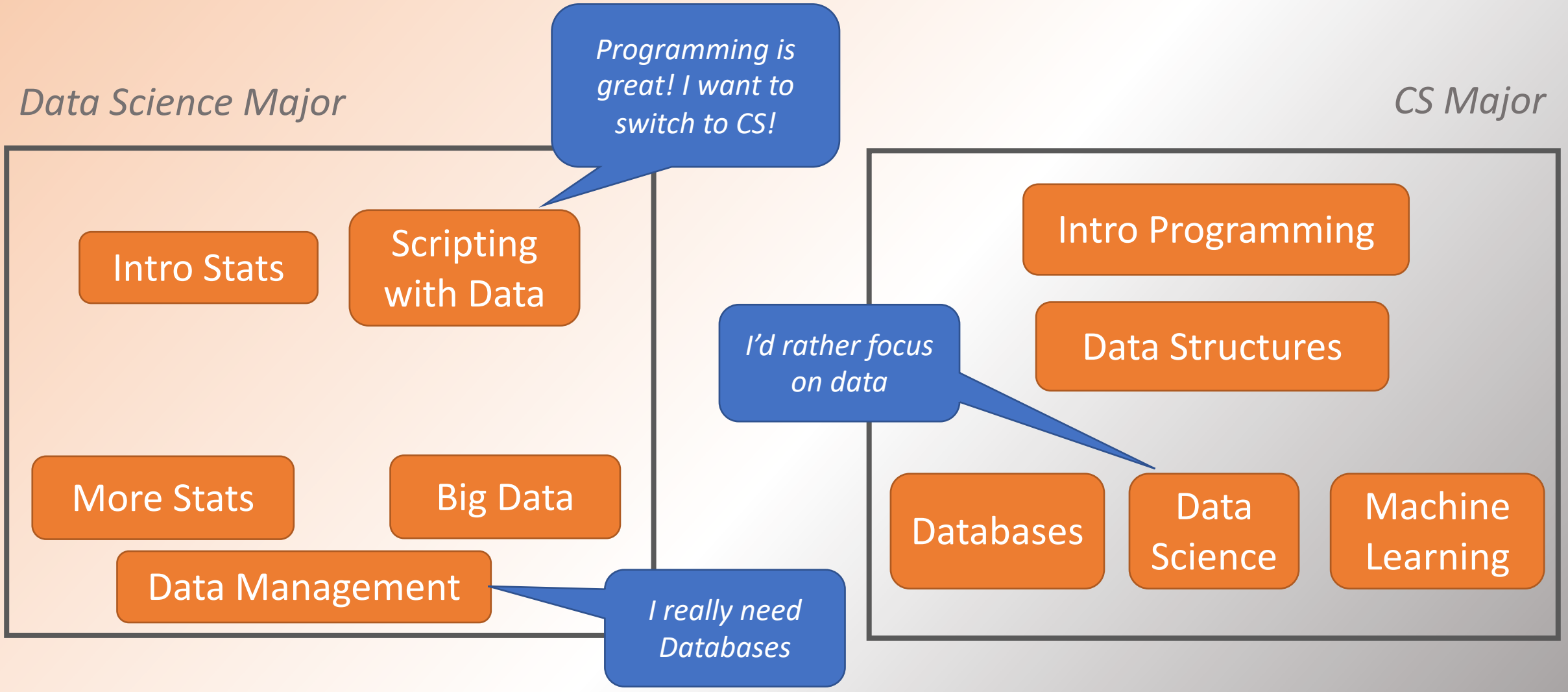
Leading from data supports computing for all

- Enables authentic tasks in many fields
- Raises impacts of computing, which matters for equity
- Can accomplish a lot with small amounts of code

Functional programming can get us there!

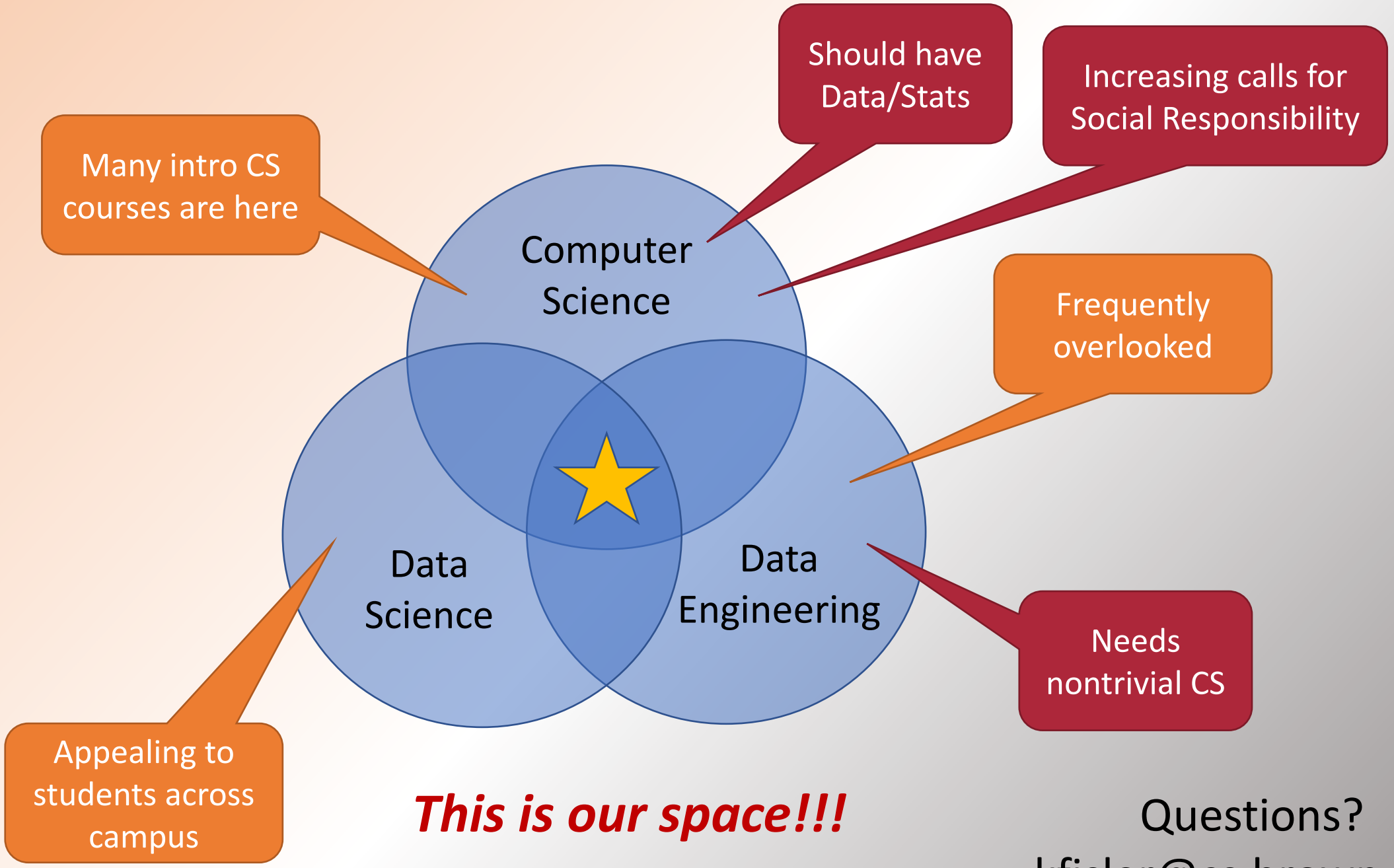
Data science gives us a new foothold

but the linguistic and pedagogic details matter
A LOT



Little content alignment, so switching requires starting over ☹

Novices don't understand these fields well enough to decide!



This is our space!!!

Questions?

kfisler@cs.brown.edu