COMPUTING WITH CORGIS: DIVERSE, REAL-WORLD DATASETS FOR INTRODUCTORY COMPUTING

Austin Cory Bart, Ryan Whitcomb, Dennis Kafura, Clifford A. Shaffer, Eli Tilevich

Virginia Tech
Overview

• Bringing real-world data into introductory computing classes
• Via a new system that manages and produces datasets
• In order to motivate non-computing majors
Computer Science For All
Diverse Majors

Animal Sciences
Chemistry
Theater Arts
Building Construction
English
Biological Sciences
Education
History

... with Rich Knowledge
(1) No Prior Background

“I’ve never done this before.”
(2) Low Self-efficacy

“I have no idea how to do this!”
(3) Unclear on Why

“Why am I doing this?”
MUSIC Model of Academic Motivation

Students are more motivated when they perceive that:

1. they are **empowered,**
2. the content is **useful** to their goals,
3. they can be **successful,**
4. they are **interested,** and
5. they feel **cared** for by others in the learning environment

Motivation ➔ Engagement

Motivation:
- eMpowerment
- Usefulness
- Success
- Interest
- Caring

Engagement Outcomes:
- Persistence
- Proactivity
- Attendance
- Learning
- ...

Interest
A spectrum

Content

FOR-EACH
Recursion
IF
Assignment
Data Structures
Dictionaries
Integers
Booleans

WHILE
Iteration
Algorithms
Arrays
Lists

Context

Games
Websites
Development
Mobile Apps
Images
Audio
Animations
Scientific Modelling
Math
Scientific Computing

10
Authenticity

• Situated Learning
• “Relevant”, “Real-world”
• Media Computation as an “Imagineered Authentic Experience”

Why *are* we teaching computing?

“A Tidal Wave of Data”
State of the Art

• Bart 2014 – Connecting to real-time APIs (RealTimeWeb)
• Hamid 2016 – More generalized framework for real-time APIs (Sinbad)
• Subramanian 2014 – Visualization of data structures with real data (BRIDGES)
• Anderson 2014 – Real world data in CS1
• Sullivan 2013 – Data Science for non-majors
Problem – We Need Data

• ICPSR – Tightly controlled datasets
• UCI Machine Learning – Only for machine learning
• Census.gov, Kaggle, etc. – Not ready for beginners
CORGIS

The Collection Of Really Great, Interesting, Situated Datasets
Metrics

42 datasets

267 mB

420,672 rows

9,365,520 values
Datasets
Connecting to Students’ Majors

- Crime
- Criminal Justice
- Aerospace
- Theater
- Theater Arts
- Construction
- Building Construction
- English
- Geological Science
- Education
- History
- Immigration
- Books
- Weather
- Education
Architecture

Raw Datasets
- CSV: ABC, XYZ, 173, 441, 0, 27, ...
- XML: ?xml version
  <catalog>
    <book>
      ...
  ...
- TXT: Never gonna give you up, never gonna let you down...
- HTML: <html>
  <body>
    <div>
      ...

Cleaned Data
- JSON: [
  {...},
  ...
]
- CORGIS: Metadata:
  ... Interfaces:
  ... Databases:

Final Datasets
- SQL
- CSV
- JSON
- BlockPy

Preprocess - Manual
Build - Automatic
Specification File
Java, Python, Racket

// Java
import corgis.crime.StateCrimeLibrary;
import corgis.crime.domain.Report;
import java.util.ArrayList;
public class Main {
    public static void main(String[] args) {
        StateCrimeLibrary scl = new StateCrimeLibrary();
        ArrayList<Report> reports = scl.getAll();
    }
}

; Racket
(require crime)
(define reports (crime-get-all))

# Python
import crime
crime_reports = crime.get_all()
BlockPy
Hypotheses

• Context provides motivation
• Students have some preference for Data Science
• The usefulness of the context connects to engagement outcomes as strongly as the content
Interventions

- Computational Thinking Course
  - Basic programming
  - Social Impacts
  - Data Science

- 6 semesters taught

- Audience
  - Non-computing majors
  - Freshmen -> Senior
  - Gender balanced
## Motivation × Course Components

<table>
<thead>
<tr>
<th>Motivational Components</th>
<th>Course Component</th>
<th>Likert</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I believe that I will have freedom to explore my own interests when I…”</td>
<td>&quot;... learn to write computer programs&quot;</td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td>&quot;I believe it will be useful to my long-term career goals to…”</td>
<td>&quot;... learn to work with abstraction&quot;</td>
<td>Disagree</td>
</tr>
<tr>
<td>“I believe I will be successful in this course when I…”</td>
<td>&quot;... learn about the social impacts of computing&quot;</td>
<td>Somewhat Disagree</td>
</tr>
<tr>
<td>“I believe it will be interesting to…”</td>
<td>&quot;... work with real-world data related to my major&quot;</td>
<td>Neither Agree nor Disagree</td>
</tr>
<tr>
<td>“I believe that my instructors and peers will care about me when I…”</td>
<td>&quot;... work with my cohort&quot;</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

- **eMpowerment**: "I believe that I will have freedom to explore my own interests when I…”
- **Usefulness**: "I believe it will be useful to my long-term career goals to…”
- **Success**: "I believe I will be successful in this course when I…”
- **Interest**: "I believe it will be interesting to…”
- **Caring**: "I believe that my instructors and peers will care about me when I…”
N = 85, 62% Female

Students’ sense of the usefulness of various course components was highest for the context, lowest for the content.
## Preference for Contexts

<table>
<thead>
<tr>
<th>Preference for Contexts</th>
<th>Likert</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Working with data sets related to your major”</td>
<td>Strongly Avoid</td>
</tr>
<tr>
<td>“Working with pictures, sounds, movies”</td>
<td>Avoid</td>
</tr>
<tr>
<td>“Making games and animations”</td>
<td>Somewhat Avoid</td>
</tr>
<tr>
<td>“Making websites”</td>
<td>Neither Prefer nor Avoid</td>
</tr>
<tr>
<td>“Making scientific models of real-world phenomenon”</td>
<td>Somewhat Prefer</td>
</tr>
<tr>
<td>“Controlling robots or drones”</td>
<td>Prefer</td>
</tr>
<tr>
<td>“Making phone apps”</td>
<td>Strongly Prefer</td>
</tr>
</tbody>
</table>
Preference for Contexts

N = 85, 62% Female

Students’ preferred a Data Science context over all others

* No significant difference with Media Computation according to KW test
Engagement (Intent to Continue)

<table>
<thead>
<tr>
<th>Intent to Continue</th>
<th>Likert</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I will try to learn more about computing, either through a course or on my own.”</td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td>“I will recommend this class to others.”</td>
<td>Disagree</td>
</tr>
<tr>
<td>“I will directly apply what I have learned in my career.”</td>
<td>Somewhat Disagree</td>
</tr>
<tr>
<td>“I will recommend this class to others.”</td>
<td>Neither Agree nor Disagree</td>
</tr>
<tr>
<td>“I will directly apply what I have learned in my career.”</td>
<td>Somewhat Agree</td>
</tr>
<tr>
<td>“I will recommend this class to others.”</td>
<td>Agree</td>
</tr>
<tr>
<td>“I will directly apply what I have learned in my career.”</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>
Engagement (Intent to Continue)

N = 85, 62% Female

Although students would recommend the course, many did not intend to continue learning more computing or applying what they learned.
Engagement vs. Components

Pearson correlation of “Student’s intent to continue learning computing” with students’ perception of each course and motivational component

<table>
<thead>
<tr>
<th>Fall 2016</th>
<th>eMpowerment</th>
<th>Usefulness</th>
<th>Success</th>
<th>Interest</th>
<th>Caring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Programming</td>
<td>.406</td>
<td>.354</td>
<td>.341</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 85, 62% Female

Intent to continue seems to be correlated with the content, not the context.
Take-aways

• Data Science seems to be a preferable context for students, across genders.

• Context, and in particular Data Science, can seem to provide motivation in ways that content cannot

• But some engagement outcomes might be more connected to content than context
Future Work

• More Datasets

• Maintenance

• Connecting motivation to learning outcomes
Thanks!

We gratefully acknowledge the support of Virginia Tech and the National Science Foundation under Grants NSF DGE 0822220, NSF IUSE 1624320, and NSF IUSE 479632.
Questions?

https://think.cs.vt.edu/corgis

Artwork by Eleonor Bart
Trends in Motivation
<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empowerment/Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empowerment/Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empowerment/Ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empowerment/Abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empowerment/Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring/Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring/Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring/Ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring/Abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring/Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful/Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful/Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful/Ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful/Abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful/Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful/Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful/Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful/Ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful/Abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Useful/Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest/Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest/Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest/Ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest/Abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest/Programs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Disagree**
- **Somewhat Disagree**
- **Neither Agree nor Disagree**
- **Somewhat Agree**
- **Agree**
- **Strongly Agree**
Other Components
<table>
<thead>
<tr>
<th>Spring 2016</th>
<th>eMpowerment</th>
<th>Usefulness</th>
<th>Success</th>
<th>Interest</th>
<th>Caring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction</td>
<td>.458</td>
<td>.699</td>
<td>.614</td>
<td>.488</td>
<td></td>
</tr>
<tr>
<td>Cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethics</td>
<td>.485</td>
<td>.418</td>
<td>.323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Programming</td>
<td>.437</td>
<td><strong>.823</strong></td>
<td>.600</td>
<td>.638</td>
<td></td>
</tr>
</tbody>
</table>

Continue Learning, Applying, and/or Recommend Course
N =36
50% female
Structure