Situating Computational Thinking with Big Data: BlockPy and CORGIS

Austin Cory Bart, Javier Tibau, Bushra Chowdhury,
Dr. Dennis Kafura, Dr. Clifford A. Shaffer, Dr. Eli Tilevich

Problem and Motivation

> *"To this end, **Virginia Tech will** comprehensively evaluate and **modify the current Curriculum** for Liberal Education to … **incorporate computational thinking** and informatics/digital fluency as basic skills **for all students**, thereby enabling our students to be engaged citizens and life-long learners."*

- **Computational Thinking**: New requirement at Virginia Tech for 24,000+ undergraduates
- Represents a **movement** in higher education to bring computing to everyone
- The primary challenges here are
  - **Engage** a diversity of different majors
  - **Scale** the learning experience to thousands

- Many **existing approaches** exist: Math, Animation, Games, Robots, Web/Mobile Dev, Simulations, Image Manipulation, … But are these really **authentic** and **useful** for all?

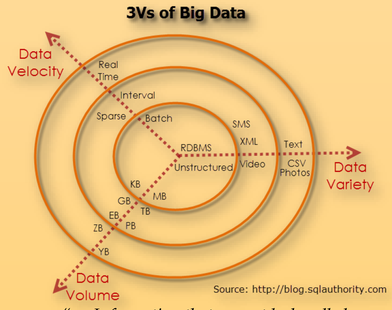
- **We join with a small but growing initiative of research initiatives that use Big Data as an authentic, motivating learning context for students.**

Relevant Theories and Background

Sources of motivation as modelled by the **MUSIC Model of Academic Motivation**
- eMpowerment
- Usefulness
- Success
- Interest
- Caring

Educational Theories
- Situated Learning
- Constructivism
- Socio-cognitivism
- Active Learning
- Problem-based Learning
- Cooperative Learning
- Mastery Learning
- Instructional Design

Theory of Big Data



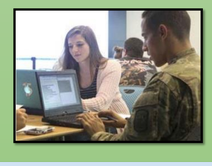
"… Information that cannot be handled with traditional methods…"[5]

Pedagogy

"Introduction to Computational Thinking"
- 2 instructors (Senior and Associate)
- **Fall 2014**: 24 students , 70% male
- **Spring 2015**: 40 students, 60% female, 3 UTAs
- **Spring 2016**: 50 students, 50% female, 4 UTAs
- 20 different majors from 5 different colleges

Class Strategy
- Staff are guides, not talking heads
- Only a quarter of classtime is lecture
- Early focus on paper-and-pencil topics
- All material situated in Big Data context

Learning Objectives
- Abstraction
- Algorithms
- Visualizations
- Social Impacts



3-phase course with repeated emphasis on core objectives

(1) NetLogo
Abstraction and Modelling

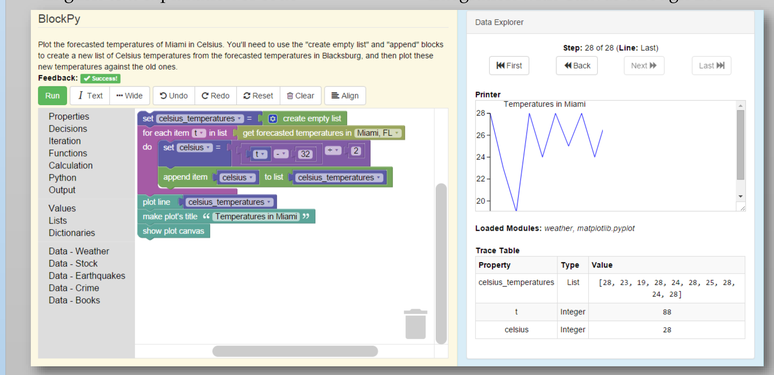
(2) BlockPy
Algorithms

(3) Python
Open-ended Final Projects

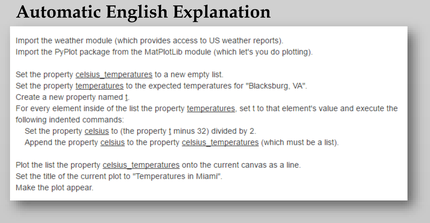
Cohort model
- Interdisciplinary teams
- 5-6 students
- Free to work together
- Support network

Technology

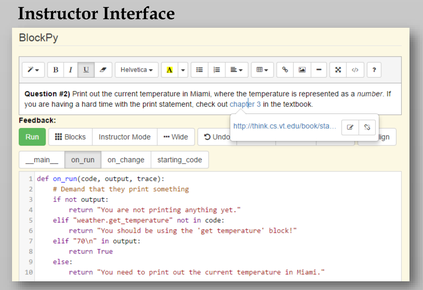
BlockPy
- Block-based programming using BlockPy
- Local Python execution with Skulpt
- Automatic, interactive feedback through static analysis and output checking
- Code-aligned Property Explorer
- English-text explanation of code
- Real-time bi-directional mapping with Python code for deep transfer
- Plot visualization blocks
- CORIGS datasets blocks
- Interaction logging for advanced analysis
- Integration with Canvas through LTI



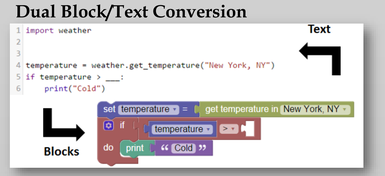
Automatic English Explanation



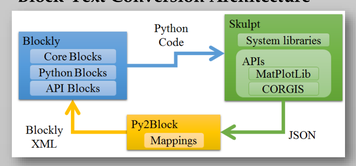
Instructor Interface



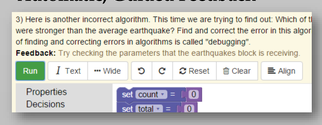
Dual Block/Text Conversion



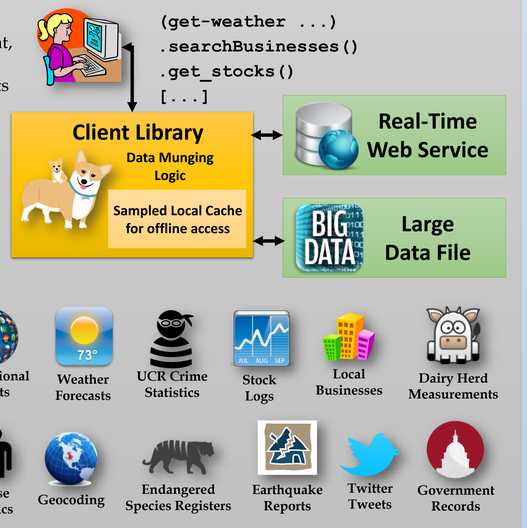
Block-Text Conversion Architecture



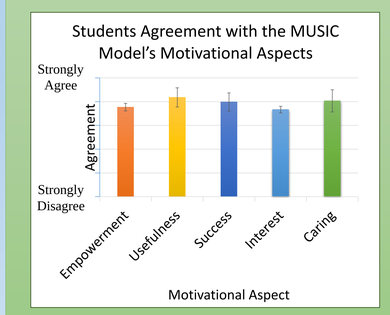
Automatic, Guided Feedback

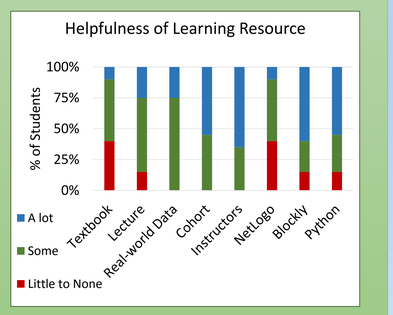


CORGIS Datasets Project
- Collection of Real-time, Giant, Interesting datasetS
- Over 35 ready-to-use datasets in a wide variety of subjects
- Free, open-source
- Beginner-friendly interface
- Goal is to support Python, Java, Racket, and Android
- Internal scaffolding for managing high velocity and high volume datasets



Motivational Data





Conclusions and Future Work

- **Overall Lessons**
  - Survey data gathered from the first offering indicates students enjoyed the experience
  - Authentic assessment of student final projects suggests positive learning gains
  - Plan to leverage the Computational Thinking Concept Inventory for context comparison
  - Longitudinal analysis to determine the course's impact on students long-term success
- **Pedagogical Lessons**
  - Social interaction is key – as class continues to scale, the human element must be retained
  - Students definitely appreciated the more active lessons (as expected from literature)
  - Success (Self-efficacy) and Self-regulation is of growing importance
- **Technology Lessons**
  - Building up problems is hard, especially if they have feedback – need more tools for this
  - NetLogo doesn't fit in well – needs to be replaced with BlockPy/Python material
  - As we embrace more and more learners, we need more and more datasets; eventually we will want to look into artificially created datasets

References

[1] A. C. Bart, et al., "*Implementing an Open-access, Data Science Programming Environment for Learners*", COMPSAC '16, Atlanta, Georgia. June 10-15, 2016.
[2] Bart, A. C., et al. "*Transforming introductory computer science projects via real-time web data*." SIGCSE. 2014.
[3] Anderson, C. "*The end of theory.*" Wired magazine, 16, 2008.
[4] Anderson, R. E., et al. "*Introductory programming meets the real world: using real problems and data in CS1.*" SIGCSE. 2014.
[5] Carter, L. "*Why students with an apparent aptitude for computer science don't choose to major in computer science.*" ACM. p. 27-31 (2006).
[6] Jones, B. D. "*Motivating Students to Engage in Learning: The MUSIC Model of Academic Motivation.*" International Journal of Teaching and Learning in Higher Education 21.2 (2009)
[7] Manyika, J. , et al. "*Big data: The next frontier for innovation, competition, and productivity.*" (2011).
[8] Office of the Senior Vice President and Provost. "*Academic implementation strategy for a plan for a new horizon: Envisioning Virginia Tech 2013-2018.*" Technical report, 2013.

This material is based upon work supported by the National Science Foundation, grant TUES-1140318, and the National Science Foundation Graduate Research Fellowship, Grant No. DGE 0822220