

Implementing an Open-access, Data-Science Programming Environment for Everyone

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

COMPSAC 2016 – CELT

June 13, 2016



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

Who am I?

- **Austin Cory Bart**
- PhD in Computer Science
- Certification in Learning Sciences
- (Software + Instructional) Design

- Graduating Spring 2017!





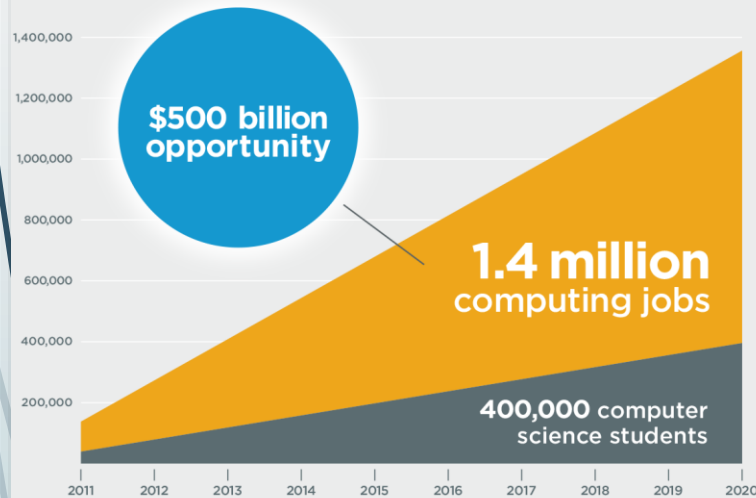
Overview



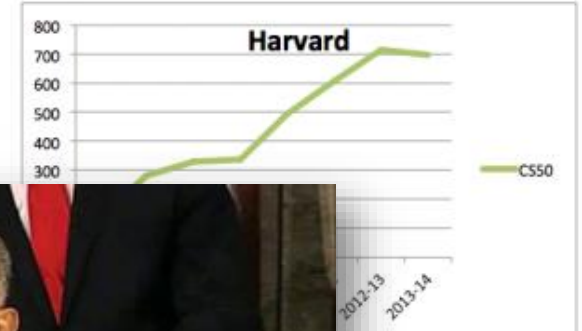
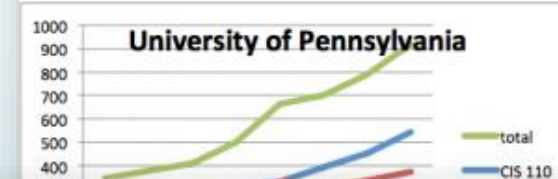
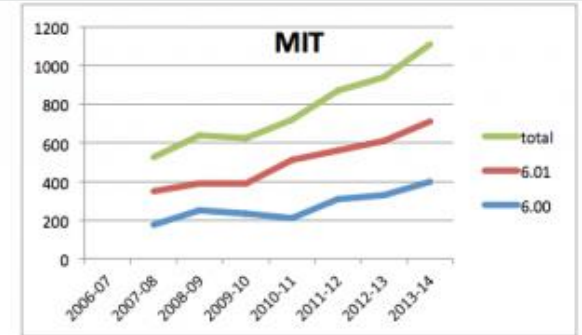
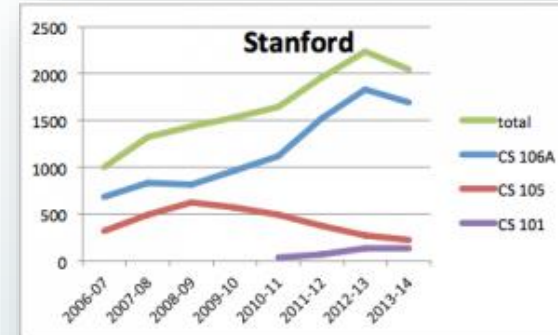
- Motivation
- Demo
- Features
- Evaluations
- Future Work

Computing For Everyone

1,000,000 more jobs than students by 2020



Computer science
is a top paying
college degree
and computer
programming
jobs are growing
at 2X the
national average.



“...offering every student the hands-on computer science and math classes that make them job-ready on day one...”

President Barack Obama
2016 State of the Union Address



The Local Interest

*"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."*



Contexts: Math and Business

Pure Math (e.g., Fibonacci)

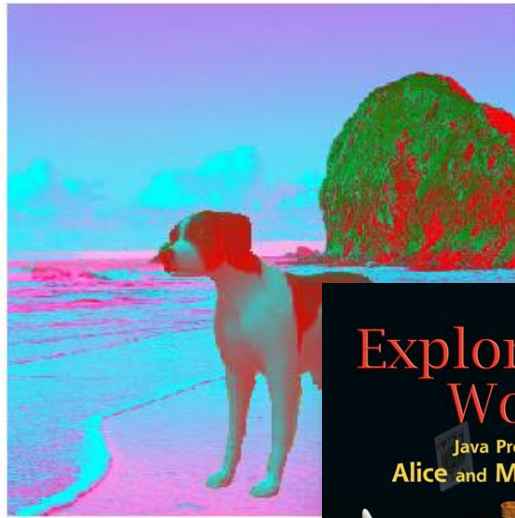
1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...

4. CONCLUSION

In this exposition, I showed how to infuse some algorithmic and mathematical aspects to guide the programming experience. The main theme is Fibonacci (and the golden ratio), which is a pleasant topic for many students. The typical paradigm that I support here is to first start with a warm up question (one that is not too trivial). then to

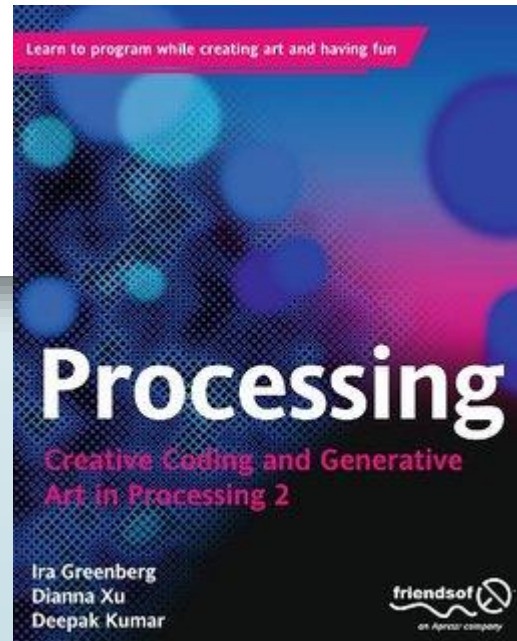
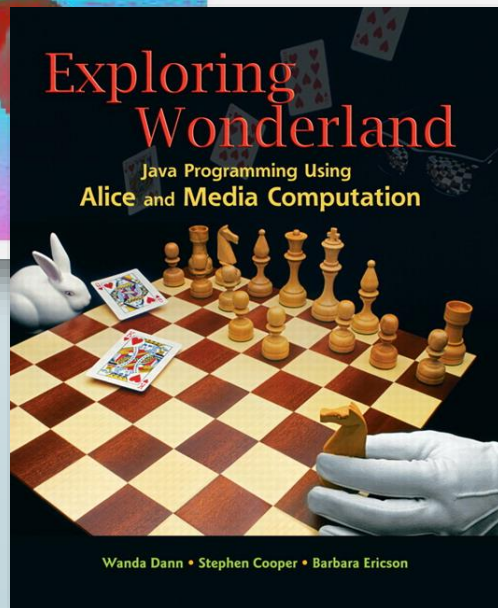
Saad Mneimneh. 2015. Fibonacci in The Curriculum: Not Just a Bad Recurrence. In Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15). ACM, New York, NY, USA, 253-258.

Fun to the rescue!



```

When run button pressed
do
  count with x and y from 0 to 399 by 1
  do
    set p to pixel at x x y on canvas 3
    change red value of pixel p to 255 - red value of pixel p
  
```

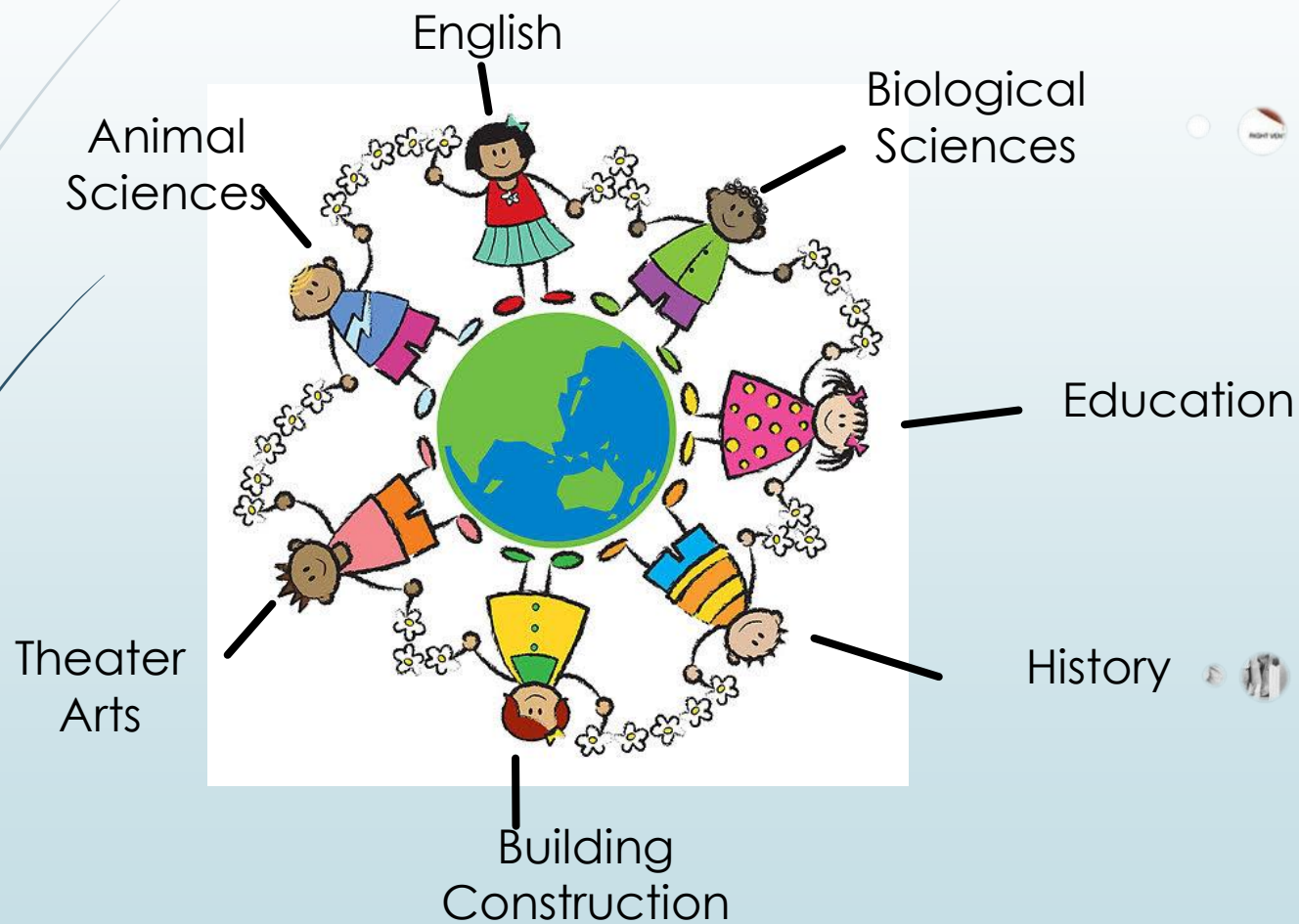


... But authentic?

*Mark Guzdial and Allison Elliott Tew. 2006. *Imagineering inauthentic legitimate peripheral participation: an instructional design approach for motivating computing education*. In *Proceedings of the second international workshop on Computing education research (ICER '06)*. ACM, New York, NY, USA, 51-58. DOI=10.1145/1151588.1151597 <http://doi.acm.org/10.1145/1151588.1151597>

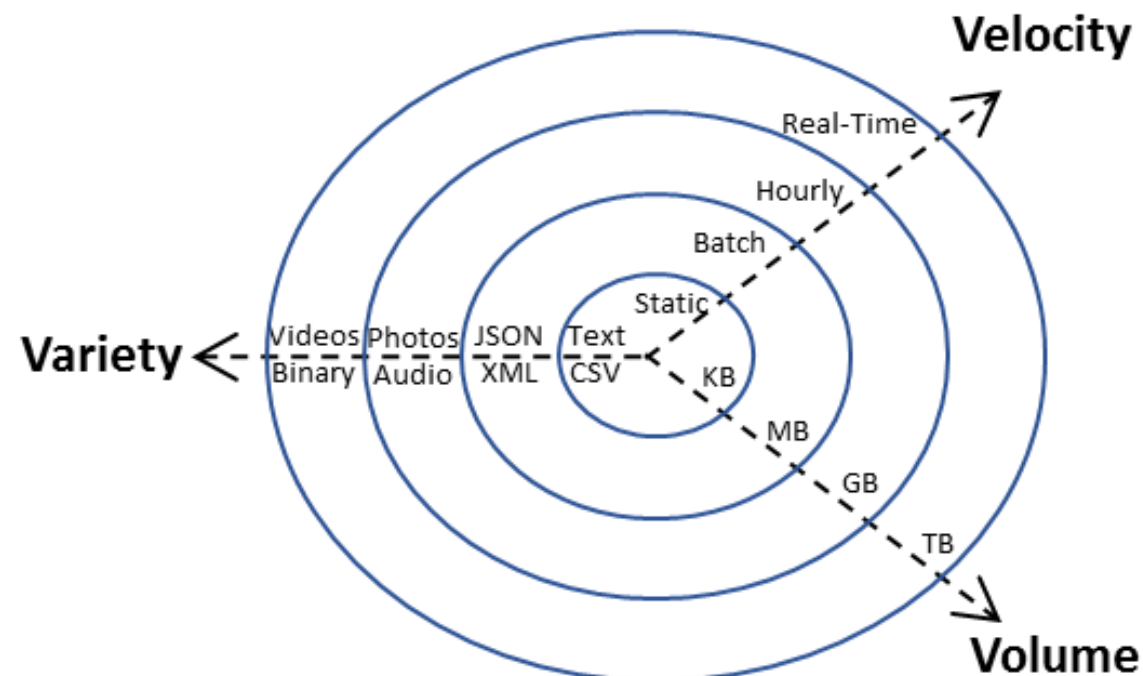
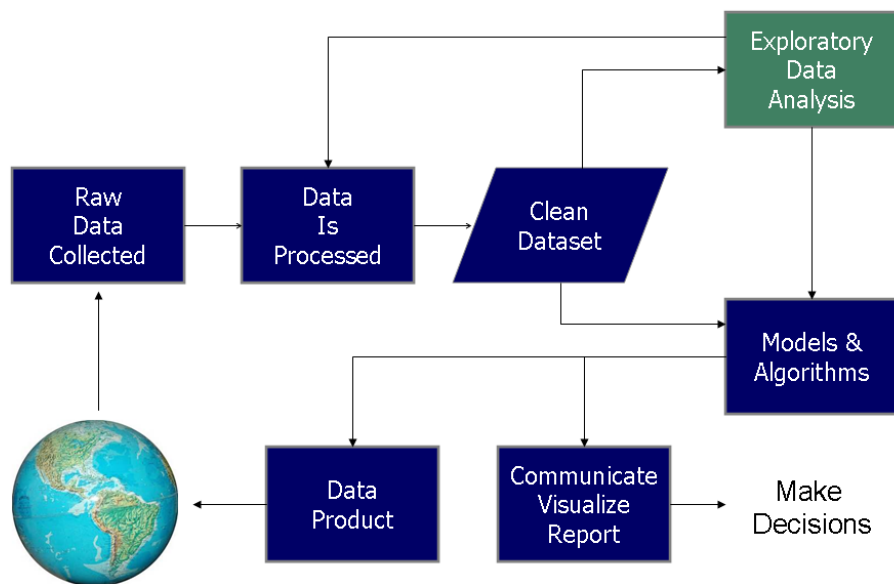
Diverse Majors

... with Rich Knowledge



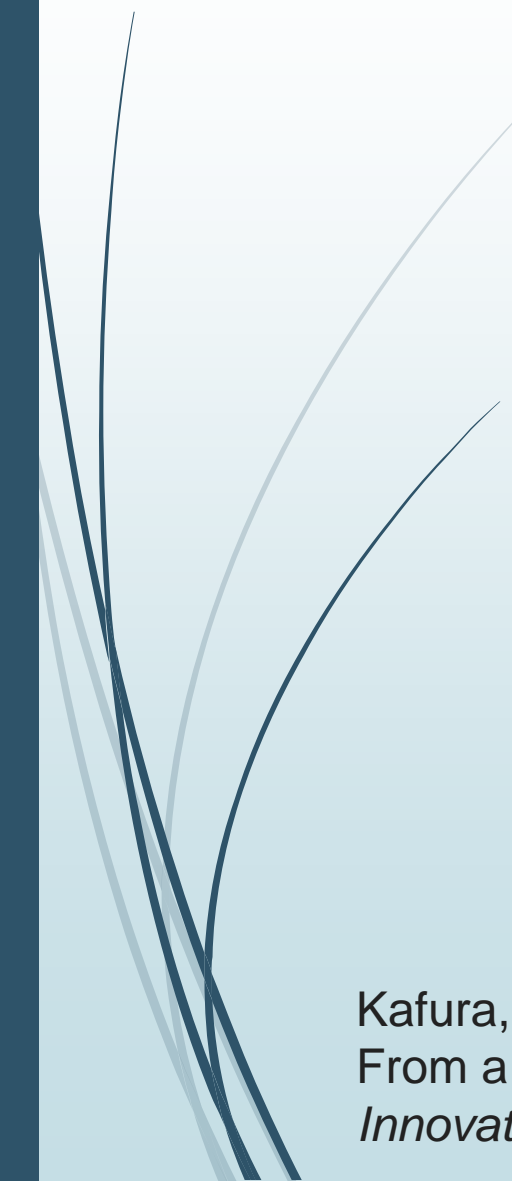
Big Idea: Real-World Data

Data Science Process





Computational Thinking via Data Science



Computational
modelling
with NetLogo



Short practice
problems in
Python

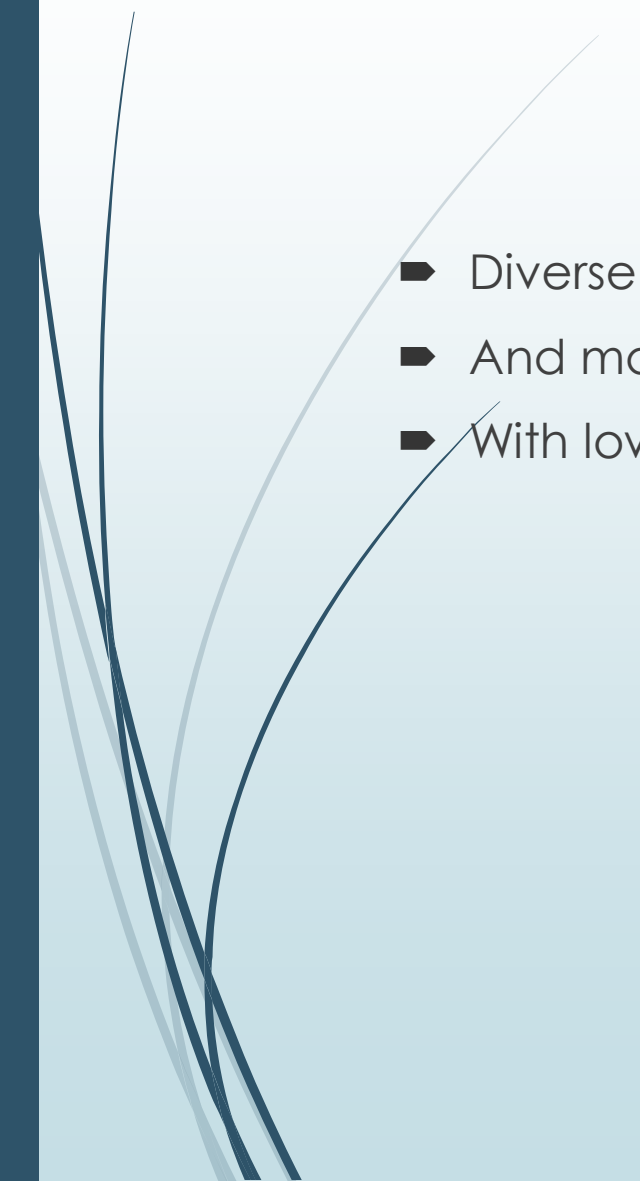


A month-long
Python project
in real IDE

Kafura, Dennis, Austin Cory Bart, and Bushra Chowdhury. "Design and Preliminary Results From a Computational Thinking Course." *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, 2015.



Design Goals

- Diverse learners
 - And many of them
 - With low self-efficacy
- 

Design Goals

- Diverse learners → 1. Give authentic problems
- And many of them → 2. Minimize human effort
- With low self-efficacy → 3. Reduce barriers
→ 4. Fade scaffolds

Python

- Popular intro language
- Popular data science language
- Many great libraries

TIOBE Index

Jun 2016	Jun 2015	Change	Programming Language	Ratings	Change
1	1		Java	20.794%	+2.97%
2	2		C	12.376%	-4.41%
3	3		C++	6.199%	-1.56%
4	6	▲	Python	3.900%	-0.10%
5	4	▼	C#	3.786%	-1.27%



Existing Solutions?

Block-based Environments

- Blockly
- Scratch
- Snap
- Greenfoot

Not Python

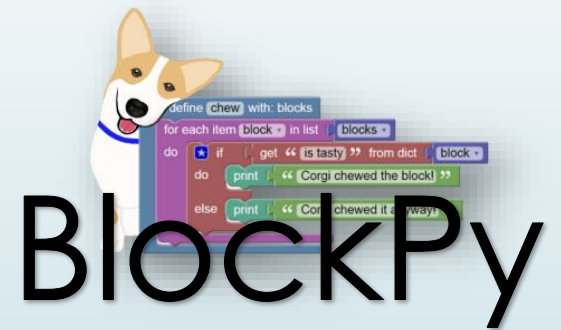
Python in the Browser

- CodeSkulptor
- Pythy
- Online Python Tutor

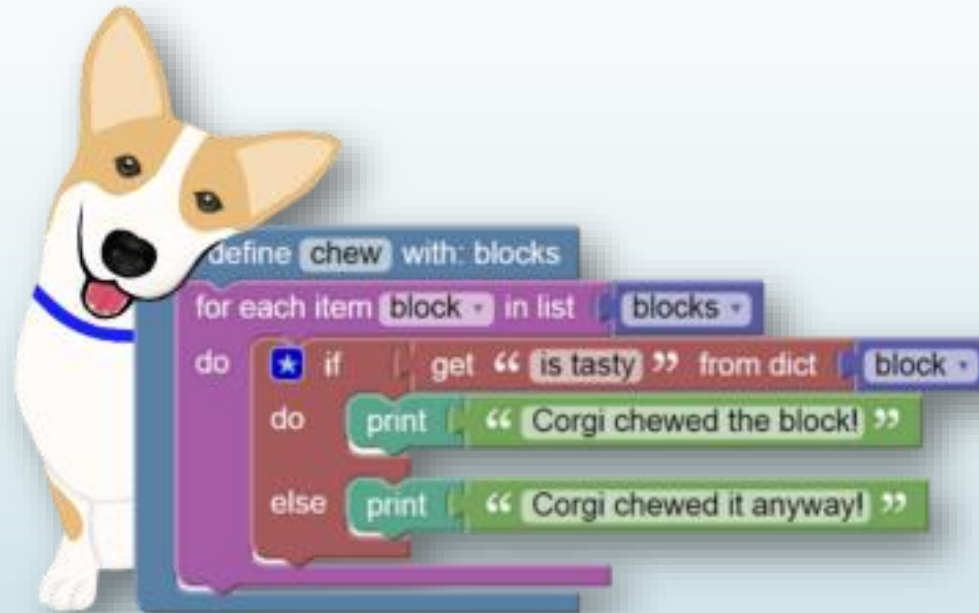
No blocks



SKULPT



BlockPy

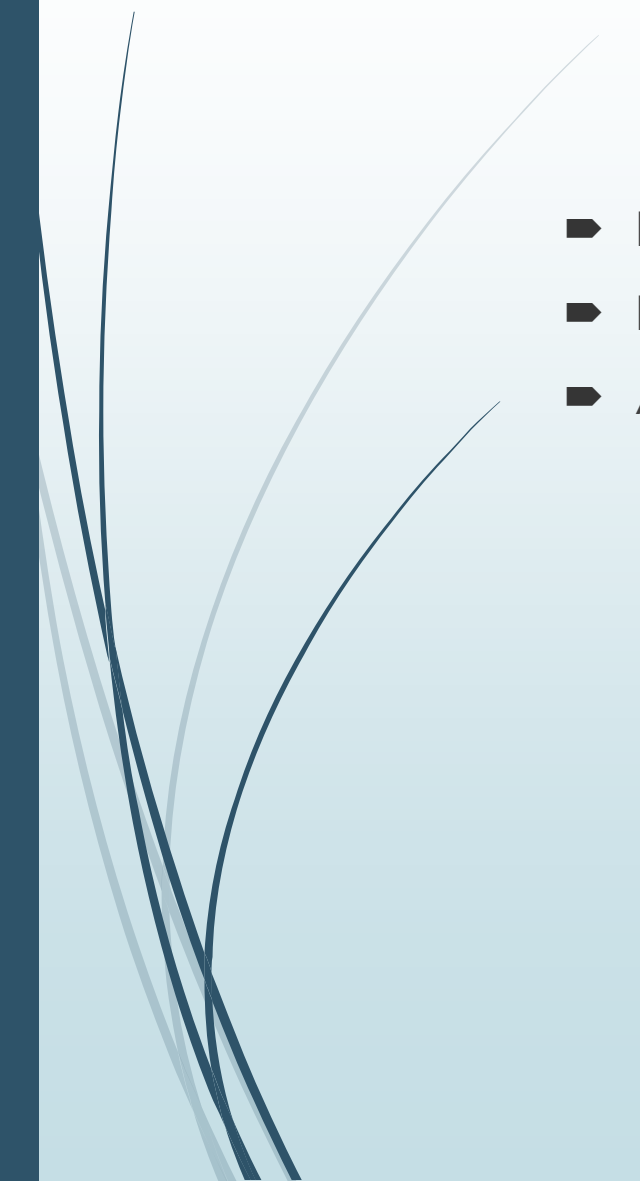


www.blockpy.com

github.com/RealTimeWeb/blockpy/

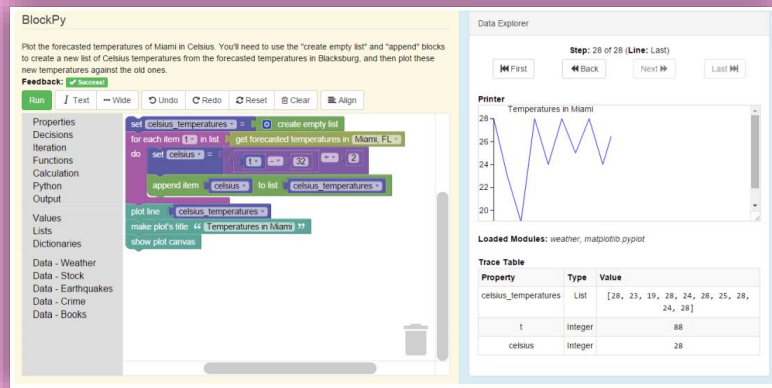


BlockPy “Secret Sauce”

- Dual text/block
 - Data Science
 - Auto Feedback
- 

Architecture

Client (Student Browser)



Blockly
(Visual Block Interface)

Skulpt
(Python Execution Engine)

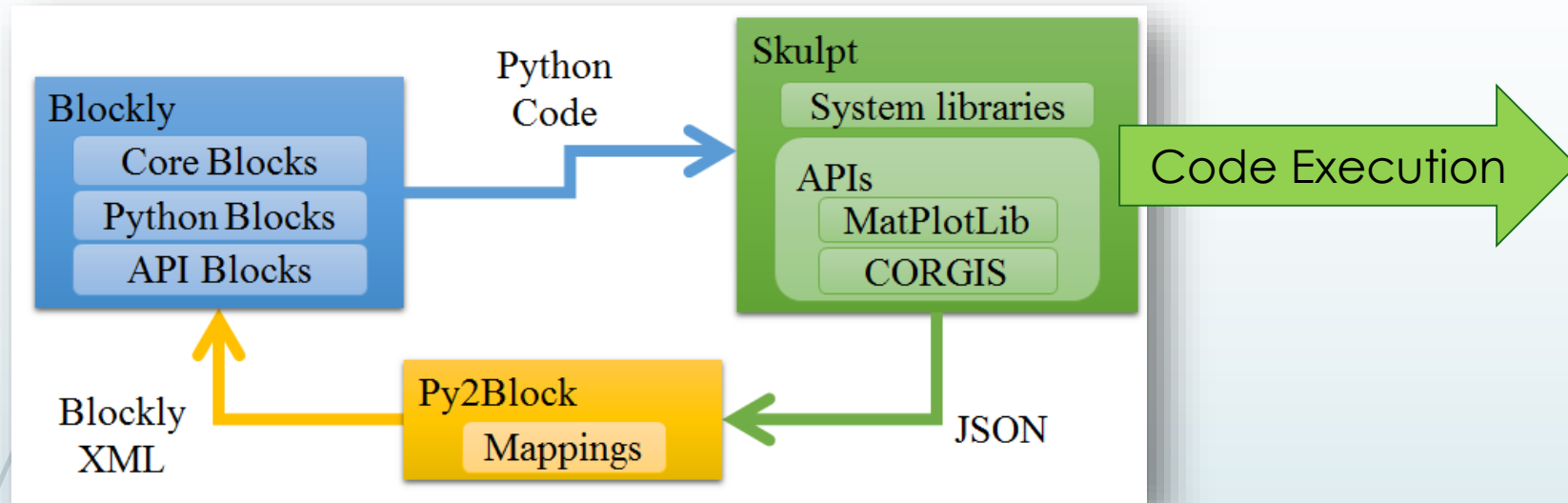
BlockPy Server

Assignment/Submission
Persistence

Learning Management System

Grade Tracking

Mutual Language Translation



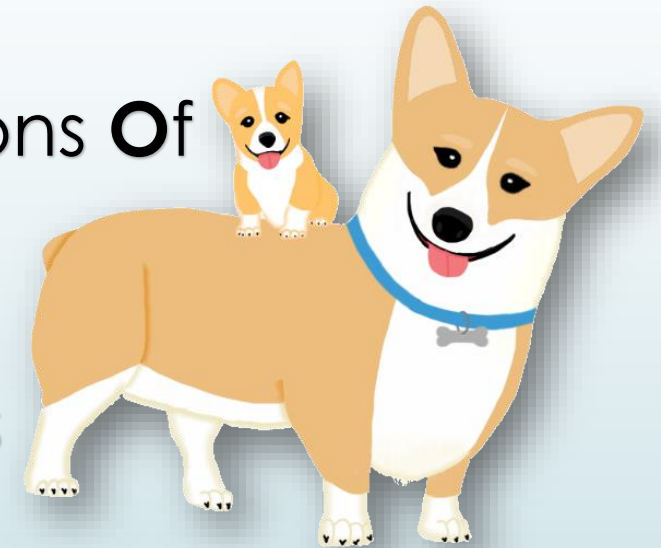
Data Blocks

Properties
Decisions
Iteration
Functions
Calculation
Python
Output
Values
Lists
Dictionaries
Data - Weather
Data - Stock
Data - Earthquakes
Data - Crime
Data - Books

get temperature in Blacksburg, VA
get weather in Blacksburg, VA
get forecasted temperatures in Blacksburg, VA
get forecasted weather in Blacksburg, VA
get all forecasted temperatures
get highs and lows in Blacksburg, VA

Collections Of Datasets

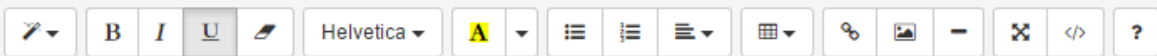
- Real-time
- Giant
- Interesting
- Situated



Tilevich, Eli, Clifford A. Shaffer, and Austin Cory Bart. "Creating stimulating, relevant, and manageable introductory computer science projects that utilize real-time, large, web-based datasets." *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. ACM, 2015.

Guided Practice Problems

BlockPy



Question #2) Print out the current temperature in Miami, where the temperature is represented as a *number*. If you are having a hard time with the print statement, check out [chapter 3](#) in the textbook.

Feedback:

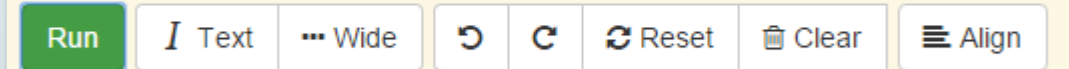


__main__ on_run on_change starting_code

```
1 def on_run(code, output, trace):
2     # Demand that they print something
3     if not output:
4         return "You are not printing anything yet."
5     elif "weather.get_temperature" not in code:
6         return "You should be using the 'get temperature' block!"
7     elif "70\n" in output:
8         return True
9     else:
10        return "You need to print out the current temperature in Miami."
```

3) Here is another incorrect algorithm. This time we are trying to find out: Which of the earthquakes were stronger than the average earthquake? Find and correct the error in this algorithm. The process of finding and correcting errors in algorithms is called "debugging".

Feedback: Try checking the parameters that the earthquakes block is receiving.



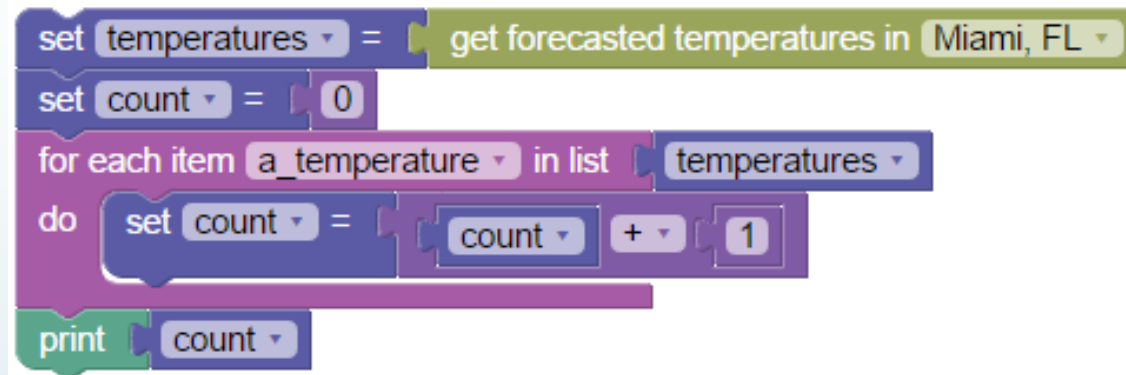
Properties
Decisions

```
set count = 0
set total = 0
```

Plot the forecasted temperatures of Miami in Celsius to create a new list of Celsius temperatures from the new temperatures against the old ones.

Feedback: ✔ Success!

Pseudo-Code



```
set temperatures = get forecasted temperatures in Miami, FL
set count = 0
for each item a_temperature in list temperatures
do
  set count = count + 1
print count
```

Pseudo-code Explanation

Import the weather module (which provides access to US weather reports).

Set the property temperatures to the expected temperatures for "Miami, FL".

Set the property count to 0.

Create a new property named a_temperature.

For every element inside of the list the property temperatures, set a_temperature to that element's value and execute the following indented commands:

Set the property count to the property count added to 1.

Print the property count to the printer.

State Explorer

Trace step
and line

Rewind
variables AND
console

State Explorer

Step: 15 of 31 (Line: 9)

First Back Next Last

Printer

```
Printing earthquakes' magnitudes
0.66
1.62
1.4
1.93
0.87
```

Step 14, Line 8


Loaded Modules: earthquakes

Trace Table

Property	Type	Value
magnitudes	List	[0.66, 1.62, 1.4, 1.93, 0.87, 3.1, 5, 0.9399, 0.2899, 0.9599, 0.54, 1.1]
total	Float	5.61
a_quake	Float	0.87

LTI Interface

Link Resource from External Tool

BlockPy

Choose an Assignment

For vs. If Last modified at 8:24PM on Wed 18, Nov 2015	Solve the problem below by writing either a FOR loop or an IF statement. Print whether the temperature outside right now in Miami "Cold" or "Hot".	<div>Select</div> <div>Edit Duplicate Delete</div>
Iteration practice Last modified at 8:59PM on Wed 18, Nov 2015	Print out all the temperatures in Seattle.	

Create new Assignment

© Copyright 2015



Evaluation

Spring 2015

- 35 students
- 4 days
- 57% female
- Version 1

Spring 2016

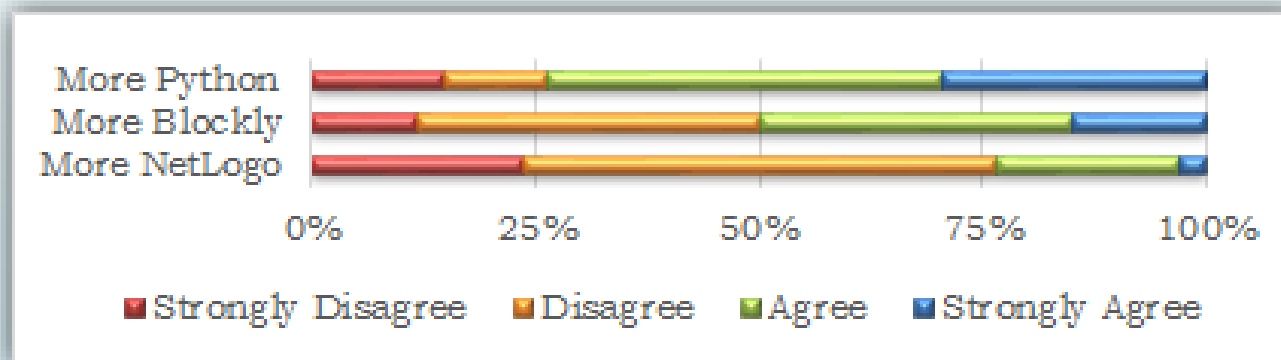
- 50 students
- 5 days
- 50% female
- Version 2

Two different post-surveys

Transition into Spyder

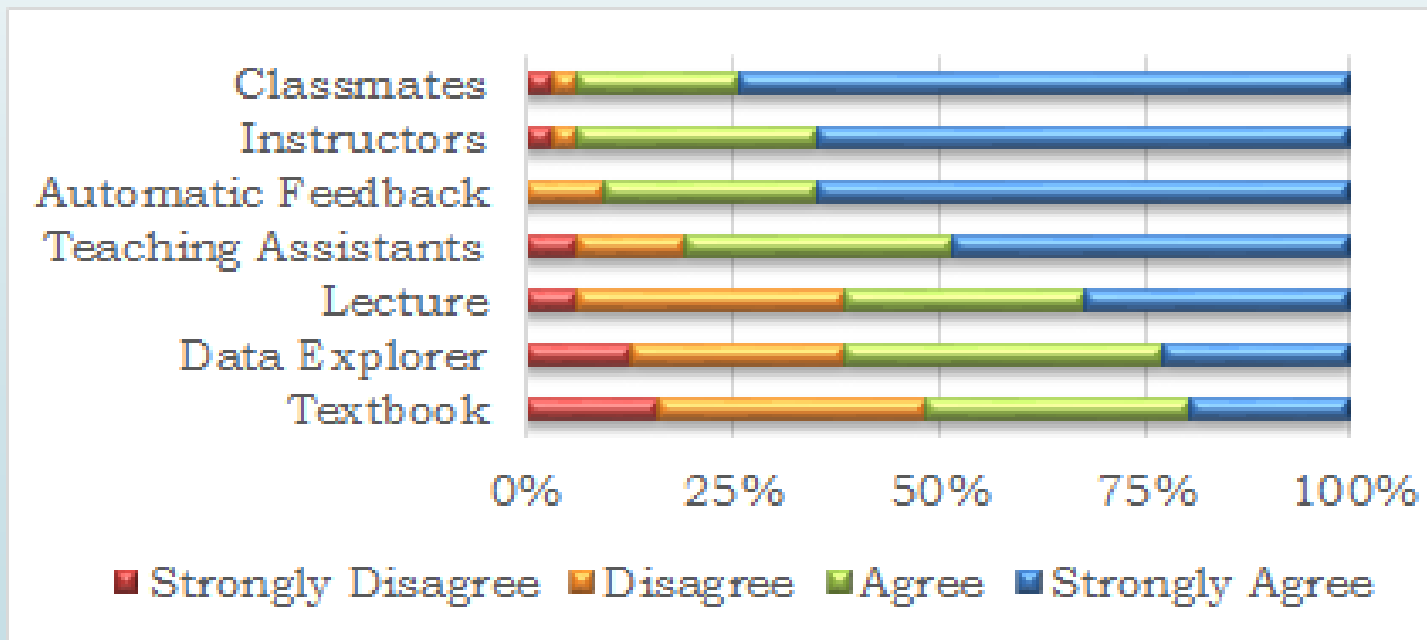
Spring 2015 - Perceptions

Do students want more time with NetLogo, BlockPy (Blockly), or Spyder (Python)?



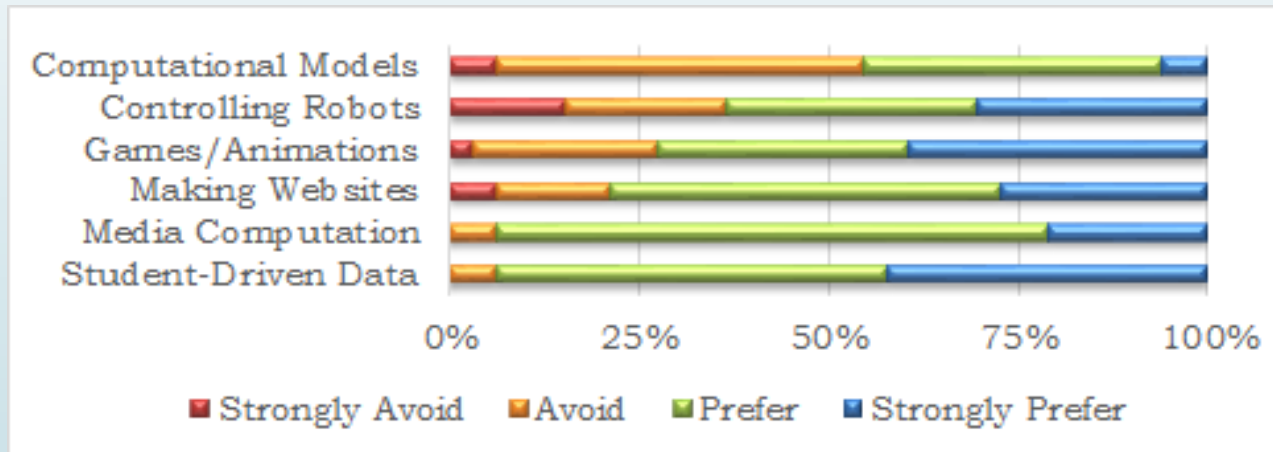
Spring 2015 – Post-Survey

How helpful to students' learning were these resources?

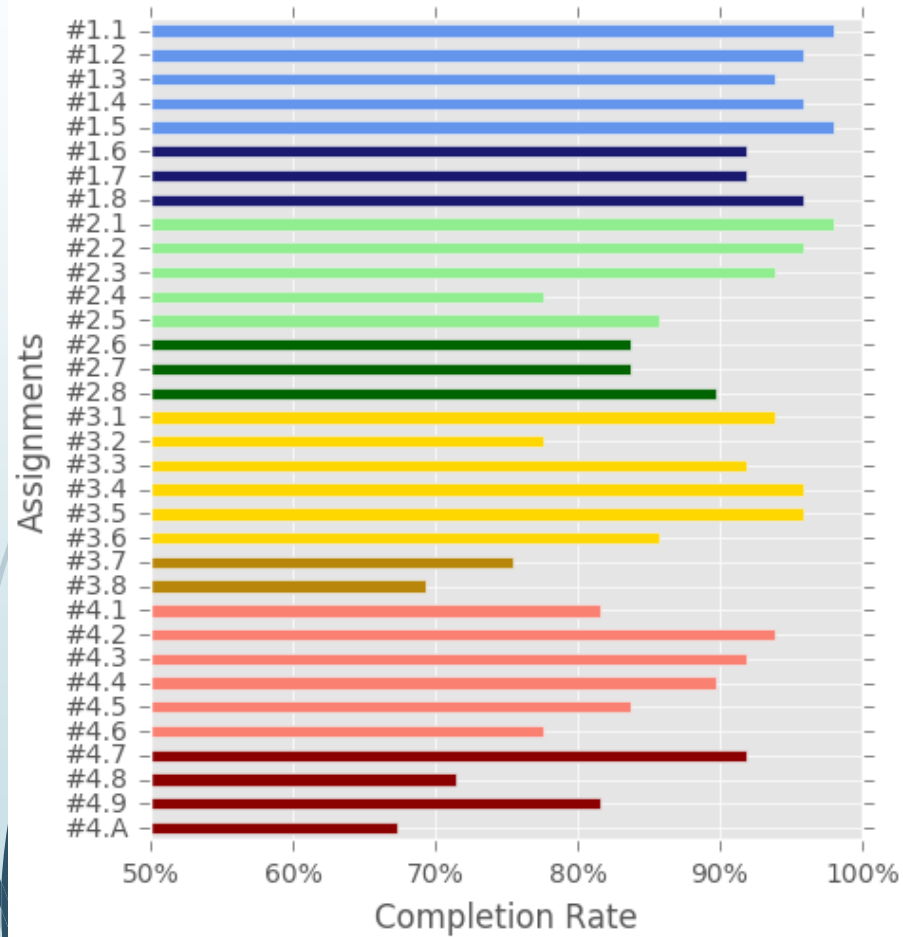


Spring 2015 – Contexts

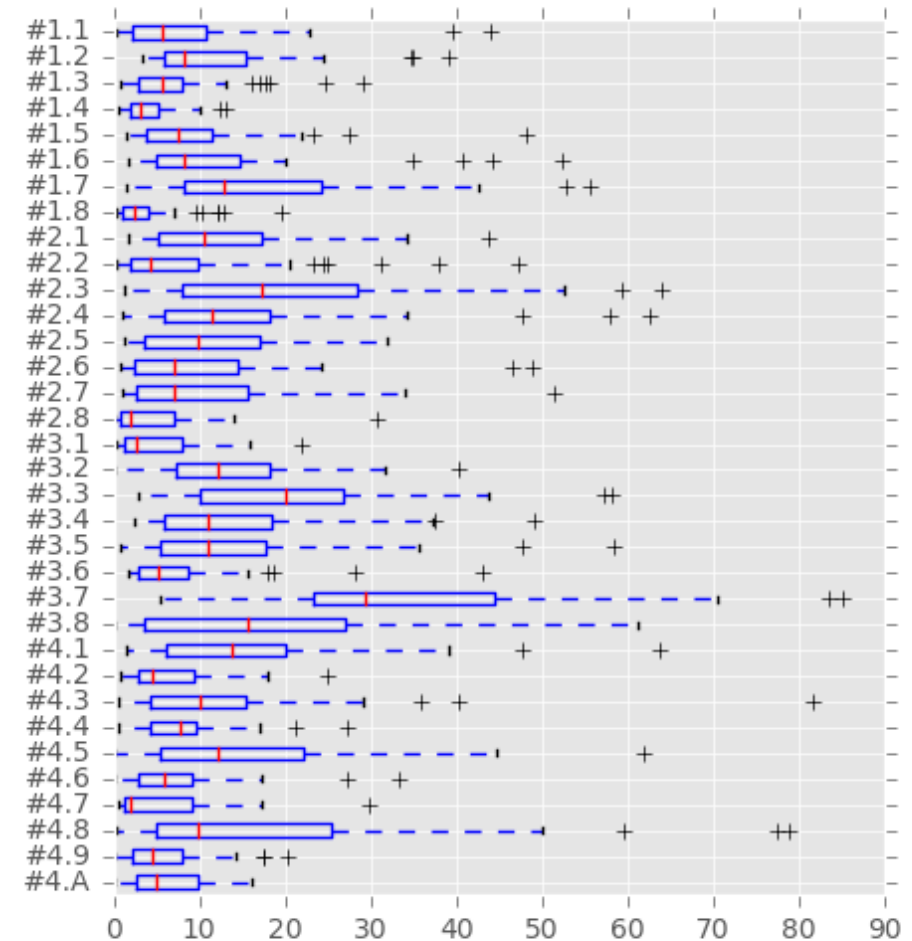
Which context seems most preferable to students?



Spring 2016 – Curriculum progress

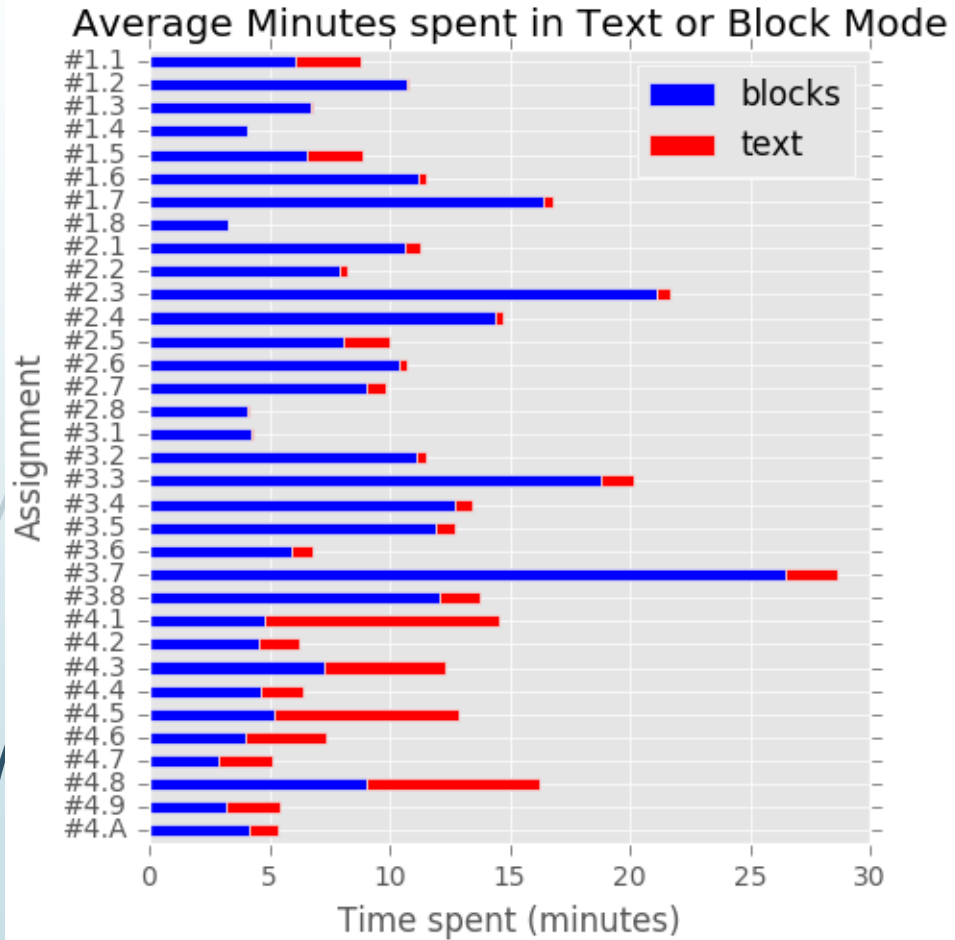


Most students completed most problems



Most problems took about 15 minutes

Spring 2016 – Dual block/text

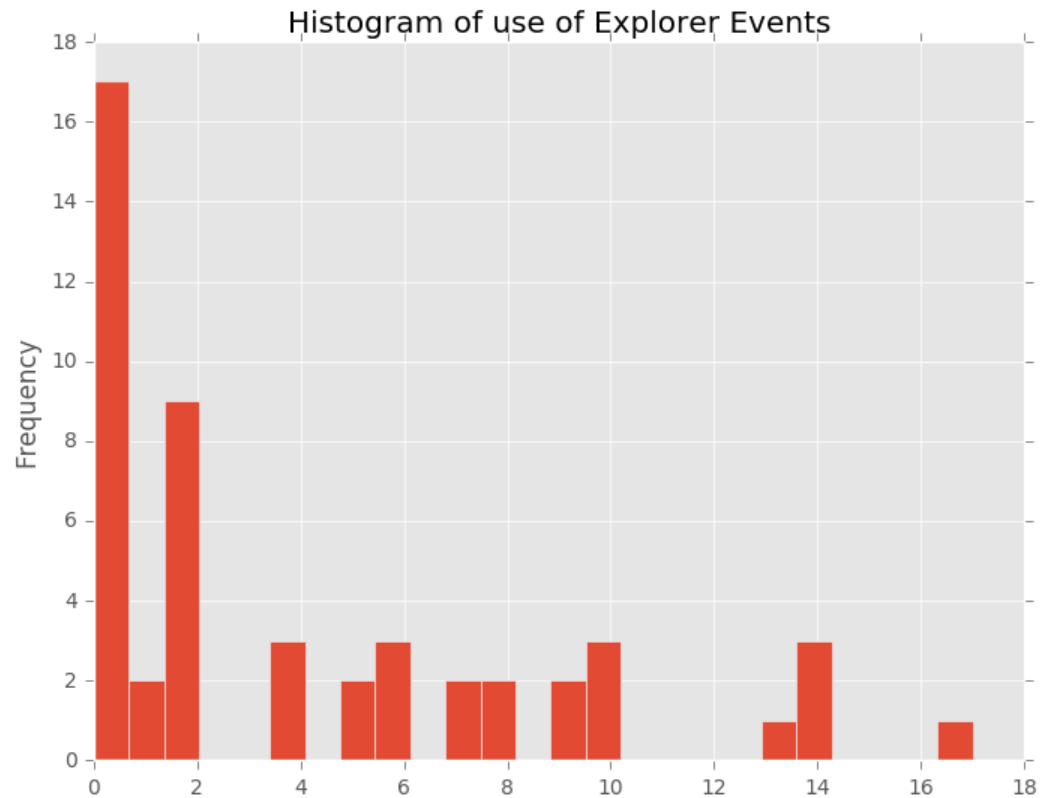


Students did transition to text



Blocks generate more code/sec

Spring 2016 – Tool events



Students use tools sparingly



Spring 2016 – Qualitative Results

- ▶ BlockPy => Spyder transition point was good
- ▶ Error reporting was better in Spyder
- ▶ 2/3 of students singled out blocks as helpful (only 12% singled out auto feedback)

Semantic Errors

- Semantic errors are slipping through

```
import stocks
```

```
stocks = stocks.get_past("FB")
```

```
new = []
```

```
if new:
```

```
    new = new + stocks
```

```
for stocks in stocks:
```

```
    print(stocks)
```

Dead codepath

Incorrect
appending

Iteration list is the
same as the
iteration variable



Looking forward

- ▶ How do we perfect the blocks?
- ▶ How can we give better feedback?
- ▶ How can we let students do better data science?

The Next Iteration

The screenshot shows a code editor interface. At the top, a feedback message is displayed: "Feedback: Semantic Error" in a red box, followed by the text "The property `a` was set on line 1, but before it could be read it was changed on line 2. It is unnecessary to change an existing variable's value without reading it first." Below the feedback message is a toolbar with buttons: "Text", "Undo", "Redo", "Align", "Reset", "Clear", "Reset", and "Capture". At the bottom, a code block is shown with three lines: "set a = 0", "set a = 5", and "print a". The second line is highlighted with a yellow border, and a pink callout bubble points to it with the text "Better error locating". Another pink callout bubble points to the feedback message with the text "Semantic Errors".

Feedback: Semantic Error
The property `a` was set on line 1, but before it could be read it was changed on line 2. It is unnecessary to change an existing variable's value without reading it first.

Semantic Errors

Text Undo Redo Align Reset Clear Reset Capture

```
set a = 0
set a = 5
print a
```

Better error locating



Conclusions

- BlockPy
 - Dual text/block
 - Auto feedback
 - Data Science
- For experimental curriculum



Thanks!

▸ Questions?

