HUMAN CENTERED DESIGN & ENGINEERING
UNIVERSITY of WASHINGTON

Community Data Science Collective

How Interest-Driven Content Creation Shapes Opportunities for Informal Learning:

A Case Study of Novices' Usage of Data Structures in Scratch

Ruijia "Regina" Cheng

Sayamindu Dasgupta

Benjamin Mako Hill

Cheng, R., Dasgupta, S., Hill, B. How Interest-Driven Content Creation Shapes Opportunities for Informal Learning in Scratch: A Case Study on Novices' Use of Data Structures. 2022. Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2022).

Background & Motivation

Computational Participation:

Learning to code in a socially supported context

Learning in online communities:

Interest-driven creation & sharing

How can online communities support learning of challenging computational concepts?

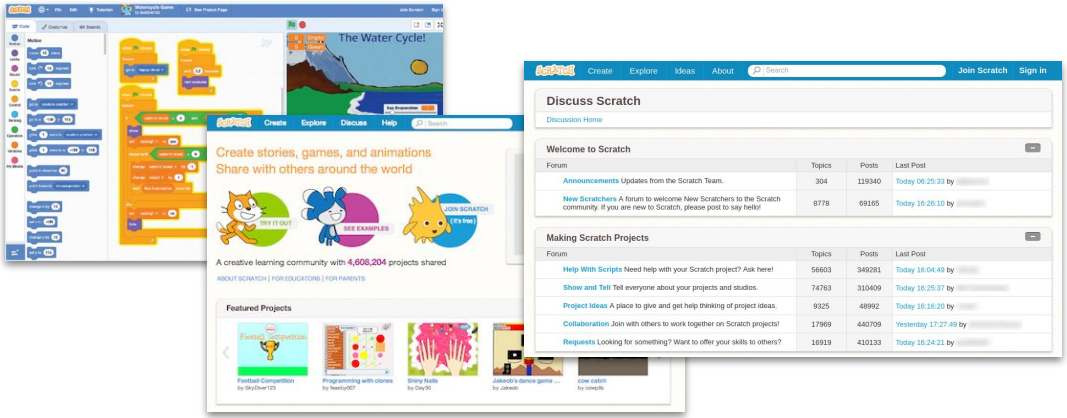
Image source: https://codeweek.eu/images/banner_training.svg

Image sources: (Left) http://womeninastronomy.blogspot.com/2021/05/how-one-astronomy-department-took-on.html
(Right) https://www.vectorstock.com/royalty-free-vector/children-using-computer-design-vector-7781415

Empirical Setting

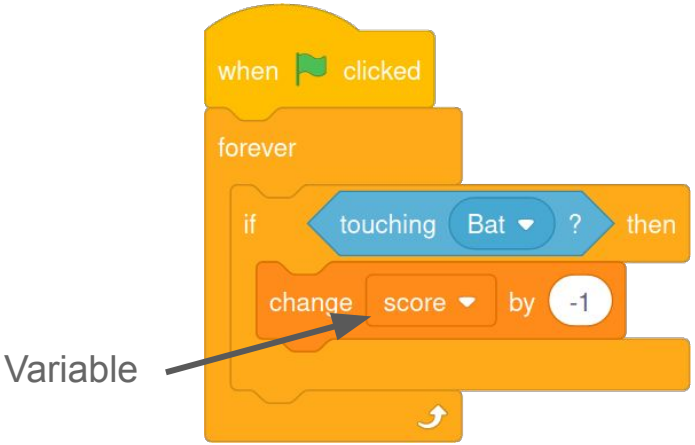
Scratch Online community

- Scratch: block-based programming language for novices
- Largest interest-driven computational learning community



Scratch Data structures

- Scalar **variables** and **lists**
- < 15% users have ever used data structures



Method: Mixed-method analysis

Study 1: Qualitative analysis of Scratch forum discussions on data structures

- Sampled 400 threads
- Grounded theory analysis

Proposed Theory

Study 2: Quantitative analysis using large-scale data on Scratch projects and user activities

Hypothesis testing analyses on 5 years of Scratch data: (241,634 projects created by 75,911 users)

Study 1 :
Qualitative Analysis of Discussions on Data Structures

Novices use specific use cases to teach each others about variables and list

- Many are game specific

User generated examples and tutorials are often framed in game-making scenarios

"Create a Variable with the name **lives** so that every time the main character touches a ghost, it would lose one life."

"You can use Lists to store the armors in the **inventory**."

Study 1 :
Qualitative Analysis of Discussions on Data Structures

Certain examples of how to use data structures become archetypes, which can constrain innovative usage

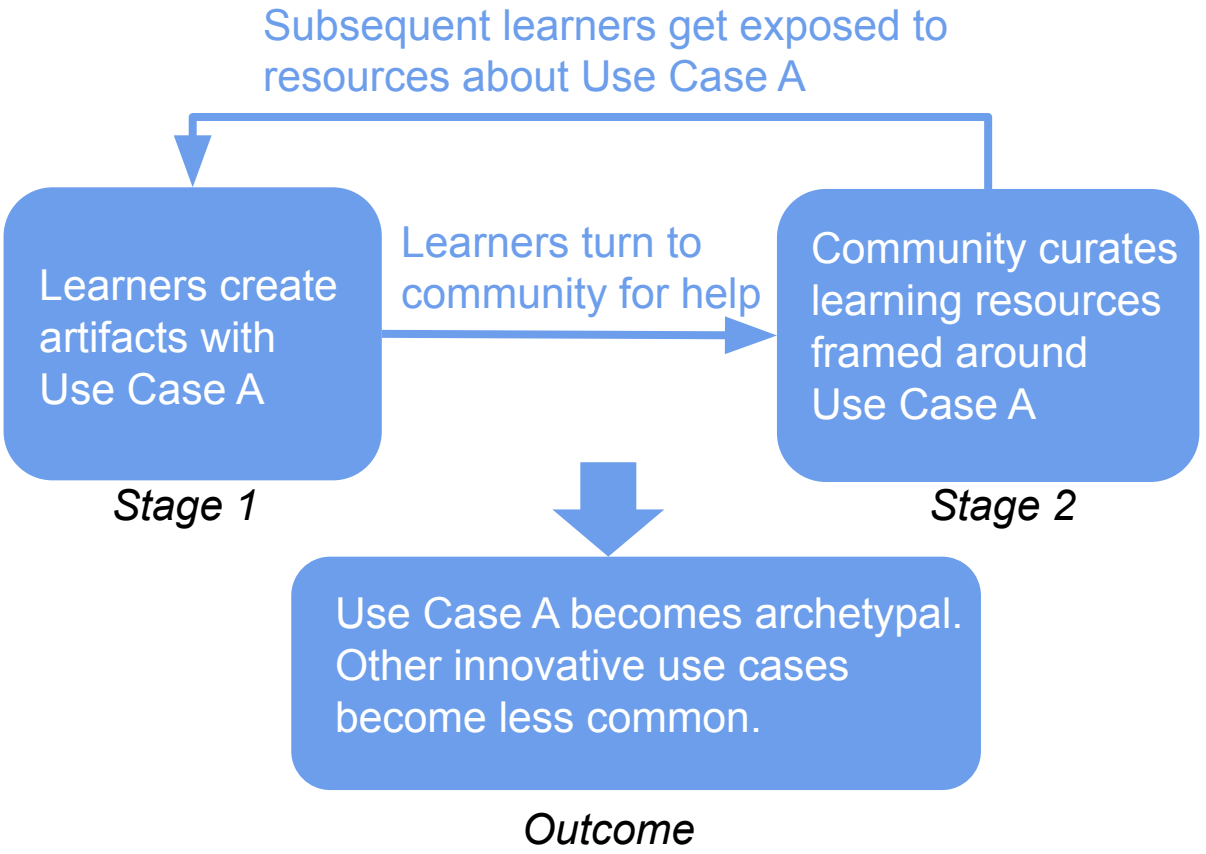
Q: "I want to use variables in my project but I don't know how?"

A: "Here is an example of how to make a score counter in your game!"

Q: "But my project is not a game. I am making a storytelling project…"

Proposed Theory

A social feedback loop of how community-generated resources may constrain innovative computational participation

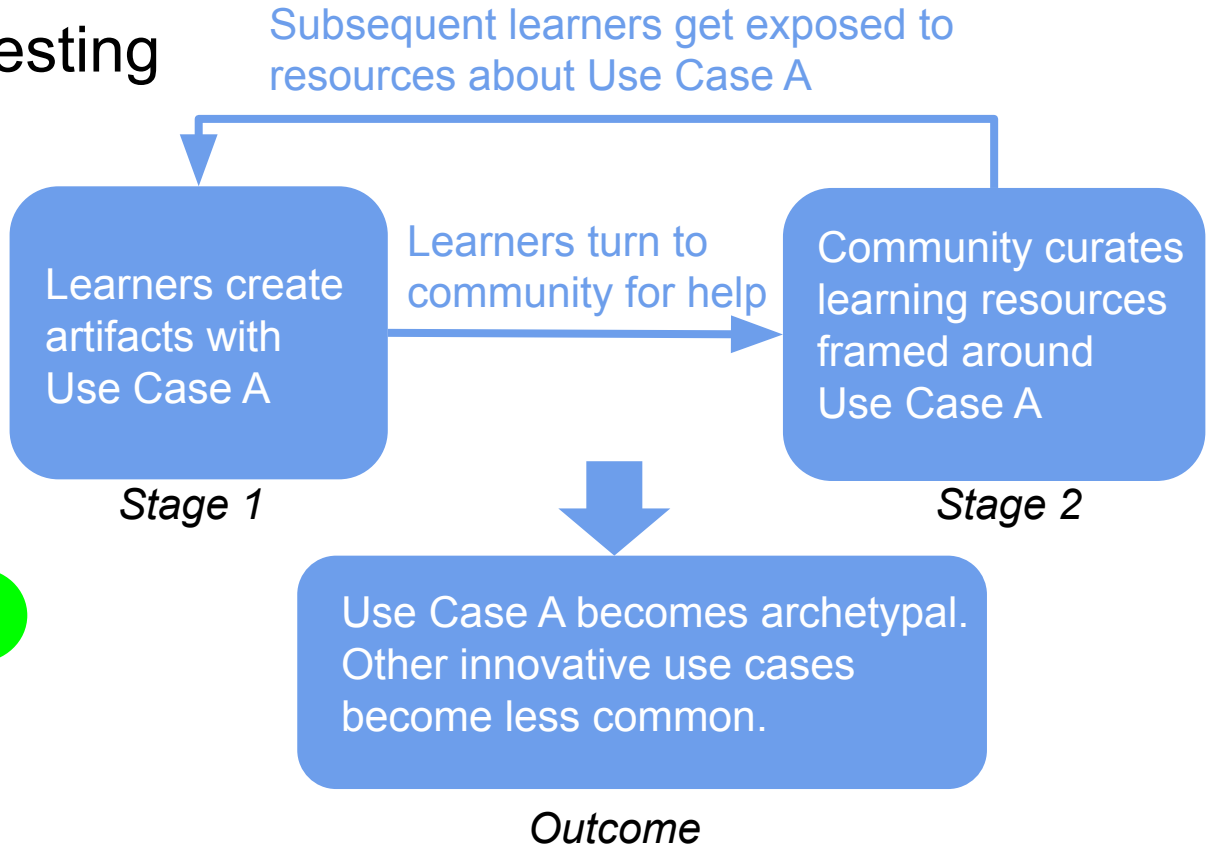


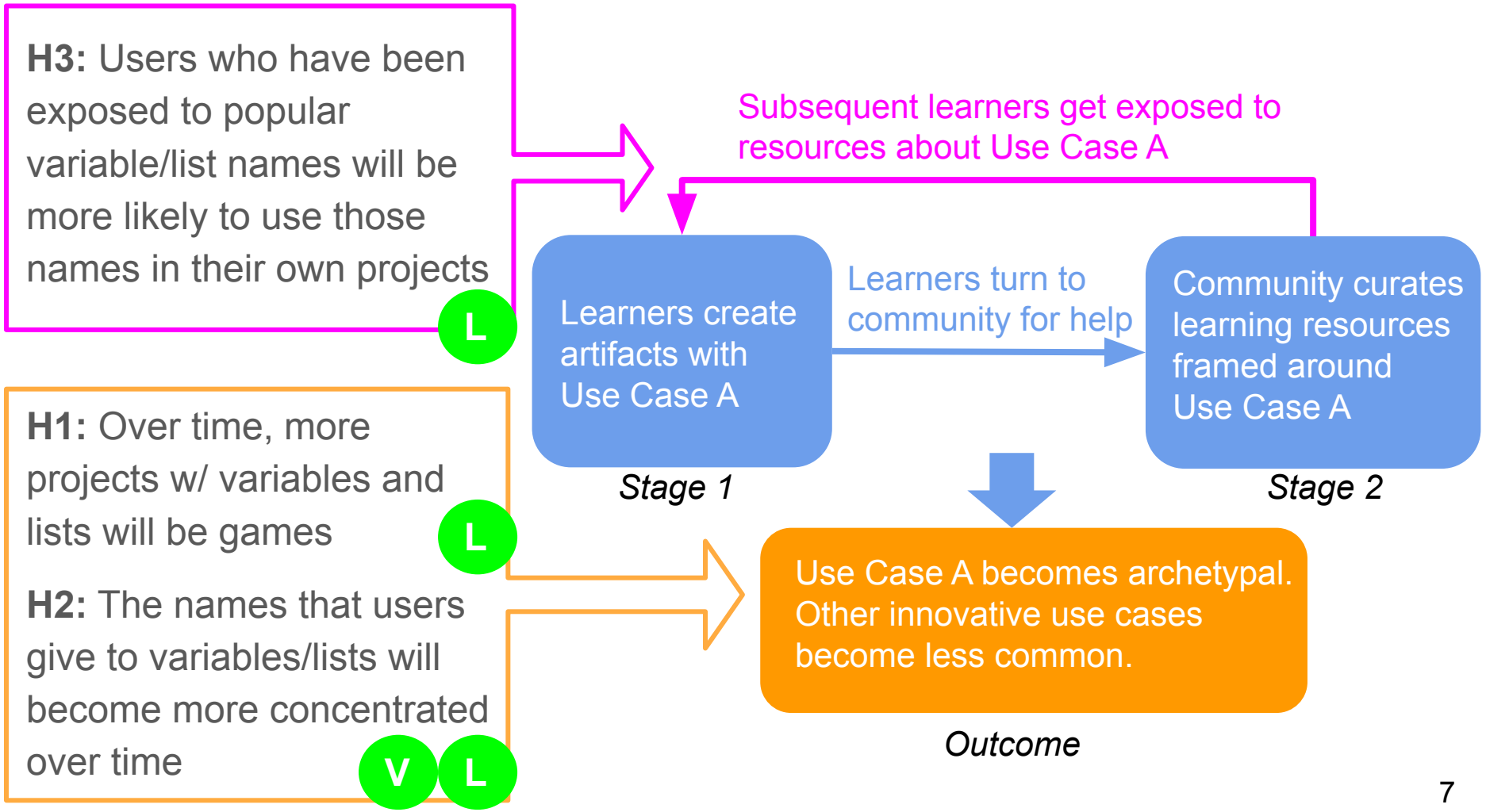
Study 2:
Quantitative Theory Testing

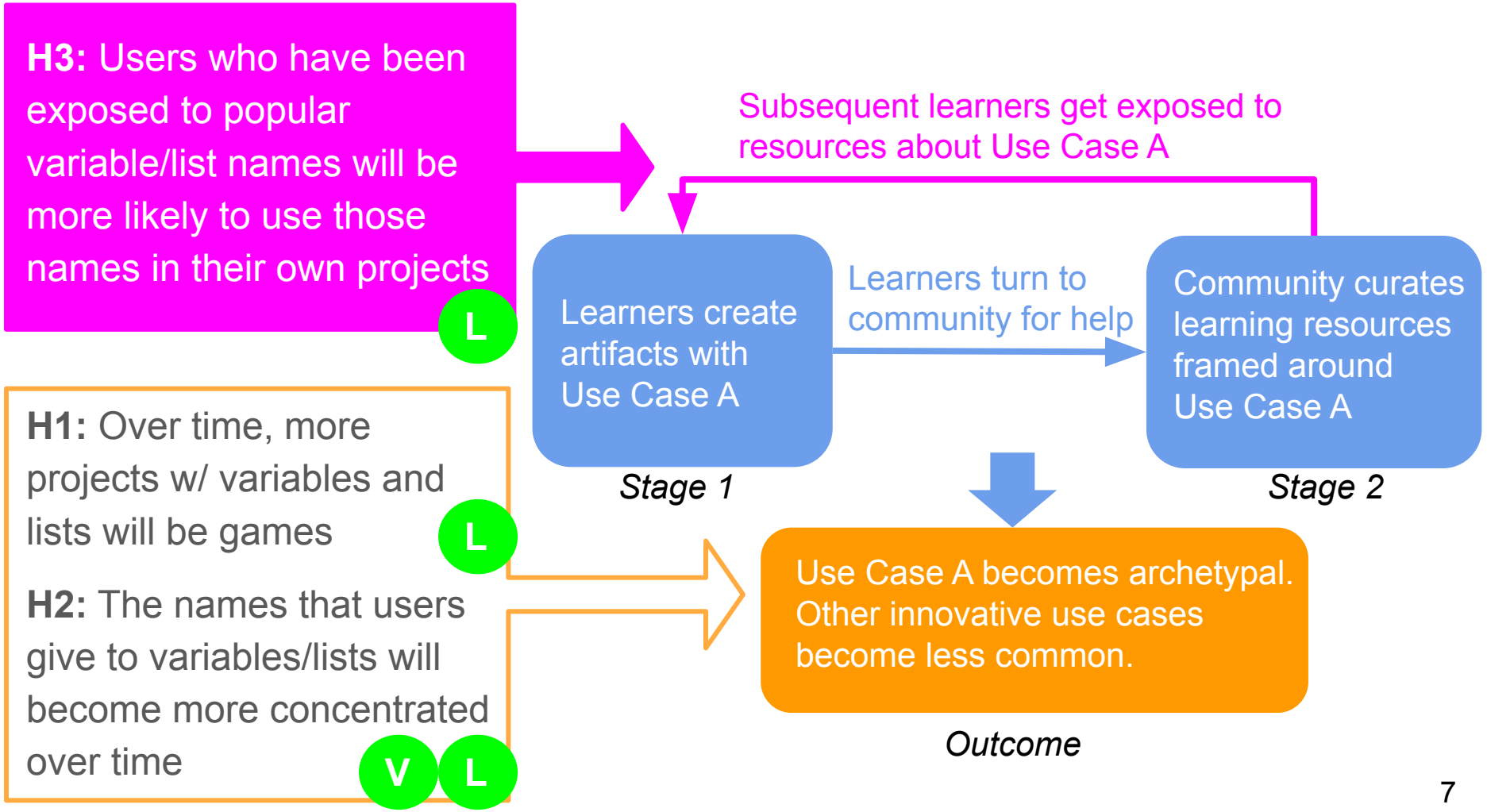
Hypothesis testing analyses on 5 years of Scratch data:

(241,634 projects created by 75,911 users)

- Projects w/ variables **V**
- Projects w/ lists **L**

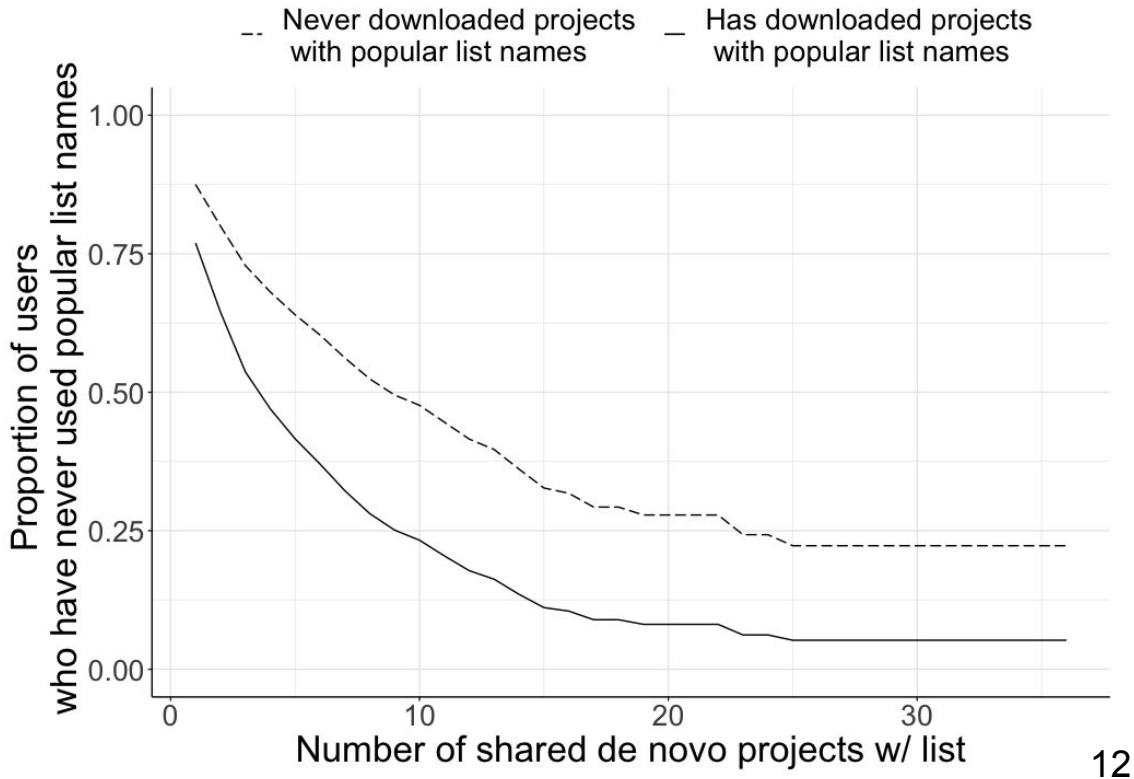



H3: Users who have been exposed to popular variable/list names will be more likely to use those names in their own projects
L
H1: Over time, more projects w/ variables and lists will be games
L
H2: The names that users give to variables/lists will become more concentrated over time
V
L
Subsequent learners get exposed to resources about Use Case A
Learners create artifacts with Use Case A
Stage 1
Learners turn to community for help
Community curates learning resources framed around Use Case A
Stage 2
Use Case A becomes archetypal. Other innovative use cases become less common.
Outcome


H3: Users who have been exposed to popular variable/list names will be more likely to use those names in their own projects
L
H1: Over time, more projects w/ variables and lists will be games
L
H2: The names that users give to variables/lists will become more concentrated over time
V
L
Subsequent learners get exposed to resources about Use Case A
Learners create artifacts with Use Case A
Stage 1
Learners turn to community for help
Community curates learning resources framed around Use Case A
Stage 2
Use Case A becomes archetypal. Other innovative use cases become less common.
Outcome

H3: Users who have been exposed to popular variable/list names will be more likely to use those names in projects

- Popular names: 20 most frequently used names for variables or lists
- Exposure: has downloaded projects w/ popular variable/list names?
- Survival analysis



Discussion & Implication

Trade-off of online community for computational learning:

- User-generated explanation and understanding that fit common interests
- Risk of superficial, restricted learning resources

Additional regulation and support needs to be implemented:

- Support a wider range of inspiration
- Broaden participation for diverse interests
- Scaffold construction of generalizable knowledge

HUMAN CENTERED DESIGN & ENGINEERING
UNIVERSITY of WASHINGTON

Community Data Science Collective

Thanks for Listening!

Ruijia “Regina” Cheng
PhD candidate at University of Washington
Research Areas: online communities, human-AI collaboration, data literacies, data science & programming support, creativity
Email: rcheng6@uw.edu
Twitter: @reginarcheng

Study 1 : Qualitative analysis of discussions on data structures

Method

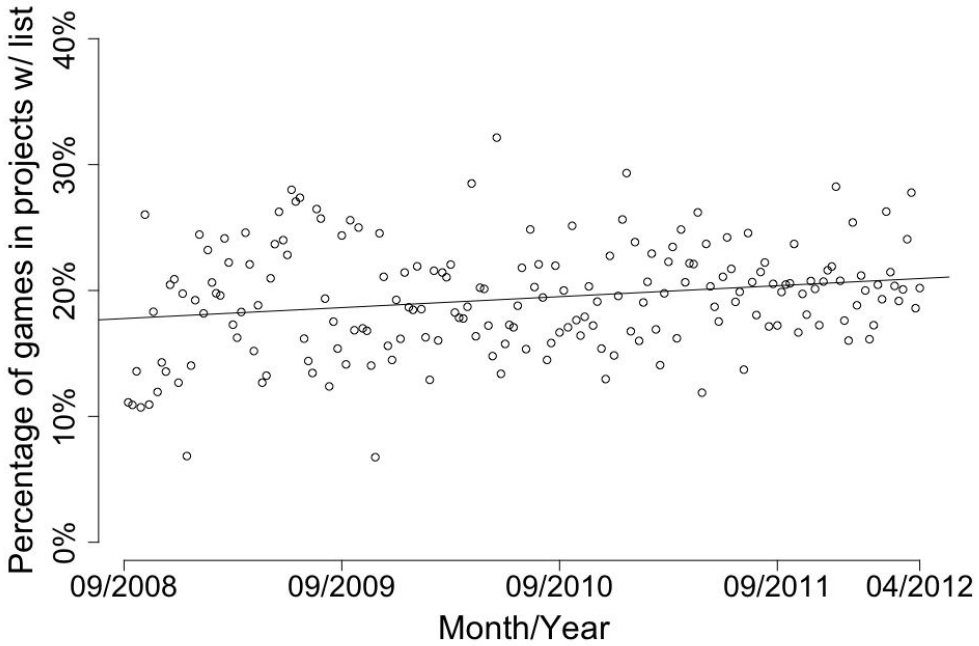
- Data from Scratch Q&A forums
- Sampled 400 text-based discussion threads using keywords:
  - “data”, “variable”, ”list”
  - Top 100 user defined variable and list names: e.g., “score”, “leaderboard”, “inventory”
- Grounded theory analysis on the threads

Study 2: Quantitative testing of the social feedback loop

H1: Over time, more projects involving variables and lists will be games

- Label games: projects w/ substring “game” or “gaming” in description/title
- Logistic regression on the odds of game over time.

(Partially supported by List)



Study 2: Quantitative testing of the social feedback loop

H2: The names that users give to variables and lists will become more concentrated over time

- Concentration: “Gini coefficient” of the distribution of variable and list names
- Linear regression of Gini over time.

(Supported by both Variables and List)