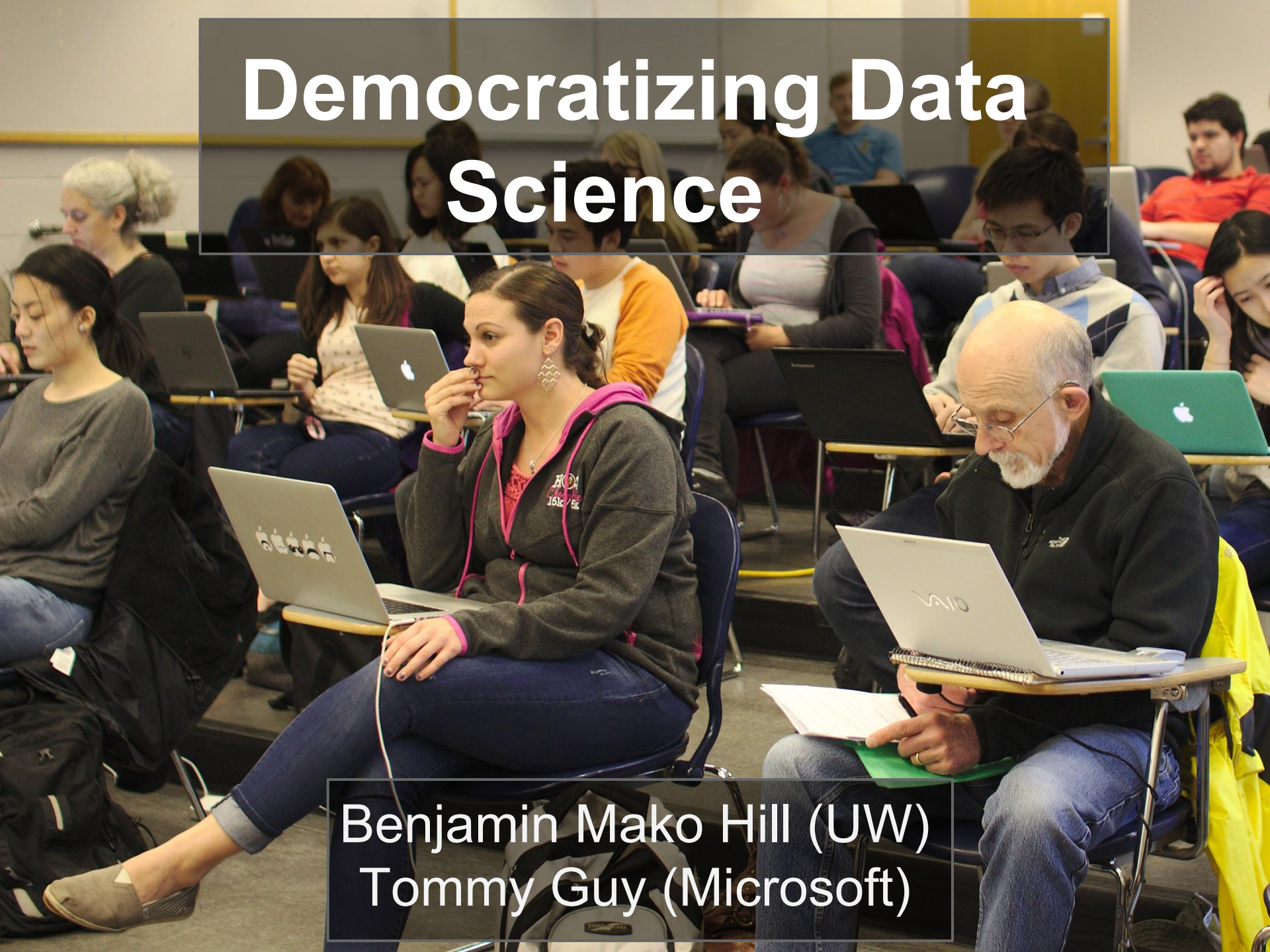
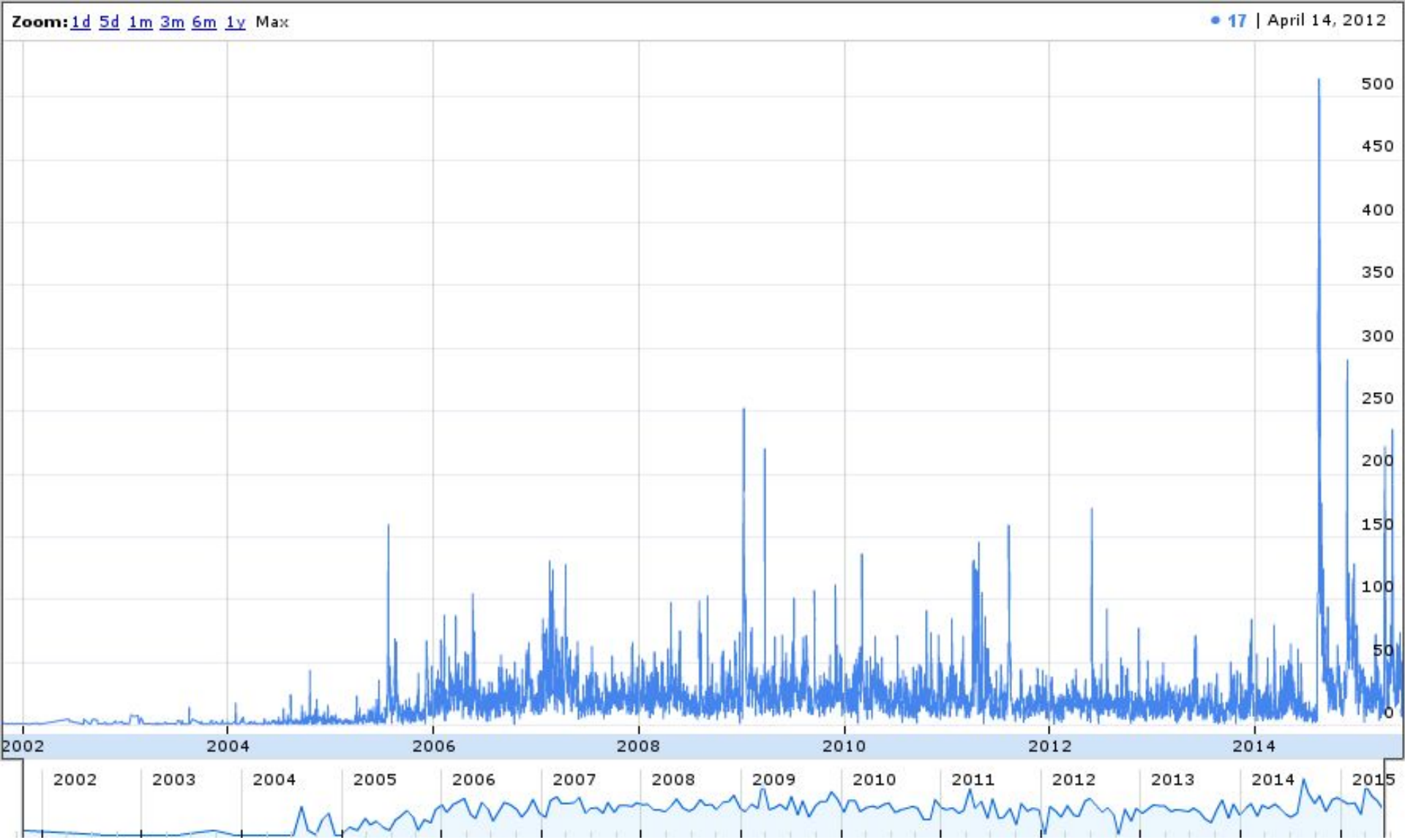


Democratizing Data Science

Benjamin Mako Hill (UW)
Tommy Guy (Microsoft)



Edit per day in English Wikipedia to all articles in the category “People shot dead by law enforcement officers.” (Work done by Mary Dickson during Session #3 of the Fall 2014 Community Data Science Workshops.)


Community
Data Science
Workshops
http://wiki.communitydata.cc/CDSW

Our Goal

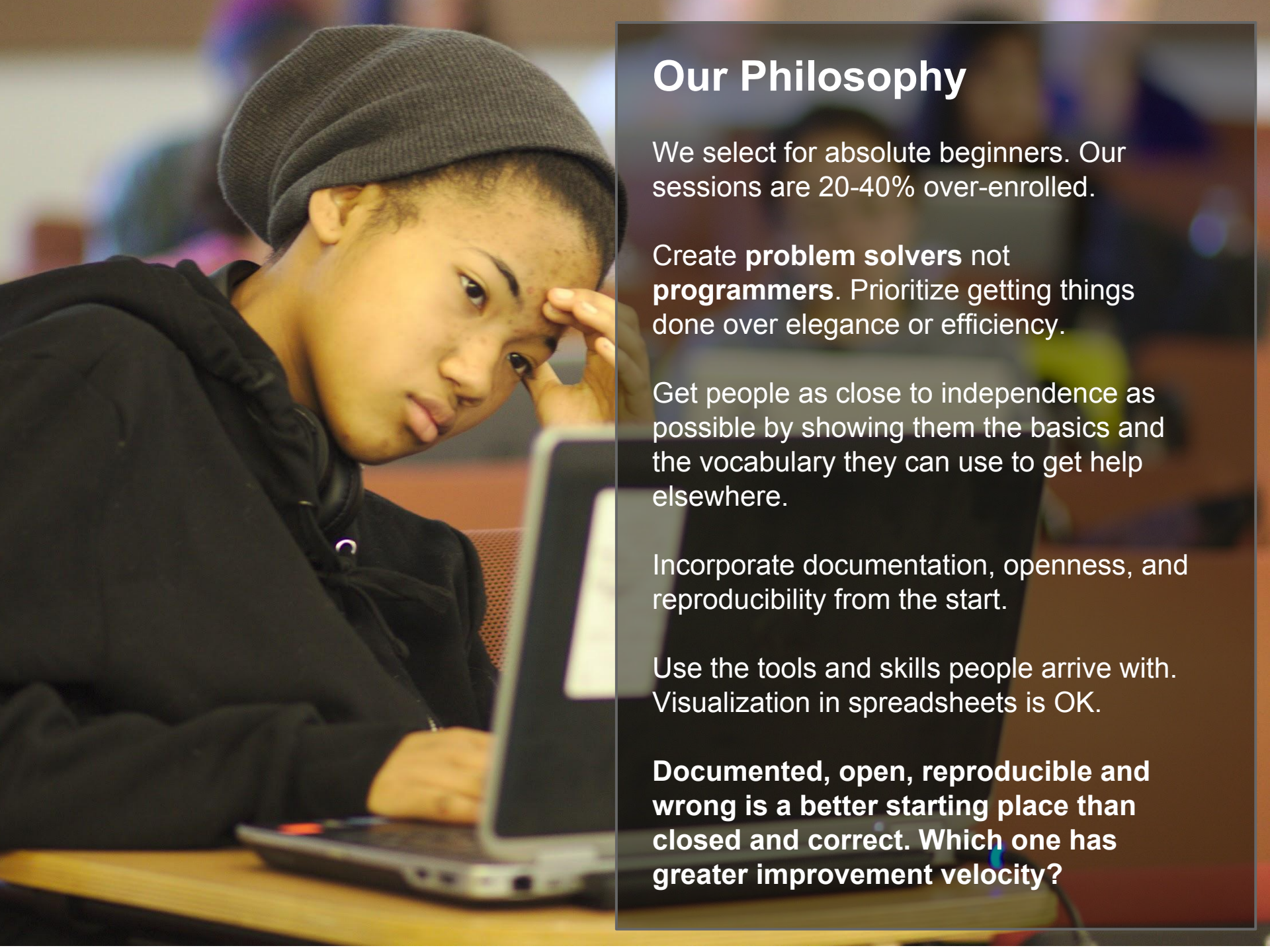
Teach non-programmers to ask and answer questions using Python and web data.

Questions Like

Are new contributors to an article in Wikipedia sticking around longer or contributing more than people who joined last year?

Who are the most active or influential users of a particular Twitter hashtag?

Are people who participated in a Wikipedia outreach event staying involved? How do they compare to people that joined the project outside of the event?



Our Philosophy

We select for absolute beginners. Our sessions are 20-40% over-enrolled.

Create **problem solvers** not **programmers**. Prioritize getting things done over elegance or efficiency.

Get people as close to independence as possible by showing them the basics and the vocabulary they can use to get help elsewhere.

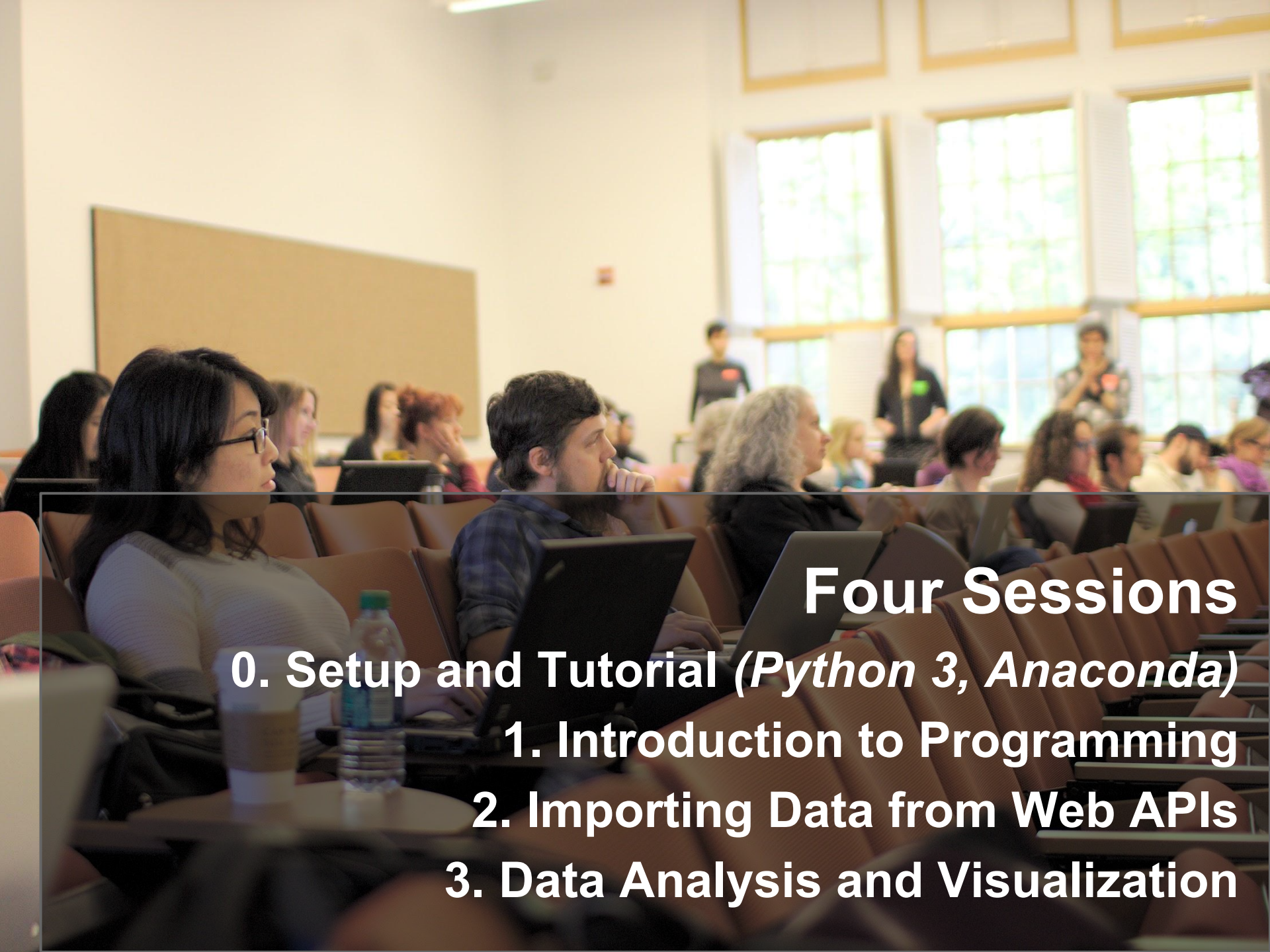
Incorporate documentation, openness, and reproducibility from the start.

Use the tools and skills people arrive with. Visualization in spreadsheets is OK.

Documented, open, reproducible and wrong is a better starting place than closed and correct. Which one has greater improvement velocity?

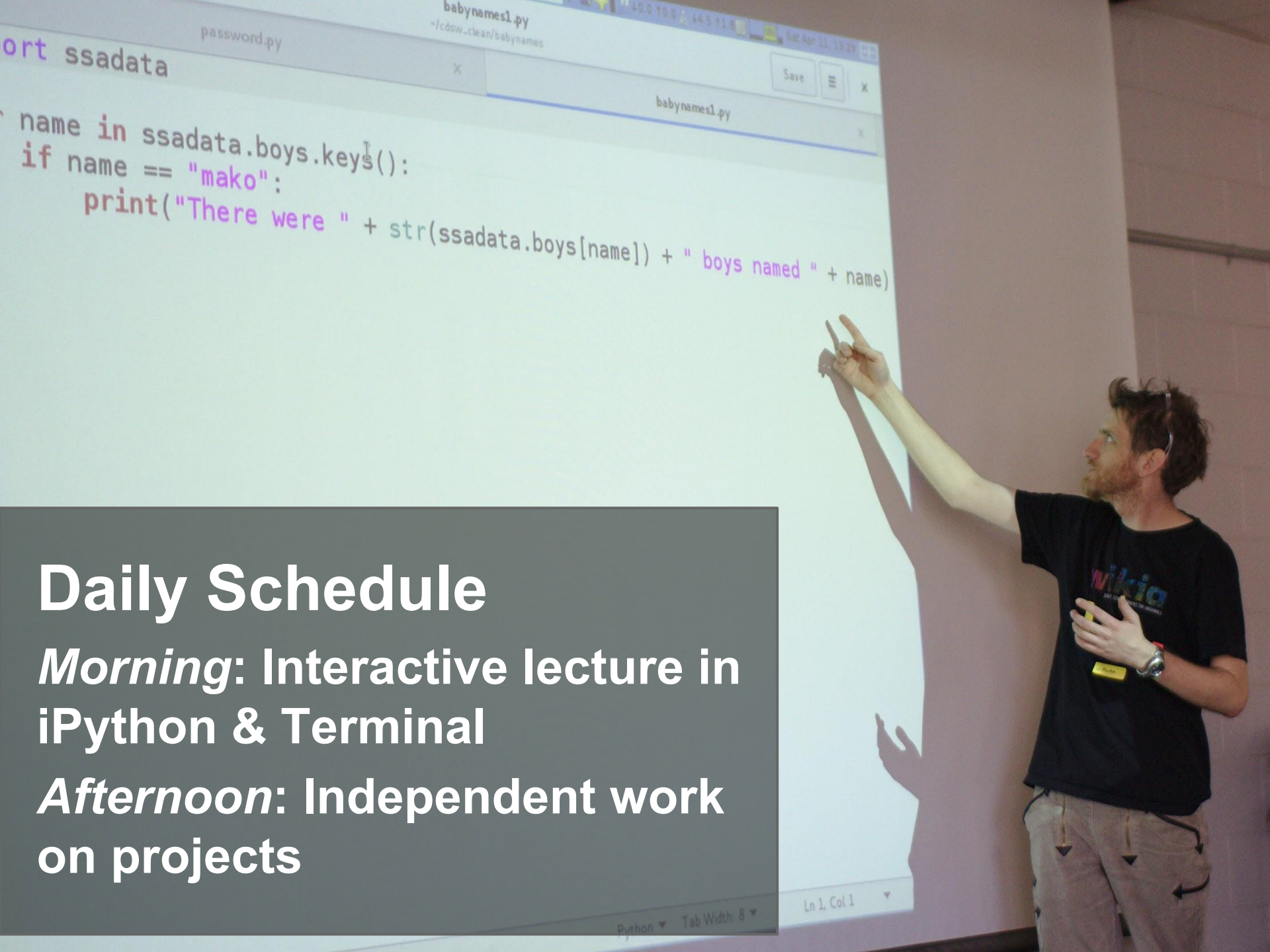
Principles

- Students write *real programs* on *their* computers.
- 5:1 learner:mentor ratio (or better!)
- Project based work.



Four Sessions

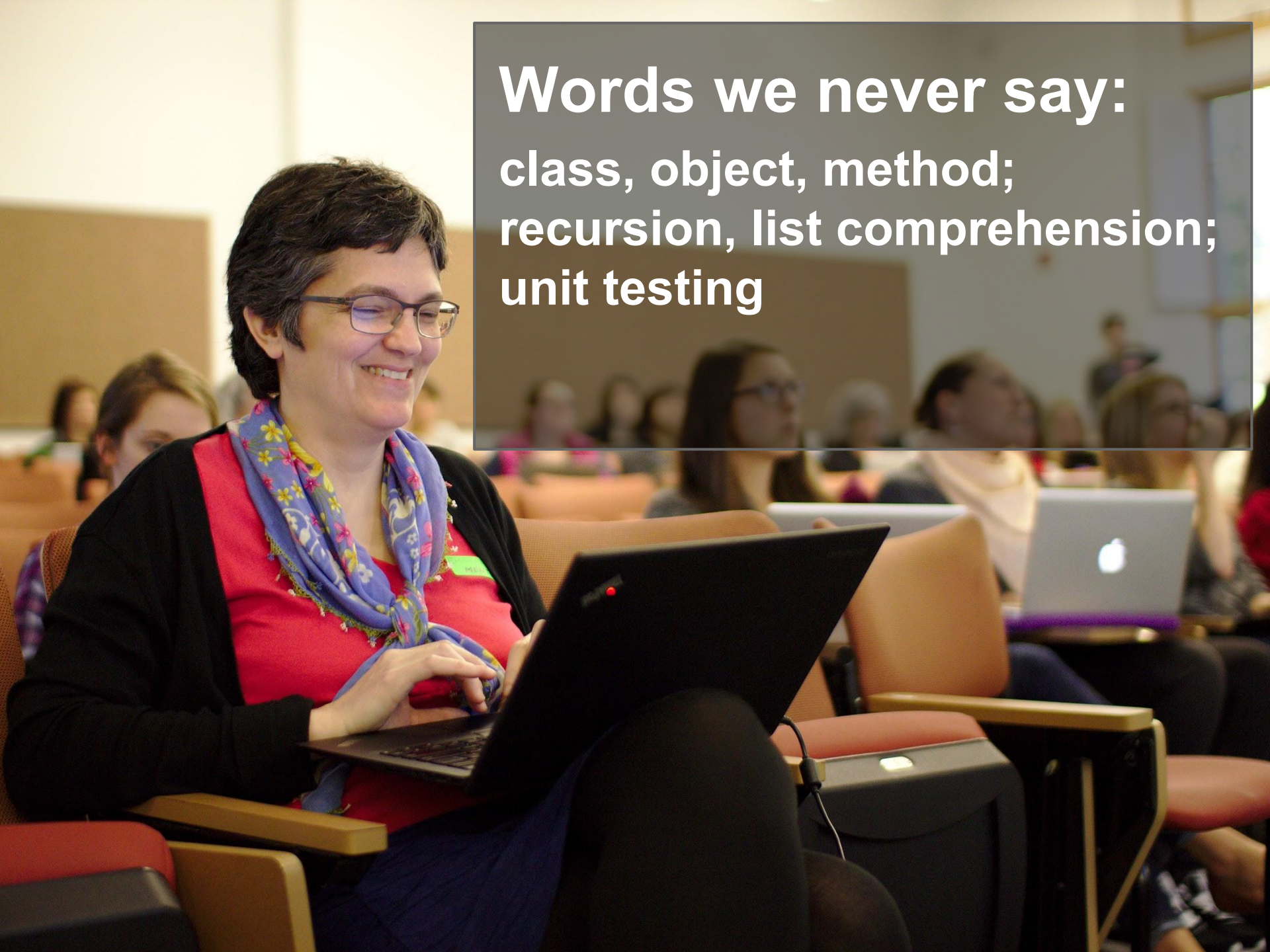
0. Setup and Tutorial *(Python 3, Anaconda)*
1. Introduction to Programming
2. Importing Data from Web APIs
3. Data Analysis and Visualization



Daily Schedule

Morning: **Interactive lecture in iPython & Terminal**

Afternoon: **Independent work on projects**


Words we never say:
class, object, method;
recursion, list comprehension;
unit testing

Many CDSWs!

Seattle (UW): Spring 2014, Fall 2014, Spring 2015

Waterloo: Fall 2014

Chicago and DC: Planning

Sign up on our mailing list
(mentors and attendees!)
http://mako.cc/0m

Run your own session
with our curriculum!
http://wiki.communitydata.cc/

UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

DEPARTMENT
OF COMMUNICATION
UNIVERSITY of WASHINGTON