Smallest font

Welcome

Calibration slide

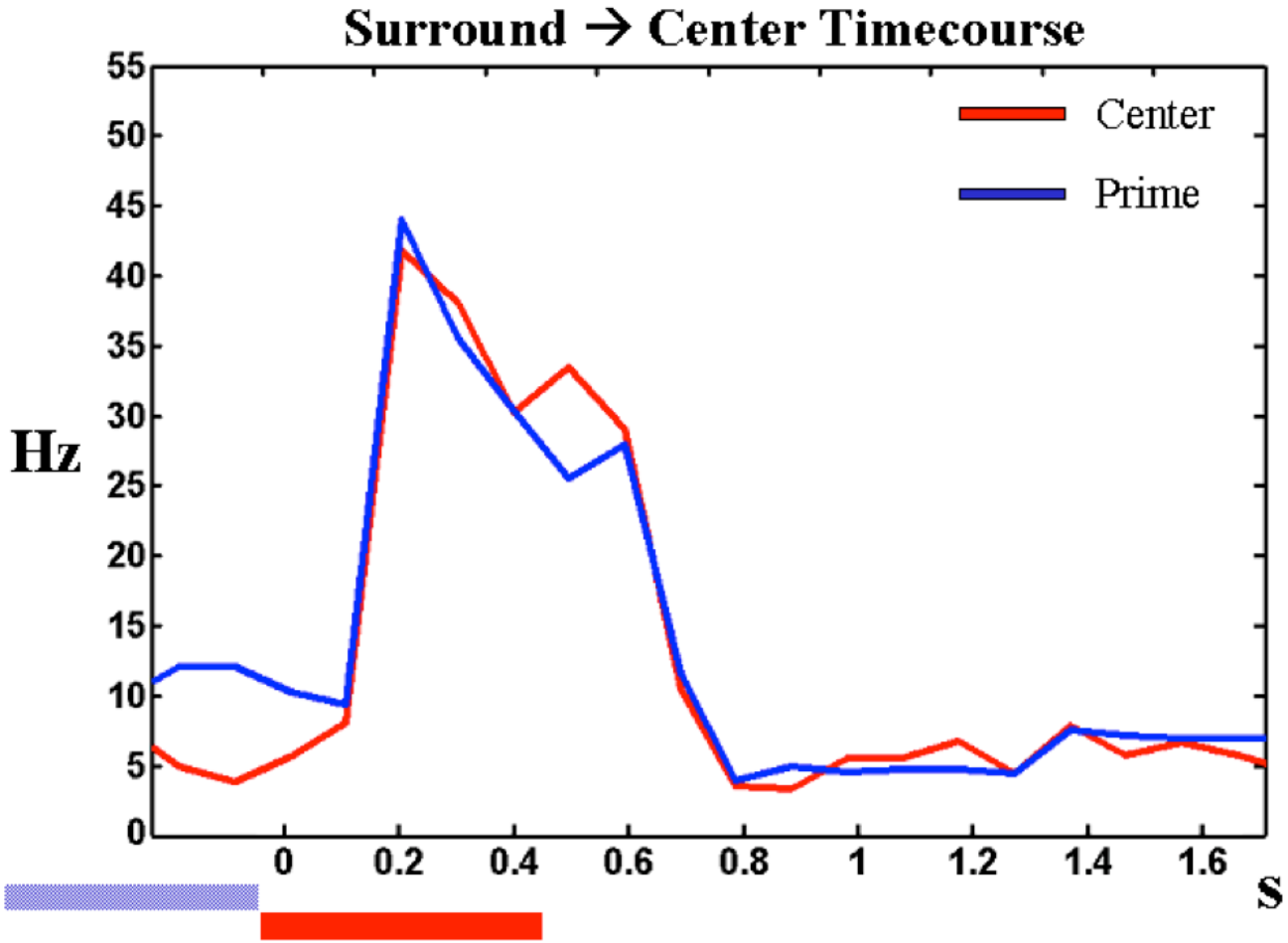
Stand by

Smallest font

Scientific Programming and Computing for the Behavioral Sciences

Welcome to the matrix!

A desperate situation



You

- Where are you coming from?
- Where do you want to go?
- How do you think scientific programming/ computing can help you to get there (what do you want to get out of this class)?

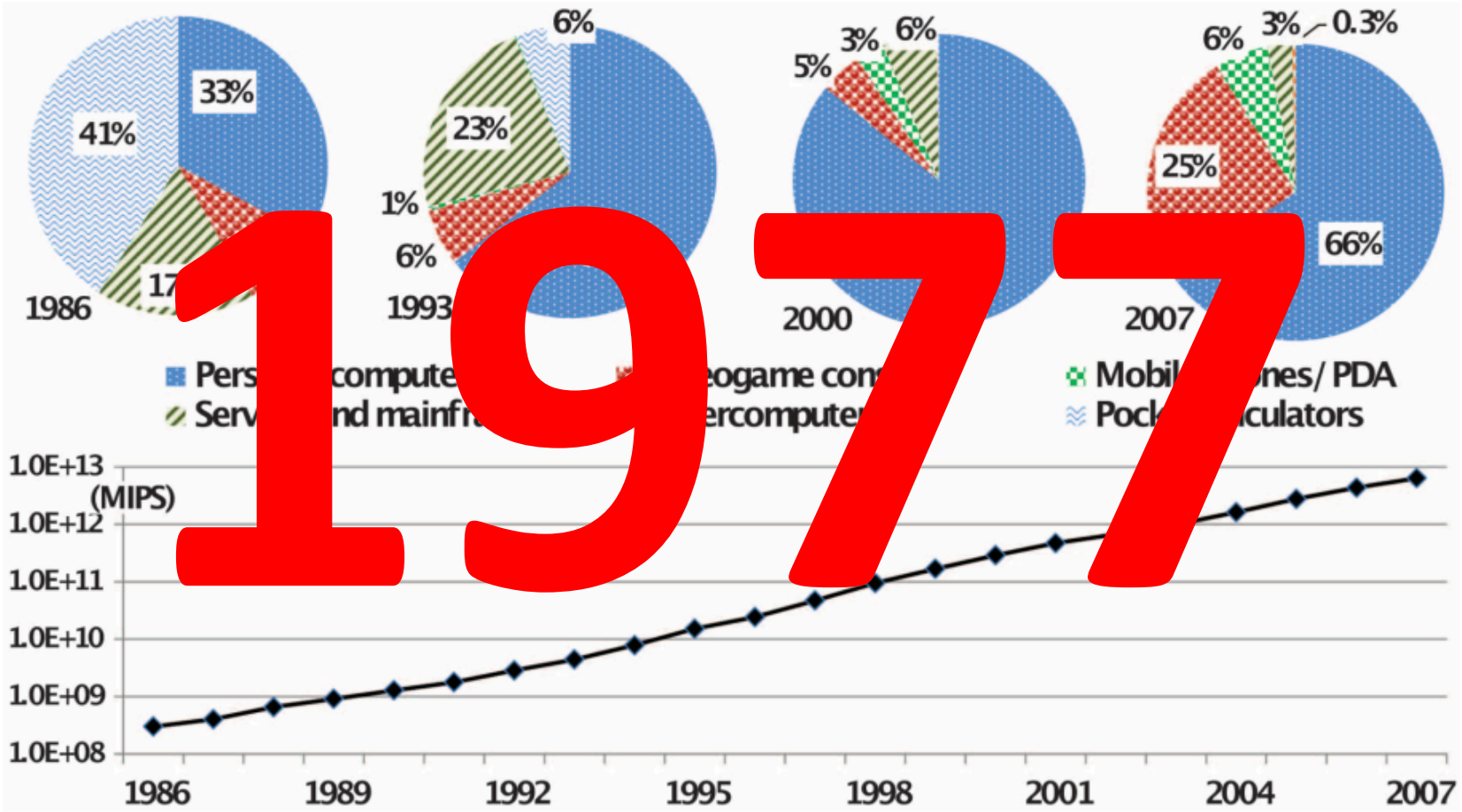
The program today

- Philosophy of this course (Why?)
- Logistics (How?)

Why scientific computing?

The current state of affairs

The computational power at your disposal

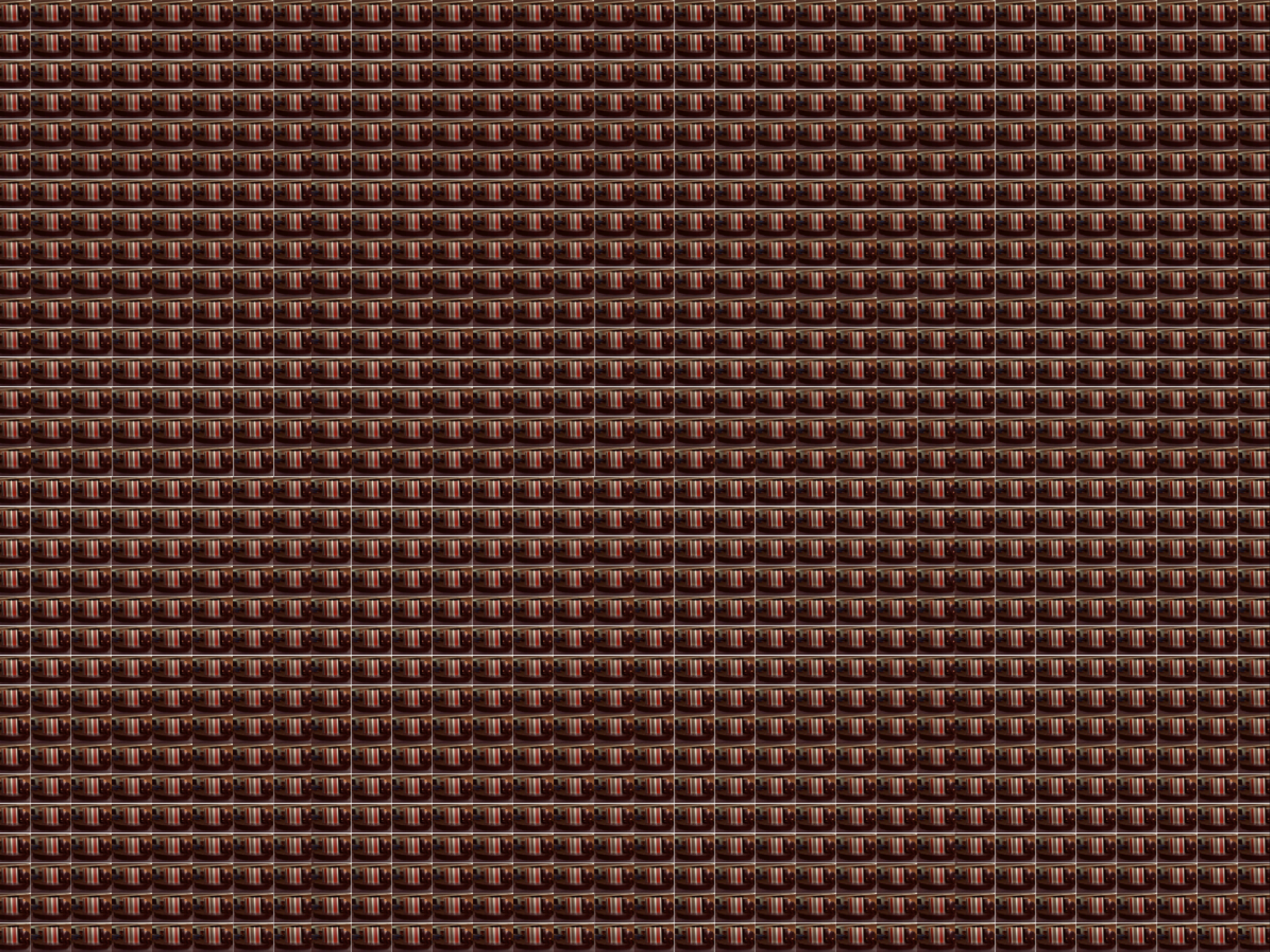


Hilbert et al., *Science*, 2011

Times have changed

Cray 1 (1977)
$8.86 million
80 MFLOP/s

iPhone 7 (2016)
<$800
>160 GFLOP/s



This is a game changer…

Inverting a matrix

- Demonstrating the revolutionary computational power of computers.
- Inverse: $\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{A}^{-1} \times \mathbf{A} = \mathbf{I}$ $\quad\quad$ $(\mathbf{A} \times \mathbf{I} = \mathbf{A})$
- Matrices are inverted in order to solve systems of linear equations (and other things)
- Say you know **A**, **B** and that **XA = B**. What's **X**?
- Can't divide!
- Can multiply with the inverse.
- Inverting a 24x24 matrix by hand takes about a thousand years of continuous work.
- How long does it take by computer?

Even a single Cray was sufficient to do some cool stuff

How most people use these capabilities today

The urgency of developing computational literacy for scientists

- An avalanche/tsunami of data is headed your way.
- Data in the behavioral sciences used to be scarce because it was so hard to record.
- This is no longer the case.
- The computational power to process this data is there.
- You need to develop the skills to use it.

Why programming?

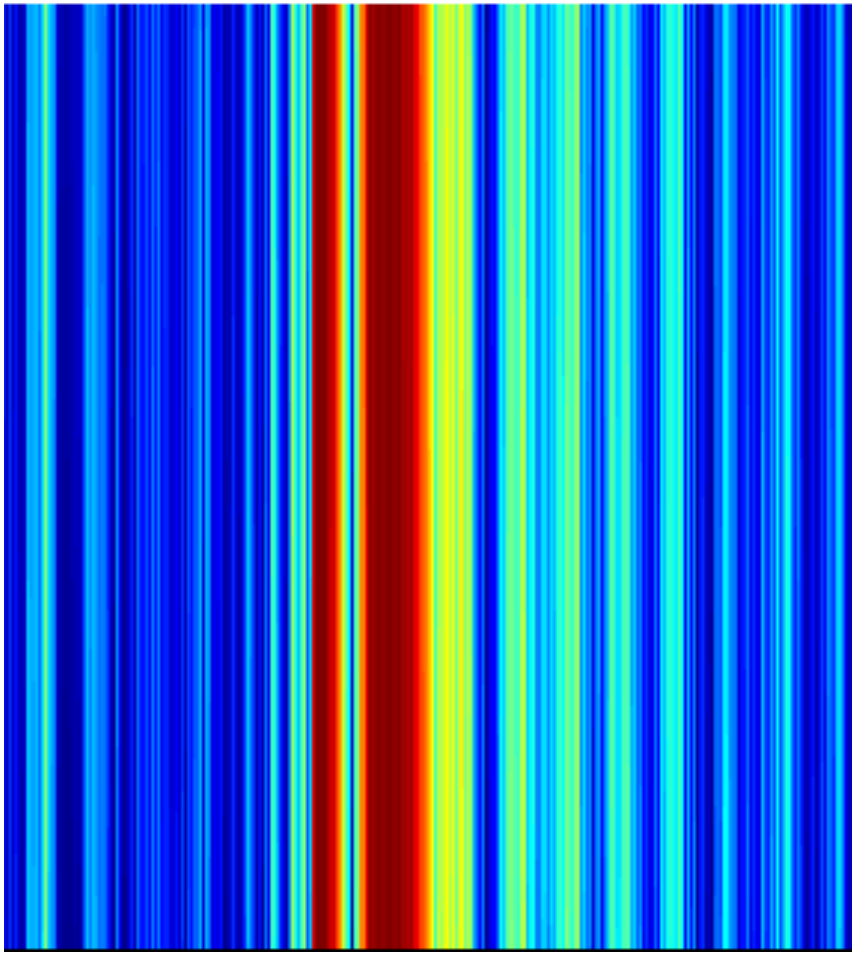
The case for programming

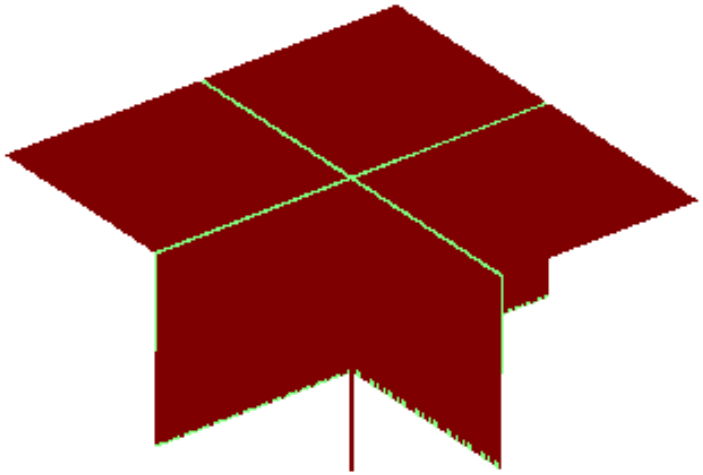
- Society provides a moral hazard
- Enables frivolous behavior due to success.
- Your beliefs don’t have to be true, they don’t have to work. You’ll be ok.
- So people start believing whatever they want.
- We see this everywhere.
- But that is dangerous.
- It can’t go on forever.
- Programming is salutary.
- It has to work.

MATH

Mental Abuse To Humans

A better interface for math





Not a trivial problem…

- "Heavy use of equations impedes communication among biologists."
- *Fawcett & Higginson, PNAS, 2012*
- 35% fewer citations of biological papers for each additional equation per page.
- (!)

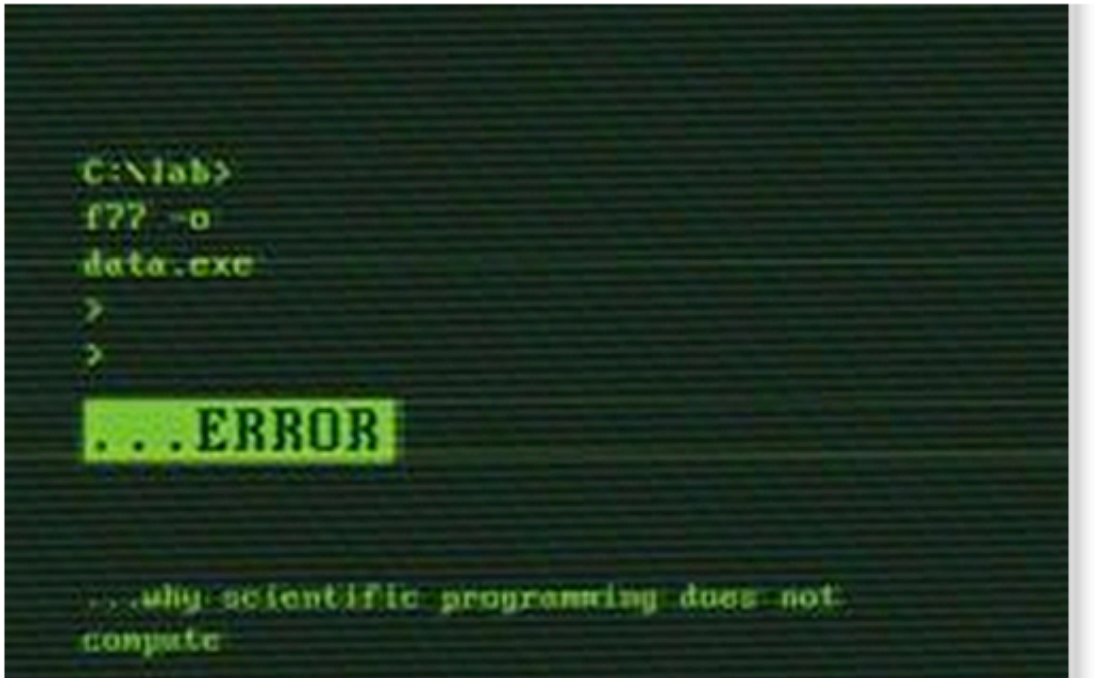
The coding problem is even worse

- Few scientists have formal instruction in programming.
- Few to no industry standards (e.g. version control, unit testing, debugging) are followed.
- Most people stop developing code if it gives them roughly the result they expect.
- Does this have consequences?

It does have dire consequences

...why scientific programming does not compute.

Zeeya Merali



Nature News, 2010

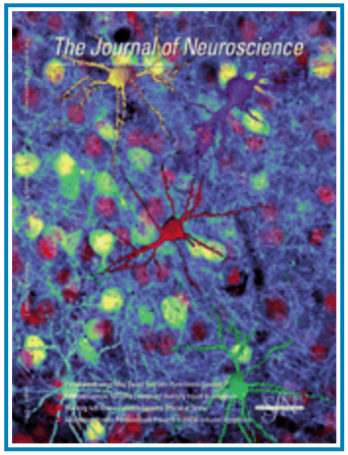
Retractions ensue

Retraction Watch

Tracking retractions as a window into the scientific process

Matlab mixup sinks Journal of Neuroscience paper

with 14 comments



A team of neuroscientists at University of Oregon and the University of California, San Diego (UCSD) have retracted a paper from *The Journal of Neuroscience* after realizing their analytic code contained an error.

The authors state in the notice that their conclusion remains accurate after correcting the mistake in the program Matlab. However, the paper — which examined the role of neuronal oscillations in working memory — still contained "some findings that we no longer believe to be robust."

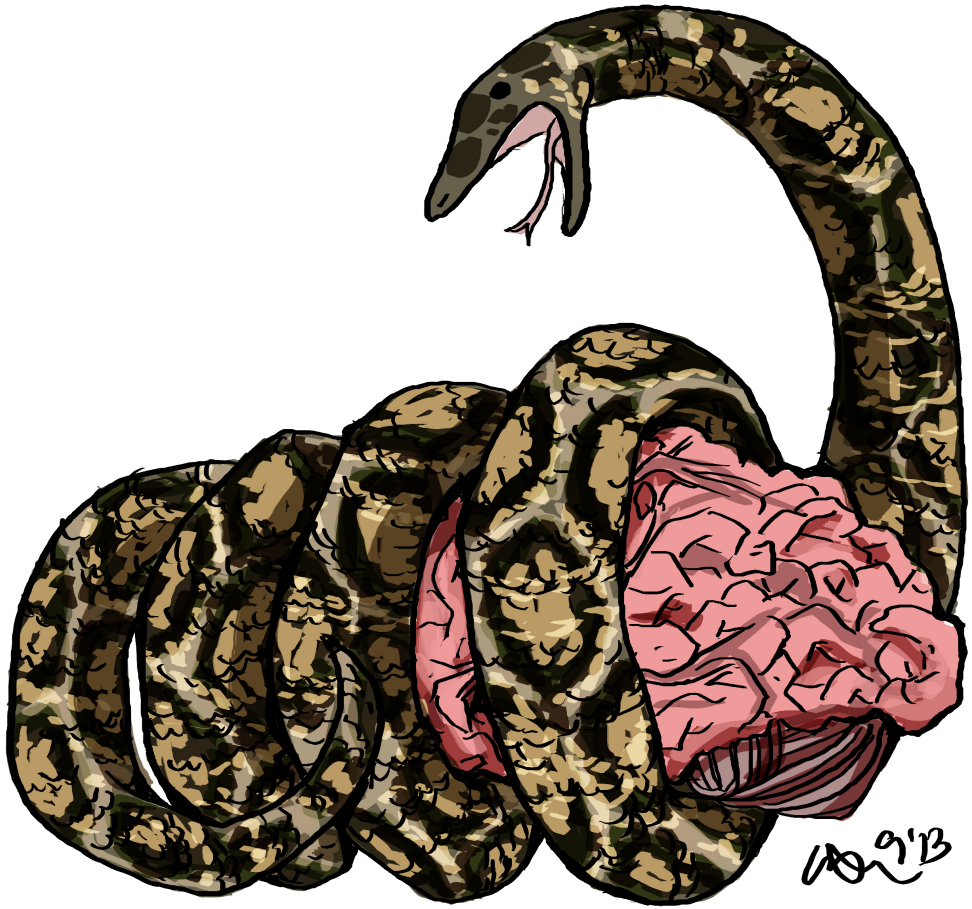
It's a very useful notice:

> At the request of the authors, *The Journal of Neuroscience* is retracting "Induced Alpha Rhythms Track the Content and Quality of Visual Working Memory Representations with High Temporal Precision" by David E. Anderson, John T. Serences, Edward K. Vogel, and Edward Awh, which appeared on pages 7587–7599 of the May 28, 2014 issue.
>
> We regret that there was an error in the analytic code used to compute oscillatory power in our article. Specifically, there was a matrix transposition error in the code (see abs(hilbert(eegfilt(data,Fs,f1,f2))).2 on page 7588, right column, end of second full paragraph). The data matrix was oriented correctly for the call to eegfilt, but the output of the call to eegfilt was not correctly transposed in the standard Matlab format before passing into the built-in Matlab 'hilbert' function, as the EEGLAB function 'eegfilt' and the built-in Matlab function 'hilbert' require the data

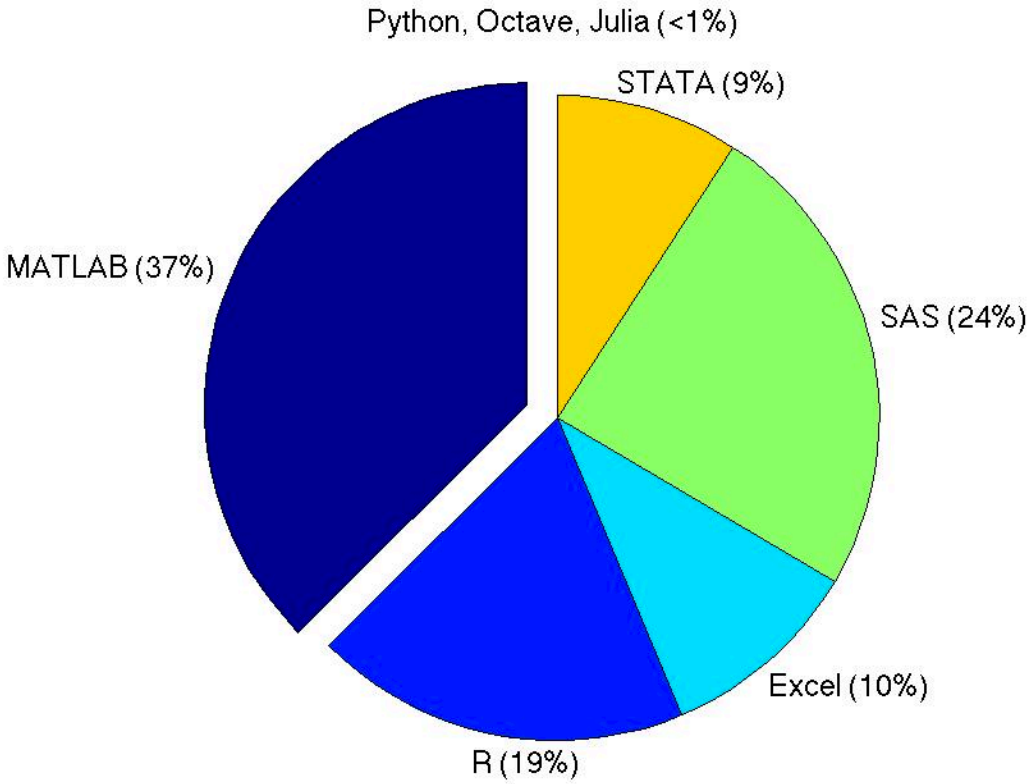
There are also practical considerations

- We will use Matlab
- The de facto lingua franca in the community
- Most labs use it
- Large user base
- Beginner friendly
- Optimized for data analysis and visualization

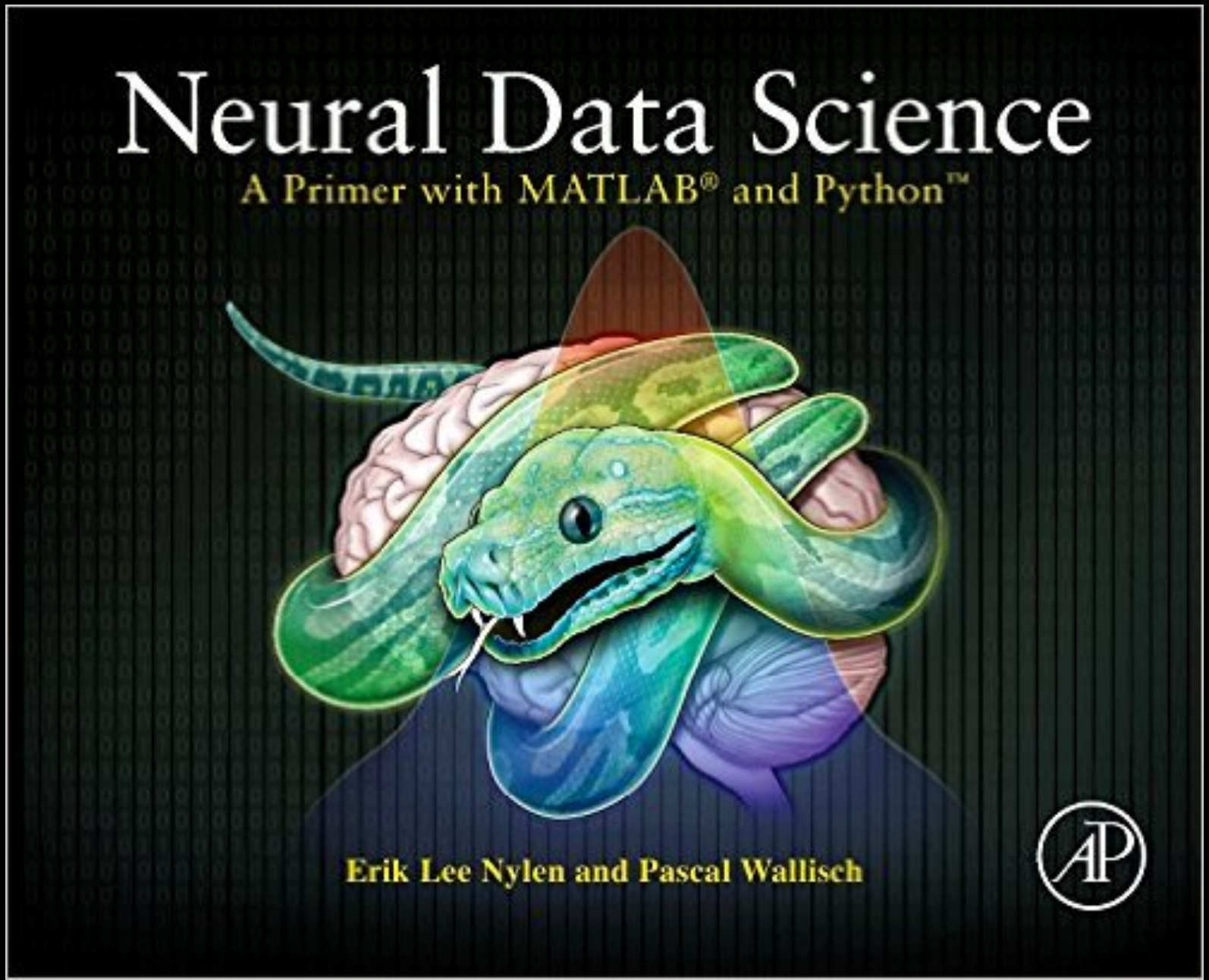
Why not Python (yet)?



Python is the (online) future

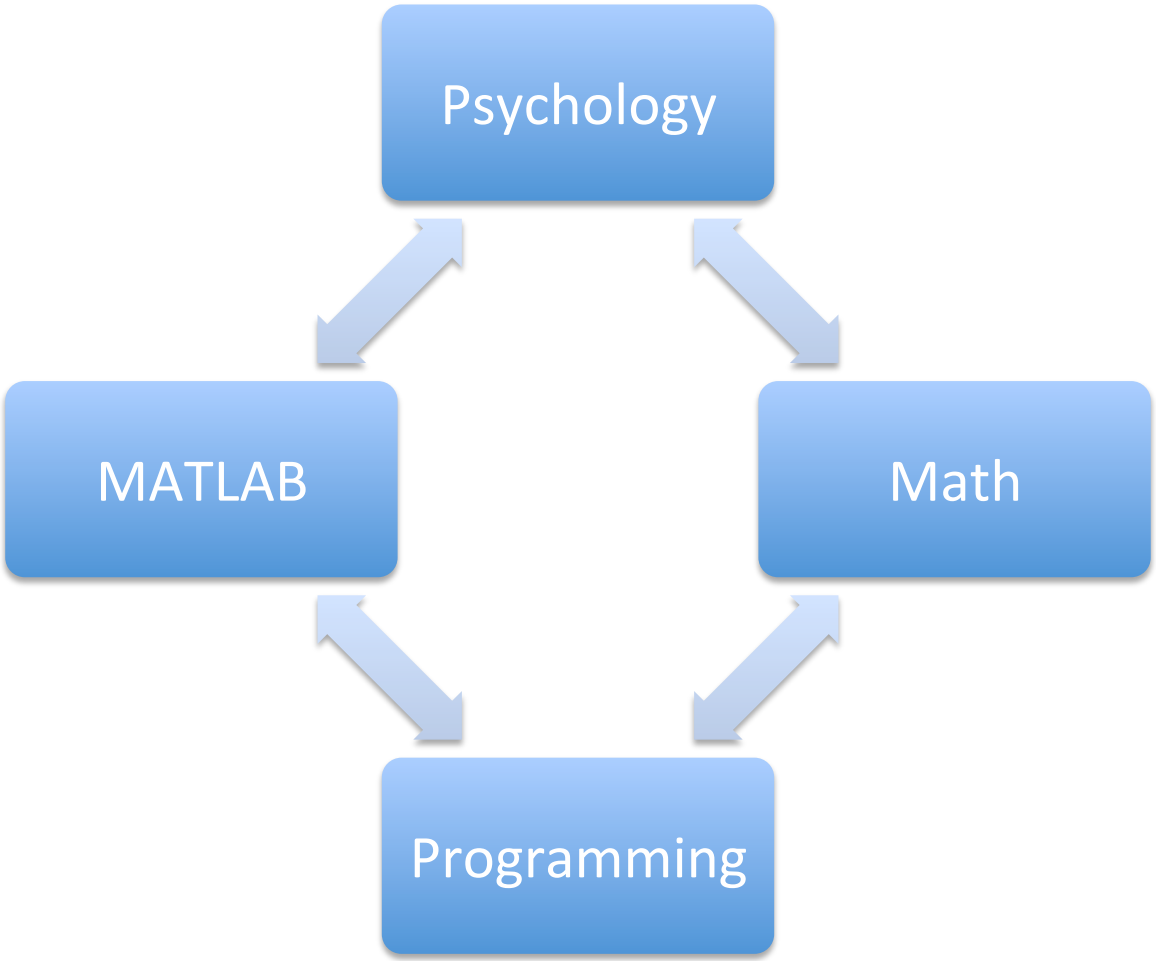


BUT… don’t fret



How to teach this class?

By juggling



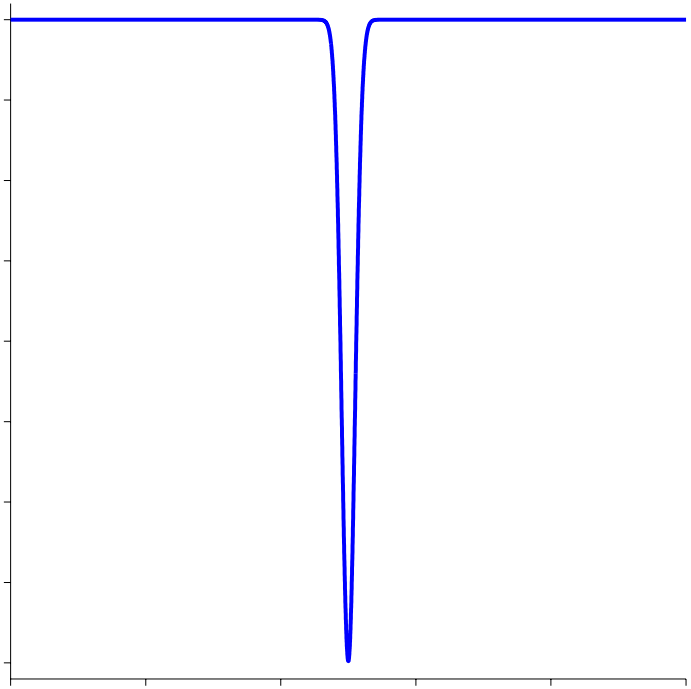
The challenge of teaching well

- Breadth vs. Depth (shallow vs. thorough)
- Abstract vs. Specific
- Reception vs. Production

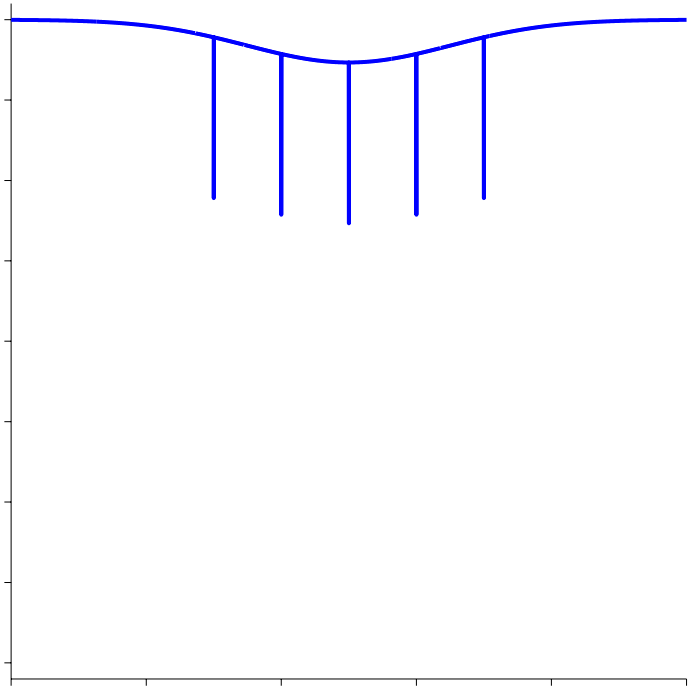
- Any teaching is a tradeoff of these three factors.
- Effective teaching is an optimal mix, given the subject matter, the students and the teaching goal.
- There is also a scale factor involved. In addition to making this tradeoff, the question is whether the teacher can motivate students and convey information.

Appropriate balance of breadth vs. depth depends on use case

Research



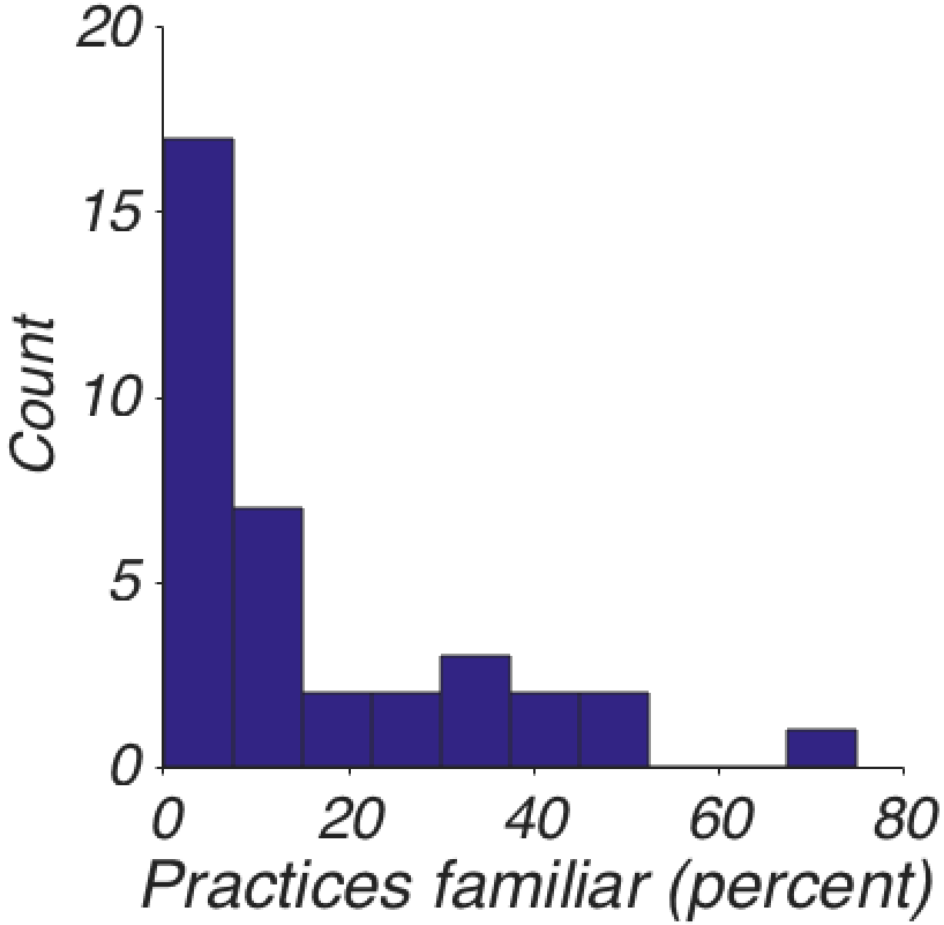
Teaching



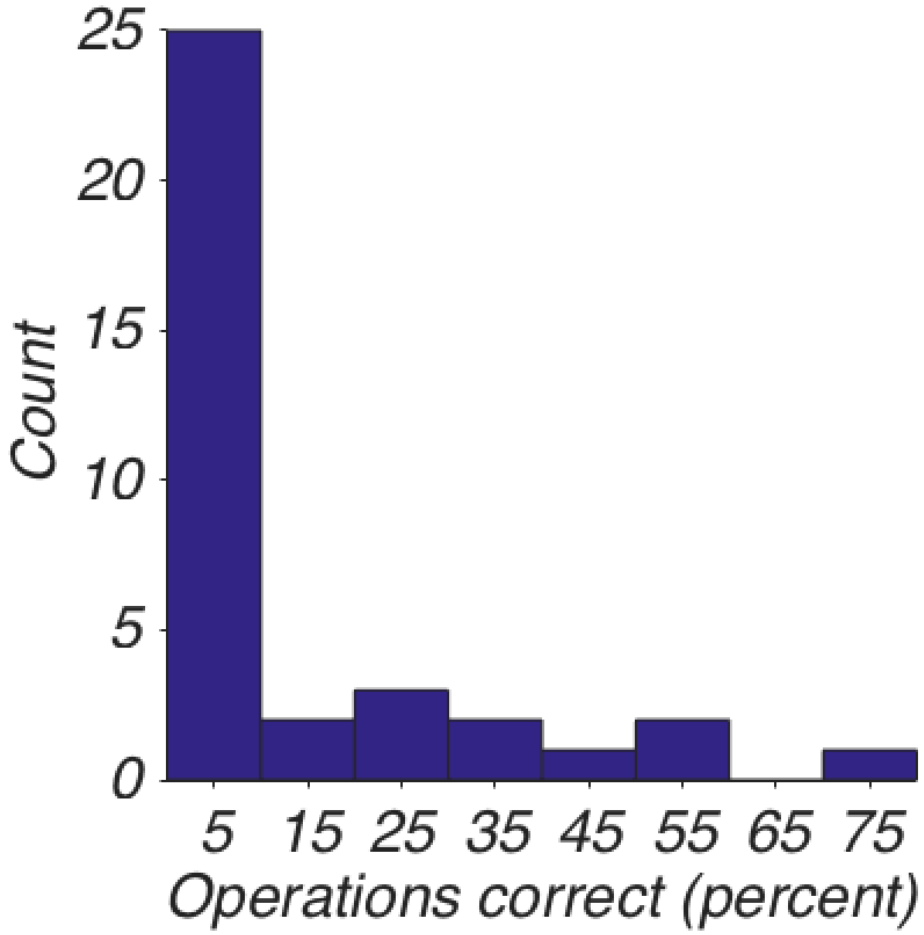
Student baselines have to be considered

- What is our starting position?

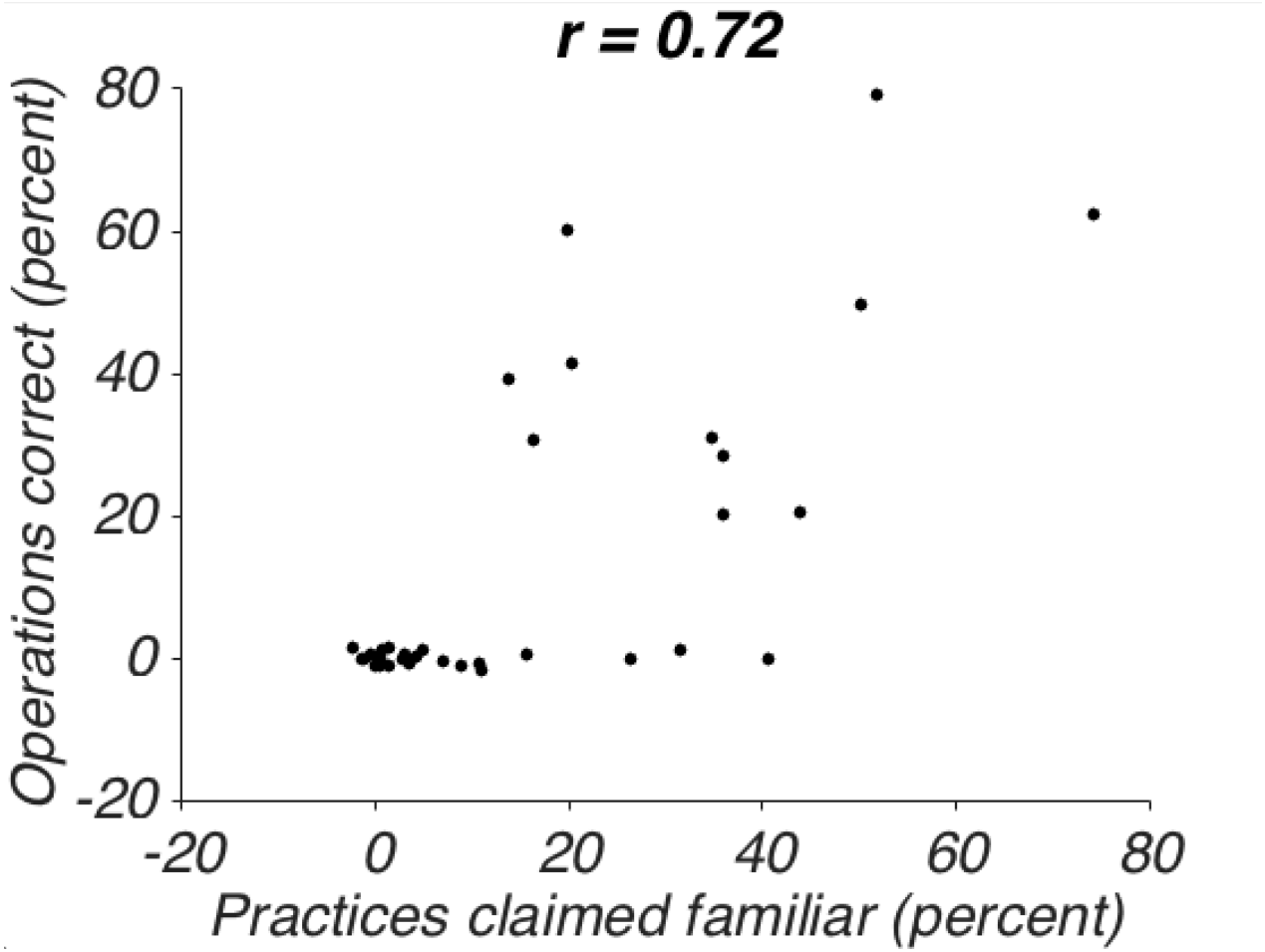
The starting position: Familiar practices?



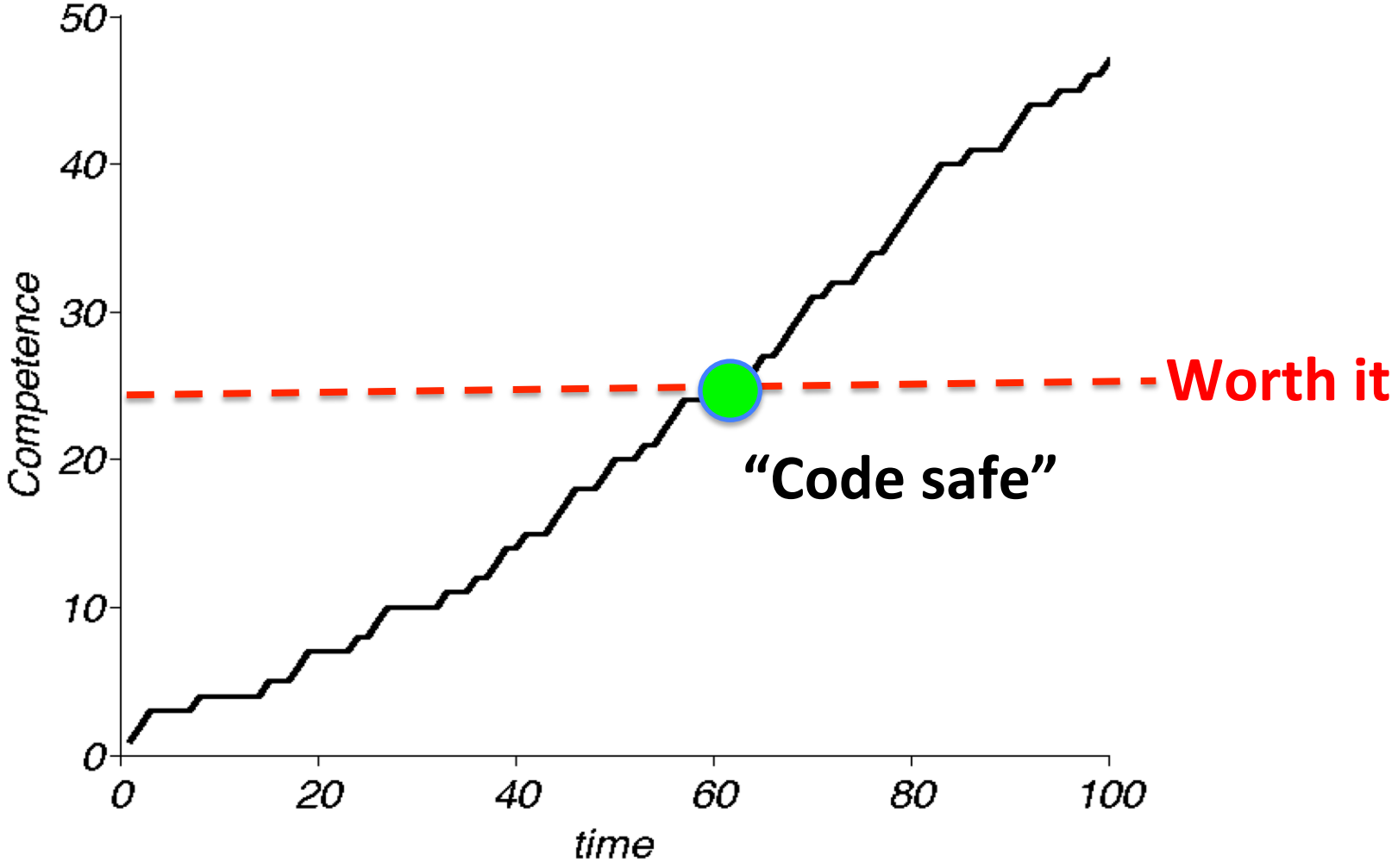
The starting position: Basic operations?



Overconfidence?



The problem of gaining expertise



Downsides:

- Effort (energy, time, money)
- Discouragement

Upsides:

- Rewards (intrinsic, extrinsic)
- Encouragement

The goal of this class

- To get you to a point where you are comfortable and able to pursue your own solutions for scientific problems with Matlab
- ~100-150 hours
- Shortcut (saving you maybe a couple hundred to a thousand) and training wheels (limiting the downside)
- “Code safe”

A final word about “the best” code…

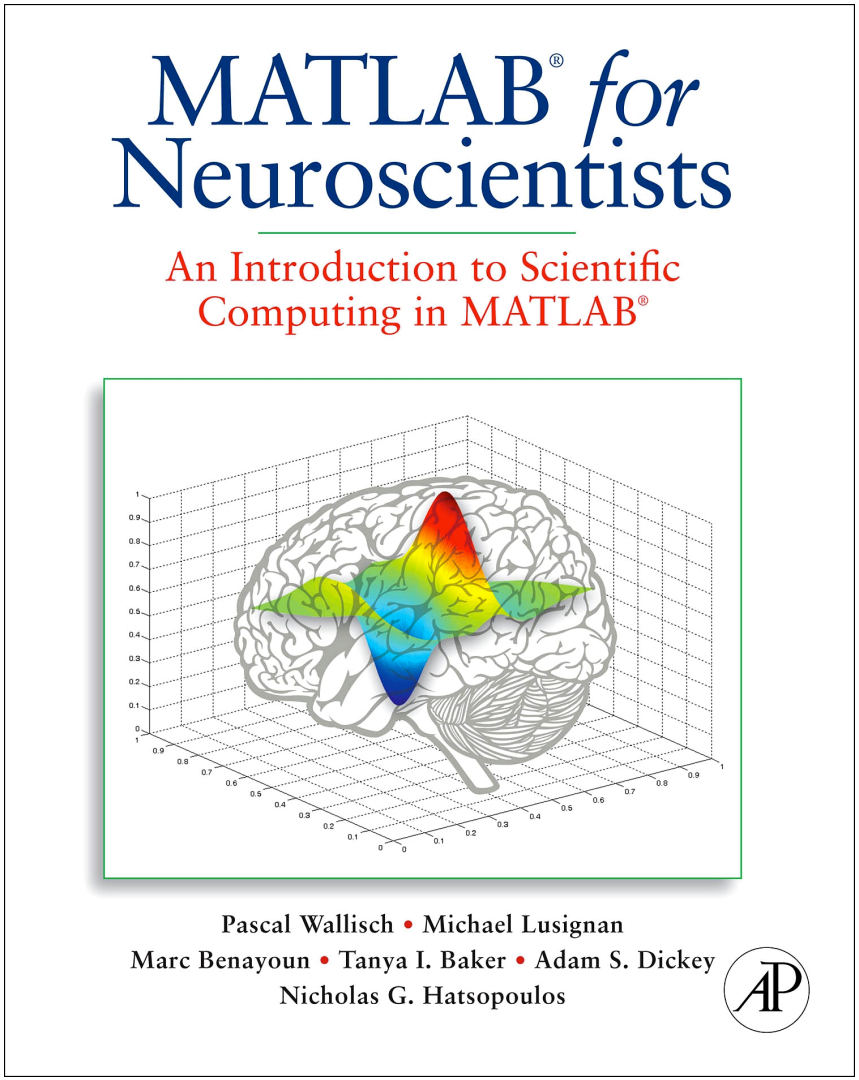
- Most efficient code (fastest?)
- Most robust code (redundancy, hard to break)
- Error-free code
- Easy to read code
- Fast to write code
- Most versatile code
- Well organized code
- …
- Can’t have it all. Priorities depend on use case.

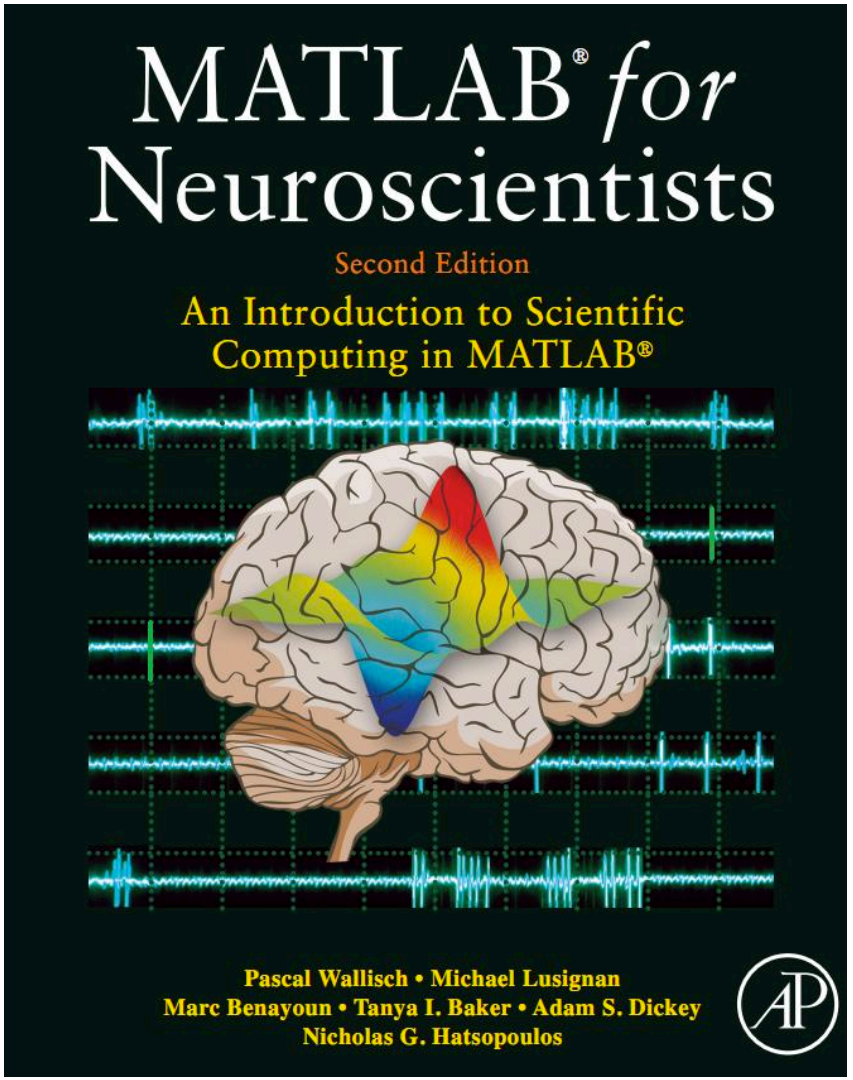
Logistics: The Sittyba

Logistics: Matlab

- Get a student version
- Use the virtual lab

Logistics: The textbook





Logistics: The labs

- Headed by Deshawn
- Offered on Thursday
- 2 hours
- Help with code and assignments.
- Not mandatory, but extremely helpful.
- Kind of mandatory
- Not (yet) this but next week.

Logistics: The buddy system

Logistics: Code peer review

- Review code of 2 peers each week.
- To gain multiple perspectives on how to implement something.
- Particularly useful – particularly for beginners – to see how else someone implements something (and practice writing readable code from the very beginning).

Logistics: The weekly workflow

- Mo, 11:59 pm: Code review due
- Tuesday 12.30 to 3: Class
- Thursday: Lab
- Saturday, 11:59 pm: Code creation due

Skills can be taught

- This class really does work – if you put in the time (18-20 hours outside of class time per week).
- Most programming classes don't enable the student to actually program at the end of it.
- This one does.
- I get a lot of testimonials afterwards.
- Like this one (just last week)

Former students

I enrolled in this class for the sole purpose of acquiring a new skill that I could put on my resume.

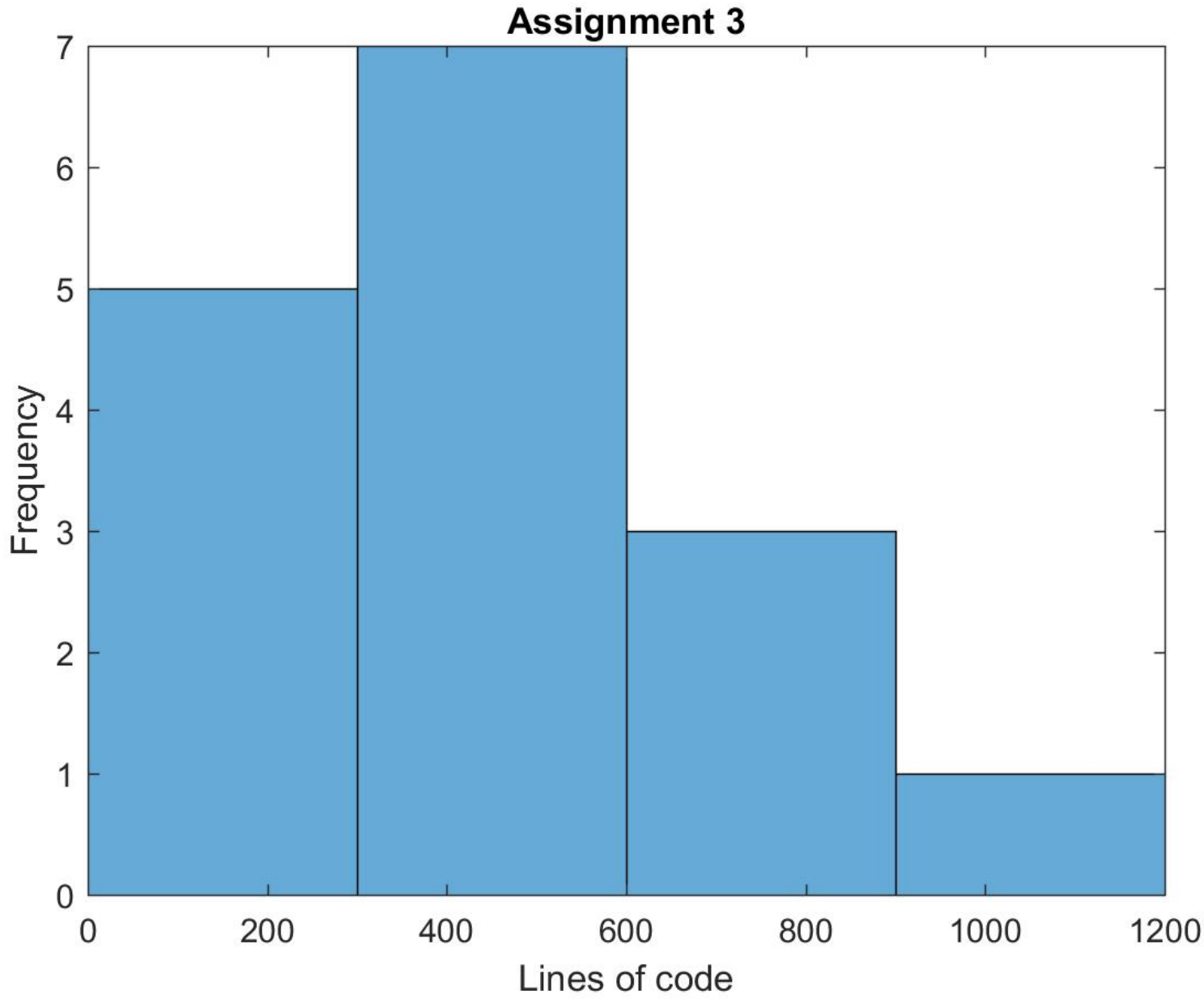
I am very interested in medical research and I felt that a class like this on my transcript would be very helpful when applying to jobs during my gap year(s) between undergrad and eventually, hopefully, medical school.

I was right. Three months after graduation I landed a full-time research position with the number one cancer hospital and research center in the country.

During my interviews I was able to talk about working with not only SPSS, but Matlab as well, in order to collect and analyze data. My interviewers expected SPSS but they seemed to be very impressed that I was familiar with Matlab. I truly think this helped to give me an edge.

Before learning how to program experiments and analyze data in Matlab, I was blind to reality.
That changed once I learned how to program in Matlab, which also allowed me to switch from collecting to analyzing data. I feel creative and in control of what I do with my data.
Now I am on track to being a published scientist at the age of 23, thanks to everything I learned from one of the world's leading experts.

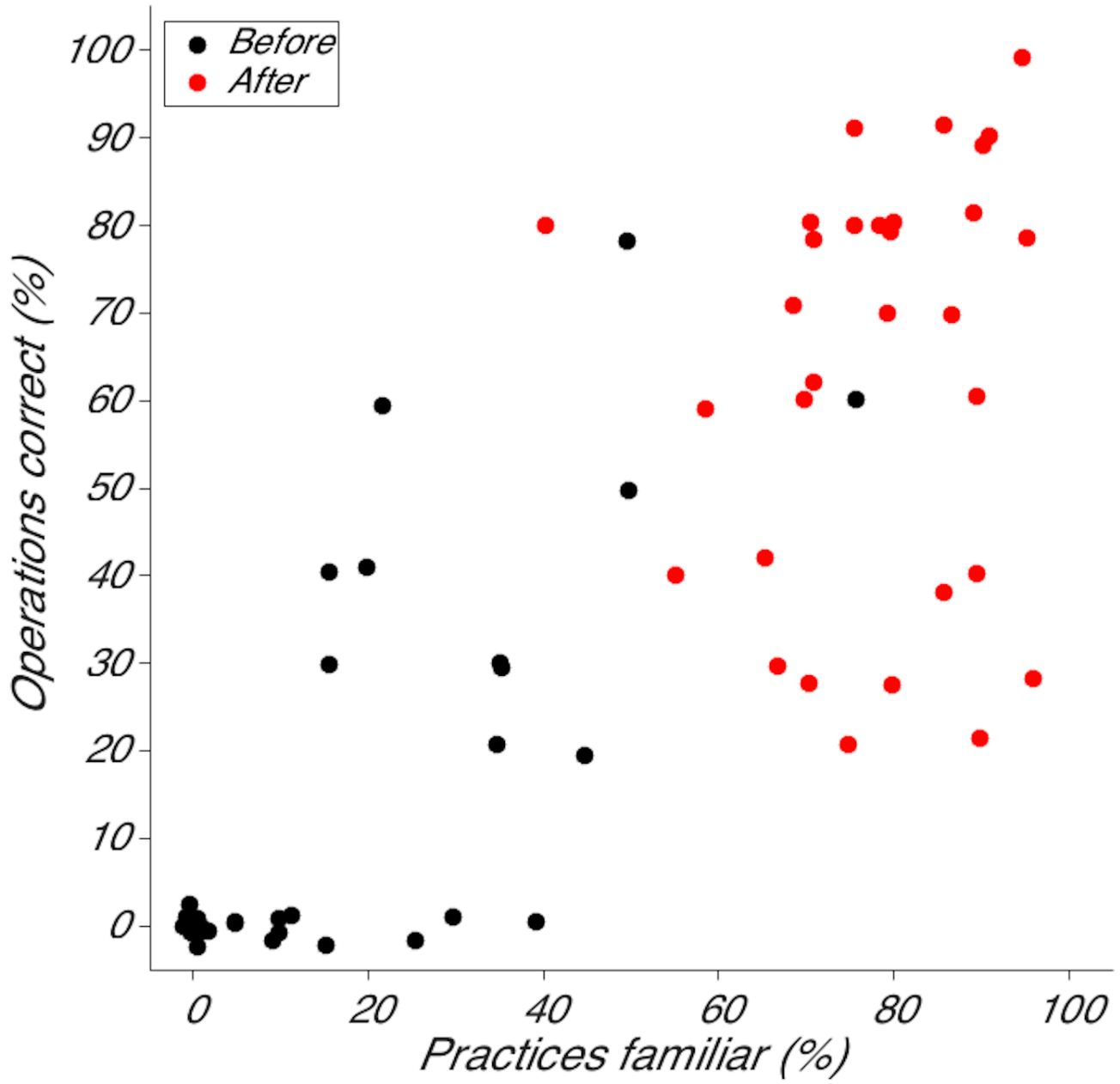
This is an intense class



Potential outcomes on the table

- “Code safety” status
- Jobs
- Educated citizenship
- Ownership of the research cascade
- Math tools stepping stone (feeder class)

Actual class outcomes



Before we begin to get started

- This is an opportunity.
- And a challenge
- If you meet it, it will be transformational.
- If you don’t, it will be a complete waste.
- Outcomes in this class are usually binary.
- If you stay, make sure to be on the right side of this divide.

Anything else?