CMPU 101 § 2

#### Computer Science



5 December 2022



How Technology Changes Our Perceptions and Decision-Making Capabilities Prairie Goodwin, Vassar College 4:30 pm, 5 December 2022 SP 105

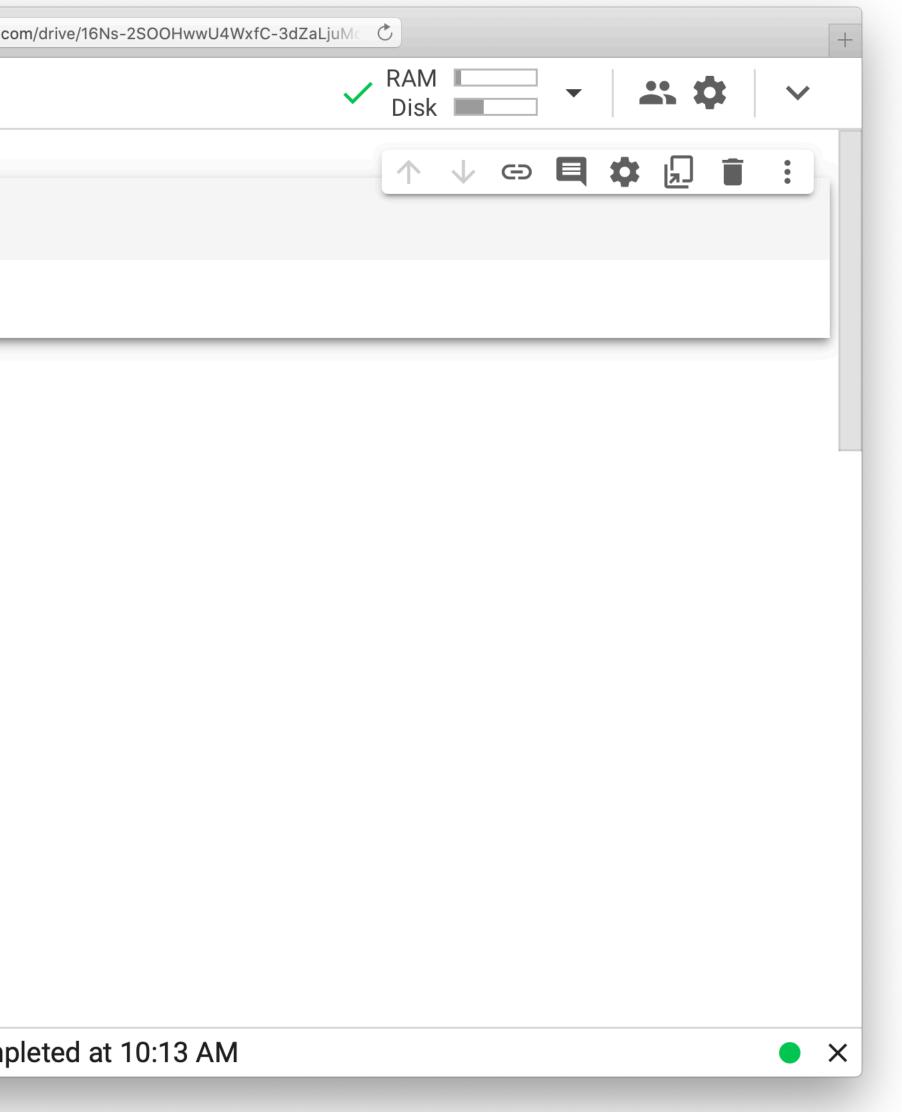


## What have we been doing this semester?



code.pyret.org/editor	C			+
	Rur	•	Stop	
"Hello,	computer!"			
>>>				

						b.researc	h.google.co
:=		+ Cod	e + Te	ext			
Q	✓ Os	0	"Hell	o, comp	uter!		
{ <i>x</i> }			'Hello	o, comput	er!'		
<>							
=							
>_							
					~	0s	comp



We're not especially interested in Pyret – *or* Python! If you're programming 20 years from now, it'll be in a different language, using different tools.

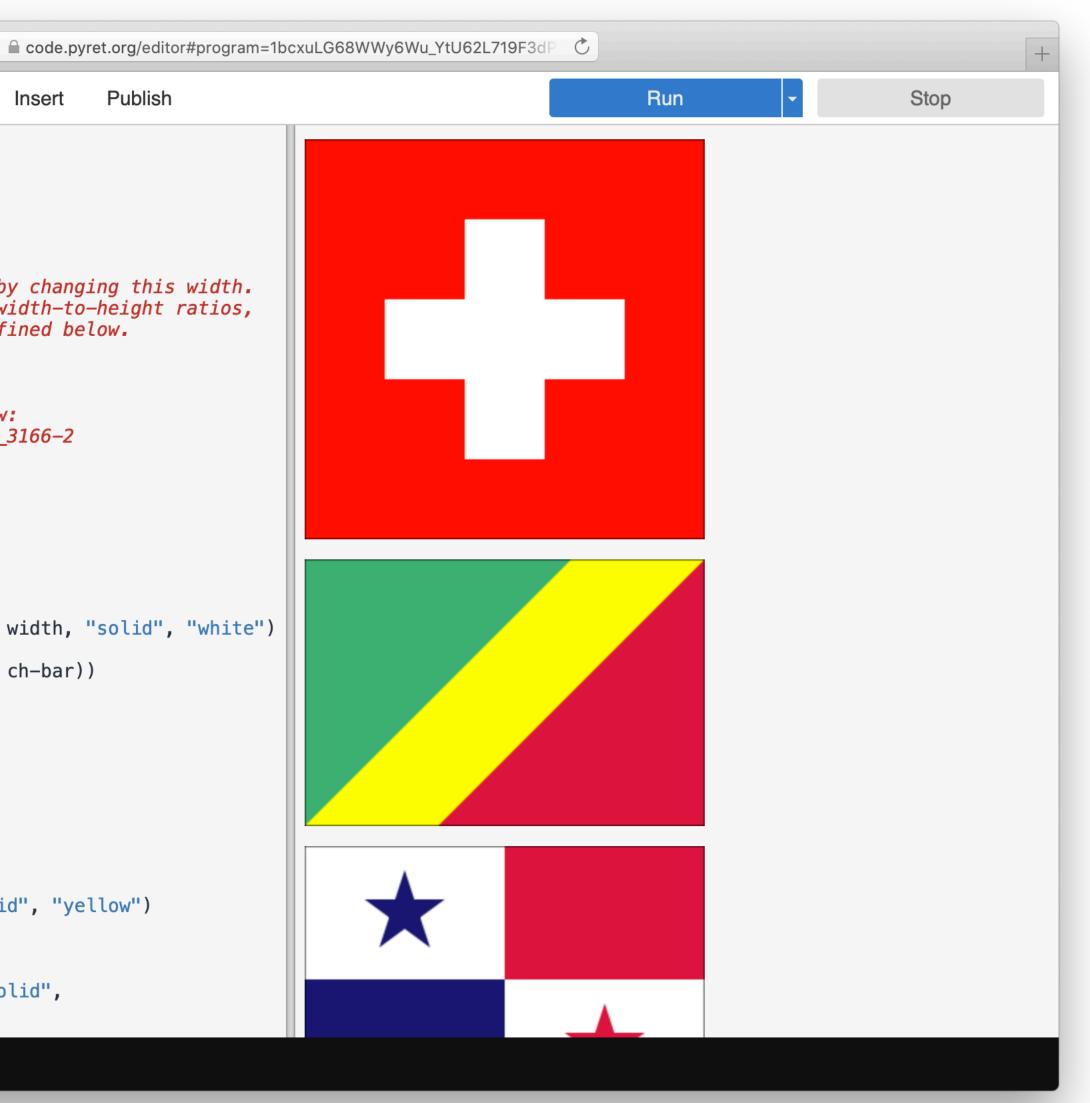
#### What have we been **doing** in these languages?

```
\bullet \bullet \bullet < >
- 🤶
                   ✓ File (asmt01-solns.arr)
                                         Insert Publish

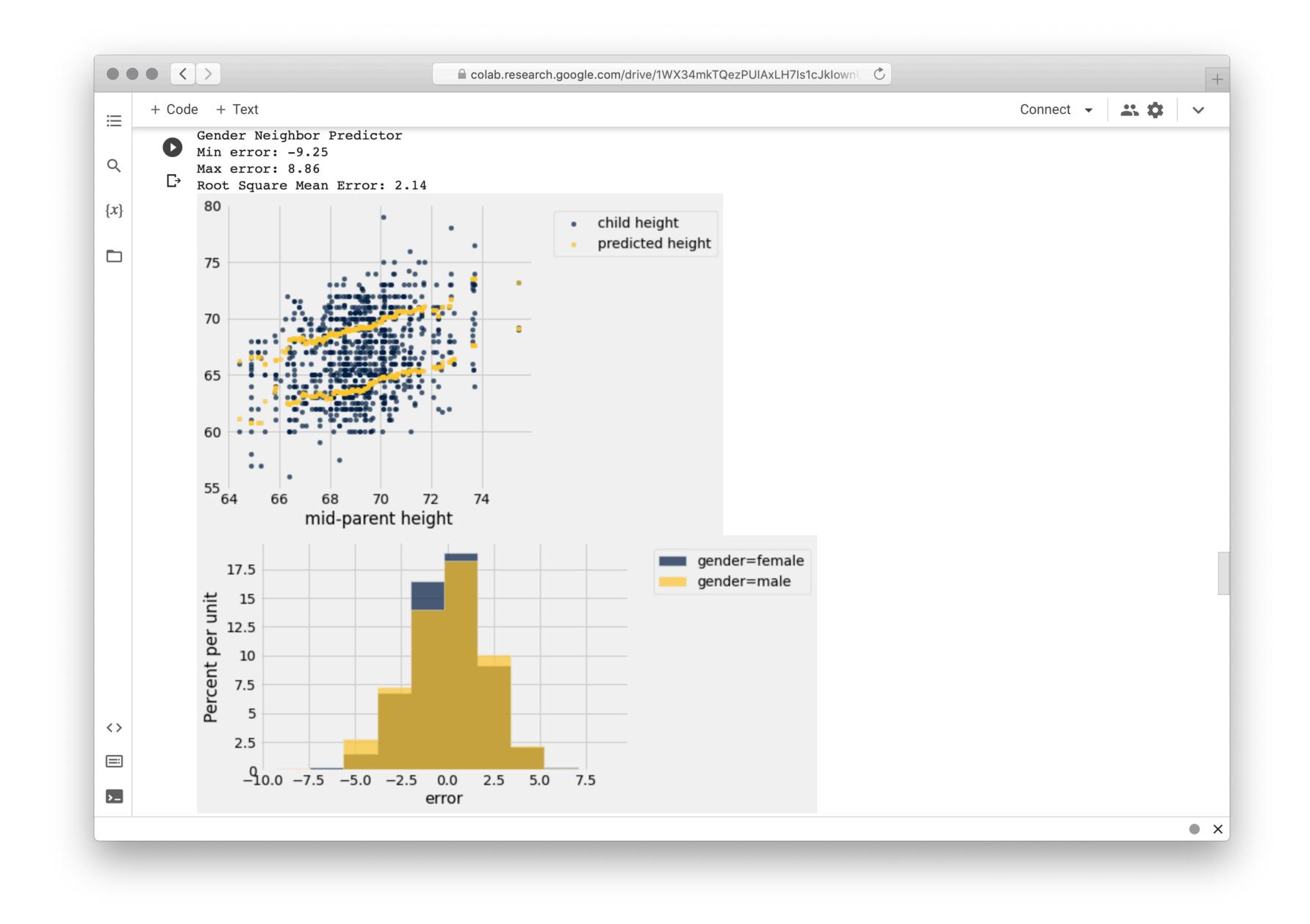
    View

1 use context essentials2021
 2
3 # Assignment 1: Fun with Flags
4 # Example Solutions
5 # CMPU 101, Fall 2022
 6
8 # You can scale the flags up or down by changing this width.
9 # Note that the flags have different width-to-height ratios,
    so they don't all use the `height` defined below.
10 width = 300
11 height = width / 1.5
12
13 # The country abbreviations used below:
14 # https://en.wikipedia.org/wiki/IS0_3166-2
15
16
17 #
18 # Switzerland
19 #
20
21 ch-bg = square(width, "solid", "red")
22
23 ch-bar = rectangle(0.6 * width, 0.2 * width, "solid", "white")
24
25 ch-cross = overlay(ch-bar, rotate(90, ch-bar))
26
27 ch = frame(overlay(ch-cross, ch-bg))
28
29 ch
30
31
32 #
33 # The Republic of the Congo
34 #
35
36 cg-bg = rectangle(width, height, "solid", "yellow")
37
38 cg-upper =
     flip-vertical(
39
        right-triangle(height, height, "solid",
40
          "medium sea green"))
41
12
```

Programming as jgordon@vassar.edu.



• • •



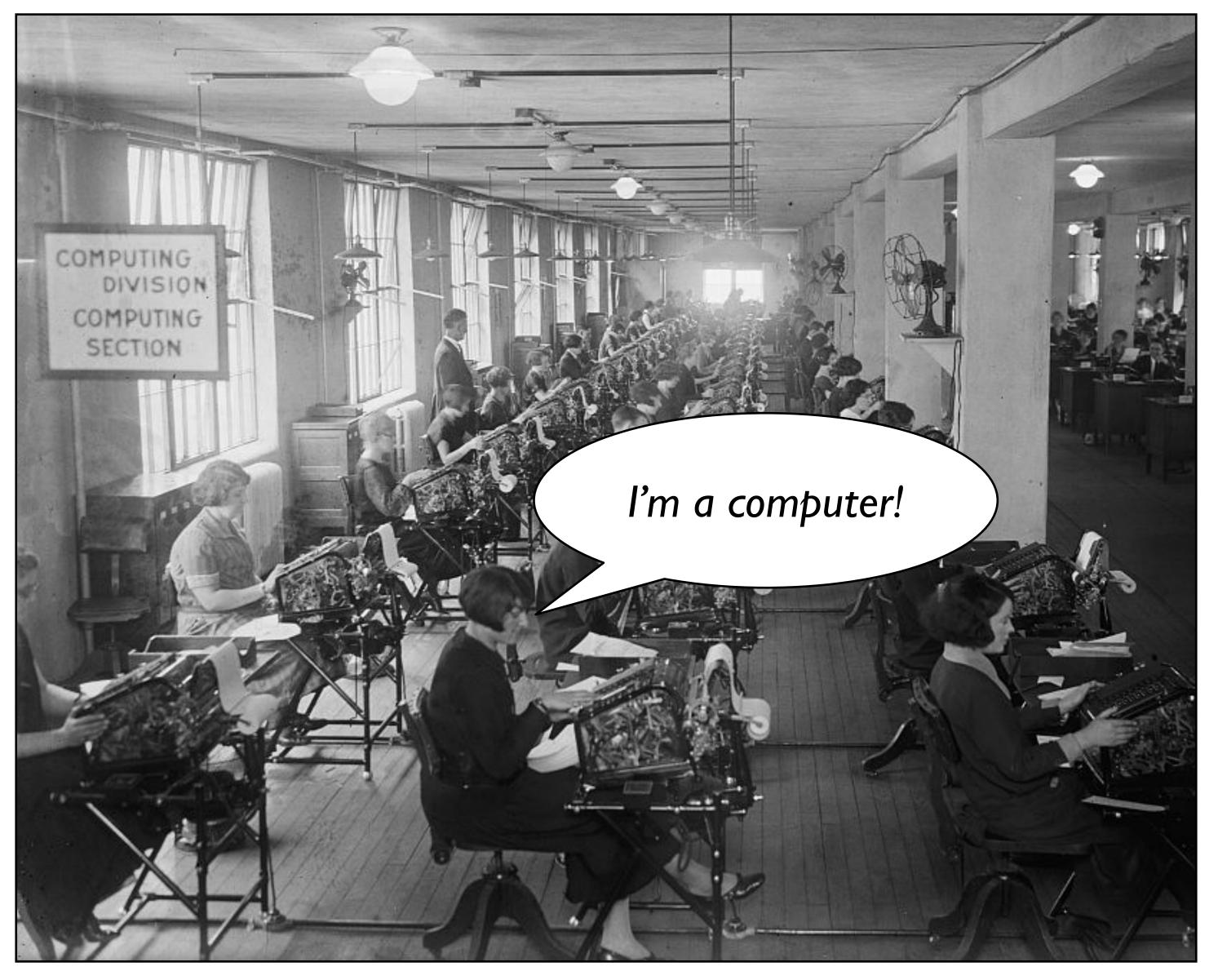
#### We've been practicing *computational thinking*.

"Modern computer science is the last 1 percent of the historical timeline of computational thinking. Computer scientists inherited and then perfected computational thinking from a long line of mathematicians, natural philosophers, scientists, and engineers all interested in performing large calculations and complex inferences without error."

Peter J. Denning & Matti Tedre, Computational Thinking

#### Origins of computational thinking

Before the modern computer age, there was a profession of mathematically trained experts who performed complex calculations as teams. They were called "computers".



Bonus Bureau, Computing Division, 1924, loc.gov/pictures/item/2016838906

Teams of human computers engaged in computational thinking long before the invention of electronic computers.

Early computational thinking can be seen going back to the records of the Babylonians, who wrote down general procedures for solving mathematical problems around starting around 1800 BCE.

Long before this class, you learned these kind of computational methods.

### 27182818284590 +31415926535897

### 27182818284590 +31415926535897

#### +31415926535897

#### 1 1 27182818284590 +31415926535897

#### 1 1 1 27182818284590 +31415926535897 0487

#### 1 1 1 1 27182818284590 +31415926535897

#### 1 1 1 1 27182818284590 +31415926535897

#### 1 1111 27182818284590 +31415926535897

#### 1 1111 27182818284590 +31415926535897 44820487

# $\begin{array}{r} 1 & 1 & 1 & 1 & 1 & 1 \\ 27182818284590 \\ +31415926535897 \\ \hline 744820487 \end{array}$

# $\begin{array}{c} \begin{smallmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 27182818284590 \\ +31415926535897 \\ 58598744820487 \end{array}$

#### Euclid's algorithm

Around 300 BCE, the Greek mathematician Euclid gave a method to find the greatest common divisor (GCD) of two numbers, which is the largest integer that divides both numbers.

#### Euclid's algorithm

Euclid noticed that the GCD of two numbers divides their difference.

So, he repeatedly replaced the larger number with their difference until both were the same.

## Euclid's algorithm

Euclid noticed that the GCD of two numbers divides their difference.

So, he repeatedly replaced the larger number with their difference until both were the same.

gcd(48, 18)  $\rightarrow$  gcd(30, 18)  $\rightarrow$  gcd(12, 18)  $\rightarrow$  gcd(12, 6)  $\rightarrow$  gcd(6, 6)  $\rightarrow$  6

## Sieve of Eratosthenes

This is another famous method dating back to the ancient Greeks, used to find all the prime numbers up to some limit.

We begin with a list of all the integers, from 2 to the limit.

## Sieve of Eratosthenes **2** 3 4 5 <del>6</del> 7 <del>8</del> 9 <del>10</del> 11 <del>12</del> 13 <del>14</del> 15 <del>16</del> 17 <del>18</del>

### We cross out all the multiples of 2.

### Then all the multiples of 3.

### Then all the multiples of 5.

And so on, leaving you with only the primes between 2 and the limit you chose.

## Sieve of Eratosthenes

After each round of elimination, a new prime will be revealed, and the next round crosses out all its multiples.

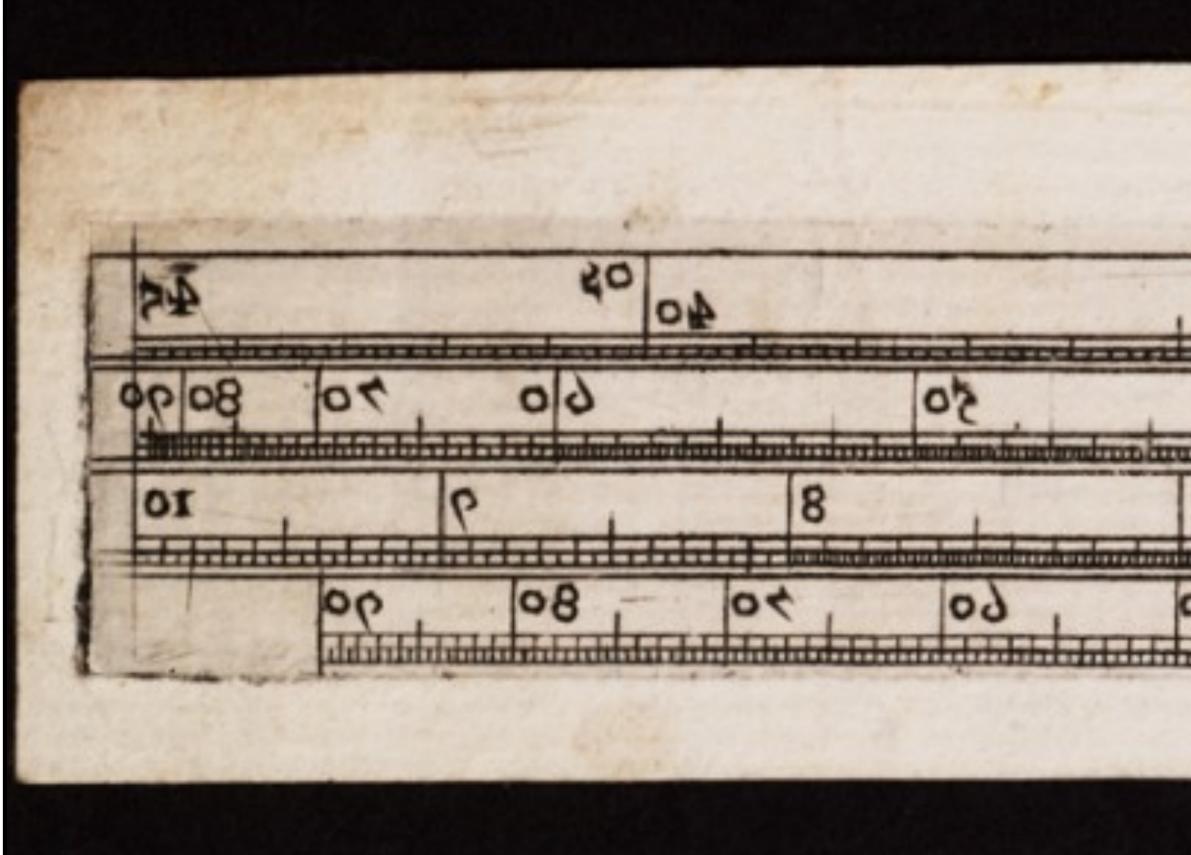
## These are computational procedures, carried out by hand!

# Programmable computers

No matter how simple and unambiguous the steps are made, human computers make mistakes – and lots of them!

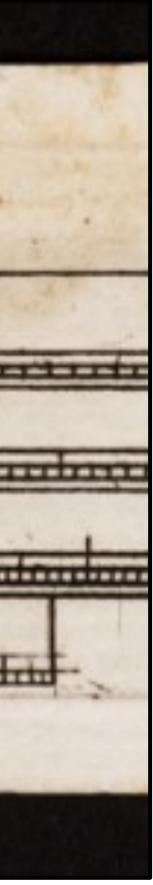
So, inventors through the ages have sought to make computing *machines* to allow people to perform longer computations with fewer errors.

## This was a slow process, taking us from...



-		40 09		
	40			30
in the second second		3		2
dunden	40	02.	20	OI

- Slide rule
  - c. 1620



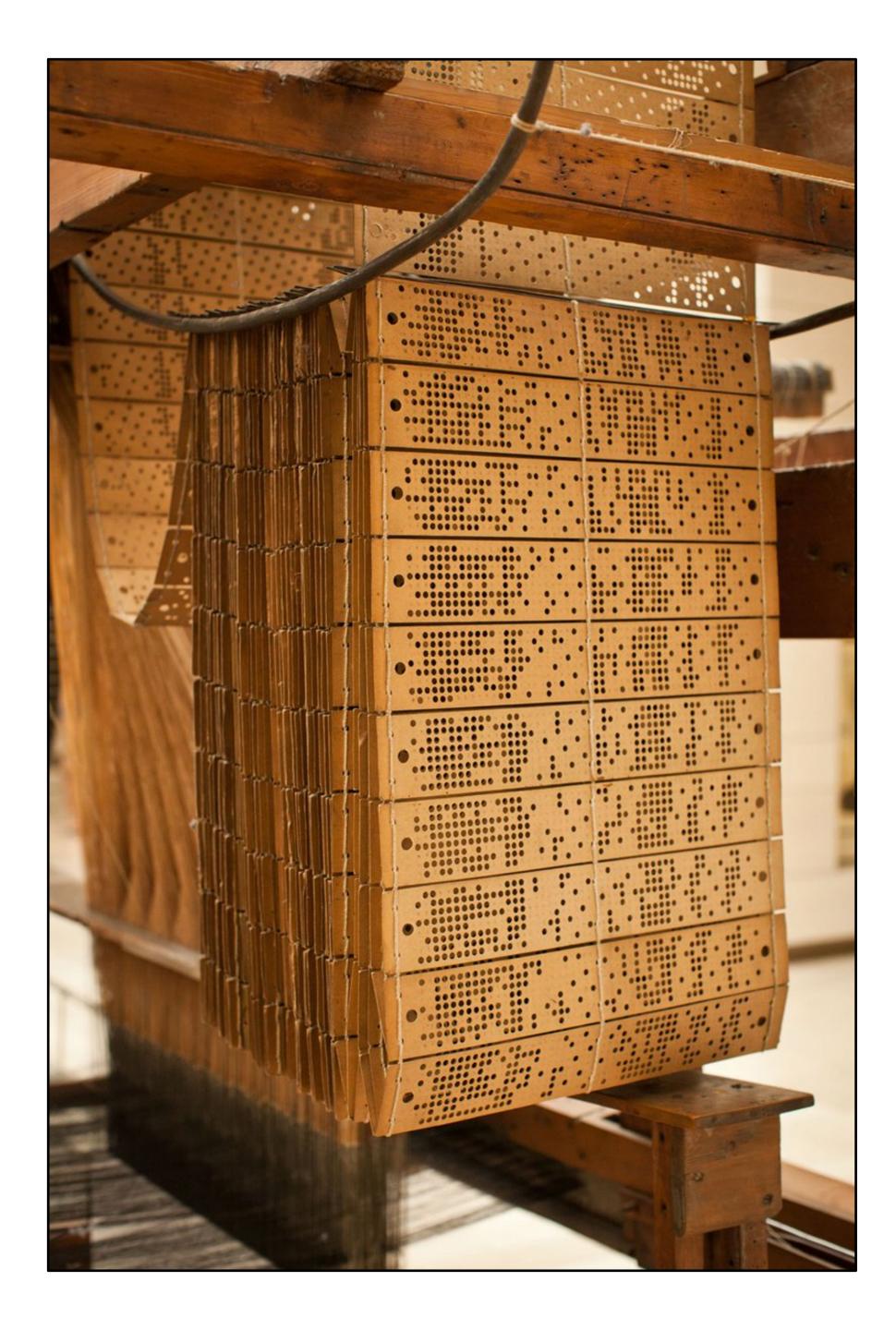


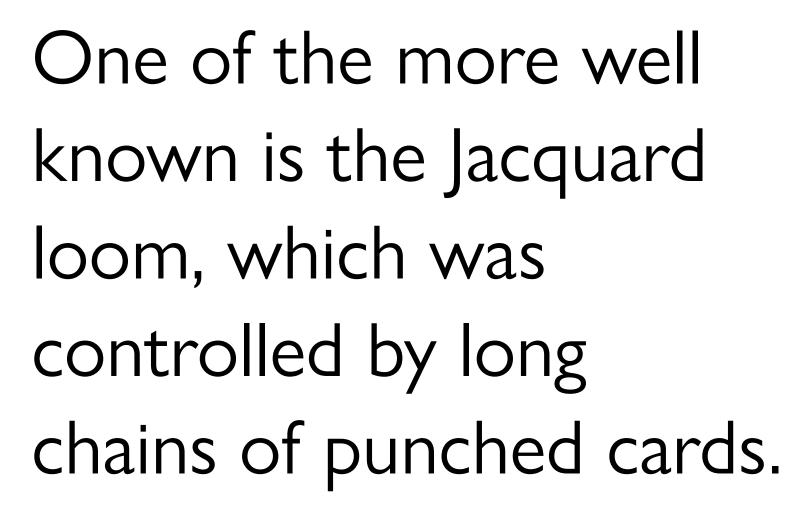
Blaise Pascal's mechanical calculator

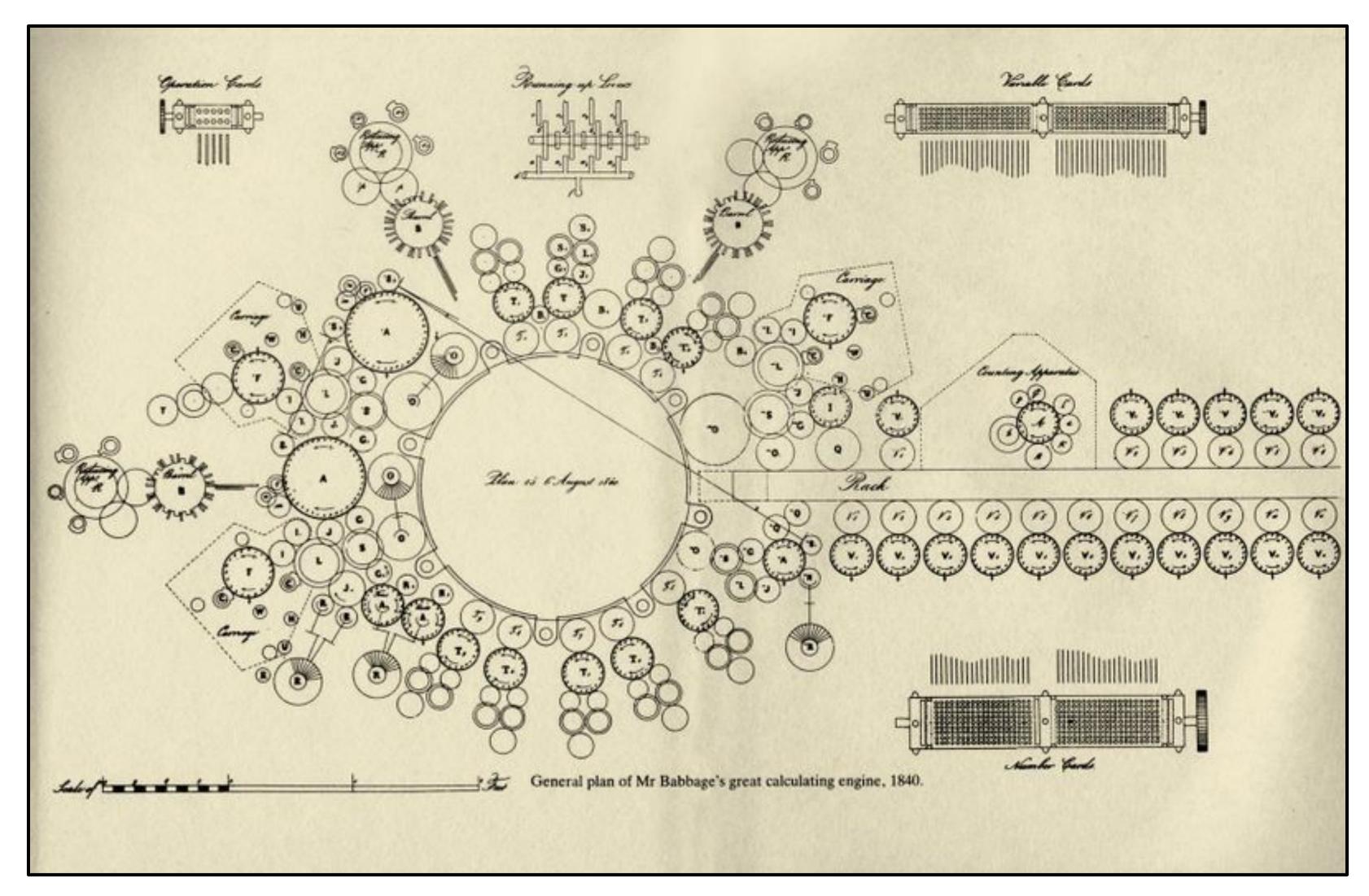
1642

Precursors to the idea of a *programmable* computer originated well before the electronic computing age.

In the early 1700s, French textile weavers experimented with machines that could weave complex patterns using an automatic loom.







### Plan for Babbage's Analytical Engine

1840

Babbage collaborated with a gifted mathematician, Ada Lovelace, who designed algorithms for the Analytical Engine, even though there was no machine to run them on.



Lovelace saw the Analytical Engine not as a mere calculator but as a processor of any information that could be encoded in symbols.

This insight, that computing programs can calculate not only over numbers but over symbols that can stand for anything in the world, anticipated by a hundred years a key tenet of the modern computer age.

Lovelace saw the computer as an *information* machine.

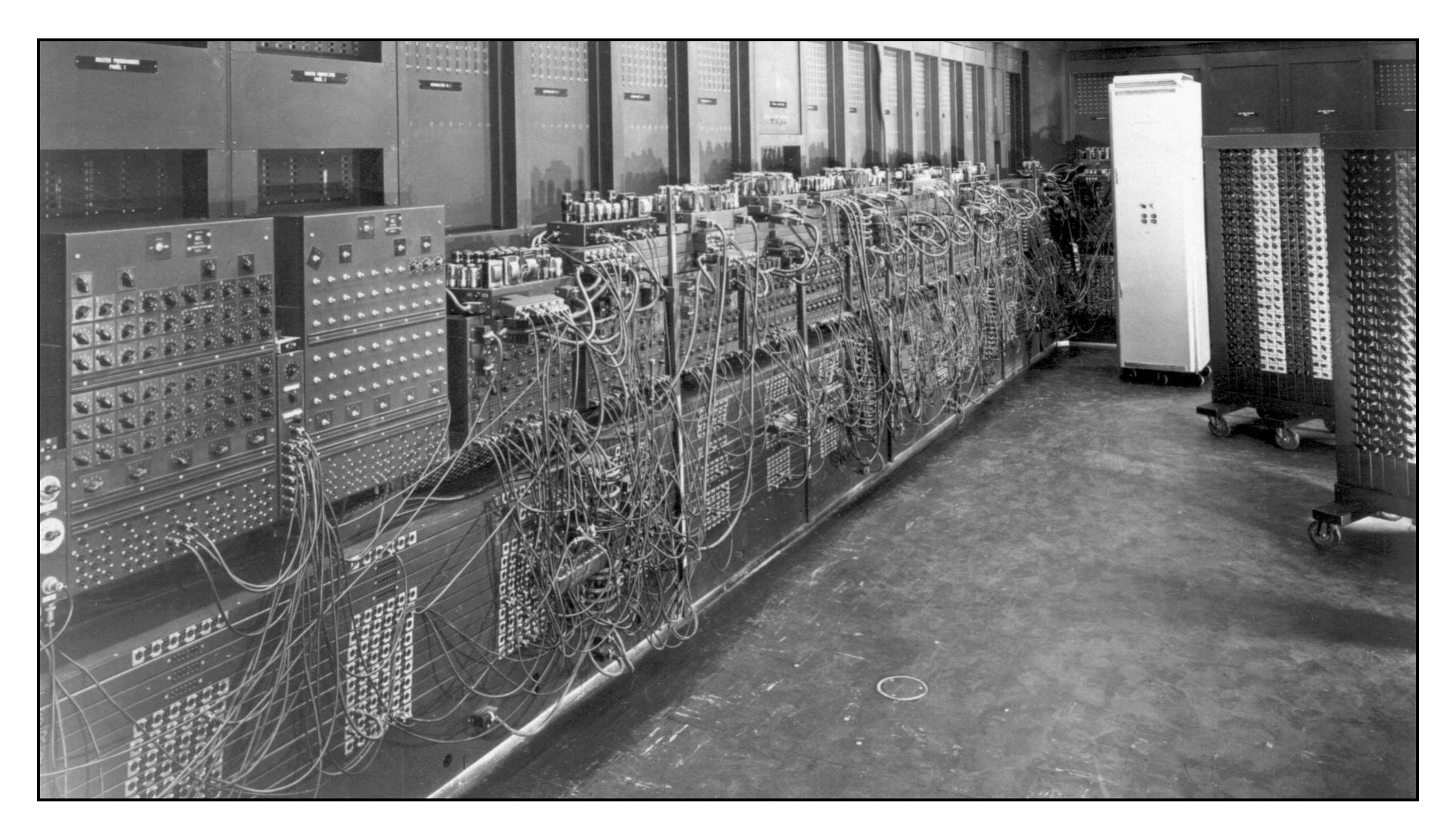
While Babbage's designs for a programmable computer weren't realized at the time, the age of electronics opened new possibilities.





### Harvard Mark I

1944



- ENIAC
- c. 1945

Early computers were very difficult to program, working in languages that were closely tied to the hardware.

Grace Hopper '28 popularized the idea of a compiler for machine-independent programming languages and defined FLOW-MATIC, the first English-like data processing language in the early 1950s.

Those ideas were later folded into the popular COBOL language (1959).



Since the 1950s, many programming languages have been defined, experienced popularity, and then been supplanted by new designs.

Today, Python is the programming language most often used for work in data science and artificial intelligence.

Programming no longer involves plugging in wires or punching cards, but it's still hard!

"The programmer, like the poet, works only slightly removed from pure thought-stuff. He builds castles in the air, from air, creating by exertion of the imagination...

Fred Brooks, The Mythical Man Month, 1975



"Few media of creation are so flexible, so easy to polish and rework, so readily capable of realizing grand conceptual structures. Yet the program construct, unlike the poet's words, is real in the sense that it moves and works, producing visible outputs separate from the construct itself...

Fred Brooks. The Mythical Man Month, 1975



"One types the correct incantation on a keyboard, and a display screen comes to life, showing things that never were nor could be... It prints results, draws pictures, produces sounds, moves arms. The magic of myth and legend has come true in our time...

Fred Brooks, The Mythical Man Month, 1975



"The computer resembles the magic of legend in this respect, too. If one character, one pause, of the incantation is not strictly in proper form, the magic doesn't work. Human beings are not accustomed to being perfect, and few areas of human activity demand it. Adjusting to the requirement for perfection is, I think, the most difficult part of learning to program."

Fred Brooks, The Mythical Man Month, 1975



# Computing with data

We've seen some cool data sets during this semester, but there are many, many more you can explore.

### $\bullet \bullet \bullet < >$

### **Data Is Plural**

... is a weekly newsletter of useful/curious datasets, published by Jeremy Singer-Vine. There have been <u>311 editions</u>, dating from <u>October 21, 2015</u> to <u>November 30</u>, 2022. To receive future editions, sign up here:

Enter your email address

### **Some Nice Things People Have Said**

- "a treasure trove of interesting datasets" Julia Silge

- "consistently fascinating" Robin Sloan
- "the only newsletter I open immediately" Paul Ford
- <u>Melody Joy Kramer</u>
- period" Melanie Walsh
- *"the best data newsletter ever"* François Briatte

### **Recent Editions**

2022.11.30 • Pills, per-pupil spending, travelers' coronavirus variants, Indonesia earthquake intensities, and more roadkill.

2022.11.23 • Presidential pardons, radioactive waste, songs of the world, semiconductor logistics, and "every Star Trek ever."

2022.11.16 • Big emitters, disease outbreaks, permissively-licensed code, impact craters, and tinned fish.

■ www.data-is-plural.com

Subscribe

Ċ

• "delivers exactly what it promises, it's delightful" - Simon Willison

• "required reading for anyone interested in data journalism" — Julia Angwin

• "the newsletter I forward most frequently to colleagues who like datasets"

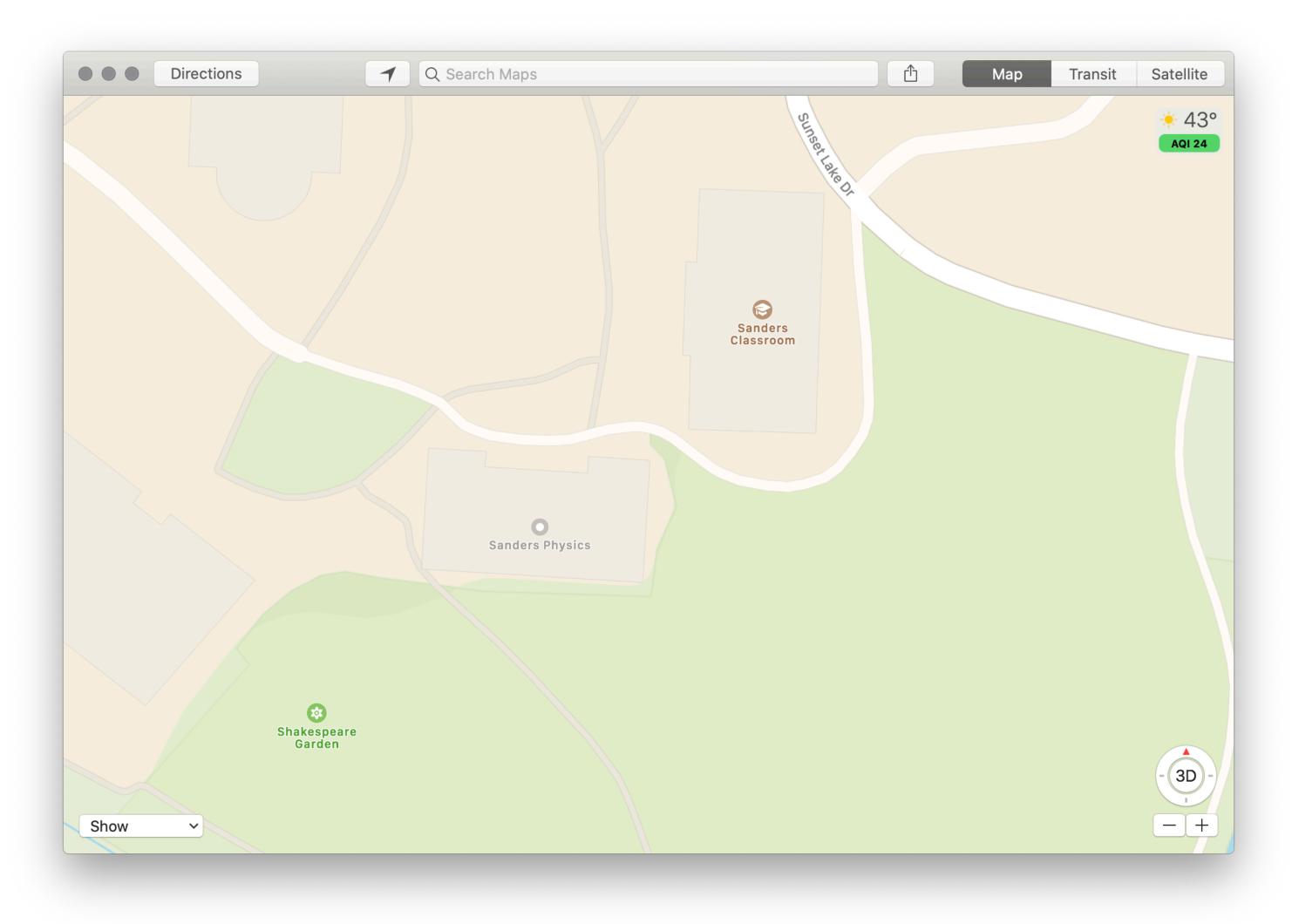
• "definitely my favorite newsletter about data and possibly my favorite newsletter

# Building data models

## Data is situated in the environment where it was gathered.

### Consider Galton's child-height data.

He gathered the data in England c. 1886. What would happen if you tried to use it to predict heights in Poughkeepsie today? In Guatemala? In China?



When we collect data, it's like making a map:

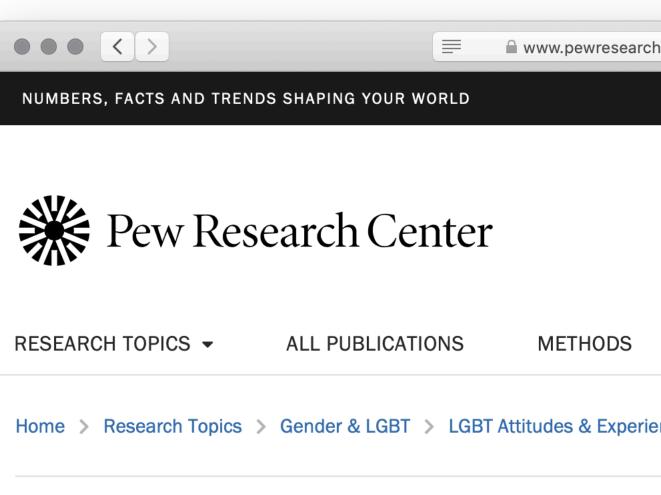
We're constructing a model, where we choose what to represent, and how to represent it.

"...most of the data and data models we have inherited deal with structures of power, like gender and race, with a crudeness that would never pass muster in a peer-reviewed humanities publication." Miriam Posner, "What's Next: The Radical, Unrealized Potential of

Digital Humanities", 2016

• <				e co	lab.research.g	oogle
+ Code	e + Text					
[]	family	father	mother	children	childNum	gei
	1	78.5	67	4	1	
	1	78.5	67	4	2	fe
	1	78.5	67	4	3	fe
	1	78.5	67	4	4	fe
	2	75.5	66.5	4	1	
	2	75.5	66.5	4	2	
	(928 rov	ws omitted	)			
we c	an see th	is differe	nce in the	se distribut	ions by look	ing
We c					ions by look er"], uni	
_	galton					
0						
0	galton -					
0	galton -					
0	galton -					
0	galton .					
0	galton 17.5 – 15 – 15 – 10 – 7.5 –					
	[ ] We s	[ ] family	family       father         family       father         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         1       78.5         2       75.5          (928 rows omitted)	family       father       mother         1       78.5       67         1       78.5       67         1       78.5       67         1       78.5       67         1       78.5       67         1       78.5       67         1       78.5       67         1       78.5       665         1       75.5       66.5         1       928 rows omitted)       66.5	family       father       mother       children         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       67       4         1       78.5       66.5       4         2       75.5       66.5       4         (928 rowsted)	$\begin{bmatrix} 1 \\ \mathbf{family} \end{bmatrix} \mathbf{father} \mathbf{hother} \mathbf{children} \mathbf{childNum}$ $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \mathbf{family} \mathbf{father} \end{bmatrix} \mathbf{father} father$

		Connect 👻 🚢 🏟	~	
der	childHeight			
nale	73.2			
nale	69.2			
nale	69			
nale	69			
nale	73.5			
nale	72.5			
	•			
hoc		↑↓	:	h
hes.	istogram of mo		:	5
	stogram of mo	↑ ↓ ⊕ ■ ➤ □ ■ other and father heights together:	:	
it a hi			:	
			•	
it a hi				
it a hi	")		•	
it a hi	")		•	
it a hi	")		•	
it a hi	")		•	
it a hi	")		•	
it a hi	")			
it a hi	")			



DECEMBER 18, 2019

### About four-in-ten U.S. adults forms should offer more than gender options

### **BY NIKKI GRAF**

In 2019, at least seven states have started offering a <u>third gender</u> option on driver's licenses for people who don't identify as male or female, and at least four more plan to do so in

### States that offer a third gender option on driver's licenses



reh era/fact tank/2010/12/12/20	ender-options		
rch.org/fact-tank/2019/12/18/ge			
	NEWSLETTER	RS   DONATE   MY	ACCOUNT
	See our research on: Climate Chan	ge   Social Media	COVID-19
	Search pewresearch.org	ğ	Q
SHORT READS	TOOLS & RESOURCES	EXPERTS	ABOUT
ences > Gender Identity			
	RELATED		
<b>G Y 🛅 </b>	REPORT   SEP 5, 2019		
say	About one-in-five U.S. a		
n two	someone who goes by a neutral pronoun	gender-	
	REPORT   JUL 27, 2021	ultaknow	
	Rising shares of U.S. ad someone who is transge		
on	by gender-neutral pron	•	
	REPORT   DEC 28, 2017		
ble in 2020 or later	10 things we learned ab	-	
NH	issues in the U.S. in 201	7	
VT MA	REPORT   DEC 14, 2017		
PA	Gender discrimination	comes in	

"I want us to be more ambitious, to hold ourselves to much higher standards when we are claiming to develop data-based work that depicts people's lives." Miriam Posner, "What's Next: The Radical, Unrealized Potential of Digital Humanities", 2016

# Data and privacy

Sentime	ent140 Search this site				
eneral Information ite Functionality or Academics PI	General Information				
Sentiment Analysis Sites Contact Us	What is Sentiment140?				
Return to Sentiment140	Sentiment140 allows you to discover the sentiment of a brand, product, or topic on Twitter.				
	How does this work?				
	You can read about our approach in our technical report: <u>Twitter Sentiment Classification using Distant</u> <u>Supervision</u> . There are also additional features that are not described in this paper.				
	How is this different?				
	Our approach is different from other sentiment analysis sites because:				
	<ul> <li>We use classifiers built from <u>machine learning</u> algorithms. Other products use a simpler keyword-based approach which may have higher precision but lower recall.</li> <li>We provide transparency for the classification results of individual tweets. Other sites only surface aggregated metrics, which makes it difficult to assess the accuracy of their classifiers.</li> </ul>				
	Who created this?				
	Sentiment140 was created by Alec Go,				

## help.sentiment140.com

~	150%	1
	10070	

i≡ ~

🔲 testdata.r

1 × 1	
-------	--

+		She	eet 1			
0	Α	В	С	D	E	
37	0	122	Sat May 16 22:42:25 UTC 2009	itchy	robloposky	<b> </b>
38	0	123	Sat May 16 22:42:44 UTC 2009	itchy	EdwinLValencia	(
39	4	124	Sat May 16 23:48:15 UTC 2009	stanford	imusicmash	F <u>k</u>
40	4	125	Sat May 16 23:58:34 UTC 2009	lyx	drewloewe	@ F
41	4	131	Sun May 17 15:05:03 UTC 2009	Danny Gokey	VickyTigger	l' r
42	4	132	Sun May 17 17:27:45 UTC 2009	sleep	babblyabbie	i
43	0	133	Sun May 17 17:27:49 UTC 2009	sleep	kisjoaquin	С
44	0	134	Sun May 17 17:28:02 UTC 2009	sleep	Whacktackular	E
45	0	135	Sun May 17 17:29:50 UTC 2009	san francisco	Adrigonzo	ç
46	2	136	Sun May 17 17:30:19 UTC 2009	san francisco	sulu34	i
47	2	137	Sun May 17 17:30:23 UTC 2009	san francisco	schuyler	j
48	2	138	Sun May 17 17:30:56 UTC 2009	san francisco	MattBragoni	S
19	0	139	Sun May 17 17:32:00 UTC 2009	aig	KennyTRoland	?
50	0	140	Sun May 17 17:32:30 UTC 2009	aig	aMild	s t
51	0	141	Sun May 17 17:32:36 UTC 2009	aig	Trazor1	S
52	4	142	Sun May 17 17:35:17 UTC 2009	star trek	mimknits	(
53	2	143	Sun May 17 17:35:28 UTC 2009	star trek	GeeRen	(
54	2	144	Sun May 17 17:35:45 UTC 2009	star trek	checkyesjess	(
55	0	145	Mon May 18 01:13:27 UTC 2009	Malcolm Gladwell	renano	a
55 SUM		145 22	AVERAGE 61	MIN 0	MAX 122	

manual	2009	.06.14	— Edited	
nanuai	.2005	.00.14	LUILUU	

Τ 📕 🖾 🗸 📕

**Qt** 

×1

Ŧ

### I'm itchy and miserable!

@sekseemess no. I'm not itchy for now. Maybe later, lol.

RT @jessverr I love the nerdy Stanford human biology videos - makes me miss school. <u>http://bit.ly/13t7NR</u>

F

@spinuzzi: Has been a bit crazy, with steep learning curve, but LyX is really good for long docs. For anything shorter, it would be insane.

I'm listening to "P.Y.T" by Danny Gokey <3 &lt;3 &lt;3 Aww, he's so amazing. I &lt;3 him so much :)

is going to sleep then on a bike ride:]

cant sleep... my tooth is aching.

Blah, blah, blah same old same old. No plans today, going back to sleep I guess.

glad i didnt do Bay to Breakers today, it's 1000 freaking degrees in San Francisco wtf

is in San Francisco at Bay to Breakers.

just landed at San Francisco

San Francisco today. Any suggestions?

?Obama Administration Must Stop Bonuses to AIG Ponzi Schemers ... http://bit.ly/2CUIg

started to think that Citi is in really deep s&^t. Are they gonna survive the turmoil or are they gonna be the next AIG?

ShaunWoo hate'n on AiG

@YarnThing you will not regret going to see Star Trek. It was AWESOME!

On my way to see Star Trek @ The Esquire.

Going to see star trek soon with my dad.

annoying new trend on the internets: people picking apart michael lewis and malcolm gladwell.

"The words and phrases we search for on Google, the times of day we are most active on Facebook, and the number of items we add to our Amazon carts are all tracked and stored as data – data that are then converted into corporate financial gain." D'Ignazio & Klein, *Data Feminism*, 2020



Antonio Bolfo/Reportage for The New York Times

By Charles Duhigg

Feb. 16, 2012

Andrew Pole had just started working as a statistician for Target in 2002, when two colleagues from the marketing department



"As Pole's computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a 'pregnancy prediction' score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

"One Target employee I spoke to provided a hypothetical example. Take a fictional Target shopper named Jenny Ward, who is 23, lives in Atlanta and in March bought cocoa-butter lotion, a purse large enough to double as a diaper bag, zinc and magnesium supplements and a bright blue rug. There's, say, an 87 percent chance that she's pregnant and that her delivery date is sometime in late August."

Kashmir Hill, "How Target Figured Out a Teen Girl was Pregnant Before Her Father Did", *Forbes*, 2012

Computing with data is complex, and it's not just "technical" issues we need to concern ourselves with!

Computing with data gives us a lot of power! We can do a lot of harm, but we can also do a lot of good!

Computer Science I —or, where do you go from here?

Data types Naming values Evaluation Conditionals (if and cases) Function signatures Testing functions Tabular data Sanitizing real-world data Visualization

Higher-order functions Linked lists data

Python lists and arrays

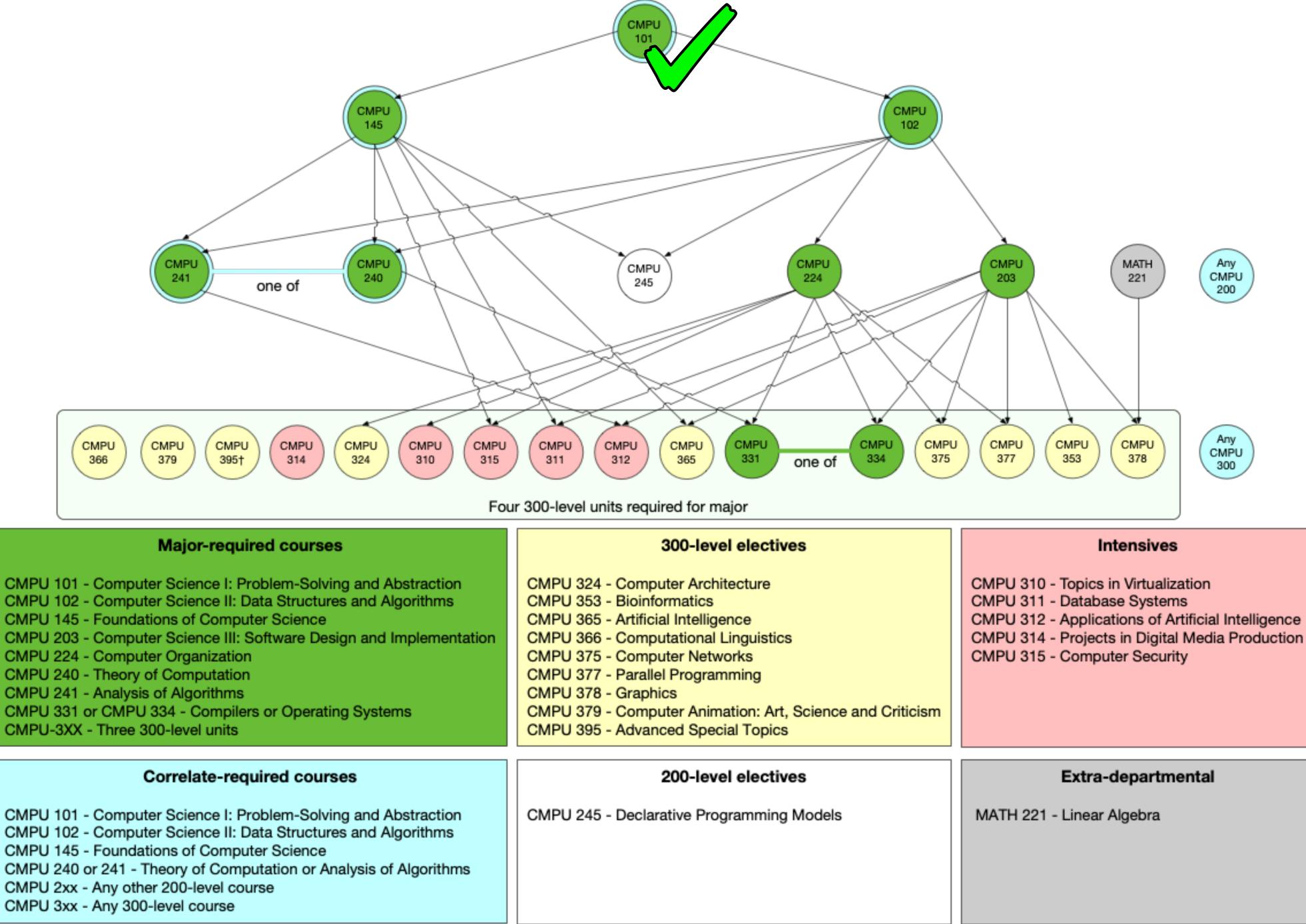
Iteration (for)

- Lambda expressions
- Defining structured
- Structurally recursive data and functions
- Trees (binary, *n*-ary)

- Side effects and mutation Debugging using print statements Functional vs imperative languages
- Accumulators
- Memory and aliasing
- Hash tables (dictionaries)
- Web APIs and JSON
- Numpy arrays

## Congratulations on making it this far!

# CS courses at Vassar



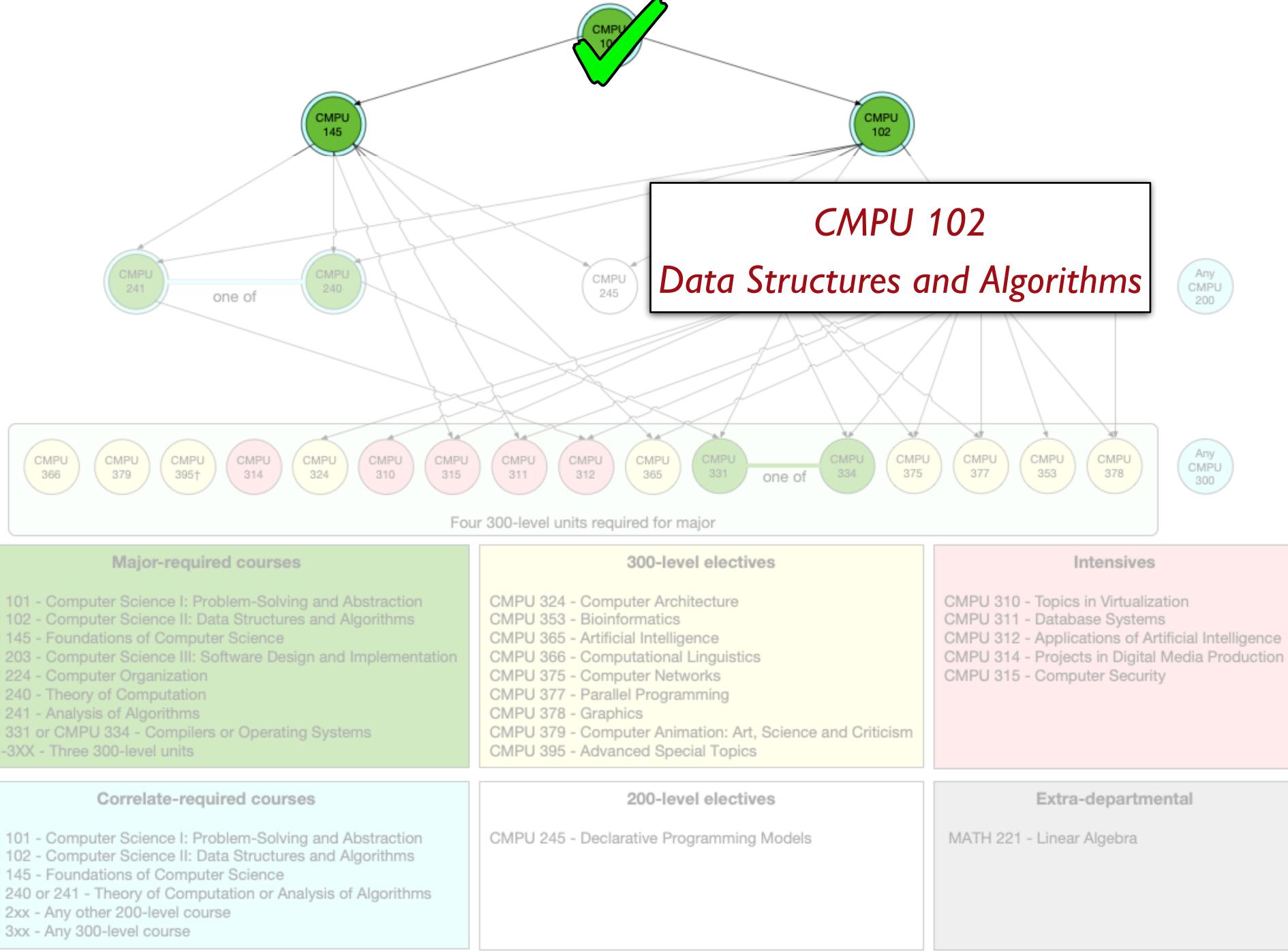
CMPU 101 - Computer Science I: Problem-Solving and Abstraction
CMPU 102 - Computer Science II: Data Structures and Algorithms
CMPU 145 - Foundations of Computer Science
CMPU 203 - Computer Science III: Software Design and Implementation
CMPU 224 - Computer Organization

CMPU 240 - Theory of Computation

CMPU-3XX - Three 300-level units

CMPU 101 - Computer Science I: Problem-Solving and Abstraction CMPU 102 - Computer Science II: Data Structures and Algorithms CMPU 145 - Foundations of Computer Science CMPU 240 or 241 - Theory of Computation or Analysis of Algorithms CMPU 2xx - Any other 200-level course CMPU 3xx - Any 300-level course

\*At least two CMPU-200 level courses required for every CMPU-300 level course. †Prerequisites vary depending on topic.



CMPU 101 ·	<ul> <li>Computer</li> </ul>	Science I: Problem-Solving and Abstracti	on
CMPU 102 ·	<ul> <li>Computer</li> </ul>	r Science II: Data Structures and Algorithm	S

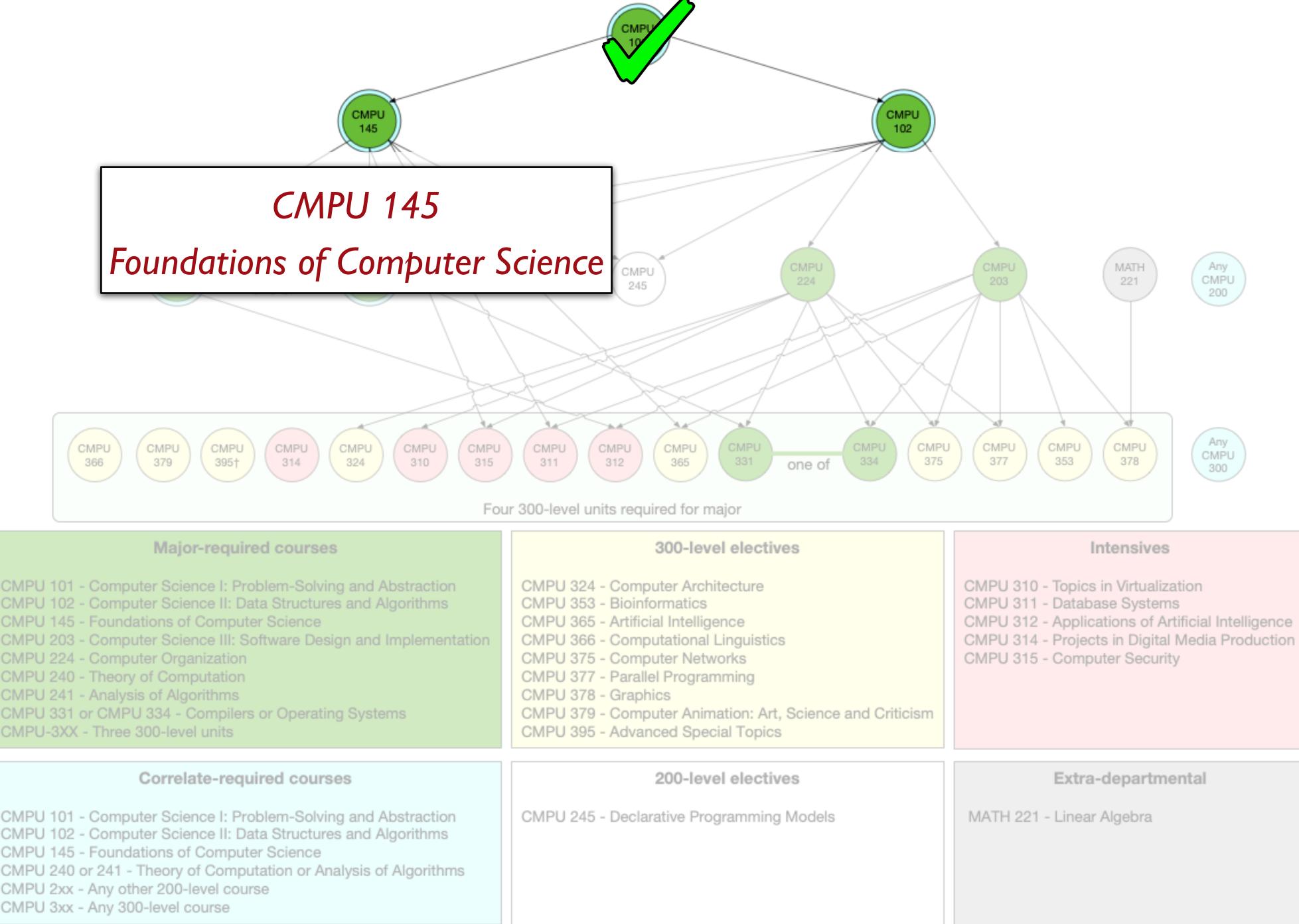
- CMPU 145 Foundations of Computer Science
- CMPU 203 Computer Science III: Software Design and Implementation
- CMPU 224 Computer Organization
- CMPU 240 Theory of Computation
- CMPU 241 Analysis of Algorithms
- CMPU 331 or CMPU 334 Compilers or Operating Systems
- CMPU-3XX Three 300-level units

CMPU 101 - Computer Science I: Problem-Solving and Abstraction CMPU 102 - Computer Science II: Data Structures and Algorithms CMPU 145 - Foundations of Computer Science CMPU 240 or 241 - Theory of Computation or Analysis of Algorithms CMPU 2xx - Any other 200-level course CMPU 3xx - Any 300-level course

\*At least two CMPU-200 level courses required for every CMPU-300 level course. †Prerequisites vary depending on topic.

### CMPU 145

### **CMPU 145**



- CMPU 101 Computer Science I: Problem-Solving and Abstraction
- CMPU 145 Foundations of Computer Science
- CMPU 203 Computer Science III: Software Design and Implementation
- CMPU 224 Computer Organization
- CMPU 240 Theory of Computation
- CMPU 241 Analysis of Algorithms
- CMPU 331 or CMPU 334 Compilers or Operating Systems
- CMPU-3XX Three 300-level units

CMPU 101 - Computer Science I: Problem-Solving and Abstraction CMPU 102 - Computer Science II: Data Structures and Algorithms CMPU 145 - Foundations of Computer Science CMPU 240 or 241 - Theory of Computation or Analysis of Algorithms CMPU 2xx - Any other 200-level course CMPU 3xx - Any 300-level course

\*At least two CMPU-200 level courses required for every CMPU-300 level course. †Prerequisites vary depending on topic.

## Try them out!

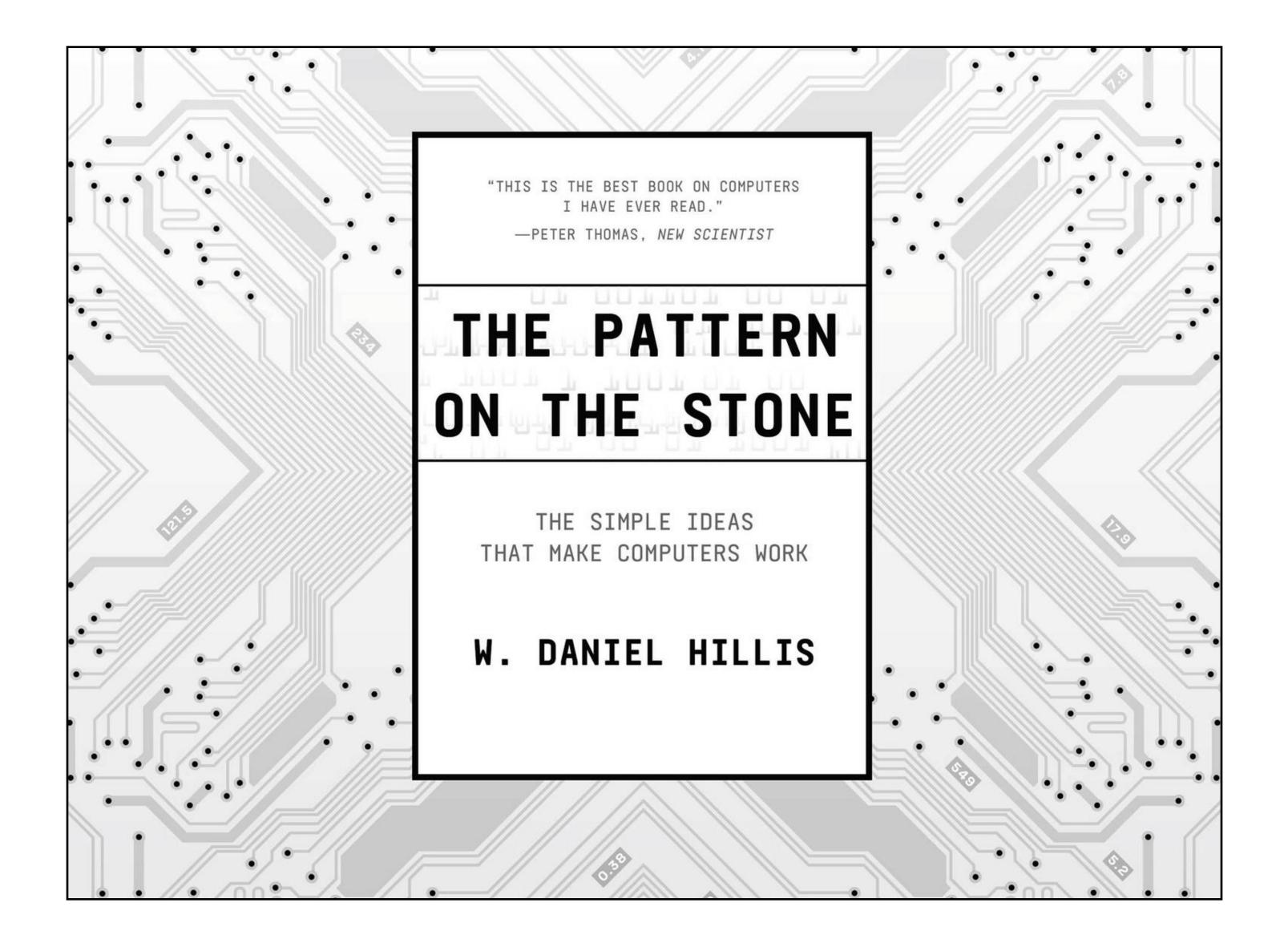
If you keep going with the CS major sequence, you work your way up to some really exciting courses, including...

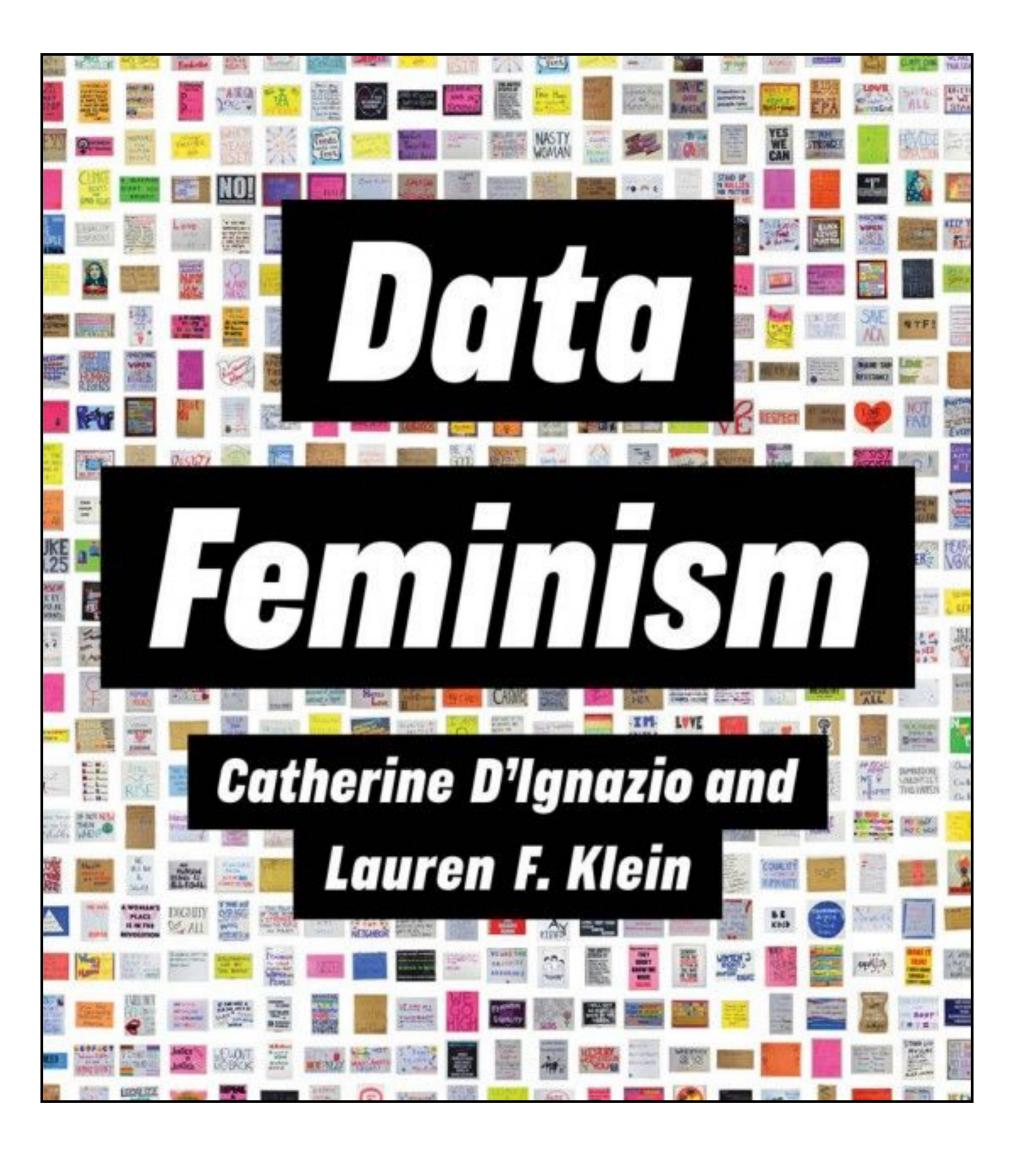
CMPU 353 Bioinformatics

CMPU 377 Parallel Programming

And, you know, probably some cool courses I don't teach as well!

# Further reading





## data-feminism.mitpress.mit.edu

# That's it!

On Wednesday, we'll have a review for Exam 3 where we'll work through practice problems and answer your questions.

go.vassar.edu/course/evals

## Acknowledgments

This lecture incorporates material from: Peter J. Denning & Matti Tedre, Computational Thinking Miram Posner, UCLA Edward Tufte, Visual Explanations Jonathan Gordon, Vassar College