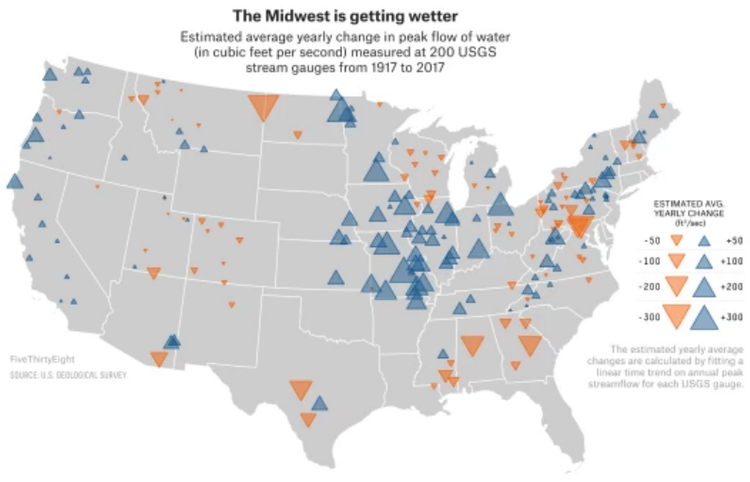
[301] Introduction

Tyler Caraza-Harter

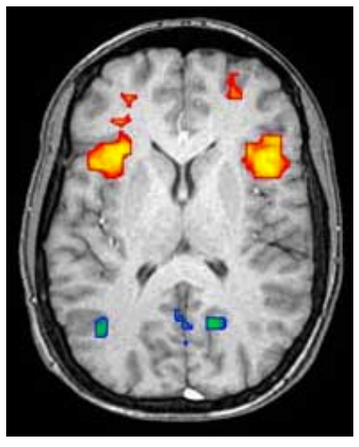
Welcome to Data Programming I!

Data is exploding in many fields

- Journalism
- Biology, physics, chemistry
- Psychology, sociology, economics, business
- Engineering (mechanical, electrical, industrial, etc)

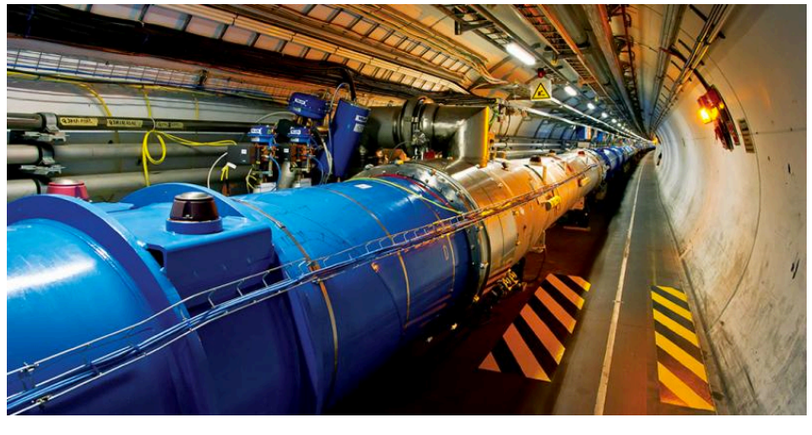


https://fivethirtyeight.com/features/the-midwest-is-getting-drenched-and-its-causing-big-problems/

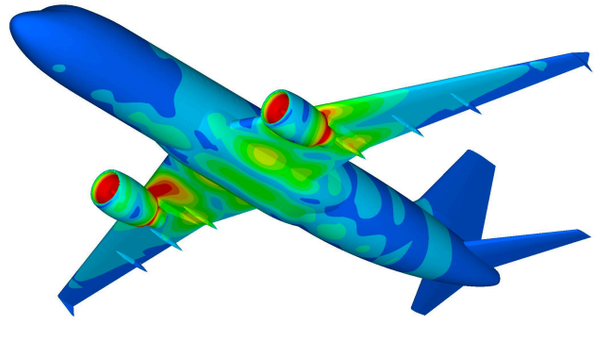


https://en.wikipedia.org/wiki/Neuroimaging

https://science.howstuffworks.com/life/genetic/gattaca-gaptacaz-adding-letters-the-genetic-alphabet.htm



https://home.cern/topics/large-hadron-collider



http://www.stressebook.com/finite-element-analysis-in-a-nut-shell/

Welcome to Data Programming I!

Data is exploding in many fields

- Journalism
- Biology, physics, chemistry
- Psychology, sociology, economics, business
- Engineering (mechanical, electrical, industrial, etc)

How can we gain insights from that data?

Welcome to Data Programming I!

Data is exploding in many fields

- Journalism
- Biology, physics, chemistry
- Psychology, sociology, economics, business
- Engineering (mechanical, electrical, industrial, etc)

How can we gain insights from that data?

- With computation

Welcome to Data Programming I!

Data is exploding in many fields

- Journalism
- Biology, physics, chemistry
- Psychology, sociology, economics, business
- Engineering (mechanical, electrical, industrial, etc)

How can we gain insights from that data?

- With computation

Approach 1: human computation

https://en.wikipedia.org/wiki/Human_computer

Welcome to Data Programming I!

Data is exploding in many fields

- Journalism
- Biology, physics, chemistry
- Psychology, sociology, economics, business
- Engineering (mechanical, electrical, industrial, etc)

How can we gain insights from that data?

- With computation

Approach 1: human computation

https://en.wikipedia.org/wiki/Human_computer

Approach 2: machine computation

http://fortune.com/2015/11/15/intel-super-7/

Welcome to Data Programming I!

CS 301 is about approach 2

- Faster, more reliable, can churn through more data

Approach 1: human computation

https://en.wikipedia.org/wiki/Human_computer

Approach 2: machine computation

http://fortune.com/2015/11/15/intel-super-7/

Welcome to Data Programming I!

CS 301 is about approach 2

- Faster, more reliable, can churn through more data
- Automate to save human effort

"Find the leverage in the world, so you can be more lazy!"

~ Larry Page

Approach 1: human computation

https://en.wikipedia.org/wiki/Human_computer

Approach 2: machine computation

http://fortune.com/2015/11/15/intel-super-7/

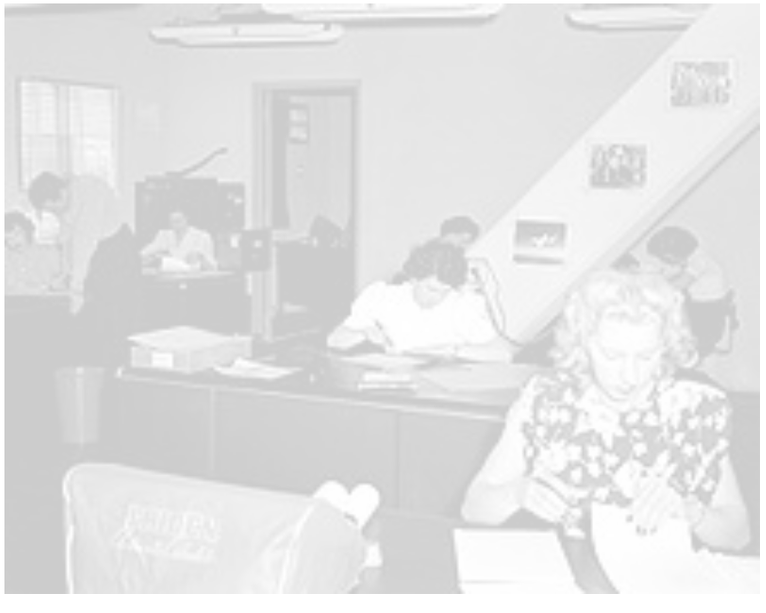
Welcome to Data Programming I!

CS 301 is about approach 2

- Faster, more reliable, can churn through more data
- Automate to save human effort
- Requires being able to tell computers what to do!

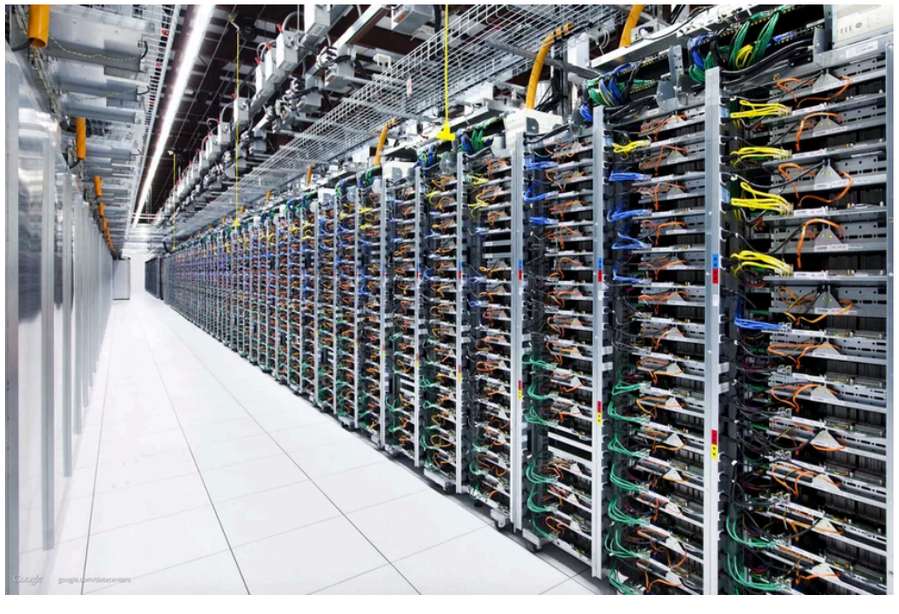
society needs more domain experts in specific fields who can write code

Approach 1: human computation



https://en.wikipedia.org/wiki/Human_computer

Approach 2: machine computation

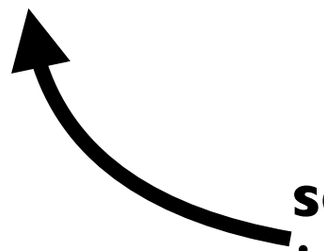


http://fortune.com/2015/11/15/intel-super-7/

Welcome to Data Programming I!

CS 301 is about approach 2

- Faster, more reliable, can churn through more data
- Automate to save human effort
- Requires being able to tell computers what to do!



society needs more domain experts in specific fields who can write code

Goal: become "bilingual"

- Speak the language of X (biology, mech eng, journalism, etc)
- Speak the language of computing

Why CS 301?

Typical intro CS

- Challenging language (e.g., C++ or Java)
- CS students and other majors together
- Heavy on theory, light on data

vs

CS 301 approach

- Python (powerful but easier to learn)
- Bring more coding into other fields
- Light on theory, heavy on data
- Emphasize questions and communication

50 Best Jobs in America for 2019

	Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1	Data Scientist	$108,000	4.3/5	6,510	View Jobs
#2	Nursing Manager	$83,000	4/5	13,931	View Jobs
#3	Marketing Manager	$82,000	4.2/5	7,395	View Jobs
#4	Occupational Therapist	$74,000	4/5	17,701	View Jobs

https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

Today's Topics

Introductions

- Who am I? Who are you?

Course overview

Computer hardware basics

Website

Who am I?

Tyler Caraza-Harter

- Long time Badger
- Email: tylerharter@gmail.com
- Just call me "Tyler"

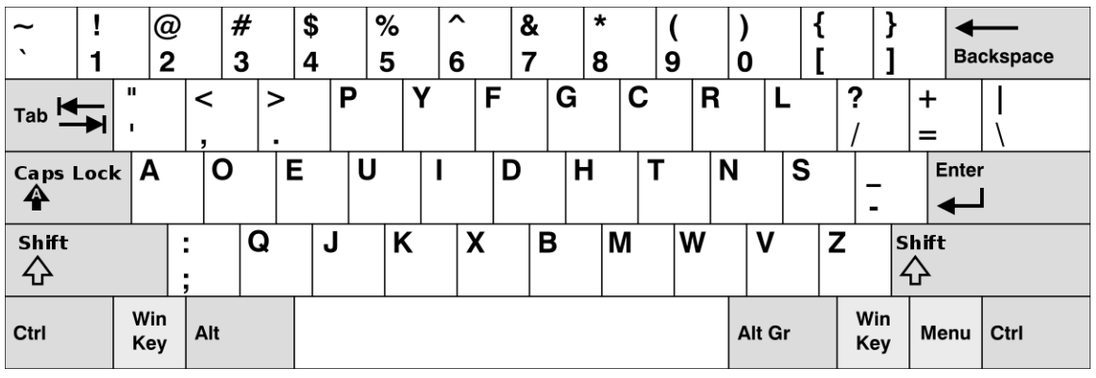
Industry experience

- Worked at Microsoft on SQL Server and Cloud
- Other internships/collaborations:
  Qualcomm, Google, Facebook, Tintri

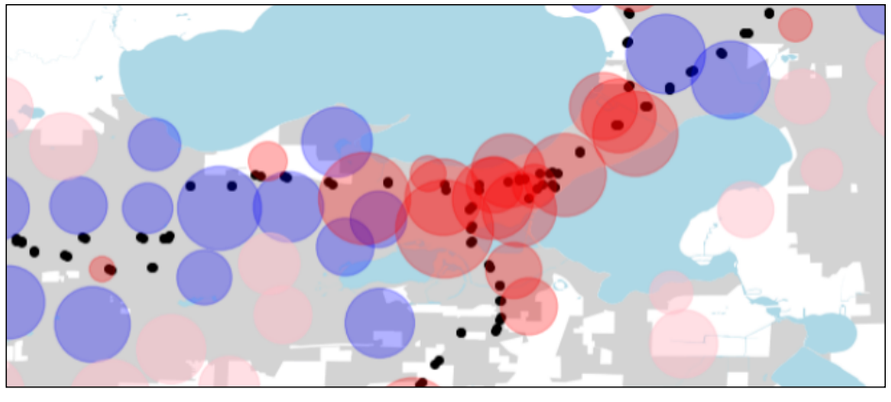
interests

typing efficiently

civic "hacking"



https://en.wikipedia.org/wiki/Dvorak_Simplified_Keyboard



Plot by Jin Woo Lee (previous CS 301 student)

More: https://wisc-ds-projects.github.io/f19/

Who are You?

Year in school?

- 1st year? 2nd? Junior/senior? Grad student?

Area of study

- Natural science, social science, engineering, other?

How many have programmed before?

- Any language? Python? Taken a class?

Survey (counts for participation)

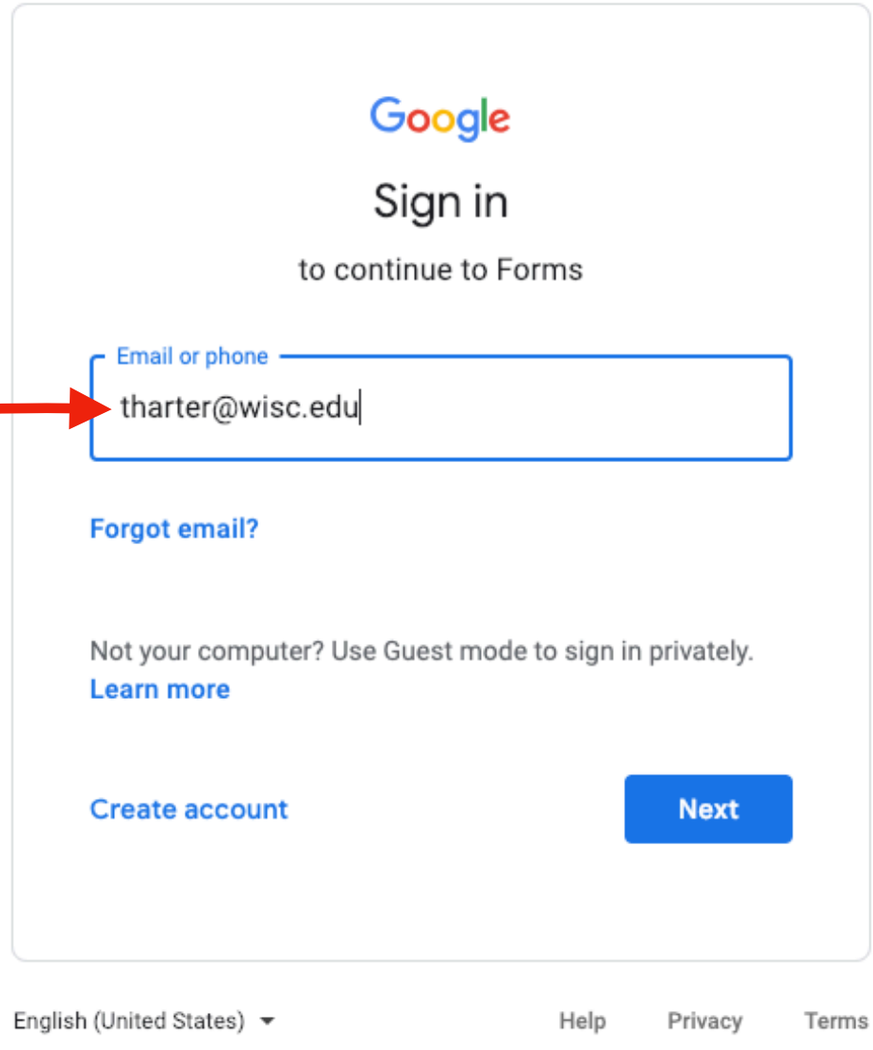
Please help us get to know you (not anonymous):

https://forms.gle/eobVbkoHzKWjTjoYA

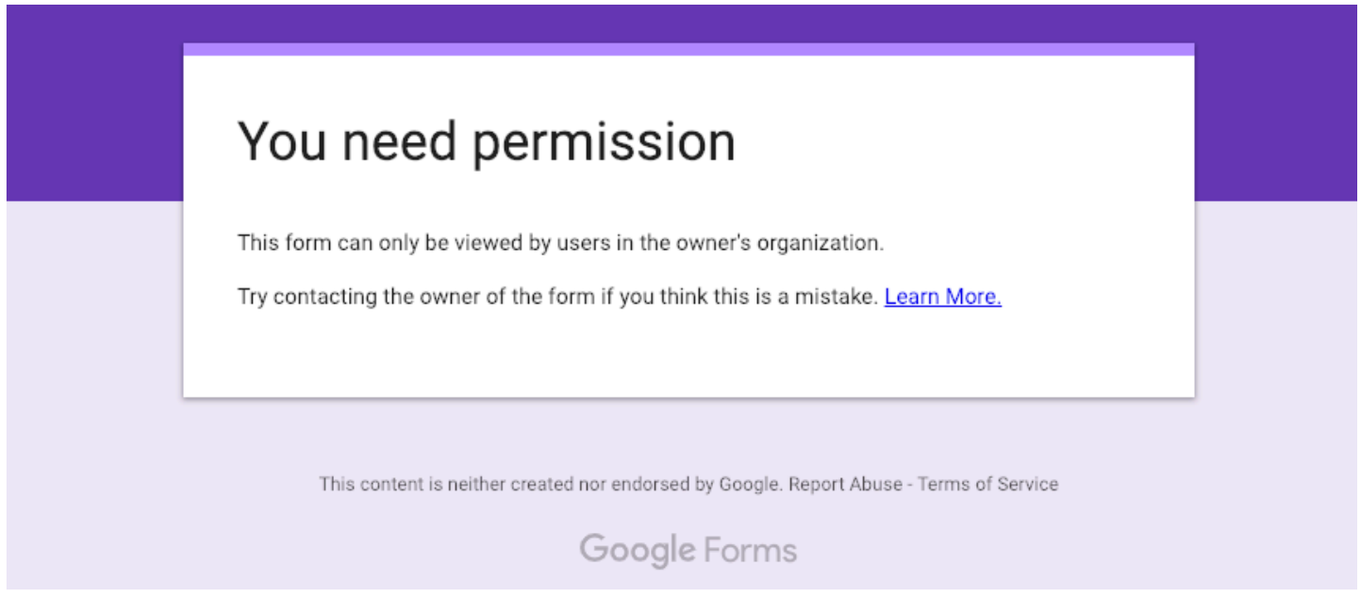
Purposes:

- gauge class interest/experience
- determine who on **waitlist** is attending (please finish by 4pm today!)
- correlate experience with later scores

be sure to use your campus email!

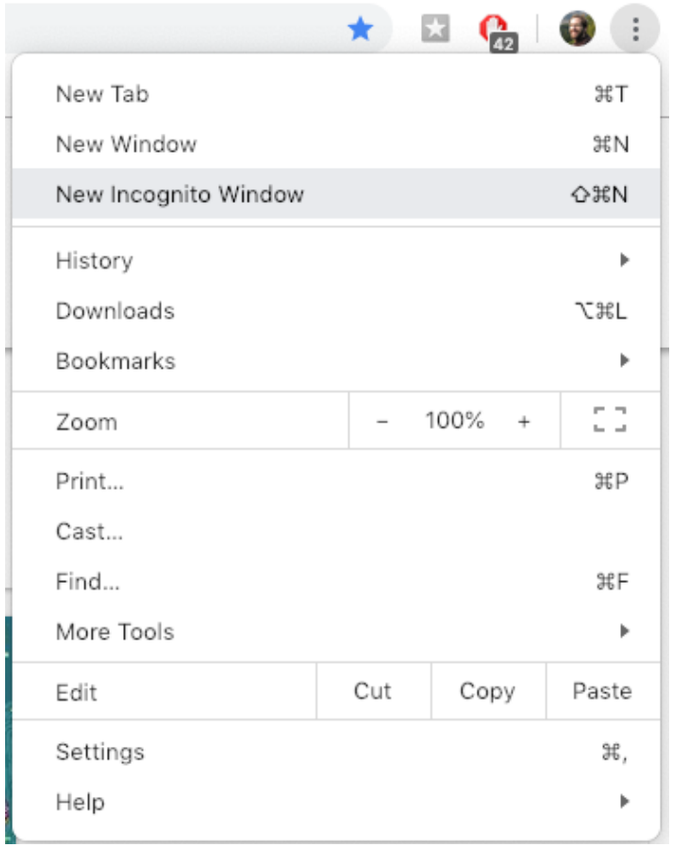


Survey: Common Technical Issues

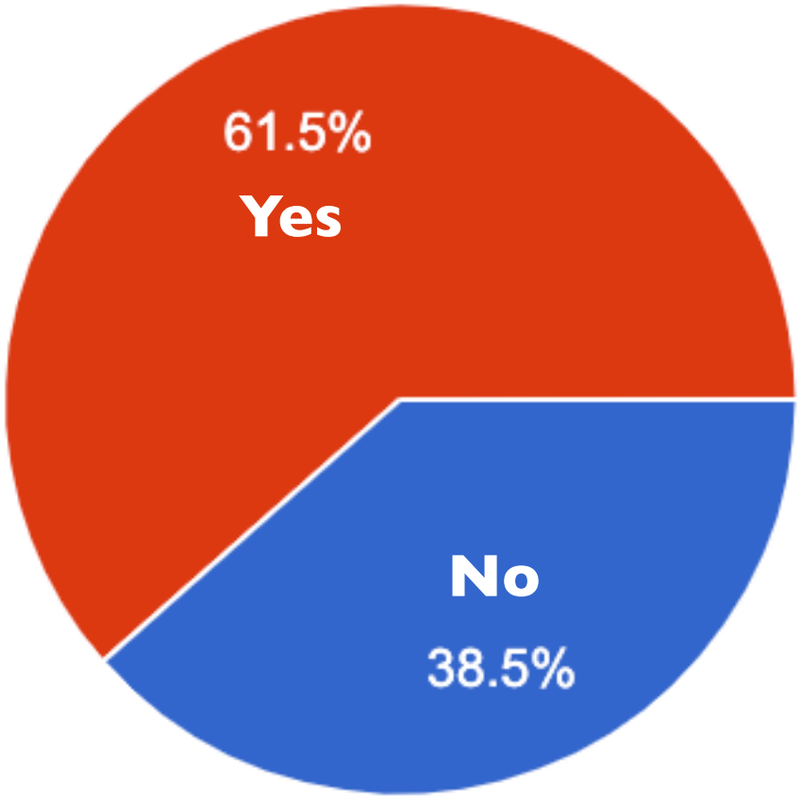


if you see this, it means you're signed in via Gmail instead of your campus email

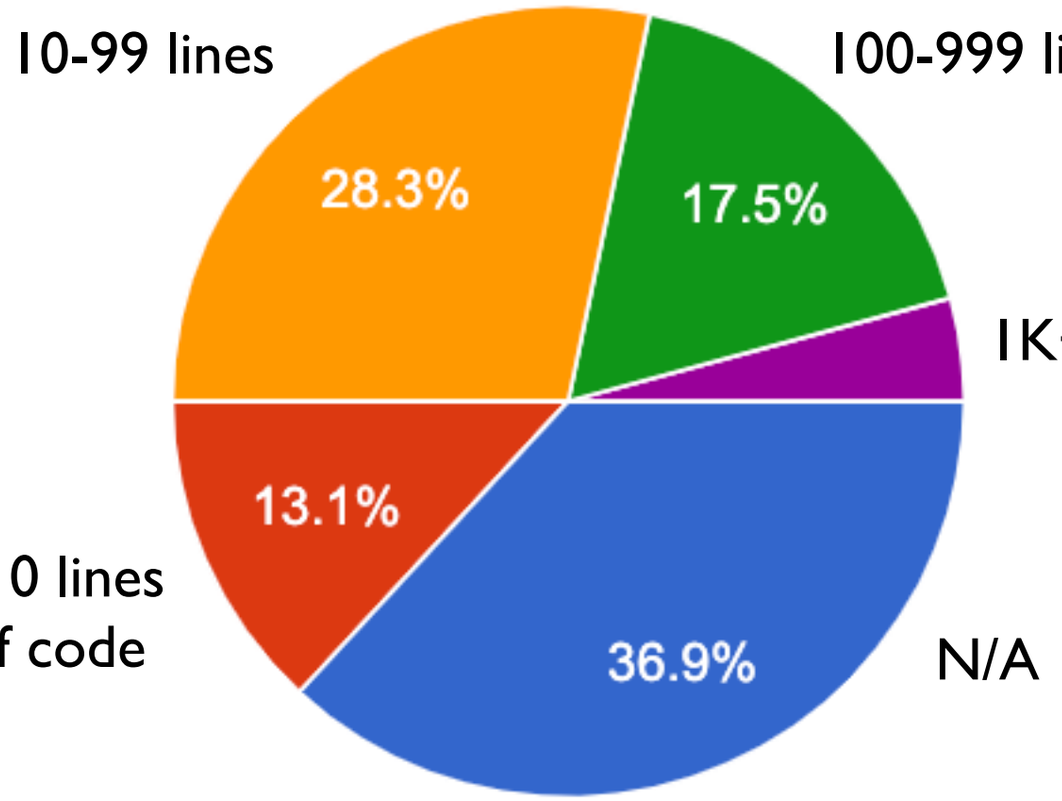
if you were automatically signed into gmail without being asked, consider clearing cookies or using an Incognito Window (in Chrome)



Some results from Spring 2019...



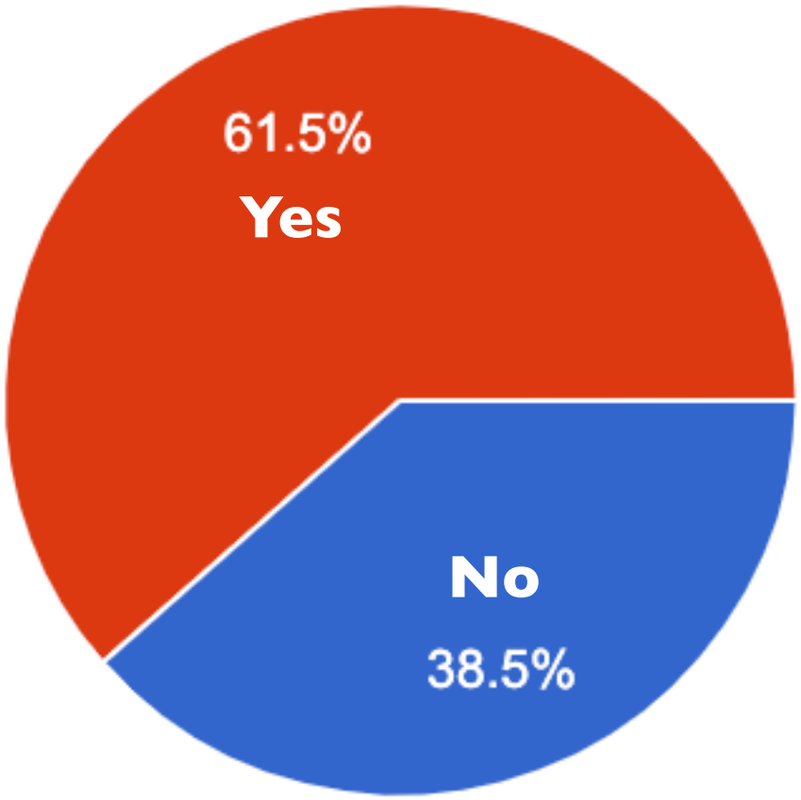
First time taking a CS course?



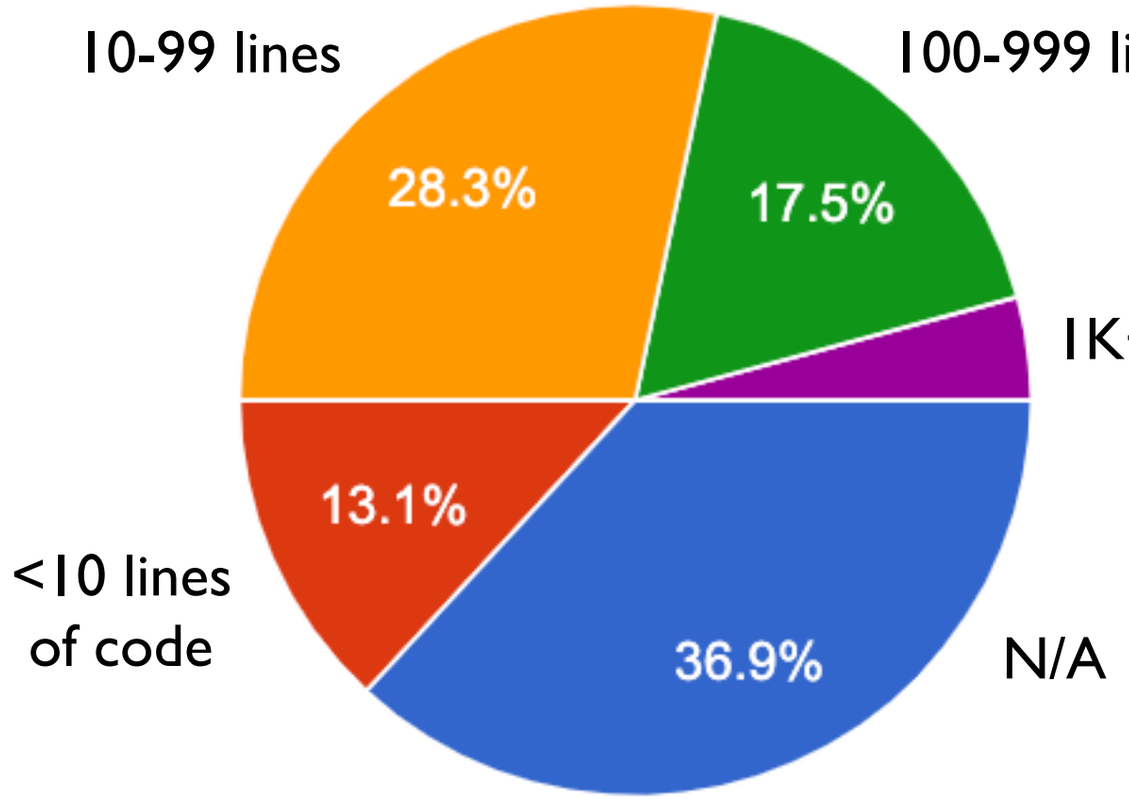
Largest program written prior?

from 548 students

Some results from Spring 2019...



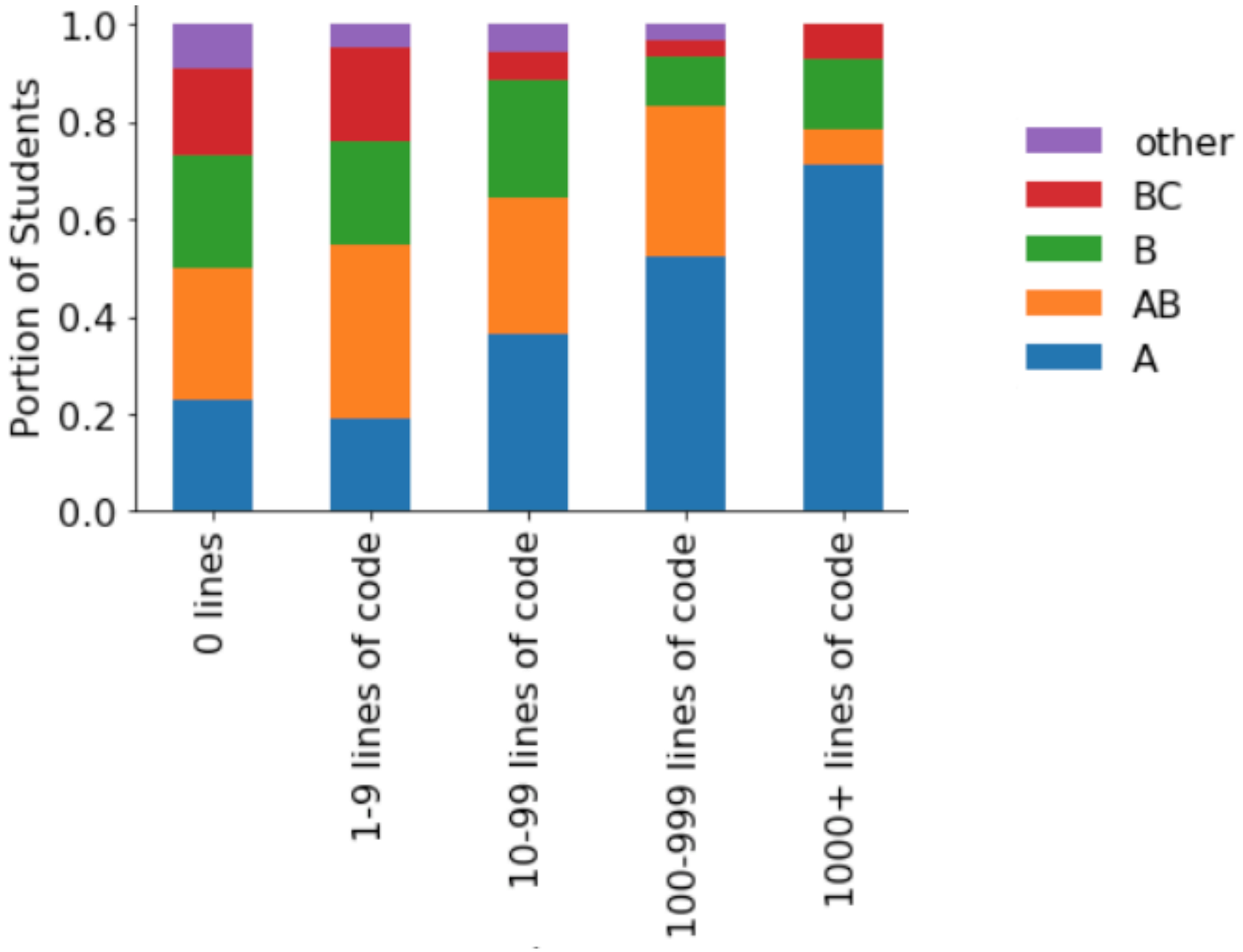
First time taking a CS course?



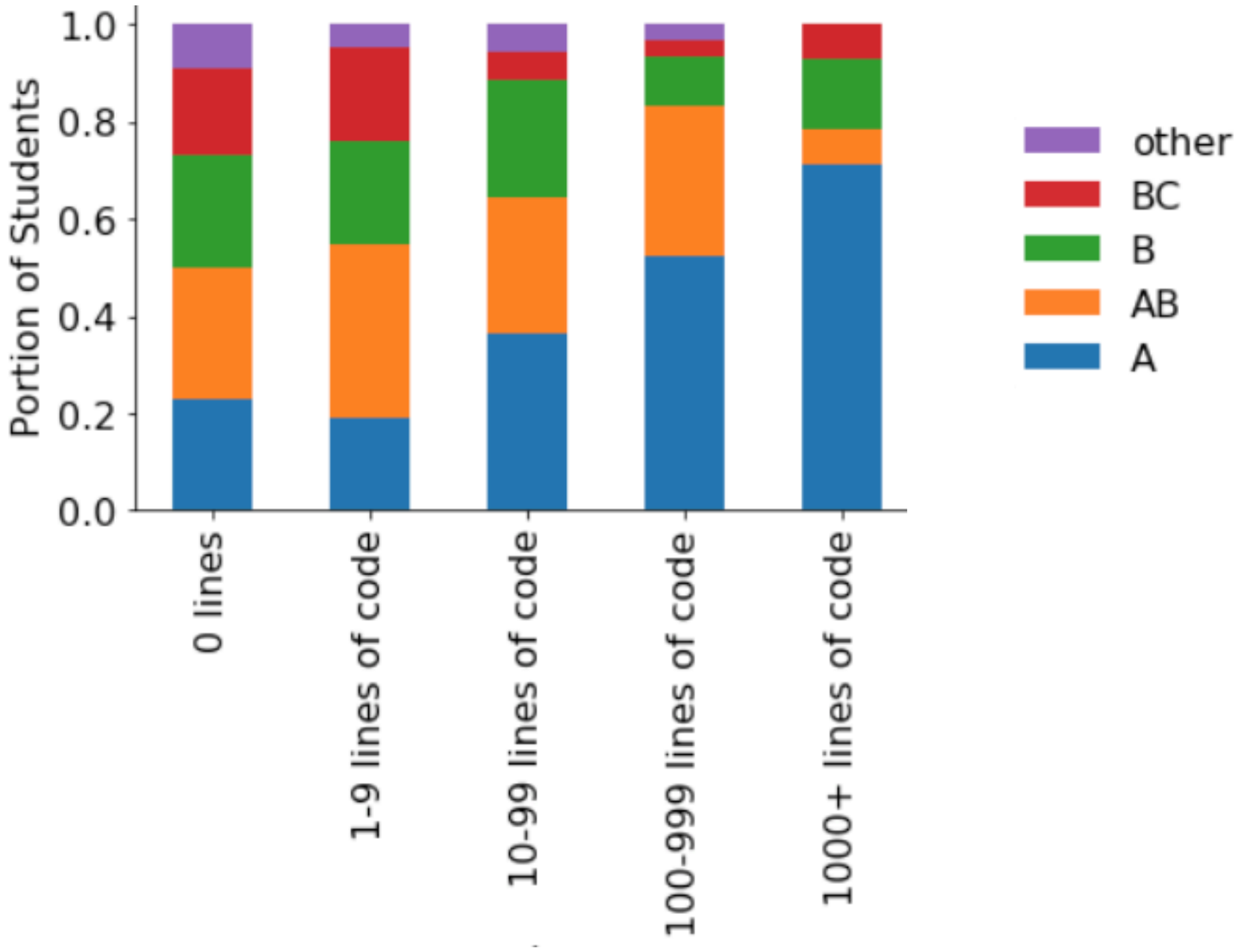
Largest program written prior?

how did students in each group do?

from 548 students



Experience and grades



Some comments on Fall 2018 course evals:

- *I am a senior CS student, this class was very easy for me*
- *Make it significantly easier. None of [us] will ever code again...*
- *Good course, I think there is a good pace for this course, speaking as someone with zero programming experience coming into the class.*

Today's Topics

Introductions

Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

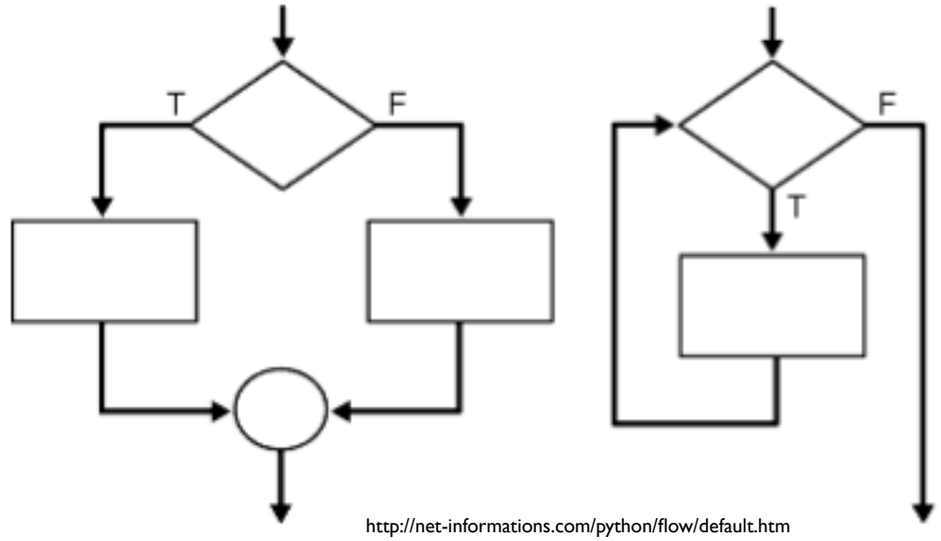
Computer hardware basics

Website

301 Topics

Part 1: Control Flow

- What step is currently executing?
- How to write functions?
- How to conditionally do something?
- How to repeat steps?



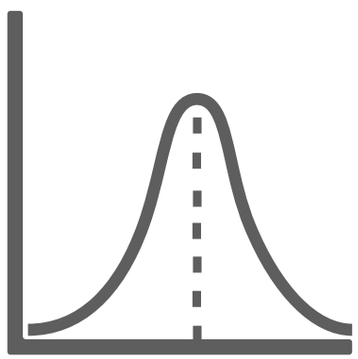
http://net-informations.com/python/flow/default.htm

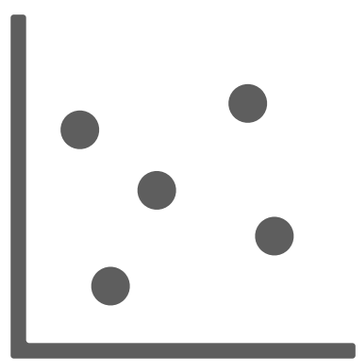
Part 2: State

- How to structure lots of data?
- How to save data in files?

Part 3: Data Science

- Tabular data
- Internet
- Databases
- Plotting





Today's Topics

Introductions

Course overview

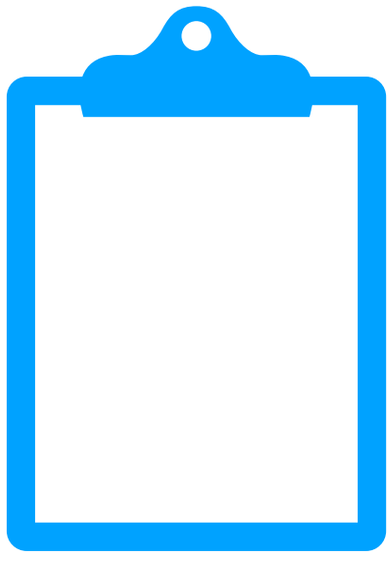
- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

Lecture Style

general concepts

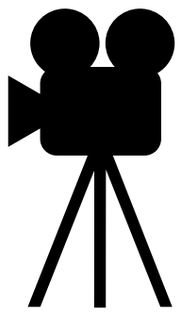


worksheet practice

live coding

Your role

- do readings before or after
- feel free to bring your laptop along!
- I'll often post recordings of lecture 2 here (but don't count on it!): https://tyler.caraza-harter.com/cs301/fall19/videos.html
- I love to get questions



Thoughts on Attendance...

Suggestions from Course Evals:

I think one MAJOR thing to do is make lecture mandatory [student from F19]

Make labs semi mandatory [student from S19]

Feedback Form:

There was someone in front of me today just watching movies and buying things online today in the very front row and it was incredibly distracting as their screen is right in front of the screen you are working on, so I had to be looking at it. I didn't really feel comfortable asking him not to do it again as just another student [student from S19]

Thoughts on Attendance...

Suggestions from Course Evals:

I think one MAJOR thing to do is make lecture mandatory [student from F19]

Make labs semi mandatory [student from S19]

There are lot's of non-mandatory things you should do
(it's up to you how to utilize the resources I'll provide)

Feedback Form:

There was someone in front of me today just watching movies and buying things online today in the very front row and it was incredibly distracting as their screen is right in front of the screen you are working on, so I had to be looking at it. I didn't really feel comfortable asking him not to do it again as just another student [student from S19]

Thoughts on Attendance...

Suggestions from Course Evals:

I think one MAJOR thing to do is make lecture mandatory [student from F19]

Make labs semi mandatory [student from S19]

There are lot's of non-mandatory things you should do
(it's up to you how to utilize the resources I'll provide)

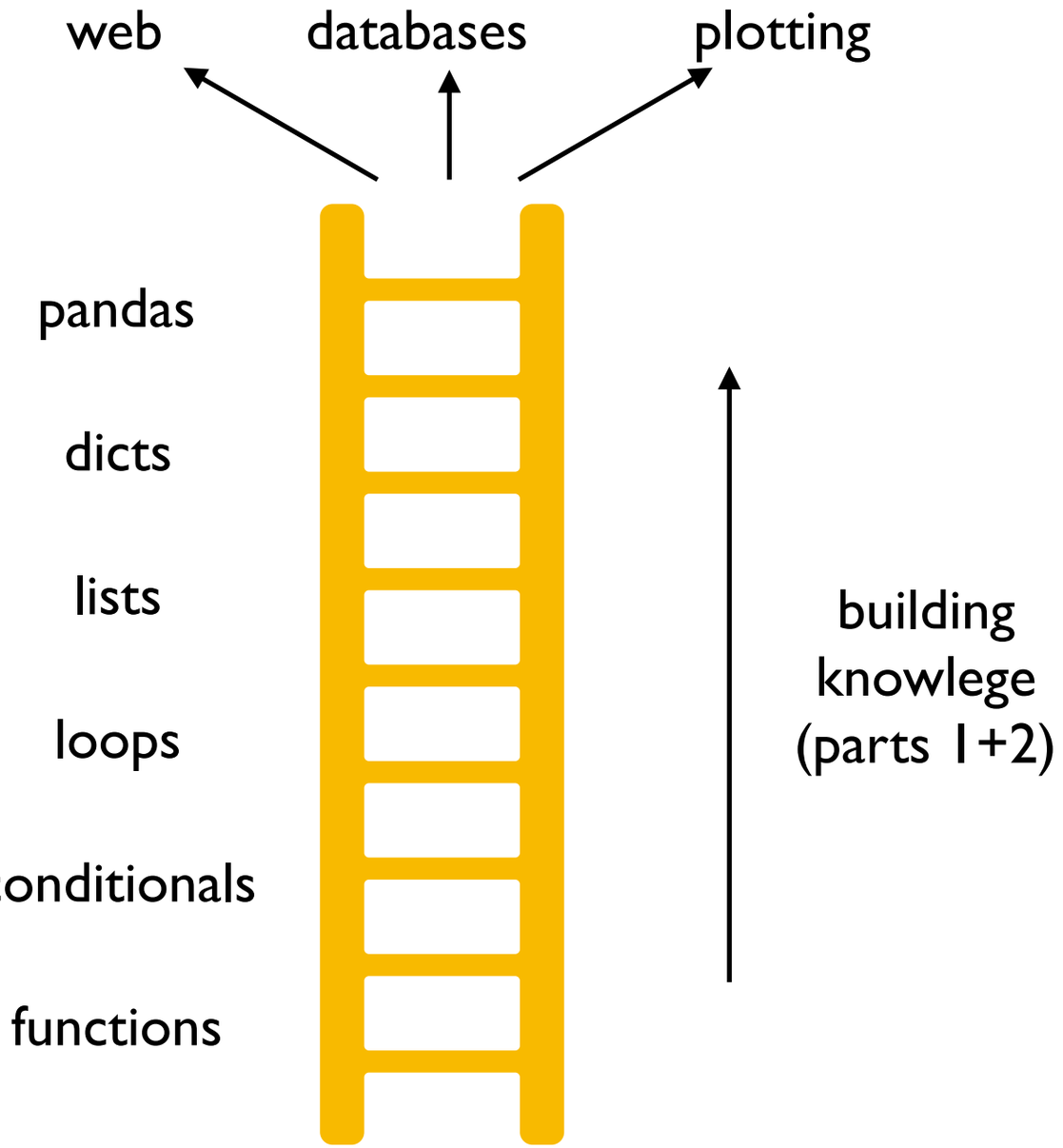
Feedback Form:

There was someone in front of me today just watching movies and buying things online today in the very front row and it was incredibly distracting as their screen is right in front of the screen you are working on, so I had to be looking at it. I didn't really feel comfortable asking him not to do it again as just another student [student from S19]

Lecture rule: *anything you do on your laptop in class must be less interesting than my lectures*

Especially Avoid Holes in Understanding in Parts 1+2 of the course

see Salman Kahn...



Today's Topics

Introductions

Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

Labs

Format

- 75 minutes on Thu or Fri, leave when you're done
- self guided, not graded
- lab document will be posted each week
- do the lab before starting the project!
- get help with projects+content too! (just ask a TA/mentor)

People

- best to do lab docs with a partner
- 1-2 TAs will be there to answer questions

Computers

- bring your laptop!
- use backup lab computers if necessary

we will have labs this first week

(also, get any help needed installing Python during this one)

Today's Topics

Introductions

Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

Readings (all free!)

Think Python, 2nd Edition

- Allen B. Downey
- Assumes no programming background
- It's very concise
- Get the 2nd edition, which is for **Python 3**!

Automate the Boring Stuff

- Al Sweigart
- Useful for some more advanced topics related to using data

CS 301 Notes

Course Notes

- 301 instructors
- Mostly for data science part of class

Today's Topics

Introductions

Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

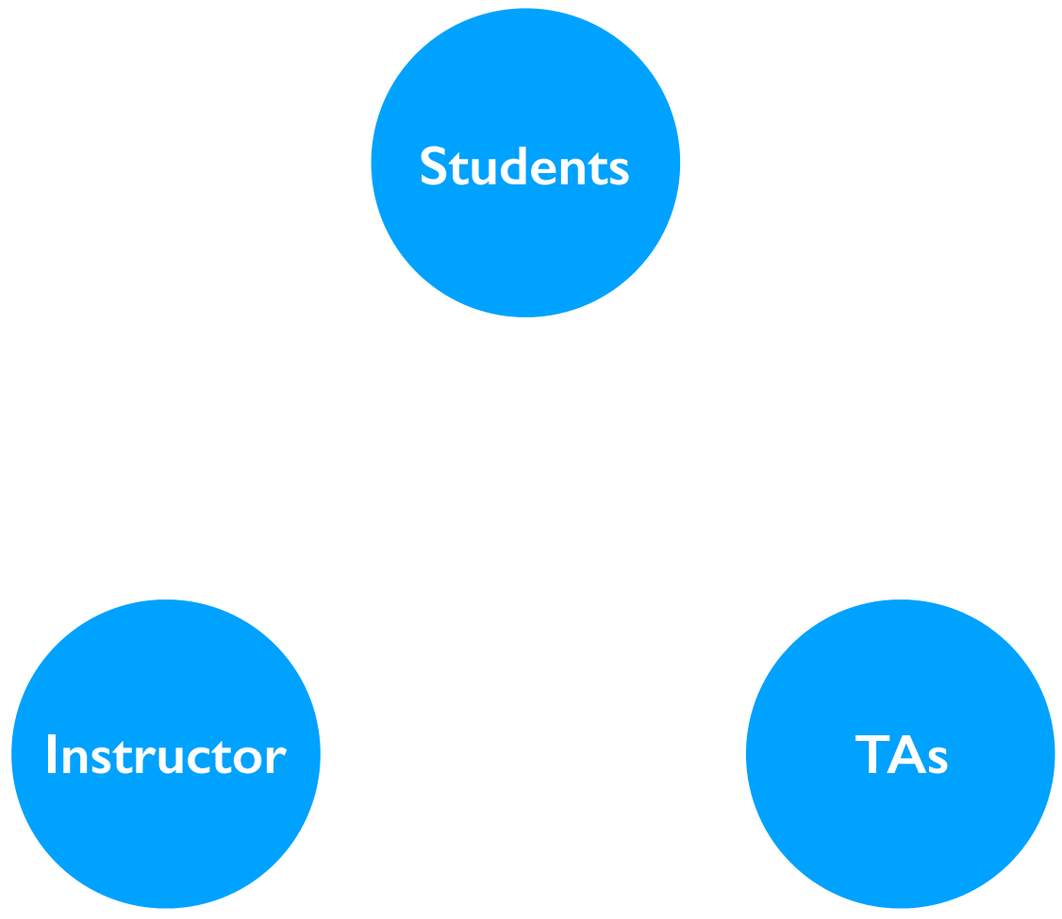
Communication is CS 301

Good communication is critical for a class of this size

- Who needs to communicate: students, TAs (+mentors!), instructors

Besides direct email, we'll use five communication tools

- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas



Communication is CS 301

Good communication is critical for a class of this size

- Who needs to communicate: students, TAs (+mentors!), instructors

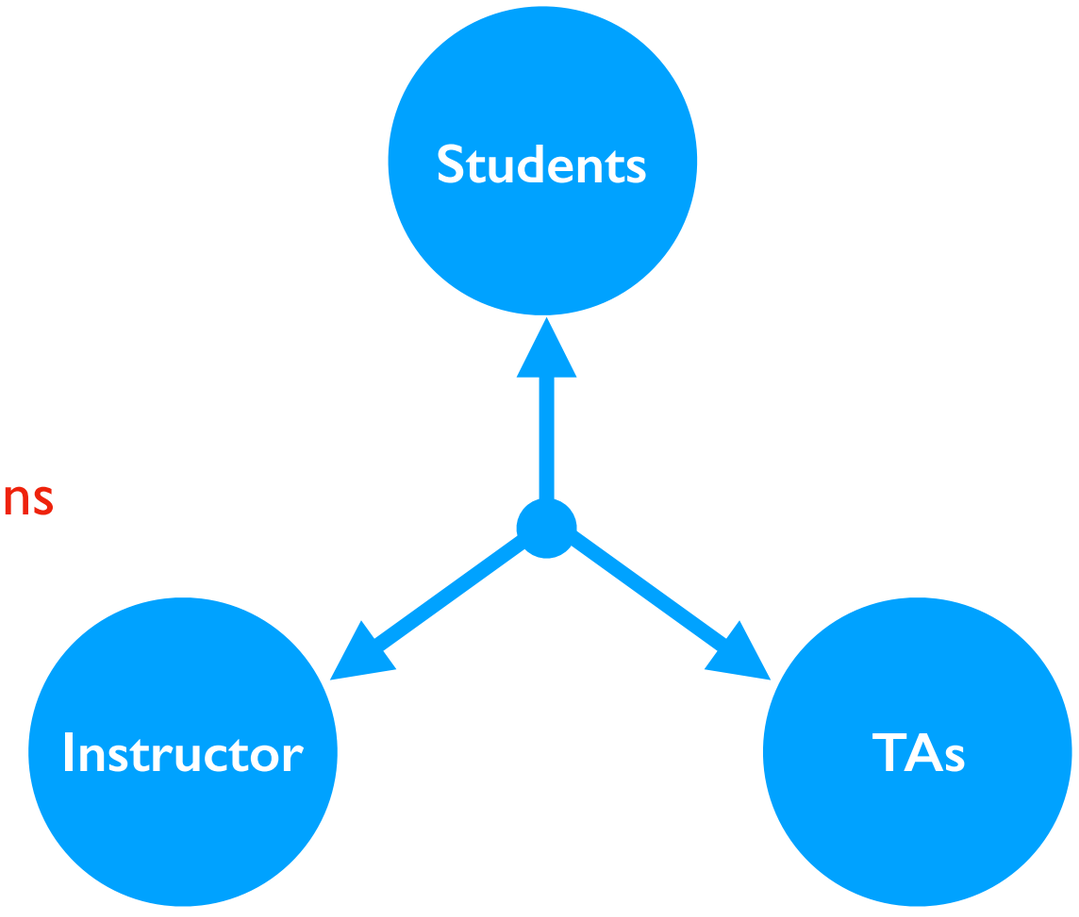
Besides direct email, we'll use five communication tools

- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas

Rule 1: don't most more than 5 lines of code

Rule 2: check other posts and project corrections to avoid repeat questions

Note: we'll keep a pinned post of current office hours here



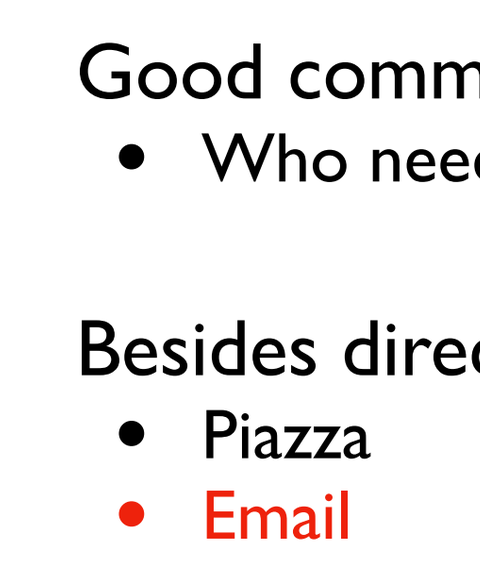
Communication is CS 301

Good communication is critical for a class of this size

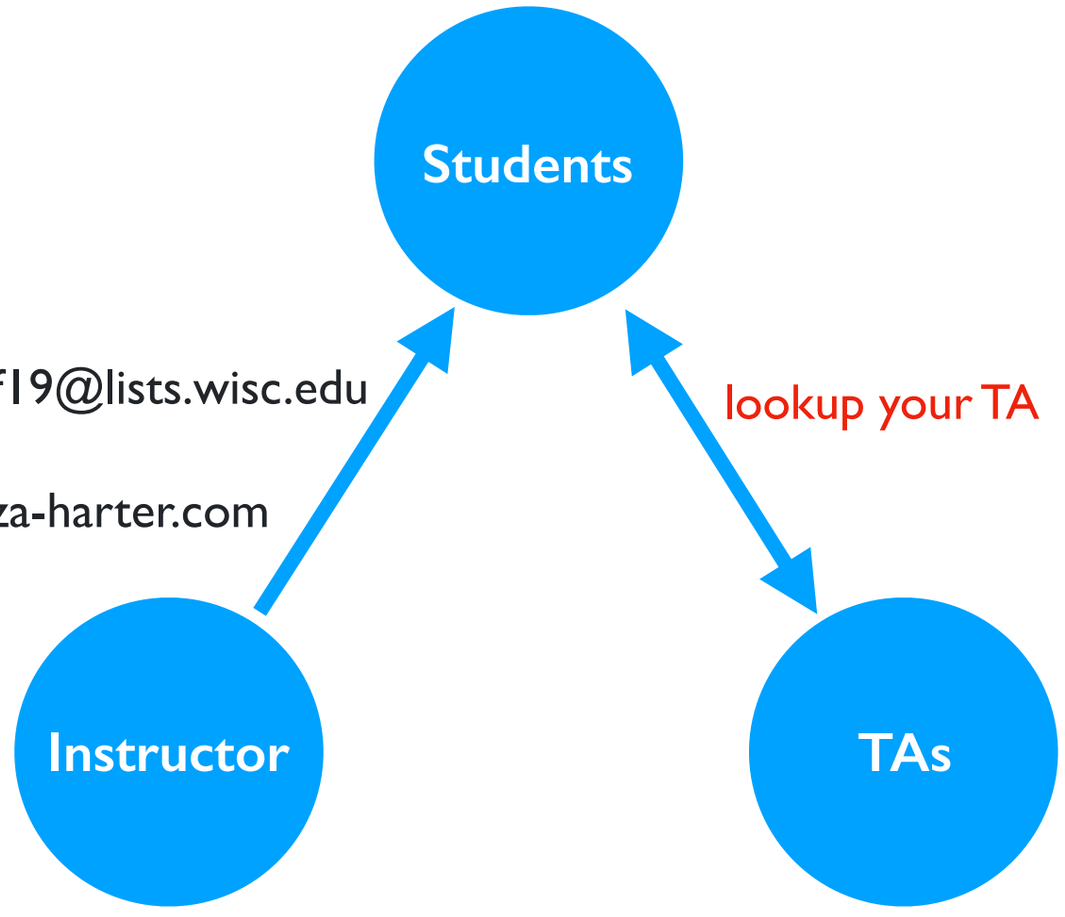
- Who needs to communicate: students, TAs (+mentors!), instructors

Besides direct email, we'll use five communication tools

- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas



please email your TA before me, though do CC me at tylerharter@gmail.com if you don't get a response within 48 hours.



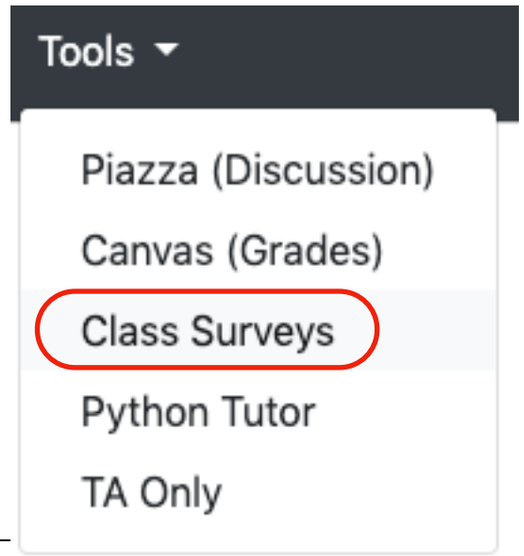
Communication is CS 301

Good communication is critical for a class of this size

- Who needs to communicate: students, TAs (+mentors!), instructors

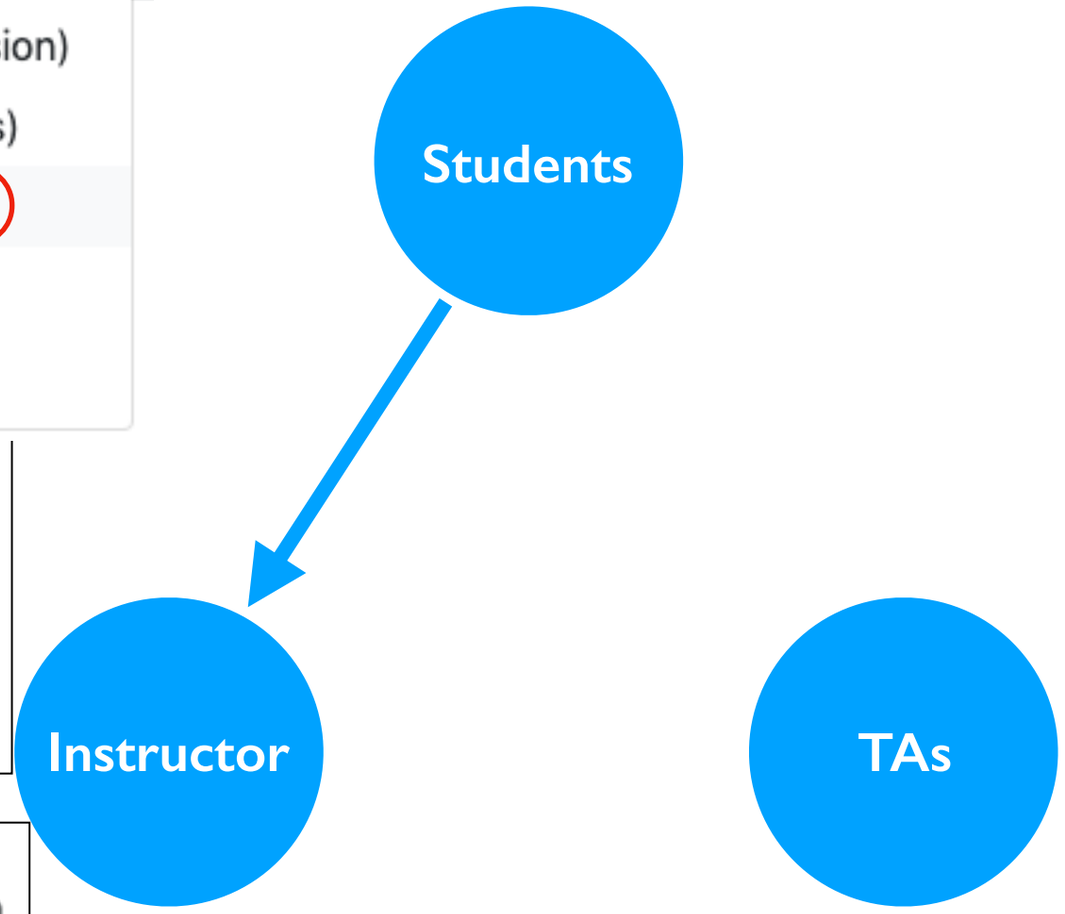
Besides direct email, we'll use five communication tools

- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas



2. Feedback Form. If you have any issues with the class or suggestions for improvement, please let us know sooner rather than later; we may be able to make changes more rapidly than you might imagine. This is optionally anonymous, but it's always nice to know who you are (sometimes it makes sense to have followup conversations).

4. Thank You! Has a TA or mentor provided exceptional help, during office hours, Shelf hours, lab, etc? Thank them by filling out this form, and I'll pass along the feedback.



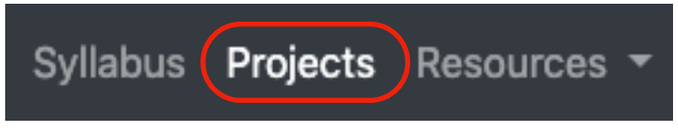
Communication is CS 301

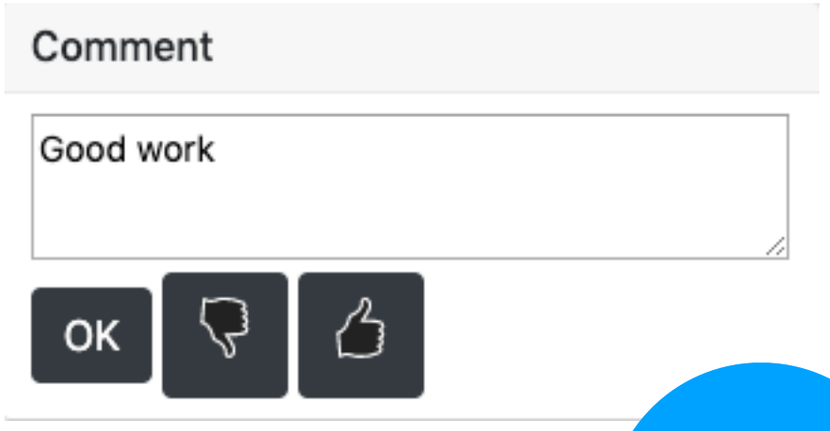
Good communication is critical for a class of this size

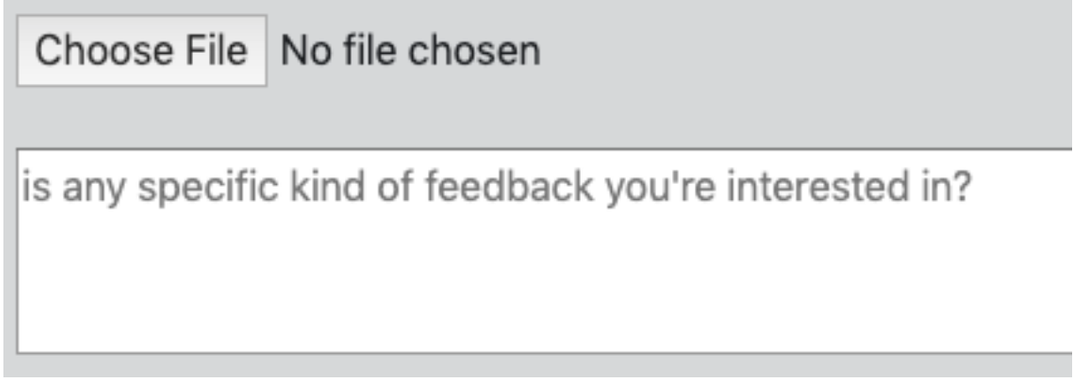
- Who needs to communicate: students, TAs (+mentors!), instructors

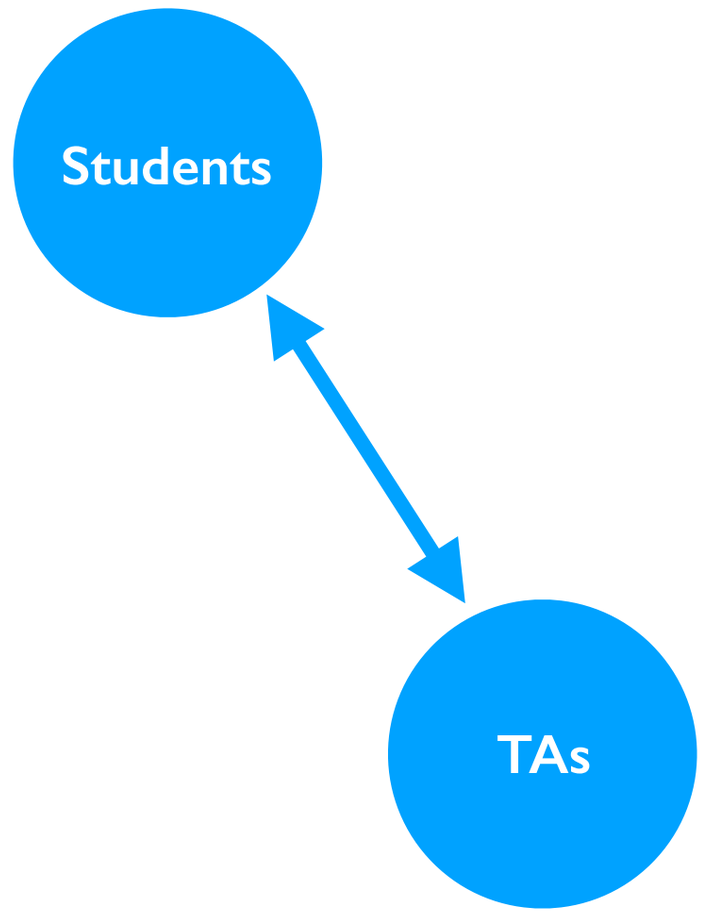
Besides direct email, we'll use five communication tools

- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas









Communication is CS 301

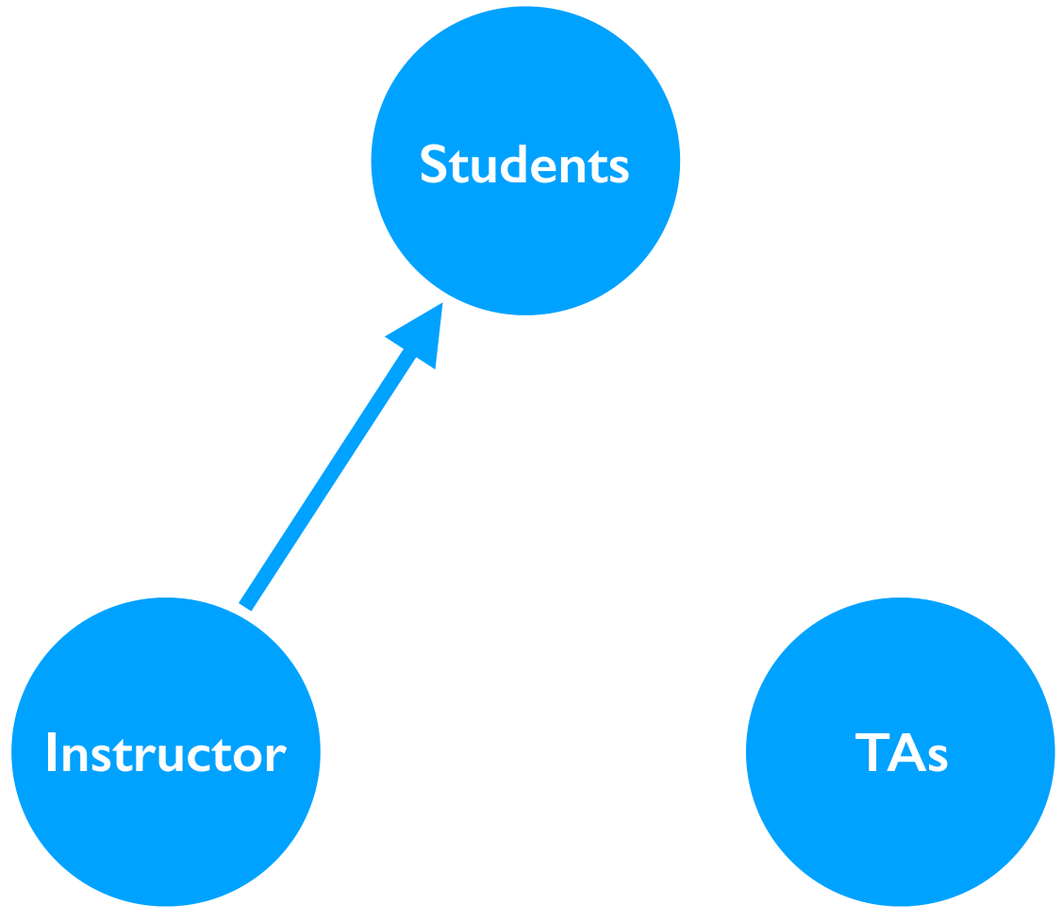
Good communication is critical for a class of this size

- Who needs to communicate: students, TAs (+mentors!), instructors

Besides direct email, we'll use five communication tools

- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas

our Canvas is
ONLY for grades



Today's Topics

Introductions

Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

Grades

49% - programming projects

- **10 projects**, not evenly weighted
- we'll share grading tests with you - avoid surprise
- learning to program is the most import part of the course

50% - exams

- 15% midterm 1 (evening)
- 15% midterm 2 (evening)
- 20% final
- finalized times coming soon

1% - participation

- filling surveys, following directions, other

The Final Curve

The curve will be set at the end of the semester, based on sum of all points earned.

I try to keep the grade distribution similar across semesters: https://registrar.wisc.edu/grade-reports/

I'll tweak to minimize students on the margin.

Guarantees:

- at least 95% guarantees an **A**
- at least 85% guarantees a **B** (or better)
- at least 70% guarantees a **C** (or better)
- at least 60% guarantees a **D** (or better)

Grades

49% - programming projects

- **10 projects**, not evenly weighted
- we'll share grading tests with you - avoid surprise
- learning to program is the most import part of the course

50% - exams

- 15% midterm 1 (evening)
- 15% midterm 2 (evening)
- 20% final
- finalized times coming soon

1% - participation

- filling surveys, following directions, other

Today's Topics

Introductions

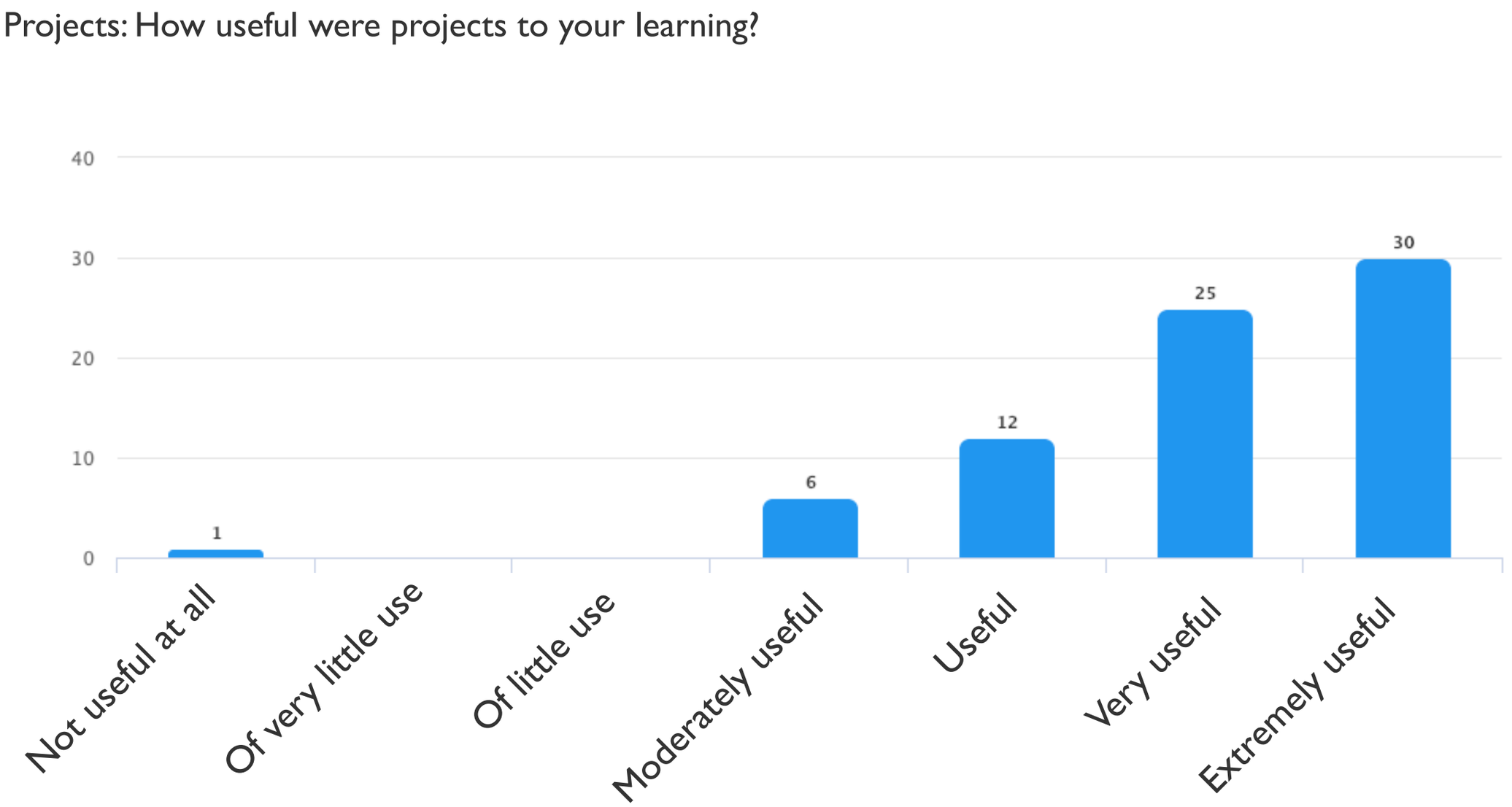
Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

Prior student reaction to projects



Projects are the heart and soul of CS 301

Project Overview

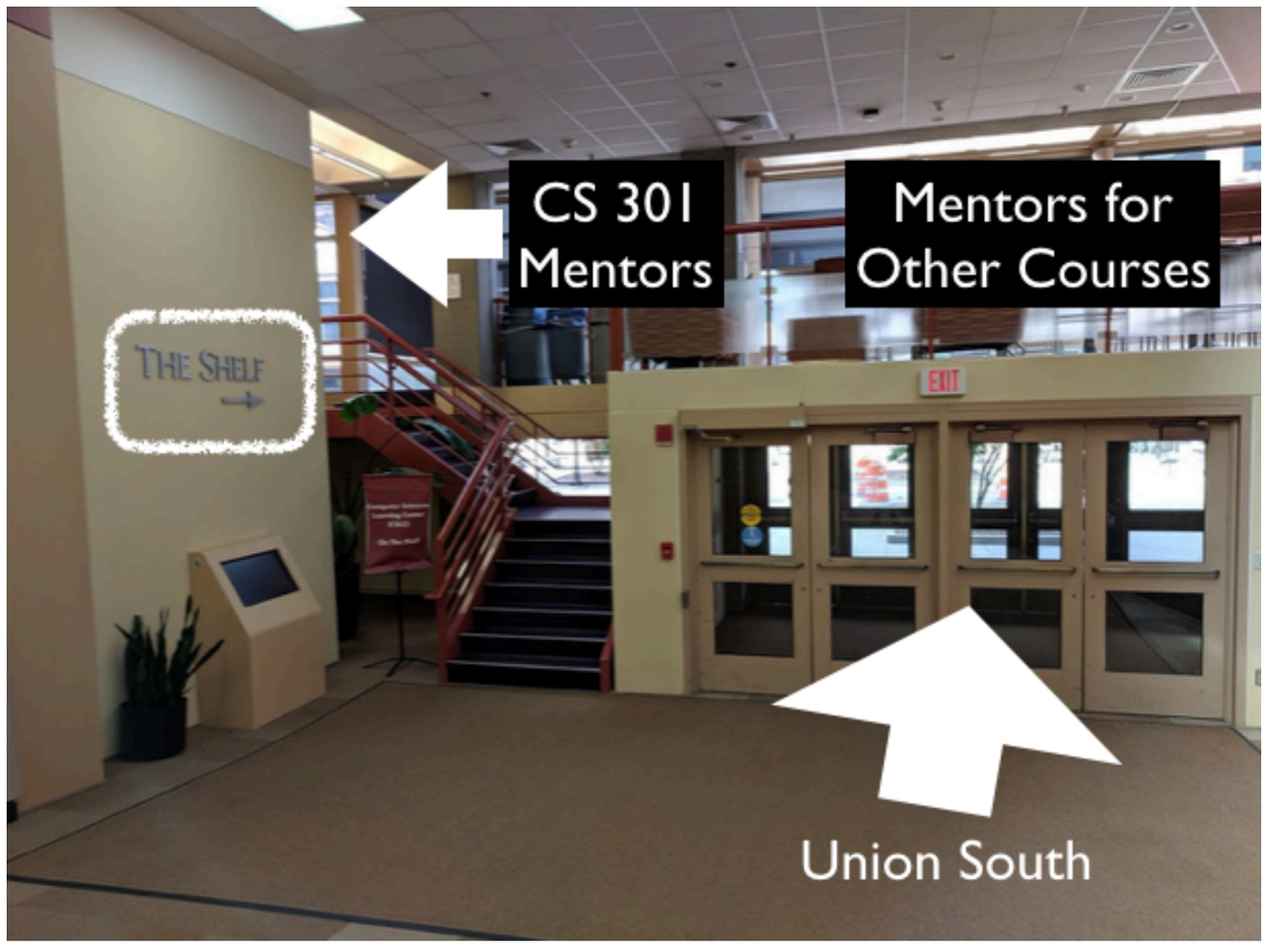
Nearly all projects will relate to some dataset

Timeline

- Projects will be due most weeks, on Wed, at midnight
- You get 7 late days, use them wisely!
- Contact us about any issues

Getting help

- Piazza/email
- Lab sessions
- Instructor or TA office hours
- Shelf Hours
  - Sun (2-8pm)
  - Mon-Wed (3-9pm)



Pair Programming

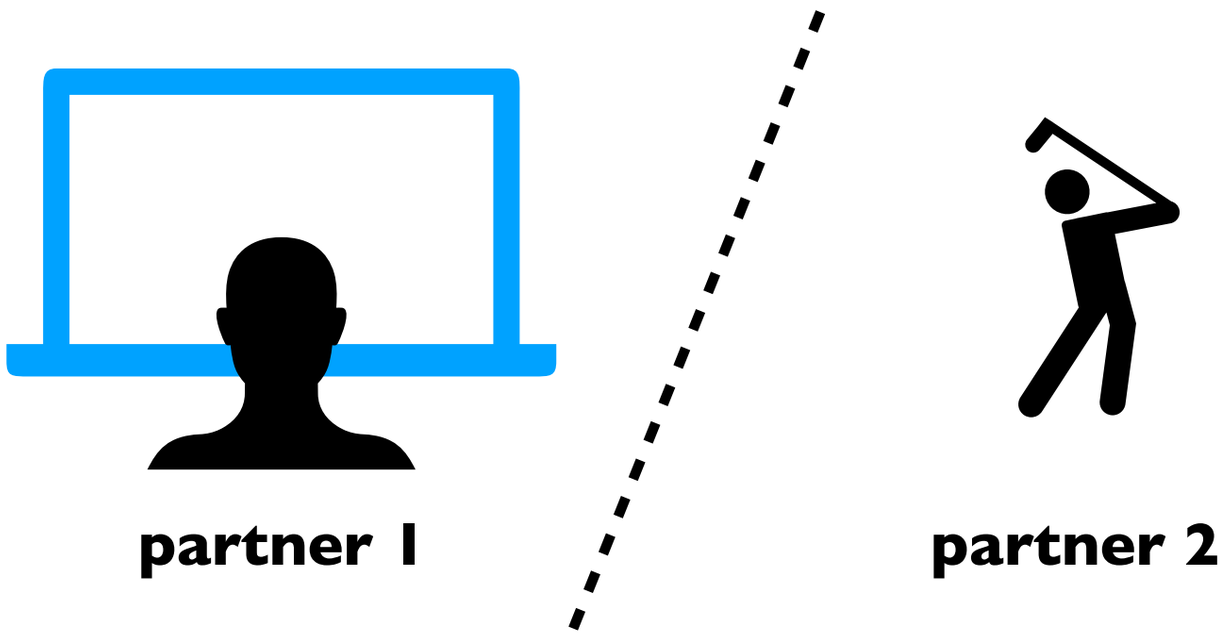
You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)

Pair Programming

You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)

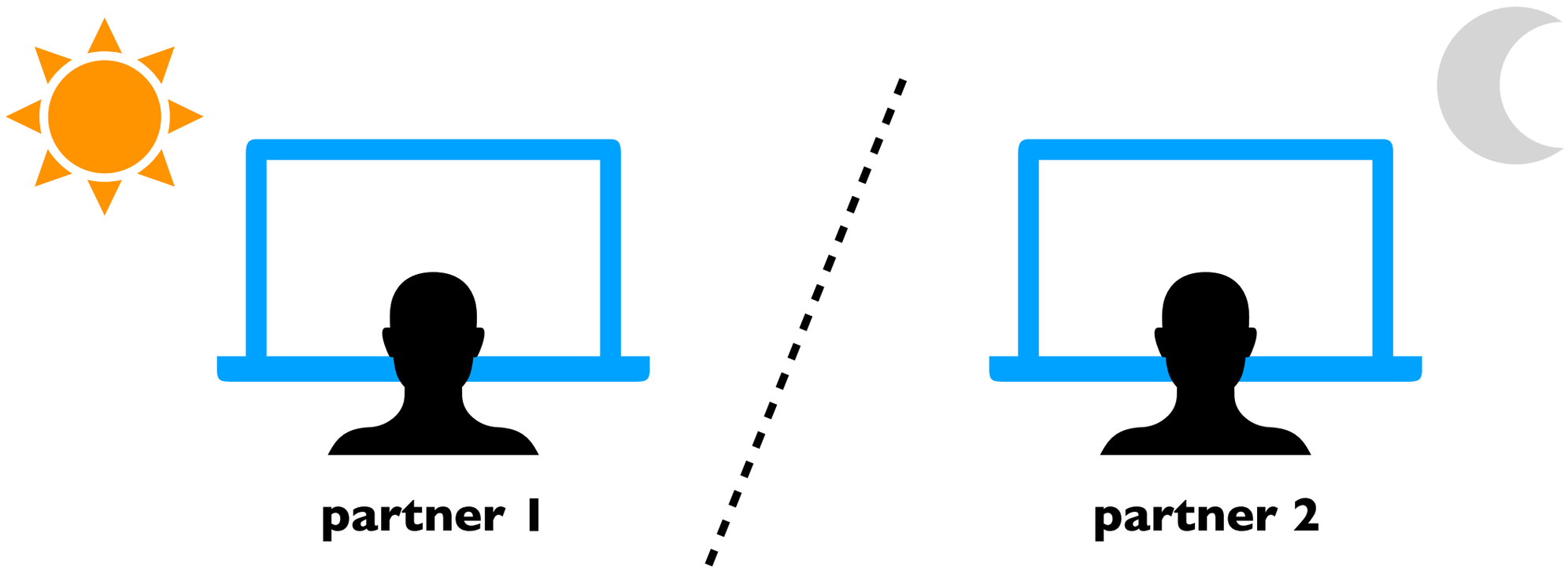


bad: partners don't share work

Pair Programming

You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)

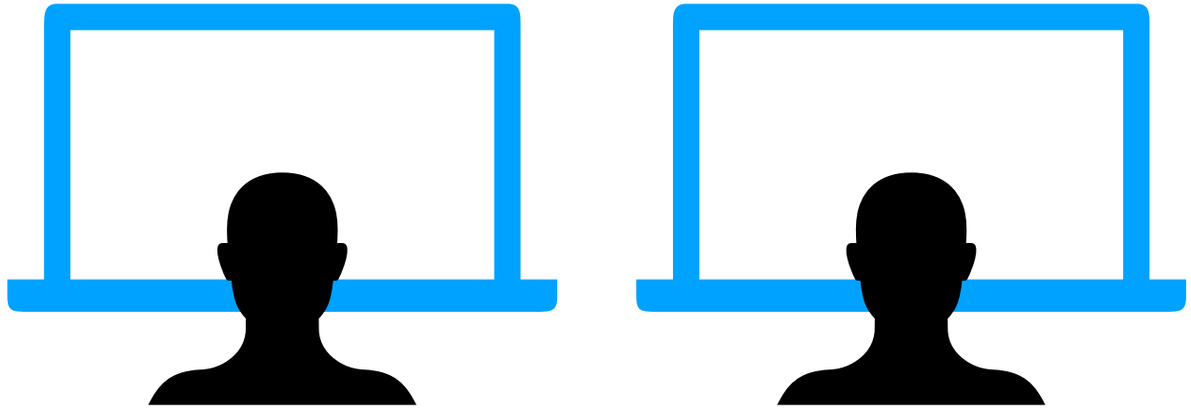


bad: working on different parts at different times

Pair Programming

You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)

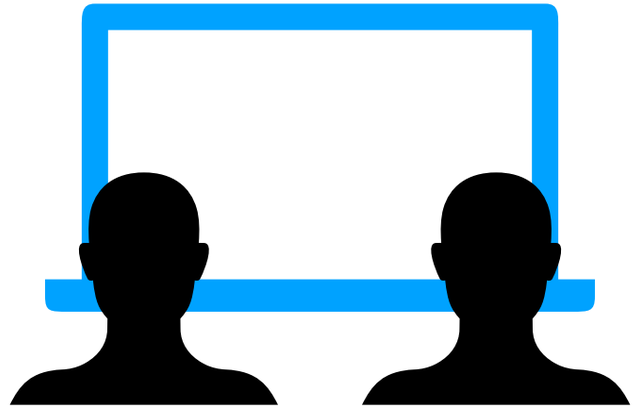


better: working alongside each other

Pair Programming

You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)

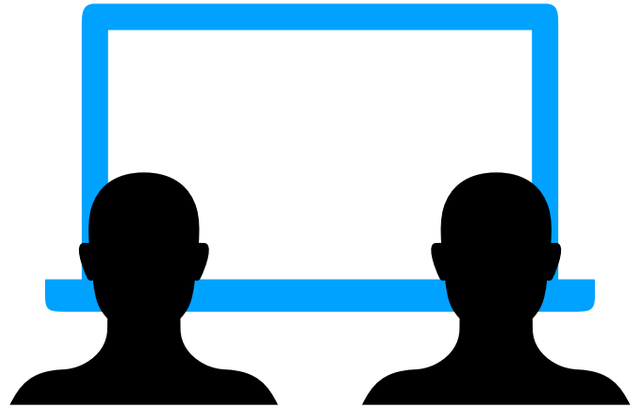


best: working on same computer

Pair Programming

You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)



Suggestions

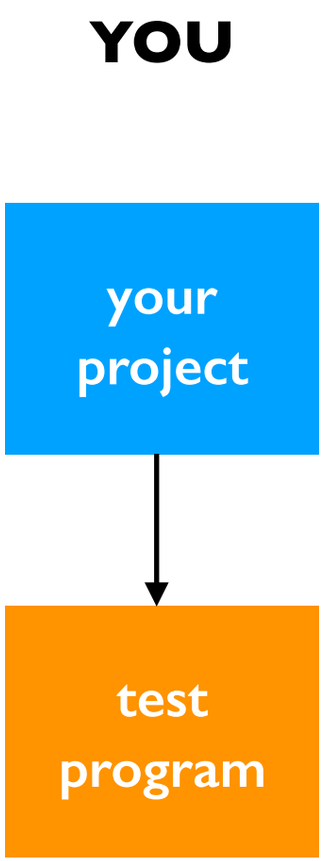
- Take turns coding (don't be greedy/aggressive!)
- One person types, other makes suggestions and thinks about design

Project Grading

YOU

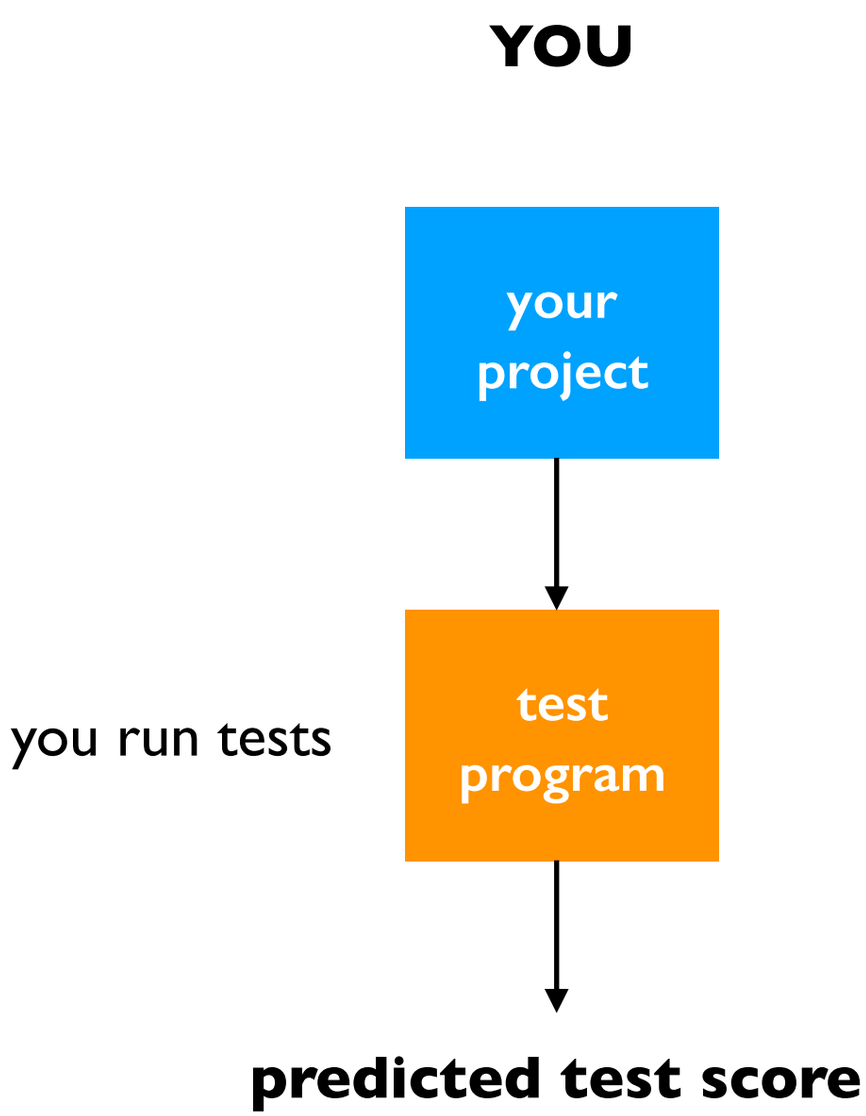
your project

Project Grading

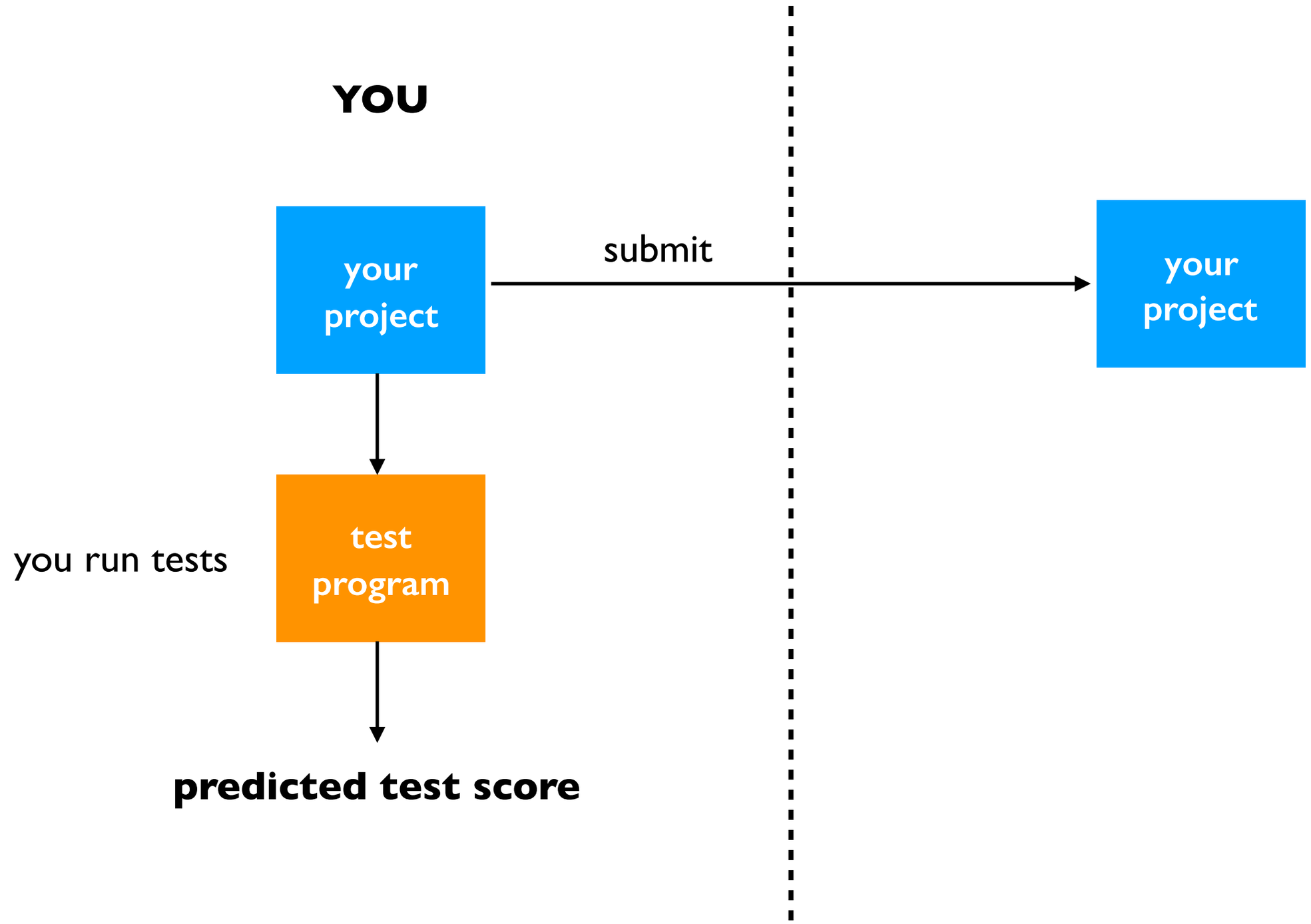


you run tests

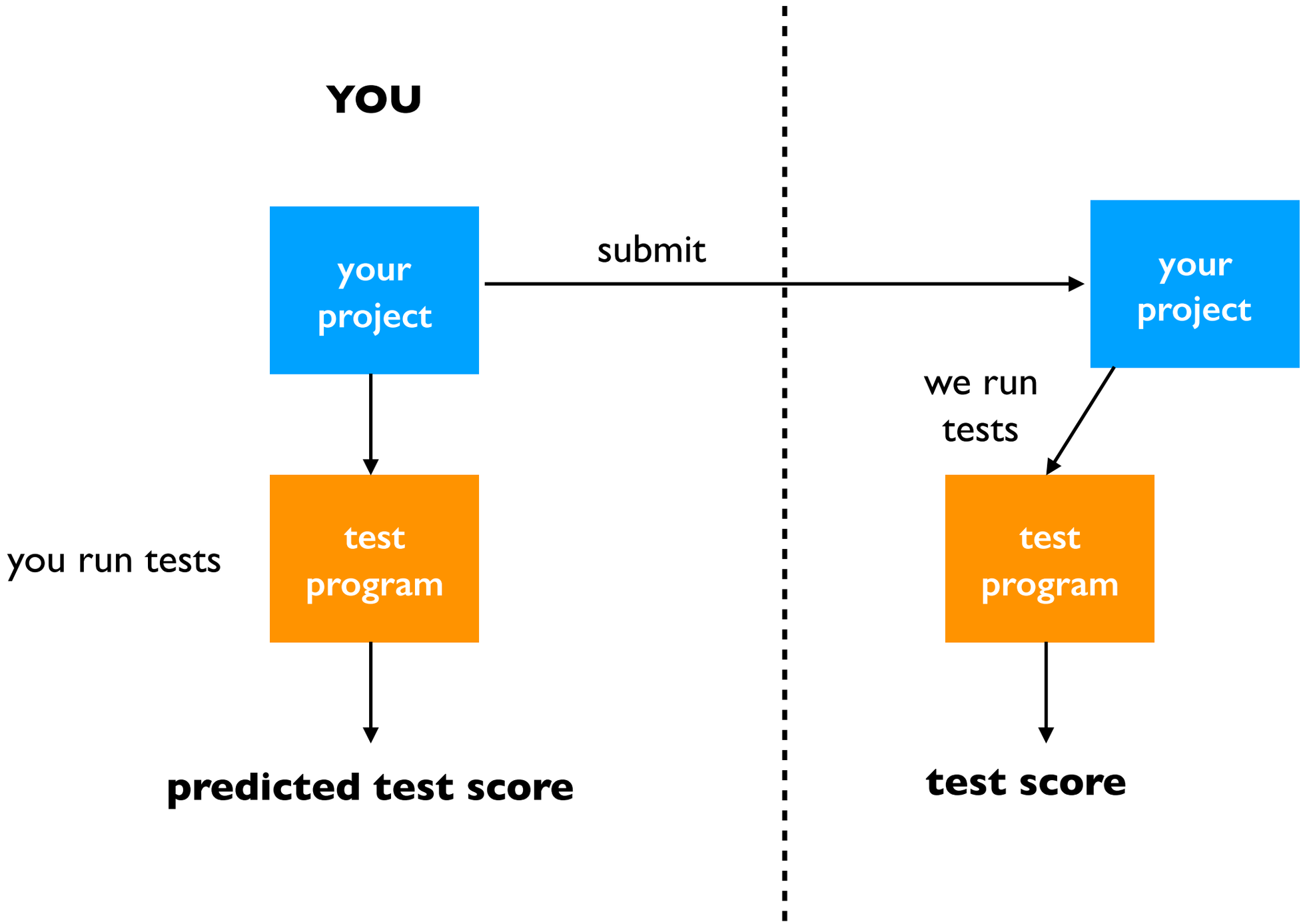
Project Grading



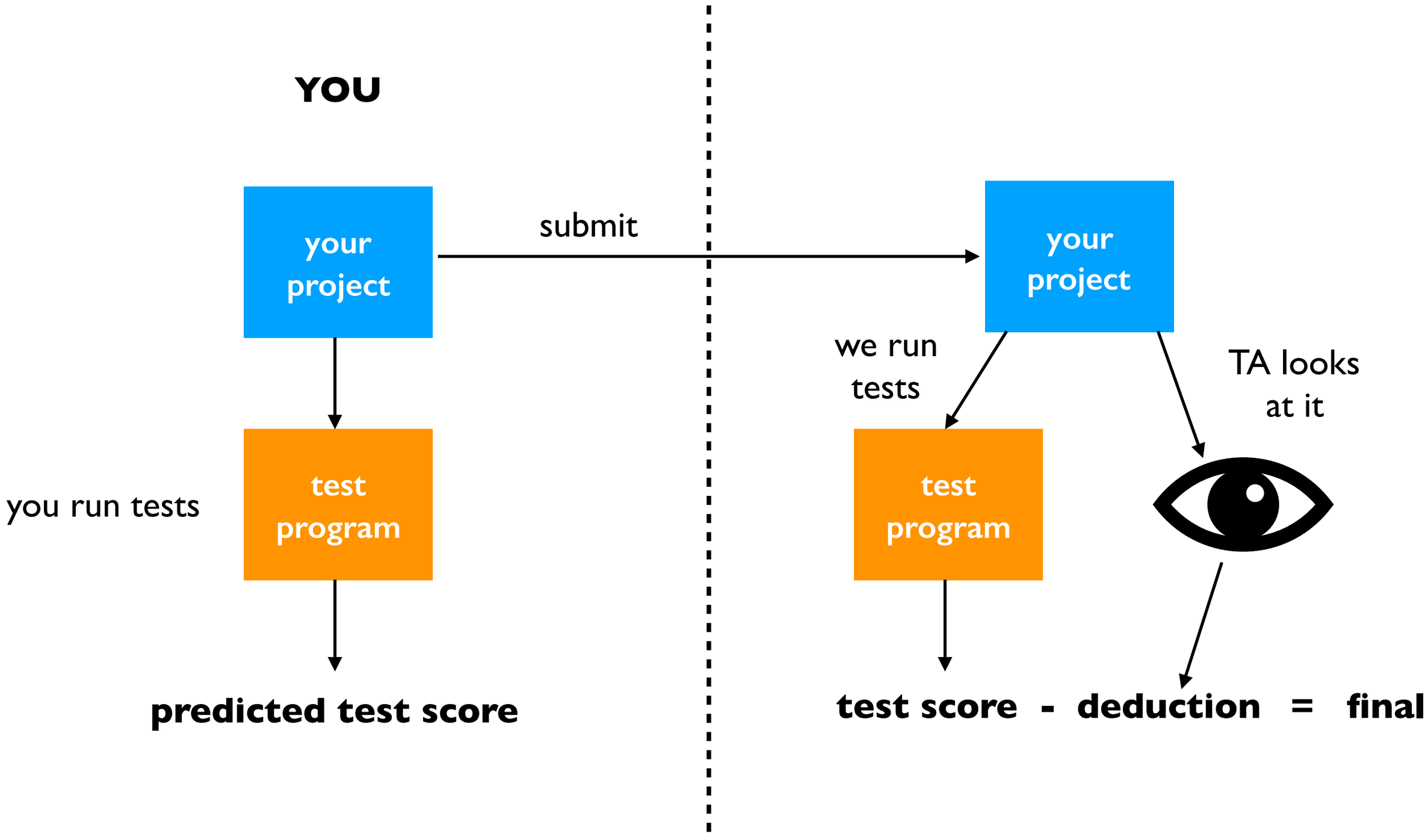
Project Grading



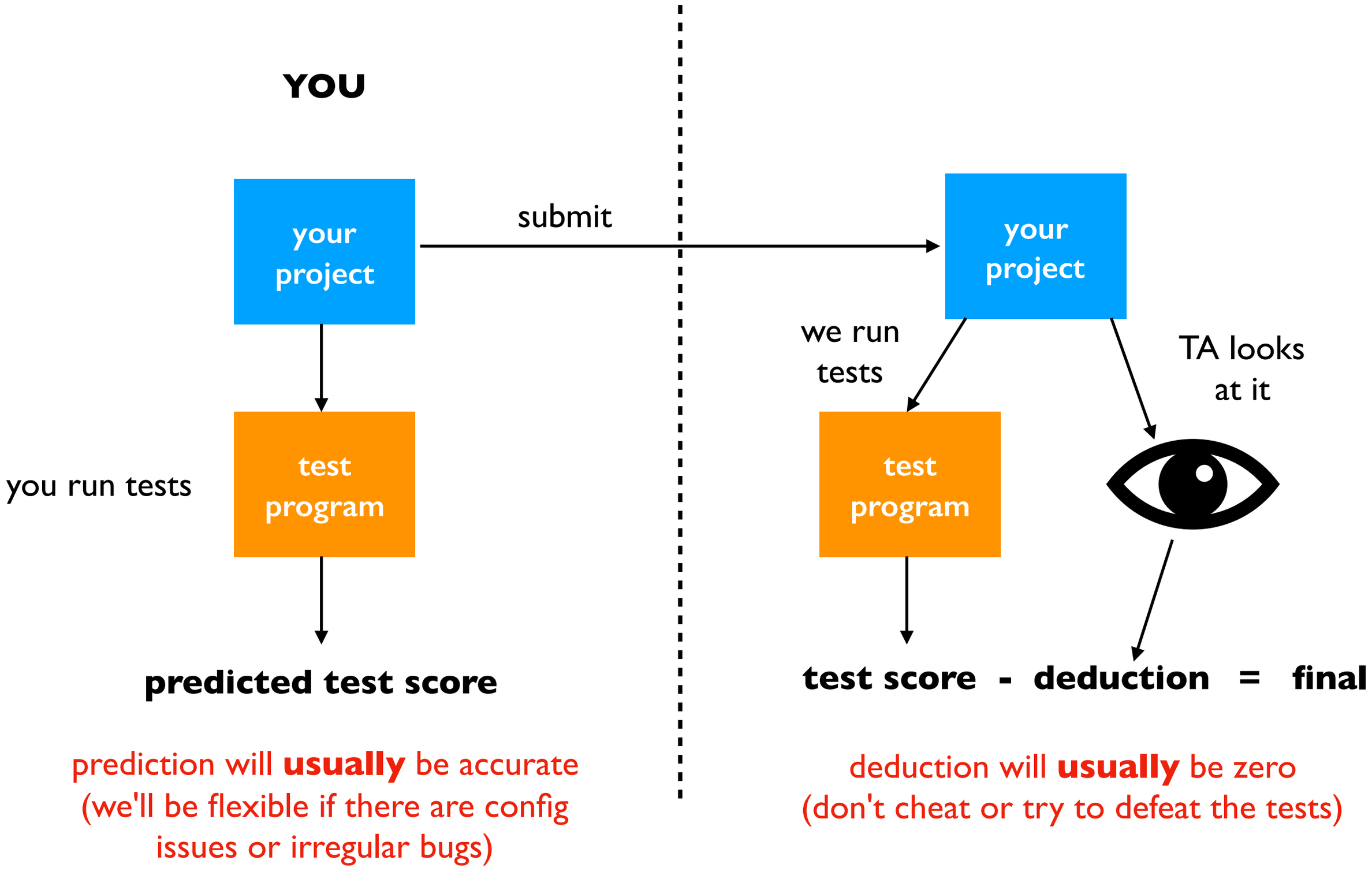
Project Grading



Project Grading



Project Grading

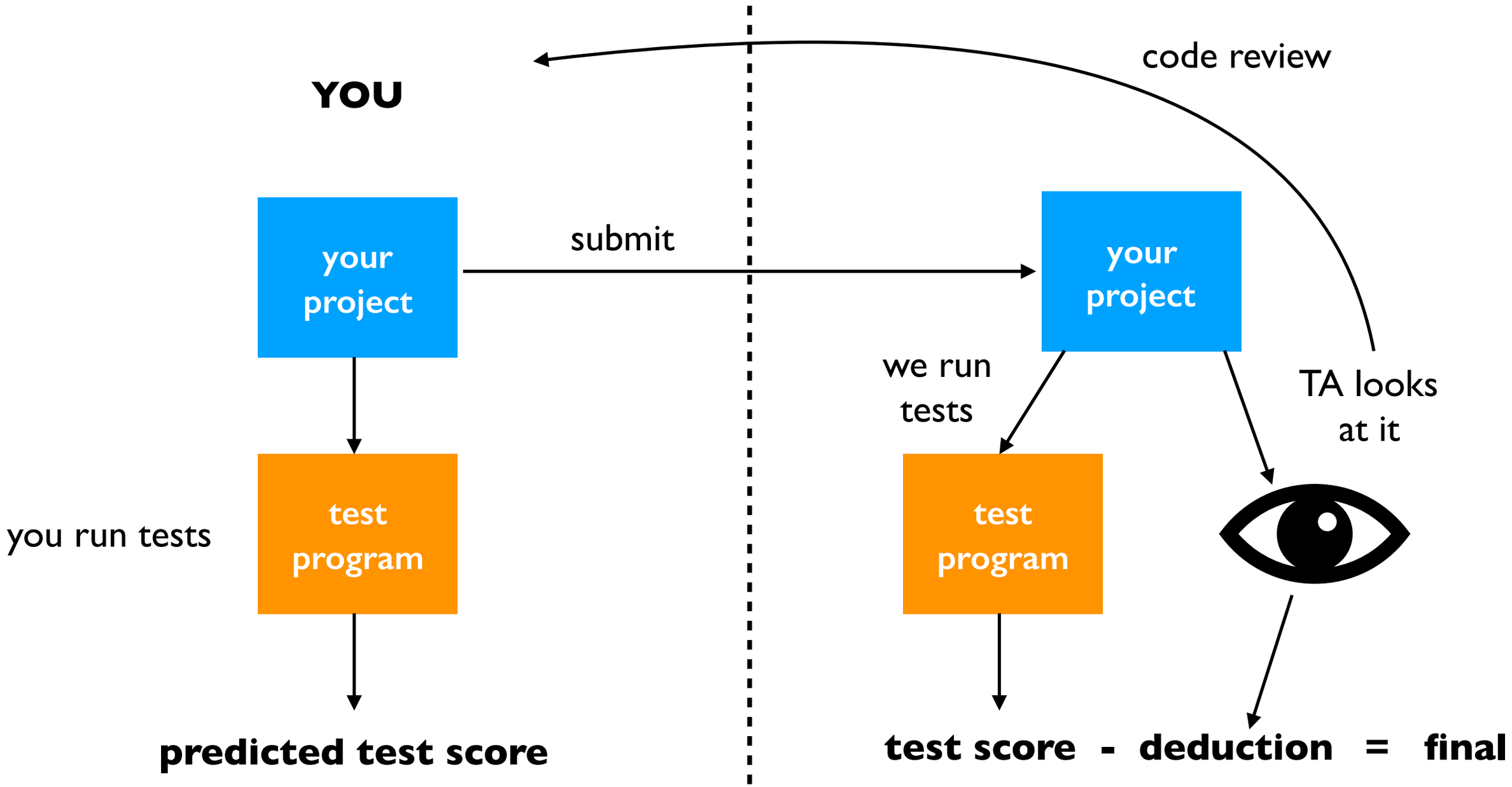


prediction will **usually** be accurate
(we'll be flexible if there are config issues or irregular bugs)

deduction will **usually** be zero
(don't cheat or try to defeat the tests)

Project Grading

feedback is mostly about how
to do things better or more simply
(valuable even if you score 100%)



prediction will **usually** be accurate
(we'll be flexible if there are config
issues or irregular bugs)

deduction will **usually** be zero
(don't cheat or try to defeat the tests)

Today's Topics

Introductions

Course overview

- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams

Computer hardware basics

Website

Exams

There will be two midterms and one final

- Check website for tentative dates/locations
- One 8.5x11 inch notesheet (both sides, printed or written) only
- Exams will be multiple choice scantron

Contents

- cumulative
- ideally not much time pressure
- one goal: reward project partners doing more work over those slacking

projects ≈ writing code

exams ≈ reading code

Today's Topics

Introductions

Course overview

Computer hardware basics

- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Today's Topics

Introductions

Course overview

Computer hardware basics

- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Input/Output

I/O (stands for input/output)

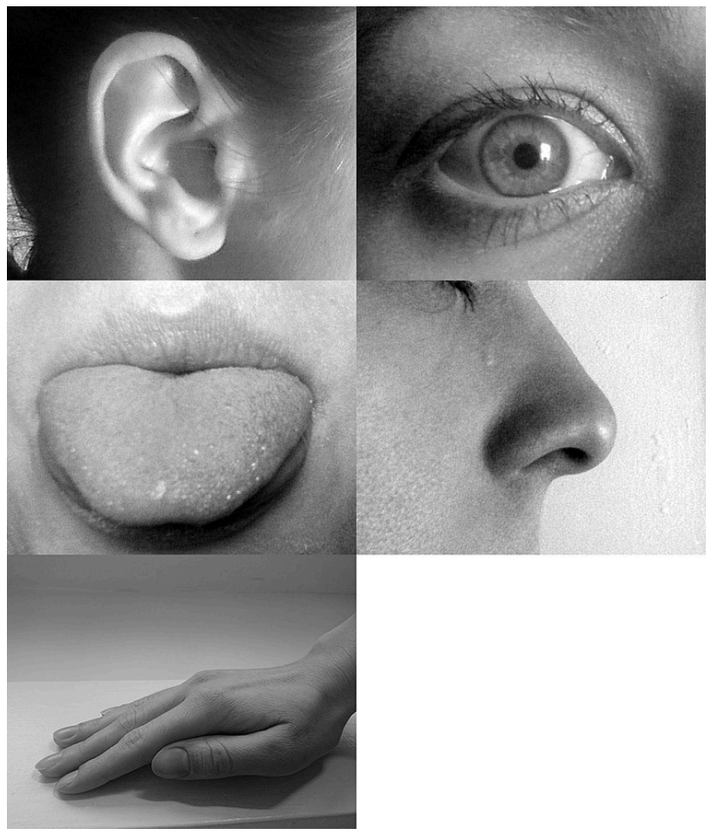
- What are examples for human?

Input/Output

I/O (stands for input/output)

- What are examples for human?

input: senses

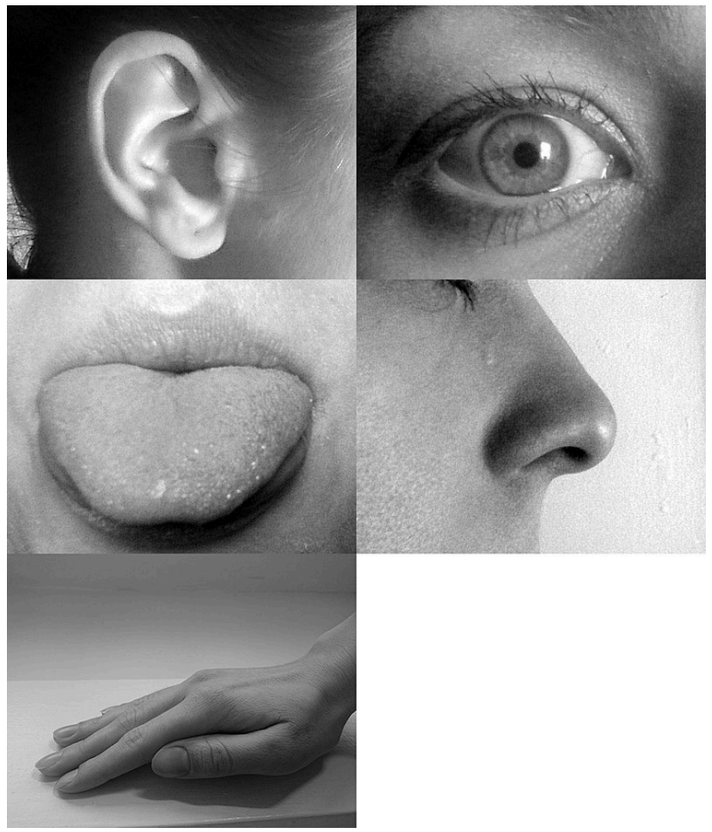


Input/Output

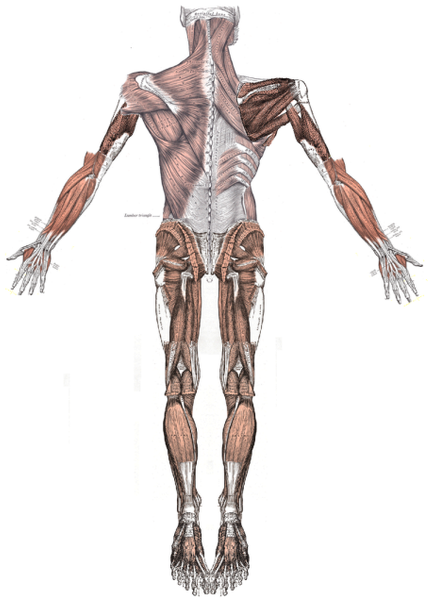
I/O (stands for input/output)

- What are examples for human?

input: senses



output: muscles, voice



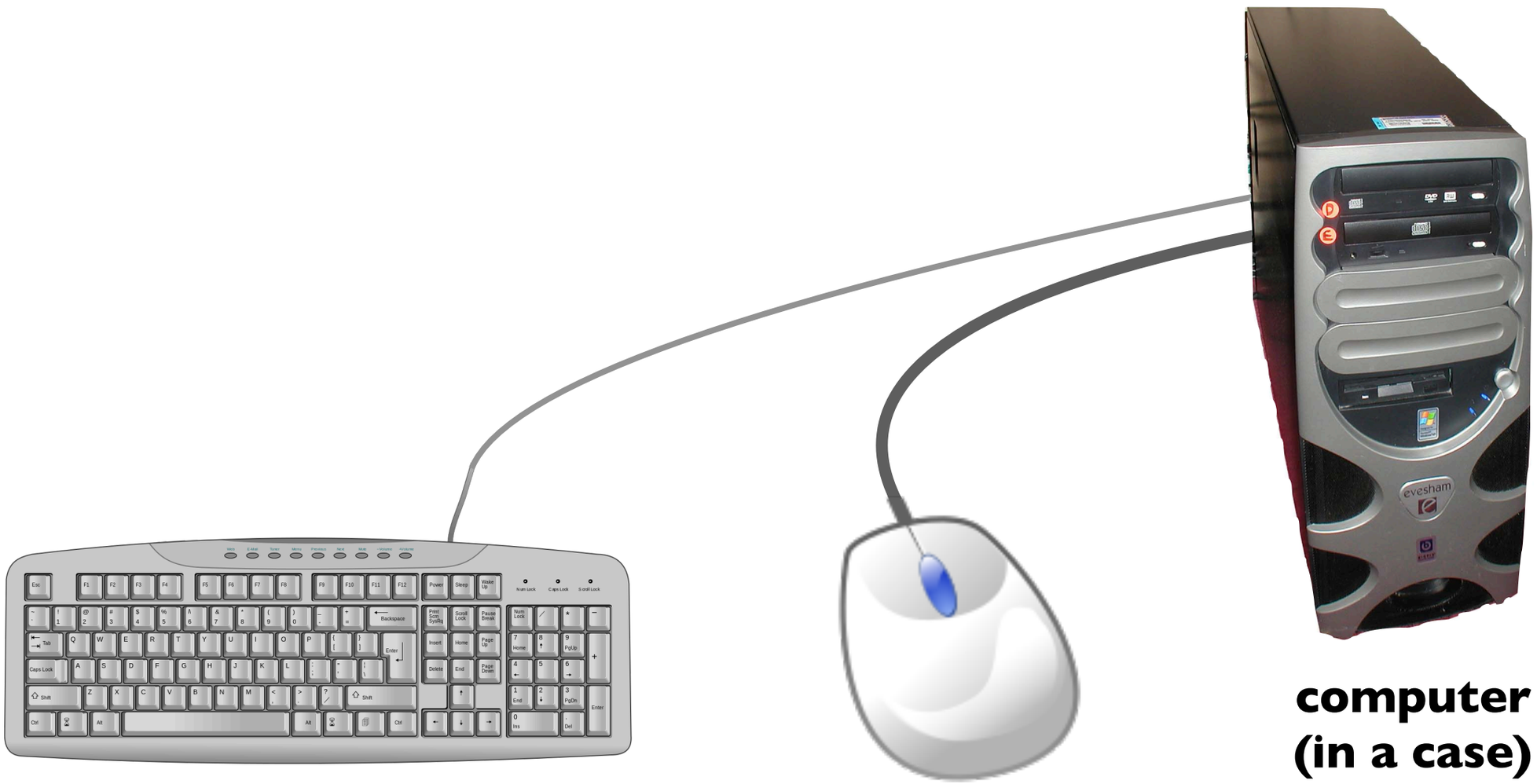
https://jasperproject.github.io/

Computer Input/Output

what are some common compute inputs?

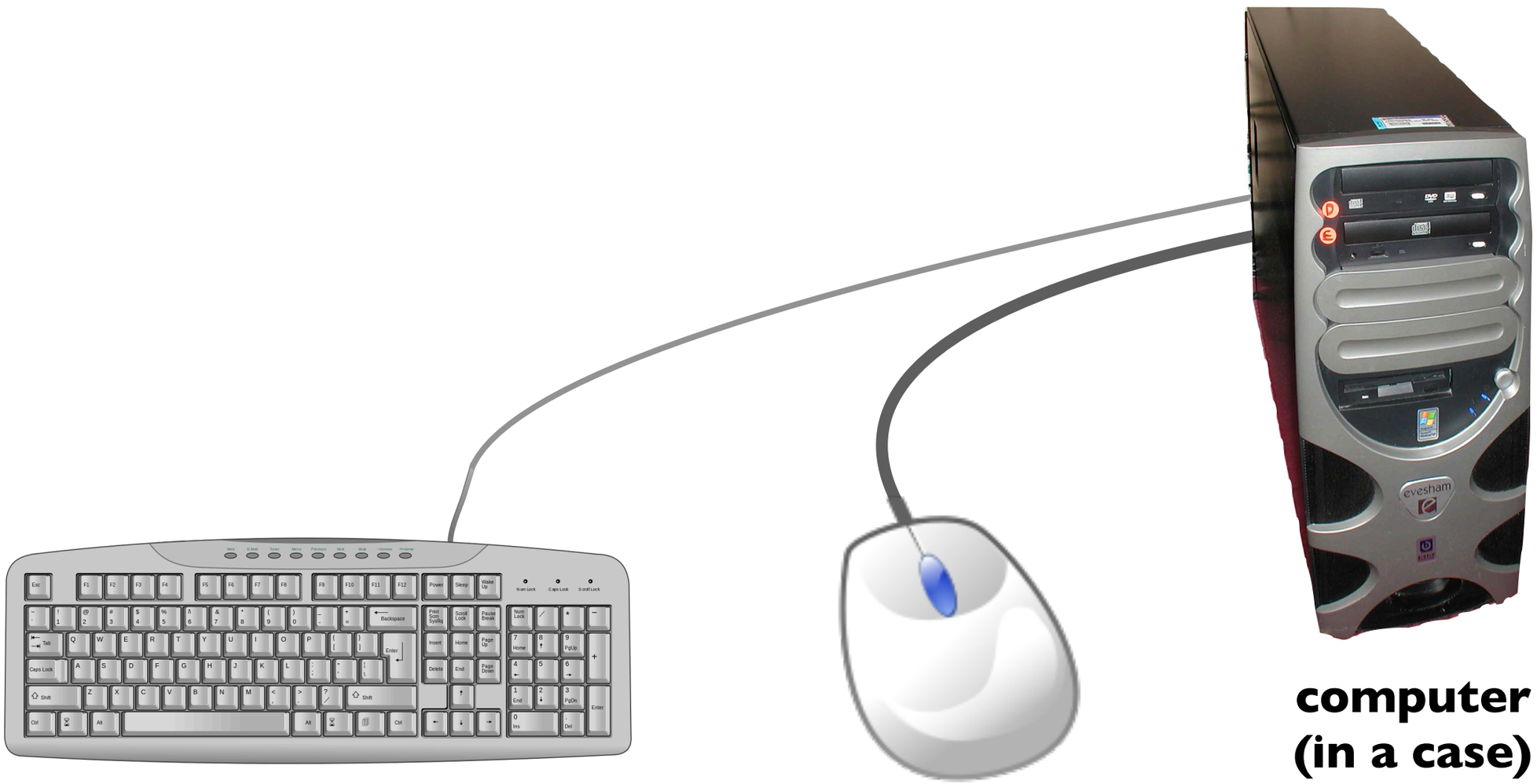
computer (in a case)

Computer Input/Output

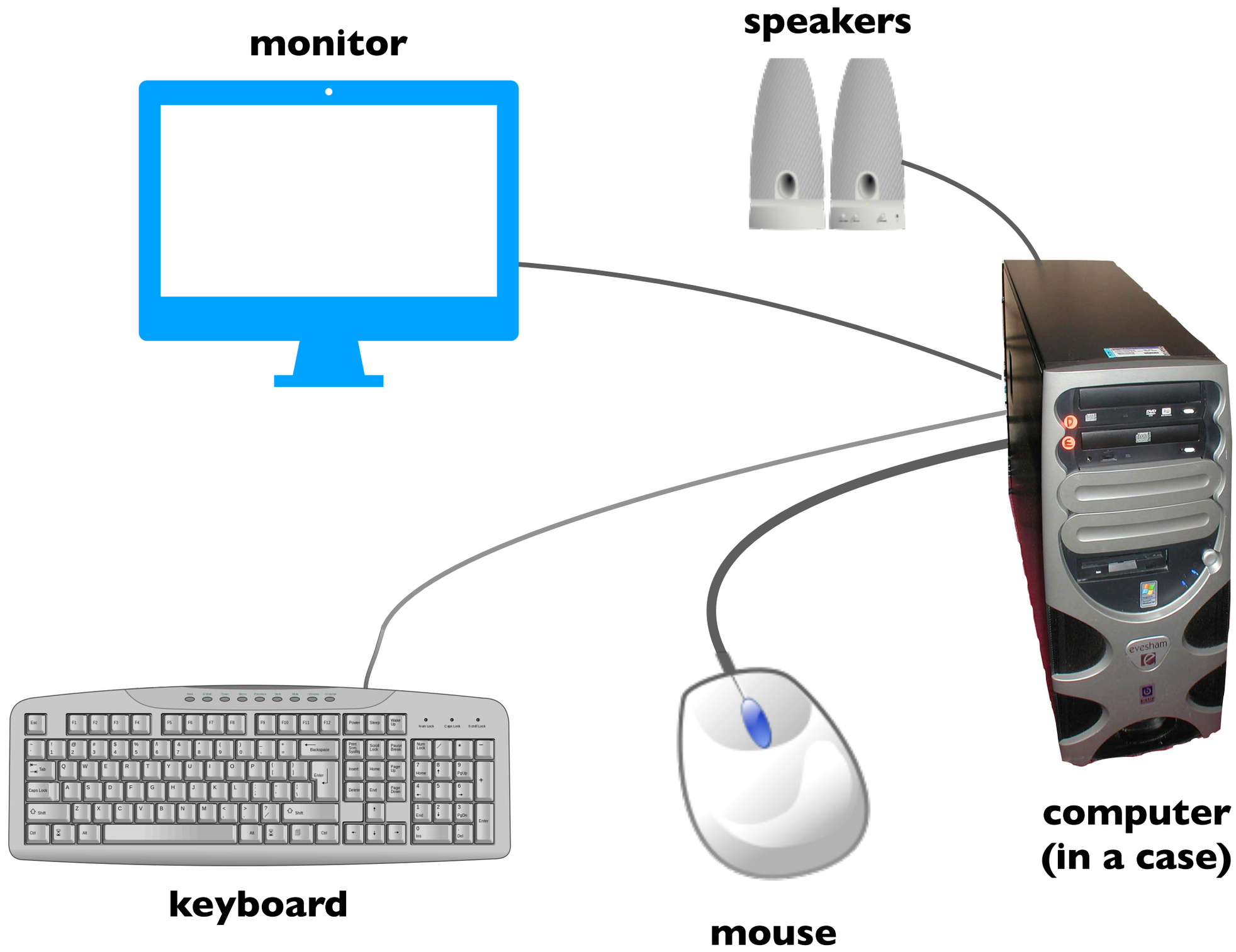


Computer Input/Output

what are some common compute outputs?



Computer Input/Output



Computer Input/Output

I/O devices attach via "ports" (e.g. USB) in back of computer

monitor

speakers

computer (in a case)

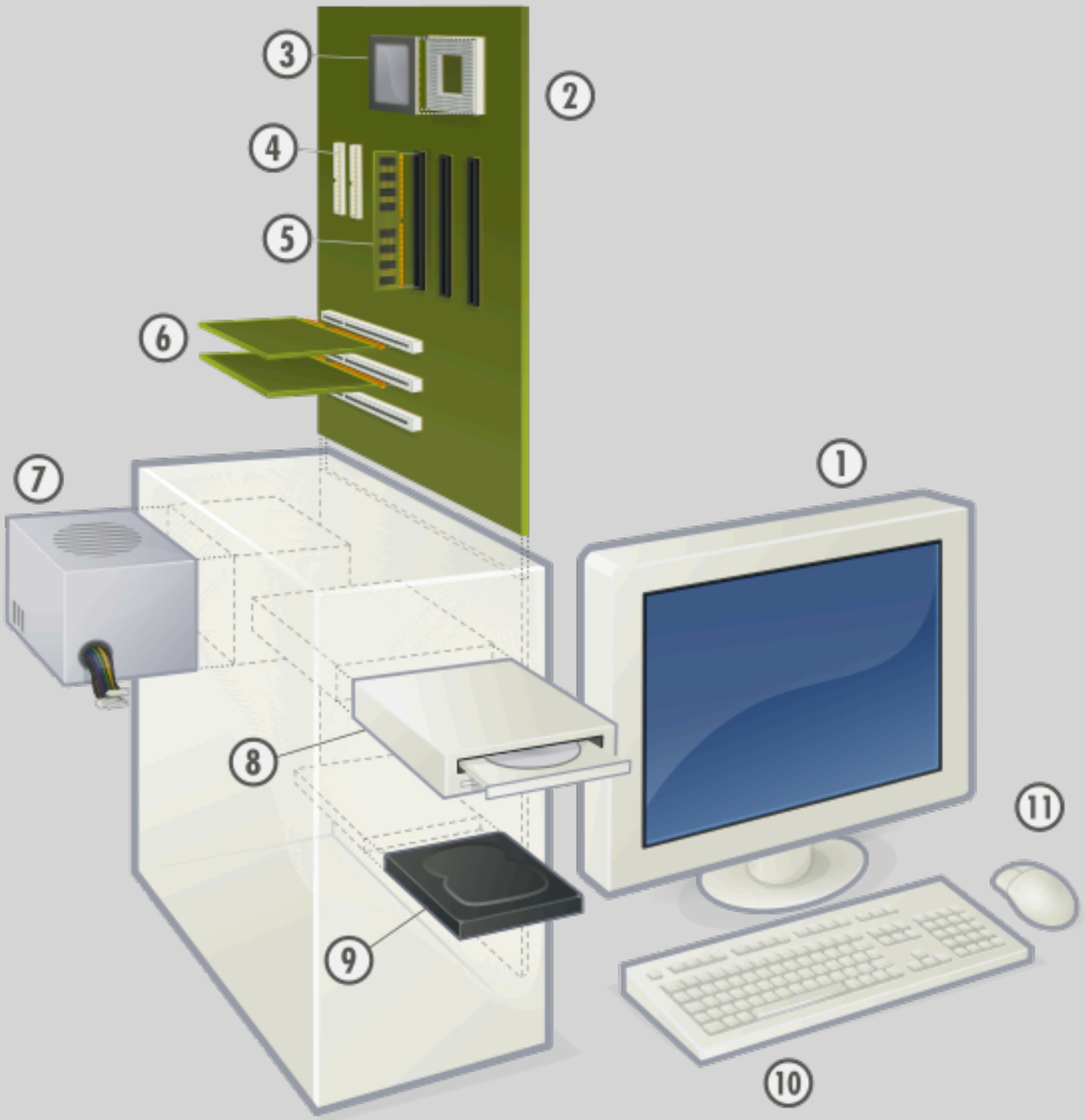
keyboard

mouse

Computer Input/Output

What's inside a computer case?

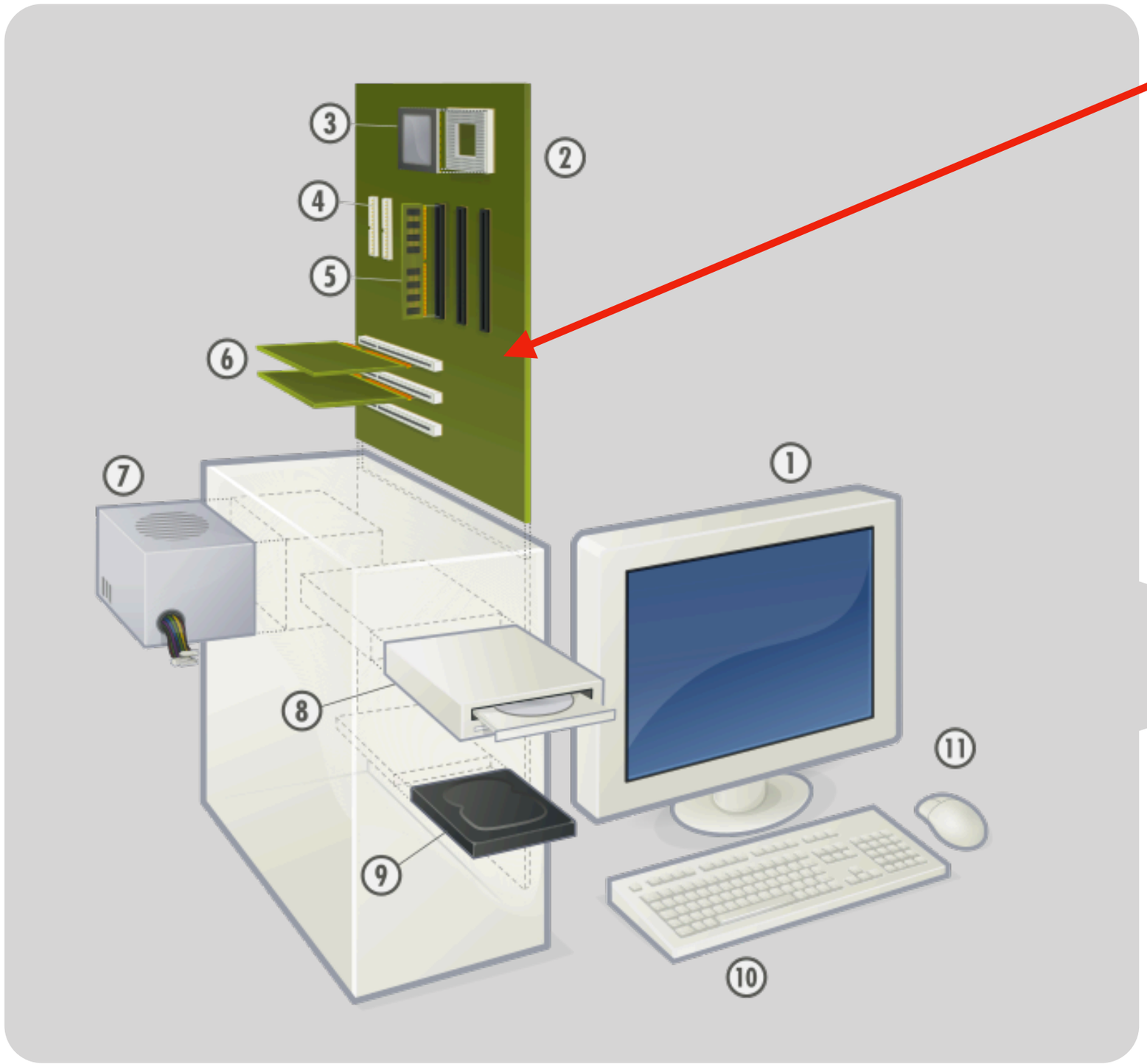
Computer Internals



https://commons.wikimedia.org/wiki/File:Personal_computer,_exploded_4.svg

Computer Internals

Motherboard: main circuit board to which other components connect, via sockets/slots



https://commons.wikimedia.org/wiki/File:Personal_computer,_exploded_4.svg

Today's Topics

Introductions

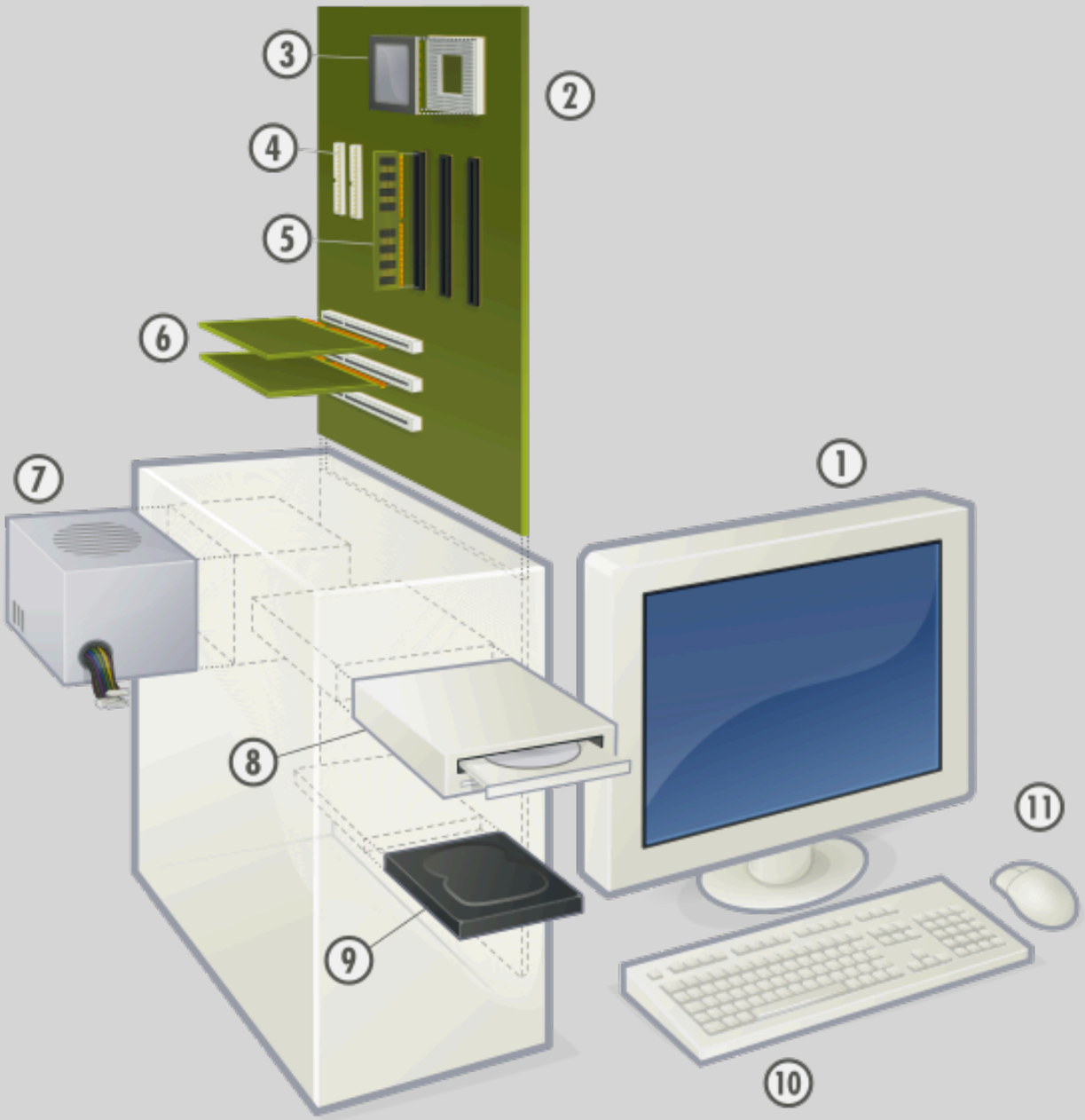
Course overview

Computer hardware basics

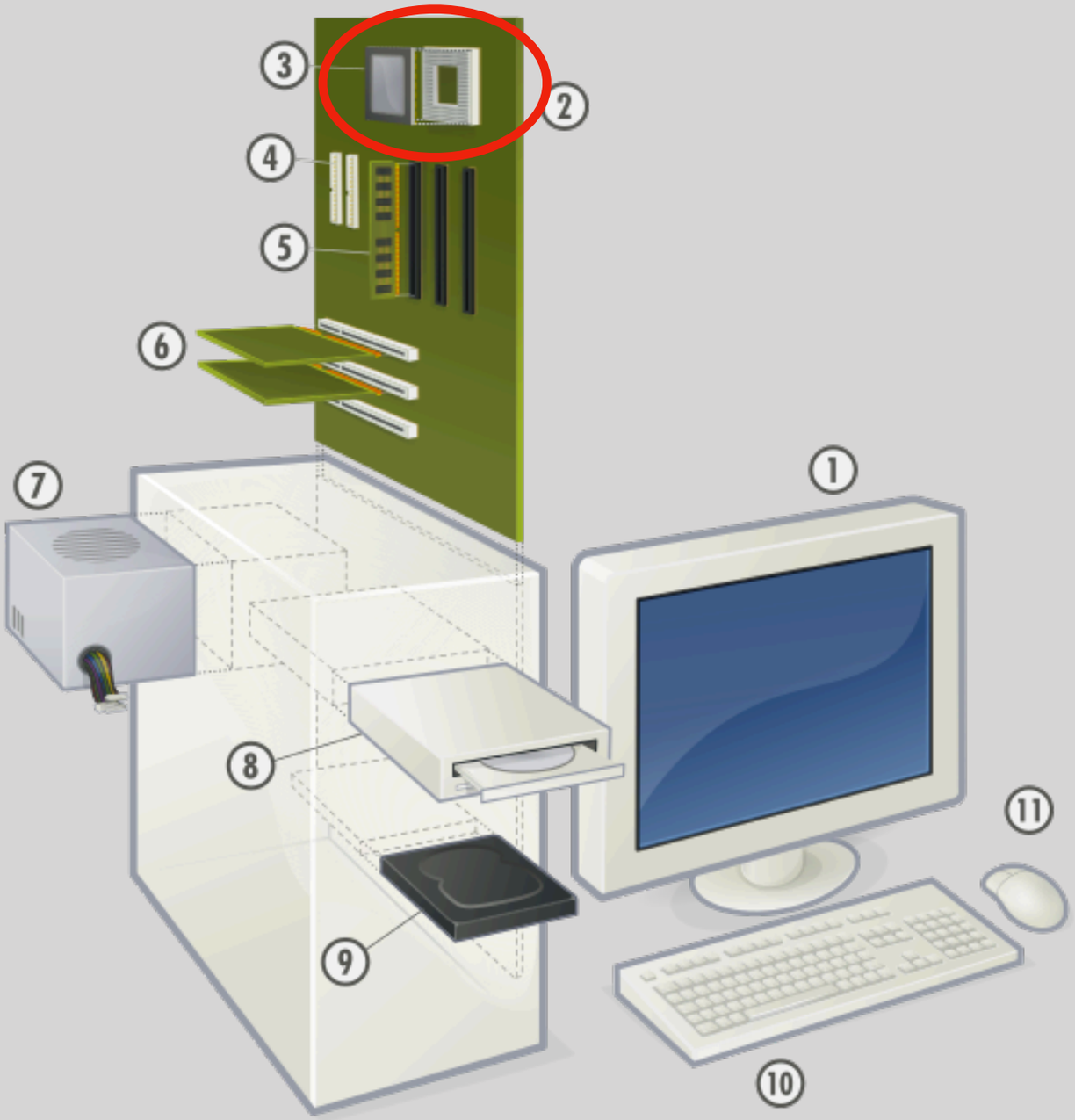
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Central Processing Unit (CPU)



Central Processing Unit (CPU)



CPU

Responsible for computation

- Runs code
- Performs addition, other math
- Compares numbers, text
- Receives input, sends output
- Some compare it to a “brain”

Runs on a clock

- Typically a couple GHz (i.e., billions of ticks per second)
- High-speed makes CPUs hot, require fans/cooling

Computers often have multiple CPUs

- Motherboard may have multiple sockets
- Single chip may contain multiple CPUs
- Allows computers to do more things simultaneously

Today's Topics

Introductions

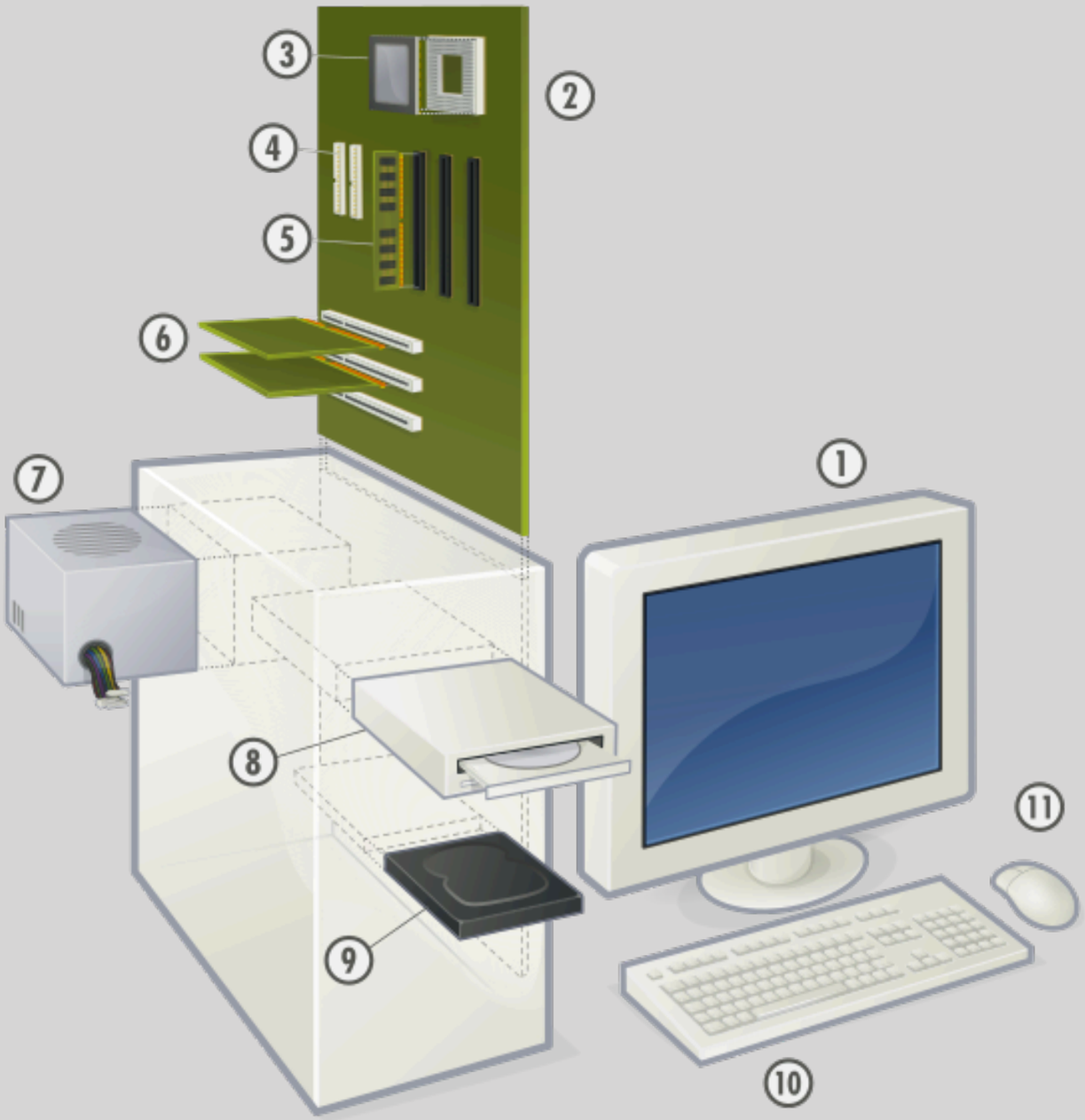
Course overview

Computer hardware basics

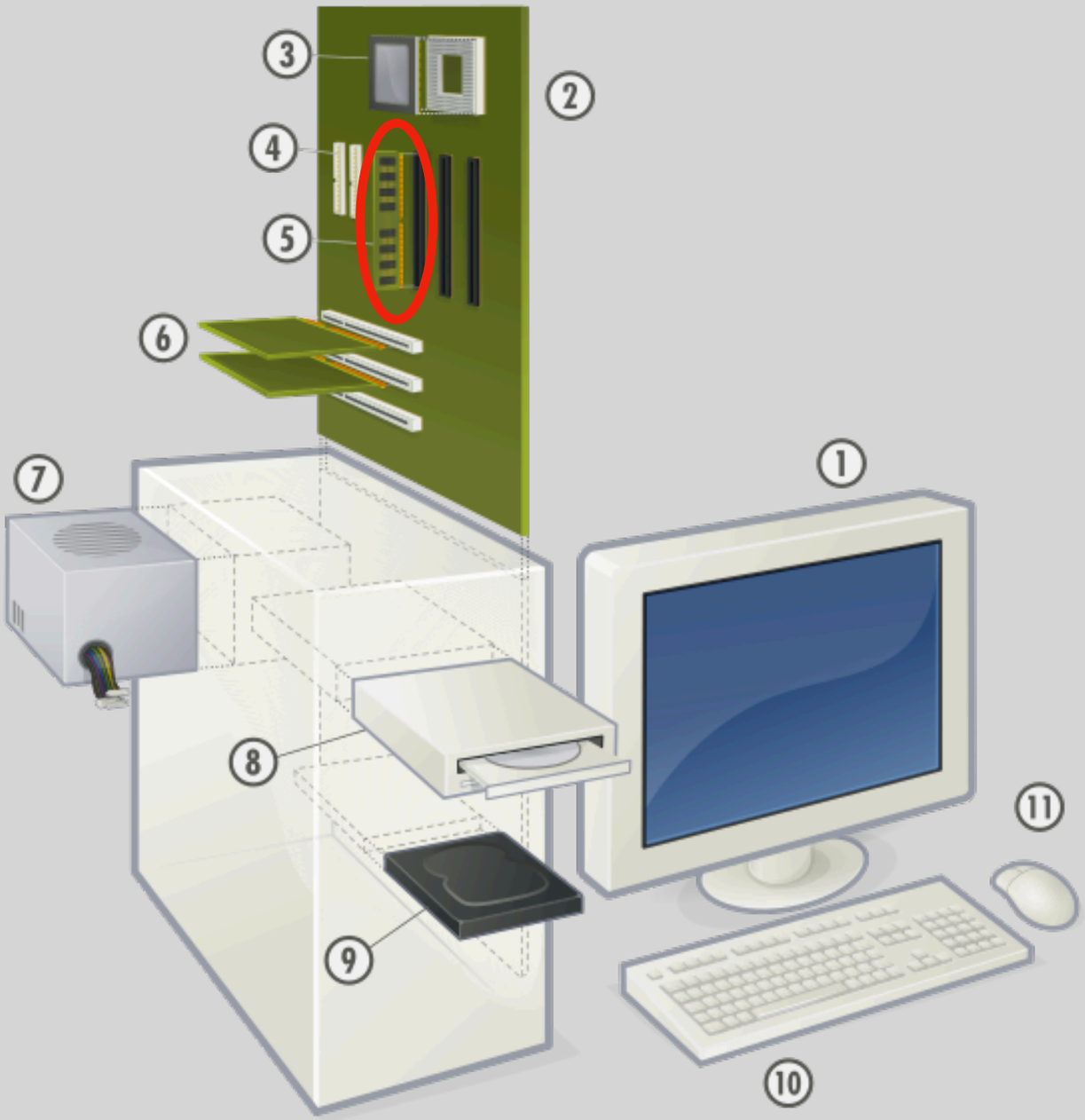
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Random Access Memory (RAM)



Random Access Memory (RAM)



Memory

Memory stores data for short term

- RAM is most common form today (don't worry about specifics)
- CPU sends data to/from memory
- Accessing it is very fast
- It is "volatile" — meaning you lose this data when you power off your computer
- You don't save "files" in memory, otherwise they would be gone!

Stores bytes of data

- One byte ≈ one letter
- The text "hello" requires 5 bytes
- Typical personal computer has few to tens of gigabytes (billion bytes) of memory

Today's Topics

Introductions

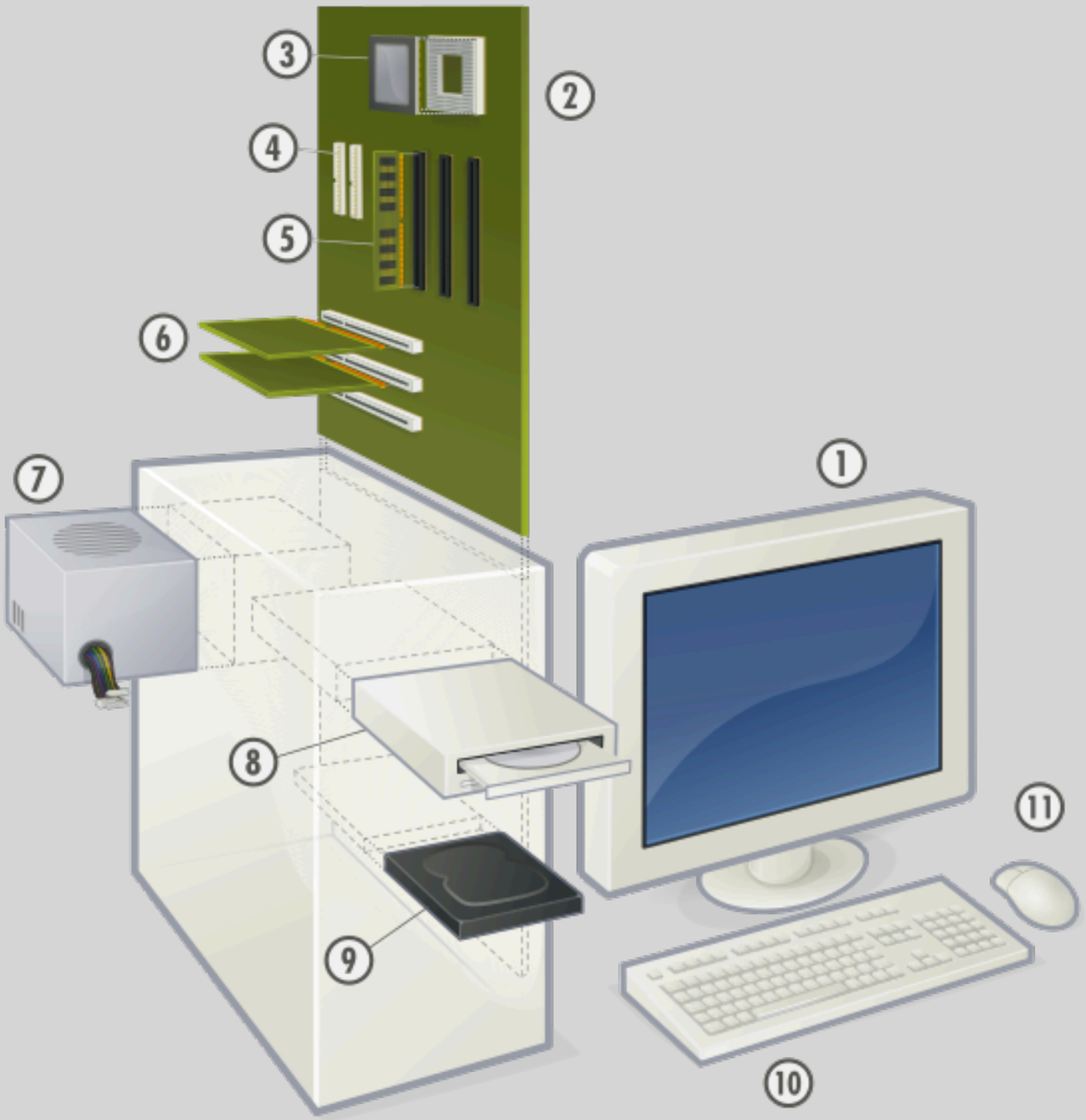
Course overview

Computer hardware basics

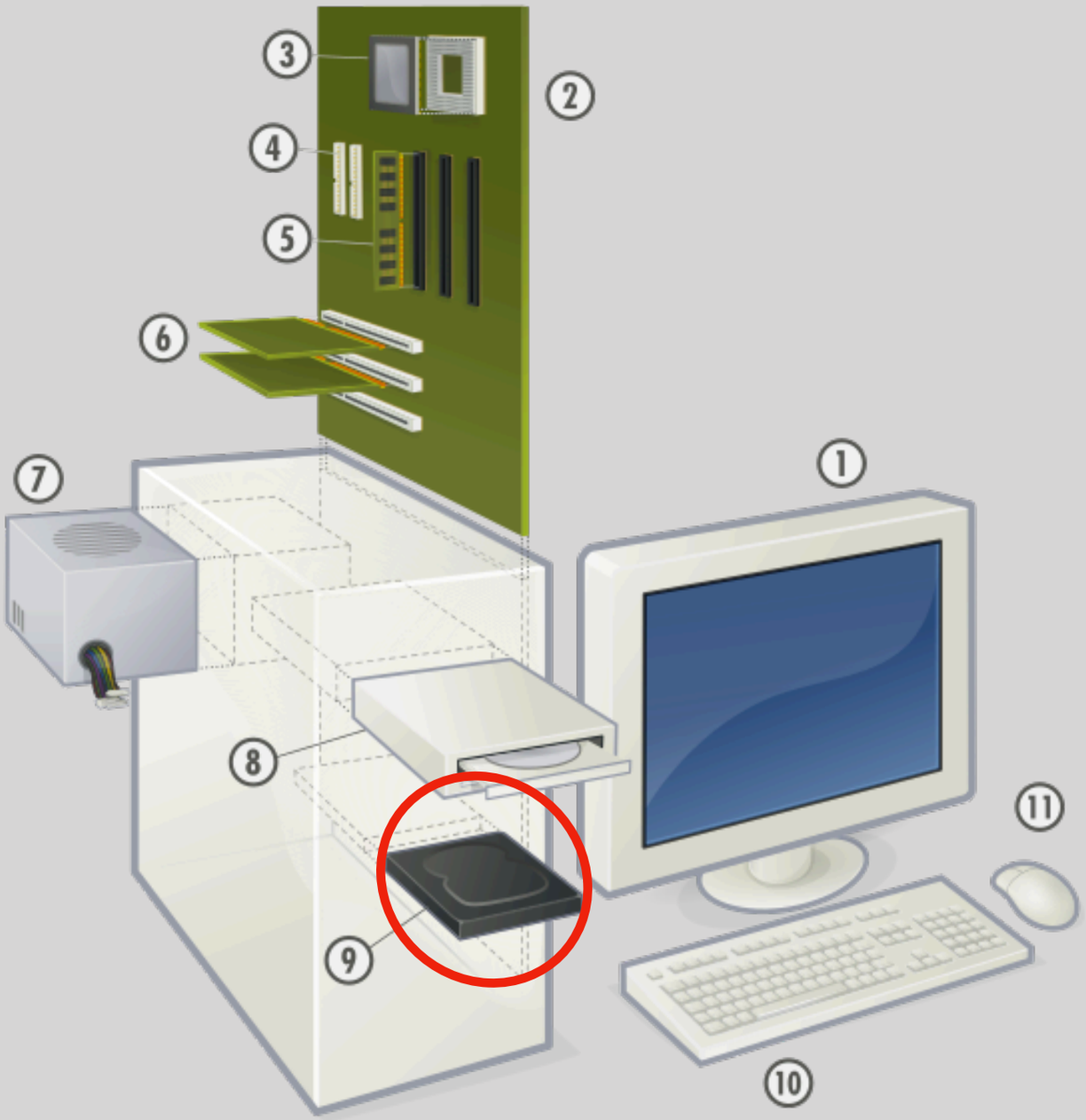
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Storage Drives



Storage Drives



Storage Drives

Two common devices

- HDD (hard disk drive), has moving parts, cheap, slow
- SSD (solid state drive), no moving parts, expensive, fast
- Both much slower than RAM...

Storage devices used to save data after power down

- Persistant medium, in contrast to volatile RAM
- Typical capacity: hundreds of gigabytes

When you make a directory/folder or save a file, that data is ultimately getting recorded to your storage device

- Sometimes computers save to RAM first, and only to the device later; power down cleanly to avoid losing your data!!!

Today's Topics

Introductions

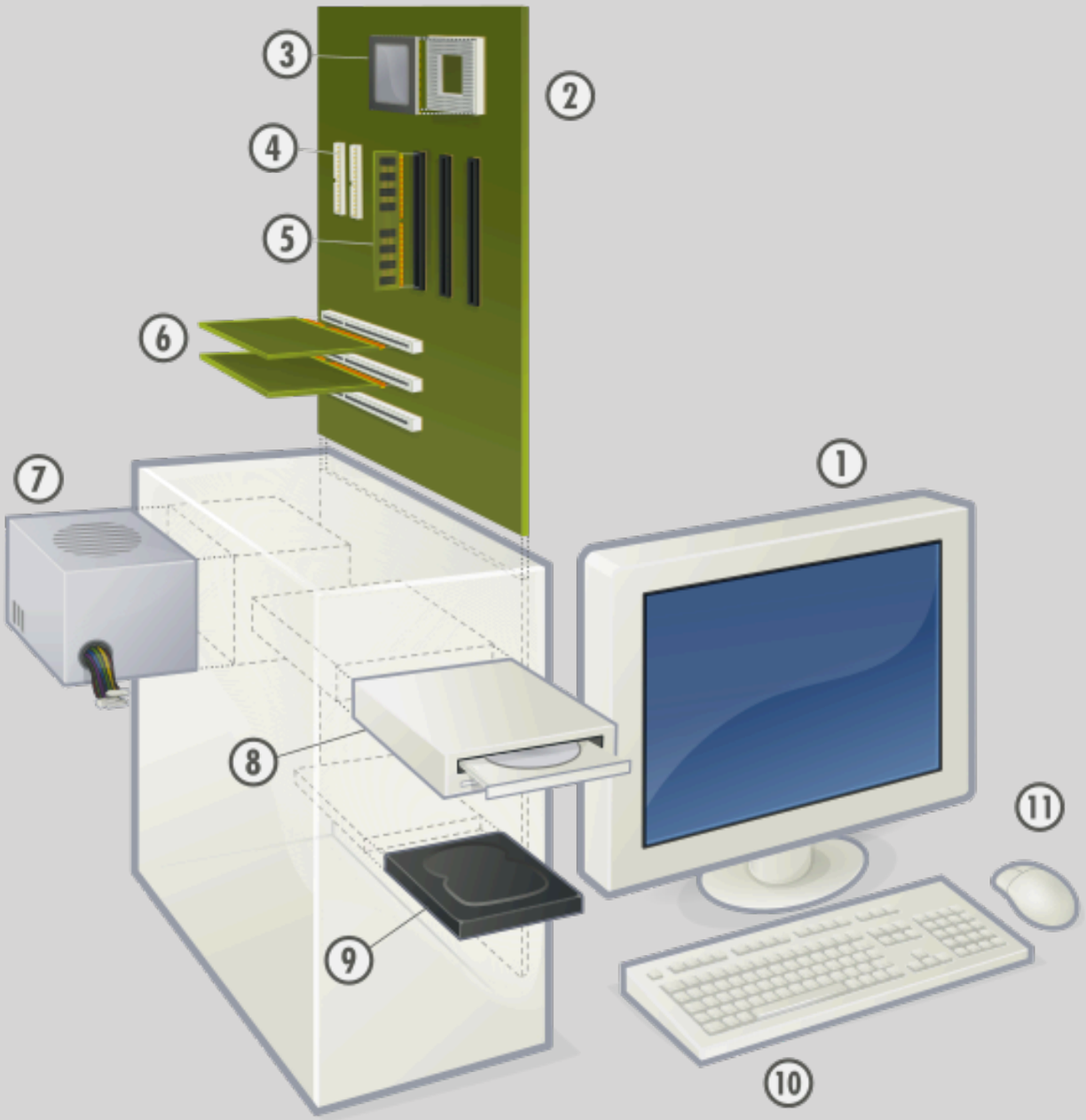
Course overview

Computer hardware basics

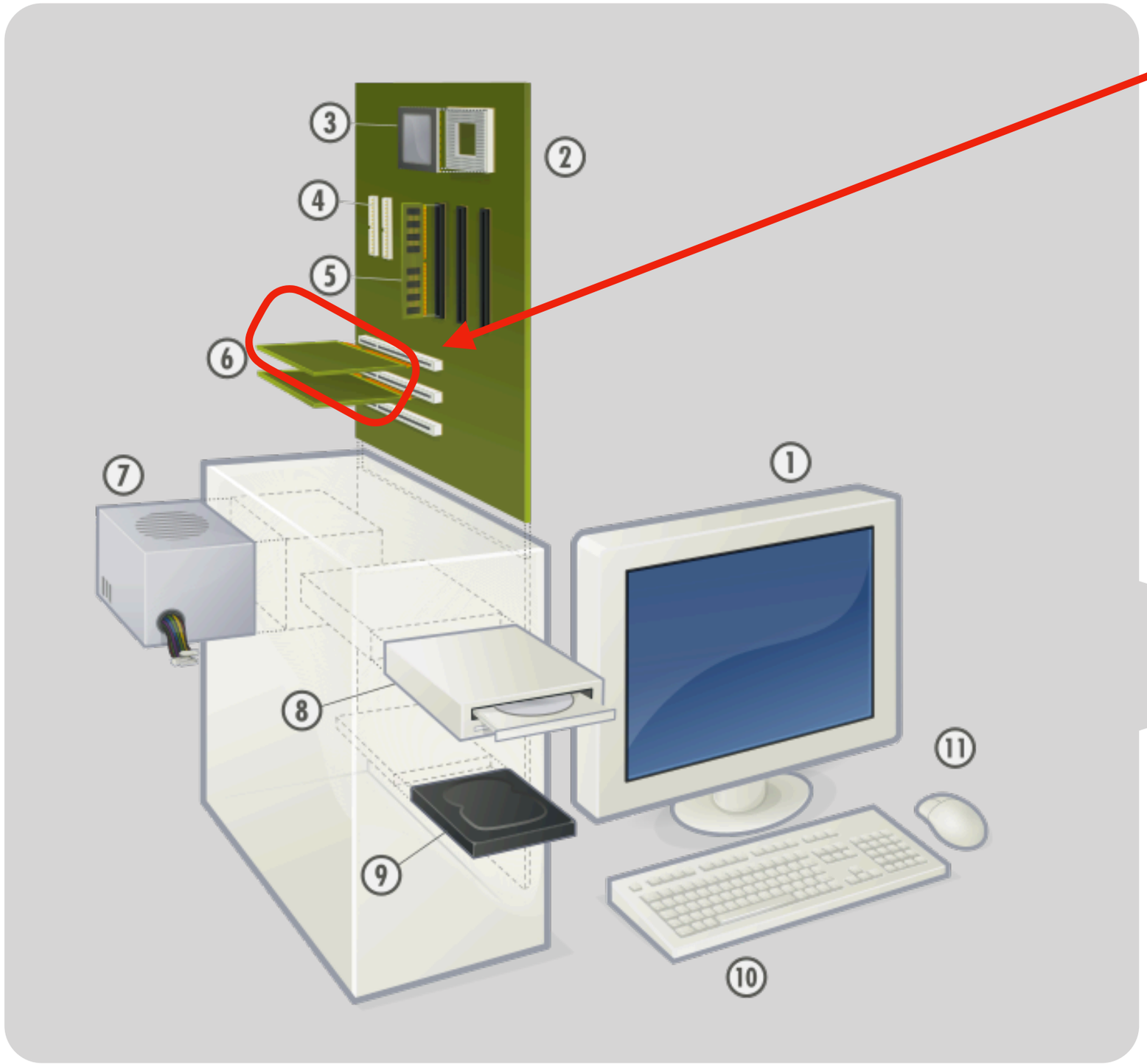
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Network Interfaces

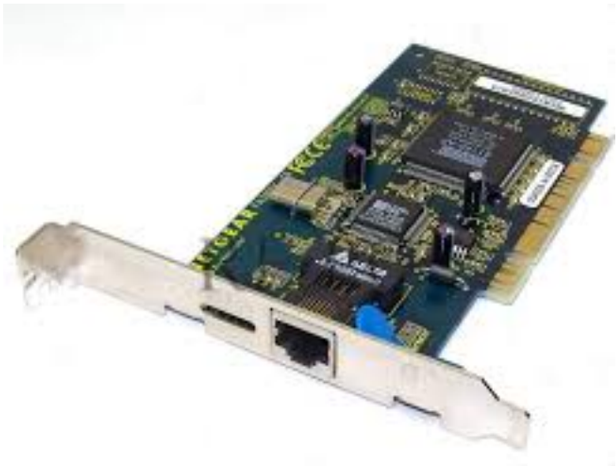


Network Interfaces



Network: often based on extension card or built into the motherboard itself

Networking



NIC (Network Interface Controller)

- Provides computer communication to other computers, and the Internet

Wired vs. Wireless

- Wired ethernet is common for cable-based connection
- Wi-Fi is common for radio-based wireless connection

Terminology

- **Server**: program/computer that runs, waiting for for incoming requests, to which it responds
- **Client**: program/compute that sends requests to a server

Today's Topics

Introductions

Course overview

Computer hardware basics

Website

Course Website

Shared website (sections 1+2):
https://tyler.caraza-harter.com/cs301/fall19/schedule.html

Walk through…

Next steps...

- take the "Who are You?" survey:
  https://tyler.caraza-harter.com/cs301/fall19/surveys.html

- read syllabus carefully:
  https://tyler.caraza-harter.com/cs301/fall19/syllabus.html

- setup Python on your computer (with videos) and do Lab-P1:
  https://github.com/tylerharter/cs301-projects/tree/master/fall19/lab-p1

- start P1 (Project 1), due next Wed:
  https://github.com/tylerharter/cs301-projects/tree/master/fall19/p1