CS220 / CS319

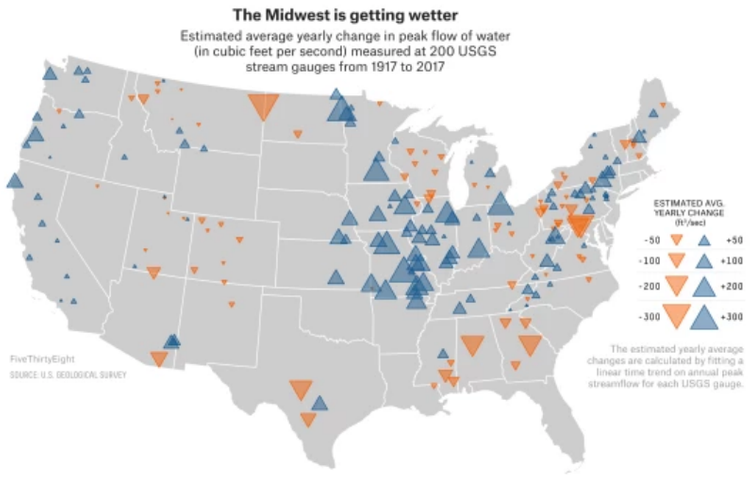
Introduction

Meena Syamkumar
Andy Kuemmel
Alexi Brooks

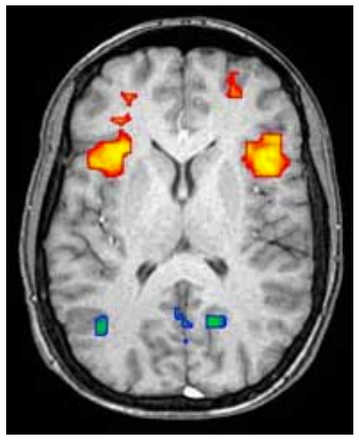
Welcome to Data Science Programming I!

Data is exploding in many fields

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

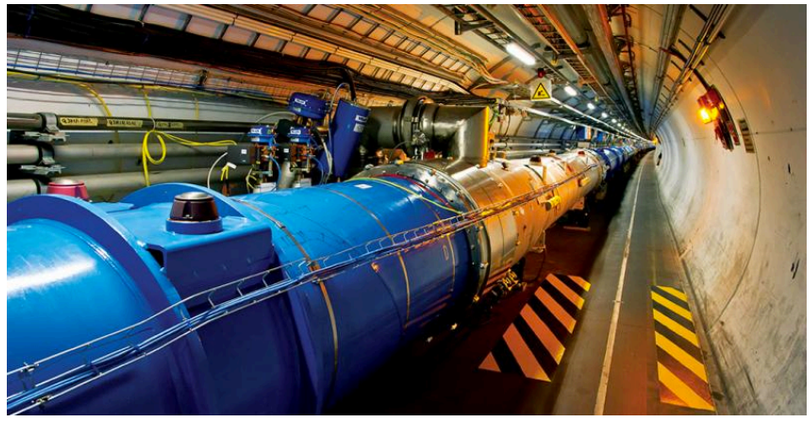


https://fivethirtyeight.com/features/the-midwest-is-getting-drenched-and-its-causing-big-problems/

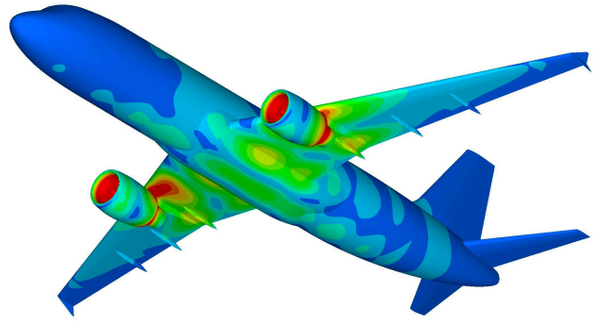


https://en.wikipedia.org/wiki/Neuroimaging

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 Science 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 Science Programming I!

CS 220 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

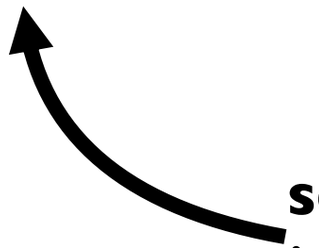
Approach 2: machine computation

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

Welcome to Data Science Programming I!

CS 220 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 220?

Typical intro CS

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

vs

CS 220 approach

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

Why CS 220?

50 Best Jobs in America for 2020

	Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1	Front End Engineer	$105,240	3.9/5	13,122	View Jobs
#2	Java Developer	$83,589	3.9/5	16,136	View Jobs
#3	Data Scientist	$107,801	4.0/5	6,542	View Jobs
#4	Product Manager	$117,713	3.8/5	12,173	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?

Meena Syamkumar

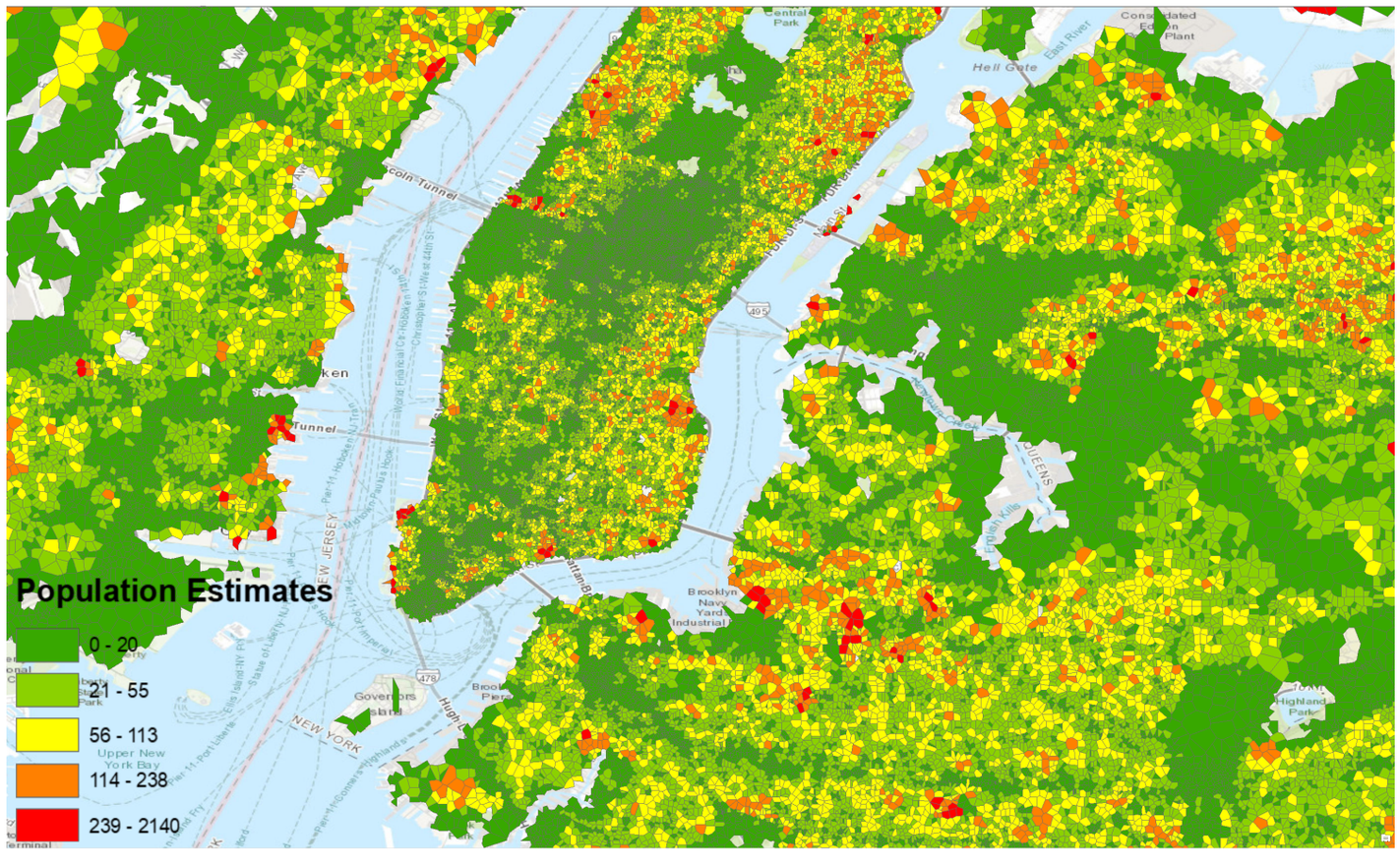
- Email: ms@cs.wisc.edu
- Please call me "Meena"

Industry and Teaching experience

- Citrix, Cisco, and Microsoft
- CS220 / CS319 (S'20, F'20), CS367 (Summer'17), guest lectures in CS640, CS740

Passion: Running

Research: Internet measurements



Who am I?

Andy Kuemmel
- Email: kuemmel@wisc.edu
- Please call me "Andy"

Work Experience
- College Board – AP Exam Committee
- AP Computer Science teacher
- Microsoft curriculum writer
- UW Madison Faculty Associate

Interests

Men's Barbershop Chorus

Running

Thanksgiving Day 10k • 5k run/walk

2021 BERBEE DERBY

Powering TECHNOLOGY EDUCATION FOUNDATION

Crazylegs CLASSIC

BRAT FEST BUN RUN

Who am I?

Alexi Brooks
- Email: albrooks@cs.wisc.edu
- Please call me "Alexi"

Industry and Teaching experience
- Epic
- CS367, CS301, CS300, CS368

Research: Learning in social media

Fun fact:

When I started my research, the term "data science" was not yet in common use at UW!

Who am I?

Tyler Caraza-Harter

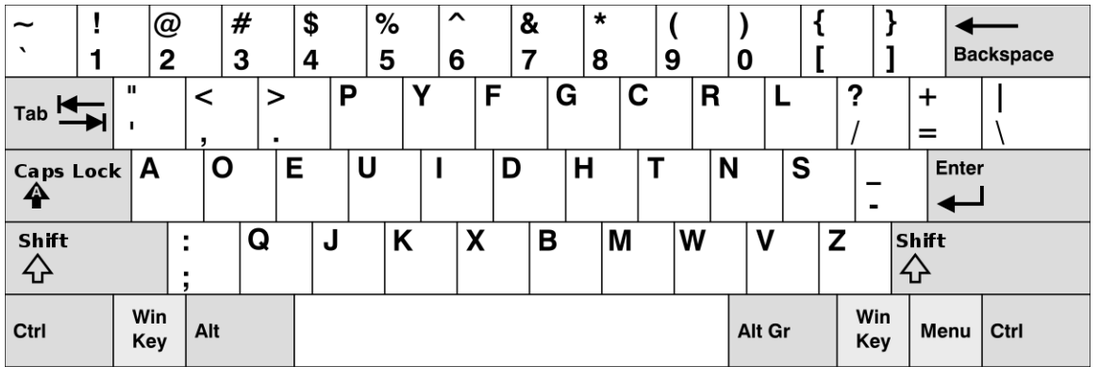
- Long time Badger
- Email: tylerharter@gmail.com
- Just call me "Tyler"

Industry experience

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

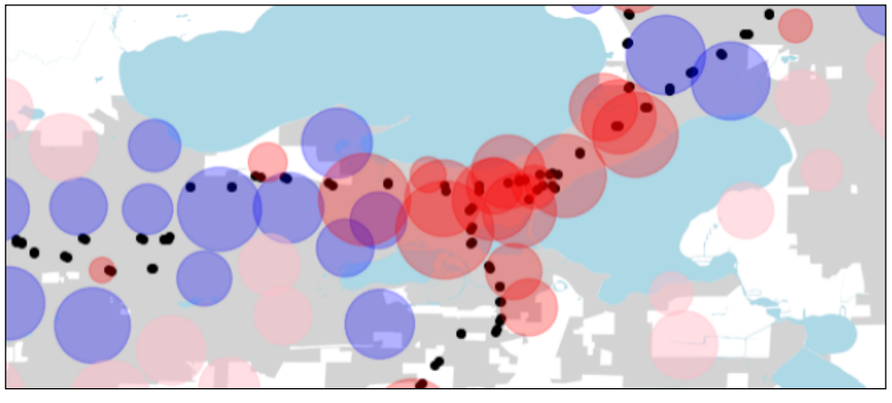
interests

typing efficiently



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

civic "hacking"



Plot by Jin Woo Lee (previous CS 301 student)

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

Who are You? Survey (counts for communication)

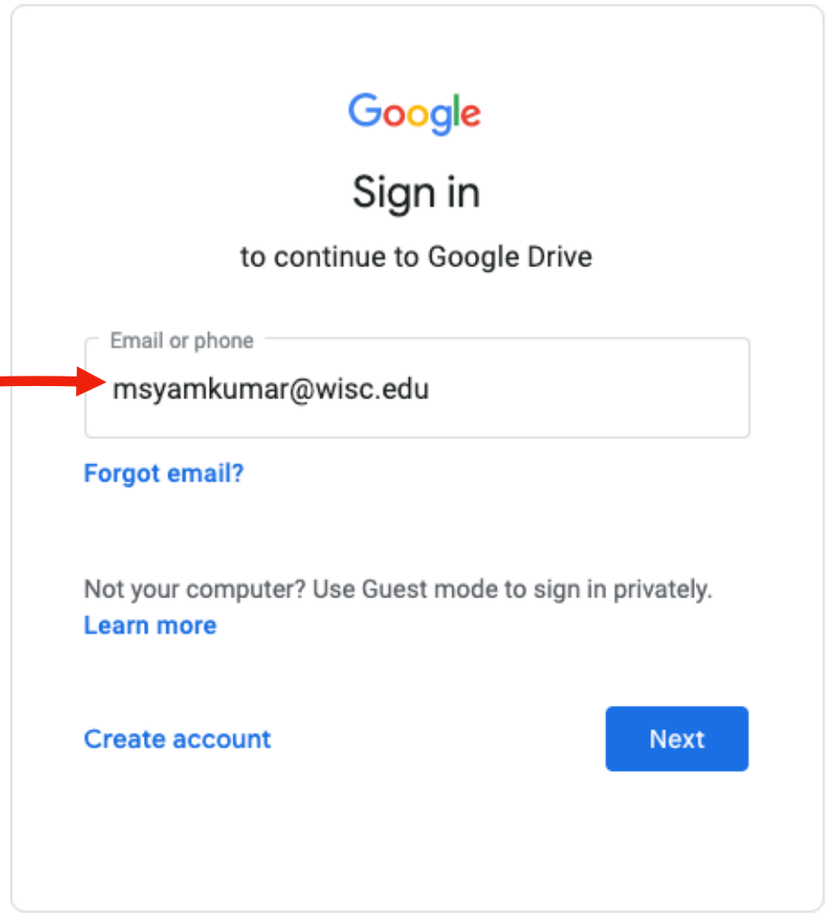
Please help us get to know you (not anonymous):

https://forms.gle/9owhe4A4TCjxmm966

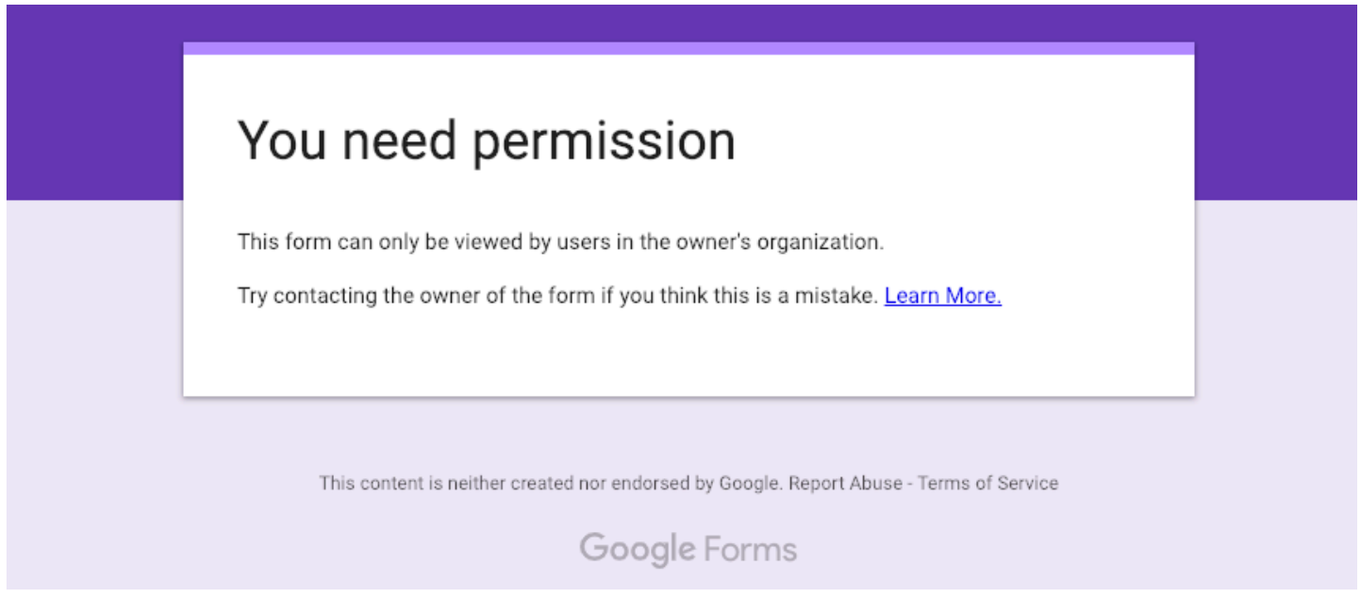
Purposes:

- gauge class interest/experience
- study course trends for internal purposes

be sure to use your campus email!

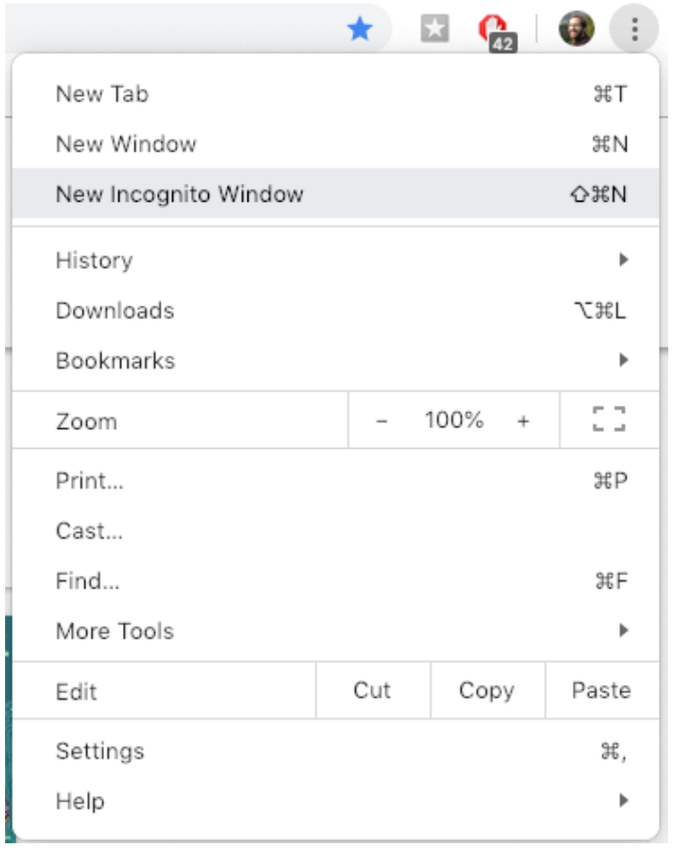


Survey: Common Technical Issues



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)



Today's Topics

Introductions

Course overview

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

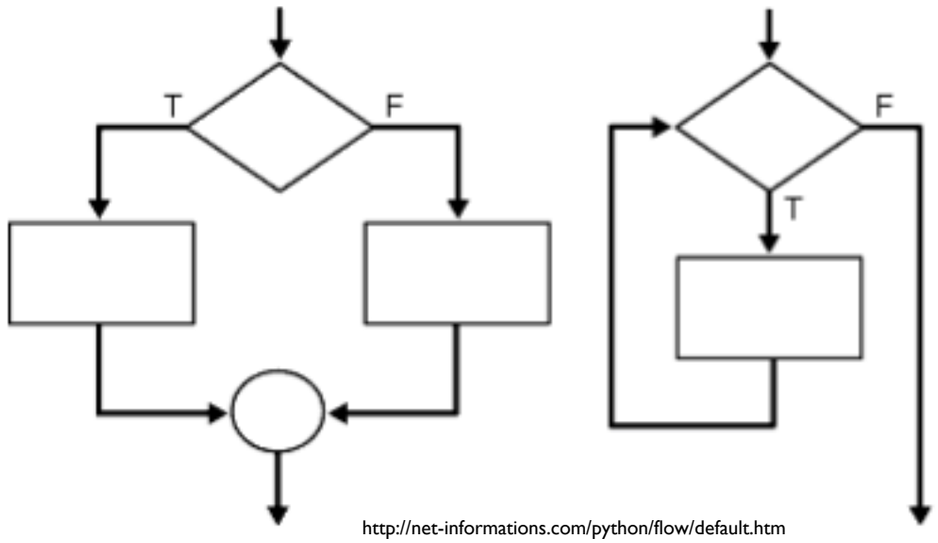
Computer hardware basics

Website

220 Topics

Part 1: Control Flow

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



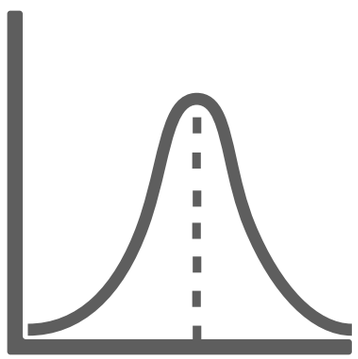
http://net-informations.com/python/flow/default.htm

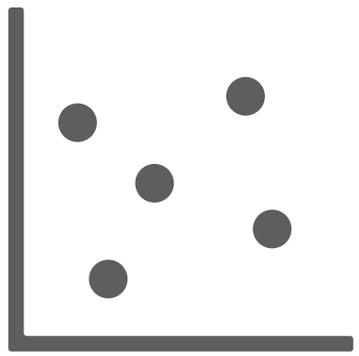
Part 2: State

- 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

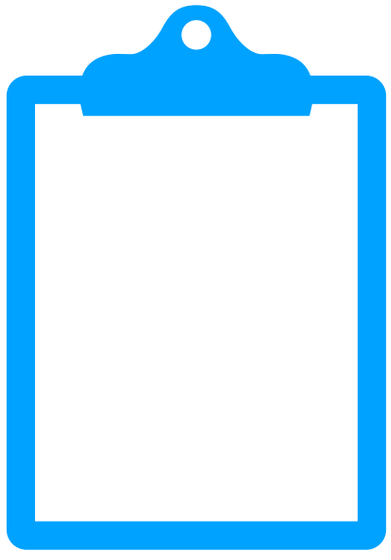
- Topics
- Lecture
- Lab
- Readings
- Class communication
- Grades
- Projects
- Exams & quizzes

Computer hardware basics

Website

Lecture Style (pre-recorded + live-streamed)

general concepts

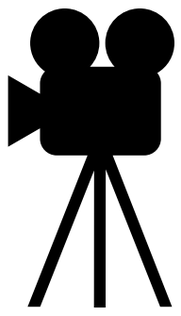


worksheet practice

live coding

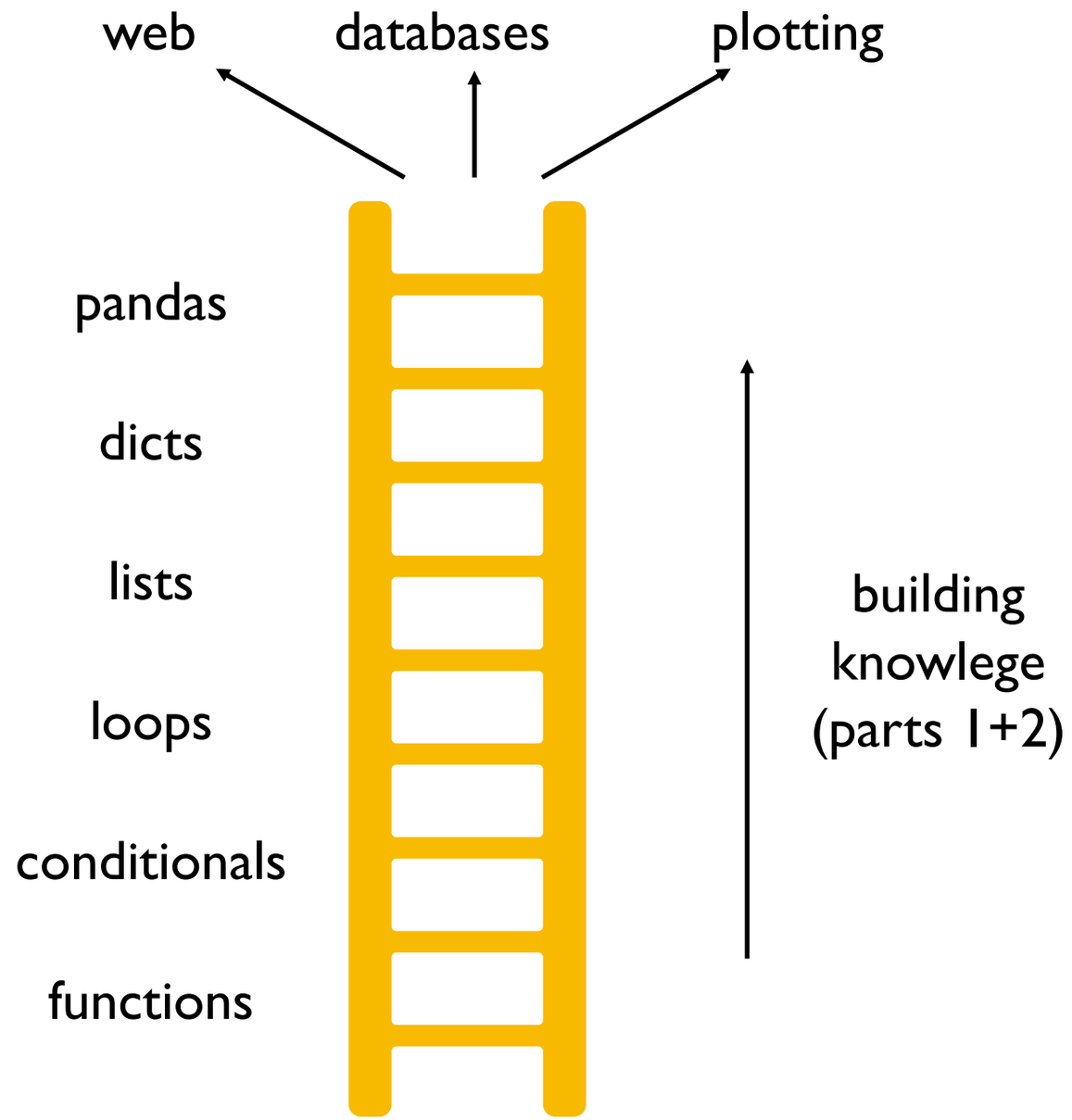
Your role

- do readings before or after
- I love to get questions, ask me during the live-stream + Q/A sessions



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 & quizzes

Computer hardware basics

Website

Labs

Format

- 75 minutes on Thu or Fri, leave when you're done
- led by TA / peer mentor or self-guided, not graded
- lab document will be posted each week
- do the lab before starting the project!

People

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

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 & quizzes

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 220 Notes

Course Notes

- 220 instructors
- Mostly for data science part of class

Today's Topics

Introductions

Course overview

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

Computer hardware basics

Website

Communication is CS 220

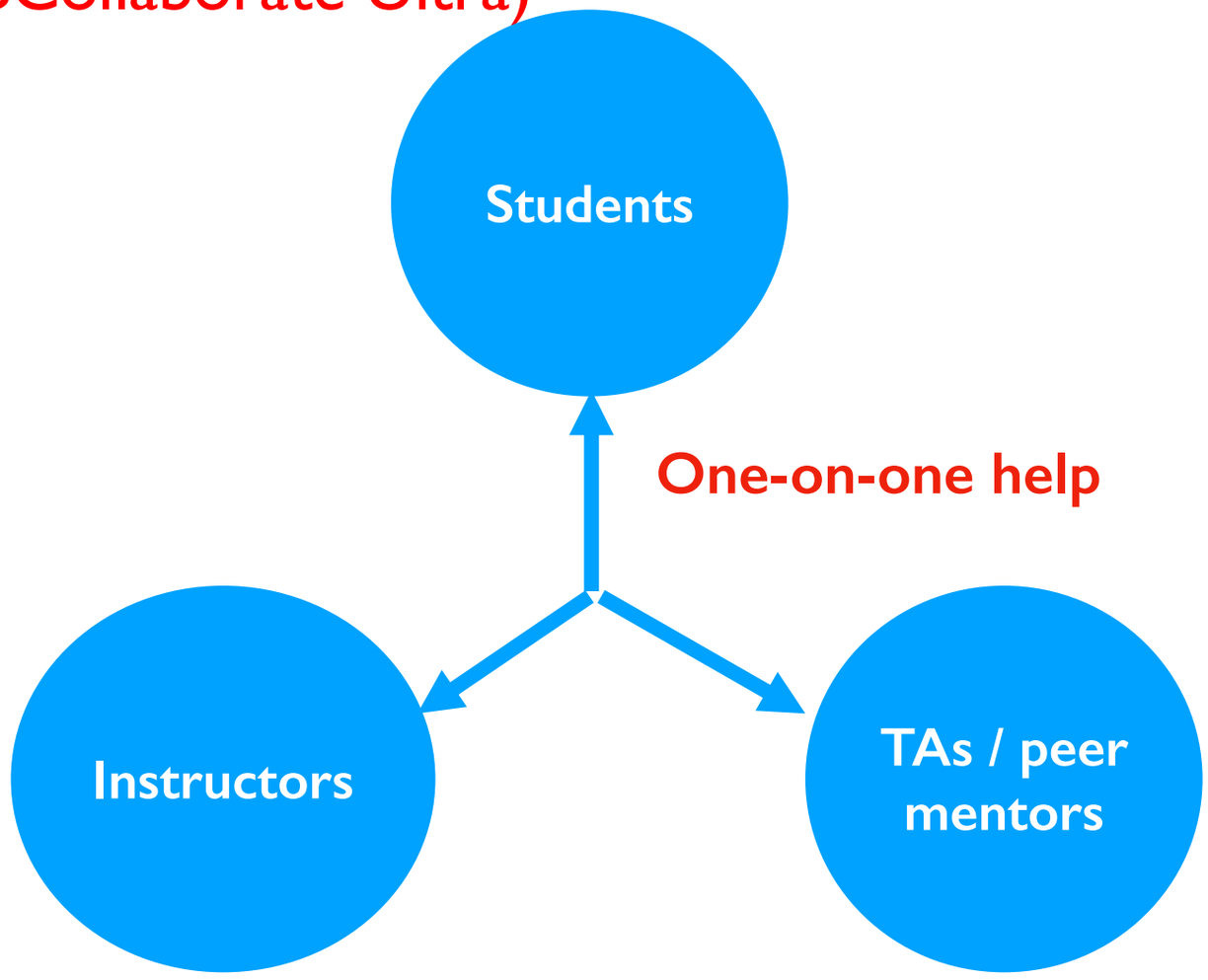
Good communication is critical for a class of this size

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

Communication tools

- Office hours (queuing system + BBCollaborate Ultra)
- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas

Read: How can I find help? page on course website.



Communication is CS 220

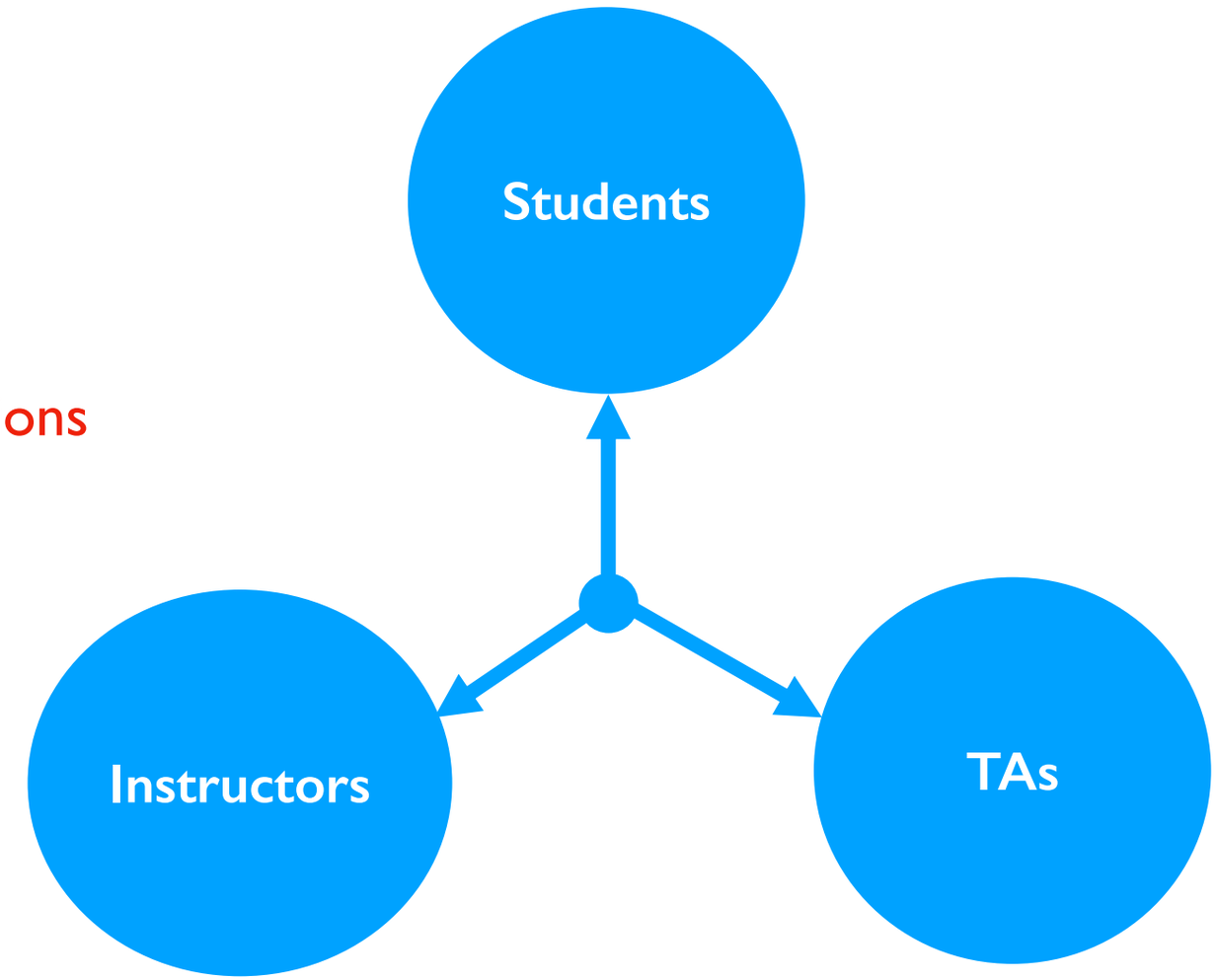
Communication tools

- Office hours (queuing system + BBCollaborate Ultra)
- 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

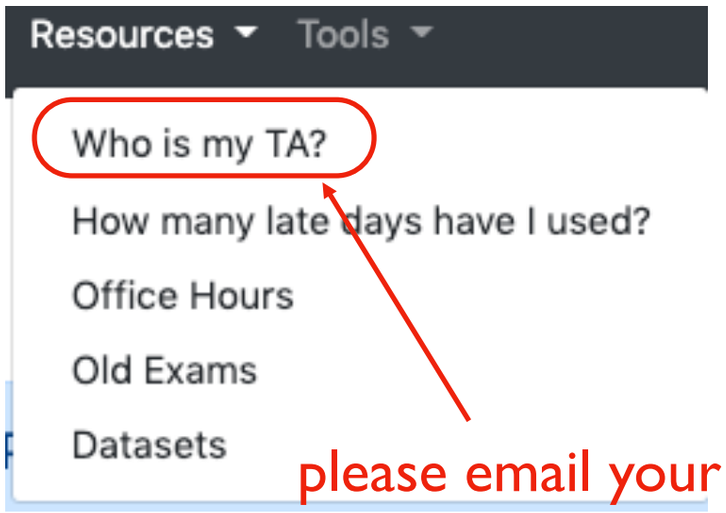
Note: we'll keep a pinned post of current office hours here



Communication is CS 220

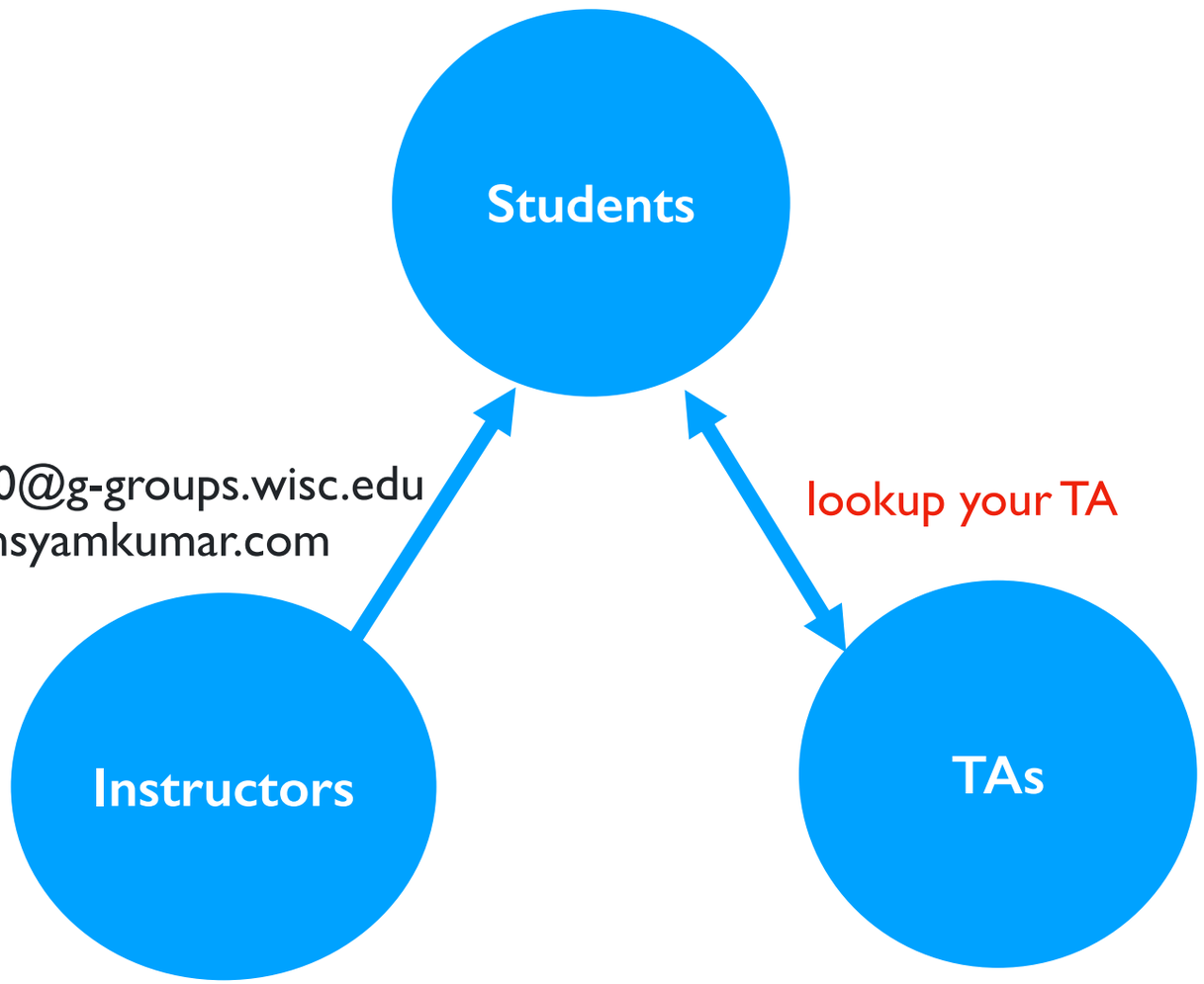
Communication tools

- Office hours (queuing system + BBCollaborate Ultra)
- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas



please email your assigned TA

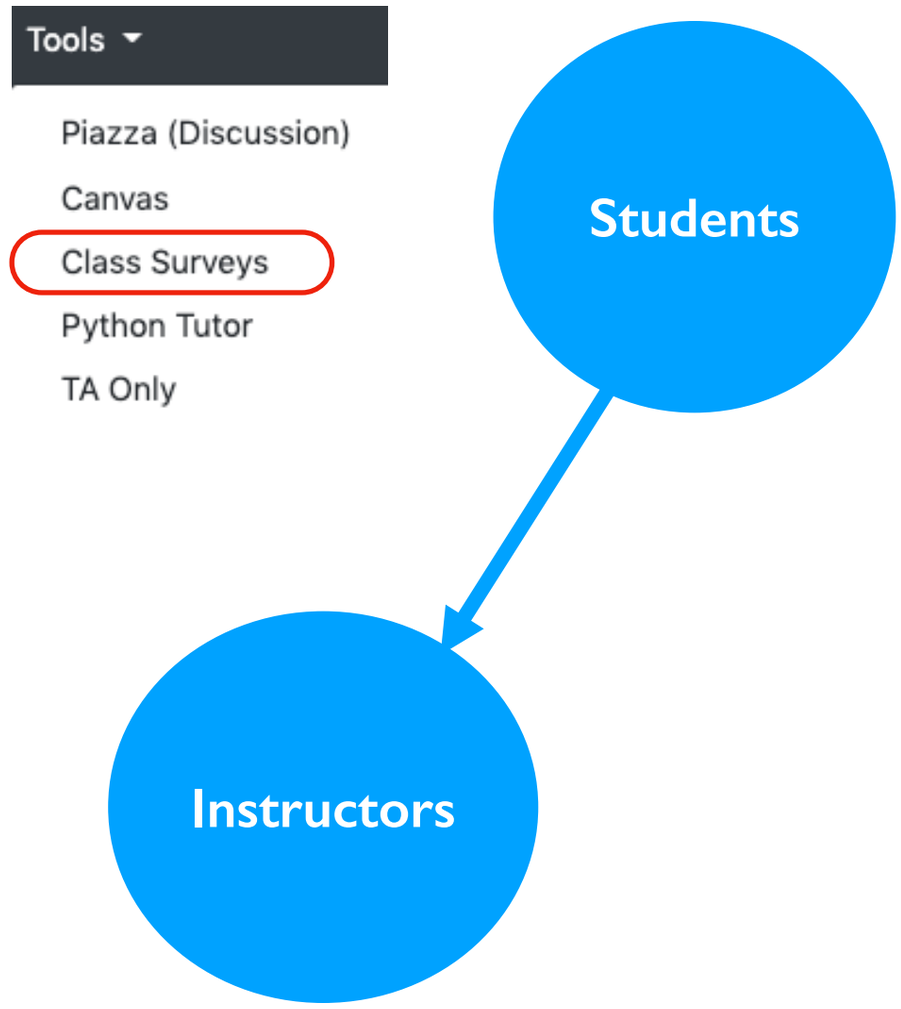
CC instructor at
ms@cs.wisc.edu
kuemmel@wisc.edu
albrooks@cs.wisc.edu
if you don't get a response within 48 hours.



Communication is CS 220

Communication tools

- Office hours (queuing system + BBCollaborate Ultra)
- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas



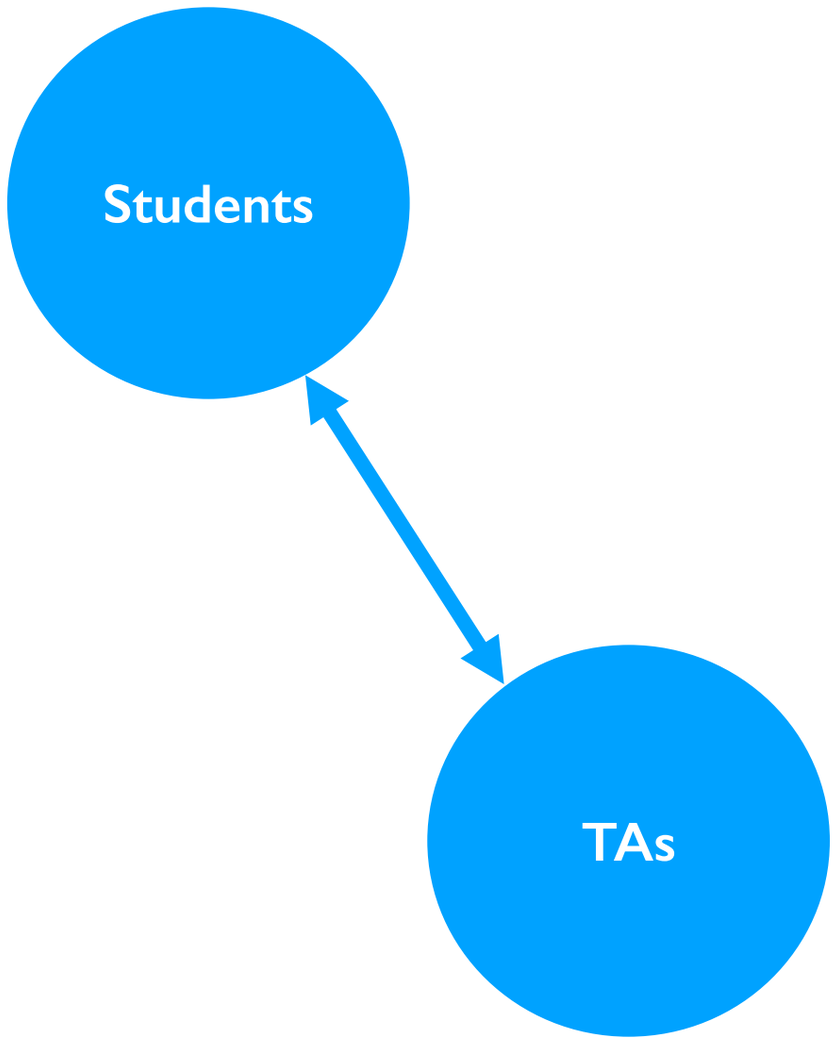
2. Feedback Form. If you have any issues with the class or suggestions for improvement, please let us know sooner rather than later; we may be able to make changes more rapidly than you might imagine. This is optionally anonymous, but it's always nice to know who you are (sometimes it makes sense to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

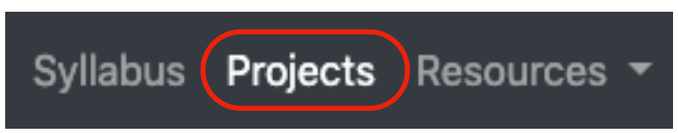
4. Thank You! Has a TA or mentor provided exceptional help, during office hours, Shelf hours, lab, etc? Thank them by filling out this form, and I'll pass along the feedback.

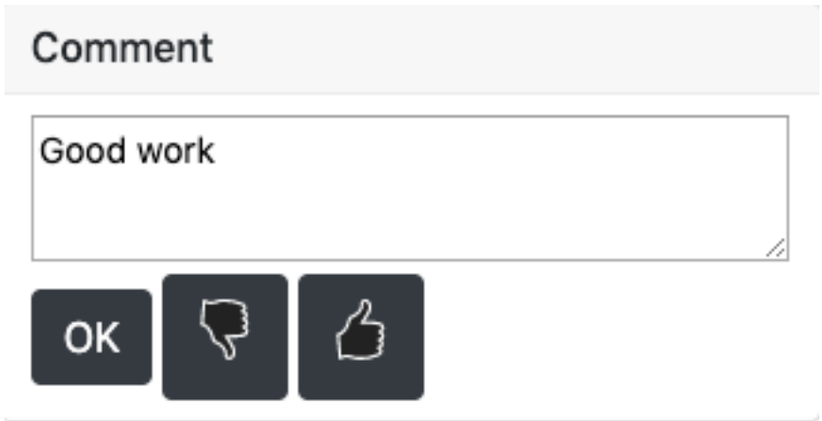
Communication is CS 220

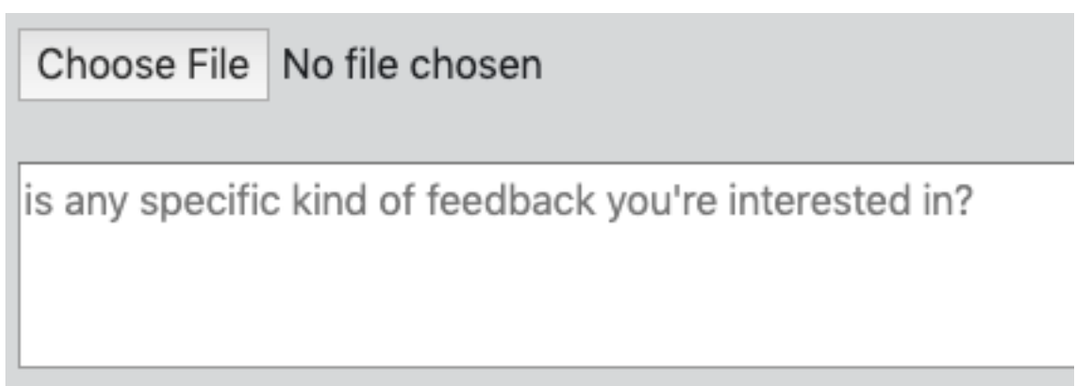
Communication tools

- Office hours (queuing system + BBCollaborate Ultra)
- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas







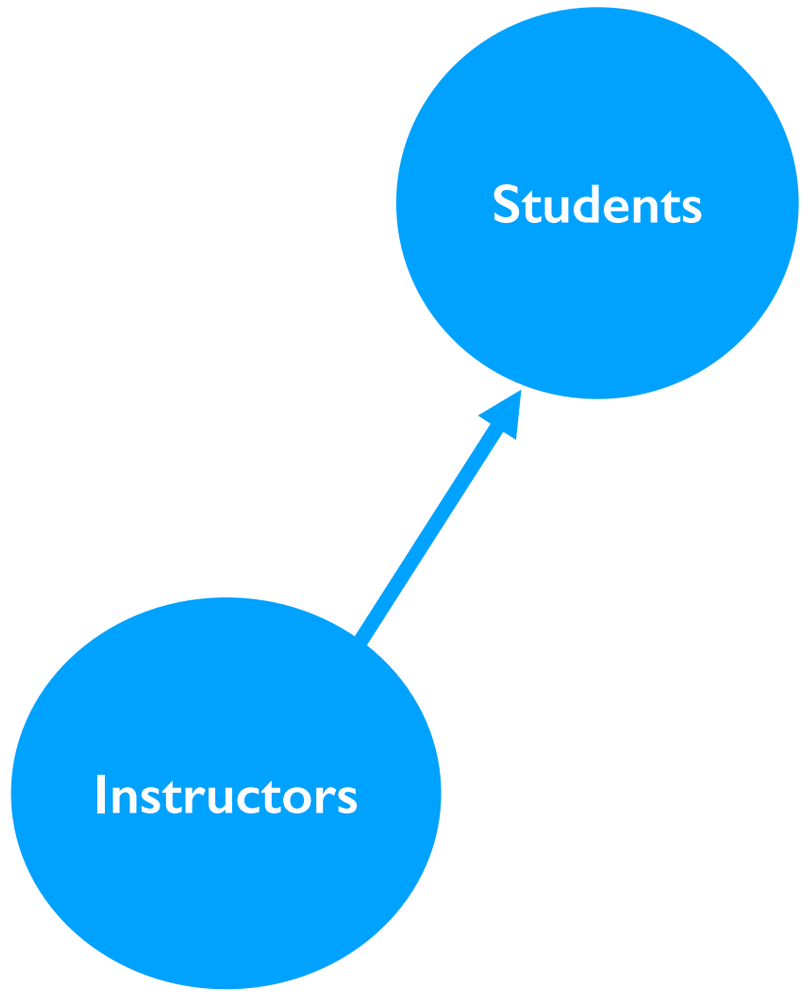


Communication is CS 220

Communication tools

- Office hours (queuing system + BBCollaborate Ultra)
- Piazza
- Email
- Feedback Forms
- Project Submission
- Canvas

Quizzes,
exams,
grades, and
BBCollaborate Ultra sessions



Today's Topics

Introductions

Course overview

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

Computer hardware basics

Website

Grades

49% - programming projects

- **13 projects**, evenly weighted except for p1
- we'll share grading tests with you - avoid surprise
- learning to program is the most import part of the course

20% - quizzes

- 12 quizzes (drop 2 lowest scores)

30% - exams

- 10% midterm 1 (24-hour window)
- 10% midterm 2 (24-hour window)
- 10% final (24-hour window)
- details coming soon

1% - communication

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

- **13 projects**, evenly weighted except for p1
- we'll share grading tests with you - avoid surprise
- learning to program is the most import part of the course

20% - quizzes

- 12 quizzes (drop 2 lowest scores)

30% - exams

- 10% midterm 1 (24-hour window)
- 10% midterm 2 (24-hour window)
- 10% final (24-hour window)
- details coming soon

1% - communication

- filling surveys, following directions, other

Today's Topics

Introductions

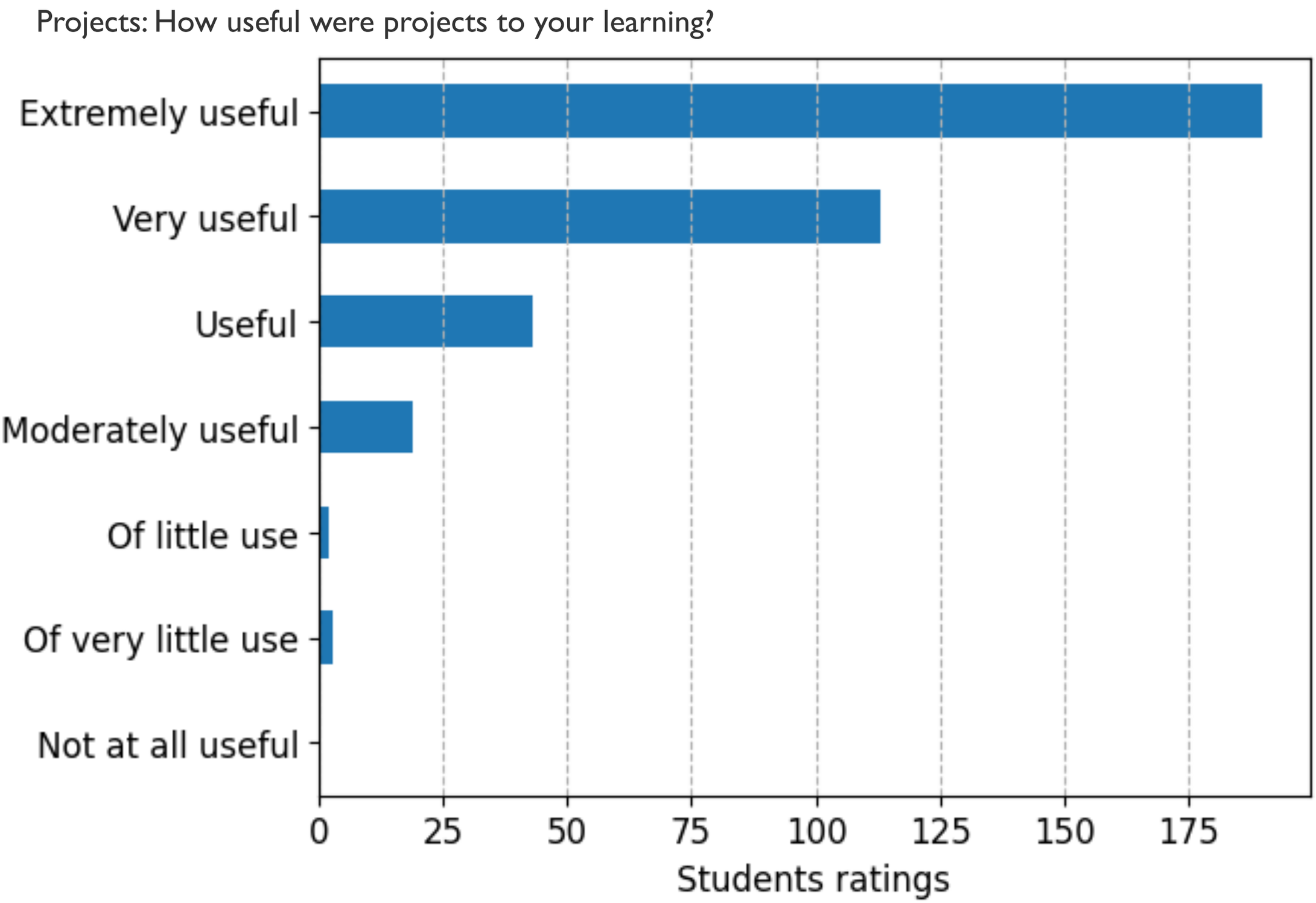
Course overview

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

Computer hardware basics

Website

Prior student reaction to projects



Project Overview

Nearly all projects will relate to some dataset

Timeline

- Projects will be due most weeks, on Wed, at midnight
- You get 10 late days, use them wisely!
- Contact us about any issues

Getting help

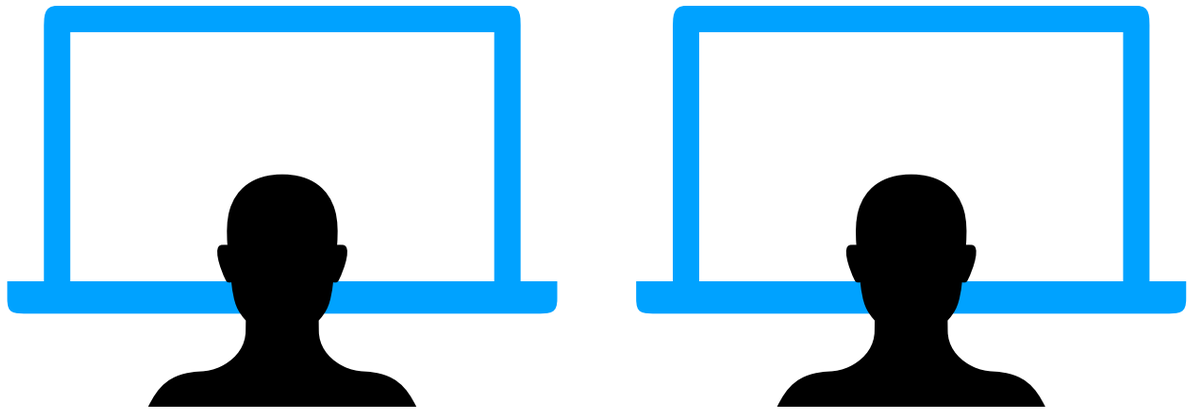
- Piazza
- Instructor or TA office hours
- Lecture Q/A sessions
- Lab sessions
- Email (least preferred)

Pair Programming

You can optionally work in pairs of two

- Partnerships across sections allowed
- Switch partners between projects (or keep with same partner)
- CS220 students can partner with any CS220 students, immaterial of section
- CS319 students can partner with any CS319 students.

Pair Programming

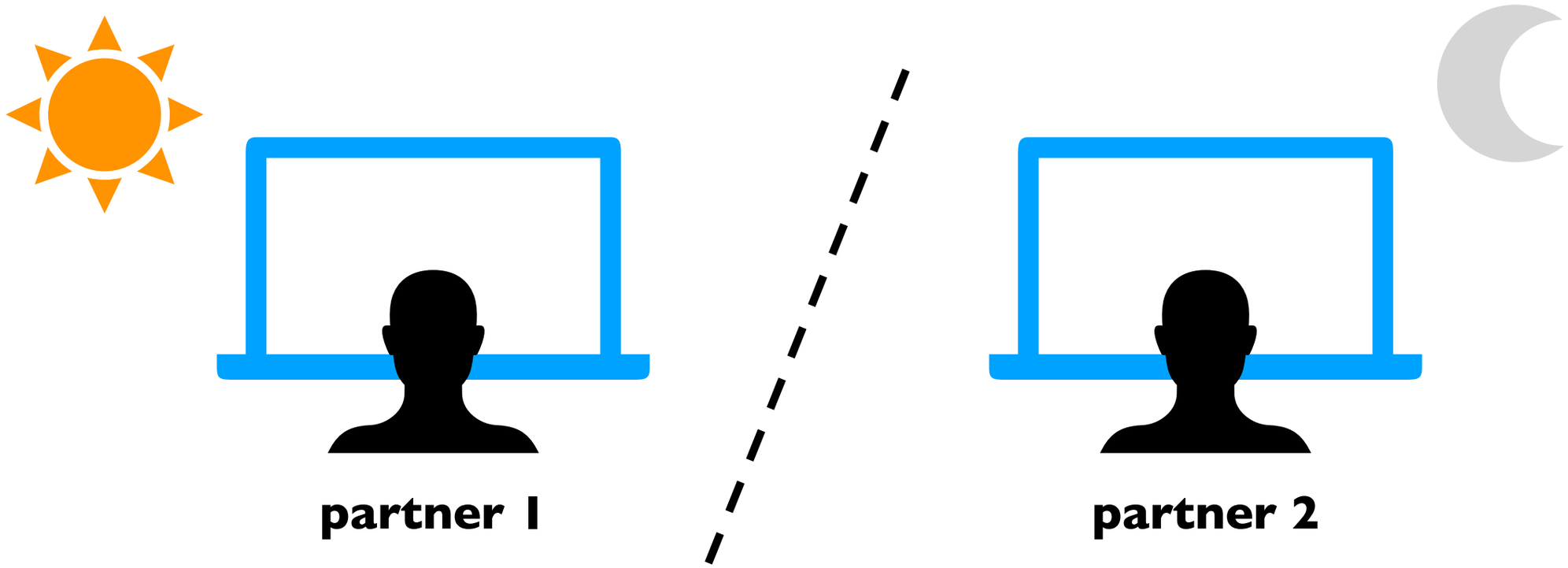


Best practice: working alongside each other

Suggestions

- Use BBC / Google Meet or some other platform for collaboration

Pair Programming

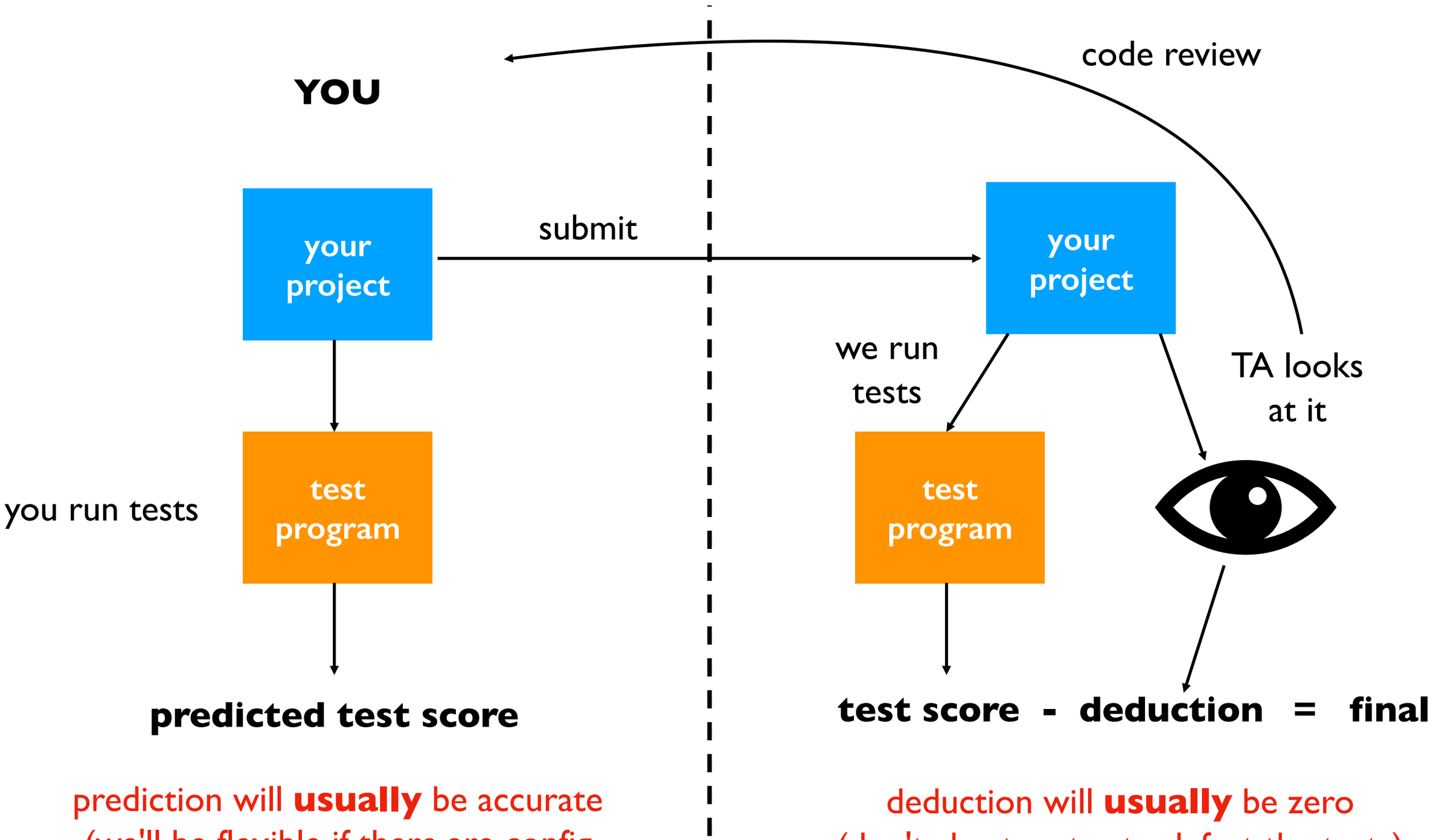


Breaks syllabus rules: working on different parts at different times

Breaks syllabus rules: working on alternate projects individually

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 & quizzes

Computer hardware basics

Website

Quizzes and Exams

Quizzes

- Weekly
- Keeps track of your progress in this course

There will be two midterms and one final

- Multiple choice
- 24-hour window period
- Midterms are already on the schedule

projects ≈ **writing code**

Exams & quizzes ≈ **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

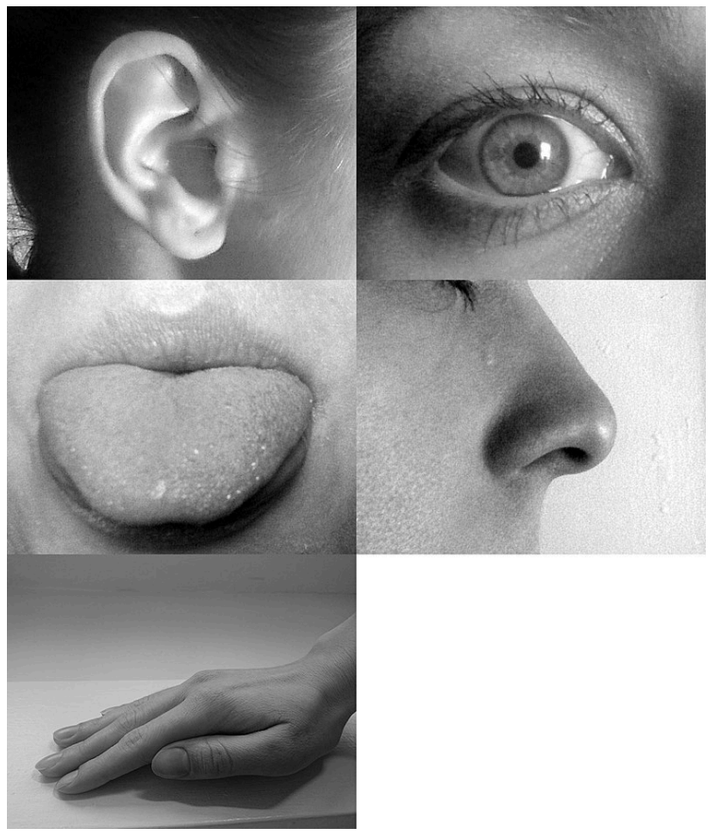
Website

Input/Output

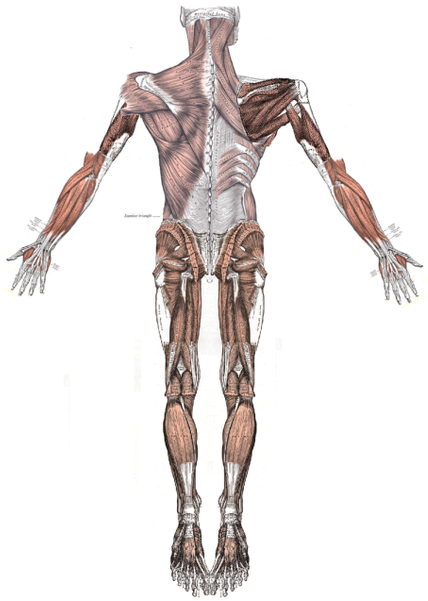
I/O (stands for input/output)

- What are examples for human?

input: senses



output: muscles, voice



https://jasperproject.github.io/

Computer Input/Output

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

monitor

speakers

computer (in a case)

keyboard

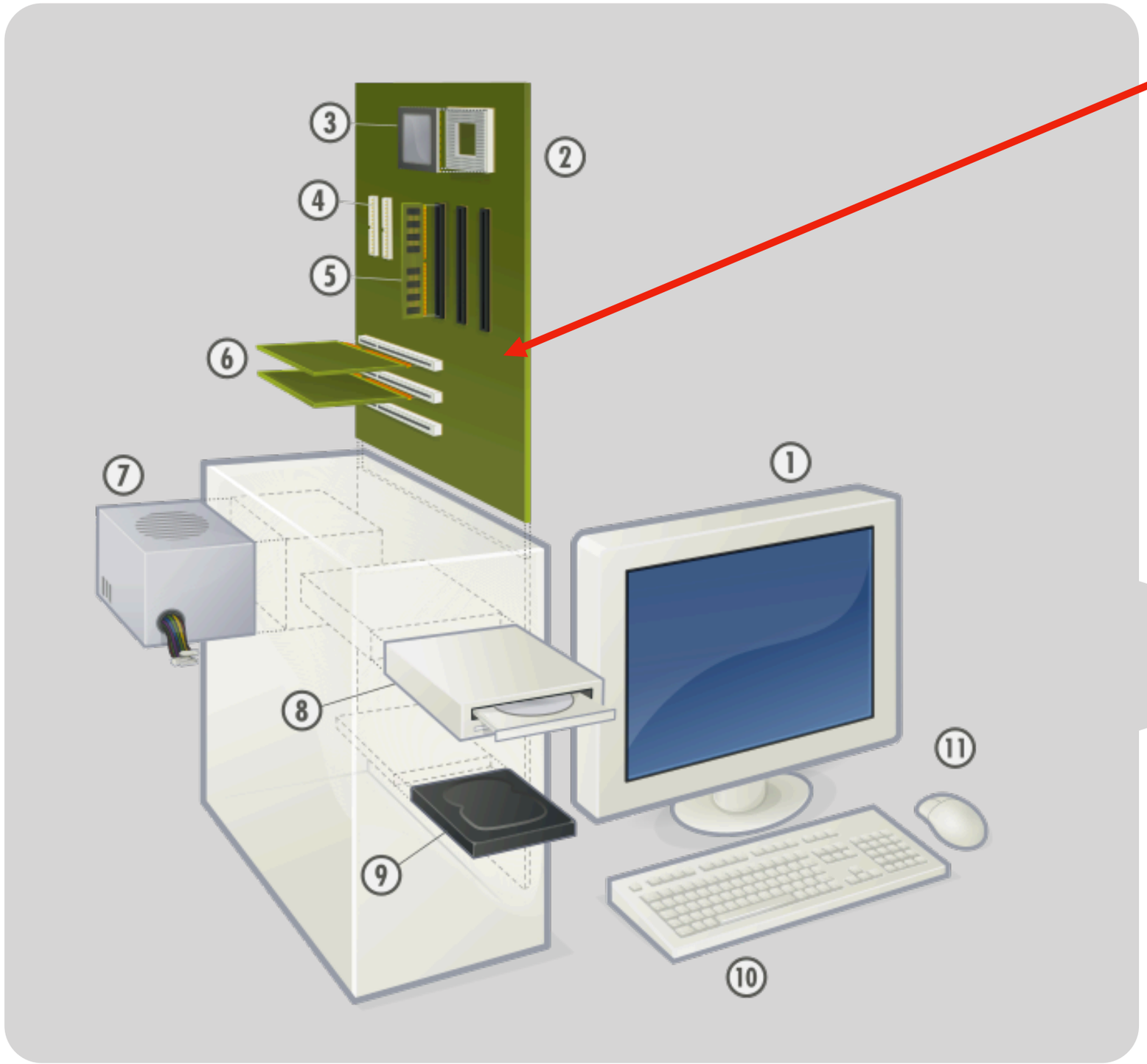
mouse

Computer Input/Output

What's inside a computer case?

Computer Internals

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

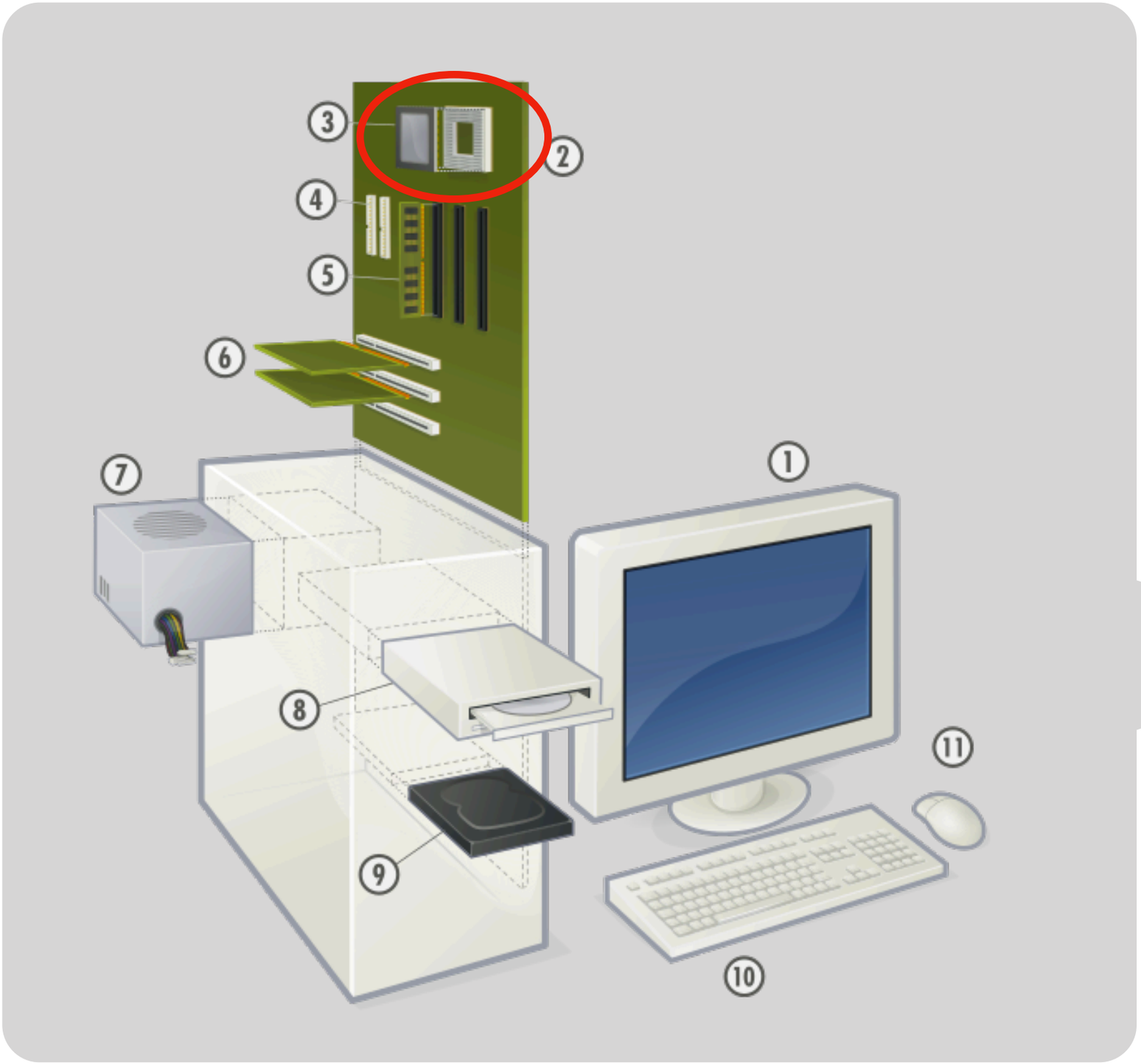
Course overview

Computer hardware basics

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

Website

Central Processing Unit (CPU)



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

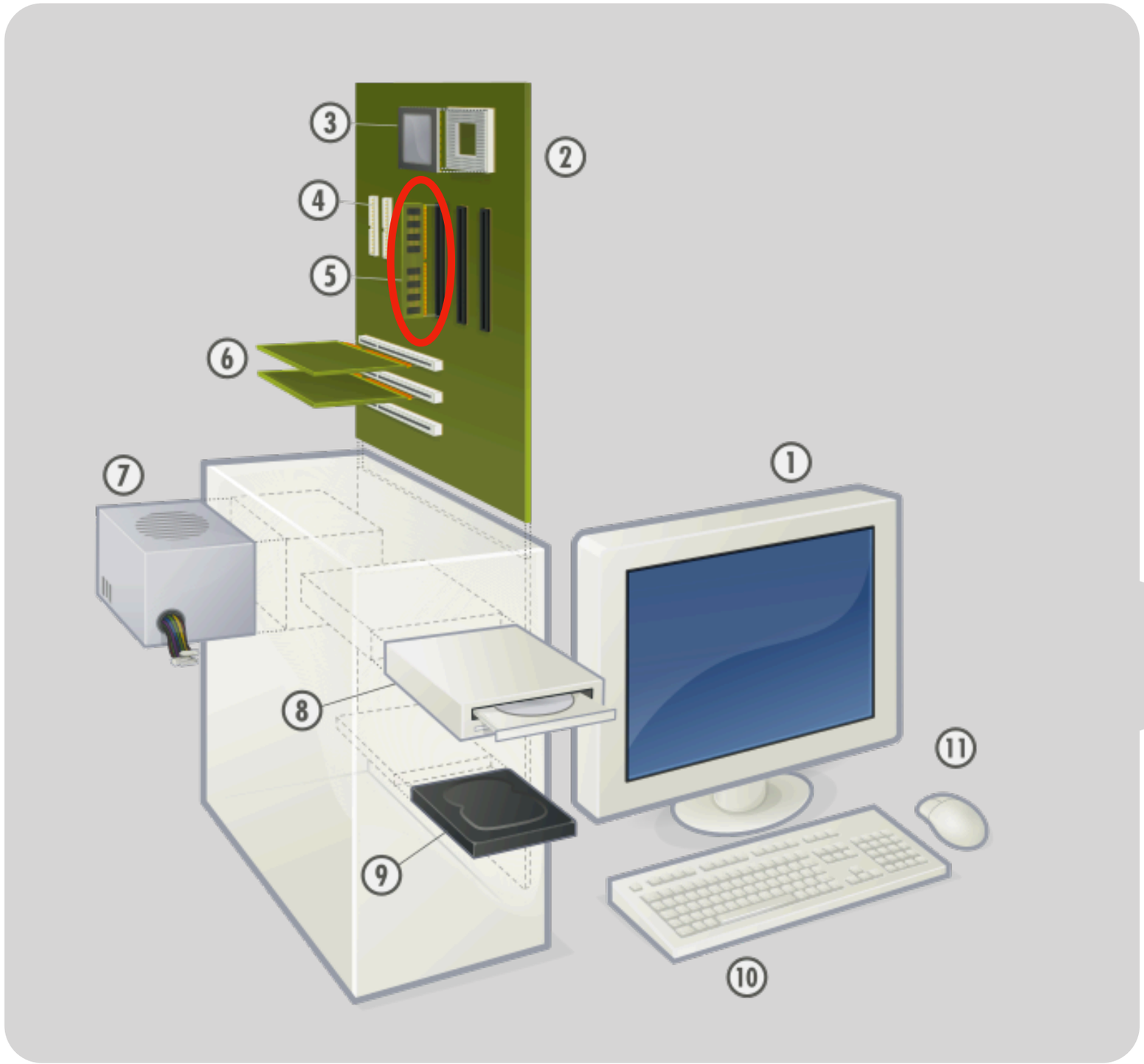
Introductions

Course overview

Computer hardware basics
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Random Access Memory (RAM)



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

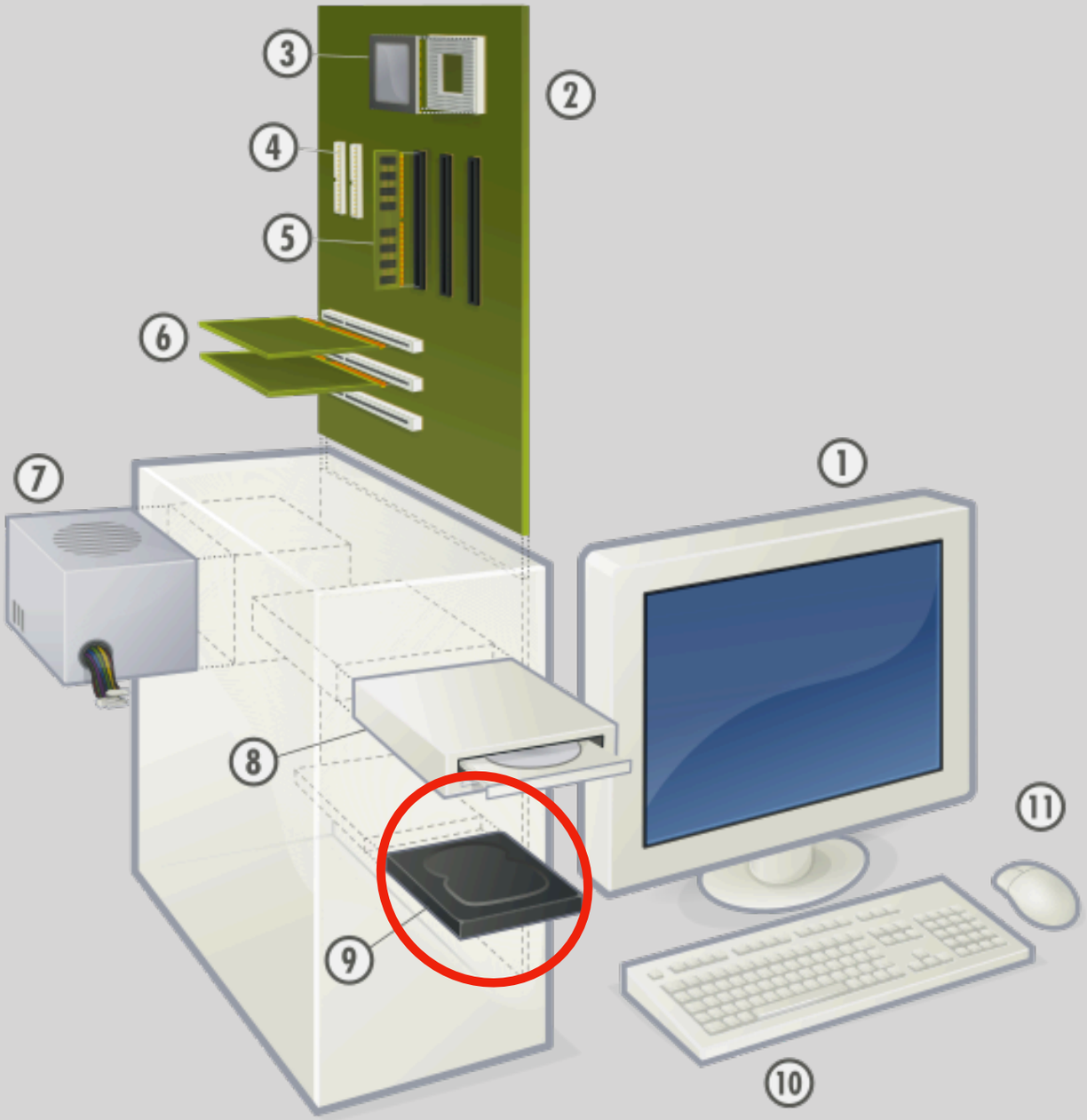
Introductions

Course overview

Computer hardware basics
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Storage Drives



Storage Drives

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

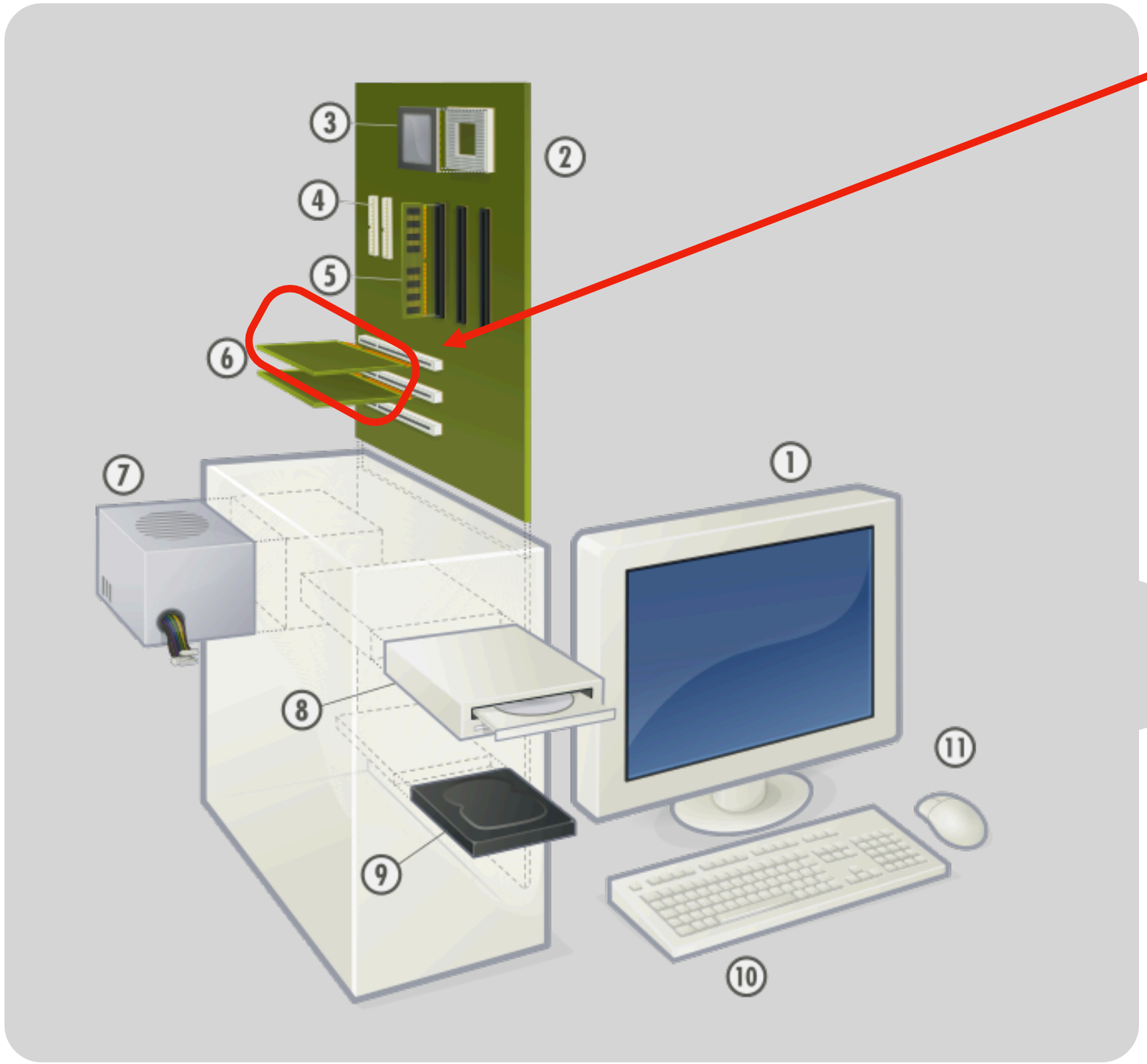
Introductions

Course overview

Computer hardware basics
- Input/Output
- CPU
- Memory
- Storage
- Networking

Website

Network Interfaces



Network: 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 through 4):
https://www.msyamkumar.com/cs220/s21/schedule.html

Walk through…

Next steps...

- take the "Who are You?" survey:
  https://www.msyamkumar.com/cs220/s21/surveys.html

- read syllabus carefully:
  https://www.msyamkumar.com/cs220/s21/syllabus.html

- setup Python on your computer (with videos) and do Lab-P1:
  (Link will be released on the course website on 01/25/2021)

- start P1 (Project 1), due next Wed:
  (Link will be released on the course website on 01/25/2021)