

group projects!

Why projects?

time to show us what you've learned in Matlab!

no tests/midterms in this class – we want to see how you put everything together

this is an opportunity to get your hands dirty using new techniques we won't cover in lecture / tutorials – lots of coding is following instructions you find online and self teaching! (but we have tutorials/documents for many common choices)

we also want you to get to be creative!

Information – see document linked in new module

Dates:

- **today (march 8):** think about groups
- **march 11th (friday):** finalize your group
- **march 14th (monday):** submit proposal
- **march 16th (wednesday):** we approve your proposal
- **march 25th (friday):** submit project!

starting today, **all class times will be spent on projects.** you are also expected to work outside of class with your groups!

Information – see document linked in new module

- groups should be 2-3 students (exceptions if you explain why)
- groups can be across classes: if you have someone you want to work with, let us know. we may add others to your group depending on topics. we can also pair you by topic in general! bit.ly/clps950_projects
- projects should represent **25 hours of work per student**. so bigger groups should have more work in their projects.
- you will use **pair programming** – and we need to see your progress as you work! this is why you'll use git. we will be able to see all your commits – so don't try to do it all at once!

Git



Agenda

1. Questions?
2. Moving to the last part of this module
3. Projects, create groups
4. Git
5. Git practice
6. Problem

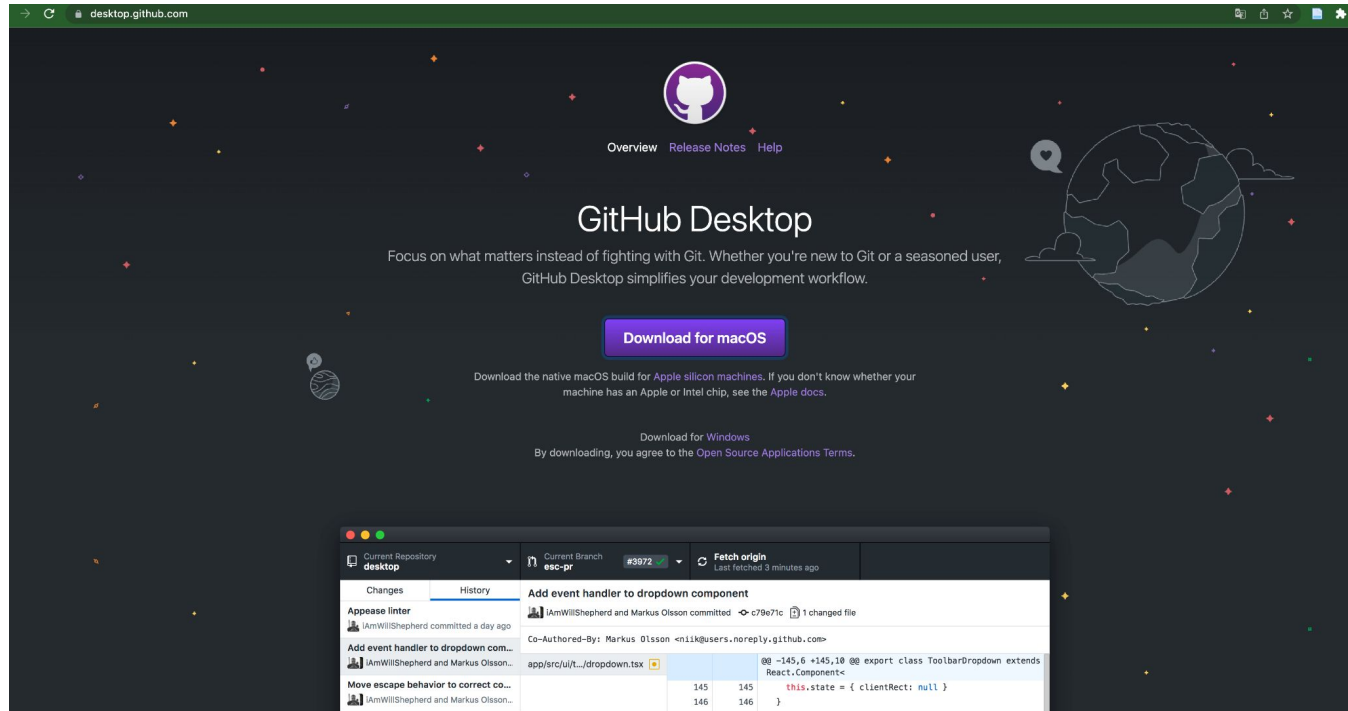
Why version control?

Why track/manage
different versions of code?

- **Backup:** Undo or refer to old stuff
- **Branch:** Maintain old release while working on new
- **Collaborate:** Work in parallel with teammates

Getting git

<https://desktop.github.com/>

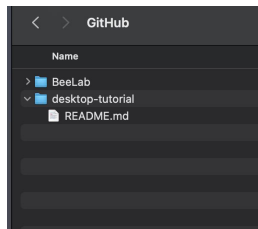


Github idea

Local

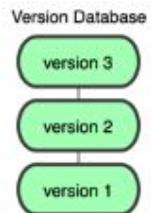


Working dir:

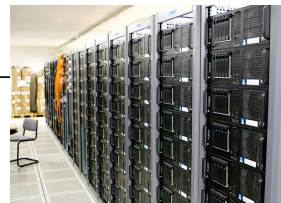


Local Repo

Version Database:



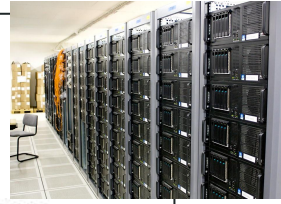
GitHub



Version Database



GitHub

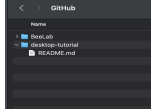


Version Database



Local 1

Working dir:

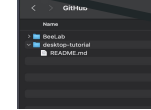


Local Repo



Local 5

Working dir:

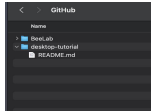


Local Repo



Local 2

Working dir:

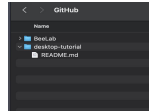


Local Repo



Local 3

Working dir:

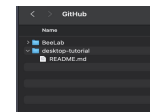


Local Repo



Local 4

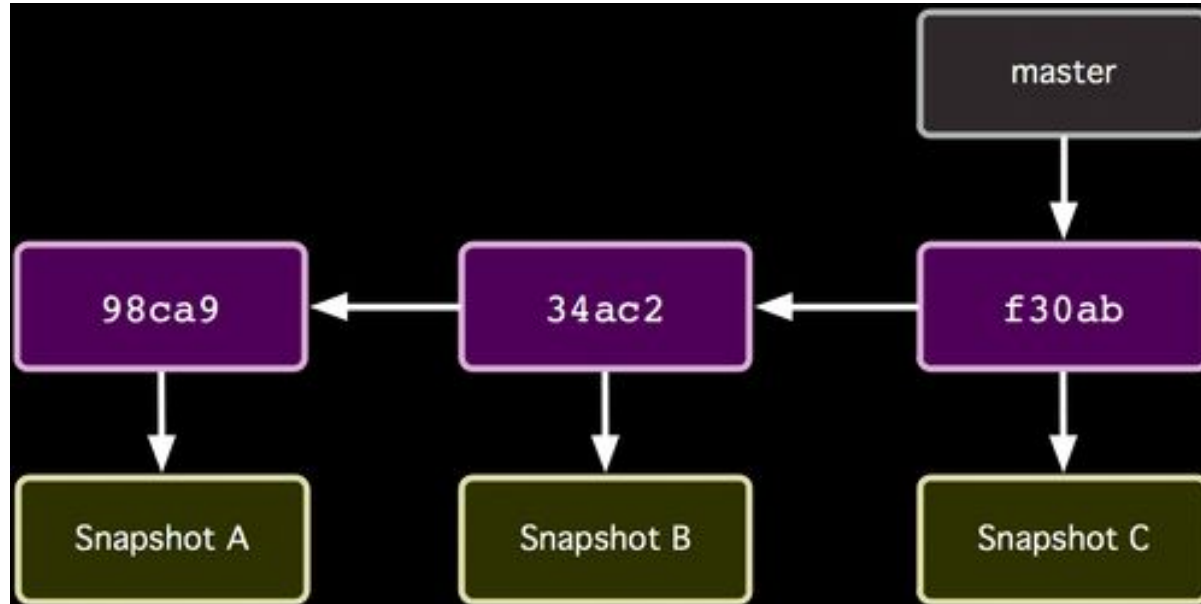
Working dir:



Local Repo

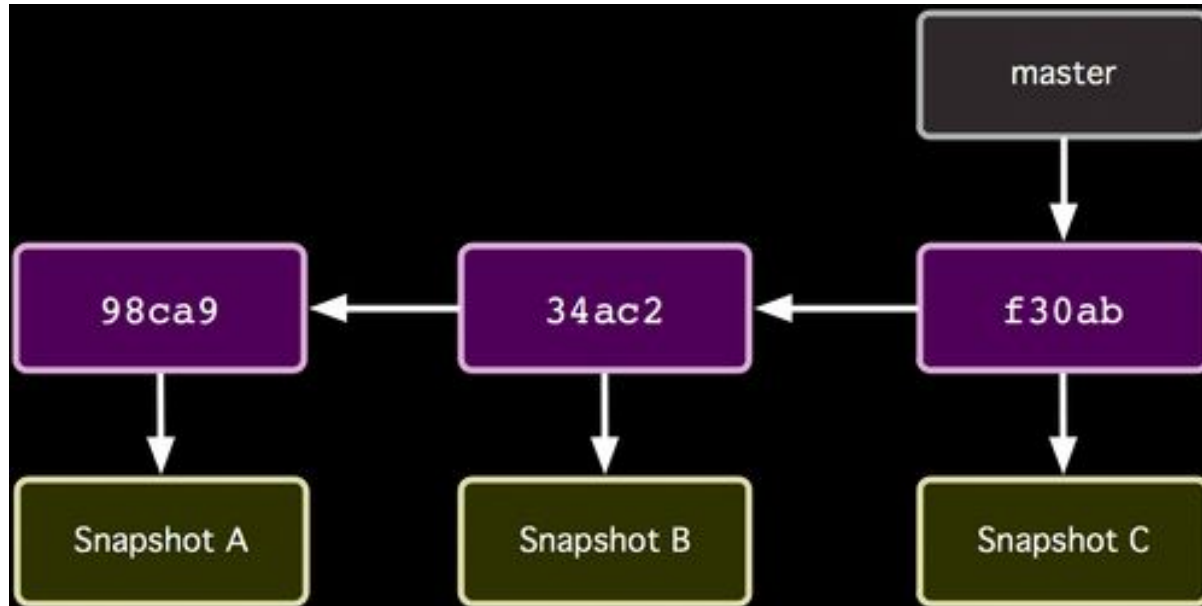


Repo organization

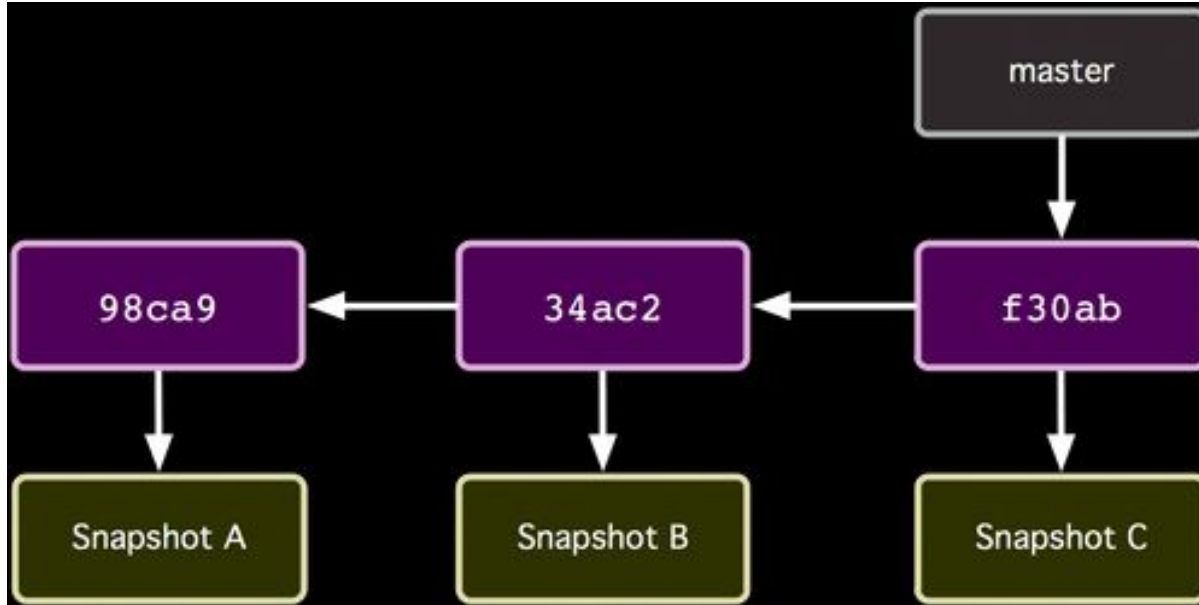


Repo Organization

Commits (oldest to newest; Hashes as commit Ids)



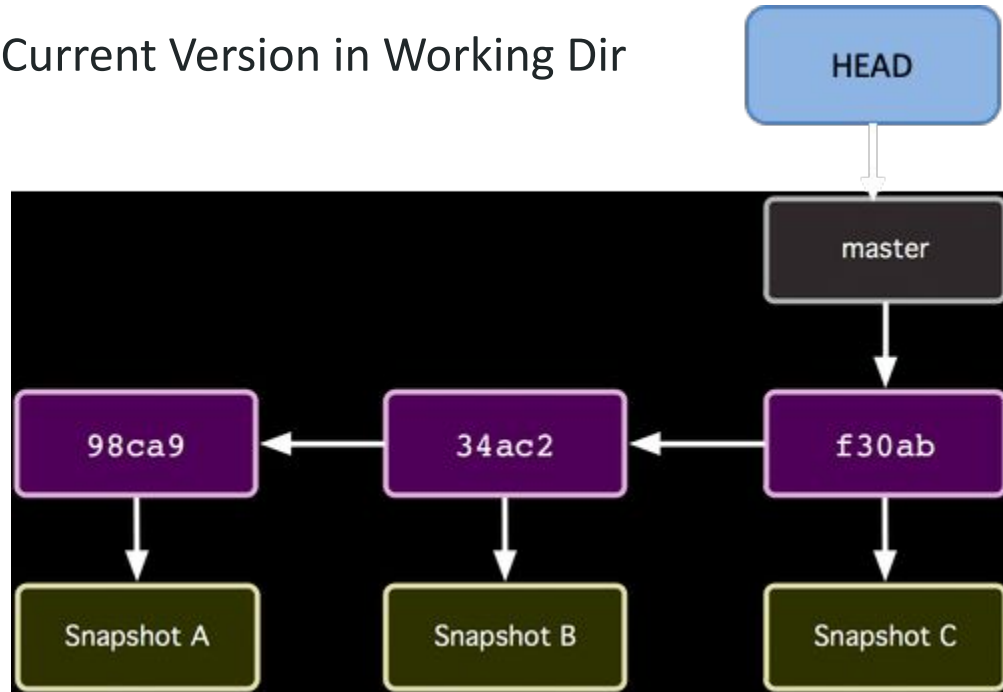
Repo Organization



Snapshots contain a copy of the current state of all the files in the commit.

Head of a Repo

Current Version in Working Dir

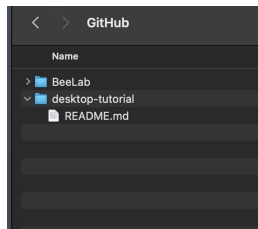


Local Repo Operations

Local



Working dir:



Local Repo

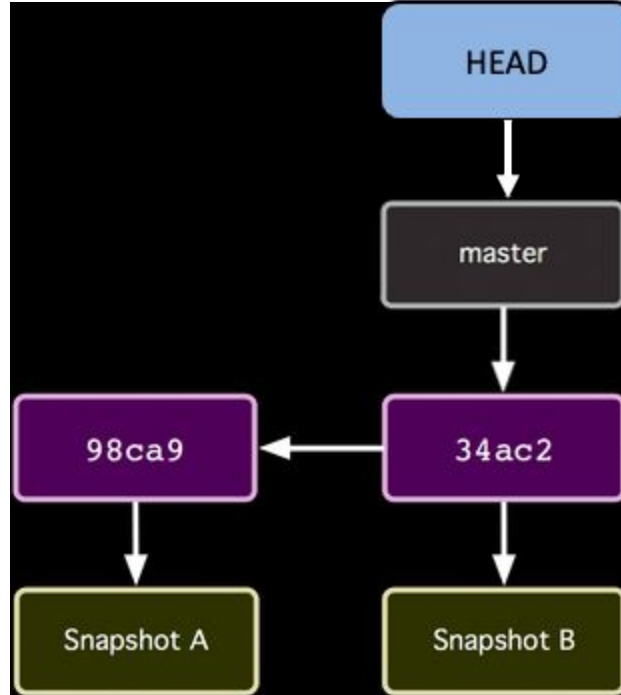
Version Database:

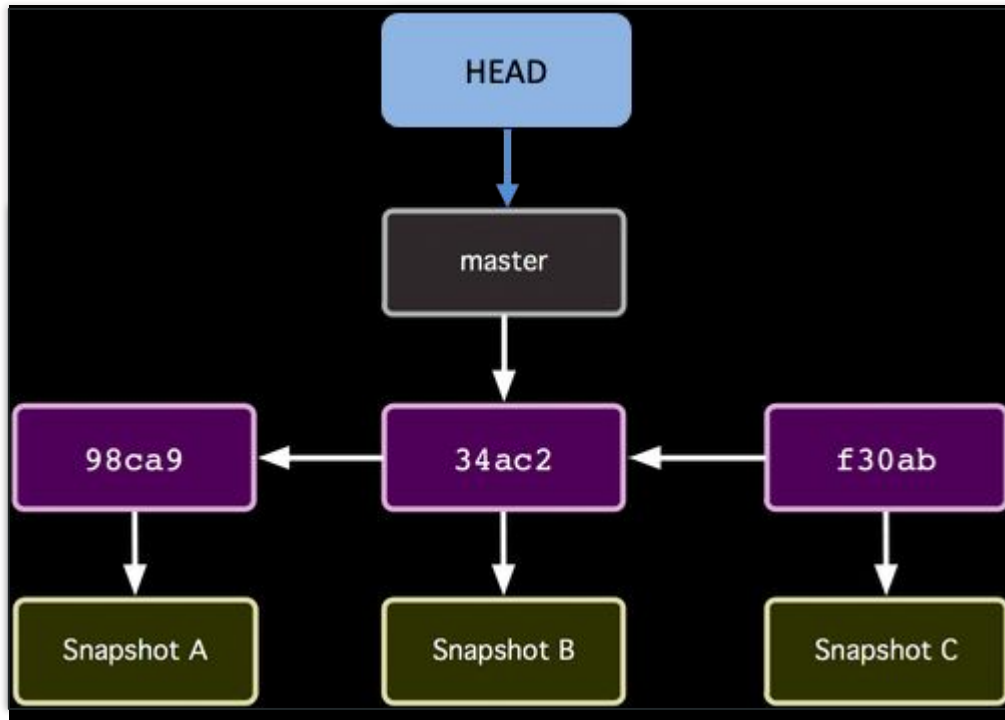


- init
- add/commit
- log
- switch/checkout
- branch
- merge
- ...

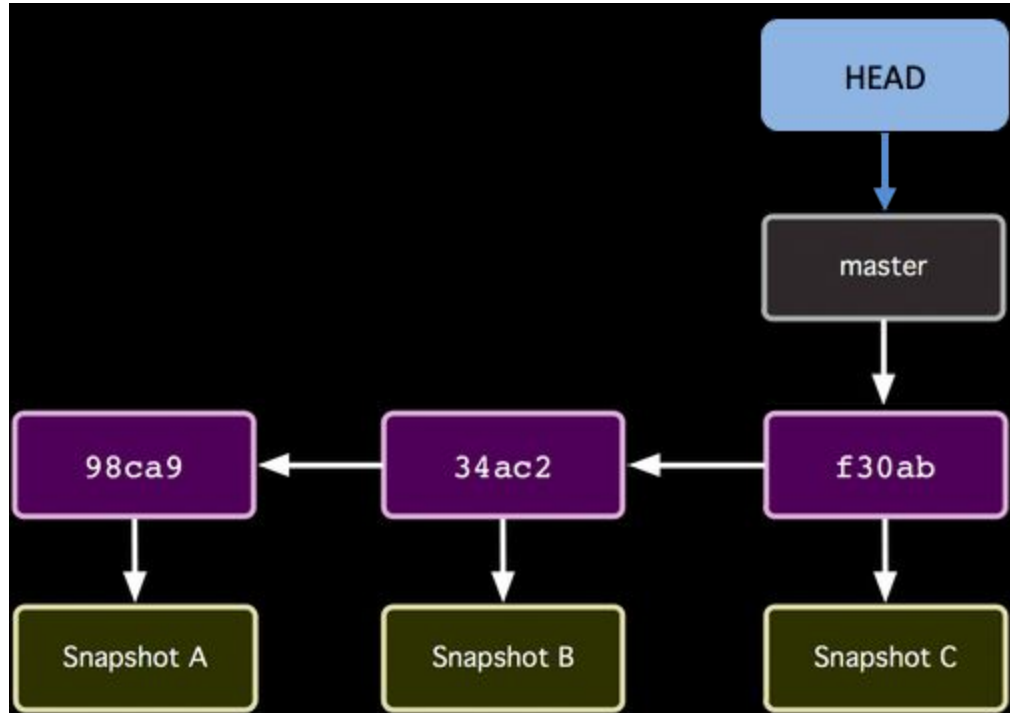
How Commit works

Before

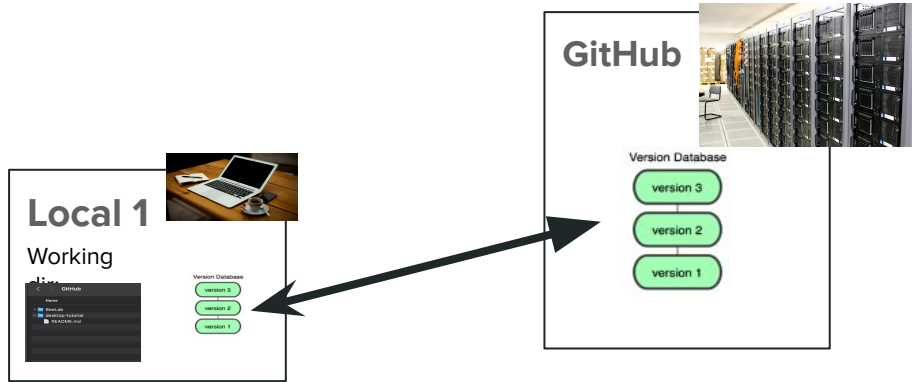




After



Interacting with the server...

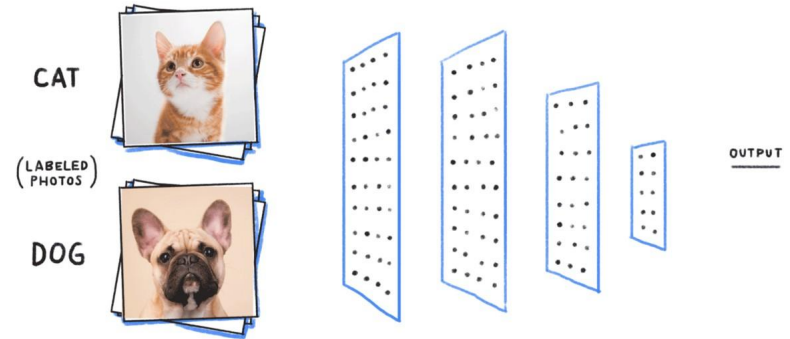
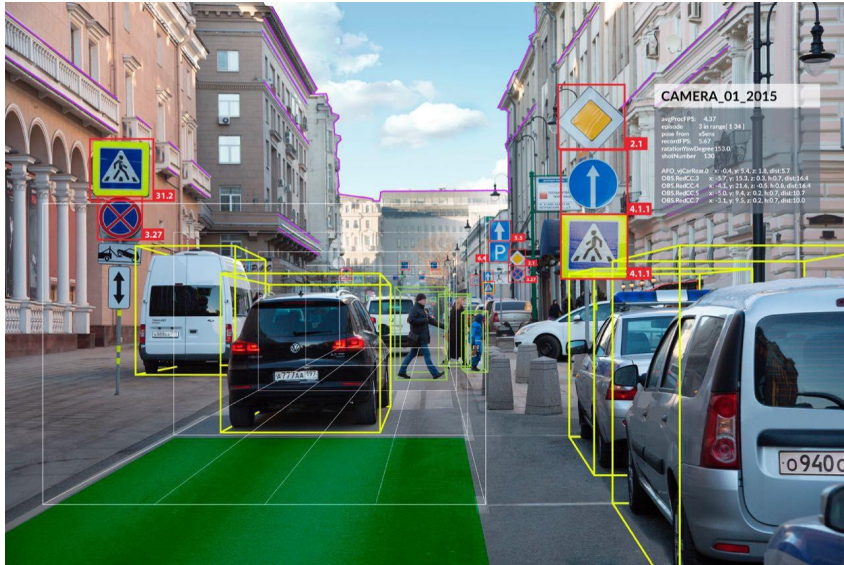


- push
- clone
- fetch
- pull
-

Project Ideas based on your responses

- computer / human vision and optical illusions

how do computers classify images? you can implement a computer vision network to classify images in two categories (dog and cat). are the features computers use to categorize images what you expect?




Project Ideas based on your responses

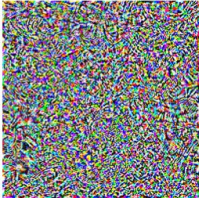
- computer / human vision and optical illusions

what types of images confuse computers? are they the same types of illusions that confuse humans?
you will take a computer vision model and try to build an ‘adversarial’ one to work against it and get it to say the wrong thing!

“pig”




+ 0.005 x



=


“airliner”



Clean image


Noise

Adversarial example




“panda”
57.7% confidence


+ ϵ



=

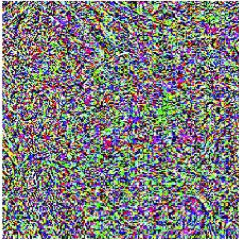



“gibbon”
99.3% confidence




tabby
47.96 %

0.01x



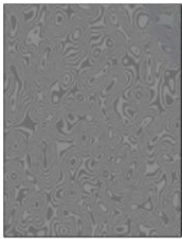


mosquito_net
99.99 %




Prediction: “Chair”
89.1% confidence

+



Adversarial noise

=



Prediction: “Person”
99.3% confidence

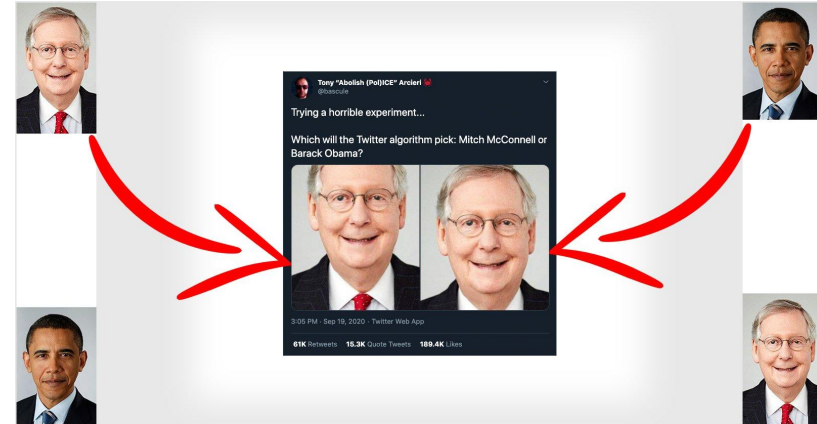
Project Ideas based on your responses

- algorithmic bias in policing

why do we hear about algorithmic bias? what does that mean? how easy is it for bias to enter into predictive encoding? can we use the same types of features we use with image processing to do predictions of crime? given a dataset of traffic stops in maryland, can you see why we shouldn't use past arrest statistics as prediction of future arrest probability? can you draw a map of where more traffic stops occur given data?



*data*Montgomery



Project Ideas based on your responses

- sorting game: hue

using psychtoolbox, can you make a game for people to sort colors based on their hue?



Project Ideas based on your responses

- character quiz using psychtoolbox

collect data about people using a game and tell them what character they are most like (using a dataset which already exists). lots are out there for our use! fun one is tv / movie characters... lots of data

Open-Source Psychometrics Project

Raw data from online personality tests

For general public edification the data collected through the [personality tests on this website](#) is dumped here. All data is anonymous. Users were informed at the beginning of the test that their answers would be used for research and were asked to confirm that their answers were accurate and suitable for research upon completion (those that did not have been removed from these datasets).

In a comparison, the quality of this data is equal to or better than data from Amazon Mechanical Turk. See [A quality comparison of data collected on this website to data collected on Amazon Mechanical Turk](#). These datasets have been used in 25+ high quality published papers (and many more dubious quality ones); see a list [here](#).

The downloadables are .zip files each containing a .csv file with the data (open with Open Office Calc, or Microsoft Office Excel) and a .txt codebook.

Updated	Description	Variables	n	Download
5/14/2014	Answers to Cattell's 16 Personality Factors Test with items from the IPIP	163 likert rated items, gender, age, country and accuracy	49159	16PF
6/05/2020	Aggregated ratings from the Statistical "Which Character" Personality Quiz . (Interesting for practicing dimension reduction techniques because of low measurement error due to aggregation!)	268 items	800 characters	characters-aggregated.zip
4/05/2021	Raw data from the Statistical "Which Character" Personality Quiz . Multiple different surveys.	400 personality items, 1600 characters, self-reports, character ratings of multiple types, demographics	~2,500,000	characters_raw.zip
9/6/2012	Answers to the Narcissistic Personality Inventory , constructed with the version from Raskin and Terry (1988).	40 multiple choice, gender, age, time elapsed	11243	NPI

Use the slider to indicate where you fall on this spectrum:

deep (50%) shallow (50%)

Project Ideas based on your responses

- public health / infectious disease

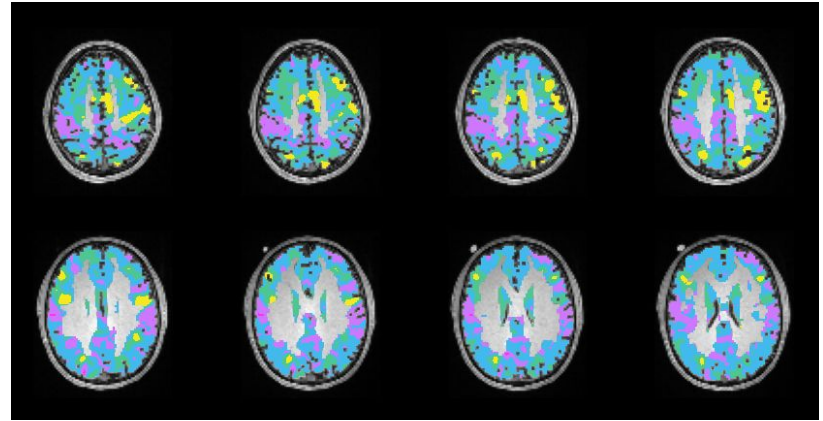
make a visualization of covid spread in providence (or at brown) with different parameters (how contagious, masking/social distancing), look at datasets of disease / hospitalizations / etc (so many cool projects!), use disease information to make a small “webmd” to tell people the most likely disease based on their symptoms ...



Project Ideas based on your responses

- MRI / neuroimaging

what does data for neuroimaging look like? if we have a 3D image with a 4th dimension for time, how can we visualize this? how could we plot an image of finding signal in the brain? how do we know where in the brain there is signal?



Project Ideas based on your responses

- data pipeline for your own projects! (set up a meeting with us)

