



LEAP

Climate Prediction Challenges

Introduction

Spring 2022

January 18, 2022



You are the
inaugural class of
LEAP CPC!



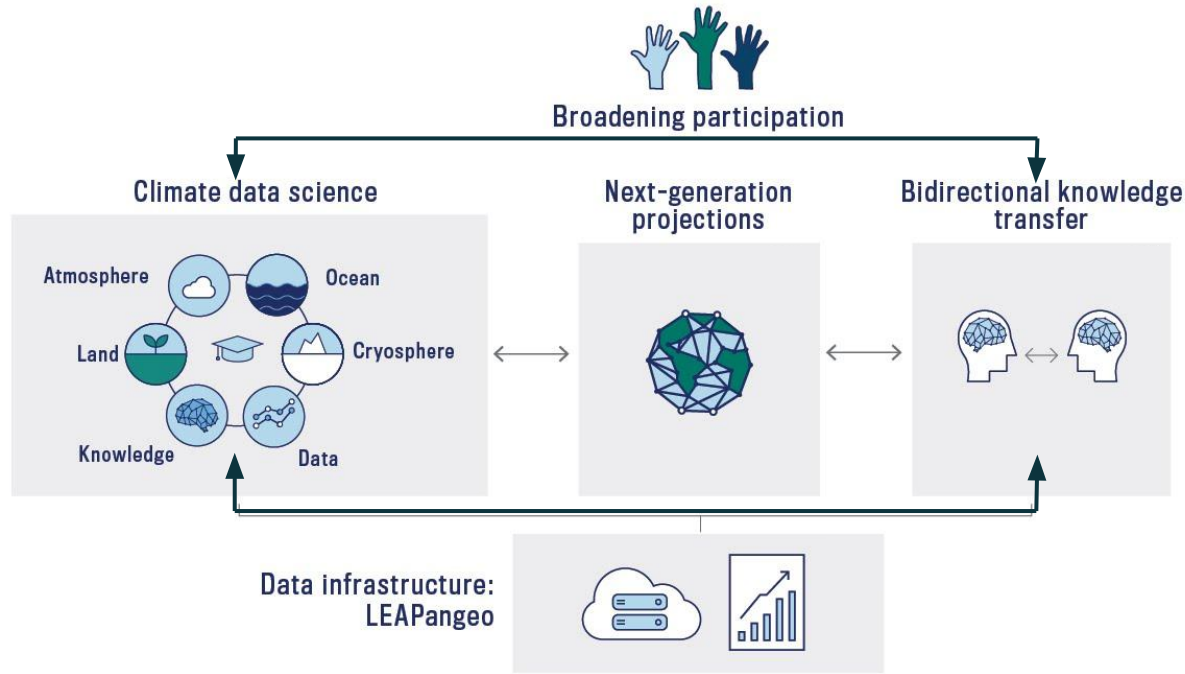
LEAP

What is LEAP?

- ❑ NSF Science and Technology Center (STC) led by Columbia
- ❑ Columbia, NYU, NCAR, UC Irvine, Minnesota, Columbia/Teachers College
- ❑ At Columbia, the LEAP community reaches SEAS, LDEO/Climate School, A&S, SSW, CBS, etc.



Vision: *LEAP forward in the **reliability**, **utility**, and **reach** of climate projections through synergistic innovations in data science and climate science*





**Chief Convergence Officer
& Education Director**
Tian Zheng

- Professor + Chair of Statistics, Columbia
- PhD, Statistics, Columbia

Expertise

- Spatiotemporal modeling
- Statistical machine learning
- Data science education
- Management

Synergistic Experiences

- Associate Director for Education, Data Science Institute, 2017-2020
- Creator, DSI Scholars Program, 2018-2020
- Leadership Team Member, Collaboratory Program
- Faculty Advisor, Columbia Statistics Club 2016-2020





Deputy Director
Galen McKinley

Professional Experience and Education

- Professor of Earth + Environmental Sciences, Columbia
- previously Prof. of Atmospheric and Oceanic Sci., Wisconsin
- PhD, Climate Physics + Chemistry, MIT

Expertise

- Ocean carbon cycle, feedbacks with climate
- Ocean and climate modeling; data science

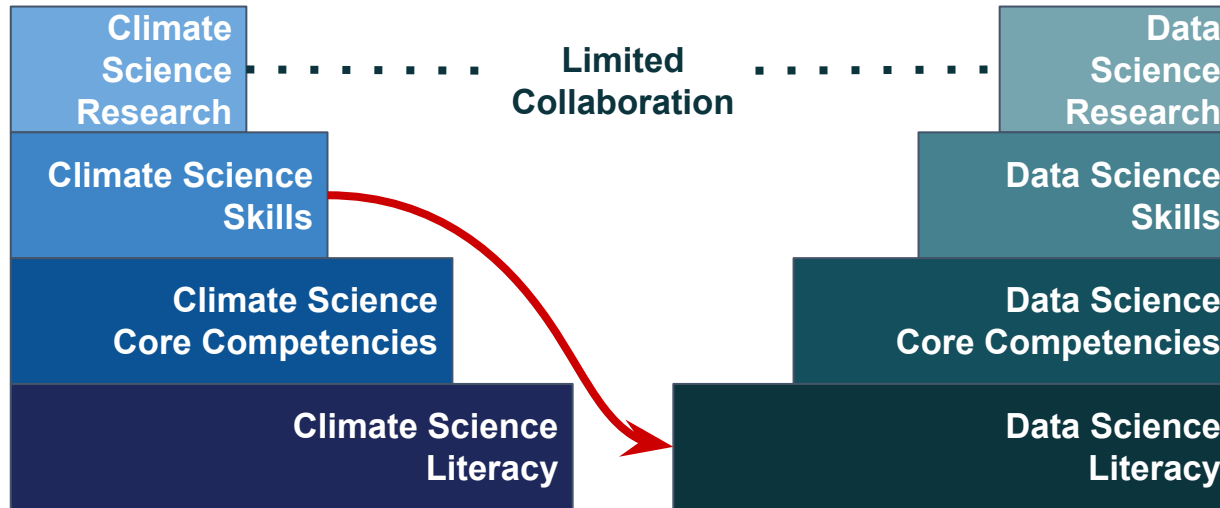
Synergistic Activities

- The Oceanography Society, Councilor, 2020-present
- JASON Advisory Group, 2015-present
- Diversity Committee chair, EES department, 2019-present
- Co-chair for undergrad education, Wisconsin AOS, 2010-15



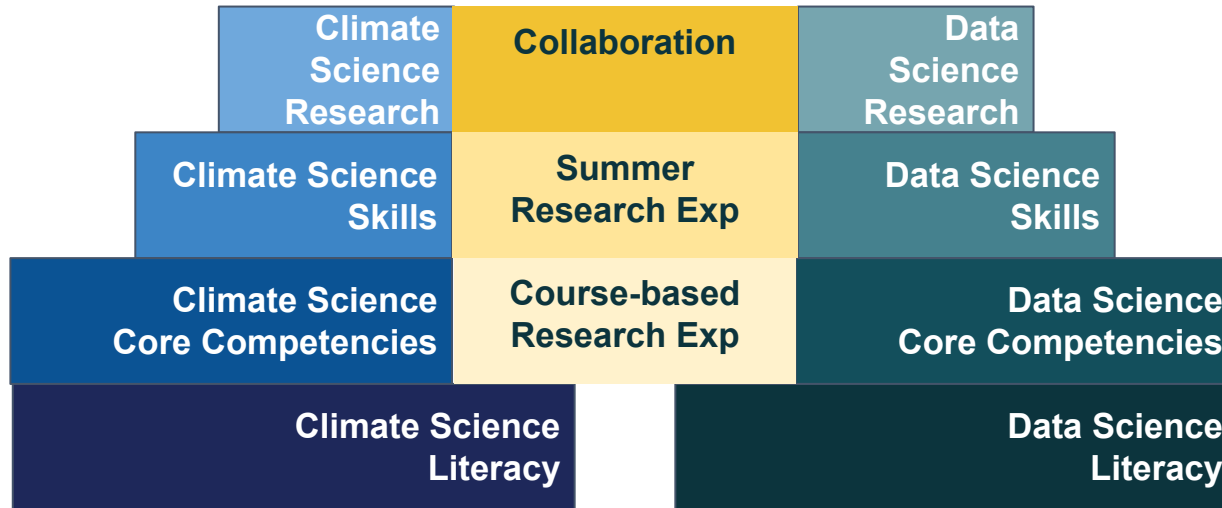
Climate Scientists + Data Scientists Overcome the Collaboration Barriers

Disconnect between core disciplines exacerbates skills gaps + narrows pathways for convergent learning + research

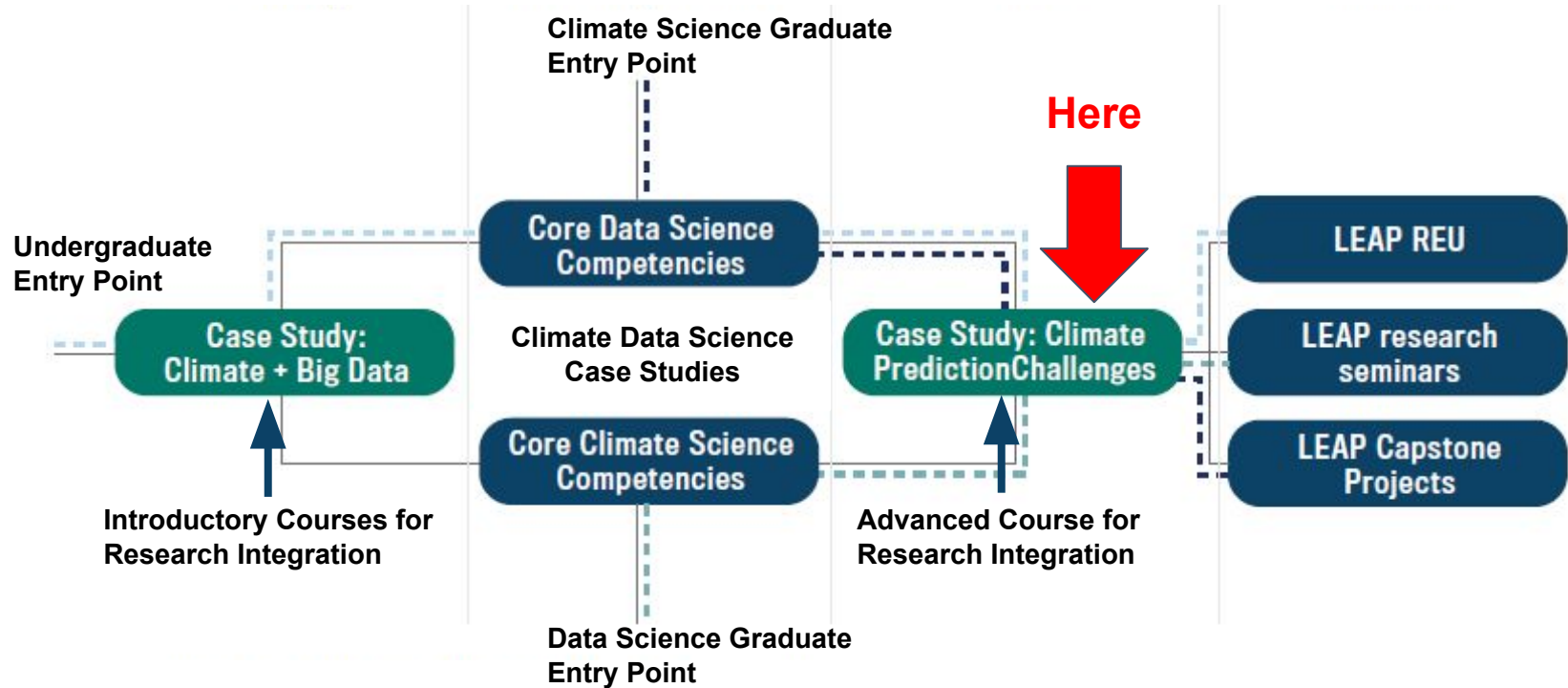


Climate Scientists + Data Scientists Overcome the Collaboration Barriers

Research Experiences started in a mixed classroom offer interdisciplinary learning and early training in collaboration skills



Where you are in LEAP's learning map?



What is key to climate science + data science collaboration?

The hardest part, I think, is the different vocabularies. And the way to overcome it, is to keep talking to each other.



- Luke Gloege (Postdoctoral Fellow at NASA Goddard Institute for Space Studies)
- ❑ January 2018 - Data Science Bootcamp
- ❑ Spring 2018 - Collaborated with a PhD in Statistics student
- ❑ Fall 2018 - Mentored a DSI capstone project
- ❑ Summer 2020 - Defended dissertation with a strong data science component.



Let's talk to each other
In this course



LEAP

What are the most useful data science skills?

Skills listed in job postings
on indeed.com



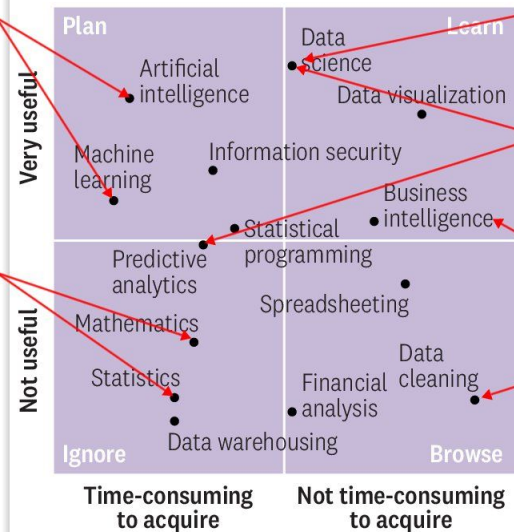
Which data skills do you actually need?
To find out, make sure you **IGNORE** this idiotic 2x2 matrix.

An Example of How to Plot Data Skills on a 2x2 Learning Matrix

How one company mapped its own internal learning needs.

I get the feeling that whoever made this chart has NO IDEA what either of these is.

Good luck doing data science without these "not useful" skills!



"Not time-consuming to acquire"

How can predictive analytics be in the "ignore" quadrant, when it's one of the big uses of data science, which is in the "learn" quadrant.

Predictive analytics (which they call "not useful", but in a suit ("very useful").

People who work with data will tell you that the task that eats up most of their time and that nobody tells you about is data cleaning.

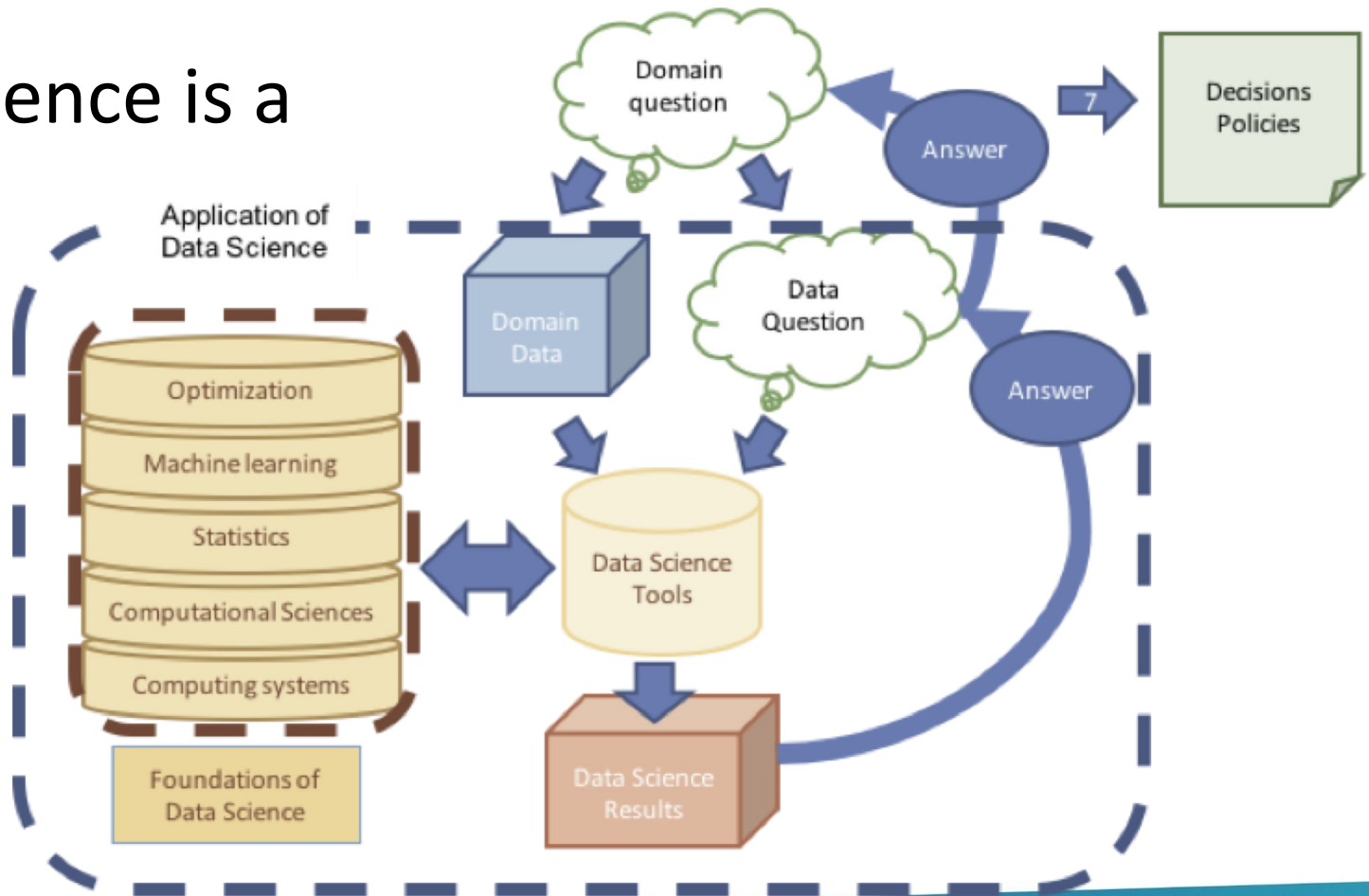
<https://www.globalnerdy.com/2018/10/29/data-science-reading-list-monday-october-29-2018/>

<http://www.kimberlycoffey.com/blog/2016/11/text-analysis-is>



LEAP

Data Science is a process



Data Science Skill Sets

- How to **think** about data versus problem:
 - Mathematics/Statistics/Machine Learning
- How to **handle** data
 - **Technologies: R/Python, Java, SQL, etc**
- Teamwork and collaboration skills - how to **work** with others.
- How to **interpret** data into knowledge and insight:
Or find **value** in your data
 - Innovation, intellectual curiosity
 - Problem-solving skills
- How to convince others about your data science results
 - Visualization, story telling
 - **Communication** skills



How this course can help?

- ❑ No formal and systematic instruction on statistics/machine learning.
- ❑ Not intended to be a comprehensive data science bootcamp.
- ❑ Project-based learning course. **Learning by doing.**
 - ❑ Problem **identification** via teamwork and discussion.
 - ❑ Problem **solving** by using existing skills or **new skills**, learn new things “on the job”, and learn from your peers.
 - ❑ Present your codes, your results and your story (try to sell them).
- ❑ There will be things we cannot answer right away but let's learn together.

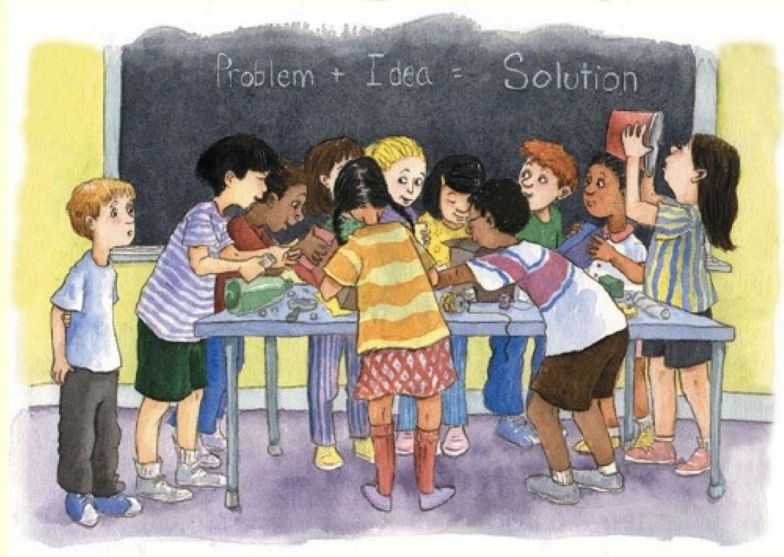


Stay Hungry. Stay Foolish.
- Steve Jobs



LEAP

Project-Based Learning Integrating 21st Century Skills

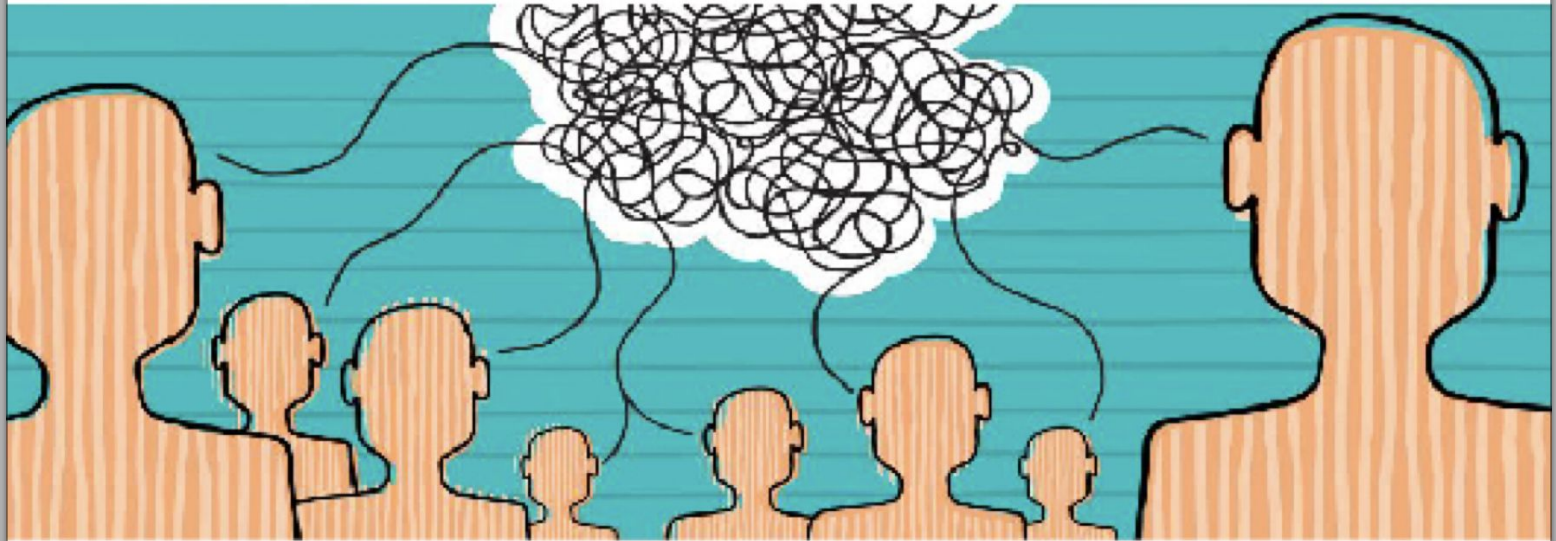


This is a student-led course

- ❑ We are not to lecture here but to facilitate active learning.
- ❑ We design open-ended challenges, each of which focuses on a slightly different area in climate data science.
- ❑ In each challenge,
 - ❑ Start with information/knowledge we already have (maybe not yours but your teammate's) about the problem.
 - ❑ Identify knowledge/skills we need to solve the problem.
 - ❑ Articulate the above thinking process in a team and implement an inquiry as a team
- ❑ We will provide case studies and tutorials to provide guidance on aspects of the above processes.



Communicate!



Communication is EVERYTHING!



LEAP

Channels of Communication

- During class time
 - Brainstorm
 - Ask questions during tutorials and presentations
- Before and after classes
- Discussion board
- Office hours
 - TBA
- Group discussion (Slack, Discord, etc)
- Group collaboration (Google drive, GitHub)



LEAP CPC Learning Systems

- ❑ Courseworks: grading, peer reviews, discussion, Q&A, official announcements
- ❑ GitHub: where the project materials are posted
- ❑ Google Colab - develop and collaborate
 - ❑ (only required for project 1)



Group projects (Randomly assigned)



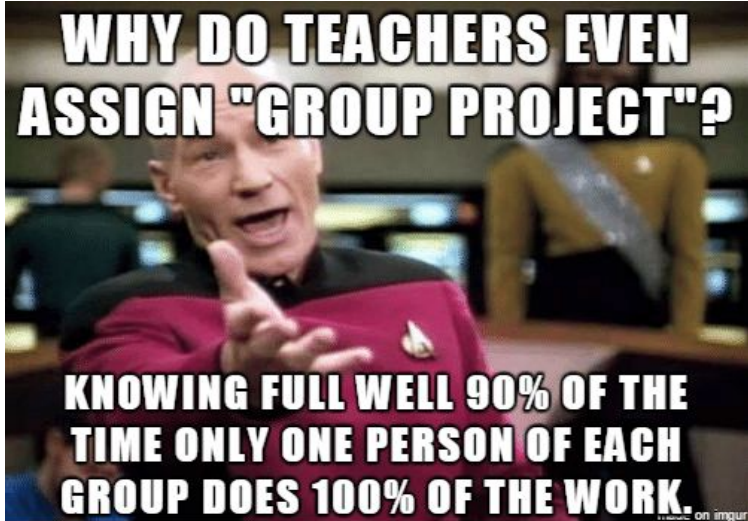
Working together

- ❑ You don't have to be in the same (zoom) room at the same time to work together.
- ❑ Here are several ways you will work together in this course
 - ❑ Face-to-face or zoom brainstorm
 - ❑ Online discussion in group forum
 - ❑ Slack/wechat is good for brainstorm but not good for assign tasks. Keep notes and project updates in group forum, google drive or GitHub
 - ❑ Be diligent on taking notes and post to a group-shared notes keeping mechanism. The team need to be “on the same page.”
- ❑ Learning is not a zero-sum game.



Group Projects

When you're doing a group project
but you end up doing all the work



It happens every year



LEAP

Intro to Google Colab



Coding TensorFlow



LEAP