

# Stat 418: Tools in Data Science

## Course Description

Stats 418: Tools in Data Science is a graduate level statistics course restricted to UCLA Masters in Applied Statistics students. The course will present current tools for data acquisition, transformation and analysis, data visualization, and machine learning and tools for reproducible data analysis, collaboration, and model deployment used by data scientists in practice. Advanced R packages and Python libraries, analytical databases, high-performance machine learning libraries, big data tools.

## MASDS Departmental Disclaimers:

For students trying to take or audit from outside the MASDS program.

Taking or auditing 400 courses is simply not permitted because this is a self-supporting program. Sorry, unfortunately, you will NOT be able to take any of the 400 level Stats courses.

There are NO exceptions that can be made by the department. These classes were designed specifically for students who applied directly to the program.

The students of this program are also not allowed to audit or enroll in classes outside of the program as it was created for working professionals.

If you would like to apply for the program, you are welcome to do so: <https://master.stat.ucla.edu/admissions/>

Information is found here: <https://master.stat.ucla.edu/> And here: <https://master.stat.ucla.edu/faq/>

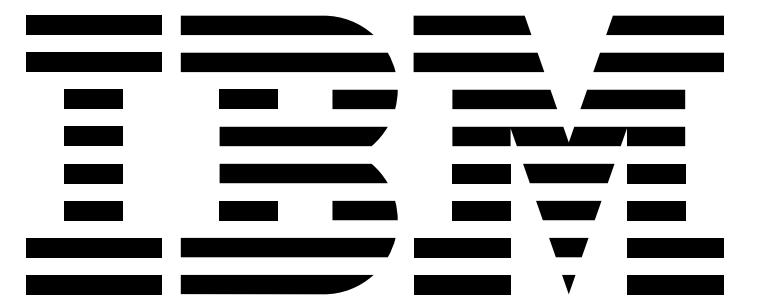
## Course Pre-reqs

- Stat 404: Statistical Computing
- Stat 405: Data Management

## Goals

- Gain a broad understanding of the many ‘tools’ in Data Science
- Practice curiosity driven data science
- Take a Data Science project from conception to completion; simulate something an industry data scientist would do

Nate Langholz, Ph.D.  
Statistician - Data Scientist



# Peiyu Yu

3rd Year Stat PhD working on  
Generative AI



Class Intros...

## Class Intros

- Name
- Background/Current Job
- Why this class? Why this program?
- Anything particular you'd like to learn?

## Course Info and Logistics

<https://github.com/natelangholz/stat418-tools-in-datasience-2025>

or ask on Slack...

<https://uclastat418cl-ubw9419.slack.com/>

[Week 1](#) [4/1]: Introduction to course and each other. Overview of Data Science tools. Introduction of containerization and installation of Docker.

[Week 2](#) [4/8]: Data Science in the command line. Learning about Unix. Reproducible research/work through git/Github and Docker.

[Week 3](#) [4/15]: More Data Science in the command line.

[Week 4](#) [4/22]: Acquiring data through APIs, web-scraping, and OCR. Rvest and Beautiful Soup in Python.

[Week 5](#) [4/29]: Tools for data visualization: shiny dashboards and plotly.

[Week 6](#) [5/6]: **Final Project proposal presentations.** Machine Learning libraries in both R and Python. Introduction Retrieval Augmented Generation.

[Week 7](#) [5/13]: Continuation of Machine Learning libraries with use of cloud services.

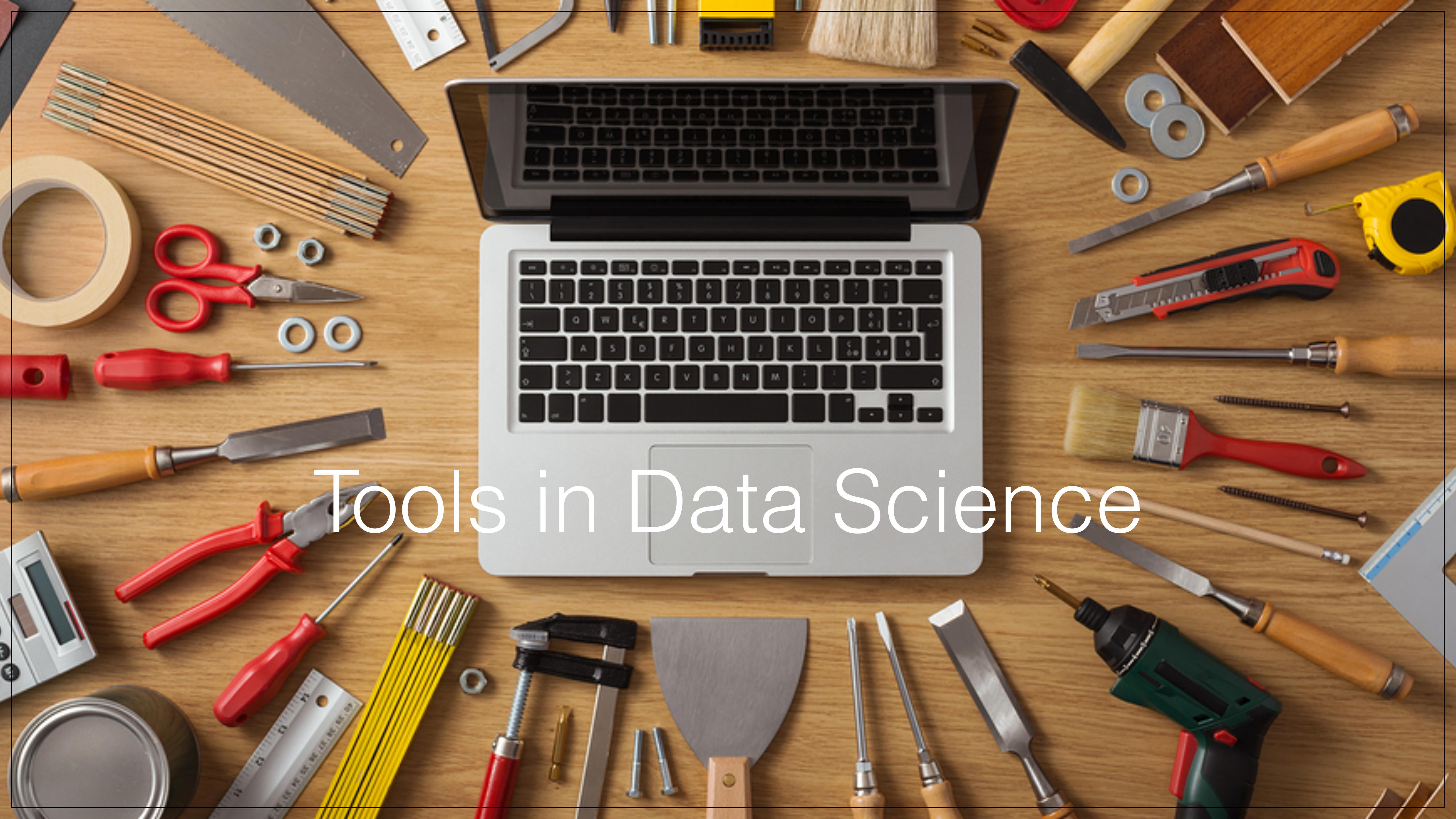
[Week 8](#) [5/20]: Building APIs for model deployment. Exploration of Plumber (R) and Flask (Python)

[Week 9](#) [5/27]: Continuation of API construction for model deployment.

[Week 10](#) [6/3]: **Student Final Presentations.**

All subject to change. Also, two confirmed external speakers with dates TBD.

Data Science?  
(and how it's changing)



# Tools in Data Science



<http://github.com>

<https://git-scm.com/book/en/v2>

<http://happygitwithr.com>



<https://www.docker.com/get-started>





Development Platforms & aids



**XGBoost**



**TensorFlow**

PyTorch

Broad-stoke machine learning  
frameworks



NLP libraries

spaCy



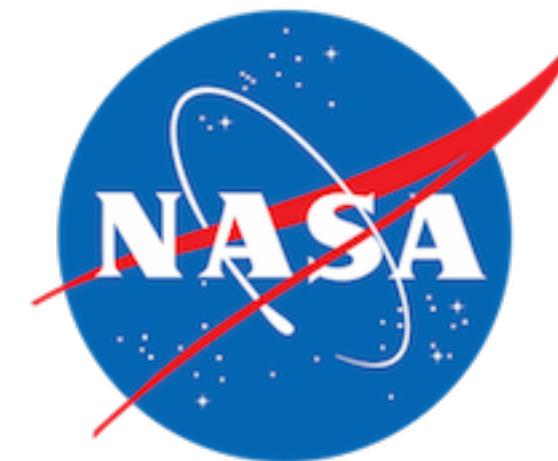


**SQLAlchemy**



Data storage, acquisition  
manipulation

Apis for acquiring data



PUBLIC APIs



BeautifulSoup



Docling

Acquiring more data



Visualization



watsonx™



Cloud based services

**Gemini**



 **deepseek**

 **LLaMA** 

 **Claude**

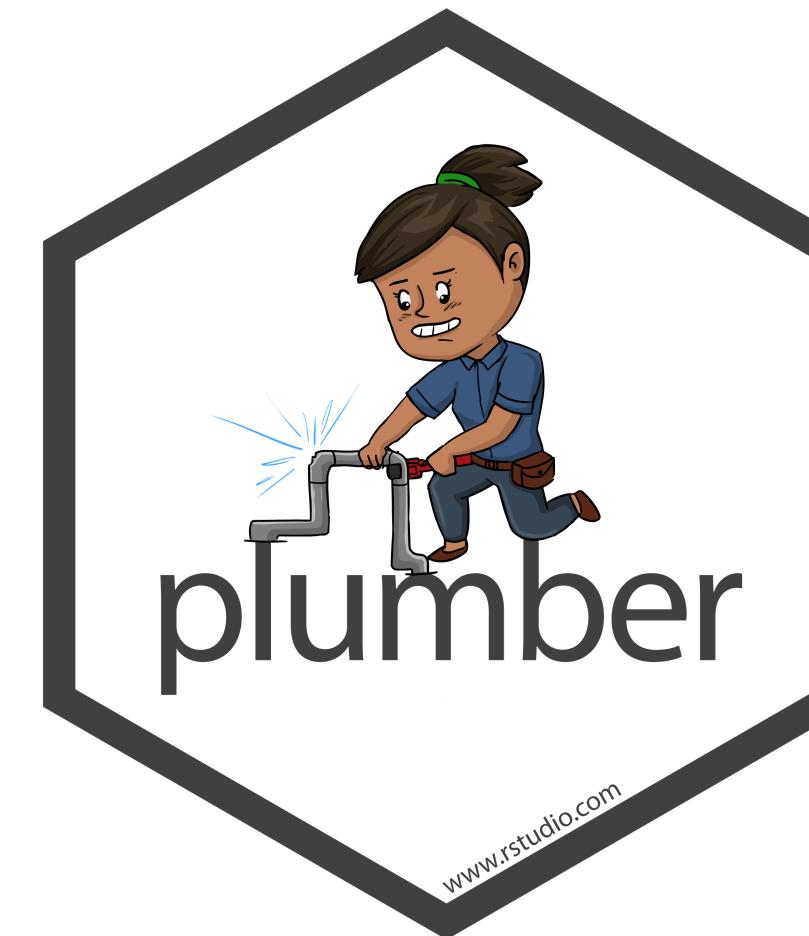
AI & LLMs



**Hugging Face**



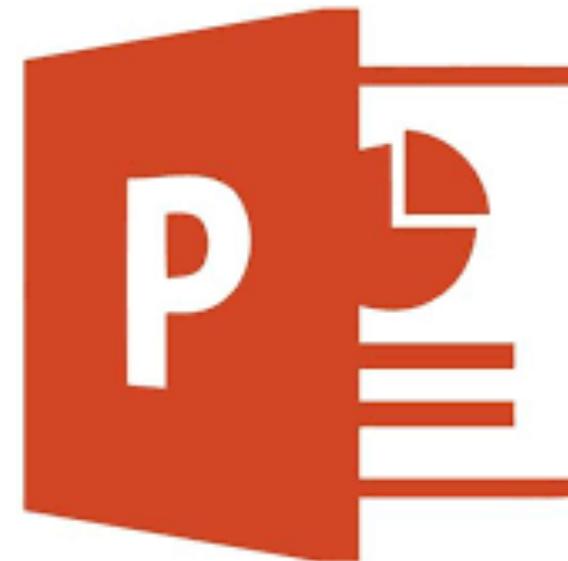
**Ollama**



creating APIs



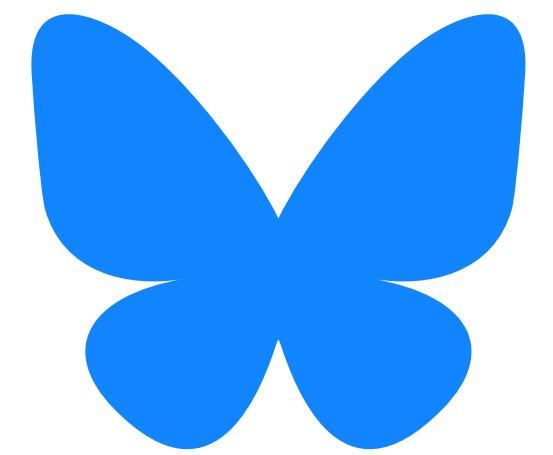
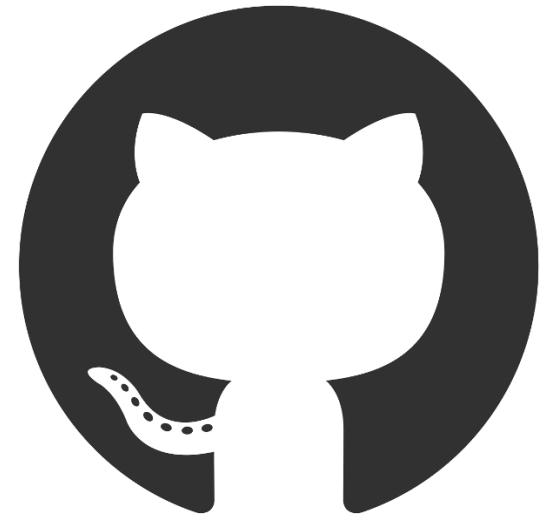
communication  
Slack Channel  
all-uclastat418-class-2025



effective communication skills

critical skill for data scientist requiring  
practice and feedback to develop

marketing



Why do public work?

# Let's get started with docker



<https://www.docker.com/get-started>

Download and install docker

Then lets try to get an RStudio container running