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.

MAS Departmental Disclaimers:

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

Taking or auditing 400 courses is simply is 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

Ruochen Jiang

Mathematician - Statistician

Class Intros…

Class Intros

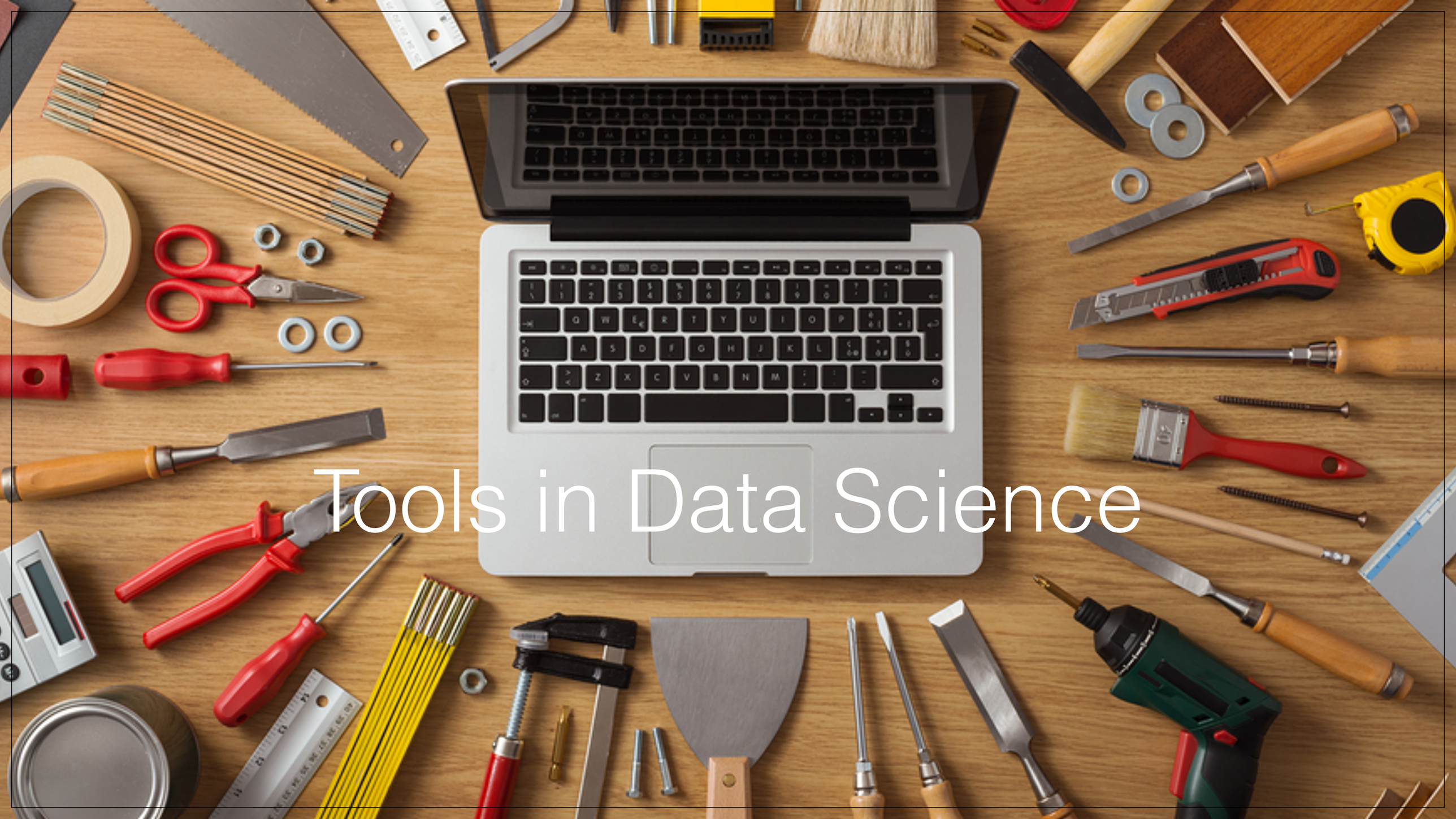
- Name
- Background/Current Job
- Why this class? Why this program?
- Anything particular you’d like to learn?

Course Info and Logistics

https://github.com/natelangholz/stat418-tools-in-datascience

or ask on Slack…
uclastat418-class

What is Data Science?
(and how it's changing)

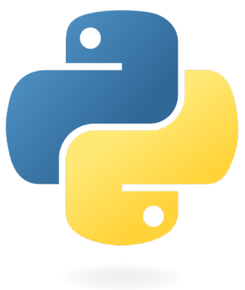

Tools in Data Science

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

http://github.com

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

http://happygitwithr.com



python™

BASH

THE BOURNE-AGAIN SHELL

SQLAlchemy

Data storage, acquisition
manipulation

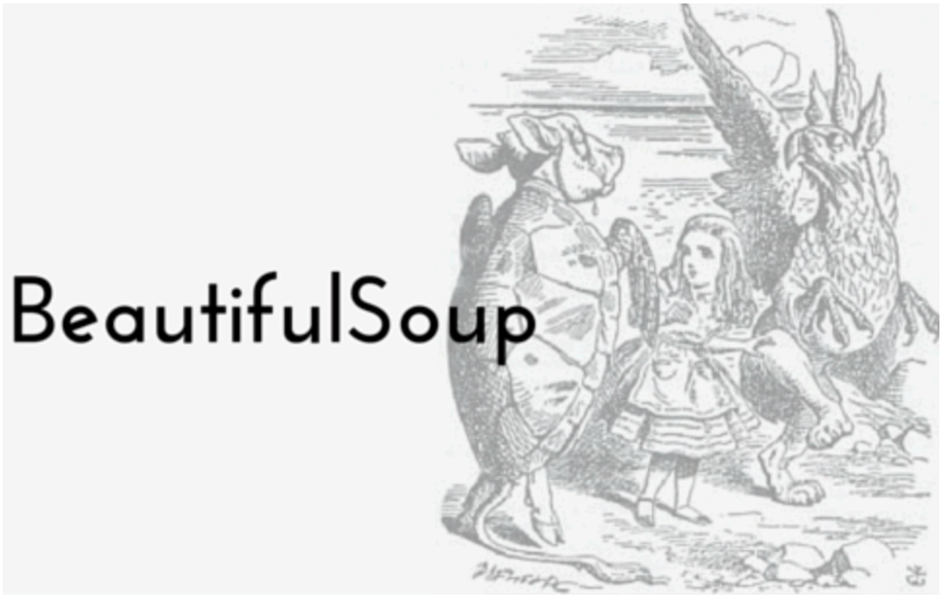
Apis for acquiring data

yelp fusion

The New York Times Developers

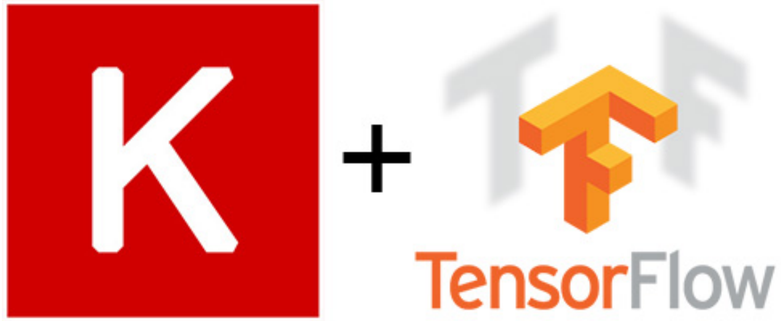
Spotify®

BreweryDB
by SteadyServ Technologies



Acquiring more data

Visualization



Broad-stoke machine learning frameworks

NLTK

Natural Language Toolkit

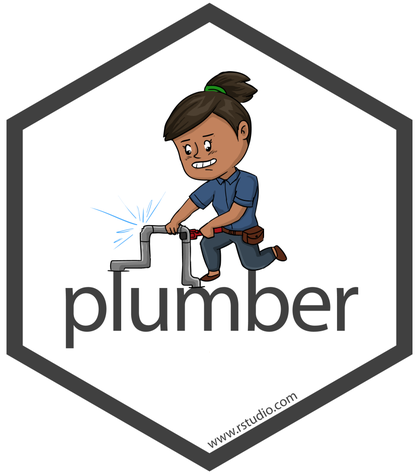
Reference Guide

NLP libraries

spaCy

Google AI

Cloud based services



creating APIs

communication

Slack Channel
uclastat418-class

slackbots?

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 try to get an RStudio container running