

Discussion: Tools

Data Science Education Community of Practice
DSECOP Workshop

June 27, 2023

Mohammad Soltanieh-ha
Clinical Assistant Professor
Information Systems Department
Boston University



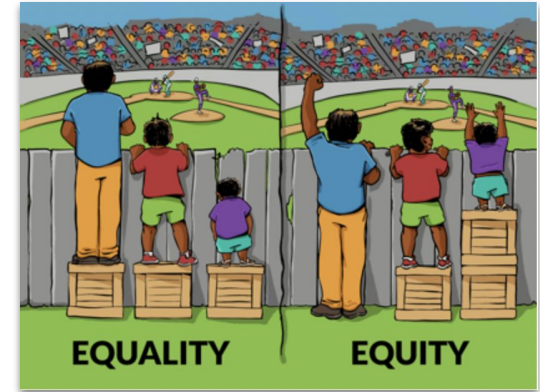
Data science education

- Hardware
 - Laptop
 - Cloud
- Software
 - Open source
 - Third party/cloud
- Materials (a few useful examples)
 - [A Whirlwind Tour of Python \(GitHub\)](#) & [Python Data Science Handbook \(GitHub\)](#) by Jake Vanderplas
 - [Deep Learning with Python, 2nd Edition \(GitHub\)](#) by Francois Chollet
 - www.DataCamp.com. Apply for a 6-month classroom premium access:
<https://www.datacamp.com/groups/classrooms>
 - cloudskillsboost.google (previously [Qwiklabs](#)). Apply [here](#) for educational credits.

The issue of equity in data science

Equity is the provision of personalized resources needed for all individuals to reach common goals. ~Dr. Laura Latta

Problem: **Inadequate hardware**, **ineffective software**, **data access**, **slow internet connection**, and **poor infrastructure** can slow down or block individuals or organizations.



Source: ([Interaction Institute for Social Change, 2016](#))



Cloud computing addresses these issues
(to some extent)

Software

- Programming (python)
- Coding environment (Jupyter notebooks on Colab)
- Version control (GitHub)
- Data processing (numpy, pandas)
- Visualization (matplotlib, seaborn)
- Machine learning (scikit-learn, tensorflow/keras/pytorch)
- Database (BigQuery)

Ad-hoc, small-data environments

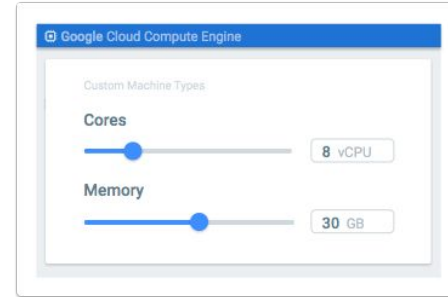
- Google Colab notebooks
 - Access to free GPU and TPU
 - [Intro to Colab](#) by Maxim Ziatdinov & Mohammad Soltanieh-ha
- mybinder.org/
- RStudio Cloud



Create VMs that are right for your workloads

- Machine type options to consider:
 - Higher proportion of memory to CPU
 - Higher proportion of CPU to memory
 - Blend of both

- With AI Platform notebooks students can customize their hardware and get pre-installed software specialized for the task at hand, in less than a minute.



ChatGPT as a educational tool

Example prompts:

1. Write three sets of code blocks in python that compare the speed of numpy against core python. For python consider both a for loop and a list comprehension. In addition, assume this is for a Colab environment (Jupyter notebook) so use the %%timeit magic function to measure the cpu time.
2. Can you provide an example that compares the execution time of matrix multiplication between numpy and standard Python using the %%timeit cell magic function in a Jupyter notebook?

Other ways to leverage ChatGPT in classroom:

- Debug code
- Comment code
- Improve readability or efficiency
- Explain code
- Write quiz problems

Code Sharing

- GitHub
 - Share your code with the students
 - Let them collaborate on projects using their own private repos
 - Project management, issue tracking, wiki, automatic and documentation tools integrated
 - Github for education free premium account
 - <https://education.github.com/>
 - GitHub now also offers unlimited private repositories on its free tier!
- GitHub integration with Colab



DSECOP: Data Science Education Community of Practice

Preparing students for multiple career paths by offering teaching materials to faculty members who teach undergraduate and graduate physics courses.



GitHub: bit.ly/DSECOP-GitHub



Table of Contents

- [Introduction to Data Science Libraries](#) by Julie Butler (2023)
- [Symbolic Regression](#) by Joseph Dominic Lap (2023)
- [Connecting MonteCarlo to Modern AI](#) by Ashley Dale (2023)
- [Time Series Analysis and Forecasting](#) by Connor Robertson (2023)
- [Intro to Classification Algorithms](#) by Richard Harry (2023)
- [Automated Object Detection](#) by Karan Shah (2023)
- [Intro to Data Processing with Histograms](#) by Radha Mastandrea (2022)
- [Intro to Deep Learning](#) by Fatima Bagheri (2022)
- [Learning the Schrodinger Equation](#) by Karan Shah (2022)
- [NMR Deep Learning](#) by Sebastian Atalla (2022)
- [Solving Differential Equations with NNs](#) by Julie Butler (2022)
- [Spectral Clustering](#) by Cunwei Fan (2022)
- [Exploratory Data Analysis](#) by Radha Mastandrea (2022)
- [Intro to Random Forest](#) by Fatima Bagheri (2022)
- [Singular Value Decomposition](#) by Sebastian Atalla (2022)
- [Machine Learning Workflow](#) by Julie Butler (2022)

Thank you!
Comments?