Lecture 01:

Introduction

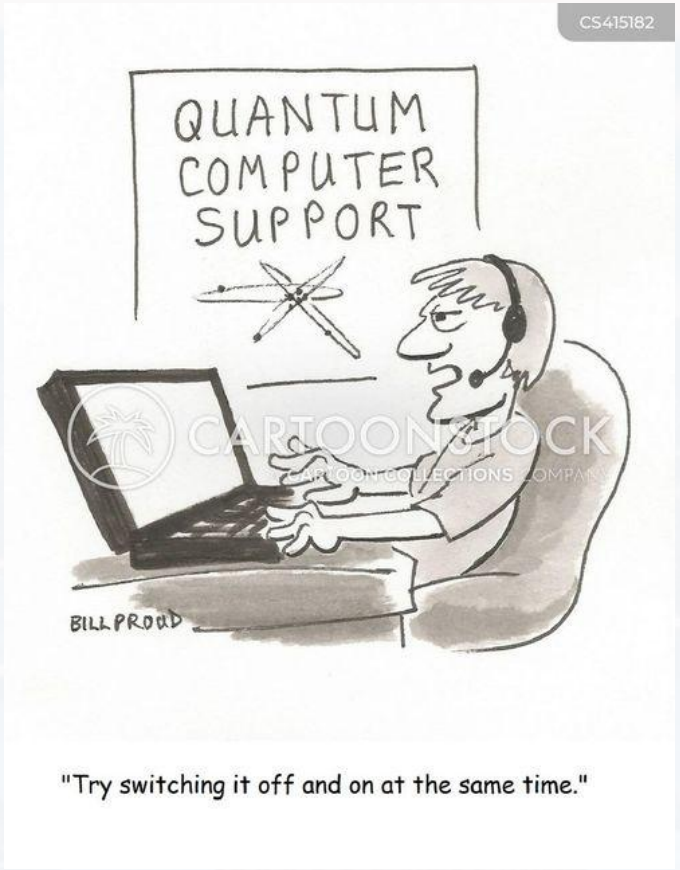
Markus Hohle

University California, Berkeley

Numerical Methods for Computational Science

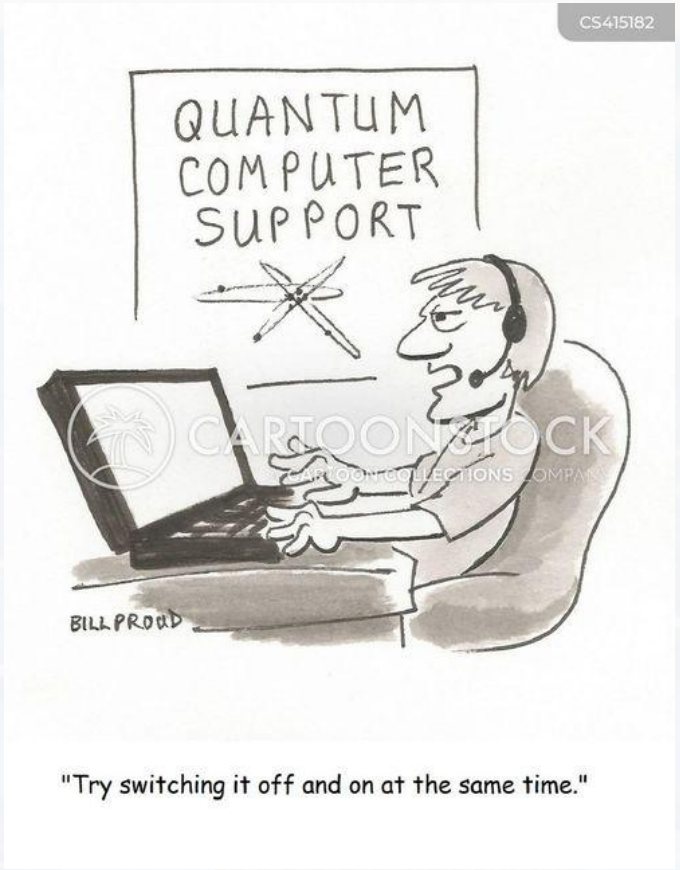
MSSE 273, 3 Units

Berkeley MSSE
MOLECULAR SCIENCE & SOFTWARE ENGINEERING



Outline

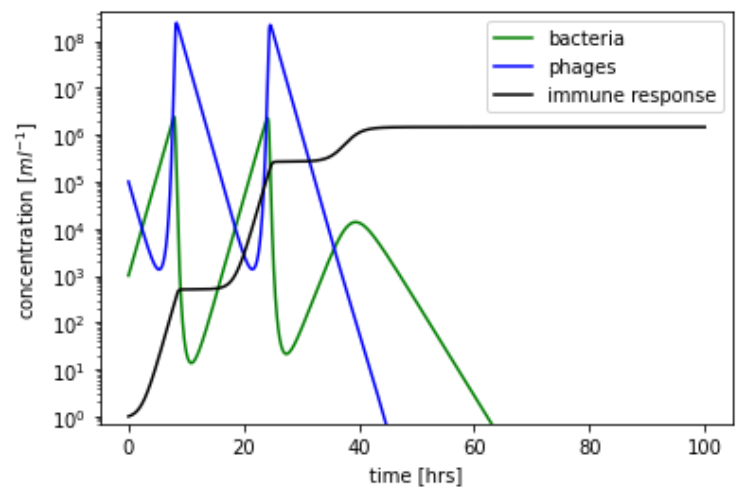
- Motivation
- Course Map
- Python Libraries
- Lecture Exercise



Outline

- Motivation
- Course Map
- Python Libraries
- Lecture Exercise

Why Numerical Methods?

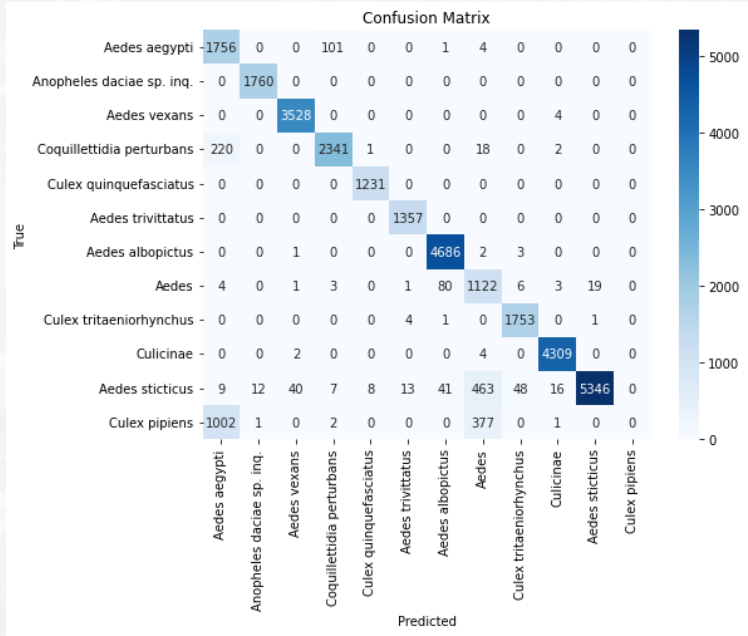


Leung & Weitz 2016

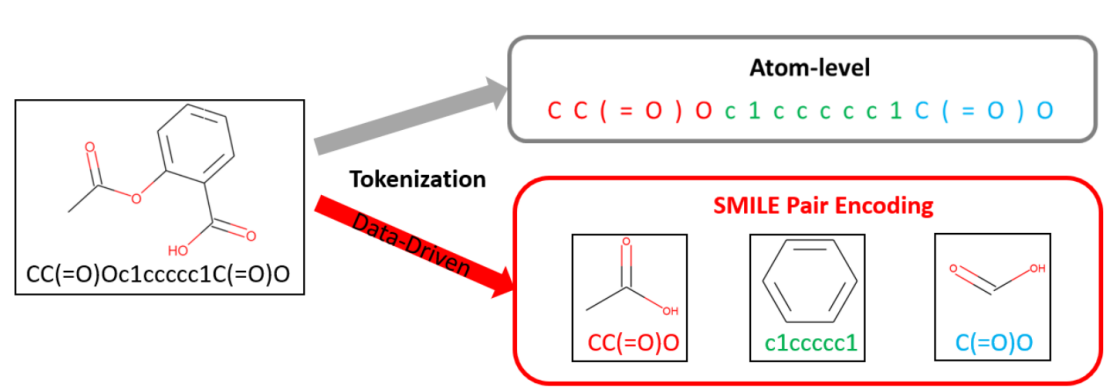
drug development

AI driven data analysis (lecture exercise, see Chem 277B)

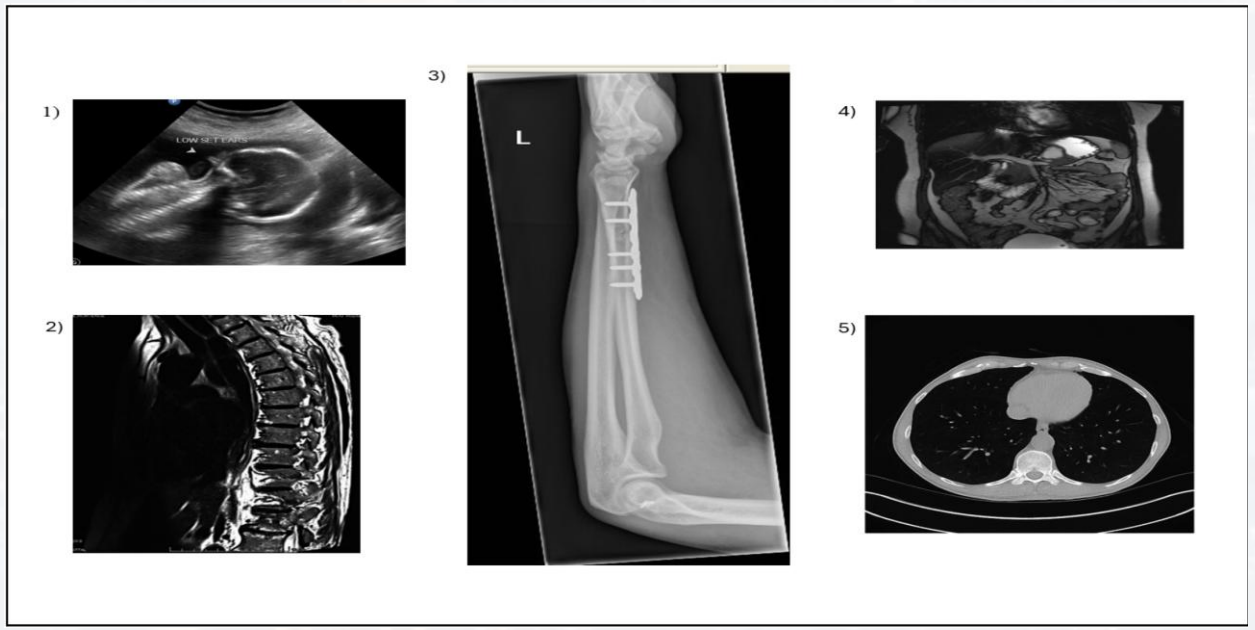
```
>BEISA025-19|Culex|COI-5P
AACATTATATTTCATTTTTGGTGCTTGAGCAGGAATAATTGGAACTTCTTTAAGTCTTCTTATTCG
AGCTGAATTAAGTCAACCAGGAGTTTTTATTGGGAATGATCAAATTTATAATGTAATTGTTACAGC
TCATGCTTTTATTATAATTTTTTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATT
AGTTCCTTTAATACTAGGAGCTCCTGATATAGCTTTTCCTCGAATAAATAATATAAGATTTTGAAT
ACTTCCCCCCTCATTAACACTTCTACTTTCTAGTAGTATAGTAGAAAATGGAGCTGGTACAGGTTG
AACAGTATATCCTCCTCTTTCTTCTGGAACAGCTCATGCTGGAGCTTCTGTTGATTTAGCTATTTT
TTCTTTACATTTAGCCGGAATTTCTTCAATTTTAGGAGCTGTAAATTTTATTACTACTGTAATTAA
TATGCGATCTTCTGGTATTACCCTTGATCGAATACCTTTATTTGTTTGATCAGTTGTAATTACTGC
TATTCTTTTATTATTATCTCTTCCTGTTTTAGCTGGAGCTATTACTATATTATTAACAGATCGTAA
TTTAAATACTTCTTTTTTCGATCCTATTGGAGGAGGAGATCCTATTTTATATCAACATTTATTT
>BEISA121-19|Anopheles|COI-5P
AACATTATATTTTATTTTCGGTGCTTGAGCAGGAATAGTAGGAACTTCTTTAAGTATTCTTATTCG
```



Why Numerical Methods?

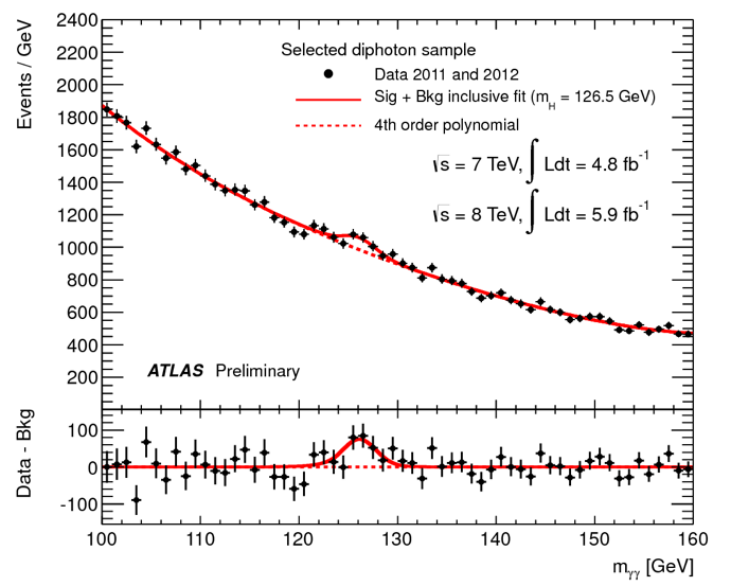


automating workflows (Mass Spec to SMILES, Chem 277B project, **Casey Tomlin, Dulce Torres, Esther Mathew, Jesse Maki, Marie Anand**)

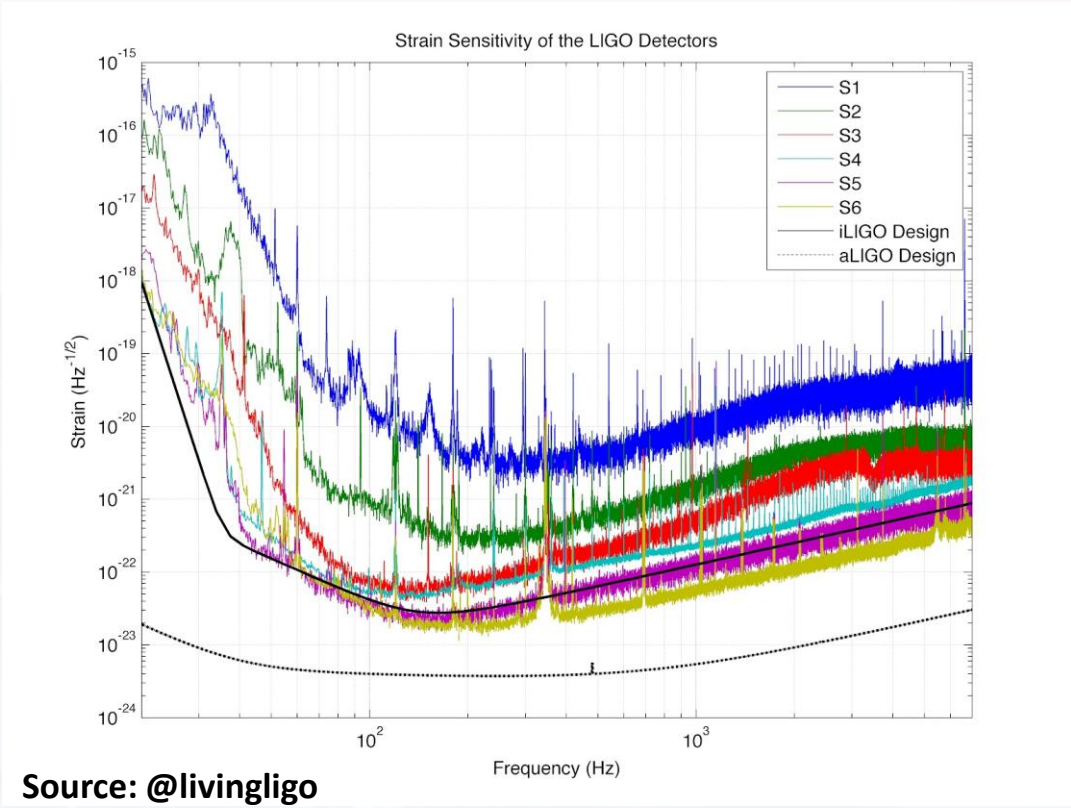


automating workflows (image to diagnosis, Chem 277B project, **Elle MacLennan, Carmen Matar, Timothy Nguyen, Brandon Ton**)

Why Numerical Methods?



Source: CERN/ATLAS

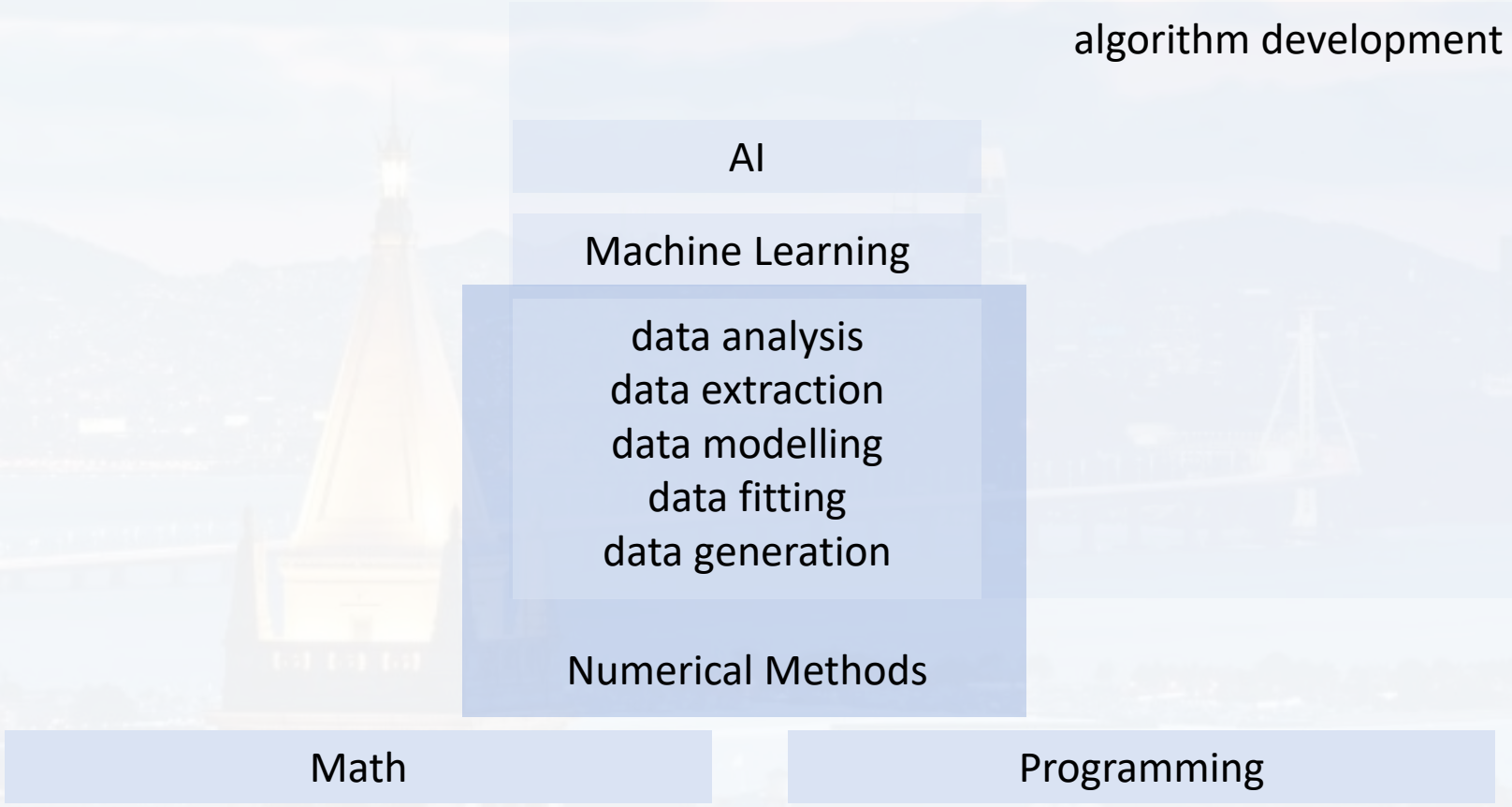


Source: @livingligo

- data analysis
- data extraction
- data modelling
- data fitting
- data generation

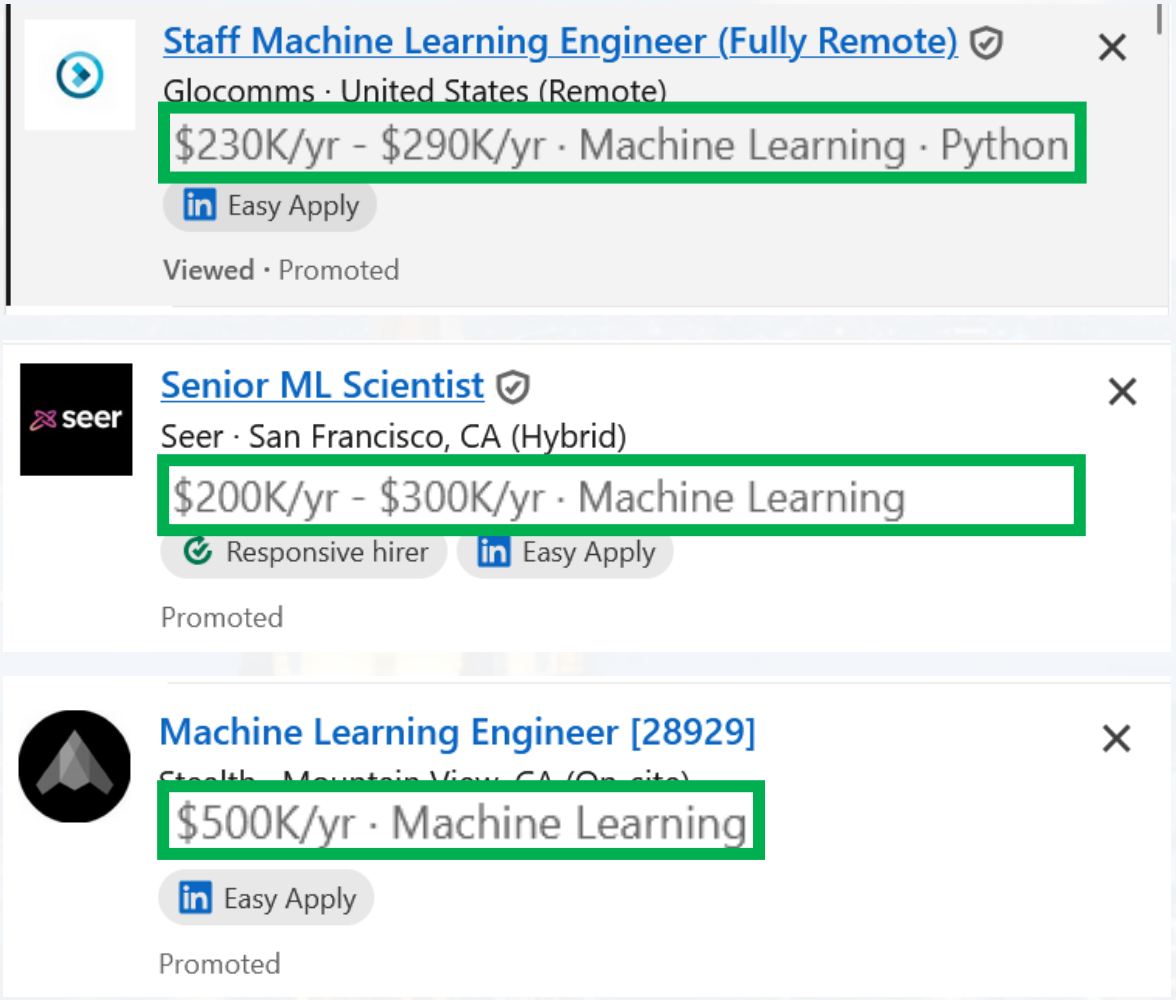
math + coding

Why Numerical Methods?



Why Computational Methods?

LinkedIn: 1st three Job Postings, Jan 10th 2025



Why Computational Methods?

Listen up son, there are three languages you need to learn in order to become successful:

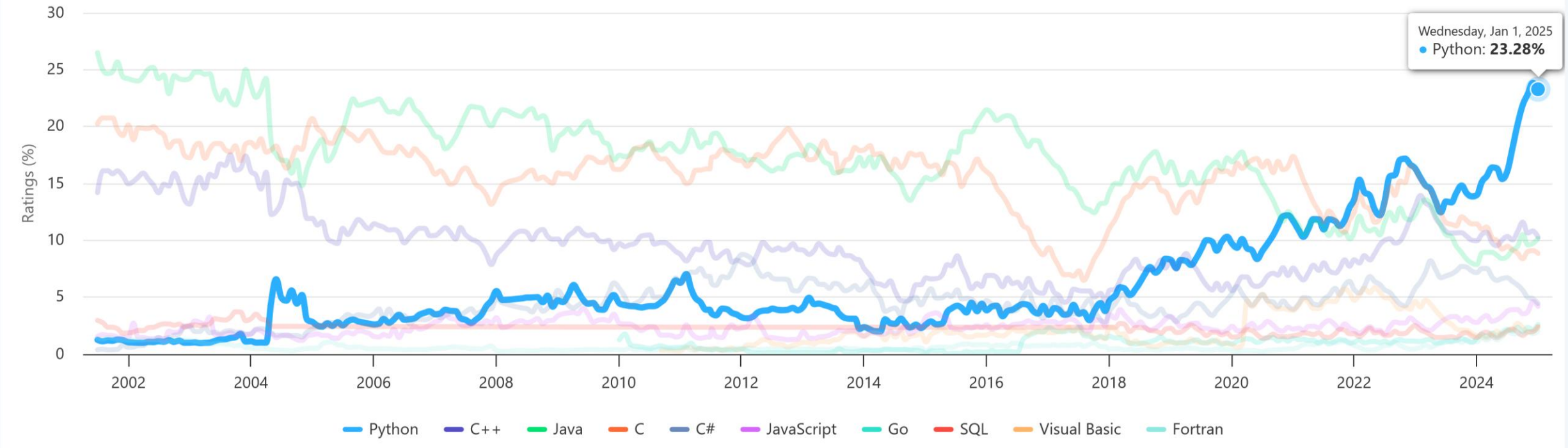
- English
- Math
- Python *)

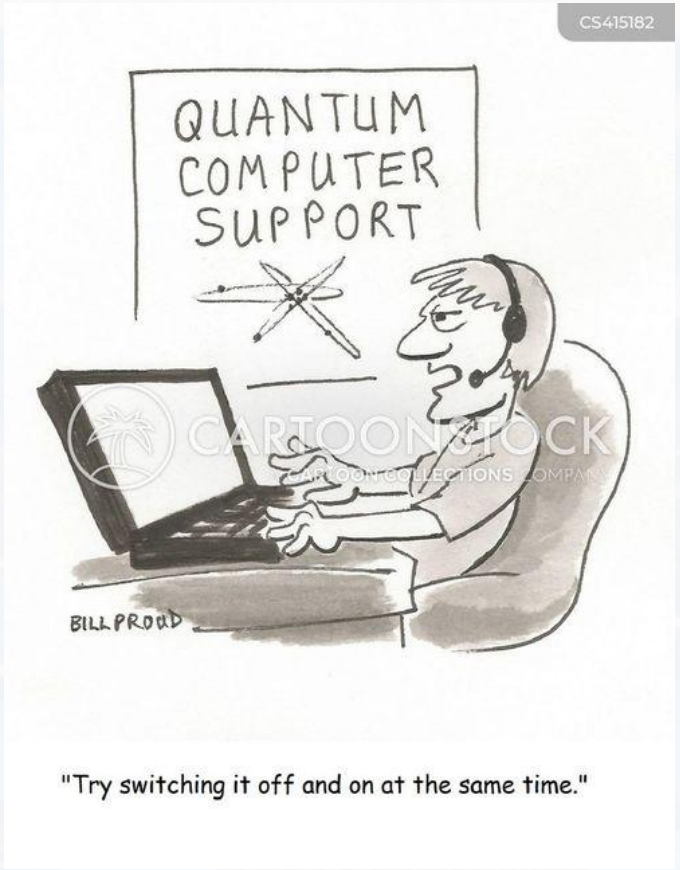
*) or C++

TIOBE index, Jan 2025

Jan 2025	Jan 2024	Change	Programming Language	Ratings	Change
1	1		Python	23.28%	+9.32%
2	3	↑	C++	10.29%	+0.33%
3	4	↑			
4	2	↓			
5	5		C#	4.45%	-2.71%
6	6		JavaScript	4.20%	+1.43%
7	11	⇈	Go	2.61%	+1.24%
8	9	↑	SQL	2.41%	+0.95%
9	8	↓	Visual Basic	2.37%	+0.77%
10	12	↑	Fortran	2.04%	+0.94%
11	13	↑	Delphi/Object Pascal	1.79%	+0.70%
12	10	↓	Scratch	1.55%	+0.11%
13	7	⇊	PHP	1.38%	-0.41%
14	19	⇈	Rust	1.16%	+0.37%
15	14	↓	MATLAB	1.07%	+0.09%

TIOBE index, Jan 2025





Outline

- Motivation
- Course Map
- Python Libraries
- Lecture Exercise

GSI: Elizabeth (Lizzie) Gilson

Toxicology Data Scientist at EPA,

UC Berkeley Alumna (MSSE)

Lecturer: Markus Hohle

Lecturer at UC Berkeley &

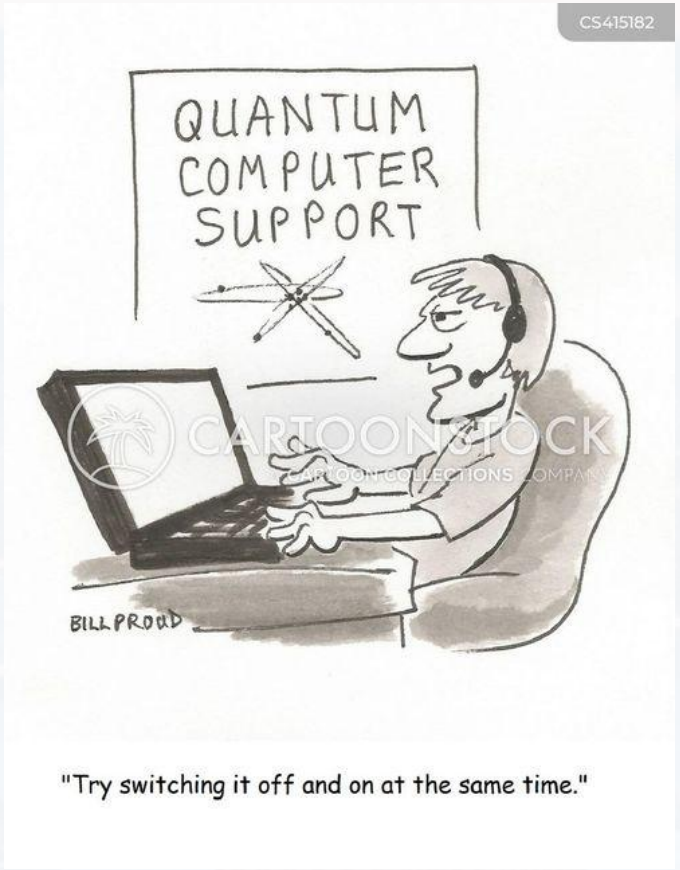
Data Analysis Consultant

PhD Physics

Lecture:	Monday,	5:00 – 8:00pm PT
Discussion (Lizzie):	Wednesday,	5:30 – 6:30pm PT
Lab Session (Markus):	Wednesday,	6:30 – 7:30pm PT
Office Hours (Markus):	Friday,	5:00 – 7:00pm PT
Material:	Recorded Lectures:	bcourses
	codes/slides	bcourses & GitHub
	HW assignments	bcourses
Grades:	HW Assignments:	40%
	Programming Projects	20%
	Lecture Exercises	20%
	Discussion & Lab Participation	20%

Course Start Date: **Monday May 19th, 2025**
Course End Date: **Friday, August 15th, 2025**

Week 1:	Introduction to Scientific Computing and Python Libraries
Week 2:	Linear Algebra Fundamentals
Week 3:	Vector Calculus
Week 4:	Numerical Differentiation and Integration
Week 5:	Solving Nonlinear Equations
Week 6:	Probability Theory Basics
Week 7:	Random Variables and Distributions
Week 8:	Statistics for Data Science
Week 9:	Eigenvalues and Eigenvectors
Week 10:	Simulation and Monte Carlo Method
Week 11:	Data Fitting and Regression
Week 12:	Optimization Techniques
Week 13:	Machine Learning Fundamentals



Outline

- Motivation
- Course Map
- Python Libraries
- Lecture Exercise

1) reading files (.xlsx, .xls, .csv, .txt, …)

`pandas` (standard)

+ analyzing/evaluating/manipulating data frames

faster than pandas, but fewer functions

- `dask`

- `polars`

- `fireducks`

1) reading files (.xlsx, .xls, .csv, .txt, …) pandas (standard)

+ analyzing/evaluating/manipulating data frames

pandas	excel	104.00sec
polars	excel	10.17sec
pandas	csv	1.39sec
polars	csv	0.14sec
dask	csv	0.02sec

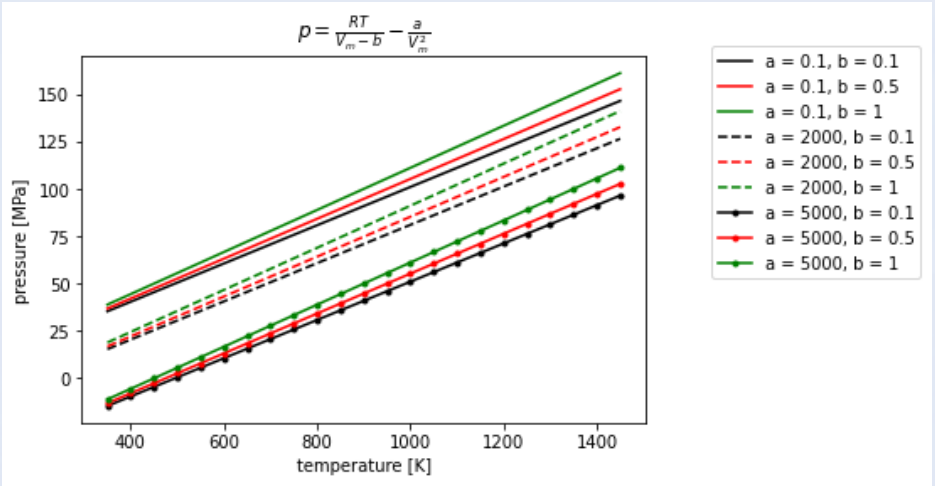
factor of 4000 – 5000!!

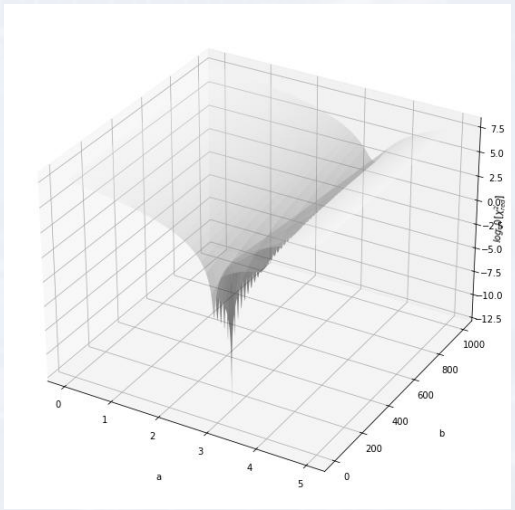
FireDucks

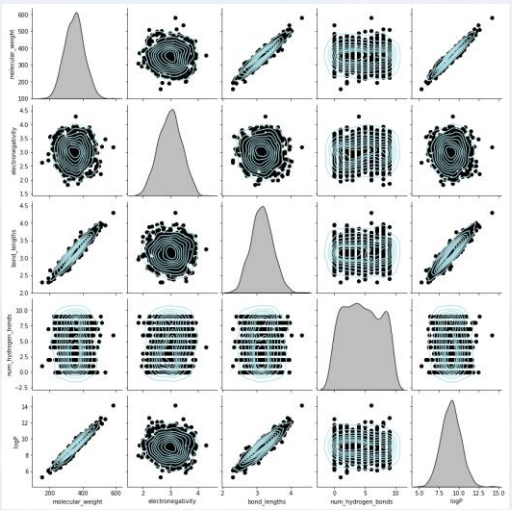
1) **reading files (.xlsx, .xls, .csv, .txt, …)** pandas (standard)

2) **plotting**

matplotlib







seaborn

1) **reading files (.xlsx, .xls, .csv, .txt, …)** `pandas` (standard)

2) **plotting** `matplotlib, seaborn`

3) **numerical methods**

`math` basics

```
math.
acos
acosh
asin
asinh
atan
atan2
atanh
```

numpy standard


```
np.random.
hypergeometric
laplace
logistic
lognormal
logseries
mtrand
multinomial
```

1) **reading files (.xlsx, .xls, .csv, .txt, …)** `pandas` (standard)

2) **plotting** `matplotlib, seaborn`

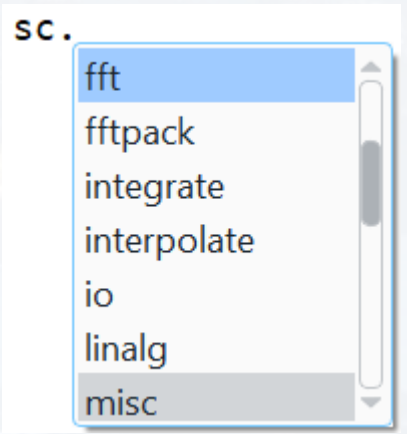
3) **numerical methods**

`math` basics

`numpy` 

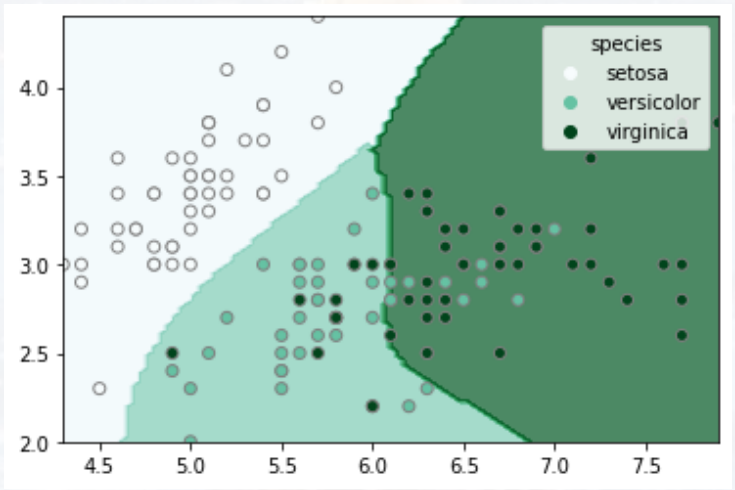
standard

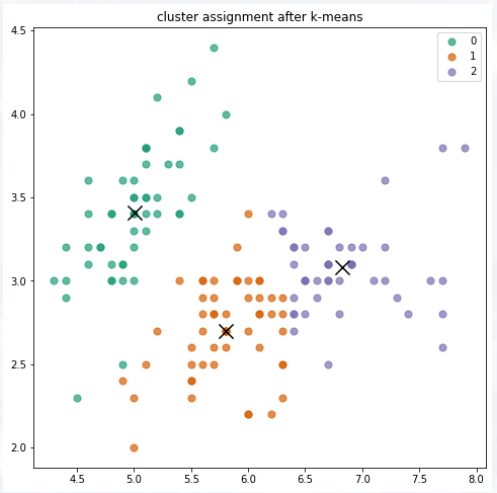
`scipy` 

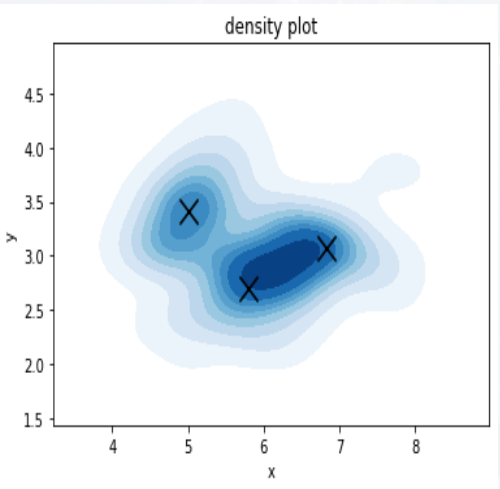


- num. integration/differentiation
- Fourier transformation
- optimization
- curve fitting …

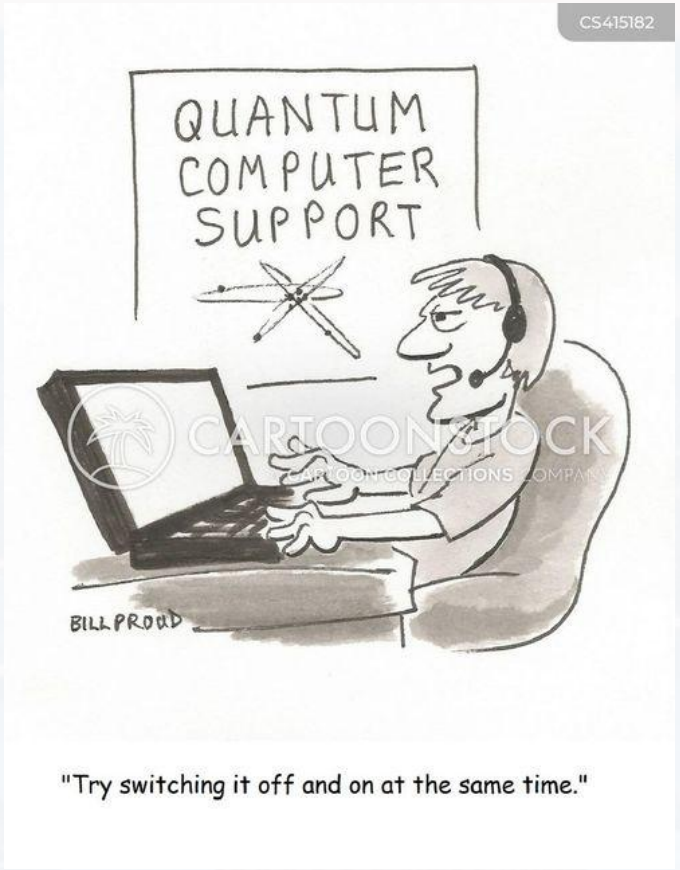
1) reading files (.xlsx, .xls, .csv, .txt, …)	`pandas` (standard)
2) plotting	`matplotlib, seaborn`
3) numerical methods	`math, numpy, scipy`
4) machine learning	`scikitlearn`







1) reading files (.xlsx, .xls, .csv, .txt, …)	`pandas` (standard)
2) plotting	`matplotlib, seaborn`
3) numerical methods	`math, numpy, scipy`
4) machine learning	`scikitlearn`
5) ANN/AI/DeepLearning	



Outline

- Motivation
- Course Map
- Python Libraries
- Lecture Exercise

- reading different file formats (same content)
- benchmarking
- coding warm-up for course
- FYI dynamic programming

see: `01_Lecture_Exercise.ipynb`

- reading different file formats (same content)
- benchmarking
- coding warm-up for course
- **FYI dynamic programming**

see: `01_Lecture_Exercise.ipynb`

```
dfPandasCSV = ReadWithAnyToolAnyMethod()
dfPandasCSV = ReadWithAnyToolAnyMethod(filename = 'Data_set_0.xlsx', my_method = 'read_excel')

dfPandasCSV = ReadWithAnyToolAnyMethod(my_tool = 'dd')

dfPandasCSV = ReadWithAnyToolAnyMethod(my_tool = 'pl')
dfPandasCSV = ReadWithAnyToolAnyMethod(filename = 'Data_set_0.xlsx', my_tool = 'pl', my_method = 'read_excel')
```

```
Total runtime: 2.937999999994645 seconds
Total runtime: 198.5779999999795 seconds
Total runtime: 0.0 seconds
Total runtime: 0.1870000000053551 seconds
Total runtime: 17.0 seconds
```

Thank you very much for your attention!