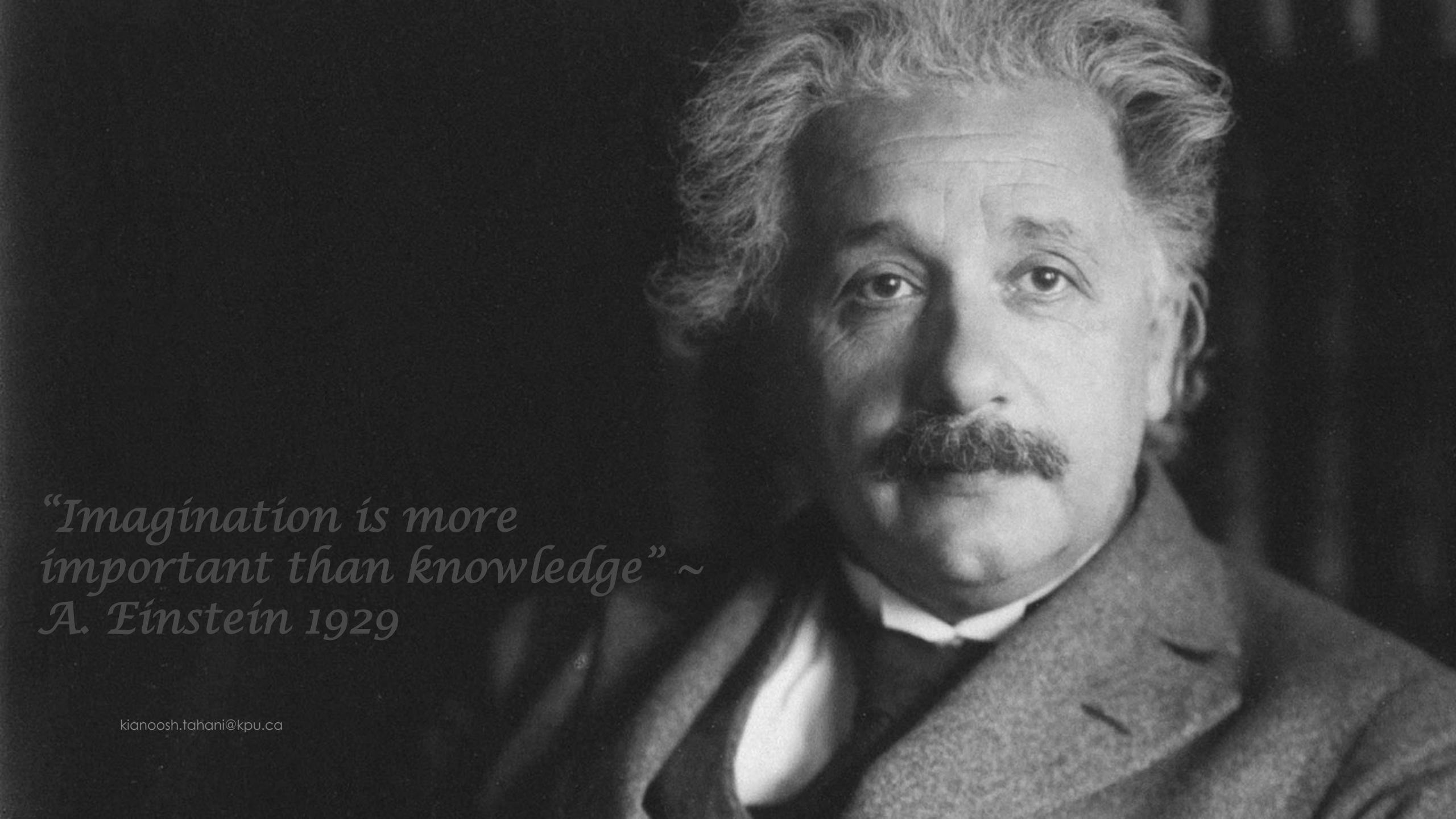


PHYS 4900

Intro to Data Science

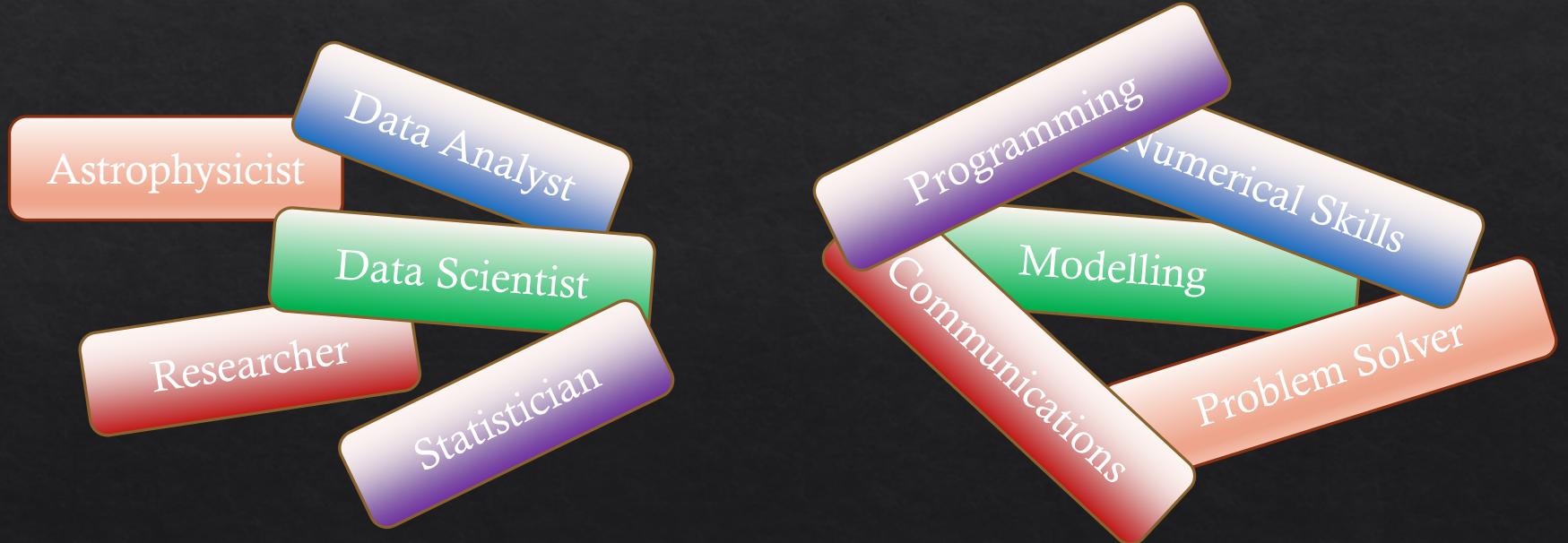
“Imagination is more important than knowledge” ~ A. Einstein 1929

A black and white portrait of Albert Einstein, showing him from the chest up. He has his characteristic wild, wavy hair and a well-groomed, dark mustache. He is looking slightly to the right of the camera with a thoughtful expression. The lighting is dramatic, with strong highlights on his forehead and nose, while the rest of his face and the background are in deep shadow.

*“Imagination is more
important than knowledge” ~
A. Einstein 1929*

To Infinity, And Beyond!

- ❖ By the end of this course, you would have basic Data Science knowledge! However, you need to use your imaginations to determine the applications of your knowledge!



Who am I?



TIMELINE OF THE INFLATIONARY UNIVERSE

Big Bang

In an infinitely dense moment 13.7 billion years ago, the Universe is born from a singularity.

Inflation

A mysterious particle or force accelerates the expansion. Some models inflate the Universe by a factor of 10^{26} in less than 10^{-32} seconds.

Cosmic microwave background

After 380,000 years, loose electrons cool enough to combine with protons. The Universe becomes transparent to light. The microwave background begins to shine.

Dark ages

Clouds of dark hydrogen gas cool and coalesce.

First stars

Gas clouds collapse. The fusion of stars begins.

Galaxy formation

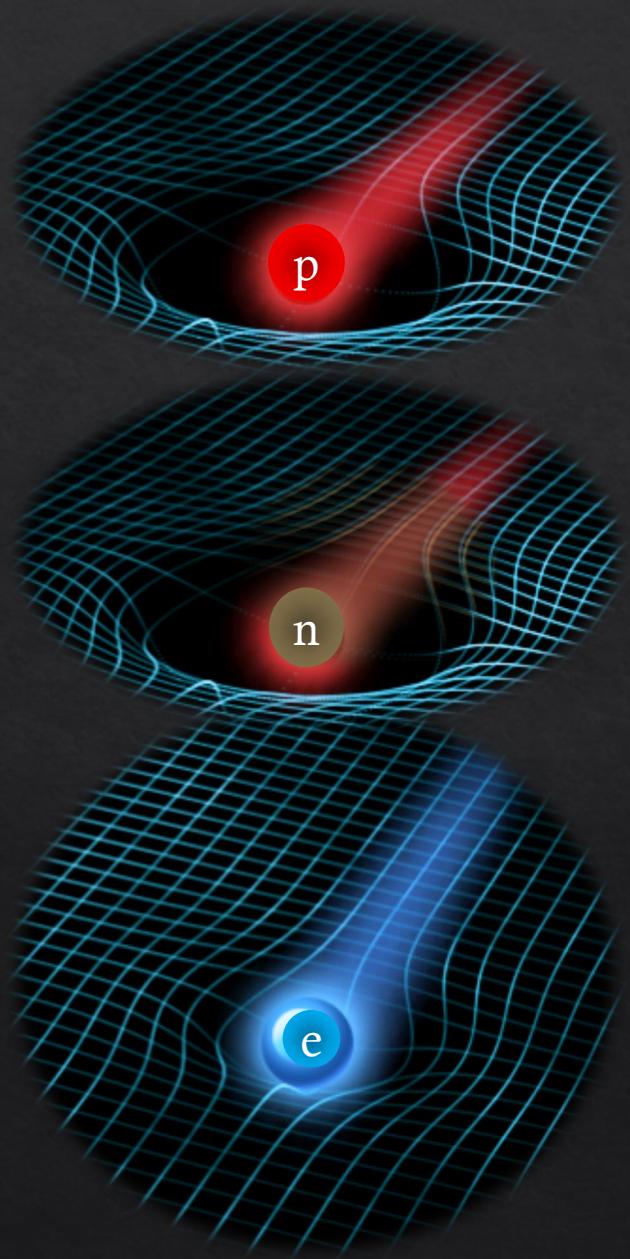
Gravity causes galaxies to form, merge and drift. Dark energy accelerates the expansion of the Universe, but at a much slower rate than inflation.

Big Bang expansion

13.7 billion years

Data/Information

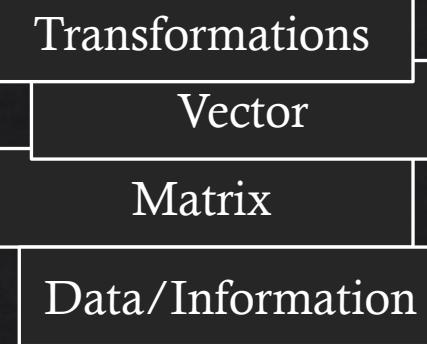


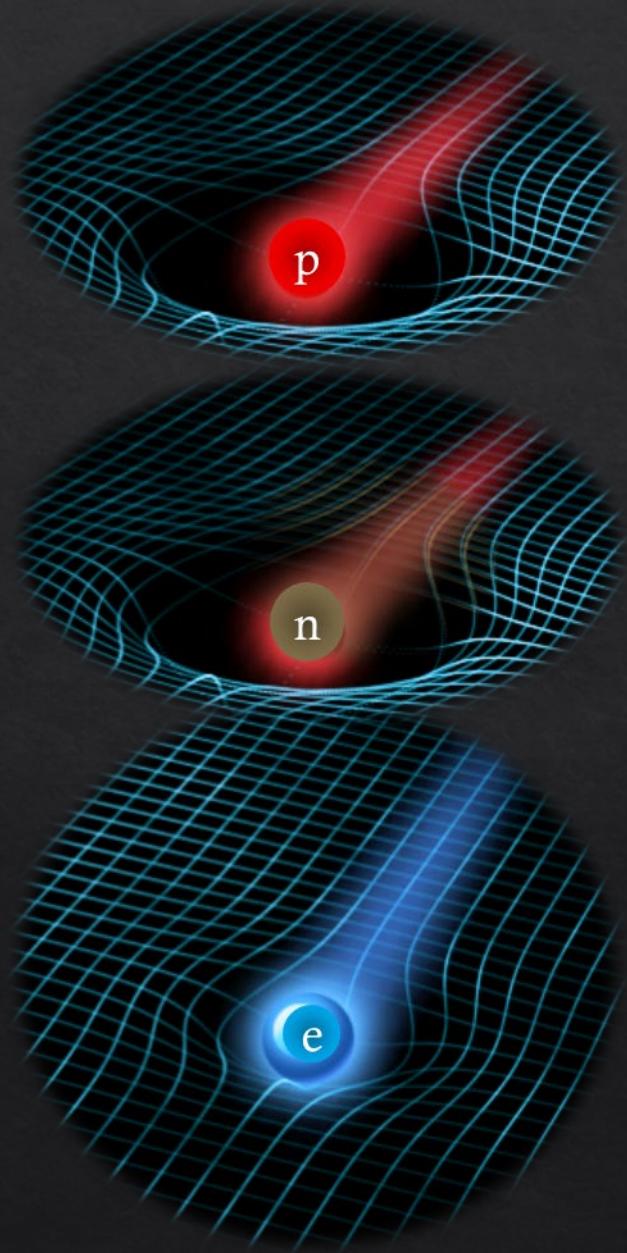


$$\begin{bmatrix} 1(u) \\ 1(u) \\ 0(d) \end{bmatrix}$$

$$\begin{bmatrix} 1(u) \\ 0(d) \\ 0(d) \end{bmatrix}$$

$$\begin{bmatrix} n \\ \ell \\ m_l \\ m_s \end{bmatrix}$$



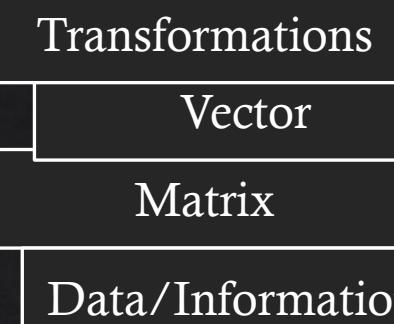
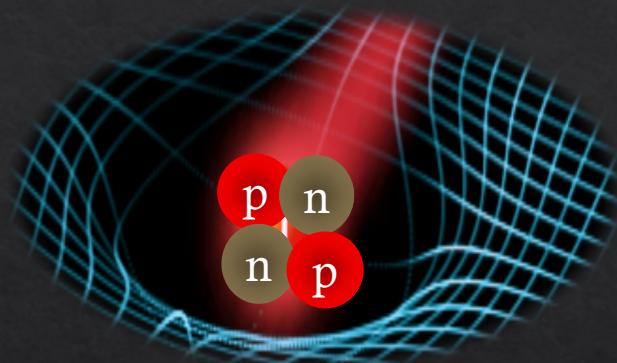


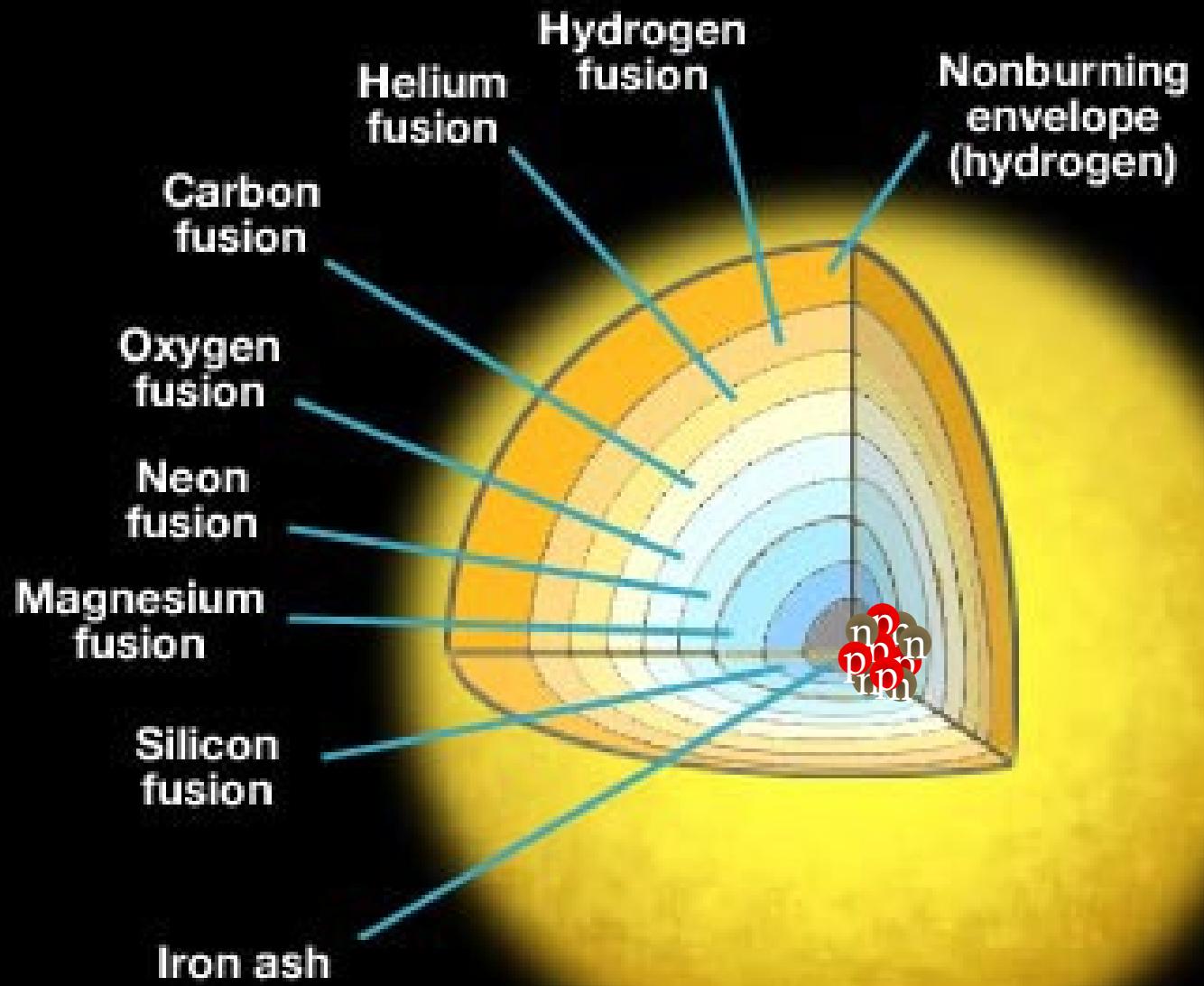
$$\begin{bmatrix} 1(u) \\ 1(u) \\ 0(d) \end{bmatrix}$$

$$\begin{bmatrix} 1(u) \\ 0(d) \\ 0(d) \end{bmatrix}$$

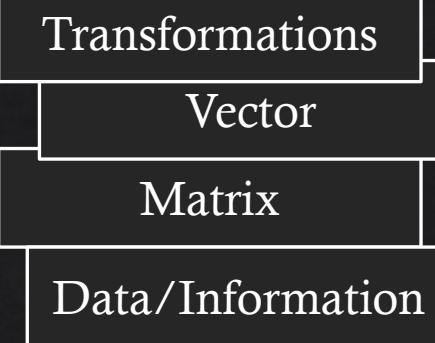
$$\begin{bmatrix} n \\ \ell \\ m_l \\ m_s \end{bmatrix}$$

$$\begin{bmatrix} 1(u) \\ 0(d) \\ 0(d) \end{bmatrix} + \begin{bmatrix} 1(u) \\ 1(u) \\ 0(d) \end{bmatrix} + \begin{bmatrix} 1(u) \\ 0(d) \\ 0(d) \end{bmatrix} + \begin{bmatrix} 1(u) \\ 1(u) \\ 0(d) \end{bmatrix}$$





10011001100110
010011001100110
10011001100110
0011001100110





Clustering

Transformations

Vector

Matrix

Data/Information



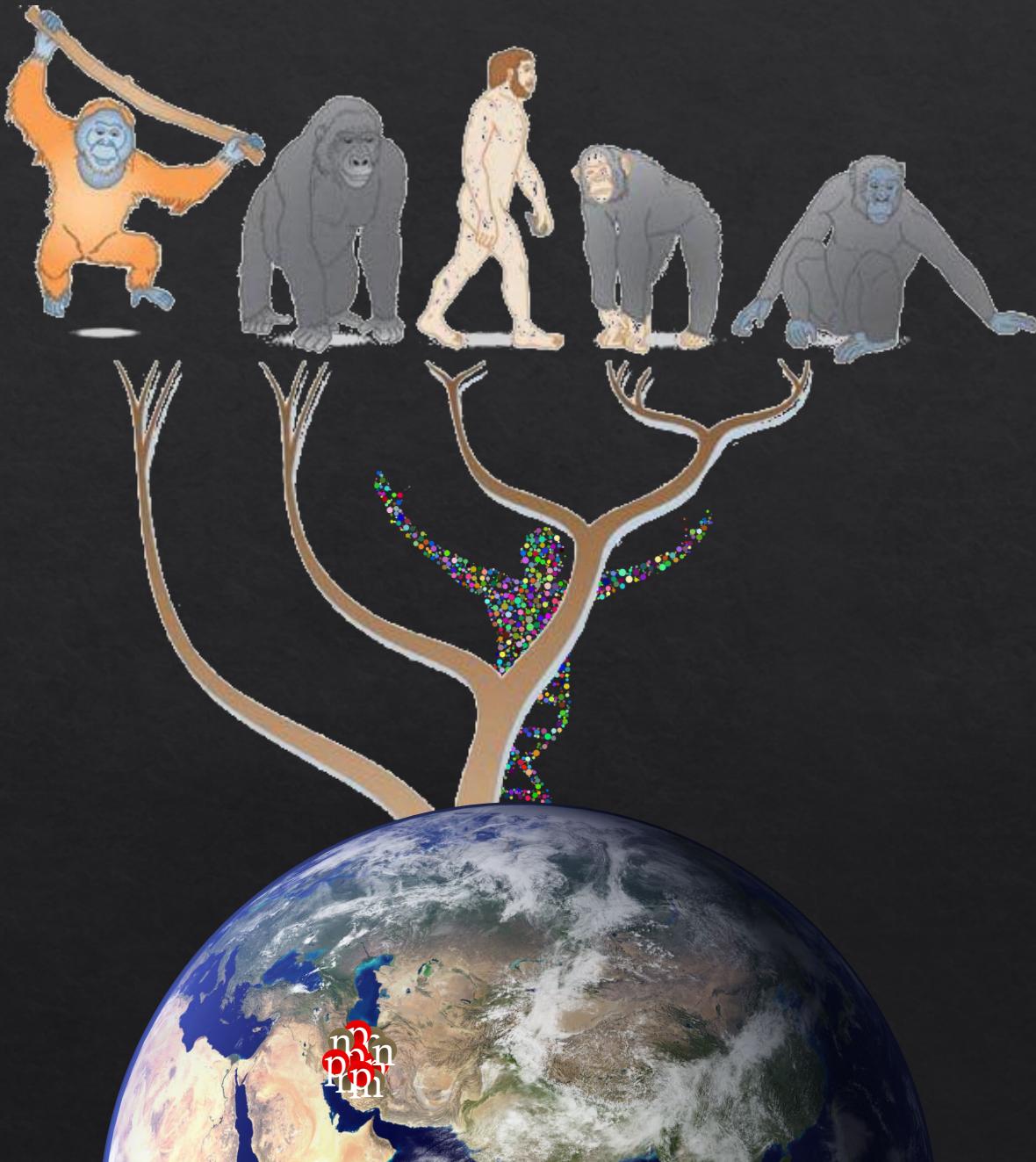
Clustering

Transformations

Vector

Matrix

Data/Information



Introduction to Data Science

WEEK	Theory	Coding
1		Python Bootcamp: <ul style="list-style-type: none"> Environment Set up (Anaconda, Jupyter Notebook) List, Tuple, Set, Dictionary, Boolean, String For loop, If, elif, else Function (args & kwargs)
2	Introducing Your Data/Information + Basics of Star Formation	Python Bootcamp: <ul style="list-style-type: none"> Numpy Pandas
3	Mean Absolute Error, Squared Error, Mean Squared Error, Vector Matrix, Determine, Transformations , Regression	Python Bootcamp: <ul style="list-style-type: none"> Matplotlib.pyplot Seaborn
4	<ul style="list-style-type: none"> Supervised Learning vs Unsupervised, Cost Function, Derivative & Gradient Descent Confusion Matrix Linear Regression	Python Data Science Toolkits: <ul style="list-style-type: none"> Scikitlearn - Cross Validation, - Bias Validation Trade-off
5	<ul style="list-style-type: none"> Logistics Regression Data Scaling Training and Test Data Over Fitting vs Underfitting	Python Data Science Toolkits: <ul style="list-style-type: none"> Scikitlearn - Logistic Regression Train-Test-Split
6	Clustering Decision Tree	Python Data Science Toolkits: <ul style="list-style-type: none"> Scikitlearn - KNN K Means Clustering
7	Neural Network	Tensorflow
8	Review	Review

Introduction to Data Science

- ❖ Kianoosh Tahani
 - ❖ Physicist
 - ❖ Specialized in Star Formation, Astrophysics
 - ❖ Data Science well, just experienced (lots of practice), and lots of **Google-ing!**
- ❖ Throughout this course
- ❖ Data Science is a new topic being offered.
So, the topics to teach are vast and methods of teaching is being developed. I am happy that PMT at KPU is the first to integrate it.



Data Science vs Data Engineer

Understands the
Science Behind the
Data

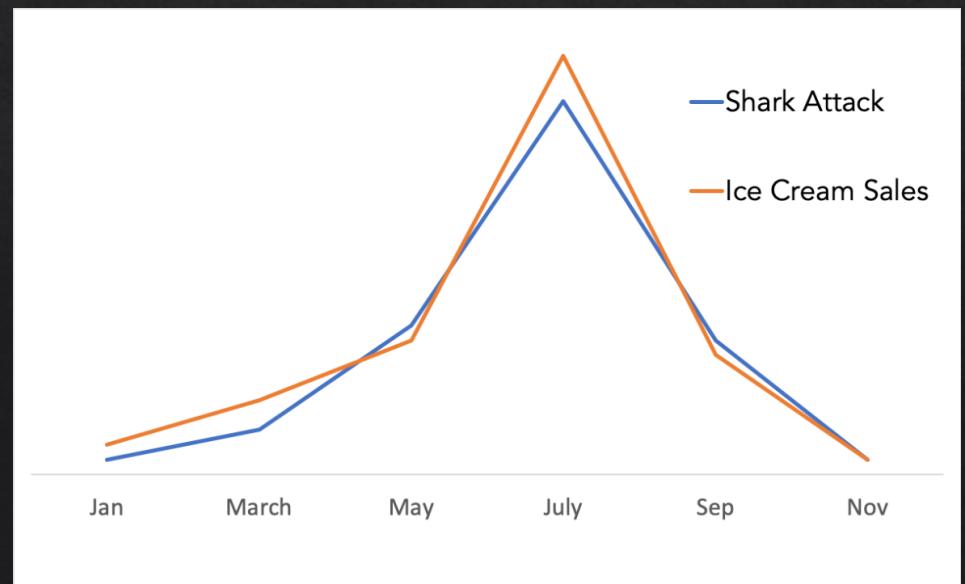
- May involve quite interesting projects
- You have at least 4 years of expertise in science (B.Sc). You developed problem solving skills and understood lots of plots!
- This is exactly what distinguishes you from others!!

Codes properly

I don't and I have no intention to ask you to do same! Except your GitHub!

My Goal in This Course

- ❖ You are among a peer group of scientist! This may cause you to think that everybody is as smart as you are!
- ❖ We live in an era of data. Tons of data is available and not enough people to analyze it! Not everybody understands data and not everybody understands it. It is the job of science major to comprehend it, interpret it, and convey it, in a way that everybody can understand it.



My Goal in This Course

- ❖ You are among a peer group of scientist! This may cause you to think that everybody is as smart as you are!
- ❖ We live in an era of data. Tons of data is available and not enough people to analyze it! Not everybody understands data and not everybody understands it. It is the job of science major to comprehend it, interpret it, and convey it, in a way that everybody can understand it.
- ❖ By the end of this course, you should identify your skillsets!
 - ❖ E,g. Science, Statistics, Programming, Data-Visualization, ...
- ❖ You should also practice on conveying the concept to someone with no background.

Build up the confidence to see your unique and distinguished ability.

Convince any hiring committee to see that.

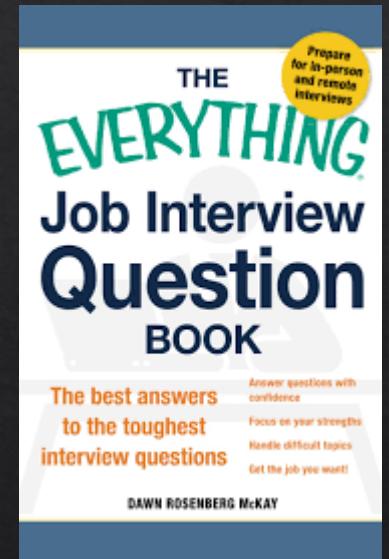
Build up such confidence to believe that you have almost earned all the proper training.

- Strong Science Background
- Strong Research Skills
- Good Computer Programming Skills
- Ability to Work in an Agile Environment
- Ability to Work in a Team
- Most Importantly, Your Willingness to Learn and Your Confidence on Getting the Job Done

These are the general transferrable skills I can think of, and you should more precisely think about your transferrable skills to the job you will be applying! I believe you should be able to convince any hiring committee to see that, and they should be happy to hire you.

Start Preparing Today!

- ❖ <https://paulsandhuindustries.com/top-7-books-to-help-with-job-interviews/>



This is a friendly environment!

- Feel free to get engaged and ask questions
 - Specially if I do not make sense

There are no stupid
questions, just an
intelligent person seeking
information.



Dates

- ❖ Class Hours: (M: 14:00-15:50) and (W: 10:00-11:50) and (W: 13:00-15:50)
- ❖ Office Hours: (M: 12:30-13:30) or (by appointment)

Grading

In class assignments	30%
Midterm	25%
Final	25%
Project	20%
TOTAL	100%

- ❖ Very much like real life where no one asks you what your grade was in that in a course, I don't care about the grades either. But I do care about your willingness, attitude, skillsets, and these are exactly what anybody needs to see in your reference letter! Hence, your project is super important!



Final Grade!

I do not round up your grades.

Oh, that's a bummer!



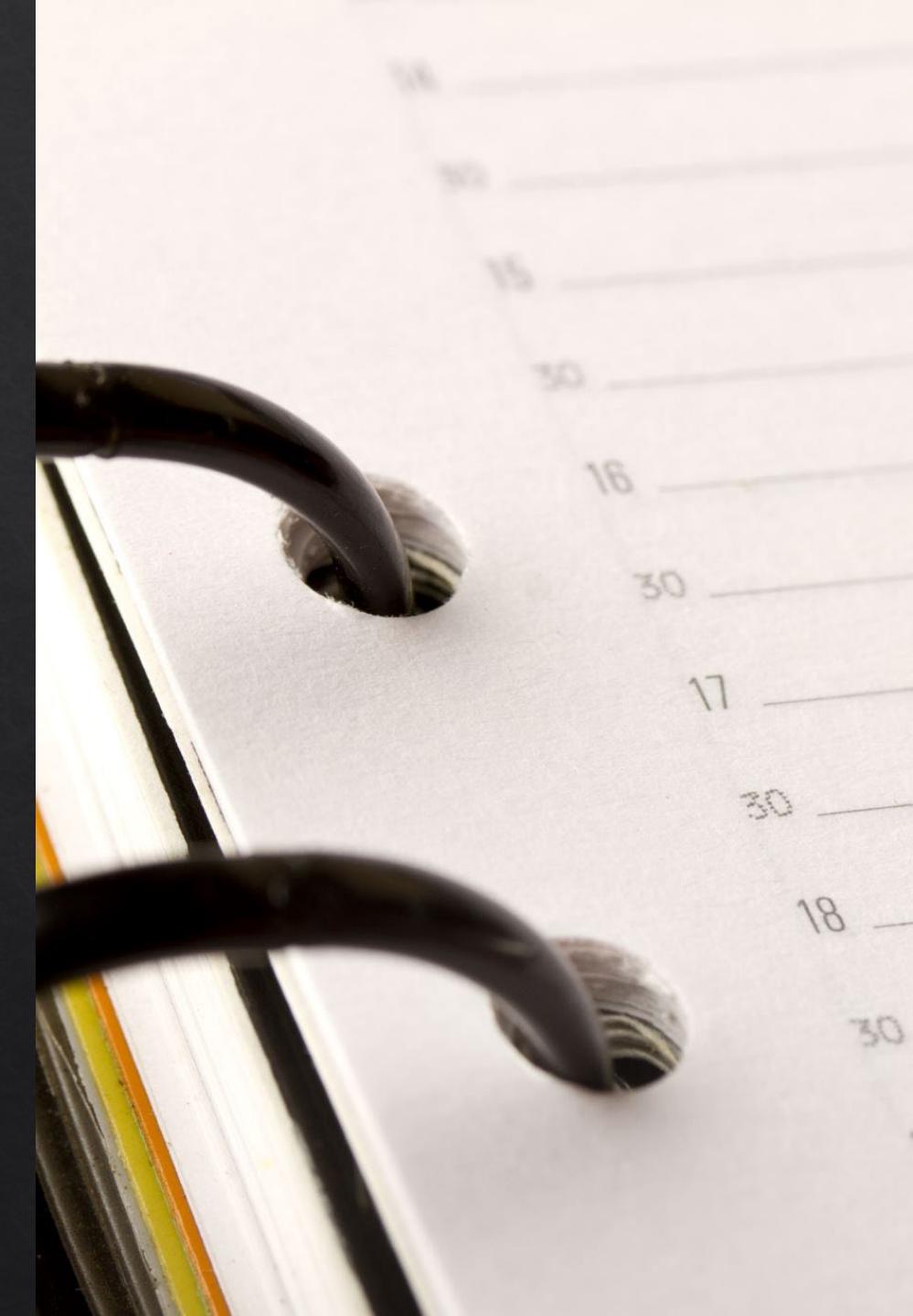
In the grade calculation I give 1% free extra credit to everyone. That means if, for example, you got 84.1 it round up to 84. However, with 1% extra it will boost you up to 85.1%. It works toward your favor.

However, if you got 83.99 the round up or the bonus will both result in 84% final.

Grade	Grade Points	Conversion Scale (%)	Definition
A+	4.33	90 - 100	Exceptional Achievement
A	4.00	85 - 89	Outstanding Achievement
A-	3.67	80 - 84	Excellent Achievement
B+	3.33	76 - 79	Very Good Achievement
B	3.00	72 - 75	Good Achievement
B-	2.67	68 - 71	Good Achievement
C+	2.33	64 - 67	Satisfactory Achievement
C	2.00	60 - 63	Satisfactory Achievement
C-	1.67	56 - 59	Minimal Achievement. This grade does not permit student to pursue another course for which the graded course was a prerequisite.
D	1.00	50 - 55	Minimal achievement. This grade does not permit student to pursue another course for which the graded course was a prerequisite.
F	0.00	0 - 49	Unsatisfactory achievement. Student did not meet minimum course requirements.

Grading

In class assignments	30%
Midterm	25%
Final	25%
Project	20%
TOTAL	100%
Bonus Activities	up to 5%



Moodle

PHYS 4900: Special Topics (R10)

Home

My courses

PHYS 4900 R10 SPRING 2023

General

General

In-Class-Assessments

Project + Presentation

Course Notes & Slides



First Week Checklist



Announcements



Discussion Forum for Students



Course Presentation



Exams



|| Online Session if Required ||

Moodle

PHYS 4900: Special Topics (R10)

Home My courses PHYS 4900 R10 SPRING 2023 In-Class-Assessments

General In-Class-Assessments Project + Presentation Course Notes & Slides

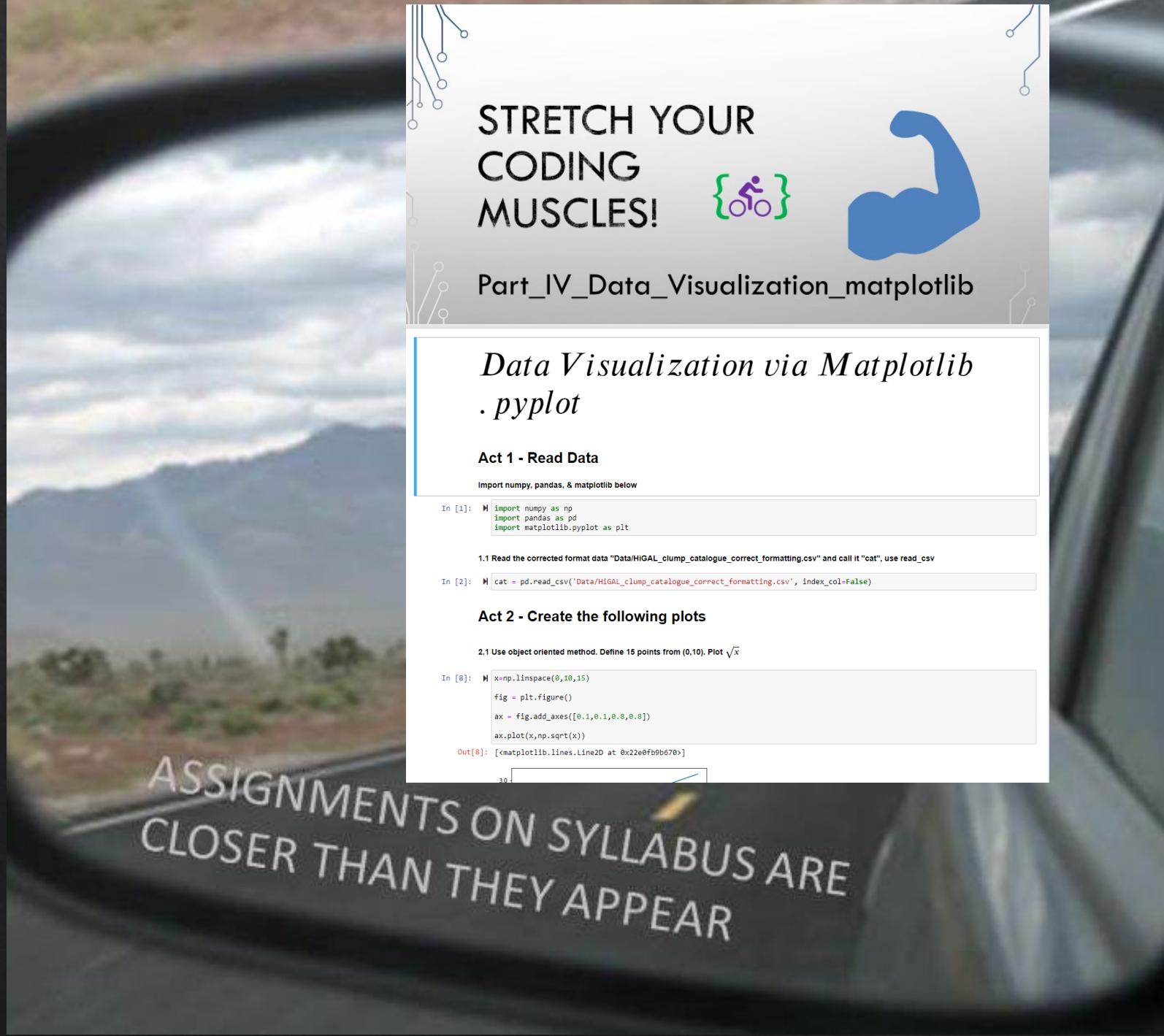
Part_I_Essentials_ToDo
Part_I_Essentials_Solutions
Part_II_np_pd_plt sns ToDo
Part_II_np_pd_plt sns Solutions

◀ General

Homework – 30%

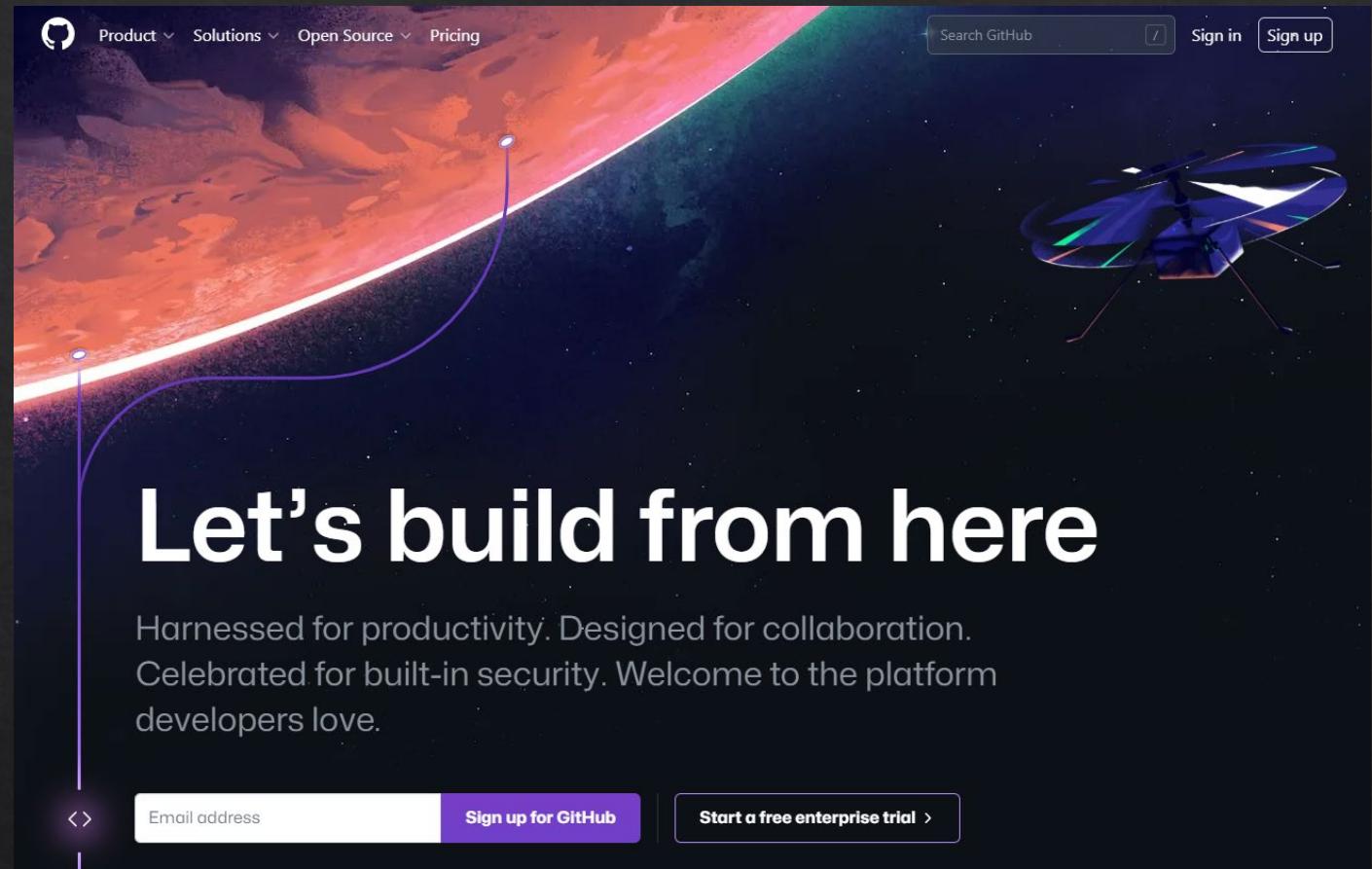
- ◆ Assignments are due in class!

Working in groups is encouraged. It helps you understand the material better. BUT you must write up your own assignment, by yourself, in your own words, using your own logical flow. PLEASE DON'T JUST COPY someone else's work.



A suggestion!

- ❖ Sign up for [GitHub](#)
- ❖ Upload the assignments on your GitHub, right after you are done with them!
- ❖ It will be very helpful one day! Trust me!



Moodle

PHYS 4900: Special Topics (R10)



Home

My courses

PHYS 4900 R10 SPRING 2023

Project + Presentation

Turn editing on

General

In-Class-Assessments

Project + Presentation

Course Notes & Slides

Download any data set you are interested in and see what you can do with this. You have plenty of expertise! I believe in it and you should as well. You need to show that you are the one who understands YOUR data up and down! Your goal should not only be confidently analyze data and explain it carefully to me and the audience but any hiring committee should be able to see this. This is going on your GitHub, hence, your portfolio!

For your final presentation, I am trying to invite an actual data scientist to attend. If their timing allows this will be your chance to show your skillsets to a real-life expert. Take it seriously!

A few websites to find your dataset:

- CADC: <https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/>
- Open Canada: <https://open.canada.ca/en/open-data>
- Open Canada: [Open Government Portal](#)
- STATCAN: <https://www.statcan.gc.ca/en/microdata/data-centres/data>
- Kaggle Datasets: <https://www.kaggle.com/datasets>

Or simply google for keywords like Magnetic field data, Acoustic data, or any other physical concept you are interested in!

Enjoy and I am so looking forward to this! Note that it is not the dataset which matters, but what you do with it and how you present it. So just pick something and start working on it asap. Remember, you are a scientist. You can interpret your results and understand them in a way that others may not be able to.

Use may use this for your files submissions

Project/Presentation 20%

- ❖ The most important part, if you ask me. Book as many as office hours you need with me!
- ❖ Download any data set you are interested in and see what you can do with this. I would suggest a catalogue of data! Table of content representing physical properties.
- ❖ You have plenty of expertise! I believe in it, and you should as well.
- ❖ You need to show that you are the one who understands YOUR data up and down! Your goal should not only to be confidently analyze data and explain it carefully to me and the audience, but any hiring committee should be able to see this.
- ❖ This is going on your GitHub, hence, your portfolio!
- ❖ For your final presentation, I am trying to invite an actual data scientist to attend. If their timing allows this will be your chance to show your skillsets to a real-life expert. Take it seriously!
- ❖ Enjoy and I am so looking forward to this! Note that it is not the dataset which matters, but what you do with it and how you present it. So just pick something and start working on it asap. Remember, you are a scientist. You can interpret your results and understand them in a way that others may not be able to

Project/Presentation 20%

A few websites to explore:

- ❖ CADC: <https://www.cadc-ccda.hia-ihc.nrc-cnrc.gc.ca/en/>
- ❖ Open Canada: <https://open.canada.ca/en/open-data>
 - ❖ Open Canada: [Open Government Portal](#)
- ❖ STATCAN: <https://www.statcan.gc.ca/en/microdata/data-centres/data>
- ❖ Kaggle Datasets: <https://www.kaggle.com/datasets>

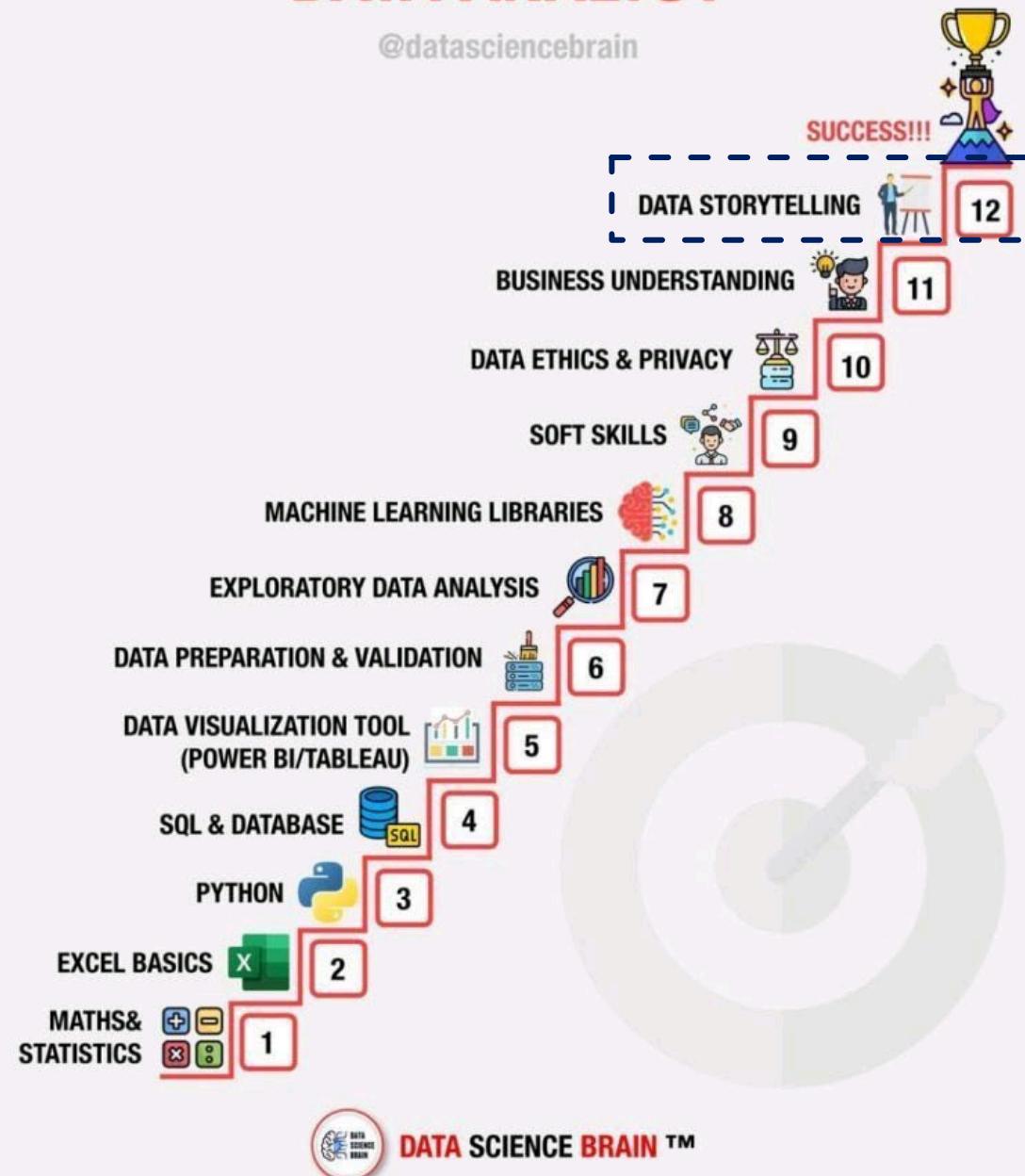
Or simply google for keywords like Magnetic field data, Acoustic data, or any other physical concept you are interested in!

Presentation

- ❖ Scientist/Physicist are not usually good at this.
- ❖ You can get this right; anybody would be happy to work with you!
- ❖ 1. Introduction, 2. Data, 3. Methods, 4. Results, 5. Conclusion → From now on these 5 steps should be part of any of your paper/presentation/interview/etc.
- ❖ “You do not really understand something unless you can explain it to your grandmother.” — Albert Einstein

SELF GUIDE TO BECOME A DATA ANALYST

@datasciencebrain



Moodle

PHYS 4900: Special Topics (R10)

Home My courses PHYS 4900 R10 SPRING 2023 Course Notes & Slides

General In-Class-Assessments Project + Presentation Course Notes & Slides

Introduction
Part_I_Essentials
Part_II_np_pd_plt sns

◀ Project + Presentation

Exams!

Just think about them like a review of materials!

Midterm: Feb 8

Final: TBA

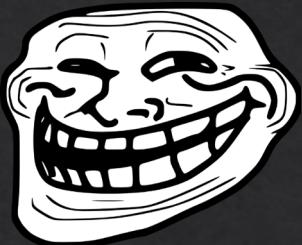


A few important suggestions!

- ❖ Download any data you have slightest interest on.
- ❖ Apply techniques we learn on them!
- ❖ Practice, practice, practice.
- ❖ Perform tons of visualizations. Plot everything vs everything!
- ❖ Spend time on your plotted data! Afterall, you are one of the few who can interpret it well!
- ❖ Two key factors: work on your writing and presentation skills. This is how you get to convince people how invaluable you are and thus hiring!
- ❖ Do lots of readings! Afterall you are a scientist, and you live on this!

FAQ

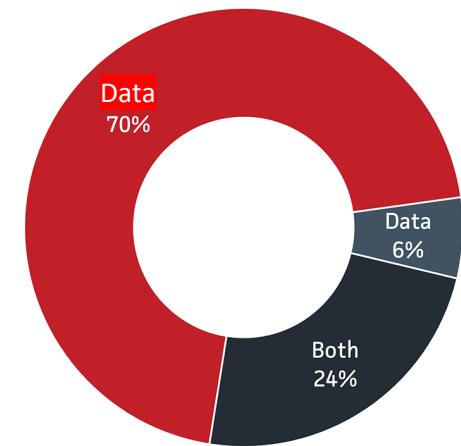
- ❖ Who teaches coding via PowerPoint?
 - ❖ Me!
 - ❖ Think about it like a note and review it to memorize it



- ❖ Is it pronounced Data or Data?
 - ❖ No Matter How You Say It,
I'm Thankful for Data



Do you say **data** or **data**?



Source: Data Trust internal poll of employees.

