IIT Bombay : Spring 2022

AE 102 : Data Analysis & Interpretation.

Amuthan Ramabathiran
amuthan @ aero. iitb. ac. in

my cats !

my brothers guitar !

Data & the scientific method

* Data Analysis goes back to the very foundations of science !

* This lecture : What is the scientific method ?
  What role does data analysis play ?

* Historical & ongoing case study :
  - Discovering Planets !

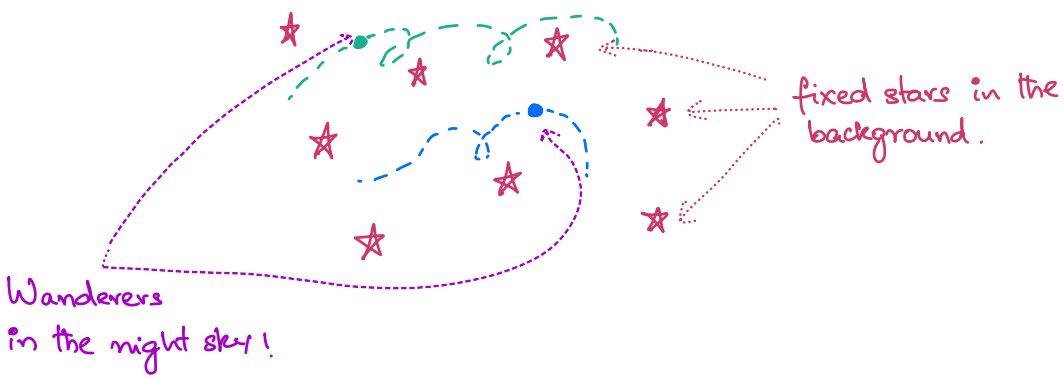
Pop Quiz !

How many planets are there in our solar system ?

Patterns in the sky !

How did we discover planets ?

- Mercury , Venus, Mars , Jupiter , Saturn are visible in the night sky.



PLANETS

The ancients knew about the five planets !

Planet	Greeks	Indians
Mercury	Ερμης	बुध
Venus	Αφροδιτη	शुक्र
Mars	Αρης	मङ्गल
Jupiter	Ζευς	बृहस्पति
Saturn	Κρονος	शनि

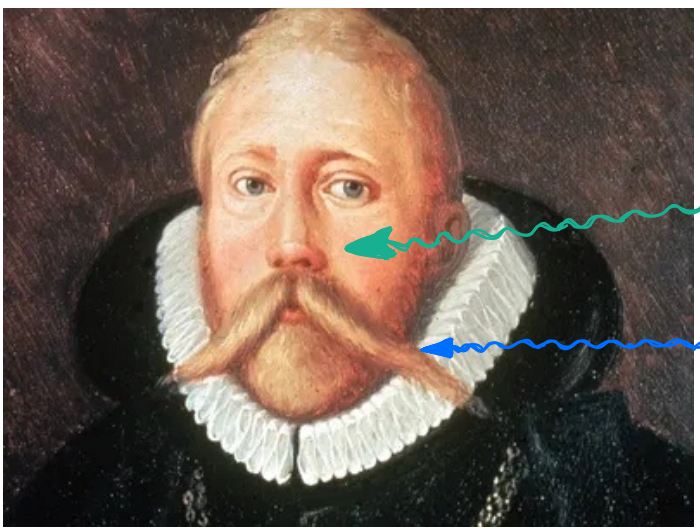
... but we have a much better understanding now !

* Ptolemy's geocentric model was the established model until the 16th century.

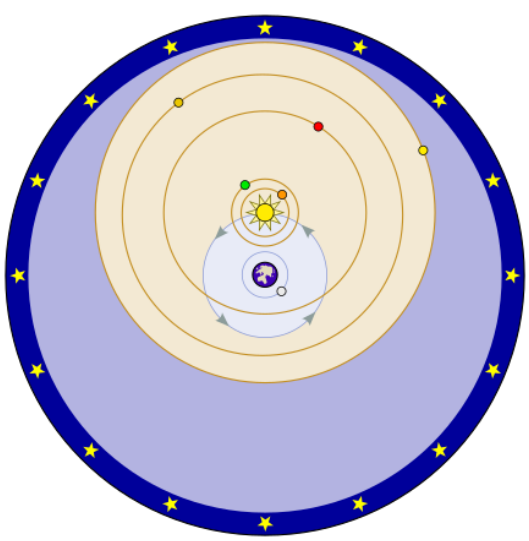
* Nicolaus Copernicus (1473 - 1543) proposed the heliocentric model to explain the observed motion of the planets.

Fast forward to the 16th century:

Tycho Brahe (1546 - 1601)



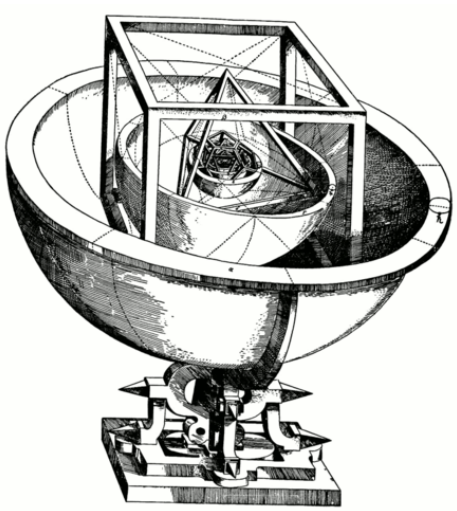
* Brahe proposed a hybrid model where the sun and the moon revolved around the earth, while other planets revolved around the sun.



* He collected lots of data & hired Johannes Kepler as his assistant.

Johannes Kepler (1571 - 1630)

* Kepler, in turn, had his own favourite version of heliocentric theory.

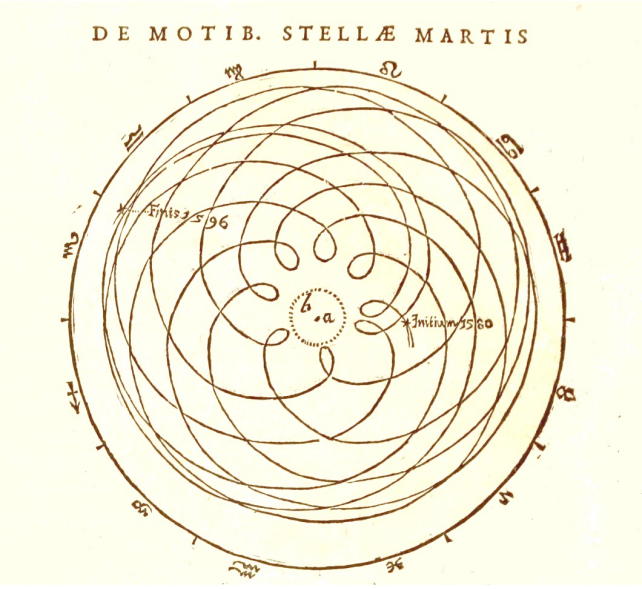


* But data (from Brahe's observations) did not agree with his pet theory!

* Kepler noticed that elliptical orbits fitted the data better, but did not publish this for more than a decade!

* Epicycles : Mars' trajectory in the sky according to Kepler.

* Kepler used this as strong evidence that planets' orbits are elliptical.



Dr. William Donahue presents a lot of evidence that Kepler partly falsified his data on Mars' orbit to artificially strengthen his conclusions !

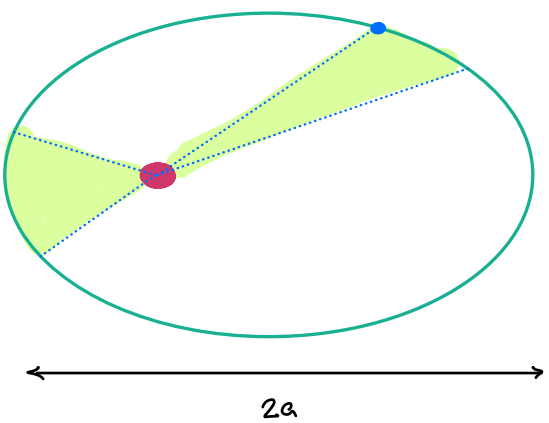
* Kepler's laws :

(i) Elliptical orbits with sun as one of its foci.

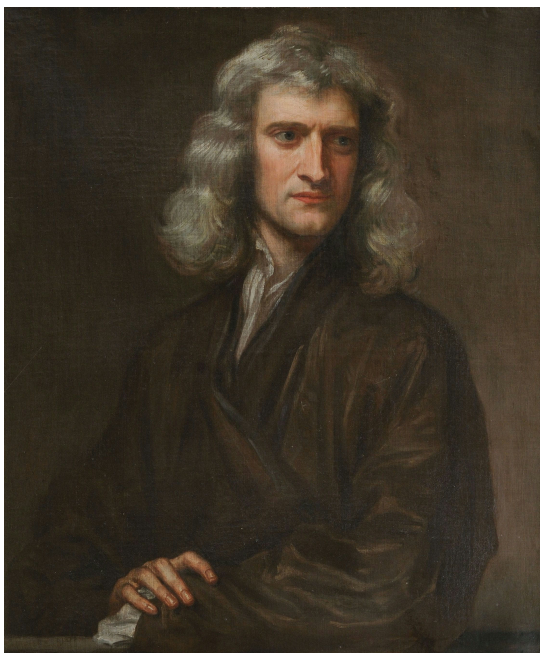
(ii) Sweeps equal areas in equal intervals of time.

(iii) $T^2 \propto a^3$

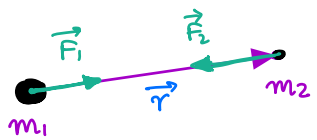
↳ orbital time period



Isaac Newton (1642 - 1726)

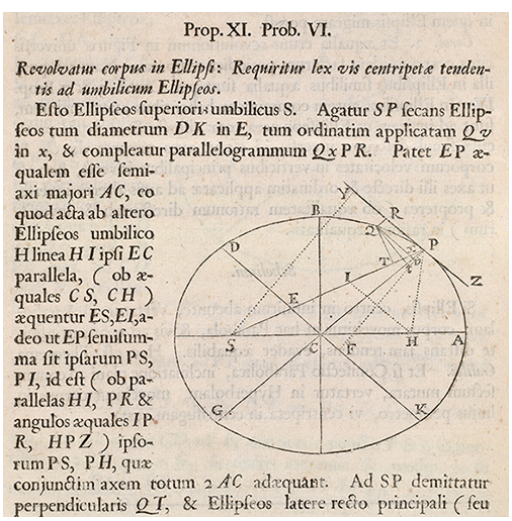


* Explained Kepler's laws in terms of his law of universal gravitation.



$$-\vec{F}_2 = \vec{F}_1 = \frac{G m_1 m_2}{\|\vec{r}\|^3}\vec{r}$$

* A proposition from Newton's magnum opus Philosophiæ Naturalis Principia Mathematica.



Prop. XI. Prob. VI.

Revolvatur corpus in Ellipſi: Requiritur lex vis centripetæ tendentis ad umbilicum Ellipſeos.

Eſto Ellipſeos ſuperioris umbilicus S. Agatur *SP* ſecans Ellipſeos tum diametrum *DK* in *E*, tum ordinatim applicatam *Qv* in *x*, & compleatur parallelogrammum *Qx*PR. Patet *EP* æqualem eſſe ſemiaxi majori *AC*, eo quod acta ab altero Ellipſeos umbilico H linea *HI* ipſi *EC* parallela, (ob æquales *CS*, *CH*) æquentur *ES*, *EI*, adeo ut *EP* ſemiſumma ſit ipſarum PS, *PI*, id eſt (ob parallelas *HI*, PR & angulos æquales *IPR*, *HPZ*) ipſorum PS, PH, quæ conjunctim axem totum 2*AC* adæquant. Ad SP demittatur perpendicularis *QT*, & Ellipſeos latere recto principali (ſeu

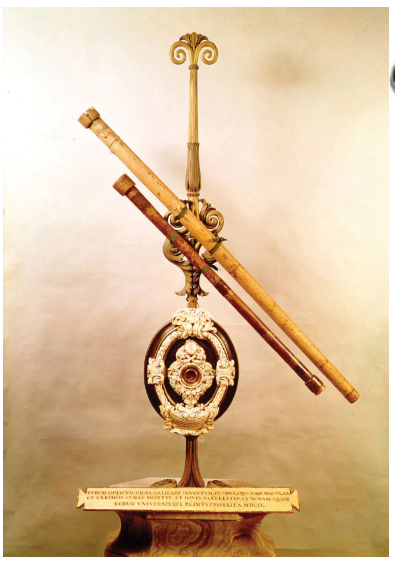
Newton was not a nice guy!

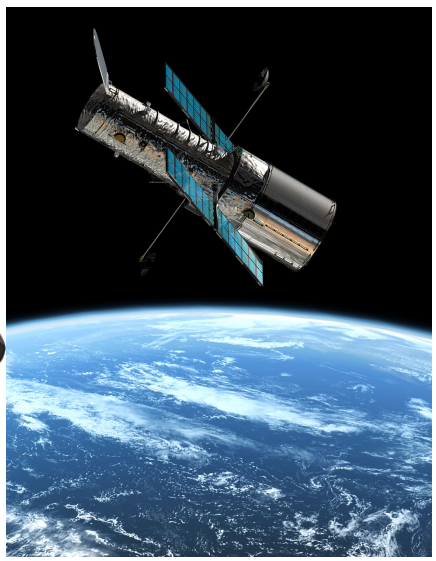
"... a nasty antagonist"

"... a bad man to have as an enemy."

} Steven Weinberg on Newton.

* Better & better telescopes continued (& continue!) being developed, giving better & better observations.





Galileo's telescope (17th century) → Herschel's telescope (18th century) → Hubble space telescope (20th century)

* William Herschel discovered Uranus in 1781 using his own hand-made telescope, from his backyard.

Herschel initially thought Uranus was a comet!

smooth!

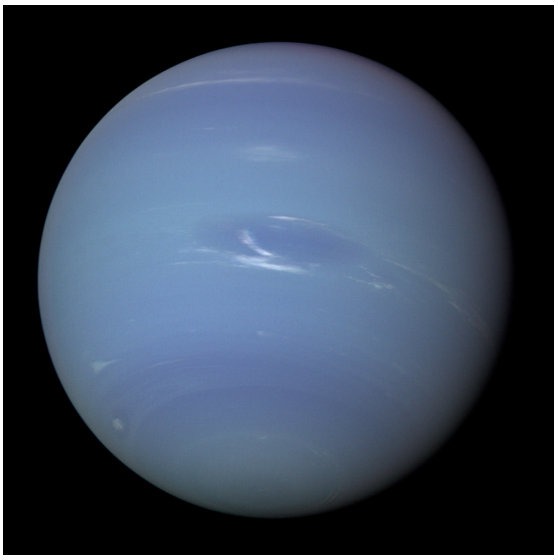
The discovery of Neptune - Triumph of Newtonian Gravity.

* Alexis Bouvard noticed in 1821 that there were inconsistencies between observed & predicted orbits of Uranus.

* John Couch Adams & Urbain le Verrier independently predicted the existence of a massive planet around 1845-46.

!!!

* Johann Galle pointed his telescope at the location predicted by Le Verrier and discovered Neptune !



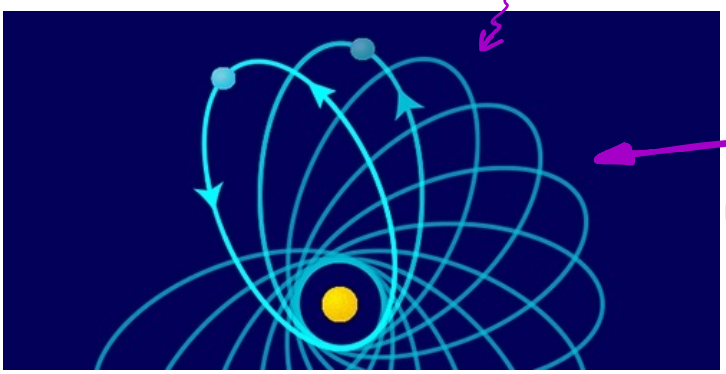
Neptune first planet to be discovered based on a prediction !*

Science is much more than data analysis !

* The data used by Le Verrier & Adams was not entirely accurate. They got lucky with their guess !

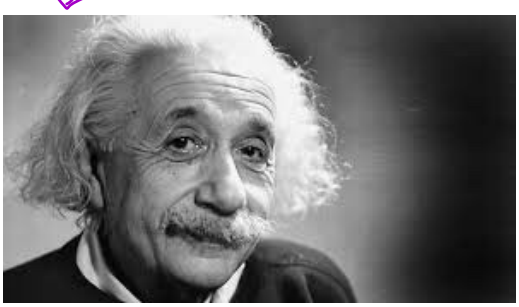
* The discovery of Neptune based on Newton's theory of gravity gave this theory solid experimental support.

* Within a hundred years of Neptune's discovery, new data on the perihelion precession of Mercury was not explainable by Newtonian gravity!



Falsified Newtonian gravity!

* This was resolved by Albert Einstein's General Theory of Relativity, which superceded Newtonian theory of gravity.



MASS DISTORTS SPACETIME!!!

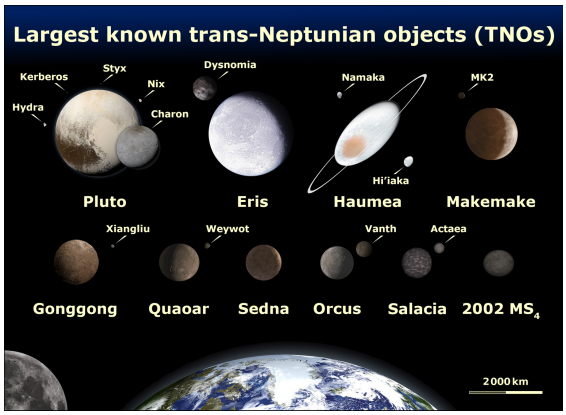
$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

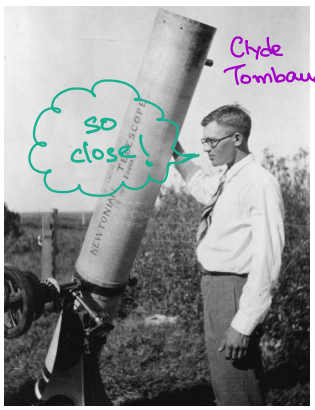
Spacetime CURVATURE — Amount of MASS

Science is self-correcting!

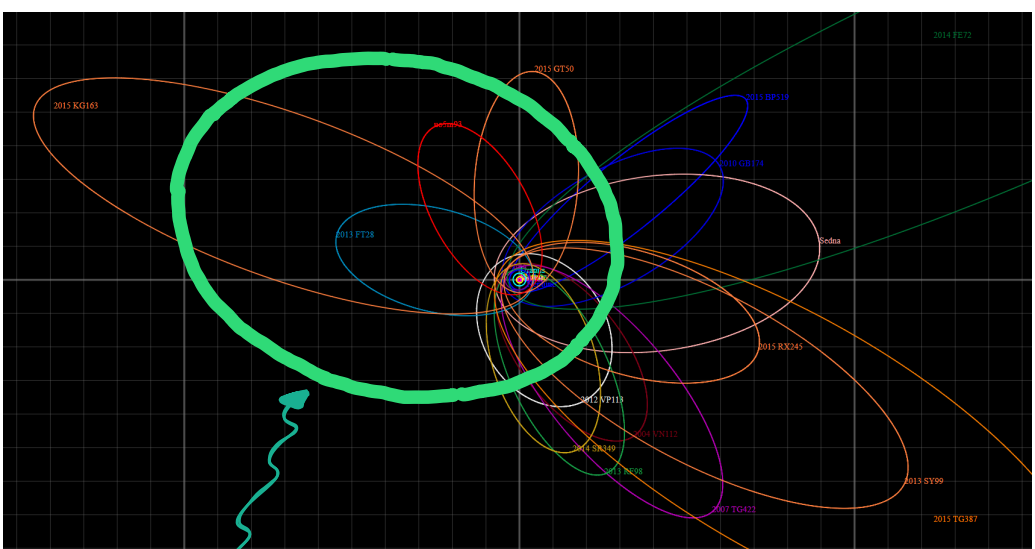
* New data crushed Pluto's status as a planet !

Too many trans-Neptunian objects!





* The planetary tale is not over yet ! Some astronomers think that there is a massive 9th planet (not Pluto !) out there, far outside the orbit of Neptune.



Predicted orbit of the yet undiscovered Planet 9 !

The hunt is on!

Key take-aways from the planetary tale.

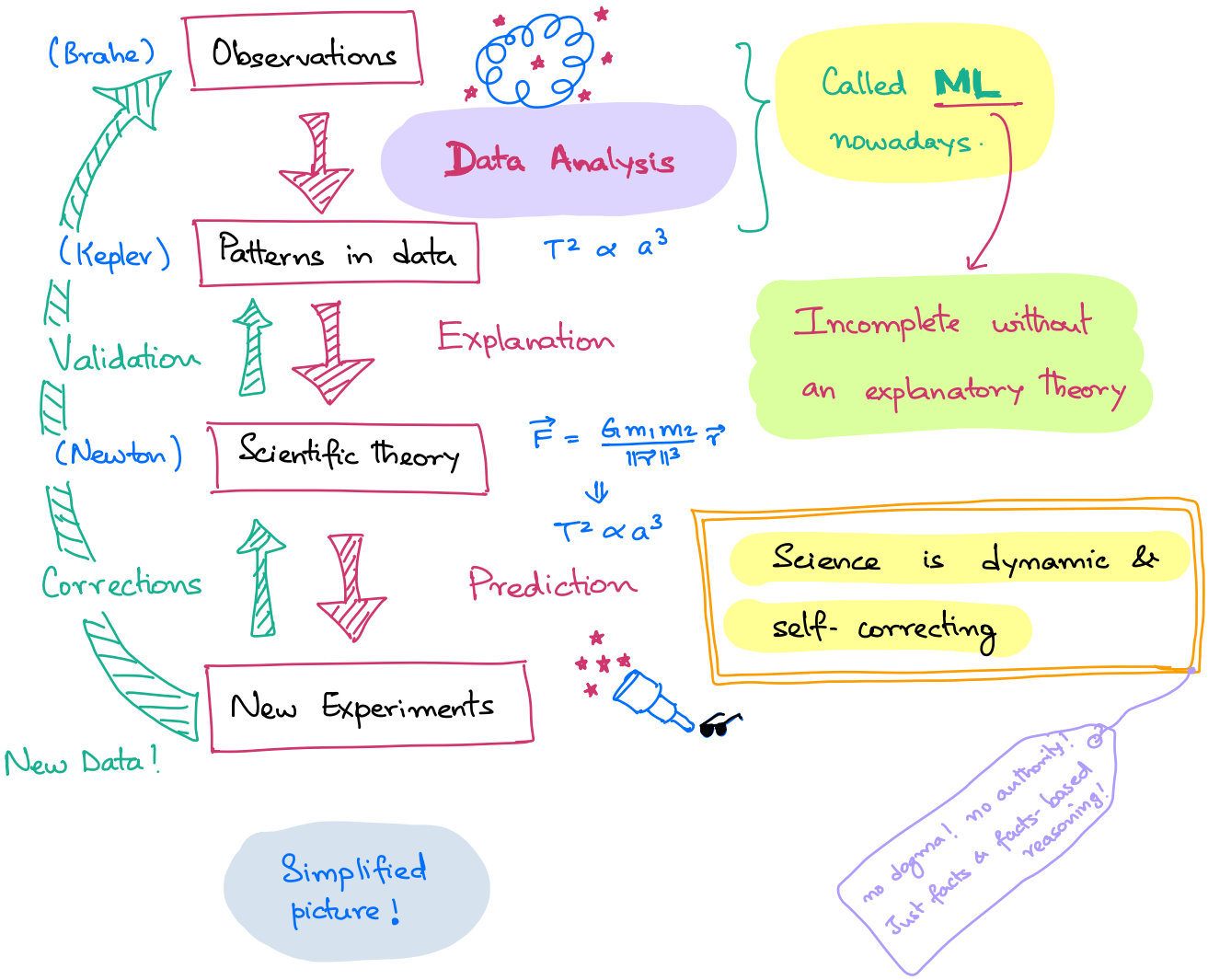
* Collecting data and extracting meaningful information from plays a central role in science; data serves as a reality check for speculations.

* Better data leads to better science!
  - Naked eye observations → epicycles.
  - Kepler's analysis of Brahe's data → Newtonian Gravity ← Kepler's laws
  - Discrepancies in the observed orbit of Uranus → Discovery of Neptune
  - Data about perihelion precession of Mercury → ~~Newtonian Gravity~~ General Theory of Relativity
  - New data about trans-Neptunian objects → ☹ Pluto is not a planet!
  - Unexplained correlations in the orbits of outer solar objects → prediction of Planet 9!

Data analysis uncovers hidden patterns in the data, and is integral to science.

Science explains those patterns & allow us to make testable predictions.

A small detour on the scientific method.



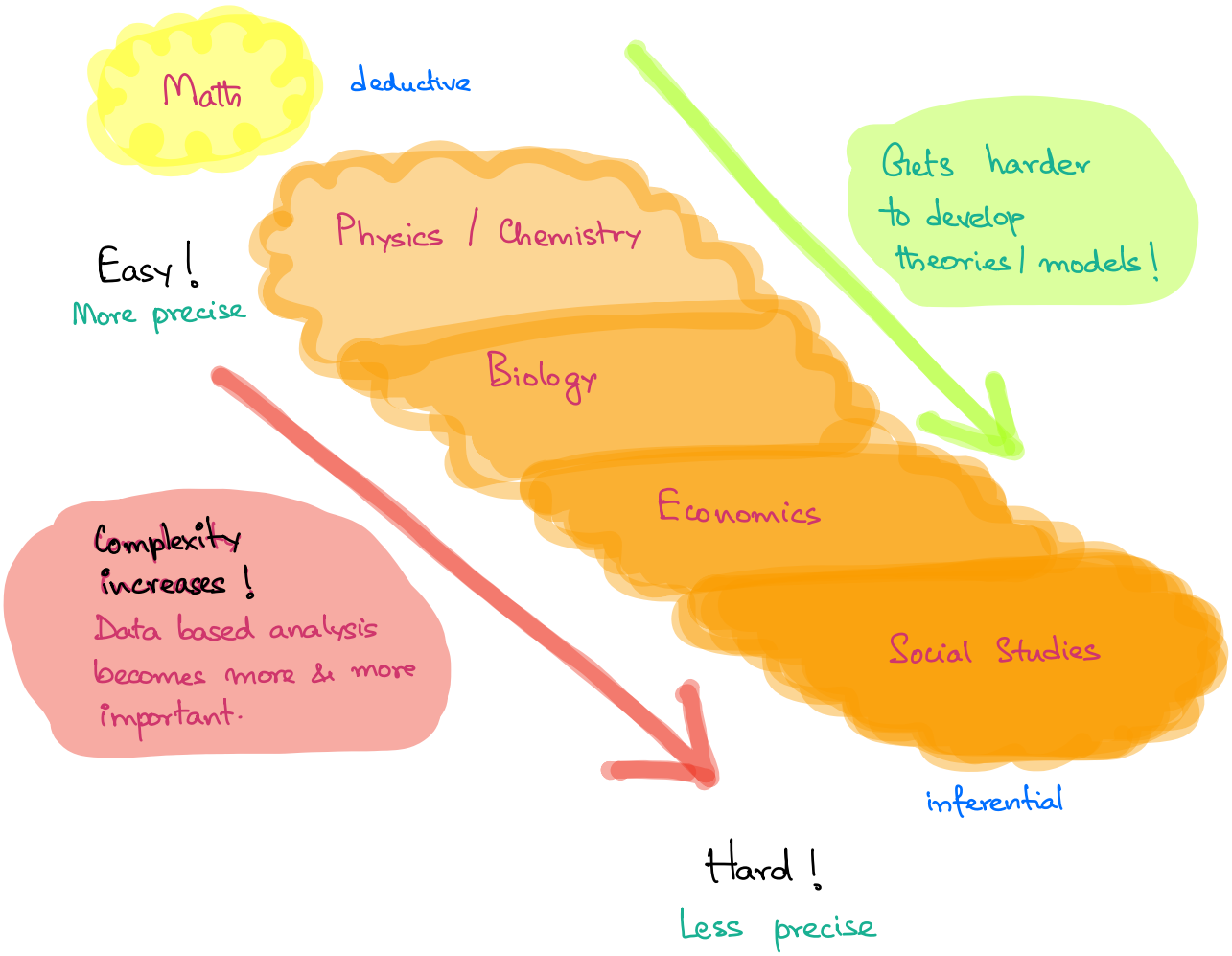
There is much more to the scientific method!

* Analysis of Data & its proper Interpretation are strict subsets of the larger scientific endeavor!

Data Analysis is useful in lots of applications!

duh!

* In general, the more complex the application, the more difficult it is to develop a theory / model for it!



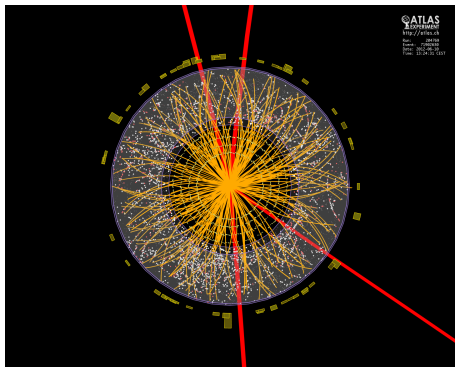
→ read Anderson's article "More is different"

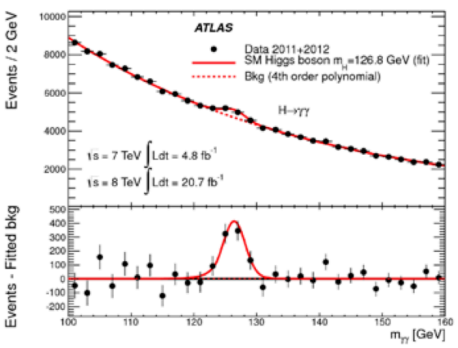
* KEEP IN MIND: for complex applications, data analysis provides clues & hints - incomplete without a proper theory!

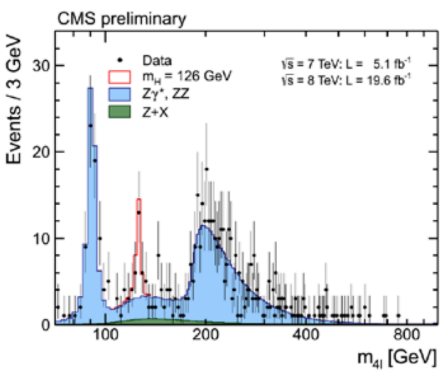
Data Analysis can take you places!

Nobel prize - Physics 2013

Discovery of the Higgs boson.



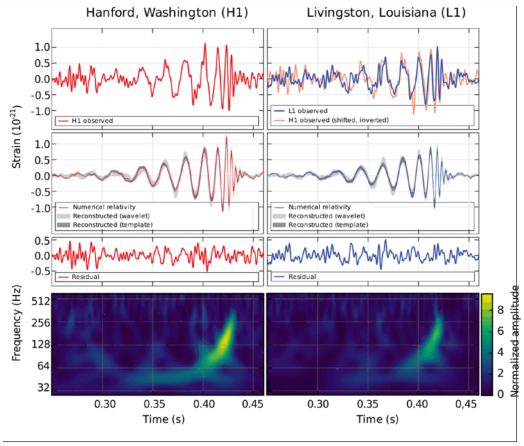




Nobel prize - Physics 2017

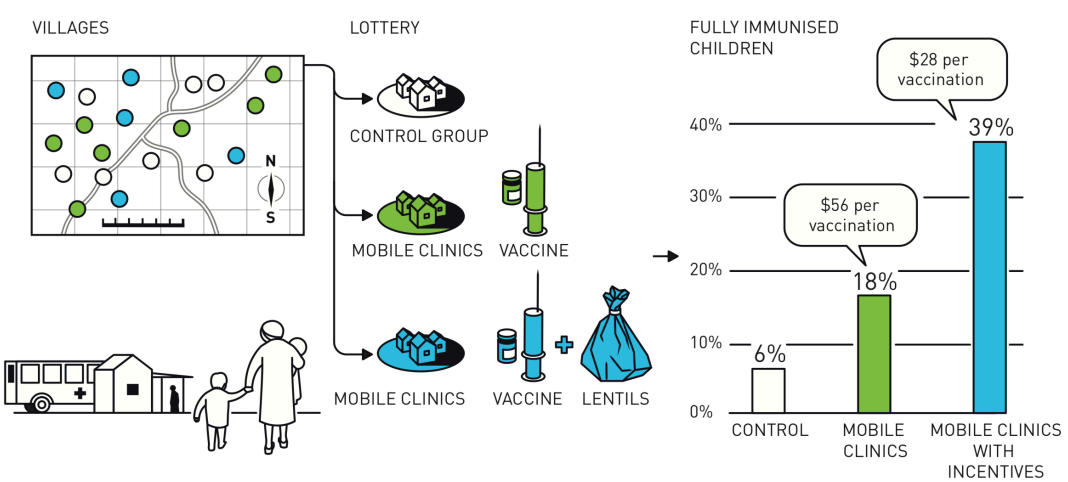
Detection of gravitational waves by the LIGO team.

I took this pic!



Nobel prize - Economics 2019

Using Randomized Controlled Trials (RCT) to inform economic policy.



There is much more!

* We will see some elementary examples of data analysis in the real world.

* You will also have an option to perform a data-based exploratory project, if you wish to.

... which brings us to some boring course logistics

* Lectures will be online, at least until May, or until you are all in campus.

* Lecture hours:

Monday 1530 - 1700 hrs

Thursday 1530 - 1700 hrs

1 hr theory + 0.5 hrs hands-on

Attendance is optional!

But I strongly recommend attending the lectures live!

* Office hours: ?

optional 30 min session every week to clarify your doubts, give feedback etc.

Teaching assistants

		@ iitb.ac.in
(Head TA)	Navaneet	204010012
	Dharamveer Kumar	204010002
	Aishwarya Kasarla	214010002
	Prince Sharma	203010006
	Neel Mani	203010036
	Amit Kumar	203010037
	Rajan Rahul	203010040
	Mohammad Aman	203010041
	Jagriti Saha	203010042
	Krishna Kishore Bokka	213010039
	Arpitha Rathod	213010041
(UG representative)	Sarthak Mehrotra	200010068

* Make sure you get in touch with at least one of the TAs in case you are not able to reach me.

Moodle

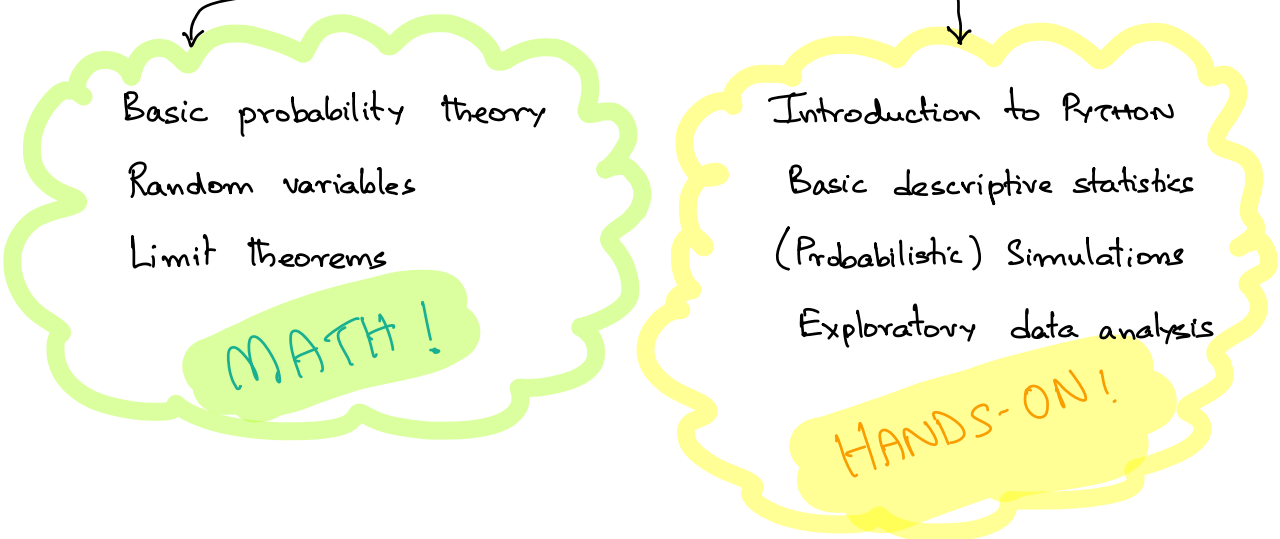
* All of you are automatically enrolled into the course Moodle; if not, talk to me or the TAs immediately!

Students with Disabilities

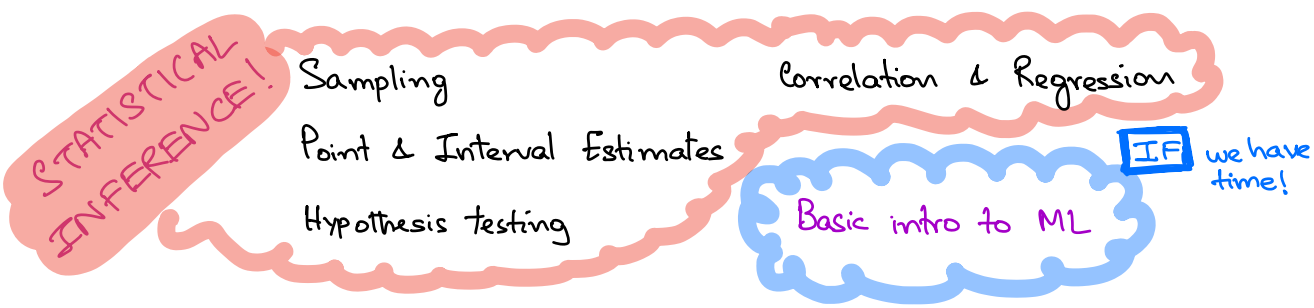
* If you have any disabilities or if you need special help, please talk to me ASAP. All your private disclosures regarding this will be confidential.

Course Contents (Tentative!)

PART-1: Basic toolkit (before midsem)



PART-2: Data Analysis (after midsem)



Learning outcomes

* Reason effectively based on (noisy) data.

* Given data about a particular phenomenon, you should be able to

  (i) Frame questions about the phenomenon that you would like to answer based on the data.

  (ii) Use appropriate random variables that will help you answer questions raised in (i) above.

  (iii) Make educated guesses about the distribution of the random variables.

  (iv) Characterize, Visualize & Estimate quantities of interest.

  (v) Formulate hypotheses about the data & test them.

  (vi) Understand correlations between random variables, and perform (linear) regression analysis.

References

* The primary reference is what is covered in the lectures.

* Optional reading :

  - Introduction to Probability and Statistics for Engineers and Scientists — Sheldon Ross.

  - Cartoon Guide to Statistics — Larry Gonick, Woollcott Smith.

  - Computational & Inferential Thinking (www.inferentialthinking.com) — Ani Adhikari, John DeNero.

  -  Stat Quest YouTube channel by Josh Starmer.

  - My lecture notes on probability theory available on my webpage https://amuthan.github.io/webpage/teaching

* Lecture notes created by you collectively. Scribes will be assigned for each lecture.

Evaluation

will be announced a week in advance!

* Short in-class Quizzes — 30%

* Mid-semester exam — 20%
  * Open notes!
* End-semester exam — 30%

* Assignments — 15%
  * ZERO plagiarism policy!

* Contributing to Class Wiki — 5%

* Optional semester-long project — 20%

(1) Individual contributions should be clearly mentioned

(2) should be completed within one week of lecture!

(1) Done in groups of 1-4.

(2) Topic should be approved by me before April 20.

(3) Should follow prescribed format (more later)

(4) Final report + Class Presentation.

Grading

Relative

90 - 95 %	AA
85 - 95 %	AB
75 - 85 %	BB
65 - 75 %	BC
55 - 65 %	CC
45 - 55 %	CD
35 - 45 %	DD
< 35 %	FR

* I usually curve scores to give you the best possible grade; so please don't haggle!

* To get AU, you need to score ≥ 50%.

Important Dates

Start of Instruction	March 21
* Choice of project topic	April 20
Mid-semester exam	May 9-14
* Approval of dataset for project	May 4
* Project submission	June 22
* Project presentation	June 24
Last day of Instruction	June 24
End-semester exam	June 27 - July 3
Final Grades	July 6.

* Only if you opted to do a project!