

Data Analytics

Mrinal Das

Introduction

hello

Associate Professor

Department of Data Science
IIT Palakkad (2017 -)

University of Massachusetts, USA
Aalto University, Finland

PhD

Indian Institute of Science, Bangalore
Machine Learning

Graduated Students

- ★ Rimmon Bhosale, MS 2022 (SMU, Singapore)
- ★ Gaurav Jain, MS 2022 (IBM Research)
- ★ Etana Fikadu, PhD 2025

Current Students

- ★ Eva Cherian, PhD, Jan 2022 -
- ★ Thomas John, PhD, Aug 2023 -
- ★ Jotipriya Das, PhD, Aug 2025 -

Current Sponsored projects

Detecting manufacturing defects in automobile parts

Learning Objective

Data Science and Analytics for Manufacturing

Data Analysis

Hypothesis testing

Decision making

Regression analysis

Data processing

Design of Experiments

Visualization of Analysis

Association Rule Mining

Statistical Analytics

Clustering

Basic Concepts, Examples, Hands-on Demonstrations

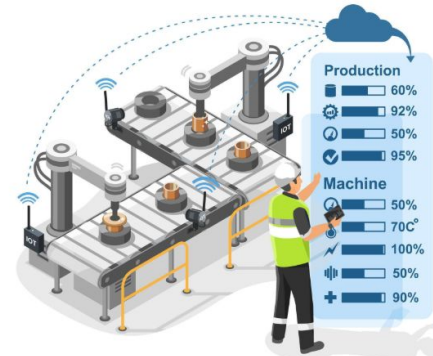
Data

Collection of discrete or continuous values (rainfall, air quality)

Outcome of set of natural or artificial events (goals in matches)

Observations recorded (attendance in a class)

Data is everywhere



What is Data Analytics

Data analytics involves
the research, discovery, and interpretation of patterns
within data.

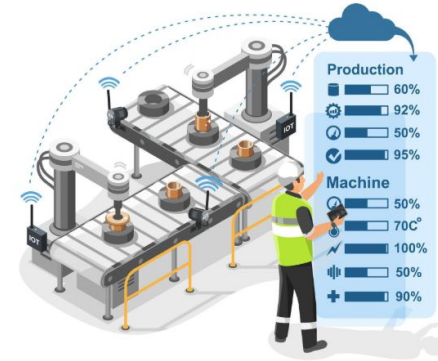
Why we need Data Analytics

Scientific ways to store, process, extract data

Data holds history, but hard to understand

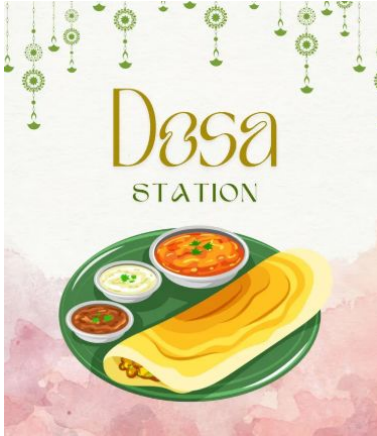
Analysing that can reveal

- Cause of failure
- Possible future
- Ways to make future better



Everyone is data analyst in daily life for simple problems

Data Analytics is Fundamental

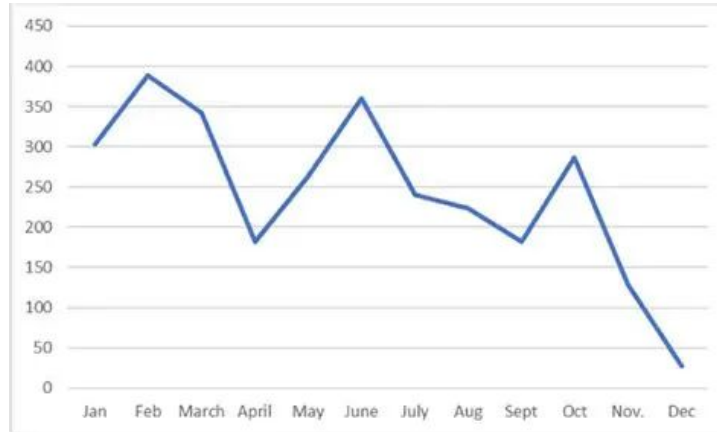


Example Scenario

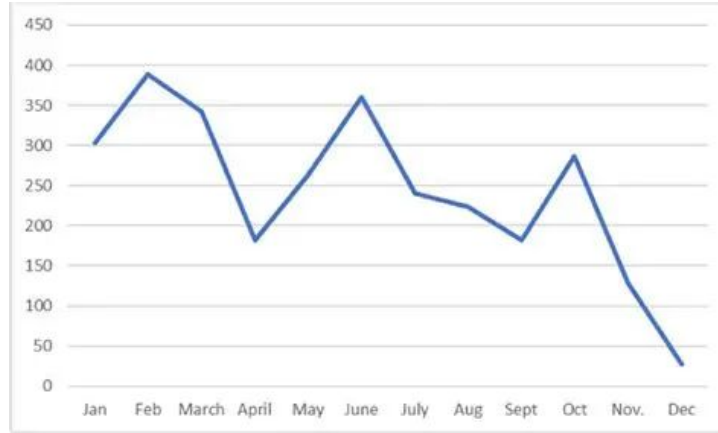


Lamborghini India has set up a DA department: LDAD

LDAD collected some sales data

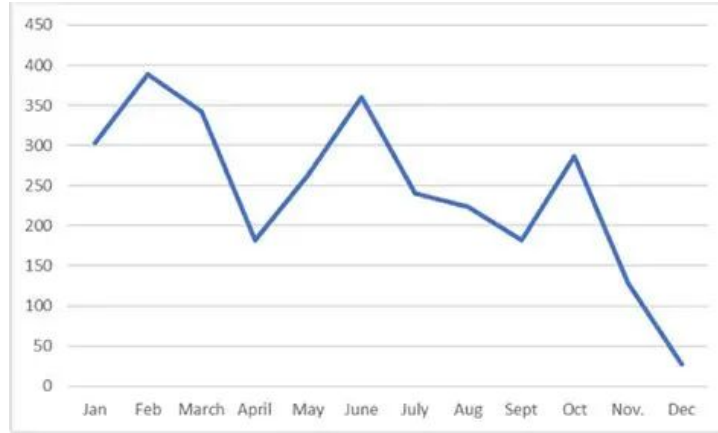


Example Scenario



- The trend in sale drop was noticed and reported
- Need to investigate the cause

Example Scenario



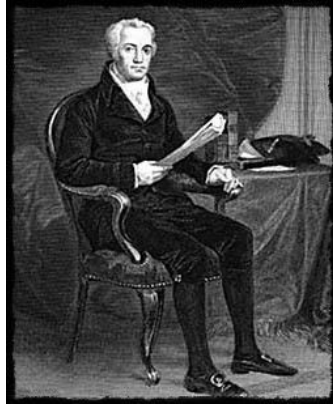
- LDAD analysed various data, current and past
- Checked correlation, dependency etc
- Finally they found the reason
- 90% demand was for black cars, whereas 20% cars manufactured were black

Pioneers



Kepler 1610 (law of
planetary motions)

Astrophysics



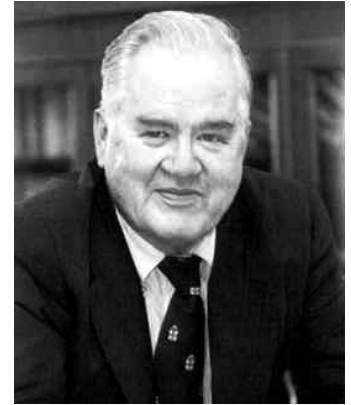
Playfair 1786 (bar
chart, pie chart)

Secret Service



Nightingale 1876 (data
visualization)

Nursing



Tukey 1965
(exploratory data
analysis)

Computer and
Software

Important Components

- ★ Read data
 - ★ Clean, slice, massage, visualize
 - ★ Compute statistical properties
 - ★ Represent findings to be understood by all
 - ★ Come up with remedial measures
- ★ Computer Program
 - ★ Mathematics
 - ★ Presentation Skill

Computer Program

```
1 threshold=5
2 height=5
3 if height >= threshold:
4     room=1
5 else:
6     room=0
7 print('room is:'+str(room))
```

```
room is:1
```

A tool to automate

Implementing logic

Beyond normal programs

The logic can not be predefined completely

Some parts depend on data

50% of 100 is different from 50% of 50

Example:

50% people go to room 1

A predefined threshold will not work

Example

Car weight vs pickup

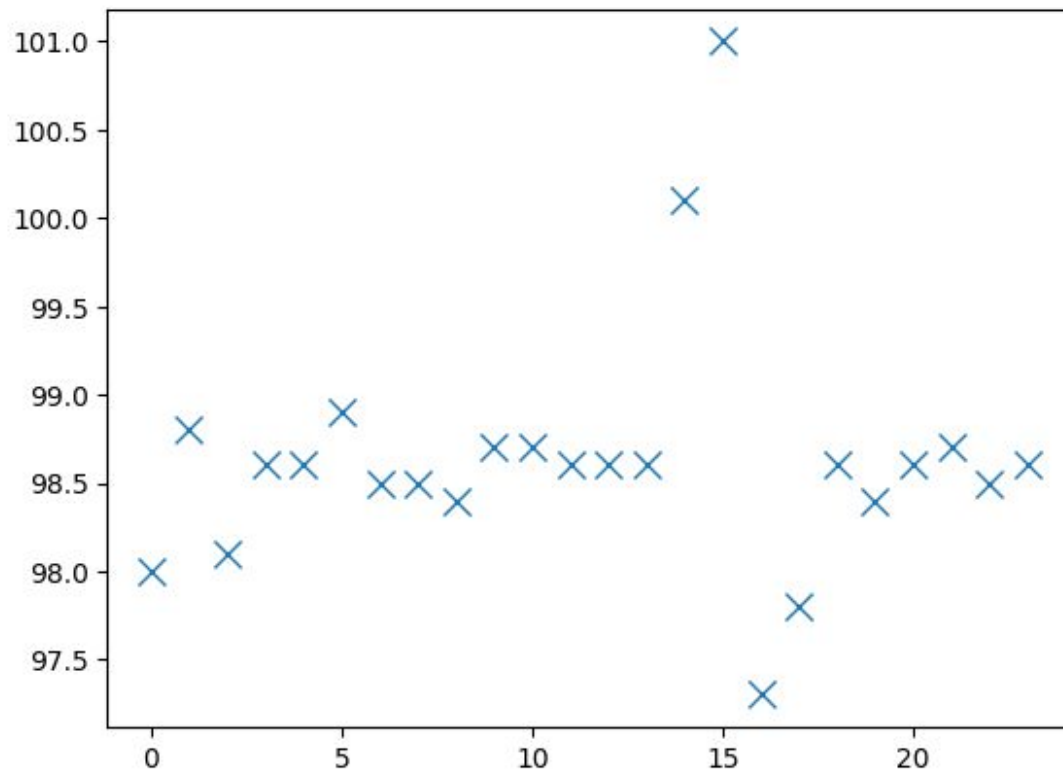
There are several models hatchback, sedan, SUV

Which cars have faster pickup ? (same engine power)

Example: car weight vs pickup

Car Model	Weight (kg)	0-100 km/h Time (s)
Hatchback A	900	10.2
Sedan B	1100	11
Sedan C	1250	11.8
SUV D	1450	12.8
SUV E	1600	13.5
Luxury F	1800	14.3
Van G	2100	16

Data Challenge 1

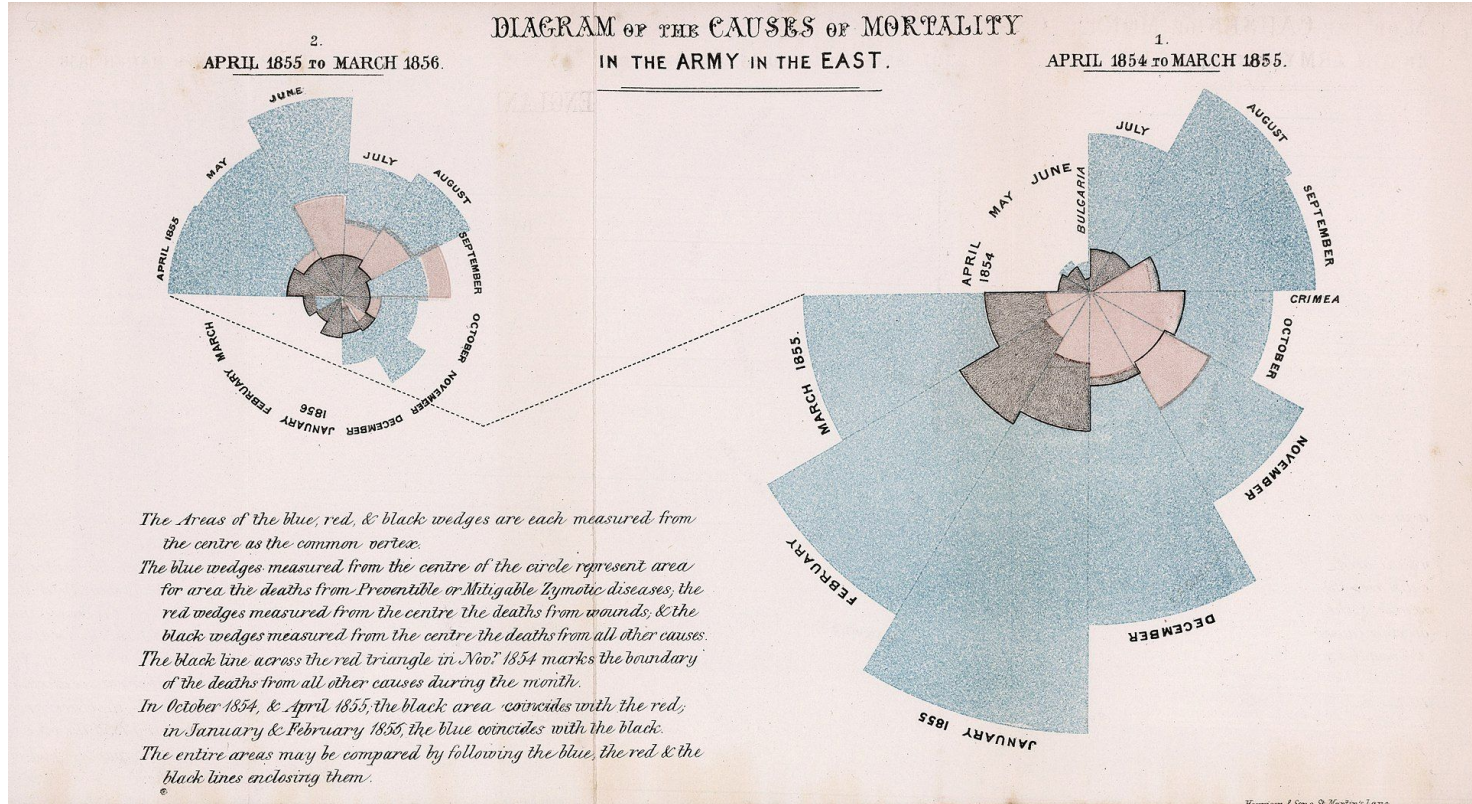


The Myth of 98.6

The 98.6 F standard dates to the mid-1800s.

German doctor Carl Wunderlich measured the armpit temperatures of about 25,000 people and came up with an average of 98.6 F.

Florence Nightingale the Lady with Data



Earth moves round the Sun

In 1515 Copernicus proposed an alternative theory--that the Earth was a planet orbiting the Sun, and that all planets moved in circles, one inside the other.

In 1610 Kepler proposed laws of planetary motions

In 1687 Thinking on Kepler's laws, Newton realized that all motion, whether it was the orbit of the Moon around the Earth or an apple falling from a tree, followed the same basic principles.

	Date, Old Style			Longitude					Latitude			Mean Longitude			
	Year	Day	Month	H	M	D	M	S	Sign	D	M	S	D	M	S
I	1580	18	November	1	31	6	28	35	Gemeni	1	40	N.	1	25	49 31
II	1582	28	December	3	58	16	55	30	Cancer	4	6	N.	3	9	24 55
III	1585	30	January	19	14	21	36	10	Leo	4	32	N.	4	20	8 9
IV	1587	6	March	7	23	25	43	0	Virgo	3	41	N.	6	0	47 40
V	1589	14	April	6	23	4	23	0	Scorpio	1	12	N.	7	14	18 26
VI	1591	8	June	7	43	26	43	0	Sagitt.	4	0	S.	9	5	43 55
VII	1593	25	August	17	27	12	16	0	Pisces	6	2	S.	11	9	49 31
VIII	1595	31	October	0	39	17	31	40	Taurus	0	8	N.	1	9	55 4
IX	1597	13	December	15	44	2	28	0	Cancer	3	33	N.	2	23	11 56
X	1600	18	January	14	2	8	38	0	Leo	4	30	N.	4	4	35 50
XI	1602	20	February	14	13	12	27	0	Virgo	4	10	N.	5	14	59 37
XII	1604	28	March	16	23	18	37	10	Libra	2	26	N.	6	27	0 12

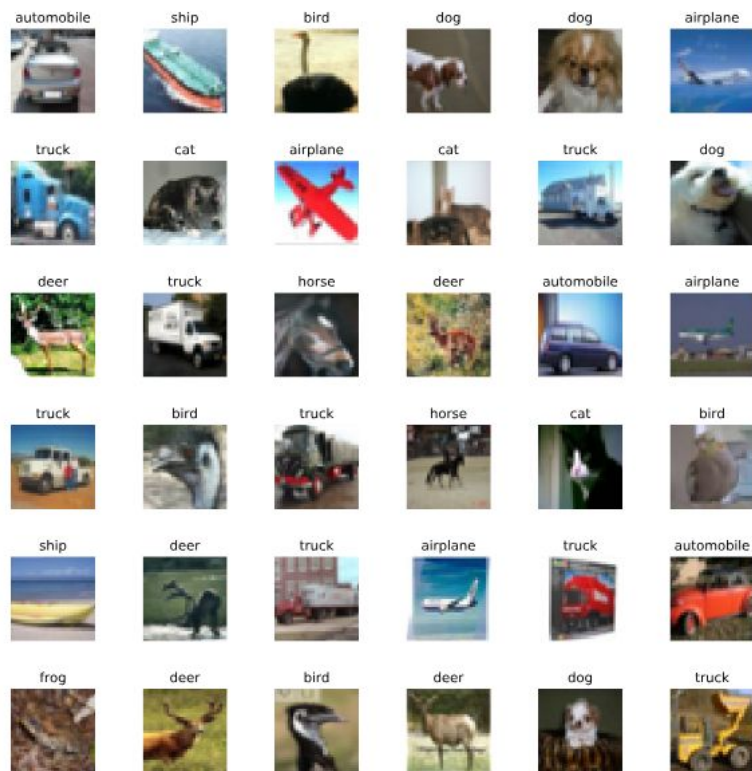
Numerical Data

Team	Span	Mat	Won▼	Lost	Tied	Draw	W/L	Ave	RPO
Australia	1877-2024	866	414	232	2	218	1.784	34.39	2.97
England	1877-2024	1071	392	324	0	355	1.209	32.19	2.78
West Indies	1928-2024	575	183	210	1	181	0.871	31.46	2.98
India	1932-2024	579	178	178	1	222	1.000	33.79	2.95
South Africa	1889-2024	464	178	161	0	125	1.105	31.56	2.82
Pakistan	1952-2024	456	148	142	0	166	1.042	32.67	2.92
New Zealand	1930-2024	470	115	185	0	170	0.621	29.58	2.71
Sri Lanka	1982-2024	316	103	121	0	92	0.851	32.50	3.11
Bangladesh	2000-2024	142	19	105	0	18	0.180	25.07	3.14
Zimbabwe	1992-2023	117	13	75	0	29	0.173	26.14	2.71
Afghanistan	2018-2024	9	3	6	0	0	0.500	22.69	3.03
Ireland	2018-2024	8	1	7	0	0	0.142	23.48	3.03
ICC World XI	2005-2005	1	0	1	0	0	0.000	16.70	3.43

Team	Span	Mat	Won▼	Lost	Tied	NR	W/L	Ave	RPO
India	2006-2024	229	149	69	5	6	2.159	31.08	8.60
Pakistan	2006-2024	245	142	92	4	7	1.543	24.46	7.84
New Zealand	2005-2024	220	111	92	10	7	1.206	24.39	8.24
Australia	2005-2024	195	105	83	3	4	1.265	26.37	8.50
South Africa	2005-2024	185	104	77	1	3	1.350	26.96	8.27
England	2005-2024	192	100	83	2	7	1.204	25.73	8.51
West Indies	2006-2024	202	88	101	3	10	0.871	23.48	8.08
Sri Lanka	2006-2024	192	86	100	4	2	0.860	21.38	7.67

Team	Span	Mat	Won▼	Lost	Tied	NR	W/L	Ave	RPO
Australia	1971-2024	1000	609	348	9	34	1.750	34.15	5.06
India	1974-2023	1055	559	443	9	44	1.261	34.20	5.12
Pakistan	1973-2023	970	512	428	9	21	1.196	30.91	4.87
Sri Lanka	1975-2024	918	421	452	5	40	0.931	29.55	4.94
West Indies	1973-2024	873	420	412	11	30	1.019	29.97	4.80
South Africa	1991-2023	672	410	235	6	21	1.744	35.89	5.17
England	1971-2023	797	400	357	9	31	1.120	31.26	4.99
New Zealand	1973-2023	824	379	395	7	43	0.959	29.55	4.88
Bangladesh	1986-2024	438	159	269	0	10	0.591	26.35	4.67
Zimbabwe	1983-2024	572	151	398	8	15	0.379	25.36	4.55
Ireland	2006-2024	201	80	103	3	15	0.776	27.85	4.89
Afghanistan	2009-2024	166	79	82	1	4	0.963	27.62	4.85
Scotland	1999-2024	156	69	79	1	7	0.873	26.81	4.72
Netherlands	1996-2024	127	43	78	2	4	0.551	25.22	4.66
Kenya	1996-2014	154	42	107	0	5	0.392	23.01	4.25
U.A.E.	1994-2024	111	37	73	1	0	0.506	24.25	4.49
Nepal	2018-2024	66	34	30	1	1	1.133	24.72	4.61
Namibia	2003-2024	52	26	25	0	1	1.040	25.79	4.78
Canada	1979-2024	89	24	63	0	2	0.380	22.84	4.41
Oman	2019-2023	46	23	21	1	1	1.095	29.77	4.74
U.S.A.	2004-2023	51	22	27	2	0	0.814	24.81	4.68
P.N.G.	2014-2023	66	14	51	1	0	0.274	21.21	4.33
Hong Kong	2004-2018	26	9	16	0	1	0.562	23.46	4.46
Bermuda	2006-2009	35	7	28	0	0	0.250	21.68	4.10
Asia XI	2005-2007	7	4	2	0	1	2.000	31.84	5.80
Africa XI	2005-2007	6	1	4	0	1	0.250	25.63	5.32
ICC World XI	2005-2005	4	1	3	0	0	0.333	24.10	5.55
Jersey	2023-2023	5	1	4	0	0	0.250	26.39	4.72
East Africa	1975-1975	3	0	3	0	0	0.000	12.21	2.03

Vision Data



Text Data



Hillary Clinton ✓
@HillaryClinton

Following

"To all the little girls watching...never doubt that you are valuable and powerful & deserving of every chance & opportunity in the world."

4:51 PM - 9 Nov 2016

675,019 Retweets 1,253,910 Likes



47K



675K



1.3M



olivia wilde ✓
@oliviawilde

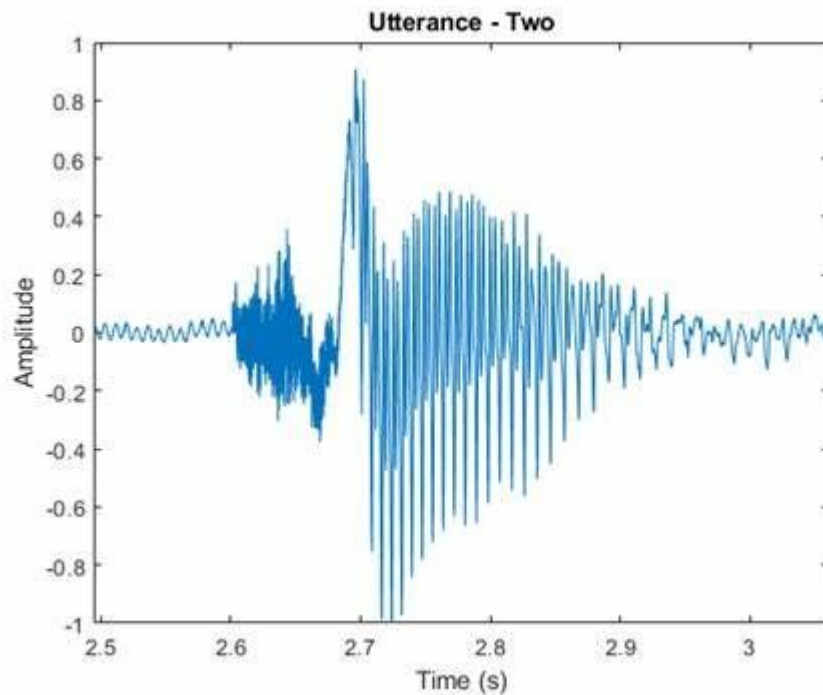
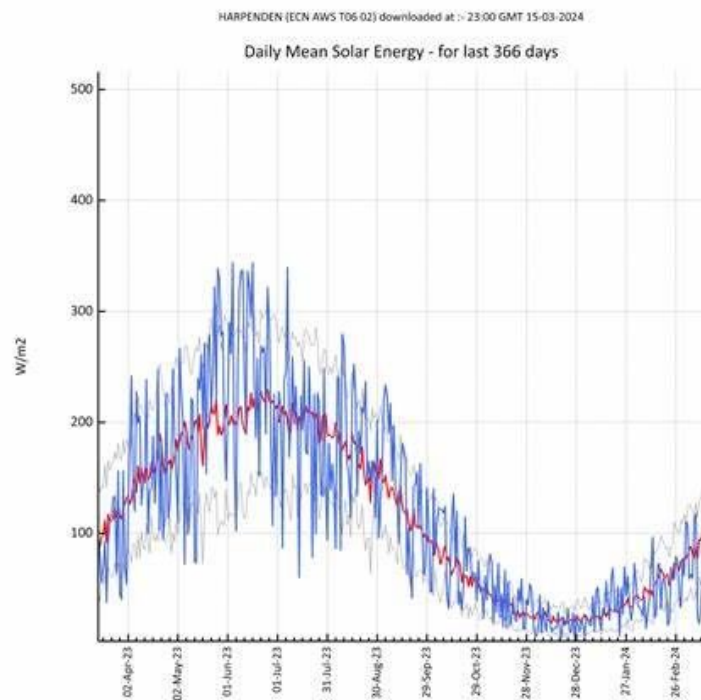
in a thousand years archeologists will dig up tanning beds and think we fried people as punishment.

9:36 PM · 5/14/11 · [Twitter Web Client](#)

2,583 Retweets 5 Quote Tweets 723 Likes



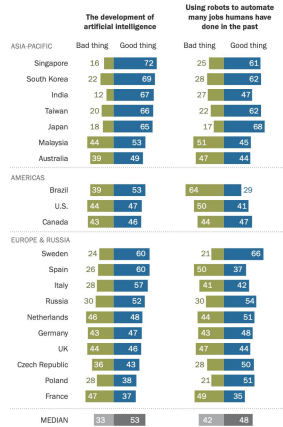
Acoustics Data



Survey Data

Majorities in most Asian publics surveyed see AI as a good thing for society

% who say each of the following has mostly been a ____ for society



Note: Respondents who gave other responses or did not give an answer are not shown.
Source: International Science Survey 2019-2020, Q1-1a%
Science and Scientists Held in High Esteem Across Global Publics

PEW RESEARCH CENTER

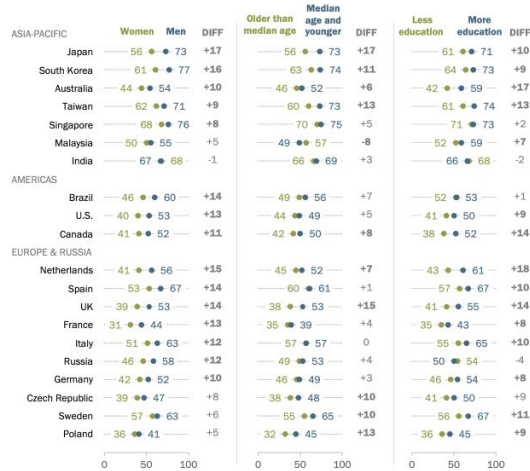
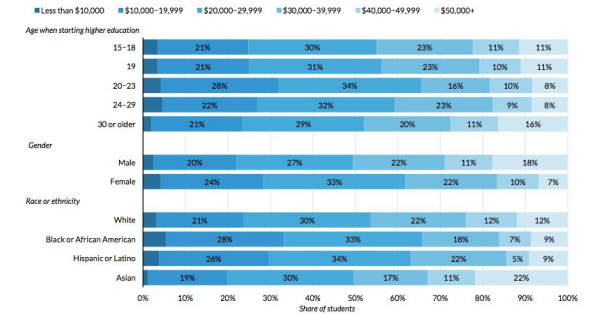


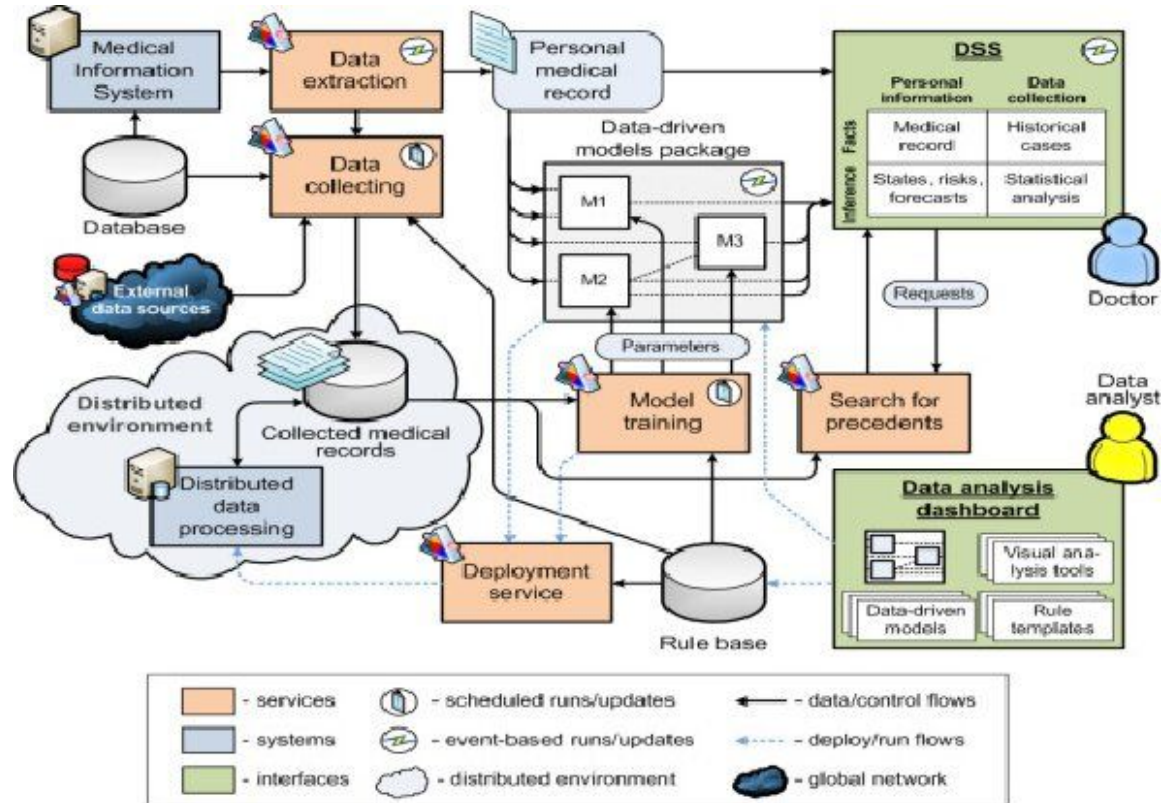
FIGURE 3
Income Six Years after Enrollment
For students who started in the 2003-04 school year



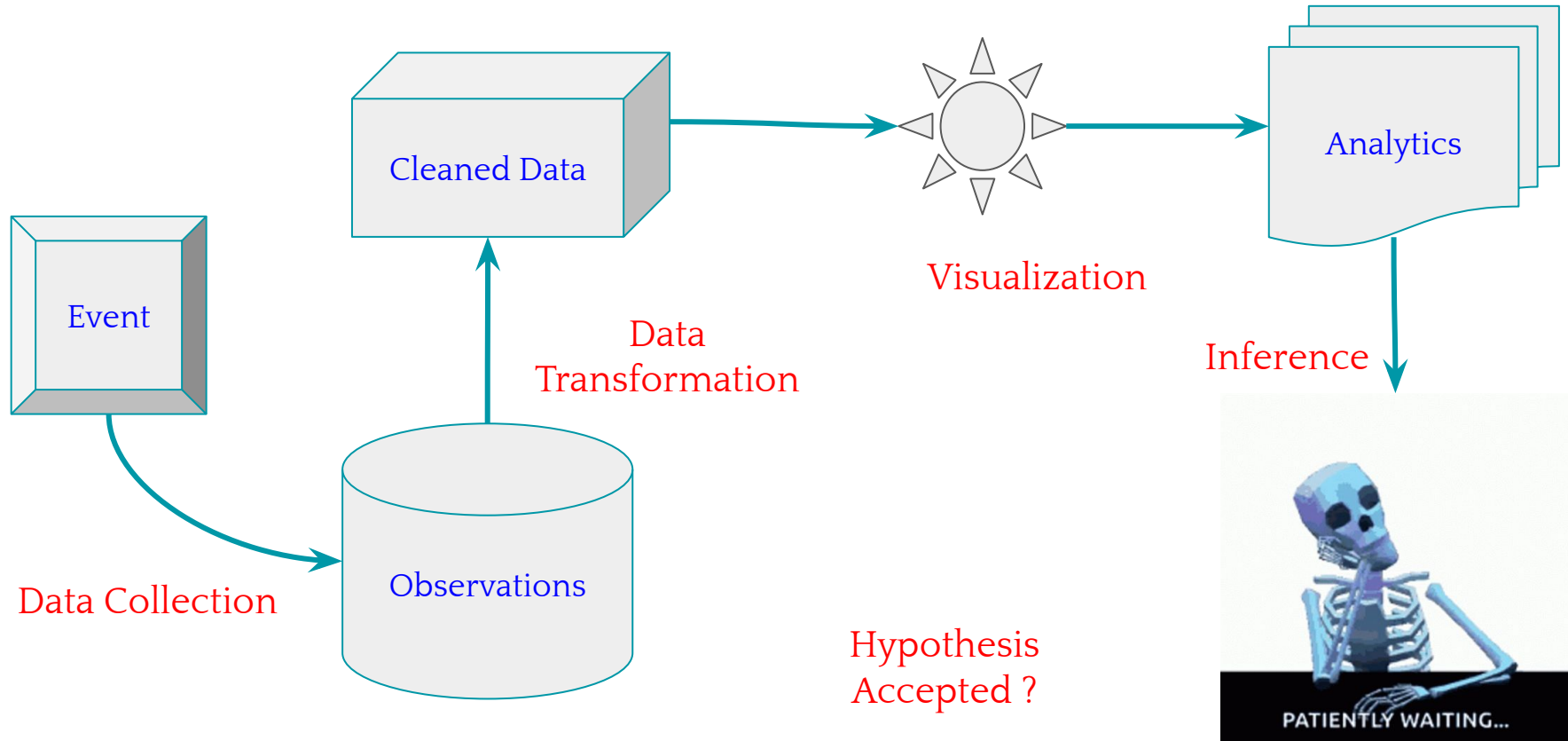
Source: Urban Institute analysis of National Center for Education Statistics, National Postsecondary Student Aid Study (NPSAS), 2004/2009 Beginning Postsecondary Students Longitudinal Study (BPS 04/09), PowerStats, <https://nces.ed.gov/surveys/ips/whatsnew.asp>.

URBAN INSTITUTE

Data Driven Systems



Flow Chart



Quiz-1

1. Write 3 types of data.
2. Write one application related to data.
3. Write 2 things that you liked/understood today.
4. Write 2 things that you did not like/understand today.