S-PLUS Southern Photometric Local Universe Survey

Introduction to Data Science

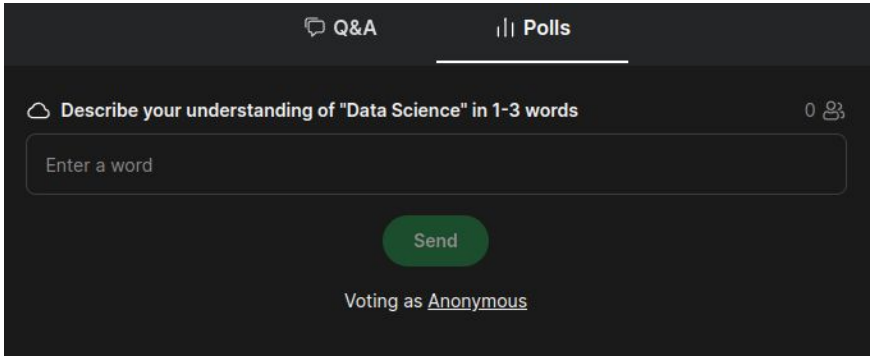
Lilianne Nakazono (lilianne.nakazono@gmail.com)
Postdoc at Instituto de Física, Universidade de São Paulo
(In May: Technology Specialist at Observatório Nacional, Rio de Janeiro)

XI LAPIS

07 April 2025

What do you understand by “Data Science”?

https://www.slido.com/
 # 1217881



Definition tentative

From Cao 2016:

Definition 2.1 (Data Science). A high-level statement is: “data science is the science of data” or **“data science is the study of data”**.

Definition 2.2 (Data Science). From the disciplinary perspective, data science is a new interdisciplinary field that synthesizes and **builds on STATISTICS, INFORMATICS, COMPUTING, COMMUNICATION, management and sociology to study data** and its environments (including domains and other contextual aspects, such as organizational and social aspects) in order to transform data to insights and decisions by following a **data-to-knowledge-to-wisdom** thinking and methodology

Definition tentative

From Cao 2016:

Definition 2.3 (Data Products). A data product is a deliverable from data, or is enabled or driven by data, and can be a **discovery, prediction, service, recommendation, decision-making insight, thinking, model, mode, paradigm, tool or system**.

The ultimate data products of value are **knowledge, intelligence, wisdom and decision**.

data science
Search term

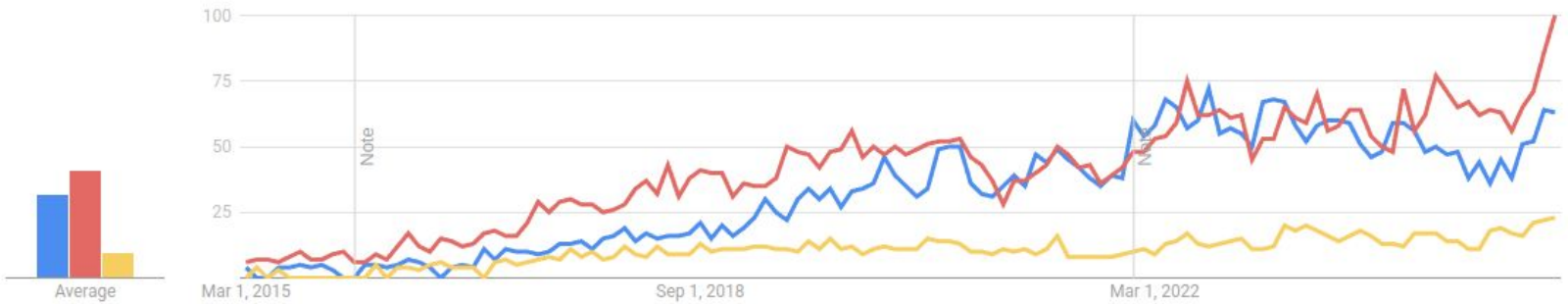
machine learning
Search term

deep learning
Search term

+ Add comparison

Argentina | 3/1/15 - 4/1/25 | All categories | Web Search

Interest over time



source: https://trends.google.com/

What is DATA?

Data can be **structured** or **unstructured**

S-PLUS: structured data

	ID	RA	DEC	A	B	KRON_RAD...	FWHM_n	MU_MAX_I...	ISOarea	SEX_FLAGS_DET	u_iso	e_u_iso	u_aper_3	e_u_aper_3	u_aper_6	e_u_aper_(
1	IDR5_3_STRIPE82-0003_0000001	0,74137	-1,37967	0,000204	0,000102	7,90908	1,96348	20,6443	2,33410E-	0	99,	99,	99,	99,	99,	99,
2	IDR5_3_STRIPE82-0003_0000002	0,70738	-1,37958	0,000204	0,000165	3,21479	1,45574	19,9723	1,86728E-	0	23,2554	1,06408	23,1281	0,957706	21,5562	31,3706
3	IDR5_3_STRIPE82-0003_0000003	0,73549	-1,37937	0,000213	0,000122	7,04948	2,20191	20,7577	9,33642E-	0	99,	99,	99,	99,	99,	99,
4	IDR5_3_STRIPE82-0003_0000004	0,74365	-1,37889	0,000217	0,00011	4,63158	2,11113	20,5659	9,33642E-	0	24,7683	3,2639	23,217	0,998706	21,5698	0,45892
5	IDR5_3_STRIPE82-0003_0000005	0,74688	-1,379	0,000272	0,000204	5,16087	2,09552	19,4571	4,43480E-	3	23,7966	2,32876	25,3006	6,78678	21,8455	0,59198
6	IDR5_3_STRIPE82-0003_0000006	0,85699	-1,37858	0,000111	6,89582E-5	1,82	1,1778	20,8298	2,33410E-	0	99,	99,	99,	99,	99,	99,
7	IDR5_3_STRIPE82-0003_0000007	0,82666	-1,37868	0,000205	0,000112	1,82	2,06868	20,3826	4,66821E-	0	22,5336	0,40192	22,7921	0,696414	99,	99,
8	IDR5_3_STRIPE82-0003_0000008	0,6603	-1,38008	0,000292	0,000256	4,19813	1,20017	18,2692	9,10301E-	1	99,	99,	99,	99,	99,	99,
9	IDR5_3_STRIPE82-0003_0000009	0,75732	-1,37836	0,00021	0,000187	3,08624	0,972412	19,2498	3,26775E-	0	22,159	0,432417	22,9856	0,81089	21,2552	0,32728
10	IDR5_3_STRIPE82-0003_0000010	0,89848	-1,37818	0,000134	0,000131	2,66863	1,03745	20,2072	1,16705E-	0	99,	99,	99,	99,	99,	99,
11	IDR5_3_STRIPE82-0003_0000011	0,94574	-1,37814	0,00022	0,000122	1,82	1,35644	20,7338	4,66821E-	0	24,0311	6,64685E16	24,0643	103,699	99,	99,
12	IDR5_3_STRIPE82-0003_0000012	0,79429	-1,37807	0,000221	0,000168	3,02718	2,63501	20,5056	1,16705E-	0	21,5566	0,213114	21,6278	0,230947	20,882	0,22876
13	IDR5_3_STRIPE82-0003_0000013	0,71954	-1,37868	0,000292	0,000255	3,23533	1,55785	18,8604	8,63619E-	0	21,3403	0,300883	22,0952	0,361057	21,3578	0,36120
14	IDR5_3_STRIPE82-0003_0000014	1,01727	-1,37766	7,69682E-5	7,55487E-5	0,	1,01823	20,8327	2,33410E-	0	99,	99,	99,	99,	99,	99,
15	IDR5_3_STRIPE82-0003_0000015	0,95699	-1,3775	0,000119	7,62368E-5	5,41854	0,	20,9444	0,	0	99,	99,	99,	99,	99,	99,
16	IDR5_3_STRIPE82-0003_0000016	0,8579	-1,37747	7,64183E-5	7,63103E-5	8,79959	0,057328	20,8067	2,33410E-	0	25,2311	2,53188	23,7288	1,55001	99,	99,
17	IDR5_3_STRIPE82-0003_0000017	0,74631	-1,37833	0,000282	0,000263	5,75973	1,78762	19,4777	6,06867E-	3	21,5906	0,337879	22,9025	0,741812	21,1216	0,28737
18	IDR5_3_STRIPE82-0003_0000018	0,9249	-1,37865	0,000498	0,000287	6,65712	5,18143	19,4769	7,93596E-	3	99,	99,	99,	99,	99,	99,
19	IDR5_3_STRIPE82-0003_0000019	0,76685	-1,37714	0,000216	8,55751E-5	7,39486	0,	20,603	2,33410E-	0	23,2296	0,634239	23,1478	0,903732	21,0223	0,25645
20	IDR5_3_STRIPE82-0003_0000020	0,75258	-1,37687	0,000207	0,000131	1,82	2,17456	20,4954	7,00231E-	0	26,0186	10,5816	24,5588	3,2554	21,4485	0,37375
21	IDR5_3_STRIPE82-0003_0000021	0,70199	-1,37828	0,000245	0,00022	4,29991	1,04647	18,5862	5,83526E-	0	99,	99,	99,	99,	99,	99,
22	IDR5_3_STRIPE82-0003_0000022	0,7408	-1,37669	0,000182	0,000146	1,82	2,19659	20,5176	4,66821E-	0	23,1692	0,747867	22,9594	0,748347	23,1256	1,73483
23	IDR5_3_STRIPE82-0003_0000023	0,72492	-1,3768	0,000164	0,000158	6,97142	2,006	20,5894	1,16705E-	0	22,9375	0,697542	23,8203	1,70471	22,1718	0,74500
24	IDR5_3_STRIPE82-0003_0000024	0,94936	-1,37737	0,000225	0,000197	3,61028	1,12194	18,9935	4,20139E-	0	99,	99,	99,	99,	99,	99,
25	IDR5_3_STRIPE82-0003_0000025	0,65759	-1,3765	7,75900E-5	7,51032E-5	3,28595	1,01948	20,6432	4,66821E-	0	99,	99,	99,	99,	99,	99,
26	IDR5_3_STRIPE82-0003_0000026	1,00225	-1,37745	0,00037	0,000285	5,18488	1,83223	18,9359	1,07369E-	1	99,	99,	99,	99,	99,	99,
27	IDR5_3_STRIPE82-0003_0000027	1,01577	-1,37639	0,000224	0,000198	1,82	1,83723	19,7703	3,03434E-	0	99,	99,	99,	99,	99,	99,
28	IDR5_3_STRIPE82-0003_0000028	1,23575	-1,37562	0,000173	4,40968E-5	10,6275	0,	20,2855	0,	1	99,	99,	99,	99,	99,	99,
29	IDR5_3_STRIPE82-0003_0000029	0,79698	-1,37584	0,000253	0,000199	4,38289	2,29447	20,1695	3,50116E-	0	21,2004	0,194846	21,6303	0,224978	20,4293	0,14685
30	IDR5_3_STRIPE82-0003_0000030	0,68581	-1,37553	0,0002	8,60425E-5	7,9019	1,78329	20,3299	2,33410E-	0	99,	99,	99,	99,	99,	99,
31	IDR5_3_STRIPE82-0003_0000031	0,97711	-1,37836	0,000363	0,000205	3,63235	1,32698	14,4733	2,07735E-	3	99,	99,	99,	99,	99,	99,
32	IDR5_3_STRIPE82-0003_0000032	0,76911	-1,37979	0,000394	0,00029	3,18298	1,45823	13,9091	5,27508E-	1	17,5836	9,14209E14	18,684	3,79314	18,0113	2,41875
33	IDR5_3_STRIPE82-0003_0000033	0,8323	-1,37848	0,000313	0,000291	3			2,54417E-	0	19,3019	1,15348E15	19,6907	0,047701	19,3707	0,06509
34	IDR5_3_STRIPE82-0003_0000034	1,21271	-1,37524	0,000439	7,63313E-5	1		2	0,	0	99,	99,	99,	99,	99,	99,
35	IDR5_3_STRIPE82-0003_0000035	1,18802	-1,37518	0,000148	0,000106	1		2	0,	0	99,	99,	99,	99,	99,	99,
36	IDR5_3_STRIPE82-0003_0000036	1,11072	-1,37503	0,000199	0,000156	4,45234	2,40854	20,7195	9,33642E-	0	99,	99,	99,	99,	99,	99,
37	IDR5_3_STRIPE82-0003_0000037	0,80183	-1,37496	0,00025	0,000189	5,55002	2,40982	20,2972	2,33411E-	0	21,8853	0,311598	22,1277	0,347355	20,9948	0,24284
38	IDR5_3_STRIPE82-0003_0000038	0,93145	-1,37554	0,000219	0,000208	3,49983	1,03275	18,5362	5,13503E-	0	99,	99,	23,8945	1,77801	99,	99,
39	IDR5_3_STRIPE82-0003_0000039	1,29028	-1,37493	0,000204	0,0001	6,90245	1,22901	18,9456	1,16705E-	1	23,7204	8,87412E16	24,265	270,153	99,	99,
40	IDR5_3_STRIPE82-0003_0000040	0,92322	-1,37771	0,000359	0,000321	5,30273	1,06517	18,1427	1,00367E-	3	99	99	99	99	99	99

qualitative

S-PLUS: structured data

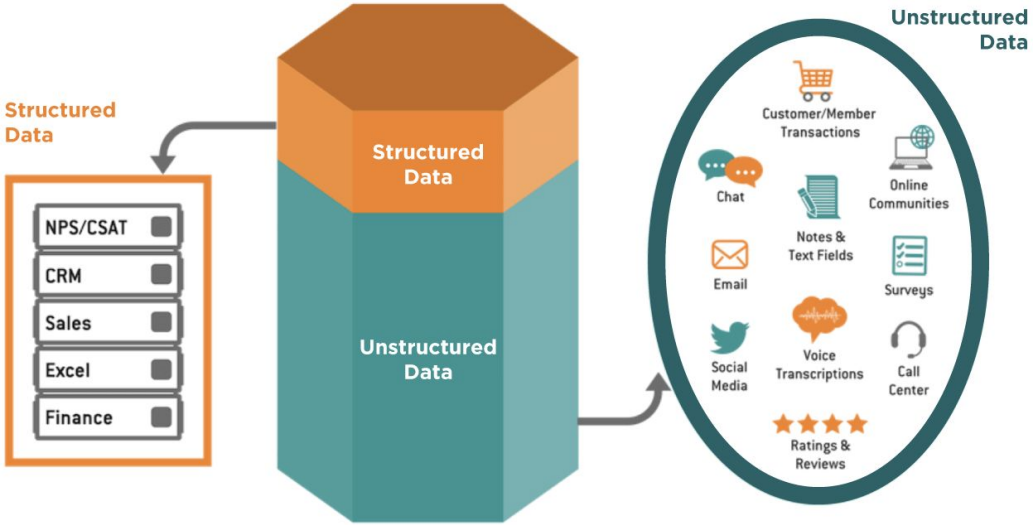
	ID	RA	DEC	A	B	KRON_RAD...	FWHM_n	MU_MAX_I...	ISOarea	SEX_FLAGS_DET	u_iso	e_u_iso	u_aper_3	e_u_aper_3	u_aper_6	e_u_aper_(
1	iDR5_3_STRIPE82-0003_0000001	0,74137	-1,37967	0,000204	0,000102	7,90908	1,96348	20,6443	2,33410E-8	0	99,	99,	99,	99,	99,	99,
2	iDR5_3_STRIPE82-0003_0000002	0,70738	-1,37958	0,000204	0,000165	3,21479	1,45574	19,9723	1,86728E-7	0	23,2554	1,06408	23,1281	0,957706	21,5562	31,3706
3	iDR5_3_STRIPE82-0003_0000003	0,73549	-1,37937	0,000213	0,000122	7,04948	2,20191	20,7577	9,33642E-8	0	99,	99,	99,	99,	99,	99,
4	iDR5_3_STRIPE82-0003_0000004	0,74365	-1,37889	0,000217	0,00011	4,63158	2,11113	20,5659	9,33642E-8	0	24,7683	3,2639	23,217	0,998706	21,5698	0,45892
5	iDR5_3_STRIPE82-0003_0000005	0,74688	-1,379	0,000272	0,000204	5,16087	2,09552	19,4571	4,43480E-7	3	23,7966	2,32876	25,3006	6,78678	21,8455	0,59198
6	iDR5_3_STRIPE82-0003_0000006	0,85699	-1,37858	0,000111	6,89582E-5	1,82	1,1778	20,8298	2,33410E-8	0	99,	99,	99,	99,	99,	99,
7	iDR5_3_STRIPE82-0003_0000007	0,82666	-1,37868	0,000205	0,000112	1,82	2,06868	20,3826	4,66821E-8	0	22,5336	0,40192	22,7921	0,696414	99,	99,
8	iDR5_3_STRIPE82-0003_0000008	0,6603	-1,38008	0,000292	0,000256	4,19813	1,20017	18,2692	9,10301E-7	1	99,	99,	99,	99,	99,	99,
9	iDR5_3_STRIPE82-0003_0000009	0,75732	-1,37836	0,00021	0,000187	3,08624	0,972412	19,2498	3,26775E-7	0	22,159	0,432417	22,9856	0,81089	21,2552	0,32728
10	iDR5_3_STRIPE82-0003_0000010	0,89848	-1,37818	0,000134	0,000131	2,66863	1,03745	20,2072	1,16705E-7	0	99,	99,	99,	99,	99,	99,
11	iDR5_3_STRIPE82-0003_0000011	0,94574	-1,37814	0,00022	0,000122	1,82	1,35644	20,7338	4,66821E-8	0	24,0311	6,64685E16	24,0643	103,699	99,	99,
12	iDR5_3_STRIPE82-0003_0000012	0,79429	-1,37807	0,000221	0,000168	3,02718	2,63501	20,5056	1,16705E-7	0	21,5566	0,213114	21,6278	0,230947	20,882	0,22876
13	iDR5_3_STRIPE82-0003_0000013	0,71954	-1,37868	0,000292	0,000255	3,23533	1,55785	18,8604	8,63619E-7	0	21,3403	0,300883	22,0952	0,361057	21,3578	0,36120
14	iDR5_3_STRIPE82-0003_0000014	1,01727	-1,37766	7,69682E-5	7,55487E-5	0,	1,01823	20,8327	2,33410E-8	0	99,	99,	99,	99,	99,	99,
15	iDR5_3_STRIPE82-0003_0000015	0,95699	-1,3775	0,000119	7,62368E-5	5,41854	0,	20,9444	0,	0	99,	99,	99,	99,	99,	99,
16	iDR5_3_STRIPE82-0003_0000016	0,8579	-1,37747	7,64183E-5	7,63103E-5	8,79959	0,057328	20,8067	2,33410E-8	0	25,2311	2,53188	23,7288	1,55001	99,	99,
17	iDR5_3_STRIPE82-0003_0000017	0,74631	-1,37833	0,000282	0,000263	5,75973	1,78762	19,4777	6,06867E-7	3	21,5906	0,337879	22,9025	0,741812	21,1216	0,28737
18	iDR5_3_STRIPE82-0003_0000018	0,9249	-1,37865	0,000498	0,000287	6,65712	5,18143	19,4769	7,93596E-7	3	99,	99,	99,	99,	99,	99,
19	iDR5_3_STRIPE82-0003_0000019	0,76685	-1,37714	0,000216	8,55751E-5	7,39486	0,	20,603	2,33410E-8	0	23,2296	0,634239	23,1478	0,903732	21,0223	0,25645
20	iDR5_3_STRIPE82-0003_0000020	0,75258	-1,37687	0,000207	0,000131	1,82	2,17456	20,4954	7,00231E-8	0	26,0186	10,5816	24,5588	3,2554	21,4485	0,37375
21	iDR5_3_STRIPE82-0003_0000021	0,70199	-1,37828	0,000245	0,00022	4,29991	1,04647	18,5862	5,83526E-7	0	99,	99,	99,	99,	99,	99,
22	iDR5_3_STRIPE82-0003_0000022	0,7408	-1,37669	0,000182	0,000146	1,82	2,19659	20,5176	4,66821E-8	0	23,1692	0,747867	22,9594	0,748347	23,1256	1,73483
23	iDR5_3_STRIPE82-0003_0000023	0,72492	-1,3768	0,000164	0,000158	6,97142	2,006	20,5894	1,16705E-7	0	22,9375	0,697542	23,8203	1,70471	22,1718	0,74500
24	iDR5_3_STRIPE82-0003_0000024	0,94936	-1,37737	0,000225	0,000197	3,61028	1,12194	18,9935	4,20139E-7	0	99,	99,	99,	99,	99,	99,
25	iDR5_3_STRIPE82-0003_0000025	0,65759	-1,3765	7,75900E-5	7,51032E-5	3,28595	1,01948	20,6432	4,66821E-8	0	99,	99,	99,	99,	99,	99,
26	iDR5_3_STRIPE82-0003_0000026	1,00225	-1,37745	0,00037	0,000285	5,18488	1,83223	18,9359	1,07369E-6	1	99,	99,	99,	99,	99,	99,
27	iDR5_3_STRIPE82-0003_0000027	1,01577	-1,37639	0,000224	0,000198	1,82	1,83723	19,7703	3,03434E-7	0	99,	99,	99,	99,	99,	99,
28	iDR5_3_STRIPE82-0003_0000028	1,23575	-1,37562	0,000173	4,40968E-5	10,6275	0,	20,2855	0,	1	99,	99,	99,	99,	99,	99,
29	iDR5_3_STRIPE82-0003_0000029	0,79698	-1,37584	0,000253	0,000199	4,38289	2,29447	20,1695	3,50116E-7	0	21,2004	0,194846	21,6303	0,224978	20,4293	0,14685
30	iDR5_3_STRIPE82-0003_0000030	0,68581	-1,37553	0,0002	8,60425E-5	7,9019	1,78329	20,3299	2,33410E-8	0	99,	99,	99,	99,	99,	99,
31	iDR5_3_STRIPE82-0003_0000031	0,97711	-1,37836	0,000363	0,000205	3,63235	1,32698	14,4733	2,07			99,	99,	99,	99,	99,
32	iDR5_3_STRIPE82-0003_0000032	0,76911	-1,37979	0,000394	0,00029	3,18298	1,45823	13,9091	5,27	6		9,14209E14	18,684	3,79314	18,0113	2,41875
33	iDR5_3_STRIPE82-0003_0000033	0,8323	-1,37848	0,000313	0,000291	3,50243	0,991016	15,638	2,54	9		1,15348E15	19,6907	0,047701	19,3707	0,06509
34	iDR5_3_STRIPE82-0003_0000034	1,21271	-1,37524	0,000439	7,63313E-5	1,82	0,	20,5332	0,	0	99,	99,	99,	99,	99,	99,
35	iDR5_3_STRIPE82-0003_0000035	1,18802	-1,37518	0,000148	0,000106	1,82	0,	20,9452	0,	0	99,	99,	99,	99,	99,	99,
36	iDR5_3_STRIPE82-0003_0000036	1,11072	-1,37503	0,000199	0,000156	4,45234	2,40854	20,7195	9,33642E-8	0	99,	99,	99,	99,	99,	99,
37	iDR5_3_STRIPE82-0003_0000037	0,80183	-1,37496	0,00025	0,000189	5,55002	2,40982	20,2972	2,33411E-7	0	21,8853	0,311598	22,1277	0,347355	20,9948	0,24284
38	iDR5_3_STRIPE82-0003_0000038	0,93145	-1,37554	0,000219	0,000208	3,49983	1,03275	18,5362	5,13503E-7	0	99,	99,	23,8945	1,77801	99,	99,
39	iDR5_3_STRIPE82-0003_0000039	1,29028	-1,37493	0,000204	0,0001	6,90245	1,22901	18,9456	1,16705E-7	1	23,7204	8,87412E16	24,265	270,153	99,	99,
40	iDR5_3_STRIPE82-0003_0000040	0,92322	-1,37771	0,000359	0,000321	5,30273	1,06517	18,1427	1,00367E-6	3	99	99	99	99	99	99

quantitative

S-PLUS: unstructured data

What is DATA?

Data can be **structured** or **unstructured**



(in business context)

“Classical” statistical analysis (Frequentist)

1. Research Planning

   Objectives and hypothesis are defined uphand, **before** data collection

2. Data collection

   Sample data assuming certain characteristics of the population (such as probability distribution)

3. Statistical Analysis

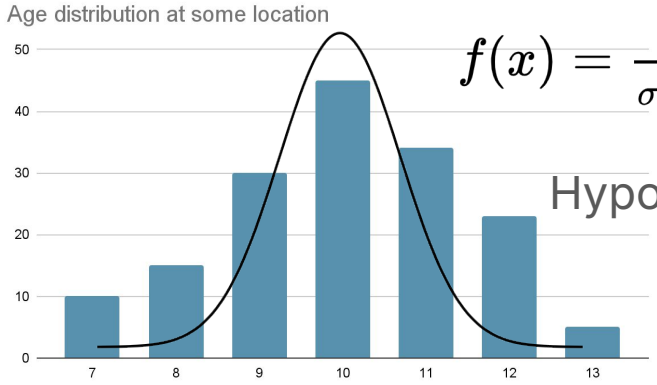
   a. Exploratory data analysis

   b. Hypothesis test → is there enough statistical evidence to reject an hypothesis?

Observation: Bayesian Statistics shifts how we understand probabilities, opposed to Frequentist Statistics. This will be taught in Laerte’s lecture in this course!

“Classical” statistical analysis (Frequentist)

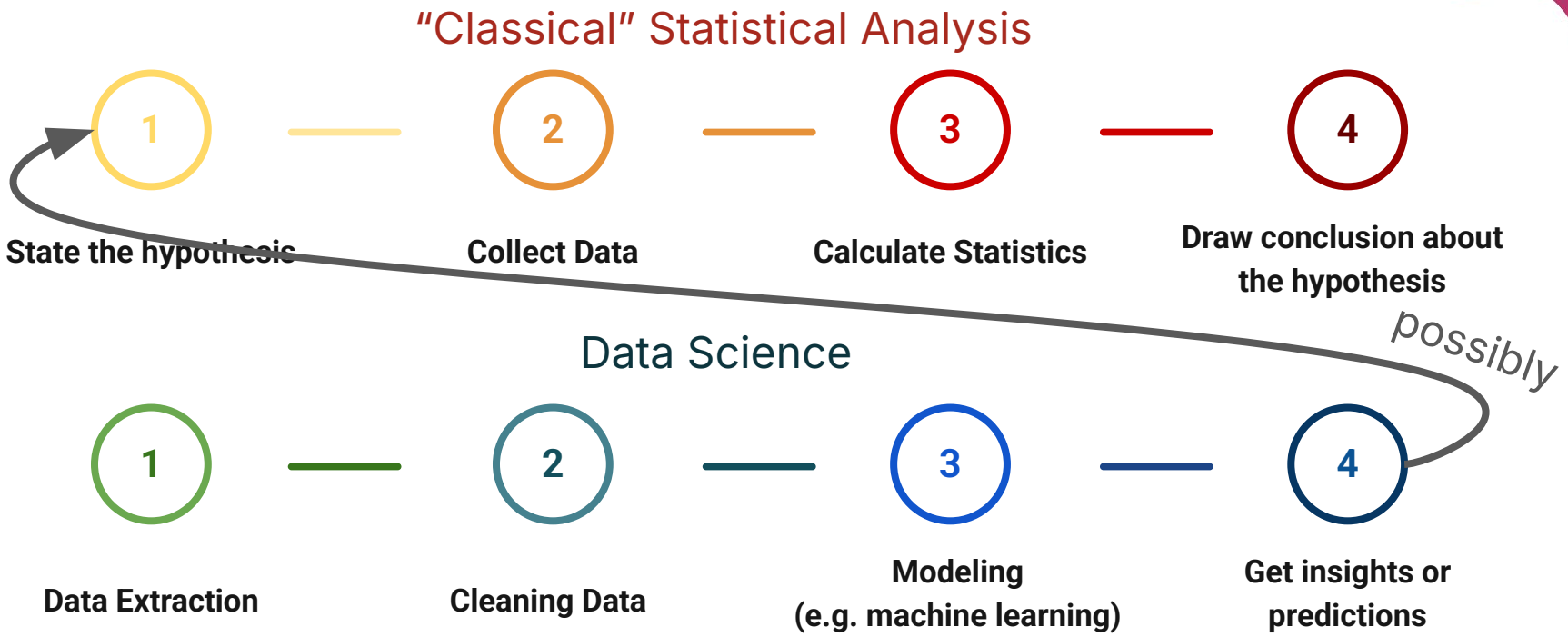
Example: On a natural population it is assumed that the age of all individuals follow a normal distribution. In a rattlesnake population age can be measured using the keratin at the tip of its tail. This hypothesis can be tested, after data collection, using Shapiro-Wilk test. **If the distribution is not following a normal distribution we can assume that some ecological imbalance may be occurring**.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Hypothesis test: 👍 or 👎 ?

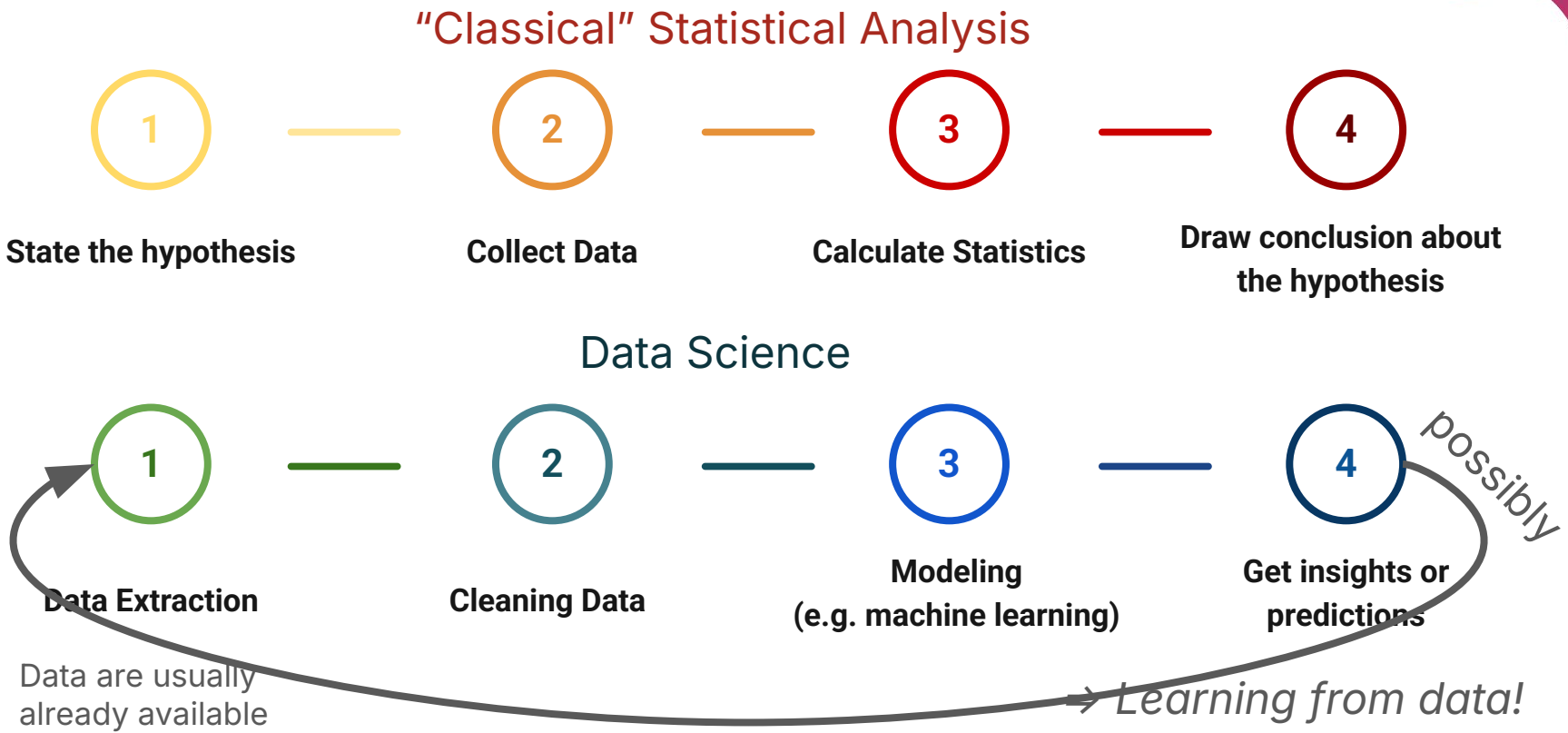
Steps (very broadly!)



Data are usually already available

⇒ *Learning from data!*

Steps (very broadly!)



Exploratory Data Analysis

Understanding your data is crucial in any case. But how?

- Measures of **Central Tendency**

  Mean, Median, Mode

- Measures of **Dispersion**

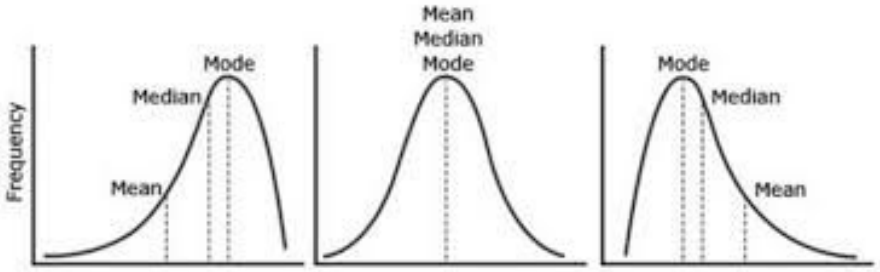
  Variance, Standard Deviation

- Measures of **Position**

  Percentiles, Quartiles

- Measures of **Shape**

  Skewness, Kurtosis



Exploratory Data Analysis

Understanding your data is crucial in any case. But how?

- Measures of **Central Tendency**

  Mean, Median, Mode

- Measures of **Dispersion**

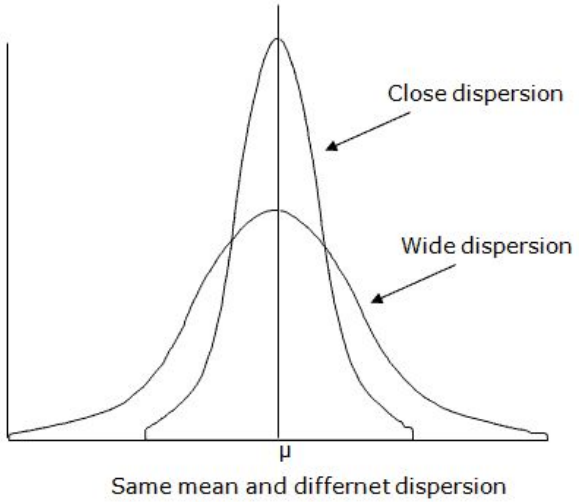
  Variance, Standard Deviation

- Measures of **Position**

  Percentiles, Quartiles

- Measures of **Shape**

  Skewness, Kurtosis



Same mean and differnet dispersion

Exploratory Data Analysis

Understanding your data is crucial in any case. But how?

- Measures of **Central Tendency**

  Mean, Median, Mode

- Measures of **Dispersion**

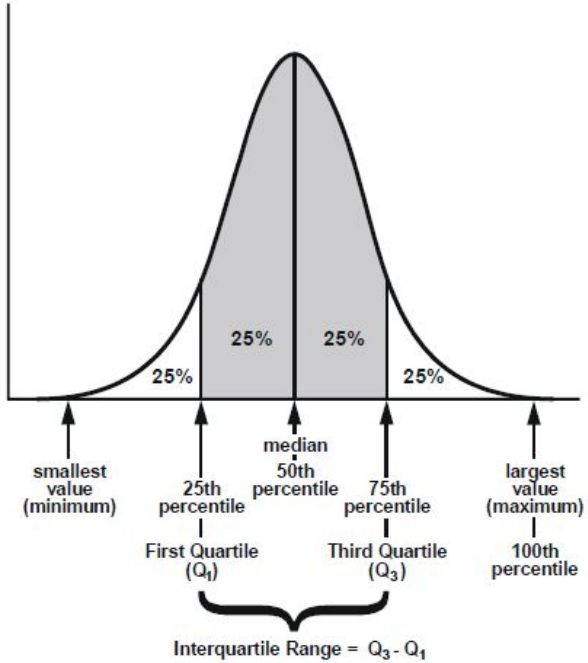
  Variance, Standard Deviation

- Measures of **Position**

  Percentiles, Quartiles

- Measures of **Shape**

  Skewness, Kurtosis



Exploratory Data Analysis

Understanding your data is crucial in any case. But how?

- Measures of **Central Tendency**

  Mean, Median, Mode

- Measures of **Dispersion**

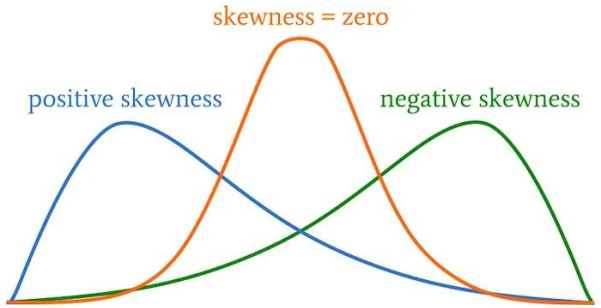
  Variance, Standard Deviation

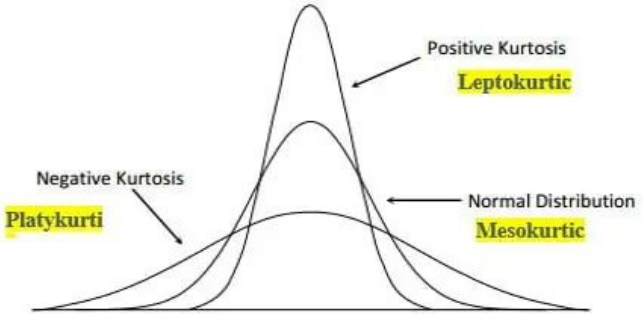
- Measures of **Position**

  Percentiles, Quartiles

- Measures of **Shape**

  Skewness, Kurtosis

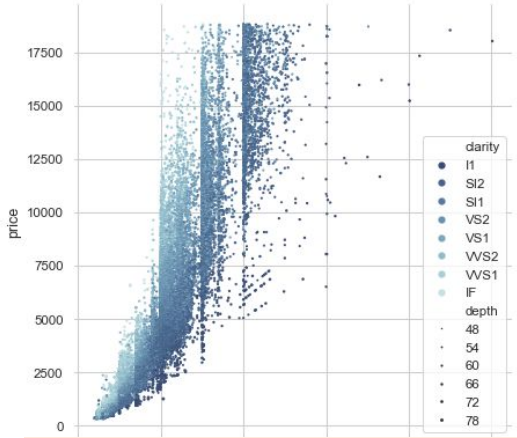




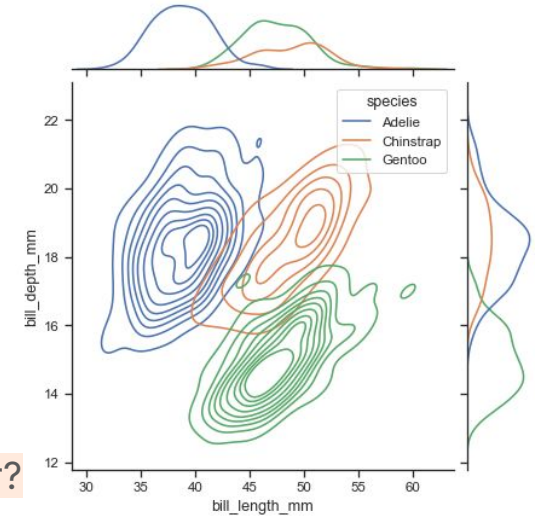
Exploratory Data Analysis: Data Visualization

Some useful plots:

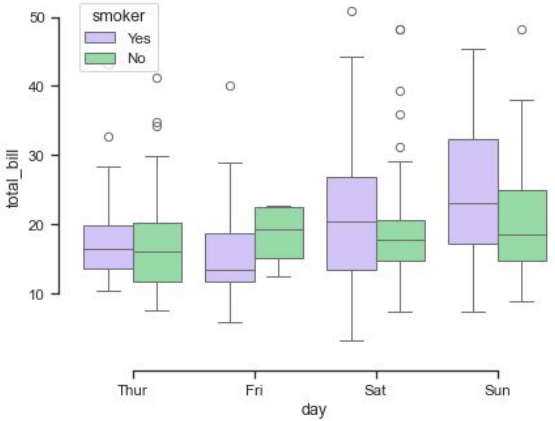
Scatterplot



Jointplot



Boxplot

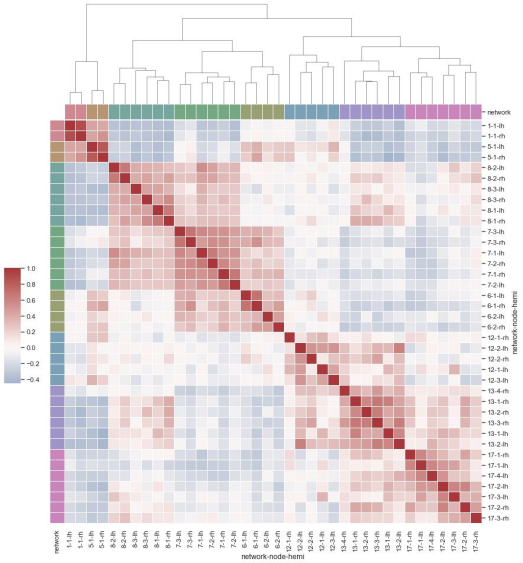


Are there any unusual behaviour?
Is anything out of expected?
Are there outliers?

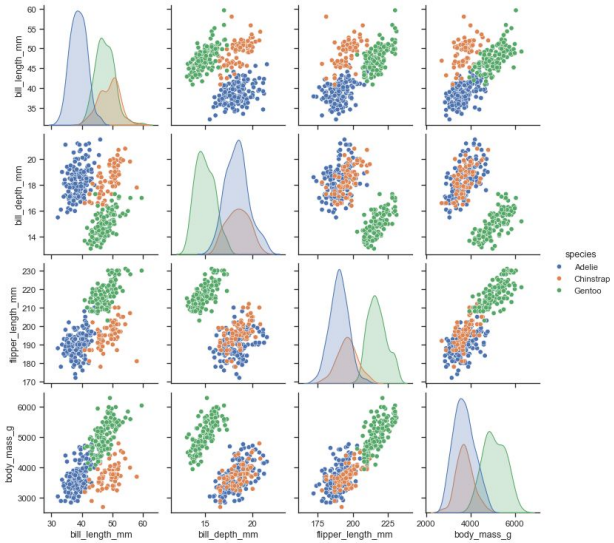
Exploratory Data Analysis: Data Visualization

Some useful plots:

Correlation Matrix



Scatterplot Matrix



Are your features strongly correlated?
Are there any indication of different behaviours for different classes?

Programming Languages

Up to date, it is the most used programming language for data science

In Astronomy, we use **ADQL** *(Astronomical Data Query Language)* to query structured data. Syntax is very similar to SQL

Python packages

Astronomy-specific

splusdata

LSDB

Data manipulation

Mathematical computing

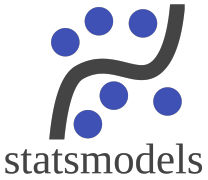
Image manipulation

OpenCV

Data Visualization

Machine Learning

Statistical Modeling



Deep Learning

Tools: code version control

Web-based platforms that hosts Git repositories:

Installing via command line in Linux: `sudo apt-get install git`

Tools: code editor

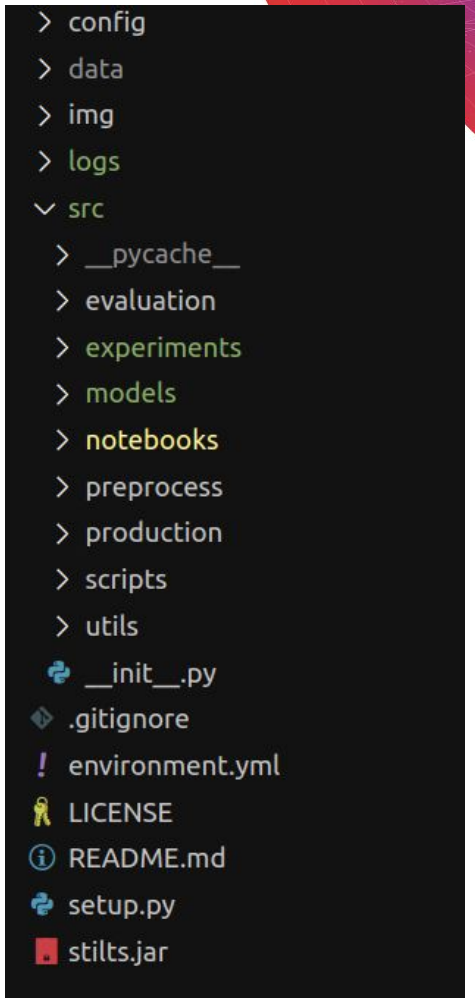
My personal recommendation:

With these extensions installed: Python, Jupyter, Remote - SSH, GitHub Copilot (plus others... Those make your programming life much much easier, trust me!)

Programming best practices

Your data science project starts with setting up code environment (e.g. pyenv, conda) and properly organizing your code files!

One of my repositories



If you ever feel lost with terminology during this course, check this extensive list of cheatsheets:

https://github.com/FavioVazquez/ds-cheatsheets?tab=readme-ov-file

Python For Data Science *Cheat Sheet*

Matplotlib

Learn Python Interactively at www.DataCamp.com

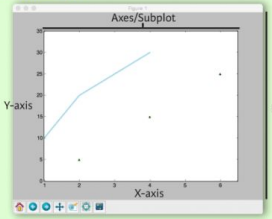
Matplotlib

Matplotlib is a Python 2D plotting library which produces publication-quality figures in a variety of hardcopy formats and interactive environments across platforms.

matplotlib

Plot Anatomy & Workflow

Plot Anatomy



Workflow

The basic steps to creating plots with matplotlib are:

1 Prepare data 2 Create plot 3 Plot 4 Customize plot 5 Save plot 6 Show plot

```
>>> import matplotlib.pyplot as plt
>>> x = [1,2,3,4]                  Step 1
>>> y = [10,20,25,30]
>>> fig = plt.figure()             Step 2
>>> ax = fig.add_subplot(111)      Step 3
>>> ax.plot(x, y, color='lightblue', linewidth=3)  Step 3, 4
>>> ax.scatter([2,4,6],
               [5,15,25],
               color='darkgreen',
               marker='^')
>>> ax.set_xlim(1, 6.5)
>>> plt.savefig('foo.png')
>>> plt.show()                     Step 6
```

1 Prepare The Data

Also see Lists & NumPy

1D Data

```
>>> import numpy as np
>>> x = np.linspace(0, 10, 100)
>>> y = np.cos(x)
>>> z = np.sin(x)
```

2D Data or Images

```
>>> data = 2 * np.random.random((10, 10))
>>> data2 = 3 * np.random.random((10, 10))
>>> Y, X = np.mgrid[-3:3:100j, -3:3:100j]
>>> U = -1 - X**2 + Y
>>> V = 1 + X - Y**2
>>> from matplotlib.cbook import get_sample_data
>>> img = np.load(get_sample_data('axes_grid/bivariate_normal.npy'))
```

2 Create Plot

```
>>> import matplotlib.pyplot as plt
```

Figure

```
>>> fig = plt.figure()
>>> fig2 = plt.figure(figsize=plt.figaspect(2.0))
```

Axes

All plotting is done with respect to an `Axes`. In most cases, a subplot will fit your needs. A subplot is an axes on a grid system.

```
>>> fig.add_axes()
>>> ax1 = fig.add_subplot(221) # row-col-num
>>> ax3 = fig.add_subplot(212)
>>> fig3, axes = plt.subplots(nrows=2,ncols=2)
>>> fig4, axes2 = plt.subplots(ncols=3)
```

4 Customize Plot

Colors, Color Bars & Color Maps

```
>>> plt.plot(x, x, x, x**2, x, x**3)
>>> ax.plot(x, y, alpha = 0.4)
>>> ax.plot(x, y, c='k')
>>> fig.colorbar(im, orientation='horizontal')
>>> im = ax.imshow(img,
                   cmap='seismic')
```

Markers

```
>>> fig, ax = plt.subplots()
>>> ax.scatter(x,y,marker=".")
>>> ax.plot(x,y,marker="o")
```

Linestyles

```
>>> plt.plot(x,y,linewidth=4.0)
>>> plt.plot(x,y,ls='solid')
>>> plt.plot(x,y,ls='--')
>>> plt.plot(x,y,'--',x**2,y**2,'-.')
>>> plt.setp(lines,color='r',linewidth=4.0)
```

Text & Annotations

```
>>> ax.text(1,
            -2.1,
            'Example Graph',
            style='italic')
>>> ax.annotate("Sine",
                xy=(8, 0),
                xycoords='data',
                xytext=(10.5, 0),
                textcoords='data',
                arrowprops=dict(arrowstyle="->",
                                connectionstyle="arc3"),)
```

Mathtext

```
>>> plt.title(r'$sigma_i=15$', fontsize=20)
```

Limits, Legends & Layouts

Limits & Autoscaling

```
>>> ax.margins(x=0.0,y=0.1)          Add padding to a plot
>>> ax.axis('equal')                 Set the aspect ratio of the plot to 1
>>> ax.set(xlim=[0,10.5],ylim=[-1.5,1.5])  Set limits for x-and y-axis
>>> ax.set_xlim(0,10.5)              Set limits for x-axis
```

Legends

```
>>> ax.set(title='An Example Axes',   Set a title and x-and y-axis labels
           ylabel='Y-Axis',
           xlabel='X-Axis')
>>> ax.legend(loc='best')            No overlapping plot elements
```

Ticks

```
>>> ax.xaxis.set(ticks=range(1,5),    Manually set x-ticks
                 ticklabels=[3,100,-12,"foo"])
>>> ax.tick_params(axis='y',          Make y-ticks longer and go in and out
                   direction='inout',
                   length=10)
```

Subplot Spacing

```
>>> fig3.subplots_adjust(wspace=0.5,  Adjust the spacing between subplots
                         hspace=0.3,
                         left=0.125,
                         right=0.9,
                         top=0.9,
                         bottom=0.1)
>>> fig.tight_layout()               Fit subplot(s) in to the figure area
```

Axis Spines

```
>>> ax1.spines['top'].set_visible(False)   Make the top axis line for a plot invisible
>>> ax1.spines['bottom'].set_position(('outward',10))  Move the bottom axis line outward
```

3 Plotting Routines

1D Data

```
>>> fig, ax = plt.subplots()
>>> lines = ax.plot(x,y)              Draw points with lines or markers connecting them
>>> ax.scatter(x,y)                   Draw unconnected points, scaled or colored
>>> axes[0,0].bar([1,2,3],[3,4,5])    Plot vertical rectangles (constant width)
>>> axes[1,0].barh([0.5,1,2.5],[0,1,2])  Plot horiontal rectangles (constant height)
>>> axes[1,1].axhline(0.45)           Draw a horizontal line across axes
>>> axes[0,1].axvline(0.65)           Draw a vertical line across axes
>>> ax.fill(x,y,color='blue')         Draw filled polygons
>>> ax.fill_between(x,y,color='yellow')  Fill between y-values and o
```

Vector Fields

```
>>> axes[0,1].arrow(0,0,0.5,0.5)      Add an arrow to the axes
>>> axes[1,1].quiver(y,z)             Plot a 2D field of arrows
>>> axes[0,1].streamplot(X,Y,U,V)     Plot a 2D field of arrows
```

Data Distributions

```
>>> ax1.hist(y)                       Plot a histogram
>>> ax3.boxplot(y)                    Make a box and whisker plot
>>> ax3.violinplot(z)                 Make a violin plot
```

2D Data or Images

```
>>> fig, ax = plt.subplots()
>>> im = ax.imshow(img,               Colormapped or RGB arrays
                   cmap='gist_earth',
                   interpolation='nearest',
                   vmin=-2,
                   vmax=2)
```

```
>>> axes2[0].pcolor(data2)            Pseudocolor plot of 2D array
>>> axes2[0].pcolormesh(data)         Pseudocolor plot of 2D array
>>> CS = plt.contour(Y,X,U)           Plot contours
>>> axes2[2].contourf(data1)          Plot filled contours
>>> axes2[2]= ax.clabel(CS)           Label a contour plot
```

5 Save Plot

Save figures

```
>>> plt.savefig('foo.png')
```

Save transparent figures

```
>>> plt.savefig('foo.png', transparent=True)
```

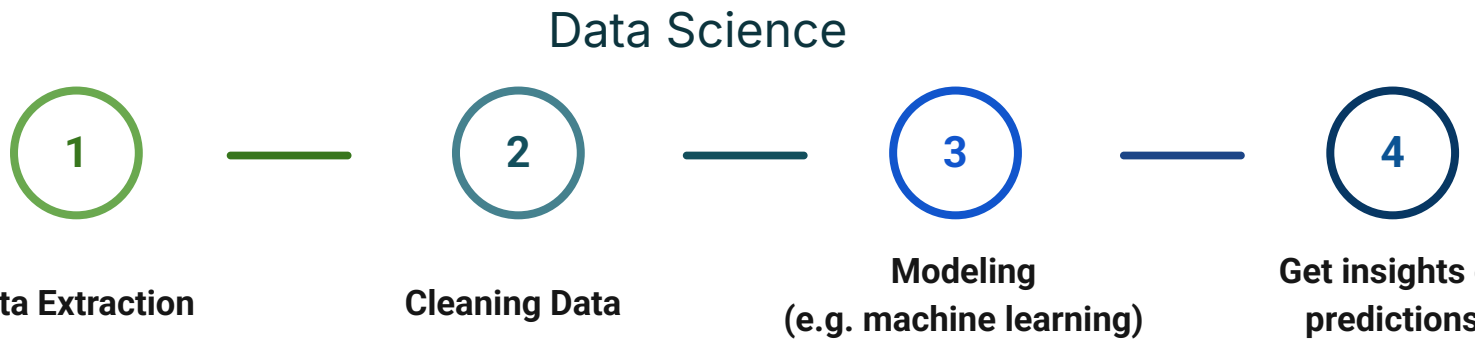
6 Show Plot

```
>>> plt.show()
```

Close & Clear

```
>>> plt.cla()                         Clear an axis
>>> plt.clf()                         Clear the entire figure
>>> plt.close()                       Close a window
```

Next Lecture: Introduction to Machine Learning



Data are usually already available!

Gustavo showed in his lecture how to extract astronomical data

We are going to focus on step 3

Contacts

Email: lilianne.nakazono@gmail.com

GitHub: https://github.com/marixko

Website: https://marixko.github.io