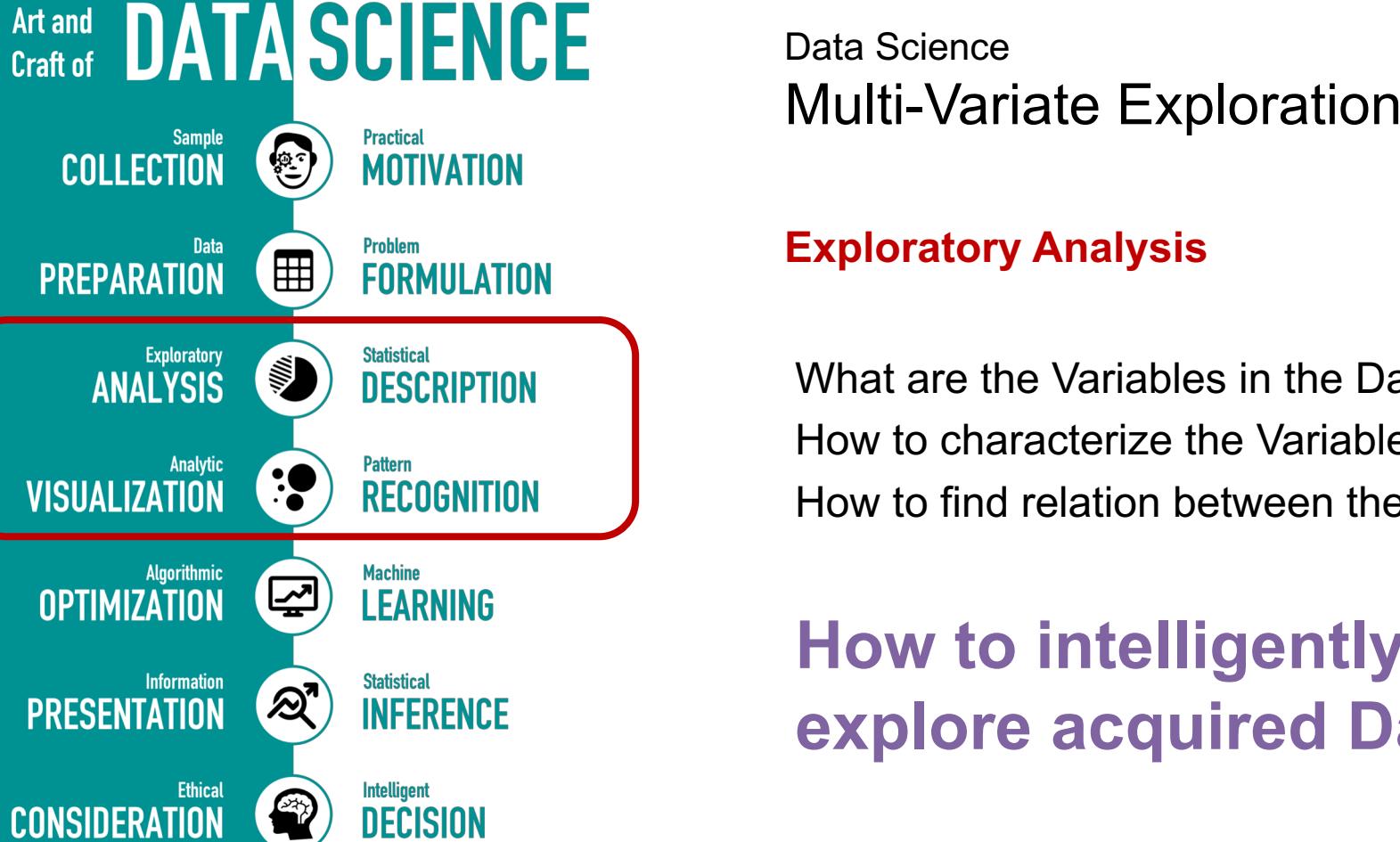


Multi-Variate Exploration

Sourav SEN GUPTA
Lecturer, SCSE, NTU







Data Science

The Pokemon Dataset

#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
430	Honchkrow	Dark	Flying	505	100	125	52	105	52	71	4	False
338	Solrock	Rock	Psychic	440	70	95	85	55	65	70	3	False
32	Nidoran♂	Poison	NaN	273	46	57	40	40	40	50	1	False
442	Spiritomb	Ghost	Dark	485	50	92	108	92	108	35	4	False
480	Uxie	Psychic	NaN	580	75	75	130	75	130	95	4	True
536	Palpitoad	Water	Ground	384	75	65	55	65	55	69	5	False
360	Wynaut	Psychic	NaN	260	95	23	48	23	48	23	3	False
478	Froslass	Ice	Ghost	480	70	80	70	80	70	110	4	False
76	Golem	Rock	Ground	495	80	120	130	55	65	45	1	False
177	Natu	Psychic	Flying	320	40	50	45	70	45	70	2	False

Source : Kaggle Datasets | **Pokemon with stats** by Alberto Barradas | <https://www.kaggle.com/abcsds/pokemon>

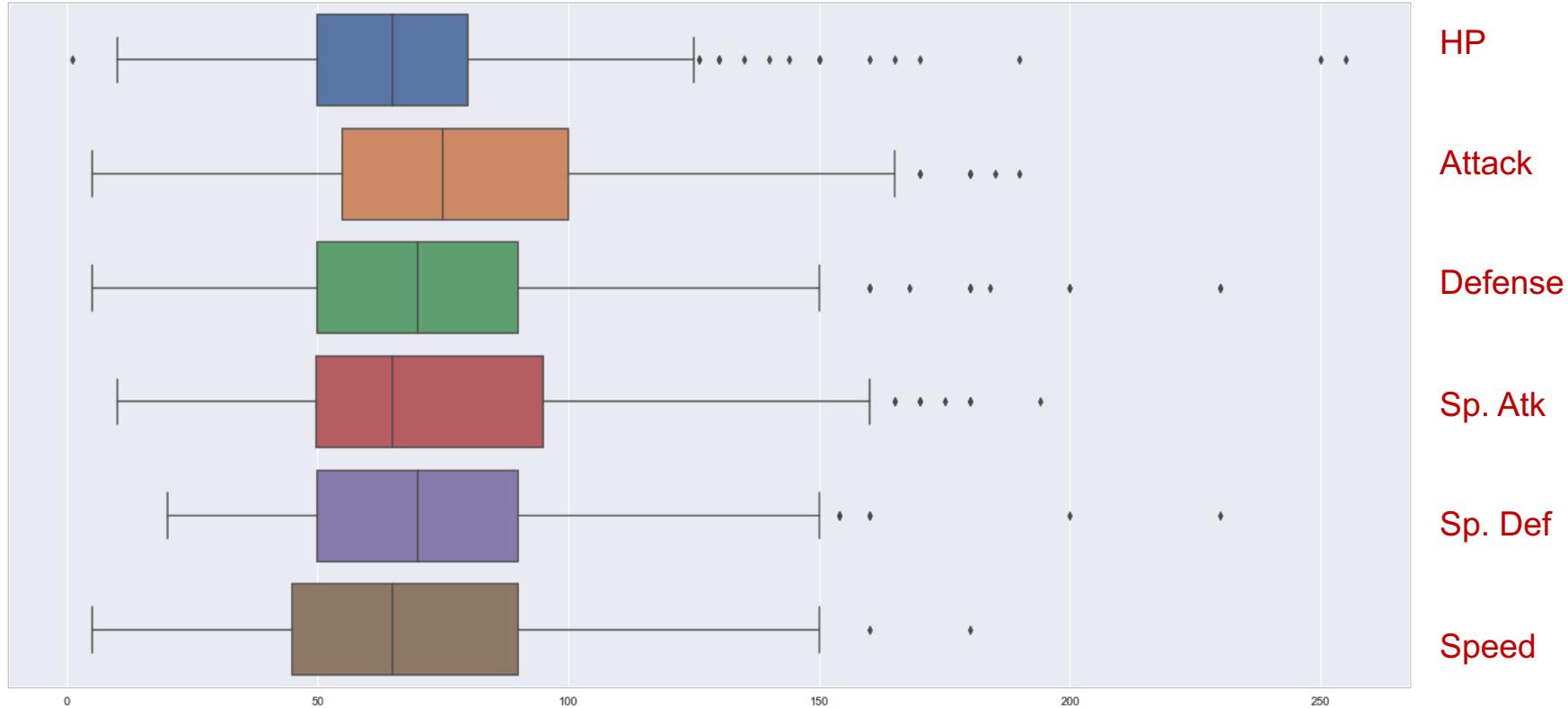
Individual Statistical Summary

Data Science Multi-Variate Statistics

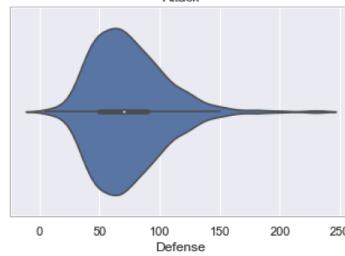
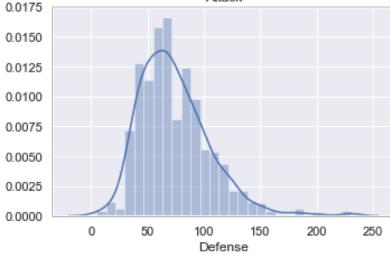
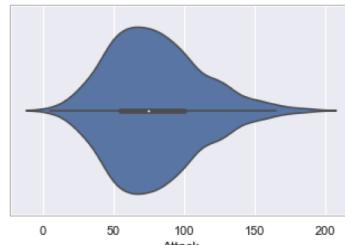
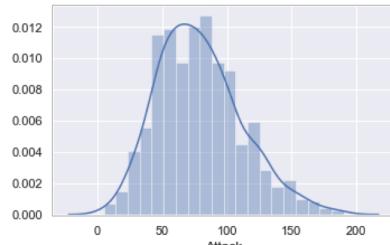
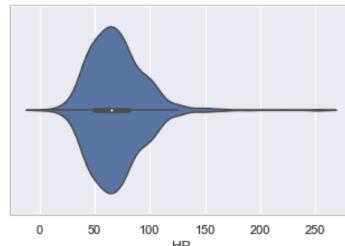
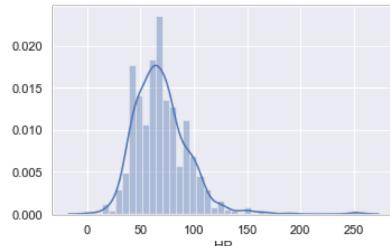
	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
count	800.000000	800.000000	800.000000	800.000000	800.000000	800.000000
mean	69.258750	79.001250	73.842500	72.820000	71.902500	68.277500
std	25.534669	32.457366	31.183501	32.722294	27.828916	29.060474
min	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000
25%	50.000000	55.000000	50.000000	49.750000	50.000000	45.000000
50%	65.000000	75.000000	70.000000	65.000000	70.000000	65.000000
75%	80.000000	100.000000	90.000000	95.000000	90.000000	90.000000
max	255.000000	190.000000	230.000000	194.000000	230.000000	180.000000

Individual Box-Plots

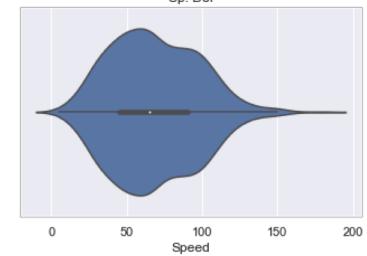
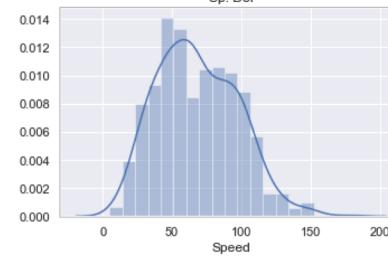
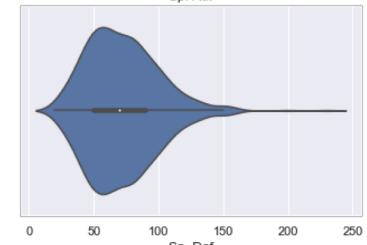
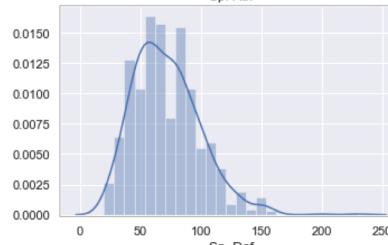
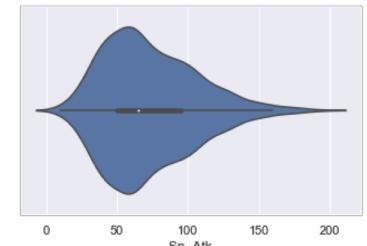
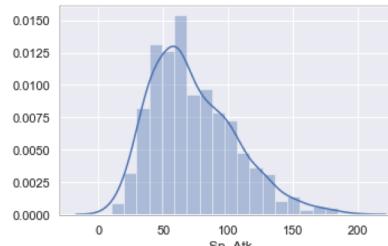
Data Science Multi-Variate Statistics



Individual Histograms and Violin Plots



Data Science Multi-Variate Statistics



Mutual Correlations

Data Science Multi-Variate Statistics

	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
HP	1.000000	0.422386	0.239622	0.362380	0.378718	0.175952
Attack	0.422386	1.000000	0.438687	0.396362	0.263990	0.381240
Defense	0.239622	0.438687	1.000000	0.223549	0.510747	0.015227
Sp. Atk	0.362380	0.396362	0.223549	1.000000	0.506121	0.473018
Sp. Def	0.378718	0.263990	0.510747	0.506121	1.000000	0.259133
Speed	0.175952	0.381240	0.015227	0.473018	0.259133	1.000000



Mutual Correlations

No Dependence

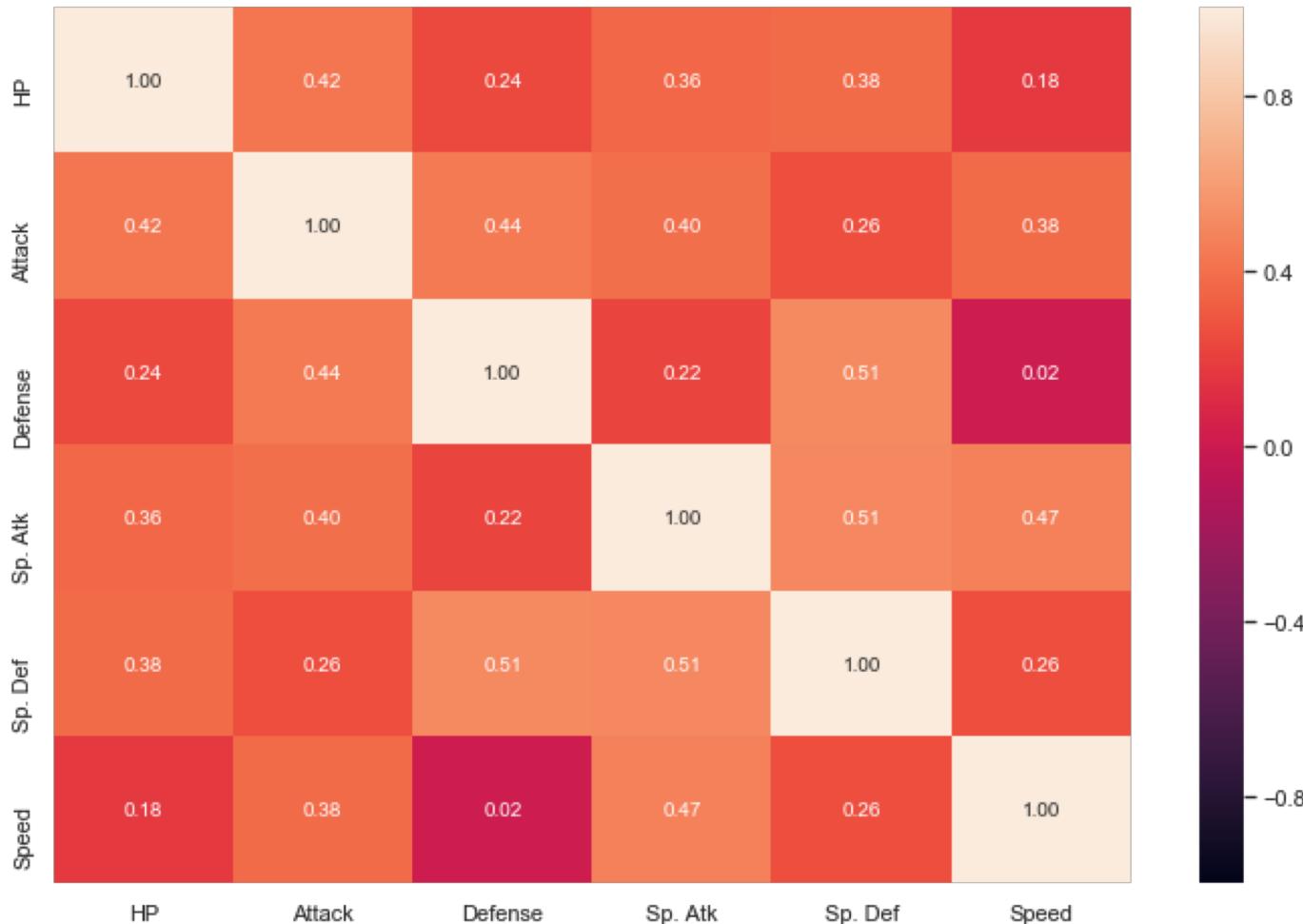
Corr = 0

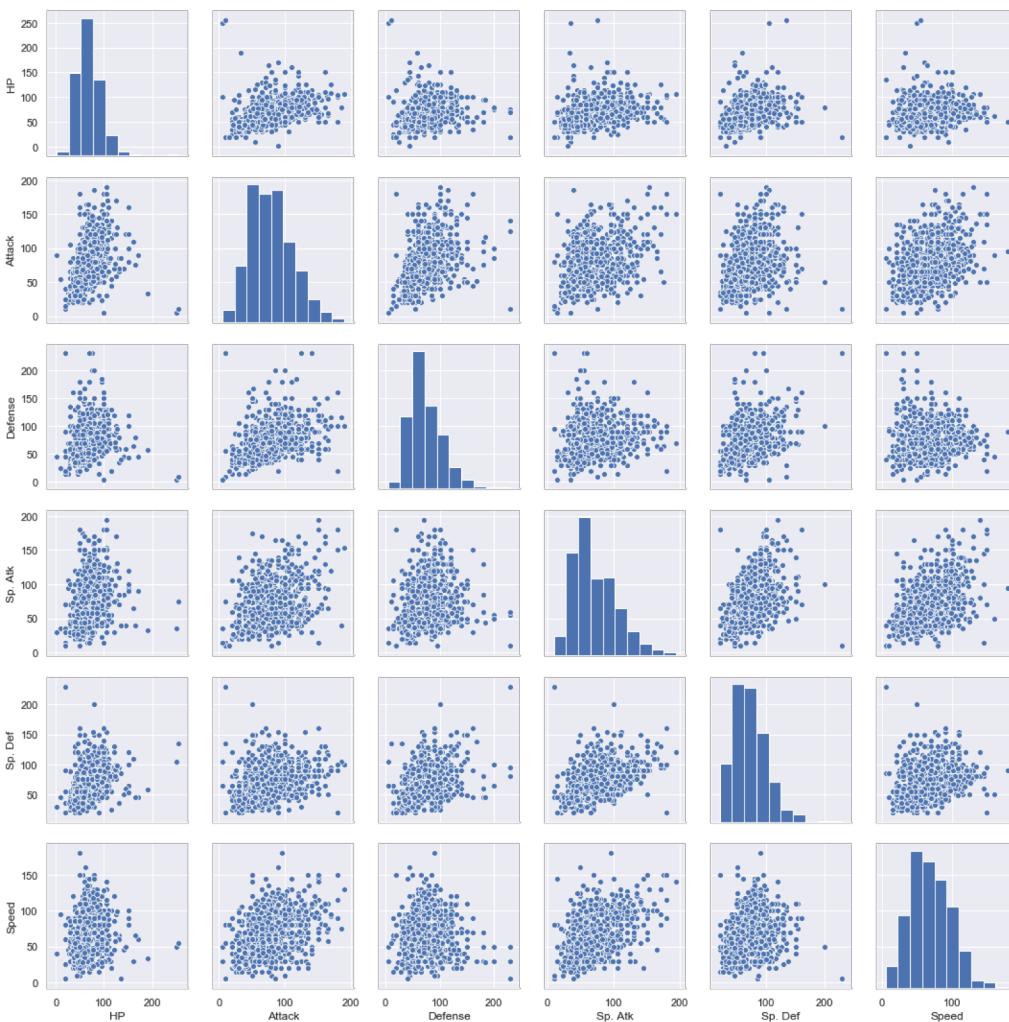
Perfect Positive

Corr = + 1

Perfect Negative

Corr = - 1





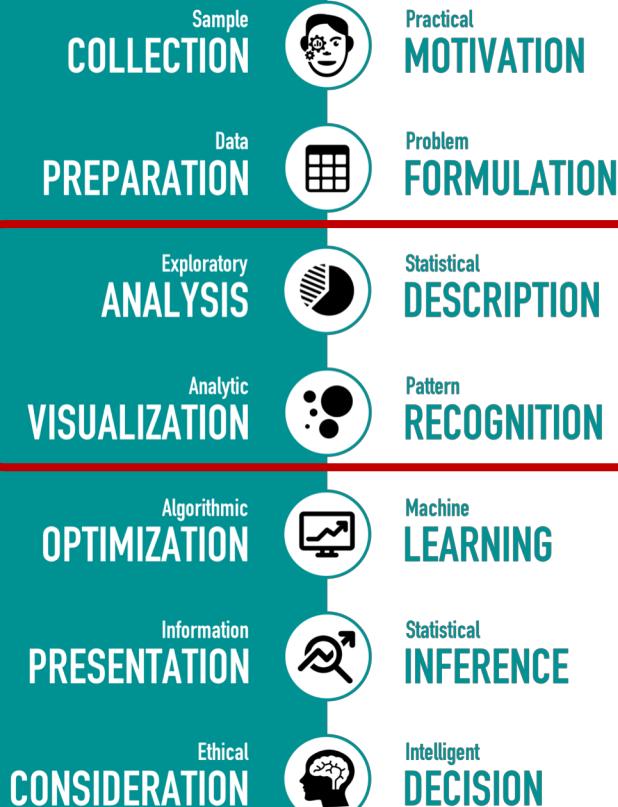
Data Science Multi-Variate Dist.

Pair-Plot of Multi-Variate Data

Histograms in the Diagonals
Scatter-Plot in Off-Diagonals

Pattern recognition

- Distributions of Variables
- Inter-Variable Dependence



Data Science Pipeline Exploratory Analysis

How to summarize the acquired Data?
How to visualize the acquired Data?
How to analyze the acquired Data?

How to intelligently
explore acquired Data?