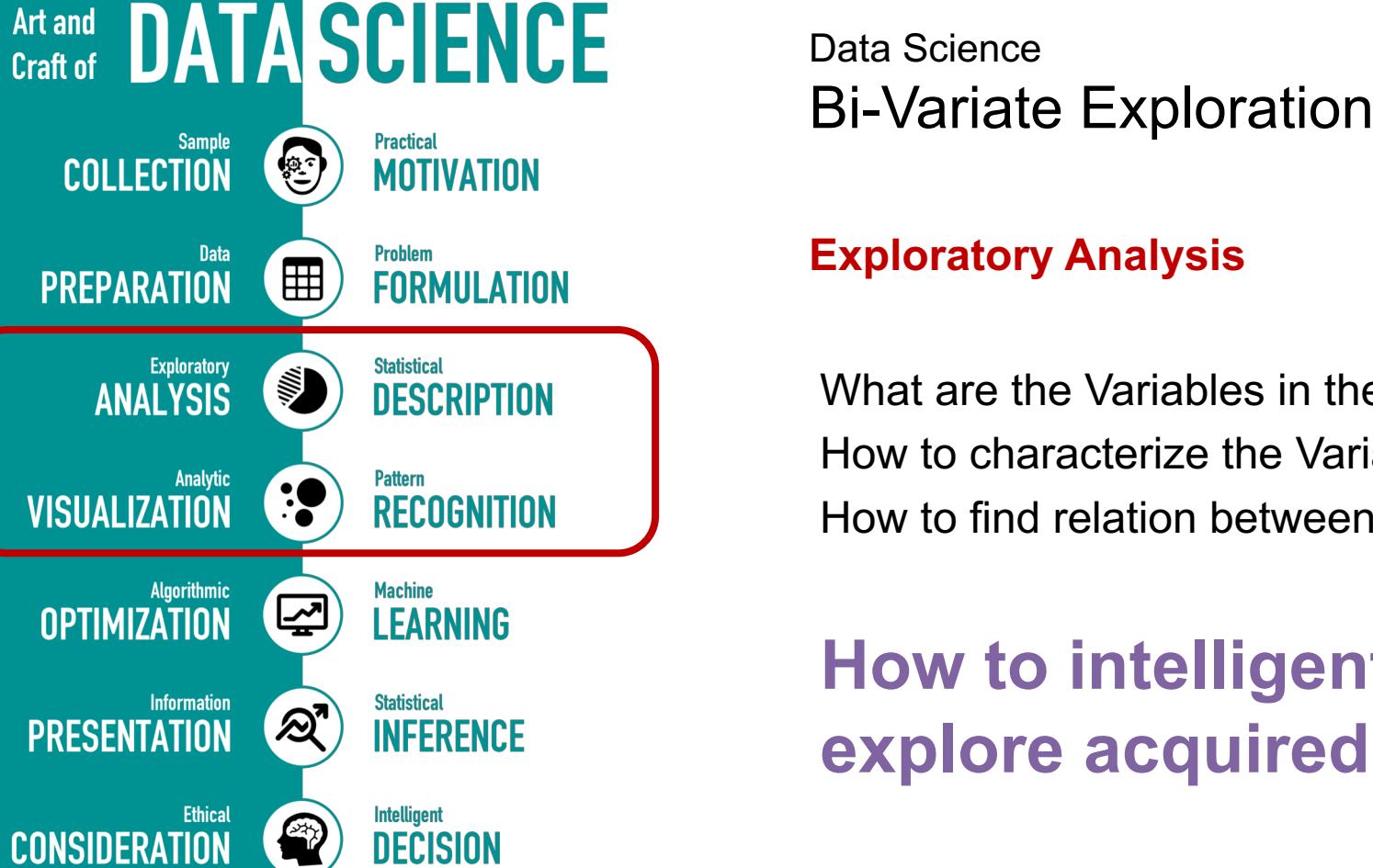


Bi-Variate Exploration

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>

Data Science

Bi-Variate Statistics

Statistical Summary

HP		Attack	
count	800.000000	count	800.000000
mean	69.258750	mean	79.001250
std	25.534669	std	32.457366
min	1.000000	min	5.000000
25%	50.000000	25%	55.000000
50%	65.000000	50%	75.000000
75%	80.000000	75%	100.000000
max	255.000000	max	190.000000

HP Hit Points of a Pokemon
Attack Base Modifier for Attack

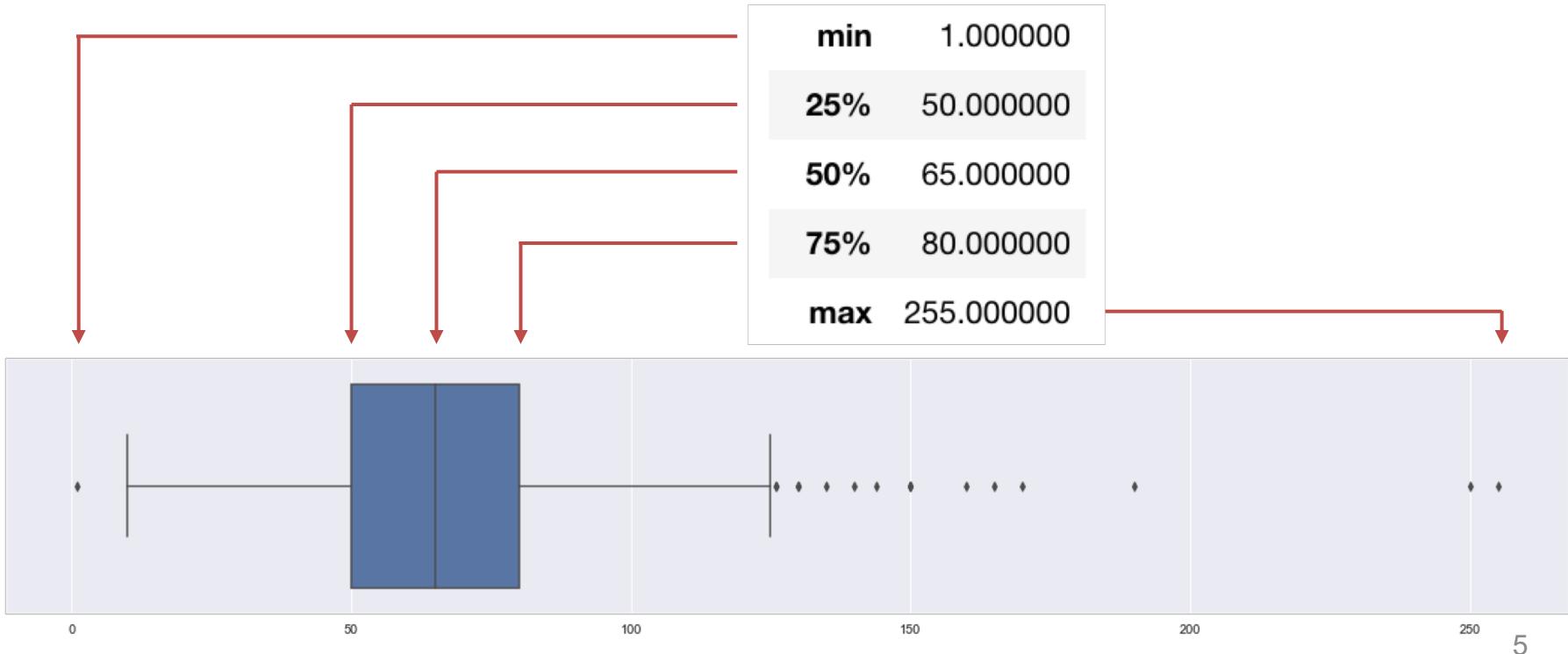
Statistical Questions

- What is the Central Tendency?
- What is the Spread of the Data?

HP

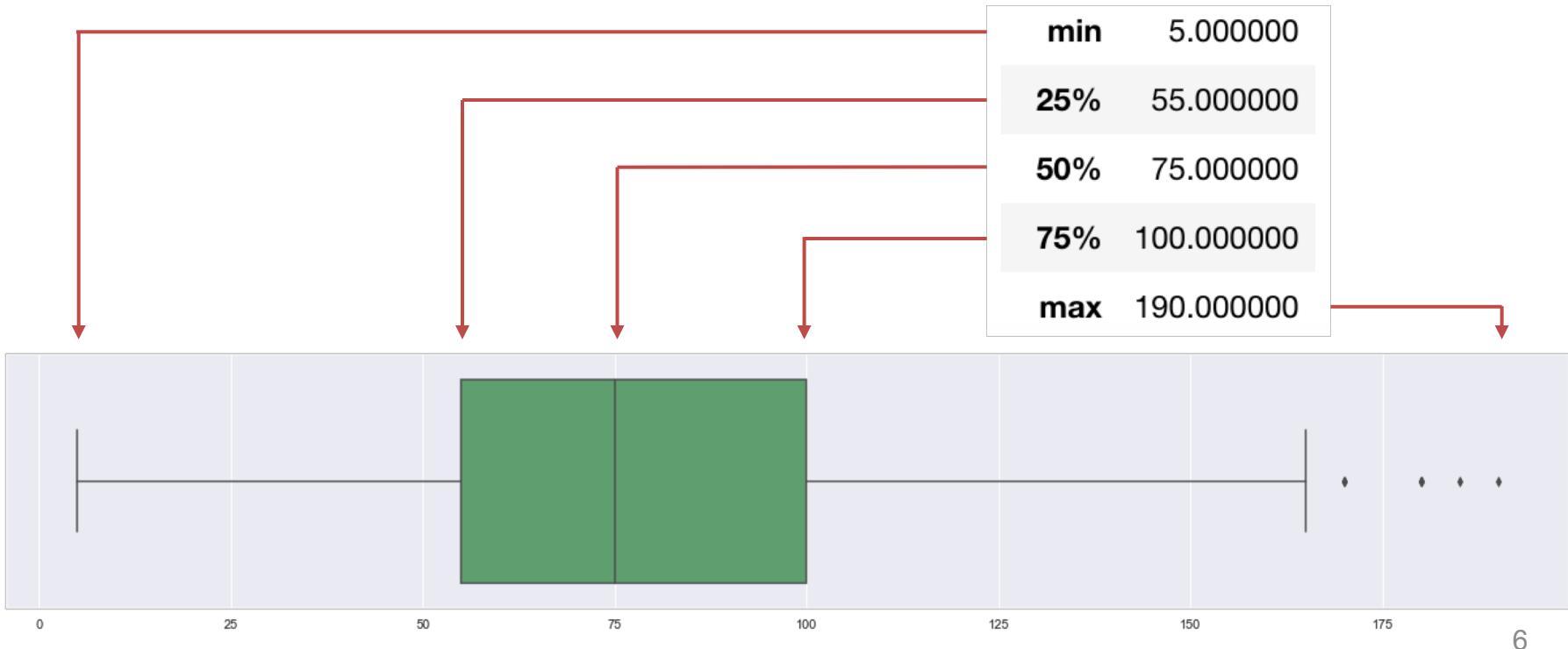
Data Science

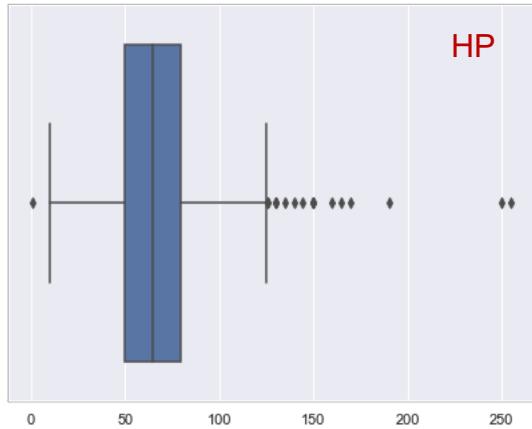
Uni-Variate Box-Plot



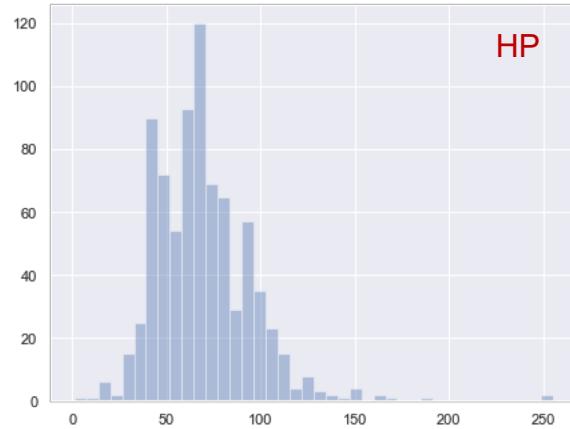
Attack

Data Science Uni-Variate Box-Plot

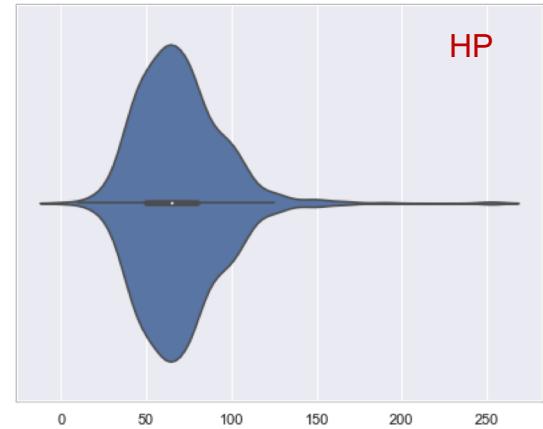




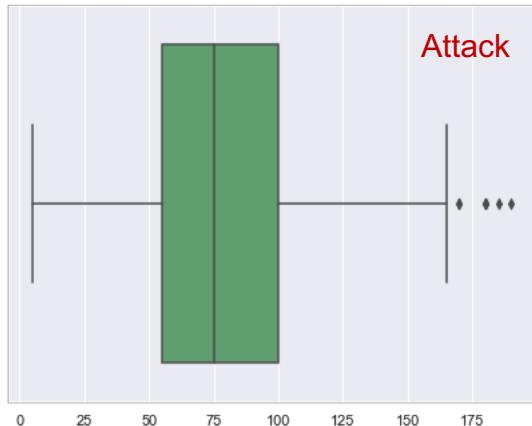
HP



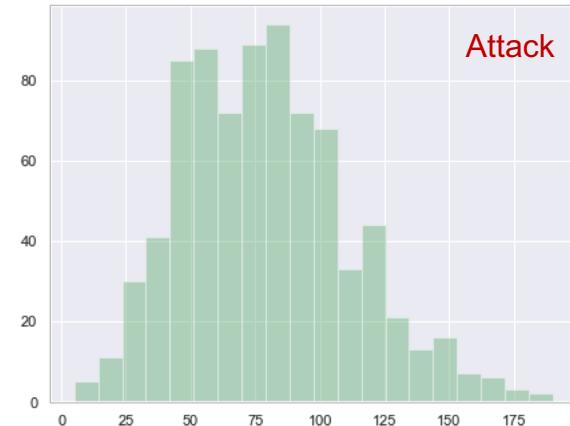
HP



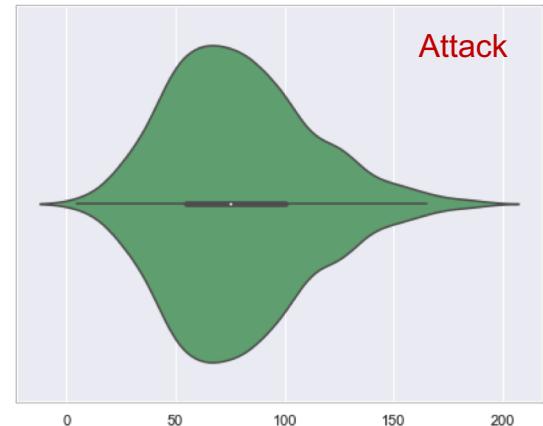
HP



Attack

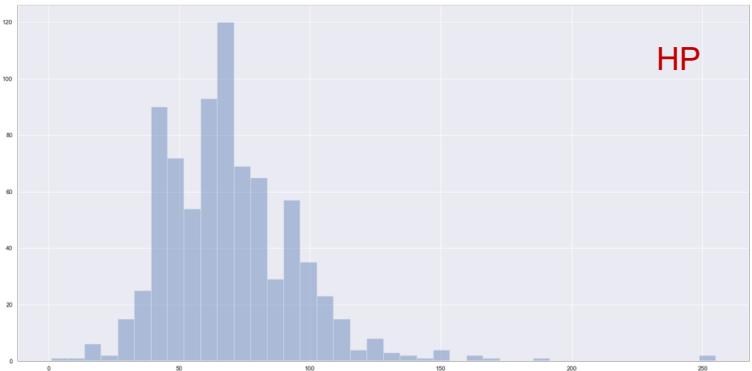


Attack



Attack



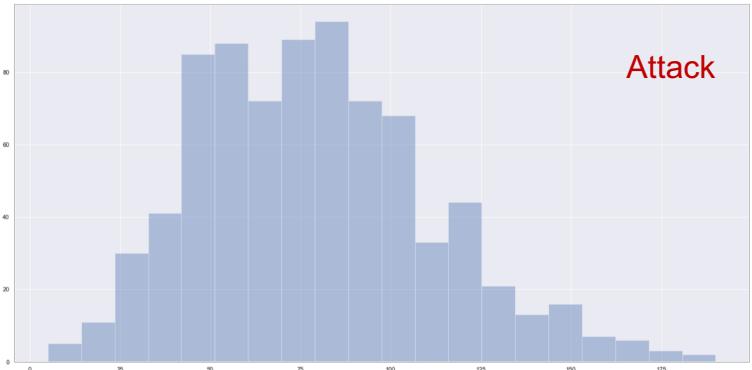


Data Science

Bi-Variate Statistics

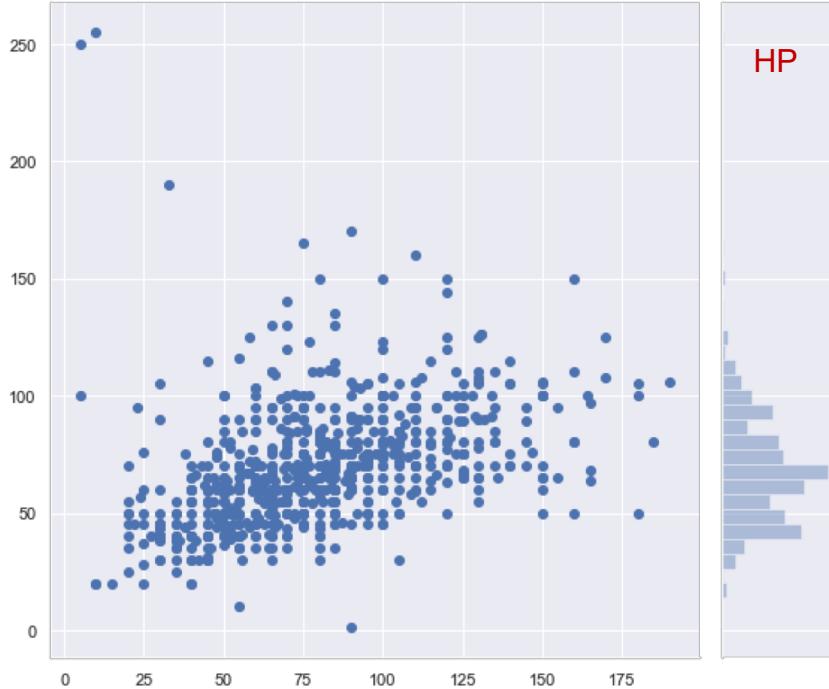
Statistical Relation

HP Hit Points of a Pokemon
Attack Base Modifier for Attack



Statistical Questions

- Is there a Mutual Dependence?
- What is the Mutual Relationship?



Data Science

Bi-Variate Joint Plot

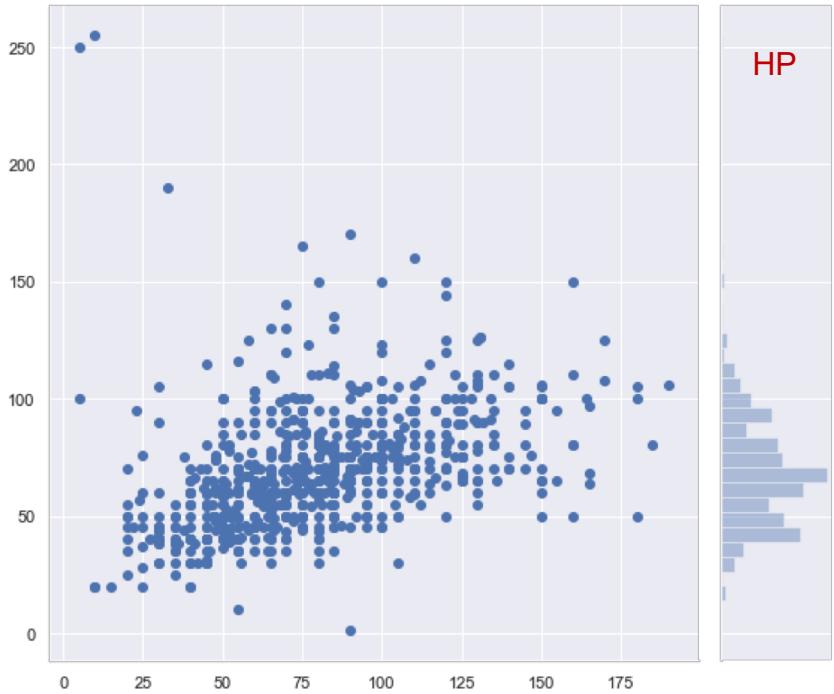
Statistical Relation

HP Plotted along Y axis

Attack Plotted along X axis

Pattern Recognition

- Is there a Mutual Dependence?
- What is the Mutual Relationship?



Data Science

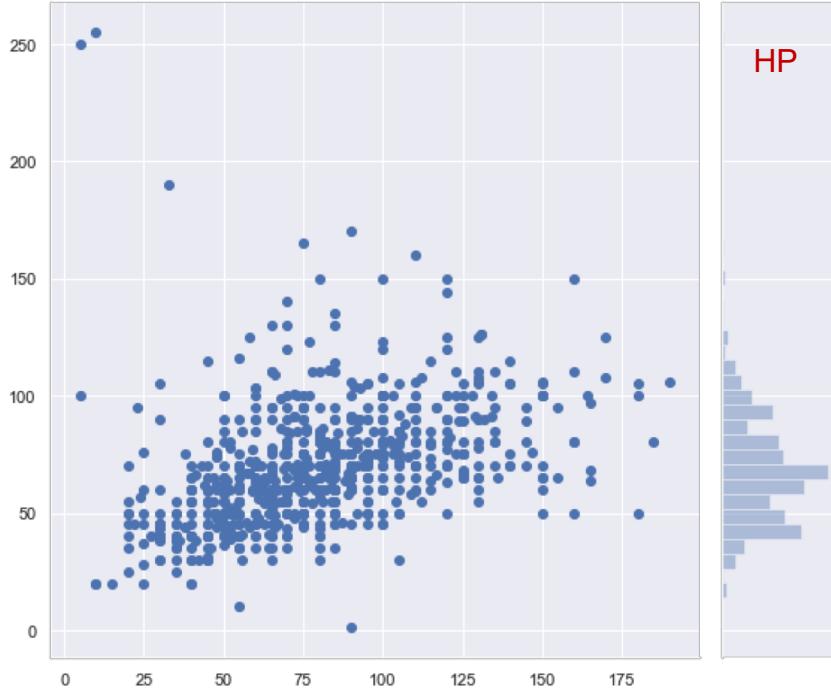
Bi-Variate Joint Plot

Statistical Relation

HP Plotted along Y axis
Attack Plotted along X axis

Pattern Recognition

- HP increases as Attack increases
- Dependence is moderately strong



Data Science Bi-Variate Relation

Correlation Coefficient

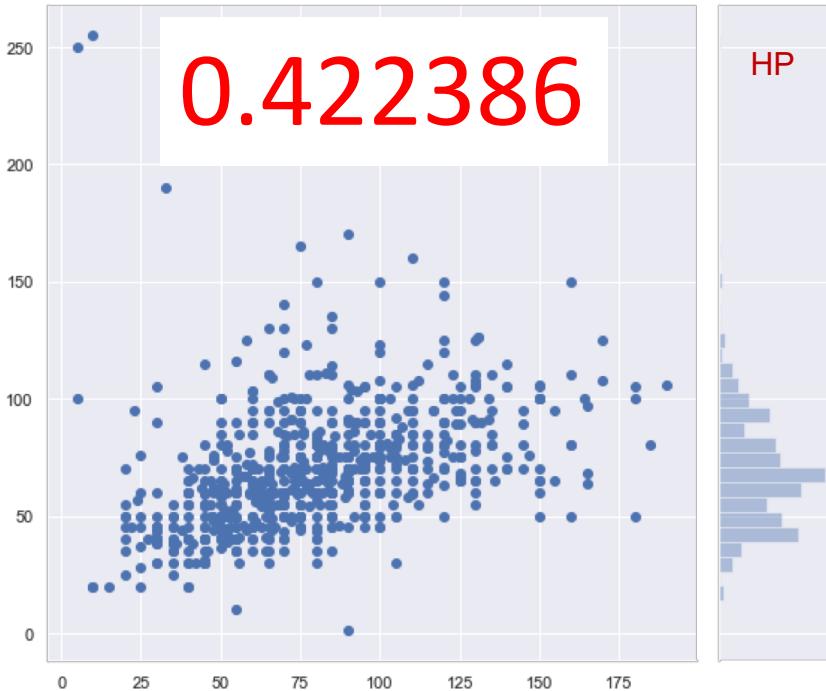
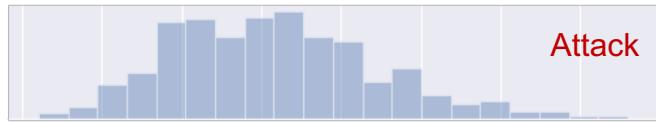
Natural Intuition

Dependence of HP and Attack

Statistical Formula

Co-Variance / St. Dev Product

$$\rho_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$



Data Science Bi-Variate Relation

Correlation Coefficient

Natural Intuition

Dependence of HP and Attack

Statistical Intuition

No Dependence

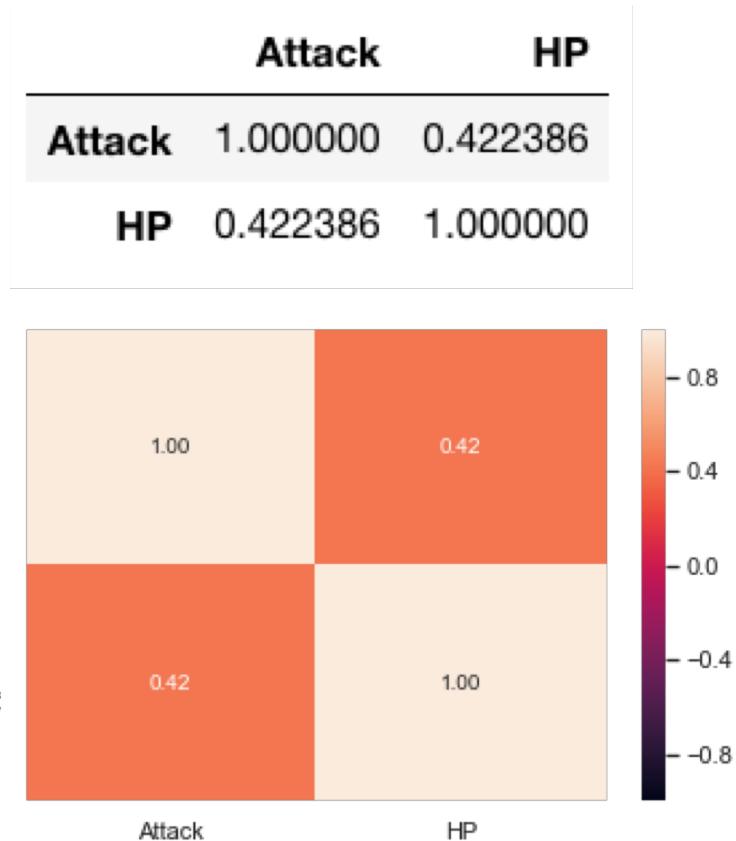
Corr = 0

Perfect Positive

Corr = + 1

Perfect Negative

Corr = - 1



Data Science Bi-Variate Relation

Correlation Matrix and Plot

Natural Intuition

Dependence of HP and Attack

Statistical Intuition

No Dependence

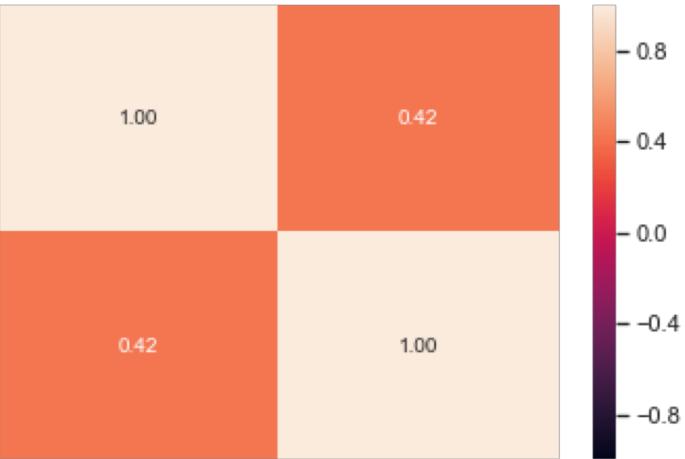
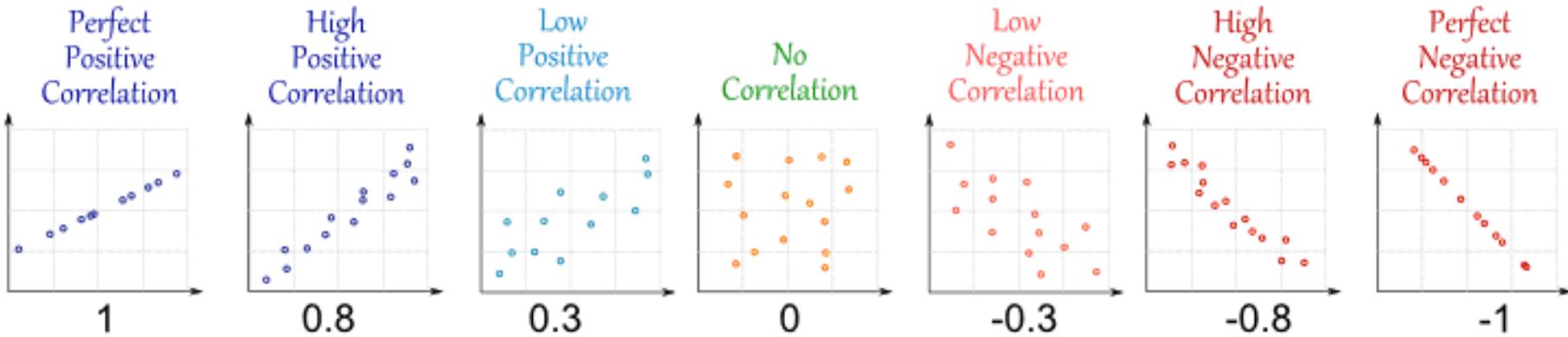
Corr = 0

Perfect Positive

Corr = + 1

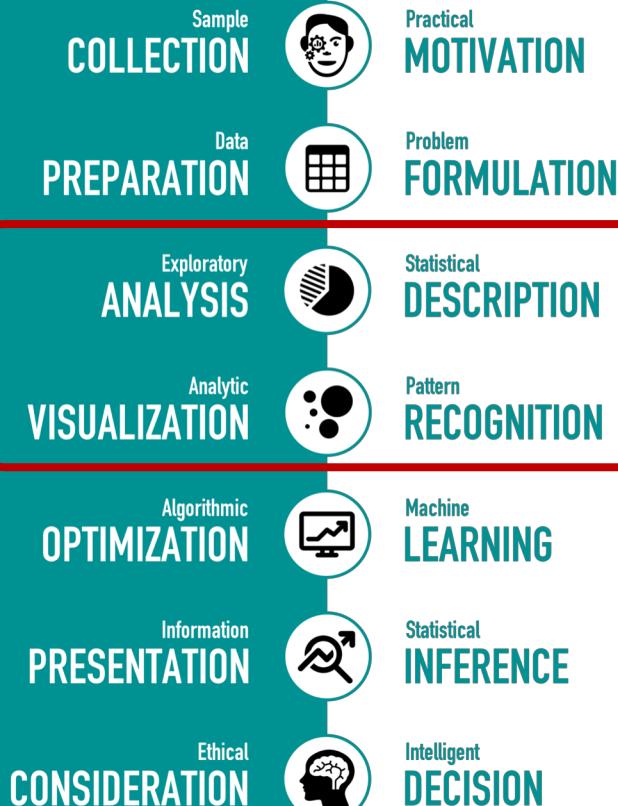
Perfect Negative

Corr = - 1



Play this to improve your
Correlation spotting skills

Guess the Correlation
<http://guessthecorrelation.com/>



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?