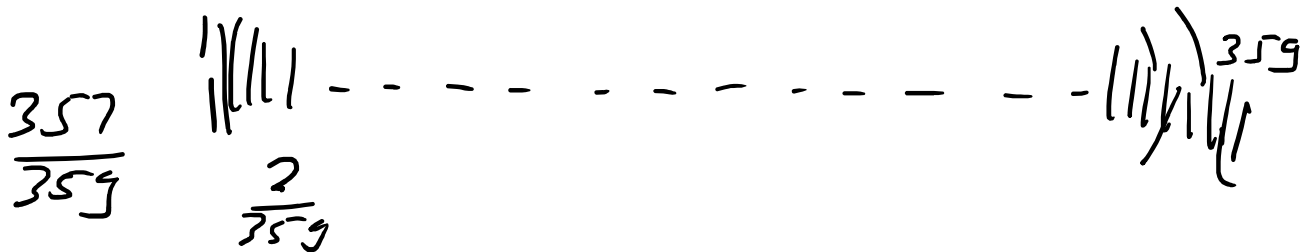


→ Visualizations: Numerical:

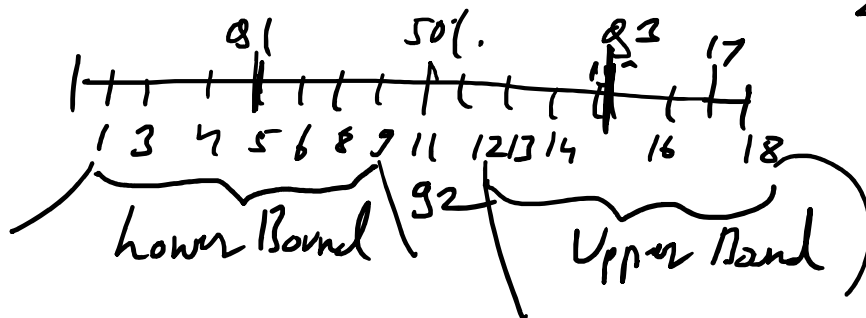
L> Box Plot & Quantile (Percentiles)

L> Histogram

→ Percentile: Segregation: (Divisible Portion)



→ Quantiles: 25% 50% 75% 100% Q1 Q2 Q3 Q4



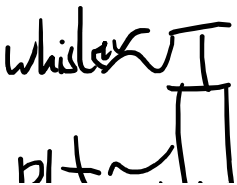
$$\frac{1+9}{2} = 5$$

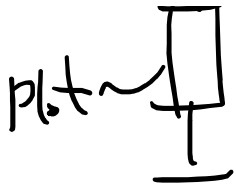
581

$$\frac{12+18}{2} = 15$$

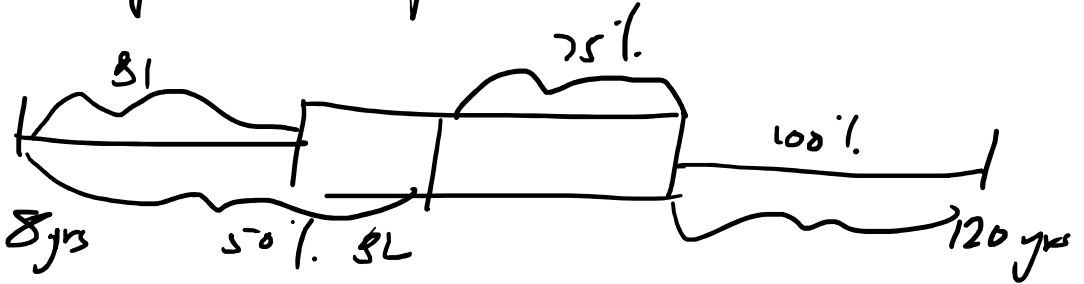
$$\frac{UP - LB}{2} = \frac{15 - 5}{2} = 5$$

→ Box Plot: Marine biologist performs a research on lifespan of blue whales. He is able to record some values. He has given a box plot





able to recover some values. He has given a boxplot



$$\frac{120-8}{2} = \frac{112}{2} = 56$$



→ Flow

→ Dispersion

→ Gaussian Plot

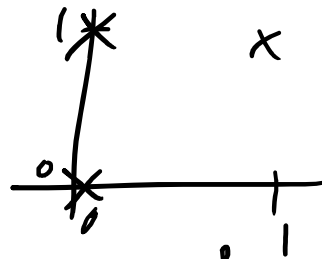
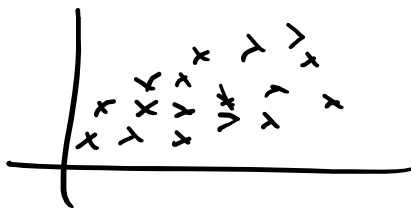
→ Categorical Variables : (Non-Numerical)

Occurrence Oriented

Relationships

A	Yes	No	1	0	only
B	No	No	0	0	
C	Yes	No	1	1	
			1	1	

→ Mode → Scatterplot



→ Variance :

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$



→ Covariance:

$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N-1}$$

→ $x(2, 4, 6, 8, 10)$ & $y(3, 7, 10, 14, 17)$

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
2	3	-4	-7.2	
4	7	-2	-3.2	
6	10	0	-0.2	
8	14	2	3.8	
10	17	4	6.8	

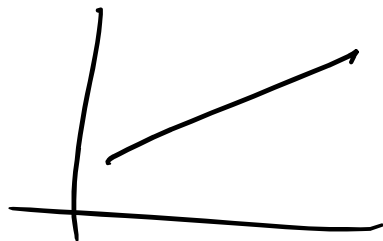
$$\bar{x} = \underline{6}$$

$$\frac{51}{5} = 10.2 = \bar{y}$$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \underline{70}$$

>

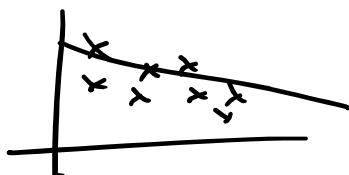
$$\text{cov}(x, y) = \frac{70}{4} = \underline{17.5}$$



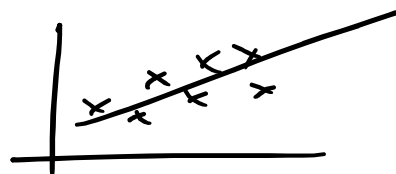
→ Name / Income / DOB / Age

→ Correlation Coefficient:

ρ



$$0 < \rho < 1$$



$$-1 > \rho > 0$$

$$r = \frac{n \sum xy - \sum x \cdot \sum y}{n \sqrt{\sum x^2 \sum y^2}}$$

$$r = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

x : 1, 2, 3, 4, 5, 6

y : 2, 4, 7, 9, 12, 14

x	y	xy	x ²	y ²
2	4	211	91	490

$$r = \frac{6 \times 211 - 21 \times 48}{\sqrt{[6 \times (91) - (21)^2] \cdot [6 \times (490) - (48)^2]}}$$

$$r = \underline{\underline{0.998}} \approx \underline{\underline{99.8\%}}$$