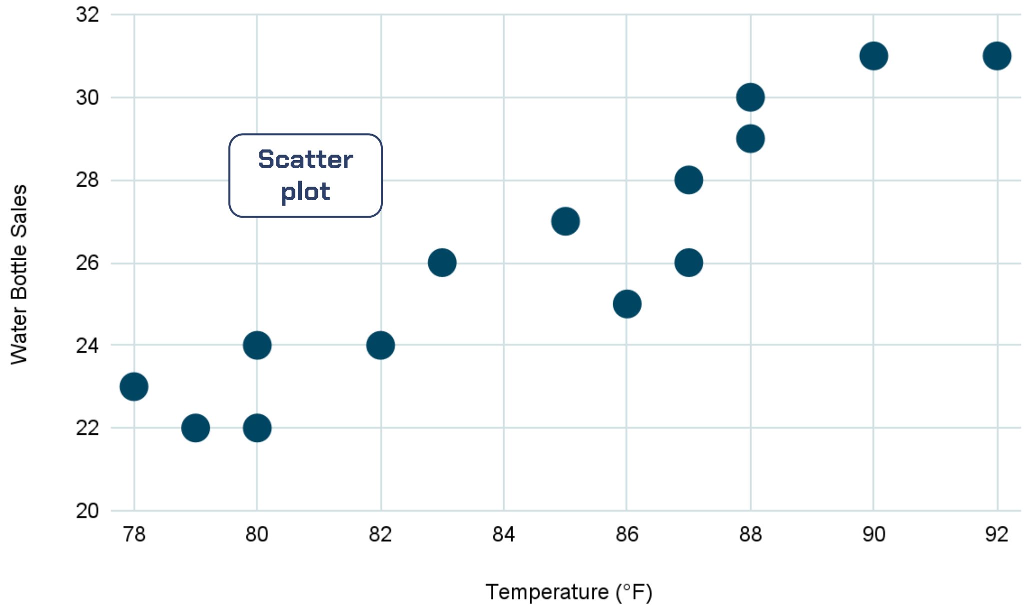
Multivariate data

Temperature (°F)	Water Bottle Sales
78	23
79	22
80	24
80	22
82	24
83	26
85	27
86	25
87	28
87	26
88	29
88	30
90	31
92	31



Methods of association

Covariance

descriptive measure of the linear association between two variables

Sample $$Cov(x,y) = \frac{\sum(x_i - \bar{x}_x)(y_i - \bar{y}_y)}{n-1}$$

Population $$Cov(X,Y) = \frac{\sum(X_i - \mu_x)(Y_i - \mu_y)}{N}$$

Employees	**Sales**
3	5
4	10
10	15
15	22
Mean (32/4) = **8**	Mean (52/4) = **13**

Cov(x,y) = (3-8)(5-13) + (4-8)(10-13) + (10-8)(15-13) + (15-8)(22-13) **/ (4-1)**
= 38.66

Review	**Inventory**
3	8
2	15
5	1
4	4
Mean (14/4) = **3.5**	Mean (28/4) = **7**

Cov(x,y) = (3-3.5)(8-7) + (2-3.5)(15-7) + (5-3.5)(1-7) + (4-3.5)(4-7) **/ (4-1)**
= -7.667

Correlation

Correlation coefficient measures the relationship between two variables

Varies b/w **-1** and **+1**

$$r_{xy} = \frac{Cov_{xy}}{S_x S_y}$$

Employees	Sales
3	5
4	10
10	15
15	22
Mean (24/4) = **6**	Mean (52/4) = **13**
Stdev = **5.59**	Stdev = **7.25**

Corr(x,y) = 38.66/ (5.59 * 7.25)
= 0.97

Strong **positive** correlation

Strong **negative** correlation

Covrr(x,y) = -7.667 / (1.29*6.05)
= -0.98

Review	Inventory
3	8
2	15
5	1
4	4
Mean (14/4) = **3.5**	Mean (28/4) = **7**
Stdev = **1.29**	Stdev = **6.05**

Regression

developing an equation for the relationship between variables

Temperature (°F)	Water Bottle Sales
78	23
79	22
80	24
80	22
82	24
83	26
85	27
86	25
87	28
87	26
88	29
88	30
90	31
92	31

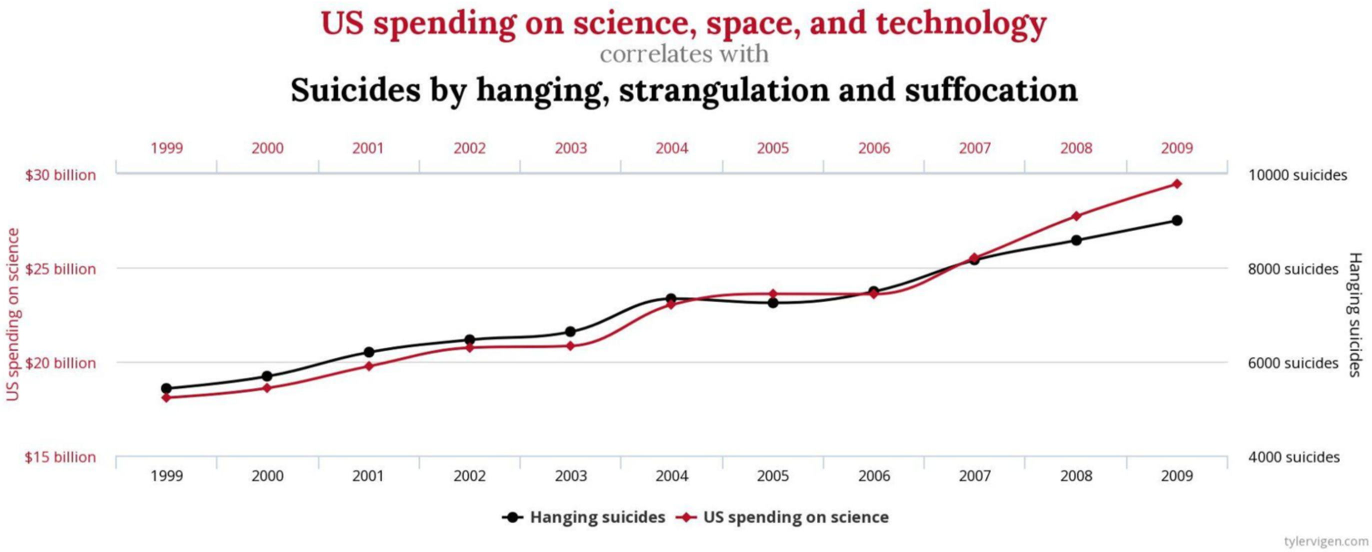
Line - **linear regression**

Polynomial function - **polynomial regression**

More than 2 variables - **multi-regression**

Causation

Correlation ≠ Causation



https://tylervigen.com/spurious-correlations