Correlation

Correlation

- Correlation coefficient is an equation that is used to determine the strength of relation between two variables.
- Correlation coefficient sometimes called as cross correlation coefficient. Correlation coefficient always lies between -1 to +1
- where -1 represents X and Y are negatively correlated and
- +1 represents X and Y are positively correlated.
- 0 means there is no correlation

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

Correlation

- Each data point in the dataset is an observation, and the features are the properties or attributes of those observations.

- Every dataset you work with uses variables and observations.
- For example, you might be interested in understanding the following:

  - How the height of basketball players is correlated to their shooting accuracy
  - Whether there's a relationship between employee work experience and salary
  - In the examples above, the height, shooting accuracy, years of experience, salary, population density, and gross domestic product are the features or variables. The data related to each player, employee, and each country are the observations.

Correlation

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

	x	y		x*y	x*x	y*y
	15	25		375	225	625
	18	25		450	324	625
	21	27		567	441	729
	24	31		744	576	961
	27	32		864	729	1024
Sum	105	140		3000	2295	3964
Sum	x	y		x*y	x*x	y*y

Corrélation coefficient
= (5 * 3000 - 105 * 140)
/ sqrt((5 * 2295 - 1052)*(5*3964 - 1402))
= 300 / sqrt(450 * 220) = 0.953463

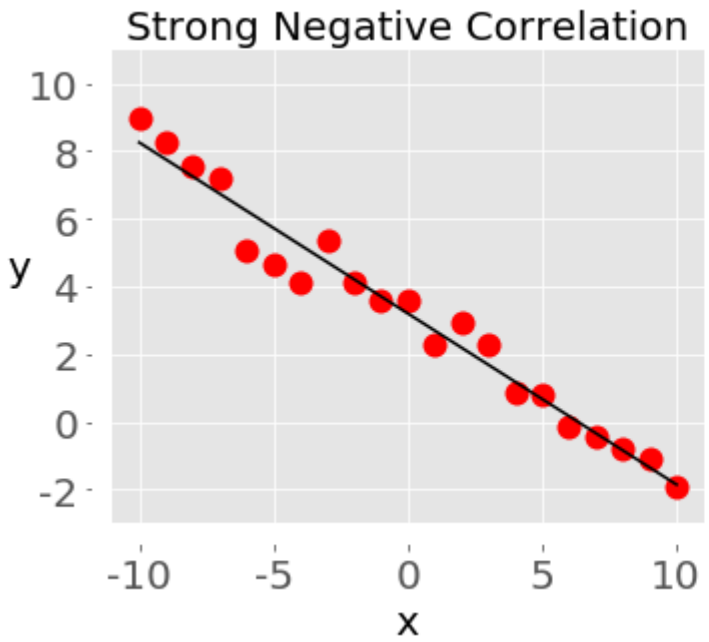
Corrélation coefficient = 0.953463

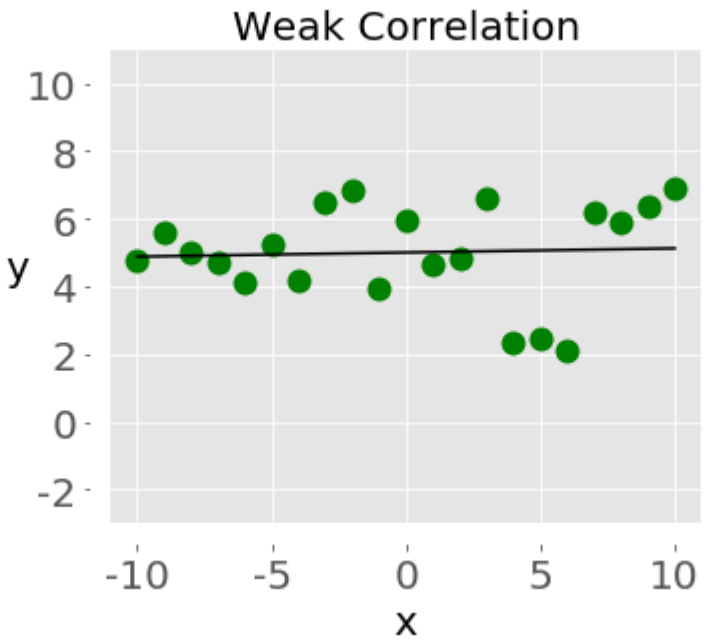
Correlation

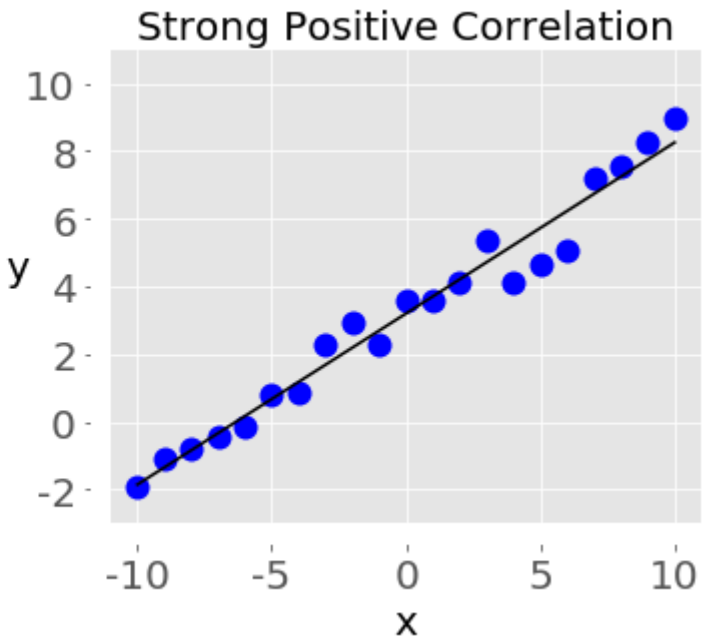
Experience	Salary
1	10000
2	9000
3	8000
4	7000
5	6000
6	5000
7	4000
8	3000
9	2000
10	1000

Experience	Salary
1	1000
2	1200
3	1200
4	1000
5	1000
6	1000
7	1200
8	1000
9	1200
10	1000

Experience	Salary
1	1000
2	2000
3	3000
4	4000
5	5000
6	6000
7	7000
8	8000
9	9000
10	10000







Correlation does not imply causation
Correlation is not causation

End