

Correlação

Prof. Altair R. Gomes Jr

A large, dark blue, abstract shape that starts as a thin line on the left and expands into a wide, solid area at the bottom right, creating a modern, minimalist design element.

Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson, também chamado de "coeficiente de correlação produto-momento" ou simplesmente de "ρ de Pearson" mede o grau da correlação entre duas variáveis aleatórias.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

- 0.9 para mais ou para menos indica uma correlação muito forte.
- 0.7 a 0.9 positivo ou negativo indica uma correlação forte.
- 0.5 a 0.7 positivo ou negativo indica uma correlação moderada.
- 0.3 a 0.5 positivo ou negativo indica uma correlação fraca.
- 0 a 0.3 positivo ou negativo indica uma correlação desprezível.

Interpretação geométrica

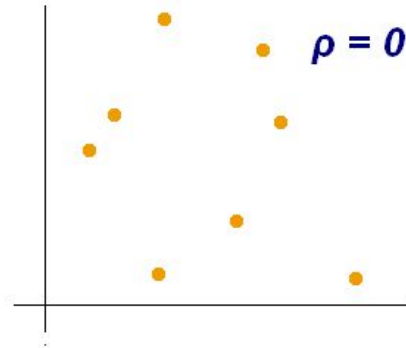
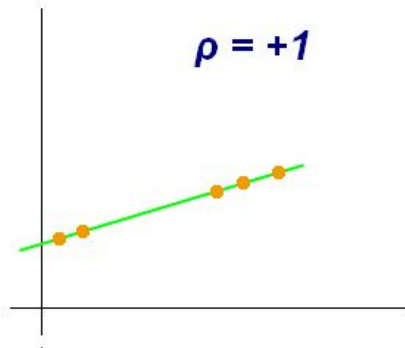
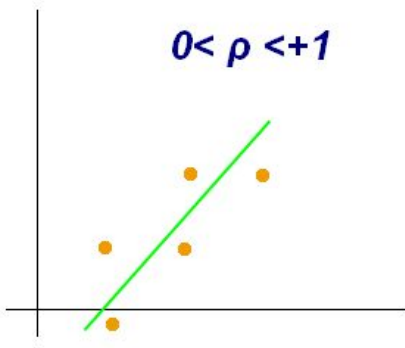
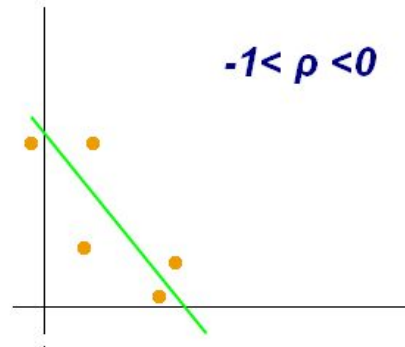
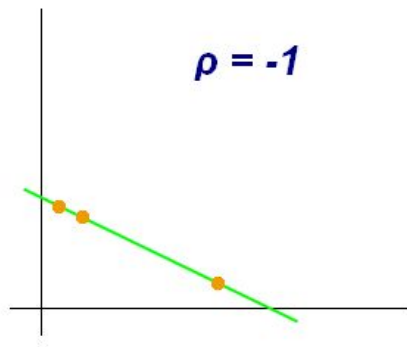
Supondo que os dois conjuntos de dados são vetores no espaço n-dimensional:

$X[x_1 - \bar{x}_M, x_2 - \bar{x}_M, \dots, x_n - \bar{x}_M]$ e $Y[y_1 - \bar{y}_M, y_2 - \bar{y}_M, \dots, y_n - \bar{y}_M]$

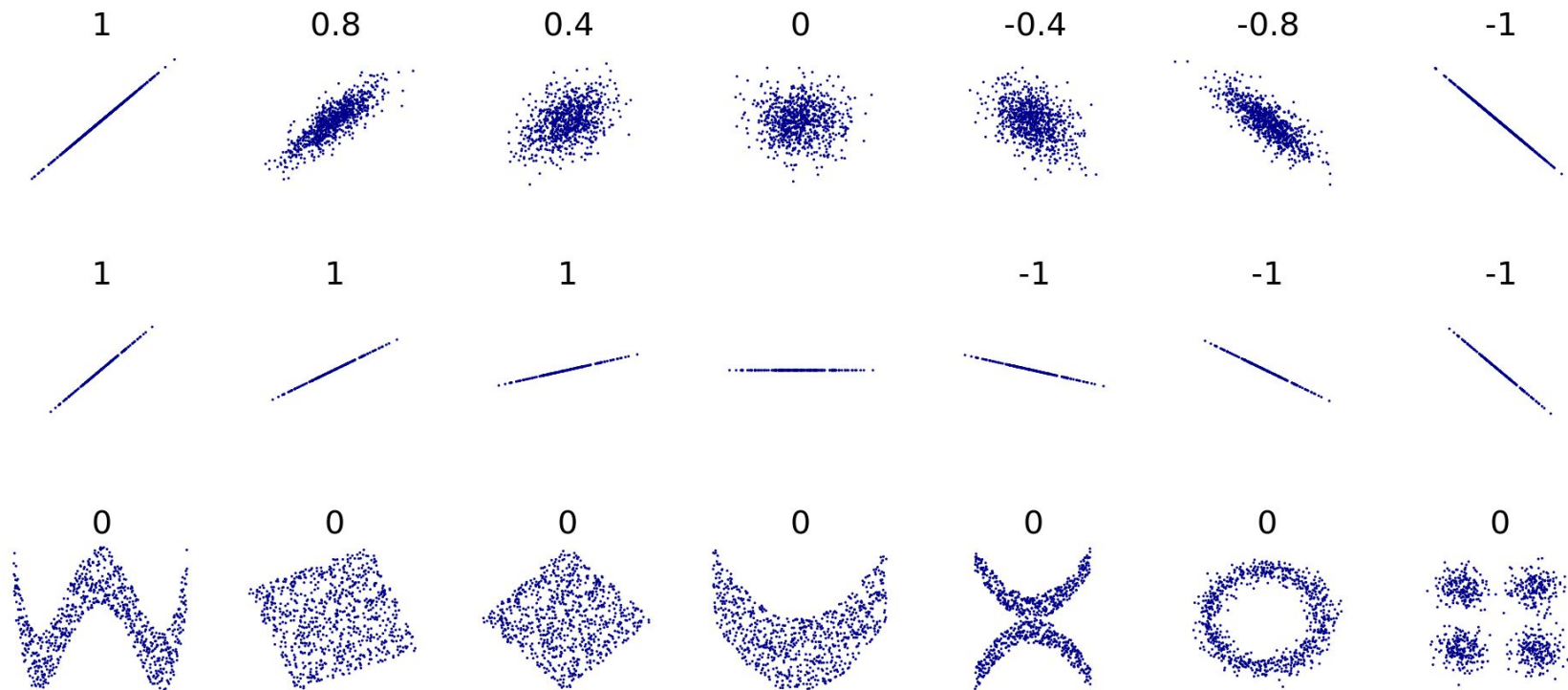
O ângulo entre os dois vetores será dado por:

$$\rho = \cos \theta = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Correlação



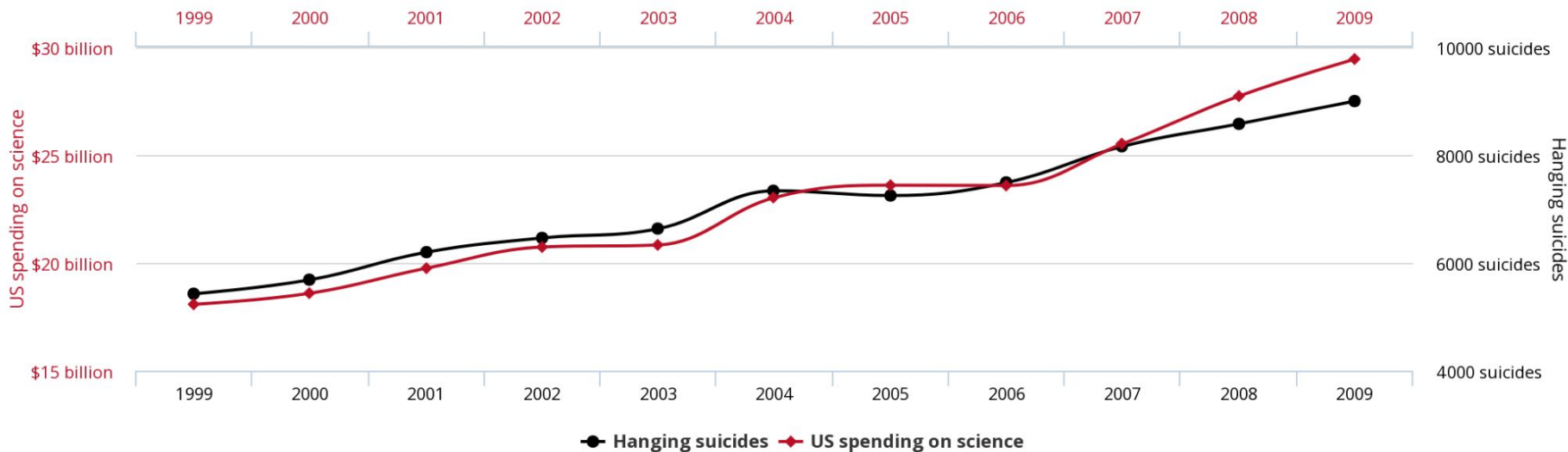
Correlação



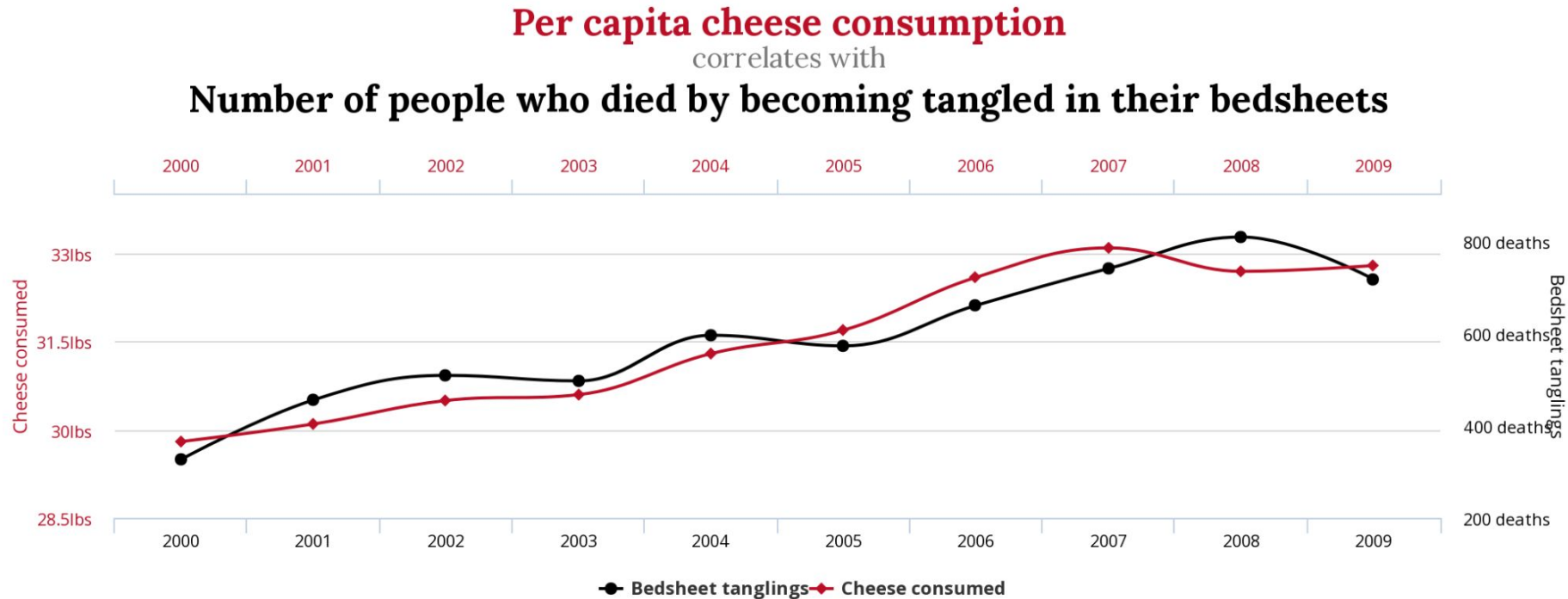
Correlação não significa
causalidade. Mas pode dar
um indício de causalidade.

Correlação: 99.8%

US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation

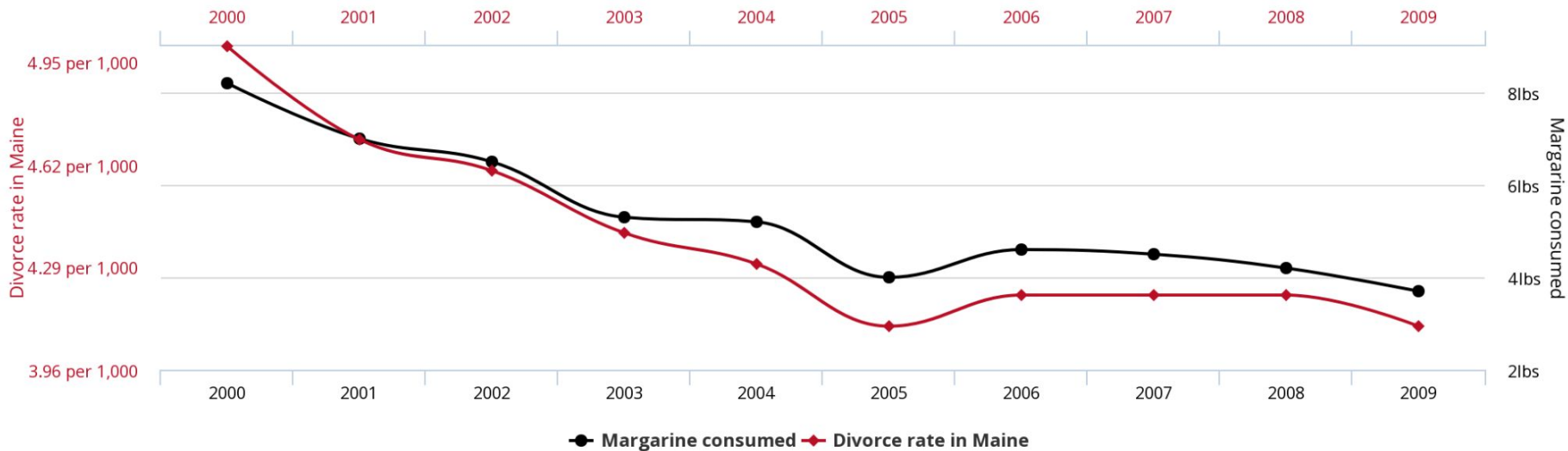


Correlação: 94.7%



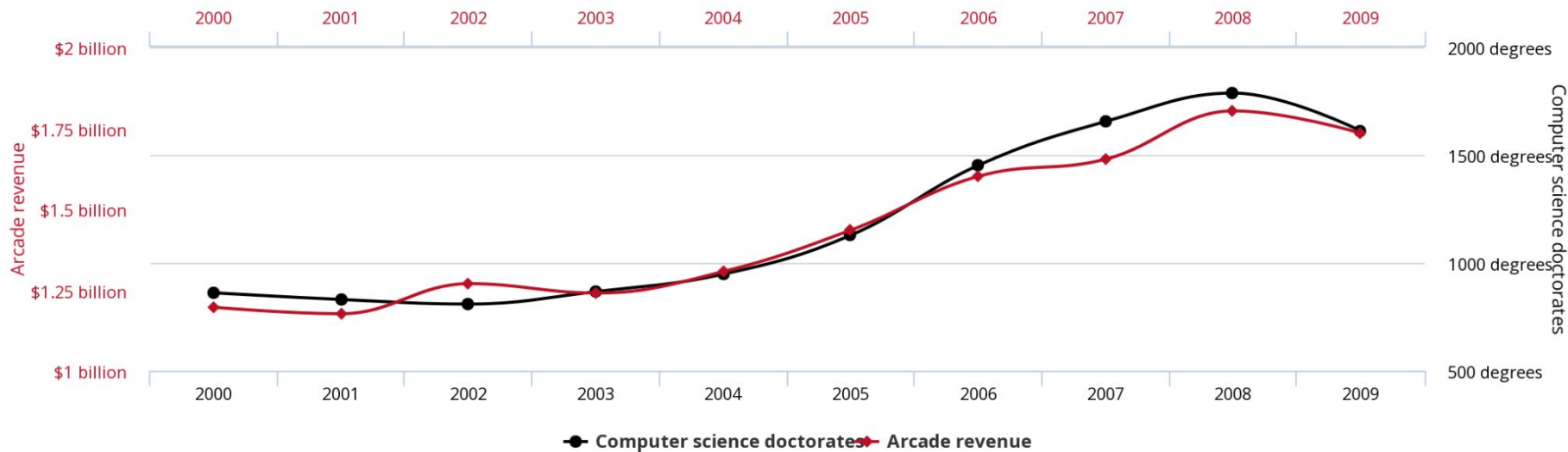
Correlação: 99.3%

Divorce rate in Maine correlates with Per capita consumption of margarine



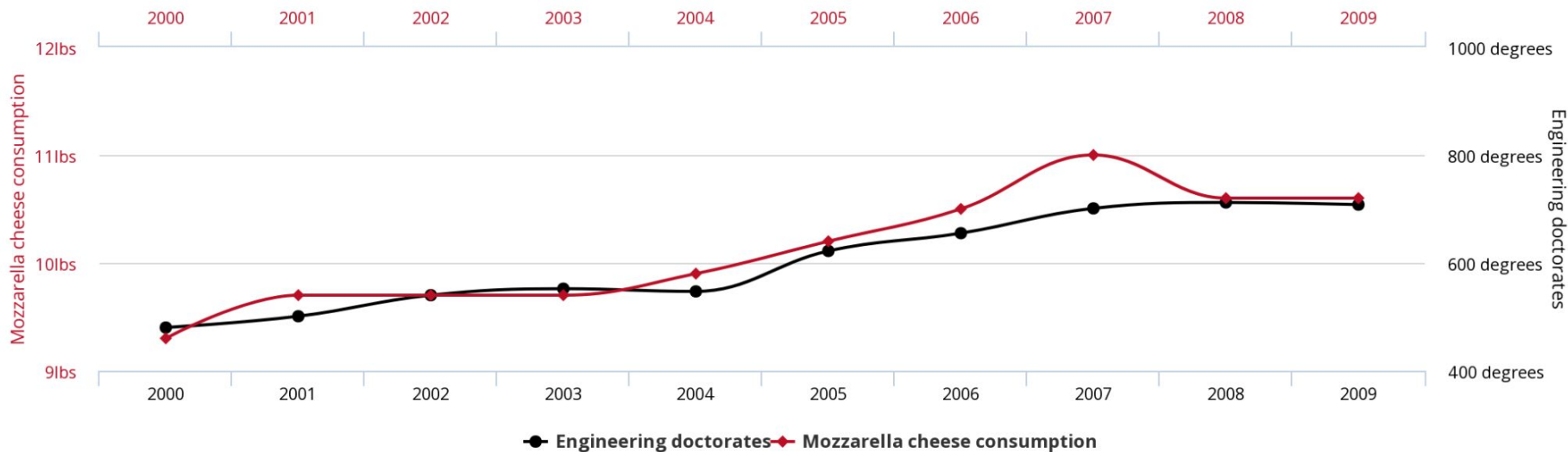
Correlação: 98.5%

Total revenue generated by arcades
correlates with
Computer science doctorates awarded in the US



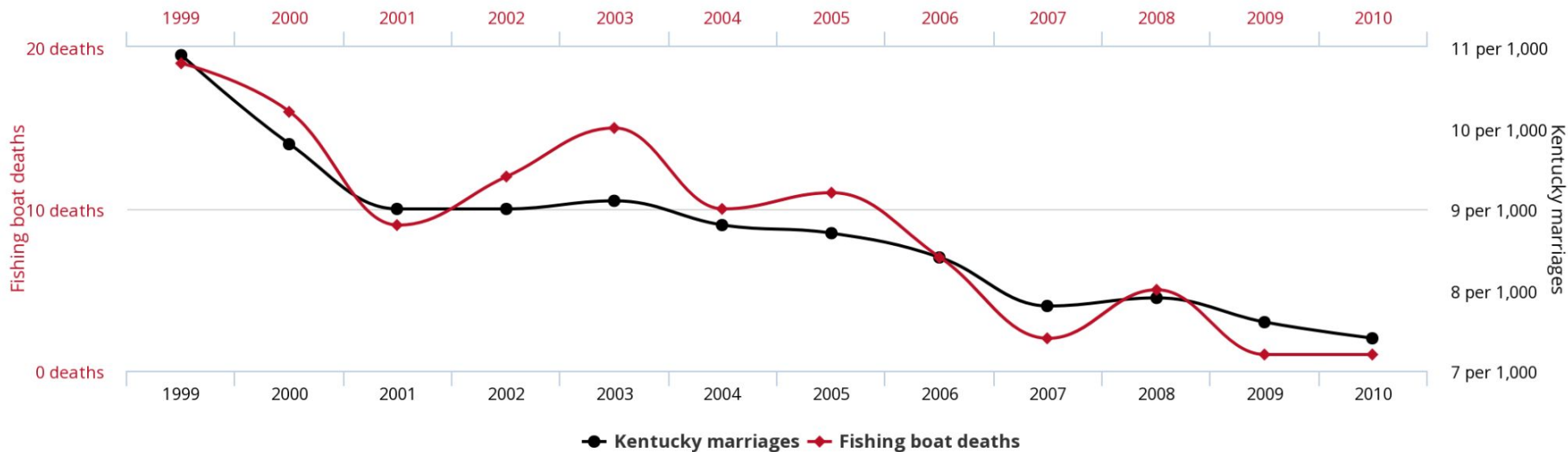
Correlação: 95.8%

Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded



Correlação: 95.2%

People who drowned after falling out of a fishing boat correlates with Marriage rate in Kentucky



Correlação não-paramétrica: Coeficiente de Spearman

- Substituir os valores pela ordem (rank) no qual estes valores crescem.
- A nova lista de valores deve ser [1, 2, 3, 4, ...]
- Se houver números iguais, substituir todos pelo valor médio do rank de todos eles.

O coeficiente de correlação de Spearman será:

$$r_s = \frac{\sum_{i=1}^N (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (R_i - \bar{R})^2} \cdot \sqrt{\sum_{i=1}^N (S_i - \bar{S})^2}}$$