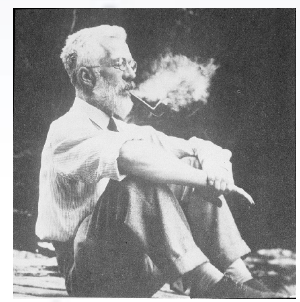
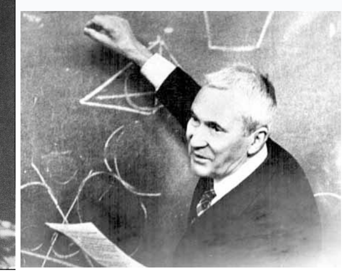
Applied Statistics

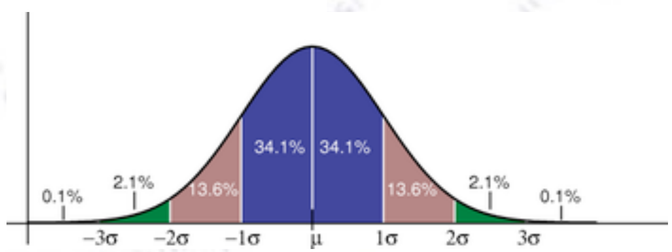
Correlations





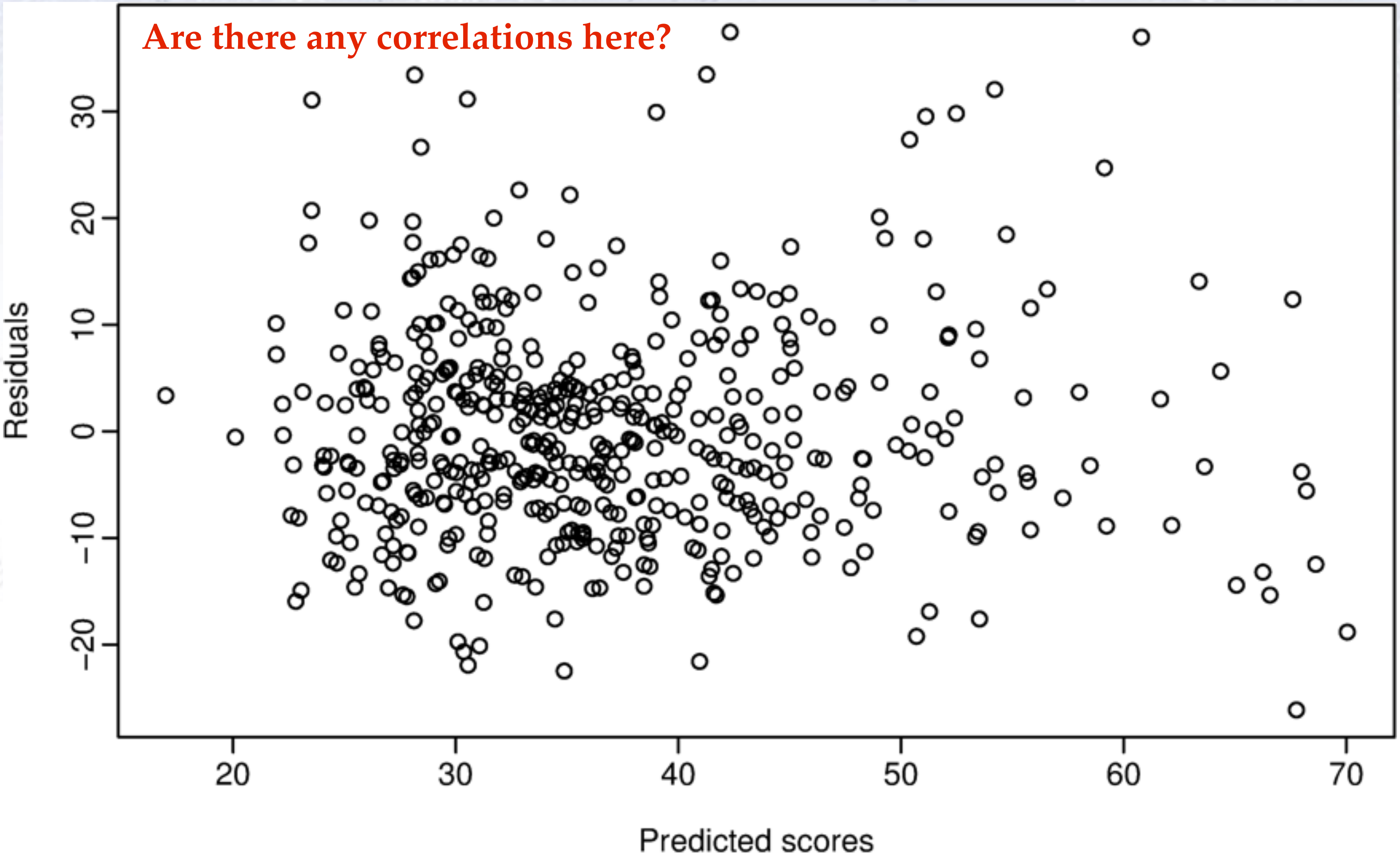



Troels C. Petersen (NBI)

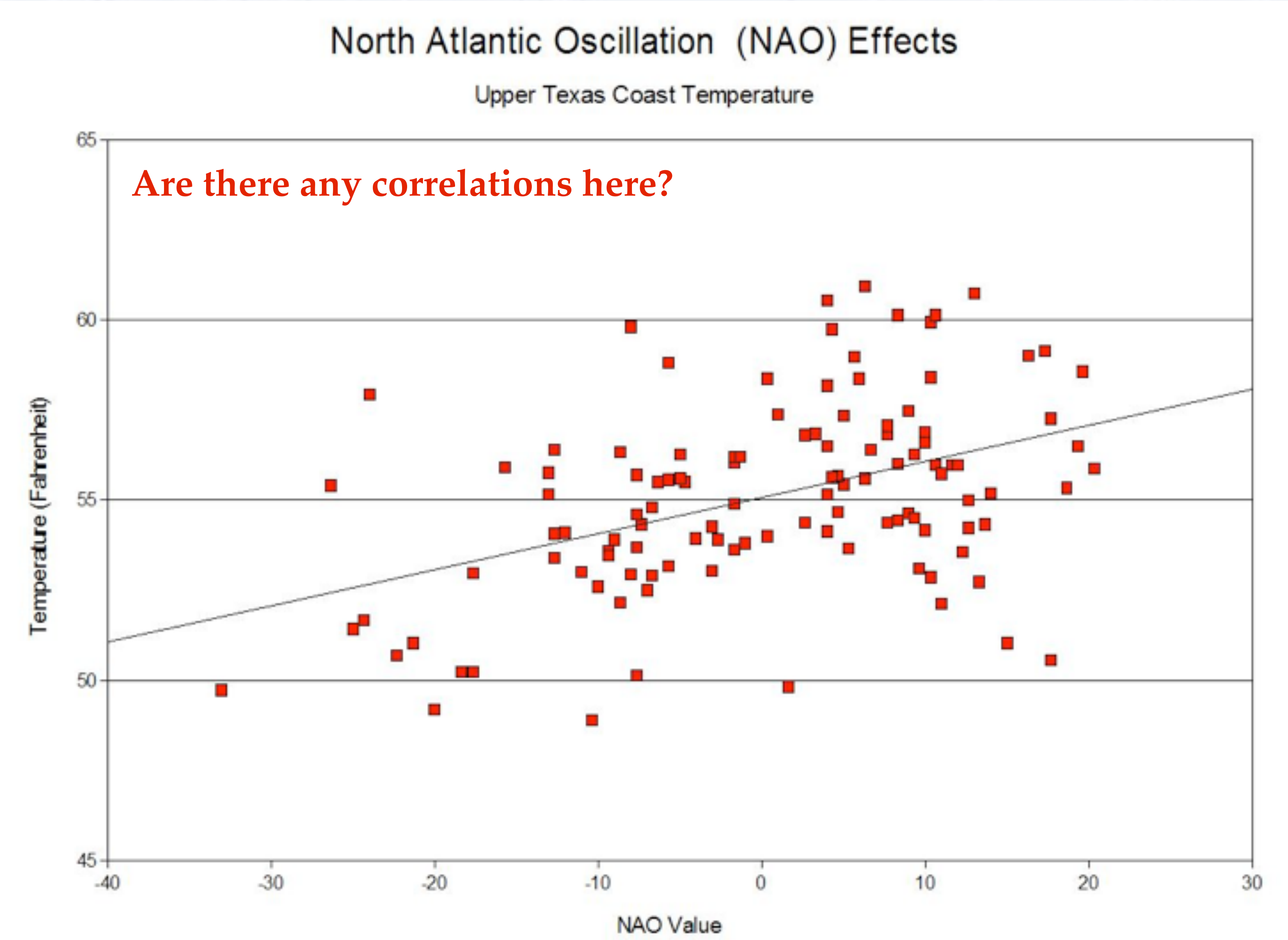


"Statistics is merely a quantisation of common sense"

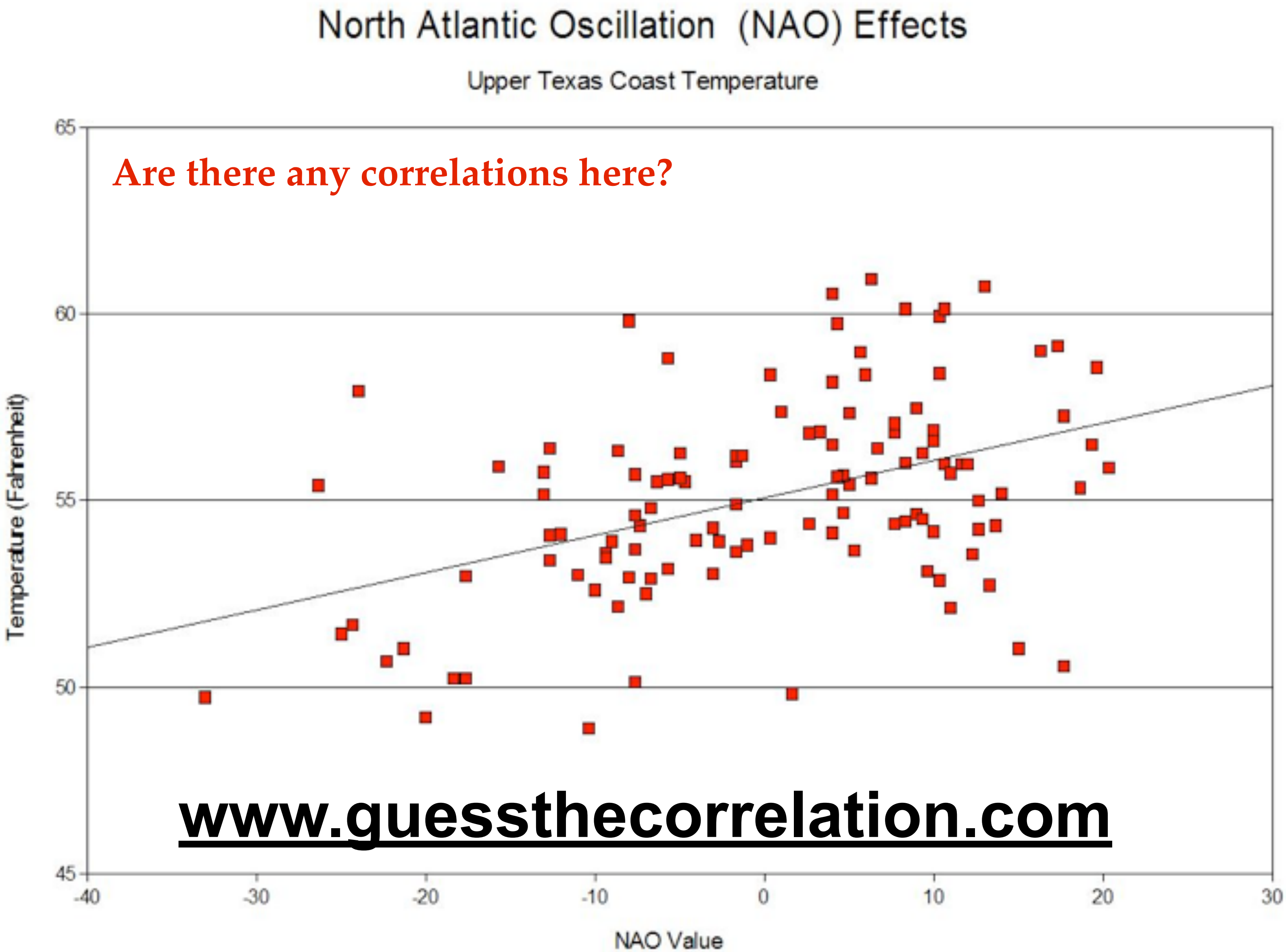
Correlation



Correlation



Correlation



www.guessthecorrelation.com

Correlation

Recall the definition of the Variance, V:

$$V = \sigma^2 = \frac{1}{N}\sum_i^n (x_i - \mu)^2 = E[(x-\mu)^2] = E[x^2] - \mu^2$$

Likewise, one defines the **Covariance, $\mathbf{V_{xy}}$:**

$$V_{xy} = \frac{1}{N}\sum_i^n (x_i - \mu_x)(y_i - \mu_y) = E[(x_i - \mu_x)(y_i - \mu_y)]$$

"Normalising" by the widths, gives the (linear) correlation coefficient:

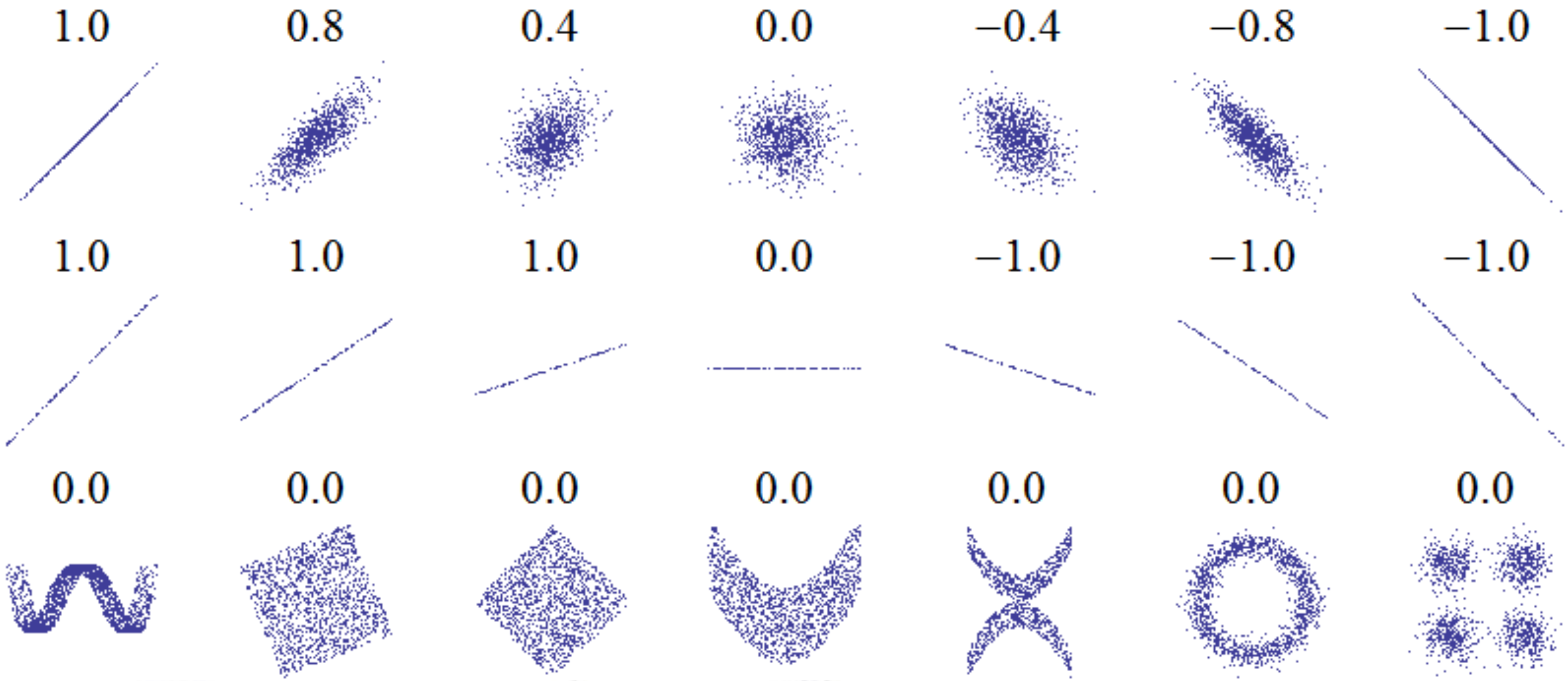
$$\rho_{xy} = \frac{V_{xy}}{\sigma_x \sigma_y}$$

$$-1 < \rho_{xy} < 1$$

$$\sigma(\rho) \simeq \sqrt{\frac{1}{n}(1-\rho^2)^2 + O(n^{-2})}$$

Correlation

Correlations in 2D are in the Gaussian case the "degree of ovalness"!



Note how ALL of the bottom distributions have $\varrho = 0$, despite obvious correlations!

Correlation

The correlation matrix V_{xy} explicitly looks as:

$$V_{xy} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1N}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2N}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_N^2 & \sigma_{N2}^2 & \cdots & \sigma_{NN}^2 \end{bmatrix}$$

Very specifically, the calculations behind are:

$$V = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}.$$

Correlation and Information

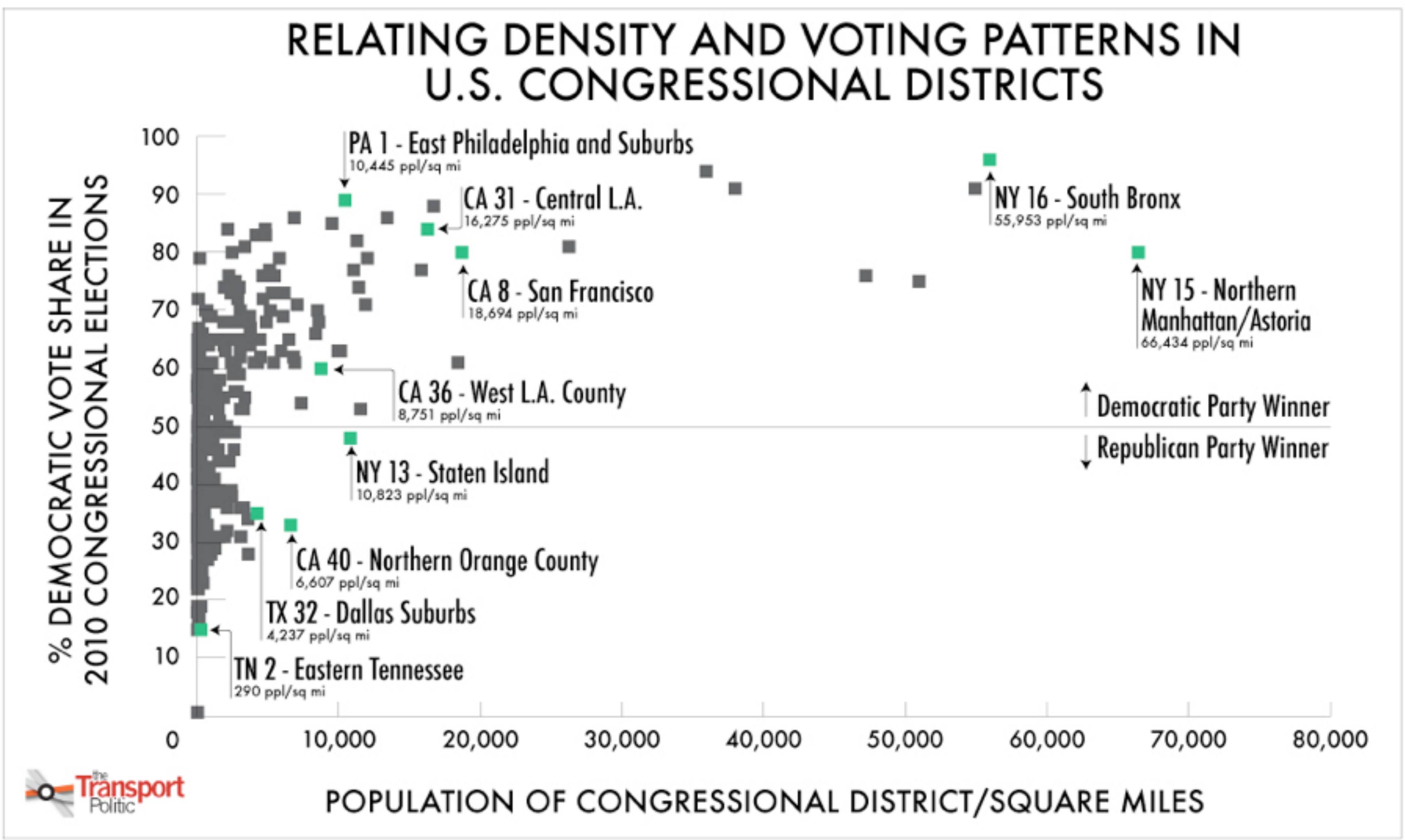
Correlations influence results in complex ways!

They need to be taken into account, for example in **Error Propagation!**

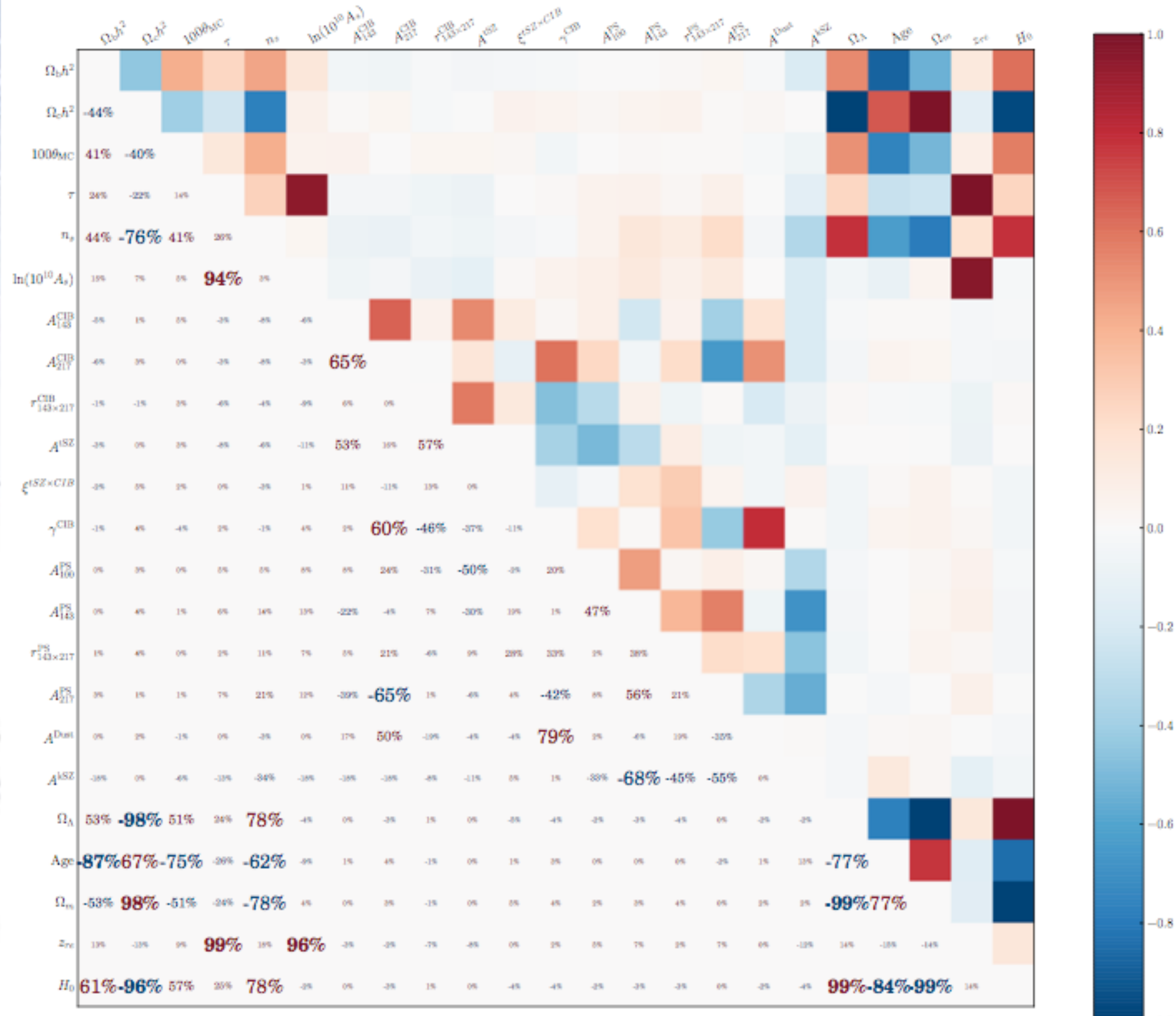
Correlations may contain a significant amount of information.

We will consider this more when we play with multivariate analysis.

Correlation example



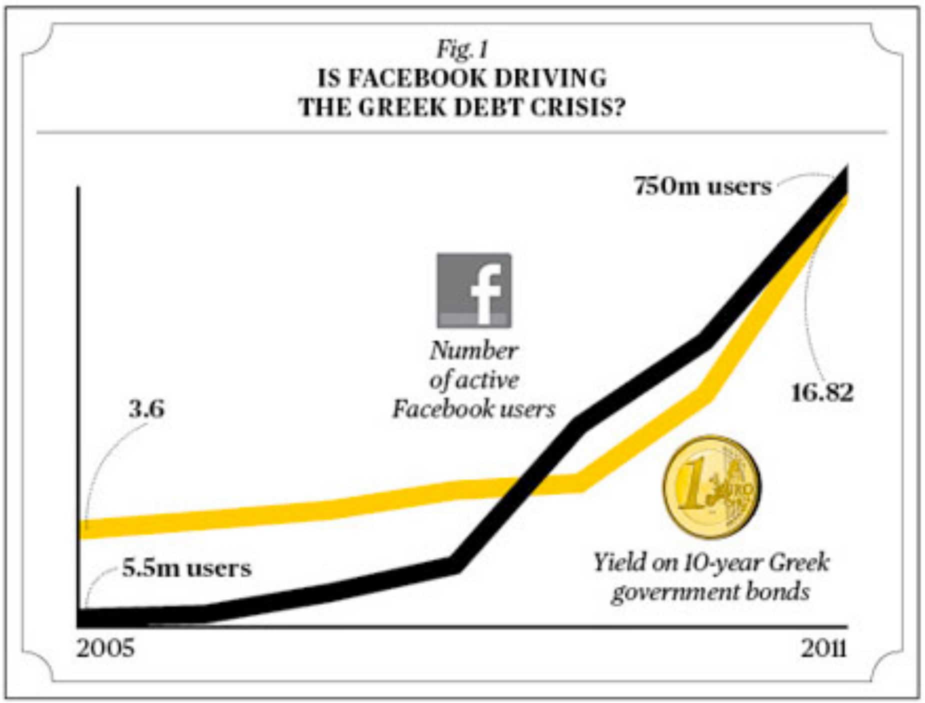
Planck example

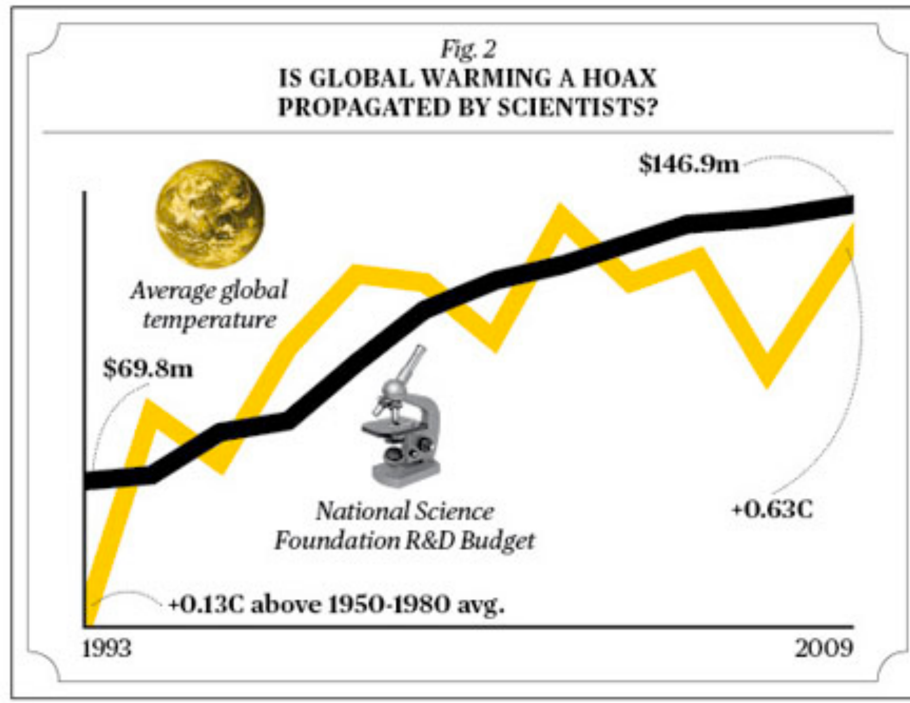


Correlation Vs. Causation

"Com hoc ergo propter hoc"

(with this, therefore because of this)





It is a common mistake to think that correlation proves causation…

Correlation Vs. Causation

"Com hoc ergo propter hoc"

(with this, therefore because of this)