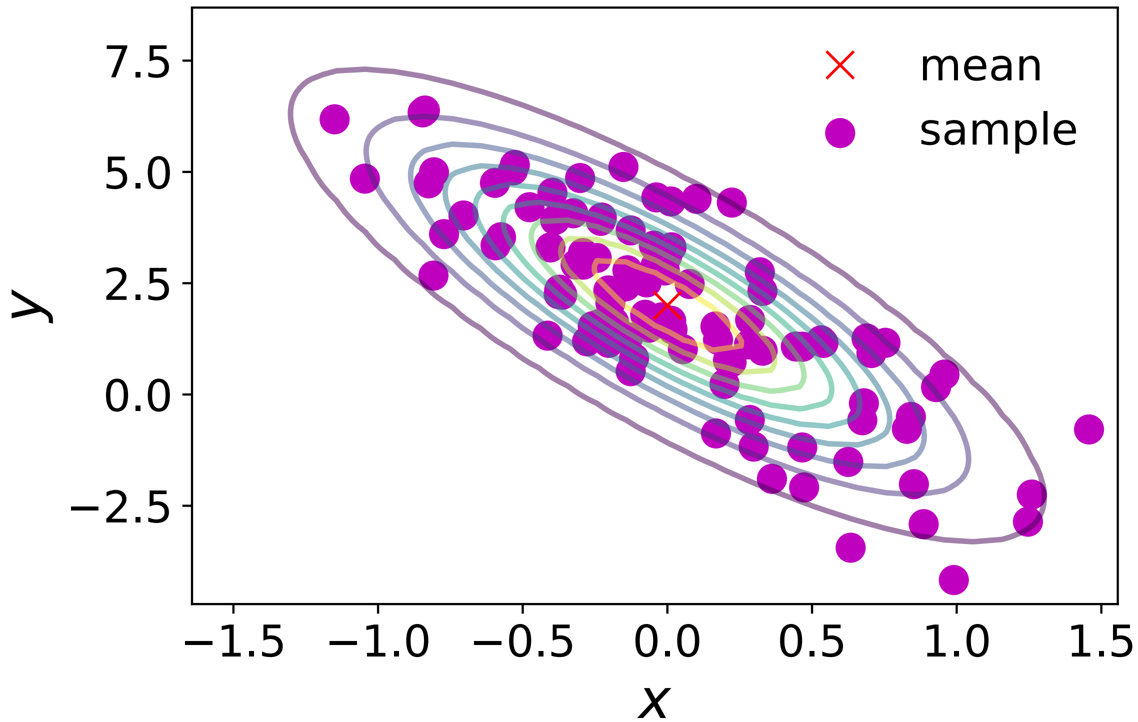
ENM 5310: Data-driven Modeling and Probabilistic Scientific Computing

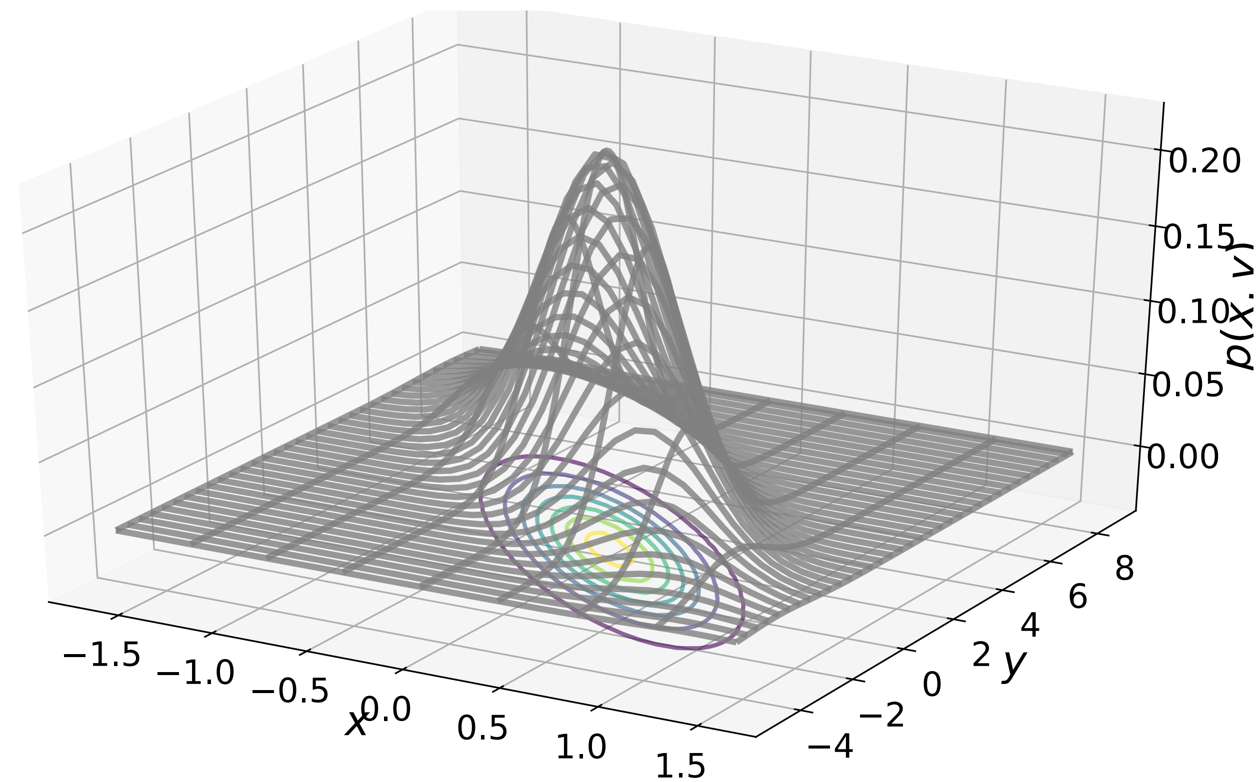
Lecture #3

Statistical Estimation

Penn
UNIVERSITY *of* PENNSYLVANIA

The multivariate Gaussian





$$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)$$

Mean, variance & high-order moments

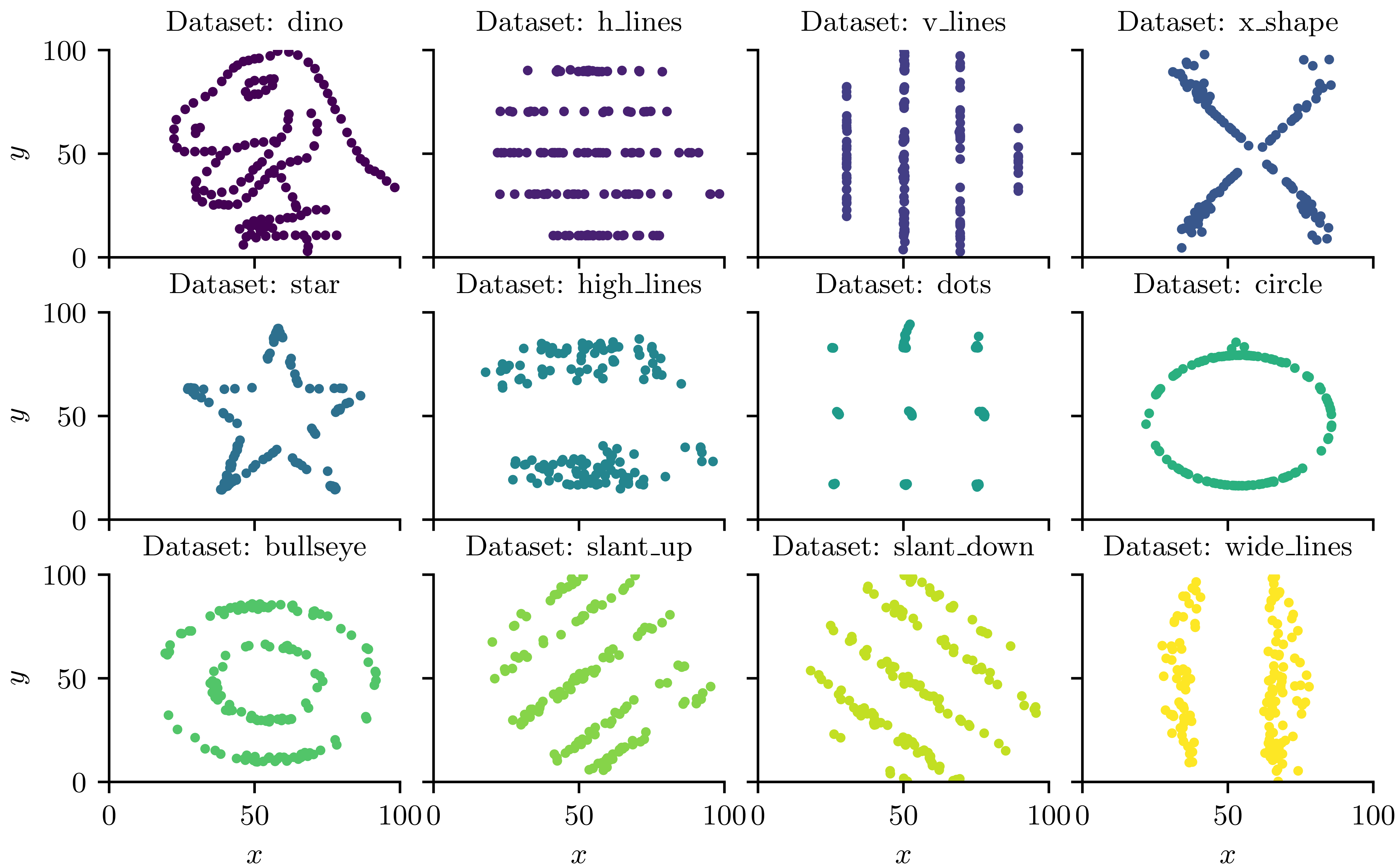
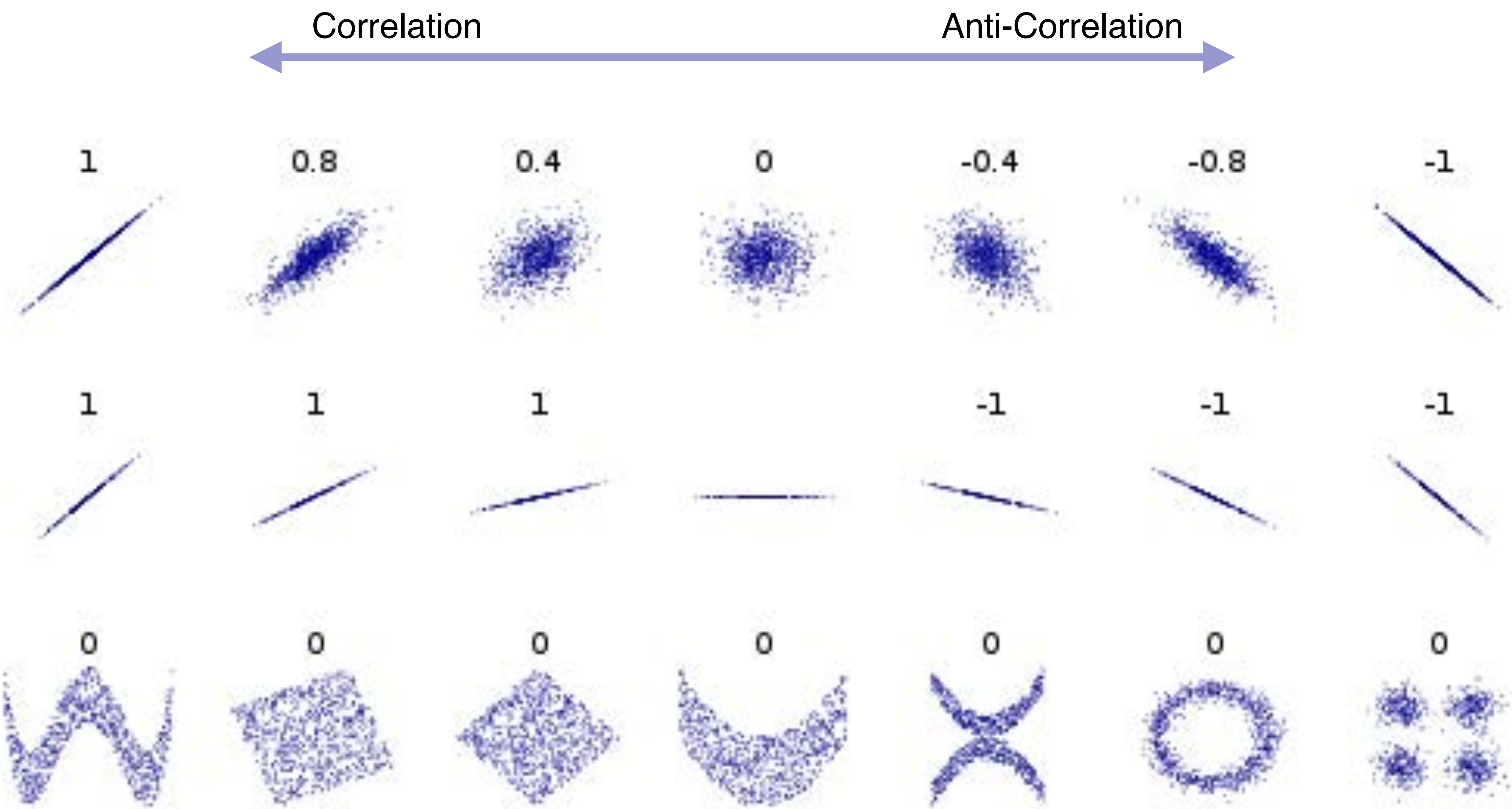


Figure 2.6: Illustration of the Datasaurus Dozen. All of these datasets have the same low order summary statistics. Adapted from Figure 1 of [MF17]. Generated by datasaurus_ dozen.ipynb.

Correlation and linear dependence



Entropy and Mutual Information

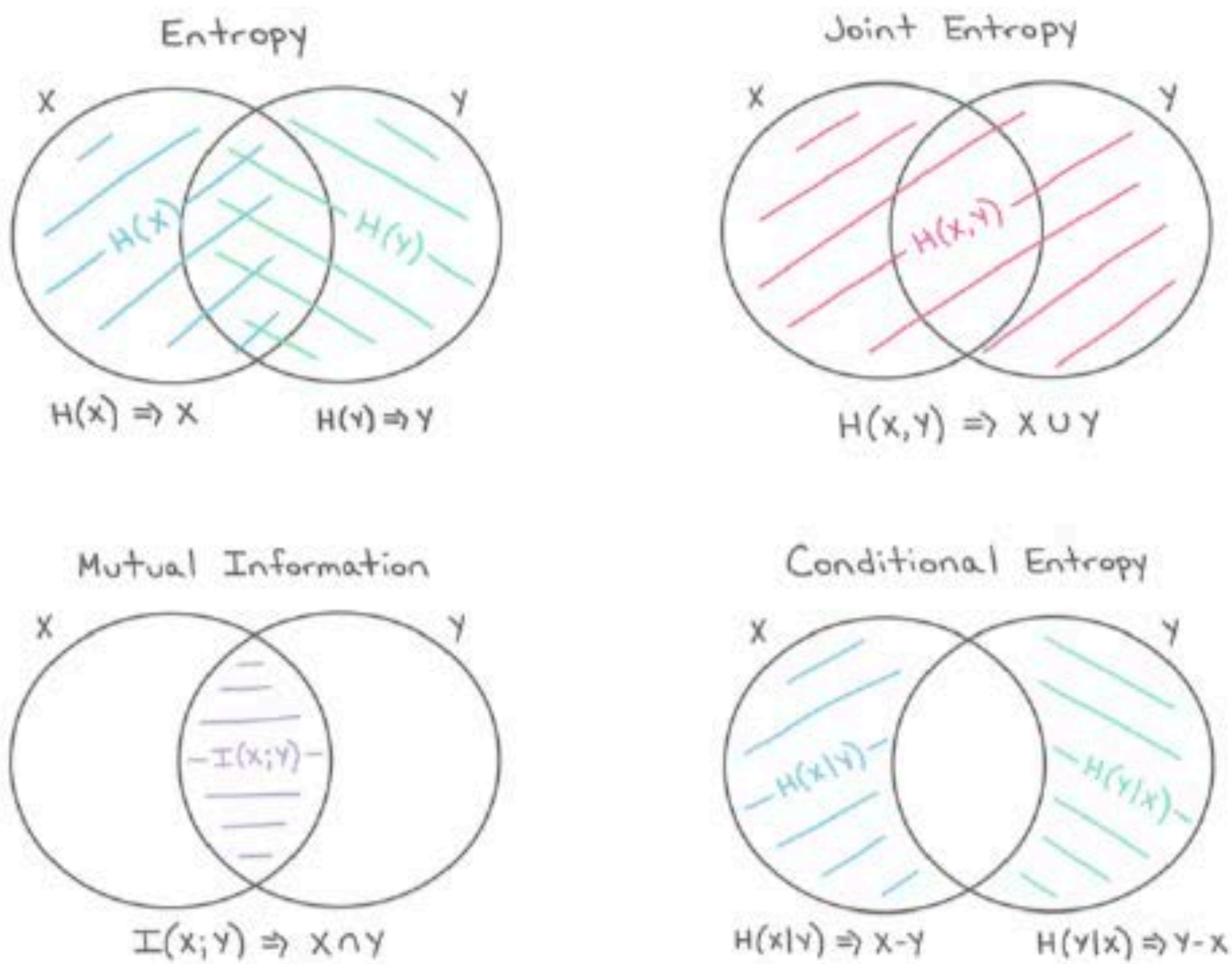
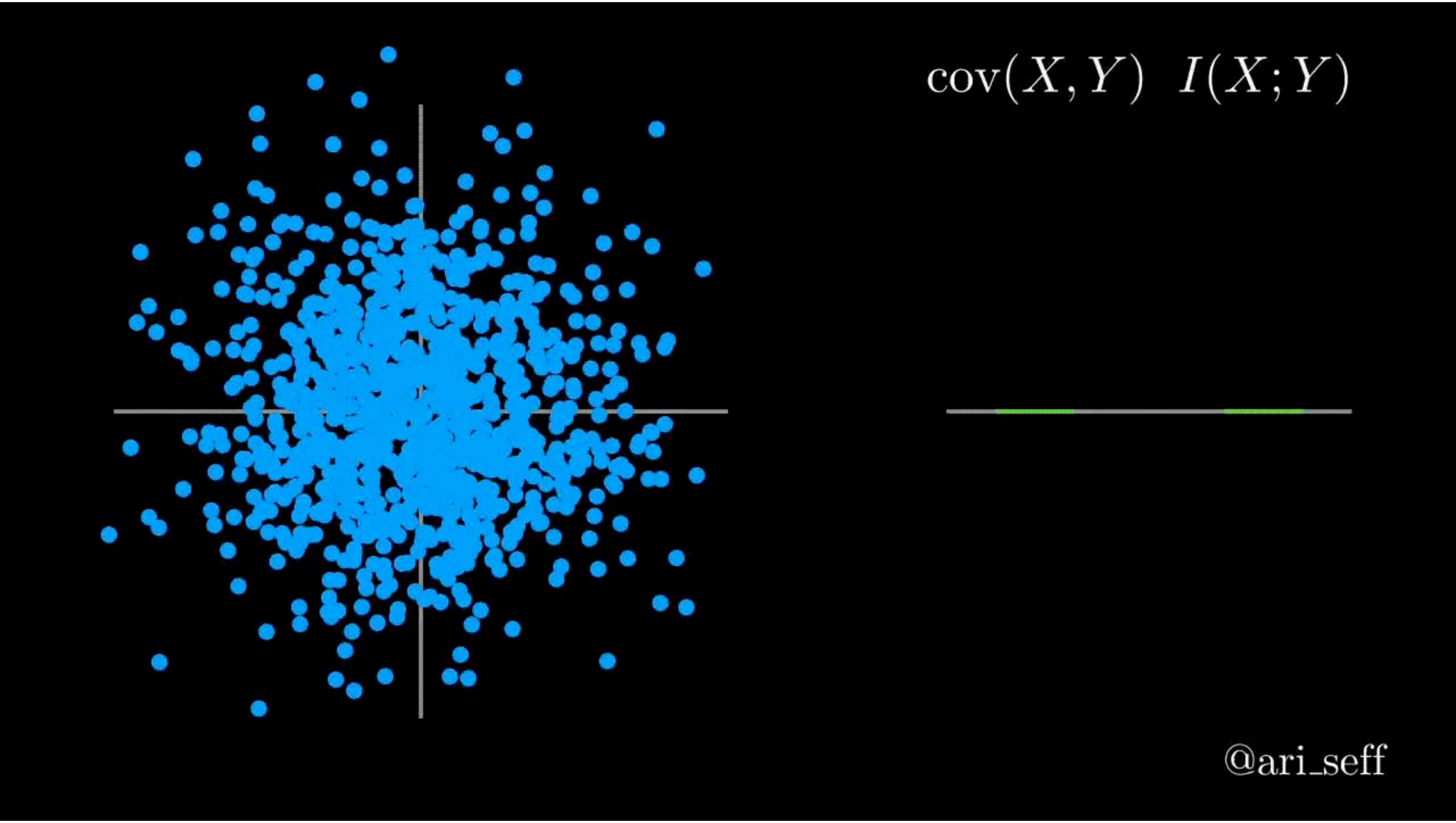
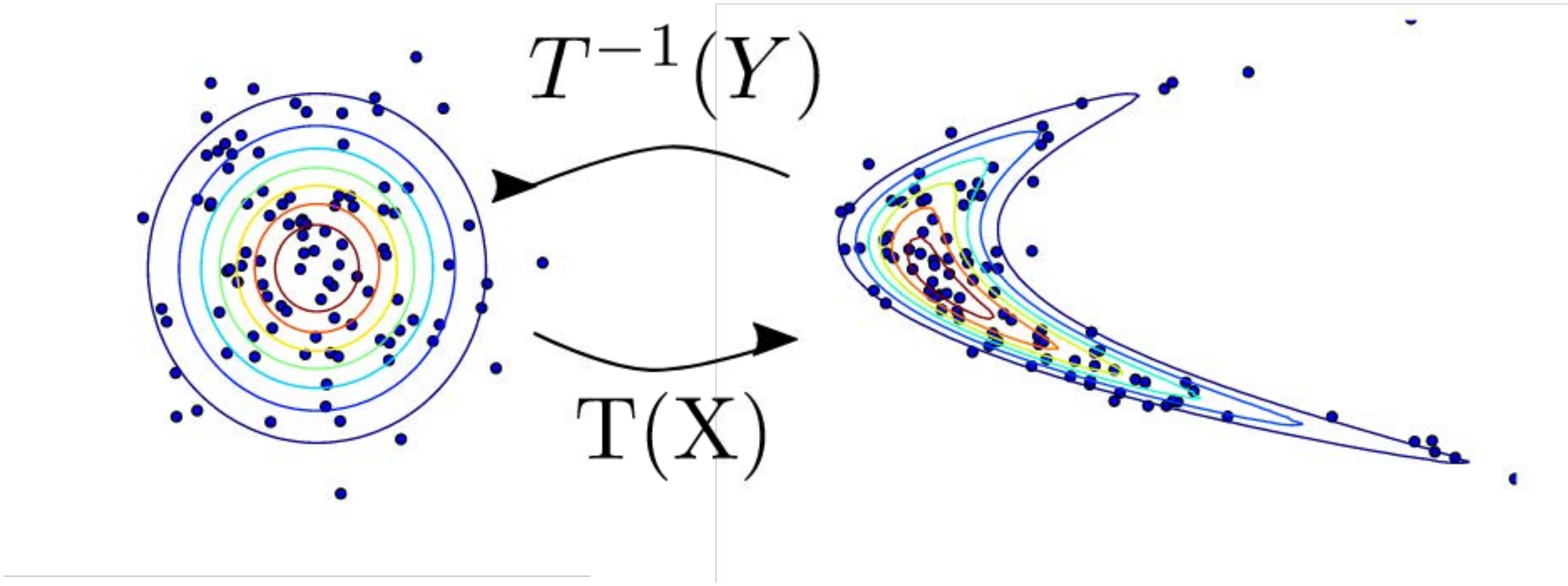


Figure 6.4: The marginal entropy, joint entropy, conditional entropy and mutual information represented as information diagrams. Used with kind permission of Katie Everett.

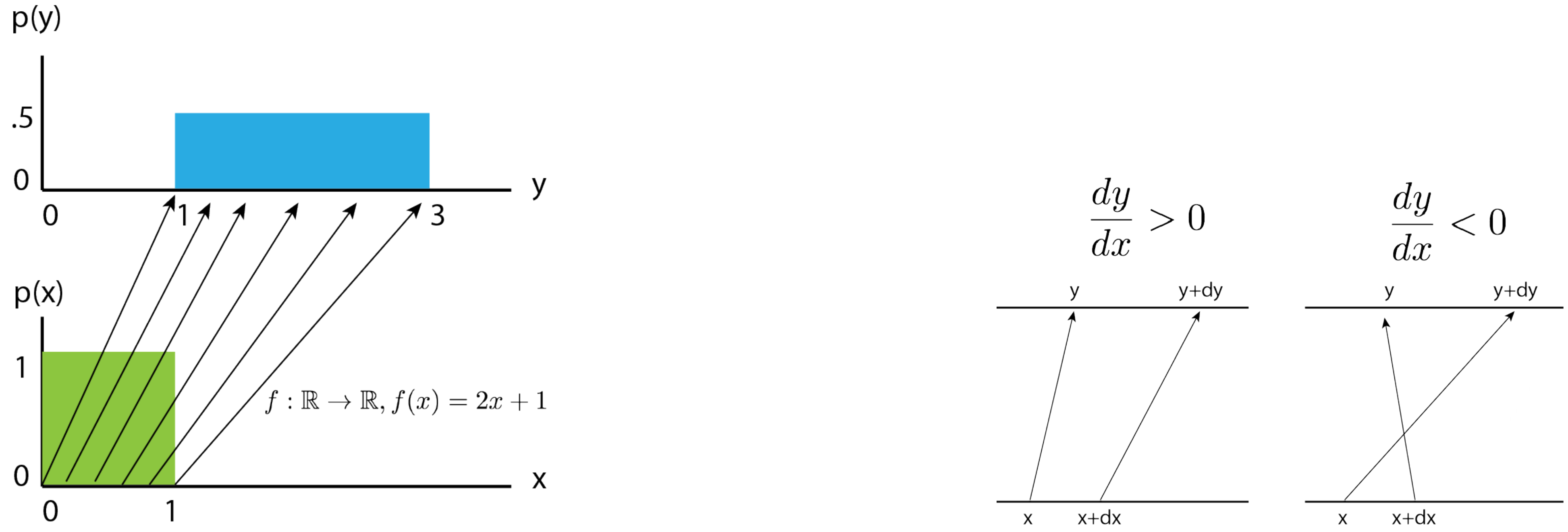
Covariance vs Mutual Information



Transformations



Change of Variables

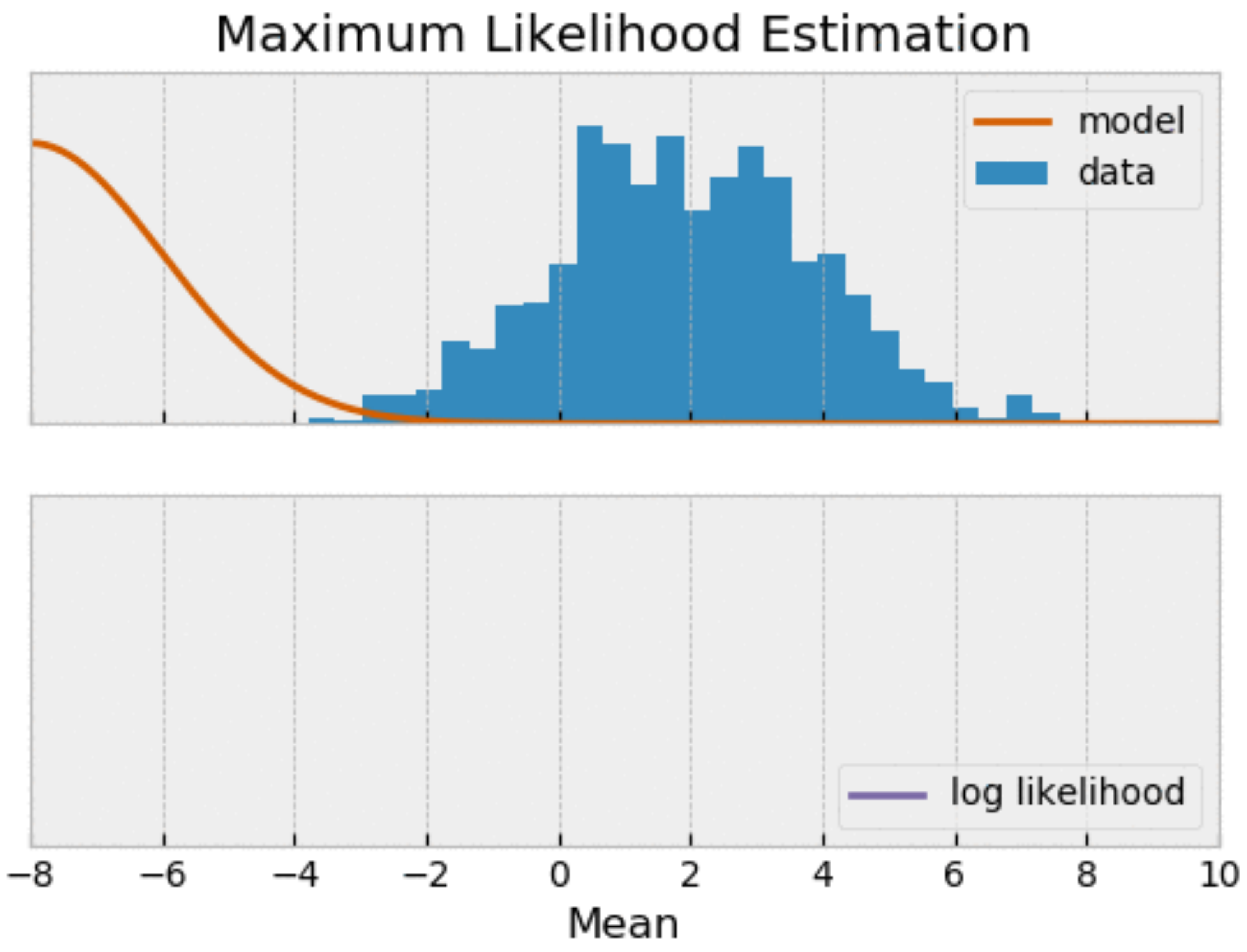


(a) *(b)*

Figure 2.19: (a) Mapping a uniform pdf through the function $f(x) = 2x + 1$. (b) Illustration of how two nearby points, x and $x + dx$, get mapped under f. If $\frac{dy}{dx} > 0$, the function is locally increasing, but if $\frac{dy}{dx} < 0$, the function is locally decreasing. From [Jan18]. Used with kind permission of Eric Jang.

Maximum likelihood estimation

$$\theta_{\mathrm{MLE}} = \arg\max_{\theta \in \Theta} p(\mathcal{D}|\theta)$$



Bayesian estimation

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$