$\{1/2, 1/4, 1/8, 1/8\}$

$$[1, 2, 3, 4] \xrightarrow{\text{no compression}} [00, 01, 10, 11]$$

$$[1, 2, 3, 4] \xrightarrow{\text{with compression}} [0\ 10\ 110\ 111]$$

$$\langle \text{length} \rangle = \tfrac{1}{2}1 + \tfrac{1}{4}2 + \tfrac{1}{8}3 + \tfrac{1}{8}3 = \tfrac{7}{4} < 2$$

$$H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8}) = -\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{2}{8}\log\tfrac{1}{8} = \tfrac{7}{4}$$

We can use any encoding which gives average length >= Shannon entropy.
if we use something <, then we might have our data corrupted

$$H(X, Y) = -\sum_{x,y} p(x,y) \log_2 p(x,y)$$

$$H(X|Y) = H(X,Y) - H(Y)$$

(Mutual information)

X, Y in common (How much?)

$$H(X:Y) = H(X) + H(Y) - H(X,Y)$$

How much do X, Y have in common
Mutual information of X & Y

Fidelity: At the level of probability

In quantum realm

$$S(\rho) = -\mathrm{tr}(\rho \log_2 \rho)$$

if λ_i are eigenvalues of ρ

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i$$

(trace = sum of diagonals)

if $S(A) = S(B)$ A & B are in pure state

$$S(A) \geq 0$$

if $$\rho = \frac{I}{d} = \begin{bmatrix} 1/d & & \\ & 1/d & \\ & & \ddots & \\ & & & 1/d \end{bmatrix}$$

$S(\rho) \leq \log_2 d$, $S(A) = S(B)$ A, B completely mixed

[READ ABOUT THIS A BIT MORE]