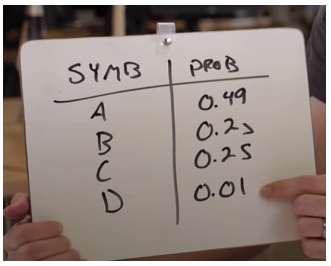
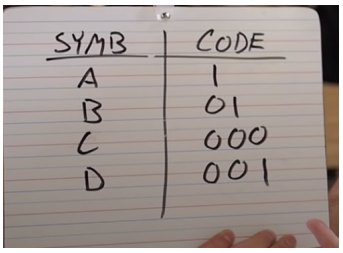
Data Compression

YouTube video:
https://www.youtube.com/watch?v=6rnF2Mo80x0





DAT-1.D.1
Data compression can reduce the size (number of bits) of transmitted or stored data.

DAT-1.D.2
Fewer bits does not necessarily mean less information.

DAT-1.D.3
The amount of size reduction from compression depends on both the amount of redundancy in the original data representation and the compression algorithm applied.

DAT-1.D.4
Lossless data compression algorithms can usually reduce the number of bits stored or transmitted while guaranteeing complete reconstruction of the original data.

DAT-1.D.5
Lossy data compression algorithms can significantly reduce the number of bits stored or transmitted but only allow reconstruction of an approximation of the original data.

For example, the above probability distribution indicates that the symbol A appears frequently (in other words, there is a lot of redundancy in the data--A is occurring basically half the time). So it makes the most sense to choose a compression algorithm that would require A's encoding to use a few bits as possible (in the case above, A just requires one bit: 1).

For the purposes of the AP exam, you do not need to be able to choose the best encoding for a situation from scratch.

So let's say you were streaming a YouTube video online.
If you have a high quality internet connection, you're not going to have any issues streaming a 4K video.

However, if your internet really sucks, then you're not going to want to stream in 4K. You're going to want the data more compressed so that the video isn't constantly pausing.

If you're streaming at 4K and the video was originally recorded in 4K, then this is an example of **lossless** data compression. (The data is sent to you in a way that guarantees you are able to reconstruct the video on your end exactly how it was originally recorded.)

If you needed to stream at a lower definition than 4K, then this would be an example of **lossy** data compression. The reason why is because you can only reconstruct a lower quality version of the originally recoded video.

Think of the word winless. Winless means you didn't ever win.
Lossless means that you didn't ever have to do something lossy.

DAT-1.D.6
Lossy data compression algorithms can usually reduce the number of bits stored or transmitted more than lossless compression algorithms.

DAT-1.D.7
In situations where quality or ability to reconstruct the original is maximally important, lossless compression algorithms are typically chosen.

DAT-1.D.8
In situations where minimizing data size or transmission time is maximally important, lossy compression algorithms are typically chosen.

Binary (Base 2)

(Decimal)
Base 10

2^3 2^2 2^1 2^0

10^3 10^2 10^1 10^0

Binary is in base 2, which means there are 2 symbols available to choose from: 0, 1
Decimal is in base 10, which means there are 10 symbols available to choose from: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Ternary (Base 3)

Ternary is in base 3, which means there are 3 symbols to choose from: 0, 1, 2

$$\frac{2}{3^3} \quad \frac{1}{3^2} \quad \frac{0}{3^1} \quad \frac{2}{3^0}$$

$$2 \cdot 3^3 + 1 \cdot 3^2 + 0 + 2 \cdot 3^0$$

$$54 + 9 + 0 + 2$$

$$65$$

$$65_{10} = 2102_3$$

Hexadecimal (Base 16)

Since we are in base 16, there are 16 symbols to choose from:
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A (10), B (11), C (12), D (13), E (14), F (15)

$$\frac{A^{(10)}}{16^4} \quad \frac{4}{16^3} \quad \frac{F^{(15)}}{16^2} \quad \frac{1}{16^1} \quad \frac{6}{16^0}$$

$$10 \cdot 16^4 + 4 \cdot 16^3 + 15 \cdot 16^2 + 1 \cdot 16^1 + 6 \cdot 16^0$$

$$A4F16_{16} = 675606_{10}$$

Base 64