Information Layer

AP CSP @ SouthLake Christian Academy

We saw that simple messages, colors, and images require many bits.

EX: A 1200x1200 pixel 24-bit color emoji requires

$$1200 \times 1200 \times 24 = 34560000 \text{ bits} = 4320000 \text{ bytes}$$

$$= 4320000 \text{ bytes}$$


54630
bytes

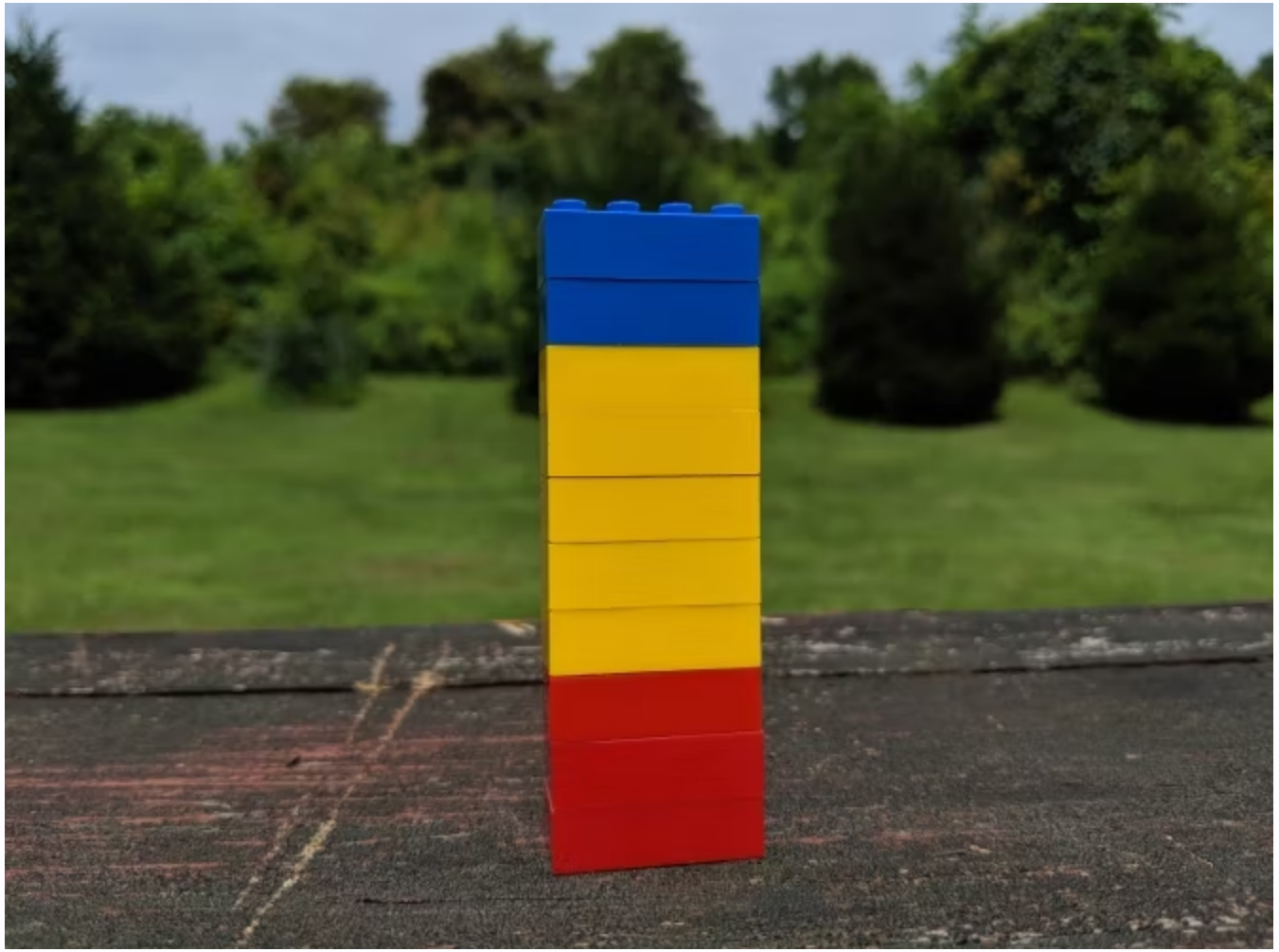
Compression

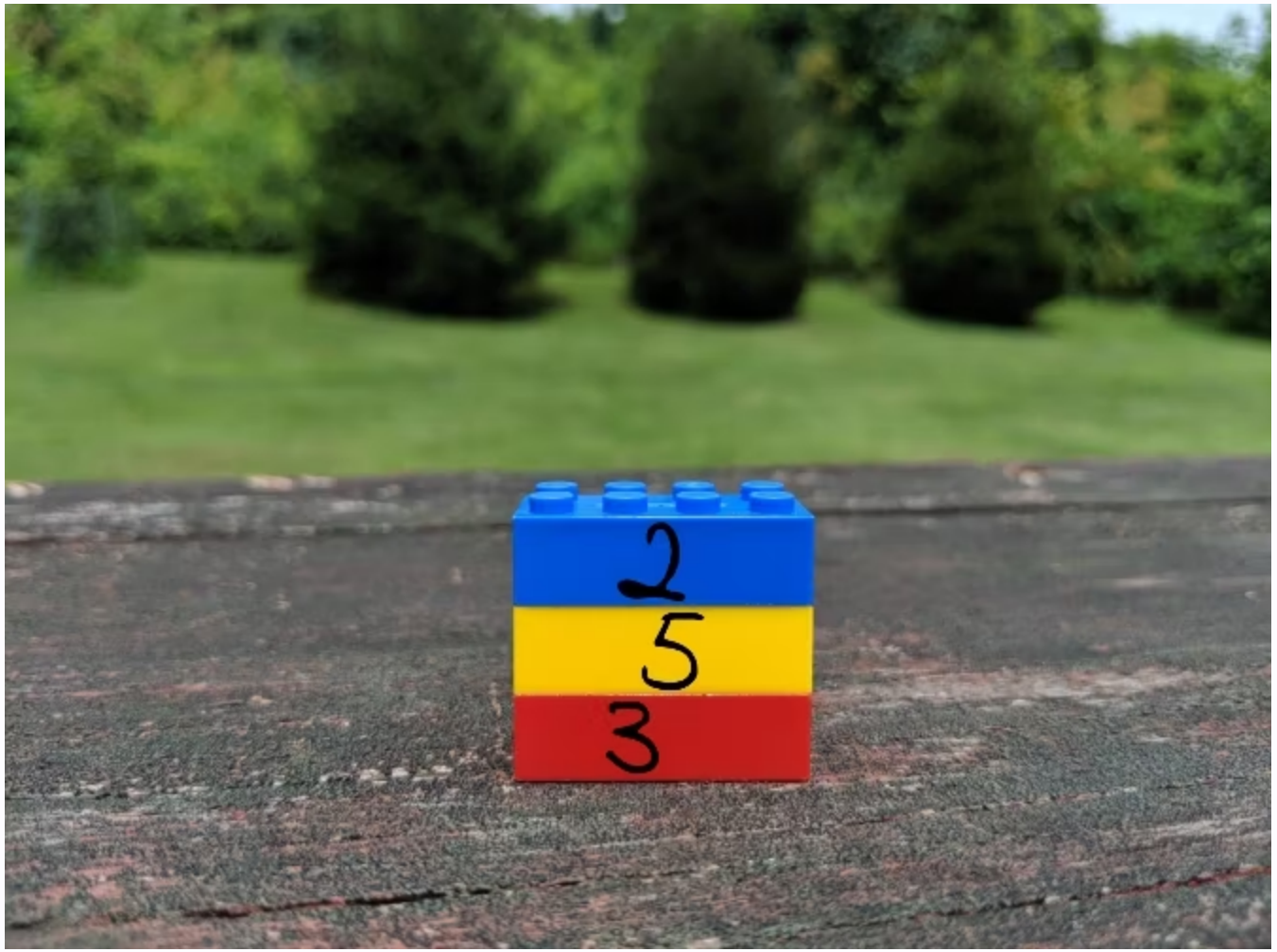
- **lossy**: compression that *loses information*
  - data gets trashed to save space
  - unable to reconstruct original data
- **lossless**: compression that does not loses information
  - able to reconstruct original data

Run Length Encoding

- a method of compression that describes how many times a character or text is repeated
- starting from left to right, we start counting how many times `1` is repeated, then how many times `0` is repeated, etc.

Instead of `000001111111` → `0 5 7`





Which type of compression?

Size: 12KB

Size: 4KB

Which type of compression?

Takeaway

- Run Length encoding is a *lossless* compression
- *Lossy* compression generally allows for faster transmission
  - since we threw out bits, we can send a message in less time

Quantifying Compression

compression rate: $1 - \frac{\text{compressed data size}}{\text{original data size}}$

also called *space saving*

original size	compressed size	compression rate
12KB	4KB	$1 - \frac{4}{12} \approx 67\%$

Example

original bin	compressed dec	compressed bin	compression rate
`0b000001111111`	`0 5 7`	`0b000101111`	$1 - \frac{9}{12} = 25\%$

Key Takeaways

1. Compressing data saves us bits!
2. Compression can remove unnecessary data (**lossy**)
    - generally faster transmission speed
3. Compression can just restructure data (**lossless**)
4. Compression methods work better for large data