Digitization

Objectives

- Understand digital data.
- Understand the process of digitization.

Chapter 1 Background

ANALOG VS. DIGITAL REPRESENTATIONS

Analog Information

Examples:

- time
- weight
- temperature
- line length
- width and length of a sheet of paper

Analog Information

More examples:

- sound loudness
- light brightness
- color saturation and hue

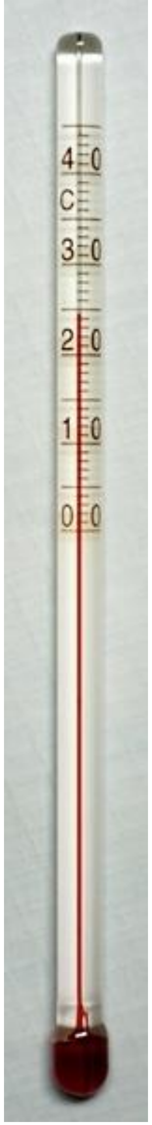
Analog Information

- Continuous information
- An infinite number of divisions exist between any two measurements

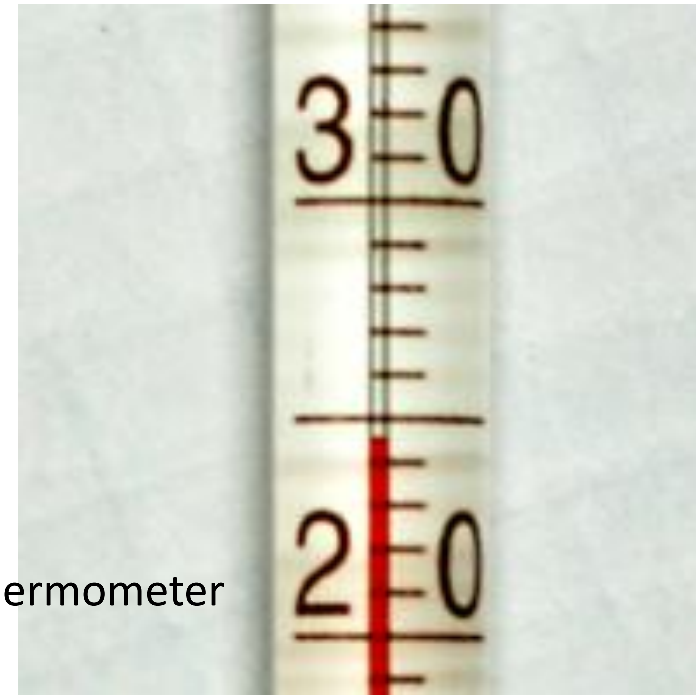
What is the length of the pencil?

What is the length of the pencil?

What is the temperature?



Analog Thermometer vs. Digital Thermometer



analog thermometer

digital thermometer

Analog Scales vs. Digital Scales

What are the characteristics of digital readings?

Analog vs. Digital

- Analog information
  - continuous
  - made up of infinite number of data points

- Digital data
  - discrete

Discrete Data

Examples:

- number of persons

  There is no in-between one person and two persons.

- choices in multiple-choice questions

  There is no in-between choice A and choice B.

Analog vs. Digital Therometers and Scales

- What are the limitations of these analog and digital devices?

- What are the advantages of these analog and digital devices?

Analog vs. Digital

- Sight and sound we peceive in our natural world are analog information--continuous and infinite number of points between any two points.

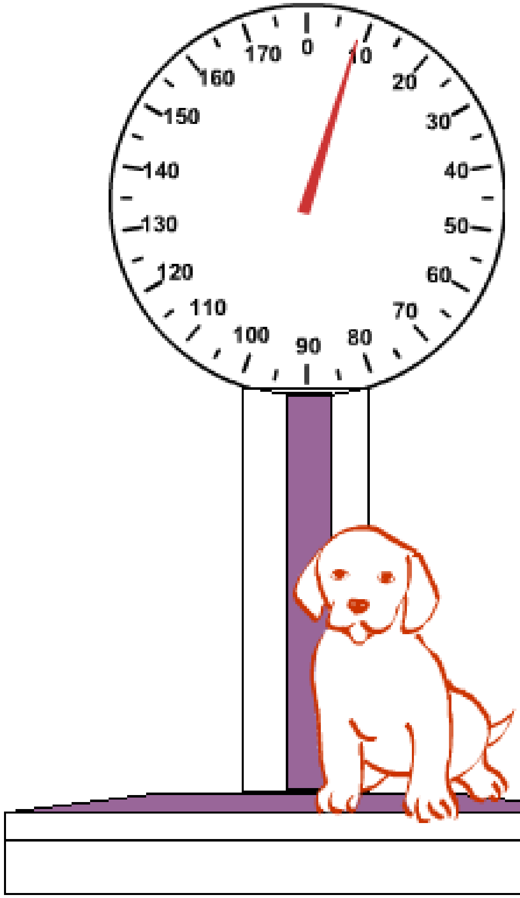
- Computers handle discrete digital data. In addition, the amount of data has to be finite.

- Sight and sound must be converted into finite discrete digital data in order for the computer to handle.

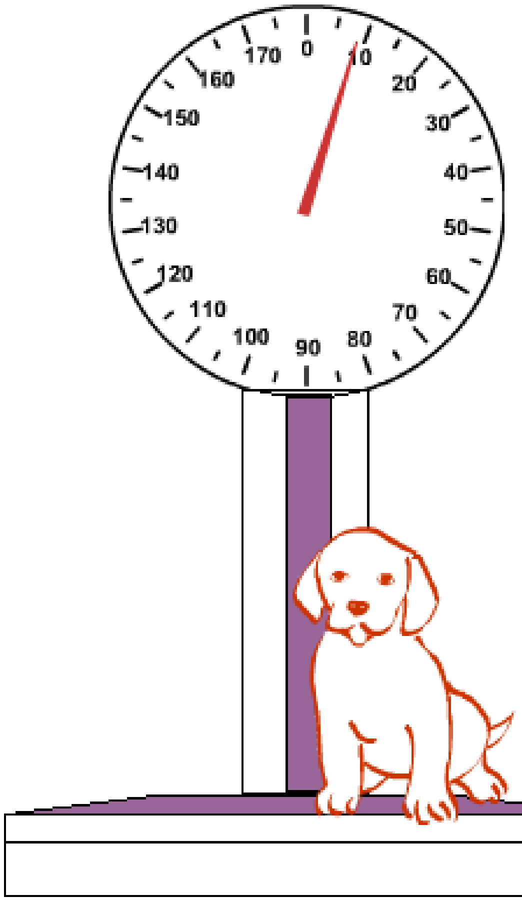
Before we talk about the conversion process, let's look at a scenario:

Monitoring a puppy's weight in his first year

Suppose you use an analog scale to weigh the puppy



Now, what is the weight you would note down for this puppy?



See the problem in picking a number to represent an analog measurement?

Number of Decimal Places

- In recording the weight, you must decide the number of decimal places to use.
- This determines the precision or exactness of the measurement.

- How many will give an exact measurement? How many is enough? How many is too many?

Using More Decimal Places

- Pros :
  - increase the precision in general
    (But how many is meaningful?)
  - Will allow finer distinction between values
    (will explain in the next slide)

- Cons:
  - Require more paper and paperwork.
  - Take longer to read through and interpret the numbers.

Distinction Between Values

With one decimal place:

– You can have 10 different values between say 2 and 3:
  2.1, 2.2, ...3.0

- You can distinct between 2.5 and 2.8.
- But 2.5 and 2.8 would have been rounded to the same value of 3 the values do not allow decimal places.

Distinction Between Values

Suppose the allowable weight read outs are these 10 levels:

0, 5, 10, 15, 20, 25, 30, 35, 40, 45

Then,

2 pounds: rounded to 0 pound

3 pounds: rounded to 5 pounds

The difference between 2 and 3 pounds is 1 pound.

But now, it become 5 pounds if we use these levels.

Now, how often would you weigh the puppy to produce a "good" monitoring of his weight over his first year?

A. once a year
B. once a month
C. every two weeks
D. every week
E. every day
F. every hour
G. every minute
H. every second

What are your considerations in deciding how often to weigh the puppy?

Considerations

- What happens if you weigh the puppy not often enough?
- What happens if you weigh the puppy too often?
- Is there one right answer?
- Will you use the same weighing schedule to monitor the weight of an adult dog?

Back to the Computer

DIGITIZATION: SAMPLING AND QUANTIZATION

Number of bits

Binary/Decimal																
Bit #	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Power	2^{15}	2^{14}	2^{13}	2^{12}	2^{11}	2^{10}	2^{9}	2^{8}	2^{7}	2^{6}	2^{5}	2^{4}	2^{3}	2^{2}	2^{1}	2^{0}
Decimal Bit Value	32678	16384	8192	4096	2048	1024	512	256	128	64	32	16	8	4	2	1
Max Value	65535_{10}															

Table 2

Digitization

- To convert analog information into digital data that computers can handle

- 2-step process:
  1. sampling
  2. quantization

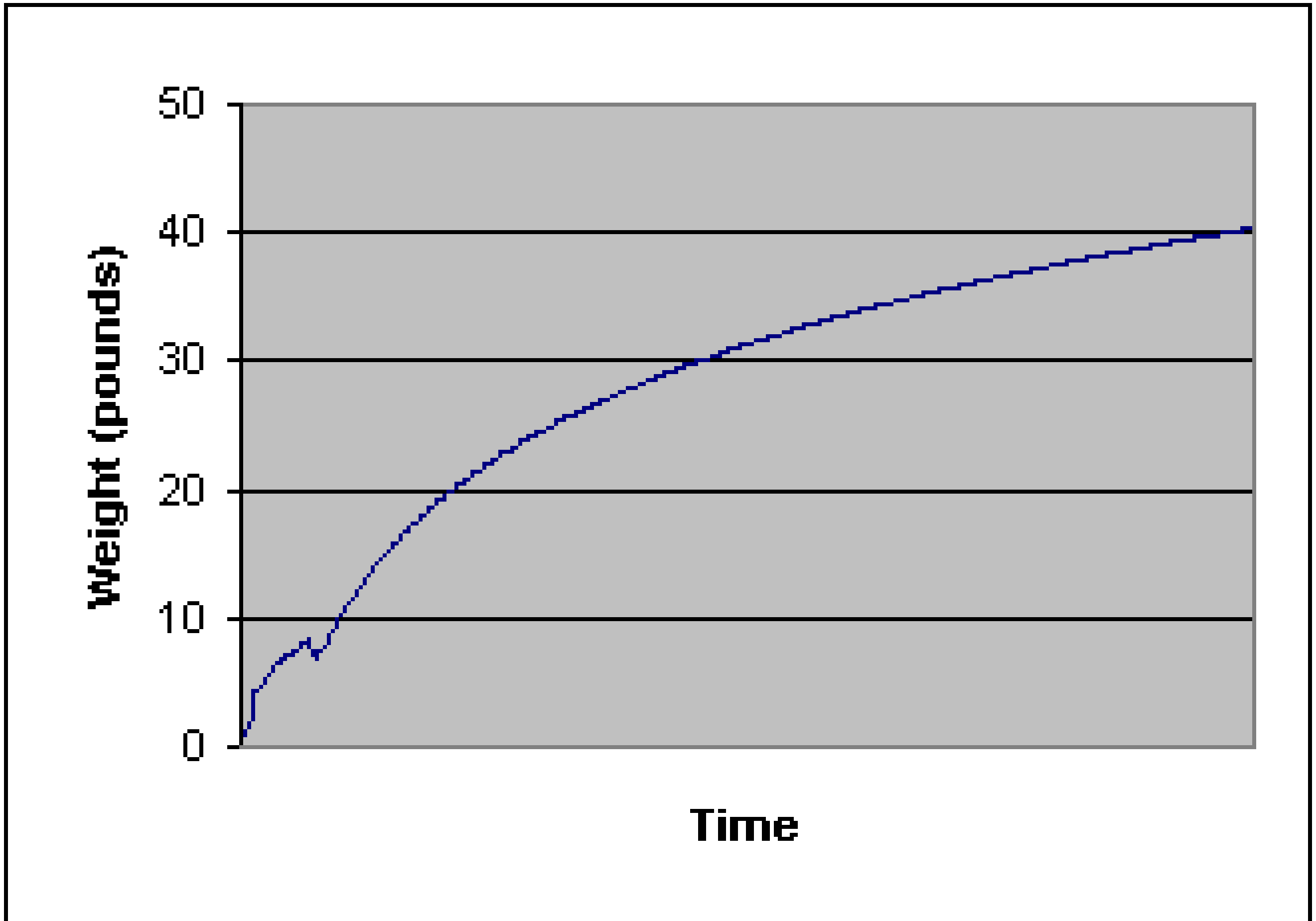
Sampling

- Analogous to weighing and recording the puppy's weight

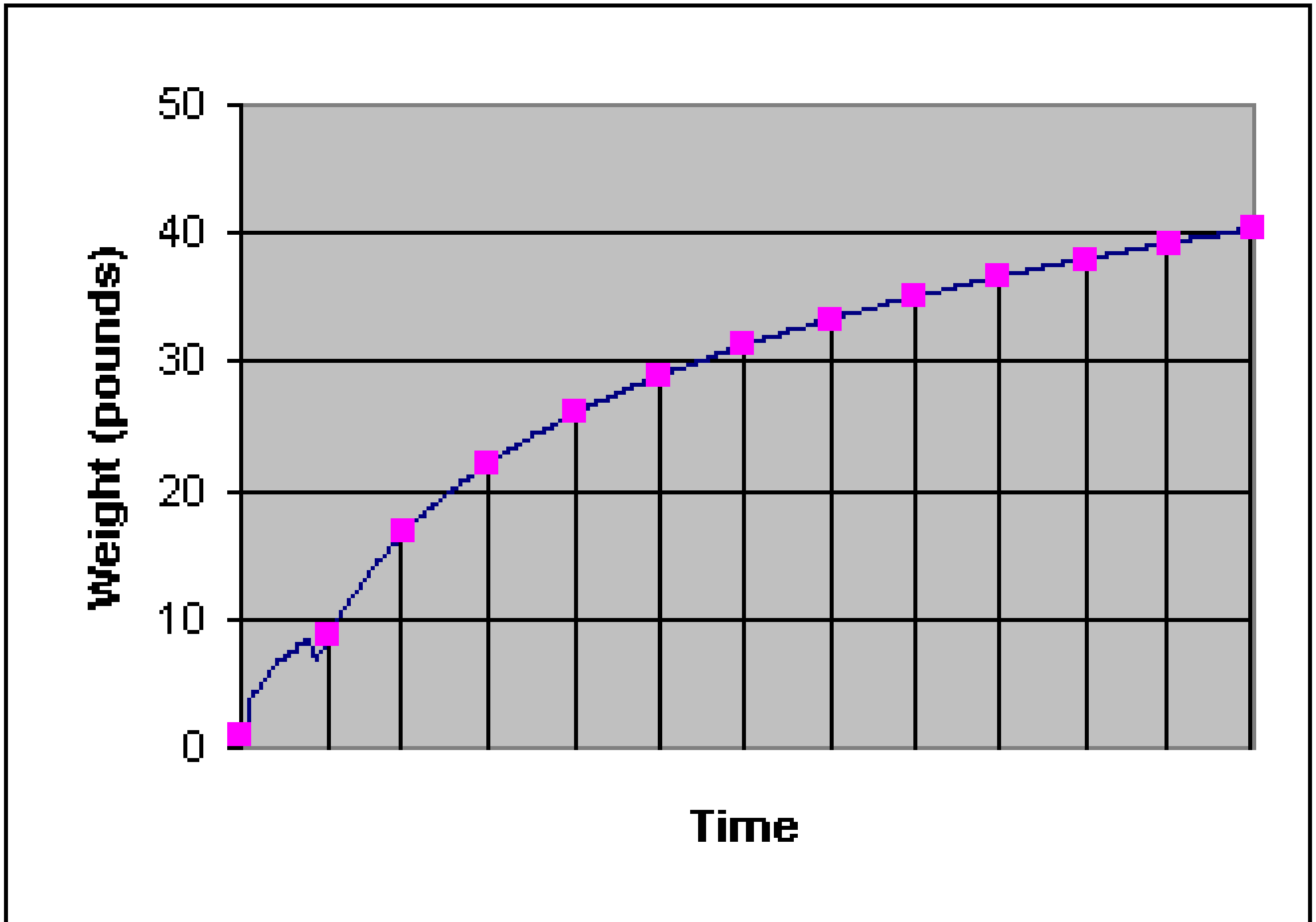
- During the sampling step, you need to set a sampling rate.

- Sampling rate: how often you take a data

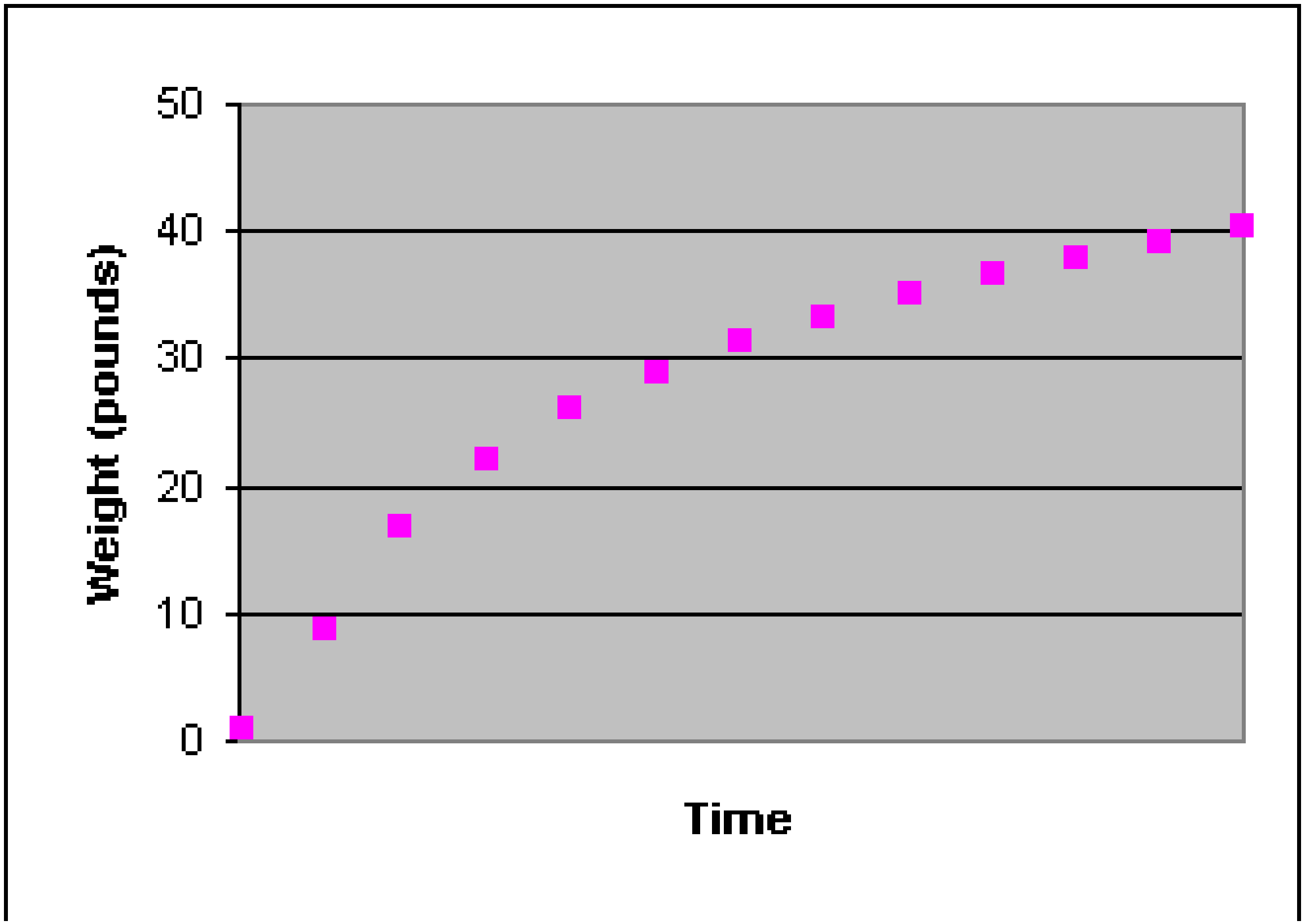
Suppose this is the true timeline of the puppy's first-year growth



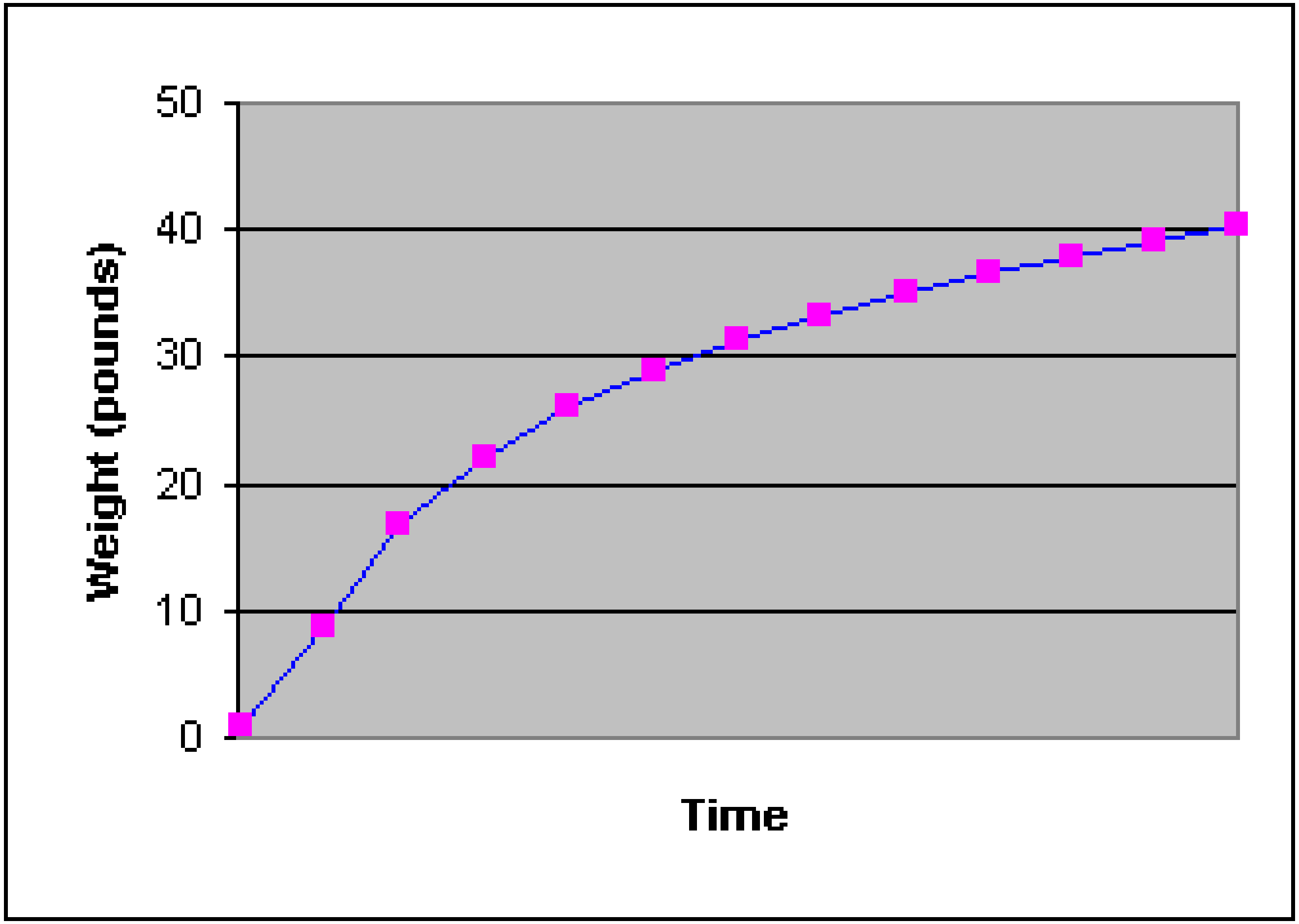
Suppose you weigh the puppy once a month



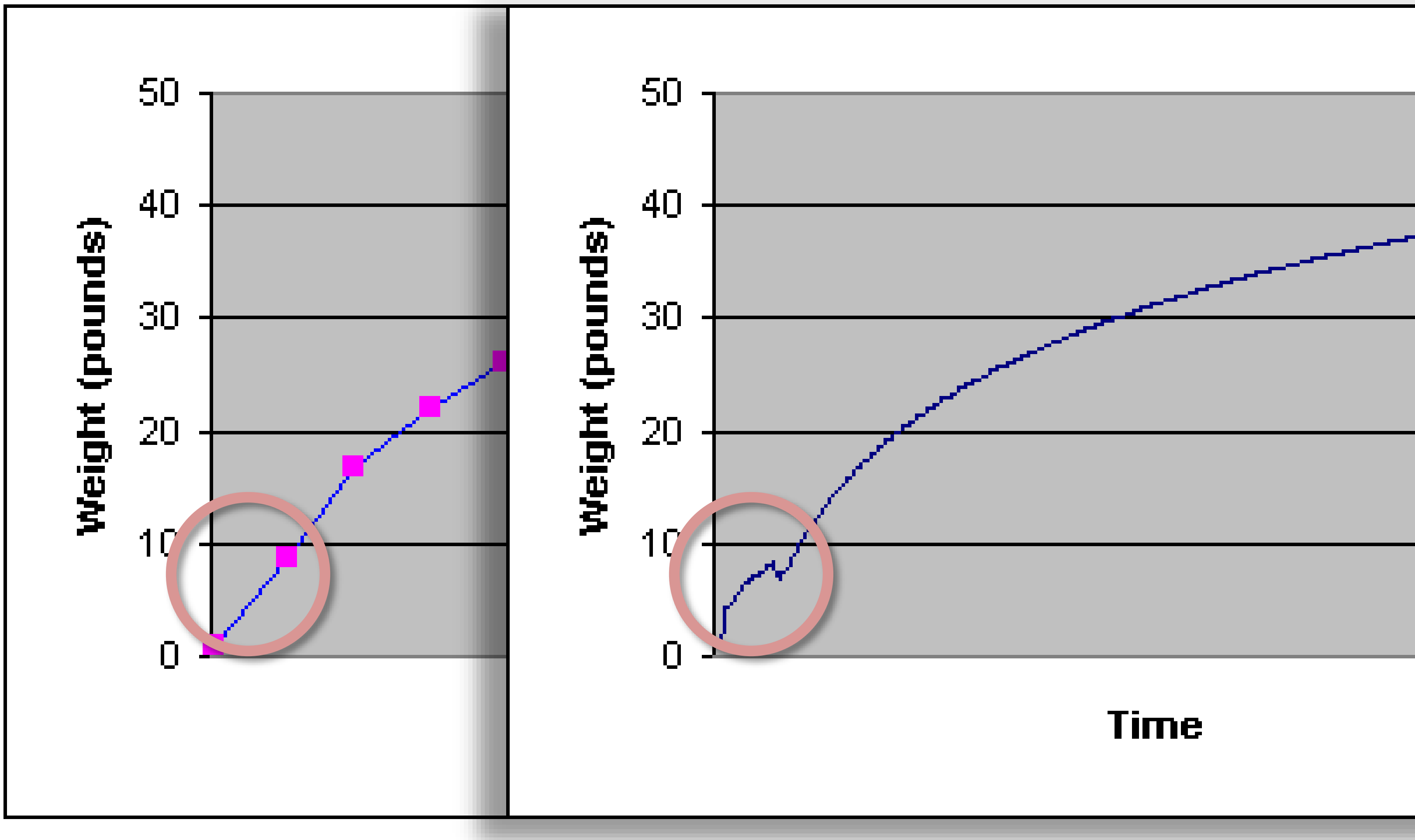
You get these data points



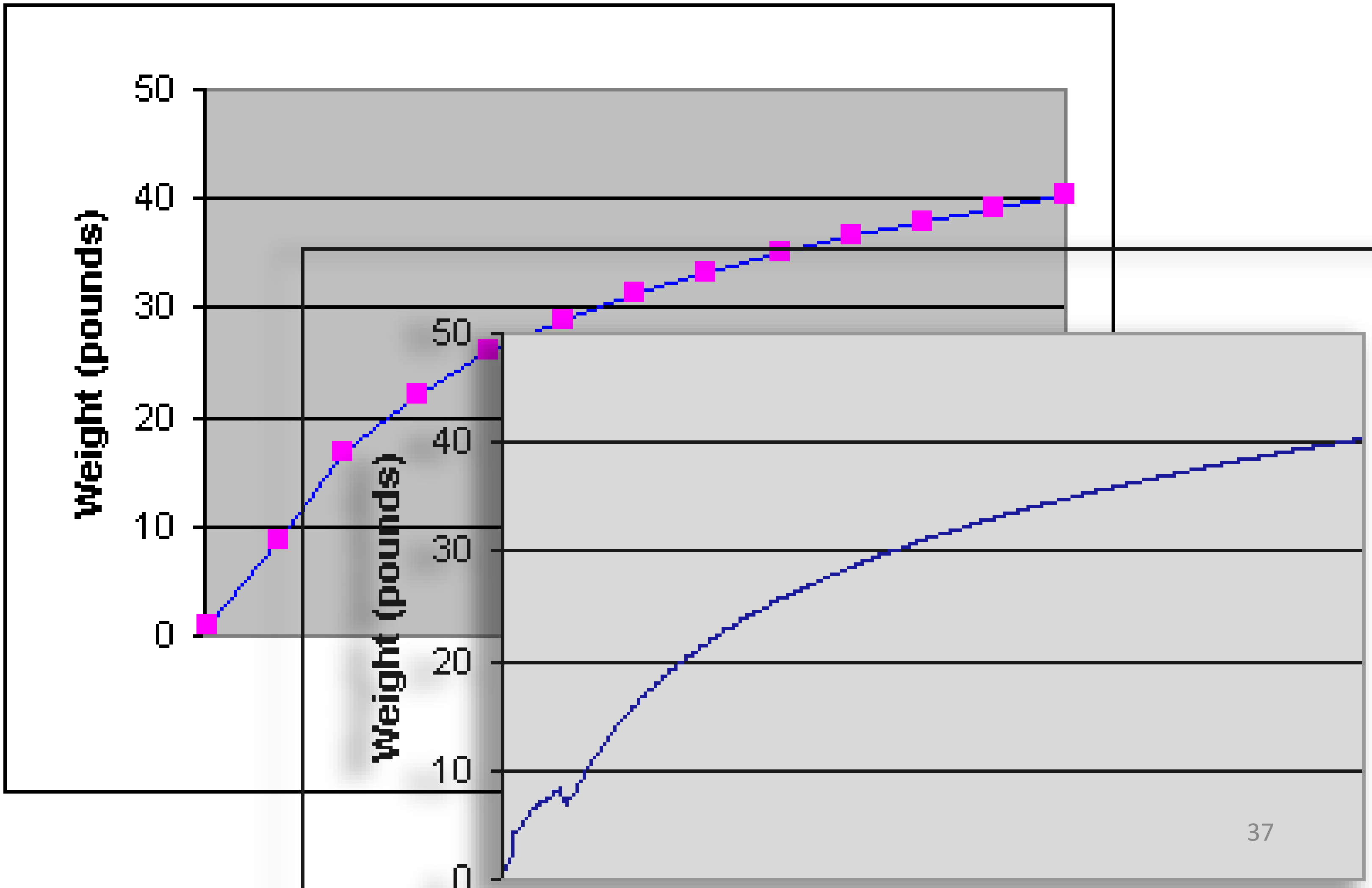
You then interpolate the points



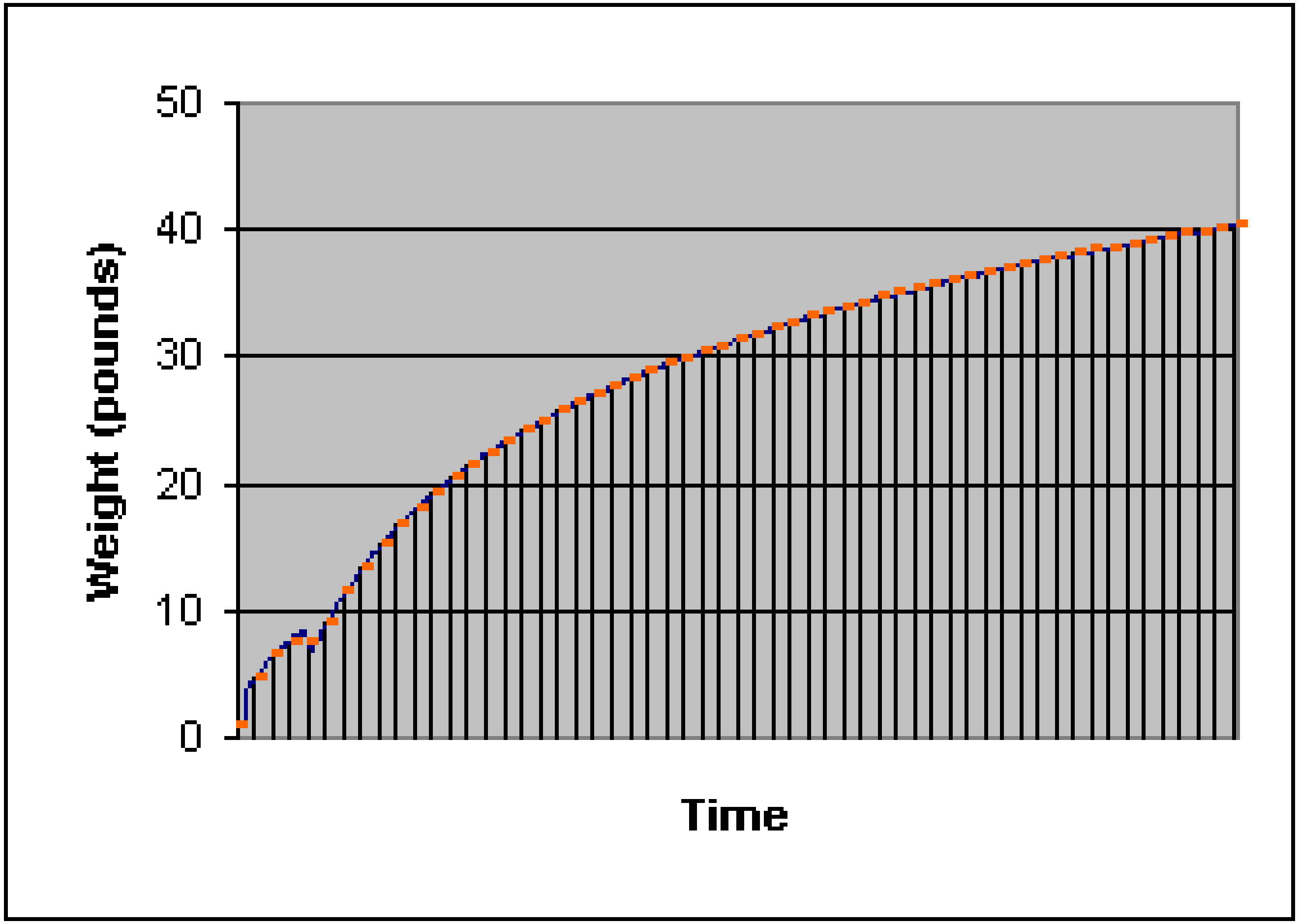
You would miss the changes that occur during the first month



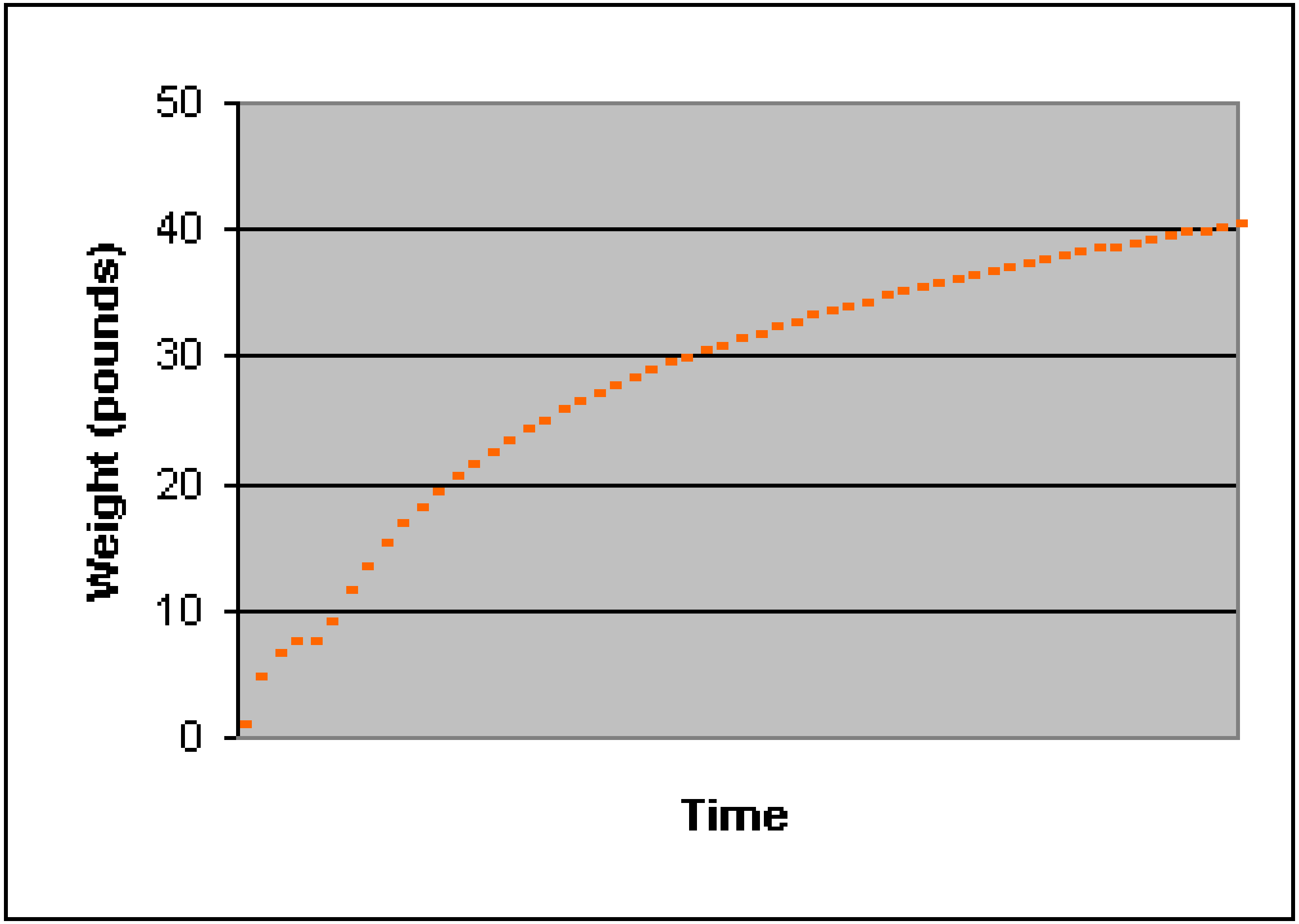
But the rest matches with the true growth pretty well



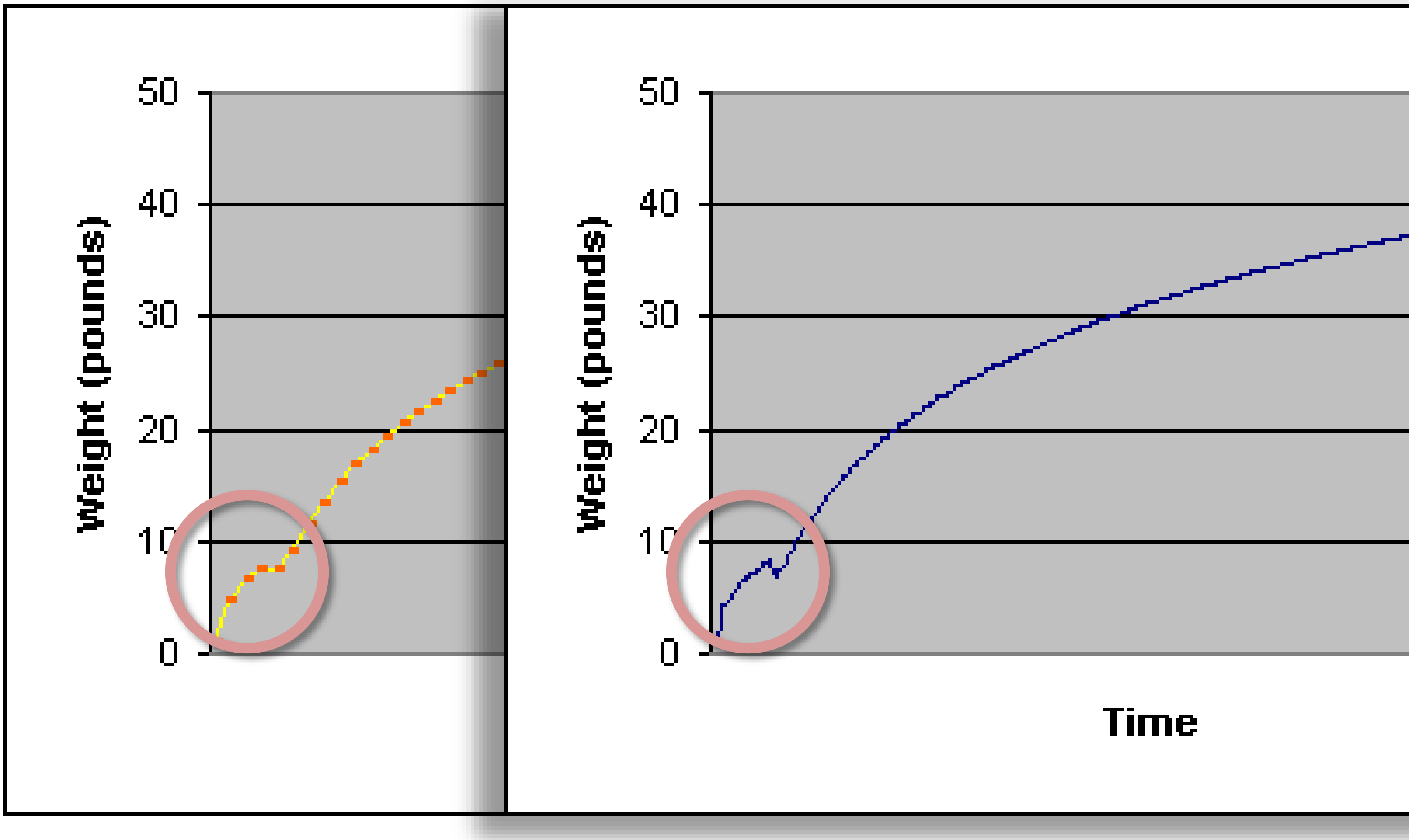
What about weighing the puppy once a week?



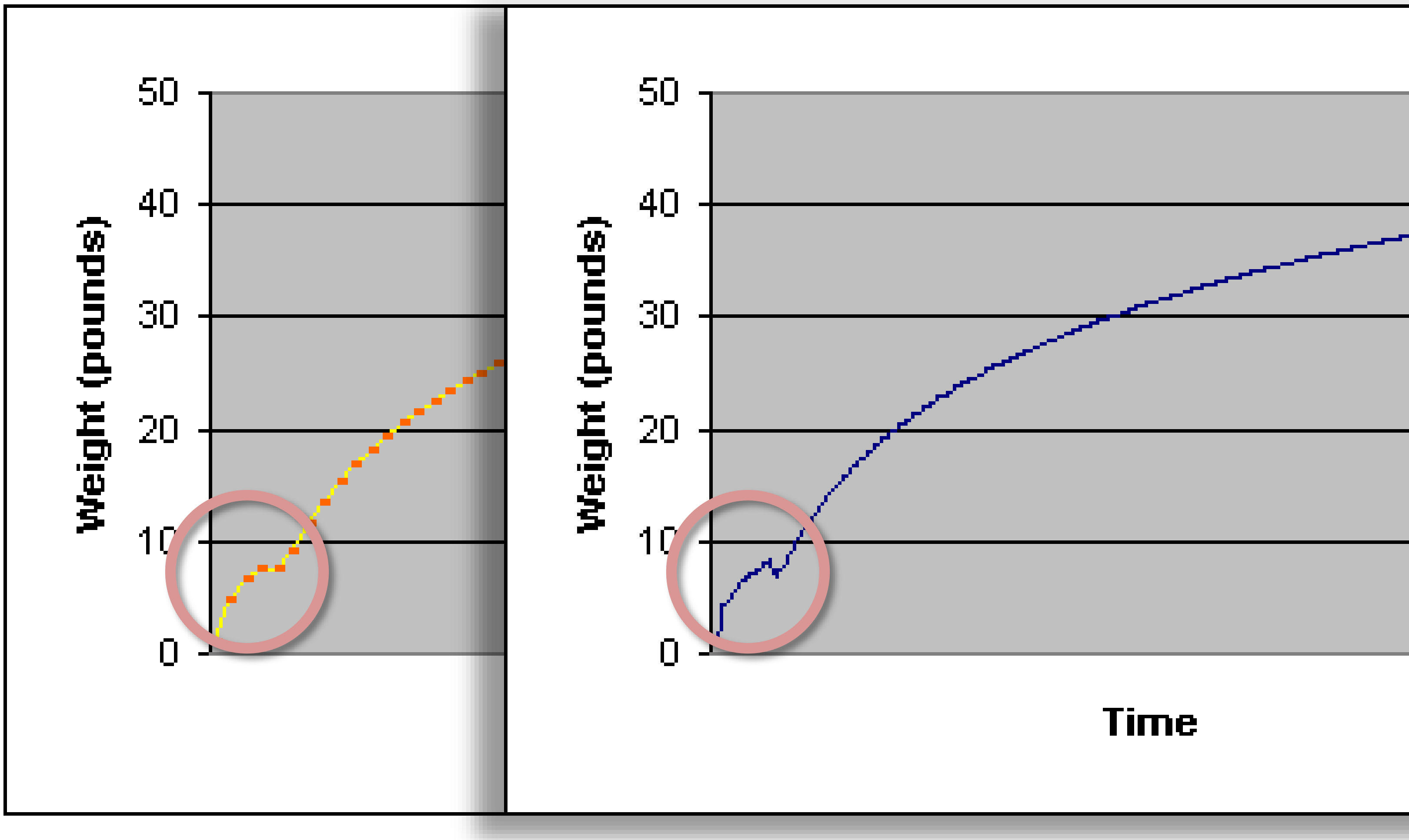
You get these data points



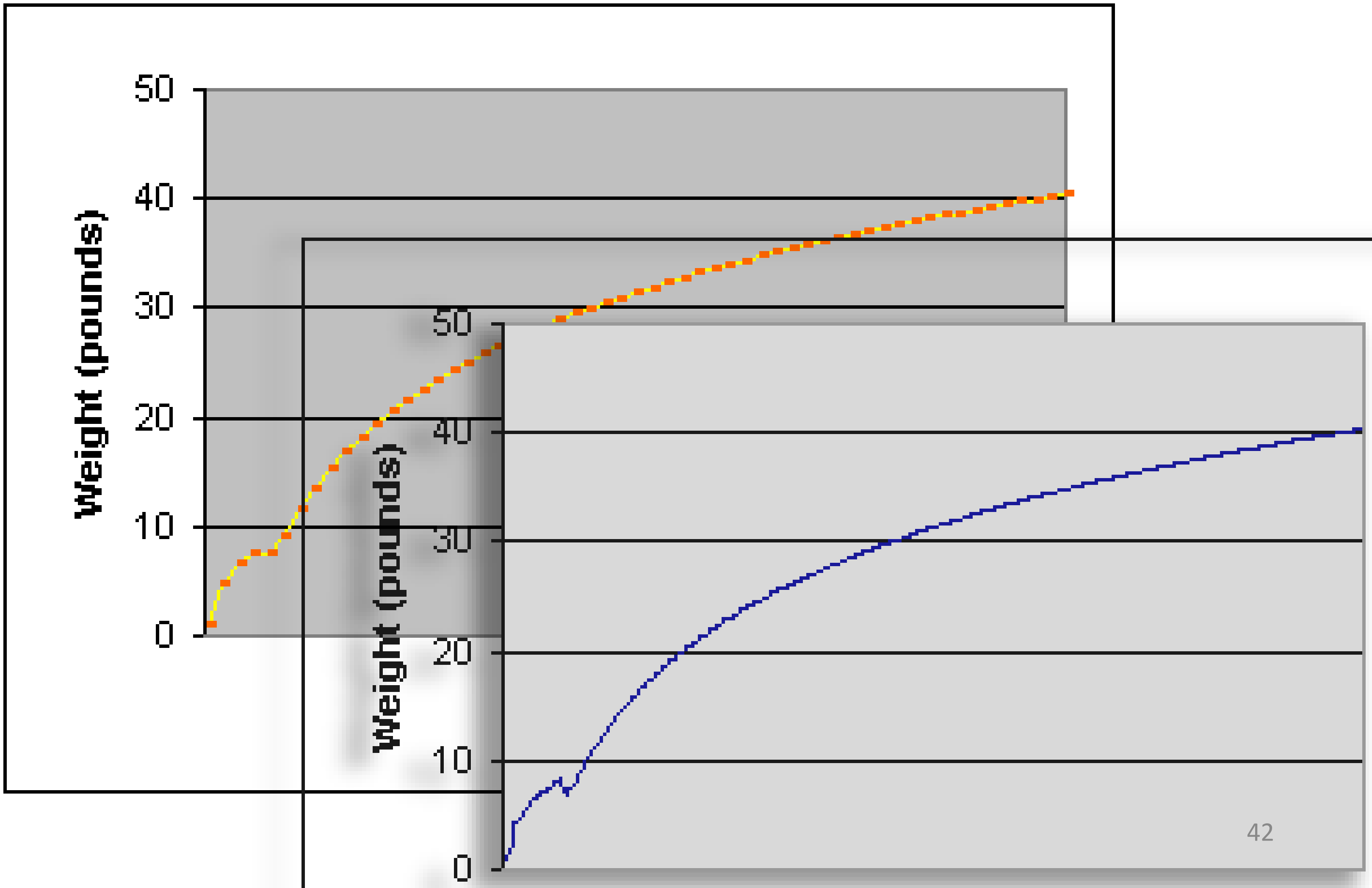
The data is catching the changes occurring in the first month better



But is it exactly?



Now for the rest of the year, the data points seem too many



Sampling Rate

	Weighing Puppy Scenario	Digitization
high (i.e. taking data often)	Pros: can catch more weight changes Cons: produce more paperwork and thus take longer to read through all the data	Pros: can capture details (e.g. some changes of color within a small region in a picture or amplitude changes in sound within a short period of time) Cons: produce larger file and thus take longer to process
low (i.e. taking data infrequently)	Pros: less paperwork and thus take shorter time to read through all the data Cons: may miss weight changes	Pros: produce smaller file and thus take shorter time to process Cons: may miss details (e.g. color changes in a picture or changes in sound)

Quantization

- Analogous to rounding the weight to fix number of digits in the weighing puppy scenario

- During the quantization step, you need to set bit depth.

- Bit depth refers to the number of allowable levels you map (or round) the values to.

Example: 10 levels of weight

For 10 discrete levels, you may have the 10 allowable values as

- 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, and 2.9
- 0, 5, 10, 15, 20, 25, 30, 35, 40, 45
- 2, 4, 6, 8, 10, 12, 14, 16, 18, 20
- 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- ... and so forth

Suppose you choose 2.0, 2.1, ..., 2.9

For 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9

- Any weight data below 2.0 will be recorded as 2.0.

- Any weight data higher than 2.9 will be capped at 2.9.

- It works well if the puppy's weight falls in this range. But it does not seem to be the case.

Suppose you choose 0, 5,..., 45

For 0, 5, 10, 15, 20, 25, 30, 35, 40, 45

- A weight of 2 pounds would be rounded to 0 and a weight of 3 pounds to 5.
- Cons:
  For example, the difference between 2 and 3 pounds is altered after they are mapped to the allowable value on this 10-level scale. The difference becomes 5 pounds not 1 pound.
- Pros:
  Wider range.
- Again, it works well if the puppy's weight falls in this range.

Well, what if we choose this:

2.0, 2.1, 2.2, ..., 44.8, 44.9, 45.0

Suppose you choose 2.0, 2.1,…,44.9, 45.0

- You have increased the number of levels from 10 to 431.

- Pros:
  - Increase precision compared to using
    0, 5, 10, 15, 20, 25, 30, 35, 40, 45
  - Increase range compared to using
    2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9

Increase Number of Allowable Levels

- There does not seem to be any cons in the weighing puppy scenario.

- However, for digitization, increasing the number of allowable levels (i.e. increasing bit depth) will increase the file size.

Sampling and Quantization

Digitizing media involves sampling and quantization regardless of the type of media:

- images
- video
- audio

Overview of how sampling rate and bit depth affect digital media file quality

	Sampling rate is related to:	Bit depth is related to:
digital images	image resolution, or number of pixels	number of allowable colors in an image
digital video	number of pixels in the video, frame rate	number of allowable colors
digital audio	sampling rate of the audio (it limits how high the pitch of the audio can be captured)	number of allowable levels of amplitude

Details will be covered in chapters for each media type.

Review Question

Digital data is ______ and analog information is ______.

A. continuous; discrete

B. discrete; continuous

Review Question

Digitization means converting ______ into ______.

Review Question

Converting from analog to digital involves a two-step process: ______ and ______.

Review Question

When analog information is converted to digital data, two properties affect the exactness of the digital representation, one from sampling and one from quantizing. Which of the following is a result of quantization?

A. sampling rate

B. bit depth

Review Question

Which of the following refers to the number of allowable levels of digitized data?

A. sampling rate

B. bit depth

Review Question

Which of the following can reduce file size of digital media? (More than one choice)

A. decrease sampling rate

B. increase sampling rate

C. decrease bit depth

D. increase big depth

Background

BITS BASIC CONCEPTS

In this lecture, you will find answers to these questions

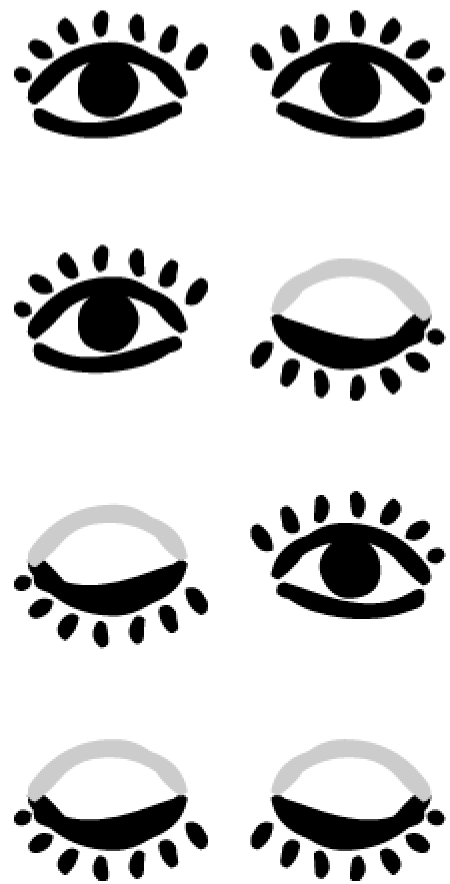
- What are *bits*?
- What do *encode* and *decode* mean?
- What is the significance of the number of bits?

Bits

- In computer systems, data is stored and represented in <u>b</u>inary dig<u>its</u>, called bits.

- To understand how bits can be used to store information, let's use eye signals as an analogy.

Two eyes, Four Combinations of Open and Closed

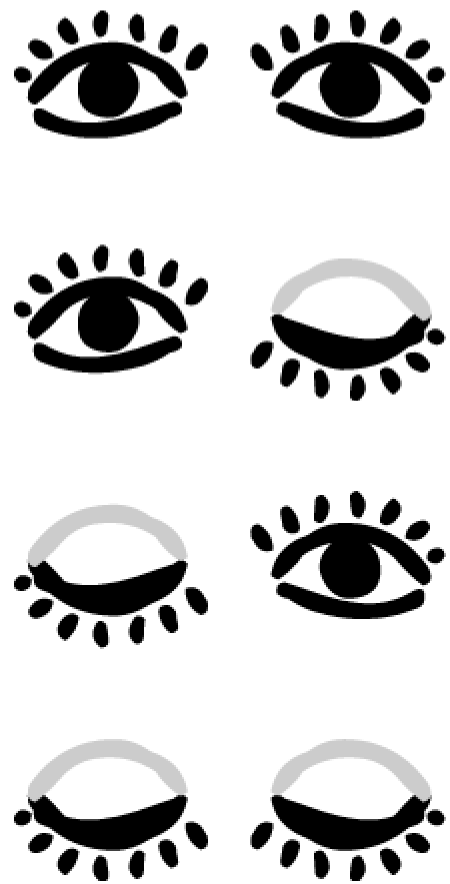


Encoding Your Eye Signals

- To communicate with your friends with your eye signals, you will need to assign meanings (or messages) to the different combinations of open and closed eyes.

- We call this process encoding the message.

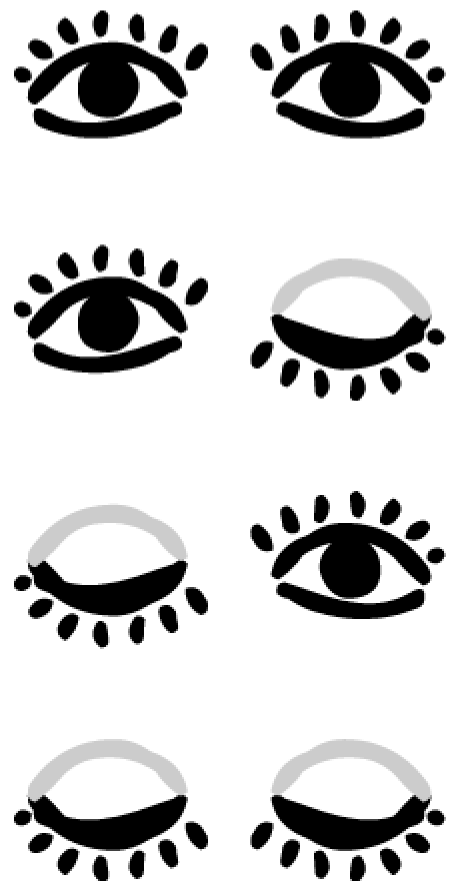
Two eyes, Four Combinations of Open and Closed



You may assign a different meaning to each combination of open and close eyes.

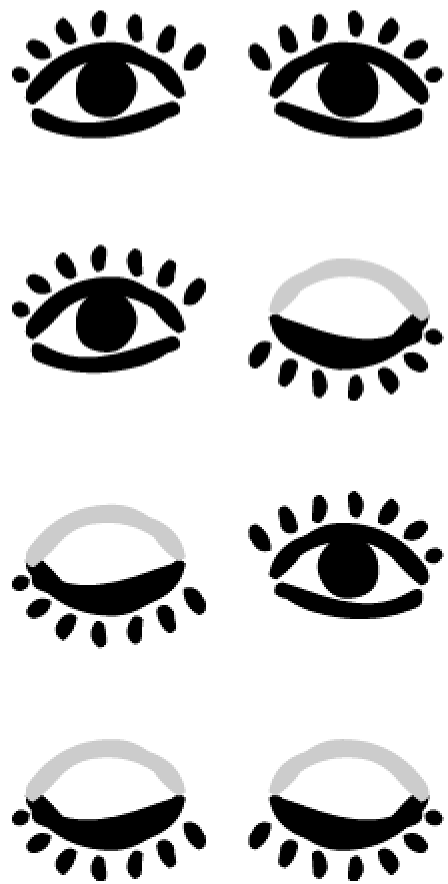
Two eyes, Four Combinations of Open and Closed

You may assign "OMG" to this

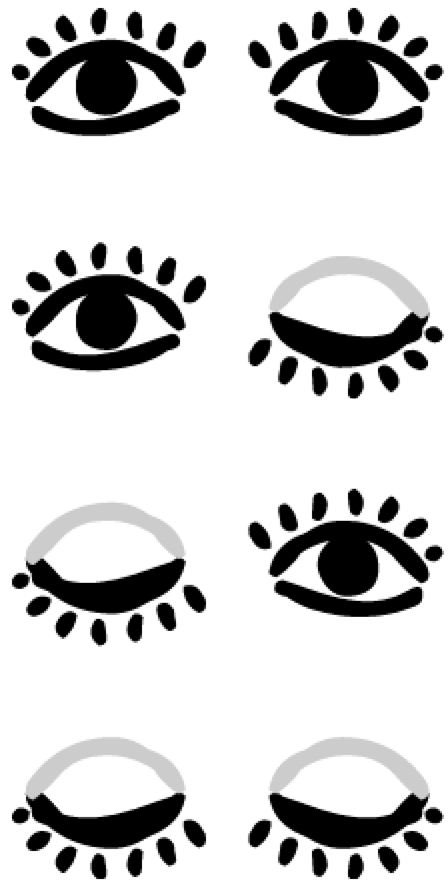


Two eyes, Four Combinations of Open and Closed

You may assign "Yes" to this

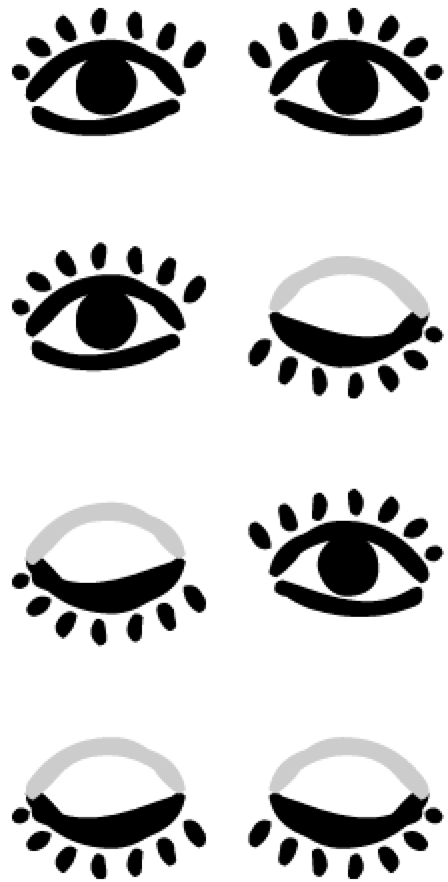


Two eyes, Four Combinations of Open and Closed



You may assign "No" to this

Two eyes, Four Combinations of Open and Closed



You may assign "May be" to this

Two eyes, Four Combinations of Open and Closed

Or, in another situation,

You may assign nothing to this

You may assign "Like" to this

You may assign "Not Like" to this

You may assign nothing to this

Two eyes, Four Combinations of Open and Closed

Or, in yet another situation,

You may assign "red" to this

You may assign "black" to this

You may assign "blue" to this

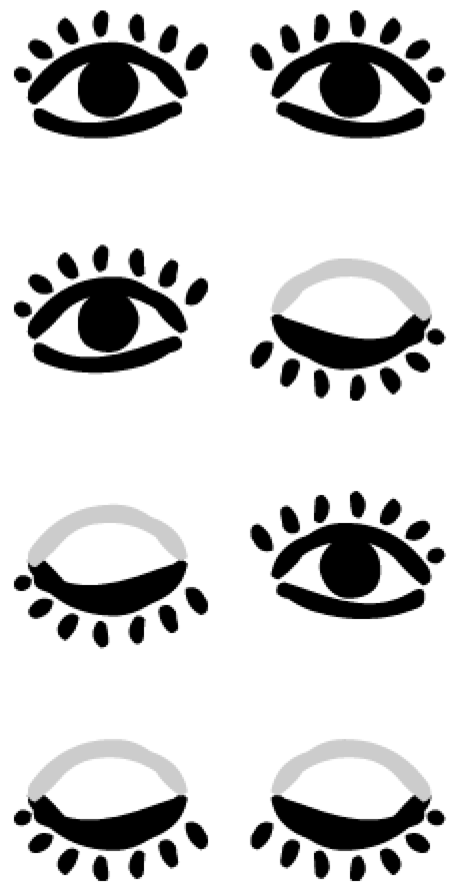
You may assign "yellow" to this

Decode Your Eye Signals

- In order to use your eye signals to communicate with your friends, they will need to know how to interpret your eye signals.
- We call it decoding your eye signals.

Two eyes, Four Combinations of Open and Closed

No matter what messages, with 2 eyes, you can encode no more than 4 different messages.



How many eyes do you need if you have 16 possible colors to signal to your friends?

Hand Signals

- Suppose we consider only two possible poses for each finger: raised up or bent down.

- How many different messages can you encode with 10 fingers?

NOW RETURNING TO BITS

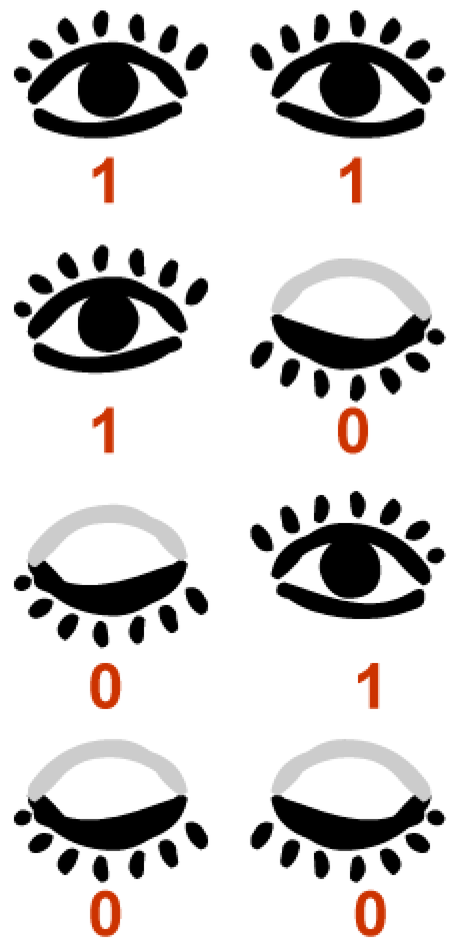
Bits

- In computer systems, data is stored and represented in binary digits, called bits.
- A bit has two possible values, 0 or 1.

Recall Our Eye Signals

Say, we give the open eye a 1 and the closed eye a 0.

We can think of each eye is a bit.



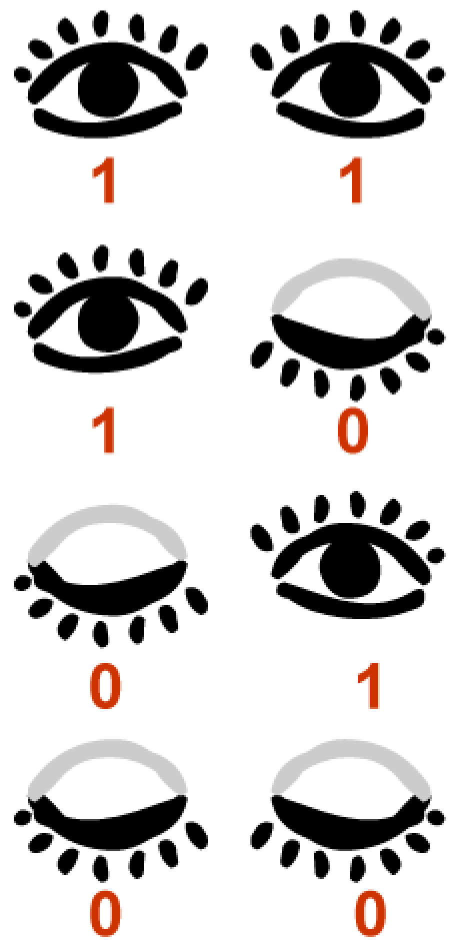
Recall Our Eye Signals

Say, we give the open eye a 1 and the closed eye a 0.

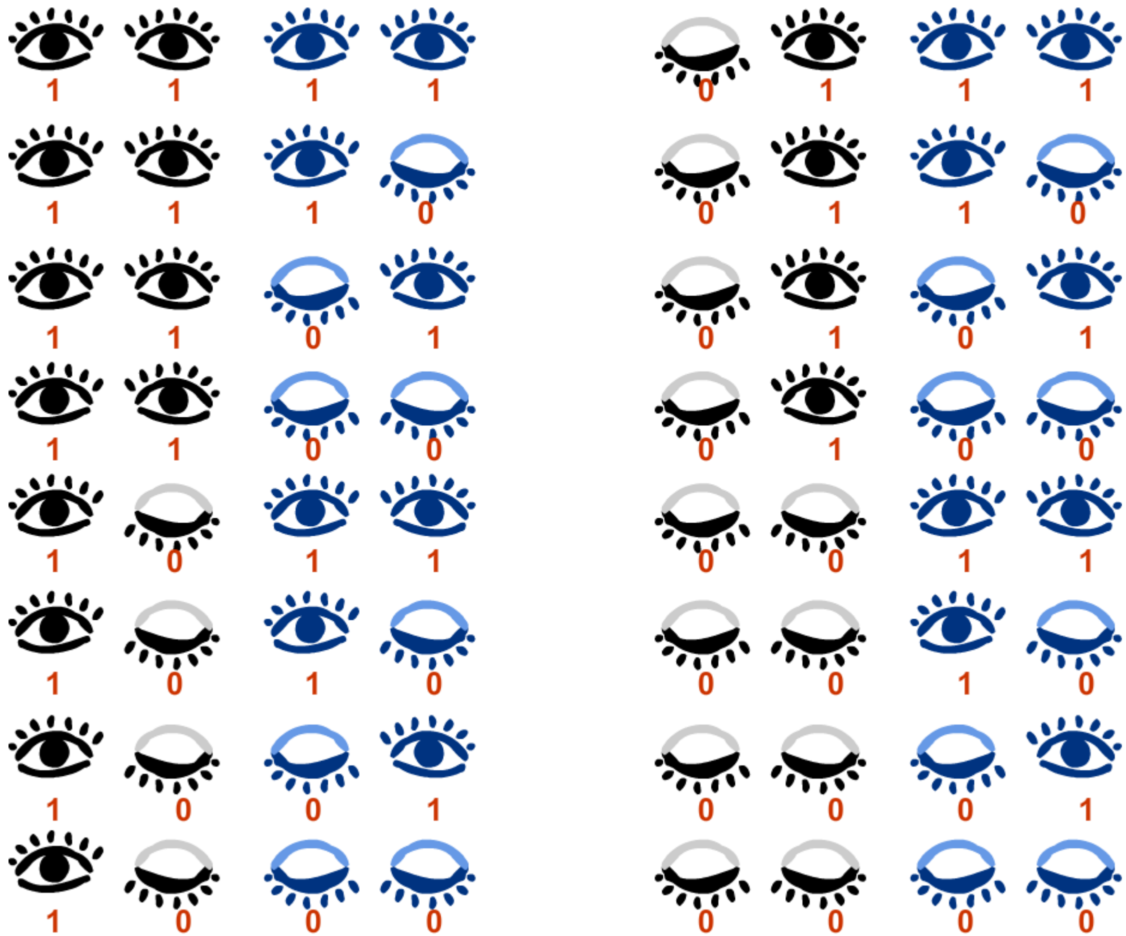
We can think of each eye is a bit.

We can think of our 2-eye signal system is a 2-bit system.

As you see, 2-bit system can encode 4 messages (or choices of things.)



4 bits can encode 16 (2^4) different messages



Number of possible values = $2^{(\text{number of bits})}$

More bits can encode more information.

More bits require more computer storage.

Bytes

1 byte = 8 bits

So Many Bits...

- The number of bits to encode information especially for digital media are very large.

- We use prefixes, such as mega and giga, to better conceive the number of bits and bytes of computer storage.

Prefixes

Prefix Name	Abbreviation	Size
Kilo	K	2^{10} = 1,024
Mega	M	2^{20} = 2,048
Giga	G	2^{30} = 1,073,741,824
Tera	T	2^{40} = 1,099,511,627,776
Peta	P	2^{50} = 1,125,899,906,842,624

Note the size is computed by the exponential of 2.
The exponent is increased in a step of 10,
i.e. 2^{10} , 2^{20} , 2^{30} , 2^{40} , 2^{50} , ...

It is NOT 10^{3} , 10^{6} , 10^{9} , 10^{12} , 10^{15} , ...

Review Question

The word bit comes from the shortening of the words ______ ______.

Review Question

The smallest unit in a binary system is a ______.

A. bit

B. byte

Review Question

A bit has these two possible values: ______ and ______.

Review Question

Eight ______ equals one ______.

A. bytes; bit

B. bits; byte

Review Question

If you want to use hand signals to communicate only two possibilities—like or not like—to your friend, what is the minimum number of finger(s) you need? ______

We can call this hand-signal system ___-bit.

Review Question

If you want to encode only 2 colors, what is the minimum number of bit(s) you need? ______

Review Question

If you want to encode only 8 colors, what is the minimum number of bit(s) you need? ______

Review Question

Most grayscale images use 8-bit color. This means there can be ______ possible different gray tones in the image.

Review Question

(i) How many number of bits in this binary notation?
0011010

(ii) How many possible values can these many bits represent?

Review Question

Which of the following sizes is the largest?

A. 24 GB

B. 24 MB

C. 240 MB

D. 2400 KB

Background

HOW BITS REPRESENT INFORMATION

In this lecture, you will find answers to these questions

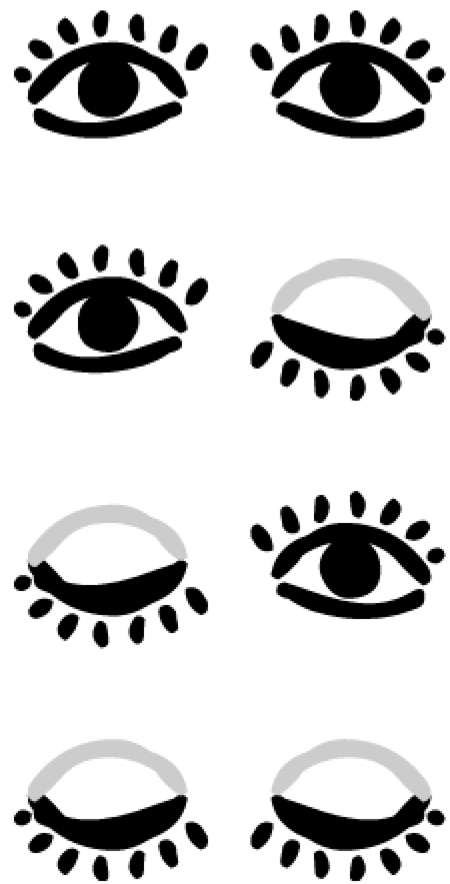
- How do bits represent digital media information?

REVIEW: BITS

Review

- In computer systems, data is stored and represented in binary digits, called bits.
- We use eyes in the eye signals as an analogy.

Two eyes, Four Combinations of Open and Closed

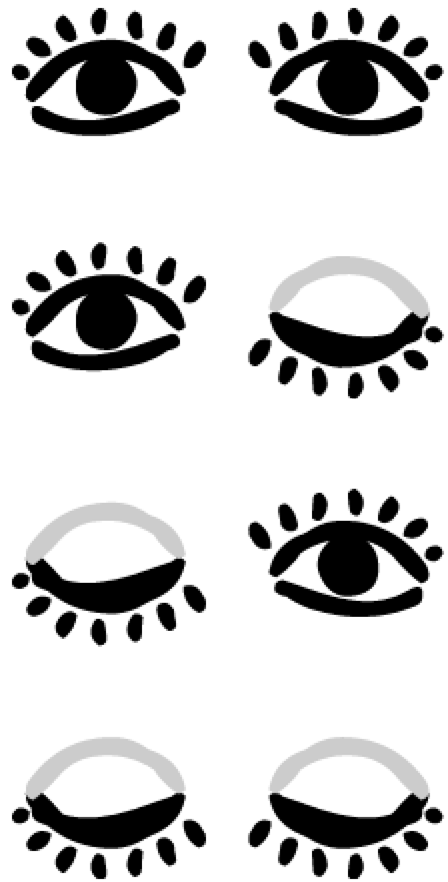


Encoding and Decoding Your Eye Signals

- To communicate with your friends with your eye signals, you will need to assign meanings (or messages) to the different combinations of open and closed eyes--*encoding* the message.

- In order to use your eye signals to communicate with your friends, they will need to know how to interpret your eye signals--*decoding* your eye signals.

Two eyes, Four Combinations of Open and Closed



You may assign a different meaning to each combination of open and close eyes.

Two eyes, Four Combinations of Open and Closed

You may assign "red" to this

You may assign "black" to this

You may assign "blue" to this

You may assign "yellow" to this

Bits

- In computer systems, data is stored and represented in <u>b</u>inary dig<u>its</u>, called bits.
- A bit has two possible values, 0 or 1.

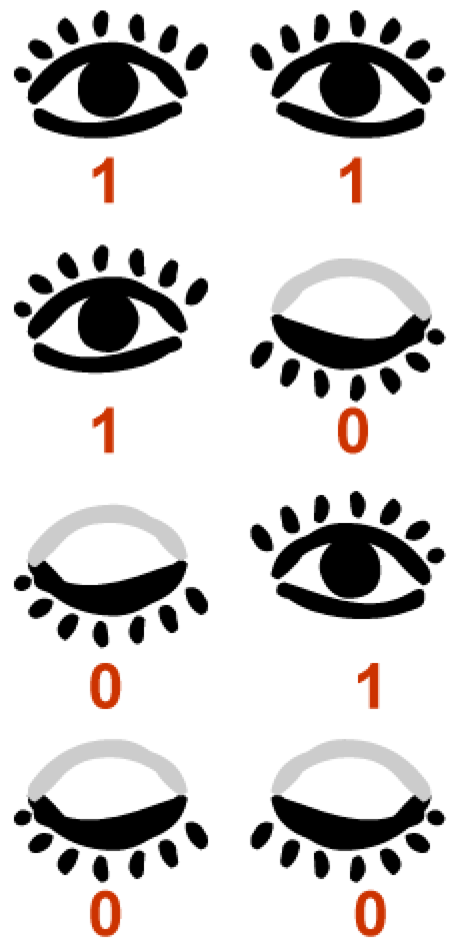
Recall Our Eye Signals

Say, we give the open eye a 1 and the closed eye a 0.

We can think of each eye is a bit.

We can think of our 2-eye signal system is a 2-bit system.

As you see, 2-bit system can encode 4 messages (or choices of things.)



Number of possible values = $2^{(\text{number of bits})}$

More bits can encode more information.

More bits require more computer storage.

DONE REVIEW: BITS

Using bits to represent numeric values

Base-10 and Base-2 Conversion

Decimal Notation
Base-10

- Commonly used in our daily life

- Use combinations of 10 different numerals to construct any values

- The 10 different numerals are:

  0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Base-10 Example

The decimal number 5872 is interpreted as follows.

$$
\begin{array}{rcccc}
 & 5 & 0 & 0 & 0 \\
 &   & 8 & 0 & 0 \\
 &   &   & 7 & 0 \\
+ & &   &   & 2 \\
\hline
 & 5 & 8 & 7 & 2
\end{array}
$$

Base-10 Example

In other words,

$$5\ 8\ 7\ 2$$

$$= 5 \times 10^3 + 8 \times 10^2 + 7 \times 10^1 + 2 \times 10^0$$

$$= 5 \times 1000 + 8 \times 100 + 7 \times 10 + 2 \times 1$$

$$= 5000 + 800 + 70 + 2$$

$$= 5\ 8\ 7\ 2$$

Binary Notation
Base-2

- Used in machine language (language that computers understand)

- Use combinations of 2 different numerals to construct any values

- The 2different numerals are:

  0, 1

Base-2 Example

The binary notation 1011 is interpreted as follows.

1 0 1 1

$$= 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

$$= 1 \times 8 + 0 \times 4 + 1 \times 2 + 1 \times 1$$

$$= 8 + 0 + 2 + 1$$

$= 11$ (eleven, in decimal notation)

Base-2 to Base-10

- The previous slide shows the base-2 to base-10 conversion method.
- 1101_2 (one one zero one) represents 11_{10} (eleven).
- The subscript indicates the base.

Base-10 to Base-2

To convert base-10 to base-2 notation:

1. repeatedly divide the decimal number by 2 until it becomes 0, noting the remainder of each division.

2. The reverse order of the sequence of the remainders is the binary representation of the decimal number.

Base-10 to Base-2 Example

To convert 19_{10} to binary notation:

19 / 2 = 9 remainder 1

9 / 2 = 4 remainder 1

4 / 2 = 2 remainder 0

2 / 2 = 1 remainder 0

1 / 2 = 0 remainder 1

Base-10 to Base-2 Example

To convert 19_{10} to binary notation:

19 / 2 = 9 remainder 1

9 / 2 = 4 remainder 1

4 / 2 = 2 remainder 0

2 / 2 = 1 remainder 0

1 / 2 = 0 remainder 1

10011_2

Examples

Decimal	Binary
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111

Using bits to represent text

By assigning unique numbers to each text character

For example, the character A is represented by 65.

TABLE 1.3 The Lower 128 ASCII Codes

0	NUL	16	DLE	32		48	0	64	@	80	P	96	`	112	p
1	SOH	17	DC1	33	!	49	1	65	A	81	Q	97	a	113	q
2	STX	18	DC2	34	“	50	2	66	B	82	R	98	b	114	r
3	ETX	19	DC3	35	#	51	3	67	C	83	S	99	c	115	s
4	EOT	20	DC4	36	$	52	4	68	D	84	T	100	d	116	t
5	ENQ	21	NAK	37	%	53	5	69	E	85	U	101	e	117	u
6	ACK	22	SYN	38	&	54	6	70	F	86	V	102	f	118	v
7	BEL	23	ETB	39	’	55	7	71	G	87	W	103	g	119	w
8	BS	24	CAN	40	(	56	8	72	H	88	X	104	h	120	x
9	TAB	25	EM	41	)	57	9	73	I	89	Y	105	i	121	y
10	LF	26	SUB	42	*	58	:	74	J	90	Z	106	j	122	z
11	VT	27	ESC	43	+	59	;	75	K	91	[	107	k	123	{
12	FF	28	FS	44	,	60	<	76	L	92	\	108	l	124	|
13	CR	29	GS	45	-	61	=	77	M	93	]	109	m	125	}
14	SO	30	RS	46	.	62	>	78	N	94	^	110	n	126	~
15	SI	31	US	47	/	63	?	79	O	95	_	111	o	127	DEL

ASCII

- stands for American Standard Code for Information Interchange

- an encoding standard for text characters, including the 26-letter English alphabets and symbols in computer programs.

ASCII

- For ASCII character set, each character uses 8 bits.

- With 8 bits, you can encode $2^8 = 256$ different characters.

Unicode

- another standard for encoding text character
- can represent a large repertoire of multilingual characters
- use more than 8 bits to encode a text character because multilingual character sets are larger than the ASCII set

Using bits to represent images

Using bits to represent images

- Bitmap images, such as digital photos
  – color value of each pixel encoded into bits

- Vector graphics, such as graphics created in Flash
  – coordinates of anchor points encoded into bits
  – tangent of each anchor points encoded into bits

- Bitmap images, vector graphics, and pixels will be explained in the digital images chapters

Using bits to represent sound

Using bits to represent sound

- sampled audio
  - amplitude for each sample encoded into bits
    For CD quality audio, it has 44,100 samples per second of the audio
- MIDI music
  - each musical instrument has an ID which can be encoded into bits
  - each musical note has an ID which can be encoded into bits
- Sampled audio and MIDI will be explained in the audio chapters

Using bits to represent program instructions

By using a sequence of bits to represent an operation

Using bits to represent program instructions

For example, an arithmetic addition may be represented by a sequence of bits: 001100

Reference:

- Wong, Y.L. (2013), Digital Media Primer, Pearson Education, Chapter 1
- https://web.stanford.edu/class/cs101/bits-bytes.html