Unit 2 Lesson 1

Bytes and File Sizes

Resources

Name(s)_____ Period _____ Date _____

Activity Guide - Bytes and File Sizes

What is a byte? A byte is a unit of data that is 8 bits long. A byte is the standard "chunk size" for binary information in most modern computers

Larger Chunks of Data: On modern computers the amount of information we can create and store has grown so large that we need new units of measurement to describe the size of our data. Use these websites for your research.

● **Stanford University - CS 101 - Kilobytes Megabytes Gigabytes**
● **Computer Hope - How much is 1 byte, kilobyte, megabyte, gigabyte, etc.?**
   http://www.computerhope.com/issues/chspace.htm

Unit	Number of Bytes (approx)	Example of File Type or Data Measured in this Unit
Kilobyte (KB)		
Megabyte (MB)		
Gigabyte (GB)		
Terabyte (TB)		
Petabyte (PB)		
Exabyte (EB)		

How big are the files I use every day? Try to determine the size of files you probably use every day. You can either research these answers online or check the size of actual files on your computer.
   ● PC/WINDOWS: Right-click and choose "Properties"
   ● MAC: Ctrl+click and choose "Get Info"

File type	Size as # of pages, minutes, seconds, or dimensions	Size of file in Bytes, KB, MB, GB, etc.
page of plain text (.txt)	About 500 words, or 2500 characters	2500 Bytes, 2.5KB
.jpg image		
animated .gif image		
.pdf file		
Audio file as .mp3		
movie file such as .mov or .mp4		

Test your knowledge!

The first 3 questions here are from:
You can check the answers there.

1. Alice has 600 MB of data. Bob has 2000 MB of data. Will it all fit on Alice's 4 GB thumb drive?

2. Alice has 100 small images, each of which is 500 KB. How much space do they take up overall in MB?

3. Your ghost hunting group is recording the sound inside a haunted classroom for 20 hours as MP3 audio files. About how much data will that be, expressed in GB?

Here are a few more.

1. A salesperson is trying to sell you a phone that has 16 GB of memory saying, "that's enough space to record an hour of high quality video!" This salesperson is probably wrong, but in which direction? Would you have more than enough memory or not enough?

2. Shakespeare's complete works have approximately 3.5 million characters. Which is bigger in file size: Shakespeare's complete works stored in plain ASCII text or a 4 minute song on mp3? How much bigger?

3. **Tricky**: Assume your Internet connection can transmit 1 million **bits** per second. Approximately how long would it take you to download 1 Terabyte of data? (Hint: first figure out how many bits a terabyte is, second be prepared to wait a long time).

Unit 2 Lesson 2

Text Compression

Resources

Name(s)_____ Period _____ Date _____

Activity Guide: Decode this message!

What's the original message?

Below is an encoded message. It's not necessarily a secret message but it does need to be decoded.
Study the clues and key to reconstruct the original message.

Encoded Message:

★listen_to☀rain_★on☀window_pane

Key:

☀	_the_
🌂	tter_
☃	Pi🌂
❄	Pa🌂
★	☃❄☃❄

Original Message:

_____

_____

_____

_____

_____

_____

Name(s)_____ Period _____ Date _____

Activity Recap: Decode this message!

Here is a breakdown of how much the "Pitter Patter" message was compressed.

How Much?   40% compressed!

Original:	93 characters
Compressed:	56 characters
Difference:	**37 characters (~40%)**

Original Message   93 characters

Pitter_patter_pitter_patter_listen_to_the_
rain_pitter_patter_pitter_patter_on_the_
window_pane

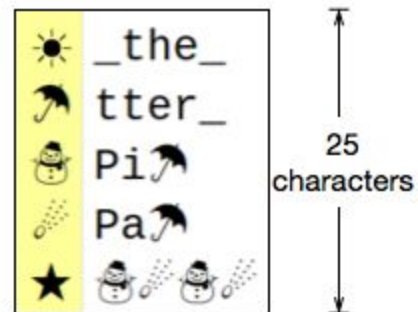**Compressed**   56 characters

★listen_to☀rain_★on☀window_pane

|← ————— 31 characters ————— →|

☀	_the_
🌂	tter_
⛄	Pi🌂
🌧	Pa🌂
★	⛄🌧⛄🌧

25 characters

Total number of characters needed to represent compressed version is:

31 (message) + 25 (key) = 56 characters

Name(s)_____ Period _____ Date _____

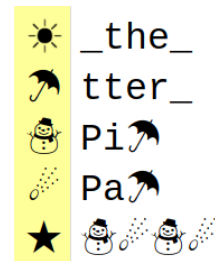## Activity Guide - Text Compression

Objectives
- Compress a piece of text using the Text Compression Widget
- Explain the factors that make compression challenging.
- Explain why the "best" compression is impossible or "hard" to identify.
- Create your own heuristic for compressing data.

Text Compression Tool
If you have not already done so, open up the Text Compression Tool in Code Studio and watch the video explaining how to use it. Then choose one poem and build a simple dictionary using the table below.

Tips:
- Look for patterns that repeat and enter each in the "dictionary"
- Look for patterns of patterns - the dictionary can refer to itself (see right) making for a powerful amount of compression!
- Compare with others compressing the same text
- Try to develop a general strategy that will lead to "good" compression,

What's your best?
Copy and paste the best compression you made. (If you have a digital copy of this activity guide you could copy/paste a screenshot here, or just copy paste the text from the tool itself.)

Poem Name: _____

Compressed Text:

Dictionary:

Compression Stats:

Reflection
Respond to these prompts.
1. What made compressing text hard to do?

2. Describe the thinking process you used in solving this challenge - what was your strategy for compressing the text. Could you explain it to someone who had never done this before?

Develop a Heuristic

Continue working on compressing your poem. As you do so, develop a set of rules, or a "heuristic" that generally seems to provide good results. Record the steps of your heuristic in some way that will allow you to exchange with another group. Make sure your rules are as clear as possible, so the other group will always know what to do.

Exchange Heuristics

Trade your heuristics with another group. Are they clear and specific enough that you always know what to do? If not, provide feedback to one another and improve your heuristics to provide clearer instructions.

Test Heuristics

You should now have another group's heuristic. Using the heuristic, attempt to compress the poems below. Record the compression rates you achieve.

Poem	Compression Rate
I Need a Dollar	
Pitter Patter	
A Tutor	
She Sells Sea Shells	
I Know an Old Lady	
Pease Porridge	

Record Your Conclusions

Use the data you collected to respond to the questions below.

1. Do you think it's possible to describe (or write) a specific set of instructions that a person could follow that would always result in better text compression than your heuristic? Why or why not?

2. Is there a way to know that a compressed piece of text is compressed the most possible? If yes, describe how you could determine it. If no, why not?

3. If you send the compressed poem would your friend be able to read it? Why is the dictionary important?

Vocabulary

Compress: to decrease the number of bits used to represent a piece of information
Algorithm: a precise sequence of instructions designed to complete a task
Heuristic: a specific type of algorithm, usually used when exact solutions are difficult or impossible. Heuristics are generally simple to use and are designed to provide reasonably good results without guaranteeing a perfect solution.

Name(s)_____ Period _____ Date _____

Activity Guide - Text Compression Heuristics

This meant as a follow-on to the text compression activity.

Vocabulary
Heuristic: a problem solving approach (typically an algorithm) to find a satisfactory solution where finding an optimal or exact solution is impractical or impossible.

Develop a Heuristic
Continue working on compressing your poem using the Text Compression Widget. As you do so, develop a set of rules, or a "heuristic" that generally seems to provide good results.
- **Record your heuristic** as a list of steps that someone else unfamiliar with the problem could follow and still end up with decent compression.
- **Trade your heuristics with another group.** Are they clear and specific enough that you always know what to do? If not, provide feedback to one another and improve your heuristics to provide clearer instructions.
- Using another group's heuristic, attempt to compress one or more of the poems in the tool. Record the amount of compression you achieve.

| Poem	Compression Rate

Record Your Conclusions
1. Do you think it's possible to describe (or write) a specific set of instructions that a person could follow that would always result in better text compression than your heuristic? Why or why not?

2. Is there a way to know that a compressed piece of text is compressed the most possible? If yes, describe how you could determine it. If no, why not?

Unit 2 Lesson 3

Encoding B&W Images

### Resources

Name(s)_____ Period _____ Date _____

Activity Guide - B&W Pixelation Widget
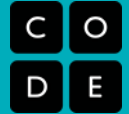
The B&W Pixelation Widget
The pixelation widget uses a file format as depicted below. For example, the 3x5 image of the letter "A," shown at right within the Pixelation tool, would be encoded as a simple stream of these bits, organized like this (color added for emphasis):

00000011 00000101 0000000110101000010010

width   height   pixel data
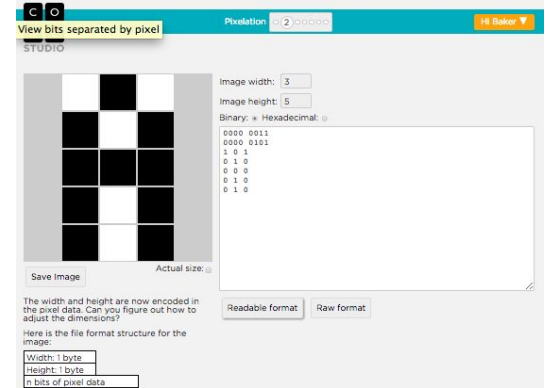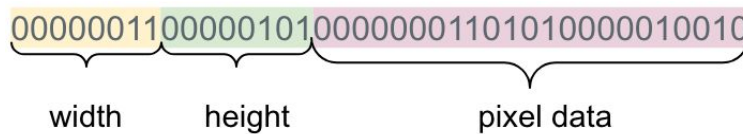
We can break it up into pieces like so:

purpose	size	example
width	1 byte	0000 0011
height	1 byte	0000 0101
pixel data	varies	00000001101010000010010

B&W Image File Format

metadata	Bits 0-7 (1 byte) = width
	Bits 8-15 (1 byte) = height

pixel data	Bits 16 - n = pixel data
	0 = black (light off) 1 = white (light on)

Student Tasks

Log into Code Studio to get started with the pixelation widget. In code studio there are a few short tasks to help you get acquainted with the tool.

- **Create a small image:** Start by trying to recreate the 3x5 letter "A" depicted (shown above) using the pixelation widget.

- **Correct the error:** Oh no! An extra bit was inserted into an image during transmission! Track it down.

- **Make your own image of any size.**
    ○ Encode an image of anything you like.
    ○ Do not simply make an abstract pattern, like a checkerboard. It should depict something, perhaps your name written out, an icon or logo of some sort.
    ○ Optional: For fun, send your image bits to a friend using the Internet Simulator

When you are finished, complete the table below and answer the following questions:

Your image: Save your image and drag-and-drop or copy/paste it below.	Your bits: Copy/paste the bits that create your image below.

Questions

1. Using the B&W file format from the pixelation widget

   ○ What are the largest dimensions (width and height) of an image we can make with the pixelation widget?

   ○ How many total bits would there be in the <u>largest</u> possible image we could make with the pixelation widget?

   ○ How many bits would it take to represent the <u>smallest</u> possible image (i.e. an image with one pixel)?

2. What would happen if we didn't include width and height bits in our protocol? Assume your friend just sent you 32 bits of pixel data (just the 0s and 1s for black and white pixels). Could you recover the original image? If so, how?
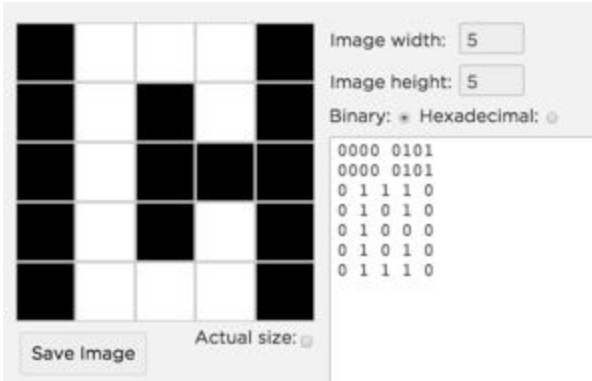
Name(s)_____ Period _____ Date _____

Activity Guide - Extension: Magnify an Image

Directions

Add bits to magnify the image below (a small letter "C") by a factor of 2. The image should look the same, but twice as big! The bits are provided and a screen capture of the resulting image is shown.

Before	After
00000101000001010111001010010000101001110 Image width: 5 Image height: 5 Binary: ● Hexadecimal: ○ 0000 0101 0000 0101 0 1 1 1 0 0 1 0 1 0 0 1 0 0 0 0 1 0 1 0 0 1 1 1 0 Actual size: ○ Save Image	

What was challenging about magnifying the image? How did you interpret the instruction to "double the size" of the image?

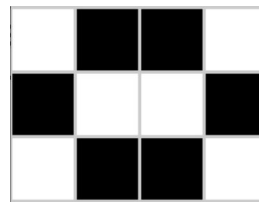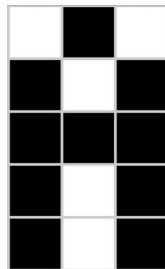How many more bits did you need to double the size?

Name(s)_____ Period _____ Date _____

Activity Guide - Invent a B&W Encoding Scheme

Develop an Encoding Scheme

Look at the simple black-and-white images below. With your partner discuss how you might encode images like these in binary.

Record your ideas for your encoding in the space below. As you develop your encoding consider:
- What information will need to be included to reconstruct the image?
- How will that information be represented in binary?
- Try out your encoding with the sample images above. Is there any aspect of your encoding that is ambiguous or could lead to improperly structured images? Could you include more information in your encoding to solve this problem.

Now consider these larger images. Does your encoding scheme still work? Why or why not? What would you have to change to make it work?
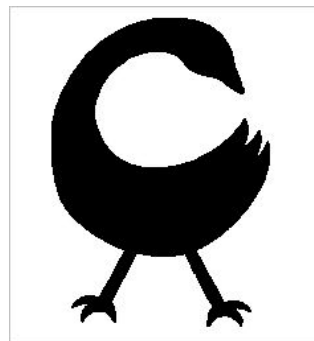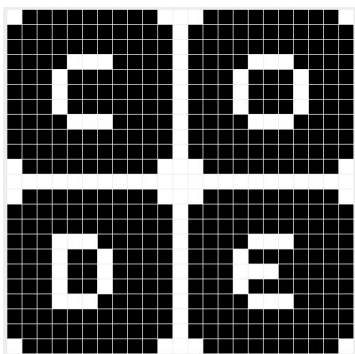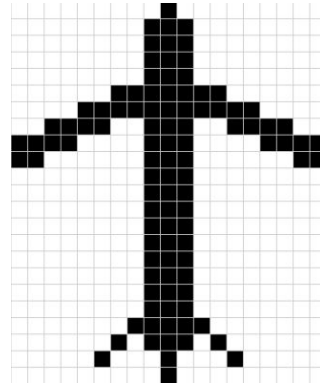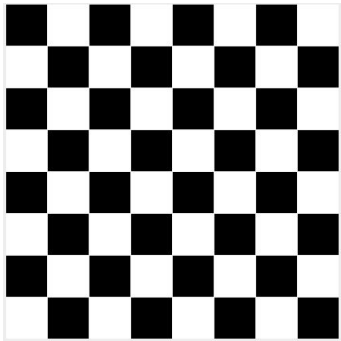
Record any new ideas you have in the space provided below.

Unit 2 Lesson 4

Encoding Color Images

Resources

Name(s)_____ Period _____ Date _____

Worksheet - Video Guide for "A Little Bit about Pixels"

VIDEO OVERVIEW

In this video, Kevin Systrom, one of the co-founders of Instagram (a popular image sharing application), and Piper Hanson, a freelance photographer, explain pixels, RGB color, how image filters work, and that all image data is ultimately represented as bits, 1's and 0's.

Concepts

- Image sharing
- Digital images as data
- RGB color
- Screen resolution and pixel density
- Digital photo filters

THINK/DISCUSS

- Do pixels have to be physically represented as squares? What else could they be and why? Why are squares preferred?
- Think about your favorite image filters. How do you think the RGB values are manipulated to produce another image? Suggest possible algorithms.
- How do you suppose image sharing services could send the image data (bits) faster from one machine to another?
- Why can you never have a perfect digital reconstruction of a real-life image no matter how many pixels are used?
- There are 8 bits in one byte. How many bits and how many bytes are used to represent one pixel's RGB components?

QUESTIONS

1. Why were images first broken down into pixels? What problem did this solve?

2. How many different values can be represented by one Red component of a pixel?

3. What are screen resolution and pixel density? How do they differ? Draw a diagram to support your answer.

4. RGB Values
   a. Why do RGB values range from 0 to 255?

   b. Can they be any other range?

   c. What happens if you increase the range?

   d. Is it possible to represent all of the colors in the world? Why or why not?

5. Describe a function that would take in any RGB value and double its intensity. What do you think happens if the input is 200, 220, 209?

6. Extra: Take out the calculator! How many different colors can be represented in the standard RGB color model?

Name(s)_____ Period _____ Date _____

Activity Guide - Encoding Color Images

Objectives:

- Practice with the color pixelation tool.
- Use the Pixelation Widget to recreate specific colors.
- Recreate simple color images and describe the difference when using 6 or 12 bits per pixel.
- Encode colors with hex.

Directions:

Go to Code Studio - Find this lesson on Code Studio
The tasks below follow the guided sequence in Code Studio which has tutorial videos followed by specific tasks. The purpose of this activity is only to become familiar with the color pixelation tool, and to practice a little bit, in order to be comfortable enough to start the **favicon assignment.**

Step 1: 3-bit color
Tutorial Video: How to use the pixelation widget to control color.
Task 1: Fill in the last two pixels with the missing colors
Problem: fill in the last two pixels.

Optional Practice: change the pixel data so that the blue and yellow pixels exchange locations in the image. If you mess up, you can always re-copy-paste the bits from this example.

Step 2: 6-bit color
Tutorial Video: more bits per pixel for more colors
Task 2: Experiment with 6-bit color
Set bits/pixel to 6 and experiment with each of the R, G, and B values. At 6 bits/pixel, it means each color channel has 2 bits assigned to it, so there are 4 possible shades of each of R, G, and B. To start you are presented with the image at right.
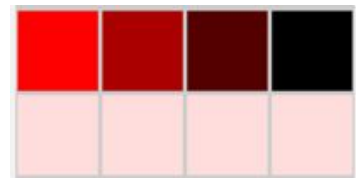
Step 3: 12-bit color and hex
Tutorial Video: Using hex to type bits more quickly
Task 3: Experiment with Hex

Make a 4x4 image with *bits-per-pixel set to 12* and the tool into *hexadecimal mode.*

NOTE: Using 12-bits-per-pixel (4096 colors!) is convenient because 12 bits maps to exactly 3 hex digits, one digit each for RGB. For example, to make a teal color (shown right) whose 12-bit value is: 001110101011 We can represent a 12-bit color in hex much more easily as hex: **3AB**. Here is the breakdown:

	R	G	B
binary	0011	1010	1011
hex	3	A	B

To make a darker shade, just choose a smaller hex digit for the channel of color you want to affect, such as 3A5 (which turns the amount of blue light down).

Recall: When writing a hex number we frequently preface it with a "#" symbol or "0x". So hex: 3AB, would be written #3AB or 0x3AB. This is to avoid confusion when hex numbers have only decimal digits. 101 is a very different number from #101.

Create an image that looks something like picture shown at right in which
- row 1 - displays 4 different "reddish" colors (with red being the dominant color).
- row 2 - display should show 4 different "greenish" colors
- row 3 - should show 4 "bluish" colors
- row 4 - show only shades of gray
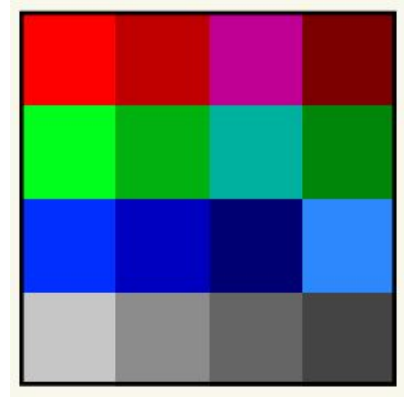
Questions:

1. Rewrite this string of numbers from raw format to readable format, and identify or label the sections of bits.

   000000100000001000000011010010101010

2. How many more colors are available with 12 bits-per-pixel than 6 bits-per-pixel?

3. What happens if you have you an image that has 6 bits-per-pixel and you change it to 12?

4. A digital artist comes up to you and says, "Help! I need a 12-bit color that's just a little bit greener than **#79B."** What would you suggest and why?

Name(s)_____ Period _____ Date _____

Activity Guide - Encoding Hexadecimal Numbers

Hexadecimal: a number system comprised of the familiar ten arabic numerals (0, 1, 2 … 9) as well as the first six English letters (A, B, C, D, E, F). Also referred to as a "base 16" number system, hexadecimal is used in the world of computing to help humans read and talk about large binary numbers. Since there are no symbols reserved for the numbers 10 through 15 in our familiar base 10 number system, the characters A through F are used to represent them.

Here are the first 16 numbers, in both binary and hexadecimal:

| Number	Binary	Hex
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

Converting from hexadecimal to binary is quite easy. One hexadecimal digit can represent any of 16 values. To do the same in binary you would need 4 bits, or 4 binary digits. Therefore, every hexadecimal digit can be replaced with its four bit equivalent in binary.

Converting Hexadecimal to Binary

1. Split the hexadecimal number up into two digits.
2. Convert each digit to decimal.

Example: Converting 1A in hexadecimal to binary

The first digit contains the number 1. The 4-bit binary representation of 1 is 0001.
The second digit contains "A" or 10. The 4-bit binary representation of 10 is 1010.
Therefore 1A in hexadecimal can be converted to 00011010 in binary.

Hexadecimal		
1	A	
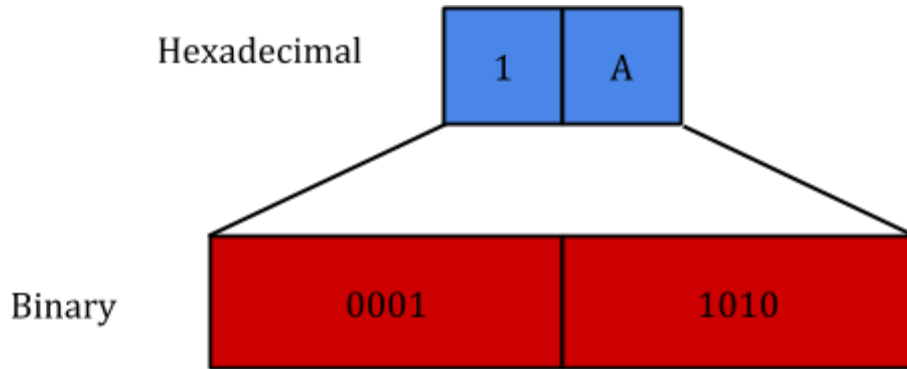
Binary	
0001	1010

Convert: Convert the following hexadecimal numbers to their binary equivalents:

Hexadecimal	Binary
3	
A	
1C	
89	
DD0	
ABC7	

Example: Converting 01001101 in binary to hexadecimal

The first four bits are 0100 or "4," which is just the character "4" in hexadecimal.
The second four bits are 1101 or "13," which is just the character "D" in hexadecimal.
Therefore 01001101 in binary can be converted to 4D in hexadecimal.
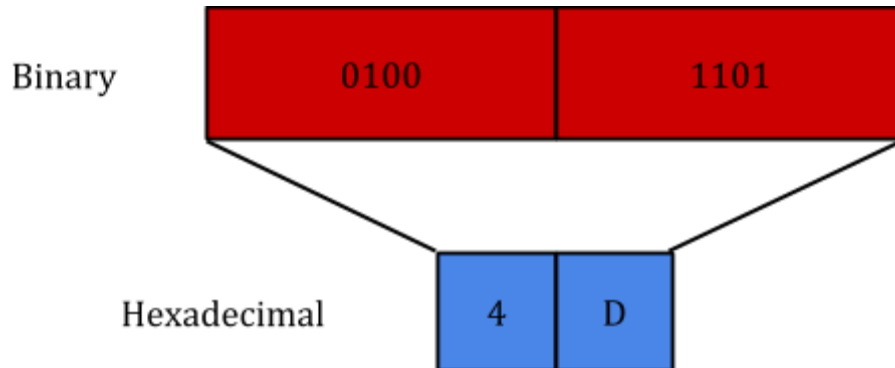
Binary	Hexadecimal
0111	
1110	
01010001	
11000111	
000101011111	
101001011110	

Hexadecimal numbers are used because this conversion allows you to quickly communicate information about large chunks of binary digits. Perhaps the most common place one will see this is in naming colors to be displayed on a screen.

Representing Colors

While monitors typically can represent thousands, if not millions of colors, each individual pixel contains only three tiny lights, one each of red, green, and blue. Every color you see on a screen is created by setting different brightnesses for each of these little lights. Amazing!

Typically each of these little lights (red, green, blue) has its brightness dictated by a single byte. The largest number that can be written using one byte is 255, which indicates that a light should be turned to full brightness. 0 indicates that a light should be turned off.

Example: If we wanted to make the reddest color our pixel could make, we would turn the red light all the way on, and the other two all the way off. This would be represented as

255
"All the way on"

0
"Off"

0
"Off"

A Single Pixel

red: 11111111 or 255
green: 00000000 or 0
blue: 00000000 or 0
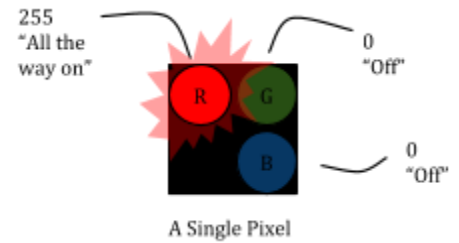
But since these will usually be read as one 24-bit chunk, we display them below in the order they would be received, first the red byte, then the green byte, then the blue byte

Reddest Red: *111111110000000000000000*

While this system makes a lot of sense for a computer to read, it can be tricky for a human to keep track of all those digits. This is where hexadecimal becomes really handy. The 24-bit sequence representing each pixel is 6 nibbles long, and so it can be represented as 6 hexadecimal digits.

Example: Our reddest red can be written in hexadecimal as "#FF0000". To do the conversion yourself, just read the bits above in 4-bit nibbles and convert from binary to hexadecimal using the table above.

Match the color to its hexadecimal equivalent:

Color	Hex
(green)	#df0060
(blue)	#ffffff
(white)	#00ffb6
(teal)	#ff00ff
(orange)	#000000
(black)	#0000ff
(magenta)	#ff9900
(pink)	#00ff00

Name(s)_____ Period _____ Date _____

Activity Guide - Personal Favicon Project

Objectives

- Encode a 16 x 16 pixel image with at least 12 bits per pixel.
- Create and encode a color image of your own design.
- Explain the benefits of using hexadecimal numbers for representing long streams of bits.

Overview

A favicon is a small image, usually 16x16 pixels, that is typically shown in a web browser's address bar next to the title of the page or web address for a particular site. It is typically a small version of a company logo or some other symbol for the site. A favicon for Code.org is shown to the right.

Favicons are designed by artists and programmed into web pages by web designers. Below are some examples of favicons—you might recognize some!

Do a google search for *Favicon* and see what comes up.

Directions

1. Create a personal 16x16 favicon and encode it using the Pixelation Widget on the final level of this lesson in Code Studio.
2. The image you make should represent your personality in some distinctive way.
3. Optional: After you have finished your favicon, share it with others in the class by sending them the bits with the Internet Simulator Widget!

Requirements

- The icon must be 16x16 pixels.
- You must use the Pixelation Widget to encode the bits of color information.
- The image must be encoded with at least 12 bits per pixel.

Things to think about

- A simple design with a few basic colors is probably the best solution. How could you use more colors?
- Plan ahead: Sketch your design before starting to encode the bits. You might want to use a tool to help you draw small images. Suggestions:
  - Favicon & App Icon Generator: http://favicon-generator.org/editor/
  - Make Pixel Art: http://makepixelart.com/free/

Tip

Consider switching the Pixelation tool into HEX mode (base-16) instead of binary. This will enable you to more easily use a greater number of bits. If you use hex, you should consider using 12-bits-per-pixel (4096 colors!) because that maps to exactly 3 hex digits, which also maps easily to RGB—1 hex digit for each. For example, to make a teal color (shown right) whose 12-bit binary value is: 001110101011, we can represent it in hex much more easily as 3AB. Here is the breakdown:

	R	G	B
binary	0011	1010	1011
hex	3	A	B

To make a darker shade, just choose a smaller hex digit for the channel of color you want to affect, such as 3A5 (which turns down the amount of blue light).

Name(s)_____ Period _____ Date _____

Rubric - Personal Favicon Project
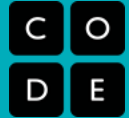
Rubric: Evaluate the success in creating a favicon according the following criteria.

Criteria	Yes	No	Comments
Favicon is 16 by 16 pixels			
Favicon is encoded in RGB color using at least 12-bits-per-pixel			
Favicon is a discernable image, and not merely a pattern			

Reflection Questions

1. What are the potential benefits of using a greater number of bits in designing your favicon? What are the potential drawbacks?

2. Your classmate claims, "Switching my favicon from binary mode to hexadecimal mode is be an example of compression". Do you agree with classmate? Justify your response.

3. Explain what your personal favicon is a representation of and why you chose to create this particular image.

Unit 2 Lesson 5

Lossy Compression and File Formats

Resources

Name(s)_____ Period _____ Date _____

Worksheet - File Formats Rapid Research

Rapid Research

You and a partner will do some rapid research and reporting on some of the most common file formats that are out there. Your teacher will assign you one or more of the file formats listed below. Use the web as your research tool.

Fill in the row of the table for the format you were assigned. TIPs:
- When you visit a web page do a text search on the page for **"compress"**
- In most browsers the keystroke Ctrl+F (Windows) or ⌘+F (Mac) will get you there.

File format name	What type of information or data is it?	Compression? uncompressed, lossy, lossless	How does it work? (just a few brief words about the theory behind the encoding or compression)
BMP (bitmap)			
JPEG (or JPG)			
WAV			
PNG			
MP3			
GIF			
ZIP			

Question: What other file type(s) have you heard of? Which one would you like to know the most about?

# Unit 2 Lesson 6

Practice PT - Encode an Experience

Resources

Practice PT: Encode an Experience

Project Description

Figuring out clever ways to encode information is one of the things computer scientists need to think about in many different contexts. There was a time before we had digital images, or music, or streaming television, or online shopping. Someone had to figure out how to encode those things in order get them onto and into computing devices. For this project you will: **Design a way to encode some human experience in binary.**
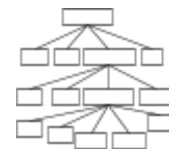
Encodings
Images:
Formatted Text: whoa there!
Pixels:
ASCII: z
Numbers: 90
Bits: 01100100
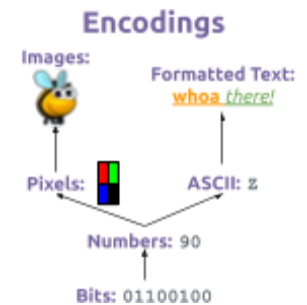
Step 1: Choose Your Experience	You should choose any type of experience you wish. Your challenge is to identify the critical aspects of that experience and determine how that can be translated into a binary encoding. You should: • **Pick a general category of experience --** "Going to a baseball game" is the general experience, as opposed to a specific game you attended once. • **Pick something that isn't already "digital" --** "Going to a concert" is a better choice than "making a playlist" since the second one is already primarily a digital process. • **Pick something you have a detailed understanding of --** "Walking on the moon" is fun if you think you know enough to design that, but "Walking to school" has a lot of interesting information as well.
Step 2: Break Down Your Topic	You will create a diagram (see example on next page) that shows how you broke down your experience into pieces you could represent with ASCII or numbers (i.e. binary). This is sometimes called **Top-Down Design** - a design process that begins by specifying complex pieces and then dividing them into successively smaller, less-complex pieces.
Step 3: Develop a Detailed Encoding	Pick **one** aspect of your experience to show with detail how you would encode it in binary. You will develop a detailed encoding protocol explaining precisely how that portion of your diagram would be encoded. You will make a table that has a row for for each component that lists the following information.

Component	Type	Number of bits/range	Description / Comments
name of the component	*ASCII or Number*	*How many bits do you need? For text: how many characters?*	*Explain how/why you encoded they way you did.*

(see full example on next page)

Step 4: Written Response	Write no more than 300 words in response to this statement: ***There are trade-offs when representing information as digital data.*** Citing examples from this project and what you've learned so far in the course, what does this mean to you? What are the considerations that go into representing information digitally? What are the trade-offs? What's easy? What's hard? Why?

Name(s)_____ Period _____ Date _____

Example: encode "Attending a Birthday Party"

Here is an annotated example of how we used a top-down design strategy to encode the experience of "attending a birthday party". Designing top-down, you **start with the experience...**
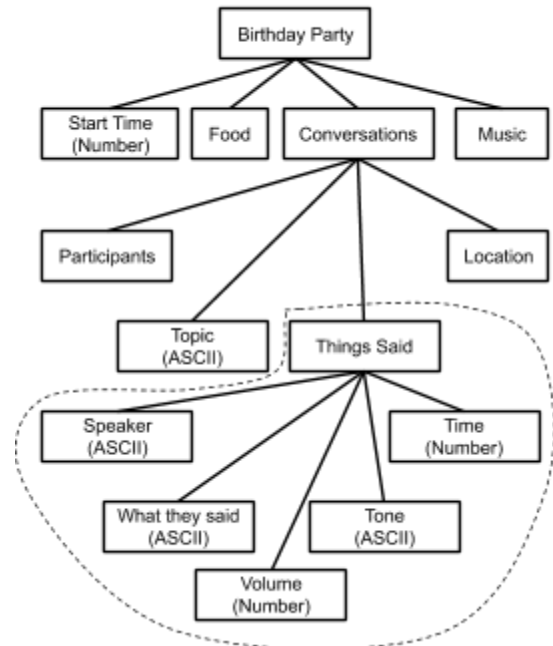- Then think first about the **major categories** that make up that experience.
- Then **consider whether any of those components can be immediately encoded into ASCII text or numbers**.
- Then for the rest of your components **keep breaking them down into sub-parts** until a path in your design tree ends in ASCII or a number.

Your diagram will end up looking like a "tree" (see right).

The example shows how we chose to break down "Attending a Birthday Party". We chose to include the start time, food, music, and conversations as the major categories (but there are certainly many others we could have added).

At the top level *Start Time* could be encoded as a number, but we couldn't figure out a way to directly encode conversations, so we kept breaking it down, dividing it into *participants*, a conversation *topic*, the *things said*, and a *location*. We even decided that the *things said* would need to be further divided to truly capture the experience. Each "Thing said" at the party was broken down into pieces that could be encoded as ASCII or a Number.

Everything in the diagram represents **choices we made** about how to encode the information. There are no right or wrong answers here. A "correct" way to encode something depends on how you justify the choices you made. Here is our **detailed encoding for *Things Said.***

Diagram: Attending a Birthday Party

Things Said: Every time a person says something in a conversation their comment, question, etc. is encoded. By combining all of the "Things Said" it is possible to recreate the conversation.

Component	Type	Number of bits/range	Description / Comments
Speaker	ASCII	256 bits (0 - 32 chars)	Name of the person speaking. 32 characters should be enough for most names. It is important to know who said what.
What they said	ASCII	8000 bits (0 - 1000 chars)	The text of what the speaker said. Contains the actual contents of the conversation. I figure that a person wouldn't speak for more than 1 minute at a time. I found out that a typical American speaks at most 150 words per minute, and on average words have 5 characters. With spaces that is 900 characters so 1000 is sufficient.
Volume	number	7 bits (0 - 127 dB)	Measured in decibels. A rock concert is 125 decibels so the range of a 7 bit number (0-127) is enough. Helps gauge how loud the party was.
Tone	ASCII	128 bits (0 - 16 chars)	How the person said something e.g. "laughing" or "serious". This will only be a short one or two-word description to help reconstruct the tone of the conversation.
Time	number	16 bits (0-65,535 secs) (0 to 18 hours)	When the thing was said, measured in seconds since the beginning of the party. 16 bits provides (0-65,535) seconds or a little more than 18 hours (long party). This is important for knowing the order of comments.

Encode an Experience- Submission Guidelines and Rubric

Submission Guidelines

For this project you will submit to your teacher:

1. **Your diagram** showing the breakdown of your experience
2. **Your detailed encoding** table showing the binary components of a bottom level of your diagram "tree"
3. **Your written response** about trade-offs in representing information as digital data

Project Rubric

Component	1	2	3	Score
Diagram: Complete	Few aspects of the experience have been accounted for or the diagram is incomplete.	The experience has been generally broken down but some key elements may have been missed.	The experience has been thoroughly broken down into its component parts. All critical elements have been accounted for.	
Diagram: Proper Encodings	Many components have been improperly assigned number / ASCII or should have been further subdivided.	Most components are subdivided when necessary and are reasonably assigned ASCII / number.	Components are consistently subdivided when necessary and are reasonably assigned ASCII / number.	
Detailed Encoding: Number of Bits	The encoding frequently uses an unreasonable number of bits to encode each component.	The encoding generally uses a reasonable number of bits to encode each component.	The encoding consistently uses a reasonable number of bits to encode each component.	
Detailed Encoding: Description / Comments	Most components require more explanation of how they will be interpreted or why they are included.	Some components could benefit from further explanation of how they will be interpreted or why they are included.	Nearly all components provide rich explanations of how they are interpreted and why they are included.	
Reflection: Tradeoffs in representing digital data	The response reflects an incomplete, incorrect, or trivial understanding of considerations that go into the digital representation of information.	The response reflects a partial understanding of considerations that go into the digital representation of information. Could benefit from better examples or better explanation of pros and cons.	The response reflects a deep understanding of considerations that go into the digital representation of information. Examples are cited, benefits and drawbacks are mentioned.	

Name(s)_____ Period _____ Date _____

Encode an Experience - Diagram Workspace

Write your topic in the space provided and subdivide it into its key components. For each rectangle you draw write "ASCII" or "number" if you believe it can be encoded in that way. If not, further subdivide that rectangle. You may need additional sheets of paper.

_____
your experience

Encode an Experience - Detailed Protocol Template

Topic:			
Component	**Type**	**Number of bits/range**	**Description / Comments**

Encode an Experience - Written Response

Write no more than 300 words in response to this statement:

There are trade-offs when representing information as digital data.

Citing examples from this project and what you've learned so far in the course, what does this mean to you? What are the considerations that go into representing information digitally? What are the trade-offs? What's easy? What's hard? Why?

Encode an Experience- Submission Guidelines and Rubric

Submission Guidelines

For this project you will submit to your teacher:

1. **Your diagram** showing the breakdown of your experience
2. **Your detailed encoding** table showing the binary components of a bottom level of your diagram "tree"
3. **Your written response** about trade-offs in representing information as digital data

Project Rubric

Component	1	2	3	Score
Diagram: Complete	Few aspects of the experience have been accounted for or the diagram is incomplete.	The experience has been generally broken down but some key elements may have been missed.	The experience has been thoroughly broken down into its component parts. All critical elements have been accounted for.	
Diagram: Proper Encodings	Many components have been improperly assigned number / ASCII or should have been further subdivided.	Most components are subdivided when necessary and are reasonably assigned ASCII / number.	Components are consistently subdivided when necessary and are reasonably assigned ASCII / number.	
Detailed Encoding: Number of Bits	The encoding frequently uses an unreasonable number of bits to encode each component.	The encoding generally uses a reasonable number of bits to encode each component.	The encoding consistently uses a reasonable number of bits to encode each component.	
Detailed Encoding: Description / Comments	Most components require more explanation of how they will be interpreted or why they are included.	Some components could benefit from further explanation of how they will be interpreted or why they are included.	Nearly all components provide rich explanations of how they are interpreted and why they are included.	
Reflection: Tradeoffs in representing digital data	The response reflects an incomplete, incorrect, or trivial understanding of considerations that go into the digital representation of information.	The response reflects a partial understanding of considerations that go into the digital representation of information. Could benefit from better examples or better explanation of pros and cons.	The response reflects a deep understanding of considerations that go into the digital representation of information. Examples are cited, benefits and drawbacks are mentioned.	

Unit 2 Lesson 7

Introduction to Data

Resources

Teacher Setup for "Data Tracker" Project

This lesson requires a one-time special setup in order to create a form for data collection with the students in your class. Once you have it setup you will use it for several weeks.

Here are the steps:

1. **You must have a Google Account**
   If you have a school Google Account you probably want to be logged in with that account.
   If you do not have a Google account or would prefer not to use Google products for this lesson, see the Alternatives to Google section at the bottom of this setup guide.

2. **Make a copy of the Google Form**
   a. Click this link: http://bit.ly/CSP_Unit4_DataFormTemplate (This redirects to a *Template* for a Google Form.)

   b. At the top-left of the screen, you should see something that looks like this:

   That will make a copy of the form and add it to your Google Drive.

   c. You will then see an editable version of the "Class Data Tracker" form.
   (NOTE: If you would like to make changes to this form to add or remove questions please see our notes about doing so below.)
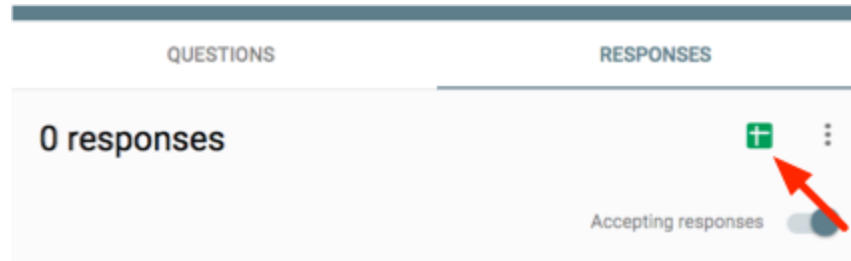   d. Add a spreadsheet to collect responses…

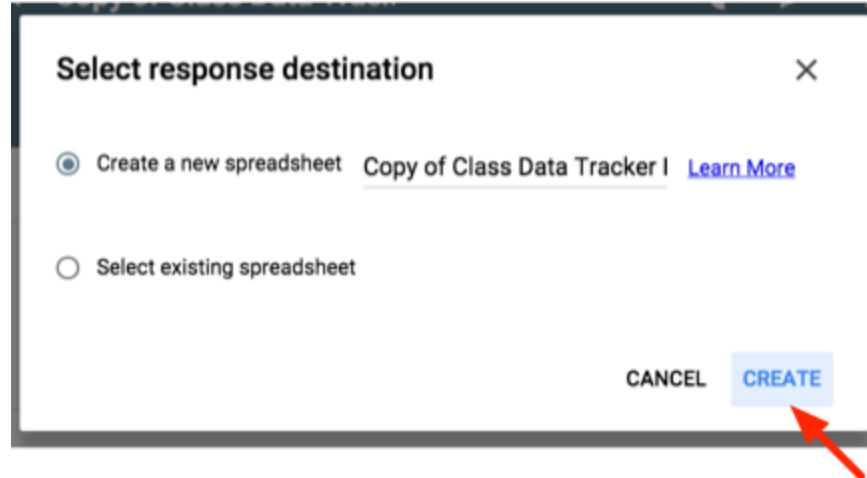   At the top of the survey click RESPONSES
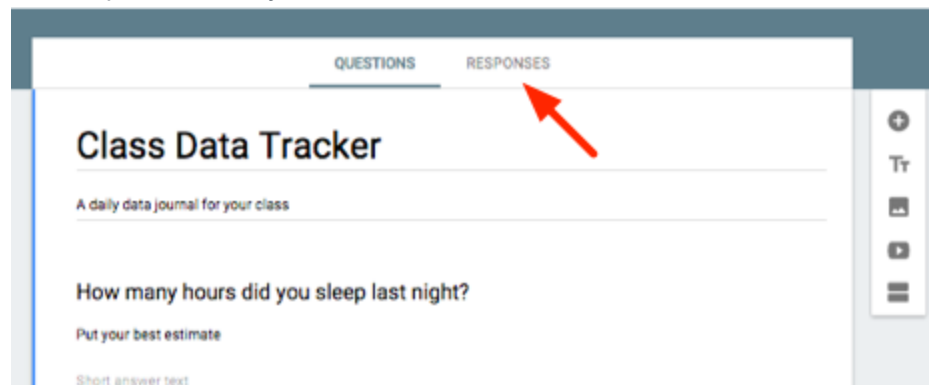
   Then click the green Icon

Then Click the CREATE button - use the default option to create a new spreadsheet
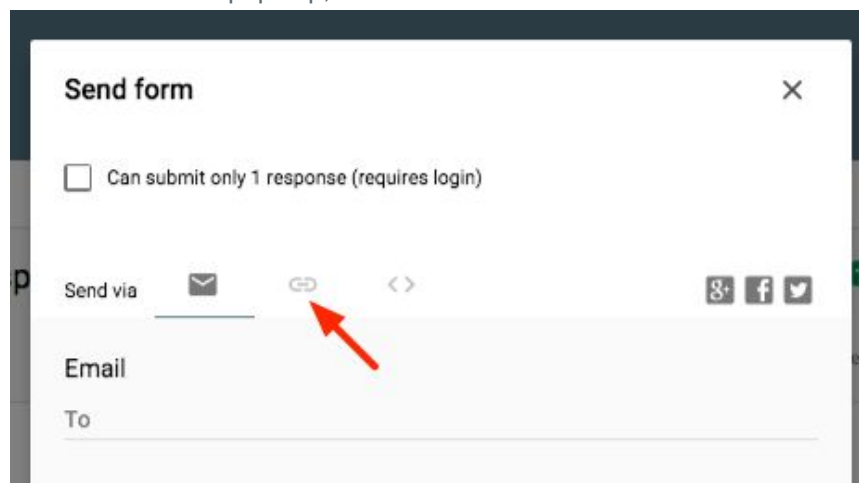

It will then open up a new empty spreadsheet where responses to the survey will be collected. You can close this spreadsheet for now. We'll get back to it later.

3. **Share a link to the form with your students**
   a. Go back to the page that was showing the form. At the top-right of the page click the button that says SEND.

   b. In the window that pops up, click the little Link Icon

   c. The window then shows a direct link to a live version of the form that anyone can fill out. This is the link to share with students. It's probably easiest to choose the short URL version and to share that link with students.

You may alternatively want to share the link via the email option if you have a list of student email addresses you can use.

Place or post this link in a place where students have quick and easy access to it.

4. Whew, done!
You should now have:
- A copy of a Google form in your Google Drive
- A spreadsheet that will collect responses from the form in your Google Drive
- A link that students should use to fill out the survey

You will have students fill out this form EVERY DAY, or as frequently as possible over the next few weeks. We will look at the results more fully in Lesson 7 of this unit. You can place the form and spreadsheet documents wherever you like in your Google Drive. They are yours now.

NOTES:
For general help with Google Forms, go here:
https://support.google.com/docs/topic/6063584?hl=en&ref_topic=1360904

Alternatives to Google Forms/Sheets
You need some way for students to anonymously submit a survey every day and for the responses to be collected in a spreadsheet somewhere. There are other survey products out there, such as **Microsoft Excel Online** or **Survey Monkey**, or you may also wish to do it with your school's LMS, if that is an option for you.

Editing the Survey Questions...
The questions in this short survey have been chosen for certain properties they will have after students have filled out the survey repeatedly over a number of weeks. If you choose to remove or change questions on the survey, or make your own completely different survey, please keep the following in mind, so that the future lessons in the unit (specifically 7, 8 and 9) will hold up.

Question	Properties we need
How many hours of sleep? How many hours of work? How many hours did you relax?	A free-form text entry that's supposed to be a number. We want this data to be free-form so that it gets a little "dirty." We have tasks and strategies for cleaning up the data later on.
What did you do to relax?	Free-form text entry that could be anything, but a directed response to a question. As a later task, students will look at these responses and try to fit them into categories so that the data is computable.

How do you feel today?	A fixed scalar response. Something easy to plot.
Which describes you (only child, oldest, middle, youngest sibling)?	A fixed response that will not change from day to day. This is in lieu of some other way to track responses over time, like asking for other personally identifiable info like eye/hair color, or even name or student ID. We want it to be anonymous but to be able to track *something* over time. If you are comfortable adding a question about gender to the survey, that can lead to interesting investigations later on.

Unit 2 Lesson 8

Finding Trends with Visualizations

Resources

Name(s)_____ Period _____ Date _____

Activity Guide - Exploring Trends

What's a Trend?

When you post information to a social network, watch a video online, or simply search for information on a search engine, some of that data is collected, and you reveal what topics are currently on your mind. When a topic is quickly growing in popularity, it is often said to be **trending**, but there are many different trends or patterns we might find in this data, including historical trends. These patterns might help us to identify, understand, and predict how our world is changing.

Using Google Trends

You will be using Google Trends, which is a tool that allows you to **visualize data** about search history across different times and locations**.** You will be **looking for interesting patterns, trends, or relationships between multiple trends** and **try to tell the story that pattern is showing**.

- Access it here:
  https://www.google.com/trends/
- To get started, you want to "Explore" a trend of your own.
- Find the Explore text box or go to the Explore section of the site
  https://www.google.com/trends/explore
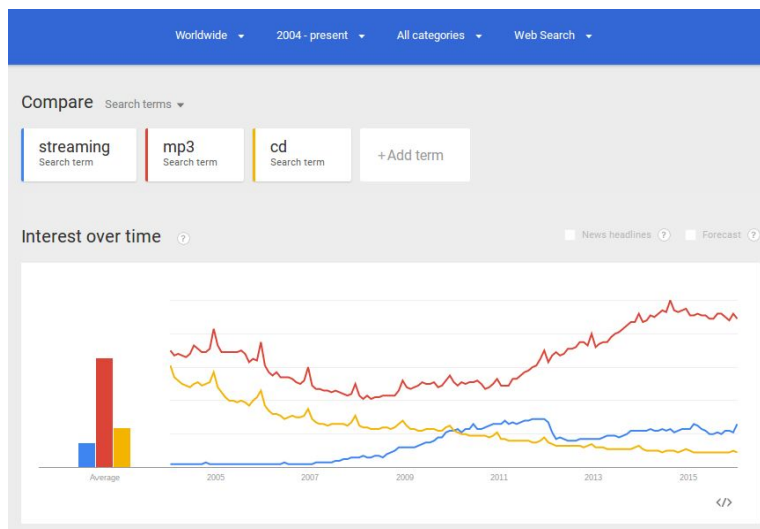
Do your first comparison of trends

You can enter any two or three things you like. This example compares **streaming**, **mp3**, and **cd**.

Understand what is being represented: Take a moment to actually understand what the Google Trends data is showing. Respond to the questions below. You can find answers on the Google Trends Help Page.

- Where does Google Trends data come from?

- How is Google Trends data adjusted? What does a value of 100 mean?

Exploring Trends on your own

Continue exploring comparisons of search trends that are interesting to you. You might start by looking up:

- **a current event / social movement / hashtag / meme**
- **your favorite hobby / movie / song / book / celebrity**
- **popular apps / businesses / products / websites**

As you try different terms, here are **some things to look for:**
- **Long-term trends:** Is your topic becoming more popular over time? Less?
- **Short-term trends:** Does your topic suddenly spike or dip in popularity?
- **Patterns:** Does your topic follow some predictable repeated pattern?
- **Relationships:** Does one topic increase or decrease in popularity when another one does?
- **Zoom-in:** You can narrow your search to particular regions, times, and categories

Tell a Story

As you explore, you should eventually **settle on one topic** you find particularly interesting. For your chosen topic, **be prepared to present or share** what you found, responding to the following:

- **Describe** what terms you compared and whether you narrowed your search by using filters.

'

- **Accurately describe** what the charts or other visualizations are showing.

- Come up with a **possible story or explanation of why** the trend you described might have happened.

Unit 2 Lesson 9

Check Your Assumptions

Resources

Name(s)_____ Period _____ Date _____

Activity Guide - Digital Divide & Checking Assumptions

The "Digital Divide"

Perhaps one of the easiest assumptions to make, when looking at data collected online, is that it actually is a good representation of what the average person is thinking, doing, or cares about. Look through this report from Pew Research, which shows the large differences in access to technology, even in the modern day. Digital Divides 2018 (http://www.pewinternet.org/fact-sheet/internet-broadband/)

1. What is the "digital divide"?

2. What groups are overrepresented or underrepresented online as a result of the digital divide?

3. What was the most surprising piece of information or visualization you found in this report?

Identifying Assumptions in Data Analysis

When you use data to make decisions, you need to be careful to **identify your assumptions** and **reflect on how those assumptions impact your analysis.**

Pick one of the scenarios below. Circle or highlight the one your chose. With a partner, respond to the questions you find there about the assumptions made to conduct that analysis.

Scenario	Data	Decision
A city would like to more efficiently locate potholes that need to be filled.	The city builds an app that allows residents to report potholes from their smartphone.	The city will use this app as the primary method of identifying potholes.
A news agency would like to predict the outcome of a coming election.	A social media company (e.g., Twitter) keeps track of how many times each candidate has been mentioned on the platform.	The news agency will use this data to predict the outcome of the election.
A state government is trying to determine which issues are most important for the upcoming year.	The government creates an online survey where citizens can vote for the issue they care most about.	The government will use the results of this survey to help prioritize issues in the coming year.
A chef is deciding where to open a new restaurant in the city.	A restaurant reviews website keeps track of the areas of the city that receive the most restaurant reviews.	The chef will use the data to choose the location for opening a new restaurant.

Respond
Answer each question below for the scenario you chose.

1. What kind of trends or patterns might the decision makers in your scenario be observing in the data when they make their decision? How might they use these patterns to help make their decision?

2. The people making decisions in each instance have not taken into account the "digital divide." How has the assumption that all people have equal access to technology **possibly biased their analysis?**

3. Are there any other assumptions being made in your scenario? How might they affect the final decision made? What additional data might you want to collect before making the decision for your scenario?

Unit 2 Lesson 10

Good and Bad Data Visualizations

Resources

Name(s)_____ Period _____ Date _____

Student Worksheet - Data Visualization Scorecard

Rate it: From Good to Hilariously Bad

Making good data visualizations is hard work, even for so-called experts. Today you and a partner are going to **collaboratively review a collection of data visualizations**. Some are strong data visualizations that create a deeper understanding of the underlying data. Others are difficult to interpret or even just hilariously bad. Your job is to rate the quality of the visualizations and keep a few notes about why they are good or bad.

Directions

You can find two collections of visualizations ("Collection A" and "Collection B") on **Code Studio.** Choose either collection and go through each visualization. For each one:

- **Examine the visualization**, as well as any included text. What is it trying to communicate?
- **Discuss with a partner**: Is the visualization clear? Is it misleading? How could it be improved?
- **Rate the visualization** and make a short note about why you gave the rating you did.

Which collection are you looking at? (circle one)   **Collection A**   **Collection B**

Number	Your Rating				Comments / Notes
1	Great	Good	Bad	Horrible	
2	Great	Good	Bad	Horrible	
3	Great	Good	Bad	Horrible	
4	Great	Good	Bad	Horrible	
5	Great	Good	Bad	Horrible	
6	Great	Good	Bad	Horrible	
7	Great	Good	Bad	Horrible	
8	Great	Good	Bad	Horrible	
9	Great	Good	Bad	Horrible	
10	Great	Good	Bad	Horrible	
11	Great	Good	Bad	Horrible	
12	Great	Good	Bad	Horrible	
13	Great	Good	Bad	Horrible	
14	Great	Good	Bad	Horrible	
15	Great	Good	Bad	Horrible	

Best and Worst Visualizations

Choose the **best and worst** visualization from the list above.

The Best One

- Which visualization was the best?
- What story or idea does it successfully convey?
- What specific aspects of this visualization makes it particularly effective?

The Worst One

- Which visualization was the worst?
- What story or idea do you think it is trying to convey?
- Why specific aspects of this visualization makes it particularly ineffective? Is it wrong? Misleading? Nonsense?

Fix It (optional)

Sketch an outline of a better way to visualize the information from **the worst visualization**. **Briefly explain** why your suggestions would improve the visualization.

Unit 2 Lesson 11

Making Data Visualizations

Resources

Unit 2 Lesson 12

Discover a Data Story

Resources

Name(s)_____ Period _____ Date _____

Activity Guide - Discover a Data Story

Exploring a Dataset with Visualizations
We collect data in order to learn more about the world around us. It is usually hard to see patterns in a dataset when viewing it in its raw form. Computational tools like summary tables and visualizations make it easier to find interesting trends and relationships within the data.

Terminology
- **README:** a document providing background information about a dataset
- **CSV:** abbreviation of "comma-separated values," this is a widely-used format for storing data
- **Raw data:** the original data as it was collected
- **Summary table:** a table of aggregate information about a dataset (e.g., the average, sum, count of some values)

Getting Started
Collaborate with a partner to **explore the datasets** and **choose one** you would like to explore further.
- For each of the 5 datasets…
  - **Read the README** to understand the **raw data** that was collected.
  - **Look at the rawData.csv file** - see if it makes sense, based on the README.
  - **Look at the other CSV files in the folder -** they are summary tables provided for your dataset.
- **Repeat these steps** with additional datasets.
- **Choose** one to explore more deeply.

Discover a Data Story
Now that you've chosen a dataset, you'll want to use visualization to help find an interesting pattern in your dataset.
- **Open a summary table** from your dataset.
- **Experiment with different chart types** for visualizing that data.
- **Discuss with your partner** what you are seeing in each chart. Have any interesting patterns appeared?
- **Repeat these steps** with additional summary tables.
- **Choose** the summary table and visualization which was most interesting or insightful.

Present Your Findings
You should have found an interesting trend or relationship within your data. Now you'll want to communicate your findings in a way that others can easily understand.
- **Use the [Data Visualization 101](http://content.visage.co/hs-fs/hub/424038/file-2094950163-pdf) guide** (http://content.visage.co/hs-fs/hub/424038/file-2094950163-pdf) to make sure you are making good choices for presenting the pattern or trend you have found.
- **On the next page,** you will find a template for presenting the pattern you found.

Name(s)_____ Period _____ Date _____

Activity Guide - Present Your Data Story

Your Visualization
Paste your visualization in the space below.

Prompts

1. Provide a brief description of the dataset you chose to visualize.
   (approximately 50 words)

2. **Describe the trend, pattern, or relationship** you found within your data. How is it shown with your visualization?
   (approximately 100 words)

3. Come up with a **possible story or explanation** for the trend you described. Make sure to note any assumptions you are making in this interpretation of the data.
   (approximately 100 words)

Name(s)_____ Period _____ Date _____

Rubric - Present Your Data Story

Criteria	Yes	No	Comments
The visualization is clear and easy to read.			
An appropriate chart type was chosen for presenting the data being visualized. (See the Data Visualization 101 guide.)			
A clear explanation of the data on which the visualization is based is provided.			
A clear description of the trend, pattern, or relationship in the data is provided.			
A plausible interpretation, or possible explanation for why the the trend, pattern, or relationship exists is provided.			
The response includes an account of any assumptions made in interpreting the trend, pattern, or relationship in the data.			

Unit 2 Lesson 13

Cleaning Data

Resources

Unit 2 Lesson 14

Creating Summary Tables

Resources

Unit 2 Lesson 15

Practice PT - Tell a Data Story

Resources

Name(s)_____ Period _____ Date _____

Practice PT Overview and Rubric - Tell a Data Story

Background

In this unit, you have collected data about you and your classmates and have learned analysis and visualization skills that should allow you to find interesting patterns within it. At this point, you will have worked with a partner to clean the data that was collected and have learned how to use computational tools to summarize this data. Now it's time to put all these skills together and tell a story with the data.

General Requirements

This task requires you to:

- Manipulate data that you have collected
- Create a computational artifact in the form of a data visualization
- Write about your artifact and its possible interpretations

Step 1: Make your own copy of the data

You will work individually for this project. You may have worked on this data with partner beforehand and you may still collaborate with them to share ideas and brainstorm. However, this project must be a reflection of your individual creativity and work. So make a separate copy of the data you cleaned with your partner so that you don't accidentally clobber each other's work.

Step 2: Find your data story

Dig into the data! Use your skills for data manipulation to find connections and trends. Remember the skills you have for doing discovery:

- Sort data
- Rearrange columns
- Filter
- Make summary (pivot) tables
- Use charts and visuals for discovery

Step 3: Visualize your data story

You will need to design a data visualization that clearly communicates the data story you found. Here are a few things to help you along the way:

- Use the Data Visualization 101 guide to help you pick good chart types.
- Make your visualization clear and easy to read.
- In some instances, **a simple table may be a better way to show your data than a chart.**
- Use text along with your visual to help explain your data story. If you need to use too much text, however, ask yourself if there's a better way to visualize it.

Step 4: Written responses

Respond to the written responses below. Note that a recommended word count is provided for each question.

1. Describe how you collected the data for your visualization including how it was collected, when it was collected, and over what period of time. Your description should be understandable by someone unfamiliar with this project.
   (approximately 100 words)

2. Describe your development process, explicitly identifying the computing tools and techniques you used to create your artifact. Your description must be detailed enough so that a person unfamiliar with those tools and techniques will understand your process.
   (approximately 100 words)

3. **Describe your findings:**
   - **Describe the trend, pattern, or relationship** you found within your data. How is it shown with your visualization?
   - Come up with a **possible story or explanation** for the trend you described. Make sure to note any assumptions you are making in this interpretation of the data.
   (approximately 200 words)

4. **Make a recommendation based on the results of your analysis.**
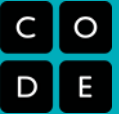   - Describe who you are making the recommendation to and what should they do.
   - Explain how the recommendation will lead to some benefit or prevent some harmful effect.
   - Explain how your recommendation is supported by your analysis of the data, or what else might need to be investigated in order to make a stronger recommendation.
   (approximately 100 words)

Step 5: Submission

Copy your visualization and written responses into a single document. Ask for specific instructions on how to submit this single document.

Rubric - Tell a Data Story

Criteria	Low	Medium	High
Visualization			
The visualization demonstrates proficiency with the computational tools used to create it.			
The visualization follows good design principles.			
The visualization provides insight into a trend or pattern in the data.			
Reflection			
The response clearly describes the source of the data used to create the visualization.			
The response clearly describes the iterative development process used to create the visualization.			
The response clearly describes the relationship, trend, or pattern shown in the data and provides a possible explanation, while noting assumptions.			
The response makes a recommendation, indicates potential benefits of that recommendation, and explains how the recommendation is supported by the provided interpretation of the visualization.			