

# Multiple Choice

## Hardware (Pick 1 of 2 questions)

You are working on an image processing project that involves adding colour values together. Your image is composed of pixels and each pixel is represented as 3 byte hex value (e.g., 34AB12). You want to add the values of two pixels together. Instead of doing this by hand, you opt to use the calculator on your computer. You enter the first colour value and then the second value. The calculator now displays the sum. Where are these three values stored in your computer?

- a. RAM
- b. Register
- c. Hard drive
- d. Cache
- e. In the SSD
- f. None of the above

While on a trip to Bali, Amar left his computer outside on the balcony for a couple of days. When he came back to it, it would not turn on. When Amar took his computer to a repair centre the next day, the technicians informed him that the motherboard on his system was permanently damaged due to overheating. Fortunately, they were able to recover many of the pictures, documents, and videos stored on his computer. Where were these files stored?

- a. Cache
- b. RAM
- c. Hard drive
- d. Register
- e. CPU
- f. None of the above

# Algorithms

Your grandmother has asked you to arrange pictures of her pets in alphabetical order (i.e., A to Z, with A being the smallest or first) based on the pet's name. The pictures are in this order: [Tim, Bob, Binny, Sue, Anne].

You decide to use insertion sort to quickly put the pictures in order. After the second iteration of insertion sort, what order are the pictures in?

- a. [Tim, Bob, Binny, Sue, Anne]
- b. [Bob, Tim, Binny, Sue, Anne]
- c. [Bob, Binny, Tim, Sue, Anne]
- d. [Binny, Bob, Tim, Sue, Anne]
- e. [Binny, Bob, Sue, Tim, Anne]
- f. [Anne, Binny, Bob, Sue, Tim]
- g. None of the above

Due to the COVID-19 pandemic, many sports leagues are uncertain about when they can resume games. You are in charge of deciding if the NHL will begin its season in January 2021. While others have relied on intuition, you decide to examine the evidence and the recommendations of public health advisors around the country and base your decision on three attributes.

Your plan is to develop a decision tree.

Each of the entries in the following table is the advice of one professional. Here are the detailed descriptions of each column:

- Salary cut: To offset the cost of empty stadiums and loss of revenue
- Infected Players: The number of players in the NHL infected with COVID-19 as of December 1<sup>st</sup> 2020.
- Location: Playing in different cities or a hub city?
- Result: Decision to start season in January 2021

Salary Cut	Infected Players	Location	Result
Yes	0 - 5	Hub City	Yes
No	6 - 20	Hub City	No
No	Over 20	Different Cities	No
No	0 - 5	Different Cities	No
No	6 - 20	Hub City	Yes
No	Over 20	Hub City	No
Yes	0 - 5	Different Cities	Yes
No	Over 20	Different Cities	No
No	Over 20	Hub City	No
Yes	0 - 5	Different Cities	Yes
No	6 - 20	Hub City	No
No	Over 20	Hub City	No
No	0 - 5	Different Cities	No
No	Over 20	Hub City	No
Yes	6 - 20	Hub City	Yes

Using the entropy algorithm presented in class and the table above, determine what is the best attribute to split on first?

### Infected Players

Using the entropy algorithm presented in class and the table above, determine what is the best attribute to split on second (i.e., after splitting on the first attribute, what would you split on next)?

## Salary Cut

In the year 3000, humans developed the ability to see additional colours. To represent these additional colours on our screens, computer scientists introduce three new colours into each pixel. I.e., Each pixel is now represented by RGBXYZ where XYZ are the three new colours.

If a pixel has some sort of intensity in 4/6 of these colours, how many bytes would be required to store the pixel?

- a. 256 bytes
- b. 6 bytes
- c. 4 bytes
- d. 12 bytes
- e. Impossible to tell

You are a graphic designer who wants to create images that will be used as promotional material. You are unsure about the scale to be used for the image (i.e., you don't know how big or small these images will need to be). What file format should you save your image as?

Assume you don't have a limitation on computer resources as you have a powerful computer and the images are of simple objects.

- a. Vector
- b. Bitmap
- c. JPG
- d. SVG
- e. GIF
- f. PNG
- g. BMP
- h. None of the above

With regards to data representation, which of the following statements are false?

- a. Adding additional key frames between two images to generate a smooth transition is referred to as tweening
- b. In comparison to paper documents, electronic documents can be easily encrypted
- c. Image pixels are stored on a computer using hexademical values
- d. A bitmap image representation must specify the number of rows and columns in a grid of pixels
- e. None of the above

## Internet

You open your web browser to explore gift options for your friend. You notice that the temperature for your location is displayed in Celsius (as opposed to Fahrenheit). You also notice that one of the ads on the page references some articles relating to climate change. Which of the following provides a reason as to why you are seeing these pieces of information?

- a. You are seeing the temperature of your location in Celsius due to third-party cookies and the list of articles from a story you were following earlier due to first-party cookies.
- b. You are seeing the temperature of your location in Celsius as well as the list of articles from a story you were following earlier due to first-party cookies.
- c. You are seeing the temperature of your location in Celsius as well as the list of articles from a story you were following earlier due to third-party cookies.
- d. You are seeing the temperature of your location in Celsius due to first-party cookies and the list of articles from a story you were following earlier due to third-party cookies.
- e. None of the above.

You create a new Google account on your brand new laptop. You start doing research for your upcoming CPSC 100 research paper on self-driving cars. You keep working on this project for a week or so before submitting it. After the project is over, you notice that you keep getting Tesla ads whenever you visit websites. This is an example of:

- a. Filter bubbles
- b. Computer bias
- c. IP address tracking
- d. Fake news
- e. Google super-intelligent AI
- f. None of the above

Throughout the COVID-19 pandemic, individuals are using the Internet for social interaction much more than before. Some of the natural aspects of in-person interaction, such as being exposed to people with all different views and opinions, have been lost online. Which of the following is an example of an online filter bubble?

- a. Finding a turkey recipe on your favourite food blog to prepare for the holidays
- b. Opening a URL in an email received from your IT department
- c. A soccer player seeing Instagram posts about soccer recommended to them
- d. Visiting the r/UBC reddit page to interact with fellow UBC students
- e. A computer science student joins a Facebook group of only computer science students to discuss your favourite CS electives
- f. None of the above

You have just sent an email from your computer to your friends when suddenly you spill water all over your computer. What happens to your email?

- a. The email is not sent as your computer has shut down
- b. The email gets stopped during transport as when your computer died so did your IP address
- c. The email is received by your friend
- d. The email is saved in packets by the router until you get a new computer and then it will send the email
- e. None of the above

As a clinical psychologist, you have decided to extend your services to include messaging. You have found a Messaging App that offers end to end encryption. As your business grows, you decide to take extra precautionary measures to ensure that conversations are not lost if your cellphone is damaged. You decide to backup the data for all your apps in Google Drive. Which of the following statements is the most true?

- a. Your data is encrypted on both platforms
- b. Your data is encrypted on neither platform
- c. Your data is encrypted in the Messaging App but not in Google Drive
- d. Your data is encrypted in Google drive but not the Messaging App
- e. None of the above

Even before taking CPSC 100, you have heard about hackers tracking your location. Now that you are finished with the course, you want to see if you can actually trace an IP address from a website on your own. By following a guide on Google, you manage to discover your school's IP address.

If `www.cs.ubc.ca` and IP address `142.103.6.5` represent the same server, the former is called [blank1] and the latter is a(n) [blank2] address.

[blank1] = domain name

[blank2] = IPv4

A client and a server can have many interactions. After the first interaction, the route your packet travels is...

- a. Always going to be the same since you have established contact with the server already
- b. Dependent on the traffic conditions of the network
- c. Dependent on whether the server is busy or not
- d. Never going to be the same because the route is dependent on the traffic in the network

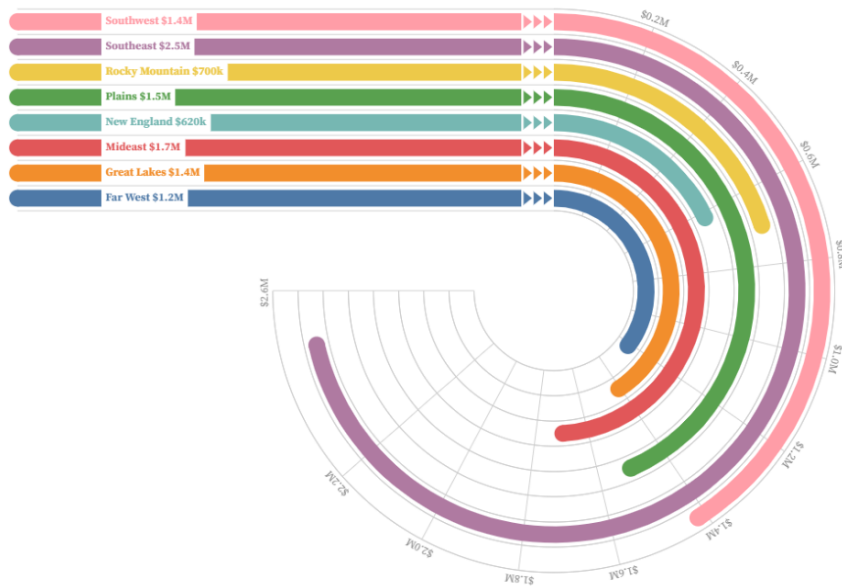
# Visualization

Over the holidays, you conducted a brief survey with family members to ascertain the amount of time they spend on social media apps. You want to present these results to them at the end of the week using a visual representation. Your goal is to show the relationship between the age of the individual and the time spent on social media apps. Which kind of visual representation should you use?

- a. Pie chart
- b. Multi-line chart
- c. Treemap
- d. Scatter plot
- e. Parallel coordinate plot
- f. Heatmap
- g. Network diagram



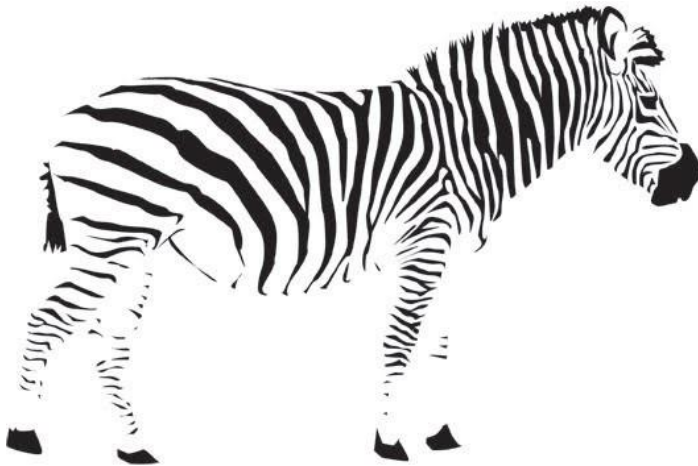
For the diagram shown below, which visual cue is used to encode quantitative data? If you do not see the figure below, then please look for Figure 1 in the exam images document.



For the diagram shown above which visual cue is used to encode quantitative data

- Position common Scale
- Position Non-Aligned scale
- Length
- Direction
- Angle
- Area
- Volume
- Curvature
- Shading
- Color Saturation

Which Gestalt principle is used in the following image?



Closure

Which Gestalt principle is used in the following image?



Continuity

# Data Mining

Hiccup the dragon handler is interested in seeing which species of dragons have similar lifespans from 0 to 100 years. He used the k-means algorithm and found 3 clusters with the x-coordinate of the centroids to be 10, 50 and 100 years. Before he could do it a second time, the dragon species that were in the 50-year cluster flew away to another world. Hiccup decided to run the algorithm again with the dragons that still remained from before. What happens to the inter-cluster distance and intra-cluster distance compared to the first result?

- a. Increases; decreases
- b. Increases; stays the same
- c. Increases; increases
- d. Decreases; increases
- e. Decreases; decreases
- f. Stays the same; decreases
- g. Stays the same; stays the same
- h. None of the above

To improve marketing efforts among UBC students, a company has gathered a dataset that includes student IDs and a list of all the social media apps that each student uses. The dataset is below.

ID	Apps Used
R1	Snapchat, Twitter, TikTok
R2	Snapchat, TikTok, Youtube, Instagram, Facebook
R3	Youtube, Snapchat, Tiktok, WeChat, Twitter
R4	Facebook, Pinterest, Tiktok, Twitter, Tumblr,
R5	Pinterest, Facebook, Twitter, Snapchat, Instagram, Tiktok
R6	Twitter, Facebook, Tumblr, WeChat, Pinterest,
R7	Twitter, Youtube, Facebook, Tiktok
R8	Youtube, Pinterest, Twitter, Tiktok, Snapchat,

R9	Twitter, Instagram, Facebook, Snapchat, Tiktok, Pinterest
R10	Youtube, Instagram, Tumblr, WeChat

<b>X</b>	<b>Y</b>	<b>Z</b>	<b>Confidence of <math>\{X, Y\} \rightarrow Z</math></b>
Youtube	TikTok	Snapchat	$\frac{3}{4}$ (Snapchat+Youtube+Tiktok / Youtube+Tiktok) (R2 + R3 + R8 / R2 +R3 +R7 + R8)
Youtube	TikTok	Instagram	$\frac{1}{4}$ (Instagram+YouTube+Tiktok / YouTube+Tiktok) (R2 / R2 +R3 +R7 + R8)
Youtube	TikTok	Facebook	$\frac{2}{4}$ (Facebook+YouTube+Tiktok / YouTube+Tiktok) (R2 + R7 / R2 +R3 +R7 + R8)
Snapchat	Twitter	Facebook	$\frac{2}{5}$ (Snapchat+Twitter+Facebook / Snapchat+Twitter) (R5 + R9 / R1+R3+R5+R8+R9)
Snapchat	Twitter	WeChat	$\frac{1}{5}$ (Snapchat+Twitter+Wechat / Snapchat+Twitter) (R3/ R1+R3+R5+R8+R9)
Snapchat	Twitter	Pinterest	$\frac{3}{5}$ (Snapchat+Twitter+Pinterest / Snapchat+Twitter) (R5+R8+R9 / R1+R3+R5+R8+R9)
Snapchat	Twitter	TikTok	$\frac{5}{5}$ (Snapchat+Twitter+TikTok / Snapchat+Twitter) (R1 + R3 + R5 + R8 + R9)/ R1+R3+R5+R8+R9)

Facebook	Twitter	TikTok	$\frac{4}{5}$ (Facebook + Twitter + TikTok / Facebook + Twitter) (R4 + R5 + R7 + R9 / R4 + R5 + R6 + R7 + R9)
Facebook	Twitter	Tumblr	$\frac{2}{5}$ (Facebook + Twitter + Tumblr / Facebook + Twitter) (R4 + R6 / R4 + R5 + R6 + R7 + R9)
Facebook	Twitter	Youtube	$\frac{1}{5}$ (Facebook + Twitter + YouTube / Facebook + Twitter) (R7 / R4 + R5 + R6 + R7 + R9)
Pinterest	Twitter	Tumblr	$\frac{2}{5}$ (Pinterest + Twitter + Tumblr / Pinterest + Twitter) (R4 + R6 / R4 + R5 + R6 + R8 + R9)
Pinterest	Twitter	Facebook	$\frac{4}{5}$ (Pinterest + Twitter + Facebook / Pinterest + Twitter) (R4 + R5 + R6 + R9 / R4 + R5 + R6 + R8 + R9)
Pinterest	Twitter	TikTok	$\frac{4}{5}$ (Pinterest + Twitter + TikTok / Pinterest + Twitter) (R4 + R5 + R8 + R9 / R4 + R5 + R6 + R8 + R9)
Pinterest	Twitter	Snapchat	$\frac{3}{5}$ (Pinterest + Twitter + Snapchat / Pinterest + Twitter) (R5 + R8 + R9 / R4 + R5 + R6 + R8 + R9)
Pinterest	Twitter	WeChat	$\frac{1}{5}$ (Pinterest + Twitter + WeChat / Pinterest + Twitter) (R6 / R4 + R5 + R6 + R8 + R9)

To improve marketing efforts among UBC students, a company has gathered a dataset that includes student IDs and a list of all the social media apps that each student uses.

In order for {Snapchat, Twitter, Tiktok} to be considered a frequent itemset, what must the **largest possible minimum support** be?

ID	Apps Used
R1	Snapchat, Twitter, TikTok
R2	Snapchat, TikTok, Youtube, Instagram, Facebook
R3	Youtube, Snapchat, Tiktok, WeChat, Twitter
R4	Facebook, Pinterest, Tiktok, Twitter, Tumblr,
R5	Pinterest, Facebook, Twitter, Snapchat, Instagram, Tiktok
R6	Twitter, Facebook, Tumblr, WeChat, Pinterest,
R7	Twitter, Youtube, Facebook, Tiktok
R8	Youtube, Pinterest, Twitter, Tiktok, Snapchat,
R9	Twitter, Instagram, Facebook, Snapchat, Tiktok, Pinterest
R10	Youtube, Instagram, Tumblr, WeChat

- a. 30%
- b. 40%
- c. 50%
- d. 60%
- e. 70%
- f. 80%

# Diversity

Steve is interviewing TA applicants for the next offering of CPSC 100. Amy is the next applicant Steve is going to interview. Amy earned 99% in the course last year, and wrote an impressive application. When Amy walks into the room and starts to speak, Steve immediately detects that she has an accent and decides that she must be an international student. Once the interview is finished, despite Amy doing a good job, Steve immediately throws out her application because she is an international student, and he knows that all international students are less capable of being good instructors. What type of bias is this and why?

- a. Unconscious bias; it's just something that Steve feels without reason
- b. **Conscious bias; Steve knows that he's biased and doesn't care how it affects Amy or the hiring process**
- c. Conscious bias; Steve has formed this bias from his experiences and can rationalize it
- d. Unconscious bias; Steve's immediately threw out the application for no apparent reason
- e. None of the above

## Pick 1 of 2 questions

Your mother and father, who are both admission officers for a computer science program at Mars University, are having a debate at dinner as to why there are more men than women in computer science.

Your mother claims women have the tendency to shy away from overly competitive fields, such as computer science, because they do not believe they are smart enough to be successful in this male-dominated field. On the other hand, your father claims that the reason is because admission officers believe women are too emotional, and thus not good enough to be successful in the field of computer science.

Which of the following theories can be used to describe your father's claim?

- a. **Bias**
- b. Image of computing
- c. Biological
- d. Socialization

Your mother and father, who are both admission officers for a computer science program at Mars University, are having a debate at dinner as to why there are more men than women in computer science.

Your mother claims women have the tendency to shy away from overly competitive fields, such as computer science, because they do not believe they are smart enough to be successful in this male-dominated field. On the other hand, your father claims that the reason is because admission officers believe women are too emotional, and thus not good enough to be successful in the field of computer science.

Which of the following theories can be used to describe your mother's claim?

- a. Imposter Syndrome
- b. Image of computing
- c. Biological
- d. Socialization

## Artificial Intelligence

Some researchers want to build a computer that can do Natural Language Processing (NLP) so that one day, it can help speed up their research. The researchers decide using word embeddings is disadvantageous and that a more structured approach would be better. What approach are the researchers taking when it comes to NLP? Why?

- a. Traditional approach; it requires programming the computer with specific rules to follow
- b. Recent/modern approach; it requires programming the computer with specific rules to follow
- c. Traditional approach; because it can make guesses about unfamiliar and erroneous input
- d. Recent/modern approach; because it is able to use statistical methods to compare different interpretations
- e. Traditional approach; because it is more reliable and will result in no bias for word associations
- f. Recent/modern approach; because it is more reliable and will result in no bias for word associations



### Pick 1 of the 2 questions

Meredith Gray and her friends at Seattle Grace hospital have just been told that AI algorithms will be implemented around the hospital to assist them with patient care and general duties. Meredith has been asked to decide which type of machine learning is appropriate for an AI algorithm that needs to be able to spot patterns of COVID-19 damage to the lungs from x-ray images. The hospital chief hasn't had time to label all the x-ray images but has a sample image of a COVID-19 x-ray the AI can compare with. He also wants the AI algorithm to think like a human as much as possible. What type of machine learning would you recommend to Meredith?

- a. Unsupervised learning
- b. Deep unsupervised learning
- c. Reinforcement learning
- d. Deep supervised learning
- e. Deep reinforcement learning
- f. None of the above

You are enrolled in a computer science course and the final project is to program an AI framework with a given number of hidden layers that is capable of successfully maneuvering a virtual car simulation into a single parking space. To do this, your AI algorithm gets fed large amounts of training data on the kinds of objects that exist within its environment while the input of the neural networks are the car's current position, speed, relative position to the target, etc. With positive and negative reward signals in the environment, you are able to train your AI to react to environment feedback accordingly and gradually solve the task. What is this process an example of?

- a. Unsupervised learning
- b. Supervised learning
- c. Deep supervised learning
- d. Deep reinforcement learning
- e. Deep unsupervised learning
- f. None of the above

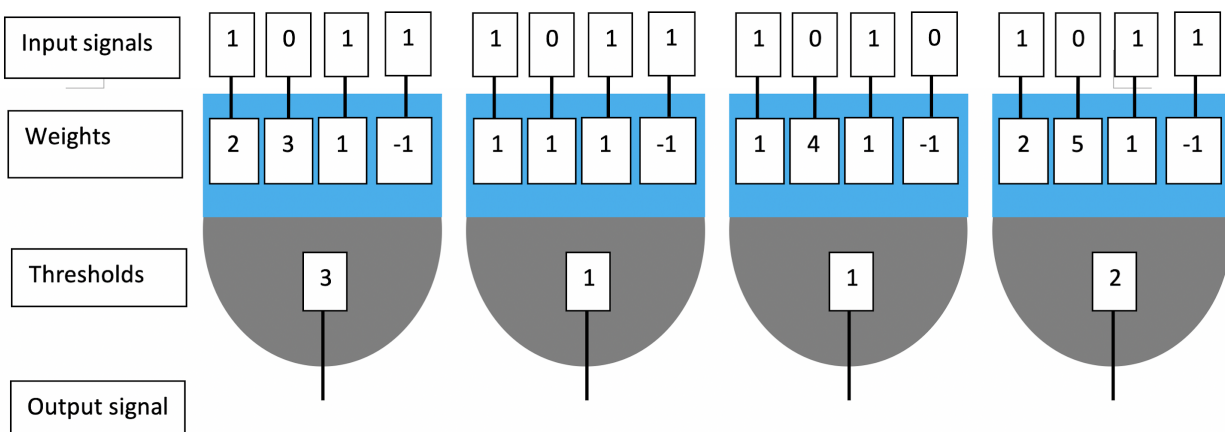
You are working for Amazon on the AI behind Alexa. One Monday morning before you've had your coffee, a student beta tester asks "Alexa, can you pull up my draft for review?" to which the AI responds, "As a citizen of Canada you are not subject to an obligatory draft." The tester gives you a cheeky grin and you pull off your glasses and pinch the skin between your brow. What went wrong here?

Note: "Obligatory draft" (also called conscription) is the mandatory enlistment of individuals into a service for the country (most often, military service).

- Pragmatics were not employed appropriately so Alexa did not pick up on clues in pronouns and adjacent words.
- Alexa made an error in parsing out the sentence structure resulting in the ambiguity.
- Clear semantic cues were missed by Alexa resulting in a misidentification of words.
- Alexa's ability to recognize speech was at fault, had the question been provided in ASCII it would have been correctly answered.
- The student is at fault, questions never come before coffee.
- None of the above

If you read **left to right**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

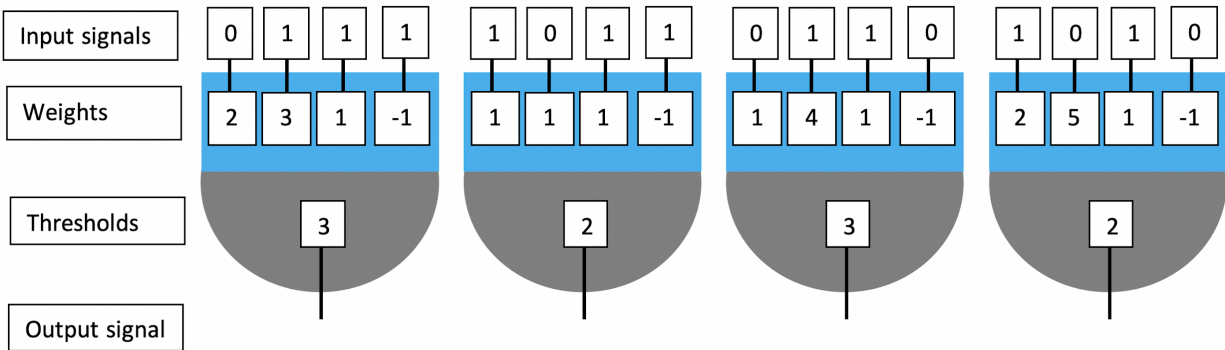
You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: 7

If you read **left to right**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

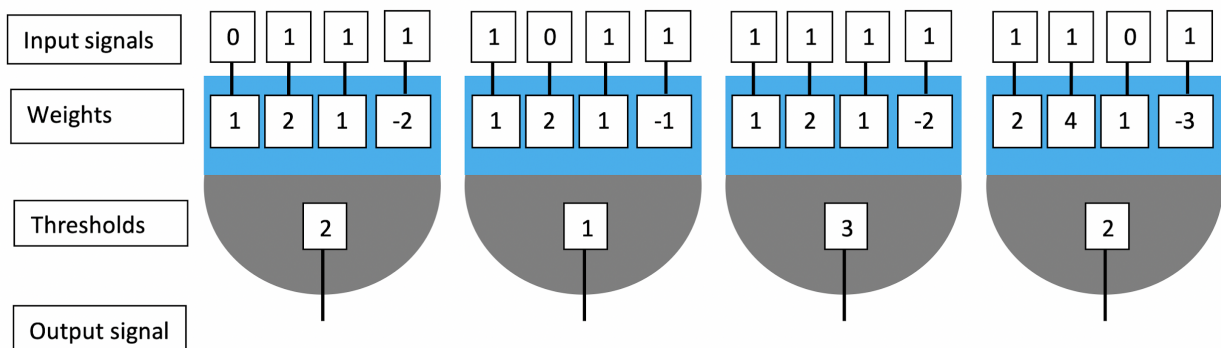
You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: B

If you read **left to right**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

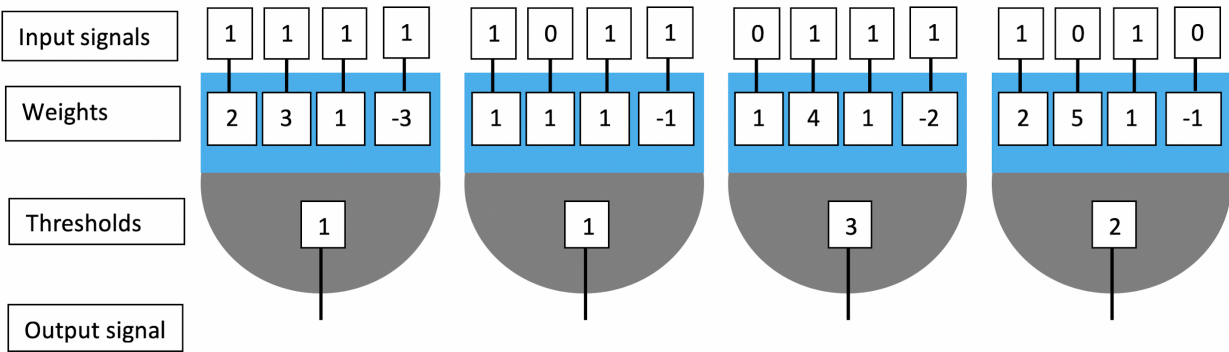
You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: 5

If you read **left to right**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

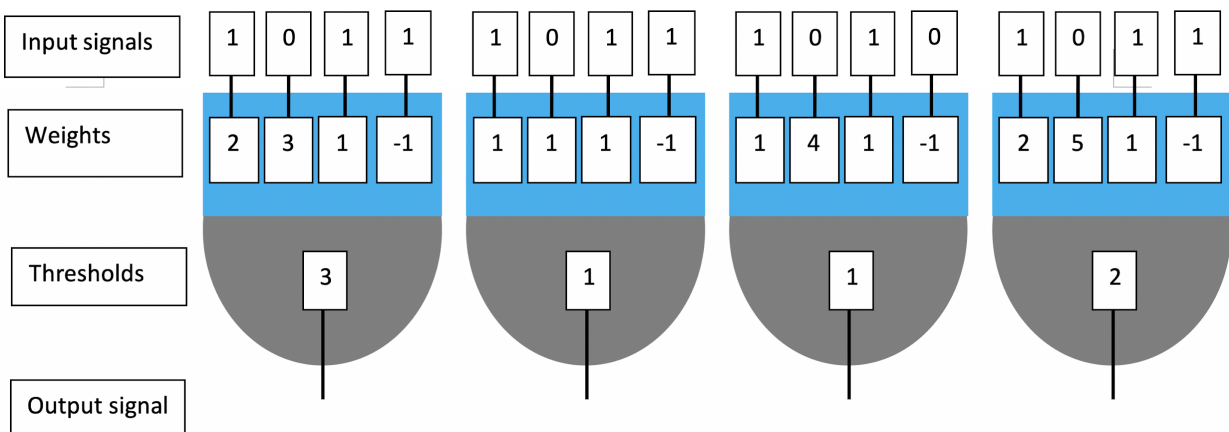
You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: F

If you read **right to left**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

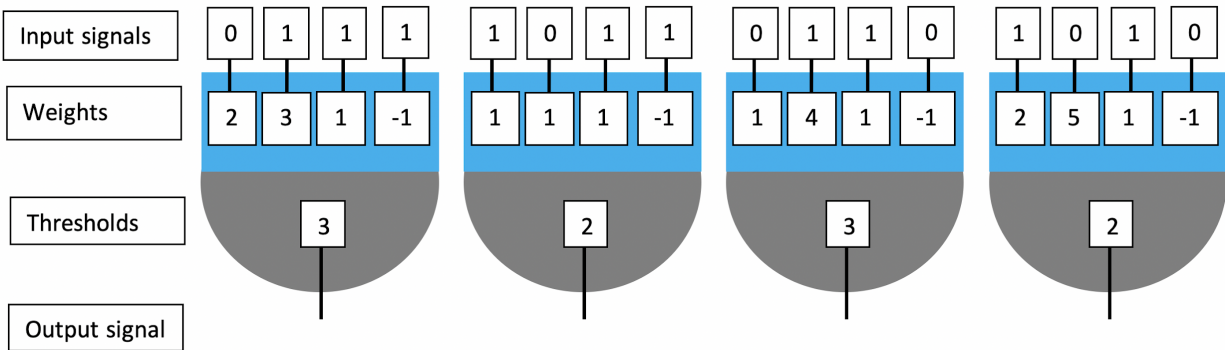
You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: E

If you read **right to left**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

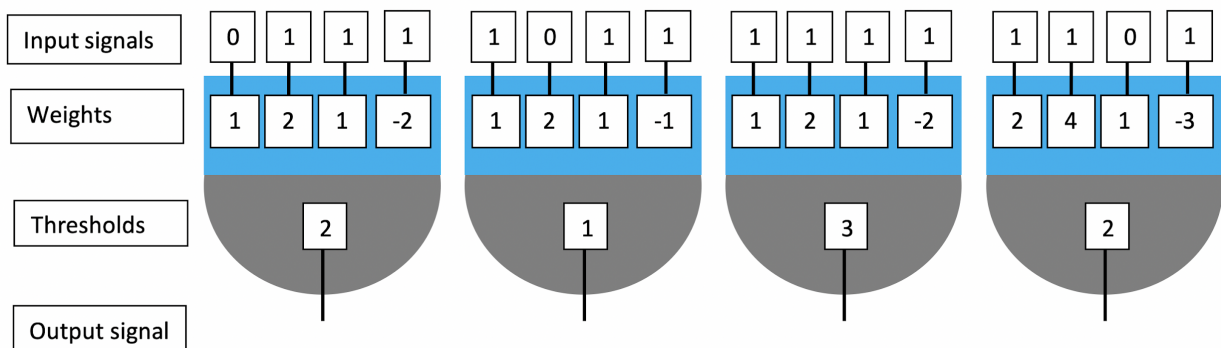
You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: D

If you read **right to left**, the outputs of the artificial neurons below provide the answer to this question in binary. If a weighted sum is greater than or equal to the threshold, then the signal should pass through.

You must translate this binary number into hexadecimal to get the correct answer to this question.



Answer: A

While in your computer science lab at UBC, you and your lab mates discover two robots, Robot 1 and Robot 2. You quickly realize that both robots are used for language translation.

Upon this realization, you and your lab mates try to figure out which machine learning algorithm each of the robots is using to look for patterns in the massive amounts of data.

You discover that Robot 1 has 2000 layers of neurons talking to each other, whereas Robot 2 has 1000 layers of neurons talking to each other. You also discover that Robot 1 is looking through the large amount of data and trying to reach a specific objective using trial and error, whereas Robot 2 is being told what to look for through labeled data.

Based on this information, what types of ML algorithms are the robots using for language translation?

- a. Robot 1: Deep unsupervised learning; Robot 2: Deep unsupervised learning
- b. Robot 1 Reinforcement learning; Robot 2: Supervised learning
- c. Robot 1: Supervised learning; Robot 2: Deep reinforcement learning
- d. Robot 1: Deep unsupervised learning; Robot 2: Deep supervised learning
- e. Robot 1: Deep reinforcement learning; Robot 2: Deep supervised learning
- f. Robot 1: Deep reinforcement learning; Robot 2: Deep unsupervised learning
- g. Robot 1: Deep reinforcement learning; Robot 2: Unsupervised learning
- h. None of the above

Haofeng calls a software company to discuss an issue he is having with his computer. Upon calling, Haofeng was able to not only discuss his computer problem; he also conversed with the computer about his day. After the conversation finished, Haofeng got an email asking about his experience with the new AI computer specialist. Haofeng was in shock as he thought he was talking to a real person the whole time.

Would this AI be considered intelligent by Searle and/or Turing's criteria?

- a. Both Searle and Turing
- b. Only Turing
- c. Only Searle
- d. Neither Searle nor Turing

# Programming

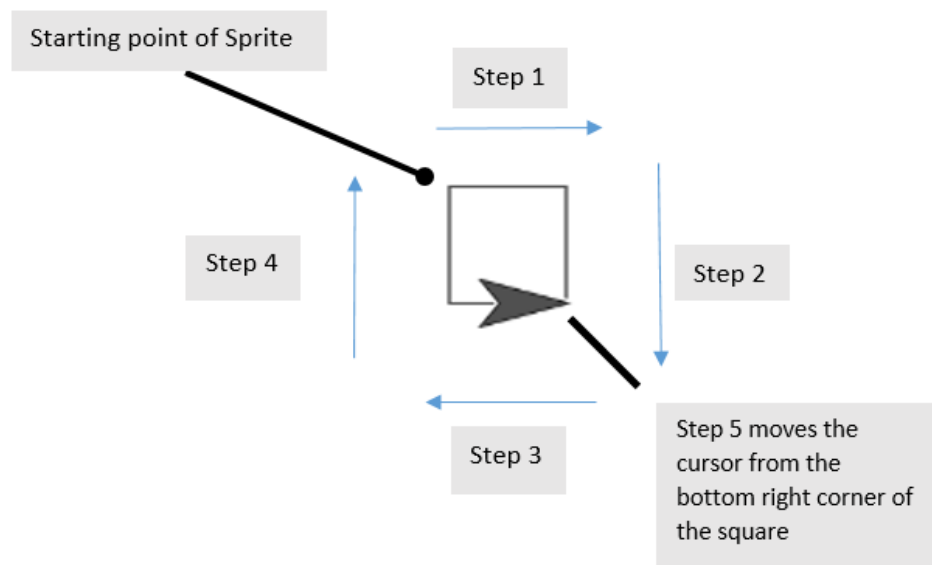
There are two custom-made blocks used in this Snap Program.

The first one **reset** does as the block name suggests, it resets the Sprite to position (0,0) facing East (i.e., 90 degrees)

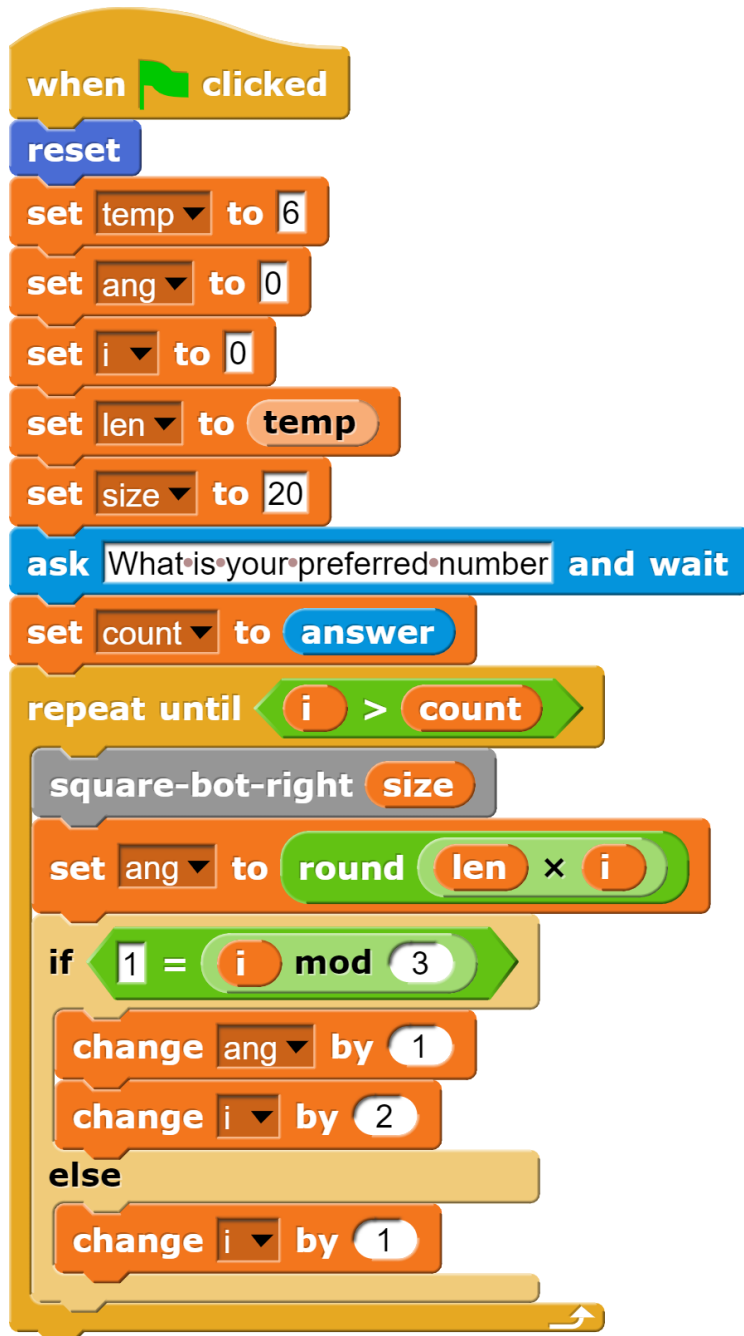


The second block **square-bot-right** consumes a side length and creates a square with the specified side-length. After the square is drawn the Sprite is posted at the bottom

right edge of the square. So,  produces



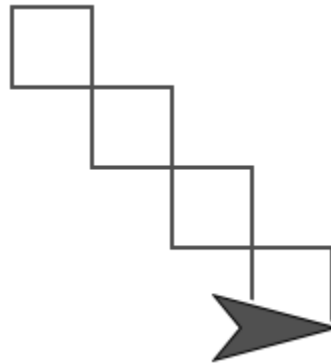
Consider the following code snippet:



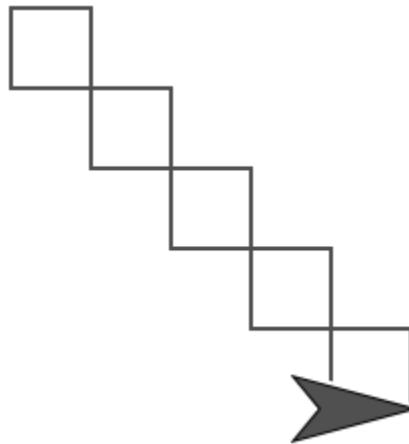


Now, consider the following options.

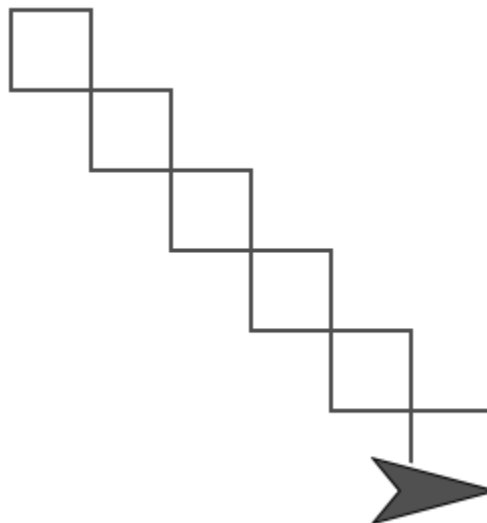
Option A



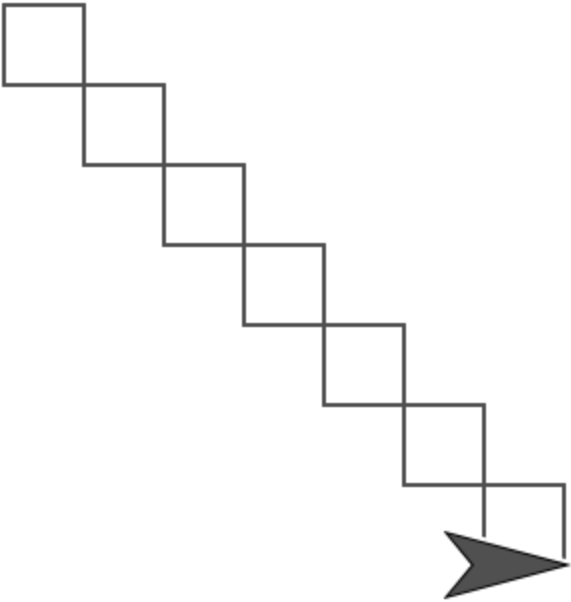
Option B



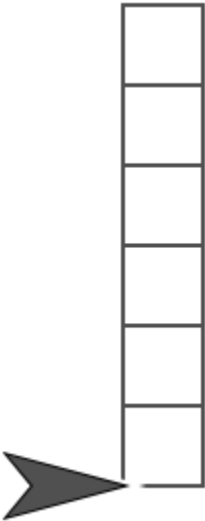
Option C



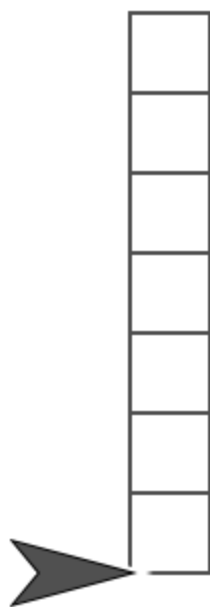
Option D



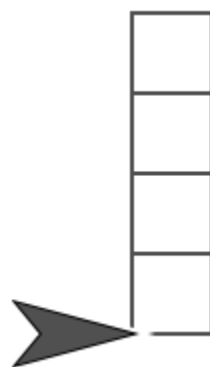
Option E



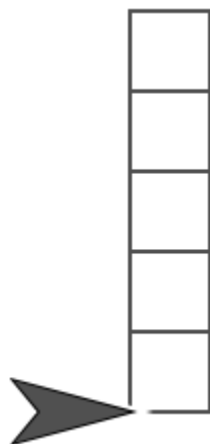
Option F



Option G



Option H



Option I



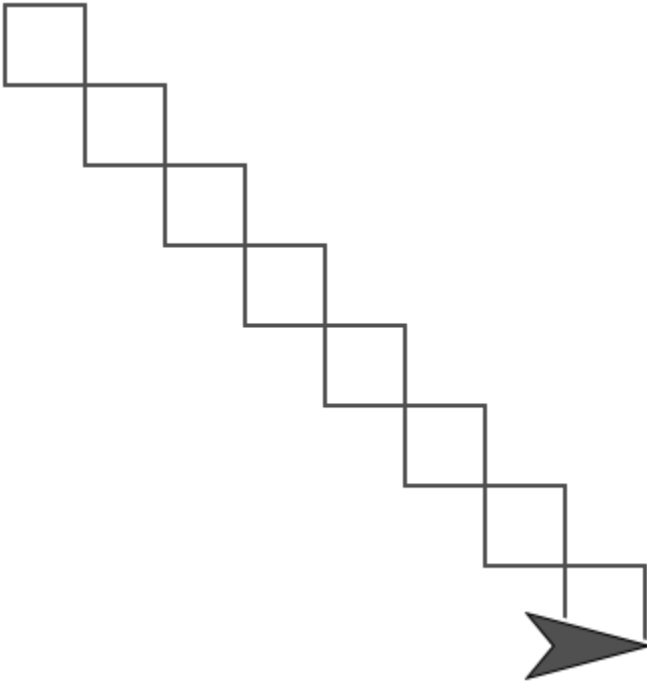
Option J



Option K



Option L



Which image most accurately represents the stage after the code above is run and the user enters 5 after the prompt? [Option A](#)

Which image most accurately represents the stage after the code above is run and the user enters 6 after the prompt? [Option B](#)

Which image most accurately represents the stage after the code above is run and the user enters 7 after the prompt? [Option C](#)

Which image most accurately represents the stage after the code above is run and the user enters 8 after the prompt? [Option C](#)

# Short Answer

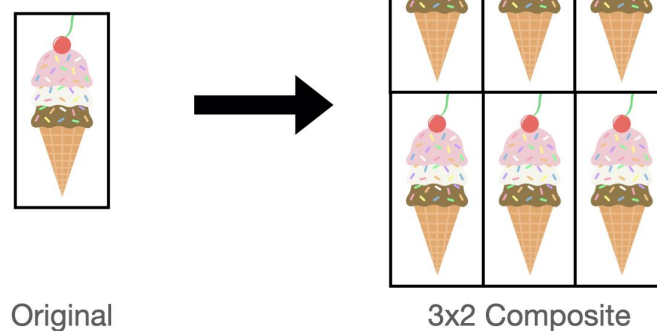
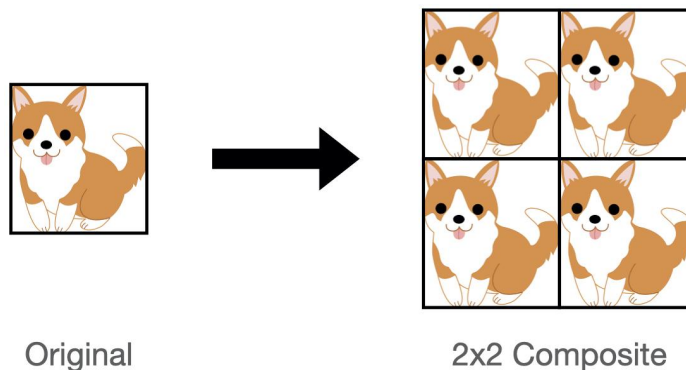
## Algorithms

Given an image, write an algorithm to create a  $n \times m$  composite of that same image.

Remember, the way a computer processes an image is pixel by pixel. Your algorithm must address the problem at the level of individual pixels. Answers that say something along the lines of “I would copy and paste the image multiple times” will be given a 0.

You **cannot** make any assumptions about the shape or the dimensions (i.e., the length or the width) of the given image. You can safely assume that  $n$  and  $m$  are going to be positive whole numbers;  $n$  and  $m$  do not have to have the same value.

Some examples of what it means to produce a composite are below.



## Setup

- Let  $n$  be the number of columns and  $m$  be the number of rows.
  - $-0.5$  for swapping the two since that violates the example
- Let `numRepetitionsRow` and `numRepetitionsColumns` be the number of times the image has been replicated in a row and in a column respectively.

## Algorithm (Option 1: the composite image is separate from the original)

1. Set `numRepetitionsRow` to 0 and `numRepetitionsColumns` to 0.
2. Start at the top left corner of the original image.
3. Copy the full row of pixels.
4. Increment `numRepetitionsColumns` to 1.
5. If `numRepetitionsColumns` is equal to  $n$ , stop. Otherwise, repeat steps 2 to 4.
6. Repeat steps 2 to 5 for each row of pixels in the image.
7. When you reach the end of the image, increment `numRepetitionsRows` by 1.
8. If `numRepetitionsRows` is equal to  $m$ , stop. Otherwise, repeat steps 2 to 7.

## Algorithm (Option 2: the original image is altered to become the composite image)

If an answer goes this route, there is a built in assumption here that  $n$  and  $m$  are at least 1.

1. Set `numRepetitionsRow` to 1 and `numRepetitionsColumns` to 1.
2. Start from the first row of pixels in the original image. Move to the end of the row.
3. If `numRepetitionsColumn` is not equal to  $n$ , copy the first row of pixels and append it to the end of the row. Increment `numRepetitionsColumn` by 1.
4. If `numRepetitionsColumn` is equal to  $n$ , move to the next row of pixels and repeat steps 2-3.
5. Once you finish processing all of the rows of pixels in the image, compare `numRepetitionsRow` against  $m$ .
6. If `numRepetitionsRow` is less than  $m$ , repeat steps 2 to 5.
7. If `numRepetitionsRow` is equal to  $m$ , stop.

Note when grading: There is no reason that an answer has to “copy” the image in the order of left to right, top to bottom. The answer could also start from top to bottom, left to right.

# Reasoning About Confidence Rules

Consider the rule  $X \rightarrow Y$ . If you know the support of  $\{X, Y\}$  and you know the confidence of  $X \rightarrow Y$ , will you be able to determine the confidence of  $Y \rightarrow X$ . Why?

You are **not** able to determine the confidence of  $Y \rightarrow X$ . When calculating confidence, the denominator of this value is the number of transactions that  $Y$  appears in (i.e., you need to know the support of  $\{Y\}$ ).

The confidence of  $X \rightarrow Y$  only tells us given the number of transactions  $X$  appears in, how many of them also contain  $Y$ . This number is different from the support of  $\{X, Y\}$ .

Support tells you how many times these two items appear together but it is possible for  $Y$  to appear in transactions without  $X$ . Therefore, there is no way to tell the exact number of transactions that  $Y$  appears in from the support of  $\{X, Y\}$  and the confidence of  $X \rightarrow Y$ .



# Connecting the Dots

In the context of what you have learned this term, what is the relationship between the concepts below.

- Algorithm
- Data
- Bias
- Word Embeddings
- Reinforcement Learning
- Programming
- Encryption
- Computer Animation

**Data and Bias and Algorithm:** The quality of the data influences the quality of an algorithm. If the data contains bias, this bias will be reflected in the algorithms produced. Algorithmic Bias can result from biased data or from bias that exists in the creator of the algorithm.

**Algorithm and Programming:** Algorithms are the building blocks of programs. Programming is the process by which algorithms are translated into languages which are close to what computers understand. High level programming languages such as Snap are then translated/compiled into machine code which is the language a computer understands.

**Reinforcement Learning and Algorithms and Word Embeddings:** Word embeddings and reinforcement learning are both words that related to machine learning such that we can employ machine learning algorithms to teach computers how to create word embeddings using previous data and then integrate these word embeddings into other algorithms such as those that can translate between languages. A way that we can teach computers how to do this is through reinforcement learning, learning by trial and error through the use of ANNs. A computer would need to be programmed to learn how to employ these algorithms.

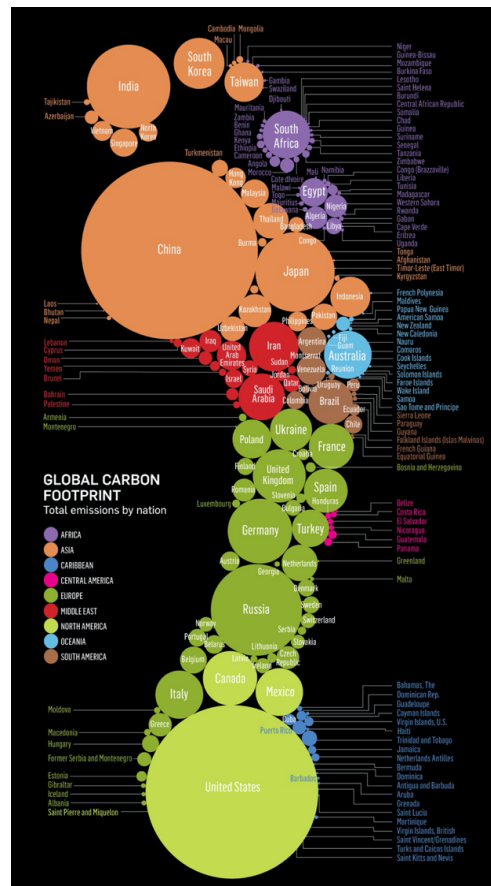
**Programming/Algorithm/Data and Encryption:** Encrypting data that is accessible by others is very important. One of the challenges of using the internet is that when data is sent, people can listen in. There are many encryption techniques that have been programmed to allow user's data to remain private.

**Programming/Algorithm/Data and Computer Animation:** All data in a computer is stored in binary, Computer Animation is a way by which data is brought to life, it takes on quality of movement. There are various algorithms that have been programmed that allow for animation to occur.

# Visualization

You were also introduced to the following principles for creating good infographics.

1. **Simplicity** – minimal text, clear message, avoid clutter with text and charts
2. **Consistency** – layout and design elements should be consistent. 2 to 3 font sizes, color scheme
3. **Visibility** – appropriate font sizes, colors that contrast
4. **Navigability (structure)** – clear order to follow, use scale/proportion to emphasize key points/headings
5. **Suitability** – right data for message, right graphic for the message, right metaphors for the audience



State how the infographic adheres to or violates each design principle. Explain your answer. In your explanation, particularly in the case where a principle is violated, suggest specific changes that can be made to fix the violation.

**Simplicity** – minimal text, clear message, avoid clutter with text and charts

Adherence: The message is very clear in that it is trying to show the contribution of each nation to the global carbon footprint.

Violation: Too much of text, to the point that we cannot easily make sense of what it says since it is so cluttered. This infographic is very cluttered since it tries to represent an abundance of countries with too many circles and labels for the smaller ones.

Fix: remove the labels for the smaller circles as this would reduce the amount of text on the infographic.

**Consistency** – layout and design elements should be consistent. 2 to 3 font sizes, color scheme

Adheres: Same font type, color and size. Consistency across continents. That is the text of the labels be the same colour as the circle they are labeling and all circles representing the same continent are the same colour.

**Visibility** – appropriate font sizes, colors that contrast

Adheres: colours of text and circles that contrast against a black background.

Violates: font size is too small (e.g., it violates this principle as the font size is way too small to be legible without zooming in). With that said though, If this were to be printed, a passerby would certainly *not* be able to read the text on the infographic from 3ft away (a way to check visibility discussed in class). Light green and dark green colours to represent 2 different continents while being located next to each other decreases visibility.

Fix: the author should remove the labels for the smaller circles all together -- this would also address the issue of simplicity.

**Navigability (structure)** – clear order to follow, use scale/proportion to emphasize key points/headings

**Adheres:** Little structure exists. The fact that the title is bolded and in large font helps with identifying and emphasizing the main idea of the infographic (global carbon footprint). As well, the legend and colour coding used provided makes it very easy to identify the different continents represented by each colour. The fact that countries from the same continent are grouped together in the footprint also helps with navigability. Scales are proportionate to the size of the carbon footprint of the country.

**Suitability** – right data for message, right graphic for the message, right metaphors for the audience

Adheres: Foot metaphor for carbon footprint. Grouping nations by continent helps the audience visualize the top contributors to this problem.

Violates: The unit and amount of total emissions reported is not shown so the title "total emissions by nation" is not technically correct.

Fix: The title should read "relative proportions of emissions by nation" that way, the author would not need to show numbers and using area to represent the data would be sufficient.