

# A Benchmark of Question-Answering Problems about Two-Dimensional Abstract Diagrams

Ernest Davis

August 18, 2024

## Summary

I have written code to generate random diagrams tiles with polygonal shapes and random questions about these within 24 different general templates, with the correct answers. I believe that the problems are mostly easy for people and I conjecture that the current (August 2024) generation of visual QA systems would not succeed at human levels. I am looking for collaborators to take this project further.

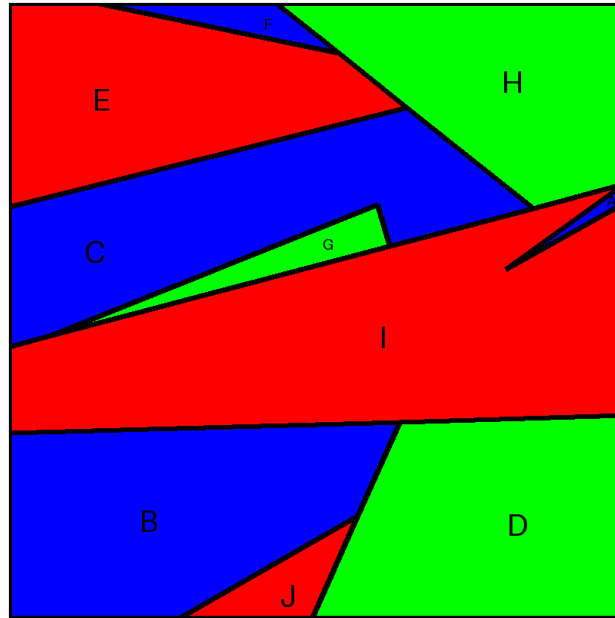
This document is structured as follows:

1. Examples of ten diagrams with questions (section 1). The first diagram has examples of all 24 types of questions: the remaining nine diagrams have 5 questions each.
2. Discussion (section 2 of the current state of the benchmark: the characteristics of the diagrams (section 2.1), the questions (section 2.2) and the state of the code (section 2.3).
3. Discussion (section 3) of various options for taking this forward.

## 1 Examples

Let me start by showing some examples, with associated questions and answers.

## 1.1 Example 1



The diagram was created with the function call `TestQuestions.randomSetup(9)`. The argument to `randomSetup` is a seed for the random number generator.

All example questions were generated with the function call `RandomQuestion.displayRandomQuestion(<question number>, 1)`. The second argument is the random seed.

**Question 1:** Which regions border region B along an edge?

**Answer:** I, J, D

**Question 2:** If you draw a straight line segment from a point in the interior of region E to a point in the interior of B, which other regions might it go through?

**Answer:** G, I, C

**Question 3:** Which if any of the regions are not convex?

**Answer:** I, C

**Question 4:** Let  $p$  be the rightmost vertex of J. Suppose that someone starts at  $p$  and goes counterclockwise around B until they have returned to  $p$ . What regions do they pass on their right in sequence? For this question, include the outside of the frame.

**Answer:** [D, I, Outside, J]

**Question 5:** Which pairs of regions, if any, meet at a vertex but not along an edge?

**Answer:** (H, A)

**Question 6:** Which pairs of regions, if any, meet along two or more disconnected edges? Do not include the outside of the frame.

**Answer:** (I, C)

**Question 7:** Which regions, if any, meet the outside of the frame along two or more disconnected edges?

**Answer:** I

**Question 8:** Which regions have 4 sides?

**Answer:** H, D

**Question 9:** How many sides does region B have?

**Answer:** 5

**Question 10:** Let  $p$  be the meeting point of regions E and C with the outside of the frame. Let  $u$  be the rightmost vertex of F. Let  $v$  be the meeting point of regions C and I with the outside of the frame. Let  $w$  be the meeting point of regions H, I, and A with the outside of the frame. Sort  $u, v, w$  in increasing order of distance from  $p$ . **Answer:** ['v', 'u', 'w']

**Question 11:** Region H has 4 vertices: the vertex at the top right of the overall diagram; (2) the leftmost vertex of H; (3) the meeting point of H, C, and I; (4) the meeting point of H, I, and A with the outside of the frame. Sort these in increasing order by the size of the interior angle at each corner.

**Answer:** [2, 1, 4, 3]

**Question 12:** Let  $m$  be the edge of J that meets region B. Suppose  $m$  is extended in both directions along straight line  $L$ . Which are the regions  $R$  in the diagram for which  $L$  passes through the interior of  $R$ ? (Do not include a region  $R$  if  $L$  only aligns with the boundary of  $R$ , and does not go through  $R$ 's interior)

**Answer:** ['D', 'I']

**Question 13:** Let  $p$  be the leftmost vertex of A. Which regions meet at  $p$ ?

**Answer:** ['A', 'I']

**Question 14:** Let  $U$  be the union of regions D and J. How many sides does  $U$  have?

**Answer:** 5

**Question 15:** Let  $U$  be the union of regions D and J. Which of the labelled regions A-J does  $U$  border on an edge?

**Answer:** B, I

**Question 16:** Let  $N$  be the union of regions E and C. Sort regions  $[N, A, H]$  in order of increasing area:

**Answer:** [A, H, N]

**Question 17:** Find all pairs of regions  $\{X, Y\}$  with  $X \neq Y$  such that the union of  $X$  and  $Y$  is convex. (There may be no such pairs.)

**Answer:** (A, I), (J, B), (F, E), (G, C)

**Question 18:** Let  $p$  be the rightmost vertex of J. Let  $q$  be the leftmost vertex of F. If you travel in a straight line from  $p$  to  $q$ , what regions do you pass through in the interior, in sequence? (If you enter and exit a region more than once, list it multiple times.)

**Answer:** [B, I, G, C, E]

**Question 19:** Let  $p$  be the topmost vertex of J. Suppose that you travel in a vertical line downward from  $p$  until you reach the bottom of the frame. What regions do you pass through in the interior, in sequence? (If you enter and exit a region more than once, list it multiple times.)

**Answer:** [D]

**Question 20:** Let  $u$  be the bottommost vertex of F. Let  $v$  be the bottommost vertex of G. Let  $w$

be the vertex of A with the widest angle. Order u, v, and w in order bottom to top.

**Answer:** [v, w, u]

**Question 21:** Let u be the meeting point of regions E and C with the outside of the frame. Let v be the vertex of H with the widest angle. Let w be the top left vertex of E. Suppose that someone travels from u to v to w back to u. Is this cycle clockwise or counterclockwise?

**Answer:** counterclockwise

**Question 22:** Let p be the vertex of F with the widest angle. Let q be the vertex at the top right of the overall diagram. Let u be the topmost vertex of J. Let v be the meeting point of regions C and I with the outside of the frame. Does the line segment between p and q cross the line segment between u and v?

**Answer:** No

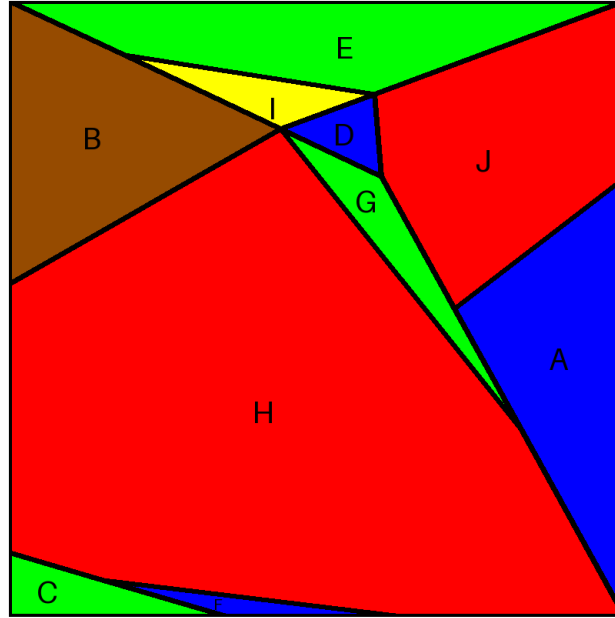
**Question 23:** List all pairs of regions  $\langle X, Y \rangle$  such that the bottom vertex of X is the same as the top vertex of Y.

**Answer:** None

**Question 24:** Which region is closer to E: B or G? (Consider the distance between two regions to be the distance between their closest points.)

**Answer:** G

## 1.2 Example 2



The diagram was created with the function call `TestQuestions.randomSetup(1)`.

The questions here and in the remaining examples were generated by a call to function `RandomQuestions.randomQuestions(5, <Example>)`. The first argument is the number of questions to be generated; the second is the random seed.

**Question 22:** Let  $p$  be the meeting point of regions E, I, D, and J. Let  $q$  be the rightmost vertex of G. Let  $u$  be the rightmost vertex of F. Let  $v$  be the vertex at the bottom left of the overall diagram. Does the line segment between  $p$  and  $q$  cross the line segment between  $u$  and  $v$ ?

**Answer:** No

**Question 1:** Which regions border region B along an edge?

**Answer:** I, E, H

**Question 23:** List all pairs of regions  $\{X, Y\}$  such that the bottom vertex of X is the same as the top vertex of Y.

**Answer:** [(I, G), (I, H), (E, D)]

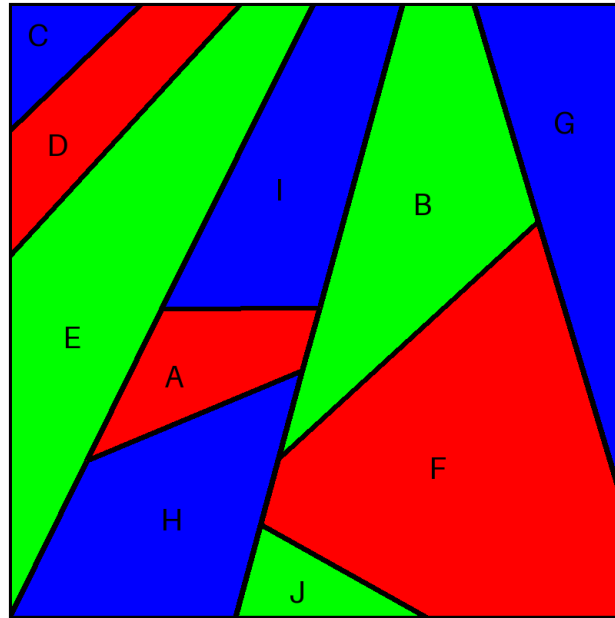
**Question 7:** Which regions, if any, meet the outside of the frame along two or more disconnected edges?

**Answer:** H

**Question 15:** Let  $U$  be the union of regions B and E. Which of the labelled regions A-J does  $U$  border on an edge?

**Answer:** I, J, H

### 1.3 Example 3



The diagram was created with the function call `TestQuestions.randomSetup(2)`.

**Question 23:** List all pairs of regions  $\langle X, Y \rangle$  such that the bottom vertex of  $X$  is the same as the top vertex of  $Y$ .

**Answer:** None

**Question 13:** Let  $p$  be the leftmost vertex of  $I$ . Which regions meet at  $p$ ?

**Answer:** ['I', 'E', 'A']

**Question 18:** Let  $p$  be the vertex of  $G$  with the widest angle. Let  $q$  be the topmost vertex of  $H$ . If you travel in a straight line from  $p$  to  $q$ , what regions do you pass through in the interior, in sequence? (If you enter and exit a region more than once, list it multiple times.)

**Answer:** [G, B]

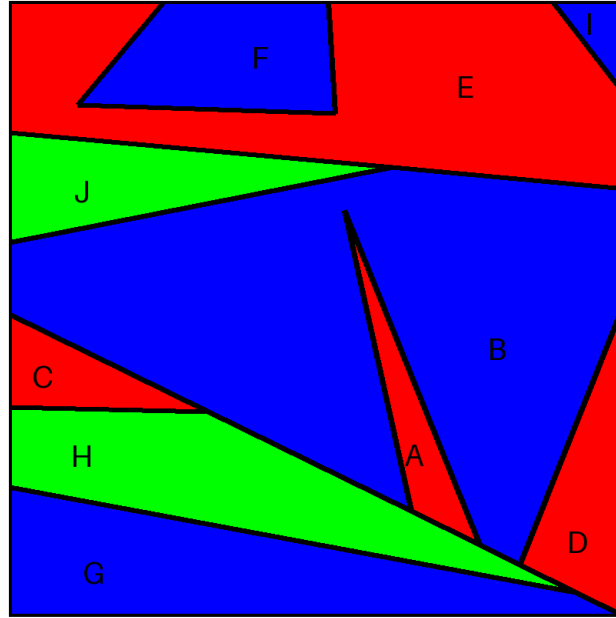
**Question 16:** Sort regions [I, F, C, D] in order of increasing area:

**Answer:** [C, D, I, F]

**Question 15:** Let  $U$  be the union of regions  $B$  and  $H$ . Which of the labelled regions  $A$ - $J$  does  $U$  border on an edge?

**Answer:** I, G, A, E, F, J

## 1.4 Example 4



The diagram was created with the function call `TestQuestions.randomSetup(3)`.

**Question 23:** List all pairs of regions  $\{X, Y\}$  such that the bottom vertex of  $X$  is the same as the top vertex of  $Y$ .

**Answer:** None

**Question 22:** Let  $p$  be the vertex of  $F$  with the widest angle. Let  $q$  be the vertex at the bottom left of the overall diagram. Let  $u$  be the vertex of  $D$  with the widest angle. Let  $v$  be the top left vertex of  $J$ . Does the line segment between  $p$  and  $q$  cross the line segment between  $u$  and  $v$ ?

**Answer:** Yes

**Question 1:** Which regions border region  $I$  along an edge?

**Answer:** E

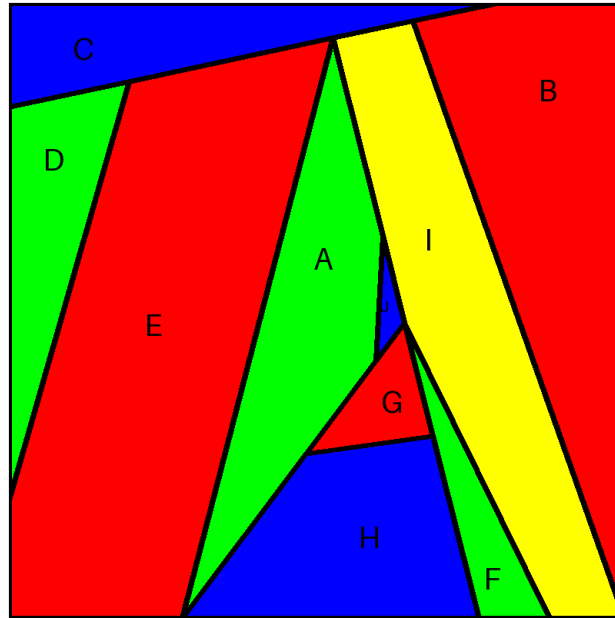
**Question 4:** Let  $p$  be the vertex at the top left of the overall diagram. Suppose that someone starts at  $p$  and goes clockwise around  $E$  until they have returned to  $p$ . What regions do they pass on their left in sequence? For this question, include the outside of the frame.

**Answer:** [Outside, F, Outside, I, Outside, B, J, Outside]

**Question 13:** Let  $p$  be the bottom right vertex of  $F$ . Which regions meet at  $p$ ?

**Answer:** ['F', 'E']

## 1.5 Example 5



The diagram was created with the function call `TestQuestions.randomSetup(4)`.

**Question 21:** Let  $u$  be the bottom left vertex of  $E$ . Let  $v$  be the vertex at the top right of the overall diagram. Let  $w$  be the vertex of  $D$  with the widest angle. Suppose that someone travels from  $u$  to  $v$  to  $w$  back to  $u$ . Is this cycle clockwise or counterclockwise?

**Answer:** counterclockwise

**Question 3:** Which if any of the regions are not convex?

**Answer:** None

**Question 19:** Let  $p$  be the bottommost vertex of  $D$ . Suppose that you travel in a horizontal line to the right from  $p$  until you reach the right side of the frame. What regions do you pass through in the interior, in sequence? (If you enter and exit a region more than once, list it multiple times.)

**Answer:** [E, A, H, F, I, B]

**Question 20:** Let  $u$  be the bottom left vertex of  $E$ . Let  $v$  be the rightmost vertex of  $F$ . Let  $w$  be the vertex of  $H$  with the sharpest angle. Order  $u$ ,  $v$ , and  $w$  in order left to right.

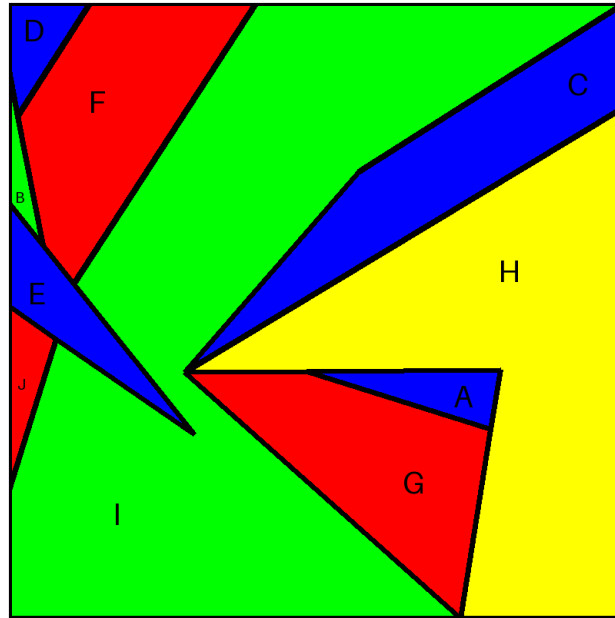
**Answer:** [u, w, v]

**Question 24:** Which region is closer to  $B$ :  $G$  or  $D$ ? (Consider the distance between two regions to be the distance between their closest points.)

**Answer:** G



## 1.6 Example 6



The diagram was created with the function call `TestQuestions.randomSetup(5)`.

**Question 22:** Let  $p$  be the leftmost vertex of  $A$ . Let  $q$  be the bottommost vertex of  $E$ . Let  $u$  be the meeting point of regions  $F$ ,  $E$ , and  $I$ . Let  $v$  be the topmost vertex of  $C$ . Does the line segment between  $p$  and  $q$  cross the line segment between  $u$  and  $v$ ?

**Answer:** No

**Question 8:** Which regions have 4 sides?

**Answer:** C, G, D

**Question 6:** Which pairs of regions, if any, meet along two or more disconnected edges? Do not include the outside of the frame.

**Answer:** (H, G)

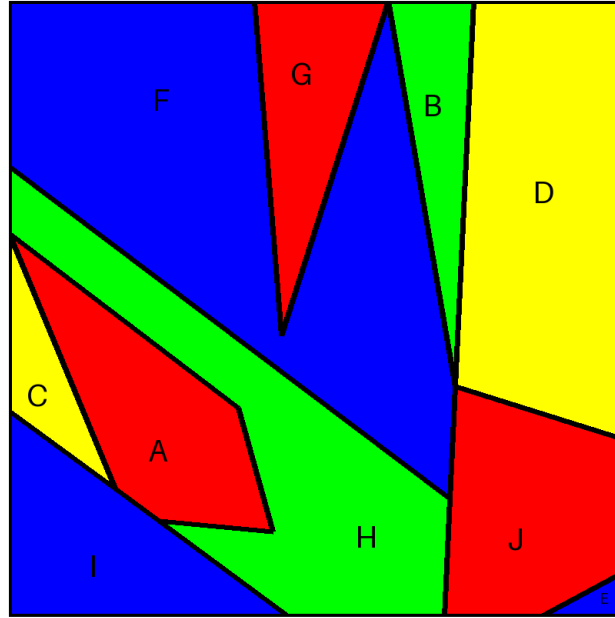
**Question 7:** Which regions, if any, meet the outside of the frame along two or more disconnected edges?

**Answer:** I

**Question 9:** How many sides does region B have?

**Answer:** 3

## 1.7 Example 7



The diagram was created with the function call `TestQuestions.randomSetup(6)`.

**Question 2:** If you draw a straight line segment from a point in the interior of region A to a point in the interior of I, which other regions might it go through?

**Answer:** H, C

**Question 12:** Let  $m$  be the edge of I that meets regions H, A, and C. Suppose  $m$  is extended in both directions along straight line  $L$ . Which are the regions  $R$  in the diagram for which  $L$  passes through the interior of  $R$ ? (Do not include a region  $R$  if  $L$  only aligns with the boundary of  $R$ , and does not go through  $R$ 's interior)

**Answer:** None

**Question 14:** Let  $U$  be the union of regions A and H. How many sides does  $U$  have?

**Answer:** 6

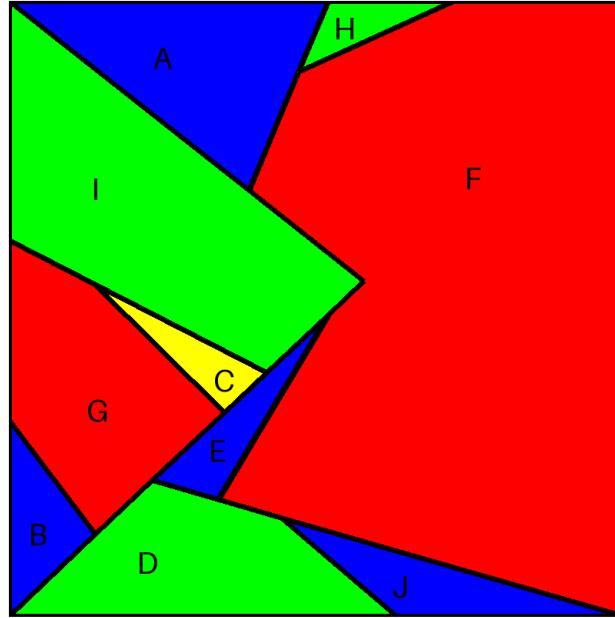
**Question 3:** Which if any of the regions are not convex?

**Answer:** H, F

**Question 22:** Let  $p$  be the meeting point of regions F, H, and J. Let  $q$  be the rightmost vertex of C. Let  $u$  be the top left vertex of G. Let  $v$  be the meeting point of regions B and D with the outside of the frame. Does the line segment between  $p$  and  $q$  cross the line segment between  $u$  and  $v$ ?

**Answer:** No

## 1.8 Example 8



The diagram was created with the function call `TestQuestions.randomSetup(7)`.

**Question 2:** If you draw a straight line segment from a point in the interior of region C to a point in the interior of J, which other regions might it go through?

**Answer:** G, E, F

**Question 12:** Let  $m$  be the rightmost edge of F. Suppose  $m$  is extended in both directions along straight line  $L$ . Which are the regions  $R$  in the diagram for which  $L$  passes through the interior of  $R$ ? (Do not include a region  $R$  if  $L$  only aligns with the boundary of  $R$ , and does not go through  $R$ 's interior)

**Answer:** None

**Question 14:** Let  $U$  be the union of regions I and E. How many sides does  $U$  have?

**Answer:** 7

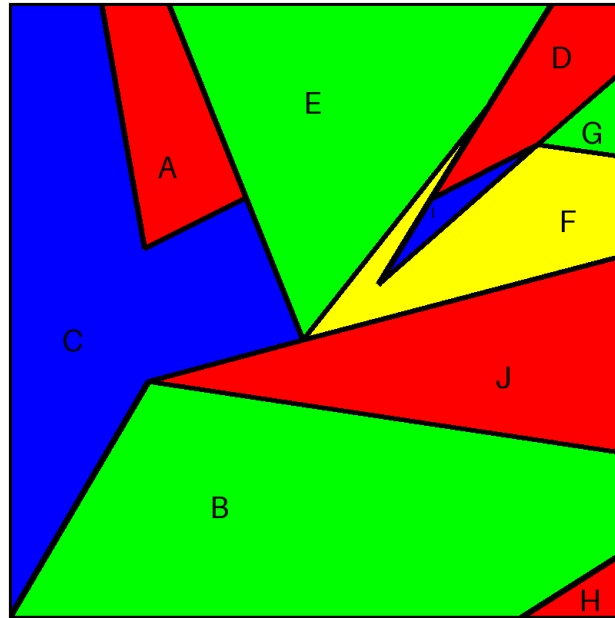
**Question 3:** Which if any of the regions are not convex?

**Answer:** F

**Question 22:** Let  $p$  be the meeting point of regions H and F with the outside of the frame. Let  $q$  be the bottommost vertex of I. Let  $u$  be the meeting point of regions B and G with the outside of the frame. Let  $v$  be the topmost vertex of J. Does the line segment between  $p$  and  $q$  cross the line segment between  $u$  and  $v$ ?

**Answer:** No

## 1.9 Example 9



The diagram was created with the function call `TestQuestions.randomSetup(8)`.

**Question 7:** Which regions, if any, meet the outside of the frame along two or more disconnected edges?

**Answer:** B

**Question 8:** Which regions have 3 sides?

**Answer:** G, J, I, H

**Question 5:** Which pairs of regions, if any, meet at a vertex but not along an edge?

**Answer:** (F, C), (E, J), (I, G)

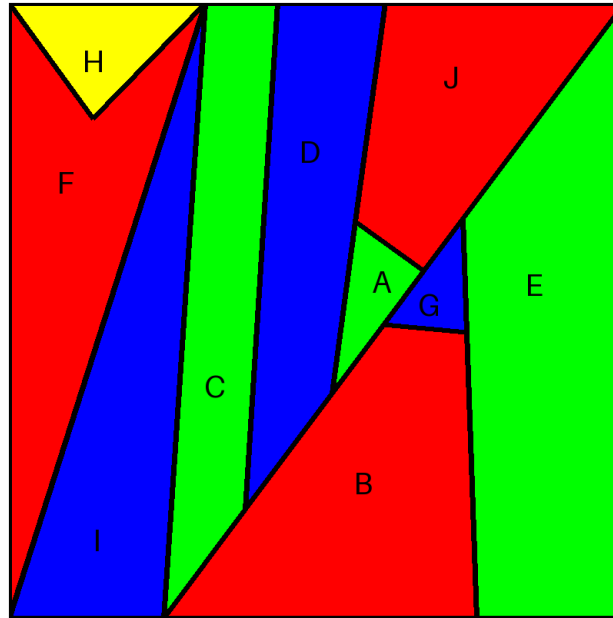
**Question 16:** Sort regions [A, J, C] in order of increasing area:

**Answer:** [A, J, C]

**Question 1:** Which regions border region B along an edge?

**Answer:** J, H, C

## 1.10 Example 10



The diagram was created with the function call `TestQuestions.randomSetup(10)`.

**Question 2:** If you draw a straight line segment from a point in the interior of region I to a point in the interior of J, which other regions might it go through?

**Answer:** D, A, C

**Question 8:** Which regions have 4 sides?

**Answer:** E, D, F, J, B, C

**Question 3:** Which if any of the regions are not convex?

**Answer:** F

**Question 12:** Let  $m$  be the edge of D that meets regions A and J. Suppose  $m$  is extended in both directions along straight line  $L$ . Which are the regions  $R$  in the diagram for which  $L$  passes through the interior of  $R$ ? (Do not include a region  $R$  if  $L$  only aligns with the boundary of  $R$ , and does not go through  $R$ 's interior)

**Answer:** ['B']

**Question 13:** Let  $p$  be the vertex of B with the widest angle. Which regions meet at  $p$ ?

**Answer:** ['A', 'B', 'G']

## 2 The Dataset

Code has been written to generate a dataset with a randomly constructed diagrams and, for each diagram, randomly constructed questions from 24 different general templates together with their correct answers, as illustrated in

Given the range of possibilities, my guess would be that, if one were to generate 800 random diagrams each with 10 regions, there would be few pairs if any that were not significantly different qualitatively. As a point of comparison, note that there are over a million planar graphs with 10 vertices, so by the birthday paradox, one would need a set of a thousand randomly generated to have two that are identical. (This argument should not be taken seriously for a number of reasons in opposite directions. On the one hand, two maps that differ only in one bordering relationship are not very different; and, though the procedure for generating random maps can generate any planar graph, there is certainly no reason to expect that it generates them equiprobably. On the other hand, two diagrams may have the same underlying connection graph and look very different qualitatively, and give very different answers to the questions.)

### 2.1 The Diagrams

Every diagram consists in a partition of an overall rectangle into irregular polyhedral tiles. In the examples in section 1, the rectangles are all squares and every diagram contains exactly 10 regions. However both the aspect ratio of the rectangles and the number of regions is a parameter of a function call.

Every region in the diagram is topologically homeomorphic to the unit disk; in other words, the boundary of every region is a topologically closed curve. Any number of regions may meet at a vertex. Two regions may meet along a consecutive series of edges (for example, regions A and B in example 4) or along multiple disconnected edges (for example, regions C and I in Example 1.) Most regions are convex and have between 3 and 5 sides, but there are some non-convex regions (apparently about 1 in 10, with the current parameter settings) and some regions with more than 5 sides. Regions are brightly colored blue, red, green, yellow, brown, and possibly gray.<sup>1</sup> The coloring observes the constraints that no two regions that border on an edge have the same color, and when easily achieved, that no two regions that share a vertex have the same color. (In the examples in section 1, the second condition is always achieved.) The regions are labelled with letters starting with 'A' in random order.

The intent of the benchmark is to create an overall map that humans can easily read with questions that they can easily answer. Therefore the code is written to avoid placing two vertices very close together, and to avoid generating angles very close to  $0^\circ$  or to  $180^\circ$ . Nonetheless, with the current parameters, we do end up with some very narrow regions that require letter labels in a smaller size that can be somewhat hard to see (e.g. region J in example 5).

The map is generated by iteratively choosing a region  $R$  at random, with probability proportional to  $\text{area}(R)$ , and then splitting it into two parts using one of the splitting methods shown in figure 1. Table 1 shows the pseudo-code.

Once the face has been selected, the random choice of a splitting method involves (a) choosing the method type, following a fixed, non-uniform distribution; (b) choosing the parameters of the method (the vertices/edges involved; the positions along the chosen edges; and the number and locations of intermediate points along the path).

---

<sup>1</sup>Of course, by the four-color theorem, the first four of these should suffice. The algorithm I have implemented certainly requires no more than six colors, and may in fact require no more than five.

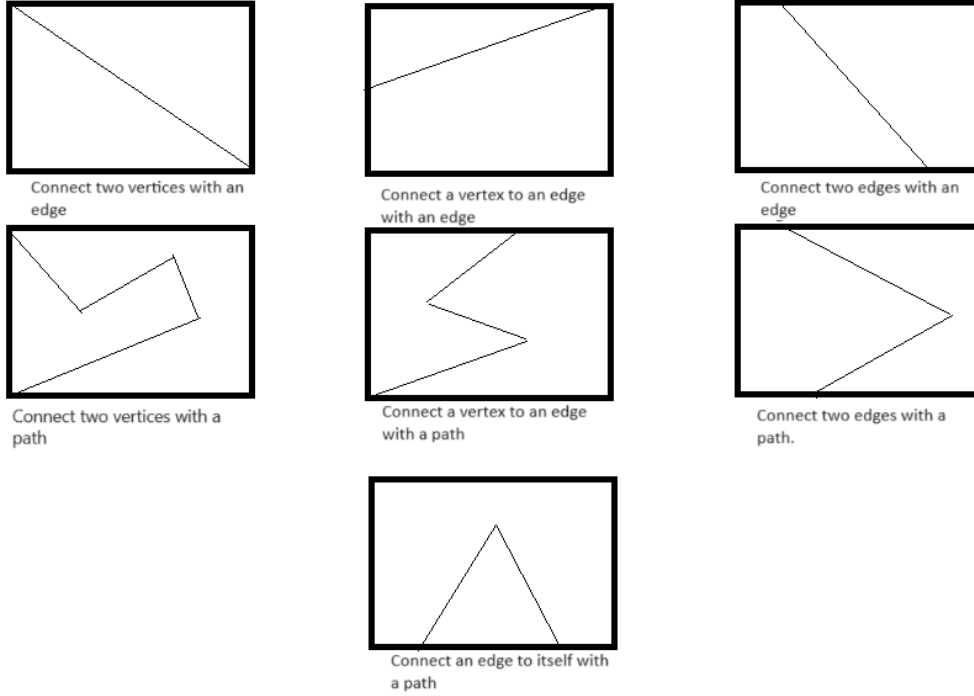


Figure 1: Ways to split a region

```

/* nRegions is the number of regions in the output map.
xSpan, ySpan are the dimensions of the overall rectanble */
function BuildMap(nRegions,xSpan,ySpan) {
  diagram  $\leftarrow$  { rectangle(xSpan,ySpan) }
  for i = 1:nRegions-1 {
    f  $\leftarrow$  randomly choose a region  $R$  in diagram,
                     with probability proportional to  $\text{area}(R)$ 
    randomly apply a splitting method to f generating new faces f1,f2
    diagram  $\leftarrow$  diagram  $\cup$  { f1,f2 }  $\setminus$  { f }
  }
  return diagram
}

```

Table 1: Pseudo-code for constructing diagram

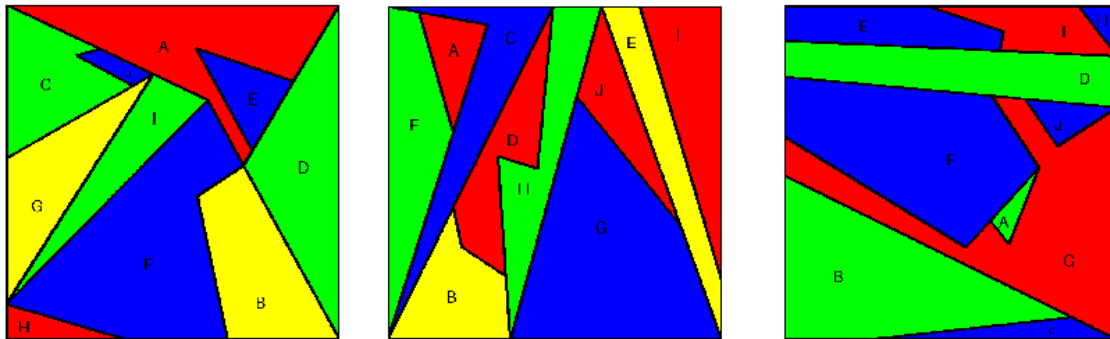


Figure 2: Complex diagram. Generated using uniform probabilities across the seven methods with random seeds 1, 2, and 3

Let  $e(f)$  be the number of edges of region  $f$ . Suppose that you split region  $f$  into  $f1$  and  $f2$ . If you use an edge-splitting method, then you are guaranteed that  $e(f1) \leq e(f) + 1$ , that  $e(f2) \leq e(f)$ , that  $e(f1) + e(f2) \leq e(f) + 4$ , and that, if  $f$  is convex, then both  $f1$  and  $f2$  are convex. However, if you used a path-splitting method with  $k$  intermediate vertices on the path, then  $e(f) + 2k + 2 \leq e(f1) + e(f2) \leq e(f) + 4$  and at least one of  $f1$  or  $f2$  must be non-convex. Therefore, one can indirectly control the number of many-sided non-convex shapes that you generate by adjusting the probabilities of the various methods. The examples shown above were generated using a probability distribution over the seven method types shown in figure 1 of  $[0.06, 0.11, 0.62, 0.06, 0.05, 0.05, 0.05]$ ; thus, there is a total probability of 0.21 of using a path-splitting method. (We keep the probability of connecting vertex-to-vertex by an edge small since these tend to oversimplify the diagram.) If you use a uniform distribution over all splitting methods, you end up with diagrams like those in figure 2; these may well be more complicated and harder to read than you would want.

(For replicability, the major functions here all include an argument to set a random seed. It should be noted that, since many calls to the random number generator are needed in the course of generating a diagram, apparently small changes to the code may result in complete loss of replicability, as, of course, would changes to the RNG in the Python numpy library.)

## 2.2 The Questions

I have implemented code that, given a diagram, can generate questions following 24 different templates. Examples of all 24 current (8/17/24) templates are given as part of Example 1, above. I give an overview below of the question template set as currently implemented, but of course all this is fairly easily modified with changes to the programming, so none of it is at all set in stone.

In the questions, regions are always identified by letter. Some questions ask, or optionally ask, about pairwise unions of regions; these are always unions of regions that meet along a consecutive sequence of edges, so that the union is topologically a disk. E.g. in Example 1, there could be a question about the union of C and G or the union of E and H, but not about the union of B and H, which are disconnected, or about the union of C and I, which meet on two disconnected edges, and whose union is therefore not simply connected.

Vertices are identified, either as the meeting place of regions, an extreme vertex of a region in some direction, or an extreme vertex of a region in terms of angle size. For instance Question 21 of Example 1 begins, “Let  $u$  be the meeting point of regions E and C with the outside of the frame.



Let  $v$  be the vertex of  $H$  with the widest angle. Let  $w$  be the top left vertex of  $E$ .” Some vertices cannot be named by any of these means; e.g. in example 7, the common vertex of  $A$  and  $H$  that is closest to  $G$  is not unique as a meeting point of  $A$  and  $H$ , nor is it extreme among the vertices of  $A$  or  $H$  in terms of angle or of any of the cardinal directions. Therefore, as the code is currently written, no question about this vertex can be formulated. (Of course it would be possible to extend the code to allow expressions like “the common vertex of  $A$  and  $H$  that is closest to  $G$ ”; the question is, what kinds of such extensions would you want to permit?) When a vertex can be named in multiple ways — e.g. in the example above, vertex  $w$  could just as well have been identified as “the meeting point of  $H$ ,  $I$ , and  $C$ ”, “the bottom vertex of  $H$ ” or “the rightmost vertex of  $C$ ” — the choice is made at random.

A number of questions refer to lines or line segments. These can be identified either as the line segment between two specified vertices; as the ray starting at a specified vertex and moving in a specified cardinal direction; as all lines between the interior of two specified regions; or as the continuation of an edge, which in turn may be specified either as the boundary between two region or as the extreme border of a region in some cardinal direction.

The form of the answer may be (a) Boolean or binary choice (b) a small integer; (c) a region; (d) a set of regions; (e) an ordered list of regions; (f) an ordered list of vertices; (g) a set of pairs of faces.

The geometric features that the questions refer to, or ask about, include topological features; number of sides; comparative distances; comparative angle; comparative position in a cardinal direction; comparative area; clockwise vs. counterclockwise triples of points; convexity; and lines crossing the interior of regions or two lines crossing each other.

Table 2 shows the features of each question, other than those involved in vertex identification.

Some questions can only be asked in one form for each diagram, e.g. question 3, “Which if any of the regions are not convex?”. Most have multiple instances per diagram; e.g. Question 1, “Which regions border region  $< identifier >$  on an edge?”

I have made a substantial effort, in writing the code, to ensure that the questions are easily answerable by humans (just using my own subjective judgment). I think I have been largely successful but certainly not entirely. An extreme example — so extreme that it really qualifies as a bug — is question 2 in example 8. It is really not obvious under visual inspection that a line from just inside the upper left corner of  $J$  to just inside the the upper left corner of  $C$  passes through the interior of  $G$ , though it is easily verified using drawing software.

## 2.3 The Code, so far

The code is in Python — currently about 4000 lines. It seems to be largely debugged, but I have no reason to suppose, and would be very surprised to learn, that it is completely debugged. It is almost completely uncommented, except for noting a few fiddly points that I myself might quickly forget. This is the first substantial program I’ve written in Python, so no doubt there are plenty of unnecessary inefficiencies and stylistic uglinesses. There is at least one deprecated and outdated method, namely the use of `numpy.random`.

## 3 Where should we take this?

Broadly speaking, I would like to use this as a benchmark for VQA AI systems. This will certainly be done much better and more usefully if I can work with collaborators who are more knowledgeable and more experienced than me (that bar is pretty much on the floor.)

Q	Answer form	Features	Unique?	Line identification
1	Set of regions	Topology	Yes	
2	Set of regions	Line crossing	No	Lines between region interior
3	Set of regions	Convexity	Yes	
4	List of regions	Topology	No	
5	Set of pairs of regions	Topology	Yes	
6	Set of pairs of regions	Topology	Yes	
7	Set of region	Topology	Yes	
8	Set of regions	Number of sides	Yes	
9	Small integer	Number of sides	Yes	
10	List of vertices	Distance	No	
11	List of vertices	Angle size	No	
12	List of vertices	Line crossing	No	Extension of edge
13	Set of regions	Topology	No	
14	Small integer	Number of sides	No	
		Union of region		
15	Set of regions	Topology Union of regions	No	
16	List of regions	Area of regions	No	
17	Set of pairs of regions	Union of regions	No	
		Convexity	No	
18	List of regions	Line crossing	No	Endpoints specified
19	List of regions	Line crossing	No	Endpoint and direction
20	List of vertices	Position in cardinal direction	No	
21	Binary choice	Clockwise	No	
22	Boolean	Line crossing	No	Endpoint specified
23	List of pairs of regions	Position in	Yes	
24	Binary choice	Distance	No	

Table 2: Features of question templates. The features used in identifying vertices are not included here. The “Unique?” column is “Yes” if there is only one instance of this question per diagram

Decisions to be made include:

- How should the answers be formulated: Free form, explicitly specified form, multiple choice, binary choice? With questions whose answer is a set or a list, is partial credit given and if so, how?
- How is the dataset to be divided into a training set and a test set? Options include:
  - (a) Use this only as a test set, not as a training set.
  - (b) Choose randomly among diagram-question pairs. Thus, the training set will contain every diagram type and every question type, but not every combination.
  - (c) Include all questions types but only some of the diagrams in the training set, thus forcing the AI to generalize to new diagrams.
  - (d) Include all of the diagrams but only some of the question types, thus forcing the AI to generalize to new question types.
  - (e) Include only some of the diagrams and some of the question types, forcing the AI to generalize in both directions.
- What should be tested? Models, training, fine-tuning, prompts, etc.
- Should we do human subject tests? If so, how? In particular, should we do tests of human subjects who can use drawing software to draw straight lines?
- Where should we set the bar in trying to make the problems easy for humans, and how?