Figure Question Answering Task

Task Description

This task is to analyze a figure and answer the questions about it. Some examples include identifying the largest slice in a pie chart, comparing two bars in a bar graph, and determining which line in a line graph appears the roughest.

Figures all include legends or labels to identify the plot elements by name.

Questions compare one element of the figure to all the others in a global fashion, or compare two elements. All questions are yes-no format, with three answer possibilities: **yes**, **no**, and **unknown**.

An answer for each question must be provided to complete the judgement. In case a question cannot be answered adequately, please select "**Unknown**".

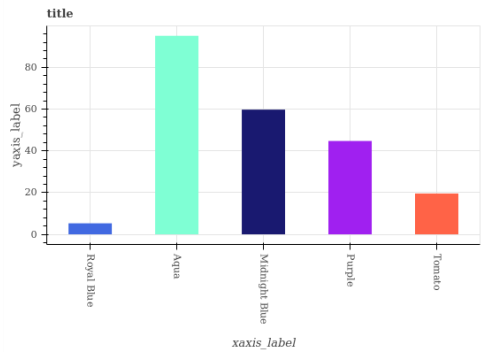
The judge should take about 1 to 2 minutes per hit based on 15-30 seconds per question, on average. This may vary if some figures have more elements, or similar plot elements. In any case, it is important that the judge not overthink their answer to each question.

Note: this task is requires that a judge is able to identify and differentiate colors. Judges with color blindness may not be suitable for this task.

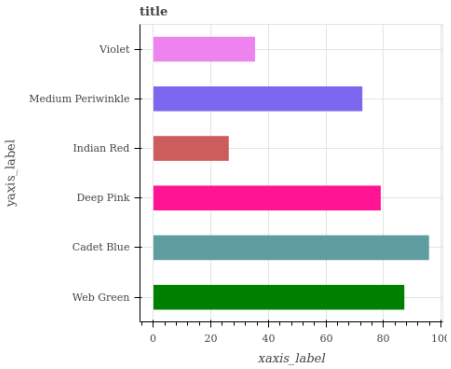
Please review the detailed disambiguation comments and guidelines at the end of this document.

Figure Overview

You will be asked to examine up to five different types of plots, see below:

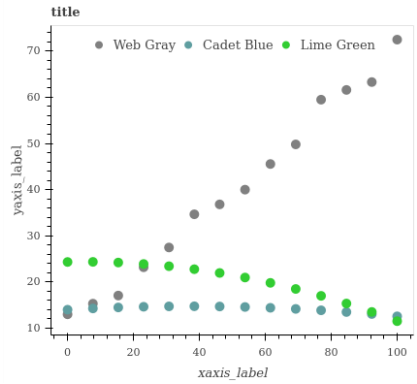


Vertical Bar Graph

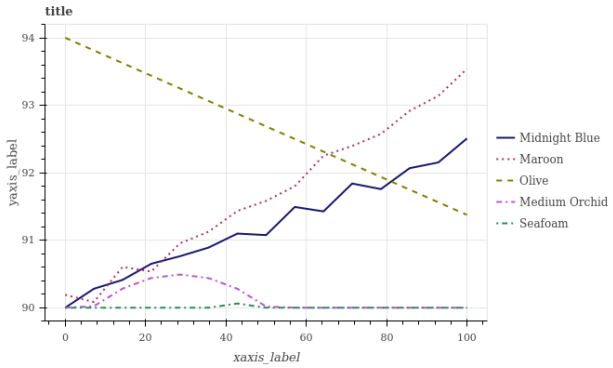


Horizontal Bar Graph

Pie Chart



Dot Line Chart



Line Chart

Questions Overview and Answering Guidelines

For global attribute questions, you may encounter instances where two or more plot elements appear to be the same. For example, two curves could each seem the roughest, some bars all have the same maximum height, or three slices of a pie chart may each be the median.

First, thoroughly re-inspect the elements to ensure that the similar elements are strictly the same.

Otherwise if all these elements are strict candidates for the criteria, then all of these plot elements satisfy the attribute, and the answer **"Yes"** would apply to all of them. Do not select **"Unknown"**.

As a concrete example, if there's two bars A and B that have strictly the maximum height, then they are both maximum bars. The question "Is A the maximum?" has answer **"Yes"** and question "Is B the maximum?" has answer **"Yes"** too. The same principle applies to other plots and question types.

If it helps, you can think of these global attribute questions as "does X have the <global attribute> value?" For example, "does Blue have the minimum value?" in a bar chart, or "does Green have the maximum area under the curve value?".

You may also encounter colors that are difficult to differentiate, for example “dark red” and “firebrick red.” These are expected and intended to be part of the task. In these cases, use your best judgement to designate colors, but be consistent.

Below are some answering guidelines to questions that may be ambiguous:

<table>
<tr><th>Figures</th><th>Question</th><th>Clarification</th></tr>
<tr><td>All figures</td><td>“Is W greater than X?”

“Is Y less than Z?”</td><td>Here ‘greater’ and ‘less’ mean <u>strictly</u> greater or less than.
If curves intersect or are equal at any point, then they are <u>not</u> strictly greater than or strictly less than. This is the same for pie slices and bars.</td></tr>
<tr><td>Vertical bar graphs,
horizontal bar graphs,
pie charts</td><td>“Is X the low median?”

“Is Y the high median?”</td><td>Low and high median are needed in cases where there are an even number of elements.

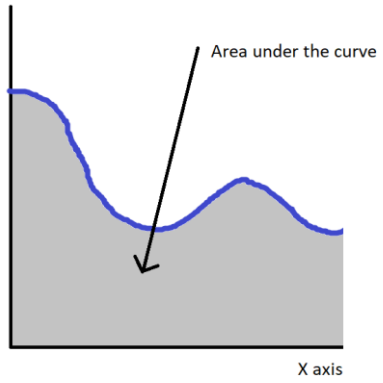
For example, in the sorted sequence [1,2,3,4,5,6], 3 is the low median and 4 is the high median.
When there is an <u>odd</u> number of elements, <u>low median and high median are the same.</u>

Additionally, if there are multiple candidates for a median, select “Yes” for all candidates that satisfy the median criteria. For example, given a bar graph

<table><tr><th>Bar Name</th><th>Height</th></tr><tr><td>Red</td><td>1</td></tr><tr><td>Green</td><td>2</td></tr><tr><td>Blue</td><td>2</td></tr></table>
both the questions “is green the median?” and
is blue the median?” have answer “Yes”.</td></tr>
<tr><td>Line charts,
dot line charts</td><td>“Is X the roughest?”

“Is Y the smoothest?”</td><td>This figure is intuitive, and the judge shouldn’t overthink this.
The ‘roughness’ of a curve refers to how noisy or jagged it is. You can also think of it as how much the direction of the curve (derivative or tangent) is changing from point-to-point.</td></tr>
<tr><td>Line charts,
dot line charts</td><td>“Does X have the maximum area under the curve?”

“Does Y have the minimum area under the curve?”</td><td>Area under the curve refers to the total area between every point on a curve and the x-axis below. For example, area under the blue curve is in grey below:

 
</td></tr>
<tr><td>Line charts,
dot line charts</td><td>“Does X have the minimum value?”

“Does Y have the maximum value?”</td><td>Here ‘minimum’ and ‘maximum’ refer to the lowest and highest points in the whole chart, among all curves/points.</td></tr>
</table>