

Chart-FR1: Visual Focus-Driven Fine-Grained Reasoning on Dense Charts

Supplementary Material

Appendix

In this appendix, we provide additional details and results to complement the main paper. Specifically, Sec. A outlines model and training details, Sec. B describes data details, Sec. C provides evaluation details, Sec. D presents cold-start data examples, Sec. E shows Chart-FR1 inference examples, and Sec. F presents HID-Chart examples.

A. Model and Training Details

We adopt Qwen2.5-VL-7B [1] as our base model. Training follows a two-stage focused reasoning paradigm, and code implementation is based on the Llama-Factory [11] and VeRL [7] frameworks. The training parameters are summarized in Table 1 and 2. The reward curves during training are shown in Fig. 1. The effect of the hyperparameter α on model performance is shown in Fig. 2. The impact of the adaptive KL penalty on reasoning length is reported in Table 3. The prompt templates are shown in Fig. 3 and 4.

Parameter	Value
Learning Rate	2×10^{-6}
Optimizer	BAdam
Global Batch Size	256
Warmup Ratio	0.1
Finetuning Type	Full
Epochs	1

Table 1. Key training parameters used for Cold-Start.

Parameter	Value
Learning Rate	1×10^{-6}
Weight Decay	1×10^{-2}
Optimizer	AdamW
Global Batch Size	512
Max Prompt Length	2048
Max Response Length	2048
Rollout	8
Temperature	1.0
Top_p	1.0
β	1×10^{-2}
α	2
τ	0.9
w_1	0.1
w_2	0.1
Epochs	3

Table 2. Key training parameters used for Focus-GRPO.



Figure 1. Reward curves during Focus-GRPO training.

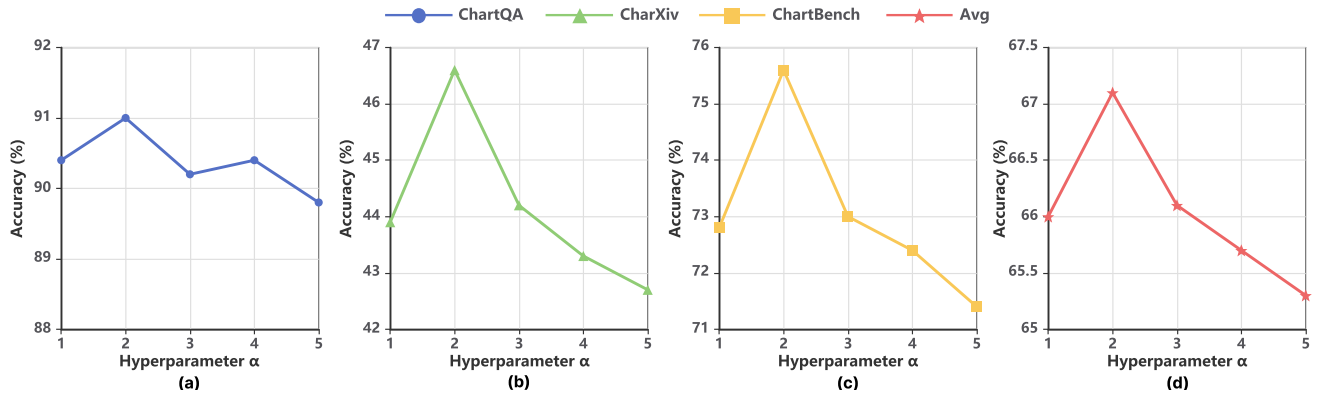


Figure 2. Effect of the hyperparameter α on model performance. Setting $\alpha = 2$ achieves the best performance in five chart benchmarks.

$N_{ocr} + N_{box}$	ChartQA	CharXiv	EvoChart	ChartBench	PlotQA	Avg
[0, 2]	218	330	228	221	247	249
[3, 4]	265	341	282	242	318	290
[5, $+\infty$]	385	459	416	314	419	399

Table 3. The effect of adaptive KL penalty on reasoning length. We incorporate an adaptive KL penalty into the GRPO algorithm, partition the number of focused visual cues into three intervals, and report the corresponding reasoning length. The results show that the adaptive KL penalty leads to longer reasoning as the number of cues increases.

```

Prompt template for Cold-Start
{
  "messages": [
    {
      "role": "user",
      "content": "<image> {question}
      You FIRST think about the reasoning process as an internal monologue and then provide the final answer.
      The reasoning process MUST BE enclosed within <think> </think> tags.
      During verification, you MUST use <focus> </focus> to simulate adaptive focus on relevant image regions and
      OCR evidence, in the following formats:
      <ocr> text evidence </ocr>, <box>{\\"bbox_2d\\": [x1,y1,x2,y2], \\"label\\": \"region description\\"} </box>.
      The output format MUST be exactly as follows:
      <think> reasoning process </think>
      <focus> Verify using <ocr> ... </ocr>, <box> {\\"bbox_2d\\": [x1,y1,x2,y2], \\"label\\": \"region description\\"} </box>
      </focus>
      <think> Continue reasoning </think>
      <answer> Final answer </answer>."
    },
    {
      "role": "assistant",
      "content": "{Focus_CoT}"
    }
  ],
  "images": [{"image_url"}]
}

```

Figure 3. The prompt template for Cold-Start.

Prompt template for Focus-GRPO

{Query} You FIRST think about the reasoning process as an internal monologue and then provide the final answer.

The reasoning process MUST BE enclosed within <think> </think> tags.

During verification, you MUST use <focus> </focus> to simulate adaptive focus on relevant image regions and OCR evidence, in the following formats:

<ocr> text evidence </ocr>, <box>{"bbox_2d": [x1,y1,x2,y2], "label": "region description"}</box>.

The output format MUST be exactly as follows:

<think> reasoning process </think>

<focus> Verify using <ocr> ... </ocr>, <box> {"bbox_2d": [x1,y1,x2,y2], "label": "region description"} </box> </focus>

<think> Continue reasoning </think>

<answer> Final answer </answer>.

Figure 4. The prompt template for Focus-GRPO.

CoT generation prompt template

{Query} First thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process and answer are enclosed within <think> </think> and <answer> </answer> tags, respectively, i.e., <think> reasoning process here </think><answer> answer here </answer>

Figure 5. The prompt template for CoT generation

LLM judgment prompt template

You are given a Question, its Ground Truth answer, and a Final Answer produced by a model.

Determine whether the Final Answer is semantically equivalent to the Ground Truth.

Matching criteria:

1. If the meaning of the Final Answer and the Ground Truth are the same, consider it a match, even if the wording differs. Minor differences such as punctuation and word capitalization are ignored.
2. For numerical answers with units:
 - If the numeric value matches exactly and the unit is correct or equivalent, consider it a match.
 - Ignore minor formatting differences.

Question: {question}

Ground Truth: {groundtruth}

Final Answer: {final_answer}

Output format:

Match: <Yes or No>

Figure 6. The prompt template for LLM judgment

B. Data Details

We design an automated Focus-CoT generation pipeline and present the prompt templates used in this section. We employ the baseline model Qwen2.5-VL-7B [1] to produce the initial CoT, with the corresponding prompt template shown in Fig. 5. The initial CoT is then conditionally reconstructed by GPT-5 [8] to obtain the Focus-CoT, and the prompt template for this step is illustrated in Fig. 7. During correctness filtering, the prompt template used for LLM-based judgment is provided in Fig. 6.

C. Evaluation Details

We conduct comprehensive evaluations on five common chart benchmarks and our HID-Chart. For ChartQA [5] and CharXiv [9], we adopt the public VLMEvalKit [2] framework. For EvoChart [3], ChartBench [10], PlotQA [6] and HID-Chart, we perform inference using vLLM [4] and evaluate the results with GPT-5 mini using a scoring prompt, shown in Fig. 8. In addition, our proposed Chart-ID information density metric is scored by GPT-5 along four dimensions, and the corresponding prompt is provided in Fig. 9.

Focus-CoT generation prompt template

User:

You are an expert in Chain-of-Thought (CoT) reasoning. You will receive the following information:

- An image.
- A question.
- The CoT reasoning process of another model for this question (without any <focus> tags).
- The correct answer to the question.

Your task is to insert <focus> tags at appropriate positions in the reasoning process based on the correctness of the model's answer, simulating adaptive "focus" analysis of the image, and modify the reasoning process accordingly. Finally, return the CoT reasoning process with the inserted <focus> tags.

Rules for inserting <focus> tags:

1. If the model's answer is incorrect, identify the error in the reasoning and focus on the relevant information, using <ocr> and <box> tags to mark information in the image within the <focus> tag. After focusing on the information, point out logical errors, missing image information, or information inconsistent with the image in the second <think> tag, and then modify the subsequent reasoning and final answer accordingly. Do not modify the reasoning before the error.
2. If the model's answer is correct, insert <focus> tags at appropriate points in the reasoning process, using <ocr> and <box> tags to mark information in the image within the <focus> tag. After focusing on the information, use the second <think> tag to verify the reasoning steps related to the information. Do not modify the reasoning content.
3. The content within the <focus> tag must include necessary <ocr> and <box> tags.
4. There must be one <focus> tag between two <think> sections.
5. During the process of reasoning and focusing on the information, avoid making statements that imply you already know the correct answer.

Format requirements (simulating adaptive focus):

- The reasoning process must be enclosed in <think> tags.
- During verification, you must use <focus> tags to simulate adaptive focus analysis of image-related areas and OCR information, in the following format:
<ocr>textual evidence</ocr>, <box>{"bbox_2d":[x1,y1,x2,y2], "label":"area description"}</box>.
- The final answer must be enclosed in <answer> tags.

--- Input ---

Image:

Question: {question}

CoT reasoning process: {response}

Correct answer: {answer}

--- Output Format ---

<think> reasoning process </think>

<focus> Verify using <ocr> ... </ocr>, <box> {"bbox_2d":[x1,y1,x2,y2], "label":"region description"} </box> </focus>

<think> Continue reasoning </think>

<answer> Final answer </answer>

--- Output Example ---

<think> The rosybrown solid line represents Romanticism. The second data point from the left on this line is at the 1880s mark. The value at this point is 277.</think>

<focus> Verify using <ocr>106</ocr>, <box>{"bbox_2d":[330,610,385,650], "label":"value label near the second rosybrown point"}</box> </focus>

<think> The focus shows the label "106" at the second rosybrown data point. The earlier step incorrectly identified the 1880s peak (labeled 277) as the second point from the left. That peak occurs later in the series, not at the second position. Correcting this, the 2nd data point from the left on the rosybrown solid line is 106.</think>

<answer>106</answer>

Figure 7. The prompt template for Focus-CoT generation

Inference result evaluation prompt

You will be given a question, an ground truth answer and a model response. You need to extract the final answer from the model response, compare it with the ground truth answer, and then assign a binary score. Avoid providing explanations in your response. If there is no provided model response, please leave the extracted answer empty and give a score of 0.

Your response must follow json formats with keys [extract_answer, score] where the value of the score is an interger in [0, 1]. You must follow the scoring rules:

1. Give a score of 1 if and only if the final answer and the ground truth answer are referring to the same term. It's acceptable to have different grammar or form (e.g., α and alpha; $R^2_{\{t,h,v,m\}}$ and $R^2_{t,h,v,m}$). It's also acceptable to have different orders of the terms when question asks for multiple terms.
2. Give a score of 0 if any term (e.g., ACC+ and ACC; P-101 and P=101) is different between the final answer and the ground truth.

Example 1 Starts

* Question: What is the name of the curve that intersects $y=\lambda$ exactly three times?

* Ground Truth: P56962

* Response: There is only one curve that intersects $y=\lambda$ exactly three times. The name of the curve is written as P55762.

{"extract_answer": "P55762", "score": 0}

Example 1 Ends

Example 2 Starts

* Question: What is the letter of the subplot where all bars are above 35?

* Ground Truth: (b)

* Response: The letter of the subplot where all bars are above 35 is b.

{"extract_answer": "b", "score": 1}

Example 2 Ends

Your Turn

* Question: {question}

* Ground Truth: {groundtruth}

* Response: {prediction}

Figure 8. The prompt template for evaluation.

Information density scoring prompt template

You are asked to evaluate the information density of a given chart image.

Information density is defined as the degree to which the chart conveys a large amount of meaningful information effectively, efficiently, and clearly within limited visual space.

Evaluate the chart across four dimensions. For each dimension, provide both a score (1–5) and a short explanation:

1. Information Richness (weight: 0.5): Amount, diversity, and complexity of information (e.g., number of data series, dimensions, annotations, subplots).
2. Information Efficiency (weight: 0.2): How effectively the information is delivered without unnecessary redundancy.
3. Information Clarity (weight: 0.2): Visual readability and ease of understanding (e.g., labeling, layout, absence of clutter).
4. Information Interactivity (weight: 0.1): How easily the user can capture and interpret the key insights from the chart (e.g., whether important trends, comparisons, or anomalies stand out quickly).

Finally, compute the overall_score as a weighted sum:

$overall_score = 0.5 * richness + 0.2 * efficiency + 0.2 * clarity + 0.1 * interactivity$

Round the result to 1 decimal place.

Your output must be in JSON format with the following fields:

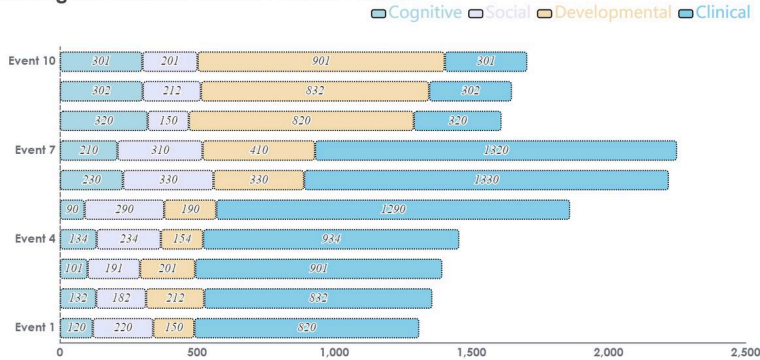
- description: one-sentence description of the chart.
- scores: an object containing all four dimensions, where each dimension includes score and reason.
- overall_score: a floating-point number with one decimal place.

Figure 9. The prompt template for calculating the Chart-ID.

D. Cold-Start Data Examples

Cold-Start Data Example 1

Psychological Studies Trends Across Events



Question: What is the value of Clinical in Event 5 ?

Answer: 1290

CoT

<think>

The chart shows the values for each category (Cognitive, Social, Developmental, Clinical) across different events. For Event 5, the Clinical value is represented by the blue bar. By examining the chart, the Clinical value for Event 5 is 1320.

</think>

<answer>1320</answer>

Focus-CoT

<think>

The chart shows the values for each category (Cognitive, Social, Developmental, Clinical) across different events. For Event 5, the Clinical value is represented by the blue bar. By examining the chart, the Clinical value for Event 5 is 1320.

</think>

<focus> Verify using <ocr>Event 5</ocr>, <box>{\bbox_2d\":[85,420,185,460], \label\":"Event 5 label on y-axis"}</box>; <ocr>1290</ocr>, <box>{\bbox_2d\":[720,420,1140,470], \label\":"Clinical (blue) bar value for Event 5"}</box> </focus>

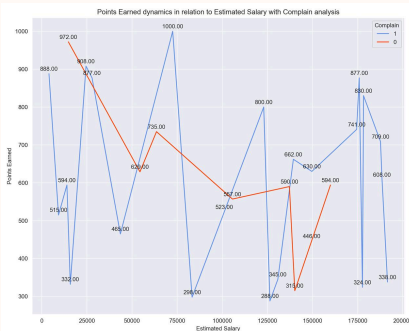
<think>

The earlier step misread the Clinical value for Event 5 as 1320. The focused region shows the Clinical bar for Event 5 labeled 1290; 1320 corresponds to a different event. Correcting this, the Clinical value for Event 5 should be 1290.

</think>

<answer>1290</answer>

Cold-Start Data Example 2



Question: What is the y-value for $x=14721.8$ and $class=0$ based on the chart data?

Answer: 972

CoT

<think>

To determine the y-value for $x=14721.8$ and $class=0$, we need to follow these steps:

1. Identify the x-axis value of 14721.8 on the graph.
2. Look at the corresponding point on the orange line ($class=0$) since it represents the points earned for non-complain cases.
3. Read the y-value (points earned) from the y-axis for that point.

From the graph:

- The x-axis value of 14721.8 is approximately aligned with the point where the orange line ($class=0$) intersects the y-axis.
- The y-value for this point is around 590.00.

Therefore, the y-value for $x=14721.8$ and $class=0$ is approximately 590.00.

</think>

<answer>590.00</answer>

Focus-CoT

<think>

To determine the y-value for $x=14721.8$ and $class=0$, we need to follow these steps:

1. Identify the x-axis value of 14721.8 on the graph.
2. Look at the corresponding point on the orange line ($class=0$) since it represents the points earned for non-complain cases.
3. Read the y-value (points earned) from the y-axis for that point.

From the graph:

- The x-axis value of 14721.8 is approximately aligned with the point where the orange line ($class=0$) intersects the y-axis.

</think>

<focus> Verify using <ocr>972.00</ocr>, <box>{"bbox_2d": [150, 95, 235, 140], "label": "annotated y-value near the left orange point ($x \approx 14721.8$)"}</box>, <ocr>0</ocr>, <box>{"bbox_2d": [870, 80, 905, 115], "label": "legend entry for class 0 (orange)"}</box> </focus>

<focus>

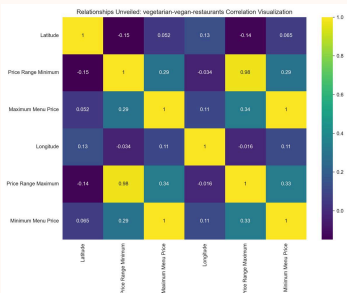
<think>

The focused annotation shows the orange point near $x \approx 14721.8$ is labeled 972.00, and the legend confirms orange corresponds to class 0. The earlier inference of 590.00 was a misreading—590 appears on a different point, not at $x=14721.8$ for class 0. Correcting this, the y-value corresponding to $x=14721.8$ and $class=0$ is 972.

</think>

<answer>972</answer>

Cold-Start Data Example 3



Question: What numerical data do `menus.amountMax` and `menus.amountMax` hold in the heatmap?

Answer: 1.0

CoT

<think>

The image provided is a heatmap showing correlation coefficients between various attributes of vegetarian-vegan restaurants. The heatmap does not contain any numerical data such as `"menus.amountMax"` or `"menus.amountMax"`. Instead, it displays correlation values ranging from -1 to 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

Since the question asks for the numerical data held by `"menus.amountMax"` and `"menus.amountMax"`, which are not present in the heatmap, we cannot provide a numerical value based on the information given in the image. The heatmap only shows correlation coefficients between different attributes like latitude, price range minimum, maximum menu price, longitude, price range maximum, and minimum menu price. Therefore, the answer cannot be derived from the image alone.

</think>

<answer>The heatmap does not contain the numerical data `"menus.amountMax"` or `"menus.amountMax"`. These are not present in the image.</answer>

Focus-CoT

<think>

The image provided is a heatmap showing correlation coefficients between various attributes of vegetarian-vegan restaurants. The heatmap does not contain any numerical data such as `"menus.amountMax"` or `"menus.amountMax"`. Instead, it displays correlation values ranging from -1 to 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

Since the question asks for the numerical data held by `"menus.amountMax"` and `"menus.amountMax"`, which are not present in the heatmap, we cannot provide a numerical value based on the information given in the image. The heatmap only shows correlation coefficients between different attributes like latitude, price range minimum, maximum menu price, longitude, price range maximum, and minimum menu price. Therefore, the answer cannot be derived from the image alone.

</think>

<focus> <ocr>Maximum Menu Price</ocr>, <box>{"bbox_2d": [60, 350, 220, 430], "label": "y-axis label 'Maximum Menu Price'"}</box>, <ocr>1</ocr>, <box>{"bbox_2d": [430, 420, 520, 510], "label": "diagonal cell value at Maximum Menu Price × Maximum Menu Price"}</box> </focus>

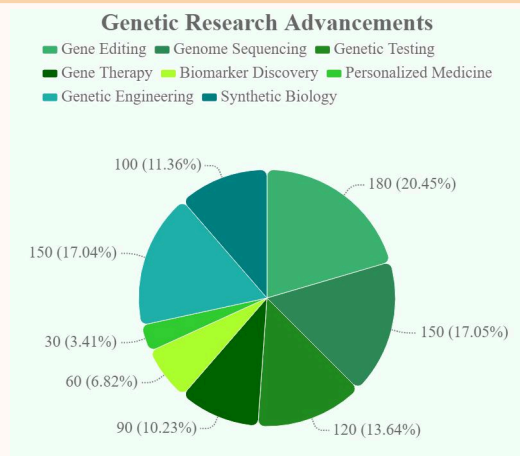
<think>

The earlier reasoning incorrectly states that the heatmap does not contain numerical data for `menus.amountMax`. In the heatmap, `"Maximum Menu Price"` corresponds to `menus.amountMax`. On the diagonal (a variable correlated with itself), the value displayed is 1, which is standard for self-correlation and is explicitly shown in the cell at the intersection of `"Maximum Menu Price"` with itself. Therefore, the numerical data shown for `menus.amountMax` vs `menus.amountMax` in the heatmap is 1.0.

</think>

<answer>1.0</answer>

Cold-Start Data Example 4



Question: What count of 'Personalized Medicine' does 'Research Areas' have in this chart?

Answer: 30

CoT

<think>

The chart is a pie chart representing different research areas in genetic research advancements. Each slice of the pie chart corresponds to a specific area, and the percentage next to each slice indicates the proportion of that area relative to the total. The area labeled "Personalized Medicine" has a value of 180 and a percentage of 20.45%. This means that "Personalized Medicine" accounts for 20.45% of the total research areas shown in the chart.

</think>

<answer>180</answer>

Focus-CoT

<think>

The chart is a pie chart representing different research areas in genetic research advancements. Each slice of the pie chart corresponds to a specific area, and the percentage next to each slice indicates the proportion of that area relative to the total. The area labeled "Personalized Medicine" has a value of 180 and a percentage of 20.45%. This means that "Personalized Medicine" accounts for 20.45% of the total research areas shown in the chart.

</think>

<focus> Verify using <ocr>Personalized Medicine</ocr>, <box>{"bbox_2d": [220, 105, 670, 160], "label": "legend entry 'Personalized Medicine' near top"}</box>, <ocr>30 (3.41%)</ocr>, <box>{"bbox_2d": [135, 520, 250, 560], "label": "small pie-slice label near lower-left of chart"}</box> </focus>

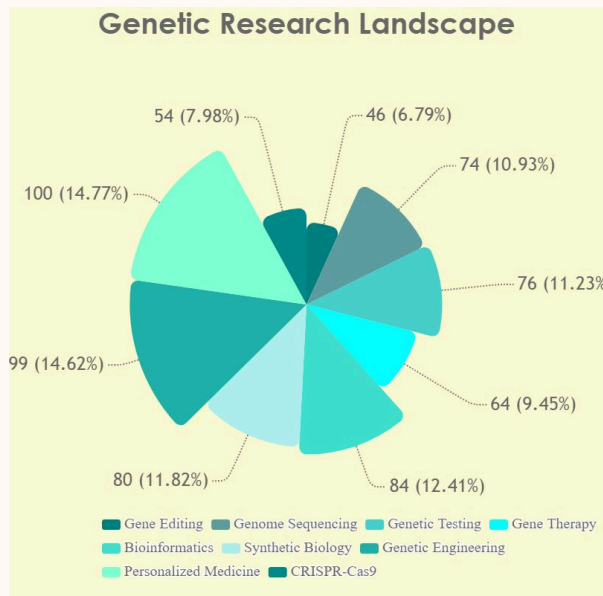
<think>

The earlier step incorrectly associated the largest slice value (180, 20.45%) with "Personalized Medicine." Focusing on the legend shows the "Personalized Medicine" category, and the nearby small slice label reads "30 (3.41%)," which indicates its count. Therefore, the previous reasoning is inconsistent with the image; the correct count for "Personalized Medicine" is 30. Updating the final answer accordingly.

</think>

<answer>30</answer>

Cold-Start Data Example 5



Question: What is the percentage of the largest sector in 'Research Areas' in the pie chart?

Answer: 14.77

CoT

<think>

The largest sector in the pie chart represents "Personalized Medicine" with 14.77%.

</think>

<answer>14.77</answer>

Focus-CoT

<think>

The largest sector in the pie chart represents "Personalized Medicine" with 14.77%.

</think>

<focus> <ocr>100 (14.77%)</ocr>, <box>{\bbox_2d\":[120,280,240,340], \label\":"largest percentage label near top-left of pie chart\"}</box>, <ocr>Personalized Medicine</ocr>, <box>{\bbox_2d\":[300,650,540,700], \label\":"legend item near bottom center\"}</box> </focus>

<focus>

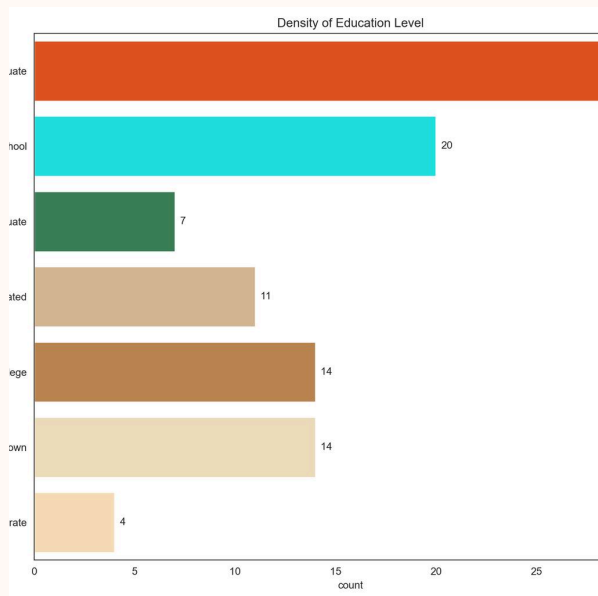
<think>

The highlighted label shows 100 (14.77%), and the legend confirms the sector corresponds to Personalized Medicine, verifying that this is the largest percentage in the chart.

</think>

<answer>14.77</answer>

Cold-Start Data Example 6



Question: Inform me of the chart value for class Unknown.

Answer: 14

<think>

CoT

The chart is a horizontal bar chart showing the density of education levels. Each bar represents a different education level, and the length of the bar corresponds to the count of individuals at that level. The bar labeled "Unknown" extends to a count of 14.

</think>

<answer>14</answer>

Focus-CoT

<think>

The chart is a horizontal bar chart showing the density of education levels. Each bar represents a different education level, and the length of the bar corresponds to the count of individuals at that level. The bar labeled "Unknown" extends to a count of 14.

</think>

<focus> Verify using <ocr>Unknown</ocr>, <box>{"bbox_2d": [65, 465, 190, 505], "label": "y-axis category label 'Unknown'"}</box>, <ocr>14</ocr>, <box>{"bbox_2d": [530, 470, 565, 505], "label": "value label at end of 'Unknown' bar"}</box> </focus>

<think>

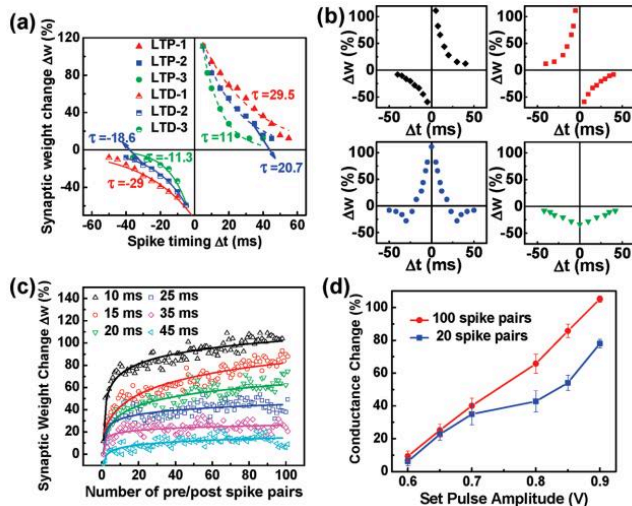
The focused regions show the category label "Unknown" and the numeric label "14" at the end of that bar, confirming that the count for "Unknown" is 14.

</think>

<answer>14</answer>

E. Chart-FR1 Inference Examples

Chart-FR1 Inference Example 1



Question: In panel (a), which LTD trace demonstrates the slowest decay?

Answer: LTD-1

<think>

To determine which LTD trace demonstrates the slowest decay in panel (a), we need to analyze the slopes of the LTD traces. The LTD traces are represented by the blue triangles and are labeled as LTD-1, LTD-2, and LTD-3. The decay rate is indicated by the value of τ (tau), which represents the time constant of the exponential decay.

- LTD-1 has a τ value of 29.5 ms.
- LTD-2 has a τ value of 20.7 ms.
- LTD-3 has a τ value of 11.3 ms.

The smaller the τ value, the faster the decay. Therefore, the LTD trace with the largest τ value will demonstrate the slowest decay.

Comparing the τ values:

- 29.5 ms (LTD-1) is the largest.
- 20.7 ms (LTD-2) is the second largest.
- 11.3 ms (LTD-3) is the smallest.

Thus, LTD-1 has the slowest decay.

</think>

<focus> $\tau = 29.5$ </focus>, "bbox_2d": [100, 100, 150, 150], "label": "legend label for LTD-1 trace" </focus>

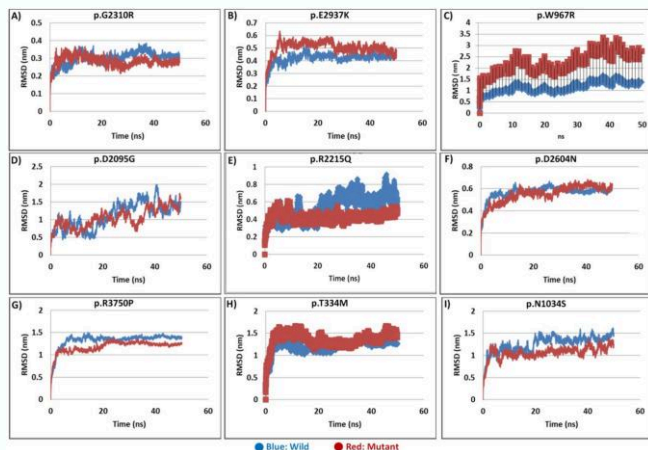
<think>

The focus on the legend label for LTD-1 trace confirms the τ value of 29.5 ms, which is the largest among the LTD traces. This supports the reasoning that LTD-1 demonstrates the slowest decay.

</think>

<answer>LTD-1</answer>

Chart-FR1 Inference Example 2



Question: Which mutation attains the global maximum RMSD, indicated by red bars exceeding 3.0 nm in its panel?

Answer: p.W967R

<think>

To determine which mutation attains the global maximum RMSD, we need to identify the red bar (representing the mutant) that exceeds approximately 3.0 nm in its respective panel.

Let's analyze each panel:

- A) p.G2310R: The red bar does not exceed 3.0 nm.
- B) p.E2937K: The red bar does not exceed 3.0 nm.
- C) p.W967R: The red bar does not exceed 3.0 nm.
- D) p.D2095G: The red bar does not exceed 3.0 nm.
- E) p.R2215Q: The red bar does not exceed 3.0 nm.
- F) p.D2604N: The red bar does not exceed 3.0 nm.
- G) p.R3750P: The red bar does not exceed 3.0 nm.
- H) p.T334M: The red bar does not exceed 3.0 nm.
- I) p.N1034S: The red bar does not exceed 3.0 nm.

None of the red bars in the panels exceed 3.0 nm. Therefore, there is no mutation that attains the global maximum RMSD indicated by red bars exceeding 3.0 nm in its panel.

</think>

<focus> <ocr> p.W967R, p.R2215Q, p.D2604N, p.R3750P, p.T334M, p.N1034S </ocr>, <box>{\\"bbox_2d\\": [500,100,600,200], \\"label\\": \"Panel C, E, F, G, H, I showing RMSD values\\\"}</box> </focus>

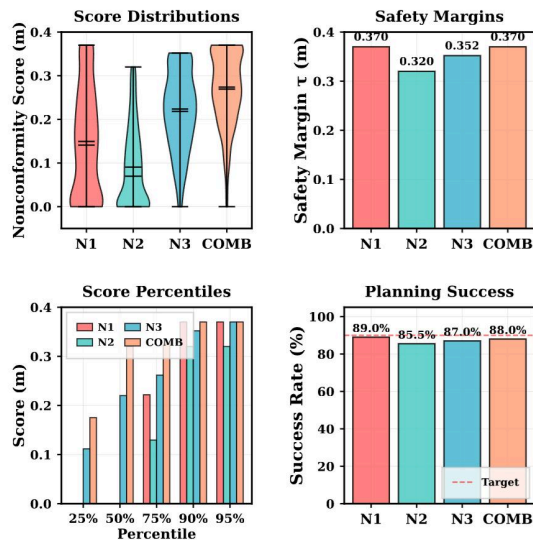
<think>

The focus is on panels C, E, F, G, H, and I, which show the RMSD values for the mutations p.W967R, p.R2215Q, p.D2604N, p.R3750P, p.T334M, and p.N1034S respectively. The previous reasoning was incorrect as it did not consider the full range of RMSD values. Upon closer inspection, the red bar in panel C (p.W967R) reaches the highest value among all the panels, exceeding 3.0 nm.

</think>

<answer>p.W967R</answer>

Chart-FR1 Inference Example 3



Question: What is the difference (in meters) between the maximum and minimum Safety Margin τ values across N1–COMB?

Answer: 0.050 meters

<think>

To find the difference between the maximum and minimum Safety Margin τ values across N1–COMB, we need to identify the highest and lowest values from the bar chart labeled "Safety Margins."

From the bar chart:

- N1: 0.370 m
- N2: 0.320 m
- N3: 0.352 m
- COMB: 0.370 m

The maximum value is 0.370 m (N1 and COMB).

The minimum value is 0.320 m (N2).

The difference between the maximum and minimum values is:

$$0.370 \text{ m} - 0.320 \text{ m} = 0.050 \text{ m}$$

</think>

<focus> <ocr>0.370, 0.320, 0.352, 0.370</ocr>, <box>{"bbox_2d": [400, 100, 600, 200], "label": "the top bar chart labeled 'Safety Margins' showing the values for N1, N2, N3, and COMB"}</box> </focus>

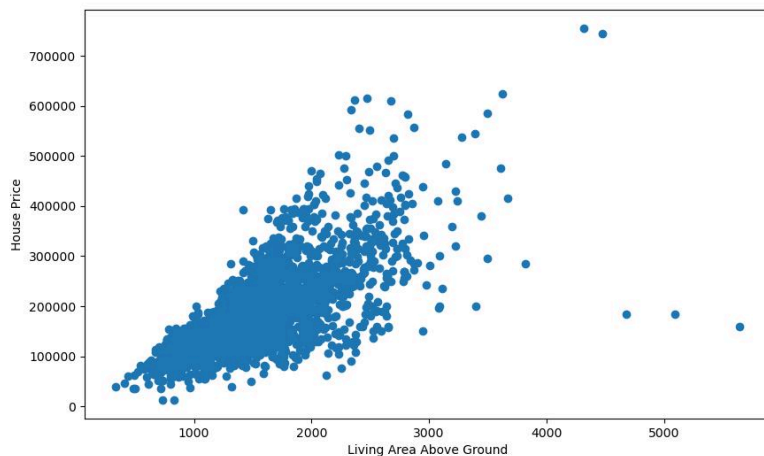
<think>

The focus is on the top bar chart labeled "Safety Margins." The values for N1, N2, N3, and COMB are clearly labeled as 0.370, 0.320, 0.352, and 0.370 respectively. The previous reasoning correctly identified these values and calculated the difference between the maximum and minimum values as 0.050 m. The focus on the chart confirms the values and the reasoning is consistent with the chart data.

</think>

<answer>0.050</answer>

Chart-FR1 Inference Example 4



Question: How many dots are there on and after the line: "Living area above ground = 4000"?

Answer: 5

<think>

To determine the number of dots on and after the line "Living area above ground = 4000," we need to:

1. Identify the x-axis label: "Living Area Above Ground."
2. Locate the horizontal line at the value 4000 on the x-axis.
3. Count all the data points (dots) that fall on or to the right of this horizontal line.

From the graph, the horizontal line at 4000 on the x-axis separates the data points into two regions:

- To the left of the line ($x < 4000$), there are many data points.
- To the right of the line ($x \geq 4000$), there are fewer data points.

By visually inspecting the graph, it is clear that there are no dots at exactly 4000, but we count all the dots to the right of this point.

</think>

<focus> Verify using `Living area above ground = 4000`, `<box>{"bbox_2d": [1000,500,1100,550], "label": "horizontal line near the 4000 label on the x-axis"}`</focus>

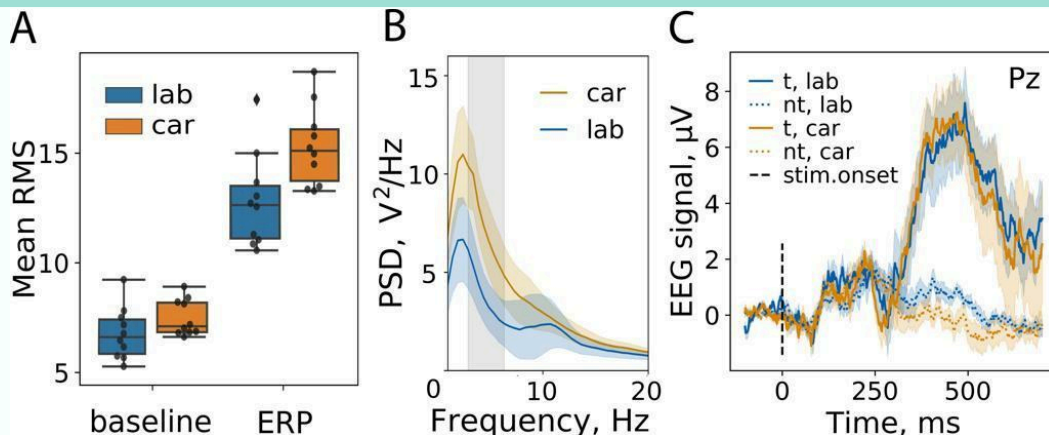
<think>

The reasoning focuses on the horizontal line at the value 4000 on the x-axis. The task is to count all data points (dots) on or to the right of this line. The labeled focus box points to the horizontal line at 4000. By visual inspection, the number of dots beyond this line is 5.

</think>

<answer>5</answer>

Chart-FR1 Inference Example 5



Question: In plot A, which condition shows a greater dispersion of Mean RMS values for 'car'?

Answer: ERP

<think>

To determine which condition shows a greater dispersion of Mean RMS values for 'car' in plot A, let's analyze the box plots:

1. **Identify the box plots for 'car':** The orange box represents the 'car' condition in both the 'baseline' and 'ERP' categories.
2. **Compare the dispersion in the 'car' box plots:**
 - In the 'baseline' category, the 'car' box plot shows a relatively narrow box and a small whisker range, indicating low dispersion.
 - In the 'ERP' category, the 'car' box plot shows a wider box and larger whisker range, indicating higher dispersion.
3. **Conclusion:** The 'ERP' category shows a greater dispersion of Mean RMS values for the 'car' condition compared to the 'baseline' category.

</think>

<focus>Verify using <ocr>Mean RMS, car, baseline, ERP</ocr>, <box>{"bbox_2d": [100,100,500,500], "label": "the legend and box plots in plot A"}</box> </focus>

<think>

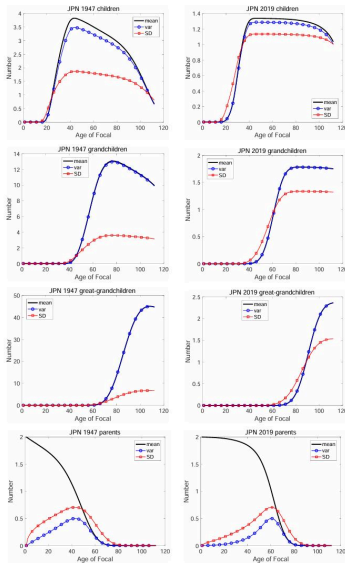
The focus on the legend and box plots in plot A shows the 'car' box plots in both 'baseline' and 'ERP' categories. The 'ERP' box plot indeed has a wider box and larger whiskers compared to the 'baseline' box plot, confirming the reasoning that the 'ERP' category shows greater dispersion for the 'car' condition.

</think>

<answer>ERP</answer>

Chart-FR1 Inference Example 6

Figure 3: (Part 1) The mean, variance, and standard deviation (SD) of the number of kin as a function of the age of Focal, under the rates for Japan in 1947 and 2019. Note different scales on the y-axes.



Question: At Age of Focal 60, which year has the higher blue “var” for grandchildren, JPN 1947 or JPN 2019?

Answer: JPN 1947

<think>

To determine which year has the higher blue “var” for grandchildren at the age of Focal 60, we need to compare the values of the blue “var” line at the 60th age mark for both the JPN 1947 and JPN 2019 graphs.

1. Locate the graph for JPN 1947 grandchildren (the second row, first column).
2. Identify the blue “var” line in this graph.
3. Find the value of the blue “var” line at the 60th age mark.
1. Locate the graph for JPN 2019 grandchildren (the second row, second column).
2. Identify the blue “var” line in this graph.
3. Find the value of the blue “var” line at the 60th age mark.

By visually comparing the two values, we can determine which one is higher.

</think>

<focus> JPN 1947 grandchildren, <box>{“bbox_2d”:[100,300,300,400], “label”:\’the second row, first column graph label\’}</box> </focus>

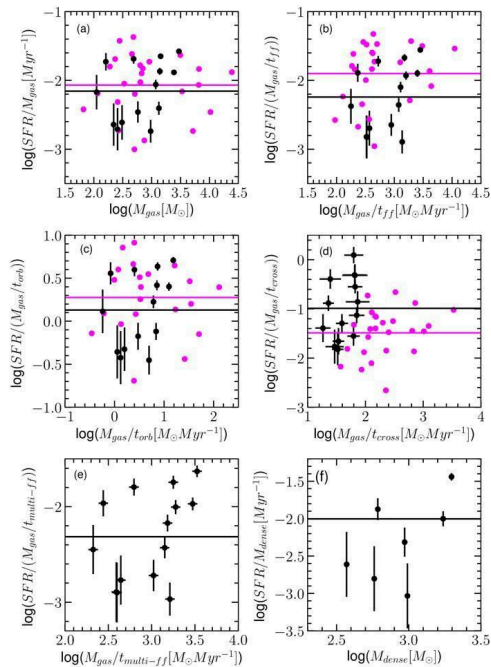
<think>

The focus is on the label of the graph for JPN 1947 grandchildren, which indicates the graph we need to examine for the blue “var” line at the 60th age mark. The reasoning process involves comparing the blue “var” values at the 60th age mark between the JPN 1947 and JPN 2019 graphs.

</think>

<answer>JPN 1947</answer>

Chart-FR1 Inference Example 7



Question: How many subplots do not contain the purple data points?

Answer: 2

<think>

To determine how many subplots do not contain the purple data points, let's analyze each subplot:

- (a) contains both black and purple data points.
- (b) contains both black and purple data points.
- (c) contains both black and purple data points.
- (d) contains both black and purple data points.
- (e) contains only black data points.
- (f) contains only black data points.

Subplots (e) and (f) do not contain the purple data points.

</think>

<focus> Verify using (e) $\log(M_{gas}/t_{multi-ff}[M_{\odot}M_{yr^{-1}}])$, $\text{\texttt{"bbox_2d": [100,800,400,1000]}}$, $\text{\texttt{"label": "the bottom-left subplot label and axis range"}}$ </focus>

<think>

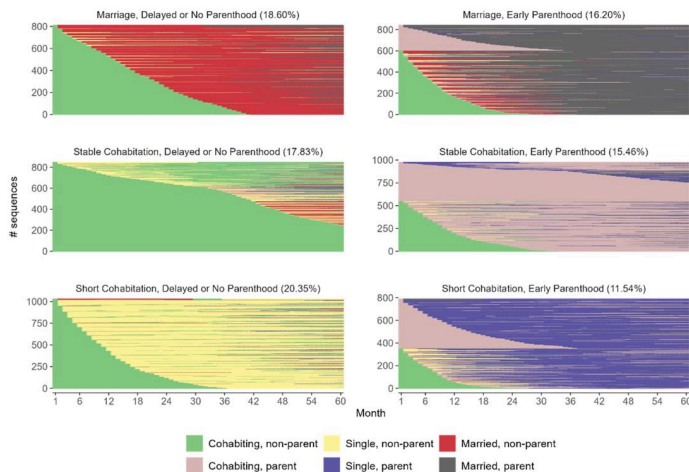
The focus is on the bottom-left subplot (e). The label indicates it is about gas mass and time, and the plot shows only black data points, confirming the reasoning that this subplot does not contain the purple data points. The reasoning is consistent with the earlier analysis.

</think>

<answer>2</answer>

Chart-FR1 Inference Example 8

Figure 1: Sequence index plot of partnership-parenthood trajectories by cluster membership



Question: Of the clusters with a 1000-sequence y-axis, which is dominated by yellow (“Single, non-parent”) between months 18 and 36?

Answer: Short Cohabitation, Delayed or No Parenthood

<think>

To determine which cluster is dominated by yellow (“Single, non-parent”) between months 18 and 36, let’s follow these steps:

1. Identify the cluster with a 1000-sequence y-axis.
2. Check the color distribution for the “Single, non-parent” category (yellow) in the specified time period (months 18 to 36).

The cluster with a 1000-sequence y-axis is the bottom-left plot labeled “Short Cohabitation, Delayed or No Parenthood (20.35%)”.

Now, let’s analyze the color distribution in this plot between months 18 and 36:

- The yellow color (Single, non-parent) is present but not dominant. The majority of the plot shows a mix of colors, with no single color dominating the area.

Therefore, the cluster with a 1000-sequence y-axis that is dominated by yellow (“Single, non-parent”) between months 18 and 36 is not present in this plot.

</think>

<focus> <ocr>Short Cohabitation, Delayed or No Parenthood (20.35%)</ocr>

<box>{\“bbox_2d\“:[100,300,500,400], \“label\“:\“the bottom-left plot label\“}</box> </focus>

<think>

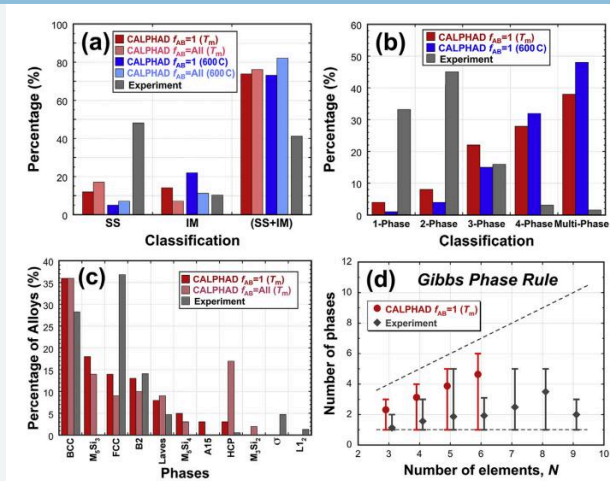
The focus is on the bottom-left plot label, which indicates the cluster with 1000 sequences. The reasoning should be rechecked for the plot itself, not just the label. The plot shows a mix of colors, and the yellow (Single, non-parent) is dominant.

</think>

<answer>Short Cohabitation, Delayed or No Parenthood</answer>

F. HID-Chart Examples

HID-Chart Example 1

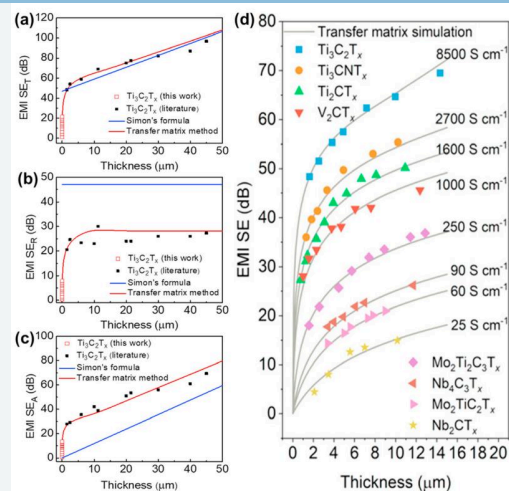


Question: According to subplot (d), for which value of N (number of elements) does the experimental number of phases most closely match the prediction made by the Gibbs Phase Rule?

Answer: $N = 5$

Information Density: 4.4

HID-Chart Example 2

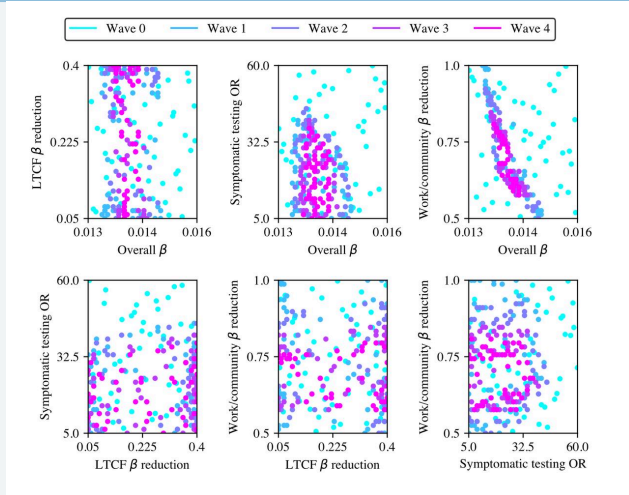


Question: Referring to (d), which conductivity curve is almost overlapped with the data points for Ti_3CNT_x for thickness up to 10 μm ?

Answer: 2700 $S\ cm^{-1}$

Information Density: 4.5

HID-Chart Example 3

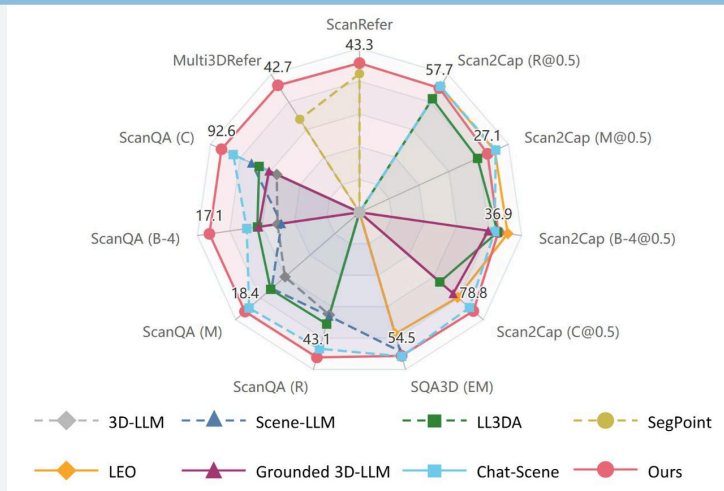


Question: Across all six panels, which pair of variables shows the strongest monotonic decrease (y falls as x rises) compared to the more scattered relationships elsewhere?

Answer: Overall β and Work/community β reduction

Information Density: 4.2

HID-Chart Example 4

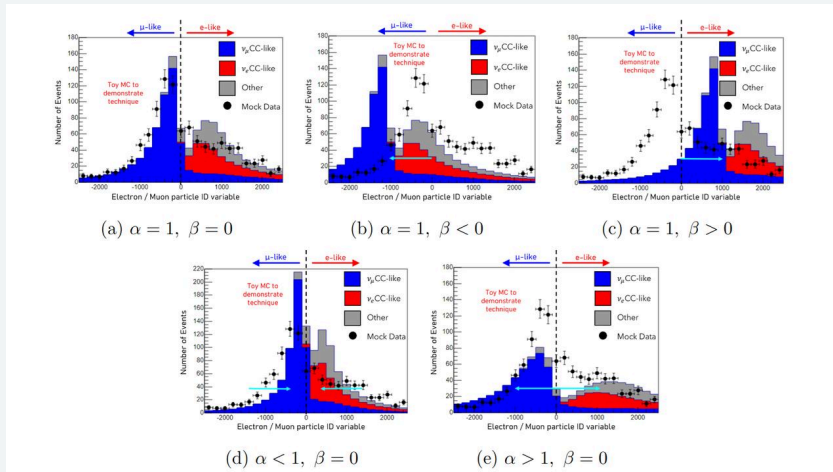


Question: Among Ours' four ScanQA metrics (C, R, M, B-4), which one corresponds to its lowest value?

Answer: ScanQA (B-4)

Information Density: 4.0

HID-Chart Example 5

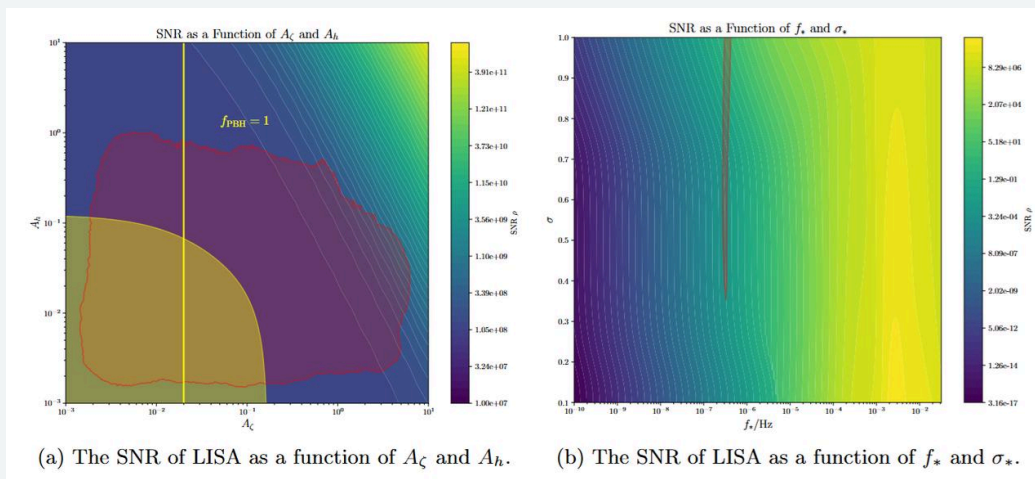


Question: Comparing panels (b) and (c), in which panel do the mock data points (black) peak closest to $x \approx +1000$ within the e-like region?

Answer: Panel (c)

Information Density: 4.2

HID-Chart Example 6

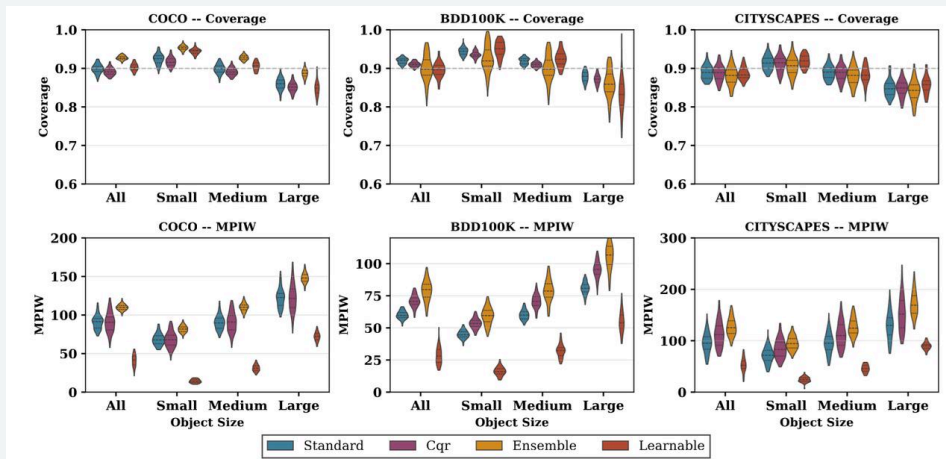


Question: In panel (a), does the vertical line labeled $f_{\text{PBH}} = 1$ intersect the red-outlined region?

Answer: Yes

Information Density: 4.2

HID-Chart Example 7



Question: In BDD100K — Coverage, for which object size does the Learnable (brown) violin fall furthest below the 0.9 dashed line?

Answer: Large

Information Density: 4.4

HID-Chart Example 8

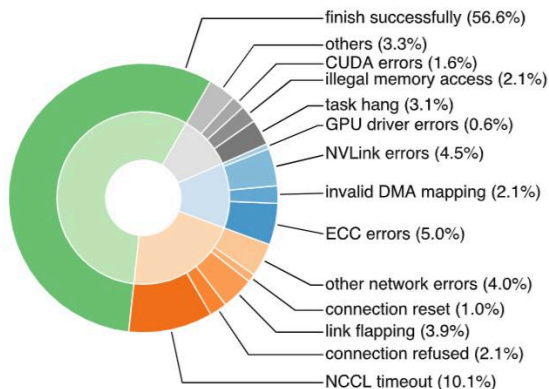


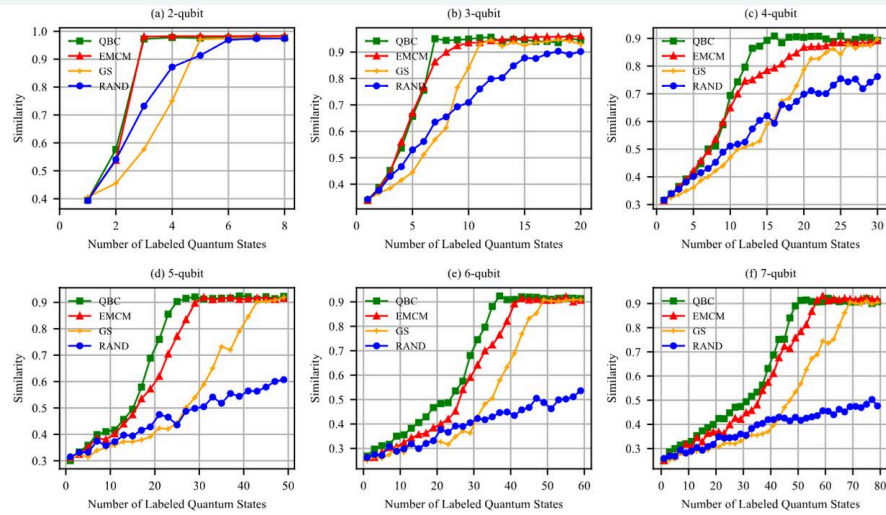
Figure 1: Distribution of task termination statistics.

Question: Within the orange network cluster, which labeled issue has the smallest percentage?

Answer: connection reset (1.0%)

Information Density: 3.7

HID-Chart Example 9

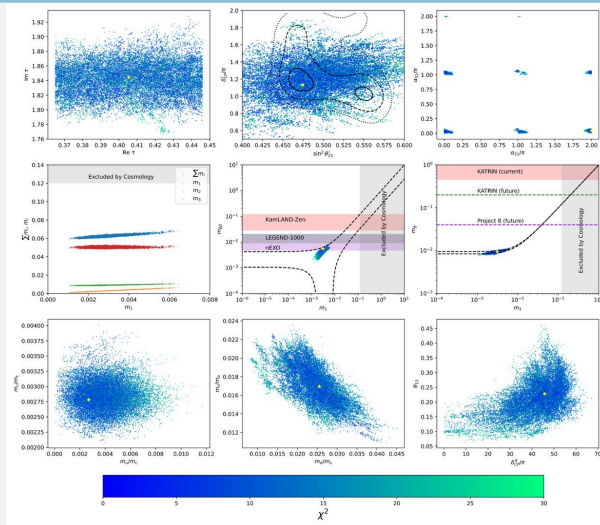


Question: Across all six subplots, which qubit setting requires the most labeled states before EMCM first reaches 0.9 Similarity?

Answer: (f) 7-qubit

Information Density: 4.5

HID-Chart Example 10

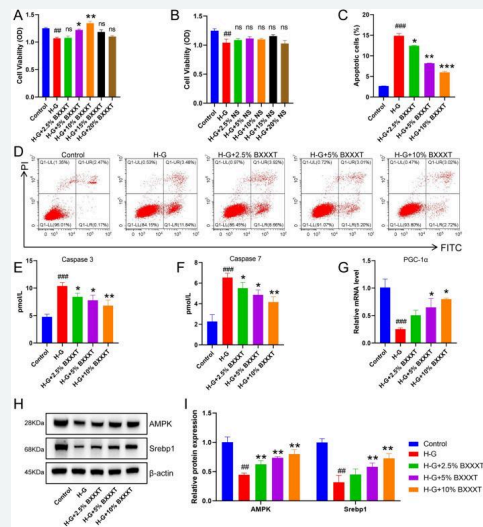


Question: In the Σm_i , m_i vs m_1 panel (with the grey region 'Excluded by Cosmology'), which eigenmass band— m_1 , m_2 , or m_3 —occupies the highest y-values?

Answer: m_3

Information Density: 4.2

HID-Chart Example 11

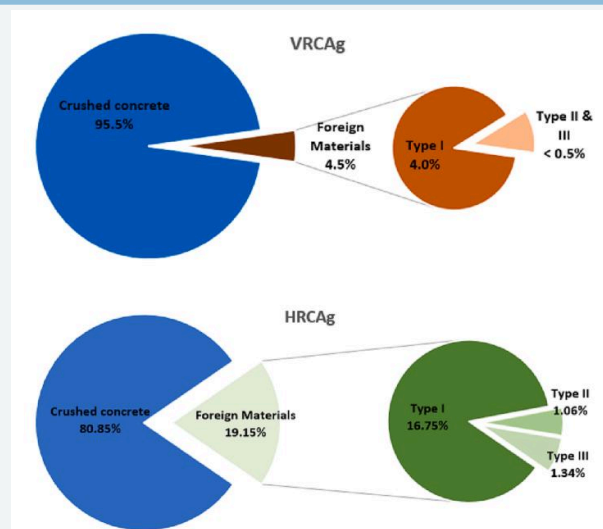


Question: Which treatment simultaneously yields the lowest Caspase 3 and Caspase 7 levels (panels E and F) and the highest PGC-1 α mRNA level (panel G) among BXXXT doses?

Answer: H-G+10% BXXXT

Information Density: 4.2

HID-Chart Example 12

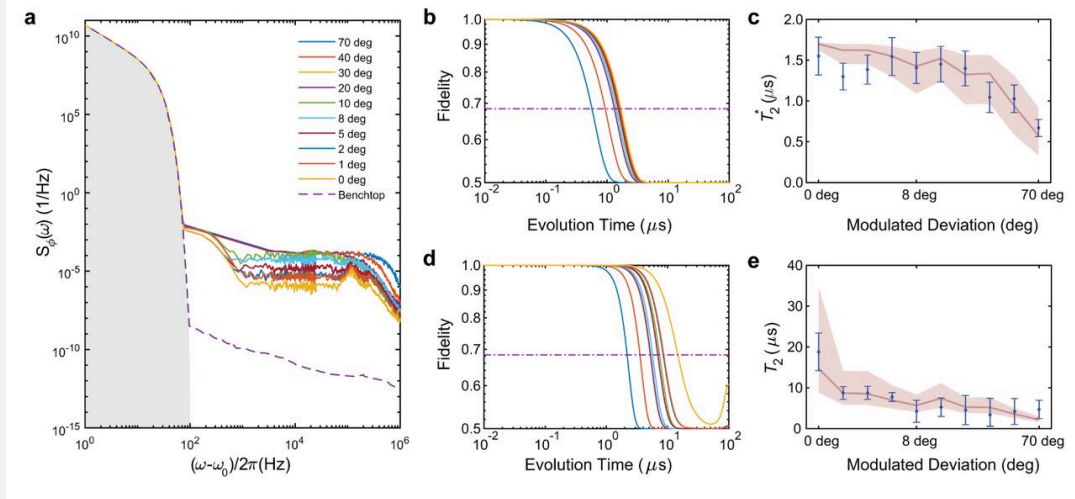


Question: Looking at all cement types (I, II, and III) together, which aggregate type (VRCAg or HRCaG) has a higher total percentage of cement types?

Answer: HRCaG

Information Density: 3.3

HID-Chart Example 13

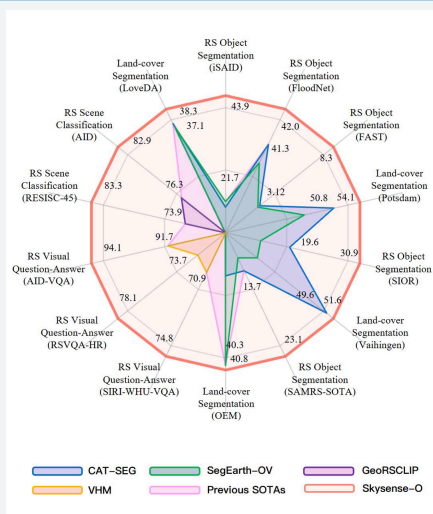


Question: In panel b, at approximately what evolution time (μs) does the 30 deg curve drop to the 0.7 dashed fidelity line?

Answer: Approximately 2.5 μs

Information Density: 4.1

HID-Chart Example 14

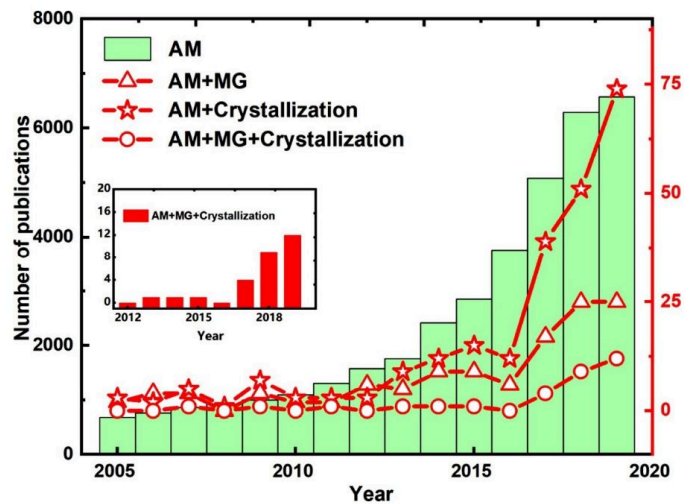


Question: Within the three RS Visual Question-Answer axes, which one has both the longest red radius and the highest numeric value on the red outline?

Answer: AID-VQA

Information Density: 4.0

HID-Chart Example 15



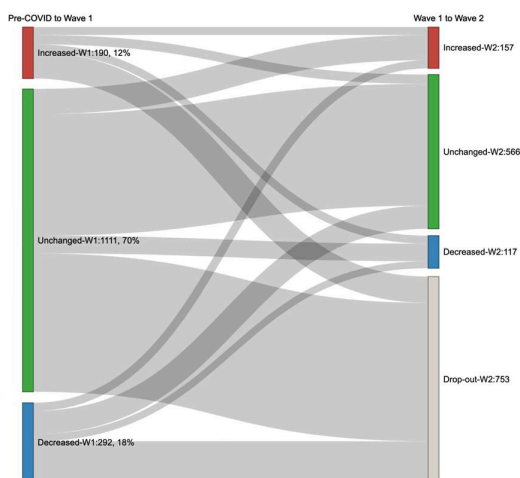
Question: After which year does the AM+Crystallization dataset (star markers) always surpass the AM+MG dataset (triangle markers) in number of publications?

Answer: 2013

Information Density: 4.0

HID-Chart Example 16

Figure 1: Flow chart of change in fertility desire across waves

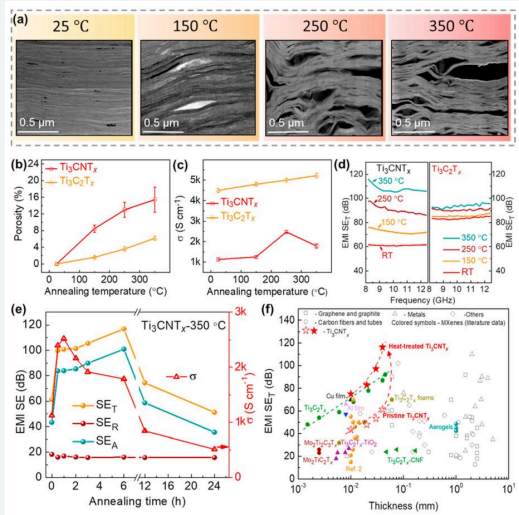


Question: Looking at the Decreased-W2 bar (117), which Wave 1 category supplies the largest share of its incoming flow?

Answer: Unchanged-W1

Information Density: 3.7

HID-Chart Example 17



Question: In (e), after how many hours of annealing does Ti₃CNT_x-350 °C reach its maximum EMI SE_T, and what is the corresponding electrical conductivity (σ) at that point?

Answer: 6 hours, 1700 S/cm

Information Density: 4.2

HID-Chart Example 18

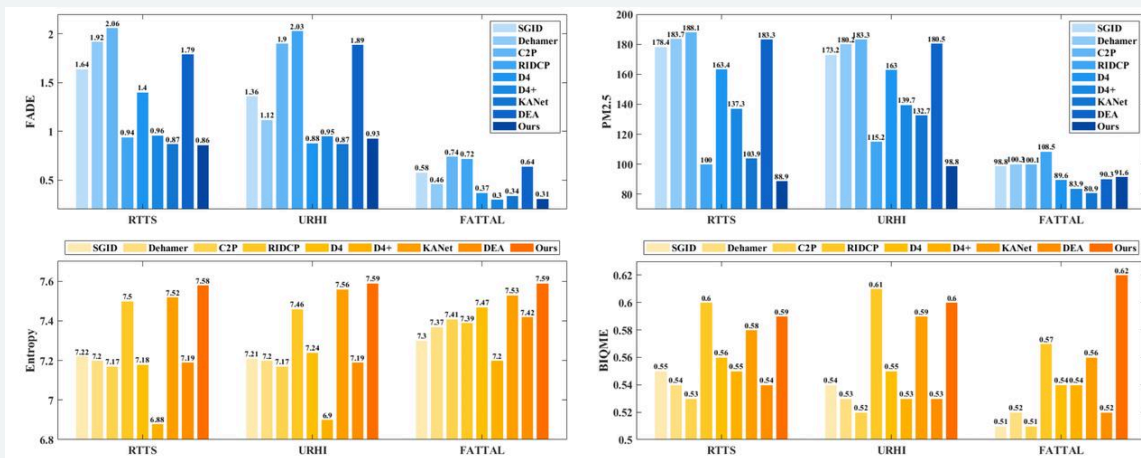


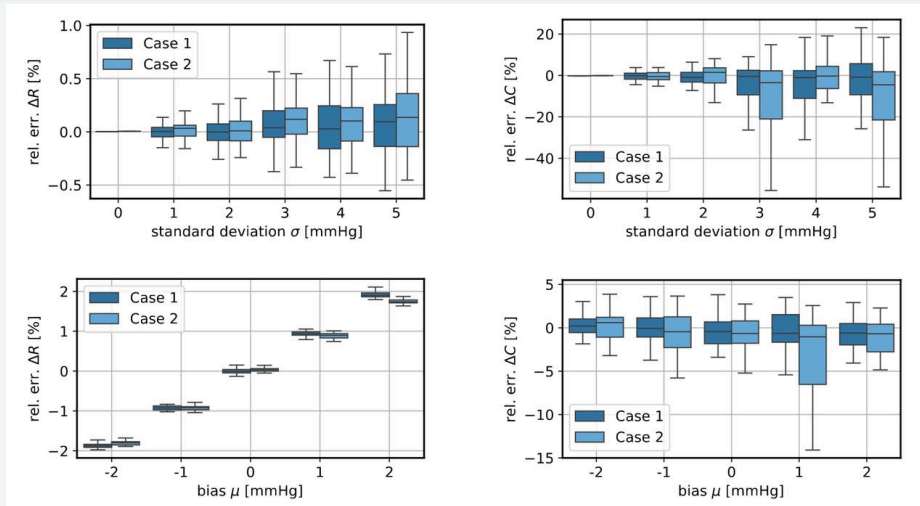
Figure 3. Quantitative comparison on three real-world datasets. Four no-reference image quality assessments were calculated.

Question: Considering the Entropy subplot, in which dataset (RTTS, URHI, FATTAL) is the gap between 'Ours' and the lowest competing method the largest?

Answer: RTTS

Information Density: 4.2

HID-Chart Example 19

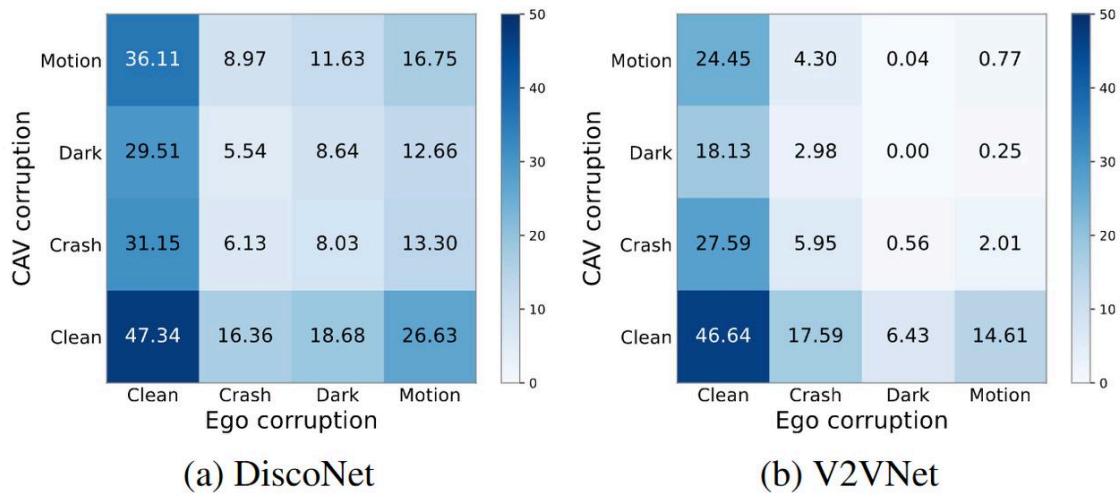


Question: In the ΔC vs σ plot (top-right), for Case 1, at what standard deviation σ [mmHg] is the most negative lower whisker observed?

Answer: 4

Information Density: 4.2

HID-Chart Example 20

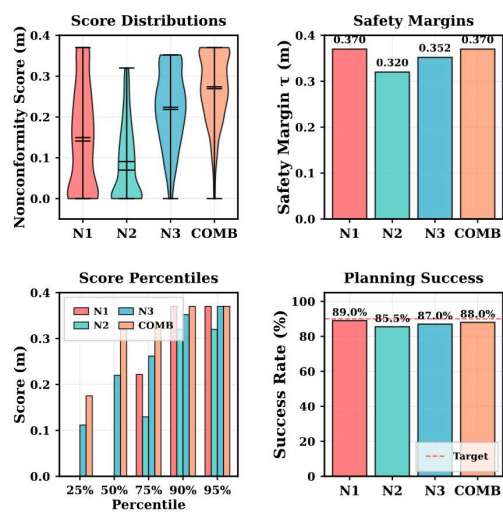


Question: What is the absolute difference between DiscoNet and V2VNet at CAV=Crash and Ego=Clean?

Answer: 3.56

Information Density: 3.7

HID-Chart Example 21



Question: What is the difference (in meters) between the maximum and minimum Safety Margin τ values across N1–COMB?

Answer: 0.050 meters

Information Density: 4.4

References

- [1] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibao Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 1, 3
- [2] Haodong Duan, Junming Yang, Yuxuan Qiao, Xinyu Fang, Lin Chen, Yuan Liu, Xiaoyi Dong, Yuhang Zang, Pan Zhang, Jiaqi Wang, et al. Vlmevalkit: An open-source toolkit for evaluating large multi-modality models. In *Proceedings of the 32nd ACM international conference on multimedia*, pages 11198–11201, 2024. 3
- [3] Muye Huang, Han Lai, Xinyu Zhang, Wenjun Wu, Jie Ma, Lingling Zhang, and Jun Liu. Evochart: A benchmark and a self-training approach towards real-world chart understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3680–3688, 2025. 3
- [4] Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pages 611–626, 2023. 3
- [5] Ahmed Masry, Do Xuan Long, Jia Qing Tan, Shafiq Joty, and Enamul Hoque. Chartqa: A benchmark for question answering about charts with visual and logical reasoning. In *Findings of the association for computational linguistics: ACL 2022*, pages 2263–2279, 2022. 3
- [6] Nitesh Methani, Pritha Ganguly, Mitesh M Khapra, and Pratyush Kumar. Plotqa: Reasoning over scientific plots. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1527–1536, 2020. 3
- [7] Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. In *Proceedings of the Twentieth European Conference on Computer Systems*, pages 1279–1297, 2025. 1
- [8] Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, et al. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*, 2025. 3
- [9] Zirui Wang, Mengzhou Xia, Luxi He, Howard Chen, Yitao Liu, Richard Zhu, Kaiqu Liang, Xindi Wu, Haotian Liu, Sadhika Malladi, et al. Charxiv: Charting gaps in realistic chart understanding in multimodal llms. *Advances in Neural Information Processing Systems*, 37:113569–113697, 2024. 3
- [10] Zhengzhuo Xu, Sinan Du, Yiyang Qi, Chengjin Xu, Chun Yuan, and Jian Guo. Chartbench: A benchmark for complex visual reasoning in charts. *arXiv preprint arXiv:2312.15915*, 2023. 3
- [11] Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, and Zheyang Luo. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 3: system demonstrations)*, pages 400–410, 2024. 1