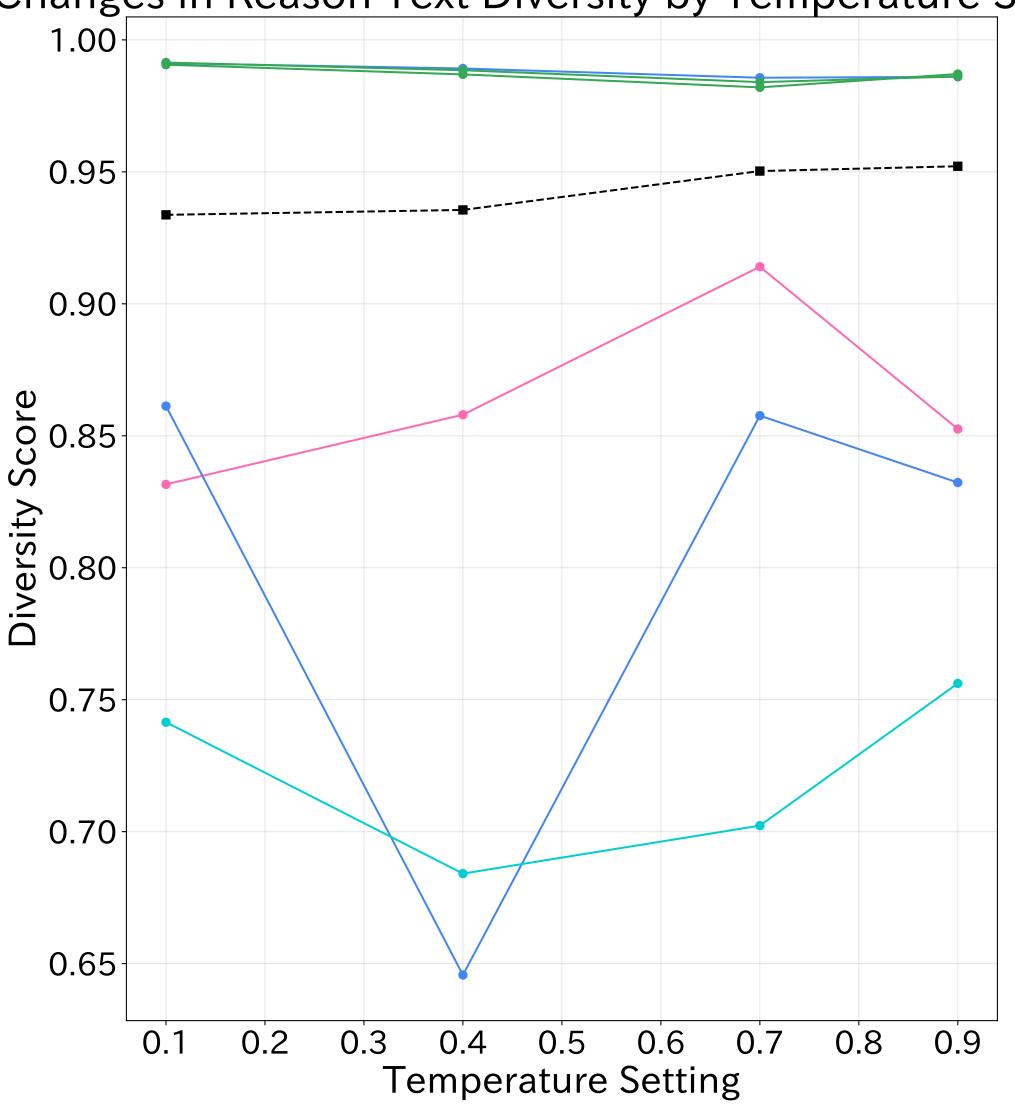
Changes in Reason Text Diversity by Temperature Setting



gpt-4o (High Diversity)
gemini-2.0-flash-thinking-exp (High Diversity)
gemini-2.5-flash-preview-04-17 (High Diversity)
gemma-3-27b-it (Low Diversity)
Llama-4-Scout-17B (Low Diversity)
Qwen2.5-VL-7B-Instruct (Low Diversity)
Overall Average