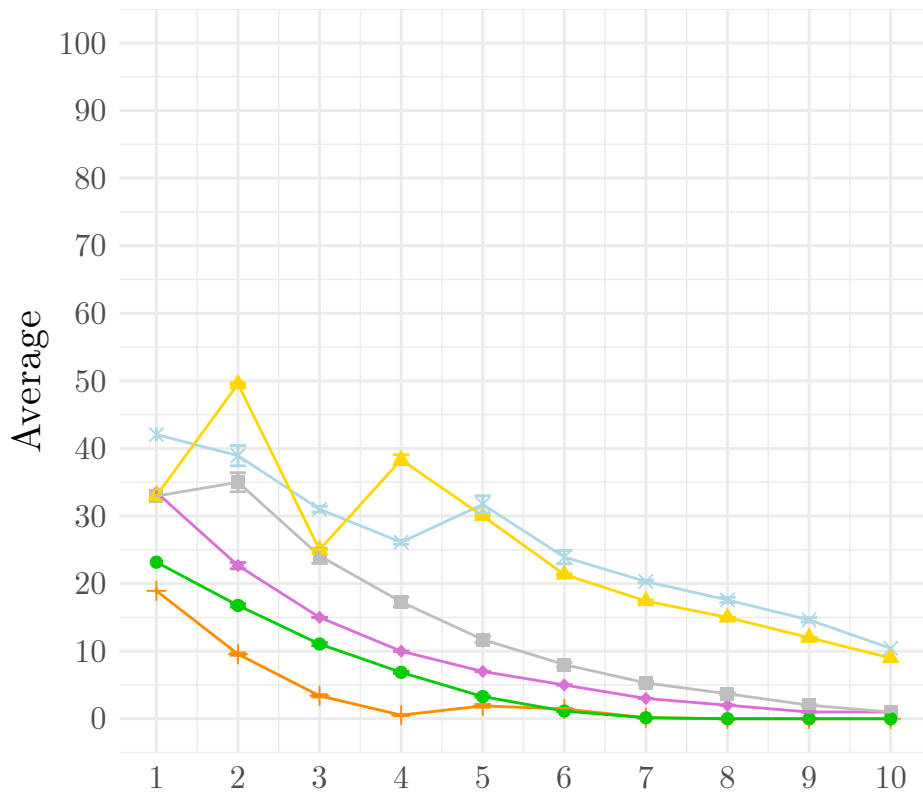


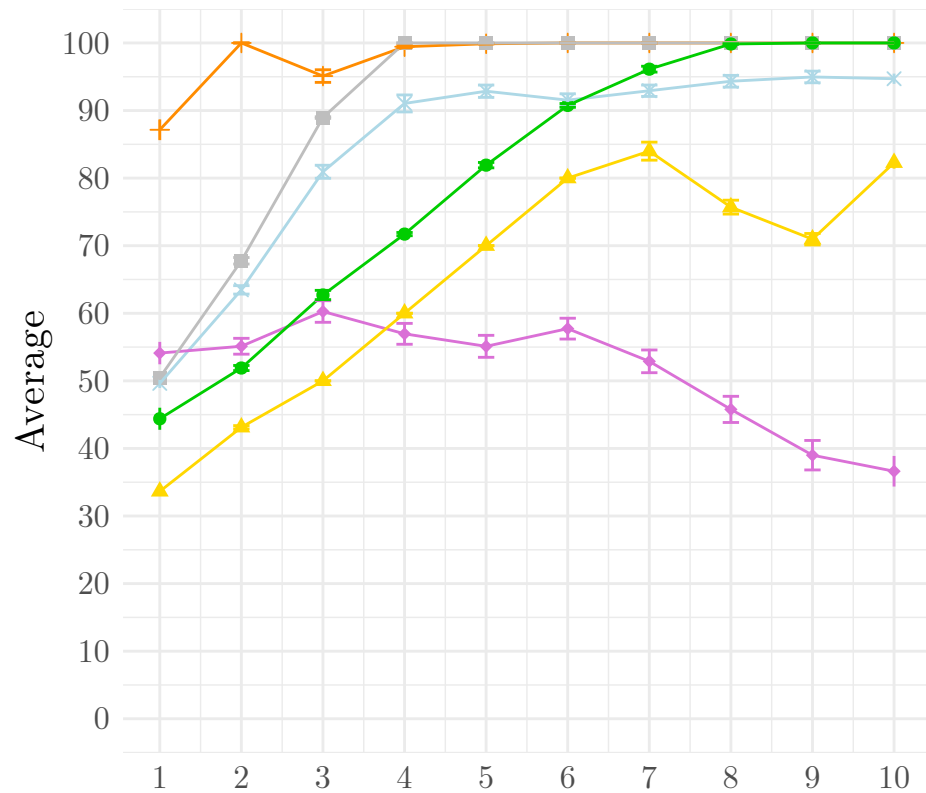
$p = 2/3$



Round

● Gemini-2.0 Flash (Thinking) ■ Claude-3.5 (Sonnet) + ChatGPT-o1
▲ Gemini-1.5 Flash ◆ Claude-2.1 × ChatGPT-4

$p = 4/3$



Round