LLM file win percentage by winner group size (type precision score)

Legend: gpt-3.5, gpt-4o, o1-mini, o3-mini, claude3-sonnet, deepseek

X-axis: Winner group size (number of top scorers)
Y-axis: Percentage of files won (%)

Group 1: gpt-3.5 16.81, gpt-4o 16.91, o1-mini 19.32, o3-mini 6.47, claude3-sonnet 17.65, deepseek 22.83
Group 2: gpt-3.5 2.94, gpt-4o 15.03, o1-mini 15.36, o3-mini 15.03, claude3-sonnet 27.78, deepseek 23.86
Group 3: gpt-3.5 2.56, gpt-4o 21.61, o1-mini 17.58, o3-mini 15.02, claude3-sonnet 21.98, deepseek 21.25
Group 4: gpt-3.5 1.71, gpt-4o 20.21, o1-mini 21.58, o3-mini 15.75, claude3-sonnet 20.55, deepseek 20.21
Group 5: gpt-3.5 3.29, gpt-4o 19.45, o1-mini 18.90, o3-mini 19.18, claude3-sonnet 20.00, deepseek 19.18
Group 6: gpt-3.5 16.67, gpt-4o 16.67, o1-mini 16.67, o3-mini 16.67, claude3-sonnet 16.67, deepseek 16.67