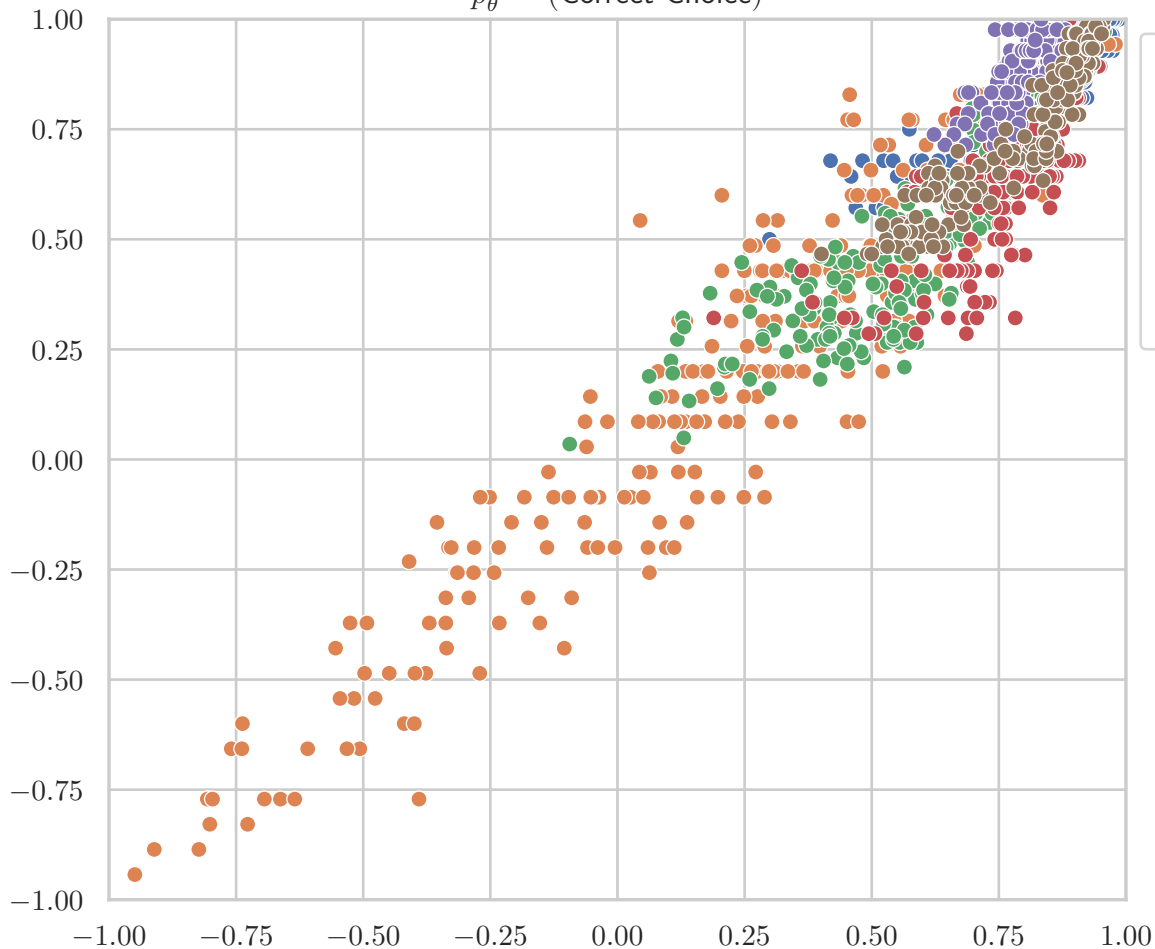


MMLU Prehistory  
 $p_{\theta}^{\text{Vocab}}(\text{Correct Choice})$

Spearman Correlation  
Between FLOPs and Scores (Per Sample)



Pearson Correlation  
Between FLOPs and Scores (Per Sample)

