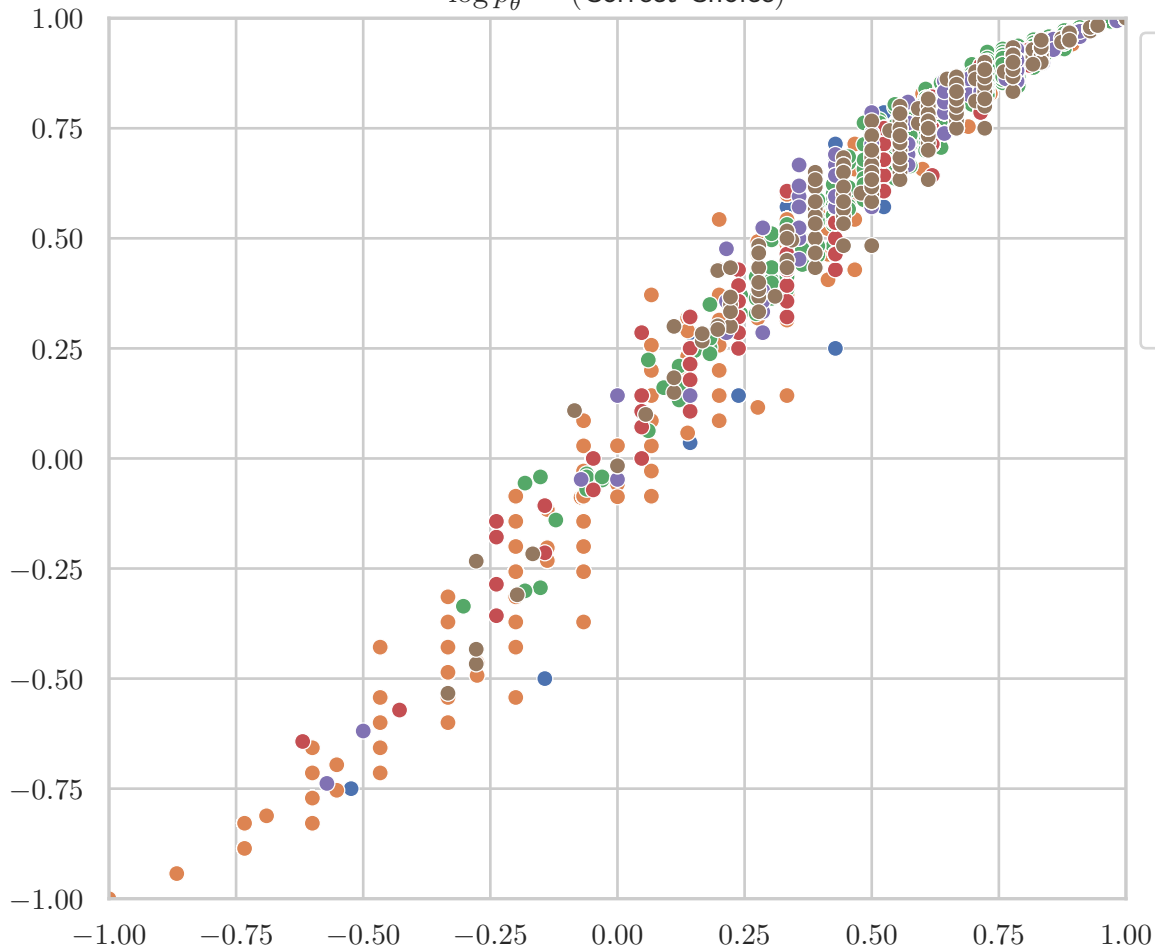


ARC-Challenge
 $\log p_{\theta}^{\text{Vocab}}(\text{Correct Choice})$

Spearman Correlation
Between FLOPs and Scores (Per Sample)



Kendall Correlation
Between FLOPs and Scores (Per Sample)

