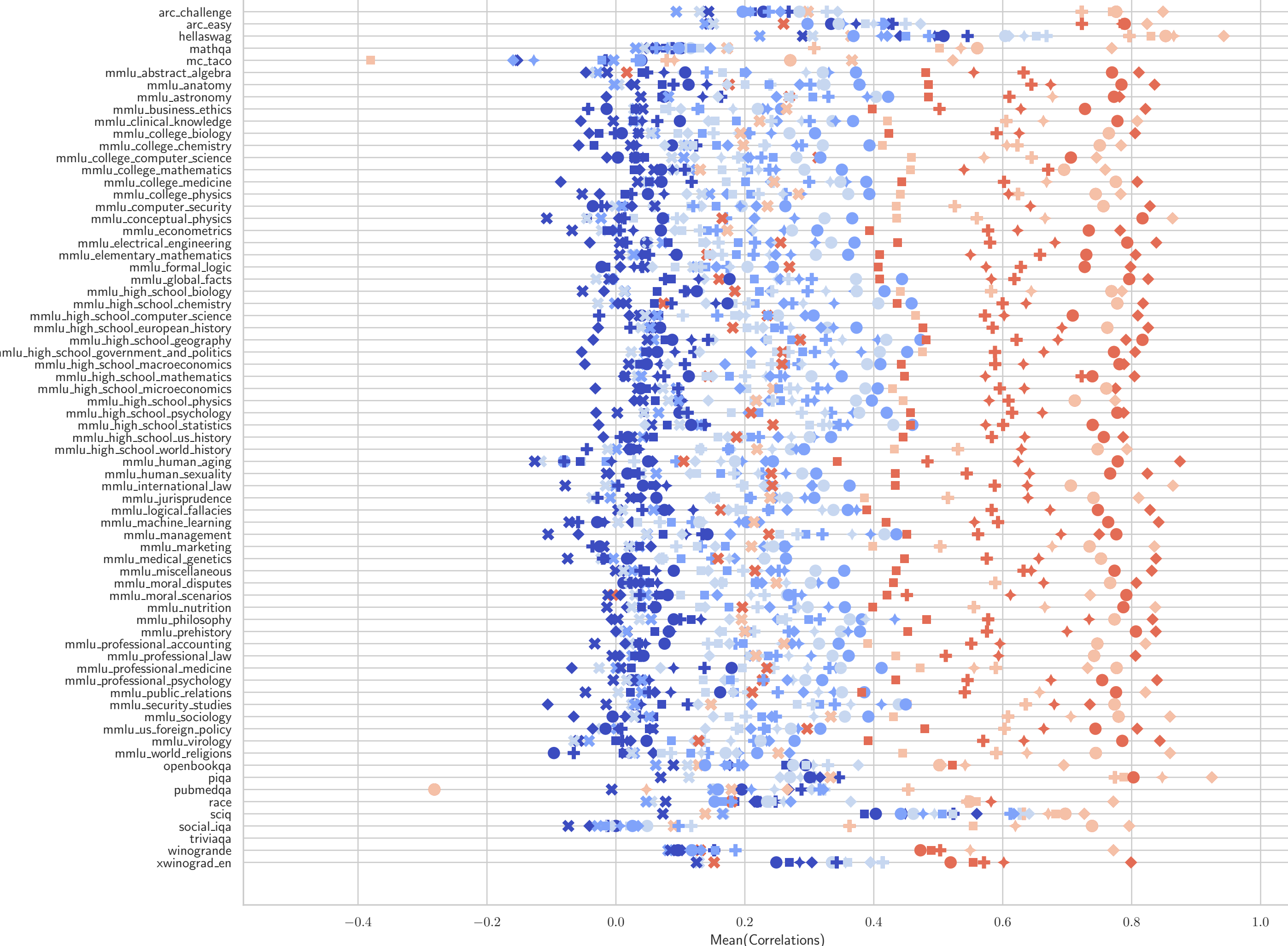
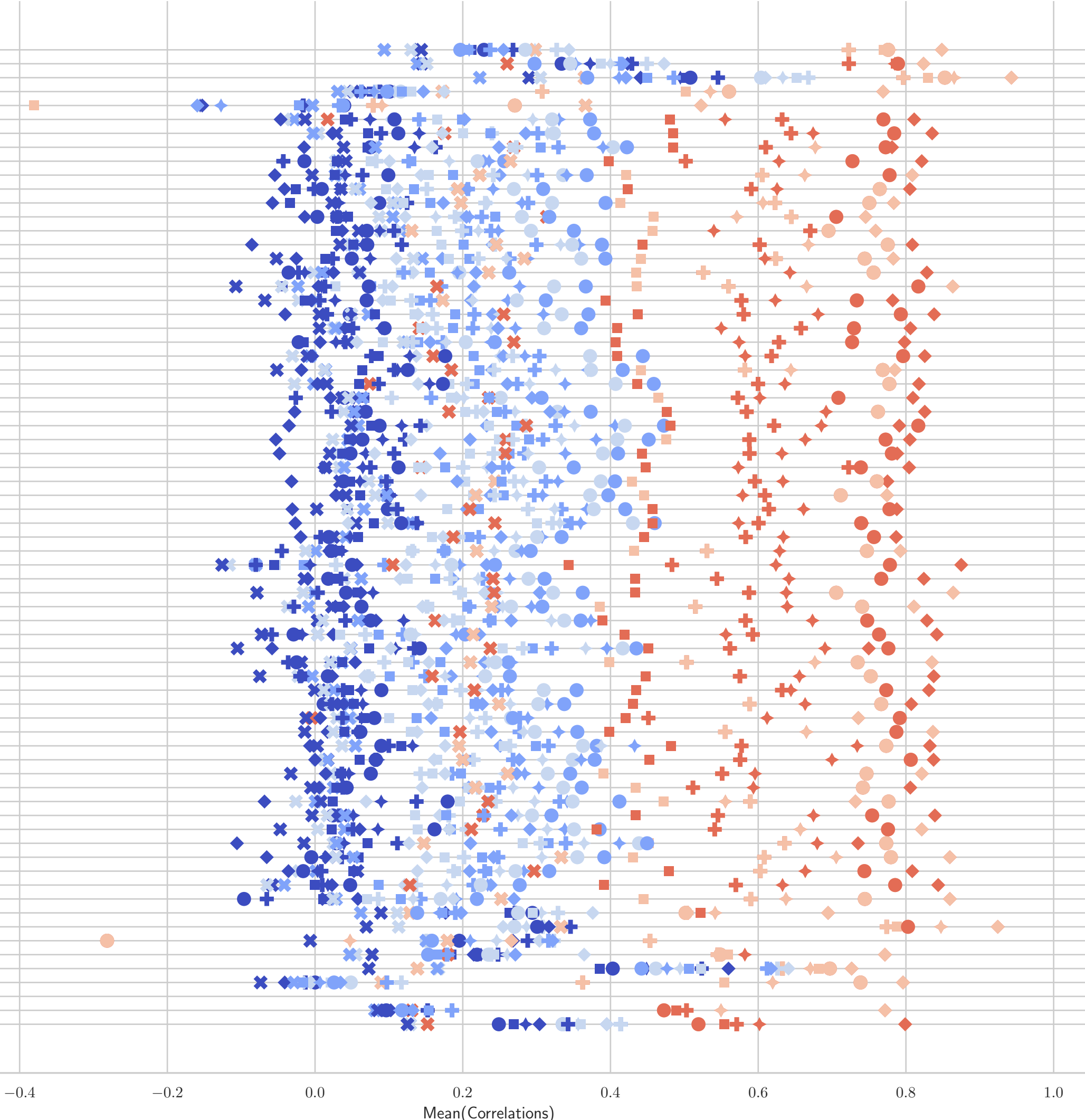


Benchmark (and Optional Task)

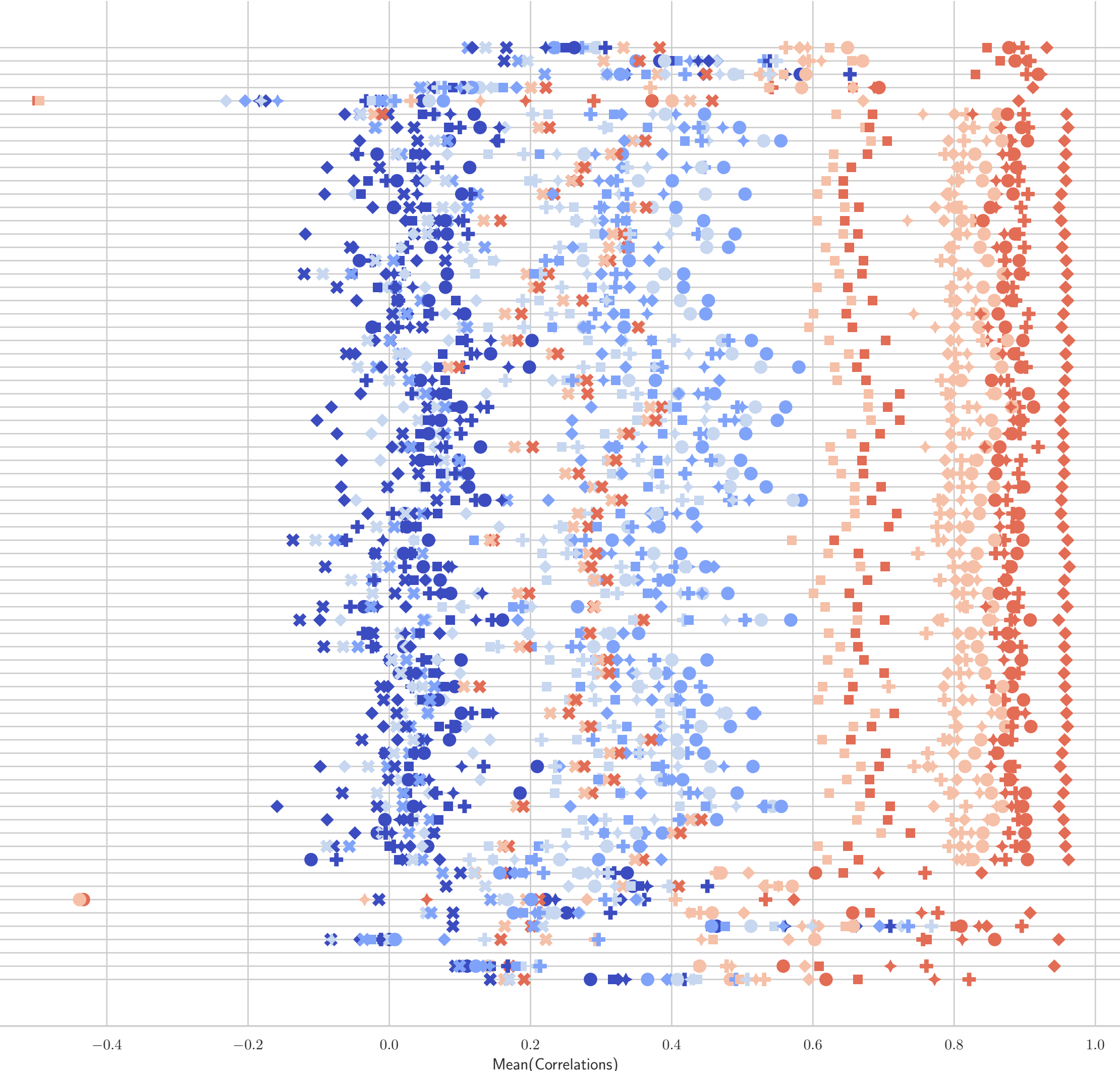


Correlation: kendall

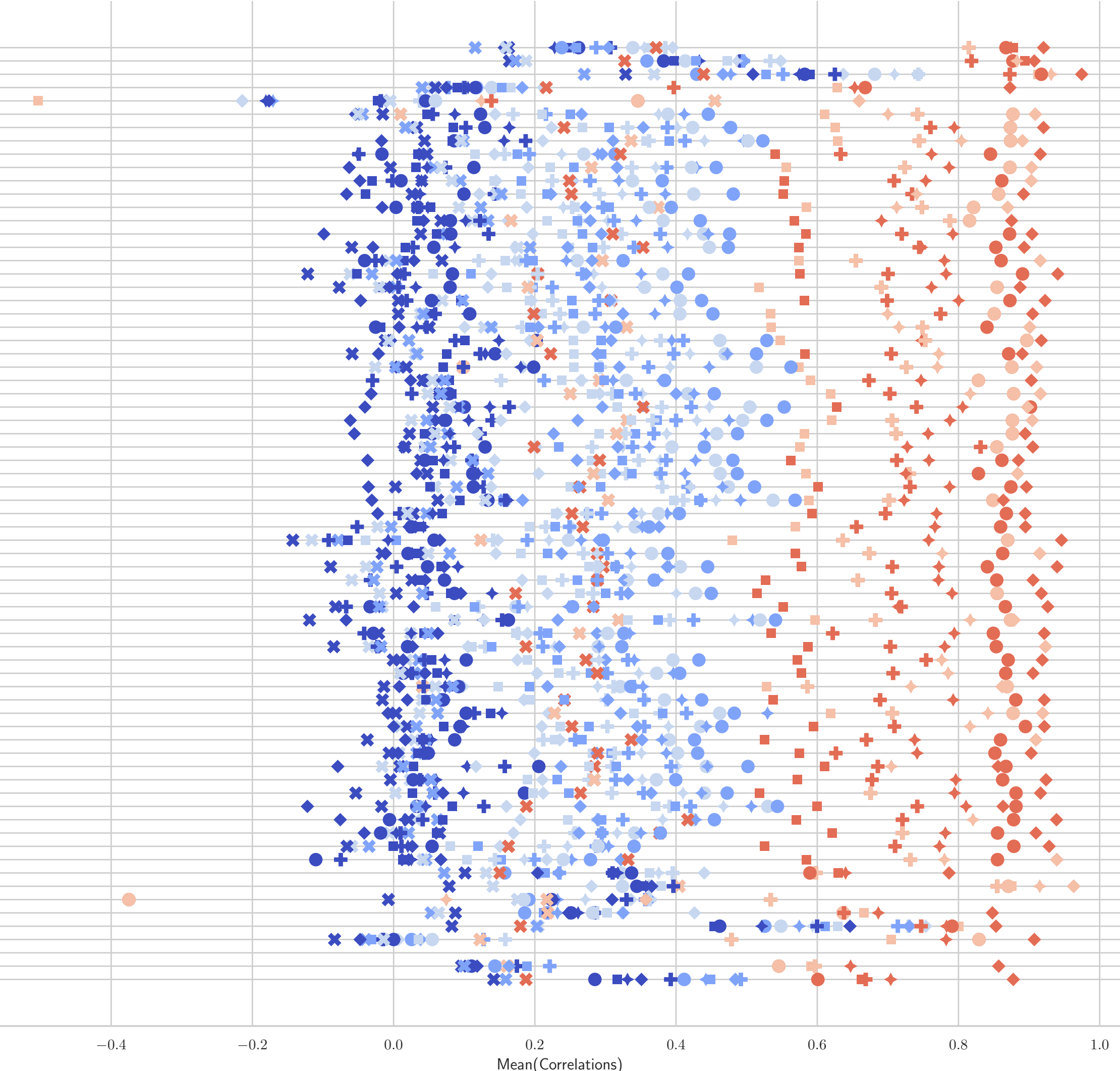


Mean(Correlations) by Benchmark and Metric

Correlation: pearson



Correlation: spearman



- metric
- log_prob_vocab_correct
 - prob_vocab_correct
 - prob_choices_correct
 - brier_score
 - acc
- Model Family
- Cerebras (Param. and Data Scaling)
 - INCITE 7B Param. (Data Scaling)
 - LLM360 Amber 7B Tokens (Param Scaling)
 - OLMo 7B Param. (Data Scaling)
 - Pythia 12B Param. (Data Scaling)
 - Pythia 300B Tokens (Param. Scaling)