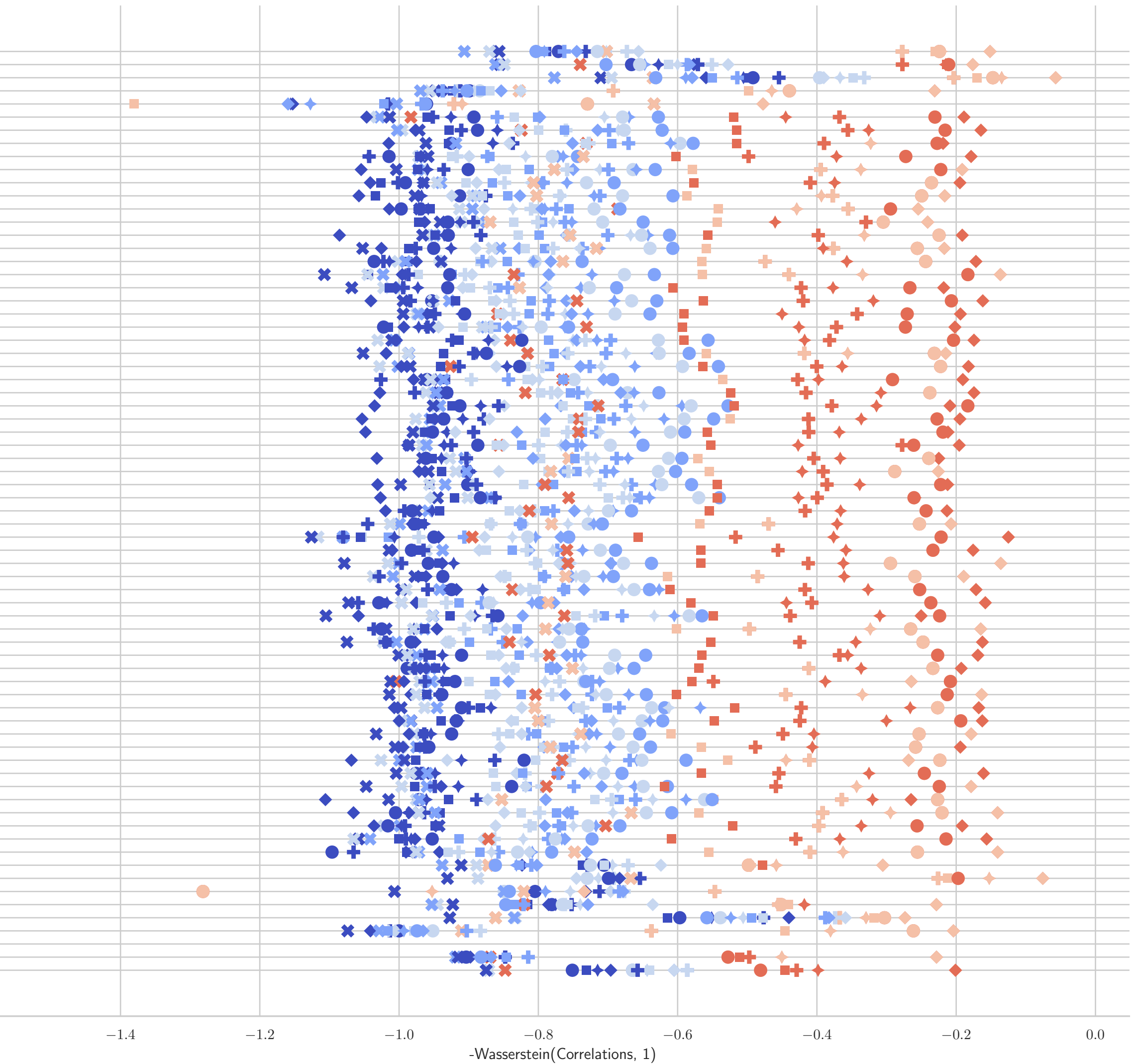


- arc_challenge
- arc_easy
- hellaswag
- mathqa
- mc_taco
- mmlu_abstract_algebra
- mmlu_anatomy
- mmlu_astronomy
- mmlu_business_ethics
- mmlu_clinical_knowledge
- mmlu_college_biology
- mmlu_college_chemistry
- mmlu_college_computer_science
- mmlu_college_mathematics
- mmlu_college_medicine
- mmlu_college_physics
- mmlu_computer_security
- mmlu_conceptual_physics
- mmlu_econometrics
- mmlu_electrical_engineering
- mmlu_elementary_mathematics
- mmlu_formal_logic
- mmlu_global_facts
- mmlu_high_school_biology
- mmlu_high_school_chemistry
- mmlu_high_school_computer_science
- mmlu_high_school_european_history
- mmlu_high_school_geography
- mmlu_high_school_government_and_politics
- mmlu_high_school_macro_economics
- mmlu_high_school_mathematics
- mmlu_high_school_microeconomics
- mmlu_high_school_physics
- mmlu_high_school_psychology
- mmlu_high_school_statistics
- mmlu_high_school_us_history
- mmlu_high_school_world_history
- mmlu_human_aging
- mmlu_human_sexuality
- mmlu_international_law
- mmlu_jurisprudence
- mmlu_logical_fallacies
- mmlu_machine_learning
- mmlu_management
- mmlu_marketing
- mmlu_medical_genetics
- mmlu_miscellaneous
- mmlu_moral_disputes
- mmlu_moral_scenarios
- mmlu_nutrition
- mmlu_philosophy
- mmlu_prehistory
- mmlu_professional_accounting
- mmlu_professional_law
- mmlu_professional_medicine
- mmlu_professional_psychology
- mmlu_public_relations
- mmlu_security_studies
- mmlu_sociology
- mmlu_us_foreign_policy
- mmlu_virology
- mmlu_world_religions
- openbookqa
- piqa
- pubmedqa
- race
- sciq
- social_lqa
- triviaqa
- winogrande
- xwinograd.en

Correlation: kendall

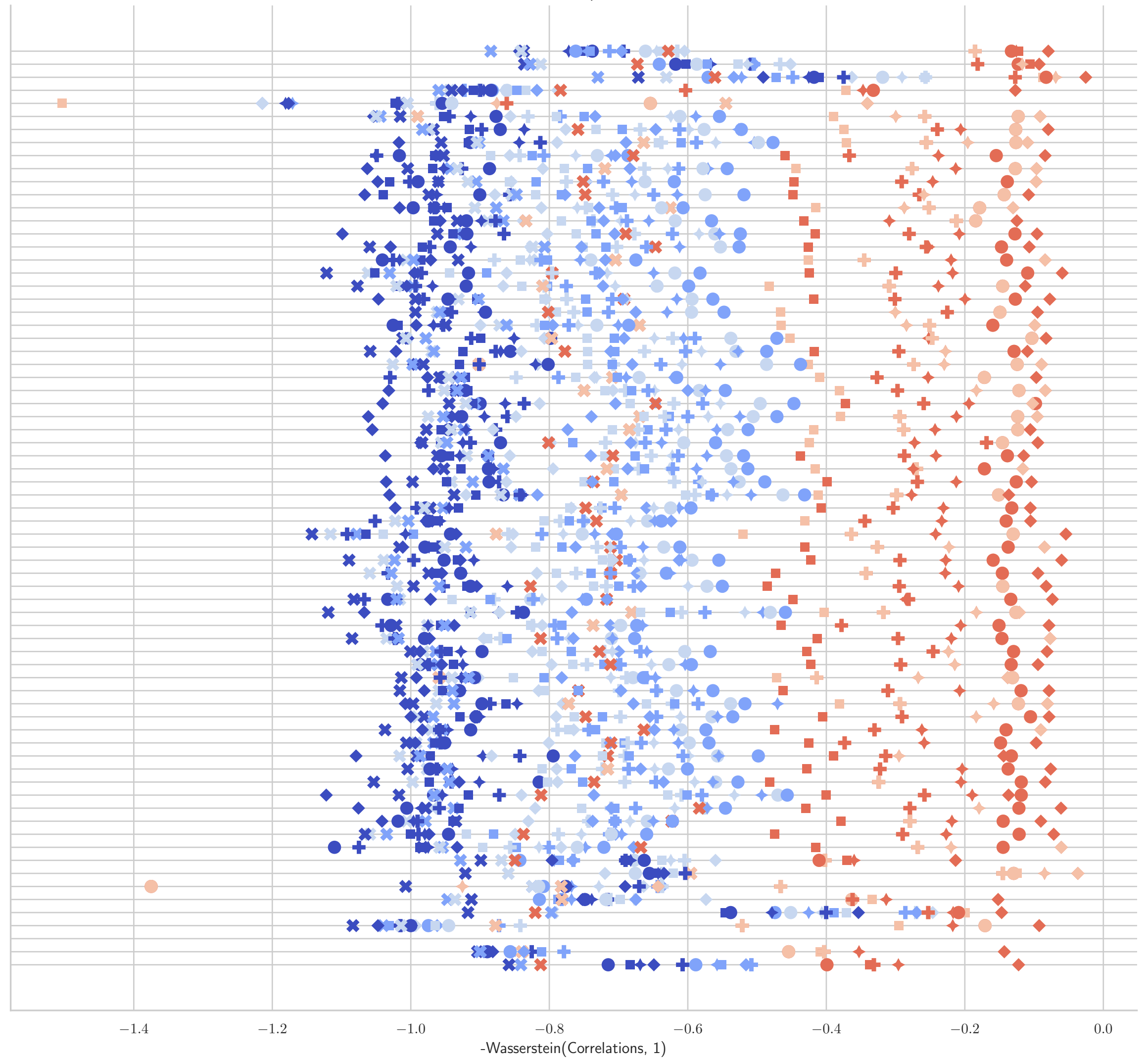


-Wasserstein(Correlations, 1) by Benchmark and Metric

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)