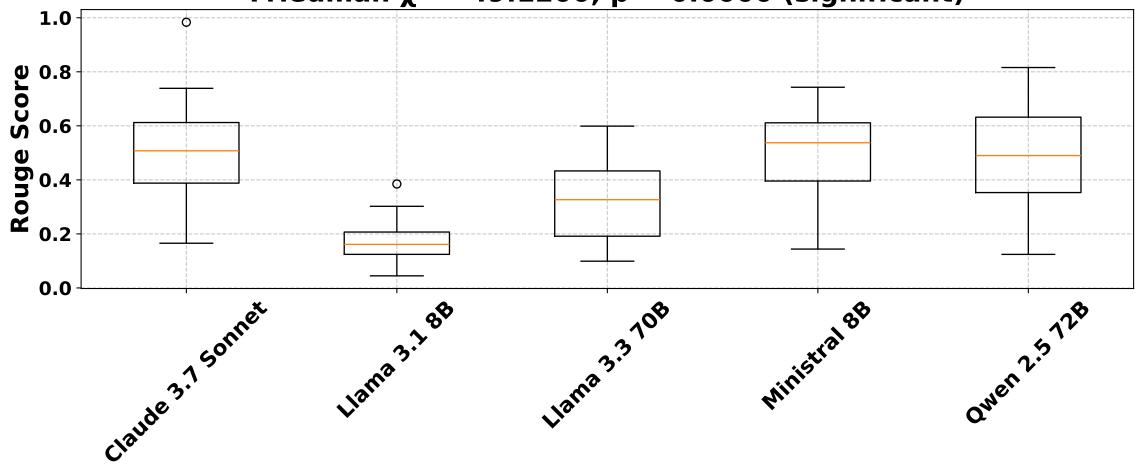Comparison of rouge_score across models
Friedman χ^2 = 49.1200, p = 0.0000 (significant)