Results

Descriptive Statistics

Descriptive Statistics

		Valid	Missing	Mean	Std. Deviation	Shapiro-Wilk	P-value of Shapiro-Wilk	Minimum	Maximum
rank	abalone	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	adult	2790	0	1.968	0.823	0.790	< .001	1.000	3.000
rank	air_quality	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	bike	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	car	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	fish_toxicity	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	forest_fires	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	housing	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	iris	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	mushroom	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	parkinsons	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	student_performance	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	wine_quality	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	bank	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
rank	diabetic	2790	0	2.000	0.817	0.793	< .001	1.000	3.000
test_loss	abalone	2790	0	2.354	0.411	0.803	< .001	1.919	6.404
test_loss	adult	2790	0	81.221	361.595	0.227	< .001	0.311	6186.153
test_loss	air_quality	2790	0	0.276	0.050	0.365	< .001	0.242	0.748
test_loss	bike	2790	0	0.103	0.071	0.575	< .001	0.046	0.663
test_loss	car	2790	0	0.393	0.356	0.711	< .001	0.077	2.845
test_loss	fish_toxicity	2790	0	0.114	0.043	0.369	< .001	0.083	0.537
test_loss	forest_fires	2790	0	0.082	0.102	0.527	< .001	0.011	0.880
test_loss	housing	2790	0	0.116	0.051	0.585	< .001	0.063	0.580
test_loss	iris	2790	0	0.285	0.335	0.606	< .001	0.014	6.272
test_loss	mushroom	2790	0	0.168	1.693	0.044	< .001	0.000	60.909
test_loss	parkinsons	2790	0	0.073	0.056	0.244	< .001	0.053	0.656
test_loss	student_performance	2790	0	0.213	0.080	0.541	< .001	0.144	0.615
test_loss	wine_quality	2790	0	1.189	0.255	0.462	< .001	1.015	2.948
test_loss	bank	2790	0	0.261	0.139	0.269	< .001	0.202	1.991
test_loss	diabetic	2790	0	0.978	0.334	0.223	< .001	0.881	8.668

ANOVA

ANOVA - rank

Cases	Sum of Squares	df	Mean Square	F	p
dataset	2.710	14	0.194	0.495	0.938
heuristic_pool	7193.497	2	3596.749	9192.243	< .001
dataset ✻ heuristic_pool	4376.103	28	156.289	399.430	< .001
Residuals	16357.497	41805	0.391		

Note. Type III Sum of Squares

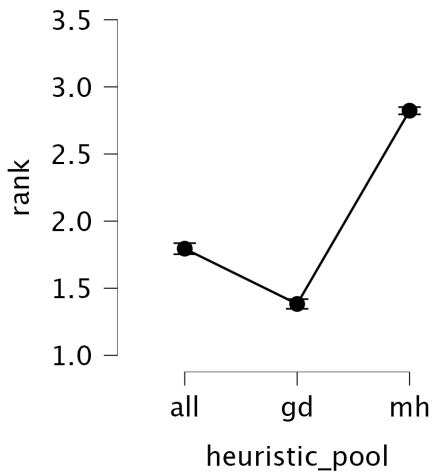
Descriptives

Descriptives - rank

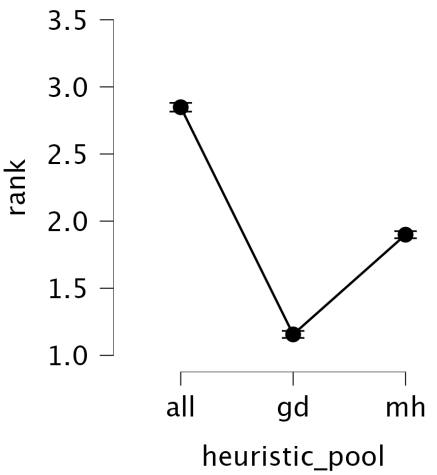
dataset	heuristic_pool	Mean	SD	N
abalone	all	1.795	0.641	930
	gd	1.383	0.560	930
	mh	2.823	0.420	930
adult	all	2.848	0.506	930
	gd	1.156	0.408	930
	mh	1.899	0.410	930
air_quality	all	2.187	0.815	930
	gd	1.683	0.751	930
	mh	2.130	0.788	930
bank	all	1.773	0.562	930
	gd	1.322	0.492	930
	mh	2.905	0.334	930
bike	all	1.625	0.553	930
	gd	1.439	0.532	930
	mh	2.937	0.281	930
car	all	1.590	0.511	930
	gd	1.441	0.518	930
	mh	2.969	0.228	930
diabetic	all	2.452	0.735	930
	gd	1.281	0.576	930
	mh	2.268	0.580	930
fish_toxicity	all	1.990	0.849	930
	gd	1.834	0.768	930
	mh	2.175	0.796	930
forest_fires	all	2.067	0.834	930
	gd	1.827	0.781	930
	mh	2.106	0.807	930
housing	all	1.604	0.687	930
	gd	1.660	0.665	930
	mh	2.735	0.524	930
iris	all	1.858	0.800	930
	gd	1.781	0.727	930
	mh	2.361	0.796	930
mushroom	all	1.520	0.581	930
	gd	1.548	0.527	930
	mh	2.931	0.289	930
parkinsons	all	1.557	0.624	930
	gd	1.601	0.591	930
	mh	2.842	0.445	930
student_performance	all	1.983	0.822	930
	gd	1.790	0.848	930
	mh	2.227	0.715	930
wine_quality	all	1.578	0.550	930
	gd	1.494	0.563	930
	mh	2.928	0.294	930

Descriptives plots

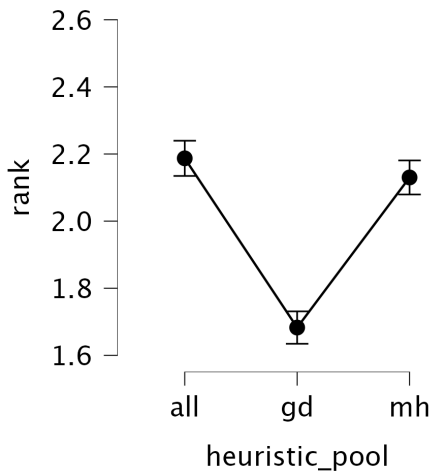
dataset: abalone



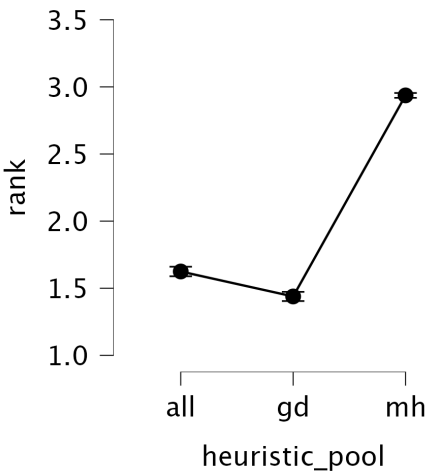
dataset: adult



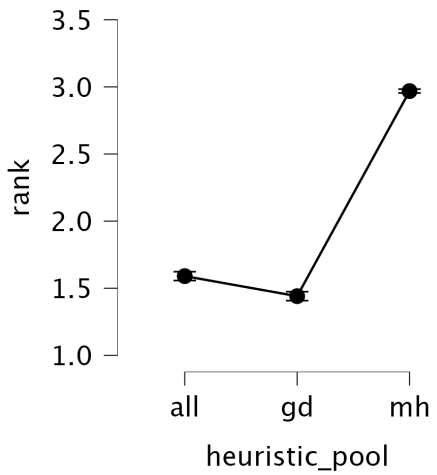
dataset: air_quality



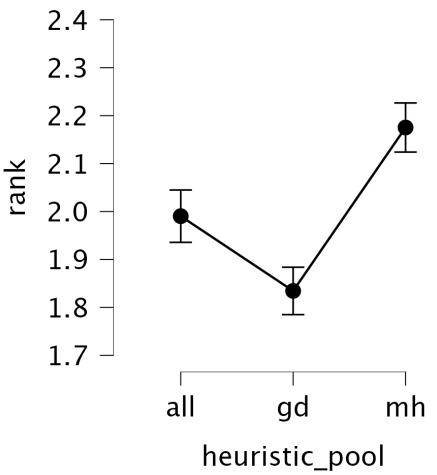
dataset: bike



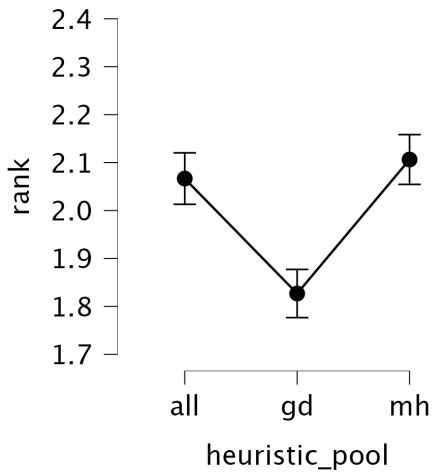
dataset: car



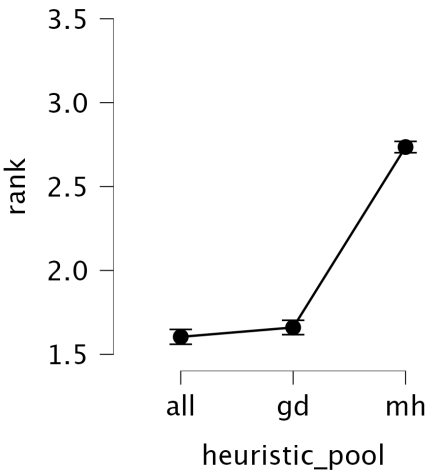
dataset: fish_toxicity



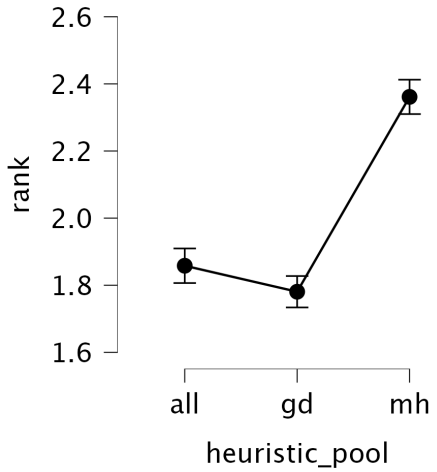
dataset: forest_fires



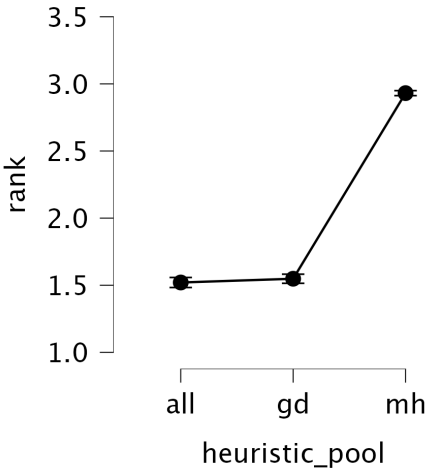
dataset: housing



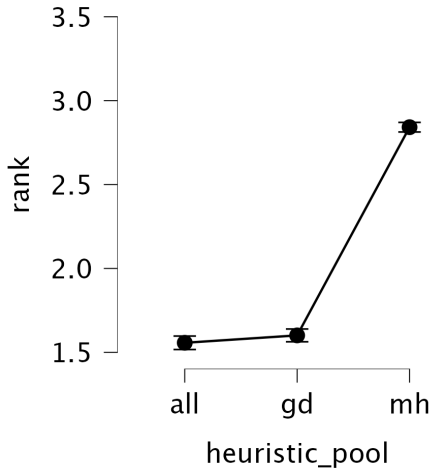
dataset: iris



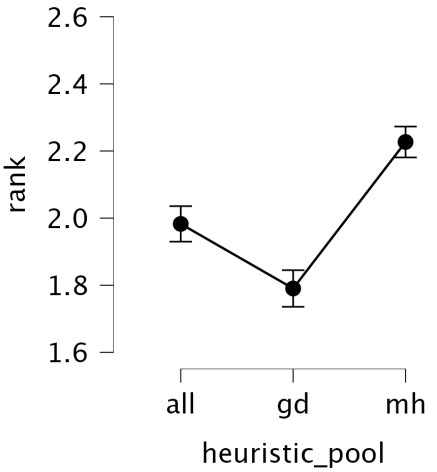
dataset: mushroom



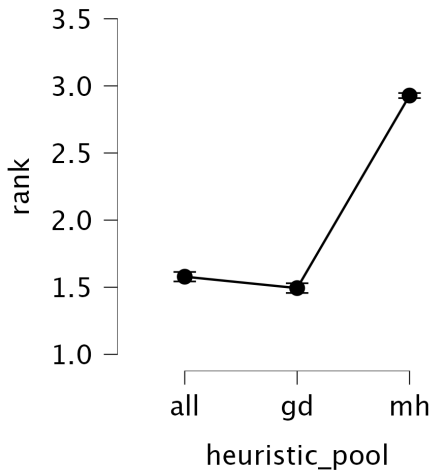
dataset: parkinsons



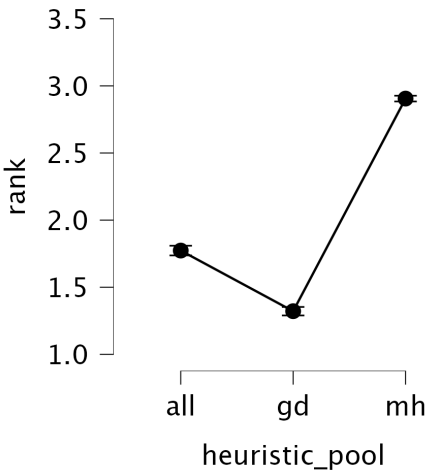
dataset: student_performance



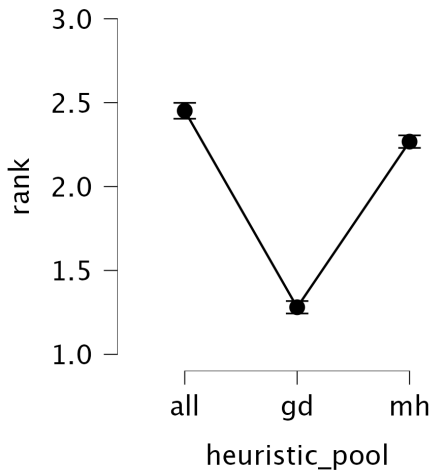
dataset: wine_quality



dataset: bank

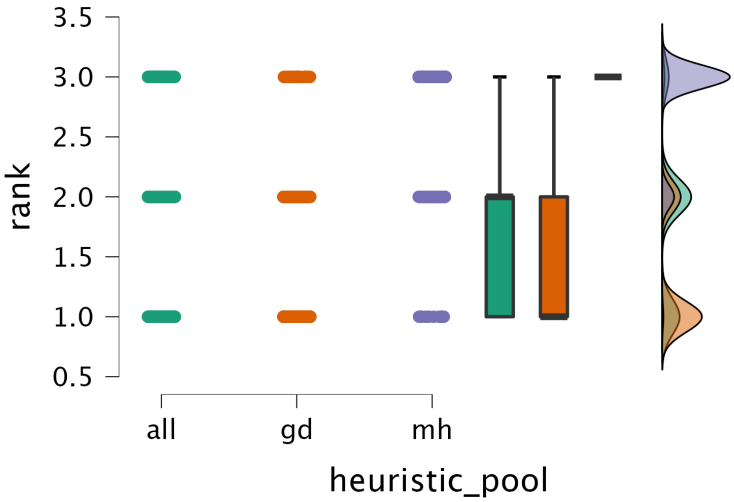


dataset: diabetic

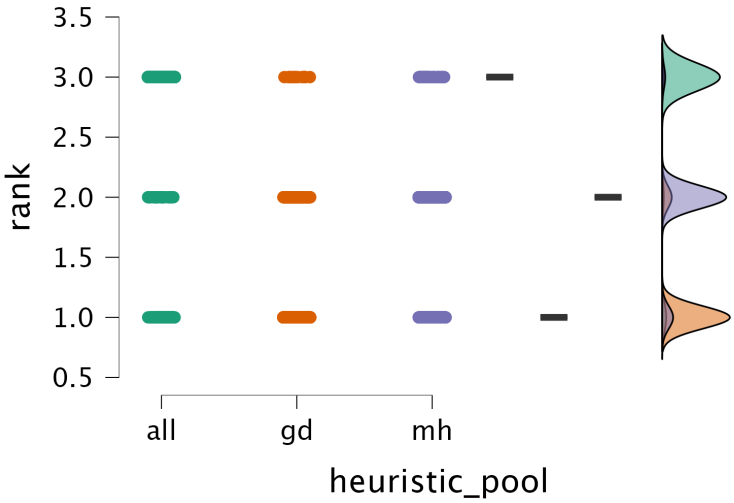


Raincloud plots

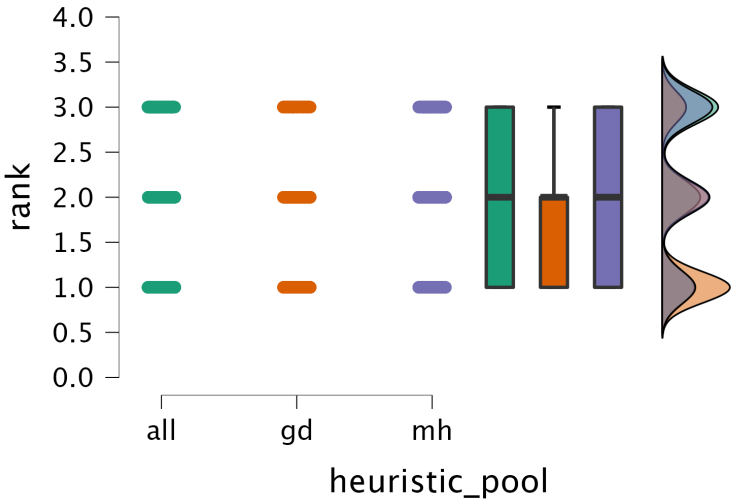
dataset: abalone



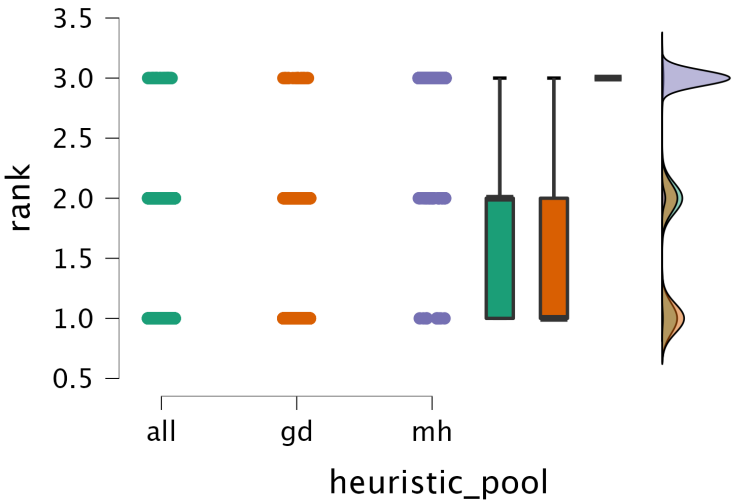
dataset: adult



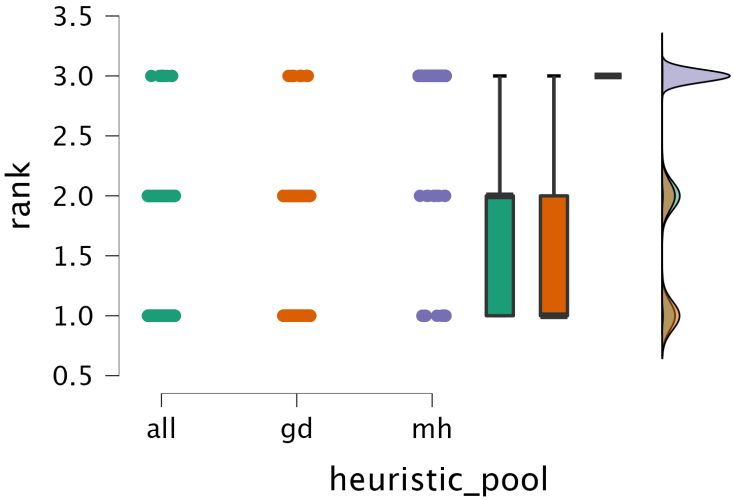
dataset: air_quality



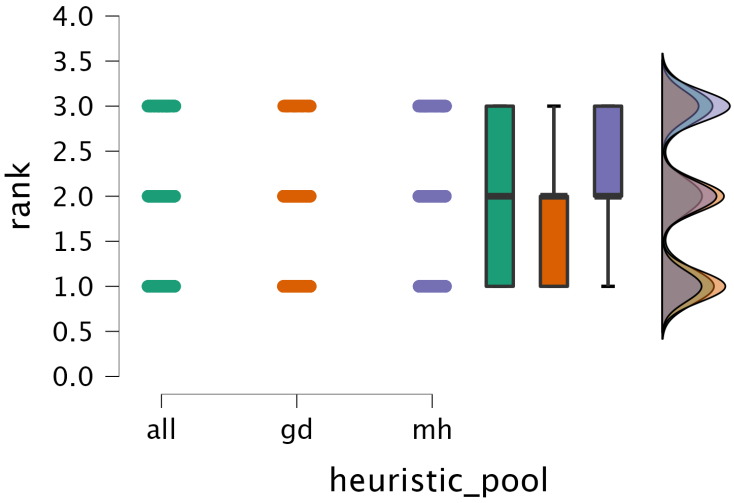
dataset: bike



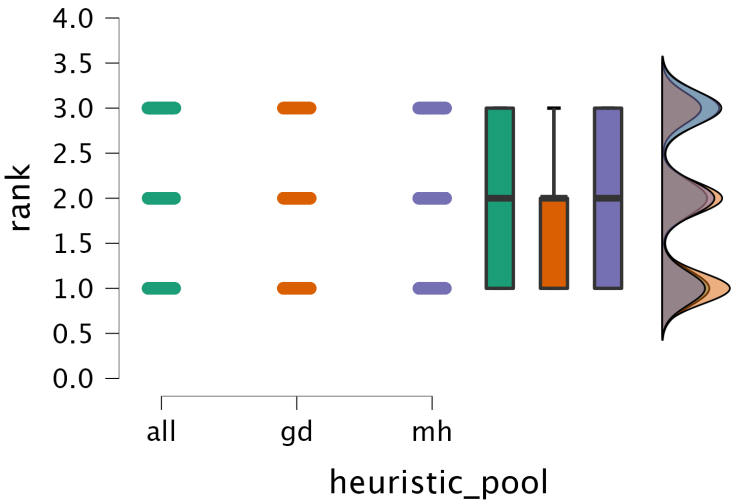
dataset: car



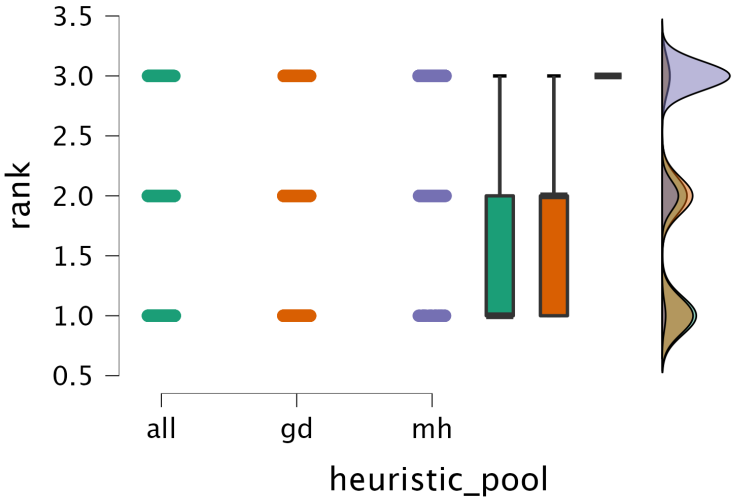
dataset: fish_toxicity



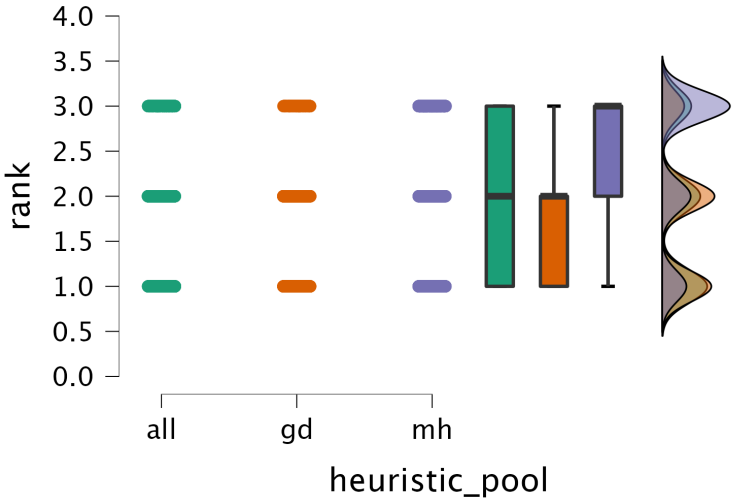
dataset: forest_fires



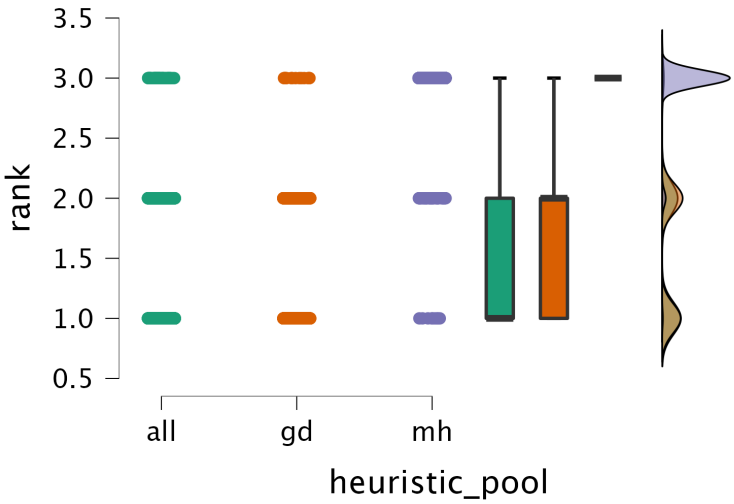
dataset: housing



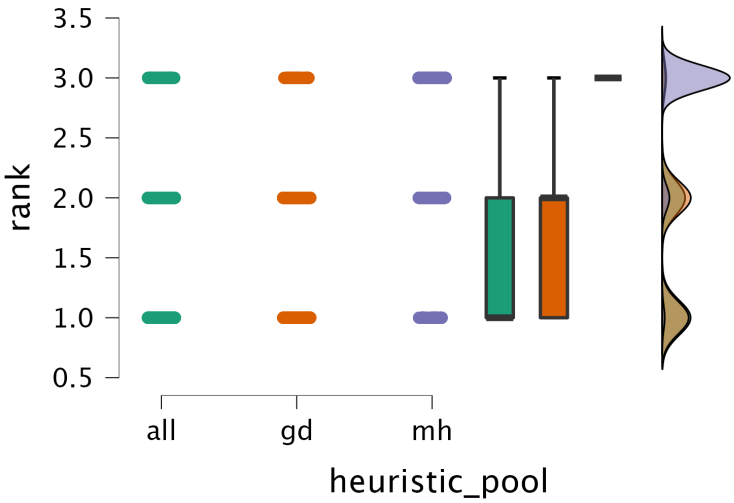
dataset: iris



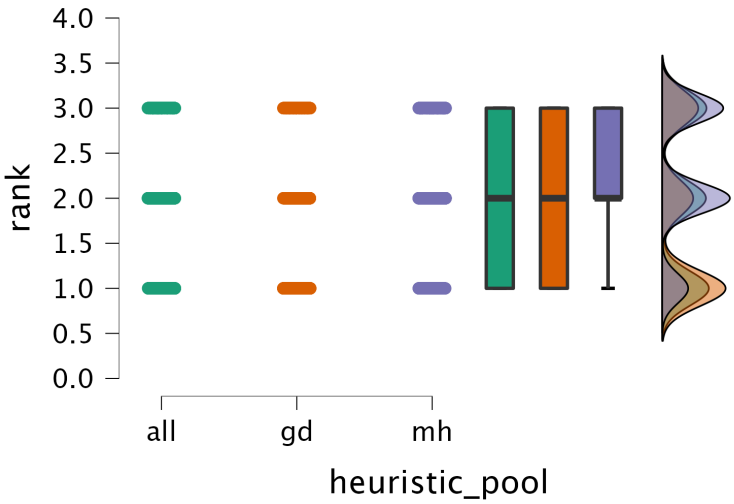
dataset: mushroom



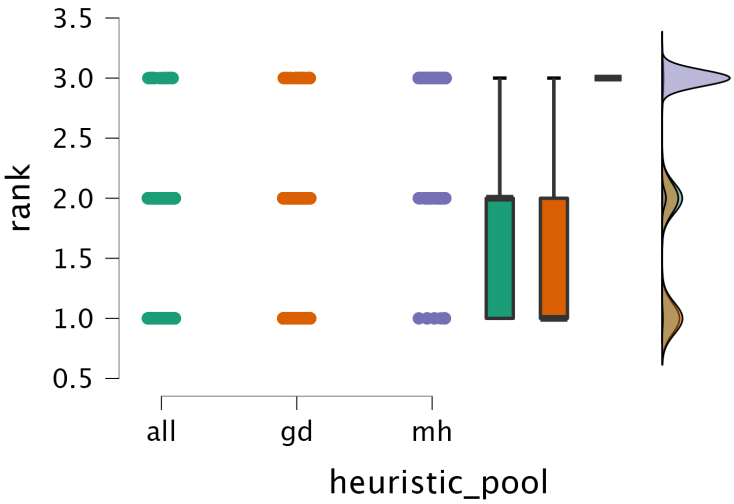
dataset: parkinsons



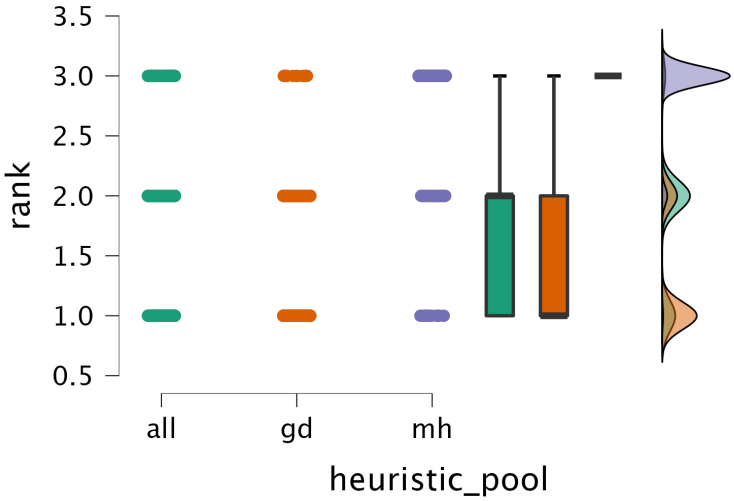
dataset: student_performance



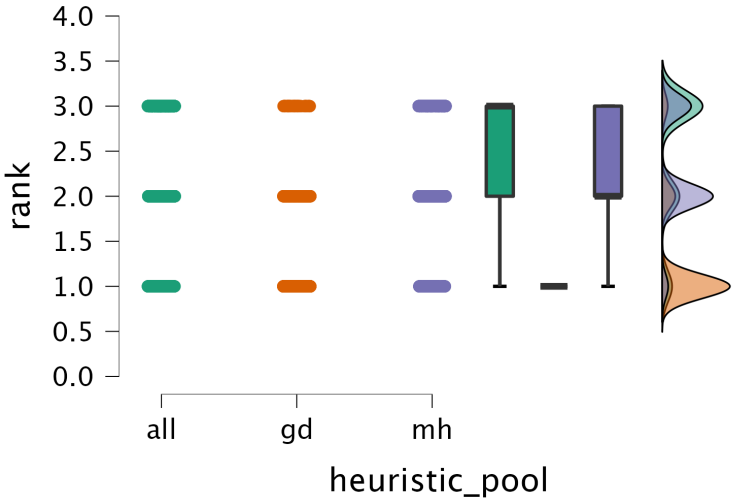
dataset: wine_quality



dataset: bank



dataset: diabetic



Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
327.821	44.000	41805.000	< .001

Contrast Tables

Simple Contrast - heuristic_pool

		95% CI for Mean Difference					
Comparison	Estimate	Lower	Upper	SE	df	t	p
gd - all	−0.346	−0.361	−0.331	0.007	41805	−46.189	< .001
mh - all	0.654	0.639	0.669	0.007	41805	87.306	< .001

Post Hoc Tests

Standard

Post Hoc Comparisons - heuristic_pool

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
all	gd	0.346	0.328	0.364	0.007	46.189	< .001***
	mh	−0.654	−0.671	−0.636	0.007	−87.306	< .001***
gd	mh	−1.000	−1.017	−0.982	0.007	−133.495	< .001***

Note. Results are averaged over the levels of: dataset
Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).
*** p < .001

Kruskal-Wallis Test

Kruskal-Wallis Test

Factor	Statistic	df	p
heuristic_pool	10774.348	2	< .001