Results

Descriptive Statistics

Descriptive Statistics

		Valid	Missing	Mean	Std. Deviation	Skewness	Std. Error of Skewness	Shapiro-Wilk	P-value of Shapiro-Wilk	Minimum	Maximum
rank	abalone	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	adult	4650	0	2.933	1.436	0.049	0.036	0.883	< .001	1.000	5.000
rank	air_quality	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	bike	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	car	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	fish_toxicity	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	forest_fires	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	housing	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	iris	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	mushroom	4650	0	2.999	1.415	−3.013e−4	0.036	0.888	< .001	1.000	5.000
rank	parkinsons	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	student_performance	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	wine_quality	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	bank	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
rank	diabetic	4650	0	3.000	1.414	0.000	0.036	0.888	< .001	1.000	5.000
test_loss	abalone	4650	0	2.283	0.476	4.099	0.036	0.619	< .001	1.915	9.874
test_loss	adult	4650	0	504.217	3561.080	17.355	0.036	0.101	< .001	0.315	96825.040
test_loss	air_quality	4650	0	0.278	0.050	5.855	0.036	0.395	< .001	0.238	0.748
test_loss	bike	4650	0	0.084	0.068	5.711	0.036	0.392	< .001	0.047	0.667
test_loss	car	4650	0	0.250	0.299	5.354	0.036	0.350	< .001	0.077	2.829
test_loss	fish_toxicity	4650	0	0.112	0.043	5.936	0.036	0.366	< .001	0.079	0.536
test_loss	forest_fires	4650	0	0.085	0.102	4.217	0.036	0.571	< .001	0.010	0.884
test_loss	housing	4650	0	0.108	0.051	5.142	0.036	0.513	< .001	0.056	0.573
test_loss	iris	4650	0	0.288	0.641	13.132	0.036	0.297	< .001	0.001	18.060
test_loss	mushroom	4650	0	1.801	83.768	63.632	0.036	0.006	< .001	0.000	5563.888
test_loss	parkinsons	4650	0	0.071	0.056	7.119	0.036	0.208	< .001	0.054	0.659
test_loss	student_performance	4650	0	0.232	0.090	2.004	0.036	0.709	< .001	0.149	0.612
test_loss	wine_quality	4650	0	1.143	0.248	5.467	0.036	0.307	< .001	0.995	2.977
test_loss	bank	4650	0	0.257	0.173	14.007	0.036	0.191	< .001	0.202	6.184
test_loss	diabetic	4650	0	1.229	1.352	16.519	0.036	0.200	< .001	0.884	43.945

ANOVA

ANOVA - rank

Cases	Sum of Squares	df	Mean Square	F	p
dataset	19.257	14	1.375	0.695	0.781
replay	109.057	4	27.264	13.784	< .001
dataset ✻ replay	1876.507	56	33.509	16.941	< .001
Residuals	137815.766	69675	1.978		

Note. Type III Sum of Squares

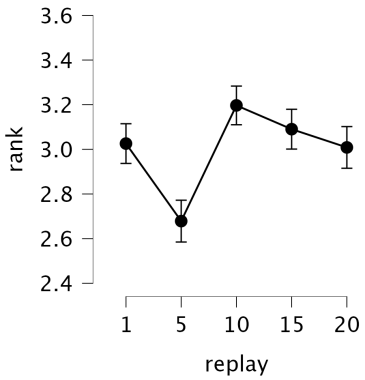
Descriptives

Descriptives - rank

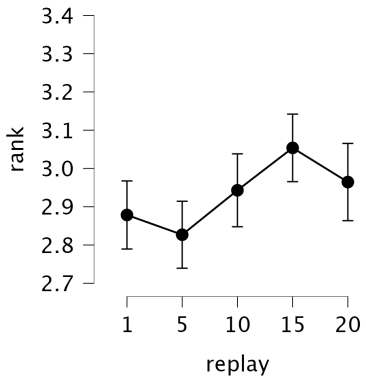
dataset	replay	Mean	SD	N
abalone	1	3.026	1.382	930
	10	3.197	1.346	930
	15	3.090	1.385	930
	20	3.009	1.450	930
	5	2.678	1.455	930
adult	1	2.878	1.382	930
	10	2.943	1.480	930
	15	3.054	1.371	930
	20	2.965	1.569	930
	5	2.827	1.360	930
air_quality	1	2.874	1.385	930
	10	2.998	1.429	930
	15	2.951	1.297	930
	20	3.162	1.443	930
	5	3.015	1.497	930
bank	1	2.948	1.460	930
	10	2.860	1.344	930
	15	3.017	1.420	930
	20	3.117	1.359	930
	5	3.057	1.473	930
bike	1	2.960	1.274	930
	10	3.145	1.448	930
	15	2.923	1.417	930
	20	3.235	1.527	930
	5	2.737	1.340	930
car	1	2.705	1.357	930
	10	3.076	1.372	930
	15	3.327	1.463	930
	20	3.054	1.409	930
	5	2.838	1.390	930
diabetic	1	3.247	1.369	930
	10	2.594	1.372	930
	15	2.808	1.384	930
	20	3.027	1.411	930
	5	3.325	1.406	930
fish_toxicity	1	2.727	1.419	930
	10	3.177	1.398	930
	15	2.989	1.386	930
	20	3.028	1.337	930
	5	3.078	1.490	930
forest_fires	1	2.805	1.493	930
	10	2.991	1.408	930
	15	3.059	1.411	930
	20	3.229	1.342	930
	5	2.915	1.381	930
housing	1	3.155	1.301	930
	10	2.855	1.450	930
	15	3.035	1.490	930
	20	2.977	1.404	930
	5	2.977	1.407	930
iris	1	3.072	1.376	930
	10	3.220	1.392	930
	15	2.618	1.396	930
	20	2.931	1.410	930
	5	3.158	1.419	930
mushroom	1	3.054	1.324	930
	10	3.042	1.414	930
	15	2.861	1.382	930
	20	2.896	1.497	930
	5	3.143	1.437	930
parkinsons	1	3.172	1.372	930
	10	2.673	1.453	930
	15	3.117	1.419	930
	20	3.195	1.498	930
	5	2.843	1.244	930
student_performance	1	3.040	1.368	930
	10	2.655	1.403	930
	15	3.094	1.323	930
	20	3.289	1.487	930
	5	2.923	1.411	930
wine_quality	1	3.174	1.430	930
	10	3.158	1.418	930
	15	3.042	1.356	930
	20	2.894	1.362	930
	5	2.732	1.456	930

Descriptives plots

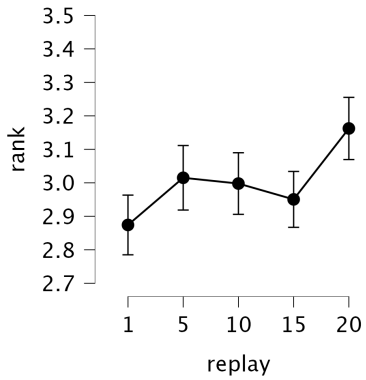
dataset: abalone



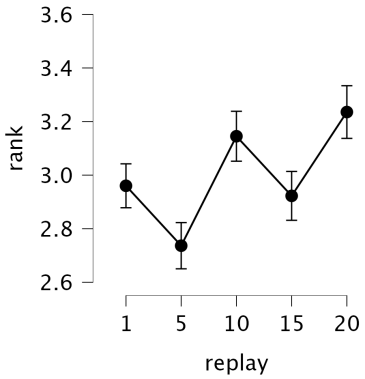
dataset: adult



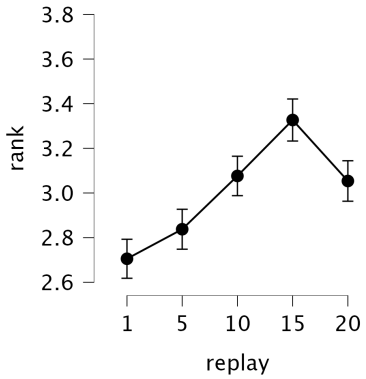
dataset: air_quality



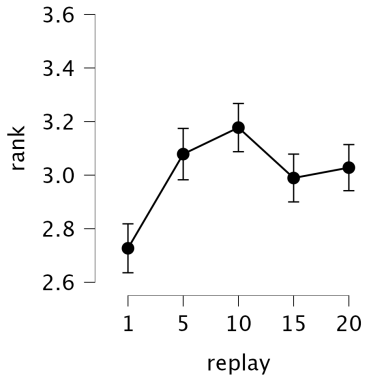
dataset: bike



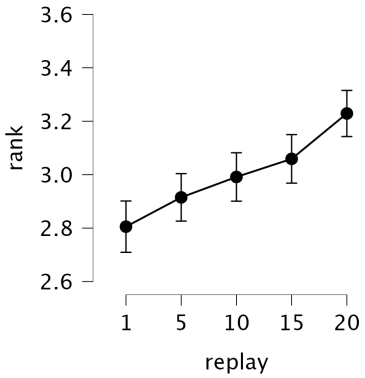
dataset: car



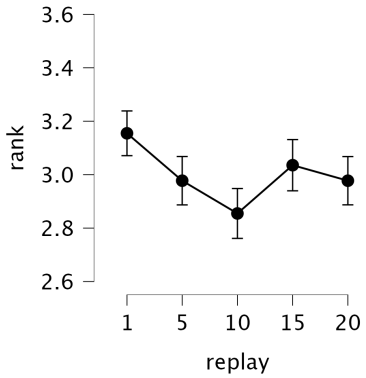
dataset: fish_toxicity



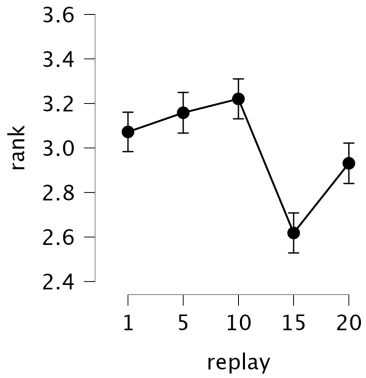
dataset: forest_fires



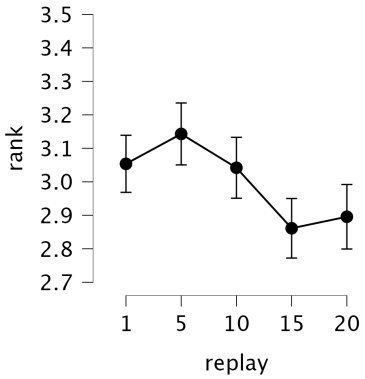
dataset: housing



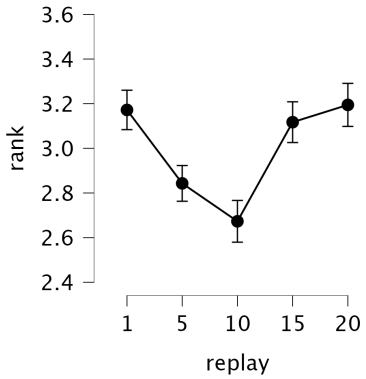
dataset: iris



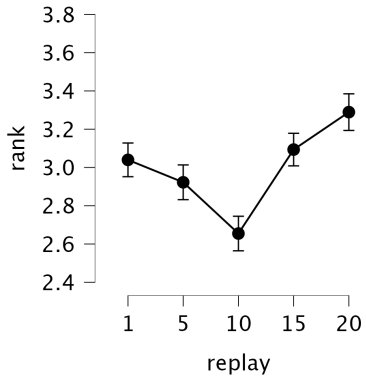
dataset: mushroom



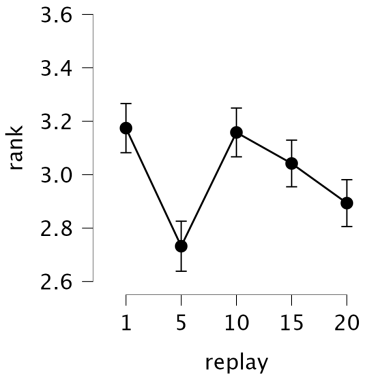
dataset: parkinsons



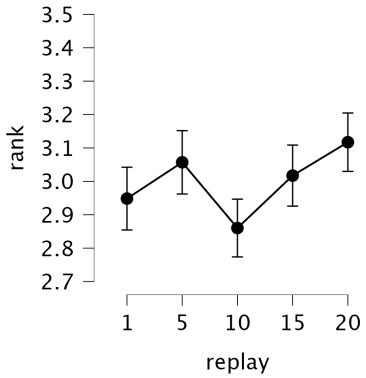
dataset: student_performance



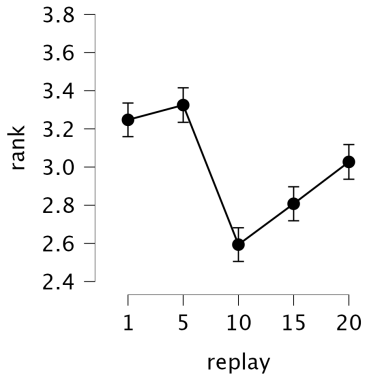
dataset: wine_quality



dataset: bank

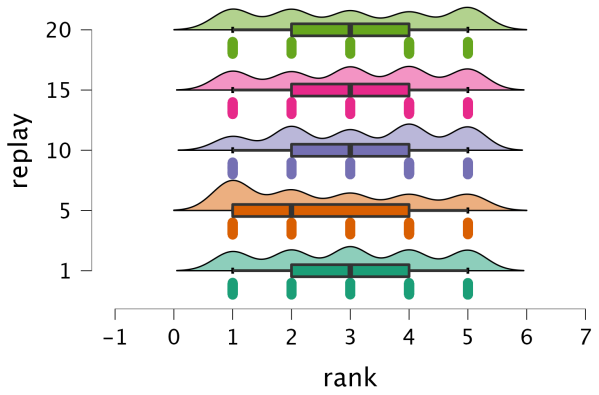


dataset: diabetic

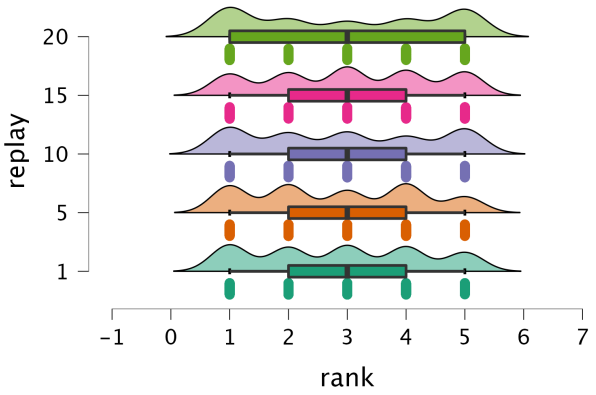


Raincloud plots

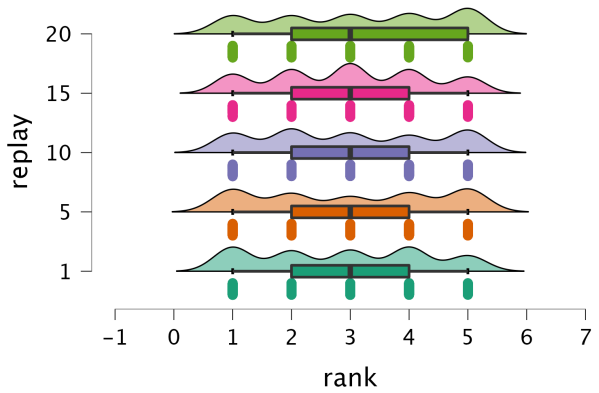
dataset: abalone



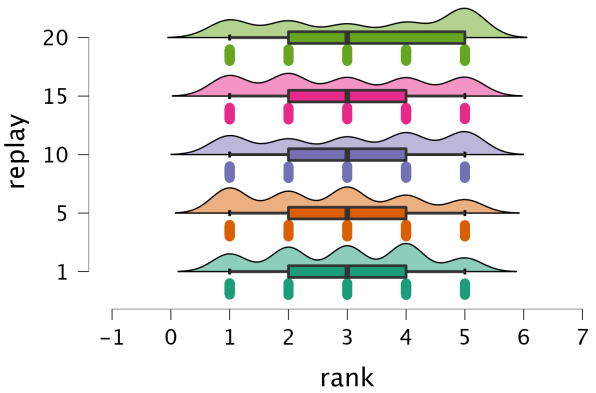
dataset: adult



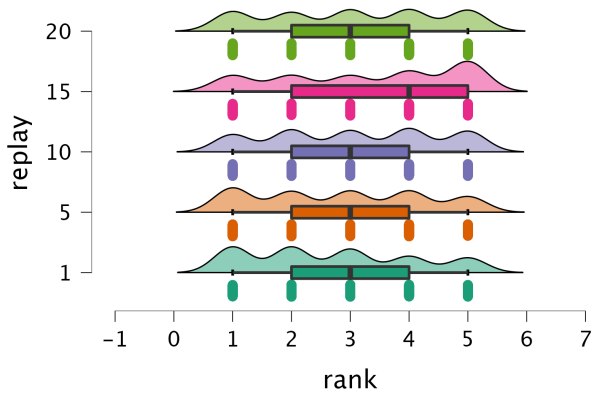
dataset: air_quality



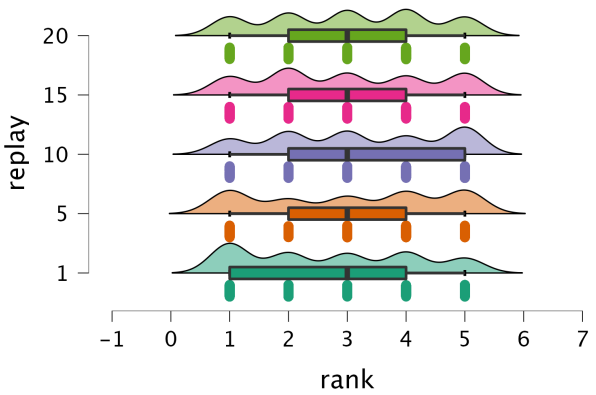
dataset: bike



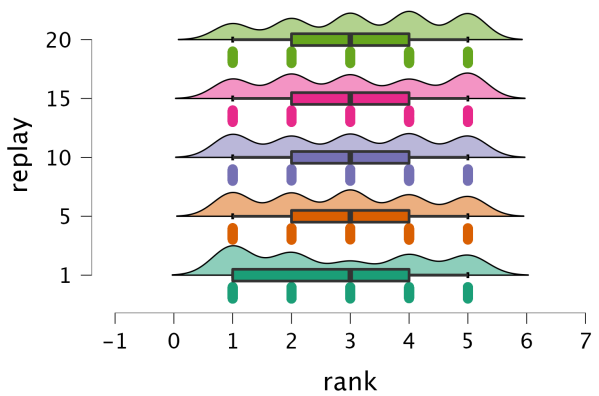
dataset: car



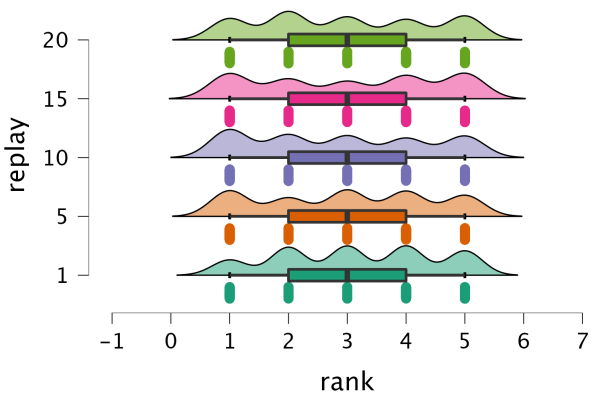
dataset: fish_toxicity



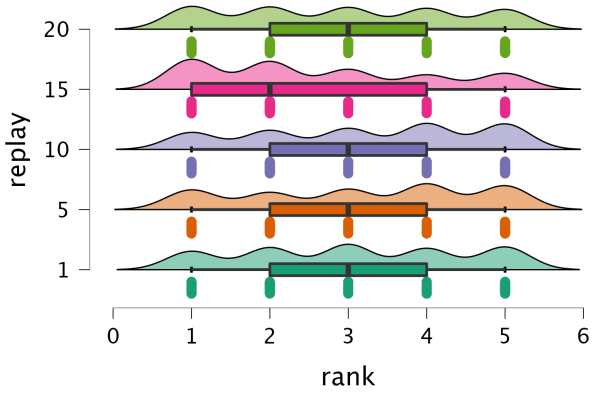
dataset: forest_fires



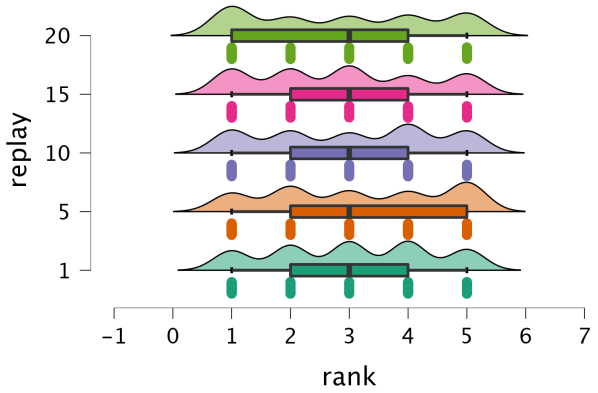
dataset: housing



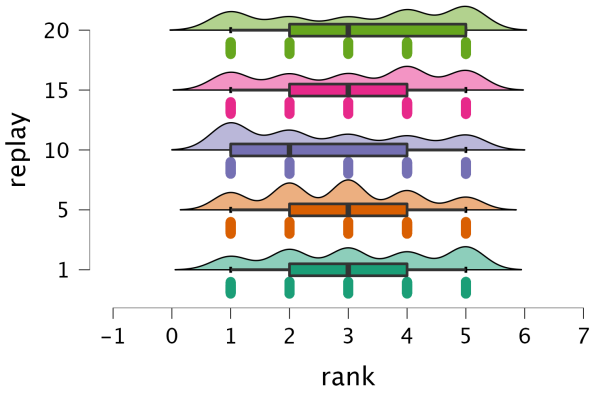
dataset: iris



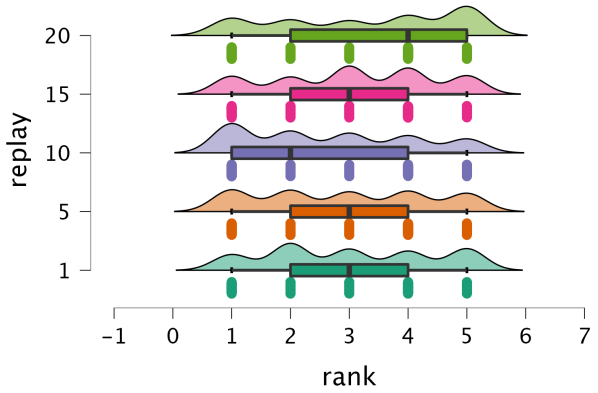
dataset: mushroom



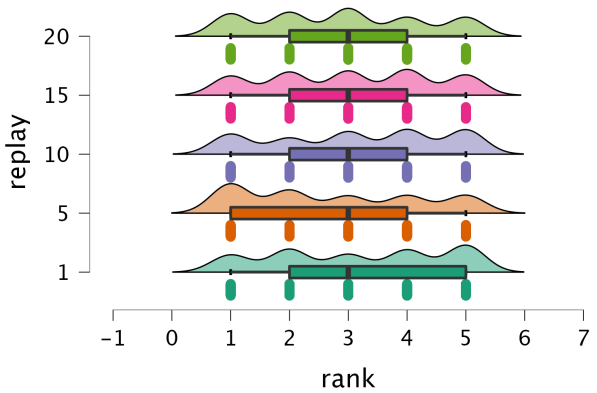
dataset: parkinsons



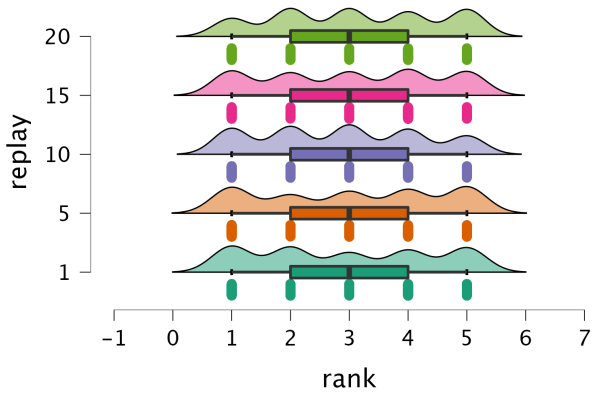
dataset: student_performance



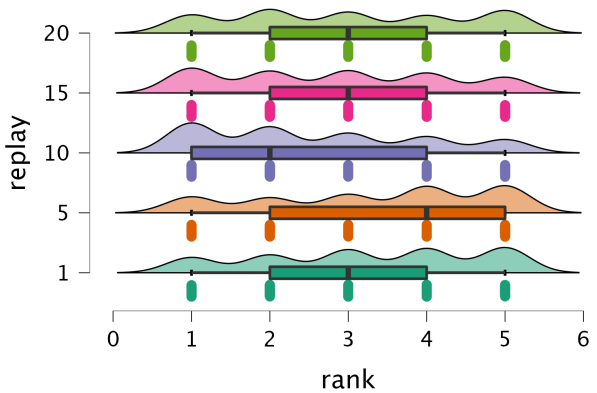
dataset: wine_quality



dataset: bank



dataset: diabetic



Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
9.255	74.000	69675.000	< .001

Contrast Tables

Simple Contrast - replay

Comparison	Estimate	SE	df	t	p
5 - 1	−0.039	0.017	69675	−2.346	0.019
10 - 1	−0.017	0.017	69675	−1.005	0.315
15 - 1	0.010	0.017	69675	0.579	0.563
20 - 1	0.078	0.017	69675	4.627	< .001

Post Hoc Tests

Standard

Post Hoc Comparisons - replay

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
1	5	0.039	−0.006	0.085	0.017	2.346	0.131
	10	0.017	−0.029	0.063	0.017	1.005	0.853
	15	−0.010	−0.056	0.036	0.017	−0.579	0.978
	20	−0.078	−0.124	−0.032	0.017	−4.627	< .001***
5	10	−0.023	−0.069	0.023	0.017	−1.341	0.666
	15	−0.049	−0.095	−0.003	0.017	−2.924	0.028*
	20	−0.117	−0.163	−0.071	0.017	−6.973	< .001***
10	15	−0.027	−0.073	0.019	0.017	−1.584	0.508
	20	−0.095	−0.141	−0.049	0.017	−5.632	< .001***
15	20	−0.068	−0.114	−0.022	0.017	−4.048	< .001***

Note. Results are averaged over the levels of: dataset
Note. P-value and confidence intervals adjusted for comparing a family of 5 estimates (confidence intervals corrected using the tukey method).
* p < .05, ** p < .01, *** p < .001

Kruskal-Wallis Test

Kruskal-Wallis Test

Factor	Statistic	df	p
replay	54.387	4	< .001