Results

Descriptive Statistics

Descriptive Statistics

		Valid	Missing	Mean	Std. Deviation	Shapiro-Wilk	P-value of Shapiro-Wilk	Minimum	Maximum
rank	abalone	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	adult	1860	0	1.483	0.500	0.636	< .001	1.000	2.000
rank	air_quality	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	bank	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	bike	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	car	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	diabetic	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	fish_toxicity	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	forest_fires	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	housing	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	iris	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	mushroom	1860	0	1.488	0.500	0.636	< .001	1.000	2.000
rank	parkinsons	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	student_performance	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
rank	wine_quality	1860	0	1.500	0.500	0.637	< .001	1.000	2.000
test_loss	abalone	1860	0	2.308	0.518	0.562	< .001	1.941	10.470
test_loss	adult	1860	0	251.524	820.521	0.286	< .001	0.316	21579.059
test_loss	air_quality	1860	0	0.280	0.050	0.391	< .001	0.244	0.744
test_loss	bank	1860	0	0.253	0.144	0.226	< .001	0.207	1.991
test_loss	bike	1860	0	0.088	0.069	0.445	< .001	0.047	0.663
test_loss	car	1860	0	0.252	0.296	0.352	< .001	0.077	2.819
test_loss	diabetic	1860	0	1.148	0.968	0.220	< .001	0.882	28.528
test_loss	fish_toxicity	1860	0	0.113	0.043	0.352	< .001	0.083	0.529
test_loss	forest_fires	1860	0	0.085	0.102	0.572	< .001	0.012	0.879
test_loss	housing	1860	0	0.108	0.052	0.518	< .001	0.062	0.583
test_loss	iris	1860	0	0.272	0.453	0.431	< .001	0.006	9.792
test_loss	mushroom	1860	0	0.218	2.214	0.065	< .001	0.000	60.909
test_loss	parkinsons	1860	0	0.071	0.056	0.204	< .001	0.054	0.659
test_loss	student_performance	1860	0	0.213	0.067	0.674	< .001	0.149	0.596
test_loss	wine_quality	1860	0	1.143	0.248	0.306	< .001	1.021	2.954

Independent Samples T-Test

Independent Samples T-Test

	W	df	p
rank	9.276e+7		< .001
test_loss	9.673e+7		0.397

Note. Mann-Whitney U test.

Assumption Checks

Test of Normality (Shapiro-Wilk)

		W	p
rank	False	NaN[a]	
	True	NaN[a]	
test_loss	False	NaN[b]	
	True	NaN[b]	

Note. Significant results suggest a deviation from normality.
[a] Number of observations is < 3 or > 5000 in rank after grouping on discounted_rewards
[b] Number of observations is < 3 or > 5000 in test_loss after grouping on discounted_rewards

Test of Equality of Variances (Levene's)

	F	df	p
rank	1.726	1	0.189
test_loss	0.679	1	0.410

Descriptives

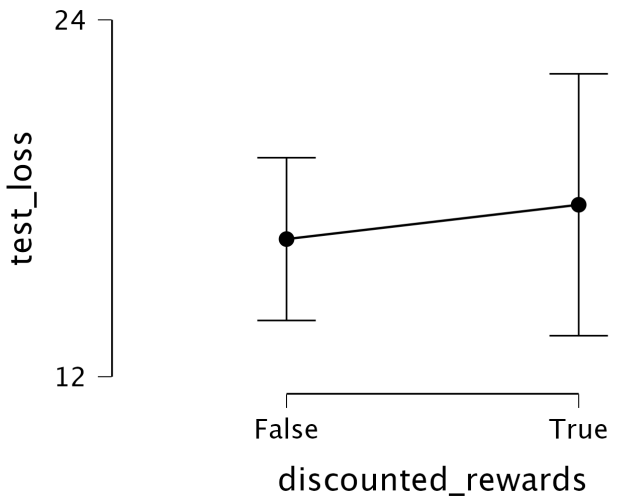
Group Descriptives

	Group	N	Mean	SD	SE
rank	False	13950	1.475	0.499	0.004
	True	13950	1.521	0.500	0.004
test_loss	False	13950	16.628	164.884	1.396
	True	13950	17.782	265.299	2.246

Descriptives Plots

rank

test_loss



Raincloud Plots

rank

test_loss

ANOVA

ANOVA - rank

Cases	Sum of Squares	df	Mean Square	F	p
dataset	0.727	14	0.052	0.210	0.999
discounted_rewards	15.190	1	15.190	61.607	< .001
dataset ✻ discounted_rewards	87.282	14	6.234	25.285	< .001
Residuals	6871.694	27870	0.247		

Note. Type III Sum of Squares

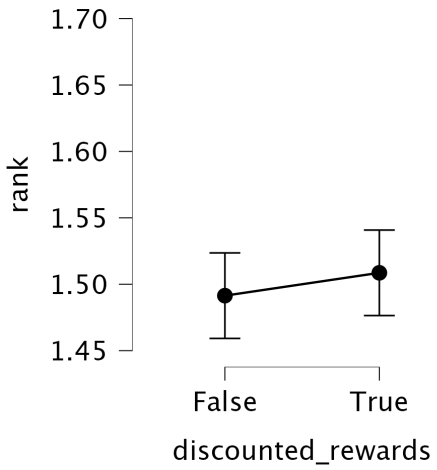
Descriptives

Descriptives - rank

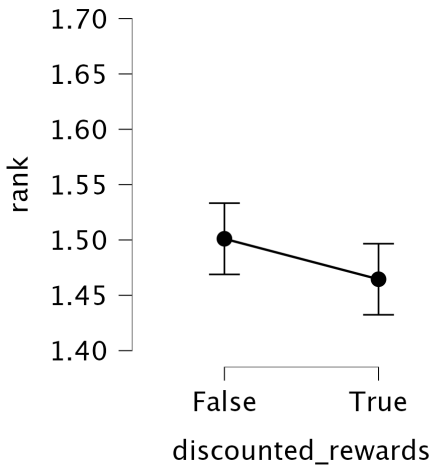
dataset	discounted_rewards	Mean	SD	N
abalone	False	1.491	0.500	930
	True	1.509	0.500	930
adult	False	1.501	0.500	930
	True	1.465	0.499	930
air_quality	False	1.461	0.499	930
	True	1.539	0.499	930
bank	False	1.516	0.500	930
	True	1.484	0.500	930
bike	False	1.501	0.500	930
	True	1.499	0.500	930
car	False	1.428	0.495	930
	True	1.572	0.495	930
diabetic	False	1.470	0.499	930
	True	1.530	0.499	930
fish_toxicity	False	1.501	0.500	930
	True	1.499	0.500	930
forest_fires	False	1.474	0.500	930
	True	1.526	0.500	930
housing	False	1.424	0.494	930
	True	1.576	0.494	930
iris	False	1.590	0.492	930
	True	1.410	0.492	930
mushroom	False	1.395	0.489	930
	True	1.581	0.494	930
parkinsons	False	1.371	0.483	930
	True	1.629	0.483	930
student_performance	False	1.440	0.497	930
	True	1.560	0.497	930
wine_quality	False	1.557	0.497	930
	True	1.443	0.497	930

Descriptives plots

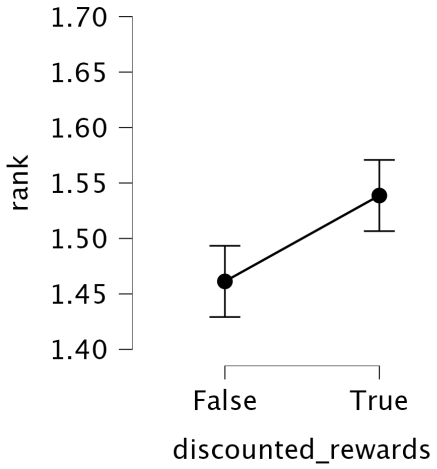
dataset: abalone



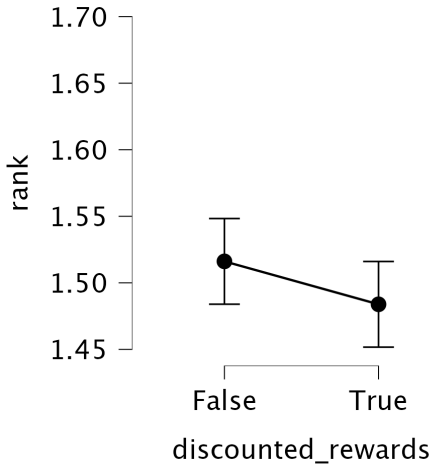
dataset: adult



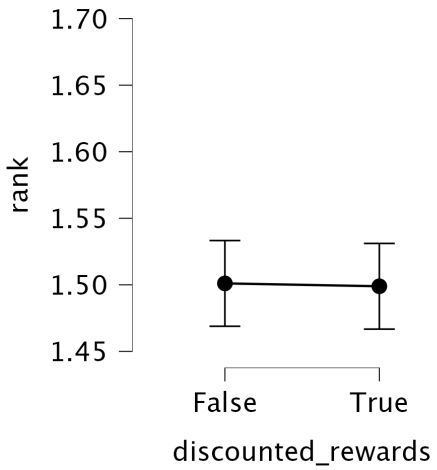
dataset: air_quality



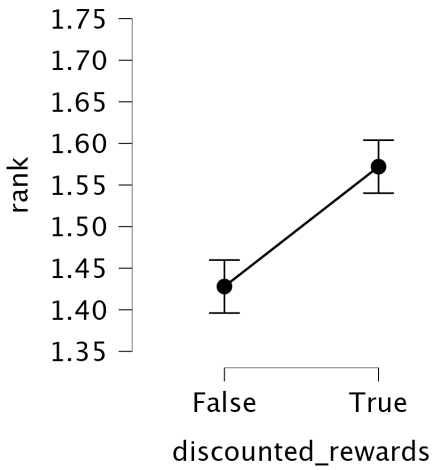
dataset: bank



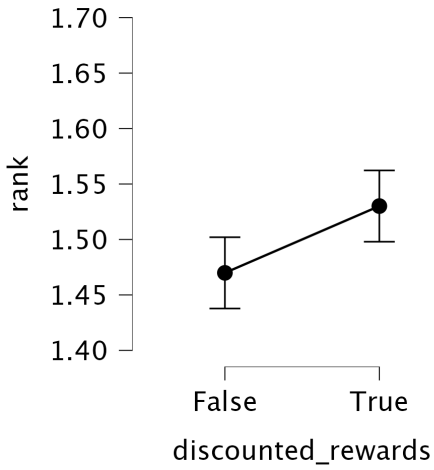
dataset: bike



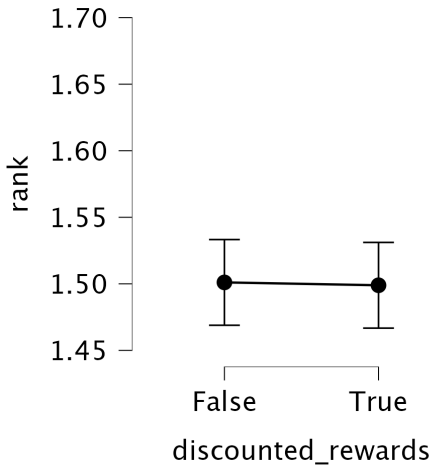
dataset: car



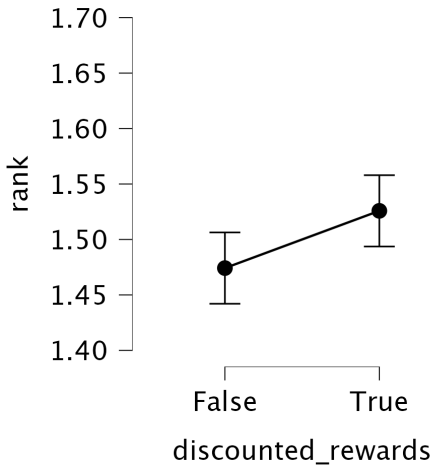
dataset: diabetic



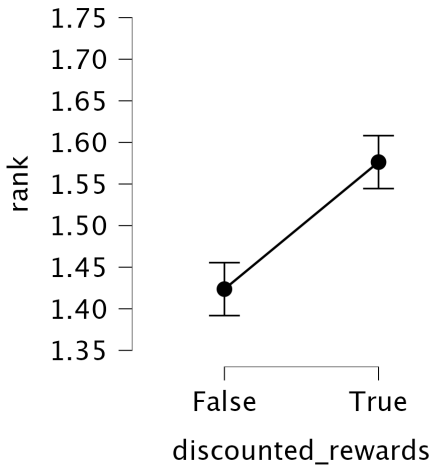
dataset: fish_toxicity



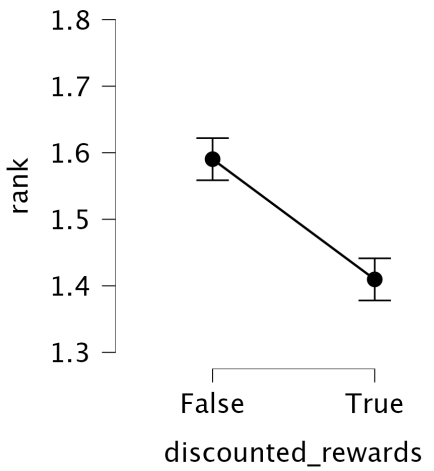
dataset: forest_fires



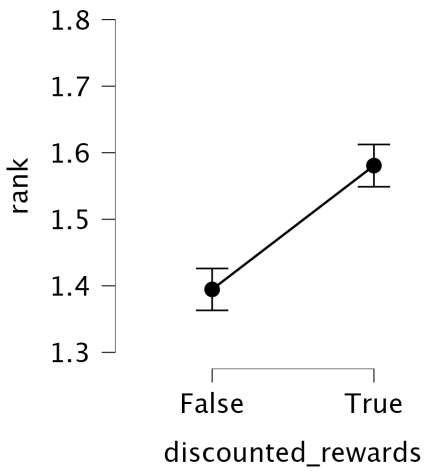
dataset: housing



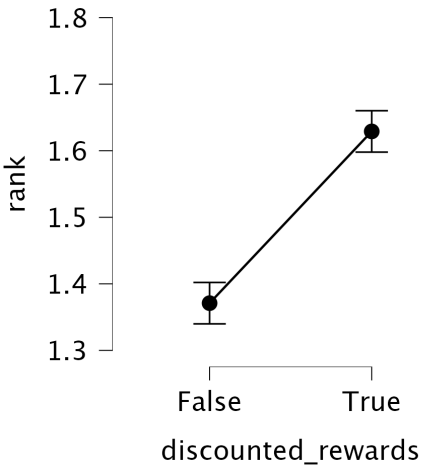
dataset: iris



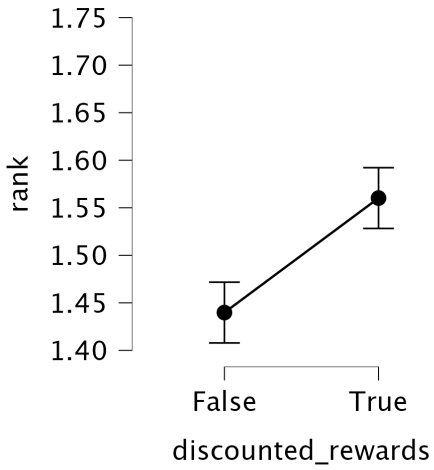
dataset: mushroom



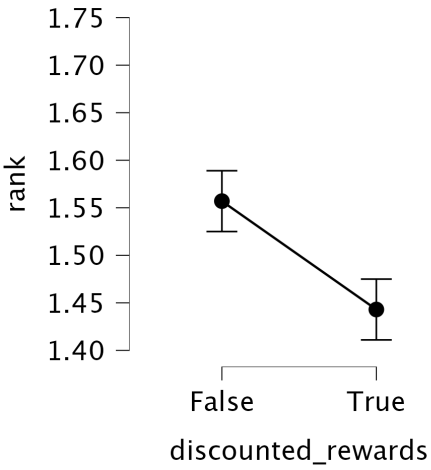
dataset: parkinsons



dataset: student_performance

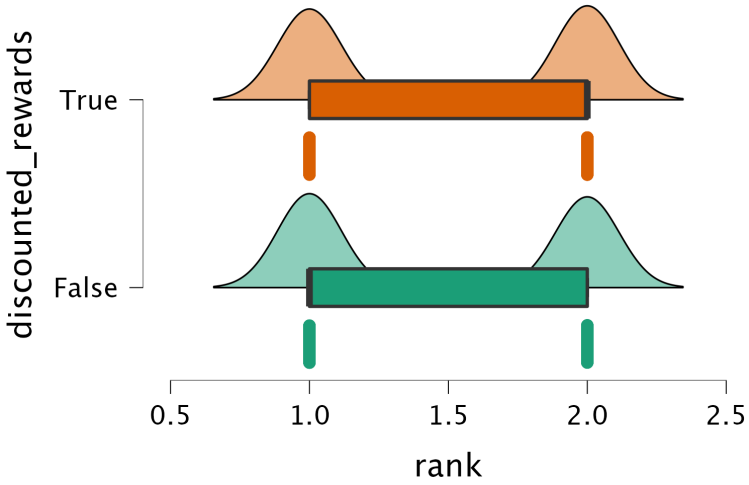


dataset: wine_quality

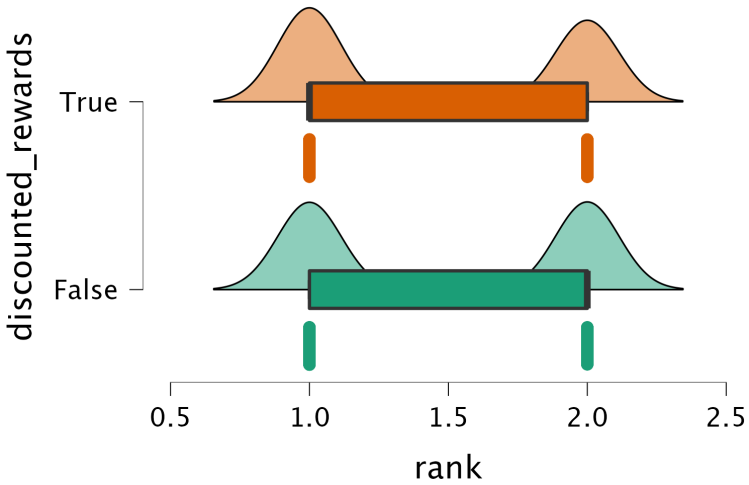


Raincloud plots

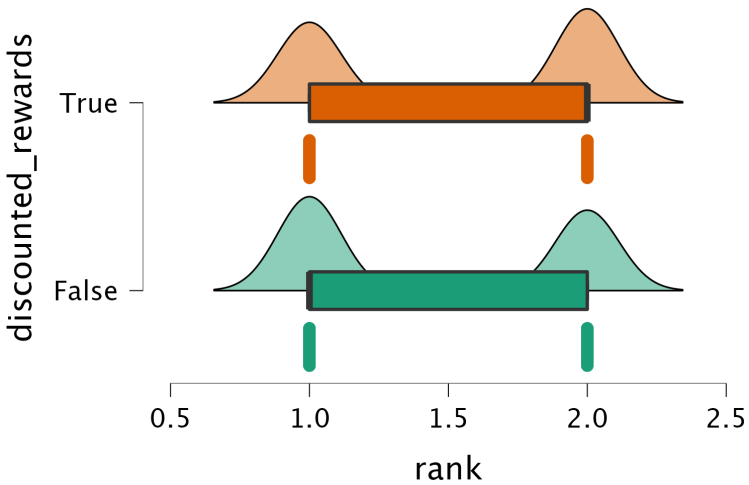
dataset: abalone



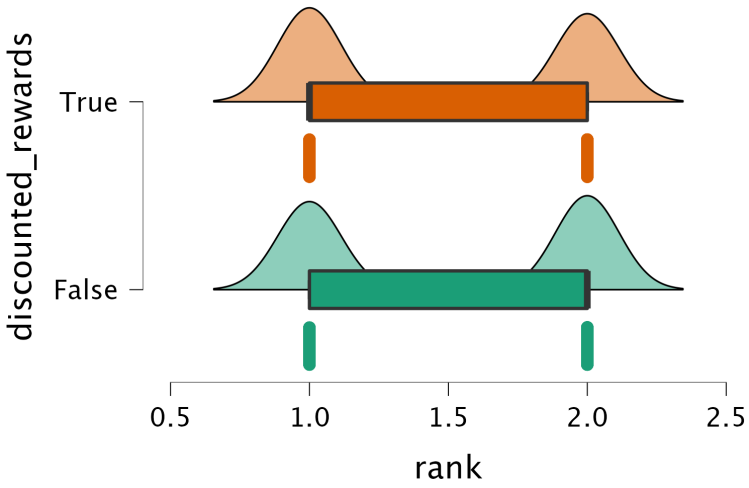
dataset: adult



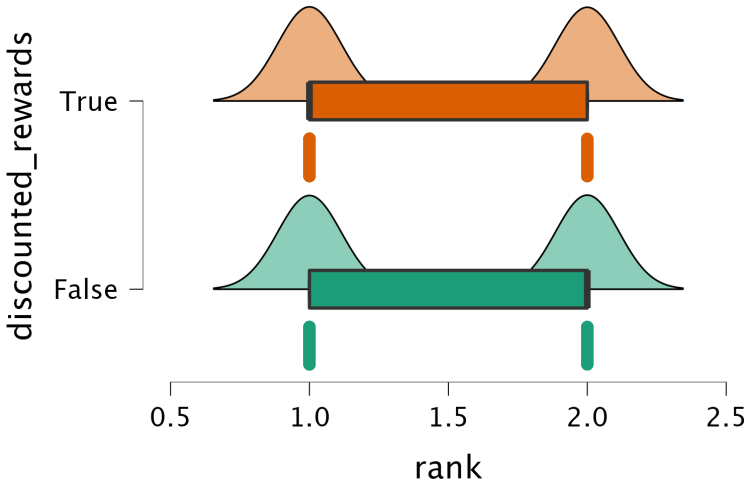
dataset: air_quality



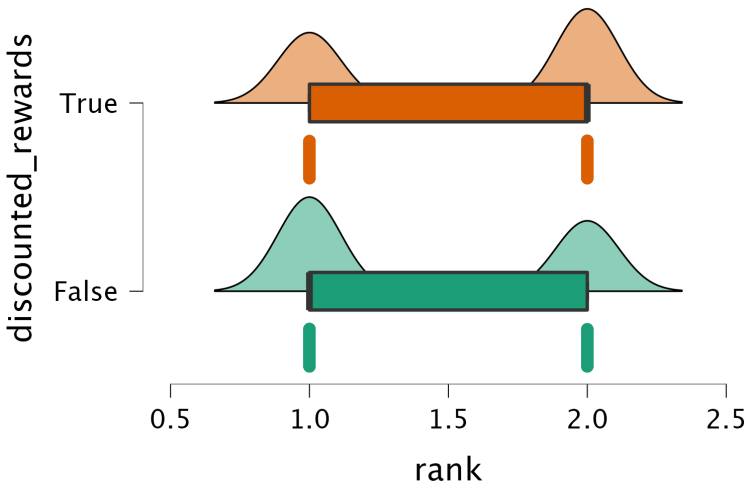
dataset: bank



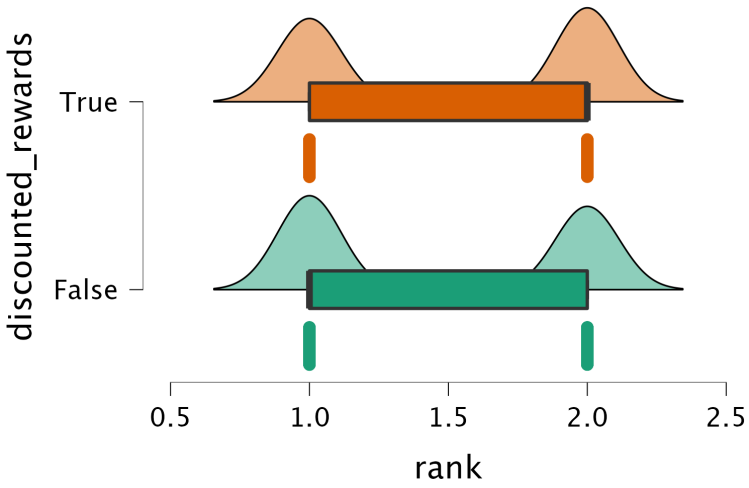
dataset: bike



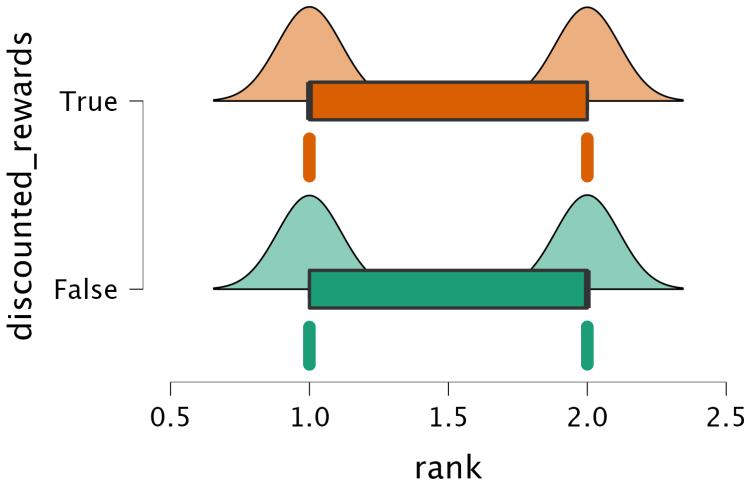
dataset: car



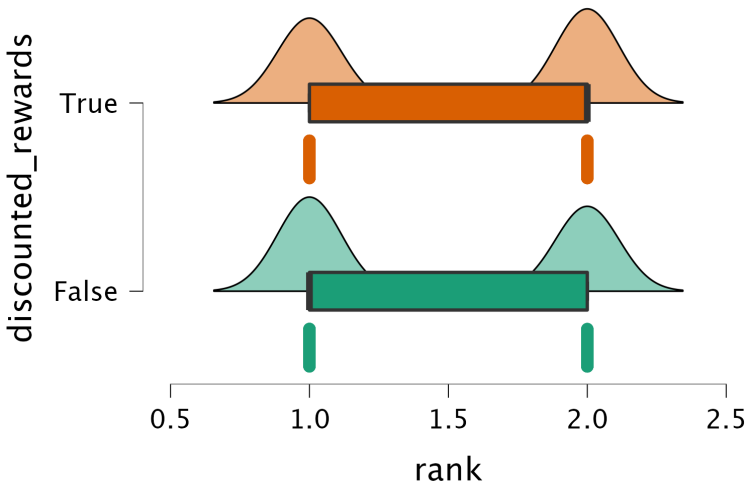
dataset: diabetic



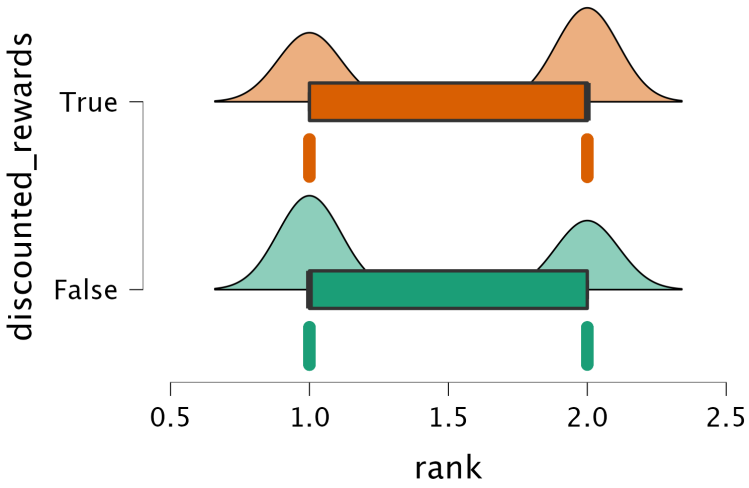
dataset: fish_toxicity



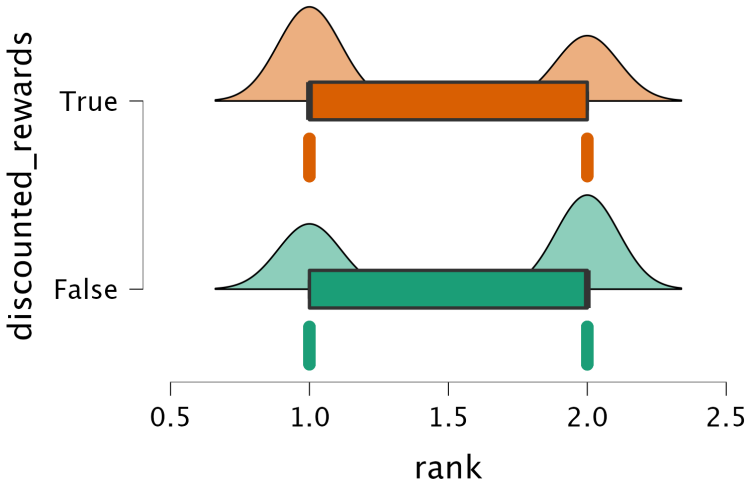
dataset: forest_fires



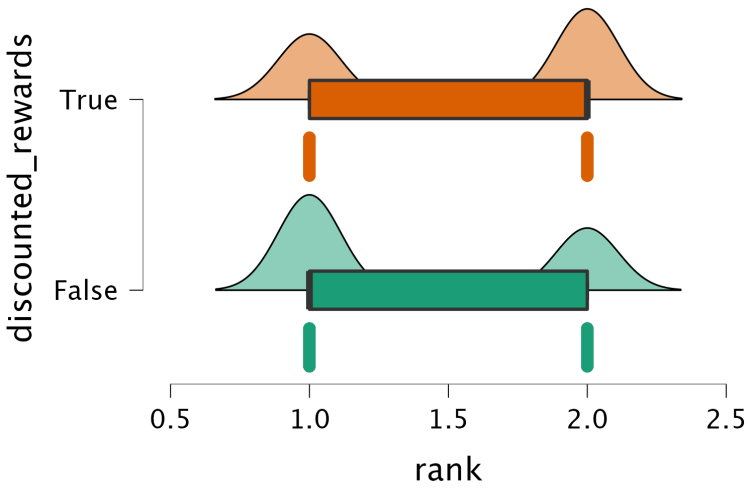
dataset: housing



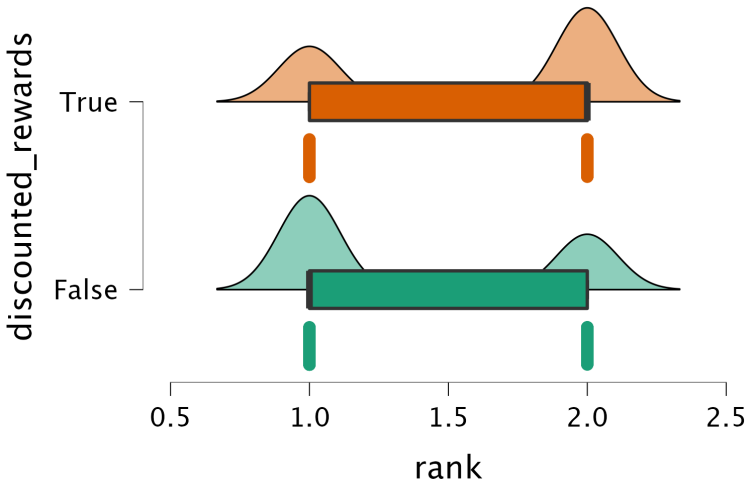
dataset: iris



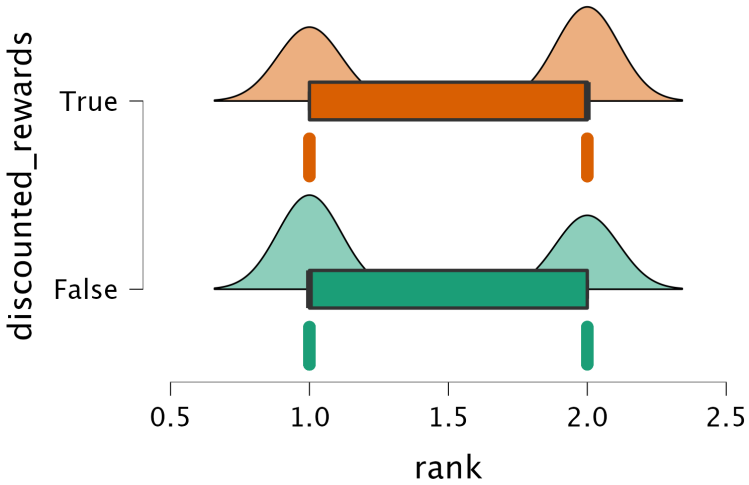
dataset: mushroom



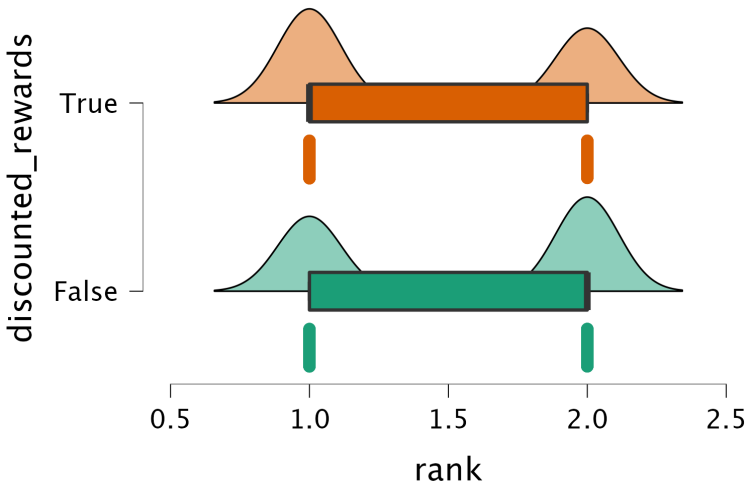
dataset: parkinsons



dataset: student_performance



dataset: wine_quality



Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
22.275	29.000	27870.000	< .001

Contrast Tables

Simple Contrast - discounted_rewards

Comparison	Estimate	SE	df	t	p
True - False	0.047	0.006	27870	7.849	< .001

Post Hoc Tests

Standard

Post Hoc Comparisons - discounted_rewards

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
False	True	−0.047	−0.058	−0.035	0.006	−7.849	< .001***

Note. Results are averaged over the levels of: dataset
*** p < .001

Kruskal-Wallis Test

Kruskal-Wallis Test

Factor	Statistic	df	p
discounted_rewards	60.759	1	< .001