Validation Study Details

This slide set provides a breakdown of the exact details of the validation study results.
The validation study was performed using ARTools by Wobbrock et al. [1] in R with the following measures:

- Repeated-measures ANOVA
- Post-hoc pairwise Bonferroni correction

For Task 1 and Task 2, the results are shown for Composition (AccuStripes vs Line Charts).
For Task 3, the results are shown for Composition, and Flaw x Composition.

The results include the evaluation of the dependent variables accuracy, time, and confidence.
For Task 1 and Task 2 accuracy is measured as error. Error is given through the continuous Earth Mover Distance (EMD) score.
In Task 1 the correct response had a value EMD = 0. In Task 2 the correct response had the smallest EMD value of all responses. Thus, the higher the EMD value, the greater the error.
For Task 3 accuracy is measured as correctness. Correctness is given by the number of correct responses in percent in the interval [0, 1].

For each dependent variable we provide following information:

- The results of the repeated-measures ANOVA
- The results of the post-hoc pairwise Bonferroni correction
- The statistical values as mean, standard deviation (sd), standard error (se), confidence interval (ci), interquartile range (iqr) etc.

For easier interpretation we added visual assistance:

1) Results which are significant are marked (see Table 1).
2) We framed the important outcomes in blue.
   We do not just provide the framed results because we want the reader to have a complete picture of the data and be able to make comparisons.
3) We highlighted the best performing technique in green and the worst performing technique in red.

Significance	p – value
***	[0, 0.001]
**	(0.001, 0.01]
*	(0.01, 0.05]
.	(0.05, 0.1]
	(0.1, 1]

Table 1: Significance Codes

T1 – Identification Task

1. Composition – Error

1. ANOVA on **ERROR**

	Df	Df.res	F value	Pr(>F)	
1 compositionFactor	1	60	6.2746	0.014982	*

2. Post Hoc on **ERROR**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	6.741935	2.691478	60	2.50492	0.01498152	*

3. Analysis over all conditions - **ERROR**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
1	LineChart	Error	31	0	0.222	0	0	0.049	0.049	0	0.045	0.084	0.015	0.031
2	AccuStripes	Error	31	0	0.311	0	0	0.000	0.000	0	0.010	0.056	0.010	0.020

1. Composition – Time

1. ANOVA on **TIME**

	Df	Df.res	F value	Pr(>F)
1 compositionFactor	1	60	0.60659	0.43914

2. Post Hoc on **TIME**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	3.580645	4.597425	60	0.7788371	0.4391366	

3. Analysis over all conditions - **TIME**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
1	AccuStripes	Time	31	3.968	31.618	7.683	5.131	9.704	4.572	3.695	8.744	5.614	1.008	2.059
2	LineChart	Time	31	3.625	74.572	7.746	5.296	14.567	9.271	5.085	12.291	13.238	2.378	4.856

1. Composition – Confidence

1. ANOVA on **CONFIDENCE**

	Df	Df.res	F value	Pr(>F)	
1 compositionFactor	1	60	3.9321	0.051956	.

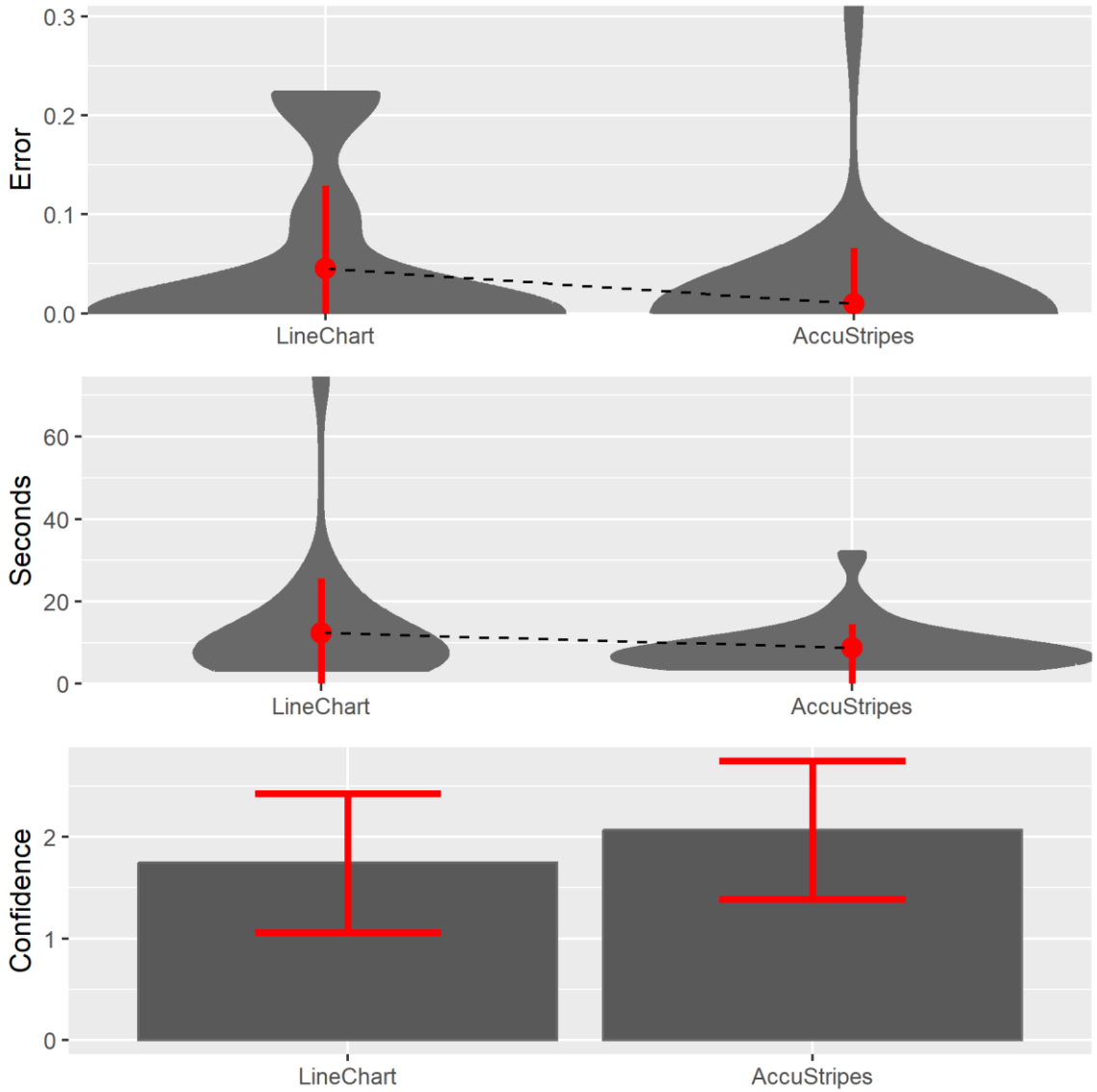
2. Post Hoc on **CONFIDENCE**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	-7.451613	3.757836	60	-1.982953	0.05195594	.

3. Analysis over all conditions - **CONFIDENCE**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
1	AccuStripes	Confidence	31	0	3	2	2.0	2	0.0	0	2.065	0.680	0.122	0.249
2	LineChart	Confidence	31	0	3	2	1.5	2	0.5	0	1.742	0.682	0.122	0.250

1. Composition – Summary



T2 – Comparison Task

1. Composition – Error

1. ANOVA on **ERROR**

		Df	Df.res	F value	Pr(>F)	
1	compositionFactor	1	60	32.697	3.613e-07	***

2. Post Hoc on **ERROR**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	20.93548	3.661252	60	5.718121	3.613023e-07	***

3. Analysis over all conditions - **ERROR**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
1	LineChart	Error	31	0.067	0.235	0.230	0.067	0.235	0.168	0.008	0.175	0.075	0.013	0.027
2	AccuStripes	Error	31	0.050	0.215	0.068	0.050	0.068	0.019	0.028	0.074	0.048	0.009	0.018

1. Composition – Time

1. ANOVA on **TIME**

		Df	Df.res	F value	Pr(>F)
1	compositionFactor	1	60	0.84824	0.36074

2. Post Hoc on **TIME**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	4.225806	4.588287	60	0.9209988	0.3607395	

3. Analysis over all conditions - **TIME**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
1	AccuStripes	Time	31	4.094	43.450	13.080	9.770	16.394	6.624	5.005	15.065	8.636	1.551	3.168
2	LineChart	Time	31	4.924	55.452	14.451	9.848	21.850	12.003	6.937	17.396	10.544	1.894	3.867

1. Composition – Confidence

1. ANOVA on **CONFIDENCE**

	Df	Df.res	F value	Pr(>F)
1 compositionFactor	1	60	0.18833	0.66587

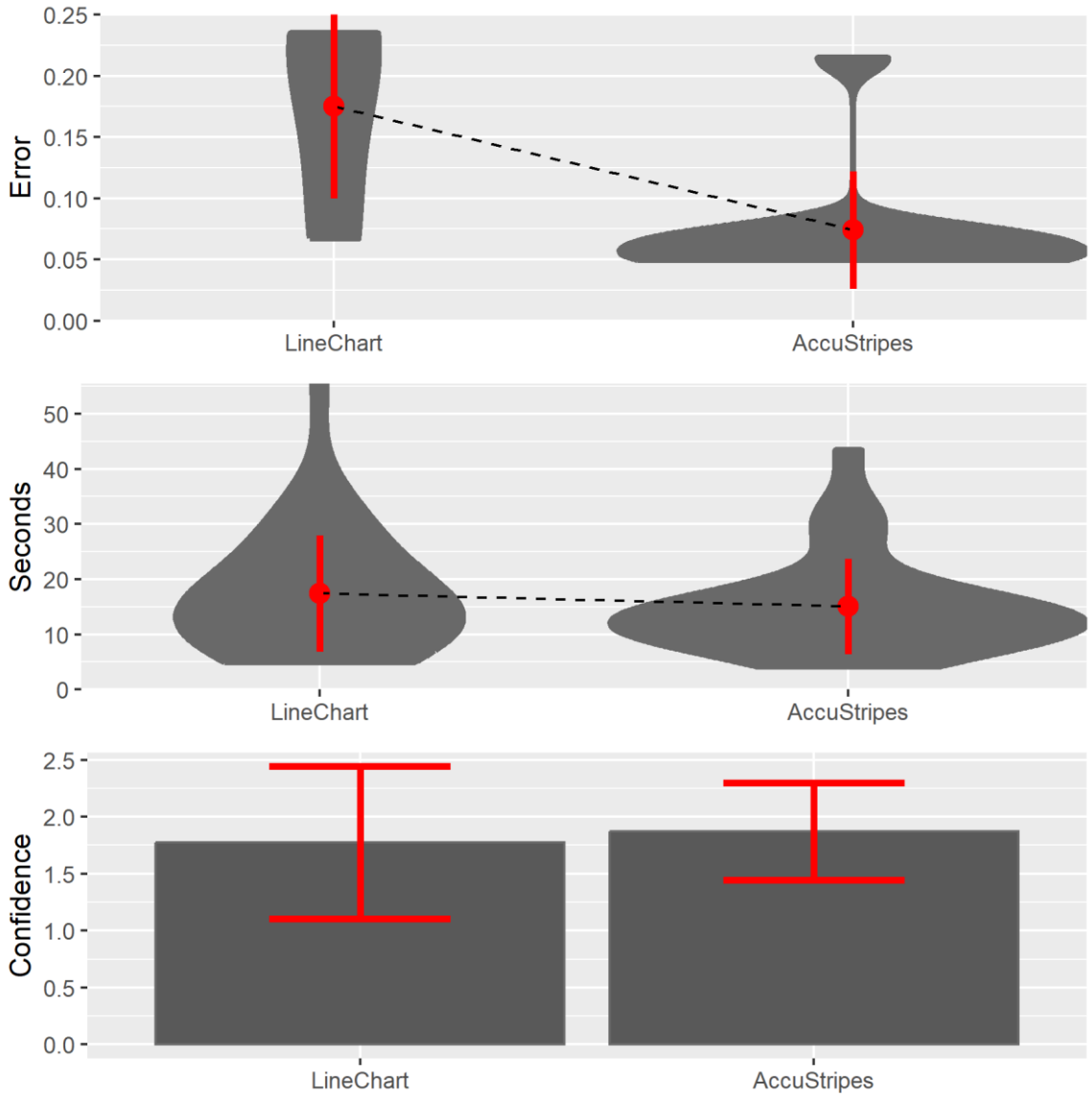
2. Post Hoc on **CONFIDENCE**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	-1.225806	2.824663	60	-0.4339656	0.6658699	

3. Analysis over all conditions - **CONFIDENCE**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci
1	AccuStripes	Confidence	31	0	2	2	2	2	0	0	1.871	0.428	0.077	0.157
2	LineChart	Confidence	31	0	3	2	2	2	0	0	1.774	0.669	0.120	0.245

1. Composition – Summary



T3 – Flaw Detection Task

1. Composition – Correctness

1. ANOVA on **CORRECTNESS**

		Error	Df	Df.res	F value	Pr(>F)	
1	compositionFactor	Withn	1	90	85.015	1.1968e-14	***
2	dataflawFactor	usF:F	2	60	123.291	< 2.22e-16	***
3	compositionFactor:dataflawFactor	Withn	2	90	25.272	1.9466e-09	***

2. Post Hoc on **CORRECTNESS**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	-48.46237	5.256006	90	-9.220379	1.196794e-14	***

3. Analysis over all conditions - **CORRECTNESS**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci	ci_min	ci_max
1	AccuStripes	Accuracy	93	0	1	1	1	1	0	0	0.753	0.434	0.045	0.089	0.664	0.842
2	LineChart	Accuracy	93	0	1	1	0	1	1	0	0.613	0.490	0.051	0.101	0.512	0.714

1. Composition – Time

1. ANOVA on **TIME**

		Error	Df	Df.res	F value	Pr(>F)	
1	compositionFactor	Withn	1	90	0.82357	0.366561	
2	dataflawFactor	usF:F	2	60	2.89452	0.063086	.
3	compositionFactor:dataflawFactor	Withn	2	90	2.34859	0.101332	

2. Post Hoc on **TIME**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	5.580645	6.149414	90	0.9075085	0.3665615	

3. Analysis over all conditions - **TIME**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci	ci_min	ci_max
1	AccuStripes	Time	93	3.532	37.270	8.838	6.675	12.307	5.632	3.976	11.070	6.723	0.697	1.385	9.685	12.455
2	LineChart	Time	93	4.864	115.263	9.706	7.309	13.464	6.155	3.924	12.815	13.465	1.396	2.773	10.042	15.588

1. Composition – Confidence

1. ANOVA on **CONFIDENCE**

		Error	Df	Df.res	F value	Pr(>F)	
1	compositionFactor	Withn	1	90	6.0450	0.0158597	*
2	dataflawFactor	usF:F	2	60	3.3972	0.0400257	*
3	compositionFactor:dataflawFactor	Withn	2	90	7.1241	0.0013422	**

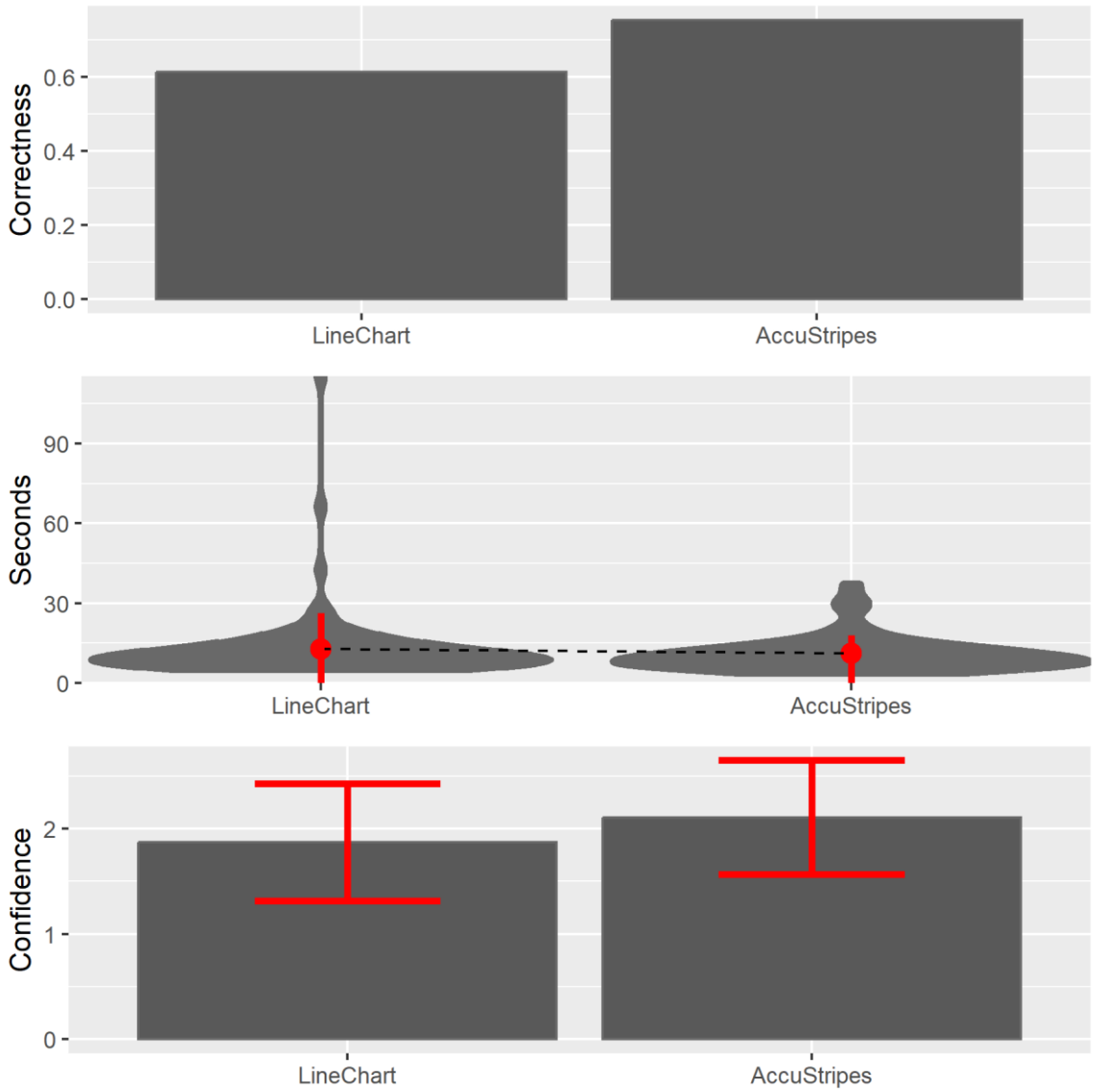
2. Post Hoc on **CONFIDENCE**

	contrast	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	-13.53763	5.506127	90	-2.458649	0.01585967	*

3. Analysis over all conditions - **CONFIDENCE**

	compositionFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci	ci_min	ci_max
1	AccuStripes	Confidence	93	0	3	2	2	2	0	0	2.108	0.541	0.056	0.111	1.997	2.219
2	LineChart	Confidence	93	0	3	2	2	2	0	0	1.871	0.556	0.058	0.115	1.756	1.986

1. Composition – Summary



2. Flaw x Composition – Correctness

1. ANOVA on **CORRECTNESS**

		Error	Df	Df.res	F value	Pr(>F)	
1	compositionFactor	Withn	1	90	85.015	1.1968e-14	***
2	dataflawFactor	usF:F	2	60	123.291	< 2.22e-16	***
3	compositionFactor:dataflawFactor	Withn	2	90	25.272	1.9466e-09	***

2. Post Hoc on **CORRECTNESS**

	compositionFactor_pairwise	dataflawFactor_pairwise	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	Gap - Outlier	83.838710	13.30878	90	6.2995047	3.227038e-08	***
2	LineChart - AccuStripes	Gap - Spike	3.935484	13.30878	90	0.2957059	1.000000e+00	
3	LineChart - AccuStripes	Outlier - Spike	-79.903226	13.30878	90	-6.0037988	1.198523e-07	***

3. Analysis over all conditions - **CORRECTNESS**

	compositionFactor	dataflawFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci	ci_min	ci_max
	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	LineChart	Gap	Accuracy	31	0	1	1	1	1	0	0	0.903	0.301	0.054	0.11	0.793	1.01
2	AccuStripes	Gap	Accuracy	31	0	1	1	1	1	0	0	0.903	0.301	0.054	0.11	0.793	1.01
3	AccuStripes	Outlier	Accuracy	31	0	1	0	0	1	1	0	0.387	0.495	0.089	0.182	0.205	0.569
4	LineChart	Outlier	Accuracy	31	0	1	0	0	0	0	0	0.032	0.18	0.032	0.066	-0.034	0.098
5	AccuStripes	Spike	Accuracy	31	0	1	1	1	1	0	0	0.968	0.18	0.032	0.066	0.902	1.03
6	LineChart	Spike	Accuracy	31	0	1	1	1	1	0	0	0.903	0.301	0.054	0.11	0.793	1.01

2. Flaw x Composition – Time

1. ANOVA on **TIME**

		Error	Df	Df.res	F value	Pr(>F)	
1	compositionFactor	Withn	1	90	0.82357	0.366561	
2	dataflawFactor	usF:F	2	60	2.89452	0.063086	.
3	compositionFactor:dataflawFactor	Withn	2	90	2.34859	0.101332	

2. Post Hoc on **TIME**

	compositionFactor_pairwise	dataflawFactor_pairwise	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	Gap - Outlier	-17.87097	14.69879	90	-1.2158121	0.68170786	
2	LineChart - AccuStripes	Gap - Spike	-31.77419	14.69879	90	-2.1616876	0.09988807	.
3	LineChart - AccuStripes	Outlier - Spike	-13.90323	14.69879	90	-0.9458755	1.00000000	

3. Analysis over all conditions - **TIME**

	compositionFactor	dataflawFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci	ci_min	ci_max
	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	LineChart	Gap	Time	31	4.86	18.1	9.31	6.82	10.5	3.72	3.27	9.53	3.68	0.66	1.35	8.18	10.9
2	AccuStripes	Gap	Time	31	3.53	37.3	8.62	6.39	15.0	8.58	5.00	13.1	9.59	1.72	3.52	9.54	16.6
3	AccuStripes	Outlier	Time	31	5.85	29.6	10.7	8.20	12.3	4.10	3.14	11.3	4.87	0.875	1.79	9.55	13.1
4	LineChart	Outlier	Time	31	6.18	66.3	10.7	8.75	14.8	6.09	4.38	14.7	12.2	2.19	4.48	10.2	19.1
5	AccuStripes	Spike	Time	31	4.50	18.8	7.73	6.24	10.5	4.22	2.86	8.82	3.68	0.661	1.35	7.47	10.2
6	LineChart	Spike	Time	31	5.22	115.	9.63	7.21	13.1	5.92	3.82	14.3	19.4	3.49	7.12	7.14	21.4

2. Flaw x Composition – Confidence

1. ANOVA on **CONFIDENCE**

		Error	Df	Df.res	F value	Pr(>F)	
1	compositionFactor	Withn	1	90	6.0450	0.0158597	*
2	dataflawFactor	usF:F	2	60	3.3972	0.0400257	*
3	compositionFactor:dataflawFactor	Withn	2	90	7.1241	0.0013422	**

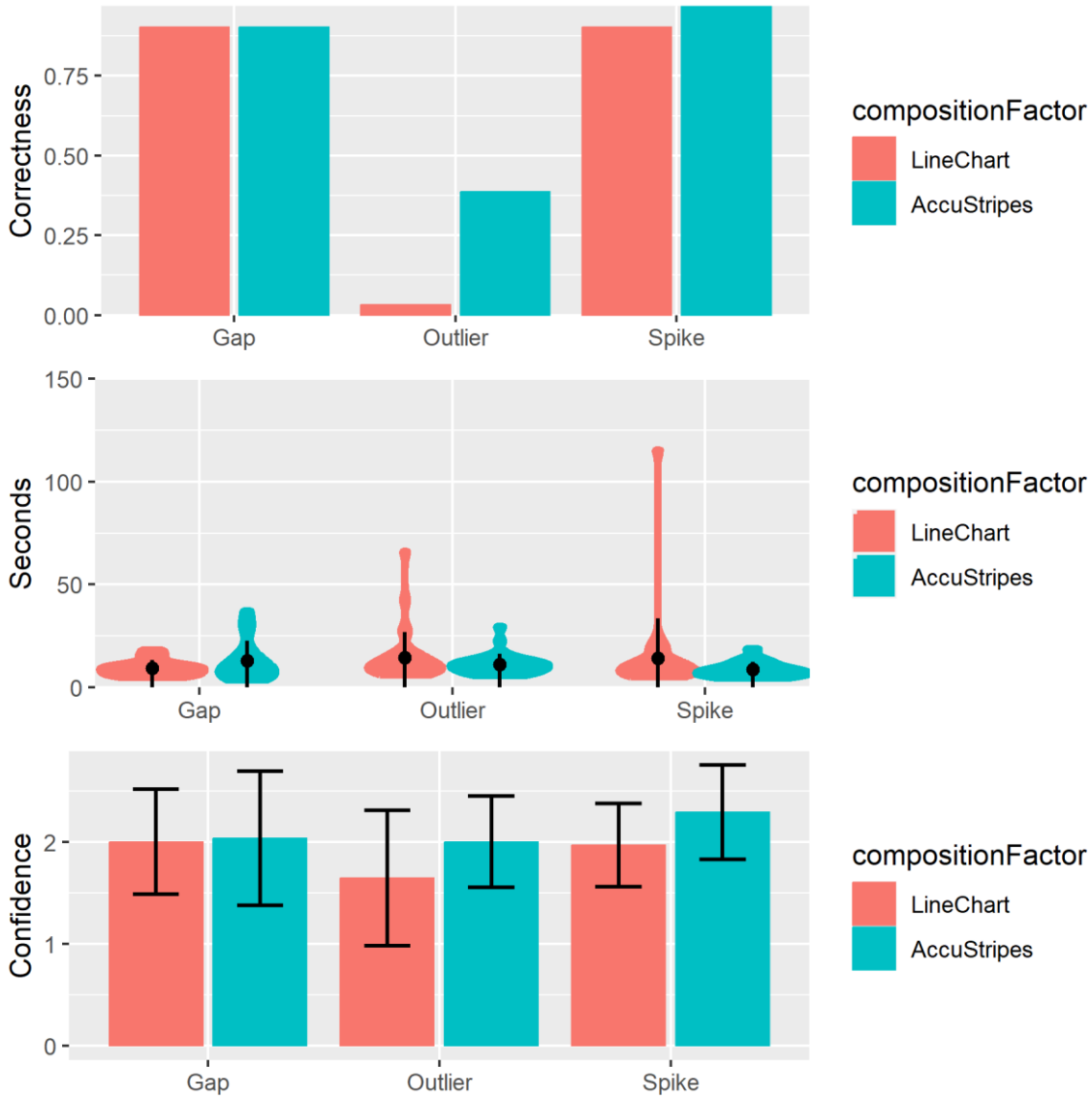
2. Post Hoc on **CONFIDENCE**

	compositionFactor_pairwise	dataflawFactor_pairwise	estimate	SE	df	t.ratio	p.value	sig.
1	LineChart - AccuStripes	Gap - Outlier	39.870968	13.85117	90	2.8785262	0.014972109	*
2	LineChart - AccuStripes	Gap - Spike	49.225806	13.85117	90	3.5539086	0.001821057	**
3	LineChart - AccuStripes	Outlier - Spike	9.354839	13.85117	90	0.6753824	1.000000000	

3. Analysis over all conditions - **CONFIDENCE**

	compositionFactor	dataflawFactor	variable	n	min	max	median	q1	q3	iqr	mad	mean	sd	se	ci	ci_min	ci_max
1	AccuStripes	Gap	Confidence	31	0	3	2	2.0	2	0.0	0	2.032	0.657	0.118	0.241	1.791	2.273
2	LineChart	Gap	Confidence	31	0	3	2	2.0	2	0.0	0	2.000	0.516	0.093	0.189	1.811	2.189
3	AccuStripes	Outlier	Confidence	31	1	3	2	2.0	2	0.0	0	2.000	0.447	0.080	0.164	1.836	2.164
4	LineChart	Outlier	Confidence	31	0	2	2	1.5	2	0.5	0	1.645	0.661	0.119	0.242	1.403	1.887
5	AccuStripes	Spike	Confidence	31	2	3	2	2.0	3	1.0	0	2.290	0.461	0.083	0.169	2.121	2.459
6	LineChart	Spike	Confidence	31	1	3	2	2.0	2	0.0	0	1.968	0.407	0.073	0.149	1.819	2.117

2. Flaw x Composition – Summary



References

[1] Wobbrock, JO, Findlater, L, Gergle, D, Higgins, JJ. The aligned rank66 transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2011,doi:10.1145/1978942.1978963.