What is the breakdown of schema change in terms of the different types of change in schema evolution?

P. Vassiliadis, A. Karakasidis [2024-11-28]

1 Schema Change Breakdown

We have studied the breakdown of the types of changes that take place in the history of the studied schemata, and relate it to the time patterns we introduce. In this Section, we start with a brief description of the respective terminology, and then move on to discuss the statistical profile of the different change types as well as their relationship to the time-related patterns of schema evolution. The long version of this analysis contains scatter-plots, box-plots and violin-plots of the presented data[1].

1.1 Terminology and how to measure change

A Schema History is a list of commits (a.k.a. versions) of the same DDL file of a database schema, ordered over time. A transition from an older version i to its subsequent version i +1 occurs at the timepoint where version i +1 is committed, and potentially incurs changes in the schema. The initial, originating version of the history is called, as shorthand, V_s0. Active commits are the commits whose sum of updates (see next) exceeds zero. Non-Active commits involve changes in comments, directives to the DBMS, INSERT statements, indexing, and other changes that do not affect the logical capacity of the schema in terms of tables, attributes, data types or primary keys. The Schema Update Period (SUP) is the time span (in human time) between the first and the last commit of the schema file. This is a very different time interval than its superset, Project Update Period (PUP) that marks the start and end of project history.

For each transition of the schema history, our tool, *Hecate*[2], automatically computes several categories of measurements. First, Hecate computes timing information, like the distance of the i +1 commit from V_s0 in days, and the running month and year. Second, the tool registers the schema size (no. of tables, attributes) of both the older and the subsequent version of the transition. Third, Hecate identifies and quantifies updates (all measured in attributes):

- attributes born with a new table,
- attributes injected into an existing table,
- attributes deleted with a removed table,
- attributes ejected from a surviving table,
- attributes having a changed data type,
- attributes with participation in a changed primary key.

[1] Available at https://github.com/DAINTINESS-Group/Schema_Evolution_Datasets/tree/master/SchemaEvolutionDatasets2020/2025_EDBT_TimeRelatedPatternsSchemaEvolution

[2] Hecate is available at https://github.com/DAINTINESS-Group/Hecate

We measure as Expansion the sum of attributes born and injected, and as Maintenance, the sum of all the other categories. Total Activity, or simply Activity, of the schema, is the sum of Expansion and Maintenance.

In what follows, remember that the fundamental unit of measurement of change in our setting is the attribute, for all categories. Also keep in mind that in all the aforementioned measures, we measure change *after* the birth of the schema (*without* incorporating the number of *attributes at birth* in the respective measurements and sums).

1.2 Change type breakdown in the corpus of studied projects, without patterns being involved

Figure 1.1 depicts discretized percentages of expansion and maintenance, along with their constituent change types in groups of 0.25. Therefore, if a project has, for example, 27% of its change as expansion and 78% of its change as maintenance, it ends up in the second bucket for the Expansion category and the fourth bucket for the Maintenance category. For each of the buckets of each category, we count (i) how many projects pertain to it, and, (ii) what percentage of the overall corpus this population constitutes. Ideally, the Expansion and Maintenance columns would have been antisymmetric, if it was not for the 27 projects that have absolutely no change whatsoever (thus a pair of <0%,0%> for expansion and maintenance), as well as some rounding errors.

	EXPANSION		Attr @ T.Ins		Attr Injected		MAINTENANCE		Attr @ T.Del		Attr Ejected		Data Type Ch.		PK Change	
%change in ...	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
[0.00-0.25)	47	31%	82	54%	110	73%	66	44%	118	78%	146	97%	123	81%	147	97%
[0.25-0.50)	24	16%	36	24%	17	11%	38	25%	26	17%	3	2%	10	7%	2	1%
[0.50-0.75)	40	26%	22	15%	8	5%	27	18%	5	3%	1	1%	3	2%	2	1%
[0.75-1.00]	40	26%	11	7%	16	11%	20	13%	2	1%	1	1%	15	10%	0	0%
Grand Total	**151**	**100%**	**151**	**100%**	**151**	**100%**	**151**	**100%**	**151**	**100%**	**151**	**100%**	**151**	**100%**	**151**	**100%**

Figure 1.1 Breakdown of change types in ranges of 25% as (a) absolute population size and (b) population percentages

We observe that overall, there is a quite clear tendency of the projects of the corpus towards expansion rather than maintenance. More than half the projects (80 projects) have expansion higher or equal to 50%, thus more projects than not, are inclined towards expansion. At the same time, the respective fraction for projects with maintenance that exceeds 50% arises to 47 projects, i.e., 31% of the overall population.

The rest of the columns of Figure 1.1 detail the constituents of Expansion and Maintenance. With the exception of *Attributes Inserted with Table Births*, the rest of the change types show a strong tendency towards the [0-25%) bucket, meaning that they are rarely of substantial contribution to the mixture of changes of a project. Overall, in terms of volume, change is mostly attained via *Attributes Inserted with Table Births*, *Attributes Injected to Existing Tables* and *Attributes Deleted with Table Deletions*.

Figure 1.2 provides a frequency-oriented picture of the situation: for each project, for each change type, we have rounded the percentage of the respective change type into buckets of 12.5%; then, in Figure 1.2 we take the histogram of frequencies for each bucket, for each change type. In terms of occurrences, the most frequent type of change is *Attributes Injected to Existing Tables,* followed by *Attributes Inserted with Table Births,* and interestingly *Attributes with Data Type Updates.* Bear in mind that this has to do with the frequency and not the volume of change, that was previously discussed. The buckets in bold demonstrate where the mass of the projects actually concentrates in the histogram of each change type. As an example, *Data Type Updates* and *Attribute Injections* are

frequent, but not massive in volume. Observe also the numerous 100% cases in these two change types, too.

INS w. Tbl	Count
0.0%	72
12.5%	6
25.0%	9
37.5%	**17**
50.0%	**23**
62.5%	9
75.0%	7
87.5%	3
100.0%	5
Σ	151
Non-Zero	79

DEL w. Tbl	Count
0.0%	100
12.5%	**13**
25.0%	**12**
37.5%	**12**
50.0%	9
62.5%	3
87.5%	1
100.0%	1
Σ	151
Non-Zero	51

Type Upd	Count
0.0%	91
12.5%	**24**
25.0%	**10**
37.5%	7
50.0%	3
62.5%	1
75.0%	2
100.0%	**13**
Σ	151
Non-Zero	60

INJECTED	Count
0.0%	68
12.5%	**28**
25.0%	**20**
37.5%	**10**
50.0%	4
62.5%	5
75.0%	1
100.0%	**15**
Σ	151
Non-Zero	83

EJECTED	Count
0.0%	110
12.5%	**32**
25.0%	4
37.5%	3
50.0%	1
100.0%	1
Σ	151
Non-Zero	41

PK	Count
0.0%	143
12.5%	**4**
25.0%	2
50.0%	1
62.5%	1
Σ	151
Non-Zero	8

EXP	Count
0.0%	43
12.5%	2
25.0%	4
37.5%	**12**
50.0%	**22**
62.5%	**20**
75.0%	**17**
87.5%	4
100.0%	**27**
Σ	151
Non-Zero	108

MNTNC	Count
0.0%	54
12.5%	6
25.0%	**13**
37.5%	**20**
50.0%	**23**
62.5%	**12**
75.0%	5
87.5%	3
100.0%	**15**
Σ	151
Non-Zero	97

Figure 1.2 Popularity of change types in detail

1.3 Change type breakdown per pattern

Based on the detailed data that we list in the Appendix on the change breakdown per change type, in Table 1.1, we present a **summary of the breakdown of change, for expansion and maintenance per pattern**. We divide the percentages in ranges, (a) treating a range of higher than 60% for an arbitrary category *C* as "*mostly C*", (b) treating values in the 40%-60% range as "middle level" mixture, and also, (c) isolate the special cases of 0% for a certain category, and, (d) of no *change* (0%,0%) whatsoever, too.

	Expansion	Maintenance	#prj	%prjs of pattern
11_FlatLiner	0%	0%	11	48%
23	0%	100%	5	22%
	(0%-40%)	(60%-100%)	0	0%
	middle	level	1	4%
	(60%-100%)	(0%-40%)	4	17%
	100%	0%	2	9%
12_RadicalSign	0%	0%	6	15%
41	0%	100%	1	2%
	(0%-40%)	(60%-100%)	6	15%
	middle	level	9	22%
	(60%-100%)	(0%-40%)	9	22%
	100%	0%	10	24%
13_Sigmoid	0%	0%	8	42%
19	0%	100%	1	5%
	(0%-40%)	(60%-100%)	1	5%
	middle	level	2	11%
	(60%-100%)	(0%-40%)	3	16%
	100%	0%	4	21%
14_LateRiser	0%	0%	2	14%
14	0%	100%	3	21%
	(0%-40%)	(60%-100%)	0	0%
	middle	level	2	14%
	(60%-100%)	(0%-40%)	1	7%
	100%	0%	6	43%
21_QuantumSteps	0%	0%	0	0%
23	0%	100%	4	17%
	(0%-40%)	(60%-100%)	4	17%
	middle	level	3	13%
	(60%-100%)	(0%-40%)	8	35%
	100%	0%	4	17%
22_RegularlyCurated	0%	0%	0	0%
14	0%	100%	0	0%
	(0%-40%)	(60%-100%)	0	0%
	middle	level	7	50%
	(60%-100%)	(0%-40%)	7	50%
	100%	0%	0	0%
31_SmokingFunnel	0%	0%	0	0%
7	0%	100%	0	0%
	(0%-40%)	(60%-100%)	0	0%
	middle	level	5	71%
	(60%-100%)	(0%-40%)	2	29%
	100%	0%	0	0%
32_Siesta	0%	0%	0	0%
10	0%	100%	1	10%
	(0%-40%)	(60%-100%)	2	20%
	middle	level	4	40%
	(60%-100%)	(0%-40%)	3	30%
	100%	0%	0	0%

Table 1.1 Summary of the breakdown of expansion and maintenance in ranges per pattern

RQ. What is the change mixture of each pattern? Are there any biases or particularities?

The data reveal some important properties for each pattern which we discuss right away. We also introduce the following terminology:

- *Monothematic change* is change achieved via exactly one mega-type of change, i.e., expansion or maintenance.
- *Monotype change* is change achieved via exactly one of the 6 types of changes.

Flatliners. Flatliners are mostly projects of zero change. The rest of the projects in this pattern also have very small, almost zero, volumes of change. Ideally, one would expect flatliners having zero change entirely. However, some of them experience change within the first month after schema birth, and thus exhibit change, albeit not visually demonstrable. These projects are equally split in a maintenance only, and, mostly expansion. With exceptions, expansion is mostly achieved via attribute injections. Maintenance is mostly manifested via type updates. With few exceptions in the flatliners category, change is typically monothematic.

Radical sign. The radical sign pattern is the most populous among time-related patterns. It comes with the archetype pattern of "schema is born with or soon after project initiation; immediately after, some changes are applied to it; then, schema is frozen". The radical sign pattern is the most diverse in terms of internal mixture of change than all others. 6 projects have zero change whatsoever. With the exception of 14 projects (1/3 of the population of the pattern) of the radical sign pattern that are oriented towards maintenance, change is focused on two or three change types, mostly of expansion nature (12 projects are expansion only).

Radical sign change is predominantly achieved via table insertions and deletions. We conjecture that this is due to table and attribute renaming, which in our modeling is represented via deletions and insertions immediately after. Maintenance is mostly due to attribute deletion.

Moreover, there are six projects that hold total attribute injection as their change and nothing else. For the projects that are mostly inclined towards expansion, expansion via table birth is predominant. Also, for the projects that are mostly inclined towards expansion, change is frequently monothematic, if not monotype.

Sigmoid. Sigmoid is a category with projects that have mostly low change and are substantially monothematic. 8 out of 19 projects have no change whatsoever. 5 of the remaining 12 projects are monothematic. With few exceptions, change is small. Sigmoid projects are slightly inclined towards expansion. There is no maintenance really, except for two large projects with too many table deletions. Expansion is mostly due to attribute injections.

Late risers. Late risers are heavily monothematic (out of 12 projects demonstrating some sort of change, 9 are monothematic). They include projects with small change (with just one exception). The pattern is slightly biased towards expansion. Expansion is mostly due to attribute injections. In terms of maintenance, sigmoid has four projects with a stronger inclination to total maintenance, achieved

mostly via type updates. In all other cases, maintenance is not existent: the rest of the projects are expansion-only projects, with very small change, manifested mainly via attribute injections.

Quantum Steps. The Quantum Steps pattern is mostly expansion-oriented. The pattern has four projects that are maintenance-only, with change manifested via type updates. In the projects of the pattern where maintenance is dominant, this is done mostly via update-type updates. In the projects of the pattern with a variety of change types, maintenance is mostly achieved via deletion of entire tables. Wherever expansion is predominant in this pattern, this is clearly done via new table births.

Regularly curated projects. This is also a time-related pattern with almost all categories appearing in almost all projects (with the exception of primary key updates). Change rises to significant volumes in the projects of this pattern. The mix of change is mostly balanced, slightly in favor of expansions. However, it is clear that, in the categories of Table 1.1 there are only two ranges hosting projects (equally divided): (a) those of balanced, middle-level bias, and, (b) of slight bias towards expansion. There is no project with bias in favor of maintenance whatsoever, and no monothematic projects either. Expansion is mostly achieved via new tables. Injections clearly exist but significantly less than attributes been born with new tables. Maintenance is mostly achieved via table deletions, with (i) attributes being ejected and (ii) data types being updated constituting a second tier of types in terms of popularity.

Smoking funnel. Smoking Funnel is a time-related pattern quite similar to the one of Regularly Curated projects, in terms of its change type breakdown, both with respect to expansion and maintenance. Again, almost all projects demonstrate change in almost all the categories of change, with a slight bias towards of expansion. Expansion in half the cases is mostly via table births and the other half of the cases is mostly via attribute injections. Maintenance is mostly achieved via table deletions with (i) attributes being ejected and (ii) data types being updated constituting a second tier of types in terms of population.

Siesta. Finally, siesta projects are also heavily mixed, with each project having a large number of change types (at least three of them) in its change mixture. In contrast to the other categories, siesta projects are slightly in favor of maintenance. Maintenance is split half and half between (a) table deletions and (b) data type updates. Expansion is mostly achieved via table insertions.

With the exception of few radical sign projects that are oriented towards maintenance, the "Be Quick or Be Dead" family involves small change, frequently being zero, and an inclination towards expansion. Due to the small volume of change, the patterns of the family are frequently monothematic in their internal breakdown of change types.

In contrast, the rest of the time-related patterns come with higher volumes of changes, which is also related to a variety of change types. In terms of volume, both expansion and maintenance are performed with the granule of change being mostly the entire table (being inserted or deleted), rather than the internal restructuring of existing tables.

1.4 Change type breakdown per Change Type

As already mentioned, for each project, for each change type, we have rounded the percentage of the respective change type into buckets of 12.5%; then, in Figure 1.3 we take the histogram of frequencies (rows) for each time-related pattern (columns), divided per group of change type, i.e., Expansion and Maintenance. Similarly, in Figure 1.4, we depict the same information, now detailed per change type.

EXPANSION

Count of | Column Labels									
Row Labe	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	16	7	10	5	4			1	43
12.5%		1			1				2
25.0%		2			1			1	4
37.5%		4		1	2	2	1	2	12
50.0%	1	7	2	1	3	3	2	3	22
62.5%	2	3	1	1	4	4	3	2	20
75.0%	2	5	2		2	4	1	1	17
87.5%		2			1	1			4
100.0%	2	10	4	6	5				27
Grand Tot	23	41	19	14	23	14	7	10	151

MAINTENANCE

Count of | Column Labels									
Row Labe	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	13	16	12	8	5				54
12.5%		3			1	2			6
25.0%	3	2	2		2	2	1	1	13
37.5%	1	6	1	1	4	3	2	2	20
50.0%	1	6	2	1	3	4	3	3	23
62.5%		4		1	1	3	1	2	12
75.0%		2			2			1	5
87.5%		1	1		1				3
100.0%	5	1	1	3	4			1	15
Grand Tot	23	41	19	14	23	14	7	10	151

Figure 1.3 Breakdown of discretized percentages, as histograms, for expansion and maintenance, for all patterns

RQ: what is the correlation/orthogonality of change types and time-related pattern?

A first interesting observation from Figure 1.3 is that neither do patterns have an intense profile of change breakdown, nor change types are mostly present in the projects of a certain pattern. *The timing of change and the mixture of change types do not seem to be related in the first place*. With the exception of Flatliners, there is no pattern demonstrating an extreme bias towards a certain region of the histograms. The histogram shows some empty areas, however, with Regularly Curated and Smoking Funnel projects avoiding polarity to high and low buckets of the histograms, but this is the only consistent behavior we can spot.

RQ: How diversified are the projects of the same pattern with respect to their change profile (i.e., mixture of change types)? Do the projects of the same profile share the same change profile?

With the exception of Smoking Funnel, which has a mixture of balanced change (without extremities), *all other patterns have projects in at least 5 buckets of the histogram*. Radical Sign and Quantum Steps are the more diversified patterns in terms of change mixture (i.e., they occupy all the buckets of the histogram with at least one project, for both expansion and maintenance). Along with the almost frozen Flatiners, these two patterns constitute the most populous one, however. It is not completely certain whether there is indeed an effect between pattern population size and diversity.

RQ: Which change type is the most voluminous or frequent?

Figure 1.3 tells us that *Expansion is more frequent and more voluminous than Maintenance*. As already mentioned, there are 27 projects without any change whatsoever. Apart from these, there are 15 projects without any expansion (2/3 of them in the Be Quick or Be Dead family), and 27 projects without any maintenance (mostly in the Be Quick or Be Dead family). Apart from the aforementioned projects, the rest of the corpus has a breakdown of change with Expansion mostly ranging in the 0.375 – 0.75 range, whereas Maintenance mostly ranges in smaller percentages, 0.25 – 062.5. Therefore, as a mixture, Expansion is both more frequent and more voluminous than Maintenance.

RQ: is there any change type that overdominates the others? More generally, is the distribution of values in the histograms of the different change types biased?

Moving on to study the most detailed types of change, in Figure 1.4 we see a common theme for all change types: with the exception of few projects that fall in the 100% category bucket, all change types are avoiding the high-percentage buckets.

Figure 1.4 tells us that in general (albeit, with exceptions), all change types are characterized by a concentration of projects in the upper buckets of the histogram, up to 0.375. If the histograms were visually depicted (x-axis: bucket, y-axis: frequency), we would observe a heavy bias to the left. *All change types are mostly biased towards 0% and with the few exceptions of 100% monothematic change, few projects escape the 37.5% barrier*.

An exception to the above is the change type that attracts most of the change, i.e., *Insertion with New Tables.* This is the only category that is mostly placed towards the middle of the histogram, and is certainly the most voluminous one. As already mentioned, the two change types that have large 100% populations are (a) the most frequent one, *Injection into Existing Tables*, and, (b) *Data Type Updates*.

General Description of Curator Practices per Change Type.

The breakdown of the maintenance group of change types is really revealing. Attributes ejected are rarely present and almost only for Regularly Curated and Smoking Funnel projects. People avoid ejecting attributes from existing tables in the general case. The avoidance of changes in Primary Keys is even more widespread. Table deletions are rather absent from the populous “Be Quick or be Dead family”, and mostly present in the two other families. The case of data type updates (empirically, we have observed that most of this change concerns enlarging VARCHAR fields) is more widespread.

Concerning the expansion group of change types (attributes injected to existing tables, attributes born with new tables and total expansion), we can restate that injection is slightly more frequent than birth with new tables, mostly due to the projects of the “Be Quick or be Dead family”. Other than that, Expansion via table birth attains higher volumes of change than Expansion via injection, a phenomenon which intensifies with change frequency (Stairway to Heaven is clearly more biased in favor of Expansion via Birth).

INSERTION with New Table

Count of PR(	Column Labels								
Row Labels	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	19	17	14	10	8		1	3	72
12.5%	1	2	1		1		1		6
25.0%		1			1	4	1	2	9
37.5%	1	6	1		2	4	2	1	17
50.0%	2	6	1	1	7	2	1	3	23
62.5%		2		1	1	3	1	1	9
75.0%		3	1	2		1			7
87.5%		2			1				3
100.0%		2	1		2				5
Grand Total	**23**	**41**	**19**	**14**	**23**	**14**	**7**	**10**	**151**

INJECTION

Count of PR(	Column Labels								
Row Labels	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	17	19	11	7	8	1	2	3	68
12.5%	1	6	1		7	8	1	4	28
25.0%		6	2	2	4	4	1	1	20
37.5%	1	3		1	1		2	2	10
50.0%	1		1		1	1			4
62.5%		1	2		1		1		5
75.0%	1								1
100.0%	2	6	2	4	1				15
Grand Total	**23**	**41**	**19**	**14**	**23**	**14**	**7**	**10**	**151**

DELETION

Count of PR(	Column Labels								
Row Labels	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	19	26	17	13	15	3	1	6	100
12.5%	1	3			3	3	3		13
25.0%	1	2			3	5	1		12
37.5%	1	4	1	1	1	2	1	1	12
50.0%		4			1	1	1	2	9
62.5%		2						1	3
87.5%			1						1
100.0%	1								1
Grand Total	**23**	**41**	**19**	**14**	**23**	**14**	**7**	**10**	**151**

EJECTION

Count of PR(	Column Labels								
Row Labels	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	20	31	15	14	14	6	3	7	110
12.5%	2	7	3		8	7	3	2	32
25.0%		2				1	1		4
37.5%			1		1			1	3
50.0%		1							1
100.0%	1								1
Grand Total	**23**	**41**	**19**	**14**	**23**	**14**	**7**	**10**	**151**

DATA TYPE UPD

Count of PR(	Column Labels								
Row Labels	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	18	27	16	10	10	3	3	4	91
12.5%	1	6	1		3	9	2	2	24
25.0%	1	3			2	1	2	1	10
37.5%		4			2	1			7
50.0%			1	1				1	3
62.5%								1	1
75.0%					2				2
100.0%	3	1	1	3	4			1	13
Grand Total	**23**	**41**	**19**	**14**	**23**	**14**	**7**	**10**	**151**

PK UPD

Count of PR(	Column Labels								
Row Labels	11_FlatLiner	12_RadicalSign	13_Sigmoid	14_LateRiser	21_QuantumSteps	22_RegularlyCurated	31_SmokingFunnel	32_Siesta	Grand Total
0.0%	22	40	18	13	22	14	5	9	143
12.5%			1		1		2		4
25.0%	1							1	2
50.0%		1							1
62.5%				1					1
Grand Total	**23**	**41**	**19**	**14**	**23**	**14**	**7**	**10**	**151**

Figure 1.4 Breakdown of discretized percentages, as histograms, for expansion and maintenance constituent change types, for all patterns

1.5 A final remark on the change breakdown

Interestingly, all these results are not fully in-sync with previous literature (cf. "References") which has traditionally been based on very active projects, where evolution was amply evident. Previous research has been performed on selected projects, who, by selection, were demonstrating significant amounts of change. In our case, exactly due to the unbiased nature of the project selection process, large numbers of low-evolving projects were included, and thus, the numbers are different.

The result can be very useful for benchmark construction.

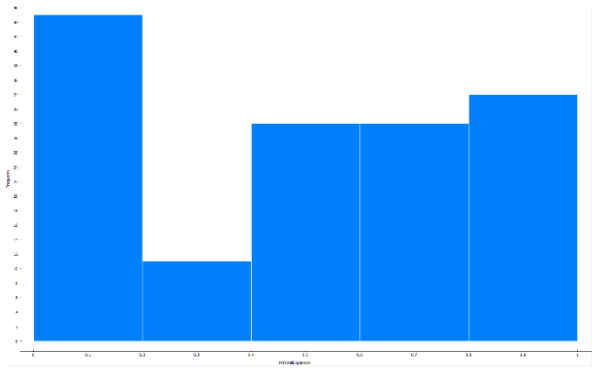
References

Curino, C.A., Tanca, L., Moon, H.J., Zaniolo, C. Schema evolution in wikipedia: toward a web information system benchmark. In: Proceedings of the ICEIS 2008

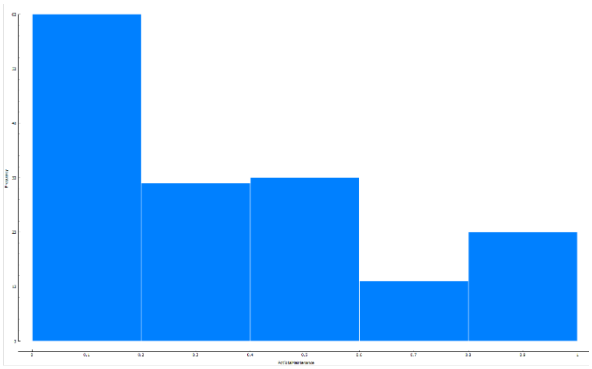
Mark Lukas Möller, Stefanie Scherzinger, Meike Klettke, Uta Störl. Why It Is Time for Yet Another Schema Evolution Benchmark - Visionary Paper. CAiSE Forum 2020: 113-125

APPENDIX A:

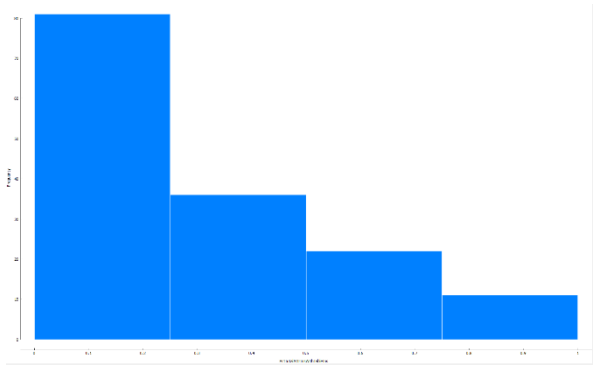
2 Overall Stats



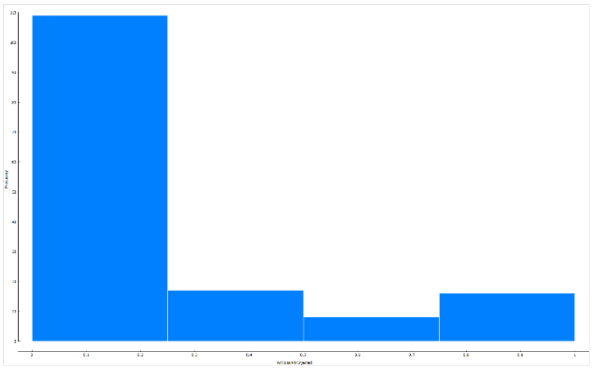
Expansion



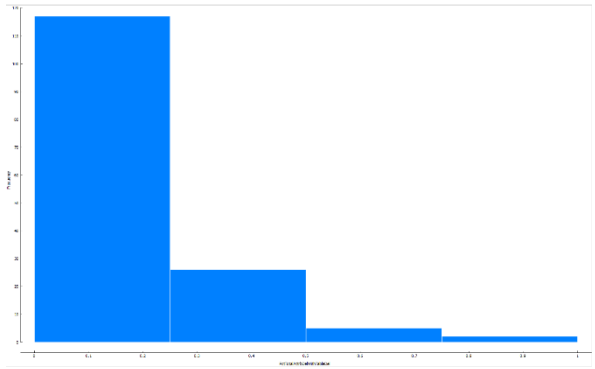
Maintenance



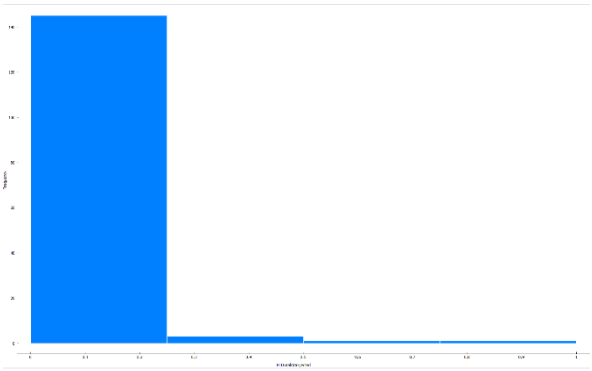
Born with Table Births



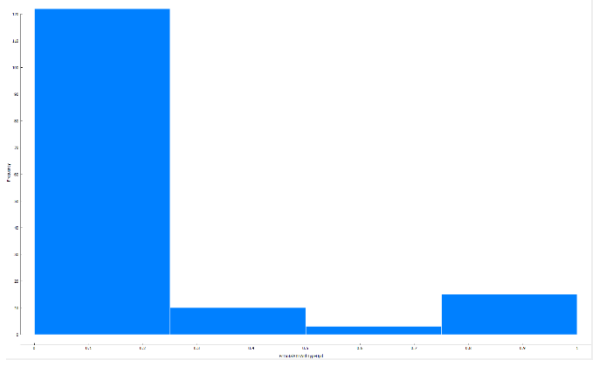
Injected in Existing Tables



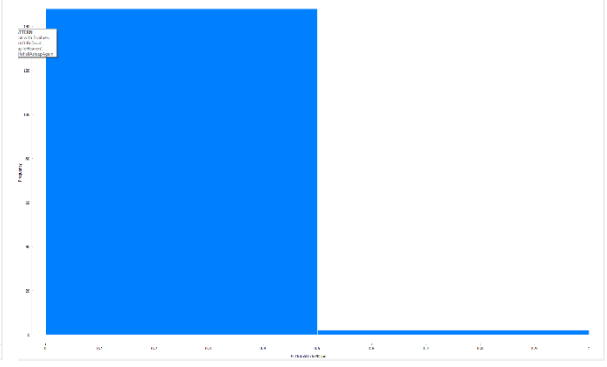
Deleted with table Deletion



Ejected from surviving tables



Data type maintenance



PK change

Figure 2.1 Histograms of the change types for the entire data set. X-axis in buckets of 0.25 within 0 to 1; Y-axis shows the number of projects for the respective bucket.

Figure 2.1 shows the histograms of the entire data set. The high numbers of the first bucket in all categories, clearly show the aversion to change.

APPENDIX B:

3 Quartiles

For each project, we have computed how total change breaks down in the different categories of change, in both absolute and ratio numbers. Here we mainly work with ratio measures, as percentages, to better show the breakdown of change. For the population of each pattern, as well as for the entire corpus, we have computed the 25%, 50% (median) and 75% quartile for each change measure. Given any arbitrary population, the interpretation of this numbers is as follows: if the value of the 25% quartile for an activity measure is large, this means that there are few projects with small volume of activity of this type in the population; similarly, if the value of the 75% quartile is small, this means that most projects up to the 75% of the population have small activity value.

Note, however, that due to the presence of predominantly small values for most of the projects, even small numbers can lead to large percentages (e.g., for a schema with just one attribute injection in its entire life, attribute injection gets 100% of the activity). For projects with zero total activity, we assign all percentages as zero, in order to illustrate the fact in the respective statistics that we use – ignoring them would be worse, in our opinion).

Figure 3.1 presents the cardinality of the corpus of each quarter of the population and Figure 3.2 presents the respective bar charts. Figure 3.3 presents the values of the quartiles. The appendices contain the scatter-plots, box-plots and violin-plots of the data.

			Expansion				Maintenance			
		#	0.00-0.25	0.25-0.50	0.50-0.75	0.75-1.00	0.00-0.25	0.25-0.50	0.50-0.75	0.75-1.00
Be Quick	11_FlatLiner	*23*	16		3	4	14	4		5
or Be Dead	12_RadicalSign	*41*	9	9	9	14	20	9	9	3
	13_Sigmoid	*19*	10		3	6	14	2	1	2
	14_LateRiser	*14*	5	1	2	6	8	2	1	3
Stairway to	21_QuantumSteps	*23*	6	4	6	7	7	5	5	6
Heaven	22_RegularlyCurated	*14*		3	8	3	3	8	3	
Scared to Fall	31_SmokingFunnel	*7*		1	6			5	2	
Asleep Again	32_Siesta	*10*	1	6	3			3	6	1
		151	**47**	**24**	**40**	**40**	**66**	**38**	**27**	**20**

			Expansion				Maintenance			
		#	0.00-0.25	0.25-0.50	0.50-0.75	0.75-1.00	0.00-0.25	0.25-0.50	0.50-0.75	0.75-1.00
Be Quick	11_FlatLiner	*23*	**69.57%**	0.00%	13.04%	17.39%	**60.87%**	17.39%	0.00%	21.74%
or Be Dead	12_RadicalSign	*41*	21.95%	21.95%	21.95%	34.15%	48.78%	21.95%	21.95%	7.32%
	13_Sigmoid	*19*	**52.63%**	0.00%	15.79%	31.58%	**73.68%**	10.53%	5.26%	10.53%
	14_LateRiser	*14*	35.71%	7.14%	14.29%	42.86%	**57.14%**	14.29%	7.14%	21.43%
Stairway to	21_QuantumSteps	*23*	26.09%	17.39%	26.09%	30.43%	30.43%	21.74%	21.74%	26.09%
Heaven	22_RegularlyCurated	*14*	0.00%	21.43%	**57.14%**	21.43%	21.43%	**57.14%**	21.43%	0.00%
Scared to Fall	31_SmokingFunnel	*7*	0.00%	14.29%	**85.71%**	0.00%	0.00%	**71.43%**	28.57%	0.00%
Asleep Again	32_Siesta	*10*	10.00%	**60.00%**	30.00%	0.00%	0.00%	30.00%	**60.00%**	10.00%
		151	**31.13%**	**15.89%**	**26.49%**	**26.49%**	**43.71%**	**25.17%**	**17.88%**	**13.25%**

Figure 3.1 Breakdown of Change-Type in Quartiles as (a) absolute numbers ad (b) population percentages

Figure 3.2 Breakdown of Change-Type in Quartiles as (a) absolute numbers ad (b) population percentages as a bar chart

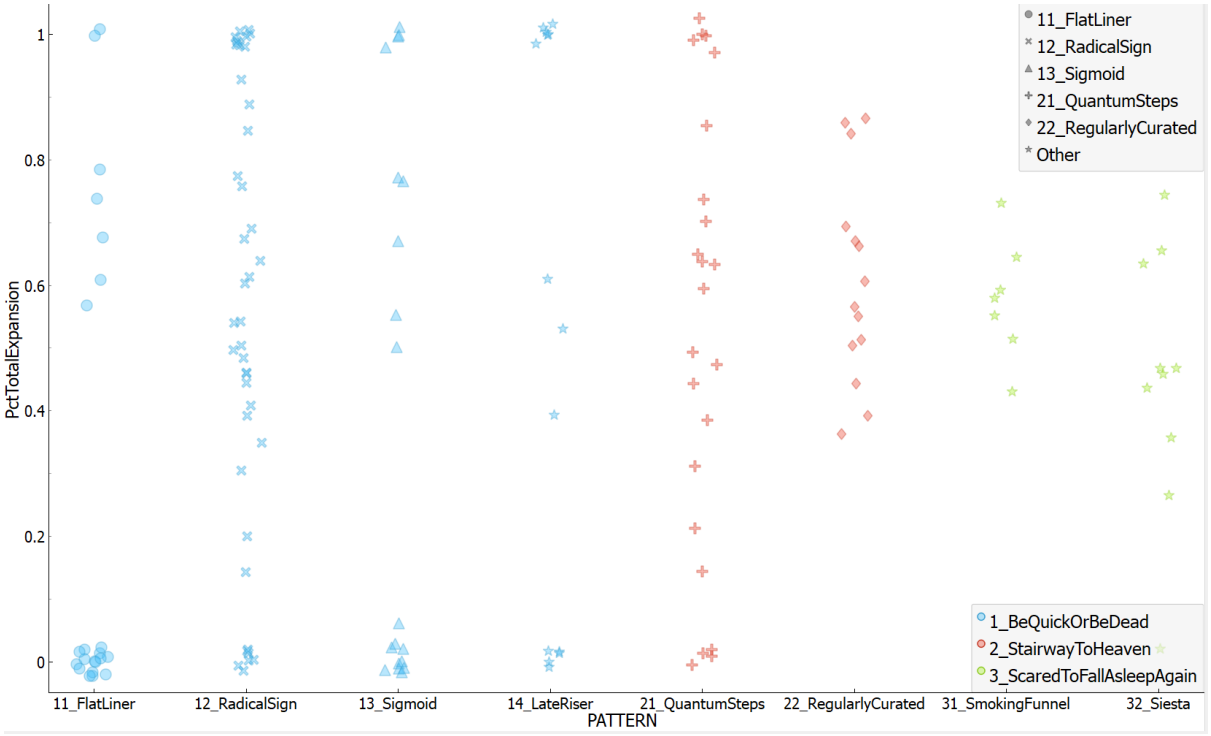
#prj	mega pattern	pattern	Total Expansion			Total Attr Ins With Table Ins			Total Attr Injected			Total Maintenance			Total Attr Del With Table Del			Total Attr Ejected			Total Attr With Type Upd			Total Attr In PK Upd		
			Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
23		**11_FlatLiner**	**0%**	**0%**	**61%**	0%	0%	0%	0%	0%	11%	**0%**	**0%**	**41%**	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%
41	1_Be Quick Or Be	**12_RadicalSign**	**36%**	**54%**	**91%**	0%	31%	54%	0%	11%	25%	**0%**	**26%**	**52%**	0%	0%	30%	0%	0%	5%	0%	0%	11%	0%	0%	0%
19	Dead	**13_Sigmoid**	**0%**	**6%**	**77%**	0%	0%	10%	0%	0%	39%	**0%**	**0%**	**29%**	0%	0%	0%	0%	0%	1%	0%	0%	1%	0%	0%	0%
14		**14_LateRiser**	**0%**	**57%**	**100%**	0%	0%	35%	0%	13%	85%	**0%**	**0%**	**57%**	0%	0%	0%	0%	0%	0%	0%	0%	35%	0%	0%	0%
23	2_Stairway To	**21_QuantumSteps**	**26%**	**61%**	**81%**	0%	39%	48%	0%	14%	24%	**19%**	**39%**	**74%**	0%	0%	17%	0%	0%	9%	0%	12%	56%	0%	0%	0%
14	Heaven	**22_RegularlyCurated**	**51%**	**59%**	**70%**	31%	41%	56%	11%	15%	20%	**30%**	**41%**	**49%**	9%	21%	27%	3%	7%	12%	6%	10%	15%	0%	0%	0%
7	3_Scared To Fall	**31_SmokingFunnel**	**52%**	**58%**	**61%**	19%	32%	44%	11%	26%	37%	**39%**	**42%**	**48%**	13%	16%	28%	4%	7%	12%	4%	8%	17%	0%	0%	5%
10	Asleep Again	**32_Siesta**	**38%**	**46%**	**60%**	6%	32%	47%	3%	12%	26%	**40%**	**54%**	**63%**	0%	0%	43%	0%	3%	7%	6%	12%	44%	0%	0%	3%
151	**TOTAL**	**TOTAL**	**0%**	**51%**	**76%**	**0%**	**15%**	**48%**	**0%**	**11%**	**28%**	**0%**	**33%**	**55%**	**0%**	**0%**	**19%**	**0%**	**0%**	**7%**	**0%**	**2%**	**17%**	**0%**	**0%**	**0%**

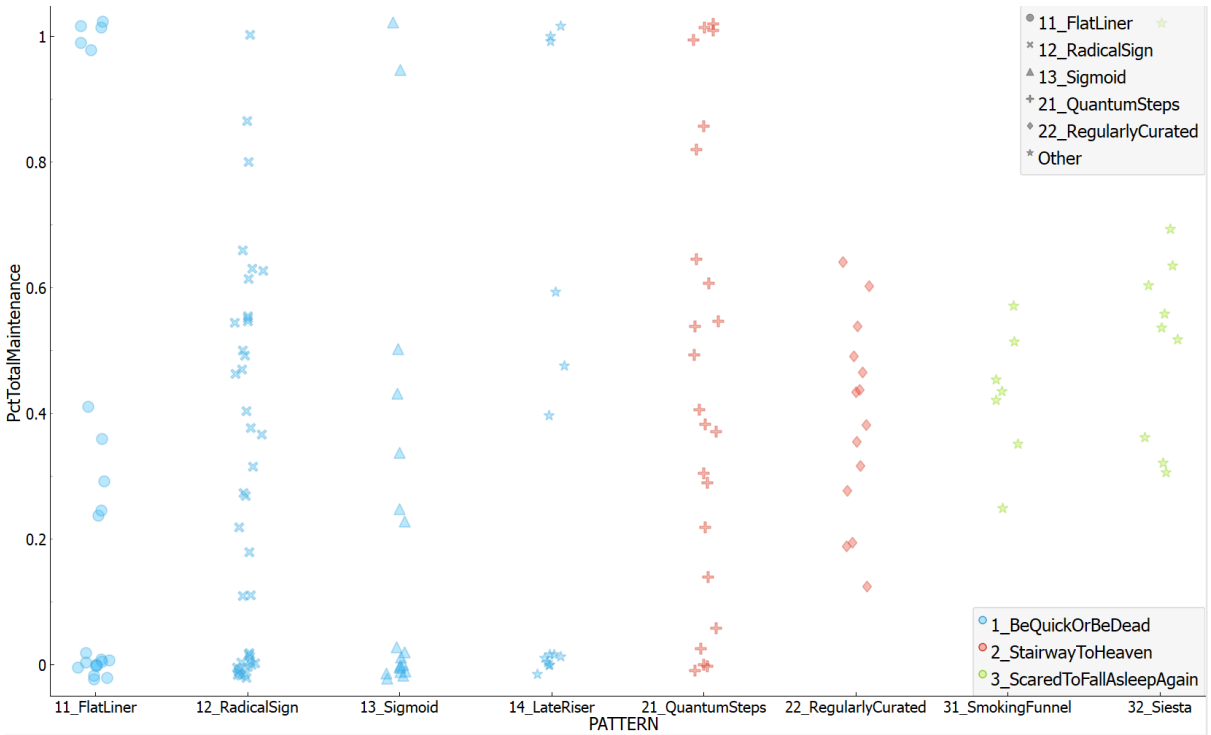
Figure 3.3 Quartiles of Change-Type Breakdown

APPENDIX C

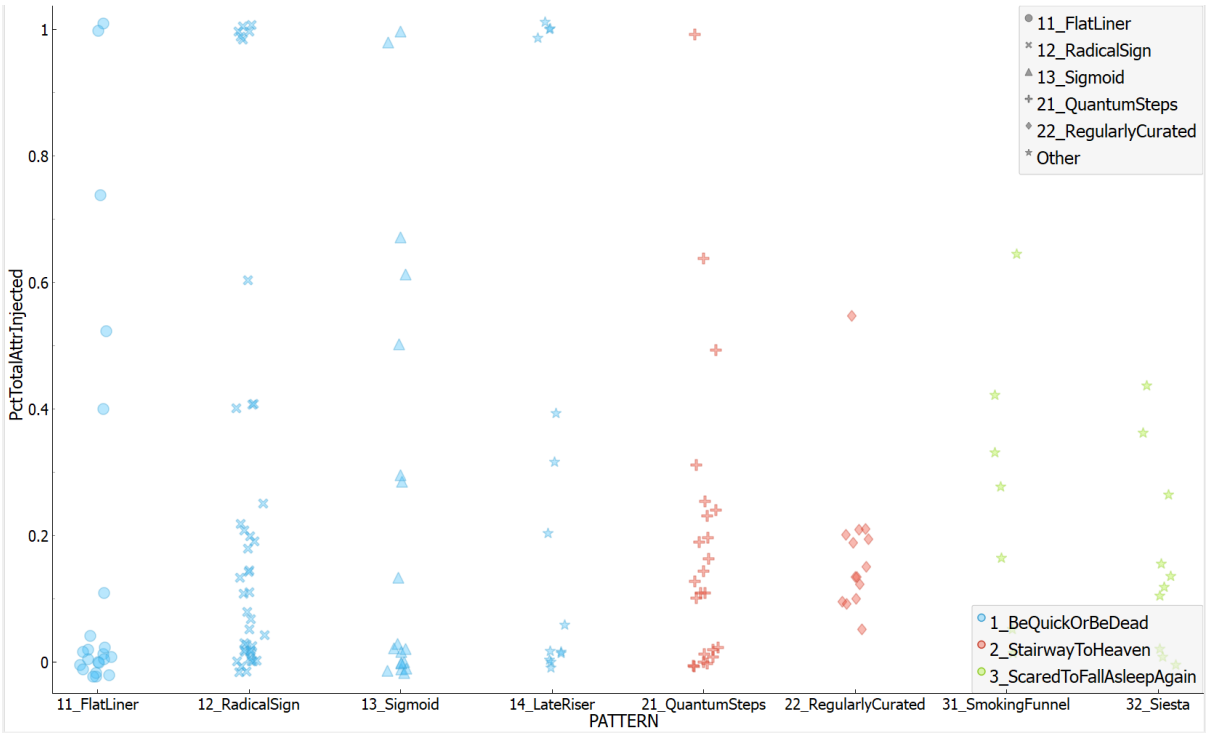
Different Charts for different Patterns

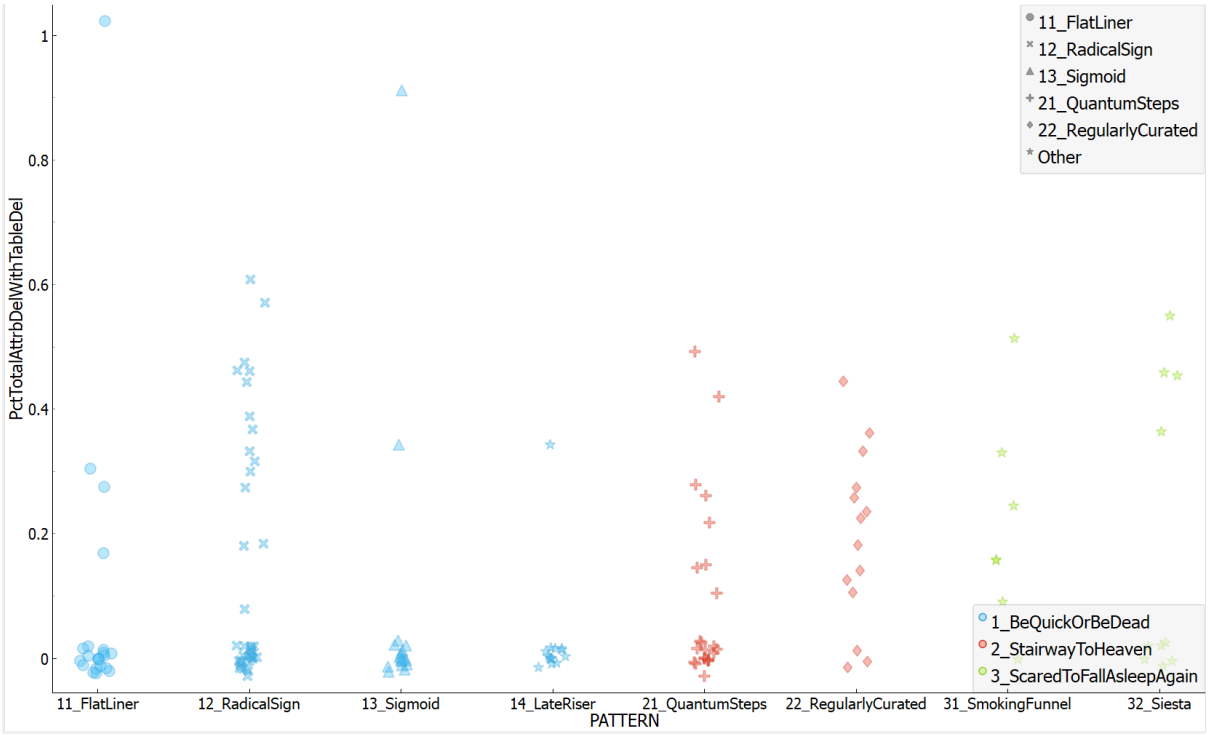
4 Scatterplots for the different categories of change



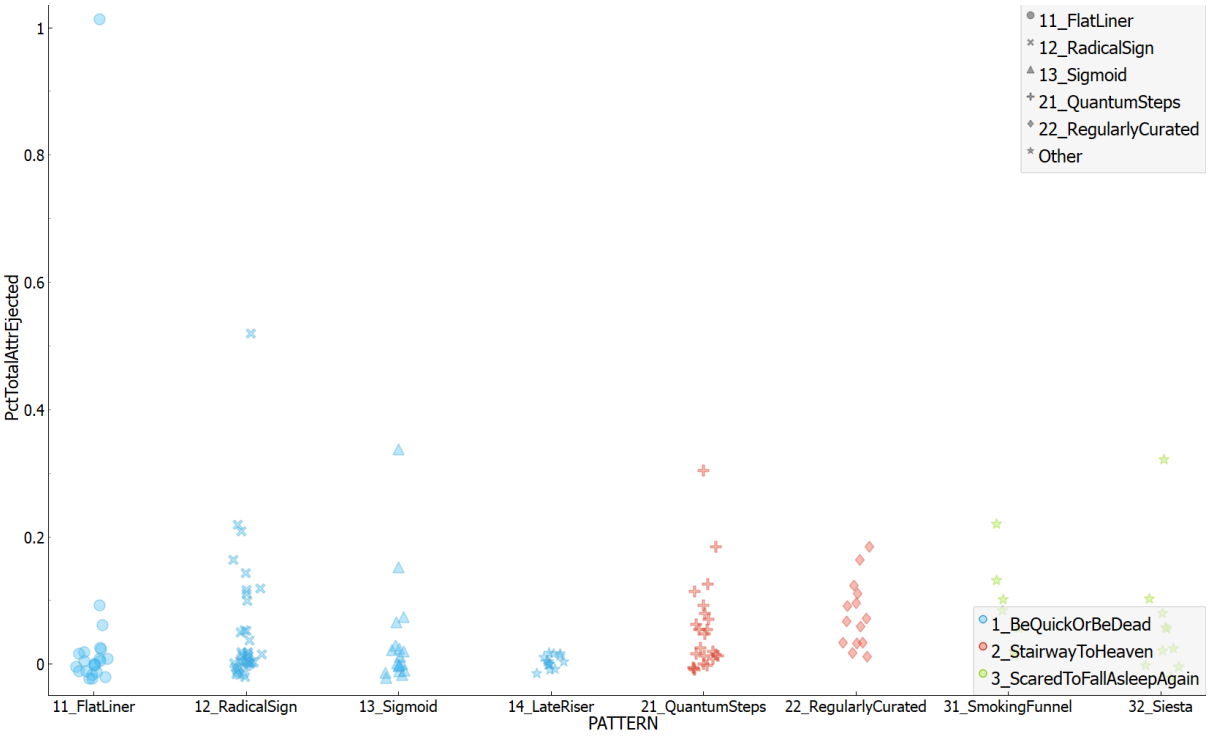


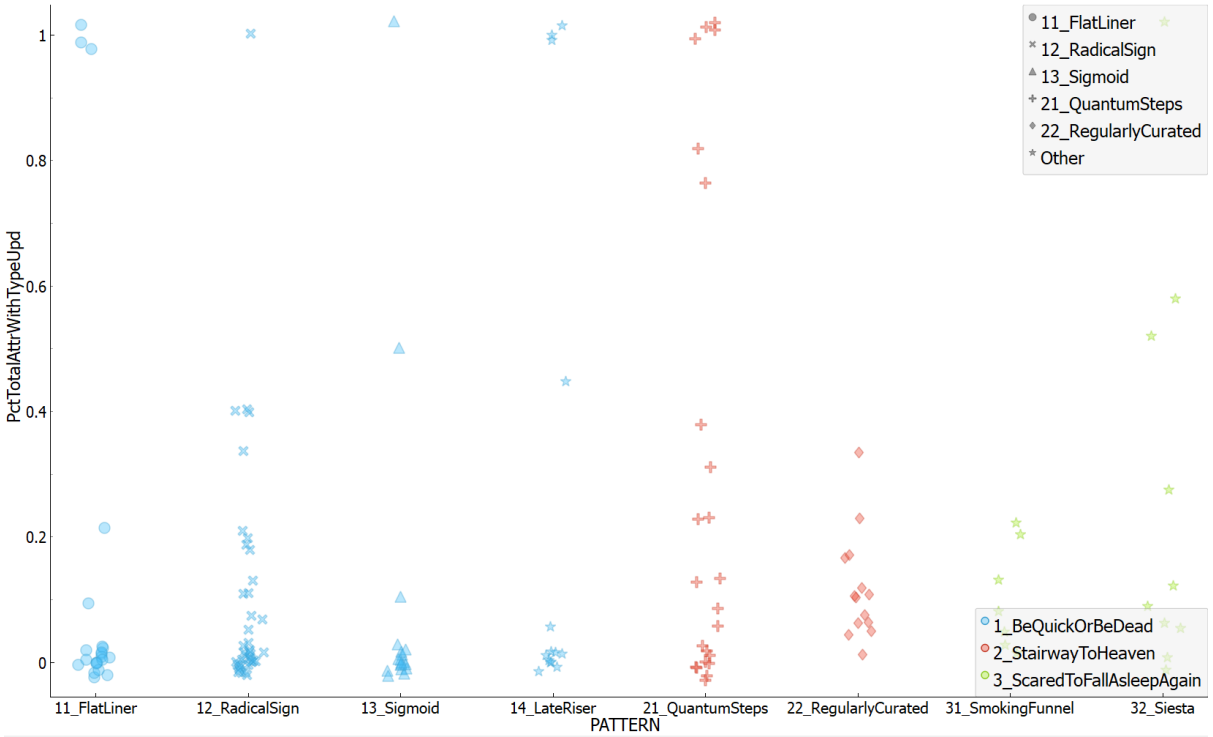
Contrast the two scatterplots: injection has (a) some “fanatics” of 100% and (b) otherwise lower levels of participation to total change compared to birth with new tables.



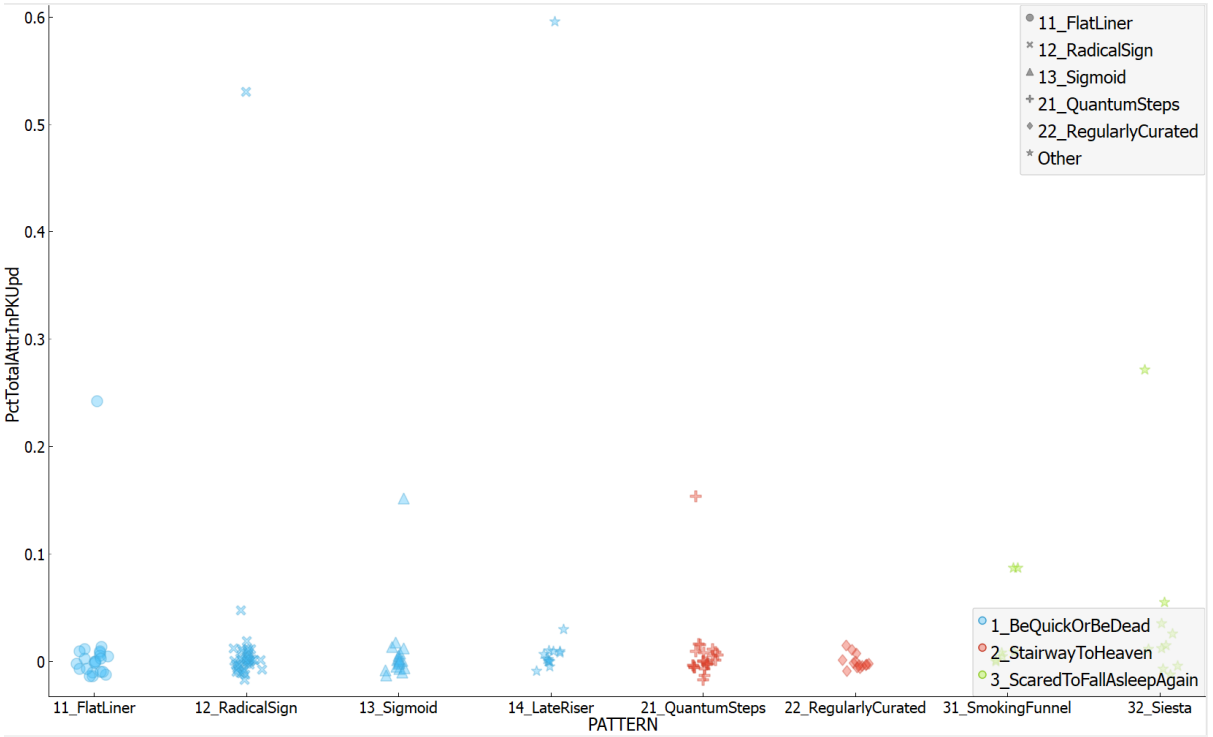


Contrast the two scatterplots: removal with entire tables is more popular than ejection from tables that continue to exist (very few cases surpass 20% of change due to ejection).



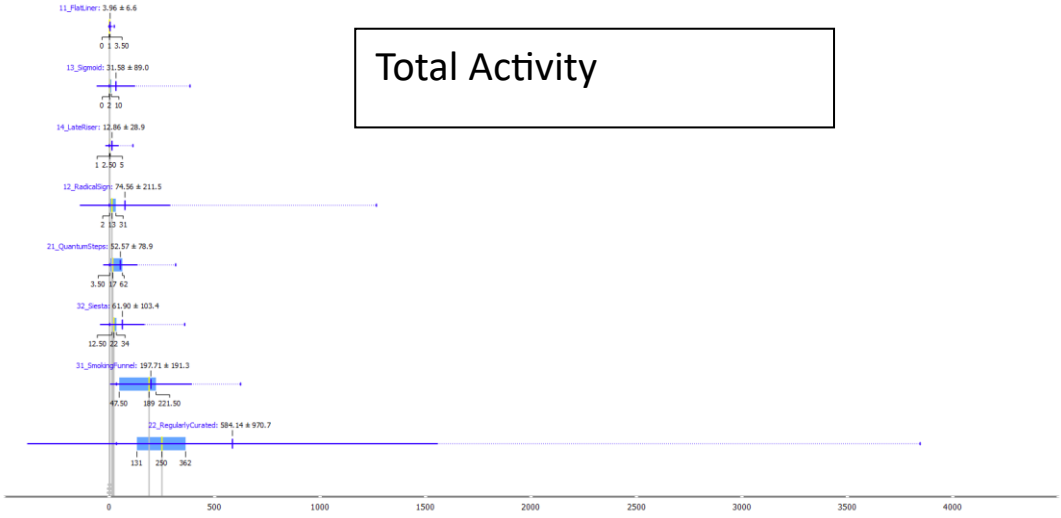


Data type changes has its “fanatics” of 100% (as well as many projects where it does not exist at all) and several projects where it exists in the 10% - 40%. PK change is practically inexistent.

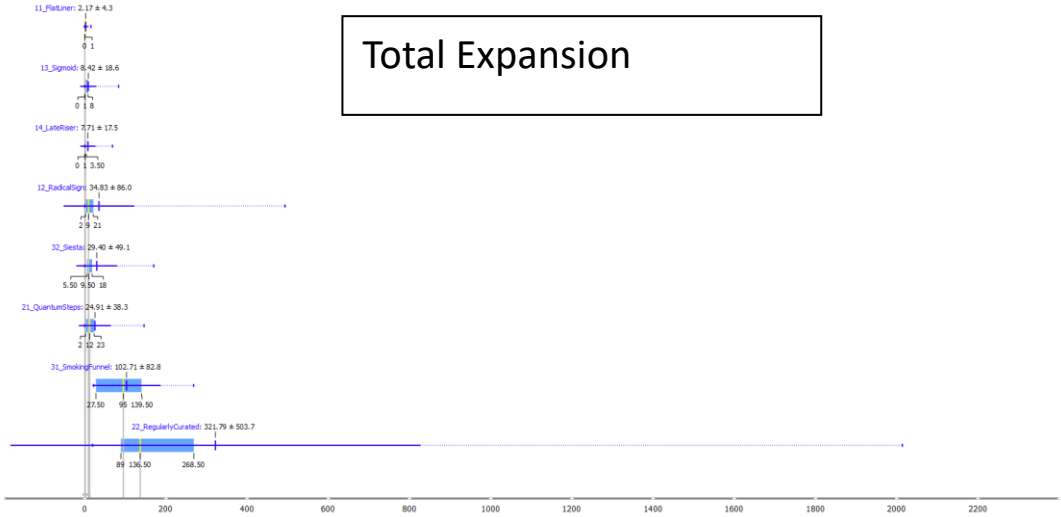


PAGE BREAK

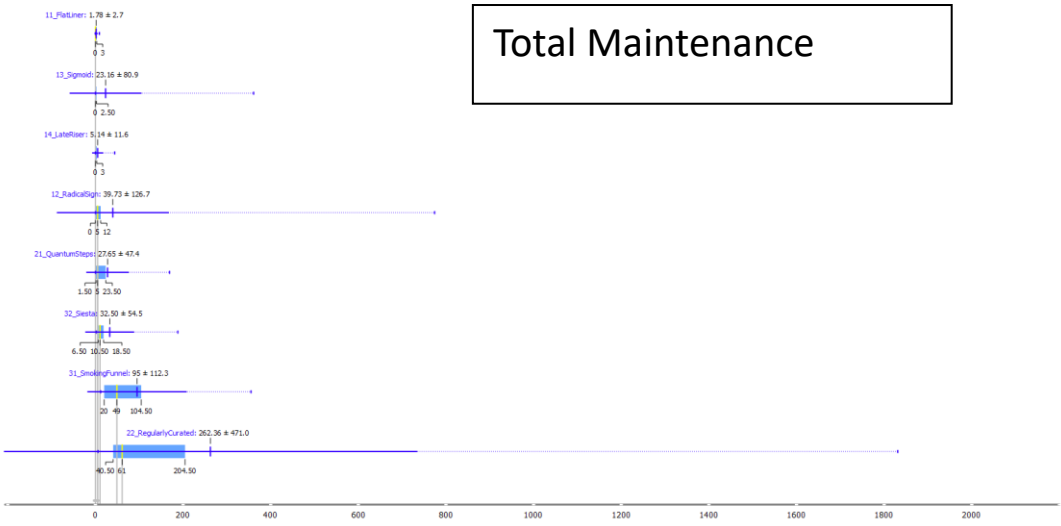
5 Box Plots of Absolute Values for the different Categories of Change



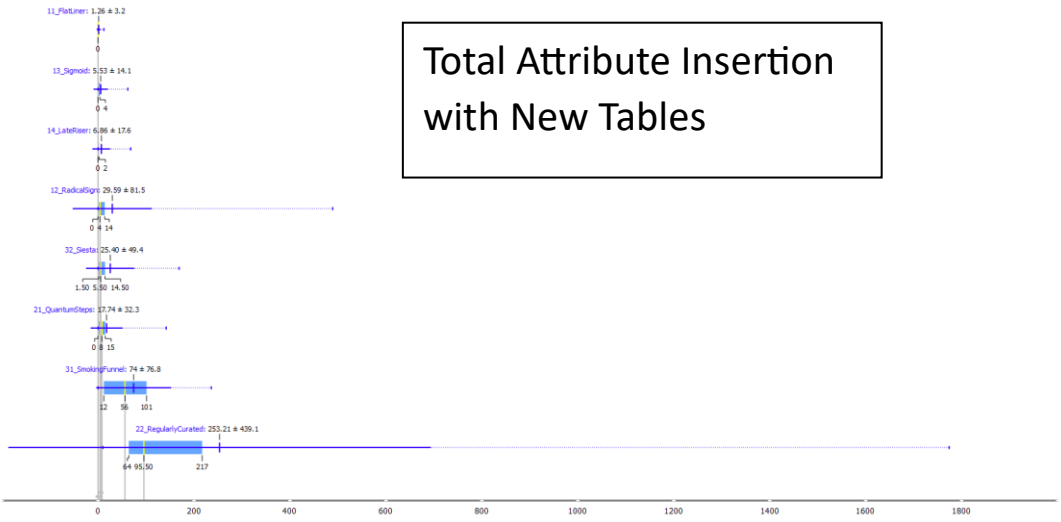
Total Activity



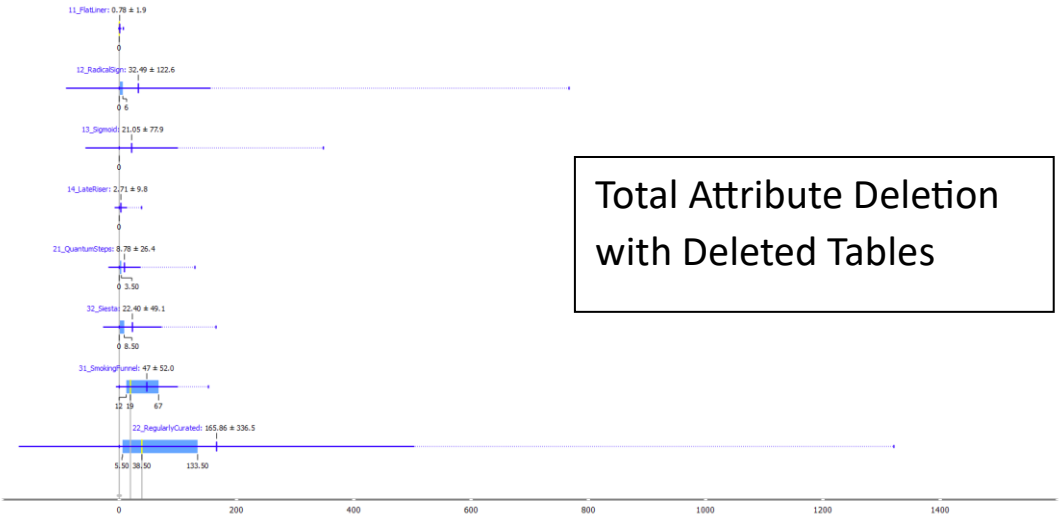
Total Expansion



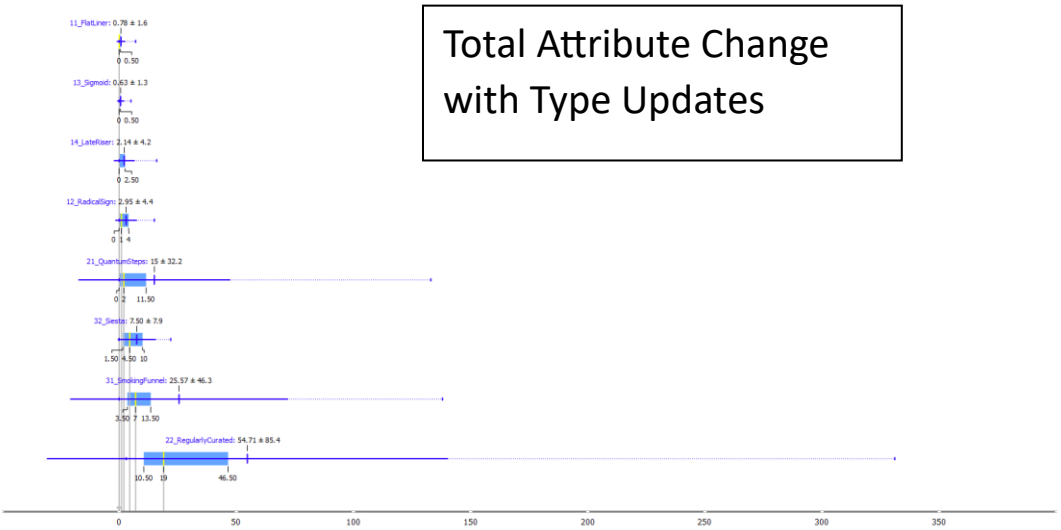
Total Maintenance



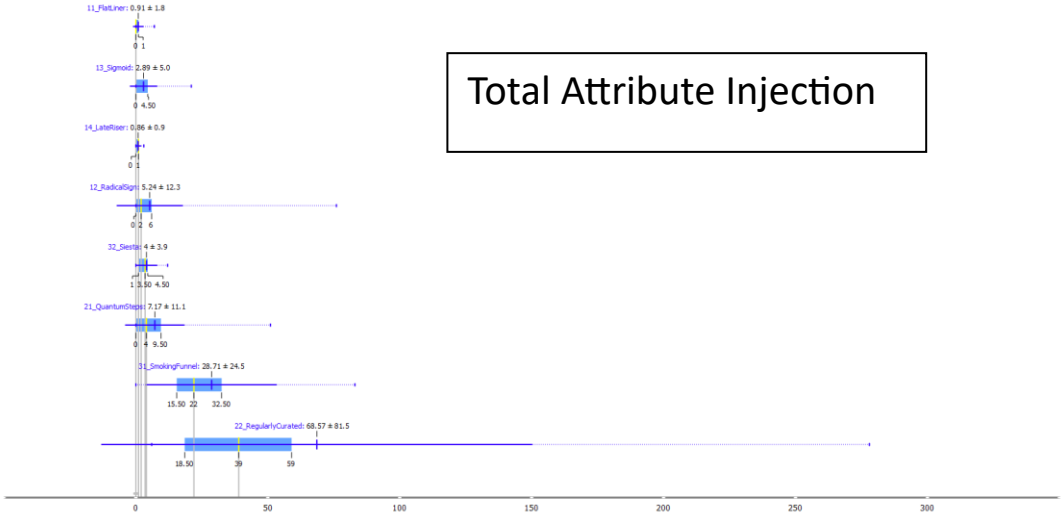
Total Attribute Insertion with New Tables



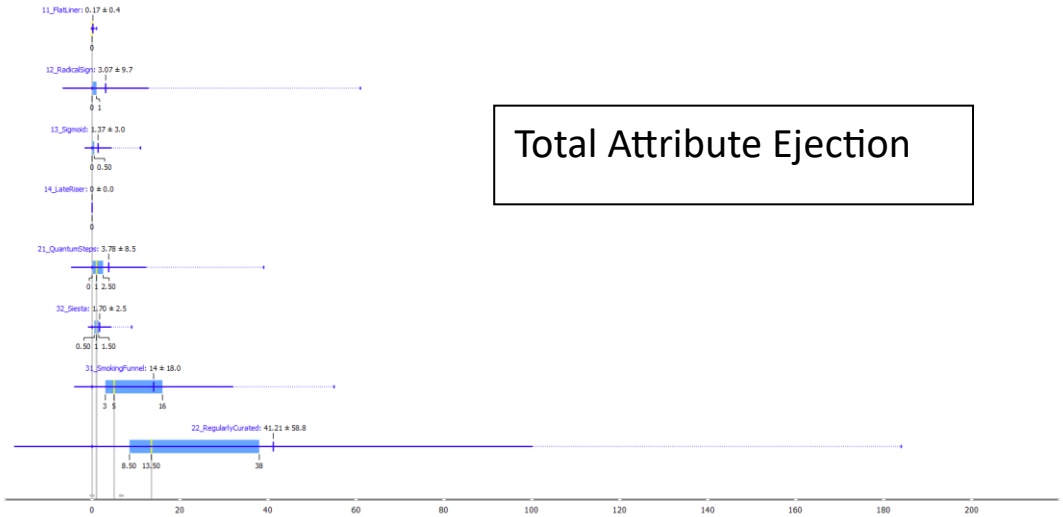
Total Attribute Deletion with Deleted Tables



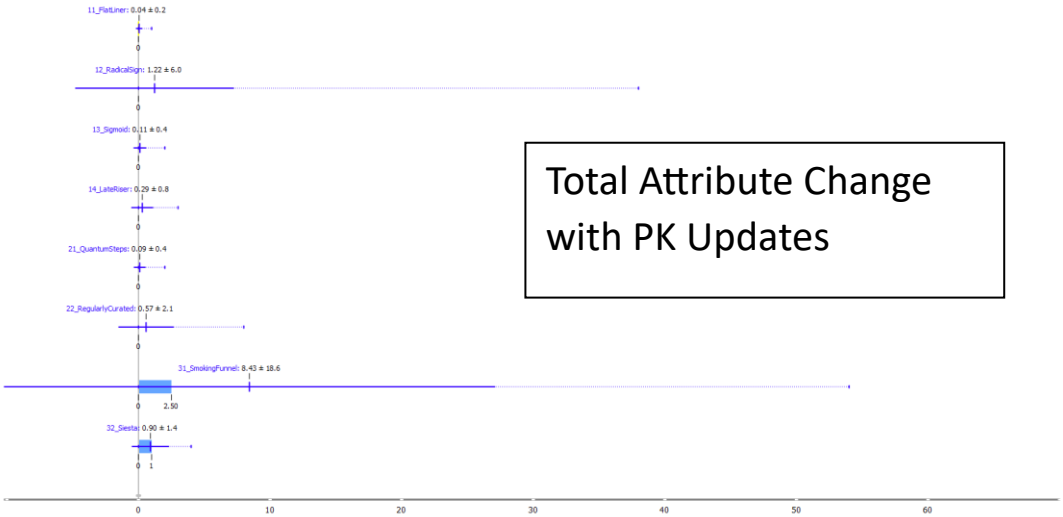
Total Attribute Change with Type Updates



Total Attribute Injection

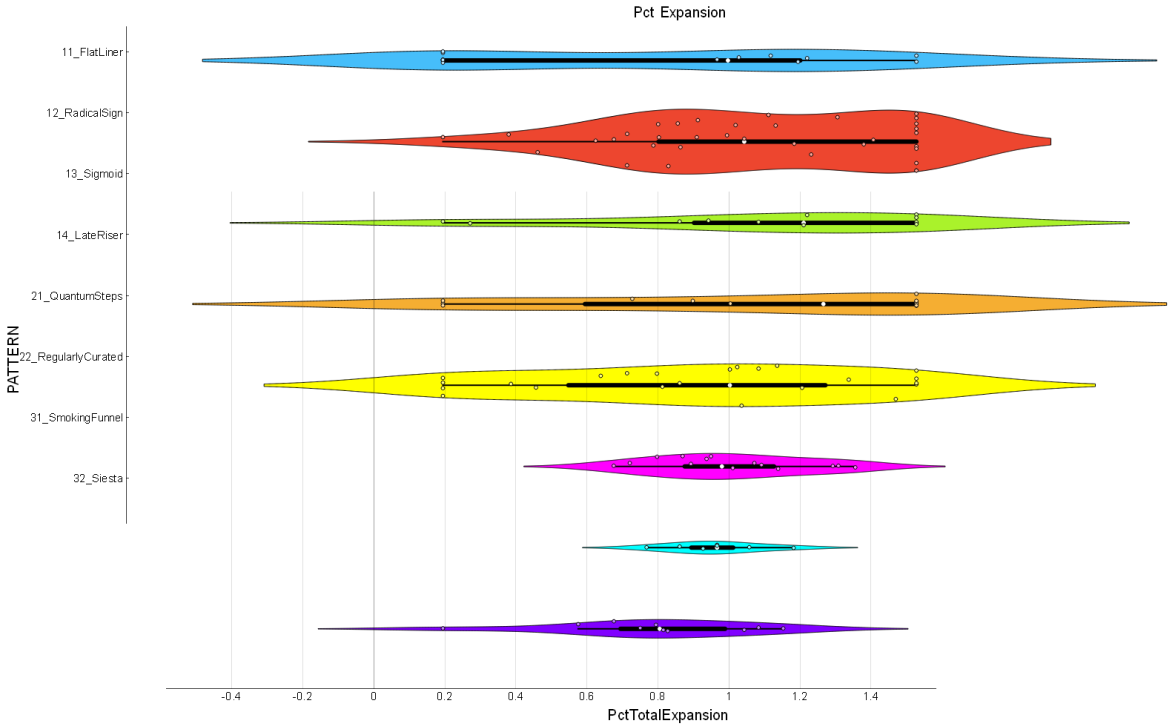


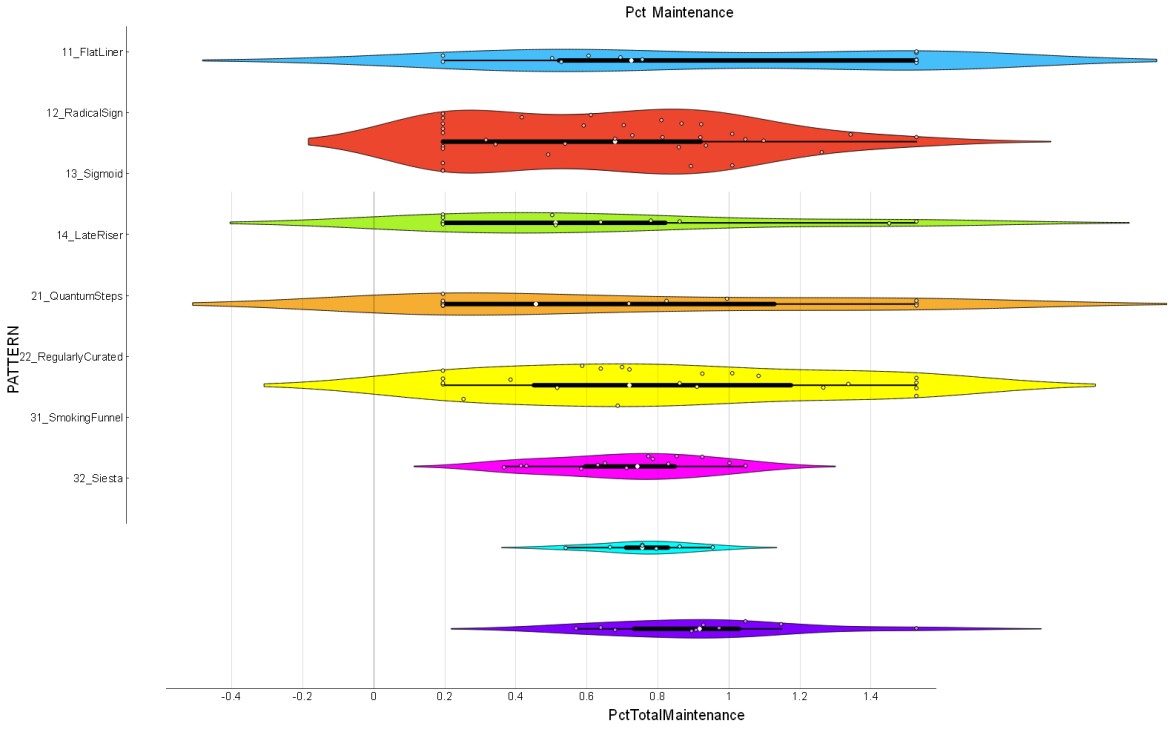
Total Attribute Ejection

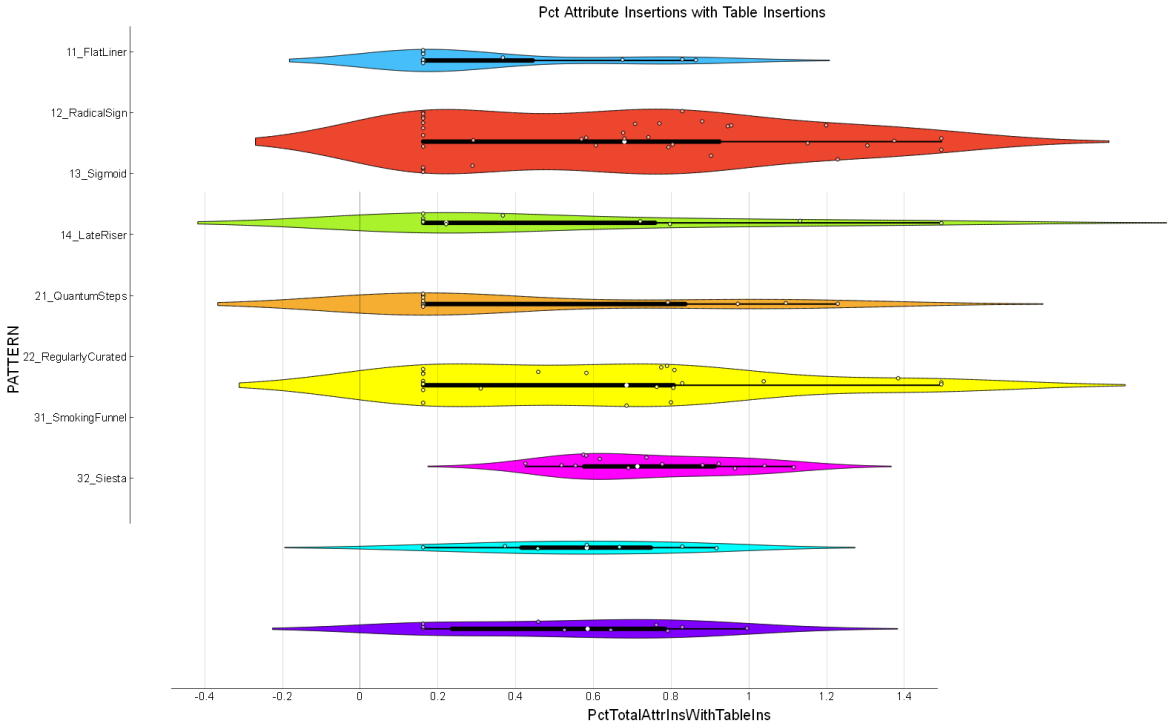


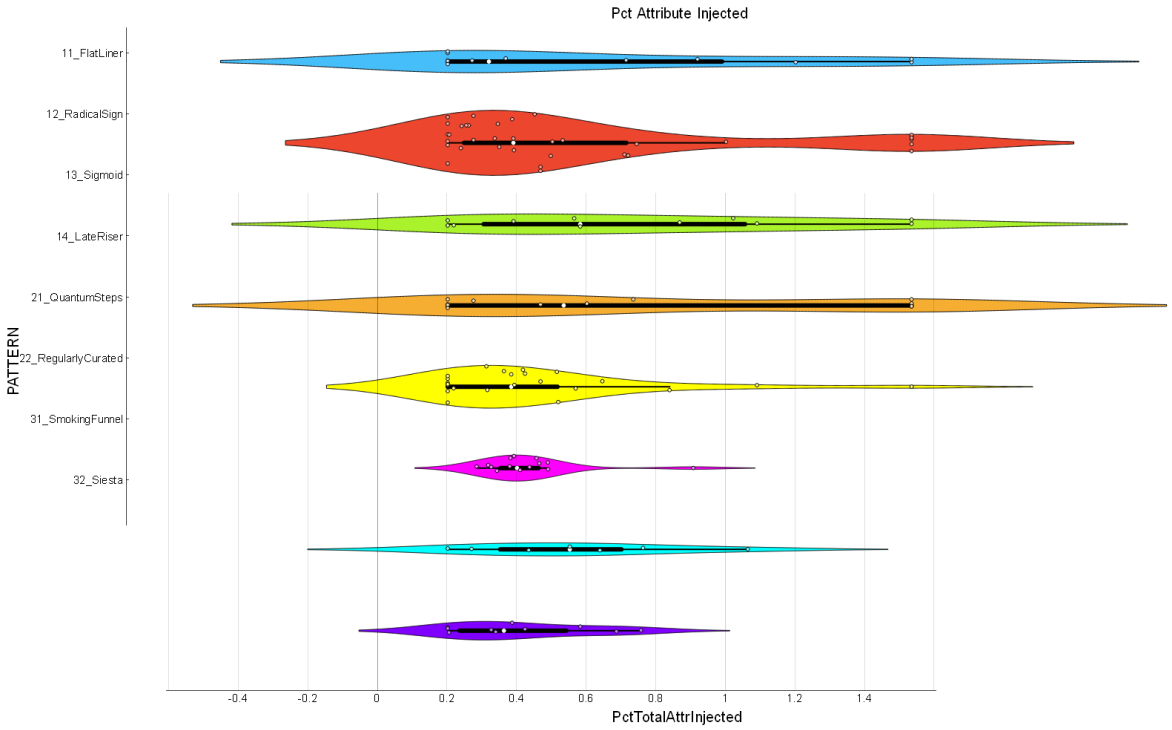
Total Attribute Change with PK Updates

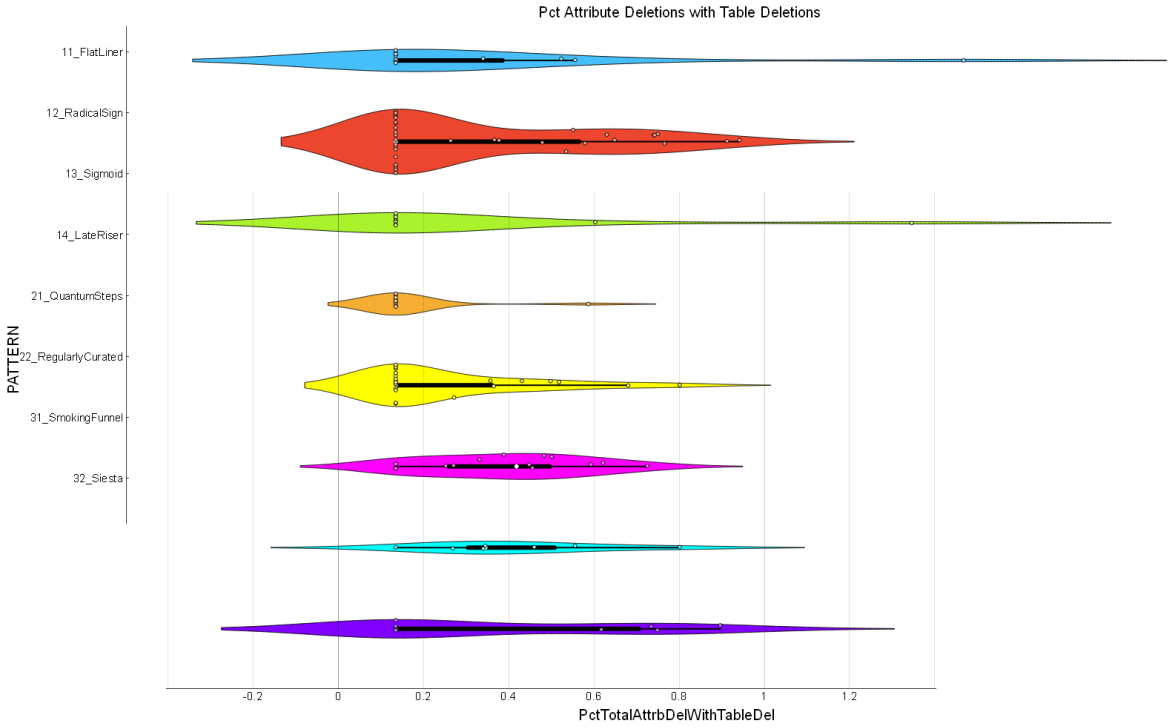
6 Violin Plots of Percentages for the different Categories of Change

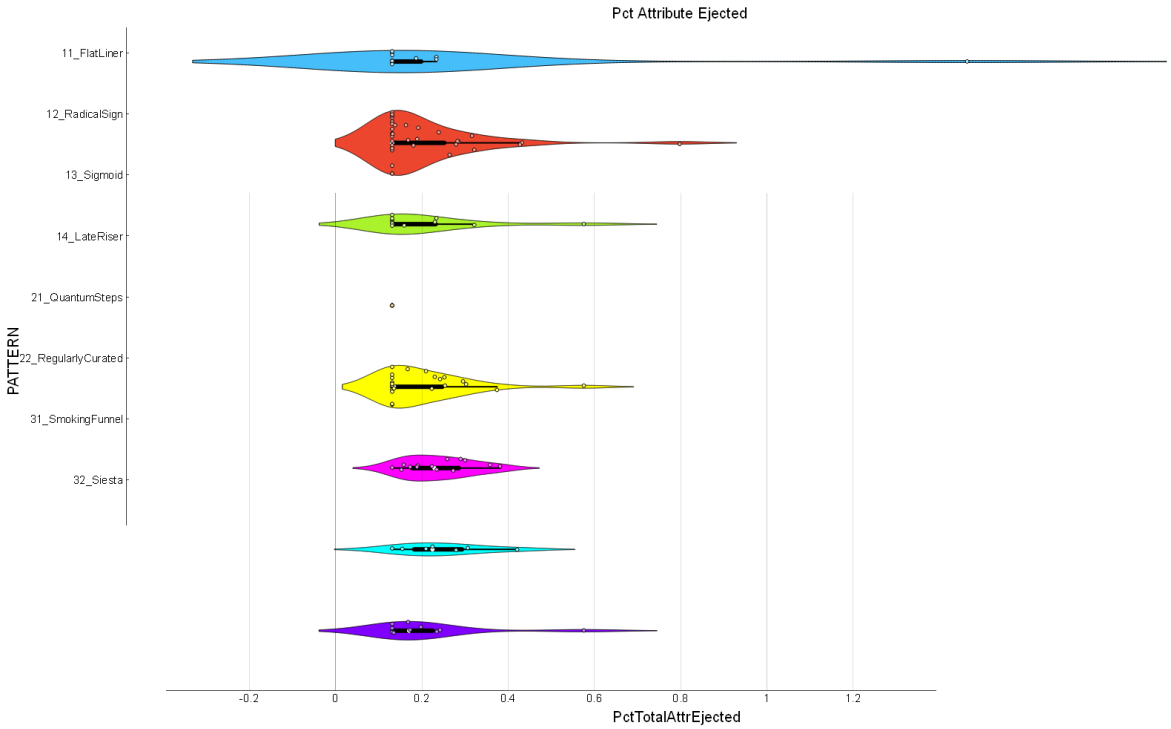


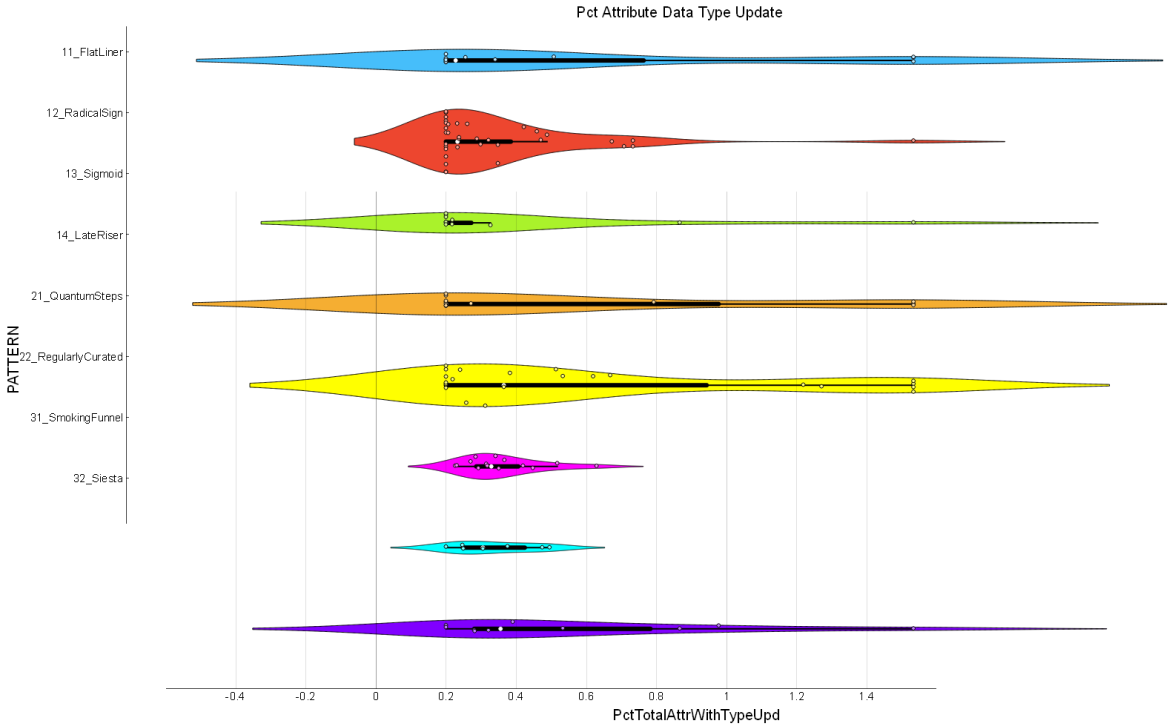


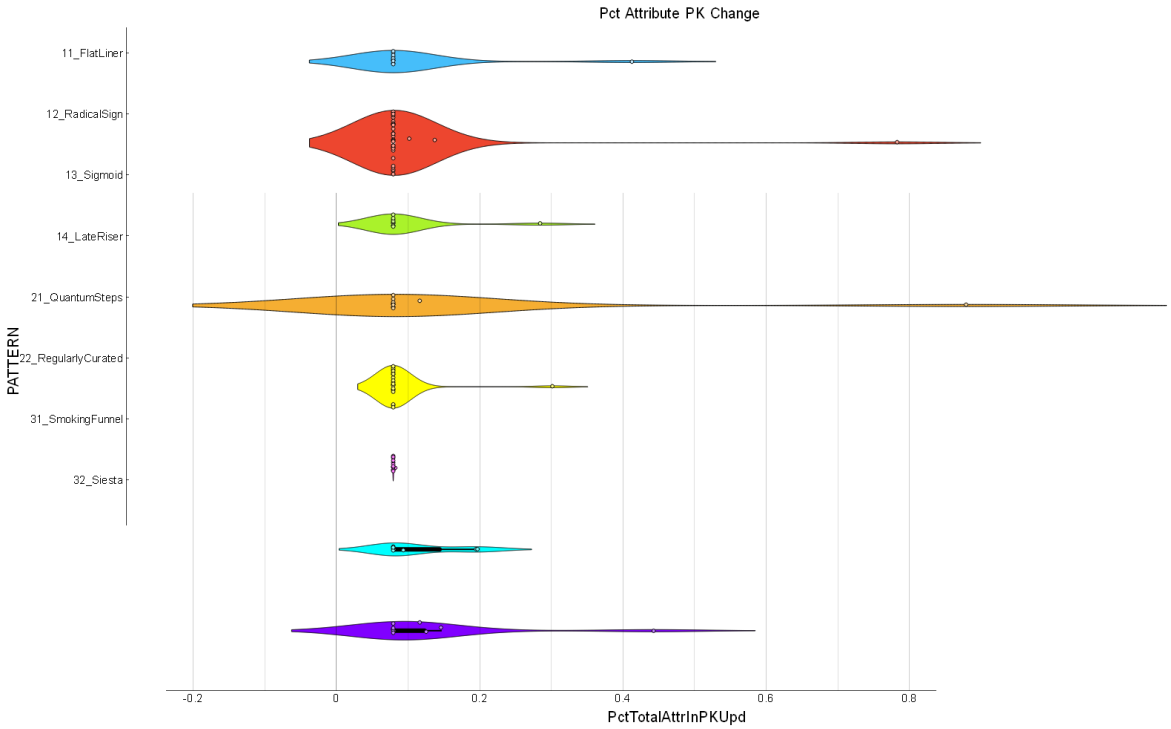

Pct Attribute Insertions with Table Insertions
11_FlatLiner
12_RadicalSign
13_Sigmoid
14_LateRiser
21_QuantumSteps
22_RegularlyCurated
31_SmokingFunnel
32_Siesta
PATTERN
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
PctTotalAttrInsWithTableIns


Pct Attribute Injected
11_FlatLiner
12_RadicalSign
13_Sigmoid
14_LateRiser
21_QuantumSteps
22_RegularlyCurated
31_SmokingFunnel
32_Siesta
PATTERN
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
PctTotalAttrInjected


Pct Attribute Deletions with Table Deletions
11_FlatLiner
12_RadicalSign
13_Sigmoid
14_LateRiser
21_QuantumSteps
22_RegularlyCurated
31_SmokingFunnel
32_Siesta
PATTERN
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
PctTotalAttrbDelWithTableDel


Pct Attribute Ejected
11_FlatLiner
12_RadicalSign
13_Sigmoid
14_LateRiser
21_QuantumSteps
22_RegularlyCurated
31_SmokingFunnel
32_Siesta
PATTERN
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
PctTotalAttrEjected


Pct Attribute Data Type Update
11_FlatLiner
12_RadicalSign
13_Sigmoid
14_LateRiser
21_QuantumSteps
22_RegularlyCurated
31_SmokingFunnel
32_Siesta
PATTERN
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
PctTotalAttrWithTypeUpd


Pct Attribute PK Change
11_FlatLiner
12_RadicalSign
13_Sigmoid
14_LateRiser
21_QuantumSteps
22_RegularlyCurated
31_SmokingFunnel
32_Siesta
PATTERN
-0.2
0
0.2
0.4
0.6
0.8
PctTotalAttrInPKUpd

7 Detailed Data

PATTERN	#Change Types	% Attr Ins w. Tbl Ins	% Attr Del w. Tbl Del	% Attr Inj.	% Attr Ej.	% Attr Type Upd	% Attr PK Upd	% Expansion	% Maintenance	Attr Ins w. Tbl Ins	Attr Del w. Tbl Del	Attr Inj.	Attr Ej.	Attr Type Upd	Attr PK Upd	Expansion	Maintenance
FlatLiner	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	1	0%	0%	0%	100%	0%	0%	0%	100%	0	0	0	1	0	0	0	1
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	2	0	0	2
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	3	0	0	3
	1	0%	100%	0%	0%	0%	0%	0%	100%	0	3	0	0	0	0	0	3
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	7	0	0	7
	4	53%	32%	5%	0%	11%	0%	58%	42%	10	6	1	0	2	0	11	8
	5	50%	29%	13%	4%	4%	0%	63%	38%	12	7	3	1	1	0	15	9
	4	15%	0%	54%	8%	23%	0%	69%	31%	2	0	7	1	3	0	9	4
	2	0%	0%	75%	0%	0%	25%	75%	25%	0	0	3	0	0	1	3	1
	4	38%	15%	38%	8%	0%	0%	77%	23%	5	2	5	1	0	0	10	3
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
12_RadicalSig	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	9	0	0	9
	4	0%	0%	14%	14%	19%	53%	14%	86%	0	0	10	10	14	38	10	62
	4	0%	30%	20%	10%	40%	0%	20%	80%	0	3	2	1	4	0	2	8
	5	10%	10%	23%	23%	35%	0%	32%	68%	3	3	7	7	11	0	10	21
	5	31%	58%	6%	3%	3%	0%	36%	64%	11	21	2	1	1	0	13	23
	3	0%	0%	39%	50%	11%	0%	39%	61%	0	0	7	9	2	0	7	11
	5	39%	61%	0%	0%	0%	0%	39%	61%	489	767	4	1	6	0	493	774
	5	33%	33%	11%	11%	11%	0%	44%	56%	3	3	1	1	1	0	4	5
	5	41%	45%	5%	5%	5%	0%	45%	55%	9	10	1	1	1	0	10	12
	6	31%	39%	14%	11%	3%	2%	46%	54%	169	207	76	61	15	9	245	292
	4	10%	0%	38%	14%	38%	0%	48%	52%	2	0	8	3	8	0	10	11

PATTERN	#Change Types	% Attr Ins w. Tbl Ins	% Attr Del w. Tbl Del	% Attr Inj.	% Attr Ej.	% Attr Type Upd	% Attr PK Upd	% Expansion	% Maintenance	Attr Ins w. Tbl Ins	Attr Del w. Tbl Del	Attr Inj.	Attr Ej.	Attr Type Upd	Attr PK Upd	Expansion	Maintenance
	5	46%	46%	4%	2%	2%	0%	50%	50%	57	57	5	3	3	0	62	63
	4	47%	47%	3%	0%	2%	0%	50%	50%	134	134	8	0	7	0	142	141
	6	43%	17%	10%	4%	20%	4%	54%	46%	30	12	7	3	14	3	37	32
	2	54%	46%	0%	0%	0%	0%	54%	46%	7	6	0	0	0	0	7	6
	2	0%	0%	60%	0%	40%	0%	60%	40%	0	0	3	0	2	0	3	2
	5	59%	37%	3%	1%	1%	0%	62%	38%	117	74	6	1	1	0	123	76
	5	39%	18%	25%	12%	7%	0%	64%	36%	47	22	30	14	8	0	77	44
	3	50%	31%	19%	0%	0%	0%	69%	31%	8	5	3	0	0	0	11	5
	4	59%	0%	11%	8%	22%	0%	70%	30%	22	0	4	3	8	0	26	11
	2	74%	26%	0%	0%	0%	0%	74%	26%	23	8	0	0	0	0	23	8
	3	56%	0%	22%	22%	0%	0%	78%	22%	15	0	6	6	0	0	21	6
	3	78%	0%	6%	0%	17%	0%	83%	17%	14	0	1	0	3	0	15	3
	4	48%	0%	41%	4%	7%	0%	89%	11%	13	0	11	1	2	0	24	3
	2	91%	0%	0%	0%	9%	0%	91%	9%	10	0	0	0	1	0	10	1
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	2	0	0	0	2	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	2	0	0	0	2	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	3	0	0	0	3	0
	2	80%	0%	20%	0%	0%	0%	100%	0%	4	0	1	0	0	0	5	0
	1	100%	0%	0%	0%	0%	0%	100%	0%	5	0	0	0	0	0	5	0
	1	100%	0%	0%	0%	0%	0%	100%	0%	9	0	0	0	0	0	9	0
	2	86%	0%	14%	0%	0%	0%	100%	0%	12	0	2	0	0	0	14	0
13_Sigmoid	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	2	0	0	2
	5	4%	91%	1%	2%	1%	0%	6%	94%	17	348	5	8	5	0	22	361
	2	0%	0%	50%	0%	50%	0%	50%	50%	0	0	1	0	1	0	1	1
	5	42%	35%	14%	7%	1%	0%	56%	44%	62	52	21	11	2	0	83	65
	2	0%	0%	67%	33%	0%	0%	67%	33%	0	0	6	3	0	0	6	3
	4	48%	0%	29%	14%	10%	0%	76%	24%	10	0	6	3	2	0	16	5
	4	15%	0%	62%	8%	0%	15%	77%	23%	2	0	8	1	0	2	10	3
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	4	0	0	0	4	0
	1	100%	0%	0%	0%	0%	0%	100%	0%	6	0	0	0	0	0	6	0

PATTERN	#Change Types	% Attr Ins w. Tbl Ins	% Attr Del w. Tbl Del	% Attr Inj.	% Attr Ej.	% Attr Type Upd	% Attr PK Upd	% Expansion	% Maintenance	Attr Ins w. Tbl Ins	Attr Del w. Tbl Del	Attr Inj.	Attr Ej.	Attr Type Upd	Attr PK Upd	Expansion	Maintenance
	2	73%	0%	27%	0%	0%	0%	100%	0%	8	0	3	0	0	0	11	0
14_LateRiser	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	0	0%	0%	0%	0%	0%	0%	0%	0%	0	0	0	0	0	0	0	0
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	2	0	0	2
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	3	0	0	3
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	3	0	0	3
	2	0%	0%	40%	0%	0%	60%	40%	60%	0	0	2	0	0	3	2	3
	4	47%	0%	6%	0%	44%	3%	53%	47%	17	0	2	0	16	1	19	17
	3	61%	34%	0%	0%	5%	0%	61%	39%	68	38	0	0	6	0	68	44
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	1	0	0	0	1	0
	2	80%	0%	20%	0%	0%	0%	100%	0%	4	0	1	0	0	0	5	0
	2	70%	0%	30%	0%	0%	0%	100%	0%	7	0	3	0	0	0	10	0
21_Quantum	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	2	0	0	2
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	2	0	0	2
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	3	0	0	3
	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	4	0	0	4
	3	0%	0%	14%	9%	76%	0%	14%	86%	0	0	25	16	133	0	25	149
	3	11%	0%	9%	0%	80%	0%	20%	80%	13	0	10	0	94	0	23	94
	5	0%	17%	33%	8%	25%	17%	33%	67%	0	2	4	1	3	2	4	8
	5	22%	22%	17%	7%	31%	0%	39%	61%	12	12	9	4	17	0	21	33
	5	32%	29%	14%	12%	14%	0%	45%	55%	23	21	10	9	10	0	33	40
	5	45%	41%	1%	0%	12%	0%	46%	54%	142	129	4	1	39	0	146	169
	2	50%	50%	0%	0%	0%	0%	50%	50%	2	2	0	0	0	0	2	2
	5	48%	27%	12%	9%	3%	0%	61%	39%	16	9	4	3	1	0	20	13
	4	46%	0%	16%	3%	35%	0%	62%	38%	17	0	6	1	13	0	23	14
	5	39%	10%	24%	18%	8%	0%	63%	37%	84	22	51	39	18	0	135	79
	2	0%	0%	67%	33%	0%	0%	67%	33%	0	0	2	1	0	0	2	1
	4	47%	0%	24%	6%	24%	0%	71%	29%	8	0	4	1	4	0	12	5
	4	48%	17%	28%	7%	0%	0%	76%	24%	14	5	8	2	0	0	22	7
	4	66%	0%	20%	13%	1%	0%	86%	14%	46	0	14	9	1	0	60	10
	3	48%	0%	48%	0%	4%	0%	96%	4%	11	0	11	0	1	0	22	1
	1	0%	0%	100%	0%	0%	0%	100%	0%	0	0	2	0	0	0	2	0
	1	100%	0%	0%	0%	0%	0%	100%	0%	3	0	0	0	0	0	3	0
	1	100%	0%	0%	0%	0%	0%	100%	0%	6	0	0	0	0	0	6	0
	2	92%	0%	8%	0%	0%	0%	100%	0%	11	0	1	0	0	0	12	0
22_Regularly	5	27%	44%	9%	3%	16%	0%	36%	64%	94	156	33	11	58	0	127	225
	5	20%	37%	20%	19%	5%	0%	39%	61%	60	111	60	57	16	0	120	184
	5	31%	19%	14%	12%	24%	0%	45%	55%	13	8	6	5	10	0	19	23
	5	31%	26%	19%	13%	11%	0%	51%	49%	457	379	278	184	154	0	735	717

PATTERN	#Change Types	% Attr Ins w. Tbl Ins	% Attr Del w. Tbl Del	% Attr Inj.	% Attr Ej.	% Attr Type Upd	% Attr PK Upd	% Expansion	% Maintenance	Attr Ins w. Tbl Ins	Attr Del w. Tbl Del	Attr Inj.	Attr Ej.	Attr Type Upd	Attr PK Upd	Expansion	Maintenance
	6	46%	34%	6%	4%	9%	0%	52%	48%	1774	1321	240	171	331	8	2014	1831
	5	34%	15%	22%	17%	13%	0%	56%	44%	30	13	19	15	11	0	49	39
	5	43%	27%	14%	10%	6%	0%	57%	43%	306	195	96	68	45	0	402	308
	5	40%	24%	22%	8%	7%	0%	61%	39%	97	59	53	19	17	0	150	95
	4	57%	0%	9%	2%	32%	0%	66%	34%	85	0	13	3	48	0	98	51
	5	54%	24%	13%	7%	2%	0%	67%	33%	117	51	29	15	5	0	146	71
	4	60%	0%	11%	11%	19%	0%	71%	29%	68	0	12	12	21	0	80	33
	4	29%	9%	53%	0%	9%	0%	82%	18%	10	3	18	0	3	0	28	6
	5	66%	10%	18%	4%	2%	0%	84%	16%	168	26	45	11	5	0	213	42
	4	72%	0%	16%	2%	11%	0%	87%	13%	266	0	58	6	42	0	324	48
31_SmokingF	6	38%	24%	5%	2%	22%	9%	43%	57%	236	152	32	11	138	54	268	355
	2	50%	50%	0%	0%	0%	0%	50%	50%	95	95	0	0	0	0	95	95
	5	22%	15%	33%	22%	8%	0%	55%	45%	56	39	83	55	20	0	139	114
	5	16%	16%	42%	13%	13%	0%	58%	42%	6	6	16	5	5	0	22	16
	5	32%	32%	26%	7%	4%	0%	58%	42%	18	18	15	4	2	0	33	24
	4	0%	0%	65%	6%	21%	9%	65%	35%	0	0	22	2	7	3	22	12
	6	57%	10%	17%	11%	4%	1%	74%	26%	107	19	33	21	7	2	140	49
32_Siesta	1	0%	0%	0%	0%	100%	0%	0%	100%	0	0	0	0	2	0	0	2
	3	0%	57%	29%	0%	14%	0%	29%	71%	0	8	4	0	2	0	4	10
	5	22%	0%	14%	3%	58%	3%	36%	64%	8	0	5	1	21	1	13	23
	3	0%	0%	42%	8%	50%	0%	42%	58%	0	0	10	2	12	0	10	14
	4	45%	45%	0%	5%	0%	5%	45%	55%	9	9	0	1	0	1	9	11
	6	36%	36%	10%	8%	6%	3%	47%	53%	42	42	12	9	7	4	54	62
	5	47%	46%	0%	0%	6%	0%	47%	53%	169	165	1	1	22	0	170	188
	4	27%	0%	36%	0%	9%	27%	64%	36%	3	0	4	0	1	3	7	4
	3	50%	0%	17%	33%	0%	0%	67%	33%	3	0	1	2	0	0	4	2
	4	63%	0%	9%	3%	25%	0%	72%	28%	20	0	3	1	8	0	23	9