PxEdit 4.0

Structural tables

Principles

Principles

- The structural table is the way to move information from other applications to PxEdit (and further to px format)
- They include only the necessary information:
  - table title
  - variable names
  - value names for each variable
  - data matrix
- Most statistical tables can easily be converted to structural tables
- The table may be a spreadsheet or a well structural text file

Good to know

- PxEdit also accepts
  - asymmetrical tables
    (i.e. tables with missing information or tables with disorder)
    - *Fill Item* is used for the missing information
  - redundant tables
    - the user will be warned about duplicated values
    - the last value will be used
- Possible problems
  - there are notes below the table (especially in Excel sheets)
  - title cells that should be empty are not
    - the tabulation recognition algorithm is based on empty areas

Short description

- The first line shows the table heading
  - the language code may be given next to the heading
- The next lines contain the column variable names in the first column
- The next line contains the row variable names side by side
  - column or even row variables may be missing
- Row variable value names are in their own columns
- Column variable value names are in their own rows
- Data are represented as a table, every figure positioned by its headers

Layouts

Example 1

Married couples with children							*table title*
Year		1992		...			*column variables*
Age		-19	20-24	25-29	30-34	35-39 ...	
Area	**Sex**						
Akaa	Males	1	20	108	238	379	...
	Females	6	33	191	305	406	
Alajärvi	Males	0	21	107	237	291	
	Females	3	45	188	272	304	
Alavieska	Males	0	4	27	45	75	
	Females	1	7	45	66	69	
Alavus	Males	0	14	102	227	345	
	Females	2	51	170	273	347	
Asikkala	Males	0	8	42	111	174	
	Females	0	13	74	150	229	
Askola	Males	0	2	26	94	99	
	Females	0	7	53	106	131	
Aura	Males	0	3	32	52	91	
...	...	...					
row variables							*data part*

Statistics Finland

Shortcuts

- Variable codes may be given in their own rows or columns (i.e. the general structural table)
  - they share the same variable name
  - the code must be given first
  - table indexing is based on value-code combinations
- There may be empty columns (not in text files) and single empty rows
  - when reading Excel tables or tables from clipboard two successive empty rows stop the table reading, because they can be used to separate the table from its metadata part
- The second cell after the heading may contain the language code

Example 2

Married couples with children *table title*

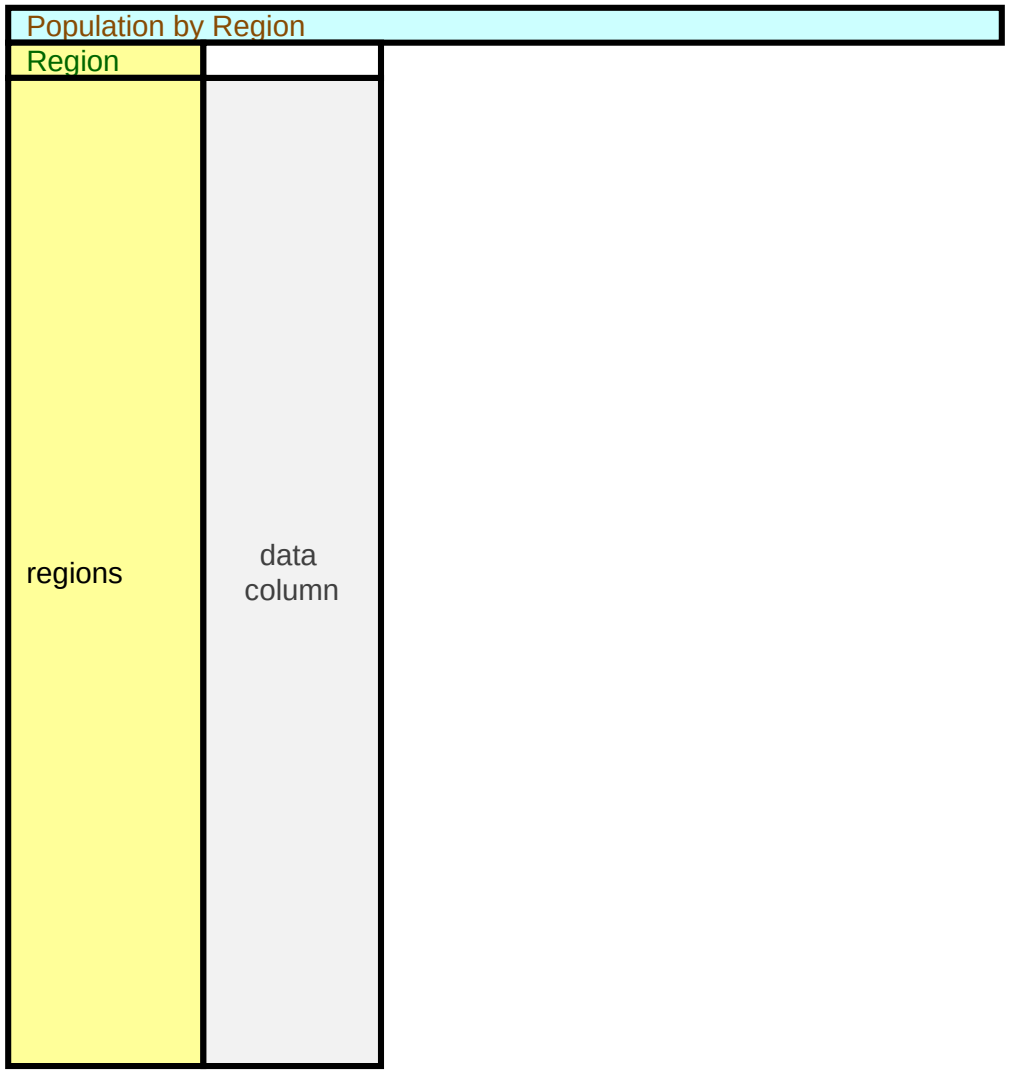
Year			1991		...	*column variables*
Year			1991		...	
Age			-19	20-24	25-29	30-34 ...
Area	**Area**	**Sex**				
020	Akaa	Males	1	20	108	238 ...
		Females	6	33	191	305
005	Alajärvi	Males	0	21	107	237
		Females	3	45	188	272
009	Alavieska	Males	0	4	27	45
		Females	1	7	45	66
010	Alavus	Males	0	14	102	227
		Females	2	51	170	273
016	Asikkala	Males	0	8	42	111
		Females	0	13	74	150
018	Askola	Males	0	2	26	94
...	...	...	...			

row variables *data part*

Example 2 in Excel with language code

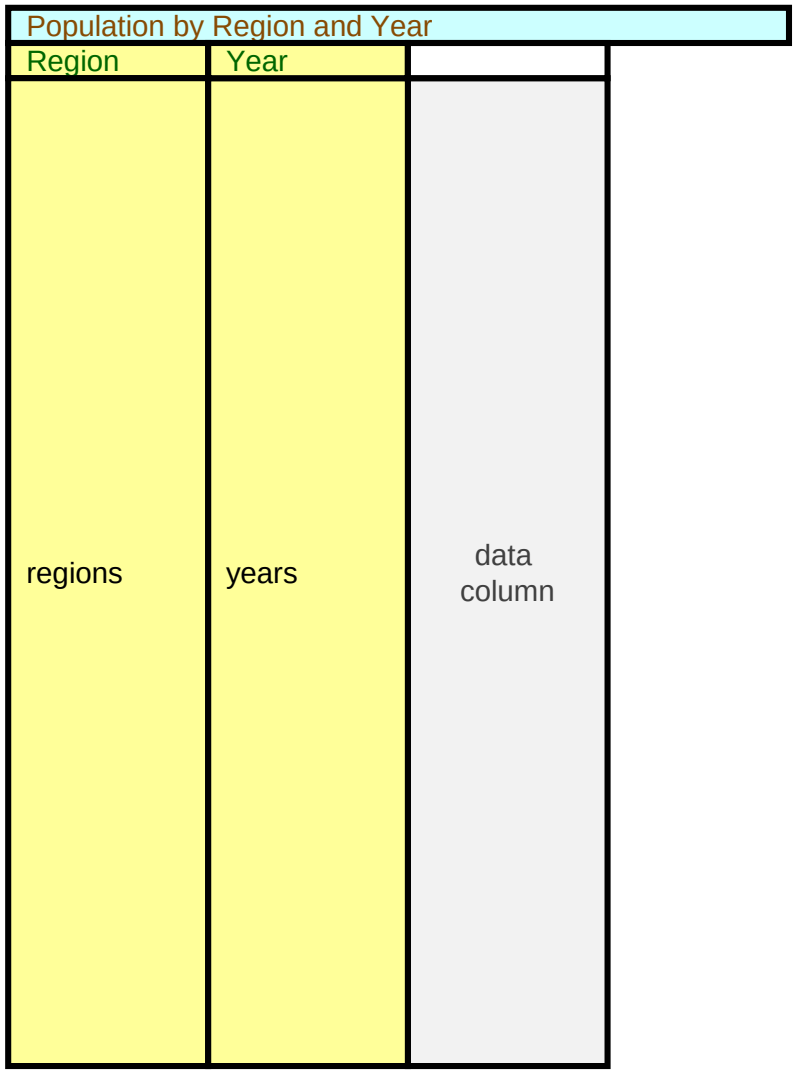
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Marrie	en														
2	Year			1992												
3	Year			1992												
4	Age			- 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74	75 -
5	Area	Area	Sex													
6	020	Akaa	Males	1	20	108	238	379	452	329	186	110	73	53	27	15
7			Females	6	33	191	305	406	429	266	146	95	58	29	21	6
8	005	Alajärvi	Males	0	21	107	237	291	363	242	160	128	92	60	34	25
9			Females	3	45	188	272	304	332	229	140	91	79	51	18	8
10	009	Alavieska	Males	0	4	27	45	75	76	50	32	30	27	16	6	9
11			Females	1	7	45	66	69	68	49	24	30	23	5	7	3
12	010	Alavus	Males	0	14	102	227	345	374	271	215	130	119	51	30	31
13			Females	2	51	170	273	347	339	272	172	117	96	42	14	14
14	016	Asikkala	Males	0	8	42	111	174	276	209	145	65	58	17	16	12
15			Females	0	13	74	150	229	241	208	97	61	26	13	14	7
16	018	Askola	Males	0	2	26	94	99	136	108	58	40	28	13	10	5
17			Females	0	7	53	106	131	120	84	53	29	16	11	6	3
18	019	Aura	Males	0	3	32	52	91	93	86	30	18	11	11	5	4
19			Females	0	10	50	75	95	94	47	29	14	12	2	5	3
20	035	Brändö	Males	0	0	0	8	8	3	12	6	9	3	2	1	2
21			Females	0	0	7	4	5	13	9	7	3	3	1	1	1
22	043	Eckerö	Males	0	0	10	7	19	19	13	9	7	5	4	2	3

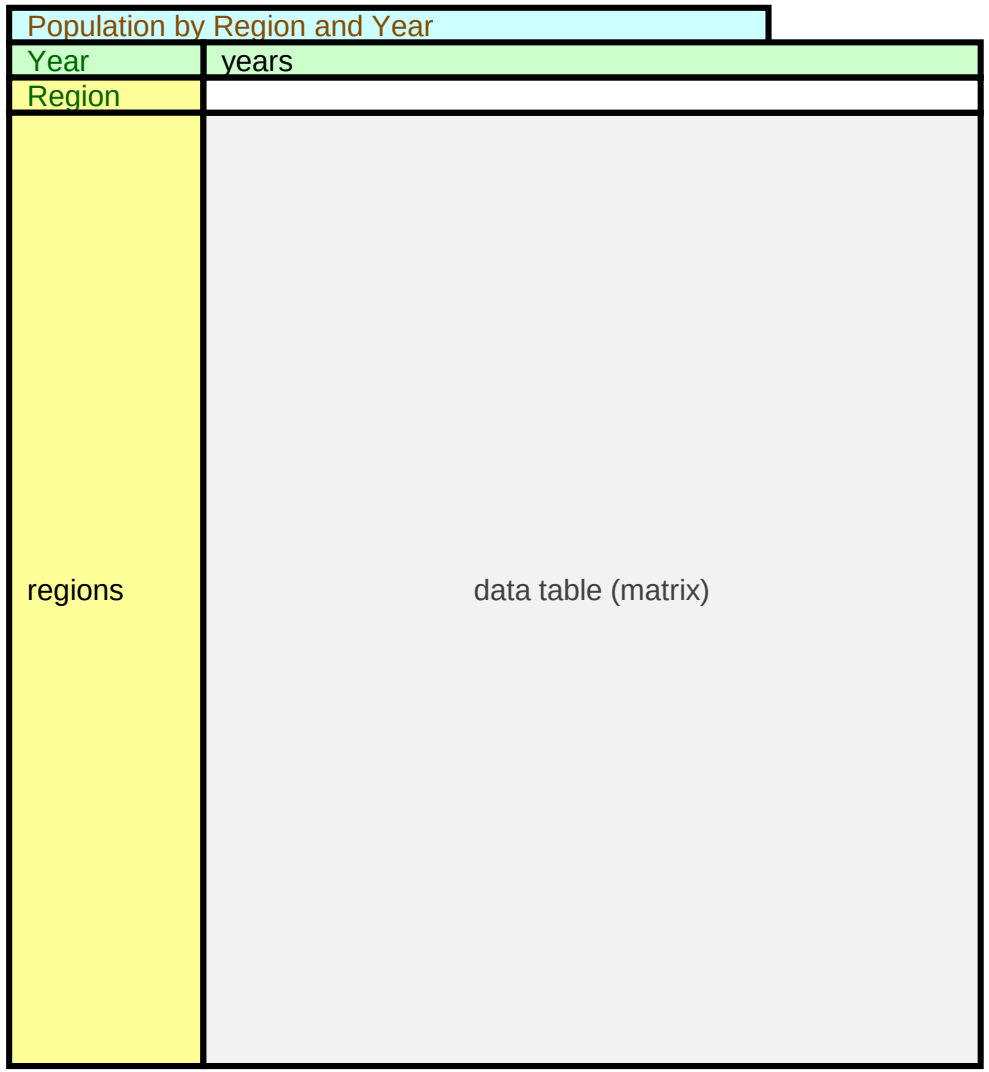
Layout: one variable table



Statistics Finland

Layouts: two variable tables





Statistics Finland

Layouts: three variable tables

Population by Region, Year and Age group			
Region	Year	Age group	
regions	years	age groups	data column

Population by Region, Year and Age group		
Age group		age groups
Region	Year	
regions	years	data table

Population by Region, Year and Age group	
Year	years
Age group	age groups
Region	
regions	data table

Keyword block

Keyword block

- The Excel tables may have a keyword setting block under the table separated by at least two empty rows
- The format of the keyword table is similar to the control csv for PxJob

	A	B	C	D	E	F	G
1	Establishment and personnel						
2	Industry			C		D	
3	Industry			C Mining and quarrying		D Manufacturing	
4	Data			Establishment	Personnel	Establishment	Personnel
5	Year	Region	Region				
6	1995	0	Whole country	1271	3528	26009	380791
7		4	Alahärmä	2	0	42	458
8		5	Alajärvi	6	6	84	548
9	1996	0	Whole country	1409	3961	27442	386001
10		4	Alahärmä	2	0	41	406
11		5	Alajärvi	6	11	88	655
12							
13							
14	UNITS	persons					
15	NOTE	table footnote	variable footnote for Industry				
16	variablename		Industry	Year			
17	VALUENOTE			value footnote for Year 1996			
18	valuetext			1996			
19							
20	CELLNOTE				cell note for year 1995, Alajärvi, establishment and all industries		
21	Industry				*		
22	Data				Establishment		
23	Year				1995		
24	Region				Alajärvi		
25							

Statistics Finland

Keyword block: table-specific keywords

- The table-specific keywords are given in a separate row (or column)
- There may be several rows (or columns)

table-specific keywords in rows			
SOURCE	Statistics Finland		
DATABASE	StatFin		
NOTE	this is a footnote		
..or in columns			
SOURCE	**DATABASE**	**NOTE**	
Statistics Finland	StatFin	this is a footnote	

Keyword block: variable-specific keywords

- The variable-specific keywords are given in separate rows (or columns)
- The corresponding variable names are needed, too

variable-specific keywords in rows			
NOTE	annual footnote	footnote for Area	
DOMAIN		region	
variablename	Year	Area	
..or in columns			
NOTE	**DOMAIN**	variablename	
annual footnote		Year	
footnote for Area	region	Area	

Statistics Finland

Keyword block: value-specific keywords

- The value-specific keywords are given in separate rows (or columns)
- The corresponding variable names and values are needed, too

value-specific keywords			
VALUENOTE	happy new year	northmost capital	
variablename	Year	Region	
valuetext	2019	Reykjavik	

Keyword block: cell-specific keywords

- The cell-specific keywords are given in separate rows or columns
- The corresponding variable names and values are needed, too
  - * means all values

cell-specific keyword syntax (in rows):				
CELLNOTE	footnote for year 2019 and Finland			
Age	*			
Year	2019			
Country	Finland			

Keyword block: combined levels

- The different levels should be given in separate rows (or columns)

SOURCE	Statistics Finland								
DATABASE	StatFin								
NOTE	this is a fo	annual footnote	footnote for Area						
DOMAIN			region						
variablename		Year	Area						
VALUENOTE				happy new year	northmost capital				
variablename				Year	Area				
valuetext				2019	Reykjavik				
CELLNOTE						footnote for year 2019 and Finland			
Age						*			
Year						2019			
Country						Finland			

Statistics Finland

Multilingual tables

Multilingual tables

- The structural table may have multiple languages
  - the languages must be given in the second cell after the title as a comma-separated list
  - the first language code is the base language
  - the value text columns and rows are given in the same language order and they share the same variable name in the base language
- The variable names for different languages, as well as the table titles are given in the keyword block under the table with keywords *VARIABLES* (or *STUB* and *HEADING*) and *CONTENTS*
- The repeated value texts (such as years) may be given only once
  - they have to be repeated if there are code columns or rows present

Multilingual example

	A	B	C	D	E	F	G	H	I	J
1	**Suomen kansalaisuuden saaneet**	fi,sv,en								
2	**Sukupuoli**				Yhteensä		Miehet		Naiset	
3	**Sukupuoli**				Totalt		Män		Kvinnor	
4	**Sukupuoli**				Total		Males		Females	
5	**Vuosi**				2017	2018	2017	2018	2017	2018
6	Syntymävaltio	**Syntymävaltio**	**Syntymävaltio**	**Syntymävaltio**						
7	SSS	Yhteensä	Totalt	Total	12219	9211	5844	4335	6375	4876
8	EUR	Eurooppa	Europa	Europe	4652	3309	1944	1423	2708	1886
9	AFR	Afrikka	Afrika	Africa	1844	1480	1008	806	836	674
10	AME	Amerikka	Amerika	America	279	223	123	102	156	121
11	AAS	Aasia	Asien	Asia	3138	2429	1578	1110	1560	1319
12	OSE	Oseania	Oceanien	Oceania	10	10	9	6	1	4
13										
14										
15	languagecode	**sv**	**en**	**sv**	**en**					
16	VARIABLES	Födelseland,Kön,År,Uppgifter	Country of birth,Sex,Year,Information							
17	CONTENTS			**Personer som erhållt finskt medborgarskap**	**Citizenships granted**					
18										

Hierarchy levels

Hierarchy levels

- The hierarchy levels of one variable may be given in a separate column or row before the corresponding code information with the same variable name
- The hierarchy levels are given as integers starting from **0** (the base level), and each sub-level is shown with a level number greater than its mother level
- If the variable codes show enough level information, they can be used, too (but that is not generally recommended)

Hierarchy level example

	A	B	C	D	E	F	G	H	I	J
1	*Consumer Price Index*									
2	**Period**			2005M08	2005M09	2005M10	2005M11	2005M12	2006M01	2006M02
3	Product	Product	**Product**							
4	0	tot	All products	279,95	281,87	282,37	281,69	281,82	279,59	280,9
5	1	01	01 Food and non-alcoholic beverages	236,97	236,57	235,89	235,94	235,94	236,96	238,58
6	2	01.1	01.1 Food	245,9	245,45	244,73	244,86	244,91	246,09	247,92
7	3	01.1.1	01.1.1 Bread and cereals	247,81	246,78	245,84	245,98	245,15	245,29	243,62
8	3	01.1.2	01.1.2 Meat	230,81	230,37	229,03	230,18	231,09	229,29	232,2
9	3	01.1.3	01.1.3 Fish	320,31	317,93	314,36	311,79	308,46	316,3	314,92
10	3	01.1.4	01.1.4 Milk, cheese and eggs	266,5	266,3	266,31	266,86	267,11	266,87	267,35
11	3	01.1.5	01.1.5 Oils and fats	211,46	210,79	210,6	210,83	210,27	208,83	211,21
12	3	01.1.6	01.1.6 Fruit	221,3	224,34	222,01	227,26	233,14	232,19	227,03
13	3	01.1.7	01.1.7 Vegetables	191,96	190,89	192,6	189,47	187,17	196,45	209,49
14	3	01.1.8	01.1.8 Sugar, jam, honey, chocolate and confectionery	292,48	292,6	291,55	290,57	289,83	290,1	290,95
15	3	01.1.9	01.1.9 Food products n.e.c.	267,12	264,16	263,68	263,11	263,57	262,74	262,17
16	2	01.2	01.2 Non alcoholic beverages	172,36	172,27	171,85	171,41	171,1	170,96	171,16
17	3	01.2.1	01.2.1 Coffee, tea and cocoa	153,85	152,76	152,28	151,97	152,23	151,57	153,05
18	3	01.2.2	01.2.2 Cold non-alcoholic beverages	223,06	223,57	223,09	222,47	221,74	221,98	221,38
19	1	02	02 Alcoholic beverages and tobacco	352,84	353,03	353,14	353,15	353,35	354,1	355,62
20	2	02.1	02.1 Alcoholic beverages	272,44	272,24	272,31	272,21	272,37	272,12	272,3
21	3	02.1.1	02.1.1 Spirits	290,8	290,88	291,32	291,22	291,23	291,49	291,48
22	3	02.1.2	02.1.2 Wine	285,38	285,48	285,65	285,65	285,65	285,69	285,61
23	3	02.1.3	02.1.3 Beer	240,39	239,59	239,23	239,02	239,49	238,34	238,97
24	2	02.2	02.2 Tobacco	503,7	504,87	505,09	505,37	505,65	507,14	511,58
25	1	03	03 Clothing and footwear	149,69	169,97	173,16	172,79	170,32	150,78	154,35

Sequential files

Sequential files

- Mask control can be used to transfer sequential files containing fixed-size records obtained from e.g. some legacy systems
  - all the information is given in columns (i.e. sequential records)
  - the record mask is given in the second row
  - the record mask uses the same character for one record
  - mask spaces will skip the corresponding columns
  - the data part is the last column, it needs only one marker (because the data part length usually varies)
  - the variable names and codes have to be changed afterwards

Column mask example

Statistical table	*table title*
aaabbbbc d	*column mask*
00019901;4847 00019902;583676 00019903;488 00019904;113653 00019905;93 00019906;1443 00019907;279 00019908;0 00019911;4823 00019912;586586 00019913;489 00019914;123399 00019915;93 00019916;1427 00019917;279 ...	*code columns, data column at the end*

Cell-specific information

DATANOTECELL input

- It is possible to attach the *DATANOTECELL* text strings with the corresponding figures or dot codes in the structural tables
- The possible strings must be given in the main settings file:

```
[System]
datanotecells=
```

  - comma separated list, may have leading or trailing blanks, may be enclosed with double quotes
- Strings will be recognised without any spaces, but they will be added to the table metadata with the same formatting as in the settings file

DATANOTECELL example

- If there is the following line in the settings file:

```
datanotecells=*,**,"   ***"
```

then the input table may contain such figures or cells as:

`123**`	123 with '**' (quotes are for visibility only)
`3.456 *`	3.456 with '*'
`..***`	dot code with '***'

Cell-specific footnotes

- The *Separator characters* for different types of footnotes is an alternative to *datanotecells*

- If one uses separator characters that are not the defaults, they must be configured in the main settings file:

```
[System]
separator_attribute=          (default |)
separator_cellnote=           (default !)
separator_datanote=           (default #)
```

- Footnote strings may now be attached to the cell figure with the separator character

Separator examples

- Separators enable multiple footnotes for a single cell, even with dot codes
  - strings containing special characters (such as spaces, commas or separators) should be enclosed in quotes
- For example the following input will be recognised (the cell separator is `;` and the other separators are coloured just for clarity)

```
1;2|A,B;3.14!"Just a comment"#a);4;..|A,C; …
```

- NB: PxEdit will copy the Excel cell comments to the corresponding px keywords