An Introduction to Tablet for PDF

Tim Bergsma

2025-12-22

Contents

Motivation

Occasionally it is useful to generate a table of summary statistics for rows of a dataset, where such rows represent sampling units and and columns may be categorical or continuous. The excellent R package table1 does exactly this, and was the inspiration for `tablet`. `table1` however is optimized for html; `tablet` tries to provide a format-neutral implementation and relies on kableExtra to handle the rendering. Support for pdf (latex) is of particular interest, and is illustrated here. See the companion vignette for a proof-of-concept html implementation.

Software

To support our examples, we load some other packages and in particular locate the melanoma dataset from boot. By the way, in the yaml header for the Rmd source file, we've added the header-includes as described on p. 4 of the `kableExtra` documentation.

```
library(tidyr)
library(dplyr)
library(magrittr)
library(kableExtra)
library(boot)
```

```
library(yamlet)
library(tablet)
```

```
x <- melanoma
x %<>% select(-time, -year)
```

Simple Case

For starters, we'll just coerce two variables to factor to show that they are categorical, and then pass the whole thing to tablet(). Then we forward to as_kable() for rendering (calls kableExtra::kbl and adds some magic).

```
x %>%
  mutate(
    sex = factor(sex),
    ulcer = factor(ulcer)
  ) %>%
  tablet %>%
  as_kable
```

	All (N = 205)
status	
Mean (SD)	1.79 (0.551)
Median (range)	2 (1, 3)
Missing	0
sex	
0	126 (61.5%)
1	79 (38.5%)
age	
Mean (SD)	52.5 (16.7)
Median (range)	54 (4, 95)
Missing	0
thickness	
Mean (SD)	2.92 (2.96)
Median (range)	1.94 (0.1, 17.4)
Missing	0
ulcer	
0	115 (56.1%)
1	90 (43.9%)

With Metadata

Now we redefine the dataset, supplying metadata almost verbatim from `?melanoma`. This is fairly easy using package `yamlet`. Note that we reverse the authors' factor order of 1, 0 for ulcer and move status 'Alive' to first position.

```
x <- melanoma

x %<>% decorate('
time:       [ Survival Time Since Operation, day ]
```

```
status:
 - End of Study Patient Status
 -
  - Alive: 2
  - Melanoma Death: 1
  - Unrelated Death: 3
sex:       [ Sex, [ Male: 1, Female: 0 ]]
age:       [ Age at Time of Operation, year ]
year:      [ Year of Operation, year ]
thickness: [ Tumor Thickness, mm ]
ulcer:     [ Ulceration, [ Absent: 0, Present: 1 ]]
')
x %<>% select(-time, -year)
x %<>% group_by(status)
x %<>% resolve
```

- group_by(status) causes statistics to be summarized in columns by group.
- resolve() disambiguates labels, units, and factor levels (actually creating factors where appropriate, such as for sex and ulcer).

Now we pass x to tablet() and as_kable() for a more informative result.

```
x %>% tablet %>% as_kable
```

	Alive (N = 134)	Melanoma Death (N = 57)	Unrelated Death (N = 14)	All (N = 205)
Sex				
Male	43 (32.1%)	29 (50.9%)	7 (50%)	79 (38.5%)
Female	91 (67.9%)	28 (49.1%)	7 (50%)	126 (61.5%)
Age at Time of Operation (year)				
Mean (SD)	50 (15.9)	55.1 (17.9)	65.3 (10.9)	52.5 (16.7)
Median (range)	52 (4, 84)	56 (14, 95)	65 (49, 86)	54 (4, 95)
Missing	0	0	0	0
Tumor Thickness (mm)				
Mean (SD)	2.24 (2.33)	4.31 (3.57)	3.72 (3.63)	2.92 (2.96)
Median (range)	1.36 (0.1, 12.9)	3.54 (0.32, 17.4)	2.26 (0.16, 12.6)	1.94 (0.1, 17.4)
Missing	0	0	0	0
Ulceration				
Absent	92 (68.7%)	16 (28.1%)	7 (50%)	115 (56.1%)
Present	42 (31.3%)	41 (71.9%)	7 (50%)	90 (43.9%)

Notice that:

- the order of variables down the left side is exactly their order in the dataset;
- the order of factor levels is exactly that in x;
- the order of groups across the top is exactly the levels (if any) of the grouping variable(s), and
- labels and titles (highest priority) are substituted for item names.

If you don't particularly care for some aspect of the presentation, you can jump in between tablet() and as_kable() to fix things up. For example, if you don't want the "All" column you can just say

- `x %>% tablet %>% select(-All) %>% as_kable`.

If you **only** want the the "All" column, you can just remove the group(s):

- `x %>% ungroup %>% select(-1) %>% tablet %>% as_kable`.

By the way, you can also pass `all = NULL` to suppress the 'All' column.

As Xtable

Some support is provided for 'xtable'. Currently, grouped columns (see next section) are not supported.

```
library(xtable)
  x %>%
    filter(!(status == 'Alive' & sex == 'Male')) %>%
    tablet %>% as_xtable(format_value = function(x,...)x) %>%
    print(
      booktabs = TRUE,
      include.rownames = FALSE
    )
```

	Alive (N = 91)	Melanoma Death (N = 57)	Unrelated Death (N = 14)	All (N = 162)
Sex				
Male		29 (50.9%)	7 (50%)	36 (22.2%)
Female	91 (100%)	28 (49.1%)	7 (50%)	126 (77.8%)
Age at Time of Operation (year)				
Mean (SD)	48.8 (15.4)	55.1 (17.9)	65.3 (10.9)	52.4 (16.7)
Median (range)	49 (4, 77)	56 (14, 95)	65 (49, 86)	54 (4, 95)
Missing	0	0	0	0
Tumor Thickness (mm)				
Mean (SD)	2.02 (2.22)	4.31 (3.57)	3.72 (3.63)	2.97 (3.08)
Median (range)	1.29 (0.1, 12.9)	3.54 (0.32, 17.4)	2.26 (0.16, 12.6)	1.94 (0.1, 17.4)
Missing	0	0	0	0
Ulceration				
Absent	68 (74.7%)	16 (28.1%)	7 (50%)	91 (56.2%)
Present	23 (25.3%)	41 (71.9%)	7 (50%)	71 (43.8%)

Grouped Columns

In tablet(), most columns are the consequences of a grouping variable. Not surprisingly, grouped columns are just a consequence of nested grouping variables. To illustrate, we follow the table1 vignette by adding a grouping variable that groups the two kinds of death.

```
x %<>% mutate(class = status)                          # copy the current group
x %<>% modify(class, label = 'class')                  # change its label
levels(x$status) <- c('Alive','Melanoma','Unrelated')  # tweak current group
levels(x$class)  <- c(' ',    'Death',   'Death')      # cluster groups
x %<>% group_by(class, status)                         # nest groups
x %>% tablet %>% as_kable                              # render
```

	Alive		Melanoma		Unrelated		
	Male (N = 43)	Female (N = 91)	Male (N = 29)	Female (N = 28)	Male (N = 7)	Female (N = 7)	All (N = 205)
Age at Time of Operation (year)							
Mean (SD)	52.5 (16.9)	48.8 (15.4)	53.9 (19.7)	56.4 (16.2)	62.4 (11.2)	68.1 (10.6)	52.5 (16.7)
Median (range)	55 (12, 84)	49 (4, 77)	52 (19, 95)	58 (14, 89)	64 (49, 76)	66 (54, 86)	54 (4, 95)
Missing	0	0	0	0	0	0	0
Tumor Thickness (mm)							
Mean (SD)	2.73 (2.49)	2.02 (2.22)	4.63 (3.47)	3.99 (3.71)	4.83 (4.19)	2.6 (2.84)	2.92 (2.96)
Median (range)	1.62 (0.16, 8.38)	1.29 (0.1, 12.9)	4.04 (0.81, 14.7)	3.14 (0.32, 17.4)	4.84 (0.65, 12.6)	1.45 (0.16, 8.54)	1.94 (0.1, 17.4)
Missing	0	0	0	0	0	0	0
Ulceration							
Absent	24 (55.8%)	68 (74.7%)	8 (27.6%)	8 (28.6%)	4 (57.1%)	3 (42.9%)	115 (56.1%)
Present	19 (44.2%)	23 (25.3%)	21 (72.4%)	20 (71.4%)	3 (42.9%)	4 (57.1%)	90 (43.9%)

		Death		
	Alive (N = 134)	Melanoma (N = 57)	Unrelated (N = 14)	All (N = 205)
Sex				
Male	43 (32.1%)	29 (50.9%)	7 (50%)	79 (38.5%)
Female	91 (67.9%)	28 (49.1%)	7 (50%)	126 (61.5%)
Age at Time of Operation (year)				
Mean (SD)	50 (15.9)	55.1 (17.9)	65.3 (10.9)	52.5 (16.7)
Median (range)	52 (4, 84)	56 (14, 95)	65 (49, 86)	54 (4, 95)
Missing	0	0	0	0
Tumor Thickness (mm)				
Mean (SD)	2.24 (2.33)	4.31 (3.57)	3.72 (3.63)	2.92 (2.96)
Median (range)	1.36 (0.1, 12.9)	3.54 (0.32, 17.4)	2.26 (0.16, 12.6)	1.94 (0.1, 17.4)
Missing	0	0	0	0
Ulceration				
Absent	92 (68.7%)	16 (28.1%)	7 (50%)	115 (56.1%)
Present	42 (31.3%)	41 (71.9%)	7 (50%)	90 (43.9%)

Transposed Groups

Categorical observations (in principle) and grouping variables are all factors, and are thus transposable. To illustrate, we drop the column group above and instead nest sex within status ...

```
x %<>% group_by(status, sex)
x %<>% select(-class)
x %>%
  tablet %>%
  as_kable %>%
  kable_styling(latex_options = 'scale_down')
```

... or nest ulceration within status ...

```
x %<>% group_by(status, ulcer)
x %>%
  tablet %>%
  as_kable %>%
  kable_styling(latex_options = 'scale_down')
```

... or where it makes sense, use multiple levels of nesting.

	Alive		Melanoma		Unrelated		
	Absent (N = 92)	Present (N = 42)	Absent (N = 16)	Present (N = 41)	Absent (N = 7)	Present (N = 7)	All (N = 205)
Sex							
Male	24 (26.1%)	19 (45.2%)	8 (50%)	21 (51.2%)	4 (57.1%)	3 (42.9%)	79 (38.5%)
Female	68 (73.9%)	23 (54.8%)	8 (50%)	20 (48.8%)	3 (42.9%)	4 (57.1%)	126 (61.5%)
Age at Time of Operation (year)							
Mean (SD)	49.3 (15.4)	51.6 (17.1)	54.9 (19.9)	55.1 (17.4)	58.4 (8.66)	72.1 (8.53)	52.5 (16.7)
Median (range)	50 (4, 83)	54.5 (12, 84)	59 (16, 83)	56 (14, 95)	56 (49, 71)	72 (60, 86)	54 (4, 95)
Missing	0	0	0	0	0	0	0
Tumor Thickness (mm)							
Mean (SD)	1.63 (1.93)	3.58 (2.58)	2.7 (3.35)	4.94 (3.5)	2.1 (1.93)	5.34 (4.33)	2.92 (2.96)
Median (range)	1.13 (0.1, 12.9)	3.06 (0.32, 12.2)	1.94 (0.32, 14.7)	4.04 (0.97, 17.4)	1.45 (0.65, 6.12)	4.84 (0.16, 12.6)	1.94 (0.1, 17.4)
Missing	0	0	0	0	0	0	0

	Alive				Melanoma				Unrelated				
	Absent		Present		Absent		Present		Absent		Present		
	Male (N = 24)	Female (N = 68)	Male (N = 19)	Female (N = 23)	Male (N = 8)	Female (N = 8)	Male (N = 21)	Female (N = 20)	Male (N = 4)	Female (N = 3)	Male (N = 3)	Female (N = 4)	All (N = 205)
Age at Time of Operation (year)													
Mean (SD)	50.4 (17)	48.9 (14.9)	55.3 (16.9)	48.7 (17)	55.2 (22.2)	54.6 (18.8)	53.3 (19.2)	57 (15.5)	54.5 (7.14)	63.7 (8.74)	73 (2.65)	71.5 (11.8)	52.5 (16.7)
Median (range)	54 (15, 83)	49 (4, 77)	56 (12, 84)	48 (19, 75)	56 (27, 83)	59 (16, 77)	52 (19, 95)	58 (14, 89)	52.5 (49, 64)	66 (54, 71)	72 (71, 76)	70 (60, 86)	54 (4, 95)
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
Tumor Thickness (mm)													
Mean (SD)	1.47 (1.72)	1.69 (2)	4.32 (2.42)	2.97 (2.59)	3.27 (4.68)	2.14 (1.18)	5.14 (2.86)	4.72 (4.13)	2.42 (2.5)	1.67 (1.14)	8.05 (4.02)	3.3 (3.71)	2.92 (2.96)
Median (range)	0.97 (0.16, 7.09)	1.29 (0.1, 12.9)	3.87 (0.81, 8.38)	1.94 (0.32, 12.2)	1.78 (0.81, 14.7)	2.02 (0.32, 3.56)	4.83 (1.62, 12.9)	3.54 (0.97, 17.4)	1.46 (0.65, 6.12)	1.45 (0.65, 2.9)	6.76 (4.84, 12.6)	2.26 (0.16, 8.54)	1.94 (0.1, 17.4)
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0

```
x %<>% group_by(status, ulcer, sex)
x %>%
  tablet %>%
  as_kable %>%
  kable_styling(latex_options = 'scale_down') # %>% landscape ?
```

Aesthetics

`tablet` tries to give rather exhaustive control over formatting. Much can be achieved by replacing elements of 'fun', 'fac', 'num', and 'lab' (see `?tablet.data.frame`). For finer control, you can replace these entirely. In this example, we will ...

- ignore categoricals (other than groups) by replacing 'fac' with something of length zero,
- drop the 'N =' header material by substituting in 'lab', and
- switch to '(min - max)' instead of '(min, max)'.

```
x %<>% group_by(status)
x %>%
  tablet(
    fac = NULL,
    lab ~ name,
    `Median (range)` ~ med + ' (' + min + ' - ' + max + ')'
  ) %>%
  as_kable
```

	Alive	Melanoma	Unrelated	All
Age at Time of Operation (year)				
Mean (SD)	50 (15.9)	55.1 (17.9)	65.3 (10.9)	52.5 (16.7)
Median (range)	52 (4 - 84)	56 (14 - 95)	65 (49 - 86)	54 (4 - 95)
Missing	0	0	0	0
Tumor Thickness (mm)				
Mean (SD)	2.24 (2.33)	4.31 (3.57)	3.72 (3.63)	2.92 (2.96)
Median (range)	1.36 (0.1 - 12.9)	3.54 (0.32 - 17.4)	2.26 (0.16 - 12.6)	1.94 (0.1 - 17.4)
Missing	0	0	0	0

Note

The default presentation includes "N =" under the header, but also has percent characters in the table. Considerable gymnastics are required to make this work! If you change the defaults you may want to consider the arguments to `?as_kable.tablet`.

- 'linebreaker' anticipates that 'lab' has inserted newlines before "N =".
- 'linebreak' invokes the eponymous kableExtra function internally.
- 'escape' is turned off to protect resulting latex markup.
- 'escape_latex' is supplied to restore the usual escaping of percent characters.

Conclusion

`tablet` gives a flexible way of summarizing tables of observations. It reacts to numeric columns, factors, and grouping variables. Display order derives from the order of columns and factor levels in the data. Result columns can be grouped arbitrarily deep by supplying extra groups. Column labels and titles are respected. Rendering is largely the responsibility of `kableExtra` and can be extended. Further customization is possible by manipulating data after calling tablet() but before calling as_kable(). Powerful results are possible with very little code.