Tables in knitr

R.W. Oldford

In the `knitr` package there is a function called `kable` which is available to help produce some nicely formatted tables from RMarkdown (for example). See `help(kable)` for details (you may need to execute `library(knitr)` first).

Here we will illustrate some of its functionality by using part of a dataset from `R`.

```
library(knitr)
data <- head(mtcars)
kable(data, caption="Motor trends' car data")
```

Table 1: Motor trends' car data

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

```
library(kableExtra)
dt <- mtcars[1:5, 1:6]
dt
```

```
##                   mpg cyl disp  hp drat    wt
## Mazda RX4         21.0   6  160 110 3.90 2.620
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875
## Datsun 710        22.8   4  108  93 3.85 2.320
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215
## Hornet Sportabout 18.7   8  360 175 3.15 3.440
```

```
kable(dt, "latex") %>%
  kable_styling("striped") %>%
  add_header_above(c(" " = 1, "Group 1" = 2, "Group 2" = 2, "Group 3" = 2))
```

	Group 1		Group 2		Group 3	
	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440

```
r     cleanWater = c(82, 66)     waterData = data.frame(c('Satisfactory', 'Unsatisfactory'),
rbind(cleanWater, 100-cleanWater))     colnames(waterData) = c('', "(%)", "(%)")     rownames(waterData
= NULL     kable(waterData, "latex", booktabs = T) %>%       kable_styling() %>%       add_header_above
Source" = 1, "2000" = 1, "2015" = 1)) %>%       add_header_above(c(" " = 1, "Year" = 2))

\begin{table}[!h]
```

```
\centering
\begin{tabular}{lrr}
\toprule
\multicolumn{1}{c}{ } & \multicolumn{2}{c}{Year} \\
\cmidrule(l{2pt}r{2pt}){2-3}
\multicolumn{1}{c}{Water Source} & \multicolumn{1}{c}{2000} & \multicolumn{1}{c}{2015} \\
\cmidrule(l{2pt}r{2pt}){1-1} \cmidrule(l{2pt}r{2pt}){2-2} \cmidrule(l{2pt}r{2pt}){3-3}
 & (\%) & (\%)\\
\midrule
Satisfactory & 82 & 66\\
Unsatisfactory & 18 & 34\\
\bottomrule
\end{tabular}
\end{table}
```

Note that `kable` already displays the numbers so that they line up positionally. Some of the functionality of `kable` can be seen in the next few examples:

```
# swap rows and columns (transpose of data = t(data) )
# Note that it matches the number of decimal places within each column
# so that the numbers line up vertically.
kable(t(data))
```

	Mazda RX4	Mazda RX4 Wag	Datsun 710	Hornet 4 Drive	Hornet Sportabout	Valiant
mpg	21.00	21.000	22.80	21.400	18.70	18.10
cyl	6.00	6.000	4.00	6.000	8.00	6.00
disp	160.00	160.000	108.00	258.000	360.00	225.00
hp	110.00	110.000	93.00	110.000	175.00	105.00
drat	3.90	3.900	3.85	3.080	3.15	2.76
wt	2.62	2.875	2.32	3.215	3.44	3.46
qsec	16.46	17.020	18.61	19.440	17.02	20.22
vs	0.00	0.000	1.00	1.000	0.00	1.00
am	1.00	1.000	1.00	0.000	0.00	0.00
gear	4.00	4.000	4.00	3.000	3.00	3.00
carb	4.00	4.000	1.00	1.000	2.00	1.00

```
# swap rows and columns and round so that there are no (digits = 0) decimal places
kable(t(data), digits=0)
```

	Mazda RX4	Mazda RX4 Wag	Datsun 710	Hornet 4 Drive	Hornet Sportabout	Valiant
mpg	21	21	23	21	19	18
cyl	6	6	4	6	8	6
disp	160	160	108	258	360	225
hp	110	110	93	110	175	105
drat	4	4	4	3	3	3
wt	3	3	2	3	3	3
qsec	16	17	19	19	17	20
vs	0	0	1	1	0	1
am	1	1	1	0	0	0
gear	4	4	4	3	3	3
carb	4	4	1	1	2	1

```
# Changing alignment in columns here
kable(t(data), digits = 0, align="lcrlcr")
```

	Mazda RX4	Mazda RX4 Wag	Datsun 710	Hornet 4 Drive	Hornet Sportabout	Valiant
mpg	21	21	23	21	19	18
cyl	6	6	4	6	8	6
disp	160	160	108	258	360	225
hp	110	110	93	110	175	105
drat	4	4	4	3	3	3
wt	3	3	2	3	3	3
qsec	16	17	19	19	17	20
vs	0	0	1	1	0	1
am	1	1	1	0	0	0
gear	4	4	4	3	3	3
carb	4	4	1	1	2	1

```
# back to the original rows and columns but reducing the number of digits displayed
kable(data/10, digits=0)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	2	1	16	11	0	0	2	0	0	0	0
Mazda RX4 Wag	2	1	16	11	0	0	2	0	0	0	0
Datsun 710	2	0	11	9	0	0	2	0	0	0	0
Hornet 4 Drive	2	1	26	11	0	0	2	0	0	0	0
Hornet Sportabout	2	1	36	18	0	0	2	0	0	0	0
Valiant	2	1	22	10	0	0	2	0	0	0	0

```
# No rownames
kable(data, row.names = FALSE)
```

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

```
# To actually change the row names we have to work on the dataset
orig.rownames <- rownames(data)  # save the rownames
rownames(data) <- LETTERS[1:nrow(data)]
kable(data, digits = 0)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
A	21	6	160	110	4	3	16	0	1	4	4
B	21	6	160	110	4	3	17	0	1	4	4
C	23	4	108	93	4	2	19	1	1	4	1
D	21	6	258	110	3	3	19	1	0	3	1

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
E	19	8	360	175	3	3	17	0	0	3	2
F	18	6	225	105	3	3	20	1	0	3	1

```
# And set the rownames back
rownames(data) <- orig.rownames

# Changing the colnames could have been done in the same way using colnames(data)
# or through the kable function
kable(data, digits = 0, col.names = LETTERS[1:ncol(data)])
```

	A	B	C	D	E	F	G	H	I	J	K
Mazda RX4	21	6	160	110	4	3	16	0	1	4	4
Mazda RX4 Wag	21	6	160	110	4	3	17	0	1	4	4
Datsun 710	23	4	108	93	4	2	19	1	1	4	1
Hornet 4 Drive	21	6	258	110	3	3	19	1	0	3	1
Hornet Sportabout	19	8	360	175	3	3	17	0	0	3	2
Valiant	18	6	225	105	3	3	20	1	0	3	1

It's also important to be able to manipulate the data before creating the table.

Here we will add column means to the data and then sort the cols in ascending order of column means.

```
# bind a row (rbind) to data of the colMeans
data1 <- rbind(data, colMeans(data))
rownames(data1) <- c(rownames(data), "Average")
colOrder <- order(colMeans(data))

# The column means
colMeans(data)
```

```
##        mpg        cyl       disp         hp       drat         wt
##  20.500000   6.000000 211.833333 117.166667   3.440000   2.988333
##       qsec         vs         am       gear       carb
##  18.128333   0.500000   0.500000   3.500000   2.166667
```

```
# the order from smallest to largest
order(colMeans(data))
```

```
##  [1]  8  9 11  6  5 10  2  7  1  4  3
```

```
# So we reorder the data as

colorder <- order(colMeans(data))
# With the column averages
kable(data1[,colorder], digits = 0)
```

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Mazda RX4	0	1	4	3	4	4	6	16	21	110	160
Mazda RX4 Wag	0	1	4	3	4	4	6	17	21	110	160
Datsun 710	1	1	1	2	4	4	4	19	23	93	108
Hornet 4 Drive	1	0	1	3	3	3	6	19	21	110	258
Hornet Sportabout	0	0	2	3	3	3	8	17	19	175	360

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Valiant	1	0	1	3	3	3	6	20	18	105	225
Average	0	0	2	3	3	4	6	18	20	117	212

```
# Without the column averages
kable(data[,colorder], digits = 0)
```

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Mazda RX4	0	1	4	3	4	4	6	16	21	110	160
Mazda RX4 Wag	0	1	4	3	4	4	6	17	21	110	160
Datsun 710	1	1	1	2	4	4	4	19	23	93	108
Hornet 4 Drive	1	0	1	3	3	3	6	19	21	110	258
Hornet Sportabout	0	0	2	3	3	3	8	17	19	175	360
Valiant	1	0	1	3	3	3	6	20	18	105	225

```
# And now order the rows in descending order of means.
roworder <- order(rowMeans(data), decreasing = TRUE)
# First order the table as desired
newtable <- data[, colorder][roworder,]
kable(newtable, digits = 0)
```

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Hornet Sportabout	0	0	2	3	3	3	8	17	19	175	360
Hornet 4 Drive	1	0	1	3	3	3	6	19	21	110	258
Valiant	1	0	1	3	3	3	6	20	18	105	225
Mazda RX4 Wag	0	1	4	3	4	4	6	17	21	110	160
Mazda RX4	0	1	4	3	4	4	6	16	21	110	160
Datsun 710	1	1	1	2	4	4	4	19	23	93	108

We might also then remove the column medians from the last table to see the deviations from the column summary (our potential model).

```
colMedians <- apply(newtable, MARGIN = 2, FUN = median)
colMedians <- t(as.matrix(colMedians))
colMedians
```

```
##       vs  am carb    wt drat gear cyl   qsec mpg  hp  disp
## [1,] 0.5 0.5  1.5 3.045  3.5  3.5   6 17.815  21 110 192.5
```

```
# Now form the new table of deviations by sweeping away the colMedians
deviations <- sweep(newtable, MARGIN = 2, STATS = colMedians)

# Note that in this table, the columns are variates that are all on different scales.
# In such cases it makes sense to use different digits for each of the columns.
kable(deviations, digits = c(rep(1,6), rep(0,5)))
```

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Hornet Sportabout	-0.5	-0.5	0.5	0.4	-0.4	-0.5	2	-1	-2	65	168
Hornet 4 Drive	0.5	-0.5	-0.5	0.2	-0.4	-0.5	0	2	0	0	66
Valiant	0.5	-0.5	-0.5	0.4	-0.7	-0.5	0	2	-3	-5	32

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Mazda RX4 Wag	-0.5	0.5	2.5	-0.2	0.4	0.5	0	-1	0	0	-32
Mazda RX4	-0.5	0.5	2.5	-0.4	0.4	0.5	0	-1	0	0	-32
Datsun 710	0.5	0.5	-0.5	-0.7	0.4	0.5	-2	1	2	-17	-84

The table shows a variety of deviation patterns. The last two variates (`hp` and `disp`) have very different deviation patterns from the previous 9, with a notable outlier in the first row. The first six variates have deviations that are all about ± 0.5 with again some notable outliers.

Given that the variates were all on different scales, in Statistics we might try to make them comparable by standardizing them (note that this removes the means from the columns and divides by the standard deviations)

```
scaledtable <- scale(newtable)
kable(scaledtable, digits = 1)
```

	vs	am	carb	wt	drat	gear	cyl	qsec	mpg	hp	disp
Hornet Sportabout	-0.9	-0.9	-0.1	1.0	-0.6	-0.9	1.6	-0.7	-1.0	2.0	1.6
Hornet 4 Drive	0.9	-0.9	-0.8	0.5	-0.7	-0.9	0.0	0.9	0.5	-0.2	0.5
Valiant	0.9	-0.9	-0.8	1.0	-1.4	-0.9	0.0	1.4	-1.4	-0.4	0.1
Mazda RX4 Wag	-0.9	0.9	1.2	-0.2	0.9	0.9	0.0	-0.7	0.3	-0.2	-0.6
Mazda RX4	-0.9	0.9	1.2	-0.8	0.9	0.9	0.0	-1.1	0.3	-0.2	-0.6
Datsun 710	0.9	0.9	-0.8	-1.4	0.8	0.9	-1.6	0.3	1.3	-0.8	-1.2