TestaggregateRMD

Dan Baldewicz

February 10, 2017

R Markdown

```
#Testaggregate

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.3.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.2
```

```
## ------------------------------------------------------------------------------
```

```
## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.3.2
```

```
library(tabplot)
```

```
## Warning: package 'tabplot' was built under R version 3.3.2
```

```
## Loading required package: bit
```

```
## Warning: package 'bit' was built under R version 3.3.2
```

```
## Attaching package bit
```

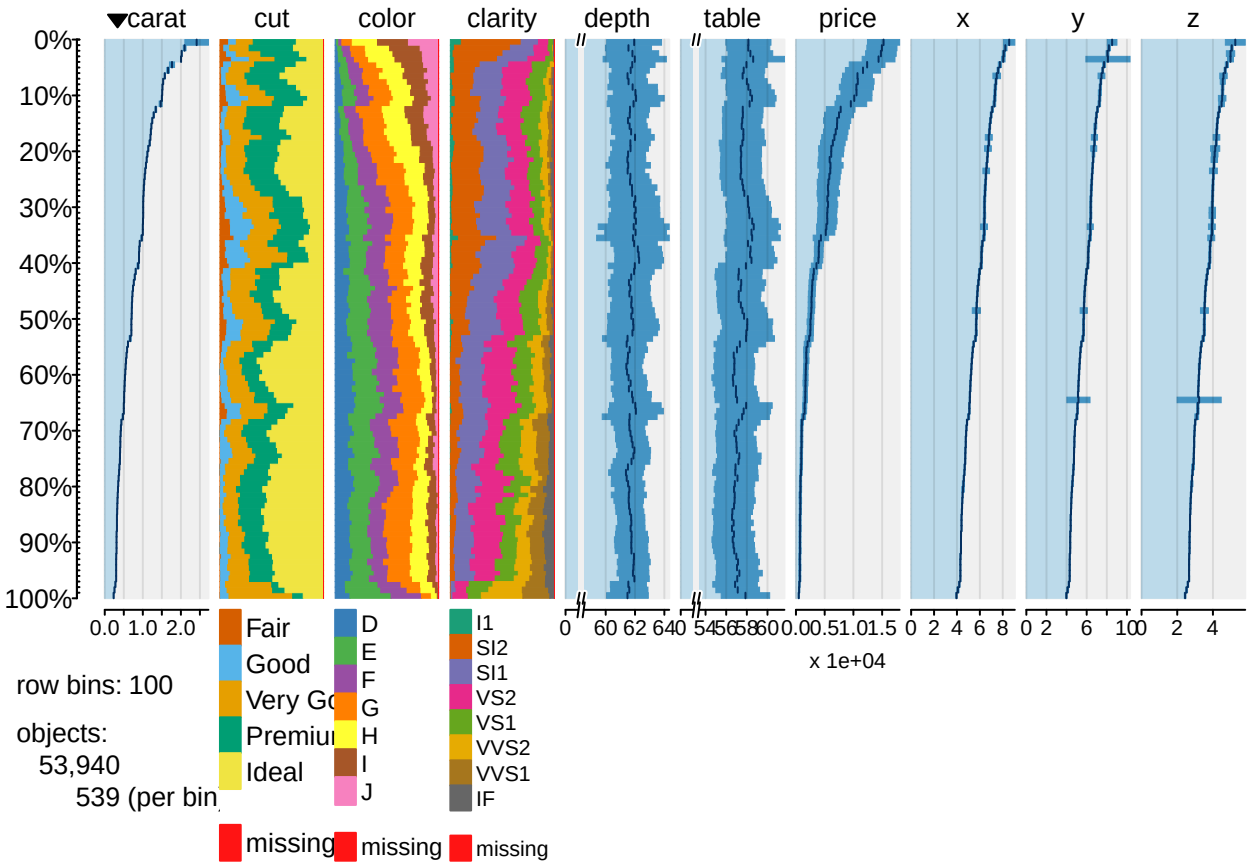
```
## package:bit (c) 2008-2012 Jens Oehlschlaegel (GPL-2)
```

```
## creators: bit bitwhich
```

```
## coercion: as.logical as.integer as.bit as.bitwhich which
## operator: ! & | xor != ==
## querying: print length any all min max range sum summary
## bit access: length<- [ [<- [[ [[<-
## for more help type ?bit
##
## Attaching package: 'bit'
## The following object is masked from 'package:data.table':
##
##     setattr
## The following object is masked from 'package:base':
##
##     xor
## Loading required package: ff
## Warning: package 'ff' was built under R version 3.3.2
## Attaching package ff
## - getOption("fftempdir")=="C:/Users/33531/AppData/Local/Temp/RtmpaSwFQX"
## - getOption("ffextension")=="ff"
## - getOption("ffdrop")==TRUE
## - getOption("fffinonexit")==TRUE
## - getOption("ffpagesize")==65536
## - getOption("ffcaching")=="mmnoflush"  -- consider "ffeachflush" if your system stalls on large writ
## - getOption("ffbatchbytes")==84724940.8 -- consider a different value for tuning your system
## - getOption("ffmaxbytes")==4236247040 -- consider a different value for tuning your system
##
## Attaching package: 'ff'
## The following objects are masked from 'package:bit':
##
##     clone, clone.default, clone.list
## The following objects are masked from 'package:utils':
##
##     write.csv, write.csv2
## The following objects are masked from 'package:base':
##
##     is.factor, is.ordered
## Loading required package: ffbase
## Warning: package 'ffbase' was built under R version 3.3.2
##
## Attaching package: 'ffbase'
```

```
## The following objects are masked from 'package:ff':
##
##     [.ff, [.ffdf, [<-.ff, [<-.ffdf

## The following objects are masked from 'package:base':
##
##     %in%, table
```

```
df<- diamonds

tableplot(df)
```

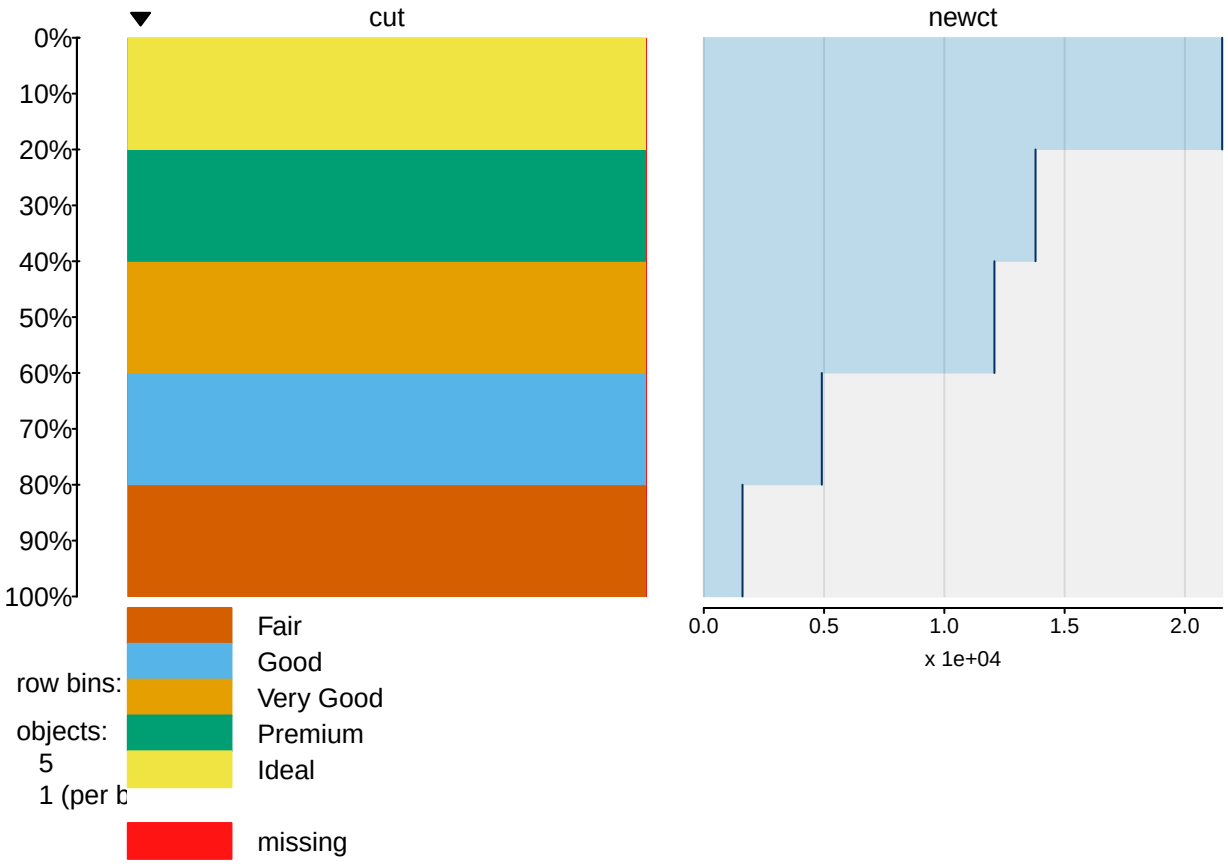



```
#Want table of prices, factor on cut

df2 <- df %>% group_by(cut) %>% summarize(newprice = mean(price))

df3 <- df %>% group_by(cut) %>% summarize(newct = n())

sum(df3$newct)
```

```
## [1] 53940
```

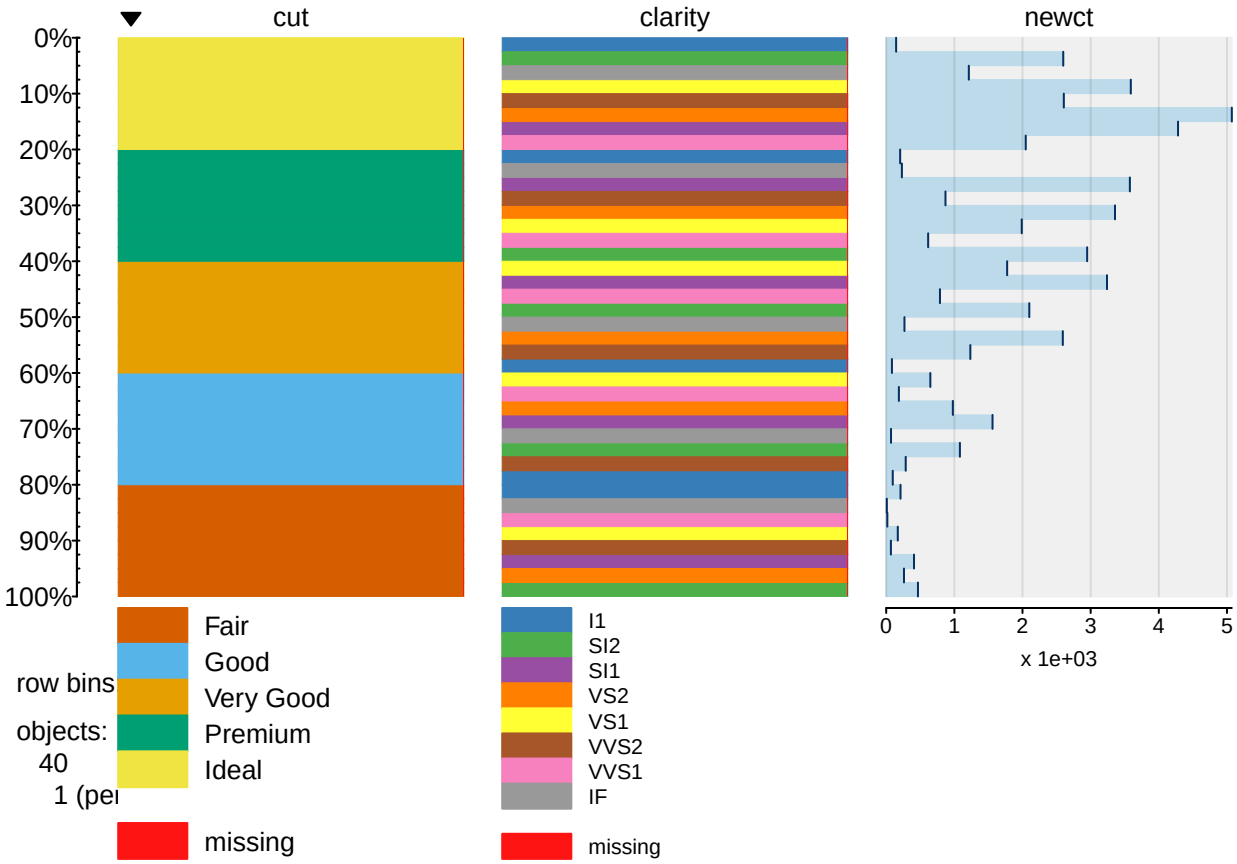
```
tableplot(df3)
```

```
## Warning in tableplot_checkBins(nBins, max(N, 2)): Setting nBins (100) to
## number of rows (5)
```



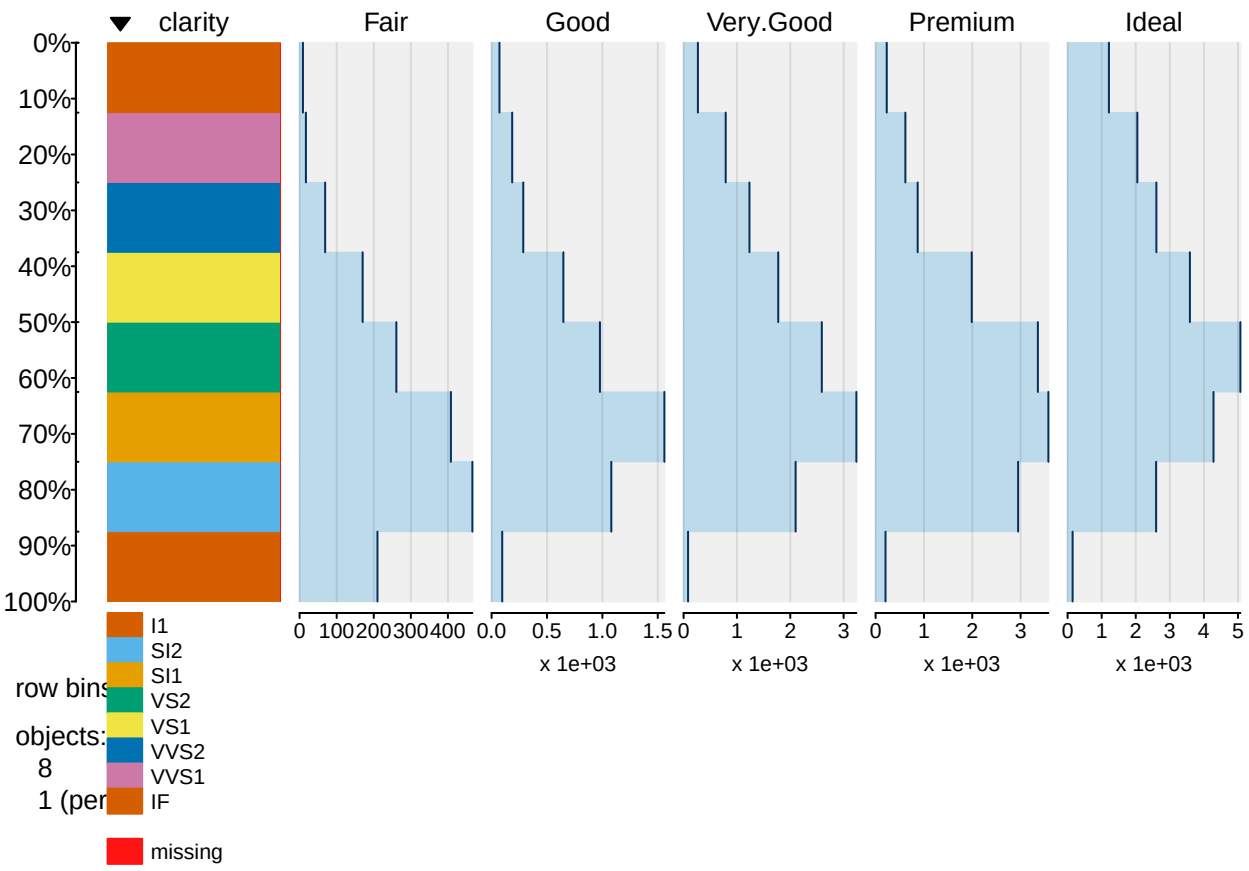

```
df4 <- df %>% group_by(cut,clarity) %>% summarize(newct = n())

tableplot(df4)
```

```
## Warning in tableplot_checkBins(nBins, max(N, 2)): Setting nBins (100) to
## number of rows (40)
```




```
df5 <- spread(df4,cut,newct)

tableplot(df5)
```

```
## Warning in tableplot_checkBins(nBins, max(N, 2)): Setting nBins (100) to
## number of rows (8)
```


clarity
Fair
Good
Very.Good
Premium
Ideal
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0 100 200 300 400
0.0 0.5 1.0 1.5
x 1e+03
0 1 2 3
x 1e+03
0 1 2 3
x 1e+03
0 1 2 3 4 5
x 1e+03
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
missing
row bins
objects:
8
1 (per