```r
library(data.table)
library(reshape2)
library(ggplot2)

# download file
download.file(
    url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data',
    destfile = 'wine.data'
)

# read data
wine <- fread('wine.data')

# first column is the wine type
setnames(wine, 1, 'wine_type')

# download documentation with column titles
download.file(
    url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.names',
    destfile = 'wine_doc.txt'
)

# read documentation
doc <- readLines('wine_doc.txt')

# extract column names
# -- first limit to section 4 of the documentation
doc <- doc[seq(grep('^4.', doc),grep('^5.', doc)-1)]

# -- now extract all characters following the numbered list [0-9])
colnames <- str_trim(str_extract(doc,"(?<=[0-9][)]).+"))

# -- finally, remove NA entries
colnames <- colnames[!is.na(colnames)]

# now assign column names do DF
setnames(wine, seq(2, ncol(wine)), tolower(gsub(' ','_',colnames)))

# rename od280/od315_of_diluted_wines column name
setnames(wine, grep('od280/od315',colnames(wine), value=T), 'wine_dilution')

# visualize table
head(wine)
```

```
##    wine_type alcohol malic_acid  ash alcalinity_of_ash magnesium
## 1:         1   14.23       1.71 2.43              15.6       127
## 2:         1   13.20       1.78 2.14              11.2       100
## 3:         1   13.16       2.36 2.67              18.6       101
## 4:         1   14.37       1.95 2.50              16.8       113
## 5:         1   13.24       2.59 2.87              21.0       118
## 6:         1   14.20       1.76 2.45              15.2       112
##    total_phenols flavanoids nonflavanoid_phenols proanthocyanins
## 1:          2.80       3.06                 0.28            2.29
## 2:          2.65       2.76                 0.26            1.28
```

```
## 3:          2.80       3.24                0.30            2.81
## 4:          3.85       3.49                0.24            2.18
## 5:          2.80       2.69                0.39            1.82
## 6:          3.27       3.39                0.34            1.97
##    color_intensity  hue wine_dilution proline
## 1:            5.64 1.04          3.92    1065
## 2:            4.38 1.05          3.40    1050
## 3:            5.68 1.03          3.17    1185
## 4:            7.80 0.86          3.45    1480
## 5:            4.32 1.04          2.93     735
## 6:            6.75 1.05          2.85    1450
```

```
# examine table
str(wine)
```

```
## Classes 'data.table' and 'data.frame':   178 obs. of  14 variables:
##  $ wine_type           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ alcohol             : num  14.2 13.2 13.2 14.4 13.2 ...
##  $ malic_acid          : num  1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...
##  $ ash                 : num  2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...
##  $ alcalinity_of_ash   : num  15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...
##  $ magnesium           : int  127 100 101 113 118 112 96 121 97 98 ...
##  $ total_phenols       : num  2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...
##  $ flavanoids          : num  3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...
##  $ nonflavanoid_phenols: num  0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...
##  $ proanthocyanins     : num  2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...
##  $ color_intensity     : num  5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...
##  $ hue                 : num  1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...
##  $ wine_dilution       : num  3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...
##  $ proline             : int  1065 1050 1185 1480 735 1450 1290 1295 1045 1045 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
# variable details
summary(wine)
```

```
##    wine_type        alcohol        malic_acid         ash
##  Min.   :1.000   Min.   :11.03   Min.   :0.740   Min.   :1.360
##  1st Qu.:1.000   1st Qu.:12.36   1st Qu.:1.603   1st Qu.:2.210
##  Median :2.000   Median :13.05   Median :1.865   Median :2.360
##  Mean   :1.938   Mean   :13.00   Mean   :2.336   Mean   :2.367
##  3rd Qu.:3.000   3rd Qu.:13.68   3rd Qu.:3.083   3rd Qu.:2.558
##  Max.   :3.000   Max.   :14.83   Max.   :5.800   Max.   :3.230
##  alcalinity_of_ash   magnesium      total_phenols     flavanoids
##  Min.   :10.60     Min.   : 70.00   Min.   :0.980   Min.   :0.340
##  1st Qu.:17.20     1st Qu.: 88.00   1st Qu.:1.742   1st Qu.:1.205
##  Median :19.50     Median : 98.00   Median :2.355   Median :2.135
##  Mean   :19.49     Mean   : 99.74   Mean   :2.295   Mean   :2.029
##  3rd Qu.:21.50     3rd Qu.:107.00   3rd Qu.:2.800   3rd Qu.:2.875
##  Max.   :30.00     Max.   :162.00   Max.   :3.880   Max.   :5.080
##  nonflavanoid_phenols proanthocyanins color_intensity       hue
##  Min.   :0.1300       Min.   :0.410   Min.   : 1.280   Min.   :0.4800
##  1st Qu.:0.2700       1st Qu.:1.250   1st Qu.: 3.220   1st Qu.:0.7825
##  Median :0.3400       Median :1.555   Median : 4.690   Median :0.9650
##  Mean   :0.3619       Mean   :1.591   Mean   : 5.058   Mean   :0.9574
##  3rd Qu.:0.4375       3rd Qu.:1.950   3rd Qu.: 6.200   3rd Qu.:1.1200
##  Max.   :0.6600       Max.   :3.580   Max.   :13.000   Max.   :1.7100
```

```
##  wine_dilution      proline     
##  Min.   :1.270   Min.   : 278.0  
##  1st Qu.:1.938   1st Qu.: 500.5  
##  Median :2.780   Median : 673.5  
##  Mean   :2.612   Mean   : 746.9  
##  3rd Qu.:3.170   3rd Qu.: 985.0  
##  Max.   :4.000   Max.   :1680.0
```