Data Mining 201 Homework 1

Idris Raja

April 13, 2012

Does this caption work?

```
wine <- read.table("data/winequality-red.csv", sep = ";", header = TRUE)
summary(wine)
```

```
##  fixed.acidity   volatile.acidity  citric.acid    residual.sugar    chlorides
##  Min.   : 4.60   Min.   :0.120    Min.   :0.000   Min.   : 0.90   Min.   :0.0120
##  1st Qu.: 7.10   1st Qu.:0.390    1st Qu.:0.090   1st Qu.: 1.90   1st Qu.:0.0700
##  Median : 7.90   Median :0.520    Median :0.260   Median : 2.20   Median :0.0790
##  Mean   : 8.32   Mean   :0.528    Mean   :0.271   Mean   : 2.54   Mean   :0.0875
##  3rd Qu.: 9.20   3rd Qu.:0.640    3rd Qu.:0.420   3rd Qu.: 2.60   3rd Qu.:0.0900
##  Max.   :15.90   Max.   :1.580    Max.   :1.000   Max.   :15.50   Max.   :0.6110
##  free.sulfur.dioxide total.sulfur.dioxide    density             pH         sulphates
##  Min.   : 1.0        Min.   :  6.0        Min.   :0.990   Min.   :2.74   Min.   :0.330
##  1st Qu.: 7.0        1st Qu.: 22.0        1st Qu.:0.996   1st Qu.:3.21   1st Qu.:0.550
##  Median :14.0        Median : 38.0        Median :0.997   Median :3.31   Median :0.620
##  Mean   :15.9        Mean   : 46.5        Mean   :0.997   Mean   :3.31   Mean   :0.658
##  3rd Qu.:21.0        3rd Qu.: 62.0        3rd Qu.:0.998   3rd Qu.:3.40   3rd Qu.:0.730
##  Max.   :72.0        Max.   :289.0        Max.   :1.004   Max.   :4.01   Max.   :2.000
##     alcohol        quality
##  Min.   : 8.4   Min.   :3.00
##  1st Qu.: 9.5   1st Qu.:5.00
##  Median :10.2   Median :6.00
##  Mean   :10.4   Mean   :5.64
##  3rd Qu.:11.1   3rd Qu.:6.00
##  Max.   :14.9   Max.   :8.00
```

```
str(wine)
```

```
## 'data.frame': 1599 obs. of  12 variables:
##  $ fixed.acidity       : num  7.4 7.8 7.8 11.2 7.4 7.4 7.9 7.3 7.8 7.5 ...
##  $ volatile.acidity    : num  0.7 0.88 0.76 0.28 0.7 0.66 0.6 0.65 0.58 0.5 ...
##  $ citric.acid         : num  0 0 0.04 0.56 0 0 0.06 0 0.02 0.36 ...
##  $ residual.sugar      : num  1.9 2.6 2.3 1.9 1.9 1.8 1.6 1.2 2 6.1 ...
##  $ chlorides           : num  0.076 0.098 0.092 0.075 0.076 0.075 0.069 0.065 0.073 0.071 ...
##  $ free.sulfur.dioxide : num  11 25 15 17 11 13 15 15 9 17 ...
##  $ total.sulfur.dioxide: num  34 67 54 60 34 40 59 21 18 102 ...
##  $ density             : num  0.998 0.997 0.997 0.998 0.998 ...
##  $ pH                  : num  3.51 3.2 3.26 3.16 3.51 3.51 3.3 3.39 3.36 3.35 ...
##  $ sulphates           : num  0.56 0.68 0.65 0.58 0.56 0.56 0.46 0.47 0.57 0.8 ...
##  $ alcohol             : num  9.4 9.8 9.8 9.8 9.4 9.4 9.4 10 9.5 10.5 ...
##  $ quality             : int  5 5 5 6 5 5 5 7 7 5 ...
```

```
head(wine)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
## 1           7.4             0.70        0.00            1.9     0.076                  11
## 2           7.8             0.88        0.00            2.6     0.098                  25
## 3           7.8             0.76        0.04            2.3     0.092                  15
## 4          11.2             0.28        0.56            1.9     0.075                  17
## 5           7.4             0.70        0.00            1.9     0.076                  11
## 6           7.4             0.66        0.00            1.8     0.075                  13
##   total.sulfur.dioxide density   pH sulphates alcohol quality
## 1                   34  0.9978 3.51      0.56     9.4       5
## 2                   67  0.9968 3.20      0.68     9.8       5
## 3                   54  0.9970 3.26      0.65     9.8       5
## 4                   60  0.9980 3.16      0.58     9.8       6
## 5                   34  0.9978 3.51      0.56     9.4       5
## 6                   40  0.9978 3.51      0.56     9.4       5
```

The first element of `x` is

```
Error in eval(expr, envir, enclos) : object 'x' not found
```

. Boring boxplots and histograms recorded by the PDF device:

```
## two plots side by side (option fig.show='hold')
par(mar = c(4, 4, 0.1, 0.1), cex.lab = 0.95, cex.axis = 0.9, mgp = c(2, 0.7, 0),
    tcl = -0.3, las = 1)
boxplot(x)

## Error: object 'x' not found

hist(x, main = "")

## Error: object 'x' not found
```

Do the above chunks work? You should be able to compile the TEX document and get a PDF file like this one: `https://github.com/downloads/yihui/knitr/knitr-minimal.pdf`. The Rnw source of this document is at `https://github.com/yihui/knitr/blob/master/inst/examples/knitr-minimal.Rnw`.