

Bruce Campbell ST-617 Discussion Group 7

Mon Jul 18 22:49:53 2016

```
rm(list = ls())
set.seed(7)

setwd("C:/st-617/")

## use read.csv2 since data fields are delimited by semicolons (;)

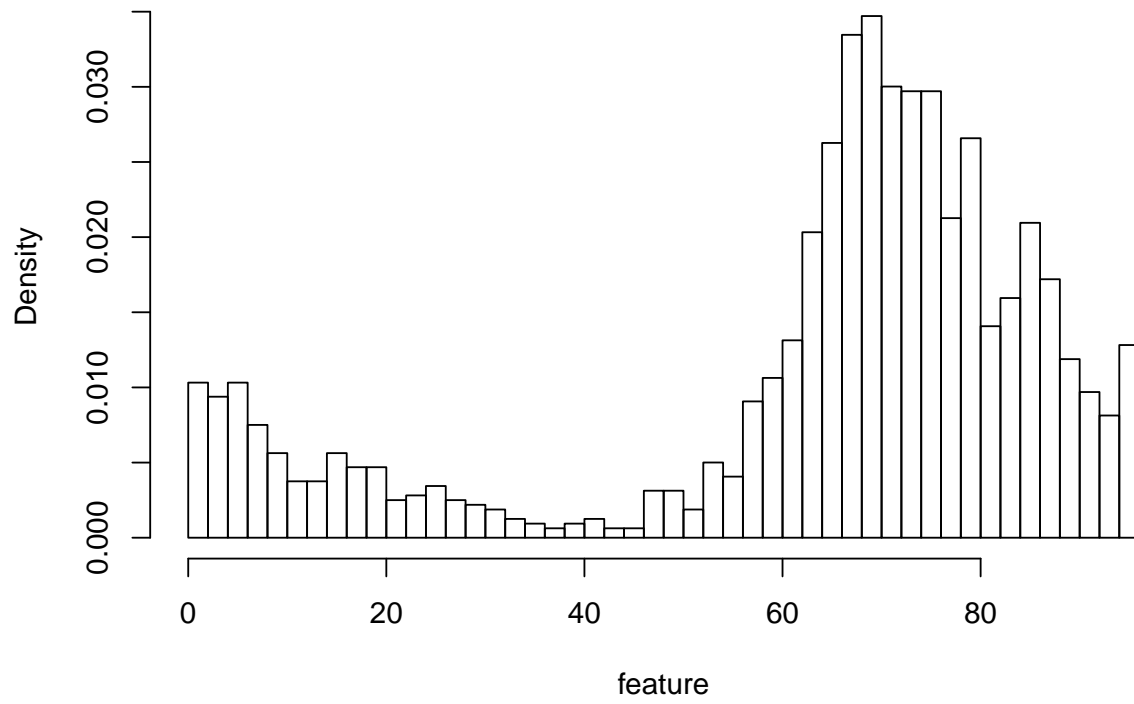
cls = c(rep("numeric", 11), "integer")

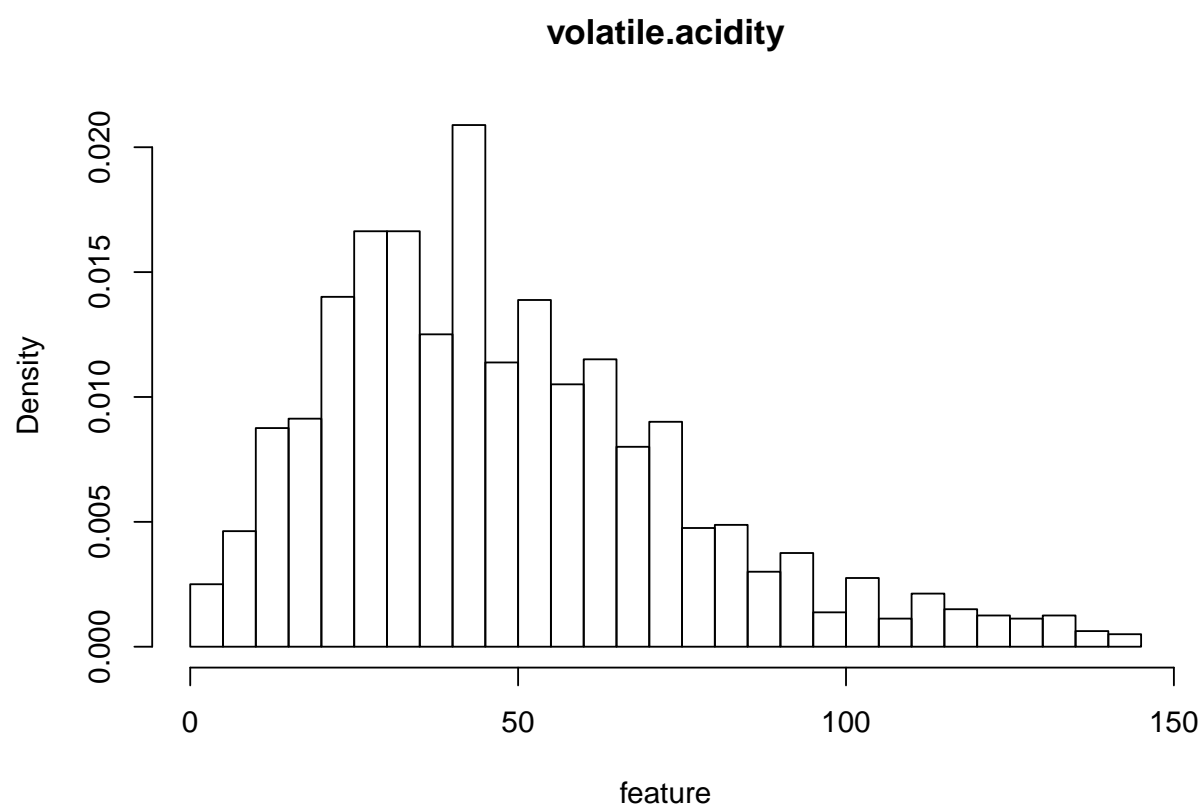
wine.data <- read.csv2("FinalProject/winequality-red.csv")

# Specifying the data type via colClasses did not work - so we supply
# as.numeric to convert the factor data to numeric. We may want to convert
# some of the variables to factors later - but for visualization of the raw
# features we should use numeric type.
wine.data[, c(1:12)] <- sapply(wine.data[, c(1:12)], as.numeric)

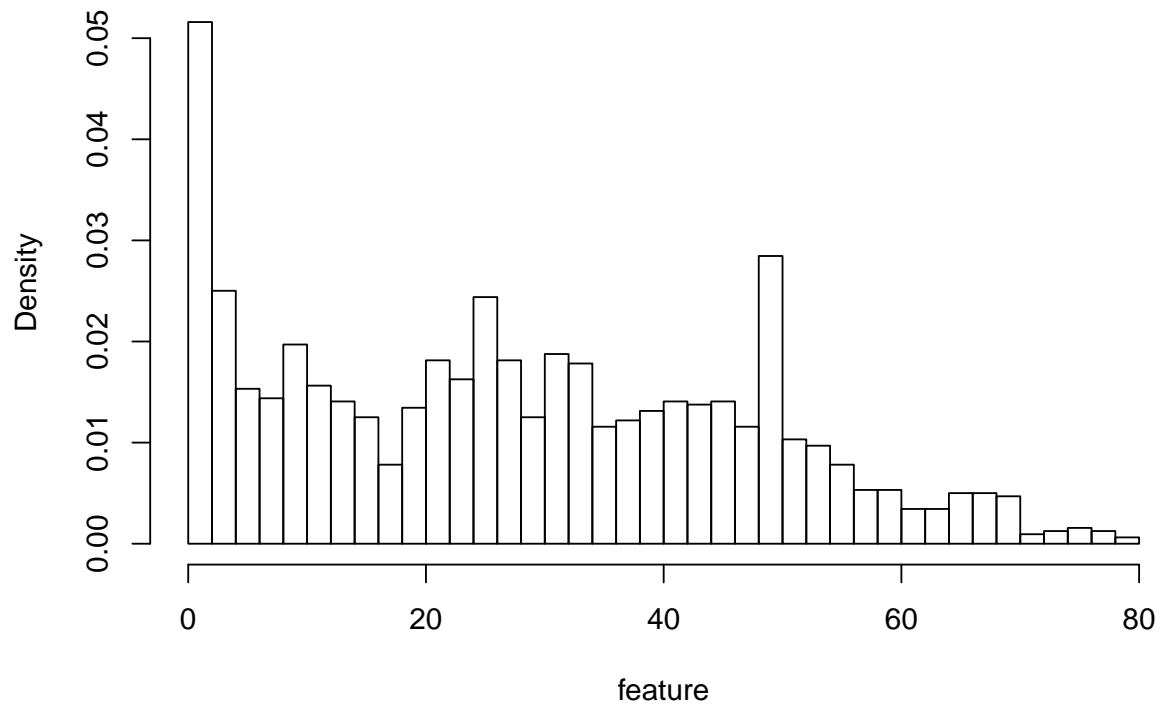
for (i in 1:(ncol(wine.data) - 1)) {
  feature = wine.data[, i]
  featureName = names(wine.data)[i]
  hist(feature, main = featureName, freq = FALSE, 50)
}
```

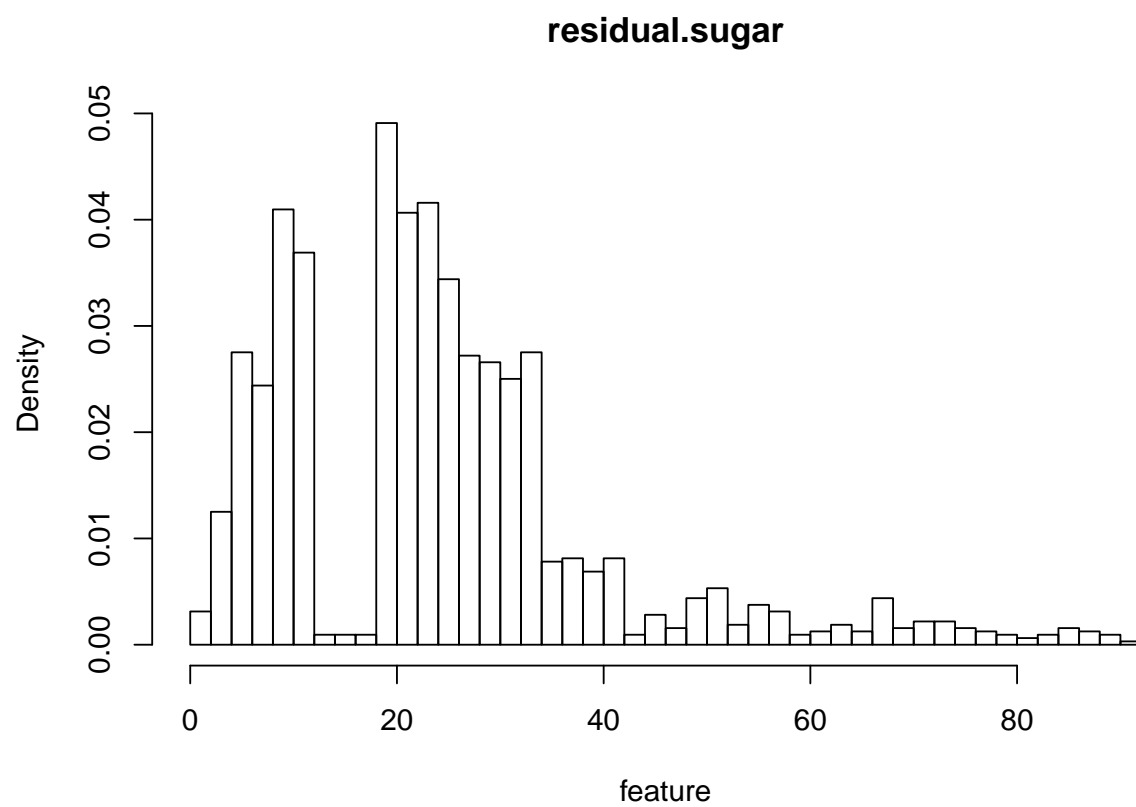
fixed.acidity



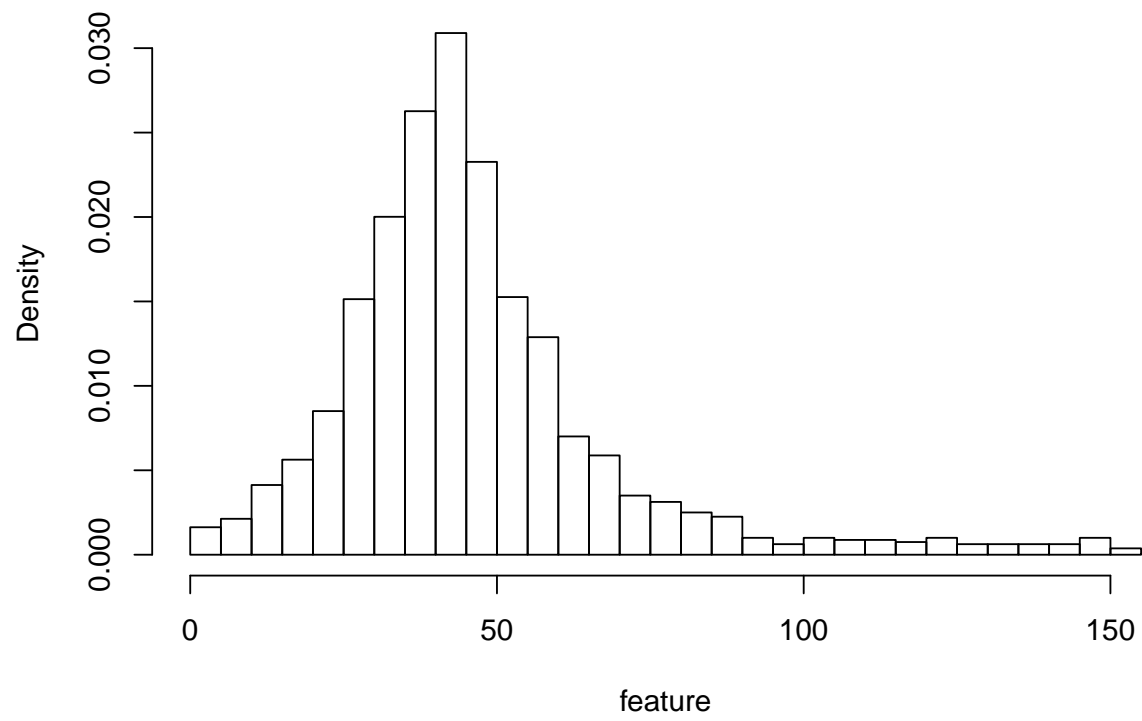


citric.acid

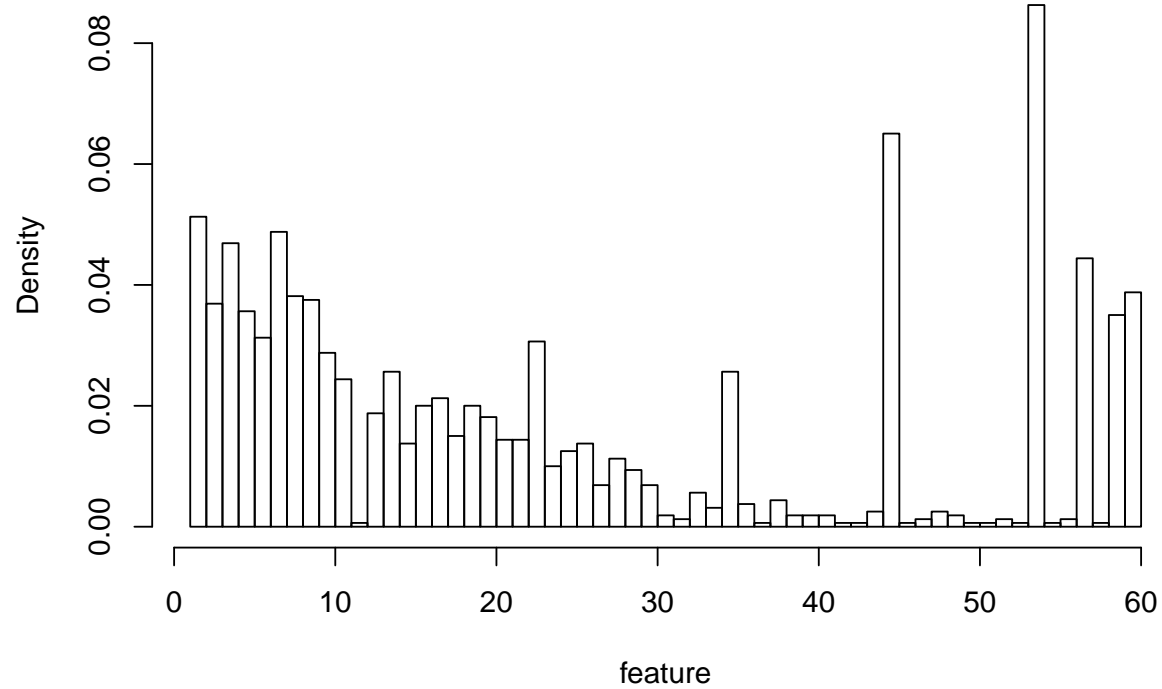


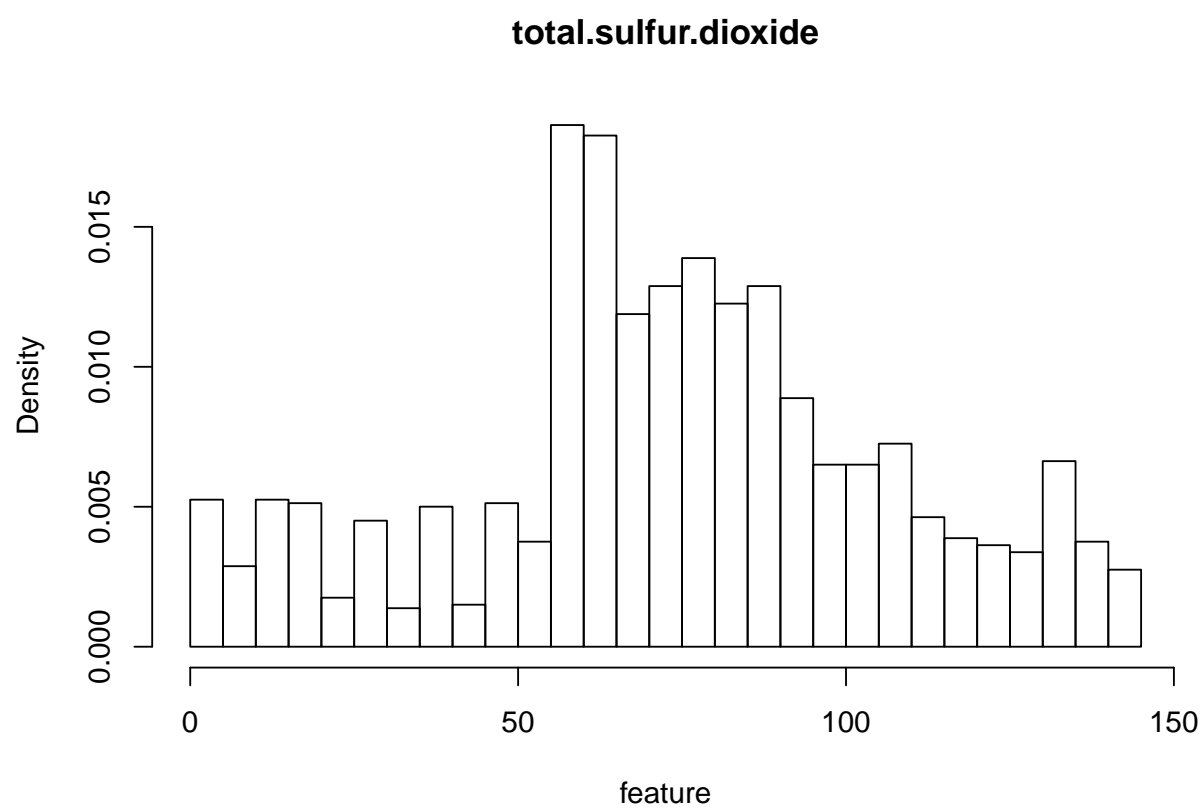


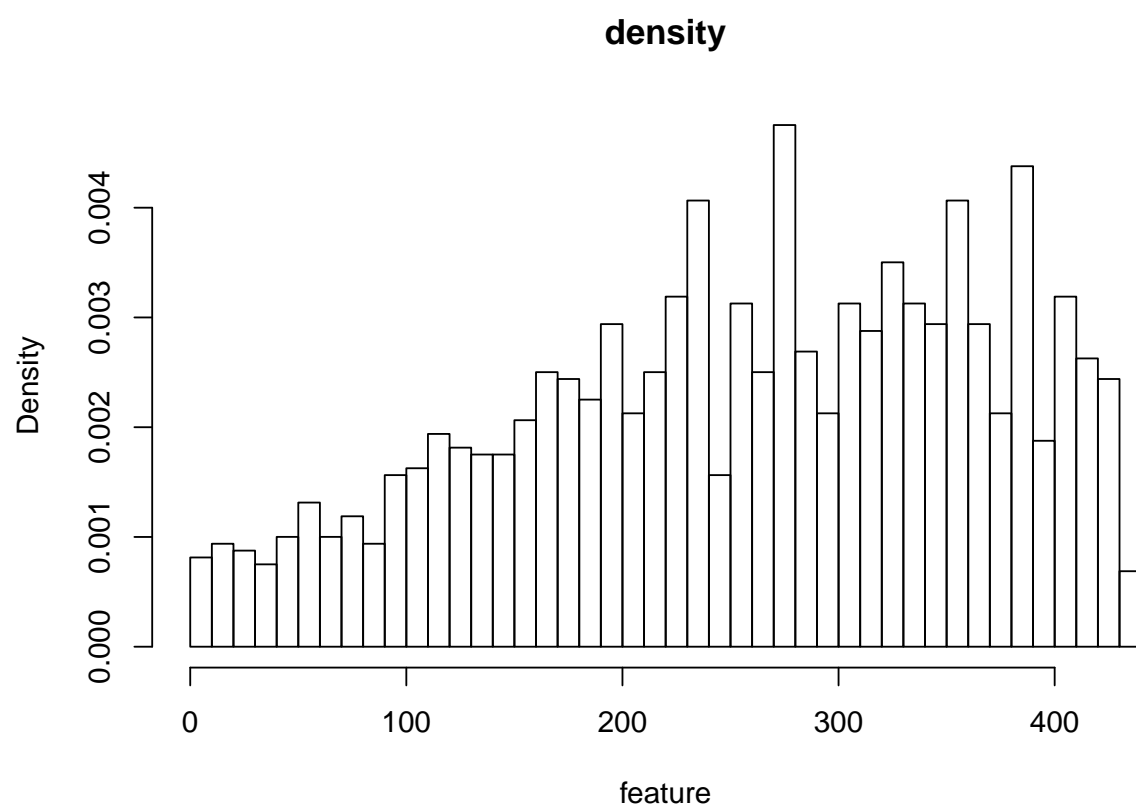
chlorides

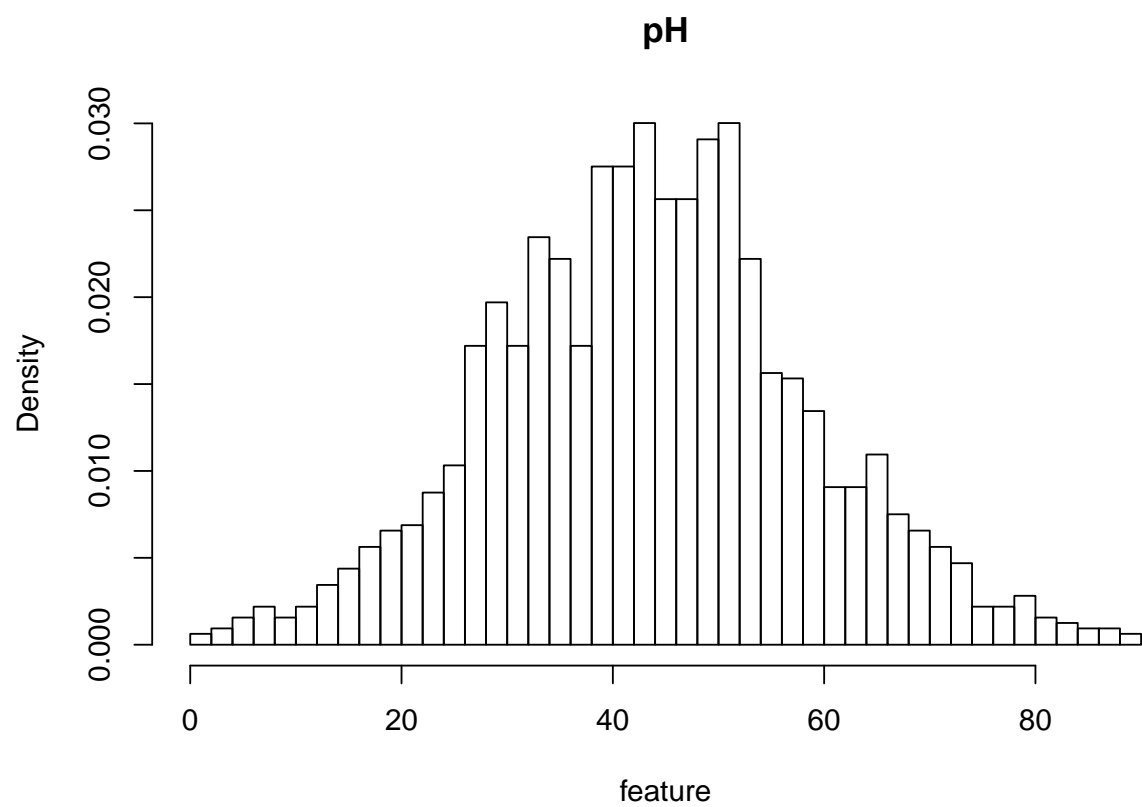


free.sulfur.dioxide

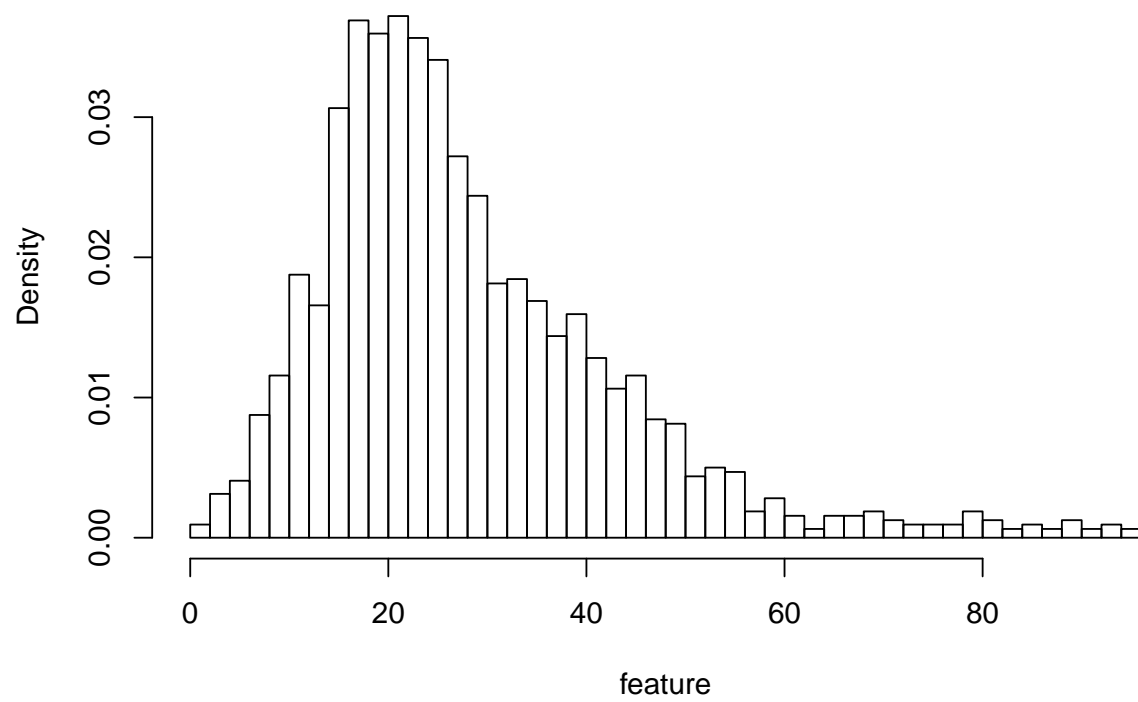




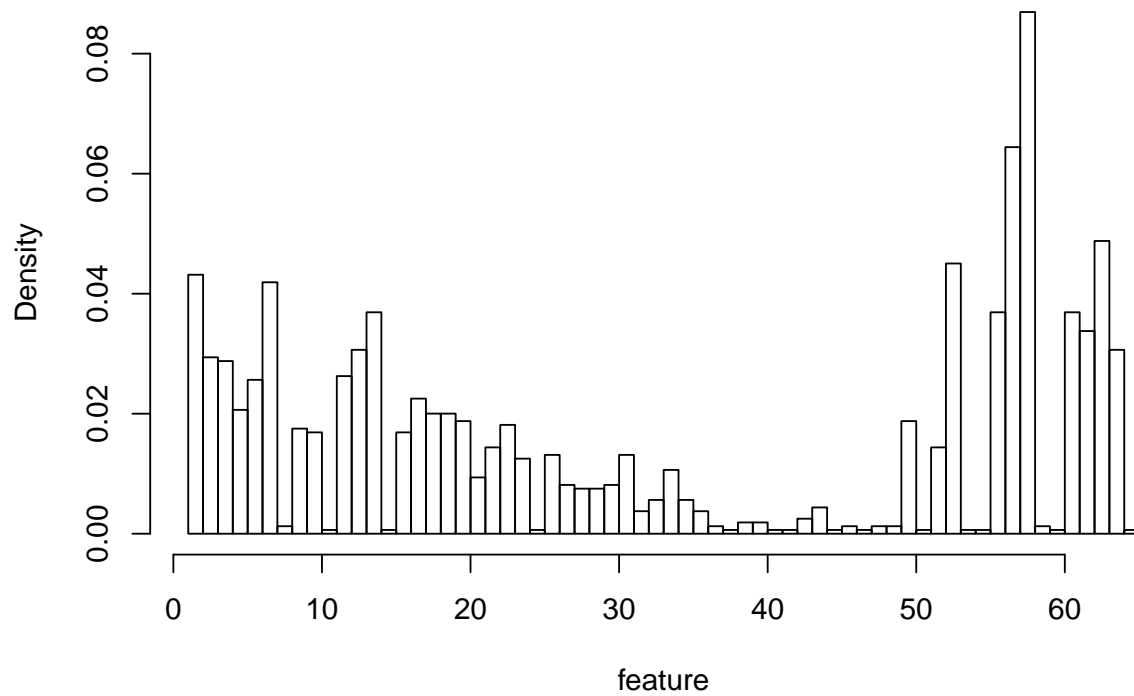




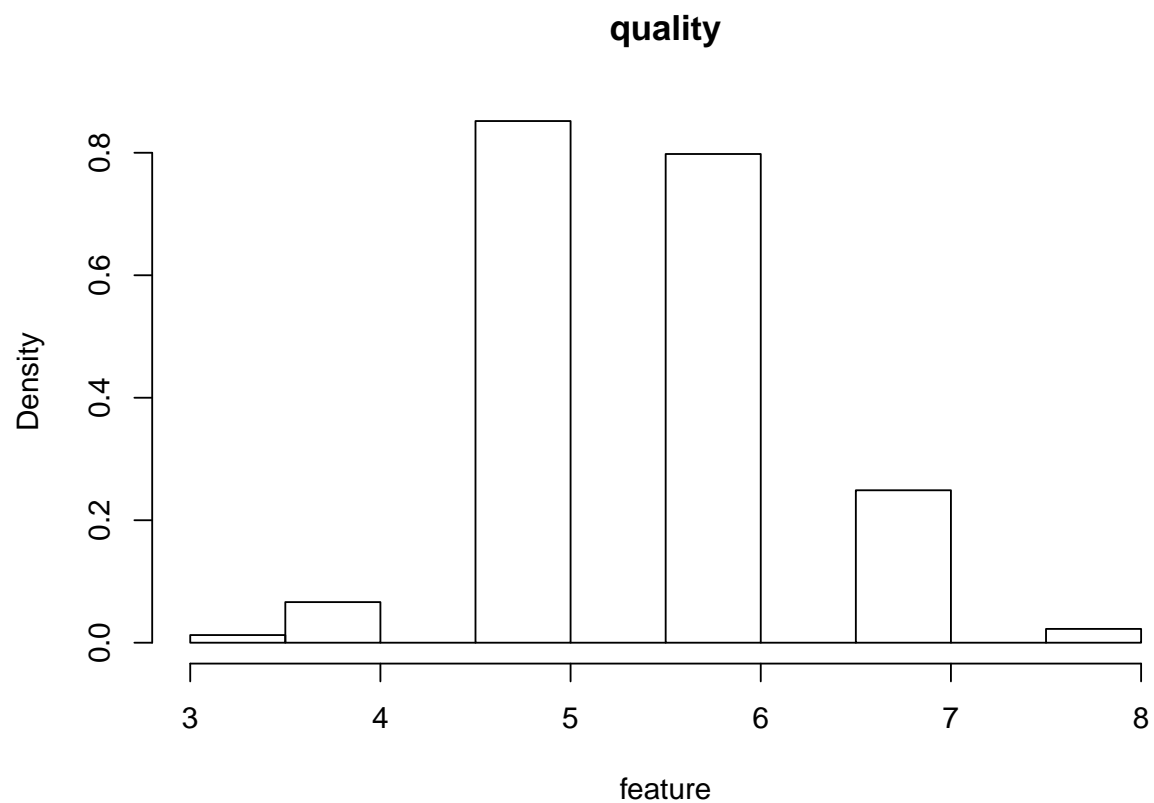
sulphates



alcohol



```
i = ncol(wine.data)
feature = wine.data[, i]
featureName = names(wine.data)[i]
hist(feature, main = featureName, freq = FALSE)
```



fixed.acidity

Slightly bimodal

volatile.acidity

Fat tail. Looks Poisson

citric.acid

residual.sugar

Bimodal, fat tail to the right

chlorides

Fat tail to the right. Looks lognormal

free.sulfur.dioxide

total.sulfur.dioxide

density

pH

Looks Normal

sulphates

Fat tail to the right - looks lognormal

alcohol

Slightly bimodal

quality

The majority of the wines have a quality factor of 5 ,6