Olive Oil Analysis R

Tanner Patrom

02/06/2024

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

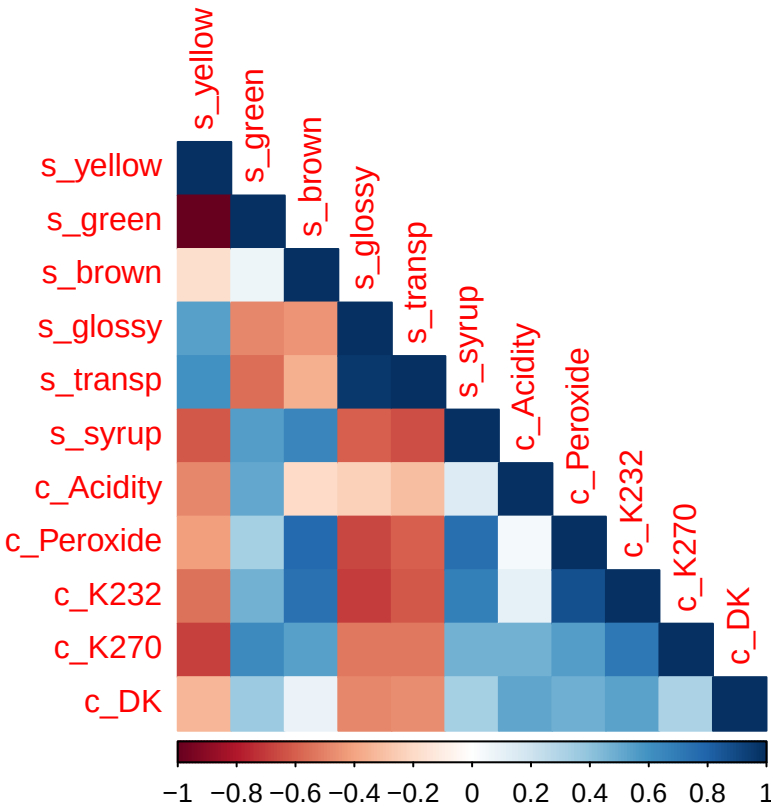
```
oo_sensory = read_csv("olive_oil_sensory.csv", show_col_types = FALSE)
```

```
oo_chemical = read_csv("olive_oil_chemical.csv", show_col_types = FALSE)
```

```
combined = left_join(oo_sensory, oo_chemical, by="region")
print(combined)
```

```
## # A tibble: 16 x 12
##    region s_yellow s_green s_brown s_glossy s_transp s_syrup c_Acidity
##    <chr>     <dbl>   <dbl>   <dbl>    <dbl>    <dbl>   <dbl>     <dbl>
##  1 G1         21.4    73.4    10.1     79.7     75.2    50.3      0.73
##  2 G2         23.4    66.3     9.8     77.8     68.7    51.7      0.19
##  3 G3         32.7    53.5     8.7     82.3     83.2    45.4      0.26
##  4 G4         30.2    58.3    12.2     81.1     77.1    47.8      0.67
##  5 G5         51.8    32.5     8       72.4     65.3    46.5      0.52
##  6 I1         40.7    42.9    20.1     67.7     63.5    52.2      0.26
##  7 I2         53.8    30.4    11.5     77.8     77.3    45.2      0.24
##  8 I3         26.4    66.5    14.2     78.7     74.6    51.8      0.3
##  9 I4         65.7    12.1    10.3     81.6     79.6    48.3      0.35
## 10 I5         45      31.9    28.4     75.7     72.9    52.8      0.19
## 11 S1         70.9    12.2    10.8     87.7     88.1    44.5      0.15
## 12 S2         73.5     9.7     8.3     89.9     89.7    42.3      0.16
## 13 S3         68.1    12      10.8     78.4     75.1    46.4      0.27
## 14 S4         67.6    13.9    11.9     84.6     83.8    48.5      0.16
## 15 S5         71.4    10.6    10.8     88.1     88.5    46.7      0.24
## 16 S6         71.4    10      11.4     89.5     88.5    47.2      0.3
## # i 4 more variables: c_Peroxide <dbl>, c_K232 <dbl>, c_K270 <dbl>, c_DK <dbl>
```

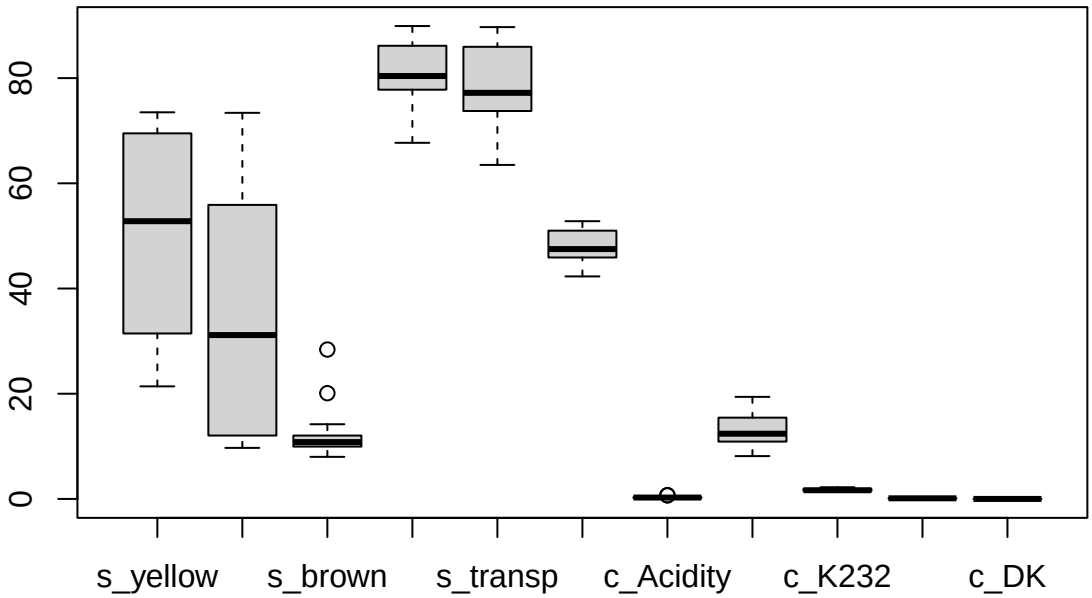
```
numeric = combined[-1]
cMatrix = cor(numeric)
corrplot(cMatrix, method = "color", type = "lower")
```



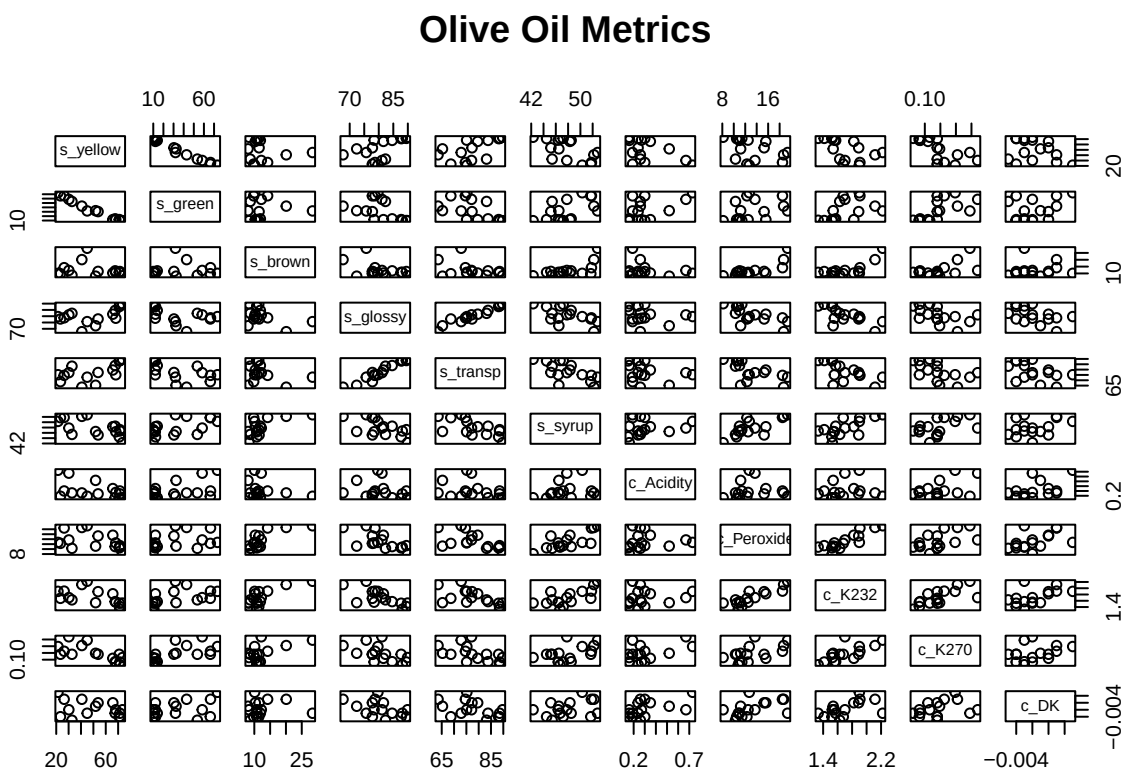

```
summary(numeric)
```

```
##     s_yellow        s_green         s_brown         s_glossy
##  Min.   :21.40   Min.   : 9.70   Min.   : 8.00   Min.   :67.70
##  1st Qu.:32.08   1st Qu.:12.07   1st Qu.:10.03   1st Qu.:77.80
##  Median :52.80   Median :31.15   Median :10.80   Median :80.40
##  Mean   :50.88   Mean   :33.51   Mean   :12.33   Mean   :80.81
##  3rd Qu.:68.80   3rd Qu.:54.70   3rd Qu.:11.97   3rd Qu.:85.38
##  Max.   :73.50   Max.   :73.40   Max.   :28.40   Max.   :89.90
##     s_transp        s_syrup         c_Acidity        c_Peroxide
##  Min.   :63.50   Min.   :42.30   Min.   :0.1500   Min.   : 8.14
##  1st Qu.:74.17   1st Qu.:46.15   1st Qu.:0.1900   1st Qu.:10.95
##  Median :77.20   Median :47.50   Median :0.2600   Median :12.40
##  Mean   :78.19   Mean   :47.98   Mean   :0.3119   Mean   :13.25
##  3rd Qu.:84.88   3rd Qu.:50.65   3rd Qu.:0.3125   3rd Qu.:15.38
##  Max.   :89.70   Max.   :52.80   Max.   :0.7300   Max.   :19.40
##      c_K232          c_K270            c_DK
##  Min.   :1.331   Min.   :0.0850   Min.   :-0.00500
##  1st Qu.:1.536   1st Qu.:0.1015   1st Qu.:-0.00325
##  Median :1.653   Median :0.1160   Median :-0.00200
##  Mean   :1.708   Mean   :0.1181   Mean   :-0.00175
##  3rd Qu.:1.893   3rd Qu.:0.1285   3rd Qu.: 0.00000
##  Max.   :2.222   Max.   :0.1680   Max.   : 0.00300
```

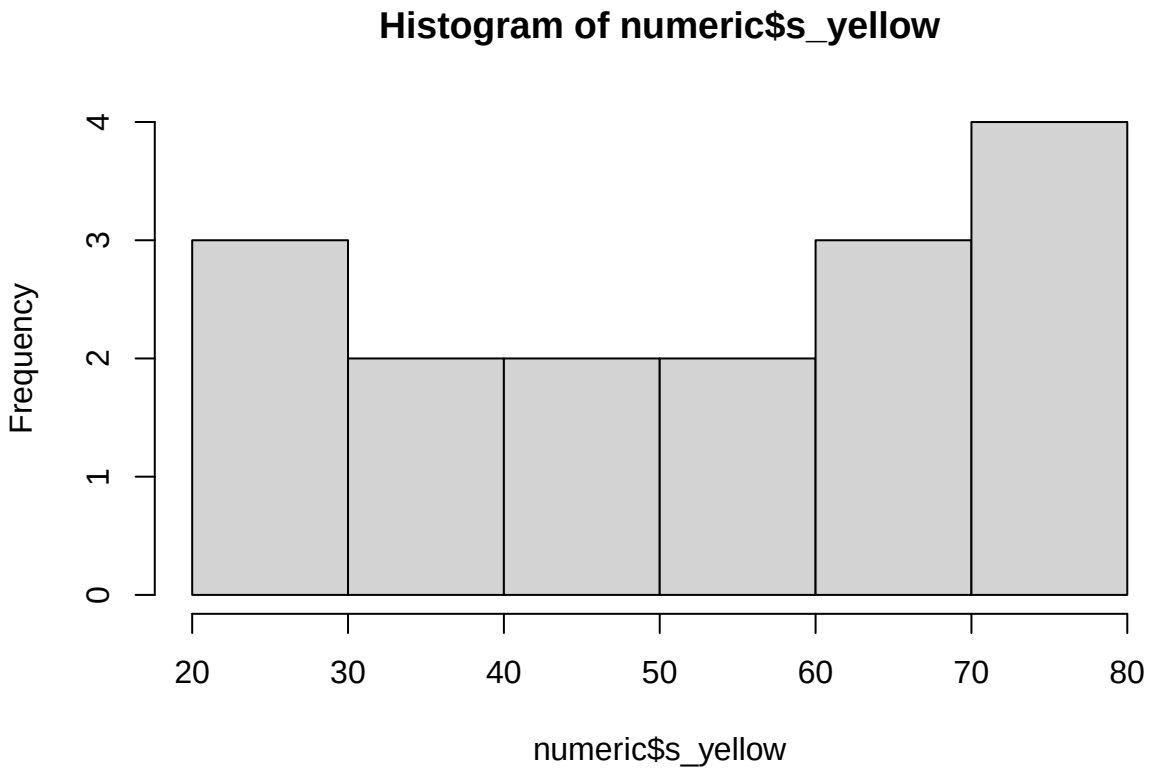
```
boxplot(numeric)
```




```
pairs(numeric, main="Olive Oil Metrics", labels = colnames(numeric)[1:11])
```




```
hist(numeric$s_yellow)
```




```
means = c()
for(column in numeric){
  means = c(means, mean(column))
}

print(means)
```

```
##  [1] 50.8750000 33.5125000 12.3312500 80.8125000 78.1937500 47.9750000
##  [7]  0.3118750 13.2525000  1.7082500  0.1181438 -0.0017500
```

```
medians = c()
for(column in numeric){
  medians = c(medians, median(column))
}

print(medians)
```

```
##  [1] 52.8000 31.1500 10.8000 80.4000 77.2000 47.5000  0.2600 12.4000  1.6535
## [10]  0.1160 -0.0020
```

```
diff = c()
for(i in 1:11){
  diff = c(diff, means[i] - medians[i])
}

print(diff)
```

```
##  [1] -1.92500000  2.36250000  1.53125000  0.41250000  0.99375000  0.47500000
##  [7]  0.05187500  0.85250000  0.05475000  0.00214375  0.00025000
```