Olive data

Ordering displays

25 marks

The **olive oil** data is a well known dataset having interesting properties,

This contains measurements on the fatty acid content of 572 different Italian olive oils; eight different fatty acids are measured.
The olive oils come from one of nine different olive growing regions in Italy.

This data are easily available from the `loon` package:

```
library(loon)
# The first three rows of which are
head(olive, 3)
```

```
##   Region         Area palmitic palmitoleic stearic oleic linoleic linolenic
## 1  South North-Apulia     1075          75     226  7823      672        36
## 2  South North-Apulia     1088          73     224  7709      781        31
## 3  South North-Apulia      911          54     246  8113      549        31
##   arachidic eicosenoic
## 1        60         29
## 2        61         29
## 3        63         29
```

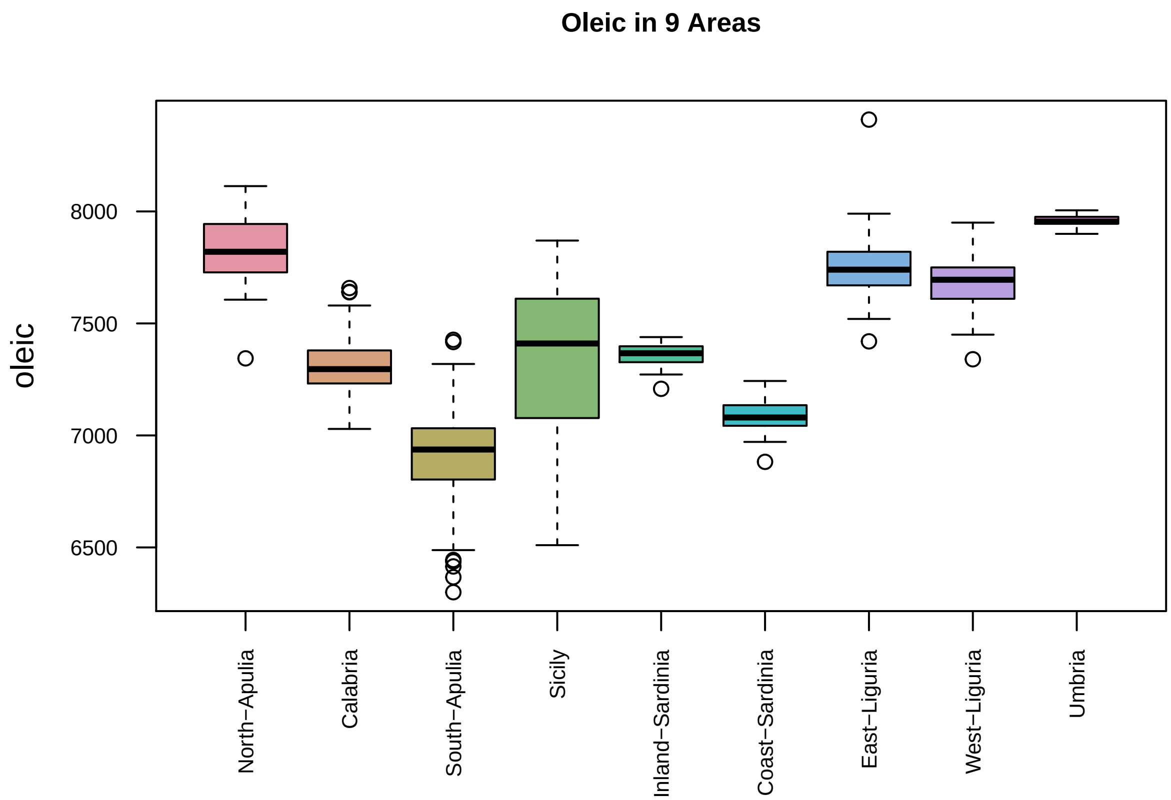
In this question you are going to focus on the fatty acid `oleic`.

a. **(2 marks)** Separate the data on `oleic` into 9 different groups as defined by the olive growing `Area`, and draw side by side boxplots of all 9 groups. Colour the boxplots uniquely using

```
library(colorspace)
cols <- rainbow_hcl(9) # Use these colours.
```

Show your code together with your output.

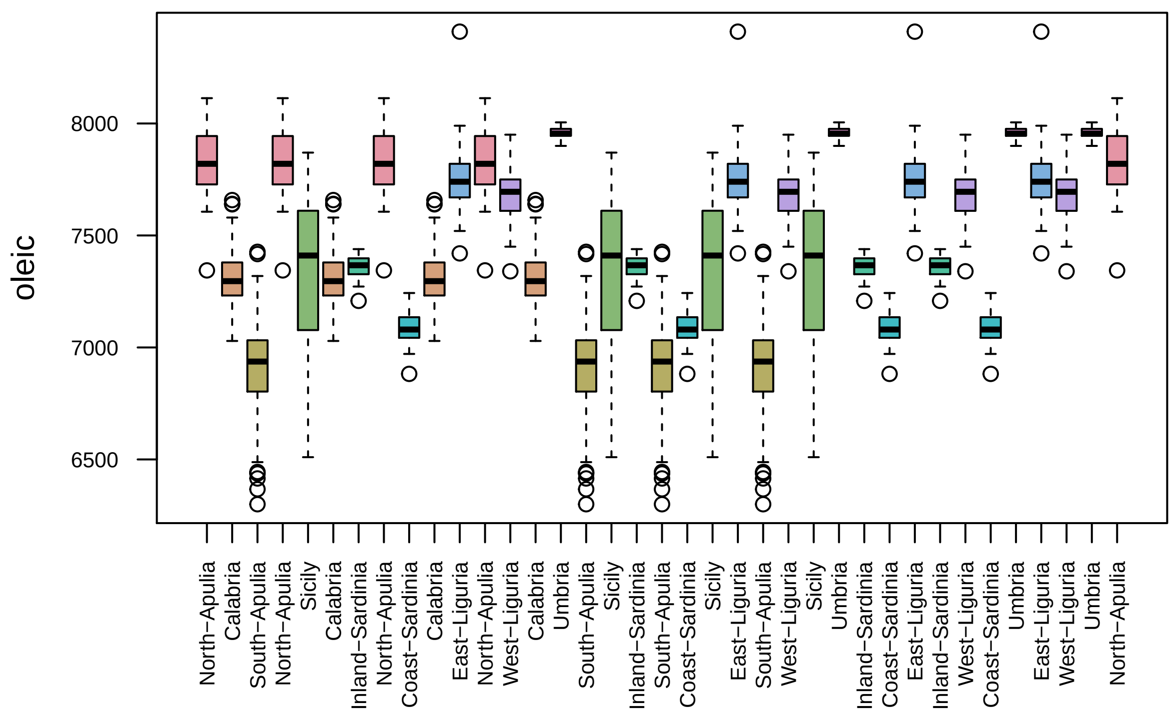
```
boxplot(oleic ~ Area, data = olive, col = cols, las=2, cex.axis=0.7
        , xlab='', main='Oleic in 9 Areas', cex.main=0.8)
```



b. Load the `R` package `PairViz` (i.e. `library(PairViz)`). Use the variate `oleic` and the same colours for the olive growing areas as in part (a) throughout the following:

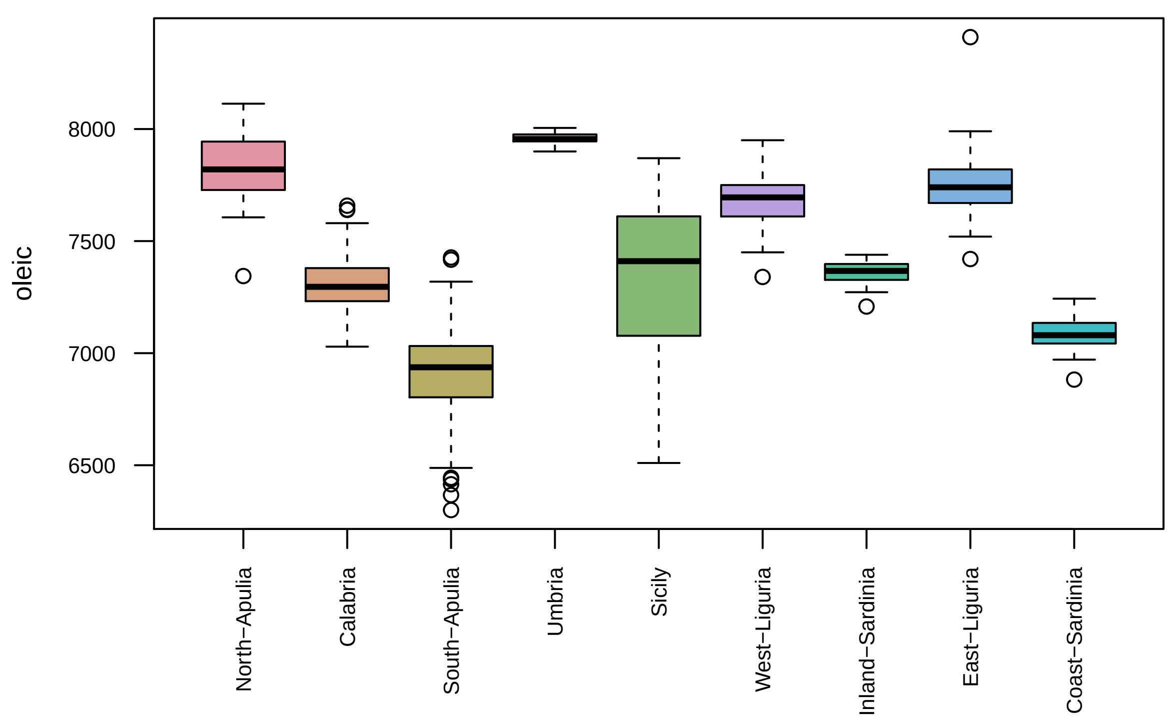
   i. It will have $\binom{9}{2}$ pairwise comparisons exist.

```
library(PairViz)
order <- eulerian(9)
area <- with(olive, split(oleic, Area))
boxplot(area[order],col=cols[order],las=2,cex.axis=0.7, xlab='',
        ylab='oleic')
```



ii.

```r
orderham <- hpaths(9)
boxplot(area[orderham[1,]],col=cols[orderham[1,]],las=2,
        cex.axis=0.7, ylab='oleic', cex.lab=0.8)
```

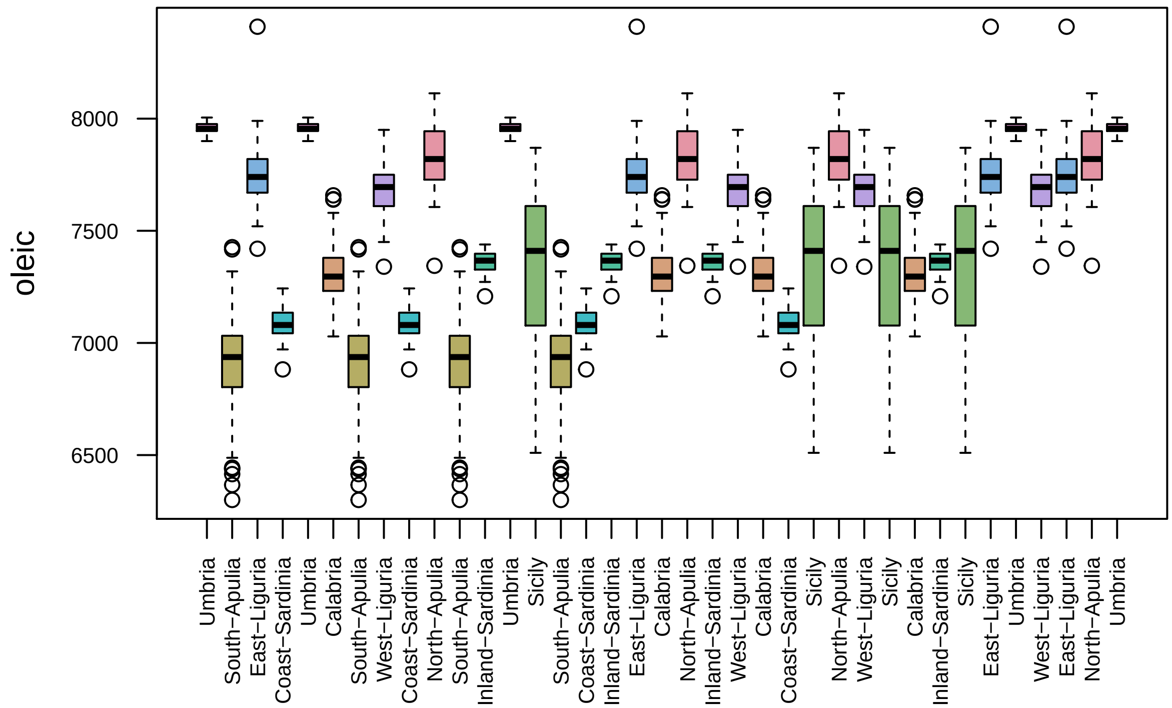


iii.

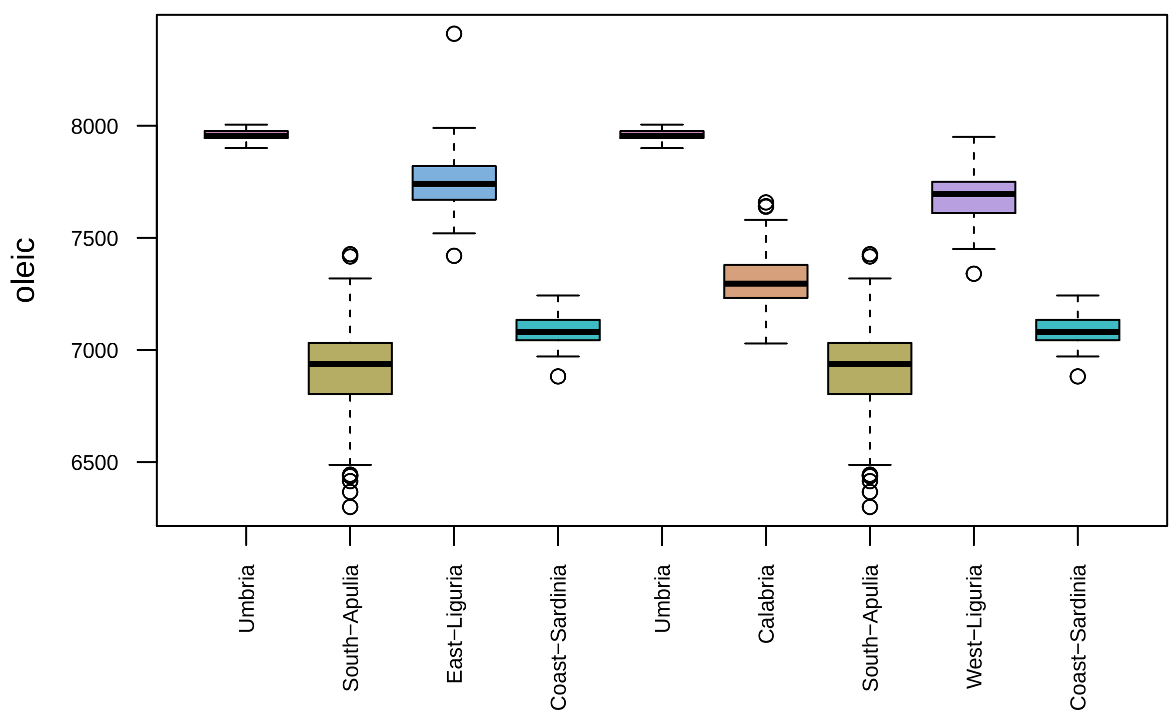
```r
test <-with(olive, pairwise.t.test(oleic, Area))
pvals <- test$p.value
areanames <- names(area)
ngrps <-length(areanames)
weights <-matrix(0, nrow=ngrps, ncol=ngrps)
rownames(weights) <- areanames
colnames(weights) <- areanames
weights[2:ngrps, 1:(ngrps-1)] <- pvals
diag(weights) <- 0
for (i in 1:(ngrps -1)) {
  for(j in (i+1):ngrps) {
    weights[i,j] <- weights[j,i]
  }
}
round(weights, 4)
```

```
##                 North-Apulia Calabria South-Apulia Sicily Inland-Sardinia
## North-Apulia          0.0000   0.0000            0 0.0000          0.0000
## Calabria              0.0000   0.0000            0 0.3058          0.2691
## South-Apulia          0.0000   0.0000            0 0.0000          0.0000
## Sicily                0.0000   0.3058            0 0.0000          0.9374
## Inland-Sardinia       0.0000   0.2691            0 0.9374          0.0000
## Coast-Sardinia        0.0000   0.0000            0 0.0000          0.0000
## East-Liguria          0.2691   0.0000            0 0.0000          0.0000
## West-Liguria          0.0024   0.0000            0 0.0000          0.0000
## Umbria                0.0053   0.0000            0 0.0000          0.0000
##                 Coast-Sardinia East-Liguria West-Liguria Umbria
## North-Apulia                 0       0.2691       0.0024 0.0053
## Calabria                     0       0.0000       0.0000 0.0000
## South-Apulia                 0       0.0000       0.0000 0.0000
## Sicily                       0       0.0000       0.0000 0.0000
## Inland-Sardinia              0       0.0000       0.0000 0.0000
## Coast-Sardinia               0       0.0000       0.0000 0.0000
## East-Liguria                 0       0.0000       0.1533 0.0000
## West-Liguria                 0       0.1533       0.0000 0.0000
## Umbria                       0       0.0000       0.0000 0.0000
```

```r
lowtohigh <- eulerian(weights)
boxplot(area[lowtohigh],col=cols[lowtohigh],las=2,cex.axis=0.7,
        xlab='', ylab='oleic')
```




```
boxplot(area[lowtohigh[1:9]],col=cols[lowtohigh[1:9]],las=2,cex.axis=0.7,
        xlab='', ylab='oleic')
```

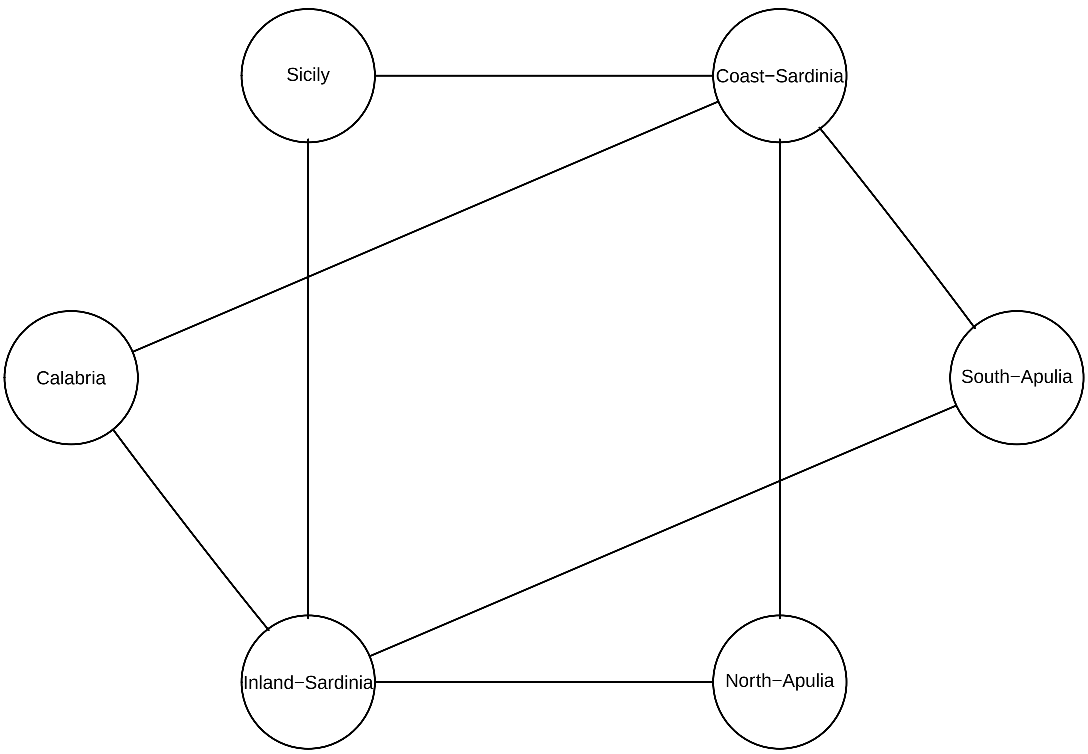


Comments: We can see that from left to right, significant differences tend to decrease. Therefore, the graph is right. The ordering is successful because We are able to find areas with huge differences quickly and we can see a clear pattern when moving from left to right.

c. The olive growing areas are divided into three different regions: North, South, and Sardinia. In this part of the question, interest lies only in comparisons between each growing area in the south and each area in Sardinia. That is, each southern area (4 areas) is to be compared to each Sardinian area (2 areas) yielding a total of 8 comparisons of interest.

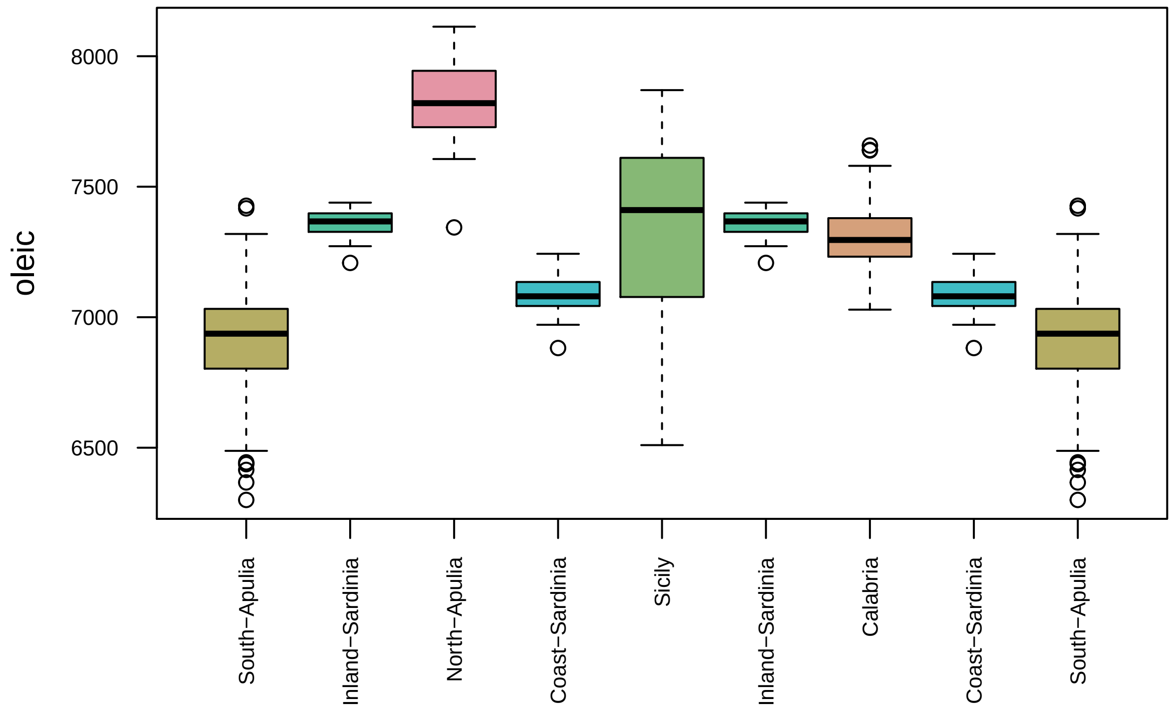
i.

```
g <- mk_complete_graph(weights[1:6,1:6])
g <- removeEdge('North-Apulia', 'Sicily', g)
g <- removeEdge('Calabria', 'Sicily', g)
g <- removeEdge('Coast-Sardinia', 'Inland-Sardinia', g)
g <- removeEdge('North-Apulia', 'South-Apulia', g)
g <- removeEdge('South-Apulia', 'Sicily', g)
g <- removeEdge('South-Apulia', 'Calabria', g)
g <- removeEdge('North-Apulia', 'Calabria', g)
plot(g, 'circo')
```



ii.

```
neword <- eulerian(g)
boxplot(area[neword],col=cols[match(neword, areanames)],las=2,cex.axis=0.7,
        xlab='', ylab='oleic')
```


oleic
8000
7500
7000
6500
South−Apulia
Inland−Sardinia
North−Apulia
Coast−Sardinia
Sicily
Inland−Sardinia
Calabria
Coast−Sardinia
South−Apulia