Olive data

Ordering displays

25 marks

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

```
library(loon)
```

```
## Warning: package 'loon' was built under R version 4.0.4
```

```
## Loading required package: tcltk
```

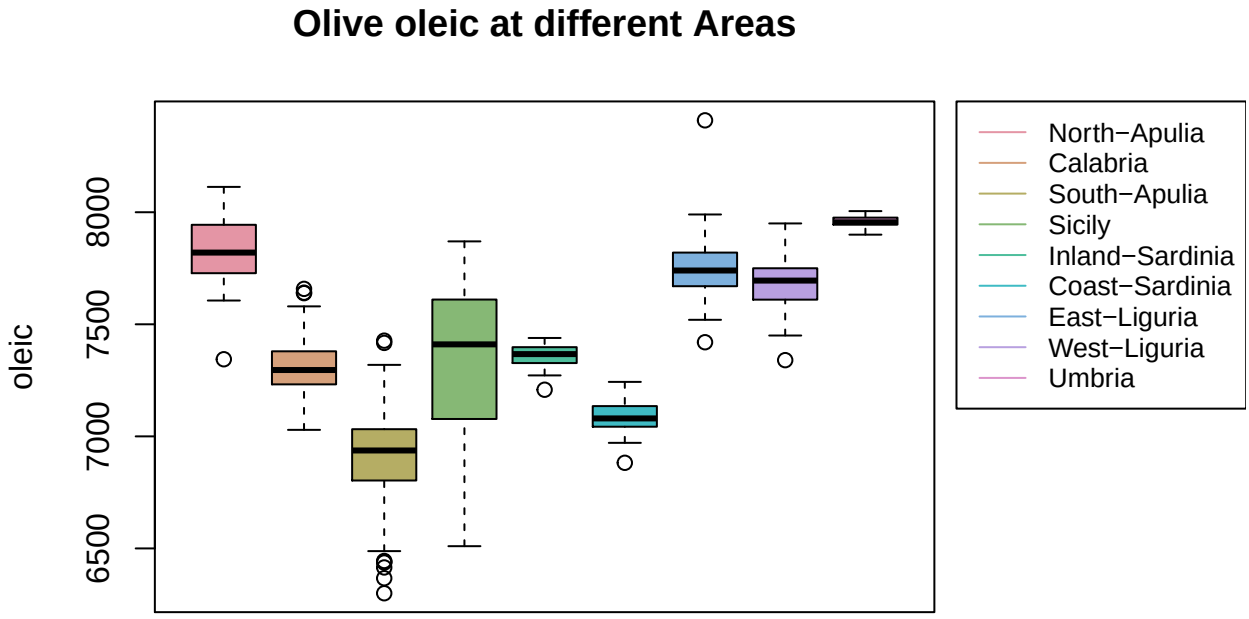
```
## loon Version 1.3.3.
## To learn more about loon, see l_web().
```

```
# 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
```

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.

data = split(olive$oleic, olive$Area)
names(cols)=names(data)
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
boxplot(data, col = cols, xaxt='n', main="Olive oleic at different Areas", ylab="oleic")
legend("topright", legend = names(cols), col=cols, lty=1,cex=0.8, inset=c(-0.4,0))
```



Show your code together with your output.

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:

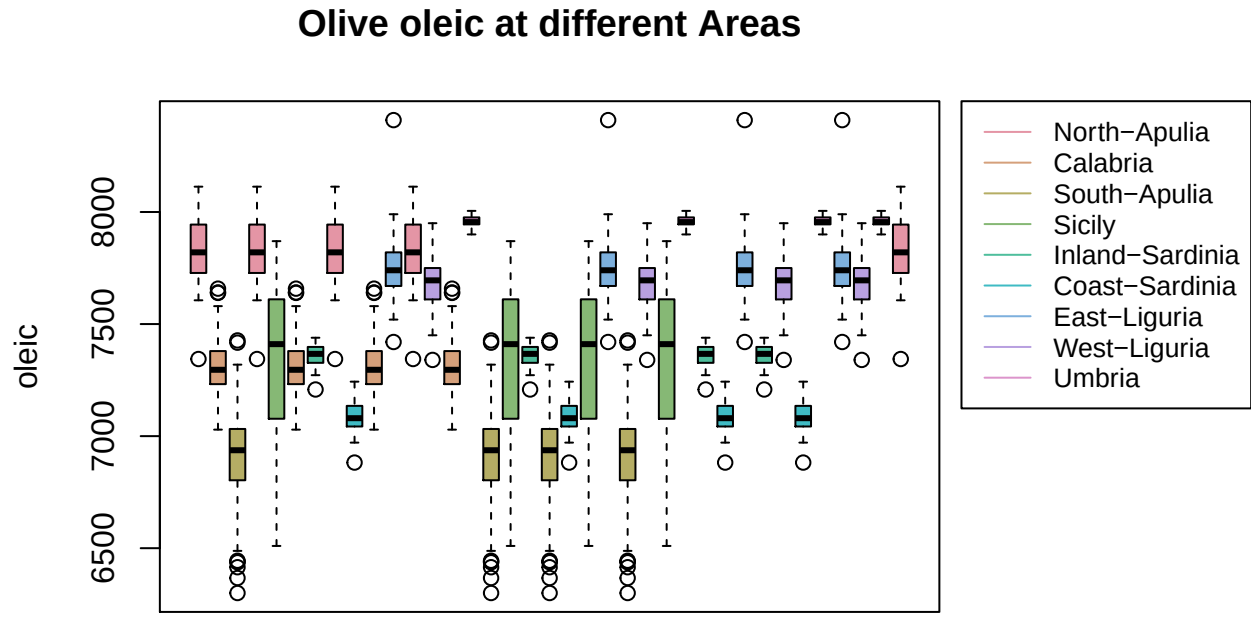
   i. **(3 marks)** Suppose we wish every pair of boxplots to appear next to one another in the same plot.
   - How many such pairwise comparisons exist?
   - Give the code that will construct this display (without any other constraint on the ordering).
   - Show the display which resulted from your code.

```
library(PairViz)
```

```
## Warning: package 'PairViz' was built under R version 4.0.4
## Loading required package: TSP
## Warning: package 'TSP' was built under R version 4.0.4
## Loading required package: gtools
## Loading required package: graph
## Warning: package 'graph' was built under R version 4.0.3
## Loading required package: BiocGenerics
## Warning: package 'BiocGenerics' was built under R version 4.0.3
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
```

```
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##     dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##     grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##     rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##     union, unique, unsplit, which.max, which.min

##
## Attaching package: 'graph'

## The following object is masked from 'package:loon':
##
##     complement
```

```
ord = eulerian(9)
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
boxplot(data[ord], col = cols[ord],xaxt="n", main="Olive oleic at different Areas", ylab="oleic")
legend("topright", legend = names(cols), col=cols, lty=1,cex=0.8, inset=c(-0.4,0))
```

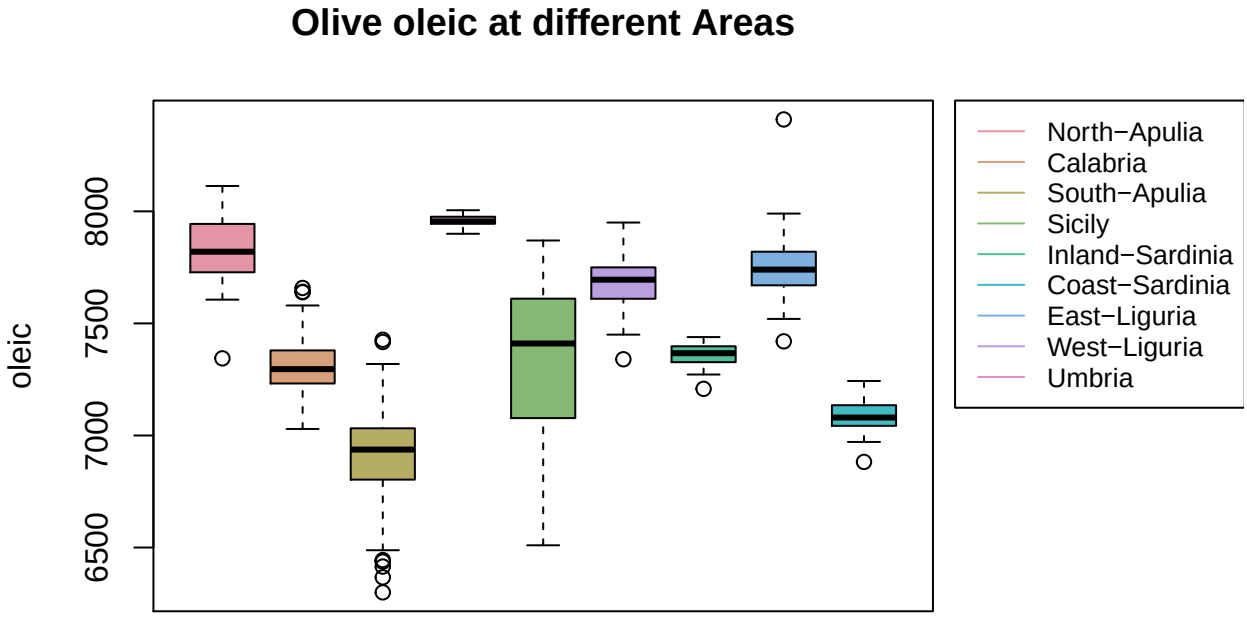


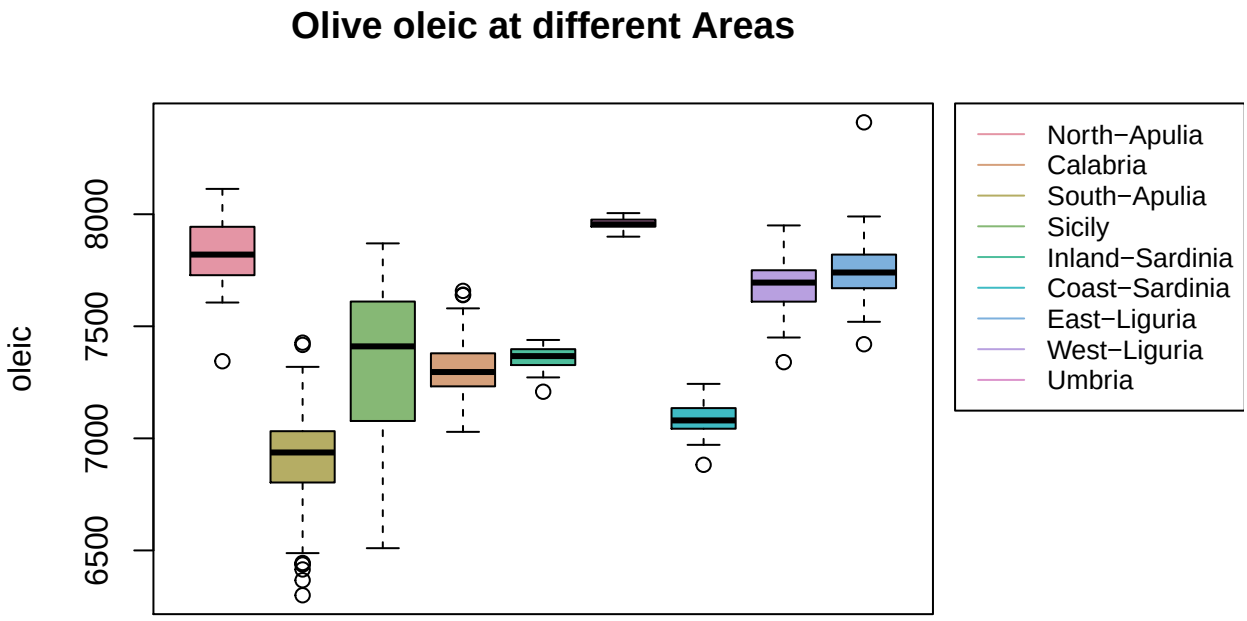
number of pairwise that exists is $\binom{9}{2} = 36$

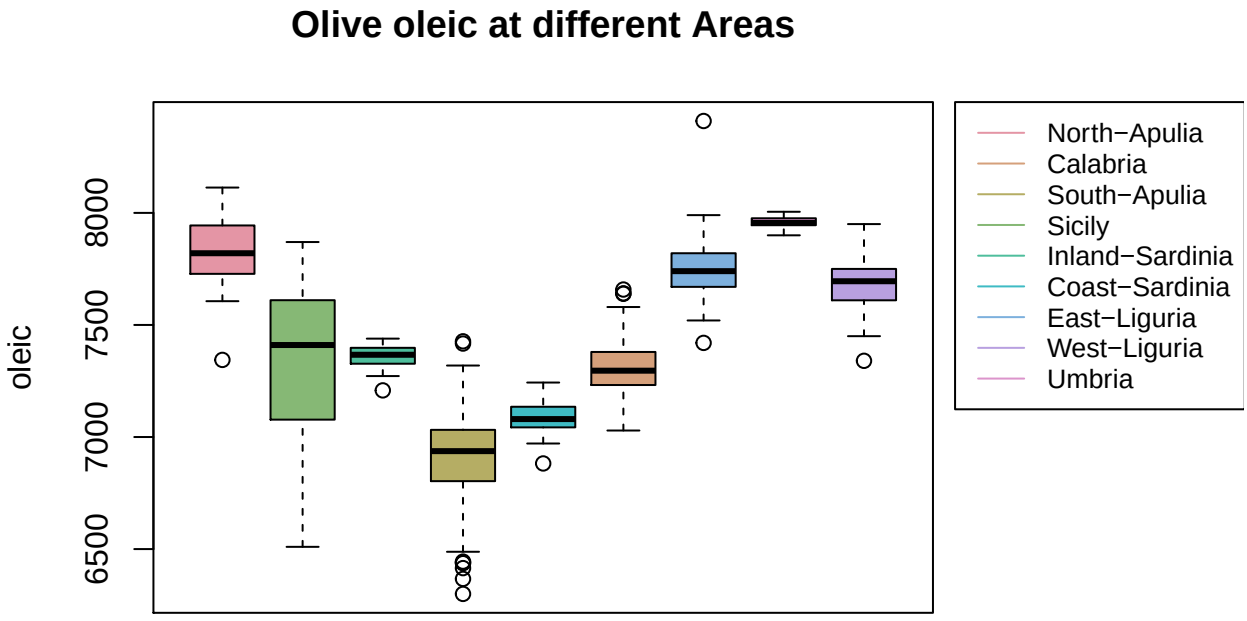
ii. **(5 marks)** Suppose we wish every pair of boxplots to appear next to one another in the same plot but
- Maintaining the same colours for the areas as before, give the code that will construct this display

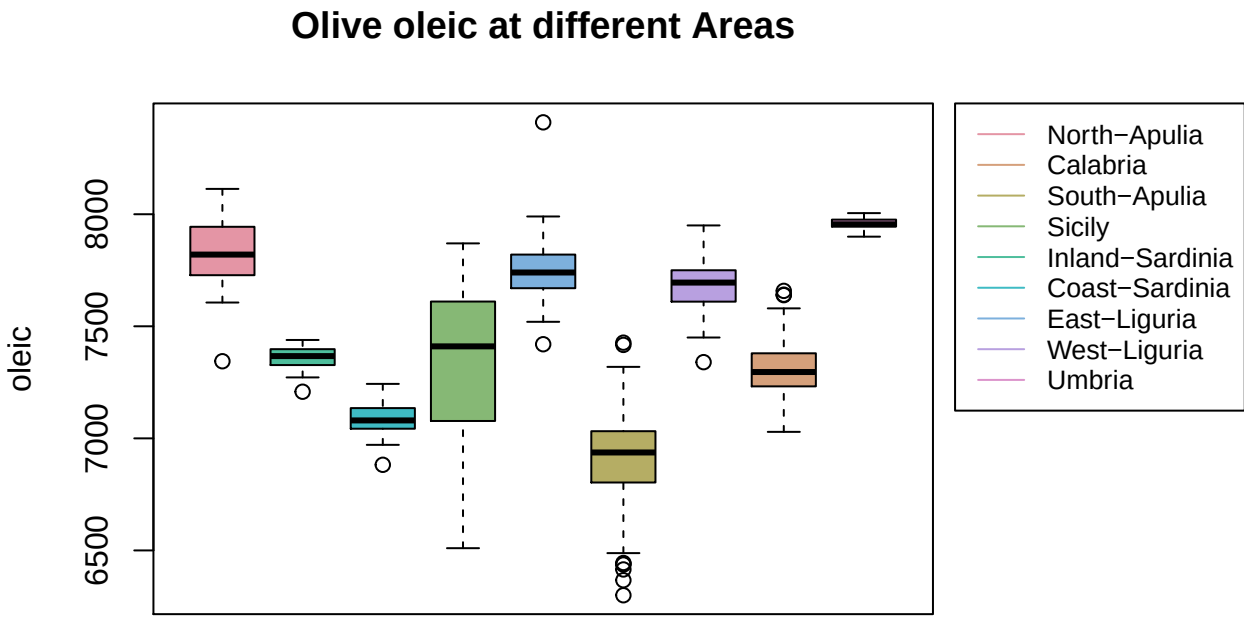
- Show the display which resulted from your code.

```
ordHam = hpaths(9)
for (i in 1:4){
  par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
  boxplot(data[ordHam[i,]], col = cols[ordHam[i,]],xaxt="n", main="Olive oleic at different Areas", ylab="o
  legend("topright", legend = names(cols), col=cols, lty=1,cex=0.8, inset=c(-0.4,0))
}
```



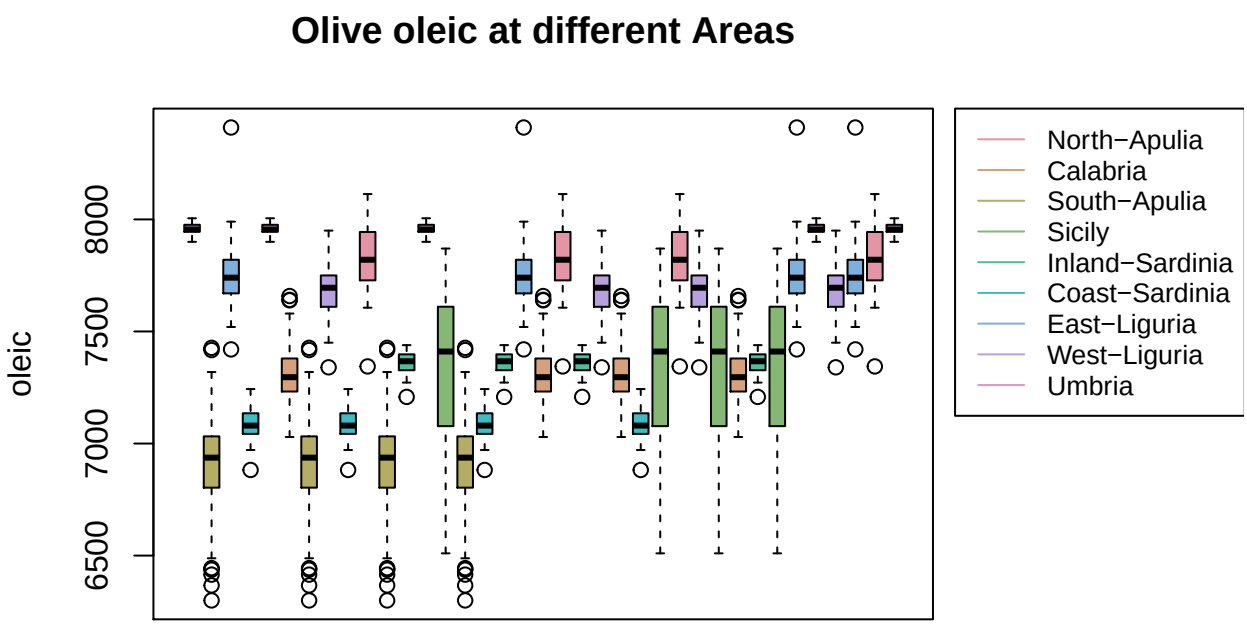

Olive oleic at different Areas
oleic
6500
7000
7500
8000
North−Apulia
Calabria
South−Apulia
Sicily
Inland−Sardinia
Coast−Sardinia
East−Liguria
West−Liguria
Umbria


Olive oleic at different Areas
oleic
6500
7000
7500
8000
North−Apulia
Calabria
South−Apulia
Sicily
Inland−Sardinia
Coast−Sardinia
East−Liguria
West−Liguria
Umbria

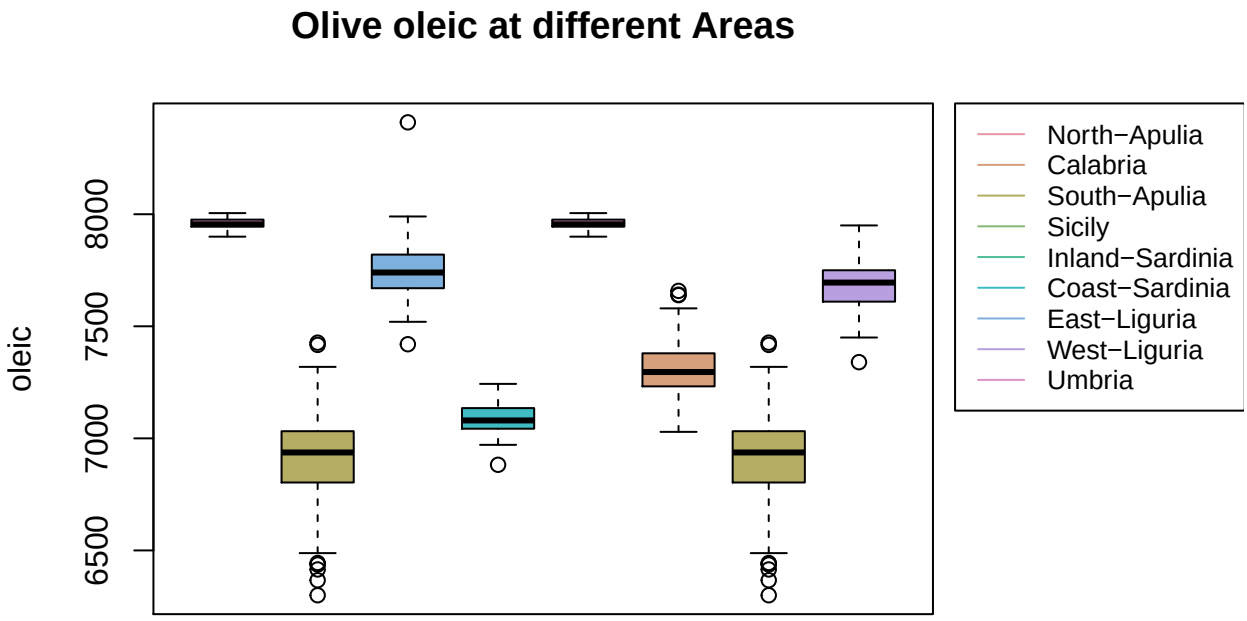


iii. **(7 marks)** Construct t tests for every pair of olive growing areas (recall `pairwise.t.test` fro
- Show your code.
- Show the resulting display.
- Does the ordering perfectly arrange the boxplots so that for any pairwise comparison, those to the le
- Explain why the ordering was successful (or unsuccessful) in this way.
- Show a display showing only the first 8 comparisons.

```
test = pairwise.t.test(olive$oleic, olive$Area)
pvals = test$p.value

Area.names = names(data)
narea = length(Area.names)

oleics = matrix(0, nrow=narea, ncol=narea)
rownames(oleics) = Area.names
colnames(oleics) = Area.names
oleics[2:narea, 1:(narea-1)] = pvals


diag(oleics) <- 0
for(i in 1:(narea -1)) {
  for(j in (i+1):narea) {
    oleics[i,j] <- oleics[j,i]
  }
}

low2high = eulerian(oleics)
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
boxplot(data[low2high],col =  cols[low2high], xaxt='n', main="Olive oleic at different Areas", ylab="ol
legend("topright", legend = names(cols), col=cols, lty=1,cex=0.8, inset=c(-0.4,0))
```




```
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
boxplot(data[low2high[1:8]], col = cols[low2high[1:8]], xaxt='n', main="Olive oleic at different Areas"
legend("topright", legend = names(cols), col=cols, lty=1,cex=0.8, inset=c(-0.4,0))
```



Does the ordering perfectly arrange the boxplots so that for any pairwise comparison, those to the left are more significant and those to the right are less significant?

No.

Explain why the ordering was successful (or unsuccessful) in this way.

Denote T score for H_0: area A and area B has the same oleic by T_{AB}

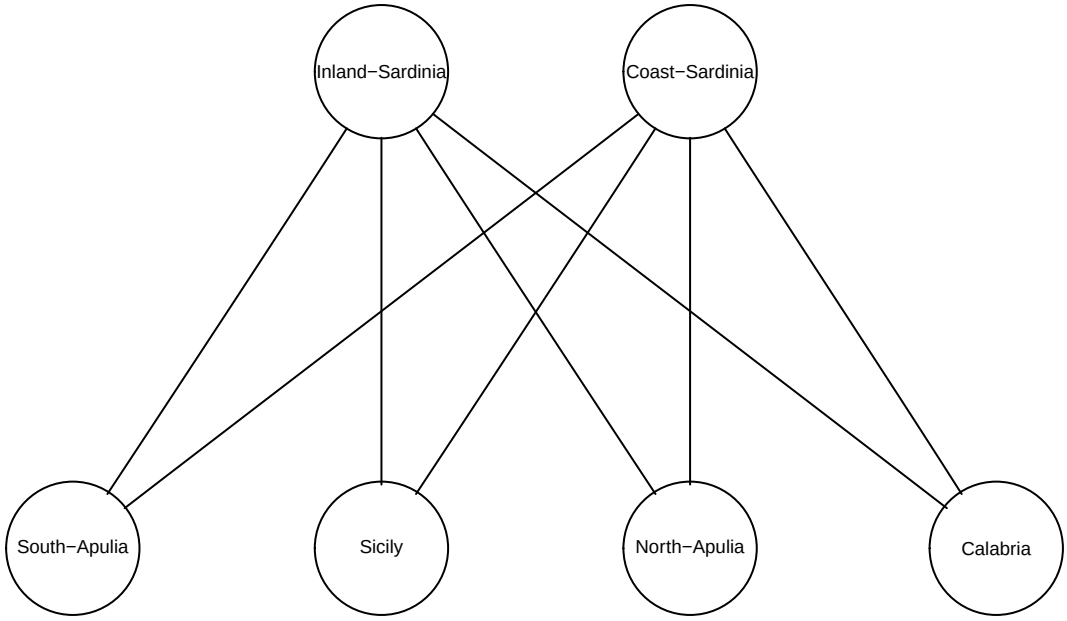
If smallest t-score is T_{AB} AND second smallest t-score is T_{CD}

Then, there is no way to put (boxplot of A or B) next to (boxplot of C or D) if ABCD are different Areas(which is entirely possible). Then what we actually put on the right of T_{AB} would be larger than T_{CD}, breaking the rule that comparison on the left are more significant.

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.

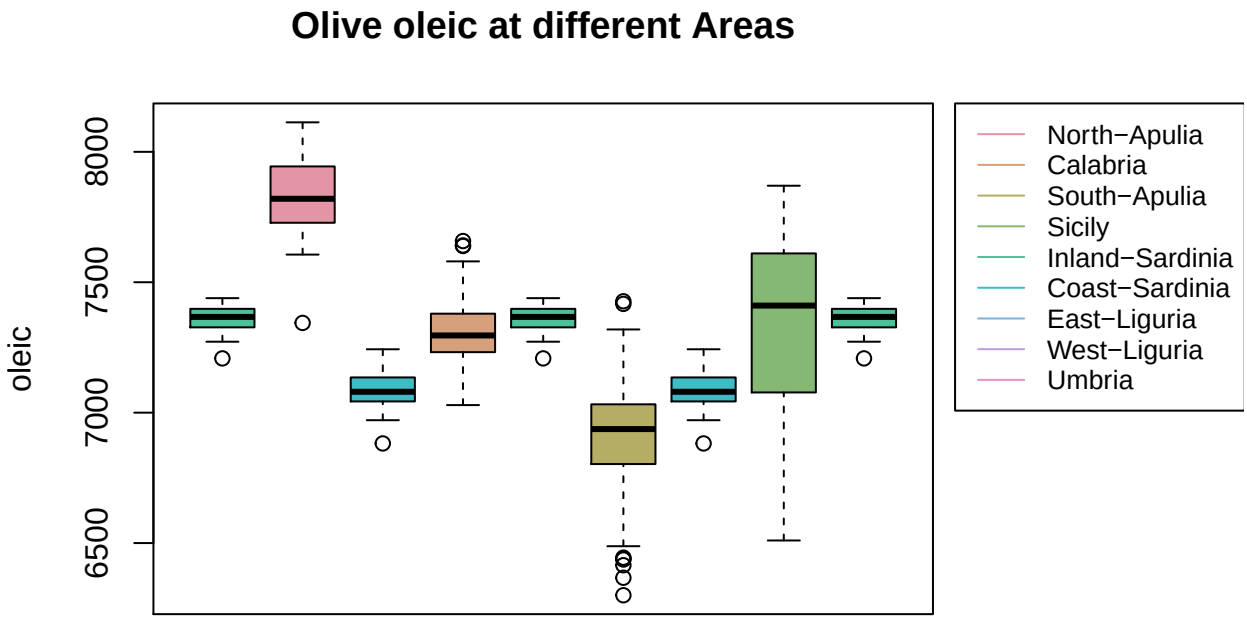
   i. **(4 marks)** Having loaded `PairViz`, create a graph having all six areas in the South and Sardinia as nodes and with edges between every pair whose comparison is of interest.
      - plot this graph
      - show the code used to create the graph and to plot it.

```
sdata = split(olive, olive$Region)
sdata.south = sdata$South
south.nodes = split(sdata.south$oleic, sdata.south$Area,drop = TRUE)
sdata.Sardinia= sdata$Sardinia
sardinia.nodes = split(sdata.Sardinia$oleic, sdata.Sardinia$Area,drop = TRUE)
p = bipartite_graph(names(sardinia.nodes), names(south.nodes))
plot(p)
```



ii. **(4 marks)** Using the graph from part (i), construct an Eulerian and use that Eulerian to produce a
    - show the boxplot display
    - show the code used to construct the Eulerian and the display.

```
ord = eulerian(p)
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
boxplot(data[ord], col=cols[ord], xaxt='n', main="Olive oleic at different Areas", ylab="oleic")
legend("topright", legend = names(cols), col=cols, lty=1,cex=0.8, inset=c(-0.4,0))
```


Olive oleic at different Areas
oleic
6500
7000
7500
8000
North−Apulia
Calabria
South−Apulia
Sicily
Inland−Sardinia
Coast−Sardinia
East−Liguria
West−Liguria
Umbria