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
```

Alternatively, if you do not have `loon`, the data are also on the course website in the data directory. You will need to download that file (`olive.Rda`) and read it into your `R` session as follows:

```
# Suppose you stored the file in a sub-directory "R" of the current directory "."
# Because it is an "Rda" file (R data), it is easily loaded
load("./R/olive.Rda")
# The data is now available as `olive' (just as in loon)
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.

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.

ii. **(5 marks)** Suppose we wish every pair of boxplots to appear next to one another in the same plot but that the boxplots should be grouped so that all areas appear only once in each group.
- Maintaining the same colours for the areas as before, give the code that will construct this display (without any other constraint on the ordering).
- Show the display which resulted from your code.

iii. **(7 marks)** Construct t tests for every pair of olive growing areas (recall `pairwise.t.test` from class). Use the significance levels from these tests to construct an ordering of the boxplot pairs, one which favours having the most significantly different pairs at the left of the display.
- Show your code. - Show the resulting display. - 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? - Explain why the ordering was successful (or unsuccessful) in this way. - Show a display showing only the first 8 comparisons.

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.

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