Nel file olive\_oil.xlsx ci sono dati relativi ad un data set di olii extra vergine di oliva italiani di diversa provenienza (1a colonna contiene l'indice di categoria): SA (sud puglia); NA (nord puglia); WL (liguria ovest); EL (liguria est); U (Umbria).

Questi sono stati analizzati mediante cromatografia e sono state determinate le concentrazioni relative in ciascun campione di olio di 7 acidi grassi (1 riga contiene i nomi).

Applicate i diversi strumenti di visualizzazione visti per il data set fisher iris , in particolare:

1. line plot (parralel coordinates)
2. gplotmatrix
3. istogrammi di frequenza per ciascuna variabile (7 in tutto)
4. boxplot distinti per ciascuna variabile evidenziando le categorie
5. imagesc dati tal quali e dati standardizzati
6. matrice di correlazione al quadrato (tra righe matrice)

Per le diverse rappresentazioni commentate i grafici rispetto a:

1. è possibile individuare una o più variabili in grado di distinguere alcune delle categorie? (specificare quali variabili e quali categorie)
2. Quali variabili da sole (diagonale) mostrano una distribuzione non normale che possa indicare la presenza di gruppi ? quali categorie sono distinguibili? Quali scatter plots (extradiagonali) mettono meglio in luce le diverse categorie o alcune di esse rispetto alle altre? Individuati un apio di scatter plot significativi fate (solo per questi) uno scatterhist. Ci sono variabili correlate (quali)?
3. Commentate le distribuzioni delle variabili e l'eventuale presenza di valori estremi
4. Quali variabili mostrano di poter distinguere le diverse categorie o alcune di esse?
5. - 6. Sono utili queste rappresentazioni per avere un idea della somiglianza tra campioni (righe della matrice)? Quale di più? Commentante le somiglianze che riuscite ad individuare.