



-
1. Johnson e Wichern - Exercício 12.3.
 2. Johnson e Wichern - Exercício 12.5.
 3. Johnson e Wichern - Exercício 12.6.
 4. Johnson e Wichern - Exercício 12.7.
 5. Johnson e Wichern - Exercício 12.10. - Pesquise sobre o método de Ward para agrupamento hierárquico e utilize o R para resolver o problema.

Importante: para os exercícios 12.11, 12.12 e 12.13 seguintes, é preciso fornecer o centro de cada grupo inicial no comando `kmeans` do R em vez do número de grupos. Exemplo: se queremos dois grupos e o centro inicial são as linhas 1 e 4 da matrix `x`, então definimos `kc = data.frame(x[c(1,4),])` e indicamos este centro em `kmeans(x,centers=kc)`.

6. Johnson e Wichern - Exercício 12.11.
7. Johnson e Wichern - Exercício 12.12.
8. Johnson e Wichern - Exercício 12.13.
9. Johnson e Wichern - Exercício 12.14. - arquivo `Brands-of-Cereal-T11-9.dat` no site do curso.
Nota: A distância Euclidiana é calculada considerando as medidas *calories*, *protein*, *fat*, *sodium*, *fiber*, *carbohydrates*, *sugar* e *potassium*.
10. Johnson e Wichern - Exercício 12.15.
11. Johnson e Wichern - Exercício 12.31. - arquivo `Crude-Oil-T11-7.dat` no site do curso.

Nota:

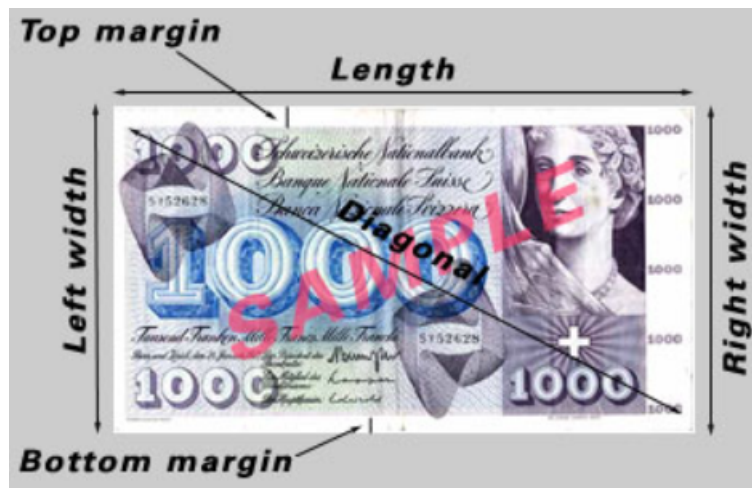
Grupos: Wilhelm, Sub-Mulinia, Upper. Medidas: $X_1 = \text{vanadium}$, $X_2 = \text{iron}$, $X_3 = \text{beryllium}$, $X_4 = \text{saturated hydrocarbons}$, $X_5 = \text{aromatic hydrocarbons}$.



12. Seis variáveis são medidas de 100 notas genuínas e 100 notas falsificadas (Flury and Riedwyl, 1988¹). Dados no pacote `gclus` através do comando `data(bank)` do R.

As colunas correspondem as seguintes variáveis:

1. Comprimento da nota (*Length*).
2. Altura da nota no lado esquerdo (*Left*).
3. Altura da nota no lado direito (*Right*).
4. Distância da moldura interna a borda superior (*Bottom*).
4. Distância da moldura interna a borda inferior (*Top*).
6. Comprimento da diagonal (*Diagonal*).



As 100 primeiras notas são genuínas e as 100 últimas são falsificadas.

- (a) Calcule as médias de cada variável em cada população (notas verdadeiras e falsificadas). Obtenha também a matriz de variância-covariâncias Σ em cada população. Obtenha um *face plot* de cada população utilizando os valores médios. Compare os resultados numéricos com os gráficos. Discuta.
Utilize o *face plot* para comparar os valores médios das variáveis para notas verdadeiras e notas falsas. Discuta os resultados.
- (b) Utilize um método de agrupamento para verificar se é possível distinguir as notas falsas das verdadeiras, imaginando que você não soubesse se 100 últimas notas eram verdadeiras.
- (c) Compare os resultados do agrupamento anterior com o agrupamento por mistura de normais.
- (d) Utilize o Índice Ajustado de Rand para comparar seu agrupamento das 200 notas com a verdadeira classificação em genuínas e falsas. Avalie a qualidade do agrupamento.

¹Referência: Flury, B. and Riedwyl, H. (1988). Multivariate Statistics, A Practical Approach, Cambridge University Press.