

Análise Multivariada Visualização de Dados

Professor: George von Borries

Departamento de Estatística
Universidade de Brasília

2023



Exemplo 1: Suponha que possuímos medidas (em polegadas) do peito, cintura e quadril de 20 indivíduos. Registramos também o sexo do indivíduo. Ver programa `medidas.R` no site do curso.

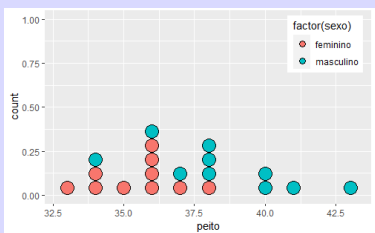
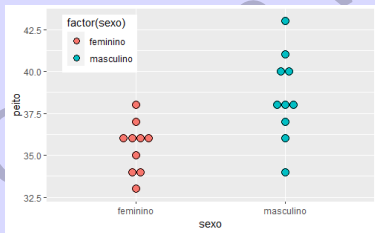
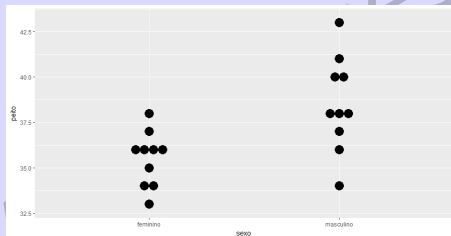
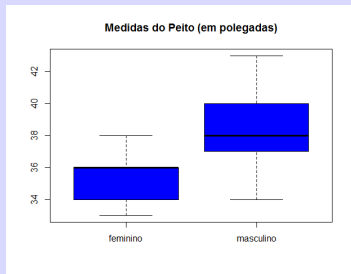
Obs	Peito	Cintura	Quadril	Sexo
1	34	30	32	masculino
2	37	32	37	masculino
3	38	30	36	masculino
⋮	⋮	⋮	⋮	⋮
18	36	26	37	feminino
19	38	28	40	feminino
20	35	23	35	feminino

(Nota: Exemplo retirado de Everitt, 2011.)

- Apresentamos a seguir alguns exemplos de representações gráficas dos dados.

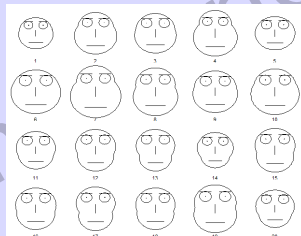
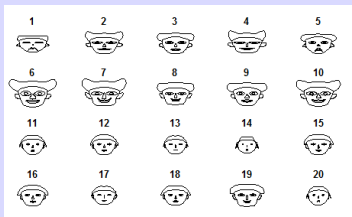


Representações gráficas univariadas: box-plot, gráfico de pontos, Ramo e folhas, Histograma.



Representações gráficas multivariadas:

Chernoff Faces veja Chernoff (1973).



Pacote TeachingDemos do R - gráficos das medidas (peito, cintura e quadril) de 20 pacientes. Comandos face e face2.

Os parâmetros implementados nos comandos face são:

1. altura da face, 2. largura da face, 3. formato da face, 4. altura da boca, 5. largura da boca,
6. curvatura do sorriso, 7. altura dos olhos, 8. largura dos olhos, ... (cabelo, nariz e orelha).



Exemplo 2: Iris Data contém dados de um gênero de plantas (IRIS) com flor, muito apreciado por suas diversas espécies, que apresentam flores de cores muito vivas. O nome popular é lírio.

O problema é classificar as flores nos três tipos denominados setosa, versicolor e virginica. As características disponíveis são: largura e comprimento da sépala (parte da flor que dá sustentação a pétala) e pétala ($p = 4$).



(a)



(b)



(c)

Três tipos de flores Iris: (a) setosa, (b) versicolor, (c) virginica.

Fonte: Murphy, K.P. *Machine Learning*, 2012.



Andrews Plot: gráfico que permite descobrir a existência de grupos. O gráfico de Andrews (David F. Andrews, 1972) permite as dimensões através de séries de Fourier Finitas ($-\pi < t < \pi$),

$$f(t) = \frac{x_1}{\sqrt{2}} + x_2 \times \text{sen}(t) + x_3 \times \cos(t) + x_4 \times \text{sen}(2t) + x_5 \times \cos(2t) + \dots,$$

O R produz este gráfico através da função Andrews do pacote Andrews.

Funções disponíveis:

$$1 : f(t) = \frac{x_1}{\sqrt{2}} + x_2 \times \text{sen}(t) + x_3 \times \cos(t) + x_4 \times \text{sen}(2t) + x_5 \times \cos(2t) + \dots$$

$$2 : f(t) = x_1 \times \text{sen}(t) + x_2 \times \cos(t) + x_3 \times \text{sen}(2t) + x_4 \times \cos(2t) + \dots$$

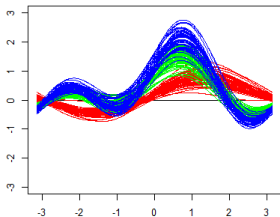
$$3 : f(t) = x_1 \times \cos(t) + x_2 \times \cos((2t)^{0.5}) + x_3 \times \cos(3t^{0.5}) + \dots$$

$$4 : f(t) = \frac{x_1}{20.5} + x_2 \times (\text{sen}(t) + \cos(t)) + x_3 \times (\text{sen}(t) - \cos(t)) \\ + x_4 \times (\text{sen}(2t) + \cos(2t)) + \dots$$

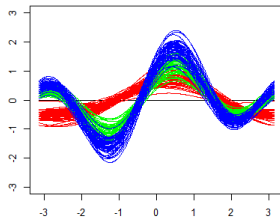


Andrews Plot (data Iris)

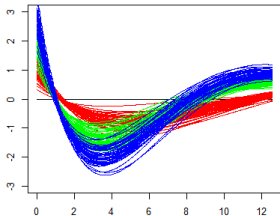
Andrews Plot 1



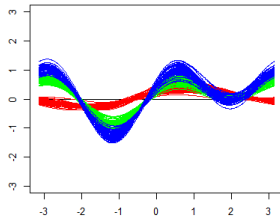
Andrews Plot 2



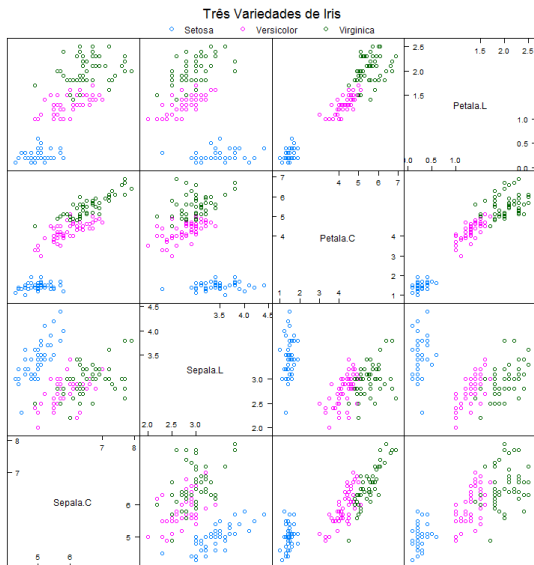
Andrews Plot 3



Andrews Plot 4



Matriz de Dispersão



O R produz gráficos de correlação bem interessantes.

Ver <https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>

Gráfico de Perfis ou *Parallel Plot*

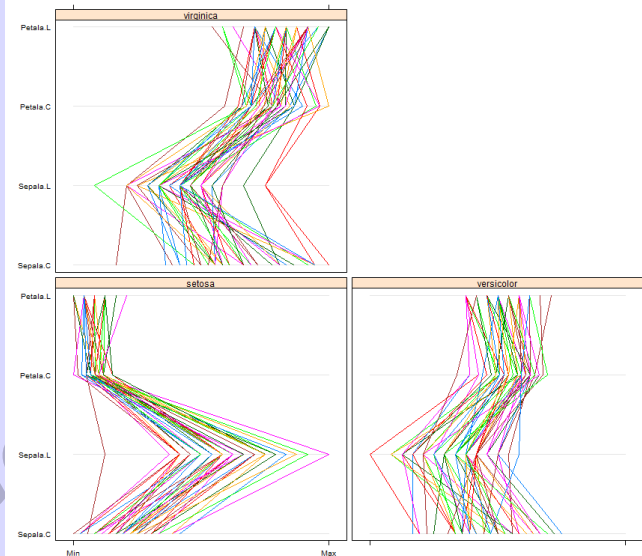


Gráfico de Perfis ou *Parallel Plot*

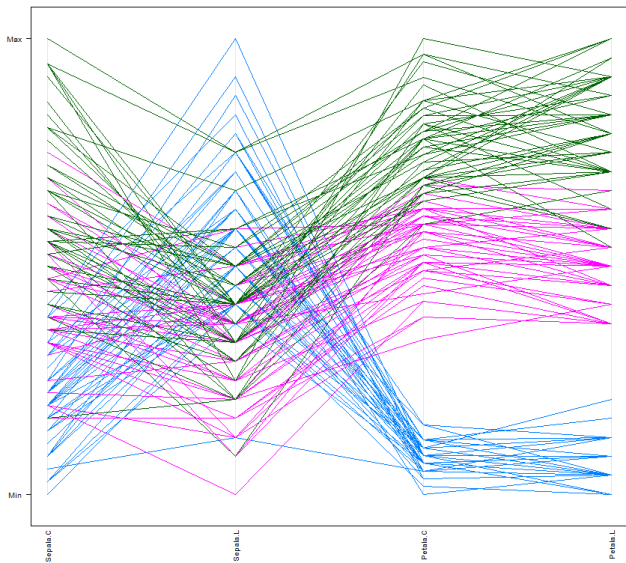
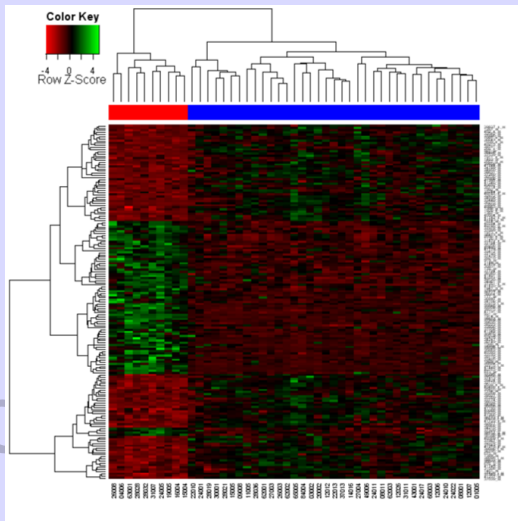


Gráfico de Calor ou *Heatmap*



Fonte: https://warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/heatmap



Gráfico de Calor ou *Heatmap*

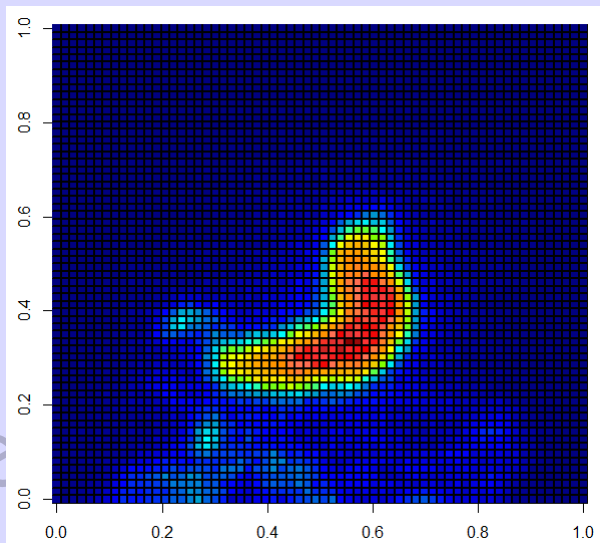


Imagem de Scintigrafia do Estomago.



Podemos ver uma foto como um Gráfico de Calor?

