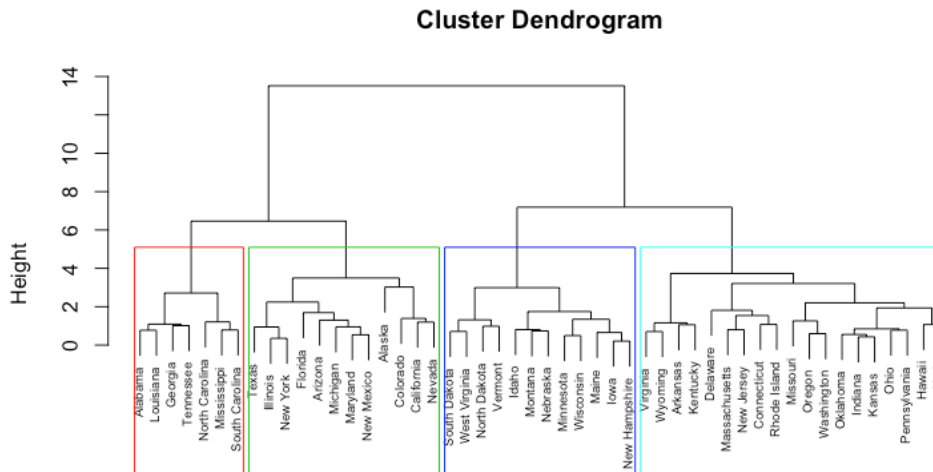


LISTA DE EXERCÍCIOS 4

1. Considere o conjunto de dados da tabela abaixo. Apresente os gráficos de coordenadas paralelas obtidos considerando: (a) os valores na tabela dos dados; (b) os valores após um pré-processamento de normalização (min-max) de todos os atributos numéricos; (c) os valores após um processamento de *standardization* dos mesmos atributos.

| Produto | Preço | Peso | Volume | Quantidade |
|---------|-------|------|--------|------------|
| P10 | 10 | 70 | 100 | 8 |
| P15 | 12 | 20 | 30 | 18 |
| P03 | 30 | 15 | 30 | 20 |
| P20 | 30 | 10 | 200 | 10 |
| P07 | 10 | 60 | 50 | 8 |
| P22 | 9 | 30 | 40 | 20 |
| P31 | 15 | 30 | 35 | 30 |
| P18 | 10 | 70 | 100 | 5 |
| P30 | 20 | 20 | 30 | 3 |

2. Descreva as conclusões que podem ser inferidas sobre os dados observando as visualizações criadas no exercício anterior.
3. Sugira uma estratégia de ordenação de eixos em Coordenadas Paralelas que seja efetiva para minimizar os cruzamentos entre as poli-linhas.
4. Explique de que maneira a visualização de dados multivariados por *heatmaps* é afetada pela presença de variáveis em escalas de magnitude distintas. Qual o problema, e como ele se manifesta na visualização? Como esse problema pode ser tratado para gerar uma visualização mais adequada? Discuta também o caso da presença de *outliers* nos dados.
5. Considerando uma tabela $X_{n \times m}$ de dados multivariados, explique como funciona um algoritmo de *clustering* hierárquico aglomerativo aplicado a esses dados. Quais os parâmetros do algoritmo, e qual a saída gerada?
6. No *clustering* hierárquico aglomerativo, explique qual a diferença entre as estratégias de aglomeração “complete linkage”, “average linkage” e “single linkage”.
7. Considere o dendograma na figura a seguir: detalhe os agrupamentos descritos considerando *height* = 10, 4, 2 e 1.



8. Descreva duas métricas adequadas para medir a qualidade do resultado gerado por um algoritmo de *clustering*. Apresente como elas são computadas, o *range* de valores possíveis e como esses valores são interpretados.
9. Sugira uma maneira alternativa ao dendrograma para exibir os resultados de um algoritmo de clustering hierárquico. Ilustre a sua solução mostrando a visualização do clustering apresentado no dendrograma da figura.
10. Porque as visualizações por *heatmaps* são tipicamente associadas a um clustering hierárquico dos dados? Explique.
11. Discuta as vantagens e limitações das visualizações por *radar plots*, em relação a Coordenadas Paralelas. Pode-se afirmar que essas técnicas têm objetivos similares em algum aspecto? De que maneiras elas diferem?
12. Explique no que consiste o problema da ambiguidade visual na visualização por RadViz, e porque ele ocorre.
13. Faça uma análise comparativa de das características das técnicas de visualização multivariada RadViz, em relação a Coordenadas Paralelas e Heatmaps, mostrando em uma tabela comparativa suas propriedades, que informações são capazes de revelar, que tipo de observações favorecem, que tipo de dado de entrada admitem. Em seguida, apresente os pontos fortes e fracos de cada uma, e que tipo de aplicações elas se mostram adequadas para apoiar.