

Data Mining

Análise de Grupos III

Prof. Dr. Joaquim Assunção

DEPARTAMENTO DE COMPUTAÇÃO APLICADA
CENTRO DE TECNOLOGIA
UFSM
2023

www.inf.ufsm.br/~joaquim



Fair user agreement

Este material foi criado para a disciplina de Mineração de Dados - Centro de Tecnologia da UFSM.

Você pode usar este material livremente*; porém, caso seja usado em outra instituição, **me envie um e-mail** avisando o nome da instituição e a disciplina.

*A maior parte deste material foi retirado do livro: “[Joaquim V. C. Assunção. Uma Breve Introdução à Mineração de Dados: Bases Para a Ciência de Dados, com Exemplos em R. 192 páginas. Novatec. 2021. ISBN-10 : 6586057507.](#)”

Prof. Dr. Joaquim Assunção.
joaquim@inf.ufsm.br

Agrupamento hierárquico

- Útil quando queremos ver:
 - Como os n grupos se formam ou
 - Qual a proximidade dos n grupos ou
 - Qual a hierarquia entre os grupos.

Agrupamento hierárquico

- O dendrograma é a representação mais comum em um agrupamento hierárquico.
 - Nele é possível ver não somente as instâncias agrupadas, mas também a ordem em que cada união ocorreu.
- Todo processo começa dividindo o conjunto de n registros em n grupos.
 - Um por um, os registros são agrupados até que um único grupo seja formado.

Agrupamento hierárquico

- Nesta categoria é importante diferenciar o tipo de método usado para métricas de distância dentre os registros:
 - Acoplamento único (single-linkage),
 - Acoplamento completo (complete-linkage) e
 - Baseada no centroide.
- A distância em que as ligações são feitas representa a ordem de criação dos grupos.

Agrupamento hierárquico - Exemplo

Tabela 7.1: Distância aérea entre capitais brasileiras.

	Porto Al.	Curitiba	Rio de Jan.	Brasília	Fortaleza	Natal
Porto Al.	0	546	1131	1611	3216	3174
Curitiba		0	676	1081	2672	2646
Rio de Jan.			0	930	2191	2086
Brasília				0	1687	1773
Fortaleza					0	433
Natal						0

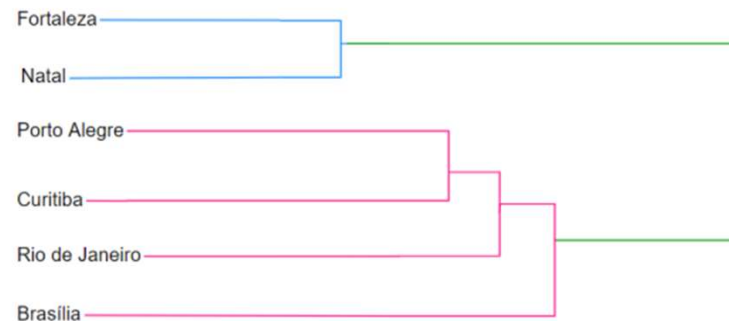
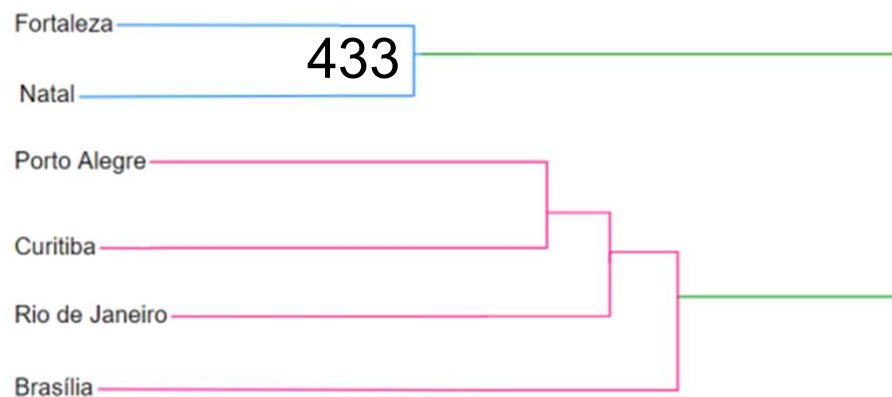


Figura 7.12: Exemplo de dendrograma para agrupamento hierárquico, single-linkage da Tabela 7.1.

Agrupamento hierárquico - Exemplo

Tabela 7.1: Distância aérea entre capitais brasileiras.

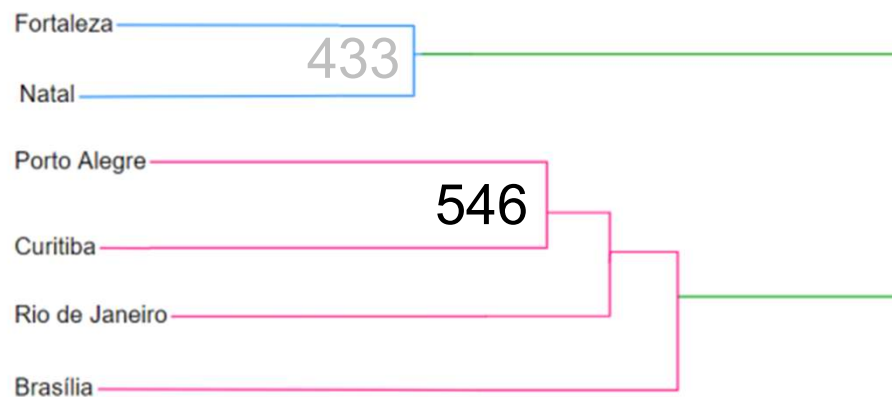
	Porto Al.	Curitiba	Rio de Jan.	Brasília	Fortaleza	Natal
Porto Al.	0	546	1131	1611	3216	3174
Curitiba		0	676	1081	2672	2646
Rio de Jan.			0	930	2191	2086
Brasília				0	1687	1773
Fortaleza					0	433
Natal						0



Agrupamento hierárquico - Exemplo

Tabela 7.1: Distância aérea entre capitais brasileiras.

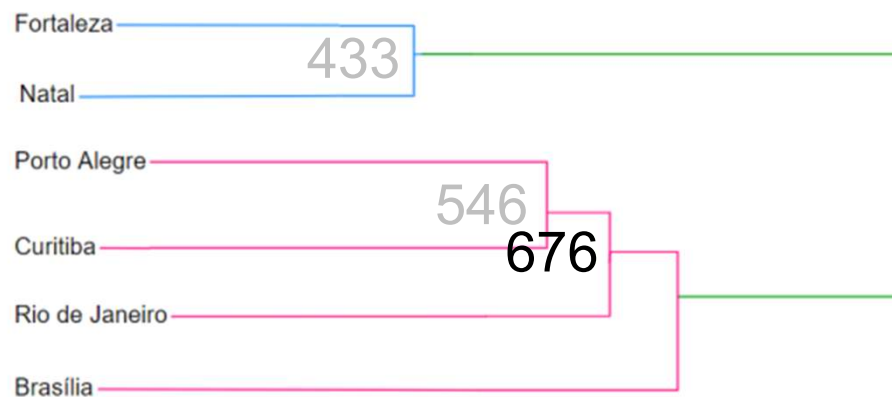
	Porto Al.	Curitiba	Rio de Jan.	Brasília	Fortaleza	Natal
Porto Al.	0	546	1131	1611	3216	3174
Curitiba		0	676	1081	2672	2646
Rio de Jan.			0	930	2191	2086
Brasília				0	1687	1773
Fortaleza					0	433
Natal						0



Agrupamento hierárquico - Exemplo

Tabela 7.1: Distância aérea entre capitais brasileiras.

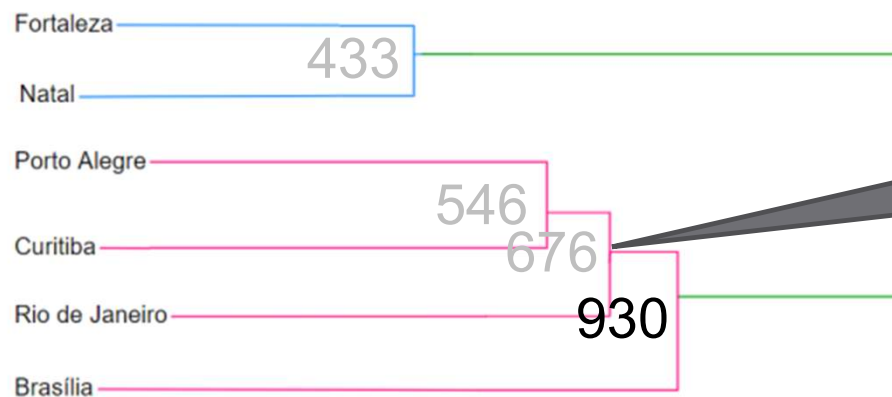
	Porto Al.	Curitiba	Rio de Jan.	Brasília	Fortaleza	Natal
Porto Al.	0	546	1131	1611	3216	3174
Curitiba		0	676	1081	2672	2646
Rio de Jan.			0	930	2191	2086
Brasília				0	1687	1773
Fortaleza					0	433
Natal						0



Agrupamento hierárquico - Exemplo

Tabela 7.1: Distância aérea entre capitais brasileiras.

	Porto Al.	Curitiba	Rio de Jan.	Brasília	Fortaleza	Natal
Porto Al.	0	546	1131	1611	3216	3174
Curitiba		0	676	1081	2672	2646
Rio de Jan.			0	930	2191	2086
Brasília				0	1687	1773
Fortaleza					0	433
Natal						0

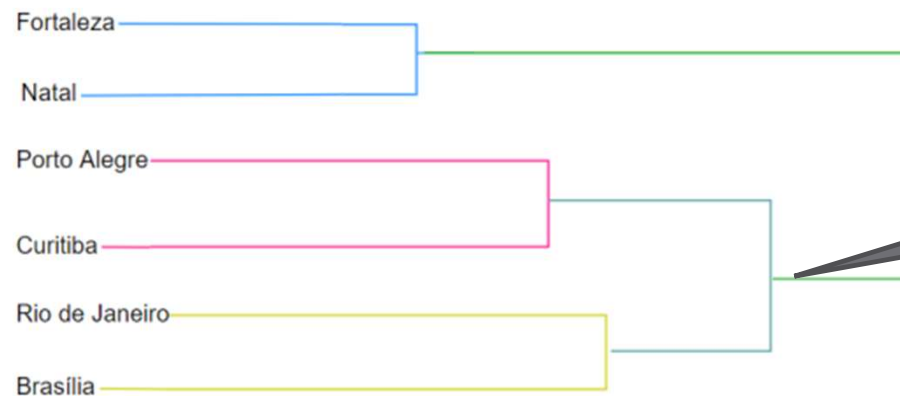


Medida por
Single-linkage

Agrupamento hierárquico - Exemplo

Tabela 7.1: Distância aérea entre capitais brasileiras.

	Porto Al.	Curitiba	Rio de Jan.	Brasília	Fortaleza	Natal
Porto Al.	0	546	1131	1611	3216	3174
Curitiba		0	676	1081	2672	2646
Rio de Jan.			0	930	2191	2086
Brasília				0	1687	1773
Fortaleza					0	433
Natal						0



Medida por
Complete-linkage

Technical demo

Captura os dados

```
> acoes_fecha2019 <- read.csv(url("http://www-usr.inf.ufsm.br/~joaquim/  
book_datasets/acoes_close2019.csv"))
```

Transforma em matriz e
transpõe para o formato
desejado.

```
> acoes_fecha2019 = t(as.matrix(acoes_fecha2019))
```

Do it! Veja o header dos dados.

Technical demo

- 1 - Coleta as distâncias pela correlação.
- 2 - Forma o cluster hierárquico.
- 3 - ...

```
> acoes_fecha2019 <- read.csv(url("http://www-usr.int.utsn.br/~joaquim/br_datasets/acoes_close2019.csv"))
```

```
> acoes_fecha2019 = t(as.matrix(acoes_fecha2019))
```

```
acoes_fecha2019.cor <- as.dist(1- cor(t(acoes_fecha2019)))  
HC_acoes.cor <- hclust(acoes_fecha2019.cor, method = "complete",  
                      members=NULL)  
plot(HC_acoes.cor, ylab="Altura", xlab = "Grupos de ações", main=NULL)
```

Do it! Escreva o código e veja o gráfico gerado.