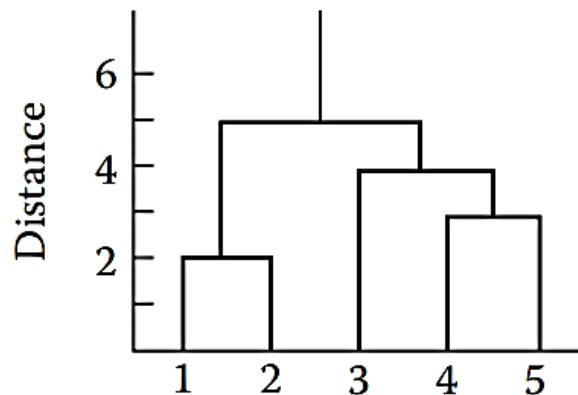


Análise de Agrupamentos

Jhessica Letícia Kirch
Universidade de São Paulo

Simpósio de Microbiologia Agrícola



Aplicação da Análise de Componentes Principais e de Agrupamento para os Indicadores de Desempenho das Universidades Federais do Brasil

Jhessica Letícia Kirch^{1†}, Rubiane P. Schoenherr², Tereza Christina M. A. Veloso³, Kuang Hongyu⁴

¹Programa de Pós-graduação em Estatística e Experimentação Agronômica, Universidade de São Paulo - ESALQ/USP.

²Universidade Federal de Mato Grosso - UFMT. email: rubi_prado@hotmail.com.

³Departamento de Alimento e Nutrição e Programa de Pós-Graduação em Educação, Universidade Federal de Mato Grosso - UFMT. email: tecmapv@terra.com.br.

⁴Departamento de Estatística, Instituto de Ciências Exatas e da Terra, Universidade Federal de Mato Grosso - UFMT. email: kuang_hongyu@hotmail.com.

Resumo: A decisão número 408/2002 do Tribunal de Contas da União determinou que as Instituições Federais de Ensino Superior (IFES) incorporassem em seus Relatórios de Gestão nove indicadores para o acompanhamento do desempenho e aprimoramento da gestão das instituições federais. O objetivo desta pesquisa é analisar os indicadores de desempenho das universidades federais do Brasil no ano de 2015 por meio da técnica estatística multivariada Análise de Componentes Principais e de Agrupamento, permitindo assim uma redução da quantidade de variáveis de estudo e o agrupamento das universidades federais do Brasil de acordo com seus indicadores de desempenho. Através dos resultados da Análise de Componentes Principais (ACP) foi possível a formação de três componentes principais que, juntos, explicavam 74,20% da variação total dos dados. O primeiro componente principal explicou 39,22%, o segundo, 22,86% e o terceiro componente principal explicou 12,12% da variação total dos dados. Observou-se ainda, pela Análise de Agrupamento, a formação de nove grupos, ou clusters, das universidades federais analisadas. Estes clusters se mostraram explicativos em relação a ACP, unindo as IFES que são similares em seus valores para cada um dos componentes principais analisados.

Palavras-chave: Análise de Agrupamento; Análise de Componentes Principais; Ensino Superior; Indicadores de desempenho.

USOS DE ANÁLISE DE AGRUPAMENTOS

Idéia: usar os valores das p variáveis para agrupar os n objetos em classes de modo que objetos similares estejam na mesma classe.

- O método precisa ser numérico;
- O número de classes não é, em geral, conhecido;
- Os diferentes grupos formados devem ser heterogêneos entre si, mas homogêneos dentro de grupo.

USOS DE ANÁLISE DE AGRUPAMENTOS

- Se resume na identificação de grupos de **objetos similares**.
- Não faz sentido utilizar a análise de agrupamento com dados como dos pardais, pois os grupos já são conhecidos.
- Exemplo: cães pré-históricos.

MÉTODOS DE AGRUPAMENTO

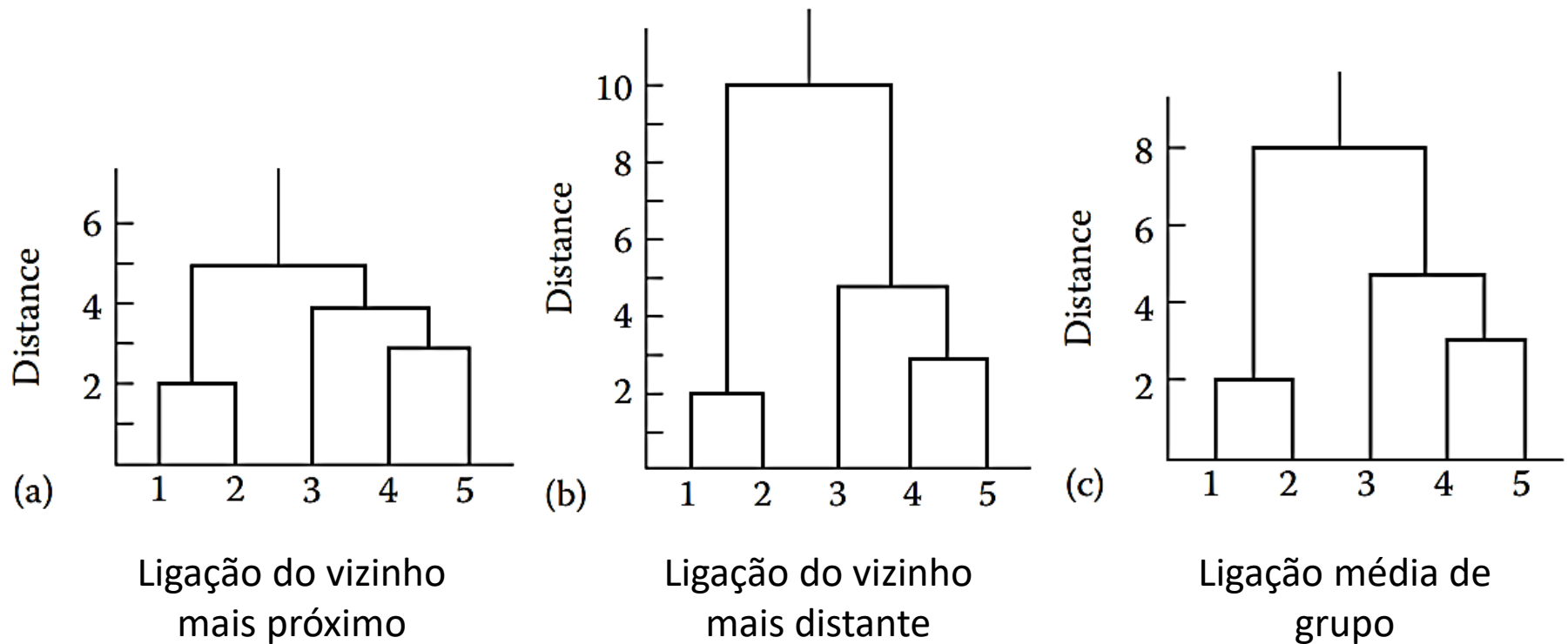
- Os métodos de agrupamento se dividem em:
 - **Hierárquicos:** são técnicas em que os dados são particionados sucessivamente, produzindo um dendrograma.
 - **Não hierárquico:** partição de objetos que podem se mover para dentro e para fora de grupos em diferentes momentos da análise.

MÉTODOS HIERÁRQUICOS

- Os métodos hierárquicos são os mais utilizados em análise de agrupamento.
- Alguns dos principais métodos são:
 - Método de ligação simples (vizinho mais próximo);
 - Método de ligação completa (vizinho mais distante);
 - Método de ligação média.

MÉTODOS HIERÁRQUICOS

Figura 1 Exemplos de dendrogramas de análise de agrupamentos de cinco objetos



MÉTODOS HIERÁRQUICOS

- Todos esses métodos operam sobre uma matriz de distâncias.

DISTÂNCIA UNIARIADA

Considere dois indivíduos nos quais foi medida a variável quantitativa X . Naturalmente, uma medida de distância entre esses indivíduos seria:

$$d_{12} = x_1 - x_2$$

Ou ainda:

$$d_{12} = |x_1 - x_2|$$

$$d_{12} = \sqrt{(x_1 - x_2)^2}$$

DISTÂNCIA MULTIVARIADA

No caso em que p variáveis quantitativas foram medidas, pode-se generalizar para o caso multivariado, conhecida como **distância euclidiana**:

$$d_{ij} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{i'j})^2}$$

DISTÂNCIA MULTIVARIADA

- Uma desvantagem dessa distância é que as variáveis **não tem o mesmo peso!**
- As variáveis cujas unidades de medidas são mais altas têm maiores pesos no cálculo da distância.
- Esse problema pode ser contornado pela **padronização** de variáveis.

PADRONIZAÇÃO DE VARIÁVEIS

- Em muitos métodos multivariados, a padronização de variáveis é um procedimento útil e, muitas vezes, necessário para eliminar influências das diferentes unidades de medida das variáveis nos resultados.
- A principal forma de padronização de variáveis é:

$$z = \frac{x - \mu}{\sigma}$$

DISTÂNCIA MULTIVARIADA

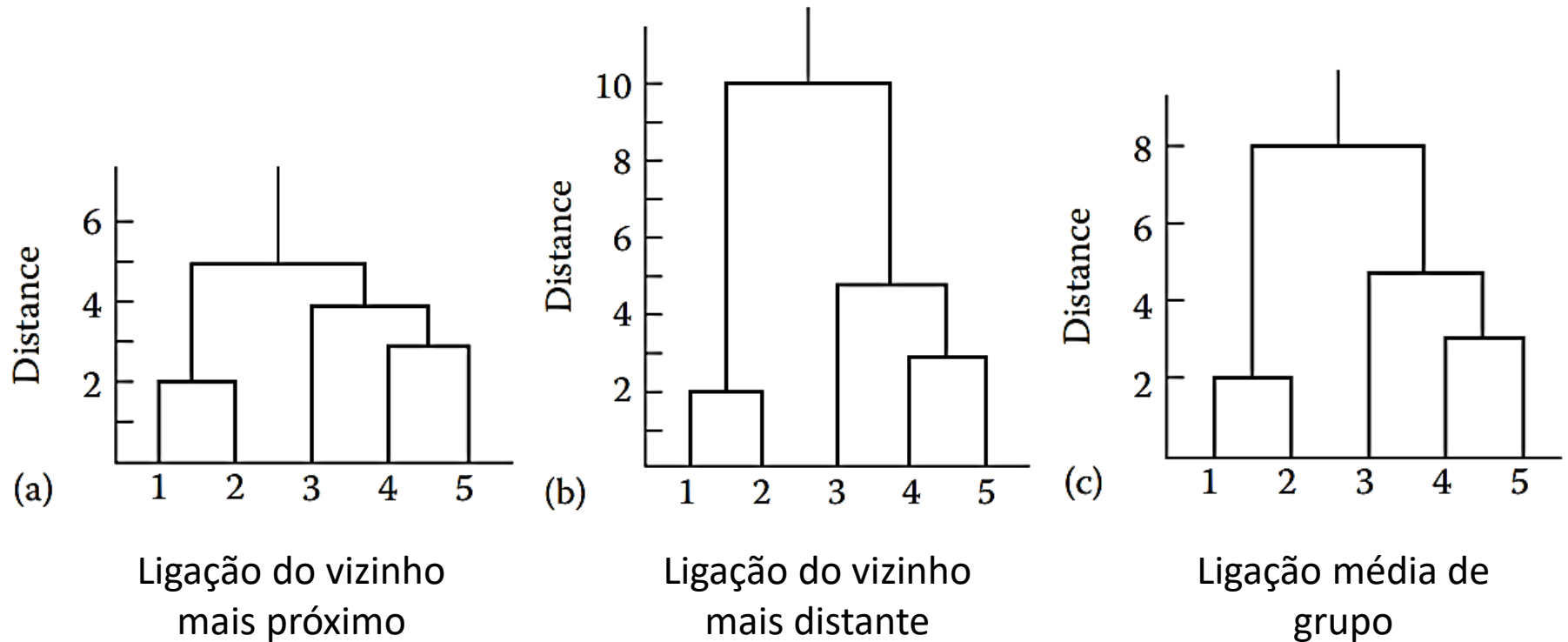
- A distância euclidiana não leva em consideração a associação entre as p variáveis.
- Uma medida de distância que leva em consideração as variâncias (escala) e covariâncias (associação) das variáveis é a **distância quadrada generalizada de Mahalanobis**, dada por:

$$D_{ii'}^2 = (\mathbf{x}_i - \mathbf{x}_{i'})^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_{i'})$$

em que \mathbf{S} é a matriz de covariância.

MÉTODOS HIERÁRQUICOS

Voltando para os métodos hierárquicos...



MÉTODOS HIERÁRQUICOS

- Inicia-se calculando uma matriz de distâncias.
- Todos os objetos começam sozinhos em grupos de tamanho um, e grupos que estão mais próximos se unem.

MÉTODOS HIERÁRQUICOS

- Para o exemplo dos cães pré-históricos, temos:

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

CM: cão moderno, CD: chacal dourado, LC: lobo chinês, LI: lobo indiano, CPH: cão pré-histórico.

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Identificar as duas espécies mais similares.

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Identificar as duas espécies mais similares.

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Primeiro agrupamento: cão moderno e cão pré-histórico.
- Eles se unem a uma distância de 0,66

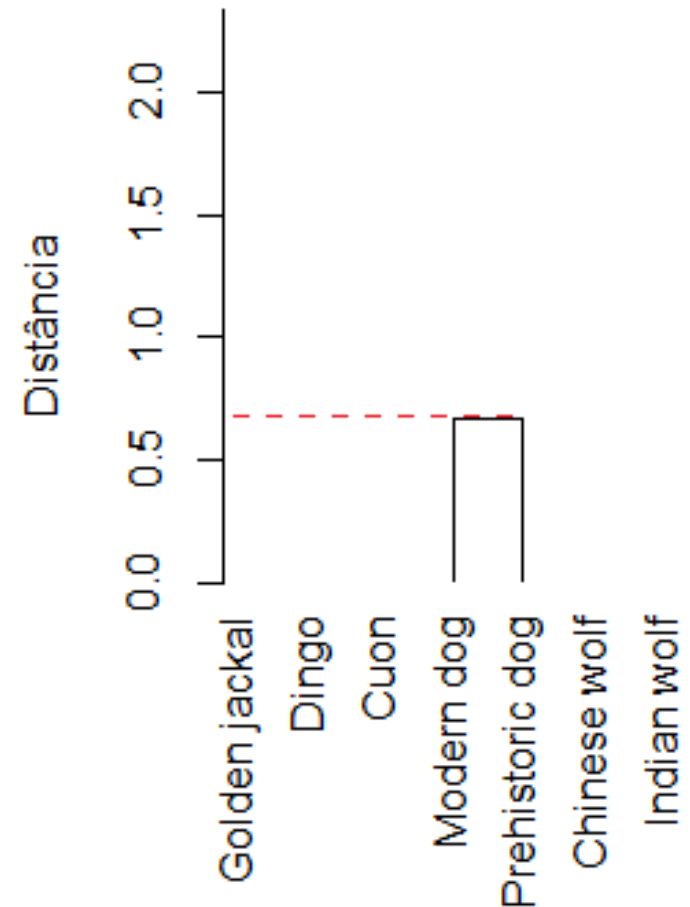


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Atualizar a matriz de distâncias pelo método do vizinho mais próximo.

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD		-	-	-	-	-	-
LC		7,12	-	-	-	-	-
LI		5,06	2,14	-	-	-	-
Cuon		3,19	4,57	2,91	-	-	-
Dingo		3,18	4,21	2,20	1,67	-	-
CPH							-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Atualizar a matriz de distâncias pelo método do vizinho mais próximo.

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	7,12	-	-	-	-
LI	3,24	5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	7,12	-	-	-	-
LI	3,24	5,06	2,14	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo (completa) pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	7,12	-	-	-	-
LI	3,24	5,06	2,14	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

Identificar a menor distância na nova matriz de distâncias:

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	7,12	-	-	-	-
LI	3,24	5,06	2,14	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

Identificar a menor distância na nova matriz de distâncias:

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	7,12	-	-	-	-
LI	3,24	5,06	2,14	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Segundo agrupamento: cão moderno, cão pré-histórico e cuon.
- Eles se unem a uma distância de 1,26.
- Atualizar novamente a matriz de distâncias.

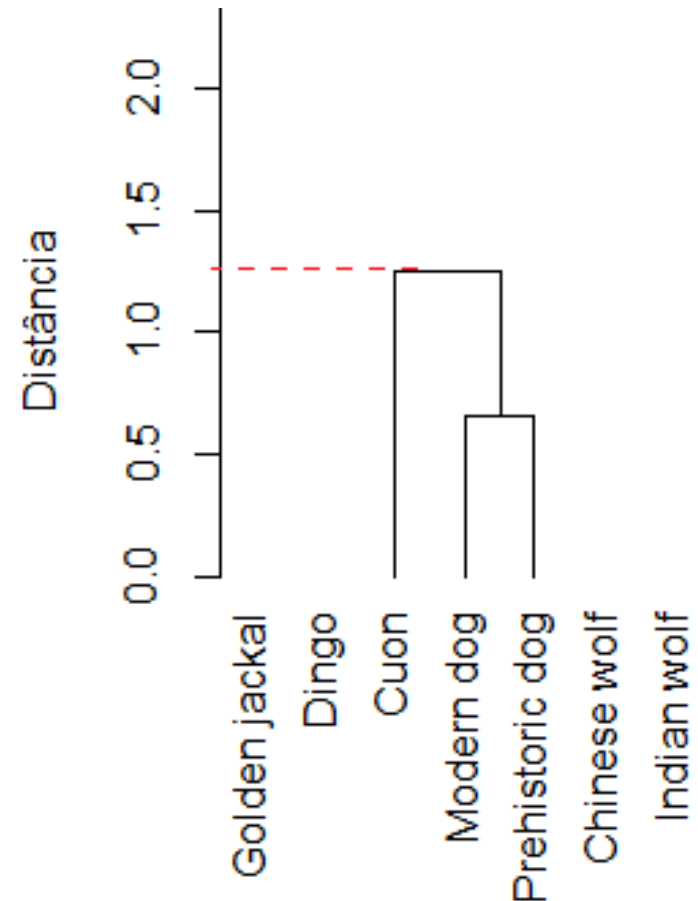


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	-	-	-	-	-
LI	3,24	-	-	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	1,67
Dingo	1,56	-	-	-	-	-

Tabela 3. Distâncias euclidianas entre sete grupos caninos para o terceiro passo (incompleta) pelo VMP.

	CM CPH Cuon	CD	LC	LI	Dingo
CM CPH Cuon	-	-	-	-	-
CD		-	-	-	-
LC		7,12	-	-	-
LI		5,06	2,14	-	-
Dingo		3,18	4,21	2,20	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	-	-	-	-	-
LI	3,24	-	-	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	1,67
Dingo	1,56	-	-	-	-	-

Tabela 3. Distâncias euclidianas entre sete grupos caninos para o terceiro passo (incompleta) pelo VMP.

	CM CPH Cuon	CD	LC	LI	Dingo
CM CPH Cuon	-	-	-	-	-
CD	1,91	-	-	-	-
LC		7,12	-	-	-
LI		5,06	2,14	-	-
Dingo		3,18	4,21	2,20	-

Tabela 2. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMP.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	1,91	-	-	-	-	-
LC	5,12	-	-	-	-	-
LI	3,24	-	-	-	-	-
Cuon	1,26	3,19	4,57	2,91	-	1,67
Dingo	1,56	-	-	-	-	-

Tabela 3. Distâncias euclidianas entre sete grupos caninos para o terceiro passo (incompleta) pelo VMP.

	CM CPH Cuon	CD	LC	LI	Dingo
CM CPH Cuon	-	-	-	-	-
CD	1,91	-	-	-	-
LC	4,57	7,12	-	-	-
LI	2,91	5,06	2,14	-	-
Dingo	1,56	3,18	4,21	2,20	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

Identificar a menor distância na nova matriz de distâncias:

Tabela 3. Distâncias euclidianas entre sete grupos caninos para o terceiro passo pelo VMP.

	CM CPH Cuon	CD	LC	LI	Dingo
CM CPH Cuon	-	-	-	-	-
CD	1,91	-	-	-	-
LC	4,57	7,12	-	-	-
LI	2,91	5,06	2,14	-	-
Dingo	1,56	3,18	4,21	2,20	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Terceiro agrupamento: cão moderno, cão pré-histórico, cuon e dingo.
- Eles se unem a uma distância de 1,56.
- Atualizar a matriz de distâncias.

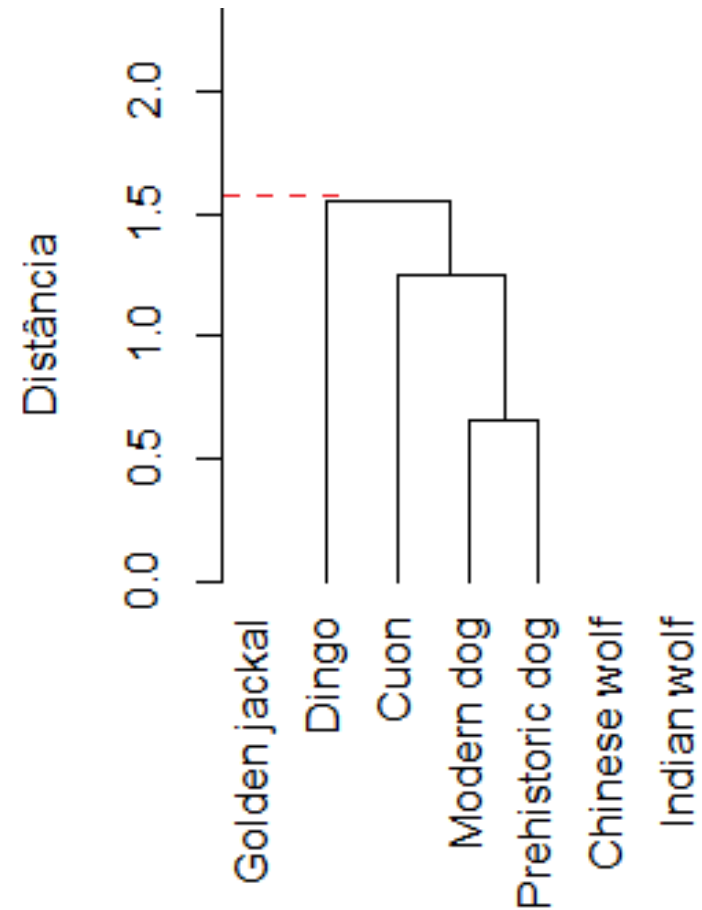


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

MÉTODOS DO VIZINHO MAIS PRÓXIMO

Identificar a menor distância na nova matriz de distâncias:

Tabela 4. Distâncias euclidianas entre sete grupos caninos para o quarto passo pelo VMP.

	CM CPH Cuon Dingo	CD	LC	LI
CM CPH Cuon Dingo	-	-	-	-
CD	1,91	-	-	-
LC	4,21	7,12	-	-
LI	2,20	5,06	2,14	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Quarto agrupamento: cão moderno, cão pré-histórico, cuon, dingo e chacal dourado.
- Eles se unem a uma distância de 1,91.

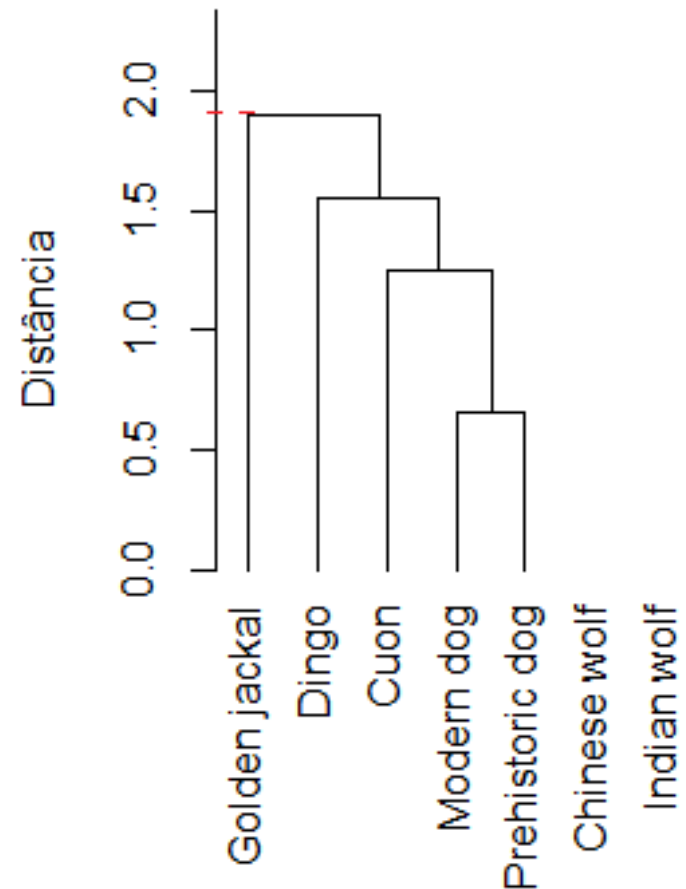


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

MÉTODOS DO VIZINHO MAIS PRÓXIMO

Identificar a menor distância na nova matriz de distâncias:

Tabela 5. Distâncias euclidianas entre sete grupos caninos para o quinto passo pelo VMP.

	CM CPH Cuon Dingo CD	LC	LI
CM CPH Cuon Dingo CD	-	-	-
LC	4,21	-	-
LI	2,20	2,14	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Quinto agrupamento: lobo chinês e lobo indiano.
- Eles se unem a uma distância de 2,14.

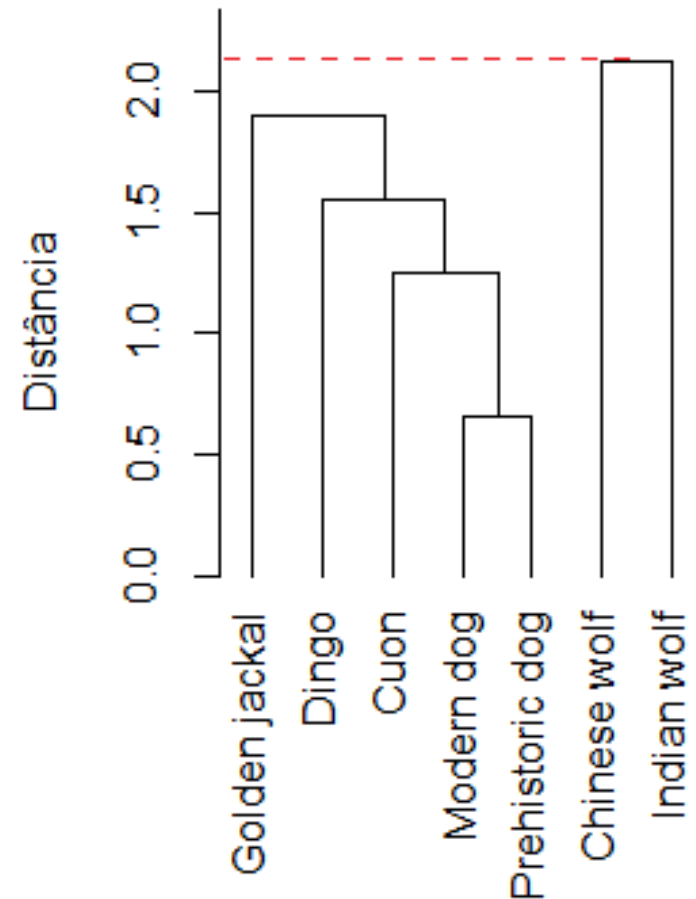


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

MÉTODOS DO VIZINHO MAIS PRÓXIMO

Identificar a menor distância na nova matriz de distâncias:

Tabela 6. Distâncias euclidianas entre sete grupos caninos para o sexto passo pelo VMP.

	CM CPH Cuon Dingo CD	LC LI
CM CPH Cuon Dingo CD	-	-
LC LI	2,20	-

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Sexto agrupamento: todos os indivíduos.
- Eles se unem a uma distância de 2,20.

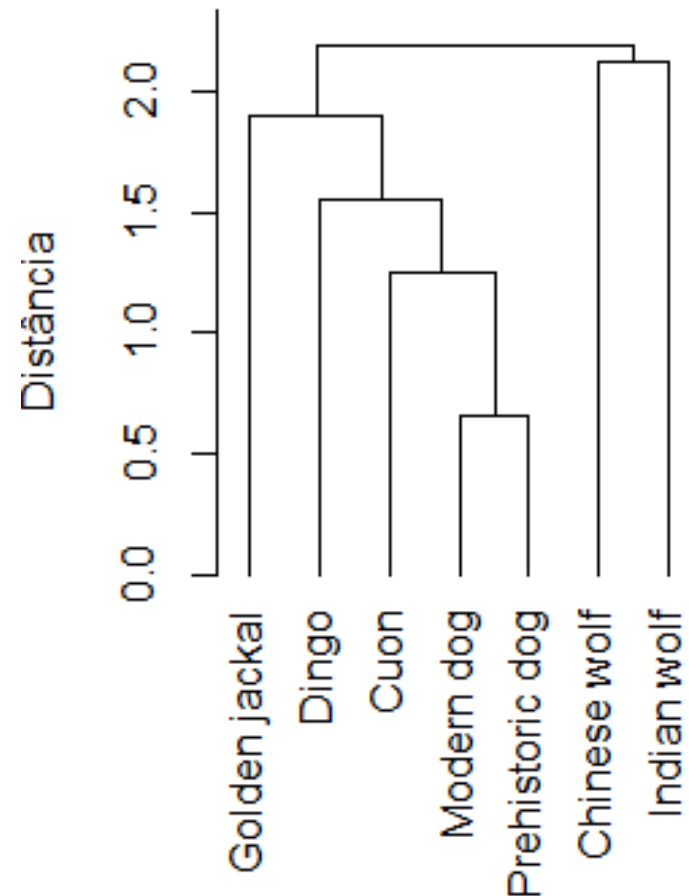


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

MÉTODOS DO VIZINHO MAIS PRÓXIMO

- Sexto agrupamento: todos os indivíduos.
- Eles se unem a uma distância de 2,20.
- E agora, quem são meus grupos?

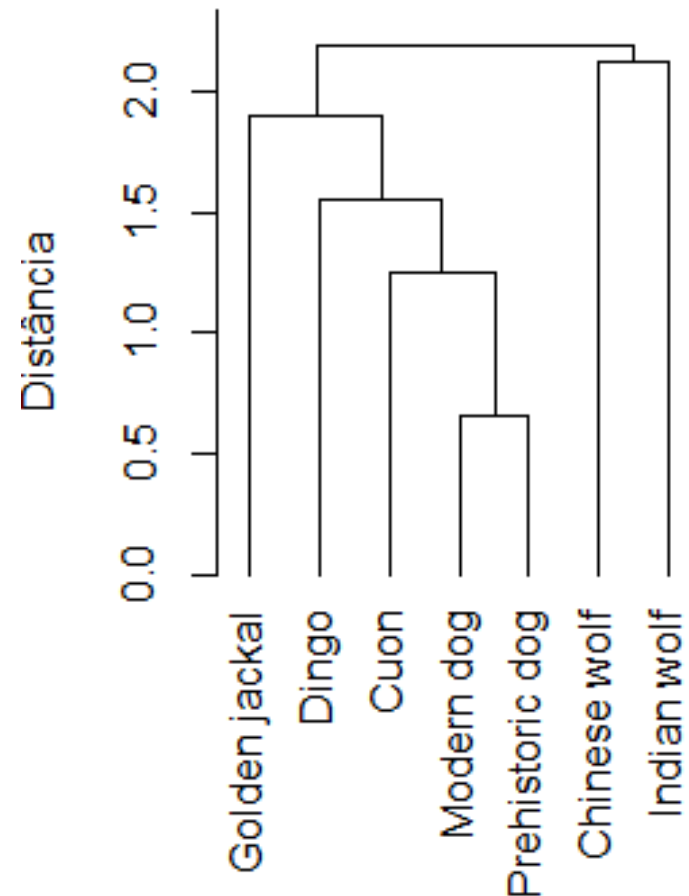


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

SELEÇÃO DE GRUPOS

- Se eu escolher arbitrariamente a distância 1,75, quais seriam os grupos?

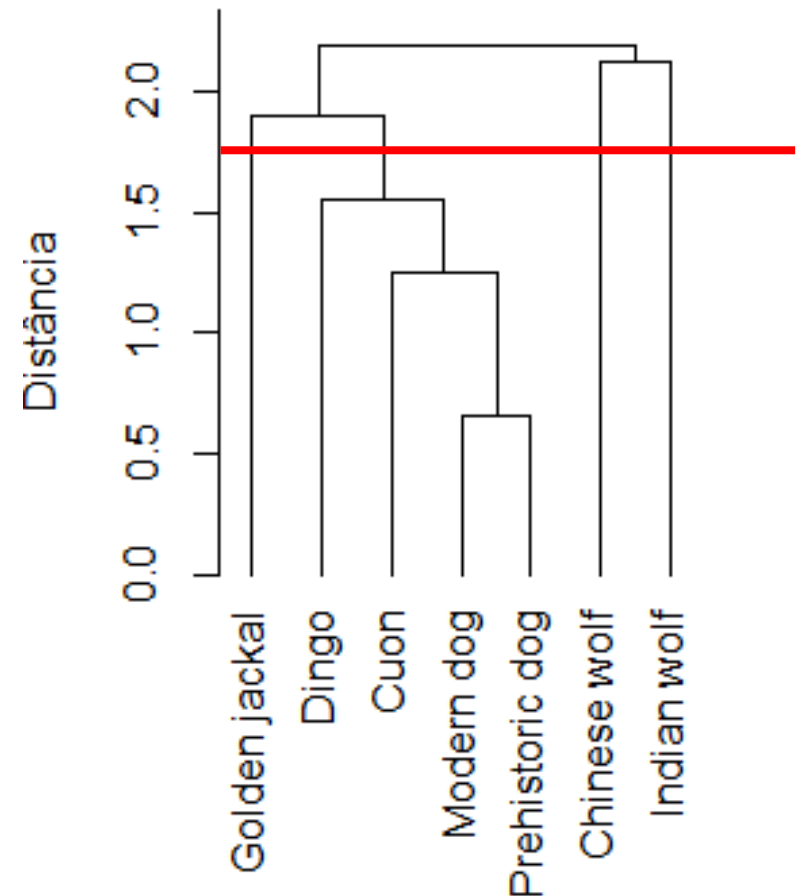


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

SELEÇÃO DE GRUPOS

- Se eu escolher arbitrariamente a distância 1,75:
- Grupo 1: CD
- Grupo 2: Dingo, Cuon, MD, CPH
- Grupo 3: LC
- Grupo 4: LI

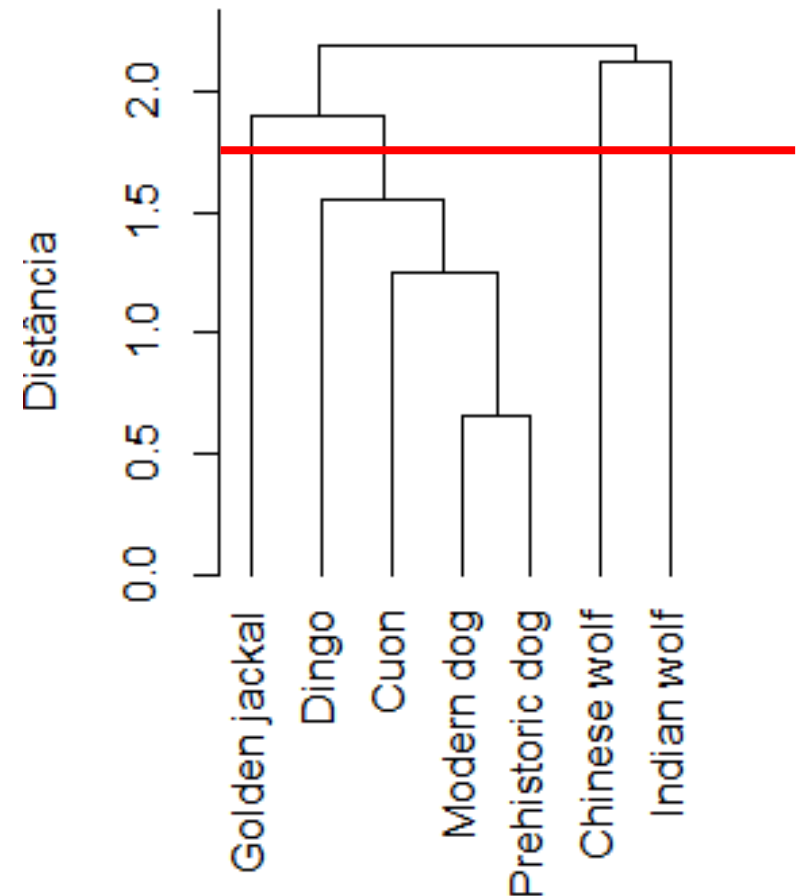


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

SELEÇÃO DE GRUPOS

- Se eu escolher arbitrariamente a distância 1:

- Grupo 1: CD
- Grupo 2: Dingo
- Grupo 3: Cuon
- Grupo 4: CM e CPH
- Grupo 5: LC
- Grupo 6: LI

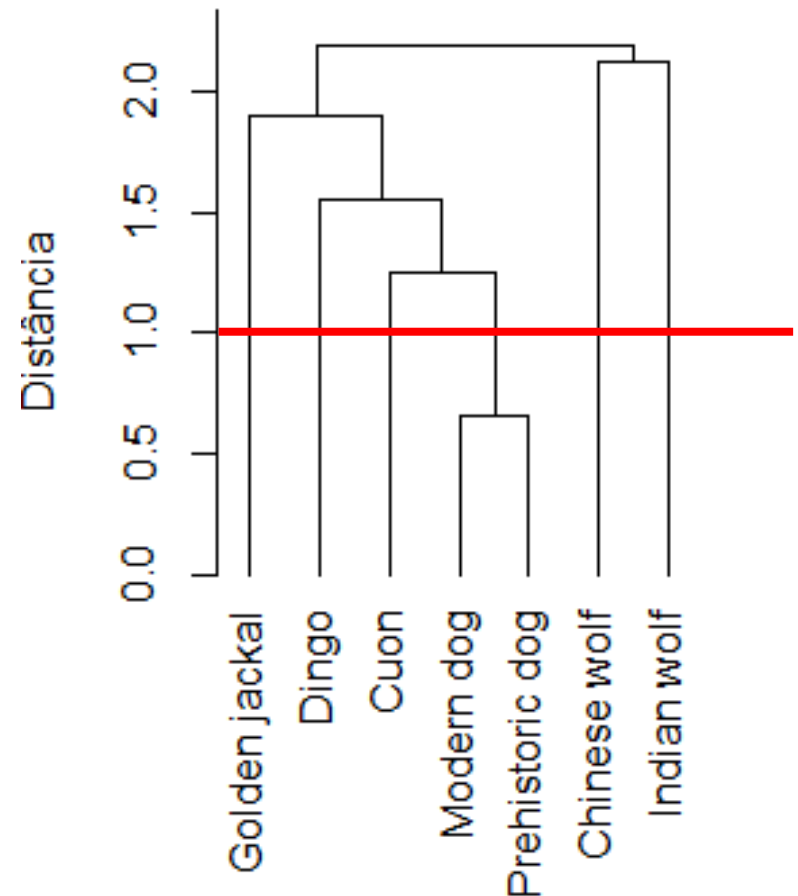


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

SELEÇÃO DE GRUPOS

- Se eu escolher arbitrariamente a distância 2,17:
- Grupo 1: CD, Dingo, Cuon, CM e CPH
- Grupo 2: LC e LI

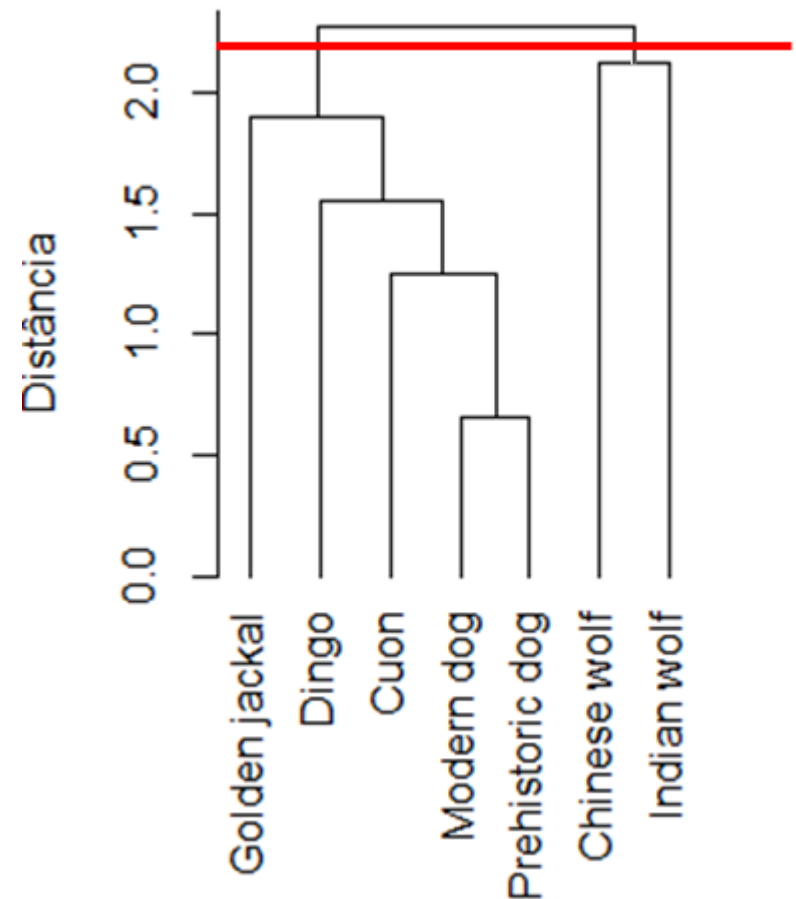
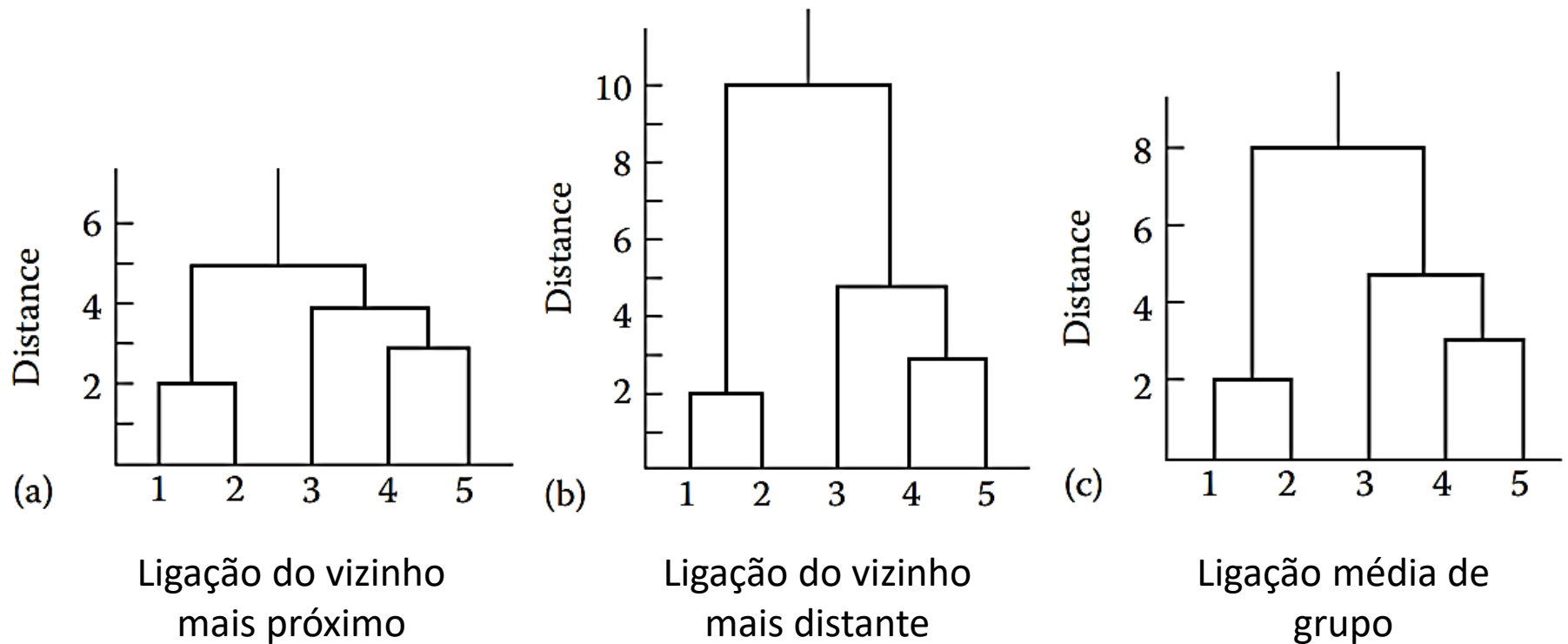


Figura 1. Dendrograma para os sete grupos caninos pelo VMP.

MÉTODOS HIERÁRQUICOS

Voltando para os métodos hierárquicos...



MÉTODOS DO VIZINHO MAIS DISTANTE

- Inicia-se calculando uma matriz de distâncias.

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

CM: cão moderno, CD: chacal dourado, LC: lobo chinês, LI: lobo indiano, CPH: cão pré-histórico.

MÉTODOS DO VIZINHO MAIS DISTANTE

- Identificar as duas espécies mais similares.

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

MÉTODOS DO VIZINHO MAIS DISTANTE

- Primeiro agrupamento formado por CM e CPH.

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC	5,38	7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC	5,38	7,12	-	-	-	-
LI	3,38	5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC	5,38	7,12	-	-	-	-
LI	3,38	5,06	2,14	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	-	-	-	-	-	-
LI	3,38	-	-	-	-	-	-
Cuon	1,51	-	-	-	-	-	-
Dingo	1,56	-	-	-	-	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo (completa) pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC	5,38	7,12	-	-	-	-
LI	3,38	5,06	2,14	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-
Dingo	1,71	3,18	4,21	2,20	1,67	-

MÉTODOS DO VIZINHO MAIS DISTANTE

Identificar a menor distância na nova matriz de distâncias:

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC	5,38	7,12	-	-	-	-
LI	3,38	5,06	2,14	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-
Dingo	1,71	3,18	4,21	2,20	1,67	-

MÉTODOS DO VIZINHO MAIS DISTANTE

Identificar a menor distância na nova matriz de distâncias:

Tabela 7. Distâncias euclidianas entre sete grupos caninos para o segundo passo pelo VMD.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,39	-	-	-	-	-
LC	5,38	7,12	-	-	-	-
LI	3,38	5,06	2,14	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-
Dingo	1,71	3,18	4,21	2,20	1,67	-

MÉTODOS DO VIZINHO MAIS DISTANTE

Identificar a menor distância na nova matriz de distâncias:

Tabela 8. Distâncias euclidianas entre sete grupos caninos para o terceiro passo pelo VMD.

	CM CPH Cuon	CD	LC	LI	Dingo
CM CPH Cuon	-	-	-	-	-
CD	3,19	-	-	-	-
LC	5,38	7,12	-	-	-
LI	3,38	5,06	2,14	-	-
Dingo	1,71	3,18	4,21	2,20	-

MÉTODOS DO VIZINHO MAIS DISTANTE

Identificar a menor distância na nova matriz de distâncias:

Tabela 9. Distâncias euclidianas entre sete grupos caninos para o quarto passo pelo VMD.

	CM CPH Cuon Dingo	CD	LC	LI
CM CPH Cuon Dingo	-	-	-	-
CD	3,19	-	-	-
LC	5,38	7,12	-	-
LI	3,38	5,06	2,14	-

MÉTODOS DO VIZINHO MAIS DISTANTE

Identificar a menor distância na nova matriz de distâncias:

Tabela 10. Distâncias euclidianas entre sete grupos caninos para o quinto passo pelo VMD.

	CM CPH Cuon Dingo	CD	LI LC
CM CPH Cuon Dingo	-	-	-
CD	3,19	-	-
LI LC	5,38	7,12	-

MÉTODOS DO VIZINHO MAIS DISTANTE

Identificar a menor distância na nova matriz de distâncias:

Tabela 11. Distâncias euclidianas entre sete grupos caninos para o sexto passo pelo VMD.

	CM CPH Cuon Dingo CD	LI LC
CM CPH Cuon Dingo CD	-	-
LI LC	7,12	-

MÉTODOS DO VIZINHO MAIS DISTANTE

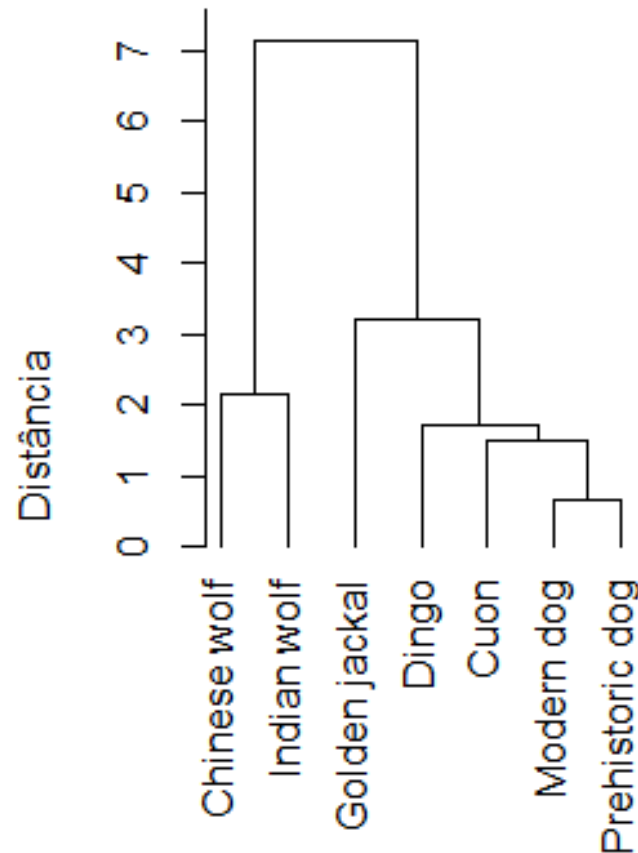


Figura 2. Dendrograma para os sete grupos caninos pelo método do vizinho mais distante.

MÉTODOS DO VIZINHO MAIS DISTANTE

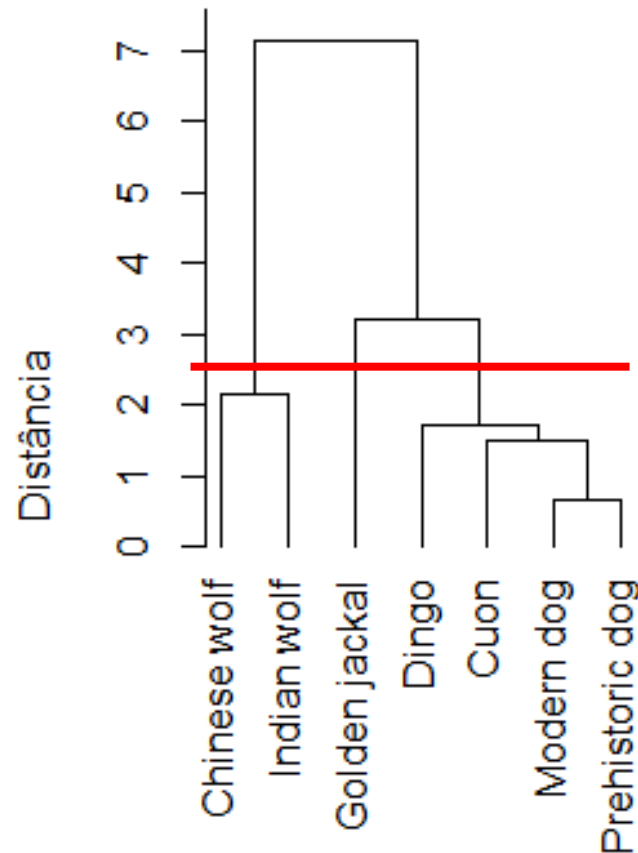


Figura 2. Dendrograma para os sete grupos caninos pelo método do vizinho mais distante.

MÉTODOS DA LIGAÇÃO MÉDIA DE GRUPO

- Identificar as duas espécies mais similares.

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 11. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo método da LMG.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD		-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

Tabela 1. Distâncias euclidianas entre sete grupos caninos.

	CM	CD	LC	LI	Cuon	Dingo	CPH
CM	-	-	-	-	-	-	-
CD	1,91	-	-	-	-	-	-
LC	5,38	7,12	-	-	-	-	-
LI	3,38	5,06	2,14	-	-	-	-
Cuon	1,51	3,19	4,57	2,91	-	-	-
Dingo	1,56	3,18	4,21	2,20	1,67	-	-
CPH	0,66	2,39	5,12	3,24	1,26	1,71	-

Tabela 11. Distâncias euclidianas entre sete grupos caninos para o segundo passo (incompleta) pelo método da LMG.

	CM CPH	CD	LC	LI	Cuon	Dingo
CM CPH	-	-	-	-	-	-
CD	2,15	-	-	-	-	-
LC		7,12	-	-	-	-
LI		5,06	2,14	-	-	-
Cuon		3,19	4,57	2,91	-	-
Dingo		3,18	4,21	2,20	1,67	-

MÉTODOS DA LIGAÇÃO MÉDIA DE GRUPO

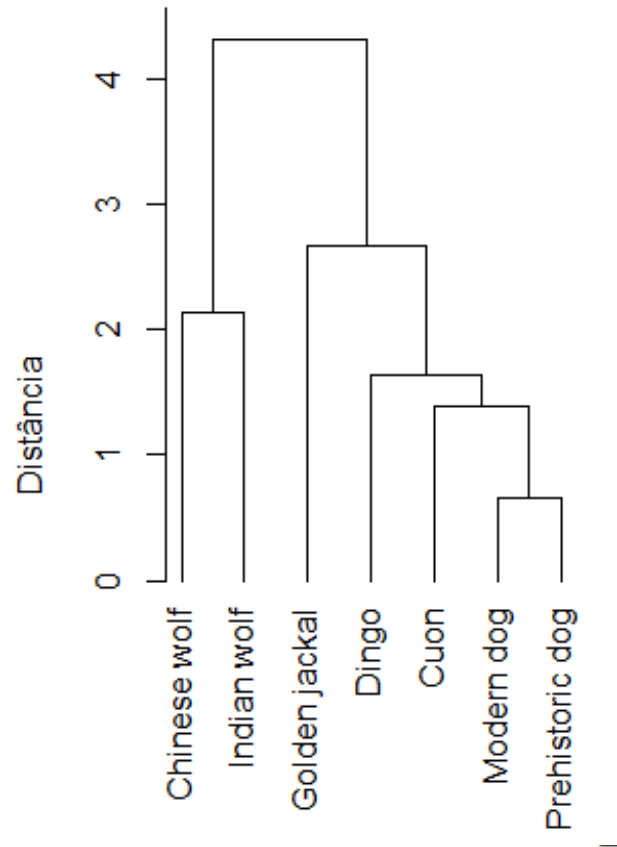
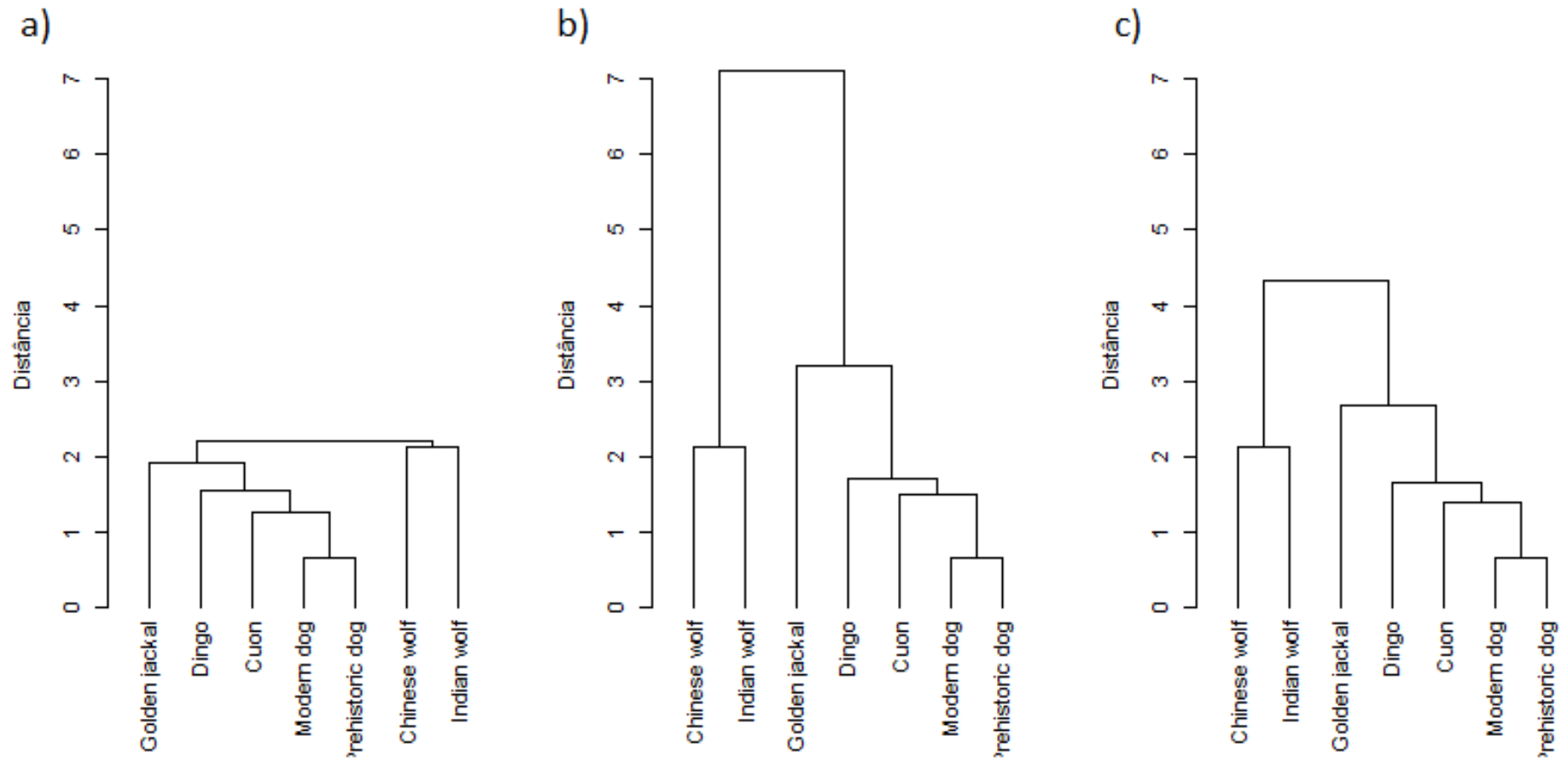


Figura 3. Dendrograma para os sete grupos caninos pelo método da ligação média de grupo.

MÉTODOS HIERÁRQUICOS



- **Figura 4.** Dendrograma para os sete grupos caninos pelos métodos do vizinho mais próximo (a), vizinho mais distante (b) e ligação média de grupo (c).

MÉTODOS HIERÁRQUICOS

Resumo:

Vizinho mais próximo:

Dois grupos são unidos se os membros dos dois grupos estão perto o suficiente.

Vizinho mais distante:

Dois grupos são unidos somente se os membros mais distantes dos dois grupos estão próximos o suficiente.

Ligação media:

Dois grupos são unidos somente se a distância média entre eles é pequena.

EXEMPLO

Emprego em países Europeus

Tabela 12: Porcentagem de força de trabalho de empregados em nove diferentes grupos de indústria em 30 países da Europa.

País	Grupo	AGR	MIN	FAB	FEA	CON	SER	FIN	SSP	TC
Bélgica	EU	2,6	0,2	20,8	0,8	6,3	16,9	8,7	36,9	6,8
Dinamarca	EU	5,6	0,1	20,4	0,7	6,4	14,5	9,1	36,3	7,0
França	EU	5,1	0,3	20,2	0,9	7,1	16,7	10,2	33,1	6,4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Albânia	Leste	55,5	19,4	0,0	0,0	3,4	3,3	15,3	0,0	3,0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Turquia	Outros	44,8	0,9	15,3	0,2	5,2	12,4	2,4	14,5	4,4

EXEMPLO

Emprego em países Europeus - legenda

AGR = agricultura, florestal e pesca;

MIN = mineração e exploração de pedreiras;

FAB = fabricação;

FEA = fornecimento de energia e água;

CON = construção;

SER = serviços;

FIN = finanças;

SSP = serviços sociais e pessoais;

TC = transportes e comunicações.

EXEMPLO

OBJETIVOS:

- Isolar grupos de países com padrões similares de empregos;
- Entender os relacionamentos entre os países;
- Diferenças entre países que são relacionados a grupos políticos (EU: União Europeia, AELC: área europeia de livre comércio, LESTE: países do leste europeu) podem ser de particular interesse.

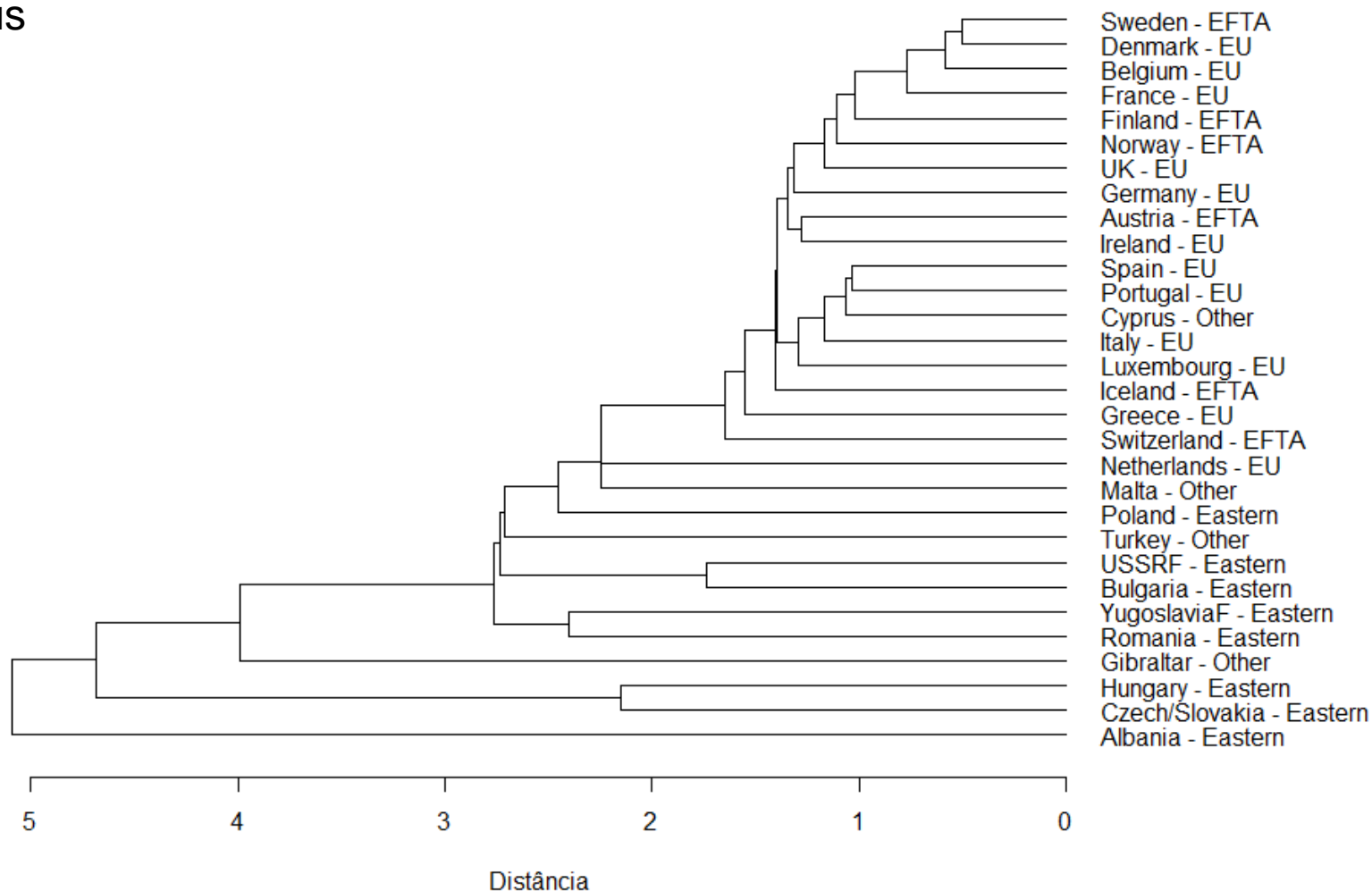
EXEMPLO

Exemplo Agrupamentos de países europeus

- Padronização das nove variáveis (média 0, desvio-padrão 1);
- Cálculo das distâncias euclidianas entre todos os pares de países;
- Dendrograma.

EXEMPLO

Figura 5. Dendrograma obtido de uma análise de agrupamentos hierárquica pelo método do vizinho mais próximo nos dados de emprego de países europeus



EXEMPLO

Interpretações:

- Existem somente quatro agrupamentos definidos por uma distância tipo vizinho mais próximo em torno de 3.
- Quatro agrupamentos definidos:
 - (1) Albânia
 - (2) Hungria e repúblicas Tcheca/Eslováquia
 - (3) Gibraltar
 - (4) Todos os outros países
- Isto separa três países do leste e Gibraltar de todo o resto, sugerindo que a classificação em UE, AELC, leste e outros países não é um bom indicador de padrões de empregos.

TIPOS DE ANÁLISE DE AGRUPAMENTOS

Não Hierárquicas:

Partição de objetos que podem se mover para dentro e para fora de grupos em diferentes momentos da análise.

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Inicia com centros arbitrários de agrupamentos, e variáveis padronizadas;
- Aloca itens ao centro mais próximo;
- Recalcula as os valores médios das variáveis para cada grupo;
- Novamente aloca indivíduos aos seus centros de grupos mais próximos para minimizar a soma total dos quadrados dentro do agrupamento, e assim por diante.

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

▪ Exemplo:

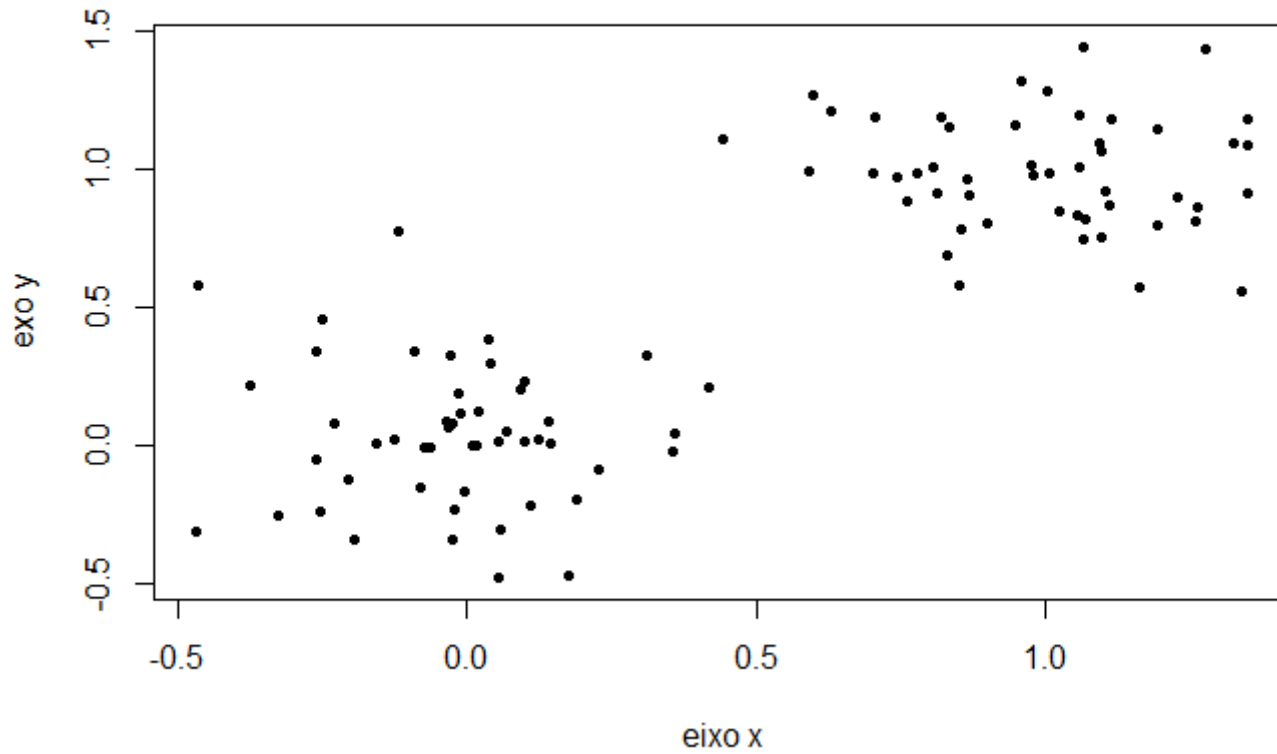


Figura 6. Gráfico de dispersão bivariado para exemplo do k-means (1).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Definir a quantidade de grupos.

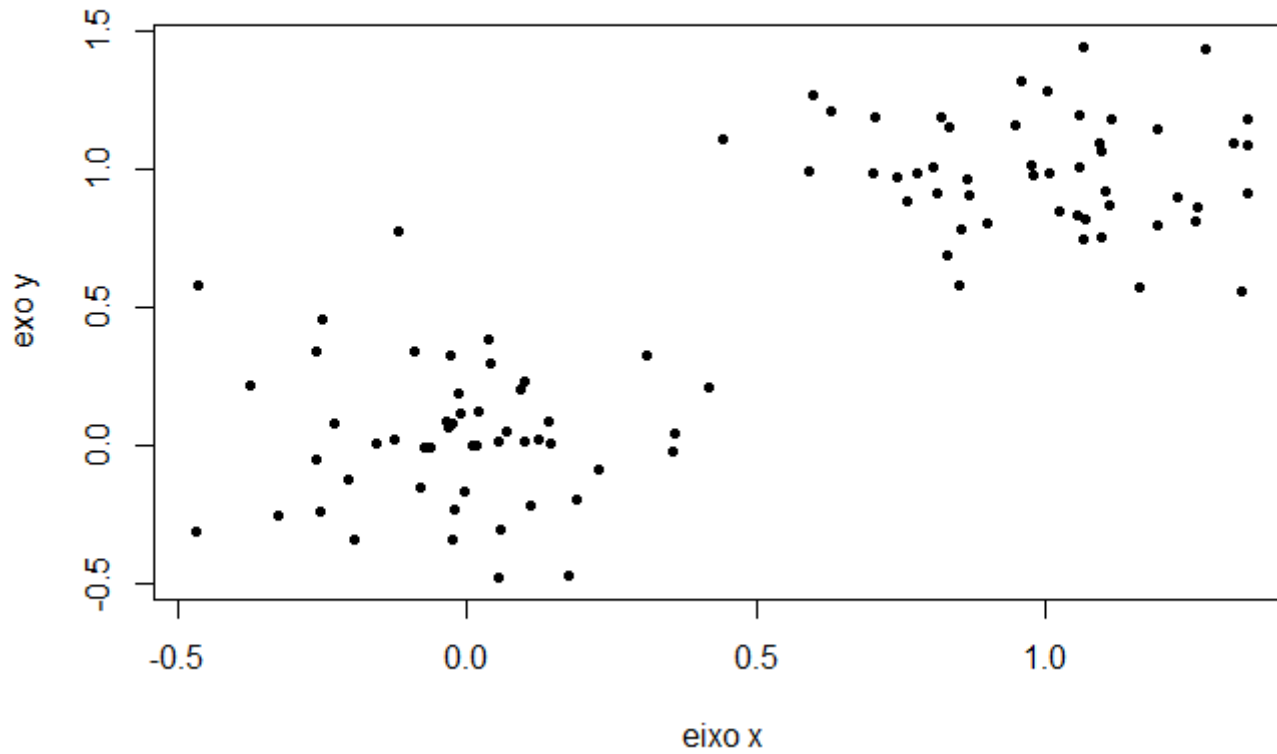


Figura 6. Gráfico de dispersão bivariado para exemplo do k-means (1).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Definir a quantidade de grupos.

K=2

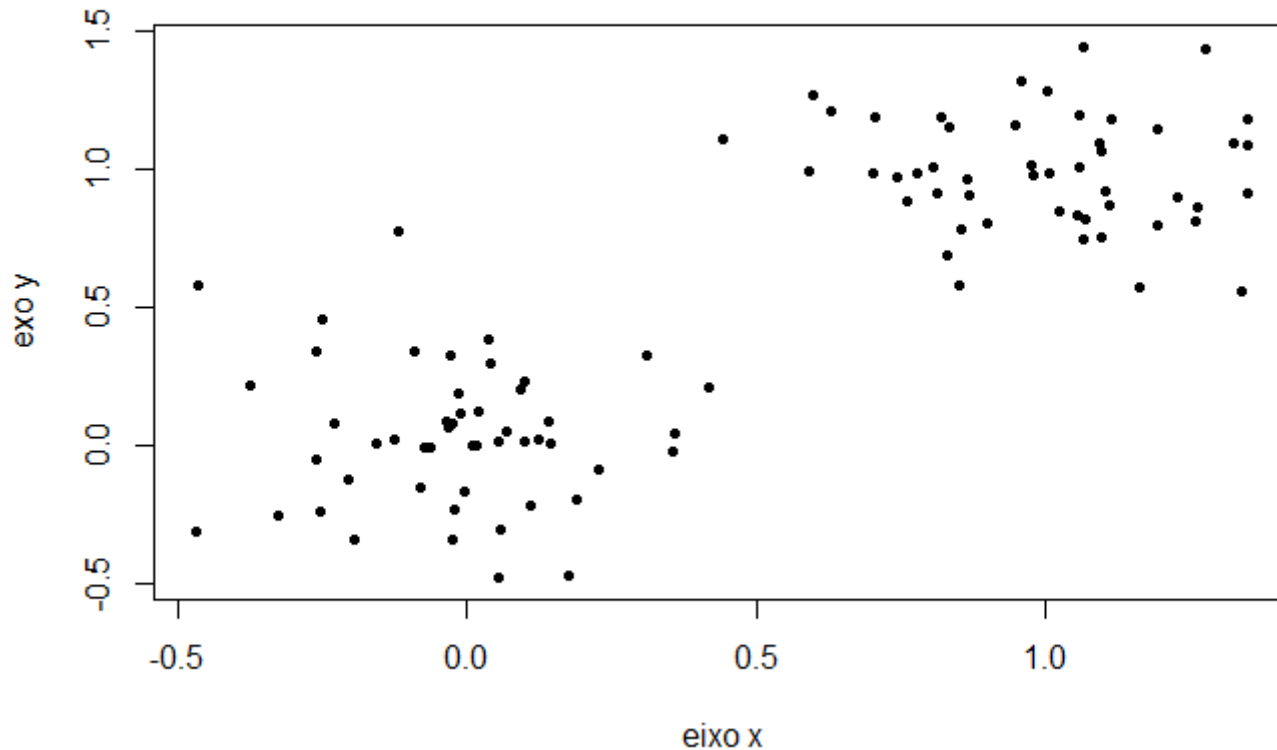


Figura 6. Gráfico de dispersão bivariado para exemplo do k-means (1).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Definir dois centroides aleatoriamente.

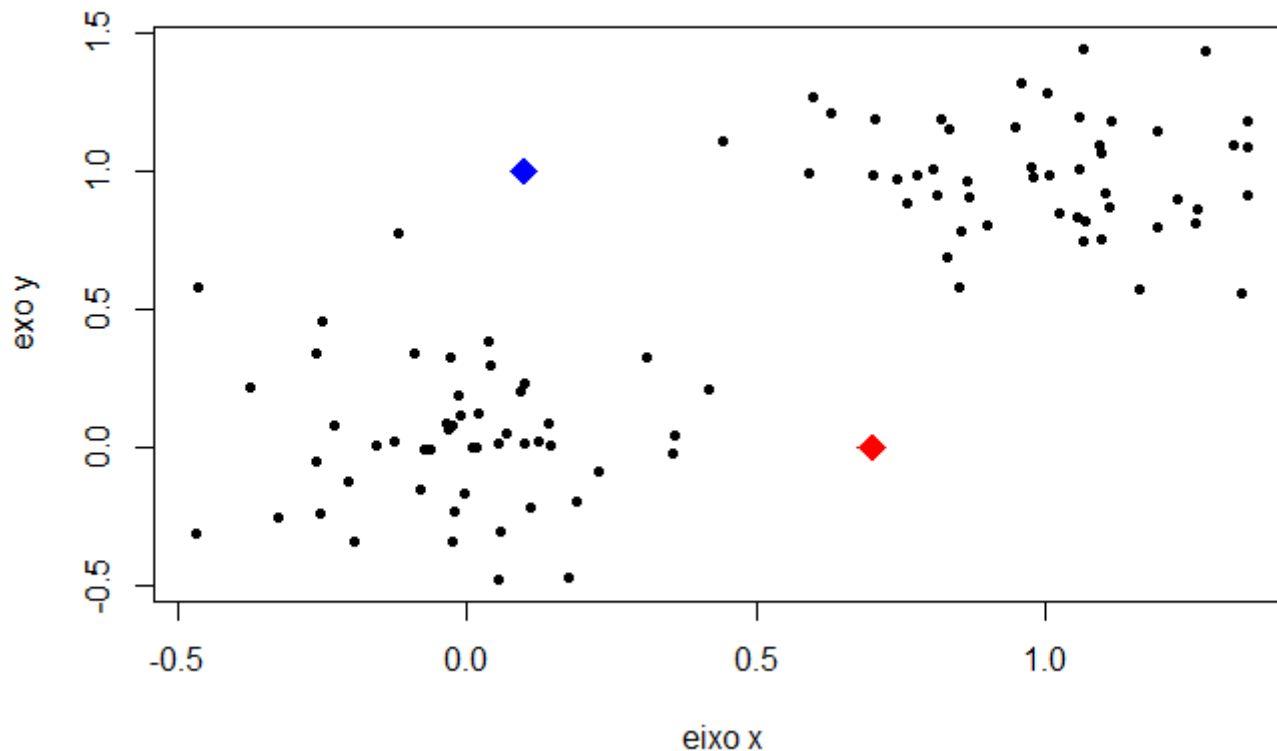


Figura 7. Gráfico de dispersão bivariado para exemplo do k-means (2).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- O algoritmo calcula a distância de cada objeto até cada centroide.

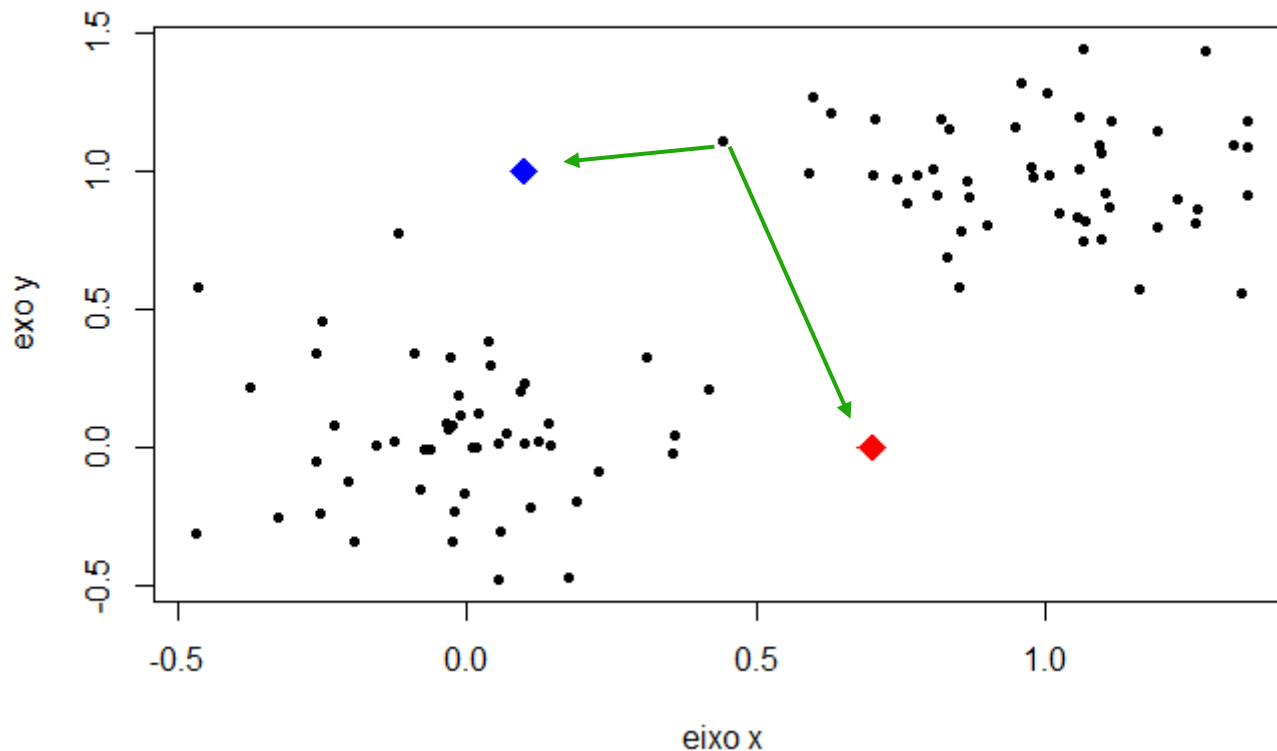


Figura 8. Gráfico de dispersão bivariado para exemplo do k-means (2).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Definir o grupo de acordo com a menor distância.

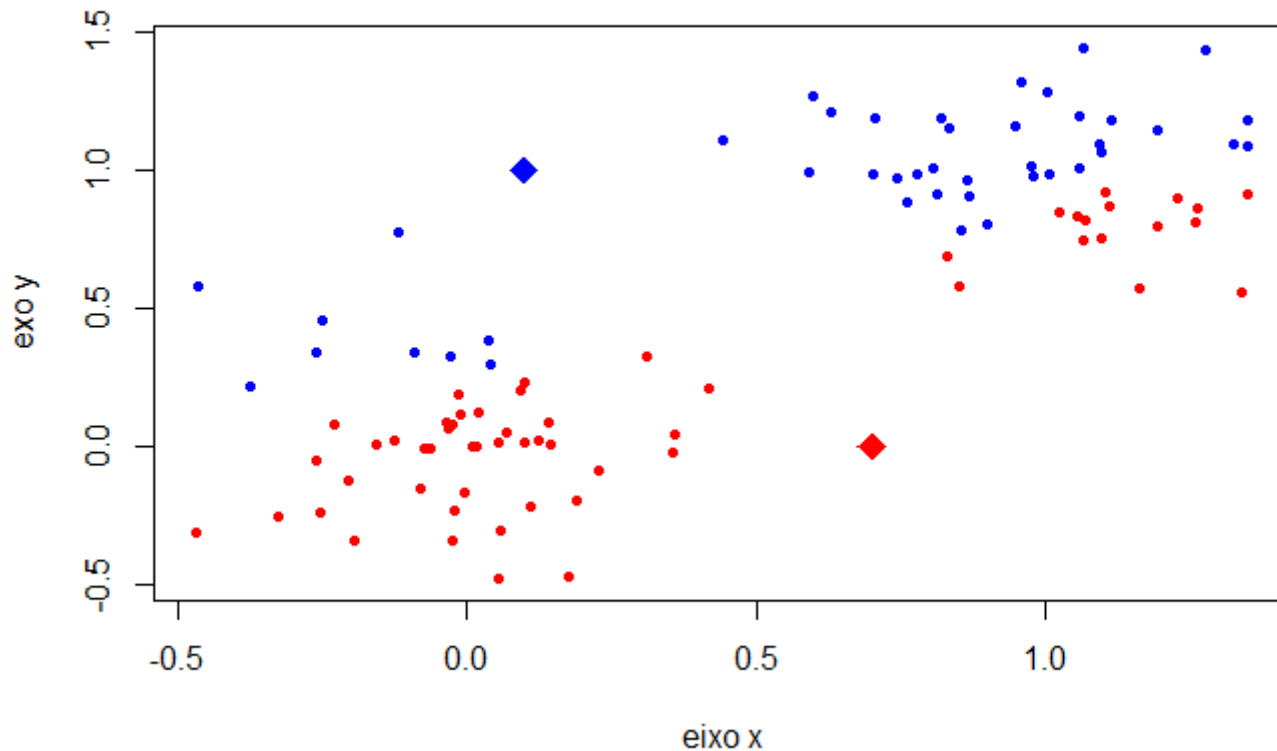


Figura 9. Gráfico de dispersão bivariado para exemplo do k-means (3).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Calcular os novos centroides dos grupos.

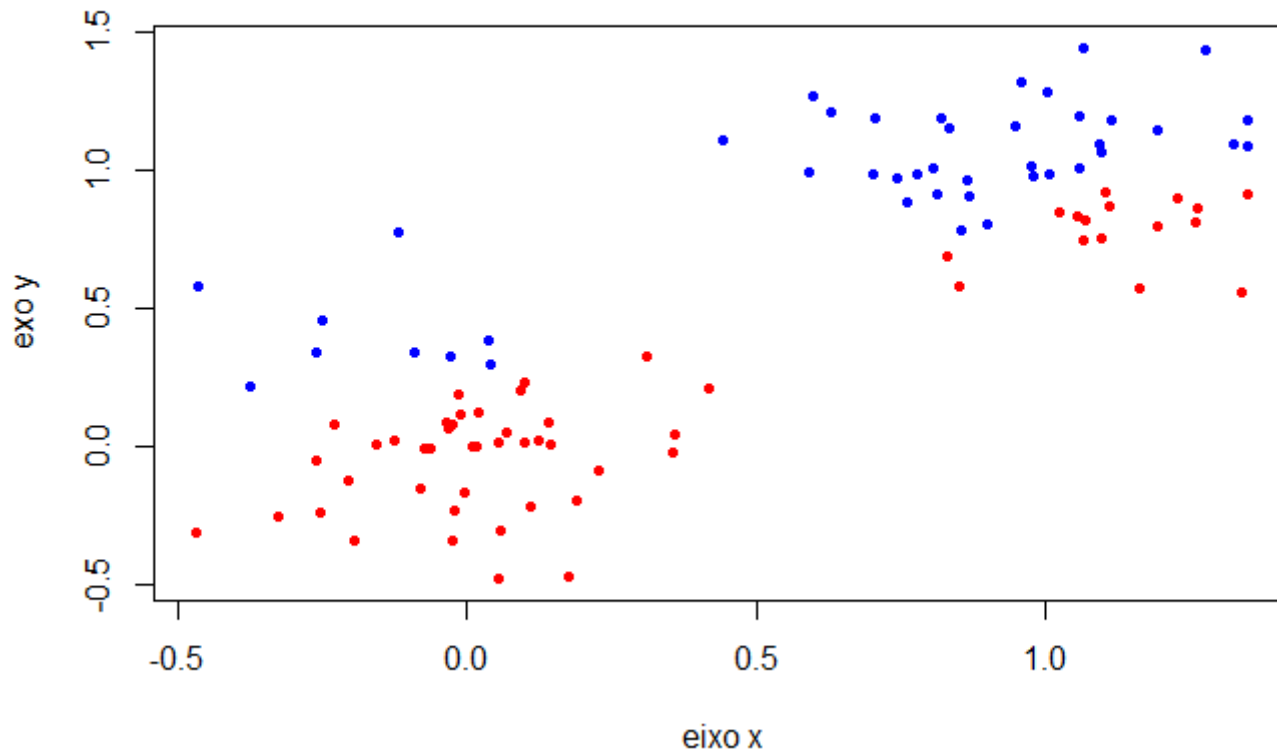


Figura 10. Gráfico de dispersão bivariado para exemplo do k-means (4).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- O algoritmo recalcula a distancia de cada objeto até cada (novo) centroide.

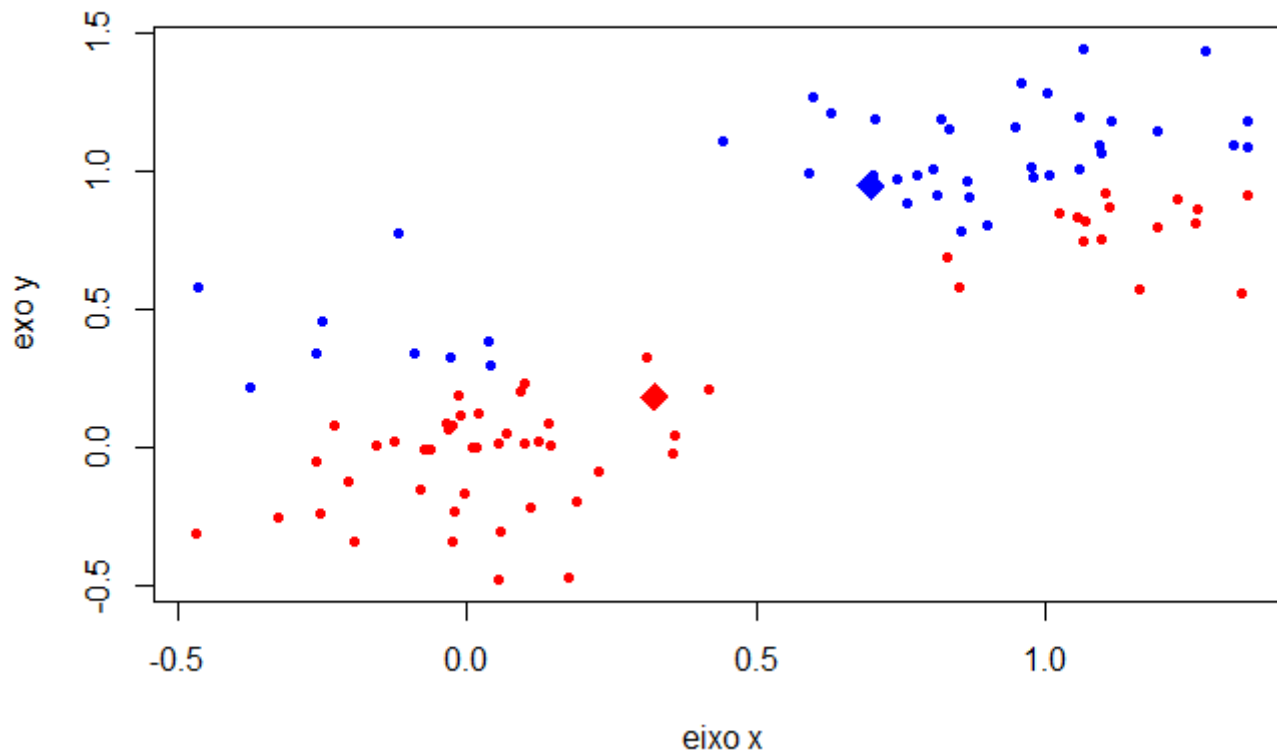


Figura 11. Gráfico de dispersão bivariado para exemplo do k-means (5).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Alguns indivíduos podem mudar de grupo!

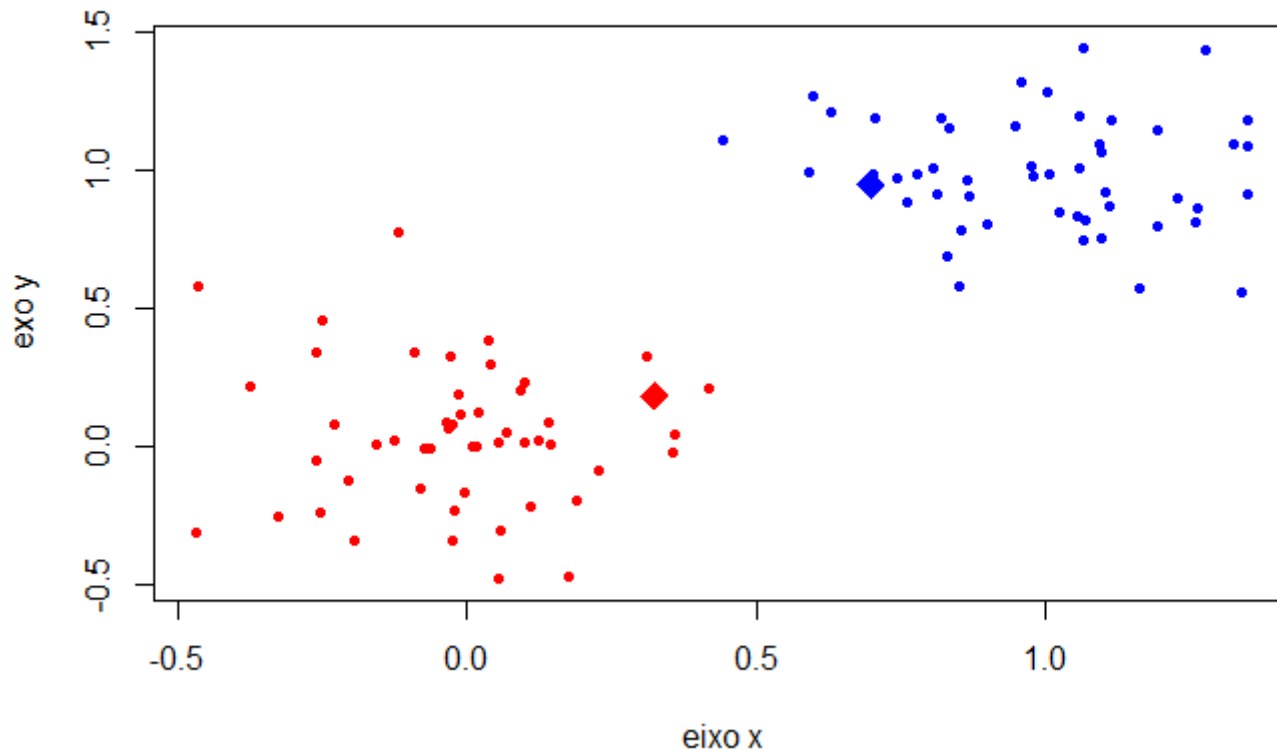


Figura 12. Gráfico de dispersão bivariado para exemplo do k-means (6).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Recalcular o centroide e identificar se existe mudanças no padrão de agrupamento.

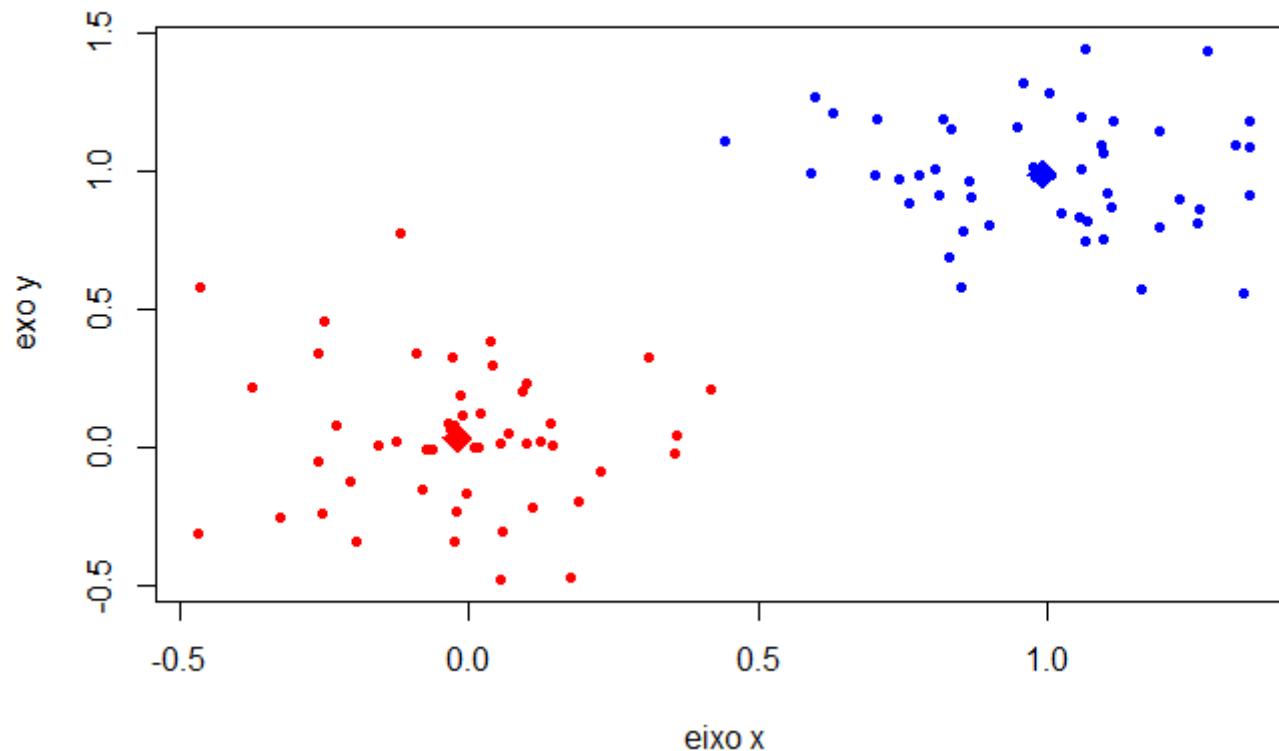


Figura 13. Gráfico de dispersão bivariado para exemplo do k-means (7).

MÉTODO NÃO-HIERÁRQUICO DE K-MÉDIAS

- Resultado final do agrupamento k-means.

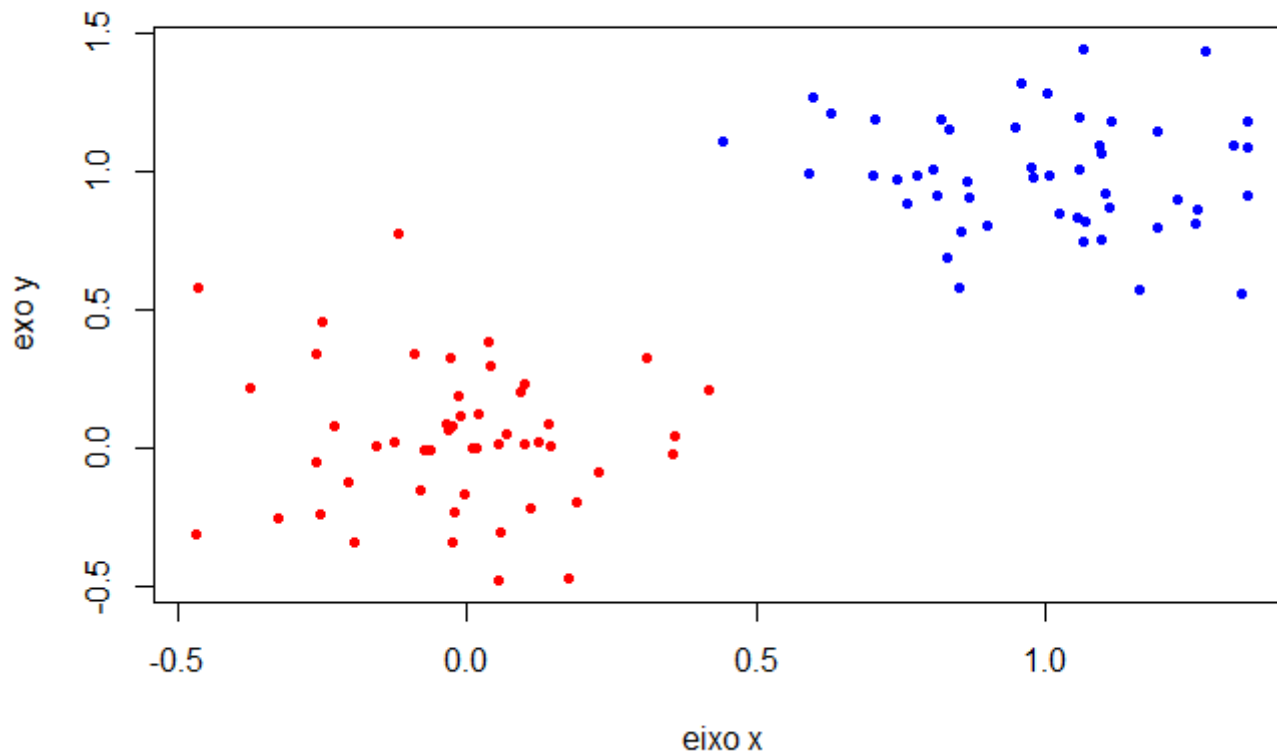


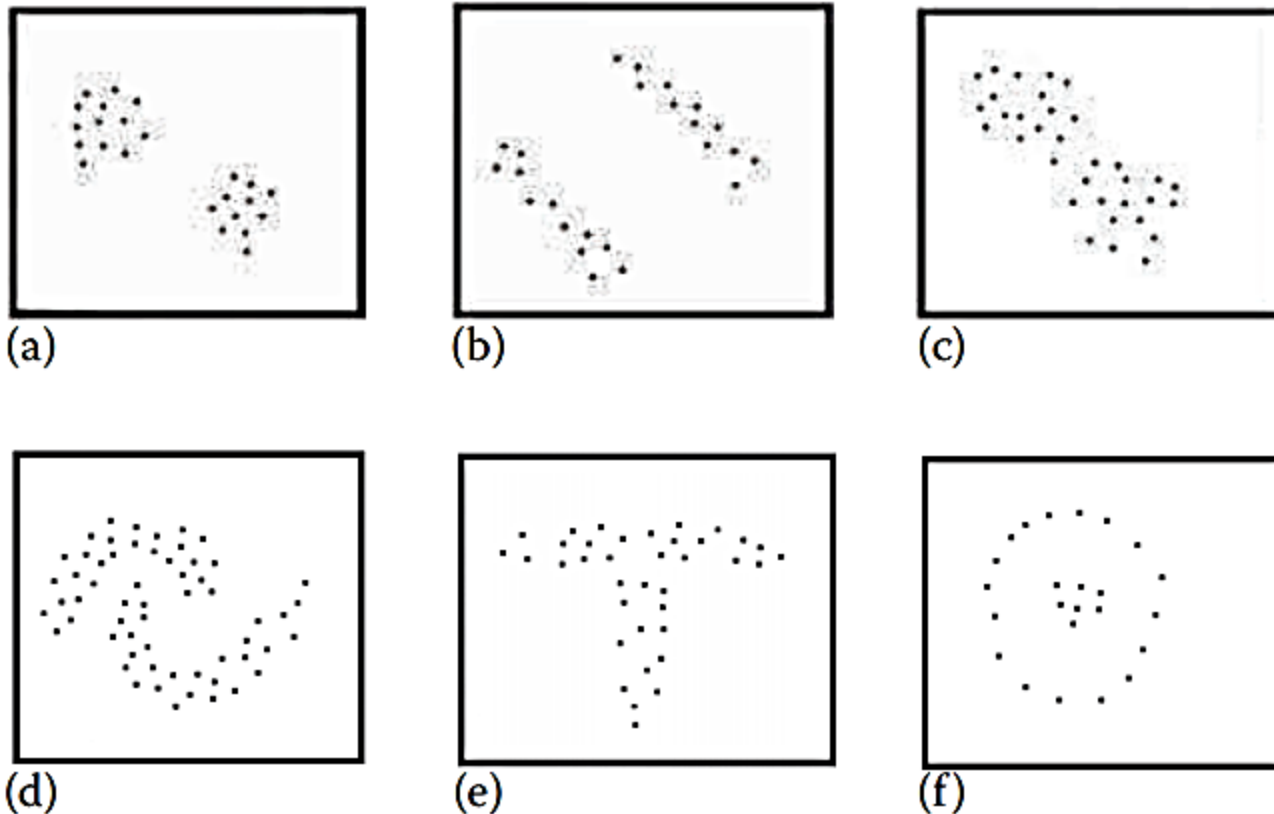
Figura 14. Gráfico de dispersão bivariado para exemplo do k-means (8).

O PROBLEMA DE ANÁLISE

- Não há um algoritmo melhor do que outro!
- Diferentes algoritmos não produzem necessariamente os mesmos resultados
- Quando há uma considerável sobreposição entre os grupos iniciais, pode conduzir a uma solução diferente da verdadeira.
- Dificuldades devido a forma dos agrupamentos:

O PROBLEMA DE ANÁLISE

Figura 15. Alguns possíveis padrões de pontos quando existem dois agrupamentos e duas variáveis X_1 e X_2



O PROBLEMA DE ANÁLISE

- (a) e (b): Provavelmente será encontrado por qualquer algoritmo razoável;
- (c): Alguns algoritmos podem falhar, devido a pontos intermediários;
- (d), (e) e (f): A maioria dos algoritmos teriam problemas.

EXEMPLO

- O conjunto de dados IRIS contém quatro características (comprimento e largura das sépalas e pétalas) de 50 amostras de três espécies de Iris (Iris setosa, Iris virginica e Iris versicolor)



Iris Versicolor



Iris Setosa



Iris Virginica

from [Machine Learning in R for beginners](#)