



**XIX COBREAP | Foz do Iguaçu**

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

AGOSTO 2017

**Prof. Dr. Norberto Hochheim**  
**UFSC**  
**IBAPE-SC**

# **Análise de Cluster aplicada à avaliação de bens**

# Análise de Clusters ou Análise de Agrupamentos

- é o processo de **agrupar** um conjunto de objetos em classes de **objetos similares**
  - de acordo com algum critério ou métrica
- ✓ Um cluster é uma coleção de objetos que são **similares uns aos outros** e **dissimilares** a objetos pertencentes **a outros clusters**

# Análise de Clusters ou Análise de Agrupamentos

Ou seja:

- deve existir o máximo de **semelhança** entre os **elementos** de um **mesmo subgrupo** e a máxima **dissimilaridade** entre elementos de **subgrupos distintos**

# Objetivos da *Análise de Clusters*

- Exploração dos dados
- Redução de dados
- Geração de hipóteses
- Predição

# Aplicações da Análise de Clusters

- **Áreas tradicionais:**
  - Arqueologia, Biologia, Geologia, Medicina, Psiquiatria...
- **Novos desafios:**
  - Análise de mercados,
  - Planejamento e Gestão Territorial
  - Avaliação de Imóveis

# Tipos de dados e variáveis

## ➤ Variáveis numéricas:

- Podem ser reais ou inteiras.

Ex: Área, Número de Quartos, etc.

## ➤ Variáveis binárias:

- Possuem somente dois estados

Ex: Prédio tem ou não elevador (1 ou 0)

## ➤ Variáveis nominais:

Ex: Cor do prédio (branco, verde, azul, etc.)

## ➤ Variáveis compostas por vários tipos

Ex: variáveis de interação

# Apresentação dos dados

- **Matriz de dados:** **colunas** representam as **variáveis** e as **linhas** representam as **observações**

$$\begin{bmatrix} x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix}$$

# Normalização de variáveis numéricas

## ➤ Objetivos:

- minimizar os problemas devidos ao uso de **unidades** ou **dispersões distintas** entre as variáveis
- deixar as **variáveis** com **mesmo peso**

# Normalização de variáveis numéricas

- Amplitude:

$$s_{ij} = \frac{x_{ij} - \min_f}{\max_f - \min_f}$$

- Z-score:

$$z_{ij} = \frac{x_{ij} - m_f}{\sigma_f}$$

- Desvio absoluto médio:

$$s_f = \frac{1}{n} (|x_{1f} - m_f| + |x_{2f} - m_f| + \dots + |x_{nf} - m_f|)$$

sendo

$$m_f = \frac{1}{n} (x_{1f} + x_{2f} + \dots + x_{nf}).$$

# Medidas de Similaridade

- A análise de **cluster** busca agrupar **dados** baseando-se na **similaridade** entre eles
- Os grupos são determinados de forma a obter-se **homogeneidade dentro** dos grupos e **heterogeneidade entre** eles
- As medidas de **similaridade** são utilizadas na análise de **cluster** de forma a determinar a **distância entre elementos**

# Medidas de Similaridade

- Matriz de distâncias (Matriz de similaridade):

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix}$$

- Propriedades:

$$d(i,j) \geq 0$$

$$d(i,i) = 0$$

$$d(i,j) = d(j,i)$$

$$d(i,j) \leq d(i,k) + d(k,j)$$

# Medidas de Similaridade

- Sejam os elementos:

$$X = [X_1, X_2, \dots, X_p] \text{ e } Y = [Y_1, Y_2, \dots, Y_p]$$

- Distância Euclidiana:**

$$d_{xy} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2} = \sqrt{\sum_{i=1}^p (X_i - Y_i)^2}$$

- Distância Euclidiana Quadrática:**

$$d_{xy} = (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2 = \sum_{i=1}^p (X_i - Y_i)^2$$

# Medidas de Similaridade

- Sejam os elementos:

$$X = [X_1, X_2, \dots, X_p] \text{ e } Y = [Y_1, Y_2, \dots, Y_p]$$

- Distância de Manhattan:**

$$d_{xy} = |X_1 - Y_1| + |X_2 - Y_2| + \dots + |X_p - Y_p| = \sum_{i=1}^p |X_i - Y_i|$$

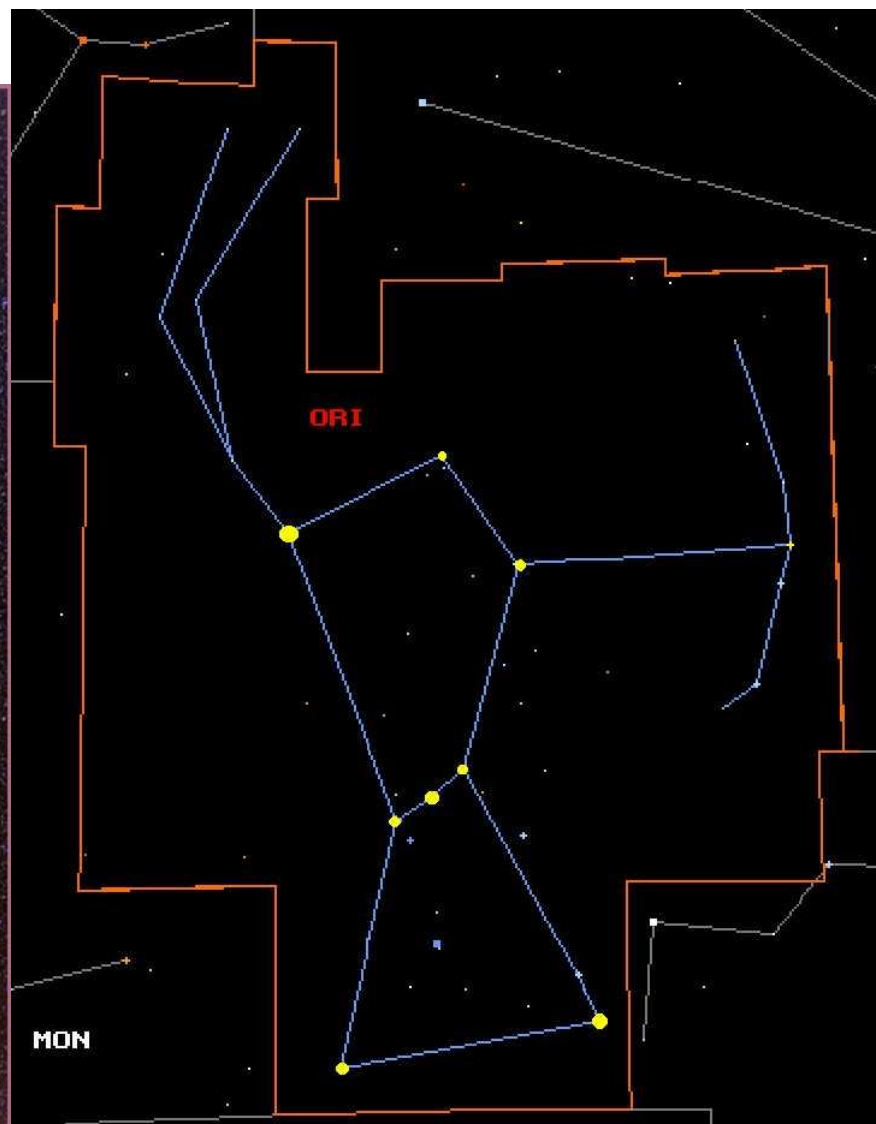
- Distância de Chebychev:**

$$d_{xy} = \text{máximo}(|X_1 - Y_1|, |X_2 - Y_2|, \dots, |X_p - Y_p|)$$

# Métodos de Clusterização

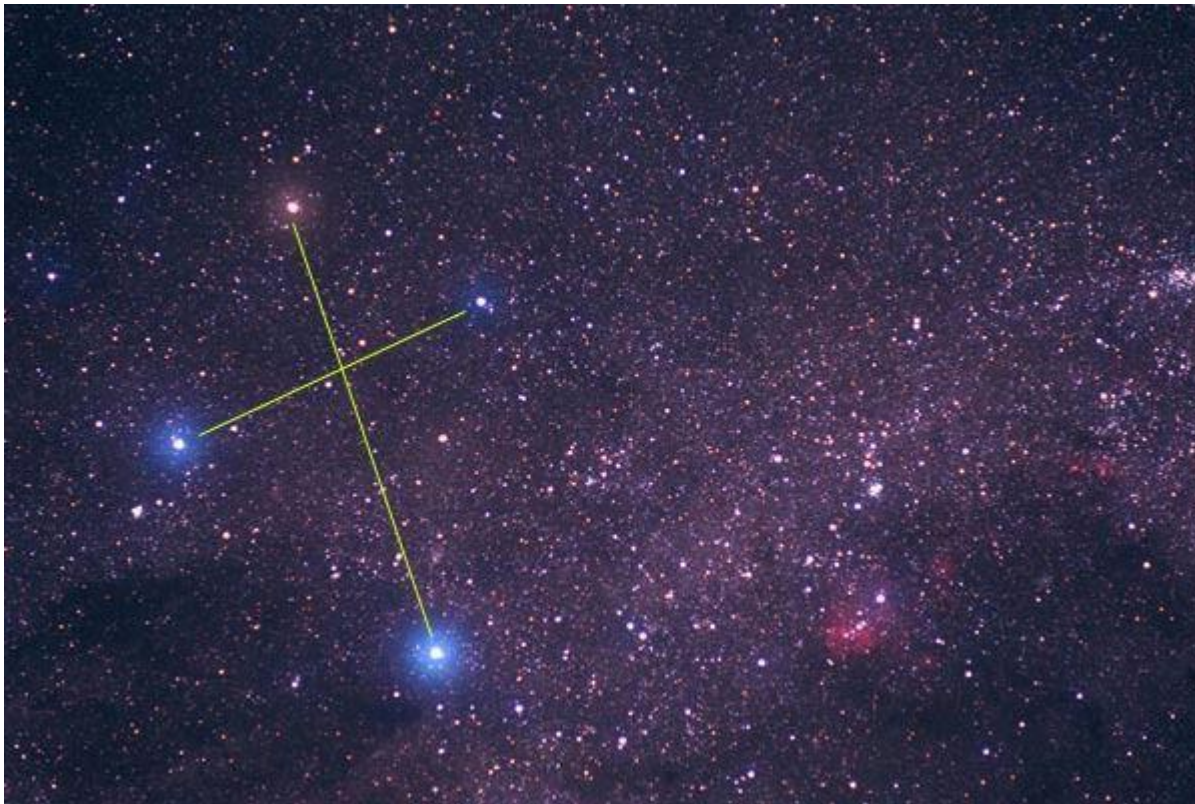
- **Gráfico:**
  - Padrões são determinados por “visualização”
- **Hierárquico:**
  - cria uma decomposição hierárquica dos objetos
- **Particionamento (método não hierárquico):**
  - Segmenta um banco de dados com  $n$  objetos num conjunto com  $k$  clusters

# Métodos Gráficos



# Métodos Gráficos

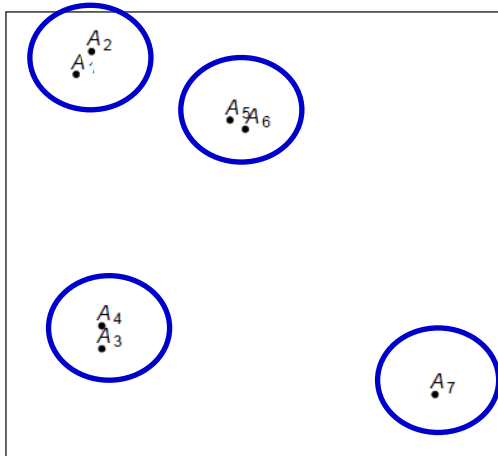
- **vizualizam os clusters a partir da representação gráfica dos objetos ou das variáveis**



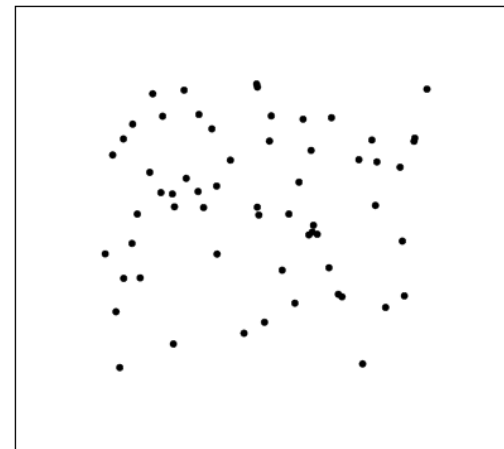
# Métodos Gráficos

- visualizam os clusters a partir da representação gráfica dos objetos ou das variáveis

Sete objetos e três clusters:

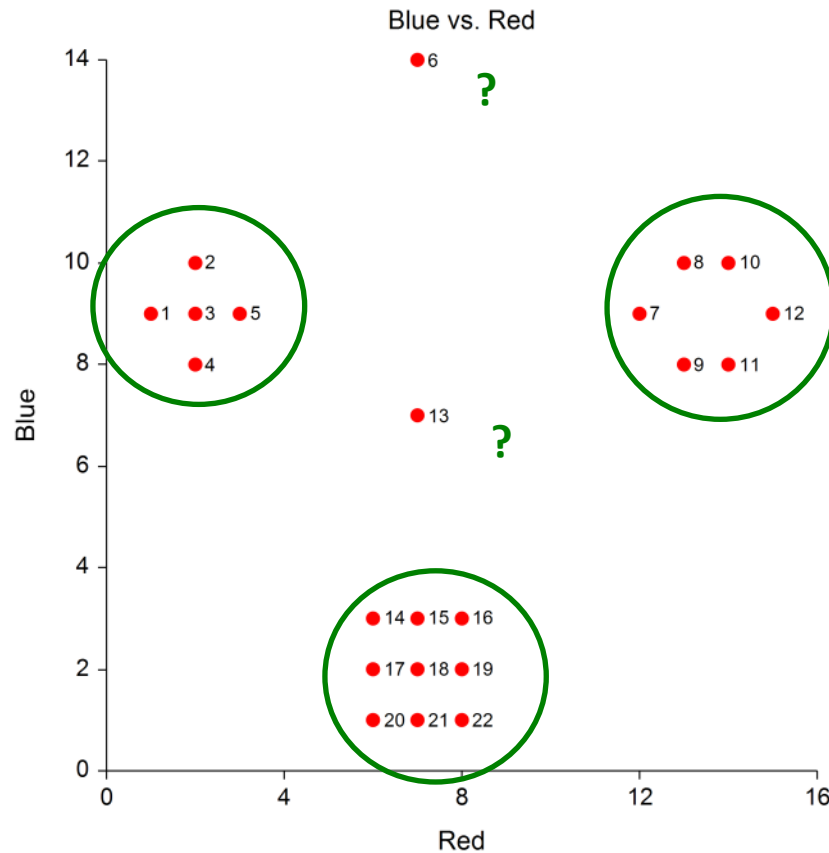


Sem estrutura aparente de grupos:



Fonte: BRANCO (----)

# Métodos Gráficos



Fonte: **NCSS Statistical Software**

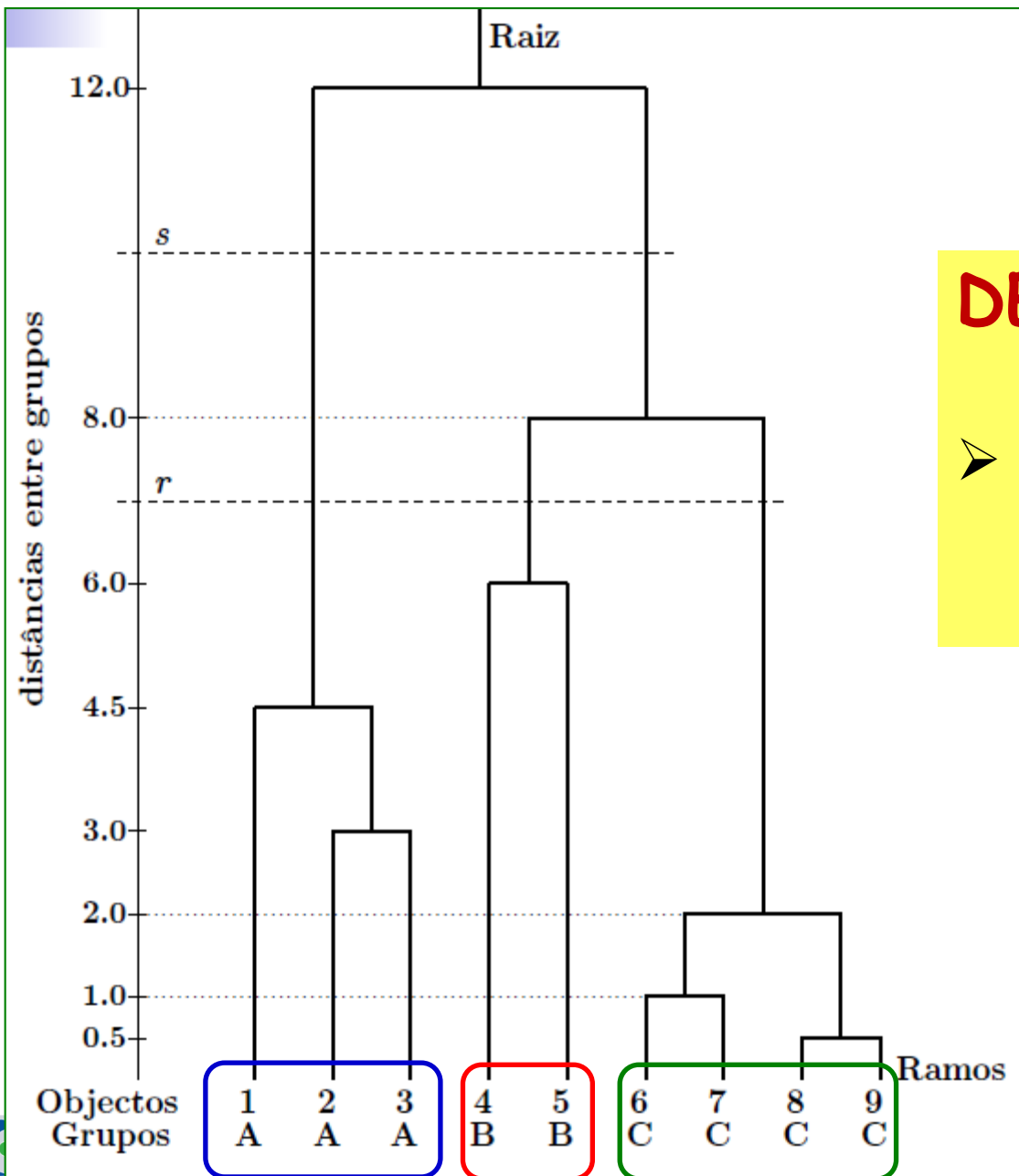
# Métodos Gráficos

## ➤ Limitações:

- Usam espaços de **dimensão  $\leq 3$**
- **Difícil** para muitos objetos
- Método **subjetivo**

# Métodos Hierárquicos

- Fazem sucessivos agrupamentos (**métodos aglomerativos**) ou divisões de elementos (**métodos divisivos**)
- Os grupos são geralmente representados por um diagrama bi-dimensional chamado de **dendrograma** ou diagrama de árvore
- Cada **ramo** representa **um elemento**, enquanto a **raiz** representa o agrupamento de **todos** os elementos



## DENDROGRAMA

- mostra a sequência de passos e os níveis de fusão

Fonte: BRANCO (---)

# Métodos Hierárquicos:

## Métodos Aglomerativos

- Formam-se os grupos reunindo-se os **objetos** em **grupos** cada vez **maiores**
- O processo continua até que todos os objetos sejam membros de um **único grupo**
- Respeita-se uma determinada **seqüência de aglomeração**

# Métodos Hierárquicos:

## Métodos Aglomerativos

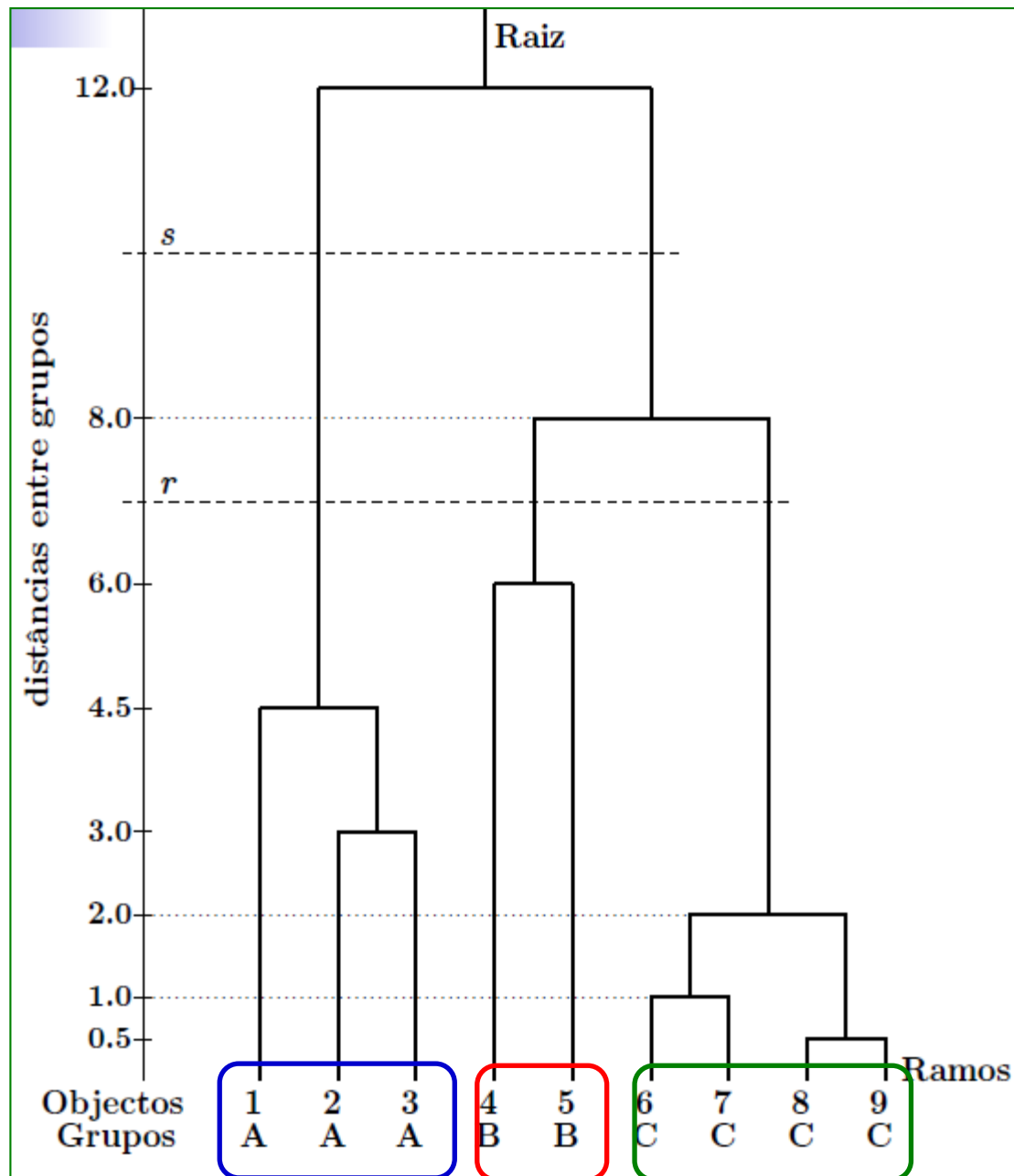
- Inicia-se com o cálculo de uma **matriz de distâncias** entre as variáveis e finaliza-se com um **dendrograma**
- Dendrograma mostra os **agrupamentos sucessivos** dos indivíduos, até formarem um **único grupo**

# Métodos Hierárquicos:

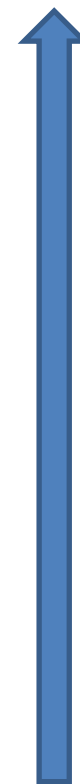
## Métodos Divisivos

- Parte-se de um **único** grupo que é subdivididos em **dois** subgrupos
- Esses subgrupos são, posteriormente, **subdivididos** em outros subgrupos dissimilares
- O processo é repetido até que haja **tantos subgrupos quantos objetos**

Métodos Divisivos



Métodos Aglomerativos



# Métodos Hierárquicos Aglomerativos

- **Critério da fusão** entre um objeto e um grupo (ou entre dois grupos): serão reunidos os grupos que têm **maior similaridade** entre si
- O problema é: **como calcular** esta similaridade?
- O método de **cálculo depende** do **método de aglomeração** escolhido

# Métodos Hierárquicos Aglomerativos

- **Diferentes métodos** de aglomeração podem conduzir a **diferentes clusters**
- Cada método tem suas características, com vantagens e desvantagens
- Métodos são caracterizados de acordo com o **critério** utilizado para **definir as distâncias** entre grupos

# Métodos Hierárquicos Aglomerativos

Conceitos de agrupamento aglomerativo:

- 1) Métodos de **ligação** (*single linkage, complete linkage, average linkage, median linkage*)
- 2) Métodos de **centróide**
- 3) Métodos de soma de erros quadráticos ou variância (método de **Ward**)

# Métodos Hierárquicos Aglomerativos

| MÉTODO                            | DISTÂNCIA                                                                                                              | CARACTERÍSTICAS                                              |
|-----------------------------------|------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| Ligação por vizinho mais próximo  | $d_{(UV)W} = \min(d_{UW}, d_{VW})$                                                                                     | Sensibilidade à ruídos.<br>Encadeamento.                     |
| Ligação por vizinho mais distante | $d_{(UV)W} = \max(d_{UW}, d_{VW})$                                                                                     | Tendência a formar grupos compactos.                         |
| Ligação por média                 | $d_{(UV)W} = \frac{(N_u \cdot d_{UW} + N_v \cdot d_{VW})}{N_u + N_v}$                                                  | Tendência a formar grupos com número de elementos similares. |
| Ligação por centróide             | $d_{(UV)W} = \frac{N_U \cdot d_{UW} + N_V \cdot d_{VW}}{N_U + N_V} - \frac{N_U \cdot N_V \cdot d_{UV}}{(N_U + N_V)^2}$ | Robustez à ruídos.<br>Reversão.                              |
| Ligação por mediana               | $d_{(UV)W} = \frac{d_{UW} + d_{VW}}{2} - \frac{d_{UV}}{4}$                                                             | Robustez à ruídos.                                           |
| Ligação de Ward                   | $d_{(UV)W} = \frac{((N_W + N_U) \cdot d_{UW} + (N_W + N_V) \cdot d_{VW} - N_W \cdot d_{UV})}{N_W + N_U + N_V}$         | Sensibilidade à ruídos.                                      |

## Métodos Hierárquicos Aglomerativos

Fonte: DONI (2004)

# Métodos não-Hierárquicos (Métodos por Particionamento)

- Operam sobre **matriz de dados**
- Aplicam-se apenas a **objetos**
- Os grupos devem satisfazer os critérios de **coesão interna** e **isolamento externo**
- O **número** de grupos **é fixado** *à priori*
- Um objeto pode **viajar** por vários clusters

# Métodos não-Hierárquicos (Métodos por Particionamento)

- Encontrar **k** clusters que otimizam um **critério de particionamento** escolhido (k é arbitrado)
- Principais heurísticas:
  - algoritmos **k-means** (k-médias) e **k-medoids** (k-medóides)
  - **k-means**: Cada cluster é representado pelo centro (**centróide**) do cluster
  - **k-medoids** ou PAM (Partition Around Medoids) : Cada cluster é representado por **um dos objetos** no cluster

# Agrupamentos Fuzzy

- Métodos por **particionamento**: cada elemento pertence à um grupo (**agrupamentos rígidos**)
- Agrupamentos **fuzzy** estabelecem **um grau de associação** de cada elemento em cada grupo
- um elemento pertence a **diferentes grupos**, com **diferentes graus** de associação

# Agrupamentos Fuzzy: Exemplo

6 Elementos,  
2 Variáveis

| ELEMENTOS | X | Y |
|-----------|---|---|
| 1         | 4 | 3 |
| 2         | 2 | 7 |
| 3         | 4 | 7 |
| 4         | 2 | 3 |
| 5         | 3 | 5 |
| 6         | 6 | 1 |

2 Grupos

| ELEMENTO | GRUPO 1 | GRUPO 2 |
|----------|---------|---------|
| 1        | 0.9515  | 0.0485  |
| 2        | 0.0518  | 0.9482  |
| 3        | 0.0700  | 0.9300  |
| 4        | 0.6605  | 0.3395  |
| 5        | 0.1451  | 0.8549  |
| 6        | 0.8902  | 0.1098  |

Elemento **1** pertence **95%** ao **Grupo 1** e **5%** ao **Grupo 2**  
Elemento **6** pertence **89%** ao **Grupo 1** e **11%** ao **Grupo 2**

Fonte: Doni (2004)

# ANÁLISE MULTIVARIADA

- **Estatística Multidimensional:**
  - Estudo das **relações** entre múltiplas **variáveis**:
    - Relações
    - Semelhanças
    - Diferenças
  - Estudo das **relações** entre **variáveis** e **indivíduos**

# ANÁLISE FATORIAL

- Método para reduzir um **conjunto extenso** de **variáveis correlacionadas** em um **número pequeno** de novas variáveis **não correlacionadas** (fatores)
- Os fatores **não são diretamente observáveis**, seguem conceitos de natureza abstrata

# Aplicações em avaliações de imóveis

Como a **ANÁLISE DE CLUSTERS** agrupa os indivíduos de uma amostra num número restrito de **classes homogêneas**, pode-se usá-la **complementarmente** à **ANÁLISE FATORIAL**

# Aplicações em avaliações de imóveis

- A partir de um conjunto extenso de variáveis **quantitativas** e **qualitativas**:
  - queremos **classificar** um **grande número** de observações

# Aplicações em avaliações de imóveis

- Proposta metodológica:
  - 1) Fazer Análise Fatorial
  - 2) Classificar os indivíduos a partir das coordenadas fatoriais
  - 3) Analisar o mercado imobiliário com as informações obtidas
  - 4) Fazer avaliações de imóveis usando modelos para cada classe

# Estudo de caso

- **Local: centro de Florianópolis (SC)**
  - **Imóveis tipo Apartamento**
  - **Amostra: 125 dados**
  - **Variáveis analisadas: 10**
- 
- **Dados foram georreferenciados**
  - **Usado SIG para auxiliar nas análises**

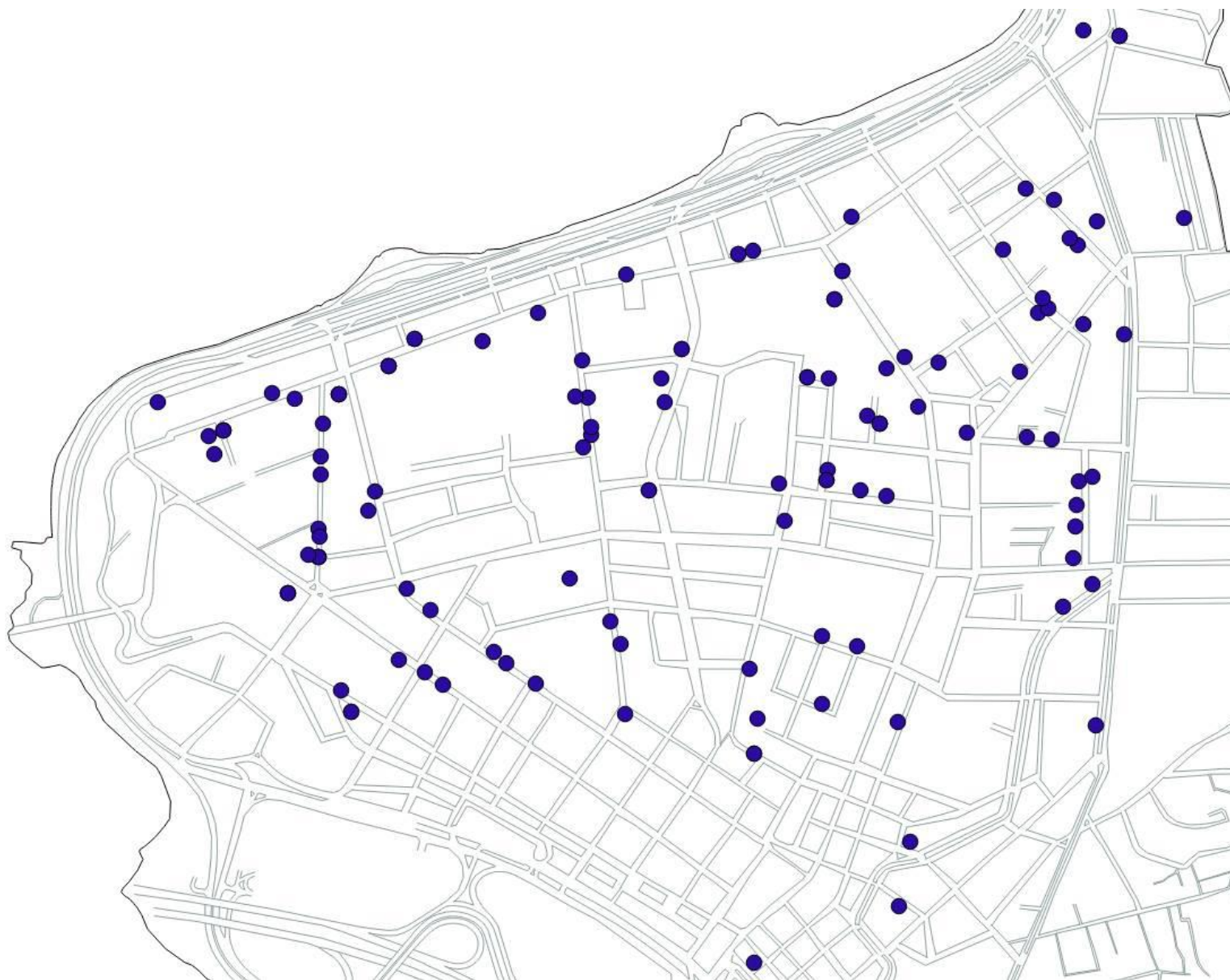
# Estudo de caso

## ➤ Variáveis analisadas:

- Área total
- Área privativa
- relação Área total/Área provativa
- nº de Quartos
- nº de Suítes
- nº de Garagens
- Idade
- Padrão de Construção
- Estado de Conservação
- Distância ao polo valorizador (Av. Beiramar)

| IDENT  | VT           | VU_Atot   | VU_Aprivat | Area_Total | Area_Privat | Aprivat/Atotal | N_Quartos | N_Suites | N_garagens | Idade | Padrão | Conservação | Dist. Beira-Mar | Coord. E   | Coord. N     |
|--------|--------------|-----------|------------|------------|-------------|----------------|-----------|----------|------------|-------|--------|-------------|-----------------|------------|--------------|
| Cen_01 | 205.000,00   | 3.474,58  | 4270,83    | 59,00      | 48,00       | 0,81356        | 1         | 1        | 1          | 31    | 1      | 1           | 670,0           | 741.087,00 | 6.945.316,00 |
| Cen_02 | 199.000,00   | 5.685,71  | 7107,14    | 35,00      | 28,00       | 0,8            | 1         | 0        | 0          | 28    | 1      | 1           | 975,0           | 741.829,43 | 6.945.239,84 |
| Cen_03 | 365.000,00   | 8.111,11  | 11406,25   | 45,00      | 32,00       | 0,71111        | 1         | 1        | 1          | 12    | 1      | 2           | 665,0           | 741.373,04 | 6.945.406,75 |
| Cen_04 | 545.400,00   | 6.222,48  | 13466,67   | 87,65      | 40,50       | 0,46207        | 1         | 0        | 1          | 7     | 3      | 3           | 190,0           | 742.503,00 | 6.946.563,00 |
| Cen_05 | 220.000,00   | 5.641,03  | 7586,21    | 39,00      | 29,00       | 0,74359        | 1         | 0        | 1          | 28    | 1      | 1           | 700,0           | 741.350,00 | 6.945.427,00 |
| Cen_06 | 480.000,00   | 6.575,34  | 9230,77    | 73,00      | 52,00       | 0,71233        | 1         | 0        | 1          | 7     | 3      | 3           | 450,0           | 741.027,00 | 6.945.602,00 |
| Cen_07 | 580.000,00   | 10.740,74 | 13809,52   | 54,00      | 42,00       | 0,77778        | 1         | 0        | 1          | 8     | 3      | 2           | 270,0           | 741.993,00 | 6.946.130,00 |
| Cen_08 | 175.000,00   | 3.500,00  | 4605,26    | 50,00      | 38,00       | 0,76           | 1         | 0        | 1          | 31    | 1      | 1           | 670,0           | 741.087,00 | 6.945.316,00 |
| Cen_09 | 170.000,00   | 4.857,14  | 6071,43    | 35,00      | 28,00       | 0,8            | 1         | 0        | 0          | 30    | 1      | 1           | 700,0           | 741.070,00 | 6.945.357,00 |
| Cen_10 | 180.000,00   | 4.186,05  | 5142,86    | 43,00      | 35,00       | 0,81395        | 1         | 0        | 0          | 28    | 1      | 1           | 975,0           | 741.829,43 | 6.945.239,84 |
| Cen_11 | 295.000,00   | 5.175,44  | 6704,55    | 57,00      | 44,00       | 0,77193        | 1         | 0        | 0          | 12    | 2      | 3           | 380,0           | 741.529,00 | 6.945.827,00 |
| Cen_12 | 200.000,00   | 4.545,45  | 6060,61    | 44,00      | 33,00       | 0,75           | 1         | 0        | 1          | 20    | 2      | 3           | 1400,0          | 742.097,00 | 6.944.958,00 |
| Cen_14 | 695.000,00   | 7.239,58  | 9928,57    | 96,00      | 70,00       | 0,72917        | 2         | 1        | 1          | 9     | 2      | 1           | 450,0           | 740.971,52 | 6.945.536,18 |
| Cen_15 | 270.000,00   | 2.727,27  | 3214,29    | 99,00      | 84,00       | 0,84848        | 2         | 1        | 0          | 30    | 1      | 1           | 680,0           | 741.176,00 | 6.945.412,00 |
| Cen_16 | 1.100.000,00 | 6.547,62  | 8461,54    | 168,00     | 130,00      | 0,77381        | 2         | 1        | 2          | 7     | 3      | 3           | 450,0           | 741.027,00 | 6.945.602,00 |
| Cen_17 | 180.000,00   | 2.571,43  | 3214,29    | 70,00      | 56,00       | 0,8            | 2         | 0        | 0          | 30    | 1      | 1           | 1600,0          | 741.831,00 | 6.944.853,00 |
| Cen_19 | 420.000,00   | 3.387,10  | 6885,25    | 124,00     | 61,00       | 0,49194        | 2         | 0        | 1          | 20    | 2      | 3           | 280,0           | 741.031,00 | 6.945.788,00 |
| Cen_20 | 1.500.000,00 | 8.823,53  | 10563,38   | 170,00     | 142,00      | 0,83529        | 2         | 2        | 1          | 20    | 2      | 3           | 60,0            | 740.731,00 | 6.945.889,00 |
| Cen_21 | 540.000,00   | 6.000,00  | 7714,29    | 90,00      | 70,00       | 0,77778        | 2         | 1        | 1          | 15    | 2      | 3           | 120,0           | 742.437,00 | 6.946.575,00 |
| Cen_23 | 695.000,00   | 7.722,22  | 9520,55    | 90,00      | 73,00       | 0,81111        | 2         | 1        | 1          | 9     | 2      | 1           | 450,0           | 740.971,52 | 6.945.536,18 |
| Cen_24 | 584.590,00   | 6.089,48  | 8993,69    | 96,00      | 65,00       | 0,67708        | 2         | 1        | 2          | 20    | 2      | 3           | 880,0           | 741.956,00 | 6.945.457,00 |
| Cen_25 | 760.000,00   | 7.755,10  | 11014,49   | 98,00      | 69,00       | 0,70408        | 2         | 1        | 2          | 9     | 3      | 3           | 450,0           | 740.971,52 | 6.945.536,18 |
| Cen_27 | 769.000,00   | 4.660,61  | 9734,18    | 165,00     | 79,00       | 0,47879        | 2         | 2        | 2          | 7     | 3      | 3           | 450,0           | 741.027,00 | 6.945.602,00 |
| Cen_28 | 1.650.000,00 | 7.932,69  | 14732,14   | 208,00     | 112,00      | 0,53846        | 2         | 2        | 2          | 5     | 3      | 3           | 160,0           | 741.064,00 | 6.945.902,00 |
| Cen_29 | 990.000,00   | 7.333,33  | 8534,48    | 135,00     | 116,00      | 0,85926        | 2         | 2        | 2          | 17    | 2      | 3           | 300,0           | 742.382,00 | 6.946.261,00 |
| Cen_30 | 1.450.000,00 | 9.415,58  | 13425,93   | 154,00     | 108,00      | 0,7013         | 2         | 2        | 2          | 5     | 3      | 3           | 160,0           | 741.064,00 | 6.945.902,00 |
| Cen_31 | 495.000,00   | 4.759,62  | 6111,11    | 104,00     | 81,00       | 0,77885        | 2         | 2        | 1          | 25    | 2      | 3           | 980,0           | 742.453,00 | 6.945.553,00 |
| Cen_32 | 750.000,00   | 4.120,88  | 6410,26    | 182,00     | 117,00      | 0,64286        | 3         | 1        | 1          | 27    | 1      | 2           | 500,0           | 741.636,00 | 6.945.726,00 |
| Cen_33 | 850.000,00   | 6.296,30  | 7727,27    | 135,00     | 110,00      | 0,81481        | 3         | 1        | 1          | 12    | 2      | 3           | 910,0           | 742.021,00 | 6.945.438,00 |
| Cen_34 | 850.000,00   | 6.538,46  | 7327,59    | 130,00     | 116,00      | 0,89231        | 3         | 1        | 1          | 21    | 2      | 3           | 870,0           | 742.424,00 | 6.945.698,00 |
| Cen_35 | 680.000,00   | 5.396,83  | 6800,00    | 126,00     | 100,00      | 0,79365        | 3         | 1        | 1          | 15    | 2      | 3           | 600,0           | 741.234,00 | 6.945.504,00 |
| Cen_36 | 750.000,00   | 6.578,95  | 7731,96    | 114,00     | 97,00       | 0,85088        | 3         | 1        | 1          | 29    | 2      | 2           | 350,0           | 741.131,00 | 6.945.723,00 |
| Cen_37 | 900.000,00   | 7.758,62  | 9090,91    | 116,00     | 99,00       | 0,85345        | 3         | 1        | 1          | 25    | 2      | 2           | 280,0           | 742.330,00 | 6.946.282,00 |
| Cen_38 | 945.000,00   | 6.176,47  | 7745,90    | 153,00     | 122,00      | 0,79739        | 3         | 1        | 1          | 10    | 2      | 3           | 170,0           | 741.802,00 | 6.946.162,00 |
| Cen_40 | 640.000,00   | 4.155,84  | 6336,63    | 154,00     | 101,00      | 0,65584        | 3         | 1        | 1          | 25    | 2      | 2           | 380,0           | 741.665,00 | 6.945.888,00 |

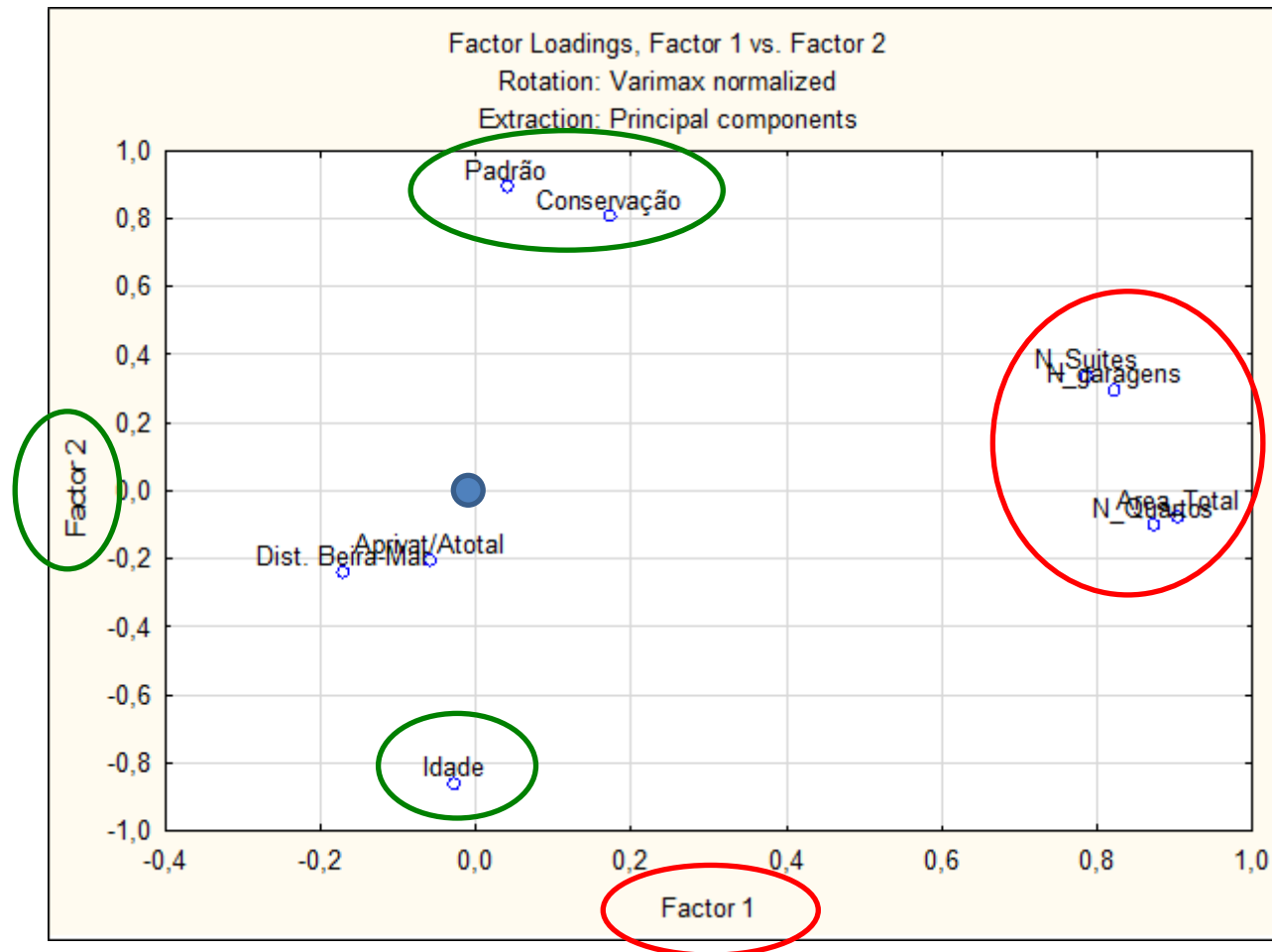




# Análise Fatorial

|                 | <b>Factor Loadings (Varimax normalized)</b><br>Extraction: Principal components<br>(Marked loadings are >,700000) |               |              |               |
|-----------------|-------------------------------------------------------------------------------------------------------------------|---------------|--------------|---------------|
| Variable        | Factor 1                                                                                                          | Factor 2      | Factor 3     | Factor 4      |
| Area_Total      | <b>0,904</b>                                                                                                      | -0,074        | -0,157       | 0,072         |
| Aprivat/Atotal  | -0,059                                                                                                            | -0,201        | <b>0,963</b> | 0,001         |
| N_Quartos       | <b>0,873</b>                                                                                                      | -0,098        | 0,057        | 0,149         |
| N_Suites        | <b>0,787</b>                                                                                                      | 0,337         | 0,119        | -0,007        |
| N_garagens      | <b>0,823</b>                                                                                                      | 0,296         | -0,134       | 0,094         |
| Idade           | -0,028                                                                                                            | <b>-0,859</b> | 0,214        | -0,105        |
| Padrão          | 0,039                                                                                                             | <b>0,897</b>  | -0,136       | 0,141         |
| Conservação     | 0,172                                                                                                             | <b>0,808</b>  | 0,024        | 0,087         |
| Dist. Beira-Mar | -0,171                                                                                                            | -0,237        | 0,000        | <b>-0,952</b> |
| Expl.Var        | 2,940                                                                                                             | 2,509         | 1,052        | 0,981         |
| Prp.Totl        | 0,327                                                                                                             | 0,279         | 0,117        | 0,109         |

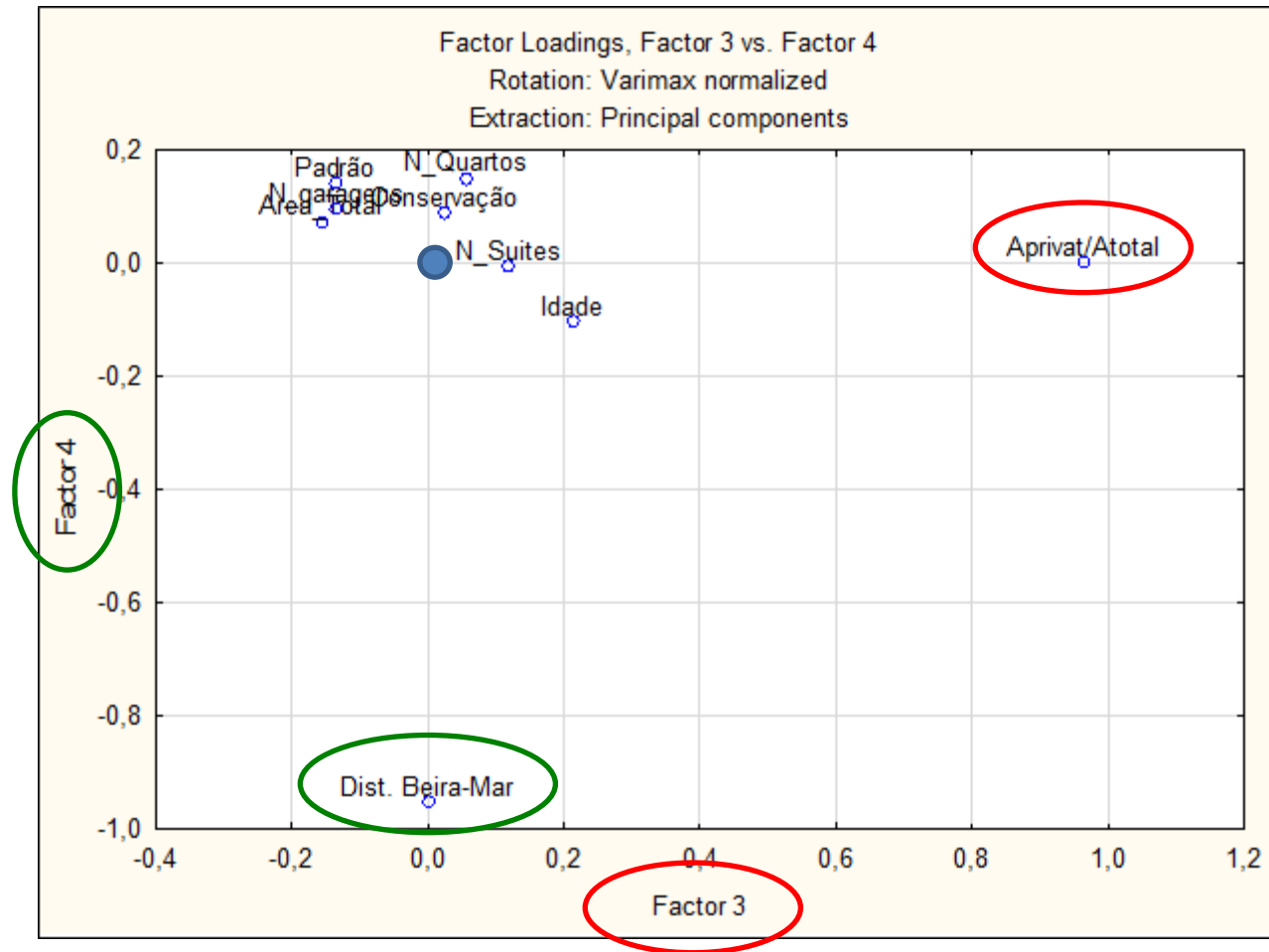
# Análise Fatorial



# Análise Fatorial

- Dois primeiros eixos fatoriais caracterizados por:
  - **EIXO 1**: variáveis de **tamanho** do apartamento
  - **EIXO 2**: variáveis de **qualidade** do condomínio

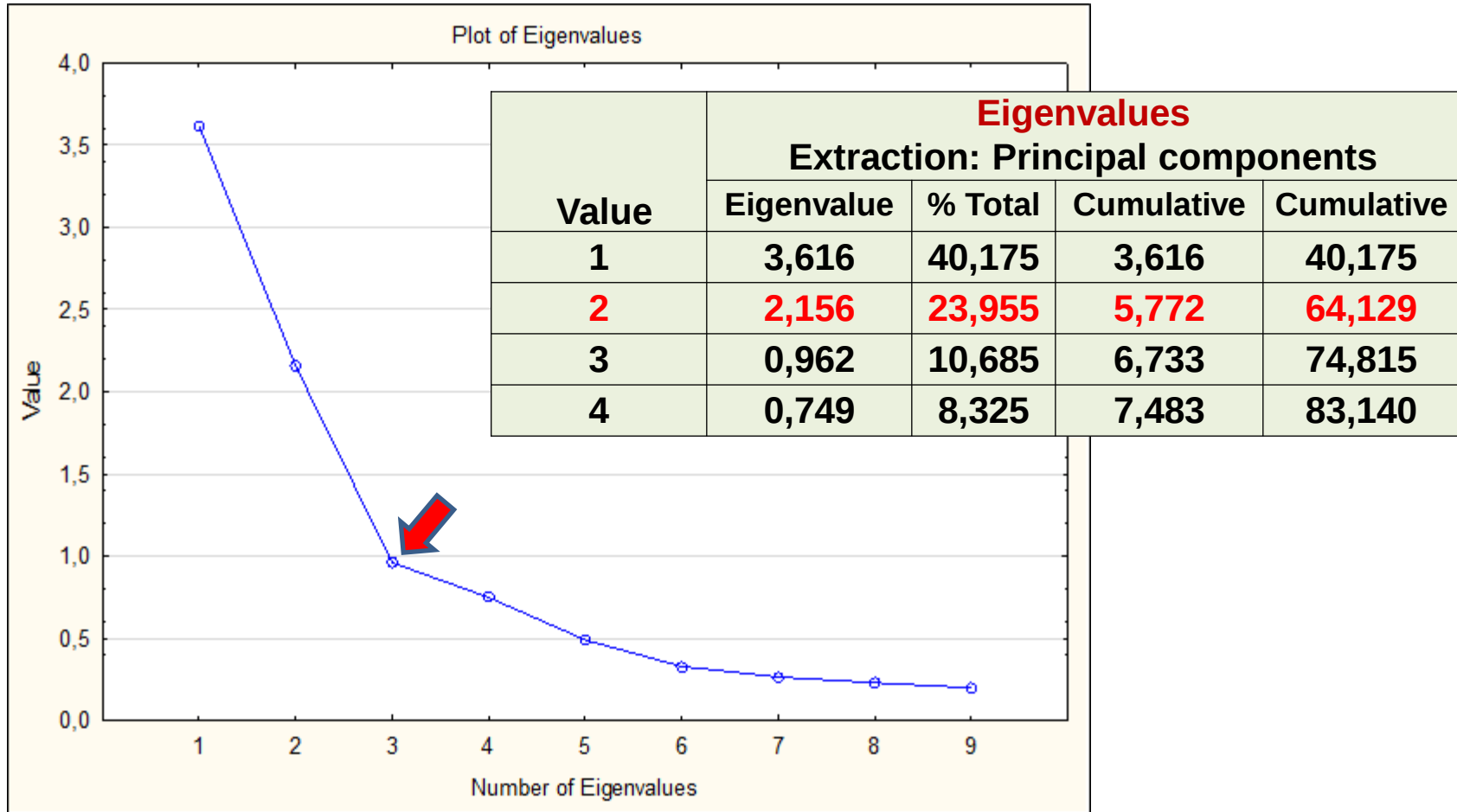
# Análise Fatorial



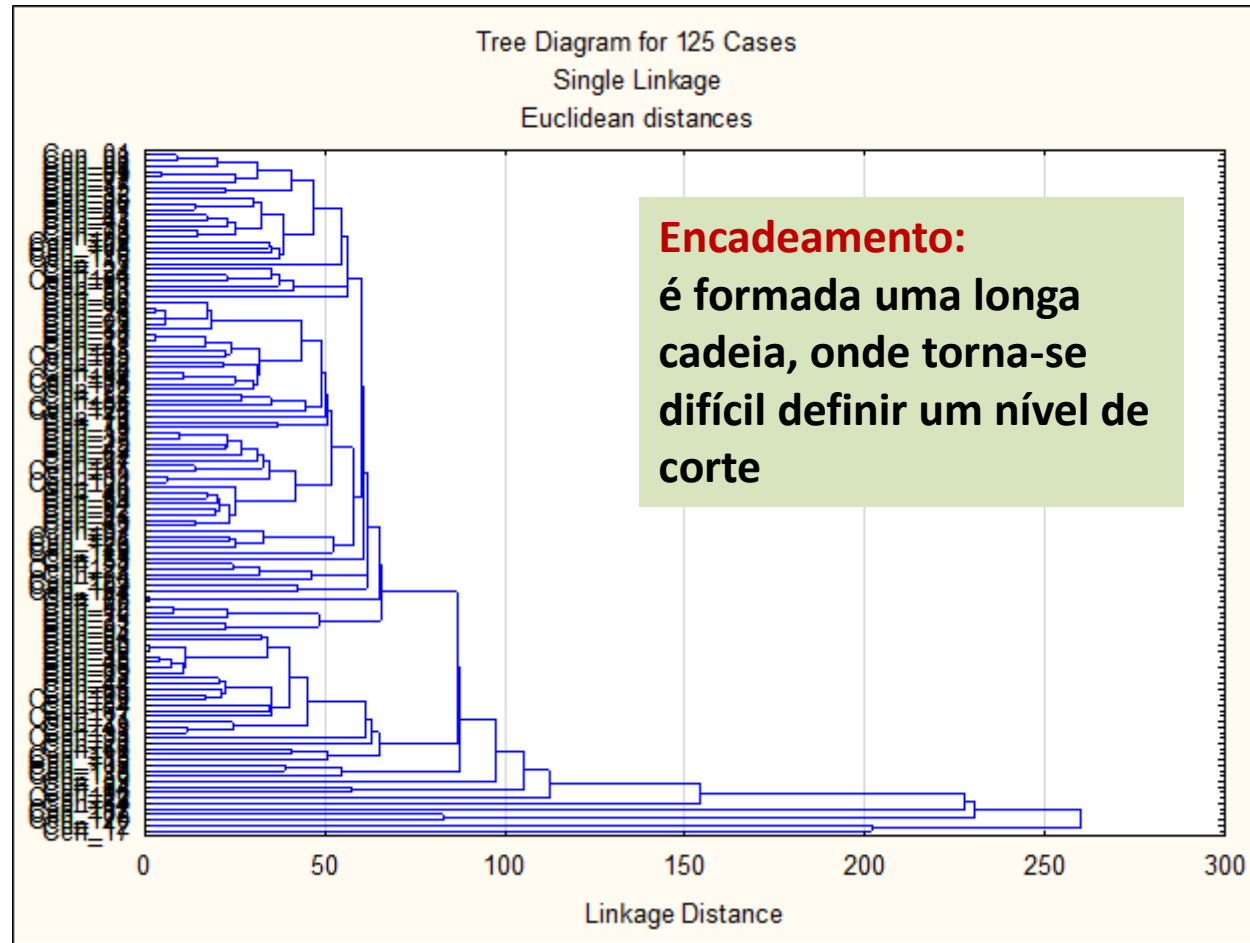
# Análise Fatorial

- Outros eixos fatoriais caracterizados por:
  - **EIXO 3:** variável de **infraestrutura** do condomínio (condomínios com maior infraestrutura tem mais áreas comuns)
  - **EIXO 4:** variável de **distância** ao polo valorizador

# Análise Fatorial

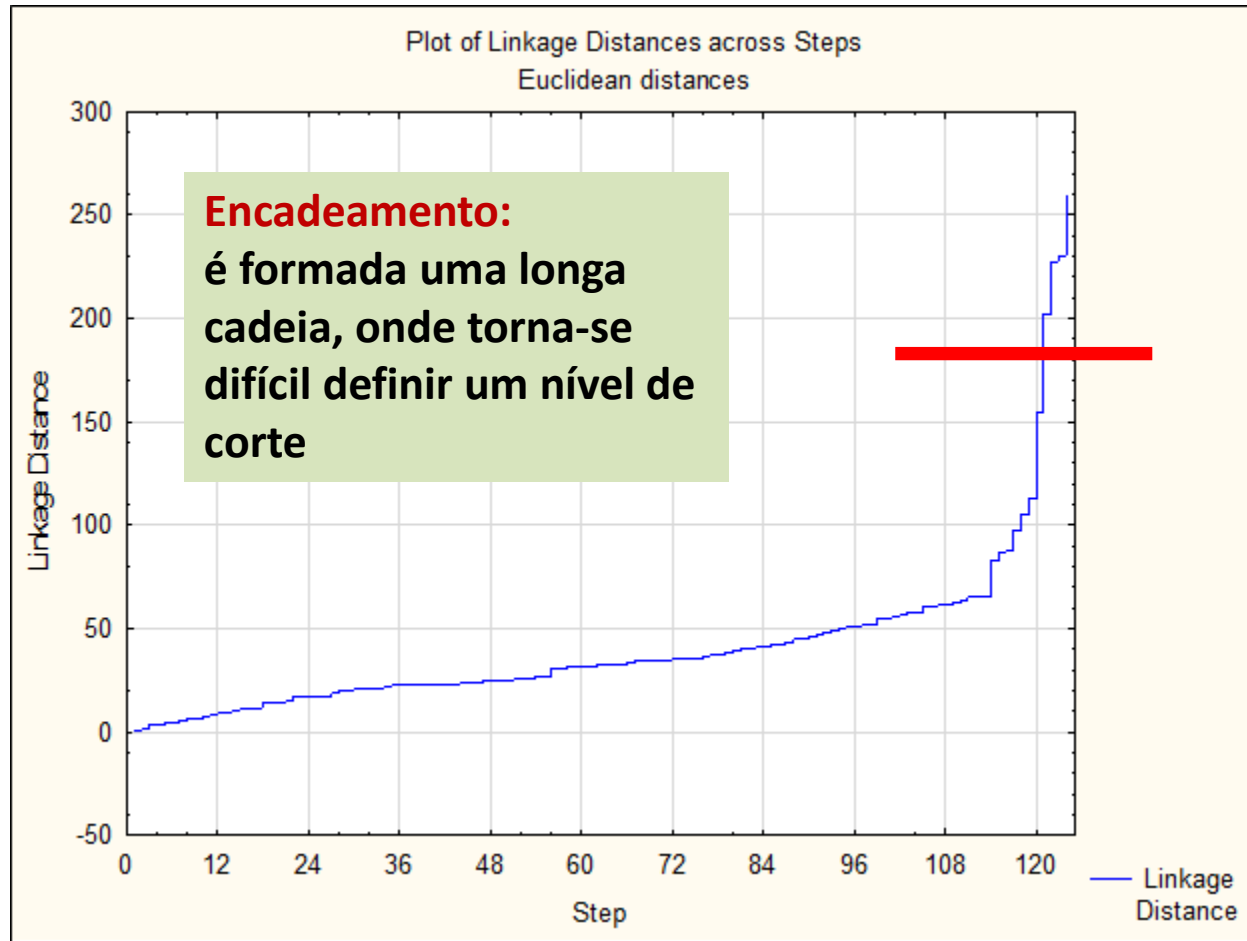


# Classificação Hierárquica



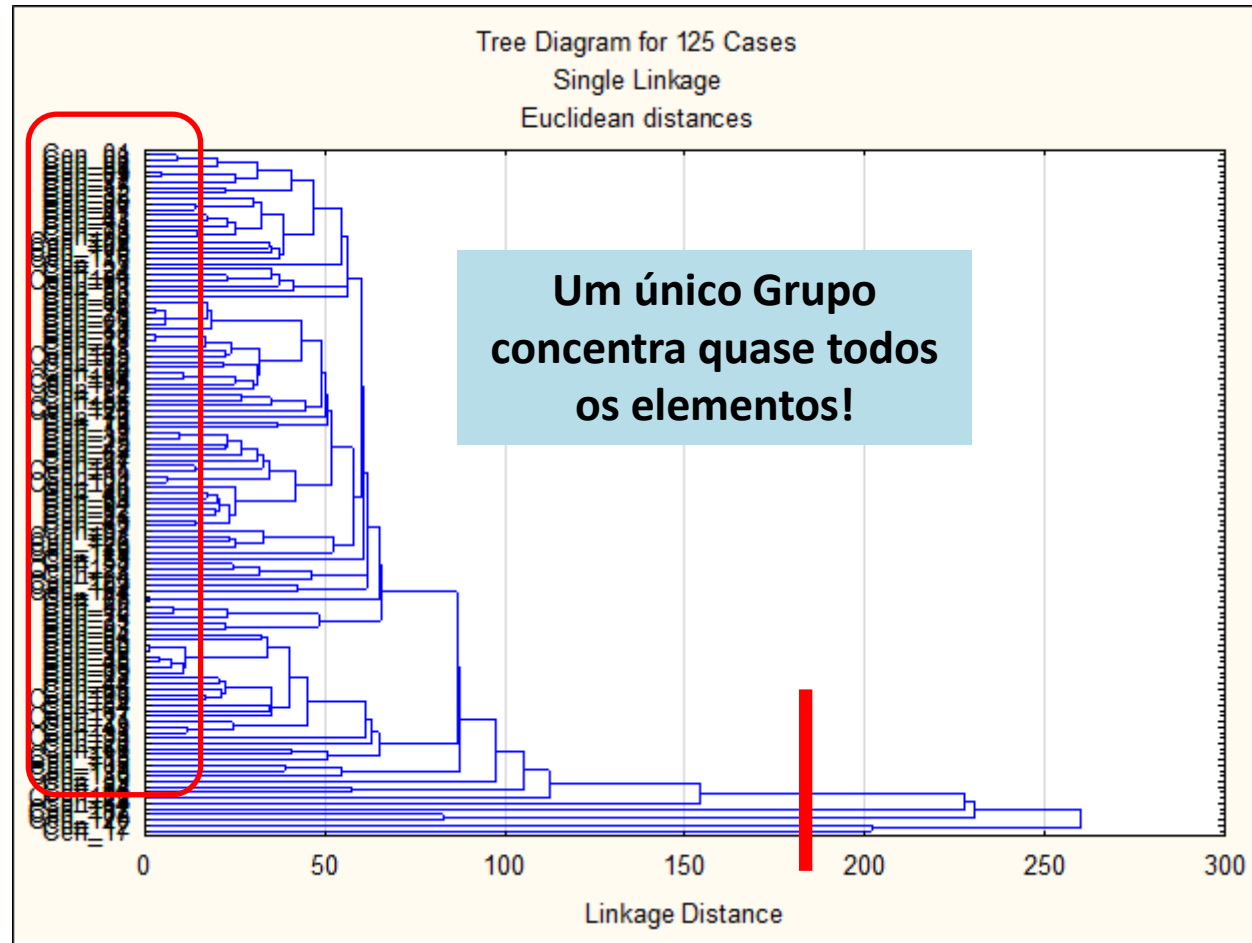
Ligação por vizinho mais próximo

# Classificação Hierárquica



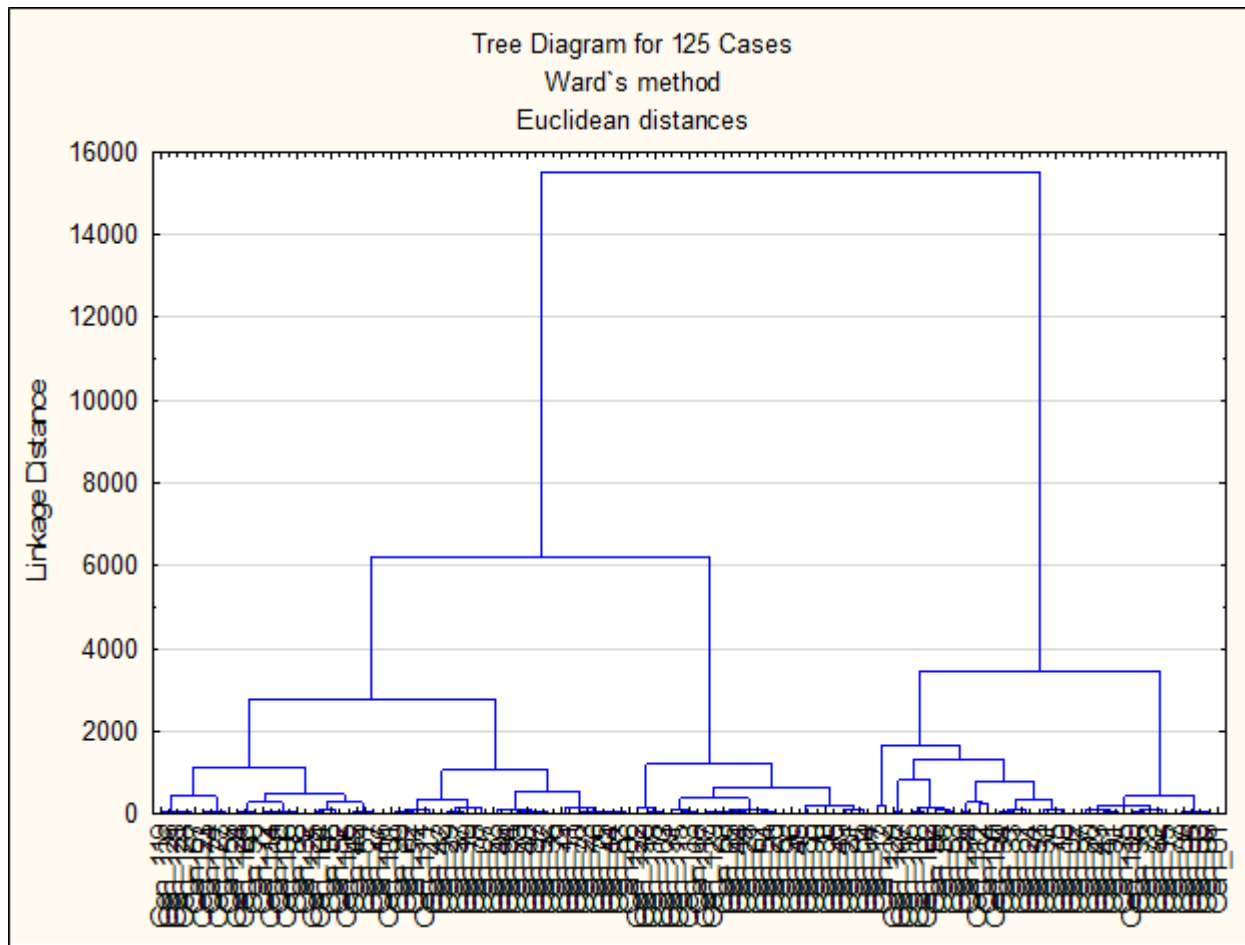
Ligação por vizinho mais próximo

# Classificação Hierárquica



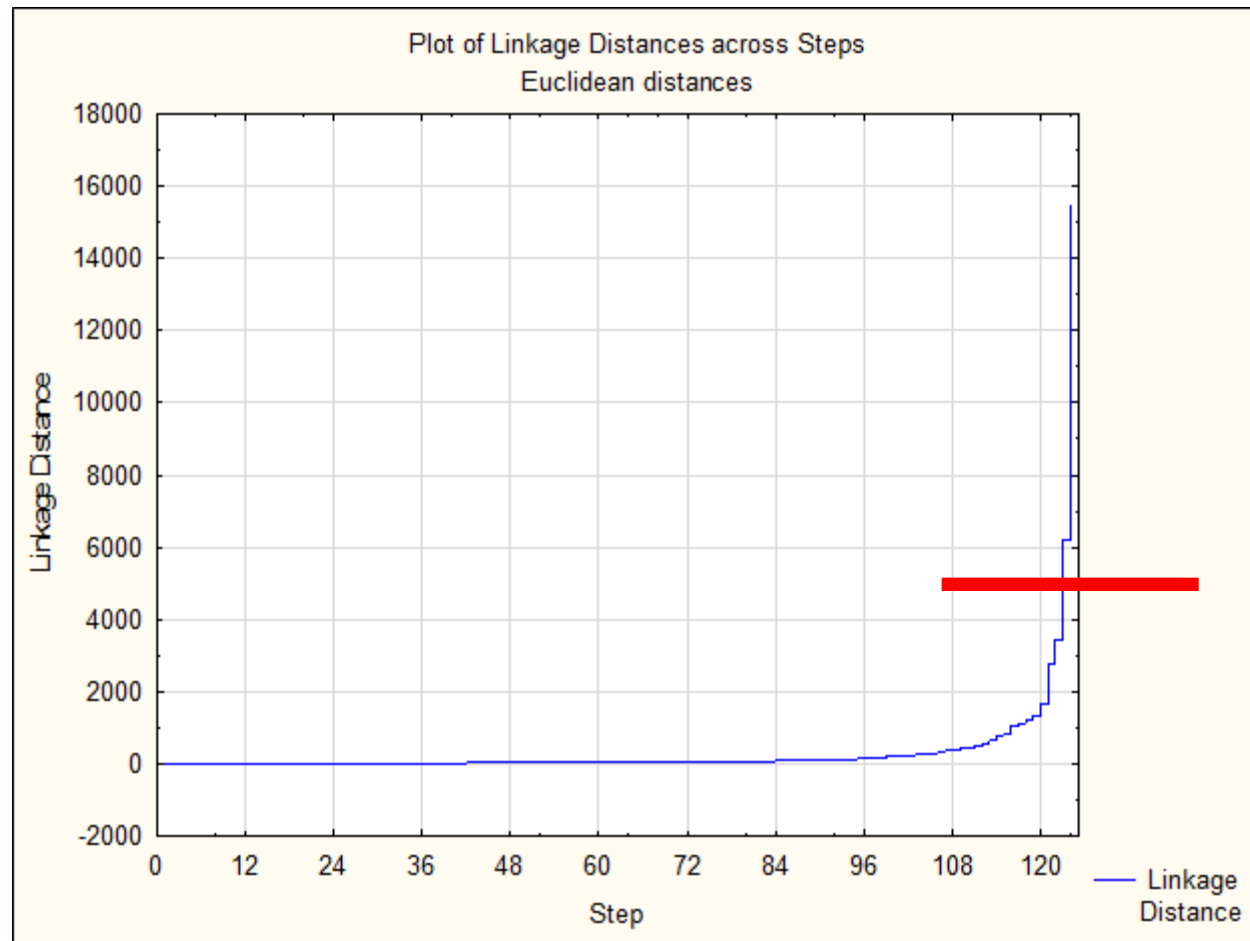
**Ligação por vizinho mais próximo**

# Classificação Hierárquica



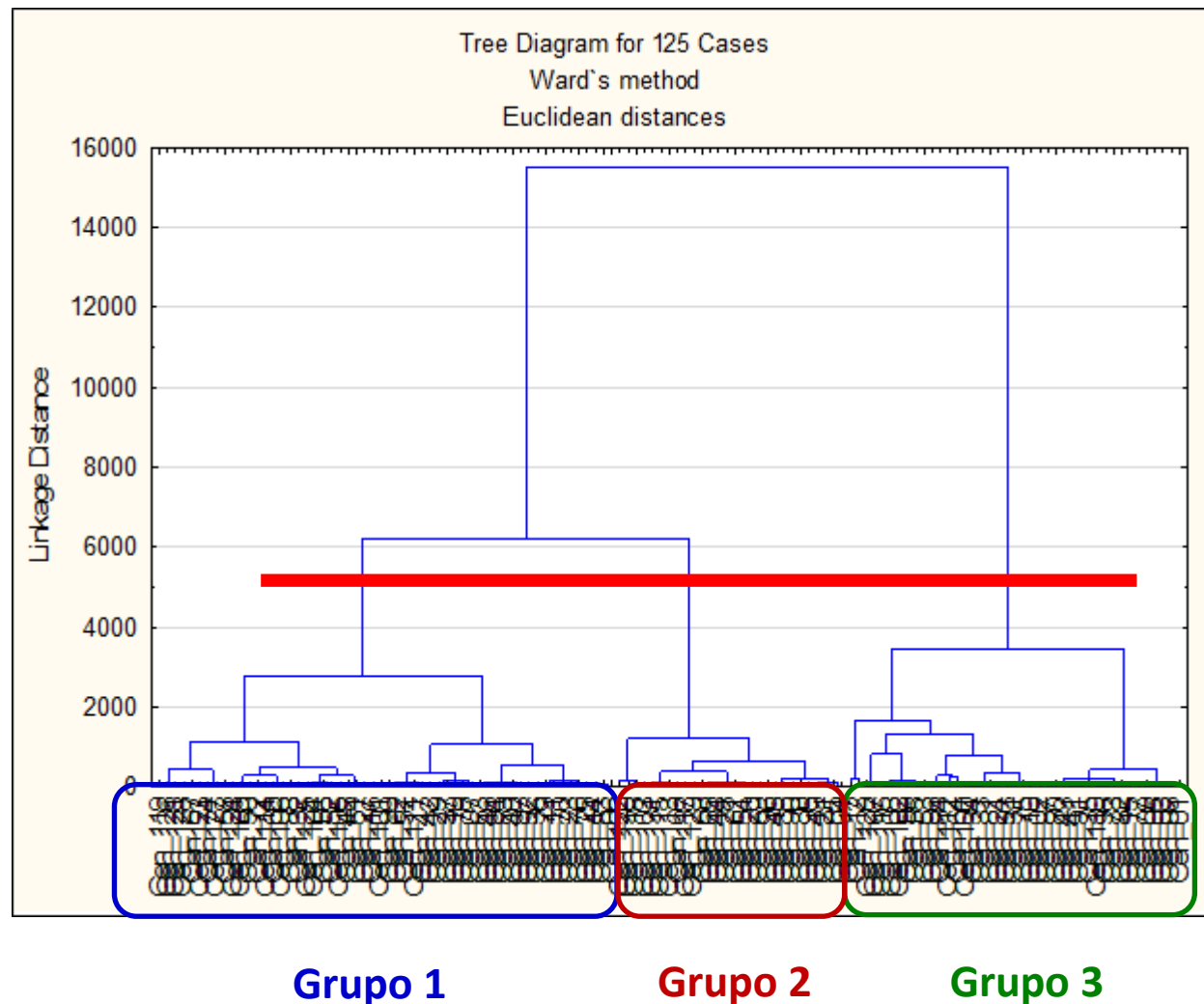
## Método de Ward

# Classificação Hierárquica



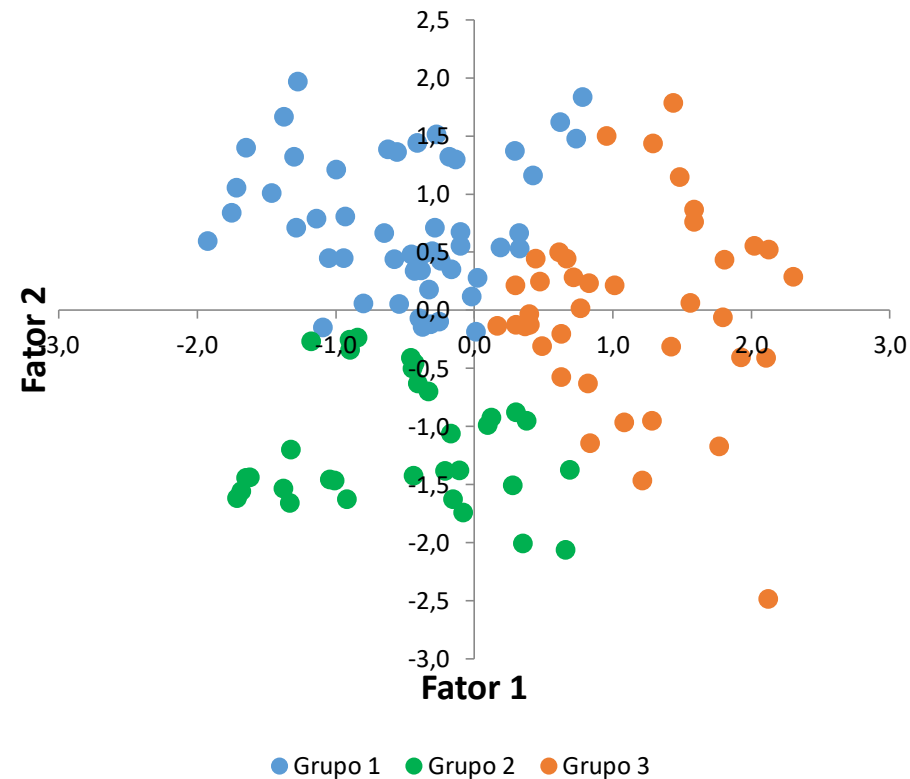
**Método de Ward**

# Classificação Hierárquica



Método de Ward

# Classificação k-médias



# Classificação k-médias

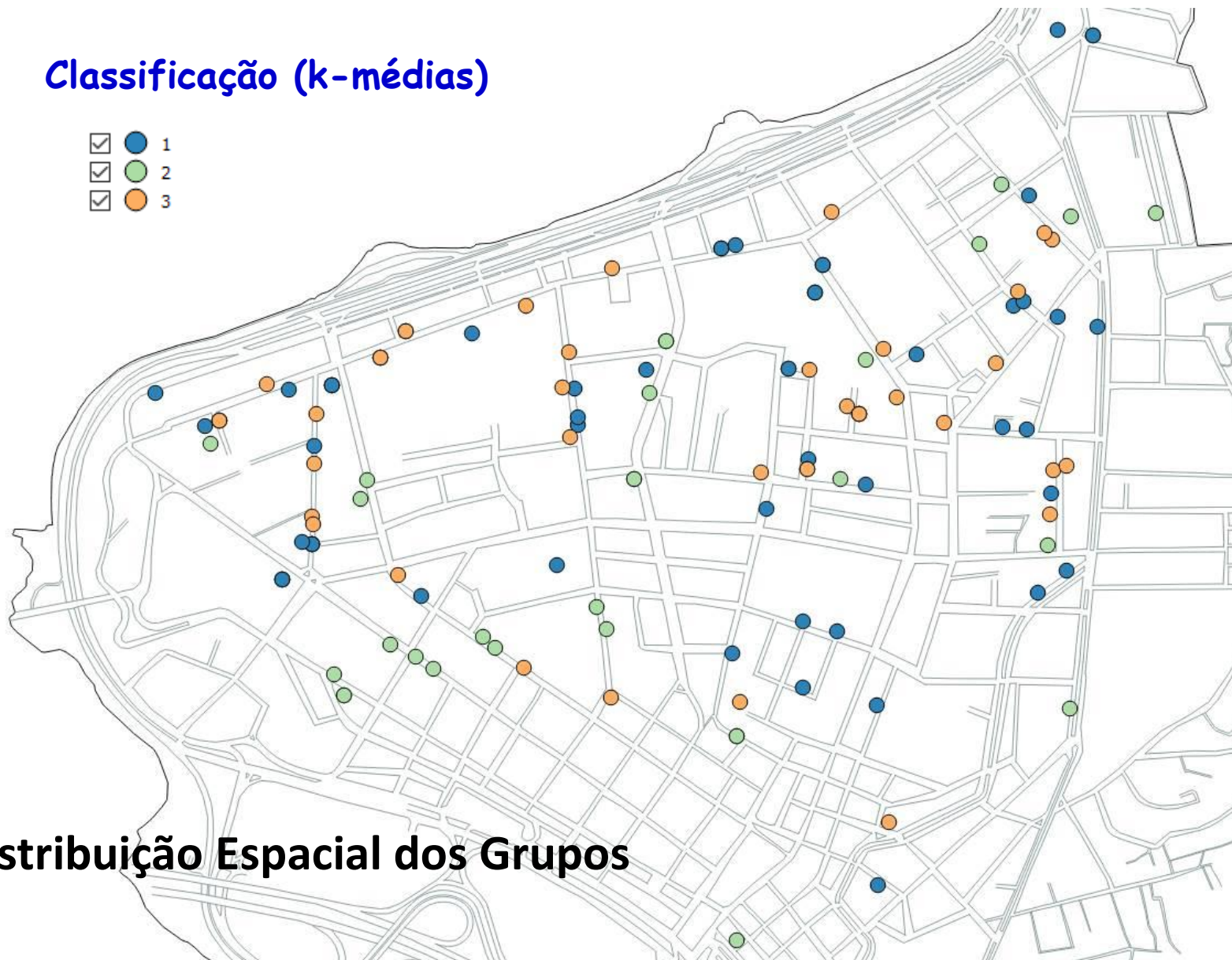
|         | Estatísticas | Area_Total | Aprivat/Atotal | N_Quartos | N_Suites | N_garagens | Idade | Padrão | Conservação | Dist. Beiramar |
|---------|--------------|------------|----------------|-----------|----------|------------|-------|--------|-------------|----------------|
| Grupo 1 | Média        | 132,28     | 0,726          | 2,3       | 1,2      | 1,5        | 11,8  | 2,4    | 2,8         | 474,3          |
|         | Máximo       | 261,00     | 0,896          | 4         | 3        | 3          | 25    | 3      | 3           | 1400           |
|         | Mínimo       | 44,00      | 0,462          | 1         | 0        | 0          | 1     | 2      | 2           | 60             |
|         | Amplitude    | 217,00     | 0,434          | 3         | 3        | 3          | 24    | 1      | 1           | 1340           |
|         | Q1           | 98,00      | 0,677          | 2         | 1        | 1          | 7     | 2      | 3           | 190            |
|         | Mediana      | 135,00     | 0,740          | 2         | 1        | 1          | 11    | 2      | 3           | 450            |
|         | Q2           | 157,00     | 0,796          | 3         | 1        | 2          | 15    | 3      | 3           | 630            |
| Grupo 2 | Média        | 121,45     | 0,765          | 2,4       | 0,82     | 0,9        | 26,5  | 1,3    | 1,5         | 578,0          |
|         | Máximo       | 240,00     | 0,914          | 4         | 2        | 2          | 41    | 2      | 2           | 1600           |
|         | Mínimo       | 35,00      | 0,601          | 1         | 0        | 0          | 9     | 1      | 1           | 120            |
|         | Amplitude    | 205,00     | 0,313          | 3         | 2        | 2          | 32    | 1      | 1           | 1480           |
|         | Q1           | 69,00      | 0,711          | 2         | 1        | 1          | 25    | 1      | 1           | 360            |
|         | Mediana      | 116,00     | 0,772          | 3         | 1        | 1          | 28    | 1      | 2           | 500            |
|         | Q2           | 180,00     | 0,814          | 3         | 1        | 1          | 30    | 2      | 2           | 700            |
| Grupo 3 | Média        | 297,69     | 0,724          | 3,8       | 2,3      | 2,3        | 16,1  | 2,1    | 2,4         | 458,5          |
|         | Máximo       | 551,00     | 0,942          | 4         | 4        | 3          | 30    | 3      | 3           | 1230           |
|         | Mínimo       | 155,00     | 0,498          | 3         | 1        | 1          | 1     | 1      | 1           | 80             |
|         | Amplitude    | 396,00     | 0,444          | 1         | 3        | 2          | 29    | 2      | 2           | 1150           |
|         | Q1           | 208,50     | 0,624          | 4         | 1        | 2          | 11    | 2      | 2           | 215            |
|         | Mediana      | 283,00     | 0,750          | 4         | 2        | 2          | 17    | 2      | 2           | 450            |
|         | Q2           | 373,50     | 0,810          | 4         | 3        | 3          | 20    | 2      | 3           | 580            |

# Classificação k-médias

|         | Estatísticas | VT           | VU_Atot   | VU_Aprivat |
|---------|--------------|--------------|-----------|------------|
| Grupo 1 | Média        | 942.222,45   | 7.056,98  | 9.879,90   |
|         | Máximo       | 2.400.000,00 | 12.972,97 | 17.142,86  |
|         | Mínimo       | 200.000,00   | 3.387,10  | 6.060,61   |
|         | Amplitude    | 2.200.000,00 | 9.585,87  | 11.082,25  |
|         | Q1           | 650.000,00   | 5.825,24  | 7.310,34   |
|         | Mediana      | 795.000,00   | 6.547,62  | 9.008,26   |
|         | Q2           | 1.060.000,00 | 7.932,69  | 11.538,46  |
| Grupo 2 | Média        | 594.666,67   | 4.962,38  | 6.554,97   |
|         | Máximo       | 1.250.000,00 | 8.111,11  | 11.406,25  |
|         | Mínimo       | 170.000,00   | 2.571,43  | 3.214,29   |
|         | Amplitude    | 1.080.000,00 | 5.539,68  | 8.191,96   |
|         | Q1           | 270.000,00   | 4.120,88  | 5.395,68   |
|         | Mediana      | 650.000,00   | 4.629,63  | 6.347,83   |
|         | Q2           | 795.000,00   | 5.685,71  | 7.586,21   |
| Grupo 3 | Média        | 1.806.333,33 | 6.265,90  | 8.769,46   |
|         | Máximo       | 3.800.000,00 | 9.354,84  | 14.338,24  |
|         | Mínimo       | 748.000,00   | 2.287,46  | 3.468,21   |
|         | Amplitude    | 3.052.000,00 | 7.067,38  | 10.870,03  |
|         | Q1           | 1.294.000,00 | 5.171,58  | 7.202,54   |
|         | Mediana      | 1.800.000,00 | 6.462,59  | 8.771,93   |
|         | Q2           | 2.145.000,00 | 7.617,50  | 10.563,44  |

## Classificação (k-médias)

- ☑ 1
- ☑ 2
- ☑ 3



**Distribuição Espacial dos Grupos**

# Aplicações em avaliações de imóveis (continuação)

- ✓ Feita a **análise exploratória do mercado imobiliário** com as informações obtidas, passamos a:
- Desenvolver **modelos** para **cada classe**

# Modelo obtido sem partição dos dados:

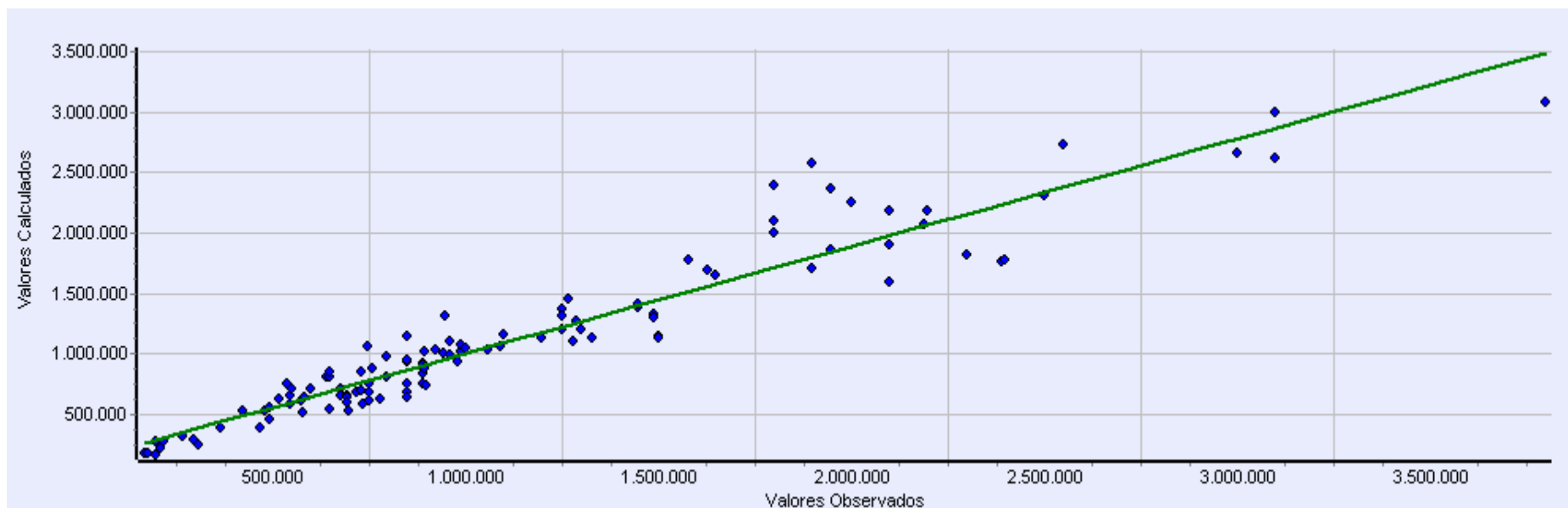
$$\begin{aligned} \ln(VT) = & 12,085157 + \\ & + 0,088254072 * \text{Area\_Total}^{1/2} + \\ & + 0,10376765 * \ln(N\_Quartos) + \\ & + 0,15814249 * N\_Suites^{1/2} + \\ & + 0,41468744 * N\_garagens^{1/2} - \\ & - 0,00017520383 * \text{Idade}^2 + \\ & + 0,1057469 * \text{Padrão} - \\ & - 0,018565 * \text{Dist. Beira-Mar}^{1/2} \end{aligned}$$

## Modelo obtido sem partição dos dados:

- $R^2$  ajustado: **0,940**
- Erro-padrão: 219.042
- Média da amostra: 1.130.831
- CV = **19,4%**
- Significância dos regressores **< 10%**  
(Grau III)
- Análise de Variância **< 0,001%** (Grau III)
- Atendeu aos requisitos de **normalidade** e **homocedasticidade** dos resíduos,  
ausência de **multicolinearidade**

# Modelo obtido sem partição dos dados:

- Bom poder de predição:



# Modelo obtido para o Grupo 1:

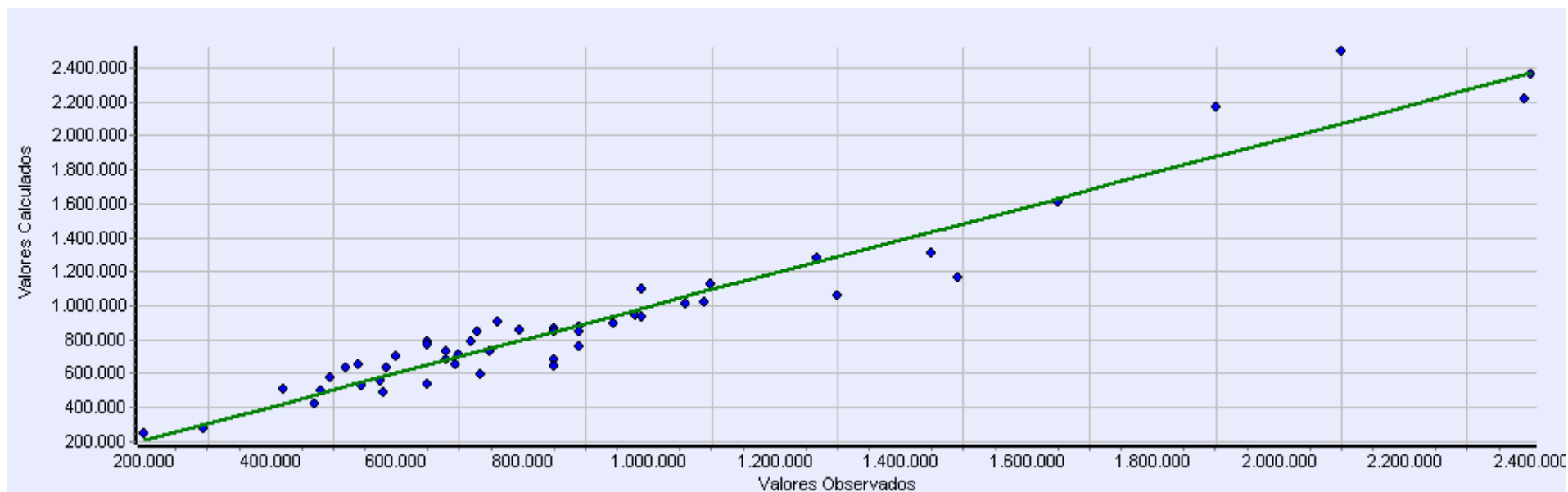
$$\begin{aligned} \ln(VT) = & 12,223217 + \\ & + 1,0388547e-05 * \text{Area\_Total}^2 + \\ & + 0,32672682 * N\_Quartos^{1/2} + \\ & + 0,28723357 * N\_Suites^{1/2} + \\ & + 0,38367535 * N\_garagens^{1/2} + \\ & - 0,11093081 * \ln(Idade) + \\ & + 0,12966241 * \text{Padrão} + \\ & - 2,3755545e-07 * \text{Dist. Beira-Mar}^2 \end{aligned}$$

## Modelo obtido para o Grupo 1:

- $R^2$  ajustado: **0,917**
- Erro-padrão: 133.998
- Média da amostra: 911.649
- CV = **14,7%**
- Significância dos regressores **< 10%**  
(Grau III)
- Análise de Variância **< 0,001%** (Grau III)
- Atendeu aos requisitos de **normalidade** e **homocedasticidade** dos resíduos,  
ausência de **multicolinearidade**

# Modelo obtido para o Grupo 1:

- Bom poder de predição:



# Modelos obtidos para os Grupos 2 e 3:

Não ficaram melhores!

Possíveis causas:

- 1) Os **eixos fatoriais 1 e 2 não foram suficientes**
  - Testar soluções usando também os eixos 3 e 4

| Value | Eigenvalues                      |         |            |            |
|-------|----------------------------------|---------|------------|------------|
|       | Extraction: Principal components |         |            |            |
|       | Eigenvalue                       | % Total | Cumulative | Cumulative |
| 1     | 3,616                            | 40,175  | 3,616      | 40,175     |
| 2     | 2,156                            | 23,955  | 5,772      | 64,129     |
| 3     | 0,962                            | 10,685  | 6,733      | 74,815     |
| 4     | 0,749                            | 8,325   | 7,483      | 83,140     |

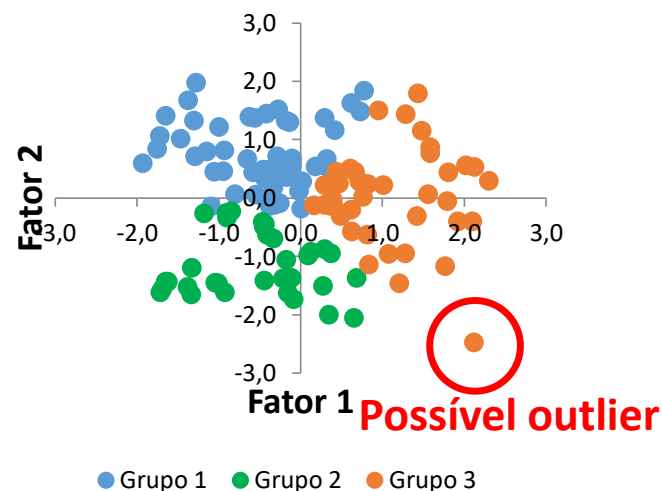
# Modelos obtidos para os Grupos 2 e 3:

Não ficaram melhores!

Possíveis causas:

2) O método de **agrupamento não foi adequado**

- Inércia inter-classes foi pequena
- Testar outros métodos



# Outro exemplo de aplicação em Avaliação de Imóveis

PERUZZO TRIVELLONI (1998):

- ✓ Conseguiu **melhoras** nos modelos para os agrupamentos em relação ao modelo original
- Em seu trabalho se observa que:
  - Modelo inicial **não** estava **bem ajustado**, apresentava muitos **outliers** e **heterocedasticidade**
  - Havia um número bem maior de variáveis
    - **Dimensionalidade foi reduzida**
  - No método aglomerativo, obteve inércia intra-classes pequena e inércia inter-classes grande
    - **Clusterização foi eficiente**

# Considerações finais

- O exemplo desenvolvido neste trabalho não está esgotado em todas suas possibilidades
  - **trabalhos em andamento...**
- O procedimento apresentado aqui nem sempre trará melhores resultados do que a regressão feita pelos procedimentos clássicos
  - **Se o modelo já está bom, fica difícil melhorá-lo...**

# Bibliografia

AMO, Sandra de. **Curso de Data Mining.**

<http://www.deamo.prof.ufu.br/arquivos/Aula13.pdf>

BRANCO, João A. **Introdução à análise de clusters.**

<https://fenix.tecnico.ulisboa.pt/downloadFile/3779579704252/SlidesACluster.pdf>

DONI, Marcelo V. **Análise de cluster: métodos hierárquicos e de particionamento.** Mackenzie, São Paulo, 2004.

NCSS, LLC. **NCSS Statistical Software.**

<https://www.ncss.com/software/ncss/ncss-documentation/>

PERUZZO TRIVELLONI, C. A. **Metodologia para avaliação em massa de apartamentos por inferência estatística e técnicas de análise multivariada – Uma análise exploratória.** Dissertação, Programa de Pós-Grauação em Engenharia Civil, Universidade Federal de Santa Catarina, 1998.

VICINI, Lorena. **Análise multivariada: da teoria à prática.** Santa Maria : UFSM, CCNE, 2005.



**XIX COBREAP | Foz do Iguaçu**

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

AGOSTO 2017

**Prof. Dr. Norberto Hochheim**

**UFSC**

**IBAPE-SC**

# **Análise de Cluster aplicada à avaliação de bens**

## **Obrigado!**