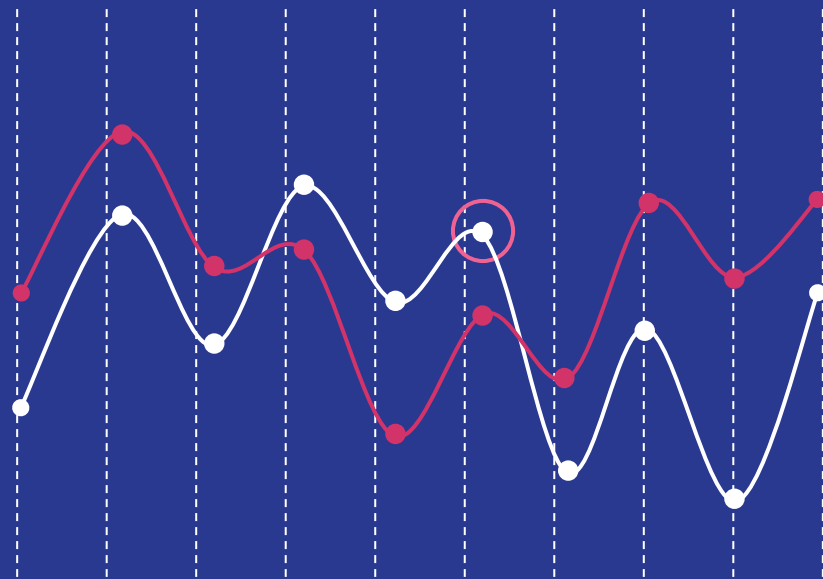


# Clusterização de Séries Temporais

Luan Menezes  
Marcus Freire

# Agenda

- Problema
- Apresentação da Base
- Preparação dos Dados
- Agrupamento
- Validação
- Interpretação



# O problema

- Padrão de uso do laboratório 148 do IME
- Coleta de dados



# Apresentação da base

- Base gerada no projeto da disciplina MATE84.

TIMESTAMP_	TEMPERATURA	SOM	LUZ	UMIDADE	QTD_PESSOAS
2018-06-08 10:21:03	26	497	950	79	4
2018-06-08 10:21:33	26	497	1000	79	4

Temperatura  $t$ :  $0 \leq t \leq 100$

Umidade  $u$ :  $0 \leq u \leq 100$

Som  $s$ :  $0 \leq s \leq 1024$

Luminosidade  $l$ :  $0 \leq l \leq 1024$

Quantidade de pessoas  $p$ :  $p \geq 0$

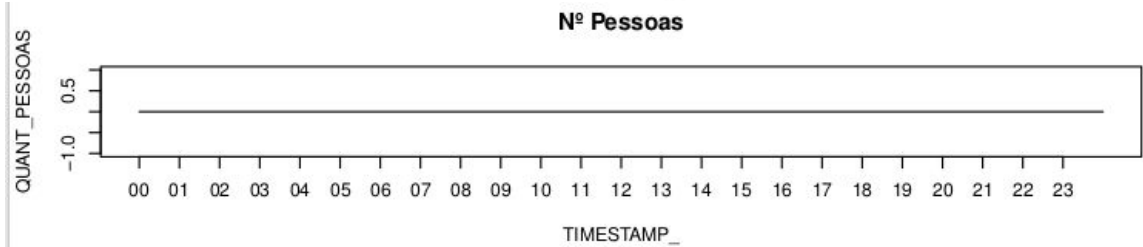
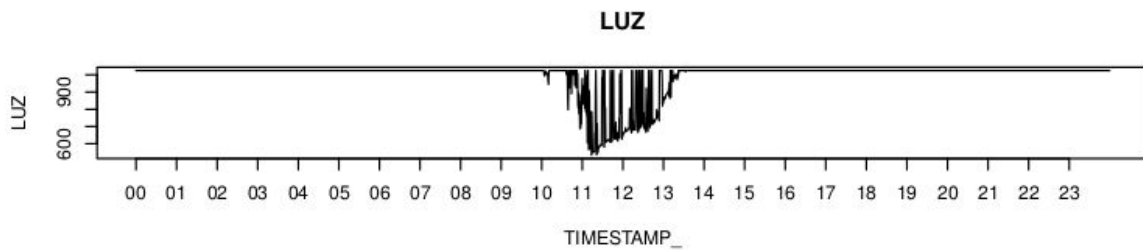
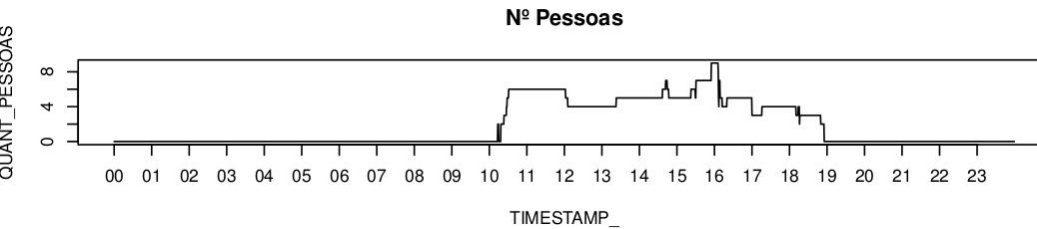
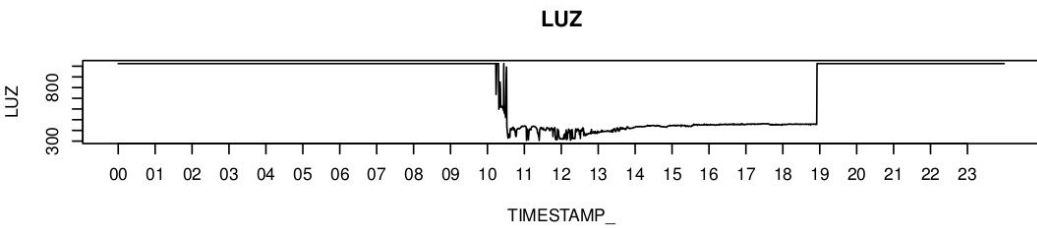
# Apresentação da base

Quantidade total de instâncias: 77230

Dias coletados: 8/06/2018 à 4/07/2018, totalizando 27 dias

Média de instâncias por dia 2860





# Preparação do Dados



# Preparação dos Dados

- Instâncias ausentes:
  - Minuto sem nenhuma instância;
  - Inserção de uma instância por minuto faltante.
- Valores errados:
  - Valores fora do intervalo definido dos sensores;
  - Quantidade de pessoas:
    - Entre 17h e 06h se a luz estiver apagada e o sensor de quantidade de pessoas indicar a presença de alunos no laboratório, o valor do sensor será corrigido;
    - Quantidade de pessoas acima de 35.
- Outliers:
  - Temperatura: abaixo de 20°C ou acima de 35°C;
- Método para inserção e substituição de dados:
  - Média das instâncias anteriores.





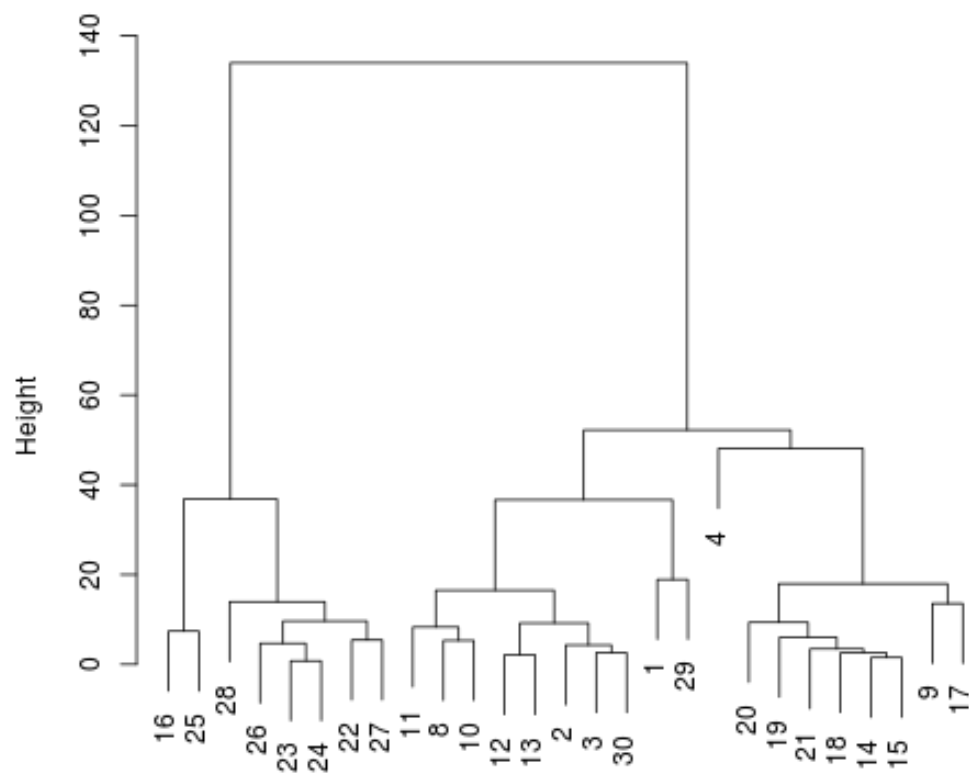
# Agrupamento

# Agrupamento

- Descobrir padrões na sequência de séries temporais.
  - Verificar as séries mais semelhantes
    - Utilizar medidas de similaridades DTW e MDDL.
- Algoritmo de agrupamento hierárquico, gera a partir de uma matriz de proximidades uma sequência de partições aninhadas.

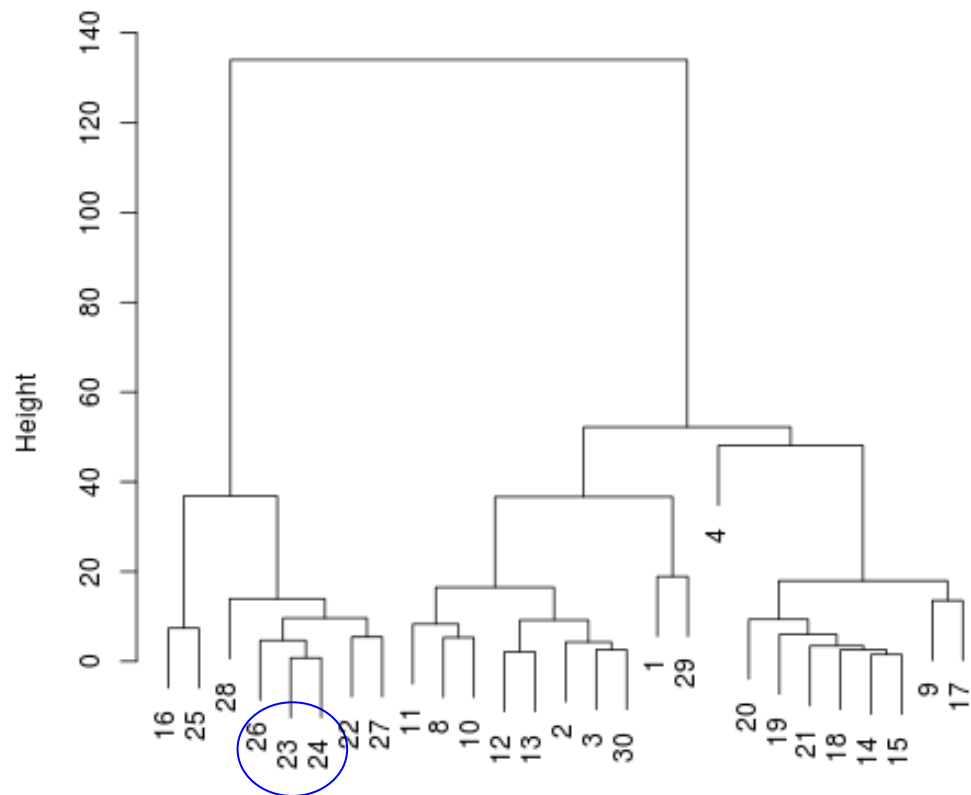


Cluster Dendrogram



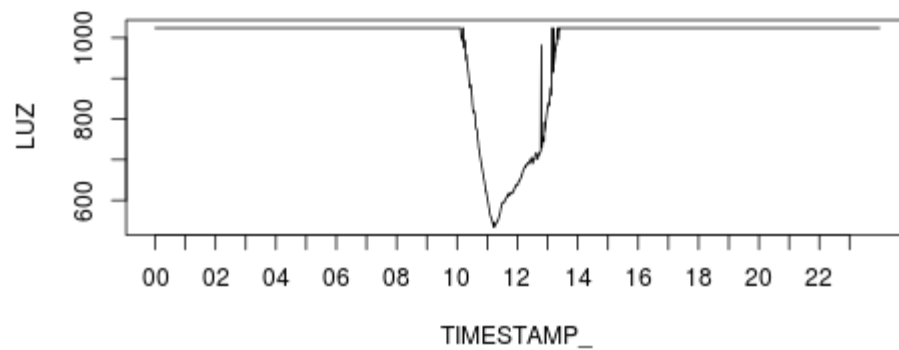
as.dist(luminosidade)  
hclust (\*, "complete")

Cluster Dendrogram

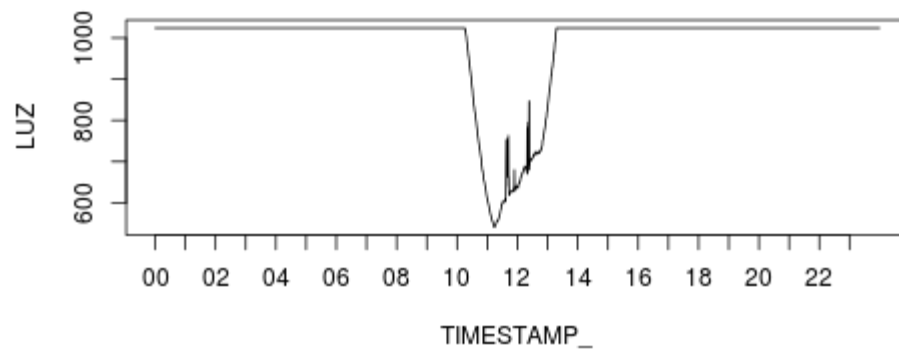


as.dist(luminosidade)  
hclust (\*, "complete")

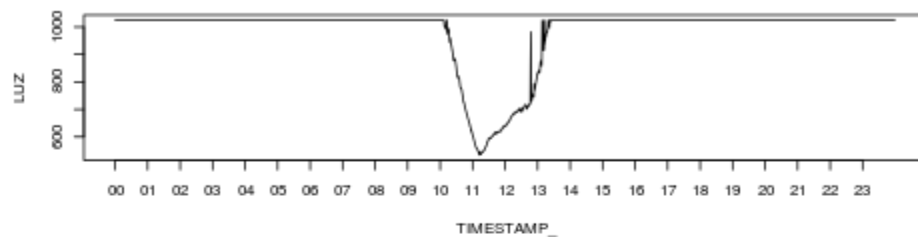
**Data: 23/06**



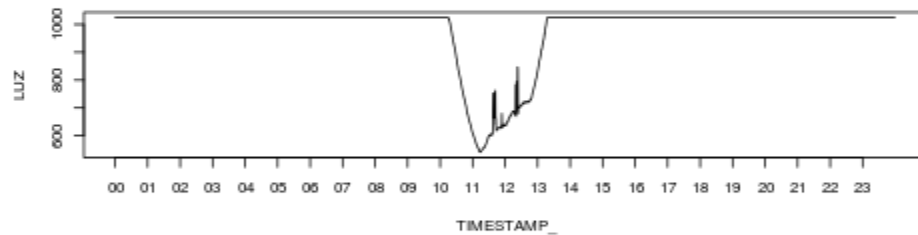
**Data: 24/06**



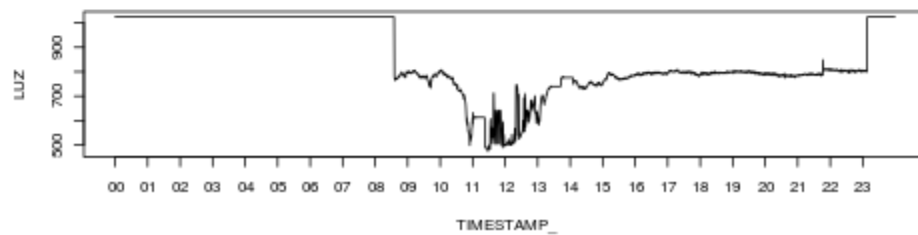
Data: 23/06



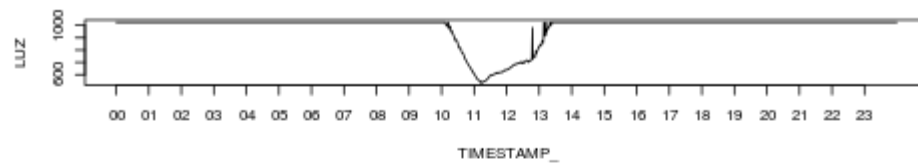
Data: 24/06



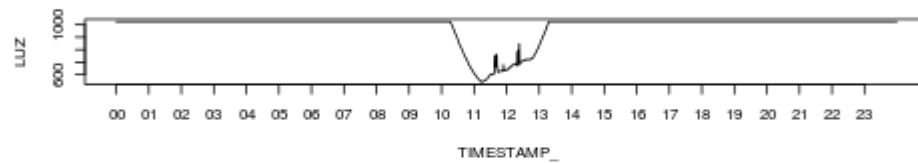
Data: 26/06



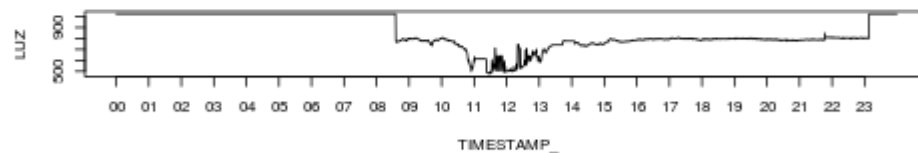
Data: 23/06



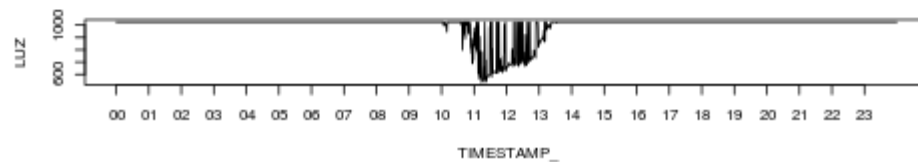
Data: 24/06



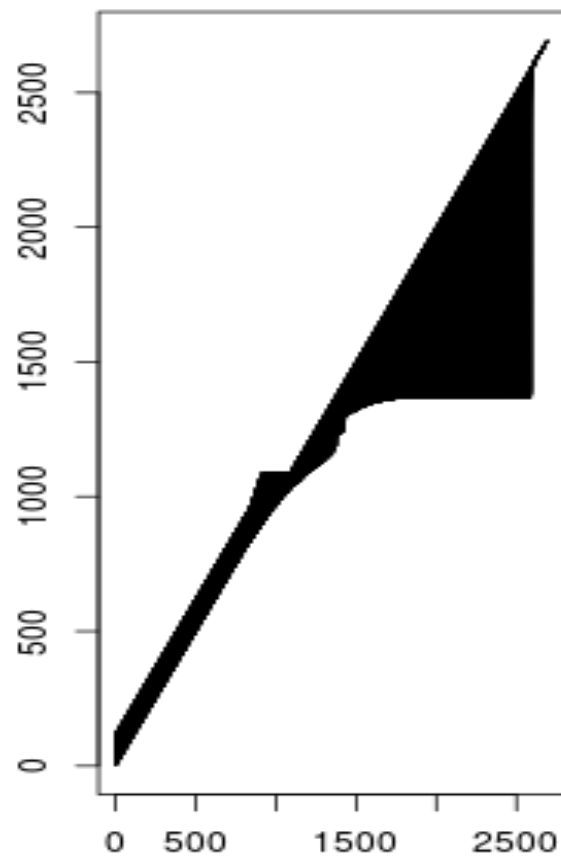
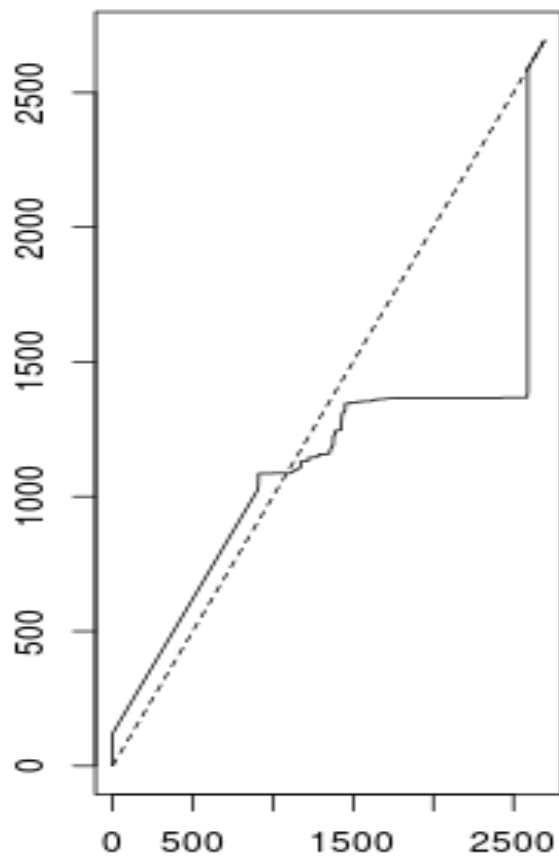
Data: 26/06



Data: 22/06

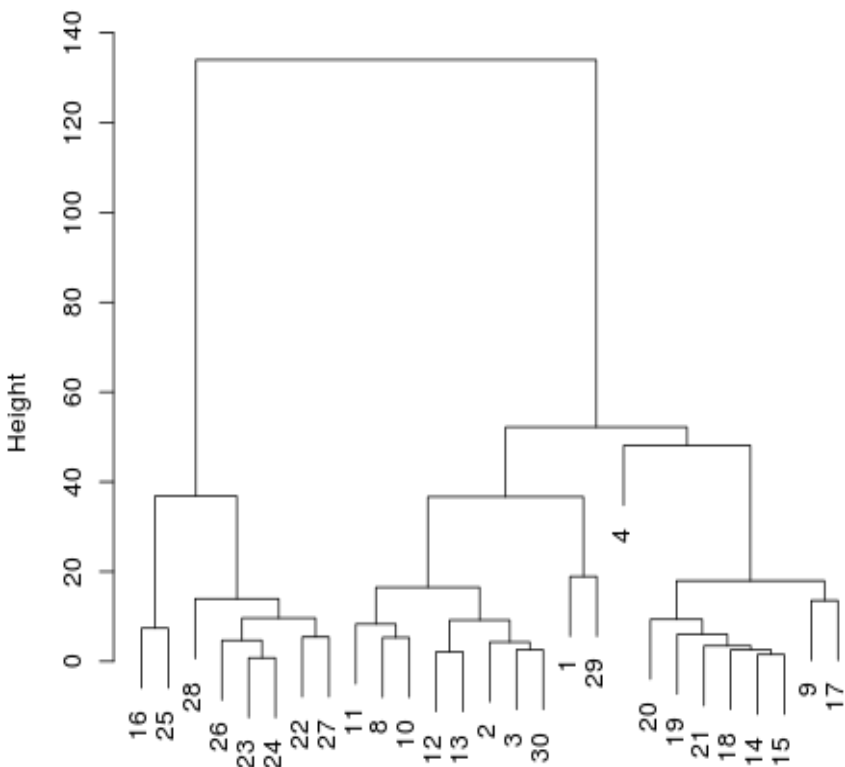


# DTW e MDDL



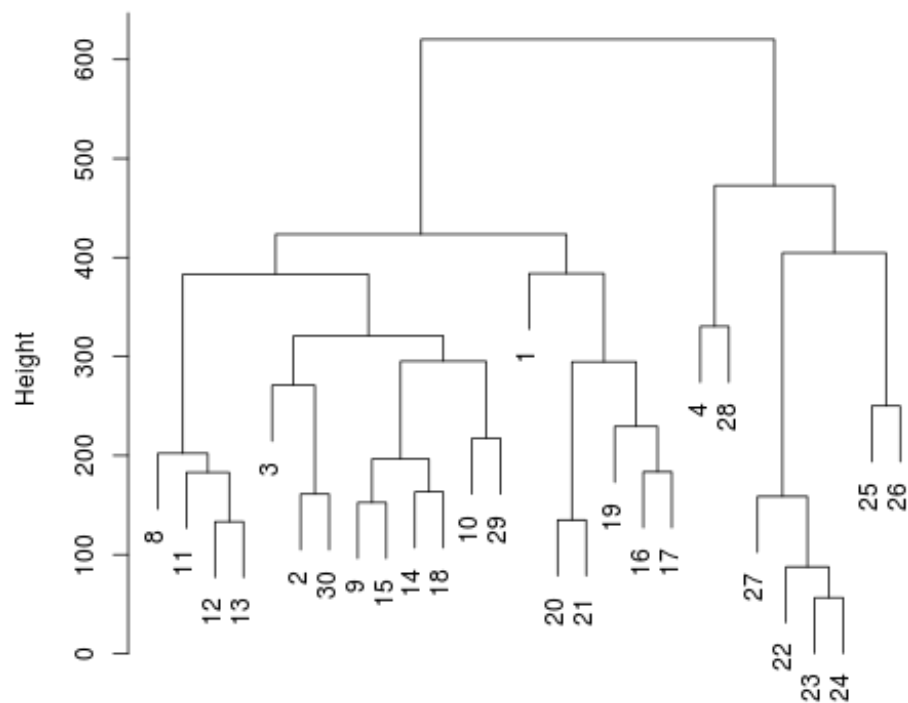


Cluster Dendrogram



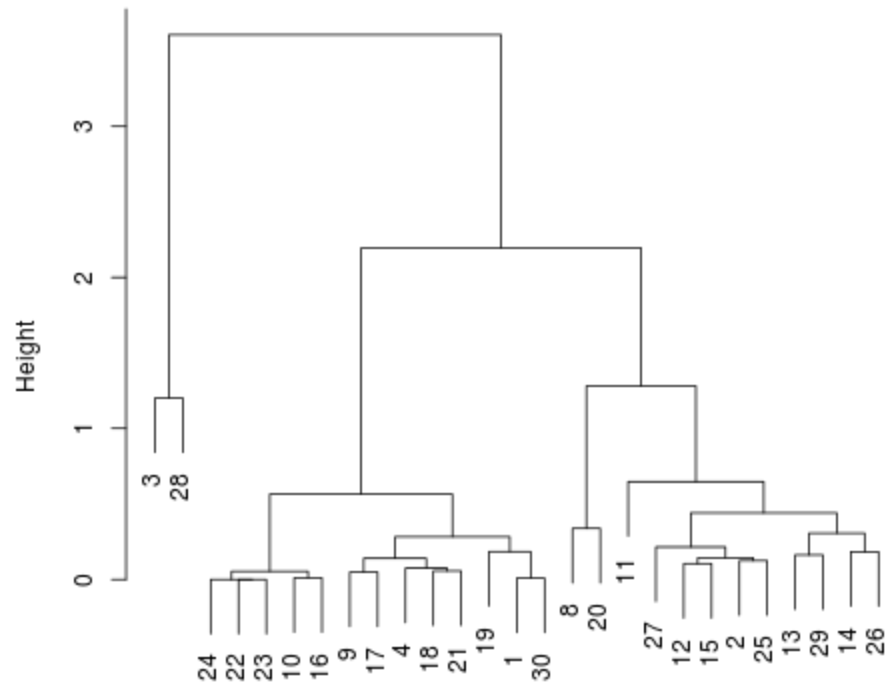
as.dist(luminosidade)  
hclust(\*, "complete")

Cluster Dendrogram



luminosidadeMDDL  
hclust(\*, "complete")

Cluster Dendrogram

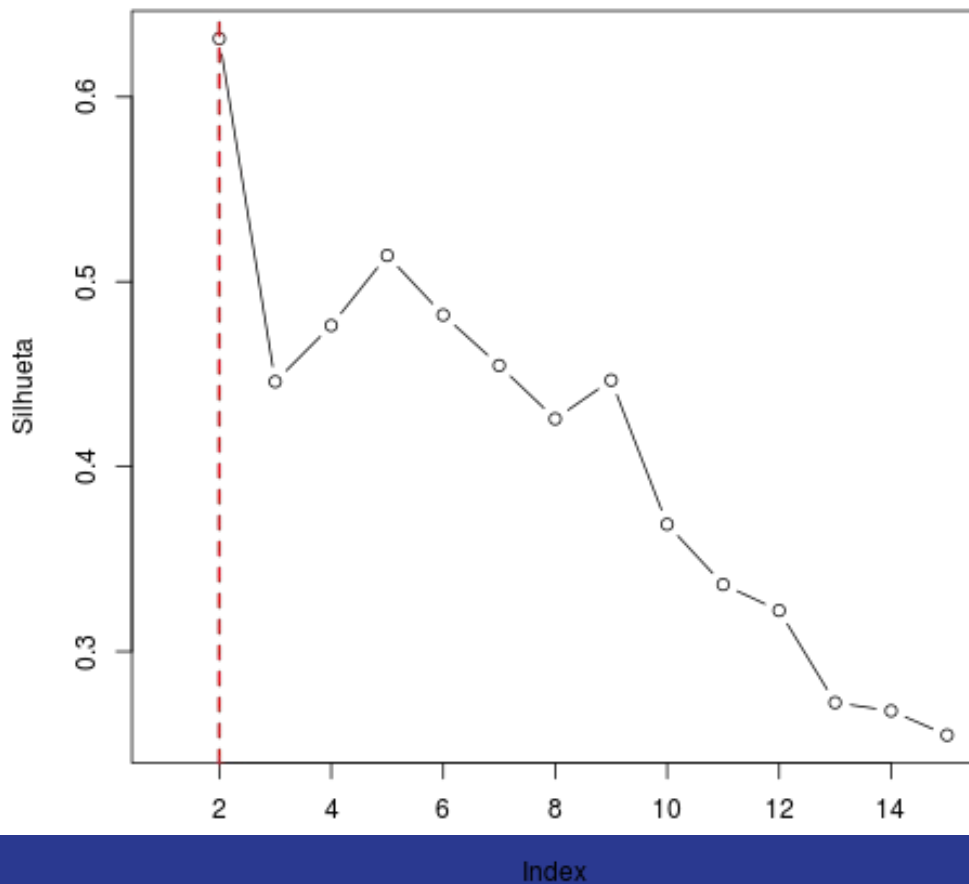


```
as.dist(qtdPessoas)  
hclust ("*", "complete")
```



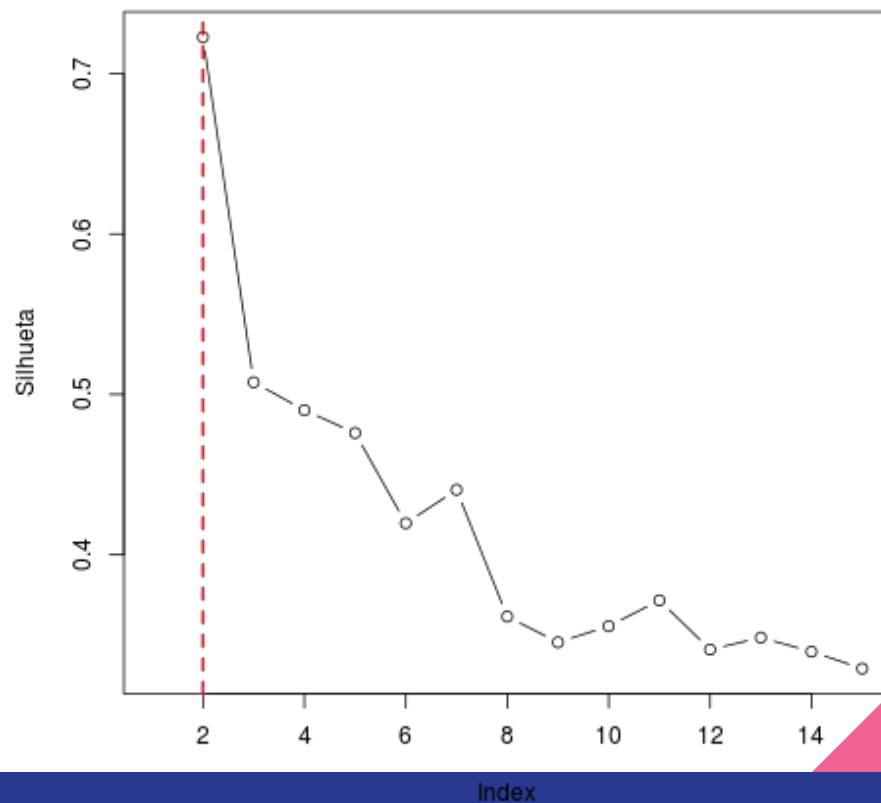
# Validação

# Validação - Luminosidade

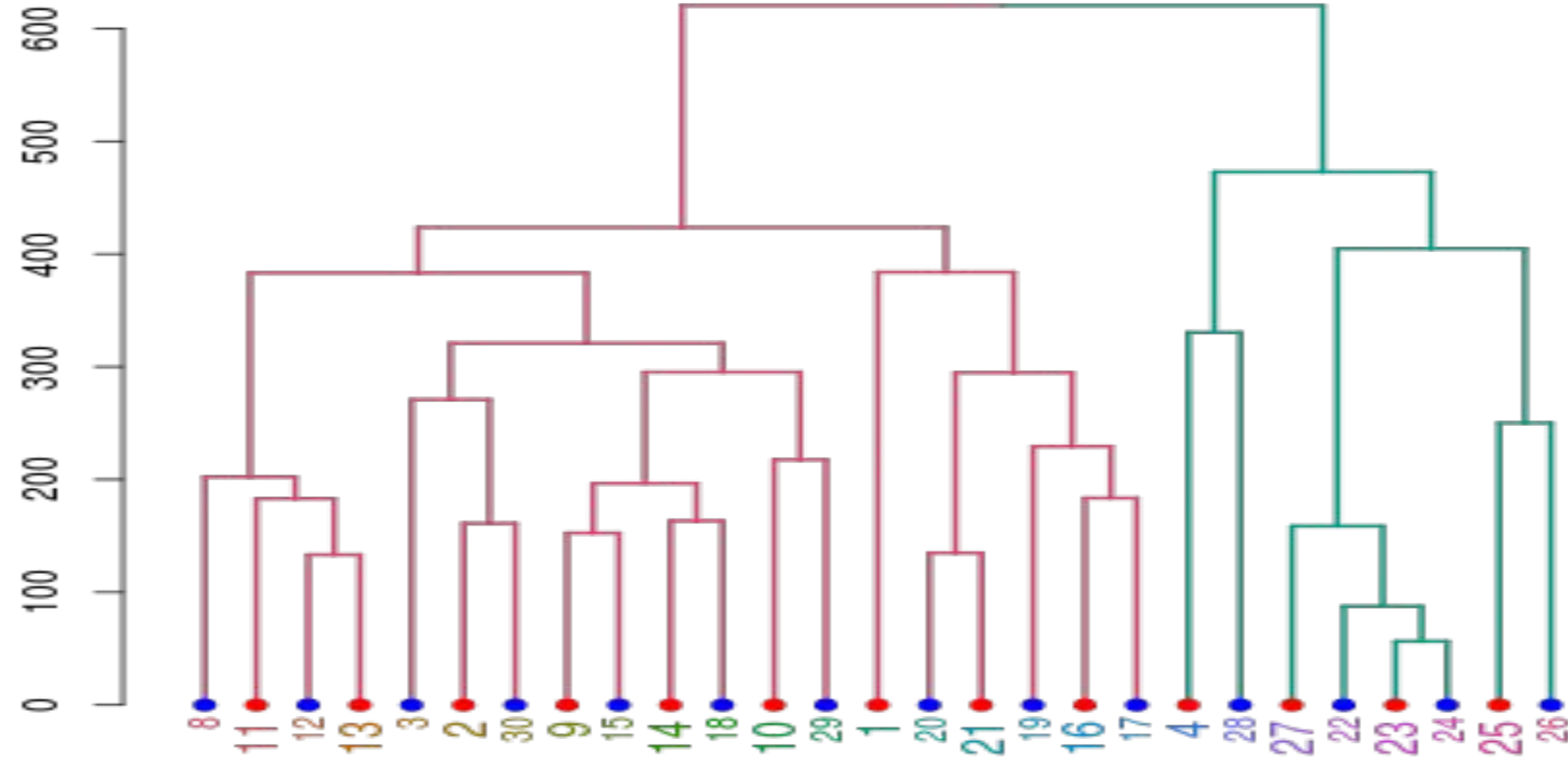


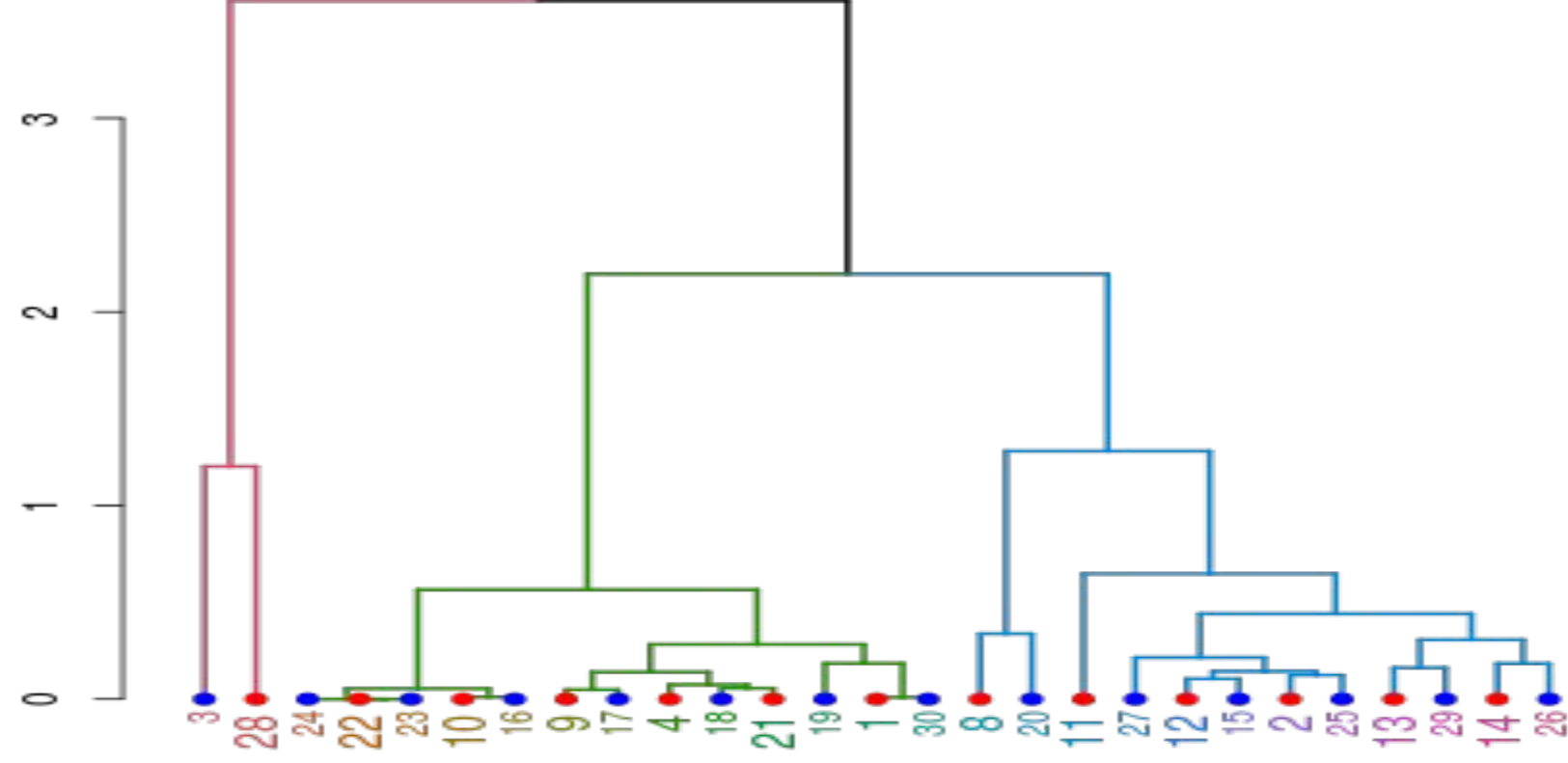
- Medir a similaridade(ou relação) entre objetos de um determinado grupo
  - Silhueta: Avaliar relativamente a qualidade das partições.

# Validação - Qtd\_Pessoas



# Interpretação









# Obrigado!

Luan Menezes  
Marcus Freire