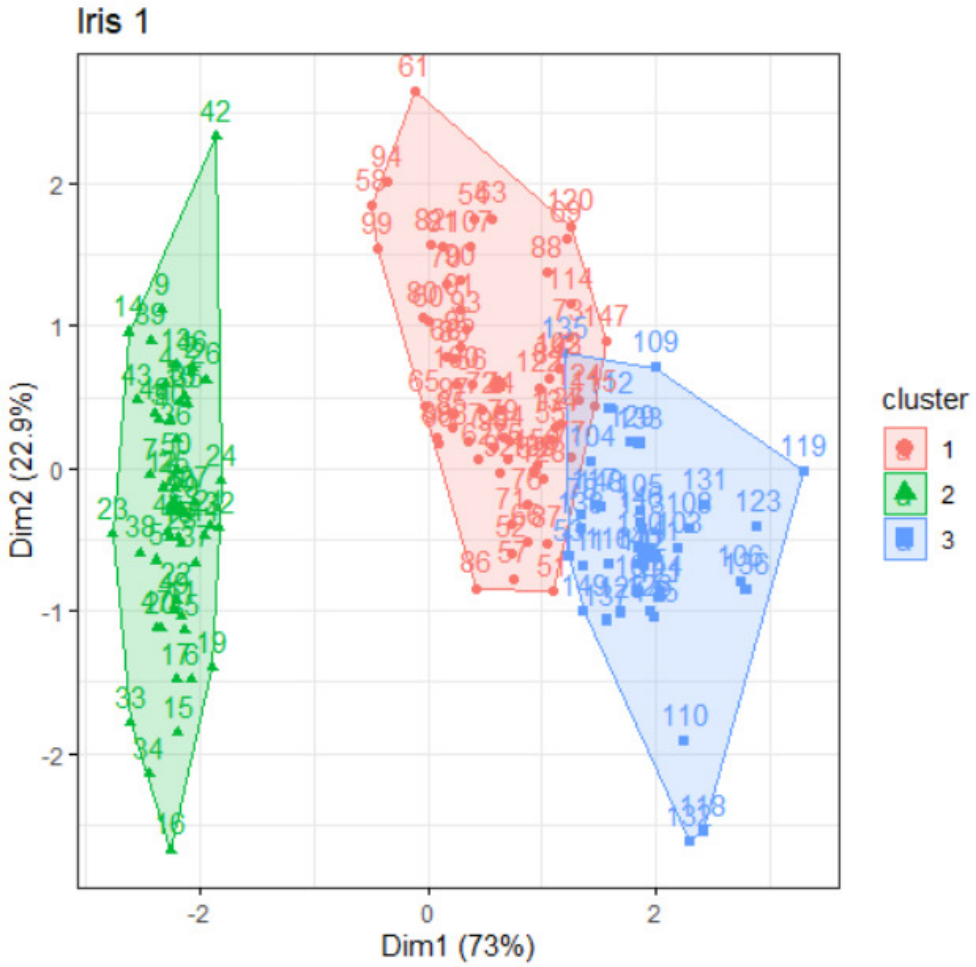
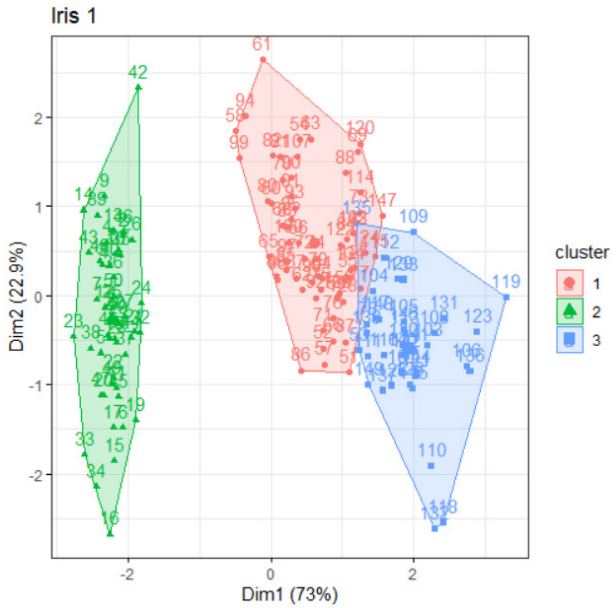
Agrupamentos

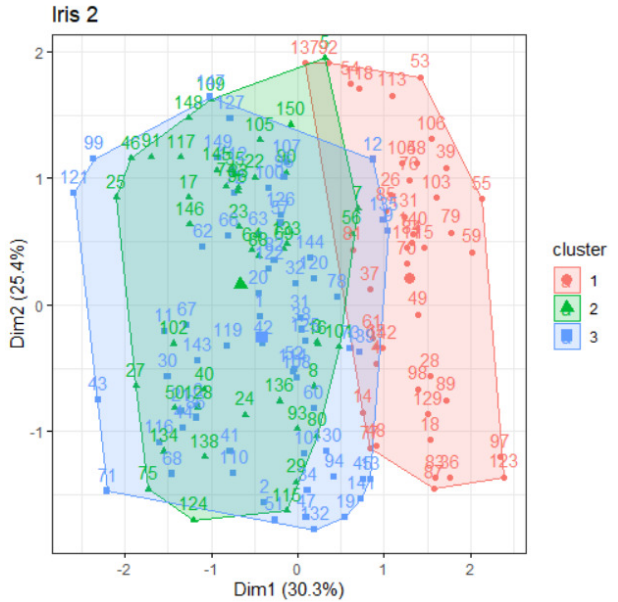


Perguntas:

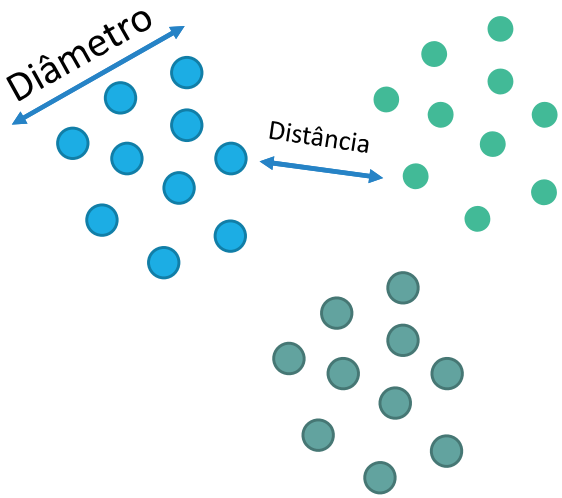
1. De fato existem Clusters nos dados?
    1. Sabemos que K-means irá agrupar sempre os dados
2. O número ideal de clusters é de fato 3?
    1. Usamos 3 porque conhecemos os labels!
3. Foram produzidos bons clusters?
4. Usamos o melhor agrupador?

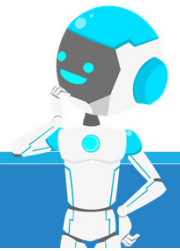
Produzindo um cluster



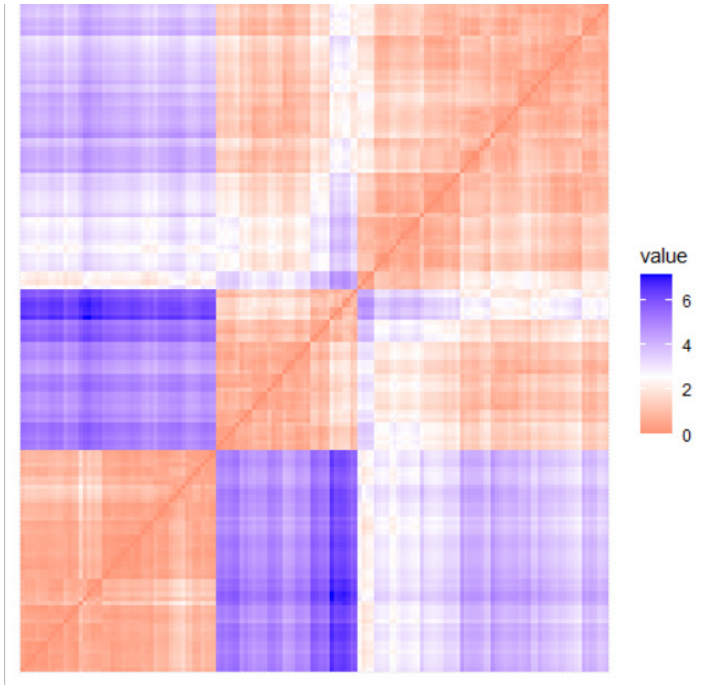


1.De fato existem Clusters nos dados?

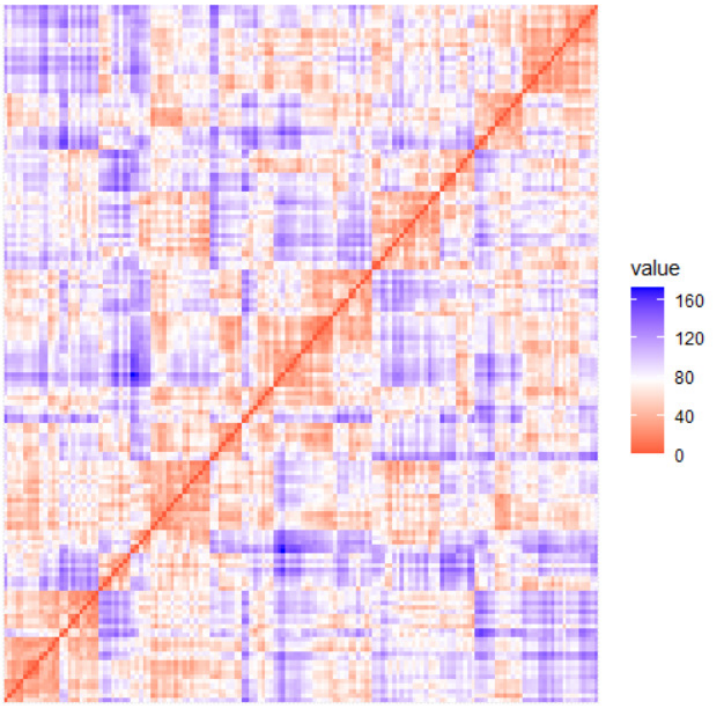




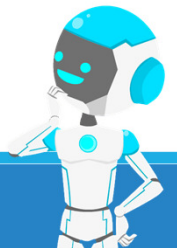
1.De fato existem Clusters nos dados?



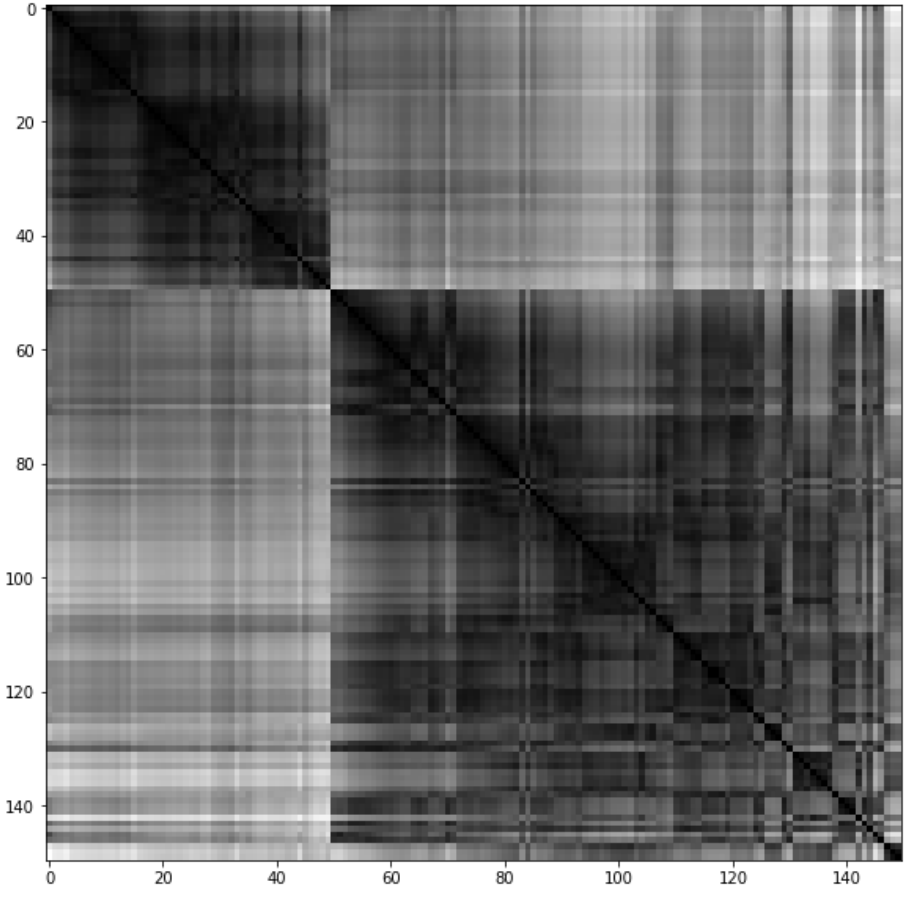
H 0.4814079

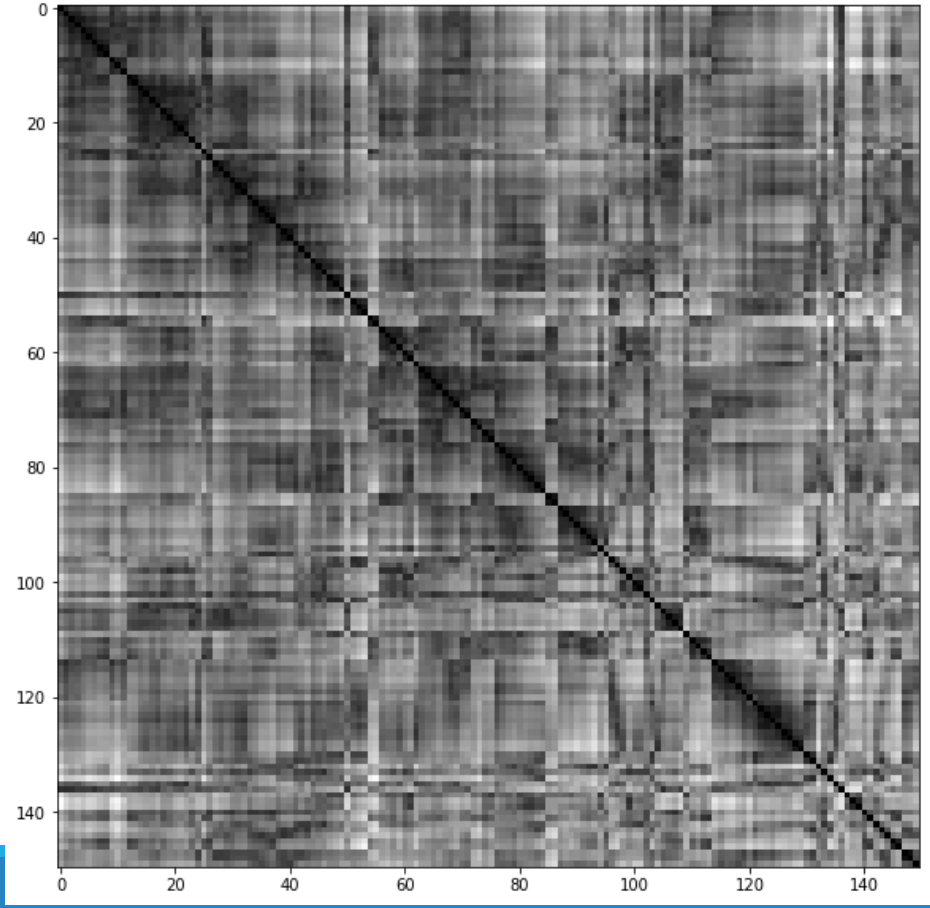


H 0.8191482

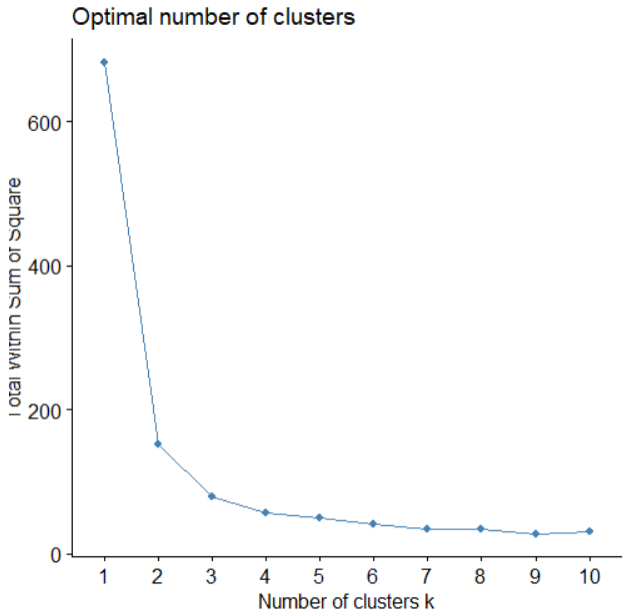


1.De fato existem Clusters nos dados?

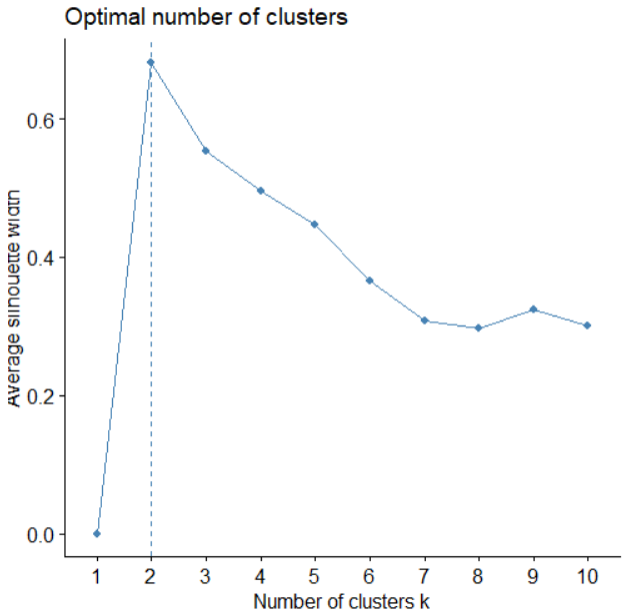




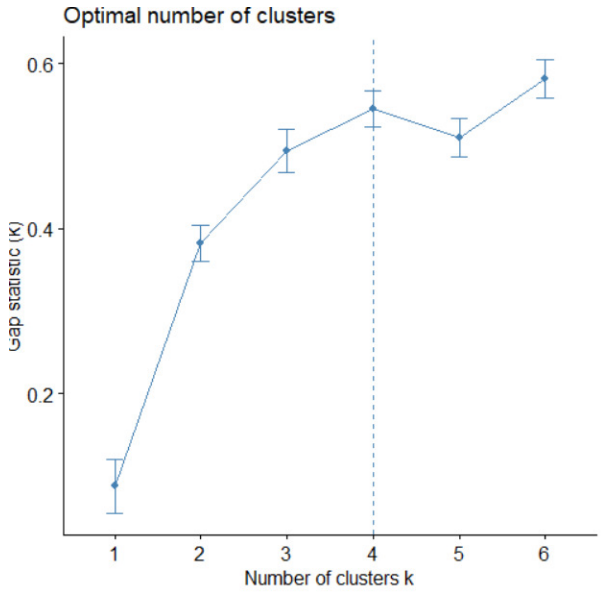
2. Qual o número ideal de clusters?



Elbow

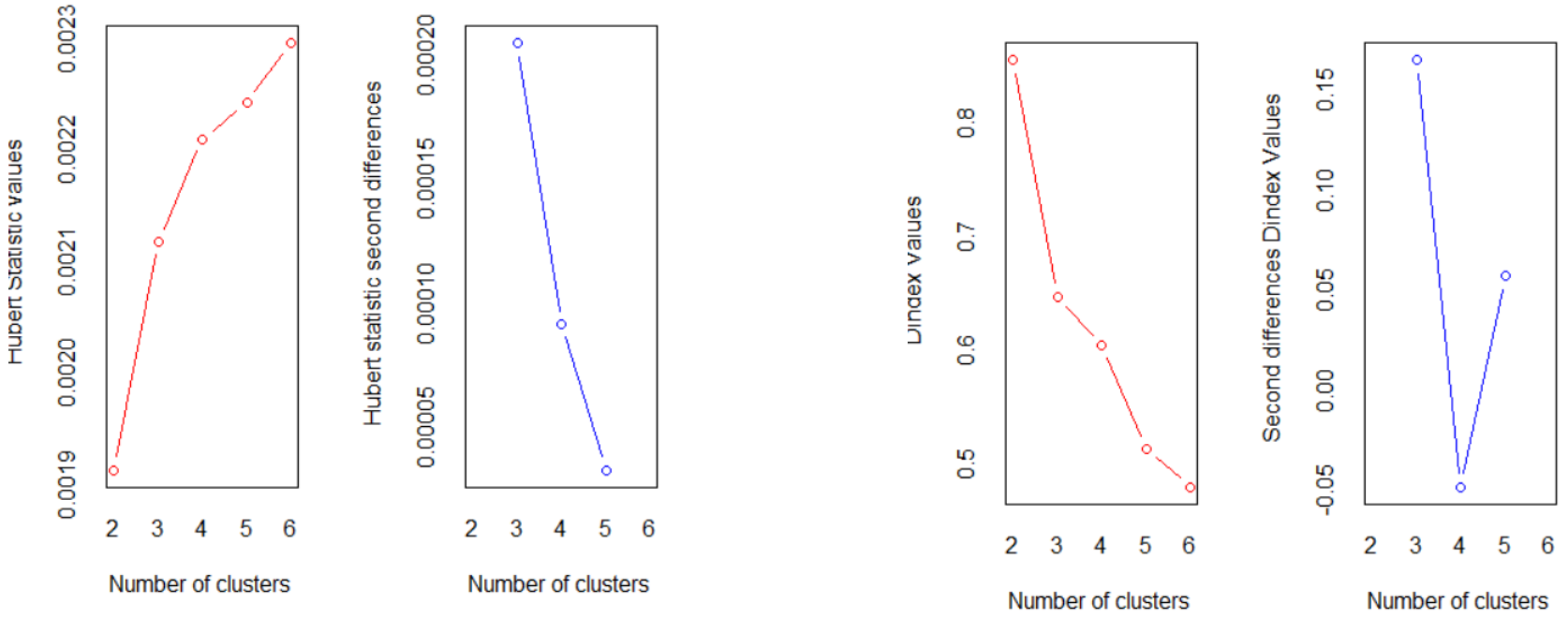


Average silhouette



Gap

Hubert index e D index



3. Foram produzidos bons clusters?

Em um bom cluster:

- Diâmetro do cluster pequeno
- Distância entre os clusters deve ser grande

Índice "Dunn" mede a qualidade do cluster

Busca-se maximizar este índice