

Exercícios sobre Clustering com R - EP3

1 Questões sobre clustering

1. Utilizando o dataset **ruspini** do pacote **cluster**, execute uma análise de cluster utilizando o algoritmo **k-means**.
2. Faça uma análise de cluster utilizando o dataset sobre **abalos sísmicos** do site ¹. Considere apenas as variáveis de **profundidade** e **magnitude**.
3. Imprima a informação dos clusters em um mapa georeferenciado. Para imprimir pontos em um mapa georeferenciado você pode utilizar o código abaixo:

```
library(maps)
library(mapdata)
map(mar = c(0.1, 0.1, 0.1, 0.1), myborder=0.00001)
points(abalos$longitude, abalos$latitude, col=2, pch=20)
```

Você identificou algum padrão que você considere relevante?

4. Faça uma análise de cluster utilizando o dataset **survey**, atributos **Age** e **Height**, do pacote **UsingR**. Implemente três cenários diferentes:
 - (a) Com os valores originais;
 - (b) Com os valores da altura (Height) em metros, e;
 - (c) Com os valores da altura (Height) e idade (Age) devidamente normalizados.Comente o que acontece em cada um dos casos.
5. Utilizando o dataset **survey**, atributos **Exer** (sobre o hábito de fazer exercícios) e **Smoke** (sobre o hábito de fumar), faça uma análise de cluster deste dataset.
6. Levando-se em consideração o dataset do item anterior, inclua o atributo de sexo (**Sex**) e faça a mesma análise.

2 Códigos que podem ser úteis

```
elbow <- function(dataset){
  wss <- numeric(15)
  for (i in 1:15)
    wss[i] <- sum(kmeans(dataset,centers=i,
                        nstart=100)$withinss)
  plot(1:15, wss, type="b", main="Elbow method",
       xlab="Number of Clusters",
       ylab="Within groups sum of squares",
       pch=8, col="red")
}
```

¹http://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_month.csv

3 Método de entrega

Este trabalho deverá ser realizado em dupla. A data máxima para entrega é no dia 26 de maio de 2015. A entrega deve ser realizada em sala de aula.