06-Agregar datos cuantitativos por factor

Adrian

17/1/2022

Funcion summary()

Aplicada a un vector numerico o a una variable cuantitativa nos devuelve un resumen estadistico con los valores minimo y maximo del vector, sus tres cuartiles y su media.

```
cangrejos = read.table("../../../data/datacrab.txt", header = T)
# Eliminar la primera fila
cangrejos = cangrejos[-1]
summary(cangrejos)
```

```
##      color           spine           width          satell           weight
##  Min.   :2.000   Min.   :1.000   Min.   :21.0   Min.   : 0.000   Min.   :1200
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:24.9   1st Qu.: 0.000   1st Qu.:2000
##  Median :3.000   Median :3.000   Median :26.1   Median : 2.000   Median :2350
##  Mean   :3.439   Mean   :2.486   Mean   :26.3   Mean   : 2.919   Mean   :2437
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:27.7   3rd Qu.: 5.000   3rd Qu.:2850
##  Max.   :5.000   Max.   :3.000   Max.   :33.5   Max.   :15.000   Max.   :5200
```

```
# Comparar numericamente los pesos y las anchuras de los cangrejos con 3 colores con los que tienen 5 c
summary(subset(cangrejos, color == 3, c("weight", "width")))
```

```
##      weight         width
##  Min.   :1300   Min.   :22.5
##  1st Qu.:2100   1st Qu.:25.1
##  Median :2500   Median :26.5
##  Mean   :2538   Mean   :26.7
##  3rd Qu.:3000   3rd Qu.:28.2
##  Max.   :5200   Max.   :33.5
```

```
summary(subset(cangrejos, color == 5, c("weight", "width")))
```

```
##      weight         width
##  Min.   :1300   Min.   :21.00
##  1st Qu.:1900   1st Qu.:23.90
##  Median :2125   Median :25.50
##  Mean   :2174   Mean   :25.28
##  3rd Qu.:2400   3rd Qu.:26.57
##  Max.   :3225   Max.   :29.30
```

Funcion by()

Aplicar una determinada funcion a algunas columnas segmentandolas segun los niveles de un factor

Sintaxis: by(columnas, factor, FUN = funcion)

```
# Comparar segun diferentes especies ciertas columnas
by(iris[,c(1,3)], iris$Species, FUN = summary)
```

```
## iris$Species: setosa
##   Sepal.Length    Petal.Length
##  Min.   :4.300   Min.   :1.000
##  1st Qu.:4.800   1st Qu.:1.400
##  Median :5.000   Median :1.500
##  Mean   :5.006   Mean   :1.462
##  3rd Qu.:5.200   3rd Qu.:1.575
##  Max.   :5.800   Max.   :1.900
## ------------------------------------------------------------
## iris$Species: versicolor
##   Sepal.Length    Petal.Length
##  Min.   :4.900   Min.   :3.00
##  1st Qu.:5.600   1st Qu.:4.00
##  Median :5.900   Median :4.35
##  Mean   :5.936   Mean   :4.26
##  3rd Qu.:6.300   3rd Qu.:4.60
##  Max.   :7.000   Max.   :5.10
## ------------------------------------------------------------
## iris$Species: virginica
##   Sepal.Length    Petal.Length
##  Min.   :4.900   Min.   :4.500
##  1st Qu.:6.225   1st Qu.:5.100
##  Median :6.500   Median :5.550
##  Mean   :6.588   Mean   :5.552
##  3rd Qu.:6.900   3rd Qu.:5.875
##  Max.   :7.900   Max.   :6.900
```

Funcion aggregate()

igual que by pero con un resultado diferente

```
#aggregate(cbind(Sepal.length, Petal.Length)~Species,                    data=iris,FUN = summary
```

NA

Las funciones anteriores no funcionan si hay valores NA.

Hay que aplicar el parametro na.rm = True.