Reglas de asociación

Minería de datos: aprendizaje no supervisado y detección de anomalías

Francisco Luque Sánchez

21/12/2019

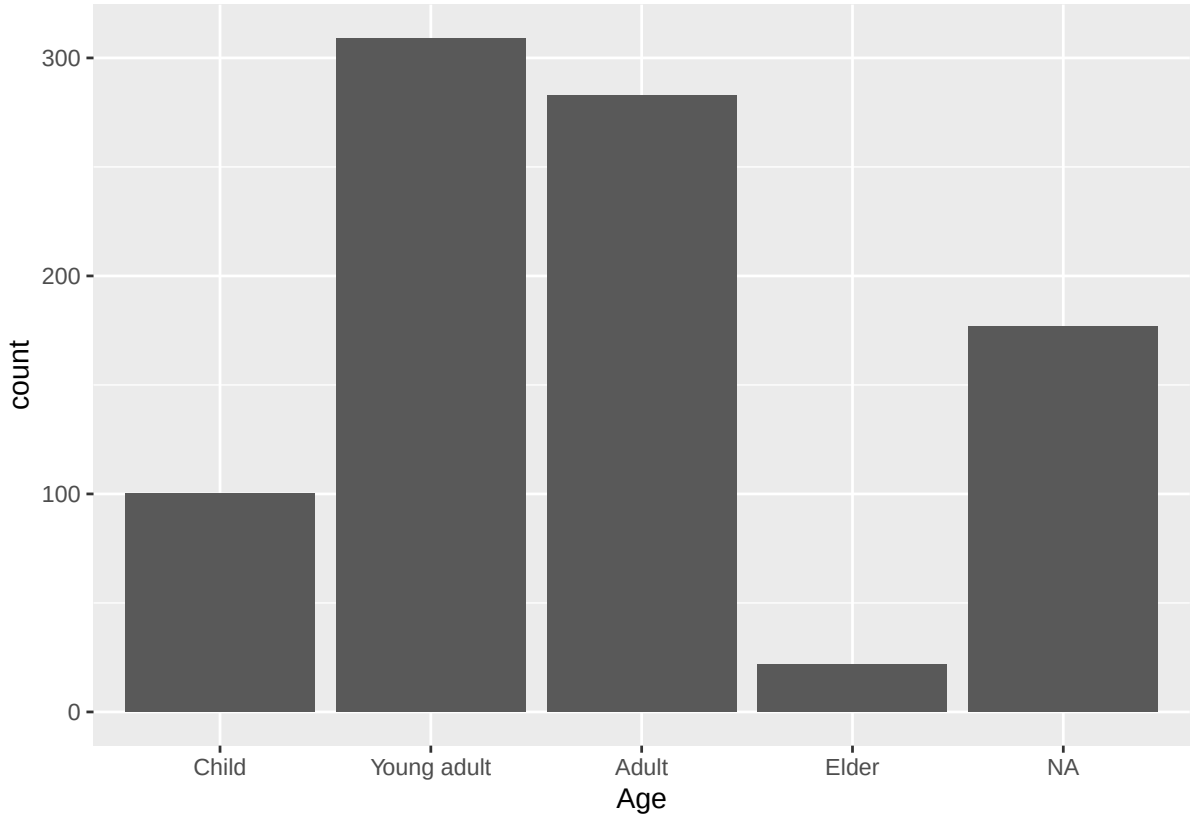
```
dataset <- read.csv("titanic.csv")

## Eliminamos variables que no nos interesan para el estudio
dataset$PassengerId <- NULL
dataset$Name <- NULL
dataset$Ticket <- NULL
dataset$Cabin <- NULL

## Cuantizamos la información de edad en intervalos
dataset$Age <- cut(dataset$Age, breaks = c(0,16, 30, 60, Inf),
                   labels=c("Child", "Young adult", "Adult", "Elder"))
summary(dataset)
```

```
##     Survived          Pclass          Sex                Age
##  Min.   :0.0000   Min.   :1.000   female:314   Child      :100
##  1st Qu.:0.0000   1st Qu.:2.000   male  :577   Young adult:309
##  Median :0.0000   Median :3.000                Adult      :283
##  Mean   :0.3838   Mean   :2.309                Elder      : 22
##  3rd Qu.:1.0000   3rd Qu.:3.000                NA's       :177
##  Max.   :1.0000   Max.   :3.000
##      SibSp           Parch             Fare        Embarked
##  Min.   :0.000   Min.   :0.0000   Min.   :  0.00    :  2
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:  7.91   C:168
##  Median :0.000   Median :0.0000   Median : 14.45   Q: 77
##  Mean   :0.523   Mean   :0.3816   Mean   : 32.20   S:644
##  3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.: 31.00
##  Max.   :8.000   Max.   :6.0000   Max.   :512.33
```

```
ggplot(dataset, aes(Age)) + geom_bar()
```




```
print("foo")
```

```
## [1] "foo"
```