

PEC2: Estudio de técnicas de visualización de datos

Hernando Hernández Mariño

16 de noviembre, 2021

1. Cargar datos

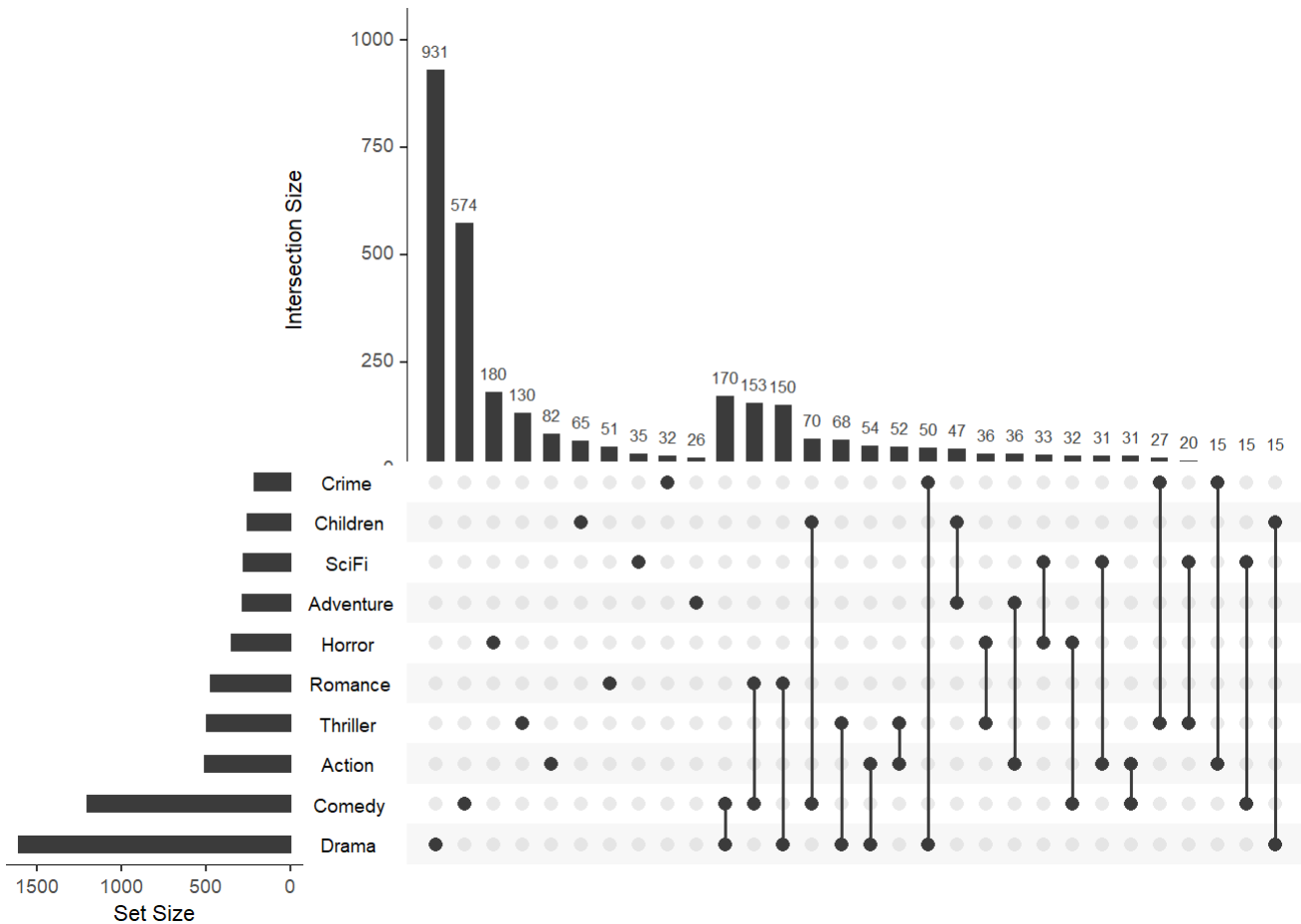
```
# Lectura de datos
movies <- read.csv(system.file("extdata","movies.csv",package = "UpSetR"), header = TRUE, sep=";")
```

```
# Estructura del conjunto de datos
str(movies)
```

```
## 'data.frame': 3883 obs. of 21 variables:
## $ Name : chr "Toy Story (1995)" "Jumanji (1995)" "Grumpier Old Men (1995)" "Waiting t
o Exhale (1995)" ...
## $ ReleaseDate: int 1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
## $ Action : int 0 0 0 0 0 1 0 0 1 1 ...
## $ Adventure : int 0 1 0 0 0 0 0 1 0 1 ...
## $ Children : int 1 1 0 0 0 0 0 1 0 0 ...
## $ Comedy : int 1 0 1 1 1 0 1 0 0 0 ...
## $ Crime : int 0 0 0 0 0 1 0 0 0 0 ...
## $ Documentary: int 0 0 0 0 0 0 0 0 0 0 ...
## $ Drama : int 0 0 0 1 0 0 0 0 0 0 ...
## $ Fantasy : int 0 1 0 0 0 0 0 0 0 0 ...
## $ Noir : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Horror : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Musical : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Mystery : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Romance : int 0 0 1 0 0 0 1 0 0 0 ...
## $ SciFi : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Thriller : int 0 0 0 0 0 1 0 0 0 1 ...
## $ War : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Western : int 0 0 0 0 0 0 0 0 0 0 ...
## $ AvgRating : num 4.15 3.2 3.02 2.73 3.01 3.88 3.41 3.01 2.66 3.54 ...
## $ Watches : int 2077 701 478 170 296 940 458 68 102 888 ...
```

2. Procedimiento a realizar.

```
upset(movies, nsets = 10, nintersects = 30, mb.ratio = c(0.5, 0.5),
      order.by = c("freq", "degree"), decreasing = c(TRUE,FALSE))
```



3. Análisis e interpretación de resultados.

En el lado izquierdo se observan unas barras que representan el tamaño total de cada conjunto. Se evidencian 10 conjunto de datos. Así por ejemplo, en el caso de “Drama, hay un poco más de 1500 películas que se identifican con el género dramático.

En la parte superior se observa el tamaño de cada conjunto que se cruza. Ahora bien, si se completa un punto negro, ese atributo se incluye en el conjunto. Si el punto es gris, entonces ese atributo no está en el conjunto; los conjuntos están separados, es decir, no se superponen.

A la izquierda del gráfico se observa los puntos de datos representados de forma única en las barras de la parte superior; por ejemplo, la segunda barra representa a 574 películas que son del género “Comedy”.

En contraste, en las barras de la izquierda se aprecian las distintas intersecciones de los géneros de las películas; por ejemplo, la barra que representa a 174 películas cuyos géneros “Comedy” y “Drama” se entre cruzan.

Es interesante observar del gráfico que las películas entre los años 1919 y 2020 son únicas en relación a su género o manejan no más de dos géneros.