

# Proyecto1

Juan Luis Solórzano (carnet: 201598)

Micaela Yataz (carnet: 18960)

2025-01-20

```
##          id          budget          genres          homePage
## Min.      :      5    Min.      :      0    Length:10000    Length:10000
## 1st Qu.: 12286    1st Qu.:      0    Class :character    Class :character
## Median :152558    Median :   500000    Mode  :character    Mode  :character
## Mean    :249877    Mean    : 18551632
## 3rd Qu.:452022    3rd Qu.: 20000000
## Max.    :922260    Max.    :380000000
## productionCompany productionCompanyCountry productionCountry
## Length:10000      Length:10000          Length:10000
## Class :character   Class :character      Class :character
## Mode  :character   Mode  :character      Mode  :character
##
##
##
##          revenue          runtime          video          director
## Min.      :0.000e+00    Min.      : 0.0    Mode :logical    Length:10000
## 1st Qu.:0.000e+00    1st Qu.: 90.0    FALSE:9430      Class :character
## Median :1.631e+05     Median :100.0    TRUE :84        Mode  :character
## Mean    :5.674e+07     Mean    :100.3    NA's :486
## 3rd Qu.:4.480e+07     3rd Qu.:113.0
## Max.    :2.847e+09     Max.    :750.0
##          actors          actorsPopularity          actorsCharacter          originalTitle
## Length:10000      Length:10000          Length:10000          Length:10000
## Class :character   Class :character      Class :character      Class :character
## Mode  :character   Mode  :character      Mode  :character      Mode  :character
##
##
##
##          title          originalLanguage          popularity          releaseDate
## Length:10000      Length:10000          Min.      :    4.258    Length:10000
## Class :character   Class :character      1st Qu.:   14.578    Class :character
## Mode  :character   Mode  :character      Median :   21.906    Mode  :character
##                                     Mean    :   51.394
##                                     3rd Qu.:   40.654
##                                     Max.    :11474.647
##          voteAvg          voteCount          genresAmount          productionCoAmount
## Min.      : 1.300    Min.      :    1    Min.      : 0.000    Min.      : 0.000
## 1st Qu.: 5.900    1st Qu.:  120    1st Qu.: 2.000    1st Qu.: 2.000
## Median : 6.500    Median :   415    Median : 3.000    Median : 3.000
## Mean    : 6.483    Mean    : 1342    Mean    : 2.596    Mean    : 3.171
## 3rd Qu.: 7.200    3rd Qu.: 1316    3rd Qu.: 3.000    3rd Qu.: 4.000
## Max.    :10.000    Max.    :30788    Max.    :16.000    Max.    :89.000
## productionCountriesAmount actorsAmount    castWomenAmount
```

```
## Min.      : 0.000           Min.      :    0   Length:10000
## 1st Qu.:  1.000           1st Qu.:   13   Class :character
## Median   :  1.000           Median   :   21   Mode  :character
## Mean     :  1.751           Mean     :  2148
## 3rd Qu.:  2.000           3rd Qu.:   36
## Max.     :155.000           Max.     :919590
## castMenAmount
## Length:10000
## Class :character
## Mode  :character
##
##
##
```

## 1. Clustering

**1.1. Haga el preprocesamiento del dataset, explique qué variables no aportan información a la generación de grupos y por qué. Describa con qué variables calculará los grupos.**

Como el algoritmo de k-medias y el clustering jerárquico necesitan de alguna medida de distancia, entre los datos, en una primera instancia vamos a tomar solo las variables numéricas y vamos a quitar el id por ser como el nombre de una película. Las variables que tomaremos en consideración son las siguientes:

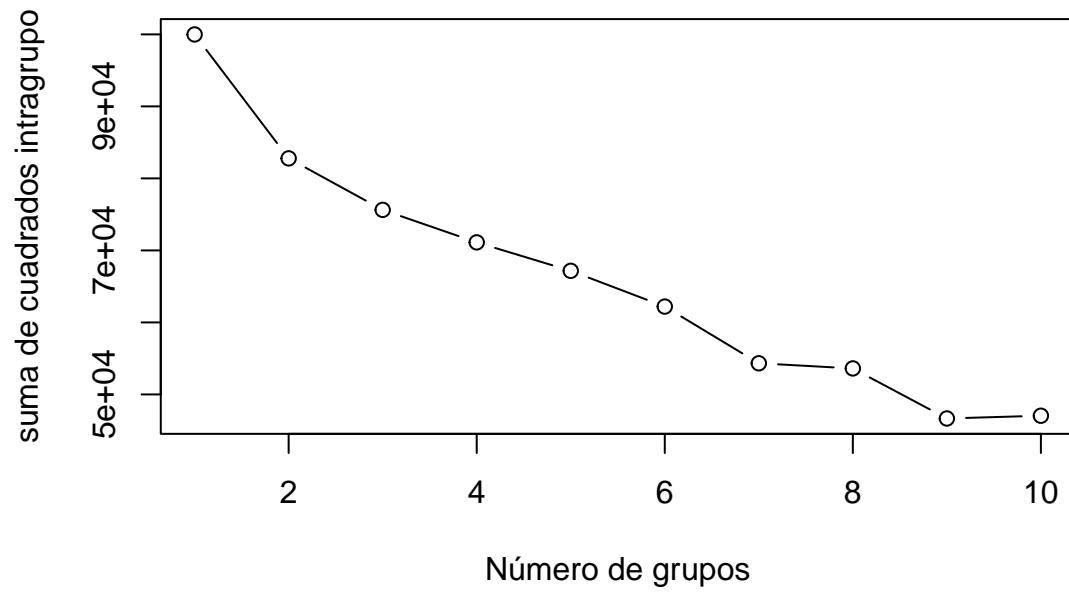
```
## 'data.frame':  10000 obs. of  10 variables:
## $ budget      : int  4000000 21000000 11000000 94000000 55000000 15000000 839727 128000000
## $ revenue     : num  4.26e+06 1.21e+07 7.75e+08 9.40e+08 6.77e+08 ...
## $ runtime     : int  98 110 121 100 142 122 119 141 126 149 ...
## $ popularity  : num  20.9 9.6 100 134.4 58.8 ...
## $ voteAvg     : num  5.7 6.5 8.2 7.8 8.5 8 8 7.9 7.5 8.2 ...
## $ voteCount   : int  2077 223 16598 15928 22045 9951 4253 1335 8726 1963 ...
## $ genresAmount : int  2 3 3 2 3 1 2 2 5 2 ...
## $ productionCoAmount : int  2 3 2 1 2 2 2 26 2 1 ...
## $ productionCountriesAmount: int  1 2 1 1 1 1 1 12 1 1 ...
## $ actorsAmount : int  25 15 105 24 76 40 152 29 117 24 ...
```

**1.2. Analice la tendencia al agrupamiento usando el estadístico de Hopkins y la VAT (Visual Assessment of cluster Tendency). Esta última hágala si es posible, teniendo en cuenta las dimensiones del conjunto de datos. Discuta sus resultados e impresiones.**

El estadístico de Hopkins es de 1 que es lejano a 0.5, entonces los datos no son aleatorios. Sin embargo no haremos un VAT por ser difícil de visualizar e interpretar con 10 variables.

**1.3. Determine cuál es el número de grupos a formar más adecuado para los datos que está trabajando. Haga una gráfica de codo y explique la razón de la elección de la cantidad de clústeres con la que trabajará.**

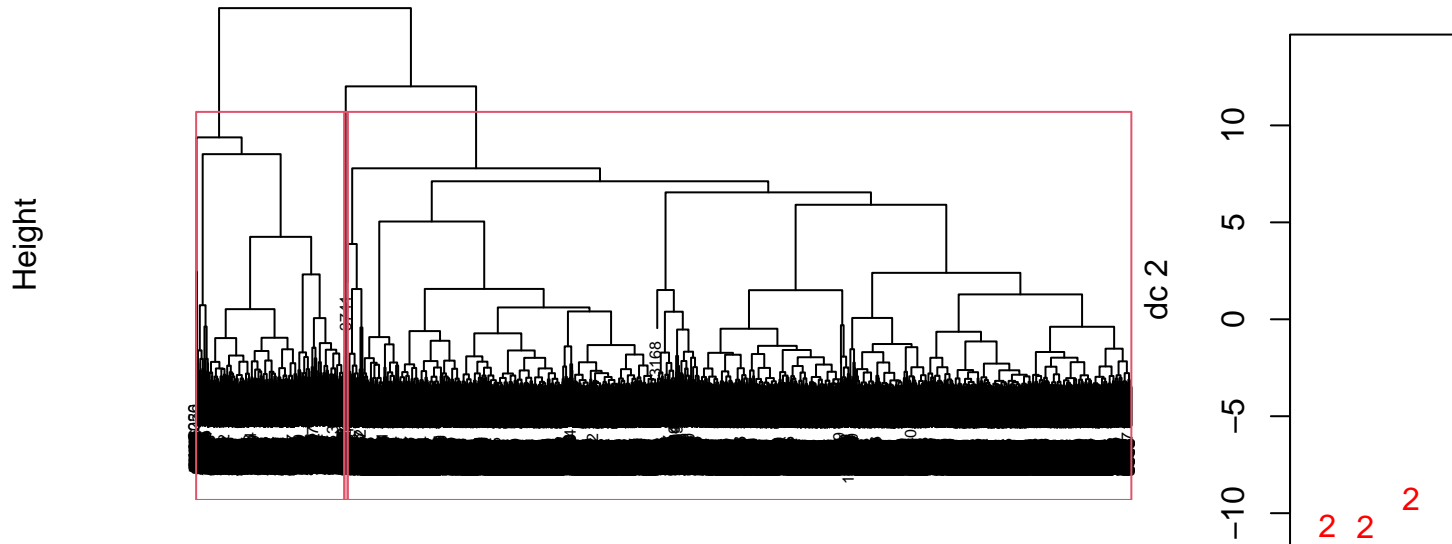
Para ello se usará el método de Codo



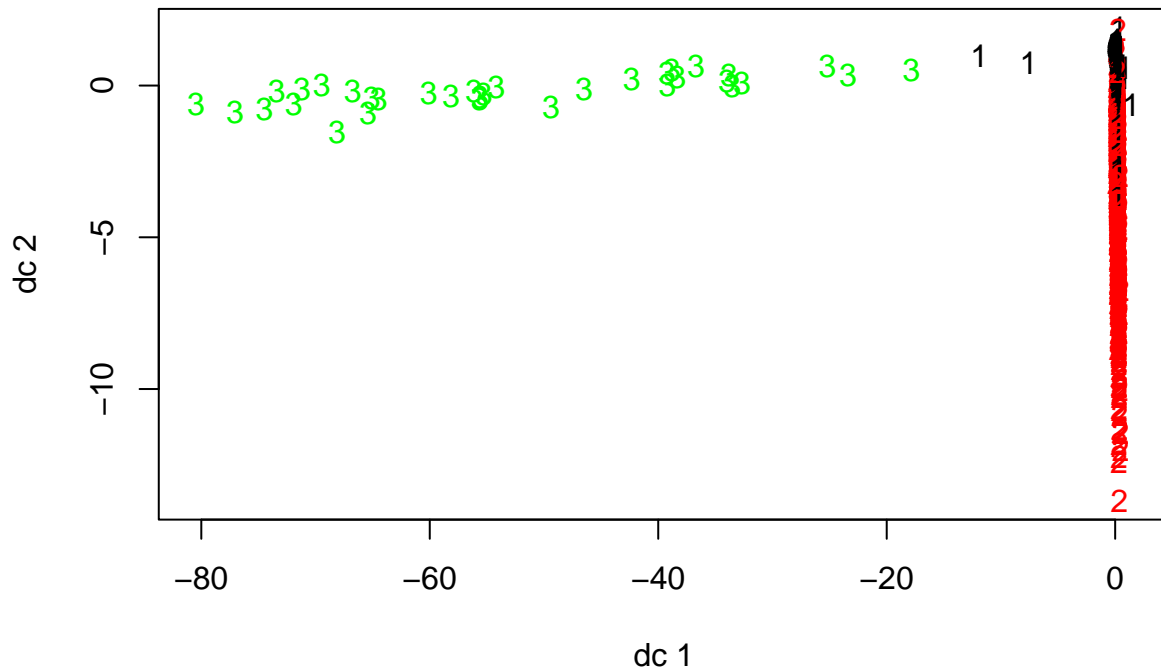
Como a partir de 3 grupos en adelante la suma de cuadrados intragrupo no disminuye tan rápido se elegirán 3 grupos.

1.4. Utilice los algoritmos k-medias y clustering jerárquico para agrupar. Compare los resultados generados por cada uno.

### Cluster Dendrogram



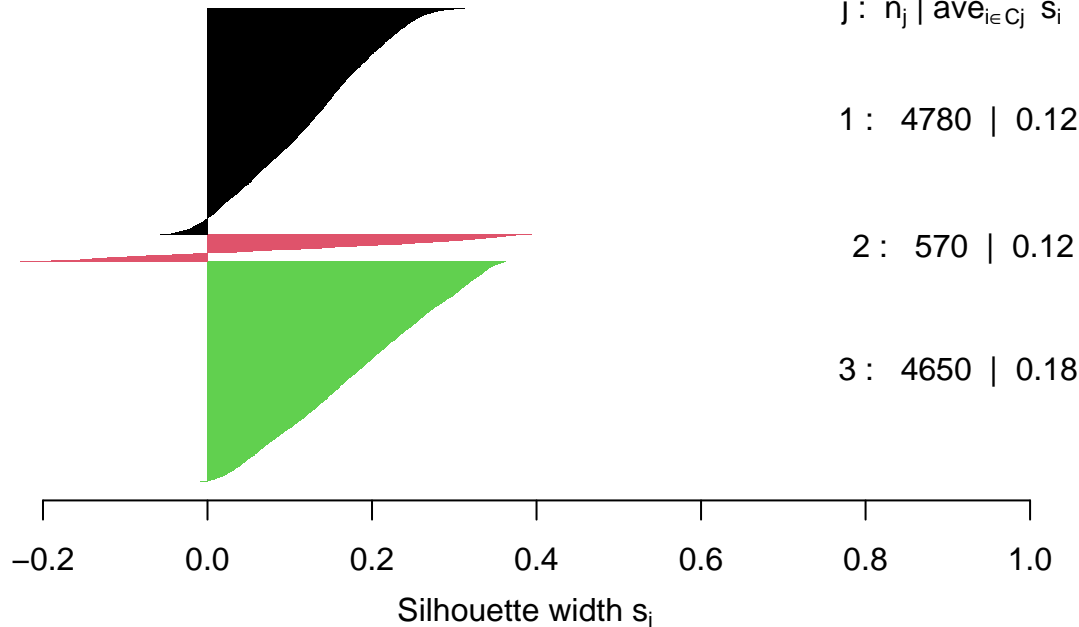
### Clusters generados por Clustering Jerárquico



1.5. Determine la calidad del agrupamiento hecho por cada algoritmo con el método de la silueta. Discuta los resultados.

### Silueta K-means

n = 10000



### Silueta Clustering Jerárquico

n = 10000



Usando el método de la silueta el clustering jerárquico que tiene una silueta promedio de  $0.35 > 0.15$  del clustering de Kmedias. Entonces para estos datos el clusterin jerárquico obtuvo un mejor resultado. Esto tiene sentido pues en la entrega anterior vimos que ninguna variable se comportaba de manera normal. En estos casos el k-medias no suele ser tan eficiente.

1.6. Interprete los grupos basado en el conocimiento que tiene de los datos. Recuerde investigar las medidas de tendencia central de las variables continuas y las tablas de frecuencia de las variables categóricas pertenecientes a cada grupo. Identifique hallazgos interesantes debido a las agrupaciones y describa para qué le podría servir.

## Tamaños de los dataframes:

## K-Means -> Cluster 1: 4780 | Cluster 2: 570 | Cluster 3: 4650

## Jerárquico -> Cluster 1: 8380 | Cluster 2: 1585 | Cluster 3: 35

`summary(df_kmeans_1)`

```
##           id           budget           genres           homePage
##  Min.      :    6   Min.      :0.00e+00   Length:4780   Length:4780
## 1st Qu.: 9952   1st Qu.:0.00e+00   Class :character   Class :character
## Median : 50781   Median :7.29e+06   Mode  :character   Mode  :character
## Mean   :204505   Mean   :1.91e+07
## 3rd Qu.:401728   3rd Qu.:3.00e+07
## Max.    :912223   Max.    :1.75e+08
## productionCompany productionCompanyCountry productionCountry
```

```

## Length:4780      Length:4780      Length:4780
## Class :character  Class :character  Class :character
## Mode :character  Mode :character  Mode :character
##
##
##
##      revenue      runtime      video      director
## Min.   :      0      Min.   : 0.0      Mode :logical  Length:4780
## 1st Qu.:      0      1st Qu.: 96.0      FALSE:4556      Class :character
## Median : 12355433      Median :108.0      TRUE :36        Mode :character
## Mean   : 47683585      Mean   :110.1      NA's :188
## 3rd Qu.: 68585560      3rd Qu.:121.0
## Max.   :742500000      Max.   :750.0
##      actors      actorsPopularity      actorsCharacter      originalTitle
## Length:4780      Length:4780      Length:4780      Length:4780
## Class :character  Class :character  Class :character  Class :character
## Mode :character  Mode :character  Mode :character  Mode :character
##
##
##
##      title      originalLanguage      popularity      releaseDate
## Length:4780      Length:4780      Min.   : 4.627      Length:4780
## Class :character  Class :character  1st Qu.: 15.330      Class :character
## Mode :character  Mode :character  Median : 22.491      Mode :character
##                                     Mean   : 43.519
##                                     3rd Qu.: 38.748
##                                     Max.   :1553.397
##      voteAvg      voteCount      genresAmount      productionCoAmount
## Min.   : 2.000      Min.   : 1.0      Min.   : 1.0      Min.   : 1.000
## 1st Qu.: 6.400      1st Qu.: 264.8      1st Qu.: 2.0      1st Qu.: 2.000
## Median : 6.900      Median : 775.0      Median : 3.0      Median : 3.000
## Mean   : 6.887      Mean   : 1377.7      Mean   : 3.1      Mean   : 3.909
## 3rd Qu.: 7.400      3rd Qu.: 1819.0      3rd Qu.: 4.0      3rd Qu.: 5.000
## Max.   :10.000      Max.   :11913.0      Max.   :16.0      Max.   :89.000
##      productionCountriesAmount      actorsAmount      castWomenAmount
## Min.   : 0.000      Min.   : 0.0      Length:4780
## 1st Qu.: 1.000      1st Qu.: 16.0      Class :character
## Median : 1.000      Median : 25.0      Mode :character
## Mean   : 1.581      Mean   : 511.2
## 3rd Qu.: 2.000      3rd Qu.: 41.0
## Max.   :28.000      Max.   :784594.0
##      castMenAmount      Cluster_KMeans      Cluster_Jerarquico
## Length:4780      1:4780      1:3812
## Class :character  2: 0      2: 964
## Mode :character  3: 0      3: 4
##
##
##

```

```
summary(df_kmeans_3)
```

```

##      id      budget      genres      homePage
## Min.   : 5      Min.   : 0      Length:4650      Length:4650
## 1st Qu.: 26284      1st Qu.: 0      Class :character  Class :character
## Median :292681      Median : 0      Mode :character  Mode :character

```

```

## Mean :310177 Mean : 4859315
## 3rd Qu.:517274 3rd Qu.: 3500000
## Max. :922260 Max. :170000000
## productionCompany productionCompanyCountry productionCountry
## Length:4650 Length:4650 Length:4650
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
## revenue runtime video director
## Min. : 0 Min. : 0.00 Mode :logical Length:4650
## 1st Qu.: 0 1st Qu.: 85.00 FALSE:4440 Class :character
## Median : 0 Median : 92.00 TRUE :48 Mode :character
## Mean : 10508480 Mean : 87.45 NA's :162
## 3rd Qu.: 1463556 3rd Qu.:100.00
## Max. :347988775 Max. :240.00
## actors actorsPopularity actorsCharacter originalTitle
## Length:4650 Length:4650 Length:4650 Length:4650
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## title originalLanguage popularity releaseDate
## Length:4650 Length:4650 Min. : 4.258 Length:4650
## Class :character Class :character 1st Qu.: 13.450 Class :character
## Mode :character Mode :character Median : 19.168 Mode :character
## Mean : 38.732
## 3rd Qu.: 33.058
## Max. :4789.705
## voteAvg voteCount genresAmount productionCoAmount
## Min. : 1.300 Min. : 1.0 Min. :0.000 Min. : 0.000
## 1st Qu.: 5.400 1st Qu.: 63.0 1st Qu.:1.000 1st Qu.: 1.000
## Median : 6.000 Median : 173.0 Median :2.000 Median : 2.000
## Mean : 6.007 Mean : 392.7 Mean :2.003 Mean : 2.326
## 3rd Qu.: 6.700 3rd Qu.: 471.0 3rd Qu.:3.000 3rd Qu.: 3.000
## Max. :10.000 Max. :6000.0 Max. :9.000 Max. :17.000
## productionCountriesAmount actorsAmount castWomenAmount
## Min. : 0.000 Min. : 0 Length:4650
## 1st Qu.: 1.000 1st Qu.: 9 Class :character
## Median : 1.000 Median : 16 Mode :character
## Mean : 1.955 Mean : 4086
## 3rd Qu.: 1.000 3rd Qu.: 26
## Max. :155.000 Max. :919590
## castMenAmount Cluster_KMeans Cluster_Jerarquico
## Length:4650 1: 0 1:4566
## Class :character 2: 0 2: 53
## Mode :character 3:4650 3: 31
##
##
##

```

```
summary(df_kmeans_2)
```

```
##          id          budget          genres          homePage
## Min.   :    11   Min.   :      0   Length:570   Length:570
## 1st Qu.: 2093   1st Qu.: 80000000   Class :character   Class :character
## Median : 51058   Median :125000000   Mode  :character   Mode  :character
## Mean   :138438   Mean   :125618804
## 3rd Qu.:274867   3rd Qu.:165000000
## Max.   :791373   Max.   :380000000
## productionCompany productionCompanyCountry productionCountry
## Length:570        Length:570        Length:570
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##          revenue          runtime          video          director
## Min.   :0.000e+00   Min.   : 81.0   Mode :logical   Length:570
## 1st Qu.:2.904e+08   1st Qu.:106.0   FALSE:434   Class :character
## Median :4.111e+08   Median :120.0   NA's :136   Mode  :character
## Mean   :5.098e+08   Mean   :122.7
## 3rd Qu.:6.524e+08   3rd Qu.:136.0
## Max.   :2.847e+09   Max.   :242.0
##          actors          actorsPopularity          actorsCharacter          originalTitle
## Length:570        Length:570        Length:570        Length:570
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##          title          originalLanguage          popularity          releaseDate
## Length:570        Length:570        Min.   :    5.165   Length:570
## Class :character   Class :character   1st Qu.:   49.747   Class :character
## Mode  :character   Mode  :character   Median :   88.116   Mode  :character
##                               Mean   :  220.728
##                               3rd Qu.:  150.984
##                               Max.   :11474.647
##          voteAvg          voteCount          genresAmount          productionCoAmount
## Min.   :4.300   Min.   : 15   Min.   :1.000   Min.   : 1.000
## 1st Qu.:6.400   1st Qu.: 5121   1st Qu.:3.000   1st Qu.: 2.000
## Median :7.000   Median : 7438   Median :3.000   Median : 4.000
## Mean   :6.984   Mean   : 8794   Mean   :3.214   Mean   : 3.879
## 3rd Qu.:7.600   3rd Qu.:11654   3rd Qu.:4.000   3rd Qu.: 5.000
## Max.   :8.700   Max.   :30788   Max.   :6.000   Max.   :12.000
## productionCountriesAmount actorsAmount          castWomenAmount
## Min.   :1.000        Min.   : 7.00   Length:570
## 1st Qu.:1.000        1st Qu.: 35.00   Class :character
## Median :1.000        Median : 50.50   Mode  :character
## Mean   :1.512        Mean   : 57.98
## 3rd Qu.:2.000        3rd Qu.: 73.75
## Max.   :8.000        Max.   :248.00
## castMenAmount          Cluster_KMeans Cluster_Jerarquico
## Length:570        1: 0        1: 2
## Class :character   2:570        2:568
```



```
## Mode :character 3: 0      3: 0
##
##
##
```

```
summary(df_jerarquico_1)
```

```
##          id          budget          genres          homePage
## Min.   :      5   Min.   :      0   Length:8380   Length:8380
## 1st Qu.: 15156   1st Qu.:      0   Class :character   Class :character
## Median :227962   Median :      0   Mode :character   Mode :character
## Mean   :273544   Mean   : 7285524
## 3rd Qu.:481873   3rd Qu.: 10000000
## Max.   :922260   Max.   :110000000
## productionCompany productionCompanyCountry productionCountry
## Length:8380      Length:8380      Length:8380
## Class :character   Class :character      Class :character
## Mode :character   Mode :character      Mode :character
##
##
##
##          revenue          runtime          video          director
## Min.   :      0   Min.   : 0.00   Mode :logical   Length:8380
## 1st Qu.:      0   1st Qu.: 88.00   FALSE:8017      Class :character
## Median :      0   Median : 97.00   TRUE :72        Mode :character
## Mean   : 17646314   Mean   : 97.17   NA's :291
## 3rd Qu.: 14818341   3rd Qu.:109.00
## Max.   :686257563   Max.   :750.00
##          actors          actorsPopularity          actorsCharacter          originalTitle
## Length:8380      Length:8380      Length:8380      Length:8380
## Class :character   Class :character   Class :character   Class :character
## Mode :character   Mode :character   Mode :character   Mode :character
##
##
##
##          title          originalLanguage          popularity          releaseDate
## Length:8380      Length:8380      Min.   : 4.258   Length:8380
## Class :character   Class :character   1st Qu.: 13.853   Class :character
## Mode :character   Mode :character   Median : 19.850   Mode :character
##                                     Mean   : 39.925
##                                     3rd Qu.: 33.760
##                                     Max.   :1699.701
##          voteAvg          voteCount          genresAmount          productionCoAmount
## Min.   : 1.300   Min.   : 1.0   Min.   :0.000   Min.   : 0.000
## 1st Qu.: 5.800   1st Qu.: 100.0   1st Qu.:2.000   1st Qu.: 1.000
## Median : 6.500   Median : 294.0   Median :2.000   Median : 2.000
## Mean   : 6.423   Mean   : 620.2   Mean   :2.512   Mean   : 3.025
## 3rd Qu.: 7.100   3rd Qu.: 805.0   3rd Qu.:3.000   3rd Qu.: 4.000
## Max.   :10.000   Max.   :7818.0   Max.   :8.000   Max.   :89.000
## productionCountriesAmount actorsAmount          castWomenAmount
## Min.   : 0.000      Min.   : 0.00   Length:8380
## 1st Qu.: 1.000      1st Qu.: 11.00   Class :character
## Median : 1.000      Median : 19.00   Mode :character
## Mean   : 1.797      Mean   : 50.36
## 3rd Qu.: 2.000      3rd Qu.: 30.00
```

```
## Max. :155.000 Max. :138550.00
## castMenAmount Cluster_KMeans Cluster_Jerarquico
## Length:8380 1:3812 1:8380
## Class :character 2: 2 2: 0
## Mode :character 3:4566 3: 0
##
##
##
```

```
summary(df_jerarquico_2)
```

```
## id budget genres homePage
## Min. : 11 Min. :0.00e+00 Length:1585 Length:1585
## 1st Qu.: 3050 1st Qu.:3.60e+07 Class :character Class :character
## Median : 15512 Median :6.60e+07 Mode :character Mode :character
## Mean :117032 Mean :7.85e+07
## 3rd Qu.:205775 3rd Qu.:1.05e+08
## Max. :860623 Max. :3.80e+08
## productionCompany productionCompanyCountry productionCountry
## Length:1585 Length:1585 Length:1585
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
## revenue runtime video director
## Min. :0.000e+00 Min. : 0.0 Mode :logical Length:1585
## 1st Qu.:8.541e+07 1st Qu.:103.0 FALSE:1389 Class :character
## Median :1.761e+08 Median :116.0 TRUE :1 Mode :character
## Mean :2.646e+08 Mean :117.7 NA's :195
## 3rd Qu.:3.450e+08 3rd Qu.:130.0
## Max. :2.847e+09 Max. :242.0
## actors actorsPopularity actorsCharacter originalTitle
## Length:1585 Length:1585 Length:1585 Length:1585
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## title originalLanguage popularity releaseDate
## Length:1585 Length:1585 Min. : 5.165 Length:1585
## Class :character Class :character 1st Qu.: 24.662 Class :character
## Mode :character Mode :character Median : 42.600 Mode :character
## Mean : 112.568
## 3rd Qu.: 85.619
## Max. :11474.647
## voteAvg voteCount genresAmount productionCoAmount
## Min. :3.200 Min. : 2 Min. :1.000 Min. : 0.000
## 1st Qu.:6.200 1st Qu.: 2003 1st Qu.:3.000 1st Qu.: 2.000
## Median :6.800 Median : 3920 Median :3.000 Median : 4.000
## Mean :6.808 Mean : 5190 Mean :3.043 Mean : 3.996
## 3rd Qu.:7.400 3rd Qu.: 6734 3rd Qu.:4.000 3rd Qu.: 5.000
## Max. :8.700 Max. :30788 Max. :6.000 Max. :13.000
## productionCountriesAmount actorsAmount castWomenAmount
## Min. :0.000 Min. : 0.00 Length:1585
```

```
## 1st Qu.:1.000          1st Qu.: 28.00   Class :character
## Median :1.000          Median : 43.00   Mode  :character
## Mean   :1.523          Mean   : 48.87
## 3rd Qu.:2.000          3rd Qu.: 61.00
## Max.   :8.000          Max.    :248.00
## castMenAmount      Cluster_KMeans Cluster_Jerarquico
## Length:1585        1:964          1: 0
## Class :character    2:568          2:1585
## Mode  :character    3: 53           3: 0
##
##
##
```

```
summary(df_jerarquico_3)
```

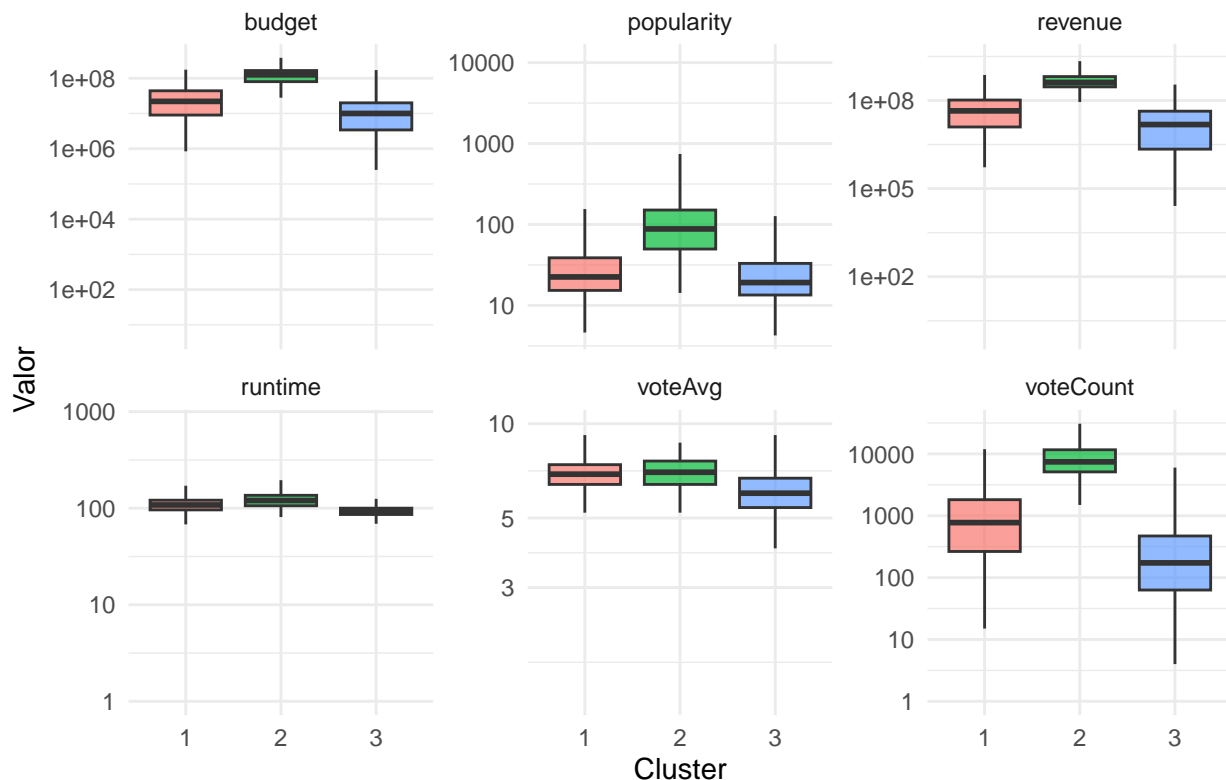
```
##          id          budget          genres          homePage
## Min.   :204450  Min.   : 0  Length:35  Length:35
## 1st Qu.:442164  1st Qu.: 0  Class :character  Class :character
## Median :640097  Median : 0  Mode  :character  Mode  :character
## Mean   :599348  Mean   :1285715
## 3rd Qu.:757708  3rd Qu.: 0
## Max.   :919590  Max.   :45000000
## productionCompany productionCompanyCountry productionCountry
## Length:35        Length:35          Length:35
## Class :character  Class :character      Class :character
## Mode  :character  Mode  :character      Mode  :character
##
##
##
##          revenue          runtime          video          director
## Min.   : 0  Min.   : 0.00  Mode :logical  Length:35
## 1st Qu.: 0  1st Qu.: 0.00  FALSE:24      Class :character
## Median : 0  Median : 16.00  TRUE :11      Mode  :character
## Mean   : 5093218  Mean   : 52.83
## 3rd Qu.: 0  3rd Qu.: 85.00
## Max.   :178262620  Max.   :182.00
##          actors          actorsPopularity          actorsCharacter          originalTitle
## Length:35        Length:35          Length:35          Length:35
## Class :character  Class :character      Class :character      Class :character
## Mode  :character  Mode  :character      Mode  :character      Mode  :character
##
##
##
##          title          originalLanguage          popularity          releaseDate
## Length:35        Length:35          Min.   : 4.627  Length:35
## Class :character  Class :character      1st Qu.:17.511  Class :character
## Mode  :character  Mode  :character      Median :26.232  Mode  :character
##                                     Mean   :27.189
##                                     3rd Qu.:36.328
##                                     Max.   :60.162
##          voteAvg          voteCount          genresAmount          productionCoAmount
## Min.   : 1.500  Min.   : 1.000  Min.   : 0.000  Min.   :0.0
## 1st Qu.: 4.000  1st Qu.: 1.000  1st Qu.: 0.000  1st Qu.:0.0
## Median : 7.000  Median : 3.000  Median : 1.000  Median :0.0
## Mean   : 6.283  Mean   : 5.914  Mean   : 2.686  Mean   :0.8
```

```
## 3rd Qu.: 8.200    3rd Qu.: 9.000    3rd Qu.: 5.000    3rd Qu.:1.0
## Max.    :10.000    Max.    :20.000    Max.    :16.000    Max.    :4.0
## productionCountriesAmount  actorsAmount    castWomenAmount
## Min.    :0.000          Min.    :204450    Length:35
## 1st Qu.:0.000          1st Qu.:442164    Class :character
## Median :0.000          Median :640097    Mode  :character
## Mean    :1.086          Mean    :599348
## 3rd Qu.:2.000          3rd Qu.:757708
## Max.    :9.000          Max.    :919590
## castMenAmount    Cluster_KMeans Cluster_Jerarquico
## Length:35        1: 4          1: 0
## Class :character  2: 0          2: 0
## Mode  :character  3:31         3:35
##
##
##
```

Como es difícil interpretar los números puros vamos a hacer unas gráficas de caja y bigote.

```
## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
## Warning: Removed 9688 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

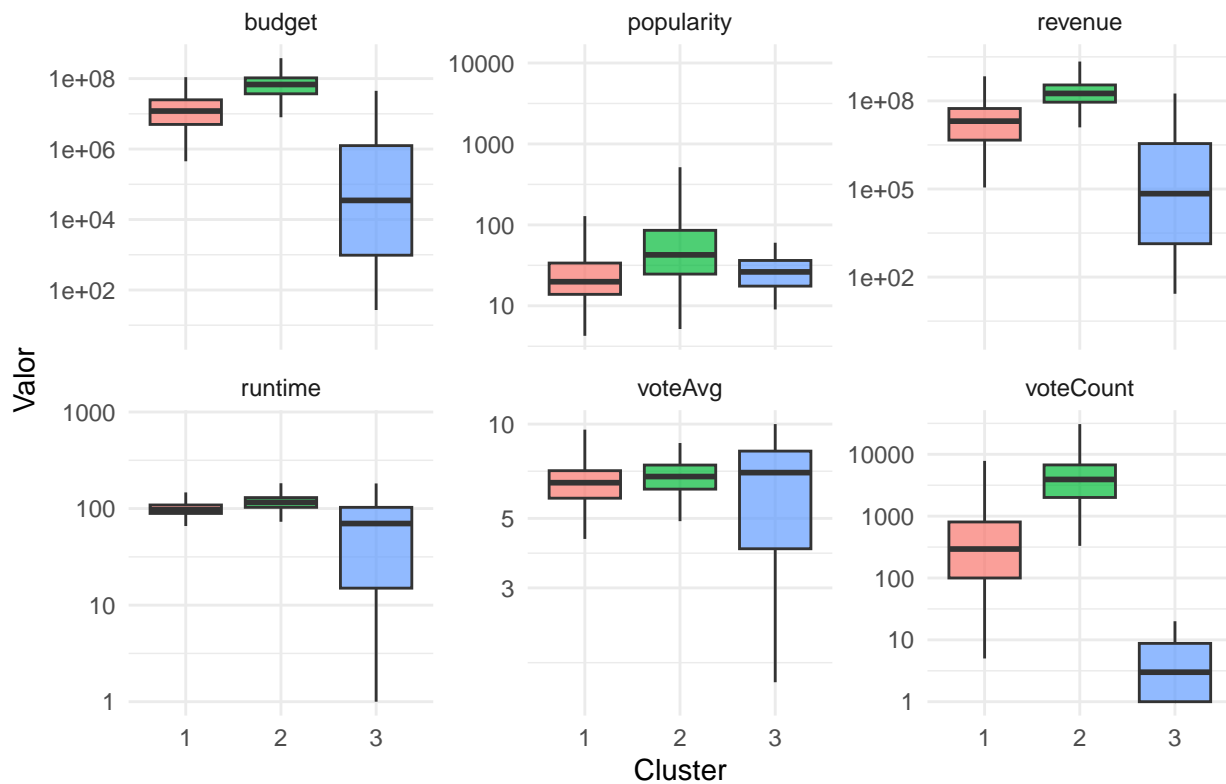
### Distribución de Variables por Clusters (K-Means)



koewsjfokerdsf

```
## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
## Warning: Removed 9688 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

## Distribución de Variables por Clusters (Jerárquico)



#### Clusters generados por K-Means:

El grupo 1: Presenta valores intermedios en presupuesto, recaudación y popularidad. Probablemente incluye películas de presupuesto moderado y popularidad moderada.  
 El grupo 2: Contiene películas con los presupuestos y recaudaciones más altos, alta popularidad y una gran cantidad de votos.  
 El grupo 3: Se caracteriza por películas con presupuestos y recaudaciones bajas, menor popularidad y pocas votos.

### Clusters generados por Clustering Jerárquico:

El grupo 1: Incluye películas con recaudación y presupuesto moderado, popularidad mas bien baja.  
 El grupo 2: Películas de alto presupuesto, altas recaudacion y popularidad  
 El grupo 3: Agrupa películas de bajo presupuesto, con popularidades relativamente bajas y recaudaciones bajas.