

HDT 1

2023-02-02

###Universidad del Valle de Guatemala ###Mineria de datos ###Elean Rivas - 19062 ###Javier Alarez - 18051

```
películas <- read.csv("C:\\Users\\javie\\Desktop\\U\\2023f\\mineria\\movies.csv")
```

Descripción del conjunto de datos. El conjunto de datos está compuesto por `rnrow(películas)` filas y `ncol(películas)` columnas

1. Resumen del conjunto de datos

```
head(películas)
```

```
##   id   budget                               genres
## 1  5  4000000                               Crime|Comedy
## 2  6 21000000          Action|Thriller|Crime
## 3 11 11000000 Adventure|Action|Science Fiction
## 4 12 94000000          Animation|Family
## 5 13 55000000          Comedy|Drama|Romance
## 6 14 15000000          Drama
##                                     homePage
## 1                https://www.miramax.com/movie/four-rooms/
## 2                                     <NA>
## 3 http://www.starwars.com/films/star-wars-episode-iv-a-new-hope
## 4                http://movies.disney.com/finding-nemo
## 5                                     <NA>
## 6                                     <NA>
##               productionCompany productionCompanyCountry
## 1                Miramax|A Band Apart                US|US
## 2 Universal Pictures|Largo Entertainment|JVC            US|US|JP
## 3                Lucasfilm|20th Century Fox            US|US
## 4                  Pixar                                US
## 5      Paramount|The Steve Tisch Company            US|
## 6   Jinks/Cohen Company|DreamWorks Pictures            |US
##               productionCountry   revenue runtime video
## 1      United States of America  4257354      98 FALSE
## 2 Japan|United States of America 12136938     110 FALSE
## 3      United States of America 775398007     121   NA
## 4      United States of America 940335536     100   NA
## 5      United States of America 677387716     142 FALSE
## 6      United States of America 356296601     122 FALSE
##                                     director
## 1 Allison Anders|Alexandre Rockwell|Robert Rodriguez|Quentin Tarantino
## 2                                     Stephen Hopkins
## 3                                     George Lucas
## 4                                     Andrew Stanton
## 5                                     Robert Zemeckis
```

```

## 6                                     Sam Mendes
##
## 1
## 2
## 3
## 4
## 5 Tom Hanks|Robin Wright|Gary Sinise|Mykelti Williamson|Sally Field|Michael Conner Humphreys|Hanna Ha
## 6
##
## 1
## 2
## 3 11.881|24.542|14.434|10.651|6.888|5.811|1.432|1.288|2.086|8.166|1.701|2.033|0.98|3.702|2.392|0.98|
## 4
## 5
## 6
##
## 1
## 2
## 3 Luke Skywalker|Han Solo|Princess Leia Organa|Grand Moff Tarkin|Obi-Wan "Ben" Kenobi|C-3P0|R2-D2|Ch
## 4
## 5
## 6
##      originalTitle      title originalLanguage popularity releaseDate
## 1      Four Rooms      Four Rooms              en      20.880 1995-12-09
## 2 Judgment Night Judgment Night              en       9.596 1993-10-15
## 3      Star Wars      Star Wars              en     100.003 1977-05-25
## 4 Finding Nemo      Finding Nemo              en     134.435 2003-05-30
## 5 Forrest Gump      Forrest Gump              en      58.751 1994-07-06
## 6 American Beauty American Beauty              en      33.589 1999-09-15
##      voteAvg voteCount genresAmount productionCoAmount productionCountriesAmount
## 1      5.7      2077          2              2              1
## 2      6.5       223          3              3              2
## 3      8.2     16598          3              2              1
## 4      7.8     15928          2              1              1
## 5      8.5     22045          3              2              1
## 6      8.0      9951          1              2              1
##      actorsAmount castWomenAmount castMenAmount
## 1          25          15          9
## 2          15           3          9
## 3         105           5         62
## 4          24           5         18
## 5          76          18         48
## 6          40          22         15

```

```
summary(peliculas)
```

```

##      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
##
##
##

```

###2. Descripción de las variables #Aquí se describen las variables:

```
class(películas$id)
```

```
## [1] "integer"
```

```
#cualitativa ordinal
```

```

class(peliculas$budget)

## [1] "integer"
#cuantitativa continua
class(peliculas$genres)

## [1] "character"
#cualitativa nominal
class(peliculas$homePage)

## [1] "character"
#cualitativa nominal
class(peliculas$productionCompany)

## [1] "character"
#cualitativa nominal
class(peliculas$productionCompanyCountry)

## [1] "character"
#cualitativa nominal
class(peliculas$productionCountry)

## [1] "character"
#cualitativa nominal
class(peliculas$revenue)

## [1] "numeric"
#cuantitativa continua
class(peliculas$runtime)

## [1] "integer"
#cuantitativa discreta
class(peliculas$video)

## [1] "logical"
#cualitativa ordinal
class(peliculas$director)

## [1] "character"
#cualitativa nominal
class(peliculas$actors)

## [1] "character"
#cualitativa nominal

```

```

class(peliculas$actorsPopularity)

## [1] "character"
#cuantitativa continua
class(peliculas$actorsCharacter)

## [1] "character"
#cualitativa ordinal
class(peliculas$originalTitle)

## [1] "character"
#cualitativa ordinal
class(peliculas$title)

## [1] "character"
#cualitativa ordinal
class((peliculas$originalLanguage))

## [1] "character"
#cualitativa ordinal
class(peliculas$popularity)

## [1] "numeric"
#cuantitativa continua
class(peliculas$releaseDate)

## [1] "character"
#cualitativa ordinal
class(peliculas$voteAvg)

## [1] "numeric"
#cuantitativa continua
class(peliculas$voteCount)

## [1] "integer"
#cuantitativa discreta
class(peliculas$genresAmount)

## [1] "integer"
#cuantitativa discreta
class(peliculas$productionCoAmount)

## [1] "integer"
#cuantitativa discreta

```

```
class(películas$productionCountriesAmount)
```

```
## [1] "integer"
```

```
#cuantitativa discreta
```

```
class(películas$actorsAmount)
```

```
## [1] "integer"
```

```
#cuantitativa discreta
```

```
class(películas$castWomenAmount)
```

```
## [1] "character"
```

```
#cuantitativa discreta
```

```
class(películas$castMenAmount)
```

```
## [1] "character"
```

```
###3 Distribución de variables cuantitativas
```

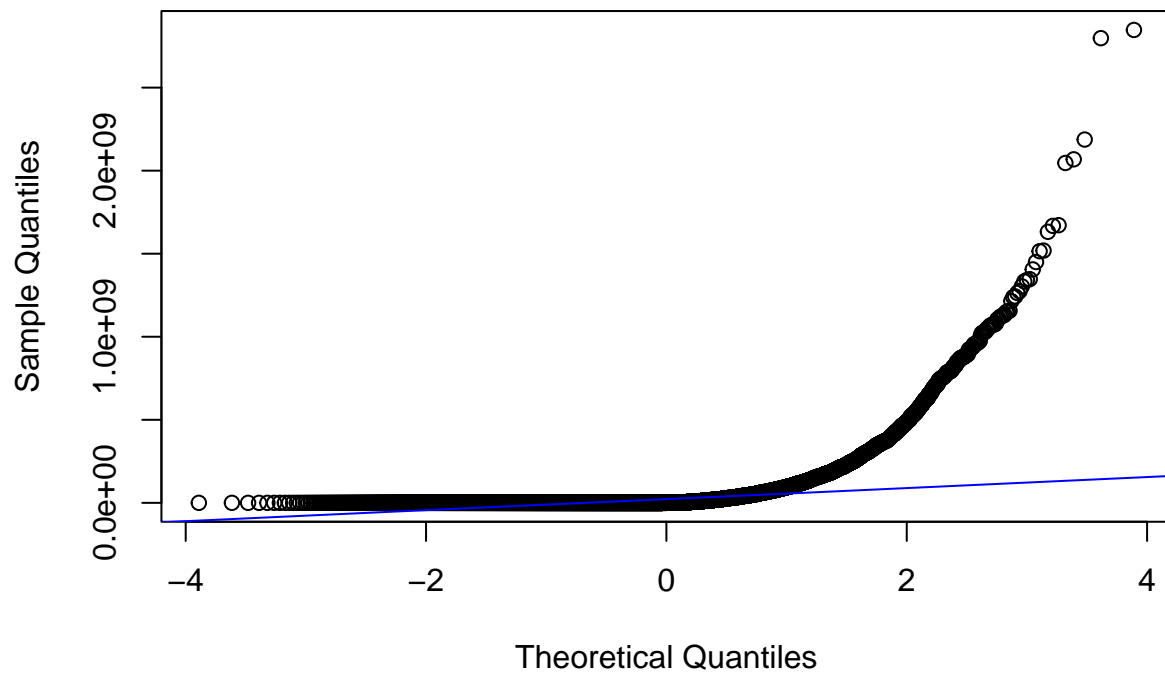
```
#revenue
```

```
library(nortest)
```

```
qqnorm(películas$revenue)
```

```
qqline(películas$revenue, col="blue")
```

Normal Q-Q Plot

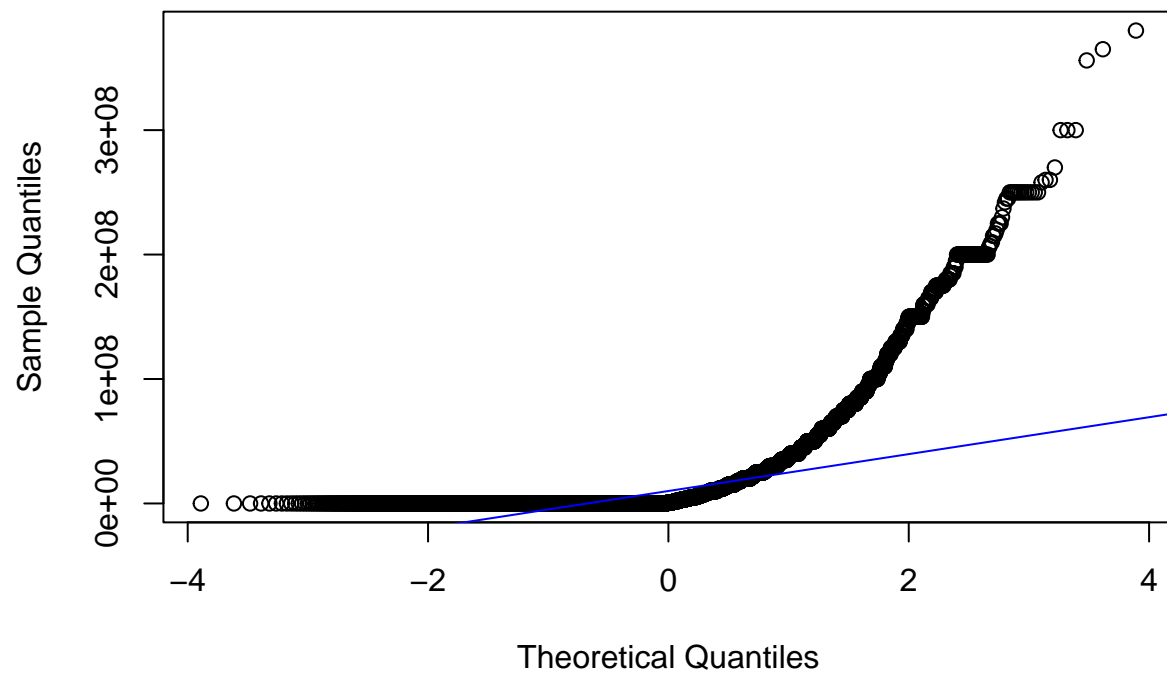


```
revenue_normal <- lillie.test(películas$revenue)
p_revenue <- revenue_normal$p.value
```

```
#Budget
```

```
library(nortest)
qqnorm(películas$budget)
qqline(películas$budget, col="blue")
```

Normal Q–Q Plot

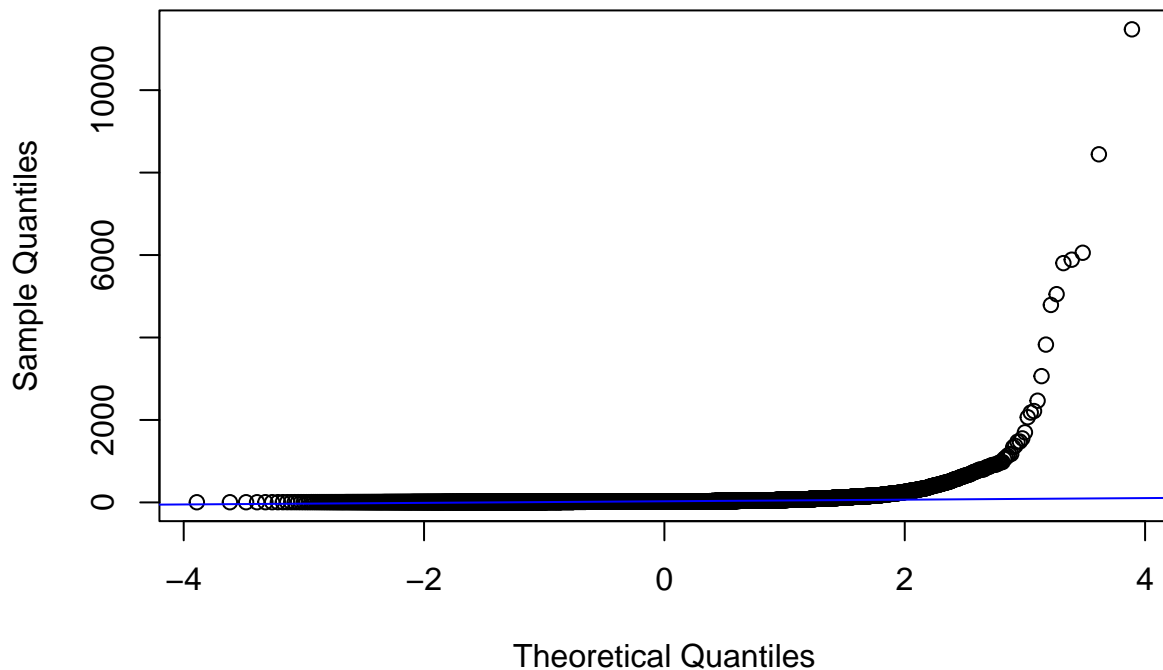


```
normal_budget <- lillie.test(películas$budget)
p_budget <- normal_budget$p.value
```

```
#popularity
```

```
library(nortest)
qqnorm(películas$popularity)
qqline(películas$popularity, col="blue")
```

Normal Q-Q Plot



```
normal_popularity <- lillie.test(peliculas$popularity)
p_popularity <- normal_popularity$p.value
```

4.1) Top 10 de películas con más presupuesto.

```
sortedByBudget <- peliculas[order(-peliculas$budget),]
top10 <- head(sortedByBudget, 10)
top10
```

##	id	budget	genres	homePage
## 717	1865	380000000	Adventure Action Fantasy	http://disney.go.com/pirates/index-on-stranger-tides.html#/video/
## 4711	99861	365000000	Action Adventure Science Fiction	http://marvel.com/movies/movie/193/avengers_age_of_ultron
## 5953	299534	356000000	Adventure Science Fiction Action	https://www.marvel.com/movies/avengers-endgame
## 164	285	300000000	Adventure Fantasy Action	http://disney.go.com/disneypictures/pirates
## 4954	141052	300000000	Action Adventure Fantasy Science Fiction	http://www.justiceleaguethemovie.com
## 5954	299536	300000000	Adventure Action Science Fiction	
## 608	1452	270000000	Science Fiction Action Adventure	
## 3792	38757	260000000	Animation Family	
## 7135	420818	260000000	Adventure Family Animation	
## 281	559	258000000	Fantasy Action Adventure	


```

## 5954      https://www.marvel.com/movies/avengers-infinity-war
## 608      https://www.warnerbros.com/movies/superman-returns
## 3792      http://disney.go.com/disneypictures/tangled/
## 7135      https://movies.disney.com/the-lion-king-2019
## 281      https://www.sonypictures.com/movies/spiderman3
##
## 717
## 4711
## 5953
## 164
## 4954 DC Comics|DC Entertainment|RatPac Entertainment|Cruel & Unusual Films|Atlas Entertainment|Lensbo
## 5954
## 608
## 3792
## 7135
## 281
##      productionCompanyCountry
## 717      US|US|GB
## 4711      US
## 5953      US
## 164      US|US|US
## 4954 US|US|US|US|US|US|US|US|US|CA
## 5954      US
## 608      US|US|US|US||
## 3792      US|US
## 7135      US|US
## 281      |US
##      productionCountry      revenue runtime video
## 717      United Kingdom|United States of America 1045713802      137 FALSE
## 4711      United States of America 1405403694      141 FALSE
## 5953      United States of America 2797800564      181 FALSE
## 164      United States of America 961000000      169 FALSE
## 4954 United Kingdom|Canada|United States of America 657924295      120 NA
## 5954      United States of America 2046239637      149 FALSE
## 608      United States of America 391081192      154 FALSE
## 3792      United States of America 592461732      100 NA
## 7135      United States of America 1667635327      118 FALSE
## 281      United States of America 890871626      139 FALSE
##      director
## 717      Rob Marshall
## 4711      Joss Whedon
## 5953      Anthony Russo|Joe Russo
## 164      Gore Verbinski
## 4954      Zack Snyder
## 5954      Anthony Russo|Joe Russo
## 608      Bryan Singer
## 3792      Byron Howard|Nathan Greno
## 7135      Jon Favreau
## 281      Sam Raimi
##
## 717
## 4711
## 5953
## 164

```

```

## 4954
## 5954
## 608
## 3792
## 7135
## 281 Tobey Maguire|Kirsten Dunst|James Franco|Thomas Haden Church|Topher Grace|Bryce Dallas Howard|R
##
## 717
## 4711
## 5953 36.077|31.514|12.766|60.833|53.109|26.075|14.289|40.686|43.886|18.701|25.641|110.474|
## 164
## 4954
## 5954
## 608
## 3792
## 7135
## 281 34.576|30.09|9.657|12.909|10.091|20.981|13.669|26.096|12.599|14.916|13.371|5.538|7.354|18.66|12
##
## 717
## 4711
## 5953
## 164
## 4954
## 5954
## 608
## 3792
## 7135
## 281 Spider-Man / Peter Parker|Mary Jane Watson|New Goblin / Harry Osborn|Sandman / Flint Marko|Venom
##
## 717 Pirates of the Caribbean: On Stranger Tides
## 4711 Avengers: Age of Ultron
## 5953 Avengers: Endgame
## 164 Pirates of the Caribbean: At World's End
## 4954 Justice League
## 5954 Avengers: Infinity War
## 608 Superman Returns
## 3792 Tangled
## 7135 The Lion King
## 281 Spider-Man 3
##
## title originalLanguage popularity
## 717 Pirates of the Caribbean: On Stranger Tides en 236.804
## 4711 Avengers: Age of Ultron en 303.290
## 5953 Avengers: Endgame en 422.152
## 164 Pirates of the Caribbean: At World's End en 159.100
## 4954 Justice League en 128.558
## 5954 Avengers: Infinity War en 648.960
## 608 Superman Returns en 43.011
## 3792 Tangled en 136.218
## 7135 The Lion King en 173.761
## 281 Spider-Man 3 en 615.655
##
## releaseDate voteAvg voteCount genresAmount productionCoAmount
## 717 2011-05-14 6.5 11580 3 3
## 4711 2015-04-22 7.3 19219 3 1
## 5953 2019-04-24 8.3 20186 3 1

```

```
## 164    2007-05-19      7.2    11650          3          3
## 4954    2017-11-15      6.2    11081          4         10
## 5954    2018-04-25      8.3    23715          3          1
## 608     2006-06-28      5.7     3336          3          6
## 3792    2010-11-24      7.6     9129          2          2
## 7135    2019-07-12      7.1     8203          3          2
## 281     2007-05-01      6.3    11137          3          2
##      productionCountriesAmount actorsAmount castWomenAmount castMenAmount
## 717              2              84              16              44
## 4711             1              72              12              33
## 5953             1             103              28              62
## 164              1              38               6              30
## 4954             3             102              23              37
## 5954             1              65              21              43
## 608              1              35              12              19
## 3792             1              36              11              19
## 7135             1              20               6              13
## 281             1             123              47              37
```

###4.2: Top 10 películas con más ingresos

```
sortedByGross <- peliculas[order(-peliculas$revenue),]
top10grossing<- head(sortedByGross, 10)
top10grossing
```

```
##      id    budget      genres
## 3211  19995 237000000 Action|Adventure|Fantasy|Science Fiction
## 5953 299534 356000000      Adventure|Science Fiction|Action
## 308    597 200000000      Drama|Romance
## 4948 140607 245000000 Action|Adventure|Science Fiction|Fantasy
## 5954 299536 300000000      Adventure|Action|Science Fiction
## 4915 135397 150000000 Action|Adventure|Science Fiction|Thriller
## 7135 420818 260000000      Adventure|Family|Animation
## 9050 634649 200000000      Action|Adventure|Science Fiction
## 3398  24428 220000000      Science Fiction|Action|Adventure
## 5088 168259 190000000      Action|Thriller|Crime|Adventure
##      homePage
## 3211      http://www.avatarmovie.com/
## 5953      https://www.marvel.com/movies/avengers-endgame
## 308      <NA>
## 4948      http://www.starwars.com/films/star-wars-episode-vii
## 5954      https://www.marvel.com/movies/avengers-infinity-war
## 4915      http://www.jurassicworld.com/
## 7135      https://movies.disney.com/the-lion-king-2019
## 9050      https://www.spidermannowayhome.movie
## 3398      http://marvel.com/avengers_movie/
## 5088      http://www.furious7.com/
##
## 3211
## 5953
## 308
## 4948
## 5954
## 4915
## 7135
```

```

## 9050
## 3398
## 5088 Abu Dhabi Film Commission|Universal Pictures|China Film Co.|Original Film|Fuji Television Network
##      productionCompanyCountry
## 3211          US|US|US|GB
## 5953          US
## 308          US|US|US
## 4948          US|US
## 5954          US
## 4915          US|US|US|JP|US
## 7135          US|US
## 9050          US|US|US
## 3398          US
## 5088 |US|CN|US|JP|JP|JP|US|US
##                                     productionCountry
## 3211          United States of America|United Kingdom
## 5953          United States of America
## 308          United States of America
## 4948          United States of America
## 5954          United States of America
## 4915          United States of America
## 7135          United States of America
## 9050          United States of America
## 3398          United States of America
## 5088 Canada|China|Japan|United Arab Emirates|United States of America
##      revenue runtime video          director
## 3211 2847246203      162 FALSE          James Cameron
## 5953 2797800564      181 FALSE Anthony Russo|Joe Russo
## 308  2187463944      194 FALSE          James Cameron
## 4948 2068223624      136  NA          J.J. Abrams
## 5954 2046239637      149 FALSE Anthony Russo|Joe Russo
## 4915 1671713208      124  NA          Colin Trevorrow
## 7135 1667635327      118 FALSE          Jon Favreau
## 9050 1631853496      148 FALSE          Jon Watts
## 3398 1518815515      143 FALSE          Joss Whedon
## 5088 1515047671      137 FALSE          James Wan
##
## 3211
## 5953
## 308
## 4948
## 5954
## 4915
## 7135
## 9050
## 3398 Robert Downey Jr.|Chris Evans|Mark Ruffalo|Chris Hemsworth|Scarlett Johansson|Jeremy Renner|Tom
## 5088
##
## 3211
## 5953
## 308
## 4948 24.542|11.881|14.434|18.836|20.734|3.588|40.852|10.079|15.329|21.788|5.811|5.965|1.288|8.914|2.
## 5954
## 4915

```

```

## 7135
## 9050
## 3398
## 5088
##
## 3211
## 5953
## 308
## 4948 Han Solo|Luke Skywalker|General Leia Organa|Kylo Ren / Ben Solo|Rey|Finn|Poe Dameron|Maz Kanata
## 5954
## 4915
## 7135
## 9050
## 3398
## 5088
##
##              originalTitle              title originalLanguage
## 3211              Avatar              Avatar              en
## 5953      Avengers: Endgame      Avengers: Endgame              en
## 308              Titanic              Titanic              en
## 4948 Star Wars: The Force Awakens Star Wars: The Force Awakens              en
## 5954      Avengers: Infinity War      Avengers: Infinity War              en
## 4915      Jurassic World      Jurassic World              en
## 7135      The Lion King      The Lion King              en
## 9050      Spider-Man: No Way Home      Spider-Man: No Way Home              en
## 3398      The Avengers      The Avengers              en
## 5088      Furious 7      Furious 7              en
##
##      popularity releaseDate voteAvg voteCount genresAmount productionCoAmount
## 3211      558.792   2009-12-10    7.5    24766           4              4
## 5953      422.152   2019-04-24    8.3    20186           3              1
## 308      186.300   1997-11-18    7.9    20447           2              3
## 4948      105.189   2015-12-15    7.3    16616           4              2
## 5954      648.960   2018-04-25    8.3    23715           3              1
## 4915      152.090   2015-06-06    6.7    17604           4              5
## 7135      173.761   2019-07-12    7.1     8203           3              2
## 9050     8443.740   2021-12-15    8.5     5630           3              3
## 3398      431.253   2012-04-25    7.7    26215           3              1
## 5088      133.032   2015-04-01    7.3     8903           4             10
##
##      productionCountriesAmount actorsAmount castWomenAmount castMenAmount
## 3211              2              37              9              21
## 5953              1             103             28             62
## 308              1             116             27             59
## 4948              1             188             24             74
## 5954              1              65             21             43
## 4915              1              48             13             31
## 7135              1              20              6             13
## 9050              1              59             13             33
## 3398              1             113             25             74
## 5088              5              57             15             28

```

###4.3: pelicula con mas votos

```

sorted <- peliculas[order(-peliculas$voteCount),]
top10votos <- head(sorted, 10)
top10votos

```

##	id	budget	genres
## 3512	27205	160000000	Action Science Fiction Adventure
## 5028	157336	165000000	Adventure Drama Science Fiction
## 88	155	185000000	Drama Action Crime Thriller
## 3398	24428	220000000	Science Fiction Action Adventure
## 5890	293660	58000000	Action Adventure Comedy
## 3211	19995	237000000	Action Adventure Fantasy Science Fiction
## 4826	118340	170000000	Action Science Fiction Adventure
## 5954	299536	300000000	Adventure Action Science Fiction
## 276	550	63000000	Drama
## 375	680	8000000	Thriller Crime
##			homePage
## 3512			http://inceptionmovie.warnerbros.com/
## 5028			http://www.interstellarmovie.net/
## 88			https://www.warnerbros.com/movies/dark-knight/
## 3398			http://marvel.com/avengers_movie/
## 5890			http://www.foxmovies.com/movies/deadpool
## 3211			http://www.avatarmovie.com/
## 4826			http://marvel.com/guardians
## 5954			https://www.marvel.com/movies/avengers-infinity-war
## 276			http://www.foxmovies.com/movies/fight-club
## 375			https://www.miramax.com/movie/pulp-fiction/
##			
## 3512			Legendary Pictures
## 5028			Legendary Pictures
## 88			DC Comics Legendary Pictures Syncopy Isobel Griffiths DC Enter
## 3398			
## 5890			20th Century Fox Marvel Entertainment The Donners' Company
## 3211			Dune Entertainment Lightstorm Entertainment 20
## 4826			
## 5954			
## 276			Regency Enterprises Fox 2000 Pictures Taurus Film Atman Entertainment Knickerbocker Films 20th C
## 375			Mir
##			
##			productionCompanyCountry productionCountry
## 3512			US GB US United Kingdom United States of America
## 5028			US GB United Kingdom United States of America
## 88			US US GB GB US US United Kingdom United States of America
## 3398			US United States of America
## 5890			US US US US US United States of America
## 3211			US US US GB United States of America United Kingdom
## 4826			US United States of America
## 5954			US United States of America
## 276			US US DE US US US Germany United States of America
## 375			US US US United States of America
##			
##			revenue runtime video director
## 3512	825532764	148	FALSE Christopher Nolan
## 5028	701729206	169	FALSE Christopher Nolan
## 88	1004558444	152	NA Christopher Nolan
## 3398	1518815515	143	FALSE Joss Whedon
## 5890	783100000	108	FALSE Tim Miller
## 3211	2847246203	162	FALSE James Cameron
## 4826	772776600	121	FALSE James Gunn
## 5954	2046239637	149	FALSE Anthony Russo Joe Russo
## 276	100853753	139	FALSE David Fincher

```

## 375 214179088 154 FALSE Quentin Tarantino
##
## 3512
## 5028
## 88
## 3398 Robert Downey Jr.|Chris Evans|Mark Ruffalo|Chris Hemsworth|Scarlett Johansson|Jeremy Renner|Tom
## 5890
## 3211
## 4826
## 5954
## 276
## 375
##
## 3512
## 5028
## 88 24.438|18.486|10.157|22.203|12.213|23.931|38.042|11.857|11.617|4.054|10.348|18.185|5.847|8.498|
## 3398
## 5890
## 3211
## 4826
## 5954
## 276
## 375
##
## 3512
## 5028
## 88
## 3398 Tony Stark / Iron Man|Steve Rogers / Captain America|Bruce Banner / The Hulk|Thor Odinson|Natasha
## 5890
## 3211
## 4826
## 5954
## 276
## 375
##
## originalTitle title originalLanguage
## 3512 Inception Inception en
## 5028 Interstellar Interstellar en
## 88 The Dark Knight The Dark Knight en
## 3398 The Avengers The Avengers en
## 5890 Deadpool Deadpool en
## 3211 Avatar Avatar en
## 4826 Guardians of the Galaxy Guardians of the Galaxy en
## 5954 Avengers: Infinity War Avengers: Infinity War en
## 276 Fight Club Fight Club en
## 375 Pulp Fiction Pulp Fiction en
## popularity releaseDate voteAvg voteCount genresAmount productionCoAmount
## 3512 200.943 2010-07-15 8.4 30788 3 3
## 5028 149.536 2014-11-05 8.4 27644 3 3
## 88 99.315 2008-07-14 8.5 26690 4 6
## 3398 431.253 2012-04-25 7.7 26215 3 1
## 5890 253.248 2016-02-09 7.6 26178 3 5
## 3211 558.792 2009-12-10 7.5 24766 4 4
## 4826 94.730 2014-07-30 7.9 23806 3 1
## 5954 648.960 2018-04-25 8.3 23715 3 1

```

##	276	70.292	1999-10-15	8.4	23338	1	6
##	375	72.724	1994-09-10	8.5	22501	2	3
##		productionCountriesAmount	actorsAmount	castWomenAmount	castMenAmount		
##	3512		2	51	12		29
##	5028		2	35	9		26
##	88		2	134	13		70
##	3398		1	113	25		74
##	5890		1	35	12		22
##	3211		2	37	9		21
##	4826		1	75	16		40
##	5954		1	65	21		43
##	276		2	79	9		41
##	375		1	54	18		26

###4.4: Top 10 películas con menos votos

```
sorted10menos <- tail(sorted, 10)
sorted10menos
```

##	id	budget	genres
##	9894	859236	0 Drama
##	9912	869488	0 Adventure Animation Fantasy TV Movie Comedy
##	9926	874960	0 Music
##	9939	878231	0 TV Movie Thriller
##	9960	888111	0 Horror
##	9972	893246	0 Documentary
##	9991	912223	60 Comedy Action Adventure Science Fiction
##	9992	913651	0 Drama
##	9995	919590	0
##	9999	922162	0
##			homePage
##	9894		<NA>
##	9912		<NA>
##	9926		<NA>
##	9939		https://lifetimeuncorked.com/2021/09/24/the-secret-lives-of-college-freshman-2021-lifetime/
##	9960		<NA>
##	9972		<NA>
##	9991		<NA>
##	9992		<NA>
##	9995		<NA>
##	9999		https://www.netflix.com/title/81425229
##		productionCompany	productionCompanyCountry
##	9894		
##	9912	Frederator Networks Cartoon Network Studios	US US
##	9926	ODEX Private Limited	
##	9939	Johnson Production Group	CA
##	9960		
##	9972	NOMADICA	
##	9991	Jepchu Producciones	
##	9992	Malvalanda	ES
##	9995		
##	9999		
##		productionCountry	revenue runtime video
##	9894		0 0 FALSE
##	9912	United States of America	0 92 TRUE

## 9926	Japan	0	114	FALSE
## 9939		0	0	FALSE
## 9960	Honduras	0	0	FALSE
## 9972	Italy Mexico	0	44	FALSE
## 9991		0	0	TRUE
## 9992	Spain	0	0	FALSE
## 9995		0	0	TRUE
## 9999	United States of America	0	59	FALSE
##				
## 9894	Samora Smallwood Paula Brancati Emmanuel Kabongo Michelle Rambharose Kaleb Alexander Rachael Cr			
## 9912				
## 9926				
## 9939				
## 9960				
## 9972				
## 9991				
## 9992				
## 9995				
## 9999				
##				
## 9894				
## 9912	Jeremy Shada John DiMaggio Olivia Olson Ava Acres Hynden Walch Steve Little Billy Brown Rebecca			
## 9926				
## 9939				
## 9960				
## 9972				
## 9991	Jes\xfas Zavala Yotzmit Ramirez Milton Wolch Negas Weena Mercator Jes\xfas Zavala Jes\xfas			
## 9992				
## 9995				
## 9999				
##				
## 9894	Lila DeMarco Mary Malone Rand Tollson Isabelle Eric DeMarco Eleanor Britt Mrs. Tollson Mr. Toll			
## 9912			9.444 5.905 7.076 11.72 7.081 5.219 7.105 16.304	
## 9926				
## 9939				16.688 7.33 4.935 3.
## 9960				
## 9972				
## 9991				0.6 0.6 0.6 0
## 9992				
## 9995				
## 9999				
##				
## 9894				
## 9912	Finn (voice) Jake (voice) Marceline (voice) Young Marceline (voice) Bonnibel (voice) Peppermint			
## 9926				
## 9939				
## 9960				
## 9972				
## 9991				Iron Man (voice) Pepper Potts (v
## 9992				
## 9995				
## 9999				
##				
## 9894			originalTitle	
			Death She Wrote	

```

## 9912 Adventure Time: Stakes
## 9926 Kimetsu Orchestra Concert
## 9939 The Secret Lives of College Freshmen
## 9960 El Regreso de La Llorona
## 9972 El tiempo del no tiempo
## 9991 Los Vengadores Chiflados
## 9992 Sexo expl\xedcito
## 9995 Phantastische Harry Potter Momente - Das gro\xdf SAT.1 Spezial
## 9999 The Witcher: Fireplace
## title
## 9894 Death She Wrote
## 9912 Adventure Time: Stakes
## 9926 Kimetsu Orchestra Concert
## 9939 The Secret Lives of College Freshmen
## 9960 The Return of La Llorona
## 9972 El tiempo del no tiempo
## 9991 Los Vengadores Chiflados
## 9992 Sexo expl\xedcito
## 9995 Phantastische Harry Potter Momente - Das gro\xdf SAT.1 Spezial
## 9999 The Witcher: Fireplace
## originalLanguage popularity releaseDate voteAvg voteCount genresAmount
## 9894 en 17.543 2021-08-06 2 1 1
## 9912 en 21.719 2015-11-16 7 1 5
## 9926 ja 43.191 2021-08-25 2 1 1
## 9939 en 66.698 2021-09-24 2 1 2
## 9960 en 17.737 2021-10-01 2 1 1
## 9972 es 26.540 2013-01-01 2 1 1
## 9991 es 43.452 2014-01-03 10 1 4
## 9992 es 28.548 2013-06-13 2 1 1
## 9995 de 38.653 2021-10-31 9 1 0
## 9999 en 9.664 2021-12-17 6 1 1
## productionCoAmount productionCountriesAmount actorsAmount
## 9894 11 5 5
## 9912 2 1 14
## 9926 1 1 0
## 9939 1 10 7
## 9960 1 0 0
## 9972 1 2 0
## 9991 1 12 0
## 9992 1 1 2
## 9995 0 0 919590
## 9999 0 0 0
## castWomenAmount
## 9894 859236
## 9912 4
## 9926 0
## 9939 2
## 9960 0
## 9972 0
## 9991 8
## 9992 1
## 9995 Phantastische Harry Potter Momente - Das gro\xdf SAT.1 Spezial
## 9999 922162
## castMenAmount

```

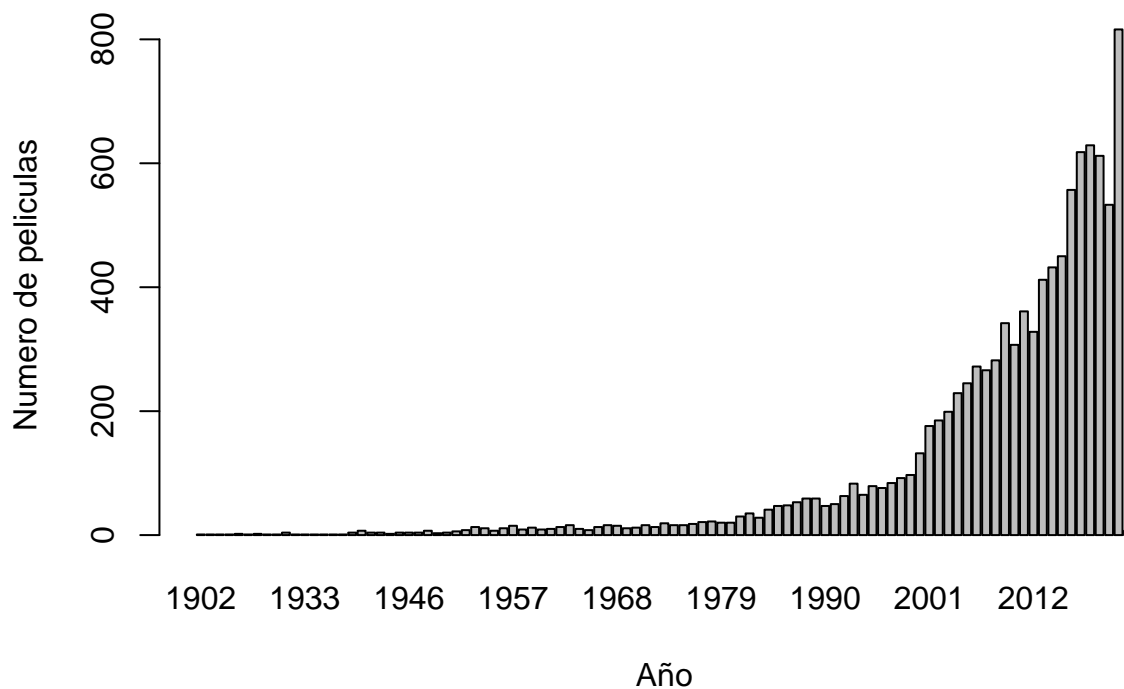
```
## 9894                                Death She Wrote
## 9912                                9
## 9926                                0
## 9939                                878231
## 9960                                888111
## 9972                                0
## 9991                                912223
## 9992                                0
## 9995 Phantastische Harry Potter Momente - Das gro\xdfte SAT.1 Spezial
## 9999                                The Witcher: Fireplace
```

###4.5: ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras

```
peliculas$releaseDate = as.Date(peliculas$releaseDate, "%Y-%m-%d")
peliculas$year = format(peliculas$releaseDate, "%Y")
peliculasPoraño = table(peliculas$year)
peliculasPoraño
```

```
##
## 1902 1920 1921 1922 1925 1926 1927 1929 1930 1931 1932 1933 1935 1936 1937 1938
##    1    1    1    1    2    1    2    1    1    4    1    1    1    1    1    1
## 1939 1940 1941 1942 1943 1944 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955
##    4    7    4    4    2    4    4    4    7    3    4    6    8    13    11    7
## 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971
##   11   15    9   12    9   10   13   16   10    8   13   16   15   11   12   16
## 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987
##   13   19   16   16   18   21   22   20   20   30   35   28   41   47   48   53
## 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003
##   59   59   47   50   63   83   65   79   76   84   92   97  132  176  185  199
## 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
##  229  245  272  266  282  342  307  361  328  412  432  450  557  618  629  612
## 2020 2021 2022
##  533  816    7
```

```
barplot(peliculasPoraño, xlab = "Año", ylab = "Numero de peliculas")
```



###4.6: ¿Cuál es el género principal de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Representélo usando un gráfico

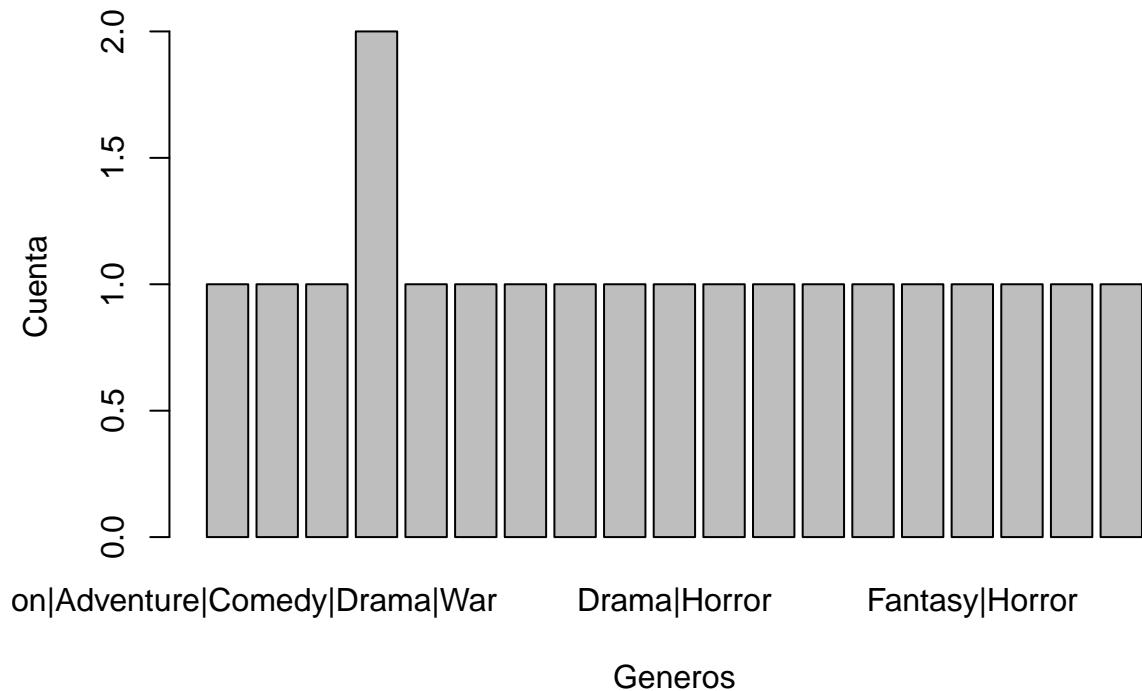
```

peliculasOrd <- peliculas[order(peliculas$releaseDate),]
peliculasRec <- peliculasOrd[1:20, ]
peliculasRec <- do.call(rbind, strsplit(as.character(peliculasRec$genres), ", "))
peliculasFreq = names(sort(table(peliculasRec), increasing=TRUE))[1]
barplot(table(peliculasRec), xlab="Generos", ylab="Cuenta", main=paste("Distribucion de los generos en "))

```

Distribucion de los generos en las 20 películas más recientes

Genero predominante: Action|Adventure|Comedy|Drama|War



###4.7: ¿Las películas de qué genero principal obtuvieron mayores ganancias?

```
top_5_g <- sortedByGross[1:5, ]
cat("El top de películas por generos es:", "\n")
```

El top de películas por generos es:

```
for (i in 1:5) {
  cat(top_5_g$title[i], ":", top_5_g$genres[i], "\n")
}
```

```
## Avatar : Action|Adventure|Fantasy|Science Fiction
## Avengers: Endgame : Adventure|Science Fiction|Action
## Titanic : Drama|Romance
## Star Wars: The Force Awakens : Action|Adventure|Science Fiction|Fantasy
## Avengers: Infinity War : Adventure|Action|Science Fiction
```

###4.8: ¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?

```
actores <- peliculas[order(-peliculas$actorsAmount),]
cut_thing <- subset(actores, select = c("id", "revenue", "actorsAmount", "releaseDate"))
head(cut_thing, 20)
```

```
##      id revenue actorsAmount releaseDate
## 9995 919590      0      919590 2021-10-31
## 9954 882878      0      882878 2021-04-11
## 9876 853253      0      853253 2006-03-21
## 9846 840964      0      840964 2006-12-31
```

```
## 9807 825491      0      825491 2017-01-01
## 9768 815762      0      815762 2018-12-31
## 9707 796822      0      796822 2008-07-01
## 9656 784594      0      784594 2021-01-07
## 9578 765207      0      765207 2009-09-15
## 9525 750209      0      750209 2020-10-02
## 9521 748501      0      748501 2020-09-04
## 9484 738562      0      738562 2017-07-13
## 9303 687920      0      687920 2004-06-18
## 9229 668297     27      668297 2014-12-06
## 9085 641026      0      641026 2016-11-16
## 9084 640103      0      640103 2019-10-07
## 9083 640100      0      640100 2018-02-12
## 9082 640097      0      640097 2017-07-17
## 9049 634646      0      634646 2019-09-28
## 8955 619022      0      619022 2015-02-03
```

#podemos ver que la revenue no influye si hay más o menos actores.

####4.10: ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

```
peliculas_na <- peliculas[!is.na(peliculas$voteCount), ]

peliculas_ord <- peliculas[order(-peliculas$voteCount), ]

# select the top 20 movies along with their directors
top_20_movies <- peliculas_ord[1:20, c("title", "director")]

# print the result
print(top_20_movies)
```

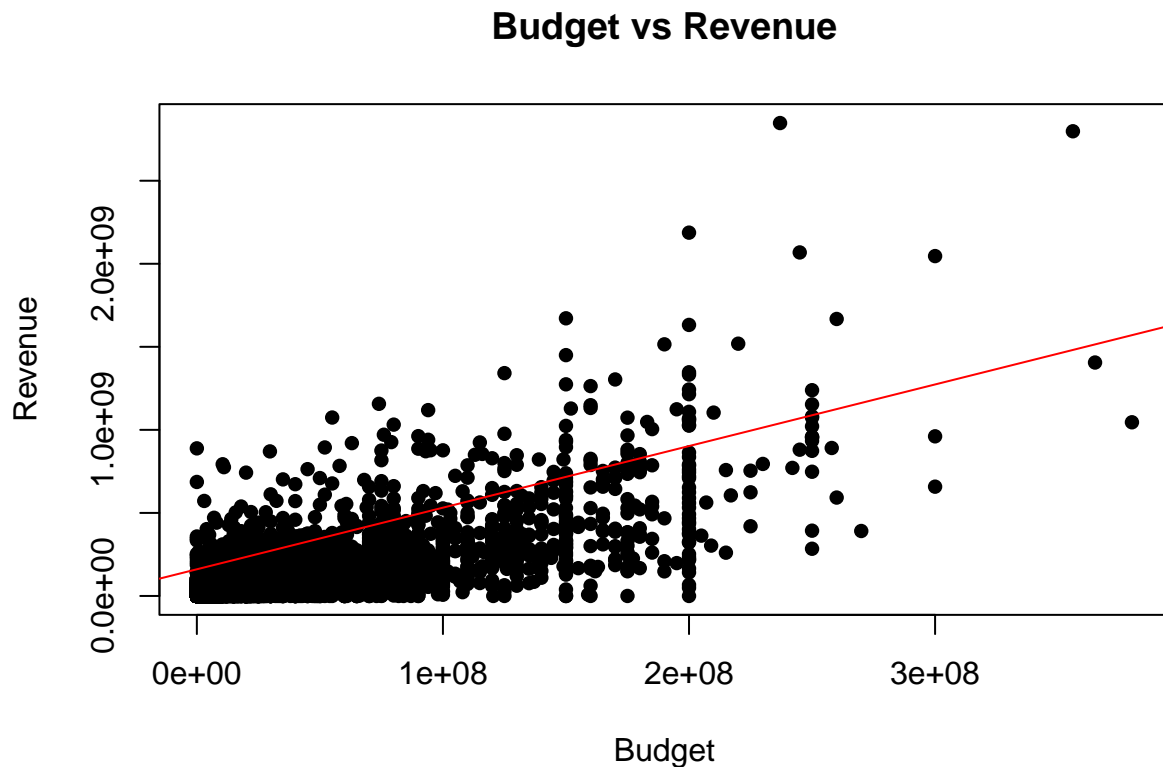
```
##                                title
## 3512                        Inception
## 5028                      Interstellar
## 88                        The Dark Knight
## 3398                      The Avengers
## 5890                      Deadpool
## 3211                      Avatar
## 4826    Guardians of the Galaxy
## 5954    Avengers: Infinity War
## 276                        Fight Club
## 375                      Pulp Fiction
## 696                      Iron Man
## 5                        Forrest Gump
## 368    Harry Potter and the Philosopher's Stone
## 4338                      Django Unchained
## 314                      The Matrix
## 160    The Shawshank Redemption
## 308                      Titanic
## 67    The Lord of the Rings: The Fellowship of the Ring
## 5953                      Avengers: Endgame
## 7765                      Joker
##                                director
## 3512    Christopher Nolan
## 5028    Christopher Nolan
```

```
## 88 Christopher Nolan
## 3398 Joss Whedon
## 5890 Tim Miller
## 3211 James Cameron
## 4826 James Gunn
## 5954 Anthony Russo|Joe Russo
## 276 David Fincher
## 375 Quentin Tarantino
## 696 Jon Favreau
## 5 Robert Zemeckis
## 368 Chris Columbus
## 4338 Quentin Tarantino
## 314 Lilly Wachowski|Lana Wachowski
## 160 Frank Darabont
## 308 James Cameron
## 67 Peter Jackson
## 5953 Anthony Russo|Joe Russo
## 7765 Todd Phillips
```

###4.11: ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos?

```
reg <- lm(peliculas$budget*20 ~ peliculas$revenue**2, data = peliculas)
```

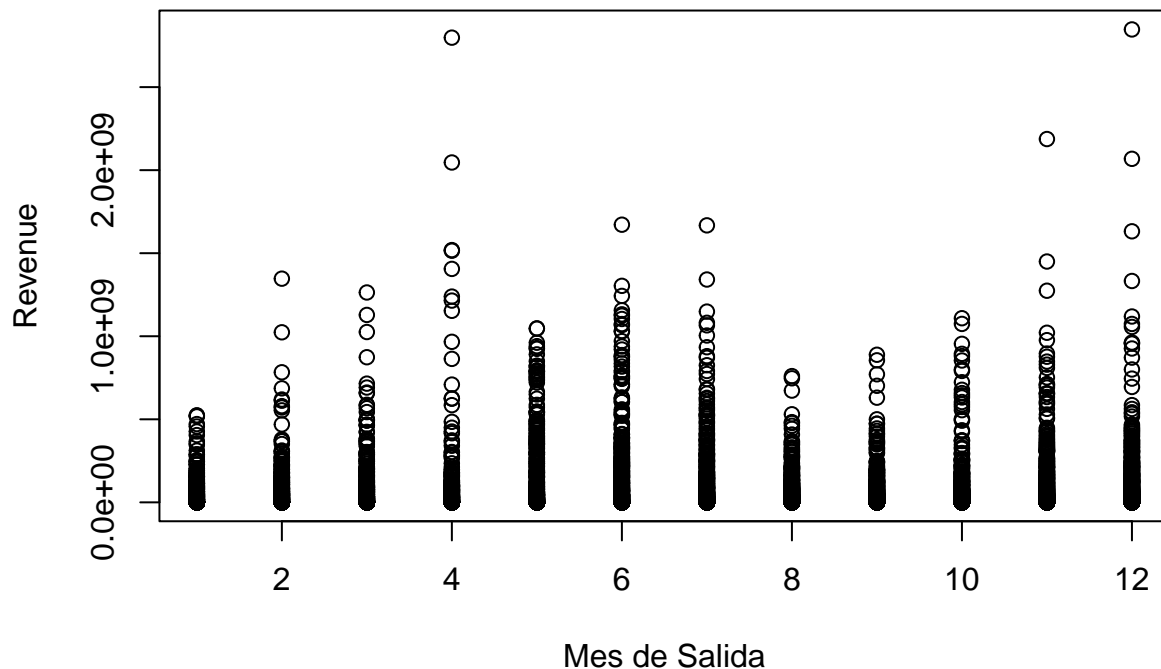
```
plot(peliculas$budget, peliculas$revenue, main="Budget vs Revenue", xlab="Budget", ylab="Revenue", pch=
abline(reg, col = "red")
```



Podemos observar que si se mantiene una ganancia en crecimiento hay un aumento en el presupuesto.

###4.12: ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?

```
peliculas$mesSalida <- as.numeric(format(as.Date(peliculas$releaseDate, format = "%Y-%m-%d"), "%m"))
plot(peliculas$mesSalida, peliculas$revenue, xlab="Mes de Salida", ylab="Revenue")
```



###4.13: ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿cuántas películas, en promedio, se han lanzado por mes?

```
peliculas$mesSalida <- as.numeric(format(as.Date(peliculas$releaseDate, format = "%Y-%m-%d"), "%m"))
revenue_mes <- tapply(peliculas$revenue, peliculas$mesSalida, sum )
mes_mas_rev <- which.max(revenue_mes)
cat("El mes con más revenue es:", mes_mas_rev, "\n")
```

```
## El mes con más revenue es: 6
```

```
cat("Este revenue es:", revenue_mes[mes_mas_rev], "\n")
```

```
## Este revenue es: 77597881637
```

```
peliculas_mensuales <- table(peliculas$mesSalida)
promedio <- mean(peliculas_mensuales)
cat("El numero promedio de peliculas que son lanzadas por mes es:", promedio, "\n")
```

```
## El numero promedio de peliculas que son lanzadas por mes es: 833.3333
```

###4.15: ¿A qué género principal pertenecen las películas más largas?

```
peliculasRun <- peliculas[order(-peliculas$runtime), ]
top_5 <- peliculasRun[1:5, ]
```



```

# print the result
cat("Los generos con las peliculas más largas son:", "\n")

## Los generos con las peliculas más largas son:
for (i in 1:5) {
  cat(top_5$title[i], ":", top_5$genres[i], "\n")
}

## How Ponyo was Born ~Hayao Miyazaki's Thought Process~ : Documentary
## Crystal Lake Memories: The Complete History of Friday the 13th : Documentary
## Napoleon : Drama|History|War
## 1900 : Drama|History
## Cleopatra : Drama|History|Romance

```