

# Limpieza y análisis de datos del Titanic

Autores: Saúl Santomé Rúa y Víctor Elías Afonso Rodríguez

Mayo 2021

```
# Cargamos los paquetes R que vamos a usar
library(ggplot2)
library(dplyr)
library(kableExtra)
# Cargamos el fichero de datos
titanicData <- read.csv('../dataset/titanicData.csv', stringsAsFactors = FALSE)
filas=dim(titanicData)[1]
columnas=dim(titanicData)[2]
```

## 1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

Este *dataset* es interesante porque permite analizar los diferentes perfiles que pudieron acceder al gran barco del momento y además da la posibilidad de localizar que variables permitieron que algunos pasajeros sobrevivieran o no. Consideramos que es un análisis relevante porque se pueden identificar los factores que separaron a los pasajeros de la vida o la muerte y esto nos lleva a entender mejor como era la sociedad de aquella época.

Vamos a visualizar la estructura del conjunto de datos.

```
## 'data.frame': 891 obs. of 12 variables:
## $ PassengerId: int 1 2 3 4 5 6 7 8 9 10 ...
## $ Survived : int 0 1 1 1 0 0 0 0 1 1 ...
## $ Pclass : int 3 1 3 1 3 3 1 3 3 2 ...
## $ Name : chr "Braund, Mr. Owen Harris" "Cumings, Mrs. John Bradley (Florence Briggs Thayer)"
## $ Sex : chr "male" "female" "female" "female" ...
## $ Age : num 22 38 26 35 35 NA 54 2 27 14 ...
## $ SibSp : int 1 1 0 1 0 0 0 3 0 1 ...
## $ Parch : int 0 0 0 0 0 0 0 1 2 0 ...
## $ Ticket : chr "A/5 21171" "PC 17599" "STON/O2. 3101282" "113803" ...
## $ Fare : num 7.25 71.28 7.92 53.1 8.05 ...
## $ Cabin : chr "" "C85" "" "C123" ...
## $ Embarked : chr "S" "C" "S" "S" ...
```

Este juego de datos está compuesto por **891 observaciones** y por **12 variables**. Los datos se encuentran en inglés.

Las **variables** son:

Table 1: Variables cualitativas, tipos interpretados

	x
Name	character
Sex	character
Ticket	character
Cabin	character
Embarked	character

```
names(titanicData)
```

```
## [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"
## [6] "Age"          "SibSp"       "Parch"       "Ticket"      "Fare"
## [11] "Cabin"        "Embarked"
```

- **PassengerId:** Es un número que corresponde con el código de cada pasajero.
- **Survived:** Indica si el pasajero sobrevivió o no. Los supervivientes se indica con un “1” y los no supervivientes con un “0”.
- **Pclass:** Indica la clase en la que se encuentra el pasajero.
- **Name:** Es el nombre.
- **Sex:** Indica si es masculino o femenino.
- **Age:** Corresponde con la edad.
- **SibSp:** Es el número de cónyuges/hermanos que viajan con el pasajero.
- **Parch:** Es el número de padres/niños que viajan junto al pasajero.
- **Ticket:** Número de ticket.
- **Fare:** Precio que ha pagado.
- **Cabin:** Número de cabina.
- **Embarked:** Puerto de embarque.

Vamos a dividir estas variables en cuantitativas y cualitativas según lo ha leído RStudio.

Las **variables cualitativas** son las siguientes:

Las **variables cuantitativas** son:

```
idQuantitative<-which(names=="PassengerId" | names=="Survived" | names=="Pclass" | names=="Age" | names=="SibSp" |
names=="Parch" | names=="Ticket" | names=="Fare")
kable( sapply(titanicData[idQuantitative],class),caption="Variables cualitativas, tipos interpretados")
```

Vamos a ver **datos estadísticos** de cada variable cuantitativa para entender mejor cada una:

Table 2: Variables cualitativas, tipos interpretados

	x
Survived	integer
Pclass	integer
Age	numeric
Parch	integer
Fare	numeric

```
titanicData<-titanicData[,-c(12)]
titanicData<-titanicData[,-c(11)]
titanicData<-titanicData[,-c(9)]
titanicData<-titanicData[,-c(5)]
titanicData<-titanicData[,-c(4)]
summary(titanicData)
```

```
## PassengerId      Survived      Pclass         Age
## Min.   : 1.0      Min.   :0.0000   Min.   :1.000   Min.   : 0.42
## 1st Qu.:223.5     1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:20.12
## Median :446.0     Median :0.0000   Median :3.000   Median :28.00
## Mean   :446.0     Mean   :0.3838   Mean   :2.309   Mean   :29.70
## 3rd Qu.:668.5     3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:38.00
## Max.   :891.0     Max.   :1.0000   Max.   :3.000   Max.   :80.00
##                                     NA's   :177
## SibSp      Parch      Fare
## Min.   :0.000   Min.   :0.0000   Min.   : 0.00
## 1st Qu.:0.000   1st Qu.:0.0000   1st Qu.: 7.91
## Median :0.000   Median :0.0000   Median :14.45
## Mean   :0.523   Mean   :0.3816   Mean   :32.20
## 3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:31.00
## Max.   :8.000   Max.   :6.0000   Max.   :512.33
##
```

```
titanicData <- read.csv('../dataset/titanicData.csv',stringsAsFactors = FALSE)
```

Gracias a este resumen podemos tener una visión del conjunto de los datos. Podemos determinar:

- **PassengerID**: Funciona como el número de identificación de cada pasajero. Lo que significa que cada pasajero tiene un número de identificación, este número de identificación es una variable prescindible ya que el valor que puede aportar es un valor en base a como asignaron cada número a cada pasajero, con lo cual no enriquece el análisis.
- **Survived**: Tiene un comportamiento binario.
- **Age**: Nos indica que hay anomalías ya que el valor mínimo es inferior a 1.
- **Parch**: Observamos que el valor oscila entre 0 y 6.
- **Fare**: Muestra que hay pasajeros que han entrado gratis y la mayor cuantía que se ha pagado.

Una vez que conocemos mejor las variables vamos a visualizar los valores que toma cada **variable cualitativa**:

Name

```
table(titanicData$Name)
```

##		
##	Abbing, Mr. Anthony	
##		1
##	Abbott, Mr. Rossmore Edward	
##		1
##	Abbott, Mrs. Stanton (Rosa Hunt)	
##		1
##	Abelson, Mr. Samuel	
##		1
##	Abelson, Mrs. Samuel (Hannah Witosky)	
##		1
##	Adahl, Mr. Mauritz Nils Martin	
##		1
##	Adams, Mr. John	
##		1
##	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	
##		1
##	Aks, Mrs. Sam (Leah Rosen)	
##		1
##	Albimona, Mr. Nassef Cassem	
##		1
##	Alexander, Mr. William	
##		1
##	Alhomaki, Mr. Ilmari Rudolf	
##		1
##	Ali, Mr. Ahmed	
##		1
##	Ali, Mr. William	
##		1
##	Allen, Miss. Elisabeth Walton	
##		1
##	Allen, Mr. William Henry	
##		1
##	Allison, Master. Hudson Trevor	
##		1
##	Allison, Miss. Helen Loraine	
##		1
##	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	
##		1
##	Allum, Mr. Owen George	
##		1
##	Andersen-Jensen, Miss. Carla Christine Nielsine	
##		1
##	Anderson, Mr. Harry	
##		1
##	Andersson, Master. Sigvard Harald Elias	
##		1
##	Andersson, Miss. Ebba Iris Alfrida	
##		1
##	Andersson, Miss. Ellis Anna Maria	
##		1
##	Andersson, Miss. Erna Alexandra	
##		1
##	Andersson, Miss. Ingeborg Constanzia	

##		1
##	Andersson, Miss. Sigrid Elisabeth	
##		1
##	Andersson, Mr. Anders Johan	
##		1
##	Andersson, Mr. August Edvard ("Wennerstrom")	
##		1
##	Andersson, Mrs. Anders Johan (Alfrida Konstantia Brogren)	
##		1
##	Andreasson, Mr. Paul Edvin	
##		1
##	Andrew, Mr. Edgardo Samuel	
##		1
##	Andrews, Miss. Kornelia Theodosia	
##		1
##	Andrews, Mr. Thomas Jr	
##		1
##	Angle, Mrs. William A (Florence "Mary" Agnes Hughes)	
##		1
##	Appleton, Mrs. Edward Dale (Charlotte Lamson)	
##		1
##	Arnold-Franchi, Mr. Josef	
##		1
##	Arnold-Franchi, Mrs. Josef (Josefine Franchi)	
##		1
##	Artagaveytia, Mr. Ramon	
##		1
##	Asim, Mr. Adola	
##		1
##	Asplund, Master. Clarence Gustaf Hugo	
##		1
##	Asplund, Master. Edvin Rojj Felix	
##		1
##	Asplund, Miss. Lillian Gertrud	
##		1
##	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)	
##		1
##	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	
##		1
##	Attalah, Miss. Malake	
##		1
##	Attalah, Mr. Sleiman	
##		1
##	Aubart, Mme. Leontine Pauline	
##		1
##	Augustsson, Mr. Albert	
##		1
##	Ayoub, Miss. Banoura	
##		1
##	Backstrom, Mr. Karl Alfred	
##		1
##	Backstrom, Mrs. Karl Alfred (Maria Mathilda Gustafsson)	
##		1
##	Baclini, Miss. Eugenie	

##		1
##	Baclini, Miss. Helene Barbara	
##		1
##	Baclini, Miss. Marie Catherine	
##		1
##	Baclini, Mrs. Solomon (Latifa Qurban)	
##		1
##	Badt, Mr. Mohamed	
##		1
##	Bailey, Mr. Percy Andrew	
##		1
##	Balkic, Mr. Cerin	
##		1
##	Ball, Mrs. (Ada E Hall)	
##		1
##	Banfield, Mr. Frederick James	
##		1
##	Barah, Mr. Hanna Assi	
##		1
##	Barbara, Miss. Saiide	
##		1
##	Barbara, Mrs. (Catherine David)	
##		1
##	Barber, Miss. Ellen "Nellie"	
##		1
##	Barkworth, Mr. Algernon Henry Wilson	
##		1
##	Barton, Mr. David John	
##		1
##	Bateman, Rev. Robert James	
##		1
##	Baumann, Mr. John D	
##		1
##	Baxter, Mr. Quigg Edmond	
##		1
##	Baxter, Mrs. James (Helene DeLaudeniére Chaput)	
##		1
##	Bazzani, Miss. Albina	
##		1
##	Beane, Mr. Edward	
##		1
##	Beane, Mrs. Edward (Ethel Clarke)	
##		1
##	Beavan, Mr. William Thomas	
##		1
##	Becker, Master. Richard F	
##		1
##	Becker, Miss. Marion Louise	
##		1
##	Beckwith, Mr. Richard Leonard	
##		1
##	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	
##		1
##	Beesley, Mr. Lawrence	

##		1
##	Behr, Mr. Karl Howell	
##		1
##	Bengtsson, Mr. John Viktor	
##		1
##	Berglund, Mr. Karl Ivar Sven	
##		1
##	Berriman, Mr. William John	
##		1
##	Betros, Mr. Tannous	
##		1
##	Bidois, Miss. Rosalie	
##		1
##	Bing, Mr. Lee	
##		1
##	Birkeland, Mr. Hans Martin Monsen	
##		1
##	Bishop, Mr. Dickinson H	
##		1
##	Bishop, Mrs. Dickinson H (Helen Walton)	
##		1
##	Bissette, Miss. Amelia	
##		1
##	Bjornstrom-Steffansson, Mr. Mauritz Hakan	
##		1
##	Blackwell, Mr. Stephen Weart	
##		1
##	Blank, Mr. Henry	
##		1
##	Bonnell, Miss. Elizabeth	
##		1
##	Bostandyeff, Mr. Guentcho	
##		1
##	Boulos, Miss. Nourelain	
##		1
##	Boulos, Mr. Hanna	
##		1
##	Boulos, Mrs. Joseph (Sultana)	
##		1
##	Bourke, Miss. Mary	
##		1
##	Bourke, Mr. John	
##		1
##	Bourke, Mrs. John (Catherine)	
##		1
##	Bowen, Mr. David John "Dai"	
##		1
##	Bowerman, Miss. Elsie Edith	
##		1
##	Bracken, Mr. James H	
##		1
##	Bradley, Mr. George ("George Arthur Brayton")	
##		1
##	Braund, Mr. Lewis Richard	

##		1
##	Braund, Mr. Owen Harris	
##		1
##	Brewe, Dr. Arthur Jackson	
##		1
##	Brocklebank, Mr. William Alfred	
##		1
##	Brown, Miss. Amelia "Mildred"	
##		1
##	Brown, Mr. Thomas William Solomon	
##		1
##	Brown, Mrs. James Joseph (Margaret Tobin)	
##		1
##	Brown, Mrs. Thomas William Solomon (Elizabeth Catherine Ford)	
##		1
##	Bryhl, Mr. Kurt Arnold Gottfrid	
##		1
##	Burke, Mr. Jeremiah	
##		1
##	Burns, Miss. Elizabeth Margaret	
##		1
##	Buss, Miss. Kate	
##		1
##	Butler, Mr. Reginald Fenton	
##		1
##	Butt, Major. Archibald Willingham	
##		1
##	Byles, Rev. Thomas Roussel Davids	
##		1
##	Bystrom, Mrs. (Karolina)	
##		1
##	Cacic, Miss. Marija	
##		1
##	Cacic, Mr. Luka	
##		1
##	Cairns, Mr. Alexander	
##		1
##	Calderhead, Mr. Edward Pennington	
##		1
##	Caldwell, Master. Alden Gates	
##		1
##	Caldwell, Mrs. Albert Francis (Sylvia Mae Harbaugh)	
##		1
##	Calic, Mr. Jovo	
##		1
##	Calic, Mr. Petar	
##		1
##	Cameron, Miss. Clear Annie	
##		1
##	Campbell, Mr. William	
##		1
##	Canavan, Miss. Mary	
##		1
##	Cann, Mr. Ernest Charles	



##		1
##	Caram, Mrs. Joseph (Maria Elias)	
##		1
##	Carbines, Mr. William	
##		1
##	Cardeza, Mr. Thomas Drake Martinez	
##		1
##	Carlsson, Mr. August Sigfrid	
##		1
##	Carlsson, Mr. Frans Olof	
##		1
##	Carr, Miss. Helen "Ellen"	
##		1
##	Carrau, Mr. Francisco M	
##		1
##	Carter, Master. William Thornton II	
##		1
##	Carter, Miss. Lucile Polk	
##		1
##	Carter, Mr. William Ernest	
##		1
##	Carter, Mrs. Ernest Courtenay (Lilian Hughes)	
##		1
##	Carter, Mrs. William Ernest (Lucile Polk)	
##		1
##	Carter, Rev. Ernest Courtenay	
##		1
##	Cavendish, Mr. Tyrell William	
##		1
##	Celotti, Mr. Francesco	
##		1
##	Chaffee, Mr. Herbert Fuller	
##		1
##	Chambers, Mr. Norman Campbell	
##		1
##	Chambers, Mrs. Norman Campbell (Bertha Griggs)	
##		1
##	Chapman, Mr. Charles Henry	
##		1
##	Chapman, Mr. John Henry	
##		1
##	Charters, Mr. David	
##		1
##	Cherry, Miss. Gladys	
##		1
##	Chibnall, Mrs. (Edith Martha Bowerman)	
##		1
##	Chip, Mr. Chang	
##		1
##	Christmann, Mr. Emil	
##		1
##	Christy, Miss. Julie Rachel	
##		1
##	Chronopoulos, Mr. Apostolos	

##		1
##	Clarke, Mrs. Charles V (Ada Maria Winfield)	
##		1
##	Cleaver, Miss. Alice	
##		1
##	Clifford, Mr. George Quincy	
##		1
##	Coelho, Mr. Domingos Fernando	
##		1
##	Cohen, Mr. Gurshon "Gus"	
##		1
##	Coleff, Mr. Peju	
##		1
##	Coleff, Mr. Satio	
##		1
##	Coleridge, Mr. Reginald Charles	
##		1
##	Collander, Mr. Erik Gustaf	
##		1
##	Colley, Mr. Edward Pomeroy	
##		1
##	Collyer, Miss. Marjorie "Lottie"	
##		1
##	Collyer, Mr. Harvey	
##		1
##	Collyer, Mrs. Harvey (Charlotte Annie Tate)	
##		1
##	Compton, Miss. Sara Rebecca	
##		1
##	Connaghton, Mr. Michael	
##		1
##	Connolly, Miss. Kate	
##		1
##	Connors, Mr. Patrick	
##		1
##	Cook, Mr. Jacob	
##		1
##	Cor, Mr. Liudevit	
##		1
##	Corn, Mr. Harry	
##		1
##	Coutts, Master. Eden Leslie "Neville"	
##		1
##	Coutts, Master. William Loch "William"	
##		1
##	Coxon, Mr. Daniel	
##		1
##	Crease, Mr. Ernest James	
##		1
##	Cribb, Mr. John Hatfield	
##		1
##	Crosby, Capt. Edward Gifford	
##		1
##	Crosby, Miss. Harriet R	

##		1
##	Culumovic, Mr. Jeso	
##		1
##	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	
##		1
##	Cunningham, Mr. Alfred Fleming	
##		1
##	Dahl, Mr. Karl Edwart	
##		1
##	Dahlberg, Miss. Gerda Ulrika	
##		1
##	Dakic, Mr. Branko	
##		1
##	Daly, Mr. Eugene Patrick	
##		1
##	Daly, Mr. Peter Denis	
##		1
##	Danbom, Mr. Ernst Gilbert	
##		1
##	Danbom, Mrs. Ernst Gilbert (Anna Sigrid Maria Brogren)	
##		1
##	Daniel, Mr. Robert Williams	
##		1
##	Danoff, Mr. Yoto	
##		1
##	Dantcheff, Mr. Ristiu	
##		1
##	Davidson, Mr. Thornton	
##		1
##	Davies, Master. John Morgan Jr	
##		1
##	Davies, Mr. Alfred J	
##		1
##	Davies, Mr. Charles Henry	
##		1
##	Davis, Miss. Mary	
##		1
##	Davison, Mrs. Thomas Henry (Mary E Finck)	
##		1
##	de Messemaeker, Mrs. Guillaume Joseph (Emma)	
##		1
##	de Mulder, Mr. Theodore	
##		1
##	de Pelsmaeker, Mr. Alfons	
##		1
##	Dean, Master. Bertram Vere	
##		1
##	Dean, Mr. Bertram Frank	
##		1
##	del Carlo, Mr. Sebastiano	
##		1
##	Denkoff, Mr. Mitto	
##		1
##	Dennis, Mr. Samuel	

##		1
##	Devaney, Miss. Margaret Delia	
##		1
##	Dick, Mr. Albert Adrian	
##		1
##	Dick, Mrs. Albert Adrian (Vera Gillespie)	
##		1
##	Dimic, Mr. Jovan	
##		1
##	Dodge, Master. Washington	
##		1
##	Doharr, Mr. Tannous	
##		1
##	Doling, Miss. Elsie	
##		1
##	Doling, Mrs. John T (Ada Julia Bone)	
##		1
##	Dooley, Mr. Patrick	
##		1
##	Dorking, Mr. Edward Arthur	
##		1
##	Douglas, Mr. Walter Donald	
##		1
##	Dowdell, Miss. Elizabeth	
##		1
##	Downton, Mr. William James	
##		1
##	Drazenoic, Mr. Jozef	
##		1
##	Drew, Mrs. James Vivian (Lulu Thorne Christian)	
##		1
##	Duane, Mr. Frank	
##		1
##	Duff Gordon, Lady. (Lucille Christiana Sutherland) ("Mrs Morgan")	
##		1
##	Duff Gordon, Sir. Cosmo Edmund ("Mr Morgan")	
##		1
##	Duran y More, Miss. Asuncion	
##		1
##	Edvardsson, Mr. Gustaf Hjalmar	
##		1
##	Eitemiller, Mr. George Floyd	
##		1
##	Eklund, Mr. Hans Linus	
##		1
##	Ekstrom, Mr. Johan	
##		1
##	Elias, Mr. Dibo	
##		1
##	Elias, Mr. Joseph Jr	
##		1
##	Elias, Mr. Tannous	
##		1
##	Elsbury, Mr. William James	

##		1
##	Emanuel, Miss. Virginia Ethel	
##		1
##	Emir, Mr. Farred Chehab	
##		1
##	Endres, Miss. Caroline Louise	
##		1
##	Eustis, Miss. Elizabeth Mussey	
##		1
##	Fahlstrom, Mr. Arne Jonas	
##		1
##	Farrell, Mr. James	
##		1
##	Farthing, Mr. John	
##		1
##	Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkinson)	
##		1
##	Fischer, Mr. Eberhard Thelander	
##		1
##	Fleming, Miss. Margaret	
##		1
##	Flynn, Mr. James	
##		1
##	Flynn, Mr. John	
##		1
##	Flynn, Mr. John Irwin ("Irving")	
##		1
##	Foo, Mr. Choong	
##		1
##	Ford, Miss. Doolina Margaret "Daisy"	
##		1
##	Ford, Miss. Robina Maggie "Ruby"	
##		1
##	Ford, Mr. William Neal	
##		1
##	Ford, Mrs. Edward (Margaret Ann Watson)	
##		1
##	Foreman, Mr. Benjamin Laventall	
##		1
##	Fortune, Miss. Alice Elizabeth	
##		1
##	Fortune, Miss. Mabel Helen	
##		1
##	Fortune, Mr. Charles Alexander	
##		1
##	Fortune, Mr. Mark	
##		1
##	Fox, Mr. Stanley Hubert	
##		1
##	Francatelli, Miss. Laura Mabel	
##		1
##	Frauenthal, Dr. Henry William	
##		1
##	Frauenthal, Mrs. Henry William (Clara Heinsheimer)	

##		1
##	Frolicher-Stehli, Mr. Maxmillian	
##		1
##	Frolicher, Miss. Hedwig Margaritha	
##		1
##	Frost, Mr. Anthony Wood "Archie"	
##		1
##	Fry, Mr. Richard	
##		1
##	Funk, Miss. Annie Clemmer	
##		1
##	Futrelle, Mr. Jacques Heath	
##		1
##	Futrelle, Mrs. Jacques Heath (Lily May Peel)	
##		1
##	Fynney, Mr. Joseph J	
##		1
##	Gale, Mr. Shadrach	
##		1
##	Gallagher, Mr. Martin	
##		1
##	Garfirth, Mr. John	
##		1
##	Garside, Miss. Ethel	
##		1
##	Gaskell, Mr. Alfred	
##		1
##	Gavey, Mr. Lawrence	
##		1
##	Gee, Mr. Arthur H	
##		1
##	Gheorgheff, Mr. Stanio	
##		1
##	Giglio, Mr. Victor	
##		1
##	Giles, Mr. Frederick Edward	
##		1
##	Gilinski, Mr. Eliezer	
##		1
##	Gill, Mr. John William	
##		1
##	Gillespie, Mr. William Henry	
##		1
##	Gilnagh, Miss. Katherine "Katie"	
##		1
##	Givard, Mr. Hans Kristensen	
##		1
##	Glynn, Miss. Mary Agatha	
##		1
##	Goldenberg, Mr. Samuel L	
##		1
##	Goldenberg, Mrs. Samuel L (Edwiga Grabowska)	
##		1
##	Goldschmidt, Mr. George B	

##		1
##	Goldsmith, Master. Frank John William "Frankie"	
##		1
##	Goldsmith, Mr. Frank John	
##		1
##	Goldsmith, Mrs. Frank John (Emily Alice Brown)	
##		1
##	Goncalves, Mr. Manuel Estanslas	
##		1
##	Goodwin, Master. Harold Victor	
##		1
##	Goodwin, Master. Sidney Leonard	
##		1
##	Goodwin, Master. William Frederick	
##		1
##	Goodwin, Miss. Lillian Amy	
##		1
##	Goodwin, Mr. Charles Edward	
##		1
##	Goodwin, Mrs. Frederick (Augusta Tyler)	
##		1
##	Graham, Miss. Margaret Edith	
##		1
##	Graham, Mr. George Edward	
##		1
##	Graham, Mrs. William Thompson (Edith Junkins)	
##		1
##	Green, Mr. George Henry	
##		1
##	Greenberg, Mr. Samuel	
##		1
##	Greenfield, Mr. William Bertram	
##		1
##	Gronnestad, Mr. Daniel Danielsen	
##		1
##	Guggenheim, Mr. Benjamin	
##		1
##	Gustafsson, Mr. Alfred Ossian	
##		1
##	Gustafsson, Mr. Anders Vilhelm	
##		1
##	Gustafsson, Mr. Johan Birger	
##		1
##	Gustafsson, Mr. Karl Gideon	
##		1
##	Haas, Miss. Aloisia	
##		1
##	Hagland, Mr. Ingvald Olai Olsen	
##		1
##	Hagland, Mr. Konrad Mathias Reiersen	
##		1
##	Hakkarainen, Mr. Pekka Pietari	
##		1
##	Hakkarainen, Mrs. Pekka Pietari (Elin Matilda Dolck)	

##		1
##	Hale, Mr. Reginald	
##		1
##	Hamalainen, Master. Viljo	
##		1
##	Hamalainen, Mrs. William (Anna)	
##		1
##	Hampe, Mr. Leon	
##		1
##	Hanna, Mr. Mansour	
##		1
##	Hansen, Mr. Claus Peter	
##		1
##	Hansen, Mr. Henrik Juul	
##		1
##	Hansen, Mr. Henry Damsgaard	
##		1
##	Harder, Mr. George Achilles	
##		1
##	Harknett, Miss. Alice Phoebe	
##		1
##	Harmer, Mr. Abraham (David Lishin)	
##		1
##	Harper, Miss. Annie Jessie "Nina"	
##		1
##	Harper, Mr. Henry Sleeper	
##		1
##	Harper, Mrs. Henry Sleeper (Myna Haxtun)	
##		1
##	Harper, Rev. John	
##		1
##	Harrington, Mr. Charles H	
##		1
##	Harris, Mr. George	
##		1
##	Harris, Mr. Henry Birkhardt	
##		1
##	Harris, Mr. Walter	
##		1
##	Harris, Mrs. Henry Birkhardt (Irene Wallach)	
##		1
##	Harrison, Mr. William	
##		1
##	Hart, Miss. Eva Miriam	
##		1
##	Hart, Mr. Benjamin	
##		1
##	Hart, Mr. Henry	
##		1
##	Hart, Mrs. Benjamin (Esther Ada Bloomfield)	
##		1
##	Hassab, Mr. Hammad	
##		1
##	Hassan, Mr. Houssein G N	



##		1
##	Hawksford, Mr. Walter James	
##		1
##	Hays, Miss. Margaret Bechstein	
##		1
##	Hays, Mrs. Charles Melville (Clara Jennings Gregg)	
##		1
##	Healy, Miss. Hanora "Nora"	
##		1
##	Hedman, Mr. Oskar Arvid	
##		1
##	Hegarty, Miss. Hanora "Nora"	
##		1
##	Heikkinen, Miss. Laina	
##		1
##	Heininen, Miss. Wendla Maria	
##		1
##	Hendekovic, Mr. Ignjac	
##		1
##	Henry, Miss. Delia	
##		1
##	Herman, Miss. Alice	
##		1
##	Herman, Mrs. Samuel (Jane Laver)	
##		1
##	Hewlett, Mrs. (Mary D Kingcome)	
##		1
##	Hickman, Mr. Leonard Mark	
##		1
##	Hickman, Mr. Lewis	
##		1
##	Hickman, Mr. Stanley George	
##		1
##	Hippach, Miss. Jean Gertrude	
##		1
##	Hippach, Mrs. Louis Albert (Ida Sophia Fischer)	
##		1
##	Hirvonen, Miss. Hildur E	
##		1
##	Hocking, Mr. Richard George	
##		1
##	Hocking, Mrs. Elizabeth (Eliza Needs)	
##		1
##	Hodges, Mr. Henry Price	
##		1
##	Hogeboom, Mrs. John C (Anna Andrews)	
##		1
##	Hold, Mr. Stephen	
##		1
##	Holm, Mr. John Fredrik Alexander	
##		1
##	Holverson, Mr. Alexander Oskar	
##		1
##	Holverson, Mrs. Alexander Oskar (Mary Aline Towner)	

##		1
##	Homer, Mr. Harry ("Mr E Haven")	
##		1
##	Honkanen, Miss. Eliina	
##		1
##	Hood, Mr. Ambrose Jr	
##		1
##	Horgan, Mr. John	
##		1
##	Hosono, Mr. Masabumi	
##		1
##	Hoyt, Mr. Frederick Maxfield	
##		1
##	Hoyt, Mr. William Fisher	
##		1
##	Hoyt, Mrs. Frederick Maxfield (Jane Anne Forby)	
##		1
##	Humblen, Mr. Adolf Mathias Nicolai Olsen	
##		1
##	Hunt, Mr. George Henry	
##		1
##	Ibrahim Shawah, Mr. Yousseff	
##		1
##	Icard, Miss. Amelie	
##		1
##	Ilett, Miss. Bertha	
##		1
##	Ilmakangas, Miss. Pieta Sofia	
##		1
##	Isham, Miss. Ann Elizabeth	
##		1
##	Ivanoff, Mr. Kanio	
##		1
##	Jacobsohn, Mr. Sidney Samuel	
##		1
##	Jacobsohn, Mrs. Sidney Samuel (Amy Frances Christy)	
##		1
##	Jalsevac, Mr. Ivan	
##		1
##	Jansson, Mr. Carl Olof	
##		1
##	Jardin, Mr. Jose Neto	
##		1
##	Jarvis, Mr. John Denzil	
##		1
##	Jenkin, Mr. Stephen Curnow	
##		1
##	Jensen, Mr. Hans Peder	
##		1
##	Jensen, Mr. Niels Peder	
##		1
##	Jensen, Mr. Svend Lauritz	
##		1
##	Jermyn, Miss. Annie	

##		1
##	Jerwan, Mrs. Amin S (Marie Marthe Thuillard)	
##		1
##	Johannesen-Bratthammer, Mr. Bernt	
##		1
##	Johanson, Mr. Jakob Alfred	
##		1
##	Johansson, Mr. Erik	
##		1
##	Johansson, Mr. Gustaf Joel	
##		1
##	Johansson, Mr. Karl Johan	
##		1
##	Johnson, Master. Harold Theodor	
##		1
##	Johnson, Miss. Eleanor Ileen	
##		1
##	Johnson, Mr. Alfred	
##		1
##	Johnson, Mr. Malkolm Joackim	
##		1
##	Johnson, Mr. William Cahoone Jr	
##		1
##	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	
##		1
##	Johnston, Miss. Catherine Helen "Carrie"	
##		1
##	Johnston, Mr. Andrew G	
##		1
##	Jonkoff, Mr. Lallio	
##		1
##	Jonsson, Mr. Carl	
##		1
##	Jussila, Miss. Katriina	
##		1
##	Jussila, Miss. Mari Aina	
##		1
##	Jussila, Mr. Eiriik	
##		1
##	Kallio, Mr. Nikolai Erland	
##		1
##	Kalvik, Mr. Johannes Halvorsen	
##		1
##	Kantor, Mr. Sinai	
##		1
##	Kantor, Mrs. Sinai (Miriam Sternin)	
##		1
##	Karaic, Mr. Milan	
##		1
##	Karlsson, Mr. Nils August	
##		1
##	Karun, Miss. Manca	
##		1
##	Kassem, Mr. Fared	

##		1
##	Keane, Miss. Nora A	
##		1
##	Keane, Mr. Andrew "Andy"	
##		1
##	Keefe, Mr. Arthur	
##		1
##	Kelly, Miss. Anna Katherine "Annie Kate"	
##		1
##	Kelly, Miss. Mary	
##		1
##	Kelly, Mr. James	
##		1
##	Kelly, Mrs. Florence "Fannie"	
##		1
##	Kent, Mr. Edward Austin	
##		1
##	Kenyon, Mrs. Frederick R (Marion)	
##		1
##	Kiernan, Mr. Philip	
##		1
##	Kilgannon, Mr. Thomas J	
##		1
##	Kimball, Mr. Edwin Nelson Jr	
##		1
##	Kink-Heilmann, Miss. Luise Gretchen	
##		1
##	Kink, Mr. Vincenz	
##		1
##	Kirkland, Rev. Charles Leonard	
##		1
##	Klaber, Mr. Herman	
##		1
##	Klasen, Mr. Klas Albin	
##		1
##	Knight, Mr. Robert J	
##		1
##	Kraeff, Mr. Theodor	
##		1
##	Kvillner, Mr. Johan Henrik Johannesson	
##		1
##	Lahoud, Mr. Sarkis	
##		1
##	Lahtinen, Mrs. William (Anna Sylfven)	
##		1
##	Laitinen, Miss. Kristina Sofia	
##		1
##	Laleff, Mr. Kristo	
##		1
##	Lam, Mr. Ali	
##		1
##	Lam, Mr. Len	
##		1
##	Landergren, Miss. Aurora Adelia	

##		1
##	Lang, Mr. Fang	
##		1
##	Laroche, Miss. Simonne Marie Anne Andree	
##		1
##	Laroche, Mr. Joseph Philippe Lemerancier	
##		1
##	Laroche, Mrs. Joseph (Juliette Marie Louise Lafargue)	
##		1
##	Larsson, Mr. August Viktor	
##		1
##	Larsson, Mr. Bengt Edvin	
##		1
##	Leader, Dr. Alice (Farnham)	
##		1
##	Leeni, Mr. Fahim ("Philip Zenni")	
##		1
##	Lefebvre, Master. Henry Forbes	
##		1
##	Lefebvre, Miss. Ida	
##		1
##	Lefebvre, Miss. Jeannie	
##		1
##	Lefebvre, Miss. Mathilde	
##		1
##	Lehmann, Miss. Bertha	
##		1
##	Leinonen, Mr. Antti Gustaf	
##		1
##	Leitch, Miss. Jessie Wills	
##		1
##	Lemberopolous, Mr. Peter L	
##		1
##	Lemore, Mrs. (Amelia Milley)	
##		1
##	Lennon, Mr. Denis	
##		1
##	Leonard, Mr. Lionel	
##		1
##	LeRoy, Miss. Bertha	
##		1
##	Lester, Mr. James	
##		1
##	Lesurer, Mr. Gustave J	
##		1
##	Levy, Mr. Rene Jacques	
##		1
##	Lewy, Mr. Ervin G	
##		1
##	Leyson, Mr. Robert William Norman	
##		1
##	Lievens, Mr. Rene Aime	
##		1
##	Lindahl, Miss. Agda Thorilda Viktoria	

##		1
##	Lindblom, Miss. Augusta Charlotta	
##		1
##	Lindell, Mr. Edvard Bengtsson	
##		1
##	Lindqvist, Mr. Eino William	
##		1
##	Lines, Miss. Mary Conover	
##		1
##	Ling, Mr. Lee	
##		1
##	Lobb, Mr. William Arthur	
##		1
##	Lobb, Mrs. William Arthur (Cordelia K Stanlick)	
##		1
##	Long, Mr. Milton Clyde	
##		1
##	Longley, Miss. Gretchen Fiske	
##		1
##	Louch, Mrs. Charles Alexander (Alice Adelaide Slow)	
##		1
##	Lovell, Mr. John Hall ("Henry")	
##		1
##	Lulic, Mr. Nikola	
##		1
##	Lundahl, Mr. Johan Svensson	
##		1
##	Lurette, Miss. Elise	
##		1
##	Mack, Mrs. (Mary)	
##		1
##	Madigan, Miss. Margaret "Maggie"	
##		1
##	Madill, Miss. Georgette Alexandra	
##		1
##	Madsen, Mr. Fridtjof Arne	
##		1
##	Maenpaa, Mr. Matti Alexanteri	
##		1
##	Maioni, Miss. Roberta	
##		1
##	Maisner, Mr. Simon	
##		1
##	Mallet, Master. Andre	
##		1
##	Mallet, Mr. Albert	
##		1
##	Mamee, Mr. Hanna	
##		1
##	Mangan, Miss. Mary	
##		1
##	Mannion, Miss. Margareth	
##		1
##	Marechal, Mr. Pierre	

##		1
##	Markoff, Mr. Marin	
##		1
##	Markun, Mr. Johann	
##		1
##	Marvin, Mr. Daniel Warner	
##		1
##	Masselmani, Mrs. Fatima	
##		1
##	Matthews, Mr. William John	
##		1
##	Mayne, Mlle. Berthe Antonine ("Mrs de Villiers")	
##		1
##	McCarthy, Mr. Timothy J	
##		1
##	McCormack, Mr. Thomas Joseph	
##		1
##	McCoy, Miss. Agnes	
##		1
##	McCoy, Mr. Bernard	
##		1
##	McDermott, Miss. Brigdet Delia	
##		1
##	McEvoy, Mr. Michael	
##		1
##	McGough, Mr. James Robert	
##		1
##	McGovern, Miss. Mary	
##		1
##	McGowan, Miss. Anna "Annie"	
##		1
##	McKane, Mr. Peter David	
##		1
##	McMahon, Mr. Martin	
##		1
##	McNamee, Mr. Neal	
##		1
##	Meanwell, Miss. (Marion Ogden)	
##		1
##	Meek, Mrs. Thomas (Annie Louise Rowley)	
##		1
##	Mellinger, Miss. Madeleine Violet	
##		1
##	Mellinger, Mrs. (Elizabeth Anne Maidment)	
##		1
##	Mellors, Mr. William John	
##		1
##	Meo, Mr. Alfonzo	
##		1
##	Mernagh, Mr. Robert	
##		1
##	Meyer, Mr. August	
##		1
##	Meyer, Mr. Edgar Joseph	

##		1
##	Meyer, Mrs. Edgar Joseph (Leila Saks)	
##		1
##	Millet, Mr. Francis Davis	
##		1
##	Milling, Mr. Jacob Christian	
##		1
##	Minahan, Dr. William Edward	
##		1
##	Minahan, Miss. Daisy E	
##		1
##	Mineff, Mr. Ivan	
##		1
##	Mionoff, Mr. Stoytcho	
##		1
##	Mitchell, Mr. Henry Michael	
##		1
##	Mitkoff, Mr. Mito	
##		1
##	Mockler, Miss. Helen Mary "Ellie"	
##		1
##	Moen, Mr. Sigurd Hansen	
##		1
##	Molson, Mr. Harry Markland	
##		1
##	Montvila, Rev. Juozas	
##		1
##	Moor, Master. Meier	
##		1
##	Moor, Mrs. (Beila)	
##		1
##	Moore, Mr. Leonard Charles	
##		1
##	Moran, Miss. Bertha	
##		1
##	Moran, Mr. Daniel J	
##		1
##	Moran, Mr. James	
##		1
##	Moraweck, Dr. Ernest	
##		1
##	Morley, Mr. Henry Samuel ("Mr Henry Marshall")	
##		1
##	Morley, Mr. William	
##		1
##	Morrow, Mr. Thomas Rowan	
##		1
##	Moss, Mr. Albert Johan	
##		1
##	Moubarek, Master. Gerios	
##		1
##	Moubarek, Master. Halim Gonios ("William George")	
##		1
##	Moussa, Mrs. (Mantoura Boulos)	



##		1
##	Moutal, Mr. Rahamin Haim	
##		1
##	Mudd, Mr. Thomas Charles	
##		1
##	Mullens, Miss. Katherine "Katie"	
##		1
##	Murdlin, Mr. Joseph	
##		1
##	Murphy, Miss. Katherine "Kate"	
##		1
##	Murphy, Miss. Margaret Jane	
##		1
##	Myhrman, Mr. Pehr Fabian Oliver Malkolm	
##		1
##	Naidenoff, Mr. Penko	
##		1
##	Najib, Miss. Adele Kiamie "Jane"	
##		1
##	Nakid, Miss. Maria ("Mary")	
##		1
##	Nakid, Mr. Sahid	
##		1
##	Nankoff, Mr. Minko	
##		1
##	Nasser, Mr. Nicholas	
##		1
##	Nasser, Mrs. Nicholas (Adele Achem)	
##		1
##	Natsch, Mr. Charles H	
##		1
##	Navratil, Master. Edmond Roger	
##		1
##	Navratil, Master. Michel M	
##		1
##	Navratil, Mr. Michel ("Louis M Hoffman")	
##		1
##	Nenkoff, Mr. Christo	
##		1
##	Newell, Miss. Madeleine	
##		1
##	Newell, Miss. Marjorie	
##		1
##	Newell, Mr. Arthur Webster	
##		1
##	Newsom, Miss. Helen Monypeny	
##		1
##	Nicholls, Mr. Joseph Charles	
##		1
##	Nicholson, Mr. Arthur Ernest	
##		1
##	Nicola-Yarred, Master. Elias	
##		1
##	Nicola-Yarred, Miss. Jamila	

##		1
##	Nilsson, Miss. Helmina Josefina	
##		1
##	Nirva, Mr. Iisakki Antino Aijo	
##		1
##	Niskanen, Mr. Juha	
##		1
##	Norman, Mr. Robert Douglas	
##		1
##	Nosworthy, Mr. Richard Cater	
##		1
##	Novel, Mr. Mansouer	
##		1
##	Nye, Mrs. (Elizabeth Ramell)	
##		1
##	Nysten, Miss. Anna Sofia	
##		1
##	Nysveen, Mr. Johan Hansen	
##		1
##	O'Brien, Mr. Thomas	
##		1
##	O'Brien, Mr. Timothy	
##		1
##	O'Brien, Mrs. Thomas (Johanna "Hannah" Godfrey)	
##		1
##	O'Connell, Mr. Patrick D	
##		1
##	O'Connor, Mr. Maurice	
##		1
##	O'Driscoll, Miss. Bridget	
##		1
##	O'Dwyer, Miss. Ellen "Nellie"	
##		1
##	O'Leary, Miss. Hanora "Norah"	
##		1
##	O'Sullivan, Miss. Bridget Mary	
##		1
##	Odahl, Mr. Nils Martin	
##		1
##	Ohman, Miss. Velin	
##		1
##	Olsen, Mr. Henry Margido	
##		1
##	Olsen, Mr. Karl Siegwart Andreas	
##		1
##	Olsen, Mr. Ole Martin	
##		1
##	Olsson, Miss. Elina	
##		1
##	Olsson, Mr. Nils Johan Goransson	
##		1
##	Olsvigen, Mr. Thor Anderson	
##		1
##	Oreskovic, Miss. Marija	

##		1
##	Oreskovic, Mr. Luka	
##		1
##	Osen, Mr. Olaf Elon	
##		1
##	Osman, Mrs. Mara	
##		1
##	Ostby, Mr. Engelhart Cornelius	
##		1
##	Otter, Mr. Richard	
##		1
##	Padro y Manent, Mr. Julian	
##		1
##	Pain, Dr. Alfred	
##		1
##	Palsson, Master. Gosta Leonard	
##		1
##	Palsson, Miss. Stina Viola	
##		1
##	Palsson, Miss. Torborg Danira	
##		1
##	Palsson, Mrs. Nils (Alma Cornelia Berglund)	
##		1
##	Panula, Master. Eino Viljami	
##		1
##	Panula, Master. Juha Niilo	
##		1
##	Panula, Master. Urho Abraham	
##		1
##	Panula, Mr. Ernesti Arvid	
##		1
##	Panula, Mr. Jaako Arnold	
##		1
##	Panula, Mrs. Juha (Maria Emilia Ojala)	
##		1
##	Parkes, Mr. Francis "Frank"	
##		1
##	Parr, Mr. William Henry Marsh	
##		1
##	Parrish, Mrs. (Lutie Davis)	
##		1
##	Partner, Mr. Austen	
##		1
##	Pasic, Mr. Jakob	
##		1
##	Patchett, Mr. George	
##		1
##	Paulner, Mr. Uscher	
##		1
##	Pavlovic, Mr. Stefo	
##		1
##	Pears, Mr. Thomas Clinton	
##		1
##	Pears, Mrs. Thomas (Edith Wearne)	

##		1
##	Peduzzi, Mr. Joseph	
##		1
##	Pekoniemi, Mr. Edvard	
##		1
##	Penasco y Castellana, Mr. Victor de Satode	
##		1
##	Penasco y Castellana, Mrs. Victor de Satode (Maria Josefa Perez de Soto y Vallejo)	
##		1
##	Pengelly, Mr. Frederick William	
##		1
##	Perkin, Mr. John Henry	
##		1
##	Pernot, Mr. Rene	
##		1
##	Perreault, Miss. Anne	
##		1
##	Persson, Mr. Ernst Ulrik	
##		1
##	Peter, Miss. Anna	
##		1
##	Peter, Mrs. Catherine (Catherine Rizk)	
##		1
##	Peters, Miss. Katie	
##		1
##	Petranec, Miss. Matilda	
##		1
##	Petroff, Mr. Nedelio	
##		1
##	Petroff, Mr. Pastcho ("Pentcho")	
##		1
##	Petterson, Mr. Johan Emil	
##		1
##	Pettersson, Miss. Ellen Natalia	
##		1
##	Peuchen, Major. Arthur Godfrey	
##		1
##	Phillips, Miss. Kate Florence ("Mrs Kate Louise Phillips Marshall")	
##		1
##	Pickard, Mr. Berk (Berk Trembisky)	
##		1
##	Pinsky, Mrs. (Rosa)	
##		1
##	Plotcharsky, Mr. Vasil	
##		1
##	Ponesell, Mr. Martin	
##		1
##	Porter, Mr. Walter Chamberlain	
##		1
##	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	
##		1
##	Quick, Miss. Phyllis May	
##		1
##	Quick, Mrs. Frederick Charles (Jane Richards)	

##		1
##	Radeff, Mr. Alexander	
##		1
##	Razi, Mr. Raihed	
##		1
##	Reed, Mr. James George	
##		1
##	Reeves, Mr. David	
##		1
##	Rekic, Mr. Tido	
##		1
##	Renouf, Mr. Peter Henry	
##		1
##	Renouf, Mrs. Peter Henry (Lillian Jefferys)	
##		1
##	Reuchlin, Jonkheer. John George	
##		1
##	Reynaldo, Ms. Encarnacion	
##		1
##	Rice, Master. Arthur	
##		1
##	Rice, Master. Eric	
##		1
##	Rice, Master. Eugene	
##		1
##	Rice, Master. George Hugh	
##		1
##	Rice, Mrs. William (Margaret Norton)	
##		1
##	Richard, Mr. Emile	
##		1
##	Richards, Master. George Sibley	
##		1
##	Richards, Master. William Rowe	
##		1
##	Richards, Mrs. Sidney (Emily Hocking)	
##		1
##	Ridsdale, Miss. Lucy	
##		1
##	Ringhini, Mr. Sante	
##		1
##	Rintamaki, Mr. Matti	
##		1
##	Risien, Mr. Samuel Beard	
##		1
##	Robbins, Mr. Victor	
##		1
##	Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)	
##		1
##	Robins, Mrs. Alexander A (Grace Charity Laury)	
##		1
##	Roebbling, Mr. Washington Augustus II	
##		1
##	Rogers, Mr. William John	

##		1
##	Romaine, Mr. Charles Hallace ("Mr C Rolmane")	
##		1
##	Rommetvedt, Mr. Knud Paust	
##		1
##	Rood, Mr. Hugh Roscoe	
##		1
##	Rosblom, Mr. Viktor Richard	
##		1
##	Rosblom, Mrs. Viktor (Helena Wilhelmina)	
##		1
##	Ross, Mr. John Hugo	
##		1
##	Roths, the Countess. of (Lucy Noel Martha Dyer-Edwards)	
##		1
##	Rothschild, Mrs. Martin (Elizabeth L. Barrett)	
##		1
##	Rouse, Mr. Richard Henry	
##		1
##	Rugg, Miss. Emily	
##		1
##	Rush, Mr. Alfred George John	
##		1
##	Ryan, Mr. Patrick	
##		1
##	Ryerson, Miss. Emily Borie	
##		1
##	Ryerson, Miss. Susan Parker "Suzette"	
##		1
##	Saad, Mr. Amin	
##		1
##	Saad, Mr. Khalil	
##		1
##	Saalfeld, Mr. Adolphe	
##		1
##	Sadlier, Mr. Matthew	
##		1
##	Sage, Master. Thomas Henry	
##		1
##	Sage, Miss. Constance Gladys	
##		1
##	Sage, Miss. Dorothy Edith "Dolly"	
##		1
##	Sage, Miss. Stella Anna	
##		1
##	Sage, Mr. Douglas Bullen	
##		1
##	Sage, Mr. Frederick	
##		1
##	Sage, Mr. George John Jr	
##		1
##	Sagesser, Mlle. Emma	
##		1
##	Salkjelsvik, Miss. Anna Kristine	

##		1
##	Salonen, Mr. Johan Werner	
##		1
##	Samaan, Mr. Youssef	
##		1
##	Sandstrom, Miss. Marguerite Rut	
##		1
##	Sandstrom, Mrs. Hjalmar (Agnes Charlotta Bengtsson)	
##		1
##	Saundercock, Mr. William Henry	
##		1
##	Sawyer, Mr. Frederick Charles	
##		1
##	Scanlan, Mr. James	
##		1
##	Sdycoff, Mr. Todor	
##		1
##	Sedgwick, Mr. Charles Frederick Waddington	
##		1
##	Serepeca, Miss. Augusta	
##		1
##	Seward, Mr. Frederic Kimber	
##		1
##	Sharp, Mr. Percival James R	
##		1
##	Sheerlinck, Mr. Jan Baptist	
##		1
##	Shellard, Mr. Frederick William	
##		1
##	Shelley, Mrs. William (Imanita Parrish Hall)	
##		1
##	Shorney, Mr. Charles Joseph	
##		1
##	Shutes, Miss. Elizabeth W	
##		1
##	Silven, Miss. Lyyli Karoliina	
##		1
##	Silverthorne, Mr. Spencer Victor	
##		1
##	Silvey, Mr. William Baird	
##		1
##	Silvey, Mrs. William Baird (Alice Munger)	
##		1
##	Simmons, Mr. John	
##		1
##	Simonius-Blumer, Col. Oberst Alfons	
##		1
##	Sinkkonen, Miss. Anna	
##		1
##	Sirayanian, Mr. Orsen	
##		1
##	Sirota, Mr. Maurice	
##		1
##	Sivic, Mr. Husein	

##		1
##	Sivola, Mr. Antti Wilhelm	
##		1
##	Sjoblom, Miss. Anna Sofia	
##		1
##	Sjostedt, Mr. Ernst Adolf	
##		1
##	Skoog, Master. Harald	
##		1
##	Skoog, Master. Karl Thorsten	
##		1
##	Skoog, Miss. Mabel	
##		1
##	Skoog, Miss. Margit Elizabeth	
##		1
##	Skoog, Mr. Wilhelm	
##		1
##	Skoog, Mrs. William (Anna Bernhardina Karlsson)	
##		1
##	Slabenoff, Mr. Petco	
##		1
##	Slayter, Miss. Hilda Mary	
##		1
##	Slemen, Mr. Richard James	
##		1
##	Slocovski, Mr. Selman Francis	
##		1
##	Sloper, Mr. William Thompson	
##		1
##	Smart, Mr. John Montgomery	
##		1
##	Smiljanic, Mr. Mile	
##		1
##	Smith, Miss. Marion Elsie	
##		1
##	Smith, Mr. James Clinch	
##		1
##	Smith, Mr. Richard William	
##		1
##	Smith, Mr. Thomas	
##		1
##	Sobey, Mr. Samuel James Hayden	
##		1
##	Soholt, Mr. Peter Andreas Lauritz Andersen	
##		1
##	Somerton, Mr. Francis William	
##		1
##	Spedden, Mrs. Frederic Oakley (Margaretta Corning Stone)	
##		1
##	Spencer, Mrs. William Augustus (Marie Eugenie)	
##		1
##	Stahelin-Maeglin, Dr. Max	
##		1
##	Staneff, Mr. Ivan	



##		1
##	Stankovic, Mr. Ivan	
##		1
##	Stanley, Miss. Amy Zillah Elsie	
##		1
##	Stanley, Mr. Edward Roland	
##		1
##	Stead, Mr. William Thomas	
##		1
##	Stephenson, Mrs. Walter Bertram (Martha Eustis)	
##		1
##	Stewart, Mr. Albert A	
##		1
##	Stone, Mrs. George Nelson (Martha Evelyn)	
##		1
##	Stoytcheff, Mr. Ilia	
##		1
##	Strandberg, Miss. Ida Sofia	
##		1
##	Stranden, Mr. Juho	
##		1
##	Strom, Miss. Telma Matilda	
##		1
##	Strom, Mrs. Wilhelm (Elna Matilda Persson)	
##		1
##	Sunderland, Mr. Victor Francis	
##		1
##	Sundman, Mr. Johan Julian	
##		1
##	Sutehall, Mr. Henry Jr	
##		1
##	Sutton, Mr. Frederick	
##		1
##	Svensson, Mr. Johan	
##		1
##	Svensson, Mr. Olof	
##		1
##	Swift, Mrs. Frederick Joel (Margaret Welles Barron)	
##		1
##	Taussig, Miss. Ruth	
##		1
##	Taussig, Mr. Emil	
##		1
##	Taussig, Mrs. Emil (Tillie Mandelbaum)	
##		1
##	Taylor, Mr. Elmer Zebley	
##		1
##	Taylor, Mrs. Elmer Zebley (Juliet Cummins Wright)	
##		1
##	Thayer, Mr. John Borland	
##		1
##	Thayer, Mr. John Borland Jr	
##		1
##	Thayer, Mrs. John Borland (Marian Longstreth Morris)	

##		1
##	Theobald, Mr. Thomas Leonard	
##		1
##	Thomas, Master. Assad Alexander	
##		1
##	Thorne, Mrs. Gertrude Maybelle	
##		1
##	Thorneycroft, Mr. Percival	
##		1
##	Thorneycroft, Mrs. Percival (Florence Kate White)	
##		1
##	Tikkanen, Mr. Juho	
##		1
##	Tobin, Mr. Roger	
##		1
##	Todoroff, Mr. Lallo	
##		1
##	Tomlin, Mr. Ernest Portage	
##		1
##	Toomey, Miss. Ellen	
##		1
##	Torber, Mr. Ernst William	
##		1
##	Tornquist, Mr. William Henry	
##		1
##	Toufik, Mr. Nakli	
##		1
##	Touma, Mrs. Darwis (Hanne Youssef Razi)	
##		1
##	Troupiansky, Mr. Moses Aaron	
##		1
##	Trout, Mrs. William H (Jessie L)	
##		1
##	Troutt, Miss. Edwina Celia "Winnie"	
##		1
##	Turcin, Mr. Stjepan	
##		1
##	Turja, Miss. Anna Sofia	
##		1
##	Turkula, Mrs. (Hedwig)	
##		1
##	Turpin, Mr. William John Robert	
##		1
##	Turpin, Mrs. William John Robert (Dorothy Ann Wonnacott)	
##		1
##	Uruchurtu, Don. Manuel E	
##		1
##	van Billiard, Mr. Austin Blyler	
##		1
##	Van der hoef, Mr. Wyckoff	
##		1
##	Van Impe, Miss. Catharina	
##		1
##	Van Impe, Mr. Jean Baptiste	

##		1
##	Van Impe, Mrs. Jean Baptiste (Rosalie Paula Govaert)	
##		1
##	van Melkebeke, Mr. Philemon	
##		1
##	Vande Velde, Mr. Johannes Joseph	
##		1
##	Vande Walle, Mr. Nestor Cyriel	
##		1
##	Vanden Steen, Mr. Leo Peter	
##		1
##	Vander Cruyssen, Mr. Victor	
##		1
##	Vander Planke, Miss. Augusta Maria	
##		1
##	Vander Planke, Mr. Leo Edmondus	
##		1
##	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	
##		1
##	Vestrom, Miss. Hulda Amanda Adolfina	
##		1
##	Vovk, Mr. Janko	
##		1
##	Waelens, Mr. Achille	
##		1
##	Walker, Mr. William Anderson	
##		1
##	Ward, Miss. Anna	
##		1
##	Warren, Mrs. Frank Manley (Anna Sophia Atkinson)	
##		1
##	Watson, Mr. Ennis Hastings	
##		1
##	Watt, Mrs. James (Elizabeth "Bessie" Inglis Milne)	
##		1
##	Webber, Miss. Susan	
##		1
##	Webber, Mr. James	
##		1
##	Weir, Col. John	
##		1
##	Weisz, Mrs. Leopold (Mathilde Francoise Pede)	
##		1
##	Wells, Miss. Joan	
##		1
##	West, Miss. Constance Mirium	
##		1
##	West, Mr. Edwy Arthur	
##		1
##	West, Mrs. Edwy Arthur (Ada Mary Worth)	
##		1
##	Wheadon, Mr. Edward H	
##		1
##	White, Mr. Percival Wayland	

##		1
##	White, Mr. Richard Frasar	
##		1
##	Wick, Miss. Mary Natalie	
##		1
##	Wick, Mrs. George Dennick (Mary Hitchcock)	
##		1
##	Widegren, Mr. Carl/Charles Peter	
##		1
##	Widener, Mr. Harry Elkins	
##		1
##	Wiklund, Mr. Jakob Alfred	
##		1
##	Wilhelms, Mr. Charles	
##		1
##	Willey, Mr. Edward	
##		1
##	Williams-Lambert, Mr. Fletcher Fellows	
##		1
##	Williams, Mr. Charles Duane	
##		1
##	Williams, Mr. Charles Eugene	
##		1
##	Williams, Mr. Howard Hugh "Harry"	
##		1
##	Williams, Mr. Leslie	
##		1
##	Windelov, Mr. Einar	
##		1
##	Wiseman, Mr. Phillippe	
##		1
##	Woolner, Mr. Hugh	
##		1
##	Wright, Mr. George	
##		1
##	Yasbeck, Mr. Antoni	
##		1
##	Yasbeck, Mrs. Antoni (Selini Alexander)	
##		1
##	Young, Miss. Marie Grice	
##		1
##	Youseff, Mr. Gerious	
##		1
##	Yousif, Mr. Wazli	
##		1
##	Yousseff, Mr. Gerious	
##		1
##	Yrois, Miss. Henriette ("Mrs Harbeck")	
##		1
##	Zabour, Miss. Hileni	
##		1
##	Zabour, Miss. Thamine	
##		1
##	Zimmerman, Mr. Leo	

```
##
```

1

Verificamos que no hay valores vacíos.

**Sex**

```
table(titanicData$Sex)
```

```
##
```

```
## female    male
```

```
##      314     577
```

Esta variable no tiene observaciones sin valor.

```
577/891*100
```

```
## [1] 64.7587
```

Casi el 65% de los pasajeros son de género masculino.

**Ticket**

```
table(titanicData$Ticket)
```

```
##
```

##	110152	110413	110465	110564
##	3	3	2	1
##	110813	111240	111320	111361
##	1	1	1	2
##	111369	111426	111427	111428
##	1	1	1	1
##	112050	112052	112053	112058
##	1	1	1	1
##	112059	112277	112379	113028
##	1	1	1	1
##	113043	113050	113051	113055
##	1	1	1	1
##	113056	113059	113501	113503
##	1	1	1	1
##	113505	113509	113510	113514
##	2	1	1	1
##	113572	113760	113767	113773
##	2	4	1	1
##	113776	113781	113783	113784
##	2	4	1	1
##	113786	113787	113788	113789
##	1	1	1	2
##	113792	113794	113796	113798
##	1	1	1	2
##	113800	113803	113804	113806
##	1	2	1	2
##	113807	11668	11751	11752
##	1	2	2	1

##	11753	11755	11765	11767
##	1	1	1	2
##	11769	11771	11774	11813
##	1	1	1	1
##	11967	12233	12460	12749
##	2	1	1	2
##	13049	13213	13214	13502
##	1	1	1	3
##	13507	13509	13567	13568
##	2	1	1	1
##	14311	14312	14313	14973
##	1	1	1	1
##	1601	16966	16988	17421
##	7	2	1	4
##	17453	17463	17464	17465
##	2	1	1	1
##	17466	17474	17764	19877
##	1	2	1	2
##	19928	19943	19947	19950
##	2	2	1	4
##	19952	19972	19988	19996
##	1	1	1	2
##	2003	211536	21440	218629
##	1	1	1	1
##	219533	220367	220845	2223
##	1	1	2	1
##	223596	226593	226875	228414
##	1	1	1	1
##	229236	230080	230136	230433
##	1	3	2	2
##	230434	231919	231945	233639
##	1	2	1	1
##	233866	234360	234604	234686
##	1	1	1	1
##	234818	236171	236852	236853
##	1	1	1	1
##	237442	237565	237668	237671
##	1	1	1	1
##	237736	237789	237798	239853
##	2	1	1	3
##	239854	239855	239856	239865
##	1	1	1	2
##	240929	24160	243847	243880
##	1	3	2	1
##	244252	244270	244278	244310
##	2	1	1	1
##	244358	244361	244367	244373
##	1	1	2	1
##	248698	248706	248723	248727
##	1	1	1	3
##	248731	248733	248738	248740
##	1	1	2	1
##	248747	250643	250644	250646
##	1	1	2	1

##	250647	250648	250649	250651
##	2	1	2	1
##	250652	250653	250655	2620
##	1	1	2	1
##	2623	2624	2625	2626
##	1	1	1	1
##	2627	2628	2629	2631
##	2	1	1	1
##	26360	2641	2647	2648
##	2	1	1	1
##	2649	2650	2651	2653
##	1	1	2	2
##	2659	2661	2662	2663
##	2	2	1	1
##	2664	2665	2666	2667
##	1	2	4	1
##	2668	2669	26707	2671
##	2	1	1	1
##	2672	2674	2677	2678
##	1	1	1	2
##	2680	2683	2685	2686
##	1	1	1	1
##	2687	2689	2690	2691
##	1	1	1	2
##	2693	2694	2695	2697
##	1	1	1	1
##	2699	2700	27042	27267
##	2	1	1	1
##	27849	28134	28206	28213
##	1	1	1	1
##	28220	28228	28403	28424
##	1	1	2	1
##	28425	28551	28664	28665
##	1	1	1	1
##	29011	2908	29103	29104
##	1	2	1	1
##	29105	29106	29108	2926
##	1	3	1	1
##	29750	29751	3101264	3101265
##	2	1	1	1
##	3101267	3101276	3101277	3101278
##	1	1	1	2
##	3101281	3101295	3101296	3101298
##	1	6	1	1
##	31027	31028	312991	312992
##	2	1	1	1
##	312993	31418	315037	315082
##	1	1	1	1
##	315084	315086	315088	315089
##	1	1	1	1
##	315090	315093	315094	315096
##	1	1	1	1
##	315097	315098	315151	315153
##	1	1	1	1

##	323592	323951	324669	330877
##	1	1	1	1
##	330909	330919	330923	330931
##	1	1	1	1
##	330932	330935	330958	330959
##	1	1	1	1
##	330979	330980	334912	335097
##	1	1	1	1
##	335677	33638	336439	3411
##	1	1	1	1
##	341826	34218	342826	343095
##	1	1	1	1
##	343120	343275	343276	345364
##	1	1	1	1
##	345572	345763	345764	345765
##	1	1	2	1
##	345767	345769	345770	345773
##	1	1	1	3
##	345774	345777	345778	345779
##	1	1	1	1
##	345780	345781	345783	3460
##	1	1	1	1
##	347054	347060	347061	347062
##	2	1	1	1
##	347063	347064	347067	347068
##	1	1	1	1
##	347069	347071	347073	347074
##	1	1	1	1
##	347076	347077	347078	347080
##	1	4	1	2
##	347081	347082	347083	347085
##	1	7	1	1
##	347087	347088	347089	3474
##	1	6	1	1
##	347464	347466	347468	347470
##	1	1	1	1
##	347742	347743	348121	348123
##	3	1	1	1
##	348124	349201	349203	349204
##	1	1	1	1
##	349205	349206	349207	349208
##	1	1	1	1
##	349209	349210	349212	349213
##	1	1	1	1
##	349214	349215	349216	349217
##	1	1	1	1
##	349218	349219	349221	349222
##	1	1	1	1
##	349223	349224	349225	349227
##	1	1	1	1
##	349228	349231	349233	349234
##	1	1	1	1
##	349236	349237	349239	349240
##	1	2	1	1



##	349241	349242	349243	349244
##	1	1	1	1
##	349245	349246	349247	349248
##	1	1	1	1
##	349249	349251	349252	349253
##	1	1	1	1
##	349254	349256	349257	349909
##	1	1	1	4
##	349910	349912	350025	350026
##	1	1	1	1
##	350029	350034	350035	350036
##	1	1	1	1
##	350042	350043	350046	350047
##	1	1	1	1
##	350048	350050	350052	350060
##	1	1	1	1
##	350404	350406	350407	350417
##	1	1	1	1
##	35273	35281	35851	35852
##	3	2	1	1
##	358585	36209	362316	363291
##	2	1	1	3
##	363294	363592	364498	364499
##	1	1	1	1
##	364500	364506	364511	364512
##	1	1	1	1
##	364516	364846	364848	364849
##	2	1	1	2
##	364850	364851	365222	365226
##	1	1	1	1
##	36568	367226	367228	367229
##	1	2	1	1
##	367230	367231	367232	367655
##	2	1	1	1
##	368323	36864	36865	36866
##	1	1	1	1
##	368703	36928	36947	36963
##	1	2	2	1
##	36967	36973	370129	370365
##	1	2	2	2
##	370369	370370	370371	370372
##	1	1	1	1
##	370373	370375	370376	370377
##	1	1	1	1
##	371060	371110	371362	372622
##	1	3	1	1
##	373450	374746	374887	374910
##	1	1	1	1
##	376564	376566	382649	382651
##	2	1	1	1
##	382652	383121	384461	386525
##	5	1	1	1
##	392091	392092	392096	394140
##	1	1	2	1

##	4133	4134	4135	4136
##	4	1	1	1
##	4137	4138	4579	54636
##	1	1	1	2
##	5727	65303	65304	65306
##	1	1	1	1
##	6563	693	695	7267
##	1	1	1	1
##	7534	7540	7545	7546
##	2	1	1	1
##	7552	7553	7598	8471
##	1	1	1	1
##	8475	9234	A./5. 2152	A./5. 3235
##	1	1	1	1
##	A.5. 11206	A.5. 18509	A/4 45380	A/4 48871
##	1	1	1	2
##	A/4. 20589	A/4. 34244	A/4. 39886	A/5 21171
##	1	1	1	1
##	A/5 21172	A/5 21173	A/5 21174	A/5 2466
##	1	1	1	1
##	A/5 2817	A/5 3536	A/5 3540	A/5 3594
##	1	1	1	1
##	A/5 3902	A/5. 10482	A/5. 13032	A/5. 2151
##	1	1	1	1
##	A/5. 3336	A/5. 3337	A/5. 851	A/S 2816
##	2	1	1	1
##	A4. 54510	C 17369	C 4001	C 7075
##	1	1	1	1
##	C 7076	C 7077	C.A. 17248	C.A. 18723
##	1	1	1	1
##	C.A. 2315	C.A. 24579	C.A. 24580	C.A. 2673
##	2	1	1	2
##	C.A. 29178	C.A. 29395	C.A. 29566	C.A. 31026
##	1	1	1	1
##	C.A. 31921	C.A. 33111	C.A. 33112	C.A. 33595
##	3	1	2	1
##	C.A. 34260	C.A. 34651	C.A. 37671	C.A. 5547
##	1	3	2	1
##	C.A. 6212	C.A./SOTON 34068	CA 2144	CA. 2314
##	1	1	6	1
##	CA. 2343	F.C. 12750	F.C.C. 13528	F.C.C. 13529
##	7	1	1	3
##	F.C.C. 13531	Fa 265302	LINE	P/PP 3381
##	1	1	4	2
##	PC 17318	PC 17473	PC 17474	PC 17475
##	1	1	1	1
##	PC 17476	PC 17477	PC 17482	PC 17483
##	1	2	1	1
##	PC 17485	PC 17558	PC 17569	PC 17572
##	2	2	2	3
##	PC 17582	PC 17585	PC 17590	PC 17592
##	3	1	1	1
##	PC 17593	PC 17595	PC 17596	PC 17597
##	2	1	1	1

##	PC 17599	PC 17600	PC 17601	PC 17603
##	1	1	1	1
##	PC 17604	PC 17605	PC 17608	PC 17609
##	2	1	2	1
##	PC 17610	PC 17611	PC 17612	PC 17754
##	1	2	1	1
##	PC 17755	PC 17756	PC 17757	PC 17758
##	3	1	4	2
##	PC 17759	PC 17760	PC 17761	PP 4348
##	1	3	2	1
##	PP 9549	S.C./A.4. 23567	S.C./PARIS 2079	S.O./P.P. 3
##	2	1	2	2
##	S.O./P.P. 751	S.O.C. 14879	S.O.P. 1166	S.P. 3464
##	1	5	1	1
##	S.W./PP 752	SC 1748	SC/AH 29037	SC/AH 3085
##	1	1	1	1
##	SC/AH Basle 541	SC/Paris 2123	SC/PARIS 2131	SC/PARIS 2133
##	1	3	1	1
##	SC/PARIS 2146	SC/PARIS 2149	SC/Paris 2163	SC/PARIS 2167
##	1	1	1	1
##	SCO/W 1585	SO/C 14885	SOTON/O.Q. 3101305	SOTON/O.Q. 3101306
##	1	1	1	1
##	SOTON/O.Q. 3101307	SOTON/O.Q. 3101310	SOTON/O.Q. 3101311	SOTON/O.Q. 3101312
##	1	1	1	1
##	SOTON/O.Q. 392078	SOTON/O.Q. 392087	SOTON/O2 3101272	SOTON/O2 3101287
##	1	1	1	1
##	SOTON/OQ 3101316	SOTON/OQ 3101317	SOTON/OQ 392076	SOTON/OQ 392082
##	1	1	1	1
##	SOTON/OQ 392086	SOTON/OQ 392089	SOTON/OQ 392090	STON/O 2. 3101269
##	1	1	1	1
##	STON/O 2. 3101273	STON/O 2. 3101274	STON/O 2. 3101275	STON/O 2. 3101280
##	1	1	1	1
##	STON/O 2. 3101285	STON/O 2. 3101286	STON/O 2. 3101288	STON/O 2. 3101289
##	1	1	1	1
##	STON/O 2. 3101292	STON/O 2. 3101293	STON/O 2. 3101294	STON/O2. 3101271
##	1	1	1	1
##	STON/O2. 3101279	STON/O2. 3101282	STON/O2. 3101283	STON/O2. 3101290
##	2	1	1	1
##	SW/PP 751	W./C. 14258	W./C. 14263	W./C. 6607
##	1	1	1	2
##	W./C. 6608	W./C. 6609	W.E.P. 5734	W/C 14208
##	4	1	1	1
##	WE/P 5735			
##	2			

Observamos que no tiene valores vacíos y que no todos los valores son únicos.

## Cabin

```
table(titanicData$Cabin)
```

##				
##		A10	A14	A16
##	687	1	1	1
##				A19
##				1

##	A20	A23	A24	A26	A31
##	1	1	1	1	1
##	A32	A34	A36	A5	A6
##	1	1	1	1	1
##	A7	B101	B102	B18	B19
##	1	1	1	2	1
##	B20	B22	B28	B3	B30
##	2	2	2	1	1
##	B35	B37	B38	B39	B4
##	2	1	1	1	1
##	B41	B42	B49	B5	B50
##	1	1	2	2	1
##	B51 B53 B55 B57 B59 B63 B66		B58 B60	B69	B71
##	2	2	2	1	1
##	B73	B77	B78	B79	B80
##	1	2	1	1	1
##	B82 B84	B86	B94	B96 B98	C101
##	1	1	1	4	1
##	C103	C104	C106	C110	C111
##	1	1	1	1	1
##	C118	C123	C124	C125	C126
##	1	2	2	2	2
##	C128	C148	C2	C22 C26	C23 C25 C27
##	1	1	2	3	4
##	C30	C32	C45	C46	C47
##	1	1	1	1	1
##	C49	C50	C52	C54	C62 C64
##	1	1	2	1	1
##	C65	C68	C7	C70	C78
##	2	2	1	1	2
##	C82	C83	C85	C86	C87
##	1	2	1	1	1
##	C90	C91	C92	C93	C95
##	1	1	2	2	1
##	C99	D	D10 D12	D11	D15
##	1	3	1	1	1
##	D17	D19	D20	D21	D26
##	2	1	2	1	2
##	D28	D30	D33	D35	D36
##	1	1	2	2	2
##	D37	D45	D46	D47	D48
##	1	1	1	1	1
##	D49	D50	D56	D6	D7
##	1	1	1	1	1
##	D9	E10	E101	E12	E121
##	1	1	3	1	2
##	E17	E24	E25	E31	E33
##	1	2	2	1	2
##	E34	E36	E38	E40	E44
##	1	1	1	1	2
##	E46	E49	E50	E58	E63
##	1	1	1	1	1
##	E67	E68	E77	E8	F E69
##	2	1	1	2	1

```
##          F G63          F G73          F2          F33          F38
##          1          2          3          3          1
##          F4          G6          T
##          2          4          1
```

Observamos que tiene 687 observaciones sin valor.

### Embarked

```
table(titanicData$Embarked)
```

```
##
##      C   Q   S
##  2 168  77 644
```

Obtenemos que esta variable puede encontrarse de 3 maneras diferentes y tiene 2 observaciones sin valor.

Una vez que conocemos mejor las variables vamos a visualizar los valores que toma cada **variable cuantitativa**:

### Survived

```
table(titanicData$Survived)
```

```
##
##    0    1
## 549 342
```

```
342/549*100
```

```
## [1] 62.29508
```

El 62% de los pasajeros fallecieron.

### Pclass

```
table(titanicData$Pclass)
```

```
##
##    1    2    3
## 216 184 491
```

Solo hay 3 clases diferentes y la menos numerosa es la segunda.

### Age

```
table(titanicData$Age)
```

```
##
## 0.42 0.67 0.75 0.83 0.92    1    2    3    4    5    6    7    8    9   10   11
##   1    1    2    2    1    7   10    6   10    4    3    3    4    8    2    4
##  12   13   14 14.5   15   16   17   18   19   20 20.5   21   22   23 23.5   24
##   1    2    6    1    5   17   13   26   25   15    1   24   27   15    1   30
```

```
## 24.5 25 26 27 28 28.5 29 30 30.5 31 32 32.5 33 34 34.5 35
## 1 23 18 18 25 2 20 25 2 17 18 2 15 15 1 18
## 36 36.5 37 38 39 40 40.5 41 42 43 44 45 45.5 46 47 48
## 22 1 6 11 14 13 2 6 13 5 9 12 2 3 9 9
## 49 50 51 52 53 54 55 55.5 56 57 58 59 60 61 62 63
## 6 10 7 6 1 8 2 1 4 2 5 2 4 3 4 2
## 64 65 66 70 70.5 71 74 80
## 2 3 1 2 1 2 1 1
```

Observamos que hay decimales. Lo que indica que nos puede resultar interesante agrupar para que los valores de las observaciones en valores enteros.

### SibSp

```
table(titanicData$SibSp)
```

```
##
## 0 1 2 3 4 5 8
## 608 209 28 16 18 5 7
```

Hay 7 elementos diferentes y nos indica que el grueso de observaciones son de 0 hermanos y/o cónyuges.

### Parch

```
table(titanicData$Parch)
```

```
##
## 0 1 2 3 4 5 6
## 678 118 80 5 4 5 1
```

Lo más habitual es que no tuvieran niños.

### Fare

```
table(titanicData$Fare)
```

```
##
## 0 4.0125 5 6.2375 6.4375 6.45 6.4958 6.75
## 15 1 1 1 1 2 2
## 6.8583 6.95 6.975 7.0458 7.05 7.0542 7.125 7.1417
## 1 1 2 1 7 2 4 1
## 7.225 7.2292 7.25 7.3125 7.4958 7.5208 7.55 7.6292
## 12 15 13 1 3 1 4 1
## 7.65 7.725 7.7292 7.7333 7.7375 7.7417 7.75 7.775
## 4 1 1 4 2 1 34 16
## 7.7875 7.7958 7.8 7.8292 7.8542 7.875 7.8792 7.8875
## 1 6 1 2 13 1 4 1
## 7.8958 7.925 8.0292 8.05 8.1125 8.1375 8.1583 8.3
## 38 18 1 43 1 1 1 1
## 8.3625 8.4042 8.4333 8.4583 8.5167 8.6542 8.6625 8.6833
## 1 1 1 1 1 1 13 1
## 8.7125 8.85 9 9.2167 9.225 9.35 9.475 9.4833
## 1 1 2 1 2 2 1 1
```

##	9.5	9.5875	9.825	9.8375	9.8417	9.8458	10.1708	10.4625
##	9	2	2	1	1	1	1	2
##	10.5	10.5167	11.1333	11.2417	11.5	12	12.275	12.2875
##	24	1	3	2	4	1	1	1
##	12.35	12.475	12.525	12.65	12.875	13	13.4167	13.5
##	3	4	1	1	1	42	1	4
##	13.7917	13.8583	13.8625	14	14.1083	14.4	14.4542	14.4583
##	1	1	1	1	1	2	7	3
##	14.5	15	15.0458	15.05	15.1	15.2458	15.5	15.55
##	7	1	1	1	1	5	8	1
##	15.7417	15.75	15.85	15.9	16	16.1	16.7	17.4
##	2	1	4	2	1	9	2	1
##	17.8	18	18.75	18.7875	19.2583	19.5	19.9667	20.2125
##	2	3	3	2	4	2	2	2
##	20.25	20.525	20.575	21	21.075	21.6792	22.025	22.3583
##	2	3	2	6	4	1	1	2
##	22.525	23	23.25	23.45	24	24.15	25.4667	25.5875
##	1	4	2	2	2	8	4	1
##	25.925	25.9292	26	26.25	26.2833	26.2875	26.3875	26.55
##	1	2	31	6	1	3	1	15
##	27	27.7208	27.75	27.9	28.5	28.7125	29	29.125
##	2	5	4	6	1	1	2	5
##	29.7	30	30.0708	30.5	30.6958	31	31.275	31.3875
##	3	6	2	5	2	3	7	4
##	32.3208	32.5	33	33.5	34.0208	34.375	34.6542	35
##	1	1	3	1	1	4	1	1
##	35.5	36.75	37.0042	38.5	39	39.4	39.6	39.6875
##	4	2	2	1	4	1	2	6
##	40.125	41.5792	42.4	46.9	47.1	49.5	49.5042	50
##	1	3	1	6	1	1	2	1
##	50.4958	51.4792	51.8625	52	52.5542	53.1	55	55.4417
##	1	1	2	7	3	5	2	1
##	55.9	56.4958	56.9292	57	57.9792	59.4	61.175	61.3792
##	2	7	2	2	2	1	1	1
##	61.9792	63.3583	65	66.6	69.3	69.55	71	71.2833
##	1	1	2	2	2	7	2	1
##	73.5	75.25	76.2917	76.7292	77.2875	77.9583	78.2667	78.85
##	5	1	1	3	2	3	2	2
##	79.2	79.65	80	81.8583	82.1708	83.1583	83.475	86.5
##	4	3	2	1	2	3	2	3
##	89.1042	90	91.0792	93.5	106.425	108.9	110.8833	113.275
##	2	4	2	2	2	2	4	3
##	120	133.65	134.5	135.6333	146.5208	151.55	153.4625	164.8667
##	4	2	2	3	2	4	3	2
##	211.3375	211.5	221.7792	227.525	247.5208	262.375	263	512.3292
##	3	1	1	4	2	2	4	3

No hay valores vacíos.

## 2. Integración y selección de los datos de interés a analizar

Hay variables que no vamos a necesitar:

- PassengerId

- Name
- Ticket - Fare - Cabin
- Embarked

```
titanicData<-titanicData[,-c(12)]
titanicData<-titanicData[,-c(11)]
titanicData<-titanicData[,-c(10)]
titanicData<-titanicData[,-c(9)]
titanicData<-titanicData[,-c(4)]
titanicData<-titanicData[,-c(1)]
```

Esto significa que nos quedamos solo con 6 variables.

```
summary(titanicData)
```

```
##      Survived      Pclass      Sex      Age
##  Min.   :0.0000   Min.   :1.000   Length:891   Min.   : 0.42
## 1st Qu.:0.0000   1st Qu.:2.000   Class :character   1st Qu.:20.12
## Median :0.0000   Median :3.000   Mode  :character   Median :28.00
## Mean   :0.3838   Mean   :2.309                Mean   :29.70
## 3rd Qu.:1.0000   3rd Qu.:3.000                3rd Qu.:38.00
## Max.   :1.0000   Max.   :3.000                Max.   :80.00
##                                     NA's   :177
##      SibSp      Parch
##  Min.   :0.000   Min.   :0.0000
## 1st Qu.:0.000   1st Qu.:0.0000
## Median :0.000   Median :0.0000
## Mean   :0.523   Mean   :0.3816
## 3rd Qu.:1.000   3rd Qu.:0.0000
## Max.   :8.000   Max.   :6.0000
##
```

### 3. Limpieza de los datos

#### 3.1 ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?

Como hemos identificado anteriormente la única variable que tiene valores nulos es *Age*.

```
colSums(is.na(titanicData))
```

```
## Survived  Pclass      Sex      Age      SibSp      Parch
##         0         0         0      177         0         0
```

```
colSums(titanicData=="")
```

```
## Survived  Pclass      Sex      Age      SibSp      Parch
##         0         0         0      NA         0         0
```



Hay un total de 177 pasajeros que no tienen edad.

```
177/891*100
```

```
## [1] 19.86532
```

Es casi un 20% de las observaciones, esto significa que la cantidad de datos que faltan son **muy significativos**, por lo que tenemos que tratarlos. Asignamos la edad media de cada pasajero teniendo en cuenta su sexo.

```
# Ponemos la media a los valores vacíos de la variable "Age"
women_df <- titanicData[titanicData$Sex == "female",]
women_mean_age <- mean(titanicData$Age, na.rm= TRUE)
print(women_mean_age)
```

```
## [1] 29.69912
```

```
men_df <- titanicData[titanicData$Sex == "male",]
men_mean_age <- mean(men_df$Age, na.rm= TRUE)
print(men_mean_age)
```

```
## [1] 30.72664
```

```
titanicData$Age[is.na(titanicData$Age) & titanicData$Sex=="female" ] <- women_mean_age
titanicData$Age[is.na(titanicData$Age) & titanicData$Sex=="male" ] <- men_mean_age
```

## 3.2. Identificación y tratamiento de valores extremos.

Vamos a analizar los valores extremos, la mejor manera de identificarlos es a través de los boxplot:

```
library(reshape)
```

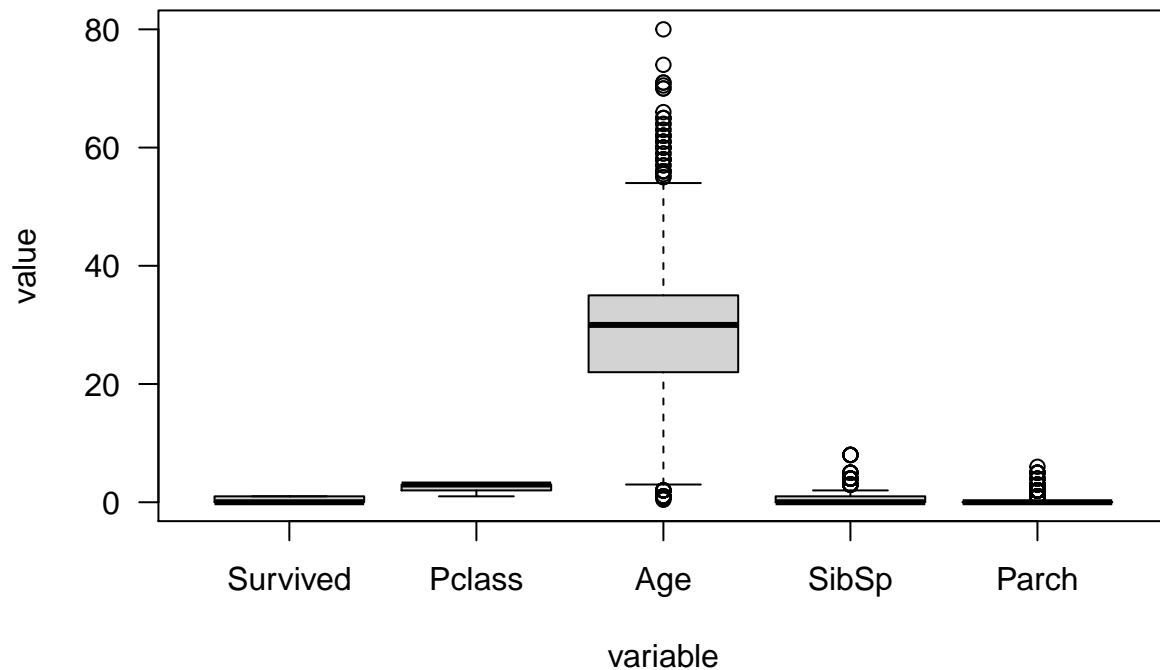
```
##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:dplyr':
##
##      rename
```

```
boxplot(data=melt(titanicData), value~variable, las=1, main="Boxplot de todos los atributos")
```

```
## Using Sex as id variables
## Using Sex as id variables
```

## Boxplot de todos los atributos



```
summary(titanicData)
```

```
##      Survived      Pclass      Sex      Age
##  Min.   :0.0000   Min.    :1.000   Length:891   Min.    : 0.42
## 1st Qu.:0.0000   1st Qu.:2.000   Class  :character 1st Qu.:22.00
## Median :0.0000   Median :3.000   Mode   :character Median :30.00
## Mean   :0.3838   Mean    :2.309                Mean   :29.84
## 3rd Qu.:1.0000   3rd Qu.:3.000                3rd Qu.:35.00
## Max.   :1.0000   Max.    :3.000                Max.   :80.00
##      SibSp      Parch
##  Min.   :0.000   Min.    :0.0000
## 1st Qu.:0.000   1st Qu.:0.0000
## Median :0.000   Median :0.0000
## Mean   :0.523   Mean    :0.3816
## 3rd Qu.:1.000   3rd Qu.:0.0000
## Max.   :8.000   Max.    :6.0000
```

Observamos que hay 3 variables con **valores extremos**. Estas variables son: - Age  
- SibSp  
- Parch

Podemos plantearnos que tanto *SibSp* como *Parch* son factores, por lo que realmente no son **valores atípicos**, por lo que realmente el enfoque tenemos que ponerlo en *Age*. *Age* tiene los valores correctos ya que corresponde con edades lógicas, por lo que procedemos a dejarla correctamente.

## 4. Análisis de los datos

### 4.1 Selección de los grupos de datos que quiere analizar/comparar (planificación de los análisis a aplicar)

Antes de tomar una decisión de los grupos a comparar vamos a mirar la correlación, para ello vamos a transformar la variable *sex* a tipo integer.

```
table(titanicData$Sex)
```

```
##  
## female    male  
##      314     577
```

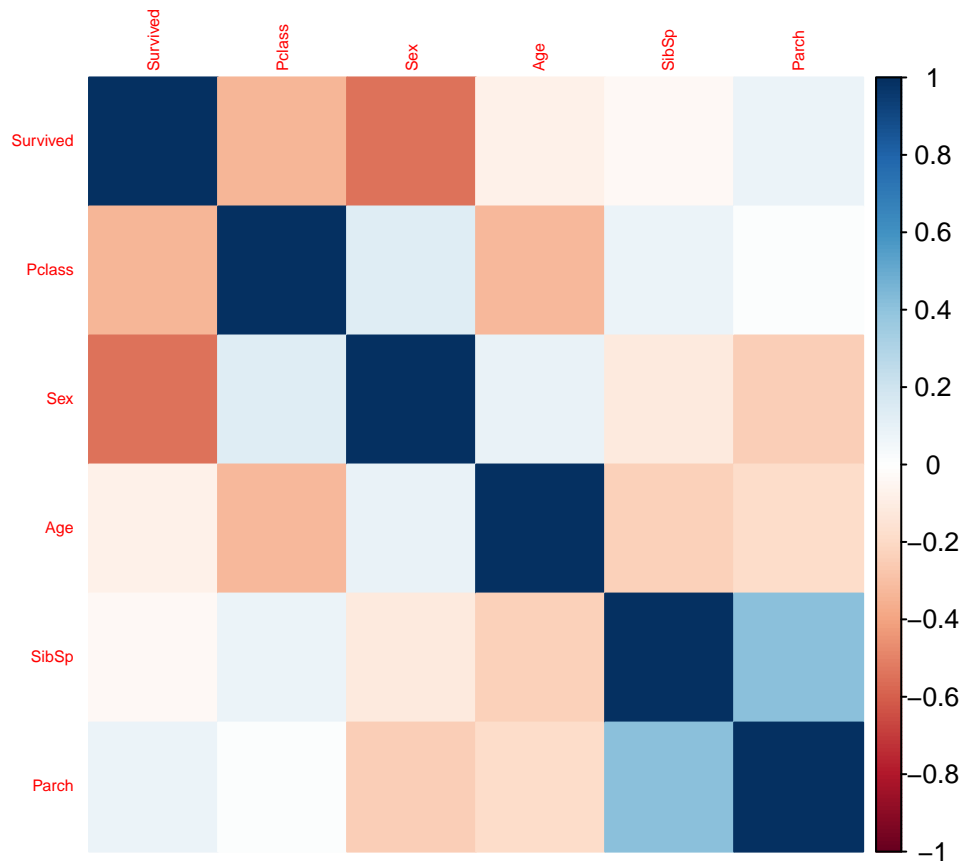
```
titanicData$Sex <- factor(titanicData$Sex, levels=c("female","male"), labels=c(1, 2))  
titanicData$Sex<-as.integer(titanicData$Sex)  
table(titanicData$Sex)
```

```
##  
##    1    2  
## 314 577
```

```
library(corrplot)
```

```
## corrplot 0.89 loaded
```

```
corrplot(cor(titanicData), method="color", tl.cex = 0.5)
```



Gracias a la correlación podemos mirar si hay cierta relación en los comportamientos de una variable respecto a las demás. La figura anterior nos indica que es interesante comparar:

- Survived con Pclass
- Survived con Sex
- Age con Pclass
- Parch con SibSp

## 4.2 Comprobación de la normalidad y homogeneidad de la varianza

### 4.2.1 Comprobación de la normalidad

Empleamos el test de Shapiro-Wilk para comprobar si se encuentra distribuida de manera normal. Generamos un contraste de Hipótesis.

$H_0$ : Nuestra hipótesis nula es que los datos provienen de una distribución normal.

$H_1$ : Nuestra hipótesis alternativa es que los datos no provienen de una distribución normal.

```
titanicData$Age<-as.numeric(titanicData$Age)
shapiro.test(titanicData$Age)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  titanicData$Age
## W = 0.9625, p-value = 2.392e-14
```

```
shapiro.test(titanicData$SibSp)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  titanicData$SibSp  
## W = 0.51297, p-value < 2.2e-16
```

```
shapiro.test(titanicData$Parch)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  titanicData$Parch  
## W = 0.53281, p-value < 2.2e-16
```

Hemos aplicado el test en las variables que tienen más de 3 tipos de valores diferentes. El test nos indica que p-value es inferior al 5%, por lo que podemos **afirmar la hipótesis alternativa**, esto significa que los datos no se distribuyen de una manera normal.

#### 4.2.2 Homogeneidad de la varianza

Para comprobar la homogeneidad de varianzas entre poblaciones empleamos el test de Fligner-Killeen.

##### Survived con Pclass

Vamos a comprobar si la varianza de hombres y mujeres es la misma dentro de la variable Survived.

$H_0$ : Nuestra hipótesis nula es que la varianza entre clases es la misma en Survived.

$H_1$ : Nuestra hipótesis alternativa es que la varianza entre clases es diferente en Survived.

```
fligner.test(Survived ~ Pclass, titanicData)
```

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data:  Survived by Pclass  
## Fligner-Killeen:med chi-squared = 35.766, df = 2, p-value = 1.712e-08
```

p-value es inferior al 5% lo que nos indica que la varianza de ambas muestras son diferentes, lo que significa que rechazamos la hipótesis nula y **nos quedamos con la hipótesis alternativa**.

##### Survived con Sex

Vamos a comprobar si la varianza de hombres y mujeres es la misma dentro de la variable Survived.

$H_0$ : Nuestra hipótesis nula es que la varianza entre hombres y mujeres es la misma en Survived.

$H_1$ : Nuestra hipótesis alternativa es que la varianza entre hombres y mujeres es diferente en Survived.

```
fligner.test(Survived ~ Sex, titanicData)
```

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data:  Survived by Sex  
## Fligner-Killeen:med chi-squared = 5.7729, df = 1, p-value = 0.01627
```

p-value es inferior al 5% lo que nos indica que la varianza de ambas muestras son diferentes, lo que significa que rechazamos la hipótesis nula y **nos quedamos con la hipótesis alternativa**.

#### Age con Pclass

Vamos a comprobar si la varianza de clases es la misma dentro de la variable Age.

$H_0$ : Nuestra hipótesis nula es que la varianza entre clases es la misma en Age.

$H_1$ : Nuestra hipótesis alternativa es que la varianza entre clases es diferente en Age.

```
fligner.test(Age ~ Pclass, titanicData)
```

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data: Age by Pclass  
## Fligner-Killeen:med chi-squared = 28.794, df = 2, p-value = 5.59e-07
```

p-value es inferior al 5% lo que nos indica que la varianza de ambas muestras son diferentes, lo que significa que rechazamos la hipótesis nula y **nos quedamos con la hipótesis alternativa**.

**Parch con SibSp** Vamos a comprobar si la varianza de número de esposas/hermanos es la misma dentro de la variable Parch.

$H_0$ : Nuestra hipótesis nula es que la varianza entre número de esposas/hermanos es la misma en Parch.

$H_1$ : Nuestra hipótesis alternativa es que la varianza entre número de esposas/hermanos es diferente en Parch.

```
fligner.test(Parch ~ SibSp, titanicData)
```

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data: Parch by SibSp  
## Fligner-Killeen:med chi-squared = 106.38, df = 6, p-value < 2.2e-16
```

p-value es inferior al 5% lo que nos indica que la varianza de ambas muestras son diferentes, lo que significa que rechazamos la hipótesis nula y **nos quedamos con la hipótesis alternativa**.

**4.3 Aplicación de pruebas estadísticas para comparar los grupos de datos.** En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.

#### 4.3.1 Regresiones

##### Survived con Pclass

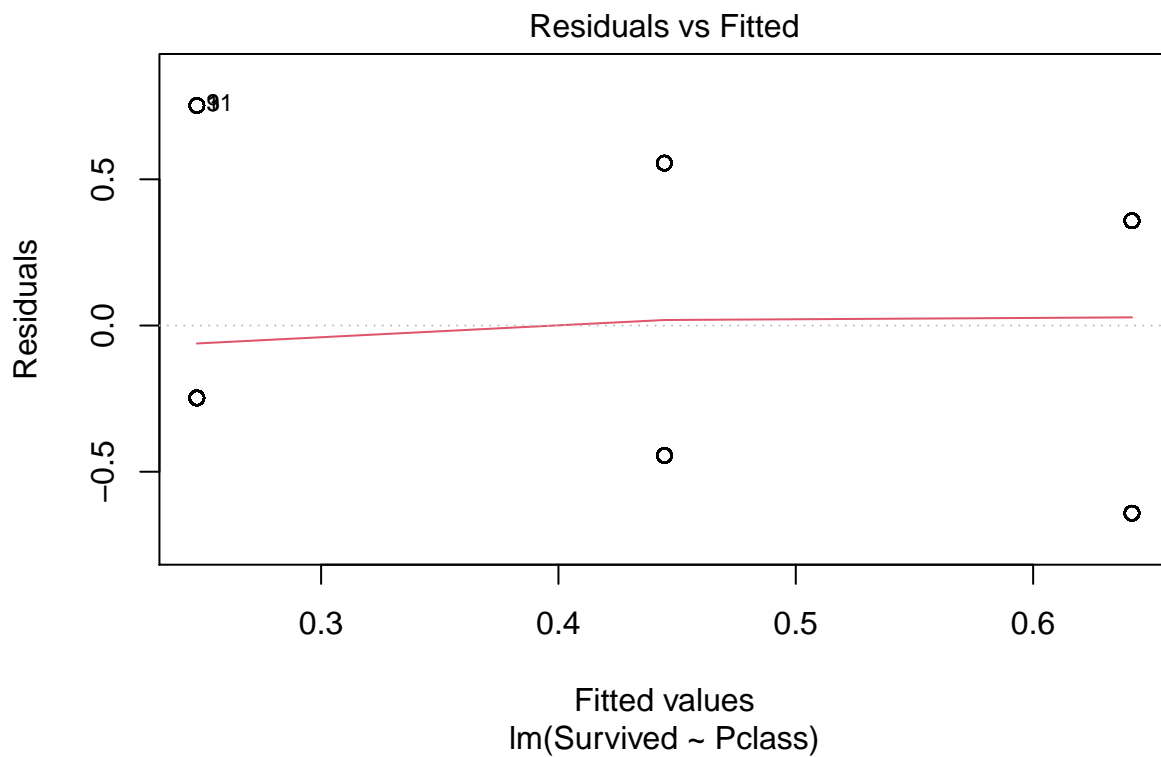
```
regsurvpclass<-lm(Survived~Pclass, titanicData)
```

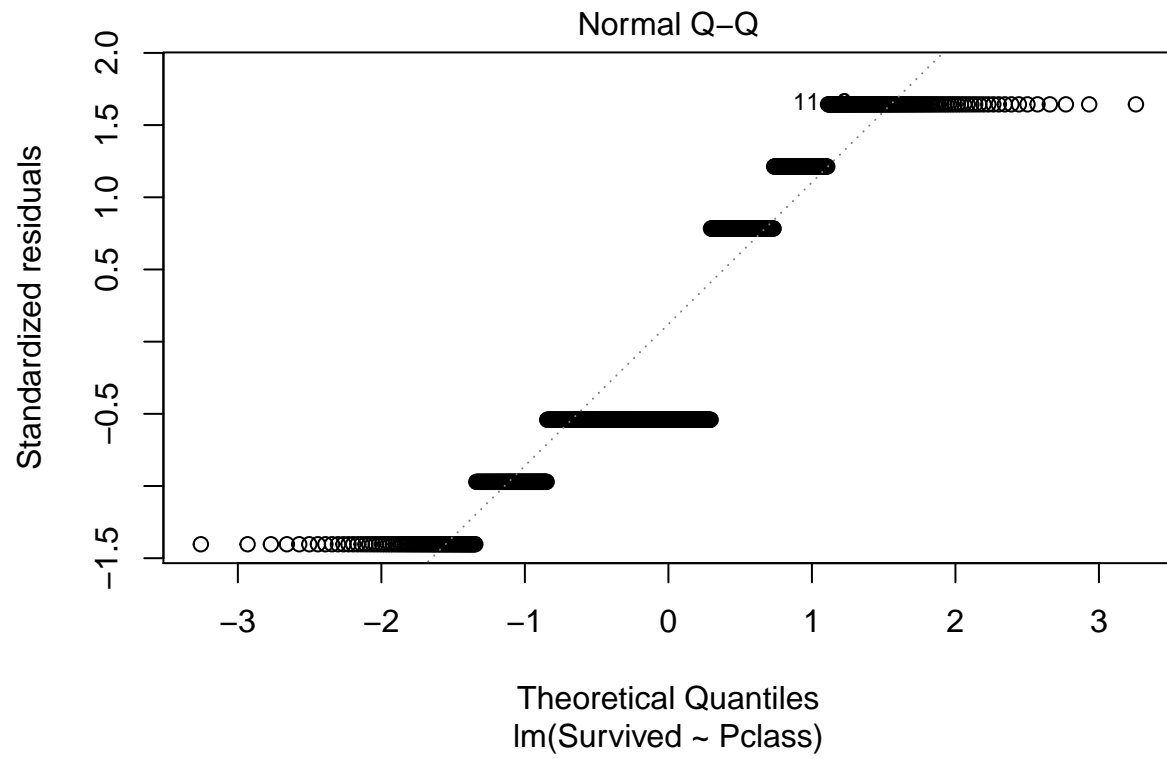
```
summary(regsurvpclass)
```

```
##  
## Call:  
## lm(formula = Survived ~ Pclass, data = titanicData)
```

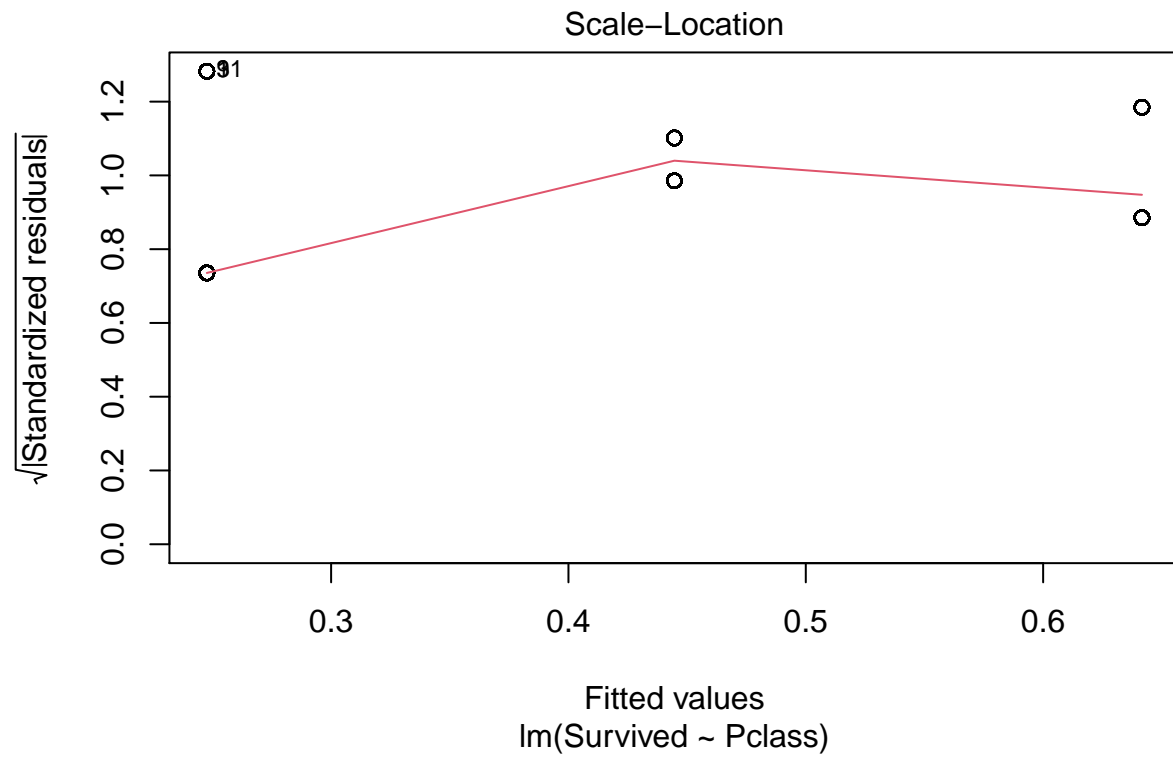
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6416 -0.2476 -0.2476  0.3584  0.7524
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.83863    0.04510   18.60  <2e-16 ***
## Pclass      -0.19700    0.01837  -10.72  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4581 on 889 degrees of freedom
## Multiple R-squared:  0.1146, Adjusted R-squared:  0.1136
## F-statistic: 115 on 1 and 889 DF, p-value: < 2.2e-16
```

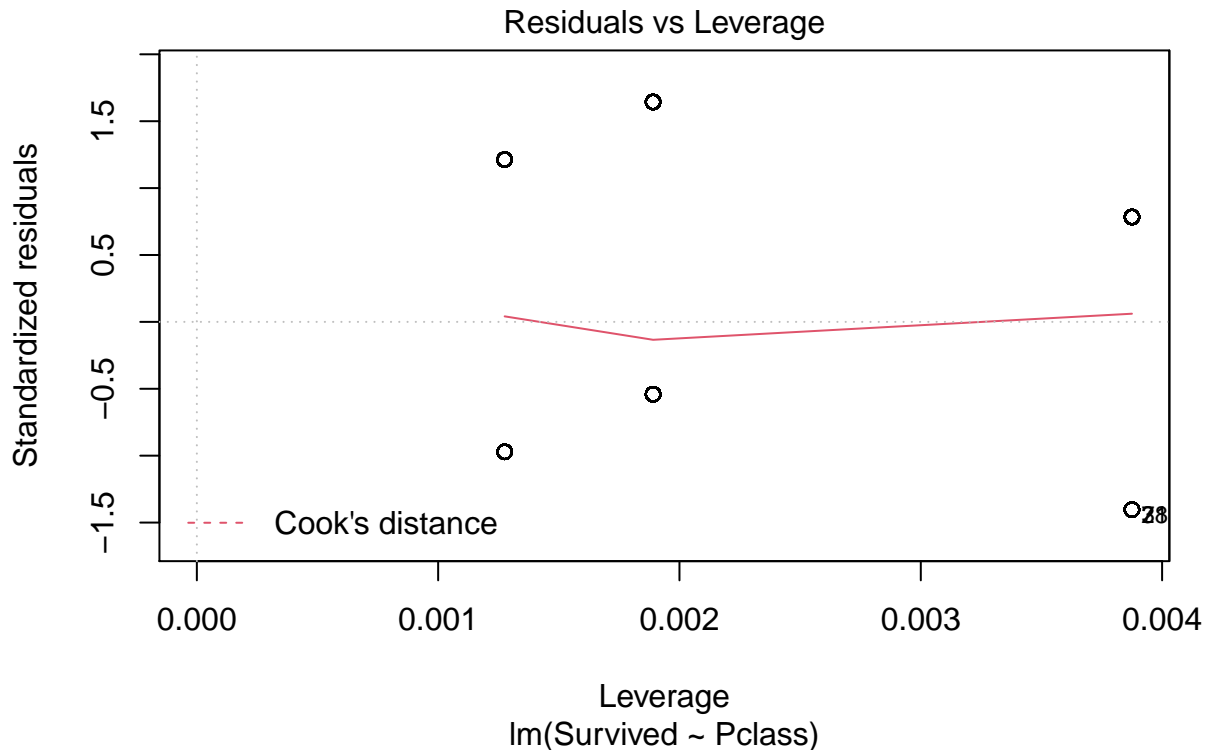
```
plot(regsurvpclass)
```











**Multiple R-squared** es 0.1146 esto significa que no podemos emplear la variable *Pclass* para predecir *Survived*. Hay que tener en cuenta que lo ideal es que sea lo más cercano a 1 y en este caso es mucho más cercano a 0.

**P-value** es inferior a  $2.2e-16$ , esto es muy buena señal, los estudios suelen realizarse con un nivel de significación del 5% y en este caso es muy inferior, por lo que podemos estar contento porque es una muy buena regresión.

### Survived con Sex

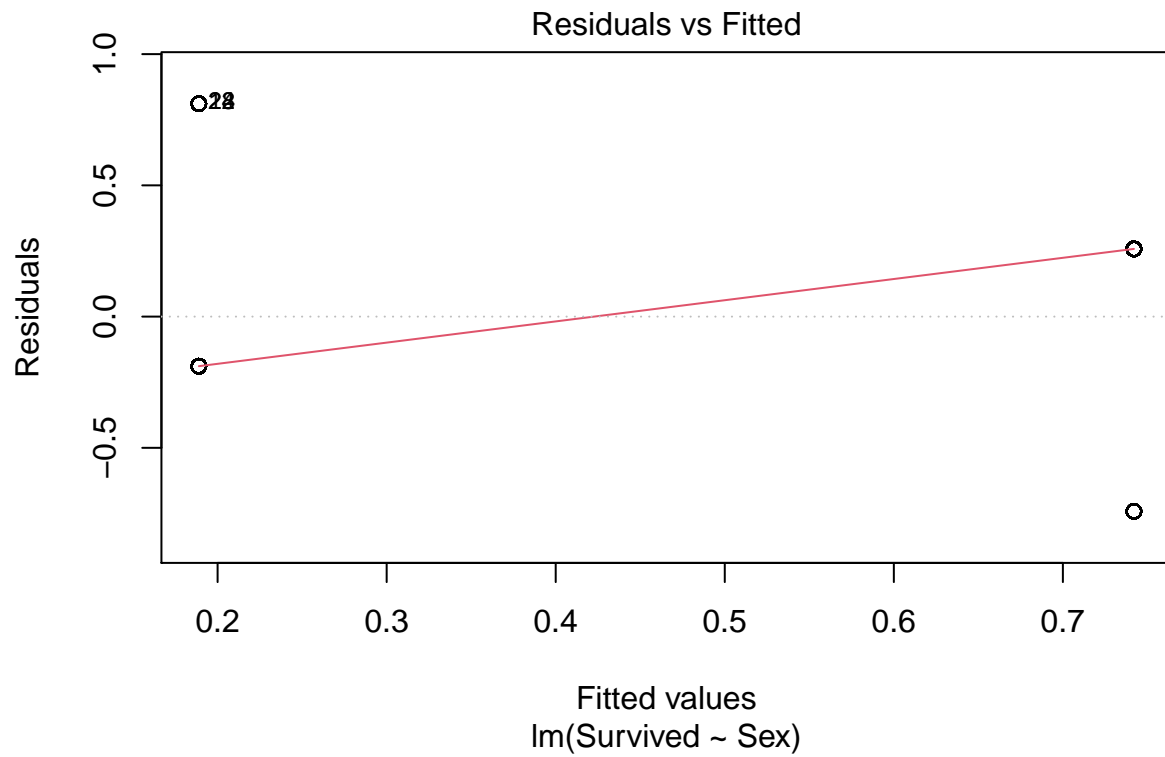
```
regsexsurv<-lm(Survived~Sex, titanicData)
```

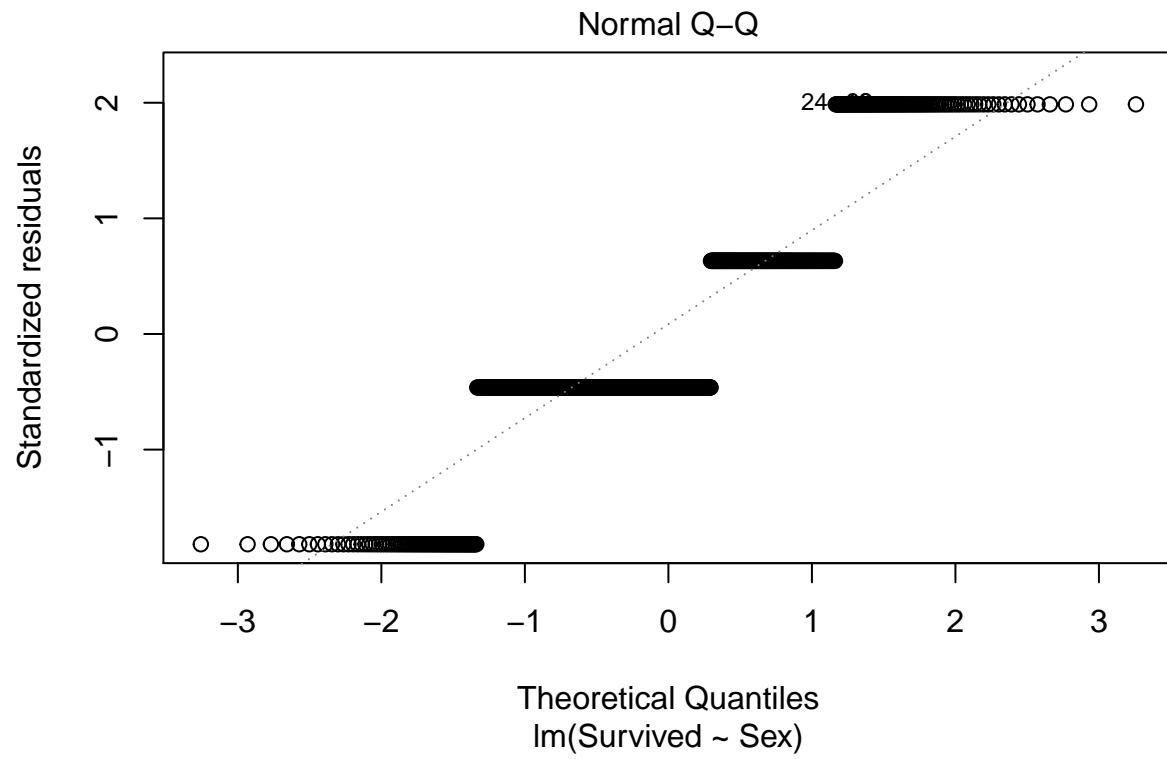
```
summary(regsexsurv)
```

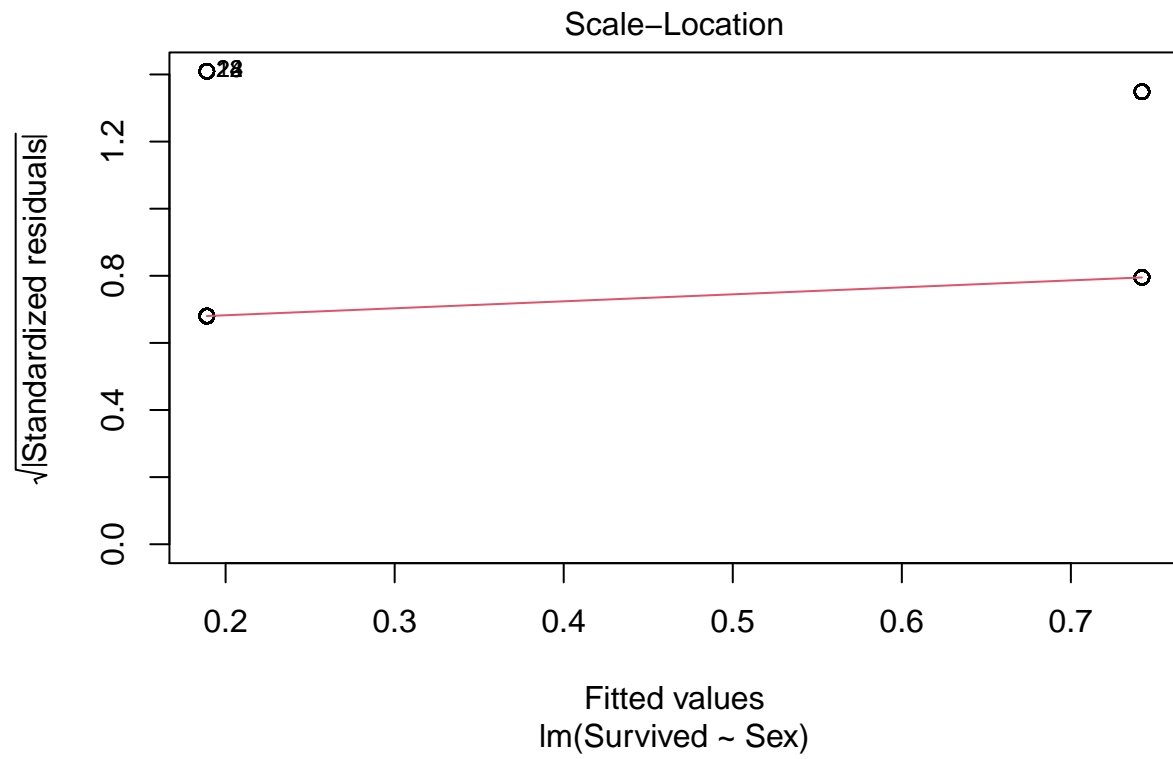
```
##
## Call:
## lm(formula = Survived ~ Sex, data = titanicData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7420 -0.1889 -0.1889  0.2580  0.8111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.29517    0.04917   26.34  <2e-16 ***
## Sex         -0.55313    0.02866  -19.30  <2e-16 ***
## ---
```

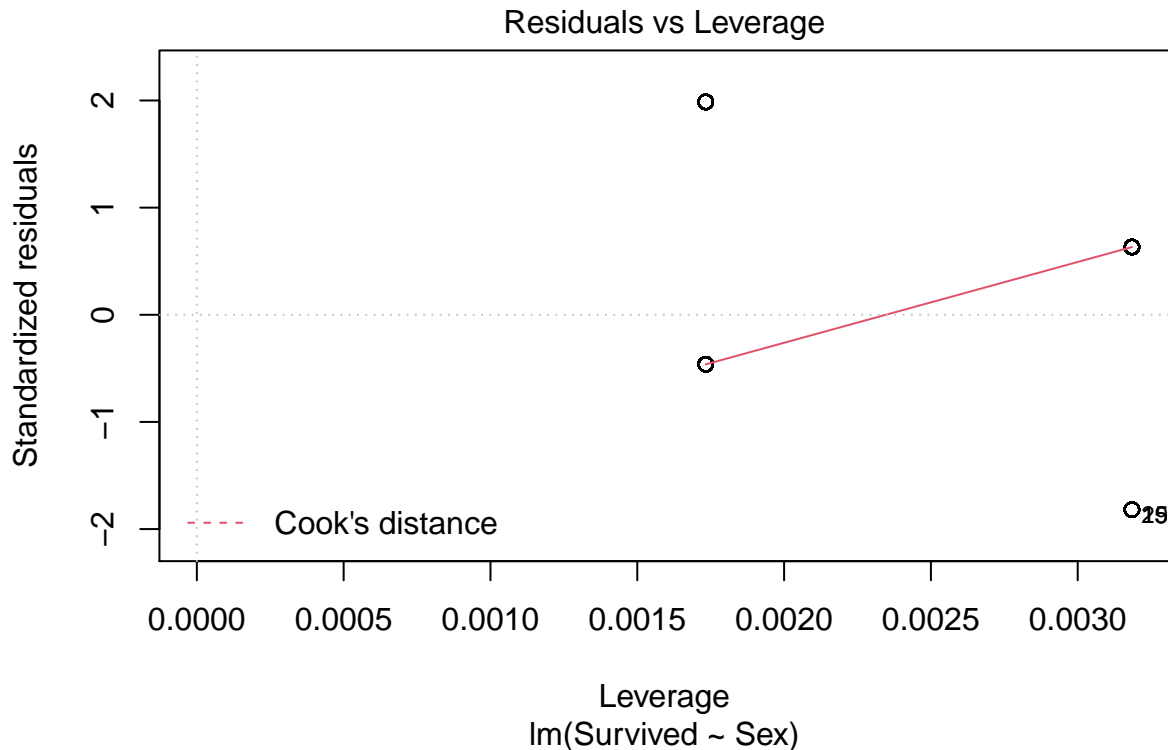
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4087 on 889 degrees of freedom
## Multiple R-squared:  0.2952, Adjusted R-squared:  0.2944
## F-statistic: 372.4 on 1 and 889 DF,  p-value: < 2.2e-16
```

```
plot(regsexsurv)
```









**Multiple R-squared** es 0.2952. Aque esta vez es un valor algo mayor que en el caso anterior, aquí seguimos sin poder emplear la variable *Sex* para predecir *Survived* de una forma precisa, ya que seguimos teniendo un valor muy lejano a 1.

**P-value** vuelve a ser inferior a  $2.2e-16$ , así que volvemos a tener una muy buena regresión al obtener un nivel de significación muy inferior al 5%.

### Age con Pclass

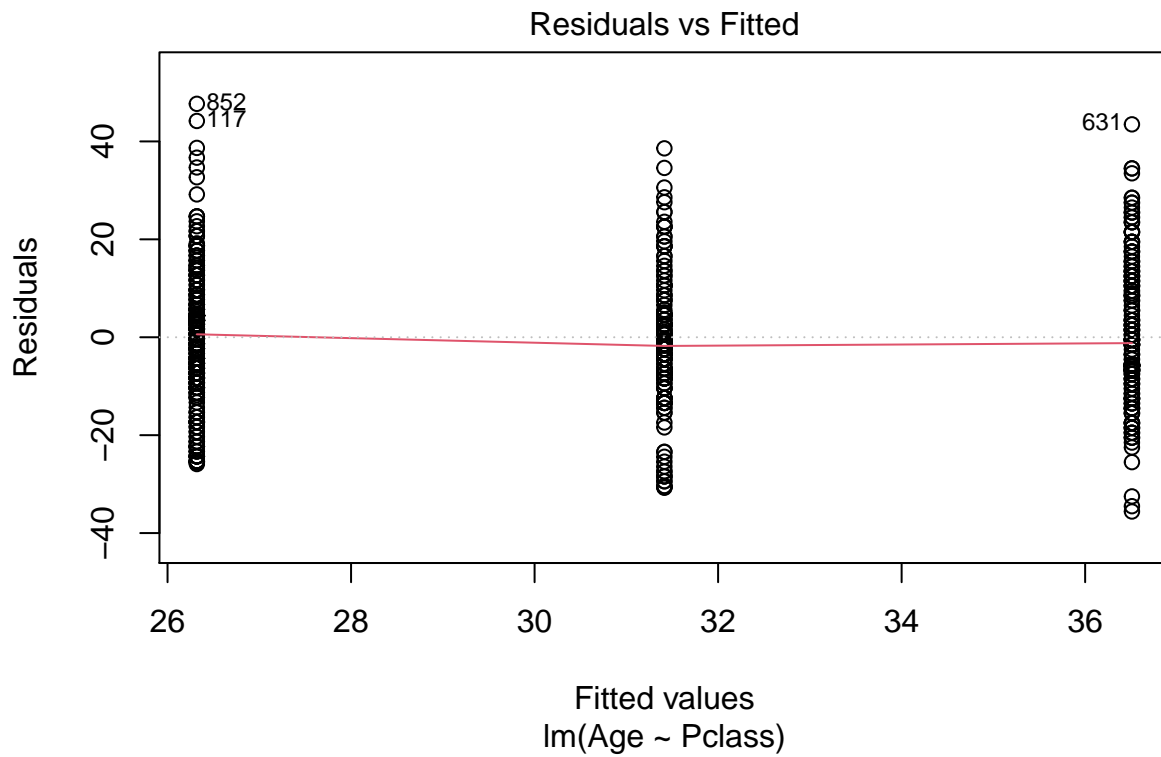
```
regagepclass<-lm(Age~Pclass, titanicData)
```

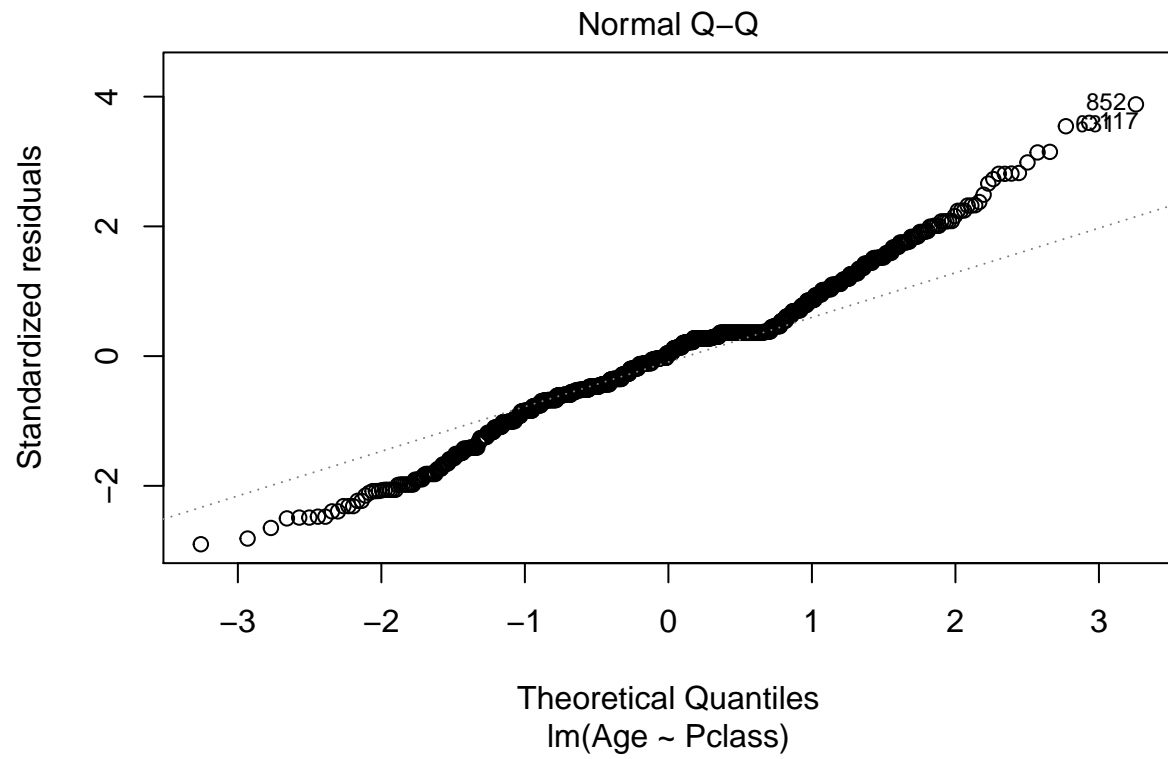
```
summary(regagepclass)
```

```
##
## Call:
## lm(formula = Age ~ Pclass, data = titanicData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -35.590  -6.811   0.585   4.585  47.681
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   41.606     1.210   34.37  <2e-16 ***
## Pclass        -5.095     0.493  -10.34  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

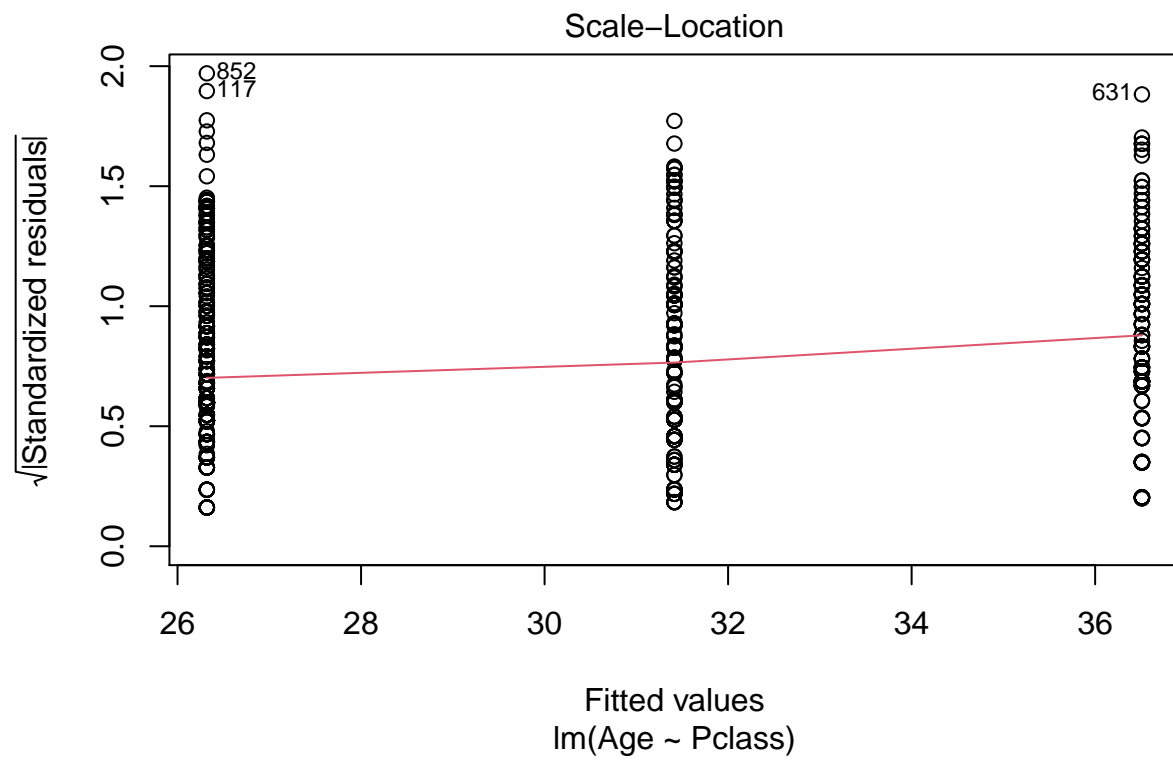
```
##
## Residual standard error: 12.3 on 889 degrees of freedom
## Multiple R-squared:  0.1073, Adjusted R-squared:  0.1063
## F-statistic: 106.8 on 1 and 889 DF,  p-value: < 2.2e-16
```

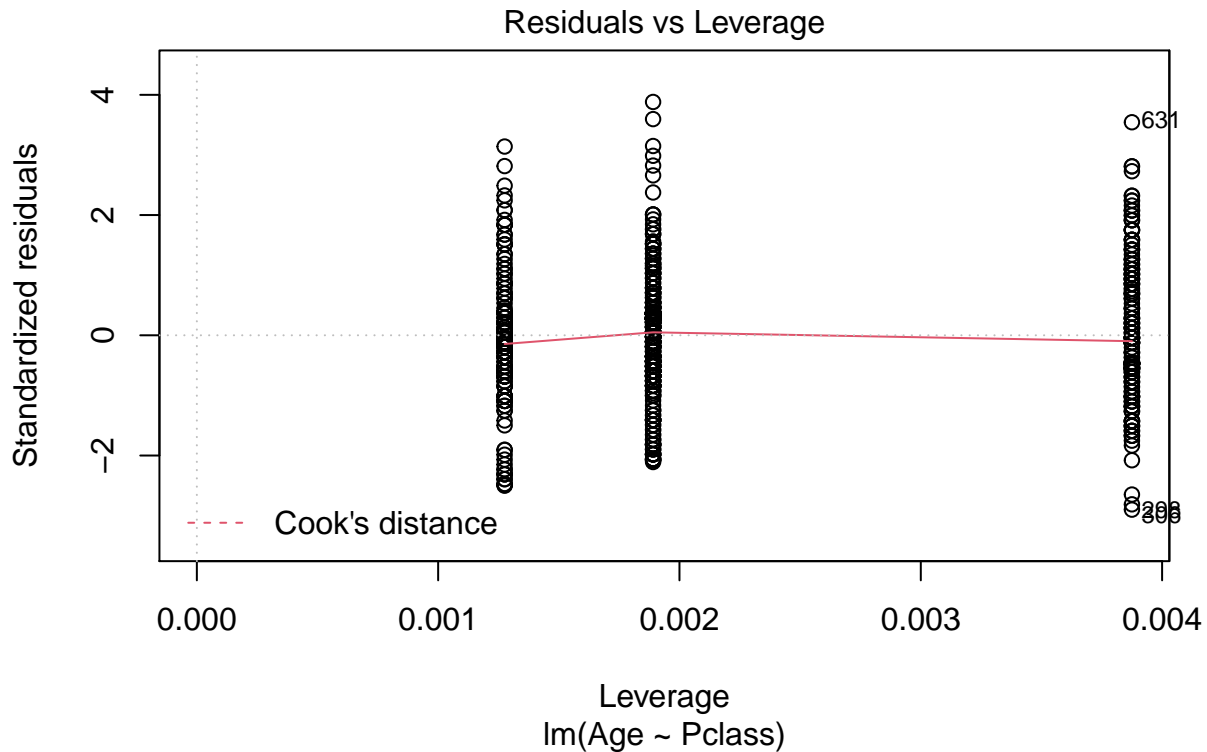
```
plot(regagepclass)
```











**Multiple R-squared** es 0.1073. Aquí obtenemos el peor valor hasta ahora, así que por descontento que no podremos emplear la variable *Age* para predecir *Pclass*.

Sin embargo, volvemos a tener una buena regresión encontrándonos con un **P-value** muy inferior a 2.2e-16.

### Parch con SibSp

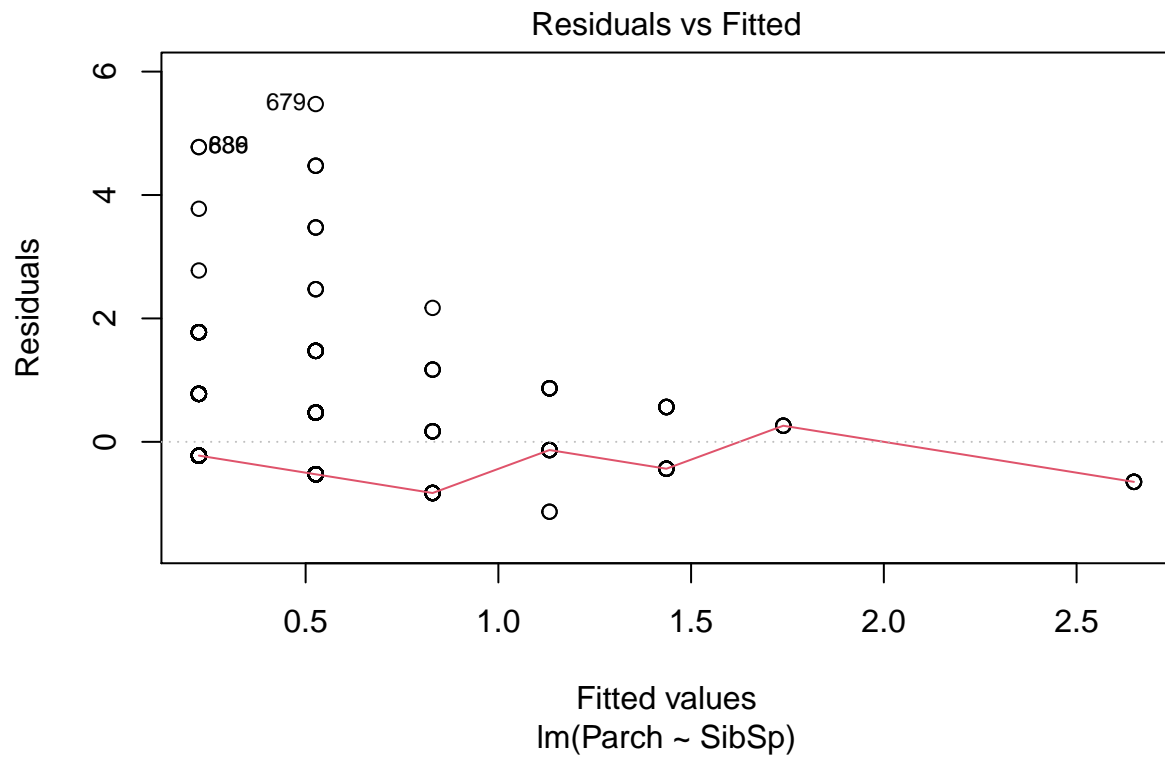
```
regparchsibsp<-lm(Parch~SibSp, titanicData)
```

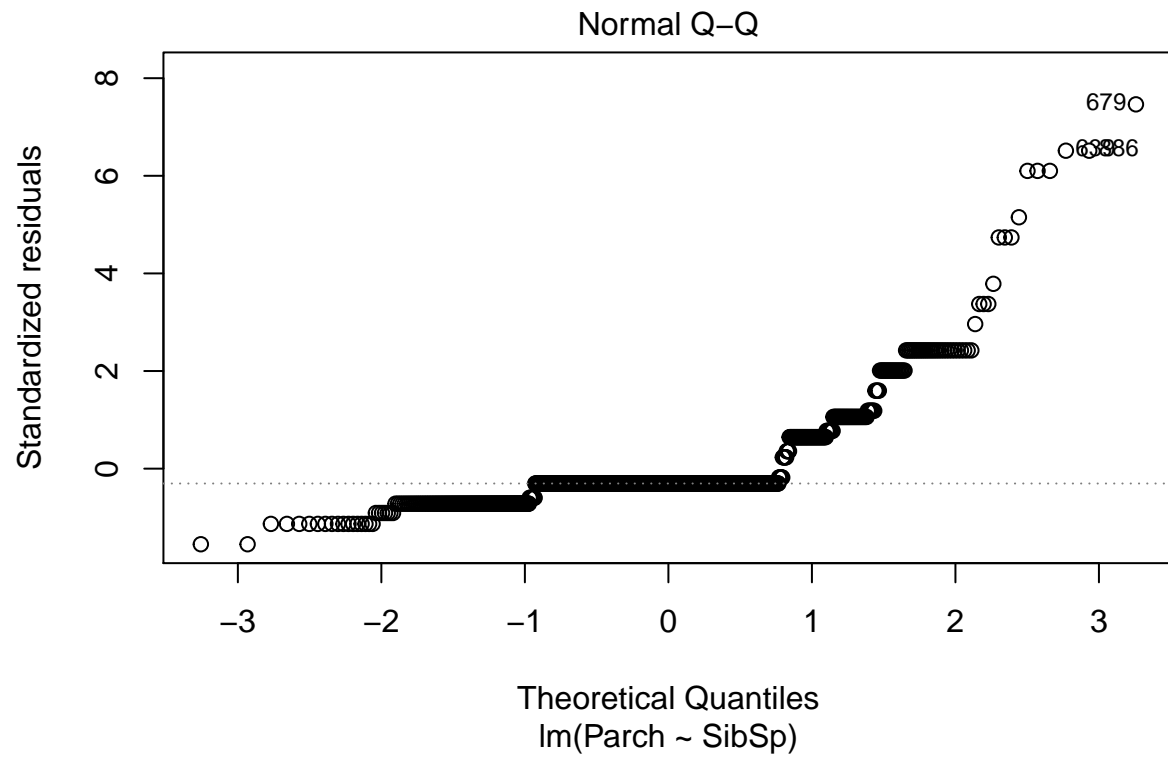
```
summary(regparchsibsp)
```

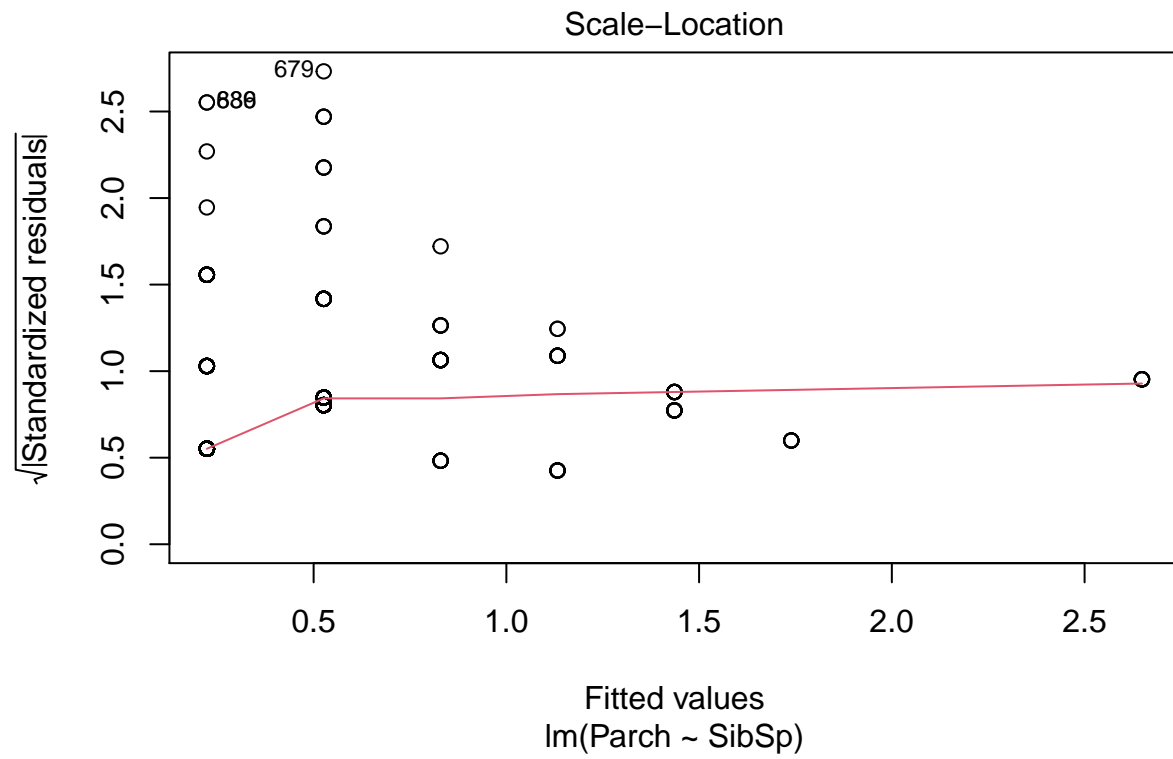
```
##
## Call:
## lm(formula = Parch ~ SibSp, data = titanicData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.133  -0.223  -0.223  -0.223   5.474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.22300    0.02721   8.195 8.7e-16 ***
## SibSp        0.30323    0.02231  13.594 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7338 on 889 degrees of freedom
```

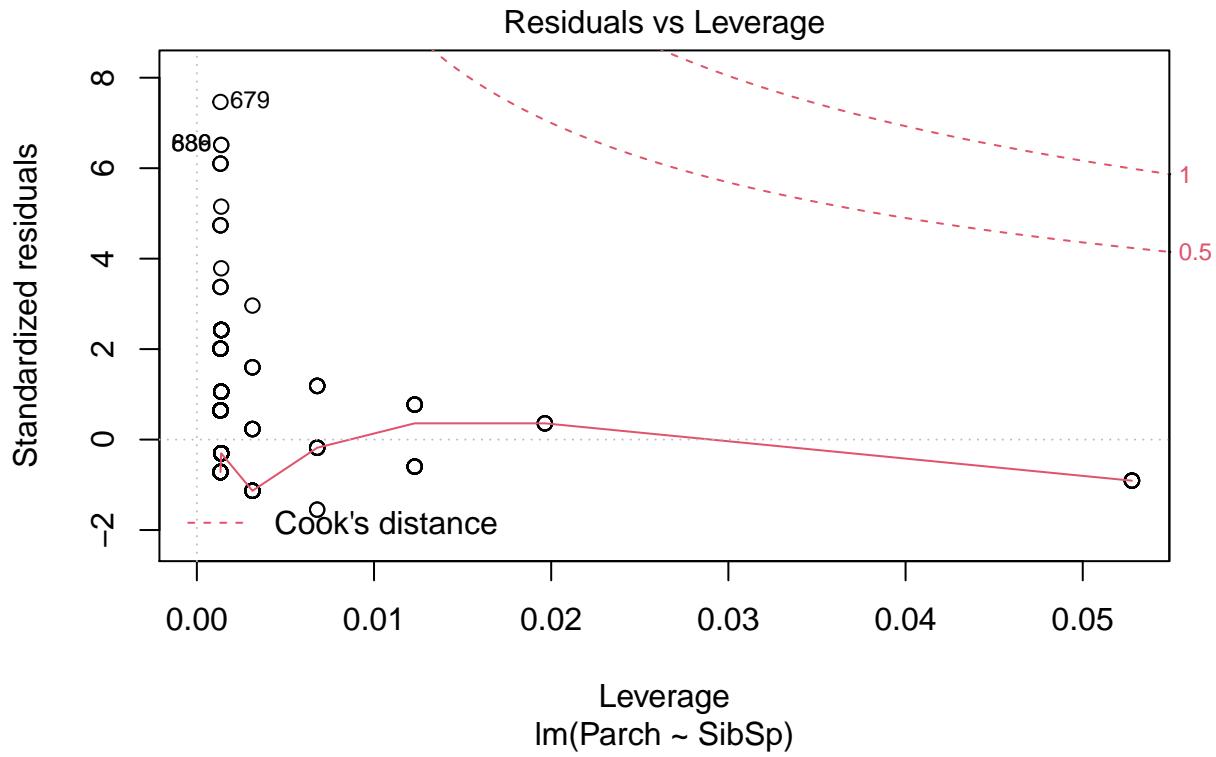
```
## Multiple R-squared:  0.1721, Adjusted R-squared:  0.1712
## F-statistic: 184.8 on 1 and 889 DF,  p-value: < 2.2e-16
```

```
plot(regparchsibsp)
```









**Multiple R-squared** es 0.1721. Al igual que en todas las regresiones anteriores, obtenemos que no podemos emplear la variable *Parch* para predecir *SibSp*.

**P-value** vuelve a ser inferior a  $2.2e-16$ , así que por lo menos obtenemos una muy buena regresión otra vez.

#### 4.3.2 Correlaciones

En este apartado vamos a llevar a cabo un análisis de correlación entre las diferentes variables del dataset, para de esta forma identificar como se influyen entre ellas.

```
cor(titanicData)
```

```
##           Survived      Pclass        Sex         Age         SibSp
## Survived  1.00000000 -0.33848104 -0.54335138 -0.07554602 -0.03532250
## Pclass    -0.33848104  1.00000000  0.13190049 -0.32752776  0.08308136
## Sex       -0.54335138  0.13190049  1.00000000  0.09223687 -0.11463081
## Age       -0.07554602 -0.32752776  0.09223687  1.00000000 -0.23381356
## SibSp     -0.03532250  0.08308136 -0.11463081 -0.23381356  1.00000000
## Parch     0.08162941  0.01844267 -0.24548896 -0.18290288  0.41483770
##
##           Parch
## Survived  0.08162941
## Pclass    0.01844267
## Sex       -0.24548896
## Age       -0.18290288
## SibSp     0.41483770
## Parch     1.00000000
```

Vamos a destacar las correlaciones superiores al  $\pm 0,30$ : Survived-Pclass, Survived-Sex, Age-Pclass y SibSp-Parch. Recordemos que el signo significa que ambas crecen o decrecen a la vez y si el símbolo es negativo significa que cuando una crece la otra decrece o a la inversa.

### 4.3.3 KMEANS

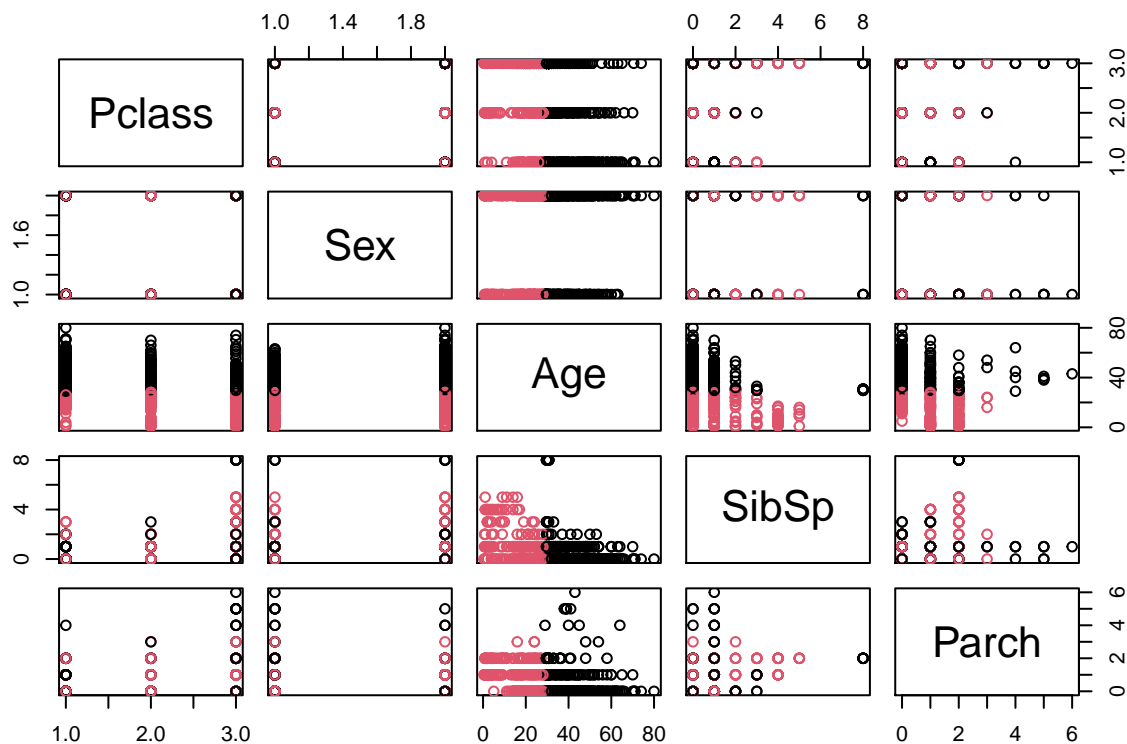
En este apartado vamos a emplear el algoritmo kmeans para agrupar a los pasajeros del barco y así tratar de predecir si sobrevivieron o no. Por descontado, como lo que queremos predecir es la supervivencia de los individuos ya sabemos de antemano que el número de clusters ideal será de 2.

```
library(cluster)

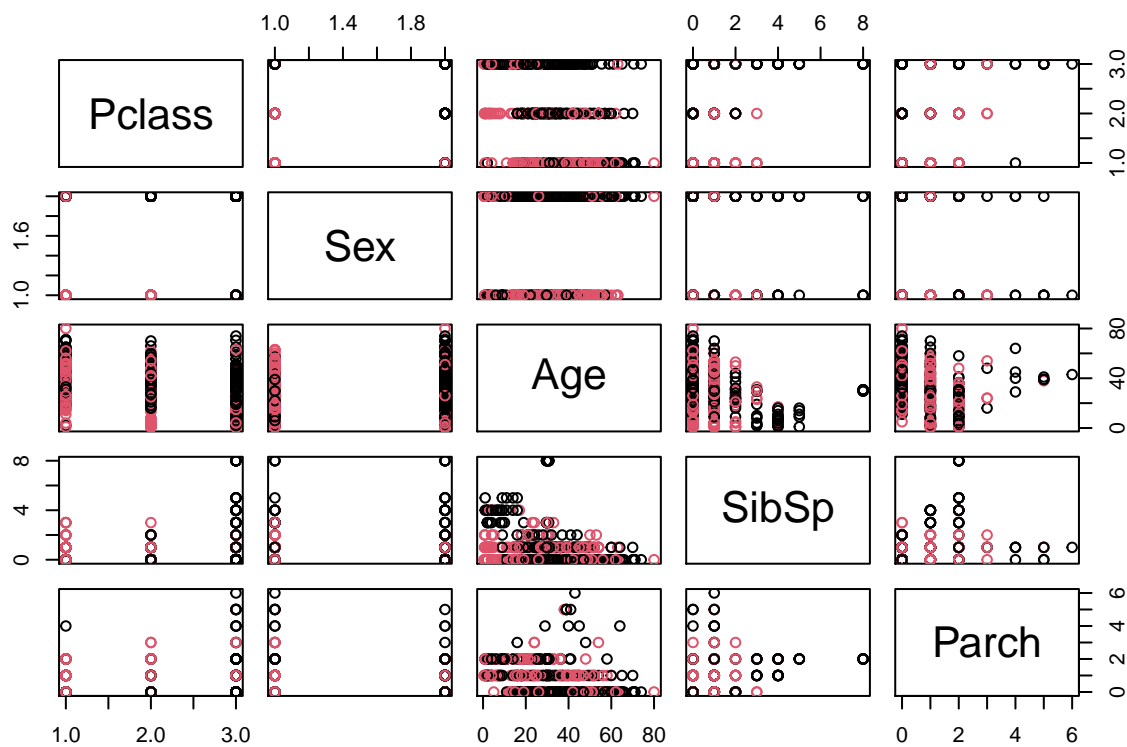
#cargamos todos los datos menos la variable survived, ya que queremos predecirla.
x <- titanicData[,2:6]

#Aplicamos el algoritmo
titanicKmenas <- kmeans(x, 2)

#se muestran los resultados de la aplicación del algoritmo
plot(x, col=titanicKmenas$cluster)
```



```
#se muestran los clusters originales con la variable survived para comparar la eficiencia del algoritmo
plot(x, col=as.factor(titanicData$Survived))
```



*#Mostramos los resultados en forma de tabla*

```
tabla <- table(titanicKmenas$cluster,titanicData$Survived)
```

```
tabla
```

```
##
##      0    1
## 1 335 194
## 2 214 148
```

*#calculamos la precisi3n*

```
if (tabla[1,1] > tabla[2,1]) { a <- tabla[1,1] } else { a <- tabla[2,1] }
```

```
if (tabla[1,2] > tabla[2,2]) { b <- tabla[1,2] } else { b <- tabla[2,2] }
```

```
100*(a+b)/dim(titanicData)[1]
```

```
## [1] 59.37149
```

Como vemos en los resultados kmeans no es la mejor opci3n para predecir la probabilidad de supervivencia que tenían los pasajeros del titanic. Ya que aunque podamos obtener una precisi3n aceptable en la creaci3n de los clusters, esta resulta ser muy aleatoria y varía muchísimo de una ejecuci3n a otra.

*#ejecutamos el algoritmo 10 veces y vemos como varia la precisi3n en cada ejecuci3n*



```

for (i in c(2,3,4,5,6,7,8,9,10))
{
  titanicKmenas <- kmeans(x, 2)
  tabla <- table(titanicKmenas$cluster,titanicData$Survived)

  if (tabla[1,1] > tabla[2,1]) { a <- tabla[1,1] } else { a <- tabla[2,1] }
  if (tabla[1,2] > tabla[2,2]) { b <- tabla[1,2] } else { b <- tabla[2,2] }

  precision <- 100*(a+b)/dim(titanicData)[1]
  print(precision)
}

```

```

## [1] 62.17733
## [1] 75.64534
## [1] 62.17733
## [1] 62.17733
## [1] 62.17733
## [1] 59.37149
## [1] 75.64534
## [1] 75.64534
## [1] 78.11448

```

## 6. Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?

Las conclusiones que podemos sacar del estudio son las siguientes:

- Los **valores extremos** encontrados son lógicos, al menos en las variables estudiadas. Si es verdad que podríamos encontrar algún valor extremo más difícil de justificar en la variable **fare**.
- Según el test de **Shapiro-Wilk** concluimos que los datos de las variables estudiadas **no muestran una distribución normal**.
- Mediante el test de **Fligner-Kill** afirmamos que en los datos de las variables estudiadas **no hay homogeneidad de la varianza**.
- Basándonos en la **regresión lineal** determinamos que el conjunto de variables estudiadas no pueden emplearse para predecir las demás.
- La prueba de **regresión lineal** indica que el conjunto de variables estudiadas tienen un nivel de significación inferior al 5% lo que significa tienen un nivel de confianza superior al 95%.
- Como ya nos adelantaban las regresiones lineales, al no tener variables capaces de predecir la supervivencia la aplicación del algoritmo **kmeans** resulta en la creación de grupos con mucha aleatoriedad y poca precisión.
- Hay una **correlación** superior al  $+0,30$  en: *Survived-Pclass*, *Survived-Sex*, *Age-Pclass* y *SibSp-Parch*. Estos resultados permiten comprender que variables están relacionadas entre ellas y especialmente con la variable *Survived*. Aunque esta correlación no resulta ser lo suficientemente fuerte como para predecir con una mínima exactitud la variable *survived*.

## 7. Tabla de contribuciones

Contribuciones	Firma
Investigación previa	VEAR, SSR
Redacción de las respuestas	VEAR, SSR
Desarrollo código	VEAR, SSR