Pràctica 2: Neteja i validació de les dades

*Víctor Espinosa Yxart*

*30 de desembre de 2018*

Taula de contingut

[Pràctica 2: Neteja i validació de les dades 1](#_Toc534234735)

[1 Descripció del dataset. 2](#_Toc534234736)

[2 Neteja de les dades. 2](#_Toc534234737)

[2.1 Carregar el fitxer 2](#_Toc534234738)

[2.2 Identificar les variables 4](#_Toc534234739)

[2.3 Corretgir el tipus de variables 4](#_Toc534234740)

[2.4 Búsqueda de valors atípics 7](#_Toc534234741)

[**2.4.1 Eliminar variables que no utilitzem** 7](#_Toc534234742)

[**2.4.2 Boxplots de les variables quantitatives** 7](#_Toc534234743)

[2.5 Valors perduts 10](#_Toc534234744)

[3 Anàlisi de les dades. 11](#_Toc534234745)

[3.1 Estudi de les companyies per sector 11](#_Toc534234746)

[**3.1.1 Industries target al sector financer** 15](#_Toc534234747)

[**3.1.2 Industries target al sector tecnològic** 16](#_Toc534234748)

[**3.1.3 Industries target al sector energètic** 17](#_Toc534234749)

[3.2 Estudi dels CEOs de les companyies 18](#_Toc534234750)

[**3.2.1 Incorporar la variable Sex** 18](#_Toc534234751)

[**3.2.2 Gràfic Homes i dones CEOs de la llista Fortune** 20](#_Toc534234752)

[**3.2.3 Model de regressió lineal múltiple** 26](#_Toc534234753)

[**3.2.4 Model de regressió logística** 39](#_Toc534234754)

[4 Resultats i conclusions. 59](#_Toc534234755)

[5 Dataset resultant amb les dades netes. 59](#_Toc534234756)

# 1 Descripció del dataset.

Treballarem amb el dataset que vam generar a la pràctica 1 amb les companyies més importants de la llista Fortune.

Aquest dataset és interessant, ja que ens dóna una visió de quines són les companyies més importants dels EUA i de les seves característiques.

* **rank** → Posició en la llista fortune
* **title** → Nom de la companyia
* **revenue** → Ingressos de l’últim any fiscal
* **ceo** → Nom del CEO
* **position** → Càrrec que ocupa el CEO
* **sector** → Sector industrial en el qual opera l’empresa
* **industry** → Indústria dins del sector en la cual opera l’empresa
* **hq** → Ubicació de la seu general
* **website** → URL de la pàgina web de la companyia
* **years** → Anys en la llista fortune
* **employees** → Nombre d’empleats
* **image** → Nom del fitxer amb la imatge corporativa

Analitzant aquest conjunt de dades intentarem estudiar quins són els sectors dominants i la diferència entre el nombre d’homes i dones que ocupen càrrecs directius.

El dataset original es troba al següent repositori: <https://github.com/victor427/PRAC1-Web-Scraping>

# 2 Neteja de les dades.

## **2.1 Carregar el fitxer**

Fem una primera inspecció del fitxer, el que podem veure és un fitxer del tipus CSV amb algunes característiques com:

* Els valors es separen amb una coma ( ; )
* Té capçalera
* Els números fan servir la notació americana i fan servir la coma ( , ) per separar els milers i el punt ( . ) com a separador decimal

Amb aquesta informació podem fer servir la funció read.csv per llegir el fitxer i transformar-lo en dades estructurades.

inputFile <- "./PRAC1-Web-Scraping-master/data/fortune500.csv"

writeLines(readLines(inputFile, n = 5))

## rank;title;revenue;ceo;position;sector;industry;hq;website;years;employees;image;

## 1;Walmart;$500,343;C. Douglas McMillon;President, Chief Executive Officer & Director;Retailing;General Merchandisers;Bentonville, Ark.;www.stock.walmart.com;24;2,300,000;walmart-fortune-5001;

## 2;Exxon Mobil;$244,363;Darren W. Woods;Chairman & Chief Executive Officer;Energy;Petroleum Refining;Irving, Texas;www.exxonmobil.com;24;71,200;exxonmobil-fortune-500;

## 3;Berkshire Hathaway;$242,137;Warren E. Buffett;Chairman, President & Chief Executive Officer;Financials;Insurance: Property and Casualty (Stock);Omaha;www.berkshirehathaway.com;24;377,000;berkshire-hathaway-fortune-5001;

## 4;Apple;$229,234;Timothy D. Cook;Chairman & Chief Executive Officer;Technology;Computers, Office Equipment;Cupertino, Calif.;www.apple.com;24;123,000;apple-fortune-500;

f500 <- read.csv(inputFile, header = TRUE, sep = ";", quote = "\"", dec = ".")

El fitxer conté 1000 companyies.

## **2.2 Identificar les variables**

Aquestes són les diferents variables que tenim a les nostres dades i el seu tipus.

* **rank** → Quantitativa discreta
* **title** → Qualitativa nominal
* **revenue** → Quantitativa contínua
* **ceo** → Qualitativa nominal
* **position** → Qualitativa nominal
* **sector** → Qualitativa nominal
* **industry** → Qualitativa nominal
* **hq** → Qualitativa nominal
* **website** → Qualitativa nominal
* **years** → Quantitativa discreta
* **employees** → Quantitativa discreta
* **image** → Qualitativa nominal

## **2.3 Corretgir el tipus de variables**

Per saber quin tipus hi ha assignat R a cada variable podem fer servir la funció class sobre el conjunt de dades. Si s’ha identificat erròniament alguna variable definirem manualment el tipus que volem.

lapply(f500, class)

## $rank

## [1] "integer"

##

## $title

## [1] "factor"

##

## $revenue

## [1] "factor"

##

## $ceo

## [1] "factor"

##

## $position

## [1] "factor"

##

## $sector

## [1] "factor"

##

## $industry

## [1] "factor"

##

## $hq

## [1] "factor"

##

## $website

## [1] "factor"

##

## $years

## [1] "factor"

##

## $employees

## [1] "factor"

##

## $image

## [1] "factor"

##

## $X

## [1] "logical"

f500[13] <- NULL

f500$revenue <- gsub('[$]','',f500$revenue)

f500$revenue <- gsub('[,]','',f500$revenue)

f500$rank <- as.integer(f500$rank)

f500$title <- as.factor(f500$title)

f500$revenue <- as.numeric(f500$revenue)

f500$ceo <- sapply(f500$ceo, toString)

f500$position <- as.factor(f500$position)

f500$sector <- as.factor(f500$sector)

f500$industry <- as.factor(f500$industry)

f500$hq <- as.factor(f500$hq)

f500$website <- as.factor(f500$website)

f500$years <- as.integer(f500$years)

f500$employees <- as.integer(f500$employees)

f500$image <- as.factor(f500$image)

lapply(f500, class)

## $rank

## [1] "integer"

##

## $title

## [1] "factor"

##

## $revenue

## [1] "numeric"

##

## $ceo

## [1] "character"

##

## $position

## [1] "factor"

##

## $sector

## [1] "factor"

##

## $industry

## [1] "factor"

##

## $hq

## [1] "factor"

##

## $website

## [1] "factor"

##

## $years

## [1] "integer"

##

## $employees

## [1] "integer"

##

## $image

## [1] "factor"

## **2.4 Búsqueda de valors atípics**

Fem un boxplot per les variables quantitatives i observem si existeixen outliers

En aquest cas els outliers de la variable revenue són dades coherents, i per la variable employees no hem trobat cap valor atípic.

### **2.4.1 Eliminar variables que no utilitzem**

Les variables website i image no les utilitzarem així que les eliminarem del dataset.

La variable website és l’URL de la pàgina web de la companyia, la variable imatge és el nom del fitxer .jpg associat amb la imatge corporativa. No són característiques que estudiarem així que les podem deixar fora.

*# website*

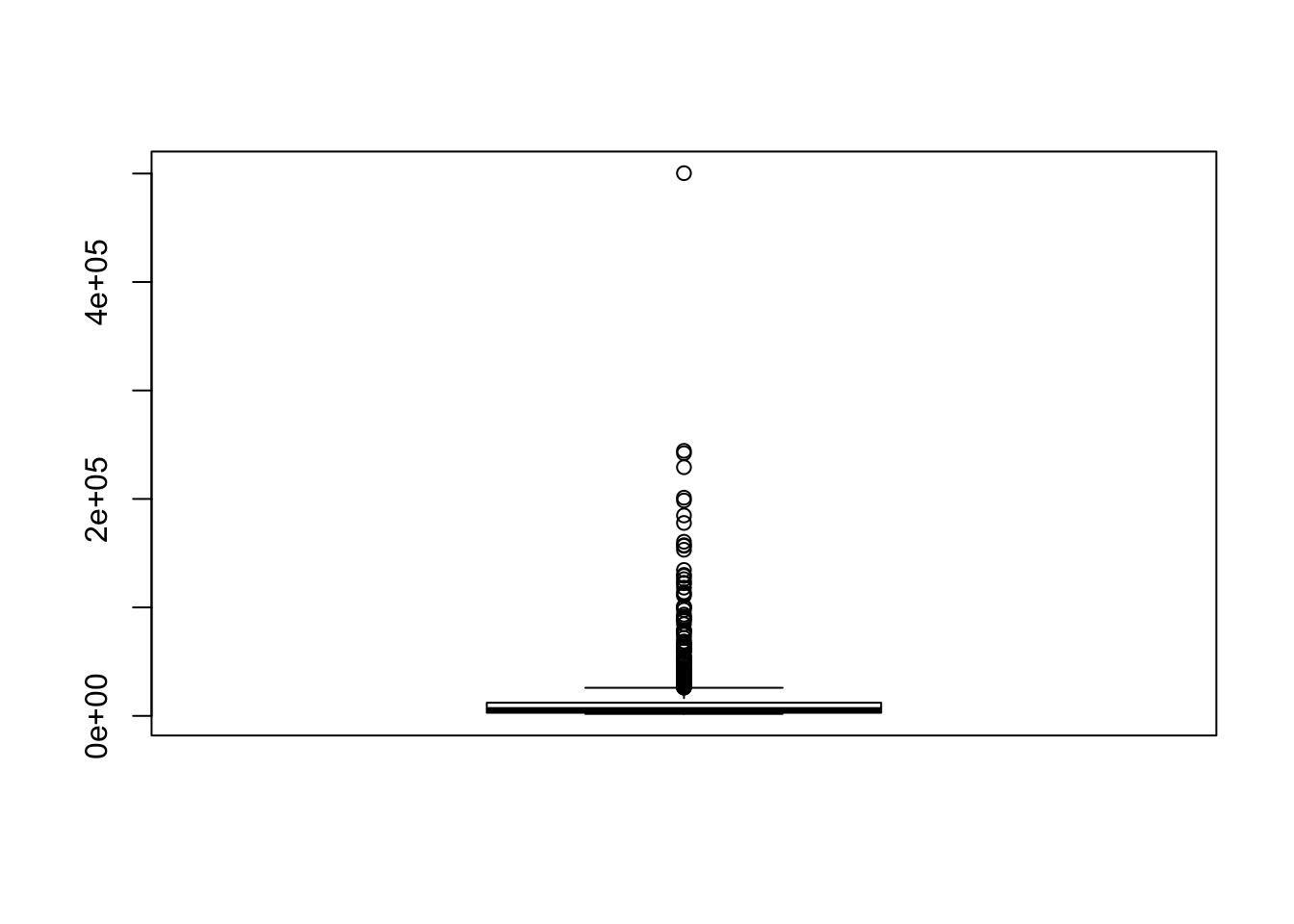
f500[12] <- NULL

*# image*

f500[9] <- NULL

### **2.4.2 Boxplots de les variables quantitatives**

boxplot(f500$revenue)



boxplot.stats(f500$revenue)$out

## [1] 500343.0 244363.0 242137.0 229234.0 201159.0 198533.0 184765.0

## [8] 177866.0 160546.0 157311.0 156776.0 153144.0 134533.0 129976.0

## [15] 129025.0 126034.0 122662.0 122274.0 118214.0 113899.0 112394.0

## [22] 110855.0 100904.0 100264.0 100064.6 97741.0 93392.0 91568.0

## [29] 90039.4 89950.0 88407.0 87966.0 84526.0 79139.0 78660.0

## [36] 78330.8 76450.0 74676.0 71879.0 68619.0 67610.0 66217.0

## [43] 66153.0 65872.0 63525.0 62761.0 62683.0 60828.0 60535.0

## [50] 60319.0 59837.0 59689.0 59678.2 55371.1 55137.0 53767.0

## [57] 52546.0 52056.0 51048.0 49520.0 48572.0 48005.0 47653.0

## [64] 47487.0 45462.0 43939.9 43642.0 42687.0 42296.0 42254.0

## [71] 42207.0 42151.0 41616.0 41581.0 41244.0 40653.0 40534.0

## [78] 40122.0 38524.0 38260.0 37736.0 37728.0 36775.0 36025.3

## [85] 35864.7 35583.0 35410.0 34836.8 34350.0 34204.0 33695.5

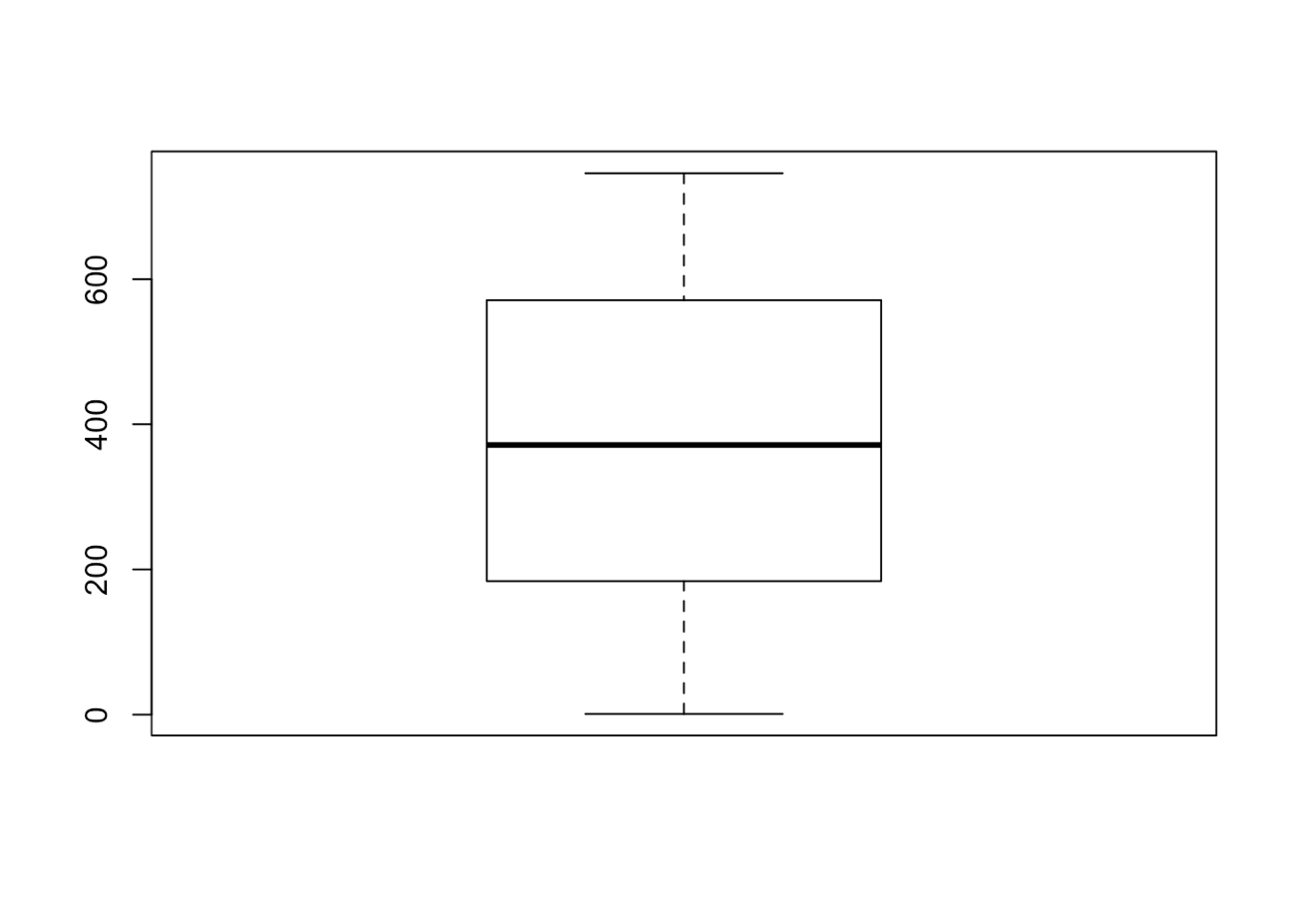
## [92] 33531.0 33495.4 32845.1 32584.0 31934.8 31657.0 31271.0

## [99] 30973.0 30015.8 29999.0 29737.7 29423.6 29331.0 29241.5

## [106] 28902.0 28871.0 28748.0 28500.0 28216.0 27390.0 26839.0

## [113] 26812.5 26232.0 26223.0 26107.0

boxplot(f500$employees)



boxplot.stats(f500$employees)$out

## integer(0)

## **2.5 Valors perduts**

S’han trobat valors perduts en les variables ceo, position, industry, hq, website, years, employees i image.

En el cas de les variables website i image no és rellevant, ja que no les farem servir en l’estudi del dataset i molts registres no tenen aquestes dades.

Després s’han torbat 14 registres que estan incomplets i depenent de les observacions que estem fent sobre el dataset haurem de tenir present treure aquestes dades.

sapply(f500, **function**(x) sum(is.na(x)))

## rank title revenue ceo position sector industry

## 0 0 0 0 14 0 14

## hq years employees

## 14 14 14

# 3 Anàlisi de les dades.

## **3.1 Estudi de les companyies per sector**

Visualitzem en un gràfic com es distribueixen les empreses de la llista per sector comercial (variable sector).

Els setors comercials presents al dataset són:

levels(f500$sector)

## [1] "Aerospace & Defense" "Apparel"

## [3] "Business Services" "Chemicals"

## [5] "Energy" "Engineering & Construction"

## [7] "Financials" "Food & Drug Stores"

## [9] "Food, Beverages & Tobacco" "Health Care"

## [11] "Hotels, Restaurants & Leisure" "Household Products"

## [13] "Industrials" "Materials"

## [15] "Media" "Motor Vehicles & Parts"

## [17] "Retailing" "Technology"

## [19] "Telecommunications" "Transportation"

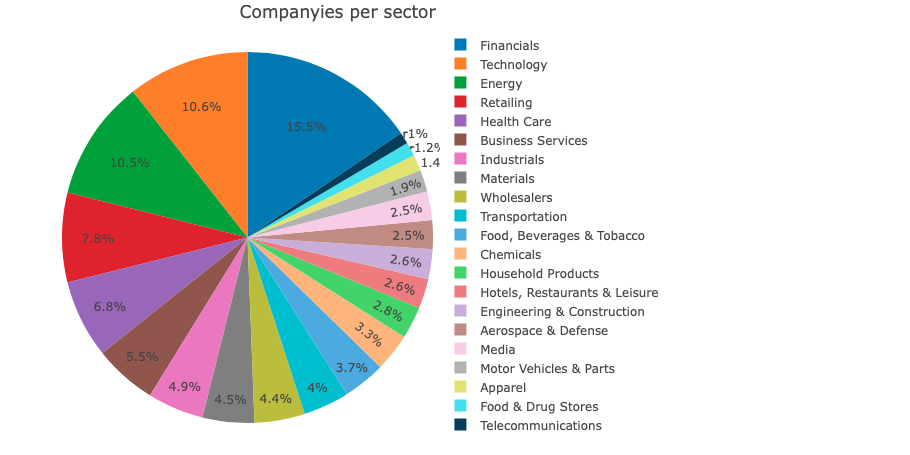
## [21] "Wholesalers"

plot\_ly(f500, labels = ~sector, type = 'pie') %>%

layout(title = 'Companyies per sector',

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))



El camp industry ens dóna informació sobre a quina indústria dins del sector comercial es dedica la companya, podríem veure aquest atribut com un nivell més de la categoria sector. Les categories que disposem són moltes, les podem veure a continuació en la llista, i en un gràfic general no ens aportaria gaire informació. Estudiarem aquesta variable en els sectors més importants (quant a volum d’empreses que hi participen).

levels(f500$industry)

## [1] "Advertising, marketing"

## [2] "Aerospace and Defense"

## [3] "Airlines"

## [4] "Apparel"

## [5] "Automotive Retailing, Services"

## [6] "Beverages"

## [7] "Building Materials, Glass"

## [8] "Chemicals"

## [9] "Commercial Banks"

## [10] "Computer Software"

## [11] "Computers, Office Equipment"

## [12] "Construction and Farm Machinery"

## [13] "Diversified Financials"

## [14] "Diversified Outsourcing Services"

## [15] "Education"

## [16] "Electronics, Electrical Equip."

## [17] "Energy"

## [18] "Engineering, Construction"

## [19] "Entertainment"

## [20] "Financial Data Services"

## [21] "Food Consumer Products"

## [22] "Food Production"

## [23] "Food Services"

## [24] "Food and Drug Stores"

## [25] "Forest and Paper Products"

## [26] "General Merchandisers"

## [27] "Health Care: Insurance and Managed Care"

## [28] "Health Care: Medical Facilities"

## [29] "Health Care: Pharmacy and Other Services"

## [30] "Home Equipment, Furnishings"

## [31] "Homebuilders"

## [32] "Hotels, Casinos, Resorts"

## [33] "Household and Personal Products"

## [34] "Industrial Machinery"

## [35] "Information Technology Services"

## [36] "Insurance: Life, Health (Mutual)"

## [37] "Insurance: Life, Health (stock)"

## [38] "Insurance: Property and Casualty (Mutual)"

## [39] "Insurance: Property and Casualty (Stock)"

## [40] "Internet Services and Retailing"

## [41] "Mail, Package, and Freight Delivery"

## [42] "Medical Products and Equipment"

## [43] "Metals"

## [44] "Mining, Crude-Oil Production"

## [45] "Miscellaneous"

## [46] "Motor Vehicles and Parts"

## [47] "Network and Other Communications Equipment"

## [48] "Oil and Gas Equipment, Services"

## [49] "Packaging, Containers"

## [50] "Petroleum Refining"

## [51] "Pharmaceuticals"

## [52] "Pipelines"

## [53] "Publishing, Printing"

## [54] "Railroads"

## [55] "Real estate"

## [56] "Scientific,Photographic and Control Equipment"

## [57] "Securities"

## [58] "Semiconductors and Other Electronic Components"

## [59] "Shipping"

## [60] "Specialty Retailers: Apparel"

## [61] "Specialty Retailers: Other"

## [62] "Telecommunications"

## [63] "Temporary Help"

## [64] "Tobacco"

## [65] "Toys, Sporting Goods"

## [66] "Transportation Equipment"

## [67] "Transportation and Logistics"

## [68] "Trucking, Truck Leasing"

## [69] "Utilities: Gas and Electric"

## [70] "Waste Management"

## [71] "Wholesalers: Diversified"

## [72] "Wholesalers: Electronics and Office Equipment"

## [73] "Wholesalers: Food and Grocery"

## [74] "Wholesalers: Health Care"

### **3.1.1 Industries target al sector financer**

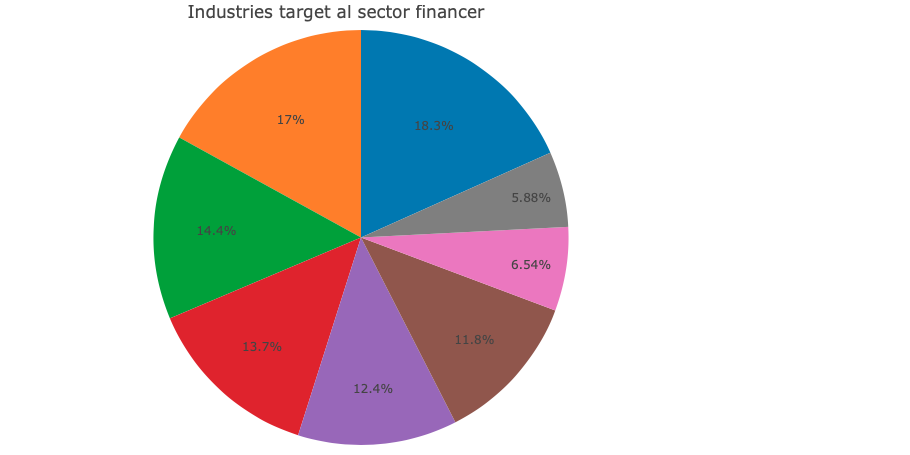
dfFinancials <- f500[ which(f500$sector=='Financials' & !is.na(f500$industry)), ]

plot\_ly(dfFinancials, labels = ~industry, type = 'pie') %>%

layout(title = 'Industries target al sector financer', showlegend = F,

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

****

### **3.1.2 Industries target al sector tecnològic**

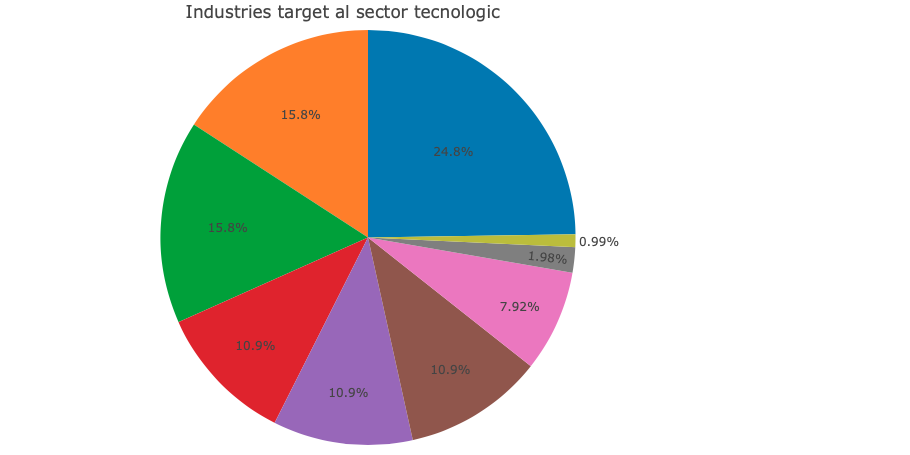
dfTechnology <- f500[ which(f500$sector=='Technology' & !is.na(f500$industry)), ]

plot\_ly(dfTechnology, labels = ~industry, type = 'pie') %>%

layout(title = 'Industries target al sector tecnologic', showlegend = F,

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

****

### **3.1.3 Industries target al sector energètic**

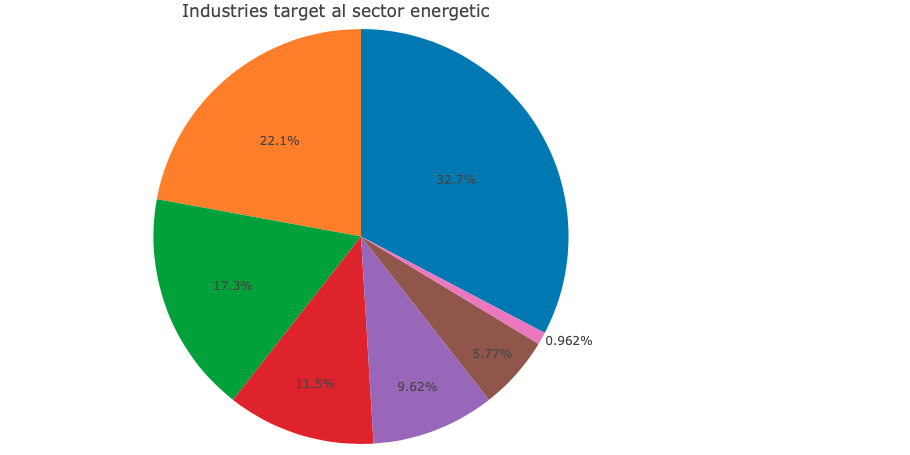
dfEnergy <- f500[ which(f500$sector=='Energy' & !is.na(f500$industry)), ]

plot\_ly(dfEnergy, labels = ~industry, type = 'pie') %>%

layout(title = 'Industries target al sector energetic', showlegend = F,

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

****

## **3.2 Estudi dels CEOs de les companyies**

Estudiarem quantes dones i homes ocupen càrrecs directius en les empreses de la llisa fortune.

### **3.2.1 Incorporar la variable Sex**

Per a realitzar l’estudi necessitem introduir una nova variable Sex que indiqui si el CEO de l’empresa és una dona (F) o un home (M). Això ho farem amb el paquet gender.

predictGender <- **function**(name) {

**if** (is.na(name)) {

**return**(NA)

}

fullName <- unlist(strsplit(name, " ", fixed = TRUE))

**if** (length(fullName) > 1) {

**if** (nchar(fullName[1]) > 2) {

name <- fullName[1]

} **else** {

name <- fullName[2]

}

} **else** {

name <- fullName[1]

}

predGender <- gender(name)

**if** (nrow(predGender) != 1) {

**return**(NA)

}

gender <- predGender$gender

**if** (gender == "female") {

**return**('F')

} **else** {

**return**('M')

}

}

f500$sex <- NA

f500$sex <- sapply(f500$ceo, predictGender)

f500$sex <- as.factor(f500$sex)

head(f500[, c(4, 11)], n = 20)

## ceo sex

## 1 C. Douglas McMillon M

## 2 Darren W. Woods M

## 3 Warren E. Buffett M

## 4 Timothy D. Cook M

## 5 David S. Wichmann M

## 6 John H. Hammergren M

## 7 Larry J. Merlo M

## 8 Jeffrey P. Bezos M

## 9 Randall L. Stephenson M

## 10 Mary T. Barra F

## 11 James P. Hackett M

## 12 Steven H. Collis M

## 13 Michael K. Wirth M

## 14 Michael C. Kaufmann M

## 15 W. Craig Jelinek M

## 16 Hans E. Vestberg M

## 17 W. Rodney McMullen M

## 18 H. Lawrence Culp Jr. M

## 19 Stefano Pessina M

## 20 James Dimon M

### **3.2.2 Gràfic Homes i dones CEOs de la llista Fortune**

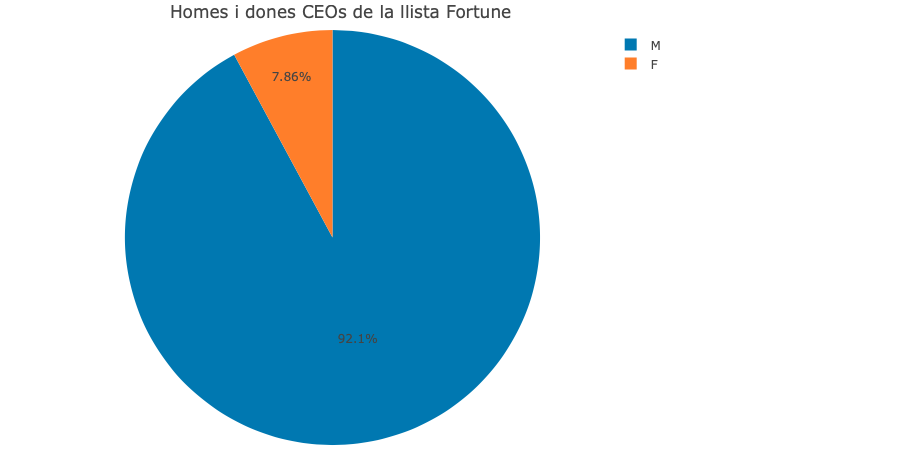
f500 <- f500[ which(!is.na(f500$sex)), ]

plot\_ly(f500, labels = ~sex, type = 'pie') %>%

layout(title = 'Homes i dones CEOs de la llista Fortune',

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))



f500[which(f500$sex == 'F'), c(2, 4, 5)]

## title ceo

## 10 General Motors Mary T. Barra

## 29 Anthem Gail K. Boudreaux

## 30 Microsoft Satya Nadella

## 34 IBM Virginia M. Rometty

## 47 DowDuPont NA

## 59 Lockheed Martin Marillyn A. Hewson

## 63 HCA Healthcare NA

## 90 Andeavor NA

## 99 General Dynamics Phebe N. Novakovic

## 112 Progressive Susan Patricia Griffith

## 125 Duke Energy Lynn J. Good

## 157 Kohl\342\200\231s Michelle D. Gass

## 159 Jabil NA

## 168 PG&E Corp. Geisha J. Williams

## 173 Synchrony Financial Margaret M. Keane

## 175 Bank of New York Mellon NA

## 209 Ross Stores Barbara Rentler

## 216 Land O\342\200\231Lakes Beth Ford

## 220 Occidental Petroleum Vicki A. Hollub

## 231 L Brands Leslie H. Wexner

## 233 Dominion Energy NA

## 234 Reinsurance Group of America Anna Manning

## 235 J.C. Penney Jill A. Soltau

## 239 Guardian Life Ins. Co. of America Deanna M. Mulligan

## 250 BB&T Corp. Kelly S. King

## 304 IQVIA Holdings NA

## 335 Hertz Global Holdings Kathryn V. Marinello

## 346 Veritiv Mary A. Laschinger

## 358 Campbell Soup Denise M. Morrison

## 372 WEC Energy Group Gale E. Klappa

## 376 Jones Financial (Edward Jones) NA

## 379 Hershey Michele G. Buck

## 402 JetBlue Airways Robin Hayes

## 412 KeyCorp Beth E. Mooney

## 421 Ralph Lauren Patrice Louvet

## 426 Graybar Electric Kathleen M. Mazzarella

## 429 CMS Energy Patricia K. Poppe

## 471 Ulta Beauty Mary N. Dillon

## 485 Avon Products Jan Zijderveld

## 506 Advanced Micro Devices Lisa T. Su

## 508 Williams-Sonoma Laura J. Alber

## 525 Commercial Metals Barbara R. Smith

## 543 Brookdale Senior Living Lucinda M. Baier

## 555 Tapestry NA

## 584 Bloomin\342\200\231 Brands Elizabeth A. Smith

## 593 Quad/Graphics NA

## 606 Encompass Health NA

## 608 Nasdaq Adena T. Friedman

## 616 Taylor Morrison Home Sheryl D. Palmer

## 652 Ventas Debra A. Cafaro

## 658 CIT Group Ellen R. Alemany

## 666 Abercrombie & Fitch Fran Horowitz-Bonadies

## 669 Puget Energy Kimberly J. Harris

## 682 Alliant Energy Patricia L. Kampling

## 690 American Water Works Susan N. Story

## 698 IAC/InterActiveCorp NA

## 708 USG Jennifer F. Scanlon

## 742 Cracker Barrel Old Country Store Sandra B. Cochran

## 761 Penn Mutual Life Insurance Eileen C. McDonnell

## 763 ArcBest Judy R. McReynolds

## 772 Convergys Andrea J. Ayers

## 778 Caleres Diane M. Sullivan

## 795 Revlon Debra G. Perelman

## 814 ITT Denise L. Ramos

## 819 Hawaiian Electric Industries Constance H. Lau

## 842 Hovnanian Enterprises Ara K. Hovnanian

## 846 Parexel International Jamie Macdonald

## 867 Cleveland-Cliffs NA

## 881 Chico\342\200\231s FAS Shelley G. Broader

## 883 Herman Miller Andi Owen

## 894 Tupperware Brands Patricia A. Stitzel

## 953 Portland General Electric Maria M. Pope

## 957 AMN Healthcare Services Susan R. Salka

## 963 M/I Homes NA

## 976 Engility Holdings Lynn A. Dugle

## 989 Aerojet Rocketdyne Holdings Eileen P. Drake

## 993 Children\342\200\231s Place Jane T. Elfers

## position

## 10 Chairman & Chief Executive Officer

## 29 President, Chief Executive Officer & Director

## 30 Chief Executive Officer & Director

## 34 Chairman, President & Chief Executive Officer

## 47 <NA>

## 59 Chairman, President & Chief Executive Officer

## 63 <NA>

## 90 <NA>

## 99 Chairman & Chief Executive Officer

## 112 President, Chief Executive Officer & Director

## 125 Chairman, President & Chief Executive Officer

## 157 Chairman & Chief Executive Officer

## 159 <NA>

## 168 President, Chief Executive Officer & Director

## 173 President, Chief Executive Officer & Director

## 175 <NA>

## 209 Chairman & Chief Executive Officer

## 216 President, Chief Executive Officer & Director

## 220 President, Chief Executive Officer & Director

## 231 Chairman, President & Chief Executive Officer

## 233 <NA>

## 234 President, Chief Executive Officer & Director

## 235 Chairman & Chief Executive Officer

## 239 President, Chief Executive Officer & Director

## 250 Chairman & Chief Executive Officer

## 304 <NA>

## 335 President, Chief Executive Officer & Director

## 346 Chairman & Chief Executive Officer

## 358 President, Chief Executive Officer & Director

## 372 Chairman & Chief Executive Officer

## 376 <NA>

## 379 President, Chief Executive Officer & Director

## 402 President, Chief Executive Officer & Director

## 412 Chairman, President & Chief Executive Officer

## 421 President, Chief Executive Officer & Director

## 426 Chairman, President & Chief Executive Officer

## 429 President, Chief Executive Officer & Director

## 471 Chairman & Chief Executive Officer

## 485 Chairman & Chief Executive Officer

## 506 President, Chief Executive Officer & Director

## 508 President, Chief Executive Officer & Director

## 525 Chairman, President & Chief Executive Officer

## 543 President, Chief Executive Officer & Director

## 555 <NA>

## 584 Chairman, President & Chief Executive Officer

## 593 <NA>

## 606 <NA>

## 608 President, Chief Executive Officer & Director

## 616 Chairman, President & Chief Executive Officer

## 652 Chairman & Chief Executive Officer

## 658 Chairwoman & CEO

## 666 Chief Executive Officer & Director

## 669 President, Chief Executive Officer & Director

## 682 Chairman & Chief Executive Officer

## 690 President, Chief Executive Officer & Director

## 698 <NA>

## 708 President, Chief Executive Officer & Director

## 742 President, Chief Executive Officer & Director

## 761 Chairman & Chief Executive Officer

## 763 Chairman, President & Chief Executive Officer

## 772 President, Chief Executive Officer & Director

## 778 Chairman, President & Chief Executive Officer

## 795 Vice Chairman & Chief Executive Officer

## 814 President, Chief Executive Officer & Director

## 819 President, Chief Executive Officer & Director

## 842 Chairman, President & Chief Executive Officer

## 846 Chief Executive Officer

## 867 <NA>

## 881 President, Chief Executive Officer & Director

## 883 President, Chief Executive Officer & Director

## 894 President, Chief Executive Officer & Director

## 953 President, Chief Executive Officer & Director

## 957 President, Chief Executive Officer & Director

## 963 <NA>

## 976 Chairman, President & Chief Executive Officer

## 989 President, Chief Executive Officer & Director

## 993 President, Chief Executive Officer & Director

Intentarem crear un model que pugui relacionar l’observació de la variable sex amb les característiques de la companya (sector, indústria, ingressos i rànquing).

### **3.2.3 Model de regressió lineal múltiple**

genderToNumber <- **function**(sex) {

**if** (sex == 'F') {

**return**(1)

} **else** {

**return**(0)

}

}

f500 <- f500[ which(!is.na(f500$industry)), ]

f500$sexCoef <- NA

f500$sexCoef <- sapply(f500$sex, genderToNumber)

f500$sexCoef <- as.numeric(f500$sexCoef)

model1 <- lm(sexCoef ~ sector + industry + revenue + rank, data = f500)

summary(model1)

##

## Call:

## lm(formula = sexCoef ~ sector + industry + revenue + rank, data = f500)

##

## Residuals:

## Min 1Q Median 3Q Max

## -0.30074 -0.08353 -0.03742 0.00007 0.96487

##

## Coefficients: (12 not defined because of singularities)

## Estimate

## (Intercept) 1.585e-01

## sectorApparel -8.299e-02

## sectorBusiness Services -1.594e-01

## sectorChemicals -1.600e-01

## sectorEnergy 8.406e-01

## sectorEngineering & Construction 6.689e-03

## sectorFinancials -1.101e-01

## sectorFood & Drug Stores -1.600e-01

## sectorFood, Beverages & Tobacco -1.600e-01

## sectorHealth Care -1.607e-01

## sectorHotels, Restaurants & Leisure -1.600e-01

## sectorHousehold Products -1.597e-01

## sectorIndustrials -1.585e-01

## sectorMaterials -1.588e-01

## sectorMedia -1.604e-01

## sectorMotor Vehicles & Parts -1.074e-01

## sectorRetailing -1.604e-01

## sectorTechnology -1.602e-01

## sectorTelecommunications -1.600e-01

## sectorTransportation -6.924e-02

## sectorWholesalers -1.605e-01

## industryAerospace and Defense NA

## industryAirlines 2.061e-02

## industryApparel NA

## industryAutomotive Retailing, Services 1.009e-01

## industryBeverages 2.327e-04

## industryBuilding Materials, Glass 8.192e-02

## industryChemicals NA

## industryCommercial Banks 7.024e-02

## industryComputer Software 7.683e-02

## industryComputers, Office Equipment 1.647e-04

## industryConstruction and Farm Machinery -1.114e-03

## industryDiversified Financials 5.941e-03

## industryDiversified Outsourcing Services 7.083e-02

## industryEducation -1.094e-03

## industryElectronics, Electrical Equip. -1.516e-03

## industryEnergy -1.001e+00

## industryEngineering, Construction -1.665e-01

## industryEntertainment 4.629e-04

## industryFinancial Data Services -6.877e-04

## industryFood Consumer Products 1.582e-01

## industryFood Production 2.052e-04

## industryFood Services 1.667e-01

## industryFood and Drug Stores NA

## industryForest and Paper Products -1.231e-03

## industryGeneral Merchandisers 1.817e-01

## industryHealth Care: Insurance and Managed Care 1.007e-01

## industryHealth Care: Medical Facilities 8.419e-02

## industryHealth Care: Pharmacy and Other Services 1.542e-01

## industryHome Equipment, Furnishings 9.058e-02

## industryHomebuilders NA

## industryHotels, Casinos, Resorts NA

## industryHousehold and Personal Products 2.140e-01

## industryIndustrial Machinery 3.536e-02

## industryInformation Technology Services 6.275e-02

## industryInsurance: Life, Health (Mutual) 5.038e-02

## industryInsurance: Life, Health (stock) 5.557e-02

## industryInsurance: Property and Casualty (Mutual) -4.983e-02

## industryInsurance: Property and Casualty (Stock) -1.404e-02

## industryInternet Services and Retailing 1.766e-04

## industryMail, Package, and Freight Delivery -9.054e-02

## industryMedical Products and Equipment 6.432e-04

## industryMetals 8.231e-02

## industryMining, Crude-Oil Production -9.551e-01

## industryMiscellaneous -8.083e-04

## industryMotor Vehicles and Parts NA

## industryNetwork and Other Communications Equipment 2.316e-04

## industryOil and Gas Equipment, Services -1.001e+00

## industryPackaging, Containers -9.691e-04

## industryPetroleum Refining -1.001e+00

## industryPharmaceuticals 1.020e-03

## industryPipelines -1.000e+00

## industryPublishing, Printing 3.497e-04

## industryRailroads -9.058e-02

## industryReal estate -4.632e-03

## industryScientific,Photographic and Control Equipment -2.545e-05

## industrySecurities NA

## industrySemiconductors and Other Electronic Components 4.364e-02

## industryShipping -9.148e-02

## industrySpecialty Retailers: Apparel 3.003e-01

## industrySpecialty Retailers: Other 6.722e-02

## industryTelecommunications NA

## industryTemporary Help -5.871e-04

## industryTobacco NA

## industryToys, Sporting Goods -1.334e-04

## industryTransportation Equipment -9.078e-02

## industryTransportation and Logistics -9.069e-02

## industryTrucking, Truck Leasing NA

## industryUtilities: Gas and Electric -7.579e-01

## industryWaste Management -3.943e-04

## industryWholesalers: Diversified 8.385e-02

## industryWholesalers: Electronics and Office Equipment 8.396e-04

## industryWholesalers: Food and Grocery 1.101e-03

## industryWholesalers: Health Care NA

## revenue 1.838e-08

## rank 2.343e-06

## Std. Error t value

## (Intercept) 5.304e-02 2.989

## sectorApparel 8.369e-02 -0.992

## sectorBusiness Services 1.800e-01 -0.885

## sectorChemicals 6.584e-02 -2.430

## sectorEnergy 3.350e-01 2.509

## sectorEngineering & Construction 8.597e-02 0.078

## sectorFinancials 7.351e-02 -1.498

## sectorFood & Drug Stores 8.610e-02 -1.859

## sectorFood, Beverages & Tobacco 1.798e-01 -0.890

## sectorHealth Care 4.004e-01 -0.401

## sectorHotels, Restaurants & Leisure 8.171e-02 -1.958

## sectorHousehold Products 3.349e-01 -0.477

## sectorIndustrials 3.359e-01 -0.472

## sectorMaterials 3.354e-01 -0.473

## sectorMedia 3.578e-01 -0.448

## sectorMotor Vehicles & Parts 7.450e-02 -1.442

## sectorRetailing 3.470e-01 -0.462

## sectorTechnology 3.081e-01 -0.520

## sectorTelecommunications 9.205e-02 -1.738

## sectorTransportation 8.865e-02 -0.781

## sectorWholesalers 1.129e-01 -1.421

## industryAerospace and Defense NA NA

## industryAirlines 1.103e-01 0.187

## industryApparel NA NA

## industryAutomotive Retailing, Services 3.523e-01 0.286

## industryBeverages 1.962e-01 0.001

## industryBuilding Materials, Glass 3.395e-01 0.241

## industryChemicals NA NA

## industryCommercial Banks 7.392e-02 0.950

## industryComputer Software 3.116e-01 0.247

## industryComputers, Office Equipment 3.132e-01 0.001

## industryConstruction and Farm Machinery 3.430e-01 -0.003

## industryDiversified Financials 7.999e-02 0.074

## industryDiversified Outsourcing Services 1.852e-01 0.382

## industryEducation 3.002e-01 -0.004

## industryElectronics, Electrical Equip. 3.393e-01 -0.004

## industryEnergy 3.364e-01 -2.974

## industryEngineering, Construction 9.627e-02 -1.730

## industryEntertainment 3.500e-01 0.001

## industryFinancial Data Services 1.810e-01 -0.004

## industryFood Consumer Products 1.820e-01 0.870

## industryFood Production 1.935e-01 0.001

## industryFood Services 9.625e-02 1.732

## industryFood and Drug Stores NA NA

## industryForest and Paper Products 3.537e-01 -0.003

## industryGeneral Merchandisers 3.513e-01 0.517

## industryHealth Care: Insurance and Managed Care 4.053e-01 0.249

## industryHealth Care: Medical Facilities 4.037e-01 0.209

## industryHealth Care: Pharmacy and Other Services 4.031e-01 0.383

## industryHome Equipment, Furnishings 3.394e-01 0.267

## industryHomebuilders NA NA

## industryHotels, Casinos, Resorts NA NA

## industryHousehold and Personal Products 3.377e-01 0.634

## industryIndustrial Machinery 3.359e-01 0.105

## industryInformation Technology Services 3.103e-01 0.202

## industryInsurance: Life, Health (Mutual) 9.510e-02 0.530

## industryInsurance: Life, Health (stock) 7.862e-02 0.707

## industryInsurance: Property and Casualty (Mutual) 9.836e-02 -0.507

## industryInsurance: Property and Casualty (Stock) 7.200e-02 -0.195

## industryInternet Services and Retailing 3.130e-01 0.001

## industryMail, Package, and Freight Delivery 1.892e-01 -0.478

## industryMedical Products and Equipment 4.021e-01 0.002

## industryMetals 3.391e-01 0.243

## industryMining, Crude-Oil Production 3.355e-01 -2.847

## industryMiscellaneous 2.237e-01 -0.004

## industryMotor Vehicles and Parts NA NA

## industryNetwork and Other Communications Equipment 3.179e-01 0.001

## industryOil and Gas Equipment, Services 3.462e-01 -2.891

## industryPackaging, Containers 3.372e-01 -0.003

## industryPetroleum Refining 3.406e-01 -2.938

## industryPharmaceuticals 4.026e-01 0.003

## industryPipelines 3.395e-01 -2.947

## industryPublishing, Printing 3.709e-01 0.001

## industryRailroads 1.321e-01 -0.686

## industryReal estate 7.560e-02 -0.061

## industryScientific,Photographic and Control Equipment 3.130e-01 0.000

## industrySecurities NA NA

## industrySemiconductors and Other Electronic Components 3.084e-01 0.141

## industryShipping 1.883e-01 -0.486

## industrySpecialty Retailers: Apparel 3.480e-01 0.863

## industrySpecialty Retailers: Other 3.465e-01 0.194

## industryTelecommunications NA NA

## industryTemporary Help 2.049e-01 -0.003

## industryTobacco NA NA

## industryToys, Sporting Goods 4.118e-01 0.000

## industryTransportation Equipment 1.320e-01 -0.688

## industryTransportation and Logistics 1.242e-01 -0.730

## industryTrucking, Truck Leasing NA NA

## industryUtilities: Gas and Electric 3.341e-01 -2.269

## industryWaste Management 2.120e-01 -0.002

## industryWholesalers: Diversified 1.139e-01 0.736

## industryWholesalers: Electronics and Office Equipment 1.336e-01 0.006

## industryWholesalers: Food and Grocery 1.427e-01 0.008

## industryWholesalers: Health Care NA NA

## revenue 3.190e-07 0.058

## rank 3.420e-05 0.069

## Pr(>|t|)

## (Intercept) 0.00288 \*\*

## sectorApparel 0.32163

## sectorBusiness Services 0.37627

## sectorChemicals 0.01528 \*

## sectorEnergy 0.01227 \*

## sectorEngineering & Construction 0.93800

## sectorFinancials 0.13460

## sectorFood & Drug Stores 0.06342 .

## sectorFood, Beverages & Tobacco 0.37383

## sectorHealth Care 0.68835

## sectorHotels, Restaurants & Leisure 0.05058 .

## sectorHousehold Products 0.63365

## sectorIndustrials 0.63708

## sectorMaterials 0.63601

## sectorMedia 0.65409

## sectorMotor Vehicles & Parts 0.14963

## sectorRetailing 0.64414

## sectorTechnology 0.60328

## sectorTelecommunications 0.08251 .

## sectorTransportation 0.43502

## sectorWholesalers 0.15571

## industryAerospace and Defense NA

## industryAirlines 0.85183

## industryApparel NA

## industryAutomotive Retailing, Services 0.77457

## industryBeverages 0.99905

## industryBuilding Materials, Glass 0.80936

## industryChemicals NA

## industryCommercial Banks 0.34221

## industryComputer Software 0.80533

## industryComputers, Office Equipment 0.99958

## industryConstruction and Farm Machinery 0.99741

## industryDiversified Financials 0.94081

## industryDiversified Outsourcing Services 0.70225

## industryEducation 0.99709

## industryElectronics, Electrical Equip. 0.99644

## industryEnergy 0.00301 \*\*

## industryEngineering, Construction 0.08399 .

## industryEntertainment 0.99895

## industryFinancial Data Services 0.99697

## industryFood Consumer Products 0.38478

## industryFood Production 0.99915

## industryFood Services 0.08366 .

## industryFood and Drug Stores NA

## industryForest and Paper Products 0.99722

## industryGeneral Merchandisers 0.60506

## industryHealth Care: Insurance and Managed Care 0.80377

## industryHealth Care: Medical Facilities 0.83485

## industryHealth Care: Pharmacy and Other Services 0.70215

## industryHome Equipment, Furnishings 0.78965

## industryHomebuilders NA

## industryHotels, Casinos, Resorts NA

## industryHousehold and Personal Products 0.52642

## industryIndustrial Machinery 0.91617

## industryInformation Technology Services 0.83977

## industryInsurance: Life, Health (Mutual) 0.59640

## industryInsurance: Life, Health (stock) 0.47988

## industryInsurance: Property and Casualty (Mutual) 0.61255

## industryInsurance: Property and Casualty (Stock) 0.84541

## industryInternet Services and Retailing 0.99955

## industryMail, Package, and Freight Delivery 0.63244

## industryMedical Products and Equipment 0.99872

## industryMetals 0.80826

## industryMining, Crude-Oil Production 0.00452 \*\*

## industryMiscellaneous 0.99712

## industryMotor Vehicles and Parts NA

## industryNetwork and Other Communications Equipment 0.99942

## industryOil and Gas Equipment, Services 0.00394 \*\*

## industryPackaging, Containers 0.99771

## industryPetroleum Refining 0.00339 \*\*

## industryPharmaceuticals 0.99798

## industryPipelines 0.00329 \*\*

## industryPublishing, Printing 0.99925

## industryRailroads 0.49308

## industryReal estate 0.95116

## industryScientific,Photographic and Control Equipment 0.99994

## industrySecurities NA

## industrySemiconductors and Other Electronic Components 0.88751

## industryShipping 0.62717

## industrySpecialty Retailers: Apparel 0.38842

## industrySpecialty Retailers: Other 0.84621

## industryTelecommunications NA

## industryTemporary Help 0.99771

## industryTobacco NA

## industryToys, Sporting Goods 0.99974

## industryTransportation Equipment 0.49171

## industryTransportation and Logistics 0.46560

## industryTrucking, Truck Leasing NA

## industryUtilities: Gas and Electric 0.02352 \*

## industryWaste Management 0.99852

## industryWholesalers: Diversified 0.46190

## industryWholesalers: Electronics and Office Equipment 0.99499

## industryWholesalers: Food and Grocery 0.99384

## industryWholesalers: Health Care NA

## revenue 0.95407

## rank 0.94540

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.2447 on 882 degrees of freedom

## Multiple R-squared: 0.1034, Adjusted R-squared: 0.01908

## F-statistic: 1.226 on 83 and 882 DF, p-value: 0.09095

*# Importància de les variables que defineixen el model:*

varImp(model1, scale = FALSE)

## Overall

## sectorApparel 9.916619e-01

## sectorBusiness Services 8.852347e-01

## sectorChemicals 2.430377e+00

## sectorEnergy 2.509397e+00

## sectorEngineering & Construction 7.780615e-02

## sectorFinancials 1.497597e+00

## sectorFood & Drug Stores 1.858554e+00

## sectorFood, Beverages & Tobacco 8.897825e-01

## sectorHealth Care 4.012327e-01

## sectorHotels, Restaurants & Leisure 1.957714e+00

## sectorHousehold Products 4.767555e-01

## sectorIndustrials 4.719500e-01

## sectorMaterials 4.734461e-01

## sectorMedia 4.482401e-01

## sectorMotor Vehicles & Parts 1.442085e+00

## sectorRetailing 4.620799e-01

## sectorTechnology 5.198784e-01

## sectorTelecommunications 1.738258e+00

## sectorTransportation 7.809947e-01

## sectorWholesalers 1.420862e+00

## industryAirlines 1.868374e-01

## industryAutomotive Retailing, Services 2.864897e-01

## industryBeverages 1.185793e-03

## industryBuilding Materials, Glass 2.413193e-01

## industryCommercial Banks 9.503245e-01

## industryComputer Software 2.465250e-01

## industryComputers, Office Equipment 5.258066e-04

## industryConstruction and Farm Machinery 3.248533e-03

## industryDiversified Financials 7.427427e-02

## industryDiversified Outsourcing Services 3.824124e-01

## industryEducation 3.644856e-03

## industryElectronics, Electrical Equip. 4.467144e-03

## industryEnergy 2.974494e+00

## industryEngineering, Construction 1.729967e+00

## industryEntertainment 1.322555e-03

## industryFinancial Data Services 3.798935e-03

## industryFood Consumer Products 8.695584e-01

## industryFood Production 1.060360e-03

## industryFood Services 1.731780e+00

## industryForest and Paper Products 3.480081e-03

## industryGeneral Merchandisers 5.173198e-01

## industryHealth Care: Insurance and Managed Care 2.485517e-01

## industryHealth Care: Medical Facilities 2.085527e-01

## industryHealth Care: Pharmacy and Other Services 3.825486e-01

## industryHome Equipment, Furnishings 2.668453e-01

## industryHousehold and Personal Products 6.337275e-01

## industryIndustrial Machinery 1.052833e-01

## industryInformation Technology Services 2.022434e-01

## industryInsurance: Life, Health (Mutual) 5.297814e-01

## industryInsurance: Life, Health (stock) 7.067889e-01

## industryInsurance: Property and Casualty (Mutual) 5.066158e-01

## industryInsurance: Property and Casualty (Stock) 1.950366e-01

## industryInternet Services and Retailing 5.642618e-04

## industryMail, Package, and Freight Delivery 4.784623e-01

## industryMedical Products and Equipment 1.599644e-03

## industryMetals 2.427433e-01

## industryMining, Crude-Oil Production 2.847005e+00

## industryMiscellaneous 3.613038e-03

## industryNetwork and Other Communications Equipment 7.283575e-04

## industryOil and Gas Equipment, Services 2.890524e+00

## industryPackaging, Containers 2.873859e-03

## industryPetroleum Refining 2.938349e+00

## industryPharmaceuticals 2.533960e-03

## industryPipelines 2.947235e+00

## industryPublishing, Printing 9.426668e-04

## industryRailroads 6.857061e-01

## industryReal estate 6.126824e-02

## industryScientific,Photographic and Control Equipment 8.130467e-05

## industrySemiconductors and Other Electronic Components 1.414996e-01

## industryShipping 4.858827e-01

## industrySpecialty Retailers: Apparel 8.629052e-01

## industrySpecialty Retailers: Other 1.940076e-01

## industryTemporary Help 2.865276e-03

## industryToys, Sporting Goods 3.238649e-04

## industryTransportation Equipment 6.878753e-01

## industryTransportation and Logistics 7.299664e-01

## industryUtilities: Gas and Electric 2.268902e+00

## industryWaste Management 1.860114e-03

## industryWholesalers: Diversified 7.360456e-01

## industryWholesalers: Electronics and Office Equipment 6.282117e-03

## industryWholesalers: Food and Grocery 7.720244e-03

## revenue 5.761537e-02

## rank 6.850666e-02

f500$sexPred1 <- predict(model1, f500, type="response")

### **3.2.4 Model de regressió logística**

model2 <- glm(sex ~ sector + industry + revenue + rank, data = f500, family = binomial(logit))

summary(model2)

##

## Call:

## glm(formula = sex ~ sector + industry + revenue + rank, family = binomial(logit),

## data = f500)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -8.4904 0.0000 0.1375 1.6237 8.4904

##

## Coefficients: (9 not defined because of singularities)

## Estimate

## (Intercept) 8.181e+01

## sectorApparel -8.397e+01

## sectorBusiness Services 2.912e+05

## sectorChemicals -6.117e+01

## sectorEnergy -4.504e+15

## sectorEngineering & Construction -8.289e+01

## sectorFinancials -8.278e+01

## sectorFood & Drug Stores -6.301e+01

## sectorFood, Beverages & Tobacco 2.013e+15

## sectorHealth Care -4.504e+15

## sectorHotels, Restaurants & Leisure 3.323e+15

## sectorHousehold Products 4.504e+15

## sectorIndustrials -1.327e+04

## sectorMaterials -2.714e+04

## sectorMedia 4.504e+15

## sectorMotor Vehicles & Parts -8.167e+01

## sectorRetailing 4.504e+15

## sectorTechnology -4.255e+05

## sectorTelecommunications 4.504e+15

## sectorTransportation -7.981e+01

## sectorWholesalers -3.971e+15

## industryAerospace and Defense NA

## industryAirlines 5.286e-01

## industryApparel NA

## industryAutomotive Retailing, Services -4.504e+15

## industryBeverages -2.013e+15

## industryBuilding Materials, Glass 2.706e+04

## industryChemicals NA

## industryCommercial Banks -3.196e+00

## industryComputer Software 4.254e+05

## industryComputers, Office Equipment 4.254e+05

## industryConstruction and Farm Machinery -1.067e+07

## industryDiversified Financials -1.759e+00

## industryDiversified Outsourcing Services -2.913e+05

## industryEducation 4.504e+15

## industryElectronics, Electrical Equip. 1.321e+04

## industryEnergy 9.007e+15

## industryEngineering, Construction 2.491e+01

## industryEntertainment 2.624e+06

## industryFinancial Data Services 1.208e+05

## industryFood Consumer Products -2.013e+15

## industryFood Production 2.490e+15

## industryFood Services -3.323e+15

## industryFood and Drug Stores NA

## industryForest and Paper Products 2.708e+04

## industryGeneral Merchandisers -4.504e+15

## industryHealth Care: Insurance and Managed Care 4.504e+15

## industryHealth Care: Medical Facilities 4.504e+15

## industryHealth Care: Pharmacy and Other Services 4.504e+15

## industryHome Equipment, Furnishings -4.504e+15

## industryHomebuilders NA

## industryHotels, Casinos, Resorts -3.323e+15

## industryHousehold and Personal Products -4.504e+15

## industryIndustrial Machinery 1.318e+04

## industryInformation Technology Services 4.254e+05

## industryInsurance: Life, Health (Mutual) -9.350e-01

## industryInsurance: Life, Health (stock) -1.673e+00

## industryInsurance: Property and Casualty (Mutual) -9.699e+02

## industryInsurance: Property and Casualty (Stock) -6.328e+00

## industryInternet Services and Retailing 2.052e+05

## industryMail, Package, and Freight Delivery -1.803e+08

## industryMedical Products and Equipment 9.007e+15

## industryMetals 2.706e+04

## industryMining, Crude-Oil Production 4.504e+15

## industryMiscellaneous 1.321e+04

## industryMotor Vehicles and Parts NA

## industryNetwork and Other Communications Equipment 4.254e+05

## industryOil and Gas Equipment, Services 4.504e+15

## industryPackaging, Containers 4.504e+15

## industryPetroleum Refining 9.007e+15

## industryPharmaceuticals 4.504e+15

## industryPipelines 4.504e+15

## industryPublishing, Printing 4.056e+06

## industryRailroads -2.509e-01

## industryReal estate -4.477e-03

## industryScientific,Photographic and Control Equipment 4.504e+15

## industrySecurities NA

## industrySemiconductors and Other Electronic Components 4.254e+05

## industryShipping 8.260e+00

## industrySpecialty Retailers: Apparel -4.504e+15

## industrySpecialty Retailers: Other -4.504e+15

## industryTelecommunications NA

## industryTemporary Help -1.772e+02

## industryTobacco -2.013e+15

## industryToys, Sporting Goods 3.294e+02

## industryTransportation Equipment 1.986e+01

## industryTransportation and Logistics 2.015e+01

## industryTrucking, Truck Leasing NA

## industryUtilities: Gas and Electric 4.504e+15

## industryWaste Management -1.677e+02

## industryWholesalers: Diversified 3.971e+15

## industryWholesalers: Electronics and Office Equipment 3.971e+15

## industryWholesalers: Food and Grocery 8.475e+15

## industryWholesalers: Health Care 8.475e+15

## revenue 1.136e-04

## rank 4.682e-03

## Std. Error

## (Intercept) 2.120e+00

## sectorApparel 2.247e+00

## sectorBusiness Services 4.745e+07

## sectorChemicals 4.949e+04

## sectorEnergy 9.087e+07

## sectorEngineering & Construction 2.205e+00

## sectorFinancials 2.322e+00

## sectorFood & Drug Stores 7.643e+04

## sectorFood, Beverages & Tobacco 5.254e+14

## sectorHealth Care 1.090e+08

## sectorHotels, Restaurants & Leisure 4.090e+14

## sectorHousehold Products 9.087e+07

## sectorIndustrials 6.127e+07

## sectorMaterials 9.087e+07

## sectorMedia 9.721e+07

## sectorMotor Vehicles & Parts 2.223e+00

## sectorRetailing 9.443e+07

## sectorTechnology 8.343e+07

## sectorTelecommunications 2.246e+07

## sectorTransportation 2.239e+00

## sectorWholesalers 8.689e+14

## industryAerospace and Defense NA

## industryAirlines 1.120e+00

## industryApparel NA

## industryAutomotive Retailing, Services 9.443e+07

## industryBeverages 5.254e+14

## industryBuilding Materials, Glass 9.087e+07

## industryChemicals NA

## industryCommercial Banks 1.194e+00

## industryComputer Software 8.343e+07

## industryComputers, Office Equipment 8.343e+07

## industryConstruction and Farm Machinery 6.570e+07

## industryDiversified Financials 1.201e+00

## industryDiversified Outsourcing Services 4.745e+07

## industryEducation 8.219e+07

## industryElectronics, Electrical Equip. 6.127e+07

## industryEnergy 9.223e+07

## industryEngineering, Construction 7.089e+04

## industryEntertainment 9.598e+07

## industryFinancial Data Services 4.956e+07

## industryFood Consumer Products 5.254e+14

## industryFood Production 5.254e+14

## industryFood Services 4.090e+14

## industryFood and Drug Stores NA

## industryForest and Paper Products 9.087e+07

## industryGeneral Merchandisers 9.443e+07

## industryHealth Care: Insurance and Managed Care 1.090e+08

## industryHealth Care: Medical Facilities 1.090e+08

## industryHealth Care: Pharmacy and Other Services 1.090e+08

## industryHome Equipment, Furnishings 9.087e+07

## industryHomebuilders NA

## industryHotels, Casinos, Resorts 4.090e+14

## industryHousehold and Personal Products 9.087e+07

## industryIndustrial Machinery 6.127e+07

## industryInformation Technology Services 8.343e+07

## industryInsurance: Life, Health (Mutual) 1.459e+00

## industryInsurance: Life, Health (stock) 1.250e+00

## industryInsurance: Property and Casualty (Mutual) 2.492e+05

## industryInsurance: Property and Casualty (Stock) 1.450e+00

## industryInternet Services and Retailing 8.585e+07

## industryMail, Package, and Freight Delivery 4.745e+07

## industryMedical Products and Equipment 1.102e+08

## industryMetals 9.087e+07

## industryMining, Crude-Oil Production 9.087e+07

## industryMiscellaneous 6.126e+07

## industryMotor Vehicles and Parts NA

## industryNetwork and Other Communications Equipment 8.343e+07

## industryOil and Gas Equipment, Services 9.087e+07

## industryPackaging, Containers 9.240e+07

## industryPetroleum Refining 9.331e+07

## industryPharmaceuticals 1.103e+08

## industryPipelines 9.087e+07

## industryPublishing, Printing 1.017e+08

## industryRailroads 1.169e+00

## industryReal estate 1.428e+00

## industryScientific,Photographic and Control Equipment 8.585e+07

## industrySecurities NA

## industrySemiconductors and Other Electronic Components 8.343e+07

## industryShipping 1.533e+02

## industrySpecialty Retailers: Apparel 9.443e+07

## industrySpecialty Retailers: Other 9.443e+07

## industryTelecommunications NA

## industryTemporary Help 5.615e+07

## industryTobacco 5.254e+14

## industryToys, Sporting Goods 1.130e+08

## industryTransportation Equipment 2.164e+04

## industryTransportation and Logistics 2.056e+04

## industryTrucking, Truck Leasing NA

## industryUtilities: Gas and Electric 9.087e+07

## industryWaste Management 5.812e+07

## industryWholesalers: Diversified 8.689e+14

## industryWholesalers: Electronics and Office Equipment 8.689e+14

## industryWholesalers: Food and Grocery 8.689e+14

## industryWholesalers: Health Care 8.689e+14

## revenue 5.727e-06

## rank 5.611e-04

## z value Pr(>|z|)

## (Intercept) 3.859e+01 < 2e-16

## sectorApparel -3.736e+01 < 2e-16

## sectorBusiness Services 6.000e-03 0.995103

## sectorChemicals -1.000e-03 0.999014

## sectorEnergy -4.956e+07 < 2e-16

## sectorEngineering & Construction -3.760e+01 < 2e-16

## sectorFinancials -3.565e+01 < 2e-16

## sectorFood & Drug Stores -1.000e-03 0.999342

## sectorFood, Beverages & Tobacco 3.832e+00 0.000127

## sectorHealth Care -4.133e+07 < 2e-16

## sectorHotels, Restaurants & Leisure 8.125e+00 4.48e-16

## sectorHousehold Products 4.956e+07 < 2e-16

## sectorIndustrials 0.000e+00 0.999827

## sectorMaterials 0.000e+00 0.999762

## sectorMedia 4.633e+07 < 2e-16

## sectorMotor Vehicles & Parts -3.674e+01 < 2e-16

## sectorRetailing 4.769e+07 < 2e-16

## sectorTechnology -5.000e-03 0.995931

## sectorTelecommunications 2.005e+08 < 2e-16

## sectorTransportation -3.564e+01 < 2e-16

## sectorWholesalers -4.571e+00 4.86e-06

## industryAerospace and Defense NA NA

## industryAirlines 4.720e-01 0.637069

## industryApparel NA NA

## industryAutomotive Retailing, Services -4.769e+07 < 2e-16

## industryBeverages -3.832e+00 0.000127

## industryBuilding Materials, Glass 0.000e+00 0.999762

## industryChemicals NA NA

## industryCommercial Banks -2.678e+00 0.007409

## industryComputer Software 5.000e-03 0.995932

## industryComputers, Office Equipment 5.000e-03 0.995932

## industryConstruction and Farm Machinery -1.620e-01 0.870947

## industryDiversified Financials -1.465e+00 0.143006

## industryDiversified Outsourcing Services -6.000e-03 0.995102

## industryEducation 5.479e+07 < 2e-16

## industryElectronics, Electrical Equip. 0.000e+00 0.999828

## industryEnergy 9.766e+07 < 2e-16

## industryEngineering, Construction 0.000e+00 0.999720

## industryEntertainment 2.700e-02 0.978186

## industryFinancial Data Services 2.000e-03 0.998056

## industryFood Consumer Products -3.832e+00 0.000127

## industryFood Production 4.740e+00 2.14e-06

## industryFood Services -8.125e+00 4.48e-16

## industryFood and Drug Stores NA NA

## industryForest and Paper Products 0.000e+00 0.999762

## industryGeneral Merchandisers -4.769e+07 < 2e-16

## industryHealth Care: Insurance and Managed Care 4.133e+07 < 2e-16

## industryHealth Care: Medical Facilities 4.133e+07 < 2e-16

## industryHealth Care: Pharmacy and Other Services 4.133e+07 < 2e-16

## industryHome Equipment, Furnishings -4.956e+07 < 2e-16

## industryHomebuilders NA NA

## industryHotels, Casinos, Resorts -8.125e+00 4.48e-16

## industryHousehold and Personal Products -4.956e+07 < 2e-16

## industryIndustrial Machinery 0.000e+00 0.999828

## industryInformation Technology Services 5.000e-03 0.995932

## industryInsurance: Life, Health (Mutual) -6.410e-01 0.521518

## industryInsurance: Life, Health (stock) -1.338e+00 0.180772

## industryInsurance: Property and Casualty (Mutual) -4.000e-03 0.996894

## industryInsurance: Property and Casualty (Stock) -4.365e+00 1.27e-05

## industryInternet Services and Retailing 2.000e-03 0.998093

## industryMail, Package, and Freight Delivery -3.801e+00 0.000144

## industryMedical Products and Equipment 8.170e+07 < 2e-16

## industryMetals 0.000e+00 0.999762

## industryMining, Crude-Oil Production 4.956e+07 < 2e-16

## industryMiscellaneous 0.000e+00 0.999828

## industryMotor Vehicles and Parts NA NA

## industryNetwork and Other Communications Equipment 5.000e-03 0.995932

## industryOil and Gas Equipment, Services 4.956e+07 < 2e-16

## industryPackaging, Containers 4.874e+07 < 2e-16

## industryPetroleum Refining 9.653e+07 < 2e-16

## industryPharmaceuticals 4.082e+07 < 2e-16

## industryPipelines 4.956e+07 < 2e-16

## industryPublishing, Printing 4.000e-02 0.968198

## industryRailroads -2.150e-01 0.830045

## industryReal estate -3.000e-03 0.997499

## industryScientific,Photographic and Control Equipment 5.246e+07 < 2e-16

## industrySecurities NA NA

## industrySemiconductors and Other Electronic Components 5.000e-03 0.995932

## industryShipping 5.400e-02 0.957022

## industrySpecialty Retailers: Apparel -4.769e+07 < 2e-16

## industrySpecialty Retailers: Other -4.769e+07 < 2e-16

## industryTelecommunications NA NA

## industryTemporary Help 0.000e+00 0.999997

## industryTobacco -3.832e+00 0.000127

## industryToys, Sporting Goods 0.000e+00 0.999998

## industryTransportation Equipment 1.000e-03 0.999268

## industryTransportation and Logistics 1.000e-03 0.999218

## industryTrucking, Truck Leasing NA NA

## industryUtilities: Gas and Electric 4.956e+07 < 2e-16

## industryWaste Management 0.000e+00 0.999998

## industryWholesalers: Diversified 4.571e+00 4.86e-06

## industryWholesalers: Electronics and Office Equipment 4.571e+00 4.86e-06

## industryWholesalers: Food and Grocery 9.753e+00 < 2e-16

## industryWholesalers: Health Care 9.753e+00 < 2e-16

## revenue 1.983e+01 < 2e-16

## rank 8.343e+00 < 2e-16

##

## (Intercept) \*\*\*

## sectorApparel \*\*\*

## sectorBusiness Services

## sectorChemicals

## sectorEnergy \*\*\*

## sectorEngineering & Construction \*\*\*

## sectorFinancials \*\*\*

## sectorFood & Drug Stores

## sectorFood, Beverages & Tobacco \*\*\*

## sectorHealth Care \*\*\*

## sectorHotels, Restaurants & Leisure \*\*\*

## sectorHousehold Products \*\*\*

## sectorIndustrials

## sectorMaterials

## sectorMedia \*\*\*

## sectorMotor Vehicles & Parts \*\*\*

## sectorRetailing \*\*\*

## sectorTechnology

## sectorTelecommunications \*\*\*

## sectorTransportation \*\*\*

## sectorWholesalers \*\*\*

## industryAerospace and Defense

## industryAirlines

## industryApparel

## industryAutomotive Retailing, Services \*\*\*

## industryBeverages \*\*\*

## industryBuilding Materials, Glass

## industryChemicals

## industryCommercial Banks \*\*

## industryComputer Software

## industryComputers, Office Equipment

## industryConstruction and Farm Machinery

## industryDiversified Financials

## industryDiversified Outsourcing Services

## industryEducation \*\*\*

## industryElectronics, Electrical Equip.

## industryEnergy \*\*\*

## industryEngineering, Construction

## industryEntertainment

## industryFinancial Data Services

## industryFood Consumer Products \*\*\*

## industryFood Production \*\*\*

## industryFood Services \*\*\*

## industryFood and Drug Stores

## industryForest and Paper Products

## industryGeneral Merchandisers \*\*\*

## industryHealth Care: Insurance and Managed Care \*\*\*

## industryHealth Care: Medical Facilities \*\*\*

## industryHealth Care: Pharmacy and Other Services \*\*\*

## industryHome Equipment, Furnishings \*\*\*

## industryHomebuilders

## industryHotels, Casinos, Resorts \*\*\*

## industryHousehold and Personal Products \*\*\*

## industryIndustrial Machinery

## industryInformation Technology Services

## industryInsurance: Life, Health (Mutual)

## industryInsurance: Life, Health (stock)

## industryInsurance: Property and Casualty (Mutual)

## industryInsurance: Property and Casualty (Stock) \*\*\*

## industryInternet Services and Retailing

## industryMail, Package, and Freight Delivery \*\*\*

## industryMedical Products and Equipment \*\*\*

## industryMetals

## industryMining, Crude-Oil Production \*\*\*

## industryMiscellaneous

## industryMotor Vehicles and Parts

## industryNetwork and Other Communications Equipment

## industryOil and Gas Equipment, Services \*\*\*

## industryPackaging, Containers \*\*\*

## industryPetroleum Refining \*\*\*

## industryPharmaceuticals \*\*\*

## industryPipelines \*\*\*

## industryPublishing, Printing

## industryRailroads

## industryReal estate

## industryScientific,Photographic and Control Equipment \*\*\*

## industrySecurities

## industrySemiconductors and Other Electronic Components

## industryShipping

## industrySpecialty Retailers: Apparel \*\*\*

## industrySpecialty Retailers: Other \*\*\*

## industryTelecommunications

## industryTemporary Help

## industryTobacco \*\*\*

## industryToys, Sporting Goods

## industryTransportation Equipment

## industryTransportation and Logistics

## industryTrucking, Truck Leasing

## industryUtilities: Gas and Electric \*\*\*

## industryWaste Management

## industryWholesalers: Diversified \*\*\*

## industryWholesalers: Electronics and Office Equipment \*\*\*

## industryWholesalers: Food and Grocery \*\*\*

## industryWholesalers: Health Care \*\*\*

## revenue \*\*\*

## rank \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for binomial family taken to be 1)

##

## Null deviance: 465.78 on 965 degrees of freedom

## Residual deviance: 9692.64 on 879 degrees of freedom

## AIC: 9866.6

##

## Number of Fisher Scoring iterations: 25

varImp(model2, scale = FALSE)

## Overall

## sectorApparel 3.736227e+01

## sectorBusiness Services 6.137501e-03

## sectorChemicals 1.235962e-03

## sectorEnergy 4.956318e+07

## sectorEngineering & Construction 3.760133e+01

## sectorFinancials 3.565395e+01

## sectorFood & Drug Stores 8.244514e-04

## sectorFood, Beverages & Tobacco 3.832234e+00

## sectorHealth Care 4.133113e+07

## sectorHotels, Restaurants & Leisure 8.124754e+00

## sectorHousehold Products 4.956318e+07

## sectorIndustrials 2.165016e-04

## sectorMaterials 2.987139e-04

## sectorMedia 4.633061e+07

## sectorMotor Vehicles & Parts 3.674323e+01

## sectorRetailing 4.769426e+07

## sectorTechnology 5.099730e-03

## sectorTelecommunications 2.004923e+08

## sectorTransportation 3.564228e+01

## sectorWholesalers 4.570545e+00

## industryAirlines 4.718014e-01

## industryAutomotive Retailing, Services 4.769426e+07

## industryBeverages 3.832234e+00

## industryBuilding Materials, Glass 2.978140e-04

## industryCommercial Banks 2.677876e+00

## industryComputer Software 5.098798e-03

## industryComputers, Office Equipment 5.098760e-03

## industryConstruction and Farm Machinery 1.624557e-01

## industryDiversified Financials 1.464689e+00

## industryDiversified Outsourcing Services 6.139239e-03

## industryEducation 5.479416e+07

## industryElectronics, Electrical Equip. 2.155893e-04

## industryEnergy 9.765774e+07

## industryEngineering, Construction 3.513649e-04

## industryEntertainment 2.734312e-02

## industryFinancial Data Services 2.436697e-03

## industryFood Consumer Products 3.832234e+00

## industryFood Production 4.739664e+00

## industryFood Services 8.124754e+00

## industryForest and Paper Products 2.980543e-04

## industryGeneral Merchandisers 4.769426e+07

## industryHealth Care: Insurance and Managed Care 4.133113e+07

## industryHealth Care: Medical Facilities 4.133113e+07

## industryHealth Care: Pharmacy and Other Services 4.133113e+07

## industryHome Equipment, Furnishings 4.956318e+07

## industryHotels, Casinos, Resorts 8.124754e+00

## industryHousehold and Personal Products 4.956318e+07

## industryIndustrial Machinery 2.151258e-04

## industryInformation Technology Services 5.098782e-03

## industryInsurance: Life, Health (Mutual) 6.410069e-01

## industryInsurance: Life, Health (stock) 1.338382e+00

## industryInsurance: Property and Casualty (Mutual) 3.892717e-03

## industryInsurance: Property and Casualty (Stock) 4.364706e+00

## industryInternet Services and Retailing 2.389839e-03

## industryMail, Package, and Freight Delivery 3.800554e+00

## industryMedical Products and Equipment 8.169950e+07

## industryMetals 2.978246e-04

## industryMining, Crude-Oil Production 4.956318e+07

## industryMiscellaneous 2.156405e-04

## industryNetwork and Other Communications Equipment 5.098982e-03

## industryOil and Gas Equipment, Services 4.956311e+07

## industryPackaging, Containers 4.873936e+07

## industryPetroleum Refining 9.652855e+07

## industryPharmaceuticals 4.081825e+07

## industryPipelines 4.956318e+07

## industryPublishing, Printing 3.986806e-02

## industryRailroads 2.146441e-01

## industryReal estate 3.134246e-03

## industryScientific,Photographic and Control Equipment 5.246141e+07

## industrySemiconductors and Other Electronic Components 5.098774e-03

## industryShipping 5.389084e-02

## industrySpecialty Retailers: Apparel 4.769426e+07

## industrySpecialty Retailers: Other 4.769426e+07

## industryTemporary Help 3.155214e-06

## industryTobacco 3.832234e+00

## industryToys, Sporting Goods 2.915807e-06

## industryTransportation Equipment 9.175981e-04

## industryTransportation and Logistics 9.799625e-04

## industryUtilities: Gas and Electric 4.956318e+07

## industryWaste Management 2.884875e-06

## industryWholesalers: Diversified 4.570545e+00

## industryWholesalers: Electronics and Office Equipment 4.570545e+00

## industryWholesalers: Food and Grocery 9.753453e+00

## industryWholesalers: Health Care 9.753453e+00

## revenue 1.982783e+01

## rank 8.343462e+00

f500$sexPred2 <- predict(model2, f500, type="response")

head(f500[, c(4, 11, 12, 13, 14)], n = 100)

## ceo sex sexCoef sexPred1 sexPred2

## 1 C. Douglas McMillon M 0 1.890554e-01 7.611609e-04

## 2 Darren W. Woods M 0 2.716634e-03 1.000000e+00

## 3 Warren E. Buffett M 0 3.882830e-02 1.000000e+00

## 4 Timothy D. Cook M 0 2.716567e-03 1.000000e+00

## 5 David S. Wichmann M 0 1.022892e-01 1.000000e+00

## 6 John H. Hammergren M 0 1.682775e-03 1.000000e+00

## 7 Larry J. Merlo M 0 1.554745e-01 9.999978e-01

## 8 Jeffrey P. Bezos M 0 1.611212e-03 1.000000e+00

## 9 Randall L. Stephenson M 0 1.474235e-03 1.000000e+00

## 10 Mary T. Barra F 1 5.399019e-02 1.000000e+00

## 11 James P. Hackett M 0 5.398270e-02 1.000000e+00

## 12 Steven H. Collis M 0 8.626237e-04 1.000000e+00

## 13 Michael K. Wirth M 0 7.238309e-04 1.000000e+00

## 14 Michael C. Kaufmann M 0 4.415029e-04 1.000000e+00

## 15 W. Craig Jelinek M 0 1.822637e-01 2.220446e-16

## 16 Hans E. Vestberg M 0 8.563350e-04 1.000000e+00

## 17 W. Rodney McMullen M 0 7.742411e-04 1.000000e+00

## 18 H. Lawrence Culp Jr. M 0 3.765187e-02 9.999897e-01

## 19 Stefano Pessina M 0 6.971764e-04 1.000000e+00

## 20 James Dimon M 0 1.207978e-01 9.998587e-01

## 21 Timothy J. Mayopoulos M 0 5.647020e-02 9.999603e-01

## 22 Larry Page M 0 5.949939e-04 2.220446e-16

## 23 Craig A. Menear M 0 6.727648e-02 9.985940e-01

## 24 Brian T. Moynihan M 0 1.205566e-01 9.993480e-01

## 25 Timothy C. Wentworth M 0 1.539600e-01 9.701039e-01

## 26 Timothy J. Sloan M 0 1.205149e-01 9.991400e-01

## 27 Dennis A. Muilenburg M 0 1.602847e-01 1.000000e+00

## 28 Greg C. Garland M 0 -3.068588e-05 1.000000e+00

## 29 Gail K. Boudreaux F 1 1.003031e-01 1.000000e+00

## 30 Satya Nadella F 1 7.688099e-02 9.999995e-01

## 31 Joseph W. Gorder M 0 -8.175417e-05 1.000000e+00

## 32 Michael L. Corbat M 0 1.203493e-01 9.974671e-01

## 33 Brian L. Roberts M 0 1.332818e-04 1.000000e+00

## 34 Virginia M. Rometty F 1 6.261575e-02 9.999932e-01

## 35 Michael S. Dell M 0 2.177975e-05 9.999535e-01

## 36 Michael L. Tipsord M 0 1.055853e-04 2.220446e-16

## 37 Alex Gorsky M 0 3.519263e-04 2.220446e-16

## 38 Donald H. Layton M 0 5.581680e-02 9.973534e-01

## 39 Brian C. Cornell M 0 1.812697e-01 2.220446e-16

## 40 Marvin R. Ellison M 0 6.672294e-02 9.516239e-01

## 41 Gary R. Heminger M 0 -4.405571e-04 1.000000e+00

## 42 David S. Taylor M 0 2.141602e-01 9.999006e-01

## 43 Steven A. Kandarian M 0 1.052986e-01 9.937826e-01

## 44 David P. Abney M 0 4.400154e-05 2.220446e-16

## 45 Ramon Laguarta M 0 1.580135e-01 9.995355e-01

## 46 Robert H. Swan M 0 4.323241e-02 9.999173e-01

## 48 Juan R. Luciano M 0 -4.168550e-05 1.000000e+00

## 49 Mark T. Bertolini M 0 9.980773e-02 9.999993e-01

## 50 Frederick W. Smith M 0 -4.400154e-05 2.220446e-16

## 51 Gregory J. Hayes M 0 1.597242e-01 1.000000e+00

## 52 John R. Strangfeld M 0 1.052009e-01 9.876577e-01

## 53 Robert G. Miller M 0 -2.990069e-04 1.000000e+00

## 54 Thomas L. Ben\357\277\275 M 0 2.653679e-04 1.000000e+00

## 55 Robert A. Iger M 0 -2.751090e-04 1.000000e+00

## 56 Bruce D. Broussard M 0 9.969973e-02 9.999984e-01

## 57 Ian C. Read M 0 -4.055398e-05 2.220446e-16

## 58 Dion J. Weisler M 0 -4.132960e-04 9.991435e-01

## 59 Marillyn A. Hewson F 1 1.595814e-01 1.000000e+00

## 60 Brian Duperreault M 0 3.542171e-02 1.994353e-01

## 61 Michael F. Neidorff M 0 9.961597e-02 9.999972e-01

## 62 Charles H. Robbins M 0 -4.114802e-04 1.000000e+00

## 64 John W. McReynolds M 0 -3.376170e-04 2.023511e-05

## 65 D. James Umpleby III M 0 -1.271278e-04 2.220446e-16

## 66 Stephen S. Rasmussen M 0 -4.562080e-04 2.220446e-16

## 67 James P. Gorman M 0 1.196167e-01 7.514953e-01

## 68 David H. Long M 0 3.531486e-02 1.063784e-01

## 69 Theodore A. Mathas M 0 9.973740e-02 9.617484e-01

## 70 David M. Solomon M 0 1.195982e-01 7.237296e-01

## 71 W. Douglas Parker M 0 1.108219e-01 9.995248e-01

## 72 Hubert B. Joly M 0 6.631144e-02 5.308273e-01

## 73 David M. Cordani M 0 9.951624e-02 9.999943e-01

## 74 Thomas M. Rutledge M 0 -5.599586e-04 1.000000e+00

## 75 Edward H. Bastian M 0 1.108136e-01 9.994798e-01

## 76 Mark Zuckerberg M 0 -5.687503e-04 2.220446e-16

## 77 Darius Adamczyk M 0 -5.909111e-04 1.000000e+00

## 78 Kenneth C. Frazier M 0 -2.197000e-04 2.220446e-16

## 79 Thomas J. Wilson M 0 3.526412e-02 7.246009e-02

## 80 Noel White M 0 -3.814996e-04 1.000000e+00

## 81 Oscar Munoz M 0 1.107632e-01 9.992469e-01

## 83 Richard T. Hume M 0 -2.702330e-04 1.000000e+00

## 84 Roger W. Ferguson Jr. M 0 9.965729e-02 9.297425e-01

## 85 Ernie L. Herrman M 0 2.992858e-01 2.220446e-16

## 86 Stephen J. Squeri M 0 5.521076e-02 8.477981e-01

## 87 James R. Quincey M 0 -3.899915e-04 2.220446e-16

## 88 Randall T. Jones Sr. M 0 -6.735761e-04 1.000000e+00

## 89 Mark G. Parker M 0 7.635021e-02 8.962419e-01

## 91 Michael J. Kasbar M 0 -6.143238e-04 1.000000e+00

## 92 Christopher M. Crane M 0 2.420212e-01 7.014055e-04

## 93 Roger W. Crandall M 0 9.963188e-02 9.119435e-01

## 94 John T. Standley M 0 -6.961258e-04 1.000000e+00

## 95 Ryan M. Lance M 0 4.485453e-02 1.426161e-04

## 96 Jay D. Debertin M 0 -4.602686e-04 1.000000e+00

## 97 Michael F. Roman M 0 -2.675269e-17 1.000000e+00

## 98 John Stankey M 0 -6.130101e-04 1.000000e+00

## 99 Phebe N. Novakovic F 1 1.593062e-01 1.000000e+00

## 100 Stuart Parker M 0 3.515694e-02 3.175881e-02

## 101 Richard D. Fairbank M 0 1.194456e-01 4.296168e-01

## 102 Samuel R. Allen M 0 -3.294479e-04 2.220446e-16

## 103 Sean M. O'Connor M 0 5.513738e-02 7.498118e-01

## 104 John E. Schlifske M 0 9.958111e-02 8.717164e-01

# 4 Resultats i conclusions.

Quant als sectors podem dir que predominen el sector financer, tecnològic i energètic.

Dintre de cada sector podem destacar:

* En el sector financer la majoria de companyies són companyies asseguradores, bancs, o de gestió de propietats immobiliàries.
* En el sector tecnològic les principals companyies s’encarreguen de la creació de components electrònics, seguidament tenim les companyies que es dediquen a crear software i després les companyies orientades a les tecnologies de la informació i comunicació.
* En el sector energètic predominen dos tipus de companyies, les que es dediquen a proveir de gas i electricitat i les companyies que es dediquen a l’extracció de minerals i petroli.

Referent al nombre de dones i homes que són CEOs de les empreses de la llista fortune, podem dir que la gran majoria d’empreses (en un 92,1%) estan dirigides per homes, mentre que el 7,86% restant tenen a una dona com CEO.

Els models predictius extrets entre el CEO i el tipus de sector i indústria no han demostrat cap relació entre aquests fets, segurament degut al tipus de variables no són les adequades o el nombre de mostres és molt petit.

# 5 Dataset resultant amb les dades netes.

write.csv(f500, file=“fortune\_clean.csv”, sep = “;”)