

OTIC ANALYZER

Alexander Cole Mora

Rafael Herrero Álvarez

Raúl Martín Morales

Daniel Ramos Acosta

Pedro Ramos Rodríguez

Moisés Yanes Carballo



OTIC ANALYZER

Práctica

Los paquetes utilizados para realizar los scripts han sido:

- **Lsa**: realizar análisis de lenguaje semántico.
- **Lattice**: crear gráficos multivariable para crear una visualización de los datos.
- **Stringr** : realizar tratamientos con cadenas.
- **Tm**: minería de texto.
- **Wordcloud**: creación de un gráfico de nube de palabras.

Script del clima

```
library (tm)
library (wordcloud)
library (lsa)

#Cargamos ficheros
raiz <- setwd("~/")
texto1 <-
readLines(paste(raiz,"/proyecto-final-TIO/datos/londres/Clima1.txt"
,sep = ""),encoding="UTF-8")
texto1 = iconv(texto1, to="ASCII//TRANSLIT")

texto2 <-
readLines(paste(raiz,"/proyecto-final-TIO/datos/londres/Clima2.txt"
,sep = ""),encoding="UTF-8")
texto2 = iconv(texto2, to="ASCII//TRANSLIT")

#Unimos todos los ficheros bajo una única lista y creamos el corpus
docs <- list(texto1,texto2)
corpus <- Corpus(VectorSource(docs))

#Limpiamos el conjunto de los documentos de numeros, espacios en
blanco, etc...
d <- tm_map(corpus, content_transformer(tolower)) #Lo convierte a
minúsculas
```

```
d <- tm_map(d, removePunctuation) #Elimina signos de puntuación
d <- tm_map(d, removeNumbers) #Elimina los números
d <- tm_map(d, stripWhitespace) #Elimina los espacios en blanco
d <- tm_map(d, removeWords,
c("london","weather","climate","temperatures","temperature","year",
"months",
"usually","although","less","day","city","cities","likely","capit
al","normally",
"always","however","sometimes","throughout","idea","common",stopw
ords("english"))) #Elimina palabras vacías en inglés
```

```
#Creamos la matriz de términos
tdm <- TermDocumentMatrix(d)
m = as.matrix(tdm)
```

```
# conteo de palabras en orden decreciente
wf <- sort(rowSums(m),decreasing=TRUE)
# crea un data frame con las palabras y sus frecuencias
dm <- data.frame(word = names(wf), freq=wf)
# Frecuencia minima igual a 20
```

```
wordcloud(dm$word, dm$freq, min.freq = 1,
          max.words=50, random.order=FALSE,
          colors=brewer.pal(8,"Dark2"))
```

Script de lugares

```
library (tm)
library (wordcloud)
library (lsa)
#Cargamos ficheros
raiz <- setwd("~/")
texto1 <-
readLines(paste(raiz,"/proyecto-final-TIO/datos/londres/Lugares1.tx
t",sep = ""),encoding="UTF-8")
texto1 = iconv(texto1, to="ASCII//TRANSLIT")

texto2 <-
readLines(paste(raiz,"/proyecto-final-TIO/datos/londres/Lugares2.tx
t",sep = ""),encoding="UTF-8")
texto2 = iconv(texto2, to="ASCII//TRANSLIT")

texto3 <-
readLines(paste(raiz,"/proyecto-final-TIO/datos/londres/Lugares3.tx
t",sep = ""),encoding="UTF-8")
texto3 = iconv(texto2, to="ASCII//TRANSLIT")
#Unimos todos los ficheros bajo una única lista y creamos el corpus
docs <- list(texto1,texto2,texto3)
corpus <- Corpus(VectorSource(docs))

#Limpiamos el conjunto de los documentos de numeros, espacios en
blanco, etc...
d <- tm_map(corpus, content_transformer(tolower)) #Lo convierte a
minúsculas
```

```
d <- tm_map(d, removePunctuation) #Elimina signos de puntuación
d <- tm_map(d, removeNumbers) #Elimina los números
d <- tm_map(d, stripWhitespace) #Elimina los espacios en blanco
d<-tm_map(d, removeWords,
c("london","londons","abbey","cast","count","end","fine","inside
","mall","new","plan","rich","sir","stay","top","war","aboard","
time","also","known","see","pauls","two","take","big","home","sto
pwords("english"))) #Elimina palabras vacías en inglés
```

```
#Creamos la matriz de términos
tdm <- TermDocumentMatrix(d)
m = as.matrix(tdm)
```

```
# conteo de palabras en orden decreciente
wf <- sort(rowSums(m),decreasing=TRUE)
# crea un data frame con las palabras y sus frecuencias
dm <- data.frame(word = names(wf), freq=wf)
# Frecuencia minima igual a 20
```

```
counts <- table (dm$word)
```

```
wordcloud(dm$word, dm$freq, min.freq = 1,
          max.words=50, random.order=FALSE,
          colors=brewer.pal(8,"Dark2"))
```

Script de procedencia

```
# Poner la ruta de donde está el repo
rutaBase <- "D:\\Documentos\\Repos\\Clases\\tio\\proyecto"

rutaCSV <- paste(rutaBase, "/datos/nuevayork/visitantes.csv", sep="")
rutaOutput <- paste(rutaBase, "/datos/nuevayork/Graficos/", sep="")

data <- read.csv(rutaCSV, header=TRUE)

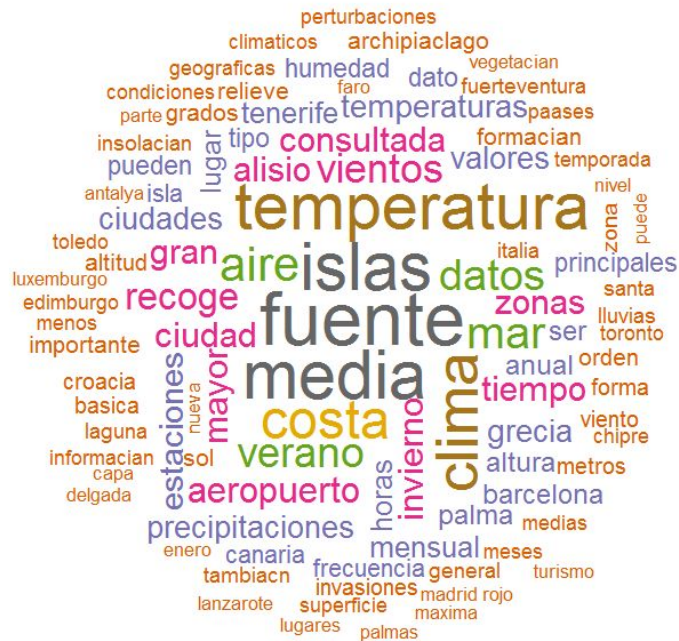
for(i in 2:ncol(data)) {
  visitantes <- data[, c(1, i)]

  png(filename=paste(rutaOutput, i-1, ".png"),
      units="px",
      width=1024,
      height=700,
      pointsize=12,
      res=72)
  plot(visitantes,type="o")
  dev.off()
}
```

Canarias. Clima

Palabras Destacables:

- ◆ Temperatura
- ◆ Precipitaciones
- ◆ Clima
- ◆ Tiempo
- ◆ Zonas
- ◆ Mar
- ◆ Costa
- ◆ Lluvias
- ◆ Alisio
- ◆ Vientos



Canarias. Lugares

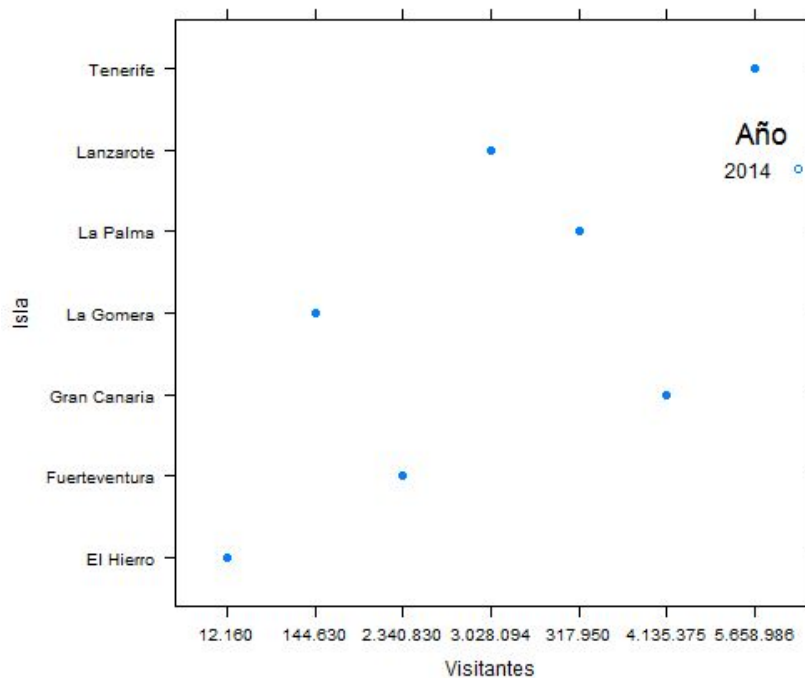
Palabras destacables:

- ◆ Tenerife
- ◆ Teide
- ◆ Playas
- ◆ Paque
- ◆ Iglesias
- ◆ Capital
- ◆ Adeje
- ◆ Garachico

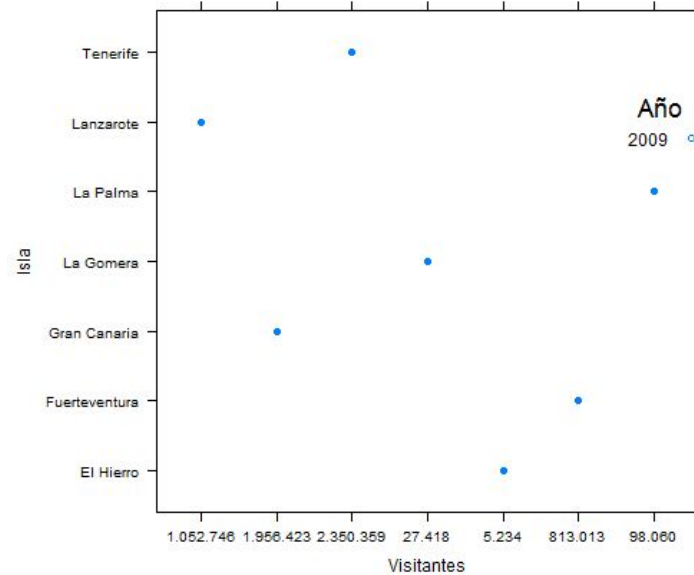


Canarias. Procedencia

Visitantes islas Canarias



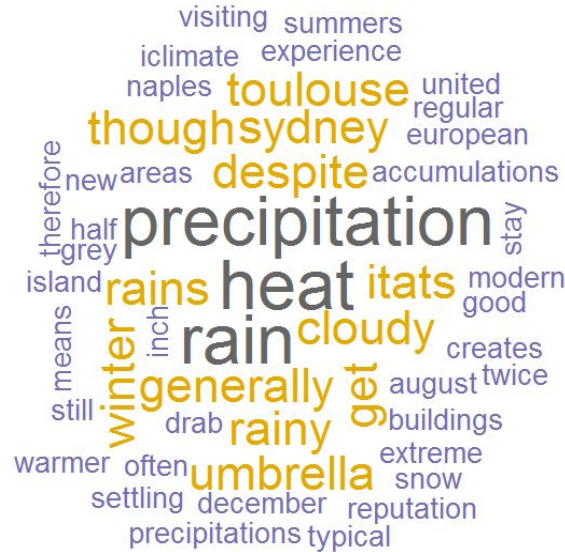
Visitantes islas Canarias



Londres. Clima

Palabras Destacables:

- ◆ Precipitation
- ◆ Rain
- ◆ Rainy
- ◆ Cloudy
- ◆ Umbrella
- ◆ December
- ◆ August
- ◆ Summers
- ◆ IClimate



Londres. Lugares

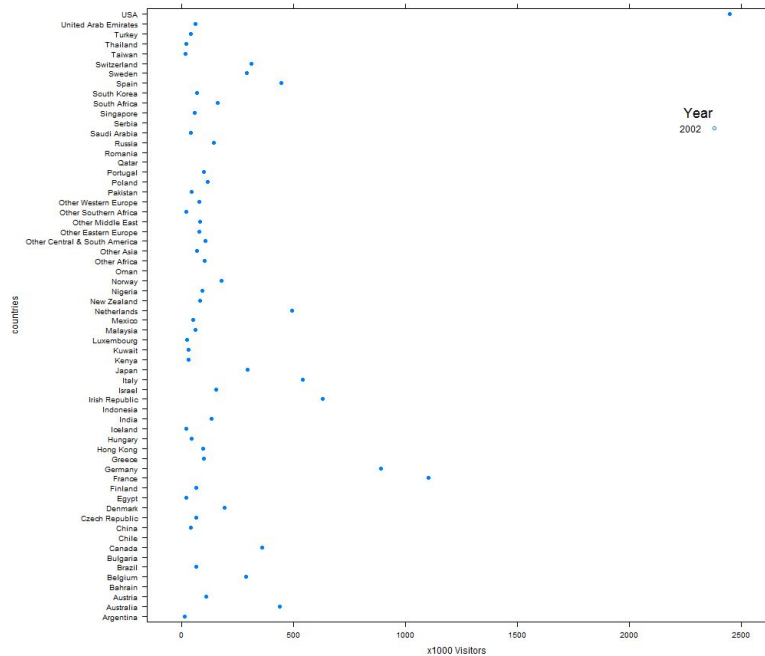
Palabras Destacables:

- ◆ Tower
- ◆ Palace
- ◆ Buckingham
- ◆ Tower
- ◆ Gallery
- ◆ Gardens
- ◆ Square
- ◆ Cathedral

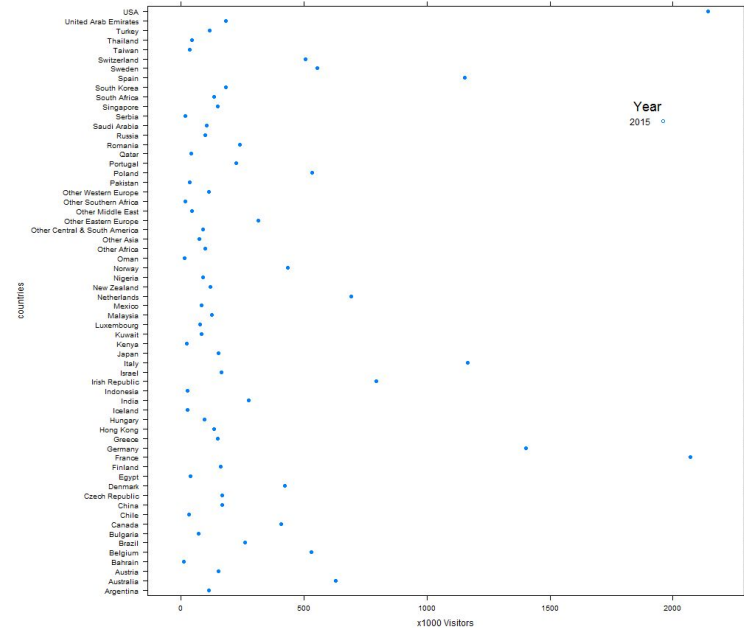


Londres. Procedencia

Visitors London (2002)



Visitors London (2015)



Nueva York. Clima

Palabras Clave:

- ◆ Clima
- ◆ Niebla
- ◆ Aire
- ◆ Tormenta
- ◆ Nubes



Nueva York. Lugares

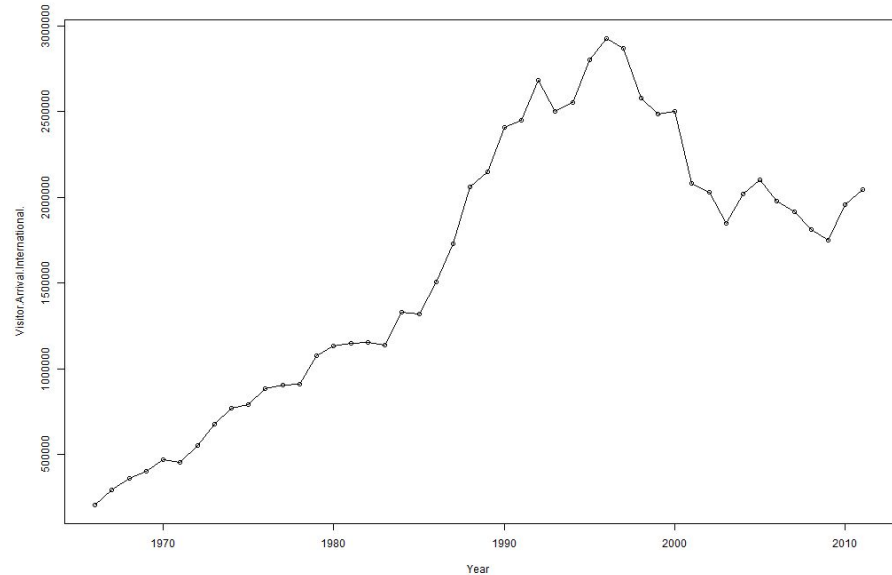
Palabras Destacables:

- ◆ Central Park
- ◆ Broadway
- ◆ Manhattan
- ◆ Museum
- ◆ Avenue
- ◆ Madison
- ◆ Brooklyn
- ◆ Square



Nueva York. Procedencia

- ◆ Crecimiento hasta 1995
- ◆ Luego empezó a descender
- ◆ ¿Posible entrada de la crisis?



Modelo de Negocio

El objetivo es obtener información útil de diferentes lugares para venderla a diferentes empresas y organizaciones:

- ◆ TourOperadores
- ◆ Compañías de vuelo, barcos
- ◆ Alquileres de coches
- ◆ Agencia de Viajes
- ◆ Gobiernos

Conclusiones

- ◆ Organización usando diagramas de Gantt.
- ◆ Analizar temporadas para asistir a los lugares.
- ◆ Lugares más visitados y su afluencia.
- ◆ Gráfico con procedencia y cantidad de turistas.
- ◆ Posibilidad de crear paquetes turísticos.



FIN

Alexander Cole Mora

Rafael Herrero Álvarez

Raúl Martín Morales

Daniel Ramos Acosta

Pedro Ramos Rodríguez

Moisés Yanes Carballo