Hemos usado un excel con las columnas: search\_term y year. Excel: data

##################Estudio de la evolución temporal de los tuits####################

install.packages("readxl")

library(readxl)

dateprova <- read\_excel("data.xlsx", sheet = 1)

str\_count(dateprova)

str(dateprova) #str para la tabla en variables independientes (para trabajar con ellas de forma independiente)

table(dateprova$search\_term) #Frecuencias por cada search\_term

d1 <- data.frame(table(dateprova$search\_term)) #lo ordenamos para verlo de forma mejor en un data.frame

##########FRECUENCIAS DE NUESTRAS DOS CATEGORIAS: RECREATIONAL FISHING Y RECREATIONAL HUNTING##################

#Vector para unir todas las palabras de fish en una nueva columna

library(dplyr)

dateprova <- dateprova %>%

mutate(nueva\_categoria = ifelse(search\_term %in% c("recreational\_fish", "recreational\_fishers", "recreational\_fisheries", "recreational\_fishery", "recreational\_fishing", "recreational\_fisher"), "recreational fishing", search\_term))

#Vector para unir todas las palabras de hunting en una nueva columna

dateprova <- dateprova %>%

mutate(nueva\_categoria2 = ifelse(search\_term %in% c("recreational\_hunt", "recreational\_hunter", "recreational\_hunting", "recreational\_hunters"), "recreational hunting", search\_term))

str(dateprova) #str para la tabla en variables independientes (para trabajar con ellas de forma independiente)

table(dateprova$nueva\_categoria)

str(dateprova) #str para la tabla en variables independientes (para trabajar con ellas de forma independiente)

table(dateprova$nueva\_categoria2)

######Para unir en una nueva columna "recreational fishing" y "recreational hunting"

dateprova$topic <-ifelse(dateprova$nueva\_categoria == "recreational fishing", "recreational fishing",

ifelse(dateprova$nueva\_categoria2 == "recreational hunting", "recreational hunting", NA))

a <- subset(dateprova, select = c("year", "topic" ))

##########################EVOLUCIÓN TUITS CON EL TIEMPO DE LAS DOS CATEGORÍAS######################################

###Arreglar datos para la gráfica de evolución temporal

install.packages("stringr")

library(stringr)

data.frame(a$topic, a$year, fish=str\_count(a$topic,"fish"), hunt=str\_count(a$topic,"hunt")) #Para que cuente con 0 y 1 los datos que empiezan por fish y por hunt, en vd no cal hacer lo anterior, se podría haber hecho directamente así, pero así tenemos un data.frame ordenado y no tenemos que cambiar los nombres otra vez

dy<-data.frame(a$topic, a$year, fish=str\_count(a$topic,"fish"), hunt=str\_count(a$topic,"hunt")) #le damos un nombre al nuevo dataframe

aggregate(dy$fish, by=list(Category=dy$a.year), FUN=sum) #aggregate=sumar para cada categoria (año)

aggregate(dy$hunt, by=list(Category=dy$a.year), FUN=sum)

tr<-aggregate(dy$fish, by=list(Category=dy$a.year), FUN=sum) #Para que sume los tuits de recreational fishing por año

trh<-aggregate(dy$hunt, by=list(Category=dy$a.year), FUN=sum) #Para que sume los tuits de recreational hunting por año

str(tr)

str(trh)

gra <- data.frame(year=tr$Category, fishing=tr$x, hunting=trh$x) #tabla resumen (data frame dónde tenemos la columna de año y una columna para fishing y otra para hunting)

##########Gráfica evolución temporal del número de tuits de cada categoría en el tiempo

install.packages("ggplot2")

library(ggplot2)

ggplot(gra, aes(year)) +

geom\_line(aes(y = fishing, colour ="fishing")) +

geom\_line(aes(y = hunting, colour = "hunting"))

###me da error, por el tipo de datos que tengo en mi data.frame

class(gra$year) #para mirar que tipo de datos tengo en la columna de year, me da un caràcter (palabra) y quiero un valor númerico

#Mi tabla estaba en formato categórico y para hacer el ggplot necesitamos datos númericos

EN INGLÉS

ggplot(gra, aes(as.numeric(year))) +

geom\_line(aes(y = as.numeric(fishing), colour = "fishing")) +

geom\_line(aes(y = as.numeric(hunting), colour = "hunting")) +

scale\_x\_continuous(limits = c(2007, 2022), breaks = seq(2002, 2022, by = 1)) + #Para que el eje de la X vaya desde el 2012 hasta el 2022, de año en año

scale\_y\_continuous(limits = c(0, 6000), breaks = seq(0, 6000, by = 500))+

labs(x="Years",y="Number of tweets", title= "Temporal evolution and variation of tweets from 2007 to 2022: a longitudinal study of social media activity", colour = "Topic") +

scale\_color\_manual(labels = c("Recreational fishing", "Recreational hunting"), values = c("cadetblue", "coral3")) +

theme\_classic() + #para ponerle un fondo blanco y sin rayas

theme(plot.title = element\_text(hjust = 0.5), text = element\_text(size = 16))

EN CASTELLANO

ggplot(gra, aes(as.numeric(year))) +

geom\_line(aes(y = as.numeric(fishing), colour = "fishing")) +

geom\_line(aes(y = as.numeric(hunting), colour = "hunting")) +

scale\_x\_continuous(limits = c(2007, 2022), breaks = seq(2002, 2022, by = 1)) + #Para que el eje de la X vaya desde el 2012 hasta el 2022, de año en año

scale\_y\_continuous(limits = c(0, 6000), breaks = seq(0, 6000, by = 500))+

labs(x="Años",y="Número de tuits", title= "Evolución temporal y variación de los tuits desde 2007 hasta 2022: un estudio longitudinal de la actividad en las redes sociales", colour = "Temática") +

scale\_color\_manual(labels = c("Pesca recrativa", "caza recreativa"), values = c("cadetblue", "coral3")) +

theme\_classic() + #para ponerle un fondo blanco y sin rayas

theme(plot.title = element\_text(hjust = 0.5), text = element\_text(size = 16))