Hemos usado un excel con las columnas: search\_term, month i year (excel boxplot\_date)

###################Gráfica de número total de tuits por mes##################

###Antes teníamos una columna de años y otra de categoría (fishing y hunting), ahora necesitamos la columna de categoría con la de los meses y años

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

b <- subset(dateprova, select = c("search\_term", "month", "year" )) #Seleccionar columnas que interesan

#Para cambiar los números por meses de la columna "month"

b$month <-ifelse(b$month == "01", "January",

ifelse(b$month == "02", "February",

ifelse(b$month == "03", "March",

ifelse(b$month == "04", "April",

ifelse(b$month == "05", "May",

ifelse(b$month == "06", "June",

ifelse(b$month == "07", "July",

ifelse(b$month == "08", "August",

ifelse(b$month == "09", "September",

ifelse(b$month == "10", "October",

ifelse(b$month == "11", "November",

ifelse(b$month == "12", "December", NA))))))))))))

#Unir las temáticas en "Recreational Fishing" y "Recreational hunting"

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

library(dplyr)

b <- b %>%

mutate(nueva\_categoria = ifelse(search\_term %in% c("recreational fish", "recreational fisher", "recreational fisheries", "recreational fishery", "recreational fishing", "recreational fishers"), "recreational fishing", search\_term))

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

b <- b %>%

mutate(nueva\_categoria2 = ifelse(search\_term %in% c("recreational hunt", "recreational hunter", "recreational hunting", "recreational hunters"), "recreational hunting", search\_term))

#Para comprobar si nos ha hecho bien los agrupamientos

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

table(b$nueva\_categoria)

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

table(b$nueva\_categoria2)

head(b) #para ver los primeros valores del data frame

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

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

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

table(b$topic) #Para ver cuántos datos tenemos de cada temática y comprovar que se nos ha hecho bien la unión

b <- subset(b, select = c("topic", "month", "year" ))

library(stringr)

data.frame(b$topic,b$month,b$year,fish=str\_count(b$topic,"fish"), hunt=str\_count(b$topic,"hunt")) #Para que cuente con 0 y 1 los datos que empiezan por fish y por hunt

dy2<-data.frame(b$topic, b$month, b$year, fish=str\_count(b$topic,"fish"), hunt=str\_count(b$topic,"hunt")) #le damos un nombre al nuevo dataframe

fish<-aggregate(dy2$fish, by=list(Category=dy2$b.month, dy2$b.year), FUN=sum)

hunt<-aggregate(dy2$hunt, by=list(Category=dy2$b.month, dy2$b.year), FUN=sum)

sum(dy2$fish)

sum(dy2$hunt)

fish<-data.frame(category="fish", month= fish$Category, year= fish$Group.2, freq= fish$x)#Reagrupar los datos anteriores: añadimos la columna fish y cambiamos el nombre del resto de columnas

hunt<-data.frame(category="hunt", month= hunt$Category, year= hunt$Group.2, freq= hunt$x)

testprova<-rbind(fish, hunt) #Para fusionar las dos matrices fish y hunt

###############################Boxplot###############################

EN INGLÉS

testprova$month<-factor(x=testprova$month,levels=month.name)

ggplot(data = testprova, aes(x = month, y = freq, fill = category)) +

stat\_boxplot(geom = "errorbar") +

geom\_boxplot(outlier.colour = NA) +

scale\_y\_continuous(limits = c(0, 600), breaks = seq(0, 600, by = 50)) +

scale\_x\_discrete(labels=c("January","February","March","April","May","June","July","August","September","October","November","December")) +

labs(x="Months",y="Number of tweets", title= "Uncovering temporal trends: a distributional analysis of monthly tweet activity from 2007 to 2022", color = "Tematica") +

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

theme\_classic() +

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

EN CASTELLANO

ggplot(data = testprova, aes(x = month, y = freq, fill = category)) +

stat\_boxplot(geom = "errorbar") +

geom\_boxplot(outlier.colour = NA) +

scale\_y\_continuous(limits = c(0, 600), breaks = seq(0, 600, by = 50)) +

scale\_x\_discrete(labels=c("Enero","Febrero","Marzo","Abril","Mayo","Junio","Julio","Agosto","Septiembre","Octubre","Noviembre","Diciembre")) +

labs(x="Meses",y="Número de tuits", title= "Descubriendo tendencias temporales: un análisis de distribución de la actividad mensual de tweets de 2007 a 2022", color = "Tematica") +

scale\_fill\_manual(values = c("cadetblue", "coral3"),name = "Temática", labels = c("Pesca recreativa", "Caza recreativa")) +

theme\_classic() +

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