Avance 2

Grupo 6

21/1/2023

##Visualización de Datos

Luego de haber scrappeado las páginas que ofrecen empleos en el ámbito de desarrolladores a nivel nacional, se obtiene el archivo “final.csv” el cual es el data frame a usarse para responder a las preguntas del proyecto por medio de una visualización con el uso de R.

library(readr)  
library(stringr)  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

datos\_final <- read.csv("final.csv")  
summary(datos\_final)

## num titulo\_empleo descripcion\_empleo etiquetas\_empleo   
## Min. : 0 Length:8556 Length:8556 Length:8556   
## 1st Qu.:1041 Class :character Class :character Class :character   
## Median :3172 Mode :character Mode :character Mode :character   
## Mean :3308   
## 3rd Qu.:5310   
## Max. :7449

## Filtramos datos por modalidad

Para realizar la filtración de modalidad, nos basamos en la descripción del empleo en el cual se obtiene información acerca del tipo de trabajo que se ofrece, en este caso nos importa la modalidad, para esto usamos la función str\_detect para poder usar su subfunción “regex” con el fin de poder escribir una expresión regular, y además colocamos el valor de TRUE en ignore\_case para que no importe si se encuentra escrita entre las mayúsculas o minúsculas. Hay que tener en cuenta que habrán descripciones en las que no se mencione la modalidad. Pero con estos datos se va a trabajar para responder algunas preguntas.

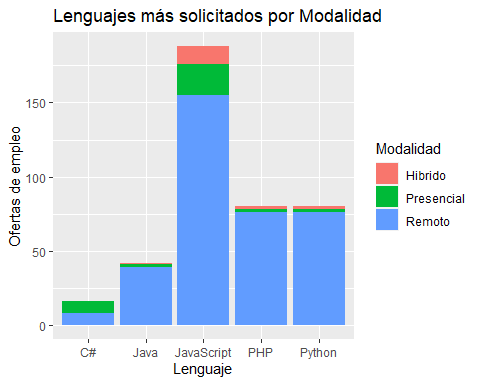
#PRESENCIAL  
descripcion <- datos\_final$descripcion\_empleo  
presencial <- str\_detect(descripcion, regex("presencial|(\\s+|,|^)on site(\\s+|,|$)", ignore\_case = TRUE))  
desc\_presencial <- descripcion[presencial]  
  
#REMOTO  
remota <- str\_detect(descripcion, regex("remote|remota|remoto", ignore\_case = TRUE))  
desc\_remota <- descripcion[remota]  
  
#HÍBRIDO  
hibrido <- str\_detect(descripcion, regex("híbrido|híbrida|hybrid|hibrido|hibrida|semipresencial", ignore\_case = TRUE))  
desc\_hibrido <- descripcion[hibrido]

## Preguntas Alan

### Pregunta 1 ¿Qué lenguajes de programación son los más solicitados por modalidad de trabajo?

#PRESENCIAL  
#PHP  
php\_pres <- str\_detect(desc\_presencial, regex("php|laravel", ignore\_case = TRUE))  
  
#Javascript  
javascript\_pres <- str\_detect(desc\_presencial, regex("javascript|js|react", ignore\_case = TRUE))  
  
  
  
#python  
python\_pres <- str\_detect(desc\_presencial, regex("python|django", ignore\_case = TRUE))  
  
#c#  
csharp\_pres <- str\_detect(desc\_presencial, regex("c#", ignore\_case = TRUE))  
  
  
#java  
java\_pres <- str\_detect(desc\_presencial,"(\\s+|,|^)[Jj]ava(\\s+|,|$)")  
  
  
  
lenguajes\_pres <- c("PHP","JavaScript","Python","C#","Java")  
valores\_pres <- c(sum(python\_pres == TRUE),sum(javascript\_pres == TRUE),sum(python\_pres == TRUE),  
 sum(csharp\_pres==TRUE),sum(java\_pres==TRUE))  
  
############DATA FRAME PRESENCIAL ##################  
df\_pres <- data.frame("lenguajes"=lenguajes\_pres,"valores"=valores\_pres)  
df\_pres <- cbind(df\_pres,Modalidad="Presencial")  
  
  
#REMOTA ######################  
#PHP  
php\_remo <- str\_detect(desc\_remota, regex("php|laravel", ignore\_case = TRUE))  
  
#Javascript  
javascript\_remo <- str\_detect(desc\_remota, regex("javascript|js|react", ignore\_case = TRUE))  
  
  
#python  
python\_remo <- str\_detect(desc\_remota, regex("python|django", ignore\_case = TRUE))  
  
#c#  
csharp\_remo <- str\_detect(desc\_remota, regex("c#", ignore\_case = TRUE))  
  
  
#java  
java\_remo <- str\_detect(desc\_remota,"(\\s+|,|^)[Jj]ava(\\s+|,|$)")  
  
######DATA FRAME ##################  
lenguajes\_remo <- c("PHP","JavaScript","Python","C#","Java")  
valores\_remo <- c(sum(python\_remo == TRUE),sum(javascript\_remo == TRUE),sum(python\_remo == TRUE),  
 sum(csharp\_remo==TRUE),sum(java\_remo==TRUE))  
  
df\_remo <- data.frame("lenguajes" = lenguajes\_remo,"valores"=valores\_remo)  
df\_remo <- cbind(df\_remo,Modalidad="Remoto")  
  
##HIBRIDO  
#PHP  
php\_hibrid <- str\_detect(desc\_hibrido, regex("php|laravel", ignore\_case = TRUE))  
  
#Javascript  
javascript\_hibrid <- str\_detect(desc\_hibrido, regex("javascript|js|react", ignore\_case = TRUE))  
  
#python  
python\_hibrid <- str\_detect(desc\_hibrido, regex("python|django", ignore\_case = TRUE))  
  
#c#  
csharp\_hibrid <- str\_detect(desc\_hibrido, regex("c#", ignore\_case = TRUE))  
  
  
#java  
java\_hibrid <- str\_detect(desc\_hibrido,"(\\s+|,|^)[Jj]ava(\\s+|,|$)")  
  
  
  
######### DATA FRAME #########  
lenguajes\_hibrid <- c("PHP","JavaScript","Python","C#","Java")  
valores\_hibrid <- c(sum(python\_hibrid == TRUE),sum(javascript\_hibrid == TRUE),sum(python\_hibrid == TRUE),  
 sum(csharp\_hibrid==TRUE),sum(java\_hibrid==TRUE))  
  
df\_hibrid <- data.frame("lenguajes"=lenguajes\_hibrid,"valores"=valores\_hibrid)  
df\_hibrid <- cbind(df\_hibrid,Modalidad="Hibrido")  
  
  
###### MERGE DE LAS 3 DATA FRAMES ###################  
df\_final <- merge(x=df\_pres,y=df\_remo, by = "lenguajes")  
  
df\_temp <- rbind(df\_pres, df\_remo)  
df\_final <- rbind(df\_temp,df\_hibrid)

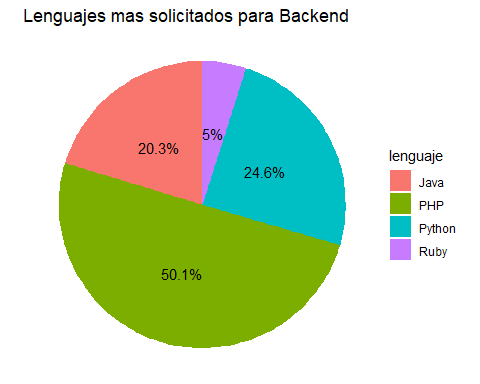
###GRAFICA####  
ggplot(df\_final, aes(x=lenguajes, y=valores,fill=Modalidad))+  
 geom\_bar(stat = "identity")+  
 labs(title="Lenguajes más solicitados por Modalidad",  
 x ="Lenguaje", y = "Ofertas de empleo")



## Pregunta 2 ¿Qué lenguajes de back-end son los más solicitados?

#Lenguajes backend  
library(dplyr)  
#PHP  
php <- str\_detect(descripcion, regex("php|laravel", ignore\_case = TRUE))  
phpCant<-sum(php == TRUE)  
#Java  
java <- str\_detect(descripcion,"(\\s+|,|^)[Jj]ava(\\s+|,|$)")  
javaCant<-sum(java==TRUE)  
#Ruby  
ruby <- str\_detect(descripcion, regex("ruby", ignore\_case = TRUE))  
rubyCant<-sum(ruby == TRUE)  
#Python  
python <- str\_detect(descripcion, regex("python|django", ignore\_case = TRUE))  
pythonCant<-sum(python == TRUE)

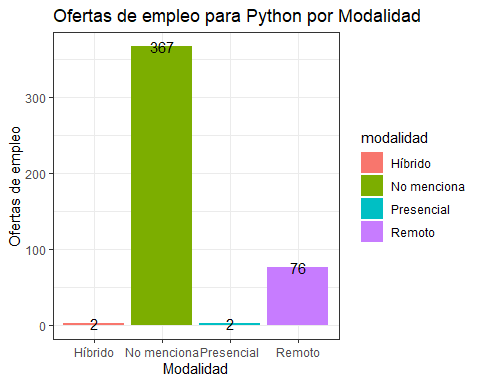
valores\_leng <-c(javaCant,phpCant,pythonCant,rubyCant)  
datospr2 <- data.frame("lenguaje" = c("Java","PHP","Python","Ruby"),  
 "valor" = valores\_leng)  
  
#Porcentaje  
porcentaje <- datospr2 %>%   
 group\_by(lenguaje) %>%   
 count() %>%   
 ungroup() %>%   
 mutate(percentage=valores\_leng/sum(valores\_leng) \* 100)   
  
##Grafica de pastel ######  
  
ggplot(porcentaje, aes(x="", y=percentage,fill=lenguaje))+  
 geom\_bar(stat="identity", width=1) +  
 geom\_text(aes(label = paste0(round(percentage,1),"%")),   
 position = position\_stack(vjust = 0.5)) +  
 coord\_polar(theta = "y") +   
 labs(title= "Lenguajes mas solicitados para Backend")+  
 theme\_void()



## Pregunta 3 ¿Para qué modalidad se solicita más el lenguaje Python?

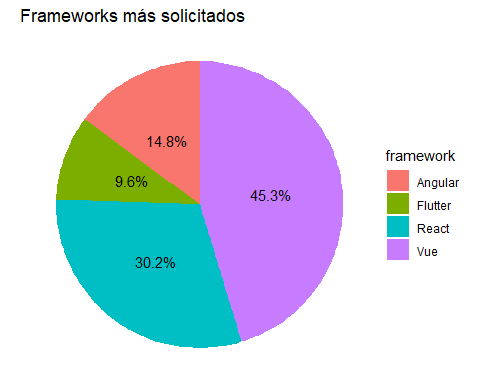
python\_presCant <- sum(python\_pres==TRUE)  
python\_remoCant <- sum(python\_remo==TRUE)  
python\_hibridCant <- sum(python\_hibrid==TRUE)  
python\_mods <- python\_presCant+python\_remoCant+python\_hibridCant  
  
#no menciona modalidad##  
python\_noMenc <- pythonCant - python\_mods  
  
python\_final <- data.frame("modalidad" = c("Presencial", "Remoto","Híbrido", "No menciona"),  
 "valor" = c(python\_presCant,python\_remoCant,python\_hibridCant, python\_noMenc))

ggplot(python\_final, aes(x=modalidad, y=valor, fill=modalidad)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = valor))+  
 theme\_bw()+  
 labs(title= "Ofertas de empleo para Python por Modalidad",  
 x="Modalidad", y="Ofertas de empleo")

 ## Pregunta 4 ¿Qué frameworks son los más solicitados?

#Lenguajes backend  
library(dplyr)  
#Vue  
vue <- str\_detect(descripcion, regex("vue|vue.js", ignore\_case = TRUE))  
vueCant<-sum(php == TRUE)  
#Angular  
angular <- str\_detect(descripcion,regex("angular|angular.js", ignore\_case= TRUE))  
angularCant<-sum(angular==TRUE)  
#Flutter  
flutter <- str\_detect(descripcion, regex("flutter", ignore\_case = TRUE))  
flutterCant<-sum(flutter == TRUE)  
  
#React  
react <- str\_detect(descripcion, regex("react", ignore\_case = TRUE))  
reactCant<-sum(react == TRUE)

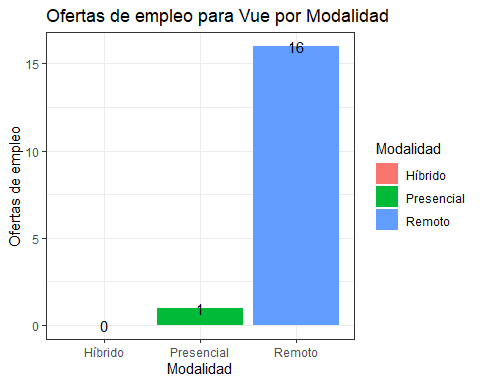
valores\_frame <- c(angularCant,flutterCant,reactCant,vueCant)  
datospr3 <- data.frame("framework" = c("Vue","Angular","Flutter","React"),  
 "valor" = valores\_frame)  
#Porcentaje  
porcentaje\_frame <- datospr3 %>%  
 group\_by(framework) %>%  
 count() %>%  
 ungroup() %>%  
 mutate(percentage=valores\_frame/sum(valores\_frame) \* 100)  
##Grafica de pastel ######  
  
ggplot(porcentaje\_frame, aes(x="", y=percentage,fill=framework))+  
 geom\_bar(stat="identity", width=1) +  
 geom\_text(aes(label = paste0(round(percentage,1),"%")),  
 position = position\_stack(vjust = 0.5)) +  
 coord\_polar(theta = "y") +  
 labs(title= "Frameworks más solicitados")+  
 theme\_void()



## Pregunta 5 ¿En qué modalidad es más solicitado el framework Vue?

vue\_pres <- str\_detect(desc\_presencial, regex("vue", ignore\_case = TRUE))  
  
vue\_remo <-str\_detect(desc\_remota, regex("vue", ignore\_case = TRUE))  
  
vue\_hibrid <-str\_detect(desc\_hibrido, regex("vue", ignore\_case = TRUE))  
  
  
vue\_noMenc <- str\_detect(descripcion, regex("vue", ignore\_case = TRUE))  
  
  
  
df\_vue <- data.frame("Modalidad"= c("Presencial","Remoto","Híbrido"),  
 "Valores" = c(sum(vue\_pres==TRUE),sum(vue\_remo==TRUE),sum(vue\_hibrid==TRUE)))

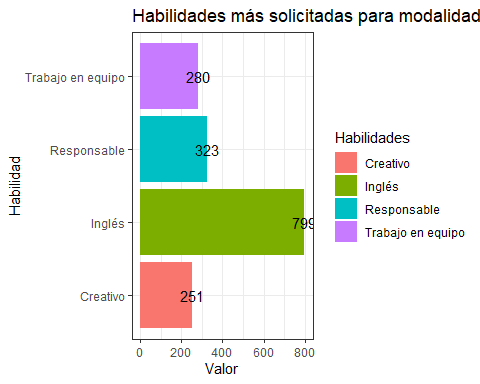
ggplot(df\_vue, aes(x=Modalidad, y=Valores,fill=Modalidad)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = Valores))+  
 theme\_bw()+  
 labs(title= "Ofertas de empleo para Vue por Modalidad",  
 x="Modalidad", y="Ofertas de empleo")



## Pregunta 6 ¿Cuáles son las habilidades más solicitadas para la modalidad presencial?

responsabilidad <- str\_detect(descripcion,regex("responsabilidad|responsability|responsable", ignore\_case = TRUE))  
creativo <- str\_detect(descripcion,regex("creatividad|creativo|creative|creativa", ignore\_case = TRUE))  
ingles <- str\_detect(descripcion,regex("inglés|ingles|english", ignore\_case = TRUE))  
equipo <- str\_detect(descripcion,regex("(\\s+|,|^)([Tt]eamwork|[Tt]rabajo en equipo|[Cc]olaboraci[oó]n)(\\s+|,|$)"))  
df\_habilidades <- data.frame("Habilidades" = c("Responsable", "Creativo","Inglés","Trabajo en equipo"),  
 "valor" = c(sum(responsabilidad == TRUE),sum(creativo == TRUE),sum(ingles == TRUE),  
 sum(equipo == TRUE)))

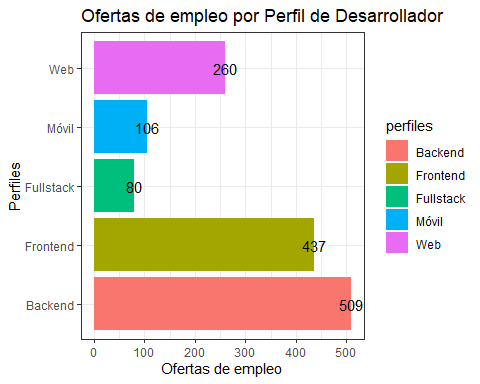
ggplot(df\_habilidades, aes(x=Habilidades, y=valor,fill=Habilidades)) +  
 geom\_bar(stat="identity")+  
 geom\_text(aes(label = valor))+  
 theme\_bw()+  
 labs(title= "Habilidades más solicitadas para modalidad presencial",  
 x="Habilidad", y="Valor")+  
 coord\_flip()



## Pregunta 7 ¿Perfiles más buscados de desarrolladores?

#Desarrollador Frontend  
frontend <- str\_detect(descripcion,  
regex("(\\s+|,|^)((([Dd]esarrollador(es)\\s\*)?[Ff]ront-?end)|([Ff]ront-?end)\\s\*([Dd]evelopers?))(\\s+|,|$)",  
ignore\_case = TRUE))  
  
#Desarrollador Backend  
backend <- str\_detect(descripcion,  
regex("(\\s+|,|^)((([Dd]esarrollador(es)\\s\*)?[Bb]ack-?end)|([Bb]ack-?end)\\s\*([Dd]evelopers?))(\\s+|,|$)",  
ignore\_case = TRUE))  
  
#Desarrollador FullStack  
fullstack <- str\_detect(descripcion,  
regex("(\\s+|,|^)((([Dd]esarrollador(es)\\s\*)?[Ff]ullstack)|([Ff]ullstack)\\s\*([Dd]evelopers?))(\\s+|,|$)",  
ignore\_case = TRUE))  
  
#Desarrollador web  
web <- str\_detect(descripcion,  
regex("(\\s+|,|^)((([Dd]esarrollador(es))\\s\*[Ww]eb)|([Ww]eb)\\s\*([Dd]evelopers?))(\\s+|,|$)",  
ignore\_case = TRUE))  
  
#Desarrollador movil  
movil <- str\_detect(descripcion,  
regex("(\\s+|,|^)[Dd]esarroll(o|ador(es)?)\\s\*[Dd]e\\s\*[Aa]p(ps?|licaciones)\\s\*([Mm][óo]vil(es)?)?(\\s+|,|$)",  
ignore\_case = TRUE))

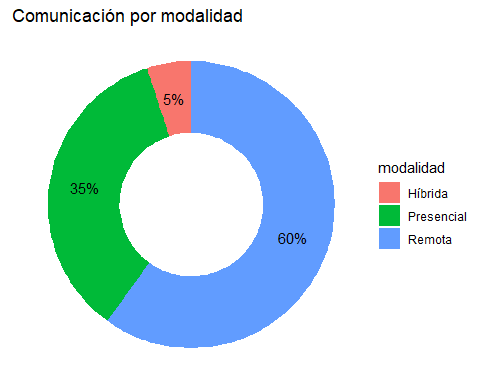
## Diagrama de barras ######  
pregunta\_7 <- data.frame("perfiles" = c("Frontend", "Backend","Fullstack", "Web", "Móvil"),  
 "valor" = c(sum(frontend==TRUE),sum(backend==TRUE),sum(fullstack==TRUE), sum(web==TRUE),sum(movil==TRUE)))  
  
ggplot(pregunta\_7, aes(x=valor, y=perfiles, fill=perfiles)) +  
 geom\_bar(stat="identity", orientation = "y")+  
 geom\_text(aes(label = valor))+  
 theme\_bw()+  
 labs(title= "Ofertas de empleo por Perfil de Desarrollador",  
 x="Ofertas de empleo", y="Perfiles")



## Pregunta 8 ¿La habilidad “comunicación” es más requerida en modalidades remotas e híbrida que en modalidades presenciales?

com\_pres <- str\_detect(desc\_presencial, regex("(\\s+|,|^)([Cc]omunicaci[oó]n|[Cc]omunication)(\\s+|,|$)", ignore\_case = TRUE))  
  
com\_rem <- str\_detect(desc\_remota, regex("(\\s+|,|^)([Cc]omunicaci[oó]n|[Cc]omunication)(\\s+|,|$)", ignore\_case = TRUE))  
  
com\_hib <- str\_detect(desc\_hibrido, regex("(\\s+|,|^)([Cc]omunicaci[oó]n|[Cc]omunication)(\\s+|,|$)", ignore\_case = TRUE))

## Gráfica de dona ######  
valores<- c(sum(com\_hib == TRUE),sum(com\_pres == TRUE),sum(com\_rem == TRUE))  
datospr8 <- data.frame("modalidad" = c("Híbrida","Presencial","Remota"),  
 "valor" = valores)  
porcentaje <- datospr8 %>%  
 group\_by(modalidad) %>%  
 count() %>%  
 ungroup() %>%  
 mutate(percentage=valores/sum(valores) \* 100)  
  
ggplot(porcentaje,aes(x=2,y=percentage, fill=modalidad))+  
 geom\_bar(stat = "identity", width=1)+  
 geom\_text(aes(label = paste0(round(percentage,1),"%")),  
 position = position\_stack(vjust = 0.5))+  
 coord\_polar(theta = "y")+  
 labs(title="Comunicación por modalidad")+  
 theme\_void()+  
 xlim(0.5,2.5)



## Pregunta 9 ¿Saber inglés es importante para desenvolverse en el ámbito del desarrollo de software?

## Menciones de inglés  
ingles <- str\_detect(descripcion, regex("(\\s+|,|^)([Ii]ngl[eé]s|[Ee]nglish)(\\s+|,|$)", ignore\_case = TRUE))  
ingles <- sum(ingles == TRUE)  
no\_ingles = (length(descripcion) - sum(ingles == TRUE))  
#no\_ingles  
  
## Gráfica de pastel ######  
valores<- c(ingles, no\_ingles)  
datospr9 <- data.frame("label" = c("Inglés","No inglés"),  
 "valor" = valores)  
porcentaje <- datospr9 %>%  
 group\_by(label) %>%  
 count() %>%  
 ungroup() %>%  
 mutate(percentage=valores/sum(valores) \* 100)  
  
ggplot(porcentaje, aes(x="", y=percentage, fill=label))+  
 geom\_bar(stat="identity", width=1) +  
 geom\_text(aes(label = paste0(round(percentage,1),"%")),  
 position = position\_stack(vjust = 0.5)) +  
 coord\_polar(theta = "y") +  
 labs(title= "Inglés en el ámbito del desarrollo de software")+  
 theme\_void()

