

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("C:\\Users\\Alan\\Downloads\\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]

```

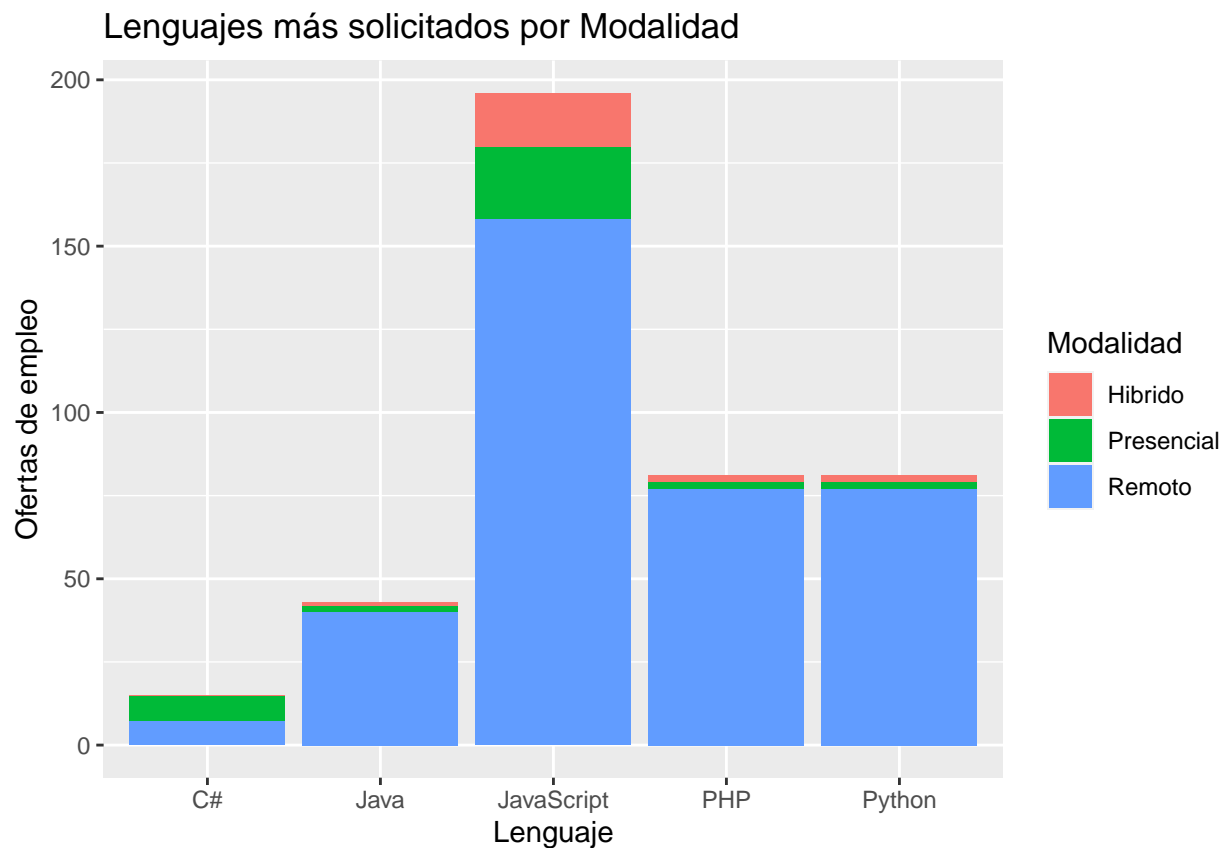
Preguntas Alan

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

```

###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")

```



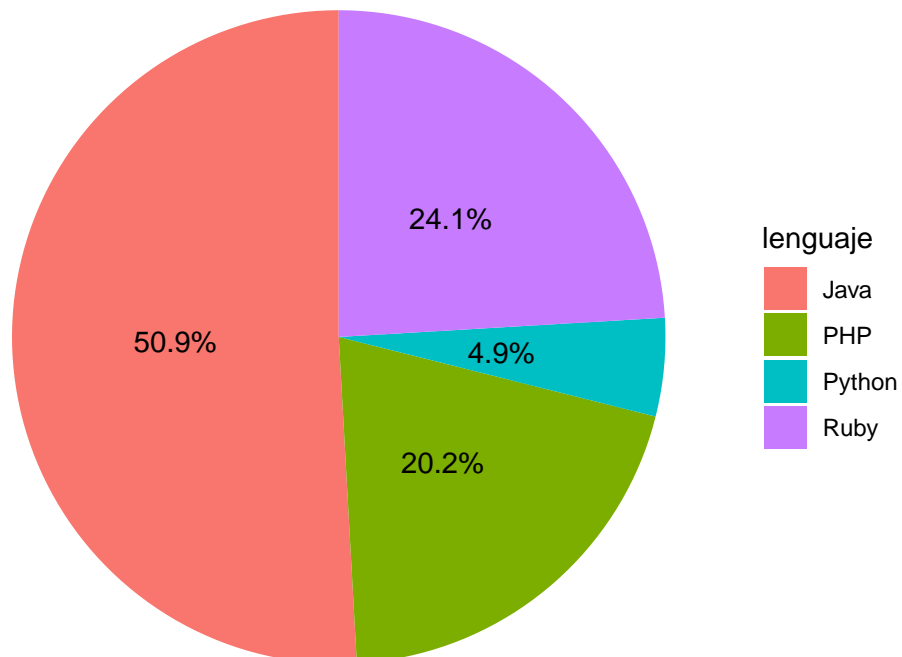
Pregunta 2 ¿Qué lenguajes de back-end son los más solicitados?

```
valor_leng<- c/phpCant,javaCant,rubyCant,pythonCant)
datospr2 <- data.frame("lenguaje" = c("PHP","Java","Ruby","Python"),
                        "valor" = valor_leng)

#Porcentaje
porcentaje <- datospr2 %>%
  group_by(lenguaje) %>%
  count() %>%
  ungroup() %>%
  mutate(percentage=valor_leng/sum(valor_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()
```

Lenguajes mas solicitados para Backend



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")
```

