



UNIVERSIDAD TÉCNICA DE AMBATO

Facultad de Ingeniería Civil y Mecánica

Carrera: Ingeniería Mecánica



Alumno: Dennis Mejía

Semestre: Tercero

Paralelo: "A"

Materia: Estadística

Fecha: 23 de Octubre del 2017

Tema: Deber sobre los verbos de dplyr y tidyr y como unir tablas

library(tidyverse)

library(tidyr)

library(dplyr)

Deber de estadística N°1

Ejercicio N° 1

#Pregunta 1.- ¿Cuáles de las siguientes expresiones valen 99 para $x = 10$ en R? Analicen la sintaxis como si estuvieran programando

$x=10$

$10x - 1$

$(x)(x) - 1$

$\text{abs}(x*x) - \text{abs}(9-x)$

$11 * x - x + 1$

#la función 1 y 2 no se ejecutan ya que no contienen símbolos *.La respuesta es la tercera

#Pregunta 2.-Un vector contiene una serie de ganancias ordenadas de manera creciente. Escriban el código que genera:

$a \leftarrow c(20, 100, 360, 700, 980)$

#La suma de todas las ganancias.

$\text{sum}(a)$

#La segunda ganancia mas grande

$x \leftarrow \text{length}(a)$

$a[x-1]$

#La diferencia mas grande entre las ganancias

$d \leftarrow \text{diff}(a)$

$\text{max}(d)$

#¿Un booleano que responda a la pregunta: La mas grande diferencia ente dos ganancias es mayor a 10?

```
v <- max(d)
```

```
nchar(v) < 10
```

#Respuesta es verdadera porque se cuentan los dígitos de la más grande diferencia que en este caso es 680 y por eso son 3 dígitos y así $3 < 10$ es verdadero.

#La menor diferencia positiva entre dos ganancias.

```
d <- diff(a)
```

```
min(d)
```

#El máximo número de ganancias que pueden sumar sin pasar de 10000.

```
cumsum(a)<10000
```

Ejercicio 2: Dplyr en los Aeorpuertos

```
library(nycflights13)
```

```
View(planes)
```

```
View(flights)
```

```
View(weather)
```

#Pregunta 3.-Instalen la librería nycflights13. Escriban el código para encontrar todos los vuelos que:

Fueron de SFO(San Francisco) hasta OAK(Oakland).

```
vuelos <- flights %>%  
  select(origin, dest)
```

```
vuelosV2 <- mutate(vuelos,  
                    dest = factor(origin, labels = c('OAK')),  
                    origin = factor(dest, labels = c('SFO')))  
View(vuelosV2)
```

#viajes en Enero

```
vuelos_enero <- flights %>%  
  select(month)
```

```
vuelos_eneroV2 <- mutate(vuelos_enero,  
                          month = factor(1))  
View(vuelos_eneroV2)
```

#Tienen demoras de más de una hora (las demoras están en minutos).

```
flights %>%  
  select(year, month, day, dep_delay) %>%  
  filter(dep_delay > 60)
```

#Salieron entre medianoche y las 5 a.m.

```
vuelos_madrugadaV2 <- filter(vuelos_madrugada, hour <= 5)
View(vuelos_madrugadaV2)
```

No existen vuelos en el rango de 1 a 5 de la mañana únicamente hay vuelos antes de la 1 y después de las 5.

#Tuvieron una demora de llegada 2 veces más grande que la de salida.

```
flights %>% select(retraso_salida = dep_time, retraso_llegada = arr_time)
%>%
  mutate(diferencia_retrasos = (retraso_llegada - retraso_salida)) %>%
  filter(retraso_salida == diferencia_retrasos) %>% View()
```

#para observar los casos que la demora de llegada es 2 veces mayor a la de salida hay que realizar las diferencias entre los retrasos de llegada y salida de cada vuelo.

#Pregunta 4.- Lean la ayuda de select().Escriban 2 formas de seleccionar las dos variables de retraso.

```
select(tiempo,arr_delay,dep_delay)
```

#Pregunta 5.- Ordenen la tabla por fecha de salida y tiempo. Cuáles fueron los vuelos que sufrieron las mayores demoras? Cuáles recuperaron la mayor cantidad de tiempo durante el vuelo?

```
flights%>% arrange(desc(year,month,day,hour))
```

Pregunta 6.- Calculen la velocidad en mph usando el tiempo (que está en minutos) y la distancia (que está en millas). Cuál fué el avión que voló mas rápido?

```

velocidad <- flights %>% select(air_time, distance) %>%
  mutate(tiempo.horas = (air_time/60),
         distancia.metros = distance*1609.34)
velocidad %>% filter(tiempo.horas, distancia.metros) %>%
  select(tiempo.horas, distancia.metros) %>%
  mutate(velocidad.mph = (distancia.metros / tiempo.horas)) %>% View()

```

#Pregunta 7.- En dplyr el comando pipeline %>% se lee entonces.
Significa:

```
flights %>% filter (! is.na(dep_delay)) %>% View()
```

```
flights %>% group_by(date, hour) %>% View()
```

```
flights %>% summarise(delay = mean(dep_delay), n = n()) %>% View()
```

```
flights %>% filter(n > 10) %>% View()
```

El código dos no corre debido a que se desconoce la columna de “date” en la tabla “flights”. - El código cuatro no corre debido a que la columna “n” no existe.

#Pregunta 8.-Cuál es la destinación que tiene las demoras promedio mas grandes?.¿ Cuántos vuelos diarios hay?.¿Cuál es la mejor hora para viajar sin retraso?

1

```
flights %>% group_by(origin, dest) %>%  
  Summarise (demoras = mean(arr_delay, na.rm=TRUE)) %>% View()
```

2

```
flights %>% group_by(day) %>%  
  summarise(vuelos = sum(day == '1')) %>% View()  
flights %>% group_by(day) %>%  
  summarise(vuelos = sum(day == '2')) %>% View()
```

#3

#Ejercicio 3.- 6,000 Años de Urbanización Global.

#Pregunta 9.- Cuántas personas han habitado Sur América? Hay alguna forma de usar los datos para saber cuál es la región que tiene las poblaciones más antiguas? Pueden usar los datos para tratar de entender si hay un patrón migratorio a lo largo de la historia?