

## ESTADÍSTICA

Nombre: Daniel Coquinche

Semestre: 3ro "A"

Fecha: 23/10/2017

Tema: R Studio.

### Ejercicio 1

Un poco de R.

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

#Pregunta 1

$x=10$

$10*x-1$

$x*x-1$

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

$11*x-x+1$  #no es correcto.

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

#Pregunta 2

`ganancias<-c(350,560,870,1022,2305,5006,6560)`

# La suma de todas las ganancias.

`sum(ganancias)`

#La segunda ganancia mas grande.

`a<-max(ganancias)`

`b<-max(diff(ganancias))`

`c<-(a-b)`

`c`

#La diferencia mas grande entre las ganancias.

`max(diff(ganancias))`

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

`gan1<-c(5000,5011)`

`gan2<-c(709,700)`

`max_dif1<-max(diff(gan1))`

`max_dif2<-max(diff(gan2))`

`max_dif1>10`

`max_dif2>10`

#La menor diferencia positiva entre dos ganancias.

`min(diff(ganancias))`

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

```
maxi_num<-max(sum(ganancias<=10000))
maxi_num
```

## Ejercicio 2

Dplyr en los Aeorpuertos.

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

```
#pregunta 03
library(dplyr)
library(tidyr)
```

```
library(nycflights13)
View(weather)
View(flights)
View(planes)
#-----
#Fueron de SFK(USA) hasta SFO(San Francisco).
SF_OK <- flights %>% filter(origin == 'JFK', dest == 'SFO') %>% View()
#Salieron en Enero.
ENERO_SALIDA<-flights %>% filter(month=='1') %>% View()
#Tienen demoras de mas de una hora (las demoras están en minutos).
DEMORAS<-flights %>% filter(dep_delay>=1,arr_delay>=1) %>% View()
#Salieron entre medianoche y las 5 a.m.
SALIDAS<-flights %>% filter(hour>=0,hour<=5) %>% View()
#Tuvieron una demora de llegada 2 veces mas grande que la de salida.
doble_vez<-flights %>% filter(arr_delay==2*dep_delay) %>% View()
```

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

```
#pregunta 04
#Lean la ayuda de select(). Escriban 2 formas de seleccionar las dos variables de retraso.
```

```
retrazos<- select(flights,c(dep_delay,arr_delay))
View(retrazos)
```

- 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(dep_time, year, month, day) %>%
View()
```

- 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?

```
flights %>% filter(air_time, distance) %>%
```

```
mutate(Velocidad=air_time*distance, air_time=air_time/60) %>%
```

```
View()
```

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

Significa:

$x \%>\% f(y) \rightarrow f(x; y):$

Es decir pasa x como primer argumento de f.

Qué significan las siguientes líneas de código:

```
flights%>% filter(! is:na(dep_delay))# Filtra los valores de elementos en los
retrasos de salida que se están perdiendo.
flights%>% group_by(date, hour)#Indica los valores agrupados entre la fecha y la
hora.
flights%>% summarise(delay = mean(dep_delay), n = n())#Hace un resumen del
promedio de los retrasos.
flights%>% filter(n > 10)#Error in n > 10 : comparison (6) is possible only for
atomic and list types
```

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

```
flights %>% group_by(dest, arr_delay) %>%
```

```
summarize(mean(arr_delay, na.rm=TRUE)) %>%
```

```
View()
```

### EJERCICIO 3

- 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 mas antiguas? Pueden usar los datos para tratar de entender si hay un patrón migratorio a lo largo de la historia?

```
TODO=alldata
```

```
View(TODO)
```

```
Países=countries_of_the_world
```

```
View(Países)
```

```
TODO$'pop'
```

```
PoblacionTodo=TODO %>% select(Poblacion='pop') %>%
```

```
mutate(Poblacion=gsub(".*$", "",Poblacion)) %>%
```

```
group_by(Poblacion) %>%
```

```
summarize(PoblacionTotal=n())
```

```
View(PoblacionTodo)
```

```
PoblacionPaises=Países %>% select(Poblacion=`X__2`,
Continente=`X__1`) %>%
filter(Continente=="LATIN AMER. & CARIB") %>%
group_by(Poblacion)
View(PoblacionPaises)
```

```
dim(PoblacionPaises)
[1] 45 2
dim(PoblacionTodo)
[1] 2460 2
```

```
PaísesTodo=inner_join(PoblacionPaises, PoblacionTodo, by="Poblacion")
View(PaísesTodo)
```

La tabla sale error porque no hay datos válidos para SUDAMERICA.

Podemos calcularlo pero a mi parecer ninguna de las 2 tablas posee datos de antigüedad.

```
Países$X__6
```

```
PatronMigratorio=Países %>% select(Migracion=`X__6`) %>%
mutate(Migracion=gsub(".*$", "", Migracion)) %>%
group_by(Migracion) %>%
summarise(mean(Migracion, na.rm=TRUE)) %>%
View()
```

Tenemos la lista de datos de migración pero no pude calcular el patron migratori o ya que el pasar de los años aumenta o disminuye dependiendo de la época.

Hecho por: Cano Kevin