



Identificación: Christian M. Lisintuña P.

Deber N.- 001

Nivel: Tercero Paralelo: "B"

Fecha de Entrega: 23 de octubre de 2017

Tema: Introducción a Rstudio

## 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.

$x = 10$  # si  $x$  vale 10 tenemos

$10 * x - 1$  # si cumple la condición

$x * x - 1$  # si cumple la condición

$\text{abs}(x * x) - \text{abs}(9 - x)$  # si cumple la condición

$11 * x - x + 1$  # No cumple la condición

### Pregunta 2

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

\*La suma de todas las ganancias

```
x <- c(10,20,30,40,50)
```

```
sum(x)
```

```
> sum(x)  
[1] 150
```

\*La segunda ganancia más grande.

UNIVERSIDAD TÉCNICA DE MABATO  
FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA

\*La diferencia más grande entre las ganancias .

```
x <- c(10,20,30,40,50)
```

```
max(x)  
[1] 50
```

\*Un booleano que responda a la pregunta: La más grande diferencia Entre dos ganancias es mayor a 10?

```
x <- c(10,20,30,40,50)  
max(x>10)  
[1] 1
```

\* La menor diferencia positiva entre dos ganancias

```
x <- c(10,20,30,40,50)  
min(x)  
[1] 10
```

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

## **Ejercicio 2 Dplyr en los Aeorpuertos.**

Vamos a estudiar los datos de los vuelos locales en Estados Unidos durante el 2011.  
Usen los verbos:

Vamos a estudiar los datos de los vuelos locales en Estados Unidos durante el 2011. Usen los verbos:

select()

filter()

mutate()

UNIVERSIDAD TÉCNICA DE MABATO  
FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA

arrange()

group\_by()

summarise()

para manipular 3 data frames: Uno con los vuelos (flights), uno con los aviones (planes) y otro con el clima (weather).

	City	OtherName	Country	Latitude	Longitude	Certainty	year	pop
1	Eridu	NA	Iraq	30.815837900	45.99607	1	-3700	6000
2	Eridu	NA	Iraq	30.815837900	45.99607	1	-3500	10000
3	Larak	Kesh	Iraq	32.312737510	45.66100	1	-3500	10000
4	Uruk	Erech	Iraq	31.322222000	45.63611	1	-3500	14000
5	Anshan	NA	Iran	30.012019580	52.40858	1	-3300	10000
6	Eridu	NA	Iraq	30.815837900	45.99607	1	-3300	10000
7	Larak	Kesh	Iraq	32.312737510	45.66100	1	-3300	10000
8	Uruk	Erech	Iraq	31.322222000	45.63611	1	-3300	40000
9	Anshan	NA	Iran	30.012019580	52.40858	1	-3000	10000
10	Eridu	NA	Iraq	30.815837900	45.99607	1	-3000	10000
11	Eridu	NA	Iraq	30.815837900	45.99607	1	-3000	10000

### Pregunta 3

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

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

\*Salieron en Enero.

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

UNIVERSIDAD TÉCNICA DE MABATO  
FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA

\*Salieron entre medianoche y las 5 a.m.

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

#### **Pregunta 4**

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

#### **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?

#### **Pregunta 6**

Calculen la velocidad en mph usando el tiempo (que está en minutos) y la distancia (que está en millas). Cuál fue el avión que voló más rápido?

#### **Pregunta 7**

En dplyr el comando `pipeline %>%` se lee entonces.

Significa:  $x \%>\% f(y) \longrightarrow f(x, y)$ .

Es decir pasa x como primer argumento de f. Qué significan las siguientes líneas de código:

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

UNIVERSIDAD TÉCNICA DE MABATO  
FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA

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

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

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

### **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?

### **Ejercicio 3**

6,000 Años de Urbanización Global.

En el artículo Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000 los autores reúnen una serie de datos sobre la población de las ciudades del mundo desde 3700 A.C. hasta el 2000 D.C. Usemos estos datos para tratar de entender algunas cosas sobre la distribución espacial de las personas a través del tiempo. Para hacer esto vamos a tener que manipular dos data frames

- uno para la población y otro con los nombres de los países y continentes - usando dplyr para eventualmente unirlos. En este ejercicio vamos a usar algunos verbos para unir data frames.

\* inner\_join(d1, d2) contiene solo las filas de d1 comunes ad2.

\* left\_join(d1, d2) contiene todas las filas de d1 y NA en las filas de d2 que no están en d1.

\* semi\_join(d1, d2) no añade columnas a d1. Contiene las filas de d1 comunes a d2.

UNIVERSIDAD TÉCNICA DE MABATO  
FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA

**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?