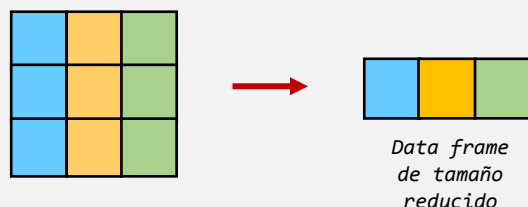




Manipulación de datos. Filtrado por renglones

Paquetería "nycflights13"

Cuando se tienen bases de datos muy grandes, es muy común plantearnos preguntas sobre los datos para obtener una cantidad reducida de los mismos.



La paquetería "nycflights13" contiene información sobre todos los vuelos que salieron de Nueva York en el año 2013; debe ser instalada previamente:

#Ejecutar el siguiente commando. Nombrar la variable "p"

```
p <- row.names(installed.packages())
```

#Ejecutar la condicional para verificar la instalación

```
if ((is.na(paq[grep("nycflights13", p)][1]) == TRUE)) {  
  install.packages("nycflights13")  
} else {  
  "Paquetería instalada"  
}
```

Una vez instalada la paquetería, la base de datos a utilizar es llamada *flights*, así que solo bastará ejecutar lo siguiente:

#Ejemplar la función View para visualizar la base de datos

```
View(flights)
```

Filtrado por renglones

La base de datos *flights* tiene 336 776 renglones y 19 columnas, cada una representa el año, hora de llegada, origen, destino, etc.

Columna 1	Columna 2	Columna 3	...	Columna 13	Columna 14	...
year	month	day	...	origin	dest	...
2013	1	1	...	EWB	IAH	...
2013	1	1	...	LGA	IAH	...
2013	1	1	...	JFK	MIA	...
2013
2013	12	2	...	LGA	LEX	...
2013	12	20	...	LGA	ORD	...
2013	12	23	...	JFK	MCO	...

Pregunta: ¿Cuántos vuelos hubo hacia México (MCO) dos días antes del 25 de diciembre?

#Condición 1: vuelos con destino a México (MCO)

```
flights$dest == "MCO" #Vector tipo lógico
```

Vector 1

T	F	T	T	T	T	T	F
---	---	---	---	---	---	---	---

#Condición 2: vuelos del mes de diciembre (12)

```
flights$month == 12 #Vector tipo lógico
```

Vector 2

F	F	T	F	T	F	T	F
---	---	---	---	---	---	---	---

#Condición 3: vuelos del día 23

```
flights$day == 23 #Vector tipo lógico
```

Vector 3

T	T	T	F	F	T	T	F
---	---	---	---	---	---	---	---

Combinación de condiciones

Cuando se tienen las condiciones, las juntamos; en este caso se realizará con el operador tipo "AND" (&).

#Unir condición 1 y 2 para generar un vector 4

```
flights$dest == "MCO" & flights$month == 12
```

Vector 1

T	F	T	T	T	T	T	F
---	---	---	---	---	---	---	---

Vector 2

F	F	T	F	T	F	T	F
---	---	---	---	---	---	---	---

Vector 4

F	F	T	F	T	F	T	F
---	---	---	---	---	---	---	---

#Unir condición 1, 2 y 3 para generar un vector 5

```
flights$dest == "MCO" & flights$month == 12 &  
flights$day == 23
```

Vector 4

F	F	T	F	T	F	T	F
---	---	---	---	---	---	---	---

Vector 3

T	T	T	F	F	T	T	F
---	---	---	---	---	---	---	---

Vector 5

F	F	T	F	F	F	T	F
---	---	---	---	---	---	---	---

El vector 5 resultante de la combinación de las tres condiciones se emplea como un vector de posiciones lógico, que se usará para indicar qué renglones deben extraerse (TRUE) y cuáles no (FALSE).

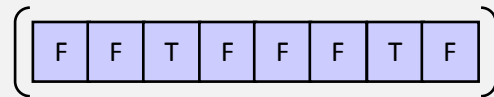


Manipulación de datos. Filtrado por renglones

Vectores lógicos como posiciones

Se emplean los corchetes simples para extraer los renglones correspondientes

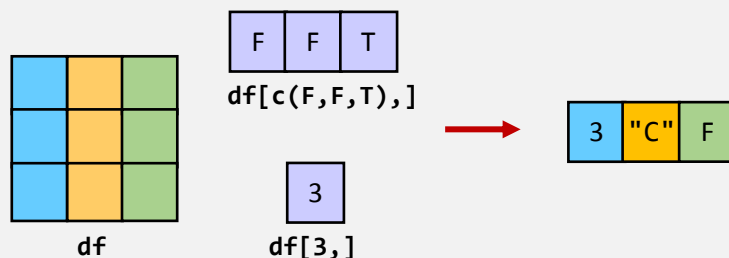
```
flights[flights$dest == "MCO" & flights$month == 12 & flights$day == 23,]
```



year	month	day	...	origin	dest	...
2013	12	23	...	EWR	MCO	...
2013	12	23	...	LGA	MCO	...
2013	12	23	...	JFK	MCO	...
2013	12	23	...	LGA	MCO	...
2013	12	23	...	LGA	MCO	...
2013	12	23	...	LGA	MCO	...
2013	12	23	...	JFK	MCO	...

El vector final es un vector de posiciones que se coloca en la parte de los renglones para extraer de forma específica los valores.

En resumen, para filtrar renglones, se pueden usar vectores numéricos o lógicos, pero el vector lógico debe tener la misma longitud que la cantidad de renglones del dataframe



Resumen: filtrado de datos

Es posible filtrar datos de diferentes maneras, pudiendo extraer una, dos o más columnas, así como uno, dos o más renglones, ello haciendo uso de vectores de posiciones de tipo numérico o lógico.

