R desde Cero

Practica Nº 1. Clase 2

Leyendo Datos

Vamos a cargar datos a R y a explorarlos.

1. Entremos a RStudio
2. Instalaremos un paquete con datos y lo cargaremos para usarlo:

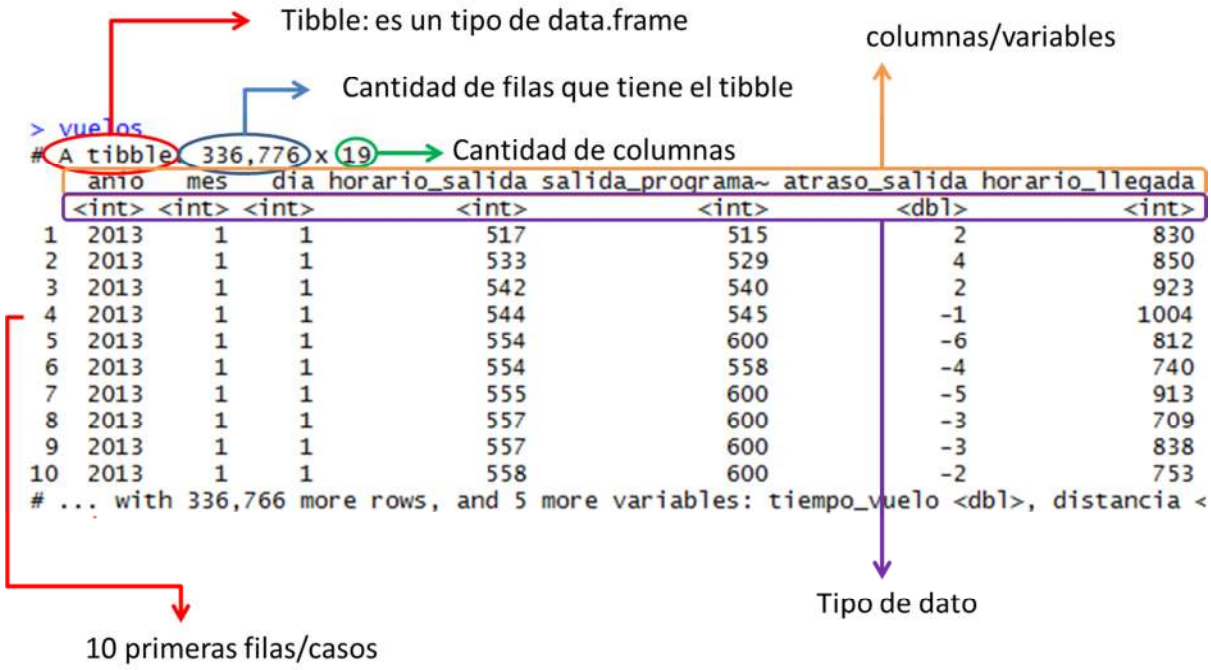
```
> devtools::install_github("cienciadedatos/datos")
> install.packages("nycflights13")

> library(datos)
```

El paquete `datos` y `nycflights13` contiene información de vuelos, aeropuertos, clima entre otras cosas, vamos a utilizarlos para practicar conceptos de manejo de datos:

3. Para ver que contiene un conjunto de datos podemos escribir su nombre en la consola:

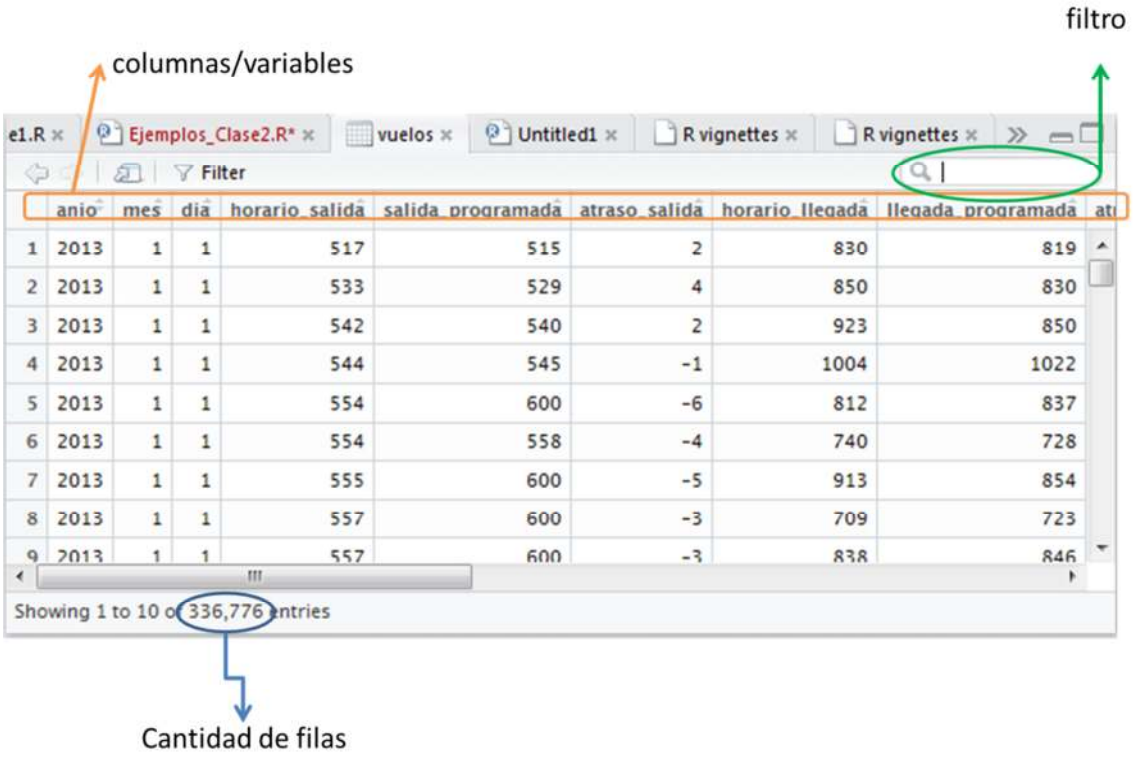
```
> vuelos
```



4. Para poder analizar mejor el set de datos existen otras funciones, probemos con:

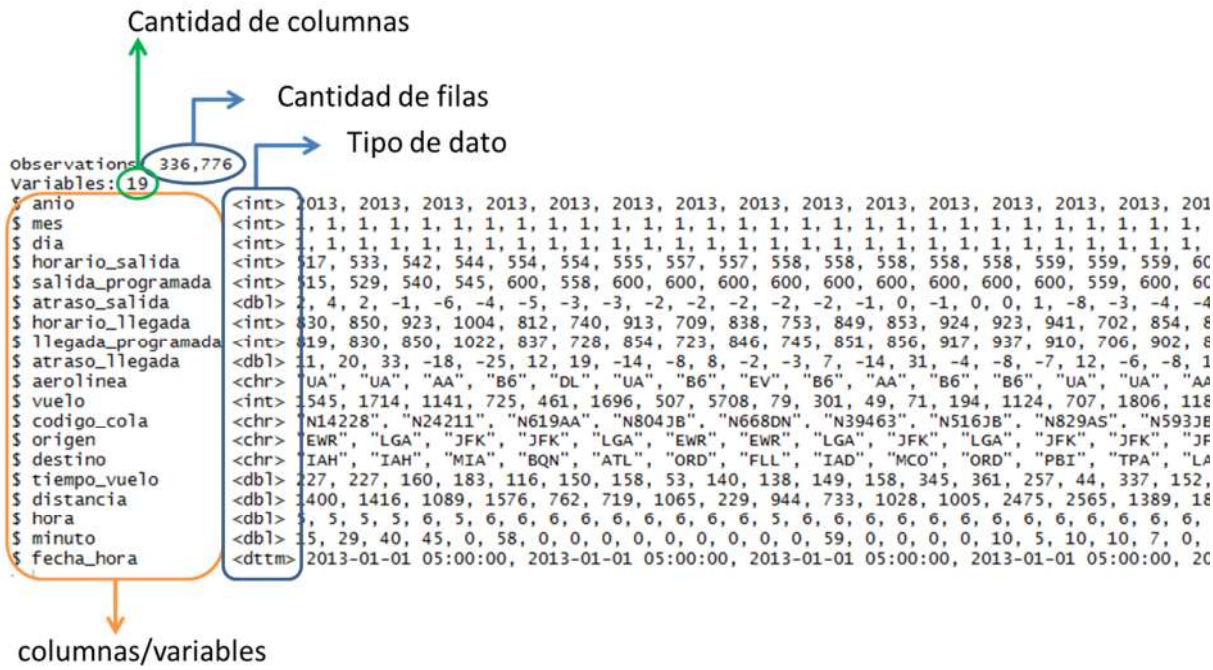
```
> View(vuelos)
```

R desde Cero



5. Otra forma de analizar el set de datos es usando la función `glimpse` del paquete `dyplr`

```
> glimpse(vuelos)
```



6. Otra opción para ver que contiene un set de datos es utilizar la función `kable` del paquete `knitr`, instalemos el paquete `knitr`, cárgalo y probemos la función `kable` con el set de datos de aerolíneas (no usamos vuelos porque tiene más de 300.000 casos y tardaría mucho en generar la salida)

```
> library(knitr)
> kable(aerolineas)
```

```
> kable(aerolineas)
```

codigo_carrier	nombre
9E	Endeavor Air Inc.
AA	American Airlines Inc.
AS	Alaska Airlines Inc.
B6	JetBlue Airways
DL	Delta Air Lines Inc.
EV	ExpressJet Airlines Inc.
F9	Frontier Airlines Inc.
FL	AirTran Airways Corporation
HA	Hawaiian Airlines Inc.
MQ	Envoy Air
OO	Skywest Airlines Inc.
UA	United Air Lines Inc.
US	US Airways Inc.
VX	Virgin America
WN	Southwest Airlines Co.
YV	Mesa Airlines Inc.

7. Por último, el operador $ nos permite explorar una sola variable dentro de un marco de datos. Por ejemplo, ejecuta lo siguiente en tu consola:

```
> aerolineas
> aerolineas$nombre
```

8. Hay una diferencia sutil entre los tipos de variables que podemos tener un set de datos: **variables de identificación y variables de medición/medidas o características.** Por ejemplo, exploremos el marco de datos de los aeropuertos mostrando la salida de glimpse (aeropuertos) a continuación:

```
> glimpse (aeropuertos)
```

9. ¿Qué columnas sirven para identificar de forma única a cada fila?

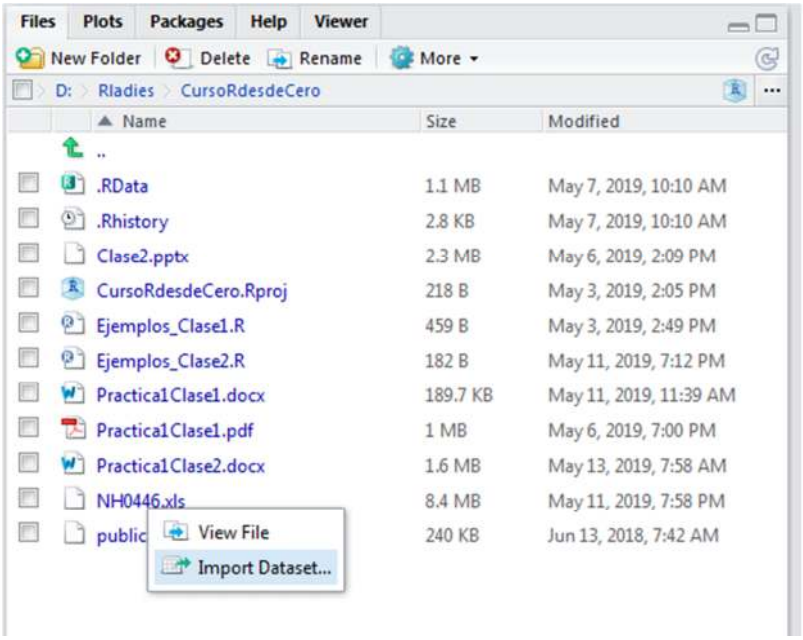
10. ¿A qué hace referencia UNA fila en este conjunto de datos de vuelos?
    a. Datos de una aerolínea
    b. Datos de un vuelo.
    c. Datos de un aeropuerto.
    d. Datos de vuelos múltiples.
11. ¿Cuáles son algunos ejemplos en este conjunto de datos de variables categóricas?

12. ¿Qué las hace diferentes a las variables cuantitativas?

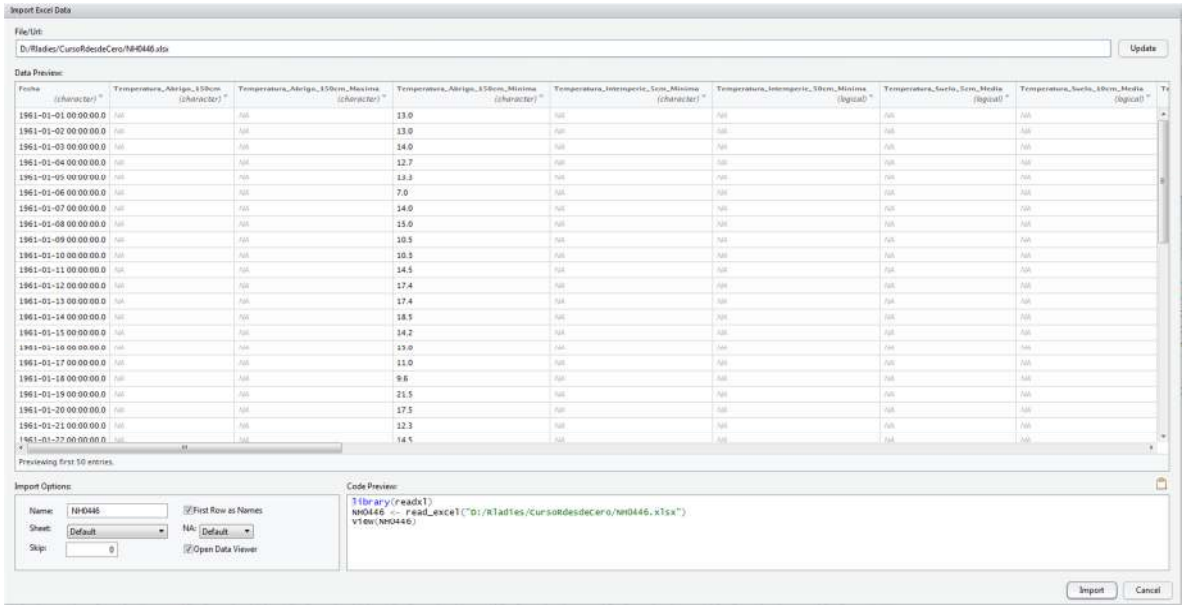
13. ¿Qué ejemplos de variables cuantitativas encontramos en vuelos?

14. Entren a su correo electrónico, a la carpeta compartida de la clase 2 y descarguen los archivos de Excel contenidos en la carpeta.

15. Vamos a cargar una serie de datos externos a R para trabajar con ellos, tenemos tres formas diferentes de hacerlo, la primera en el Panel de Archivos, hacer click sobre el archivo a importar NH0446.xlsx y seleccionar la opción Import DataSet:



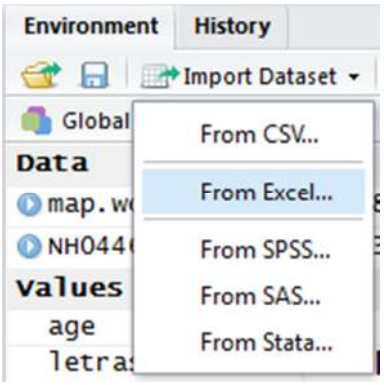
La pantalla muestra una vista pevia de los datos a importar, el tipo de dato, el nombre de las columnas, nos muestra el código de R para poder importar ese archivo, vamos a copiar el código y luego presionamos el botón Import:



16. Vamos a crear un nuevo archivo R Script y allí vamos a pegar el código que se generó cuando importamos el archivo NH0446.xlsx:

```
library(readxl)
NH0446 <- read_excel("D:/Rladies/CursoRdesdeCero/NH0446.xlsx")
View(NH0446)
```

17. Ahora vamos a importar de otra manera, utlizando el botón Import DataSet del panel de Entorno/Historial, nos presenta varias opciones de tipos de archivos a importar, seleccionamos Excel.



18. La pantalla para importar es la misma que con la opción anterior, pero en este caso debemos indicar cual es el archivo que tenemos que importar, para eso presionamos en el botón Browse y allí elegimos el archivo publicaciones_propias_2013.xls.
    Nuevamente nos presenta la misma pantalla. Vamos a copiar el código y presionamos el botón Importar.
    Pegamos el código debajo del código anterior y guarda el archivo como PracticaClase2.R.

```
library(readxl)
publicaciones_propias_2013 <- read_excel("D:/Rladies/CursoRdesdeCero/publ
icaciones_propias_2013.xlsx")
View(publicaciones_propias_2013)
```

¿Hay alguna parte que se repita del código anterior?. Acomodar el código para que el paquete `readxl` solo se cargue una vez

¿En qué objetos se guardaron los datos que importamos?

¿Cómo podemos explorar estos dos set de datos?

¿Qué cantidad de columnas y de casos tiene cada set de datos?