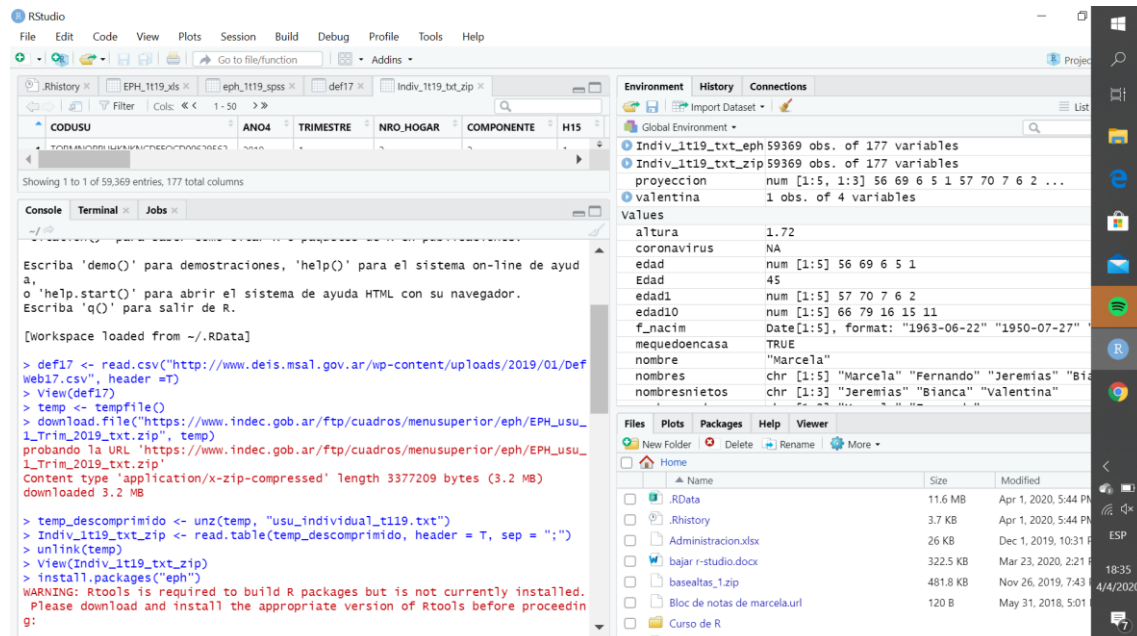


Sigo bajando archivos de distintos orígenes

Bajar datos desde la web – por ejemplo datos de la web de DEIS

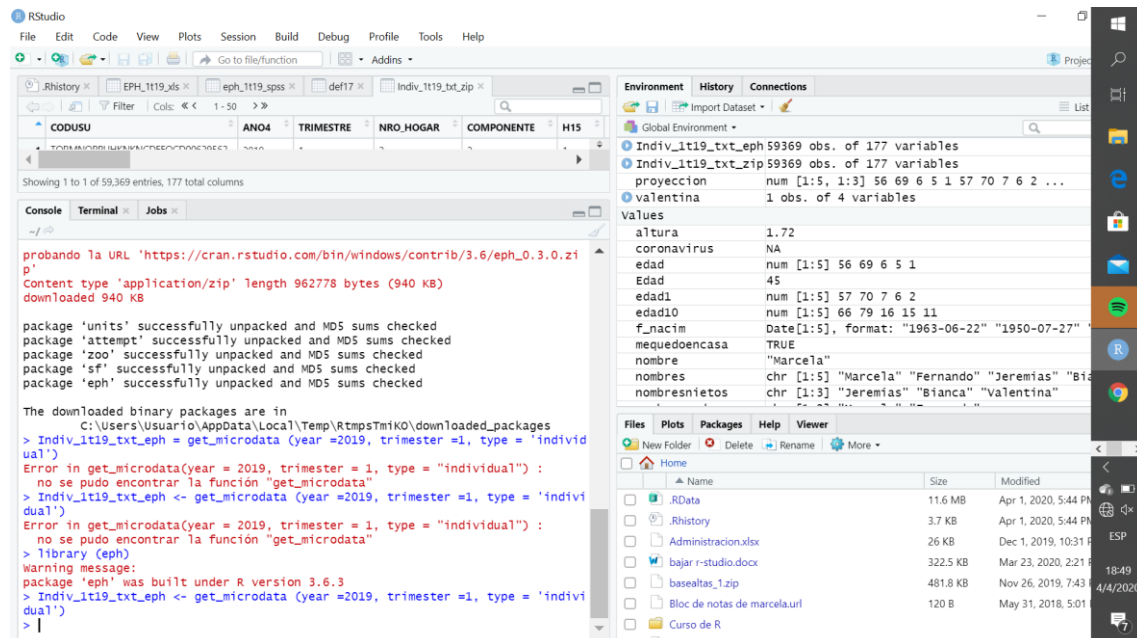
También se puede buscar un carpeta comprimida y descomprimirla y seleccionar el archivo deseado.-



```
def17 <- read.csv("http://www.deis.msa.gov.ar/wp-content/uploads/2019/01/Def17.csv", header = T)
View(def17)
temp <- tempfile()
download.file("https://www.indec.gov.ar/ftp/cuadros/menusuperior/eph/EPH_usu_1Trim_2019_txt.zip", temp)
probando la URL 'https://www.indec.gov.ar/ftp/cuadros/menusuperior/eph/EPH_usu_1Trim_2019_txt.zip'
Content type 'application/x-zip-compressed' length 3377209 bytes (3.2 MB)
downloaded 3.2 MB

temp_descomprimido <- unz(temp, "usu_individual_t119.txt")
Indiv_1t19_txt_zip <- read.table(temp_descomprimido, header = T, sep = ";")
unlink(temp)
View(Indiv_1t19_txt_zip)
install.packages("eph")
WARNING: Rtools is required to build R packages but is not currently installed.
Please download and install the appropriate version of Rtools before proceeding:
```

También puede instalarse el paquete eph – y leer una base de datos de un trimestre



```
probando la URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/eph_0.3.0.zip'
Content type 'application/zip' length 962778 bytes (940 KB)
downloaded 940 KB

package 'units' successfully unpacked and MD5 sums checked
package 'attempt' successfully unpacked and MD5 sums checked
package 'zoo' successfully unpacked and MD5 sums checked
package 'sf' successfully unpacked and MD5 sums checked
package 'eph' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\Usuario\AppData\Local\Temp\RtmpsTm1KO\downloaded_packages
> Indiv_1t19_txt_eph <- get_microdata(year = 2019, trimestre = 1, type = 'individual')
Error in get_microdata(year = 2019, trimestre = 1, type = "individual") :
  no se pudo encontrar la función "get_microdata"
> Indiv_1t19_txt_eph <- get_microdata(year = 2019, trimestre = 1, type = 'individual')
Error in get_microdata(year = 2019, trimestre = 1, type = "individual") :
  no se pudo encontrar la función "get_microdata"
> library(eph)
Warning message:
package 'eph' was built under R version 3.6.3
> Indiv_1t19_txt_eph <- get_microdata(year = 2019, trimestre = 1, type = 'individual')
> |
```

Prueben y vean como carga las bases y el tiempo que demora... son bases chicas relativamente hablando.

También se pueden descargar datos desde un servidor especifico como por ejemplo un servidor con bases sql.

Cargar el archivo C2010_ToyExample_CABA

Paquete dplyr

dplyr es un paquete dentro del paquete tidyverse

Su fin es generar un entorno de herramientas de análisis de datos en R.

Una de sus ventajas es el uso del pipe %>%, que concatena operaciones sobre el mismo objeto.

Activamos la librería y cargamos la base bajo el nombre CABA2010br.

Elegimos algunas de las variables. Primero debemos elegir la tabla y luego las variables que seleccionamos y luego renombramos algunas de las variables.

The screenshot shows the RStudio interface with the following content:

- Console:**

```
> library("tidyverse")
-- Attaching packages ---- tidyverse 1.3.0 --
v dplyr 2.1.1 v purrr 0.3.3
v tidyr 1.2.1 v dplyr 0.8.3
v readr 1.3.1 v stringr 1.4.0
v forcats 0.4.0
-- Conflicts ---- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
Warning message:
package 'tidyverse' was built under R version 3.6.3
> library(dplyr)
> CABA2010br <- read.csv(file="C:/Users/Usuario/OneDrive/Documentos/Curso de R/C2010_ToyExample_CABA.csv", header=T,
  sep=";")
> CABA2010 <- select(CABA2010br, P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO)
> # OTRA FORMA DE HACER LO MISMO CON LOS %>%
> CABA2010 <- CABA2010br %>% select(P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO)
Error in select(., P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO) :
  no se pudo encontrar la función "select"
> CABA2010 <- CABA2010br %>% select(P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO)
Error in select(., P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO) :
  no se pudo encontrar la función "select"
> CABA2010 <- CABA2010br %>% select(P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO)
Error in select(., P01, P02, P03, P05, HOGAR_ID, VIVIENDA_ID, IDOPTO) :
  no se pudo encontrar la función "select"
> CABA2010_1 <- CABA2010 %>% rename(Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05)
Error in rename(., Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05) :
  no se pudo encontrar la función "rename"
> CABA2010_1 <- CABA2010 %>% rename(Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05)
Error in rename(., Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05) :
  no se pudo encontrar la función "rename"
> # también puede pedirse todas las variables de persona "starts_with('P') o también se puede escribir P01:P05
> # ver la cantidad de observaciones
> n_personas <- CABA2010_1 %>% summarise(n())
> View(n_personas)
```
- Environment:** Shows the loaded datasets: CABA2010 (527991 obs. of 7 variables), CABA2010_1 (527991 obs. of 7 variables), CABA2010br (527991 obs. of 15 variables), and n_personas (1 obs. of 1 variable).

The screenshot shows the RStudio interface with the following content:

- Console:**

```
> # ver la cantidad de viviendas
> n_viviendas <- CABA2010_1 %>% summarise(n_distinct(VIVIENDA_ID))
Error in n_distinct(VIVIENDA_ID) :
  no se pudo encontrar la función "n_distinct"
> n_viviendas <- CABA2010_1 %>% summarise(n_distinct(VIVIENDA_ID))
> n_viviendas
n_distinct(VIVIENDA_ID)
1 196293
> # ¿cuál es el promedio de personas por vivienda? ¿cuántos hogares hay en la base? ¿cuántas personas por hogar, en promedio? ¿cuántos hogares por vivienda, en promedio? ¿podemos incluir todo en una sola sentencia?
> CABA2010_1 %>% summarise(Personas = n(), Viviendas = n_distinct(VIVIENDA_ID), Hogares = n_distinct(HOGAR_ID))
# Personas/Viviendas, Personas/Hogares, Hogares/Viviendas
1 527991 196293 210401 2.662681 2.509451 1.061061
```
- Environment:** Shows the updated datasets: CABA2010, CABA2010_1, CABA2010br, n_personas, and n_viviendas (1 obs. of 1 variable).

También podemos agrupar en función de alguna variable

The screenshot shows the RStudio interface. The top-left pane displays a data table with columns **IDOPTO**, **Personas**, and **Viviendas**. The data is as follows:

IDOPTO	Personas	Viviendas
1	37666	14485
2	28754	13058
3	34590	13833

The bottom-left pane shows the R console with the following code and output:

```
C:/Users/Usuario/OneDrive/Escritorio/ >
CABA2010_1 <- CABA2010 %>% rename (Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05)
Error in rename(., Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05) :
  no se pudo encontrar la función "rename"
> CABA2010_1 <- CABA2010 %>% rename (Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05)
Error in rename(., Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05) :
  no se pudo encontrar la función "rename"
> # ver la cantidad de observaciones
n_personas <- CABA2010_1 %>% summarise(n())
> View(n_personas)
> # ver la cantidad de viviendas
n_viviendas <- CABA2010_1 %>% summarise (n_distinct(VIVIENDA_ID))
Error in n_distinct(VIVIENDA_ID) :
  no se pudo encontrar la función "n_distinct"
> n_viviendas <- CABA2010_1 %>% summarise (n_distinct(VIVIENDA_ID))
> n_viviendas
# A tibble: 1 x 1
  n_viviendas
    <dbl>
1     196293
> # ¿Cuál es el promedio de personas por vivienda? ¿Cuántos hogares hay en la base? ¿Cuántas personas por hogar, en promedio?
# ¿Cuántos hogares por vivienda, en promedio? ¿Podemos incluir todo en una sola sentencia?
CABA2010_1 %>% summarise (Personas = n(), Viviendas = n_distinct (VIVIENDA_ID), Hogares = n_distinct (HOGAR_ID), Personas/Vivien
das, Personas/Hogares, Hogares/Viviendas)
Personas/Viviendas Hogares/Viviendas
1 527991 196293 210401 2.662681 2.509451 1.061061
> PersViv_Dpto <- CABA2010_1 %>% group_by(IDOPTO) %>% summarise(Personas = n(), Viviendas = n_distinct (VIVIENDAS_ID))
Error: inesperado "in"
> PersViv_Dpto <- in "PersViv_Dpto" <- CABA2010_1 %>% group_by(IDOPTO) %>% summarise(Personas = n(), Viviendas = n_distinct (VIVIENDAS_ID))
Error in n_distinct_multi(1:1, ..., na.rm) :
  objeto "VIVIENDAS_ID" no encontrado
> View(CABA2010_1)
> PersViv_Dpto <- CABA2010_1 %>% group_by(IDOPTO) %>% summarise(Personas = n(), Viviendas = n_distinct (VIVIENDA_ID))
> View(PersViv_Dpto)
```

The screenshot shows the RStudio interface. The top-left pane displays a data table with columns **Rel_par**, **Sexo**, **Edad**, **Nativo**, **HOGAR_ID**, **VIVIENDA_ID**, and **IDOPTO**. The data is as follows:

Rel_par	Sexo	Edad	Nativo	HOGAR_ID	VIVIENDA_ID	IDOPTO
1 cónyuge	MUJER	61	1	1	1	1
2 cónyuge	VARON	48	1	2	1	1
3 cónyuge	VARON	65	1	3	1	1

The bottom-left pane shows the R console with the following code and output:

```
C:/Users/Usuario/OneDrive/Escritorio/ >
no se pudo encontrar la función "rename"
CABA2010_1 <- CABA2010 %>% rename (Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05)
Error in rename(., Rel_par = P01, Sexo = P02, Edad = P03, Nativo = P05) :
  no se pudo encontrar la función "rename"
> # ver la cantidad de observaciones
n_personas <- CABA2010_1 %>% summarise(n())
> View(n_personas)
> # ver la cantidad de viviendas
n_viviendas <- CABA2010_1 %>% summarise (n_distinct(VIVIENDA_ID))
Error in n_distinct(VIVIENDA_ID) :
  no se pudo encontrar la función "n_distinct"
> n_viviendas <- CABA2010_1 %>% summarise (n_distinct(VIVIENDA_ID))
> n_viviendas
# A tibble: 1 x 1
  n_viviendas
    <dbl>
1     196293
> # ¿Cuál es el promedio de personas por vivienda? ¿Cuántos hogares hay en la base? ¿Cuántas personas por hogar, en promedio?
# ¿Cuántos hogares por vivienda, en promedio? ¿Podemos incluir todo en una sola sentencia?
CABA2010_1 %>% summarise (Personas = n(), Viviendas = n_distinct (VIVIENDA_ID), Hogares = n_distinct (HOGAR_ID), Personas/Vivien
das, Personas/Hogares, Hogares/Viviendas)
Personas/Viviendas Hogares/Viviendas
1 527991 196293 210401 2.662681 2.509451 1.061061
> PersViv_Dpto <- CABA2010_1 %>% group_by(IDOPTO) %>% summarise(Personas = n(), Viviendas = n_distinct (VIVIENDAS_ID))
Error: inesperado "in"
> PersViv_Dpto <- in "PersViv_Dpto" <- CABA2010_1 %>% group_by(IDOPTO) %>% summarise(Personas = n(), Viviendas = n_distinct (VIVIENDAS_ID))
Error in n_distinct_multi(1:1, ..., na.rm) :
  objeto "VIVIENDAS_ID" no encontrado
> View(CABA2010_1)
> PersViv_Dpto <- CABA2010_1 %>% group_by(IDOPTO) %>% summarise(Personas = n(), Viviendas = n_distinct (VIVIENDA_ID))
> View(PersViv_Dpto)
> View(CABA2010br)
> View(CABA2010_1)
> # LA VARIABLE V01 me da referencia la tipo de vivienda - se podrá saber si el promedio de personas difiere por tipo de vivienda
```

Para tener una devolución Espero la sentencia...