

Introducción a R

Contenido

1	Información básica	3
1.1	Descargar e Instalar R y RStudio	3
1.2	¿Qué es R?	3
1.3	¿Qué es RStudio?	3
1.4	RMarkdown y RScript	3
1.4.1	Rscript	3
1.4.2	RMarkdown	4
2	R: Básico.	5
2.1	Paquetes - Librerías	5
2.1.1	Instalar Paquetes.	5
2.1.2	Llamar Paquetes.	5
2.1.3	Otras Paqueterías importantes.	5
2.1.4	Ayuda - Información sobre Paquetes.	6
2.2	Comandos básicos	7
2.2.1	Operaciones matemáticas clásicas	7
2.2.2	Nombres de Objetos	7
2.2.3	Funciones	8
2.3	Dataframe: Estructura fundamental para manipular Conjuntos de Datos.	9
2.3.1	Datasets	9
2.3.2	Dataframe: Estructura base para manipular Conjuntos de Datos	9
2.3.3	Clases de Datos	9
2.3.4	Operador Pipeline: %>%	11
2.3.5	Filtro y selección de Datos: SELECT y FILTER.	11
2.3.6	Ordenar y Agrupar Datos: ARRANGE, GROUP_BY y SUMMARISE	13
2.3.7	Crear nuevas columnas: MUTATE	14

1 Información básica

1.1 Descargar e Instalar R y RStudio

1. Descargar R (<https://cran.r-project.org/>).
2. Descargar RStudio (<https://rstudio.com/products/rstudio/download/>).

1.2 ¿Qué es R?

R es un lenguaje y entorno de libre acceso para **computación estadística** y **gráficos**. R cuenta con una amplia variedad de técnicas estadísticas y gráficas: modelado lineal y no lineal, pruebas estadísticas, análisis de series de tiempo, clasificación, agrupamiento, etc.[1]

1.3 ¿Qué es RStudio?

Es una interfase que permite crear una interacción más fluida con el programa R. Sus principales ventajas son: **el orden** y **la visualización de los procesos** que son llevados a cabo con R. [2]

1.4 RMarkdown y RScript

RStuidip tiene la opción de generar diferentes tipos de documentos, los que se van a abarcar en este documento son: **Rmarkdown** y **RScript**. Para crear cualquiera de estos dos tipos de documentos es necesario: abrir *la aplicación de RStudio*, en la barra principal seleccionar *file*, y seleccionar **Rmarkdown** o **Rscript**.

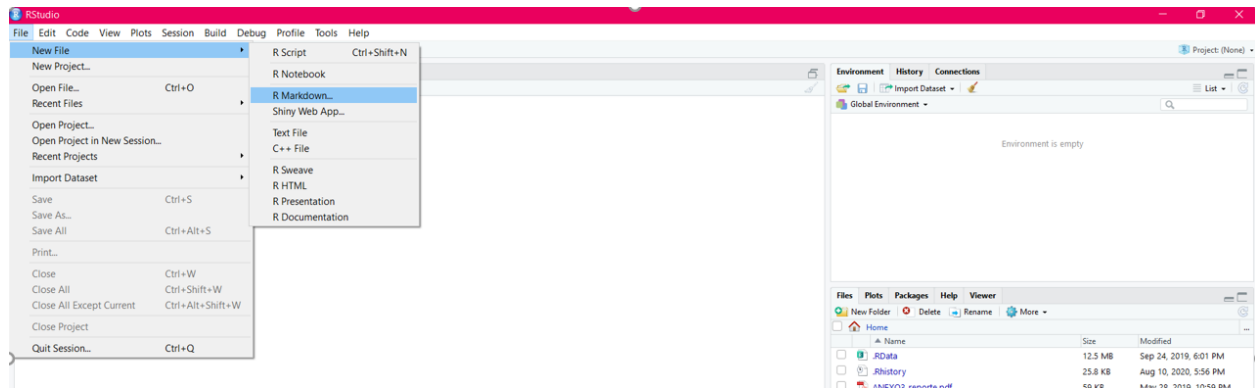


Figura 1: Crear un documento nuevo.

1.4.1 Rscript

Un **script** es un archivo de texto que contiene diversos comandos y comentarios.



Figura 2: Rscript en RStudio.

1.4.2 RMarkdown

R Markdown es un formato que permite una fácil creación de documentos (HTML, PDF, Presentaciones y Documentos de Word). Dentro de estos documentos se puede incluir texto, imágenes, fórmulas y ecuaciones, gráficas y código. [4]

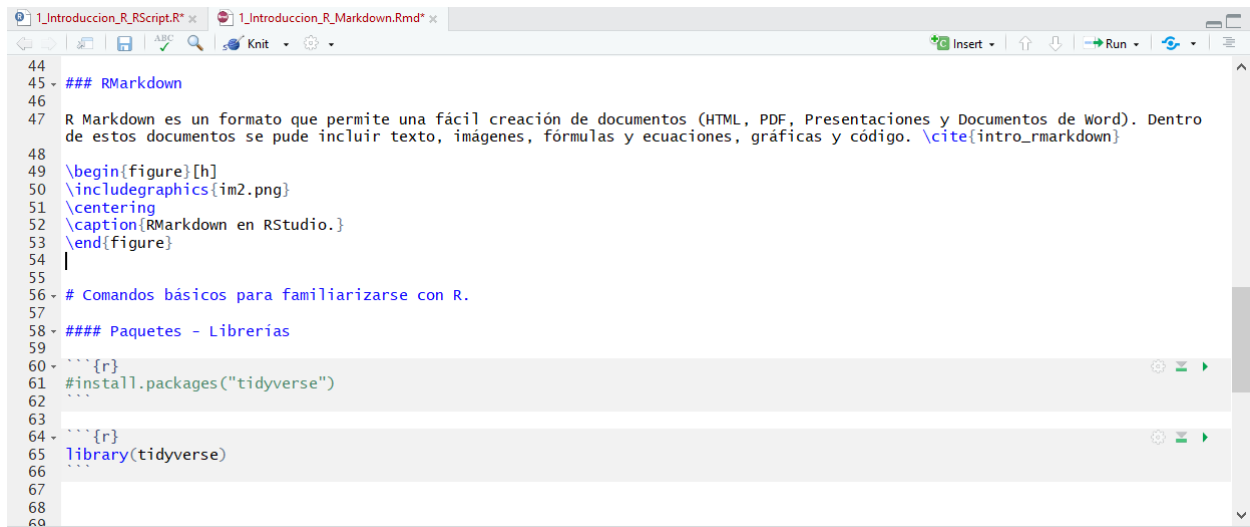


Figura 3: RMarkdown en RStudio.

2 R: Básico.

2.1 Paquetes - Librerías

Actualmente, existen diversos paquetes (librerías) que permiten tener acceso a diversas funciones y datos, para facilitar el análisis y manejo de datos en R.

2.1.1 Instalar Paquetes.

Opción 1:

```
#install.packages("tidyverse")
```

Opción 2:

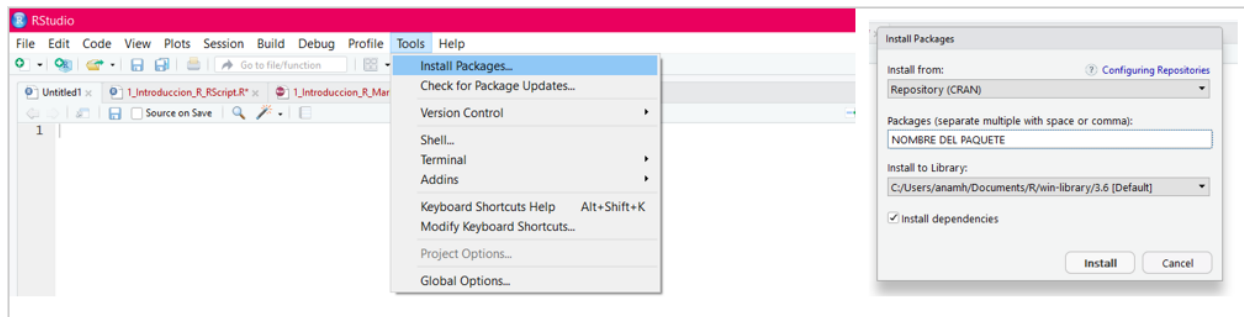


Figura 4: Instalar paquetes.

2.1.2 Llamar Paquetes.

```
library(tidyverse)
```

¡Importante! Tidyverse es un paquete que ayuda en el proceso de importar, transformar, visualizar, modelar y comunicar toda la información que normalmente utilizamos en procesos de ciencia de datos.[5]

2.1.3 Otras Paqueterías importantes.

```
#dplyr - Data manipulation (arrange, filter, group_by, join, lag, select, etc.)  
#readr - Read Rectangular Text Data (read_csv, write_csv)  
#ggplot2 - Create Elegant Data Visualisations Using the Grammar of Graphics.  
#lubridate - Make Dealing with Dates a Little Easier.  
#readxl - Read excel files.
```

2.1.4 Ayuda - Información sobre Paquetes.

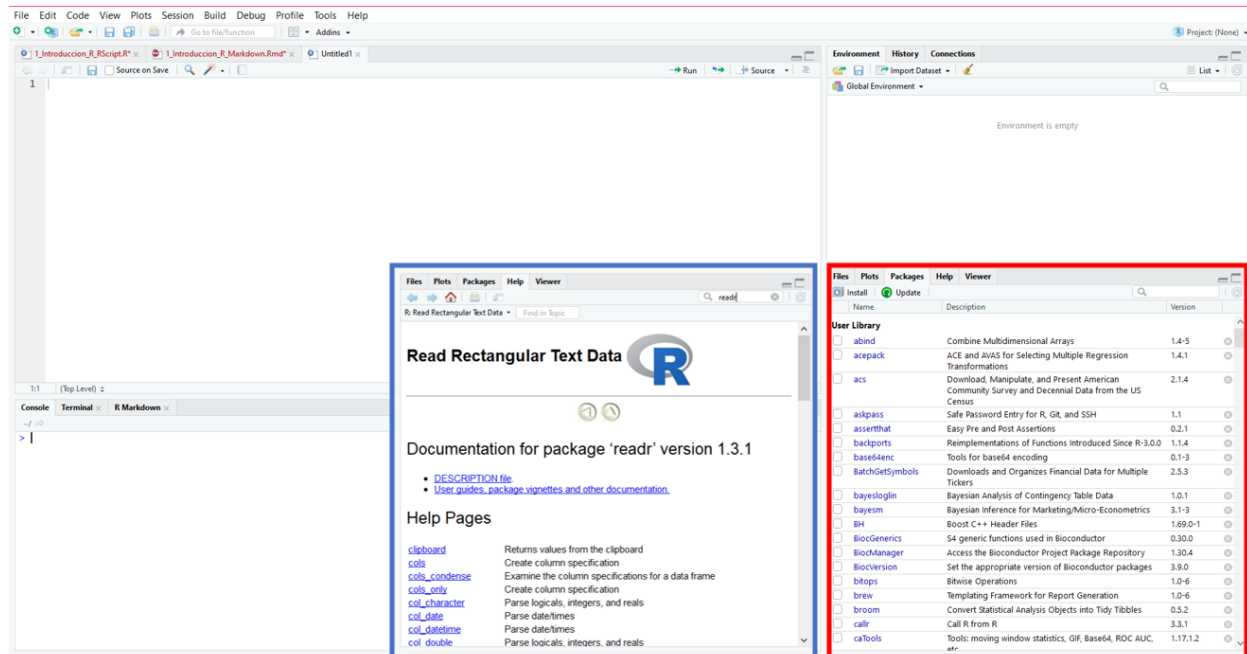


Figura 5: Sección de paquetes y ayuda para obtener más información.



Ejercicio 1: Introducción a R

1. Crear un Rmarkdown para desarrollar **TODOS** los ejercicios de esta sesión. Nombre del documento: *“Actividad_04_A0Matrícula_Nombre_Apellido”*
2. Instalar tidyverse y llamar la librería (incluir los comandos que se utilizaron).

#Respuesta

2.2 Comandos básicos

2.2.1 Operaciones matemáticas clásicas

```
suma <- 10 + 5  
suma
```

```
## [1] 15
```

```
resta <- 10 - 5  
resta
```

```
## [1] 5
```

```
multiplicacion <- 10 * 5  
multiplicacion
```

```
## [1] 50
```

```
division <- 10 / 5  
division
```

```
## [1] 2
```

```
exponencial <- 10 ^ 5  
exponencial
```

```
## [1] 1e+05
```

```
raiz <- 16 ^ (1/2)  
raiz
```

```
## [1] 4
```

```
raiz2 <- sqrt(16)  
raiz2
```

```
## [1] 4
```

```
valores <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)  
promedio <- mean(valores)  
promedio
```

```
## [1] 5.5
```

2.2.2 Nombres de Objetos

- Minúsculas.
- Sin espacios (en lugar de espacios colocar `_`).
- Nombres significativos al contexto.

```
#Convertir grados fahrenheit a celsius
```

```
grados_fahrenheit <- 50  
grados_celsius <- (5 / 9) * (grados_fahrenheit - 32)  
grados_celsius
```

```
## [1] 10
```

2.2.3 Funciones

```
fahrenheit_a_celsius <- function(grados_fahrenheit) {  
  grados_celsius <- (5 / 9) * (grados_fahrenheit - 32)  
  grados_celsius  
}
```

```
fahrenheit_a_celsius(32)
```

```
## [1] 0
```



Ejercicio en Clase

.....

Ejercicio 2: Funciones y Operaciones básicas.

(15 minutos)

1. Programar la función para transformar grados celsius a grados fahrenheit. Probar con `grados_celsius = 0` y otros 2 números que quieras probar.

```
#Respuesta
```

2. Programar la función para transformar grados celsius a grados Kelvin. Probar con `grados_celsius = 0` y otros 2 números que quieras probar.

```
#Respuesta
```

3. Programar la función para transformar grados fahrenheit a grados Kelvin, utilizando las dos funciones anteriores. Probar con `grados_fahrenheit = 32` y otros 2 números que quieras probar.

```
#Respuesta
```

.....

2.3 Dataframe: Estructura fundamental para manipular Conjuntos de Datos.

2.3.1 Datasets

Un dataset es un conjunto de datos. En R hay muchos conjuntos de datos que pueden ser utilizados.

```
#data() - para visualizar la lista con todos los datasets precargados.
```

2.3.2 Dataframe: Estructura base para manipular Conjuntos de Datos

```
#Instalar "titanic" con botones (opción 2)
```

```
library(titanic)
```

```
titanic1 <- titanic_train
```

```
head(titanic1,3)
```

```
## PassengerId Survived Pclass
## 1          1         0      3
## 2          2         1      1
## 3          3         1      3
##
##                               Name      Sex Age SibSp
## 1                               Braund, Mr. Owen Harris   male  22      1
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38      1
## 3                               Heikkinen, Miss. Laina female  26      0
## Parch          Ticket     Fare Cabin Embarked
## 1      0          A/5 21171  7.2500      S
## 2      0          PC 17599 71.2833   C85      C
## 3      0 STON/O2. 3101282  7.9250      S
```

```
class(titanic1)
```

```
## [1] "data.frame"
```

```
#Dimensión de los datos
```

```
dim(titanic1)
```

```
## [1] 891  12
```

2.3.3 Clases de Datos

Hay muchas clases de datos, pero las más importantes/comunes son las siguientes:

1. **INT (Integer)**. Números enteros.
2. **CHAR (Character)**. Variables con contenido no numérico.
3. **NUM (Numeric)**. Números.
4. **FACT (Factor)**. Es el tipo de datos para designar a variables cualitativas.

```
str(titanic1)
```

```
## 'data.frame': 891 obs. of 12 variables:
## $ PassengerId: int 1 2 3 4 5 6 7 8 9 10 ...
## $ Survived : int 0 1 1 1 0 0 0 0 1 1 ...
## $ Pclass : int 3 1 3 1 3 3 1 3 3 2 ...
## $ Name : chr "Braund, Mr. Owen Harris" "Cumings, Mrs. John Bradley (Florence Briggs Thayer)"
## $ Sex : chr "male" "female" "female" "female" ...
## $ Age : num 22 38 26 35 35 NA 54 2 27 14 ...
## $ SibSp : int 1 1 0 1 0 0 0 3 0 1 ...
## $ Parch : int 0 0 0 0 0 0 0 1 2 0 ...
## $ Ticket : chr "A/5 21171" "PC 17599" "STON/O2. 3101282" "113803" ...
## $ Fare : num 7.25 71.28 7.92 53.1 8.05 ...
## $ Cabin : chr "" "C85" "" "C123" ...
## $ Embarked : chr "S" "C" "S" "S" ...
```

```
#Cambiar tipo de datos en dataframe
titanic1$PassengerId <- as.character(titanic1$PassengerId)
class(titanic1$PassengerId)
```

```
## [1] "character"
```

```
titanic1$Sex <- as.factor(titanic1$Sex)
class(titanic1$Sex)
```

```
## [1] "factor"
```

```
levels(titanic1$Sex)
```

```
## [1] "female" "male"
```



Ejercicio en Clase

Ejercicio 3: Tipos de Datos.

(15 minutos)

1. Instalar el paquete “titanic” en RStudio, llamar la librería y asignar a un objeto (con el nombre de tu elección) los datos de entrenamiento que vienen en esa paquetería (“titanic_train”).

#Respuesta

2. Cambiar los tipos de datos a las variables necesarias y justificar por qué.

#Respuesta

2.3.4 Operador Pipeline: %>%

En R cuando uno hace varias operaciones es difícil leer y entender el código; el operador pipeline facilita la creación de código con varias operaciones, al permitir leer las operaciones que vamos aplicando de izquierda a derecha y de arriba hacia abajo.

2.3.5 Filtro y selección de Datos: SELECT y FILTER.

Existen varios comandos para filtrar y seleccionar datos.

2.3.5.1 Visualizar Renglones, Variables y Valores específicos de un Dataframe utilizando Corchetes.

```
#Selecciar un renglón en específico de un dataframe.
titanic1[1, ]
```

```
##   PassengerId Survived Pclass                Name Sex Age SibSp Parch
## 1           1         0       3 Braund, Mr. Owen Harris male  22     1     0
##   Ticket Fare Cabin Embarked
## 1 A/5 21171 7.25          S
```

```
#Seleccionar una columna en específico de un dataframe.
head(titanic1[, 1],3)
```

```
## [1] "1" "2" "3"
```

```
#Seleccionar un valor específico de un dataframe.
titanic1[1,4]
```

```
## [1] "Braund, Mr. Owen Harris"
```

2.3.5.2 Comando: SELECT

Select: Comando de la librería *tidyverse* que sirve para selecciona columnas de acuerdo a algún nombre de las variables.

```
#Selección de la columna completa
select1 <- titanic1 %>% select("PassengerId")
head(select1, 3)
```

```
##   PassengerId
## 1           1
## 2           2
## 3           3
```

```
select2 <- titanic1 %>% select(-Age)
dim(select2)
```

```
## [1] 891  11
```

```
select3 <- titanic1 %>% select(-Survived, -Age)
dim(select3)
```

```
## [1] 891 10
```

```
select4 <- titanic1 %>% select(Name, Survived, Age)
head(select4, 3)
```

```
##                               Name Survived Age
## 1                        Braund, Mr. Owen Harris      0  22
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer)      1  38
## 3                        Heikkinen, Miss. Laina      1  26
```

2.3.5.3 Comando: FILTER

Operadores importantes:

- `x > 1;`
- `x >= 1;`
- `x < 1;`
- `x <= 1;`
- `x != 1;`
- `x == 1`

Filter: Comando de la librería *tidyverse* que sirve para obtener un subconjunto de las filas de acuerdo a algún criterio.

```
filtro1 <- titanic1 %>% filter(Survived==1)
#dim(titanic1) #891 12
dim(filtro1)
```

```
## [1] 342 12
```

```
head(filtro1,3)
```

```
##   PassengerId Survived Pclass
## 1           2         1      1
## 2           3         1      3
## 3           4         1      1
##
##                               Name      Sex Age SibSp
## 1 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female   38     1
## 2                               Heikkinen, Miss. Laina female   26     0
## 3 Futrelle, Mrs. Jacques Heath (Lily May Peel) female   35     1
##   Parch      Ticket    Fare Cabin Embarked
## 1     0   PC 17599  71.2833   C85        C
## 2     0 STON/O2. 3101282   7.9250        S
## 3     0   113803  53.1000  C123        S
```

```
filtro2 <- titanic1 %>% filter(Age <=10)
dim(filtro2)
```

```
## [1] 64 12
```

```
head(filtro2,3)
```

```
##   PassengerId Survived Pclass                Name    Sex Age
## 1         8         0       3  Palsson, Master. Gosta Leonard  male   2
## 2        11         1       3 Sandstrom, Miss. Marguerite Rut  female  4
## 3        17         0       3      Rice, Master. Eugene    male   2
##   SibSp Parch  Ticket   Fare Cabin Embarked
## 1     3     1  349909 21.075         S
## 2     1     1 PP 9549 16.700      G6         S
## 3     4     1  382652 29.125         Q
```

2.3.6 Ordenar y Agrupar Datos: ARRANGE, GROUP_BY y SUMMARISE

2.3.6.1 Comando: Arrange

```
#Orden ascendente
orden1 <- titanic1 %>% arrange(PassengerId)
head(orden1, 3)
```

```
##   PassengerId Survived Pclass                Name    Sex
## 1          1         0       3      Braund, Mr. Owen Harris  male
## 2         10         1       2 Nasser, Mrs. Nicholas (Adele Achem) female
## 3        100         0       2      Kantor, Mr. Sinai    male
##   Age SibSp Parch  Ticket   Fare Cabin Embarked
## 1  22     1     0 A/5 21171  7.2500         S
## 2  14     1     0  237736 30.0708         C
## 3  34     1     0  244367 26.0000         S
```

```
#Orden ascendente
orden2 <- titanic1 %>% arrange(desc(PassengerId))
head(orden2, 3)
```

```
##   PassengerId Survived Pclass                Name    Sex
## 1         99         1       2 Doling, Mrs. John T (Ada Julia Bone) female
## 2         98         1       1  Greenfield, Mr. William Bertram  male
## 3         97         0       1  Goldschmidt, Mr. George B    male
##   Age SibSp Parch  Ticket   Fare  Cabin Embarked
## 1  34     0     1  231919 23.0000         S
## 2  23     0     1 PC 17759 63.3583 D10 D12         C
## 3  71     0     0 PC 17754 34.6542      A5         C
```

2.3.6.2 Comandos: group_by y summarise

Estos dos comandos trabajan muy bien juntos, dado que permite generar resúmenes de variables.

```
resumen1 <- titanic1 %>%
  group_by(Sex) %>%
  summarise(total_survived = sum(Survived))
resumen1
```

```
## # A tibble: 2 x 2
##   Sex      total_survived
##   <fct>         <int>
## 1 female           233
## 2 male             109
```

2.3.7 Crear nuevas columnas: MUTATE

```
mutate_1 <- titanic1 %>% mutate(valor_individual = 1)
dim(mutate_1)
```

```
## [1] 891 13
```

```
head(mutate_1, 3)
```

```
##   PassengerId Survived Pclass
## 1           1         0       3
## 2           2         1       1
## 3           3         1       3
##
##                               Name      Sex Age SibSp
## 1                               Braund, Mr. Owen Harris   male  22     1
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38     1
## 3                               Heikkinen, Miss. Laina female  26     0
##   Parch      Ticket    Fare Cabin Embarked valor_individual
## 1     0   A/5 21171  7.2500      S          1
## 2     0    PC 17599 71.2833   C85      C          1
## 3     0 STON/O2. 3101282  7.9250      S          1
```



Ejercicio en Clase

Ejercicio 4: Comandos generales de R para contestar Preguntas.

A continuación, se presentan unas preguntas; estas preguntas deben de tener una respuesta clara y concisa, y además se debe de incluir el código que se utilizó para llegar a las respuestas.

¡Importante! Si se desea revisar nuevamente las descripción de cada una de las variables, favor de dar click en “help” y en la sección de buscar, colocar “titanic_train”.

1. ¿Cuántos hombres y mujeres había abordo? (Tip: Utilizar la variable creada “valor individual”)

#Respuesta

2. ¿Cuántas clases había designadas para los pasajeros?

#Respuesta

3. ¿Cuántos pasajeros estaban en cada una de las clases?

#Respuesta

5. ¿Cuál fue la clase que tuvo más decesos?

#Respuesta

6. ¿Cuál fue la clase que tuvo menos decesos?

#Respuesta

7. ¿Cuántos pasajeros menores de 15 años sobrevivieron?

#Respuesta

8. ¿Cuántos pasajeros mayores de 60 años sobrevivieron?

#Respuesta

9. ¿Cuántos pasajeros menores de 10 años iban a bordo?

#Respuesta

10. ¿Cómo se llama el pasajero cuyo PassengerId es igual a 175?

#Respuesta

11. ¿Cuál es el nombre de la person que tenía el número de ticket PC 17593? Busca el nombre en internet y describe brevemente quién era.

#Respuesta

.....

Referencias

- [1] Cran.RProject (2020). The Comprehensive R Archive Network. 13 de agosto de 2020, de Sitio web: <https://cran.r-project.org/>
- [2] Boccardo, G.y Ruiz, F. (2019). RStudio para Estadística Descriptiva en Ciencias Sociales. 13 de agosto de 2020, de Sitio web: <https://bookdown.org/gboccardo/manual-ED-UCH/>
- [3] Aleshunas, J. (s.f.). R learning infrastructure. 13 de agosto de 2020, de Sitio web: http://mercury.webster.edu/aleshunass/R_learning_infrastructure/R%20scripts.html
- [4] Rstudio. (2014). RMarkdown hoja de referencia. 13 de agosto de 2020, de Sitio web: <https://rstudio.com/wp-content/uploads/2015/03/rmarkdown-spanish.pdf>
- [5] Gonzalez, R. (2020). ¿Qué es tidyverse? 8 paquetes para ciencia de datos. 13 de agosto de 2020, de Sitio web: https://gonzalezgouveia.com/que-es-tidyverse-8-paquetes-para-ciencia-de-datos/#Que_es_Tidyverse
- [6] Ortiz, T. (2014).Sección 3 Manipulación y agrupación de datos. 13 de agosto de 2020, de Sitio web: <https://tereom.github.io/est-computacional-2018/>