



Actividad

Te propongo las siguientes actividades como práctica de los contenidos vistos a lo largo de esta semana, en el cual estuvimos conociendo las funciones más relevantes de los paquetes del ecosistema tidyverse como **readr**, **readxl**, **tidyr**, **forcats** y **lubridate**.

Antes de iniciar:

- Abre RStudio

- Crea un script incorporando:

En las primeras líneas el borrado completo del entorno.

Configura el Directorio de trabajo

Escribe el código necesario para resolver cada punto planteado.

Al finalizar guárdalo con el nombre **E03_ActividadNiños**

1- Activa los paquetes necesarios del entorno Tidyverse

2- En la plataforma encontrarás un libro de cálculo llamado NiñosLP.xlsx. Escribe el código para importarlo a R y abre la hoja llamada "GENERAL" y asígnaselo a un objeto llamado **datos**. Se trata de un conjunto de datos ficticios, colectados durante una encuesta en el año 2011, en diversas escuelas de la Ciudad de La Plata y Gran La Plata, compuesto por 481 observaciones y 8 variables, adaptado para fines docentes.

3- Escribe el código de comandos para observar la estructura del conjunto de datos.

4- Escribe el código de comandos para observar el nombre de las variables.

5- Muestra en la consola si hay valores NA dentro del conjunto de datos y cuántos son en total. En qué variables están ubicados y cuántos por cada variable.

6- Observa la clase de la variable Sexo y las categorías (niveles). Transforma la variable sexo en factor. Ten en cuenta que 0 es el código de Hombre y 1 de Mujer

7- Observa las variables Fecha de Nacimiento (FechaNac) e indica su clase. Redondea las fechas



8- Genera una nueva variable llamada Edad_actual como diferencia entre la fecha de nacimiento (FechaNac) y la fecha de la entrevista (Fecha)

9) Teniendo en cuenta que la variable Peso se encuentra en Kilogramos y la variable Talla en centímetros, incorpora una nueva variable calculada con los valores del índice de masa corporal, según la fórmula $IMC = \text{Peso} / (\text{Talla})^2$. Las unidades deben ser kilogramo para el peso y metro para la talla.

b) Agrupa los valores de IMC creando una nueva variable gIMC y asigna el nivel tomando en cuenta la siguiente tabla:

Valor IMC	Agrupamiento
<13	Bajo Peso
>=13 < 19	Normal
>=19 < 24	Sobrepeso
>= 24	Obesidad

Observa nuevamente si hay valores NA dentro del conjunto de datos y cuántos son en total. En qué variables están ubicados y cuántos por cada variable

10- a) Verifica la clase de la variable Zona. Transfórmala en factor. Cambia la etiqueta de Casco Urbano por Centro_LP.

Nota: conserva este scripts ya que en las actividades de los s próximos encuentros, continuaremos trabajando sobre el mismo conjunto de datos

RESPUESTAS

```
> # 1) activamos el paquete-----
> require(tidyverse)
> library(readr)
> library(tidyverse)
> library(readxl)
> require(tibble)
> require(lubridate)
> # 2) lectura de datos-----
> datos<-read_excel("NiñosLP.xlsx",sheet="GENERAL",col_names = T)
```



```
> datos

> # 3) estructura de datos-----
> str(datos) # tengo 481 observaciones y 8 variables
> glimpse(datos) # otra forma de hacerlo

> # 4) nombre de las columnas/variables-----
> ncol(datos)
> colnames(datos)

> # 5) valores NA-----
> na.fail(datos) # pregunto si tiene NA
Error in na.fail.default(NiñosLP_Brote) : missing values in object
> sum(is.na(datos)) # cuantos hay?
> sapply(datos,function(datos)sum(is.na(datos)))

> # 6) variable Sexo-----
> class(datos$Sexo) # numeric
> unique(datos$Sexo) # 1 0
> # Transformo la variable sexoa a factor
> datos$Sexo<- factor(datos$Sexo,labels= c("Hombre","Mujer"))
> class(datos$Sexo)
> unique(datos$Sexo)

> # 7) variable Fecha Nacimiento y Fecha de Entrevista-----
> class(datos$FechaNac)
> round_date(datos$FechaNac)
> class(datos$Fecha)
> round_date(datos$Fecha)

> # 8) genero variable edad al momento de la entrevista-----
> datos$Edad_actual<- round((datos$Fecha - datos$FechaNac)/365.25)
> edad$Edad_actual

> # 9) crear variable IMC-----
> head(datos$Peso) # vemos la cabecera de la variable Peso
> head(datos$Talla) # vemos con al cabecera que la talla está en cm
```



```
>datos<- transform(datos,Talla=Talla/100) # transformo a metro porque tengo dato en cm
>datos$IMC<- round(datos$Peso/(datos$Talla*datos$Talla),2) # creo la nueva variable
>head(datos$IMC) # miro la cabecera
# agrupamiento de la variable IMC
>datos$gIMC<- cut(datos$IMC,breaks = c(0,13,19,24,+Inf),
                  labels = c("Bajo Peso","Normal","Sobrepeso","Obesidad"),
                  right=FALSE)
>datos$gIMC
[1] Normal Normal Normal Sobrepeso Normal Normal Normal Normal
[9] <NA> Normal Normal Normal Obesidad Normal Sobrepeso Normal
[17] Normal Normal Normal Obesidad Sobrepeso Normal Normal Normal
[25] Normal Sobrepeso Obesidad Normal Normal Obesidad Normal Normal
[33] Sobrepeso Normal Sobrepeso Sobrepeso Normal Normal Normal Bajo Peso
[41] Normal Normal Normal Normal Normal Normal Sobrepeso Normal
[49] Normal Normal Sobrepeso Normal Normal Bajo Peso Sobrepeso Normal
[57] Normal Sobrepeso Sobrepeso Normal Normal Normal Normal Normal
[65] Obesidad Normal Normal Sobrepeso Normal Normal Sobrepeso Normal
[73] Normal Normal Normal Normal Normal Normal Sobrepeso Normal
[81] Normal Normal Normal Sobrepeso Normal Normal Normal Normal
[89] Normal Normal Normal Normal Normal Normal Normal Sobrepeso
[97] Normal Normal Normal Normal Normal Normal Normal Sobrepeso
.....
```

10) variable zona----

```
>class(datos$Zona)
>datos$Zona<- as.factor(datos$Zona)
>class(datos$Zona)
>unique(datos$Zona)

>require(forcats)
>datos$Zona<-fct_recode(datos$Zona, Centro_LP="Casco Urbano")
>unique(datos$Zona)
```

