

# Manejo de Na y otros datos

Armando Ocampo

## Not available

Los Na (not available) son objetos que carecen de información dentro de nuestro data set. Esta falta de información puede ser un problema al utilizar algunas funciones. De la misma forma, existen herramientas de trabajo que nos permiten remover estos elementos o trabajar a pesar de ellos. Estos datos se pueden representar de la siguiente manera Na, NA, NAs. Algunos data sets los colocan de forma automática o pueden ser colocados por default en nuestro data set al encontrar una celda vacía.

Vamos a comenzar con algunos ejemplos con vectores completos y vectores con NAs

```
vector_cito <- c(1:10, 20:30,51)
vector_cito

not_a <- c(1:21, NA, 22:30)
not_a

vect_na <- c(1:5, NA, 14:20, NA, 21:25)
vect_na
```

Antes de comenzar a trabajar, te recomiendo hacer un resumen de tu data set

```
summary(vector_cito)
str(vector_cito)

summary(not_a)
str(not_a)

summary(vect_na)
str(vect_na)
```

Al utilizar funciones de estadística como *mean()*, *sum()*, *median()*, etc. Podemos encontrarnos con el problema que generan los Na

```
# sum
sum(not_a)

# mean
mean(not_a)

# median
median(not_a)
```

El tener Na en un conjunto de datos es tan frecuente que algunas funciones tienen la opción de ignorarlos y realizar la función pertinente

```
?sum  
  
sum(not_a, na.rm = TRUE)  
  
mean(not_a, na.rm = TRUE)  
  
median(not_a, na.rm = TRUE)
```

El argumento *na.rm=*, omite los Na para realizar la función. No los elimina del conjunto de datos original

## Removiendo Na

No obstante, en algunas ocasiones es preferible determinar el numero de NAs que se presentan en un data set y eliminarlos. Para conocer si nuestros datos presentan NA se utiliza la función *is.na()*

```
is.na(not_a)  
  
is.na(vect_na)
```

En conjunto con *which()* podemos conocer la posición en la que se encuentra cada Na

```
which(is.na(not_a))  
  
which(is.na(vect_na))
```

Eliminarlos es fácil, esto se puede realizar a partir de la función *na.omit()*, el cual elimina todos los Na de nuestro conjunto de datos

```
na.omit(not_a)  
  
# este codigo elimina los NA, para guardar el resultado lo podemos asignar a un  
# objeto nuevo o renombrar el objeto not_a  
  
obj1_sin_na <- na.omit(not_a)  
obj1_sin_na  
  
sum(obj1_sin_na)  
mean(obj1_sin_na)  
var(obj1_sin_na)  
sd(obj1_sin_na)
```

Nota: la función *na.exclude()* realiza un proceso similar para eliminar los Na del conjunto de datos. De hecho, se utiliza de forma similar. ej. *na.exclude(not\_a)*