

PRÁCTICAS BIOINFORMÁTICA II

Marina Paz (143353)

Paula Guevara (145921)

Ejercicio 1: filtrar y seleccionar datos

- ¿Cuántos registros cumplen las condiciones finales?
 - Ser sólo humanos: **35**
 - Ser sólo humanos del planeta Tatooine: **8**
 - Todas las especies menos los Droides: **77**

Ejercicio 2: calcular algunos estadísticos

- ¿Cómo calcularías la desviación estándar (sd) de esos parámetros?

Para calcular la desviación estándar de los parámetros para cada especie, utilizamos esta función y lo introducimos dentro de un data frame al que llamamos “desviación” para poder ver mejor el resultado de todas las especies.

```
desviacion <- starwars %>% group_by(species) %>% summarise(sd(height, na.rm = T), sd(mass, na.rm = T))
```

	species	sd(height, na.rm = T)	sd(mass, na.rm = T)		species	sd(height, na.rm = T)	sd(mass, na.rm = T)
1	Aleena	NA	NA	18	Mon Calamari	NA	NA
2	Besalisk	NA	NA	19	Muun	NA	NA
3	Cerean	NA	NA	20	Nautolan	NA	NA
4	Chagrian	NA	NA	21	Neimodian	NA	NA
5	Clawdite	NA	NA	22	Pau'an	NA	NA
6	Droid	49.149771	51.031853	23	Quermian	NA	NA
7	Dug	NA	NA	24	Rodian	NA	NA
8	Ewok	NA	NA	25	Skakoan	NA	NA
9	Geonosian	NA	NA	26	Sullustan	NA	NA
10	Gungan	14.189198	11.313708	27	Tholothian	NA	NA
11	Human	12.045889	19.332163	28	Togruta	NA	NA
12	Hutt	NA	NA	29	Toong	NA	NA
13	Iktotchi	NA	NA	30	Toydarian	NA	NA
14	Kaleesh	NA	NA	31	Trandoshan	NA	NA
15	Kaminoan	11.313708	NA	32	Twi'lek	1.414214	NA
16	Kel Dor	NA	NA	33	Vulptereen	NA	NA
17	Mirialan	2.828427	4.384062	34	Wookiee	4.242641	16.970563
35	Xexto	NA	NA				
36	Yoda's species	NA	NA				
37	Zabrak	2.828427	NA				
38	NA	12.355835	31.192948				

En caso de querer calcular la desviación estándar de la altura y la masa teniendo en cuenta todos los datos, utilizamos:

```
sd(starwars$height, na.rm = T)
```

```
34.77416
```

```
sd(starwars$mass, na.rm = T)
```

```
169.4572
```

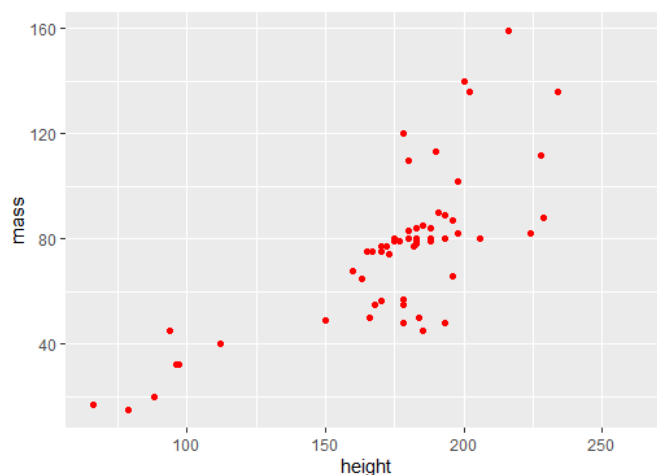
Ejercicio 3: crear gráficos y modificar algunos elementos

1. Al crear los gráficos puedes observar que hay un punto que corresponde a un personaje con una masa muy grande. Crea un nuevo dataframe sin ese personaje y crea de nuevo el gráfico final. Primero creamos un data frame al que llamamos "starwars2" y usamos "filter" para filtrar todos los datos de starwars excluyendo al personaje cuya masa es notablemente superior al resto.

```
starwars2 <- starwars %>% filter(name != "Jabba Desilijic Tiure")
```

Una vez creado el nuevo data frame, usamos el comando ggplot para elaborar la gráfica.

```
ggplot(starwars2, aes(height, mass)) + geom_point(colour = "red") + theme_light()
```



2. Inspecciona el dataset, haz un resumen de la media de las variables (Peso, Altura, IMC, IAS, CCintura). Agrupando por sexo.

Utilizamos el comando group_by para agrupar las medias de todos los parámetros en función del sexo. También, usamos la función na.rm para descartar de la operación todos los datos no asociados (NA). Todo ello lo metemos en un nuevo data frame al que llamamos "medias" para observar los resultados.

```
medias <- toy %>% group_by(Sex) %>% summarise(mean_Weight_Kg = mean(Weight_Kg, na.rm = T), mean_Height_cm = mean(Height_cm, na.rm = T), mean_IMC = mean(IMC, na.rm = T), mean_IAS = mean(IAS, na.rm = T), mean_Ccintura = mean(Ccintura, na.rm = T))
```

	Sex	mean_Weight_Kg	mean_Height_cm	mean_IMC	mean_IAS	mean_Ccintura
1	Men	78.08333	176.4167	25.13333	59.59000	81.72727
2	Women	63.04310	162.3448	23.92414	63.51154	80.88571

3. Haz una tabla sólo con los pacientes femeninos

Primero creamos un nueva data frame al que llamamos "toy2" que contiene solo los datos de las mujeres del archivo "toy"

```
toy2 <- toy %>% filter(Sex=="Women")
```

Después, del data frame "toy2" filtramos solo los personajes que tienen sobrepeso y lo introducimos en un nuevo data frame denominado "toy2sobrepeso".

```
toy2sobrepeso <- toy2 %>% filter(IMC_clas=="Overweight")
```

¿Cuántos registros cumplen las condiciones?

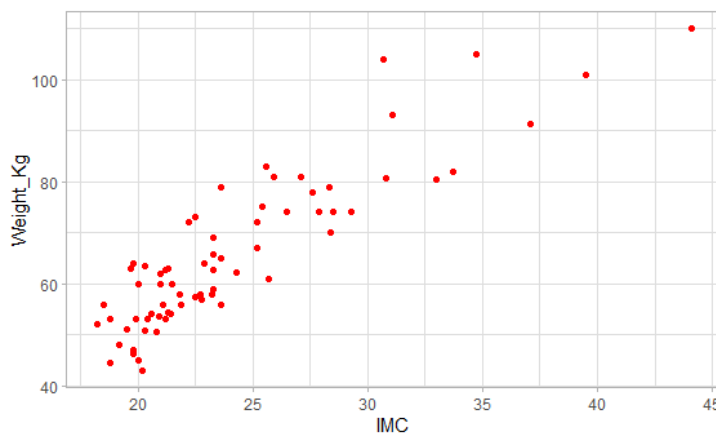
58 personajes son mujeres

¿De estos cuantos tienen Sobrepeso (Overweight)?

9 de las mujeres tienen sobrepeso

4. Haz un gráfico usando ggplot relacionando el IMC (Índice de masa corporal) con el peso (Weight_Kg) de todos los pacientes.

```
ggplot(toy, aes(IMC, Weight_Kg)) + geom_point(colour = "red") + theme_light()
```



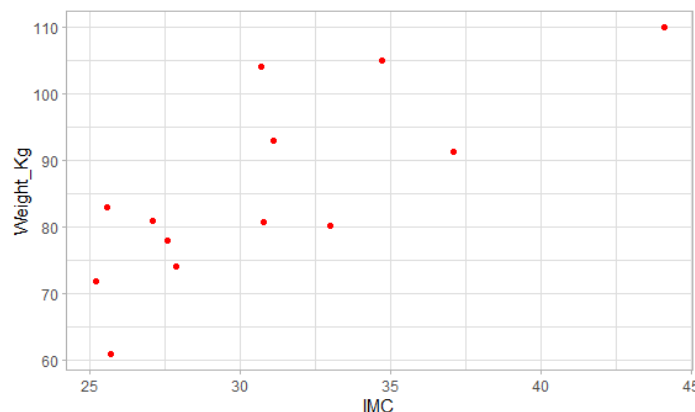
5. Repítelo filtrando sólo los pacientes categorizados como "Overweight" y "Obesity".

Del archivo toy filtramos todos los personajes que tengan sobrepeso y obesidad y los introducimos en un nuevo data frame denominado toysobrepeso. `toysobrepeso <- toy %>%`

```
filter(IMC_clas==c("Overweight", "Obesity"))
```

Luego lo representamos en una gráfica.

```
ggplot(toysobrepeso, aes(IMC, Weight_Kg)) + geom_point(colour = "red") + theme_light()
```



SCRIPT

```
library(tidyverse)
```

```
data("starwars")
```

```
#Ejemplos usando select
```

```
# Seleccionar todas las columnas menos el nombre
```

```
starwars %>% select(-name)
```

```
#Seleccionar sólo las columnas que tienen subraya (_)
```

```
starwars %>% select(contains("_"))
```

```
#Seleccionar sólo las columnas que empiezan con "s"
```

```
starwars %>% select(starts_with("s"))
```

```
#Crear un data frame con los nombres y planeta de origen (homeworld)
```

```
homeworld <- starwars %>% select(name, homeworld)
```

```
#Filtrar datos
```

```
#Filtrar por especies: sólo humanos
```

```
human <- starwars %>% filter(species == "Human")
```

```
#Filtrar por especies: sólo humanos del planeta Tatooine
```

```
Tatooine <- starwars %>% filter(species == "Human", homeworld == "Tatooine")
```

```
#Crear un nuevo dataframe con todas las especies menos los Droides
```

```
starwars_nodroids <- starwars %>% filter(species != "Droid")
```

```
#Seleccionar y agrupar datos
```

```
#Usamos group_by y tally
```

```
starwars %>% group_by(species) %>% tally()
```

```
#Añadiendo otra variable
```

```
starwars %>% group_by(species, gender) %>% tally()
```

```
#Si lo quieres guardar en el environment recuerda asignarle un nombre
```

```
table_gender <- starwars %>% group_by(species, gender) %>% tally()
```

```
#Calcular algunos estadísticos
```

```
#na.rm=T quiere decir que elimina los NA (valores No Asignados o sin datos)
starwars %>% group_by(species) %>% summarise(mean_height = mean(height, na.rm = T), mean_mass =
mean(mass, na.rm = T))
#Desviación de la altura
sd(starwars$height, na.rm = T)
#Desviación de la masa
sd(starwars$mass, na.rm = T)
desviacion <- starwars %>% group_by(species) %>% summarise(sd(height, na.rm = T), sd(mass, na.rm = T))
#Crear gráficos y modificar algunos elementos
```

```
#Hacer un gráfico de la altura vs. la masa de los personajes
ggplot(starwars, aes(height, mass)) + geom_point()
#Puedes modificar el color
ggplot(starwars, aes(height, mass)) + geom_point(colour = "red")
#Modificando el color y el punto
ggplot(starwars, aes(height, mass)) + geom_point(colour = "purple", pch = 3)
#Modificando el color y el fondo
ggplot(starwars, aes(height, mass)) + geom_point(colour = "red") + theme_light()
```

```
#haz un data frame en el que excluyas la fila que contenga el nombre:
starwars2 <- starwars %>% filter(name != "Jabba Desilijic Tiure")
ggplot(starwars2, aes(height, mass)) + geom_point(colour = "red") + theme_light()
```

```
toy <- read_csv("C:/Users/Paula/OneDrive/Documentos/BIOINFORMÁTICA/toy.csv")
medias <- toy %>% group_by(Sex) %>% summarise(mean_Weight_Kg = mean(Weight_Kg, na.rm =
T), mean_Height_cm = mean(Height_cm, na.rm = T), mean_IMC = mean(IMC, na.rm = T), mean_IAS =
mean(IAS, na.rm = T), mean_Ccintura = mean(Ccintura, na.rm = T))
toy2 <- toy %>% filter(Sex=="Women")
toy2sobrepeso <- toy2 %>% filter(IMC_clas=="Overweight")
```

```
#Haz un gráfico usando ggplot relacionando el IMC (Índice de masa corporal) con el peso (Weight_Kg) de
todos los pacientes.
ggplot(toy, aes(IMC, Weight_Kg)) + geom_point(colour = "red") + theme_light()
#Repítelo filtrando sólo los pacientes categorizados como "Overweight" y "Obesity".
toysobrepeso <- toy %>% filter(IMC_clas==c("Overweight", "Obesity"))
```

```
ggplot(toysobrepeso, aes(IMC, Weight_Kg)) + geom_point(colour = "red") + theme_light()
```

```
install.packages("ape")
```

```
install.packages("phangorn")
```

```
install.packages("phytools")
```