

Laboratorio 2

Filtrar y seleccionar datos:

¿Cuántos registros cumplen las condiciones finales?

Teniendo en cuenta solo los dataframes, en el caso de Humans solo 35 objetos cumplen con los requerimientos.

En nodroids cumplen las condiciones 77 objetos.

Si contáramos Humans y Tatooine hubo 8 objetos que cumplieron.

human	35 obs. of 14 variables
starwars_nodroids	77 obs. of 14 variables

```
> starwars %>% filter(species == "Human", homeworld == "Tatooine")
# A tibble: 8 x 14
  name          height mass hair_color skin_color eye_color birth_year sex gender
<chr>         <int> <dbl> <chr>      <chr>      <chr>      <dbl> <chr> <chr>
1 Luke Skywal...   172    77 blond     fair       blue        19  male masculi...
2 Darth Vader     202   136 none      white      yellow     41.9  male masculi...
3 Owen Lars       178   120 brown, gr... light      blue        52  male masculi...
4 Beru Whites...  165    75 brown     light      blue        47  fema... femin...
5 Biggs Darkl...  183    84 black     light      brown       24  male masculi...
6 Anakin Skyw...  188    84 blond     fair       blue       41.9  male masculi...
7 Shmi Skywal...  163    NA black     fair       brown       72  fema... femin...
8 Cliegg Lars     183    NA brown     fair       blue        82  male masculi...
```

Calcular algunos estadísticos:

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

```
starwars %>% group_by(species) %>% summarise(sd_height = sd(height, na.rm = T), sd_mass = sd(mass, na.rm = T))
```

Al crear los gráficos puedes observar que hay un punto que corresponde a un personaje con una masa muy grande. Inspecciona el dataset, filtra usando las funciones de tidyverse, crea un nuevo dataframe sin ese personaje y crea de nuevo el gráfico final. (Exporta el gráfico con la opción exportar en el panel derecho y adjúntalo en el pdf)

Primero cree un dataset con la variable peso y nombre para identificarlo:

```
pesos <- starwars %>% select(name, mass)
```

	name	mass
16	Jabba Desilijic Tiure	1358.0

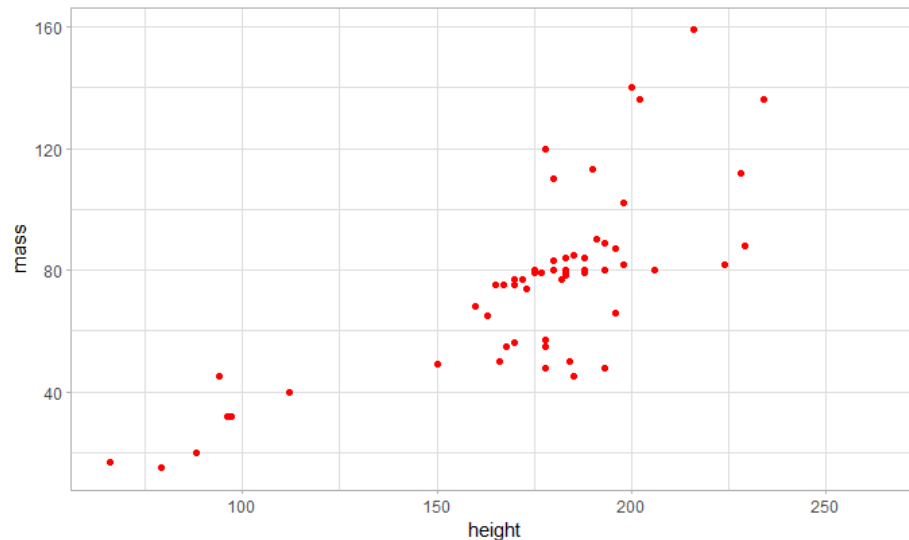
Luego ya se crea un dataset sin este personaje:

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

no_heavy	86 obs. of 14 variables
----------	-------------------------

Se crea el gráfico:

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



Ejercicio:

Descarga el dataset toy.csv cargalo en R studio usando la función read_csv de la libreria tidyverse. Tienes que poner la dirección donde has guardado el archivo descargado. En el ejemplo, el archivo está en la carpeta "Descargas"

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

```
tabla_toy <- toy %>% group_by(Sex) %>% summarise(mean_height = mean(Height_cm, na.rm = T), mean_weight = mean(Weight_Kg, 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_height	mean_weight	mean_imc	mean_ias	mean_ccintura
1	Men	176.4167	78.08333	25.13333	59.59000	81.72727
2	Women	162.3448	63.04310	23.92414	63.51154	80.88571

Haz una tabla sólo con los pacientes femeninos ¿Cuántos registros cumplen las condiciones? ¿De estos cuantos tienen Sobrepeso (Overweight)? Usa select y filter.

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

tabla_femeina 58 obs. of 10 variables

Un total de 58 cumplen las condiciones.

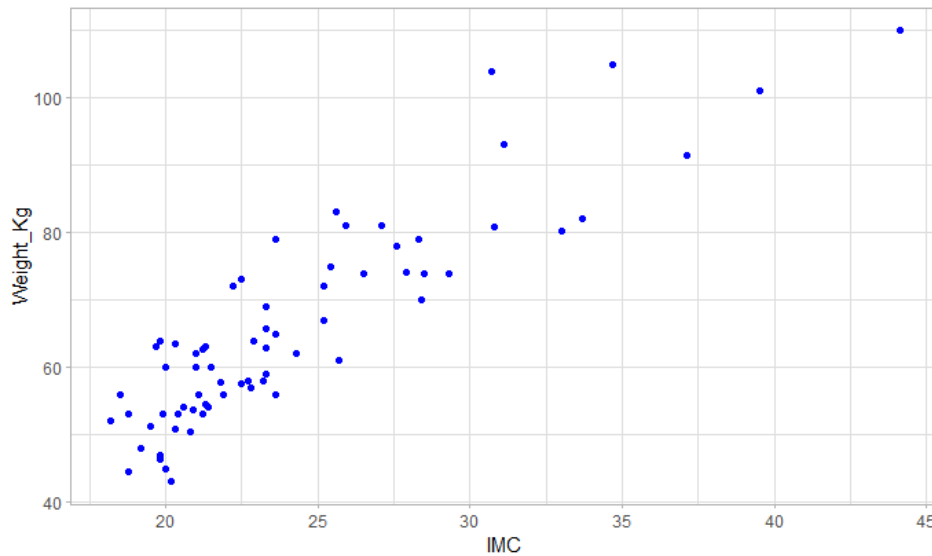
```
women_overweight <- tabla_femeina %>% filter(IMC_clas == "Overweight")
```

women_overweight 9 obs. of 10 variables

Un total de 9 tienen 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 = "blue") + theme_light()
```



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

Primero se crea un dataset con overweight y obesity con el siguiente comando:

```
obesity_overweight <- toy %>% filter(IMC_clas == "Overweight" | IMC_clas == "Obesity")
```

🔍 obesity_overweight 23 obs. of 10 variables

Luego se inserta el comando para el gráfico:

```
ggplot(obesity_overweight, aes(IMC, Weight_Kg)) + geom_point(colour = "orange") + theme_light()
```

