Assignment-1

Amine Ceyda TANDOGAN

21/01/2022

# Q1 : How many characters do have at least one starship? List the names of the characters having at least one starship

## [1] "List the names of the characters having at least one starship:"

## [1] "Luke Skywalker" "Darth Vader" "Biggs Darklighter"  
## [4] "Obi-Wan Kenobi" "Anakin Skywalker" "Chewbacca"   
## [7] "Han Solo" "Wedge Antilles" "Jek Tono Porkins"   
## [10] "Boba Fett" "Lando Calrissian" "Arvel Crynyd"   
## [13] "Nien Nunb" "Ric Olié" "Darth Maul"   
## [16] "Plo Koon" "Gregar Typho" "Grievous"   
## [19] "Poe Dameron" "Padmé Amidala"

# Q2 : Get the frequencies of the eye color of the characters. Rank them from most to least

## # A tibble: 15 x 2  
## eye\_color frequency\_  
## <chr> <int>  
## 1 brown 21  
## 2 blue 19  
## 3 yellow 11  
## 4 black 10  
## 5 orange 8  
## 6 red 5  
## 7 hazel 3  
## 8 unknown 3  
## 9 blue-gray 1  
## 10 dark 1  
## 11 gold 1  
## 12 green, yellow 1  
## 13 pink 1  
## 14 red, blue 1  
## 15 white 1

# Q3 : According to the data available, what are the mean (average) age values across each species? Find the 3 oldest species. Note that in the data, there is no age column, instead there is a birth\_year column. This column represents how many years before the Battle of Yavin the character was born. So take this column as the ages of characters at the Battle of Yavin and calculate the mean.

## # A tibble: 38 x 2  
## species avg  
## <chr> <dbl>  
## 1 Yoda's species 896   
## 2 Hutt 600   
## 3 Wookiee 200   
## 4 Cerean 92   
## 5 <NA> 62   
## 6 Zabrak 54   
## 7 Human 53.4  
## 8 Droid 53.3  
## 9 Trandoshan 53   
## 10 Gungan 52   
## # ... with 28 more rows

## [1] "Three oldest species"

## [1] "Yoda's species" "Hutt" "Wookiee"

# Q4 : Create a new data set by adding a new observation to this data. This observation should be based on your own character (your name or nickname, your weight and height, your homeworld, your starships etc). Note that you can pick one or more than one Star Wars films in which you played as a movie star.

## [1] "The last row of the new data."

## # A tibble: 1 x 14  
## name height mass hair\_color skin\_color eye\_color birth\_year sex gender  
## <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 Amine C~ 167 49 brown white brown 198 femin~ femin~  
## # ... with 5 more variables: homeworld <chr>, species <chr>, films <list>,  
## # vehicles <list>, starships <list>

# Q5 : Calculate the body mass index (BMI) values (dividing the mass value in kg to the square of the height value in meter) for all observations and categorize the observations as underweight (BMI below 18.5), healthy (BMI between 18.5-24.99), overweight (BMI between 25.0-29.99) and obese (BMI above 30.0). Add these two variables to your new data created at the 4th question.

## # A tibble: 88 x 2  
## bmi bmi\_group  
## <dbl> <chr>   
## 1 26.0 obese   
## 2 26.9 obese   
## 3 34.7 obese   
## 4 33.3 obese   
## 5 21.8 healthy   
## 6 37.9 obese   
## 7 27.5 obese   
## 8 34.0 obese   
## 9 25.1 obese   
## 10 23.2 healthy   
## # ... with 78 more rows

# Q6 : Plot the distribution of ages less than 100 by BMI groups. (i.e. use filter function to select the ages less then 100)

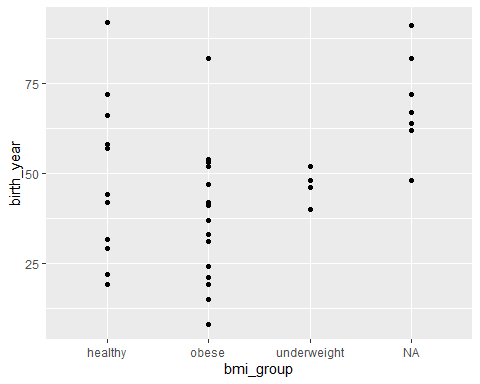


Figure 1. Distribution of ages less than 100 by BMI groups

# Q7 : By plotting a graph, show the relationship between age and BMI values (use point and line at the same time). Re-plot the same graph after filtering the data as both age and BMI less than 100.

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 51 rows containing non-finite values (stat\_smooth).

## Warning: Removed 51 rows containing missing values (geom\_point).

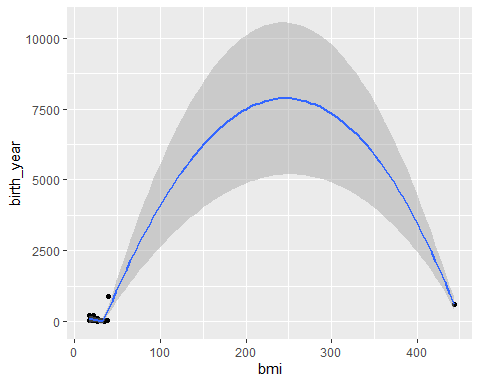


Figure2. Relationship between age and BMI values.

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

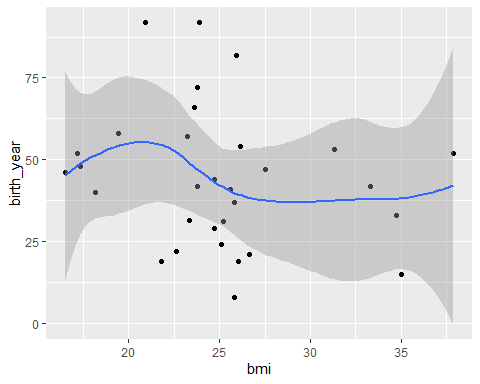


Figure3. The same graph after filtering the data as both age and BMI less than 100.

**Amine Ceyda Tandogan 2018555061 CEN429 morning**