# REpart2

```
if(!require(tidyverse)){
  install.packages("tidyverse", dependencies = TRUE)
  library(tidyverse)
}
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data1 <- read.csv("starwars.csv")
class(data1)
```

```
## [1] "data.frame"
```

```
dim(data1)
```

```
## [1] 87 12
```

```
#I first need to remove the columns with unwanted variables
data2 <- data1[ , ! names(data1) %in% c("hair_color", "eye_color", "skin_color")]

#now to remove rows that are not human
data<-subset(data2, species == "Human")

#Adding new columns
data <- data %>%
  add_column(BMI = (.$mass / (.$height /100) ^2),
             Range = if_else(BMI>24.88711, "ABOVE", "BELOW"))

#Filtering out rows with missing values
any(is.na(data))
```

```
## [1] TRUE
```

```
data <- na.omit(data)

#Finding the median
median(data[,10])
```

```
## [1] 24.88711
```

```
#Export the cleaned dataset
?write.table
```

```r
  write.table(data, file = "Final_Cleaned.csv", sep=",")

print(data)
```

```
##     X               name height  mass birth_year    sex    gender    homeworld
## 1   1     Luke Skywalker    172  77.0       19.0   male masculine     Tatooine
## 4   4        Darth Vader    202 136.0       41.9   male masculine     Tatooine
## 5   5        Leia Organa    150  49.0       19.0 female  feminine     Alderaan
## 6   6          Owen Lars    178 120.0       52.0   male masculine     Tatooine
## 7   7 Beru Whitesun lars    165  75.0       47.0 female  feminine     Tatooine
## 9   9  Biggs Darklighter    183  84.0       24.0   male masculine     Tatooine
## 10 10     Obi-Wan Kenobi    182  77.0       57.0   male masculine      Stewjon
## 11 11   Anakin Skywalker    188  84.0       41.9   male masculine     Tatooine
## 14 14           Han Solo    180  80.0       29.0   male masculine     Corellia
## 17 17     Wedge Antilles    170  77.0       21.0   male masculine     Corellia
## 20 20           Palpatine    170  75.0       82.0   male masculine        Naboo
## 21 21           Boba Fett    183  78.2       31.5   male masculine       Kamino
## 24 24   Lando Calrissian    177  79.0       31.0   male masculine      Socorro
## 25 25               Lobot    175  79.0       37.0   male masculine       Bespin
## 48 48         Mace Windu    188  84.0       72.0   male masculine   Haruun Kal
## 64 64              Dooku    193  80.0      102.0   male masculine      Serenno
## 66 66         Jango Fett    183  79.0       66.0   male masculine Concord Dawn
## 87 87      Padmé Amidala    165  45.0       46.0 female  feminine        Naboo
##    species      BMI Range
## 1    Human 26.02758 ABOVE
## 4    Human 33.33007 ABOVE
## 5    Human 21.77778 BELOW
## 6    Human 37.87401 ABOVE
## 7    Human 27.54821 ABOVE
## 9    Human 25.08286 ABOVE
## 10   Human 23.24598 BELOW
## 11   Human 23.76641 BELOW
## 14   Human 24.69136 BELOW
## 17   Human 26.64360 ABOVE
## 20   Human 25.95156 ABOVE
## 21   Human 23.35095 BELOW
## 24   Human 25.21625 ABOVE
## 25   Human 25.79592 ABOVE
## 48   Human 23.76641 BELOW
## 64   Human 21.47709 BELOW
## 66   Human 23.58984 BELOW
## 87   Human 16.52893 BELOW
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.