Coding Assignment 3

Noah Foilb

9/17/2020

#### Getting Started. Run the next two code chunks to install your libraries and create a dataset called SW that contains data on Starwars characters

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

SW <- starwars  
head(SW, 10)

## # A tibble: 10 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~ 172 77 blond fair blue 19 male mascu~  
## 2 C-3PO 167 75 <NA> gold yellow 112 none mascu~  
## 3 R2-D2 96 32 <NA> white, bl~ red 33 none mascu~  
## 4 Dart~ 202 136 none white yellow 41.9 male mascu~  
## 5 Leia~ 150 49 brown light brown 19 fema~ femin~  
## 6 Owen~ 178 120 brown, gr~ light blue 52 male mascu~  
## 7 Beru~ 165 75 brown light blue 47 fema~ femin~  
## 8 R5-D4 97 32 <NA> white, red red NA none mascu~  
## 9 Bigg~ 183 84 black light brown 24 male mascu~  
## 10 Obi-~ 182 77 auburn, w~ fair blue-gray 57 male mascu~  
## # ... with 5 more variables: homeworld <chr>, species <chr>, films <list>,  
## # vehicles <list>, starships <list>

#### 1) Explore the SW dataset using the glimpse() command and answer the following questions. Recall you explored the dataset with str() in Coding Assignmnet 2. You need to decide which you prefer.

* How many numerical variables are in the dataframe (these would be listed as int for integer or dbl for double precision)?
* How many character variables (these would be listed as chr)?
* How many categorical variables (these would be listed as fct for factor)?
* How many rows?
* How many columns?

#Using Glimpse we get more information about SW  
glimpse(SW)

## Rows: 87  
## Columns: 14  
## $ name <chr> "Luke Skywalker", "C-3PO", "R2-D2", "Darth Vader", "Leia...  
## $ height <int> 172, 167, 96, 202, 150, 178, 165, 97, 183, 182, 188, 180...  
## $ mass <dbl> 77.0, 75.0, 32.0, 136.0, 49.0, 120.0, 75.0, 32.0, 84.0, ...  
## $ hair\_color <chr> "blond", NA, NA, "none", "brown", "brown, grey", "brown"...  
## $ skin\_color <chr> "fair", "gold", "white, blue", "white", "light", "light"...  
## $ eye\_color <chr> "blue", "yellow", "red", "yellow", "brown", "blue", "blu...  
## $ birth\_year <dbl> 19.0, 112.0, 33.0, 41.9, 19.0, 52.0, 47.0, NA, 24.0, 57....  
## $ sex <chr> "male", "none", "none", "male", "female", "male", "femal...  
## $ gender <chr> "masculine", "masculine", "masculine", "masculine", "fem...  
## $ homeworld <chr> "Tatooine", "Tatooine", "Naboo", "Tatooine", "Alderaan",...  
## $ species <chr> "Human", "Droid", "Droid", "Human", "Human", "Human", "H...  
## $ films <list> [<"The Empire Strikes Back", "Revenge of the Sith", "Re...  
## $ vehicles <list> [<"Snowspeeder", "Imperial Speeder Bike">, <>, <>, <>, ...  
## $ starships <list> [<"X-wing", "Imperial shuttle">, <>, <>, "TIE Advanced ...

# There are three numerical variables, Eight Character variables, and No caregorical varaibles (UNless its "List,which there would be three)  
# There are 87 Rows and 14 Columns

#### 2) Use the proper command to create a new dataframe called SW1 that contains only the columns: name, height, mass, hair\_color, skin\_color, eye\_color, birth\_year, sex, gender, homeworld, and species. Display the top 10 rows.

* Use the pipe operator to send the SW dataset to the command you need to choose your columns
* Your new dataframe should contain 11 columns.

#use "<-" to assign to a variable  
SW1 <- SW %>%  
 select(name,height,mass,hair\_color,skin\_color,eye\_color,birth\_year,sex,gender,homeworld,species) # Use select to which variables to keep  
  
head(SW1,10) #Display top ten values

## # A tibble: 10 x 11  
## name height mass hair\_color skin\_color eye\_color birth\_year sex gender  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 Luke~ 172 77 blond fair blue 19 male mascu~  
## 2 C-3PO 167 75 <NA> gold yellow 112 none mascu~  
## 3 R2-D2 96 32 <NA> white, bl~ red 33 none mascu~  
## 4 Dart~ 202 136 none white yellow 41.9 male mascu~  
## 5 Leia~ 150 49 brown light brown 19 fema~ femin~  
## 6 Owen~ 178 120 brown, gr~ light blue 52 male mascu~  
## 7 Beru~ 165 75 brown light blue 47 fema~ femin~  
## 8 R5-D4 97 32 <NA> white, red red NA none mascu~  
## 9 Bigg~ 183 84 black light brown 24 male mascu~  
## 10 Obi-~ 182 77 auburn, w~ fair blue-gray 57 male mascu~  
## # ... with 2 more variables: homeworld <chr>, species <chr>

#### 3) Starting with SW1, use the proper command to create a new dataframe called SWDroid that only contains characters that are Droids (in other words, the species is Droid). Display the entire dataframe.

* Your new dataframe should contain 6 rows.

SWDroid <- SW1 %>%  
 filter(sex == 'none')  
  
SWDroid

## # A tibble: 6 x 11  
## name height mass hair\_color skin\_color eye\_color birth\_year sex gender  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 C-3PO 167 75 <NA> gold yellow 112 none mascu~  
## 2 R2-D2 96 32 <NA> white, bl~ red 33 none mascu~  
## 3 R5-D4 97 32 <NA> white, red red NA none mascu~  
## 4 IG-88 200 140 none metal red 15 none mascu~  
## 5 R4-P~ 96 NA none silver, r~ red, blue NA none femin~  
## 6 BB8 NA NA none none black NA none mascu~  
## # ... with 2 more variables: homeworld <chr>, species <chr>

#### 4) Starting with SW1, use the proper commands to create a dataframe called SWred containing the names, homeworlds, and species of all characters with red eyes. Display the entire dataframe.

* Your new dataframe should contain 5 rows.
* If you are having trouble, make sure your commands are in the proper order.

SWred <- SW1 %>%  
 filter(eye\_color == "red" ) %>%  
 select(name,homeworld,species)  
  
SWred

## # A tibble: 5 x 3  
## name homeworld species   
## <chr> <chr> <chr>   
## 1 R2-D2 Naboo Droid   
## 2 R5-D4 Tatooine Droid   
## 3 IG-88 <NA> Droid   
## 4 Bossk Trandosha Trandoshan  
## 5 Nute Gunray Cato Neimoidia Neimodian

#### 5) Starting with SW1, use the proper commands to create a dataframe called SWyellow containing the names, eye color, skin color, homeworlds, and species of all characters with yellow eyes OR yellow skin. Display the entire dataframe.

* Your new dataframe should contain 13 rows.

SWyellow <- SW1 %>%  
 filter(eye\_color == "yellow" | skin\_color == "yellow" ) %>%  
 select(name,eye\_color, skin\_color, homeworld, species)  
  
SWyellow

## # A tibble: 13 x 5  
## name eye\_color skin\_color homeworld species   
## <chr> <chr> <chr> <chr> <chr>   
## 1 C-3PO yellow gold Tatooine Droid   
## 2 Darth Vader yellow white Tatooine Human   
## 3 Palpatine yellow pale Naboo Human   
## 4 Watto yellow blue, grey Toydaria Toydarian   
## 5 Darth Maul yellow red Dathomir Zabrak   
## 6 Dud Bolt yellow blue, grey Vulpter Vulptereen  
## 7 Ki-Adi-Mundi yellow pale Cerea Cerean   
## 8 Yarael Poof yellow white Quermia Quermian   
## 9 Poggle the Lesser yellow green Geonosis Geonosian   
## 10 Luminara Unduli blue yellow Mirial Mirialan   
## 11 Barriss Offee blue yellow Mirial Mirialan   
## 12 Zam Wesell yellow fair, green, yellow Zolan Clawdite   
## 13 Dexter Jettster yellow brown Ojom Besalisk

#### 6) Make a two-way contingency table of hair and eye color for all characters in the SW1 dataframe.

* What is/are the most common combination(s)?

table(SW1$hair\_color,SW1$eye\_color)

##   
## black blue blue-gray brown dark gold green, yellow hazel orange  
## auburn 0 1 0 0 0 0 0 0 0  
## auburn, grey 0 1 0 0 0 0 0 0 0  
## auburn, white 0 0 1 0 0 0 0 0 0  
## black 0 2 0 9 1 0 0 0 0  
## blond 0 3 0 0 0 0 0 0 0  
## blonde 0 0 0 0 0 0 0 0 0  
## brown 0 7 0 9 0 0 0 2 0  
## brown, grey 0 1 0 0 0 0 0 0 0  
## grey 0 0 0 0 0 0 0 0 0  
## none 9 3 0 1 0 1 1 1 7  
## unknown 0 0 0 0 0 0 0 0 0  
## white 0 1 0 2 0 0 0 0 0  
##   
## pink red red, blue unknown white yellow  
## auburn 0 0 0 0 0 0  
## auburn, grey 0 0 0 0 0 0  
## auburn, white 0 0 0 0 0 0  
## black 0 0 0 0 0 1  
## blond 0 0 0 0 0 0  
## blonde 0 0 0 0 0 1  
## brown 0 0 0 0 0 0  
## brown, grey 0 0 0 0 0 0  
## grey 0 0 0 0 0 1  
## none 1 3 1 2 1 6  
## unknown 0 0 0 1 0 0  
## white 0 0 0 0 0 1

# there is 9 people with no eye color and black hair. 9 with brown eyes and black hair, 9 people with brown eyes and brown hair.

#### 7) Staring with SW1, create a new dataframe called SWage containing only the names and birth years of all of the characters from oldest to youngest. List the 20 oldest characters.

* The birth year is listed as the BBY = Before Battle of Yavin, so the oldest character (Yoda) has the largest birth year and the youngest character has the smallest birth year.

SWage <- SW1 %>%  
 select(name,birth\_year) %>%  
 arrange(desc(birth\_year))  
  
head(SWage,20)

## # A tibble: 20 x 2  
## name birth\_year  
## <chr> <dbl>  
## 1 Yoda 896  
## 2 Jabba Desilijic Tiure 600  
## 3 Chewbacca 200  
## 4 C-3PO 112  
## 5 Dooku 102  
## 6 Qui-Gon Jinn 92  
## 7 Ki-Adi-Mundi 92  
## 8 Finis Valorum 91  
## 9 Palpatine 82  
## 10 Cliegg Lars 82  
## 11 Shmi Skywalker 72  
## 12 Mace Windu 72  
## 13 Bail Prestor Organa 67  
## 14 Jango Fett 66  
## 15 Wilhuff Tarkin 64  
## 16 Quarsh Panaka 62  
## 17 Luminara Unduli 58  
## 18 Obi-Wan Kenobi 57  
## 19 Darth Maul 54  
## 20 Bossk 53

#### 8) Starting with SW1, create a new dataframe called SWtall that lists the name, height, and species for all characters over 200 cm in height and shows the tallest character first.

* Your final dataset should have 10 rows showing Yarael Poof is the tallest.

SWtall <- SW1 %>%  
 filter(height>200) %>%  
 select(name,height,species) %>%  
 arrange(desc(height))  
  
SWtall

## # A tibble: 10 x 3  
## name height species   
## <chr> <int> <chr>   
## 1 Yarael Poof 264 Quermian  
## 2 Tarfful 234 Wookiee   
## 3 Lama Su 229 Kaminoan  
## 4 Chewbacca 228 Wookiee   
## 5 Roos Tarpals 224 Gungan   
## 6 Grievous 216 Kaleesh   
## 7 Taun We 213 Kaminoan  
## 8 Rugor Nass 206 Gungan   
## 9 Tion Medon 206 Pau'an   
## 10 Darth Vader 202 Human

#### 9) Starting with SW1, create a new dataframe called SWstats that includes each character’s name, height, mass, and species and includes a new column called BMI (body mass index), where BMI is mass in kg divided by the square of height in meters.

* BMI=mass/(height/100)^2
* Display the top six rows
* If done correctly, Luke Skywalker’s BMI is 26.02758

SWstats <- SW1 %>%  
 mutate(name,height,mass,species,BMI = (mass/(height/100)\*\*2))  
  
SWstats

## # A tibble: 87 x 12  
## name height mass hair\_color skin\_color eye\_color birth\_year sex gender  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 Luke~ 172 77 blond fair blue 19 male mascu~  
## 2 C-3PO 167 75 <NA> gold yellow 112 none mascu~  
## 3 R2-D2 96 32 <NA> white, bl~ red 33 none mascu~  
## 4 Dart~ 202 136 none white yellow 41.9 male mascu~  
## 5 Leia~ 150 49 brown light brown 19 fema~ femin~  
## 6 Owen~ 178 120 brown, gr~ light blue 52 male mascu~  
## 7 Beru~ 165 75 brown light blue 47 fema~ femin~  
## 8 R5-D4 97 32 <NA> white, red red NA none mascu~  
## 9 Bigg~ 183 84 black light brown 24 male mascu~  
## 10 Obi-~ 182 77 auburn, w~ fair blue-gray 57 male mascu~  
## # ... with 77 more rows, and 3 more variables: homeworld <chr>, species <chr>,  
## # BMI <dbl>

#### 10) Round the BMI column in SWstats to 1 decimal place and display the first 6 rows of SWstats.

SWstats$BMI <- round(SWstats$BMI,1)  
head(SWstats)

## # A tibble: 6 x 12  
## name height mass hair\_color skin\_color eye\_color birth\_year sex gender  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 Luke~ 172 77 blond fair blue 19 male mascu~  
## 2 C-3PO 167 75 <NA> gold yellow 112 none mascu~  
## 3 R2-D2 96 32 <NA> white, bl~ red 33 none mascu~  
## 4 Dart~ 202 136 none white yellow 41.9 male mascu~  
## 5 Leia~ 150 49 brown light brown 19 fema~ femin~  
## 6 Owen~ 178 120 brown, gr~ light blue 52 male mascu~  
## # ... with 3 more variables: homeworld <chr>, species <chr>, BMI <dbl>

#### 11) Staring with SWstats, create a new dataframe called SWstats2 which:

* reorders the columns so that the display order is name, height, mass, BMI, species (hint: use the select() command)
* renames the column called mass to weight,
* and arranges the dataset by decreasing BMI.
* Display the top 10 rows. Jabba the Hut should be at the top!

SWstats2 <- SWstats %>%  
 select(name,height,mass,BMI,species) %>%  
 rename(Weight = mass) %>%  
 arrange(desc(BMI))   
  
head(SWstats2)

## # A tibble: 6 x 5  
## name height Weight BMI species   
## <chr> <int> <dbl> <dbl> <chr>   
## 1 Jabba Desilijic Tiure 175 1358 443. Hutt   
## 2 Dud Bolt 94 45 50.9 Vulptereen   
## 3 Yoda 66 17 39 Yoda's species  
## 4 Owen Lars 178 120 37.9 Human   
## 5 IG-88 200 140 35 Droid   
## 6 R2-D2 96 32 34.7 Droid

#### 12) Create a new column in the SWstats2 dataset called bmi\_rating where:

* characters with a BMI>100 are rated as Obese,
* characters with a BMI between 30 and 100 are rated as Heavy, and
* characters with a BMI of 30 or below are rated as Normal.
* Display the top 15 rows.
* If done correctly, Jabba the Hut will be rated Obese, Dud Bolt will be the first character rated as Heavy, and Beru Whitesun lars the first rated as Normal.

SWstats2$bmi\_rating <- if\_else(SWstats2$BMI>100, "Obese",  
 if\_else(SWstats2$BMI>=30, "Heavy","Normal"))  
  
head(SWstats2,15)

## # A tibble: 15 x 6  
## name height Weight BMI species bmi\_rating  
## <chr> <int> <dbl> <dbl> <chr> <chr>   
## 1 Jabba Desilijic Tiure 175 1358 443. Hutt Obese   
## 2 Dud Bolt 94 45 50.9 Vulptereen Heavy   
## 3 Yoda 66 17 39 Yoda's species Heavy   
## 4 Owen Lars 178 120 37.9 Human Heavy   
## 5 IG-88 200 140 35 Droid Heavy   
## 6 R2-D2 96 32 34.7 Droid Heavy   
## 7 Grievous 216 159 34.1 Kaleesh Heavy   
## 8 R5-D4 97 32 34 Droid Heavy   
## 9 Jek Tono Porkins 180 110 34 Human Heavy   
## 10 Darth Vader 202 136 33.3 Human Heavy   
## 11 Sebulba 112 40 31.9 Dug Heavy   
## 12 Bossk 190 113 31.3 Trandoshan Heavy   
## 13 Beru Whitesun lars 165 75 27.5 Human Normal   
## 14 C-3PO 167 75 26.9 Droid Normal   
## 15 Wedge Antilles 170 77 26.6 Human Normal