

# Homework 3

Aditi Madhok

2/8/21

We will continue working with the `starwars` dataset that is loaded with tidyverse.

```
library("tidyverse")
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.0.6      v dplyr  1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
starwars
```

```
## # A tibble: 87 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 masculin
## 2 C-3P0   167    75 <NA>      gold        yellow       112 none masculin
## 3 R2-D2    96    32 <NA>      white, bl~ red         33 none masculin
## 4 Dart~   202   136 none      white      yellow       41.9 male masculin
## 5 Leia~   150    49 brown     light      brown        19 fema~ feminin
## 6 Owen~   178   120 brown, gr~ light      blue        52 male masculin
## 7 Beru~   165    75 brown     light      blue        47 fema~ feminin
## 8 R5-D4    97    32 <NA>      white, red red         NA none masculin
## 9 Bigg~   183    84 black     light      brown        24 male masculin
## 10 Obi~-   182    77 auburn, w~ fair      blue-gray    57 male masculin
## # ... with 77 more rows, and 5 more variables: homeworld <chr>, species <chr>,
## #   films <list>, vehicles <list>, starships <list>
```

USE PIPES TO ANSWER EVERY QUESTION IN THIS ASSIGNMENT.

1. Create a tibble `bio` that contains the name, height and mass for each character in Star Wars. Only include rows where the height and mass are defined (not NA). Add a column called `ratio` which is the ratio of height to mass, and order your rows so that this new column is increasing.

```
bio <- starwars %>%
  select (name,height,mass) %>%
  filter(!is.na(height), !is.na(mass))%>%
  mutate(ratio=(height/mass)) %>%
  arrange(ratio)
bio
```

```
## # A tibble: 59 x 4
##   name                height mass ratio
##   <chr>              <int> <dbl> <dbl>
## 1 Jabba Desilijic Tiure    175  1358 0.129
## 2 Grievous                216   159 1.36
## 3 IG-88                   200   140 1.43
## 4 Owen Lars               178   120 1.48
## 5 Darth Vader             202   136 1.49
## 6 Jek Tono Porkins        180   110 1.64
## 7 Bossk                   190   113 1.68
## 8 Tarfful                 234   136 1.72
## 9 Dexter Jettster         198   102 1.94
## 10 Chewbacca              228   112 2.04
## # ... with 49 more rows
```

2. What are the most common eye colors among Star Wars characters? To answer this, create a tibble `eyes` with columns `eye_color` and `number` that gives the number of characters with each eye color, sorted so that the most common eye color is first.

```
eyes<-starwars %>%
  group_by(eye_color) %>%
  summarize(number=n()) %>%
  arrange(desc (number))
eyes
```

```
## # A tibble: 15 x 2
##   eye_color      number
##   <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
```

3. Create a tibble `sex_mass` with columns `sex`, `min_mass`, `max_mass`, and `mass_diff`. The `min_mass` and `max_mass` columns should be the minimum and maximum mass among all characters of the same

sex (watch out for missing mass values). The `mass_diff` column should give the difference of these two values. Remove rows with NA or NaN values for the sex (your tibble should have 4 rows). Order rows in decreasing order of `mass_diff`.

```
sex_mass <-starwars %>%
  group_by(sex)%>%
  filter(!is.na (sex))%>%
  summarize(min_mass = min(mass, na.rm = TRUE),max_mass = max(mass, na.rm = TRUE), mass_diff=(max_mass-min_mass))
  arrange(desc(mass_diff))
sex_mass
```

```
## # A tibble: 4 x 4
##   sex          min_mass max_mass mass_diff
##   <chr>          <dbl>   <dbl>   <dbl>
## 1 male             15     159     144
## 2 none             32     140     108
## 3 female           45      75      30
## 4 hermaphroditic 1358    1358      0
```

4. Create a tibble called `brown` with columns `homeworld` and `all_brown`. The `all_brown` column should be a boolean variable (TRUE or FALSE), with a value that tells you if every character from that homeworld has brown eyes.

```
brown <-starwars %>%
  group_by(homeworld) %>%
  summarize(all_brown=all(eye_color== "brown"))
brown
```

```
## # A tibble: 49 x 2
##   homeworld    all_brown
##   * <chr>      <lgl>
## 1 Alderaan    TRUE
## 2 Aleen Minor FALSE
## 3 Bespin      FALSE
## 4 Bestine IV  FALSE
## 5 Cato Neimoidia FALSE
## 6 Cerea       FALSE
## 7 Champala    FALSE
## 8 Chandrila   FALSE
## 9 Concord Dawn TRUE
## 10 Corellia   FALSE
## # ... with 39 more rows
```

5. Create a tibble called `Human_worlds` with columns `homeworld` and `number`. The `number` column should show how many humans were from each homeworld. DO NOT list any world for which only 1 Human calls home. DO list how many humans were from unknown homeworlds (`homeworld == NA`).

```
Human_worlds <- starwars %>%
  group_by(homeworld) %>%
  filter(species == "Human") %>%
  summarize(number=sum(species == "Human")) %>%
  filter(number>1)
Human_worlds
```

```
## # A tibble: 6 x 2
##   homeworld number
##   <chr>      <int>
## 1 Alderaan      3
## 2 Coreellia      2
## 3 Coruscant      2
## 4 Naboo          5
## 5 Tatooine       8
## 6 <NA>           5
```

6. Create a tibble `tallest` that has columns `name` and `homeworld` that gives the name of the tallest character from each homeworld. (Watch out for NA heights.) Sort your table so that the tallest characters in your tibble are listed first. (Hint: someone who is the tallest from their homeworld has a height that is the maximum among all heights from that world.)

```
tallest <- starwars %>%
  group_by(homeworld) %>%
  filter(rank(desc(height), ties.method = "first") == 1) %>%
  arrange(desc(height)) %>%
  select(name, height)
```

```
## Adding missing grouping variables: 'homeworld'
```

```
tallest
```

```
## # A tibble: 49 x 3
## # Groups:   homeworld [49]
##   homeworld name      height
##   <chr>      <chr>      <int>
## 1 Quermia    Yarael Poof      264
## 2 Kashyyyk   Tarfful          234
## 3 Kamino     Lama Su          229
## 4 Naboo      Roos Tarpals     224
## 5 Kalee      Grievous         216
## 6 Utapau     Tion Medon       206
## 7 Tatooine   Darth Vader      202
## 8 <NA>       IG-88            200
## 9 Cerea      Ki-Adi-Mundi     198
## 10 Ojom      Dexter Jettster   198
## # ... with 39 more rows
```

7. Create a tibble called `tallest2` with columns `name` and `homeworld` that gives the name of the second tallest character from each homeworld. Sort your table so that the tallest characters in your tibble are listed first.

```
tallest2 <- starwars %>%
  group_by(homeworld) %>%
  filter(rank(desc(height)) == 2) %>%
  arrange(desc(height)) %>%
  select(name)
```

```
## Adding missing grouping variables: 'homeworld'
```

```
tallest2
```

```
## # A tibble: 10 x 2
## # Groups:   homeworld [10]
##   homeworld name
##   <chr>      <chr>
## 1 Kashyyyk   Chewbacca
## 2 Kamino     Taun We
## 3 Naboo      Rugor Nass
## 4 <NA>       Qui-Gon Jinn
## 5 Tatooine   Anakin Skywalker
## 6 Alderaan   Raymus Antilles
## 7 Ryloth     Ayla Secura
## 8 Corellia   Wedge Antilles
## 9 Coruscant  Finis Valorum
## 10 Mirial    Barriss Offee
```

8. Create a tibble called `shortest3_avg` with columns `homeworld` and `avg` that gives the average of the heights of the three shortest characters from each homeworld. Only include homeworlds with at least three characters. Do not include unknown homeworlds (`homeworld==NA`). Sort your table so that the shortest averages are listed first.

```
shortest3_avg<- starwars %>%
  group_by(homeworld) %>%
  filter(!is.na(homeworld)) %>%
  slice(height, n=3) %>%
  summarize(avg=mean(height)) %>%
  arrange(avg) %>%
  select(homeworld, avg)
shortest3_avg
```

```
## # A tibble: 5 x 2
##   homeworld avg
##   <chr>      <dbl>
## 1 Coruscant  167
## 2 Alderaan   188
## 3 Naboo      196
## 4 Tatooine   202
## 5 Kamino     213
```