Assignment 03

This homework will help you to practice and self assess your knowledge and ability around the course concepts. Use the answer key to assess your work.

1. Load the **tidyverse** package.
2. Review of `select` & `mutate`:
   A. Drop the last three columns of `starwars`.
   B. Convert cm to inches (`cm x 0.3937007874`) and kg to pounds (`kg x 2.2046226218`) and add columns for height and weight that are in inches and pounds.
   C. Add a BMI column [`(lbs/(in^2)) * 703`] based on lbs. and in.
   D. Use the table below to add a column to the previous output bucketing the BMIs into risk levels (hint: use `case_when()`).
   E. Assign to a `starwars2` object. We will use this for the rest of the assignment.

Risk Level	Range
Underweight	less than 18.5
Normal Weight	18.5 to less than 25
Overweight	25 to less than 30
Obese	30 or greater

3. What is the mean (`mean`), standard deviation (`sd`), median (`median`), maximum (`max`), minimum (`min`), & count (`length`) for the `height` (cm) variable? Remember to set `na.rm = TRUE` on the math functions (e.g., `mean(height, na.rm = TRUE)`). Try it without setting `na.rm = TRUE`. What happens?
4. Filter out where `height` (cm) is missing and then determine what is the mean (`mean`), standard deviation (`sd`), median (`median`), maximum (`max`), minimum (`min`), & count (`length`) for the `height` (cm) variable? Do you get the same values as above?
5. What is the mean (`mean`), standard deviation (`sd`), median (`median`), maximum (`max`), minimum (`min`), and count for the `bmi` variable within species using those that BMI could be calculated for. Order from highest to lowest average BMI. What species has the worst BMI? Does that surprise you?
6. Remove those summaries with a count of 1 in the analysis in #5 above.
7. What is the proportion of those missing a value for `birth_year` within each `risk_level` group? Order by proportion missing. Is the proportion missing birth year equal across risk levels?
8. What is the distribution of eye colors across hair colors? In other words, you should be able to determine how many brown haired people have blue eyes. Remove any groups with a count of 1 and order within hair color by descending count.
9. Use the code from #8 and find the most dominant eye color within each hair color (hint: use `mutate` with `max` to create a new column and `filter` on that column).
10. This sparks a question. Who are the characters with a `hair_color` of `none`?
11. Going back to the `starwars2` dataset: What is the percent breakdown of the `gender` groups? Order by by percent. Considering the gender breakdown in the U.S., does the casting seem fair? (hint: this likely requires a `mutate` after the summary)