553.283 Introduction to R
Homework 3
Owing to the length of this problem set, it is due by 1:30PM on Friday, 17 January, 2020
Note 1: If a question asks you for a numerical answer, your submission for that question must consist of the R command that produces that answer followed immediately by the output.
Note 2: Please label all axes on any plots you create.

1. The dataset *starwars* (*dplyr*) contains information on 87 Star Wars movie characters. One's Body Mass Index (BMI) is defined as their weight in kg divided by (height in meters)^2. Using *dplyr* functions (but without using the piping operator %>%), create a tibble that displays only the names and BMIs of those characters whose BMI exceeds 30, with the characters listed in descending order of BMI.

   Hint: Be sure to inspect the dataset by viewing it in the console and using the command *?starwars*.

2. Perform the same task as above, but use the piping operator %>% to do it all in one command. Which *Star Wars* character has the highest BMI?

3. The dataset *iris* contains the petal and sepal lengths and widths for three species of iris flowers. This dataset is merely a data frame, not a tibble. Using the function *as_tibble()*, save *iris* as a tibble, then use *dplyr* functions to produce a tibble that outputs the mean sepal length and width for each of the three species.

4. (Same dataset as the previous problem.) Use *ggplot2* to create a scatterplot that plots sepal length versus sepal width, with differently colored points for each species.

5. (Same dataset as the previous two problems.) Use both *dplyr* and *ggplot2* to produce side-by-side boxplots of the ratios of sepal length to sepal width for each of the three

species.

6. Download the file *Popular_Baby_Names.csv* from the course webpage, and load it into R using the *read.csv()* function (which works similarly to *read.table()*). This dataset contains information on the number of US-born babies given each name each year from 2011 to 2016.

   Use *dplyr* to produce a tibble that only includes names with counts greater than 1000 in the year 2011, ranked in descending order of frequency. Then use *ggplot2* to produce a barplot where the heights of the bars are the respective frequencies of the same subset of names. What baby name was most popular in 2011?

   Hint: Note that each count is for a specific name in a specific ethnicity. Some names are common in multiple ethnicities!