Stat Computing - Exercises 03 - Billboard Charts

On canvas, there is a file called `Hot_100.csv`, which contains data from the Billboard 100 music charts going back to 1958. Below are some exercises on functions and subsetting that make use of the data. For all questions, include 3 examples to demonstrate calling your function with different arguments.

1. Write a function that takes in a character string indicating a date, an optional format argument that has a default date format. The function should return the #1 song and its performer during the week that contains the date argument. Assume that the date in the dataset is the date of the beginning of the week. Your function should use the default date format if a format is not provided, and it should use the user-supplied format if it is supplied. Test your function with a few different dates and different formats.

2. Modify your function to add a `top` argument, indicating that the function should return the top `top` songs on the specified date. Make sure that the returned songs are listed in order according to the weekly ranking. Test your function by printing the top 10 songs on the day you were born.

3. Write a function to search for matching performer names. The function should take in a character string as its argument, find all unique performer names that contain the argument, and return them. Include an argument to determine whether or not to do exact matching on uppercase vs. lowercase. It may be a good idea to look at the documentation `grep`, `tolower`, and `toupper`.

4. Write a function to take in a performer name, and return all of the songs belonging to that performer that appeared on the Billboard list. Use exact matching for performer names.

5. Song statistics: write a function that takes in a song id, and returns the following statistics for the song: number of weeks in the top 100, number of weeks in the top 40, number of weeks in the top 10, number of weeks at #1, lowest position (i.e. best position). The return type should be a data frame with columns for the song name, the performer, the first week it appeared on the charts, and all of the statistics.

6. Song rankings: write a function that ranks all of the songs according to the statistics in the previous question. The function should take a statistic name as its argument, and return a data frame whose first column is the song name, second column is the performer, third column is the week the song first appeared, and the fourth column is the statistic. The data frame should be ordered with the best value of the statistic at the top of the data frame. Print out the rankings of the top 10 songs for 3 of the statistics.

7. (Optional) Artist statistics: write a function that takes in a performer name, and returns the following statistics for the artists: number of unique songs on the Billboard chart, number of songs in the top 40, number of songs in the top 10, number of songs at #1, number of weeks with a song at #1.

8. (Optional) Artist rankings: write a function that ranks all of the artists according to your statistics in the previous question. The function should take a statistic name as its argument, and return a data frame whose first column is the artist, second column is the statistic. The data frame should be ordered with the best value of the statistic at the top of the data frame. Print out the rankings of the top 10 artists for 3 of the statistics.