CS544 Module 6 Assignment

General Rules for Homework Assignments

- You are strongly encouraged to add comments for the code portions. Doing so will help your facilitator to understand your programming logic and grade you more accurately.
- You must work on your assignments individually. You are not allowed to copy the answers from the others.
- Each assignment has a strict deadline. However, you are still allowed to submit your assignment within 2 days after the deadline with a penalty. 15% of the credit will be deducted unless you made previous arrangements with your facilitator and professor. Assignments submitted 2 days after the deadline would not be graded.
- When the term *lastName* is referenced in an assignment, please replace it with your last name.

Part1) Strings (60 points)

Use the ***stringr*** functions for the following:

Initialize the vector of words from Lincoln's Gettysburg address with the following code:

```
file <- "http://kalathur.com/cs544/data/lincoln.txt"
words <- scan(file, what=character())
```

a) Detect and show all the words that have a punctuation symbol.
b) Replace all the punctuations in the corresponding words with an empty string. Make this the new ***words*** data.
c) Show the frequencies of the word lengths in the above data. Plot the distribution of these frequencies.
d) What are the words with the longest length?
e) Show all the words that start with the letter ***p***.
f) Show all the words that end with the letter **r.**
g) Show all the words that start with the letter ***p*** and end with the letter ***r***.

Part2) Data Wrangling (40 points)

Use the ***tidyverse*** library for the following:

Download the following csv file,
http://people.bu.edu/kalathur/usa_daily_avg_temps.csv
locally first and use read.csv to load the data into a data frame.

a) Convert the data frame into a tibble and assign it to the variable *usaDailyTemps*.
b) What are the maximum temperatures recorded for each year? Show the values and also the appropriate plot for the results.
c) What are the maximum temperatures recorded for each state? Show the values and also the appropriate plot for the results.
d) Filter the Boston data and assign it to the variable *bostonDailyTemps*.
e) What are the average monthly temperatures for Boston? Show the values and also the appropriate plot for the results.

Submission:

Create a folder, CS544_HW6_lastName and place the following file in this folder.

Provide the R code, **HW6_lastName.R**, with each portion of the code clearly identified by the corresponding question. Prepare a corresponding word document by pasting the output for each question (**HW6_lastName.docx)**

Archive the folder (CS544_HW6_lastName.zip). Upload the zip file to the Assignments section of Blackboard.