GBA 464 – Assignment #3
Due Thursday, Feb 25 @ 10:00 AM

DATA DESCRIPTION:
Raw data file: 175 files named 001.csv through 175.csv
Raw data description: call log data for 175 call center employees in one of 10 locations.
***The raw data files are saved in a zip file on Blackboard. You will need to unzip the file and create a directory containing the 175 files somewhere on your pc.

Field	Description
date	Date
id	Employee ID # (matches the name of the file)
totalCalls	Number of total calls placed on day
answeredCalls	Number of calls that were answered
sales	Number of calls that resulted in sale

***Note: each call log contains daily data from 01-01-2015 to 12-31-2017. There are many empty cells in the data (in other words, call center employees do not work on every day).

A file call empData.csv is also supplied. This file is a cross reference table for the location (city and state) of each call center employee by ID number.

ASSIGNMENT INSTRUCTIONS: Please include the R output that answers the questions posed on the last page of the assignment. Separately, please include a .pdf of your (cleaned up) R code for the two functions callMeans.R and callVolume.R

Function 1: callMeans.R

Write a function to summarize the statistics (mean, std. dev, max and min) for a specified subset of call types (totalCalls, answeredCalls and/or sales) for a specified subset of id numbers.

Inputs: The function should accept two arguments – callTypes and id

- callTypes: a character vector indicating one or more of the types of call (totalCalls, answeredCalls, sales) to summarize. "All" may be entered, in which case all three call types are summarized.
- id: an integer vector indicating the employee id numbers to be included in the summary

Output: The function should return a dataframe that has a row for each callType specified and a column for each summary statistic (mean, std. dev., max and min). See sample output for more details.

- The function should verify that all of the elements of the callTypes argument are valid entries and return an error message if this condition is not met.

**Note: for full credit on this question, your functions should only read the .csv files necessary to create the output for the argument inputs. If you combine the files ahead of time/outside of the functions you will lose some points.

```
callMeans <- function(callTypes = "all", id = 1:175){
        ## Check for valid input
        ## Read all .csv files associated with id numbers included in the id argument
        ## Calculate summary statistics for all callTypes included in the callTypes argument
        ## Return dataframe of summary
}
```

Sample output for callMeans:

```
> callMeans(callTypes = "all", id =  1:5)
       callType mean   sd max min
1    totalCalls 47.5 16.4  90  16
2 answeredCalls 18.1  9.7  50   2
3         sales  4.3  3.4  20   0
> callMeans(callTypes = c("totalCalls", "sales"), id = 100:125)
    callType mean   sd max min
1 totalCalls 47.0 16.1  89  15
2      sales  4.1  3.0  20   0
> callMeans(callTypes = c("Calls"), id = 1:175)
[1] "invalid call type"
NULL
> callMeans(callTypes = c("answeredCalls"), id = c(2,6,17,29,100))
       callType mean  sd max min
1 answeredCalls 20.3 9.9  51   3
```

Function 2: callVolume.R

Write a function to summarize the statistics on employee productivity for a specified subset of locations (states).

Inputs: The function should accept two arguments –empData and state

- empData: the empData dataframe loaded from empData.csv
- state: a character vector indicating one or more states to summarize. "All" may be entered, in which case all states are summarized.

Output: The function should return a dataframe that has a row for each city in each of the states specified in the state argument and columns containing the following summary statistics.

- Employees: the total number of employees for the city
- totalCalls: the total number of "totalCalls" for all employees in the city
- totalDays: the total number of days on which calls were placed by all employees in the city
- avgCallsPerEmployee: totalCalls / # of Employees
- avgDaysPerEmployee: totalDays / # of Employees
- avgCallsPerDay: totalCalls / totalDays

```
callVolume <- function(empData, state = "all"){
        ##Check for at least one valid state entered
        ##Read all .csv files associated with employee ids in all states specified in argument
        ## Calculate summary statistics and return dataframe of summary
}
```

Sample output for callVolume:

```
> callVolume(empData = emp, state = c("TX", "PA", "GA"))
         city state employees totalCalls totalDays avgCallsPerEmployee avgDaysPerEmployee avgCallsPerDay
1     Atlanta    GA        22     386261     11868               17557               539           32.5
2      Dallas    TX        24     555505     10871               23146               453           51.1
3     Houston    TX        15     246259      4228               16417               282           58.2
4  Pittsburgh    PA        16     331153      7142               20697               446           46.4
5 San Antonio    TX        14     217842      3604               15560               257           60.4
> callVolume(empData = emp, state = "FL")
     city state employees totalCalls totalDays avgCallsPerEmployee avgDaysPerEmployee avgCallsPerDay
1 Orlando    FL        16     415081      8393               25943               525           49.5
2   Tampa    FL        26     475919      9293               18305               357           51.2
> callVolume(empData = emp, state = "NY")
[1] "no valid states entered"
NULL
> callVolume(empData = emp)
          city state employees totalCalls totalDays avgCallsPerEmployee avgDaysPerEmployee avgCallsPerDay
1      Atlanta    GA        22     386261     11868               17557               539           32.5
2    Charlotte    NC        16     273781      5559               17111               347           49.3
3       Dallas    TX        24     555505     10871               23146               453           51.1
4      Houston    TX        15     246259      4228               16417               282           58.2
5      Orlando    FL        16     415081      8393               25943               525           49.5
6      Phoenix    AZ        14     244276      6392               17448               457           38.2
7   Pittsburgh    PA        16     331153      7142               20697               446           46.4
8  San Antonio    TX        14     217842      3604               15560               257           60.4
9     St. Louis    MO        12     374437      8015               31203               668           46.7
10       Tampa    FL        26     475919      9293               18305               357           51.2
```

QUESTIONS:

Include the R output for the following 4 questions in your completed assignment. Please also include a .pdf of the R code for your two functions.

1. callMeans(callTypes = c("totalCalls", "answeredCalls"), id = 75:105)

2. callMeans summary for sales calls for every id number that is a multiple of 6.

3. callVolume for states AZ and MO.

4. callVolume for all cities in Texas (TX).