STAT 6012: Linear Models for Data Science
Class Activity 1

Due date: Tuesday, July 16 by 10:50 am Via Canvas.

Complete the following questions an R Markdown file and submit your compiled HTML file. If you are working in a group, list the names (last, first) of the group members in alphabetical order of last names.

The attached dataset, `ClassData.csv`, contains data from of you (recall this was collected from you during class yesterday).

and use it to answer the following questions:

1. [1] Import `ClassData.csv` dataset in R and save the resulting dataframe as `OurData`. Use the `head()` to display the first 5 rows of `OurData`.

2. [3] Add two columns, namely `Slp` and `Phone_Hrs`, to `OurData` that properly converts `Sleep_Hrs` and `Phone_Time` to numeric, respectively. Save the resulting dataframe as `OurData2`. Use the `head()` to display the first 5 rows of `OurData2`.

3. [3] Use `OurData2` to make a scatterplot with `Slp` on the y-axis and `Phone_Hrs` on the x-axis. Does there appear to be a linear relationship between the number of hours spent on phone and sleep hours?

4. [3] Assuming you all are representative of all UVA graduate students, find and interpret a 95% confidence interval to estimate the average number of hours all UVA graduate students spend on their phone per day.

Homework/practice question (ungraded).

- Write a code to add a new column, namely `Bfast`, to `OurData` to simultaneously: 1) remove all the $ signs from all the entries of the `Bfast_Amount`, and 2) convert it to numeric. Save the resulting dataframe as `OurData3`. Use the `head()` to display the first 5 rows of `OurData3`.

  [Hint: the `gsub()` function might be useful here. Or, there are functions in the `stringr` package that could be helpful.]