

## STA-234 Data Analysis Project

The goal of this project is to use techniques learned in the class on a sizeable and messy dataset, including data types, data cleaning, organization, getting help in R, problem-solving using R and data visualization with some basic data analysis to answer research questions. Data cleaning, visualization, and simple modeling are major parts along with communicating your findings in a meaningful and concise fashion.

#####

**Step-1 (2/4) [25 points]:** Share your broad areas of interest, and some specific research questions of interest in these areas with me. Investigate the datasets available in your area(s) of interest. I will share resources. Share the potential datasets with me.

**Step-2 (2/11) [25 points]:** Report recent research in the areas of interest and how is your dataset relevant for this research.

**Step-3 (2/18) [25 points]:** Perform EDA on your data and report the results directly addressing the research question of interest.

### **Presentation-1 (2/25): [50 points]**

- **Introduction:** Tell us what problem of interest. Why is this problem interesting and important? Introduce recent research done in area related to your problem. You can pack all this together to motivate us. Do keep it short, to the point, and interesting.
- **Data:** Tell us about the data resource and explain dimensions of the data, variables in the data, and how does this data relate to your research questions.
- **Initial findings:** Perform exploratory data analysis (EDA) including summaries and data visualizations for one research goal. Show main plot(s) and findings. Make sure to add labels, titles etc. to make your tables and graphs informative.
- **Time:** Each individual will get 5 minutes to present and 1 minutes for Q&A, so keep it to the point in this introductory presentation.