**STAT6730 Project Proposal**

Team member: Polina Berezina, Shan Tang

Dataset: “Park Scores” from Tidy Tuesday

<https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-06-22/readme.md>

Overview:

This dataset contains the scores/points of more than 700 public parks across the United States. The points were determined by park size, accessibility to the residents, investment and amenities. The corresponding rank and city for each park are also presented. We will investigate this dataset mainly by graphics (ggplot2), data manipulation and summarization (dplyr), and Bootstrap to answer the following questions:

* Some basic stats summaries on the overall park features.
* Which city has the most top-ranked parks?
* Which factors influence the park rank most?
* How specific data/factor influence the points? For example, in the section of park size, we have the size in acres and the points. We will explore is there a linear relationship exists between these two variables.