ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: "Big" Data Analytics
Proposal
"Yik Yak Traffic" by Mary Alyce Martin

I have decided to study the traffic at various times of the day on the popular Yik Yak app here on Georgia Tech's campus. This app is is centered around a feed of anonymous statuses and statements, more often than not jokes about the difficulty of classes here at Tech or the ineptitude of whatever college we happen to be playing in football that week, posted by students and then "up-voted" or "down-voted" by other users who like or dislike each post respectively. College students are notorious for having strange sleep and wake cycles. Staying up late, pulling all nighters, waking up mid afternoon, and countless naps all characterize the stereotypical college sleep schedule. With such strange routines in mind, I am interested to see at what times of the day most users are active on the app and plan to gather data to answer this question. I am interested to see not only how many posts are up at a given time, but also how well-received these posts are by other users. The number of up-votes for each of these posts will also be helpful in suggesting the number of users currently active on the app. Paying special attention to the top yaks featured on the app and each of their particular number of up-votes, I should be able to even better quantify the usage of the app at a given time.

To collect data for my study, I will check the usage of the app on our campus by monitoring the number of posts ("yaks") and the popularity of the top posts at various times of the day for a minimum of thirty days, eight of which will be weekend days. This can easily be done by simply opening the application and looking at the total yak count in the top left corner of the interface. Yaks disappear after a certain period of time, around three hours, and therefore this number will be almost completely reset between one recording and the next. I will check at 10:00 am, 2:00 pm, 6:00 pm, and 12:00 am for each of the days I randomly select to check on. To randomly select the twenty-two week days I plan to gather data from I will draw numbers out of a hat, each corresponding to a weekday date in November that I will be sampling. I will conduct a similar procedure to decide which eight weekend dates between now the end of this study I will use in my project. I will record the total number of yaks at the time of each of these trials and also the number of up-votes corresponding to the top three yaks listed on the popular feed, the tally of which are displayed to the right for each of these top posts. Keeping careful track of this data in a spread sheet, I will then be able to analyze my findings.

After collecting my data, I will construct four confidence intervals: one for each of the time periods I chose to study. Each interval will indicate with 95 percent confidence how many yaks one can expect for the respective time of day on any given day. Due to the reasonably large sample size of 30 days, I should be able to use the standard normal Z table with little caution despite not knowing the actual population variance. I will also construct four separate 95 percent confidence intervals for weekend days, which will inevitably be wider than the general intervals due to the smaller sample size. This will also require me to use the student t-table. I imagine that my results will indicate that midnight is a very popular time for students to use the Yik Yak app, and ten in the morning is likely too early for many of them to be active on this form of social media. I also imagine that weekdays will have heavier traffic than weekends

due to students complaining about classes and homework during the week. I hope to gain insight into the social media habits of college students through my work on this bonus project.