Assignment #3

MSAN 593 - Summer 2017

***DUE**: Thursday, August 7th, 2018, 23.59*

Instructions

Upload **both** your `*.Rmd` file as well as the knitted `pdf` to Canvas by the due date and time. Late submissions will receive a grade of zero.

1. This homework is intended to be completed and submitted individually.
2. All code should be commented in a neat, concise fashion, explaining the objective(s) of individual lines of code.
3. When making reference(s) to *summary* results, include all relevant output in text of the deliverable where it is being discussed, not in an appendix at the back of the deliverable.
4. Do not include a copy of the raw data in the body of the deliverable unless there is a compelling reason.
5. `R` can generate hundreds of graphs and statistical output extremely easily. Only include *relevant* graphs and output in the deliverable. All graphs and statistical output included in the deliverable should be referenced in the text of the deliverable.
6. **There should be no orphaned figures or graphs**. Everything should be orderly and easy for a grader to read.
7. All code should be visible in the `pdf` version of the homework, i.e., for each code chunk, be sure to set `echo = TRUE`
8. Homework **may not be emailed to the instructor**.

Question 1

1.1 Using a single line graph, generate superimposed probability density functions of the beta distribution where $\{\alpha, \beta\} \in \{\{0.5, 0.5\}, \{5, 1\}, \{1, 3\}, \{2, 2\}, \{2, 5\}\}$. The pdf of each each $\{\alpha, \beta\}$ tuple should be differentiated using color, and a legend should be included.

1.2 Using a single line graph, generate superimposed cumulative distribution functions of the beta distribution where $\{\alpha, \beta\} \in \{\{0.5, 0.5\}, \{5, 1\}, \{1, 3\}, \{2, 2\}, \{2, 5\}\}$. The pdf of each each $\{\alpha, \beta\}$ tuple should be differentiated using color, and a legend should be included.

1.3 Combining output from 1.1 and 1.2, create a single graphical output with all *pdf*s and *cdf*s. They will all be line graphs and will be faceted using two columns and five rows. The title of each facet column will be `pdf` and `cdf`. The titles of the row facets, located on the right hand side of the graph, will have the $\{\alpha, \beta\}$ tuple values, i.e., $\{\alpha, \beta\} = \{x, y\}$, where x and y are the tuple values. The titles **must** have the correct greek symbols α and β in LaTeX (not some weird, ugly, pixelated character), and the tuple values must be generated dynamically based on their values (not hard coded). The output will be a single graphical object with 10 sub-graphs in two columns and five rows.

Question 2

`Officer_Traffic_Stops.csv` is a record of traffic stops by police officers in Charlotte, North Carolina. Employing coding techniques from the `tidyverse`, piping notation, and high-quality, supporting graphs in `ggplot2`, explore the data and report on your findings. Your use of the aforementioend packages—or lack therefof—will impact your grade on this question. There is no right answer to this question. Your job is to explore and report with words, graphs and tables. You need not report on all variables. Use your judgement to distill the data and report back the most interesting information. Your work should not exceed 15 pages, but is not required to be that long. Length is not a criterion on which you will be graded. The best reports are often concise and direct.

Grades for this question will be assigned on a competitive basis, i.e., the student(s) who offer the best insight and report sets the bar for a grade of A, and everyone else will get an inferior (but scaled) grade.