Assignment 6
Submission deadline: Friday Nov 11, 2016 by 8 PM
Submission format: upload document in Canvas

1. **(20 points) Using knitr, format the answers, R code, and R output for the following problems into a pdf file. You can practice generating HTML and RHTML if you like too but that is for your own purposes.**

   Please submit the R Markdown file (.Rmd or .Rnw) that you used to in your write-up.

2. **(20 points) Barplots using ggplot2. You will need the R packages ggplot2 and gcookbook for problems 2-5.**

- Load the **cabbage_exp** dataset in R, and describe the dataset in your own words, in 2-3 lines. Note that is a summarized version of the **cabbages** dataset in the **MASS** package, and you can get detailed information from the **MASS** package.
- Construct a side-by-side bar plot with *Date* in the x-axis, *Weight* in the y-axis, and different colors for different cultivar. Add labels to each bar with the corresponding *Weight*.

3. **(20 points) Area graphs using ggplot2.**

- Load the **uspopage** dataset in R, and describe the dataset in your own words, in 2-3 lines.
- Construct a stacked area graph with Year in the x-axis, population (in thousands) in the y-axis, and different age groups in different layers.
- Next, construct a *proportional* stacked area graph, i.e., for each Year, compute the contribution from each age group to the total population as a fraction of the total population. Construct a stacked area graph as before, with the proportions.

4. **(20 points) Scatterplots using ggplot2.**

- Load the **heightweight** dataset in R (in package **gcookbook**), and describe the dataset in your own words, in 2-3 lines.
- Construct a scatterplot with age in the x-axis, height in the y-axis, and different colors for male and female students. Add linear regression line through the scatterplot using stat_smooth().
- Note that when using stat_smooth() in the previous problem, two regression lines are generated, one for male students and one for female students. Construct the same scatterplot as before, but with a single regression line for all students.

5. **(20 points) Graphical parameters using ggplot2.**

- Load the **faithful** dataset in R, and describe the dataset in your own words, in 2-3 lines.

- Construct a scatterplot with eruption time in the x-axis, waiting time in the y-axis. What interesting features do you see in the plot?
- Add visual guides (e.g. text annotations, arrows, rectangles) to the scatterplot to emphasize the interesting features noted in the previous part.

Assignment instructions:

1. **Honor code:** The Virginia Tech honor pledge for assignments is as follows:
   "I have neither given nor received unauthorized assistance on this assignment."

   The pledge is to be written out on all graded assignments at the university and signed by the student. Type up your name to sign.

2. Submit your assignment as a document (word, pdf or similar) to Canvas, clearly marked with student's name and assignment number, eg. Sengupta_Srijan_HW6.pdf. Your submission should include R code and answers to problems.

3. Late assignments will not be accepted. Check Canvas regularly for assignments and submission dates.

4. You are free to discuss assignment problems with your classmates, but submitted work (answers and codes) **must** be your own work. Students are not allowed to copy computer codes or answers from each other, and must write their own codes and answers.