# MSDS 6306: Introduction to Data Science (403)

# Case Study 01

Due: June 30, 2017 (Friday)

Submit the link to the file in GitHub via the space provided for in the Case Study 01 page in 2DS.

# Description

Beers dataset contains a list of 2410 US craft beers and Breweries dataset contains 558 US breweries. The datasets descriptions are as follows.

#### Beers.csv:

Name: Name of the beer.

Beer\_ID: Unique identifier of the beer. ABV: Alcohol by volume of the beer.

IBU: International Bitterness Units of the beer. Brewery ID: Brewery id associated with the beer.

Style: Style of the beer. Ounces: Ounces of beer.

#### Breweries.csv:

Brew\_ID: Unique identifier of the brewery.

Name: Name of the brewery.

City: City where the brewery is located. State: State where the brewery is located.

# Questions

- 1. How many breweries are present in each state?
- 2. Merge beer data with breweries data by brewery id. Print first 6 observations and the last six observations to check the merged file.
- **3.** Report the number of NA's in each column.
- **4.** Compute the median alcohol content and international bitterness unit for each state. Plot a bar chart to compare.
- 5. Which state has the maximum alcoholic beer? Which state has the most bitter beer?
- **6.** Summary statistics for ABV (Alcohol by volume) variable.
- 7. Is there a relationship between the bitterness of the beer and its alcoholic content? Draw a scatter plot.

You can use ggplot2 library for graphs. Please ignore missing values in your analysis. Comment on your analysis.

# Instructions

Deliverable: Markdown file uploaded to GitHub containing the following

- a. Introduction to the project. The introduction should not start with For my project I.
- **b.** The introduction needs to be written as if you are presenting the work to someone who has given you the data to analyze and wants to understand the result. In other words, pretend its not a case study for a course. Pretend its a presentation for a client.
- c. Brief explanations of the purpose of the code. The explanations should appear as a sentence or two before or after the code chunk. Even though you will not be hiding the code chunks (so that I can see the code), you need to pretend that the client cant see them.
- **d.** Code to answer the seven questions above (plus the answers) in the same R Markdown file.
- e. Clear answers to the questions. Just the code to answer the questions is not enough, even if the code is correct and gives the correct answer. You must state the answer in a complete sentence outside the code chunk.
- f. Conclusion to the project. Summarize your findings from this exercise. The file must be readable in GitHub. In other words, dont forget to keep the md file!!

# Note

R Markdown files often do not render graphics and output from R Markdown chunks in GitHub. Keep the MD (markdown) file and upload it along with the R Markdown file.