**Case Study 1 for MSDS 6306 Using Beers and Breweries data sets**

**Objective:**

Our client Founders Brewery wants to identify new market opportunities and we are to help them determine through finding number of breweries per state, the median ABV (alcohol %) and IBU (bitterness rating) as well as beer with the most alcohol and the most bitter beer in the US and the state(s) where it is produced. The goal is to identify an association between ABV and IBU.

**Data:**

We were given Beers and Breweries data sets but we need to combine, clean and identify any missing values and the percentage missing. If there is a high percentage of data missing, then the assumption made on associating it with other variable may be misleading.

**Read\_Me:**

Both sets are csv files, comma delimited.   
Files parsed into R using read.csv(“filepath://folder\_name//file\_name.csv”)

Data Sets are merged by the brewery\_ID in Beers,csv and Brew\_ID in Vreweries.csv via merge(df.x, df.y, by.x=”specific column\_name from df.x”, by.y=”specific column\_name from df.y”)

**Code Book:**

**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**= *U.S. State where the brewery is located*.

**Questions of Interest (QoIs):**

Identify the number breweries within each state.

Determine the median ABV and IBU from each state.

Identify the states with the highest ABV and IBU.

Relationship association between IBU and ABV.

**R Libraries:**

library(dplyr)

library(tidyr)

library(ggplot2)

library(grid)

library(gridExtra)

library(GGally)

library(maps)

library(openintro)

library(mosaic)

library(kableExtra)

library(stringr)

library(stringi)

*Compiled on a x86\_64 machine, Windows 10 Version 1803*

*R version 3.5.2 (2018-12-20),”Eggshell Igloo"*