Beers and Breweries Analysis

Case Study 01 - MSDS 6306 Section 402

Luke Pierce lepierce@smu.edu
Lokesh Maganti lmaganti@smu.edu
Andrew Walch awalch@smu.edu
MJKelleher mikek@smu.edu

February 14, 2018

Contents

Introduction 1
Beer and Brewery Data Analysis 1
Conclusion 4
Codebook 4
Appendix 9

GitHub Repository: https://github.com/allthebits/msds6306-case-study-01

Introduction

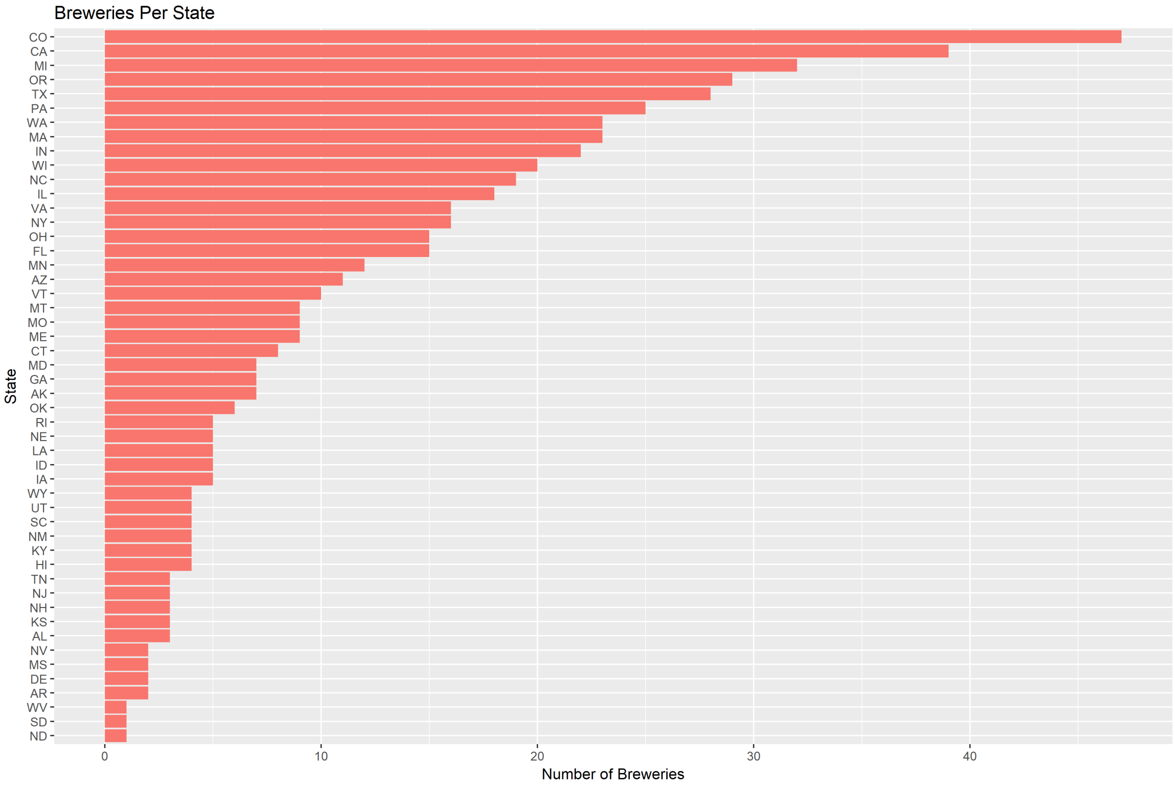
In order to assist NuBrew Inc. in determining the proper product niche for their line of distinctive craft brews, our initial study will focus upon the most general beer metrics of Alcohol by Volume (ABV) and International Bitterness Units (IBV) for product formulation purposes. Ultimately, the market success of NuBrew's product will also depend on this as well as a variety of geographical, demographic and product data analysis which should also be considered when developing a new craft beer in this highly competitive market.

This project is limited to the analysis of the major metrics of Beer and Brewery data from several breweries in the United States. The bulk of our work centers on the measurement of bitterness and alcohol content, with the ultimate goal of trying to identify a causal relationship between the two.

Before we can do this our data must be merged and cleaned. This process results in a "tidy" dataset, the steps of which are outlined below this report. In short, this involves renaming some of the data fields, as well as fixing many data deficiencies that would otherwise cause problems with the analysis. Additionally, we ensure that the analysis is limited to breweries in the United States only. The District of Columbia is not included.

Beer and Brewery Data Analysis

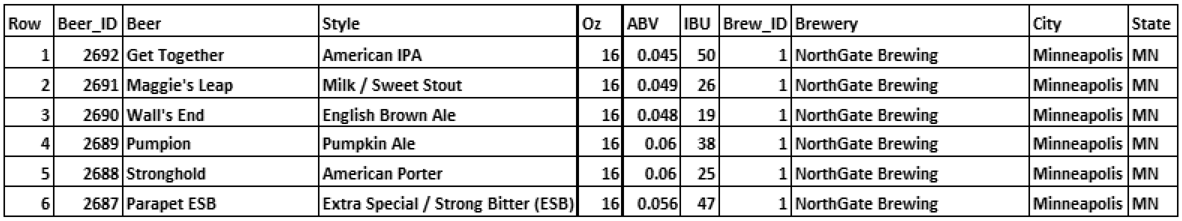
1. First, we examined how many breweries there were in each state so we could get a sense of the relative geographical popularity of craft brewed beers:



In this dataset, craft beer breweries are most numerous in Colorado, California and Michigan, while relatively rare in several states such as West Virginia, South Dakota and North Dakota. More information is needed to ascertain whether this is primarily a function of demographics (an easy answer since craft beer consumers tend to skew younger and urban) or there may be more subtle factors involved.

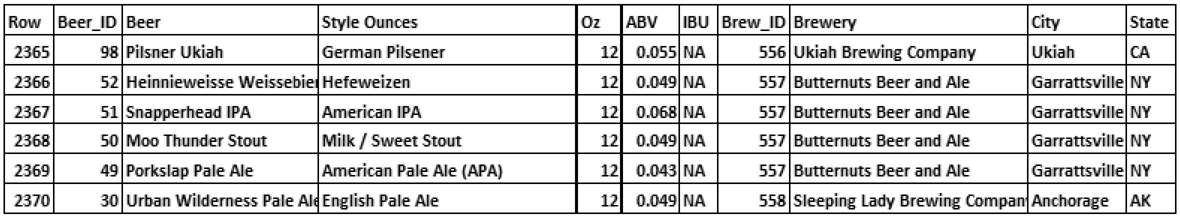
2. After cleaning the data, we used the Brew_ID as a key variable to perform the merge operation. As a quality control measure, here is a listing of the first and last 6 observations in the merged dataset.

First 6 Observations:



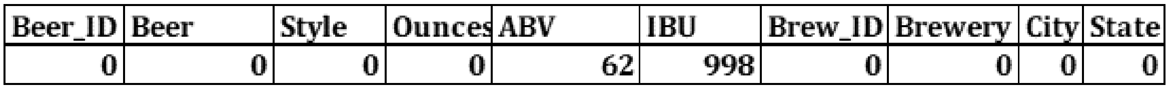
Row	Beer_ID	Beer	Style	Oz	ABV	IBU	Brew_ID	Brewery	City	State
1	2692	Get Together	American IPA	16	0.045	50	1	NorthGate Brewing	Minneapolis	MN
2	2691	Maggie's Leap	Milk / Sweet Stout	16	0.049	26	1	NorthGate Brewing	Minneapolis	MN
3	2690	Wall's End	English Brown Ale	16	0.048	19	1	NorthGate Brewing	Minneapolis	MN
4	2689	Pumpion	Pumpkin Ale	16	0.06	38	1	NorthGate Brewing	Minneapolis	MN
5	2688	Stronghold	American Porter	16	0.06	25	1	NorthGate Brewing	Minneapolis	MN
6	2687	Parapet ESB	Extra Special / Strong Bitter (ESB)	16	0.056	47	1	NorthGate Brewing	Minneapolis	MN

Last 6 Observations:



Row	Beer_ID	Beer	Style Ounces	Oz	ABV	IBU	Brew_ID	Brewery	City	State
2365	98	Pilsner Ukiah	German Pilsener	12	0.055	NA	556	Ukiah Brewing Company	Ukiah	CA
2366	52	Heinnieweisse Weissebier	Hefeweizen	12	0.049	NA	557	Butternuts Beer and Ale	Garrattsville	NY
2367	51	Snapperhead IPA	American IPA	12	0.068	NA	557	Butternuts Beer and Ale	Garrattsville	NY
2368	50	Moo Thunder Stout	Milk / Sweet Stout	12	0.049	NA	557	Butternuts Beer and Ale	Garrattsville	NY
2369	49	Porkslap Pale Ale	American Pale Ale (APA)	12	0.043	NA	557	Butternuts Beer and Ale	Garrattsville	NY
2370	30	Urban Wilderness Pale Ale	English Pale Ale	12	0.049	NA	558	Sleeping Lady Brewing Compan	Anchorage	AK

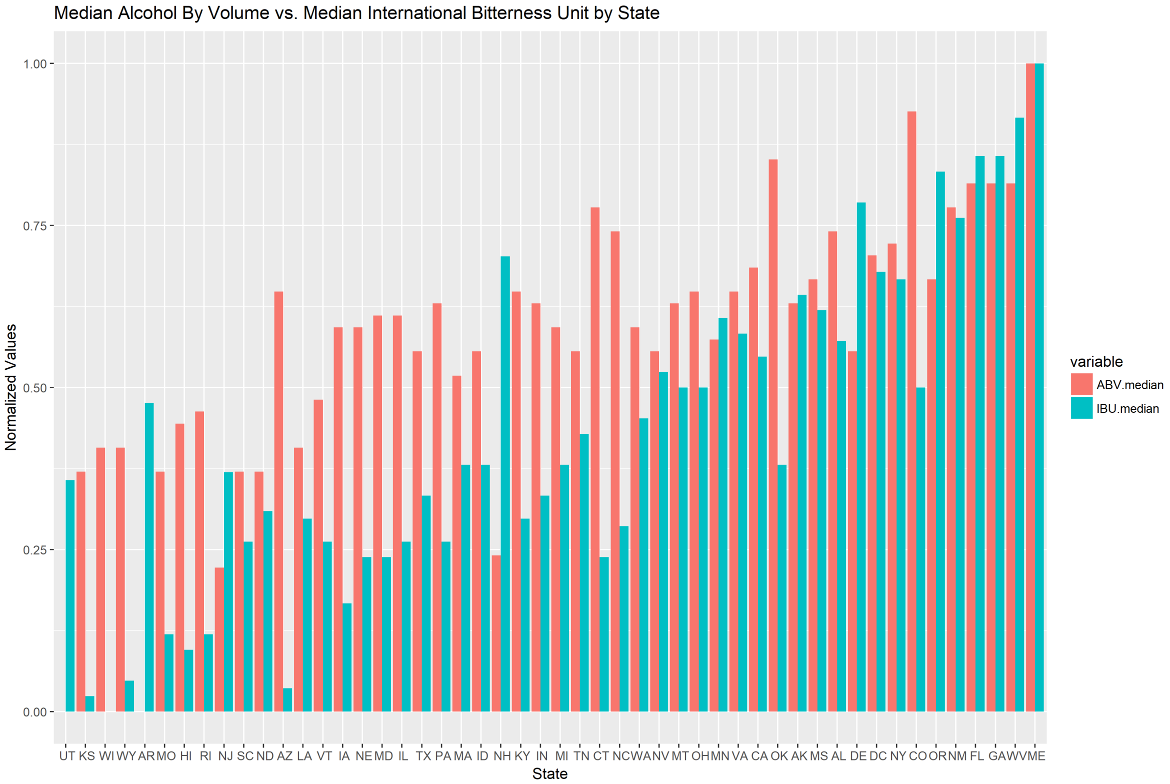
3. Further screening of the merged dataset revealed a significant number of missing ABV and IBU measurements.

- 

Beer_ID	Beer	Style	Ounces	ABV	IBU	Brew_ID	Brewery	City	State
0	0	0	0	62	998	0	0	0	0

In a dataset with 2370 observations after removing duplicate values, the missing ABV values represent 2.6% of the observations, while the missing IBU values represent 42% of the sample data provided. For the purposes of our analysis these observations were omitted in the later part of the study. This lack of data, particularly the IBU values, could impact the reliability of the ABV/IBU comparison. Future measures should be taken to source this information if available.

4. We also looked at the ABV (Alcohol by Volume) and IBU (International Bitterness Units) by state (see below):

- 

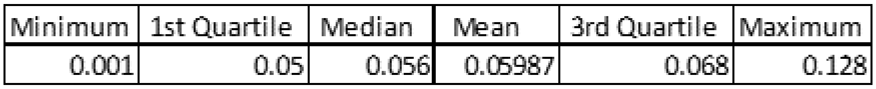
The bar plot above was derived by normalizing median values of ABV (Alcohol By Volume) and IBU (International Bitterness Units) to a range of 0 to 1 and plotting them in pairs, ordered by increasing values per state. There appears to be a positive correlation between median alcohol content by volume and the bitterness unit values observed in this dataset, although in some states there can be significant deviations from this trend.

5. We also looked at the beers with the highest ABV and IBU values:

The beer with the highest alcoholic content (with an ABV value of 0.128 or 12.8%) is “Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale”, brewed by Upslope Brewing Company in Boulder, CO.

The bitterest beer (with an IBU value of 138) is “Bitter Bitch Imperial IPA American Double”, brewed by the Astoria Brewing Company in Astoria, OR.

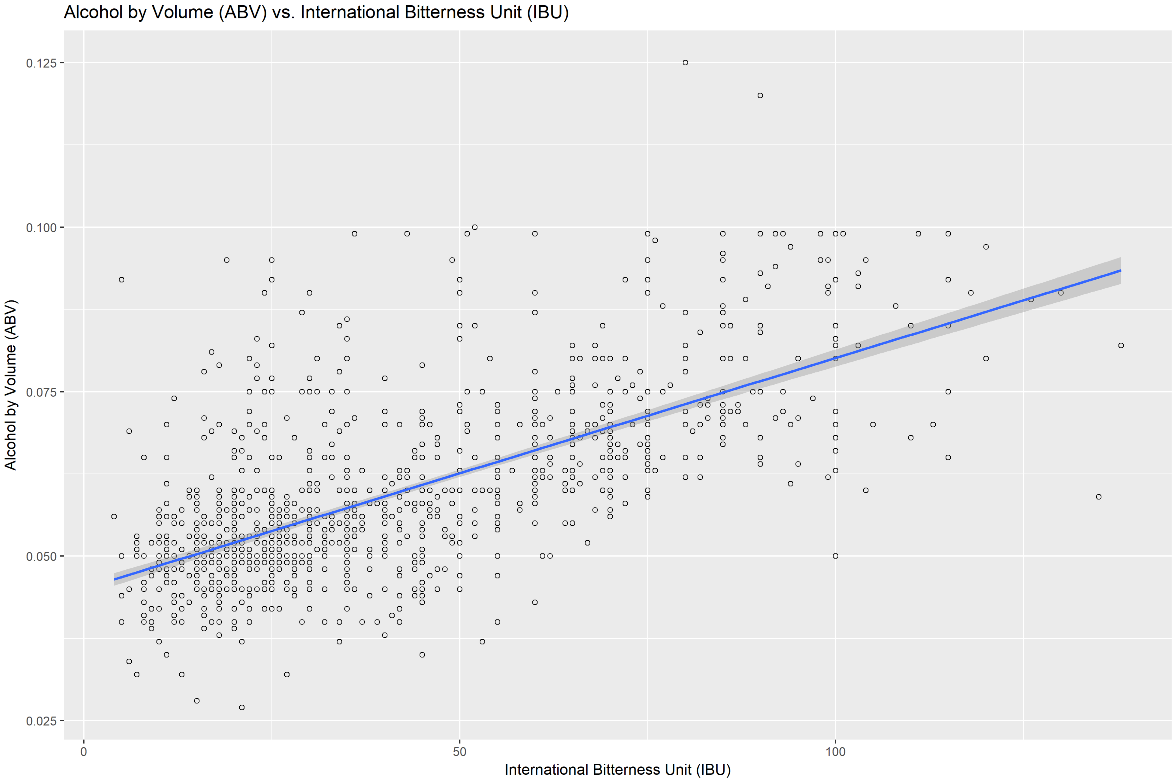
6. The summary statistics for the ABV (Alcohol by Volume) variable are listed below:

- 

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.001	0.05	0.056	0.05987	0.068	0.128

There is quite a variation between the 0.1% almost non-alcoholic brew and the 12.8% Belgian Ale, with the average ABV around 6%.

7. Finally, we plot the relationship between the bitterness of the beer and its alcoholic content.



-

The scatter plot above was derived by plotting median values of ABV (Alcohol By Volume) against median values of IBU (International Bitterness Units). The linear regression line includes a shaded 95% confidence interval. When looking at the scatterplot of ABV vs IBU, there looks to be a linear relationship between the two variables. A correlation analysis would be recommended to measure this relationship. When performing the linear correlation analysis, we have found a moderate positive linear correlation between ABV and IBU (p-value < 0.0001 95% confidence interval, 2-sided t-test). R-squared for the fit is 44.8%. So ABV can predict almost 45% of the IBU scores. This indicates there are other factors involved not accounted for in this analysis and might suggest further investigation.

Conclusion

From our analysis, it does appear that there is a weak positive correlation between the bitterness of a beer and its corresponding alcoholic content. These data also suggest the presence of additional factors which influence this relationship. Additional areas of investigation may include studying constituents (amount of hops and other ingredients and, brewing techniques (aging and post processing methods) which also may influence the character of the product.

Codebook

1. Raw Data
    - Beer Names and Metrics
        - Variable names
            - Beer

  - The name of the beer
  - String
  - Contains non UTF-8 characters
  - No NA’s
- Beer_ID
  - Unique Identifier of the beer
  - Integer - Range: (1 - 2692)
  - No NA’s
- ABV
  - Alcohol by volume of the beer
  - Real Number - Range (0.001 - 0.128)
  - Contains NA’s
- IBU
  - International Bitterness Units of the beer
  - Integer - Range (4 - 138)
  - Contains NA’s
- Brew_ID
  - Brewery ID associated with the beer
  - Integer - Range (1 - 558)
  - No NA’s
- Style
  - Style of the beer
  - String
  - Contains non UTF-8 characters
  - No NA’s
- Ounces
  - Ounces of the beer
  - Real Number - Values: (8.4, 12.0, 16.0, 16.9, 19.2, 24.0, 32.0)
  - Contains NA’s

- Breweries By State
  - Variable names
    - Brew_ID
      - Unique identifier of the brewery
      - Integer (1 - 558)
      - No NA’s
    - Brewery
      - Name of the brewery
      - String
      - Contains non UTF-8 characters
      - No NA’s
    - City
      - City where the brewery is located
      - String
      - No NA’s
    - State
      - US state where the brewery is located
      - 2 Characters
      - No NA’s
      - 51 Unique Values

2. Final Merged Data
   - Beer Names and Metrics
     - Info
       - Dataframe: beer
       - CSV name: data/tidy/beer.csv

  - Variable names
    - Beer
      - The name of the beer
      - String
      - UTF-8 characters only
      - No NA’s
    - Beer_ID
      - Unique Identifier of the beer
      - Integer - Range: (1 - 2692)
      - No NA’s
    - ABV
      - Alcohol by volume of the beer
      - Real Number - Range (0.001 - 0.128)
      - Contains NA’s
    - IBU
      - International Bitterness Units of the beer
      - Integer - Range (4 - 138)
      - Contains NA’s
    - Brew_ID
      - Brewery ID associated with the beer
      - Integer - Range (1 - 558)
      - No NA’s
    - Style
      - Style of the beer
      - String
      - UTF-8 characters only
      - No NA’s
    - Ounces
      - Ounces of the beer
      - Real Number - Values: (8.4, 12.0, 16.0, 16.9, 19.2, 24.0, 32.0)
      - No NA’s
- Breweries By State
  - Info
    - Dataframe: brewery
    - CSV name: data/tidy/brewery.csv
  - Variable names
    - Brew_ID
      - Unique identifier of the brewery
      - Integer (1 - 558)
      - No NA’s
    - Brewery
      - Name of the brewery
      - String
      - UTF-8 characters only
      - No NA’s
    - City
      - City where the brewery is located
      - String
      - No NA’s
    - State
      - US state where the brewery is located
      - 2 Characters
      - No NA’s
      - 51 Unique Values

- Combined Beer and Brewery Names and Metrics
  - Info
  - Dataframe: beerbrew_NA
  - Variable names
    - Beer_ID
      - Integer - Range: (1 - 2692)
      - No NA’s
    - Beer
      - String - Range: (1 - 1372)
      - No NA’s
    - Style
      - String
      - No NA’s
    - Ounces
      - Real Number - Values: (8.4, 12.0, 16.0, 16.9, 19.2, 24.0, 32.0)
      - No NA’s

    - ABV
      - Real Number - Range (0.027 - 0.125)
      - 62 NA’s
    - IBU
      - Integer - Range (4 - 138)
      - 998 NA’s
    - Brew_ID
      - Integer - Range (1 - 547)
      - No NA’s
    - Brewery
      - String
      - No NA’s
    - City
      - String
      - No NA’s
    - State
      - 2 Characters - 50 Unique Values
      - No NA’s
  - Dataframe: beerbrew
  - Info
    - CSV name: data/tidy/beerbrew.csv
  - Variable names
    - Beer_ID
      - Integer - Range: (1 - 2692)
      - No NA’s
    - Beer
      - String - Range: (1 - 1372)
      - No NA’s
    - Style
      - String
      - No NA’s
    - Ounces
      - Real Number - Values: (8.4, 12.0, 16.0, 16.9, 19.2, 24.0, 32.0)
      - No NA’s

    - ABV
      - Real Number - Range (0.027 - 0.125)

        - No NA’s
      - IBU
        - Integer - Range (4 - 138)
        - No NA’s
      - Brew_ID
        - Integer - Range (1 - 547)
        - No NA’s
      - Brewery
        - String
        - No NA’s
      - City
        - String
        - No NA’s
      - State
        - 2 Characters - 50 Unique Values
        - No NA’s
  - Breweries by State
    - Info
      - Dataframe: breweryByState
      - CSV name: data/tidy/BreweryByState.csv
    - Variable names
      - State
        - 2 Characters - 50 Unique Values
        - No NA’s
      - Breweries
        - Integer - Range (1 - 47)
        - No NA’s

3. Data Modifications
  - Load the Beer and Breweries datasets, rename the columns, sort beer dataset
    - Define variables to be used throughout the document
    - Define base/root URL to load the data
    - Define the String URL’s for the Beer and Brewery datasets
    - Load the Beer and Brewery Datasets
  - Convert UTF-8 format character data in Dataframes for Beer and Brewery
    - Convert UTF-8 format in “beer$Name”
    - Convert UTF-8 format in “beer$Style”
    - Convert UTF-8 format in “beer$Style”
  - Modify column/variable names on the Dataframes for Beer and Brewery
    - Rename column “Name” to “Beer” in beer df
    - Rename column “Brewery_id”" to “Brew_ID” in beer df
    - Rename column “Name” to “Brewery” in brewery df
    - Arrange beer df by Brew_ID
    - Remove duplicates with all columns other than Brew_ID as criteria for removal
    - Remove row.names column

4. Tidy dataset
  - beer
  - beerbrew
    - Merge data frames
    - Sort columns
    - NA removal (all)
  - brewery
  - BreweryByState

        - Count breweries per state
        - Remove Washington DC
        - Rename column "Brewery_id"" to "Brew_ID" in beer df
        - Sort by most to least

5. Recipe for Tidy Dataset
    - Commented R script file
        - Tidy_Recipe.R

Appendix

File Definitions

- Project Directories, Documentation, Source Code, Data
    - /
        - MSDS_6306_Sect_402_Case_Study_01.Rmd
            - The main RMarkdown file. Use this file to knit the report.
        - MSDS_6306_Sect_402_Case_Study_01.html
            - This is the knitted report.
        - README.md
            - The file you are looking at right now.
        - Solution.R
            - This is the same R code to produce the results, minus all of the RMarkdown.
        - Writeup_CaseStudy1_LEP_v1.txt
            - TODO: define what this is
        - Tidy_Recipe.R
            - If you want to just build the tidy dataset, run this file.
        - Case_Study_code_file_v1
            - TODO: define what this is, move it to archive, or delete it.
        - resources.txt
            - TODO: define what this is, move it to archive, or delete it.
        - CaseStudy1.pdf
            - TODO: define what this is, move it to archive, or delete it.
    - archive/
        - Exactly what is says.
    - code/
        - Re-usable chunks of code used in both the main project Rmd file, the Solution.R file and Tidy_Recipe.R.
        - 00_Setup.R
            - This sets the R environment up for the all of the analyses to be run.
        - 00_LoadAndPrepare.R
            - Loads the data, and prepares it for analysis.
        - 01_Question_01.R
            - Performs the operations to answer the first question.
        - 01_Question_01.tidy.R
            - Performs the oprations to create the tidy dataset needed for Question 1, and all other questions.
        - 02_Question_02.R
            - Performs the operations to answer the second question.
        - 02_Question_02.tidy.R
            - Performs the oprations to create the tidy dataset needed for Question 2, and all other questions.
        - 03_Question_03.R

      - Performs the operations to answer the third question.
    - 03_Question_03.tidy.R
      - Performs the oprations to create the tidy dataset needed for Question 3, and all other questions.
    - 04_Question_04.R
      - Performs the operations to answer the fourth question.
    - 05_Question_05.R
      - Performs the operations to answer the fifth question.
    - 06_Question_06.R
      - Performs the operations to answer the sixth question.
    - 07_Question_07.R
      - Performs the operations to answer the seventh question.
  - rmd/
    - Files in this directory are included into the main file when the document is knitted together. They are here mainly for organizational purposes.
    - Answers.Rmd
      - Human readable, textual answers in prose form. These are the formal answers the the posed questions.
    - Codebook.Rmd
      - The codebook, defining the uncleaned data, as well as the tidy data.
    - Introduction.Rmd
      - Project introduction.
  - data/
    - Breweries.csv
    - Beers.csv Directories not managed by Git
- data/tidy/ : This is where the tidy dataset are stored when they are created by running the Analysis. This is transient storage.
- tmp/ : Transient storage for graphics and other intermediary files.

Source Code

Required Libraries

1. deplyr - To Install: `install.packages("deplyr")`
2. ggplot2 - To Install: `install.packages("ggplot2")`
3. doBy - To Install: `install.packages("doBy")`
4. stringr - To Install: `install.packages("stringr")`
5. reshape2 - To Install: `install.packages("reshape2")`
6. gridExtra - To Install: `install.packages("gridExtra")`

Case Study Solution

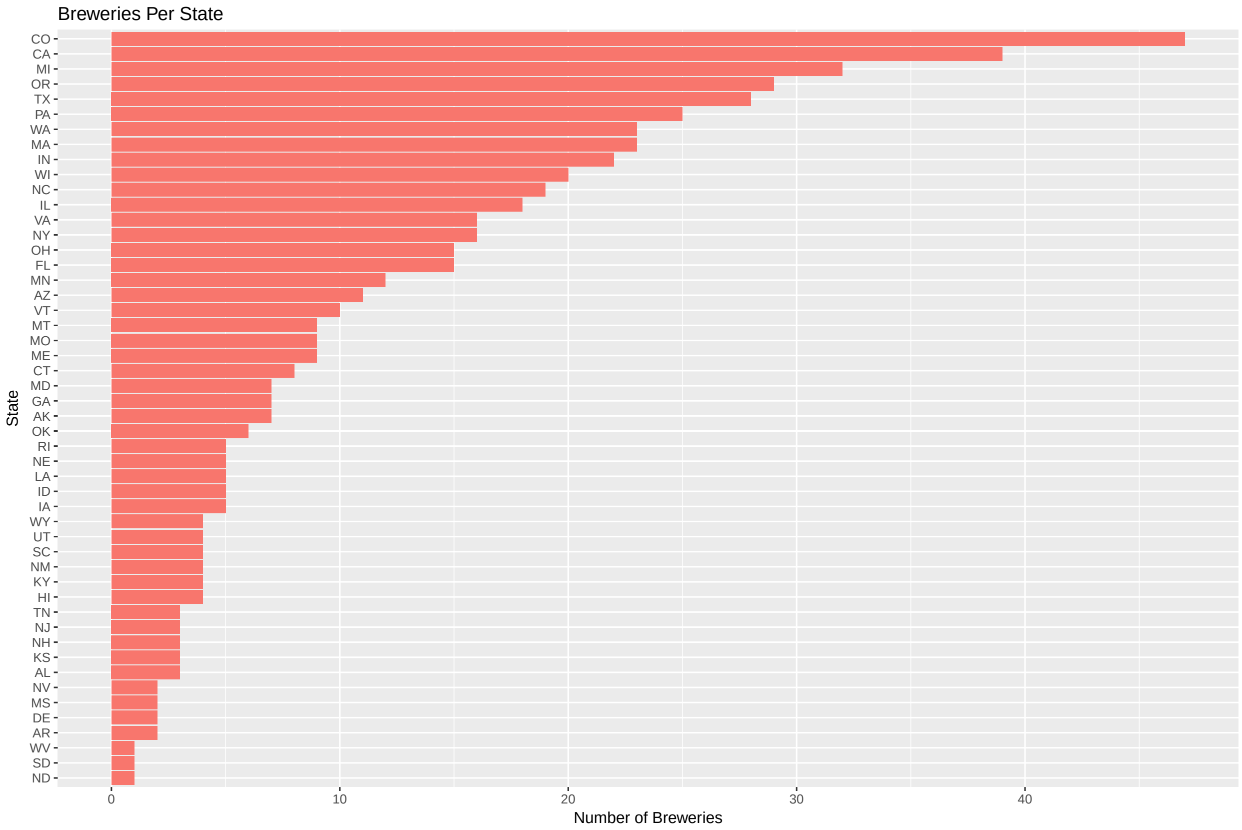
```r
#
# NOTE:
# If you add any libraries to this file, make sure you add the library to the
# 'ENVIRONMENT' section of the file: code/00_LoadAndPrepare.R
#
library(dplyr)
library(ggplot2)
library(doBy)
library(stringr)
library(reshape2)
library(gridExtra)
```

```
#
#

#===============================================================================

## Load the Beer and Breweries datasets, rename the columns, sort beer dataset

# Define variables to be used throughout the document
# The base/root URL to load the data from
data_root_url <- "https://raw.githubusercontent.com/allthebits/msds6306-case-study-01/master/data/"

# Define the String URL's for the Beer and Brewery datasets
beer_url <- paste(data_root_url, "Beers.csv", sep="");
brewery_url <- paste(data_root_url, "Breweries.csv", sep="");

# Load the Beer and Brewery Datasets
beer <- read.csv(url(beer_url), header = TRUE, sep=",", row.names = NULL)
brewery <- read.csv(url(brewery_url), header = TRUE, sep=",", row.names = NULL)

#===============================================================================

# Convert UTF-8 format character data in Dataframes for Beer and Brewery

# Beer file

# Convert UTF-8 format in "beer$Name"
beer$Name <- str_conv(beer$Name, "UTF-8")

# Convert UTF-8 format in "beer$Style"
beer$Style <- str_conv(beer$Style, "UTF-8")

# Brewery file

# Convert UTF-8 format in "beer$Style"
brewery$Name <- str_conv(brewery$Name, "UTF-8")

# ==============================================================================

# Modify column/variable names on the Dataframes for Beer and Brewery

# in beer df rename column "Name" to "Beer"
beer <- rename(beer, Beer = Name)

# in beer df rename column "Brewery_id"" to "Brew_ID"
beer <- rename(beer, Brew_ID = Brewery_id)

# in brewery df rename column "Name" to "Brewery"
brewery <- rename(brewery, Brewery = Name)

# Arrange beer df by Brew_ID
beer <- arrange(beer, (Brew_ID))

# Remove duplicates with all columns other than Brew_ID as criteria for removal
```

```r
beer <- beer[!duplicated(beer[c('Beer', 'ABV', 'IBU', 'Style', 'Ounces')]),]

# Remove row.names column
row.names(beer) <- NULL
#
#
# Case Study 01 : Question 01) Breweries per state?
# Requires the library: 'ggplot2'

# The source path MUST include the "code" directory because the context
#   when the source statement executes is within the RMarkdown file and that
# is one directory 'up' from here
source('code/01_Question_01.tidy.R')

summary(BreweryByState)
```

```
##      State      Breweries    
##   AK    : 1   Min.   : 1.00  
##   AL    : 1   1st Qu.: 4.00  
##   AR    : 1   Median : 7.00  
##   AZ    : 1   Mean   :11.14  
##   CA    : 1   3rd Qu.:16.00  
##   CO    : 1   Max.   :47.00  
##  (Other):44
```

```r
q1_plot <- ggplot(BreweryByState, aes(x=reorder(State, Breweries), y=(Breweries), fill = "red")) + geom_

grid::grid.draw(q1_plot)
```



```
ggsave(q1_plot, filename="tmp/q1_plot.png")
```

```
## Saving 12 x 8 in image
```

```
#
#
# Case Study 01 : Question 02) Merge beer and brewery data. Print first and last
#                 6 observations
# QC

source('code/02_Question_02.tidy.R')

summary(beerbrew)
```

```
##     Beer_ID           Beer              Style               Ounces
##  Min.   :   4.0   Length:2370        Length:2370        Min.   : 8.40
##  1st Qu.: 813.2   Class :character   Class :character   1st Qu.:12.00
##  Median :1457.5   Mode  :character   Mode  :character   Median :12.00
##  Mean   :1432.6                                         Mean   :13.59
##  3rd Qu.:2073.8                                         3rd Qu.:16.00
##  Max.   :2692.0                                         Max.   :32.00
##
##       ABV               IBU            Brew_ID        Brewery
##  Min.   :0.00100   Min.   :  4.00   Min.   :  1.0   Length:2370
##  1st Qu.:0.05000   1st Qu.: 21.00   1st Qu.: 94.0   Class :character
##  Median :0.05600   Median : 35.00   Median :206.0   Mode  :character
##  Mean   :0.05987   Mean   : 42.78   Mean   :232.8
##  3rd Qu.:0.06800   3rd Qu.: 64.00   3rd Qu.:367.8
##  Max.   :0.12800   Max.   :138.00   Max.   :558.0
##  NA's   :62        NA's   :998
##            City          State
##  Grand Rapids:  66    CO     : 259
##  Chicago     :  55    CA     : 180
##  Portland    :  53    MI     : 162
##  Indianapolis:  43    IN     : 139
##  Boulder     :  41    TX     : 130
##  San Diego   :  41    OR     : 114
##  (Other)     :2071   (Other):1386
```

```
str(beerbrew)
```

```
## 'data.frame':    2370 obs. of  10 variables:
##  $ Beer_ID: int  2692 2691 2690 2689 2688 2687 2686 2685 2684 2683 ...
##  $ Beer   : chr  "Get Together" "Maggie's Leap" "Wall's End" "Pumpion" ...
##  $ Style  : chr  "American IPA" "Milk / Sweet Stout" "English Brown Ale" "Pumpkin Ale" ...
##  $ Ounces : num  16 16 16 16 16 16 16 16 16 16 ...
##  $ ABV    : num  0.045 0.049 0.048 0.06 0.06 0.056 0.08 0.125 0.077 0.042 ...
##  $ IBU    : int  50 26 19 38 25 47 68 80 25 42 ...
##  $ Brew_ID: int  1 1 1 1 1 1 2 2 2 2 ...
##  $ Brewery: chr  "NorthGate Brewing " "NorthGate Brewing " "NorthGate Brewing " "NorthGate Brewing "
##  $ City   : Factor w/ 384 levels "Abingdon","Abita Springs",..: 228 228 228 228 228 228 200 200 200
##  $ State  : Factor w/ 51 levels " AK"," AL"," AR",..: 24 24 24 24 24 24 18 18 18 18 ...
```

```
# Check beerbrew

# q2_out1 <- capture.output(head(beerbrew,6))
# q2_out2 <- capture.output(tail(beerbrew,6))

# text <- paste0(q2_out1, q2_out2)
# textplot(text, valign="top")

# png(file="tmp/q2_plot.png")
# textplot(text, valign="top")
# dev.off();
#
#
# Case Study 01 : Question 03) Report NA in each column
sapply(beerbrew, function(x) sum(is.na(x)))
```

```
## Beer_ID    Beer   Style  Ounces     ABV     IBU Brew_ID Brewery    City
##       0       0       0       0      62     998       0       0       0
##   State
##       0
```

```
# q3_out <- capture.output(sapply(beerbrew, function(x) sum(is.na(x))))
# q3_out

#
source('code/03_Question_03.tidy.R')
#
#
# Case Study 01 : Question 04) Median ABV and IBU by state. Plot barchart.
# Requires the library:  'doBy'
```

```r
# Calculate median values for each obs of ABV and IBU by state using DoBy

MedianABV <- summaryBy(ABV ~ State, data = beerbrew, FUN = median)
MedianIBU <- summaryBy(IBU ~ State, data = beerbrew, FUN = median)

# Merge into one df
ABV_IBU_median <- dplyr::inner_join(MedianABV, MedianIBU, by = "State")

summary(ABV_IBU_median)
```

```
##      State      ABV.median        IBU.median
##   AK    : 1   Min.   :0.04000   Min.   :19.00
##   AL    : 1   1st Qu.:0.05262   1st Qu.:30.00
##   AR    : 1   Median :0.05625   Median :35.00
##   AZ    : 1   Mean   :0.05557   Mean   :37.05
##   CA    : 1   3rd Qu.:0.05838   3rd Qu.:44.25
##   CO    : 1   Max.   :0.06700   Max.   :61.00
##  (Other):44
```

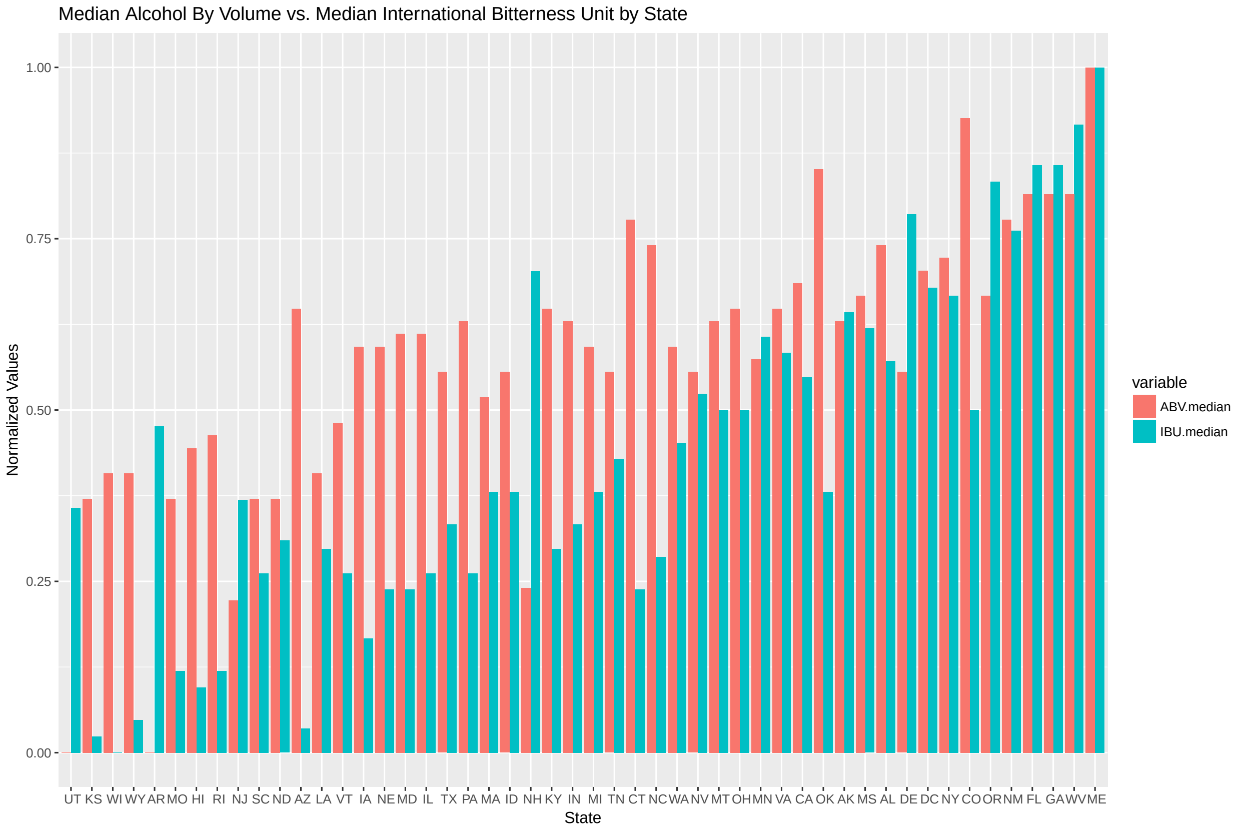
```r
str(ABV_IBU_median)
```

```
## 'data.frame':    50 obs. of  3 variables:
##  $ State     : Factor w/ 51 levels " AK"," AL"," AR",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ ABV.median: num  0.057 0.06 0.04 0.0575 0.0585 0.065 0.061 0.059 0.055 0.062 ...
##  $ IBU.median: num  46 43 39 20.5 42 40 29 47.5 52 55 ...
```

```r
# Normalize ABV and IBU values for direct comparison
ABV_IBU_median_norm <- as.data.frame(apply(ABV_IBU_median[, 2:3], 2, function(x) (x - min(x))/(max(x)-mi

# Add back State column
ABV_IBU_median_norm <- cbind(State = ABV_IBU_median$State, ABV_IBU_median_norm)

# Melt data frame (ABV and IBU in one column) for ggplot
ABV_IBU_median_long <- melt(ABV_IBU_median_norm)
```

```
## Using State as id variables
```

```r
# Plot dual barplots with ggplot2
q4_plot <- ggplot(ABV_IBU_median_long,aes(x = reorder(State,value), y = value,fill=variable)) + geom_ba


grid::grid.draw(q4_plot)
```



```
ggsave(q4_plot, filename="tmp/q4_plot.png")
```

```
## Saving 12 x 8 in image
```

```
#
#
# Case Study 01 : Question 05) State with highest ABV? Highest IBU?
# Most alcoholic beer (Kentucky)
dplyr::top_n(beerbrew_NA, 1, ABV)
```

```
##   Beer_ID                                                  Beer
## 1    2565 Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale
##              Style Ounces   ABV IBU Brew_ID                 Brewery
## 1 Quadrupel (Quad)   19.2 0.128  NA      52 Upslope Brewing Company
##      City State
## 1 Boulder    CO
```

```
# Most bitter beer (Oregon)
dplyr::top_n(beerbrew_NA, 1, IBU)
```

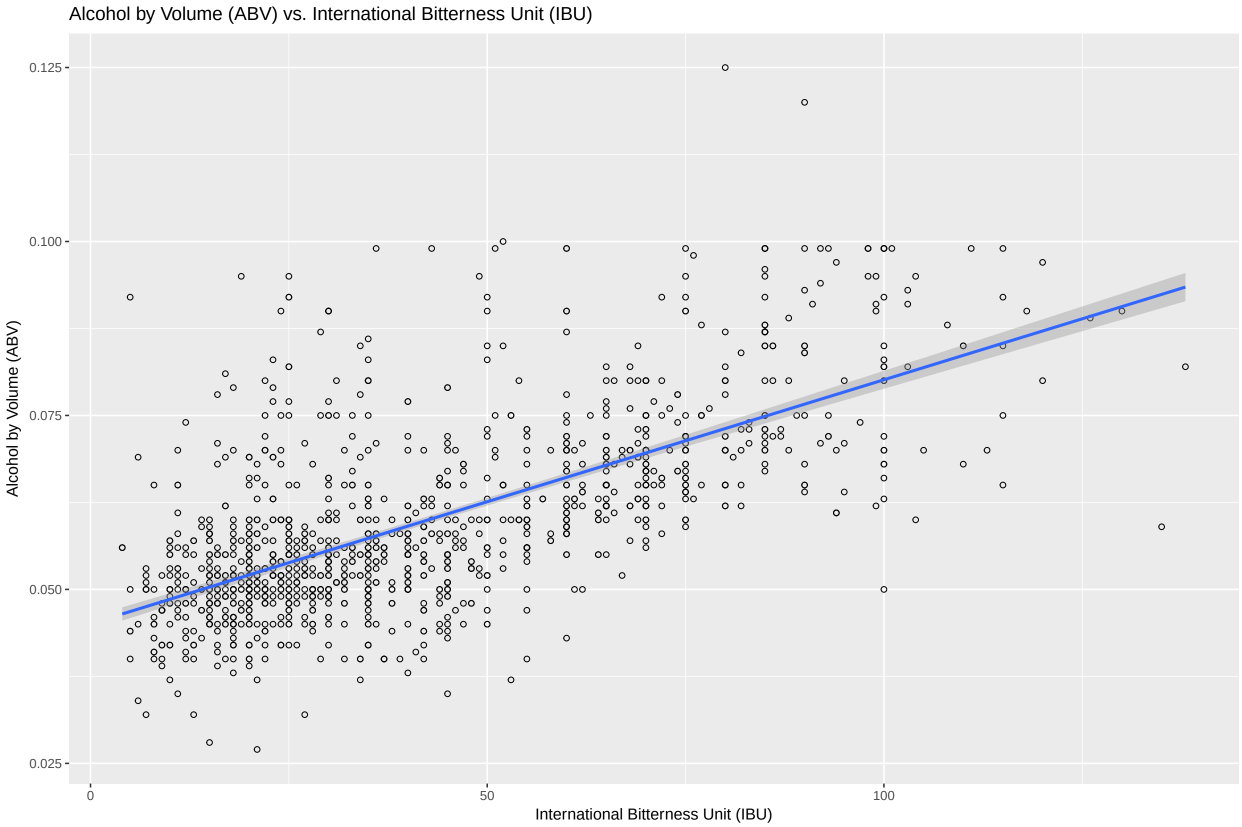
```
##   Beer_ID                      Beer                           Style Ounces
## 1     980 Bitter Bitch Imperial IPA American Double / Imperial IPA      12
##     ABV IBU Brew_ID                 Brewery    City State
## 1 0.082 138     375 Astoria Brewing Company Astoria    OR
```

```
#
# Remove NA values only for ABV variable
beerbrew_ABV <- beerbrew_NA[complete.cases(beerbrew_NA[ , 5]),]
#
```

```r
# Case Study 01 : Question 06) Summary Statistics for the ABV variable
summary(beerbrew_ABV$ABV)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00100 0.05000 0.05600 0.05987 0.06800 0.12800
```

```r
#
#
# Case Study 01 : Question 07) Is there an apparent relationship between the
#                              bitterness of the beer and its alcoholic content?
#                              Draw a scatter plot.

q7_plot <- ggplot(beerbrew, aes(x=IBU, y=ABV)) +
    geom_point(shape=1) +    # Use hollow circles
    geom_smooth(method=lm) +  # Add linear regression line (by default includes 95% confidence region)
    labs(title = "Alcohol by Volume (ABV) vs. International Bitterness Unit (IBU)") + labs(x = "Interna

grid::grid.draw(q7_plot)
```



```r
ggsave(q7_plot, filename="tmp/q7_plot.png")
```

```
## Saving 12 x 8 in image
```