MSDS Project 1 Codebook
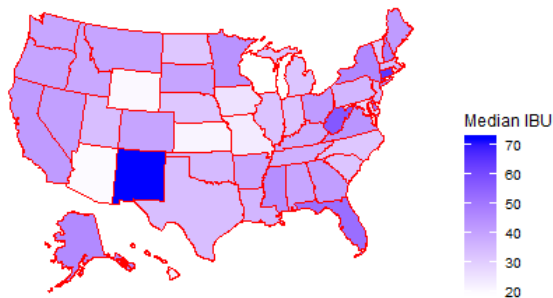
The FrankenBeer Team

- Allen Crane
- Nick Cellini
- Chris Graves
- Heber Nielsen
- Quincy Roundtree

Median State Bitterness Values
Median from Brews Within Each State

Group Github Repository: https://github.com/nickcellini/DDS-Case-Study-ACQN

Repository contents:

- **FOLDER "Original Data"**. This folder contains two files: "beers.csv" and "breweries.csv".
    - Beers.csv

        Description: Original data file containing information on beers. File obtained from course professor, Dr. Jacquelyn Cheun. This file has the following columns of data :

        - Column – "Name". This column lists individual beers by the name of the specific beer.
        - Column – "Beer_ID". This column provides an identification number specific to each beer in the "Name" column. Our research did not identify the source of the beer ID number. The number is identical to that listed in the "Brew_ID" variable in "Breweries.csv".
        - Column - _"ABV". This column lists the alcoholic content of the beer, as alcohol by volume (ABV). The number is the fraction of a total volume (1.00) that is alcohol. Multiplying this number by 100% gives the more common "percent alcohol".
        - Column – "IBU". This column lists the bitterness rating of each beer, as measured by industry standards. The rating is in "International Bitterness Units".
        - Column – "Brewery_ID". This is a number designating the specific brewery which brews the given beer. Our research did not identify the source of the brewery ID number.
        - Column – "Style". A designation of the class which describes the type of beer each entry belongs to. Source is the beer industry (generally the brewer's description of the beer).
        - Column – "Ounces". The volume of the standard individual container for purchase of that specific beer, in fluid ounces,

    - Breweries.csv
    Description: Original data file containing information on breweries which brew the beers listed in the Beers.csv file. File obtained from course professor, Dr. Jacquelyn Cheun. The file has the following columns of data:

- Column – "Brew_ID". This column provides an identification number specific to each beer in the "Name" column. Our research did not identify the source of the beer ID number. The number is identical to that listed in the "Beer_ID" variable in "Beers.csv".
- Column – "Name". This column lists the individual breweries which brew the beers identified in the "Brew_ID" column.
- Column – "City". This is the city where the breweries in "Name" are located.
- Column – "State". This lists the states associated where the associated city in "City" are located.

- **FOLDER "DATA_SETS"**. This folder contains two files: "Enhanced_beers.csv" and "Enhanced_beer_data_consolidated.csv".

  - enhanced_beers.csv
    Description: This is the data set which resulted from tidying "Beers.csv". The data file has the same columns and identities as "Beers.csv". See that file entry in this copybook for explanations. The tidying consisted primarily of adding and correcting ABV and IBU values to replace most of the "NA" values found in "Beers.csv".

  - enhanced_beer_data_consolidated.csv
    Description: This data set is the result of merging "enhanced_beers.csv" and "breweries.csv". Merging was done using the beer_ID columns in both data sets (the column title in "breweries.csv" was changed to agree with the title in "Beers.csv". The data set consists of 7 variables and 212 lines (2411 lines of data values, one header line of column names). The columns are:
    - Name – same as in "Beers.csv".
    - Beer_ID – same definition as in "Beers.csv".
    - ABV - same definition as in "Beers.csv".
    - IBU - same definition as in "Beers.csv".
    - Brewery_id - same definition as in "Breweries.csv".
    - Style - same definition as in "Breweries.csv".
    - Ounces - same definition as in "Beers.csv".

- **ADDITIONAL FILES IN THE "MAIN DIRECTORY"**

- Project 1 Enhanced.rmd
  Description: A markdown file, prepared in R, that gives the group's preferred answers to the questions posed for this exercise. NOTE: Our group prefers these answers as these use the "enhanced_beers.csv" data set (see explanation accompanying that data set). The contents in this file used only the more complete data we located for ABV and IBU values through our research. We consider the material in this file to be the better responses to the questions.

- Project_1_Enhanced.html
  Description: An html product of the markdown file in "Project 1 Enhanced.rmd". The features of this product are described under the entry for "Project 1 Enhanced.rmd".

- Project 1.Rmd
  Description: A markdown file, prepared in R, that gives the group's answers to the questions posed for this exercise. NOTE: the results shown in this file use only tidy data from the original Beers.csv and Breweries.csv data sets. As our group was able to use other sources to provide more data to fill in for NA values in these original sets, we consider this file only answers the questions in part. Please see the files titled "Project 1 Enhanced.rmd" and "Project_1.html" for our complete and preferred answers that use the additional data.

- Project_1.html
  Description: An html product of the markdown file in "Project 1.rmd. The features of this product are described under the entry for "Project 1.rmd". , including the rationale for considering this product as a limited response to the questions.. Please see the files titled "Project 1 Enhanced.rmd" and "Project_1.html" for our complete and preferred answers using the enhanced data.

:

- Project 1 References
  Description: A .txt file of the major sources used for carrying out the various components of this project.

- Project 1 Codebook.pdf
  Description: This document.