Codebook

**Data dictionary:**

Description

The 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: U.S. State where the brewery is located.

Machine/Operating system: R 4.3.2

Key ID:

* Brew\_ID
* Brewery\_ID

Changed Brew\_ID to Brewery\_ID so Key ID’s could match for full merge of both datasets.

When looking at missing values we created new columns in order to get the percentage of missingness. This was done because there is a large spread of observations throughout