Analysis of Beer and Brewery Dataset

### Introduction

#### Dataset contains:

- Breweries and beers from all 50 states
- ~550 Breweries
- ~2400 Beers
- Beers and Breweries are in separate files.
- Datasets were merged on Brewery ID

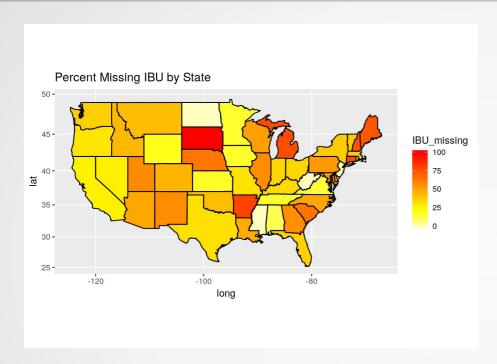
Goal: Find something interesting in this data

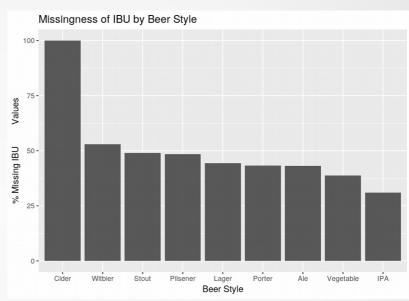
### **Null Values in Dataset**

Column Name	Number of Missing Values	Percent of Values Missing
IBU	1005	41.70%
ABV	62	2.57%
Style	5	0.21%

- A large portion of the IBU values are missing. This is a concern for taking insights from the data.
- The portion of missing data for ABV and Style is relatively small.

# Missing IBU Data

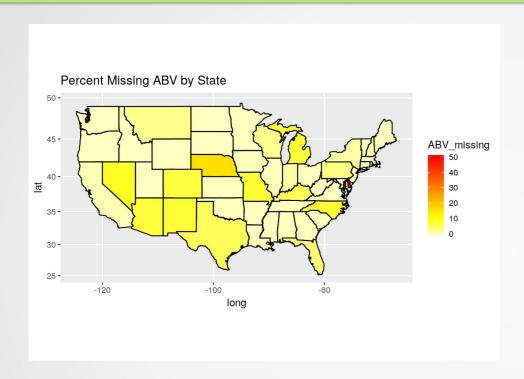




Missing IBU values appear to related to

- States
- Beer Style

# Missing ABV Data

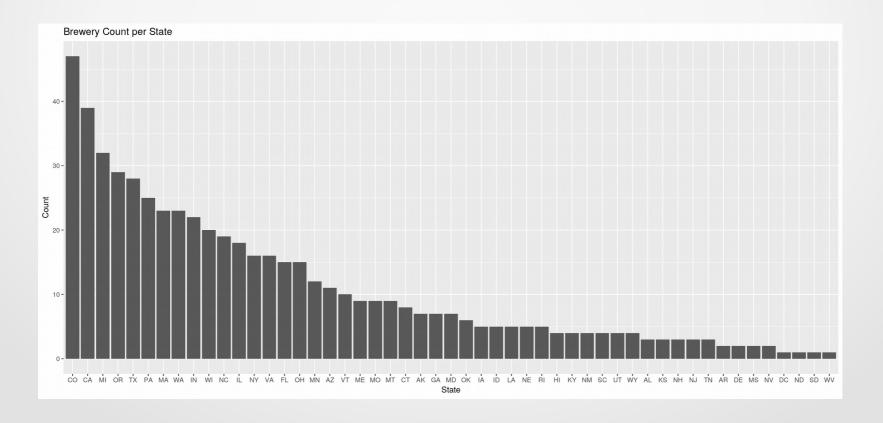


- Generally small variation across states
- Generally low amount missing per state

# Breweries by State

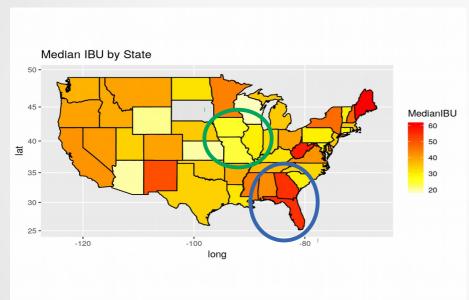
Bulk of the breweries in one-thirds of states.

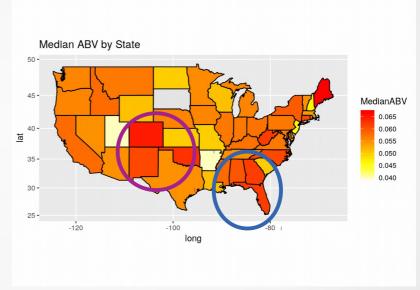
Most breweries are in Colorado.



# Median IBU and ABV by State

Plotting median IBU and ABV by state shows some geographic clustering.





Higher median IBU/ABV in deep south

High ABV in south central states

Low IBU in midwestern states

## Maximum IBU and ABV by State

 The most alcoholic beer is from Kentucky, followed by Indiana.

These two beers have a substantially higher ABV than others.

The most bitter beer is from Oregon

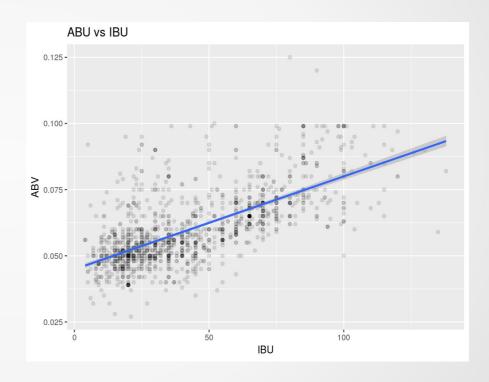
### Statistics of ABV

Statistic	Value
Min	0.1%
1st Quartile	5%
Mean	5.98%
Median	5.60%
3rd Quartile	6.7%
Max	12.8%
standard deviation	1.35 percentage points
range	12.7 percentage points

- Median and mean are not close in value
- The bulk of the data is between 5% and 6.7%

## Relationship of ABV and IBU

- Strong evidence of a linear relationship between IBU and ABV.
- Linear Regression (p-value < 0.0001)</li>



### Conclusion

Insights from the data exploration:

There is evidence that the missing data is not 'at random'.

There appears to be a geographic clustering of median IBU and median ABV of beers.

Most beers have a moderate ABV (5% - 6.7%).

There appears to be evidence of a linear relationship between IBU and ABV.

#### Thanks!

#### For more information:

See the github repo for this project https://github.com/KThompson0308/beeranalysis

#### **Presentation Video**

https://youtu.be/E9f5CHflCzM