

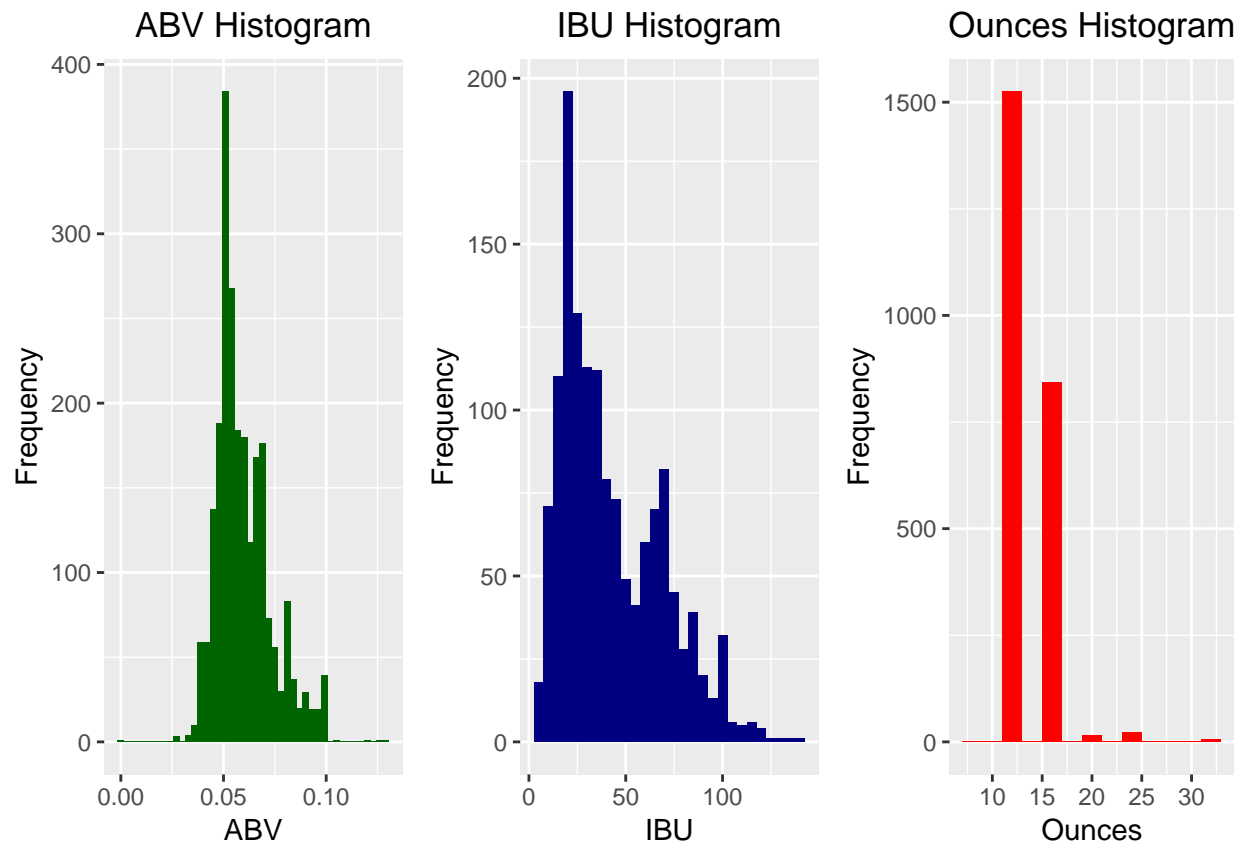
Craft Beer

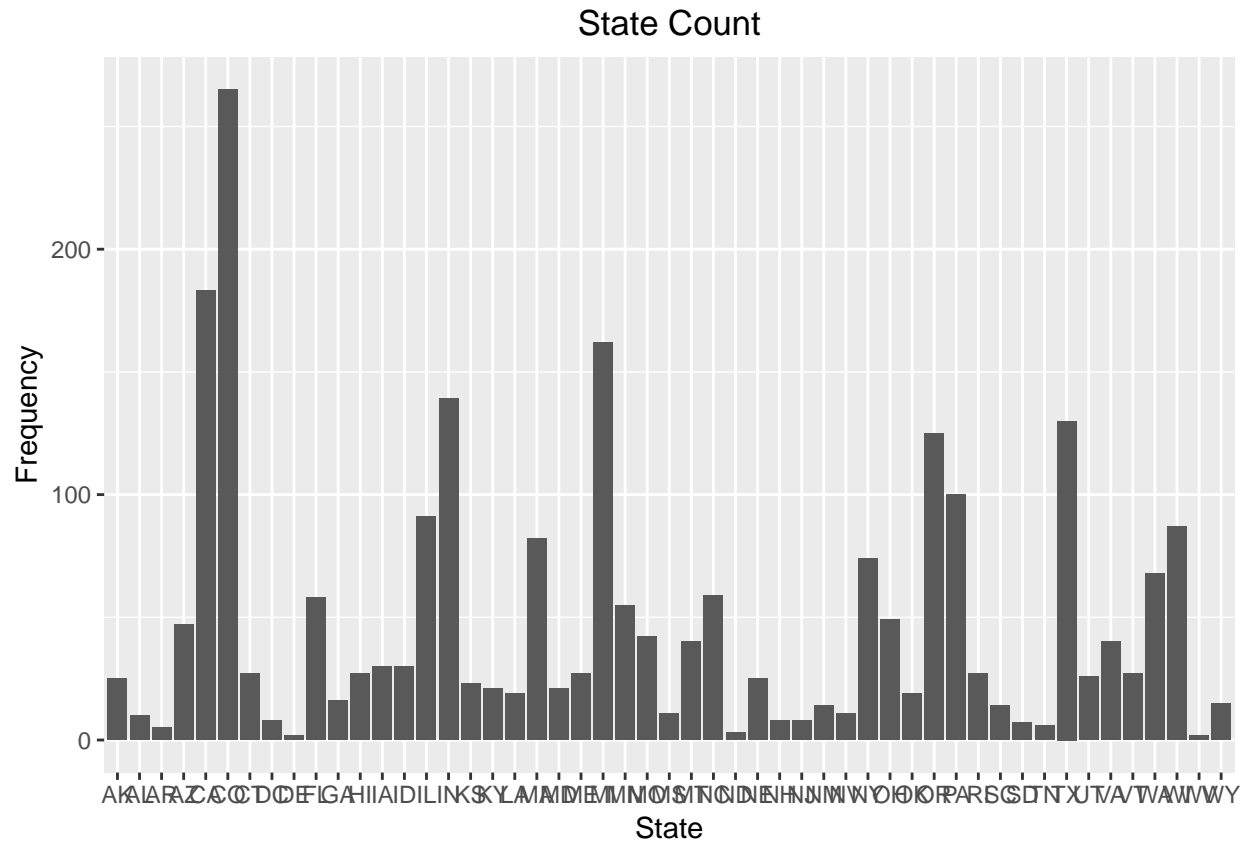
Santiago Tellez

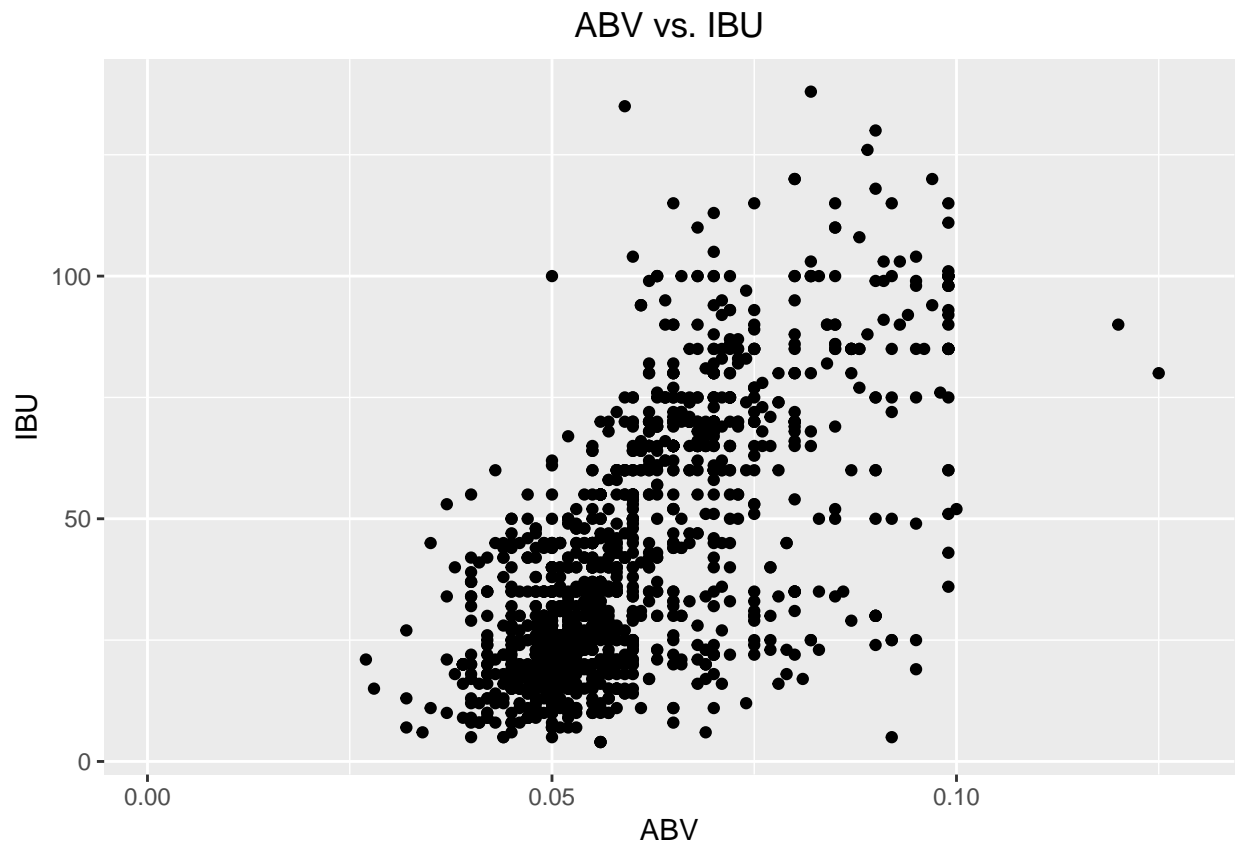
1/25/2019

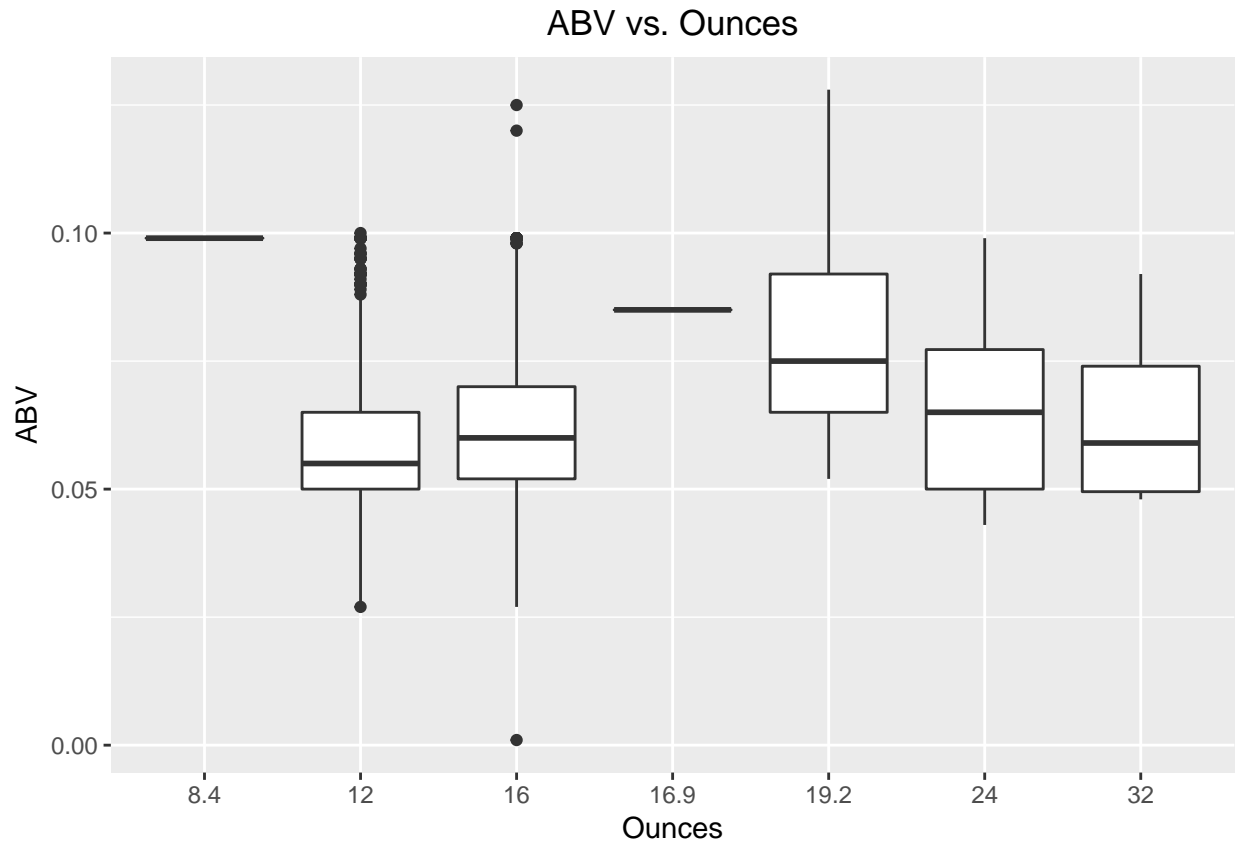
Loading Data

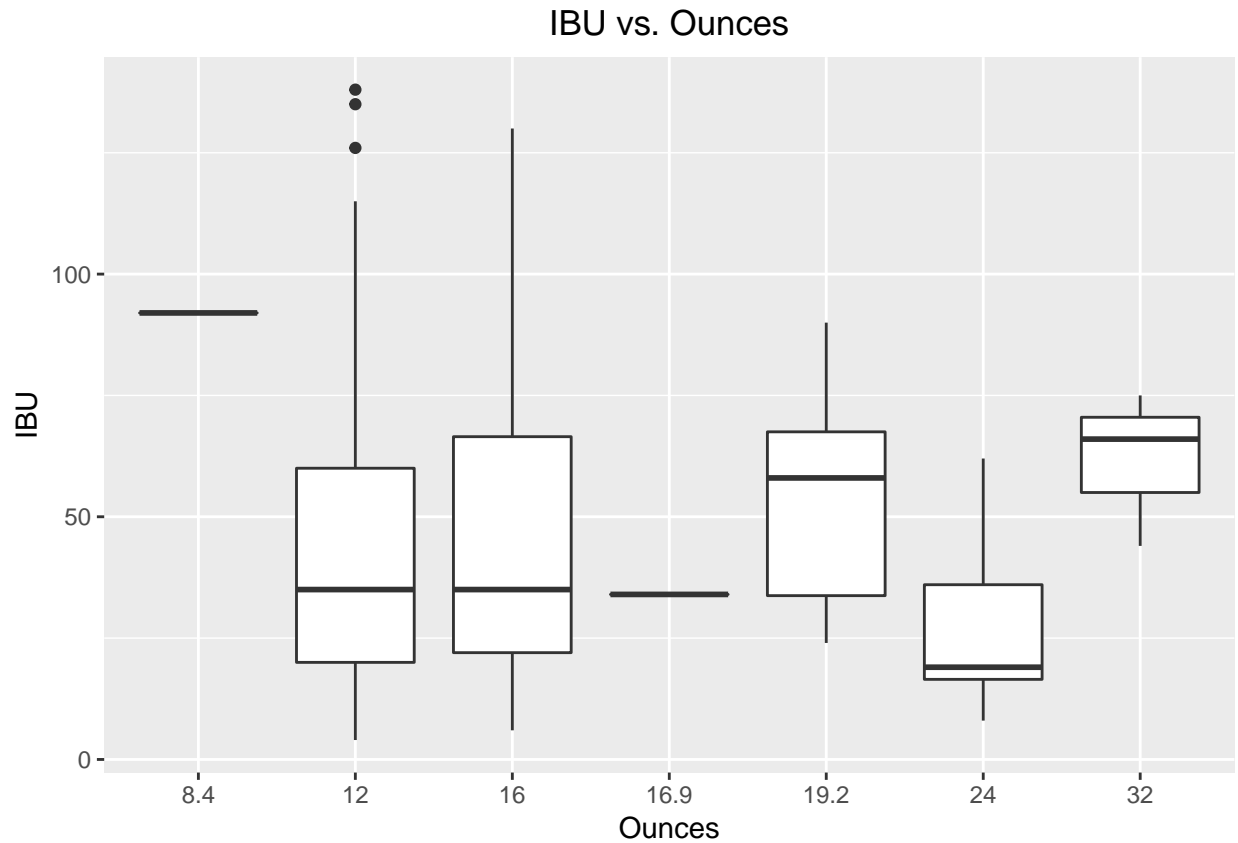
EDA











Label Creation - Reducing Factor Levels in Style

We created 8 main categories for beer type:

- Ale
- India Pale Ale
- Lager
- Kölsch
- Stout
- Wheat
- Fruit
- Other

```
##
##   ale  fruit   ipa kolsch  lager  other  stout  wheat
##   963   49   571   42    249   271   174    91
```

We also grouped the states into 4 main regions:

- West
- Northeast
- South
- North Central

Count of beers by region:

```
##
```

```
## North Central    Northeast    South    West
##           713           380           433           876
```

Count of breweries by region:

```
##
## North Central    Northeast    South    West
##           141           100           123           187
```

Is there a relationship between brewery location and beer types?

- We will conduct a chi-squared test for independence

```
##
## Pearson's Chi-squared test
##
## data:  regions and beer_type
## X-squared = 52.731, df = 21, p-value = 0.0001503
```

The result is a very low p-value, so we can conclude that there is dependence between regions and beer types.

To analyze the relationship between characteristics and region, let's map the beer alcohol content and IBU by region.

Random Forest

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    251     0    101     0      2      5      0      0

##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    143     4     93     7     36     36     29     11

## [1] 0.5682451
```

Using Names of Beers

Using only the ABV, IBU, Ounces and Region of each beer, the model is able to predict the type of beer with 57.26% on the unseen test set. To try to improve this, I will perform a sentiment analysis on the names of the beer using the .

Random Forest

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    183     3    116     3     31     17     7      5

##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    127     5    110     4     50     34     19     16

## [1] 0.5917808
```

Including the sentiment of the names of the beers improves the accuracy of the model by about 5%.

Correspondance Analysis - Random Forest

```
##
##   ale  fruit   ipa kolsch  lager  other  stout  wheat
##   175     3   109     2    18    16     6     3

##
##   ale  fruit   ipa kolsch  lager  other  stout  wheat
##   117     8   111     3    28    26    21    18

## [1] 0.6054217
```