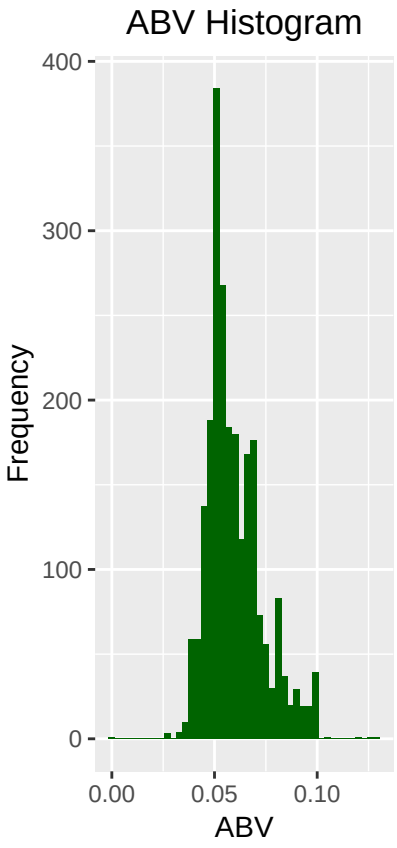
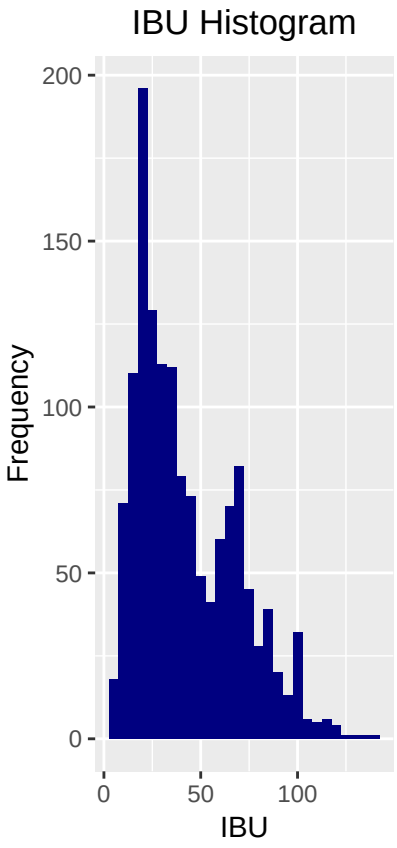
Craft Beer

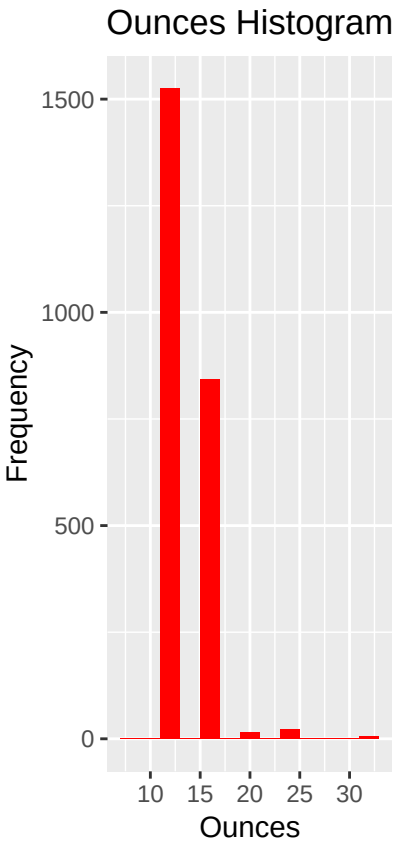
Santiago Tellez

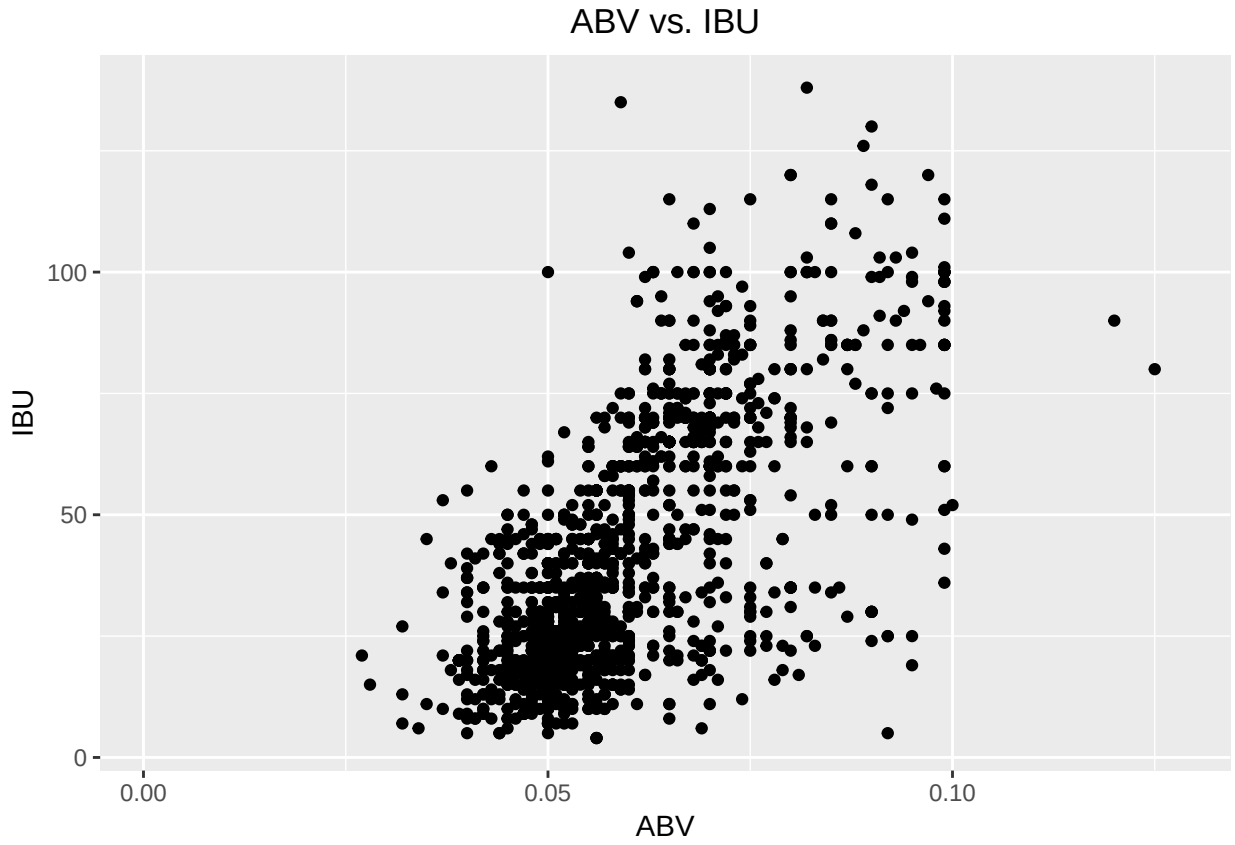
1/25/2019

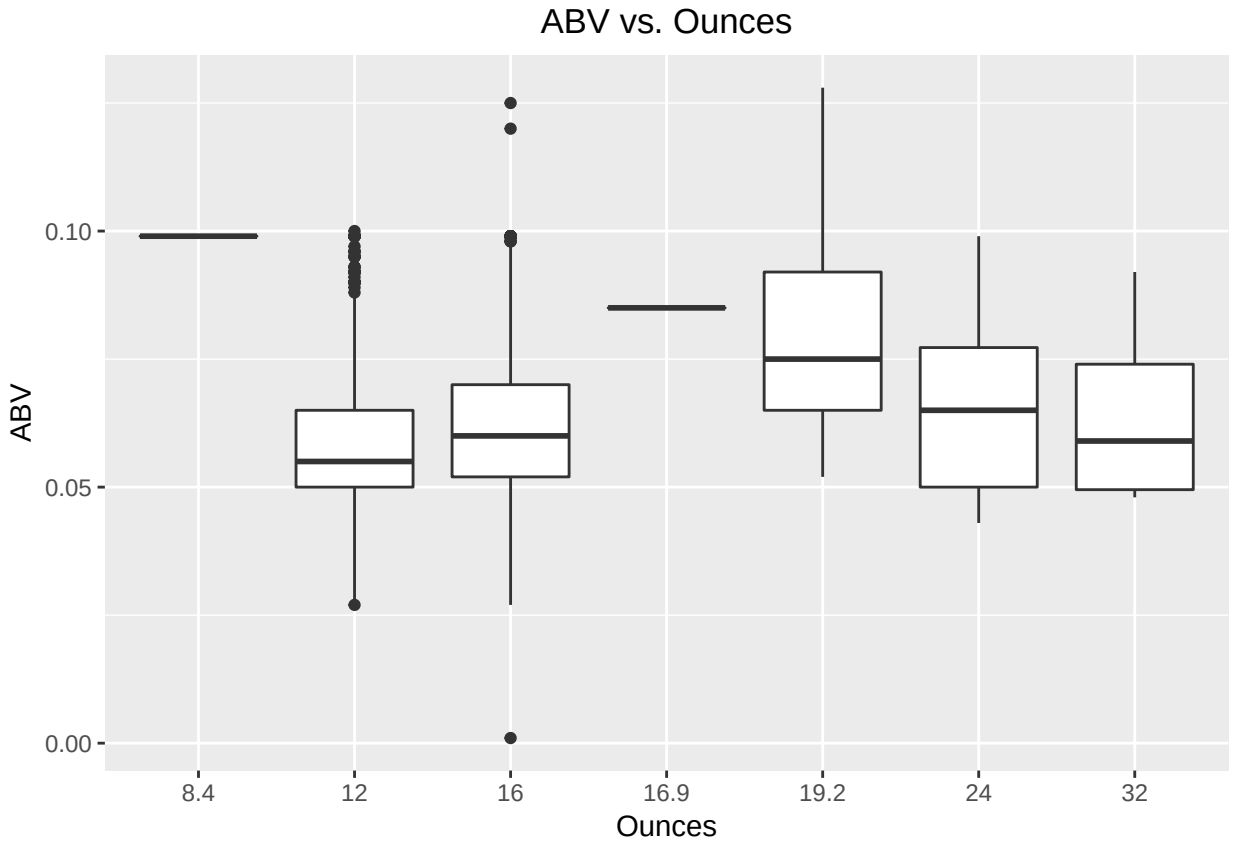
EDA

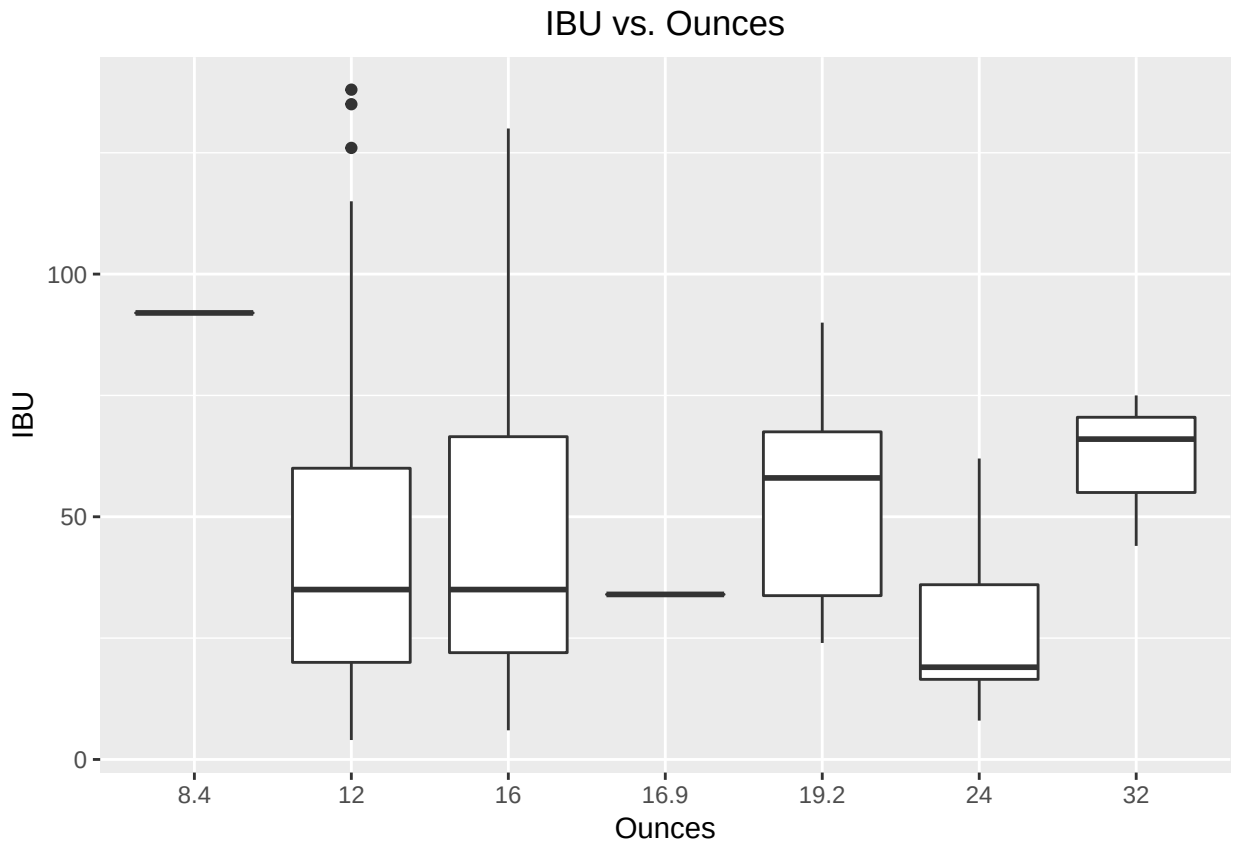







ABV vs. IBU
IBU
100
50
0
0.00
0.05
0.10
ABV


ABV vs. Ounces
ABV
0.10
0.05
0.00
8.4
12
16
16.9
19.2
24
32
Ounces



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 alochol content and IBU by region.

Random Forest

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    254      0    107      0      3      2      0      1
```

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    142      5    100      8     40     37     24     11
```

```
## [1] 0.5722071
```

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
##    188      4    102      2     25     18      8      9
```

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    140      7    102      9     28     36     18     16
```

```
## [1] 0.5983146
```

Including the sentiment of the names of the beers improves the accuracy of the model by about 5%. Next I will see if a correspondance analysis on the sentiments of each beer name adds predictive power to the model.

Correspondance Analysis - Random Forest

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    193      3    105      1     13     10      1      3
```

```
##
##    ale  fruit    ipa kolsch  lager  other  stout  wheat
##    127      9    101      8     30     30     16      8
```

```
## [1] 0.6534954
```

The accuracy on the test set drops to 60%, still an improvement on the model without using sentiment analysis but not better than the previous model.