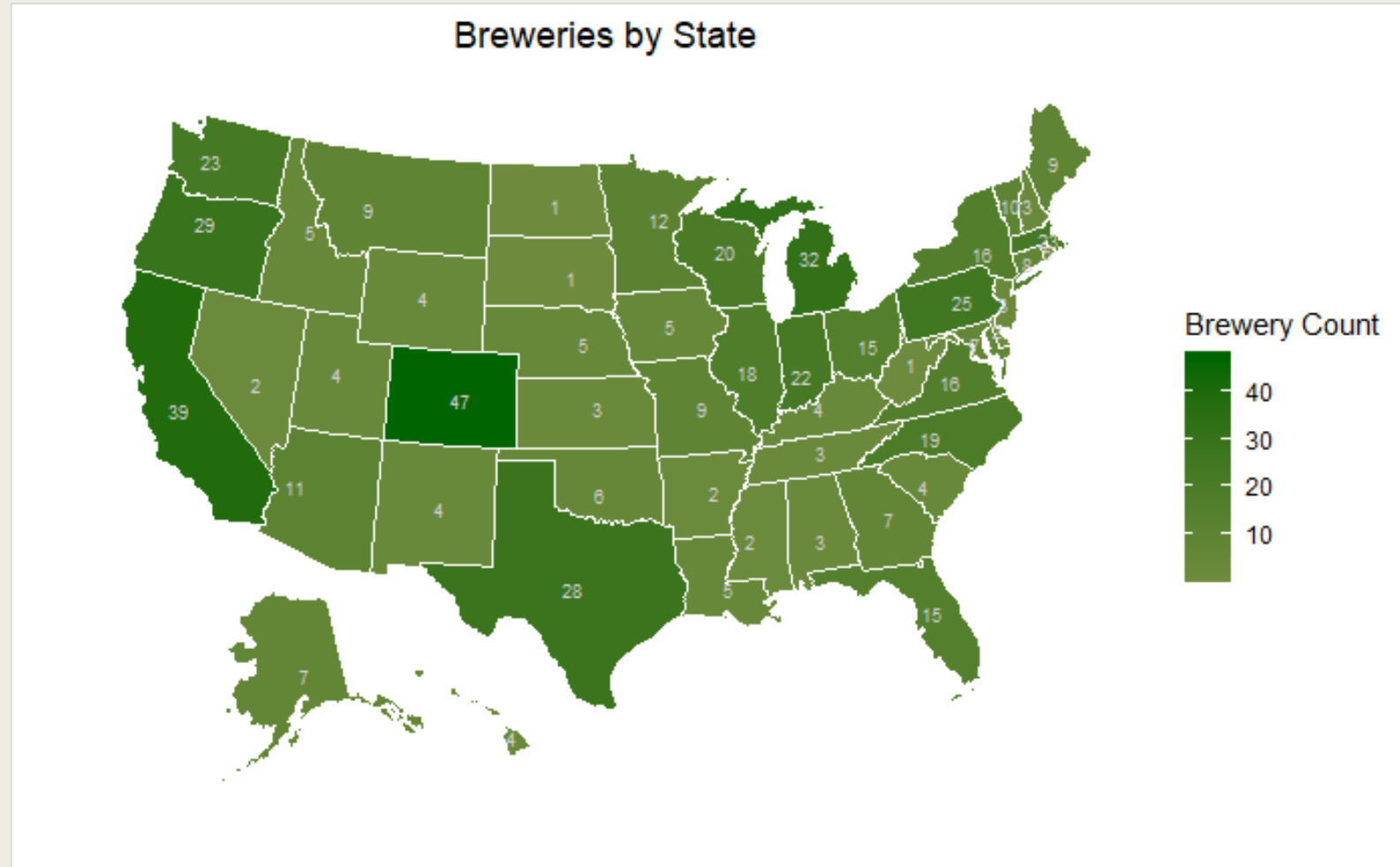# US CRAFT BEERS & BREWERIES

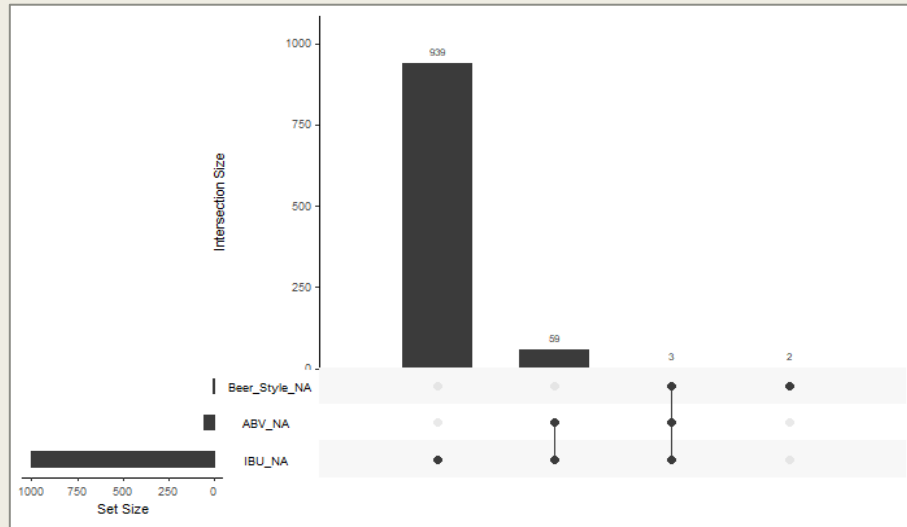## EXPLORATORY DATA ANALYSIS

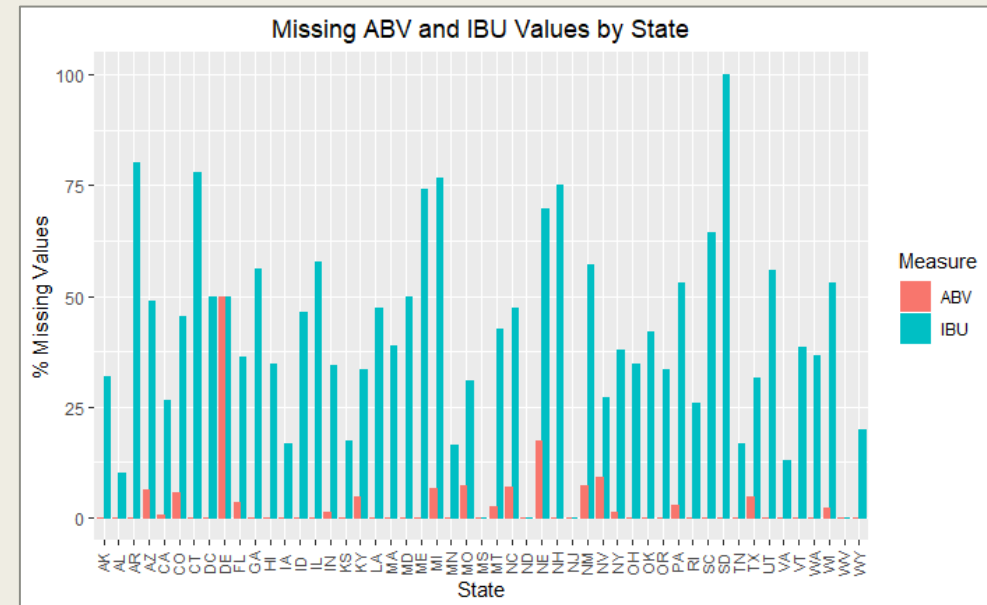Anh Phan

Where are the craft beer breweries located?

We have a total of 558 Breweries that make 2,377 different beers. Top 10 States account for more than 90% of total beers produced and represent more than 40 % of breweries in the US market.
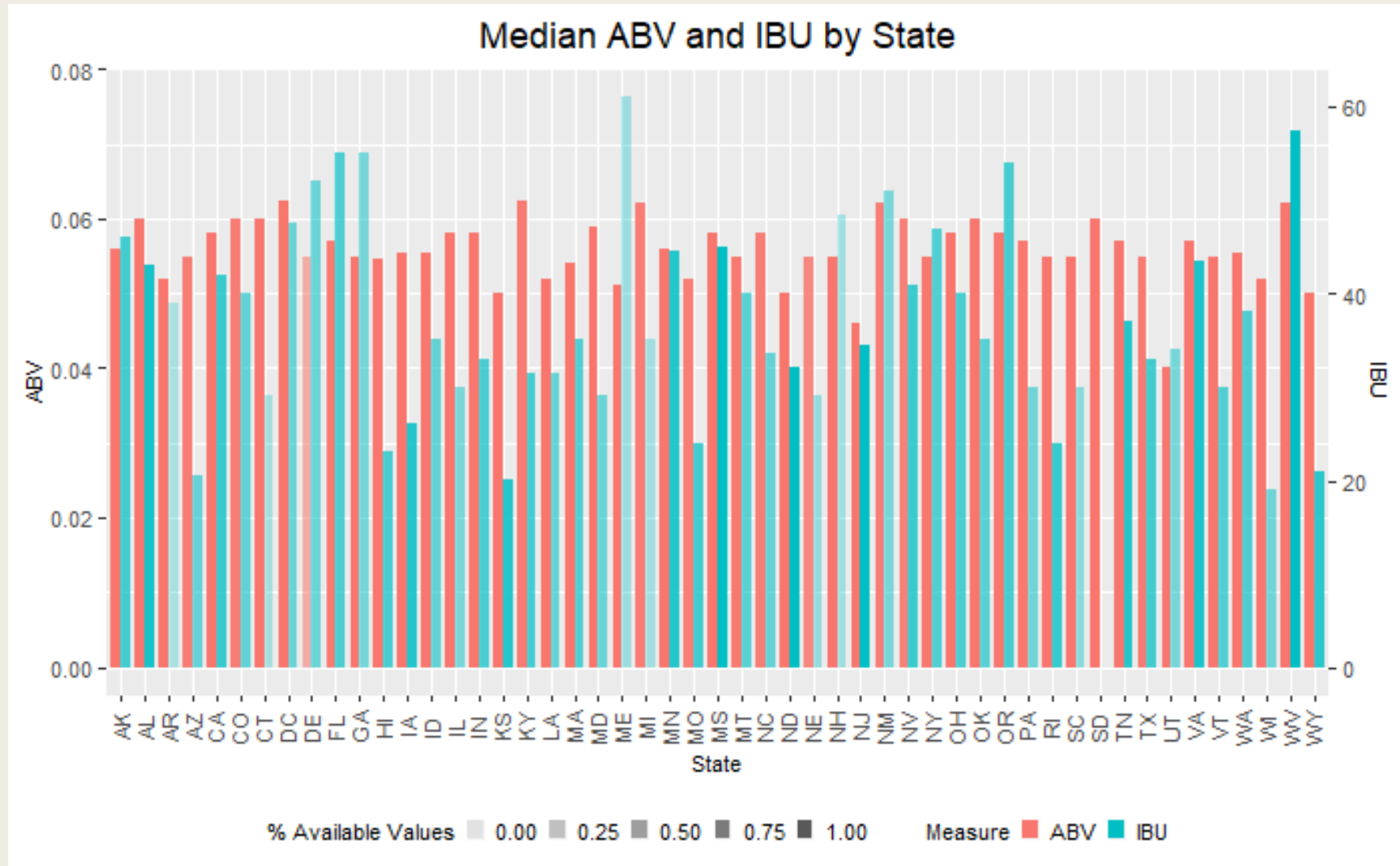
What's the quality of data we have?

- We have significant number of values missing (~42%) for IBU measure; we need to keep this in mind for any models developed using IBU
- There are a few values missing for ABV measure (less than 3%; these should not impact our inferences much
- A very few values are missing for Beer Style, but we do not need this attribute for our current analysis

- A significant number of states have more than 50% of the observations missing value for IBU; this significantly impacts any inferences based on state level IBU values
- For South Dakota, all the observations are missing value for IBU
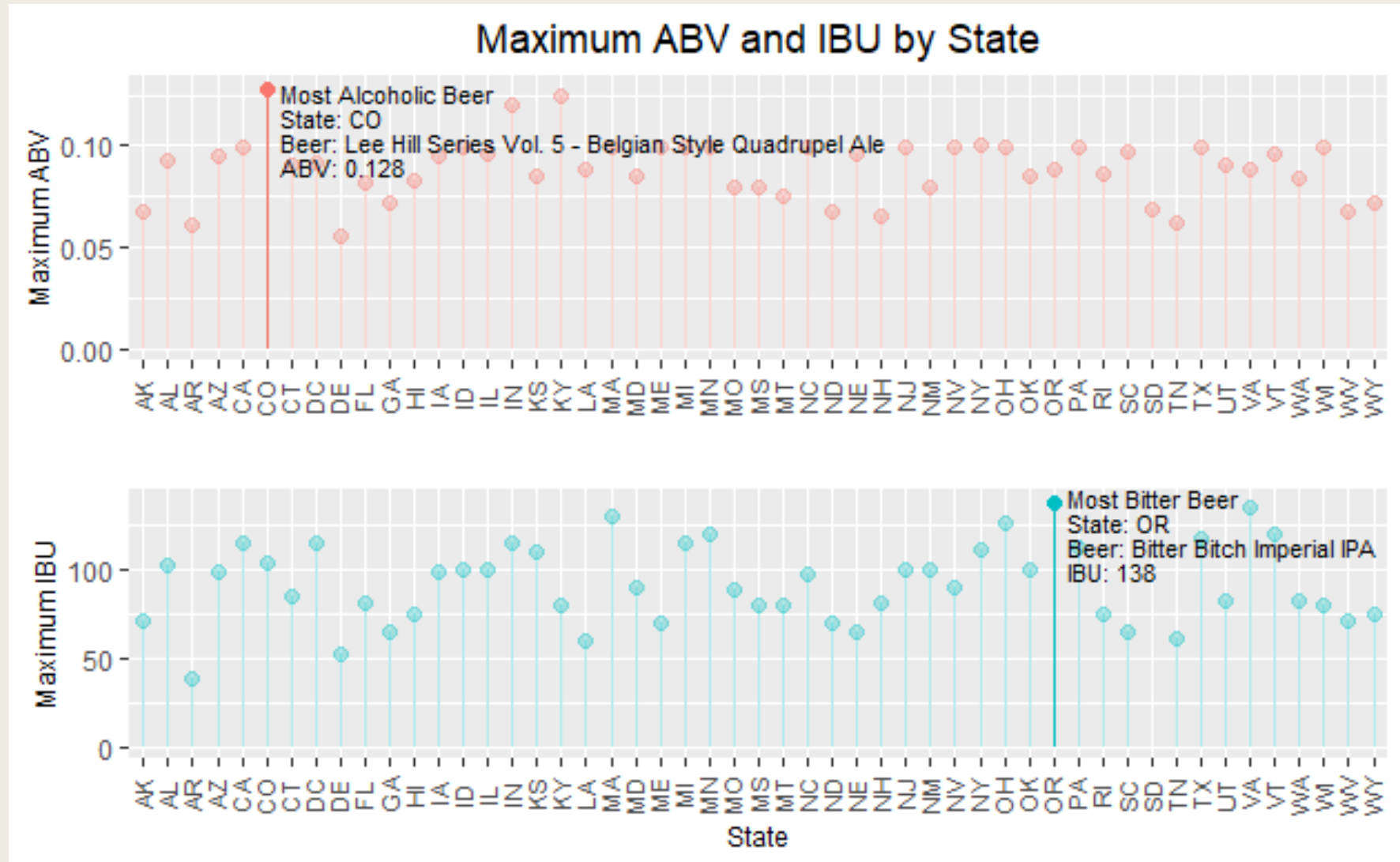- We intend to ignore the missing values for further analysis purposes
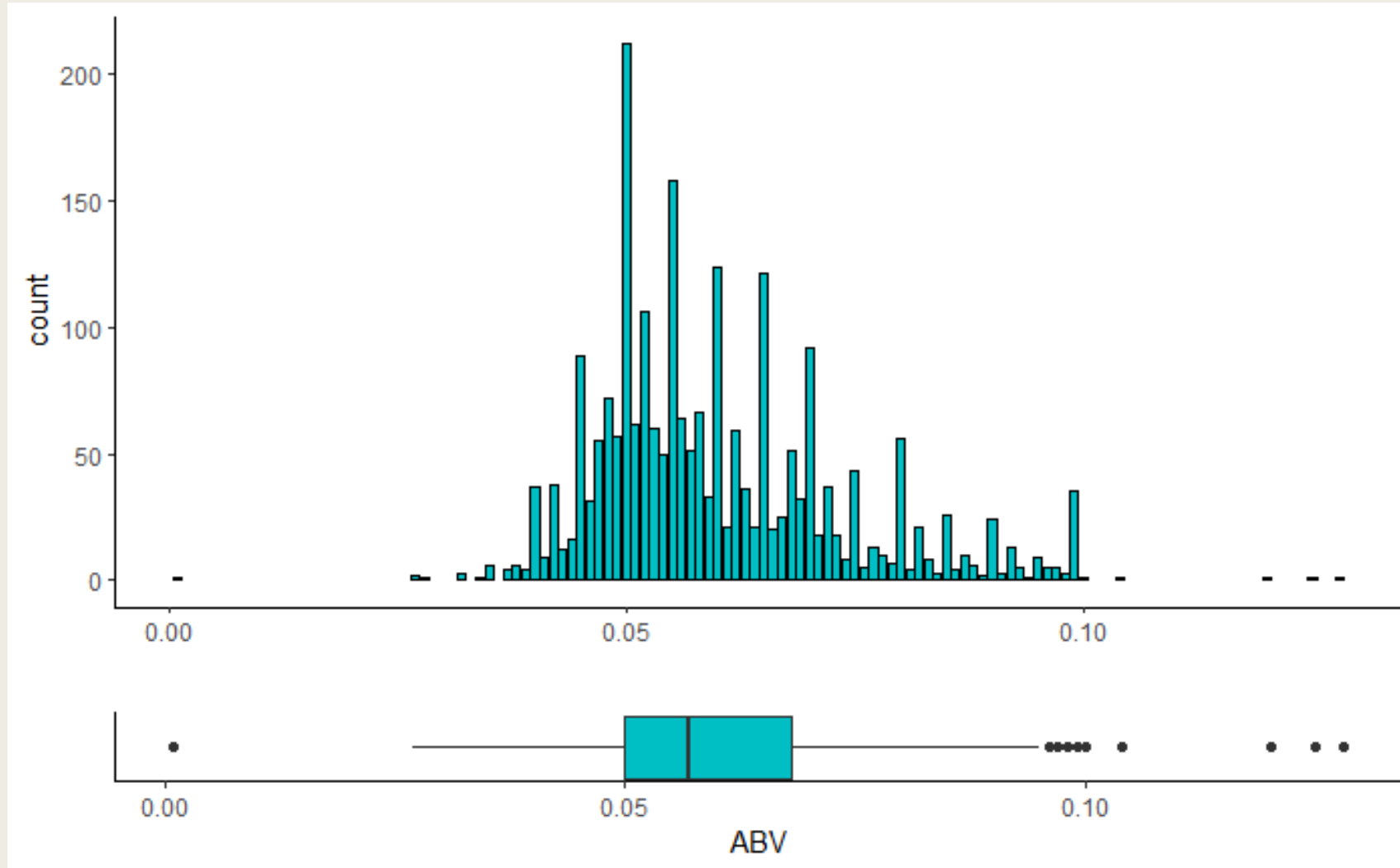
What are median beer measures by State?

Median values of ABV do not seem to vary a lot across States; however, median IBU values seem to have a bit of variance.
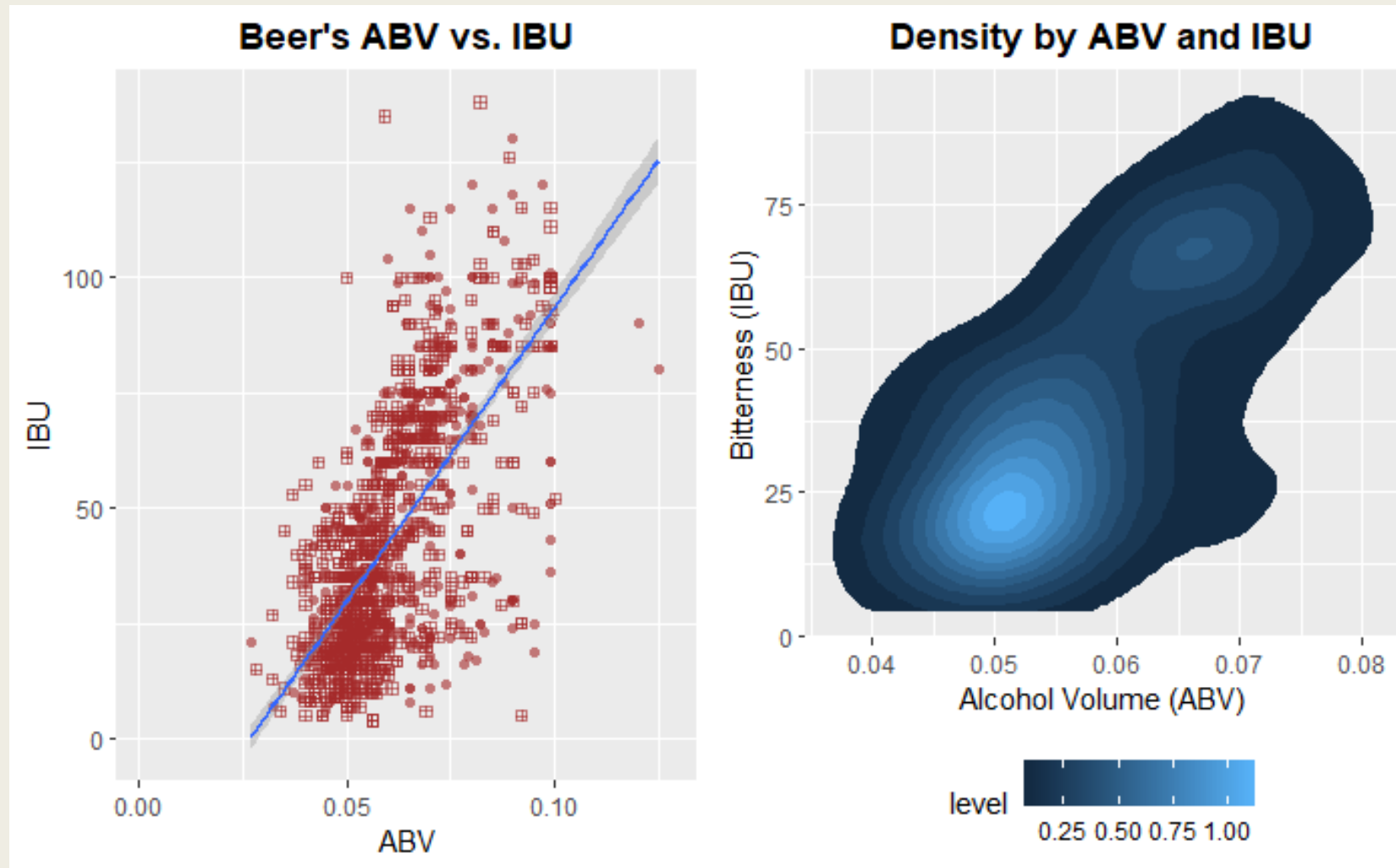
Which are our most strong and bitter beers?

Even though Maine has a highest value of median IBU, the most bitter beer is produced in Oregon.

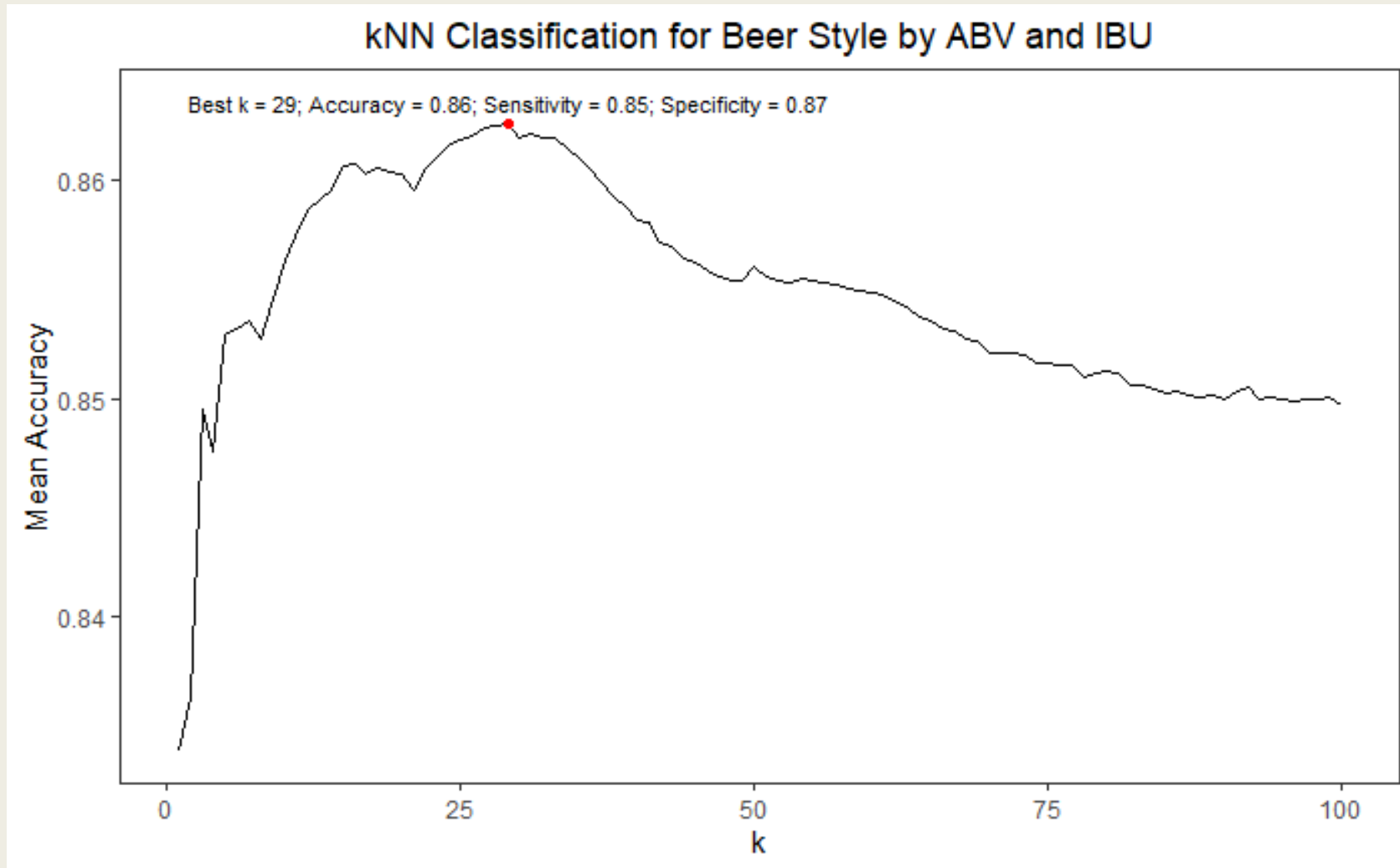# What is the distribution of ABV of our beers?

The national median ABV value is around 5.7%; as we can see from the chart above, most of the beers are concentrated around 4%-7% ABV range. There are a few beers with outlier values which impact the mean ABV value. As such, median ABV is a better center measure for any statistical analyses.
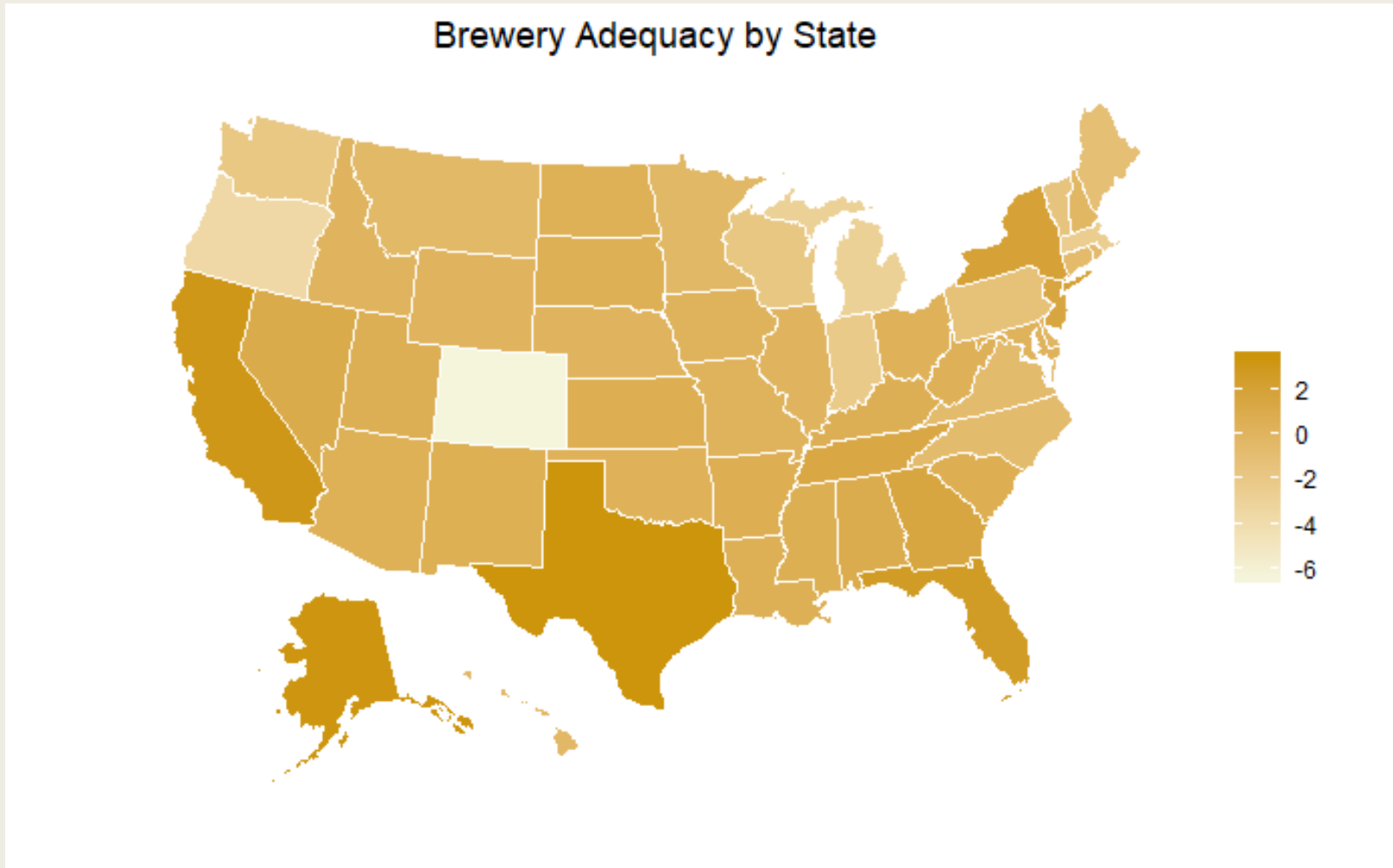
Is there a correlation between ABV and IBU?

There is a moderate positive correlation between ABV and IBU measures for the beers nationwide. In addition, most of the beers are concentrated around 5% ABV, and very few beers go beyond 10% ABV level. The density plot shows two clusters that are centered around (5% ABV, ~24 IBU) and (6% ABV, ~73 IBU); it will be beneficial to focus on pricing and promotion strategies for beers with these measures.

Can we classify beer style using ABV and IBU?

A KNN classification model based on ABV and IBU provides us IPA or Other Ale classification results with good accuracy, sensitivity, and specificity.

Where can we setup additional breweries?

Brewery Adequacy by State

We have compared states' relative population and area vs. the percentage of breweries in the state to assess if the number of breweries in the state are adequate. Based on the above, there is a case for setting up additional breweries in TX, FL, NY, and NJ. On the other end, CO seems to have too many breweries for their population

Thank You

Anh Phan – anh.phan1101@gmail.com