U.S.A BEERS AND BREWERIES EXPLORATORY DATA ANALYSIS

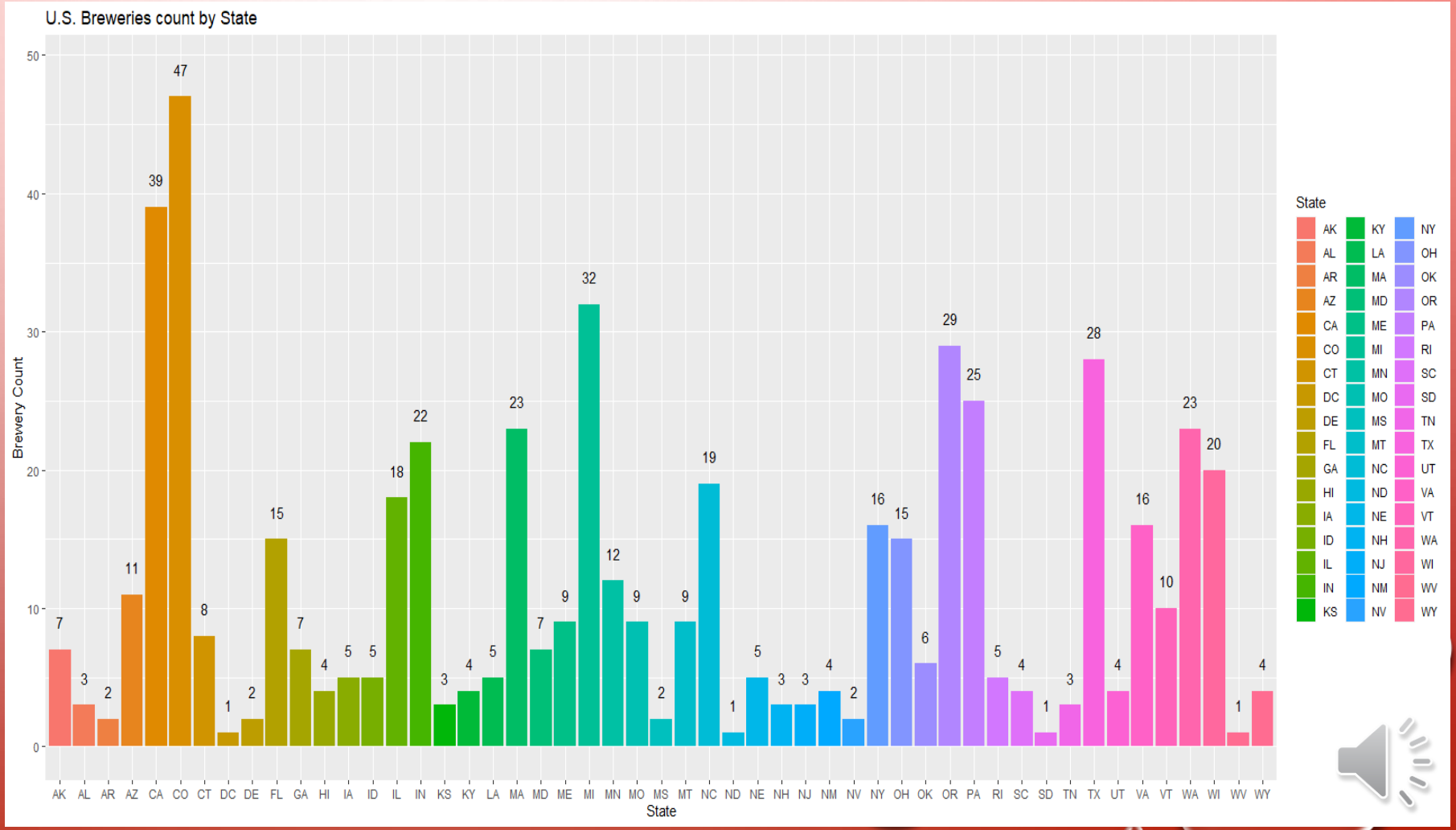
BY

DAWN BOWERMAN

U.S. BREWERIES COUNT BY STATE

Median and Mean Breweries per State in the U.S.:

State	n
Length:51	Min. : 1.00
Class :character	1st Qu.: 3.50
Mode :character	Median : 7.00
	Mean :10.94
	3rd Qu.:16.00
	Max. :47.00

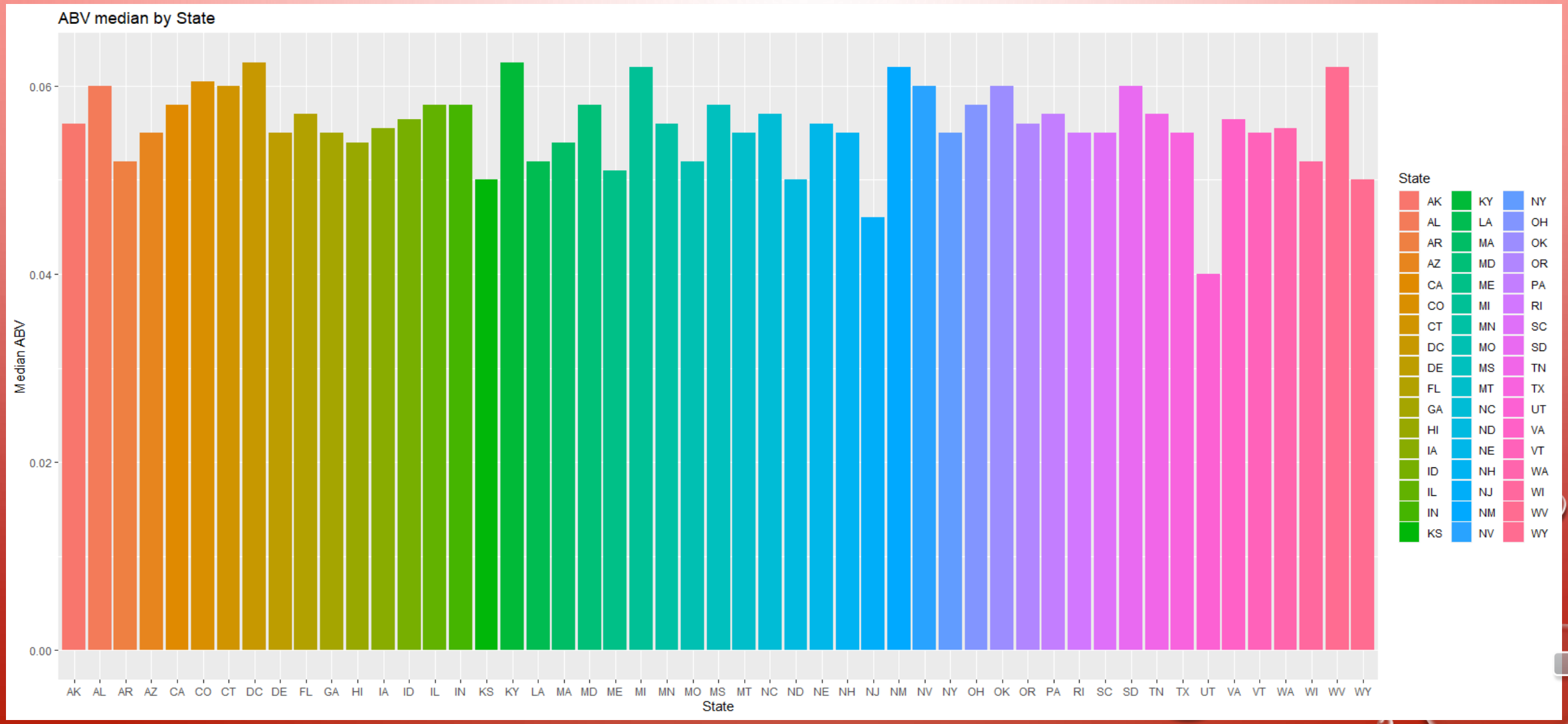


ADDRESSING THE MISSING VALUES (NAS)

- AFTER VIEWING THE AREAS WHERE THERE IS MISSING DATA, I HAVE DETERMINED TO DELETE THE NAS OUT OF SOME OF THE ANALYSIS (IBU - ALE COMPARISON, MEDIAN ALCOHOL) WHERE IT DOESN'T EFFECT THE RESULT AS MUCH AND LEAVE IN WHERE IT DOES (SUCH AS NUMBER OF BREWERIES.)
- THE MISSING VALUES ARE MCAR (MISSING COMPLETELY AT RANDOM) IN MY DETERMINATION AS THEIR MISSING VALUE HAS NOTHING TO DO WITH THEIR HYPOTHETICAL VALUE AND WITH THE VALUES OF OTHER VARIABLES.

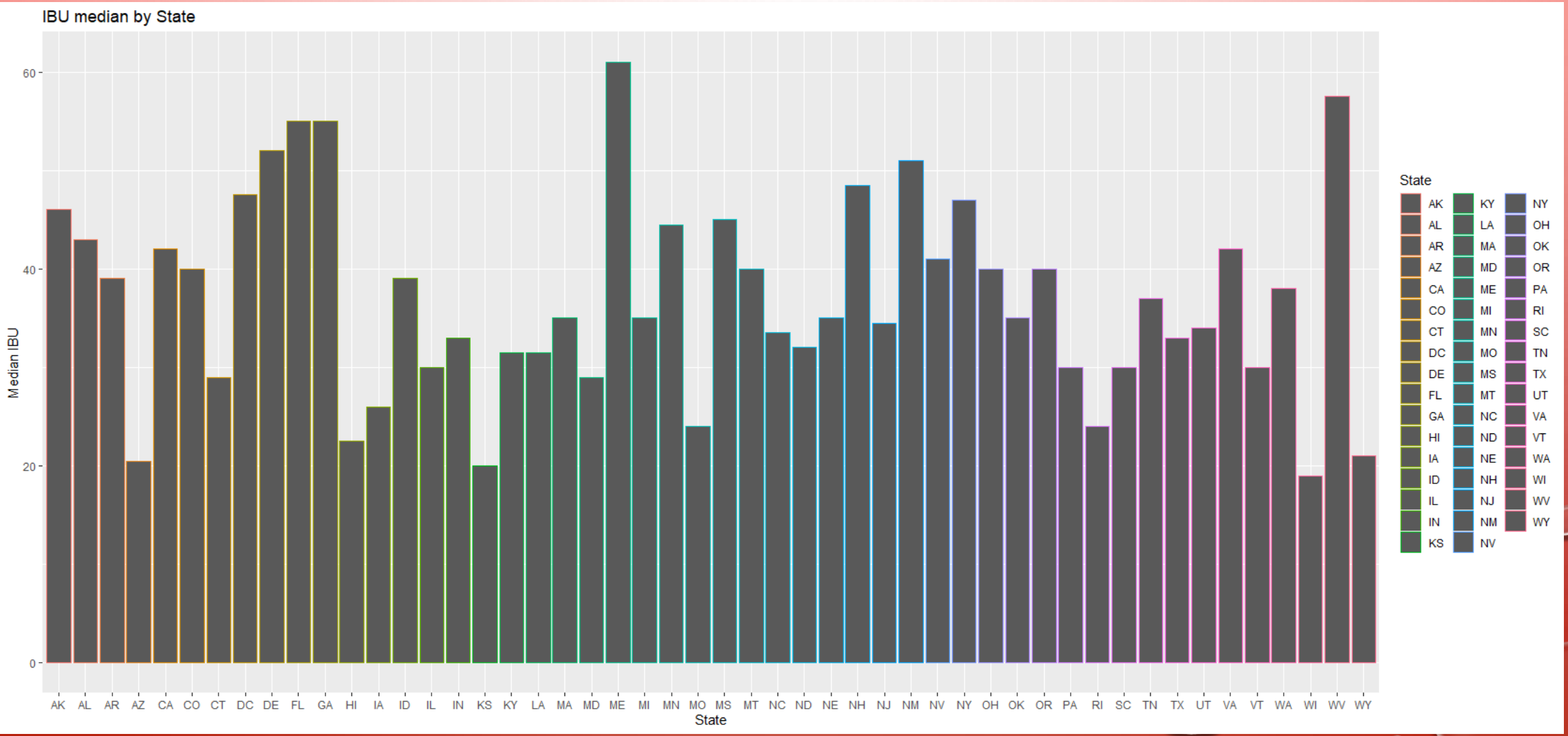
MEDIAN ALCOHOL CONTENT FOR EACH STATE;

UTAH AND NEW JERSEY ARE UNCHARACTERISTICALLY LOW, WHERE MOST ARE AROUND 0.055



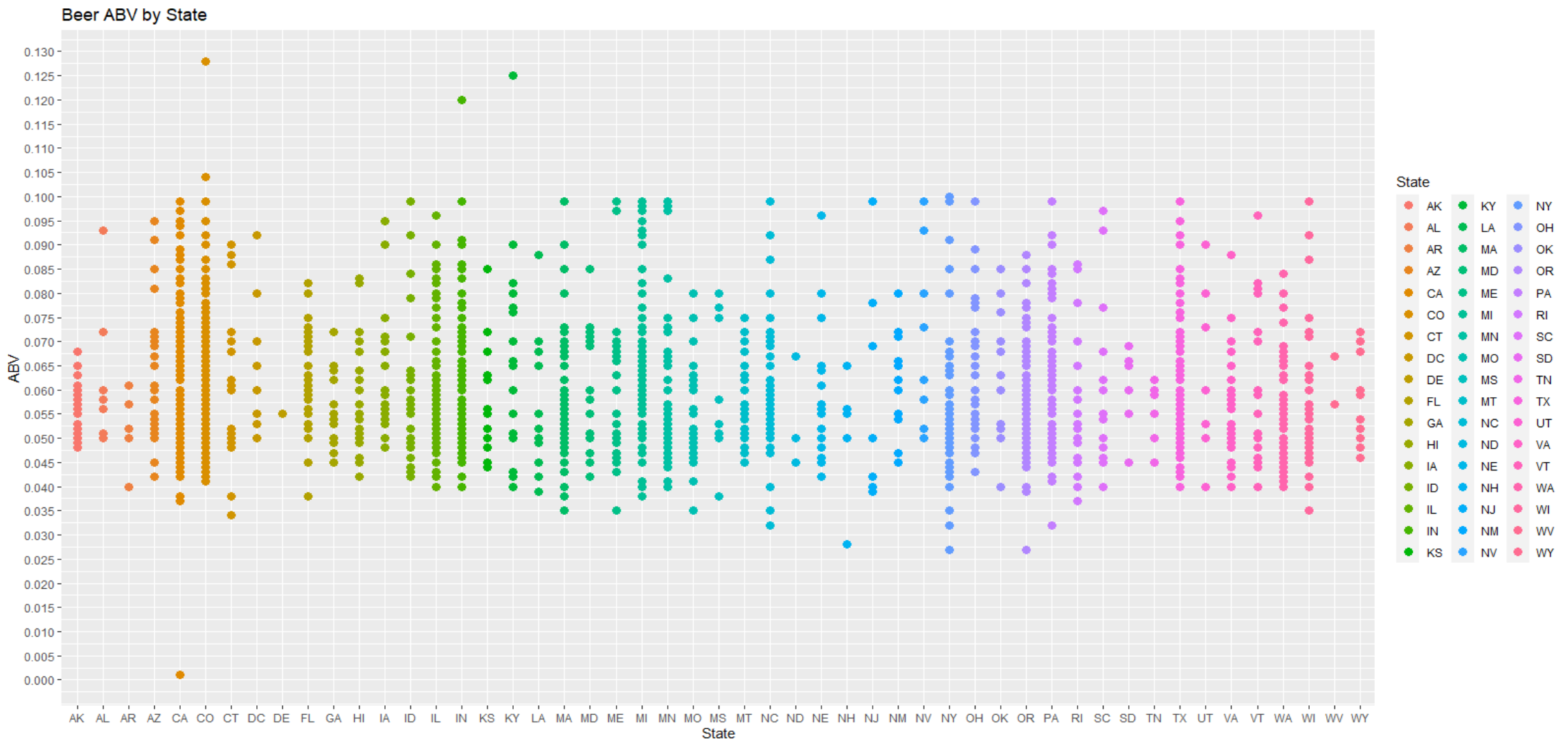
MEDIAN INTERNATIONAL BITTERNESS UNIT FOR EACH STATE;

MAINE AND WEST VIRGINIA ARE ATYPICALLY HIGH, WISCONSIN AND WYOMING ARE UNUSUALLY LOW



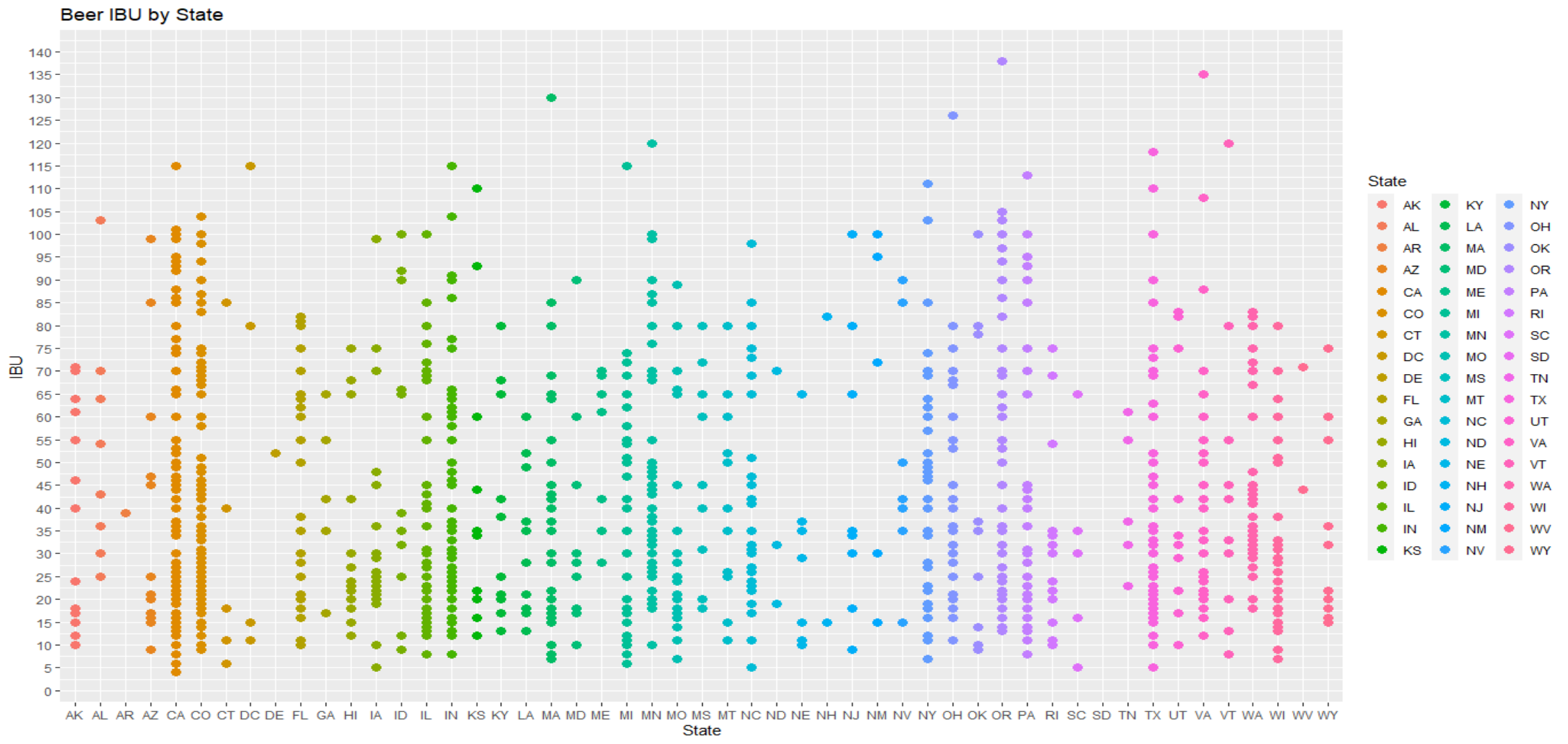
BEER ALCOHOL CONTENT BY STATE

STATE WITH THE MAXIMUM ALCOHOLIC (ABV) BEER IS COLORADO, WITH KENTUCKY, AND INDIANA CLOSE BEHIND. MOST BREWERIES TOP OUT AT 0.100 ABV.



BEER INTERNATIONAL BITTERNESS UNIT BY STATE

STATE WITH THE MOST BITTER(IBU) BEER IS OREGON, WITH VIRGINIA, AND MAINE RUNNERS UP. QUITE A BIT OF RANGE IN MOST STATES.

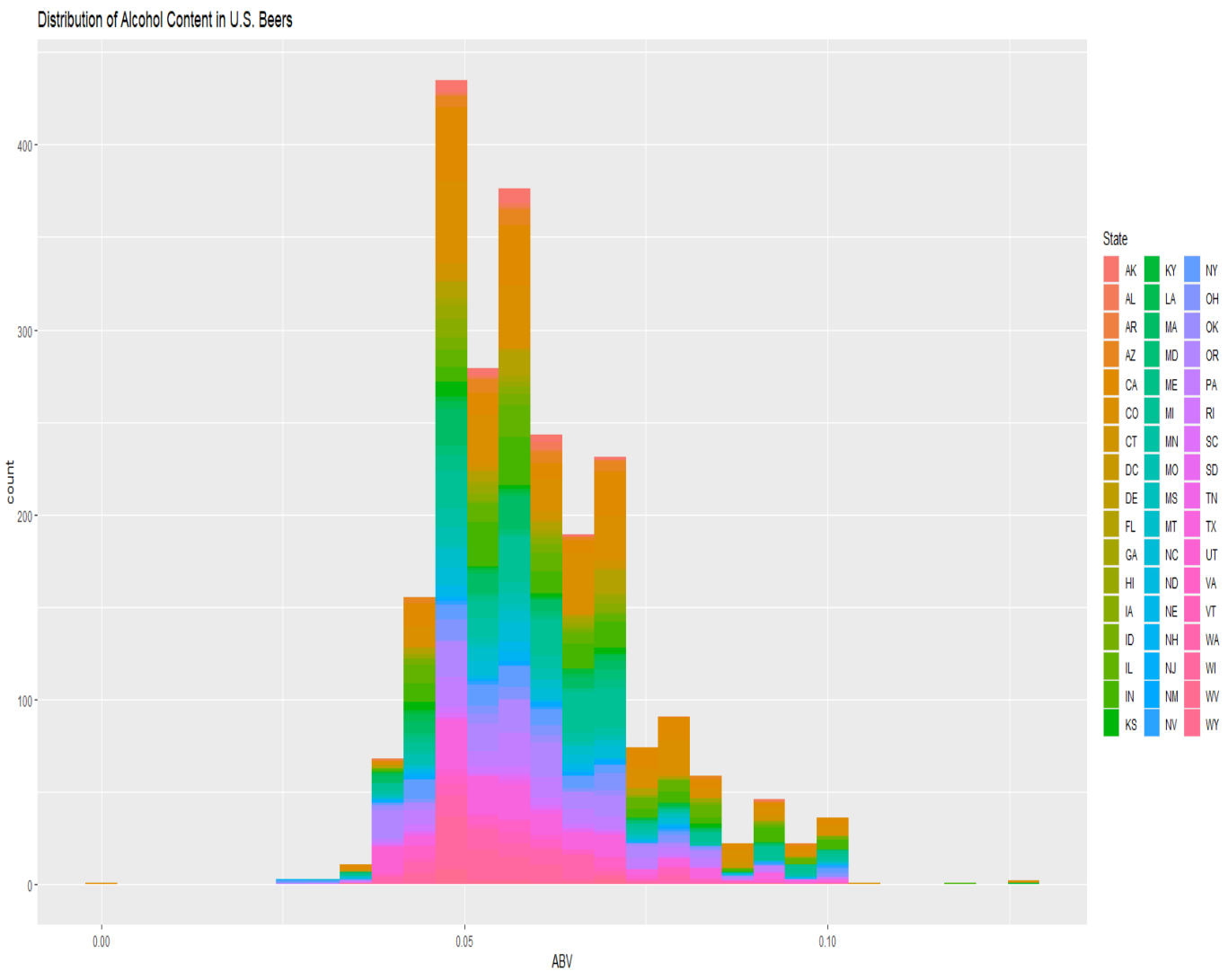


SUMMARY STATISTICS AND DISTRIBUTION OF THE ABV VARIABLE

Summary Statistics for Alcohol by Content of U.S. Breweries:

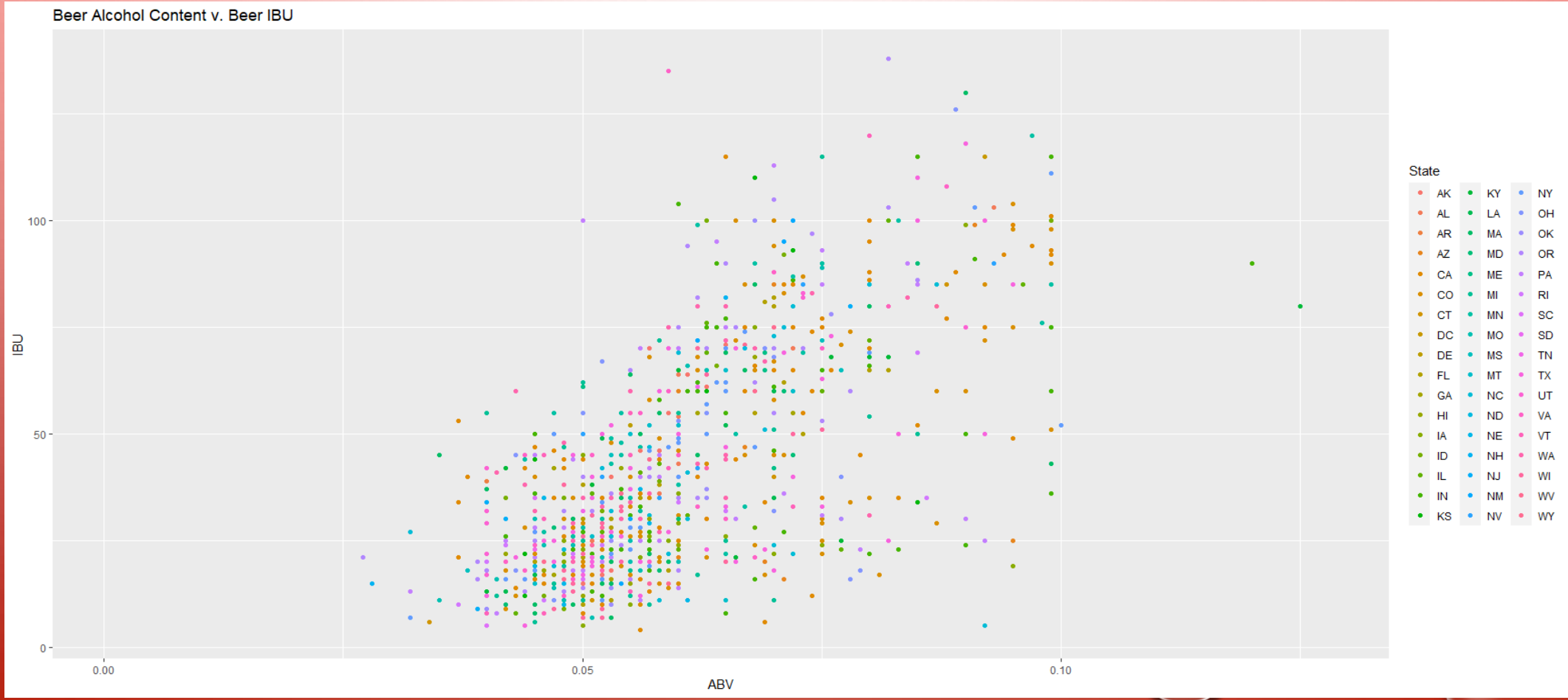
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.00100	0.05000	0.05600	0.05977	0.06700	0.12800	62

The distribution of the Alcohol by Volume variable is right-skewed and the majority ends at a little over 0.10 and then has a few outliers after that.



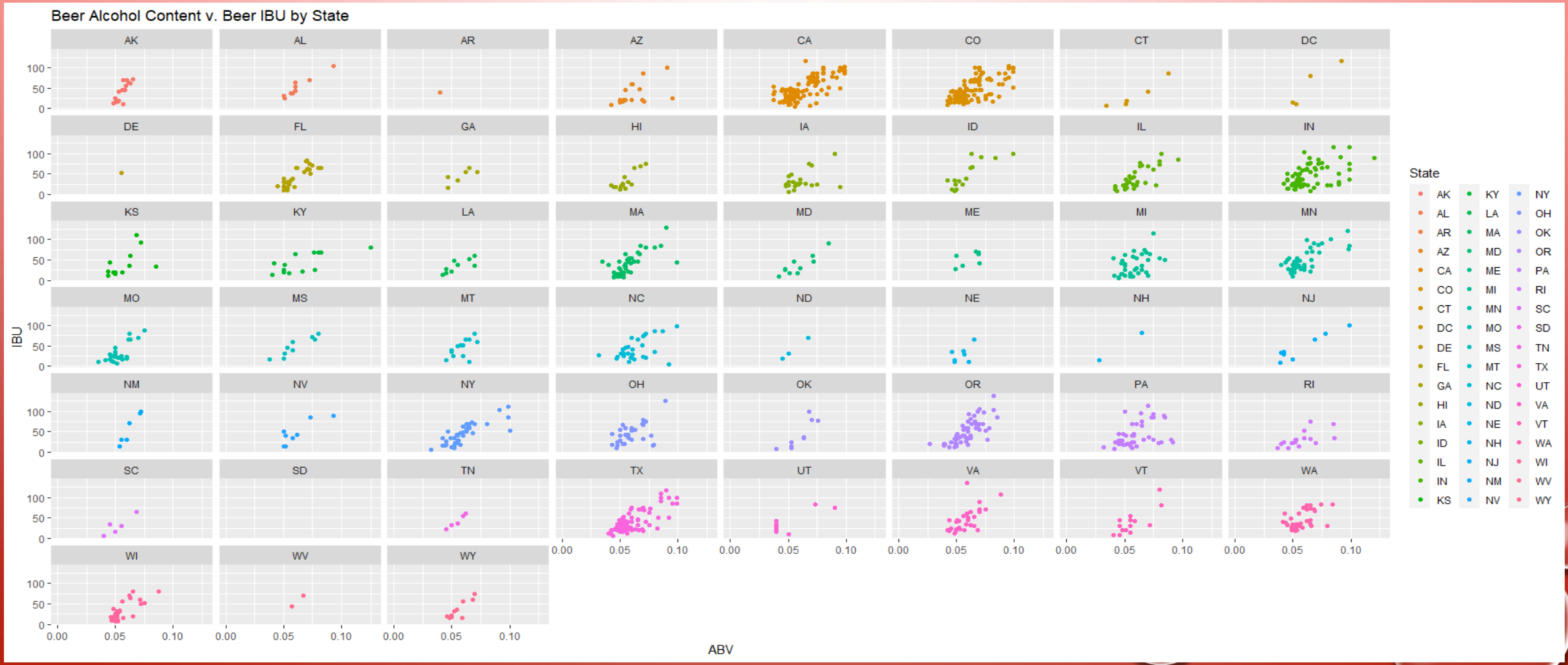
BEER ALCOHOL CONTENT VS. BEER IBU

LOOKING AT THE GRAPH, THE RELATIONSHIP BETWEEN ABV AND IBU DOES LOOK LINEAR, HOWEVER THERE IS A DENSE NODE AT 0.05 ABV AND 20-25 IBU INDICATING THAT MOST BEERS ARE CREATED WITH THESE CHARACTERISTICS. NOTICE ALSO, THAT THERE ARE SOME HIGH ABV BEERS THAT ARE NOT AS BITTER; THIS COULD BE BEERS THAT DON'T HAVE AS MUCH HOPS IN THEM OR WERE CREATED TO BE LESS BITTER.



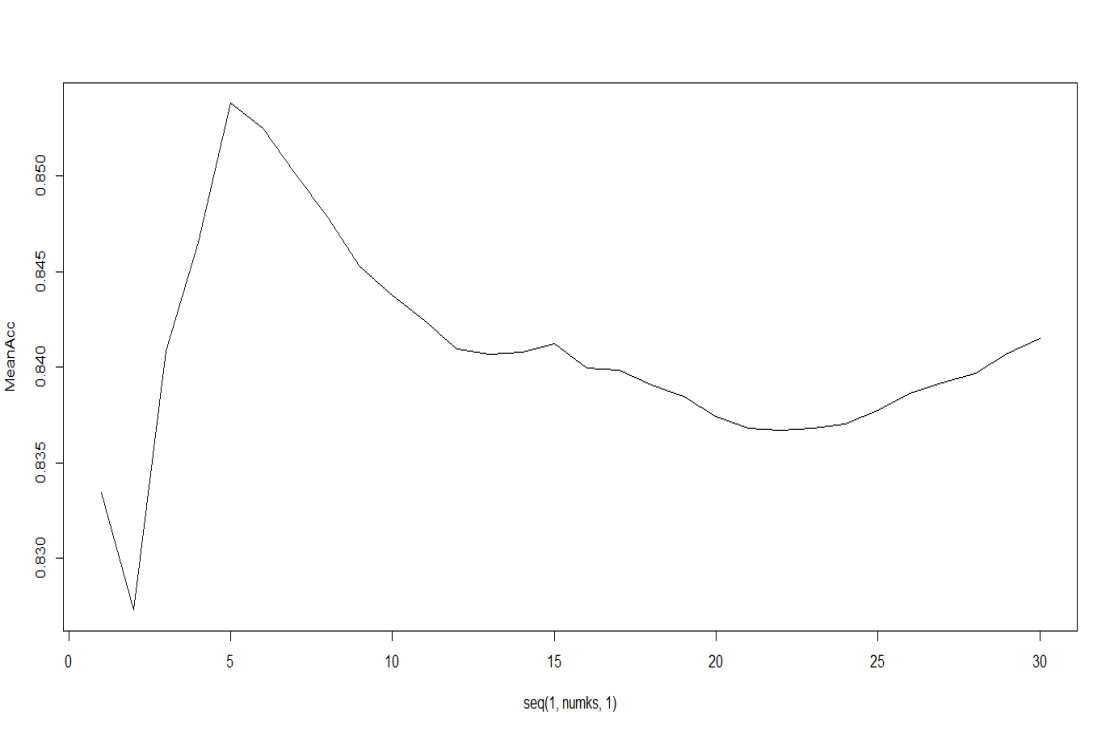
ABV VS. IBU BY STATE

IF YOU LOOK AT EACH STATES' PLOT SOME ARE VERY LINEAR WHERE OTHERS TEND TO DEVIATE. EACH STATE HAS A UNIQUE VERSION.



KNN CLASSIFICATION OF INDIA PALE ALES AND OTHER TYPES OF ALE BY THE CONTENT OF ALCOHOL BY VOLUME (ABV) AND THE INTERNATIONAL BITTERNESS UNIT (IBU) – SLIDE 1

- To investigate the difference with respect to IBU and ABV between IPAs (India Pale Ales) and other types of Ale, we use KNN classification to investigate this relationship.
- First, taking 500 tests to check for the best k value (nearest neighbors) to get the best results for our classification.
- The best k value to use for the KNN test is the highest point on the graph, which is 6.



KNN CLASSIFICATION OF INDIA PALE ALES AND OTHER TYPES OF ALE BY THE CONTENT OF ALCOHOL BY VOLUME (ABV) AND THE INTERNATIONAL BITTERNESS UNIT (IBU) –SLIDE 2

- We use the best determined k value to create a table and a confusion matrix that analyzes the ABV and IBU data to predict the "Ale" beers and the IPAs.

- The KNN classification analysis is very robust and shows that there are consistent differences in alcohol content and IBU between the two beer types.

classifications

	Ales	IPAs
Ales	**130**	17
IPAs	18	**73**

Confusion Matrix and Statistics

classifications

	Ales	IPAs
Ales	130	17
IPAs	18	73

Accuracy : 0.8529
95% CI : (0.8015, 0.8954)
No Information Rate : 0.6218
P-Value [Acc > NIR] : 3.175e-15

Kappa : 0.688

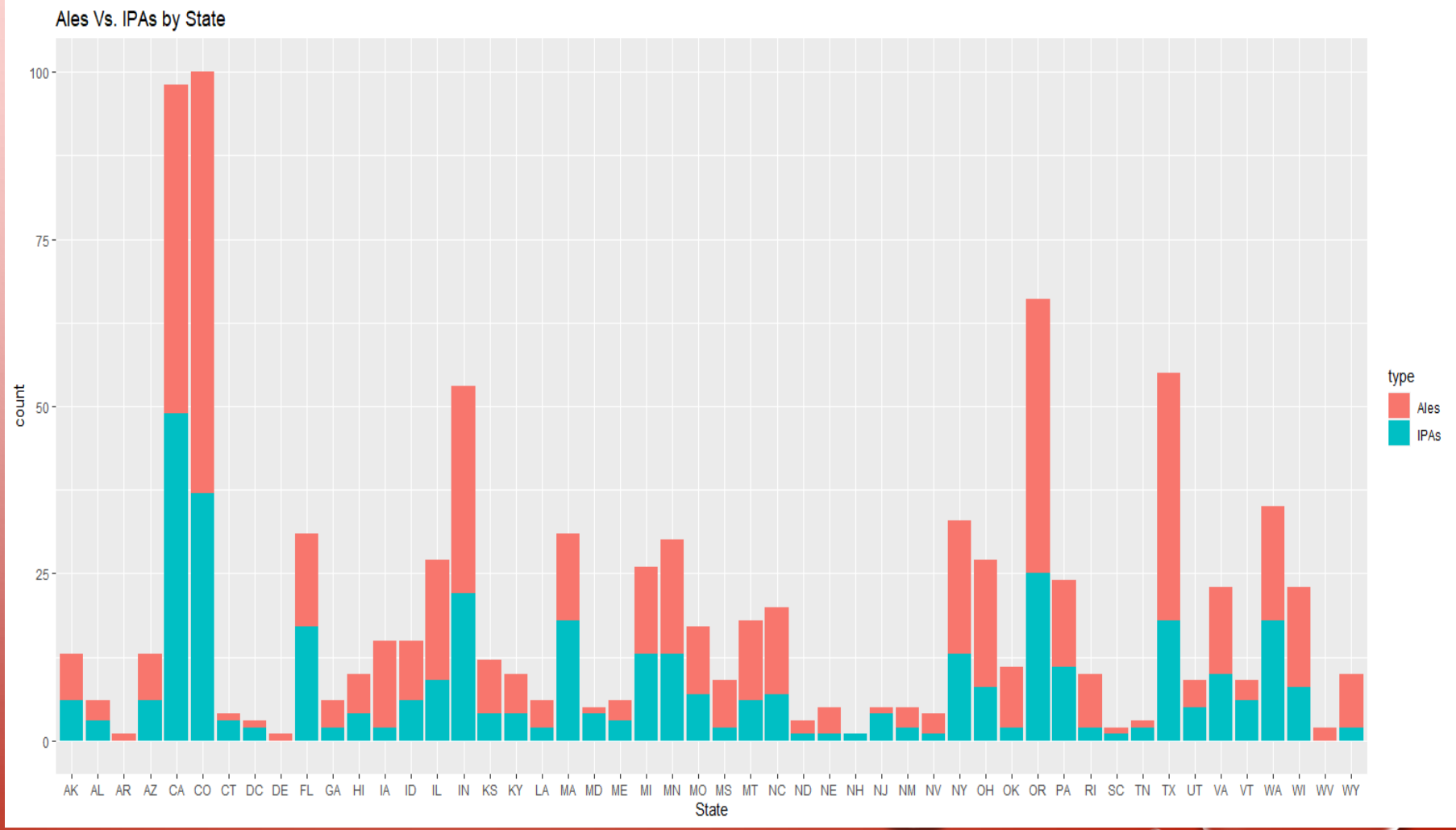
Mcnemar's Test P-Value : 1

Sensitivity : 0.8784
Specificity : 0.8111
Pos Pred Value : 0.8844
Neg Pred Value : 0.8022
Prevalence : 0.6218
Detection Rate : 0.5462
Detection Prevalence : 0.6176
Balanced Accuracy : 0.8447

'Positive' Class : Ales

ALES VS. IPAS BY STATE

Most states favor slightly Ales over IPAs. However California, Connecticut, DC, Massachusetts, Maryland, New Hampshire, New Jersey, Vermont, Tennessee, and Washington state have more IPAs than Ales. These happen to be most of the coastal states with the exception of Tennessee.



SUMMARY

- MOST BEERS ARE AROUND 0.05 ABV AND 20-25 IBU.
- THE RELATIONSHIP BETWEEN ABV AND IBU IS SOMEWHAT LINEAR WITH SOME DEVIATIONS.
- COLORADO AND CALIFORNIA HAVE THE MOST BREWERIES AND D.C. AND NORTH DAKOTA AND SOUTH DAKOTA HAVE THE LEAST.
- COLORADO HAS THE MAXIMUM ABV AND OREGON HAS THE MOST BITTER.
- MEDIAN ABV IS 0.056, WITH MIN = 0.001 AND MAX 0.128.

THE END