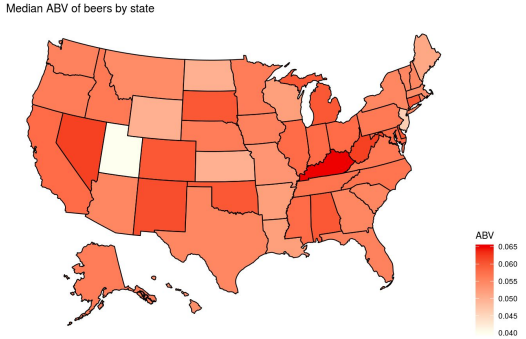


Beers in the United States

Preliminary Analysis prepared for
Budweiser CEO and CFO
17 October 19
Steven Garrity and Richard Palmer

Budweiser

DataScience@SMU

Introduction

We were presented with data on 2410 beers produced by 558 breweries in the United States.

For each brewery, we have its name, and the city and state it is located in.

For each beer, we have its style as well as numeric values for ABV (alcohol by volume), IBU (international bitterness units), and the serving size in ounces.

In this dataset, there were 100 different styles of beer represented.

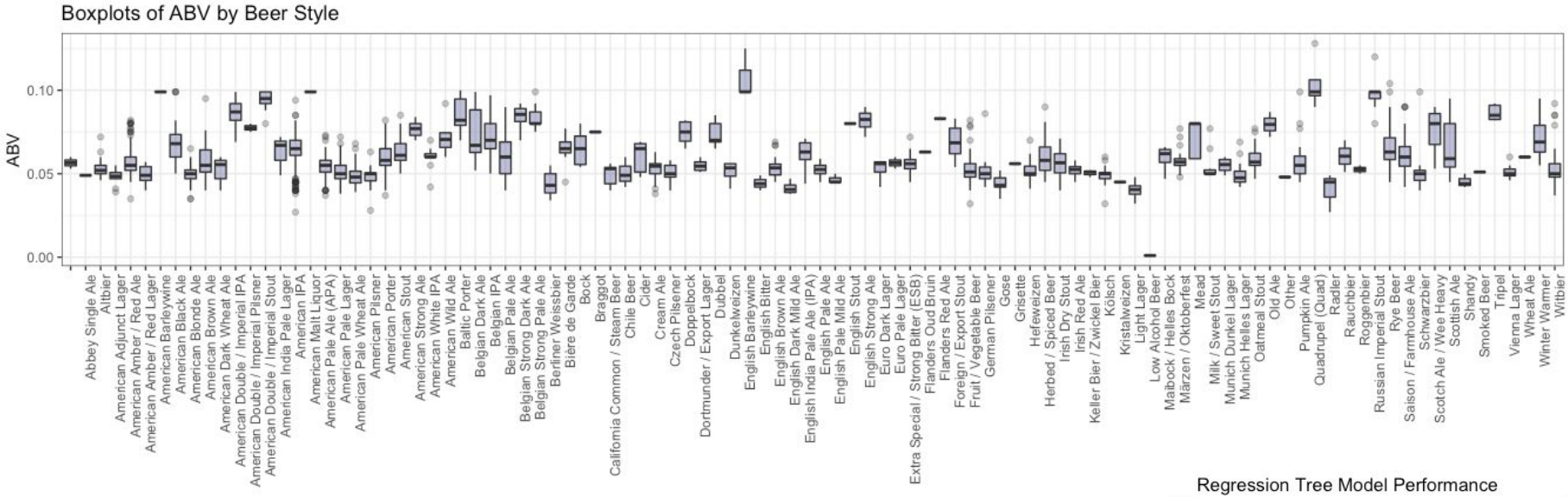
There were beers produced in every state of the nation, as well as in the District of Columbia.

Today we will present a summary of what we have found in the data so far, including

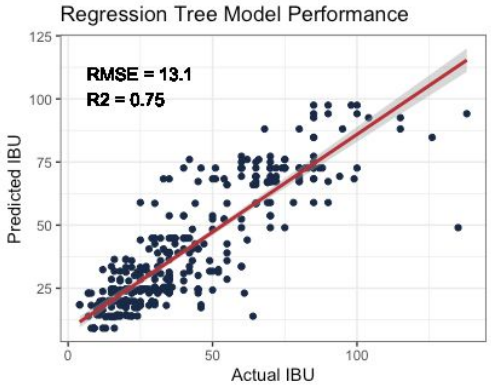
- Count of breweries by state
- Study of ABV and IBU by state
- Distribution of ABV and its relationship to IBU
- Examination of missing data and actions to address it

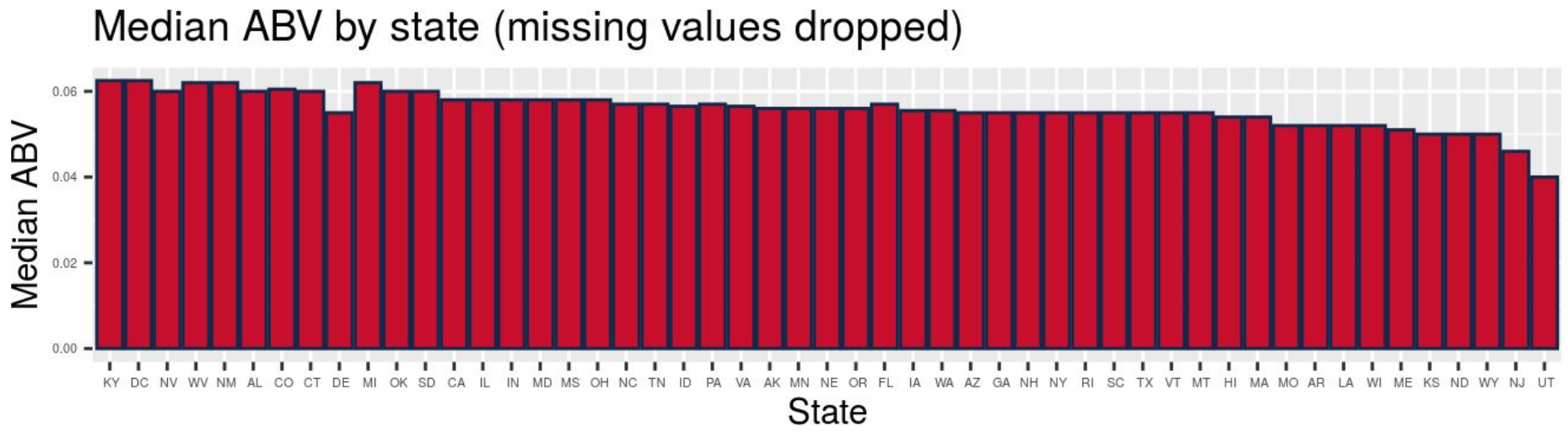
QUESTION 1 - BREWERIES BY STATE

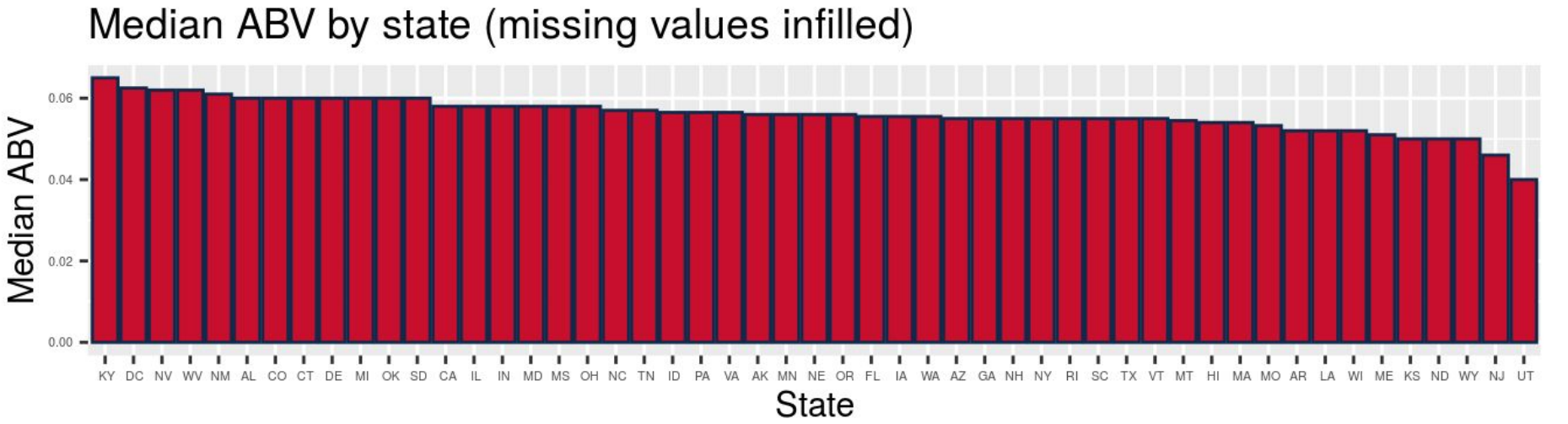
QUESTION 3 - ADDRESS THE MISSING VALUES

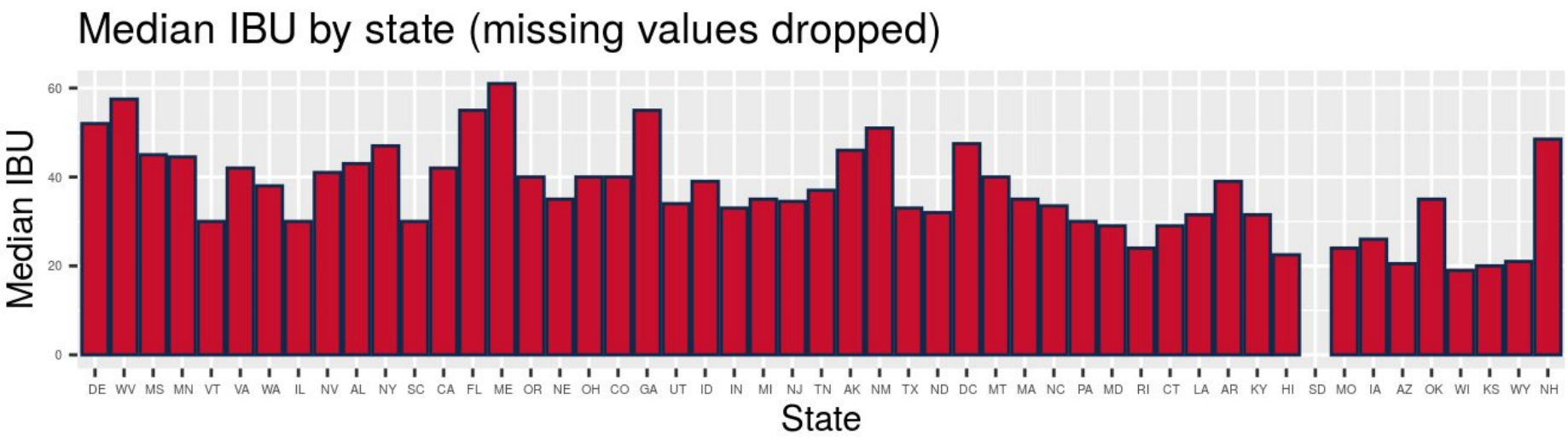


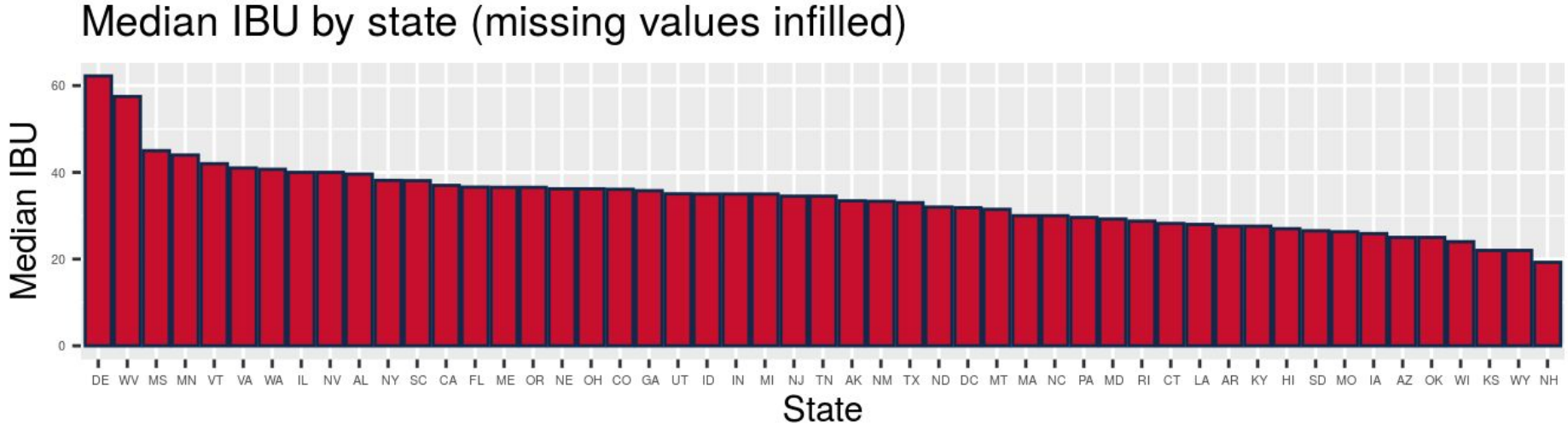
We found that ABV had 62 NAs and IBU had 1005 NAs. All of the ABV NAs also had NA IBU. We decided to fill ABV values with the median ABV from beers of the same type. To fill missing values in IBU, we fit a regression tree to observations without missing values and then used that model to predict each missing IBU value.



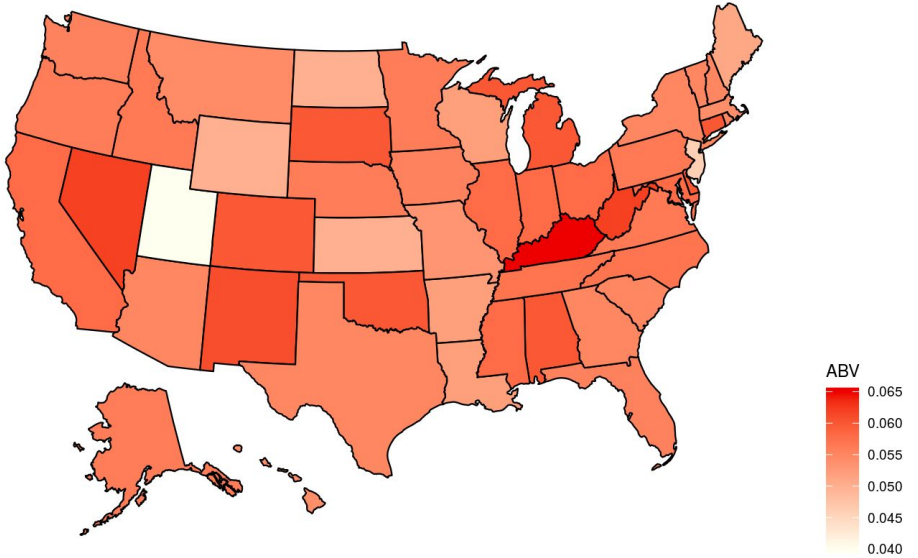

Median ABV by state (missing values dropped)
Median ABV
0.06
0.04
0.02
0.00
KY DC NV WV NM AL CO CT DE MI OK SD CA IL IN MD MS OH NC TN ID PA VA AK MN NE OR FL IA WA AZ GA NH NY RI SC TX VT MT HI MA MO AR LA WI ME KS ND WY NJ UT
State


Median ABV by state (missing values infilled)
Median ABV
0.06
0.04
0.02
0.00
KY DC NV WV NM AL CO CT DE MI OK SD CA IL IN MD MS OH NC TN ID PA VA AK MN NE OR FL IA WA AZ GA NH NY RI SC TX VT MT HI MA MO AR LA WI ME KS ND WY NJ UT
State

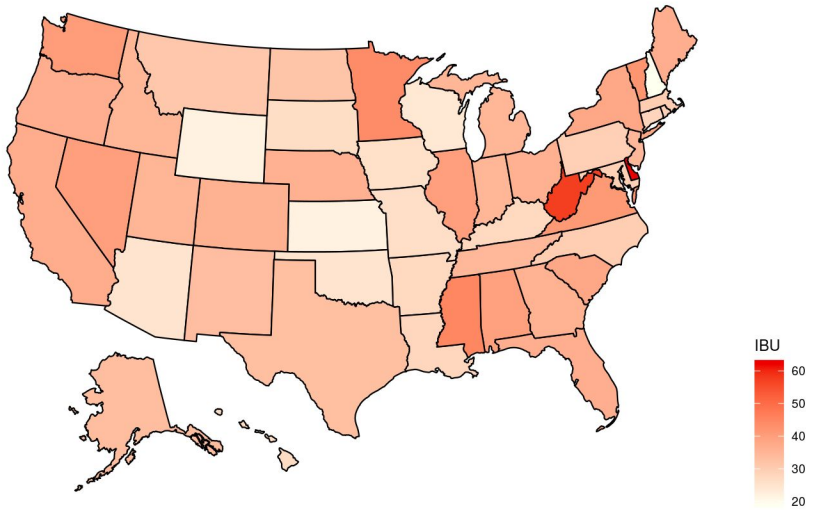

Median IBU by state (missing values dropped)
Median IBU
60
40
20
0
DE WV MS MN VT VA WA IL NV AL NY SC CA FL ME OR NE OH CO GA UT ID IN MI NJ TN AK NM TX ND DC MT MA NC PA MD RI CT LA AR KY HI SD MO IA AZ OK WI KS WY NH
State


Median IBU by state (missing values infilled)
Median IBU
60
40
20
0
DE WV MS MN VT VA WA IL NV AL NY SC CA FL ME OR NE OH CO GA UT ID IN MI NJ TN AK NM TX ND DC MT MA NC PA MD RI CT LA AR KY HI SD MO IA AZ OK WI KS WY NH
State

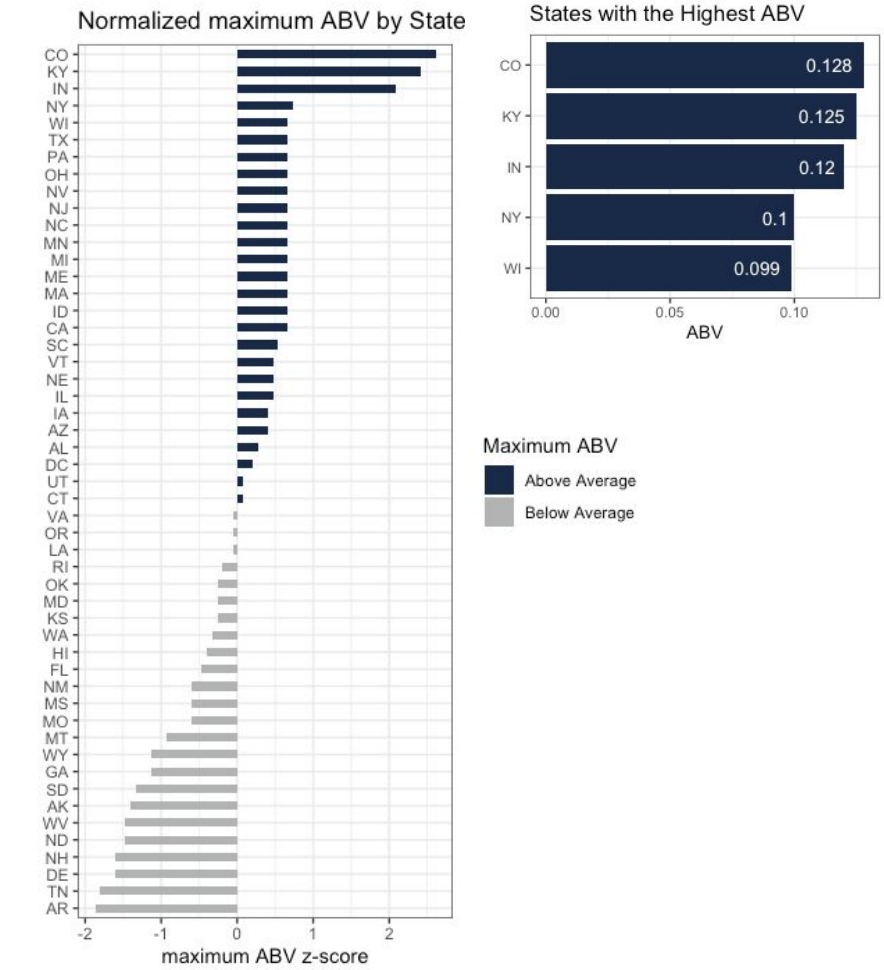
Median ABV of beers by state

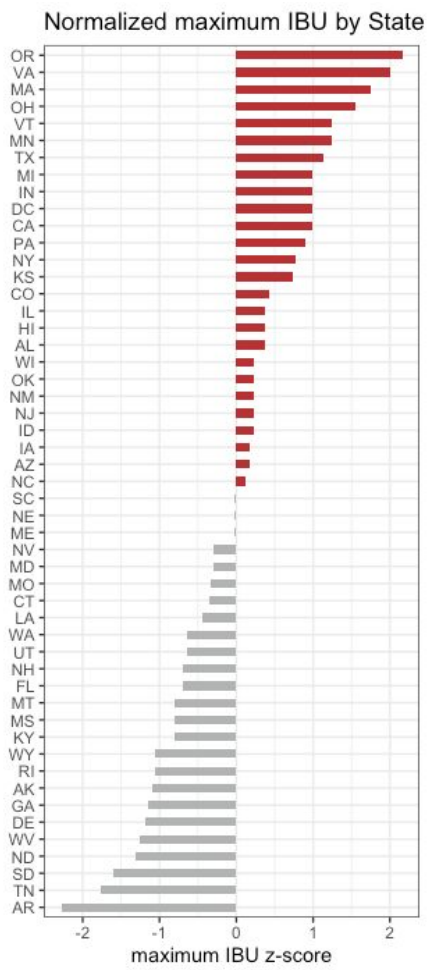


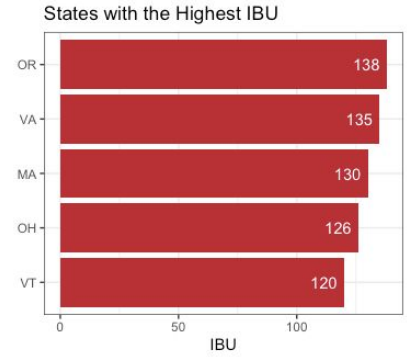
Median IBU of beers by state



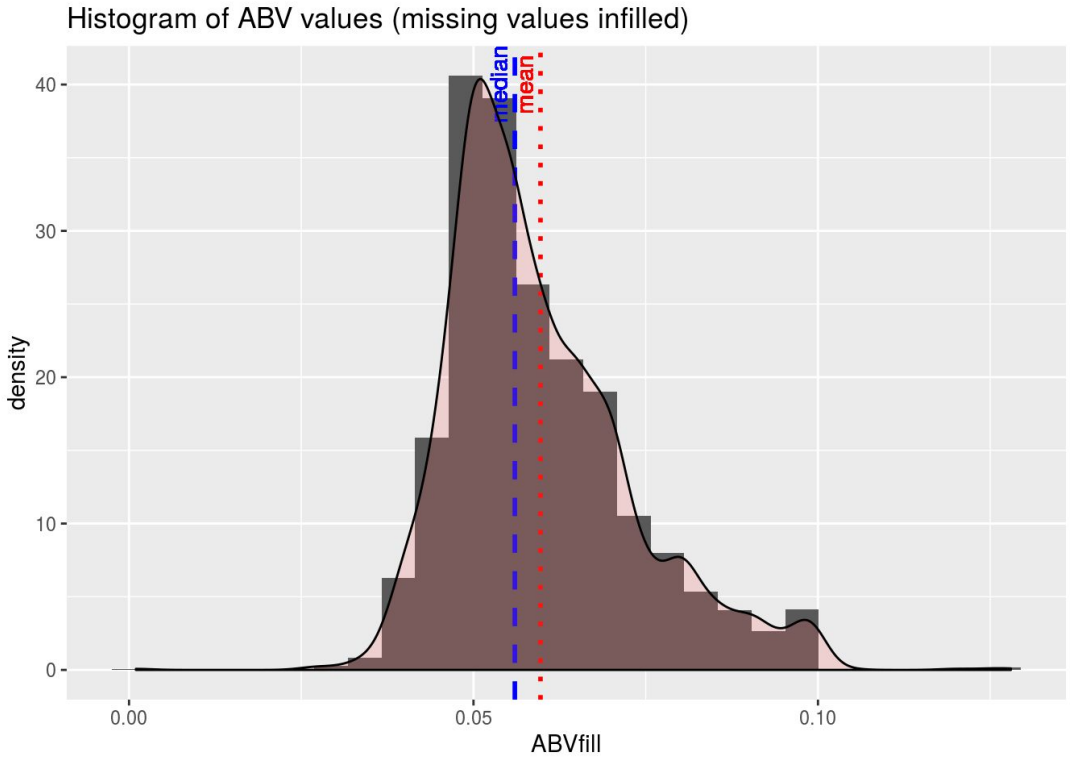
QUESTION 5 - Max ABV and IBU by state

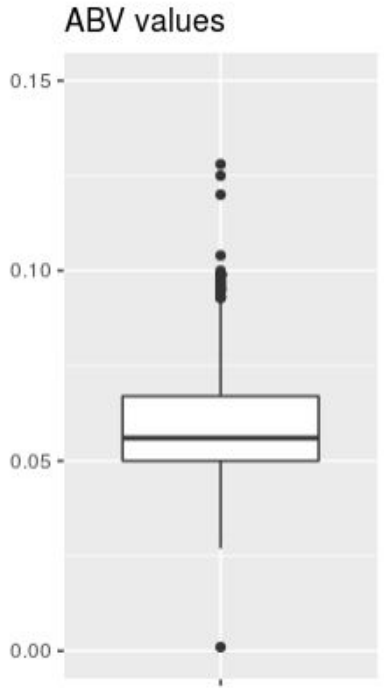






QUESTION 6 - Examine distribution of ABV values.

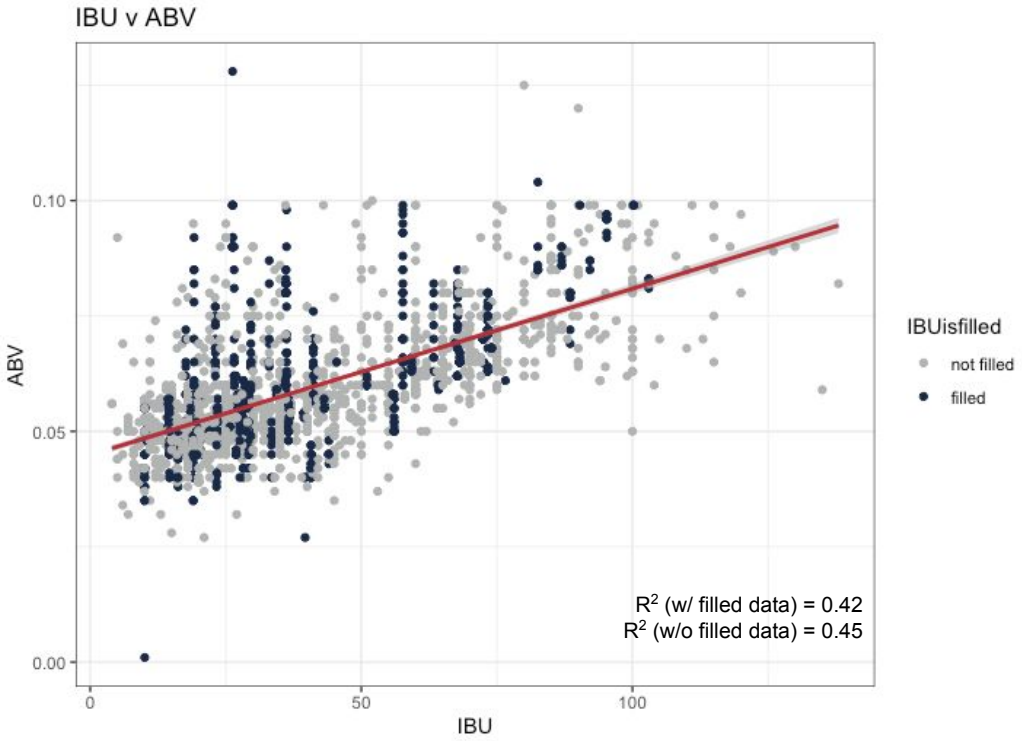




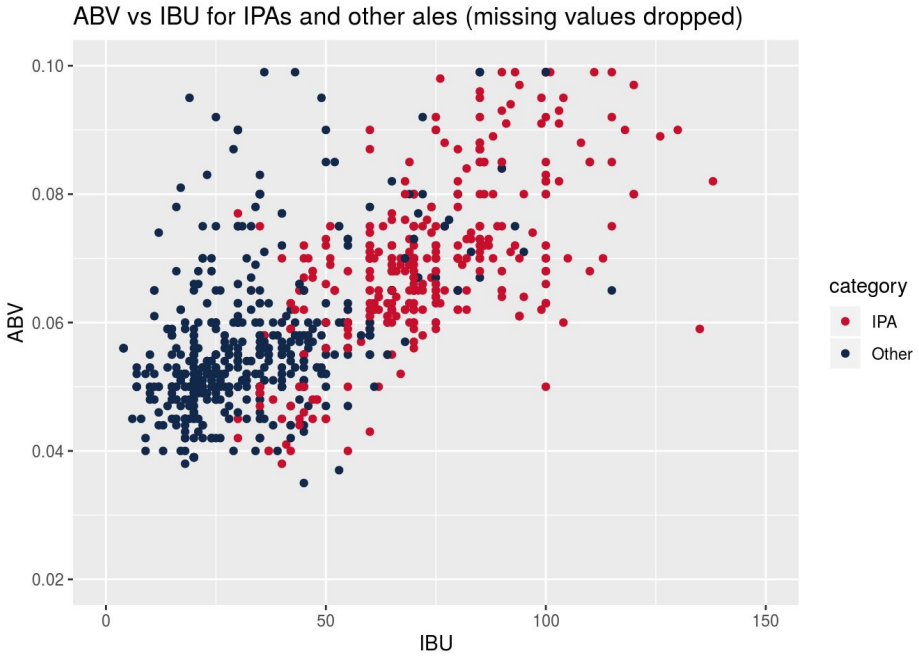
ABV values have a slightly right skewed distribution, indicating a small number of high-value outliers. Otherwise, the values are closely grouped around the mean and median.

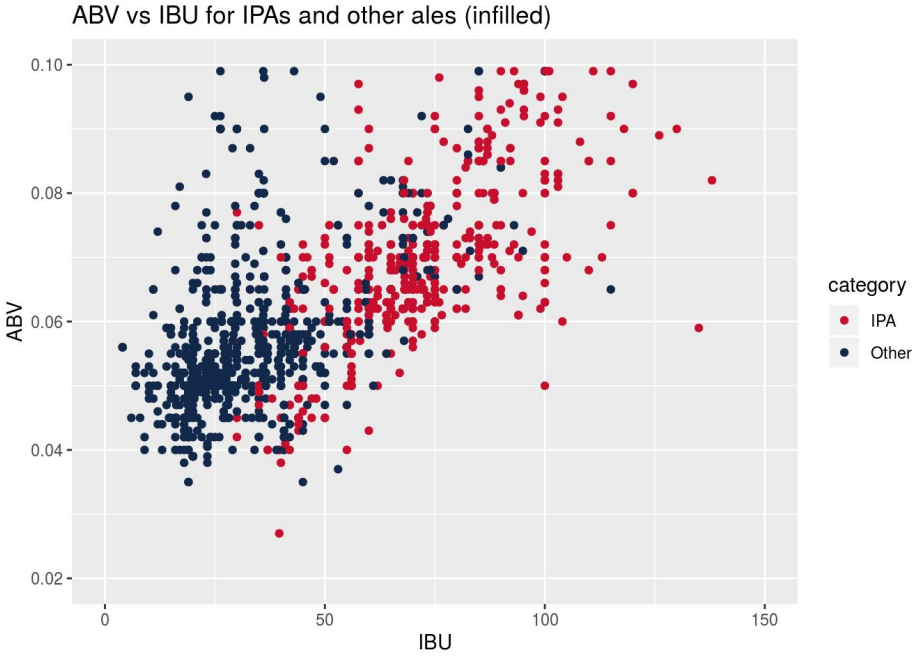
QUESTION 7 - SCATTERPLOT OF IBU VS. ABV

There is a moderate, positive relationship between IBU and ABV.

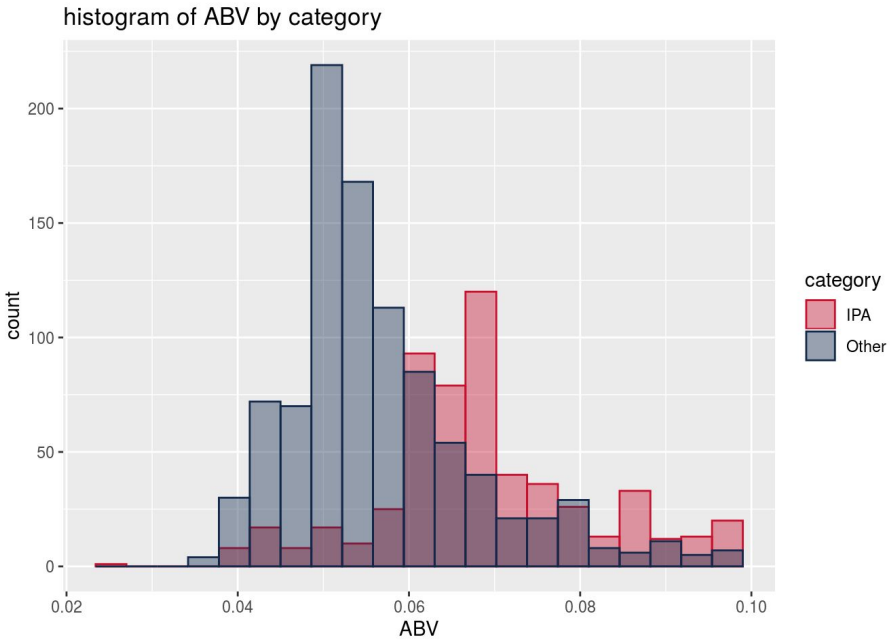


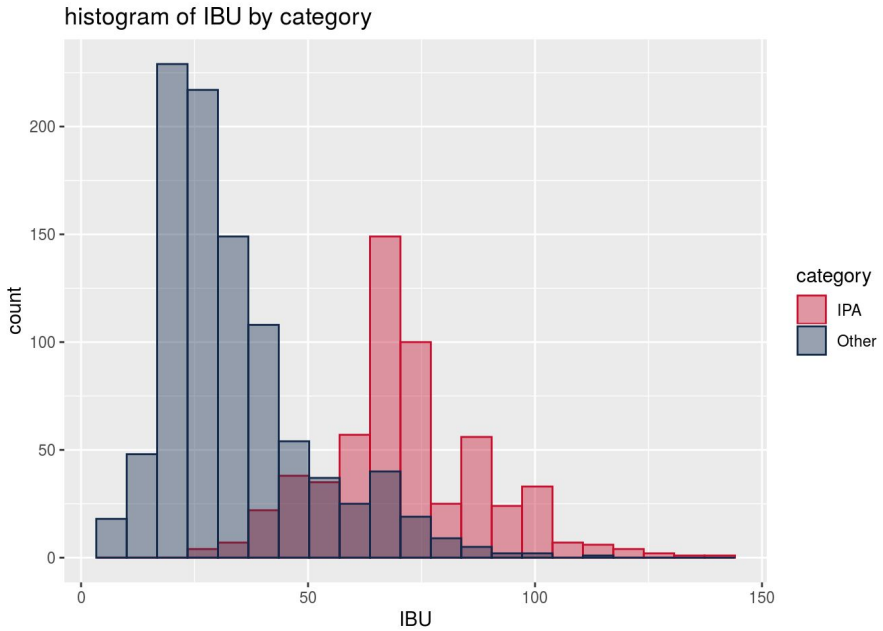
Question 8 - Investigate difference in ABV and IBU between IPAs and other ales.





Infilling missing values for ABV and IBU does not change the character of their relationship, so subsequent analysis will use infilled datasets





There is a statistically significant difference in both ABV and IBU between IPA's and other ales

	ABV (%)	IBU (Intl Bitterness Units)
IPA	5.67	71.2
Other Ales	6.88	33.9
95% Conf Interval for diff	[1.08 .. 1.33]	[35.4 .. 39.1]

Directions for Additional Work

Our study of the difference betweeen IPAs and other ales with respect to ABV and IBU continues.

Ideas for question 9: **a)** Group states into geographic regions and repeat analysis for mean/max ABV and IBU by region, **b)** do some feature engineering (IBU:ABV ratio) or external data (census data, etc.), **c)** deep dive into one state...maybe one with lots of breweries...and provide some state-level insights. I'm currently most inclined toward option "b".