



# Data Analysis of Breweries and Beer

---

**Andre Mauldin, Lijju Mathew, Michael  
Olheiser (Team Floating Point)**

---

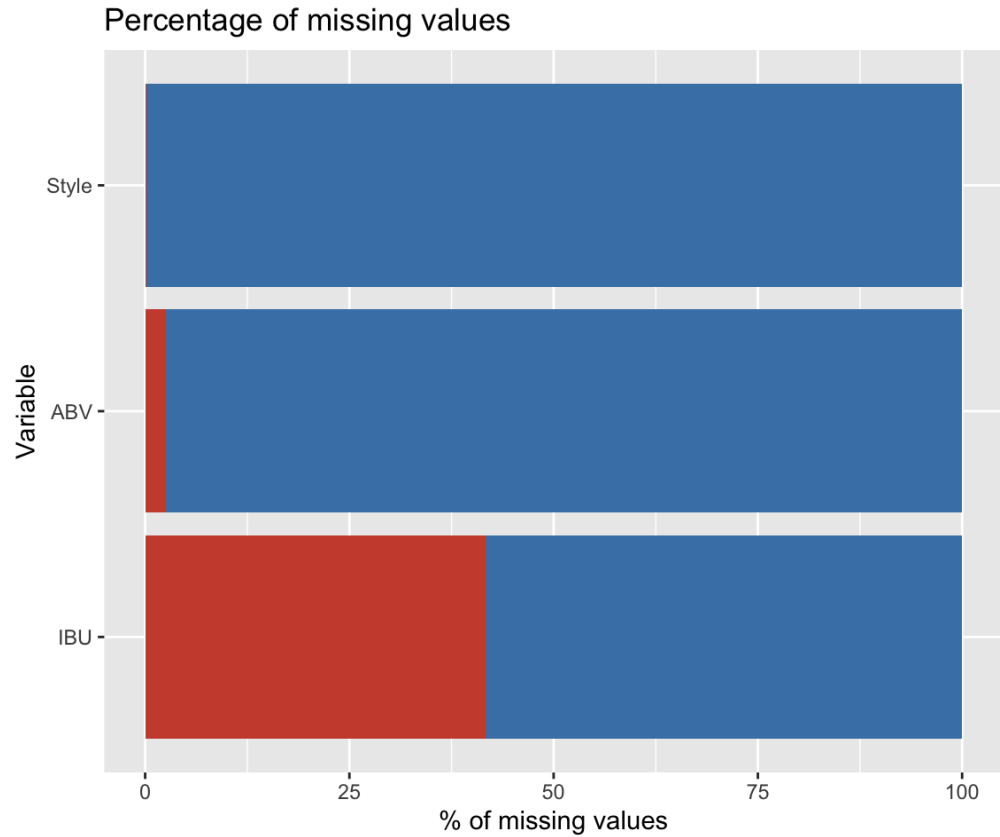
July 1st, 2020



1. Breweries by State
2. Handling Missing Values
3. ABV and IBU by State
4. Highest ABV and IBU Beers
5. Distribution of ABV of IBU
6. Relationship between ABV and IBU
7. KNN Classification of Beer Types
8. Top 5 Most Common Beers
9. Conclusions



## Handling Missing Values

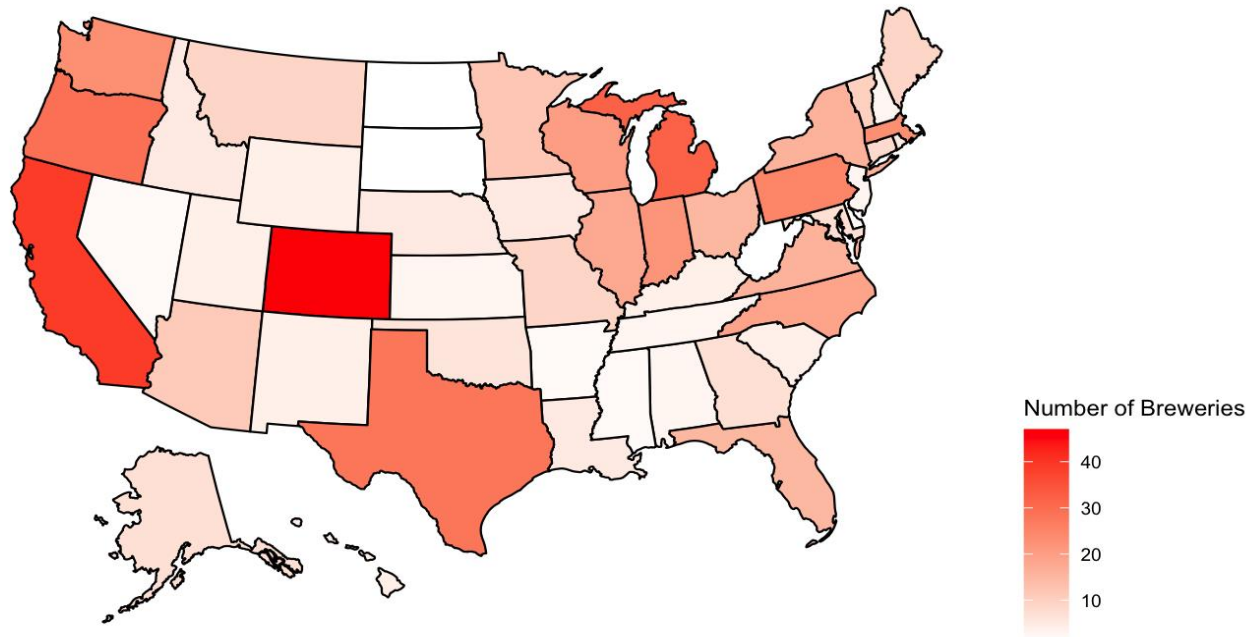


- IBU is not regularly reported per beer
- Replace missing values with average IBU/ABV in that style
- 9 styles of beer do not contain any reported IBUs including 37 Ciders
- Set IBU to zero for Cider. Drop remaining styles with no IBUs.
- Result: Drops 20 beers out of 2410.

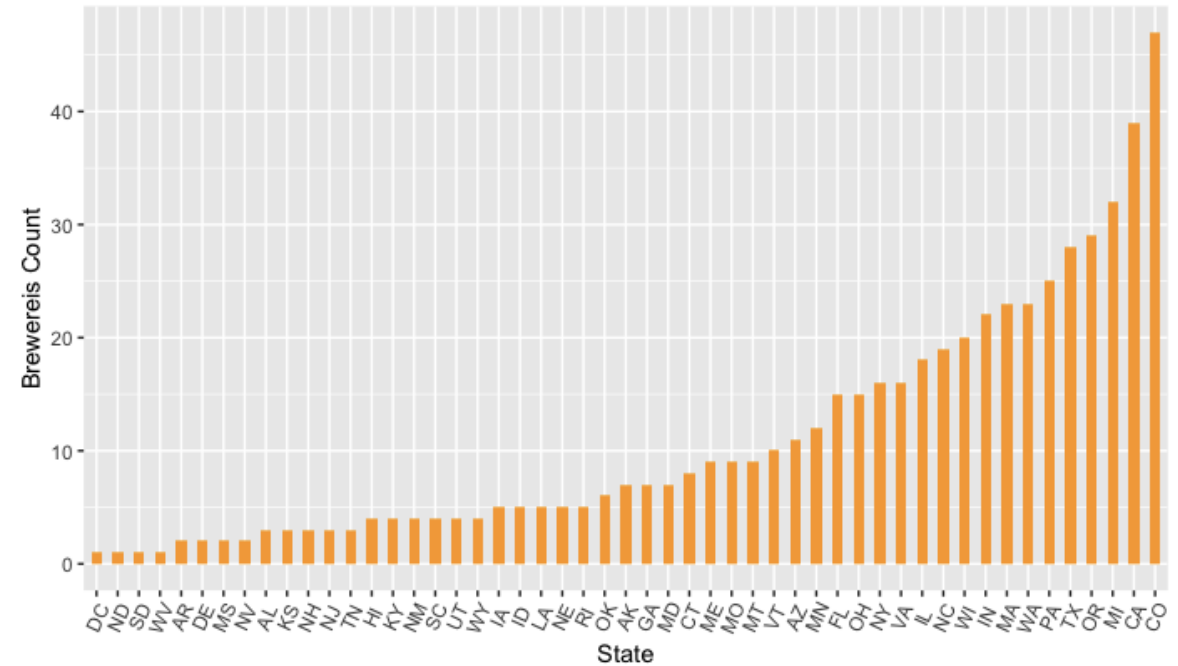


## Distribution of Breweries across the United States

Breweries in the United States



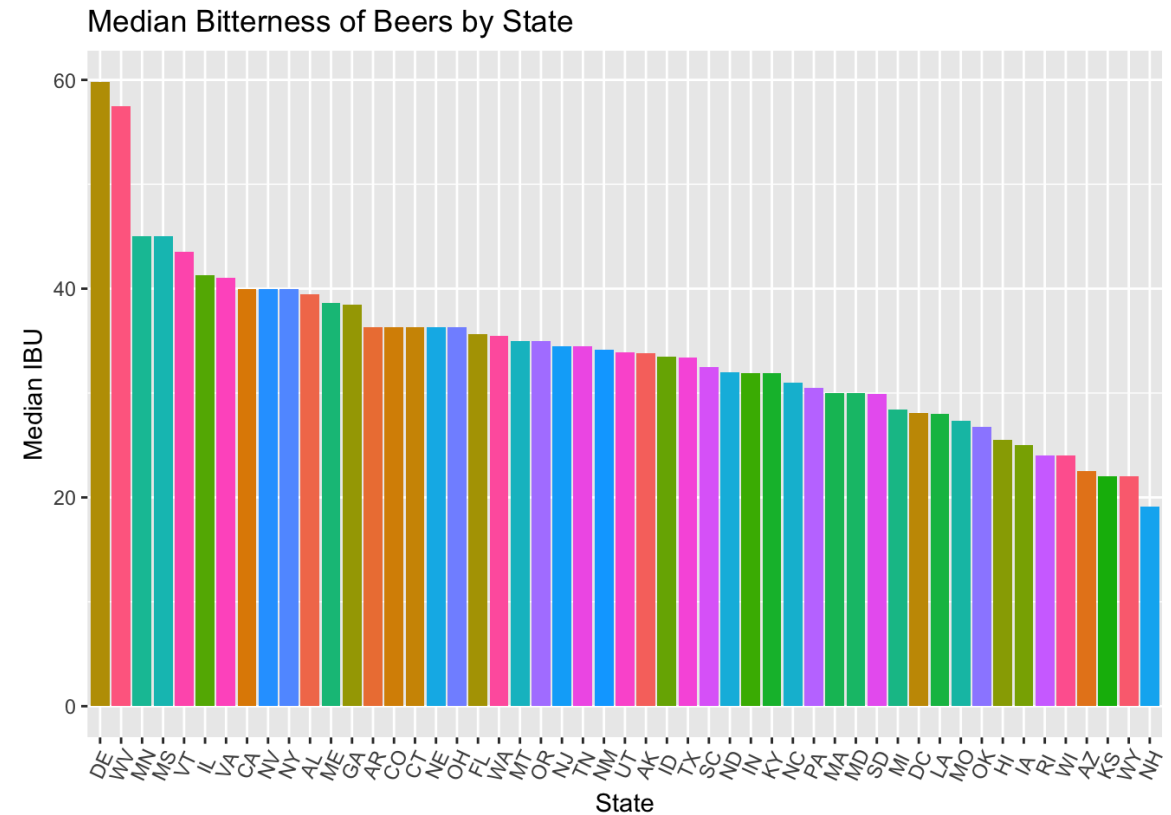
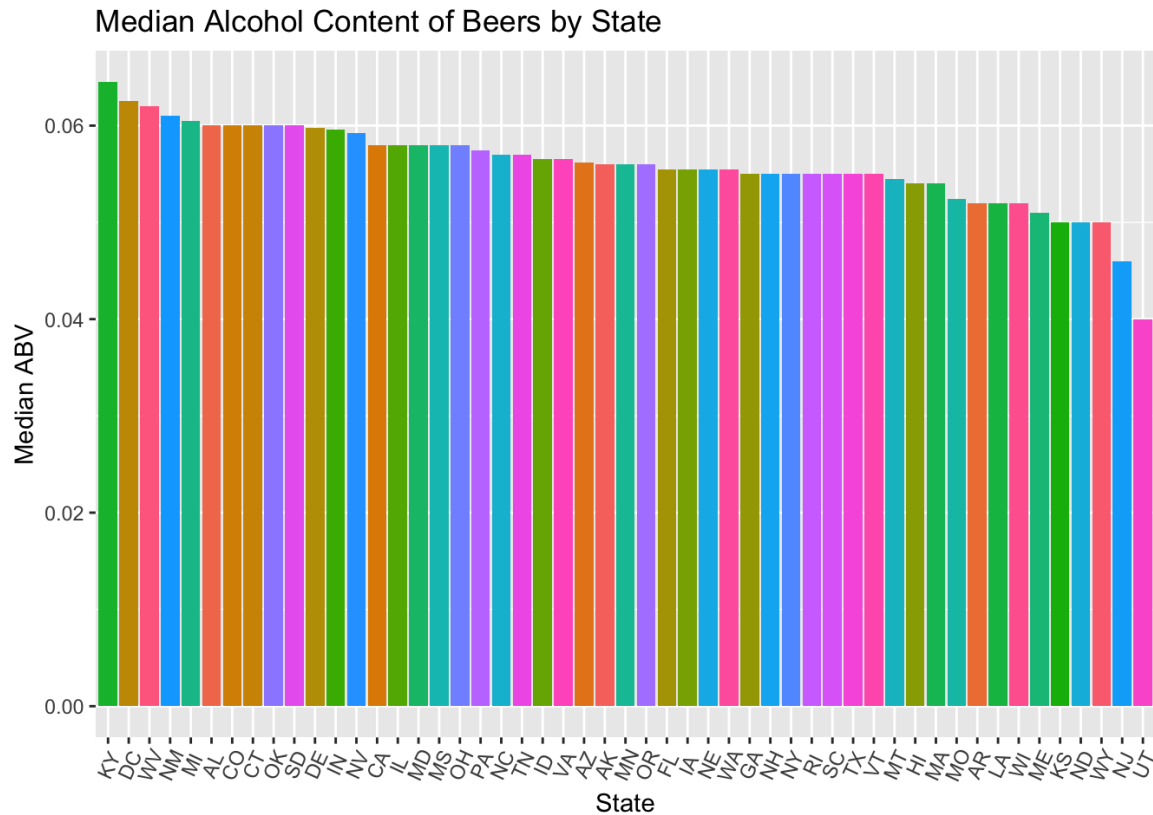
Number of Breweries Per State



- California, Colorado, Michigan contain the most breweries in the US
- No specific region that is dominated by breweries
- Population is not the only factor affecting the presence of breweries



## Median IBU and ABV by State



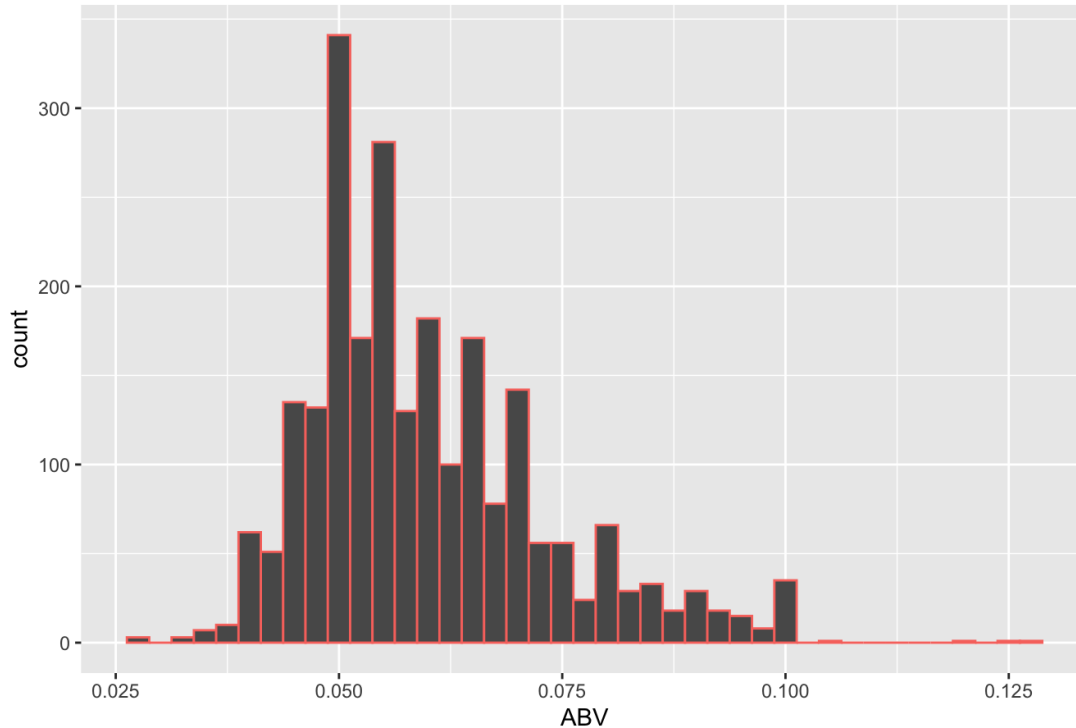
- Median ABV varies less across states than median IBU
- No obvious geographical/regional pattern.
- A higher Median ABV does not necessarily indicate higher ABV ( see next slide)





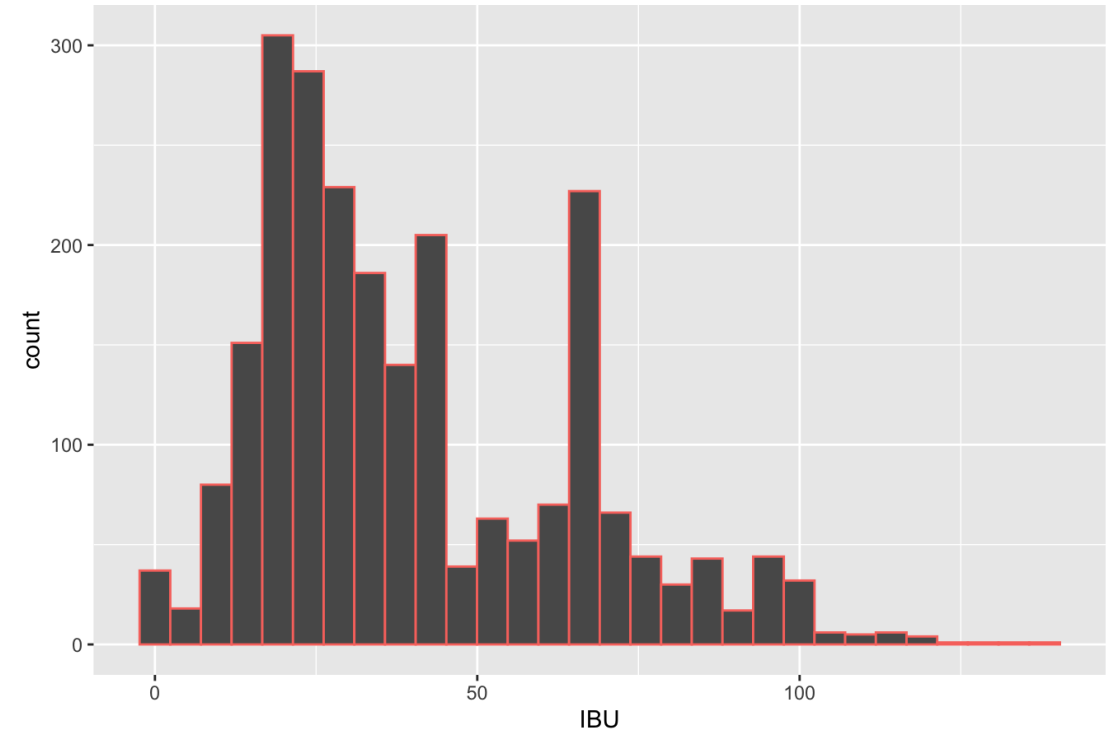
## Distribution of Alcoholic Content of Beers

Distribution of ABV for All beers



- Median = 0.05600
- Mean = 0.05974
- Distribution is slightly right-skewed as beers are not typically strong alcoholic drinks.

Distribution of IBU for All beers



- Median = 33.81
- Mean = 40.27
- Distribution may be bi-modal.



## Top IBU and ABV Beers in the US

---

### Top 5 IBU Beers

beer_name	Style	brewery_name	City	State	IBU	ABV
Bitter Bitch Imperial IPA	American Double / Imperial IPA	Astoria Brewing Company	Astoria	OR	138	0.082
Troopers Alley IPA	American IPA	Wolf Hills Brewing Company	Abingdon	VA	135	0.059
Dead-Eye DIPA	American Double / Imperial IPA	Cape Ann Brewing Company	Gloucester	MA	130	0.090
Bay of Bengal Double IPA (2014)	American Double / Imperial IPA	Christian Moerlein Brewing Company	Cincinnati	OH	126	0.089
Abrasive Ale	American Double / Imperial IPA	Surly Brewing Company	Brooklyn Center	MN	120	0.097

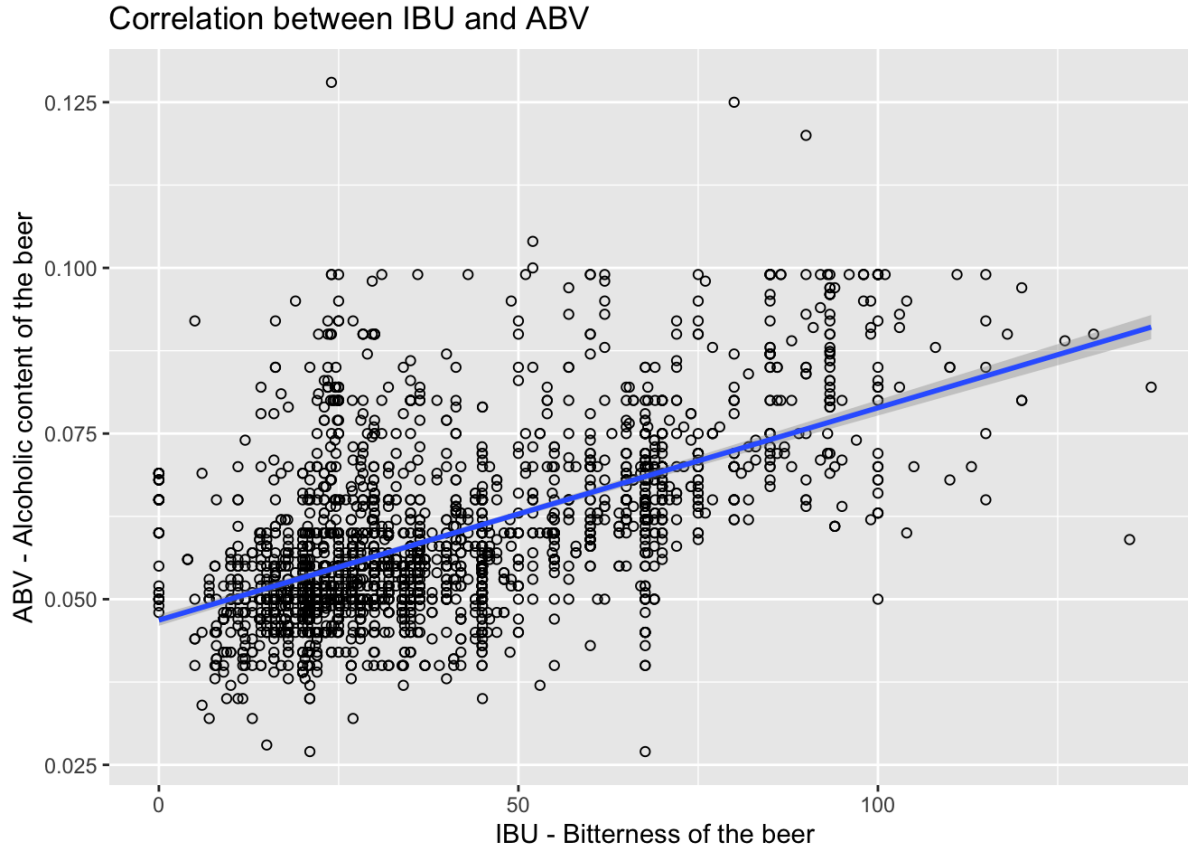
- Boulder, CO has 2 of the top 5 ABV beers
- Top 5 IBU beers are all American IPAs

### Top 5 ABV Beers

beer_name	Style	brewery_name	City	State	IBU	ABV
Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale	Quadrupel (Quad)	Upslope Brewing Company	Boulder	CO	24	0.128
London Balling	English Barleywine	Against the Grain Brewery	Louisville	KY	80	0.125
Csar	Russian Imperial Stout	Tin Man Brewing Company	Evansville	IN	90	0.120
Lee Hill Series Vol. 4 - Manhattan Style Rye Ale	Rye Beer	Upslope Brewing Company	Boulder	CO	52	0.104
4Beans	Baltic Porter	Sixpoint Craft Ales	Brooklyn	NY	52	0.100



## Relationship between IBU and ABV



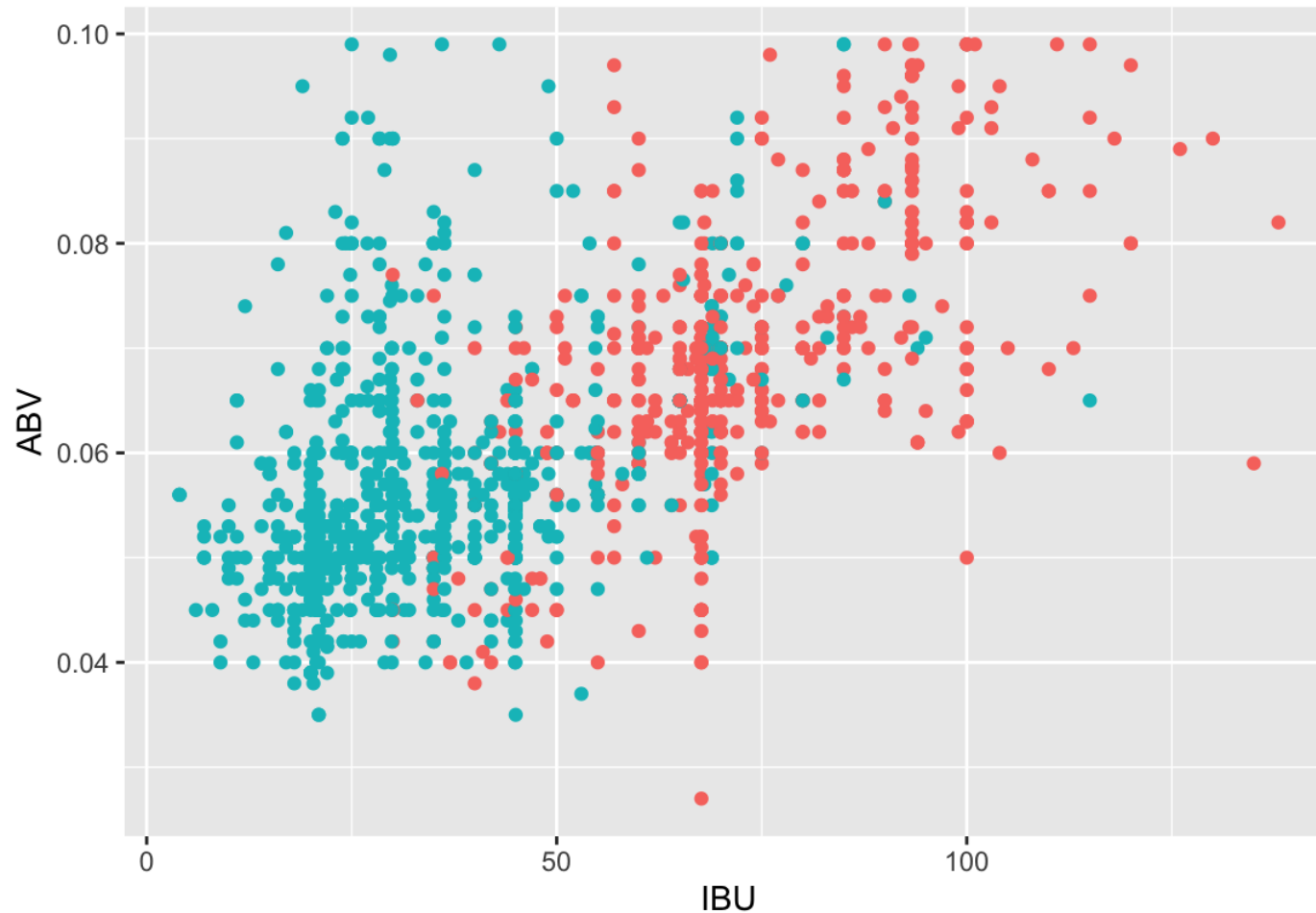
- Positive correlation
- Correlation value = 0.58
  - 95% CI (0.55 , 0.61)
- Data is sparse beyond ~60 IBU.
- There is not strong evidence against equal variances indicating a linear regression may be an appropriate model here.





## Classification of Beer Styles by IBU and ABV

IBU vs ABV by Beer Type



Confusion Matrix ( Columns are predictions)

Beer Type	IPA	Other Ales
IPA	156	20
Other Ales	30	254

beer\_type

- IPA
- Other Ales

Accuracy

0.891

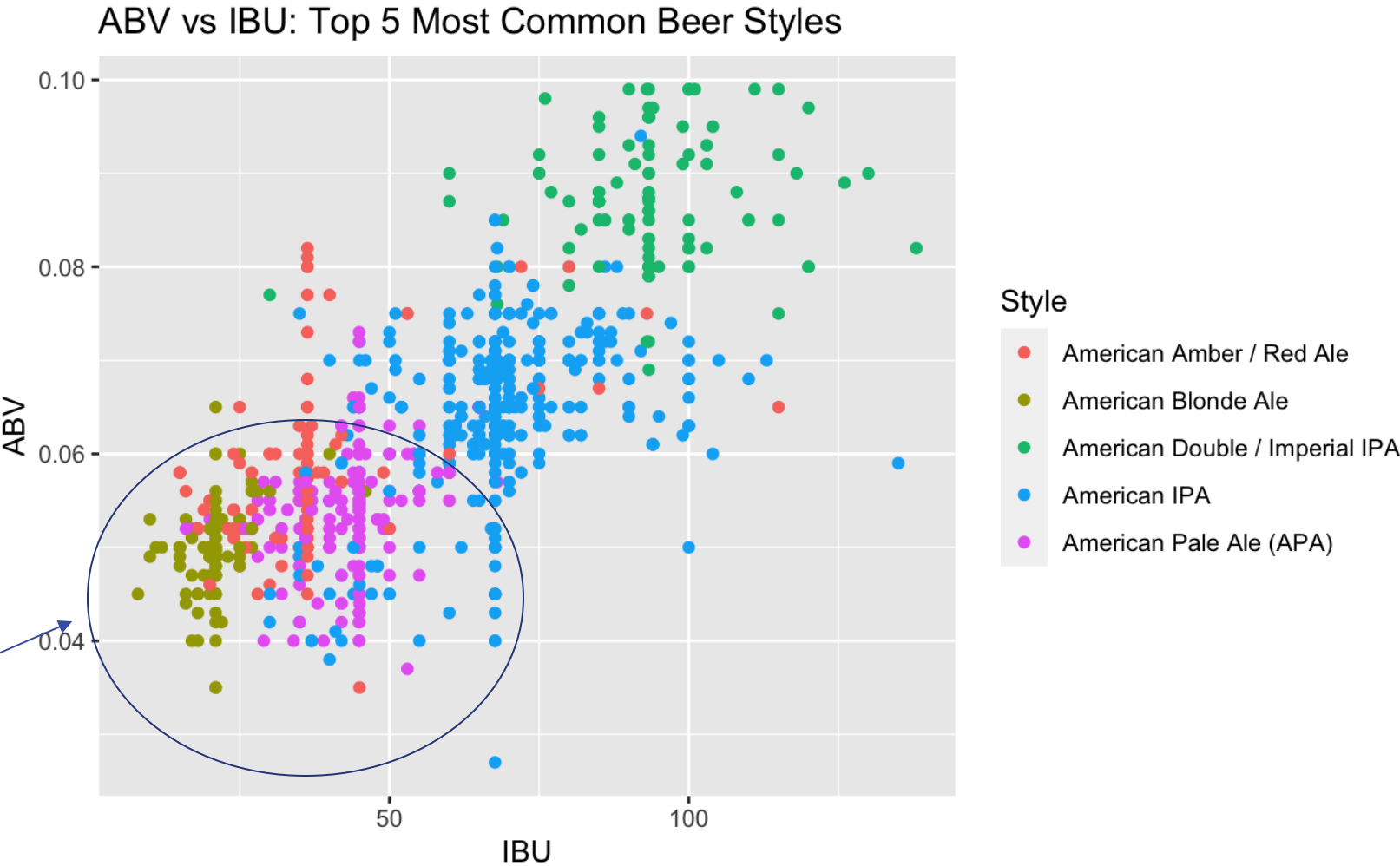
- IBU are clearly higher for IPA than other Ales and span a larger range.
- IPAs tend to be higher ABV than Other Ales



Classification of Top 5 Most Common Beer Styles

Top 5 Most Common Beer Styles  
(Columns are Predictions)

Style	Count
American IPA	424
American Pale Ale (APA)	245
American Amber / Red Ale	133
American Blonde Ale	108
American Double / Imperial IPA	105



Considerable overlap  
among Styles



## KNN Classification Results for Top 5 Most Common Beer Styles

### Confusion Matrix (Columns are Predictions)

	American Amber / Red Ale	American Blonde Ale	American Double / Imperial IPA	American IPA	American Pale Ale (APA)
American Amber / Red Ale	14	10	0	6	13
American Blonde Ale	6	20	0	0	1
American Double / Imperial IPA	0	1	23	11	1
American IPA	2	0	11	102	8
American Pale Ale (APA)	7	3	0	9	57

- American Amber/Red and Blonde Ales are difficult to classify due to the overlap with other styles.
- Pale Ale, IPA, and Imperial IPAs are more distinctly separated by IBU/ABV hence
- In general, styles are more distinct as IBU and ABV is increased.

### Model Results by Style (Overall Accuracy = 0.71)

Beer Style	Balanced Accuracy per Style
American Amber / Red Ale	0.6888431
American Blonde Ale	0.7812025
American Double / Imperial IPA	0.8142501
American IPA	0.8391155
American Pale Ale (APA)	0.8140278



# Conclusion

---

- Key takeaways

- No obvious regional patterns in the US of brewery locations
- IBU not regularly reported
- Positive linear relationship between IBU and ABV
- Bi-modal distribution of IBU – possibly different markets?
- Not all styles of beer can be accurately classified by IBU/ABV

- Further research

- Distribution/Sales data – where is the beer going?
- Demographics, economics of states

