

MSDS 6306: Case Study 1

Et tu, Brewte?

Rick Fontenot, Joe Lazarus, and Jason Herbaugh





Executive Summary

Budweiser is exploring the craft brewery market. We will review current trends for beer in this industry and provide insight for growth opportunities in the market



Agenda

Existing craft brewery locations

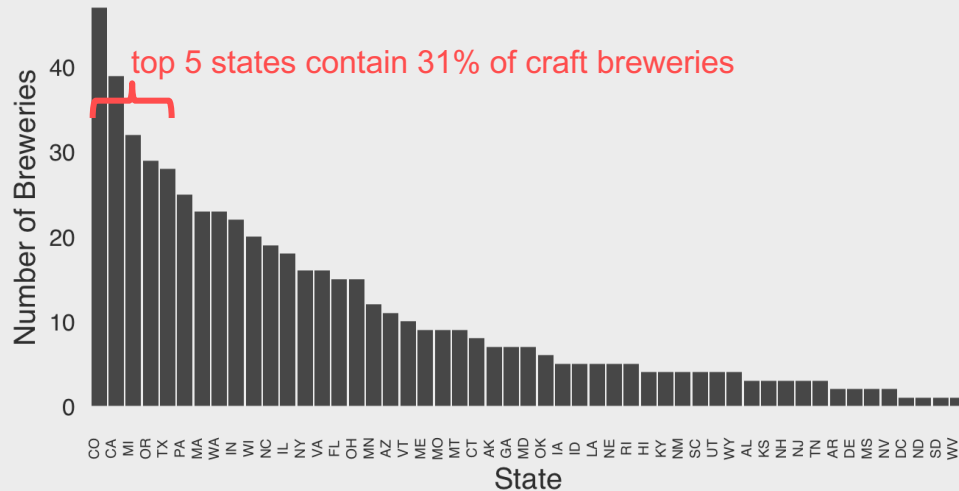
Types of craft beers

Summary of Alcohol by Volume (ABV) and Bitterness (IBU)
profiles and characteristic

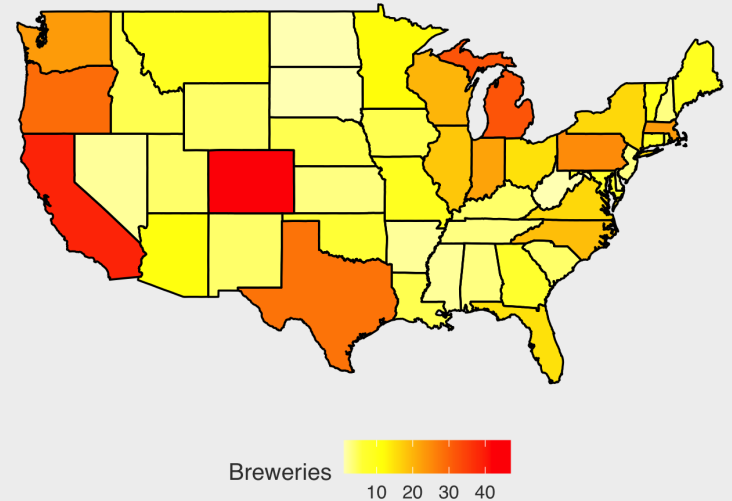
US Craft Brewery Locations

- ▶ Colorado has the most breweries with 47
- ▶ Other than Colorado, states with most breweries are clustered along US border and coasts

Number of Breweries by State

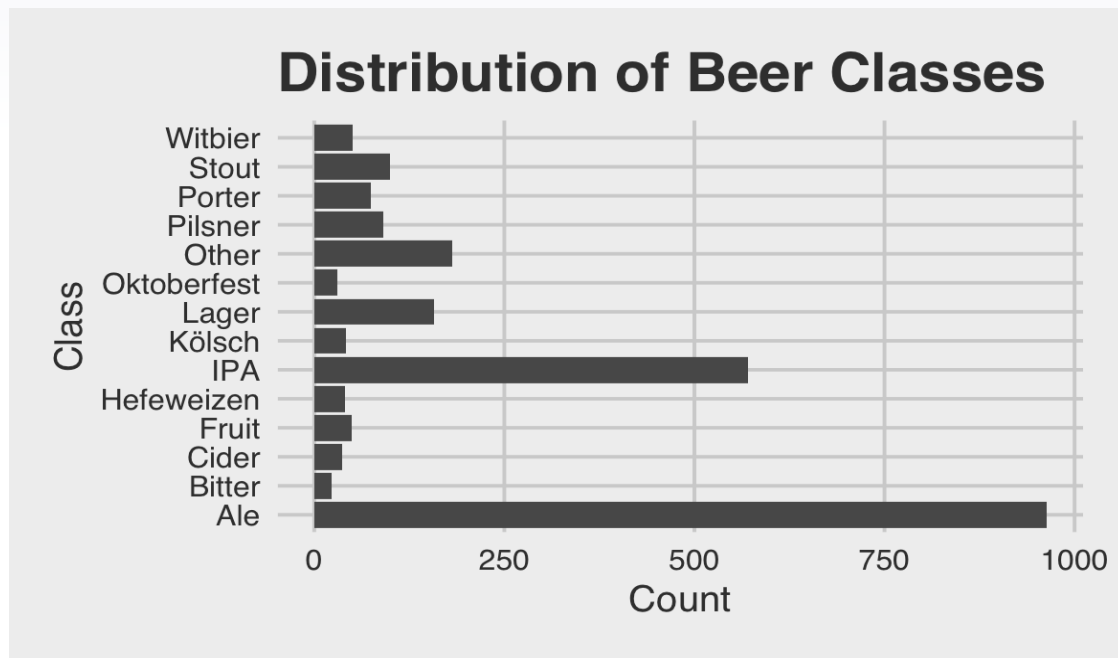


Number of Breweries by State



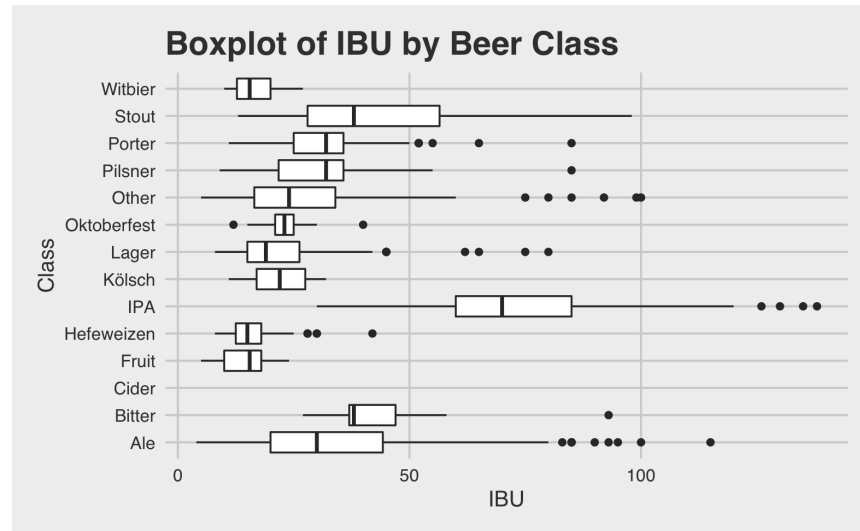
Types of Craft Beers

- ▶ There are 100 different style of craft beers brewed in the US
- ▶ Classes of Beer have style variants i.e., Ales include English Pale, Irish Red, etc.
- ▶ When grouped into classes, IPA's and Ales account for 64% of all craft beers



Addressing Missing Values

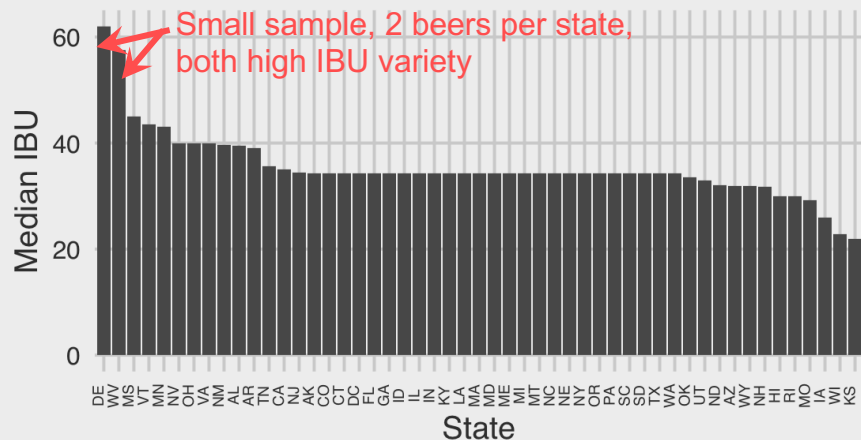
- ▶ 42% of the beers are missing IBU information, 2.6% are missing ABV information
- ▶ Should these data points be excluded or estimated from other information available?
- ▶ IBU is very different for each style of beer. If overall mean IBU=50, that would not a good estimate for Witbiers or IPA's
- ▶ We imputed missing values with the mean for their respective Class of Beer



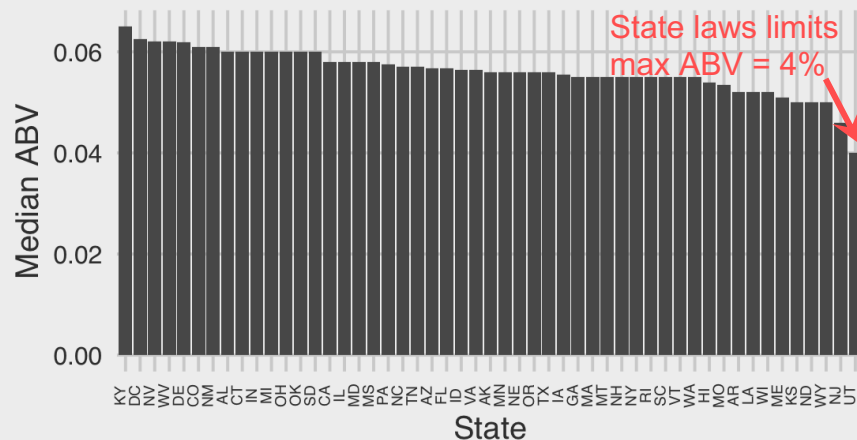
IBU and ABV by State

- ▶ For states with sufficiently larger sample sizes median IBU in the mid-30's
- ▶ Almost all states have a median ABV in the 5% to 6% Range

Median IBU by State



Median ABV by State



* Medians calculated using beer class-mean method of imputing missing values

Maximum IBU and ABV Beers

ABV = 12.8%

Upslope Brewing Company, Colorado

Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale

Brewed with a select strain of Belgian yeast and traditional Belgian candy syrup, the beer matured over six months to mellow and soften its character.



IBU = 138

Astoria Brewing Company, Oregon

Bitter Bitch Imperial IPA

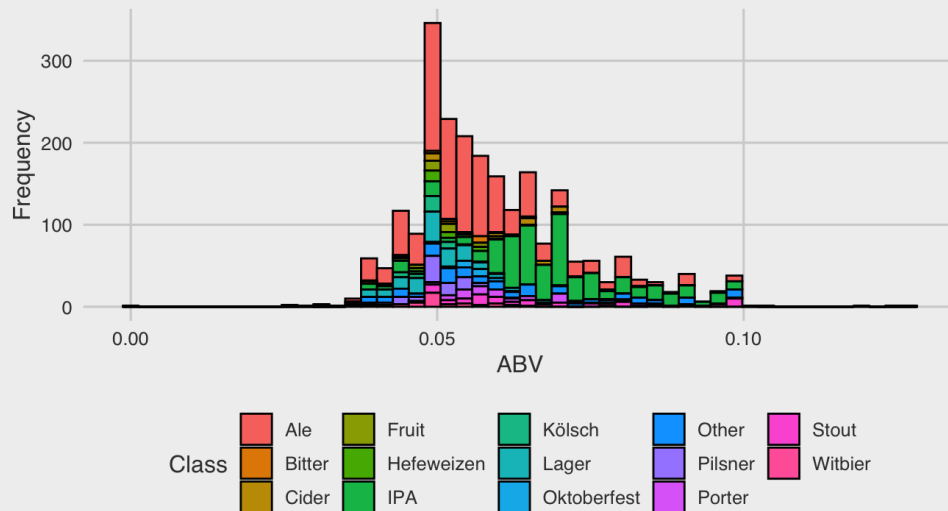
A big IPA with a huge bite. Peoples Award Winner 3 Years running at the Spring Beer & Wine Festival, Portland, OR.



Summary of ABV Distribution

- ▶ ABV distribution is relatively normal with a right skew
- ▶ IPA's contribute most to the right tail of high ABV values, explains mean > median
- ▶ The minimum value is a non-alcoholic beer, excluding this one reduces range to 2.7% - 12.8%

Distribution of Alcohol By Volume



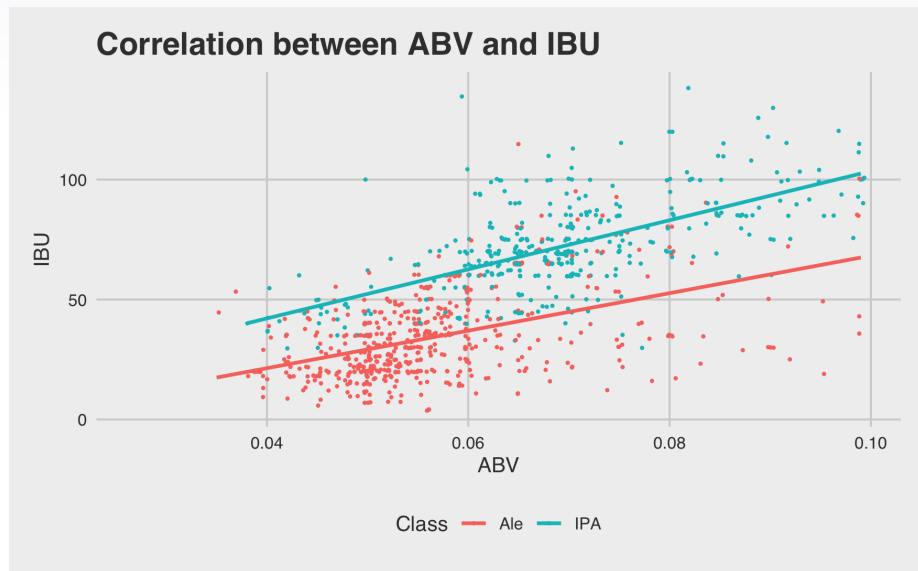
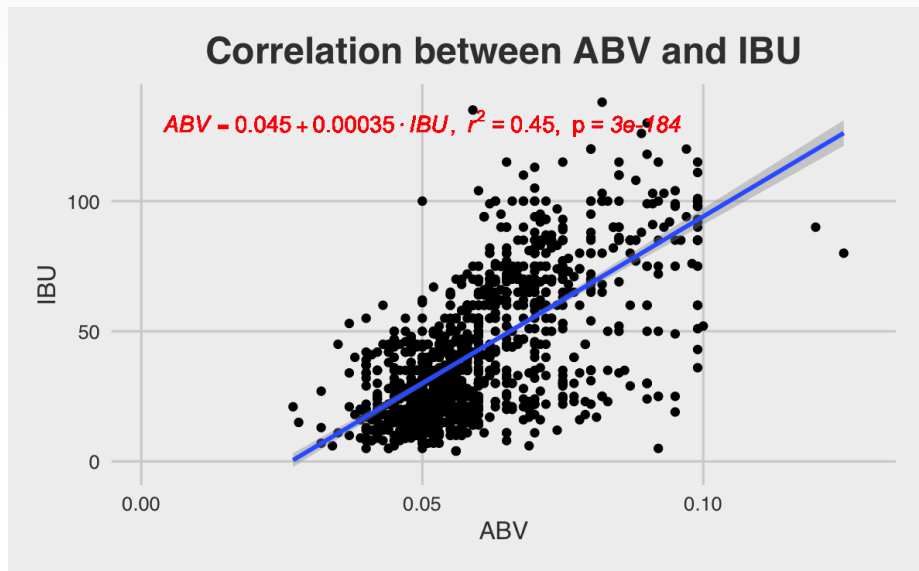
Overall
(N=2410)

ABV

Mean (SD)	0.0598 (0.0135)
Median [Min, Max]	0.0560 [0.00100, 0.128]
Missing	62 (2.6%)

IBU and ABV Relationship

- ▶ ABV and IBU are positively correlated with 45% of the variation in ABV explained by IBU
- ▶ The IBU/ABV relationship is different or shifted for different styles of beer
- ▶ For similar alcohol level, IPA is more bitter than Ale



KNN model for Ale vs. IPA

- ▶ ABV and IBU relationships are distinctly different between IPA's and Ales
- ▶ Given these two characteristics and this relationship we can predict whether a beer is of the IPA or Ale style with high accuracy, 95%
- ▶ A k-nearest-neighbor model (with NA's omitted) and tuned with optimal k=1

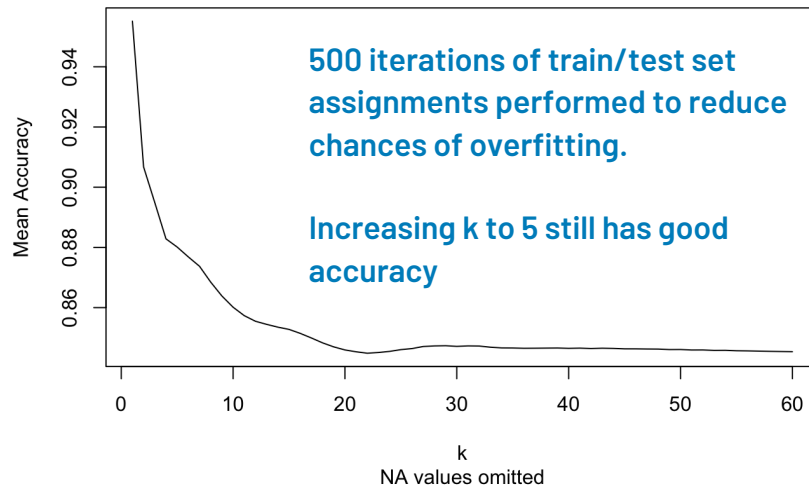
CONFUSION MATRIX

		Actual	
		Ale	IPA
Predicted	Ale	367	7
	IPA	20	253

DETAILS

Sensitivity 0.948	Specificity 0.973	Precision 0.981	Recall 0.948	F1 0.965
Accuracy 0.958		Kappa 0.914		

Optimal k for max Accuracy



Further Studies and recommendations

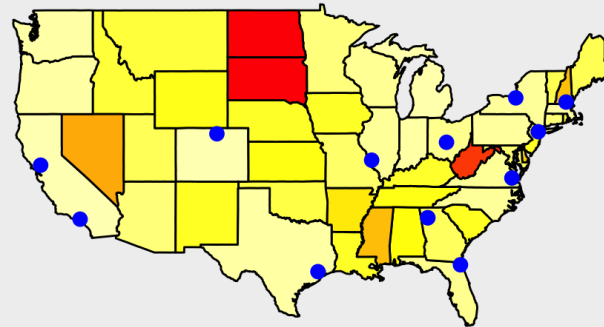
- ▶ Some states with the highest alcohol consumption rate have very few breweries
- ▶ Opportunity to launch new beer with a brewery in these areas to create regional pride and increase demand for your new beer in a ripe market



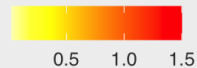
BREWERIES AND TOURS

- BALDWINVILLE, NY
- CARTERSVILLE, GA
- COLUMBUS, OH
- LOS ANGELES, CA
- NEWARK, NJ
- WILLIAMSBURG, VA
- FAIRFIELD, CA
- FORT COLLINS, CO
- HOUSTON, TX
- JACKSONVILLE, FL
- MERRIMACK, NH
- ST. LOUIS, MO

Consumption Per Capita / Brewery Count



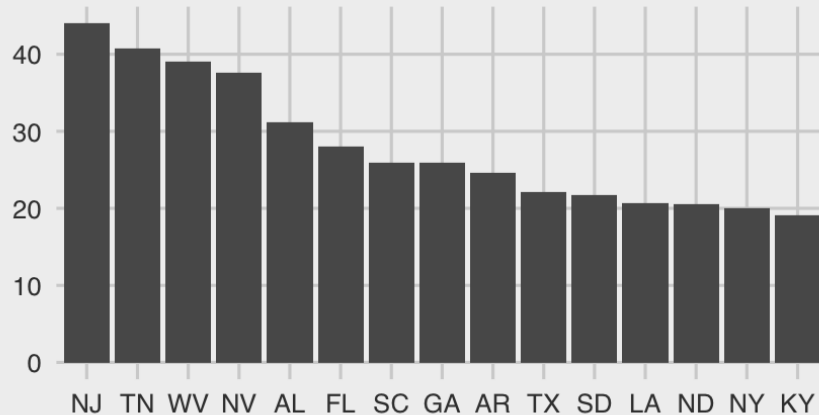
Breweries_v_Consumption



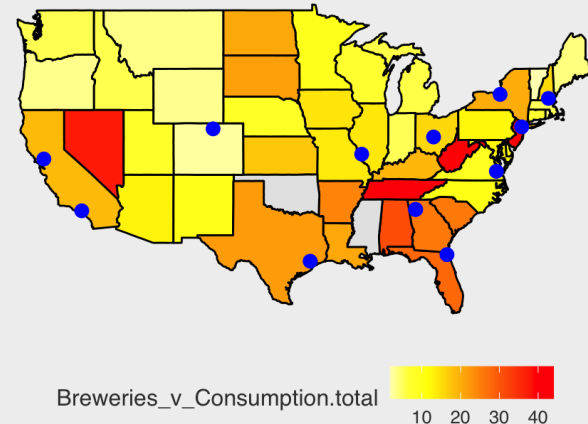
Further Studies and recommendations

- ▶ Consumption per capita may represent high interest, but total consumption (larger population states) may better represent potential sales volume
- ▶ Nevada is high in both categories, North/South Dakota represent large geographic region

Total Consumption / Breweries

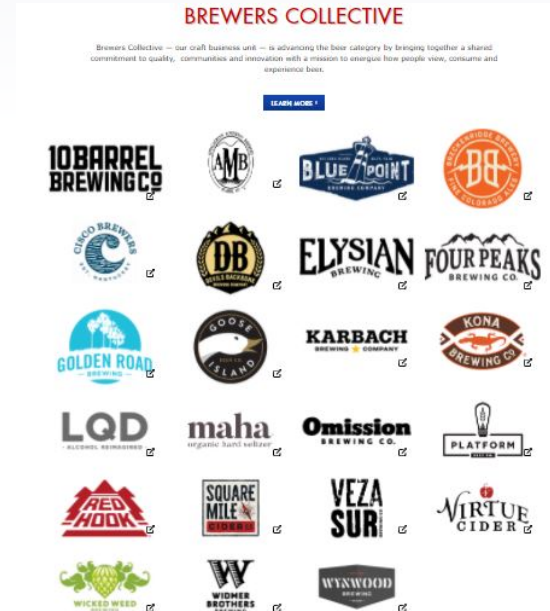


Consumption Total / Brewery Count



Conclusion

- ▶ Preliminary Analysis shows some states warrant further investigation in order to expand Budweiser's share of craft brew market.
- ▶ These states demonstrate potential to support Budweiser Brewery Experiential Site
- ▶ Budweiser could increase market share by tapping into underdeveloped craft brew market through her subsidiary breweries



Assumptions

- ▶ Per Capita Alcohol consumption can approximate taste preference for Craft Beer
- ▶ Data is complete with regards to Breweries
- ▶ Budweiser is looking for market expansion through craft styles of beer such as IPA's

Recommendations

- ▶ Add beer sales data provided by Budweiser
- ▶ Et Tu Brewte refines market analysis with more precise data
- ▶ Include consumer taste preference for Beer with respect to IBU / ABV



Appendix

Source Data

Decisions on addressing missing data

Source Data

Provided by Budweiser

Beers (2410 data points):

Name of the beer

Unique identifier of the beer.

ABV: Alcohol by volume of the beer.

IBU: International Bitterness Units of the beer.

Brewery id associated with the beer.

Style: Style of the beer.

Ounces: Ounces of beer.

Breweries (558 data points):

Unique identifier of the brewery.

Name of the brewery.

City where the brewery is located.

U.S. State where the brewery is located.

Additional External data

US Beer Consumption:

From the National Institute on Alcohol Abuse and Alcoholism

<https://vinepair.com/articles/map-states-drink-beer-america-2020/> Provide Link

Population:

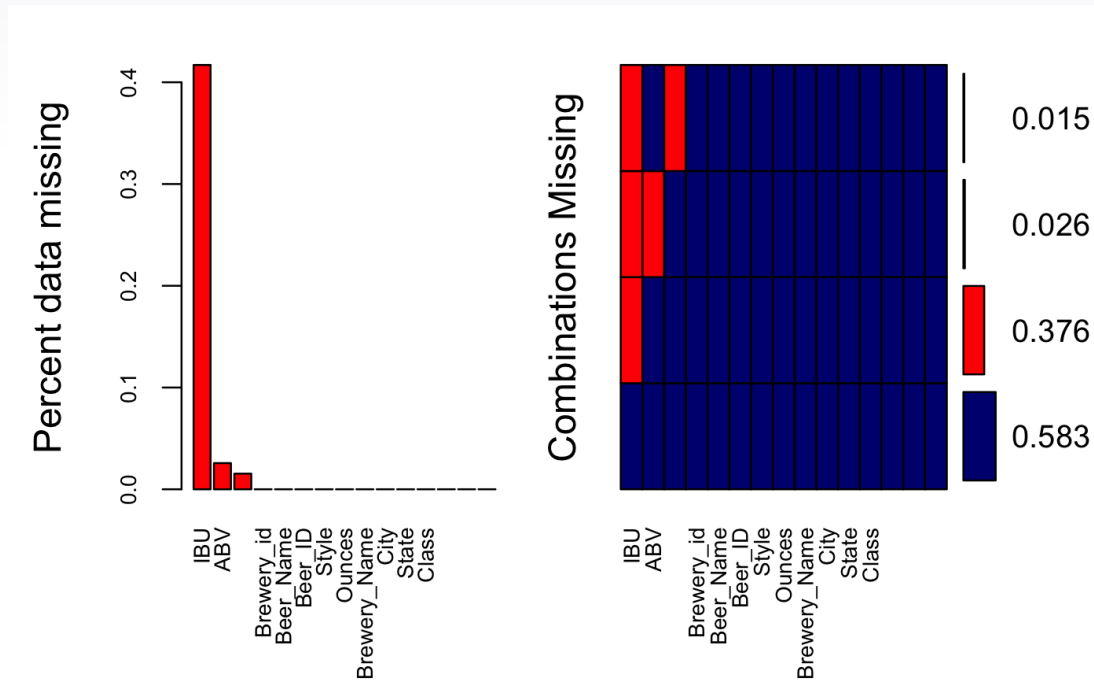
From the United States Census Bureau

<https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-total.html> Provide Link



Addressing Missing Values

- ▶ 42% of the beers are missing IBU information, 2.6% are missing ABV information
- ▶ Should these data points be excluded from analysis or estimated from other information available?

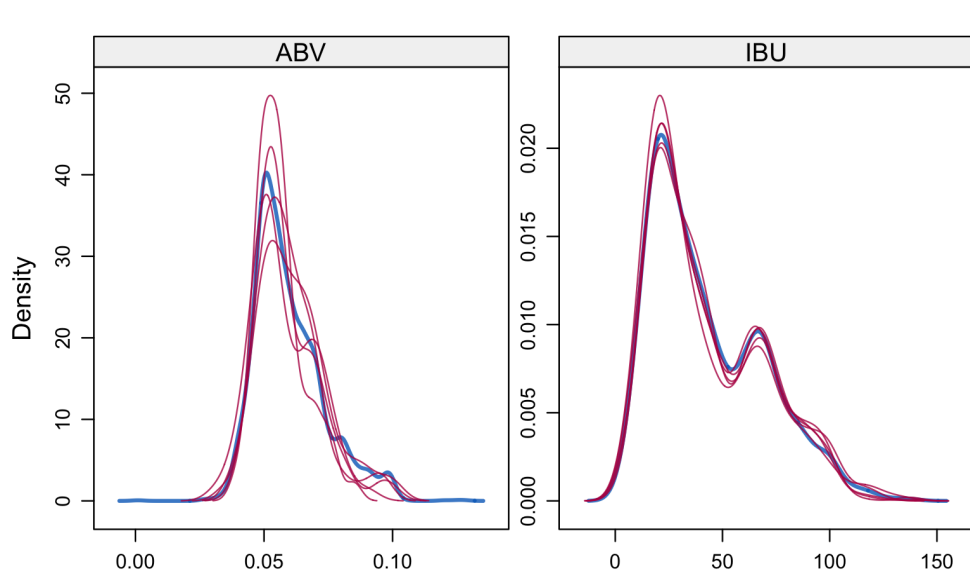


Method 1: Impute missing values based on Predictive Mean Matching (PMM)

- ▶ This method replaces missing values such that the new completed data set mimics the overall distribution of existing IBU and ABV data available
- ▶ Does this make sense to use? Is there a better way?

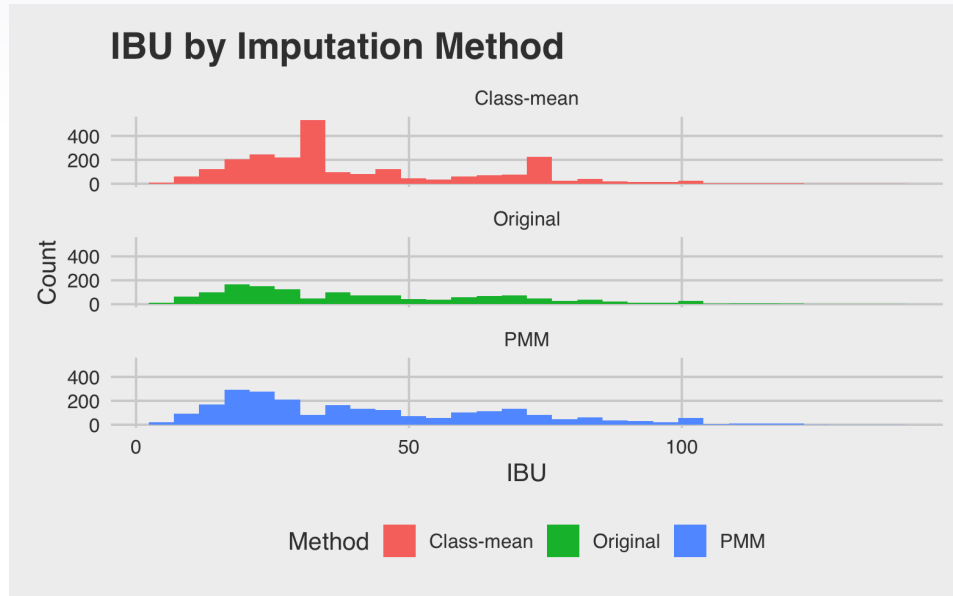
— Distribution of non-missing data

— Distributions for PMM data replacing NA's (5 iterations)



Method 2: Replace missing values with mean of respective Class of Beer

- ▶ PMM mimics original distribution, Class-Mean stacks up at Ales & IPA's (most missing values)
- ▶ Are each of these methods more valid for certain types of analysis?



Summary Statistics for different imputation methods

- ▶ All South Dakota beers are missing IBU data
- ▶ PMM method replaces with values in tail of distributions for their style of beer
- ▶ Class Mean method replaces with IBU values more representative for type of beer
- ▶ SD mean with PMM = 50 and with Class Mean = 34
- ▶ Calculating median values by State, style of beer etc. make more sense with Class Mean replacement

● PMM method

● Class-mean method

