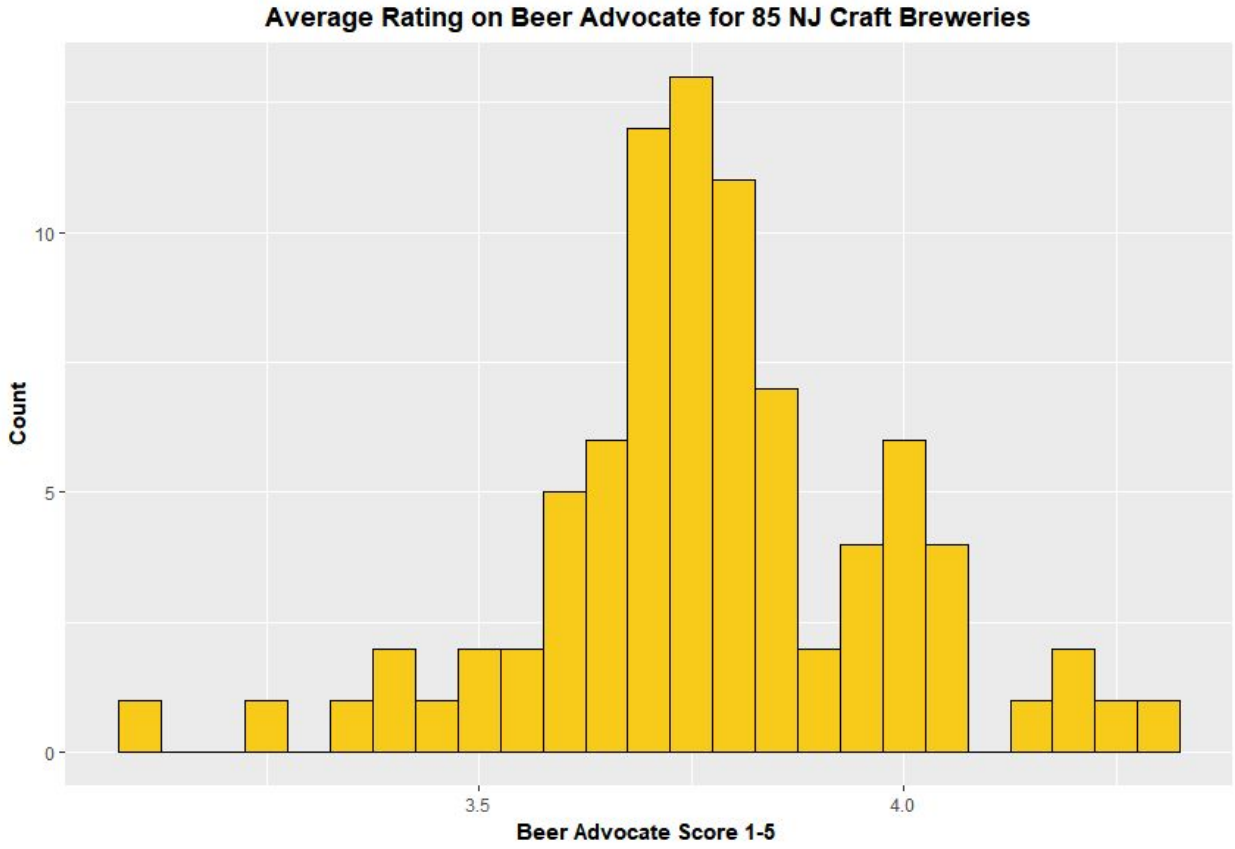
Michael Candella
Sample Exploratory Data Analysis Report

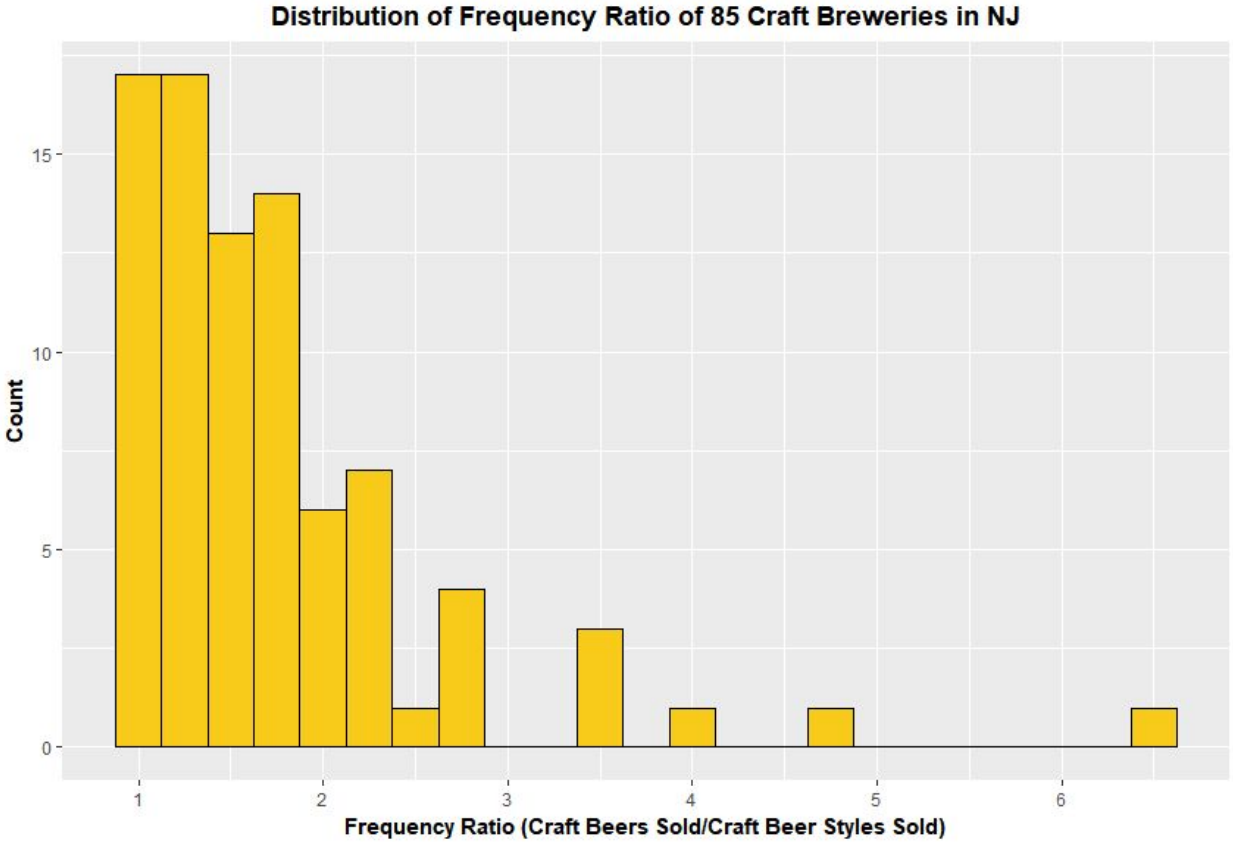
Introduction: In this report I plan to look at the relationship between beer styles, and brewery ratings for craft breweries in New Jersey. Craft beer is known for its variety of styles , part of the experience is trying all the different styles a brewery has to offer. Despite the differences in styles, some breweries choose to keep style variation to a minimum and choose the standard fan favorites such as the India Pale Ale. In this report I plan to use linear regression and exploratory data analysis to find out if there is a relationship between a Brewery's Rating and Beer Style Variety.

For the beer ratings I used Beer Advocate and combined the scores of 85 NJ Breweries and the distribution is as follows.



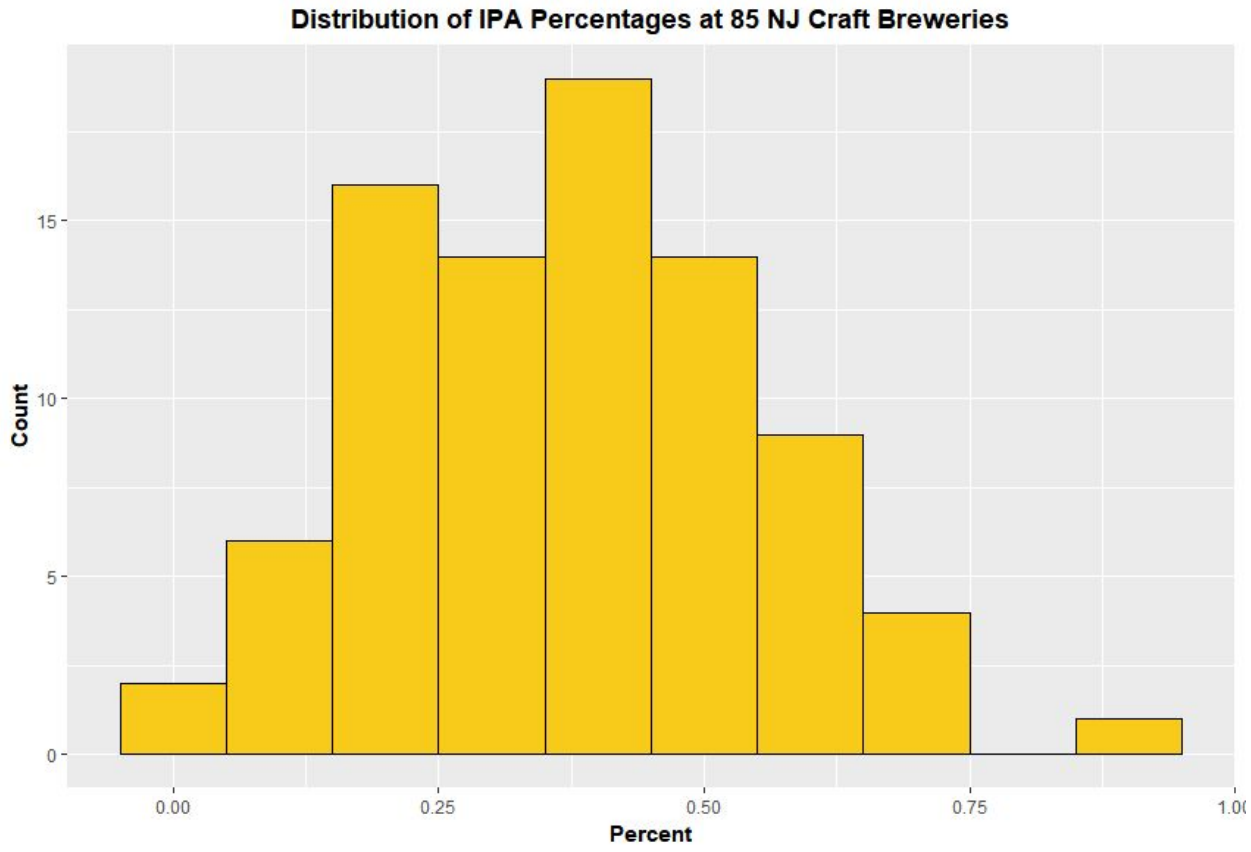
The data ranges from 3.11 to 4.30 with a median of 3.770, mean of 3.774, and a standard deviation of .21. The Data seems perfectly centered around the 3.77 mark with a handful of breweries lingering off in both directions.

Now let's consider the distribution of Beer Styles and the Number of IPAS at a given brewery. To do this I created two separate metrics to account for brewery size. One is the frequency ratio which evaluates beer style diversity relative to how many beers have available, that is (# of Beers/# of Styles). The lower the beer frequency ratio, the more diverse beers a company has with 1 being a perfect 1 to 1 beer/style ratio. Secondly, there is the Pale Percentage, simply the Percent of beers at a brewery that are of the Pale Ale Variety. The higher the percentage the larger the proportion of the brewery consists of Pale Ales.



The Summary Statistics for the frequency ratio are
Min:1
Q1:1.22
Med:1.5
Mean:1.76
Q3:2.05
Max:6.385
Std Dev:.88

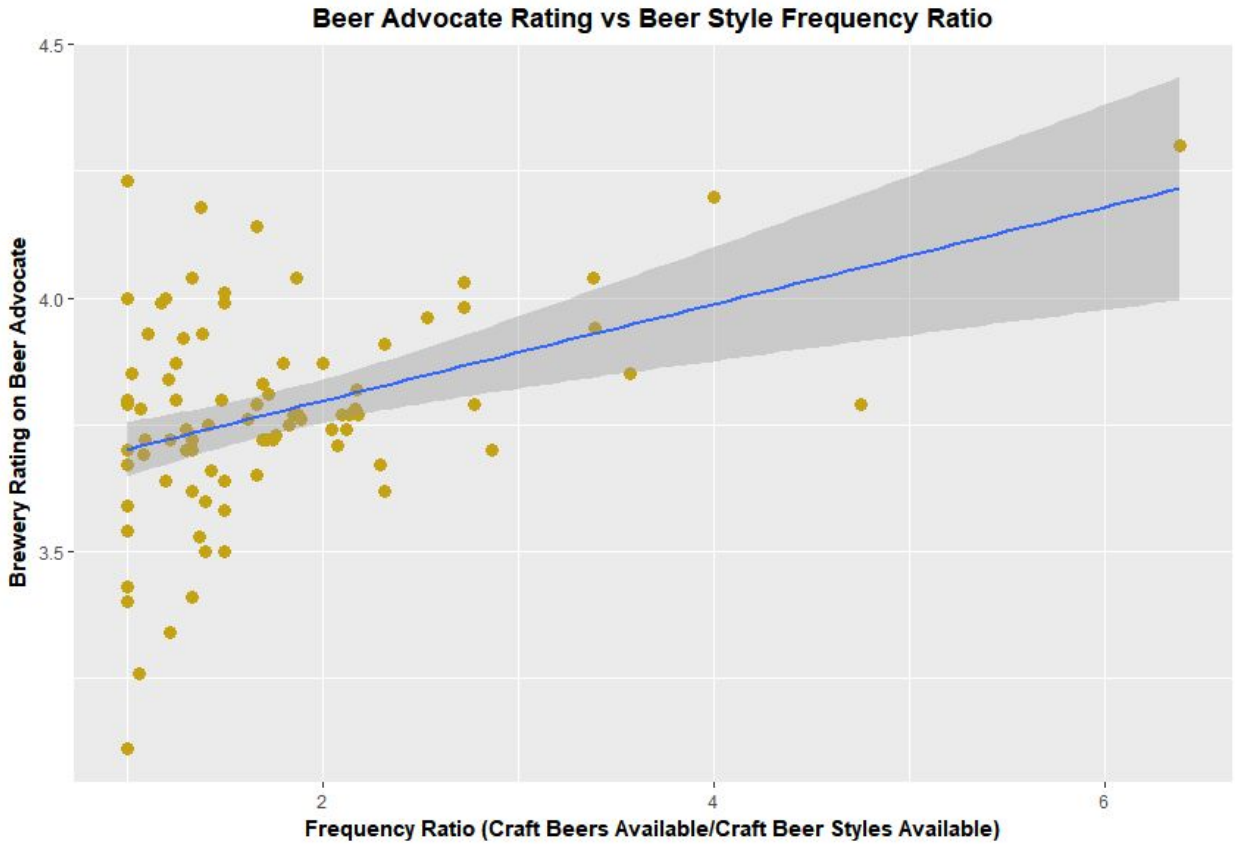
The Average Frequency Ratio shows that for every 1 beer style a brewery has, we expect about 1.76 beers in general so approximately 2 beers per beer style available. These results are relatively diverse considering how many breweries fall on the tail end of this distribution having much less beer-style diversity.



The Distribution here is a lot more evenly distributed with the summary statistics as follows:
Min: 0%
Q1: 25%
Med: 38%
Mean: 37%
Q3: 50%
Max: 88%

The average Pale Ale Percentage being 38% is quite large considering how much other beer styles are out there.For a single style to take up nearly 40% of all the beers that NJ Craft breweries have to offer, there must be a large incentive for brewers to produce beers of this style.

Now to run a standard regression model alongside some scatter plots to determine if there is a linear relationship between Average Beer Advocate Rating and Style Diversity and IPA Percentages.



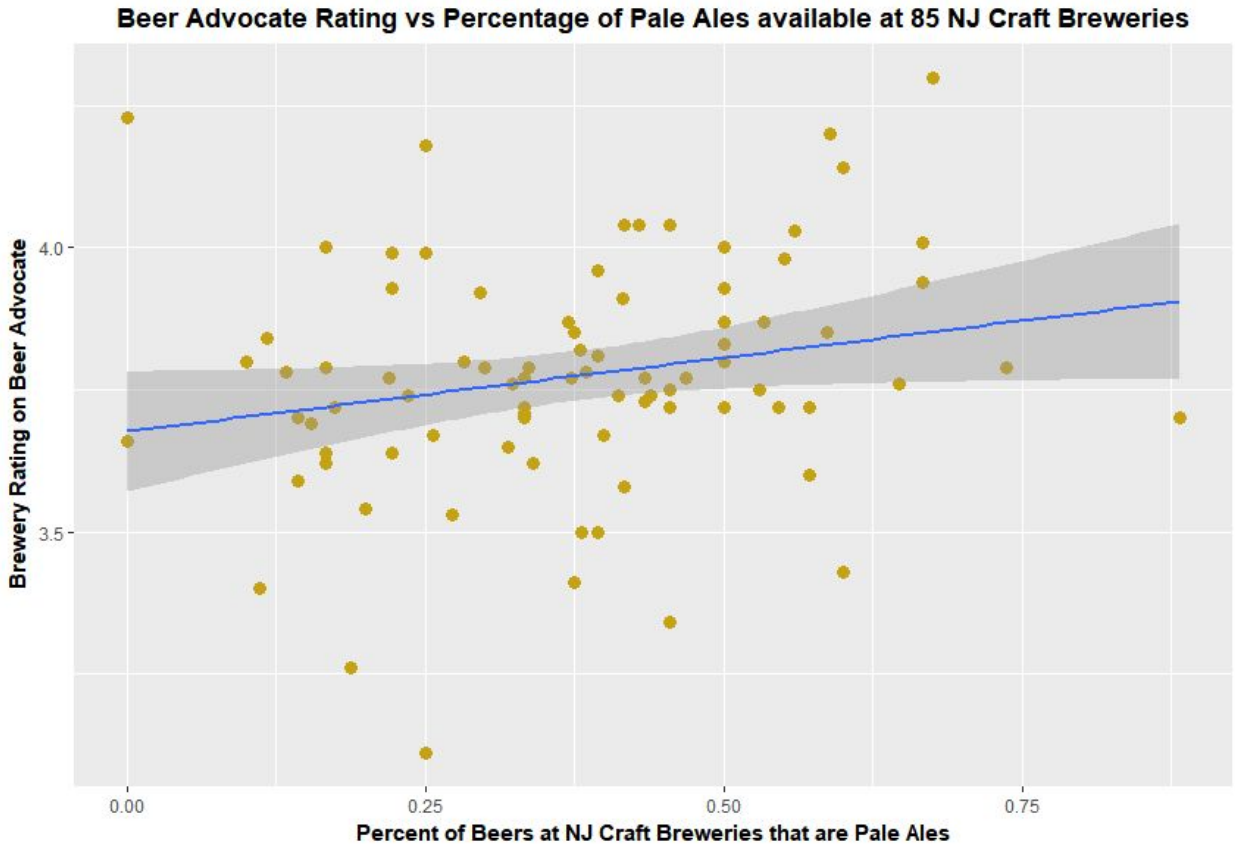
Model: Avg_Rate=3.61+.095(Freq_Ratio)

Frequency Ratio Elasticity on Average Beer Advocate Rating: .205

	Estimate	Standard Error	t-value	p-value
Intercept	3.605	.0462	78.067	<2e-16
Freq_Ratio	.09550	.0235	4.065	.000108

Looking at the estimates and significantly low p-values, it is clear that there is a positive relationship between the Beer Frequency Ratio and Brewery Rating. This means that the more beers you have per Beer Style the more likely you are to have a higher rating. This means that those who travel to breweries favor breweries that focus in on a specific style much more than those who thinly diversify. Elasticity is calculated by taking the coefficient of your independent variables and multiplying it the mean for the dependent variable over the mean of the independent variables. The elasticity determines how much a dependent variable will change

when you increase the independent variables by 1 point. The elasticity for the Frequency Ratio on the Average Beer Advocate Rating is .205 which means that for every 1 point increase in the Frequency Ratio, your Beer Advocate rating is expected to go up by .205 points.



Model: Avg_Rate=3.68+.2598(Pale_Percent)

Frequency Ratio Elasticity on Average Beer Advocate Rating: 2.61

	Estimate	Std. Error	T-value	P-value
Intercept	3.68	.0529	69.568	<2e-16
Pale_Percent	.25981	.128	2.023	.0463

Looking at the P-Values and Estimates for the Regression it is clear that there is a statistically significant relationship between the Percentage of Beers that are Pale Ales and Brewery Rating. Intuitively this makes sense as IPAs dominate the craft beer market so breweries that have more IPAs will ultimately have higher ratings. The Rating Elasticity of the Pale_Percentage is 2.61, however since Pale_Percent is a percent value it's best to divide it by 10 and say that for every 10 percent rise in the amount of Pale Ales you have available, your brewery rating goes up by .261 on Beer Advocate.

Conclusion: What this Data shows is that the strategy of making as much IPAs as possible and specializing on a specific beer style (as opposed to moving onto a different styles per new beer) is a valid one if you're looking to increase your ratings among craft beer aficionados. Beer Advocate is a website dedicated to those heavily engaged in the craft beer industry as opposed to your average beer drinker so it is interesting to see that the most knowledgeable fanbase also aligns heavily with the average craft beer market.