

Do objective measures of product quality appear in taste driven consumer sentiment?

Examining associations between
award winning beer and online
ratings

Eric Janson

MSU Data Analytics Bootcamp

November 2023



Question:

- In the context of a product (beer) where consumers have varying levels of 'evaluation expertise' and perception is heavily influenced by individual taste, can we see any association between objective product quality and consumer opinion?
- Measure of objective quality: significant awards by peer-jury/expert judged panels
- Expectations:
 1. Objective measures of product quality are consistently evident in consumer ratings
 2. Consumer opinion is positively influenced by expert-based evidence of objective quality (opinion shifts after expert evidence of quality)



The Holland Sentinel – Holland, MI

Dataset: Stanford Network Analysis Platform (SNAP) RateBeer Study 2013

- **J. McAuley and J. Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. WWW, 2013. <https://snap.stanford.edu/snap/>**
- Dataset of ~2.8 million individual beer ratings from RateBeer 2000-2012
 - Beer rating website
 - Anyone can create a reviewer profile
 - Beers rated by: appearance (scale of 1-5), aroma (1-10), palate (mouthfeel) (1-5), taste (1-10), and overall (1-20); written description of beer also provided
 - Multiple reviews by same person, but no more than one per product per reviewer
- Independent beer quality indicator: beers awarded any World Beer Cup medal in years 2002-2010
 - Biennial (at the time) international beer competition judged by panels of industry experts
 - Broken into style categories and awarded gold, silver, and bronze (not always winners)
 - Manually coded by hand from Brewers Association website
- What about RateBeer API, BeerAdvocate API, Untappd API, Sales Data (Drizly)?
 - Little to no public access or simply doesn't exist

Input data formatting and processing

```
beer/name: John Harvards Simcoe IPA
beer/beerId: 63836
beer/brewerId: 8481
beer/ABV: 5.4
beer/style: India Pale Ale &#40;IPA&#41;
review/appearance: 4/5
review/aroma: 6/10
review/palate: 3/5
review/taste: 6/10
review/overall: 13/20
review/time: 1157587200
review/profileName: hopdog
review/text: On tap at the Springfield, PA location.
```

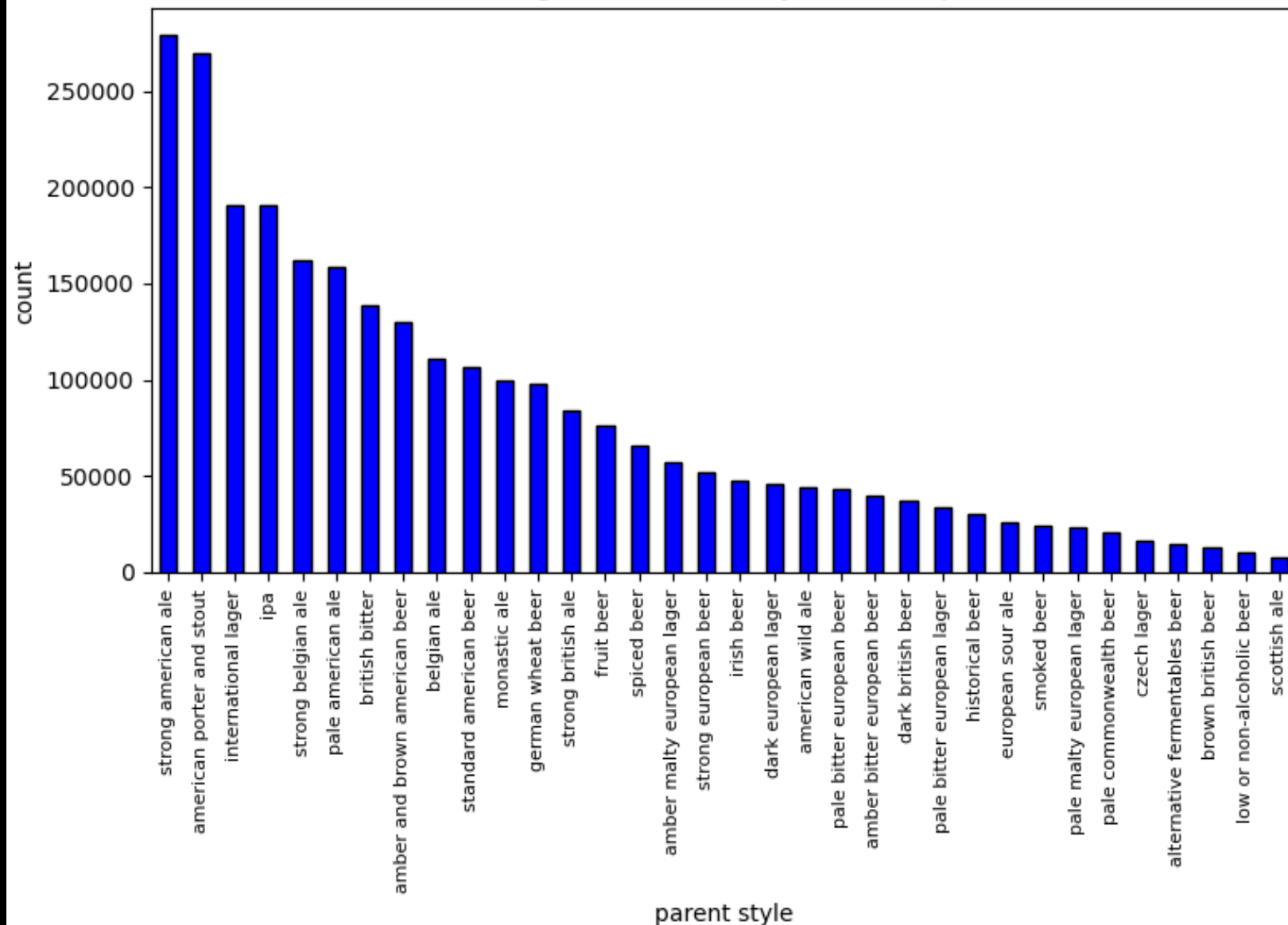
Data processing and cleanup

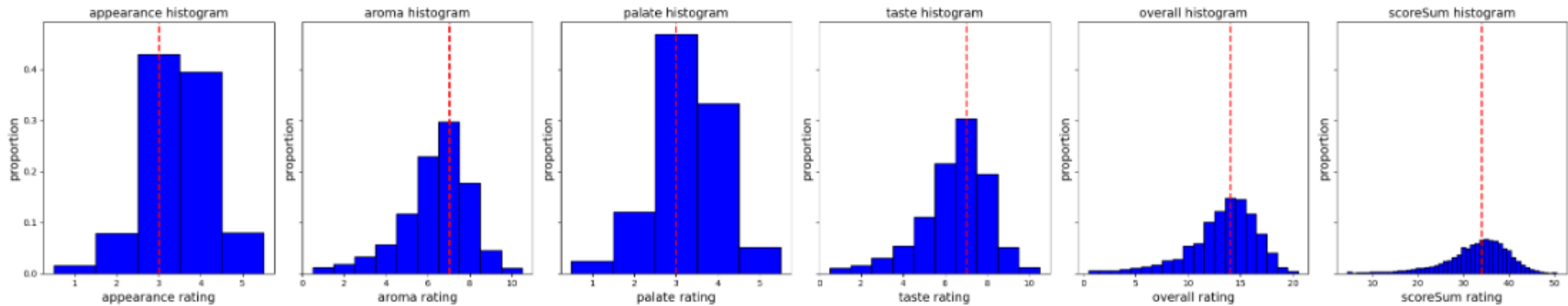
- ~1.6 GB text file
- Ordinal data (subjective ranked ratings)
- All lines were extracted (except review text) for each review
- An unweighted cumulative rating was added to every review (sum of all other ratings)
- Year of review was extracted from “time”
- “Parent style” was added to help reduce dataset complexity

Summary Statistics

Data years	2000-2012
Total ratings	2,750,509
Average ratings per beer	33
Average beer ABV	5.78%
Average ratings per parent style	80,897
Average ratings per year	211,587

Histogram of Parent Styles in Sample

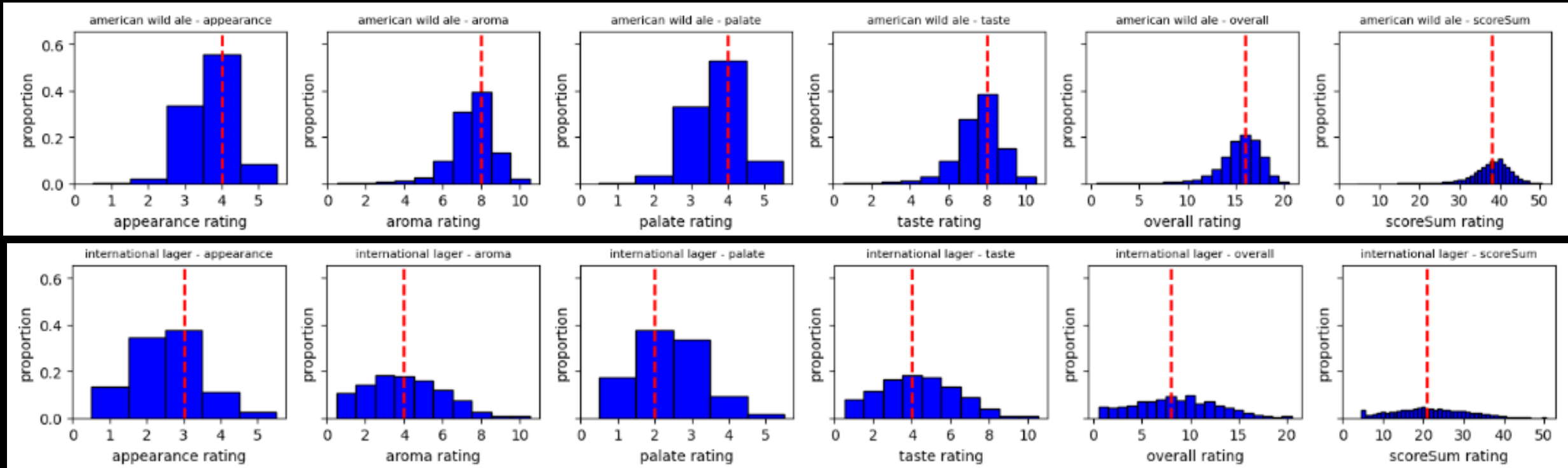




Rating Category	Median	Mode	Range
Appearance	3	3	4
Aroma	7	7	9
Palate (mouthfeel)	3	3	4
Taste	7	7	9
Overall	14	14	19
Unweighted sum of ratings	34	35	45

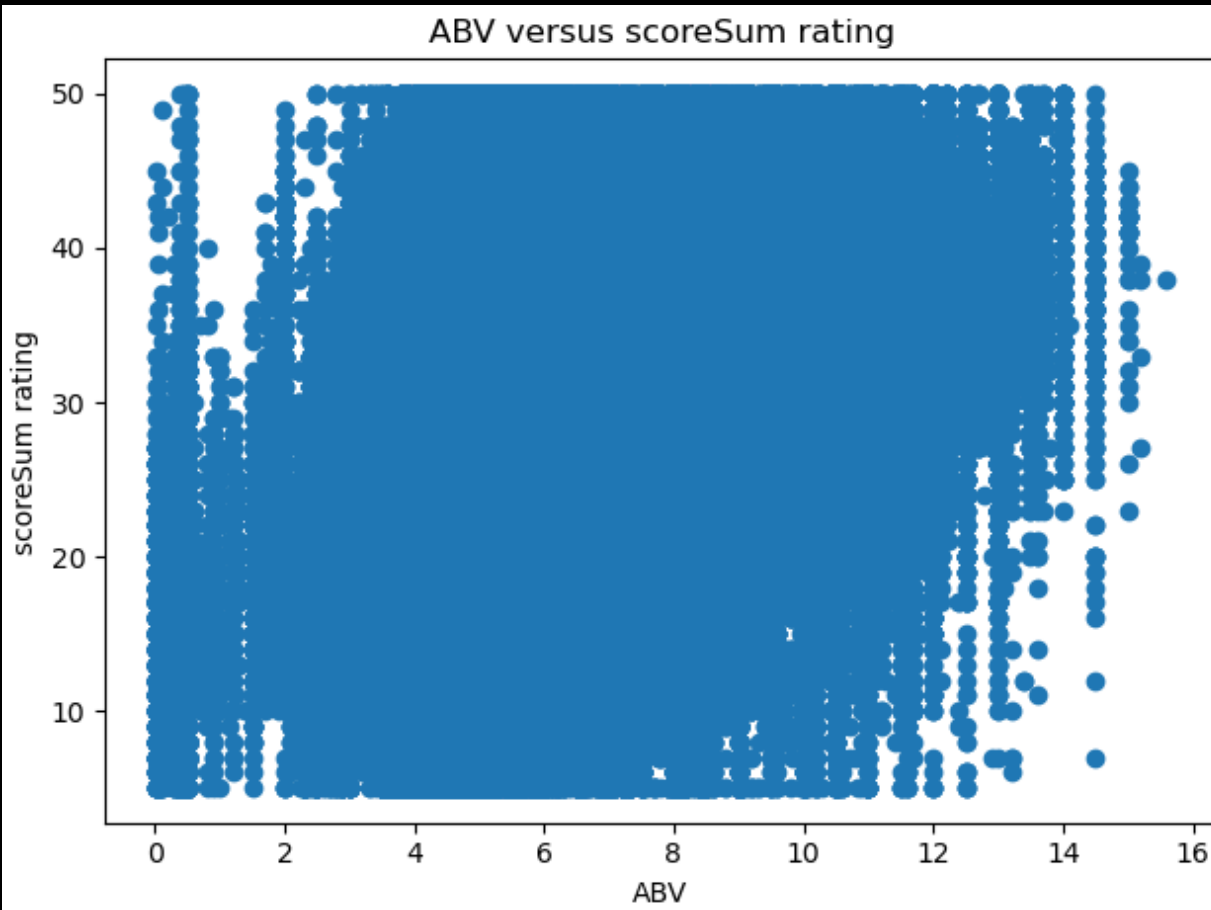
Parent style comparisons – a summary

Kruskal-Wallis H Test Statistic:
828630
P-Value: < 0.0001
Unweighted summed rating

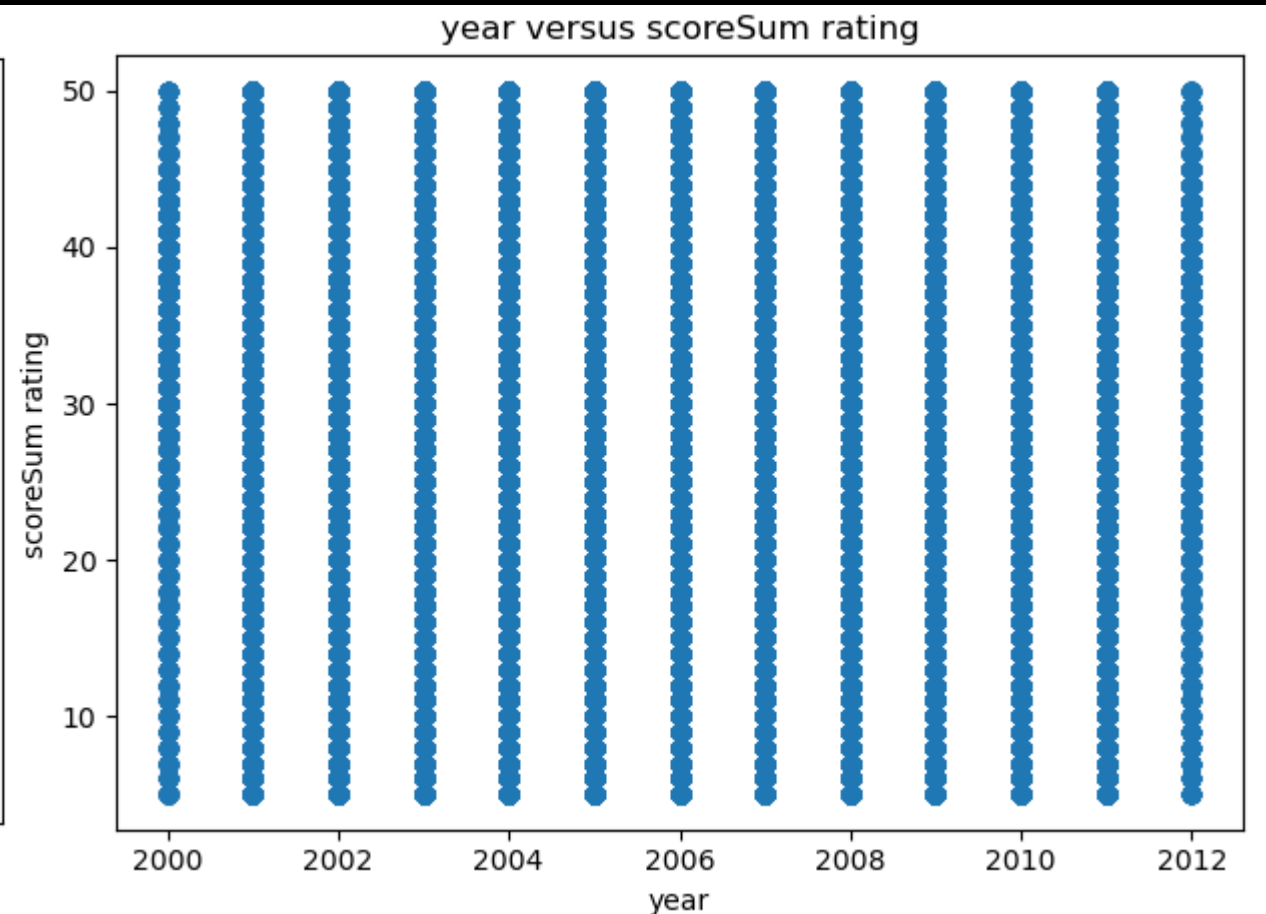


- Thirty-four total parent styles
- Top three styles (highest median summed score): American wild ale, American strong ale, American porter and stout
- Bottom three styles (lowest median summed score): low or non-alcoholic beer, international lager, American standard beer
 - These categories includes some of the most consumed beers in the world. Uh oh!

Correlation of summed rating vs. ABV/year of rating



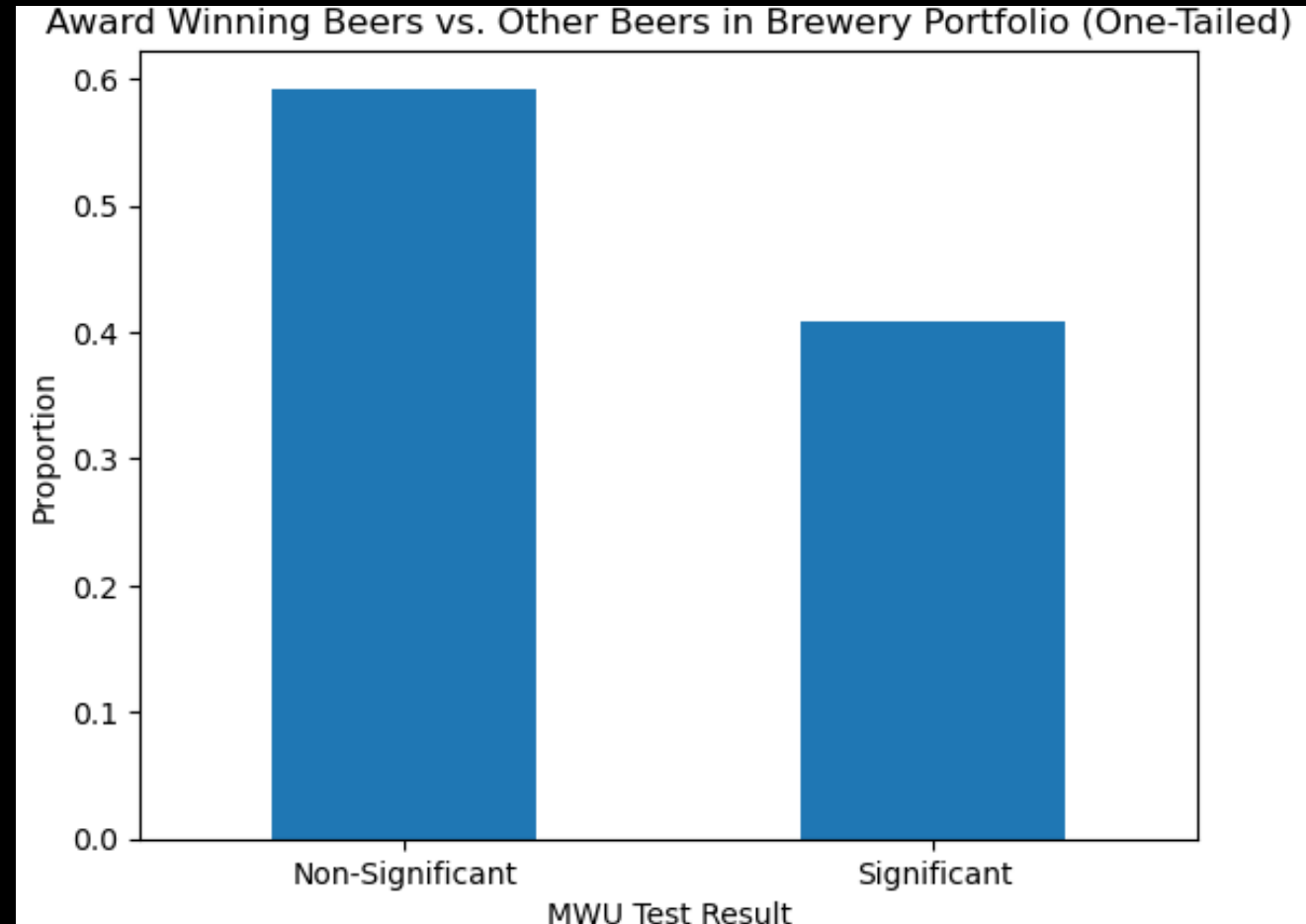
ABV outliers (within style) removed
Spearman rank correlation: 0.50
P-value: < 0.0001



ABV outliers (within style) removed
Spearman rank correlation: 0.06
P-value: < 0.0001

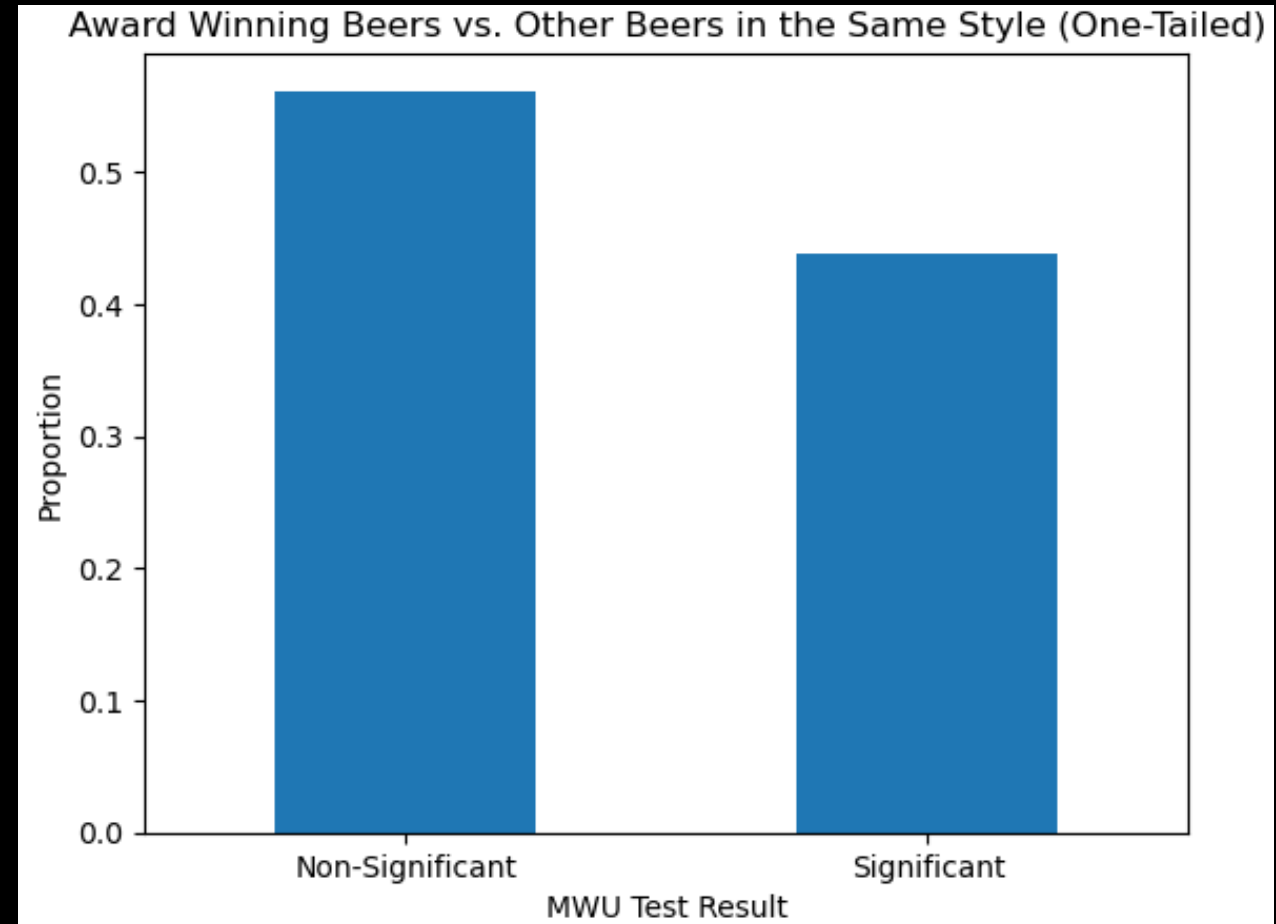
Do award winning beers rate higher than other beers by same brewery?

- Medaling beers within subset of both over- and under- performing style categories (176 beers total)
- Tested: individual medal winning beers unweighted rating sum vs. non-medal winning beers brewed by the same brewery (Bonferroni corrected)
- No consistent evidence stochastic dominance for medal-winning beers vs. all other by same brewery



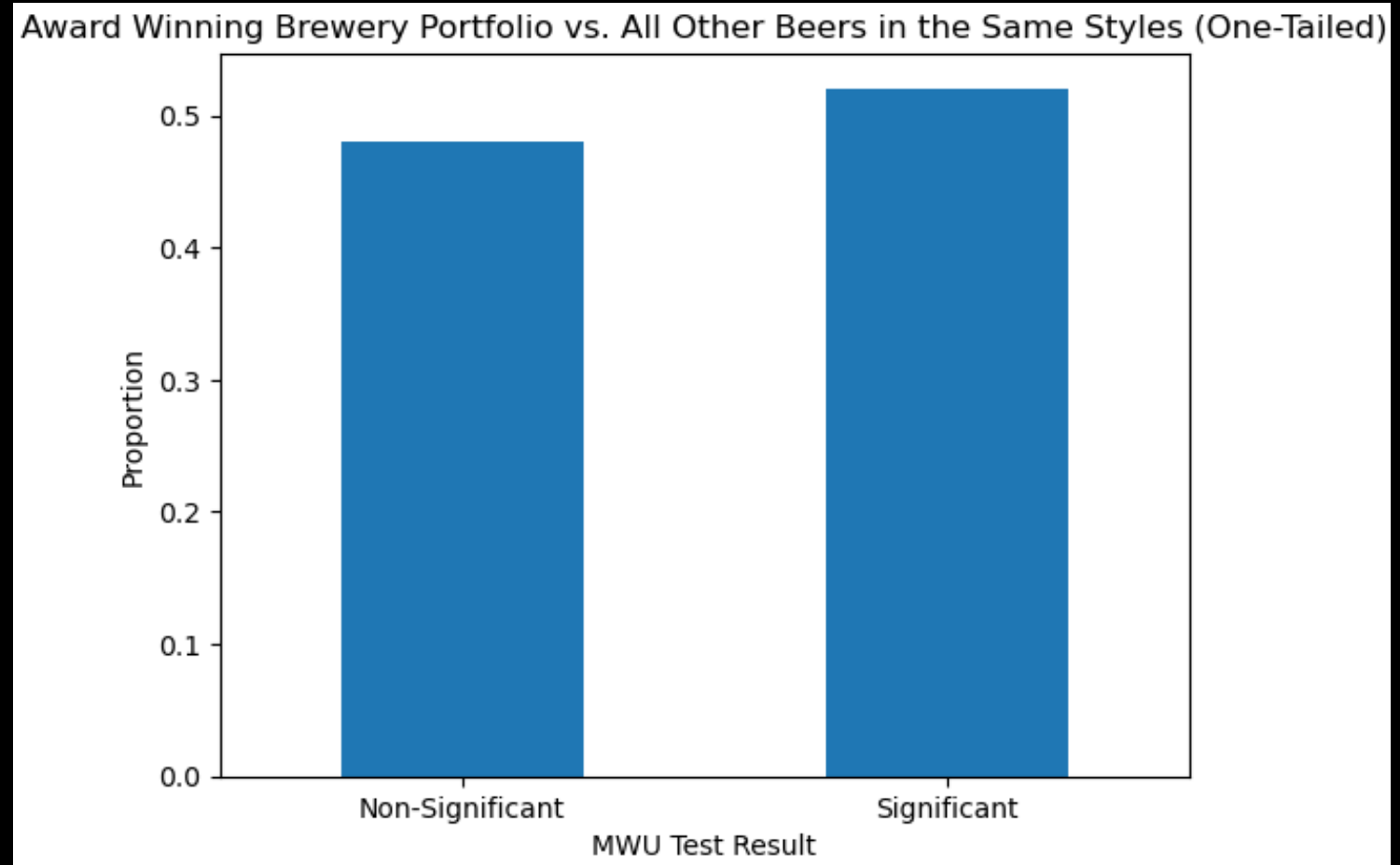
Do award winning beers rate higher than all other beers in the same style category?

- Medaling beers within subset of parent styles chosen for comparisons from both over- and under-performing style categories (176 beers total)
- Tested: individual medal winning beers unweighted rating sum vs. all other beers in the same style category (Bonferroni corrected)
- No consistent evidence of stochastic dominance for medal-winning beers



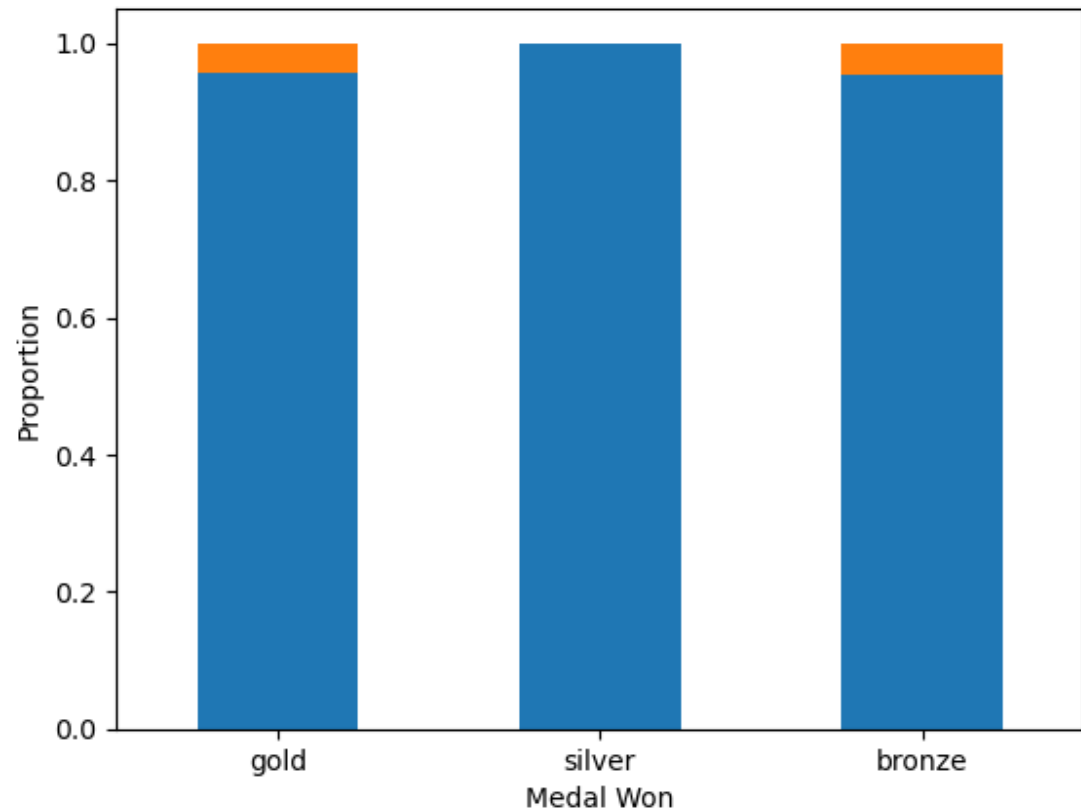
Do all of beers brewed by an award-winning brewery rate higher than all other beers (in the same style categories)?

- Medaling beers within subset of parent styles chosen for comparisons from both over- and under- performing style categories (176 beers total)
- Tested: medal winning brewery's entire portfolio vs. all other beers in the same style categories (Bonferroni corrected)
- No consistent evidence of stochastic dominance for medal-winning beers
 - Strongest evidence of signal so far

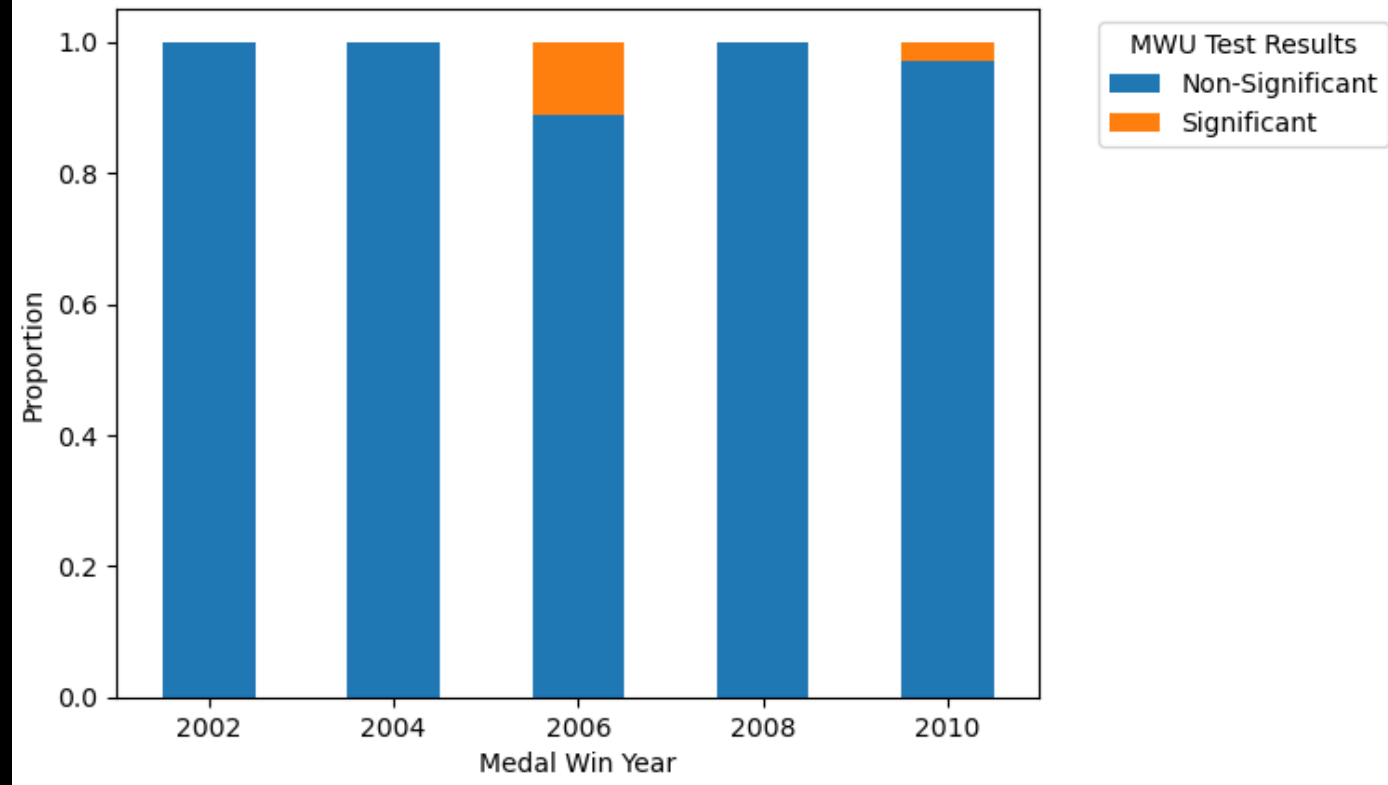


One-Tailed Mann-Whitney U: Before vs After Medal Win

MWU Test Results by Medal Won (One-Tailed)



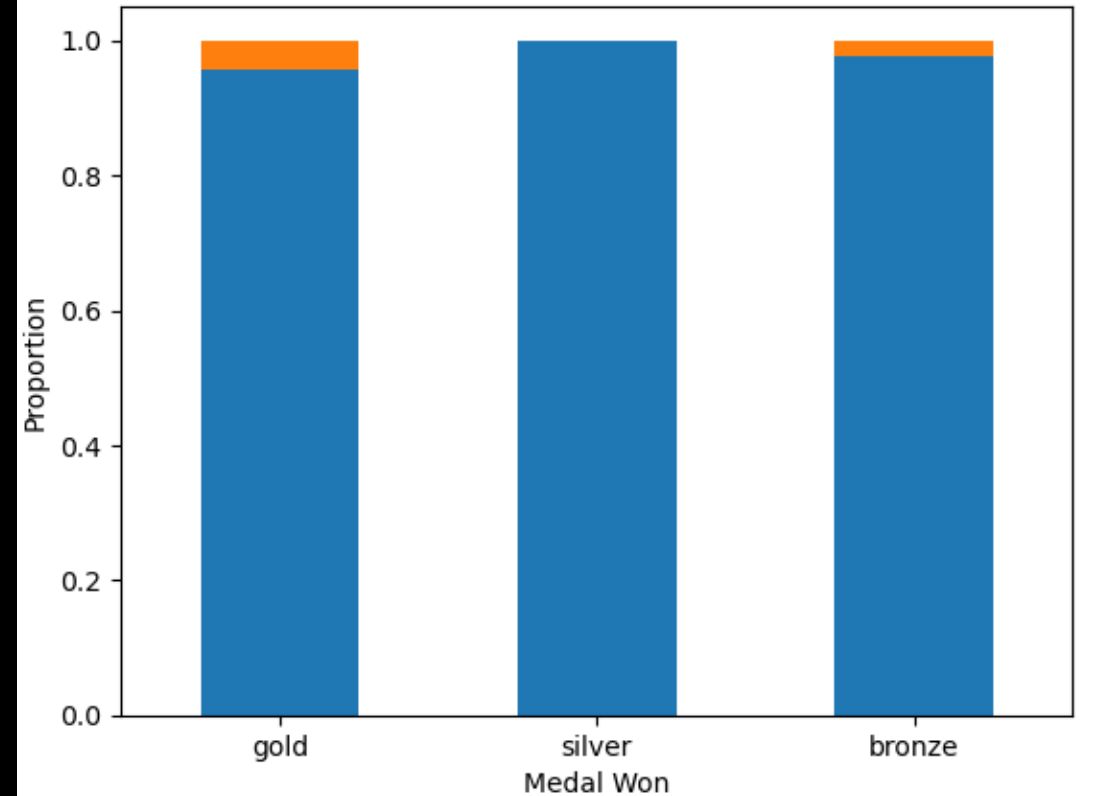
MWU Test Results by Win Year (One-Tailed)



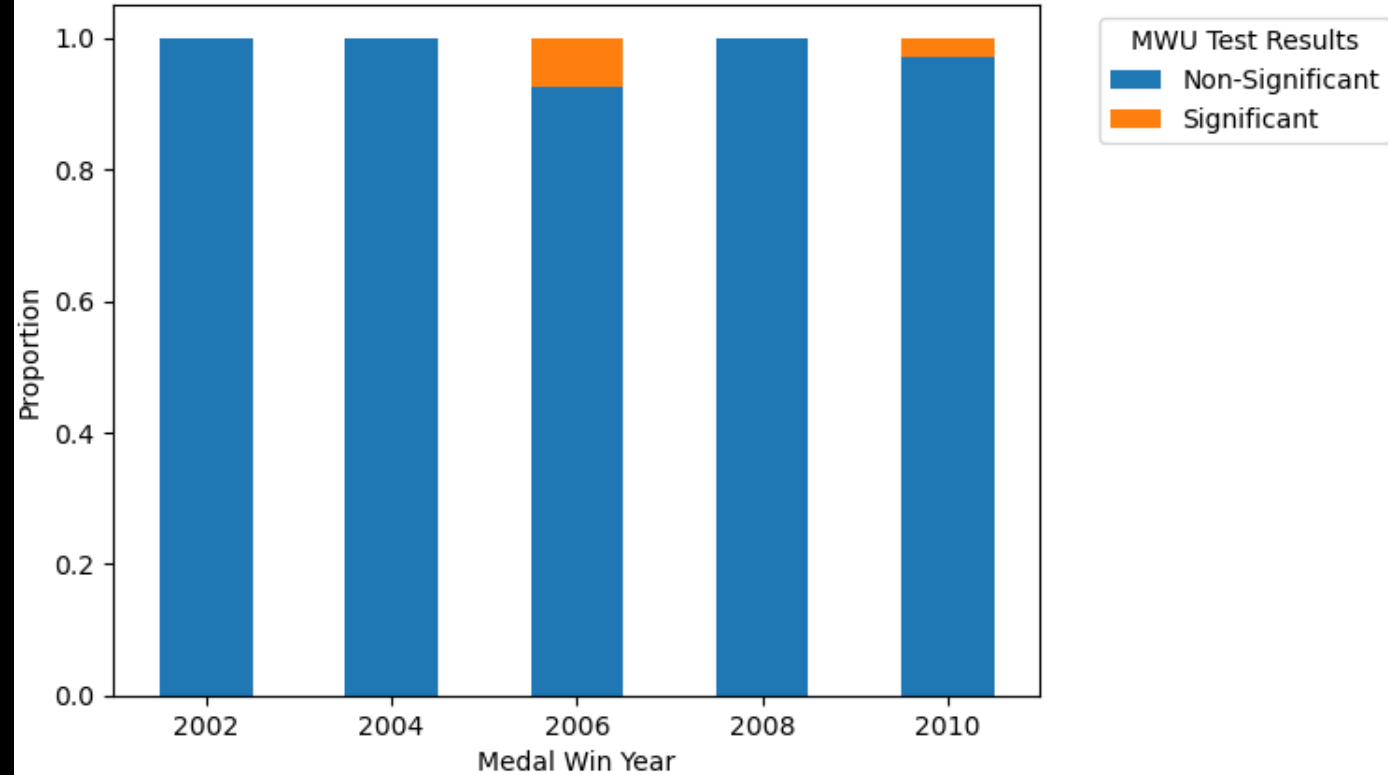
- Sample: medaling beers within subset of both over- and under- performing style categories (176 beers total)
- Tested: individual medal winning beers unweighted rating sum before and after medal win (Bonferroni corrected)
- No evidence of consistent increase in stochastic dominance for after medal win

One-Tailed Mann-Whitney U: \leq Two Years After Medal Win

MWU Test Results by Medal Won with Short Window (One-Tailed)



MWU Test Results by Win Year with Short Window (One-Tailed)



- Sample: medaling beers within subset of parent styles of both over- and under- performing style categories (176 beers total)
- Tested: individual medal winning beers unweighted rating sum before and after medal win within 2 years of winning (Bonferroni corrected)
- No compelling evidence of consistent increase in stochastic dominance \leq 2 years after medal win

Summary conclusions

- Little evidence of a strong and consistent relationship between consumer opinion and objective beer quality
 - Sales data may bear it out (probably not)? → Highest consumed beers are also some of the lowest rated by style category
- Why?
 - Consumers are not objectively evaluating each beer to strict guidelines (like in a beer competition) → too much personal taste influencing ratings
 - e.g., why is there correlation between alcohol content and ratings and significant differences in rating distributions by beer style
 - Attempting to look for signal overlap between two mostly non-overlapping populations: experts and non-experts
- Awards are nice for brewers → probably less useful as a long-term marketing tool
- Then, what factors drive some beers/breweries to rate statistically higher than others?
 - Some specific combination of branding/marketing influence, product rarity, sampling error, and sensory/chemical factors (objective quality) not directly measured here
 - Ratings likely very different for experts vs. above-average knowledge consumers vs. an average consumer
 - Don't make purchasing decisions based on online ratings for products that are a matter of taste!