

# An Analysis of the Correlation Between Alcohol Consumption and Sports Betting Volume

Data Spirits: amanj, jeveleth, jvandevo

## Hypothesis

We sought to determine whether there is a statistically significant correlation between state alcohol consumption and sports betting. We tested the hypothesis that there is a strong positive correlation between alcohol consumption and sports betting.

## Data

We had four major tables: population, alcohol, sales, and bets. We used interpolation and machine learning regression models to create a table with columns state, year, month, population, beer (gallons), spirits (gallons), wine (gallons), handle (\$), gross\_revenue (\$), and hold (\$).

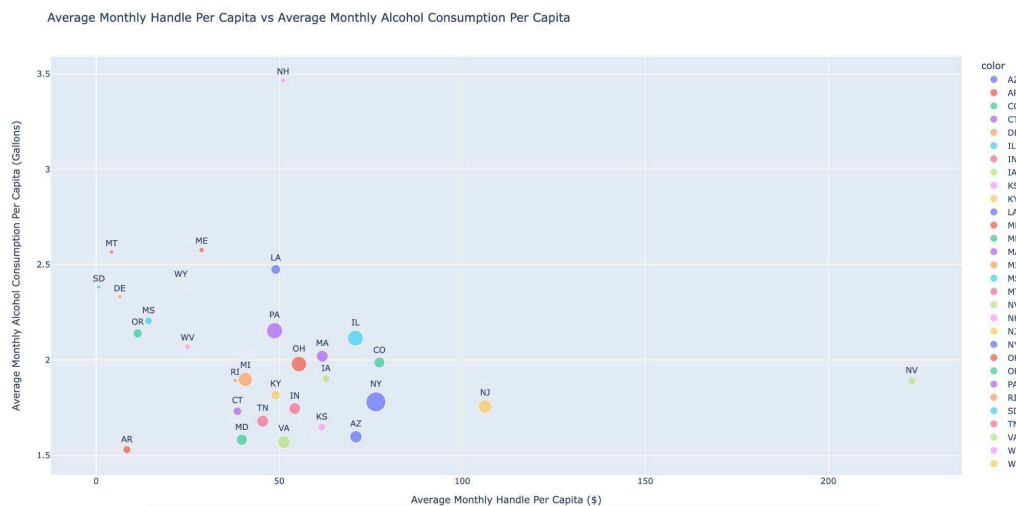
## Findings

**Claim #1:** There is not a significant strong positive correlation between alcohol consumption and sports betting, state-wise.

**Support for Claim #1:** 360 Pearson's correlation tests were performed for 12 pairs of variables across 30 states. Only 26 were statistically significant, with 12 of them being positive correlations (and the others negative). West Virginia and Massachusetts have significant weak positive correlations, but no other states had significant positive correlations.

**Claim #2:** There is a weak negative correlation between alcohol consumption and sports betting, nationally.

**Support for Claim #2:**

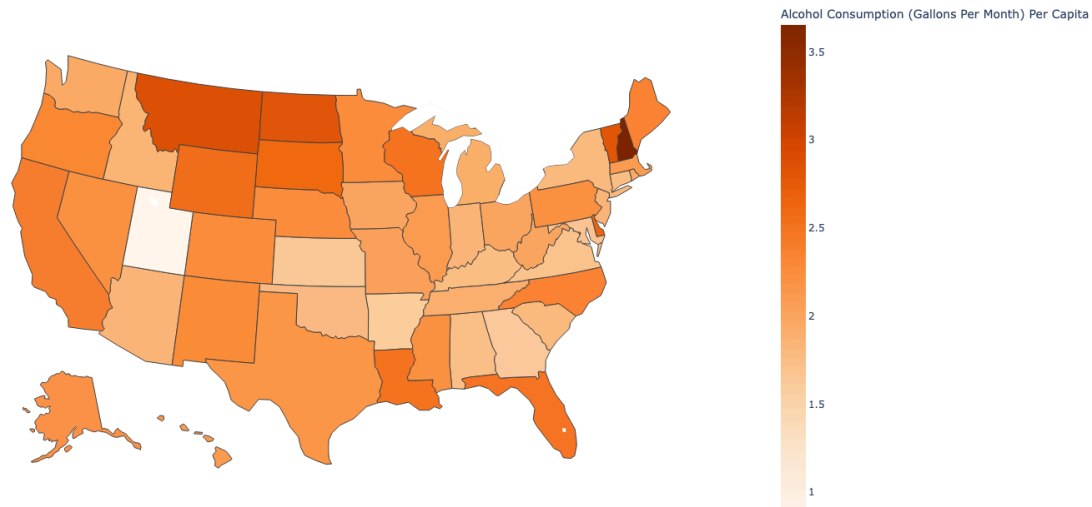


There is a  $r = -0.36$ ,  $p = 0.05$  Pearson correlation between the variables, a significant weak negative correlation. This is supported by what we see in the scatter plot above.

**Claim #3:** There is a statistically significant difference in alcohol consumption across states.

**Support for Claim #3:**

United States Alcohol Consumption (Gallons)

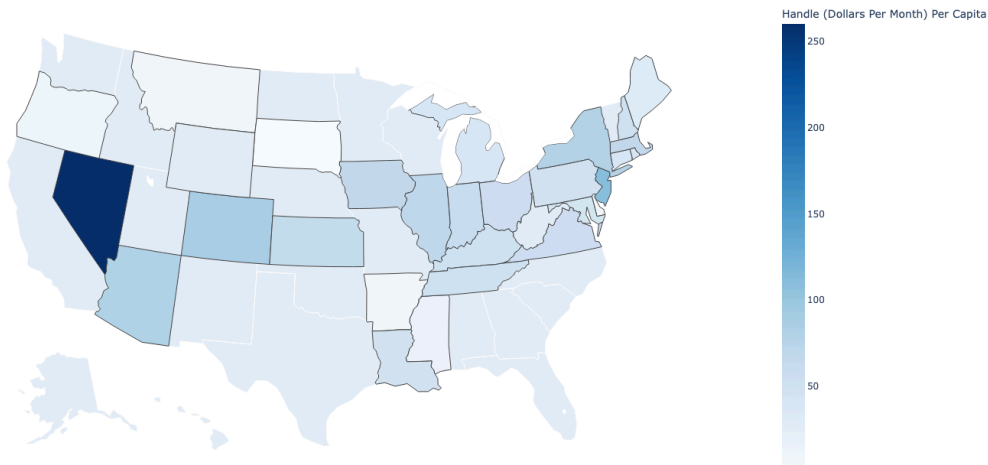


We conducted 7,650 t-tests between every pair of states for each of the beer, spirits, and wine variables per capita and 81.8% of the tests had a  $p$ -value of less than 0.01. A heatmap that colors states by their total consumption per capita is above.

**Claim #4:** There is a statistically significant difference in bet handle across states.

**Support for Claim #4:**

United States Bet Handle (\$)



We conducted 870 t-tests between every pair of the 30 states that had bet handle data and 79.0% of the tests had a  $p$ -value of less than 0.05. A heatmap that colors states by their bet handle per capita is above.

## Socio-historical Context

One of the main societal factors that affected our data was the legality of online sports betting. Our analysis contains months where we only have the data for some states because it wasn't legal yet. According to CBS News, North Carolina still cannot bet online (while they can in-person), and Tennessee can only bet online (and cannot bet in-person) [4]. As more states start to legalize betting, this will likely increase the amount of sports betting throughout the US. As a result, analysis like ours is likely to be more fruitful in a few years.

Another societal factor is the effect of casinos and drinking on sports betting. Areas that have more casinos are likely to have increased drinking and sports betting. For example, Native American reservations and Las Vegas are both likely to have more betting. We also see this reflected in our dataset, with Nevada having the highest per capita betting handle by a large margin. In California for example, gambling has been outlawed (except for the lottery and similar) and it only occurs on Native American reservations [5]. These are strong societal factors in the prevalence and distribution of betting and drinking. This is important because Native Americans and other groups are a higher likelihood of substance dependence [6], so they are especially susceptible to gambling and other effects of casinos.

The most immediate stakeholders for this project are the online gambling website like BetMGM and Draft Kings [7]. They could benefit from our analysis by gleaning trends in their customers and learning about the habits of potential new customers. They could also use this information to target and potentially manipulate people into substance abuse. Another stakeholder are people with substance addiction issues, since drinking and gambling are often associated with this, they are a major stakeholder. Another stakeholder is liquor stores, wine & spirits, and bars. They are all related to drinking and are affected by the trends we see in our analysis.

One study sampled 4363 adults about their drinking and sports betting habits and found that drinking was correlated with sports betting [1]. This provides a strong basis for our hypothesis and shows that our research question is worthwhile. This study has large societal impact because it demonstrates that drinking and sports betting are indeed related. Our findings are important for law makers and the sports betting industry.

## Ethical Considerations

Our data may have some underlying biases against people of lower socioeconomic standing. In fact, it has been found that it is harder to recruit and obtain data from individuals of low socioeconomic standing due to issues with accessibility and trust [2]. As a result, the data collection process specifically for alcohol consumption is most likely based on individuals of higher socioeconomic standing. In addition, the alcohol consumption data from our visualizations can also be misinterpreted. In our visualizations, we plot alcohol consumption in gallons, but not by total ethanol volume. This may lead individuals to make inaccurate assumptions on which states drink more. This can be particularly harmful if an individual arrives with some biases against a particular state, where they can interpret the data incorrectly to reinforce their biases.

We made a great effort to ensure a particular community privacy was not affected in this project. Since we used state level data, and there are no identifying features to any particular region or individual in a state, we feel confident that no privacy was harmed in this study. That being said, the data and visualizations can be misinterpreted. In the capstone portion, the heat maps demonstrate state clusterings, but the colors could be misinterpreted for other meanings. As a result, we made a note that the colors do not represent anything other than grouping. We also visualize aspects in differing quantities (total, per capita, etc ...), which if the user is not paying attention may not recognize the different scales. Finally, in the capstone component, we also include a bar chart race. Bar charts have been known to be misinterpreted [3], so we made efforts to ensure the scaling, coloring, units, and values are all easily accessible to limit confusion.

## Works Cited

- [1]<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2816784#>.
- [2]<https://journals.sagepub.com/doi/full/10.1177/25152459231193044>
- [3]<https://www.sciencedaily.com/releases/2022/02/220203102536.htm>
- [4]<https://www.cbssports.com/general/news/u-s-sports-betting-here-is-where-all-50-states-currently-stand-on-legalizing-online-sports-betting-sites/#>
- [5] [https://lao.ca.gov/2007/tribal\\_casinos/tribal\\_casinos\\_020207.aspx](https://lao.ca.gov/2007/tribal_casinos/tribal_casinos_020207.aspx)
- [6] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3603686/>
- [7]<https://www.cnbc.com/2024/03/27/largest-us-sportsbooks-join-forces-to-tackle-problem-gambling.html>