

Coffee Chronicles: Brewing Health or Harm?

Hanen Zeddini

Site Xiang

Matthew Garcia

Tatiana Oropel



Does the amount of coffee a person consumes influence their overall health?

H0 - There is no statistical significance between the amount
of coffee consumed and health problems

H1 - There is a statistical significance between the amount
of coffee consumed and health problems

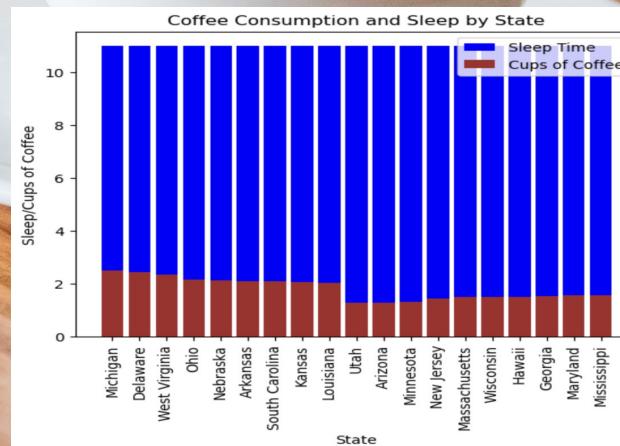
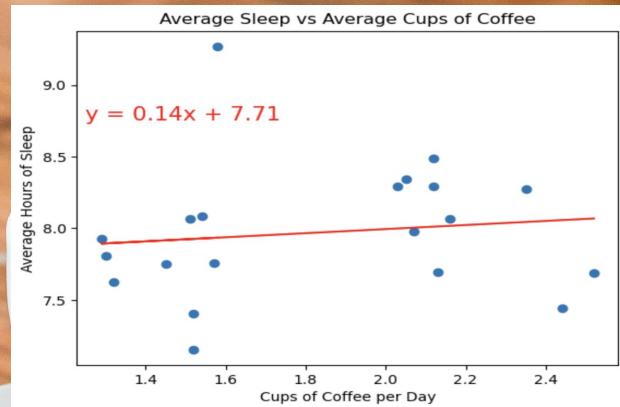


What is the correlation between coffee consumption rates and insomnia?

- The p value ($7.424251262648384e-34$) is very small and suggests strong evidence against the null hypothesis.
- This means the observed result is highly unlikely to occur by chance, which supports the rejection of the null hypothesis.
- The r value of about 0.14 suggests a very weak positive linear correlation.
- Overall the t statistic and p value together suggest that the result is significant, that it is unlikely due to random chance, and we may reject the null hypothesis
- the r value suggest a positive, but very weak trend in the data.

```
t_statistic, p_value = stats.ttest_ind(coffee_df["Average Sleep"], coffee_df["Cups of Coffee"], equal_var=False)
print(f"T Statistic: {t_statistic}")
print(f"P-value: {p_value}")
alpha = 0.05
if p_value < alpha:
    print("The correlation is statistically significant.")
else:
    print("The correlation is not statistically significant.")
```

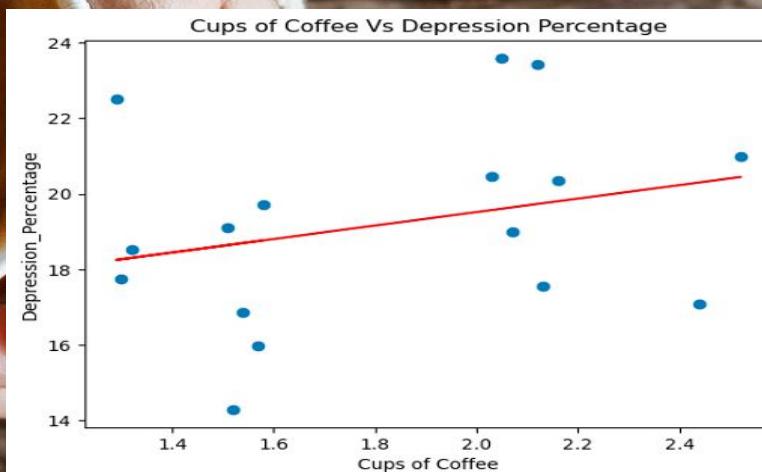
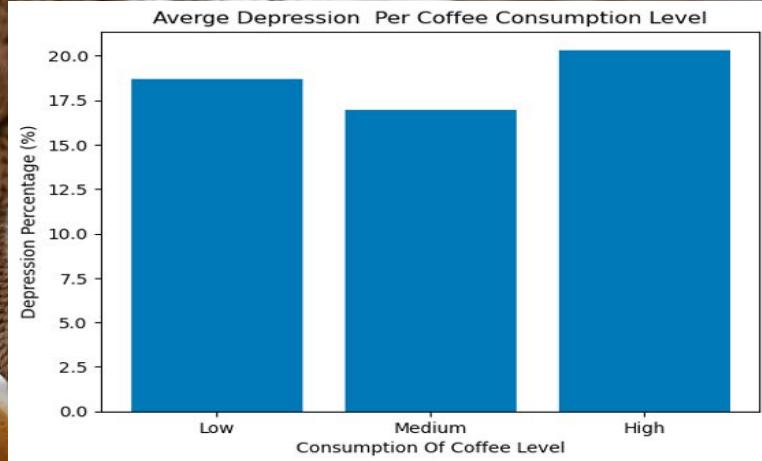
T Statistic: 44.43784243695134
P-value: 7.424251262648384e-34
The correlation is statistically significant.
The r-value is: 0.14371332329726605



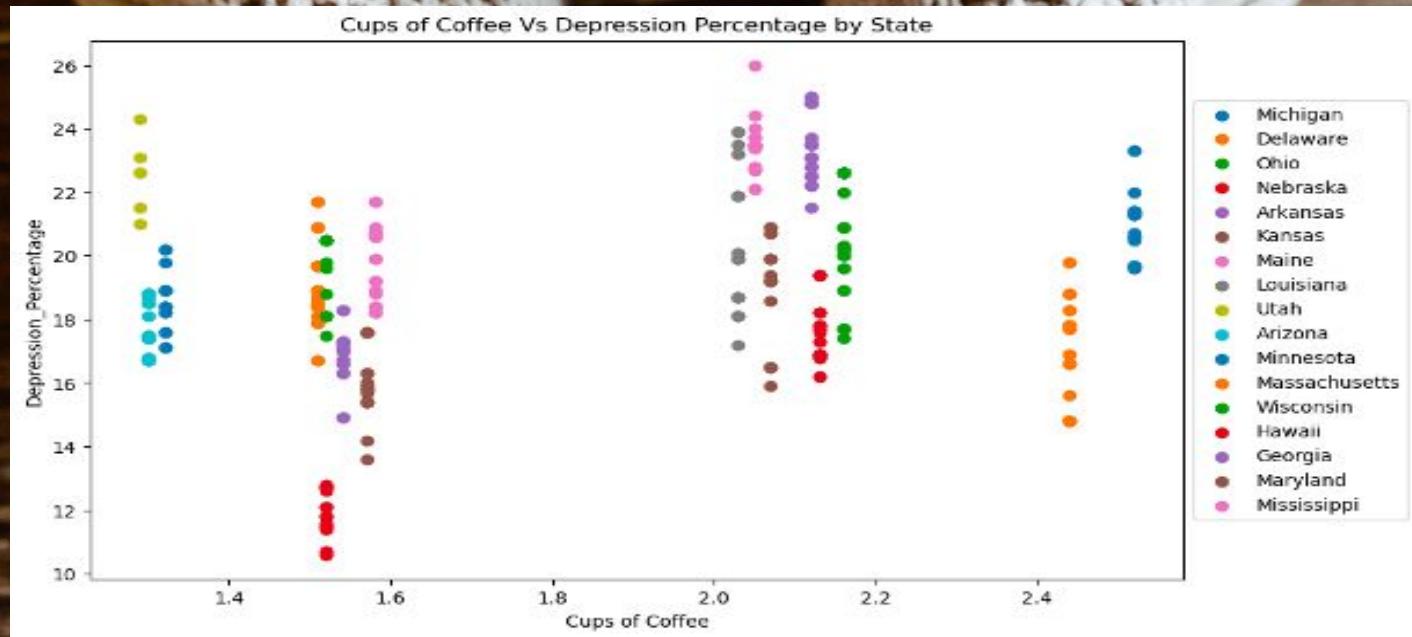
Is there a Relation between Coffee Consumption & Depression Percentage?

The bar plot presents the mean depression percentages across three categories of coffee consumption: Low, Medium, and High.

The correlation between cups of coffee and the depression percentage is 0.28.



Relation Between Coffee consumption & Depression Percentage by State

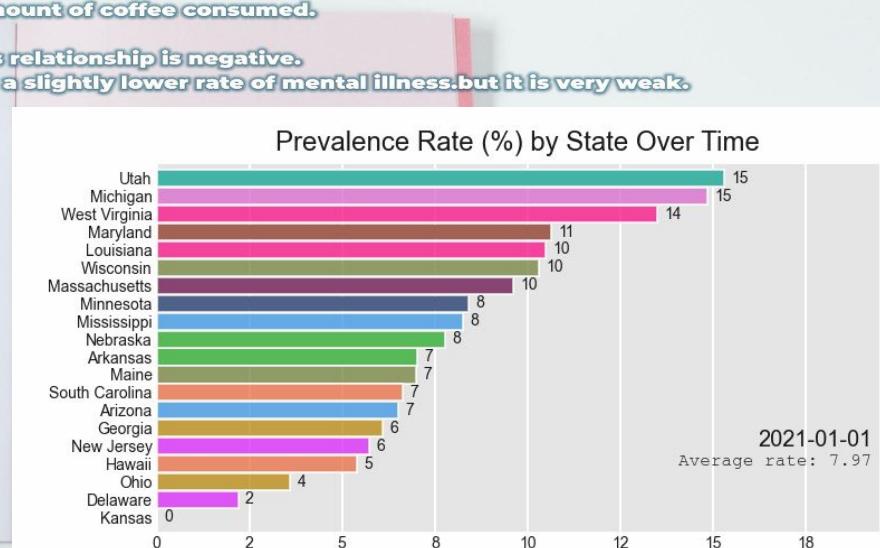
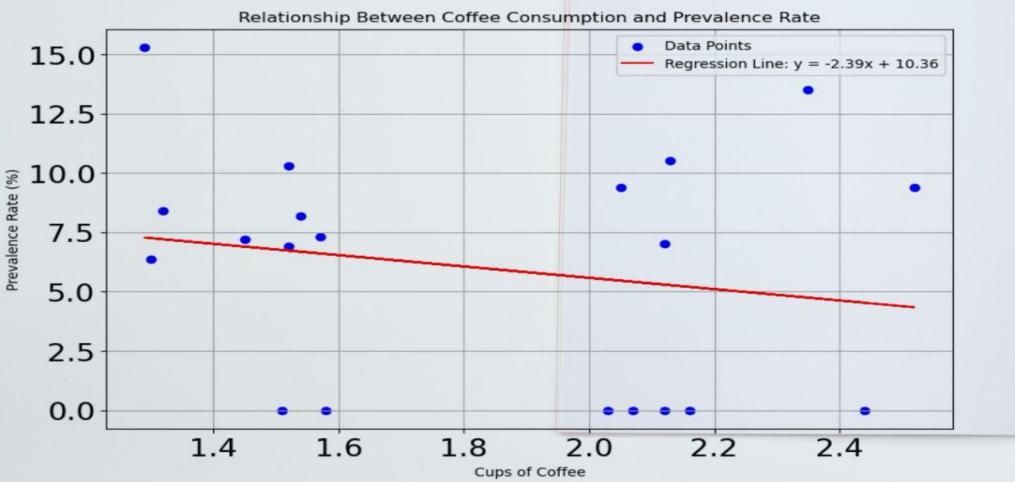


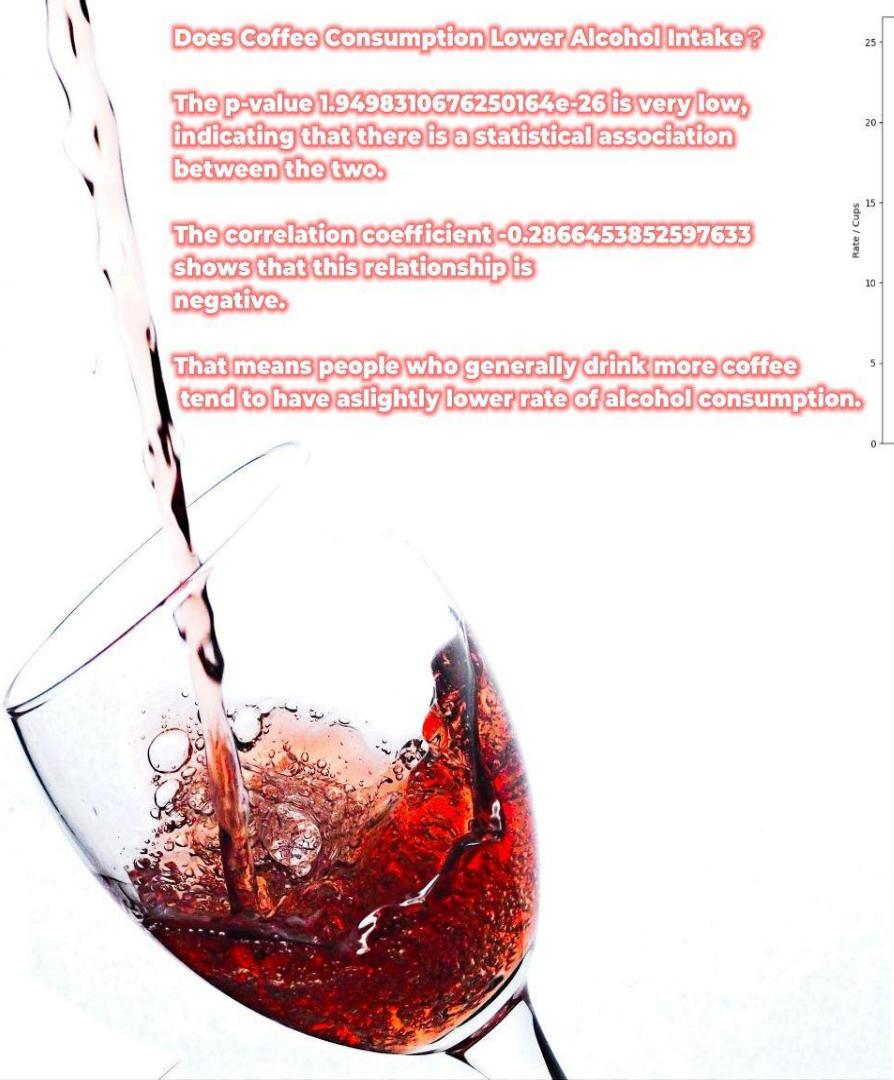


Can Coffee reduce the risk of mental illness?

Based on the p-value (1.0392807291702183e-06), which is very low, there appears to be some kind of association between the rate of mental illness and the amount of coffee consumed.

The correlation coefficient is -0.19392172921165388 shows that this relationship is negative. That means people who generally drink more coffee tend to have a slightly lower rate of mental illness. but it is very weak.



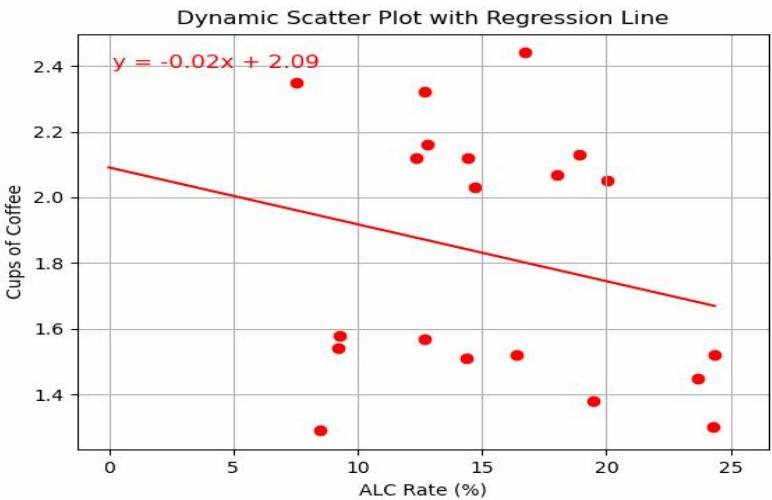
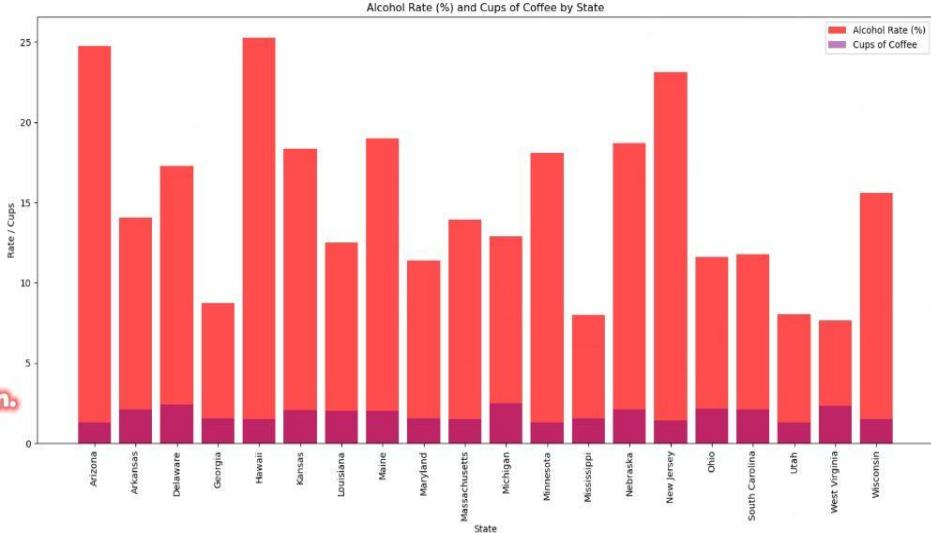


Does Coffee Consumption Lower Alcohol Intake?

The p-value $1.9498310676250164e-26$ is very low, indicating that there is a statistical association between the two.

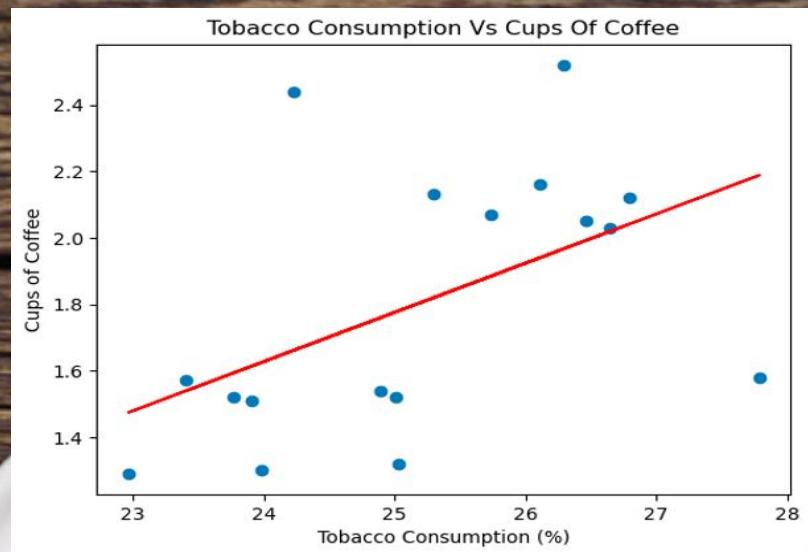
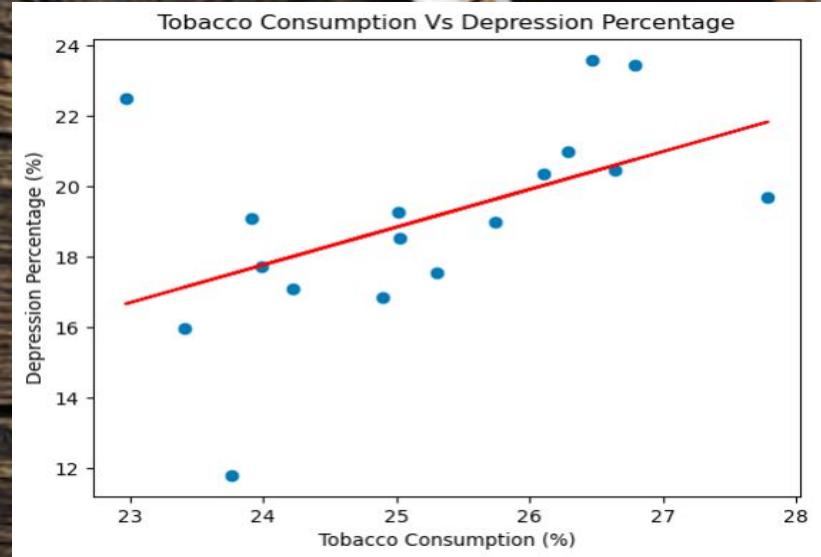
The correlation coefficient -0.2866453852597633 shows that this relationship is negative.

That means people who generally drink more coffee tend to have a slightly lower rate of alcohol consumption.



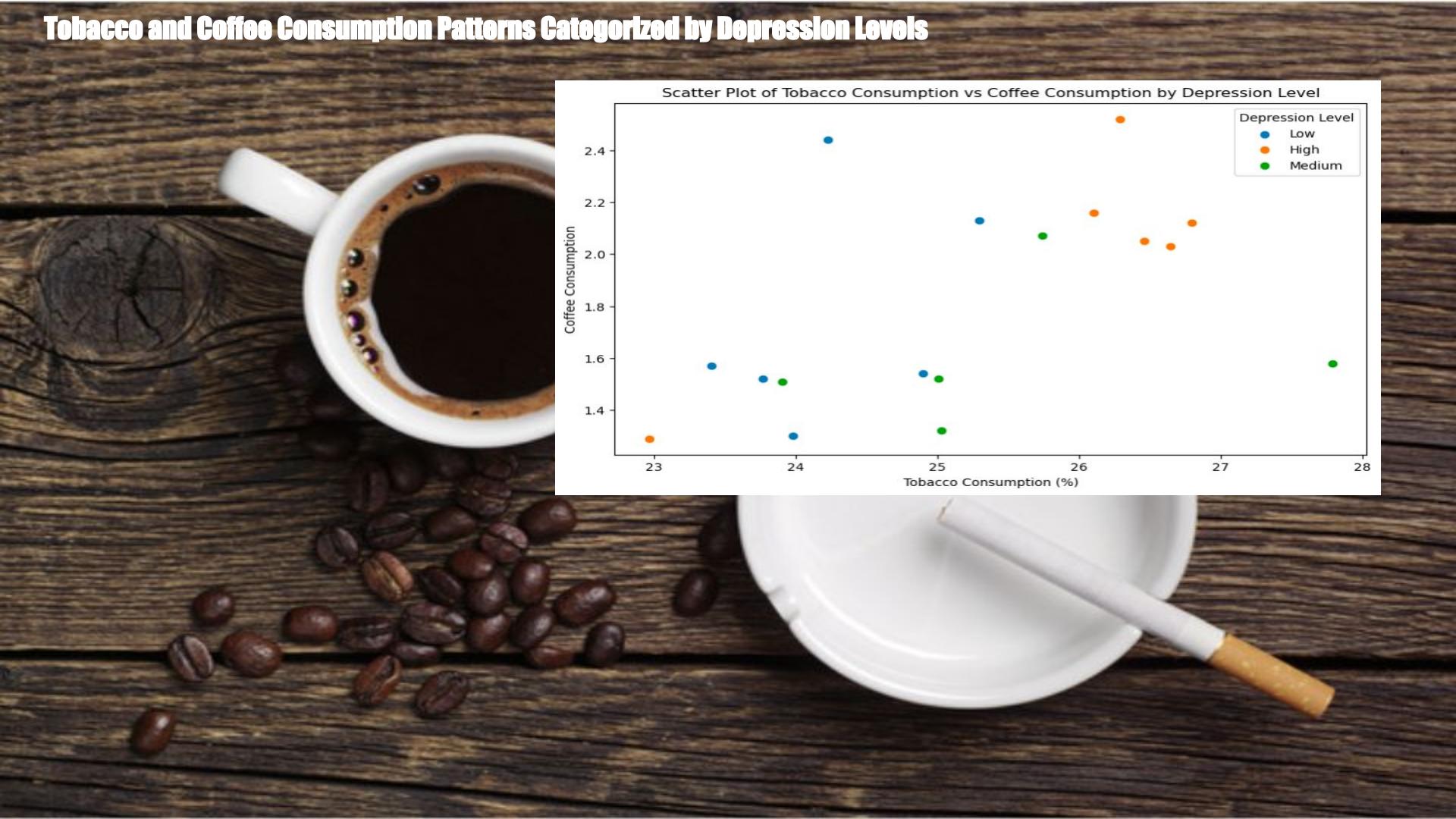
Tobacco Consumption, Coffee Intake, and Depression Rates

The correlation coefficient between Tobacco consumption and depression percentage is 0.50, indicating a moderate positive correlation.



The coefficient is 0.50, which indicates a moderate positive correlation. This means that, generally, as tobacco consumption increases, coffee consumption also tends to increase in the states observed.

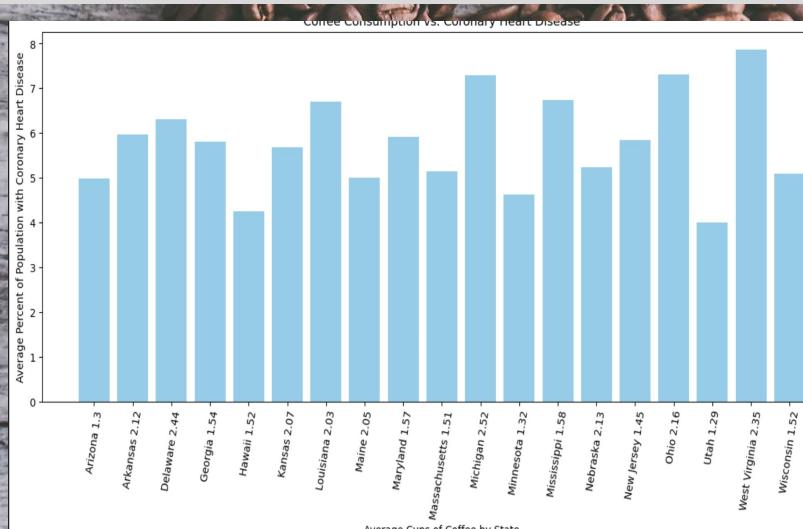
Tobacco and Coffee Consumption Patterns Categorized by Depression Levels



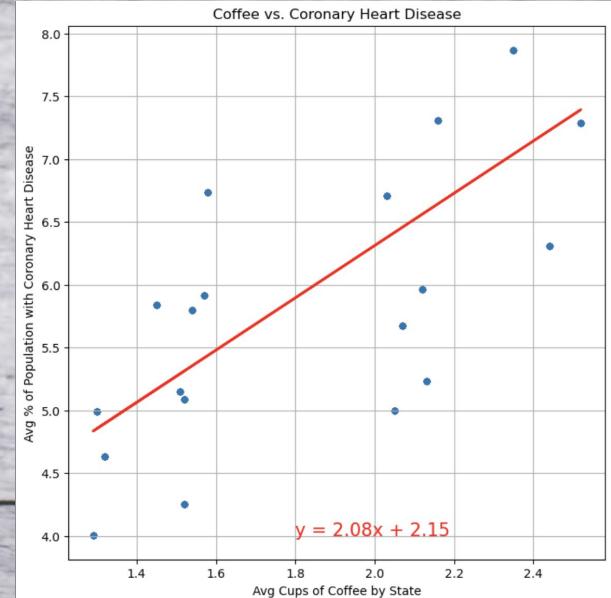
Does coffee consumption increase the risk of coronary heart disease?

The Pearson correlation coefficient of 0.8426, suggesting a strong positive linear relationship between coffee consumption and coronary heart disease.

A p-value of 0 indicates that you can reject the null hypothesis because there is a significant correlation between the variables. In other words, the positive correlation probably did not occur by chance. We could say that the more coffee you drink, the higher the possibility of having coronary heart disease.



```
# Calculate the Pearson correlation coefficient and the p-value in order to identify:  
# the strength (closer to -1 or 1 indicates stronger correlation) and the direction (positive or negative correlation) of the relationship.  
# import pearsonr  
from scipy.stats import pearsonr  
  
corr_coefficient, p_value = pearsonr(coffee_consumption_hd_df["Percent of Population (%)"], coffee_consumption_hd_df["Cups of Coffee"])  
  
#correlation_coef, p_value = avg_coffee.corr(avg_percent)  
  
print("Pearson Correlation Coefficient: (corr_coefficient)")  
print("P-value: (p_value)")  
[] ✓ 0s  
Pearson Correlation Coefficient: 0.8425843692904935  
P-value: 0.0
```





Reject the null hypothesis



According to our calculations, there is in fact a positive statistical significance between the amount of coffee consumed and overall health problems such as insomnia, tobacco addiction, heart disease, and mental health issues. One exception seems to be alcoholism, which cannot be proven to be related to the amount of coffee consumed.

Although these calculations show a positive statistical significance between coffee consumption and overall health, there are still many other factors that contribute to the overall health levels of the observed populations that were not considered.

Data Sources

Main Api or Datasets:

<https://data.cdc.gov>

<https://simpletexting.com/blog/most-coffee-obsessed-states-survey/>

Specific data sources:

Mental health:

<https://data.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-Mental-Health>

Alcohol use:

<https://data.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-Alcohol>

Depression Rate:

<https://usafacts.org/data/topics/people-society/health/health-risk-factors/depression/>

Insomnia: <https://www.kaggle.com/datasets/aemreusta/brfss-2020-survey-data>

Coronary Heart Disease:

<https://data.cdc.gov/500-Cities-Places/500-Cities-Coronary-heart-disease-among-adults-age/cqcq-r6f8>

Tobacco Use:

<https://data.cdc.gov/Survey-Data/Behavioral-Risk-Factor-Data-Tobacco-Use-2011-to-pr-wsas-xwh5>

