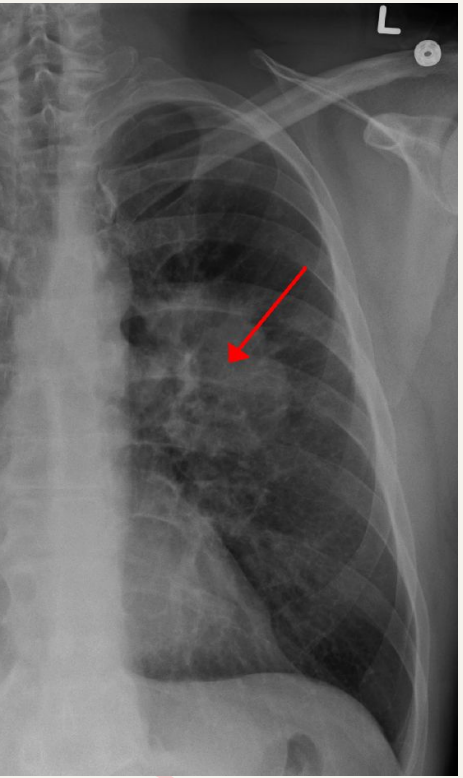
STA 199
Project green apple

Leo Lee, Jacob Han, Young Chun, Andrew Park

Topic/Research Question



Cancer is the second leading cause of death in the US, lung cancer being the leading cancer cause.

- Smoking is a primary risk factor of lung cancer
- 226,000+ cases per year

Lung cancer is by far the leading cause of cancer death in the US, accounting for about 1 in 5 of all cancer deaths. Each year, more people die of lung cancer than of colon, breast, and prostate cancers combined.

American Cancer Society

We wanted to explore whether alcohol and smoking truly have a correlation to lung cancer rates.

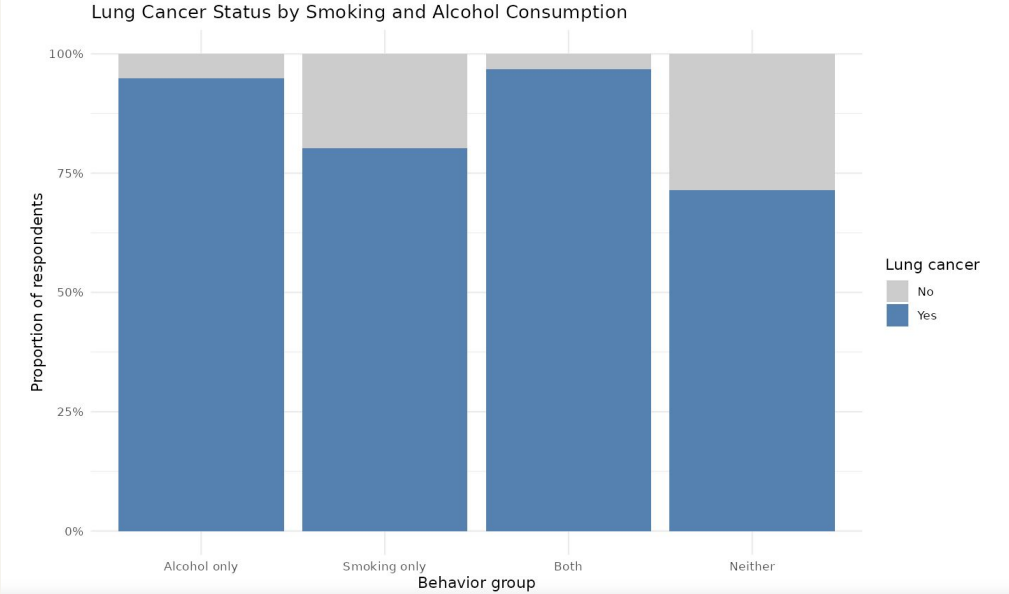
Our Dataset

- **Dataset contains information about patients (gender & age) and whether they smoke or drink alcohol.**
  - Known side effects of smoking
  - Any indicators of their surroundings that could affect their decision making.
  - Most importantly, whether they have lung cancer.

- Published on LA Open Data by the LA County Data of Public Health.

- The data was published in January 4th, 2022.

Data and Graphs



statistic	p.value	parameter	method
<dbl>	<dbl>	<int>	<chr>
28.2	0.00000331	3	Pearson's Chi-squared test

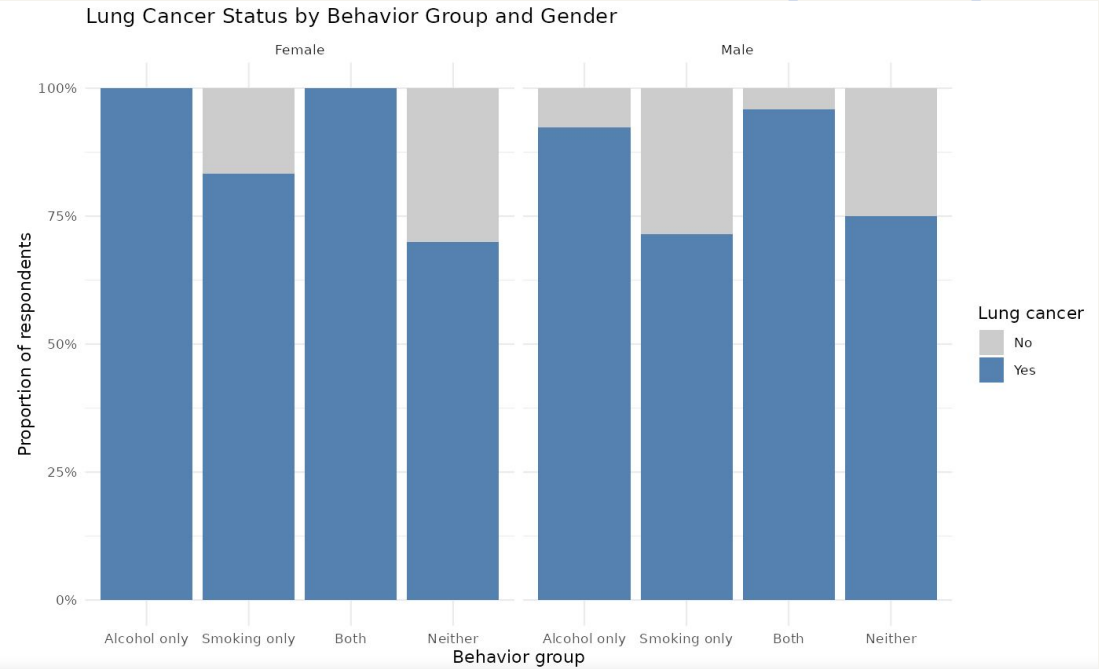
Correlation between alcohol consumption & smoking and having lung cancer

- **Alcohol only > neither** by a significant margin
- **Smoking only > neither** by less of a margin.
- **Both > neither** by a significant margin

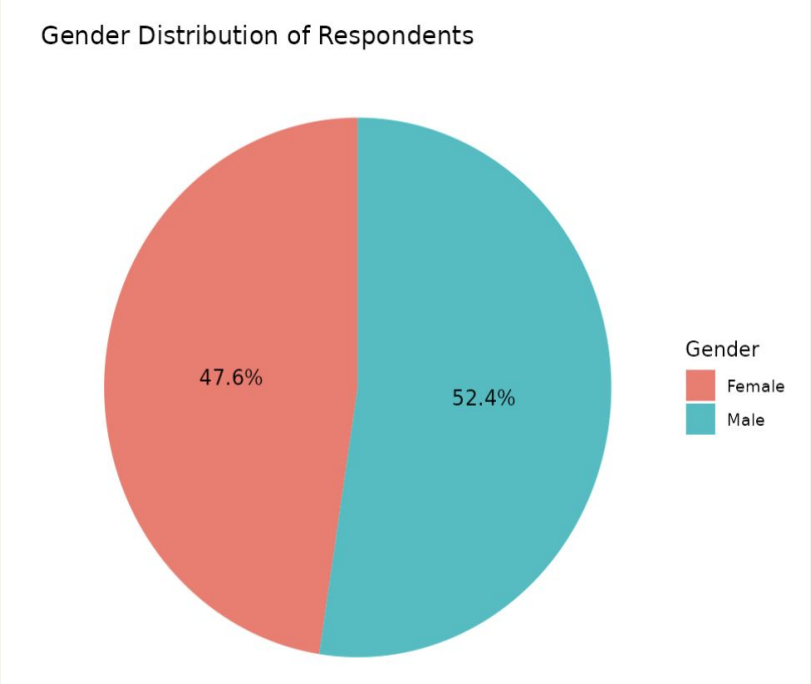
Our p.value from our Chi-squared test shows that there is **a statistically significant correlation** between behavior group and lung cancer status ($p < .001$).

Data and Graphs

- Female Data
  - Both Alcohol only and Alcohol and Smoking have a 100% chance for lung cancer
  - Smoking only has around 80% for lung cancer
  - Neither shows around 70% with lung cancer
- Male Data
  - Alcohol only shows 90% of having lung cancer
  - Smoking only shows less than 75% of getting lung cancer - conflicting.
  - Both has around 95% chance of lung cancer, and neither shows around a 75% chance.

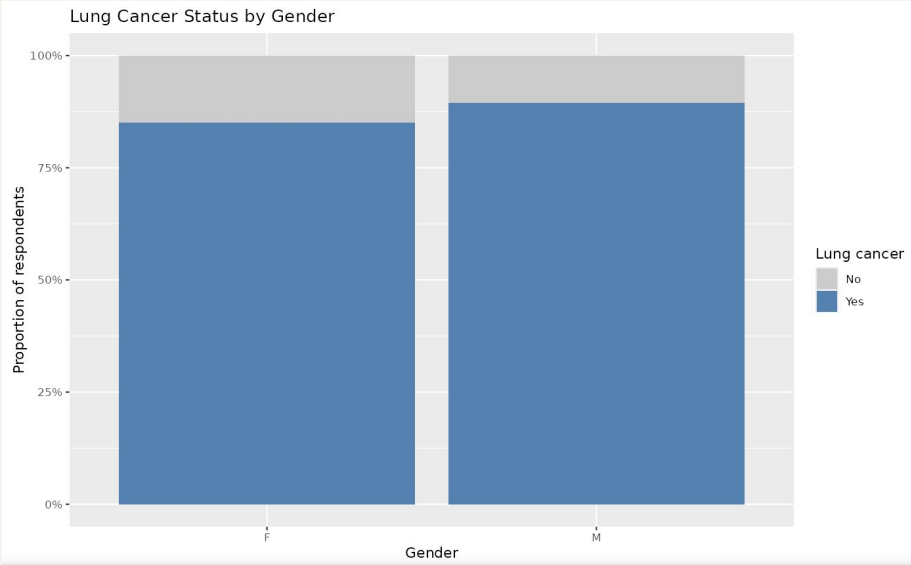


Data and Graphs



The gender distribution was relatively balanced, with male respondents accounting for 52.4% of the sample and female respondents accounting for 47.6%.

Men were more likely have lung cancer, providing evidence against that assumption that women being more likely to have lung cancer is interfering with our analysis.

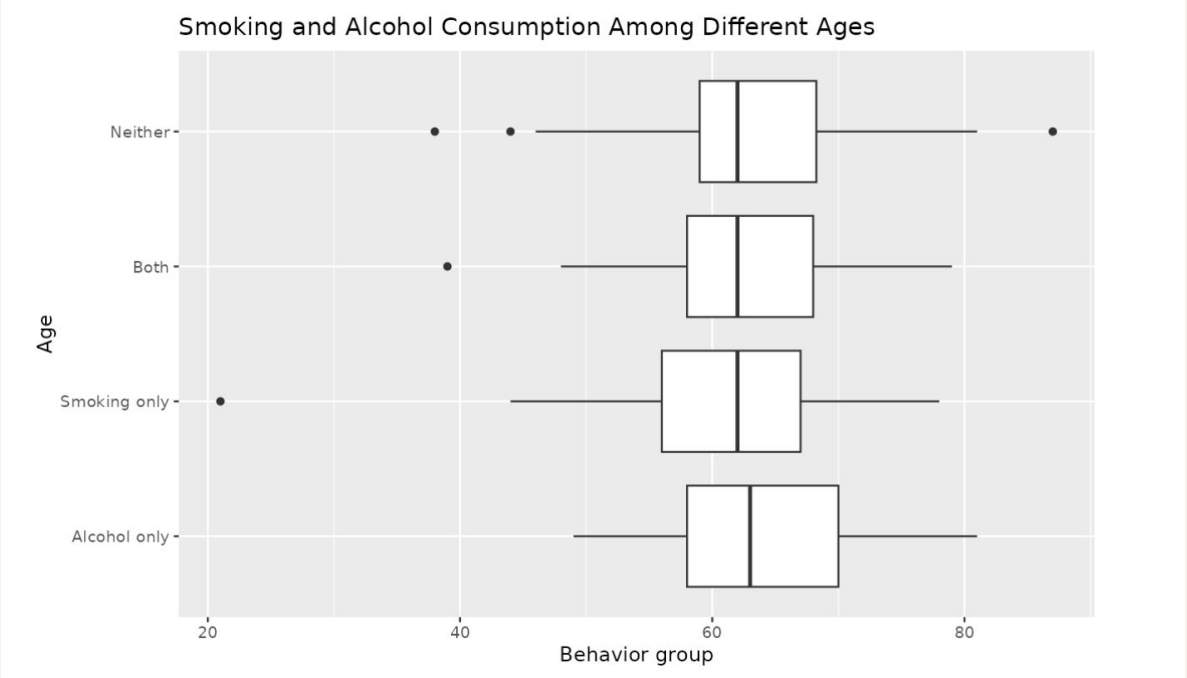


Data and Graphs

Could certain age groups smoke and consume alcohol differently?

- Perhaps **older** people were more likely to smoke and drink
- **Confounding variable**

All behavior groups relatively **equal** ages



Conclusion

Through our analysis, we provided evidence that both alcohol consumption and smoking were correlated with lung cancer, with alcohol consumption demonstrating a higher correlation, through our stacked bar graph and chi squared analysis. We were able to provide evidence against the influence of confounding variables, such as gender and age, as well.

However, our dataset is relatively small, with only 309 individuals recorded. Our dataset also demonstrated bias sampling, as the data was only based on patients, and a significantly higher percentage of individuals have cancer as compared to the actual population. To that extent, our results may be misleading, as 100% of women who drink alcohol do not have lung cancer.

While our conclusion of alcohol and smoking having a correlation with lung cancer is generally well-accepted, we cannot advocate for it due to the flaws of our dataset, and the idea that alcohol has a greater correlation than smoking is widely believed to be false.