Visualization Project

2023-09-30

Lung Cancer

Kironmoy Roy

Roll No: MDS202332

DG1107: Visualization

Email: kironmoy.mds2023@cmi.ac.in

Chennai Mathematical Institute

Table of Contents

0.1 INTRODUCTION

- A cancer that begins in the lungs and most often occurs in people who smoke.
- **Types of Lung Cancer** : Two major types of lung cancer are non-small cell lung cancer and small cell lung cancer.
- **Symptoms** : A cough that doesn't go away or gets worse over time. Trouble breathing or shortness of breath (dyspnea). Chest pain or discomfort. Wheezing. Coughing up blood (hemoptysis). Hoarseness. Loss of appetite. Unexplained weight loss
- **Main Causes of Lung Cancer** : Smoking causes the majority of lung cancers — both in smokers and in people exposed to secondhand smoke. But lung cancer also occurs in people who never smoked and in those who never had prolonged exposure to secondhand smoke.
- **Death Count** : GLOBOCAN 2020 estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer (IARC) show as lung cancer remains the leading cause of cancer death, with an estimated 1.8 million deaths (18%) in 2020.

0.2 DATASET DESCRIPTION

Click to find the source for the data set. The present work relied on a public data set. The number of participants is 276, and total 16 attributes are there in the data set.

0.2.1 Detailed description for the columns of the data set:

- **Gender** : This feature shows if the person's sex is male or female.
- **Age** (years) : This feature captures the person's age.
- **Smoking** : This feature indicates if the participant is a smoker or not.
- **Yellow fingers** : This feature refers to whether the participant has yellow fingers or not.
- **Anxiety** : This feature shows if the participant is anxious or not.
- **Peer pressure** : This feature captures if the participant feels peer pressure or not.
- **Chronic disease** : This feature expresses if the participant suffers from a chronic disease or not.
- **Fatigue** : This feature manifests if the participant suffers from fatigue or not.
- **Allergy** : This feature refers to whether the participant has an allergy or not.
- **Wheezing** : This feature declares if the participant suffers from wheezing or not.
- **Alcohol** : This feature shows if the participant consumes alcohol or not.
- **Coughing** : This feature refers to whether the participant suffers from coughing or not.
- **Shortness of breath** : This feature refers to whether the participant has shortness of breath or not.
- **Swallowing difficulty** : This feature indicates if the participant has difficulty swallowing or not.
- **Chest pain** : This feature captures whether the participant has chest pain or not.
- **Lung Cancer**: This feature shows if the participant has been diagnosed with lung cancer or not.

All the features are nominal except for age, which is numerical.

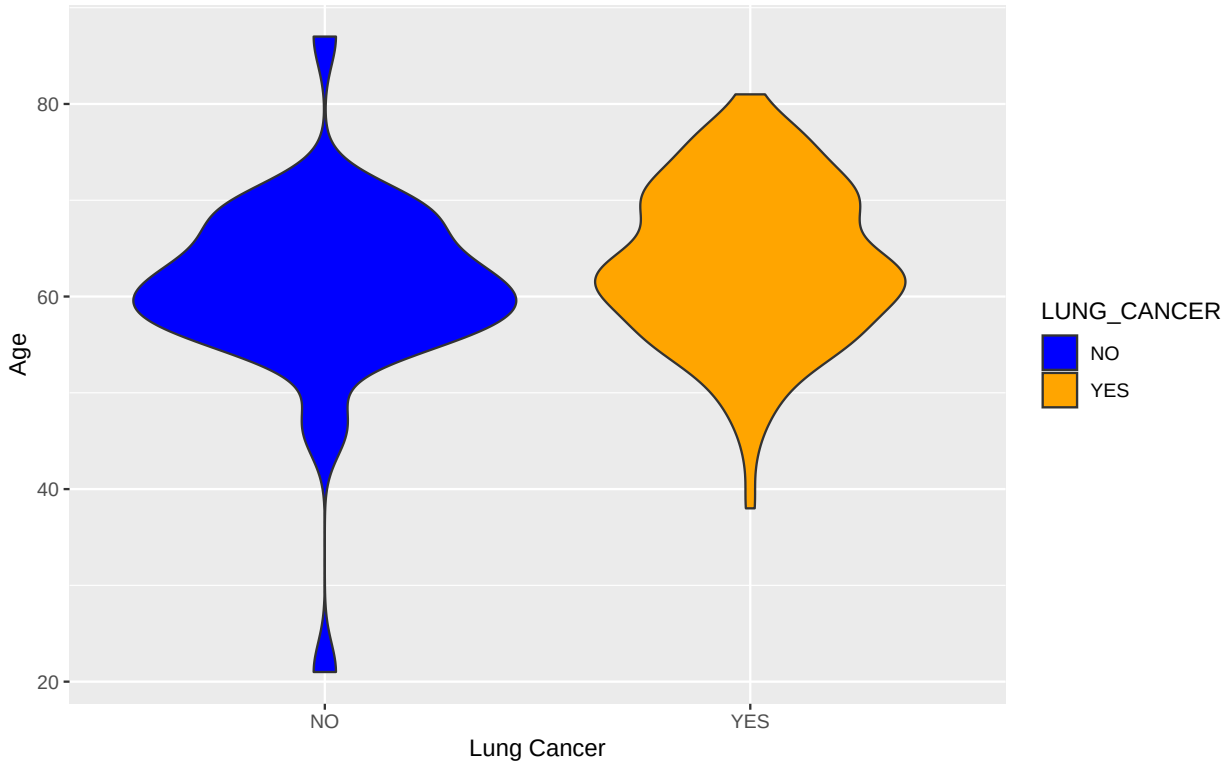
0.2.2 objectives

Using the above data set, we are trying to determine the following:

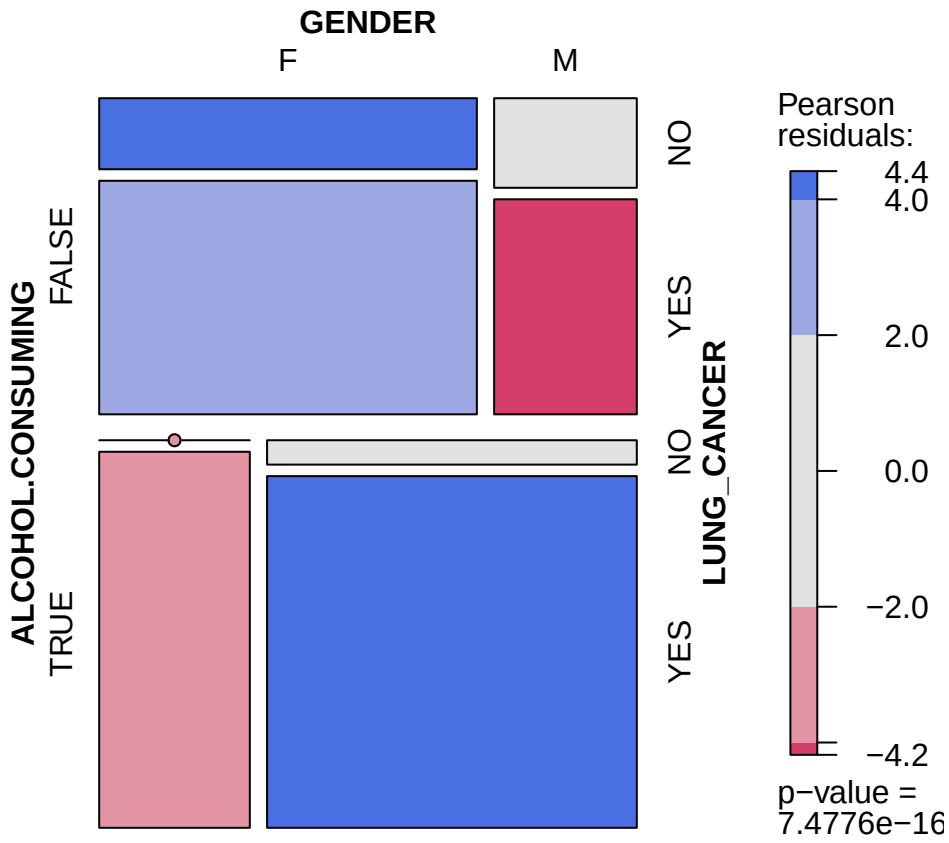
- Which Age group has the highest chance of being affected by Lung Cancer.
- which gender has the highest chance of being affected by Lung Cancer.
- Which are the main symptoms of Lung Cancer.
- Relationship between Smoking and Lung Cancer.
- Relation between smoking and chest pain and having Lung Cancer.
- Relation between Alcohol Consuming,Smoking and Peer pressure.

The above objectives can be achieved using histograms, bar plots, piechart and box plots for univariate analysis, scatter plots and line plots for bivariate analysis, and gradient scatter plots for multivariate analysis.

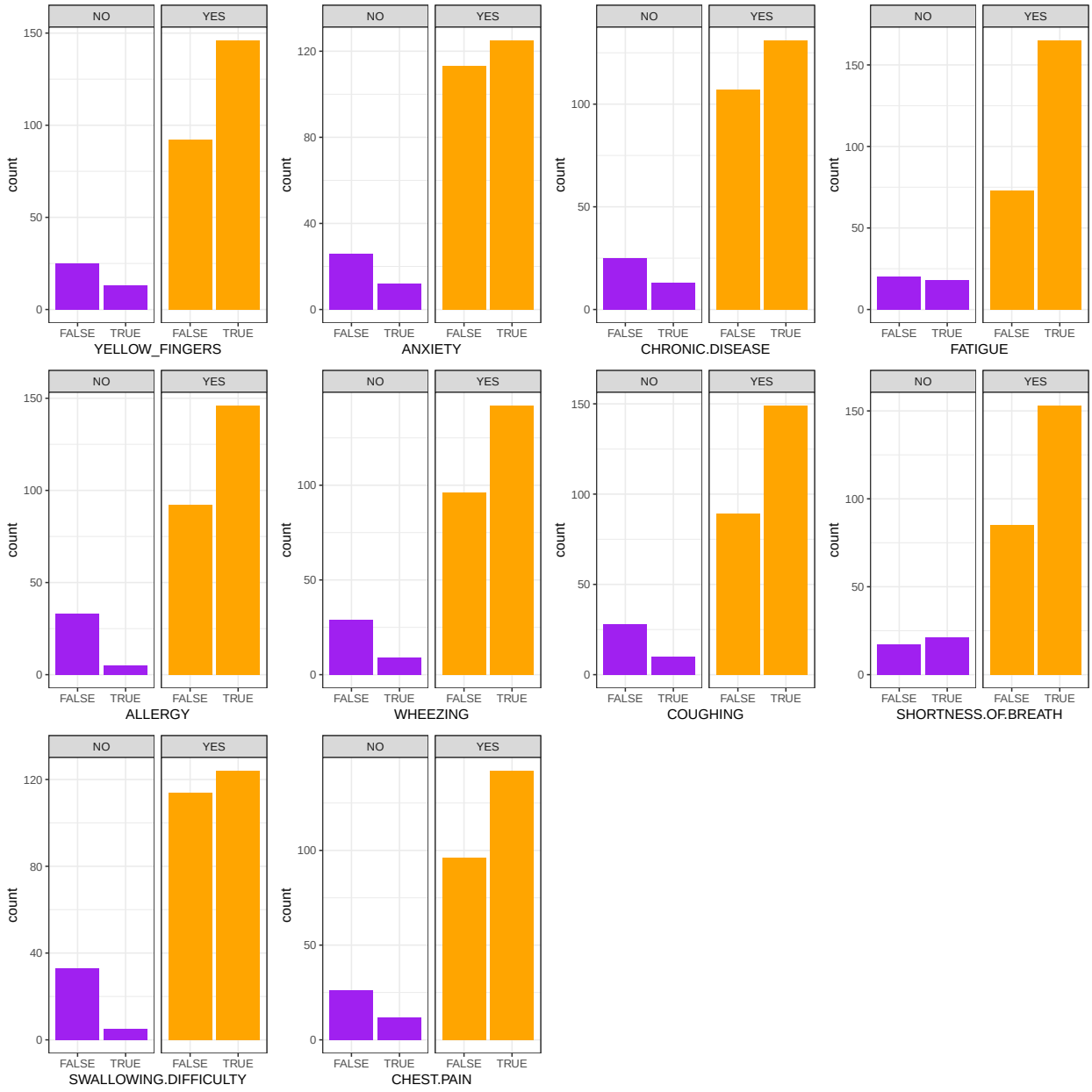
0.3 EXPLORATORY DATA ANALYSIS



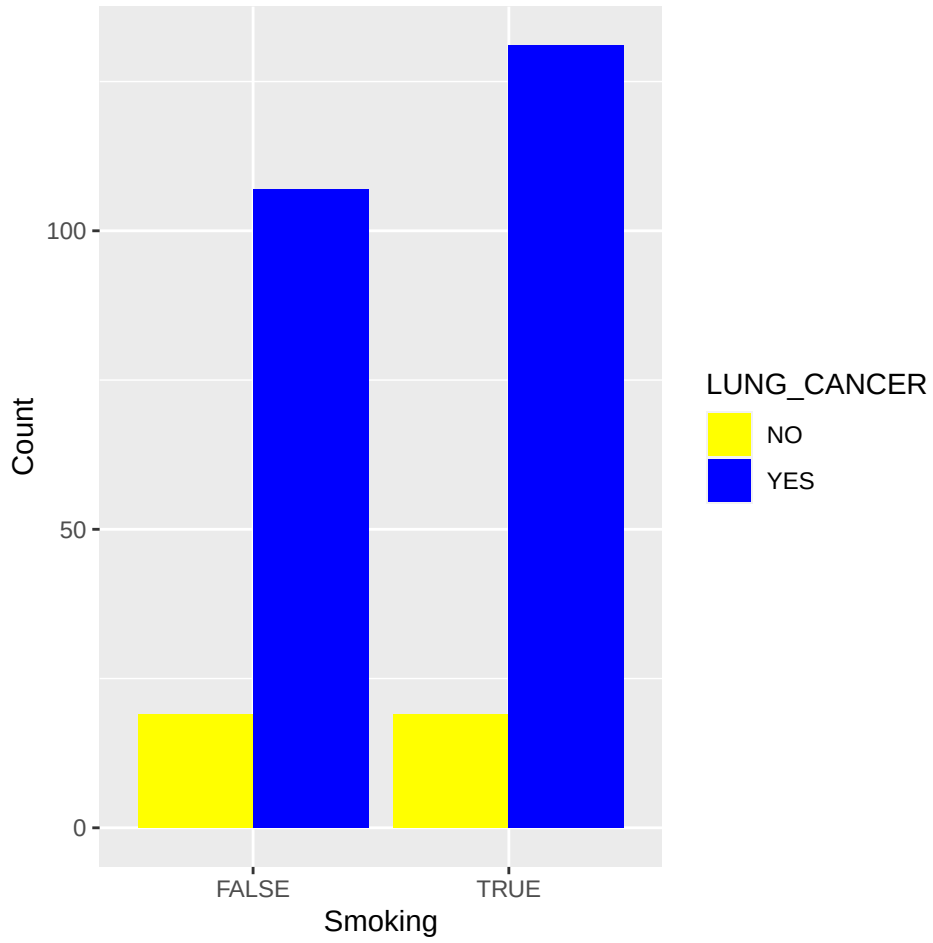
0.3.0.1 ***FIG 1: Violin Plot Representation:*** FIG1 shows the participants' distribution of having Lung Cancer with Age. We observe that lung cancer mostly concerns people between 55 and 79 years old, where the age group 60–64 is the one with the highest frequency.



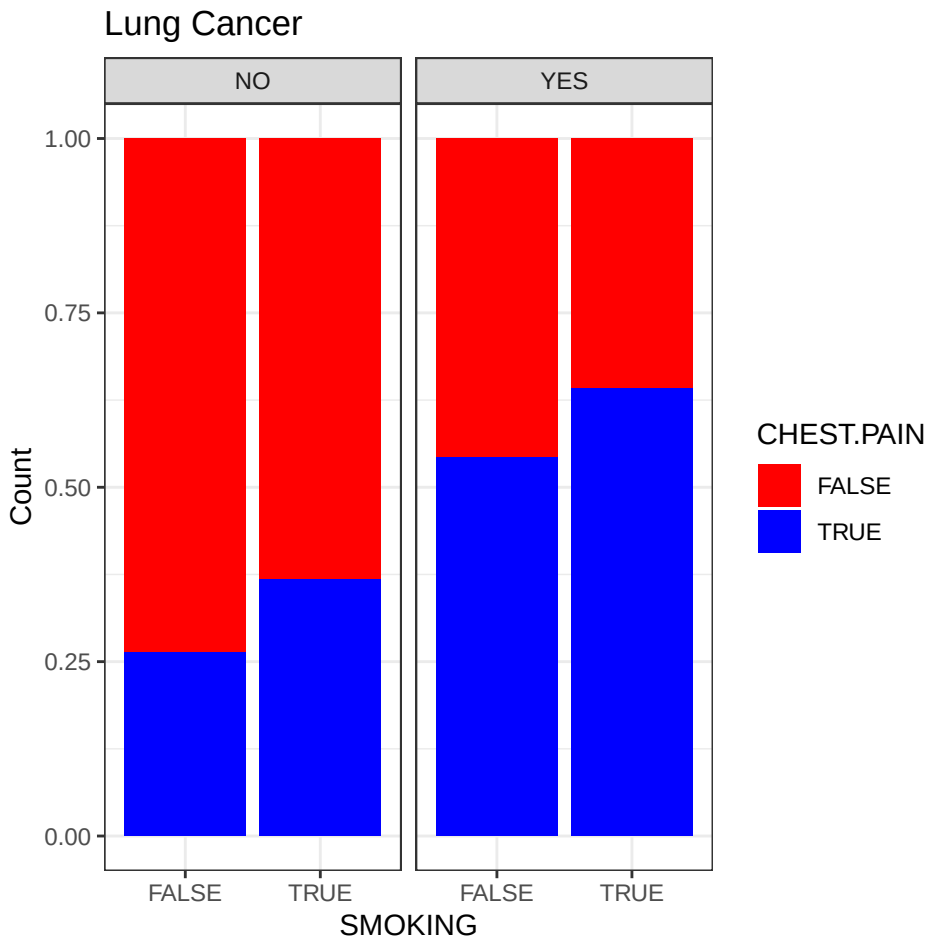
0.3.0.2 ***FIG 2: Mosaic Plot Representation:*** In the above diagram we have plotted Gender Vs Alcohol Consuming With Lung Cancer. We can see that among all drinkers, men have more chances to have Lung Cancer.



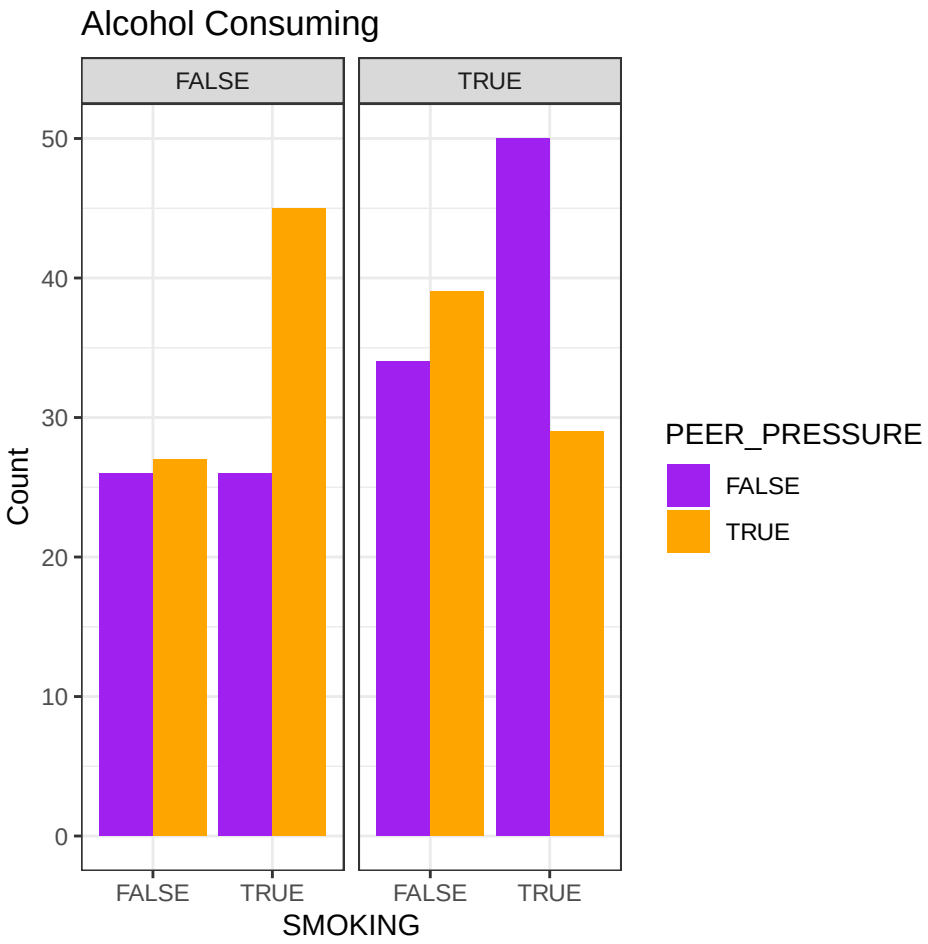
0.3.0.3 ***FIG 3: Stacked Bar Plot Representation:*** In the above Fig there are 10 plots, each of them has a symptoms of Lung Cancer vs Lung Cancer.We can conclude from the above plots that Fatigue,Yellow Fingers,Allergy,Shortness of Breathe are the main symptoms of Lung Cancer.



0.3.0.4 ***FIG 4: Bar Plot Representation:*** The above diagram is a bar plot between Smoking and Lung Cancer. As we can see that Lung Cancer is more likely to be in People who smokes.



0.3.0.5 ***FIG 5: Stacked Bar Plot Representation:*** The above figure is between Smoking and Chest Pain with Lung Cancer.We can see that a smoker has more chances to have chest pain.And if a smoker is having chest pain then he is more likely to be affected by Lung Cancer.



0.3.0.6 ***FIG 6: Grouped Bar Plot Representation:*** The above figure is between Smoking and Alcohol Consuming with Peer Pressure.We can see that if a person is a smoker then he has more chances of having peer pressure.

0.4 RESULTS

- We observe that lung cancer mostly concerns people between 55 and 79 years old, where the age group 60–64 is the one with the highest frequency.
- We can see that among all drinkers, men have more chances to have Lung Cancer.
- We can conclude from the above plots that Fatigue,Yellow Fingers,Allergy,Shortness of Breathe are the main symptoms of Lung Cancer.
- we can see that Lung Cancer is more likely to be in People who smokes.
- We can see that a smoker has more chances to have chest pain.And if a smoker is having chest pain then he is more likely to be affected by Lung Cancer.
- We can see that if a person is a smoker then he has more chances of having this bad habit by peer pressure.

0.5 CONCLUSION

From the above results we can conclude that if someone has Lung Cancer then it's not necessary that he or she is a smoker but if someone has a smoking habit then the chance of being affected by Lung Cancer is more than a non smoking person. we can also conclude that that if a person is a smoker then he has more chances of having this bad habit by peer pressure.We can see that a smoker has more chances to have chest pain.And if a smoker is having chest pain then he is more likely to be affected by Lung Cancer.

We observe that Lung Cancer is more common to those people who are in 55 to 79 years in age.

The main symptoms that a Lung Cancer affected person can have are Fatigue,Yellow Fingers,Allergy,Shortness of Breathe.

We can see that among all drinkers, men have more chances to have Lung Cancer.