Visualisation Project_part 1

2023-10-01

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Introduction

Lung cancer is a fatal disease that begins in the cells of the lungs. Lung cancer often identified in a later stage which makes it incurable then. But an analysis of possible causes and observable symptoms makes people less vulnerable to the risks of lung cancer. In this project I am visualising the data of various possible reasons and early symptoms of lung cancer with its relationship with the occurence of lung cancer.

Data Description

The original data had 16 columns describing symptoms and possible causes and 309 rows. But it had some duplicates among rows. So by using

```
Data<-distinct(Data)
```

we get 276 rows.

```
Data[Data==1]<-0
Data[Data==2]<-1
Data$LUNG_CANCER <- gsub("YES", 1, Data$LUNG_CANCER)
Data$LUNG_CANCER <- gsub("NO", 0, Data$LUNG_CANCER)
Data$LUNG_CANCER <- as.numeric(as.character(Data$LUNG_CANCER))

glimpse(Data)
```

```
## Rows: 276
## Columns: 16
## $ GENDER                <chr> "M", "M", "F", "M", "F", "F", "M", "F", "F", "M"~
## $ AGE                   <int> 69, 74, 59, 63, 63, 75, 52, 51, 68, 53, 61, 72, ~
## $ SMOKING               <dbl> 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, ~
## $ YELLOW_FINGERS        <dbl> 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, ~
## $ ANXIETY               <dbl> 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, ~
## $ PEER_PRESSURE         <dbl> 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, ~
```

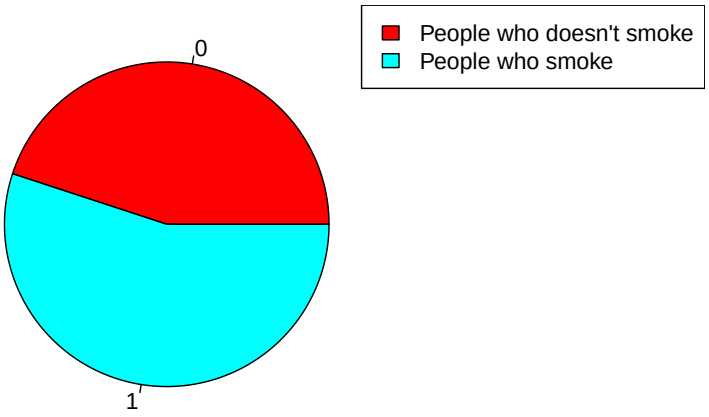
```
## $ CHRONIC.DISEASE       <dbl> 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, ~
## $ FATIGUE               <dbl> 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, ~
## $ ALLERGY               <dbl> 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, ~
## $ WHEEZING              <dbl> 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, ~
## $ ALCOHOL.CONSUMING     <dbl> 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, ~
## $ COUGHING              <dbl> 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, ~
## $ SHORTNESS.OF.BREATH   <dbl> 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, ~
## $ SWALLOWING.DIFFICULTY <dbl> 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, ~
## $ CHEST.PAIN            <dbl> 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, ~
## $ LUNG_CANCER           <dbl> 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, ~
```

Exploratory Data Analysis

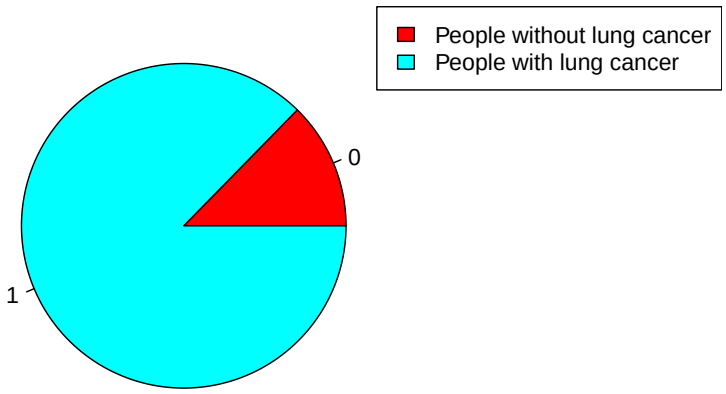
Interpretations on Possible Reasons for Lung cancer

a)Smoking

smokers and non-smokers among people with lung cancer

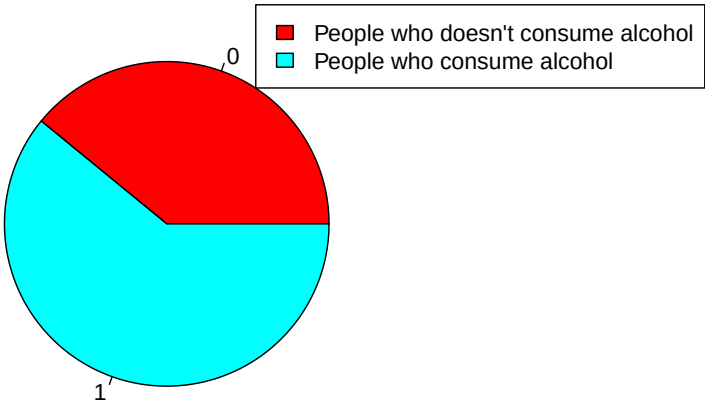


People with lung cancer among smokers

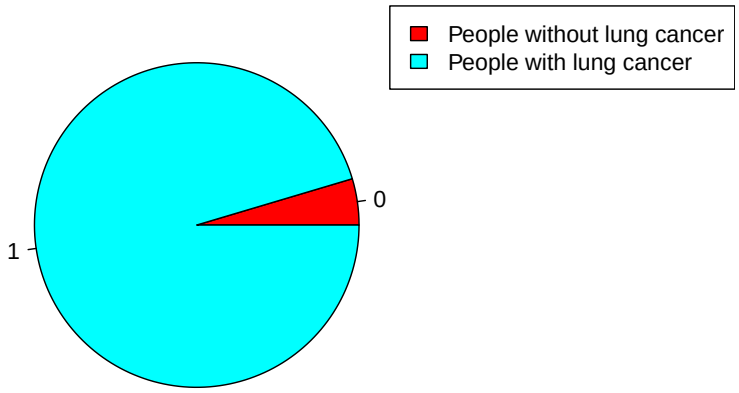


b)Alcohol consumption

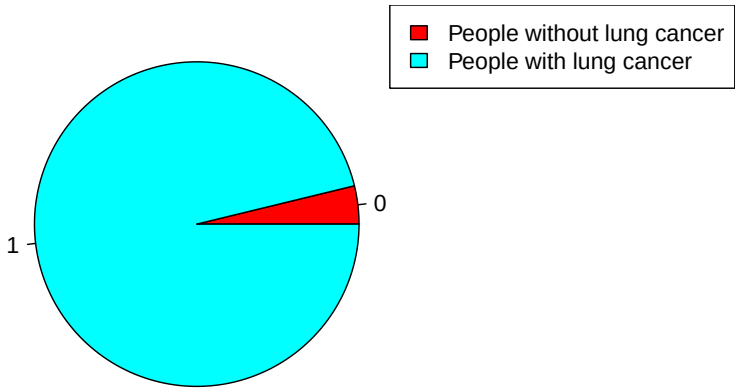
Alcoholic and non alcoholic among people with lung cancer



People with and without lung cancer among alcoholics

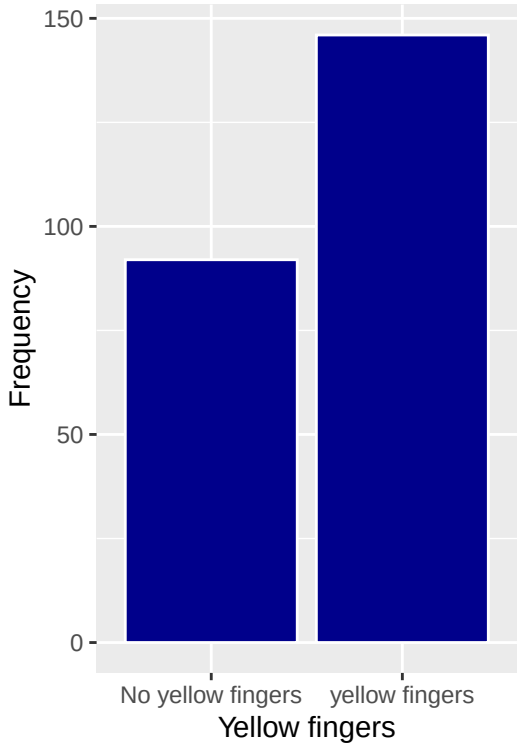


People with and without lung cancer among people who drink and smoke

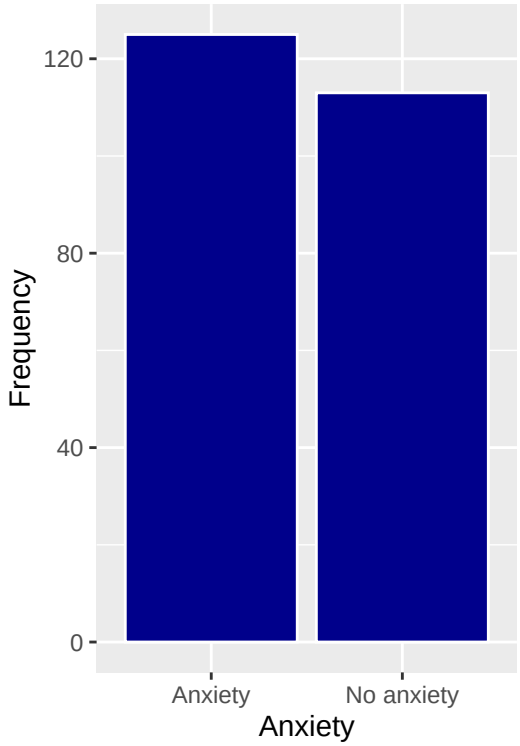


Various symptoms and its relationship with lung cancer

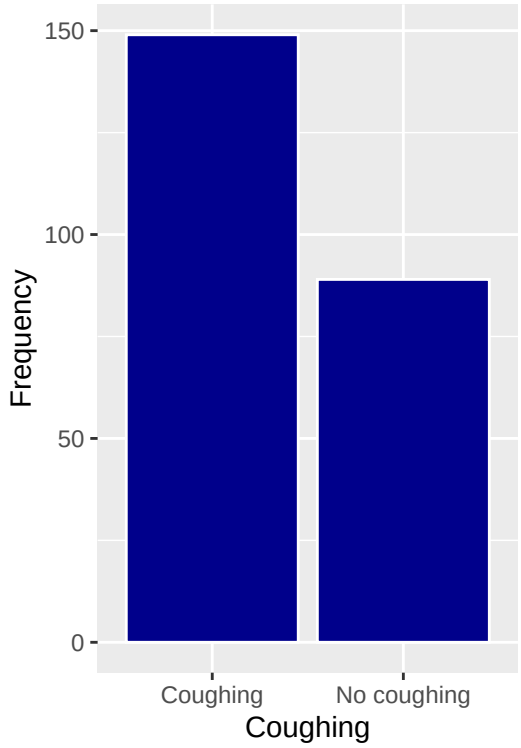
a)Presence of yellow fingers among people with lung cancer



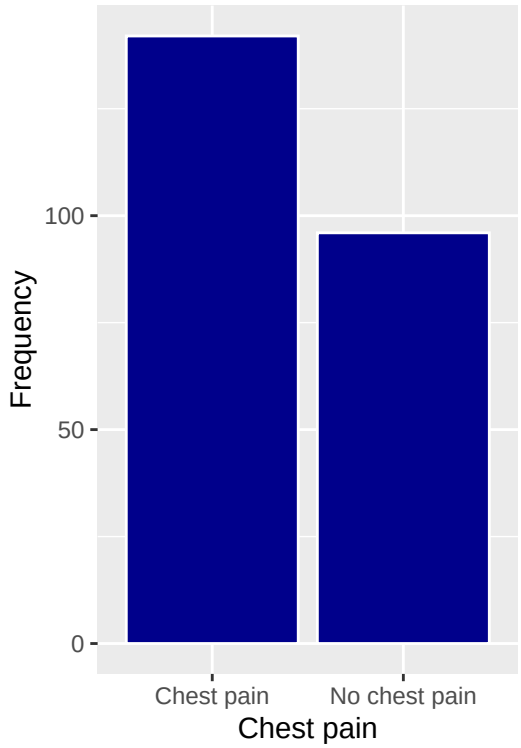
b)Presence of Anxiety among people with lung cancer



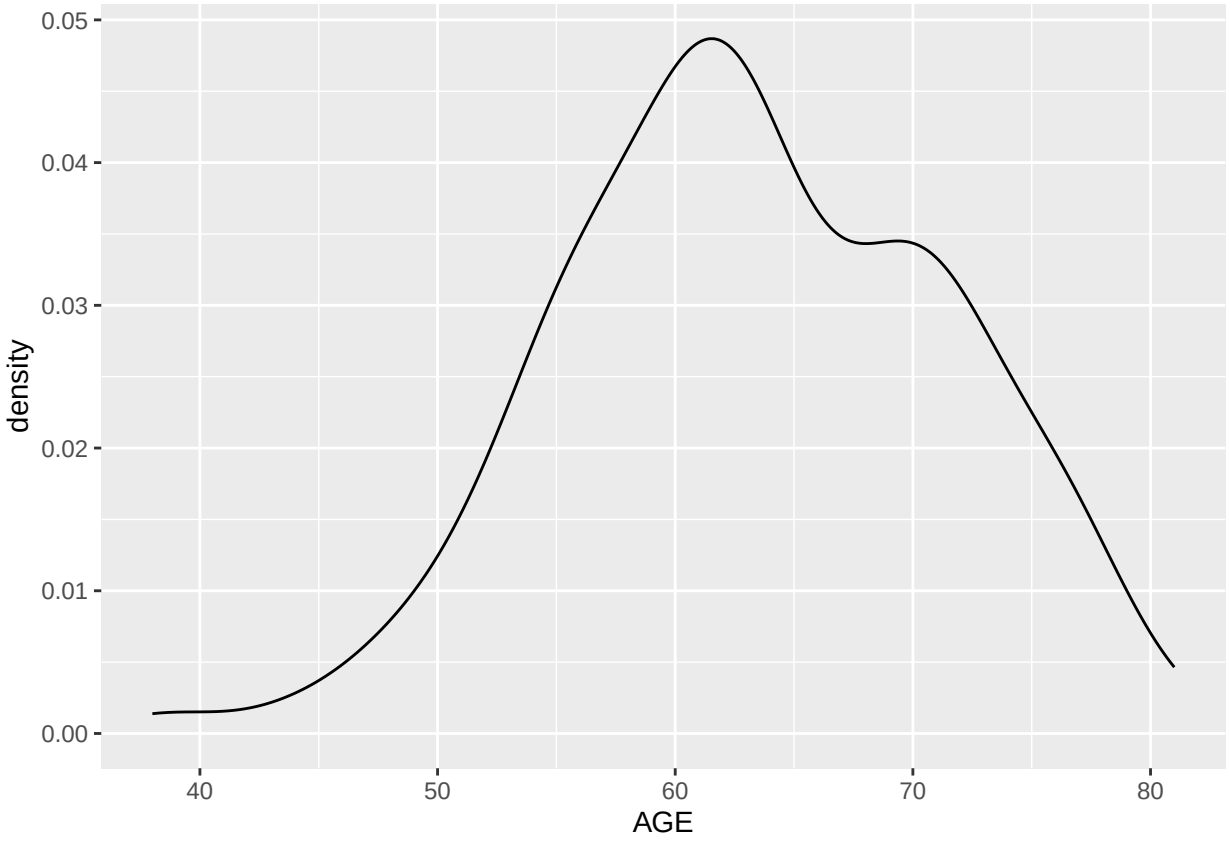
c)Presence of coughing among people with lung cancer



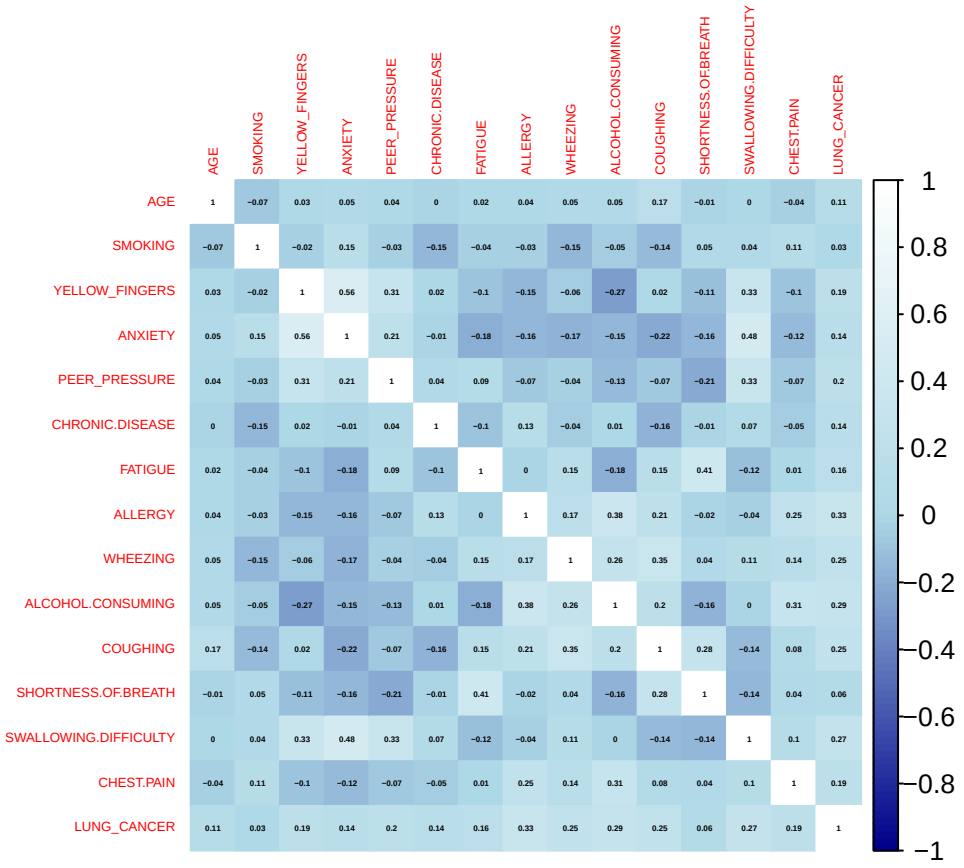
c)Presence of chest painamong people with lung cancer



###Distribution of lung cancer among various age groups




```
## corrplot 0.92 loaded
```



Results

*From the analysis of data it is evident that smoking and alcohol consumption has impact on the occurence of lung cancer. Alcohol consumption has a high correlation 0.29 with lung cancer.

*Among the people who both consume alcohol and smoke, the frequency of lung cancer is very high

*Presence of allergy, wheezing, coughing, swallowing difficulty has high correlation with lung cancer. Hence those can be used for the prediction of lung cancer.

*From the data the most vulnerable population to lung cancer is the age group around 60.

Conclusion

From the data and its analysis we could conclude that it is definitely possible to minimise the occurence of lung cancer by limiting the consumption of alcohol and tobacco. Also we can conclude that, upto some extent it is possible to predict lung cancer with early symptoms.