EDAProjects Lung Data set- Raelene Ramirez

Raelene Ramirez

2024-09-30

Introductions

This data set look at the relationship between the effect made by smog in Los Angeles and symptoms of chronic respiratory disease. We will be using the data set to see if different cities within Los Angeles are impacted more by smog. Within each city we will also be looking at how age and gender to determine if any one of those variables plays a role to being more likely to have symptoms of chronic respiratory.

1. Univariate Exploration

AREA:

```
table(Lung$AREA, useNA = "always")
```

```
##
##    1    2    3    4 <NA>
##   24   49   19   58    0
```

Above is a summary table that shows the statistics of areas located in Los Angeles. (1) represents the city of Burbank, (2) is Lancaster, (3) is Long Beach, and (4) representing Glendora. Based on the data it is found that Glendora had the highest sample size and Long Beach with the lowest.

FAGE:

```
table(Lung$FAGE, useNA = "always")
```

```
##
##   26   28   29   30   31   32   33   34   35   36   37   38   39   40   41   42
##    2    3    1    5    3    5    9    7    7    8    8    6    8   12    4    8
##   43   44   45   46   47   48   49   50   51   52   53   54   59 <NA>
##    7    9    7    7    2    2    1    4    3    4    3    4    1    0
```
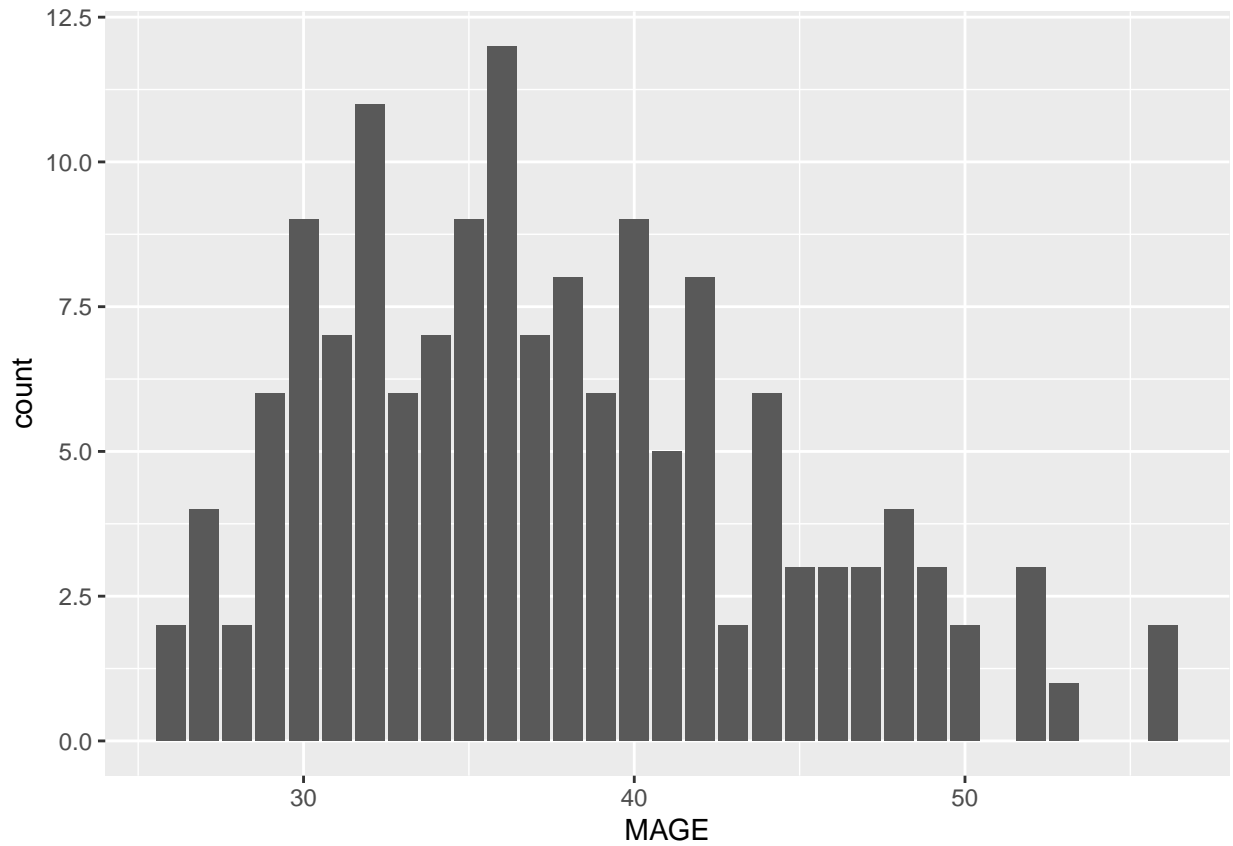
```
ggplot(Lung, aes(x=FAGE)) + geom_bar()
```
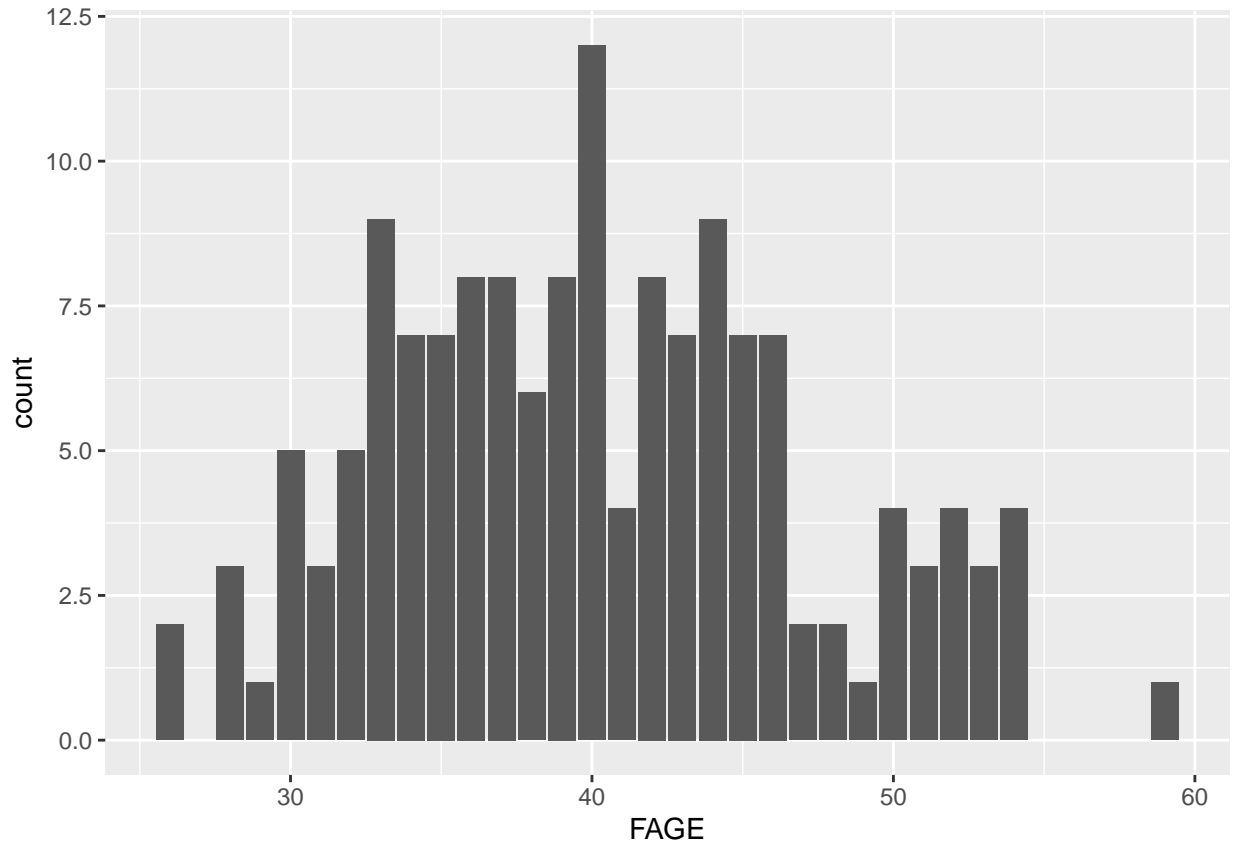
```
mean(Lung$FAGE)
```

```
## [1] 40.13333
```

MAGE:

```
table(Lung$MAGE, useNA = "always")
```

```
##
##   26   27   28   29   30   31   32   33   34   35   36   37   38   39   40   41
##    2    4    2    6    9    7   11    6    7    9   12    7    8    6    9    5
##   42   43   44   45   46   47   48   49   50   52   53   56 <NA>
##    8    2    6    3    3    3    4    3    2    3    1    2    0
```

```
ggplot(Lung, aes(x=MAGE)) + geom_bar()
```



```
mean(Lung$MAGE)
```

```
## [1] 37.56
```

Provided above is a graph to help distinguish the ages, both Male and Female of people sampled in the area of Los Angeles who have experienced symptoms of chronic respiratory disease. The mean found for both was Female: 40.13 and Male: 37.56.

Gender:

```
table(Lung$FSEX, Lung$MSEX)
```

```
##
##        2
##   1 150
```

Above describes there ia a sample of 150 people both, Male and Female that were accounted for in the whole data set. This means that there is an equal amount of chance that both Male and Female could be affected by symptoms of Chronic Respiratory.

Bivariate Comparison

```
table(Lung$AREA, Lung$MAGE)
```

```
##
##      26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
##   1   1  2  0  1  3  1  2  0  0  2  2  1  1  1  0  1  0  1  1  1  0  1  1  1  0
##   2   1  0  0  1  1  2  4  2  2  1  5  0  4  4  3  1  4  1  2  1  2  1  2  0  1
##   3   0  1  1  0  0  1  1  1  2  1  1  0  0  0  2  2  1  0  0  0  1  1  1  0  0
##   4   0  1  1  4  5  3  4  3  3  5  4  6  3  1  4  1  3  0  3  1  0  0  0  2  1
##
##      52 53 56
##   1   0  0  0
##   2   2  1  1
##   3   1  0  1
##   4   0  0  0
```

```
table(Lung$AREA, Lung$FAGE)
```

```
##
##      26 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
##   1   1  0  0  2  2  1  1  2  1  0  2  0  0  3  0  0  0  1  2  2  0  0  0  1  1
##   2   1  0  0  0  0  2  2  2  4  3  0  3  3  4  2  3  3  1  2  2  1  1  1  3  1
##   3   0  2  0  0  0  0  2  1  1  0  1  2  1  0  2  1  1  0  1  1  0  0  0  0  1
##   4   0  1  1  3  1  2  4  2  1  5  5  1  4  5  0  4  3  7  2  2  1  1  0  0  0
##
##      52 53 54 59
##   1   0  1  1  0
##   2   2  0  2  1
##   3   0  1  1  0
##   4   2  1  0  0
```
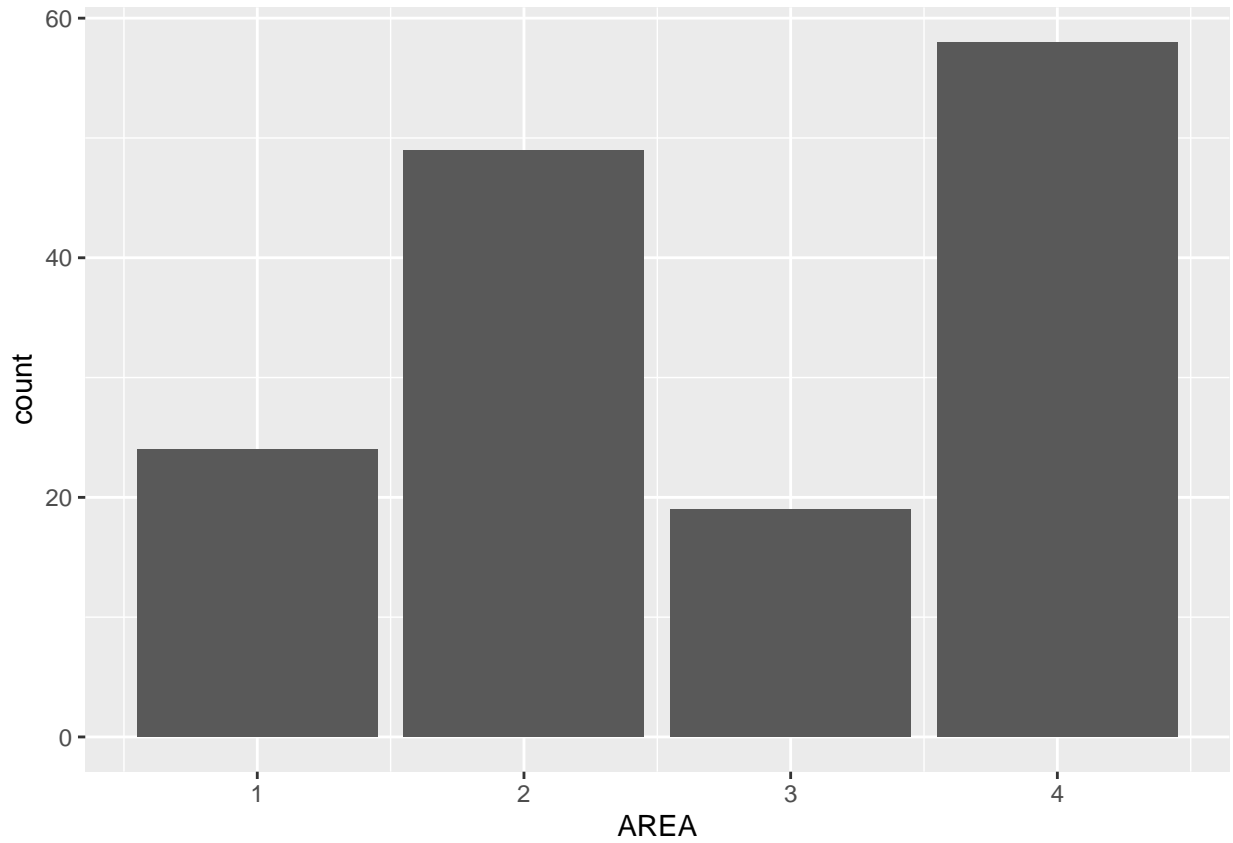
```
table(Lung$AREA, Lung$FSEX)
```

```
##
##       1
##   1 24
##   2 49
##   3 19
##   4 58
```
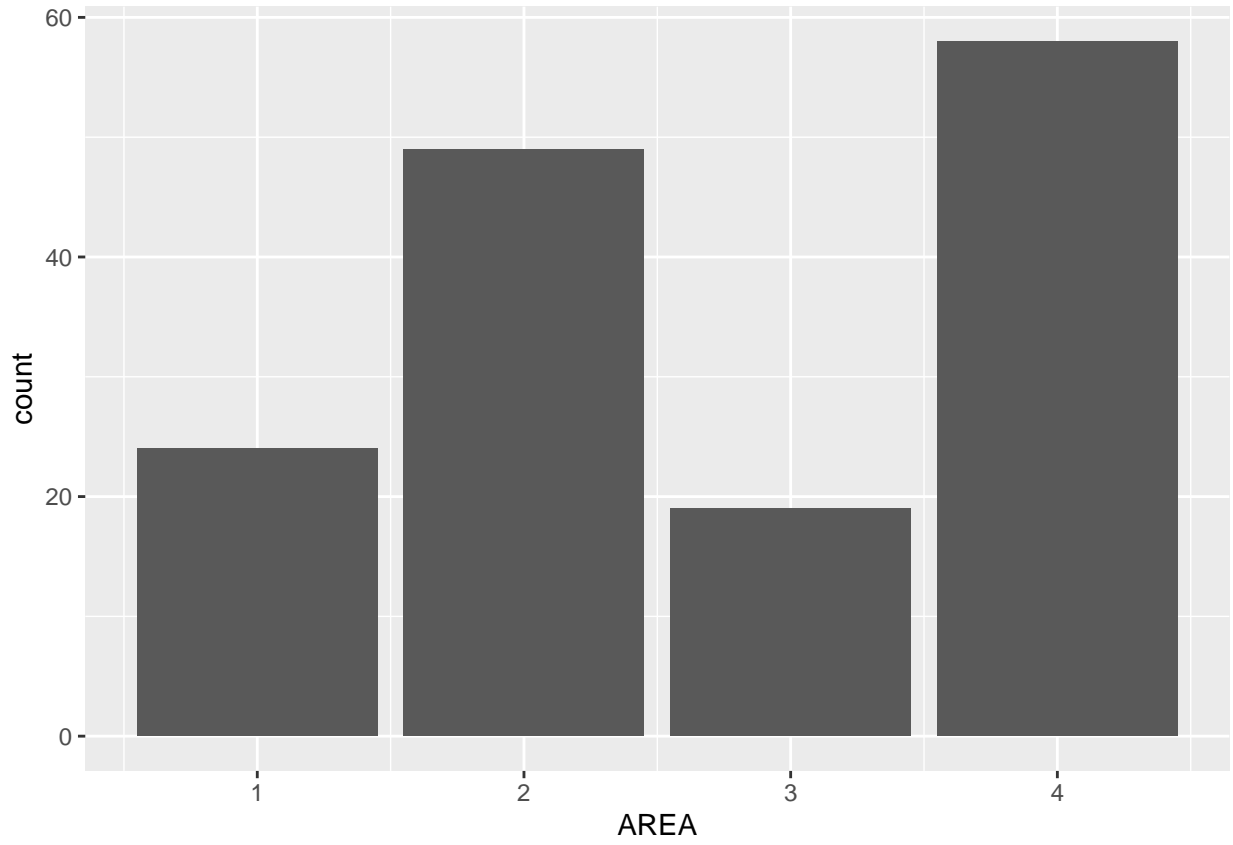
```
table(Lung$AREA, Lung$MSEX)
```

```
##
##       2
##   1 24
##   2 49
##   3 19
##   4 58
```

4

For this table we can find that both samples taken have an equal distribution between symptoms of Chronic Respirtory in all four areas of Los Angeles and Gender.
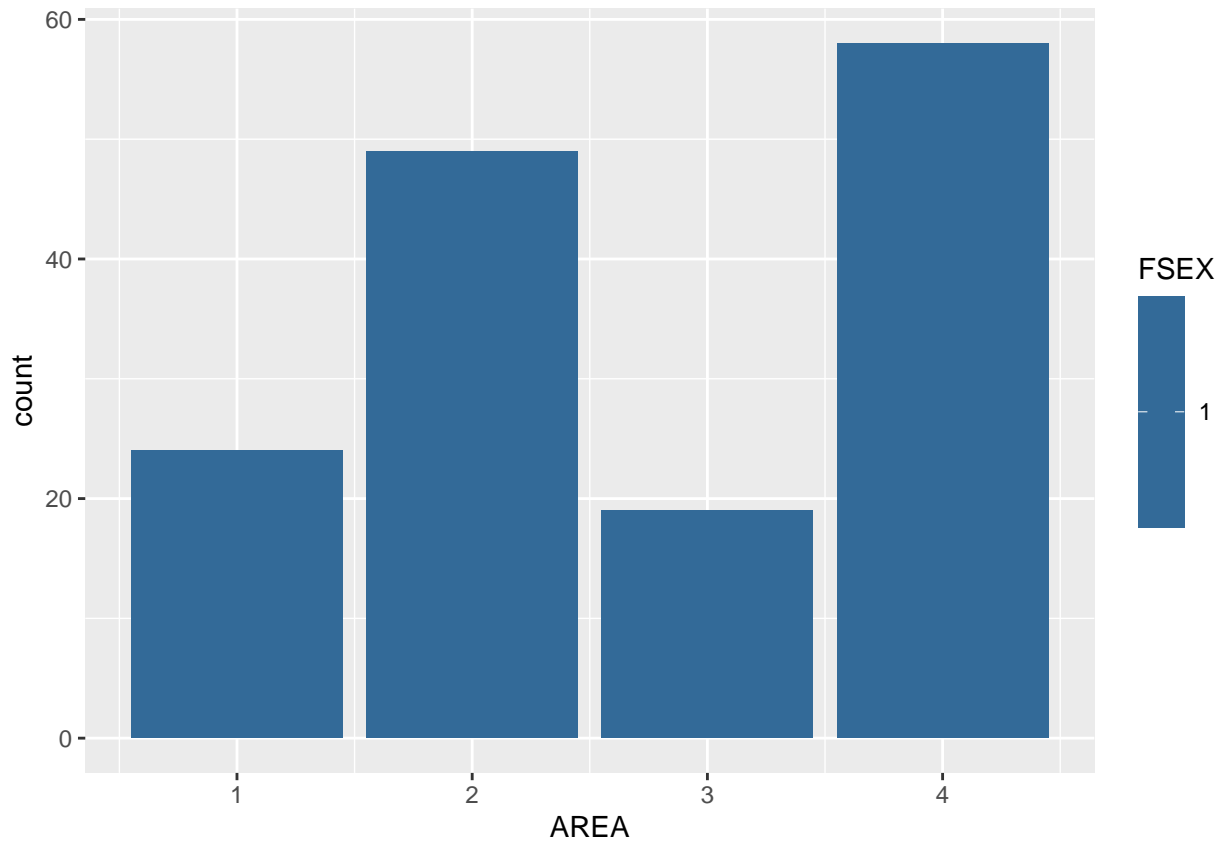
```
ggplot(Lung, aes(x= AREA, fill= FAGE)) + geom_bar(position = "dodge")
```
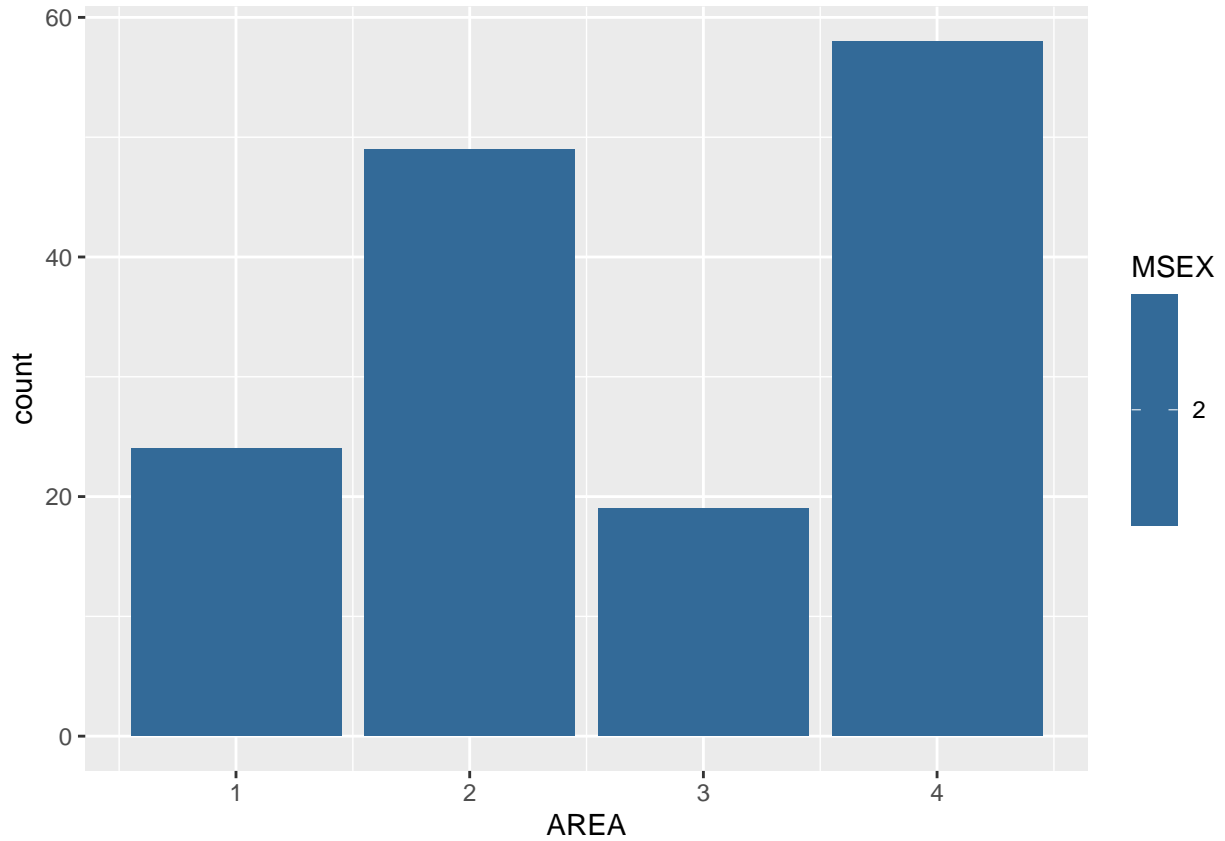


```
ggplot(Lung, aes(x= AREA, fill= MAGE)) + geom_bar(position = "dodge")
```

```
ggplot(Lung, aes(x= AREA, fill= FSEX)) + geom_bar(position = "dodge")
```

```
ggplot(Lung, aes(x= AREA, fill= MSEX)) + geom_bar(position = "dodge")
```

Above described the relation of Area between each variable, however based on the boxplot, it show that each individual area and variable compared to it are the same for both gender and ages.

Conclusion:

Based on the data of Lung function it was found that the relationship between the Area, Age, and Gender had no significance between one another. Based on the statistical data gather it was found that the only data found different it the mean in age between Male and Female. While using the boxplots it is also shown that the result are about identical.