**Final Project**

Sierra Squier

25 September 2020

Lung <-read.delim("c:/Users/Sierra/Desktop/Math 130/Data/Lung.txt")

**A short introduction/description of the data:**

The Lung Function data explores the effects of different types of smog on families in the Los Angeles area. Los Angeles is well known for having a significant amount of air pollution and this study explores how air pollution could impact the lung health of LA residents. The study takes data from subjects in four different areas of Los Angeles.

Variables being studied**:**

* Area
* Forced Vital Capacity (FVC) of the individuals

**A Univariate description of each of the variables under consideration:**

Area:

There were four areas in Los Angeles that subjects from the study resided in. Those areas were Burbank, Glendora, Lancaster, and Long Beach.

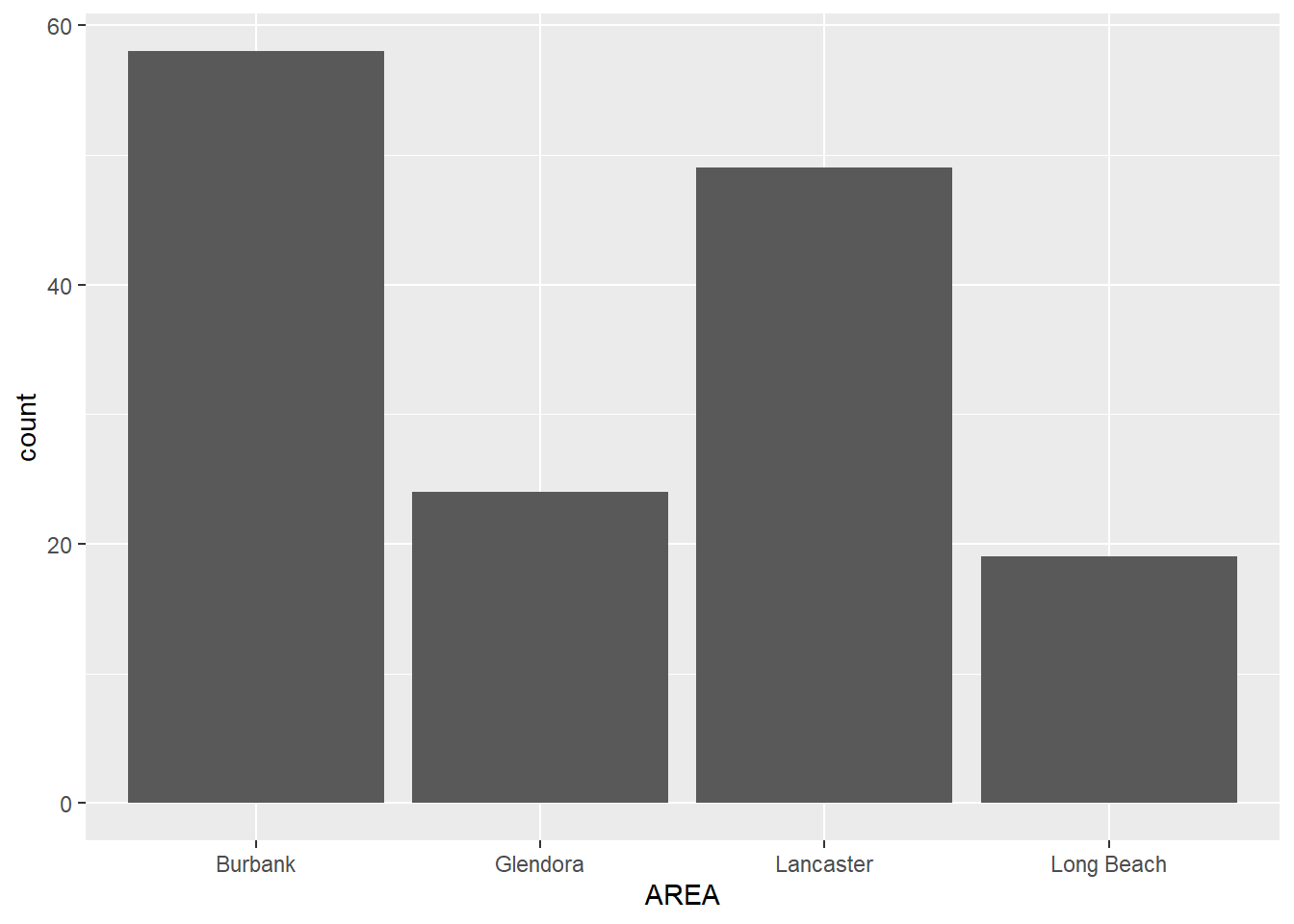
The table below shows the number of adult men and women whose data was used in the study. The count is taken from each area of Los Angeles that the study drew subjects from. The area that the subject was from is where their residence resides.

table (Lung$AREA)

## 1  2  3  4

## 24 49 19 58

The bar graph is a visual representation of the table so the data may be better understanding.



**ggplot**(L1, aes(x=AREA), main= "Subject Count in Each Area of Los Angeles”) + **geom\_bar** ()

Forced Vital Capacity (FVC) Between Adult Men and Women:

An FVC or forced vital capacity is the total amount of air exhaled during an FEV test. An FEV test can indicate a person’s lung health, especially if an FVC measurement is lower than normal. A normal FVC is between 300-500 centiliters.

Both means below are taken from the total FVC of each of the men and women from all the areas of Los Angeles in the study. The first mean represents the total men FVC, which you can see in higher than the mean for the total women FVC (the second mean).

1. mean(Lung$FFVC)

## [1] 495.2333

1. mean(Lung$MFVC)

## [1] 350.2333

The below histogram, density graphs are the visual representation of the FVC taken from both men and women across all areas of Los Angeles. The graph in purple represents the men’s FVC, while the pink represents the women. Since the men have a higher FVC on average, it could mean that the women in Los Angeles are more impacted by the smog. This of course is just speculation as not all factors are known why the women have a lover average FVC than men.

Chart, histogram

Description automatically generated

**ggplot**(L1, aes(x=FFVC), main="Forced Vital Capacity in Adult Men") +

**geom\_histogram**(aes(y=..density..), bins=15, alpha=.5, colour="purple", fill="lavender") +

**geom\_density**(col="black") +

**labs**(title="Forced Vital Capacity in Adult Men", x="Volume of Air (cl)")

Chart, line chart, histogram

Description automatically generated

**ggplot**(L1, aes(x=MFVC), main="Forced Vital Capacity in Adult Women") +

**geom\_histogram**(aes(y=..density..), bins=15, alpha=.5, colour="red", fill="pink") +

**geom\_density**(col="black") +

**labs**(title="Forced Vital Capacity in Adult Women", x="Volume of Air (cl)")

**A Bivariate comparison between two variables of interest:**

While the difference in the average FVC between men in women was interesting to see, I was more interested in looking at the difference between the average FVC of all the areas in Los Angeles that the study took place in.

The means below are an average of FVC of each area in Los Angeles for both men and women

Looking at each set of averages, it is obvious to see that the Lancaster area has the lowest average FVC. This could suggest that the people who resided hear are potentially more impacted by the smog. Though, it is important to note that this data does not include all possible factors that could be impacting people’s FVC.

**print**(mean\_data<- **list**(Men\_LG,Men\_LL,Men\_LB,Men\_LLB,Women\_LG,Women\_LL,Women\_LLB,Women\_LB))

## [[1]] -Men FVC mean Lancaster-

## [1] 485.1667

## [[2]] -Men FVC mean Glendora-

## [1] 481.7959

## [[3]] -Men FVC mean Long Beach-

## [1] 506.2414

## [[4]] -Men FVC mean Burbank-

## [1] 509

## [[5]] -Women FVC mean Lancaster-

## [1] 351.0417

## [[6]] -Women FVC mean Glendora-

## [1] 337.8367

## [[7]] -Women FVC mean Long Beach-

## [1] 340.2632

## [[8]] -Women FVC mean Burbank

## [1] 363.6379

The two panel graphs below are a visual representation of FVC averages in each area of Los Angeles for both men and women.

Chart, histogram

Description automatically generated

**ggplot**(L1, aes(x=MFVC), main=" Forced Vital Capacity in Adult Women") +

**geom\_histogram**(aes(y=..density..), bins=15, alpha=.5, colour="red", fill="pink") +

**geom\_density**(col="black") +

**labs**(title=" Forced Vital Capacity in Adult Women", x="Volume of Air (cl)") + facet\_wrap(~AREA)

Chart, histogram

Description automatically generated

**ggplot**(L1, aes(x=MFVC), main=" Forced Vital Capacity in Adult Men") +

**geom\_histogram**(aes(y=..density..), bins=15, alpha=.5, colour="purple", fill="lavender") +

**geom\_density**(col="black") +

**labs**(title=" Forced Vital Capacity in Adult Men", x="Volume of Air (cl)") + facet\_wrap(~AREA)