Analysis of relationships in pulmonary data

This program assesses the relationships among variables in a study of pulmonary function in children. There is a data dictionary that provides more details about the data. The program was written by Steve Simon on 2024-09-07 and is placed in the public domain.

Libraries

The tidyverse library is the only one you need for this program.

```
library(tidyverse)
```

List variable names

Since the variable names are not listed in the data file itself, you need to list them here.

```
pulmonary_names <- c(
    "age",
    "fev",
    "ht",
    "sex",
    "smoke")
```

Reading the data

Here is the code to read the data and show a glimpse.

```
pulmonary <- read_csv(
  file="../data/fev.csv",
  col_names=pulmonary_names,
```

```r
    col_types="nnncc")
  glimpse(pulmonary)
```

```
Rows: 654
Columns: 5
$ age   <dbl> 9, 8, 7, 9, 9, 8, 6, 6, 8, 9, 6, 8, 8, 8, 8, 7, 5, 6, 9, 9, 5, 5…
$ fev   <dbl> 1.708, 1.724, 1.720, 1.558, 1.895, 2.336, 1.919, 1.415, 1.987, 1…
$ ht    <dbl> 57.0, 67.5, 54.5, 53.0, 57.0, 61.0, 58.0, 56.0, 58.5, 60.0, 53.0…
$ sex   <chr> "F", "F", "F", "M", "M", "F", "F", "F", "F", "F", "F", "M", "F",…
$ smoke <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", "N",…
```

Question 1: Update the program to calculate descriptive statistics (mean, standard deviation, minimum, and maximum) for ht. Interpret these statistics.

```r
  summary(pulmonary$ht)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  46.00   57.00   61.50   61.14   65.50   74.00 
```

```r
  sd(pulmonary$ht)
```

```
[1] 5.703513
```

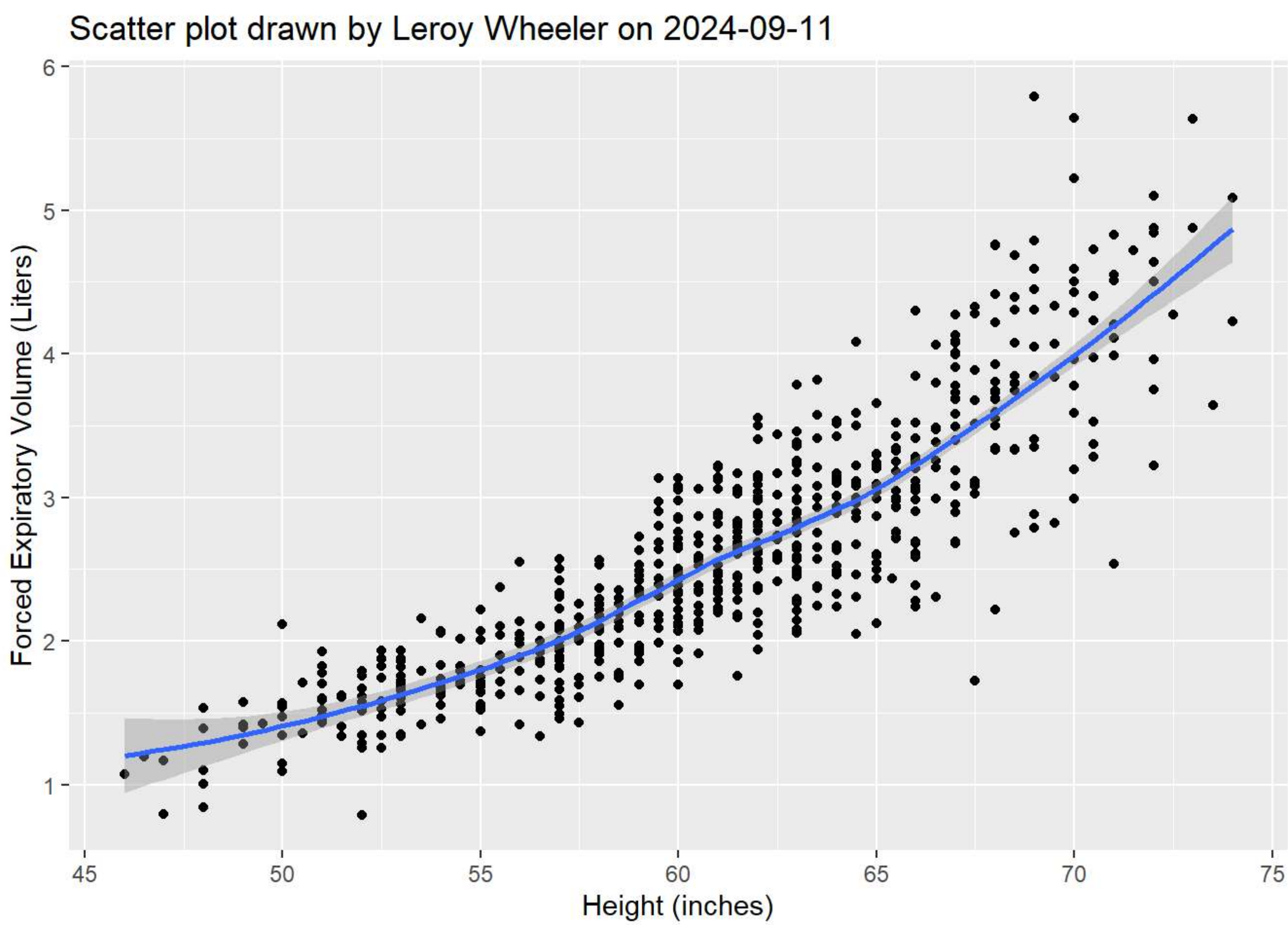
The mean height is about 61 inches with a standard deviation of almost 6 inches. Height ranges from 46 to 74 inches, which is consistent with a pediatric population.

Question 2: Draw a scatterplot of ht versus fev. Place ht on the x-axis and fev on the y-axis. Interpret this plot.

```r
  pulmonary |>
    ggplot(aes(ht, fev)) +
```

```
         geom_point() +
         geom_smooth() +
         xlab("Height (inches)") +
         ylab("Forced Expiratory Volume (Liters)") +
         ggtitle("Scatter plot drawn by Leroy Wheeler on 2024-09-11")
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



There is a positive linear association between height and fev. Calculation of r will likely confirm this observation.

Question 3: Calculate the correlation between ht and fev. Interpret this correlation.

```
cor(pulmonary$ht, pulmonary$fev)
```

```
[1] 0.868135
```

A correlation value of r=0.87 confirms the strong positive relationship between height and fev in this data set.

Question 4: Calculate counts and percentages for sex. Please be sure to convert sex from the numeric codes into a factor. Interpret these statistics.

```
pulmonary |>
  count(sex) |>
  mutate(total=sum(n)) |>
  mutate(pct=round(100*n/total))
```

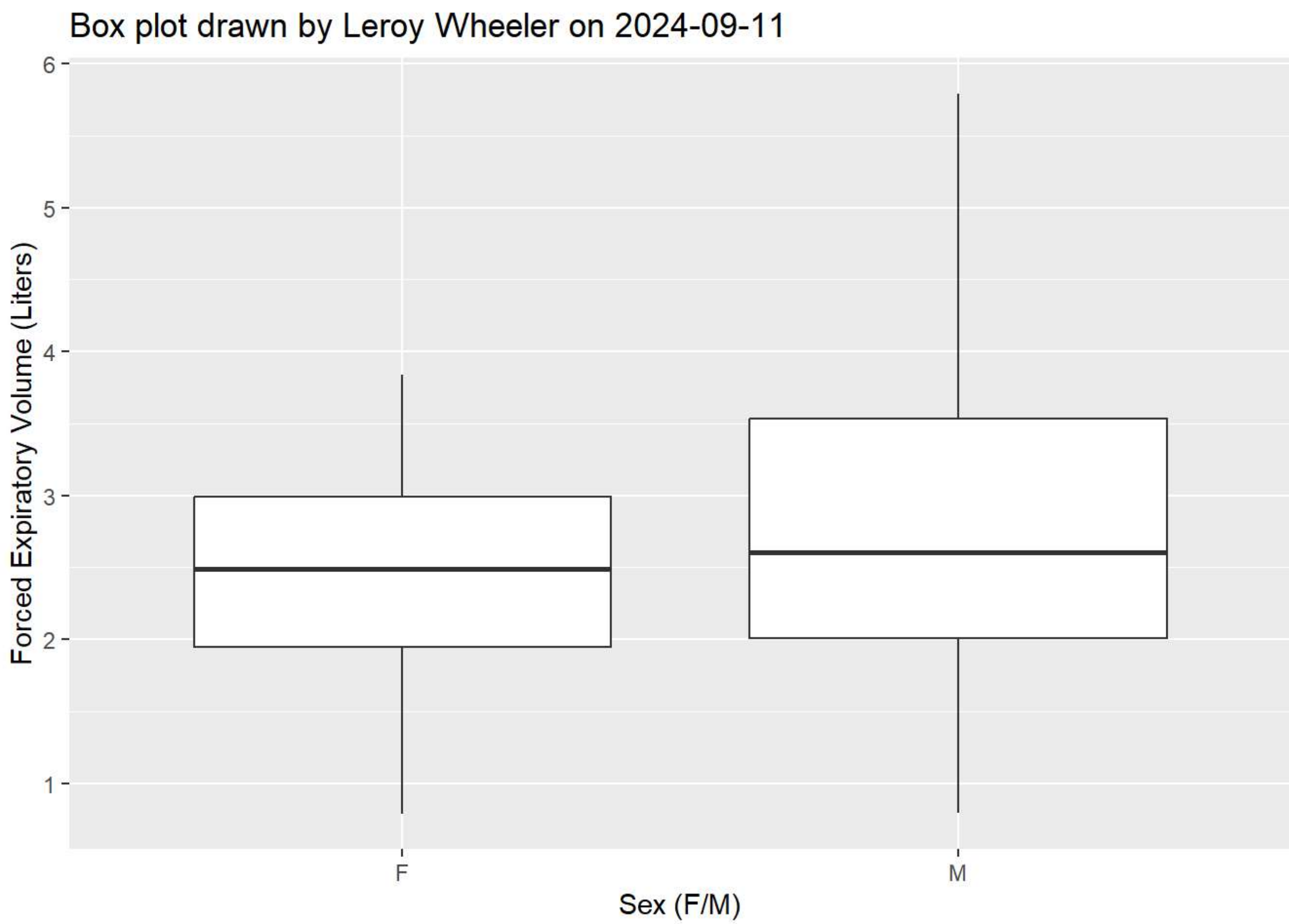
```
# A tibble: 2 × 4
  sex       n total   pct
  <chr> <int> <int> <dbl>
1 F       318   654    49
2 M       336   654    51
```

The data set was roughly split in half according to sex with 51% males and 49% females.

Question 5: Draw a boxplot for sex and fev. Interpret this boxplot

```
pulmonary |>
  ggplot(aes(sex, fev)) +
    geom_boxplot() +
    xlab("Sex (F/M)") +
```

```
        ylab("Forced Expiratory Volume (Liters)") +
        ggtitle("Box plot drawn by Leroy Wheeler on 2024-09-11")
```



The fev values are a little larger for males when compared to females. The variability for the male data is also slightly higher as well. These results are not surprising.

Question 6: Calculate the difference in average fev values between males and females. Is this a large or a small difference? Calculate the

effect size by dividing by the standard deviation of the females. Is this a small, medium, or large effect size?

```
pulmonary |>
  group_by(sex) |>
  summarize(
    mean_fev=mean(fev),
    sd_fev=sd(fev))
```

```
# A tibble: 2 × 3
  sex   mean_fev sd_fev
  <chr>    <dbl>  <dbl>
1 F         2.45  0.646
2 M         2.81  1.00
```

The average fev values for males is 2.8 which is larger than that observed in females, which is 2.5. Males also have a standard deviation of 1, which is also higher than the standard deviation of 0.6 seen in females.

The effect size between males and females is approximately 0.6 standard deviations.