Statistics Week 2 R Homework Question 1.1

In an experiment to investigate the heat of sublimation of iridium, the following 27 measurements were made, listed across the rows in the order they were taken. The data is contained in the Statistics 1 data set `iridium`.

```
load('C:/.A - Uni/Year 1/Maths/Probability and Stats/Y1-TB2-STATS-R/stats1.RData')
labelforaxis=(expression(paste("Enthalpy of Sublimation of Iridium (kJ mol"^"-1", ")")))
iridium
```

```
##  [1] 136.6 145.2 151.5 162.7 159.1 159.8 160.8 173.9 160.1 160.4 161.1 160.6
## [13] 160.2 159.5 160.3 159.2 159.3 159.6 160.0 160.2 160.1 160.0 159.7 159.5
## [25] 159.5 159.6 159.5
```

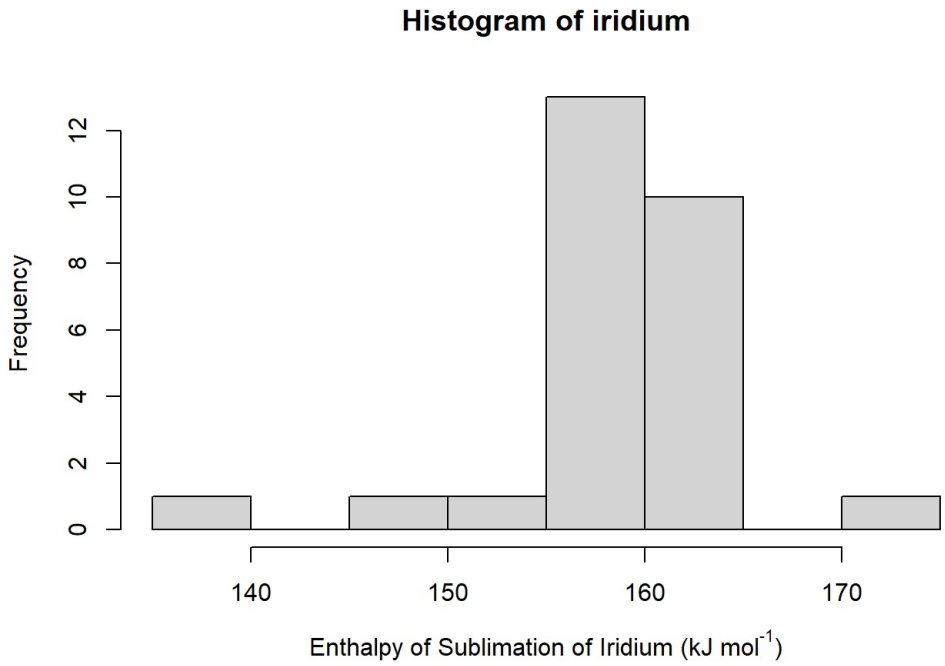
(a)

Use the `R` commands `stem`, `hist`, `boxplot`, and `plot` to make a stem-and-leaf plot, a histogram, a boxplot and a plot of the observations in the order they were taken. Print your plots and comment on the overall pattern and any striking features.

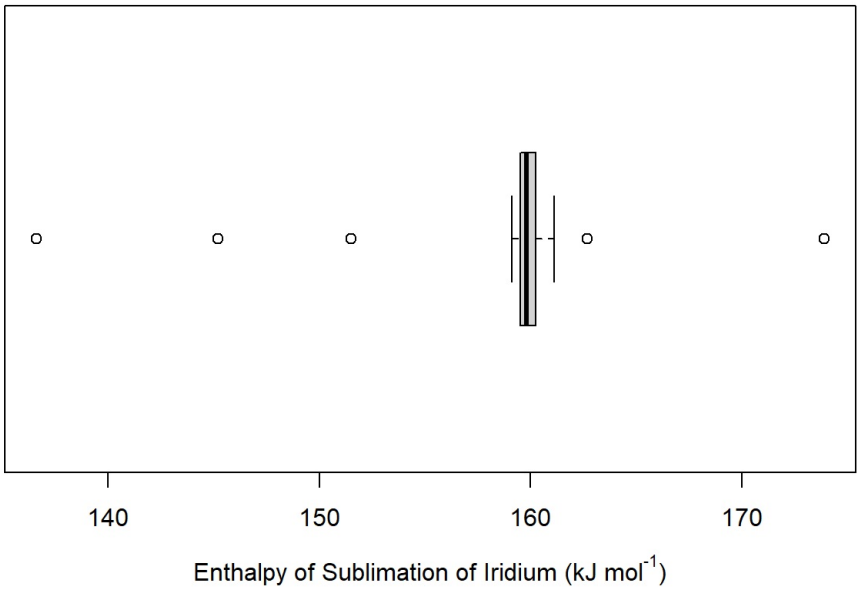
```
stem(iridium)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   13 | 7
##   14 |
##   14 | 5
##   15 | 2
##   15 | 999
##   16 | 00000000000000001113
##   16 |
##   17 | 4
```

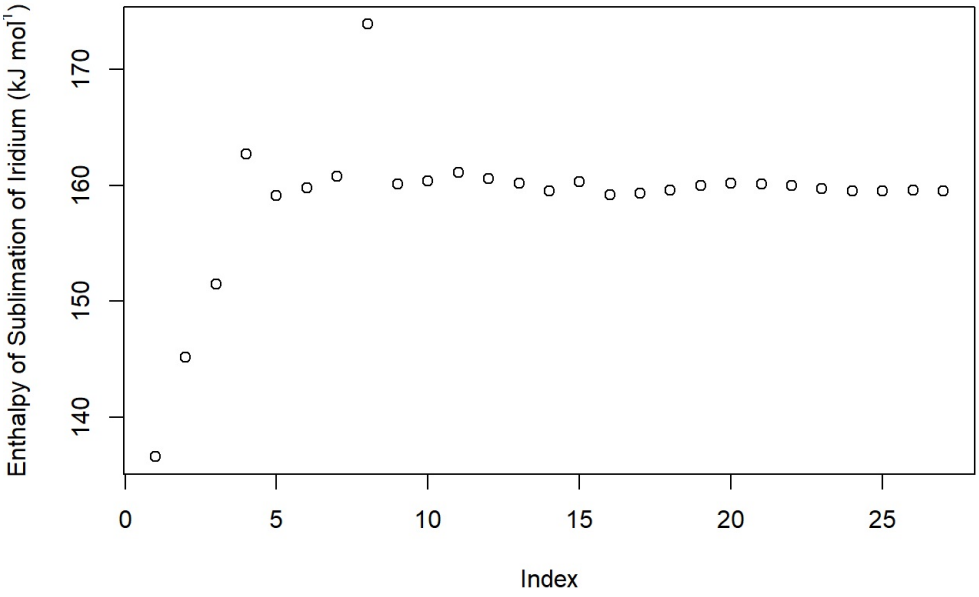
```
hist(iridium, xlab = labelforaxis)
```




```
boxplot(iridium, xlab = labelforaxis, horizontal = TRUE)
```




```
plot(iridium, ylab = labelforaxis)
```




```
Overall, the data values are very consistent, with most values being very close to 160.
There are, however, a few anomalous points.
```

(b)

Use the `R` commands `median` and `mean` to find the median and the mean.

```
median(iridium)
```

```
## [1] 159.8
```

```
mean(iridium)
```

```
## [1] 158.8148
```

Use `mean` in `R` to see how to compute a trimmed mean.

Compute the 10% and 20% trimmed means for the iridium data set.

```
mean(iridium, trim=0.1)
```

```
## [1] 159.5478
```

```
mean(iridium, trim=0.2)
```

```
## [1] 159.8412
```

Compare how well the the mean and median and trimmed means represent the centre of this dataset.

As there are a few extreme values, the mean is likely to be the least accurate representation of the centre of the dataset. The median, 10% and 20% trimmed means, all seem to be fairly good representations of the centre of the dataset.

(c)

Use the `R` commands `var` and `sd` to find the sample variance and standard deviation.

```
var(iridium)
```

```
## [1] 38.74516
```

```
sd(iridium)
```

```
## [1] 6.224561
```

Use the `R` commands `fivenum` and `summary` to find the hinges and the sample quartiles, and use `R` command `IQR` to find the interquartile range (but see comments on 'Hinges and Quartiles' overleaf).

```
fivenum(iridium)
```

```
## [1] 136.60 159.50 159.80 160.25 173.90
```

```
summary(iridium)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   136.6   159.5   159.8   158.8   160.2   173.9
```

So:
H_1=159.50
H_3=160.25
Q_1=159.5
Q_3=160.2

```
IQR(iridium)
```

```
## [1] 0.75
```

Again, compare how these values represent the spread of the data.

The data is fairly consistent, in particular it has a very low IQR. However, since there are some extreme values, the variance and standard distribution are quite large.

(d)

What conclusions do you draw from your plots and numerical summaries? What effect do the outliers have on the numerical and graphical summaries? What would the corresponding results look like if the outliers were removed?

The enthalpy of sublimation of iridium is most likely to be about 159.8 kJ mol^{-1} (the median of the data set).
Since the data is negatively skewed, the outliers lead to a smaller mean.
If the outliers were removed, the mean would be larger.