5. {hand calc mean, variance and standard deviation, simple summary stats with R}
The following table shows literacy rates among the youths (age 15- 24) for some of the countries in Sub Saharan Africa

Country	Literacy Rate (%)
Angola	72.2
Burundi	73.3
Democratic Republic of Congo	70.4
Rwanda	77.6
Kenya	80.3

a. Calculate mean and median. Also calculate variance and standard deviation.

1A) Mean:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(x_i) = \frac{1}{5}(72.2 + 73.3 + 70.4 + 77.6 + 80.3) = 74.76\%$$

Median:
1) sort them $\Longrightarrow 80.3, 77.6, 73.3, 72.2, 70.4$
2) pick the one in the middle: 73.3%

Variance:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$
$$= \frac{1}{4}((80.3 - 74.76)^2 + (77.6 - 74.76)^2 + (73.3 - 74.76)^2 + (72.2 - 74.76)^2 + (70.4 - 74.76)^2) = 16.613\%$$

Standard Deviation:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{16.613} = 4.07590481\%$$

b. Now let's add one more country to the data

Chad, Literacy Rate = 37.6

Repeat the exercise in A.

Now, our sample is $80.3, 77.6, 73.3, 72.2, 70.4, 37.6$

$$\hat{\mu} = \frac{1}{6}(72.2 + 73.3 + 70.4 + 77.6 + 80.3 + 37.6) = 68.5666667\%$$

Median $= (73.3 + 72.2)/2 = 72.75\%$

$$\hat{\sigma}^2 = \frac{1}{5}((80.3 - 68.5666667)^2 + (77.6 - 68.5666667)^2 + (73.3 - 68.5666667)^2 + (72.2 - 68.5666667)^2 + (70.4 - 68.5666667)^2 + (37.6 - 68.5666667)^2) = 243.434667$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = 15.602393\%$$

Result: Note that by adding Chad (a country whose lit rate is significantly lower than the five African countries above, the means drops from 75% to 69%, while the median moves only from

73.3% to 72.75%
Also note the significant increase in the standard deviation adding Chad, going from 4% to 17%

c. Now calculate the mean, median, mode, variance and standard deviation in R. Do this for a and b. Show both sets of observations with a histogram as well. Show your code.

R code for Problem 1) (FYI: `red is what you type in R`, green font is output)

Input Initial Five Countries:

```
literacy = c(72.2, 73.3, 70.4, 77.6, 80.3)
 literacy
[1] 72.2 73.3 70.4 77.6 80.3
 mean(literacy)
[1] 74.76


median(literacy)
 [1] 73.3
var(literacy)
[1] 16.613
sd(literacy)
[1] 4.075905
var(literacy)^0.5
[1] 4.075905
```

Now, adding in Chad

```
literacy_f = c(72.2, 73.3, 70.4, 77.6, 80.3, 37.6)
mean(literacy_f)
[1] 68.56667
median(literacy_f)
[1] 72.75
var(literacy_f)
[1] 243.4347
sd(literacy_f)
[1] 15.60239
```

Now, visualizing these observations

```
hist(literacy,main="Literacy in Africa", xlab="Literacy")
Output: a nice histogram
hist(literacy_f,main="Literacy in Africa", xlab="Literacy including Chad")
Output: a nice histogram
hist(literacy_f,20,main="Literacy in Africa", xlab="Literacy including Chad")
Output: a nice histogram, with a  total of 20 bin
```