Exercise 1

(5 points) Under what circumstances would you use the mean as a measure of central tendency instead of the median? and vice-versa?

We use the mean as a measure of central tendency when the distribution is symmetric. On the other hand, we use the median as a measure of central tendency when the distribution is skewed.

Exercise 2

(5 points) A basketball team free throw percentages are: 50, 98, 25, 76, 88, 75, 80. Compute the range and IQR.

```
throw_pcts = c(50, 98, 25, 76, 88, 75, 80)
range = max(throw_pcts) - min(throw_pcts)
IQR = as.numeric(quantile(throw_pcts, 0.75)) - as.numeric(quantile(throw_pcts, 0.25))
```

The range is 73, and the IQR is 21.5

Exercise 3

(8 points) Consider the statistics of two teams from the English Premier League in the 2019-2020 season:

Team	Total Goals	Wins	Losses
Manchester City	102	26	9
Liverpool	85	32	3

which team has the higher variability in terms of game outcome?

```
p_wm = 26 / 35
p_lm = 9 / 35
entropy_m = -(p_wm*log(p_wm) + p_lm*log(p_lm)) / log(2)

p_wl = 32 / 35
p_ll = 3 / 35
entropy_l = -(p_wl*log(p_wl) + p_ll*log(p_ll)) / log(2)
```

The standardized entropy of Manchester City is 0.82 while the standardized entropy of Liverpool is 0.42; thus, Manchester City has a higher variability.

Exercise 4

Consider the batting average from 2009 to 2015 baseball players:

Player	Position	2009	2010	2011	2012	2013	2014	2015
Marlon Byrds	Outfielder	0.283	0.293	0.276	0.210	0.291	0.264	0.247
Sam Fuld	Outfielder	0.299	0.143	0.240	0.255	0.199	0.239	0.197

In R, answer the following:

(a) (5 points) Create two vectors with called: `byrd` and `fuld` to store their corresponding batting average.

```
byrd = c(0.283, 0.293, 0.276, 0.210, 0.291, 0.264, 0.247)
fuld = c(0.299, 0.143, 0.240, 0.255, 0.199, 0.239, 0.197)
```

(b) (5 points) Report the average batting average of both players. What player has the higher average batting average from 2009 to 2015?

```
## Computing the mean batting average of Byrd
mean(byrd)

## Computing the mean batting average of Fuld
mean(fuld)
```

Marlon Byrds has a higher average batting average.

(c) (5 points) Report the standard deviation of the batting average of both players. What player has the higher variability in their batting average from 2009 to 2015?

```
## Computing the standard deviation of the batting average of Byrd
sd(byrd)

## Computing the standard deviation of the batting average of Fuld
sd(fuld)
```

Sam Fuld has the higher variability in his batting average.

(d) (5 points) Report the coefficient of variation of the batting average of both players. What player has the higher variability in their batting average from 2009 to 2015 in terms of their CVs?

```
## Computing the CV of the batting average of Byrd
sd(byrd) / mean(byrd)

## Computing the CV of the batting average of Fuld
sd(fuld) / mean(fuld)
```

Sam Fuld has the higher variability in his batting average in terms of the coefficient of variation.

(e) (5 points) Report the IQR of the batting average of both players. What player has the higher variability in their batting average from 2009 to 2015 in terms of their IQR?

```
## Computing the IQR of the batting average of Byrd
as.numeric(quantile(byrd, 0.75)) - as.numeric(quantile(byrd, 0.25))

## Computing the IQR of the batting average of Fuld
as.numeric(quantile(fuld, 0.75)) - as.numeric(quantile(fuld, 0.25))
```

Sam Fuld has the higher variability in his batting average in terms of the IQR.