Assignment 2

Luis Nicolas Luarte Rodriguez

```
library(tidyverse)
library(knitr)
library(ggplot2)
library(knitr)
library(kableExtra)
```

```
# load the dataset
dataSet <- as_tibble(mtcars)
head(dataSet)
```

```
## # A tibble: 6 x 11
##     mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21       6   160   110  3.9   2.62  16.5     0     1     4     4
## 2  21       6   160   110  3.9   2.88  17.0     0     1     4     4
## 3  22.8     4   108    93  3.85  2.32  18.6     1     1     4     1
## 4  21.4     6   258   110  3.08  3.22  19.4     1     0     3     1
## 5  18.7     8   360   175  3.15  3.44  17.0     0     0     3     2
## 6  18.1     6   225   105  2.76  3.46  20.2     1     0     3     1
```

Compute the following measures for the variable that you selected for plotting in the previous assignment (I used mtcars dataset and the variable 'mpg')

Statistics	Values
Mean	20.090625
Median	19.200000
Range	23.500000
Standard Deviation	6.026948
Variance	36.324103
Percentile 10	14.340000
Percentile 90	30.090000
Quartile 1	15.425000
Quartile 3	22.800000
Inter-quartile range	7.375000

What is the percentage of the data between the mean +/- 1 std?

Criteria	Percent
1 sd	0.7500
2 sd	0.9375
3 sd	1.0000

Is the distribution of your data close to a normal distribution?

If we compare the values obtained with the ones coming from a normal distribution we get the following differences:

Normal	MPG	Difference
68.27	75	6.73
95.45	93	2.45
99.73	100	0.27

The differences are quite small, and this is indication that is close enough to a normal distribution

```
# We can test for normality using the shapiro-wilk test
shapiro.test(d)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  d
## W = 0.94756, p-value = 0.1229
```

```
# The distribution of the data are not significantly different from normal distribution
```