Week 4 in class exercise

1. Download the grades data from " http://yegingenc.com/lectures/data/SampleStudentGrades.txt (http://yegingenc.com/lectures/data/SampleStudentGrades.txt)" and save it to a '.csv' file.

2. Load the data to R.

```
data=read.delim('http://yegingenc.com/lectures/data/SampleStudentGrades.txt')
```

3. Which variables are numerical?

Grades are numerical.

4. What are average and standard deviations for each semester?

```
semesters=unique(data$Semester)
semesters
```

```
## [1] 14_Fall   15_Fall   15_Spring
## Levels: 14_Fall 15_Fall 15_Spring
```

```
Fall14<-data[data$Semester==semesters[1],"Grades"]  ## Using the the semesters vector I ju
st created to minimize the possibilty of a type
Fall15<-data[data$Semester==semesters[2],"Grades"]
Spring15<-data[data$Semester==semesters[3],"Grades"]

mean(Fall14) ;sd(Fall14)
```

```
## [1] 80.51042
```

```
## [1] 14.98811
```

```
mean(Fall15)  ;sd(Fall15)
```

```
## [1] 77.41824
```

```
## [1] 17.16678
```

```
mean(Spring15)  ;sd(Spring15)
```

```
## [1] 74.64286
```

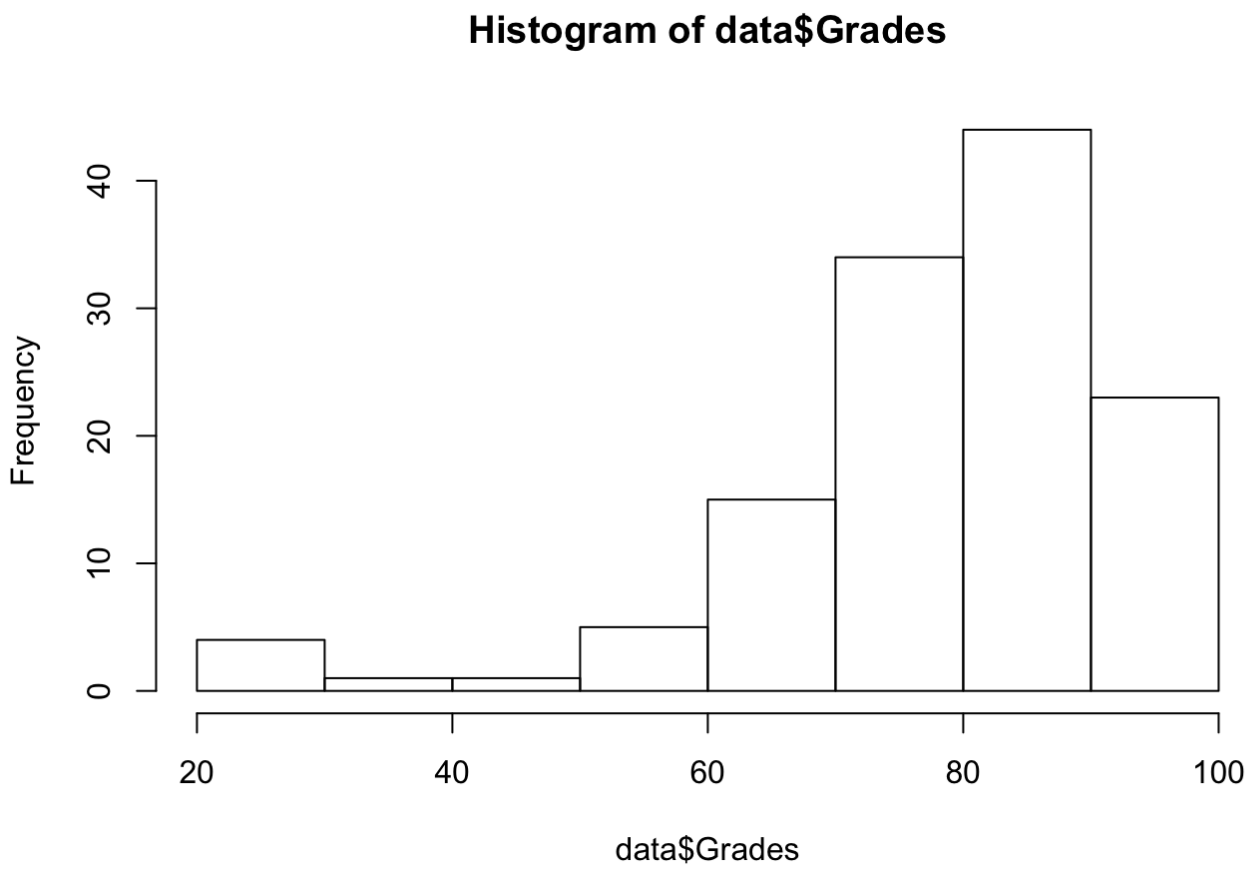
```
## [1] 12.1189
```

```
##Alternatively we can take advantage of libraries like dplyr and magrittr
library(magrittr)
library(dplyr)
data %>%
  group_by(Semester) %>%
  summarise(Avg = mean(Grades), SD = sd(Grades))
```

```
## # A tibble: 3 x 3
##    Semester      Avg       SD
##      <fctr>    <dbl>    <dbl>
## 1   14_Fall 80.51042 14.98811
## 2   15_Fall 77.41824 17.16678
## 3 15_Spring 74.64286 12.11890
```

5. Plot a histogram for the grades?

```
hist(data$Grades)
```



6. Judging by the histogram you just created what can you say about the distribution?

The data is negatively skewed in histogram.

7. Calculate the skewness of the data.

```
skewness <- function(x) {
return(3 * (mean(x) - median(x))/sd(x))
}

skewness(data$Grades)
```

```
## [1] -0.6865464
```

8. Using transformation techniques you just learnt try to normalize the grades distribution.