BASM - HW 1

Jason Tang 102034038

0. Load the exam data

```
exam <-read.table("c:/Users/tsunh/Desktop/exam_results.txt", header=TRU
E)
exam
```

```
##    scores
## 1      79
## 2      52
## 3      89
## 4      58
## 5      65
## 6      82
## 7      67
## 8      48
## 9      71
## 10     51
## 11     18
## 12     65
## 13     91
## 14     74
## 15     77
## 16     54
## 17    100
## 18     84
## 19     70
## 20     60
## 21     49
## 22     54
## 23     84
## 24     94
## 25     68
## 26     65
## 27     57
## 28     80
## 29     82
## 30     67
## 31     69
## 32     61
## 33     90
## 34     95
## 35     71
## 36     72
## 37     10
## 38     57
```

```
## 39     70
## 40     75
## 41     75
## 42     76
## 43     52
## 44     88
## 45     59
## 46     65
## 47     57
## 48     75
## 49     53
## 50     67
```

```
#Use head() to see just few rows
head(exam)
```

```
##   scores
## 1     79
## 2     52
## 3     89
## 4     58
## 5     65
## 6     82
```

1. What is the 5th element in the original list of correct grades?

```
exam$scores[5]
```

```
## [1] 65
```

2. What is the fifth lowest grade?

```
sort(exam$scores)[5] #the fifth lowest grade
```

```
## [1] 51
```

3. Extract the five lowest grades together

```
sort(exam$scores)[1:5]
```

```
## [1] 10 18 48 49 51
```

4. Get the five highest scores by first sorting exam$scores in decreasing order.

```
sort(exam$scores,decreasing = T)[1:5]
```

```
## [1] 100  95  94  91  90
```

5. What is the standard deviation of scores? (guess or google the standard deviation command)

```
sd(exam$scores) # the standard deviation of scores
```

```
## [1] 17.23826
```

6. Make a new variable called scores_diff, with the difference between each grade and the mean grade

```
scores_diff <- exam$scores -mean(exam$scores)
scores_diff
```

```
##  [1]  11.16 -15.84  21.16  -9.84  -2.84  14.16  -0.84 -19.84   3.16
-16.84
## [11] -49.84  -2.84  23.16   6.16   9.16 -13.84  32.16  16.16   2.16
 -7.84
## [21] -18.84 -13.84  16.16  26.16   0.16  -2.84 -10.84  12.16  14.16
 -0.84
## [31]   1.16  -6.84  22.16  27.16   3.16   4.16 -57.84 -10.84   2.16
  7.16
## [41]   7.16   8.16 -15.84  20.16  -8.84  -2.84 -10.84   7.16 -14.84
 -0.84
```

7. What is the average difference between each grade and the mean of all grades?

```
mean(scores_diff)
```

```
## [1] -3.410741e-15
```

```
# It might be close zero since we just centralized the data and make th
e distribution to zero mean. Sometimes we do this as one of the step fo
r data normalization.
```

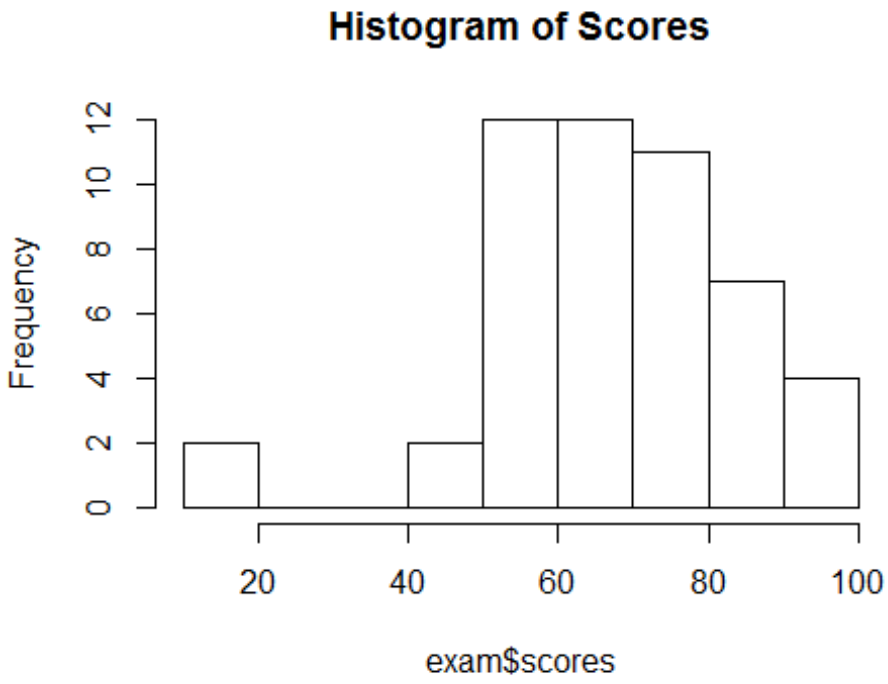
```
mean(abs(scores_diff))
```

```
## [1] 12.6864
```

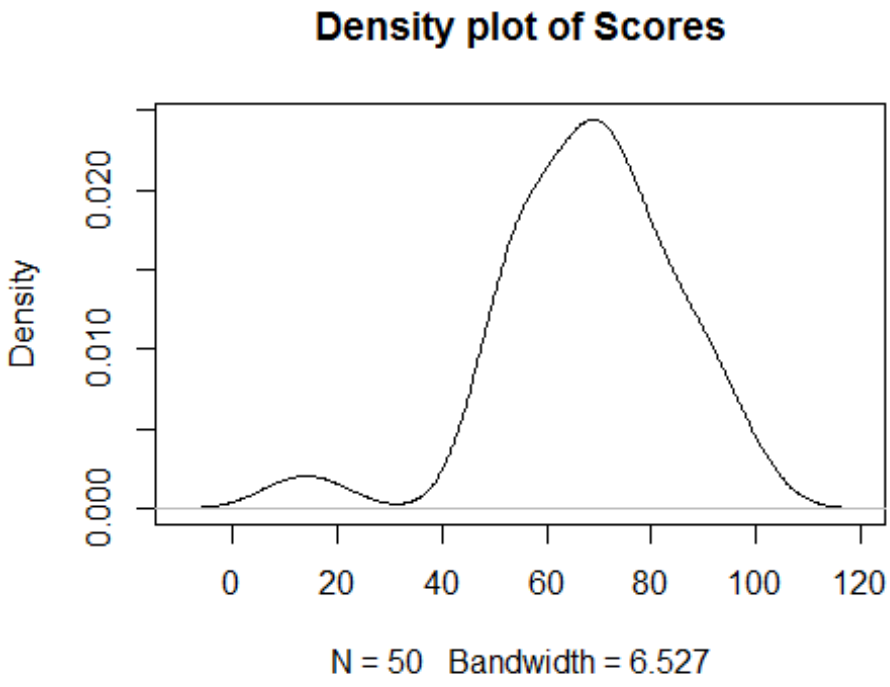
```
# We can also caculate the average of absolute differences, which is so
-call mean absolute deviation (MAD), to observe the dispersion of data.
```

8. Visualize the data as we did in class: histogram, density plot, boxplot+stripchart

```
hist(exam$scores,main="Histogram of Scores")
```




```r
plot(density(exam$scores),main="Density plot of Scores")
```



```r
boxplot(exam$scores, horizontal = TRUE)
stripchart(exam$scores, method = "stack", add = TRUE)
```