

Lecture 7 - R Software

Ashutosh Rajput
Department of Mathematics

Rajdhani College
University of Delhi

How to explore relationships

We can use R to learn about the relationship between two variables. Some of the functions that summarize one set of data can work on multiple sets of data. For example, if we have data about two exams in a Calculus class, we can compare the box plots to get an idea about the relative difficulty of the exams.

Example: We look at box plots of exam scores from two exams.

```
> exam1 = scan()  
1 : 89 64 72 91 74  
6 : 94 99 78 67 85  
11 : 71 97 80 72 79  
16 : 91 86 92 91 86  
21 : 64 88 89  
24 :  
Read 23 items
```

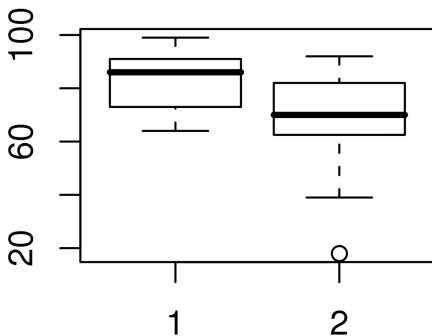
```
> exam2 = scan()  
1 : 73 48 86 88 69  
6 : 89 92 64 39 46  
11 : 68 81 75 70 75  
16 : 61 55 85 73 66
```

```
21 : 18 68 83
```

```
24 :
```

```
Read 23 items
```

```
> boxplot( exam1, exam2 )
```

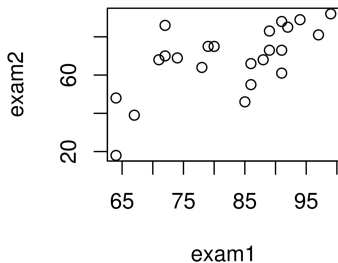


- It would seem that the second exam was generally harder than the first.
- The median is visibly lower on the second exam.
- The range of values extends much lower than the first.

Another way to find the quick and close relationship between two datas, we use `plot()` function.

Therefore, to see if individual students score consistently from one exam to the next, we can use a scatter plot, generated by the `plot()` function.

```
> plot( exam1, exam2 )
```



If there is a trend(variation) in these scores, it seems to be weak.

However, it seems to be some pattern. Students who scored lower on exam 1 also appear to score lower on exam 2, at least in a broad sense. It looks like there could potentially be a weak linear relationship between exam scores on exam 1 and corresponding scores on exam 2.