Unofficial Kahn Academy R Supplement: Range, IQR, and MAD / Box and Whisker Plots

Similar to the R commands for mean and median, finding the range is a simple matter of using the *range* command (which gives you the highest and lowest value). You can also calculate the range by hand. As is probably not surprising to you at this point, *max()* finds the maximum value in a variable, so in our gas mileage example it gives us the mpg for the car with the highest mpg, or best gas mileage. Conversely *min()* gives us the minimum value. So if we wanted to calculate the range by hand, we could simply subtract one from the other.

```
max(mtcars$mpg)-min(mtcars$mpg)
```

```
## [1] 23.5
```

```
range(mtcars$mpg)
```

```
## [1] 10.4 33.9
```

Similarly, the interquartile range is found using the *IQR* command. Once again, you can easily do it by hand using R as a calculator. The *quantile* command gives us the different quartiles.

```
quantile(mtcars$mpg)
```

```
##     0%    25%    50%    75%   100%
## 10.400 15.425 19.200 22.800 33.900
```

```
# Now we will subtract the 75th percentile from the 25th percentile.
22.800-15.425
```

```
## [1] 7.375
```

```
IQR(mtcars$mpg)
```

```
## [1] 7.375
```

AYCS, both the *IQR* command, as well as the answer we get from the arithmetic are identical.

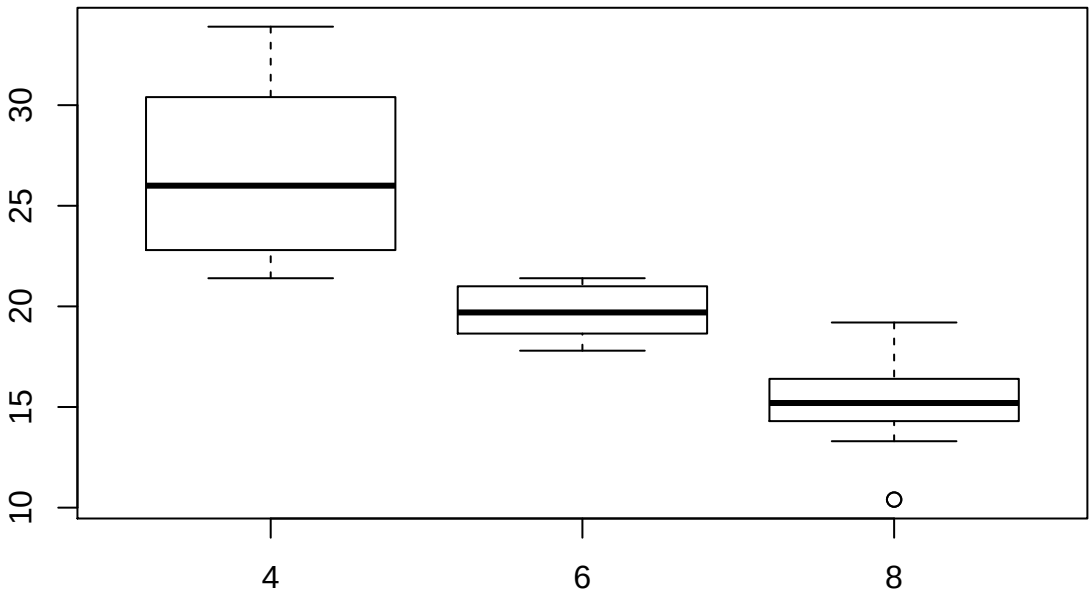
Finally, to find the mean absolute deviation one uses the *mad* command. The *mad* command, however, requires that you specify that you want the mean absolute deviation and not the median absolute deviation by including a *, center= mean(X)* subcommand. If you do not include it, then the *mad* command will automatically calculate the median absolute deviation.

```
mad(mtcars$mpg, center=mean(mtcars$mpg))
```

```
## [1] 6.37518
```

As noted in KA, Another way to visualize these themes is through a box-and-whisker plot (or "boxplot"). A basic boxplot can be created using the *boxplot* command, where the variable being measured is placed first in the command, followed by a ~, then the variable by which the different plots are organized. Finally, the dataset the variables are drawn from is referred to.

```
boxplot(mpg~cyl,data=mtcars)
```



As with other graphs, with a little Googling it's relatively easy to add X-axes labels, titles, etc.

Finally, the standard deviation of a variable is easy to find with the *sd* command. Because the variance is the standard deviation squared, we can find the variance by squaring the standard deviation. Alternatively, we can find the variance using the *var* command, but this will give us the same result

```
sd(mtcars$mpg)
```

```
## [1] 6.026948
```

```
(sd(mtcars$mpg))^2
```

```
## [1] 36.3241
```

```
var(mtcars$mpg)
```

```
## [1] 36.3241
```