Numerical Summaries with R

Do men and women differ on risk ratings for appliances? Planes?

```
> mean(mappliance,na.rm=T)
[1] 18.38636
> mean(fappliance, na.rm=T)
[1] 20.70822
> sd(mappliance,na.rm=T)
[1] 20.62296
> sd(fappliance,na.rm=T)
[1] 20.98631
```

So women rate appliances as being a little riskier then men. By the empirical rule, we can expect 68% of the men to be within 18.3 +- 20.7. But because negative scores aren't allowed, we can infer that this distribution is not symmetric. (We knew that already.)

Men and women averages differ by about 2 points, which is roughly 10% of a SD, which means that there is really very little difference or, put differently, plenty of overlap between the distributions.

We can visualize this:

```
> boxplot(appliances ~ gender)
```
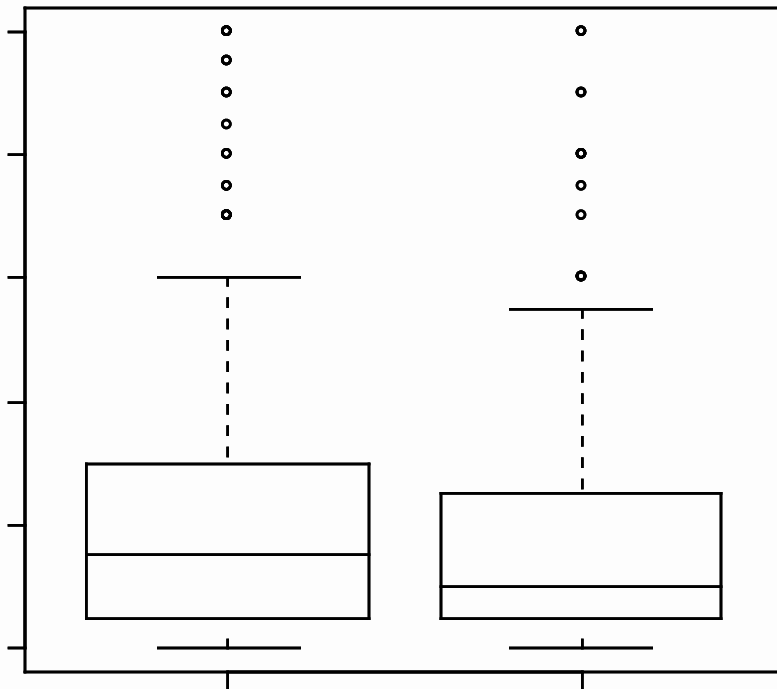
Men are on the left, women on the right. (The wordprocessor doesn't copy the axes labels for some reason.) The tickmarks on the y-axis are at 0,20,40,etc.

The command boxplot(mappliance,fappliance) would work just as well.

We see here that that the distributions are greatly skewed; both have "tails" on the upper end. The men's median is slightly higher than the women's, and there is more diversity of opinion in the men as well. (The box is bigger).
The 5-number summary:
```
> summary(mappliance)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    5.00   10.00   18.39   25.00  100.00    6.00
> summary(fappliance)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    5.00   15.00   20.71   30.00  100.00    8.00
```
Which doesn't really tell us anything we didn't know from the boxplot.