DATA ANALYTICS WITH R, EXCEL AND TABLAEU

ASSIGNMENT 8.2 ANSWERS

By ASHISH S SHANBHAG

ashishshanbhag108@gmail.com

Question no:

5)

1a) Compute the measures of central tendency for salary and reduction which variable has highest center?

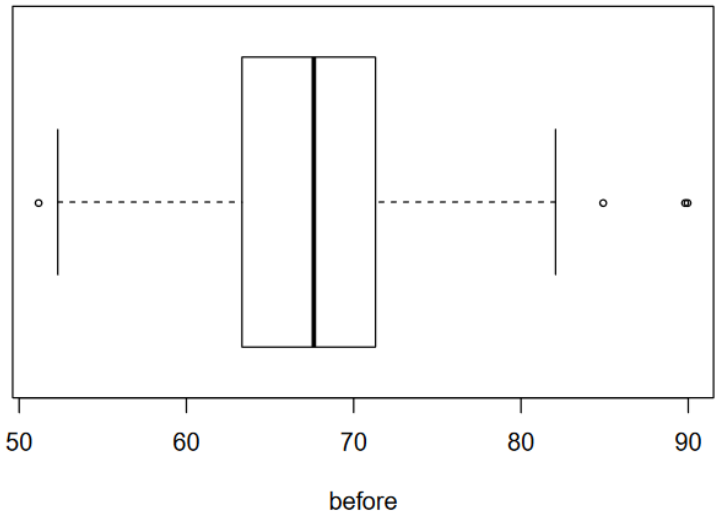
Ans > c(mean(before), median(before))

[1] 67.36338 67.61824
> c(mean(after), median(after))
[1] 66.85215 66.93608

The idea is to look at the two measures and compare them to make a decision. In a nice world, both the mean and median of one variable will be larger than the other which sends a nice message. If We get a mixed message, then we should look for other information, such as extreme values in one of the variables, which is one of the reasons for the next part of the problem

2) Which measure of center is more appropriate for before and after?

Ans: The boxplot of before is shown below.



We want to watch out for extreme values (shown as circles separated from the box) or large departures from symmetry. If the distribution is fairly symmetric then the mean and median should be approximately the same. But if the distribution is highly skewed with extreme values then we should be skeptical of the sample mean, and fall back to the median which

is resistant to extremes. By design, the before variable is set up to have a fairly symmetric distribution.
A boxplot of after is shown next.