

For comments, corrections, etc...Please contact Ahnaf Abbas: ahnaf@uaemath.com

This is an open source document. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, <http://www.gnu.org/copyleft/fdl.html> Version 1.2 or any later version published by the Free Software Foundation.

International Institute for Technology and Management



Unit 76: Management Mathematics

Handout #13a

Exploratory Data Analysis

Box Plots (also called Box-and-Whisker Plot)

Offers same information as histogram, but in more condensed form.

Revision :

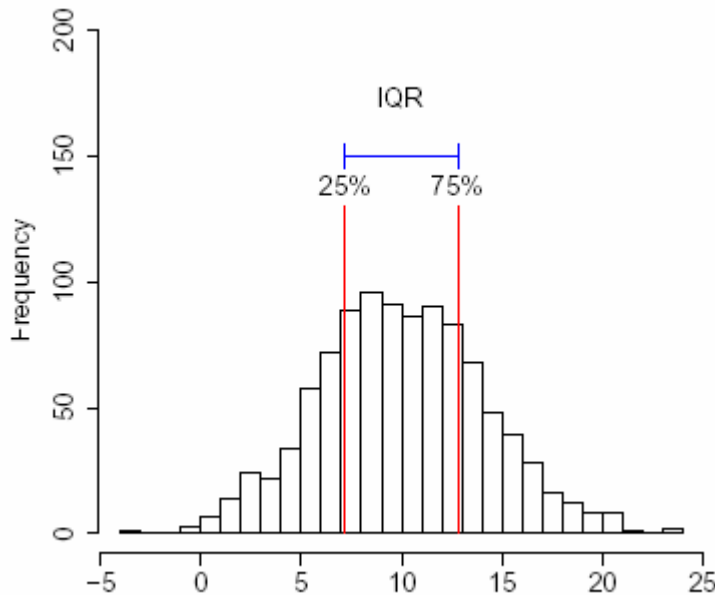
-1st Quartile : the 25% point of the dataset, its location is defined to be the $(N+1)/4$ th data point.

-3rd Quartile : the 75% point of the dataset, its location is defined to be the $3(N+1)/4$ th data point.

-Inter Quartile Range: $IQR = 3^{\text{rd}} \text{ quartile} - 1^{\text{st}} \text{ quartile}$

-Semi Inter Quartile Range: SIQR = IQR/2

Remark : The 2nd quartile is the 50% point of the dataset i.e. the **median**.



Example: Consider the set of 11 numbers :

15,18,92,33,10,51,73,86,34,80,36

Find the 1st, the 3rd quartiles and the inter quartile range and SIQR.

Rearrange the numbers : 10,15,18,33,34,36,51,73,80,86,92

1st quartile = $11 + 1/4 = 3^{\text{rd}}$ data point = 18

3rd quartile = $3(11+1)/4 = 9^{\text{th}}$ data point = 80

$IQR = 80 - 18 = 62$

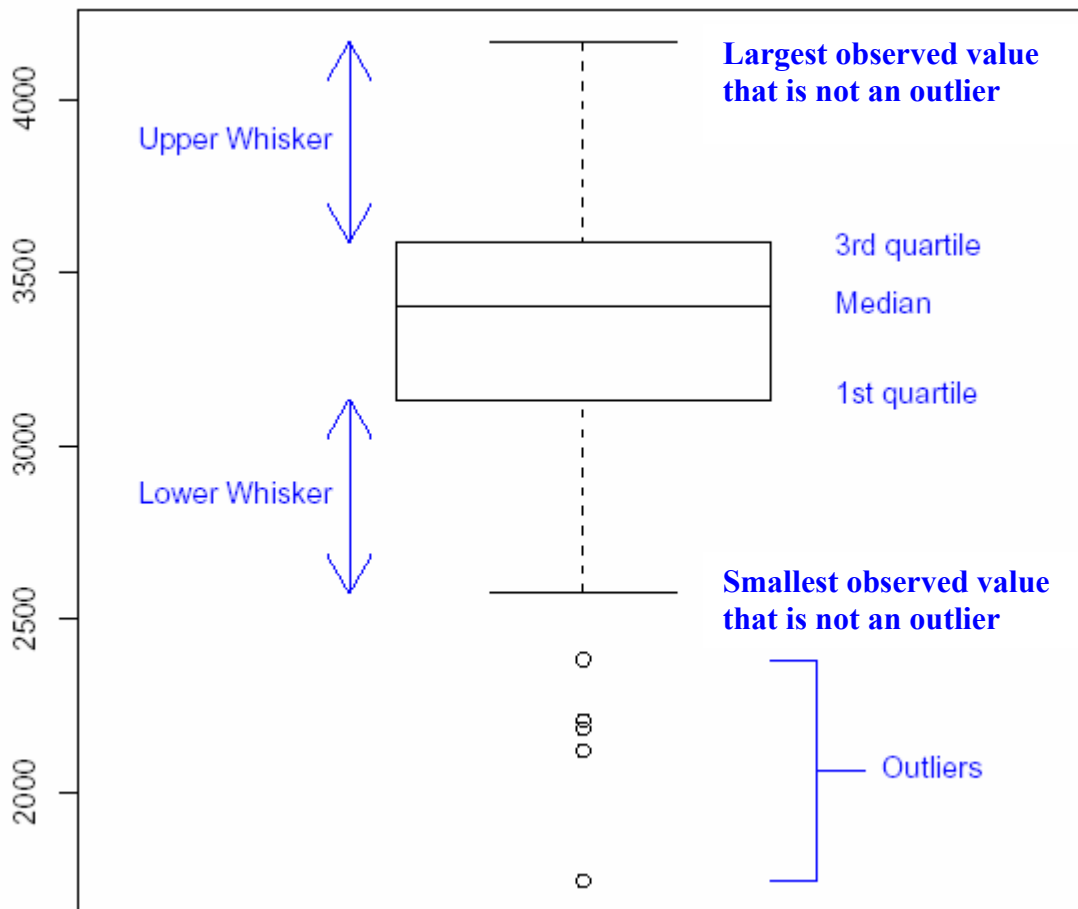
$SIQR = IQR/2 = 62/2 = 31$

For comments, corrections, etc...Please contact Ahnaf Abbas: ahnaf@uaemath.com

This is an open source document. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, <http://www.gnu.org/copyleft/fdl.html> Version 1.2 or any later version published by the Free Software Foundation.

A Box-Plot consists of three main parts :

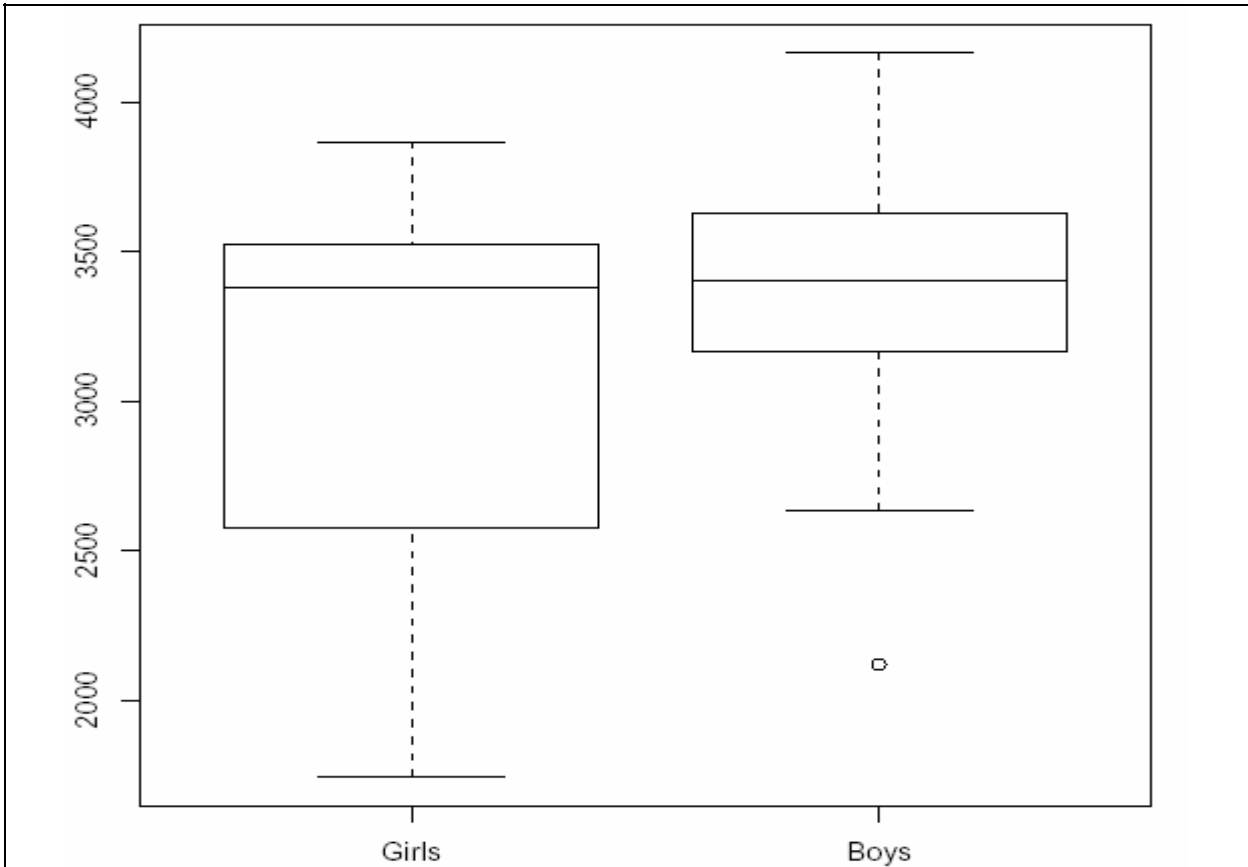
1. A box that covers the middle 50% of data. The edges of the box are the 1st and the 3rd quartiles. A line is drawn in the box at the median value.
2. Whiskers that extend out from the box to indicate how far the data extend either side of the box (largest and smallest observed values). The whiskers should extend **no** further than *1.5 times the length of the box* i.e. **the maximum length of the whisker is $1.5 \times \text{IQR}$**
3. All points that lie outside the whiskers are plotted individually as outlying observations.



Plotting box plots of measurements in different groups side by side can be illustrative. For example the following shows the box plots of birth weight for each gender side by side and indicates that distributions have quite different shapes:

For comments, corrections, etc...Please contact Ahnaf Abbas: ahnaf@uaemath.com

This is an open source document. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, <http://www.gnu.org/copyleft/fdl.html> Version 1.2 or any later version published by the Free Software Foundation.



Scatter Plots are useful when we wish to visualize the relationship between two measurement variables. To draw a scatter plot : we assign one variable to each axis and plot one point for each pair of measurements.

