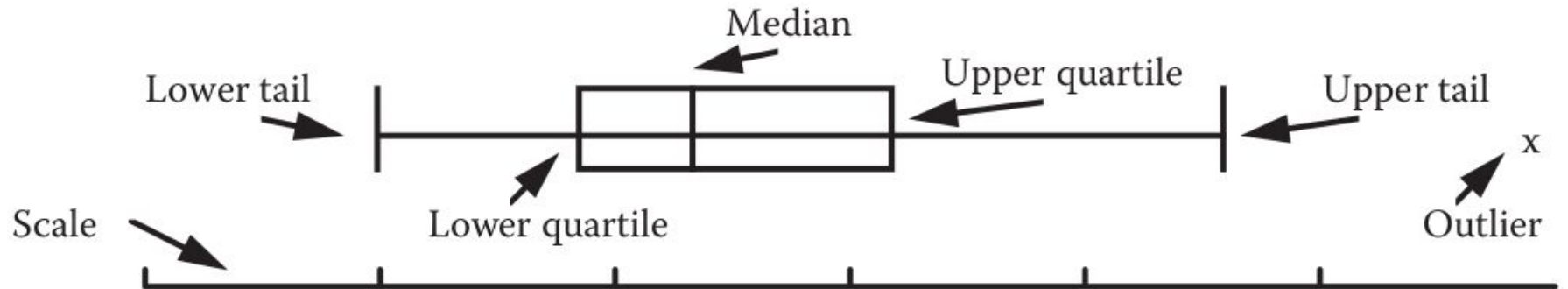


# Exploring Relationships using Plots

# Box Plot

- Software measurement datasets are often not normally distributed, and the measurements may not be on a ratio scale
- Hence, median and quartiles to define the central location and spread of the component values, rather than the more usual mean and variance
- Used to check the shape/ skewness of the distribution, central tendency and variability

# Box Plot Components



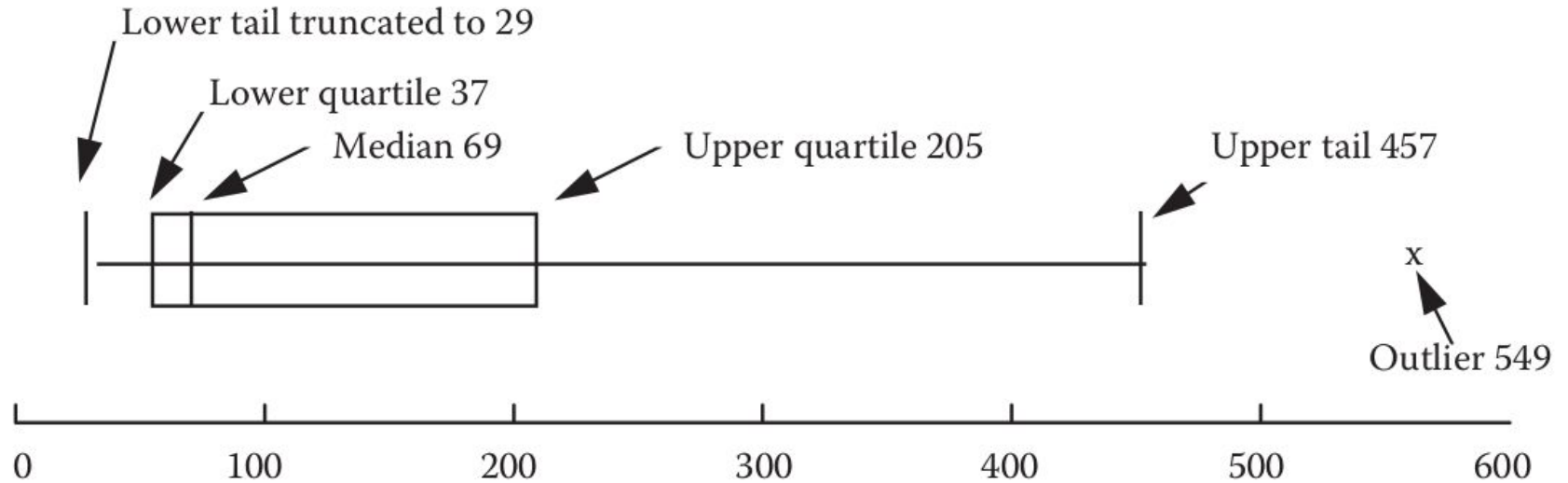
# Box Plot Components

- *Upper quartile*  $u$  is the median of the values that are more than  $m$ , and the *lower quartile*  $l$  is the median of the values that are less than  $m$
- $l$ ,  $m$ , and  $u$  split the dataset into four parts
- Interquartile Range (IQR),  $d = u - l$
- Theoretical upper tail  $u + 1.5d$  and lower tail  $l - 1.5d$
- Any values outside these points are outliers

# Box Plot : Example

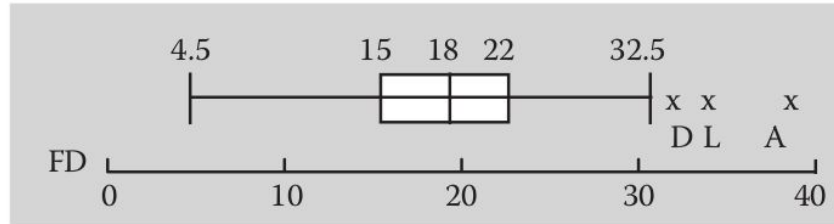
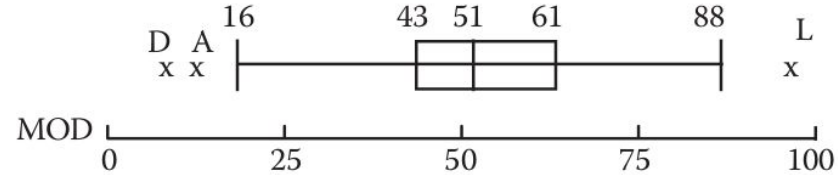
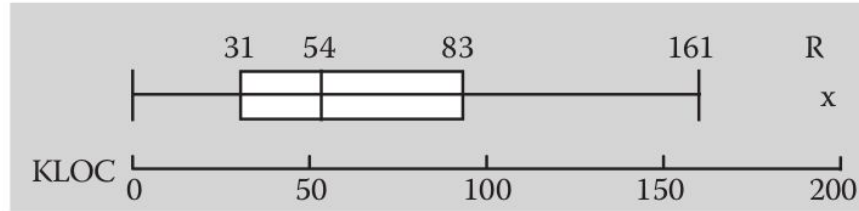
Module Size	Module Fan-Out	Module Fan-In	Module Control Flow Paths	Module Faults
29	4	1	4	0
29	4	1	4	2
32	2	2	2	1
33	3	27	4	1
37	7	18	16	1
41	7	1	14	4
55	1	1	12	2
64	6	1	14	0
69	3	1	8	1
101	4	4	12	5
120	3	10	22	6
164	14	10	221	11
205	5	1	59	11
232	4	17	46	11
236	9	1	38	12
270	9	1	80	17
549	11	2	124	16

# Box Plot : Example

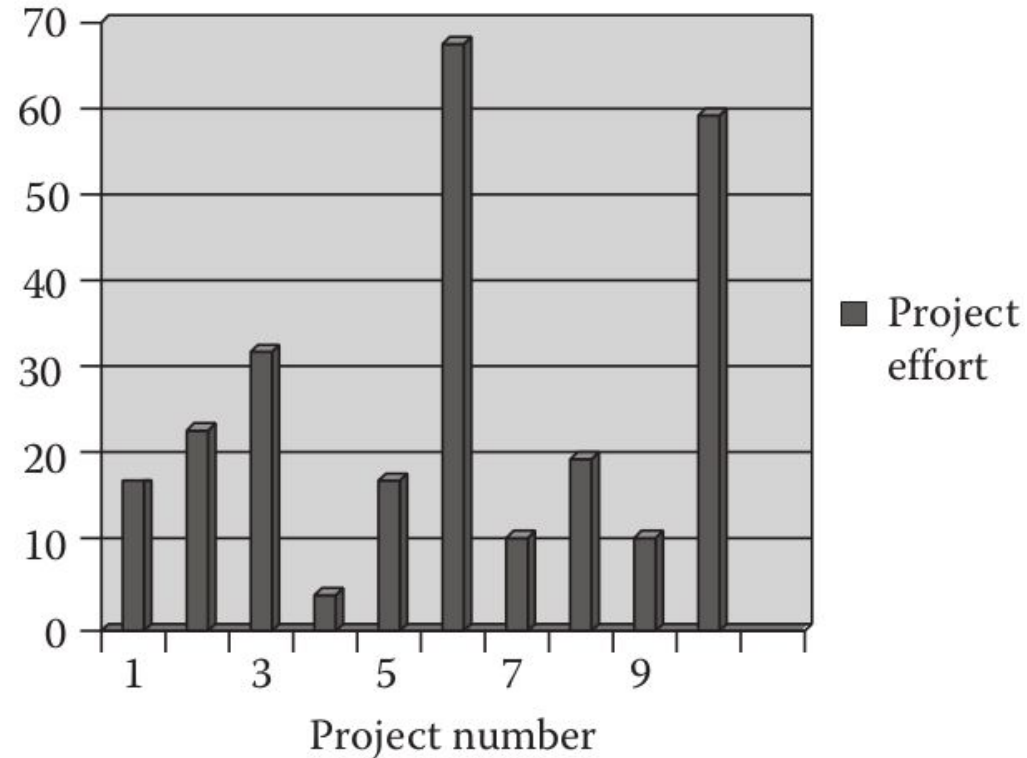


# Box Plot : Example

System	KLOC	MOD	FD
A	10	15	36
B	23	43	22
C	26	61	15
D	31	10	33
E	31	43	15
F	40	57	13
G	47	58	22
H	52	65	16
I	54	50	15
J	67	60	18
K	70	50	10
L	75	96	34
M	83	51	16
N	83	61	18
P	100	32	12
Q	110	78	20
R	200	48	21



## Bar Chart : Example





# Control Chart

- Helps to see when your data are within acceptable bounds
- By watching the data trends over time, you can decide whether to take action to prevent problems before they occur

# Control Charts : Example

Component Number	Preparation Hours/ Inspection Hours
1	1.5
2	2.4
3	2.2
4	1.0
5	1.5
6	1.3
7	1.0
Mean	1.6
Standard deviation	0.5
Upper control limit (UCL)	2.6
Lower control limit (LCL)	0.4

# Control Charts : Example

