Data Analytics

D Samanta

Topic 2:

Earthquake with 8086 observations

Any discrepancies: If timestamp was to be taken only in seconds or minutes so that time observation graph could also be drawn and other experiments could have been done. Otherwise found no discrepancy in the data given in respect to missing datas or any unlabelled features

Five Point Summary:

To calculate the minimum, first quartile, second quartile(median), third quartile and maximum value.

Five Point Summary does’nt exists for Time stamp

For other features namely,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Depth(Km) | Latitude | Longitude | Magnitude |
| Minimum | 0.2 | 41.58 | 12.54 | 2 |
| First Quartile | 8.9 | 42.75 | 122.10 | 2.1 |
| Median | 9.9 | 42.82 | 12.16 | 2.3 |
| Third Quartile | 10.8 | 42.91 | 12.22 | 2.6 |
| Maximum | 47.9 | 43.93 | 12.85 | 6.5 |

As per ITR calculations,

IQR = Q3-Q1

Data points which are above 1.5\*IQR than Q3 and less than Q1 are considered outliers and thus removed.

Using the ITR calculation which are readily built in the package “box plot”, we can draw the boxplot without outliers.

Although we can know about the outliers through “boxplot.stats”