Ex - STA-3163 - Box Plot

\[ Y = \text{AQI (Air Quality Index)} \]

2001 data from 15 cities \((n=15)\)

Data set

\[ Y: 24, 3, 5, 7, 11, 18, 34, 30, 12, 5, 27 \]

1) List data values from L to H.

\[3, 4, 6, 8, 12, 14, 19, 24, 27, 28, 31, 33, 34, 50, 81\]

Location for median \(Q_2 = \frac{n+1}{2} = 8^{th}\) (position data value)

\[ M = Q_2 = 24\]

Lower half (below 8th position) has 7 data values

location for \(Q_1 = \frac{7+1}{2} = 4^{th}\) data value

\[ Q_1 = 8\]

Similarly upper half has also 7 data values

and \(Q_3's\) location is 4th from top,

\[ Q_3 = 33\]

IQR = \(Q_3 - Q_1 = 33 - 8 = 25\)

Scale for Box plot (5 numbers min=3, max=81)

\[ \begin{array}{cccccccc}
0 & 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80 \\
& \text{Box Plot} & & & & & & & \\
\end{array} \]

\(Q_1, Q_2, Q_3\)
Distn is skewed to R

IQR = 1.5 \times 25 = 37.5

To determine outliers (Mild)

y > Q_3 + 1.5 \times IQR \iff y > 33 + 37.5 = 70.5 = upper bound

and

y < Q_1 - 1.5 \times IQR \iff y < 8 - 37.5 = -29.5 = lower bound

Since lower bound is negative we don't have any outliers on the lower side of data
On the upper side
max value = 81 > 70.5 \Rightarrow 81 is an outlier
next higher value = 50 which is < 70.5
This is not an outlier
There is only one outlier = 81 (mild)

For finding extreme outliers

y < Q_1 - 3 \times IQR (lower bound) = 8 - 75 = -67

and

y > Q_3 + 3 \times IQR (upper bound) = 33 + 75 = 108

No data y value is < -67 \Rightarrow no outlier on lower side

max y = 81 which is not > upper bound 108
\Rightarrow no extreme outlier on the upper side

This data set has no extreme outliers.