Find the weird value ?

5, –2, 20, 3, 4, –3, 15, 8, 10, 12, 6, 60, 1

Here we will talk about need for quartiles and how do we calculate quartiles.

Why quartiles matter ?

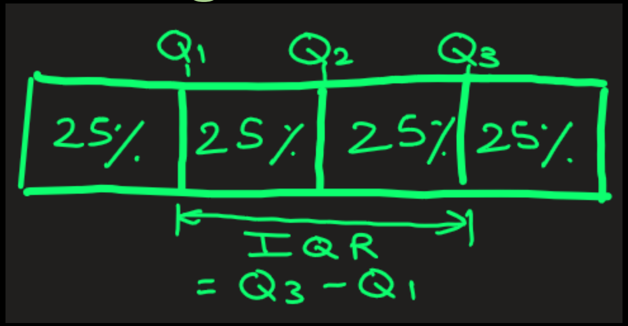
1. They Break Data Into Manageable Chunks

20, 3, 9, 9, 88, 43, 9, 5, 95, 7, 99, 3, 66, 1903

Quartiles divide data into four equal parts:

- 25% of the data below Q1
- 25% between Q1 and Q2
- 25% between Q2 and Q3
- 25% above Q3

This gives a **structured view of distribution** rather than just looking at all the numbers at once.



Why quartiles matter ?

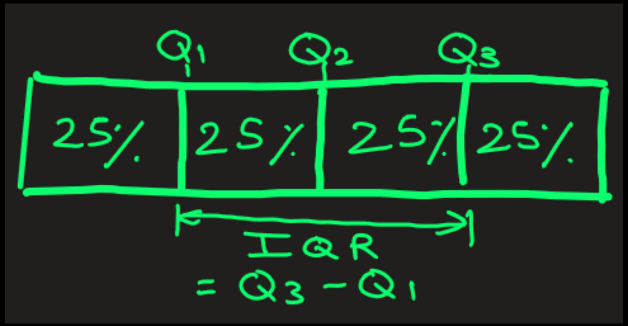
2. They Show the Spread of Data

- The **range** (max – min) can be misleading if there are outliers.
- Quartiles focus on the **middle portion of data** where most values lie.

Example: If only one student scores 0 in a test while rest score 70–90,

0, 71, 77, 70, 89, 73, 87, 84, 90, 89

the range exaggerates variability, but quartiles give a **fairer** picture.



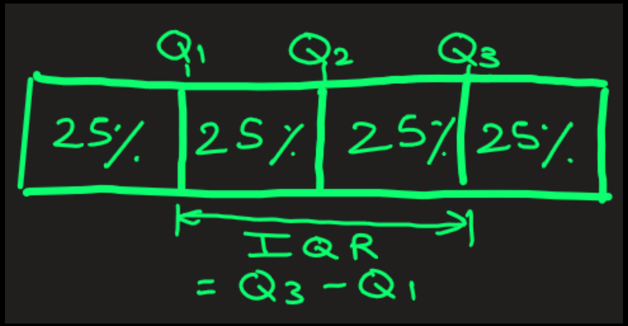
Why quartiles matter ?

3. They Help Spot Outliers

- By calculating the **Interquartile Range IQR (Q3 – Q1)**, we can set boundaries to detect unusual values AKA outliers.

2, 3, 99, 89, 88, 89, 90, 98, 95, 87, 99, 93, 89, 1903

Useful in finance (fraud detection), medicine (abnormal lab results), and business (detecting odd customer behavior).

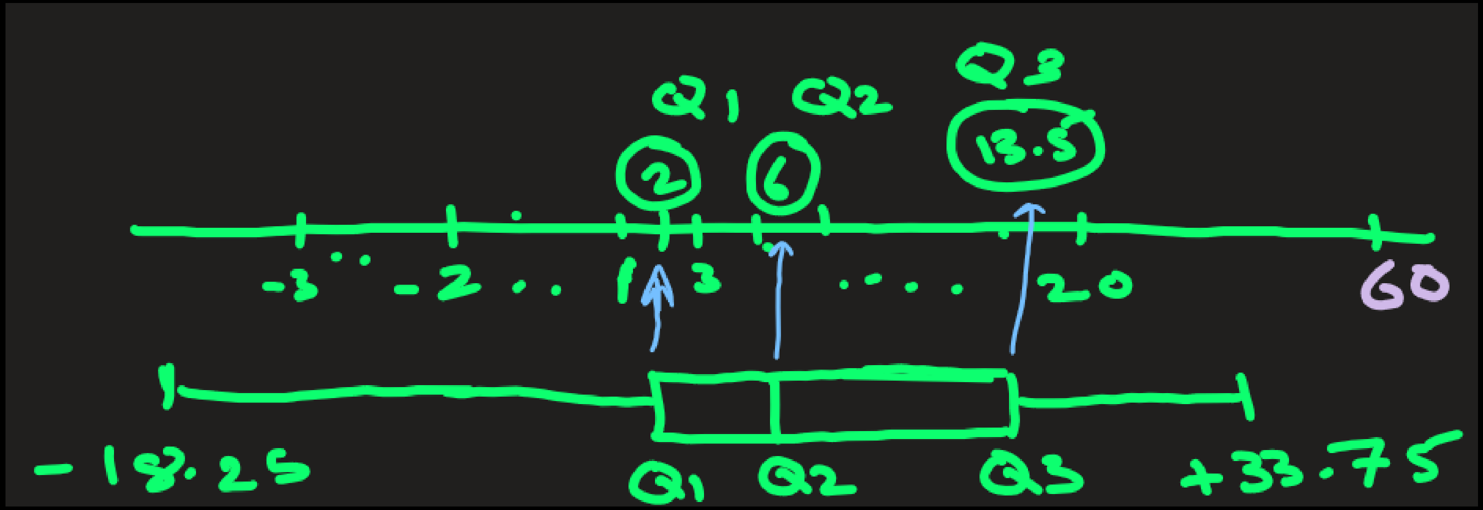


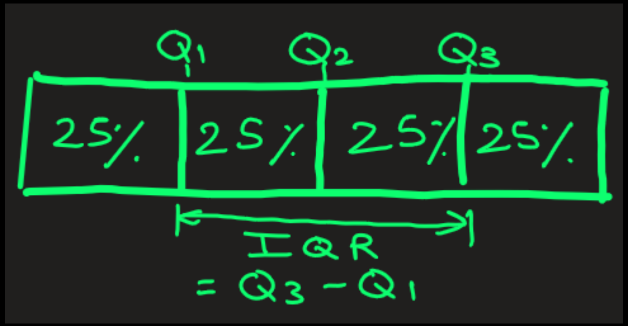
Why quartiles matter ?

4. They Feed into Visualization (Box Plots)

- Box-and-whisker plots use quartiles to summarize an entire dataset visually.

This makes it easy to **compare multiple datasets** (e.g., salaries of different professions).



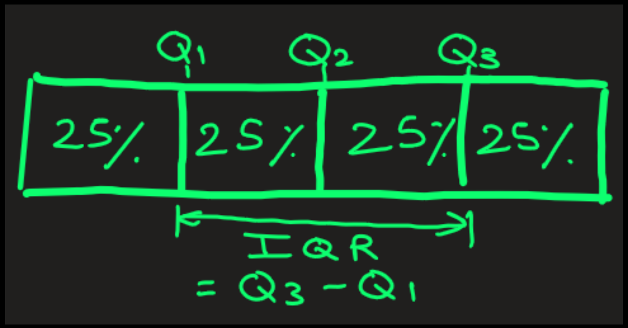


Why quartiles matter ?

5. They Are Robust to Skewed Data

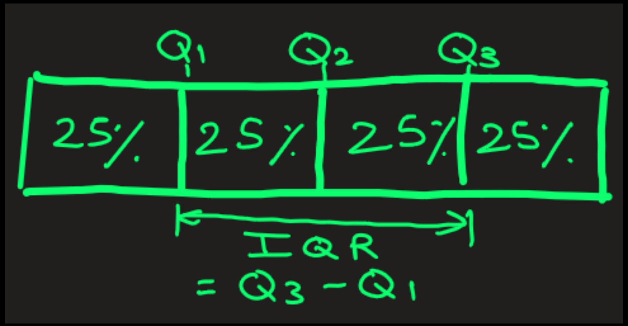
- Unlike mean, which is distorted by extreme values, quartiles (like the median) are **resistant to outliers**.

2, 3, 99, 89, 88, 89, 90, 98, 95, 87, 99, 93, 89, 1903



Quartiles matter because they

- tell us the center of the data
- show us how the data is spread,
- where most values lie, and
- whether something looks unusual.



How do you find Q1, Q2, Q3, IQR ?

Step 1: Arrange the Data

First, we always **arrange the data in ascending order** — from smallest to largest.
Quartiles only make sense when the data is sorted.

Example dataset, say temperature:

5, −2, 20, 3, 4, −3, 15, 8, 10, 12, 6, **60,** 1

After sorting:

-3, -2, 1, 3, 4, 5, 6, 8, 10, 12, 15, 20, **60**

Step 2: Find Q2 (the Median)

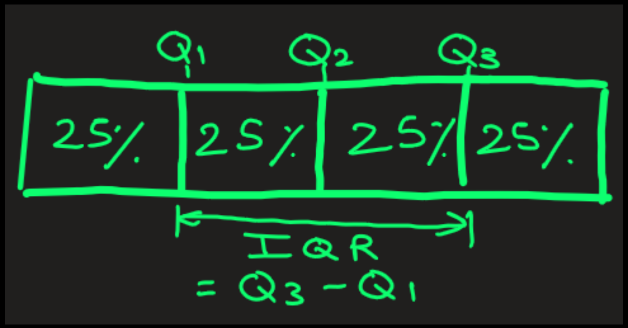
Q2 is the **median** of the entire dataset— the middle value that splits the data into two equal halves.

- If the number of data points is **odd**, median = middle value.
- If it's **even**, median = average of the two middle values.

Example:

Here we have 13 numbers -> Odd.
So, Q2 = middle value = 7^{th} number = 6

-3, -2, 1, 3, 4, 5, **6**, 8, 10, 12, 15, 20, **60**



Interpretation: **50% of data is below 6**

How do you find Q1, Q2, Q3, IQR ?

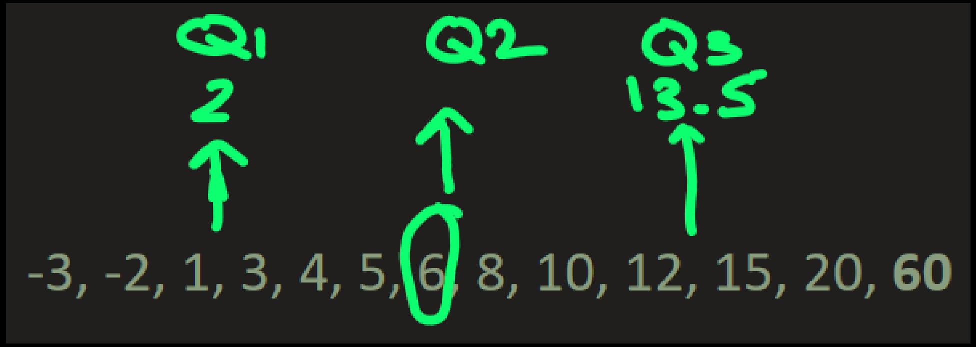
Step 3: Find Q1 (the Lower Quartile)

Q1 is the **median of the lower half** of the data — the 25th percentile

Lower half (below Q2):-3, -2, 1, 3, 4, 5

Q1, median of this part = (1 + 3)/2 = **2**

Interpretation: **25% of data is below 2**



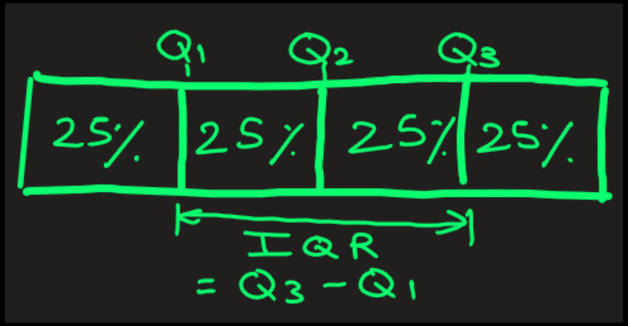
Step 4: Find Q3 (the Upper Quartile)

Q3 is the **median of the upper half** of the data — the 75th percentile.

Upper half (above Q2): 8, 10, 12, 15, 20, **60**

Q3, median of this part = (12 + 15)/2 = **13.5**

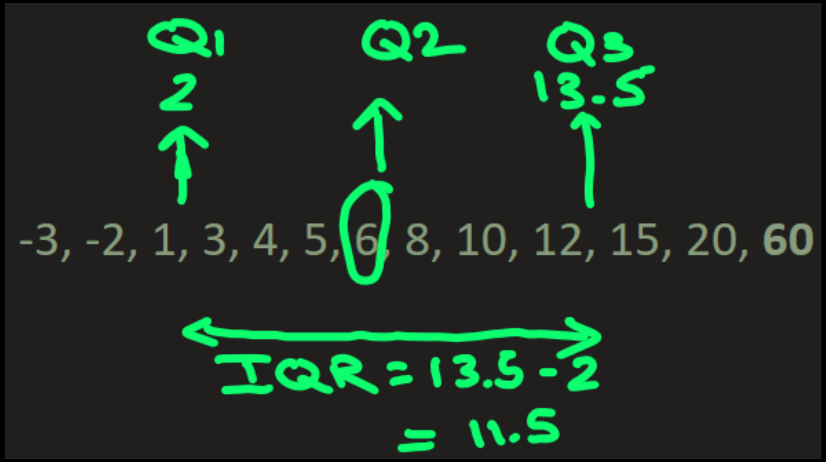
Interpretation: **75% of data is below 13.5**



How do you find Q1, Q2, Q3, IQR ?

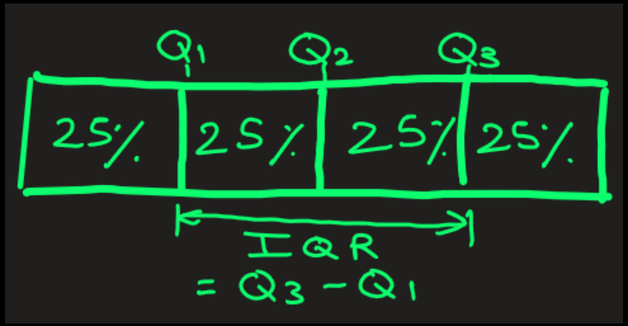
Step5) Find Interquartile Range IQR
IQR measure the spread of the middle 50% of data.
Interquartile Range (IQR) = Q3 – Q1 = 13.5 – 2 = **11.5**

Interpretation: **Central 50% of the data values lie within an interval of 11.5 units.**



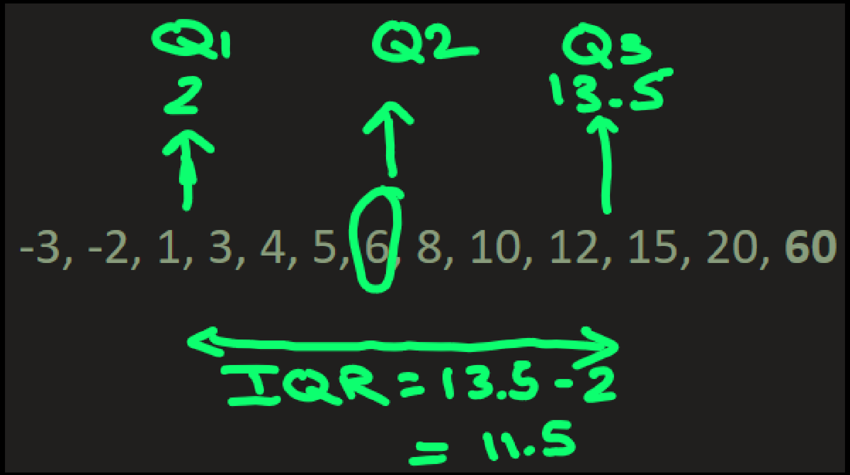
What IQR *Tells Us* ?

A **small IQR** → data is **closely packed** (less variability).
A **large IQR** → data is **spread out** (more variability).
It focuses on the middle part of the data.
So, it **ignores extreme values or outliers**. In above case, it ignored 60.



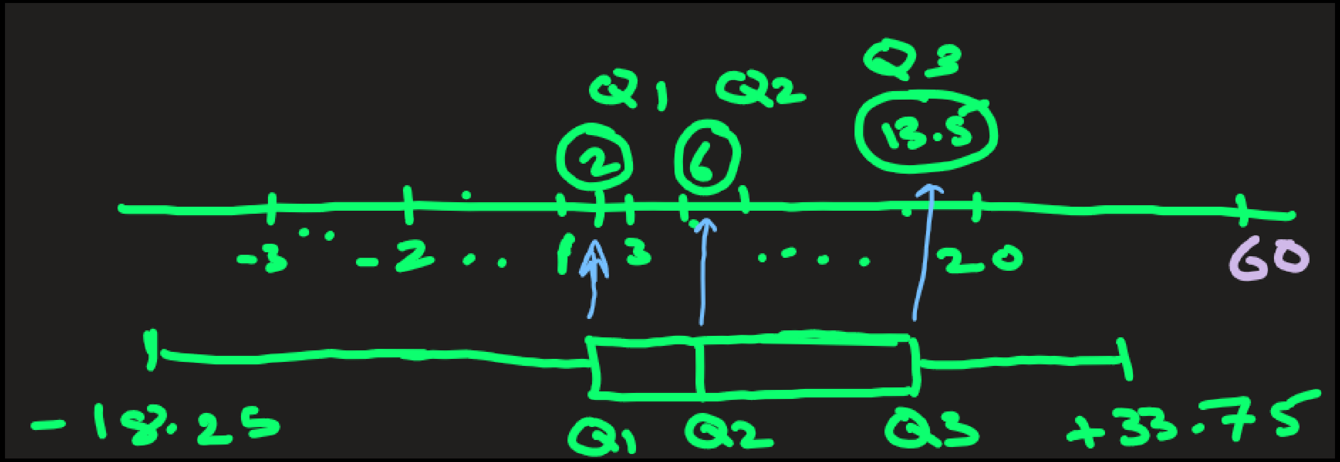
Use case: Outlier Detection

Values are considered **outliers** if they lie
below Q1 – 1.5×IQR , or
above Q3 + 1.5×IQR
That's why IQR is also used in **box plots** to identify unusual data points.

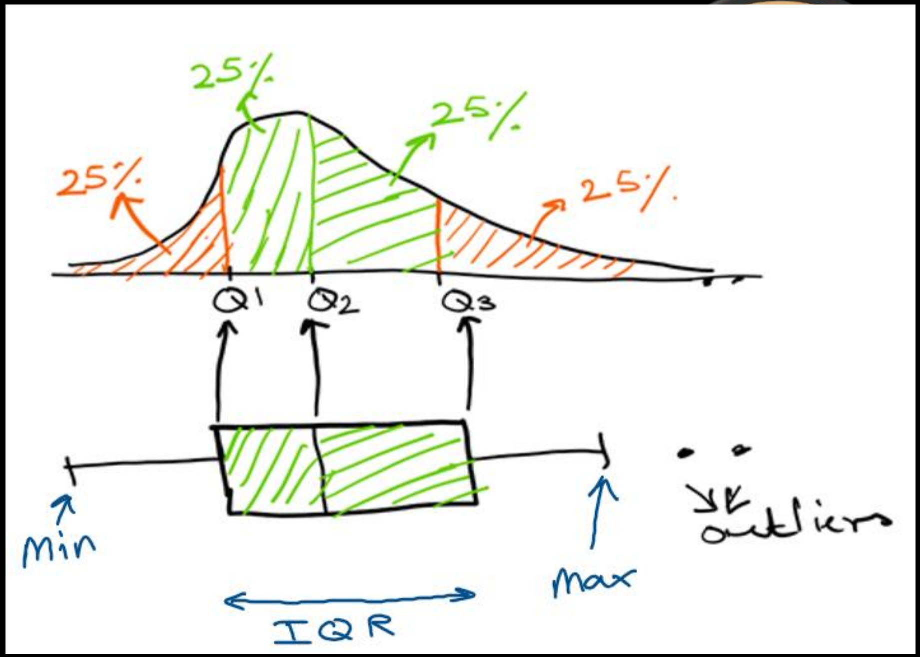


In above case,
lower boundary = Q1 – 1.5×IQR
= 2 – 1.5x13.5 = -18.25
upper boundary = Q3 + 1.5×IQR
= 13.5 + 1.5x13.5 = +33.75

The value 60 lies outside the range: It is an outlier.



STOP



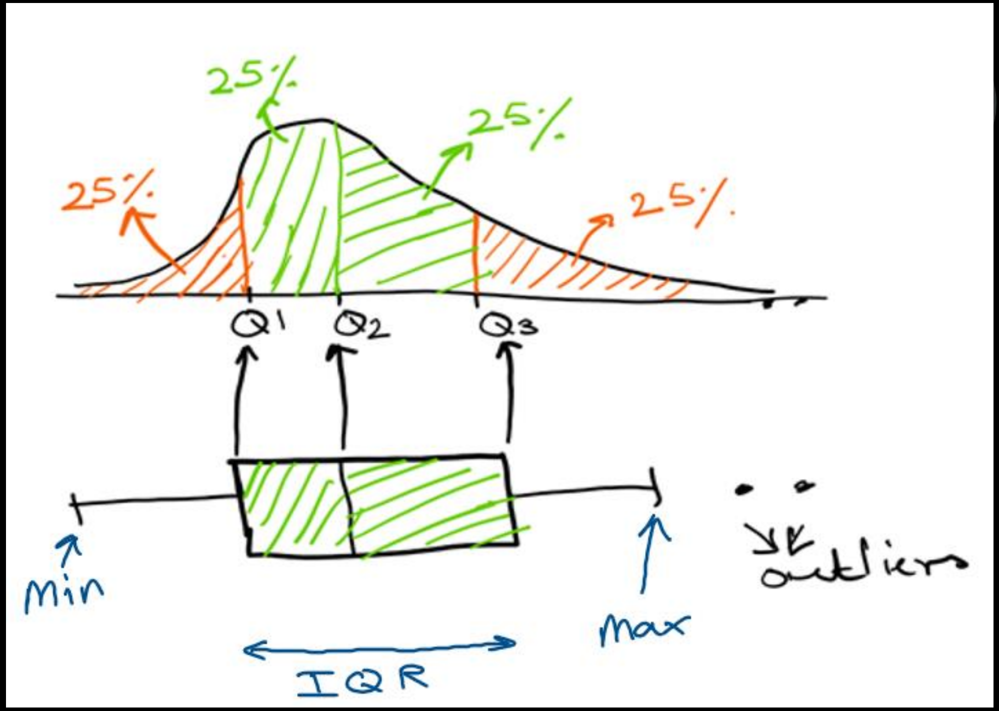
25%
25%
25%
25%
Q1
Q2
Q3
outliers
Min
Max
IQR

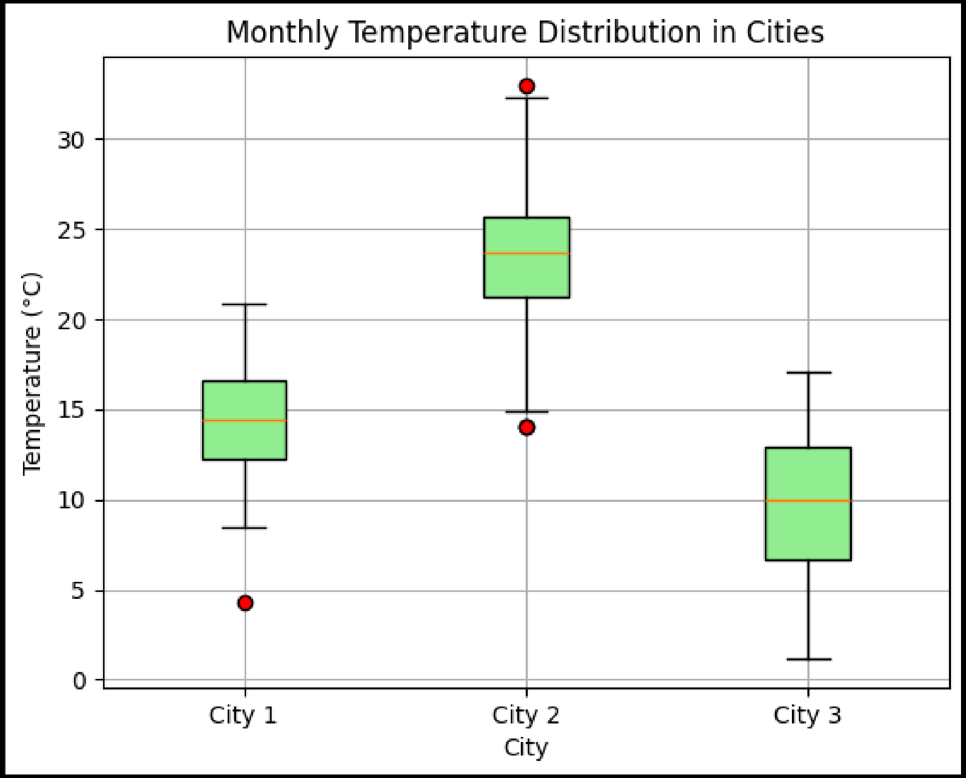
Box plots

- Show **distributions of numeric data** values, especially when you want to compare them **between multiple groups**.
- Provide visuals on **data's symmetry, skew, variance, and outliers**.

4, 3, 5, 2, 4, 3, 6, 7, 8, 3, 5, 2, 3, 4, **78**, 3, 2,-**30**, 3, 4, 5, 3, 2: here -30 and 78 seem outliers

- Easy to see where the main bulk of the data is, and make that comparison between different groups.
- 25% of data falls below Q1 (quartiles)
- 50% of data falls below Q2
- 75% of data falls below Q3





For city1:

- most of temp is between 13 to 16. There is one outlier, temp = 4
- Q1 = 13. So 25% of temp data falls below 13.
- Q2 =14. So 50% of temp data falls below 14
- Q3 = 17.