Why 1.5 Times the IQR?

The choice of 1.5 times the IQR to define outliers is based on a balance between identifying unusually large or small data points while avoiding too many false positives. Here's the reasoning:

- **Spread and Distribution:** In a normal distribution, the IQR captures the central 50% of the data, with 25% of the data falling below Q1 and 25% above Q3. The multiplier of 1.5 extends this range to capture data that is reasonably expected but still potentially unusual.
- **Identifying Outliers:** The 1.5 multiplier is chosen because it creates a range that includes about 99.3% of the data in a normal distribution (assuming a symmetric distribution without long tails). This means that any data points outside this range are likely to be true outliers—those that don't conform to the pattern of the rest of the data.
- **Flexibility Across Distributions:** The factor of 1.5 works well for various data distributions (not just normal distributions). It's a rule of thumb that is robust enough to apply to a wide range of datasets, which is why it's commonly used.