Precise summary of Fruit Quality Data Analysis (Orange)

This report analyses the quality of 24 distinct varieties of oranges using a dataset from Kaggle. Key parameters considered include weight, size, Brix (sweetness), pH (acidity), softness, harvest time, and maturity. Using Python libraries (pandas, matplotlib, seaborn, NumPy), various visualizations were applied:

- **Histograms** showed the distribution and range of quality parameters.
- **Pie Chart** depicted the proportion of five major orange colour types.
- **KDE Plot** indicated that sweeter, high-quality oranges have concentrated sweetness in a narrower range.
- **Correlation Heatmap** revealed:
    - A weak positive correlation between size, weight, and pH.
    - A weak negative correlation between size and Brix (larger oranges tend to be more acidic and less sweet).
    - Moderate negative correlation between pH and harvest time, softness, and quality (fresher, firmer, and less acidic oranges are rated higher).
- **Box Plot** illustrated blemish effects:
    - Minor blemishes Y (Minor) correlated with sweeter, firmer oranges.
    - Mould spots Y (Mould Spot) indicated lower sweetness and higher softness.
    - Blemish-free (N) oranges had better overall quality and slightly higher acidity.
- **Violin Plot** confirmed that higher sweetness is linked to higher quality ratings.

Conclusion:

High-quality oranges are typically blemish-free, sweeter, firmer, fresher, and less acidic. In contrast, larger oranges are generally more acidic and less sweet.