

Histogram

Overview

A histogram is a type of bar graph displaying the frequency distribution of measurements grouped into user-defined ranges or bins. Histograms differ from other types of bar chart, which can display groups with qualities or attributes which have no particular order. The data for a histogram has an intrinsic order which is reflected in the chart.

The information can be collected with a checklist initially and then displayed as a histogram to highlight the most frequent category and to understand the variability in your data i.e. location, spread, shape and patterns of data.



When to use a histogram

A histogram is a good tool for understanding your system and determining areas of focus for improvement. Use a histogram if you want to understand the distribution of scaled data, for example:

- age
- workload
- time
- satisfaction
- money
- ratings

This type of data is known as continuous data. The histogram will give you an idea of the location, shape and spread of the data. The shape and width of the distribution helps identify the cause of problems in a process.

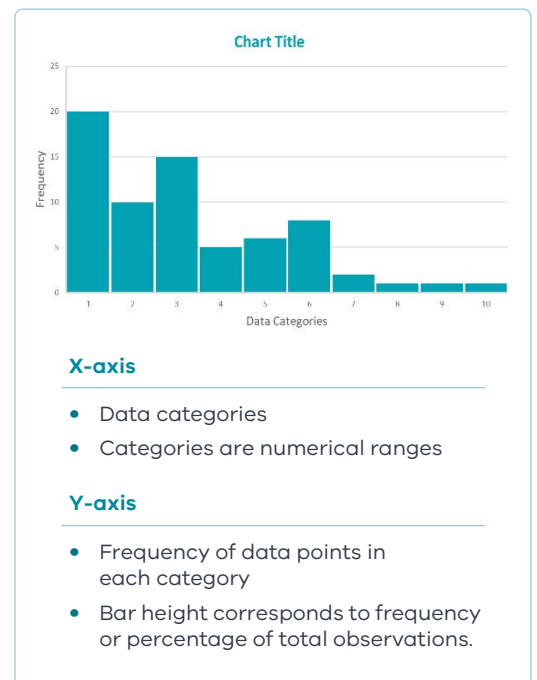


Figure 1: Histogram Structure

How to create a histogram

1. Gather data on a process you are interested in.
2. Split the data points into several non-overlapping categories of equal width.
3. Decide how many categories (or bins) you want in your histogram. Table 1 suggests how many categories you could use based on the number of data points you have collected.
4. Identify how many of your data points fall into each class and record this information in a table. This will be your frequency table.
5. Plot your graph. Categories are plotted on the x-axis (horizontal) and their frequency is plotted on the y-axis (vertical). Transfer the data from your frequency table onto the graph, so that you have a vertical bar for each category.

Table 1: Data reference

Number of data points	Number of categories
Under 50	5–7
50–100	6–10
100–250	7–12
Over 250	10–20

Example histogram

This example histogram displays the number of falls that occurred over a 12-month period on an aged care ward by patient age range. This example demonstrates a negative/right skew indicating that the older patients on the aged care ward have the greatest number of falls, particularly for those aged between 85–89 years.

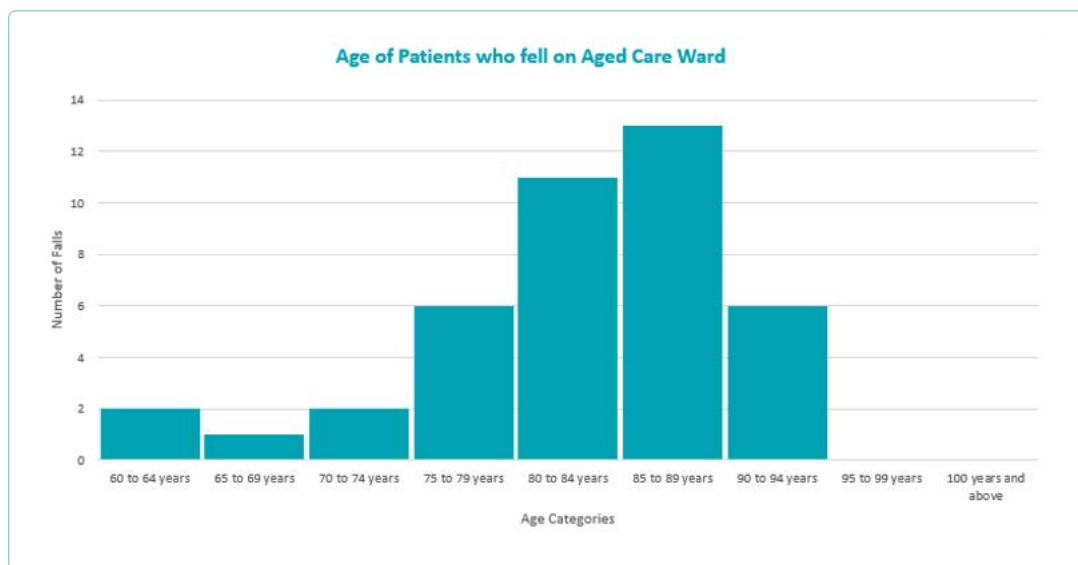


Figure 2: Histogram example - Falls on an aged care ward

Additional resources

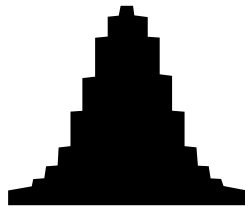
To learn more about Quality Improvement you can access the following resources:

- [SCV Quality Improvement Toolkit](#)
- [Institute for Healthcare Improvement website](#)
- [NSW Clinical Excellence Commission Quality Improvement Tools](#)

Content adapted with permission from the Institute for Healthcare Improvement (IHI) and the Clinical Excellence Commission (CEC)

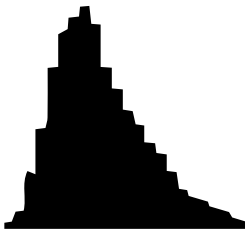
Interpreting a histogram

The shape of the histogram provides insights into how the data is distributed.



Normal distribution

A common pattern is the bell-shaped curve known as the 'normal distribution'. In a normal or 'typical' distribution, points are as likely to occur on one side of the average as on the other. If your data is 'normally distributed' the mean and median is the same.



Skewed distribution

The skewed distribution is asymmetrical because a natural limit prevents outcomes on one side. The distribution's peak is off centre toward the limit and a tail stretches away from it. Skews can be referred to as either right/positively skewed or left/negatively skewed. They suggest that a few data points are significantly higher/lower than the majority.



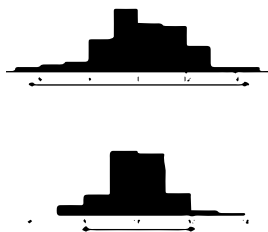
Double peaked or bimodal

The bimodal distribution looks like a two-humped camel. The outcomes of two processes with different distributions are combined in one set of data.



Symmetric/random/plateau distribution

A random distribution lacks an apparent pattern and has several peaks. In a random distribution histogram, it can be the case that different data properties were combined. Therefore, the data should be separated and analysed separately.



Spread or dispersion

The width of the distribution in the histogram indicates the spread or dispersion of the data. A wider distribution suggests higher variability, meaning the data points are spread out over a larger range of values. Conversely, a narrower distribution suggests lower variability, indicating that the data points are clustered closer together around the central tendency.



Outliers

Are data points that significantly deviate from the rest of the data. These can appear as isolated bars in the histogram, located far away from the bulk of the data. Outliers may indicate errors in data collection or measurement, or they may represent important but unusual observations.