How to Easily Choose Between a Pareto Chart and a Histogram

Authored by

stats writer

December 3, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Choose Between a Pareto Chart and a Histogram*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=103952

In the realm of statistical process control and data analysis, visualization tools are essential for transforming raw numbers into actionable insights. Among the most fundamental graphical representations are the Pareto Chart and the Histogram. While both charts utilize bars to display frequencies and share a superficial visual resemblance, their underlying purpose, structure, and application differ significantly. Understanding these distinctions is crucial for selecting the appropriate tool for effective decision-making.

The Pareto Chart, rooted in the 80/20 principle, is fundamentally designed for prioritization, helping analysts pinpoint the vital few factors that drive the majority of an outcome. Conversely, the Histogram is employed primarily to understand the underlying data distribution, revealing central tendencies, variance, and the presence of outliers within a continuous dataset.

This detailed comparison will clarify the unique characteristics of each visualization, demonstrate their construction through practical examples, and explain when to deploy a **Pareto Chart** versus a **Histogram** in professional contexts.

Defining the Pareto Chart: Focusing on Prioritization and the 80/20 Rule

A Pareto Chart is a specialized type of bar chart that incorporates both bars and a line graph. The bars display individual values (such as defect types or product categories) in descending order of frequency or magnitude, making it immediately clear which categories contribute most significantly to the total effect. This design is directly inspired by the Pareto Principle, or the 80/20 Rule, which posits that, for many outcomes, roughly 80% of consequences come from 20% of the causes.

The primary objective of using a **Pareto Chart** is to identify the "vital few" problems or factors that warrant the most attention and resource allocation. By visualizing data in this specific highest-to-lowest frequency format, managers and analysts can quickly prioritize quality improvement efforts, focusing on the handful of factors responsible for the largest share of issues, rather than spreading resources thinly across numerous minor causes.

Structurally, the chart features categories (which can be either quantitative data or qualitative data) along the x-axis, typically representing discrete events or classes. The left y-axis represents the frequency count for each category, while the secondary right y-axis represents the cumulative percentage. This dual-axis structure is essential, as the cumulative line allows the user to easily identify the point where, for instance, the top three categories account for 75% or 80% of the total counts.

Defining the Histogram: Understanding Continuous Data Distribution

The Histogram, unlike the Pareto Chart, is a visualization tool specifically designed to represent the data distribution of continuous, quantitative data. It groups continuous data into predefined ranges, known as bins or classes, and then displays the frequency (or count) of data points that fall into each bin using adjacent bars. The core function of a **Histogram** is not prioritization, but rather exploration of the data's shape and characteristics.

When interpreting a histogram, analysts are looking for several key statistical attributes. They assess the central tendency (where the peak lies), the spread or variability (how wide the base of the distribution is), and the symmetry or skewness of the data. Furthermore, histograms are invaluable for identifying potential **outliers**--data points far removed from the main body of the distribution--and determining if the process data conforms to a known distribution, such as a normal (bell-shaped) distribution.

A critical structural feature of the histogram is that the bars must touch, signifying that the data on the x-axis is continuous and that the bins represent a range without gaps. The x-axis always represents numerical intervals of quantitative data, such as time, weight, temperature, or length, displayed in ascending numerical order. The height of the bar (y-axis) indicates the frequency of observations within that specific numerical interval.

Core Structural Differences: Sorting, Axis Data Types, and Continuity

The most fundamental distinction between these two charts lies in how the data on the x-axis is categorized and ordered. A **Pareto Chart** is strictly ordered by frequency--meaning the bar height determines the sequence from left to right, regardless of the intrinsic value or label of the category. This ordering is intentional, driving immediate focus to the largest contributors.

Conversely, a **Histogram** is ordered strictly by the numerical sequence of the data intervals. The order of the bars is fixed from the smallest numerical range to the largest numerical range. The height of the bars will vary, but the position of the bins is non-negotiable, as altering the bin order would destroy the visualization of the underlying data distribution.

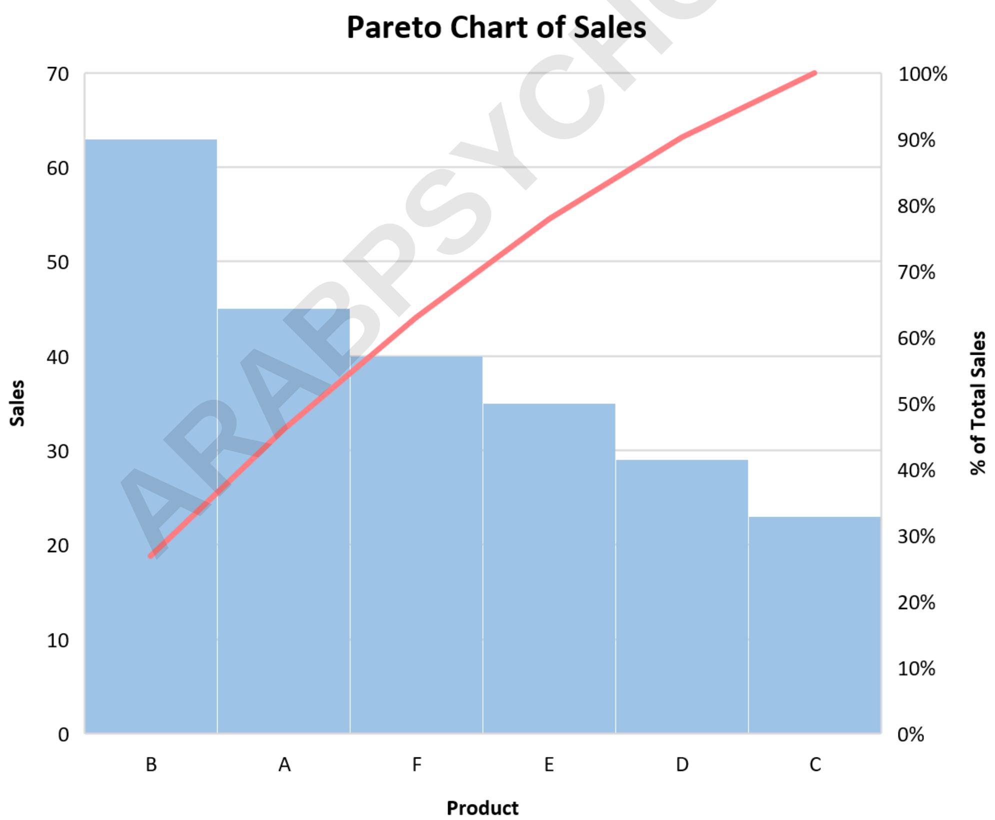
Furthermore, the type of data represented on the x-axis is different. Pareto charts commonly deal with discrete categories, often qualitative data (e.g., defect reasons, country names, product types), although they can handle discrete numerical data as well. Histograms, however, are exclusively used for continuous, high-volume quantitative data. This difference in data type also explains the bar separation: Pareto bars are usually separated to denote distinct, non-continuous categories, whereas histogram bars touch to represent contiguous numerical ranges.

Example 1: Creating and Interpreting a Pareto Chart

To properly illustrate the application of the Pareto Chart, consider a hypothetical scenario where a company is analyzing the sales contribution of its different product lines. We start with a raw dataset detailing the total sales figures associated with each discrete product category. This is the initial input data:

Product	Sales
A	45
B	63
C	23
D	29
E	35
F	40

The creation of the Pareto Chart requires the data to be sorted by sales volume (frequency) in descending order before plotting. This crucial step immediately reorganizes the data to emphasize the highest contributors. The resulting visualization clearly identifies the products that generate the majority of revenue:



Analysis of this chart confirms that **Product B** drives the highest total sales, followed by Product A, and then Product F. The descending order makes this prioritization intuitive. More importantly, the superimposed line graph represents the **cumulative percentage** of total sales, which is read against the secondary y-axis on the right. This line is the defining feature that allows for direct application of the 80/20 rule.

By tracking the cumulative line, we can derive powerful insights into resource allocation:

Product B alone accounts for approximately 25% of total sales.
The combination of **Products B and A** reaches roughly 50% of total sales.
The top three products (B, A, and F) contribute approximately 65% of the total revenue, indicating that efforts focused on these three categories will yield the greatest impact on overall sales performance.

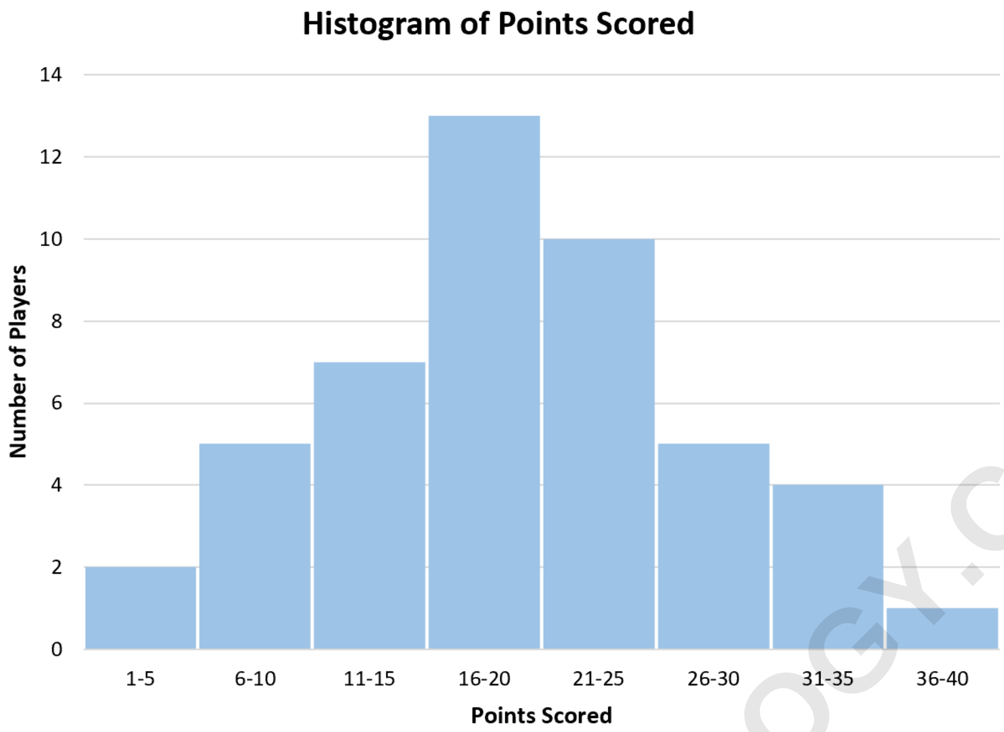
This analysis demonstrates how the Pareto Chart translates raw frequency into a tool for strategic prioritization, highlighting the few items (products) that require the most attention.

Example 2: Creating and Interpreting a Histogram

In contrast to prioritization charts, the Histogram is used when we need to understand the characteristics of continuous data. Let us consider a dataset representing the total number of points scored by a large group of basketball players throughout a season. This quantitative data must first be organized into contiguous numerical bins:

Points Range	Number of Players
1-5	2
6-10	5
11-15	7
16-20	13
21-25	10
26-30	5
31-35	4
36-40	1

When constructing the histogram, the points scored are grouped into fixed-width bins (e.g., 0-5 points, 6-10 points, etc.), and the height of the bar represents the count (frequency) of players falling within that score range. The resulting visualization shows the overall data distribution:



Interpreting this histogram provides immediate insight into the performance patterns of the player population. We can quickly ascertain the central tendency: the majority of players cluster in the central ranges, specifically between 16 and 25 points. The shape of the distribution reveals that it is roughly **symmetrical** and "bell-shaped," which often suggests the data is approximately normally distributed.

Crucially, the histogram allows for the visual detection of variance and potential skewness. In this basketball example, the bars at the extreme ends--those scoring less than 5 points or 36 points or more--are very low, indicating that extreme high or low scores are rare within this population. This visual confirmation of the distribution's bounds helps confirm the consistency and performance spread of the players.

Statistical Utility: When to Choose Which Chart

Choosing between a Pareto Chart and a Histogram depends entirely on the analytical question being posed. If the goal is to drive improvement actions by focusing resources, the Pareto Chart is the superior choice. Its inherent structure of descending frequency forces the analyst to prioritize causes based on their impact. This makes it a foundational tool in quality management methodologies like Six Sigma, where identifying and eliminating the largest source of defects is paramount.

If, however, the goal is to understand the inherent variation, stability, or conformity of a process, the Histogram is indispensable. It answers questions about process capability, uniformity, and the shape of variation over time. For example, engineers use histograms to verify if manufactured product dimensions fall within specified tolerance limits or if machine performance data is consistent. It is a descriptive statistical tool that characterizes the entire dataset rather than prioritizing specific categories.

In summary, the Pareto Chart is inherently prescriptive, directing action based on magnitude. The Histogram is descriptive, providing diagnostic information about the underlying data behavior. Using the wrong chart can lead to serious analytical errors--for example, trying to apply Pareto analysis to continuous measurement data will obscure critical information about process stability and inherent variation.

Summary: Key Differences in Visualization Approach

While both chart types employ vertical bars to illustrate frequency, their design choices diverge significantly to serve their specialized functions. Below is a structured breakdown of the critical distinctions between the **Pareto Chart** and the **Histogram**:

Difference #1: Type of Data Analyzed

The Pareto Chart is highly flexible, capable of handling discrete categories derived from both quantitative data (e.g., specific counts) and qualitative data (e.g., categories of defects). Conversely, the **Histogram** is restricted exclusively to continuous quantitative data, where data points can fall anywhere within a given range.

Difference #2: Ordering Principle

The **Pareto Chart** utilizes a prioritization sort: bars are ordered strictly from the highest frequency count to the lowest frequency count. This is a deliberate manipulation of the data order to facilitate decision-making. The **Histogram** uses a numerical sort: bars are ordered by the increasing value of the bin ranges (smallest numerical range to largest). The bar height does not dictate the position on the x-axis.

Difference #3: Visualization of Cumulative Data

A defining feature of the Pareto Chart is the inclusion of a secondary line graph that plots the **cumulative frequency** across the categories. This feature is vital for applying the 80/20 principle. The Histogram does not include a cumulative line, focusing solely on the frequency count within each bin to reveal the shape of the data distribution.

Difference #4: Bar Continuity and Meaning

In a Pareto Chart, bars are typically separated, visually emphasizing that each bar represents a discrete, independent category (e.g., Product A, Product B). In a Histogram, the bars must touch, illustrating that the numerical ranges represented by the bins are continuous and contiguous along the x-axis.

Conclusion and Further Resources

Mastery of statistical visualization requires recognizing that every chart is a purpose-built tool. The Pareto Chart is the instrument of choice for quality control and root cause analysis, demanding focus on the largest problems first. The Histogram serves as the diagnostic tool for process capability, variation analysis, and understanding the inherent structure of continuous data. By selecting the correct visualization method, analysts ensure that their data insights are both accurate and maximally impactful.

For those seeking to implement these charts in analytical environments, the following resources provide detailed technical guides:

How to Create a Pareto Chart in Python

The following tutorials offer additional information on histograms: