Scatter Plots

A scatter plot does not prove causation. It is a visual tool that shows the relationship between two variables, but it only reveals correlation, not a cause and effect connection. While a scatter plot may indicate that two variables are related (correlated), this does not imply that one variable causes the other. Often, other hidden factors, known as confounding variables, may influence both variables.

Correlation is a relationship where two variables move together in a predictable pattern. A scatter plot helps illustrate how one variable changes in relation to another, providing insights into the direction, strength, and pattern of their relationship. It also highlights outliers, or data points that are different from the general trend.

An example can be created using the mtcars data frame in R. A scatter plot could show a negative correlation between horsepower and miles per gallon: as horsepower increases, miles per gallon decreases. This suggests that cars with higher horsepower tend to have lower fuel efficiency.

Although the scatter plot would show that horsepower and fuel efficiency are related, this does not mean increasing horsepower causes lower fuel efficiency. There could be confounding variables, such as vehicle weight or engine type, that influence both horsepower and fuel consumption.

In summary, while scatter plots are helpful for understanding correlations, they do not establish a direct cause and effect relationship between variables.