# Cách chọn biểu đồ trực quan hóa dữ liệu

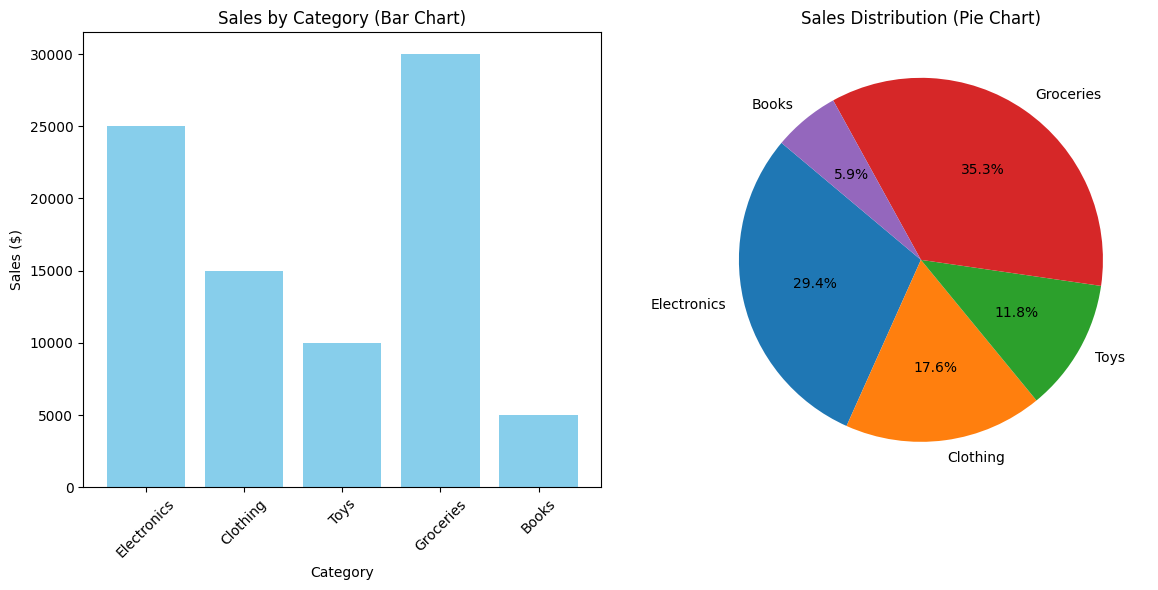
Nhiều người mới bắt đầu học khoa học dữ liệu thường bối rối về cách chọn biểu đồ trực quan hóa dữ liệu để trực quan hóa dữ liệu họ đang làm việc. Việc chọn biểu đồ trực quan hóa dữ liệu phù hợp tùy thuộc vào bản chất dữ liệu của bạn và thông tin chi tiết cụ thể mà bạn đang muốn rút ra. Vì vậy, nếu bạn muốn tìm hiểu cách chọn biểu đồ trực quan hóa dữ liệu thì bài viết này là dành cho bạn. Trong bài viết này, tôi sẽ hướng dẫn bạn cách chọn biểu đồ trực quan hóa dữ liệu.

Việc chọn biểu đồ trực quan hóa dữ liệu phù hợp là rất quan trọng trong việc truyền đạt hiệu quả những hiểu biết sâu sắc về dữ liệu của bạn. Mỗi biểu đồ phục vụ một mục đích khác nhau và có thể làm nổi bật các khía cạnh khác nhau của dữ liệu của bạn, từ so sánh và thành phần đến xu hướng, phân phối và mối quan hệ. Hãy xem xét tất cả các yếu tố có thể mà chúng ta có thể xem xét để quyết định cách chọn biểu đồ trực quan hóa dữ liệu.

## Khi bạn muốn so sánh

So sánh trong trực quan hóa dữ liệu đề cập đến việc đánh giá và làm nổi bật sự khác biệt hoặc tương đồng giữa các danh mục, nhóm hoặc điểm dữ liệu khác nhau. Nó giúp hiểu các phần tử khác nhau trong tập dữ liệu có liên quan với nhau như thế nào về độ lớn, kích thước hoặc giá trị.

Vì vậy, để so sánh các giá trị giữa các danh mục hoặc nhóm khác nhau, bạn có thể chọn:

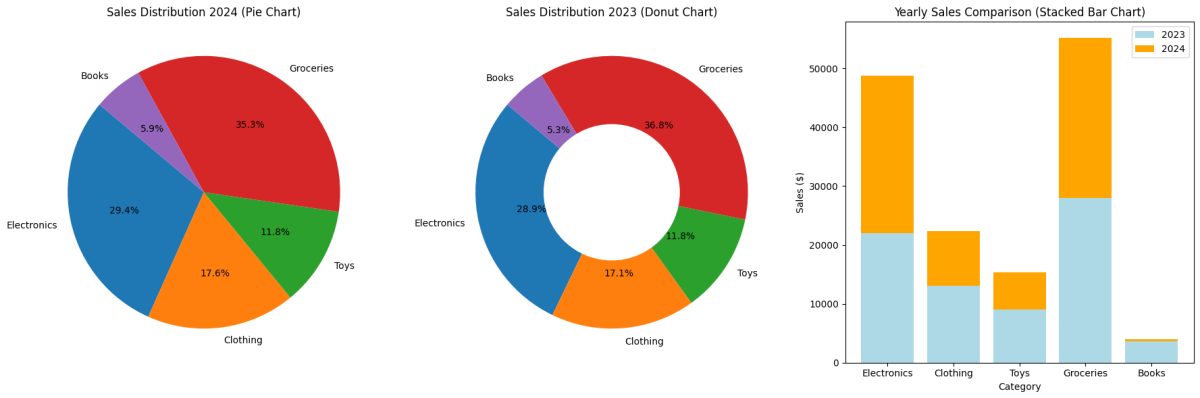


1. **Biểu đồ thanh (Bar Chart):** Biểu đồ thanh rất phù hợp để so sánh các giá trị giữa các danh mục hoặc nhóm. Ví dụ: bạn có thể sử dụng biểu đồ thanh để so sánh số liệu bán hàng giữa các khu vực khác nhau hoặc so sánh hiệu suất của các sản phẩm khác nhau.
2. **Biểu đồ hình tròn (Pie Chart):** Biểu đồ hình tròn có hiệu quả trong việc so sánh các phần của tổng thể. Chúng hoạt động tốt khi bạn có số lượng danh mục hạn chế nhưng chúng trở nên kém hiệu quả hơn khi bạn có nhiều danh mục hoặc có sự khác biệt nhỏ về dữ liệu.

## Khi bạn muốn kiểm tra thành phần

Thành phần trong trực quan hóa dữ liệu liên quan đến việc minh họa các phần tạo nên tổng thể. Đó là về việc hiển thị tỷ lệ hoặc sự đóng góp tương đối của các thành phần cho một thực thể hoặc tập dữ liệu lớn hơn.

Vì vậy, để hiểu các phần khác nhau tạo nên tổng thể như thế nào, bạn có thể chọn:



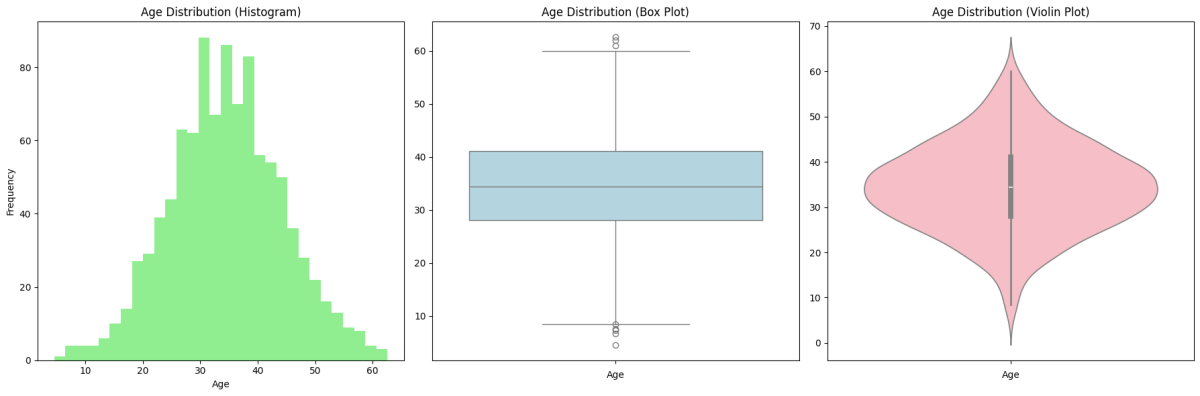
1. **Biểu đồ hình tròn (Pie Chart) hoặc Biểu đồ vành khuyên (Donut Charts):** Những biểu đồ này minh họa các phần khác nhau tạo nên tổng thể như thế nào. Chúng đặc biệt hiệu quả khi bạn có một số ít danh mục tạo thành toàn bộ tập dữ liệu.
2. **Biểu đồ thanh xếp chồng (Stacked Bar Charts):** Những biểu đồ này rất hữu ích để hiển thị thành phần của các phần khác nhau trong mối tương quan với tổng thể. Chúng hiển thị cả tổng số và tỷ lệ đóng góp của từng danh mục.

## Khi bạn muốn hiểu sự phân phối

Phân phối trong trực quan hóa dữ liệu đề cập đến việc trực quan hóa cách các điểm dữ liệu được trải rộng hoặc phân bổ trên các giá trị hoặc danh mục khác nhau. Nó giúp hiểu được hình dạng và đặc điểm của dữ liệu, chẳng hạn như liệu nó có bị lệch, đối xứng hay lưỡng kim hay không.

Vì vậy, để xem mức độ lan truyền của dữ liệu và hiểu sự phân bố của nó, bạn có thể chọn:

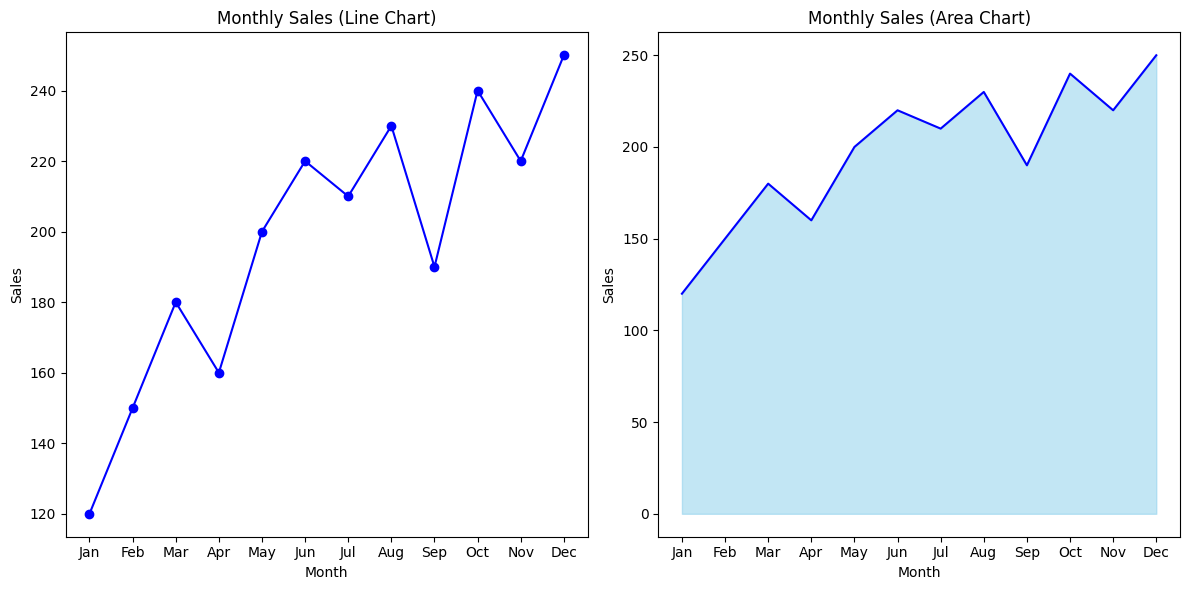
1. **Histogram:** Histogram là lý tưởng để hình dung sự phân bố của một biến duy nhất. Chúng hiển thị tần suất của các khoảng dữ liệu, giúp bạn hiểu được mức độ phân tán và hình dạng của dữ liệu.
2. **Biểu đồ hình hộp (Box Plots):** Biểu đồ hình hộp cung cấp bản tóm tắt ngắn gọn về việc phân bổ dữ liệu, bao gồm các thước đo như giá trị trung bình, phần tư và các giá trị ngoại lệ tiềm năng.
3. **Biểu đồ vĩ cầm (Violin Plots):** Biểu đồ vĩ cầm kết hợp các khía cạnh của biểu đồ hình hộp với ước tính mật độ hạt nhân, cung cấp một cách để trực quan hóa sự phân bố và mật độ xác suất của dữ liệu.



## Khi bạn muốn phân tích xu hướng theo thời gian

Xu hướng theo thời gian liên quan đến việc theo dõi cách các điểm dữ liệu thay đổi hoặc phát triển trong một khoảng thời gian liên tục, thường là theo trình tự thời gian. Đó là việc trực quan hóa các mẫu và biến động của dữ liệu trong một khung thời gian cụ thể.

Vì vậy, để quan sát các điểm dữ liệu thay đổi như thế nào theo thời gian, bạn có thể chọn:



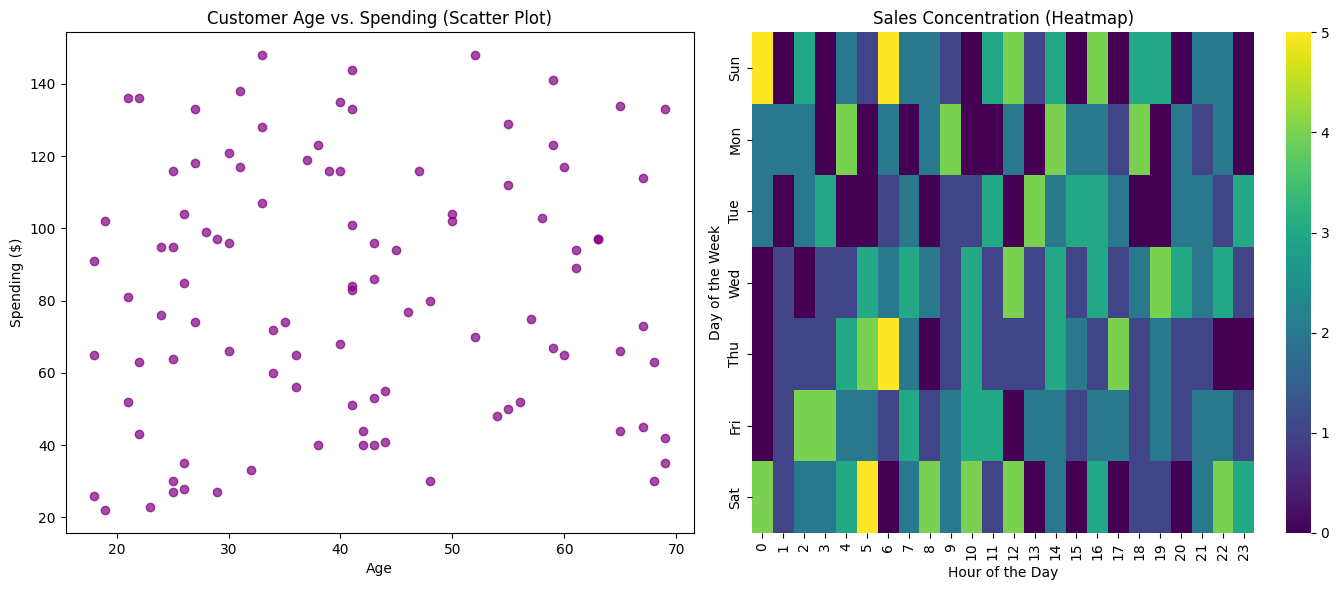
1. **Biểu đồ đường (Line Chart):** Biểu đồ đường là lựa chọn phù hợp để hiển thị xu hướng theo thời gian. Chúng đặc biệt hữu ích cho việc phân tích dữ liệu liên tục và chuỗi thời gian.
2. **Biểu đồ vùng (Area Chart):** Biểu đồ vùng, tương tự như biểu đồ đường, nhấn mạnh tầm quan trọng của xu hướng theo thời gian bằng cách lấp đầy khu vực bên dưới đường.

## Khi bạn muốn phân tích các mối quan hệ

Các mối quan hệ trong trực quan hóa dữ liệu tập trung vào việc trực quan hóa cách các biến hoặc điểm dữ liệu phụ thuộc lẫn nhau. Đó là việc khám phá các kết nối, liên kết hoặc sự phụ thuộc giữa các biến khác nhau.

Vì vậy, để hiểu mối quan hệ giữa hai hoặc nhiều biến, bạn có thể chọn:

1. **Biểu đồ phân tán (Scatter Plots):** Biểu đồ phân tán rất cần thiết để hình dung các mối quan hệ hoặc mối tương quan giữa hai biến số. Chúng giúp xác định các mẫu và xu hướng trong dữ liệu.
2. **Bản đồ nhiệt (Heatmaps):** Bản đồ nhiệt sử dụng mã màu để hiển thị mối quan hệ giữa hai biến. Chúng rất phù hợp để xác định các mẫu và cụm trong tập dữ liệu lớn.



### **Kết luận**

Vì vậy, việc lựa chọn biểu đồ là rất quan trọng và thay đổi tùy theo tính chất của dữ liệu và thông tin chi tiết dự kiến. Biểu đồ thanh và biểu đồ tròn là lý tưởng để so sánh, minh họa sự khác biệt giữa các danh mục hoặc các phần của tổng thể. Để hiểu rõ bố cục, biểu đồ hình tròn hoặc biểu đồ vành khuyên và biểu đồ thanh xếp chồng hiển thị một cách hiệu quả cách các phần khác nhau đóng góp vào tổng thể. Histogram, biểu đồ hình hộp và biểu đồ vĩ cầm phù hợp để phân tích phân bố dữ liệu và tiết lộ tần số, mức độ lan truyền và mật độ. Biểu đồ đường và vùng được ưu tiên sử dụng cho các xu hướng theo thời gian, ghi lại các thay đổi dữ liệu liên tục. Cuối cùng, biểu đồ phân tán và bản đồ nhiệt là công cụ tuyệt vời để kiểm tra mối quan hệ giữa các biến và làm nổi bật các mẫu cũng như mối tương quan trong dữ liệu số và phân loại.