# Quy tắc trực quan hóa dữ liệu để không bao giờ sai

Trực quan hóa dữ liệu một cách hiệu quả là rất quan trọng để phân tích và diễn giải. Có một số quy tắc trực quan hóa dữ liệu mà bạn phải luôn tuân theo để đảm bảo trực quan hóa dữ liệu của bạn chính xác và mang tính thông tin. Vì vậy, nếu bạn muốn biết các quy tắc trực quan hóa dữ liệu cần thiết mà bạn phải luôn tuân theo, thì bài viết này là dành cho bạn. Trong bài viết này, tôi sẽ hướng dẫn bạn một số quy tắc trực quan hóa dữ liệu để không bao giờ sai sót khi trực quan hóa hoặc trình bày dữ liệu của bạn.

## **Quy tắc trực quan hóa dữ liệu để không bao giờ sai**

Dưới đây là một số quy tắc trực quan hóa dữ liệu mà bạn phải luôn tuân theo để không bao giờ mắc sai lầm trong trình bày hoặc trong khi trực quan hóa dữ liệu của mình:

1. Trục biến phụ thuộc và biến độc lập (Axis of Dependent and Independent Variables)
2. Chọn loại biểu đồ phù hợp (Chart Type)
3. Nhãn trục (Labelling Axes)
4. Sử dụng thang đo thích hợp
5. Chú giải và Từ khóa (Legends và Keys)
6. Không bao giờ quên tiêu đề (Titles) và chú thích (Annotations)
7. Sử dụng độ dày (Weights) và kiểu đường (Styles) khi cần thiết

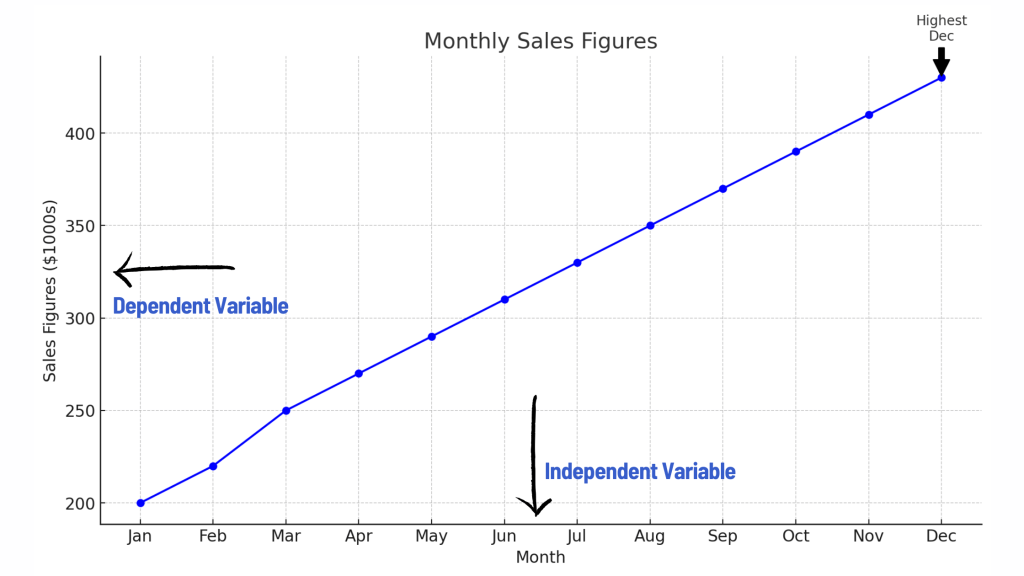
Hãy cùng điểm qua tất cả các quy tắc này để hiểu chúng một cách chi tiết nhé!

## **Trục biến phụ thuộc và biến độc lập**

Vị trí của các biến trên biểu đồ là nền tảng cho việc giải thích nó. Theo quy ước, biến độc lập, là biến bạn thao tác hoặc coi là đường cơ sở, được định vị trên trục x. Biến phụ thuộc thay đổi theo biến độc lập, được đặt trên trục y.

Thiết lập này phản ánh ký hiệu hàm toán học cổ điển **y=f(x)**, trong đó y được coi là hàm của x.

Ví dụ, nhìn vào hình ảnh dưới đây. Nếu bạn đang vẽ đồ thị cho thấy doanh số bán hàng thay đổi như thế nào qua các tháng thì các tháng (biến độc lập) phải nằm trên trục x và số liệu bán hàng (biến phụ thuộc) phải nằm trên trục y.



## **Chọn loại biểu đồ phù hợp**

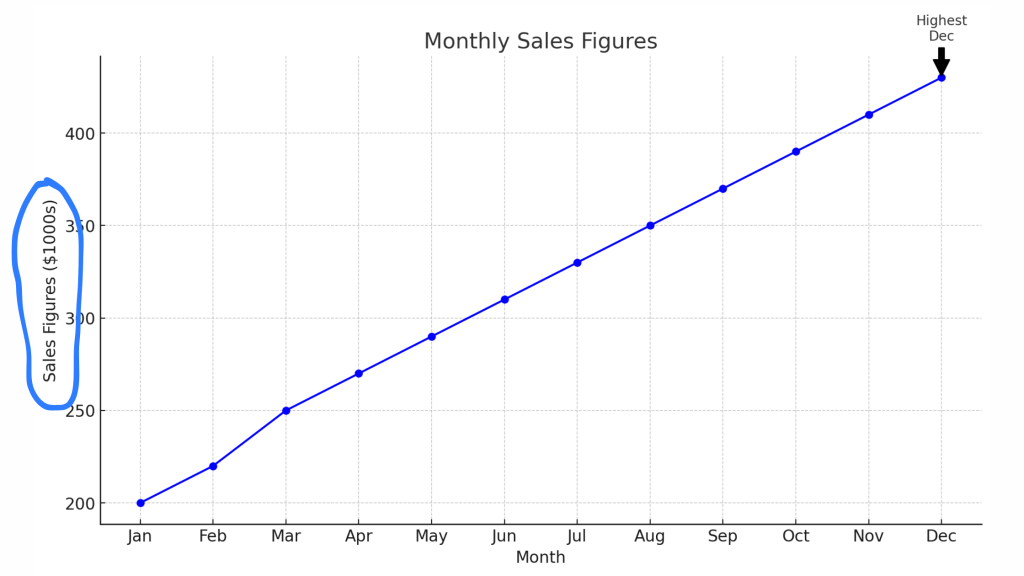
Việc chọn loại biểu đồ thích hợp là rất quan trọng để truyền đạt câu chuyện dữ liệu một cách hiệu quả. Việc chọn đúng loại biểu đồ không chỉ giúp trình bày rõ ràng mà còn thu hút người xem bằng cách trình bày dữ liệu một cách sâu sắc nhất.

Biểu đồ thanh (Bar Charts) và biểu đồ hình tròn (Pie Charts) 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 (Donut Charts) và biểu đồ thanh xếp chồng (Stacked Bar Charts) hiển thị một cách hiệu quả cách các phần khác nhau đóng góp vào tổng thể. Histograms, biểu đồ hình hộp (Box Plots) và biểu đồ vĩ cầm (Violin Plots) phù hợp để phân tích phân bố dữ liệu và tiết lộ tần số, mức độ lan truyền và mật độ. Biểu đồ đường (Line Plots) và biểu đồ vùng (Area Charts) 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 (Scatter Plots) và bản đồ nhiệt (Heatmaps) 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.

## **Nhãn trục**

Nhãn trục là các yếu tố thiết yếu cung cấp bối cảnh và quy mô cho dữ liệu được trình bày. Các nhãn rõ ràng, bao gồm tên biến và đơn vị đo lường của nó, giúp biểu đồ có nhiều thông tin và dễ hiểu. Quy tắc này đảm bảo rằng người xem có thể hiểu biểu đồ mà không cần tham khảo văn bản bên ngoài, làm cho hình ảnh trực quan trở thành một công cụ độc lập để truyền dữ liệu.

Ví dụ: Nếu trục y biểu thị số liệu bán hàng thì trục y phải được gắn nhãn là “Sales Figures ($1000s)” (Số liệu bán hàng ($1000)) nếu các giá trị tính bằng hàng nghìn đô la Mỹ, như được minh họa trong biểu đồ bên dưới.

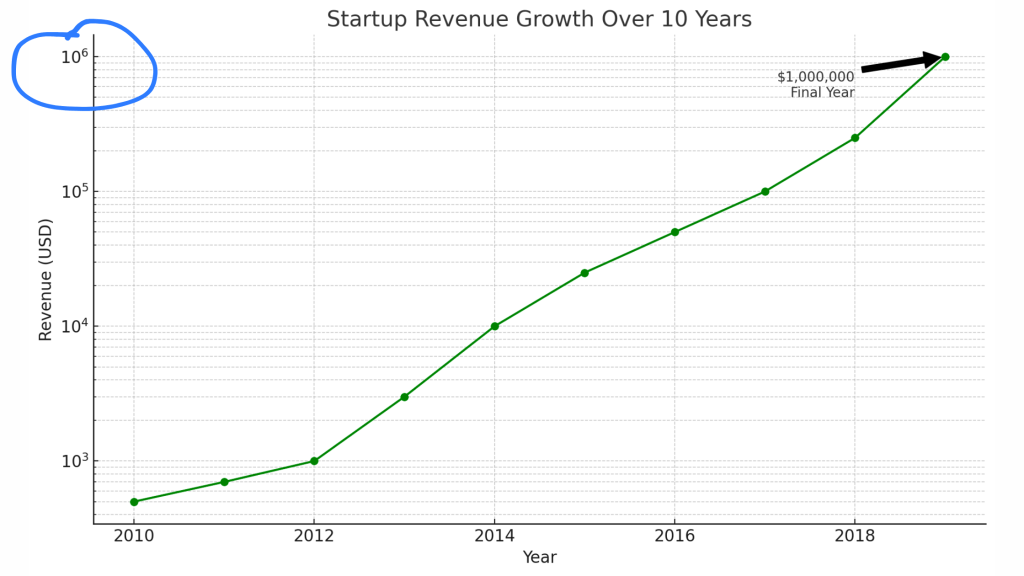


## **Sử dụng thang đo thích hợp**

Tỷ lệ của các trục xác định khoảng cách giữa các điểm dữ liệu và có thể tác động đáng kể đến việc diễn giải dữ liệu. Thang đo tuyến tính rất đơn giản và hoạt động tốt đối với dữ liệu được phân bổ đồng đều.

Ngược lại, thang logarit có lợi cho dữ liệu bao gồm nhiều giá trị vì chúng có thể nén các giá trị lớn và mở rộng các giá trị nhỏ hơn, làm cho xu hướng trở nên rõ ràng hơn. Việc lựa chọn tỷ lệ ảnh hưởng đến khả năng đọc biểu đồ và khả năng đưa ra so sánh và kết luận của người xem.

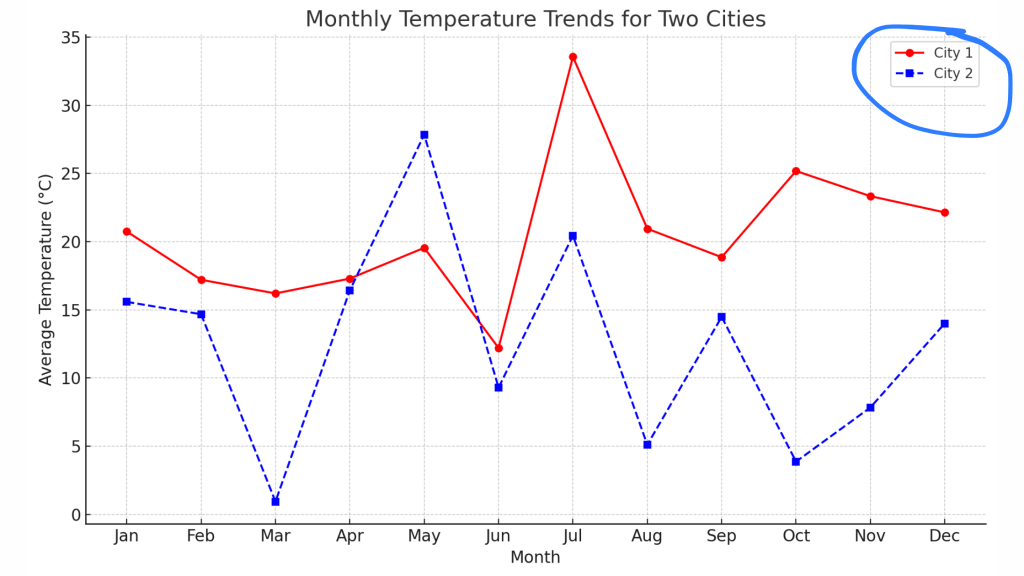
Ví dụ: biểu đồ bên dưới thể hiện mức tăng trưởng doanh thu của một công ty khởi nghiệp trong 10 năm, sử dụng thang logarit trên trục y để chứa nhiều giá trị, từ vài trăm đến một triệu đô la.



## **Chú giải và Từ khóa**

Chú giải và từ khóa loại bỏ sự mơ hồ bằng cách giải thích các ký hiệu, màu sắc và mẫu được sử dụng trong biểu đồ. Điều này đặc biệt quan trọng khi trực quan hóa nhiều tập dữ liệu hoặc danh mục trong một biểu đồ vì nó cho phép người xem dễ dàng phân biệt giữa chúng. Chú giải hoặc phím được đặt đúng vị trí sẽ nâng cao khả năng diễn giải dữ liệu một cách chính xác của người xem, giúp việc trực quan hóa hiệu quả hơn.

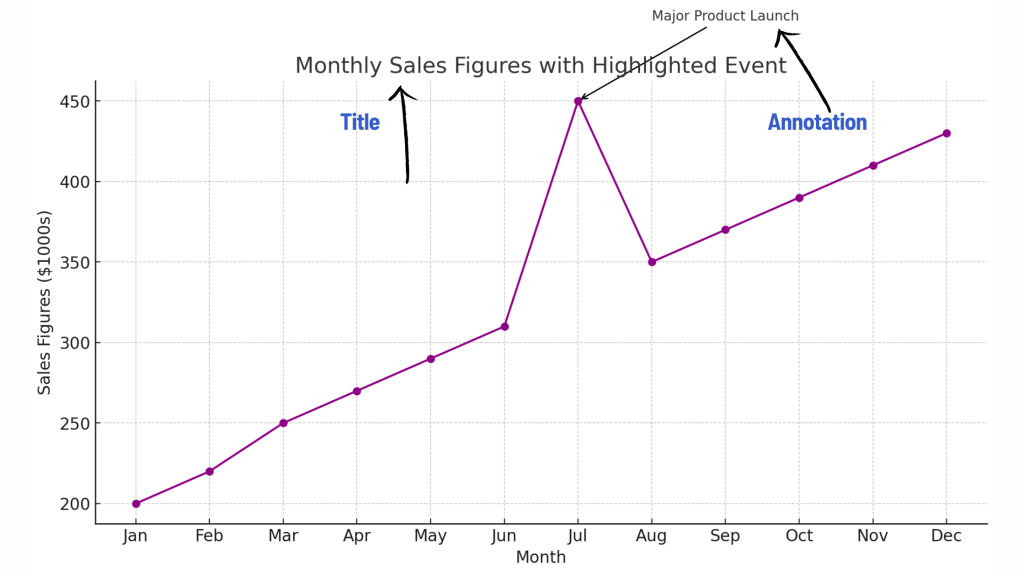
Ví dụ, nhìn vào hình ảnh dưới đây. Trong biểu đồ đường hiển thị xu hướng nhiệt độ trong một năm ở hai thành phố khác nhau, hãy sử dụng các màu khác nhau cho mỗi thành phố và bao gồm chú giải để xác định màu nào tương ứng với thành phố nào.



## **Không bao giờ quên tiêu đề và chú thích**

Tiêu đề đóng vai trò là tiêu đề cho câu chuyện dữ liệu của bạn, thu hút sự chú ý của người xem đến mục đích của biểu đồ, trong khi chú thích có thể làm nổi bật các điểm dữ liệu, xu hướng hoặc điểm bất thường cụ thể, cung cấp thông tin chi tiết sâu hơn. Việc sử dụng hiệu quả các tiêu đề và chú thích sẽ biến một biểu đồ đơn giản thành một câu chuyện hướng dẫn người xem tìm hiểu dữ liệu một cách có cấu trúc và sâu sắc.

Ví dụ, hãy nhìn vào biểu đồ bên dưới. Biểu đồ hiển thị số liệu bán hàng hàng tháng trong suốt một năm, tập trung cụ thể vào mức tăng đáng chú ý trong tháng 7, do ra mắt sản phẩm lớn. Chú thích “Major Product Launch” (Ra mắt sản phẩm chính) được đặt một cách chiến lược để thu hút sự chú ý đến mức tăng đột biến này, cung cấp ngữ cảnh cho dữ liệu và nêu bật tác động của chiến dịch tiếp thị đến số liệu bán hàng.

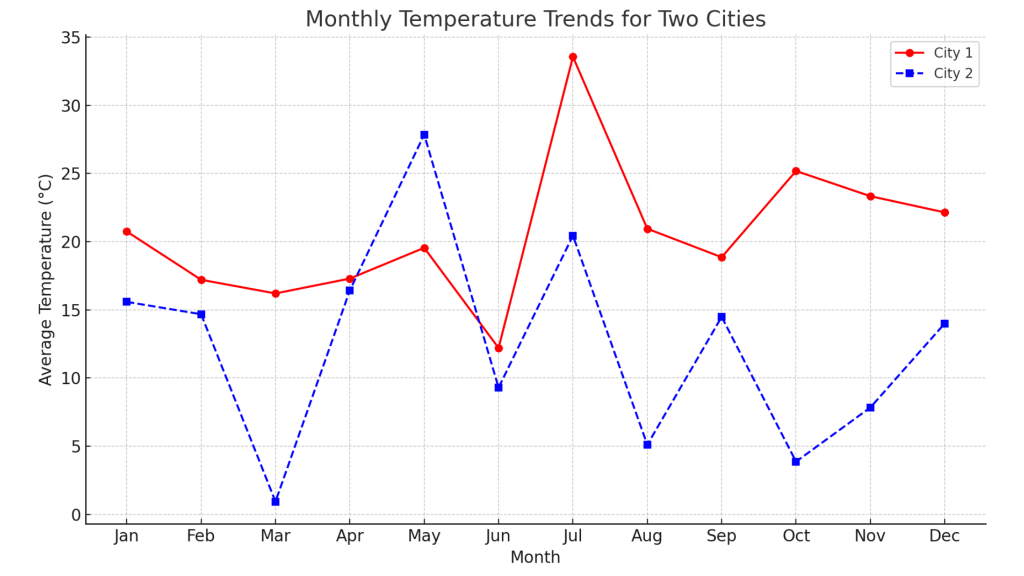


Hình ảnh trực quan này truyền đạt một cách hiệu quả câu chuyện đằng sau những con số, nâng cao khả năng diễn giải và mức độ tương tác của việc trình bày dữ liệu.

## **Sử dụng độ dày và kiểu đường khi cần thiết**

Việc phân biệt các đặc điểm trong biểu đồ là rất quan trọng để có được sự rõ ràng. Việc sử dụng các độ dày và kiểu đường khác nhau (ví dụ: nét liền, nét đứt, chấm) cho phép mỗi danh mục hoặc đặc điểm trở nên khác biệt và dễ nhận biết, đảm bảo rằng hình ảnh trực quan vẫn hiệu quả trong các điều kiện xem khác nhau.

Ví dụ: trong biểu đồ bên dưới, chúng tôi đang so sánh nhiệt độ của hai thành phố khác nhau được đánh dấu bằng các kiểu đường khác nhau.



## **Kết luận**

Đây là những quy tắc trực quan hóa dữ liệu mà bạn phải luôn tuân theo khi trình bày hoặc trực quan hóa dữ liệu của mình. Việc tuân thủ các quy tắc trực quan hóa dữ liệu này đảm bảo rằng cách trình bày dữ liệu trực quan của bạn không chỉ hấp dẫn về mặt trực quan mà còn chính xác, nhiều thông tin và có thể truy cập được. Bằng cách xem xét cẩn thận vị trí của các biến, lựa chọn loại biểu đồ, độ rõ ràng của nhãn, mức độ phù hợp của tỷ lệ, bao gồm chú giải, cách sử dụng tiêu đề và chú thích một cách chiến lược cũng như sự khác biệt của bộ dữ liệu, bạn có thể tạo trực quan hóa để truyền đạt hiệu quả dữ liệu phức tạp.