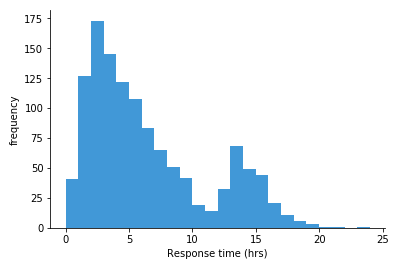
**ÔN TẬP LÝ THUYẾT**

* *Trực quan hóa dữ liệu có vai trò gì trong phân tích dữ liệu? Tại sao nó quan trọng trong EDA?*

Trực quan hóa dữ liệu giúp biến những con số thành hình ảnh dễ hiểu, từ đó hỗ trợ việc quan sát, so sánh và phát hiện xu hướng, mối quan hệ hay bất thường trong dữ liệu. Trong giai đoạn khám phá dữ liệu (Exploratory Data Analysis – EDA), trực quan hóa đặc biệt quan trọng vì nó cho phép kiểm chứng giả thuyết, hiểu được đặc trưng của dữ liệu và định hướng cho các bước phân tích tiếp theo.

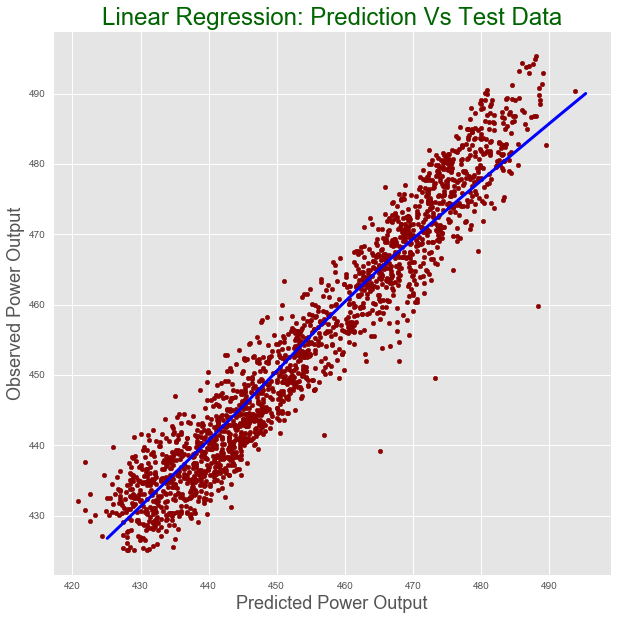
* *Các loại biểu đồ phổ biến (như histogram, scatter plot, boxplot, bar chart) được sử dụng trong các trường hợp nào?*
* **Histogram:** thể hiện sự phân bố tần suất của một tập dữ liệu liên tục thông qua các cột, giúp phân tích và trực quan hóa dữ liệu một cách hiệu quả.

*Ví dụ*:



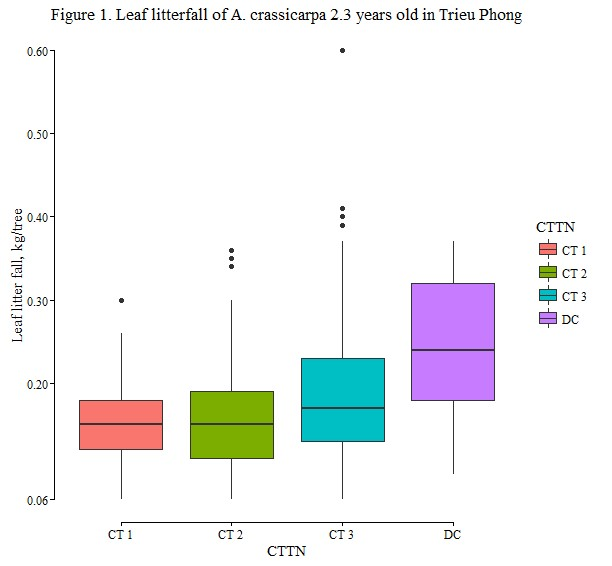
* **Scatter plot:** là dạng biểu đồ thể hiện tương quan giữa hai biến trên hệ tọa độ. Trong đó, trục Oy là biến phụ thuộc, trục Ox là biến độc lập. Dữ liệu thu thập được sẽ hiển thị trên biểu đồ bằng các dấu chấm tại các điểm tương ứng.

*Ví dụ*:

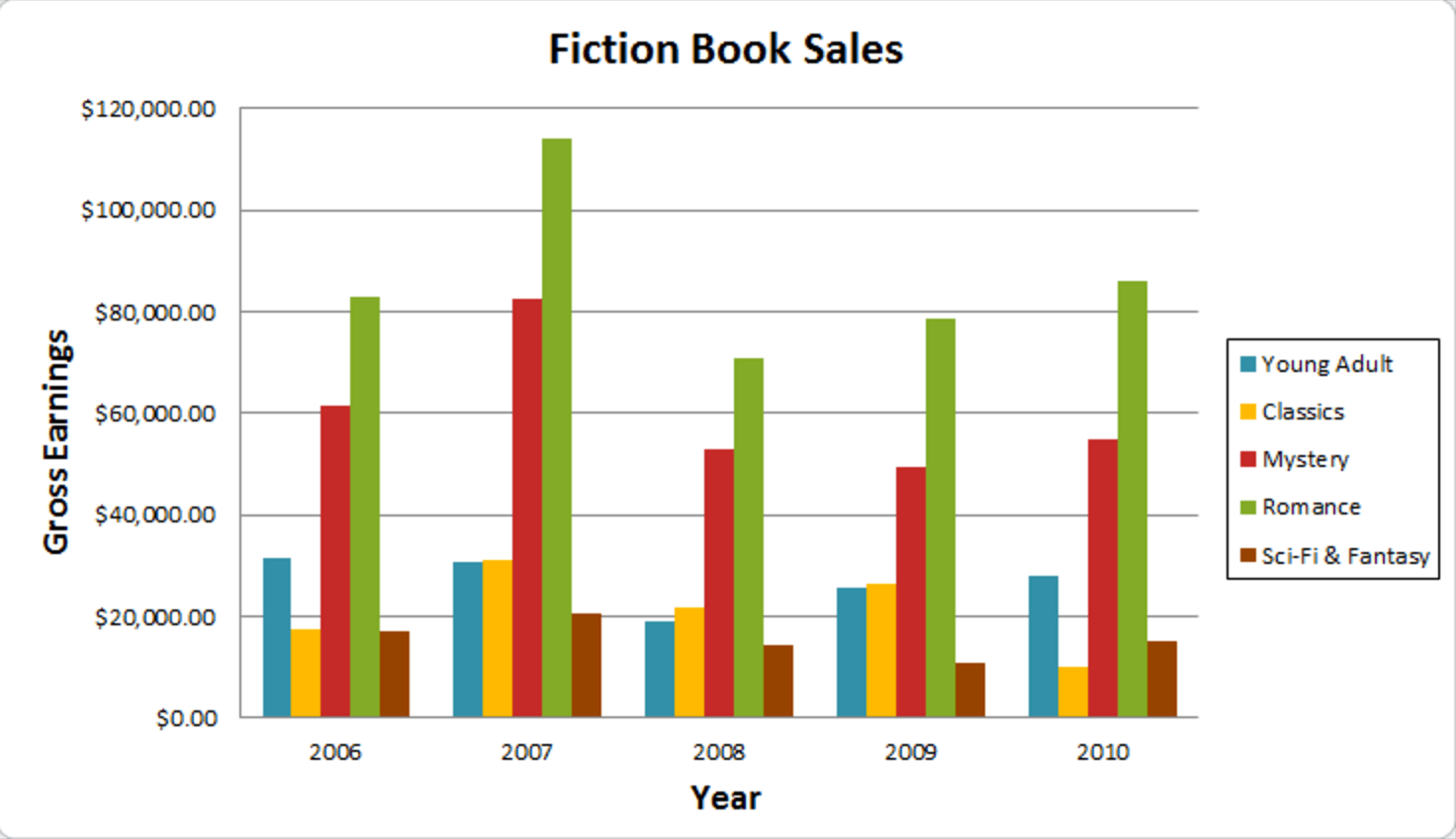


* **Boxplot:** là dạng biểu đồ mô tả sự phân bố giá trị định lượng của một hoặc nhiều nhóm dữ liệu dạng phân loại. Trong đó:
* Độ dài của hộp biểu thị phạm vi của 50% dữ liệu trung tâm
* Đường kẻ giữa hộp là giá trị trung vị của tập dữ liệu
* Các đường kẻ bên ngoài (râu của biểu đồ) mô tả phạm vi của những giá trị dữ liệu ngoài khoảng 25% và 75% dữ liệu trung tâm
* Các dấu cấm bên ngoài là các giá trị ngoại lai

*Ví dụ*:



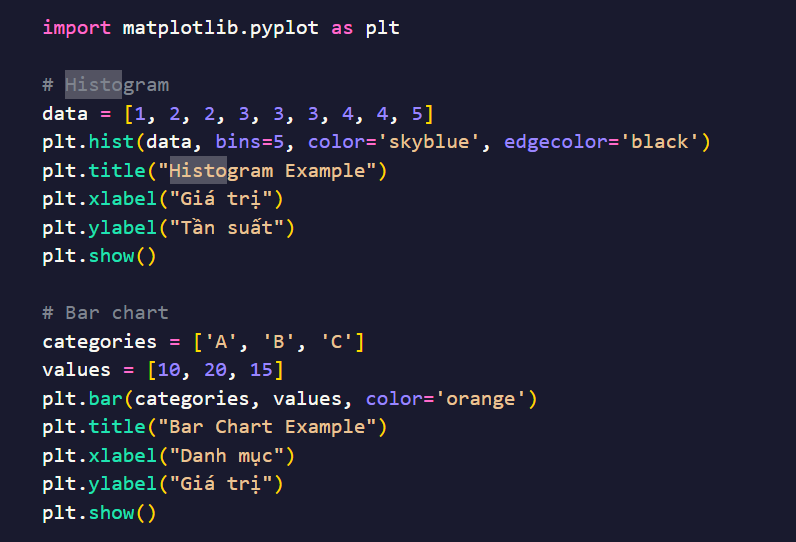
* **Bar chart:** Dùng để biểu diễn dữ liệu phân loại hoặc so sánh giữa các nhóm.

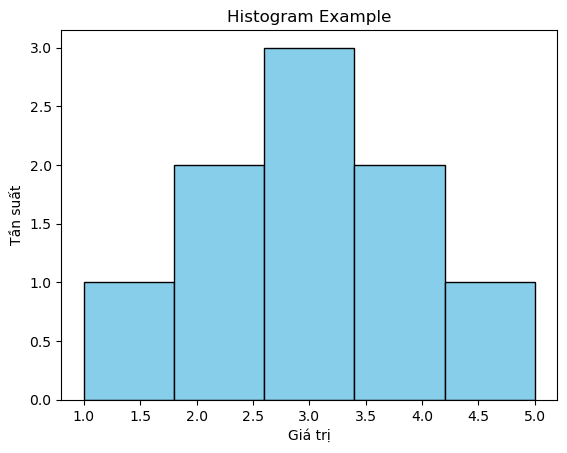
*Ví dụ*: 

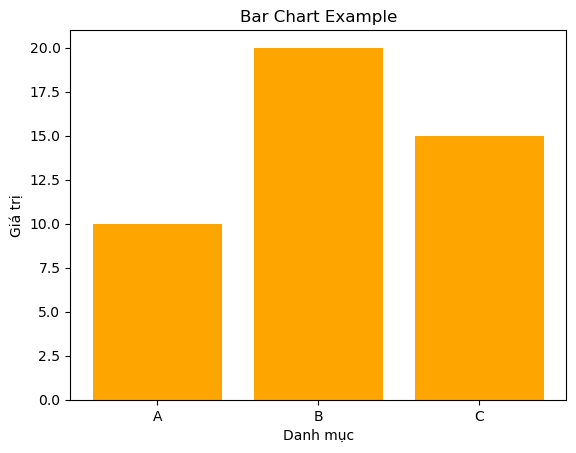
* *Làm thế nào để chọn loại biểu đồ phù hợp với đặc điểm của dữ liệu (ví dụ: dữ liệu phân loại, dữ liệu số, dữ liệu thời gian)?*
* **Dữ liệu phân loại:** Biểu đồ cột hoặc tròn.
* **Dữ liệu số:** Histogram, boxplot, scatter plot.
* **Dữ liệu chuỗi thời gian:** Line chart để thể hiện xu hướng theo thời gian.
* *Sự khác biệt giữa các thư viện trực quan hóa trong Python như Matplotlib, Seaborn và Plotly là gì?*
* **Matplotlib:** Thư viện cơ bản, linh hoạt, cho phép tùy chỉnh chi tiết nhưng cần nhiều dòng lệnh.
* **Seaborn:** Xây dựng trên Matplotlib, hỗ trợ trực quan thống kê, tạo biểu đồ đẹp mắt chỉ với ít lệnh.
* **Plotly:** Mạnh về biểu đồ tương tác, phù hợp cho dashboard hoặc ứng dụng web.
* *Những nguyên tắc thiết kế nào cần tuân thủ để tạo ra một biểu đồ trực quan hóa dễ hiểu và hiệu quả?*

Đảm bảo biểu đồ đơn giản, dễ hiểu.  
 Sử dụng màu sắc, tỷ lệ hợp lý.  
 Nhấn mạnh thông tin quan trọng.  
 Thêm nhãn trục, tiêu đề, chú thích rõ ràng.

* *Làm thế nào để tạo một biểu đồ đơn giản như histogram hoặc bar chart bằng Matplotlib? Bạn có thể chia sẻ đoạn code mẫu không?*







* *Làm thế nào để xuất biểu đồ từ Python ra các định dạng như PNG, PDF hoặc HTML để sử dụng trong báo cáo?*
* Với Matplotlib: dùng *plt.savefig("ten\_file.png")* để lưu ra PNG hoặc đổi thành *.pdf.*
* Với Plotly: dùng *fig.write\_html("ten\_file.html")* để xuất file HTML có tương tác.