### **Các biểu đồ nâng cao 2**

### **Chi tiết bài đọc**

Trong bài này, chúng ta sẽ tìm hiểu về 3 loại biểu đồ: heat map, box plot, và pair plot.

**Heat map:**

Heat map là một loại biểu đồ một biểu diễn đồ họa hai chiều của dữ liệu trong đó các giá trị riêng lẻ chứa trong ma trận được biểu diễn dưới dạng màu sắc. Để biểu diễn các giá trị phổ biến hơn hoặc các hoạt động cao hơn, các màu sáng hơn được sử dụng và để biểu thị các giá trị hoạt động hoặc ít phổ biến hơn, các màu tối hơn được ưu tiên.

Tạo dữ liệu ngẫu nhiên theo phân phối đều để vẽ:

import numpy as np

uniform\_data = np.random.rand(10, 12)

Vẽ heatmap với dữ liệu đã sinh ra:

sns.heatmap(uniform\_data)

A picture containing text, checker, colorful, colors

Description automatically generated

Thay đổi giới hạn các màu sắc:

sns.heatmap(uniform\_data, vmin=0, vmax=1)

A picture containing text, checker, colorful, colors

Description automatically generated

**Box plot:**

Box plot biểu diễn trực quan của các nhóm dữ liệu số được mô tả thông qua các phần tư của chúng. Boxplot cũng được sử dụng để phát hiện giá trị ngoại lai trong tập dữ liệu như đã giới thiệu ở bài 4. Nó biểu diễn phân bố của dữ liệu một cách hiệu quả với một hộp đơn giản và râu và cho phép chúng ta so sánh dễ dàng giữa các nhóm. Boxplot biểu diễn dữ liệu mẫu bằng cách sử dụng phần trăm thứ 25, 50 và 75. Những phần trăm này còn được gọi là phần tư dưới, trung vị và phần tư trên.

Mỗi box plot gồm 5 phần:

* Minimum
* Phần tư thứ nhất (25%)
* Trung vị (50%)
* Phần tư thứ ba (75%)
* Maximum

Chúng ta sẽ sử dụng bộ dữ liệu “tips” để ví dụ:

Vẽ biểu đồ box plot cho thuộc tính “total\_bills”

sns.boxplot(x=tips["total\_bill"])

Chart

Description automatically generated

Vẽ box plot theo chiều dọc được nhóm theo một biến phân loại:

sns.boxplot(x="day", y="total\_bill", data=tips)

Chart, box and whisker chart

Description automatically generated

Vẽ box plot theo chiều dọc được nhóm theo hai biến phân loại:

sns.boxplot(x="day", y="total\_bill", hue="smoker",

data=tips, palette="Set3")

Chart, box and whisker chart

Description automatically generated

Vẽ box plot theo chiều dọc được nhóm theo hai biến phân loại khi một số nhóm bị thiếu:

sns.boxplot(x="day", y="total\_bill", hue="time",

data=tips, linewidth=2.5)

Chart, box and whisker chart

Description automatically generated

Chúng ta có thể tham khảo thêm tại: *https://seaborn.pydata.org/generated/seaborn.boxplot.html*

**Pair plot:**

Để vẽ nhiều phân phối hai biến theo cặp trong một tập dữ liệu, chúng ta có thể sử dụng hàm biểu đồ cặp, pair plot. Biểu đồ cặp cho thấy mối quan hệ đối với kết hợp của biến trong DataFrame dưới dạng ma trận các ô và các ô đường chéo là các ô đơn biến.

Chúng ta vẫn sẽ sử dụng bộ dữ liệu “tips” để ví dụ:

sns.pairplot(tips, hue ='sex')

Chart

Description automatically generated with medium confidence

Thay đổi biến hue để thay đổi việc mapping các thuộc tính trong biểu đồ:

sns.pairplot(tips, hue ='day')

Chart

Description automatically generated

### **Tổng kết**

Qua bài đọc này, chúng ta đã tìm hiểu về 3 loại biểu đồ: heat map, box plot, và pair plot.