Bài tuần 2

Bài 1

**import** **pandas** **as** **pd // thư viện thao tác trên csdl**

**from** **matplotlib** **import** pyplot **as** plt **// thư viện trực quan hóa**

%matplotlib inline

**import** **seaborn** **as** **sns** **// thư viện trực quan hóa nhưng nâng cao của matplotlib**

df = pd.read\_csv('Pokemon.csv', index\_col=0,encoding= 'unicode\_escape') // tải file csv từ dataframe

df.head()// in ra 5 dòng đầu tiên

sns.lmplot(x='Attack', y='Defense', data=df)// trực quan hóa biểu đồ phân tán với dữ liệu 'Attack' và 'Defense'

sns.lmplot(x='Attack', y='Defense', data=df, fit\_reg=False, hue='Stage')   trực quan hóa biểu đồ phân tán với dữ liệu 'Attack' và 'Defense' thêm màu và bỏ đi đường hồi quy.

sns.lmplot(x='Attack', y='Defense', data=df,

           fit\_reg=False,

           hue='Stage')

# Tweak using Matplotlib

plt.ylim(0, None)

plt.xlim(0, None) // giới hạn trên và giới hạn dưới

sns.boxplot(data=df)// tạo sơ đồ hộp

stats\_df = df.drop(['Total', 'Stage', 'Legendary'], axis=1)

# New boxplot using stats\_df

sns.boxplot(data=stats\_df) // loại bỏ các cột 'Total', 'Stage', 'Legendary' và tạo sơ đồ hộp

sns.set\_style('whitegrid') // thiệt lập màu nền

# Violin plot

sns.violinplot(x='Type 1', y='Attack', data=df) // trực quan hóa sự phân bố

sns.violinplot(x='Type 1', y='Attack', data=df,

               palette=pkmn\_type\_colors) # Set color palette

sns.swarmplot(x='Type 1', y='Attack', data=df,

              palette=pkmn\_type\_colors) // trực quan hóa sự phân bố biễu đồ chấm

plt.figure(figsize=(10,6)) // Đặt kích thước hình với matplotlib

# Create plot

sns.violinplot(x='Type 1', // tạo plot

               y='Attack',

               data=df,

               inner=None, # Bỏ các thanh bên trong đàn viôlông

               palette=pkmn\_type\_colors)

sns.swarmplot(x='Type 1',

              y='Attack',

              data=df,

              color='k', // làm các điểm màu den

              alpha=0.7) //and slightly transparent

plt.title('Attack by Type') // đặt tiêu đề

stats\_df.head()// in 5 dòng đầu stats\_df

melted\_df = pd.melt(stats\_df,

                    id\_vars=["Name", "Type 1", "Type 2"], # Variables to keep

                    var\_name="Stat") # Name of melted variable

melted\_df.head() // tạo dataframe tan chảy

print( stats\_df.shape )

print( melted\_df.shape ) // đếm bao nhiêu dòng và bao nhiêu cột

sns.swarmplot(x='Stat', y='value', data=melted\_df,

              hue='Type 1') // trực quan hình ảnh



|  |
| --- |
| Enlarge the plot  plt.figure(figsize=(10,6))    sns.swarmplot(x='Stat',                y='value',                data=melted\_df,                hue='Type 1',                split=True, # 2. Separate points by hue                palette=pkmn\_type\_colors) # 3. Use Pokemon palette    # 4. Adjust the y-axis  plt.ylim(0, 260)    # 5. Place legend to the right  plt.legend(bbox\_to\_anchor=(1, 1), loc=2) |