



Міністерство освіти і науки України  
Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
Факультет інформатики та обчислювальної техніки  
Кафедра інформаційних систем та технологій

Лабораторна робота №4  
з дисципліни  
Аналіз даних з використанням мови Python

Виконав:

студент групи ІА-24:  
Криворучек В.С.

Перевірила:

ст. викладач  
Тимофєєва Ю.С.

**Тема:** Візуалізація даних за допомогою matplotlib та Seaborn

**Мета роботи:** Ознайомитись з основними діаграмами та графіками, що використовуються при аналізі даних. Навчитись будувати їх за допомогою бібліотек matplotlib та Seaborn.

### Хід роботи

Завдання:

Файл penguins.csv.

1. Побудувати стовпчикові діаграми, на яких відобразити а) кількість пінгвінів на кожному острові; б) медіанну вагу пінгвінів на кожному острові; в) середню довжину ласт пінгвінів на кожному острові з розподілом за статтю.
2. Побудувати гістограму довжини ласт пінгвінів, загальну і для кожного виду.
3. Побудувати діаграму розмаху ваги пінгвінів (загальну і в залежності від острова), визначити чи присутні викиди.
4. За допомогою діаграм розсіювання зробити висновки щодо залежності між а) довжиною дзьобу і ласт; б) вагою і довжиною дзьобу. Порахувати коефіцієнт кореляції за допомогою відповідних функцій.

Код програми:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Завантаження даних
df = pd.read_csv("penguins.csv")

# Показати оригінальні назви колонок
print("Оригінальні назви колонок:", df.columns.tolist())

# 1.a) Кількість пінгвінів на кожному острові
plt.figure(figsize=(6, 4))
sns.countplot(data=df, x="island", hue="island", palette="viridis",
              legend=False)
plt.title("Кількість пінгвінів на кожному острові")
plt.xlabel("Острів")
plt.ylabel("Кількість")
plt.tight_layout()
plt.show()

# 1.b) Медіанна вага пінгвінів на кожному острові
median_mass = df.groupby("island",
```

```

as_index=False)["body_mass_g"].median()
print("Медіанна вага пінгвінів за островами:\n", median_mass)

plt.figure(figsize=(6, 4))
sns.barplot(data=median_mass, x="island", y="body_mass_g",
hue="island", palette="magma", legend=False)
plt.title("Медіанна вага пінгвінів на островах")
plt.xlabel("Острів")
plt.ylabel("Медіанна вага (г)")
plt.tight_layout()
plt.show()

# 1.c) Середня довжина ласт за островом та статтю
mean_flipper = df.groupby(["island", "sex"],
as_index=False)["flipper_length_mm"].mean()
print("Середня довжина ласт пінгвінів за островами та статтю:\n",
mean_flipper)

plt.figure(figsize=(8, 5))
sns.barplot(data=mean_flipper, x="island", y="flipper_length_mm",
hue="sex", palette="coolwarm")
plt.title("Середня довжина ласт за островами та статтю")
plt.xlabel("Острів")
plt.ylabel("Середня довжина ласт (мм)")
plt.tight_layout()
plt.show()

# 2. Гістограма довжини ласт
plt.figure(figsize=(6, 4))
sns.histplot(data=df, x="flipper_length_mm", bins=20, kde=True)
plt.title("Гістограма довжини ласт (загальна)")
plt.xlabel("Довжина ласт (мм)")
plt.tight_layout()
plt.show()

plt.figure(figsize=(8, 5))
sns.histplot(data=df, x="flipper_length_mm", hue="species", bins=20,
kde=True, palette="Set2")
plt.title("Гістограма довжини ласт за видами")
plt.xlabel("Довжина ласт (мм)")
plt.tight_layout()
plt.show()

# 3. Boxplot ваги пінгвінів
plt.figure(figsize=(6, 4))
sns.boxplot(data=df, y="body_mass_g", color="lightblue")
plt.title("Діаграма розмаху ваги (загальна)")
plt.ylabel("Вага (г)")
plt.tight_layout()
plt.show()

plt.figure(figsize=(7, 5))
sns.boxplot(data=df, x="island", y="body_mass_g", hue="island",
palette="pastel", legend=False)
plt.title("Діаграма розмаху ваги за островами")
plt.xlabel("Острів")
plt.ylabel("Вага (г)")
plt.tight_layout()
plt.show()

```

```
# Описова статистика ваги по островах
print("Описова статистика ваги пінгвінів по островах:\n",
df.groupby("island")["body_mass_g"].describe())

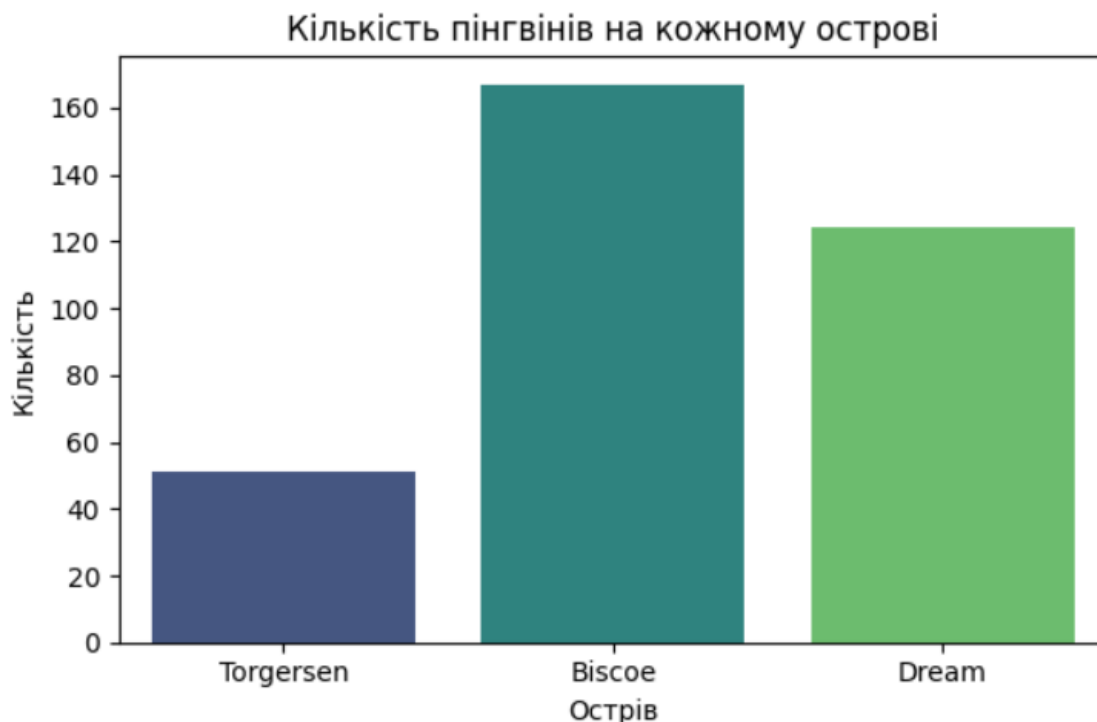
# 4.a) Залежність між довжиною дзьобу і ласт
plt.figure(figsize=(6, 5))
sns.scatterplot(data=df, x="culmen_length_mm", y="flipper_length_mm",
hue="species", palette="deep")
plt.title("Залежність: довжина дзьобу та ласт")
plt.xlabel("Довжина дзьобу (мм)")
plt.ylabel("Довжина ласт (мм)")
plt.tight_layout()
plt.show()

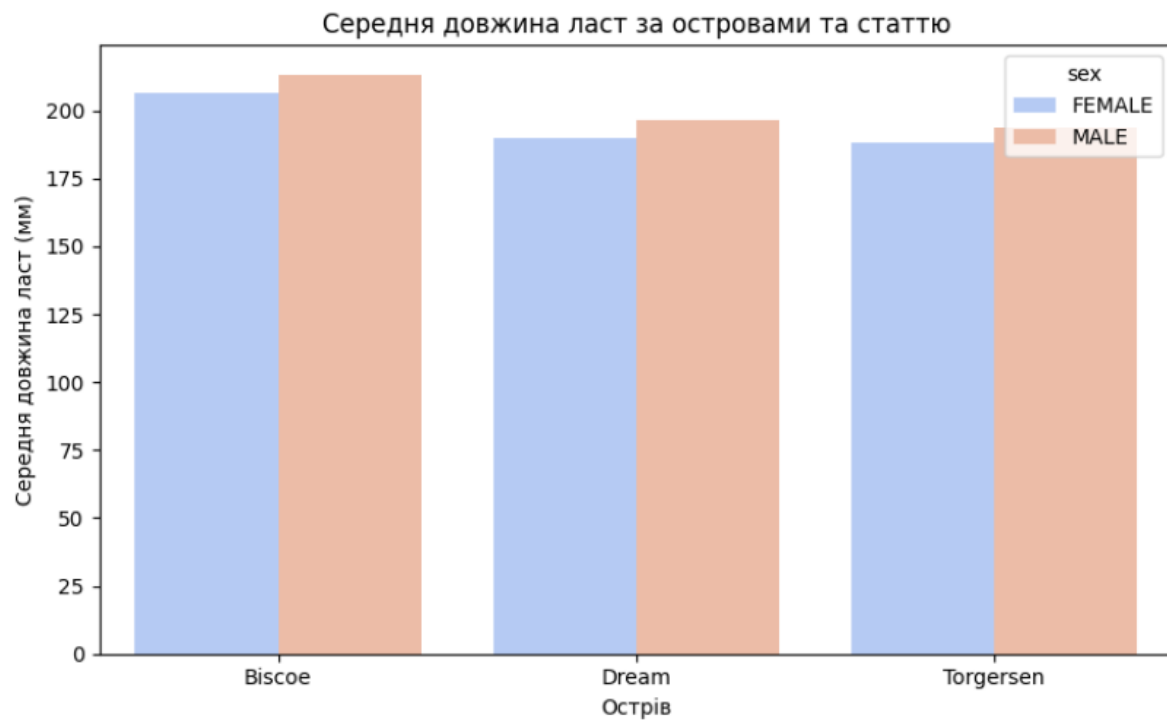
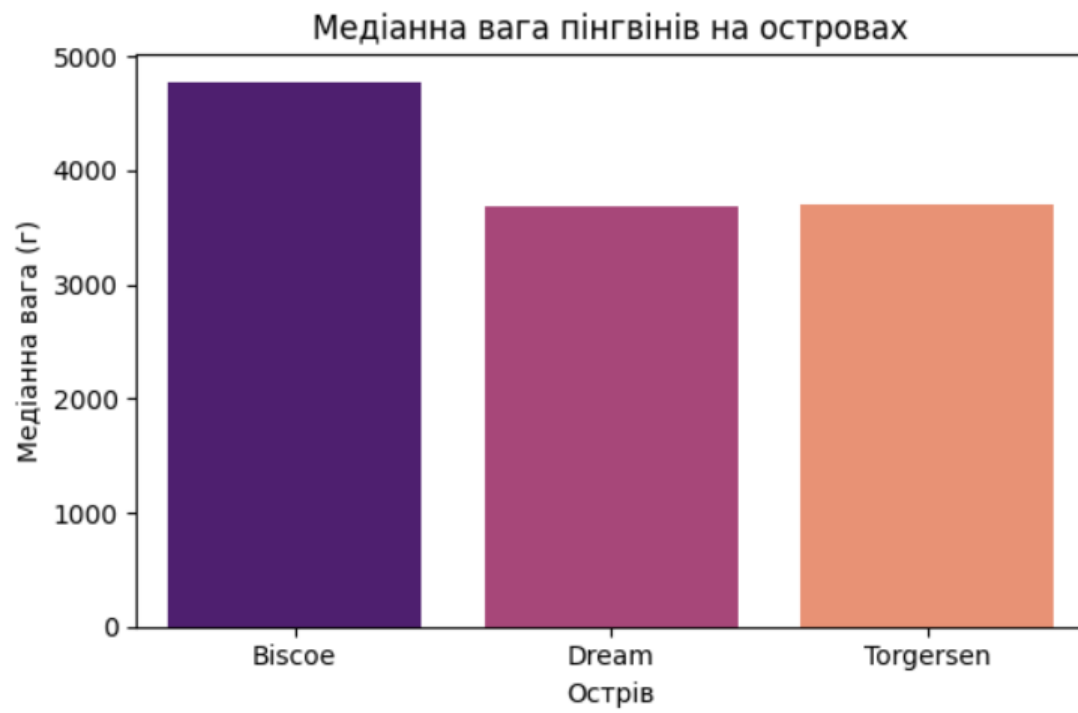
# 4.b) Залежність між вагою і довжиною дзьобу
plt.figure(figsize=(6, 5))
sns.scatterplot(data=df, x="culmen_length_mm", y="body_mass_g",
hue="species", palette="muted")
plt.title("Залежність: вага та довжина дзьобу")
plt.xlabel("Довжина дзьобу (мм)")
plt.ylabel("Вага (г)")
plt.tight_layout()
plt.show()

# Коефіцієнти кореляції
corr1 = df["culmen_length_mm"].corr(df["flipper_length_mm"])
corr2 = df["culmen_length_mm"].corr(df["body_mass_g"])

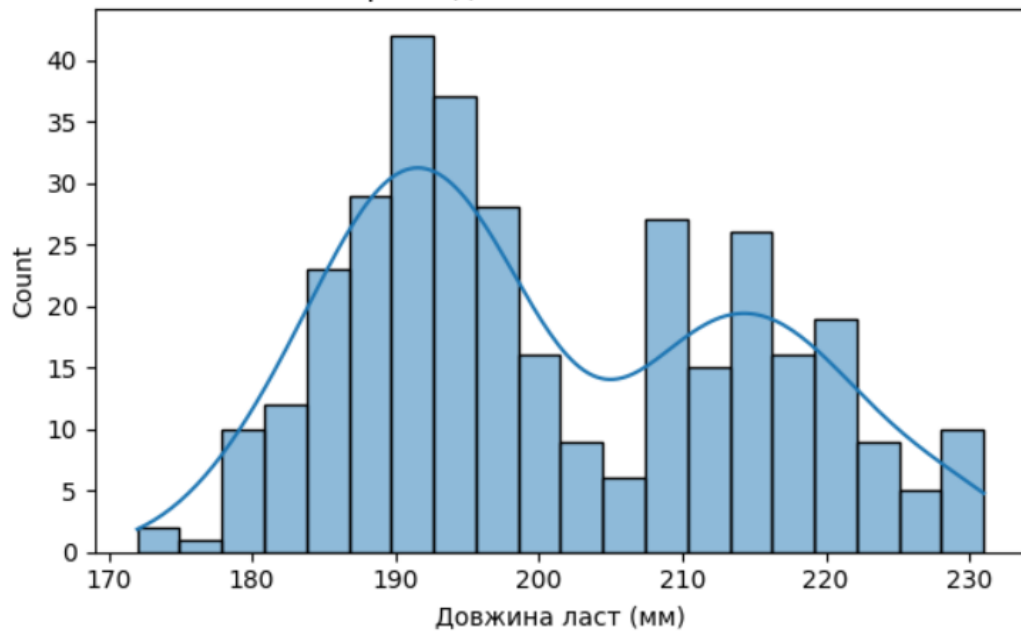
print(f"\nКоефіцієнт кореляції між довжиною дзьобу і ласт:
{corr1:.3f}")
print(f"Коефіцієнт кореляції між довжиною дзьобу і вагою: {corr2:.3f}")
```

Результат виконання:

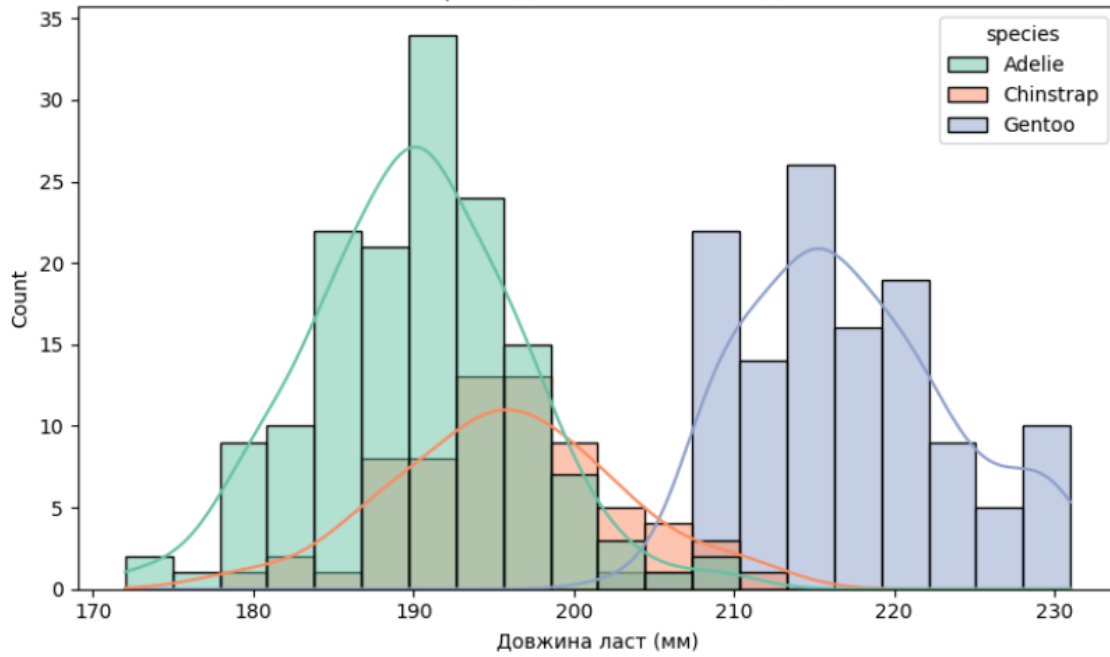




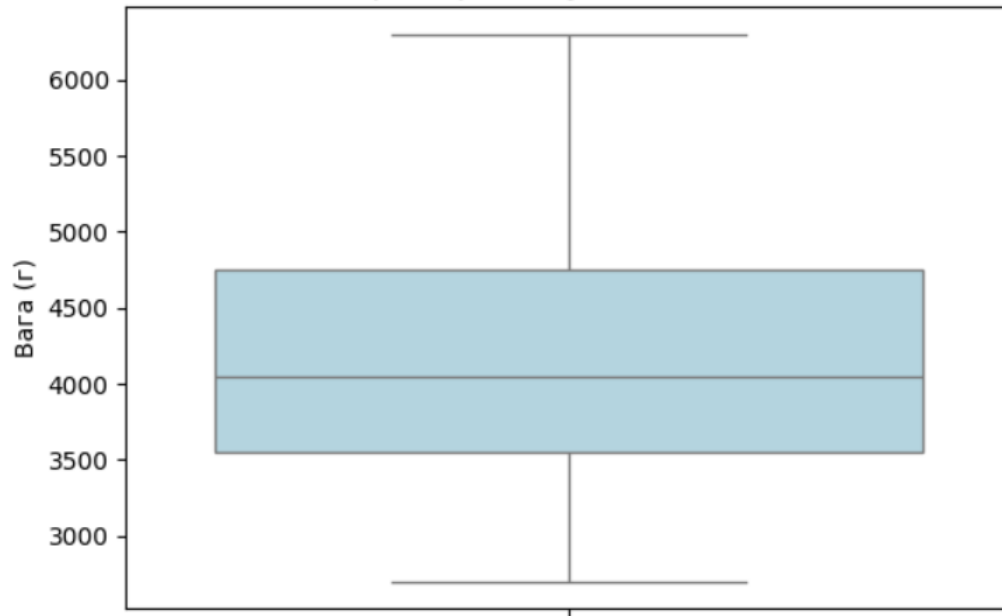
Гістограма довжини ласт (загальна)



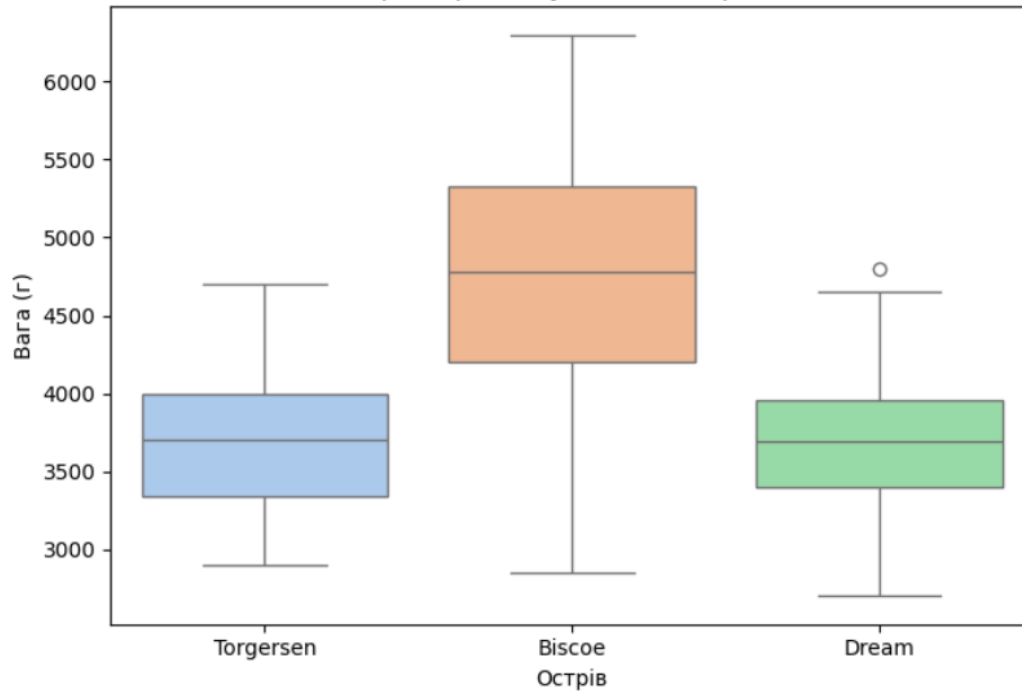
Гістограма довжини ласт за видами

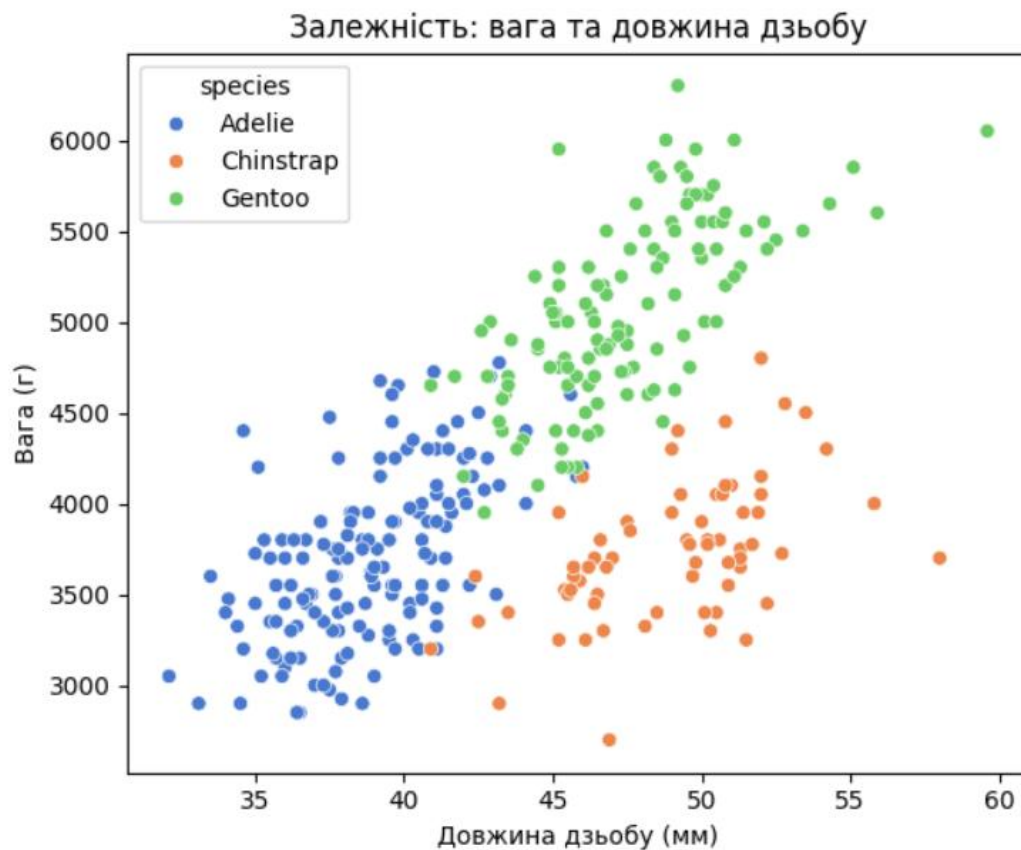
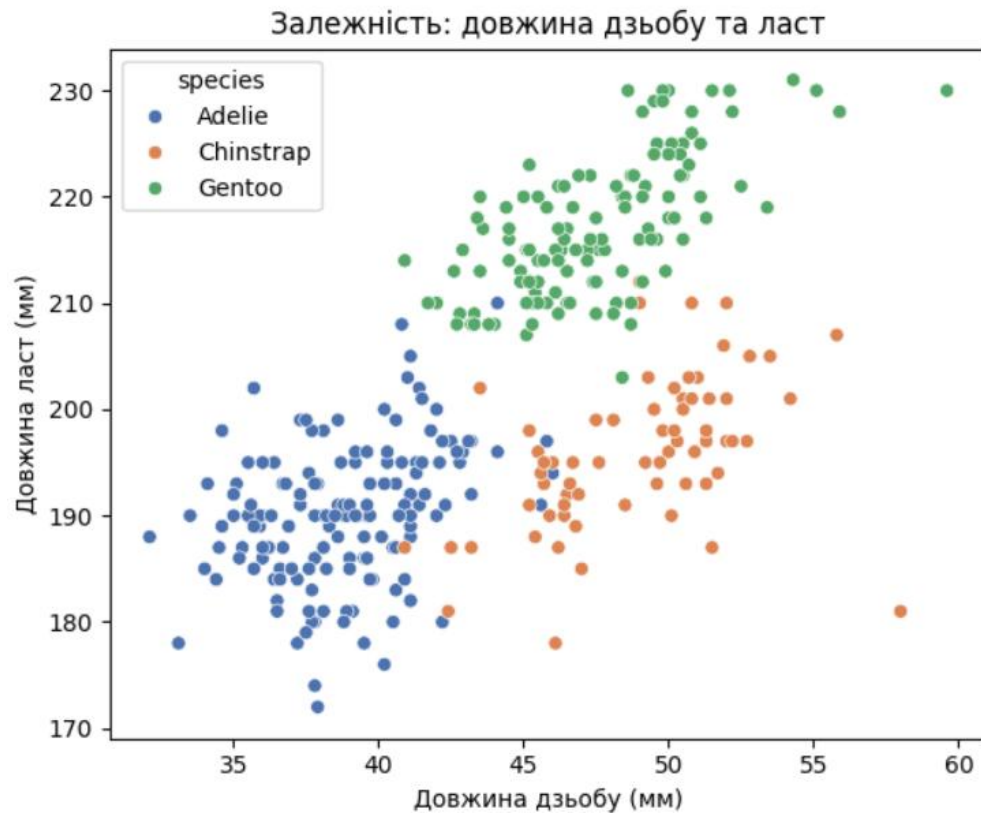


Діаграма розмаху ваги (загальна)



Діаграма розмаху ваги за островами





**Висновок:** У ході виконання даної лабораторної роботи я ознайомився з основними діаграмами та графіками, що використовуються при аналізі даних. Навчився будувати їх за допомогою бібліотек matplotlib та Seaborn.