

Київський національний університет
імені Тараса Шевченка

Звіт

до лабораторної роботи 12,
дисципліни: “Технології аналізу та візуалізації даних”,
студента: Євчика Олексія,
групи Інформатика, Магістри

1. Тема:

Візуалізація даних з використанням пакету ggplot2 програми R

2. Мета роботи:

Метою роботи є ознайомлення з побудовою графіків у пакеті ggplot2, заснованому на концепції *Grammar of Graphics*. Студент має навчитися створювати різні види візуалізацій, комбінувати естетики, шари (geoms), фасети та теми, а також обирати відповідні графічні елементи для аналізу структури та розподілу даних. Особлива увага приділяється побудові кривих густини для порівняння групових розподілів.

3. Теоретичні відомості

3.1. Принцип *Grammar of Graphics*

Grammar of Graphics — це концепція, що розглядає візуалізацію як комбінацію незалежних компонентів: даних, естетик, геометричних примітивів, статистичних перетворень, фасетів, шкал і тем. У ggplot2 кожен графік створюється шляхом послідовного додавання цих елементів у вигляді "шарів". Такий підхід дає гнучкість і стандартизований спосіб створення практично будь-якої візуалізації.

3.2. Естетики (aesthetics)

Естетики визначають, **які властивості даних** відображаються у графіку: x, y, color, fill, size, shape, alpha, linetype тощо.

Кожна естетика керує способом мапування змінної на графічний параметр.

3.3. Геоми (geoms)

Геом — це **геометрична форма**, яка показує дані:

geom_point(), geom_line(), geom_density(), geom_bar(), geom_boxplot(),

`geom_histogram()` тощо.

Кожен `geom` відображає дані по-своєму, тому вибір геоба залежить від аналітичної задачі.

3.4. Фасети (facets)

Фасети дозволяють **розділити дані на підграфіки**.

Використовуються функції:

- `facet_wrap(~ variable)`
- `facet_grid(row ~ column)`

Фасетування зручно для порівняння групових структур.

3.5. Теми (themes)

Теми налаштовують **візуальний стиль** графіка: шрифти, сітку, підписи, фон. Популярні теми:

`theme_minimal()`, `theme_bw()`, `theme_classic()`, `theme_light()`.

4. Хід роботи

Завдання: Density plot (варіант 6)

Мета

Оцінити густину розподілу та порівняти її між групами за допомогою накладених кривих.

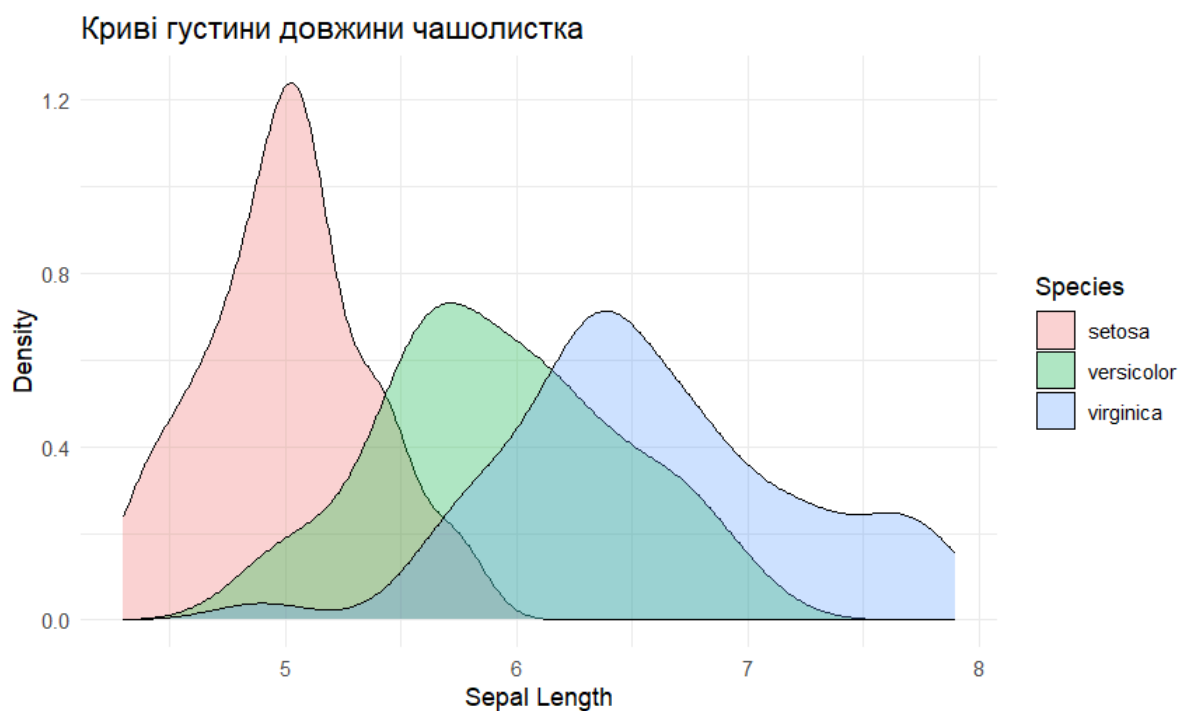
Код R

```

lab12_density.R* x
Source on Save
1 library(ggplot2)
2
3 data(iris)
4
5 ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
6   geom_density(alpha = 0.3) +
7   labs(
8     title = "Криві густини довжини чашолистка",
9     x = "Sepal Length",
10    y = "Density"
11  ) +
12  theme_minimal()
13

```

Отриманий графік



Коротка інтерпретація (2–5 речень)

Графік показує три накладені криві густини, які відповідають різним видам ірисів. Вид *setosa* характеризується меншими значеннями *Sepal.Length* та вузьким розподілом. *Versicolor* і *virginica* мають розподіли, зміщені праворуч, причому *virginica* демонструє найбільші середні

значення. Додавання $\alpha = 0.3$ дозволяє бачити значення всіх груп одночасно.

Виконаю додатково ще кілька варіантів.

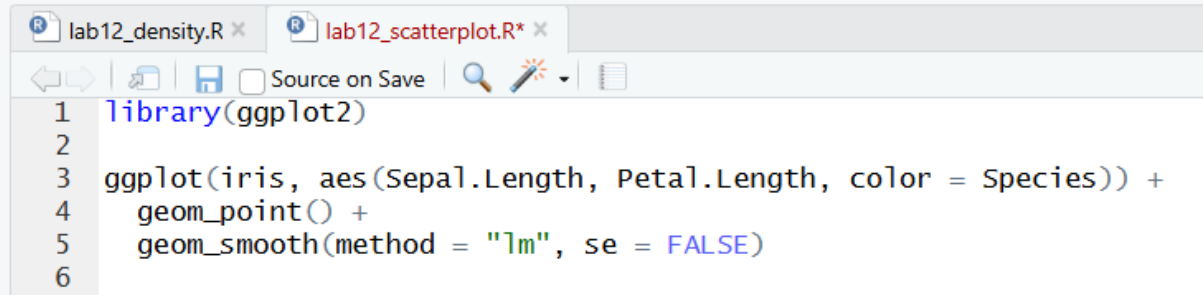
Варіант 2. Scatterplot + лінія тренду

Мета: показати взаємозв'язок між змінними.

Кроки: додати `geom_smooth(method="lm")`.

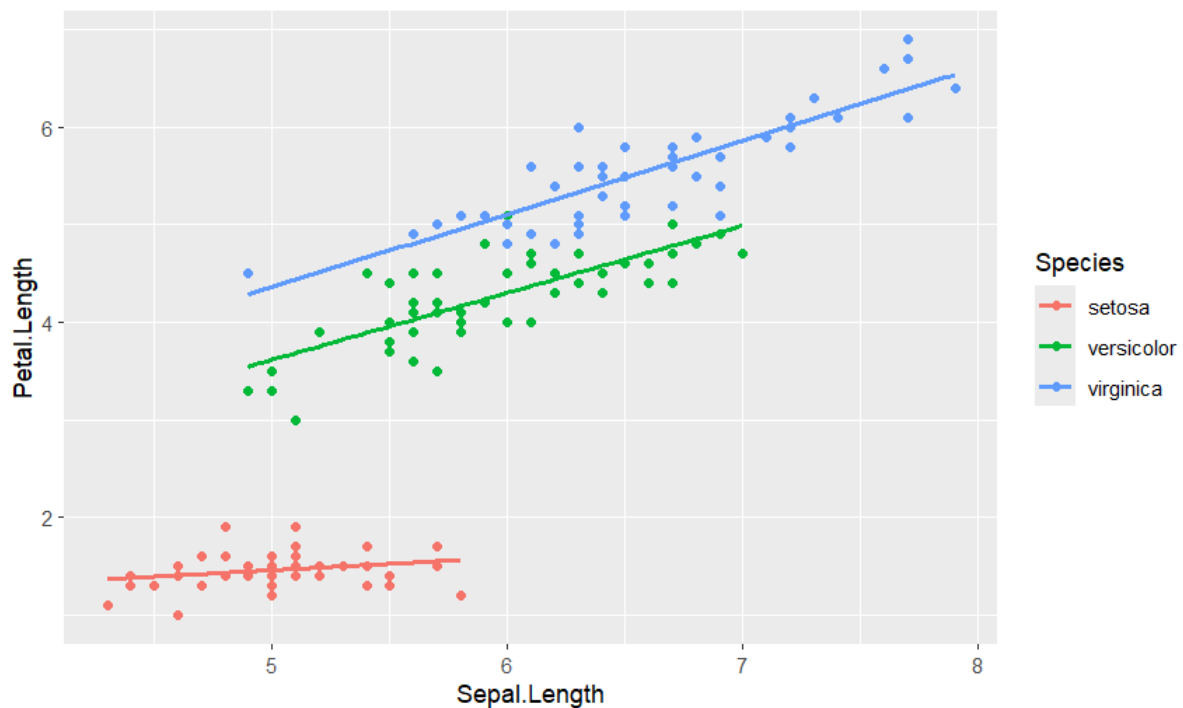
Результат: точкова діаграма + пряма тренду.

Код мовою R:



```
1 library(ggplot2)
2
3 ggplot(iris, aes(Sepal.Length, Petal.Length, color = Species)) +
4   geom_point() +
5   geom_smooth(method = "lm", se = FALSE)
6
```

Отримано:



Інтерпретація

Графік показує чітку позитивну залежність між довжиною чашолистка та довжиною пелюстки. Для всіх видів тренд майже лінійний, але нахил відрізняється між видами. Модель `lm` добре демонструє відмінності у співвідношенні змінних.

Варіант 3. Побудова `boxplot` за групами

Мета: дослідити розподіл по категоріях.

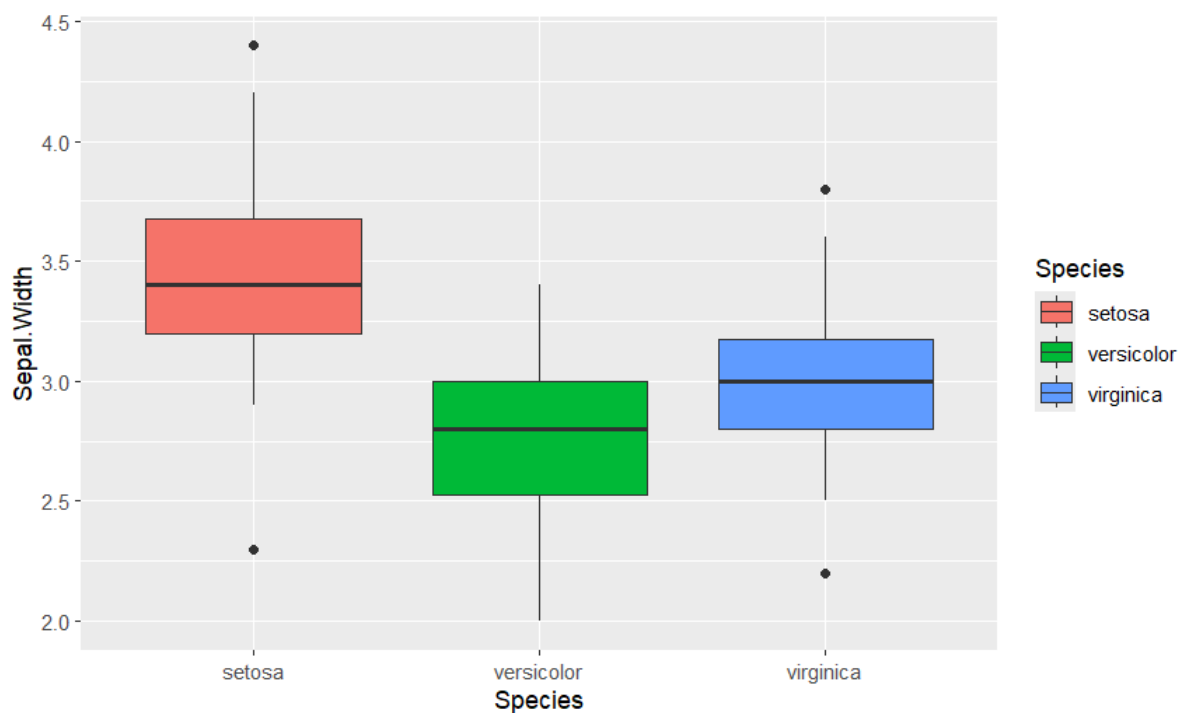
Кроки: `geom_boxplot()`.

Результат: `boxplot` для Setosa, Versicolor, Virginica.

Код

```
lab12_density.R x lab12_scatterplot.R x lab12_box_groups.R x
Source on Save
1 ggplot(iris, aes(Species, Sepal.Width, fill = Species)) +
2   geom_boxplot()
3
```

Отримано



Інтерпретація

Boxplot показує різницю в розподілі ширини чашолистка між видами. У Setosa медіана найбільша, а у Virginica — найменша. Видно також варіацію та наявність потенційних викидів.

Варіант 4. Violin plot

Мета: порівняти розподіли із збереженням форми.

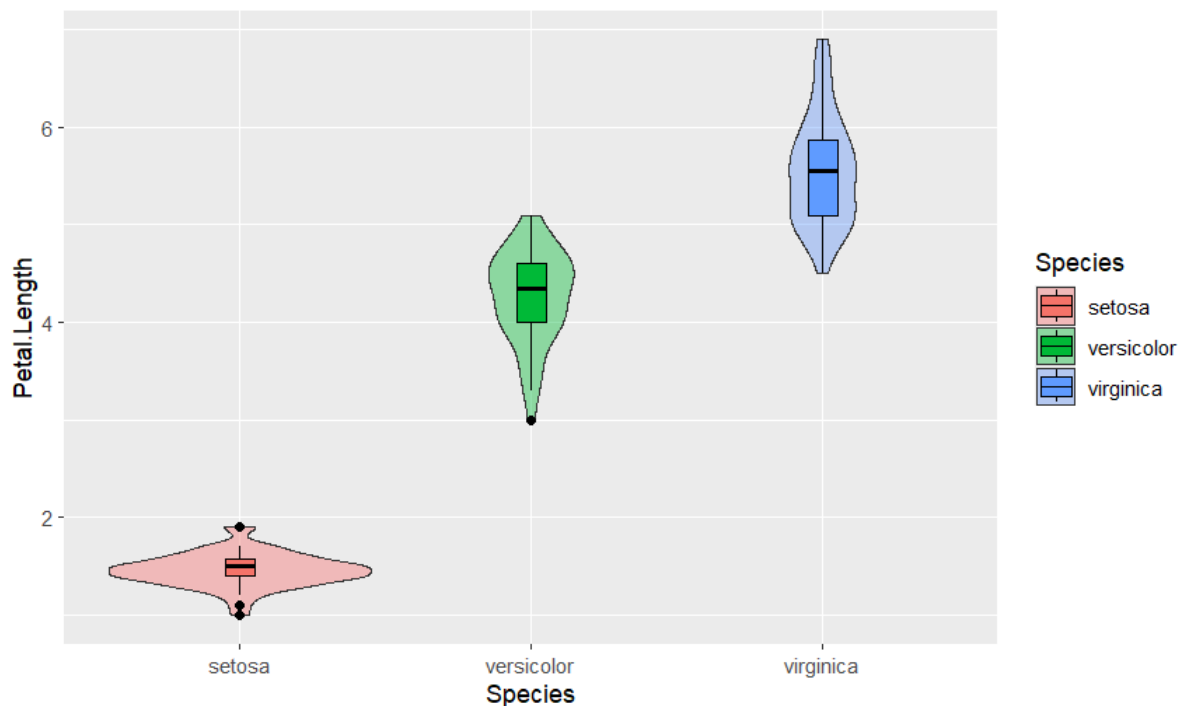
Кроки: `geom_violin(fill=..., alpha=...)`.

Результат: “скрипка” з медіаною.

Код

```
lab12_density.R x lab12_scatterplot.R x lab12_box_groups.R x lab12_violin_plot.R x
Source on Save
1 ggplot(iris, aes(Species, Petal.Length, fill = Species)) +
2   geom_violin(alpha = 0.4) +
3   geom_boxplot(width = 0.1, color = "black")
4 |
```

Отримано



Інтерпретація

Violin plot показує форму розподілу довжини пелюсток у кожного виду. Setosa має компактний і однорідний розподіл, тоді як Versicolor і Virginica демонструють ширший діапазон значень. Додавання boxplot допомагає точно бачити медіану та інтерквартильний розмах.

Варіант 5. Histogram

Мета: побудувати гістограму.

Кроки: `geom_histogram(bins=...)`.

Результат: гістограма ширини чашолистка.

Код

```
lab12_density.R x lab12_scatterplot.R x lab12_box_groups.R x lab12_violin_plot.R x lab12_histo.R x
Source on Save Run
1 ggplot(iris, aes(Sepal.Width)) +
2   geom_histogram(bins = 20, fill = "steelblue", color = "white")
3
```

Отримано



Інтерпретація

Гістограма показує розподіл ширини чашолистка в наборі даних. Розподіл близький до нормального, але має асиметрію зі зміщенням ліворуч. Видно основні моди, що відповідають різним видам квіток.

5. Висновки

У ході роботи було побудовано **density plot**, який дозволяє оцінити форму розподілу та порівняти його між кількома групами. Студент навчився застосовувати основні компоненти ggplot2: естетики, геомі, теми, а також комбінувати їх у шарову структуру. Найбільш корисними прийомами виявилися:

- використання естетики fill для порівняння груп,
- застосування geom_density() для аналізу розподілів,
- використання прозорості (alpha) для накладених графіків,
- застосування тем (theme_minimal()) для покращення читабельності.

6. Використана література

1. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer.
2. Wickham H., Grolemund G. R for Data Science. O'Reilly.
3. Chang W. R Graphics Cookbook. O'Reilly.
4. Kabacoff R. R in Action.
5. CRAN documentation: help(ggplot2), vignette("ggplot2-specs").