

Київський національний університет
імені Тараса Шевченка

Звіт

до лабораторної роботи 11,
дисципліни: “Технології аналізу та візуалізації даних”,
студента: Євчика Олексія,
групи Інформатика, Магістри

Тема:

Практичні приклади використання функцій пакету *lattice* для візуалізації даних

Мета роботи:

Метою роботи є ознайомлення з можливостями побудови багатовимірних та порівняльних графіків у середовищі R за допомогою пакету *lattice*. Студент має навчитися використовувати різні типи візуалізацій (*density plot*, *bwplot*, *barchart*, *contourplot*, *heatmap*) для аналізу даних, виявлення відмінностей між групами та структури розподілів. Робота спрямована на формування навичок вибору відповідного графіка залежно від аналітичної задачі.

Теоретична частина**Пакет *lattice***

lattice — це пакет у R, створений для побудови багатопанельних (*trellis*) графіків. Його головна особливість — декларативний підхід: користувач вказує, *що* потрібно зобразити (формулою), а вже *як* саме це буде оформлено — вирішує сама система *trellis graphics*.

Пакет дозволяє легко будувати графіки, розбиті на групи, категорії або панелі.

Основні види графіків у lattice та їх призначення

Графік	Функція	Призначення
Density plot	densityplot()	Порівняння розподілів між групами
Boxplot (BWplot)	bwplot()	Аналіз медіани, кuartилів, викидів
Dotplot	dotplot()	Ранжування та порівняння категорій
Histogram	histogram()	Оцінка форми розподілу
Barchart	barchart()	Порівняння значень між категоріями
Levelplot (heatmap)	levelplot()	Відображення двовимірних матриць/теплокарт
Contourplot	contourplot()	Рівневі лінії функції у 2D
XYplot	xyplot()	Тренди, лінійні й точкові діаграми
Stripplot	stripplot()	Альтернатива boxplot для порівняння значень

Хід виконання

Завдання: Density plot з overlay (варіант 6)

Мета

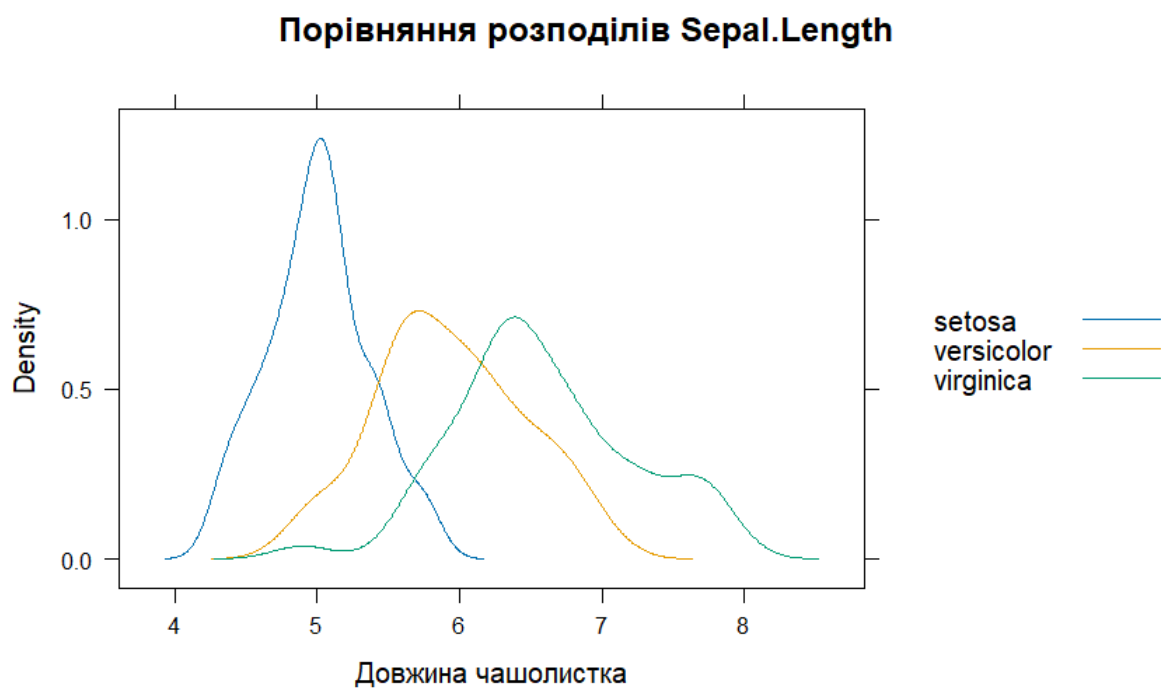
Порівняти розподіли числової змінної між категоріями за допомогою накладених кривих густини.

Як дані використаю простий вбудований датасет іріс.

Код мовою R:

```
lab11_density_plot.R x
Source on Save
1 library(lattice)
2
3 # Використовуємо вбудований набір даних iris
4 data("iris")
5
6 # Density plot з overlay
7 densityplot(~ Sepal.Length,
8             group = Species,
9             data = iris,
10            auto.key = TRUE,
11            plot.points = FALSE,
12            main = "Порівняння розподілів Sepal.Length",
13            xlab = "Довжина чашолистка")
14 |
```

Отримано:



Коротка інтерпретація

Графік відображає три криві густини, які відповідають трьом видам ірисів. Видно, що вид *setosa* має найбільш компактний розподіл і найменші значення. *Versicolor* та *virginica* демонструють зміщення розподілів праворуч, причому *virginica* загалом має найбільші значення. Такий графік

зручний для порівняння форм, ширини та положення розподілів різних груп.

Висновки

У ході роботи було розглянуто можливості пакету *lattice* для побудови багатопанельних графіків та порівняння групових розподілів. Студент навчився будувати density plot з накладанням кривих, що дозволяє ефективно аналізувати відмінності між розподілами. Найзручнішими у практиці виявилися графіки, що добре відображають групові відмінності: densityplot, bwplot та heatmap, оскільки вони чітко передають структуру даних і допомагають виявити статистичні закономірності.

Використана література:

1. Sarkar D. *Lattice: Multivariate Data Visualization with R*. Springer.
2. Deepayan Sarkar. *lattice* package documentation (CRAN).
3. Chang W. *R Graphics Cookbook*. O'Reilly.
4. Kabacoff R. *R in Action*. Manning.
5. Venables W., Ripley B. *Modern Applied Statistics with S*.
6. R documentation: help(lattice), vignette("lattice").