

Київський національний університет
імені Тараса Шевченка

Звіт

до лабораторної роботи 10 (додаткове завдання),
дисципліни: “Технології аналізу та візуалізації даних”,

студента: Євчика Олексія,
групи Інформатика, Магістри

Тема:

Візуалізація даних з використанням пакету lattice у R

Мета роботи:

Побудувати теплокарту, що показує зміну значень у двох вимірах — наприклад, за місяцями та роками. Метод дозволяє візуально виявити області з високою та низькою активністю.

Теоретичні відомості

Система trellis graphics

Trellis graphics — це система багатопанельних графіків для аналізу багатовимірних даних, реалізована у пакеті lattice у R.

Основна ідея: графік автоматично розбивається на “решітку” (trellis) панелей, кожна з яких відповідає одній підгрупі даних.

Це дозволяє легко досліджувати залежності між змінними у різних категоріях, часових інтервалах чи рівнях факторів.

Особливості trellis graphics:

- автоматичне фасетування за умовами (|),
- єдина шкала для всіх панелей (за замовчуванням),
- декларативний стиль: спочатку описуються що візуалізувати, а не як.

Основні функції пакету lattice

Функція	Призначення
xyplot()	Діаграма розсіювання, лінійні графіки
bwplot()	Boxplot у trellis-стилі
dotplot()	Dotplot — точкове порівняння категорій

densityplot()	Оцінка щільності розподілу
histogram()	Гістограма
levelplot()	Теплокарта (heatmap)
contourplot()	Контурні графіки
barchart()	Стовпчикові діаграми

Пакет використовує формули типу:

$y \sim x \mid \text{condition}$

що означає «побудувати графік y від x окремо для кожного рівня condition ».

Хід роботи

Завдання: Heatmap (levelplot)

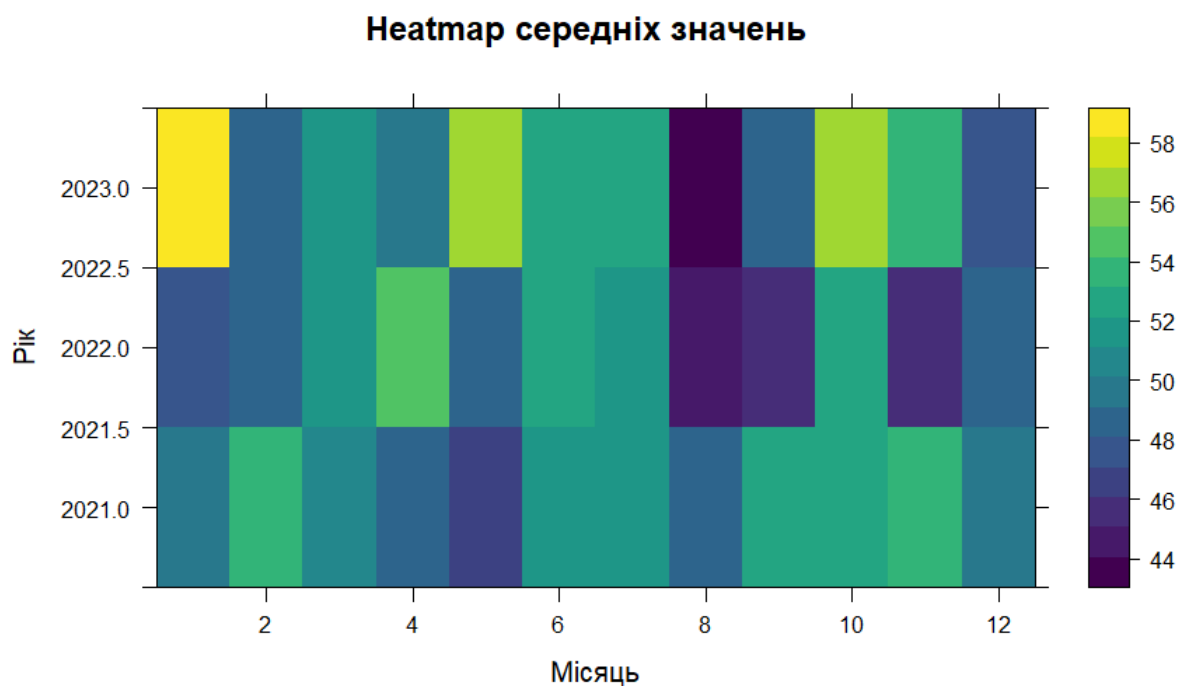
Мета

Побудувати теплокарту, що показує зміну значень у двох вимірах — наприклад, за місяцями та роками. Метод дозволяє візуально виявити області з високою та низькою активністю.

Код мовою R

```
lab10_dotplot.R x lab10_heatmap.R x
Source on Save
1 library(lattice)
2 library(tidyverse)
3 library(lubridate)
4
5 # Прикладові дані
6 set.seed(123)
7 df_hm <- tibble(
8   date = seq(as.Date("2021-01-01"), as.Date("2023-12-31"), by = "day"),
9   value = abs(rnorm(length(date), mean = 50, sd = 20))
10 )
11
12 # Агрегування за місяцем та роком
13 hm_data <- df_hm %>%
14   mutate(
15     year = year(date),
16     month = month(date)
17   ) %>%
18   group_by(year, month) %>%
19   summarise(mean_value = mean(value), .groups = "drop")
20
21 # Heatmap
22 levelplot(mean_value ~ month * year, data = hm_data,
23           col.regions = colorRampPalette(viridisLite::viridis(20)),
24           xlab = "Місяць", ylab = "Рік",
25           main = "Heatmap середніх значень")
26 |
```

Отримано графік



Коротка інтерпретація

Теплокарта показує зміни середніх значень між місяцями та роками. Світліші тони відповідають вищим значенням, темніші — нижчим. Видно,

що деякі місяці мають стабільно підвищені показники, тоді як інші зберігають низьку активність. Такий тип графіка добре підходить для виявлення сезонних патернів.

Завдання 2. Contourplot

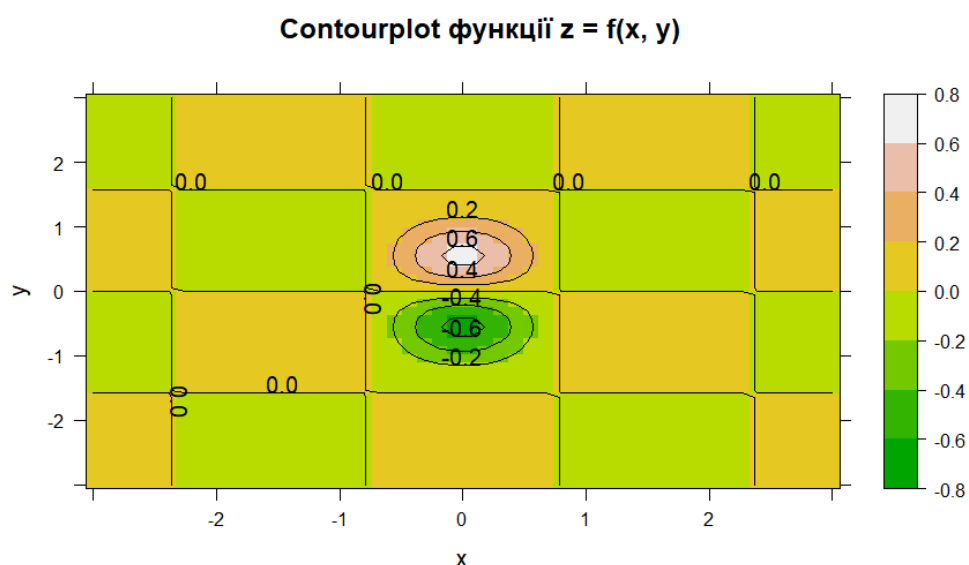
Мета:

Побудувати контурну діаграму для відображення рівнів значень у двохвимірному просторі. Метод дозволяє побачити “лінії однакових значень” і локальні максимуми/мінімуми.

Код

```
lab10_dotplot.R x lab10_heatmap.R x lab10_counturplot.R x
1 library(lattice)
2
3 # Штучні дані для contourplot
4 x <- seq(-3, 3, length = 50)
5 y <- seq(-3, 3, length = 50)
6 grid <- expand.grid(x = x, y = y)
7 grid$z <- with(grid, exp(-(x^2 + y^2)) * cos(x*2) * sin(y*2))
8
9 # Contourplot
10 contourplot(z ~ x * y, data = grid,
11             region = TRUE,
12             col.regions = terrain.colors(100),
13             main = "Contourplot функції  $z = f(x, y)$ ",
14             xlab = "x", ylab = "y")
15
```

Отримано:



Коротка інтерпретація

Контурний графік демонструє поведінку функції у двовимірному просторі. Регіони різного кольору показують рівні значень змінної **Z**, а лінії рівня позначають області з однаковим значенням. Видно наявність “хвильових” структур та симетрії. Contourplot зручний для аналізу складних поверхонь та градієнтів.

Завдання 3. Barchart

Мета:

Побудувати порівняльну стовпчикову діаграму для різних категорій. Метод дозволяє оцінити різницю між групами у числових показниках.

Код R

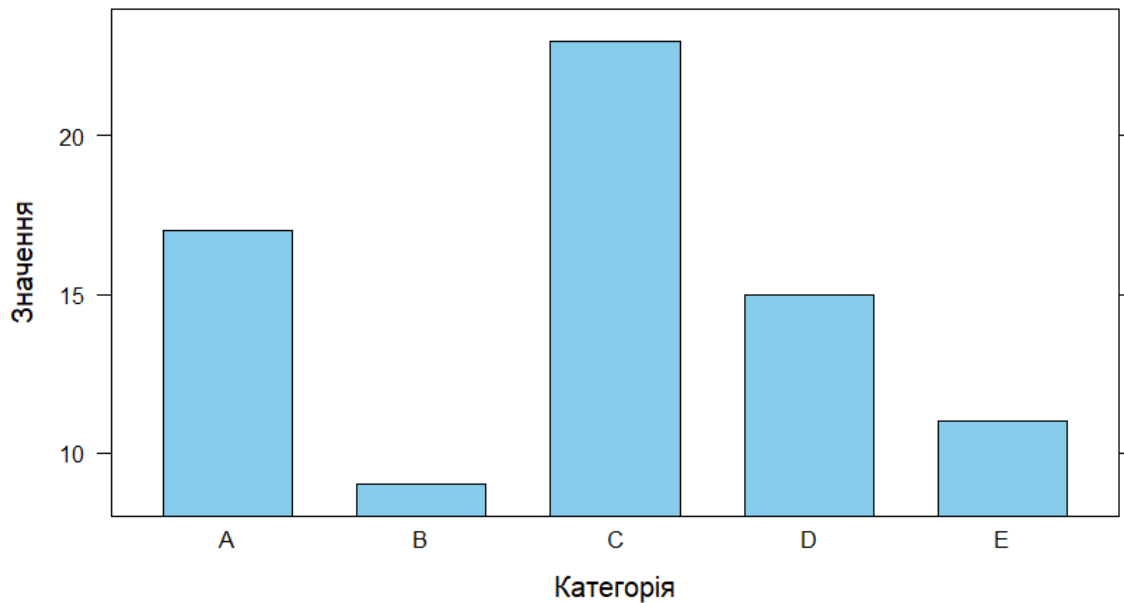
```
library(lattice)
library(tidyverse)

# Дані
df_bar <- tibble(
  Category = c("A", "B", "C", "D", "E"),
  Value = c(17, 9, 23, 15, 11)
)

# Barchart
barchart(Value ~ Category, data = df_bar,
  col = "skyblue",
  main = "Порівняльна стовпчикова діаграма",
  xlab = "Категорія", ylab = "Значення")
|
```

Отримано:

Порівняльна стовпчикова діаграма



Коротка інтерпретація

Стовпчикова діаграма показує, як значення змінюються між категоріями. Найбільший показник спостерігається у групи “С”, тоді як група “В” має найнижче значення. Bar chart дозволяє швидко оцінити структуру та ранжування категорій. Цей тип графіка добре підходить для порівняння груп у якісних змінних.

Завдання 4. BWplot (boxplot)

Мета

Побудувати boxplot, що відображає розподіл числових даних у кількох групах, включаючи медіану, квантілі та потенційні викиди.

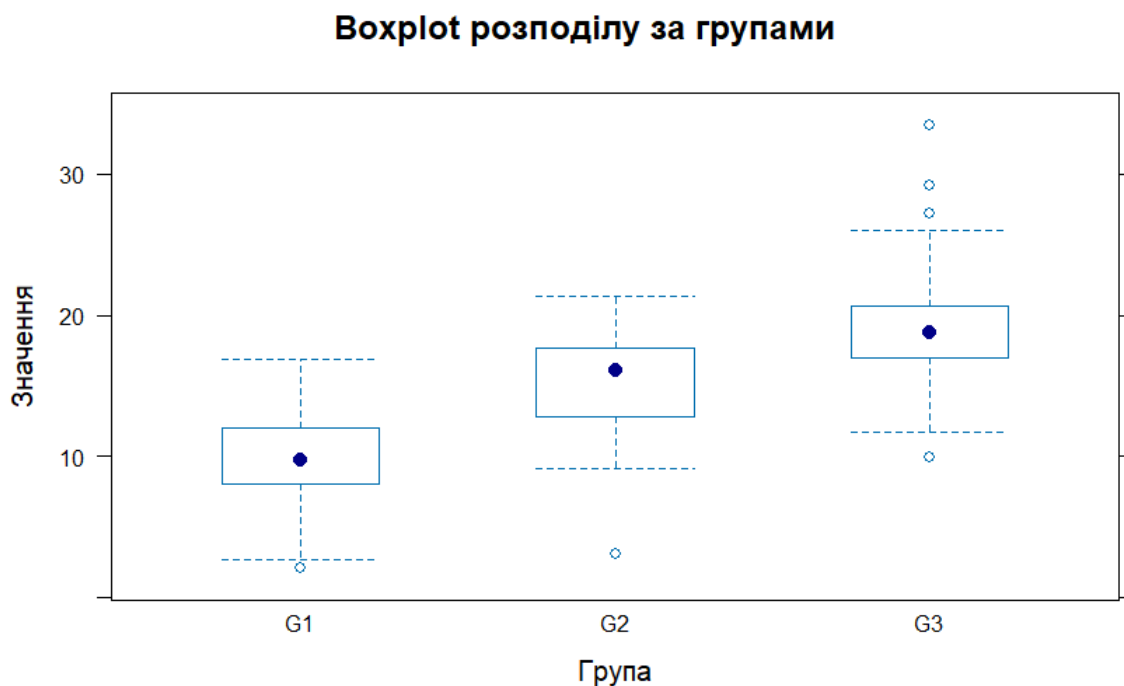
Код мовою R

```

1 library(lattice)
2 library(tidyverse)
3
4 # Дані для boxplot
5 set.seed(42)
6 df_box <- tibble(
7   Group = rep(c("G1", "G2", "G3"), each = 50),
8   Value = c(
9     rnorm(50, 10, 3),
10    rnorm(50, 15, 4),
11    rnorm(50, 20, 5)
12  )
13 )
14
15 # Bwplot
16 bwplot(Value ~ Group, data = df_box,
17         main = "Boxplot розподілу за групами",
18         xlab = "Група", ylab = "Значення",
19         col = "darkblue")
20

```

Отримано:



Коротка інтерпретація

Boxplot демонструє розподіл значень у кожній групі. Група G3 має найбільшу медіану та найширший розкид даних, тоді як G1 має найнижчі значення та меншу варіативність. Окремі точки можуть показувати

наявність викидів. Такий графік дозволяє легко порівнювати розподіли між кількома групами.

Висновки

У ході роботи я(як студент) навчився будувати багатопанельні графіки у стилі trellis graphics, застосовувати фасетування та працювати з категоріальними і числовими змінними у пакеті *lattice*. Розглянуті методи — dotplot, heatmap (levelplot), contourplot — дозволяють ефективно досліджувати структурні властивості даних, порівнювати групи та виявляти закономірності. Отримані навички є базовими для аналізу багатовимірних та часових даних.

Використана література

1. Sarkar D. *Lattice: Multivariate Data Visualization with R*. Springer.
2. Deepayan Sarkar. *lattice* documentation (CRAN).
3. Wickham H. *Elegant Graphics for Data Analysis*.
4. Venables W., Ripley B. *Modern Applied Statistics with S*.
5. Chang W. *R Graphics Cookbook*.
6. R Documentation: `help(lattice)`, `vignette("lattice")`.