

Київський національний університет
імені Тараса Шевченка
Звіт
до лабораторної роботи 8,
дисципліни: “Технології аналізу та візуалізації даних”,
студента: Євчика Олексія,
групи Інформатика, Магістри

Тема:

Тривимірні діаграми. Матриці діаграм розсіювання. Корелограми.

Мозаїчні діаграми. Карти дерева

Мета роботи:

Ознайомлення з основними видами багатовимірних графічних візуалізацій у середовищі R, зокрема з тривимірними графіками, матрицями діаграм розсіювання, корелограмами, мозаїчними діаграмами та картами дерева. Набуття практичних навичок побудови, налаштування, інтерпретації таких графіків і застосування їх для аналізу багатовимірних даних.

1. Теоретична довідка

1.1 Тривимірні діаграми (3D Charts)

Тривимірні діаграми — це графічні візуалізації даних у просторі з трьома вимірами (x , y , z), які дозволяють одночасно відобразити залежність між трьома змінними.

Основні типи 3D-діаграм:

- 3D Scatterplot — тривимірна діаграма розсіювання для візуалізації залежностей між трьома кількісними змінними.
- 3D Surface plot — показує поверхню, що утворюється значеннями функції $z = f(x, y)$.
- 3D Wireframe — каркасна модель поверхні.
- 3D Barplot — стовпчаста діаграма з глибиною.

Переваги:

- можливість бачити більш складні залежності між змінними;
- інтуїтивне уявлення про форму поверхні або кластери.

Недоліки:

- перевантаженість візуалізації;
- складність оцінки значень через перспективу;
- інколи зменшення читабельності порівняно з 2D.

1.2. Матриці діаграм розсіювання (Scatterplot Matrix, SPLOM)

Матриця діаграм розсіювання — це таблиця, у якій кожен стовпець і кожен рядок відповідає змінній, а на перетині будеться діаграма розсіювання між двома змінними.

Призначення:

- дослідити усі попарні залежності між змінними;
- виявити лінійні та нелінійні зв'язки;
- знайти викиди, кластери та кореляції.

Особливості:

- по діагоналі часто розташовуються гістограми або щільності окремих змінних;
- використовується переважно для багатовимірних наборів даних ($p \leq 10-12$).

Переваги:

- повна картина взаємодій між змінними;
- добра основа для попереднього аналізу даних (EDA).

1.3. Корелограми (Correlogram)

Корелограма — це графічне відображення кореляційної матриці між змінними.

Формати корелограм:

- Кольорові матриці (heatmap), де колір відповідає значенню кореляції;
- Кола різного розміру, де радіус і колір показує силу та знак кореляції;
- Числові кореляції з кольоровим анотуванням;
- Комбіновані (графічні + числові).

Опис та використання:

- дозволяє швидко оцінити, які змінні сильно корельовані;
- допомагає виявити мультиколінеарність у моделях;
- корисна у Feature Engineering та Feature Selection.

Плюси:

- висока інформативність;
- проста інтерпретація.

1.4. Мозаїчні діаграми (Mosaic Plot)

Мозаїчна діаграма — це спосіб візуалізації залежності між категоріальними змінними, де площа прямокутників пропорційна частоті комбінацій рівнів факторів.

Як побудована:

- перша змінна ділить площу на великі блоки;
- друга — ділить кожен блок на підблоки;
- третя — аналогічно (можливо декілька рівнів вкладеності).

Інтерпретація:

- ширина прямокутника відповідає частоті рівня першої змінної;
- висота — частоті рівня другої змінної в межах першої;
- колір часто показує відхилення від незалежності.

Застосування:

- аналіз таблиць спряженості;
- оцінка статистичної незалежності категоріальних змінних;
- використовується у медичних, соціологічних та економічних дослідженнях.

1.5. Карты дерева (Treemap)

Treemap — це спосіб представлення ієрархічних або вкладених категоричних даних у вигляді прямокутників, площа яких відповідає числовому значенню.

Механіка роботи:

- кореневий елемент ділиться на прямокутники згідно з вагами;
- кожен прямокутник може містити вкладені прямокутники (деревоподібна структура);
- колір може відображати ще один параметр (групу або величину).

Переваги:

- чудово працює для великої кількості категорій;
- дозволяє одразу порівнювати розмір та вклад різних груп;
- дуже ефективний для фінансових ринків, бюджетів, структури дисків.

Недоліки:

- важко точніше зчитувати розміри (особливо коли співвідношення сильне);
- менш інтуїтивні у порівнянні з бар-чартами при невеликій кількості категорій.

Індивідуальне завдання (варіант 6)

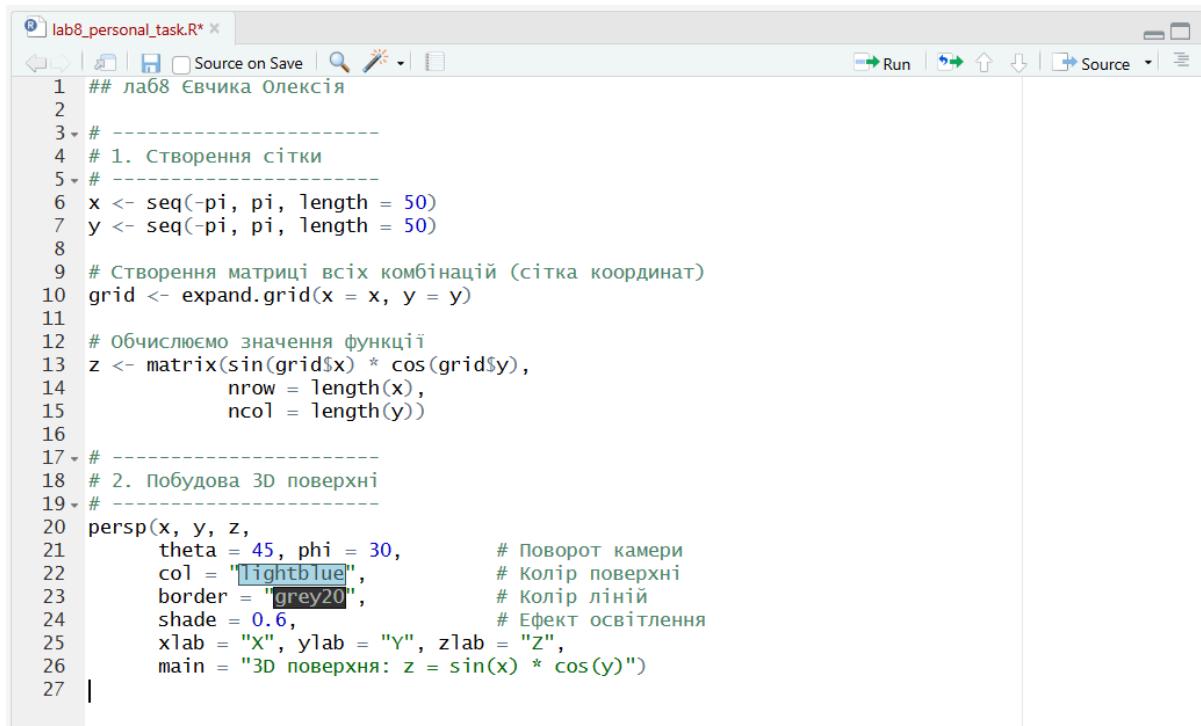
Завдання:

Побудувати 3D поверхню функції $z = \sin(x) * \cos(y)$ за допомогою persp().

1. Мета

- ознайомитися з функцією persp() у R;
- навчитися створювати 3D-поверхні;
- зробити інтерпретацію отриманої синусоїдальної поверхні.

2. Код для побудови 3D-поверхні

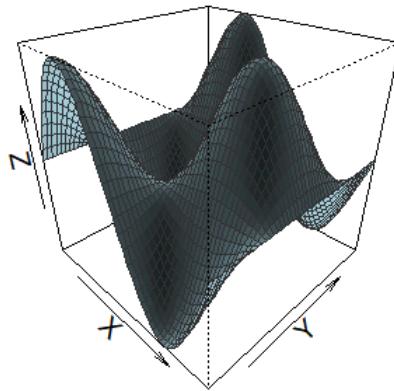


The screenshot shows the RStudio interface with the code editor open. The file is named 'lab8_personal_task.R'. The code generates a 3D surface plot of the function $z = \sin(x) * \cos(y)$. It starts by creating two vectors of x and y values from -pi to pi, each with 50 points. It then creates a grid of all combinations of these points. The z values are calculated as the matrix product of sin(x) and cos(y). Finally, it uses the persp() function to create a 3D perspective plot with specific parameters like theta and phi angles, and various colors and shading effects.

```
## лаб8 Євчика Олексія
# 1. Створення сітки
# 2. Побудова 3D поверхні
# Обчислюємо значення функції
# Створення матриці всіх комбінацій (сітка координат)
grid <- expand.grid(x = x, y = y)
# 2. Побудова 3D поверхні
persp(x, y, z,
      theta = 45, phi = 30,          # Поворот камери
      col = "lightblue",            # Колір поверхні
      border = "grey20",             # Колір ліній
      shade = 0.6,                  # Ефект освітлення
      xlab = "X", ylab = "Y", zlab = "Z",
      main = "3D поверхня: z = sin(x) * cos(y)")
```

Результат:

3D поверхня: $z = \sin(x) * \cos(y)$



3. Пояснення змінних

- x, y — рівномірно розподілені значення у діапазоні $[-\pi, \pi]$.
- z — значення функції $\sin(x) \cdot \cos(y)$ для кожної точки сітки.
- `persp()` — базова функція R для 3D-візуалізації поверхонь.

4. Що показує графік (інтерпретація)

- Отримана поверхня — це об'ємна синусоїdalна хвиля, що коливається у двох напрямках: по осі X та по осі Y.
- Форма нагадує хвилясту поверхню, де:
 - максимуми виникають коли $\sin(x)=1$ і $\cos(y)=1$,
 - мінімуми — коли значення наближаються до -1.
- Сітка створює регулярні гребені та западини, які утворюють періодичний патерн.

5. Налаштування параметрів (можна порівняти)

Кут огляду

theta = 20; phi = 40

Колір та освітлення

col = "tomato"

shade = 0.3

Густота сітки

x <- seq(-pi, pi, length = 100)

y <- seq(-pi, pi, length = 100)

6. Збереження графіка

PNG

png("surface.png", width = 800, height = 600)

persp(x, y, z,

theta = 45, phi = 30,

col = "lightblue", shade = 0.6)

dev.off()

PDF

pdf("surface.pdf")

persp(x, y, z,

theta = 45, phi = 30,

col = "lightblue", shade = 0.6)

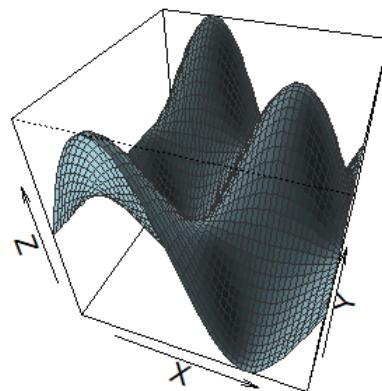
dev.off()

7. Результат

Побудована **об'ємна синусоїдальна поверхня** $z=\sin(x) \cdot \cos(y)$, яка демонструє періодичні хвилі в обох напрямках — X та Y.

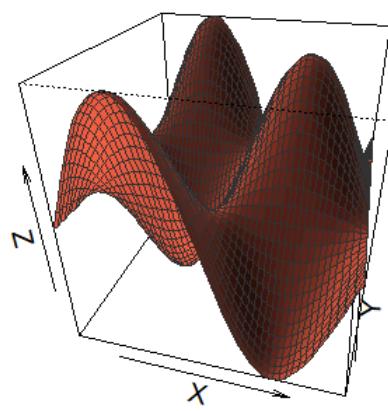
Поворот камери:

3D поверхня: $z = \sin(x) * \cos(y)$



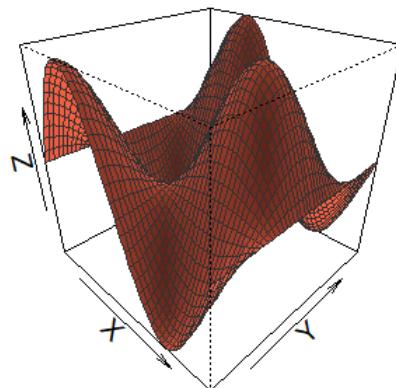
Зміна кольору:

3D поверхня: $z = \sin(x) * \cos(y)$



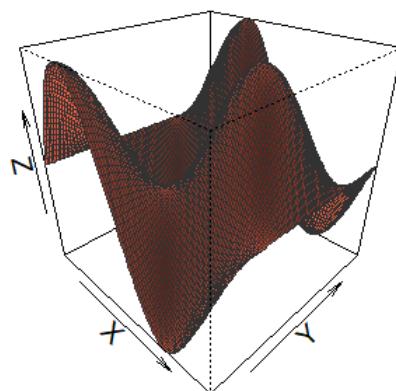
Зменшення тіні

3D поверхня: $z = \sin(x) * \cos(y)$



Збільшення сітки:

3D поверхня: $z = \sin(x) * \cos(y)$



Можемо збільшити об'єм

```

3 # -----
4 # 1. Створення сітки
5 # -----
6 x <- seq(-3*pi, 3*pi, length = 100)
7 y <- seq(-3*pi, 3*pi, length = 100)
8
9 # Створення матриці всіх комбінацій (сітка координат)
10 grid <- expand.grid(x = x, y = y)
11
12 # Обчислюємо значення функції
13 z <- matrix(sin(grid$x) * cos(grid$y),
14               nrow = length(x),
15               ncol = length(y))
16
17 # -----
18 # 2. Побудова 3D поверхні
19 # -----
20 persp(x, y, z,
21        theta = 45, phi = 30,           # Поворот камери
22        expand = 0.3,
23        col = "tomato",                # Колір поверхні
24        border = "grey20",              # Колір ліній
25        shade = 0.3,                   # Ефект освітлення
26        xlab = "X", ylab = "Y", zlab = "Z",
27        main = "3D поверхня: z = sin(x) * cos(y)")
28

```

3D поверхня: $z = \sin(x) * \cos(y)$

