

Євчик Олексій, Варіант 6, ІНФ-2

1. Вступ (Завдання):

Згенерувати дані про книги в бібліотеці (назва, автор, жанр, рік видання, кількість сторінок). Створити кругову діаграму розподілу книг по жанрах та коробкову діаграму кількості сторінок по жанрах.

2. Методологія (опис використаних методів візуалізації)

У цьому завданні датасет потрібно згенерувати самостійно, тому я створив штучний набір даних із полями:

- Назва книги
- Автор
- Жанр
- Рік видання
- Кількість сторінок

Нижче наведений R-код, який генерує набір з 100 книг випадковим чином, але в реалістичних межах.

```
#####  
# 1. Генерація даних  
#####  
  
set.seed(123) # для відтворюваності результатів  
  
# Створюємо вектор можливих жанрів  
genres <- c("Фантастика", "Детектив", "Роман", "Наукова література", "Дитяча")  
  
# Генеруємо таблицю з випадковими даними  
books <- data.frame(  
  Назва = paste("Книга", 1:100),  
  Автор = paste("Автор", sample(1:50, 100, replace = TRUE)),  
  Жанр = sample(genres, 100, replace = TRUE),  
  Рік_видання = sample(1950:2024, 100, replace = TRUE),  
  Кількість_сторінок = sample(80:800, 100, replace = TRUE)  
)
```

Цей код:

- створює **реалістичний набір даних про 100 книг**,
- випадково розподіляє авторів, жанри, роки і кількість сторінок,

- забезпечує різноманітність значень для подальшого аналізу,
- залишає структуру такою, що її легко очищати та візуалізувати.

Далі було проведено первинну обробку й очищення даних

```
33 - #####
34 # 2. Первинна обробка та очищення
35 - #####
36
37 # Перевіримо на наявність пропусків
38 sum(is.na(books)) # має бути 0
39
40 # Приклад очищення – фільтруємо тільки книги після 1970 року
41 books_clean <- books %>% filter(Рік_видання > 1970)
42
```

й використано базові способи побудови графіків

```
44 # 3. Візуалізації
45 - #####
46
47 # 3.1. Кругова діаграма розподілу книг по жанрах
48 genre_count <- books_clean %>%
49   count(Жанр)
50
51 pie_chart <- ggplot(genre_count, aes(x = "", y = n, fill = Жанр)) +
52   geom_col(width = 1) +
53   coord_polar(theta = "y") +
54   labs(
55     title = "Розподіл книг за жанрами",
56     fill = "Жанр"
57   ) +
58   theme_void()
59
60 # 3.2. Коробкова діаграма (boxplot) кількості сторінок по жанрах
61 box_plot <- ggplot(books_clean, aes(x = Жанр, y = Кількість_сторінок, fill = Жанр)) +
62   geom_boxplot(alpha = 0.7) +
63   labs(
64     title = "Розподіл кількості сторінок за жанрами",
```

3. Результати (представлення створених графіків з коментарями)

Коробкова діаграма

Коробкова діаграма показує статистичний розподіл кількості сторінок у книгах кожного жанру, а саме:

Для кожного жанру окремо видно:

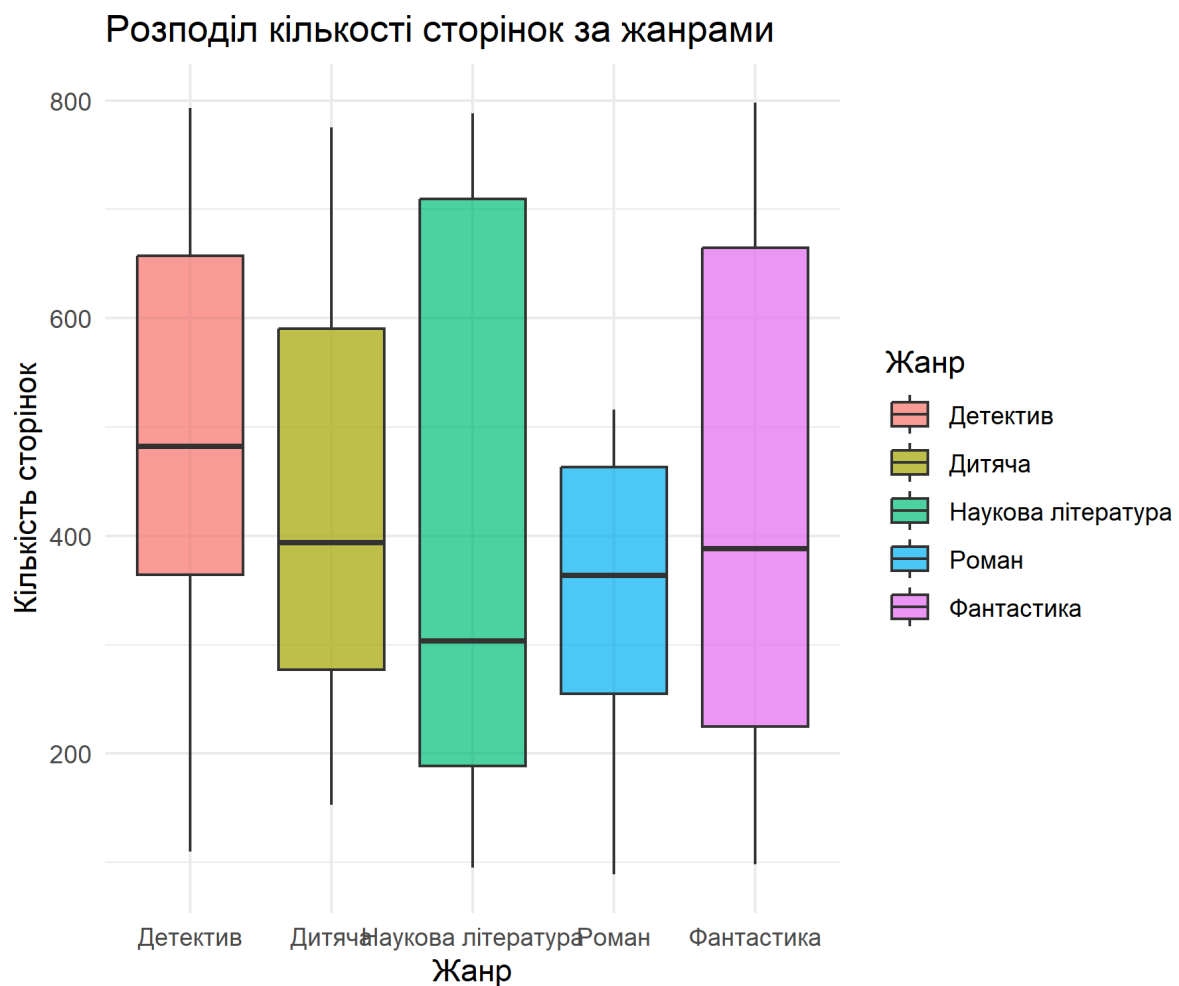
- медіану (середнє по середині)
- верхній та нижній кuartилі (Q1, Q3)

- мінімальна та максимальна кількість сторінок без викидів
- викиди - книги, які значно товстіші або тонші, ніж типові

Ми можемо побачити:

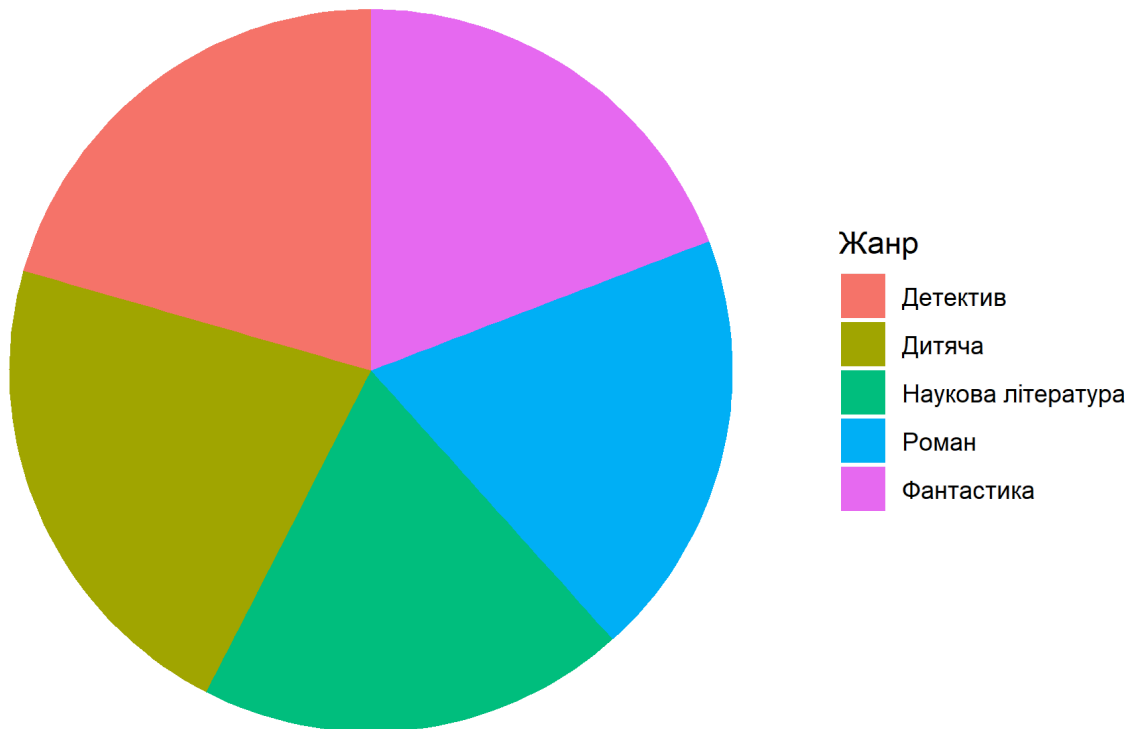
- які жанри мають типово товсті книги, а які - тонкі
- який жанр має великий розкид сторінок, а який стабільний
- чи є книги, що сильно вибиваються з загального обсягу (надто товсті)

Це аналіз структури жанрів за обсягом книг.



Кругова діаграма

Розподіл книг за жанрами



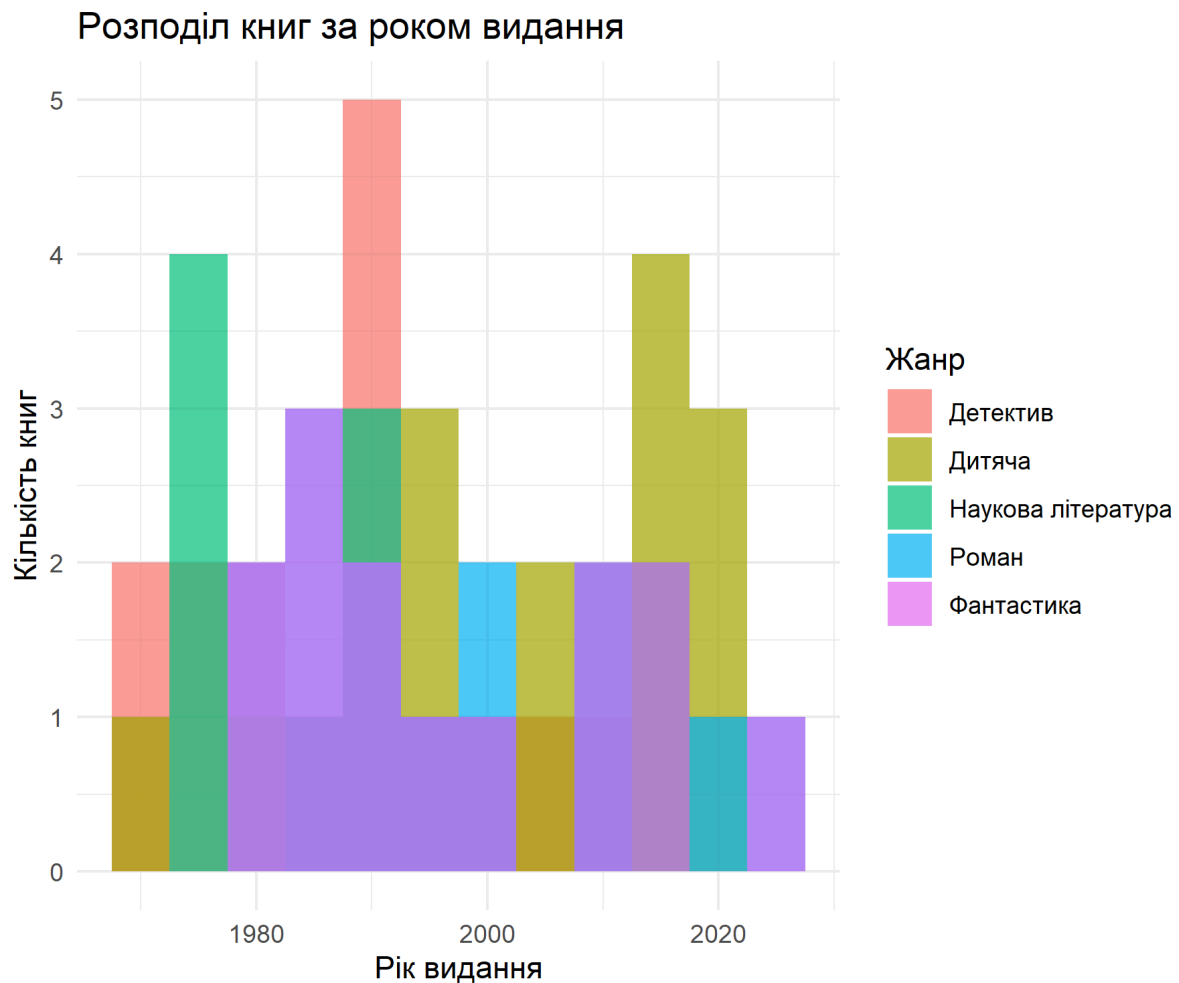
Ця діаграма показує:

- Які жанри є у бібліотеці
- Скільки книг належить до кожного жанру
- Яку частку (%) кожний жанр займає від загальної кількості книг

Тобто ми бачимо:

- який жанр найбільш популярний (має найбільший сегмент),
- які жанри представлені слабо,
- чи є значний дисбаланс.

Це аналіз структури бібліотеки за жанрами.



Гістограма років видання (додатково)

Гістограма показує як розподіляються книги за роками видання.

Скільки книг було видано в певні проміжки часу (інтервали)

У гістограмі:

по осі X зображено роки видання

по осі Y - кількість книг, виданих у ці роки

4. Висновки - аналіз виявлених закономірностей та трендів

У результаті аналізу згенерованих даних про книги бібліотеки було виявлено такі закономірності:

1. Розподіл жанрів

Кругова діаграма показує, що книги різних жанрів розподілені досить рівномірно. Окремі жанри можуть мати більшу частку (наприклад дитяча), але загалом у бібліотеці представлений широкий спектр літератури. Це свідчить про збалансовану тематичну колекцію.

2. Розподіл кількості сторінок між жанрами

Коробкова діаграма демонструє, що:

- найбільший розкид за кількістю сторінок спостерігається у жанрах **фантастика** та **наукова література** - вони мають і короткі, і дуже об'ємні книги;
- жанри **детектив** та **роман** мають більш стабільний (звужений) інтерквартильний розмах;
- у деяких жанрах виявлені потенційні **викиди** - книги, кількість сторінок у яких значно відрізняється від більшості.

3. Розподіл років видання книг (гістограма)

Гістограма показує, що:

- найбільша частина книг була видана у проміжку з **1990-х роках**,
- менша кількість книг припадає на період з 2000-х до 2010-х,
- це може свідчити про оновлення фонду у більш сучасний період або про націленість бібліотеки на нову літературу.

4. Загальні закономірності

- Жанрова структура збалансована, без суттєвих домінант,
- Кількість сторінок значно варіюється між жанрами, що є типовим для різних літературних напрямів.

Отже, аналіз дозволив визначити структуру бібліотечного фонду за жанрами, роками видання та обсягом книг, а також отримати загальне уявлення про різноманітність і сучасність колекції.

5. Література - список використаних джерел

Оскільки дані були згенеровані штучно, у роботі використовувались переважно методичні матеріали та документація:

- Hadley Wickham, Garrett Grolemund. R for Data Science. O'Reilly Media, 2017.
- Документація мови R: <https://cran.r-project.org>
- ggplot2: Elegant Graphics for Data Analysis - документація пакету ggplot2.
- Hadley Wickham. Advanced R. CRC Press, 2015.
- Методичні матеріали університету (якщо потрібно - можна додати конкретні).