

Київський національний університет
імені Тараса Шевченка

Звіт

до лабораторної роботи 9,
дисципліни: “Технології аналізу та візуалізації даних”,
студента: Євчика Олексія,
групи Інформатика, Магістри

Тема:

Візуалізація часових рядів у г.

Мета роботи:

Метою роботи є побудова теплокарти (heatmap), яка відображає сезонну активність даних у розрізі років та місяців. Для цього необхідно сформувати таблицю у форматі *tibble*, виконати попередню обробку часових даних та візуалізувати їх за допомогою `geom_tile()` у R. В результаті студент має навчитися виявляти та інтерпретувати сезонні патерни в часових рядах.

Теоретичні відомості

Часові ряди складаються з кількох основних компонентів:

- **Тренд (trend)** – довгострокова зміна рівня ряду.
- **Сезонність (seasonality)** – регулярні коливання, які повторюються з однаковим періодом (місяці, дні, години).
- **Циклічність (cycles)** – зміни, схожі на сезонні, але з непостійним періодом.
- **Випадкові коливання (noise)** – нерегулярні зміни, не пов'язані з трендом чи сезонністю.

Для аналізу сезонності існує кілька популярних типів візуалізацій:

- **Seasonal subseries plot**
- **Boxplot за місяцями / днями**
- **Теплокарта сезонності (seasonality heatmap)**
- **Polar plo**
- **Calendar heatmap**

Теплокарта дозволяє побачити інтенсивність або значення показника за двома вимірами (у нашому випадку – рік × місяць). `geom_tile()` з пакету **ggplot2** використовується для побудови кольорових прямокутників, де колір відображає величину активності.

Хід роботи

Завдання: побудувати heatmap сезонності (варіант 6)

Мета: відобразити активність за роками–місяцями.

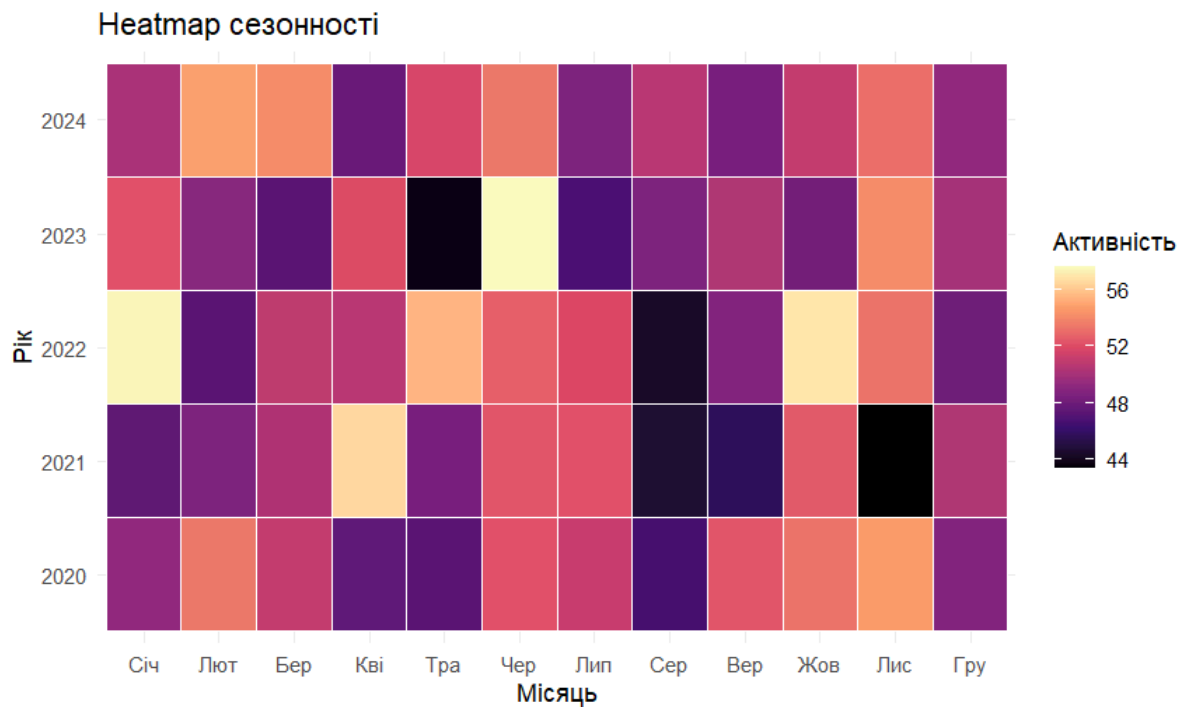
Кроки: створити tibble, використати `geom_tile()`.

Результат: теплокарта.

```
lab9_heatmap.R* x
Source on Save
Run

1 library(tidyverse)
2 library(lubridate)
3
4 # Прикладні дані
5 set.seed(123)
6 df <- tibble(
7   date = seq(as.Date("2020-01-01"), as.Date("2024-12-31"), by = "day"),
8   value = abs(rnorm(length(date), mean = 50, sd = 20))
9 )
10
11 # Підготовка даних для heatmap
12 season_df <- df %>%
13   mutate(
14     year = year(date),
15     month = month(date, label = TRUE, abbr = TRUE)
16   ) %>%
17   group_by(year, month) %>%
18   summarise(activity = mean(value), .groups = "drop")
19
20 # Теплокарта
21 ggplot(season_df, aes(x = month, y = factor(year), fill = activity)) +
22   geom_tile(color = "white") +
23   scale_fill_viridis_c(option = "magma") +
24   labs(
25     title = "Heatmap сезонності",
26     x = "Місяць",
27     y = "Рік",
28     fill = "Активність"
29   ) +
30   theme_minimal()
31
```

Отриманий графік



Короткий аналіз

Теплокарта дозволяє побачити виражені сезонні коливання за роками. У деякі місяці спостерігається підвищена активність (виділена більш теплими кольорами), тоді як інші місяці мають нижчі значення. Візуалізація добре демонструє повторюваність сезонних патернів та відхилення між роками. Колірна шкала полегшує оцінку інтенсивності.

Висновки

У ході роботи було створено tibble з агрегованими значеннями за місяцями та роками та побудовано теплокарту із застосуванням `geom_tile()`. Студент навчився працювати з сезонністю в часових рядах, формувати необхідні групування та візуалізувати результати. Найефективнішим методом для оцінки сезонності у даній роботі виявилась теплокарта, оскільки вона дозволяє швидко визначити патерни та аномальні періоди у даних.

Використана література

1. Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. (Free online: OTexts).
2. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer.
3. Shumway R., Stoffer D. Time Series Analysis and Its Applications.
4. Cowpertwait P., Metcalfe A. Introductory Time Series with R.
5. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning (розділ TS).
6. R Documentation: plot.ts, forecast, ggseasonplot, tsibble, fpp3.