

ГЕОДЕЗИЧНІ ДАНІ В КУРСІ «ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»

Хомич Василь

GEODETTIC DATA IN THE COURSE «INTELLIGENT DATA ANALYSIS»

Khomych Vasyi

Волинський національний університет імені Лесі Українки, просп. Волі, 13, Луцьк,
43025, Україна

Abstract. *This article presents methodological materials for a practical lesson involving geographic maps using the R software. The article describes the process of integrating search data from Google with a contour map of the world. Special attention is paid to the preparation, processing and formatting of data, their comparison with geographical coordinates and the creation of comparative visualisations.*

Візуалізація є важливим аспектом для аналізу статистичних даних. Це не просто допоміжний інструмент, а невід’ємна частина аналітичного процесу, що поєднує точність статистики з інтуїтивністю візуального сприйняття. Суміщення географічних карт та статистичних даних потрібно для досліджень у сфері географії та політології, сільського господарства, охорони здоров’я, медіа, екології, кліматології, бізнес-аналітиків, маркетингових спеціалістів, логістів. Тому тема візуалізація статистики на географічних картах включено до курсу «Інтелектуальний аналіз даних».

Метою є створення географічної карти світу, наповнення її даними засобами програми R із залученням відповідних бібліотек. Дана програма — потужне середовище для статистичного аналізу, різної обробки та візуалізації даних. Вона дозволяє проводити як і прості математичні операції, так і складні процеси моделювання, машинного навчання та аналізу різного роду рядів.

Було поставлене завдання: вивести статистичні дані з Google Trends на географічній карті для відображення певних закономірностей. Для аналізу було взяті дані за останні 12 місяців за проходження тесту на COVID-19 та вакцинацією. Це дані не про саме тестування і вакцинацію, а про пошукову активність — тобто, наскільки часто люди такі пошукові запити в Google. Для них присвоюємо імена змінних — covid test та covid vaccine відповідно.

Перед початком роботи необхідно встановити певні пакети для роботи:

1. rnatuarearth + sf — щоб отримати контурну карту світу у форматі sf (по суті розширена data.frame);
2. dplyr/tidyr – для детальнішої роботи з таблицями;
3. countrycode – перетворення англійських назв країн у дволітерні ISO-коди;
4. ggplot2 – реалізація “grammar of graphics” (граматики графіків), яка дає можливість будувати візуалізації за допомогою чіткого, послідовного синтаксису.

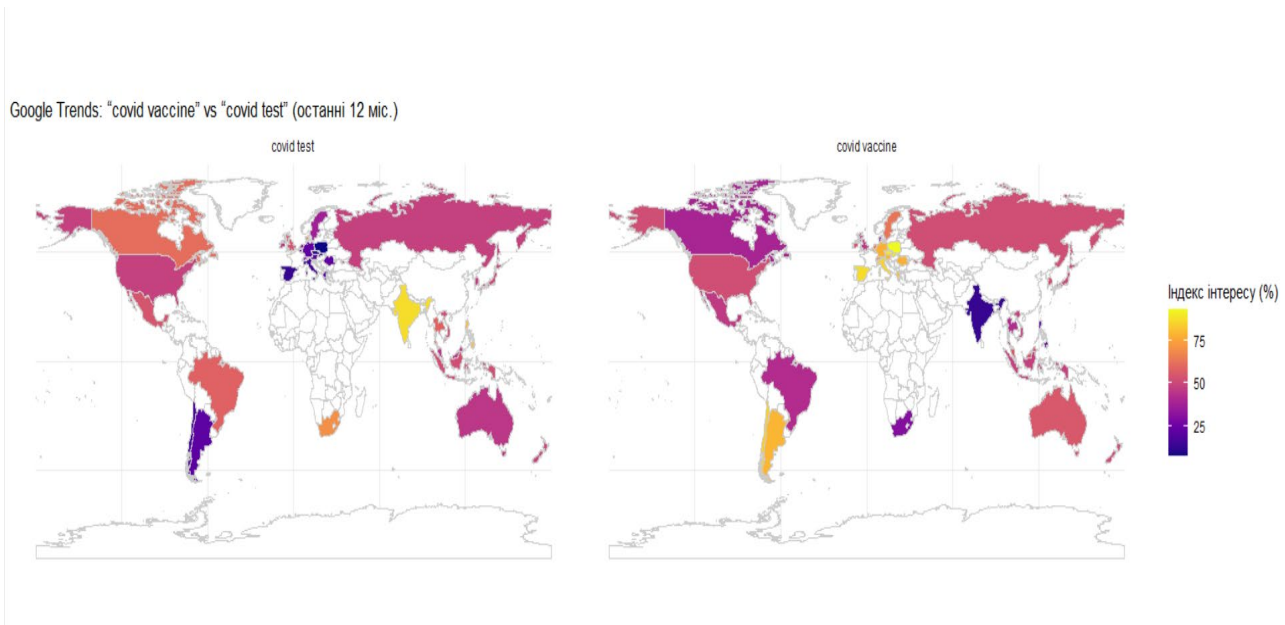
Робота розпочинається з завантаження даних з Google Trends та зчитування табличних даних. Це виконується командою read.csv та colnames(). Colnames() - присвоєння назв стовпцям таблиці задля універсальності (адже в залежності від мовної версії Google Trends можливі різні імена стовпців таблиці з даними).

Надалі ми працюємо над форматуванням даних в long для подальшої обробки (одна змінна — один рядок, команда pivot_longer(...)). Після цього ми очищаємо дані: прибираємо %, обрізаємо пробіли, заміняємо <1 на 0, перетворюємо в числовий формат. Також додаємо дволітерні ISO-коди країн для поєднання з геоданими.

Починаємо підготовку даних у wide формат для поєднання з картою. В результаті цього кожна країна має два стовпці: covid test та covid vaccine. Далі ініціалізуємо карту світу та пов'язуємо її з нашим статистичним даними командою `left_join(...)` та за ISO-кодами країн.

Після поєднання повертаємо назад в long формат для порівняльного відображення обох термінів — covid test та covid vaccine.

Надалі основний і останній крок: візуалізація. Будуємо 2 різні карти для 2 типів даних, об'єднуємо їх. Додаємо легенду для наочності відображення. Ранжування даних реалізовано кольоровою шкалою. Командою `geom_sf()` будуємо повноцінне зображення, яке відображає проаналізовані дані.



Бібліографія

1. *The Comprehensive R Archive Network*. URL: <https://cran.r-project.org/web/packages/sf/sf.pdf>.
2. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. *The Comprehensive R Archive Network*. URL: <https://cran.r-project.org/web/packages/ggplot2/index.html>
3. *The Comprehensive R Archive Network*. URL: <https://cran.r-project.org/web/packages/sf/sf.pdf>.