

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА
ШЕВЧЕНКА
ЕКОНОМІЧНИЙ ФАКУЛЬТЕТ**

**КАФЕДРА СТАТИСТИКИ,
ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМ І ДЕМОГРАФІЇ**

ПРИКЛАДНИЙ ПАКЕТ АНАЛІЗУ «R-STUDIO»

Навчально-методичний комплекс

(програма, завдання для лабораторних занять і самостійної роботи, система оцінки знань студентів 05 Соціальні та поведінкові науки, освітньої програми «Економічна аналітика та статистика»)

КИЇВ – 2025

УДК 311

Затверджено Вченою радою економічного факультету Київського національного університету імені Тараса Шевченка, протокол № __ від 15 квітня 2025 р.

Автори: Горна М.О., к.е.н., Київський національний університет імені Тараса Шевченка

Рецензенти: Іщук Я.В. к.е.н., доцент, доцент кафедри математичного моделювання та статистики Інституту інформаційних технологій в економіці Київського національного економічного університету імені Вадима Гетьмана
Федоренко І.К. к.е.н., доцент, доцент кафедри економічної кібернетики, Київський національний університет імені Тараса Шевченка

Прикладний пакет аналізу «R-studio»: Навчально-методичний комплекс (програма курсу, плани лабораторних занять, завдання для самостійної роботи і система оцінки знань студентів спеціальності 051 «Економіка») / М.О. Горна. – Київ: 2025. – 48 с.

Навчально-методичний комплекс розроблений з урахуванням освітніх вимог до організації навчання в очній та дистанційній формі з використанням сучасних інформаційно-аналітичних технологій.

Комплекс містить деталізовані рекомендації по опануванню курсу, які включають: анотацію, програму та план лабораторних занять для формування навичок аналізу; список рекомендованої навчальної літератури, посилання на зовнішні онлайн джерела; рекомендовану наукову літературу, яка демонструє застосування прикладного пакету аналізу R-Studio для аналізу даних.

© Горна М.О.
2025

ЗМІСТ

Вступ.....	4
I. Зміст дисципліни.....	6
II. Навчально-тематичний план.....	9
III. Плани лекцій.....	10
IV. Плани лабораторних занять.....	17
V. Індивідуальна самостійна робота	40
VI. Результати навчання	42
VII. Список рекомендованої літератури	45

Вступ

Дисципліна «Прикладний пакет аналізу «R-Studio» є обов'язковою навчальною дисципліною для студентів спеціальності 051 «Економіка», яка викладається протягом третього семестру ОС Магістр в обсязі 6 кредитів (за Європейською Системою ECTS), або 180 годин, з яких: 16 годин лекційні заняття, 44 години – лабораторні заняття, 120 годин – самостійної роботи.

Мета і задачі дисципліни. Мета дисципліни – формування у здобувачів необхідних теоретичних знань та практичних навичок з ефективного застосування прикладного пакету аналізу даних «R-Studio» з використанням спеціальних бібліотек аналізу та візуалізації даних для подальшого застосування в професійній діяльності.

Завдання дисципліни: опанувати основи роботи з прикладним пакетом «R-Studio», навчитися розгортати середовище для аналізу даних, зрозуміти основні принципи обробки, аналізу та візуалізації даних, освоїти використання спеціальних бібліотек для аналізу, сформувати навички побудови графіків і таблиць, а також отримати базове розуміння програмування із застосуванням «R-Studio» у професійній діяльності. Зокрема отримати чи вдосконалити наступні компетенції:

Загальні компетенції: здатність до абстрактного мислення, аналізу, синтезу та встановлення взаємозв'язків між явищами та процесами; здатність розробляти проекти та управляти ними; здатність проводити дослідження та презентувати результати.

Фахові компетентності спеціальності: здатність використовувати сучасні інформаційні технології та економікоматематичні методи і моделі для дослідження економічних та соціальних процесів; здатність розуміти ключові тренди соціально-економічного та демографічного розвитку; вміння планувати, проводити та презентувати наукові дослідження, готувати результати наукових робіт до оприлюднення.

Навчальна дисципліна «Прикладний пакет аналізу «R-Studio» дозволить здобувачам вищої освіти опанувати основи аналізу даних у середовищі «R-Studio», набути практичних навичок роботи з даними, а також розвинути здатність до самостійного виконання економічних досліджень із застосуванням

сучасних методів аналізу та візуалізації інформації, що забезпечить здобувачів необхідними інструментами для розв'язання прикладних завдань у професійній діяльності.

Значна увага приділена розгляду особливостей роботи із даними в тому числі із використанням бібліотек Tidyverse, dplyr, ggplot2 та інших, які дозволяють ефективно очищувати, трансформувати, аналізувати та візуалізувати дані.

Дисципліна передбачає навчання у формі лабораторних занять та самостійної роботи студентів. Для практичного засвоєння основних тем дисципліни лабораторні заняття, індивідуальна робота та консультації проводяться із застосуванням персональних комп'ютерів, локальної мережі та мережі Інтернет в комп'ютерних класах або дистанційно. Всі види занять забезпечуються необхідними методичним матеріалами в електронній формі.

Попередні вимоги до опанування навчальної дисципліни.

1. **розуміння** сутності статистичних понять та методів аналізу соціально-економічних явищ і процесів та основ вищої математики;
2. **вміння** застосовувати математичні та статистичні методи аналізу, працювати у команді, виконувати командні ролі, звітуватися та захищати проекти;
3. **володіння** елементарними навичками користування комп'ютерною технікою, включаючи роботу з електронними таблицями, практичні аспекти статистичних обчислень, зокрема програмування та обробку даних у середовищі R.

Система контролю знань та підсумкового оцінювання. Навчальна дисципліна «Прикладний пакет аналізу «R-Studio» оцінюється за Європейською Кредитно- Трансферною Системою ECTS. Протягом семестру набуті знання та навички оцінюються в межах кожної теми окремо. Семестрову оцінку формують бали, отримані студентом у процесі засвоєння матеріалу з усіх тем дисципліни. Вивчається дисципліна протягом одного семестру, по завершенню яких студенти складають іспит.

I. ЗМІСТ ДИСЦИПЛІНИ

ЗМІСТОВИЙ МОДУЛЬ I. ПРИКЛАДНИЙ

ПАКЕТ АНАЛІЗУ «R-STUDIO»

Тема 1. Вступ до R та Data Science.

Ознайомлення з основними поняттями Data Science, його структурою та значенням у сучасному аналізі даних. Розгляд особливостей мови програмування R, її переваг для статистичного аналізу та візуалізації даних. Встановлення та налаштування R Studio, знайомство з його інтерфейсом і основними інструментами. Зазначення ролі R у професійній діяльності аналітика. Обговорення перспектив використання R у різних галузях, включаючи економіку, фінанси та соціальні науки.

Тема 2. Аналіз даних з використанням R Studio.

Розгляд типів даних у R та основних операцій із ними: імпорту, обробки та експорту. Робота з різними джерелами даних, такими як файли CSV, Excel, бази даних. Зазначення важливості попередньої обробки даних, очищення від шумів і аномалій. Використання базових інструментів для аналізу даних, включаючи підсумовування, фільтрацію та групування. Практичне застосування інструментів R Studio для розв'язання завдань обробки даних.

Тема 3. Вступ у Tidyverse.

Ознайомлення з набором пакетів Tidyverse, які спрощують роботу з даними в R. Розгляд основних пакетів, таких як dplyr для маніпуляцій із даними, tidyr для упорядкування та ggplot2 для побудови графіків. Навчання використанню pipe-оператора, який забезпечує зручний синтаксис для обробки даних. Розгляд практичних прикладів аналізу даних за допомогою Tidyverse. Порівняння підходів до аналізу з використанням базових функцій R і бібліотеки Tidyverse.

Тема 4. Описова статистика в R Studio.

Описова статистика в середовищі R Studio. Вивчення методів описової статистики, зокрема розрахунок середнього значення, медіани, моди, дисперсії, стандартного відхилення. Використання вбудованих функцій для підсумовування даних, таких як `summary()`, `mean()`, `sd()` та `var()`. Застосування описової статистики для аналізу характеристик вибірок даних. Створення частотних таблиць та побудова графіків розподілу, таких як гістограми та Q-Q графіки, а також діаграм `Box-plot`. Розгляд методів оцінки тенденцій і варіацій у вибірках та нормальності розподілу. Вивчення основ роботи з бібліотеками для автоматизації статистичного аналізу, зокрема пакету `psych`. Практичні завдання з аналізу статистичних показників у R Studio для реальних наборів даних, включаючи їх опис, візуалізацію та порівняння між групами.

Тема 5. Перевірка гіпотез в R Studio.

Розгляд основ перевірки статистичних гіпотез, таких як t-тести, аналіз дисперсії (ANOVA) та χ^2 -тести. Вивчення етапів формулювання гіпотез: нульової та альтернативної, вибору критерію перевірки та інтерпретації результатів. Використання вбудованих функцій для статистичних тестів, таких як `t.test()`, `anova()` та `chisq.test()`. Обговорення припущень для коректного застосування кожного тесту. Практичне застосування тестів для аналізу реальних даних, зокрема порівняння середніх значень між групами, перевірка незалежності змінних та оцінка впливу факторів. Особлива увага приділяється аналізу рівня значущості, довірчих інтервалів та відмови від нульової гіпотези. Використання бібліотек, таких як `easystats` та `stats`, для автоматизації процесу перевірки гіпотез.

Тема 6. Регресійні моделі в R Studio.

Ознайомлення з основними типами регресійних моделей, такими як лінійна, множинна та логістична регресія. Навчання створенню та аналізу моделей у R Studio за допомогою функцій `lm()` та `glm()`. Розгляд критеріїв якості

моделі: коефіцієнт детермінації (R^2), р-значення, AIC (Інформаційний критерій Акаїке) та BIC (Байєсівський інформаційний критерій), та аналіз залишків. Демонстрація роботи з візуалізацією регресійних моделей за допомогою бібліотеки ggplot2. Практичні завдання на побудову регресійних моделей для прогнозування та інтерпретації зв'язків між змінними з використанням бібліотек, таких як mlr3 або caret. Аналіз впливу мультиколінеарності на результати моделювання та її усунення. Використання моделей для розв'язання прикладних задач у соціально-економічному аналізі.

Тема 7. Візуалізація даних в R Studio.

Вивчення основ візуалізації даних за допомогою бібліотеки ggplot2. Розгляд різних типів графіків: гістограми, діаграми розсіювання, стовпчикові графіки, лінійні діаграми та комбіновані графіки. Зазначення особливостей інтерактивної візуалізації з використанням бібліотек, таких як plotly та shiny. Налаштування кольорів, шрифтів і підписів для підвищення інформативності графіків. Створення багатошарових графіків для комплексного аналізу даних. Навчання застосуванню фільтрації та агрегування даних для візуалізації тенденцій. Практичні завдання з побудови візуалізацій для реальних наборів даних із різних сфер, зокрема економіки.

Тема 8. Часові ряди та їх візуалізація.

Розгляд специфіки аналізу часових рядів у R Studio, включаючи підготовку даних та оцінку динаміки. Ознайомлення з пакетами forecast та tsibble для аналізу часових рядів. Вивчення методів трендового та сезонного аналізу, а також автокореляції. Огляд методів оцінки стаціонарності часових рядів за допомогою тестів, таких як Augmented Dickey-Fuller (ADF). Ознайомлення з моделями прогнозування, такими як ARIMA, ETS та їх використанням для прогнозування. Практичне використання інструментів R Studio для аналізу часових рядів, створення графіків їхньої динаміки та інтерпретації результатів.

II. НАВЧАЛЬНО-ТЕМАТИЧНИЙ ПЛАН

Назви розділів і тем	Кількість годин у тому числі		
	Лекції	Лабораторні	Самостійна робота
Тема 1. Вступ до R та Дата Саенс. 1.1 Встановлення, огляд інтерфейсу середовища R Studio. 1.2 Data Science: визначення, етапи та місце R Studio у аналізі даних. 1.3 Основи роботи з R Studio.	2 -	- 2	4 8
Тема 2. Аналіз даних з використанням R Studio. 2.1.Завантаження, зчитування та попередній огляд даних. 2.2. Підготовка даних до аналізу.2.3.Використання базових функцій.	2 - -	- 2 4	4 6 4
Тема 3. Вступ у Tidyverse. 3.1. Загальний огляд Tidyverse. 3.1.Маніпуляція даними за допомогою dplyr.3.3.Трансформація даних за допомогою tidyr.	2 - -	 2 4	4 6 4
Тема 4. Описова статистика в R Studio. 4.1 Основні показники описової статистики в RStudio. 4.2.Частотний аналіз даних: 4.3.Аналіз аномалій.	2 - - -	- 2 2 2	4 6 4 6
Тема 5. Перевірка гіпотез в R Studio. 5.1 Концепція перевірки гіпотез. 5.2.Проведення t-тестів, ANOVA та χ^2 -тестів.5.3.Інтерпретація результатів тестів.	2 - -	- 2 4	4 6 6
Тема 6. Регресійні моделі в R Studio. 6.1.Основи лінійної регресії. 6.2. Множинна регресія. 6.3.Логістична регресія.	2 - -	- 2 4	6 4 6
Тема 7. Візуалізація даних в R Studio. 7.1. Основи візуалізації в R Studio. 7.2. Створення графіків за допомогою ggplot2. 7.3. Інтерактивна візуалізація даних.	2 - -	- 2 4	6 4 4
Тема 8. Часові ряди та їх візуалізація. 8.1. Вступ до часових рядів у R Studio та підготовка даних. 8.2.Аналіз характеристик часових рядів. 8.3.Прогнозування та оцінка моделей.	2 - -	- 2 4	6 4 4
Разом годин:	16	44	120

Загальний обсяг 160 год, в тому числі:

Лекцій – 16 год.

Лабораторні заняття – 44 год.

Самостійна робота - 120 год.

III. ПЛАНИ ЛЕКЦІЙ

Тема 1. Вступ до R та Дата Саенс.

1. Встановлення R та R Studio, огляд інтерфейсу середовища R Studio.
2. Концепція Data Science: визначення, етапи та місце R у аналізі даних.
3. Основи роботи з R: типи даних, оператори, базові функції та обробка простих наборів даних.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

Онлайн ресурси:

1. Офіційний сайт R: <https://www.r-project.org/>
2. RStudio: <https://www.rstudio.com/products/rstudio/download/>
3. Основи прогнозування в R. Вступ до R. Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_forecasting/intro_r.html.
4. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

Тема 2. Аналіз даних з використанням R Studio.

1. Завантаження, зчитування та попередній огляд даних: робота з форматами CSV, Excel та базами даних.
2. Підготовка даних до аналізу: очищення, обробка пропусків, трансформація даних.
3. Використання базових функцій та пакетів для швидкого аналізу та попередньої візуалізації даних.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

Онлайн ресурси:

1. Data Analysis with R Programming by Google. [Електронний ресурс] - Режим доступу: <https://www.coursera.org/learn/data-analysis-r>

2. Prometheus: Аналіз даних та статистичне виведення на мові R. [Електронний ресурс] - Режим доступу: <https://prometheus.org.ua/prometheus-free/data-analysis-statistics/>

3. Основи економетрики в R. Вступ до R. Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_econometrics/introR.html

4. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

Тема 3. Вступ у Tidyverse.

1. Загальний огляд Tidyverse: структура, основні бібліотеки.

2. Маніпуляція даними за допомогою dplyr: фільтрація, сортування, групування, створення нових змінних.

3. Трансформація даних за допомогою tidyr: зведення, об'єднання та розділення даних.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404.

Онлайн ресурси:

1. Introduction to the Tidyverse [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-the-tidyverse>

2. Основи економетрики в R. Маніпуляції з даними за допомогою dplyr і не тільки. Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_econometrics/dplyr.html

3. Data Manipulation with dplyr [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/data-manipulation-with-dplyr>

Тема 4. Описова статистика в R Studio.

1. Основні показники описової статистики в R Studio.

2. Частотний аналіз даних: створення таблиць частот, гістограм та діаграм Box-plot.

3. Аналіз аномалій та тенденцій: використання Q-Q графіків та виявлення аутлайєрів у вибірках.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

4. An introduction to Statistical Learning with Applications in R / Springer, 2021

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

2. Introduction to Statistics in R [Електронний ресурс]:дистанційний курс.
Режимдоступу:<https://app.datacamp.com/learn/courses/introduction-to-statistics-in-r>

3. Prometheus: Аналіз даних та статистичне виведення на мові R.
[Електронний ресурс] - Режим доступу: <https://prometheus.org.ua/prometheus-free/data-analysis-statistics/>

Тема 5. Перевірка гіпотез в R Studio.

1. Концепція перевірки гіпотез: нульова та альтернативна гіпотези, рівень значущості, р-значення.

2. Проведення t-тестів, ANOVA та χ^2 -тестів у R Studio: використання відповідних функцій та бібліотек.

3. Інтерпретація результатів тестів: довірчі інтервали, відмова від нульової гіпотези, аналіз результатів.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]:
дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

2. Introduction to Statistics in R [Електронний ресурс]:дистанційний курс.
Режим доступу:<https://app.datacamp.com/learn/courses/introduction-to-statistics-in-r>

3. Prometheus: Аналіз даних та статистичне виведення на мові R.
[Електронний ресурс] - Режим доступу: <https://prometheus.org.ua/prometheus-free/data-analysis-statistics/>

Тема 6. Регресійні моделі в R Studio.

1. Основи лінійної регресії: побудова моделей, оцінка параметрів, аналіз якості моделі.

2. Множинна регресія: робота з декількома предикторами, оцінка взаємозв'язків змінних.

3. Логістична регресія: застосування для класифікації, оцінка точності моделі та залишковий аналіз.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404.

4. Основи статистичного моделювання: навч. посібник / за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП «Рута», 2022. 604 с.

5. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

2. Introduction to Regression in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-regression-in-r>

3. Intermediate Regression in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/intermediate-regression-in-r>

4. Основи економетрики в R. Ігор Мірошниченко. [Електронний ресурс] - Режим доступу: https://araur.github.io/r_econometrics/index.html

Тема 7. Візуалізація даних в R Studio.

1. Основи візуалізації в R Studio.
2. Створення графіків за допомогою ggplot2.
3. Інтерактивна візуалізація даних.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. Claus O. Wilke. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. 1st Edition. O'Reilly Media. 2019. 387 p. — ISBN 1492031089, 978-1492031086.
4. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>
2. Introduction to Data Visualization with ggplot2 [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-data-visualization-with-ggplot2>
3. Intermediate Data Visualization with ggplot2 [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/intermediate-data-visualization-with-ggplot2>
4. ggplot2: Elegant Graphics for Data Analysis (3e). [Електронний ресурс] - Режим доступу: <https://ggplot2-book.org/>

Тема 8. Часові ряди та їх візуалізація.

1. Вступ до часових рядів у R Studio та підготовка даних.

2. Аналіз характеристик часових рядів. Моделювання часових рядів.
3. Прогнозування та оцінка моделей.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.
4. Hyndman R.J., Athanasopoulos G. "Forecasting: Principles and Practice". OTexts. 2013. 292 p. — ISBN 0987507109, 978-0987507105.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>
2. Time Series Analysis in R [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/time-series-analysis-in-r>
3. Основи прогнозування в R. Часові ряди та їх візуалізація. Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_forecasting/ts.html

IV. ПЛАНИ ЛАБОРАТОРНИХ РОБІТ

Лабораторна робота 1. Вступ до R та Дата Саенс.

1. Встановити R та RStudio на свій комп'ютер. Ознайомитися з основними елементами інтерфейсу RStudio: консоль, редактор скриптів, панель змінних, візуалізація графіків.

2. Створити вектори для цін на Bitcoin та Litecoin за тиждень. Додати імена елементів у векторах. Розрахувати середню ціну, визначити, в які дні була найвища та найнижча ціна на Bitcoin та Litecoin. Розрахувати суму ваших щоденних торгів по двом криптовалютам та запишіть результати у вектор `total_daily`. Розрахувати суму ваших торгів для кожної криптовалюти та запишіть результати у вектори `total_BTC` та `total_LTC`. Порівняйте результати торгів по двом криптовалютам. Чи заробили Ви більше з Bitcoin у порівнянні з Litecoin?

3. Побудувати лінійні графіки зміни цін за тиждень для обох криптовалют за допомогою базових візуалізацій R Studio. Додати до графіків назви осей та заголовки. Обчислити коефіцієнт кореляції між цінами Bitcoin та Litecoin. Побудувати діаграму розсіювання для візуалізації цієї залежності, додати назви осей та заголовки.

4. Створити дата фрейм, який містить дані про ціни обох криптовалют та день тижня. Додати до дата фрейму стовпчик з відносною зміною ціни.

Приклад коду:

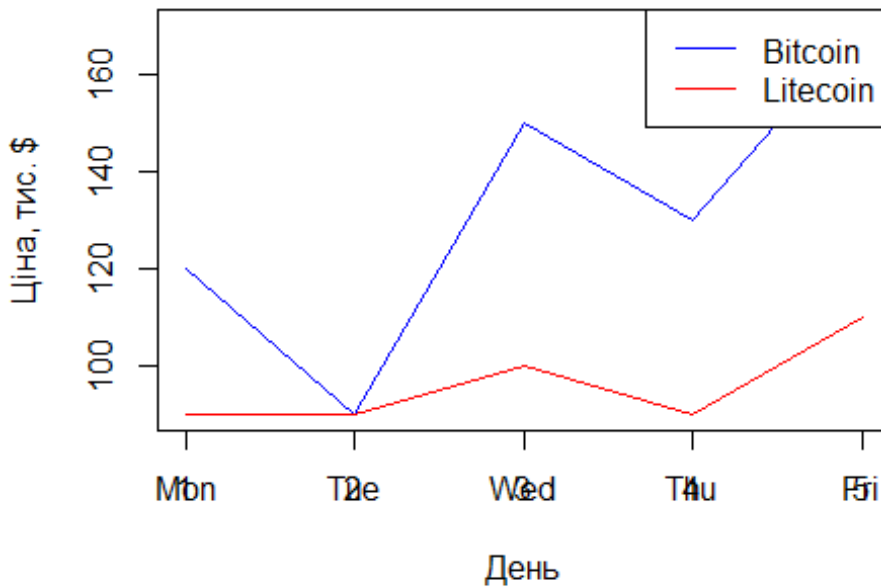
```
# Створення векторів з цінами криптовалют, тис. $
bitcoin <- c(120, 90, 150, 130, 170)
litecoin <- c(90, 90, 100, 90, 110)

# Додавання імен до елементів векторів
days <- c("Mon", "Tue", "Wed", "Thu", "Fri")
names(bitcoin) <- names(litecoin) <- days

# Побудова лінійних графіків зміни цін
plot(bitcoin, type = "l", col = "blue", main = "Динаміка цін Bitcoin та Litecoin",
      xlab = "День", ylab = "Ціна, тис. $")
lines(litecoin, col = "red")
legend("topright", legend = c("Bitcoin", "Litecoin"), col = c("blue", "red"), lty = 1)

# Додавання назв днів на вісь X
axis(1, at = 1:length(days), labels = days)
```

Динаміка цін Bitcoin та Litecoin

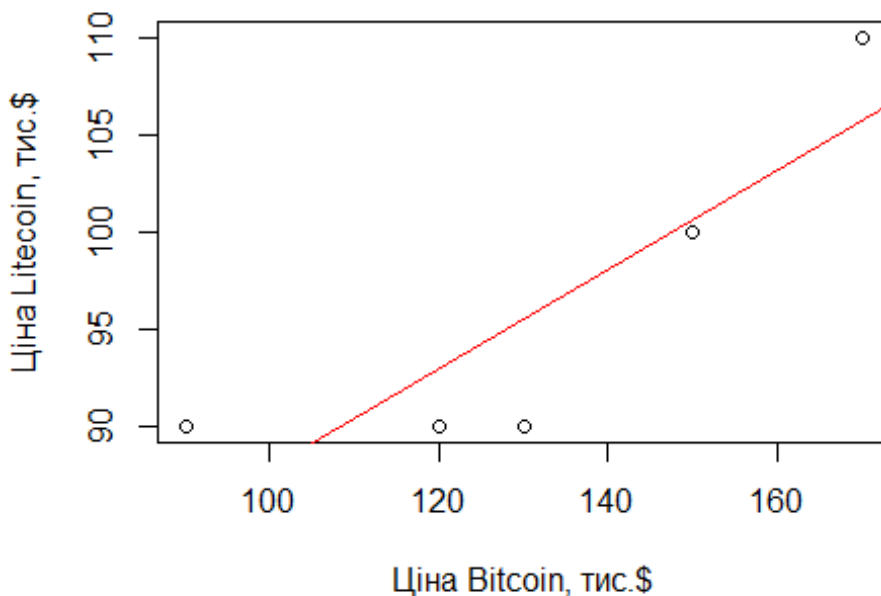


```
# Розрахунок коефіцієнта кореляції
correlation <- cor(bitcoin, litecoin)
print(paste("Коефіцієнт кореляції між Bitcoin та Litecoin:", correlation))

## [1] "Коефіцієнт кореляції між Bitcoin та Litecoin: 0.866221492410021"

# Діаграма розсіювання
plot(bitcoin, litecoin, main = "Залежність між цінами Bitcoin та Litecoin",
      xlab = "Ціна Bitcoin, тис.$", ylab = "Ціна Litecoin, тис.$")
abline(lm(litecoin ~ bitcoin), col = "red")
```

Залежність між цінами Bitcoin та Litecoin



```

# Розрахунок відносної зміни ціни (у відсотках)
bitcoin_change <- c(NA, diff(bitcoin) / bitcoin[-length(bitcoin)] * 100)
litecoin_change <- c(NA, diff(litecoin) / litecoin[-length(litecoin)] * 100)

# Створення дата фрейму з цінами та змінами
crypto_data <- data.frame(
  Day = c("Mon", "Tue", "Wed", "Thu", "Fri"),
  Bitcoin_Price = bitcoin,
  Litecoin_Price = litecoin,
  Bitcoin_Change = bitcoin_change,
  Litecoin_Change = litecoin_change
)

# Вивід дата фрейму
print(crypto_data)

##      Day Bitcoin_Price Litecoin_Price Bitcoin_Change Litecoin_Change
## Mon Mon           120             90              NA              NA
## Tue Tue            90             90          -25.00000           0.00000
## Wed Wed           150            100           66.66667           11.11111
## Thu Thu            130             90          -13.33333          -10.00000
## Fri Fri            170            110           30.76923           22.22222

```

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

Онлайн ресурси:

1. Офіційний сайт R: <https://www.r-project.org/>
2. RStudio: <https://www.rstudio.com/products/rstudio/download/>
3. Основи прогнозування в R. Вступ до R. Ігор Мірошниченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_forecasting/intro_r.html.
4. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

Лабораторна робота 2. Аналіз даних з використанням R Studio.

1. Завантажте R Studio та встановіть необхідні пакети: readr, readxl, data.table. Зчитайте дані з файлу CSV за допомогою функції read.csv(). Виведіть

перші 6 рядків набору даних. Зчитайте дані з Excel-файлу за допомогою функції `read_excel()`. Виведіть загальну інформацію про набір даних `str()`, `summary()`, перші 10 та 10 останніх рядків набору даних. Завантажте набір даних із зовнішнього джерела URL: https://raw.githubusercontent.com/agricolamz/2020-2021-ds4dh/master/data/heroes_information.csv

2. Перевірте наявність пропущених значень `is.na()`. Заповніть пропуски середнім/медіанним значенням або видаліть їх. Виконайте перетворення змінних (наприклад, змініть формат дати, об'єднайте категорії). Виконайте групування та агрегування даних (наприклад, знайдіть середнє значення певного показника по категоріях).

3. Базові візуалізації даних в Rstudio. Побудуйте гістограму для однієї з числових змінних, використовуйте функцію `hist()`. Додайте заголовок та підписи до осей. Створіть графік розсіювання (scatter plot) за допомогою функції `plot()`. Додайте заголовок, підписи до осей, а також змініть колір точок. Побудуйте Q-Q графік для кількісної змінної, використовуйте функцію `qqnorm()` для побудови Q-Q графіку та `qqline()` для додавання теоретичної лінії. Поясніть, чи розподіл змінної є нормальним, зважаючи на те, як точки розміщуються відносно теоретичної лінії. Опишіть отримані результати аналізу. Які висновки можна зробити на основі попередньої візуалізації даних?

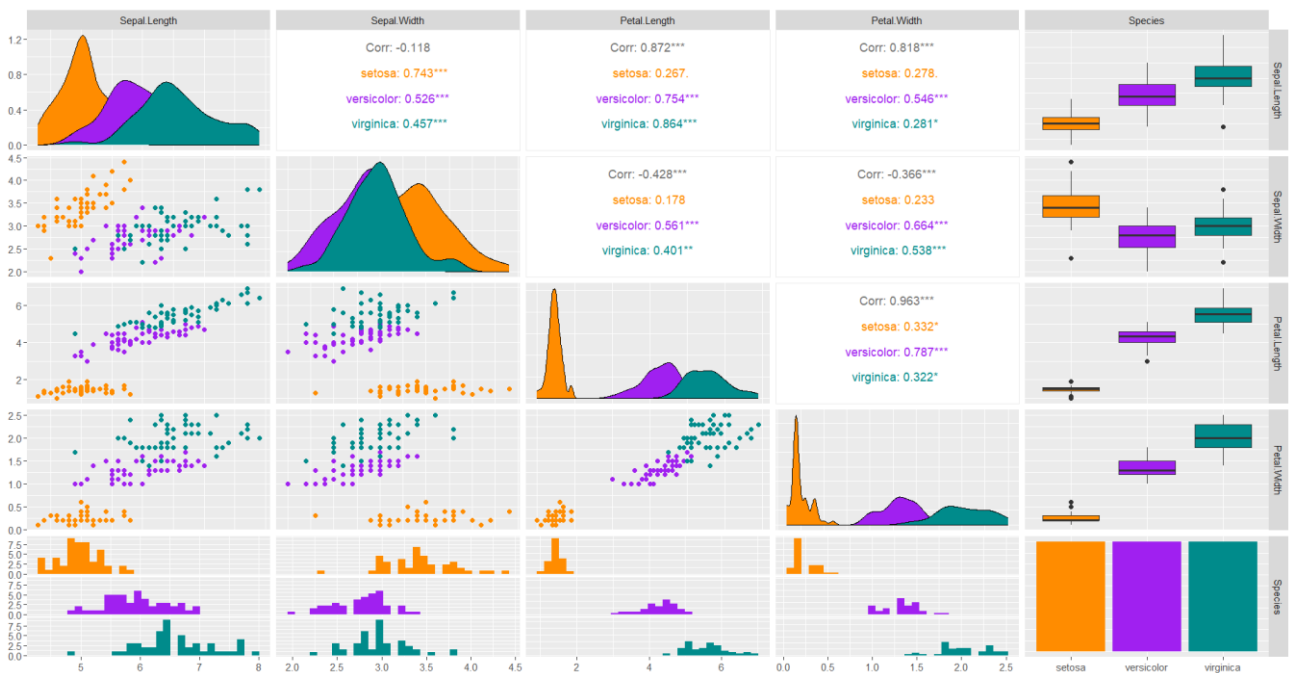
4. Ознайомитися з пакетом `explore` та його можливостями для проведення експрес-аналізу на прикладі набору даних `iris` (Іриси Фішера), використовуючи функції `describe()`, `explore()`, `explore_all()`, `explain_tree()` для отримання ключових інсайтів.

Приклад коду:

```
# Комплексна візуалізація розподілу та взаємозв'язку змінних в наборі даних iris з пакетом explore:
```

```
library(explore)

ggpairs(iris, aes(color = Species)) +
  scale_colour_manual(values = c("darkorange", "purple", "cyan4")) +
  scale_fill_manual(values = c("darkorange", "purple", "cyan4"))
```



5. Лабораторна робота має бути оформлена у вигляді звіту у форматі **html** за допомогою Quarto або Rmarkdown.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

Онлайн ресурси:

1. Data Analysis with R Programming by Google. [Електронний ресурс] - Режим доступу: <https://www.coursera.org/learn/data-analysis-r>

2. Prometheus: Аналіз даних та статистичне виведення на мові R. [Електронний ресурс] - Режим доступу: <https://prometheus.org.ua/prometheus-free/data-analysis-statistics/>

3. Основи економетрики в R. Вступ до R. Ігор Мірошниченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_econometrics/introR.html

4. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

Лабораторна робота 3. Вступ у Tidyverse. Маніпуляції з даними за допомогою пакету dplyr

1. Ознайомлення з пакетом Tidyverse. Встановіть та підключіть необхідні пакети (tidyverse, gapminder, dplyr, ggplot2). Ознайомтеся з набором даних gapminder (структура, типи змінних, загальний опис).

2. Маніпуляції з даними у dplyr. Відібрати дані за 1957 рік. Відфільтрувати дані для США у 2002 році. Виконати фільтрацію з використанням операторів &, |, ! та функції xor() (виключне АБО). Впорядкувати дані за очікуваною тривалістю життя (lifeExp) у зростаючому порядку. Відсортувати дані за ВВП на душу населення (gdpPerCap) у спадному порядку. Додати новий стовпчик lifeExpMonths, що містить очікувану тривалість життя в місяцях. Розрахувати ВВП країни (GDP_billion) у мільярдах доларів. Відібрати лише змінні country, year, gdpPerCap. Виконати вибір змінних за номерами стовпців. Знайти медіану очікуваної тривалості життя lifeExp у 2007 році. Обчислити середнє значення lifeExp та максимальний gdpPerCap для кожного континенту у 2007 році. Знайти середню та медіанну очікувану тривалість життя та кількість країн на кожному континенті у 2007 році.

3. Лабораторна робота має бути оформлена у вигляді звіту у форматі **html** за допомогою Quarto або Rmarkdown та включати: код виконаних завдань із коментарями, короткий аналіз результатів.

Приклад коду:

Інсталюйте необхідні пакети

```
install.packages("dplyr")  
install.packages("gapminder")
```

Підключимо необхідні пакети

```
library(gapminder)
```

```
library(dplyr)
```

Перегляньте набір даних `gapminder`

```
gapminder
```

```
## # A tibble: 1,704 × 6
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>
## 1 Afghanistan Asia      1952  28.8  8425333    779.
## 2 Afghanistan Asia      1957  30.3  9240934    821.
## 3 Afghanistan Asia      1962  32.0 10267083    853.
## 4 Afghanistan Asia      1967  34.0 11537966    836.
## 5 Afghanistan Asia      1972  36.1 13079460    740.
## 6 Afghanistan Asia      1977  38.4 14880372    786.
## 7 Afghanistan Asia      1982  39.9 12881816    978.
## 8 Afghanistan Asia      1987  40.8 13867957    852.
## 9 Afghanistan Asia      1992  41.7 16317921    649.
## 10 Afghanistan Asia      1997  41.8 22227415    635.
## # [i] 1,694 more rows
```

```
View(gapminder)
```

```
str(gapminder)
```

```
## tibble [1,704 × 6] (S3: tbl_df/tbl/data.frame)
## $ country : Factor w/ 142 levels "Afghanistan",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ continent: Factor w/ 5 levels "Africa","Americas",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ year     : int [1:1704] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
## $ lifeExp  : num [1:1704] 28.8 30.3 32 34 36.1 ...
## $ pop      : int [1:1704] 8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 16317921 22227415 ...
## $ gdpPercap: num [1:1704] 779 821 853 836 740 ...
```

Виконаємо просту фільтрацію з однією умовою

```
gapminder %>% # for %>% press Ctrl+Shift+M
  filter(year == 1957)
```

```
## # A tibble: 142 × 6
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>
## 1 Afghanistan Asia      1957  30.3  9240934    821.
## 2 Albania      Europe    1957  59.3  1476505    1942.
## 3 Algeria      Africa    1957  45.7 10270856    3014.
## 4 Angola       Africa    1957  32.0  4561361    3828.
## 5 Argentina    Americas  1957  64.4 19610538    6857.
## 6 Australia    Oceania   1957  70.3  9712569    10950.
## 7 Austria      Europe    1957  67.5  6965860    8843.
## 8 Bahrain      Asia      1957  53.8  138655     11636.
## 9 Bangladesh   Asia      1957  39.3 51365468     662.
## 10 Belgium      Europe    1957  69.2  8989111    9715.
## # [i] 132 more rows
```

Фільтрація з декількома умовами

```
gapminder %>%
  filter( year == 2002, country == 'United States' )

## # A tibble: 1 × 6
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>
## 1 United States Americas   2002   77.3 287675526 39097.
```

Створювати складні фільтрації зручно з операторами: “ТА” (&), “АБО” (|), “НЕ” (!), а також функцію хог(), яка виключає АБО

```
filter(gapminder, year == 2007 & country == "China")

## # A tibble: 1 × 6
##   country continent  year lifeExp      pop gdpPercap
##   <fct>    <fct>    <int> <dbl>    <int>    <dbl>
## 1 China   Asia      2007   73.0 1318683096 4959.
```

Сортування рядків в порядку зростання (для числових змінних) або в алфавітному порядку (для категоріальних змінних)

```
gapminder %>%
  arrange(lifeExp)

## # A tibble: 1,704 × 6
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>
## 1 Rwanda      Africa    1992   23.6 7290203    737.
## 2 Afghanistan Asia      1952   28.8 8425333    779.
## 3 Gambia      Africa    1952    30 284320     485.
## 4 Angola      Africa    1952   30.0 4232095   3521.
## 5 Sierra Leone Africa    1952   30.3 2143249    880.
## 6 Afghanistan Asia      1957   30.3 9240934    821.
## 7 Cambodia    Asia      1977   31.2 6978607    525.
## 8 Mozambique  Africa    1952   31.3 6446316    469.
## 9 Sierra Leone Africa    1957   31.6 2295678   1004.
## 10 Burkina Faso Africa    1952   32.0 4469979    543.
## # [i] 1,694 more rows
```

Для сортування в порядку зменшення використаємо функцію desc() або знак “-” перед змінною

```
gapminder %>%
  arrange(desc(lifeExp))

## # A tibble: 1,704 × 6
##   country      continent  year lifeExp      pop gdpPercap
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>
## 1 Japan      Asia      2007   82.6 127467972 31656.
## 2 Hong Kong, China Asia      2007   82.2 6980412 39725.
## 3 Japan      Asia      2002    82 127065841 28605.
## 4 Iceland    Europe    2007   81.8 301931 36181.
## 5 Switzerland Europe    2007   81.7 7554661 37506.
## 6 Hong Kong, China Asia      2002   81.5 6762476 30209.
## 7 Australia  Oceania   2007   81.2 20434176 34435.
## 8 Spain      Europe    2007   80.9 40448191 28821.
## 9 Sweden     Europe    2007   80.9 9031088 33860.
```

```
## 10 Israel Asia 2007 80.7 6426679 25523.
## # [i] 1,694 more rows
```

```
arrange(gapminder, -pop)
```

```
## # A tibble: 1,704 × 6
##   country continent year lifeExp      pop gdpPerCap
##   <fct>    <fct>    <int> <dbl>    <int>    <dbl>
## 1 China   Asia      2007  73.0 1318683096  4959.
## 2 China   Asia      2002  72.0 1280400000  3119.
## 3 China   Asia      1997  70.4 1230075000  2289.
## 4 China   Asia      1992  68.7 1164970000  1656.
## 5 India   Asia      2007  64.7 1110396331  2452.
## 6 China   Asia      1987  67.3 1084035000  1379.
## 7 India   Asia      2002  62.9 1034172547  1747.
## 8 China   Asia      1982  65.5 1000281000   962.
## 9 India   Asia      1997  61.8  959000000  1459.
## 10 China  Asia      1977  64.0  943455000   741.
## # [i] 1,694 more rows
```

Створення нових змінних (стовпчиків)

```
gapminder %>%
  mutate(lifeExpMonths = 12*lifeExp)
```

```
## # A tibble: 1,704 × 7
##   country      continent year lifeExp      pop gdpPerCap lifeExpMonths
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>    <dbl>
## 1 Afghanistan Asia      1952  28.8  8425333    779.    346.
## 2 Afghanistan Asia      1957  30.3  9240934    821.    364.
## 3 Afghanistan Asia      1962  32.0 10267083    853.    384.
## 4 Afghanistan Asia      1967  34.0 11537966    836.    408.
## 5 Afghanistan Asia      1972  36.1 13079460    740.    433.
## 6 Afghanistan Asia      1977  38.4 14880372    786.    461.
## 7 Afghanistan Asia      1982  39.9 12881816    978.    478.
## 8 Afghanistan Asia      1987  40.8 13867957    852.    490.
## 9 Afghanistan Asia      1992  41.7 16317921    649.    500.
## 10 Afghanistan Asia      1997  41.8 22227415    635.    501.
## # [i] 1,694 more rows
```

```
gapminder %>%
  mutate(GDP_billion = gdpPerCap * pop / 10^9) %>%
  filter(year == 2007) %>%
  arrange(desc(GDP_billion))
```

```
## # A tibble: 142 × 7
##   country      continent year lifeExp      pop gdpPerCap GDP_billion
##   <fct>        <fct>    <int> <dbl>    <int>    <dbl>    <dbl>
## 1 United States Americas  2007  78.2 301139947  42952.  12934.
## 2 China        Asia      2007  73.0 1318683096  4959.   6540.
## 3 Japan        Asia      2007  82.6 127467972  31656.  4035.
## 4 India        Asia      2007  64.7 1110396331  2452.   2723.
## 5 Germany      Europe    2007  79.4  82400996  32170.  2651.
## 6 United Kingdom Europe    2007  79.4  60776238  33203.  2018.
## 7 France       Europe    2007  80.7  61083916  30470.  1861.
## 8 Brazil       Americas  2007  72.4 190010647  9066.   1723.
## 9 Italy        Europe    2007  80.5  58147733  28570.  1661.
```

```
## 10 Mexico      Americas  2007    76.2 108700891  11978.    1302.
## # [i] 132 more rows
```

Збереження відсортованого набору даних

```
gapminder_2007 <- filter(gapminder, year == 2007)
View(gapminder_2007)
```

Відбір змінних (стовпчиків)

```
select(gapminder, country, year, gdpPercap)
```

```
## # A tibble: 1,704 × 3
##   country      year gdpPercap
##   <fct>      <int>     <dbl>
## 1 Afghanistan  1952      779.
## 2 Afghanistan  1957      821.
## 3 Afghanistan  1962      853.
## 4 Afghanistan  1967      836.
## 5 Afghanistan  1972      740.
## 6 Afghanistan  1977      786.
## 7 Afghanistan  1982      978.
## 8 Afghanistan  1987      852.
## 9 Afghanistan  1992      649.
## 10 Afghanistan 1997      635.
## # [i] 1,694 more rows
```

```
gapminder %>%
  select(1, 3, 4)
```

```
## # A tibble: 1,704 × 3
##   country      year lifeExp
##   <fct>      <int>     <dbl>
## 1 Afghanistan  1952     28.8
## 2 Afghanistan  1957     30.3
## 3 Afghanistan  1962     32.0
## 4 Afghanistan  1967     34.0
## 5 Afghanistan  1972     36.1
## 6 Afghanistan  1977     38.4
## 7 Afghanistan  1982     39.9
## 8 Afghanistan  1987     40.8
## 9 Afghanistan  1992     41.7
## 10 Afghanistan 1997     41.8
## # [i] 1,694 more rows
```

Групування та агрегація даних - group_by(), summarise()

Знайдемо медіанне значення очікуваної тривалості життя при народженні у 2007 р.

```
gapminder %>%
  filter(year == 2007) %>%
  summarize(median_lifeExp = median(lifeExp))
```

```
## # A tibble: 1 × 1
##   median_lifeExp
##             <dbl>
## 1             71.9
```

Знайдемо середню тривалість життя та максимальний ВВП на душу населення на кожному континенті у 2007 р.

```
gapminder %>%
  filter(year == 2007) %>%
  group_by(continent) %>%
  summarize(median_lifeExp = median(lifeExp), max_gdpPercap = max(gdpPercap))

## # A tibble: 5 × 3
##   continent median_lifeExp max_gdpPercap
##   <fct>          <dbl>          <dbl>
## 1 Africa          52.9          13206.
## 2 Americas        72.9          42952.
## 3 Asia            72.4          47307.
## 4 Europe          78.6          49357.
## 5 Oceania         80.7          34435.
```

Згрупуємо дані за континентом та знайдемо середнє та медіанне значення очікуваної тривалості життя при народженні, а також кількість країн в кожній групі

```
gapminder %>%
  filter(year == 2007) %>%
  group_by(continent) %>%
  summarise(mean_lifeExp = mean(lifeExp),
            median_lifeExp = median(lifeExp),
            count = n()) %>%
  arrange(-mean_lifeExp)

## # A tibble: 5 × 4
##   continent mean_lifeExp median_lifeExp count
##   <fct>          <dbl>          <dbl> <int>
## 1 Oceania         80.7           80.7     2
## 2 Europe          77.6           78.6    30
## 3 Americas        73.6           72.9    25
## 4 Asia            70.7           72.4    33
## 5 Africa          54.8           52.9    52
```

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404.

Онлайн ресурси:

1. Introduction to the Tidyverse [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-the-tidyverse>

2. Основи економетрики в R. Маніпуляції з даними за допомогою dplyr і не тільки. Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_econometrics/dplyr.html

3. Data Manipulation with dplyr [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/data-manipulation-with-dplyr>

Лабораторна робота 4. Описова статистика в R Studio.

1. Завантажити один із вбудованих наборів даних у R Studio (наприклад, mtcars, iris або gapminder) або власний набір даних. Обчислити основні статистичні показники для числових змінних: середнє значення mean(), медіана median(), стандартне відхилення sd(), мінімальне та максимальне значення min(), max(), квантили quantile(), медіанне абсолютне відхилення (median absolute deviation) mad(), створити власну функцію для розрахунку лінійного відхилення. Перевірити розподіл змінних за допомогою функції summary(). З використанням бібліотеки psych обчислити коефіцієнти асиметрії та ексцесу за допомогою відповідних функцій skew() та kurtosi(), проаналізувати показники описової статистики обчислені за допомогою функції describe(). Ознайомитися з пакетом skimr та його можливостями для швидкого аналізу описової статистики, використовуючи функцію skim() для отримання ключових характеристик.

2. Побудувати частотну таблицю для категоріальної змінної table(). Візуалізувати частоти за допомогою: гістограми hist() для числових змінних, стовпчикової діаграми barplot() для категоріальних змінних, боксплоту boxplot() для аналізу розподілу змінної та виявлення викидів.

3. Побудувати Q-Q графік qqnorm() для оцінки нормальності розподілу числової змінної. Додати теоретичну лінію нормального розподілу qqline(). Виявити аномальні значення (аутлайери) у вибірці: використати boxplot.stats() для отримання викидів. Виконати перевірку за правилом $1.5 * IQR$ (межі: $Q1 - 1.5 * IQR$ і $Q3 + 1.5 * IQR$). Описати отримані результати аналізу.

4. Перевірити відповідність вибірки нормальному розподілу за допомогою

статистичних критеріїв: Шапіро-Уїлка (`shapiro.test()`), Колмогорова-Смирнова (`ks.test()`), а також Ліллієфорса (`lillie.test()` з пакету `portest`). Проаналізувати результати тестів і зробити висновки щодо нормальності розподілу даних.

5. Лабораторна робота має бути оформлена у вигляді звіту у форматі `html` за допомогою Quarto або Rmarkdown.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.

2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.

3. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

4. An introduction to Statistical Learning with Applications in R / Springer, 2021

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

2. Introduction to Statistics in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-statistics-in-r>

3. Prometheus: Аналіз даних та статистичне виведення на мові R. [Електронний ресурс] - Режим доступу: <https://prometheus.org.ua/prometheus-free/data-analysis-statistics/>

Лабораторна робота 5. Перевірка гіпотез в R Studio.

1. Виберіть та імпортуйте набір даних для аналізу (використовуйте вбудовані набори R або завантажте власні дані). Проведіть очищення даних, перевірку на пропуски за допомогою пакету `dplyr`.

2. Перевірка гіпотез в R Studio.

2.1. Проведіть t-тести: перевірка гіпотези про середнє значення однієї вибірки; порівняння середніх двох незалежних вибірок; порівняння середніх двох залежних вибірок. t-тести для порівняння середніх в R Studio:

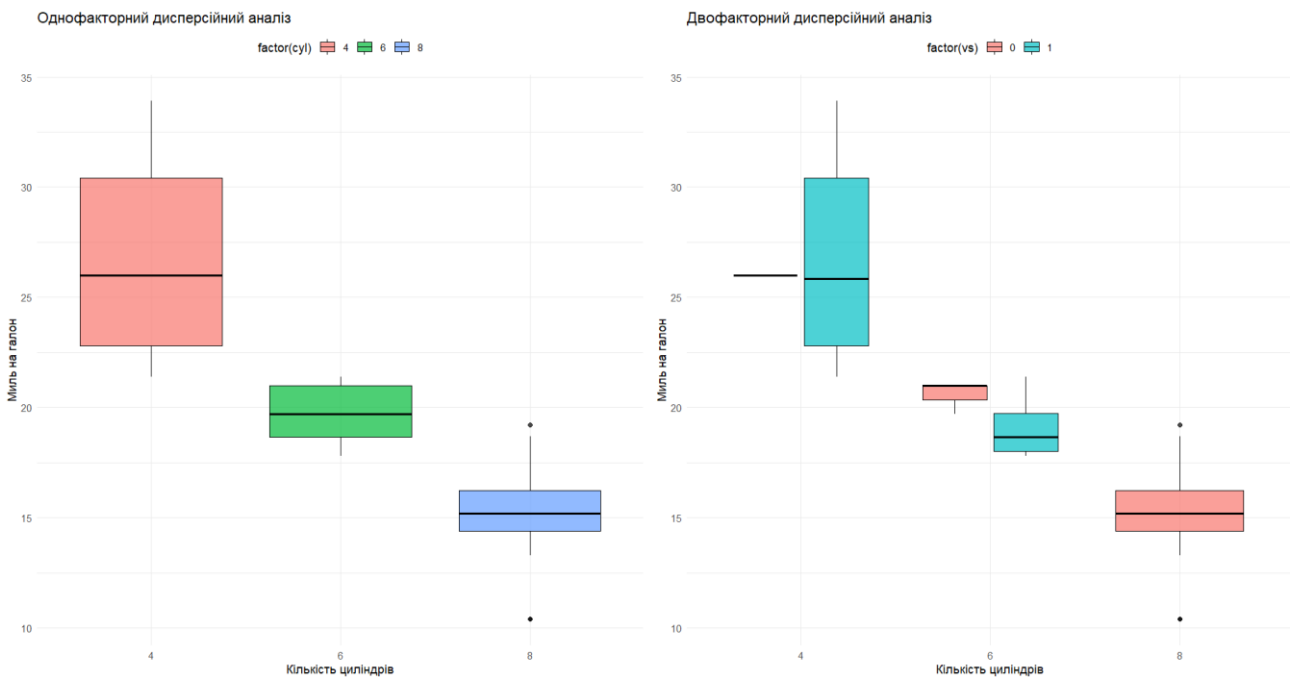
- ✓ Одновибірковий t-тест: `t.test(x, mu = ...)`.
- ✓ Двовибірковий t-тест (незалежні вибірки): `t.test(x, y, var.equal = TRUE)`. Якщо дисперсії двох вибірок відрізняються, слід використовувати аргумент `var.equal = FALSE` у функції `t.test()`.
- ✓ Парний t-тест: `t.test(x, y, paired = TRUE)`.

** Припущення про нормальний розподіл даних або достатньо великий розмір вибірки для застосування центральної граничної теореми.*

2.2. Дисперсійний аналіз (ANOVA) для порівняння більше ніж двох середніх в R Studio:

- ✓ Проведення однофакторного ANOVA: `aov(response ~ factor, data = dataset)`. Використання функції `summary()` для аналізу результатів.
- ✓ Перевірка припущень ANOVA (гомогенність дисперсій, нормальність залишків). Пост-хок (Post-hoc) тести: після проведення ANOVA, якщо виявлено значущі відмінності між групами, слід застосувати пост-хок тести (наприклад, Tukey HSD, Dunnett's test) для визначення, які саме групи відрізняються між собою.
- ✓ Двофакторний ANOVA: виконайте ANOVA для порівняння середніх значень більше ніж двох груп. Проведіть пост-хок тести для виявлення відмінностей між парами груп.

** Використання візуальних методів (гістограми, box-plotів, Q-Q графіки, графіки залишків) поряд зі статистичними тестами дозволяє отримати більш повну картину про розподіл даних та виконання припущень ANOVA. Візуалізація результатів з використання пакету `ggplot2` для створення більш складних графіків:*



2.3. Проведіть χ^2 -тест (хи-квадрат тест) для перевірки незалежності двох категоріальних змінних. χ^2 -тест для аналізу частот в R Studio:

- ✓ Перевірка незалежності категорійних змінних: `chisq.test(table(x, y))`.
- ✓ Аналіз значення р-значення та відповідність очікуваних і спостережуваних частот.

3. Виконайте тести Манна-Уїтні, Крускала-Уолліса для даних, що не відповідають нормальному розподілу.

4. Інтерпретуйте результати: аналіз р-значення та прийняття рішення про відхилення або невідхилення нульової гіпотези, обчислення довірчих інтервалів для оцінки параметрів.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>

2. Introduction to Statistics in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-statistics-in-r>

3. Prometheus: Аналіз даних та статистичне виведення на мові R. [Електронний ресурс] - Режим доступу: <https://prometheus.org.ua/prometheus-free/data-analysis-statistics/>

Лабораторна робота 6. Регресійні моделі в R Studio.

Побудувати моделі лінійної, множинної та логістичної регресії в R Studio, оцінити їх параметри та якість, виконати аналіз взаємозв'язків змінних у множинній регресії та дослідити застосування логістичної регресії для класифікаційних задач, оцінюючи точність моделі та аналіз залишків:

1. Обрати та завантажити набір даних (наприклад, скориставшись наступними джерелами: сервіс з пошуку даних від Google Dataset Search, репозитарій з машинного навчання UCI, Дані Всесвітнього банку, платформа Kaggle). Набір даних має включати більше 3 числових змінних та 2-3 категоріальні. За потреби приведіть дані до охайного вигляду, проведіть попередній аналіз (описова статистика, візуалізація).

2. Побудувати в R Studio модель лінійної регресії, яка пояснює залежність між двома числовими змінними. Побудуйте точкову діаграму та обчисліть коефіцієнт кореляції. Зробіть висновки. Побудуйте модель простої лінійної регресії за допомогою функції `lm()`. Запишіть рівняння регресії, інтерпретуйте отримані коефіцієнти регресії. Відобразіть на графіку точкової діаграми лінію регресії. Проведіть аналіз якості моделі. Обчисліть стандартні похибки оцінок параметрів моделі. Перевірте значущість коефіцієнтів регресії за допомогою t-тесту. Зробіть висновки щодо гіпотез. Побудуйте довірчі інтервали до оцінок параметрів моделі. Обчисліть коефіцієнт детермінації (R^2). Зробіть

висновок щодо його значення. Побудуйте точковий та інтервальний прогноз за побудованою моделлю за довільними значеннями незалежної змінної.

3. Розширення моделі: додайте до моделі з попереднього завдання ще один предиктор. Порівняйте моделі з одним і двома предикторами за допомогою AIC і BIC. Оберіть найкращу модель на основі отриманих результатів. Перевірте наявність мультиколінеарності між предикторами. Додайте до моделі взаємодію між предикторами (якщо доцільно).

4. Побудуйте логістичну модель для прогнозування ймовірності неповернення позики згідно свого варіанту на основі наявних предикторів:

Наявні наступні дані:	
x1	вік позичальника
x2	частка одночасно діючих угод від їх загальної кількості
x3	кількість поточних запитів в БКІ стосовно кредитної історії позичальника
x4	частка вчасно завершених угод від їх загальної кількості
x5	середнє значення кількості прострочених періодів за останні 3 місяці по всіх завершених попередніх угодах
Клас позичальника - значення 0 встановлюється для позичальників, які вчасно виконали зобов'язання за кредитом, 1 - встановлюється для позичальників, які мають борг.	

Переносимо дані з Excel в R у вигляді датасету Logic. В даному випадку незалежними змінними (факторами) будуть коефіцієнти від x1 до x5, залежною буде бінарна змінна Class. Побудуємо логістичну модель залежної бінарної змінної від факторів:

#Побудова моделі логістичної регресії

```
model <- glm(Class ~ x1 + x2 + x3 + x4 + x5, data = Logic, family = binomial)
summary(model)
```

Проаналізуйте детальний звіт, який містить показники, що характеризують якість моделі та значимість предикторів. Визначте передбачені даною моделлю значення залежної змінної, зробіть перетворення її в бінарну змінну (якщо значення більше 0,5, то клас буде 1, якщо менше – 0) та додайте її в таблицю значень. Встановіть та завантажте пакет "lmtest", після чого виконайте перевірку адекватності логістичної моделі за допомогою тесту Вальда, розраховуючи статистику Хі-квадрат. Для оцінки якості визначте відсоток правильно класифікованих об'єктів.

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404.
4. Основи статистичного моделювання: навч. посібник / за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП «Рута», 2022. 604 с.
5. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>
2. Introduction to Regression in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-regression-in-r>
3. Intermediate Regression in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/intermediate-regression-in-r>
4. Основи економетрики в R. Ігор Мірошніченко. [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_econometrics/index.html

Лабораторна робота 7. Візуалізація даних в R Studio.

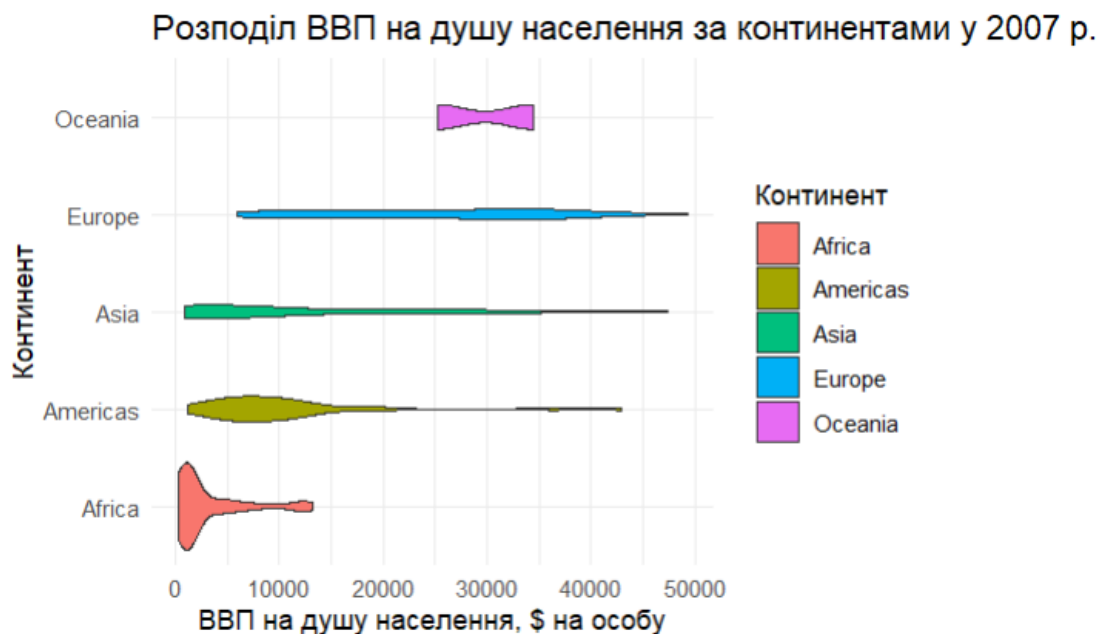
1. Основи візуалізації в R Studio. Завантажте власний набір даних. Побудуйте гістограму для числових даних. Побудуйте коробковий графік (box plot) для порівняння розподілів. Створіть точкову діаграму для візуалізації зв'язку між двома змінними. Додайте назви графіків, підписи осей, змініть кольори та розміри елементів графіків.
2. Інсталюйте та завантажте пакет ggplot2. Ознайомтеся з основними елементами графіків в ggplot2: дані, естетика, геометричні об'єкти, фасети, теми.

Створіть різноманітні типи графіків за допомогою ggplot2 (scatter plots, line plots, histograms, box plots, violin plots, density plots, bar plots, pie charts, heatmaps, mosaic plots, network graphs тощо). Додайте до графіків легенди, підписи, фасети, змініть кольори, розміри та інші параметри.

Приклад коду:

```
# Завантажуємо пакети
library(gapminder)
library(dplyr)
library(ggplot2)

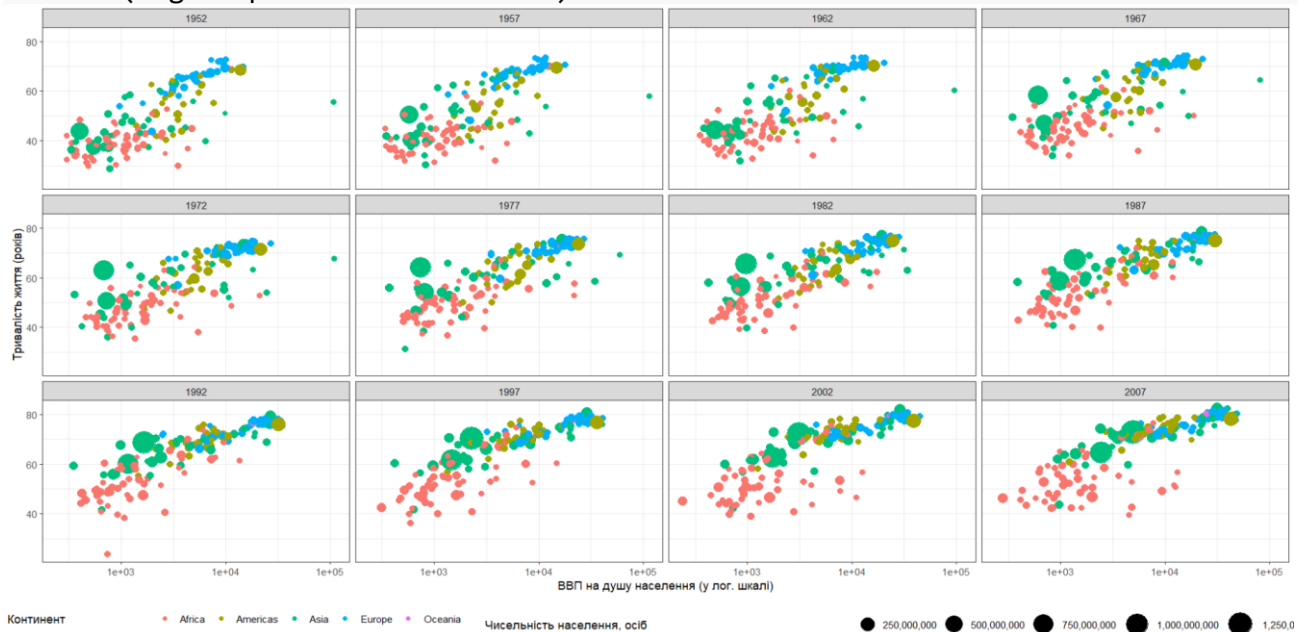
# Скрипичний графік
ggplot(gapminder_2007, aes(x = gdpPercap, y = continent, fill = continent)) +
  geom_violin() +
  labs(
    x = "ВВП на душу населення, $ на особу",
    y = "Континент",
    fill = "Континент",
    title = "Розподіл ВВП на душу населення за континентами у 2007 р."
  ) +
  theme_minimal()
```



Зв'язок між ВВП на душу населення та тривалістю життя, кольором представлений континент, розмір точок - чисельність населення, у розрізі (фасети) за роками

```
ggplot(gapminder, aes(x = gdpPercap, y = lifeExp, color = continent, size = pop
)) +
  geom_point() +
  scale_x_log10() +
  facet_wrap(~year) +
  labs(x = "ВВП на душу населення (у лог. шкалі)", y = "Тривалість життя (років)", color = "Континент", size = "Чисельність населення, осіб") +
```

```
scale_size_continuous(labels = scales::comma, range = c(2, 10)) + # Обмежити
розмір точок
theme_bw() +
theme(legend.position = "bottom")
```



3. Інтерактивна візуалізація даних. Інсталюйте та завантажте пакет `plotly`. Створіть інтерактивні графіки, які дозволяють користувачам взаємодіяти з даними (наприклад, переглядати деталі при наведенні курсору). Дослідіть інші пакети для інтерактивної візуалізації (наприклад, `shiny`, `leaflet`).

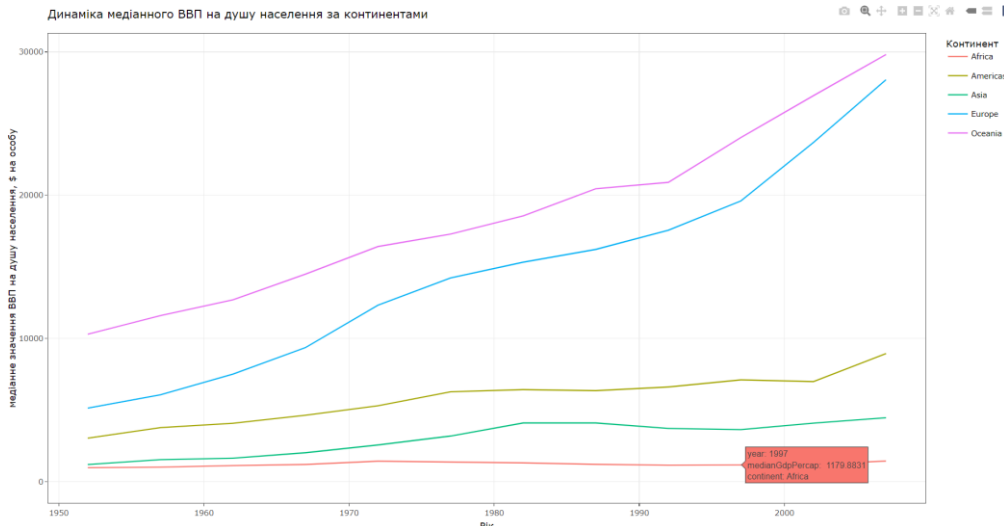
Приклад коду:

```
# Завантажуємо пакети
library(gapminder)
library(dplyr)
library(ggplot2)
library(plotly)

# Знайдемо медіанне значення ВВП на душу населення за роками та континентами
by_year_continent <- gapminder %>%
  group_by(year, continent) %>%
  summarize(medianGdpPerCap = median(gdpPerCap))

# Лінійний графік ggplot2
plot <- ggplot(by_year_continent, aes(x = year, y = medianGdpPerCap, color = co
ntinent)) +
  geom_line() +
  expand_limits(y = 0) +
  labs(title = "Динаміка медіанного ВВП на душу населення за континентами",
       x = "Рік",
       y = "Медіанне значення ВВП на душу населення, $ на особу",
       color = "Континент")

# Перетворення в інтерактивний графік за допомогою plotly
ggplotly(plot)
```



Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. Claus O. Wilke. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. 1st Edition. O'Reilly Media. 2019. 387 p. — ISBN 1492031089, 978-1492031086.
4. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>
2. Introduction to Data Visualization with ggplot2 [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-data-visualization-with-ggplot2>
3. Intermediate Data Visualization with ggplot2 [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/intermediate-data-visualization-with-ggplot2>

4. ggplot2: Elegant Graphics for Data Analysis (3e). [Електронний ресурс] -
Режим доступу: <https://ggplot2-book.org/>

Лабораторна робота 8. Часові ряди та їх візуалізація.

1. Вступ до часових рядів у R та підготовка даних. Завантажте набір даних, що містить часовий ряд (наприклад, дані про щомісячні продажі, температуру, кількість відвідувачів сайту). Перетворіть дані в об'єкт `tsibble` за допомогою відповідних функцій (наприклад, `as_tsibble()`). Ознайомтеся з основними функціями для роботи з об'єктами `tsibble` (наприклад, `index()`, `key()`, `filter_index()`). Виконайте операції фільтрації, агрегації та трансформації даних. Перевірте дані на наявність пропущених значень та аномалій. Використовуйте методи інтерполяції або заміни для обробки пропущених значень (наприклад, функція `na.interp()` з пакету `forecast`, функція `na.replace()` з пакету `imputeTS`).

2. Аналіз характеристик часових рядів та моделювання. Побудуйте графік часового ряду для візуалізації його динаміки. Використовуйте різні типи графіків для дослідження різних аспектів часового ряду (наприклад, графік розкладання, графік автокореляції). Визначте наявність тренду та сезонності в часовому ряді за допомогою візуального аналізу та статистичних методів (наприклад, `decompose()`, `acf()`). Проаналізуйте автокореляцію часового ряду за допомогою функції `acf()`. Побудуйте базові моделі часових рядів (наприклад, модель ковзного середнього, модель експоненційного згладжування). Оцініть якість побудованих моделей за допомогою відповідних статистичних показників (наприклад, AIC, BIC).

3. Прогнозування та оцінка моделей. Використайте методи прогнозування часових рядів (наприклад, метод експоненційного згладжування, ARIMA) для прогнозування майбутніх значень. Розділіть дані на навчальну та тестову вибірки для оцінки якості прогнозування. Оцініть точність прогнозів за допомогою відповідних статистичних показників (наприклад, RMSE, MAE). Порівняйте результати прогнозування різних моделей. Візуалізуйте прогнозовані значення разом з фактичними значеннями на одному графіку.

Використовуйте різні типи графіків для візуалізації результатів прогнозування (наприклад, графік з довірчими інтервалами).

Список рекомендованих джерел:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 2nd Edition. O'Reilly. 2023. 576 p. — ISBN 1492097403, 978-1492097402.
2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition. O'Reilly Media. 2019. 598 p. — ISBN 1492040681, 978-1492040682.
3. Shumway, Robert & Stoffer, David. (2011). Time Series Analysis and Its Applications With R Examples. 10.1007/978-1-4419-7865-3.
4. David Spiegelhalter. The Art of Statistics: Learning from Data / Basic books, 2019. 448 p. — ISBN 1541618513, 978-1541618510.
5. Hyndman R.J., Athanasopoulos G. "Forecasting: Principles and Practice". OTexts. 2013. 292 p. — ISBN 0987507109, 978-0987507105.

Онлайн ресурси:

1. R Programming for Statistics and Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>
2. Time Series Analysis in R [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/time-series-analysis-in-r>
3. Manipulating Time Series Data in R [Електронний ресурс]:дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/manipulating-time-series-data-in-r>
4. Основи прогнозування в R. Часові ряди та їх візуалізація. Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_forecasting/ts.html

V. ІНДИВІДУАЛЬНА САМОСТІЙНА РОБОТА

Загальні вимоги щодо виконання самостійної роботи

Індивідуальна розрахункова робота виконується за допомогою прикладного пакету аналізу «R-studio», подається у вигляді звіту у форматі html у вигляді презентації із зазначенням основних здобутків дослідження. Виконання самостійної роботи передбачає опрацювання студентом матеріалів лекційних і лабораторних занять, рекомендованих по темах курсу. Під час виконання самостійної роботи студенти мають керуватись програмою курсу та Методичними рекомендаціями до самостійного вивчення тем курсу «Прикладний пакет аналізу «R-studio». Самостійна робота складається з виконання та захисту індивідуального розрахункового проекту з послідовними завданнями навчально – дослідницького характеру.

Завдання

1. Вибір набору даних:

Самостійно обрати набір даних для аналізу. Набір даних повинен бути релевантним до спеціальності та сфери інтересів. Набір даних повинен містити достатню кількість змінних та спостережень для проведення аналізу. Джерелом даних можуть бути відкриті бази даних, наукові статті, веб-сайти тощо.

2. Підготовка даних:

Завантажити набір даних в R-studio. Провести очищення даних: обробка пропущених значень, видалення дублікатів, виявлення та обробка аномалій. Виконати трансформацію даних: створення нових змінних, перетворення типів даних тощо. Описати структуру та основні характеристики набору даних.

3. Розвідувальний аналіз даних:

Виконати візуалізацію даних за допомогою різних типів графіків (гістограми, діаграми розсіювання, коробкові діаграми тощо). Розрахувати основні статистичні показники (середнє, медіана, стандартне відхилення, кореляція тощо). Виявити та описати закономірності та зв'язки між змінними.

4. Статистичний аналіз:

Вибрати та застосувати відповідні статистичні методи для аналізу даних (t-тест, ANOVA, регресійний аналіз, кластерний аналіз тощо). Сформулювати статистичні гіпотези та перевірити їх за допомогою обраних методів. Інтерпретувати результати статистичного аналізу та зробити висновки.

5. Візуалізація результатів:

Створити графіки та діаграми для наочного представлення результатів аналізу. Використовувати інтерактивні графіки (за допомогою пакету plotly) для покращення візуалізації.

6. Оформлення роботи:

Робота має бути оформлена у вигляді звіту у форматі html, використовуючи Quarto або R Markdown. Звіт повинен містити:

- ✓ Опис набору даних та методів аналізу.
- ✓ Результати дослідницького та статистичного аналізу.
- ✓ Інтерпретацію результатів та висновки.

7. Захист роботи:

Підготувати презентацію для захисту роботи. Під час захисту продемонструвати розуміння використаних методів та результатів аналізу. Відповісти на запитання.

Рекомендації

Використовуйте матеріали лекційних та лабораторних занять. Використовуйте рекомендовану літературу та онлайн-ресурси. Консультуйтеся з викладачем у разі виникнення питань. Дотримуйтеся термінів виконання роботи.

VI. РЕЗУЛЬТАТИ НАВЧАННЯ. СИСТЕМА ОЦІНКИ ЗНАТЬ

Результат навчання (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у Підсумковій оцінці з дисципліни
Код	Результат навчання			
ЗМІСТОВНИЙ МОДУЛЬ 1. ПРИКЛАДНИЙ ПАКЕТ АНАЛІЗУ «R-STUDIO»				
1.1	Знати прикладний пакет аналізу даних R studio	Лекція, лабораторні заняття	Тест, 60% правильних відповідей, бліц опитування, творчі завдання, Case study, презентація, іспит	4
1.2	Знати як завантажувати, зчитувати, отримувати та очищати дані в R Studio	Лекція, лабораторні заняття		4
1.3	Знати як працювати з пакетами Tidyverse	Лекція, лабораторні заняття		4
1.4	Знати як проводити розвідувальний аналіз в R Studio	Лекція, лабораторні заняття		4
1.5	Знати як перевіряти статистичні гіпотези в R Studio	Лекція, лабораторні заняття		4
1.6	Знати як проводити аналіз часових рядів в R Studio	Лекція, лабораторні заняття		4
1.7	Знати проводити регресійний аналіз в R Studio	Лекція, лабораторні заняття	Індивідуальний захист виконання завдання, тестування, іспит	4
1.8	Знати як візуалізувати дані в R Studio	Лекція, лабораторні заняття		4
2.1	Вміти працювати з R Studio	Лабораторна робота №1		4
2.2	Вміти завантажувати, зчитувати, отримувати та очищати дані в в R Studio	Лабораторна робота №2		4
2.3	Вміти проводити розвідувальний аналіз з R Studio	Лабораторна робота №3		4
2.4	Вміти перевіряти гіпотези в R Studio	Лабораторна робота №4		4
2.5	Вміти проводити регресійний аналіз в R Studio	Лабораторна робота №5		4
2.6	Вміти проводити аналіз часових рядів в R Studio	Лабораторна робота №6	4	
2.7	Вміти візуалізувати дані в в R Studio	Лабораторна робота №7	4	
3.1	Статистичний аналіз в середовищі R Studio.	Лабораторна робота №7	Індивідуальний захист виконання завдання	20
4.	Продемонструвати розуміння особистої відповідальності за професійні та управлінські рішення чи надані пропозиції, які можуть впливати на результати роботи з прикладним пакетом аналізу даних R studio	Самостійна робота	Захист результатів виконання самостійної роботи	До 20

Схема формування підсумкової оцінки

1. Виконання лабораторних робіт, презентація та захист самостійної індивідуальної роботи (РН 1.1 -1.6; 2.1-2.7; 3; 4) – 48 балів / 30 балів;

2. Самостійна робота – розрахункове індивідуальне завдання (РН 4) – 12 балів / 6 балів;

- підсумкове оцінювання у формі іспиту.

Студент не допускається до складання іспиту, якщо протягом семестру отримав менше, ніж 36 балів. Якщо оцінка студента на іспиті є нижчою від мінімального порогового рівня (24 бала), то бали за іспит не додаються до семестрової оцінки (вважаються рівними нулю), а підсумкова оцінка із дисципліни є незадовільною.

На іспит виноситься 2 блока завдань:

- 1 блок – тестове опитування;
- 2 блок – практична задача в R Studio.

	Семестрова кількість балів	Іспит	Підсумкова оцінка
Мінімум	36	24	60
Максимум	60	40	100

Організація оцінювання:

Семестрову оцінку формують бали, отримані студентом у процесі засвоєння матеріалу з усіх тем дисципліни. Протягом семестру проводяться електронні контрольні роботи із тестовими питаннями. Семестрове оцінювання таких видів робіт, як: виступи з обговорення теоретичних питань, виконання практичних вправ та розрахунково-аналітичних завдань, контрольні роботи здійснюється під час лабораторних занять за розкладом, у визначеній в ньому аудиторії. Самостійна робота (розрахунково-аналітична) подається викладачу на перевірку на останньому тижні семестру. Критерії оцінювання: підготовка набору даних, якість проведеного аналізу, обґрунтованість висновків, якість візуалізації даних та результатів, оформлення звіту, презентації та якість захисту

роботи.

Результатом виконання студентом лабораторних робіт є звіт, що підлягає обов'язковому захисту шляхом усного опитування за темою роботи. Виконання та захист лабораторних робіт відбувається протягом семестру відповідно до затвердженого розкладу занять. Лабораторна робота оцінюється за 6-бальною шкалою, максимальний бал – якісно оформлений звіт з лабораторної роботи, під час захисту якого студент демонструє високий рівень знань теоретичного та практичного матеріалу за темою дослідження, правильну інтерпретацію результатів розрахунків. При оцінюванні виконання лабораторних робіт увага також приділяється якості, самостійності та своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу).

Відпрацювання пропущених занять здійснюється шляхом виконання студентом додаткових розрахункових завдань з дисципліни.

** У випадку виникнення додаткових питань та складних ситуацій під час навчального процесу, їх вирішення здійснюється з посиланням на засади Положення про організацію освітнього процесу у Київському національному університеті імені Тараса Шевченка.*

Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Основна література

- 1.1. An introduction to Statistical Learning with Applications in R / Springer, 2021
- 1.2. R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics 2nd Edition
- 1.3. Sarah Stowell. Using R for Statistics. Apress, 2014. ISBN 978-1484201404
- 1.4. Vikram Dayal. An Introduction to R for Quantitative Economics: Graphing, Simulating and Computing. Springer, 2015. ISBN 978-81-322-2340-5
- 1.5. David Spiegelhalter The Art of Statistics: Learning from Data / Basic books, 2019

2. Додаткова література

- 2.1. Victor A. Bloomfield. Using R for Numerical Analysis in Science and Engineering. Chapman & Hall/CRC, 2014. ISBN 978-1439884485
- 2.2. Torsten Hothorn and Brian S. Everitt. A Handbook of Statistical Analyses Using R. Chapman & Hall/CRC Press, Boca Raton, Florida, USA, 3rd edition, 2014. ISBN 978-1-4822-0458-2
- 2.3. Michael J. Crawley. Statistics: An Introduction using R. Wiley, 2nd edition, 2014. ISBN 978-1-118-94109-6
- 2.4. Yihui Xie. Dynamic Documents with R and knitr. Chapman & Hall/CRC, 2013. ISBN 978-1482203530.
- 2.5. Osypova, O., Horna, M., Vashchaiev, S., Ishchuk, Y., & Pomazun, O. (2023). Convergence of food consumption across Ukrainian regions: approach using spatial panel data models. Agricultural and Resource Economics: International Scientific E-Journal, 9(1), 28-43
- 2.6. C. Sun. Empirical Research in Economics: Growing up with R. Pine Square, Starkville, Mississippi, USA, 1st edition, 2015
- 2.7. Горна М.О. «Можливості R та Python у статистичному аналізі, машинному навчанні та візуалізації даних» XXII міжнародна науково-практична конференція з нагоди Дня працівників статистики «Сучасна

статистика: проблеми та перспективи розвитку», Київ, 6 грудня 2024 р.

3. Електронні ресурси

- 3.1. Prometheus: Аналіз даних та статистичне виведення на мові R [Електронний ресурс]: дистанційний курс . Режим доступу: https://courses.prometheus.org.ua/courses/IRF/Stat101/2016_T3/about
- 3.2. Introduction to Statistics in R [Електронний ресурс]: дистанційний курс. Режим доступу: <https://app.datacamp.com/learn/courses/introduction-to-statistics-in-r>
- 3.3. Основи економетрики в R, автор Ігор Мірошніченко [Електронний ресурс] - Режим доступу: https://aranaur.github.io/r_forecasting/
- 3.4. Udey: R Programming for Statistics and Data Science — курс для початківців і тих, хто хоче використовувати R у статистиці та Data Science [Електронний ресурс]: дистанційний курс . Режим доступу: <https://ua.udemy.com/course/r-programming-for-statistics-and-data-science>.
- 3.5. Coursera: Data Science: Foundations using R (Johns Hopkins University) — серія з 5 курсів, що охоплює основи Data Science та програмування в R. [Електронний ресурс]: дистанційний курс. Режим доступу: <https://www.coursera.org/specializations/data-science-foundations-r>
- 3.6. R for Data Science (Hadley Wickham & Garrett Golemund) — безкоштовна онлайн-книга, яка охоплює tidyverse, візуалізацію даних, очистку даних та моделювання. [Електронний ресурс]. Режим доступу: <https://r4ds.had.co.nz/>
- 3.7. Swirl: Learn R, in R — навчальний пакет, який дозволяє вивчати R безпосередньо у консолі RStudio. [Електронний ресурс]. Режим доступу: <https://swirlstats.com/>
- 3.8. R-bloggers — платформа зі щоденними статтями, туторіалами, порадами та прикладами застосування R. [Електронний ресурс]. Режим доступу: <https://www.r-bloggers.com/>
- 3.9. R Graph Gallery — колекція прикладів побудови графіків в R (зокрема

- ggplot2, plotly, base R). [Електронний ресурс]. Режим доступу: <https://r-graph-gallery.com/>
- 3.10.Kaggle [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/>
- 3.11.Stanford Online: Statistical Learning (with R) — безкоштовний курс Стенфордського університету, автори — Trevor Hastie та Rob Tibshirani; містить відеолекції та практику з використанням R. [Електронний ресурс]. Режим доступу: <https://online.stanford.edu/courses/sohs-ystatslearning-statistical-learning-r>
- 3.12.HarvardX: Data Science: R Basics (edX) — курс з основ програмування на R для аналізу даних, частина сертифікатної програми Data Science від Гарварду. [Електронний ресурс]. Режим доступу: <https://www.edx.org/learn/r-programming/harvard-university-data-science-r-basics>
- 3.13.Advanced R (Hadley Wickham) — безкоштовна онлайн-книга для просунутого вивчення R. [Електронний ресурс]. Режим доступу: <https://adv-r.hadley.nz/>
- 3.14.TidyTuesday — спільнота та серія щотижневих викликів із використанням реальних наборів даних у R для практики tidyverse, візуалізації та звітів. [Електронний ресурс]. Режим доступу: <https://github.com/rfordatascience/tidytuesday>

Навчальне видання

ГОРНА Марина Олексіївна

ПРИКЛАДНИЙ ПАКЕТ АНАЛІЗУ «R-STUDIO»

Навчально-методичний комплекс

(для студентів спеціальності - 05 Соціальні та поведінкові науки, освітньої програми «Економічна аналітика та статистика»)

(Укр. мовою)

Друкується за авторською редакцією

Тираж 100 прим.