

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**КАФЕДРА ЕКОНОМІЧНОЇ КІБЕРНЕТИКИ І СИСТЕМНОГО АНАЛІЗУ**

Рівень вищої освіти	Перший (бакалаврський)
Спеціальність	Системний аналіз
Освітня програма	Управління складними системами
Група	6.04.124.020.19.1

## **ДИПЛОМНИЙ ПРОЄКТ**

**на тему: «Розроблення інтерактивного застосунку оцінювання рівня енергетичної безпеки»**

**Виконав: студент Ілля ТОРОПОВ**

**Керівник: к.е.н., доцент Роман ЯЦЕНКО**

**Рецензент: доцент кафедри інформаційних систем ХНЕУ ім. С. Кузнеця  
кандидат технічних наук Парфьонов Юрій Едуардович**

**Харків – 2023 рік**

## РЕФЕРАТ

Звіт про дипломну роботу: 67 сторінки, 3 розділи, 19 рисунків, 2 таблиці, 3 додатки, 40 джерел.

Об'єктом дослідження є процеси управління енергетичною безпекою.

Мета дослідження – проведення кластеризації країн за рівнем енергетичної безпеки, з використанням методів ієрархічної кластеризації і кластеризації методом k-середніх, та розробка інтерактивного застосунку для відображення країн за їх рівнем енергетичної безпеки.

Розглянуто сутність та поняття енергетичної безпеки, інтерпретація поняття енергетичної безпеки різними видавництвами. Розглянуто основні фактори впливу на рівень енергетичної безпеки та фактори оцінки рівня енергетичної безпеки країн. Переглянуто інформацію про мову програмування R, застосування кластерного аналізу у мові програмування R, його методи та способи реалізації. Розглянуто та проаналізовано інформацію щодо потрібних даних для проведення ієрархічного кластерного аналізу та кластеризації методом k-середніх. Розроблена моделі класифікації країн за рівнем енергетичної безпеки у мові програмування R. Створено інтерактивний застосунок оцінки рівня енергетичної безпеки за допомогою пакета даних Shiny, з використанням результатів проведеного кластерного аналізу. Проведено аналіз отриманих результатів кластерного аналізу за різними методами.

**КЛЮЧОВІ СЛОВА:** ЕНЕРГОБЕЗПЕКА, КЛАСТЕРНИЙ АНАЛІЗ, РІВЕНЬ, ІНТЕРАКТИВНИЙ ЗАСТОСУНОК, ІМПОРТ.

## ABSTRACT

Dissertation report: 67 pages, 3 chapters, 19 figures, 2 tables, 3 appendices, 40 sources.

The object of research is the processes of energy security management.

The purpose of the study is to cluster countries by the level of energy security, using hierarchical clustering and k-means clustering methods, and to develop an interactive application to display countries by their level of energy security.

The essence and concept of energy security, the interpretation of the concept of energy security by different publishers are considered. The main factors influencing the level of energy security and factors for assessing the level of energy security of countries are considered. The information about the R programming language, the use of cluster analysis in the R programming language, its methods and methods of implementation are reviewed. The information on the necessary data for hierarchical cluster analysis and clustering by the k-means method is considered and analyzed. Models of classification of countries by the level of energy security in the R programming language are developed. An interactive application for assessing the level of energy security using the Shiny data package, using the results of the cluster analysis, was created. The results of the cluster analysis are analyzed by different methods.

**KEYWORDS:** ENERGY SECURITY, CLUSTER ANALYSIS, LEVEL, INTERACTIVE APPLICATION, IMPORT.

## ЗМІСТ

ВСТУП .....	9
РОЗДІЛ 1. ТЕОРЕТИЧНА СУТНІСТЬ ЕНЕРГЕТИЧНОЇ БЕЗПЕКИ КРАЇНИ .....	11
1.1 Поняття енергетичної безпеки.....	11
1.2 Фактори впливу на рівень енергетичної безпеки .....	14
1.3 Фактори оцінки рівня енергетичної безпеки країни .....	17
РОЗДІЛ 2. ІНТЕРАКТИВНІ МЕТОДИ ТА МОДЕЛІ КЛАСТЕРИЗАЦІЇ В СЕРЕДОВИЩІ R.....	22
2.1 Методи кластерного аналізу в середовищі R.....	22
2.2 Інтерактивне середовище Shiny.....	35
РОЗДІЛ 3. ІНТЕРАКТИВНИЙ ЗАСТОСУНОК ОЦІНЮВАННЯ РІВНЯ ЕНЕРГЕТИЧНОЇ БЕЗПЕКИ .....	42
3.1 Моделі класифікації країн за рівнем енергетичної безпеки .....	42
3.2 Створення інтерактивного застосунку за допомогою Shiny .....	48
3.3 Порівняння результатів класифікації в інтерактивному застосунку.....	52
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	62
ДОДАТКИ.....	67

## ВСТУП

У нинішніх умовах очевидно, що енергетичний сектор є важливою частиною економічної структури кожної країни і має значний вплив на конкурентоспроможність її економіки. Тому підтримання енергетичної безпеки має важливе значення для загальної безпеки країни. З точки зору військової безпеки, уряди, які мають енергетичну галузь, здатну задовольнити всі потреби за будь-яких обставин, мають перевагу. Як наслідок, під час кожної військової операції першочерговим завданням ворога є виведення з ладу життєво важливої інфраструктури енергопостачання. Адже через нестачу електроенергії може значно погіршитися життя звичайних споживачів та навіть статися крах економіки.

Після широкомасштабного вторгнення Росії, що призвело до масових обстрілів енергетичної інфраструктури, енергетична безпека стала вирішальним питанням для України. Внаслідок цього, багато будинків та споруд стали зазнавати постійних перебоїв з електропостачанням, які можуть тривати до 10-15 годин на день. Також це вплинуло на енергетичну проблему і в Європі, так як, навіть за відсутності ракетних ударів, Європа досить сильно залежить від російського експорту електроенергії. Тому розвинені країни змушені шукати інші методи забезпечення енергетичної безпеки.

Одним із способів покращення рівня енергетичної безпеки країни, є аналіз більш розвинених країн, їх переваг та недоліків. Таким чином, кластерний аналіз допоможе класифікувати країни за їх рівнем енергетичної безпеки, та визначити, за якими ключовими факторами країни з високим рівнем енергетичної безпеки відрізняються від країн з низьким рівнем енергетичної безпеки. В цьому нам допоможуть методи та моделі кластерного аналізу та мова програмування R.

Мета дослідження – розробка інтерактивного застосунку оцінки рівня енергетичної безпеки більшості країн світу з використанням методів

кластерного аналізу, а саме ієрархічної кластеризації та кластеризації методом k-середніх у мові програмування R.

Для досягнення поставленої мети в роботі поставлені та вирішені наступні завдання:

1. розглянуто сутність та поняття енергетичної безпеки;
2. розглянуто основні фактори впливу на рівень енергетичної безпеки
3. переглянуто інформацію про кластерний аналіз та способи його реалізації;
4. розглянуто та проаналізовано інформацію щодо потрібних даних для проведення ієрархічного кластерного аналізу та кластеризації методом k-середніх;
5. розроблена моделі класифікації країн за рівнем енергетичної безпеки у мові програмування R;
6. створено інтерактивний застосунок оцінки рівня енергетичної безпеки з використанням результатів проведеного кластерного аналізу.

Об'єкт дослідження – процеси управління енергетичною безпекою.

Предмет дослідження – інтерактивний застосунок оцінювання рівня енергетичної безпеки.

Результати дослідження можуть бути використані у практичній діяльності підприємств енергетичного сектору та у подальших наукових дослідженнях.

## РОЗДІЛ 1. ТЕОРЕТИЧНА СУТНІСТЬ ЕНЕРГЕТИЧНОЇ БЕЗПЕКИ КРАЇНИ

### 1.1 Поняття енергетичної безпеки

Енергетична безпека — стан електроенергетики, який гарантує технічно та економічно безпечне задоволення поточних і перспективних потреб споживачів в енергії та охорону навколишнього природного середовища; енергогенеруючі компанії — учасники оптового ринку електричної енергії України, які володіють чи користуються генеруючими потужностями, виробляють і продають електричну енергію.

Енергетична безпека - це вимога до органів державної влади задовольняти гарантовані мінімальні потреби країни, окремих районів, міст, населених пунктів або об'єктів паливно-енергетичного комплексу, а також забезпечувати споживачів електроенергією в достатній кількості та належній якості як у звичайних умовах, так і в надзвичайних ситуаціях. Конкретним виміром забезпечення енергетичної безпеки, прийнятим за стандарт у західних країнах, є створення тримісячного стратегічного резерву нафти та нафтопродуктів.

Кожен автор наукової публікації, присвяченої дослідженню рівня енергетичної безпеки країни [1,2,3,4,5,6,7], має своє розуміння того, що він вкладає в поняття "енергетична безпека" (див. табл. 1.1). Однак усі науковці сходяться на думці, що енергетична безпека - це динамічне явище, яке має як довгострокову, так і короткострокову складові. Безпека для довкілля та надійне енергопостачання є довгостроковими міркуваннями. З іншого боку, орієнтація енергетичної системи на швидке реагування на непередбачувані загрози або коригування загального енергетичного балансу є однією з короткострокових характеристик.

Таблиця 1.1

## Трактування терміну енергобезпеки

Автор	Визначення терміну
Шидловський А.К., Кавалко М.П.	Енергетична безпека як одна із найважливіших складових економічної безпеки проявляється, по-перше, як стан забезпечення держави паливно-енергетичними ресурсами, що гарантують її повноцінну життєдіяльність і, по-друге, як стан безпеки енергетичного комплексу та здатність енергетики забезпечити нормальне функціонування економіки, енергетичну незалежність країни. Політична і енергетична незалежність є взаємообумовленими [1]
Сайт Світової енергетичної ради	Енергетична безпека – це впевненість у тому, що енергія буде в наявності і в тій кількості та якості, які вимагаються за даними економічних умов [2]
Денчев К.	Економічна безпека для країн-імпортерів – це передусім забезпечення надійності їх енергопостачання, диверсифікація джерел постачання енергоресурсів, забезпечення безпеки енергетичної інфраструктури, впровадження нових технологій для зниження залежності від імпорту енергоресурсів. Для країн-експортерів – це закріплення на стратегічних ринках за економічно вигідними цінами, забезпечення капіталу та фінансування інвестицій в інфраструктуру та розробку ресурсів [3]
Морозов В.В.	Енергетична безпека – це надійне і безперебійне забезпечення споживачів електричною і паливною енергією [4]



Закінчення табл. 1.1

Автор	Визначення терміну
Микитенко В.В.	Енергетична безпека – система поєднання потенціалів – економічного, політичного, техніко-технологічного, ресурсного і, власне, енергетичного, а також факторів наукового, географічного, організаційного, управлінського тощо, без урахування яких аналіз будь-якої безпеки неможливий [5]
Земляний М. Г.	Енергетична безпека країни – стан її захищеності від загроз енергетичного характеру [6]
Сухін Є.І.	Енергетична безпека - спроможність держави забезпечити максимально надійне, технічно безпечне, екологічно прийнятне та обґрунтовано достатнє енергозабезпечення економіки й населення, а також гарантоване забезпечення можливостей керівництва держави у формуванні і здійсненні політики захисту національних інтересів у сфері енергетики без надмірного зовнішнього та внутрішнього тиску в сучасних та прогнозованих умовах [7]
Енергетична Стратегія України на період до 2030 р.	Енергетична безпека є невід’ємною складовою економічної і національної безпеки, необхідною умовою існування і розвитку держави. У сучасному розумінні гарантування енергетичної безпеки – це досягнення стану технічно надійного, стабільного, економічно ефективного та екологічно прийнятного забезпечення енергетичними ресурсами економіки і соціальної сфери країни, а також створення умов для формування і реалізації політики захисту національних інтересів у сфері енергетики [8]

Наведені вище визначення поняття "енергетична безпека" ґрунтуються на декількох точках зору, що ускладнює дослідження цієї сфери економічної

науки. Враховуючи, що Україна є країною-імпортером енергоресурсів, доцільно визначити енергетичну безпеку як стан захищеності держави або регіону в енергетичній сфері від наявних та потенційних внутрішніх і зовнішніх загроз шляхом забезпечення безпеки енергетичної інфраструктури, впровадження нових технологій для зменшення залежності від імпорту енергоресурсів та розвитку нетрадиційної енергетики.

## 1.2 Фактори впливу на рівень енергетичної безпеки

На сучасному етапі свого економічного зростання Україна, як і багато інших країн, залежить від енергоносіїв. Енергетичній безпеці країни перешкоджає значна залежність від імпорту природного газу, нафти та ядерного палива від одного постачальника, хоча він може задовольнити близько 50% потреб країни в енергоресурсах з власного видобутку. Відсутність диверсифікації джерел постачання та неефективне використання імпортованих паливно-енергетичних ресурсів є двома з цих проблем. У світлі цього можна сказати, що фундаментальним аспектом, який впливає на рівень енергетичної безпеки країни, є здатність країни виробляти власні паливно-енергетичні ресурси та задовольняти всі свої потреби.

Виснаження природних ресурсів є однією зі змінних, що впливають на енергетичну безпеку. Це означає, що цінність ресурсу залежить від того, наскільки легко його можна знайти в природі. Виснаження природних ресурсів є значною екологічною проблемою. Крім того, складним елементом, який може вплинути на енергетичну безпеку, є стан справ у воєнний час. Внаслідок авіаударів можна повністю виключити можливість використання інших ресурсів для швидкого поповнення запасів. Через дефіцит і відсутність замінників палива зростає необхідність використання вугілля, нафти та інших видів викопного палива, які часто імпортуються з інших країн. Дуже важливо розглядати це питання з точки зору енергоефективності національної

економіки, щоб зменшити вплив вищезгаданого аспекту. Цей показник відображає сучасність, розвиток та конкурентоспроможність економічної системи. Рівень енергетичної безпеки економічної системи зростає, коли на одиницю продукції використовується менше енергетичних одиниць, що потребує менше енергетичних ресурсів. Крім того, важливо враховувати знос основних фондів у тих секторах економіки, які безпосередньо впливають на видобуток, переробку та транспортування необхідних енергоресурсів до кінцевого споживача. Показники енергоефективності знижуються, а виснаження природних ресурсів зростає, якщо ці підприємства мають високий рівень зношеності.

Основні принципи енергетичної безпеки країни можна розділити на декілька аспектів:

- укладення договорів про постачання енергоносіїв між державою, постачальниками та кінцевими споживачами в умовах конкуренції;
- політика, спрямована на запобігання монополізації ринку енергоресурсів, диверсифікацію джерел постачання та забезпечення дотримання норм експлуатації енергетичної інфраструктури;
- ефективне використання енергії, зменшення енергоємності національної економіки, оновлення основних виробничих та інфраструктурних фондів, впровадження енергозберігаючих технологій;
- розвиток альтернативних джерел енергії, таких як зелена енергетика та передові технології, що не вимагають спалювання традиційних енергоресурсів для виробництва електроенергії.
- Модернізація та розвиток транспортної інфраструктури, що відповідає сучасним вимогам та стандартам безпеки, зокрема, транспортування енергоносіїв;
- Розробка та впровадження програм енергетичної ефективності для промислових та комунальних споживачів;
- Підвищення енергетичної свідомості та екологічної культури населення, зокрема, за допомогою впровадження енергоефективних технологій в

побуті.

Врахування та реалізація цих аспектів допоможе забезпечити стійку та надійну енергетичну безпеку країни, зменшити залежність від імпорту енергоресурсів та сприятиме сталому економічному розвитку.

На рівень енергетичної безпеки економічної системи значною мірою впливають наступні фактори:

1. частка власних джерел енергії в загальному обсязі паливно-енергетичних ресурсів країни;
2. частка країни з найбільшим обсягом імпорту в загальному обсязі імпорту нафтопродуктів;
3. ступінь імпортої незалежності від ключових ресурсів енергетичного балансу;
4. Інвестиції в основний капітал у галузі постачання електроенергії, газу, пари та кондиційованого повітря;
5. ВВП на душу населення;
6. рівень зносу основних фондів у секторах постачання електроенергії, газу, пари та кондиційованого повітря;
7. частка відновлюваних джерел енергії в первинному енергобалансі;
8. енергоємність економіки по відношенню до ВВП.

Для класифікації факторів, що впливають на регіональну енергетичну безпеку, можна виділити дві категорії:

- аспекти, які знаходяться в зоні управління, а отже, в межах досяжності управління в теперішній і стратегічній епохах (такі як дефіцит енергії, індуковані аварії і фінансові питання).
- елементи, які контролюються за допомогою принципів розвитку і тривалість яких перевищує тривалість стратегічних планів (тобто, період гіпотези); приклади включають ресурсні обмеження, екологічні проблеми тощо.

Розробка системи управління енергетичною безпекою має на меті врахувати можливість впливу на другу групу елементів на додаток до

безпосереднього управління першою групою факторів. Гарантією такого впливу може стати науково-технічний прогрес, зокрема, створення нових джерел енергії та технологій її перетворення. Особливої уваги заслуговує пошук різноманітних підходів до впливу на другу категорію елементів, які контролюються в ході реалізації стратегічних програм.

### 1.3 Фактори оцінки рівня енергетичної безпеки країни

Мною було обрано 3 критерія для кластеризації країн за їх рівнем енергобезпеки, а саме: імпорт електроенергії, енергоємність та виснаження енергетичних запасів. Для того, щоб задовольнити попит на електроенергію в певній країні або регіоні, електроенергію необхідно купувати в іншій країні або регіоні - цей процес називається імпортом електроенергії. Коли країна або регіон не може виробляти достатньо електроенергії для внутрішнього споживання або потребує більше електроенергії для задоволення своїх енергетичних потреб, виникає така ситуація. Для того, щоб отримати необхідну кількість електроенергії, необхідно укласти угоди або контракти з постачальниками електроенергії в інших країнах або регіонах. Ці угоди можуть бути двосторонніми або досягнуті через регіональні міждержавні з'єднання, ринки обміну електроенергією, або і те, і інше. Електрична мережа країни-імпортера згодом з'єднується з імпортованою електроенергією за допомогою транскордонних ліній електропередач або інтерконекторів.

Існує кілька причин, чому країні або регіону може знадобитися імпорт електроенергії:

- Недостатнє внутрішнє виробництво: Якщо внутрішніх потужностей з виробництва електроенергії в країні або регіоні недостатньо для задоволення попиту на електроенергію, імпорт електроенергії може допомогти подолати розрив між попитом і пропозицією.
- Сезонні коливання: Країнам із сезонними коливаннями попиту на

енергію, наприклад, вищим споживанням електроенергії в спекотне літо або холодну зиму, може знадобитися імпорт електроенергії в пікові періоди, щоб забезпечити стабільне енергопостачання.

- Енергетична диверсифікація: Імпорт електроенергії може бути частиною стратегії диверсифікації джерел енергії та зменшення залежності від одного виду палива або технології виробництва електроенергії. Це забезпечує доступ до електроенергії, виробленої з різних джерел енергії, таких як гідроелектростанції, атомні електростанції або відновлювані джерела енергії, які можуть бути недоступні на місцевому рівні.
- Надзвичайні ситуації або відключення електроенергії: Під час надзвичайних ситуацій, таких як аварії на електростанціях, стихійні лиха або перебої в роботі ліній електропередач, імпорт електроенергії може допомогти підтримати надійність електромережі та запобігти масовим відключенням електроенергії.
- Економічні міркування: Імпорт електроенергії іноді може бути більш економічно ефективним, ніж інвестиції в нову вітчизняну інфраструктуру виробництва електроенергії. Це дозволяє країнам або регіонам отримувати доступ до електроенергії з сусідніх країн або регіонів з надлишковими генеруючими потужностями за конкурентними цінами.

Важливо зазначити, що імпорт електроенергії - це складний процес, що включає технічні, комерційні та регуляторні аспекти. Для забезпечення безперешкодного перетоку електроенергії через кордони необхідні системи передачі, з'єднувальні лінії та нормативно-правова база. Імпортована електроенергія також повинна відповідати стандартам якості та проходити необхідне перетворення напруги або синхронізацію для узгодження з вимогами енергосистеми країни-імпортера.

Енергоємність - це показник, який кількісно вимірює кількість енергії, що використовується на одиницю економічного продукту або діяльності. Це

коефіцієнт, який виражає споживання енергії відносно певного показника, такого як валовий внутрішній продукт (ВВП), промислове виробництво або чисельність населення. Енергоємність дає уявлення про ефективність використання енергії в економіці і може допомогти оцінити прогрес у сфері енергоефективності з плином часу або порівняти споживання енергії в різних секторах чи країнах.

Енергоємність розраховується за наступною формулою: Енергоємність = Споживання енергії / Економічний випуск, де:

- Споживання енергії: Загальна кількість енергії, спожитої протягом певного періоду, зазвичай вимірюється в таких одиницях, як кіловат-години (кВт-год), британські теплові одиниці (BTU) або джоулі (Дж).
- Економічний випуск: Показник, що використовується для вимірювання економічної діяльності, наприклад, ВВП, промислової доданої вартості або інших відповідних економічних показників.

Енергоємність часто виражається в одиницях енергії на одиницю економічного випуску, наприклад, кВт-год на долар ВВП або БТЕ на одиницю промислового виробництва. Це нормалізований показник, який дозволяє проводити порівняння між різними періодами часу, регіонами або секторами.

Ключові аспекти енергоємності включають

- Ефективність використання енергії: Енергоємність відображає ефективність використання енергії в економіці. Нижча енергоємність свідчить про більш ефективне використання енергоресурсів для досягнення економічних результатів, що означає, що на одиницю економічної діяльності потрібно менше енергії.
- Технологічний прогрес: Підвищення енергоефективності та впровадження чистіших і ефективніших технологій може призвести до зниження енергоємності.
- Галузеві відмінності: Різні сектори економіки мають різний рівень

енергоємності. Наприклад, промислові сектори, такі як переробна або важка промисловість, як правило, мають вищу енергоємність порівняно зі сферою послуг.

- Наслідки для політики: Він допомагає визначити сектори або регіони з високим рівнем енергоспоживання та сприяє формулюванню політики, цілей і програм з енергоефективності, спрямованих на зниження енергоємності та просування практики сталого енергетичного розвитку.

Зниження енергоємності є спільною метою для країн та організацій, які прагнуть підвищити енергоефективність та пом'якшити вплив на навколишнє середовище. Стратегії зниження енергоємності можуть включати заходи з енергозбереження, енергоефективні технології, впровадження відновлюваних джерел енергії, політичні стимули та інформаційні кампанії з просування енергоефективних практик.

Виснаження енергетичних запасів означає поступове скорочення або вичерпання наявних енергетичних ресурсів з плином часу. Це відбувається, коли видобуток або споживання енергоресурсів перевищує швидкість, з якою вони можуть бути поповнені природним шляхом або за рахунок нових відкриттів. Виснаження є серйозною проблемою для невідновлюваних джерел енергії, таких як викопне паливо (вугілля, нафта і природний газ), оскільки вони мають обмежену кількість і не можуть бути поновлені в межах людських часових рамок.

Нижче представлені ключові аспекти вичерпання енергетичних запасів:

1. Обмеженість природних ресурсів: Невідновлювані енергетичні ресурси формуються протягом мільйонів років в результаті геологічних процесів. Їх видобуток і споживання відбуваються набагато швидше, ніж їх утворення, що призводить до їх остаточного вичерпання.
2. Зниження темпів видобутку: У міру вичерпання запасів енергоресурсів їх видобуток стає все складнішим і дорожчим. Темпи



видобутку енергоресурсів з часом мають тенденцію до зниження, що вимагає більш масштабних зусиль з розвідки та видобутку для доступу до запасів, що залишилися.

3. Динаміка покладів: Видобуток викопного палива зазвичай починається з легкодоступних і високоякісних запасів. У міру вичерпання цих запасів, ресурси, що залишилися, стають важчими для видобутку, мають нижчу якість або розташовані у більш віддалених або складних умовах (наприклад, глибоководне буріння нафти, видобуток нетрадиційного газу).
4. Екологічні та економічні наслідки: Виснаження енергетичних запасів може мати значні екологічні та економічні наслідки. Оскільки легкодоступні запаси зменшуються, можуть застосовуватися більш шкідливі для довкілля методи видобутку, що збільшує ризик руйнування середовища існування, забруднення води та викидів парникових газів.
5. Перехід на альтернативні джерела: Виснаження запасів енергоресурсів зумовлює необхідність переходу на альтернативні джерела енергії, в тому числі відновлювані джерела енергії. Диверсифікуючи енергетичний баланс і зменшуючи залежність від обмежених ресурсів, суспільство може забезпечити більш стійке і безпечне енергетичне майбутнє.

Розуміння та вирішення проблеми вичерпання енергетичних запасів має важливе значення для забезпечення енергетичної безпеки, зменшення впливу на навколишнє середовище та створення умов для сталого енергетичного переходу. Впроваджуючи диверсифікований і сталий енергетичний портфель, суспільство може мінімізувати свою залежність від обмежених ресурсів і прокласти шлях до більш стійкого та екологічно безпечного енергетичного майбутнього.

## РОЗДІЛ 2. ІНТЕРАКТИВНІ МЕТОДИ ТА МОДЕЛІ КЛАСТЕРИЗАЦІЇ В СЕРЕДОВИЩІ R

### 2.1 Методи кластерного аналізу в середовищі R

Росс Іхака та Роберт Джентльмен з Оклендського університету в Новій Зеландії розробили інтерпретовану мову програмування R. Наразі R розробляється командою Core R Team. Крім того, вона забезпечує програмне середовище для моделювання даних, звітування та аналізу графічних і статистичних даних. Семантика лексичного скорингу об'єднана з мовою програмування S, яка реалізована в R. Розгалуження, зациклення та модульне програмування з використанням функцій - все це можливо в R. R підтримує інтеграцію з процесами, створеними на C, C++, .Net, Python та FORTRAN для підвищення ефективності. Для отримання, очищення, аналізу, візуалізації та представлення даних у сучасному світі R є одним з найважливіших інструментів, що використовуються дослідниками, аналітиками даних, статистиками та маркетологами.

Коли ми розглядаємо, як R використовується в різних галузях (Рис. 2.1), то на першому місці стоїть академічна сфера. Для ведення статистики використовують мову R. У секторі охорони здоров'я - це три найпоширеніші варіанти: наука, уряд та консалтинг [40].

### Visits to R by industry

Based on visits to Stack Overflow questions from the US/UK in January-August 2017.  
The denominator in each is the total traffic from that industry.

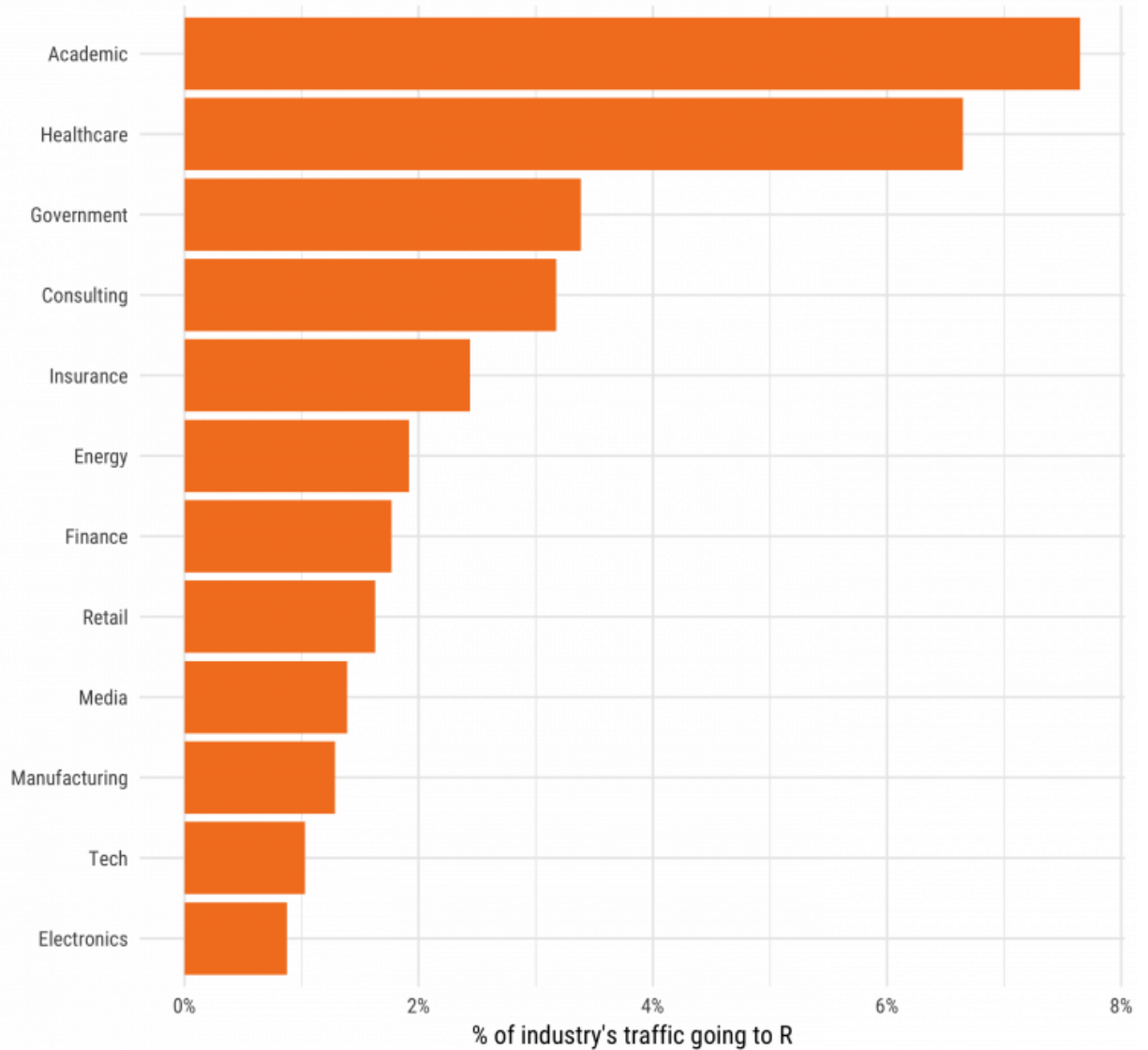


Рис. 2.1 - Використання R за галузями

R використовується і завжди буде використовуватися здебільшого для статистичних, візуальних цілей та машинного навчання. Який з пакетів R згенерував найбільше запитів на переповнення стеку, показано на зображенні нижче (Рис. 2.2). Більшість з топ-10 пов'язані з робочим процесом data scientist'a: підготовкою даних та передачею результатів [40].

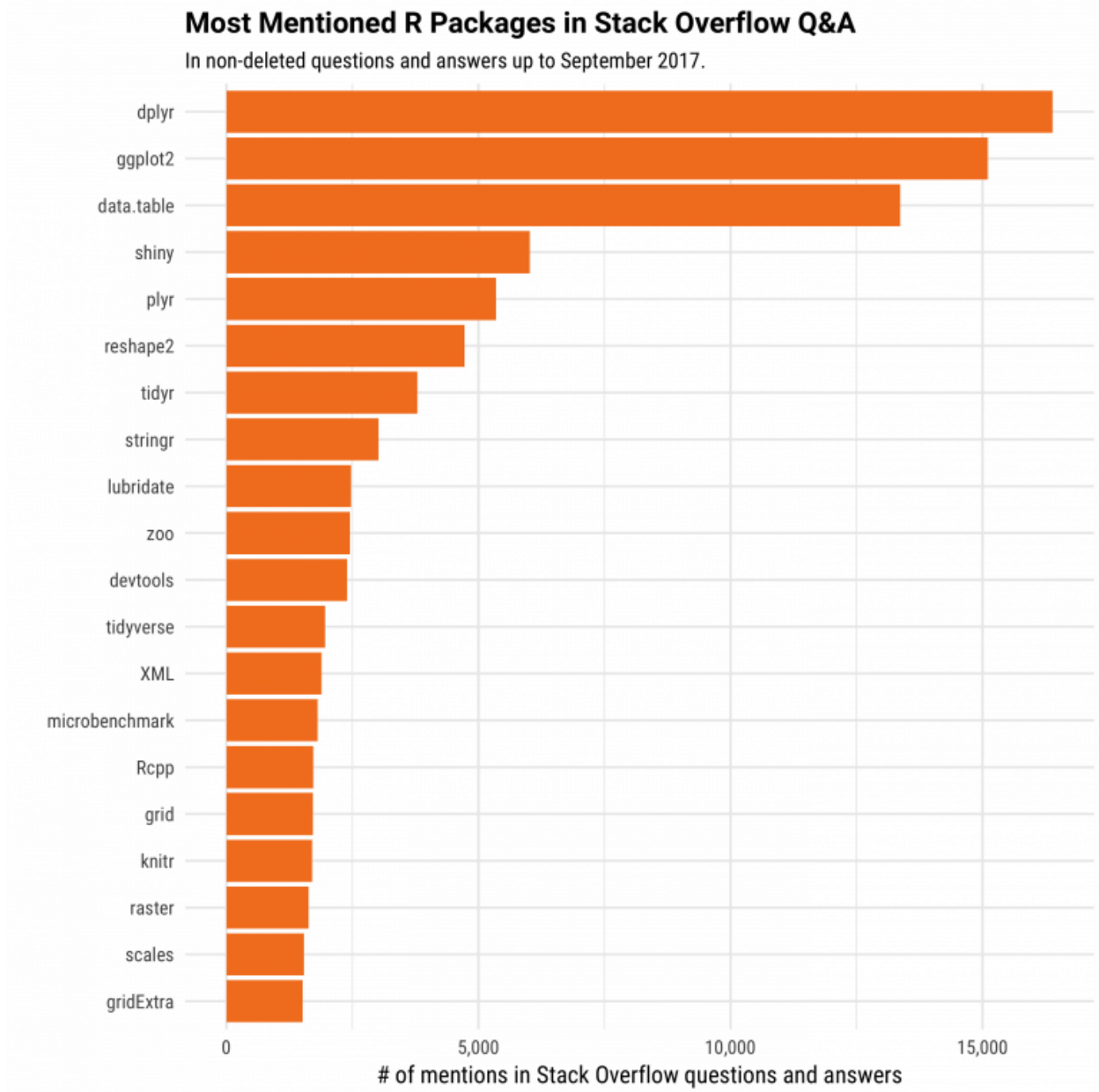


Рис 2.2 - Пакети R з найбільшою кількістю запитів

Ключовими аспектами мови програмування R можна виділити:

1. Обчислювальна статистика: R пропонує повний спектр можливостей для статистики та аналізу даних. Для обробки даних, описової статистики, перевірки гіпотез, регресійного аналізу, аналізу часових рядів, кластеризації, машинного навчання та інших завдань вона має широкий спектр вбудованих функцій і пакетів. Завдяки цим статистичним функціям R є ефективним інструментом для дослідницького аналізу даних і моделювання.

2. З відкритим вихідним кодом: R - це мова, яка доступна для використання, модифікації та розповсюдження безкоштовно. Завдяки своїй відкритій природі, R привернула увагу значної та активної спільноти користувачів та розробників, які постійно підтримують її розширення та вдосконалення. Крім того, вона надає користувачам доступ до значної колекції пакунків, створених користувачами, які розширюють можливості R.
3. Структури даних: R надає різноманітні структури даних, включаючи вектори, матриці, масиви, фрейми даних та списки, які підходять для статистичних досліджень. Ці структури даних забезпечують такі операції, як індексування, підмножини, злиття та перетворення, що дозволяє користувачам ефективно організовувати та змінювати дані.
4. Графіка та візуалізація: R пропонує потужні інструменти для створення красивої графіки та візуалізації. Функції для створення різноманітних графіків, зокрема діаграм розсіювання, гістограм, лінійних графіків, гістограм тощо, включено до основного пакету R. Існують також спеціалізовані інструменти, такі як ggplot2, які пропонують складні та адаптивні графічні опції.
5. Відтворювані дослідження: R дозволяє користувачам документувати і ділитися своїми методами аналізу даних і результатами, підтримуючи концепцію відтворюваних досліджень. Дослідники можуть поєднувати код, текст і графіку в одному документі, використовуючи розмітку R або грамотні підходи до програмування, що сприяє прозорості, співпраці та відтворюваності.
6. Інтеграція та інтероперабельність: Завдяки інтеграції з іншими мовами програмування, такими як Python, Java та C/C++, користувачі можуть скористатися перевагами кількох різних мов у своїх робочих процесах для аналізу даних. R також пропонує сильну сумісність з базами даних, електронними таблицями та форматами файлів, що дозволяє без проблем імпортувати та експортувати дані.

7. Ресурси: Спільнота R відома своєю підтримкою та процвітанням екології. Користувачам доступні різноманітні матеріали, в тому числі інструкції з експлуатації, публікації, онлайн-дискусійні форуми та програмне забезпечення, створене користувачами. Основним сховищем програм на R є всеосяжна мережа архівів R (CRAN), яка пропонує великий вибір пакунків для різноманітних статистичних програм та програм аналізу даних.

Багатьом користувачам мови програмування R подобається те, що її можна безкоштовно завантажити, вона пропонує складні можливості аналізу даних і має активну спільноту користувачів в Інтернеті, до якої вони можуть звернутися за підтримкою.

Оскільки вона існує вже багато років і користується популярністю протягом усього часу свого існування, мова є досить зрілою. Користувачі можуть завантажити додаткові пакети, які розширюють базову функціональність мови. Ці пакети дозволяють користувачам візуалізувати дані, підключатися до зовнішніх баз даних, географічно картографувати дані та виконувати розширені статистичні функції. Існує також популярний користувацький інтерфейс під назвою RStudio, який спрощує кодування мовою R.

Мову R критикують за повільний аналіз великих масивів даних. Це пов'язано з тим, що мова використовує однопотокową обробку, а це означає, що базова версія з відкритим вихідним кодом може використовувати лише один процесор одночасно. Для порівняння, сучасна аналітика великих даних процвітає завдяки паралельній обробці даних, одночасно використовуючи десятки процесорів на кластері серверів для швидкої обробки великих обсягів даних. На додаток до обмежень однопотоковой обробки, середовище програмування R є додатком, що працює в пам'яті. Всі об'єкти даних зберігаються в оперативній пам'яті машини під час сеансу роботи. Це може обмежувати кількість даних, з якими R може працювати одночасно.

Статистичний метод аналізу даних - це кластерний аналіз. Залежно від

того, наскільки тісно вони пов'язані, він організовує або об'єднує об'єкти в групи. Подібно до факторного аналізу, кластерний аналіз має справу з матрицями даних, в яких змінні ще не були розділені на підмножини для критеріїв і предикторів. Метою кластерного аналізу є пошук порівнянних груп осіб, де "схожість" між кожною парою суб'єктів відноситься до загальної міри за всією сукупністю атрибутів.

Метод сегментації даних, який називається кластеризацією, поділяє дані на кілька груп відповідно до того, наскільки вони схожі. В основному, ми використовуємо статистичний процес для групування даних. Кластери - це більш компактні групи, які створюються з більших даних. Ці кластери мають наступні властивості:

- Вони виявляються під час виконання процедури, і їхня кількість не відома заздалегідь.
- Кластери - це колекції пов'язаних речей, які мають схожі риси.

Найпоширенішим і найулюбленішим методом аналізу даних та інтелектуального аналізу даних є кластеризація. Вона застосовується, коли потрібно знайти порівняні підмножини, які можна проаналізувати різними способами, через величезний обсяг вихідних даних.

Застосування R-кластеризації полягає в наступному:

- Маркетинг - у сфері маркетингу ми використовуємо кластеризацію для дослідження та відбору клієнтів, які є потенційними покупцями продукту. Це дозволяє відрізнити найпривабливіших клієнтів від тих, хто має найменшу схильність до купівлі продукту. Після створення кластерів підприємства можуть відстежувати своїх клієнтів і приймати необхідні рішення, щоб утримати їх у цьому кластері.
- Роздрібна торгівля - Роздрібна торгівля використовує кластеризацію для групування покупців на основі їхніх уподобань, стилю, вибору одягу, а також уподобань щодо магазинів. Це дозволяє їм управляти своїми магазинами набагато ефективніше.

- Медицина - медицина та охорона здоров'я використовують алгоритми кластеризації для полегшення ефективної діагностики та лікування своїх пацієнтів, а також для відкриття нових ліків. Враховуючи вік, групу, генетичне кодування пацієнтів, ці організації краще розуміють діагноз завдяки надійній кластеризації.
- Соціологія - Кластеризація використовується в операціях інтелектуального аналізу даних для поділу людей на основі їх демографічних характеристик, способу життя, соціально-економічного статусу тощо. Це може допомогти правоохоронним органам згрупувати потенційних злочинців і навіть ідентифікувати їх за допомогою ефективного застосування алгоритму кластеризації.

У різних галузях R-кластеризація має різні назви, наприклад:

- Маркетинг - У маркетингу для позначення кластеризації використовується термін "сегментація" або "типологічний аналіз".
- Медицина - кластеризація в медицині відома як нозологія.
- Біологія - в біології це називається числовою таксономією.

Для визначення правильних критеріїв кластеризації та використання ефективних алгоритмів, загальна формула виглядає наступним чином:

$$Vn(\text{кількість розбиття для } n \text{ об'єктів}) > \exp(n)$$

Складність кластеризації можна визначити за кількістю можливих комбінацій об'єктів. Від цього числа залежить складність кластера. Основою для об'єднання або розділення об'єктів є відстань між ними. Ці відстані є несхожістю (коли об'єкти знаходяться далеко один від одного) або схожістю (коли об'єкти знаходяться поруч).

Для обчислення відстані між об'єктами в K-середніх ми використовуємо наступні типи методів:

- Евклідова відстань - це найпоширеніший метод вимірювання відстані між об'єктами, які присутні в багатовимірному просторі.



У загальному випадку, для  $n$ -вимірному простору відстань дорівнює

- Квадратна евклідова відстань - отримується шляхом піднесення до квадрату евклідової відстані. Об'єктам, які знаходяться на більшій відстані, присвоюється більша вага.
- Відстань між містом і кварталом (Манхеттен) - за допомогою цього методу обчислюється різниця між двома точками у всіх вимірах. У багатьох випадках вона подібна до Евклідової відстані, але має додаткову функцію зменшення ефекту в крайніх об'єктах, які не мають квадратичних координат.

Одним з найпопулярніших алгоритмів розбиття даних на групи є кластерний аналіз  $K$ -середніх в  $R$ . Це алгоритм некерованого навчання. Він намагається кластеризувати дані на основі їхньої схожості. Крім того, ми вказали кількість кластерів і хочемо, щоб дані були згруповані в однакові кластери. Алгоритм відносить кожне спостереження до кластера, а також знаходить центроїд кожного кластера.

Алгоритм  $K$ -середніх:

1. Вибирає  $K$  центроїдів ( $K$  рядків, вибраних випадковим чином).
2. Потім ми повинні віднести кожен точку даних до найближчого до неї центроїда.
3. Крім того, він перераховує центроїди як середнє значення всіх точок даних у кластері.
4. Присвоює точкам даних їхні найближчі центроїди.
5. Крім того, ми повинні продовжувати кроки 3 і 4, поки спостереження не будуть перепризначені.

$R$  містить приголомшливий набір функцій кластерного аналізу. Далі буде розглянуто три з них - ієрархічний агломеративний, метод розбиття та метод на основі моделей. Хоча не існує найкращих рішень для проблеми визначення кількості кластерів, які потрібно виділити, нижче наведено кілька підходів.

Перед кластеризацією даних ми можемо видалити або оцінити відсутні дані, а також змінити масштаб змінних для їхньої порівнянності, за допомогою наступних функцій:

```
# Prepare Data
mydata <- na.omit(mydata)
```

У цьому фрагменті коду `mydata` - це фрейм даних або матриця, що містить набір даних, з яким ми працюємо. Функція `na.omit()` використовується для видалення з `mydata` всіх рядків з пропущеними значеннями (NA). Це називається "списковим видаленням" або повним аналізом, коли будь-який рядок, що містить хоча б одне пропущене значення, повністю видаляється з набору даних.

```
mydata <- scale(mydata)
```

До `mydata` застосовується функція `scale()`, яка стандартизує змінні в наборі даних. Стандартизація, також відома як нормалізація за z-критерієм, перетворює змінні так, щоб їхнє середнє значення дорівнювало 0, а стандартне відхилення - 1. Цей процес відбувається шляхом віднімання середнього значення кожної змінної від її початкових значень, а потім ділення на стандартне відхилення. Цей процес майже завжди є обов'язковим для кластерного аналізу, адже данні для кластеризації зазвичай мають різний характер.

Кластеризація за методом К-середніх є найпопулярнішим методом розбиття. Він вимагає від аналітика вказати кількість кластерів, які потрібно виділити. Визначити відповідну кількість кластерів може допомогти графік суми квадратів всередині груп за кількістю виділених кластерів. Аналітик шукає вигин на графіку, подібно до скрі-тесту у факторному аналізі.

```
# Determine number of clusters
wss <- (nrow(mydata)-1)*sum(apply(mydata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(mydata,
  centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
```

```
ylab="Within groups sum of squares")
```

```
# K-Means Cluster Analysis
```

```
fit <- kmeans(mydata, 5) # 5 cluster solution
```

```
aggregate(mydata,by=list(fit$cluster),FUN=mean)
```

```
mydata <- data.frame(mydata, fit$cluster)
```

Наданий вище код виконує кластерний аналіз методом К-середніх на наборі даних (mydata) і складається з двох основних частин:

1) Визначення кількості кластерів:

- Перша частина коду обчислює внутрішню групову суму квадратів (wss) для різної кількості кластерів від 2 до 15.
- Внутрішню групову суму квадратів є мірою того, наскільки компактно розташовані точки даних у кожному кластері.
- Оцінюючи внутрішню групову суму квадратів для різної кількості кластерів, можна визначити відповідну кількість кластерів для використання в подальшому кластерному аналізі К-середніх.

2) Кластерний аналіз К-середніх:

- Друга частина коду виконує кластеризацію за методом К-середніх на наборі даних з використанням обраної кількості кластерів (у цьому випадку 5).
- Функція kmeans() застосовується до mydata з вказаною кількістю кластерів (centers = 5), в результаті чого створюються кластери і кожне спостереження відноситься до одного з кластерів на основі їхньої схожості.
- Результуючий об'єкт fit містить інформацію про рішення кластеризації, включаючи центри кластерів, призначення кластерів та інші пов'язані з ними атрибути.

Потім код продовжує витягувати додаткові відомості з результатів кластеризації:

- Середні значення кластерів: Функція `aggregate()` використовується для обчислення середнього значення для кожної змінної в `mydata` в межах кожного кластера. Це дає інформацію про середні значення змінних у кожному кластері, що дозволяє охарактеризувати та інтерпретувати кожен кластер.
- Приналежність до кластера: Столпчик кластерного розподілу з `fit` додається до `mydata` за допомогою `data.frame()`. Це додає інформацію про розподіл кластерів до вихідного набору даних, що полегшує подальший аналіз, візуалізацію або порівняння з іншими змінними.

Отже, цей код допомагає визначити оптимальну кількість кластерів, оцінюючи суму квадратів всередині груп для різних кластерних рішень. Потім він виконує кластеризацію за методом К-середніх з обраною кількістю кластерів, витягує середні значення кластерів і додає розподіл кластерів до набору даних. Мета полягає в тому, щоб згрупувати подібні спостереження разом, зрозуміти характеристики кожного кластера і отримати уявлення про дані на основі результатів кластеризації.

Тепер можемо розглянути наступний метод, а саме ієрархічний агломеративний.

```
# Ward Hierarchical Clustering
d <- dist(mydata, method = "euclidean")
fit <- hclust(d, method="ward")
plot(fit)
groups <- cutree(fit, k=5) # cut tree into 5 clusters
rect.hclust(fit, k=5, border="red")

# Ward Hierarchical Clustering with Bootstrapped p values
library(pvclust)
fit <- pvclust(mydata, method.hclust="ward",
```

```

method.dist="euclidean")
plot(fit)
pvrect(fit, alpha=.95)

```

Наданий код виконує ієрархічну кластеризацію методом Уорда на наборі даних (`mydata`) і складається з двох основних частин:

1) Ієрархічна кластеризація методом Уорда:

- Спочатку програма обчислює матрицю відстаней (`d`), використовуючи евклідову міру відстані. Матриця відстаней представляє попарні відстані між спостереженнями в наборі даних.
- Потім до матриці відстаней (`d`) застосовується функція `hclust()` з використанням методу зв'язування Уорда (`method = "ward"`). Цей метод мінімізує дисперсію всередині кластерів для створення ієрархічних кластерів.
- Результат ієрархічної кластеризації зберігається в об'єкті `fit`.
- Функція `plot()` використовується для відображення дендрограми, яка візуалізує структуру ієрархічної кластеризації.

2) Ієрархічна кластеризація методом Уорда з підвантаженими  $p$ -значеннями:

- Код спочатку завантажує бібліотеку `pvclust`, яка надає додаткову функціональність для ієрархічної кластеризації.
- Потім використовується функція `pvclust()` для виконання ієрархічної кластеризації методом Уорда з завантаженими  $p$ -значеннями. Метод `hclust = "ward"` задає метод зв'язування Уорда, а метод `dist = "euclidean"` встановлює евклідову міру відстані.
- Результат кластеризації зберігається в об'єкті `fit`.
- Функція `plot()` використовується для відображення дендрограми з  $p$ -значеннями, пов'язаними з кластерами.
- Функція `pvrect()` використовується для додавання прямокутників

навколо груп, які добре підтримуються даними. Аргумент `alpha` задає рівень довіри для визначення значущих кластерів.

Загалом, код надає можливість виконати ієрархічну кластеризацію методом Уорда на наборі даних, візуалізувати структуру кластеризації за допомогою дендрограм та оцінити значущість отриманих кластерів за допомогою бутстрепованих  $p$ -значень. Цей підхід допомагає виявити ієрархічні зв'язки та значущі групи в даних, що дає змогу зрозуміти основну структуру та закономірності, присутні в наборі даних.

Третій розглянутий метод, став метод на основі моделей. Підходи на основі моделей припускають наявність різних моделей даних і застосовують оцінку максимальної правдоподібності та критерії Байєса для визначення найбільш вірогідної моделі та кількості кластерів.

```
# Model Based Clustering
library(mclust)
fit <- Mclust(mydata)
plot(fit) # plot results
summary(fit) # display the best model
```

Наведений код виконує кластеризацію на основі моделі на наборі даних (`mydata`) за допомогою пакету `Mclust` на мові R. Ось що робить кожен рядок коду:

- `library(mclust)`: Цей рядок завантажує пакет `mclust`, який надає функції для кластеризації на основі моделей.
- `fit <- Mclust(mydata)`: Функція `Mclust()` використовується для підгонки різних моделей до набору даних і вибору найкращої моделі на основі Байєсівського інформаційного критерію (BIC). Функція автоматично визначає оптимальну кількість кластерів і вибирає відповідну модель для даних. Результат кластеризації зберігається в об'єкті `fit`.
- `plot(fit)`: Цей рядок будує графік результатів кластеризації на основі моделі. Зазвичай він включає візуалізації, такі як діаграма розсіювання або інші відповідні діаграми для відображення кластерів та їхніх

характеристик.

- `summary(fit)`: Функція `summary()` використовується для відображення резюме найкращої моделі, вибраної функцією `Mclust()`. Вона надає інформацію про параметри моделі, лог-вірогідність, BIC та інші відповідні статистичні дані. Цей звіт допомагає інтерпретувати результати і зрозуміти характеристики знайдених кластерів.

Загалом, цей код виконує кластеризацію на основі моделей на наборі даних, автоматично вибираючи найкращу модель і кількість кластерів на основі критерію BIC. BIC розшифровується як Байєсівський інформаційний критерій. BIC - це критерій, який використовується для порівняння різних моделей і вибору оптимальної моделі серед них. Він базується на принципах байєсівської статистики і надає міру відповідності моделі, одночасно караючи моделі з більшою кількістю параметрів. BIC враховує два компоненти: ймовірність даних, що відповідають моделі, та кількість параметрів у моделі. Він має на меті досягти балансу між складністю моделі та її відповідністю даним. Чим нижче значення BIC, тим краще модель відповідає даним. Отримані розподіли моделей і кластерів можна візуалізувати за допомогою функції `plot()` і далі проаналізувати за допомогою функції `summary()`. Кластеризація на основі моделей забезпечує гнучкий і потужний підхід до виявлення кластерів і розуміння базової структури даних на основі статистичних моделей і критеріїв.

## 2.2 Інтерактивне середовище Shiny

Фреймворк веб-додатків для R під назвою Shiny дозволяє створювати інтерактивні онлайн-додатки. Замість використання статичних статей, додатки Shiny можна використовувати для передачі інформації через інтерактивні дослідження даних. Додаток Shiny складається з функції `server()`, яка надає інструкції для генерації об'єктів, представлених в інтерфейсі

користувача, та інтерфейсу користувача (UI), який керує дизайном і зовнішнім виглядом додатку.

До ключових особливостей та концепції, пов'язані з Shiny можна віднести:

1. Реактивне програмування: Shiny побудовано на концепції реактивного програмування. Реактивне програмування дозволяє створювати динамічні та інтерактивні додатки, визначаючи зв'язки між входами, виходами та реактивними виразами. За допомогою Shiny ви можете вказати реактивні залежності між різними частинами вашого додатку, і система автоматично оновить вихідні дані при зміні входів.
2. Інтерфейс користувача (UI): Компонент UI додатку Shiny визначає структуру та зовнішній вигляд веб-інтерфейсу. Він використовує комбінацію коду на R та HTML-подібного синтаксису для створення макета, елементів керування введенням (наприклад, повзунки, прапорці, випадаючі меню) та елементів виведення (наприклад, графіки, таблиці), з якими взаємодіють користувачі.
3. Логіка сервера: Серверний компонент Shiny-додатку - це місце, де визначаються обчислення, обробка даних та реактивна поведінка. Він містить R-код, який реагує на вхідні дані користувача, виконує обчислення, оновлює реактивні значення і генерує вихідні дані для відображення в інтерфейсі користувача. Серверний код також може взаємодіяти із зовнішніми джерелами даних, базами даних або іншими пакетами R.
4. Реактивні вирази: Реактивні вирази у Shiny - це функції, які автоматично оновлюються, коли змінюються їхні базові залежності. Вони використовуються для виконання обчислень або перетворень вхідних значень і повернення оновлених результатів. Реактивні вирази можна використовувати для створення динамічної та чутливої поведінки у вашому додатку.
5. Керування сеансами: Shiny керує сеансами користувачів, дозволяючи



декільком користувачам взаємодіяти з додатком одночасно. Він відстежує стан сеансу, обробляє дані, що вводяться користувачем, і відповідно оновлює інтерфейс і результати. Ви можете керувати поведінкою сеансу та обробляти дані сеансу за допомогою об'єкта сеансу Shiny.

6. Розгортання: Додатки Shiny можна розгортати локально або розміщувати на веб-серверах. Ви можете запускати Shiny-додаток локально у вашому середовищі R, що дозволить вам розробляти і тестувати додаток на вашому комп'ютері. Щоб зробити програму доступною для інших, ви можете розгорнути її на веб-сервері або хмарній платформі, зробивши її доступною через Інтернет.

На рисунку 2.3 зображено додаток Shiny в дії. Ця програма містить числове поле, куди можна ввести число, і гістограму, яка будується за значеннями, випадково згенерованими з нормального розподілу. Кількість значень, що використовуються для побудови гістограми, задається числом, вказаним у числовому полі. Кожного разу, коли ми вводимо інше число в числове поле, гістограма перебудовується з новими згенерованими значеннями.

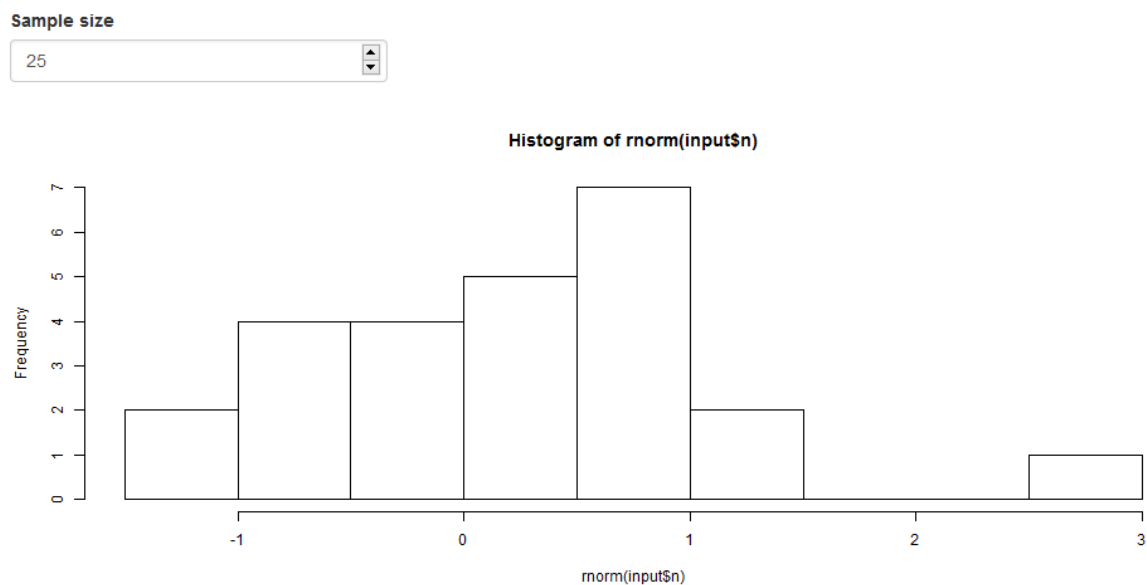


Рис. 2.3 – Проста числова програма в Shiny

На рисунку 2.4 показано другий додаток Shiny. Ця програма показує гістограму з кількістю телефонів у певному регіоні світу за кілька років. Регіон можна вибрати зі спадного меню, яке містить усі можливі регіони. Гістограма перебудовується щоразу, коли вибирається інший регіон.

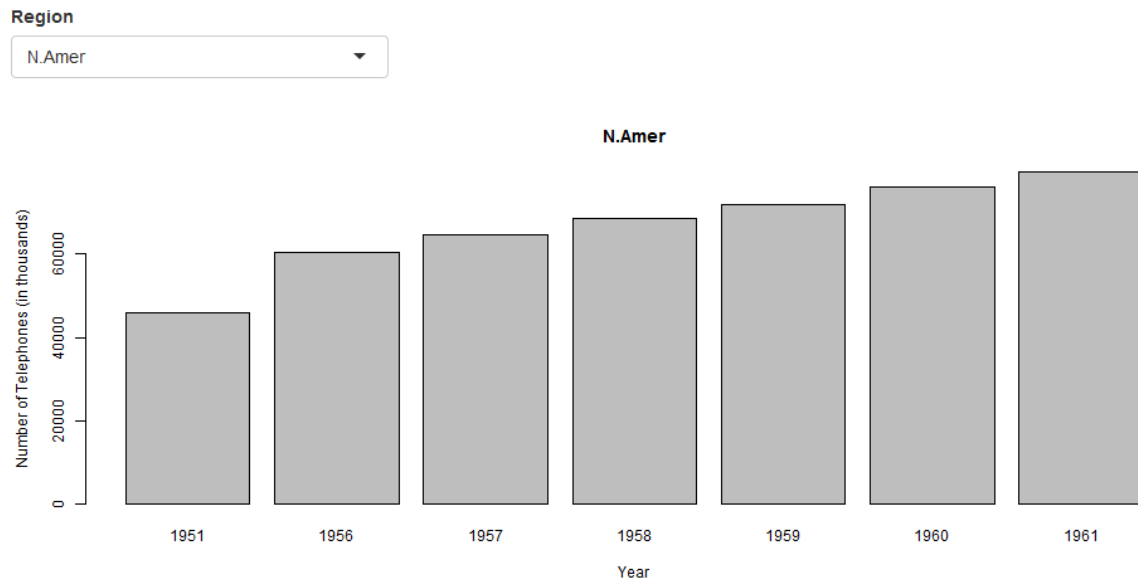


Рис. 2.4 – Проста числова програма зі спадним меню

Додаток Shiny можна створити, створивши каталог (наприклад, `appdir`), який містить R-файл (наприклад, `app.R`) з трьома компонентами:

- об'єкт інтерфейсу користувача (`ui`), який керує компонуванням та зовнішнім виглядом програми,
- функція `server()`, яка містить інструкції для побудови об'єктів, що відображаються у користувацькому інтерфейсі, та
- виклик функції `shinyApp()`, яка створює додаток з пари `ui/сервер`.

```
# load the shiny package
```

```
library(shiny)
```

```
# define user interface object
```

```
ui <- fluidPage()
```

```
# define server function
server <- function(input, output) { }

# call to shinyApp() which returns a Shiny app object from an
# explicit UI/server pair
shinyApp(ui = ui, server = server)
```

Додатки Shiny включають веб-елементи, які називаються входами, з якими користувачі можуть взаємодіяти, змінюючи їхні значення. Коли користувач змінює значення входу, інші елементи Shiny-додатку, які використовують значення входу, оновлюються. Додатки Shiny у попередніх прикладах містять кілька входів. Перший додаток показує числовий ввід для введення числа і гістограму, яка будується за цим числом (рис. 2.3). Другий додаток показує вхідні дані для вибору регіону та гістограму, яка змінюється, коли змінюється регіон (рис. 2.4).

Shiny-додатки можуть містити різноманітні вхідні дані, корисні для різних цілей, зокрема тексти, числа і дати (рис. 2.5). Користувач може змінювати ці дані, і при цьому оновлюються об'єкти в додатку, які їх використовують[28, 35].

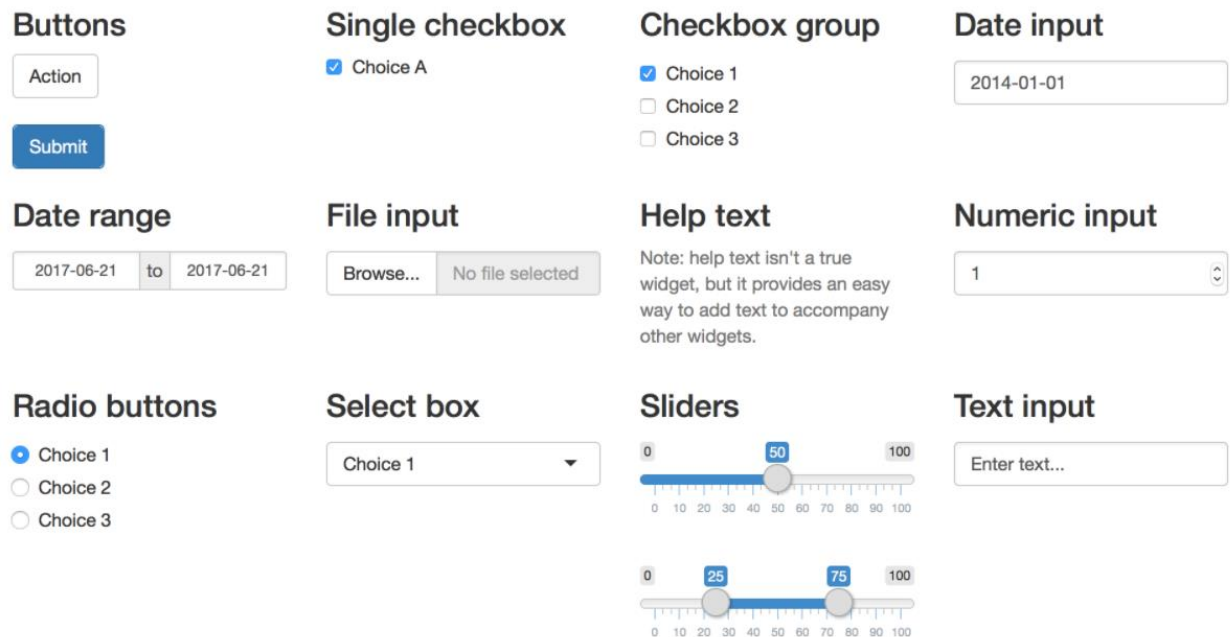


Рис. 2.5 - Приклади вхідних даних

Shiny-додатки можуть містити різноманітні елементи виводу, зокрема графіки, таблиці, тексти, зображення та HTML-віджети (рис. 2.6). HTML-віджети - це об'єкти для інтерактивної візуалізації веб-даних, створені за допомогою бібліотек JavaScript. Прикладами HTML-віджетів є інтерактивні веб-карти, створені за допомогою пакета leaflet, та інтерактивні таблиці, створені за допомогою пакета DT. HTML-віджети вбудовуються в Shiny за допомогою пакета htmlwidgets. Ми можемо використовувати значення вхідних даних для побудови вихідних елементів, і це призводить до того, що вихідні дані оновлюються щоразу, коли змінюються вхідні значення.

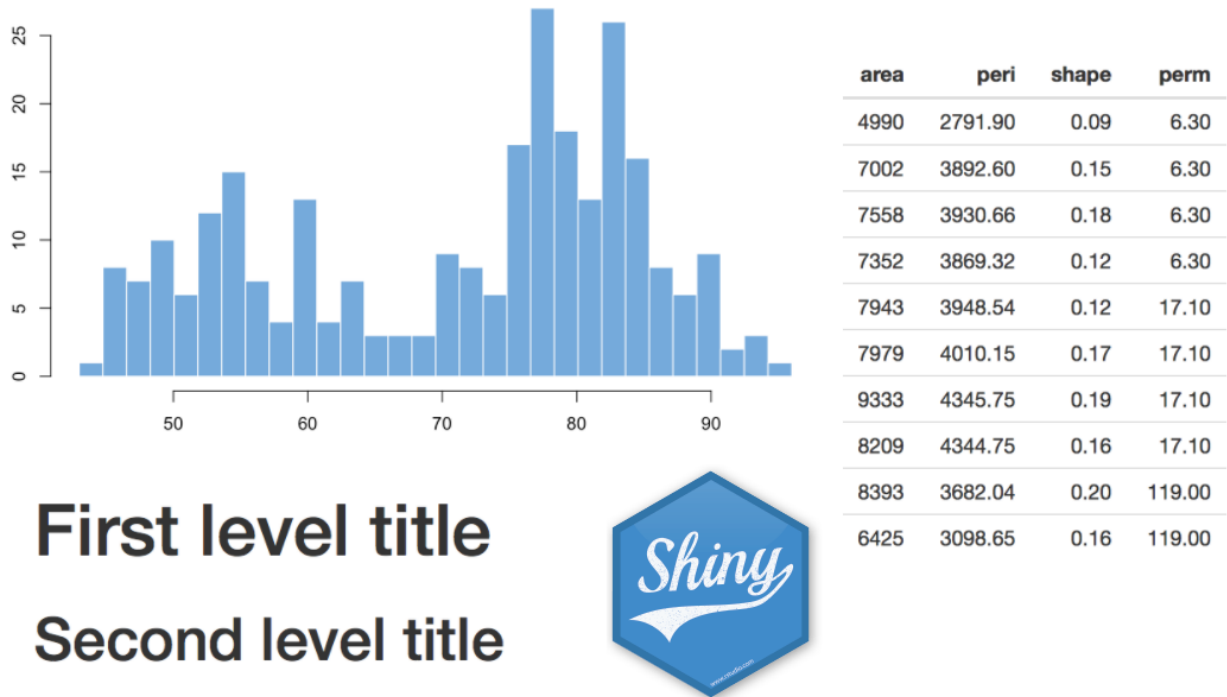


Рис. 2.6 - Приклади вихідних даних

Отже, Shiny - чудовий вибір для розробки інтерактивного додатку для кластеризації країн на основі їхнього рівня енергетичної безпеки. Він дозволяє користувачам досліджувати та візуалізувати результати кластеризації у зручний та інтерактивний спосіб, що сприяє кращому розумінню та аналізу моделей енергетичної безпеки. Однією з ключових переваг використання Shiny є його здатність подолати розрив між аналізом даних і кінцевими користувачами, надаючи інтерактивну платформу для дослідження даних. Користувачі можуть інтерактивно маніпулювати параметрами кластеризації, такими як кількість кластерів або метрики відстані, і миттєво візуалізувати вплив цих змін на результати кластеризації. Це сприяє глибшому розумінню основних закономірностей і взаємозв'язків у даних.

## РОЗДІЛ 3. ІНТЕРАКТИВНИЙ ЗАСТОСУНОК ОЦІНЮВАННЯ РІВНЯ ЕНЕРГЕТИЧНОЇ БЕЗПЕКИ

### 3.1 Моделі класифікації країн за рівнем енергетичної безпеки

Для проведення кластерного було вирішено обрати таку досліджувану сукупність даних енергетичної безпеки як імпорт енергії, енергоємність та виснаження природних запасів. Дані були обрані для 100 країн світу за даними сайту Світового банку. [Додаток А, 11].

Таблиця 3.1

Фактори впливу на енергобезпеку

Назва фактору
Імпорт енергії (%)
Енергоємність (MJ/\$)
Виснаження енергетичних запасів (% від ВНД)

Для початку проведемо кластеризацію країн за рівнем енергобезпеки в R Studio без використання пакету Shiny. Перш за все необхідно встановити та активувати усі потрібні нам пакети для виконання кластеризації .

```
install.packages(«ggplot2»)
install.packages(«factoextra»)
install.packages(«cluster»)
install.packages(«NbClust»)
install.packages(«clValid»)
install.packages(«psych»)
install.packages(«MASS»)
install.packages(«gplots»)
```

```
library(ggplot2)
library(factoextra)
library(cluster)
library(NbClust)
library(clValid)
library(psych)
library(MASS)
library(gplots)
```

Наступними кроками після підключення всіх пакетів є читка CSV-файлу, перетворення його у фрейм даних, та виконання стандартизації числових стовпців. Стандартизація допомагає забезпечити однаковий масштаб змінних, зменшує вплив викидів, покращує продуктивність алгоритму та покращує інтерпретацію результатів.

```
Countries<- read.csv(«C:/Users/ilato/OneDrive/Робочий
стол/countries.txt», header = T, sep = «\t»)
View(countries)
countries<- as.data.frame(countries)
countries_scale <- scale(countries[-1], center = T, scale =
T)#стандартизація даних
View(countries_scale)
```

Результати стандартизації даних показані на рисунку 3.1.

	Energy_intensity	Energy_depletion	Energy_import
1	-0.954003135	-0.262460742	0.219123048
2	-0.548347604	-0.146351626	0.213305438
3	-0.438710974	-0.480535914	0.654668460
4	0.043690197	-0.217530653	-1.338682043
5	-0.822439179	-0.467540453	0.587171775
6	-0.268774198	-0.480530128	0.691889433
7	0.367118255	-0.467324752	0.391339011
8	1.085238181	-0.332352042	0.771521647
9	-0.175583063	0.109984213	-1.232488324
10	1.529266532	-0.366827304	-0.399454823
11	-0.356483502	-0.478291777	0.578577267
12	1.140056496	-0.334117707	0.228386214
13	-1.107494416	0.418418016	-1.960227291
14	-0.016609949	-0.469162364	0.334110032
15	-0.789548190	-0.474400996	0.575455906
16	-1.189721889	-0.368371763	0.121834037
17	-1.233576541	-0.480535916	0.771047837
18	-0.663466066	0.314013578	-0.753569791
19	-0.394856322	0.622961101	0.058740017
20	-0.844366505	-0.480364656	0.639670377
21	0.674100819	-0.418068963	0.140399653
22	-0.499011121	-0.479065017	0.443826188
23	-1.041712439	-0.349014198	0.414983458
24	-0.268774198	-0.478557881	0.635070387
25	-0.762139032	0.434975849	0.052680202
26	-0.718284380	-0.471567076	0.583742142
27	-0.647020571	-0.391054869	0.461798176
28	-0.614129582	-0.004178919	-0.665611906
29	0.049172028	-0.355199407	0.374344687
30	-1.589895588	-0.456358134	0.752418282
31	-0.910148483	-0.465290201	0.622178033
32	-1.003339618	-0.464008232	0.682370989
33	-0.268774198	-0.479505693	0.847458370
34	-0.471601963	-0.479065672	0.826030249
35	0.860483089	-0.473354371	0.489689699

Showing 1 to 35 of 100 entries, 3 total columns

Рис. 3.1 – стандартизовані дані



Кластерний аналіз методом К-середніх передбачає спочатку визначення кількості кластерів. У випадку, якщо у вас немає встановленої кількості кластерів на які потрібно розділити об'єкти визначимо оптимальне їх число, скориставшись методом ліктя (рис. 3.2):

```
fviz_nbclust(countries_scale, kmeans, method = «wss») # метод ліктя
```

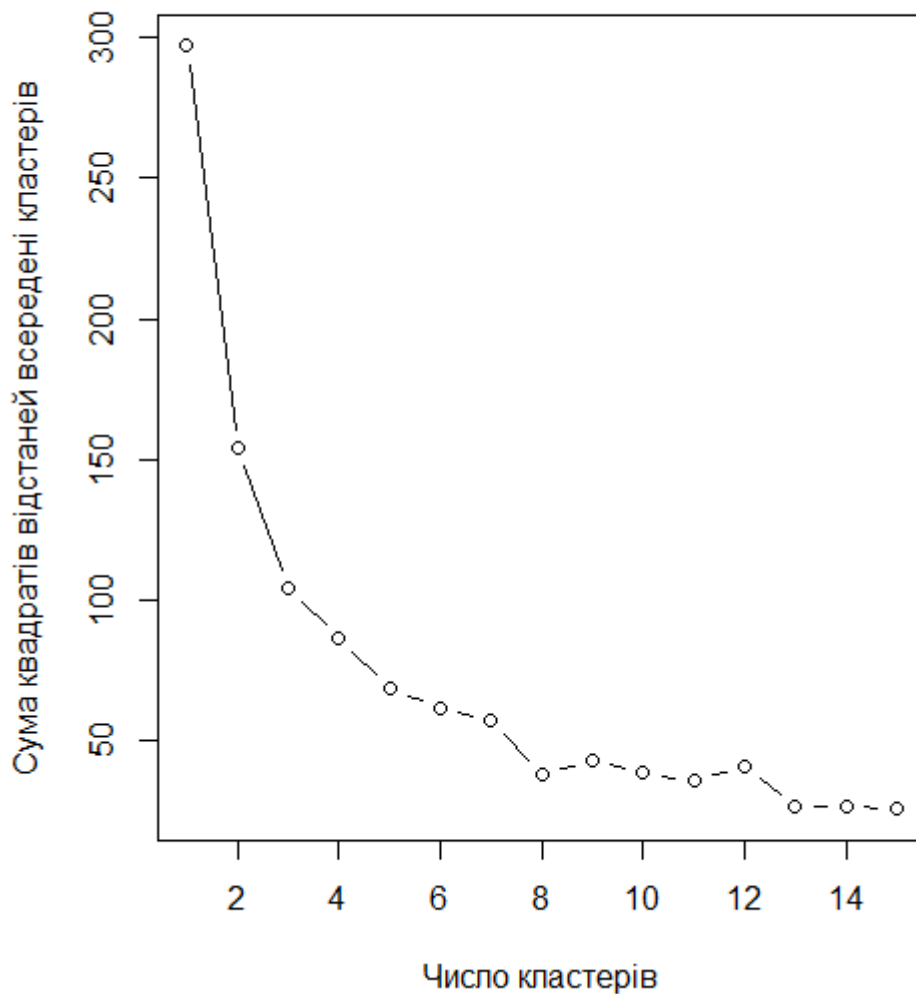


Рис. 3.2 – Визначення оптимальної кількості кластерів

Також, для наочності, скористаємось методом із пакету NbClust та

наступним фрагментом коду. Результати пошуку оптимальної кількості кластерів показані на рисунку 3.3:

```
> clust_optimal<- NbClust(countries_scale, distance = "euclidean", min.nc = 2, max.nc = 10, method = "complete", index = "all")# пошук оптимальної кількості кластерів
*** : The Hubert index is a graphical method of determining the number of clusters.
      In the plot of Hubert index, we seek a significant knee that corresponds to a
      significant increase of the value of the measure i.e the significant peak in Hubert
      index second differences plot.

*** : The D index is a graphical method of determining the number of clusters.
      In the plot of D index, we seek a significant knee (the significant peak in Dindex
      second differences plot) that corresponds to a significant increase of the value of
      the measure.

*****
* Among all indices:
* 7 proposed 2 as the best number of clusters
* 8 proposed 3 as the best number of clusters
* 1 proposed 5 as the best number of clusters
* 2 proposed 6 as the best number of clusters
* 1 proposed 7 as the best number of clusters
* 2 proposed 9 as the best number of clusters
* 3 proposed 10 as the best number of clusters

***** conclusion *****

* According to the majority rule, the best number of clusters is 3

*****
```

Рис. 3.3 – Пошук оптимальної кількості кластерів

Як міра відстані використовується евклідова відстань (`distance = «euclidean»`). Мінімальна кількість кластерів (`min.nc`) дорівнює 2, а максимальна кількість кластерів (`max.nc`) – 10. Метод кластеризації – ієрархічна кластеризація з повним зв'язком (`method = «complete»`). Опція «all» для параметра `index` вказує на те, що для оцінки рішень кластеризації буде обчислено декілька індексів кластеризації. Результат цього фрагменту коду вказує на те, що оптимальна кількість кластерів дорівнює 3.

Наступним кроком виконуємо кластеризацію методом k-середніх:

```
countries_kmeans<- kmeans(countries_scale, 3, nstart = 100)# метод к-середніх
```

```
fviz_cluster(countries_kmeans, data = countries_scale, frame.type = "convex") + theme_minimal()# візуалізація
```

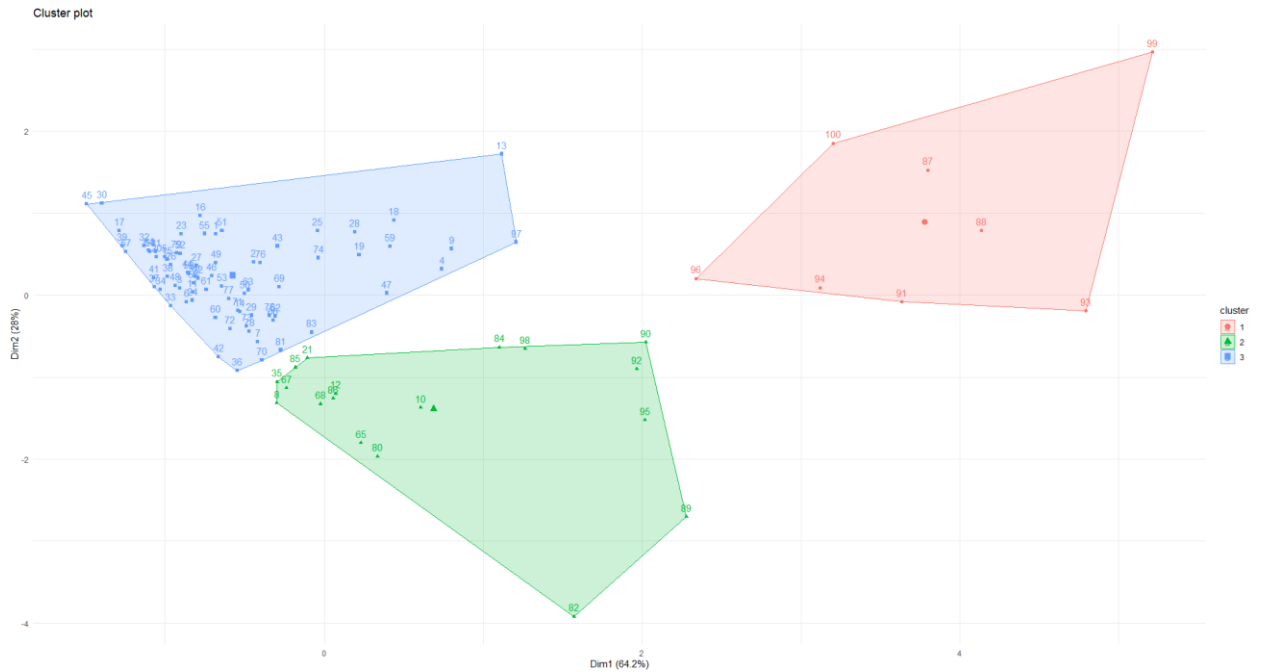


Рис. 3.4 – Візуальне відображення кластерів

Результати кластеризації методом К-середніх представлені на рисунку

3.5:

1 кластер				2 кластер	3 кластер
Albania	Georgia	North Macedonia	Tunisia	Belarus	Azerbaijan
Argentina	Ghana	Malta	Turkey	Canada	Brunei Darussalam
Armenia	Greece	Montenegro	United States	China	Kuwait
Australia	Croatia	Malaysia	Brazil	Estonia	Oman
Austria	Indonesia	Namibia	Finland	Kyrgyz Republic	Qatar
Belgium	India	Netherlands	Guatemala	Ukraine	Saudi Arabia
Bulgaria	Ireland	New Zealand	Honduras	Benin	Angola
Bolivia	Israel	Peru	Cambodia	Bosnia and Herzegovina	Norway
Chile	Italy	Philippines	Myanmar	Ethiopia	
Colombia	Jordan	Poland	Pakistan	Mozambique	
Czech Republic	Japan	Portugal	Paraguay	Nigeria	
Germany	Korea, Rep.	Romania	Sweden	Nepal	
Denmark	Lebanon	Senegal	Tajikistan	Tanzania	
Dominican Republic	Lithuania	Singapore	Uruguay	Iran, Islamic Rep.	
Ecuador	Luxembourg	El Salvador	Kenya	Kazakhstan	
Egypt, Arab Rep.	Latvia	Suriname	Niger	Libya	
Spain	Morocco	Slovak Republic	United Arab Emirates	Russian Federation	
France	Moldova	Slovenia		Mongolia	
United Kingdom	Mexico	Thailand			

Рис.3.5 – Результати кластеризації методом К-середніх

Можемо побачити, що більшість країн світу, які являються імпортерами енергоресурсів знаходяться у першому кластері. Адже імпорт електроенергії негативно впливає на рівень енергобезпеки. А в другому та третьому кластері знаходяться країни з дуже високим показником експорту електроенергії, що

має позитивний вплив на енергетичну безпеку.

### 3.2 Створення інтерактивного застосунку за допомогою Shiny

Тепер за допомогою фреймворку веб-додатків Shiny створимо інтерактивний застосунок для динамічного відображення результатів кластерного аналізу. Для початку підключаємо усі необхідні пакети, які можуть бути використані в подальшій розробці застосунку:

```
install.packages("shiny ")  
install.packages("datasets ")  
install.packages("cluster")  
install.packages("ggplot2")  
install.packages("factoextra ")  
install.packages("magrittr ")  
install.packages("dplyr ")
```

```
library(shiny)  
library(datasets)  
library(cluster)  
library(ggplot2)  
library(factoextra)  
library(magrittr)  
library(dplyr)
```

Наступний блок зчитує CSV-файл під назвою "countries.txt" з країнами та їхніми показниками у фрейм даних з назвою data. Функція "scale" використовується для стандартизації даних, за винятком першого стовпчика.

```
data <- read.csv("C:/Users/ilato/OneDrive/Робочий стол/countries.txt", header =  
TRUE, sep = "\t")  
dataframe <- as.data.frame(data)
```

```
data_scale <- scale(dataframe[-1], center = TRUE, scale = TRUE) # Standardize the
data
```

Наступним кроком нам необхідно зазначити як буде виглядати користувацький інтерфейс. Для цього ми скористаємось таким блоком коду. Він включає панель заголовка з текстом " Cluster Analysis " і макет, що складається з бічної панелі та головної панелі. Панель бічної панелі містить елементи керування для вибору методу кластеризації та кількості кластерів, а також кнопку запуску кластеризації. Основна панель містить область виведення графіка та панель вкладок з двома вкладками: " Cluster Membership " та " Number of Countries " (рис. 3.6):

```
ui <- fluidPage(
  titlePanel("Cluster Analysis"),
  sidebarLayout(
    sidebarPanel(
      selectInput("method", "Select clustering method:",
        choices = c("K-means", "Hierarchical", "PAM")),
      numericInput("clusters", "Number of clusters:", value = 3),
      actionButton("clusterButton", "Run Clustering")
    ),
    mainPanel(
      plotOutput("clusterPlot"),
      tabsetPanel(
        tabPanel("Cluster Membership", tableOutput("clusterMembershipTable")),
        tabPanel("Number of Countries", tableOutput("numberOfCountriesTable"))
      )
    )
  )
)
```

The image shows a web-based interface for cluster analysis. At the top, the title 'Cluster Analysis' is displayed. Below it, there is a form with three main components: a dropdown menu labeled 'Select clustering method:' with 'K-means' selected, a text input field labeled 'Number of clusters:' containing the number '3', and a button labeled 'Run Clustering'. Below the form, there are two tabs: 'Cluster Membership' and 'Number of Countries'. The 'Number of Countries' tab is currently selected.

Рис. 3.6 – Користувацький інтерфейс

Далі йде основний та самий великий блок коду під назвою “server” (наведено у Дод. Б). Цей блок визначає функцію сервера для додатка Shiny. Він містить реактивні вирази та функції рендерингу для генерації динамічного контенту на основі даних, введених користувачем. Реактивний вираз `clusters` обчислює результат кластеризації на основі обраного методу та кількості кластерів. Реактивний вираз `clusterPlot` генерує графік на основі результату кластеризації. Функція рендерингу `output$clusterPlot` рендерить графік у реактивному виразі `clusterPlot`. Реактивний вираз `clusterMembershipTable` обчислює таблицю приналежності країн до визначених кластерів на основі результату кластеризації. Функція рендерингу `output$clusterMembershipTable` виводить таблицю приналежності країн до визначених кластерів. Реактивний вираз `clusterMembershipTableSorted` сортує таблицю приналежності до кластерів за номером кластера, а не за назвою країни, для зручного перегляду. Реактивний вираз `numberOfCountriesTable` обчислює таблицю з кількістю країн у кожному кластері та динамічно змінюється в залежності від обраної кількості кластерів. Функція рендерингу `output$numberOfCountriesTable` виводить таблицю з кількістю країн. Запуск програми відбувається за

допомогою “ shinyApp ”. Цей блок запускає додаток Shiny, викликаючи функцію shinyApp з визначеним інтерфейсом і сервером. Результати виконання динамічної кластеризації за допомогою Shiny представлені на рисунках 3.7, 3.8 та 3.9.

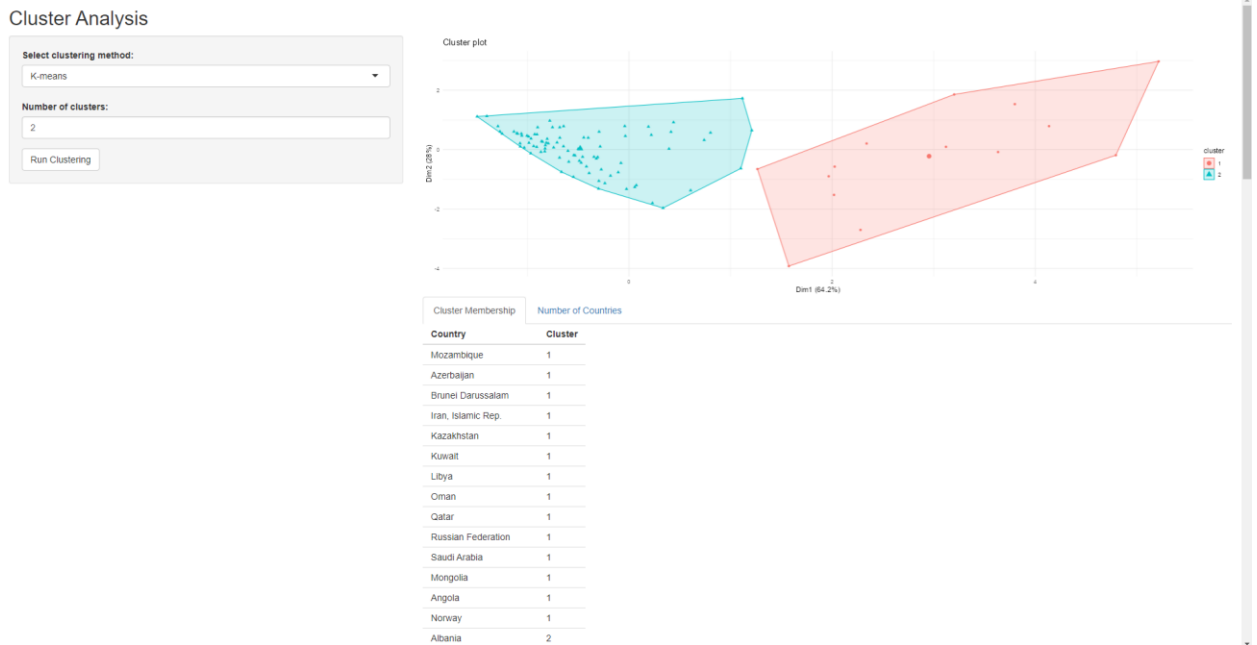


Рис. 3.7 – Приклад виконання кластерного аналізу методом К-середніх з двома кластерами



Рис. 3.8 – Приклад виконання кластерного аналізу методом К-середніх з

чотирма кластерами та відображенням кількості країн в кожному кластері

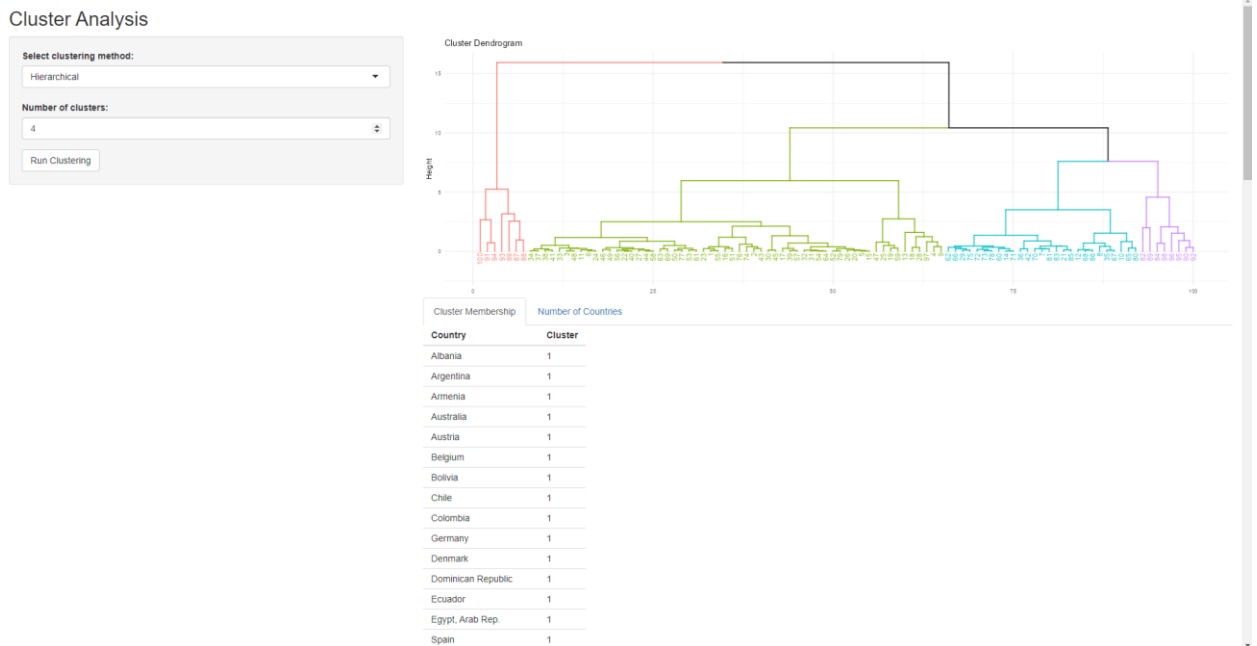


Рис. 3.9 – Приклад виконання кластерного аналізу ієрархічним методом Уорда з чотирма кластерами

Підсумовуючи усі блоки коду, можна дійти до висновку, що цей застосунок створює додаток Shiny для кластерного аналізу. Він надає користувацький інтерфейс для вибору методу кластеризації та кількості кластерів. На основі введених користувачем даних програма виконує кластеризацію, генерує графік і відображає таблиці приналежності до кластерів та кількості країн. Таблиця приналежності до кластерів відсортована за номером кластера.

### 3.3 Порівняння результатів класифікації в інтерактивному застосунку

Для підсумовування результатів кластерного аналізу країн за їх рівнем енергобезпеки візьмемо результат кластеризації за методом К-середніх із кількістю кластерів = 3. Результати цієї кластеризації представлені на



рисунках 3.10 та 3.11:

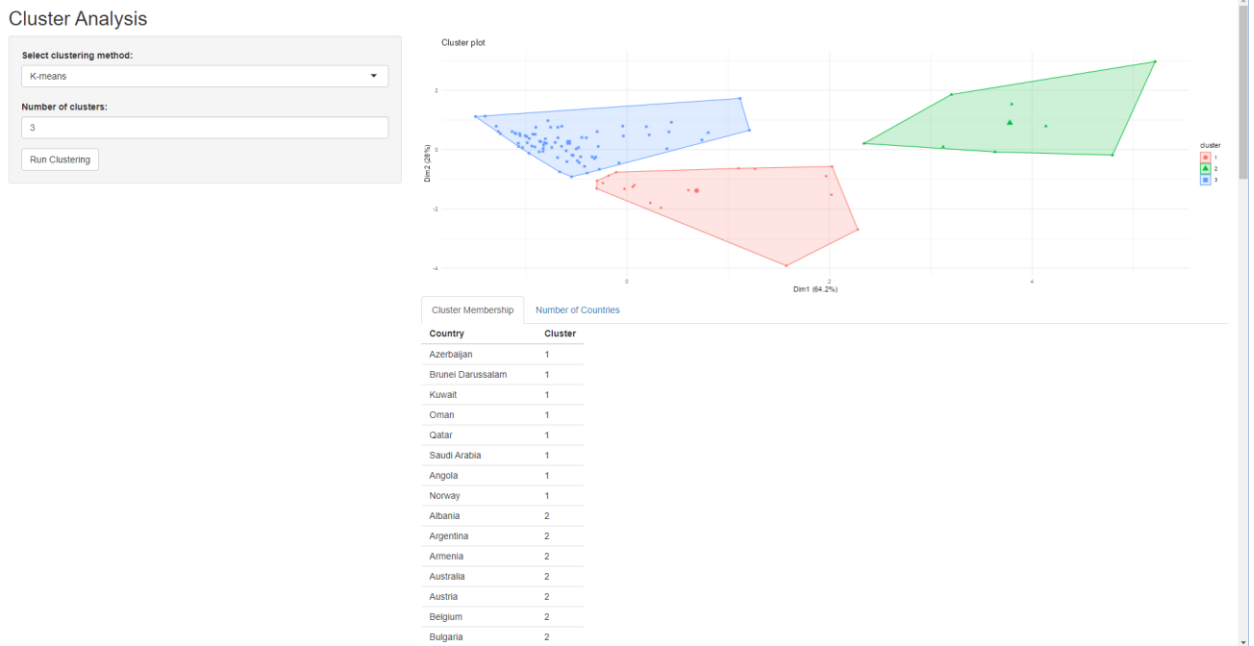


Рис. 3.10 – Результати кластеризації методом К-середніх з трьома кластерами

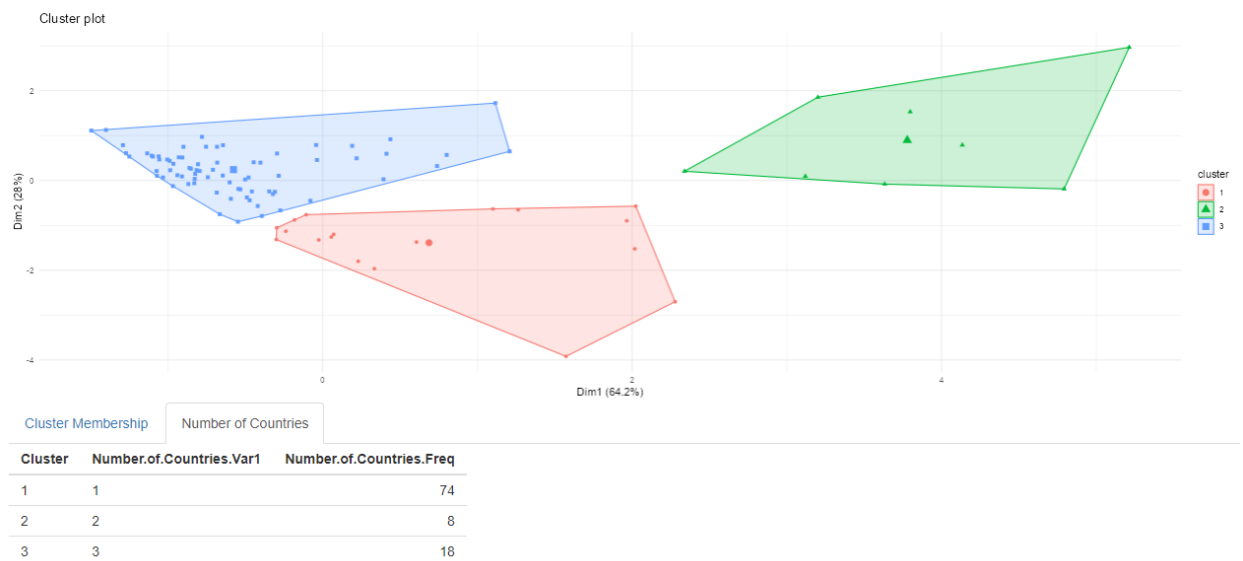


Рис. 3.11 – Результати кластеризації методом К-середніх з трьома кластерами

Тепер нам необхідно дізнатись основну інформацію отриману за рахунок кластерного аналізу:

K-means clustering with 3 clusters of sizes 18, 74, 8

Cluster means:

Energy\_intensity Energy\_depletion Energy\_import

1	1.5475393	0.0711944	-0.06863732
2	-0.4445632	-0.3108897	0.30904582
3	0.6302462	2.7155419	-2.70423986

Clustering vector:

```
[1] 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 2
2 2 2 2 1
[81] 2 1 2 1 1 1 3 3 1 1 3 1 3 3 1 3 2 1 3 3
```

within cluster sum of squares by cluster:

```
[1] 29.38862 45.39250 29.41566
(between_SS / total_SS = 64.9 %)
```

Почнемо аналізувати отримані дані з кінця. Кластерна сума квадратів: Сума квадратів кластерів являє собою суму квадратів відстаней між кожною точкою даних та центроїдом (центром) відповідного кластера. Це міра того, наскільки добре точки даних у кожному кластері згруповані разом. Менша сума квадратів кластера вказує на те, що точки даних у кластері розташовані ближче до свого центроїда, що свідчить про більш щільний та однорідний кластер. З іншого боку, більша сума квадратів кластера вказує на більшу варіабельність або дисперсію точок даних у кластері. Кластерний вектор: Кластерний вектор вказує на призначення кластера для кожного спостереження або точки даних у наборі даних. У вашому випадку вектор кластерів присвоює номер кластера кожній країні. Наприклад, якщо країні присвоєно номер кластера 2, це означає, що вона належить до кластера 2. Кластерний вектор допомагає визначити приналежність кожної країни до відповідного кластеру.

У наших конкретних результатах внутрішньокластерна сума квадратів за кластерами виглядає як [1] 29.38862, 45.39250, 29.41566, а співвідношення between\_SS/total\_SS 64,9% вказує на те, що алгоритм кластеризації пояснює значну частку загальної мінливості даних. У цьому випадку можна припустити, що алгоритм кластеризації зміг охопити значну кількість основних закономірностей і структури в даних.

Тепер зробимо висновки на рахунок кожного кластеру окремо, за результатами кластеризації.

### Кластер 1:

Енергоемність: -0.4445632

Виснаження енергетичних запасів: -0.3108897

Імпорт енергії: 0.30904582

### Кластер 2:

Енергоемність: 1.5475393

Виснаження енергетичних запасів: 0.0711944

Імпорт енергії: -0.06863732

### Кластер 3:

Енергоемність: 0.6302462

Виснаження енергетичних запасів: 2.7155419

Імпорт енергії: -2.70423986

На основі цих результатів ми можемо зробити наступні спостереження:

Кластер 1: Цей кластер має відносно нижчу енергоемність, що вказує на те, що країни цього кластеру мають нижчі витрати енергії на одиницю економічного продукту порівняно з іншими кластерами. Показники виснаження енергоресурсів та імпорту енергоресурсів також відносно нижчі, але позитивні. Це свідчить про те, що ці країни мають помірний рівень енергетичної безпеки з відносно меншою залежністю від імпорту енергоносіїв та більш ефективним використанням енергоресурсів.

Кластер 2: Цей кластер має найвищий показник енергоемності, що свідчить про те, що країни цього кластеру споживають більше енергії на одиницю економічного продукту. Показник виснаження енергоресурсів є позитивним, але відносно низьким, тоді як показник імпорту енергоресурсів є від'ємним. Це свідчить про те, що ці країни мають вищий рівень споживання енергії і можуть бути більш залежними від імпорту енергоносіїв. Від'ємне

значення імпорту енергоресурсів може свідчити про те, що ці країни є експортерами енергоресурсів, що може сприяти їхній енергетичній безпеці за рахунок отримання доходів від експорту енергоресурсів.

Кластер 3: Цей кластер має помірне значення енергоемності, вищий показник виснаження енергоресурсів і значно від'ємне значення імпорту енергоресурсів. Вищий показник виснаження енергоресурсів вказує на те, що країни цього кластеру можуть зіткнутися з проблемами, пов'язаними з виснаженням їхніх енергетичних ресурсів. Значно від'ємне значення імпорту енергоресурсів свідчить про те, що ці країни мають потужний потенціал для експорту енергоресурсів, що може сприяти їхній енергетичній безпеці, приносячи дохід. Однак вищий рівень виснаження енергетичних ресурсів викликає занепокоєння щодо довгострокової стійкості їхніх енергоресурсів.

За результатом кластерного аналізу методом К-середніх можемо дійти до висновку, що в третій кластер входять країни-експортери енергоресурсів, їх імпортозалежність найнижча, а виснаження енергетичних запасів найбільша, що позитивно впливає на економіку. Тому можна сказати, що країни в третьому кластері мають досить високий рівень енергобезпеки. До другого кластеру також відносяться країни-експортер (або країни, які скуповують не велику кількість електроенергії), але рівень їх експорту вже значно менший ніж у країн з першого кластеру. Рівень їх енергобезпеки можна назвати задовільним. Ну і в першому кластері знаходиться найбільша кількість країн. Всі ці країни сильно залежні від імпорту електроенергії, а рівень виснаження їх енергетичних запасів найменший, це означає, що вони не використовують усі доступні ресурси, а скуповують їх з інших країн.

Для порівняння скористаємось кластерним аналізом методом РАМ також на три кластери. РАМ розшифровується як Partitioning Around Medoids. Це алгоритм кластеризації, який належить до сімейства методів розбиття, схожий на кластеризацію за методом k-середніх. Основна відмінність між k-середніх і РАМ полягає в тому, як визначаються центри кластерів, або "медоїди". При кластеризації за методом k-середніх центри кластерів

представлені середніми значеннями точок даних у кожному кластері. Однак у методі РАМ центри кластерів представлені фактичними точками даних з набору даних, які називаються "медоїдами". Медоїди обираються як найбільш центрально розташовані точки в межах кожного кластера, мінімізуючи розбіжність (відстань) між медоїдом та іншими точками в тому ж кластері. Результати методу РАМ представлені нижче (рис. 3.12):



Рис. 3.12 - Результати кластеризації методом РАМ з трьома кластерами

Medoids:

```

ID Energy_intensity Energy_depletion Energy_import
[1,] 27      -0.6470206      -0.3910549      0.4617982
[2,] 81       0.5206095      -0.4799097      0.2446452
[3,] 96       0.6960281       1.8346958     -1.3349578

```

Clustering vector:

```

[1] 1 1 1 2 1 1 2 2 2 2 1 2 3 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2
1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 2 2 1 2 2 2 2 1 2
1 1 2 1 2
[81] 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3

```

Objective function:

```

build      swap
0.8569021 0.7927090

```

Тепер подивимось на загальні результати кластерного аналізу за допомогою двох різних методів та зробимо висновки:

K-середніх					
1 кластер			2 кластер		3 кластер
Albania	Georgia	North Macedonia	Tunisia	Belarus	Azerbaijan
Argentina	Ghana	Malta	Turkey	Canada	Brunei Darussalam
Armenia	Greece	Montenegro	United States	China	Kuwait
Australia	Croatia	Malaysia	Brazil	Estonia	Oman
Austria	Indonesia	Namibia	Finland	Kyrgyz Republic	Qatar
Belgium	India	Netherlands	Guatemala	Ukraine	Saudi Arabia
Bulgaria	Ireland	New Zealand	Honduras	Benin	Angola
Bolivia	Israel	Peru	Cambodia	Bosnia and Herzegovina	Norway
Chile	Italy	Philippines	Myanmar	Ethiopia	
Colombia	Jordan	Poland	Pakistan	Mozambique	
Czech Republic	Japan	Portugal	Paraguay	Nigeria	
Germany	Korea, Rep.	Romania	Sweden	Nepal	
Denmark	Lebanon	Senegal	Tajikistan	Tanzania	
Dominican Republic	Lithuania	Singapore	Uruguay	Iran, Islamic Rep.	
Ecuador	Luxembourg	El Salvador	Kenya	Kazakhstan	
Egypt, Arab Rep.	Latvia	Suriname	Niger	Libya	
Spain	Morocco	Slovak Republic	United Arab Emirates	Russian Federation	
France	Moldova	Slovenia		Mongolia	
United Kingdom	Mexico	Thailand			
PAM					
1 кластер			2 кластер		3 кластер
Albania	Ireland	Poland	Australia	Finland	Colombia
Argentina	Israel	Portugal	Bulgaria	Guatemala	Azerbaijan
Armenia	Italy	Romania	Belarus	Honduras	Brunei Darussalam
Austria	Jordan	Senegal	Bolivia	Cambodia	Iran, Islamic Rep.
Belgium	Japan	Singapore	Canada	Pakistan	Kazakhstan
Chile	Lebanon	El Salvador	China	Tajikistan	Kuwait
Germany	Lithuania	Suriname	Czech Republic	Ethiopia	Libya
Denmark	Luxembourg	Slovenia	Estonia	Kenya	Oman
Dominican Republic	Latvia	Tunisia	India	Mozambique	Qatar
Ecuador	Morocco	Turkey	Kyrgyz Republic	Niger	Russian Federation
Egypt, Arab Rep.	Mexico	Brazil	Korea, Rep.	Nigeria	Saudi Arabia
Spain	North Macedonia	Myanmar	Moldova	Nepal	United Arab Emirates
France	Malta	Paraguay	Malaysia	Tanzania	Angola
United Kingdom	Montenegro	Sweden	Slovak Republic	Mongolia	Norway
Georgia	Namibia	Uruguay	Thailand		
Ghana	Netherlands		Ukraine		
Greece	New Zealand		United States		
Croatia	Peru		Benin		
Indonesia	Philippines		Bosnia and Herzegovina		

Рис. 3.13 – Результати кластерного аналізу методом K-середніх та PAM

Можемо одразу помітити, що різні методи дали схожий між собою результат. Але також можемо побачити і деякі відмінності, наприклад Іран, Росія, Об'єднані Арабські Емірати, Лівія, Колумбія та Казахстан перемістилися з першого-другого кластерів до третього кластера, в той час як Австралія, Болгарія, Болівія і декілька інших країн перемістились з першого кластера до другого. Це відбувається через різність алгоритмів кластеризації у цих двох методів. У методі k-середніх кластери формуються шляхом мінімізації суми квадратів відстаней між точками даних і центроїдами кластерів. Цей метод спрямований на мінімізацію внутрішньокластерної варіабельності. Він є обчислювально ефективним і широко використовується. Однак k-середні можуть бути чутливими до початкового вибору центроїдів кластерів і припускають, що кластери мають сферичну форму і подібну

дисперсію.

З іншого боку, РАМ (Partitioning Around Medoids) є варіантом кластеризації за  $k$ -медоїдами. Замість того, щоб використовувати центроїди кластерів, РАМ вибирає репрезентативні точки даних, які називаються медоїдами, як центри кластерів. РАМ мінімізує розбіжність (відстань) між точками даних і вибраними медоїдами. РАМ є більш стійким до викидів і може обробляти несферичні кластери. Однак він може бути дорожчим в обчислювальному плані, ніж  $k$ -середні.

Характеристики кластерів, такі як енергоємність, виснаженість енергоресурсів та рівень імпорту енергоресурсів, є дещо схожими між цими двома методами, однак РАМ ефективніше показує себе з невеликою кількістю даних, ніж з великою. Причиною цього є те, що РАМ вимагає попарних обчислень відстані між кожною точкою даних у наборі даних, що може бути обчислювально дорогим і займати багато часу для великих наборів даних. У РАМ алгоритм повинен обчислювати розбіжність (відстань) між кожною парою точок даних, щоб визначити найкращі медоїди. Зі збільшенням кількості точок даних кількість обчислень парних відстаней зростає експоненціально, що призводить до значного збільшення обчислювальних вимог. Крім того, РАМ передбачає ітеративний процес оновлення медоїдів і перепризначення точок даних до найближчого медоїда, що також може вимагати значних обчислювальних витрат для великих наборів даних. Як наслідок, можна вважати, що метод  $K$ -середніх в даному випадку дав більш адекватний результат.

## ВИСНОВОК

Отже, аналіз, проведений у цій роботі, показує, що енергетична безпека наразі стикається зі значними викликами в усьому світі. Вивчення індикаторів енергетичної безпеки для 100 країн демонструє, що стан енергетичної безпеки не є оптимальним.

Результати кластерного аналізу проливають світло на чіткі закономірності та характеристики в наборі даних. Завдяки застосуванню алгоритмів кластеризації, таких як k-середнє та метод Уорда, країни були згруповані на основі їхніх профілів енергетичної безпеки. Визначені кластери демонструють відмінності в показниках енергоємності, виснаження енергоресурсів та імпорту енергоресурсів, надаючи цінну інформацію про різноманітний ландшафт енергетичної безпеки в різних країнах. Більша половина із 100 країн вимушені імпортувати енергоресурси, не використовують повний потенціал свої запасів, від чого їх енергобезпека знаходиться у поганому стані. Таку саму проблему має і Україна, яка також страждає від імпорту енергії, а в військові часи ця проблема особливо актуальна. До позитивних моментів можна віднести поступове зростання частки власних джерел енергії, зокрема газу, в загальному балансі паливно-енергетичних ресурсів. Той факт, що наразі ми видобуваємо близько 69% газу, який споживаємо, є надзвичайно важливим з огляду на поточний стан справ. Зменшення енергоємності національної економіки - ще одна позитивна тенденція.

Використовуючи можливості мови програмування R та бібліотеки Shiny, було розроблено інтерактивний додаток для полегшення вивчення та візуалізації результатів кластерного аналізу. Аналіз, проведений у цій роботі, слугує основою для подальших досліджень і пошуків шляхів пом'якшення викликів, з якими стикаються країни у досягненні оптимальної енергетичної безпеки.

Кластерний аналіз має кілька переваг для оцінки рівня енергетичної



безпеки. Він дозволяє виявити закономірності, спростити та візуалізувати складні дані. Кластерний аналіз полегшує порівняльний аналіз і підтримку прийняття рішень шляхом виявлення кластерів зі схожими характеристиками енергетичної безпеки. Він є гнучким і адаптивним до різних наборів даних і дослідницьких контекстів. Однак кластерний аналіз також має обмеження. Він передбачає суб'єктивну інтерпретацію та прийняття рішень. Результати можуть бути чутливими до попередньої обробки даних і можуть призвести до надмірного узагальнення. Кластерний аналіз не встановлює причинно-наслідкових зв'язків і не пояснює глибинні механізми. Обмеження та якість даних можуть впливати на точність результатів кластеризації. Загалом можна зробити висновок, що кластерний аналіз підходить для оцінки рівня енергетичної безпеки. Об'єднуючи країни в кластери, кластерний аналіз забезпечує корисну основу для розуміння відмінностей в енергетичній безпеці між різними регіонами або групами країн.

Shiny - це пакет R, який дозволяє створювати інтерактивні веб-додатки. Він дозволяє розробляти зручні інтерфейси та інтерактивні візуалізації. До переваги використання Shiny можна віднести: додатки Shiny забезпечують інтерактивний досвід для користувачів; Shiny дозволяє гнучко налаштовувати зовнішній вигляд та функціональність програми; додатки Shiny можна легко поширювати та відтворювати, що полегшує спільну роботу та обмін результатами. Але, як і будь-яка програма, Shiny має свої мінуси: Додатки Shiny можуть стикатися з обмеженнями продуктивності, особливо при роботі з великими наборами даних; також, що я особисто виділив, це відсутність достатньої кількості прикладів використання додатків Shiny, що ускладнює процес навчання та практики розробки таких інтерактивних застосунків. Загалом Shiny підходить для розробки інтерактивного додатку для кластерного аналізу та візуалізації енергетичної безпеки. Він дозволяє користувачам досліджувати та аналізувати результати кластеризації в інтерактивному режимі, покращуючи їх розуміння та процес прийняття рішень.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Паливно-енергетичний комплекс України на порозі третього тисячоліття [За заг. ред. А.К. Шидловського, М.П. Ковалка]. – К.: УЕЗ, -2001. – 398 с.
2. Сайт світової енергетичної ради. [Електронний ресурс] . – Режим доступу до ресурсу: [http://www.worldenergy.org/document/ethiopia\\_june\\_30\\_v\\_gbeddy\\_security.pdf](http://www.worldenergy.org/document/ethiopia_june_30_v_gbeddy_security.pdf).
3. Гійон, Ж.-М., Лерше, Ж.-М., & Нуссель, П. (2018). Енергетична безпека в епоху глобалізації. Київ: Видавництво "Інститут стратегічних досліджень"
4. Хінклі, Р. Е., Калландер, Дж. Е., & Лалонд, Ж.-Б. (2016). Енергетична безпека: концепції, виклики та стратегії. Науковий журнал енергетичної безпеки, 25(3), 45-62.
5. Микитенко В.В. На чому базується енергетична безпека держави/В.В. микитенко// Вісник НАН України. – 2005. - № 3. – С.41-47.
6. Земляний М.Г. До оцінки рівня енергетичної безпеки. Концептуальні підходи/ М.Г. Земляний// Стратегічна панорама. – 2009. - № 2. – С.56 – 64.
7. Сухін Є.І. Нетрадиційна енергетика як фактор економічної безпеки держави: автореф. дис...д-ра екон. наук: спец. 21.04.01 «Економічна безпека держави» / Сухін Євген Ілліч // Рада національної безпеки та оборони України; Національний ін-т проблем міжнародної безпеки. – К., 2005. – 38 с.
8. Розпорядження КМУ «Про схвалення Енергетичної стратегії України на період до 2030 року» № 145-р від 15.03.2006 р. //Сайт Верховної Ради України [Електронний ресурс] . – Режим доступу до ресурсу: <http://zakon1.rada.gov.ua/laws/show/145-2006-%D1%80>
9. Державна фіскальна служба України [Електронний ресурс]. – Режим доступу до ресурсу: <https://mof.gov.ua/uk/state-fiscal-service-state-fiscal-service>
10. Дейна А. Ю. Енергетична незалежність України [Електронний ресурс]. – Режим доступу до ресурсу: <http://nasoa.edu.ua/wp->

[content/uploads/zah/deina\\_dis.pdf](content/uploads/zah/deina_dis.pdf)

11. The World Bank [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.worldbank.org/en/home>

12. Сайт світової енергетичної ради: [Електронний ресурс] . – Режим доступу: <https://www.worldenergy.org/>

13. Енергетична безпека [Електронний ресурс]. – Режим доступу до ресурсу:

[https://uk.wikipedia.org/wiki/%D0%95%D0%BD%D0%B5%D1%80%D0%B3%D0%B5%D1%82%D0%B8%D1%87%D0%BD%D0%B0\\_%D0%B1%D0%B5%D0%B7%D0%BF%D0%B5%D0%BA%D0%B0](https://uk.wikipedia.org/wiki/%D0%95%D0%BD%D0%B5%D1%80%D0%B3%D0%B5%D1%82%D0%B8%D1%87%D0%BD%D0%B0_%D0%B1%D0%B5%D0%B7%D0%BF%D0%B5%D0%BA%D0%B0)

14. Енергетична безпека України: методологія системного аналізу та стратегічного планування [Електронний ресурс]. – Режим доступу до ресурсу: [https://niss.gov.ua/sites/default/files/2020-](https://niss.gov.ua/sites/default/files/2020-12/sukhodolia_energy_security_sayt-1.pdf)

[12/sukhodolia\\_energy\\_security\\_sayt-1.pdf](https://niss.gov.ua/sites/default/files/2020-12/sukhodolia_energy_security_sayt-1.pdf)

15. Електронний журнал «Ефективна економіка» [Електронний ресурс]. – Режим доступу до ресурсу: <http://www.economy.nayka.com.ua/>

Халатов А.А енергетична безпека України: методологічні засади оцінки рівня безпеки та порівняльний аналіз поточного стану [Електронний ресурс]. –

Режим доступу до ресурсу:

[https://ela.kpi.ua/bitstream/123456789/38239/1/TeplofizTeploenerget-42-2\\_p18-30.pdf](https://ela.kpi.ua/bitstream/123456789/38239/1/TeplofizTeploenerget-42-2_p18-30.pdf)

16. Класифікація кластер-процедур. Ієрархічні агломеративні й ітеративні кластер-процедури [Електронний ресурс]. – Режим доступу до ресурсу: <http://ebooks.git-elt.hneu.edu.ua/babap/4-1-id4-1.html>

17. Проблема вичерпання природних ресурсів [Електронний ресурс]. – Режим доступу до ресурсу:

[https://uk.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%B1%D0%BB%D0%B5%D0%BC%D0%B0\\_%D0%B2%D0%B8%D1%87%D0%B5%D1%80%D0%BF%D0%B0%D0%BD%D0%BD%D1%8F\\_%D0%BF%D1%80%D0%B8%D1%80%D0%BE%D0%B4%D0%BD%D0%B8%D1%85\\_%D1%80%D0%B5](https://uk.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%B1%D0%BB%D0%B5%D0%BC%D0%B0_%D0%B2%D0%B8%D1%87%D0%B5%D1%80%D0%BF%D0%B0%D0%BD%D0%BD%D1%8F_%D0%BF%D1%80%D0%B8%D1%80%D0%BE%D0%B4%D0%BD%D0%B8%D1%85_%D1%80%D0%B5)

[D1%81%D1%83%D1%80%D1%81%D1%96%D0%B2](#)

18. Виснаження природних ресурсів, причини, наслідки та рішення [Електронний ресурс]. – Режим доступу до ресурсу: <https://environmentgo.com/uk/depletion-of-natural-resources/>

19. Енергоємність ВВП [Електронний ресурс]. – Режим доступу до ресурсу: [https://uk.wikipedia.org/wiki/%D0%95%D0%BD%D0%B5%D1%80%D0%B3%D0%BE%D1%94%D0%BC%D0%BD%D1%96%D1%81%D1%82%D1%8C\\_%D0%92%D0%92%D0%9F](https://uk.wikipedia.org/wiki/%D0%95%D0%BD%D0%B5%D1%80%D0%B3%D0%BE%D1%94%D0%BC%D0%BD%D1%96%D1%81%D1%82%D1%8C_%D0%92%D0%92%D0%9F)

20. Статистика щодо розвитку відновлюваної енергетики [Електронний ресурс]. – Режим доступу до ресурсу: <https://sae.gov.ua/uk/content/informatsiyni-materialy>

21. Міжнародне енергетичне агентство [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.iea.org/>

22. Кластерний аналіз в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://habr.com/ru/articles/685040/>

23. Оцінка якості кластеризації [Електронний ресурс]. – Режим доступу до ресурсу: <https://rpubs.com/AllaT/clust3>

24. Shiny [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.paulamoraga.com/book-geospatial/sec-shiny.html>

25. Shiny [Електронний ресурс]. – Режим доступу до ресурсу: <https://teachdatascience.com/shiny1/>

26. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://data-flair.training/blogs/clustering-in-r-tutorial/>

27. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.analyticsvidhya.com/blog/2021/04/beginners-guide-to-clustering-in-r-program/>

28. Хэдли Уикем Вивчаємо Shiny - 2022. – С.24-89

29. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://techvidvan.com/tutorials/cluster-analysis-in-r/>
30. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.statmethods.net/advstats/cluster.html>
31. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.dominodatalab.com/blog/clustering-in-r>
32. Мова програмування R [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.techtarget.com/searchbusinessanalytics/definition/R-programming-language>
33. Мова програмування R [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.guru99.com/r-programming-introduction-basics.html>
34. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу:  
[https://agricolamz.github.io/DS\\_for\\_DH/%D0%BA%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F.html](https://agricolamz.github.io/DS_for_DH/%D0%BA%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F.html)
35. Shiny: <https://shiny.posit.co/>
36. Кластеризація в R [Електронний ресурс]. – Режим доступу до ресурсу: <https://coderlessons.com/tutorials/mashinnoe-obuchenie/r-programmirovaniye/31-k-oznachaet-klasterizatsiiu-v-r>
37. R Programming Language [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/r-programming-language-introduction/>
38. The R Project for Statistical Computing [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.r-project.org/>
39. R Programming Language [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.sciencedirect.com/topics/mathematics/r-programming-language>

40. R Programming Tutorial [Электронный ресурс]. – Режим доступа до ресурсу: <https://stackoverflow.blog/2017/10/10/impressive-growth-r/>

## ДОДАТКИ

Додаток А

Таблиця А

Показники, які були використані для кластерного аналізу

Country_Name	Energy_intensity	Energy_depletion	Energy_import
Albania	2.57	0.99161618	13.7979709
Argentina	3.31	1.51957943	13.02936074
Armenia	3.51	9.05E-09	71.34130452
Australia	4.39	1.19591914	-192.0159244
Austria	2.81	0.05909206	62.42378603
Belgium	3.82	2.632E-05	76.2588603
Bulgaria	4.98	0.06007288	36.55077772
Belarus	6.29	0.67381135	86.77969893
Bolivia	3.99	2.68517183	-177.9858361
Canada	7.1	0.51704785	-67.92722174
Chile	3.66	0.01020439	61.28829787
China	6.39	0.66578264	15.02180074
Colombia	2.29	4.08766045	-274.1331607
Czech Republic	4.28	0.05171702	28.98980679
Germany	2.87	0.02789628	60.87591028
Denmark	2.14	0.51002499	0.944353805
Dominican Republic	2.06	0	86.71710022
Ecuador	3.1	3.61291989	-114.7121382
Egypt, Arab Rep.	3.59	5.01774446	-7.391493976
Spain	2.77	0.00077874	69.35978962
Estonia	5.54	0.28404536	3.397203439
France	3.4	0.00668837	43.48527185
United Kingdom	2.41	0.59804636	39.67463181
Georgia	3.82	0.00899438	68.75204873
Ghana	2.92	4.16295107	-8.192103873
Greece	3	0.04078248	61.97067022
Croatia	3.13	0.40688196	45.8596927
Indonesia	3.19	2.16605722	-103.0913295
India	4.4	0.56992141	34.30552375
Ireland	1.41	0.10993952	84.25580305
Israel	2.65	0.06932425	67.04873828
Italy	2.48	0.07515353	75.0013037
Jordan	3.82	0.00468456	96.81229729

Japan	3.45	0.00668539	93.98125952
Kyrgyz Republic	5.88	0.03265542	49.54466148
Korea, Rep.	5.46	0.00260727	81.70429565
Lebanon	3.36	0	97.86553522
Lithuania	3.22	0.02152809	75.04470837
Luxembourg	2.37	0	96.00213678
Latvia	3.35	0	45.16008775
Morocco	3.18	0.00409628	90.72183409
Moldova	5.09	0.00933538	90.01107669
Mexico	3.09	2.07995165	-10.79360098
North Macedonia	3.23	0.07999433	51.77099874
Malta	1.36	0	98.38148617
Montenegro	3.42	0.03267161	27.5691369
Malaysia	4.5	4.95627551	-5.505028941
Namibia	3.44	0	74.44166879
Netherlands	3.17	0.32329651	19.76928935
New Zealand	3.93	0.37282579	17.08455259
Peru	2.56	0.29899202	-14.89958814
Philippines	2.81	0.17369544	45.76800308
Poland	3.68	0.11898543	28.39101101
Portugal	2.63	0	71.66828631
Romania	2.52	0.67224944	16.77549802
Senegal	3.48	0.00334746	52.69248699
Singapore	2.51	0	97.68200129
El Salvador	3.27	0	49.23546573
Suriname	3.57	5.15366476	-43.75983539
Slovak Republic	4.28	0.0041036	58.82671105
Slovenia	3.67	0.00993296	44.49892058
Thailand	4.5	1.69734069	41.56815174
Tunisia	3.85	1.29115808	36.19587142
Turkey	2.59	0.08807288	74.20811136
Ukraine	7.51	0.53846581	27.20850822
United States	4.61	0.351032	9.214465501
Benin	6.05	0.09225986	46.59070713
Bosnia and Herzegovina	6.54	0.04205698	22.73008923
Brazil	3.93	1.69334708	11.87013338
Finland	5.37	0	46.19727551
Guatemala	4.25	0.13728355	32.84420958
Honduras	4.58	0	52.99643629
Cambodia	4.61	0.00022236	33.11629751
Myanmar	3.52	2.48709455	-32.97079407
Pakistan	4.48	0.88197687	24.12316878
Paraguay	3.4	0	-36.93956313



Sweden	3.97	0	28.2711851
Tajikistan	4.73	0.05228934	36.24549558
Uruguay	2.79	0	44.43350088
Ethiopia	7.88	2.84E-06	5.926314417
Kenya	5.26	0.00284768	17.16990331
Mozambique	12.07	0.39188977	-54.60162754
Niger	5.03	0.90302309	-5.809741918
Nigeria	6.16	4.45741078	-93.02721929
Nepal	5.68	0.00225765	16.67706471
Tanzania	6.49	0.26730982	10.72806025
Azerbaijan	4.32	18.4282563	-310.3846944
Brunei Darussalam	5.85	16.9321625	-357.3943792
Iran, Islamic Rep.	10.39	8.43993685	-33.39527219
Kazakhstan	6.65	9.31611994	-116.8904268
Kuwait	7.07	10.5418497	-391.0584858
Libya	7.16	8.64003757	-102.964112
Oman	7.79	23.7647134	-206.1947121
Qatar	6.46	7.62013121	-398.9911867
Russian Federation	8.25	8.05861887	-83.67054072
Saudi Arabia	5.58	10.5276597	-191.5238831
United Arab Emirates	4.11	5.1240922	-183.8439704
Mongolia	6.37	2.73850159	-168.0933932
Angola	2.95	22.4087853	-540.9974712
Norway	3.66	6.04029351	-582.8993394

Додаток Б

Код програми Б.1

```
install.packages("ggplot2")
install.packages("factoextra")
install.packages("cluster")
install.packages("NbClust")
install.packages("clValid")
install.packages("psych")
install.packages("MASS")
install.packages("gplots")
```

```
library(ggplot2)
library(factoextra)
library(cluster)
library(NbClust)
library(clValid)
library(psych)
library(MASS)
library(gplots)

countries<- read.csv("C:/Users/ilato/OneDrive/Рабочий стол/countries.txt",
header = T, sep = "\t")
View(countries)

countries<- as.data.frame(countries)

countries_scale <- scale(countries[-1], center = T, scale = T)#стандартизація
даних
head(countries_scale)
View(countries_scale)

fviz_nbclust(countries_scale, kmeans, method = "wss")# метод ліктя
fviz_nbclust(countries_scale, kmeans, method = "silhouette")# силуетний
метод
fviz_nbclust(countries_scale, kmeans, method = "gap_stat")# геп статистика

clust_optimal<- NbClust(countries_scale, distance = "euclidean", min.nc = 2,
max.nc = 10, method = "complete", index = "all")# пошук оптимальної
кільності кластерів

#k-means
```

```
countries_kmeans<- kmeans(countries_scale, 3, nstart = 100)# метод к-середніх
```

```
fviz_cluster(countries_kmeans, data = countries_scale, frame.type = "convex") + theme_minimal()# візуалізація
```

```
#Ієрархічна
```

```
countries_res<- hclust(dist(countries_scale, method = "euclidean"), method = "ward.D2")
```

```
graphs_vision<- cutree(countries_res, k = 2)
```

```
plot(countries_res, cex = 0.8)
```

```
View(countries[countries_kmeans$cluster==1,])# перший кластер
```

```
View(countries[countries_kmeans$cluster==2,])# другий кластер
```

```
View(countries[countries_kmeans$cluster==3,])# третій кластер
```

Код програми Б.2

```
library(shiny)
```

```
library(datasets)
```

```
library(cluster)
```

```
library(ggplot2)
```

```
library(factoextra)
```

```
library(magrittr)
```

```
library(dplyr)
```

```
# Load sample data
```

```
data <- read.csv("C:/Users/ilato/OneDrive/Рабочий стол/countries.txt", header = TRUE, sep = "\t")
```

```
dataframe <- as.data.frame(data)
```

```
data_scale <- scale(dataframe[-1], center = TRUE, scale = TRUE) #
Standardize the data
```

```
# Define UI
```

```
ui <- fluidPage(
  titlePanel("Cluster Analysis"),
  sidebarLayout(
    sidebarPanel(
      selectInput("method", "Select clustering method:",
        choices = c("K-means", "Hierarchical", "PAM")),
      numericInput("clusters", "Number of clusters:", value = 3),
      actionButton("clusterButton", "Run Clustering")
    ),
    mainPanel(
      plotOutput("clusterPlot"),
      tabsetPanel(
        tabPanel("Cluster Membership",
          tableOutput("clusterMembershipTable")),
        tabPanel("Number of Countries",
          tableOutput("numberOfCountriesTable"))
      )
    )
  )
)
```

```
# Define server
```

```
server <- function(input, output, session) {
```

```
  # Create reactive object for clustering result
```

```

clusters <- reactive({
  if (input$method == "K-means") {
    kmeans(data_scale, input$clusters, nstart = 100)
  } else if (input$method == "Hierarchical") {
    hclust(dist(data_scale), method = "ward.D2")
  } else if (input$method == "PAM") {
    pam(data_scale, input$clusters)
  }
})

# Create reactive object for cluster plot
clusterPlot <- reactive({
  if (input$clusterButton == 0) {
    return()
  }

  if (input$method == "K-means") {
    result <- kmeans(data_scale, input$clusters, nstart = 100) # K-means
    clustering
    fviz_cluster(result, data_scale, geom = "point", stand = FALSE) +
    theme_minimal()
  } else if (input$method == "Hierarchical") {
    result <- hclust(dist(data_scale), method = "ward.D2") # Hierarchical
    clustering
    fviz_dend(result, k = input$clusters) + theme_minimal()
  } else if (input$method == "PAM") {
    result <- pam(data_scale, input$clusters) # PAM clustering
    fviz_cluster(result, data_scale, geom = "point", stand = FALSE) +
    theme_minimal()
  }
})

```

```

# Render cluster plot
output$clusterPlot <- renderPlot({
  if (input$clusterButton == 0) {
    return()
  }
  clusterPlot()
})

# Create reactive object for cluster membership table
clusterMembershipTable <- reactive({
  if (input$clusterButton == 0) {
    return()
  }
  if (input$method == "K-means") {
    result <- kmeans(data_scale, input$clusters, nstart = 100) # K-means
    clustering
    data.frame(Country = data$Country, Cluster = factor(result$cluster))
  } else if (input$method == "Hierarchical") {
    result <- hclust(dist(data_scale), method = "ward.D2") # Hierarchical
    clustering
    clusterLabels <- cutree(result, k = input$clusters)
    data.frame(Country = data$Country, Cluster = factor(clusterLabels))
  } else if (input$method == "PAM") {
    result <- pam(data_scale, input$clusters) # PAM clustering
    data.frame(Country = data$Country, Cluster = factor(result$clustering))
  }
})

```

```
# Sort the cluster membership table by cluster number
clusterMembershipTableSorted <- reactive({
  clusterMembershipTable() %>%
    arrange(Cluster)
})
```

```
# Render cluster membership table
output$clusterMembershipTable <- renderTable({
  if (input$clusterButton == 0) {
    return()
  }
  clusterMembershipTableSorted()
})
```

```
# Create reactive object for number of countries table
numberOfCountriesTable <- reactive({
  if (input$clusterButton == 0) {
    return()
  }
  if (input$method == "K-means") {
    result <- kmeans(data_scale, input$clusters, nstart = 100) # K-means
    clustering
    clusterCounts <- table(result$cluster)
    data.frame(Cluster = factor(names(clusterCounts)), "Number of
    Countries" = clusterCounts)
  } else if (input$method == "Hierarchical") {
```

```

    result <- hclust(dist(data_scale), method = "ward.D2") # Hierarchical
clustering
    clusterLabels <- cutree(result, k = input$clusters)
    clusterCounts <- table(clusterLabels)
    data.frame(Cluster = factor(names(clusterCounts)), "Number of
Countries" = clusterCounts)
  } else if (input$method == "PAM") {
    result <- pam(data_scale, input$clusters) # PAM clustering
    clusterCounts <- table(result$clustering)
    data.frame(Cluster = factor(names(clusterCounts)), "Number of
Countries" = clusterCounts)
  }
})

# Render number of countries table
output$numberOfCountriesTable <- renderTable({
  if (input$clusterButton == 0) {
    return()
  }
  numberOfCountriesTable()
})
}

# Run app
shinyApp(ui, server)

```