

Course: R Language in Computational Probability and Statistics

Lecture 1: Basics of working with the R. RStudio

Lecturer: Nataliia Kruhlova

Що таке R і чому саме він потрібен аналітикам даних?

З поширенням ПК та доступу до Інтернету ми стикаємося з надзвичайно великим потоком інформації. У той же час, представлення цієї інформації у зручному для сприйняття форматі стало складнішим завданням.



Рис.1. Згенероване зображення за запитом "Людина сидить за комп'ютером. Навколо цифри та інформація" в системі ШІ Bing, Січень, 2024, [1]. Джерело: bing.com.

Перед користувачами постає складна задача вибору програмного забезпечення для аналізу даних, оскільки існує безліч прикладних програм, що використовуються у цій сфері.



Рис. 2. Критерії вибору програмного забезпечення, <https://www.canva.com/> [3]

На рис. 3 відображено деякі комерційні продукти для аналізу даних.

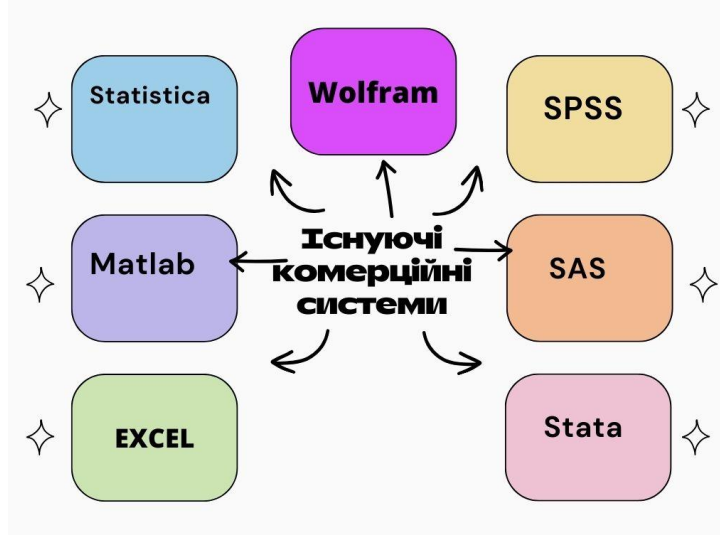


Рис. 3. Деякі комерційні програмні засоби, <https://www.canva.com/> [3]

Програмне середовище **R** є прекрасною альтернативою комерційним програмам, оскільки:

- Має безкоштовний та відкритий код.
- Середовище містить велику кількість пакетів та бібліотек, які включають в себе широкий спектр функцій для обробки, візуалізації, аналізу та моделювання даних.
- **R** володіє потужними інструментами для візуалізації даних як, наприклад, пакет **ggplot2**, який дозволяє створювати вражаючі графіки та діаграми для представлення результатів.
- **R** відомий своїм потужним масивом статистичних методів та алгоритмів, що включають регресійний аналіз, кластерний аналіз, аналіз часових рядів та багато іншого.
- **R** має велику та активну спільноту користувачів, яка постійно розвиває нові пакети та функції для аналізу даних. Це забезпечує підтримку та можливість обміну знаннями серед спеціалістів з аналізу даних.
- **R** легко інтегрується з іншими мовами програмування, такими як Python чи Java, а також з різними базами даних.

З рис. 4 можна дізнатися прізвища розробників мови програмування **R**.



Рис. 4. Розробники мови програмування, <https://www.canva.com/> [3]

Створено архів CRAN (Comprehensive R archive network), де зібрані усі відомі офіційні пакети **R**.

В [2] можна знайти усі необхідні програмні засоби, пов'язані з середовищем програмування **R**, та звернутися з питаннями стосовно розв'язання задач.

Домашня сторінка сайту www.r-project.org зображена на рис. 5:

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred **CRAN mirror**.

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 4.3.3 (Angel Food Cake)** has been released on 2024-02-29.
- **Registration for useR! 2024** has opened with early bird deadline March 31 2024.
- **R version 4.3.2 (Eye Holes)** has been released on 2023-10-31.
- **R version 4.2.3 (Shortstop Beagle)** has been released on 2023-03-15.
- You can support the R Foundation with a renewable subscription as a [supporting member](#).

News via Mastodon

[Home]

Download
CRAN

R Project
About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Get Involved: Contributing
Developer Pages
R Blog

Рис. 5. Сайт www.r-project.org

З прикладами класів задач та методів, що реалізовані в середовищі **R**, можна ознайомитись з рис. 6.

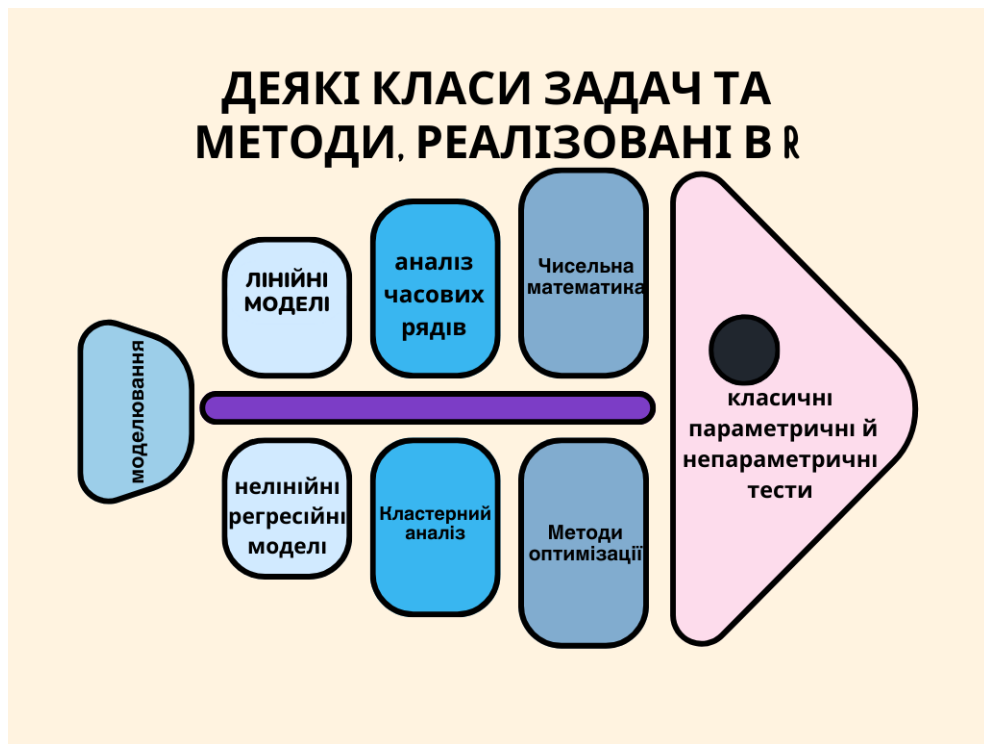


Рис. 6. Приклади класів задач, реалізованих в R, <https://www.canva.com/> [3]

RStudio

RStudio - це інтегроване середовище розробки, яке дозволяє працювати з мовою програмування R за допомогою графічного інтерфейсу. Хоча основна робота в R відбувається через командний рядок, наявність **RStudio** значно полегшує використання R, особливо для початківців. **RStudio** надає зручний інтерфейс для написання коду, виконання команд, візуалізації даних та інших завдань. На рис. 7 зображено інтерфейс цього середовища розробки.

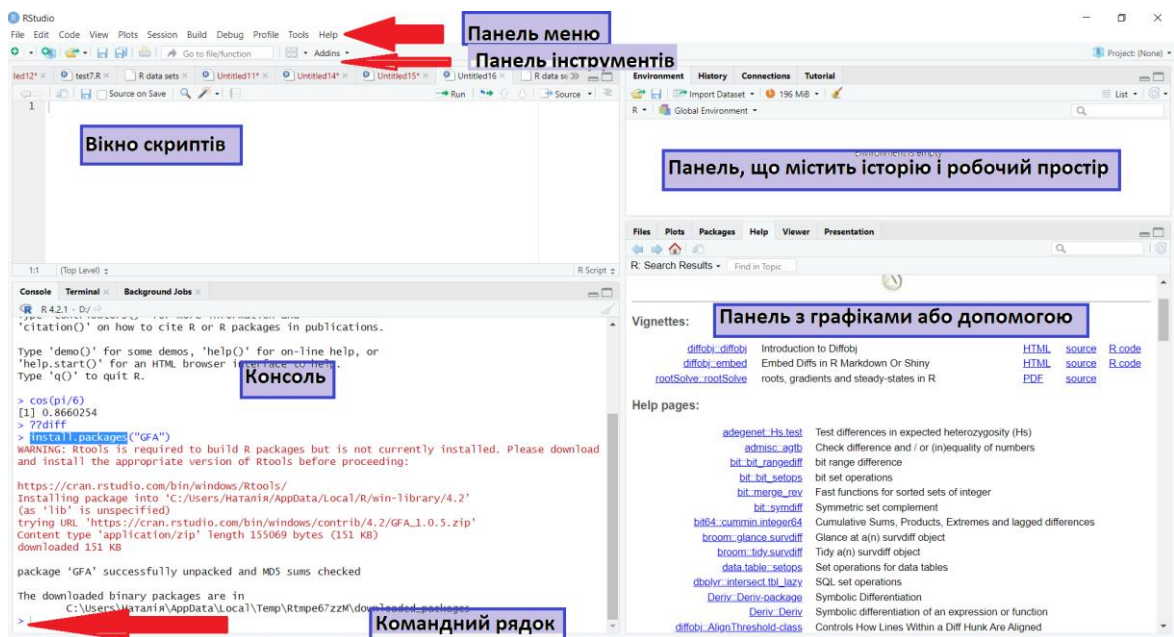


Рис. 7. Інтерфейс RStudio

Консолі RStudio властиві деякі опції, що полегшують програмування на R.

Наприклад, коли користувач вводить певний набір символів, система пропонує свої варіанти продовження команди.

Гарячі клавіші **Ctrl + ↑** та **Ctrl + ↓** викликають попередню та наступну відповідно команду. У таблиці 1 наведено деякі опції пунктів меню.

Табл.1

Дія	Пункт меню або гарячі клавіші
Перегляд всіх гарячих клавіш	HELP – >Keyboard Shortcuts Help
Перегляд гарячих клавіш задачі	Code->Потрібний пункт меню View-> Потрібний пункт меню
знайти та/або замінити частину тексту	Edit -> Find and Replace (Ctrl + F)
коментування/зняття коментування з фрагменту коду	виділити фрагмент, натиснути пункт меню Comment/Uncomment Lines

Консоль

Коли запускається **R** або **RStudio**, у верхній частині консолі виводиться інформація про встановлену версію мови **R**, інструкції щодо отримання від середовища допомоги. Впродовж сеансу у консолі відображається вся послідовність виконаних команд та результатів. Можна безпосередньо писати команди та виконувати їх.

Всі команди у консолі вводяться після символу **>**. Наприклад, введемо `cos(pi/6)` і натиснемо **Enter**. Одразу у консолі буде відобразитися результат:

```
> cos(pi/6)
[1] 0.8660254
> |
```

Результати виконання команди виводяться безпосередньо після інструкції, після чого у консолі з'являється знак $>$, що означає, що система очікує введення нової команди.

За допомогою команди **Open File** можна завантажити скрипт, створений раніше. Скрипти мають стандартне розширення *.r.

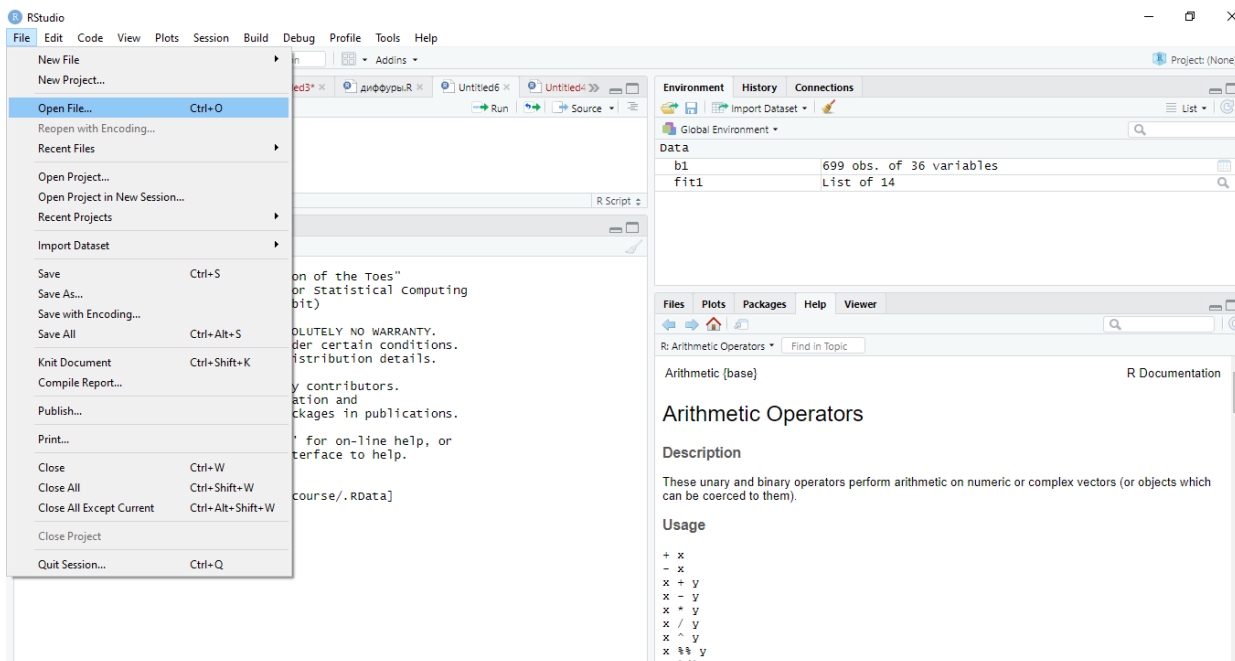


Рис. 8. Робота з меню RStudio

На рис. 8 проілюстровано роботу меню RStudio

Для виконання команди в скрипті в блоку ноті середовища R потрібно натиснути клавішу **F5** після кожної команди, в консолі – **Enter**, в **RStudio** - клавішу **Run**.

Щоб зберегти скрипт, потрібно:

- а) вибрати скрипт;
- б) або натиснути клавішу **“Save script”** на панелі інструментів, або вибрати пункт меню **“File->Save as...”** (**“File->Save”** для файлу, який вже зберігали під певною назвою).

Якщо команди у файлі розміщені у окремих рядочках, то між ними не потрібно ставити ніяких розділових знаків. Команди розділяють крапкою з комою, якщо вони розміщуються в одному рядку.

Якщо команда занадто довга і не вміщається в один рядок, її можна розділити на декілька рядків. Незавершена команда при переході в інший

рядок автоматично продовжується у наступному рядку із символом продовження **+**. **R** може розпізнати, що команда не закінчена, тому деякі синтаксичні помилки, як-от відсутність дужок, можуть призвести до того, що **R** сприйматиме їх як незакінчені команди. У такому випадку **R** виведе **+** на початку наступного рядка та перейде у режим очікування. Щоб перейти у режим введення нової команди без продовження аналізу попередньої, можна натиснути **Esc**.

Все, що йде після символу **#** до кінця рядка, вважається середовищем **R** за коментар.

Історія та робочий простір зберігаються у робочому каталозі програми (**Working Directory**). Історія має розширення ***.RHistory**, робочий простір - ***.RData**.

Збереження робочої пам'яті: пункт меню **"File->Save Workspace"**.

Збереження історії: пункт меню **"File->Save History"**.

Дізнатися робочий каталог програми можна командою **getwd()**.

```
> getwd()  
[1] "C:/Диск/R_course"
```

Робочий каталог може бути змінено командою **setwd()**; при цьому шлях вказується за допомогою символу слеш (/) або подвійний бекслеш (\\).

```
> setwd("D:/R_course")
```

Демонстрації

Можливості мови **R** демонструє команда **demo**.

```
>demo (graphics)# Демонстрація пакету графічного виводу  
>demo (persp)  
>demo (image) # інші приклади.
```

На рис. 9 зображено приклад графічного зображення, що виводиться при

запуску демонстрацій.

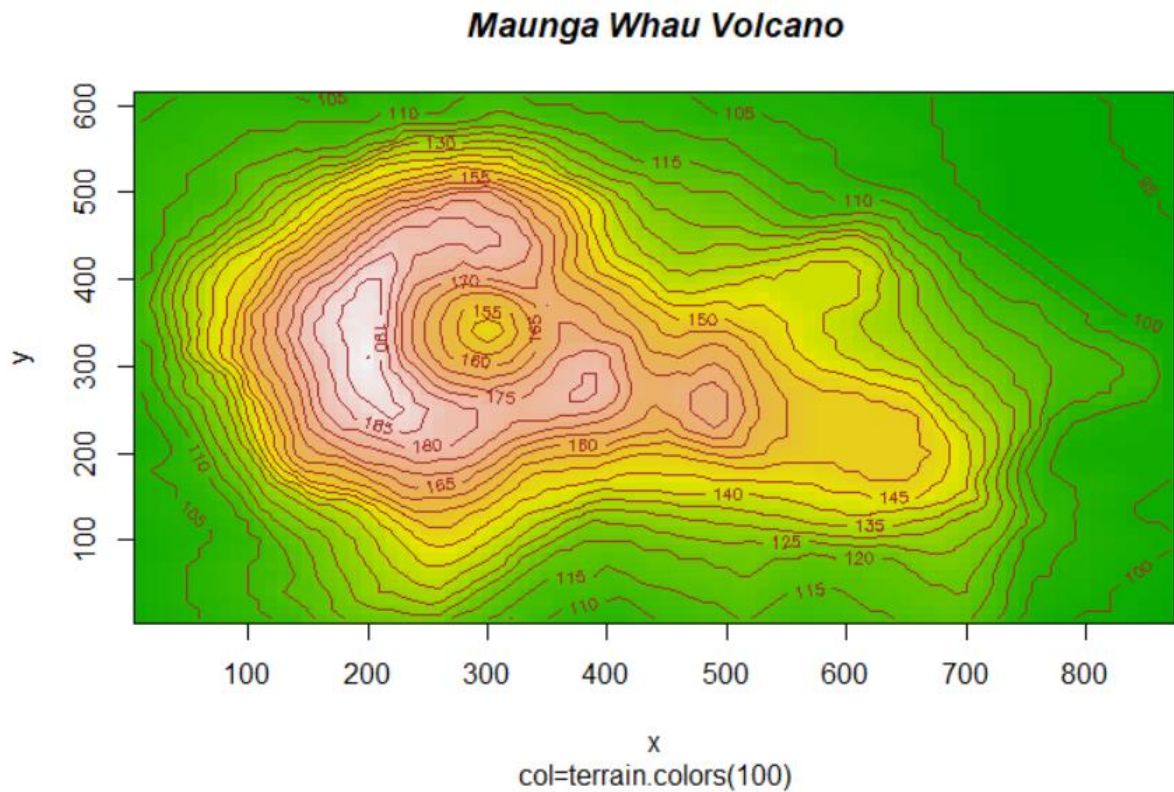


Рис.9. Приклад графічного зображення згенерованого в RStudio

Графіки відображаються у вікні **Graph**. Щоб попередній графік не видалявся з вікна, можна відмітити у верхньому меню **History -> Recording**.

Коли зміст робочого каталогу зберігається у файлі, зміст графічного вікна теж зберігається.

Отримання довідки

Існує декілька способів отримати допомогу з використанням R. Один з них - перейти до верхнього меню та вибрати шлях: **Help -> R functions (some text)**. Після цього відкриється вікно запиту, куди потрібно ввести ім'я функції, для якої ви шукаєте допомогу. Інші можливості отримати допомогу включають безпосереднє застосування консолі. Наприклад, інформацію про функцію "diff" можна знайти, виконавши одну з наступних команд.

```
?diff  
> help(diff)
```

Якщо користувач шукає функцію, назва якої містить деяке словосполучення, можна ввести словосполучення після двох питальних знаків. Наприклад,

```
> ??diff
```

Після чого виводиться перелік всіх функцій, з іменами, що містять послідовність символів " diff". На рис. 10 наведено елемент з вікна допомоги **RStudio** після виконання цієї команди.

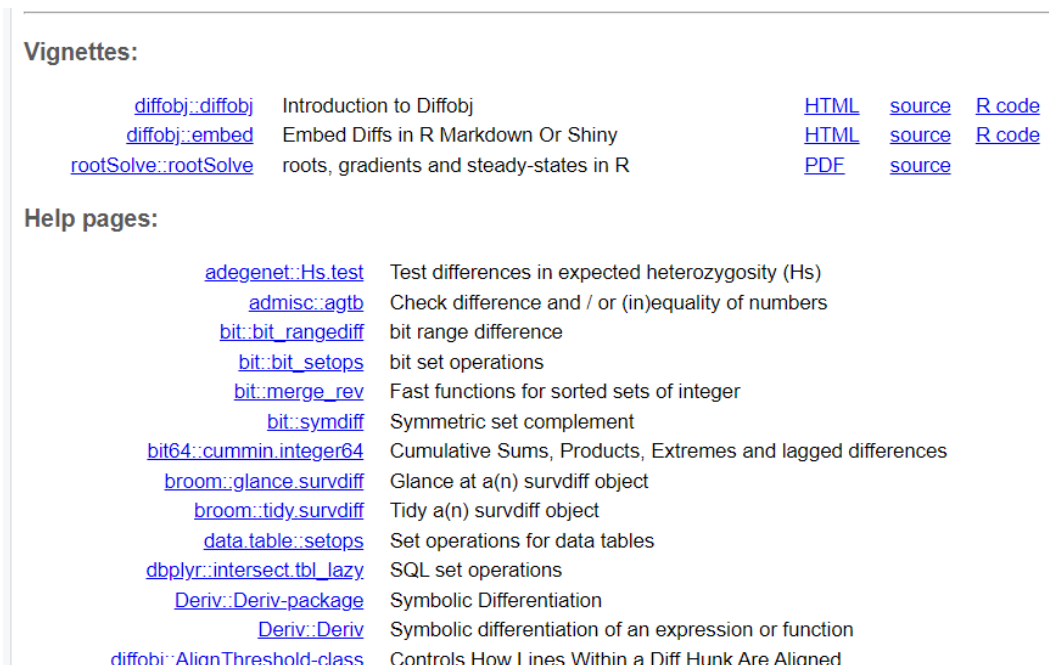


Рис. 10. Список функцій, що містять символічну послідовність " diff", створено в RStudio

Або можна так:

```
> apropos("max")
[1] "cummax"      "max"         "max.col"     "mem.maxNSize" "mem.maxVSize"
[6] "pmax"        "pmax.int"    "promax"      "varimax"      "which.max"
```

Одержати загальну інформацію середовища можна за допомогою команди:

```
> help.start()
```

На рис. 11 відображено інтерфейс **RStudio** після виконання останньої команди. У вікні допомоги середовища розробки виводиться домашня сторінка допомоги R.

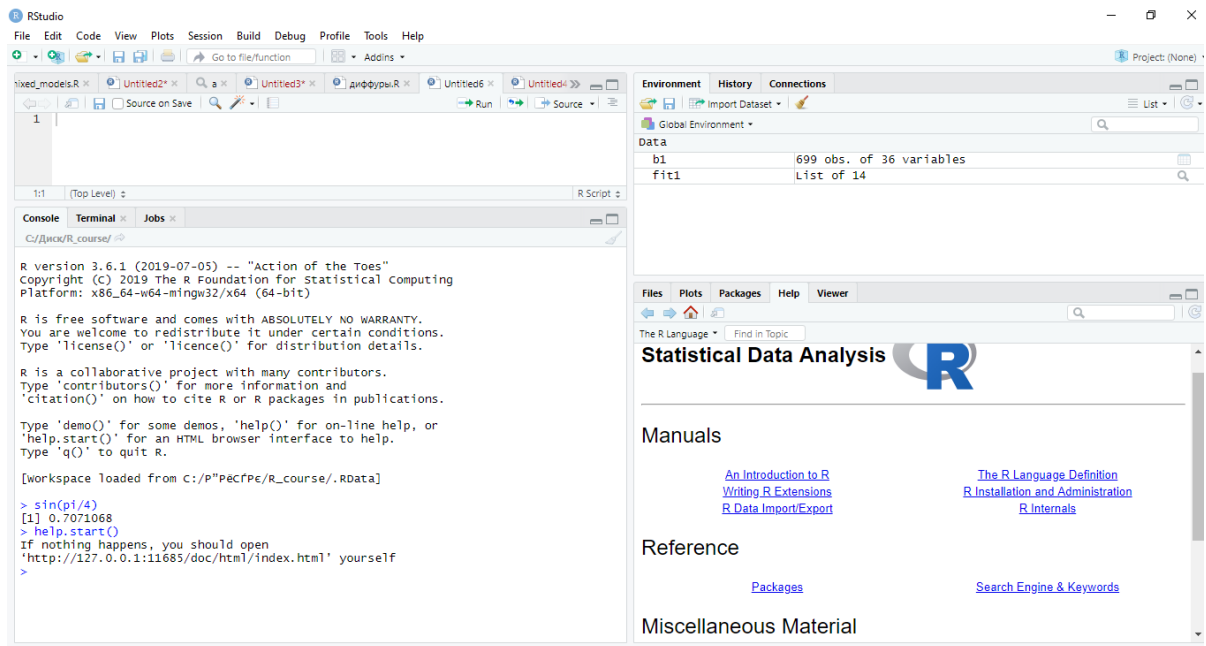


Рис. 11. Стартова сторінка допомоги R

Пакети

Множину функцій, які пов'язані спільною тематикою, та набори даних, об'єднують у пакети (бібліотеки).

Під час встановлення програми **R** деякі з основних пакетів завантажуються автоматично.

Якщо пакет, який вам потрібен, не було завантажено на комп'ютер, то його можна встановити з інтернет-архіву, використовуючи пункт головного меню **Packages->Installpackages**.

Спочатку програма пропонує обрати інтернет-архів для завантаження. За умовчанням пропонується варіант 0-cloud. Після цього вибирається необхідний пакет. Всі пакети, які потрібні для коректної роботи даного пакету і відсутні у комп'ютері, будуть також автоматично завантажуватись. Після інсталяції вони зберігаються у комп'ютері, але при кожному новому запуску програми ці пакети знову потрібно підключати, якщо ми з ними будемо працювати. Підключення пакету виконується функцією **library**.

```
> library(ggplot2)
```

У прикладі ми підключили пакет **ggplot2**, який використовується для побудови графічних об'єктів.

```
> library()
```

Ця інструкція надає перелік усіх завантажених пакетів разом з коротким поясненням їх функцій (в мінімальному варіанті цього переліку може бути кілька десятків пакетів). Усі об'єкти, що були створені або

завантажені протягом сеансу роботи з **R** і не були видалені спеціальною командою, зберігаються у робочому просторі. Під час завершення сеансу R запитує, чи потрібно зберегти робочий простір на диску? Якщо користувач погоджується, то цей робочий простір відновлюється на початку наступного сеансу. Проте варто користуватись цією можливістю дуже обережно, оскільки "старі" об'єкти можуть спотворити роботу R у новому сеансі.

За допомогою **Help-> R help->Packages** виводиться перелік основних пакетів за алфавітом.

Help -> R help -> Task Views відображається інформація про класифікацію пакетів по відношенню до типу завдання. Далі слідує перелік тем з іменами пакетів для роботи з відповідними задачами.

Можна аналогічні дії виконати у консолі.

```
> available.packages()#перелік доступних пакетів.  
>length(available.packages())# виводиться кількість пакетів на сервері.  
>install.packages ("package1", "package2")# вказані пакети завантажаться і встановляються на комп'ютері.
```

```
> data() # список наборів даних, підключених до шляху пошуку.  
>install.packages ("package",dependensies=T)#вказаний пакет завантажиться з повним набором пакетів, які необхідні для коректної роботи даного пакету;  
>search()# список пакетів, прикріплених до шляху пошуку.  
>library (help="ggplot2")# інформація про пакет ggplot2: список функцій, що містяться у вказаному статистичному пакеті, автори, назва, версія R, яка може працювати з цим пакетом.
```

Список джерел

1. Microsoft Bing(2024). Зображення за запитом " Людина сидить за комп'ютером. Навколо цифри та інформація ". <https://www.bing.com> [22.01.2024]
2. The R Project for Statistical Computing. <https://www.r-project.org>
3. Canva. <https://www.canva.com/>