



(<https://www.bigdatauniversity.com>)

R BASICS | R. Базовий рівень

Welcome! | Вітаємо!

By the end of this notebook, you will have learned the basics of R! / До кінця цього зошита, ви зможете опанувати базовий рівень мови програмування R

Table of Contents / Зміст

- [About the Dataset / Про набір даних](#)
- [Simple Math in R / Проста математика в R](#)
- [Variables in R / Змінні в R](#)
- [Vectors in R / Вектори в R](#)
- [Strings in R / Текстові ряди в R](#)

Estimated Time Needed: **15 min** / Очікувана тривалість: **15 хв**

About the Dataset / Про набір даних

Which movie should you watch next? / Який фільм подивитись наступного разу?

Let's say each of your friends tells you their favorite movies. You do some research on the movies and put it all into a table. Now you can begin exploring the dataset, and asking questions about the movies. For example, you can check if movies from some certain genres tend to get better ratings. You can check how the production cost for movies changes across years, and much more. /

Уявімо, що ви попросили своїх друзів назвати їхні улюблені фільми. Ви збираєте інформацію про ці фільми і заповнюєте таблицю. Тепер ви можете почати вивчення набору даних і задавати питання про ці фільми. Наприклад, ви можете перевірити, чи схильні фільми певного жанру отримувати вищі рейтинги. Ви також можете дізнатися, як вартість виробництва фільмів змінюється протягом років та багато іншого.

Movies dataset / Набір даних про фільми

The table gathered includes one row for each movie, with several columns for each movie characteristic:

У створеній таблиці кожному рядкові відповідає один фільм з кількома колонками для характеристик відповідного фільму:

- **name** - Name of the movie / **name** - Назва фільму
- **year** - Year the movie was released / **year** - Рік релізу фільму
- **length_min** - Length of the movie (minutes) / **length_min** - Тривалість фільму (хвилини)
- **genre** - Genre of the movie / **genre** - Жанр фільму
- **average_rating** - Average rating on [IMDB \(http://www.imdb.com/\)](http://www.imdb.com/) / **average_rating** - Середній рейтинг на сайті [IMDB \(http://www.imdb.com/\)](http://www.imdb.com/)
- **cost_millions** - Movie's production cost (millions in USD) / **cost_millions** - Вартість фільму (млн доларів США)
- **foreign** - Is the movie foreign (1) or domestic (0)? / **foreign** - Фільм є зарубіжним (1) чи вітчизняним (0)?
- **age_restriction** - Age restriction for the movie / **age_restriction** - Вікові обмеження для перегляду фільму

name	year	length_min	genre	average_rating	cost_millions	foreign	age_restriction
Toy Story	1995	81	Animation	8.3	30	0	0
Akira	1998	125	Animation	8.1	10.4	1	14
The Breakfast Club	1985	97	Drama	7.9	1	0	14
The Artist	2011	100	Romance	8	15	1	12
Modern Times	1936	87	Comedy	8.6	1.5	0	10
Fight Club	1999	139	Drama	8.9	63	0	18
City of God	2002	130	Crime	8.7	3.3	1	18
The Untouchables	1987	119	Drama	7.9	25	0	14
Star Wars	1977	121	Action	8.7	11	0	10
American Beauty	1999	122	Drama	8.4	15	0	14
Room	2015	118	Drama	8.3	13	1	14
Dr. Strangelove	1964	94	Comedy	8.5	1.8	1	10
The Ring	1998	95	Horror	7.3	1.2	1	18
Monty Python and the Holy Grail	1975	91	Comedy	8.3	0.4	1	18
High School Musical	2006	98	Comedy	5.2	4.2	0	0
Shaun of the Dead	2004	99	Horror	8	6.1	1	18
Taxi Driver	1976	113	Crime	8.3	1.3	1	14
The Shawshank Redemption	1994	142	Crime	9.3	25	0	16
Interstellar	2014	169	Adventure	8.6	165	0	10
Casino	1995	178	Biography	8.2	50	0	18
The Goodfellas	1990	145	Biography	8.7	25	0	14
Blue is the Warmest Colour	2013	179	Romance	7.8	4.5	1	18
Black Swan	2010	108	Thriller	8	13	0	16
Back to the Future	1985	116	Sci-fi	8.5	19	0	0
The Wave	2008	107	Thriller	7.6	5.5	1	16
Whiplash	2014	106	Drama	8.5	3.3	1	12
The Grand Hotel Budapest	2014	100	Crime	8.1	25.5	0	14
Jumanji	1995	104	Fantasy	6.9	65	0	12
The Eternal Sunshine of the Spotless Mind	2004	108	Drama	8.3	20	0	14
Chicago	2002	113	Comedy	7.2	45	0	12

We can use R to help us explore the dataset / Ми можемо використовувати R для вивчення набору даних

But to begin, we'll need to start from the basics, so let's get started! /

Але для старту потрібно розпочати з основ, тому гайда до роботи!

Simple Math in R / Проста математика в R

Let's say you want to watch *Fight Club* and *Star Wars: Episode IV (1977)*, back-to-back. Do you have enough time to **watch both movies in 4 hours**? Let's try using simple math in R. /

Скажімо, ви хочете подивитися *Бійцівський клуб* і *Зоряні війни: IV (1977)*. Чи буде у вас достатньо часу, щоб **подивитися обидва фільми за 4 години**? Для пошуку відповіді спробуємо використати просту математику в R.

What is the **total movie length** for *Fight Club* and *Star Wars: IV (1977)*? /

Яка **загальна тривалість фільмів** *Бійцівський клуб* і *Зоряні війни: IV (1977)*?

- **Fight Club**: 139 min / **Бійцівський клуб**: 139 хв
- **Star Wars: IV**: 121 min / **Зоряні війни: IV**: 121 хв

Tip: To run the grey code cell below, click on it, and press Shift + Enter. /

Лайфхак: Щоб запустити зелену ячейку з кодом нижче, клікніть по ній і натисніть Shift + Enter.

In []:

```
139 + 121
```

Great! You've determined that the total number of movie play time is **260 min.** /

Чудово! Ви обчислили, що загальна тривалість фільмів складає **260 хв.**

What is 260 min in hours? /

Скільки буде 260 хв у годинах?

In []:

```
260 / 60
```

Well, it looks like it's **over 4 hours**, which means you can't watch *Fight Club* and *Star Wars (1977)* back-to-back without missing out on some of the action! /

Окей, це виглядає як **понад 4 години**, що означає одне: ви не зможете переглянути *Бійцівський клуб* і *Зоряні війни: IV (1977)* один за одним, не пропустивши окремих епізодів!

[Tip] Simple math in R

You can do a variety of mathematical operations in R including:

- addition: **2 + 2**
- subtraction: **5 - 2**
- multiplication: **3 * 2**
- division: **4 / 2**
- exponentiation: **4 ^ 2**

[Лайфхак] Проста математика в R

Ви можете виконувати багато математичних дій в R, включаючи:

- додавання: **2 + 2**
- віднімання: **5 - 2**
- множення: **3 * 2**
- ділення: **4 / 2**
- піднесення до степені: **4 ^ 2**

Variables in R / Змінні в R

We can also **store** our output in **variables**, so we can use them later on. For example: /

Ми також можемо **зберегти** наш результат, прив'язавши його до **змінних** для майбутнього використання. Наприклад:

In []:

```
x <- 139 + 121
```

To return the value of **x**, we can simply run the variable as a command: /

Для виводу значення змінної **x**, можна просто запустити змінну як команду:

In []:

```
x
```

We can also perform operations on **x** and save the result to a **new variable**: /

Також можна виконувати операції з **x** і зберігати отриманий результат у **новій змінній**:

In []:

```
y <- x / 60  
y
```

If we save something to an **existing variable**, it will **overwrite** the previous value: /

Якщо ми збережемо щось в уже **існуючій змінній**, він буде **перезаписаний** на місце попереднього значення:

In []:

```
x <- x/60  
x
```

It's good practice to use **meaningful variable names**, so you don't have to keep track of what variable is what: /

Хорошим тоном вважається використання **зрозумілих назв змінних**, щоб не доводилося постійно відслідковувати що значить та чи інша змінна:

In []:

```
total <- 139 + 121  
total
```

In []:

```
total_hr <- total / 60  
total_hr
```

You can put this all into a single expression, but remember to use **round brackets** to add together the movie lengths first, before dividing by 60. /

Ви можете описати ці дії у вигляді єдиного виразу, але не забувайте використовувати **круглі дужки**, щоб спочатку додати тривалість фільмів перед тим, як поділити її на 60.

In []:

```
total_hr <- (139 + 121) / 60  
total_hr
```

[Tip] Variables in R

As you just learned, you can use **variables** to store values for repeated use. Here are some more **characteristics of variables in R**:

- variables store the output of a block of code
- variables are typically assigned using **<-**, but can also be assigned using **=**, as in **x <- 1** or **x = 1**
- once created, variables can be removed from memory using **rm(my_variable)**

Vectors in R

What if we want to know the **movie length in *hours***, not minutes, for *Toy Story* and for *Akira*?

- **Toy Story (1995)**: 81 min
- **Akira (1998)**: 125 min

In []:

```
c(81, 125) / 60
```

As you see above, we've applied a single math operation to both of the items in `c(81, 125)`. You can even assign `c(81, 125)` to a variable before performing an operation.

In []:

```
ratings <- c(81, 125)
ratings / 60
```

What we just did was create **vectors**, using the combine function `c()`. The `c()` function takes multiple items, then combines them into a **vector**.

It's important to understand that **vectors** are used everywhere in R, and vectors are easy to use.

In []:

```
c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
c(1:50)
```

In []:

```
c(10:1) # 10 to 1
```

[Tip] # Comments

Did you notice the **comment** after the `c(10:1)` above? Comments are very useful in describing your code. You can create your own comments by using the `#` symbol and writing your comment after it. R will interpret it as a comment, not as code.

Strings in R / Текстові ряди в R

R isn't just about numbers -- we can also have strings too. For example: /

R - це не тільки числа -- ми тут можемо також оперувати і текстовими рядками. Наприклад:

In []:

```
movie <- "Toy Story"
movie
```

In R, you can identify **character strings** when they are wrapped with **matching double (") or single (')** quotes. /

В мові R, ви можете визначити **текстові рядки** вставивши їх у **подвійні (") або одинарні (')** лапки.

Let's create a **character vector** for the following **genres**:

- Animation
- Comedy
- Biography
- Horror
- Romance
- Sci-fi

Створимо **текстовий вектор** для таких **жанрів**:

- Мультфільм
- Комедія
- Біографія
- Фільм жахів
- Мелодрама
- Наукова фантастика

In []:

```
genres <- c("Animation", "Comedy", "Biography", "Horror", "Romance", "Sci-fi")
genres

genres <- c("Мультфільм", "Комедія", "Біографія", "Фільм жахів", "Мелодрама", "Наукова фантастика")
genres
```

Excellent! You have just completed the R basics notebook! /

Прекрасно! Ви щойно завершили зошит "R. Базовий рівень" !

[Tip] Saving the Notebook

Your notebook saves automatically every two minutes. You can manually save by going to **File > Save and Checkpoint**. You can come back to this notebook anytime by clicking this notebook under the **"Recent Notebooks"** list on the right-hand side.

[Tip] Notebook Features

Did you know there are other **notebook options**? Click on the > symbol to the left of the notebook:

▼ Recent Notebooks

▼ Notebook 1.ipynb ⚡

📄

 Rename • Clone • Download • Delete

🔗

 Submit to Spark Cluster

🏷

 untagged ✎

⚡

 Active • Stop

🔗

 Share

←

 Insert Path • Insert Link • Insert Import

➤ Notebook 2.ipynb ⚡

➤ Notebook 3.ipynb ⚡

About the Author: | Про автора:

Hi! It's Helly Patel (<https://ca.linkedin.com/in/helly-patel-90344750>), the author of this notebook. I hope you found R easy to learn! There's lots more to learn about R but you're well on your way. Feel free to connect with me if you have any questions.

Привіт! Це Helly Patel (<https://ca.linkedin.com/in/helly-patel-90344750>), автор цього зошиту. Я сподіваюся, що ви побачили наскільки легко вивчати R! Є ще багато чого дізнатись про R, але ви вже рухаєтесь у вірному напрямку. Якщо у вас виникли запитання, можете без проблем зв'язатися зі мною.

About the Translator: | Про перекладача

Hello! It's Roman Kornyliuk (<https://www.linkedin.com/in/kornelli>), the translator of this notebook in Ukrainian. Feel free to connect with me if you have any questions too.

Привіт! Мене звуть Roman Kornyliuk (<https://www.linkedin.com/in/kornelli>), перекладач цього зошиту на українську мову. Без вагань пишіть мені також, якщо у вас виникли запитання з приводу R.

Copyright © 2016 Big Data University (https://bigdatauniversity.com/?utm_source=bducopyrightlink&utm_medium=dswb&utm_campaign=bdu).

Copyright © 2018 Roman Kornyliuk (<https://www.linkedin.com/in/kornelli>).

This notebook and its source code are released under the terms of the MIT License (<https://bigdatauniversity.com/mit-license/>).