Лабораторна робота 1 Завантаження даних та їх візуалізація на мові R

Хід роботи

1. Ознайомтесь з основними меню програм R та RStudio

За допомогою команд library() та dir(.libPaths()) перегляньте, які пакети встановлені у програмі

2. У разі необхідності встановіть бібліотеки *dplyr*, *readr*, *ggplot2* та інші, бібліотеки, які необхідні для роботи названих бібліотек.

Встановлення бібліотеки можна зробити у меню *Inslall package(s)*, обравши необхідний репозиторий (за замовчуванням *CRAN*).

3. Підключіть вказані бібліотеки за допомогою функції library(), вказавши у дужках назву необхідної бібліотеки.

4. Встановіть бібліотеку *wbstats*, яка дозволяє завантажувати дані Світового банку.

5. Для прикладу поглянемо на зміну частки населення України, яка використовує власний Інтернет (*Individuals using the Internet (% of population)*). Знайдемо сам показник, а потім, дізнавшись його код, висавимо інші параметри – країну та період часу, який нас цікавить.

```
wbsearch(pattern = "Individuals using the Internet", fields="indicator") #
Шукаємо точу назву (код) показника
```

indicatorID indicator 8540 IT.NET.USER.ZS Individuals using the Internet (% of population) Internet <- wb(country = c("UA"), indicator = c("IT.NET.USER.ZS"), startdate = 2000, enddate = 2023) View(Internet)

6. Ознайомтесь зі структурою об'єкту Internet, використовучи команду str().

str(Internet)
'data.frame': 22 obs. of 7 variables:
\$ iso3c : chr "UKR" "UKR" "UKR" "UKR" ...
\$ date : chr "2021" "2020" "2019" "2018" ...
\$ value : num 79.2 75 70.1 62.6 58.9 ...
\$ indicatorID: chr "IT.NET.USER.ZS" "IT.NET.USER.ZS" "IT.NET.USER.ZS"
"IT.NET.USER.ZS" ...
\$ indicator : chr "Individuals using the Internet (% of population)"
"Individuals using the Internet (% of population)" "Individuals using the
Internet (% of population)" "Individuals using the Internet (% of
population)" ...
\$ iso2c : chr "UA" "UA" "UA" "UA" ...
\$ country : chr "Ukraine" "Ukraine" "Ukraine" "Ukraine" ...

7. Підрахуйте кількість стовпців (команда ncol()), рядків (команда nrow()) та виведіть назви стовпців (команда colnames())

ncol(Internet)
[1] 7

nrow(Internet) [1] 22 colnames(Internet) [1] "iso3c" "date" "value" "indicatorID" "indicator" "iso2c" [7] "country"

8. Виведіть на екран перші 6 рядків файлу (команда head())

```
head(Internet)
iso3c date
                value
                          indicatorID
indicator iso2c
    UKR 2021 79.21829 IT.NET.USER.ZS Individuals using the Internet (% of
population)
               UA
    UKR 2020 75.03791 IT.NET.USER.ZS Individuals using the Internet (% of
population)
               UA
   UKR 2019 70.12484 IT.NET.USER.ZS Individuals using the Internet (% of
population)
               UA
    UKR 2018 62.55316 IT.NET.USER.ZS Individuals using the Internet (% of
5
population)
               UA
    UKR 2017 58.88948 IT.NET.USER.ZS Individuals using the Internet (% of
6
population)
               UA
    UKR 2016 53.00097 IT.NET.USER.ZS Individuals using the Internet (% of
population)
               UΔ
  country
 Ukraine
3 Ukraine
4 Ukraine
5 Ukraine
6 Ukraine
7 Ukraine
```

9. Побудуйте графік динаміки частки користувачів Інтернет у форматі стовпчикової діаграми, лінійного графіка та графіку розсіювання.

Для візуалізації даних будемо використовувати бібліотеку ggplot2. Для побудови графіків використовується функція ggplot(). Після виконання коду ви побачите графік у вкладці *Plots* у нижній правій панелі RStudio.

Першим аргументом цієї функції є набір даних.

Далі ми вказуємо змінні з набору даних як параметр *aesthetic*, які будуть відображатися, наприклад по осях X та Y.

Наступним кроком ми додаємо ще один рівень (об'єднавши їх знаком +) щоб задати *geometric* об'єкт. Наприклад, для графіка posciювання це *geom_point*, для лінійного графіка *geom_line*, для стовпчикової діаграми *geom_col*.

Стовпчикова діаграма

```
ggplot(Internet, aes(x=date,y=value))+geom_col(fill="lightblue",
col="gray")+ylab('Відсоток від населення')
```



Лінійний графік

ggplot(Internet, aes(date,value, group=1))+geom_line()+ylab('Відсоток від населення')



Графік розсіювання

ggplot(Internet, aes(x=date,y=value))+geom_point()+ylab('Відсоток від населення')



Індивідуальне завдання

Для обраного соціально-економічного показника з бази даних Світового банку для обраної країни виконати п. 5-9 лабораторної роботи та оформити звіт у системі Moodle.