МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

ЗВІТ

до індивідуального завдання №1

з дисципліни «Моделі статистичного навчання»

Виконали

студенти групи ПМіМ-12:

Бордун Михайло

Зелінський Олександр

Перевірив:

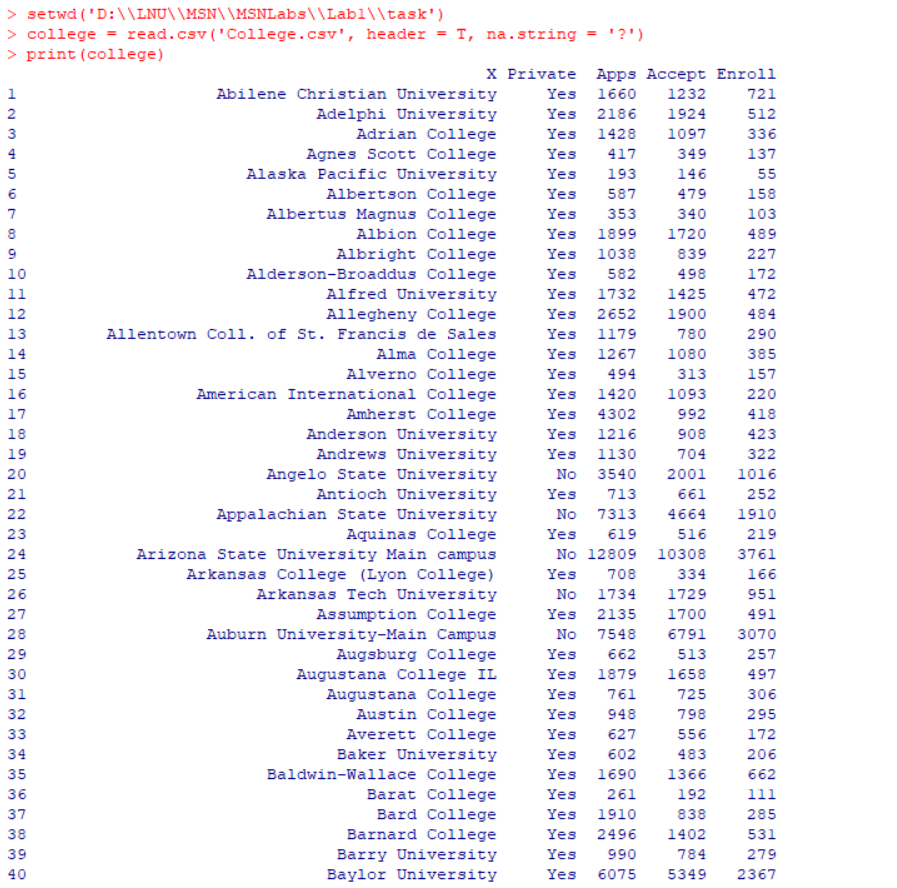
Проф. Заболоцький Т. М.

Львів – 2021

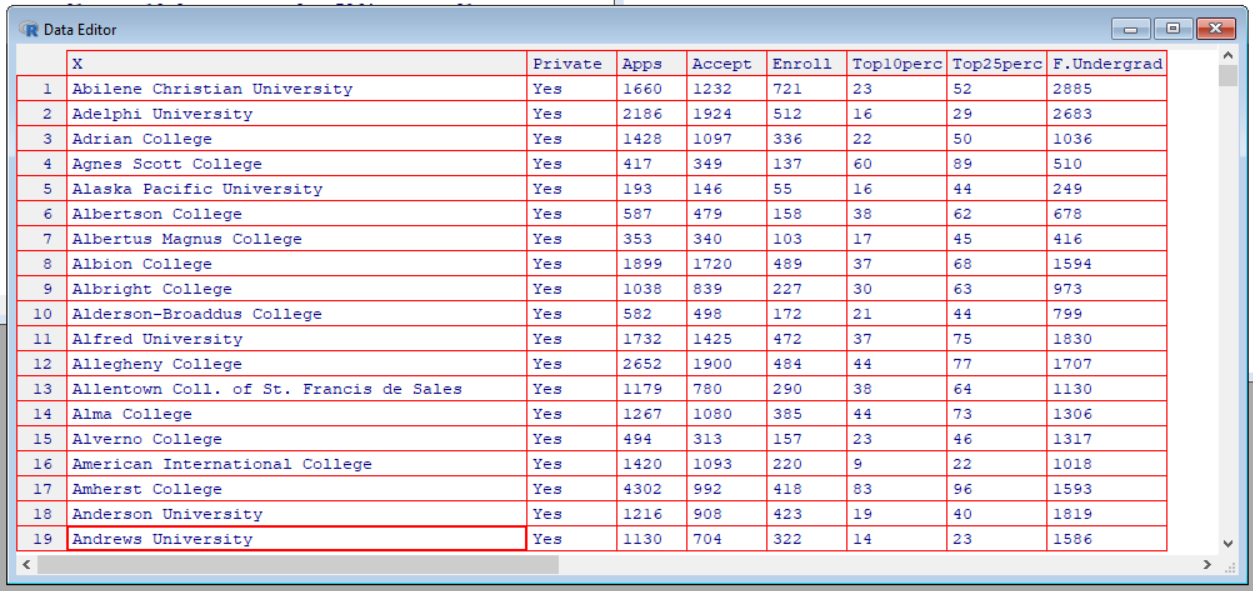
**Хід виконання**

**Пункт 1 (College)**

**1.1** Завантажені та викликані дані College.



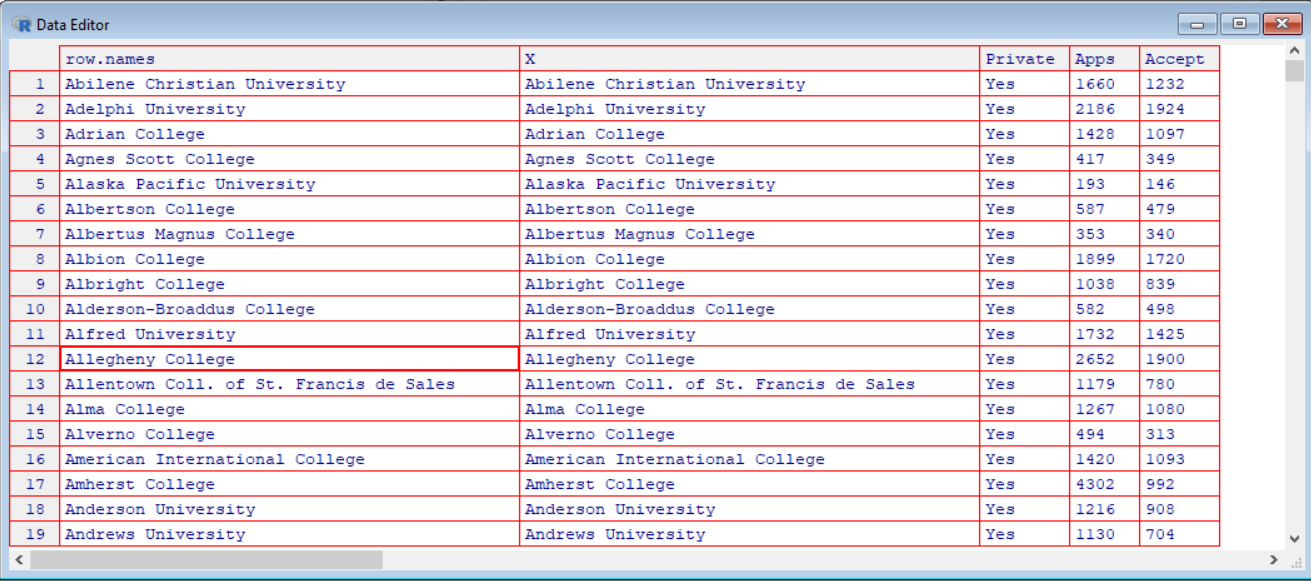
**1.2** Дані переглянуті за допомогою функції fix.



Після виконання команд

> rownames (college )=college [,1]

> fix (college )

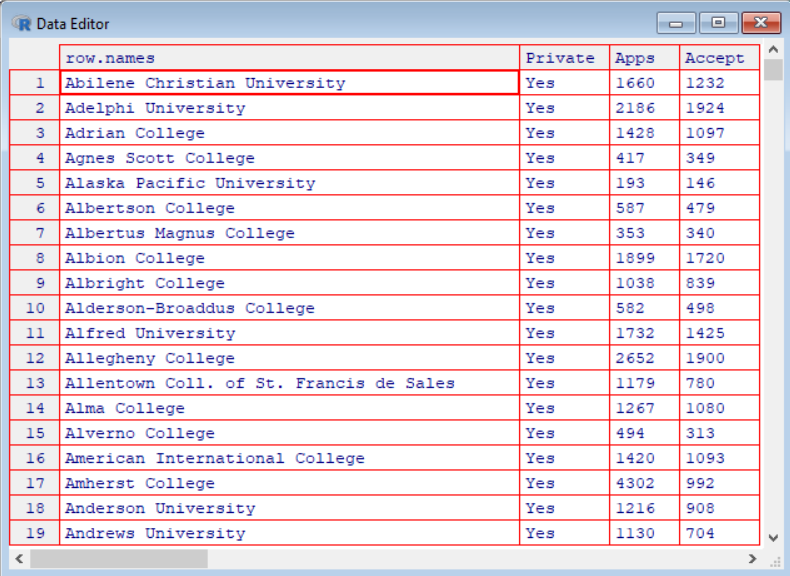


Можемо помітити, що назви факультетів здублювались в row.names, а після виконання команд

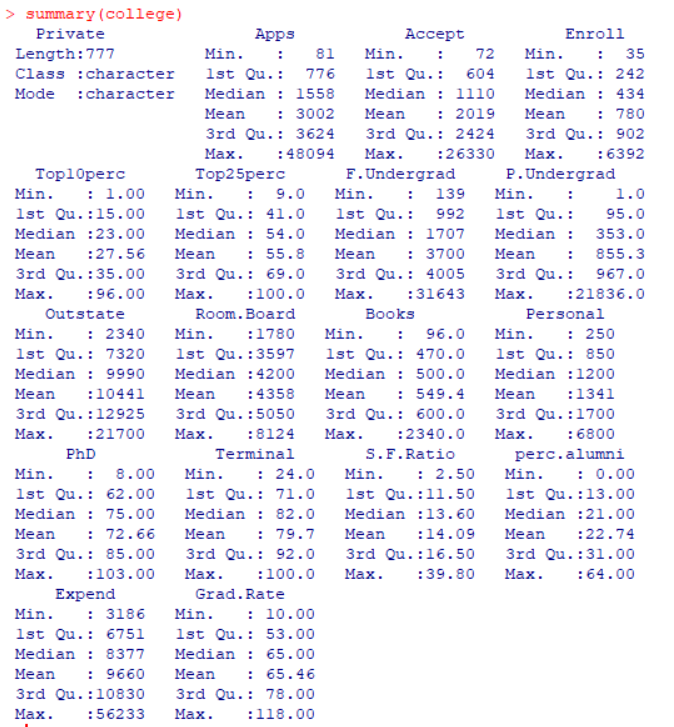
> rownames (college )=college [,1]

> fix (college )

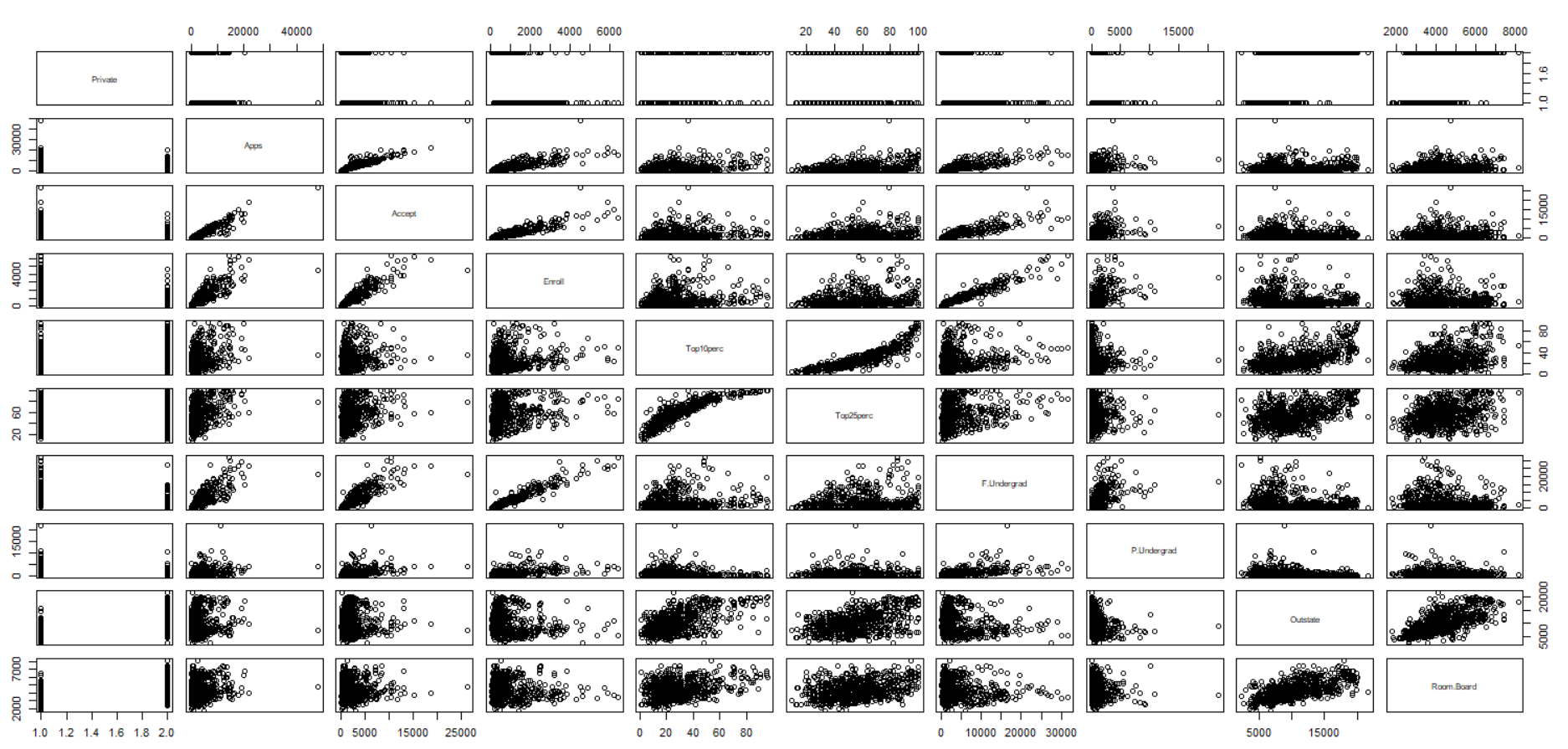
рядок з іменами університетів зник з даних і залишився лише у вигляді назв рядків (row.names).



**1.3.1** За допомогою функції summary можемо побачити деякі статистичні величини по кожному з рядків. Також можемо побачити, що всі значення числові окрім Private (приватний університет чи ні).

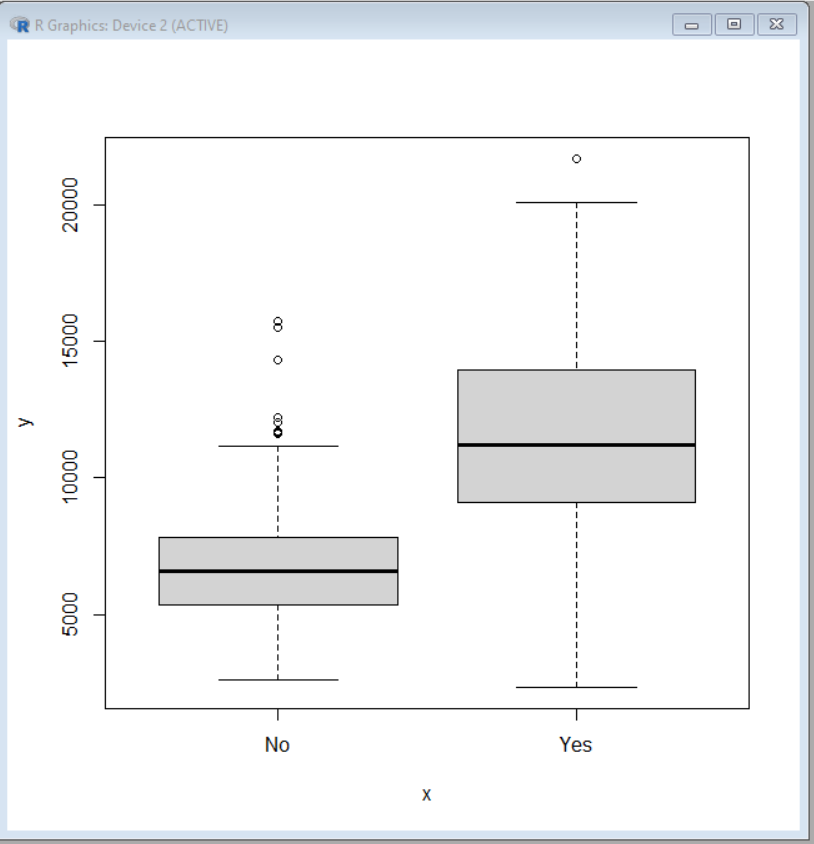


**1.3.2** Для побудови матриці графіків перших 10 стовпців використано функцію pairs(college[1:10]).



Проаналізувати пару графіків

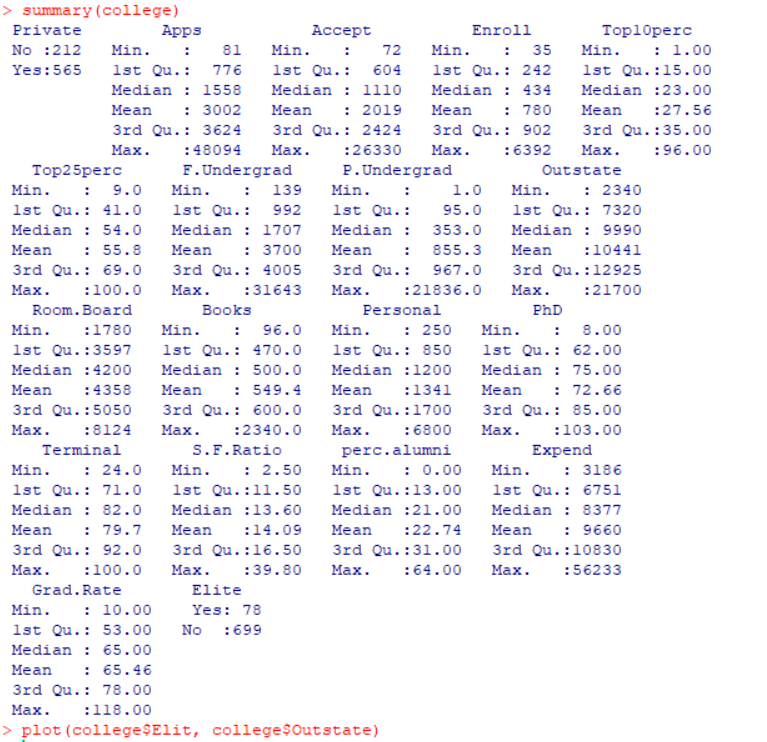
**1.3.3** Для побудови графіку залежності college$Private та college$Outstate використану функцію plot.



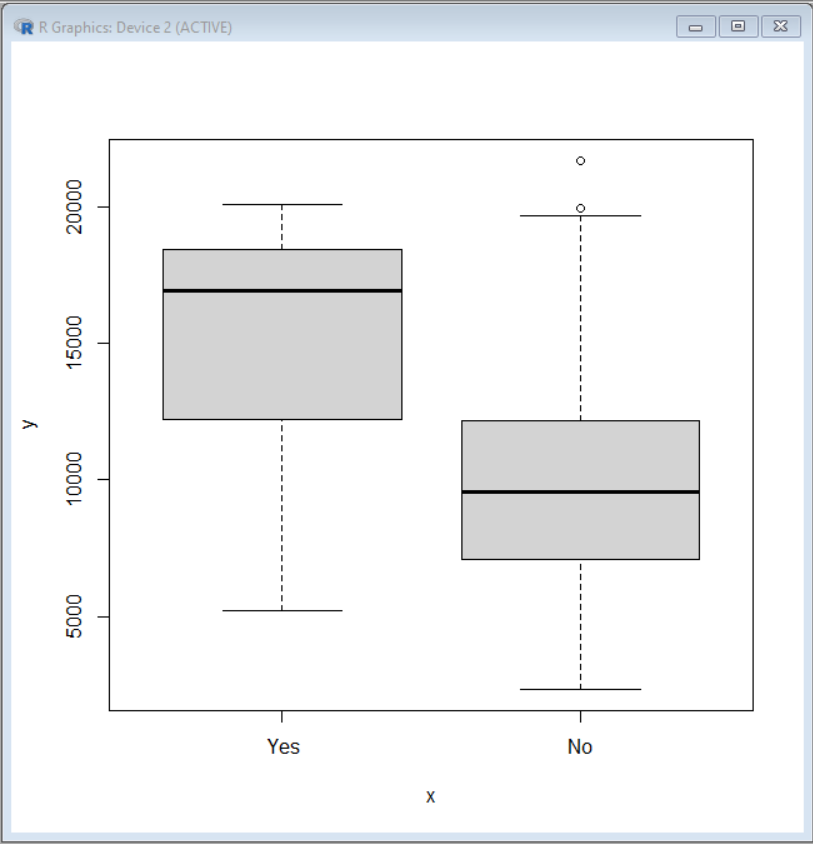
Проаналізувати графік

**1.3.4** Створено новий показник Elit використовуючи Top10perc. Поділимо всі університети на дві групи в залежності чи перевищує відсоток студентів з топ 10% шкіл 50% чи ні.

За допомогою функції summary було визначено, що елітних шкіл 78, а не елітних 699.



Також було побудовано графік залежності college$Elit віж college$Outstate.



Проаналізувати пару графік

**1.3.5** За допомогою функції hist було побудовано 4 гістограми: Витрати на навчання на одного студента, вартість навчання іноземних студентів, відсоток факультетів з PhD та Відсоток професорсько-викладацького складу.

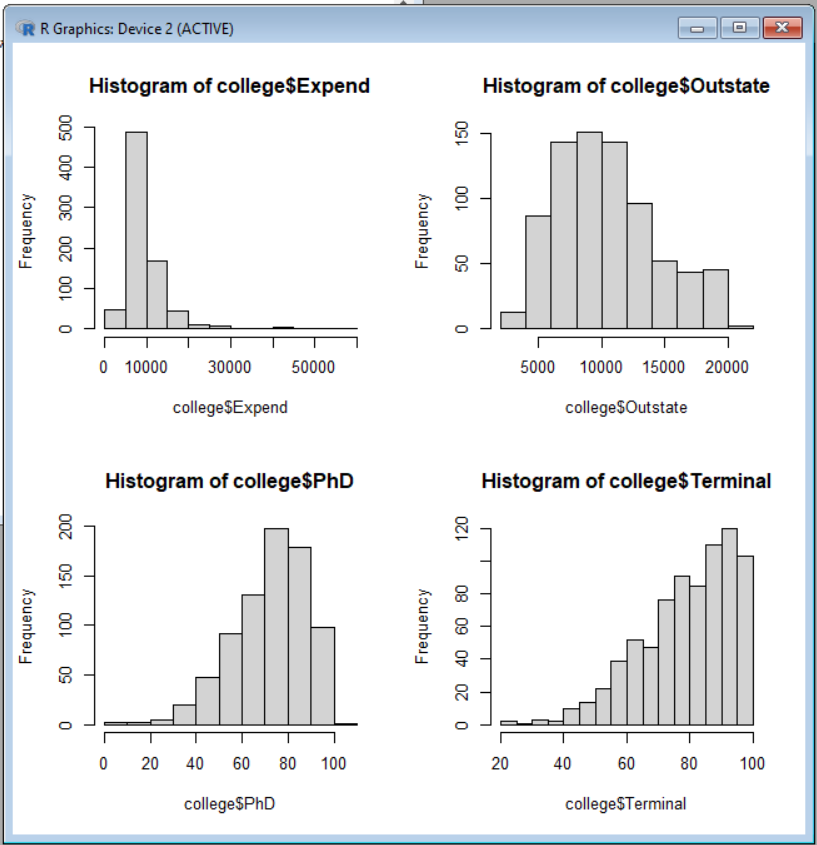
> par(mfrow=c(2,2))

> hist(college$Expend)

> hist(college$Outstate)

> hist(college$PhD)

> hist(college$Terminal)



**1.3.6** Отже спираючись на дані можна зробити такі висновки:

TODO

**Пункт 2 (Auto)**

**Пункт 3 (Boston)**