1. Набір даних College (College.csv) складається з даних про 777 різних університетів та коледжів.

*•* Private : державний/приватний

*•* Apps : кількість отриманих заяв

*•* Accept : кількість прийнятих заяв

*•* Enroll : кількість нових студентів

*•* Top10perc : нові студенти з топ 10 % шкіл

*•* Top25perc : нові студенти з топ 25 % шкіл

*•* F.Undergrad : кількість студентів денної форми навчання

*•* P.Undergrad : кількість студентів заочної форми навчання

*•* Outstate : вартість навчання іноземних студентів

*•* Room.Board : вартість проживання та харчування

*•* Books : оціночна вартість книжок

*•* Personal : оціночні персональні витрати

*•* PhD : відсоток факультетів з PhD

*•* Terminal : Відсоток професорсько-викладацького складу

*•* S.F.Ratio : відношення студенти/факультети

*•* perc.alumni : Відсоток випускників, які роблять внести

*•* Expend : Витрати на навчання на одного студента

*•* Grad.Rate : відсоток успішного закінчення

1.1. З допомогою функції read.csv() завантажте дані в R. Викличте завантажені дані college.

1.2. Перегляньте дані з допомогою функції fix(). Зауважте, що перший стовпець це просто назви університетів. Введіть наступні команди:

> rownames (college )=college [,1]

> fix (college )

Що сталося? Введіть:

> college =college [,-1]

> fix (college )

А тепер?

1.3.1. З використанням функції summary() отримайте підсумок відносно даних.

1.3.2. Використайте pairs() для побудови матриці графіків перших 10 стовпців даних.

1.3.3. Використайте plot() для побудови діаграми Outstate vs Private.

1.3.4. Створіть новий якісний показник Elite, використовуючи Top10perc. Поділимо всі університети на дві групи в залежності чи перевищує відсоток студентів з топ 10% шкіл 50% чи ні.

> Elite =rep ("No",nrow(college ))

> Elite [college$Top10perc >50]=" Yes"

> Elite =as.factor (Elite)

> college =data.frame(college ,Elite)

Використовуючи summary(), визначте чи багато таких університетів. З допомогою plot() побудуйте діаграму Outstate vs Elite.

1.3.4. Використовуючи hist(), побудуйте декілька гістограм для кількісних показників. Команда par(mfrow=c(2,2)) може бути корисною: вікно графіків буде поділено на 4 частини і можна побудувати 4 графіки одночасно.

1.3.5. Зробіть певні короткі висновки щодо даних.

2. Переконайтеся, що в даних Auto видалені пропущені значення.

2.1. Які з показників є кількісні, а які якісні?

2.2. Які межі для кожного кількісного показника? Використайте функцію range().

2.3. Обчисліть середнє і стандартне відхилення кожного кількісного показника?

2.4. Видаліть спостереження з 10-го по 85-те. Як зміняться середнє і стандартне відхилення?

2.5. Використовуючи повний набір даних проаналізуйте показники графічно. Обгрунтуйте певні взаємозв’язки між показниками.

2.6. Нехай ми хочемо передбачити розхід пального (mpg) на основі інших показників. Чи буде це можливо?

3. Дослідимо дані Boston.

3.1. Завантажте дані Boston, які є частиною бібліотеки MASS.

> library (MASS)

Дані знаходяться в об’єкті Boston.

> Boston

Перегляньте інформацію про дані:

> ?Boston

Скільки рядків і стовпців міститься в множині?

3.2. Побудуйте декілька попарних графіків показників.

3.3. Чи пов'язаний якийсь із показників із рівнем злочинності на душу населення?

3.4. Чи є в кварталах Бостона особливо високий рівень злочинності? Податкові ставки? Співвідношення учнів і вчителів? Прокоментуйте діапазон кожного показника.

3.5. Скільки кварталів в даній множині межують з річкою Charles?

3.6. Обчисліть медіану для відношення учні-вчителі для міста загалом?

3.7. Який квартал (квартали) міста має найменшу медіану кількості зайнятих помешкань? Які значення інших показників цього (цих) кварталів? Як ці показники співвідносяться зі значеннями показників інших кварталів.