

Питання до заліку
«Методи пошуку та обробки інформації»

Загальні питання.

1. Data Science, data scientists. Визначення.
2. Питання аналізу даних. Типи питань.
3. Описовий (descriptive) аналіз.
4. Дослідний (exploratory) аналіз.
5. Дедуктивний (inferential) аналіз.
6. Прогностичний (predictive) аналіз.
7. Причинно-наслідковий (causal) аналіз.
8. Механістичний (mechanistic) аналіз.
9. Визначення даних. Якими вони можуть бути?
10. Що таке великі дані (Big data)?
11. Спотворювання (confounding). Шляхи боротьби з ним.
12. Що таке апостеріорний аналіз (data dredging)?
13. CLI. Основні команди.
14. Контроль версій (version control) та Git.
15. Github.

Мова R.

1. Особливості та недоліки мови R. Що таке free software?
2. Базові (atomic) класи об'єктів. Атрибути об'єктів.
3. Створення векторів. Приведення (coercion). Явне приведення.
4. Матриці. Створення матриць.
5. Lists, factors, missing values.
6. Data frames. Names.
7. Зчитування та запис даних.
8. Subsetting, partial matching.
9. Removing NA.
10. Vectorized operations.
11. Control structures: If.
12. Control structures: For.
13. Control structures: Repeat, Break, Next, Return.
14. Функції: Defining a function, function arguments, argument matching.
15. Функції: lazy evaluation, «...» аргумент.
16. Прив'язка значень до символу в R.
17. Lexical scoping.
18. Стандарти кодування.

19. Дати та час в R.

20. Операції з датами та часом в R.

Практичні питання. Приклад

Прочитайте файл «hw1_data.csv». Виведіть частину набору даних (subset) зі значенням «Ozone» > 31 та «Temp» > 90. Яке середнє (mean) значень «Solar.R» в цьому наборі даних?