

Знакомство с языком R

1. Установить и запустить программу R.
2. Настроить рабочую директорию с помощью функции `setwd()`.
3. Сгенерировать вектор $X \sim U(0, 1)$ и построить гистограмму.
4. Сохранить результаты моделирования в текстовый файл (*.txt, *.csv), гистограмму в графический файл (*.jpg, *.svg) с использованием функций записи в файл.
5. Познакомиться с основными структурами данных `numeric()`, `matrix()`, `data.frame()`, `list()`.
6. Познакомиться со справочной системой.
7. Установка и загрузка дополнительных пакетов с помощью функций `install.packages()` и `library()`.

Литература

1. <https://ru.stackoverflow.com/questions/506597/Книги-и-учебные-ресурсы-по-языку-R>

Лабораторная работа 2

Стратегии работы с большими массивами данных

1. Сгенерировать большой массив данных и записать в один файл [1]. Установить пакет `rugpt`. Записать массив данных по частям в несколько файлов [2]. Сформировать репрезентативную выборку ограниченного размера.
2. Выполнить загрузку данных с использованием различных стратегий [1]. Сделать выводы. Установить пакеты `data.table`, `sqldf`, `ff`.
3. Установить и загрузить библиотеки `sqldf` и `nycflights13`. Ознакомиться со структурой набора данных `flights`. Вычислить количество наблюдений для всех перевозчиков `carrier` в таблице `flights`. Отобразить в консоли значения полей `dep_time`, `dep_delay`, `arr_time`, `carrier`, `tailnum` из таблицы `flights` (первые и последние 5 строк). Вычислить среднее время задержки прибытия (`mean_arr_delay`) и отправления (`mean_dep_delay`) для различных перевозчиков (`carrier`) [3].
4. Сгенерировать `data.frame` с тремя столбцами и 100 строками. Преобразовать данные из широкого в длинный формат. Установить пакет `reshape2` [4].

Литература

1. <https://stackoverflow.com/questions/1727772/quickly-reading-very-large-tables-as-dataframes/>
2. <https://stackoverflow.com/questions/57047338/split-dataset-per-rows-into-smaller-files-in-r>
3. https://www.tutorialspoint.com/big_data_analytics/introduction_to_sql.htm
4. Кабаков Р.И. R в действии. Анализ и визуализация данных на языке R (2014).