

## Лабораторная работа №2

### Загрузка данных в R. Библиотеки

**Цель:** изучить способы загрузки и подключения пакетов, загрузки различных типов данных, сохранения данных в файл, сохранение набора данных с заданным типом в **R**, ознакомиться со способами хранения и представления данных, исследовать структуру источников открытых данных.

### Задания

1. Создать новый проект.
2. Сделать папку «Lab12», созданной в лабораторной работе №1, рабочей.
3. Загрузить в среду **RStudio** три таблицы из лабораторной работы №1, средствами **RStudio**.
4. Удалить из **RStudio** все загруженные данные.
5. Загрузить три таблицы, используя функции **read\_xlsx()**, **read.table()** и **read.csv()**: создать переменные **company**, **company1** и **company2**, в которые поместить данные из файлов **company.xlsx**, **company1.txt** и **company2.csv** соответственно.
6. Изучить команду **read.table()**: **?read.table**. Выписать названия и назначение аргументов функции (продолжить таблицу 1.5).
7. С помощью команды **read.table()** загрузить в переменную **company512** данные, состоящие из 5-12 строк таблицы **company2.csv**.
8. Посмотреть данные трёх переменных: **company**, **company1** и **company2**.
9. Проверить какой тип данных у переменной **company**.
10. Избавиться от дополнительных типов данных в переменной **company**.
11. Проверить, одинаковые ли наборы данных **company**, **company1** и **company2**, сравнив их попарно.
12. Если таблицы **company**, **company1** и **company2** попарно не равны, найти несоответствия (посмотреть структуру переменных, например, используя команду **str(<имя\_таблицы>)**) и исправить их, используя соответствующие функции (**as.factor**, **as.numeric** и т.д.).
13. Исследовать наборы данных с портала открытых данных **data.gov.ru**.
  - 13.1. В табл. 1.7 указаны варианты заданий. Выбрать свой вариант согласно журналу студенческой группы.
  - 13.2. Выбрать произвольный набор данных в формате **csv** с портала **data.gov.ru** по тематике варианта. Таблица должна содержать

минимум две числовые переменные.

13.3. Сохранить выбранный набор данных, его паспорт в рабочую папку «Lab12».

13.4. Загрузить данные в переменную **tab**.

13.5. Определить количество переменных и строк в наборе данных **tab**.

Таблица 1.7. Данные портала data.gov.ru

Вариант	Номер студента в журнале	Тема	Адрес данных по тематике в сети Интернет
1	1, 17	Безопасность	<a href="http://data.gov.ru/rubriki/bezopasnost">http://data.gov.ru/rubriki/bezopasnost</a>
2	2, 18	Государство	<a href="http://data.gov.ru/rubriki/gosudarstvo">http://data.gov.ru/rubriki/gosudarstvo</a>
3	3, 19	Досуг и отдых	<a href="http://data.gov.ru/rubrics/leisure-and-entertainment">http://data.gov.ru/rubrics/leisure-and-entertainment</a>
4	4, 20	Здоровье	<a href="http://data.gov.ru/rubriki/zdorove">http://data.gov.ru/rubriki/zdorove</a>
5	5, 21	Картография	<a href="http://data.gov.ru/rubrics/cartography">http://data.gov.ru/rubrics/cartography</a>
6	6, 22	Культура	<a href="http://data.gov.ru/rubrics/culture">http://data.gov.ru/rubrics/culture</a>
7	7, 23	Метеоданные	<a href="http://data.gov.ru/rubrics/weather">http://data.gov.ru/rubrics/weather</a>
8	8, 24	Образование	<a href="http://data.gov.ru/rubriki/education">http://data.gov.ru/rubriki/education</a>
9	9, 25	Спорт	<a href="http://data.gov.ru/rubrics/sport">http://data.gov.ru/rubrics/sport</a>
10	10, 26	Строительство	<a href="http://data.gov.ru/rubriki/stroitelstvo">http://data.gov.ru/rubriki/stroitelstvo</a>
11	11, 27	Торговля	<a href="http://data.gov.ru/rubriki/torgovlya">http://data.gov.ru/rubriki/torgovlya</a>
12	12, 28	Транспорт	<a href="http://data.gov.ru/rubriki/transport">http://data.gov.ru/rubriki/transport</a>
13	13, 29	Туризм	<a href="http://data.gov.ru/rubrics/tourism">http://data.gov.ru/rubrics/tourism</a>
14	14, 30	Экология	<a href="http://data.gov.ru/rubrics/ecology">http://data.gov.ru/rubrics/ecology</a>
15	15, 31	Экономика	<a href="http://data.gov.ru/rubriki/ekonomika">http://data.gov.ru/rubriki/ekonomika</a>
16	16, 32	Электроника	<a href="http://data.gov.ru/rubrics/electronics">http://data.gov.ru/rubrics/electronics</a>

14. Найти файл с данными в сети Интернет и загрузить его в **R**, используя адрес.

15. Загрузить любой файл в переменную **tab1** используя интерактивный выбор.

16. Сохранить все созданные данные в файл **company.Rda**.

17. Сохранить данные переменной **company1** в формате **.csv**.

18. Сохранить данные переменной **company2** в формате **.txt**.

19. Сохранить данных в **Excel**-книгу с помощью пакета **openxlsx**. В первый лист таблицу **company**, во второй – **company1**, в третий – **company2**, в четвертый – **tab**. Назвать листы «Первый», «Второй», «Третий» и «Четвертый» соответственно.

20. Изучить аргументы функций **write.csv()** и **write.table()**. Выписать названия и назначение основных аргументов функций.

## Домашнее задание

1. Сделать отчет по лабораторной работе.
2. Закрепить навыки установки библиотек:
  - 2.1. Установить на домашнем ПК с помощью графического интерфейса следующие библиотеки: **rattle**, **Rcmdr** и **Deducer**.
  - 2.2. Установить все сопутствующие пакеты, предложенные **R**.
  - 2.3. Загрузить библиотеки в среду **RStudio**.
  - 2.4. Установить с сайта **GitHub** библиотеку **readr** (не забудьте предварительно установить пакет **devtools**: `install.packages("devtools")`).

## Отчет по лабораторной работе №2

1. Титульный лист, включая название работы, вариант, ФИО студента, учебную группу.
2. Оглавление.
3. Отчет:
  - 3.1. Номер и описание задания.
  - 3.2. Код решения.
  - 3.3. Результат в консоли.
  - 3.4. Скрин результата из консоли **RStudio**.
  - 3.5. Для задания 12:
    - 3.5.1. Скрин набора данных, скаченных с сайта **data.gov.ru**, открытых в программе **Excel**.
    - 3.5.2. Заполнить таблицу 1.8 «Описание набора данных».
4. Список использованных источников литературы.

Таблица 1.8.

Показатель	Значение
Название набора данных	
Формат данных	
Ссылка в сети Интернет	
Количество записей	
Количество переменных	
в том числе текстовых	
в том числе числовых	

## Контрольные вопросы

1. С помощью каких команд устанавливаются библиотеки?
2. Что такое **CRUN**?
3. Назовите команды для загрузки библиотек.
4. Перечислите способы загрузки данных в **R**.
5. Какие команды предназначены для сохранения данных из среды **R** в файл?
6. Как импортировать текстовые данные, хранящиеся на веб-сайтах?
7. С помощью какой команды можно очистить активное пространство?