

# Практическая работа №1.

## Подготовка данных для анализа

### Открытие и настройка проекта в RStudio

**Цель:** изучить способы подготовки данных, приведения их к табличному виду с помощью операций устранения несоответствий данных: кодирования, отбора, нормализации и структурирования. Научиться устанавливать **R** и **RStudio** на ПК, открывать в **RStudio** новый проект и настраивать его, устанавливать рабочую папку.

#### 1. Подготовка данных

Дана таблица данных о сотрудниках компании (табл. 1.6, файл **company.xlsx**).

Таблица 1.6

	A	B	C	D	E	F	G	H	I	J	K
1			Name	Возраст	Семейное положение	Общий стаж работы	Стаж в ООО	Должность	Пол		
2			Анна	35	замужем	14	3	менеджер	ж		
3			Petr	двадцать два	married	2	1	driver	м		
4			Ildar	33	холост	девять	5	вторая	male		
5			Anna	двадцать три		2	2	секретарь	female		
6			Ульяна	27	single	4	три	помощник дир.	ж		
7			Максим	42	женат	двадцать	5	ген.директор	м		
8			Alina	27	замужем	5	5	гл. бухгалтер	female		
9			Vasilisa	тридцать	вдова	семь	4	менеджер	ж		
10											
11											
12				position		Sex	имя	Age	marital	yer.service	yer.service.OOO
13				manager		female	Екатерина	32	married	10	4
14				оператор		male	Илья	34	single	13	4
15				водитель		м	Vladimir	старая	птица	five	2
16				третья		female	Nina	29	divorced	6	1
17				secretary		female	Olga	двадцать	single	2	1
18				operator		female	Arina	37	замужем	16	4
19				driver		male	Виталий	35	married	13	5
20				manager		female	Mariia	тридцать	single	девять	3
21				менеджер		ж	Надежда	25	married	5	4

#### Задания

1. В программе **Excel** открыть файл **company.xlsx**. Привести данные к «правильному» формату, оставив переменную «Семейное положение» текстовой.

! Имена переменных не должны содержать пробелов, так как при загрузке таблиц в среду R пробел может быть заменён, например, на точку или нижнее подчеркивание, в зависимости от типа загружаемого набора данных.

2. Сохранить файл в трёх форматах: **.xlsx**, **.csv** и **.txt**:

**company.xlsx, company1.txt и company2.csv**




Создать файлы **.txt** и **.csv** в программе **Excel**:

- В верхнем левом углу программы нажать на вкладку «**Файл**».
- Выбрать «**Экспорт**».
- Выбрать «**Изменить тип файла**». Появится список форматов.
- Выбрать необходимый формат:  
**.csv** (разделитель – запятая),  
**.txt** (разделитель – знак ‘**tab**’).
- Прописать имя файла и его расположение.

3. Создать папку «Lab12» на рабочем столе и поместить в неё файлы **company.xlsx**, **company1.txt**, **company2.csv**.

## 2. Установка R, RStudio

### Задания

1. Скачать с сайта дистрибутив языка **R** по ссылке:  
<https://cran.r-project.org/>
2. Скачать с сайта инсталлятор интегрированной среды разработки **RStudio** для работы с языком программирования **R**:  
<https://www.rstudio.com/products/rstudio/download/>
3. Установить **R** и **RStudio** на ПК.
4. Изучить окна **RStudio**.
5. Изучить графические возможности **R**. Для этого ввести в консоли команды: **demo(graphics)**, **demo(image)**, **demo(persp)**, **demo(Hershey)** и нажимать «**enter**».
6. В консоли решить простые примеры на сложение, умножение и др.
7. Открыть новый скрипт: нажать на  в левом верхнем углу панели **RStudio** и выбрать вкладку  R Script  ⌘N.
8. Прописать в редакторе кода примеры из элементарной математики.

## 3. Создание проекта

### Задания

1. Создать в **RStudio** новый проект.

2. Настроить созданный проект.
3. Проверить какая папка является рабочей.
4. Установить рабочей папкой «Lab12».
5. Убедиться, что папка «Lab12» является рабочей папкой проекта.
6. Узнать список файлов в рабочей папке проекта.

### **Домашнее задание**

1. Скачать дистрибутивы и установить на домашнем ПК **R** и **RStudio**.
2. Создать новый проект.
3. Изучить настройки проекта.
4. Создать новую папку и сделать её рабочей.

### **Контрольные вопросы:**

1. Охарактеризуйте операции предобработки данных: нормирование, кодирование и др.
2. Назовите основные форматы хранения данных и основные источники открытых данных.
3. Назовите способы создания проекта в **RStudio**.
4. С каким форматом данных предпочтительнее работать в **R**?
5. Назовите команду, предназначенную для проверки того, какая папка является рабочей.
6. Как изменить расположение рабочей папки?
7. Какой командой можно предотвратить проблемы с данными в случае, когда они могут быть прочитаны в **R** как факторы и могут неправильно переконвертироваться в числа?