

## Лабораторная работа 1. Основы работы в среде RStudio

**Цель занятия** — ознакомиться с интерфейсом IDE RStudio; создать проект; установить необходимые пакеты; выполнить импорт, очистку и экспорт данных; провести простой анализ данных; научиться создавать функции.

**Вспомогательный материал:** Лекции 2-3.

**Дополнительные файлы:** <https://goo.gl/7aDHYV>

### Общие указания:

1. В качестве отчета о выполнении практического занятия студент представляет преподавателю программный код (*Ib1.R*) и созданные файлы в соответствии с заданиями.
2. Для импорта, экспорта и обработки данных нельзя использовать возможности интерфейса RStudio — можно только писать программный код.
3. В ходе выполнения заданий студент должен обязательно использовать следующие функции R:

<code>apply()</code>	<code>do.call()</code>	<code>strsplit()</code>
<code>as.data.frame()</code>	<code>gsub()</code>	<code>substr()</code>
<code>cbind()</code>	<code>rbind()</code>	<code>tbl_df()</code>
<code>colnames()</code>	<code>read.table()</code>	
<code>diff()</code>	<code>regexpr()</code>	
4. Можно использовать другие функции из установленных пакетов.
5. Нельзя использовать операции трансформации.
6. Задания должны быть выполнены наиболее оптимальным образом (автоматизация, минимизация строк кода, универсальность и др.).
7. Программный код должен быть оформлен в соответствии с Google's R Style Guide.

### Задание 1

—> *Google's R Style Guide.xml*

Ознакомиться с Google's R Style Guide

### Задание 2

`<— Folders`

`<— mitoad.Rproj`

`<— Ib1.R`

Создать проект в RStudio:

- На диске D: создать папку, названную номером своей учебной группы (например, 141701s).

- В созданной папке создать рабочую папку, названную своей фамилией (например, hedranovich).
- Запустить RStudio и ознакомиться с интерфейсом программы.
- Установить тему Cobalt: Tools — Global Options... — Appearance — Editor Theme.
- Создать новый проект mitoad.Rproj, сохранить его в своей рабочей папке.
- Создать в папке проекта следующие папки:
  - code (для сохранения файлов \*.R);
  - data (для сохранения файлов с данными). Скопировать в неё файлы:
    - List of species EuroBirdwatch-2014 Belarus.csv
    - Tablitsa-zagryaznyajushih-veshestv-po-gorodam.csv
    - Tablitsa-zagryaznyajushih-veshestv-v-atmosfernom-vozduhe-po-gorodam.csv
  - docs (для сохранения других файлов). Скопировать в неё остальные файлы для ЛБ1.
- Создать новый R Script (Ctrl+Shift+N), сохранить его под именем lb1.R в соответствующей папке. Весь программный код для выполнения заданий должен содержаться в этом файле.

### Задание 3

`<— packages.txt`

Установка пакетов:

- Установите пакеты: plyr, ggplot2, ggthemes, knitr, highcharter, stringi, sp.
- Убедитесь, что все пакеты из следующего списка установлены:

cluster	knitr	stats
dplyr	lubridate	stringi
ggplot2	plyr	tidyr
ggthemes	rmarkdown	xts
highcharter	sp	

- Установите отсутствующие пакеты.
- Создайте файл packages.txt в папке docs. Файл должен содержать полный список установленных пакетов.

### Задание 4

`—> List of species EuroBirdwatch-2014 Belarus.csv`

`<— List of species EuroBirdwatch-2014 Belarus_updated.csv`

`<— List of species EuroBirdwatch-2014 Belarus_analysis.txt`

Проанализируйте data set “List of species EuroBirdwatch-2014 Belarus”:

- Добавьте к таблице данных колонки и заполните их соответствующими значениями: individualsPerObservation, speciesShare.
- Отсортируйте данные в порядке убывания individualsNumber.
- Сохраните модифицированную таблицу в файл List of species EuroBirdwatch-2014 Belarus\_updated.csv

- Получите следующие данные и запишите их в файл List of species EuroBirdwatch-2014 Belarus\_analysis.txt:
  - Общее количество наблюдаемых птиц.
  - Медианное значение числа наблюдений.
  - Перечни названий птиц, число наблюдаемых особей которых попадает в диапазоны (0:10], (10:50], (50:100], (100:500], (500:1000], (1000:...).
  - Название птицы, число особей которой было максимальным в одном наблюдении.
  - Получите ещё три метрики (insights) на своё усмотрение.

## Задание 5

—> *Tablitsa-zagryaznyajushih-veshestv-po-gorodam.csv*

—> *Tablitsa-zagryaznyajushih-veshestv-v-atmosfernom-vozdue-po-gorodam.csv*

<— *Tablitsa-zagryaznyajushih-veshestv-po-gorodam\_cleaned.csv*

<— *Tablitsa-zagryaznyajushih-veshestv-po-gorodam\_trend.csv*

Проанализируйте data set “Tablitsa-zagryaznyajushih-veshestv-po-gorodam”:

- Ознакомьтесь со структурой данных (см. contamination.pdf) и проведите очистку данных:
  - Данные, разделенные наклонной чертой необходимо представить в разных столбцах.
  - Для данных, представленных в виде диапазона (1 – 30), вычислите среднее и замените им диапазон (создайте для этого функцию).
- Сохраните полученную таблицу в файл *Tablitsa-zagryaznyajushih-veshestv-po-gorodam\_cleaned.csv*.

Проанализируйте data set “Tablitsa-zagryaznyajushih-veshestv-v-atmosfernom-vozdue-po-gorodam”:

- Ознакомьтесь со структурой данных и проведите очистку и подготовку данных.
- Добавьте к таблице два столбца:
  - С абсолютным трендом загрязнения (increase, decrease, mixed).
  - С относительным изменением уровня загрязнения.
 Для вычисления значений столбцов напишите функции.
- Создайте список с тремя элементами: increase, decrease, mixed. Каждый элемент должен хранить массив соответствующих городов и загрязняющих веществ.
- Сохраните полученную таблицу в файл *Tablitsa-zagryaznyajushih-veshestv-po-gorodam\_trend.csv*.

Data Sets получены из Базы открытых публичных данных <https://opendata.by/>