**Лабораторная работа № 9**

**«Предобработка данных средствами пакетов tidyr и dplyr»**

Задание 1. Пусть есть некоторый набор данных. В этом наборе отражены данные по 4 людям, участвующим в исследовании. Этим людям предлагается лекарство типа «А» или лекарство типа «B» в разное время T1 и T2. При этом измеряется температура тела этих людей.

Дата-фрейм создается с помощью следующих команд:

> set.seed(1122)

> dd<-data.frame(Name=c("John","Peter","Mary","Caroline"),

+ DrugA\_T1=runif(4,35,36),

+ DrugA\_T2=runif(4,36,39),

+ DrugB\_T1=runif(4,36,36.6),

+ DrugB\_T2=runif(4,37,38.5))

Такой фрейм не является tidy-data, потому что здесь в рамках одной и той же переменной неявным образом спрятан и тип лекарства (A и B), и время (T1 и T2), поэтому необходимо использовать пакет tidyr, чтобы привести эти данные к нужному формату.

Выполнить:

1. Используя функцию gather, преобразовать данные из широкого в длинный формат.
2. Name V Temp
3. 1 John DrugA\_T1 35.07729
4. 2 Peter DrugA\_T1 35.01392
5. 3 Mary DrugA\_T1 35.67363
6. 4 Caroline DrugA\_T1 35.14321
7. 5 John DrugA\_T2 37.41580
8. 6 Peter DrugA\_T2 37.41355
9. 7 Mary DrugA\_T2 38.92508
10. 8 Caroline DrugA\_T2 36.52195
11. 9 John DrugB\_T1 36.17455
12. 10 Peter DrugB\_T1 36.00115
13. 11 Mary DrugB\_T1 36.49993
14. 12 Caroline DrugB\_T1 36.05723
15. 13 John DrugB\_T2 37.44987
16. 14 Peter DrugB\_T2 37.88969
17. 15 Mary DrugB\_T2 38.25544
18. 16 Caroline DrugB\_T2 38.08802

2. С помощью функции separate разделить переменную, содержащую тип лекарства и время.

Name DrugType Time Temp

1 John DrugA T1 35.07729

2 Peter DrugA T1 35.01392

3 Mary DrugA T1 35.67363

4 Caroline DrugA T1 35.14321

5 John DrugA T2 37.41580

6 Peter DrugA T2 37.41355

7 Mary DrugA T2 38.92508

8 Caroline DrugA T2 36.52195

9 John DrugB T1 36.17455

10 Peter DrugB T1 36.00115

11 Mary DrugB T1 36.49993

12 Caroline DrugB T1 36.05723

13 John DrugB T2 37.44987

14 Peter DrugB T2 37.88969

15 Mary DrugB T2 38.25544

16 Caroline DrugB T2 38.08802

3. Используя функцию select пакета dplyr, выберите из дата-фрейма значения столбцов “Time” и “Temp”.

4. Используя функцию filter, выберите из дата-фрейма те наблюдения, в которых значение температуры превышает 37, а имя пациента либо «John», либо «Mary».

5. Используя функцию arrange пакета dplyr, отсортируйте данные дата-фрейма по имени (по возрастанию) и по температуре (по убыванию).

6. Используя функцию summarize пакета dplyr, найдите среднюю температуру для каждого момента времени. Данные группируются по времени.

Задание 2. Допустим, что в нашем распоряжении некий дата фрейм dg из пяти столбцов в указанном порядке:

* first\_name (имя)
* last\_name (фамилия)
* email\_address (электронная почта)
* postal\_address (почтовый адрес)
* date\_added (дата добавления записи в базу данных).

Создайте данный дата-фрейм в R. Заполните его данными (10 -15 записей).

С помощью функции select пакета dplyr напишите 3 различных выражения, которые выбирают данные по столбцам, содержащим имя, фамилию и дату.