

ЛАБОРАТОРНАЯ РАБОТА №9

Работа со строковыми данными с помощью пакетов `stringr` и `rebus`

Цель работы: научиться работать со строковыми данными с использованием пакетов `stringr` и `rebus`.

ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Некоторые сведения о пакете `babynames`

`babynames` представляет собой пакет, содержащий сведения об именах детей рожденных в США в период с 1880 по 2017 гг. В рамках выполнения лабораторной работы необходимо работать с набором данных аналогичным по названию пакету.

В наборе содержатся сведения об именах (`name`), поле (`sex`), годе (`year`), количестве детей (`n`) и частоте встречаемости имени (`prop`). Данные хранятся в виде `tibble` датафрейма.

ПРАКТИЧЕСКАЯ ЧАСТЬ

Задание 1 – С помощью пакетов `stringr` и `rebus`, используя набор данных `babynames` пакета `babynames` реализуйте код, дающий ответы на вопросы, согласно варианту. При решении задач также может быть полезен пакет `drplyr`.

Вариант – 1

Какое самое частое женское имя в период с 1920 по 1950 год?

Сколько букв имеет самое короткое женское имя?

Какое самое длинное мужское имя? (если таких имен несколько программа должна выводить наиболее часто встречающееся имя)

Сколько имен соответствуют шаблону: ГСГСГ? (где Г – гласная буква, С – согласная)

Вариант – 2

Какое самое частое мужское имя в период с 1900 по 1920 год?

Сколько букв в самом длинном мужском имени?

Какое самое длинное женское имя? (если таких имен несколько программа должна выводить первое встречающееся в лексикографическом порядке)

Сколько имен соответствуют шаблону: СГГСГГ? (где Г – гласная буква, С – согласная)

Вариант – 3

Сколько различных мужских имен встречается в период с 2000 по 2017 год, начинающихся с гласной буквы?

Сколько различных имен имеют длину 10 букв?

Сколько в среднем букв в наиболее часто встречающемся мужском имени в период с 1950 по 2000 год?

Сколько имен соответствуют шаблону: СГГСГГ? (где Г – гласная буква, С – согласная)

Задание 2 – Используя пакет `rebus` постройте шаблон, находящий совпадения в тексте по заданному варианту.

Вариант – 1

Поиск e-mail адреса

Вариант – 2

Поиск мобильного номера телефона в разных форматах (например: +7-XXX-XXX-XXXX)

Вариант – 3

Поиск даты рождения в разных форматах (например: ДД/ММ/ГГГГ)

Контрольные вопросы:

1. Перечислите основные функции пакета `stringr` и их назначение.
2. Опишите логику построения паттернов с использованием пакета `rebus`.