Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**ЛАБОРАТОРНАЯ РАБОТА №5**

**Дисциплина: Обработка больших данных**

Работу выполнила: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.В. Гончаренко

Направление подготовки: 02.03.02 Фундаментальная информатика и информационные технологии

Преподаватель: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Т.А.Приходько

**Тема:** Извлечение данных с WEB-страниц. Пакет rvest.

**Цель:** научиться работать извлекать информацию сWEB-страниц с помощью инструментов языка R.

**Вариант:**

|  |  |  |
| --- | --- | --- |
| № | ФИО | Страны |
| 5 | Гончаренко Валентина Викторовна | Китай, Индия, Япония, Австралия, Австрия |

Функций библиотеки rvest, используемые в данной лабораторной работе:

* read\_html() - Создание документа html из url
* html\_nodes() - Выделение фрагментов документа с помощью CSS-селекторов
* html\_table() - Разбор таблиц в рамках

После получения данных построим графики по каждому индексу из таблицы.

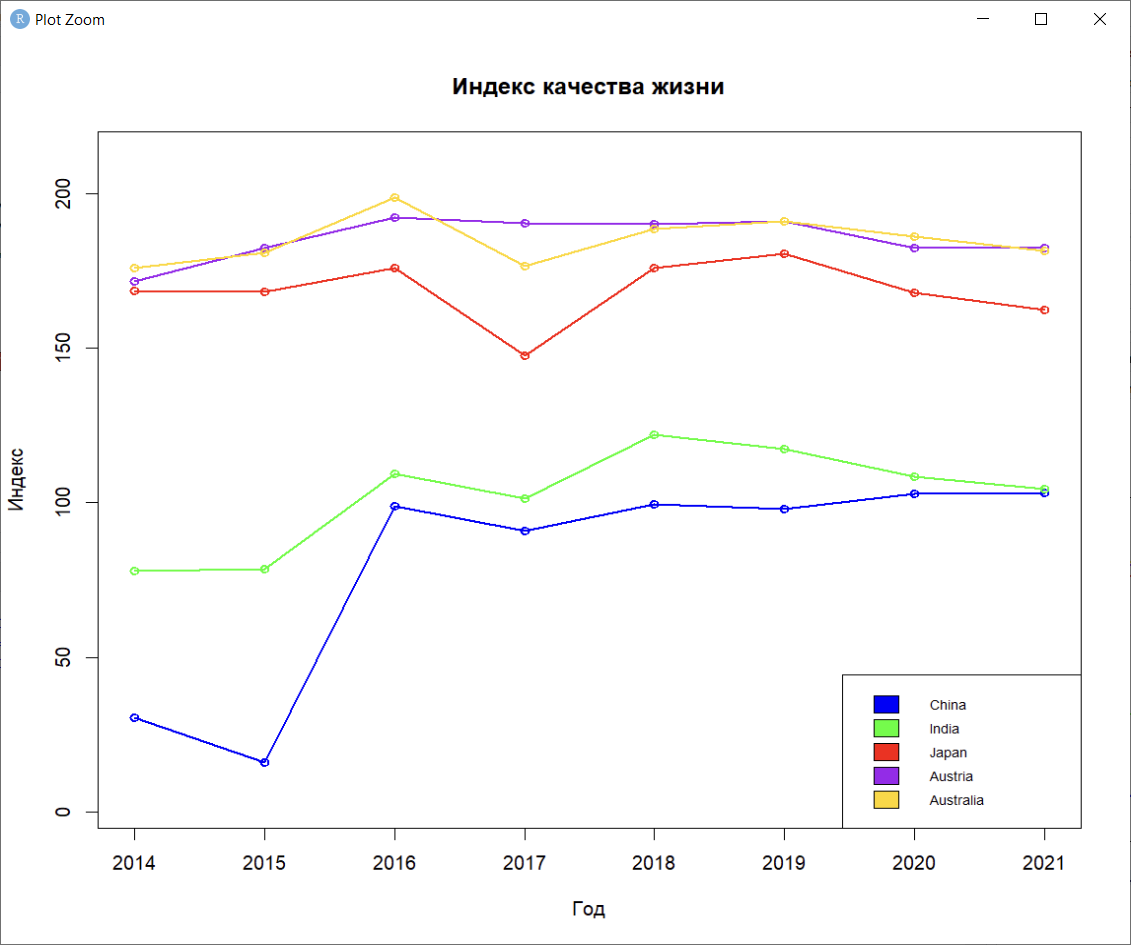


Рисунок 1 – Рейтинг стран по индексу качества жизни

По данной диаграмме можно сделать вывод о том, что в Китае и Индии индекс качества жизни ниже, чем в остальных представленных странах.

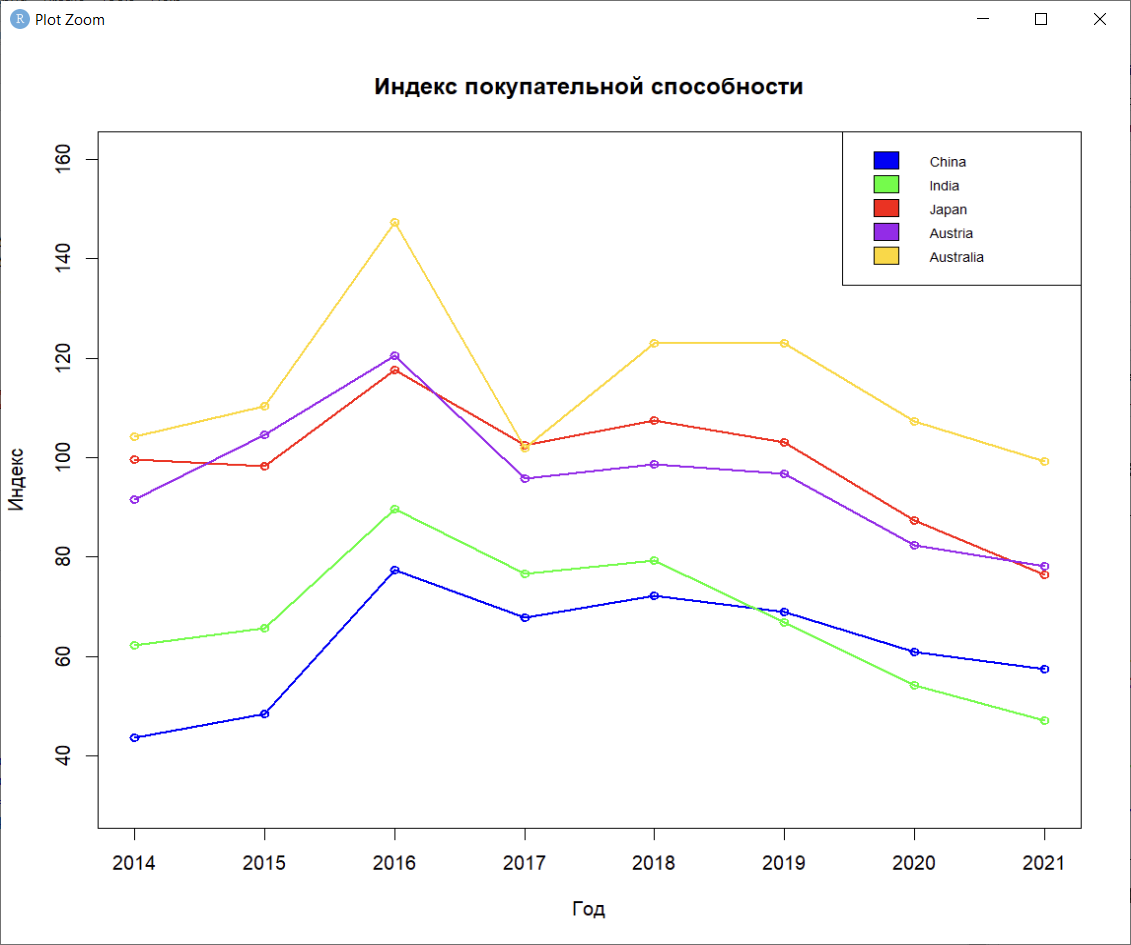


Рисунок 2 – Рейтинг стран по индексу покупательной способности

По данному графику можно сделать вывод о том, что самый большой индекс покупательной способности у Австралии, однако в 2017 году он был таким как в Японии. В Китае и Индии самые низкие индексы, а Япония и Австрия по данному показателю довольно близки.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 3 – Рейтинг стран по индексу безопасности

По данному графику можно сделать вывод о том, что самый высокий индекс безопасности принадлежит Японии, однако в 2017 году данная страна уступила Австрии.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 4 – Рейтинг стран по индексу здравоохранения

По данному графику можно сделать вывод о том, что самый высокий индекс здравоохранения принадлежит Японии, однако в 2017 году данная страна уступила Австрии. Худшие показатели принадлежат Китаю и Индии.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 5 – Рейтинг стран по индексу стоимости жизни

По данному графику можно сделать вывод о том, что самый низкий индекс принадлежит Индии, Австрия, Австралия и Япония преуспевают с большим отрывом.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 6 – Рейтинг стран по индексу соотношения цены недвижимости к доходу

По данному графику видно, что в Китае высокий индекс соотношения цены недвижимости к доходу, это означает, что цены на недвижимость в данной стране выше, чем может позволить себе большинство населения.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 7 – Рейтинг стран по индексу времени пути на транспорте

Данный график говорит нам о том, что в Австрии самый низкий индекс времени в пути на транспорте. В данной стране доступ к различным местам из определенного местоположения более быстрый.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 8 – Рейтинг стран по индексу загрязнения

Китай и Индия с приличным отрывом «преуспевают» в загрязнении.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 9 – Рейтинг стран по климатическому индексу

По данному графику видно, что все страны начиная с 2018 года имеют стабильные показатели, однако в 2017 году индекс резко упал.

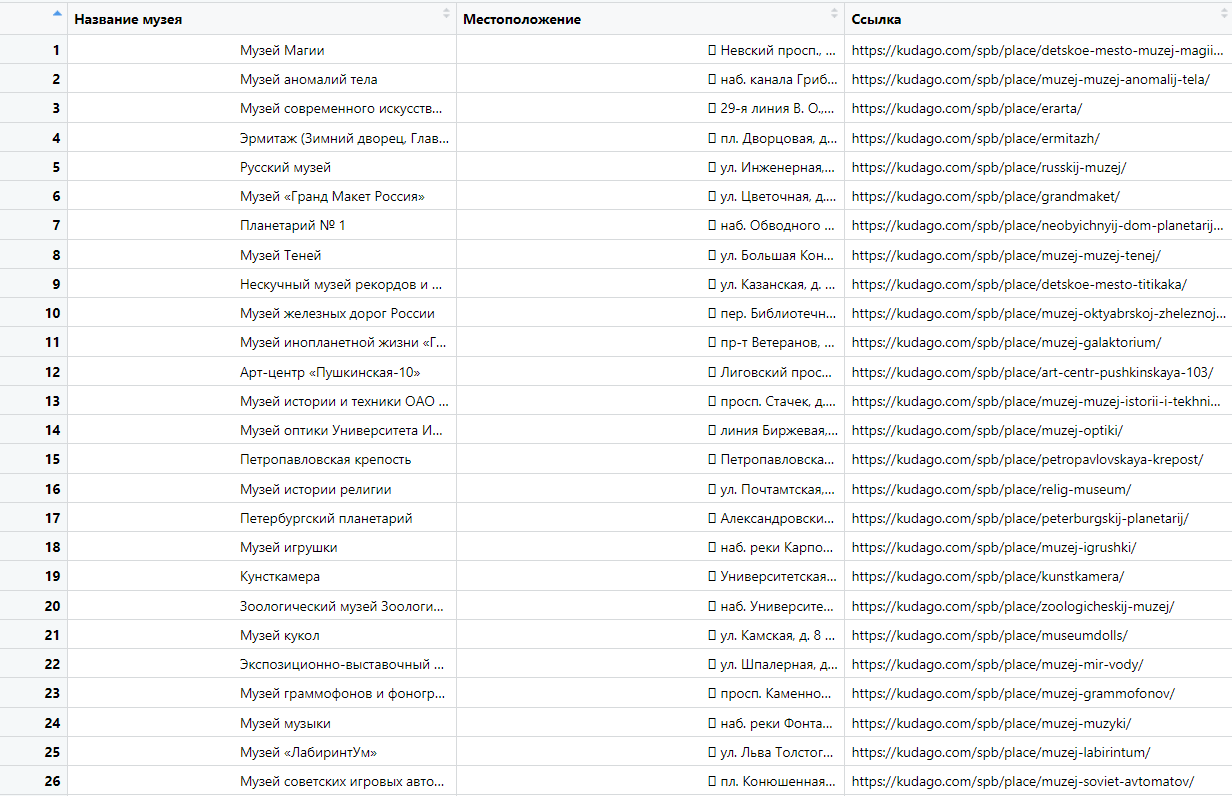


Рисунок 10 – Таблица дата фрейма, содержащего названия музея, его адрес и ссылку.

**Вывод:** В данной лабораторной работе был изучен пакет rvest и получены навыки по извлечению данных с WEB-страниц.