Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования «Кубанский государственный университет»

Кафедра вычислительных технологий

**ОТЧЕТ**

о выполнении лабораторной работы №4

по дисциплине «Обработка больших данных»

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

Выполнил: ст. гр. 36/2

Бачурин И.А.

Проверил: преподаватель

Яхонтов А.А.

Краснодар

2025

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

1. В ходе лабораторной работы, необходимо собрать информацию об уровне жизни стран мира из таблиц сайта https://www.numbeo.com/quality-of-life/rankings\_by\_country.jsp?title=2021 с 2014 по 2021гг:

Это оценка общего качества жизни с использованием

эмпирической формулы, которая учитывает:

— индекс покупательной способности (чем выше, тем лучше),

— индекс загрязнения (чем ниже, тем лучше),

— отношение цены на жилье к доходу (ниже). лучше),

— индекс прожиточного минимума (чем ниже, тем лучше),

— индекс безопасности (чем выше, тем лучше),

— индекс медицинского обслуживания (чем выше, тем лучше),

— индекс времени движения на дороге (чем ниже, тем лучше)

— климатический индекс (чем выше, тем лучше).

2. Каждый студент должен взять 5 стран (по варианту).

3. Составить data.frame (возможно для каждой страны) так, чтобы иметь возможность. Проанализировать с помощью графиков изменение рейтингов для всех 10 показателей. Для всех своих 5-ти стран, прокомментировать в отчете результат. Необходимо нарисовать на одном и том же графике рейтинг всех 5 стран, проанализировать результат, анализ словесно отразить в отчете. Проанализировать изменение во времени всех показателей указанных стран, подобрать наилучший (с вашей точки зрения) способ визуализации.

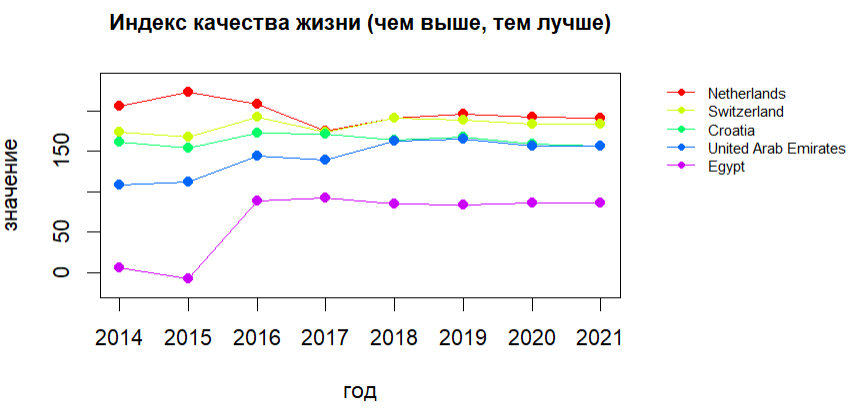


4. С одной из страниц (по варианту): вариант брать по формуле №пп % №ссылки + 1.



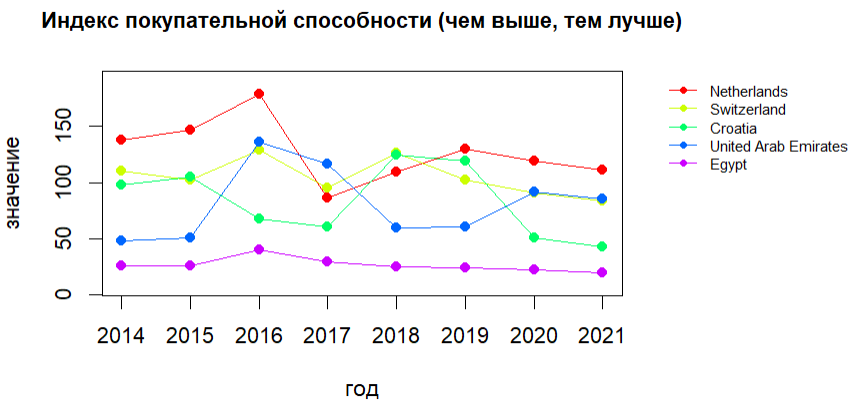
**Ход работы**

1. Индекс качества жизни:



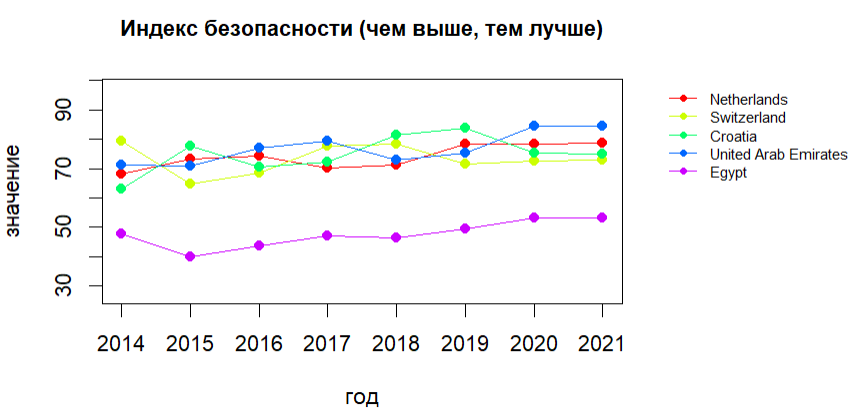
Чем выше данный показатель, тем лучше. Исходя из полученных данных можно сделать вывод, что с индекс качества жизни во всех представленных странах примерно одинаковый, только у ОАЭ он чуть ниже, чем у остальных. Хуже всего данный показатель у Египта. Но все равно он улучшился после 2015 года, где показатель был даже отрицательным.

2. Индекс покупательной способности:



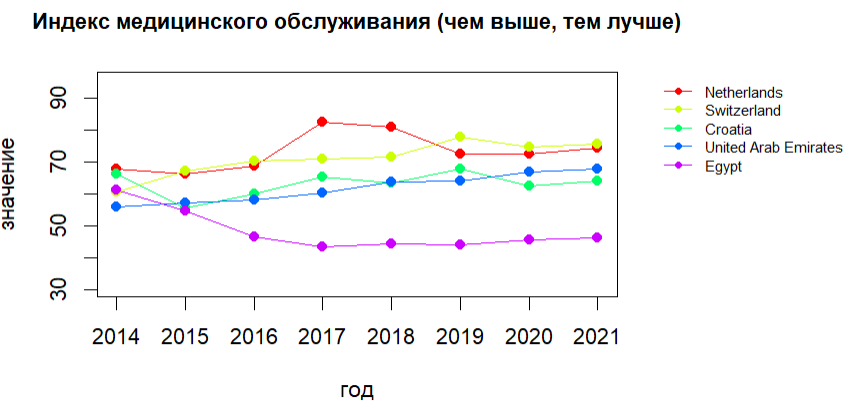
По индексу покупательной способности все годы, кроме 2017 и 2018 лидирует Голландия. В целом показатели держатся близко друг к другу. ОАЭ отстает немного, Египет по данному критерию сильно ниже.

3. Индекс безопасности:



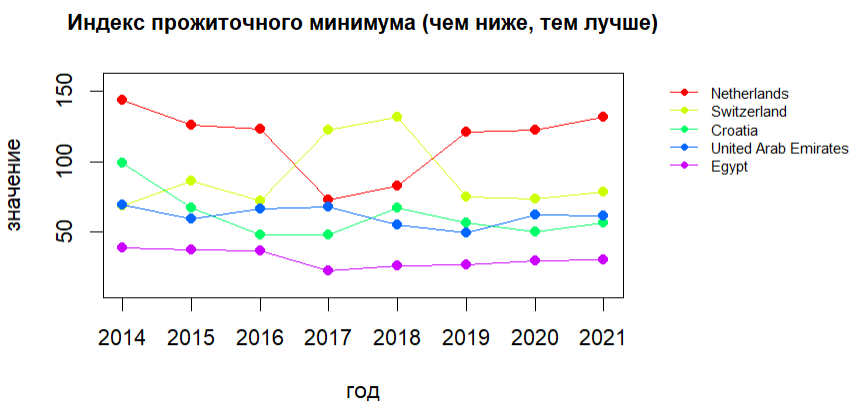
Данный показатель у всех стран держится на одном уровне, за исключением Египта, он сильно отстает во всех годах. Также видно, что в странах нет никаких резких изменений по данному индексу.

4. Индекс медицинского обслуживания:



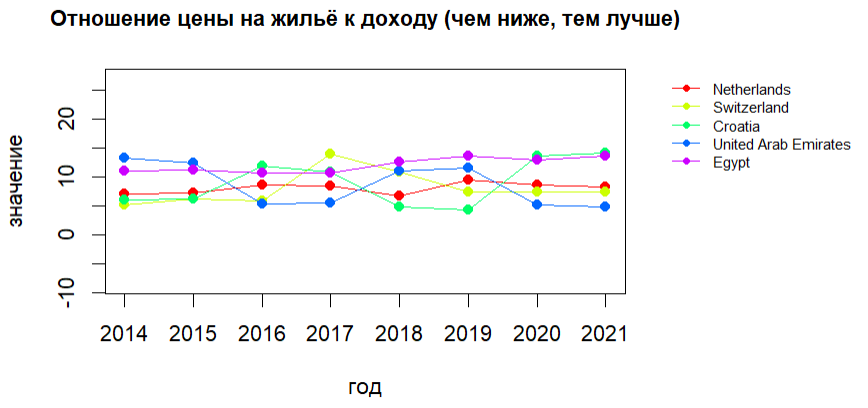
Самый высокий показатель наблюдается у Нидерландов и Швейцарии, немного ниже Хорватия и ОАЭ. Показатель у Египта ухудшается на протяжении лет.

5. Индекс прожиточного минимума:



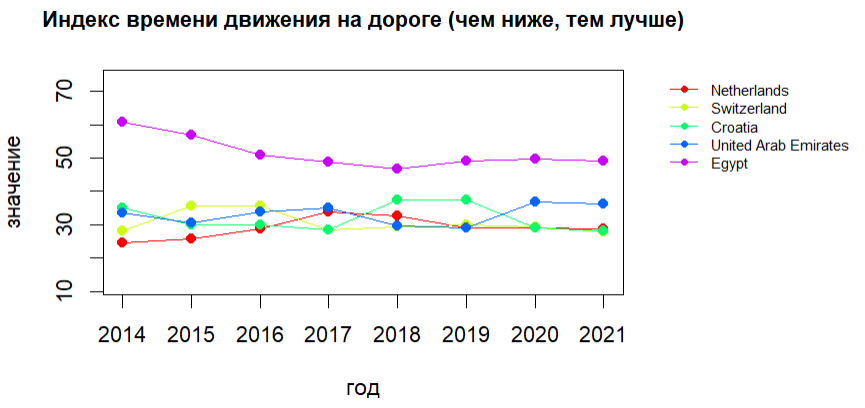
Исходя из графика, можно сделать вывод, что самый высокий индекс наблюдается в Египте на протяжении всех лет. Показатели ОАЭ, Хорватии держатся близко к друг другу. Дороже всего в Швейцарии и Нидерландах.

6. Отношение цены на жильё к доходу:



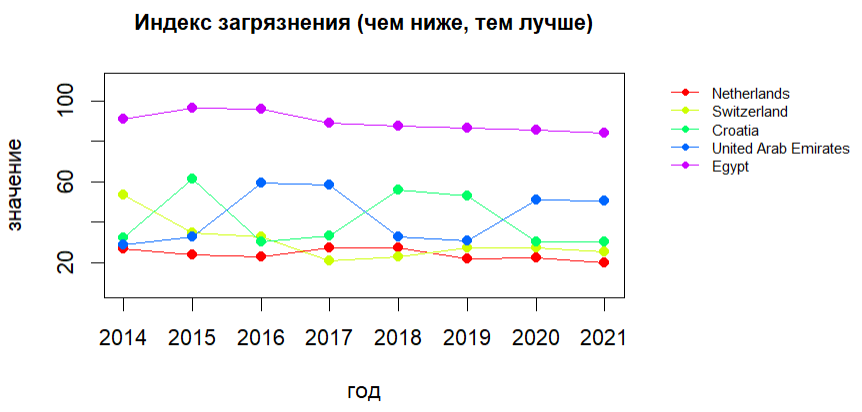
Данный показатель у всех стран находится плюс минус на одном уровне, проще всего купить жилье в Египте, сложнее всего в ОАЭ.

7. Индекс времени движения на дороге:



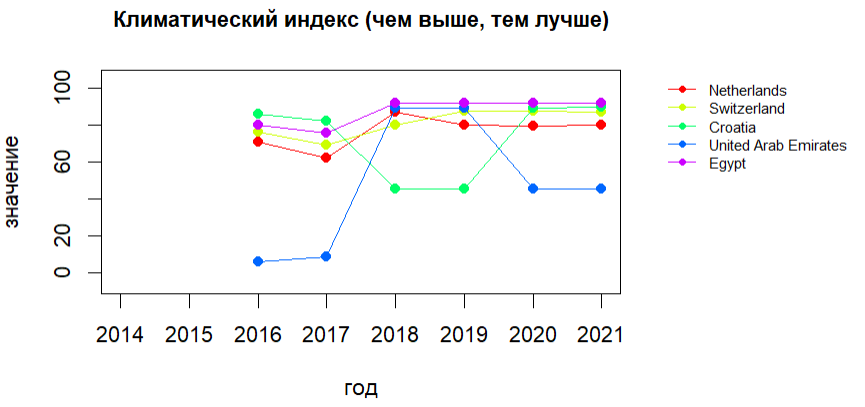
В Египте люди тратят наибольшее количество времени на дорогу среди всех пяти стран. Быстрее всего добираться на транспорте в Швейцарии и Нидерландах.

8. Индекс загрязнения:



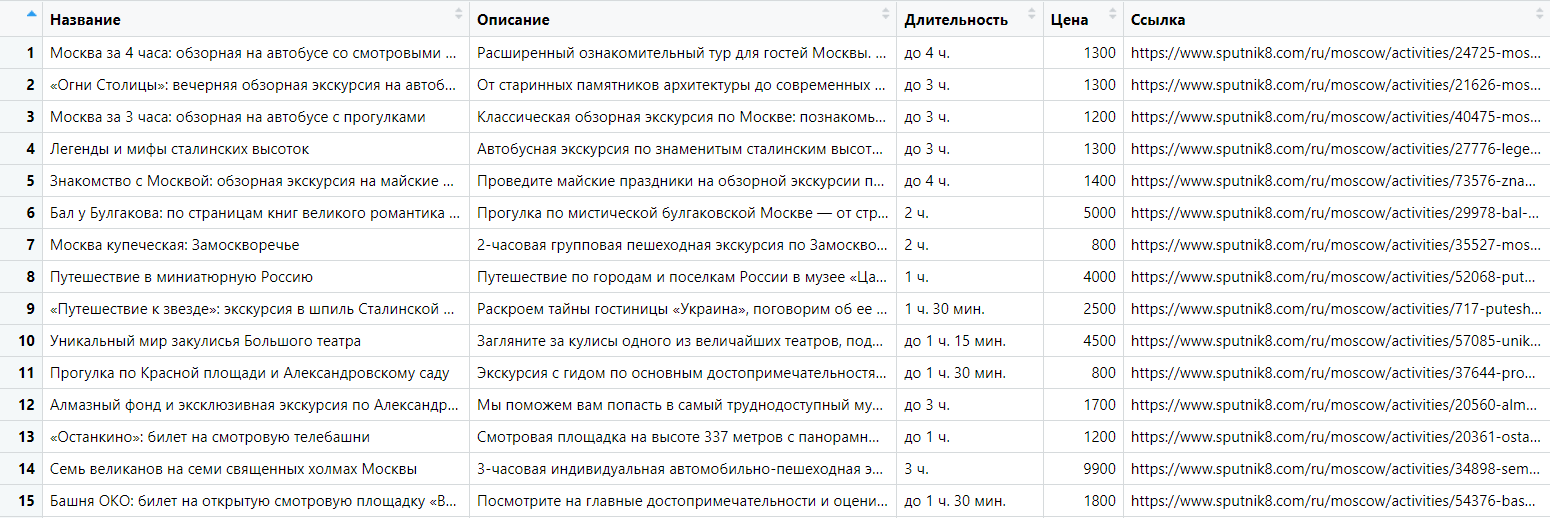
Видно, что наиболее загрязненная страна – Египет, причем отрыв значительный. Наименее загрязненная страна – Швейцария и Нидерланды. Индекс у Хорватии и ОАЭ нестабилен, сильно меняется в зависимости от года.

9. Климатический индекс:



Данные по климату до 2016 года отсутствуют. Наилучший показатель у Египта, наихудший у ОАЭ. Наибольший рост показателя датируется 2018 годом, после этого показатель практически не менялся.

10. Извлечение данных с сайта **https://www.sputnik8.com/ru/moscow:**



**Вывод:** в ходе выполнения лабораторной работы были изучены основы библиотеки rvest для веб-скрейпинга. Библиотека позволяет напрямую работать с данными, представленными на сайтах. В ней имеется удобная возможность использования CSS-селекторов, функция для парсинга HTML таблиц и др. Также были построены графики на основе данных из Интернета, что позволило легко проанализировать разные аспекты жизни нескольких стран.