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. 39/1

Козлов Э.Д.

Проверил: преподаватель

Яхонтов А.А.

Краснодар

2025

**Цель:** Научиться работать извлекать информацию сWEB-страниц с помощью инструментов

языка R.

### **Задание**

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:

Для дополнительной информации по Рейтинг стран по

уровню жизни можно использовать ссылку

https://tyulyagin.ru/ratings/rejting-stran-mira-po-urovnyu-

zhizni-2021.html

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, прокомменитровать в отчете результат.

Необходимо нарисовать на одном и том же графике рейтинг всех 5 стран, проанализировать

резутьтат, анализ словесно отразить в отчете. Проанализировать изменение во времени всех

показателей указанных стран, подобрать наилучший (с вашей точки зрения) способ

визуалазации.

4. С одной из страниц (по варианту) : вариант брать по формуле Noпп % Noссылки +1.

**Ход работы**

Данные:



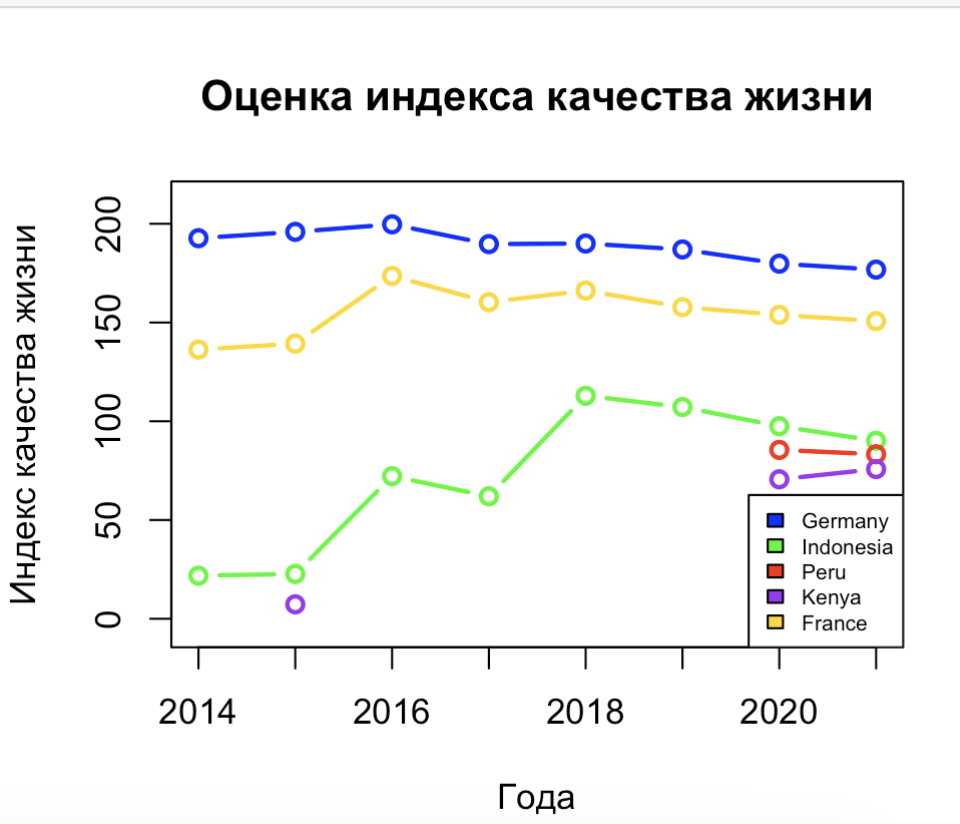


Рисунок 1 - Оценка индекса качества жизни.

Вывод: Германия и Франция лидируют по качеству жизни, Индонезия показывает нестабильный рост, а Перу и Кения существенно отстают.

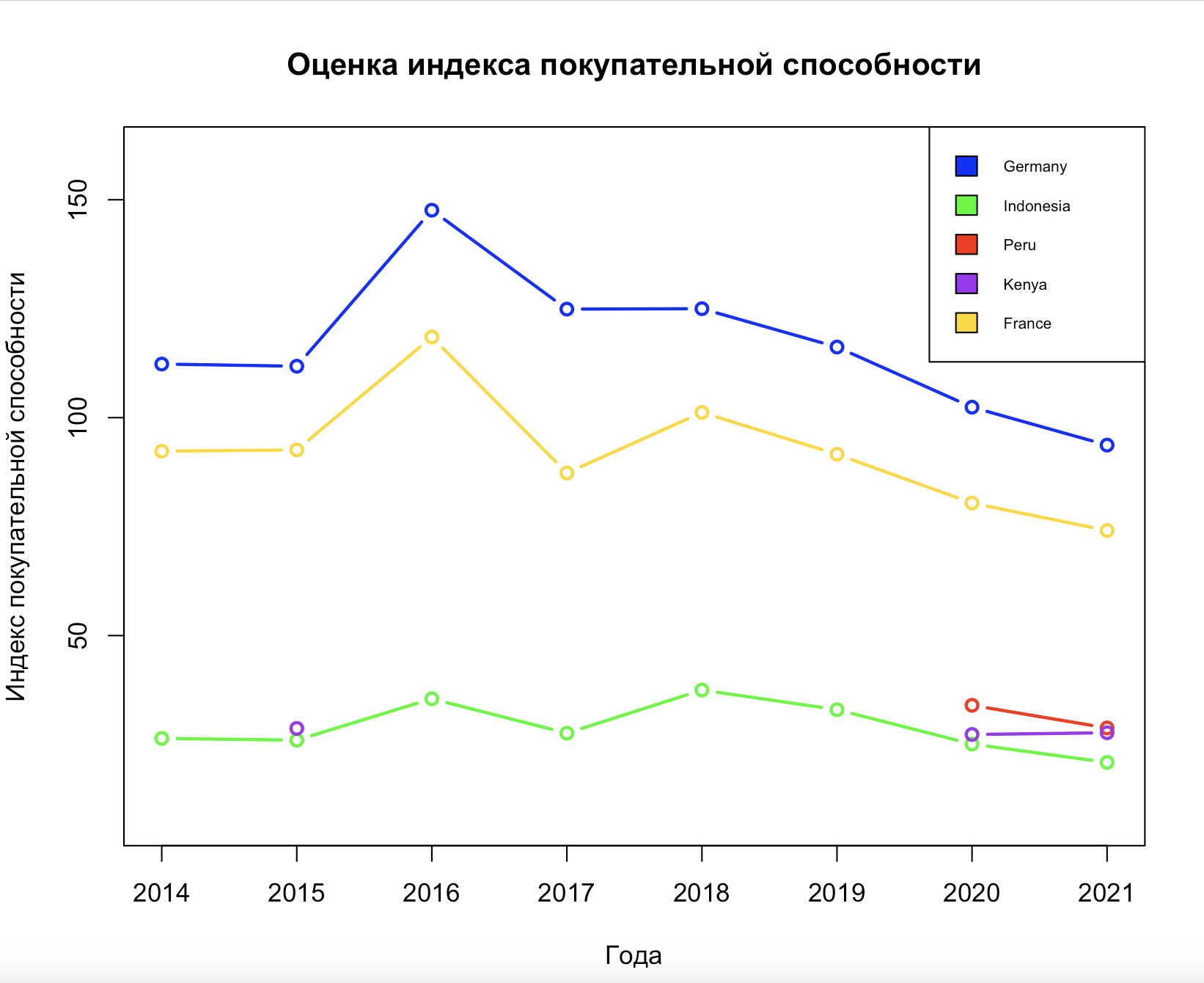


Рисунок 2 – Оценка индекса покупательской способности

Вывод: Германия и Франция имеют наибольшую покупательную способность, несмотря на спад в последние годы. Индонезия — стабильна, а Перу и Кения значительно отстают.

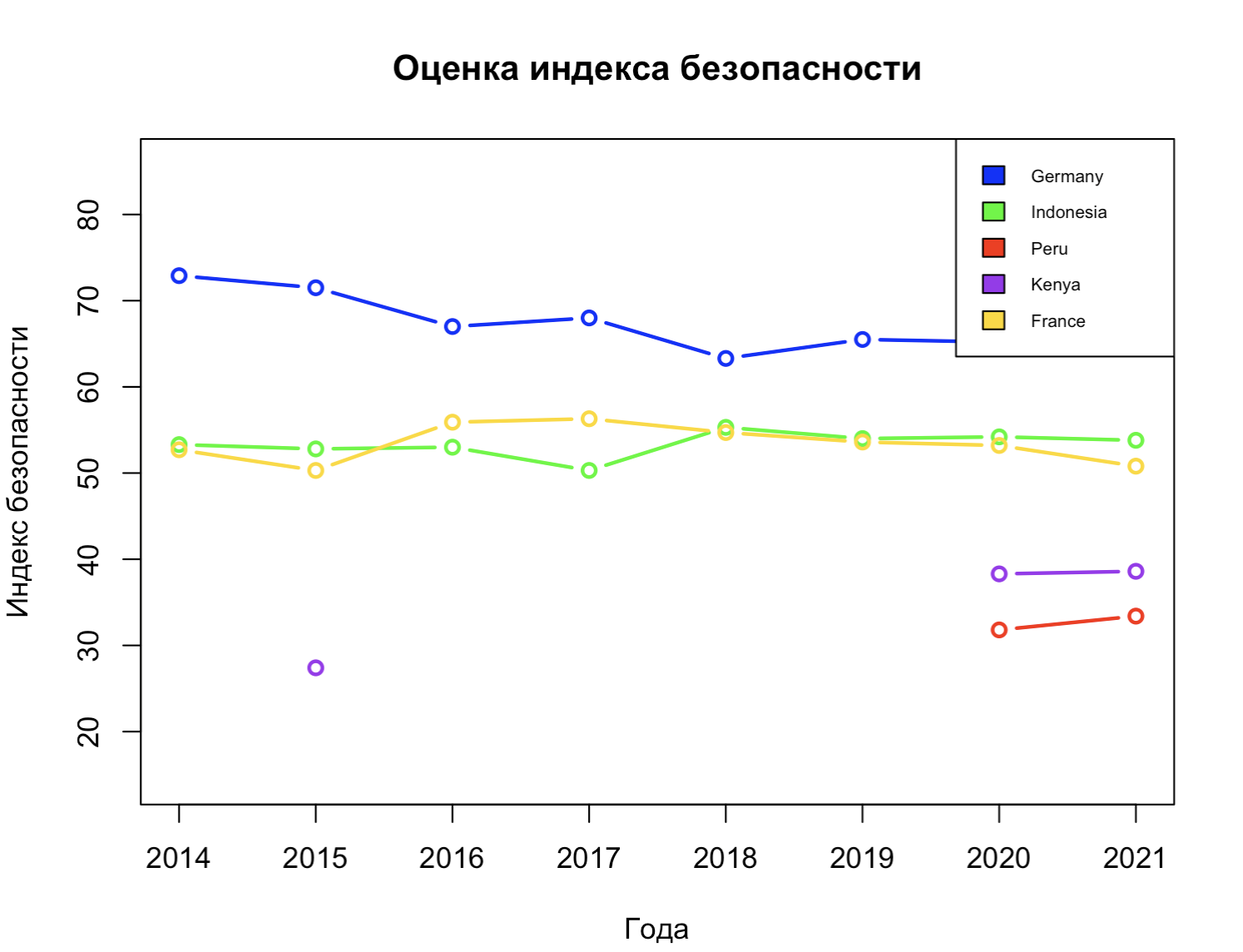


Рисунок 3 – Оценка индекса безопасности

Вывод: Германия — самая безопасная страна из представленных. Франция и Индонезия — средний уровень. Перу и Кения — наименее безопасны, но отмечается положительная динамика.

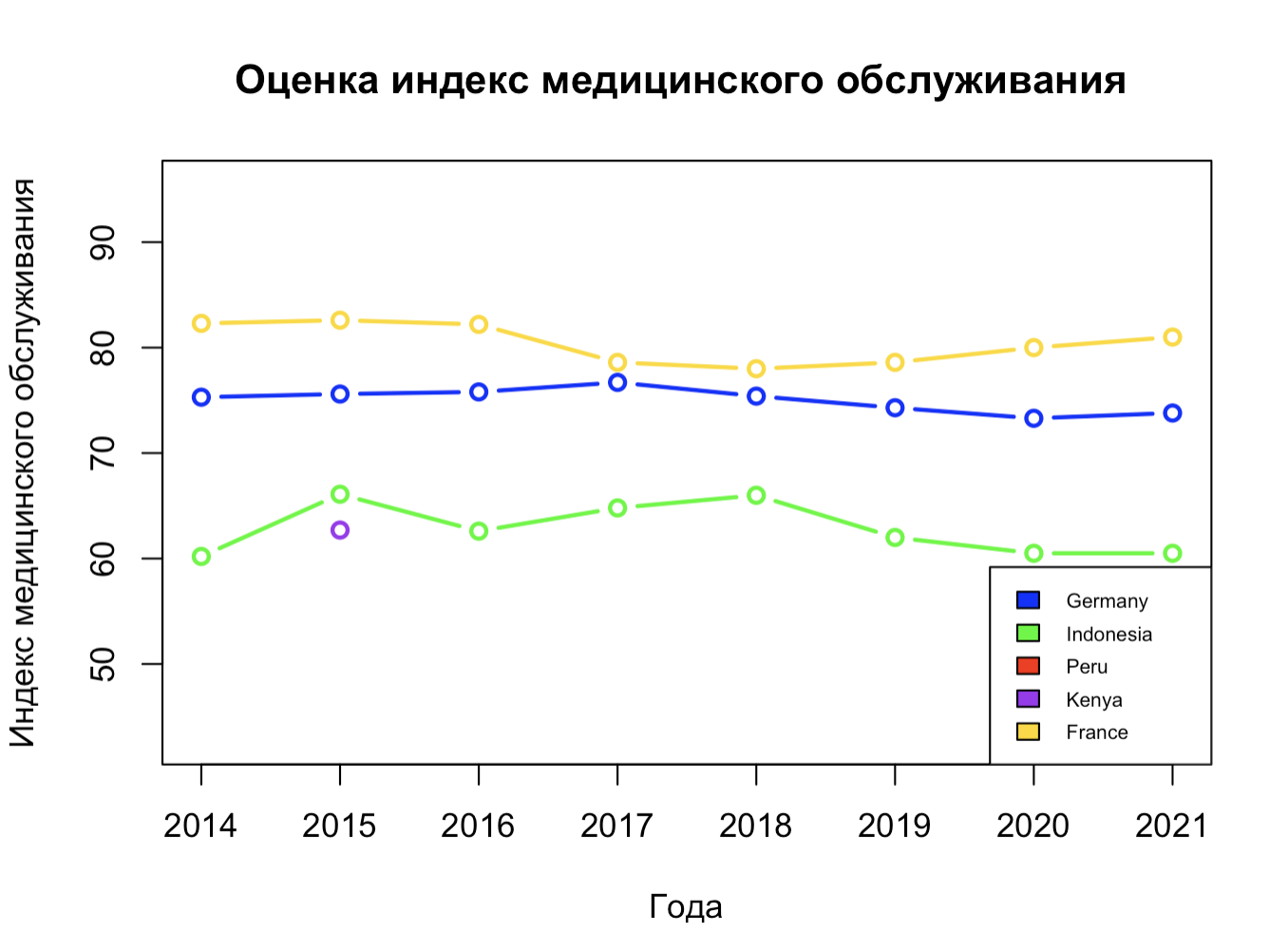


Рисунок 4 – Оценка индекса медицинского обслуживания

Вывод: Лучшее медицинское обслуживание среди пяти стран — во Франции и Германии. В Индонезии — стабильное, но умеренное. Перу и Кения отстают.

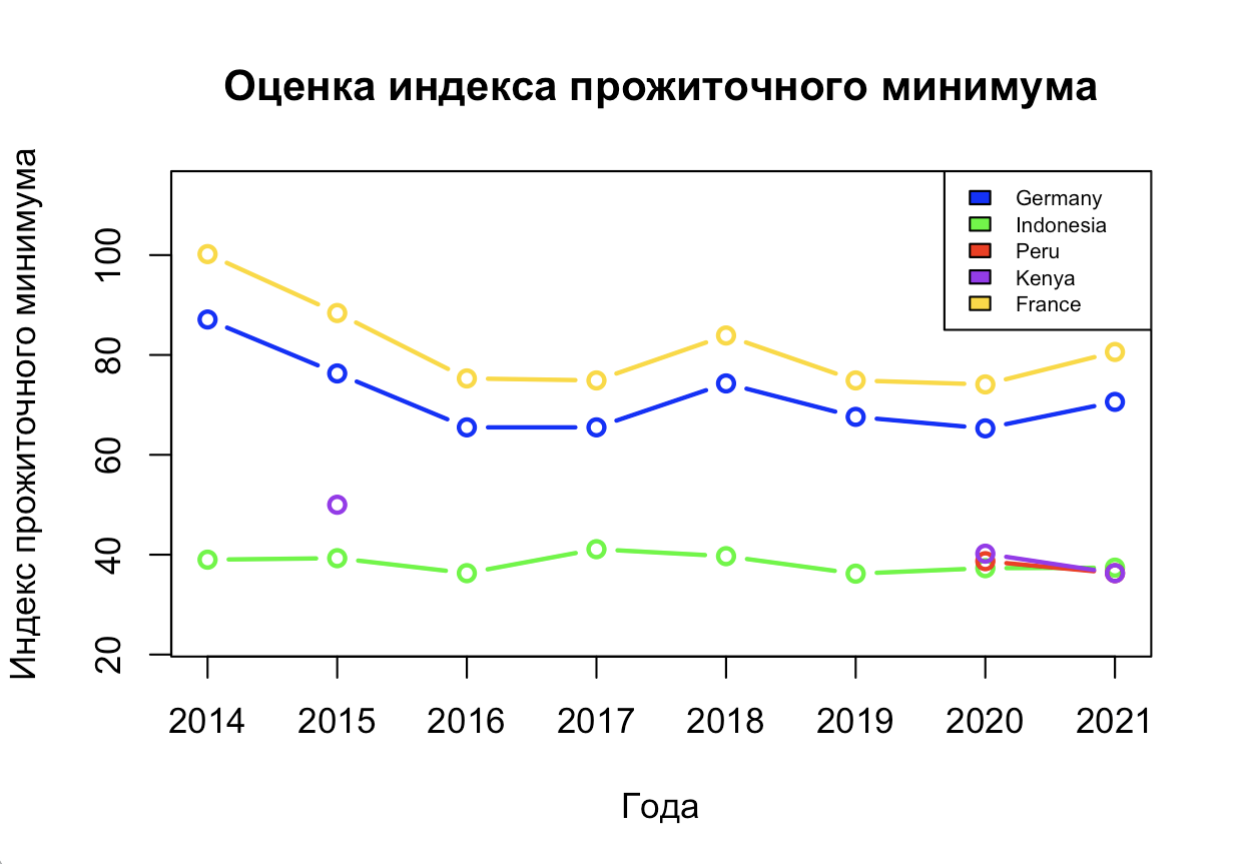


Рисунок 5 – Оценка индекса прожиточного минимума

Вывод: Индонезия, Перу и Кения имеют более низкий прожиточный минимум, что говорит о большей доступности жизни. Франция и Германия демонстрируют снижение индекса, что также позитивная динамика.

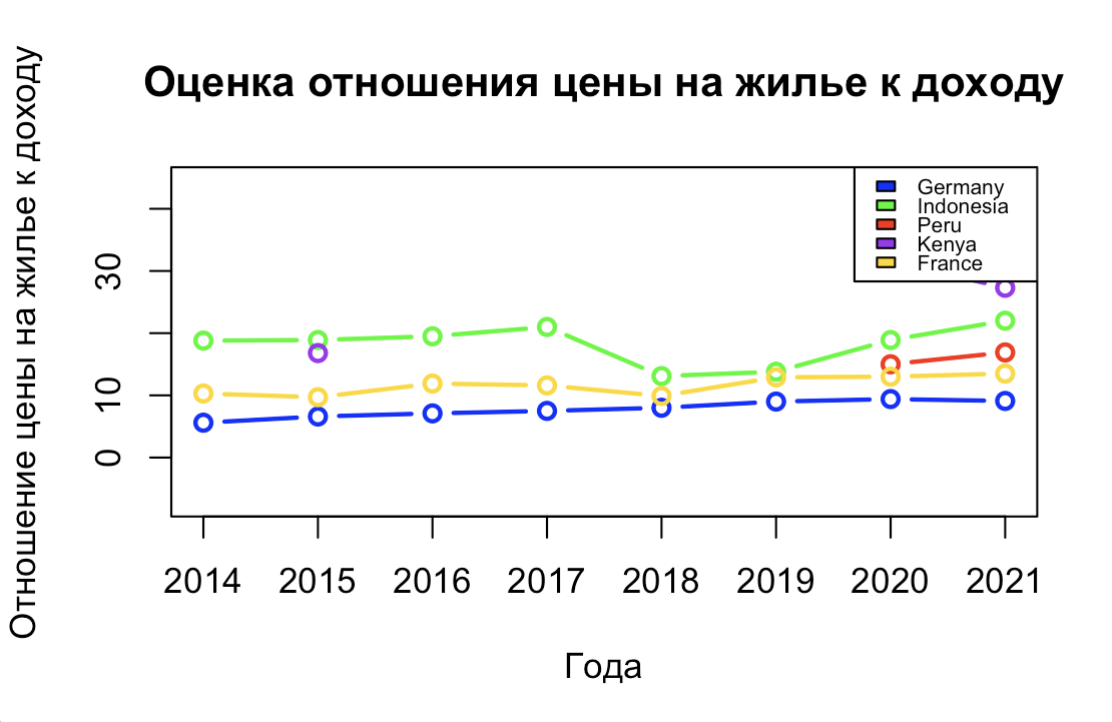


Рисунок 6 – Оценка отношения цены на жилье к доходу

Вывод: Германия — лидер по доступности жилья, в то время как в Индонезии и Кении на жильё уходит сравнительно больше средств, что делает его менее доступным.

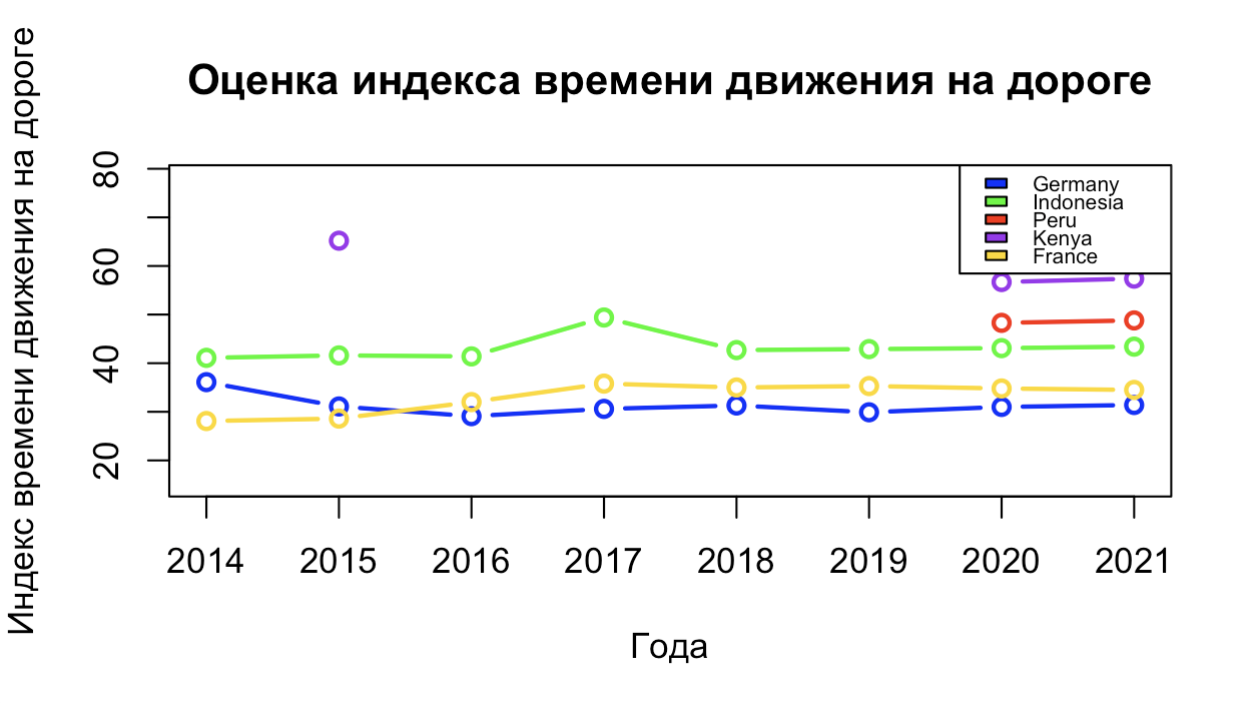


Рисунок 7 – Оценка индекса времени движения на дороге

Вывод: Германия и Франция обеспечивают лучшие условия по времени передвижения, тогда как Кения и Индонезия сталкиваются с серьёзными транспортными затруднениями.

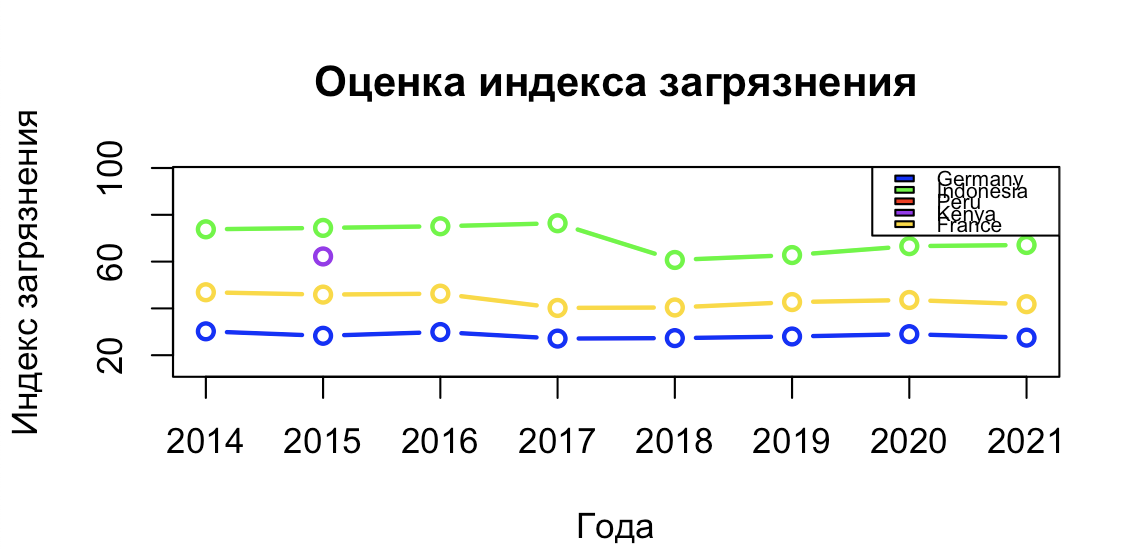


Рисунок 8 – Оценка уровня загрязнения

Вывод: Германия и Франция обладают наилучшей экологической обстановкой, в то время как Индонезия и развивающиеся страны сохраняют высокие уровни загрязнения.

**WEB – скрэпинг**

Начальные данные:

Сайт (по варианту) – (7) <https://www.culture.ru/museums/institutes/location-krasnodarskiy-kray-krasnodar>

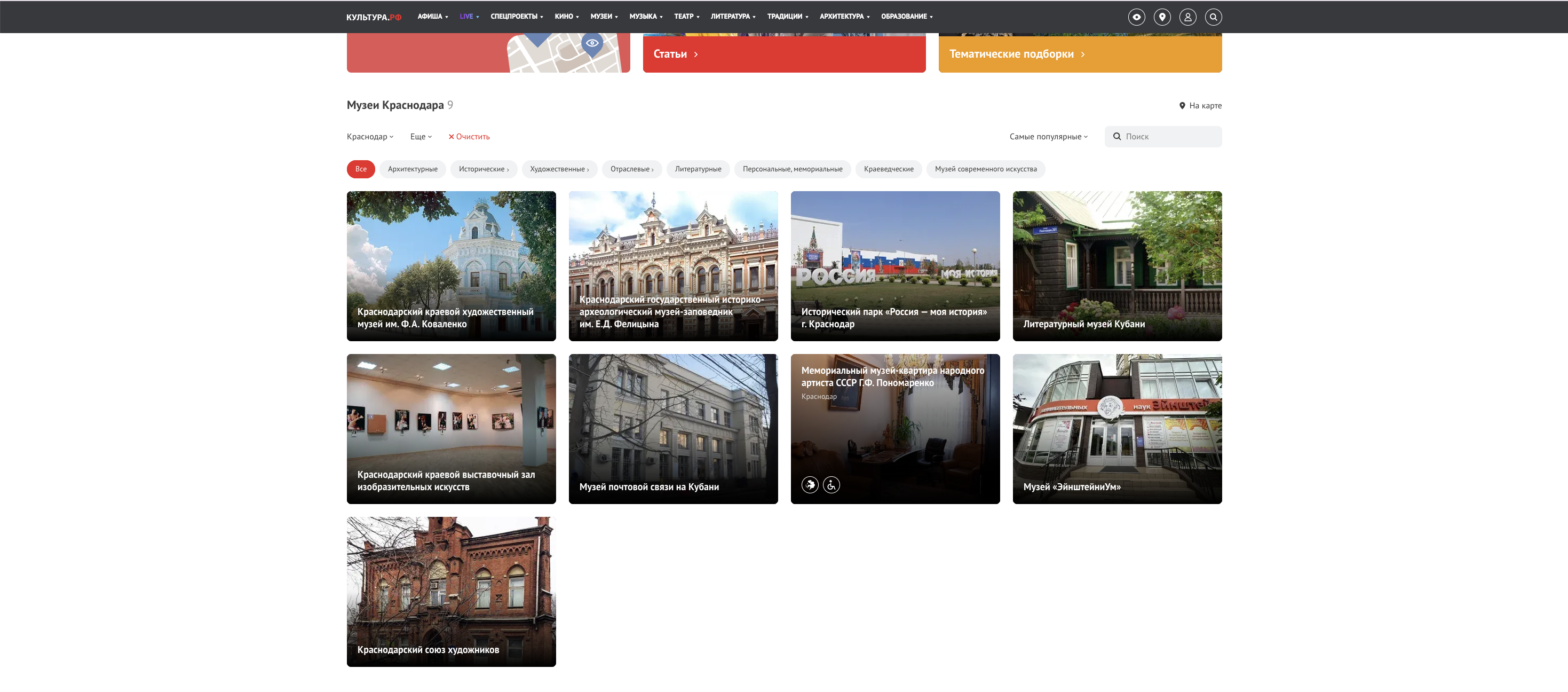


Рисунок 9 – Сайт с музеями

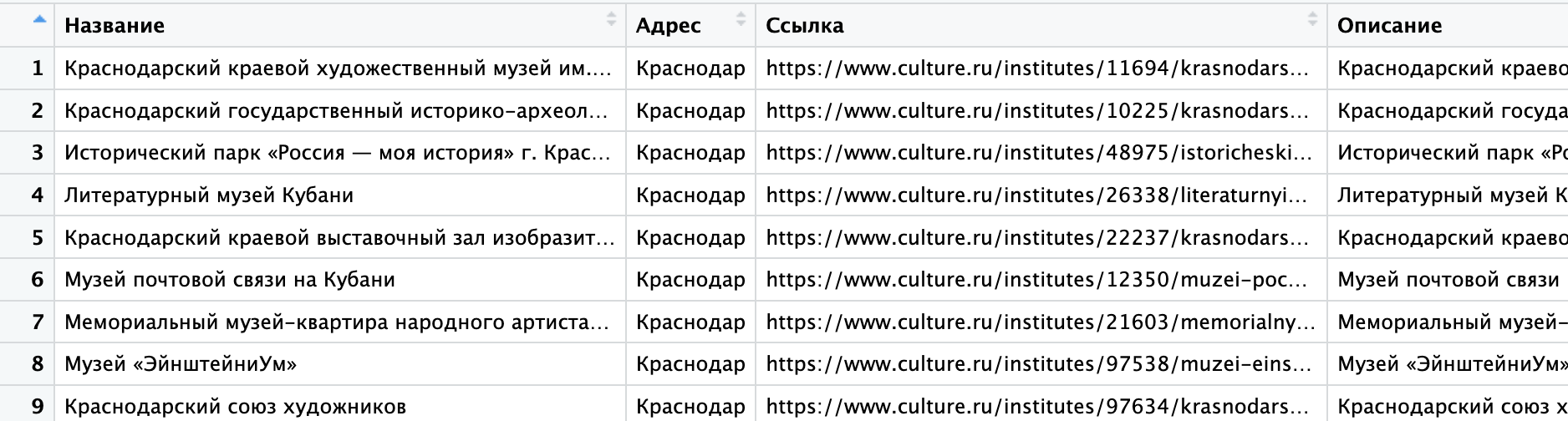


Рисунок 10 – Итоговый dataframe

**Код для скэпинга.**

library(rvest)

page <- read\_html("https://www.culture.ru/museums/institutes/location-krasnodarskiy-kray-krasnodar")

cards <- html\_nodes(page, ".CHPy6")

titles <- c()

addresses <- c()

links <- c()

descriptions <- c()

for (card in cards) {

title <- html\_node(card, ".p1Gbz") %>% html\_text(trim = TRUE)

address <- html\_node(card, ".TxTRy") %>% html\_text(trim = TRUE)

href <- html\_node(card, "a") %>% html\_attr("href")

full\_link <- paste0("https://www.culture.ru", href)

museum\_page <- read\_html(full\_link)

description <- html\_node(museum\_page, "meta[name='description']") %>% html\_attr("content")

titles <- c(titles, title)

addresses <- c(addresses, address)

links <- c(links, full\_link)

descriptions <- c(descriptions, description)

}

df <- data.frame(

Название = titles,

Адрес = addresses,

Ссылка = links,

Описание = descriptions,

stringsAsFactors = FALSE

)

View(df)