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: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д. Н. Баева

Направление подготовки: 02.03.02 Фундаментальная информатика и информационные технологии

Преподаватель: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В. И. Шиян

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

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

**Вариант:** 14

**Страны:** Испания, Италия, Румыния, Греция, Дания.

Устанавливаем пакет rvest в R Studio. Далее создаем документ html из URL, а затем используя функцию из пакета выбираем все элементы таблицы на странице и преобразуем в датафрейм.

Необходимо было выполнить сбор информации об уровне жизни в разных странах согласно варианту в период с 2014 по 2021 года.

Сперва была выполнена оценка индекса качества жизни.

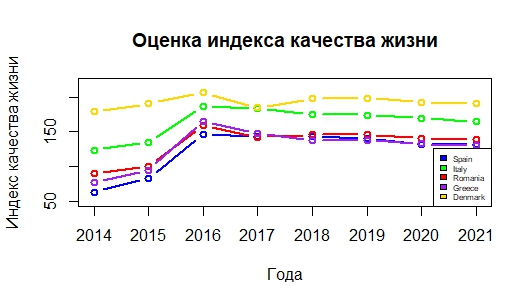


Рисунок 1 – Изменение индекса качества жизни.

Принято считать, чем выше данный показатель, тем лучше. По полученным данным можно сделать вывод, что самый высокий индекс качества жизни в Дании на протяжении всего периода. Самый низкий показатель у Испании. Испания, Румыния и Греция имеют каждый год приблизительно близкие индексы. Однако, начиная с 2016 года у большинства стран наблюдается тенденция роста показателя.

Далее анализируем индекс покупательской способности.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 2 – Изменение индекса покупательской способности.

Принято считать, чем выше данный показатель, тем лучше. По графику видно, что вновь лидирует Дания. Самый низкий показатель у Греции. В 2016 году у большинства стран наблюдался резкий скачок индекса, а затем спад.

Рассмотрим тенденции изменения индекса безопасности.

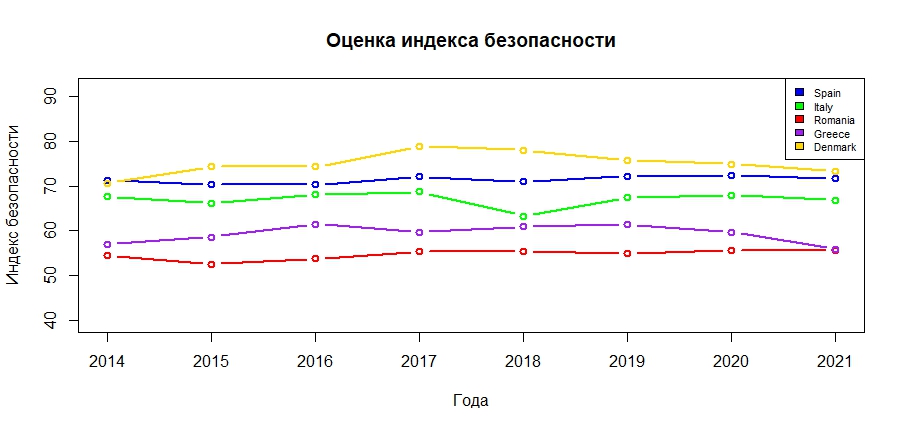


Рисунок 3 – Изменение индекса безопасности.

Анализируя представленный график, можно сказать, у всех стран показатели меняются незначительно и находятся на близких уровнях. Не наблюдаются резкие скачки или падения. Самый высокий показатель, по‑прежнему, у Дании, а самый низкий у Румынии.

Далее будем оценивать индекс здравоохранения.

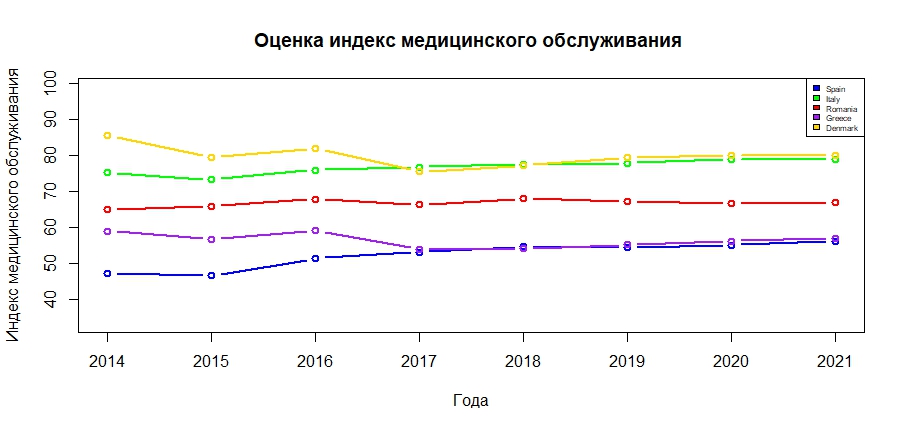


Рисунок 4 – Изменение индекса медицинского обслуживания.

Невооруженным глазом заметно, что у всех стран наблюдается спад индекса. Считается, что чем выше данный показатель, тем лучше. Лидером является снова Дания, а замыкает рейтинг Испания, с небольшим отрывом от Греции.

Следующий критерий — это индекс прожиточного минимума.

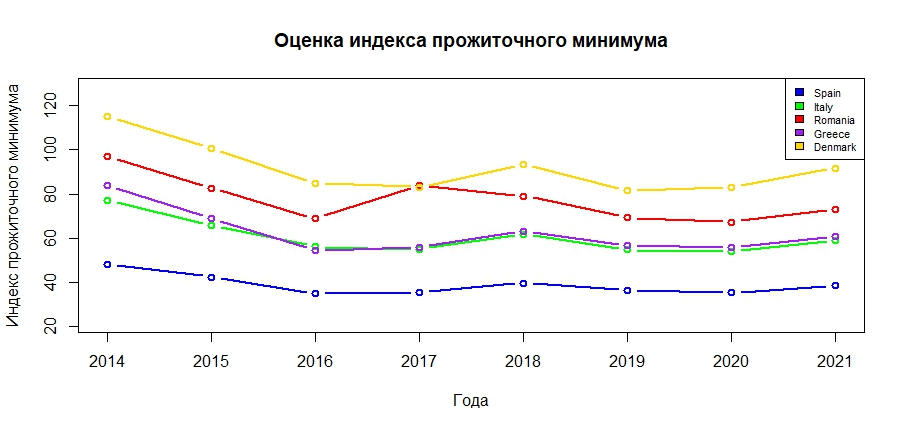


Рисунок 5 – Изменение индекса прожиточного минимума.

Показатель должен быть чем ниже, тем лучше. Отсюда следует, что лидером данного рейтинга является Испания, которая заметно отрывается от других стран. Самый высокий индекс прожиточного минимума у Дании. Остальные страны находятся на одном уровне.

Рассмотрим отношение цены на жилье к доходу.

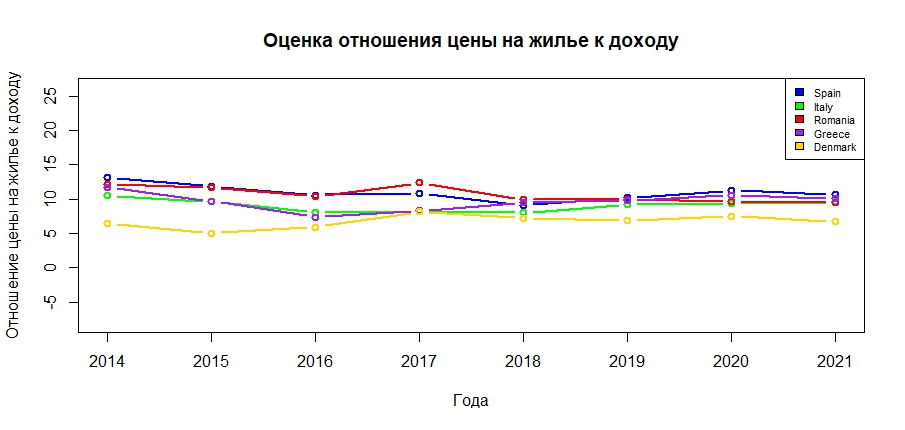


Рисунок 6 – Изменение отношения цены на жилье к доходу.

Начиная с 2017 года, все страны имеют небольшой разрыв по отношению к друг другу. Чем ниже данный показатель, тем лучше. Самым лучшим отношением обладает Дания, а самым худшим –– Румыния.

Далее оцениваем время движения на дороге на транспорте.

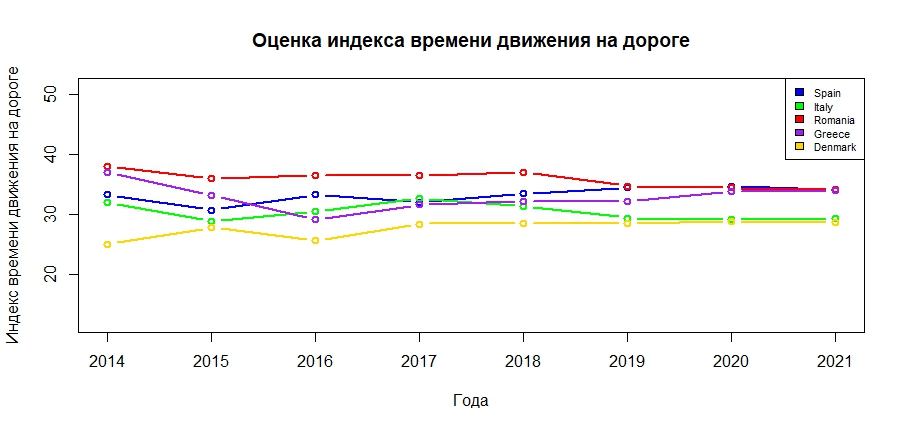


Рисунок 7 – Изменение показателя времени в пути на транспорте.

Исходя из графика и того, что чем ниже показатель, тем лучше, можно однозначно сказать, что больше всего времени уходит на путь в транспорте в Дании. В остальных же странах этот показатель примерно одинаковый.

Затем проанализируем индекс загрязнения.

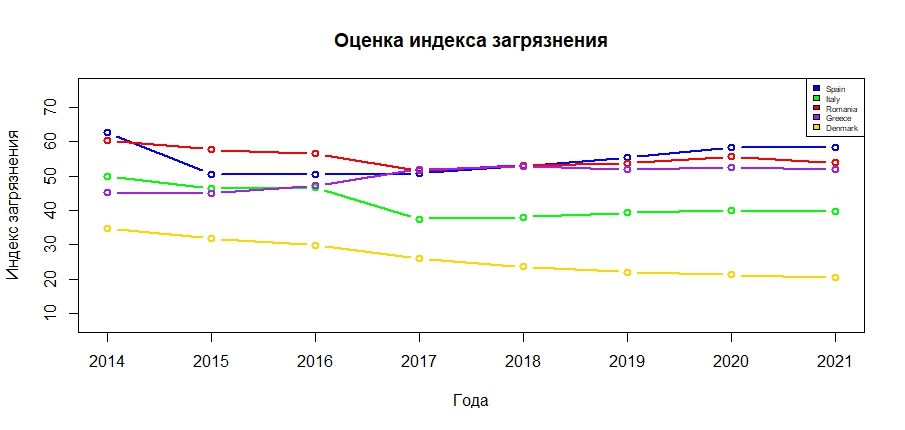


Рисунок 8 – Изменение индекса загрязнения.

Чем ниже индекс, тем лучше. Таким образом, лучшие показатели каждый год принадлежат Дании. Худшие показатели у оставшихся трех стран. Также видно, что в каждой из стран данный показатель меняется с каждым годом незначительно.

Последним критерием является индекс климата. Данные за 2014 год и 2015 год отсутствуют.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 9 – Изменения климатического индекса.

Можно заметить, что показатели всех стран уменьшились в 2017 году и резко возросли в 2018 году, после чего стабильны. Чем выше показатель, тем он лучше. Отсюда, лучший показатель имеет Италия, а худший - Испания.

В следующем задании необходимо было собрать информацию со страницы <https://kudago.com/spb/list/33-luchshih-muzeya-peterburga/> в датафрейм с названием, адресом и ссылкой на всю информацию о данном музее.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 10 – Получившийся датафрейм.

**Вывод:** В данной работе были получены знания по извлечению информации с WEB-страниц с помощью инструментов языка R.