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**

**КУРСОВАЯ РАБОТА**

**АНАЛИТИКА ТУРПОТОКОВ В РОССИИ С ОПРЕДЕЛЕНИЕМ ФАКТОРОВ ВЛИЯНИЯ.**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.М. Нагалевский

(подпись)

Направление подготовки 02.03.02 — «Фундаментальная информатика и\_\_\_\_\_

(код, наименование)

информационные технологии»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_курс\_\_\_\_\_\_\_\_3\_\_\_\_\_\_\_\_

Направленность (профиль) \_\_\_\_\_\_\_\_\_\_\_\_Математическое и программное обеспечение компьтерных технологий\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Научный руководитель

канд. техн. наук, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Т.А. Приходько

(подпись, дата)

Нормоконтролер

ассистент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Е.А. Нигодин

(подпись, дата)

Краснодар

2023

**РЕФЕРАТ**

Курсовая работа 42 стр., 4 ч., 20 рис., 5 источников.

АНАЛИЗ БОЛЬШИХ ДАННЫХ, ТУРПОТОКИ, R.

Объектом исследования являются турпотоки и туриндустрия в России за несколько лет.

Цель работы: целью данной работы является проведение анализа турпотоков в России за определенный период времени с помощью языка R, а также изучение туриндустрии с целью выявления основных тенденций, паттернов и факторов, влияющих на развитие отрасли.

В качестве методов исследования использовались сбор и обработка данных, их анализ, включая статистический, и визуализация данных. В результате работы были изучены динамика въездных и выездных турпотоков в России за последние шесть лет, были выделены популярные регионы и области, которые привлекают туристов. Также были выделены тенденции и факторы в развитии туриндустрии.

Научная новизна работы заключается в том, что она представляет комплексный анализ турпотоков в России за несколько лет, а также исследование туриндустрии и ее финансовых показателей, что позволяет получить глубокое понимание динамики туристической активности и факторов, влияющих на развитие отрасли.

Практическая ценность заключается в том, что полученные результаты и выводы могут быть использованы для разработки стратегий развития туристической индустрии, принятия управленческих решений и планирования мероприятий для привлечения туристов и повышения эффективности туристических организаций в России.

По результатам исследования были выявлены основные тенденции и паттерны в турпотоках в России за рассматриваемый период, определены популярные регионы и области, привлекающие внутренних и въездных туристов, а также проанализирована финансовая состоятельность туриндустрии.

**СОДЕРЖАНИЕ**

[Введение 5](#_Toc135339913)

[1 Теория и методы интеллектуального анализа данных 7](#_Toc135339914)

[1.1 Понятие Data Mining 7](#_Toc135339915)

[1.2 Данные и требования к их обработке 7](#_Toc135339916)

[1.2.1 Подготовка данных 9](#_Toc135339917)

[1.3 Виды данных и критерии их оценивания 10](#_Toc135339918)

[2 Анализируемая сфера исследования — турпотоки в России. 13](#_Toc135339919)

[2.1 Виды туристических потоков 14](#_Toc135339920)

[2.2 Туриндустрия 14](#_Toc135339921)

[3 Анализ турпотоков 16](#_Toc135339922)

[3.1 Въездные турпоездки 16](#_Toc135339923)

[3.2 Выездные турпоездки 19](#_Toc135339924)

[3.3 Внутренние потоки 22](#_Toc135339925)

[3.4 По сезонам 26](#_Toc135339926)

[4 Туриндустрия 29](#_Toc135339927)

[4.1 Организации (прибыльные, убыточные) 29](#_Toc135339928)

[4.2 Выручка, прибыль, убыток и НФР туристических организаций 31](#_Toc135339929)

[4.3 Ввод в действие объектов туризма 33](#_Toc135339930)

[Заключение 35](#_Toc135339931)

[Список использованных источников 36](#_Toc135339932)

[Приложение А Листинг программы 37](#_Toc135339933)

# ВВЕДЕНИЕ

Актуальность данной работы состоит в том, что анализ данных играет все более важную роль в различных сферах человеческой деятельности, включая туризм. С развитием информационных технологий и доступностью больших объемов данных, становится возможным проводить глубокий и комплексный анализ туристических потоков и факторов, влияющих на развитие туристической индустрии. При этом важно учитывать, что история анализа данных имеет долгую и богатую историю, начиная от примитивных статистических методов до современных подходов, основанных на машинном обучении и искусственном интеллекте.

Вместе с тем, современные технологии сбора, хранения и анализа данных открывают новые возможности для изучения и понимания туристической деятельности. Анализ данных позволяет выявить тенденции, зависимости и факторы, влияющие на развитие туризма, а также определить эффективные стратегии и меры для развития туристической индустрии.

В настоящее время исследования и разработки в области анализа данных представляют собой активное и важное направление в информационных технологиях. С начала 2000-х годов наблюдается значительный рост интереса к анализу данных и применению современных методов машинного обучения и статистического анализа для извлечения ценной информации из больших объемов данных. Методы анализа данных находят применение в различных сферах, включая бизнес, науку, медицину, финансы, социальные исследования и другие. Они позволяют обнаруживать скрытые закономерности, выявлять зависимости и тренды, классифицировать и кластеризовать данные, делать прогнозы и принимать решения на основе данных.

По учебным планам направление ФИИТ изучает BigData и анализ данных на 3 курсе бакалавриата, поэтому было принято решение глубже углубится в эту сферу и изучить как различные методы анализа данных, так и язык R.

Основная цель работы – проведение комплексного анализа данных с использованием современных методов и инструментов анализа данных.

Для реализации поставленной цели предполагается решить следующие задачи.

1. изучить принципы работы анализа данных.
2. изучить анализ данных на языке R.

Объектом исследования в работе являются данные, связанные с турпотоками и туристической индустрией.

Предметом исследования являются различные аспекты и характеристики турпотоков и туристической индустрии.

Информационная база исследования включает в себя несколько видов учебных материалов таких как книги по анализу данных, статьи от авторитетных сервисов и документация. В качестве методов исследования использовался сбор и обработка данных, их анализ, включая статистический и визуализация данных.

Научная новизна работы заключается в том, что она представляет комплексный анализ турпотоков в России за несколько лет, включая въездные и выездные потоки, а также исследование туриндустрии и ее финансовых показателей, что позволяет получить глубокое понимание динамики туристической активности и факторов, влияющих на развитие отрасли.

Практическая значимость исследования курсовой работы заключается в результатах исследования, которые могут быть использованы студентами при изучении туриндустрии и дальнейший анализ этих данных, для выявления тенденций и стратегий развития туризма в России.

Практическая ценность работы заключается в том, что она предоставляет детальный анализ данных и статистики путешествий и туризма в России.

# Теория и методы интеллектуального анализа данных

## 1.1 Понятие Data Mining

Data mining (сбор данных) — мультидисциплинарная область, возникшая и развивающаяся на базе прикладной статистики, распознавание образов, искусственного интеллекта, теории баз данных, визуализации, машинного обучения и других дисциплин. Разберем некоторые из них более подробно.

Технология Data Mining заключается в поиске неподготовленных, безызвестных ранее данных, а также полезных и доступных к использованию знаний для принятия решений. Так как данная технология предназначена для функционирования в глобальных базах данных, то на практике она имеет несколько закономерностей, а именно: неочевидных, объективных и практически полезных.

Неочевидные закономерности подразумевают под собой закономерности, которые нельзя обнаружить стандартным путем анализа.

Объективные — данные закономерности полностью соответствую действительности, чего нельзя сказать об экспертном мнении, которое является субъективным.

Практически полезные закономерности говорят нам о том, что конкретным значениям можно найти прямое практическое применение.

## 1.2 Данные и требования к их обработке

Данные — это факты, тексты, графики числовые выражения, документы, и другого рода сегменты. Все это может быть получено в результате практической деятельности. Данные должны быть конвертируемыми, то есть, способны менять свою форму для необходимого типа хранения. Можно сказать, данные — набор необработанной информации.

Данные включают в себя 2 категории: объекты и атрибуты. Объектом может являться просто строка чисел. Атрибутами могут являться различные характеристикам объекта, то есть, это своего рода переменная. В результате перехода из неопределенного состояния в определенное получается набор переменных изучаемого понятия. Обратим внимание на понятие переменной — это общее свойство всех объектов, которое может меняться от перехода от одного к другому. Главным значением переменной является наличие признака.

При анализе базы данных обычно нет возможности рассмотреть общую совокупности объектов в виду огромного потока данных, но вполне достаточно рассмотреть лишь некоторую ее часть. Размер данной части должен зависеть от разнообразия имеющихся объектов, представленных в генеральной совокупности. Генеральная совокупность — совокупность элементов необходимая исследователю. Выборка — часть генеральной совокупности, которая была отобрана по определенным параметрам. А параметры — числовые характеристики.

Также, наряду с закономерностями существует и определенная специфика требований к обработке данных:

1. неограниченный объем,
2. разнородность,
3. корректные результаты,
4. простые инструменты для обработки.

Традиционный процесс сбора данных состоит из следующих пунктов:

1. анализ предметной области,
2. постановка задачи,
3. подготовка данных,
4. построение моделей,
5. проверка и оценка,
6. выбор,
7. применение,
8. коррекция и обновление.

Анализ предметной области включает в себя отделение значительных и незначительных факторов, влияющих на проводимое исследование. Существенность выбранных данных зависит от первоначально поставленной задач. В процессе ее исследования мы составляем ее элементарную модель. Она может включать в себя специализированные графические нотации, диаграммы потоков и различные структурные анализы. Модель в свою очередь, описывает процессы, происходящие в области и данные, которые для этих процессов используются. От того, насколько корректно составлена модель, зависит успех дальнейшего функционирования приложения Data Mining.

### 1.2.1 Подготовка данных

Подготовка данных является важнейшим этапом, который влияет на конечный результат всех процессов. Данный этап включает в себя несколько шагов:

1. Определение и анализ требований к данным. Это и есть непосредственно создание модели данных, которые необходимы для функционирования Data Mining. При этом изучаются все вопросы распределения данных, доступа к ним, необходимость дополнительных источников и общие аналитические характеристики системы.
2. Сбор данных. Наличие централизованного хранилища данных весомо облегчает работу с ними. Также, это значительно дешевле, чем использование отдельных витрин и баз данных. Конечно, использование централизованного хранилища не всегда предоставляется возможным, тогда используют архивную систему базы данных, что так же значительно проще работы с отдельными частями. Если данные уже упорядочены, то следует определить имею ли они сезонность или цикличность компонентов. Если нет, то есть, компоненты не связаны временным диапазоном и не имеют хронологического порядка, то необходимо иметь в виду, что: недостаточное количество записей в наборе может стать причиной некорректного функционирования составленной модели, возможно, что некоторые из данных являются устаревшими и их нужно исключить из имеющегося списка, алгоритмы, которые используются для моделирования должны быть масштабируемыми, должно быть определенное соотношение входных переменных и наблюдений, а количество записей должно значительно их превышать, сам набор данных должен быть широкого спектра всевозможных ситуаций, чтобы иметь свойство адаптивности.

Предварительная обработка данных включает в себя анализ данных. Анализировать можно как качественные, так и некачественные данные, но результат будет достигнут и в том, и в другом случае. Чтобы разграничить данные существует критерий качества.

## 1.3 Виды данных и критерии их оценивания

Качество данных — критерий, определяющий полноту и точность предоставляемых данных, а также, возможность их интерпретации. Данные высокого качества — это полные и точные данные, обеспечивающие исследователю на их фоне составить полномерную корректно функционирующую модель способную к интерпретации и принятию решений. Данные низкого качества, соответственно, таких характеристик не имеют, включая неверный формат.

Наиболее распространенные виды так называемых «грязных» данных:

1. пропущенные значения,
2. дубликаты,
3. шумы и выбросы.

Пропущенные значения. Некоторые из значений могут быть пропущены из-за того, что данные попросту не были собраны или некоторые атрибуты априори не могут быть применены к данным.

Дублирование данных. Сами по себе дубликаты являются записи имеющие одинаковую информацию по всем параметрам. Есть всего два варианта обработки дубликатов: удаление всех дубликатов (только в случае обесценивании информации) или замена группы в одну уникальную запись.

Шумы и выбросы представляют собой особую группу характеристик некачественных данных. Выбросы — абсолютно отличающиеся от остальных данных объекты, по типу данных и виду их формата. Они могут быть как самостоятельными, так и иметь целую объединенную группу. Их важно не только обнаружить, но и оценить их дальнейшее влияние на данные. Различные методы интеллектуального анализа имеют различную чувствительность к шумам и выбросам. С ними нужно быть особенно осторожными, так как обработка таких данных не может нести в себе ни полезности, ни надёжности.

Некачественные данные подлежат очистке, то есть происходит выявление и удаление ошибок. Метод должен удовлетворять определенным критериям:

* 1. удаление всех основных ошибок,
  2. поддержание определенных инструментов,
  3. не затрагивать основные данные,
  4. функции очистки данных должны быть адаптивными и к другим базам данных,
  5. надёжность выполнения всех этапов.

Этапы очистки данных включают в себя 5 шагов: анализ данных, определение порядка и правил преобразования, подтверждение, преобразования и противоток очищенных данных.

* 1. Подтверждение. На данном этапе определяется эффективность предыдущего этапа. Это происходит путем тестирования копии источника.
  2. Преобразования. Просто обновление хранилища данных при ответе на запросы множества источников.
  3. Противоток очищенных данных. После очистки данных нужно произвести замену данных на новые, качественные, во всех возможных источниках и в хранилище непосредственно. Этот процесс требует больших объемов метаданных.

# 2 Анализируемая сфера исследования — турпотоки в России.

Турпотоки в России представляют собой потоки туристов, как въезжающих в страну, так и выезжающих из нее. Россия, с ее разнообразными природными достопримечательностями, культурным наследием и историческими местами, привлекает множество туристов со всего мира.

Анализ турпотоков в России является важным исследовательским направлением, поскольку позволяет понять и оценить динамику и характер туристического потока в разных регионах страны. Это помогает выявить популярные направления и места, наиболее востребованные виды туризма, а также понять предпочтения и поведение туристов.

Анализируя турпотоки, исследователи могут выявить сезонные колебания в притоке туристов, анализировать изменения в предпочтениях и тенденциях, а также оценить экономическую значимость туризма для определенных регионов. Эти данные помогают разработать эффективные стратегии развития туристической индустрии, принимать решения по улучшению инфраструктуры и продвижению туристических направлений.

Кроме того, анализ турпотоков в России также может включать изучение мотиваций и предпочтений туристов, их потребностей и ожиданий, что помогает создать более персонализированные и качественные туристические услуги.

В целом, анализ турпотоков в России позволяет получить глубокое понимание динамики туристической деятельности в стране, выявить факторы успеха и вызовы, а также разработать эффективные стратегии для развития туризма в России.

## 2.1 Виды туристических потоков

Виды турпотоков включают различные направления и характеристики движения туристов. В контексте анализа турпотоков в России выделим несколько основных видов:

1. Въездные турпотоки: Это потоки туристов, приезжающих в Россию из других стран. Въездные турпотоки могут быть связаны с различными целями путешествия, включая туризм, деловые визиты, культурный обмен и другие.
2. Выездные турпотоки: Это потоки туристов, покидающих Россию и отправляющихся в другие страны. Выездные турпотоки также отражают предпочтения и потребности российских туристов.
3. Внутренние турпотоки: Это потоки туристов, путешествующих внутри России. Внутренние турпотоки охватывают поездки российских граждан по различным регионам страны, включая популярные курорты, исторические и культурные достопримечательности, природные парки и другие места.

Каждый из этих видов турпотоков имеет свои особенности и требует отдельного анализа. Анализ различных видов турпотоков помогает выявить тенденции, предоставляет информацию для принятия управленческих решений в сфере туризма, а также способствует разработке мер по привлечению и удержанию туристов.

## 2.2 Туриндустрия

Туристическая индустрия включает в себя разнообразные организации и предприятия, которые осуществляют деятельность в сфере туризма. В контексте исследования турпотоков, анализ туристической индустрии позволяет оценить ее экономическую эффективность, выявить прибыльные и убыточные организации, а также оценить общую динамику развития этой отрасли. Рассмотрим некоторые аспекты туристической индустрии.

1. Прибыльные и убыточные организации: Анализ прибыльности туристических организаций позволяет определить их финансовую стабильность и успех на рынке. Прибыльные организации генерируют положительную прибыль, в то время как убыточные организации терпят финансовые потери.
2. Выручка: Анализ выручки в туристической индустрии позволяет оценить общий объем денежных средств, полученных от туристических услуг и продуктов. Это включает доходы от проживания, питания, транспортировки, экскурсий и других туристических активностей.
3. Прибыль и убыток: Оценка прибыли и убытка туристических организаций является ключевым аспектом анализа их финансового состояния. Прибыль является разницей между выручкой и затратами, показывая финансовый успех организации. Убыток, напротив, свидетельствует о негативных финансовых результатах.
4. НФР (Сальдированный финансовый результат): НФР включает в себя инвестиции в развитие и поддержку туристической инфраструктуры, включая строительство, ремонт и обновление объектов туризма.
5. Ввод в действие объектов туризма: Анализ ввода в действие объектов туризма оценивает динамику развития туристической инфраструктуры и предоставляемых услуг. Это включает оценку строительства и открытия новых отелей, гостиниц, курортных комплексов, музеев, парков и других объектов, которые привлекают туристов.

Анализ туриндустрии и ее компонентов является важной составляющей исследования турпотоков. Он позволяет оценить экономическую эффективность отрасли, выявить успешные бизнес-модели, идентифицировать факторы успеха и определить меры для улучшения развития туристической индустрии в целом.

# Анализ турпотоков

Сейчас мы рассмотрим основной анализ въездных и выездных турпотоков, что является ключевым аспектом исследования. Анализ турпотоков позволяет получить важную информацию о посещаемости и мобильности туристов, их предпочтениях и тенденциях. Мы рассмотрим динамику въездных и выездных туристических потоков за определенный период, исследуем факторы, влияющие на эти потоки. Полученные результаты позволят нам лучше понять особенности туристической активности и внести ценные выводы для разработки стратегий и принятия решений в сфере туризма.

## 3.1 Въездные турпоездки

Теперь рассмотрим отдельно въездные турпоездки, которые представлены на следующем графике (рисунок 1). График отражает динамику количества въезжающих туристов с 2014 по 2022 года. Анализ этой информации позволит нам выявить тенденции и изменения в потоке въезжающих туристов, исследовать факторы, влияющие на их приезд.

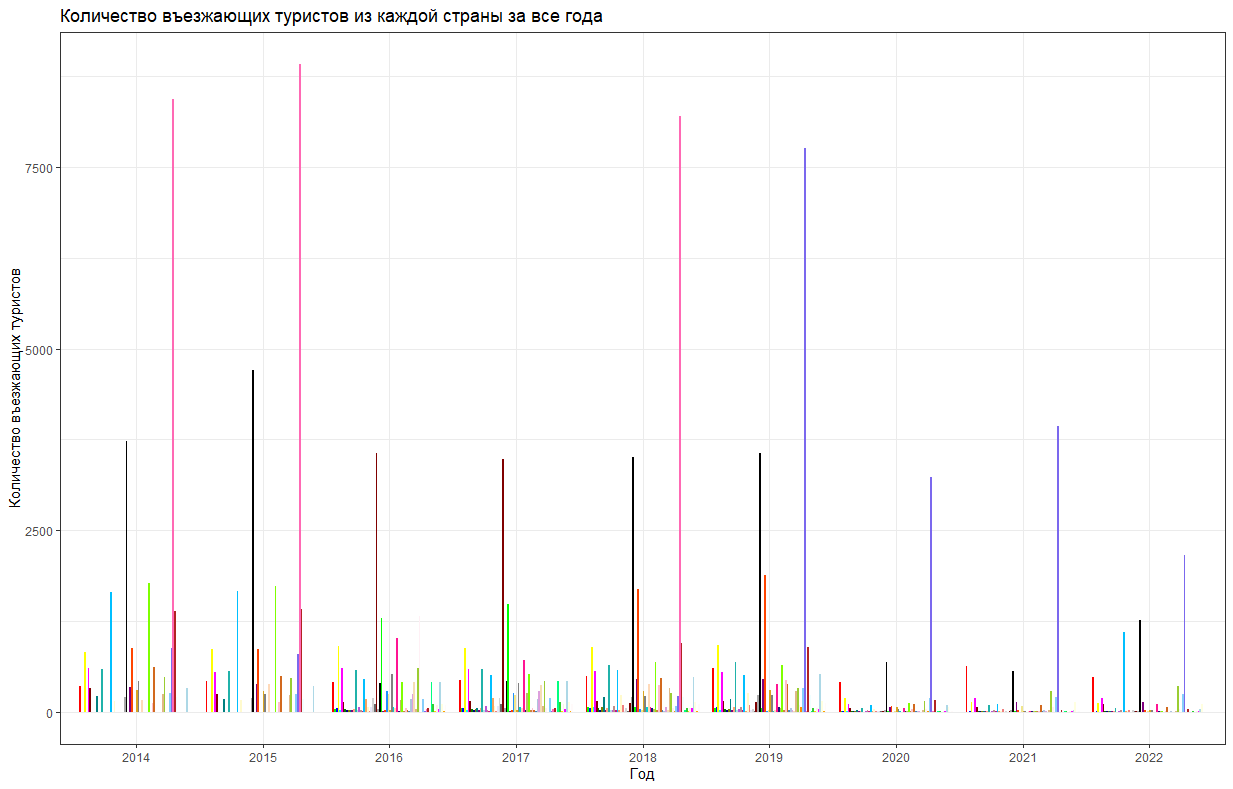


Рисунок 1 - Количество въезжающих туристов из каждой страны за последние 9 лет.

На рисунке 2 изображена легенда для данного графика.



Рисунок 2 – Легенда для графика въезжающих туристов из каждой страны за последние 9 лет.

На графике представлена визуализация количества въезжающих туристов в Россию, и заметно, что ведущие роли в этом показателе играют такие страны, как Украина, Казахстан и Узбекистан. Их значительный вклад в общий поток туристов подчеркивает их важность как источников въездных турпоездок в страну.

Украина, на первом месте среди въезжающих туристов, привлекает внимание своей географической близостью, культурными и историческими связями с Россией, а также благодаря разнообразным туристическим предложениям, таким как популярные курорты, исторические достопримечательности и широкий выбор развлечений. К сожалению, в связи с последними событиями, в 2022 году количество въезжающих с Украины серьезно сократилось.

Казахстан, занимающий второе место, является важным туристическим партнером России, обладая значительным потенциалом для развития взаимных туристических связей. Благоприятные экономические и культурные отношения между двумя странами способствуют активным туристическим пересечениям, включая посещение достопримечательностей, бизнес-туризм и шопинг.

Узбекистан, занимающий третье место, благодаря развитию туристической индустрии и улучшению транспортных связей с Россией, жители Узбекистана становятся все более частыми гостями в страну. Конечно же, не малую роль играет жители, которые приезжают в Россию на заработки и, в большинстве случае, остаются на постоянное место жительство.

Также на рисунке 3 мы можем наблюдать общее количество приезжих за последние 9 лет.

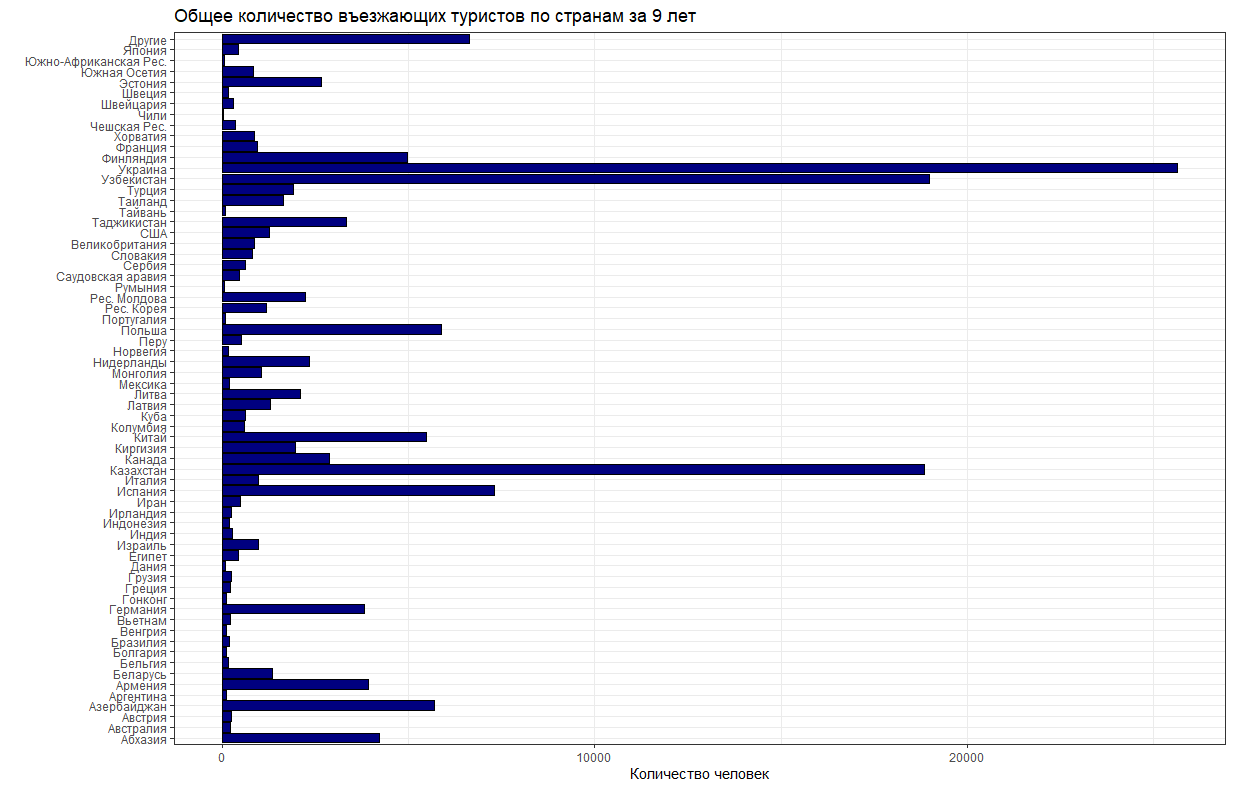


Рисунок 3 - Общее количество въезжающих туристов по странам за 9 лет.

На графике въездных турпотоков также наблюдается активность граждан других стран, хотя в более небольших объемах. Граждане Финляндии, Польши и Испании являются частыми посетителями России и активно совершают выездные поездки в нашу страну. Это можно объяснить близостью к границам России и интересу к местной культуре.

## 3.2 Выездные турпоездки

Теперь перейдем к рассмотрению выездных турпоездок, которые представлены на следующем графике (рисунок 4). Данный график отражает динамику количества выезжающих Российских туристов за последние 9 лет. Анализ выездных турпотоков позволяет нам получить важную информацию о мобильности российских туристов, их предпочтениях в выборе направлений путешествий и изменениях в туристических тенденциях.

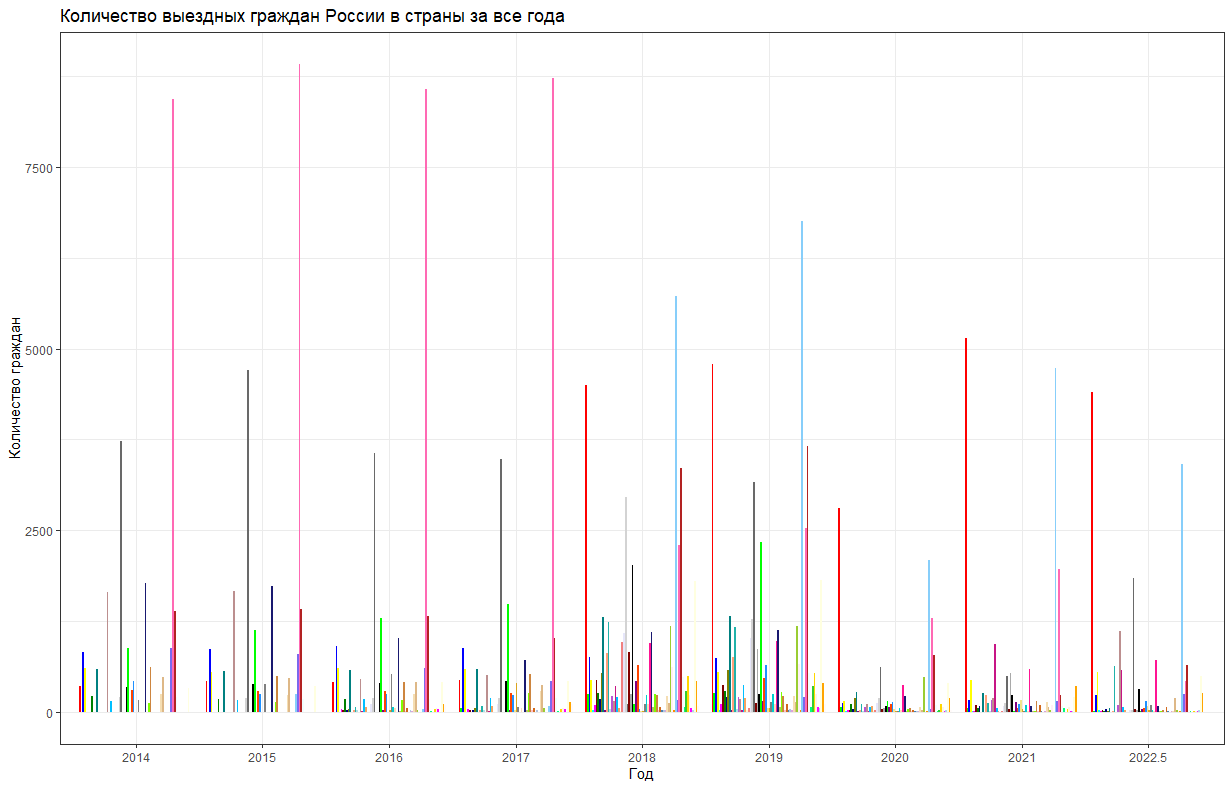


Рисунок 4 – Количество выездных граждан России в страны за все года.

На рисунке 5 изображена легенда для данного графика.

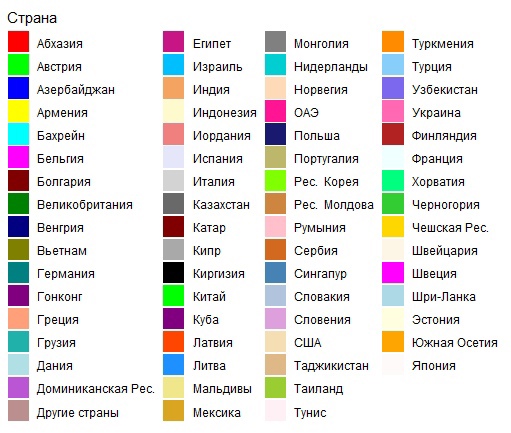


Рисунок 5 – Легенда для графика количества выездных граждан России в страны за все года

Изучив подробнее данный график, можно также выделить страны рекордсмены.

Одним из ведущих направлений, которое привлекает выезжающих российских туристов, является Украина. Близость границ и культурные связи между двумя странами обеспечивают удобство и привлекательность этого направления. Украина предлагает разнообразные туристические возможности, от культурных поездок по историческим городам до отдыха на курортах и наслаждения богатым культурным наследием.

Еще одним интересным трендом на графике является растущая популярность Турции среди российских туристов начиная с 2018 года. Турция привлекает своим прекрасным климатом, привлекательными курортами, богатой историей и гостеприимством своих жителей. Это направление становится все более предпочтительным для российских туристов, предлагая им широкий выбор отдыха на море, познавательных экскурсий и разнообразных развлечений.

Заметной тенденцией на графике является также выездные поездки в Абхазию, особенно с 2018 года. Россияне оценивают близость курортов Абхазии к Сочи, что упрощает переход в эту страну. Там они могут насладиться чистым морем и прекрасным горным воздухом. Это направление привлекает внимание российских туристов своим уникальным природным наследием, а также возможностью погружения в местную культуру и традиции.

Также на рисунке 6 мы можем наблюдать общее количество выездных граждан за последние 9 лет.

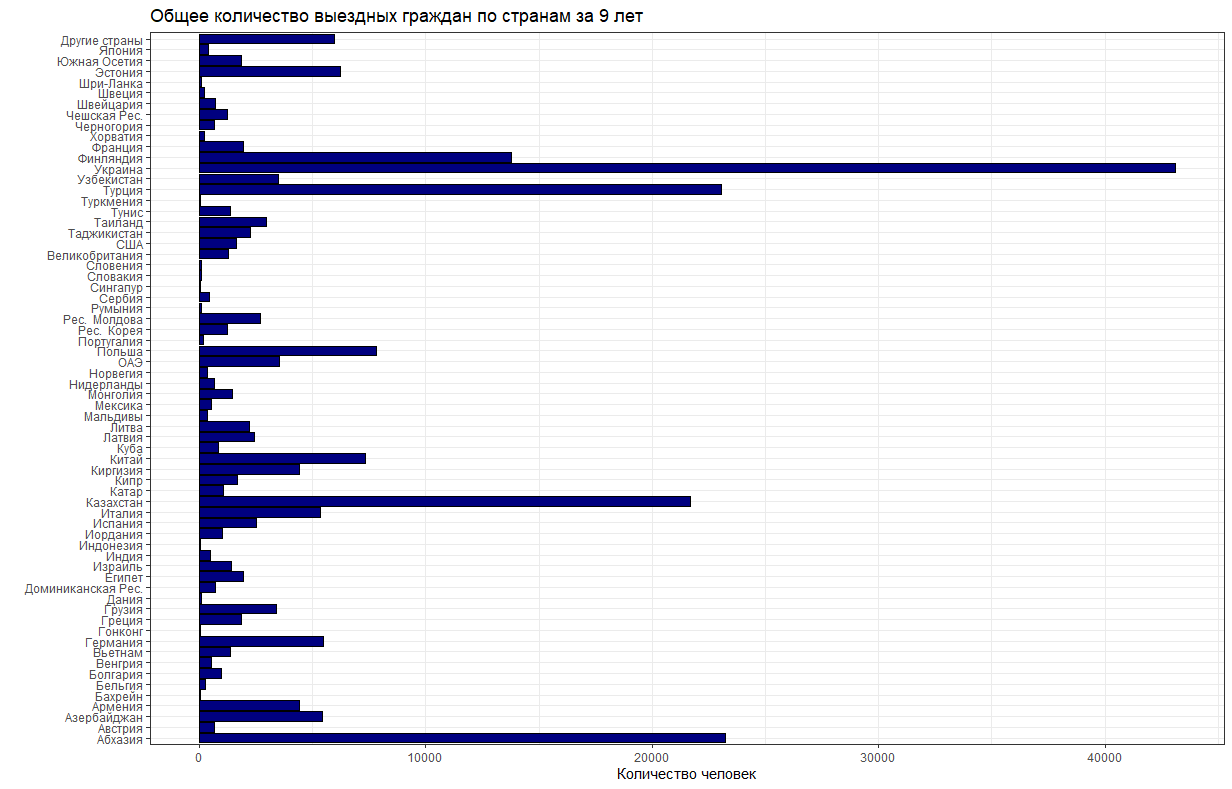


Рисунок 6 - Общее количество выезжающих туристов по странам за 9 лет.

На графике выездных турпотоков также наблюдается популярность выезда туристов в Финляндию. Это обусловлено тем, что Финляндия граничит с Санкт-Петербургом, что упрощает въезд страну. По моим личным наблюдениям и опросам Петербуржцев, некоторые горожане могут каждые выходные на машине выезжать в Финляндию на отдых и прогулки.

## 3.3 Внутренние потоки

Далее представлен график (рисунок 7), отражающий путешествия внутри России в течение 2022 года. На графике указаны различные области страны и количество посетителей, которые посетили каждую из них. График позволяет наглядно увидеть распределение туристического потока внутри России и выделить области, привлекающие большее число посетителей.

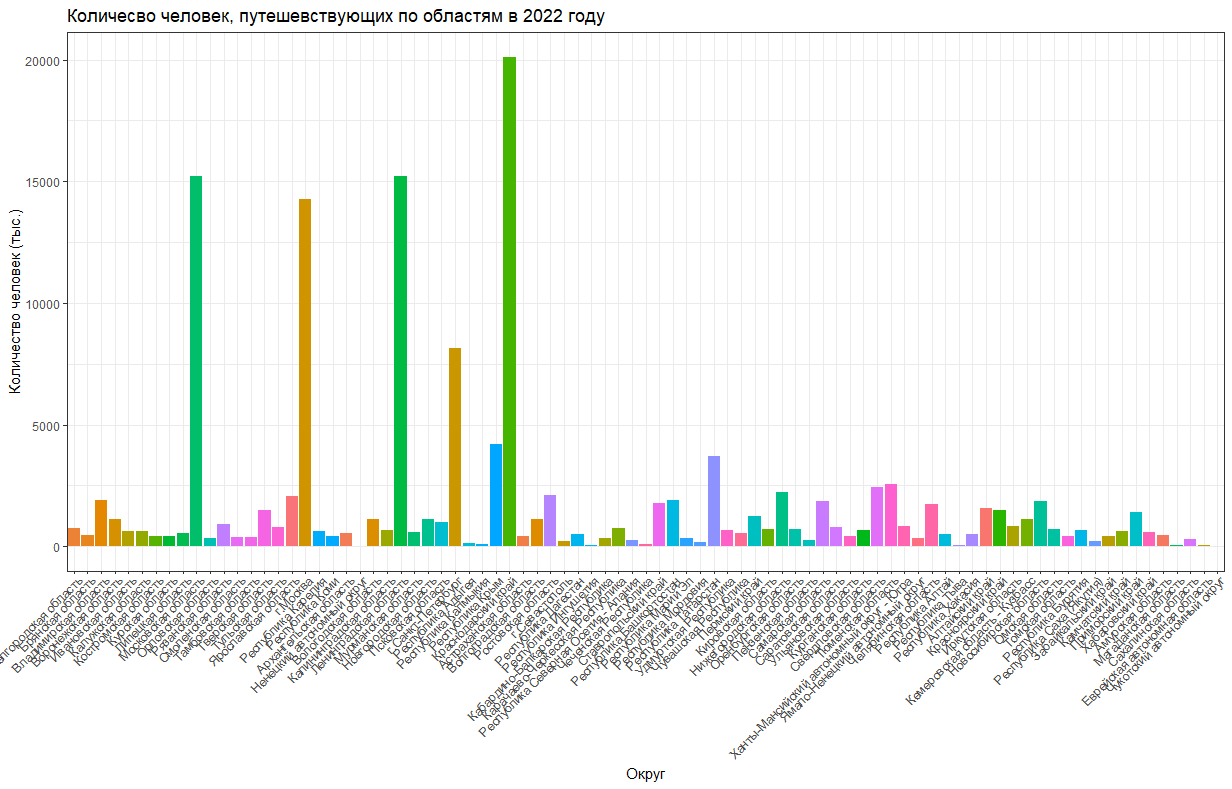


Рисунок 7 - Общее количество въезжающих туристов по странам за 9 лет.

Анализируя представленный график, можно выделить несколько регионов, которые являются особенно популярными среди путешественников внутри России.

Московская область, в частности город Москва, притягивает огромное количество туристов. Богатая история, культурное наследие, величественные архитектурные сооружения и динамичная атмосфера делают Москву незабываемым туристическим направлением.

Ленинградская область, с прекрасным городом Санкт-Петербургом во главе, также привлекает внимание множества путешественников. Исторические достопримечательности, великолепная архитектура, культурные мероприятия и уникальная атмосфера делают Санкт-Петербург одним из самых живописных и привлекательных городов России.

Краснодарский край, с его прекрасными курортами и природными ландшафтами, также является магнитом для туристов. Солнечные пляжи Черного моря, уникальные горные пейзажи Кавказа и гостеприимство местных жителей делают этот регион идеальным местом для отдыха и путешествий.

Таким образом, Московская область, Ленинградская область и Краснодарский край выделяются на графике путешествий внутри России как популярные и востребованные регионы, привлекающие туристов своими уникальными достопримечательностями и возможностями для активного отдыха.

Далее проведем кластеризацию данных о путешествиях внутри России. В ходе анализа алгоритм выделил три кластера.

1. Наиболее посещаемые области.
2. Не очень популярные, но все равно привлекающие внимание путешественников.
3. Редко посещаемые области.

Ниже представлена дендограмма, построенная на основе алгоритма (рисунок 8).

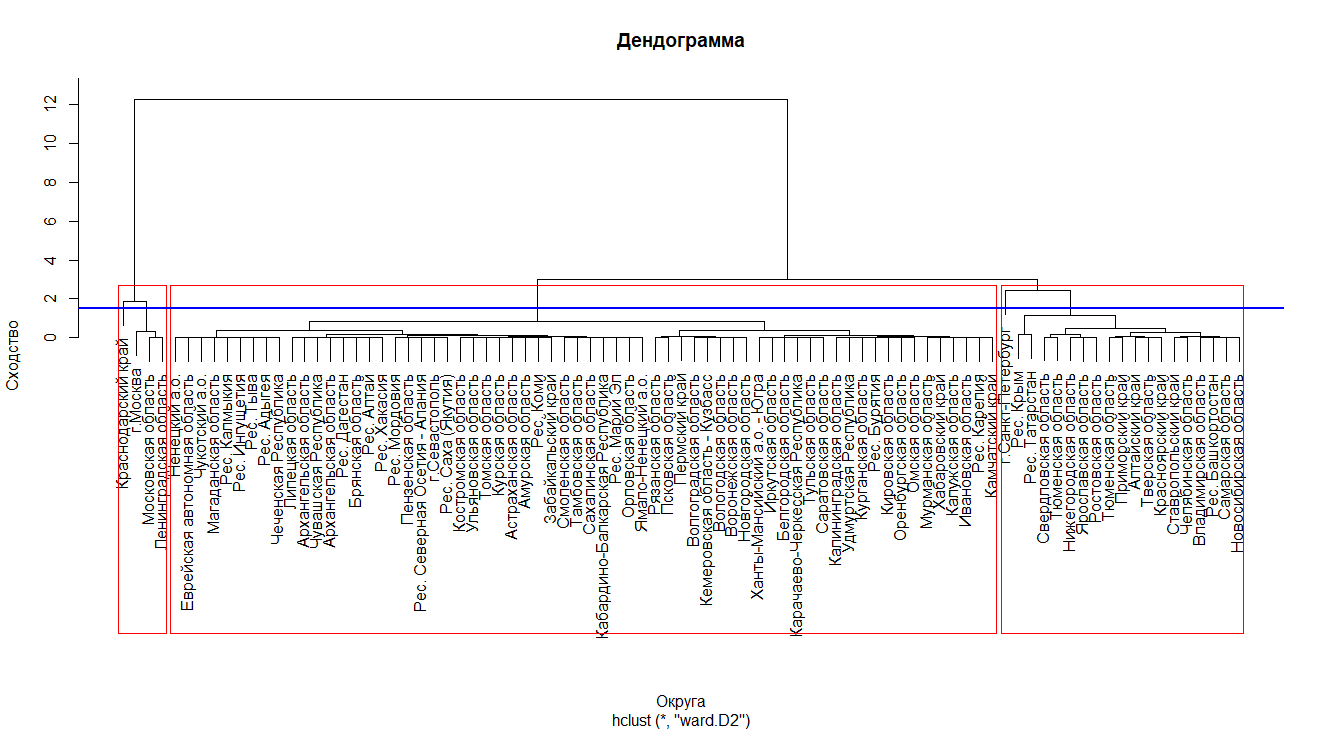


Рисунок 8 – Дендограмма.

Как было сказано ранее, наиболее часто посещаемыми области являются Московская область, Ленинградская область и Краснодарский край. Во вторую группу входят, такие области как Владимирская область, Ставропольский край, Тюменская область и многие другие. Северные области, такие как Алтайский край, Новосибирская область и Красноярский край, выделяются как менее популярные направления для туристов в России. Учитывая анализ туриндустрии (раздел 4), можно предположить, что дополнительное развитие туризма в этих регионах может быть полезным для данных областей. Открытие новых отелей и других объектов туризма в каждом году свидетельствует о возможном потенциале для развития и привлечения большего числа посетителей в непопулярные округа.

На рисунке 10 изображен график, показывающий количество туристов по общим округам.

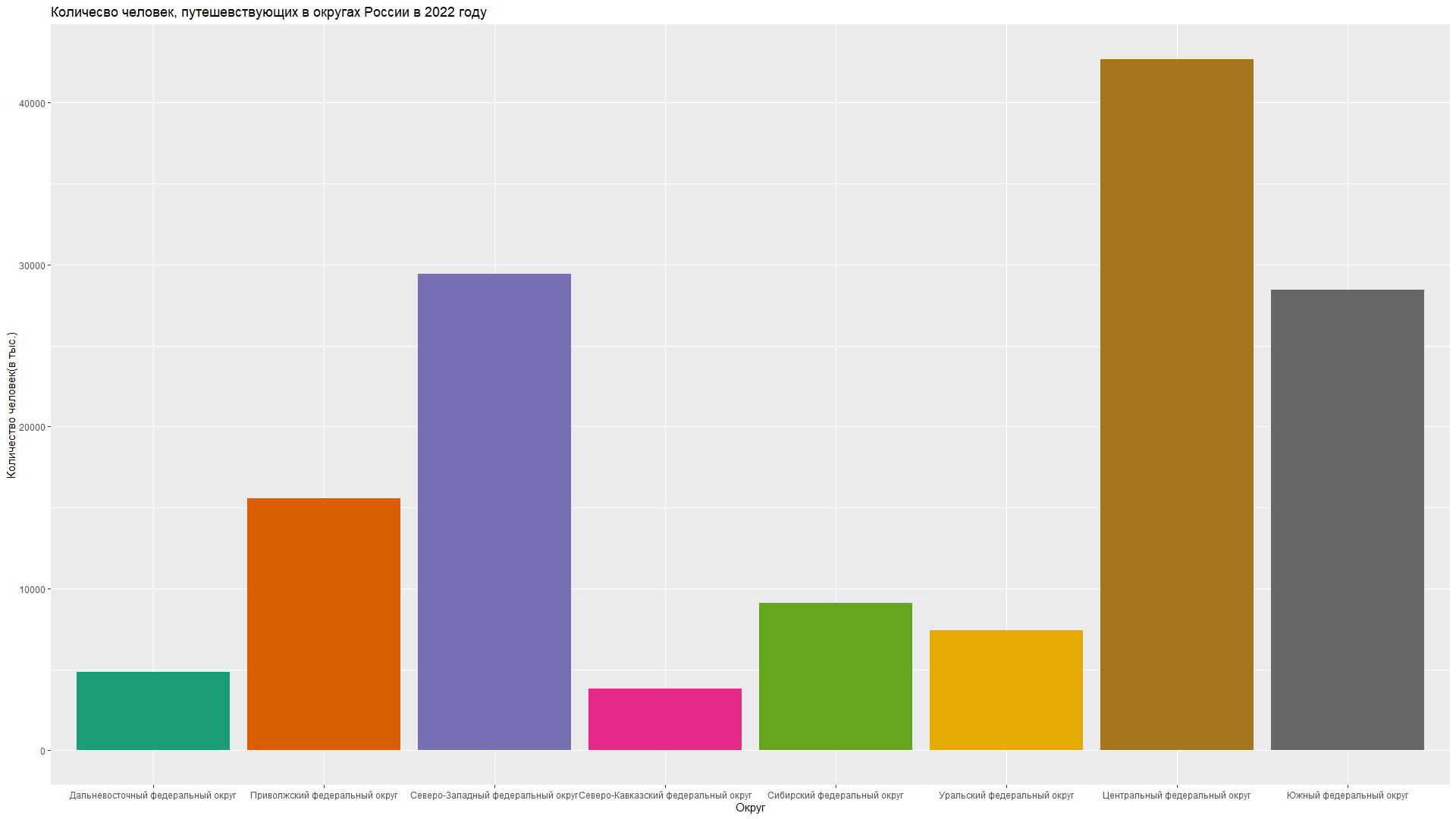


Рисунок 10 – Количество человек, путешествующих по России в 2022 году.

Мы также видим, что Южный и Центральный федеральные округа занимают лидирующие позиции.

## 3.4 По сезонам

Для более детального анализа путешествий по округам, можно разделить год на четыре сезона, с учетом климатических условий. Четвертый и первый кварталы, включающие октябрь, ноябрь, декабрь, январь, февраль, март, можно отнести к холодным сезонам, в то время второй и третий - апрель, май, июнь, июль, август, сентябрь - можно отнести к теплым сезонам.

Анализируя данные о путешествиях по округам в разные сезоны, можно выделить определенные тенденции. В холодные сезоны, люди склонны предпочитать поездки в округа с зимними активностями, такими как горные курорты или регионы с возможностью занятий лыжным спортом. Отдых в горах или посещение горнолыжных курортов являются популярными вариантами для путешествий в эти сезоны.

С другой стороны, в теплые сезоны, люди стремятся отправиться в округа с пляжами, курортами и прекрасными природными ландшафтами. Путешествия по округам с морскими побережьями, озерами или живописными парками и заповедниками востребованы в теплое время года, когда можно наслаждаться открытым воздухом и природными достопримечательностями.

Таким образом, анализ путешествий по округам в разные сезоны позволяет выявить предпочтения и тенденции путешественников. На представленном графике (рисунок 11) отображается количество людей, путешествующих по разным округам в холодный и теплый периоды года.

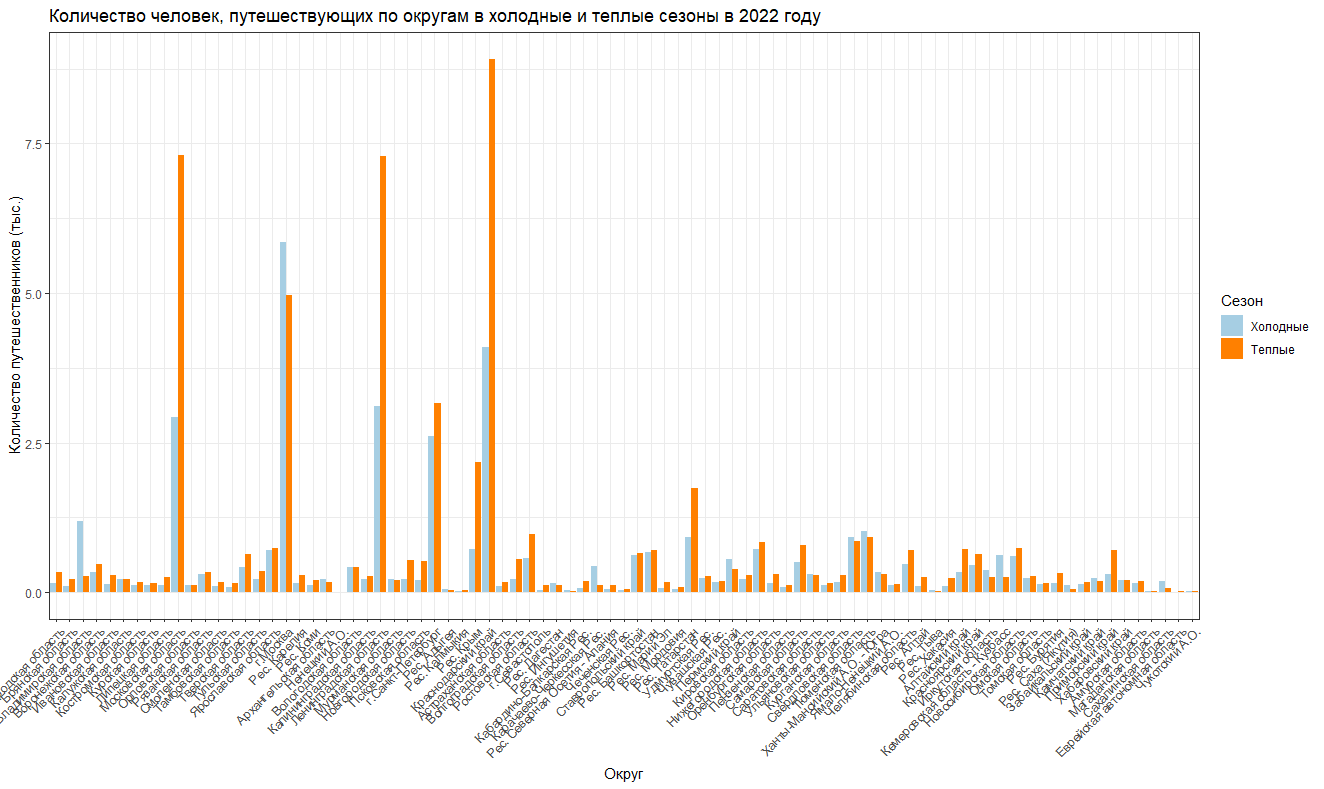


Рисунок 11 – Количество человек, путешествующих по округам в холодные и теплые сезоны.

В теплые сезоны выделяются несколько округов, которые занимают лидирующие позиции в популярности среди путешественников.

Краснодарский край привлекает своим теплым климатом, прекрасными морскими и горными пейзажами, создавая идеальные условия для отдыха и активного времяпрепровождения.

Ленинградская область, особенно город Санкт-Петербург, радует приятной погодой в это время года, предоставляя возможность насладиться прогулками по историческим улицам и наследием культуры.

Московская область, включая Москву, также пользуется популярностью благодаря своим достопримечательностям и разнообразным развлечениям для посетителей.

Эти регионы предлагают разнообразные возможности для приятного и запоминающегося отдыха в теплое время года.

В холодное время года Московская и Ленинградская области продолжают привлекать туристов, уступая лишь небольшую долю популярности в сравнении с теплыми сезонами. Краснодарский край также остается привлекательным благодаря своим горнолыжным курортам. В то же время, другие регионы не привлекают такого большого числа посетителей как в теплое, так и в холодное время года, в частности северные регионы не являются желанными точками остановки для путешественников, в чем мы можем еще раз удостоверится, увидев график на рисунке 12.

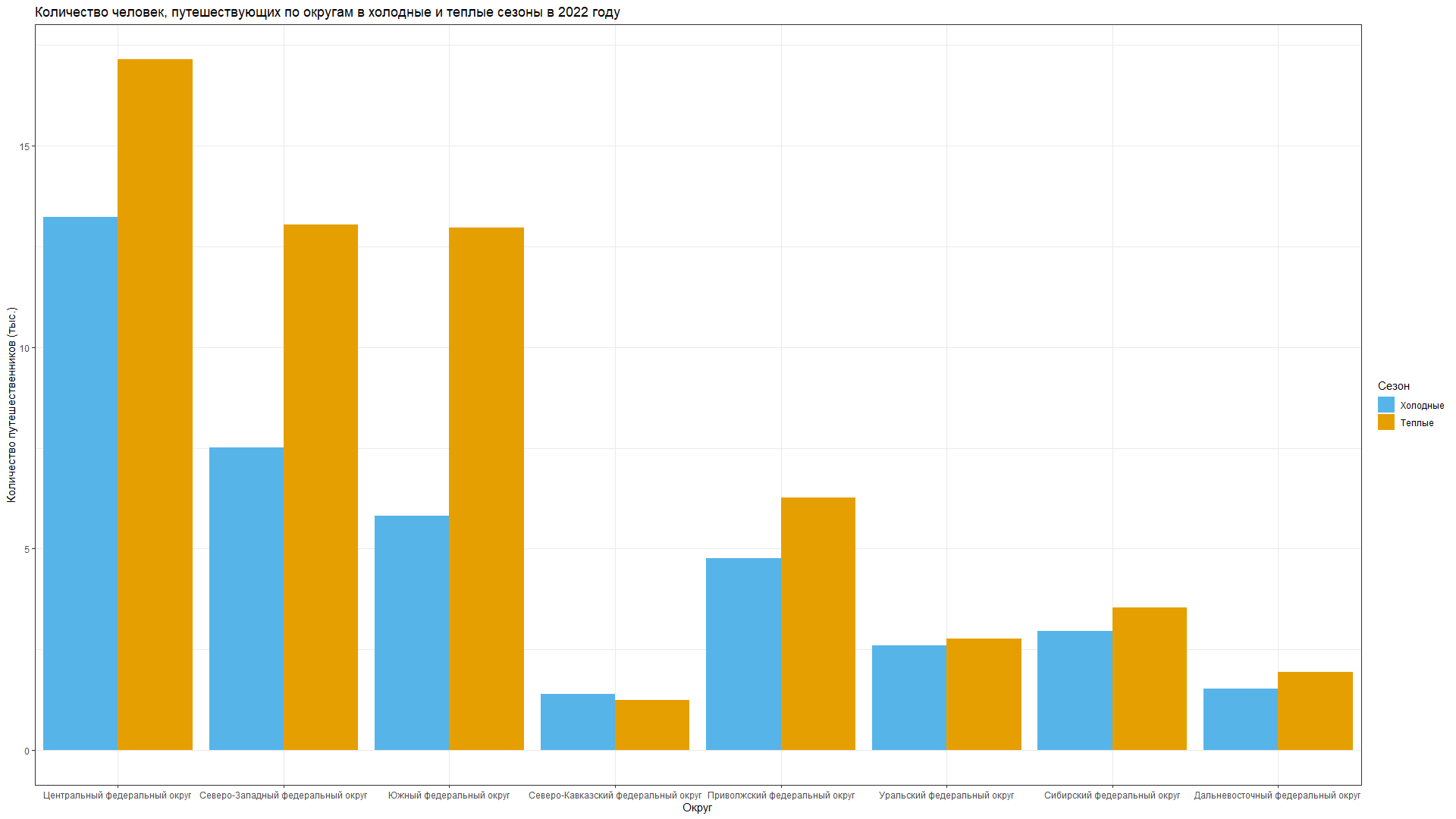


Рисунок 12 – Количество человек, путешествующих по округам в холодные и теплые сезоны.

# Туриндустрия

Туриндустрия включает в себя широкий спектр организаций, включая отели, гостиницы, туристические агентства, туристические операторы и другие предприятия, связанные с туризмом. Анализ данных о прибыльности, выручке, убытках и других показателях этих организаций позволяет получить представление о состоянии и развитии туристической отрасли.

Такой анализ предоставляет ценную информацию для принятия решений в сфере туризма и позволяет выявить факторы, влияющие на успех или неудачу туристических организаций.

## 4.1 Организации (прибыльные, убыточные)

Теперь перейдем к более подробному рассмотрению данных и графиков. На рисунке 13 представлены все организации, действующие на территории РФ с 2014 года. На рисунке 14 - прибыльные организации, а на рисунке 15 – убыточные.

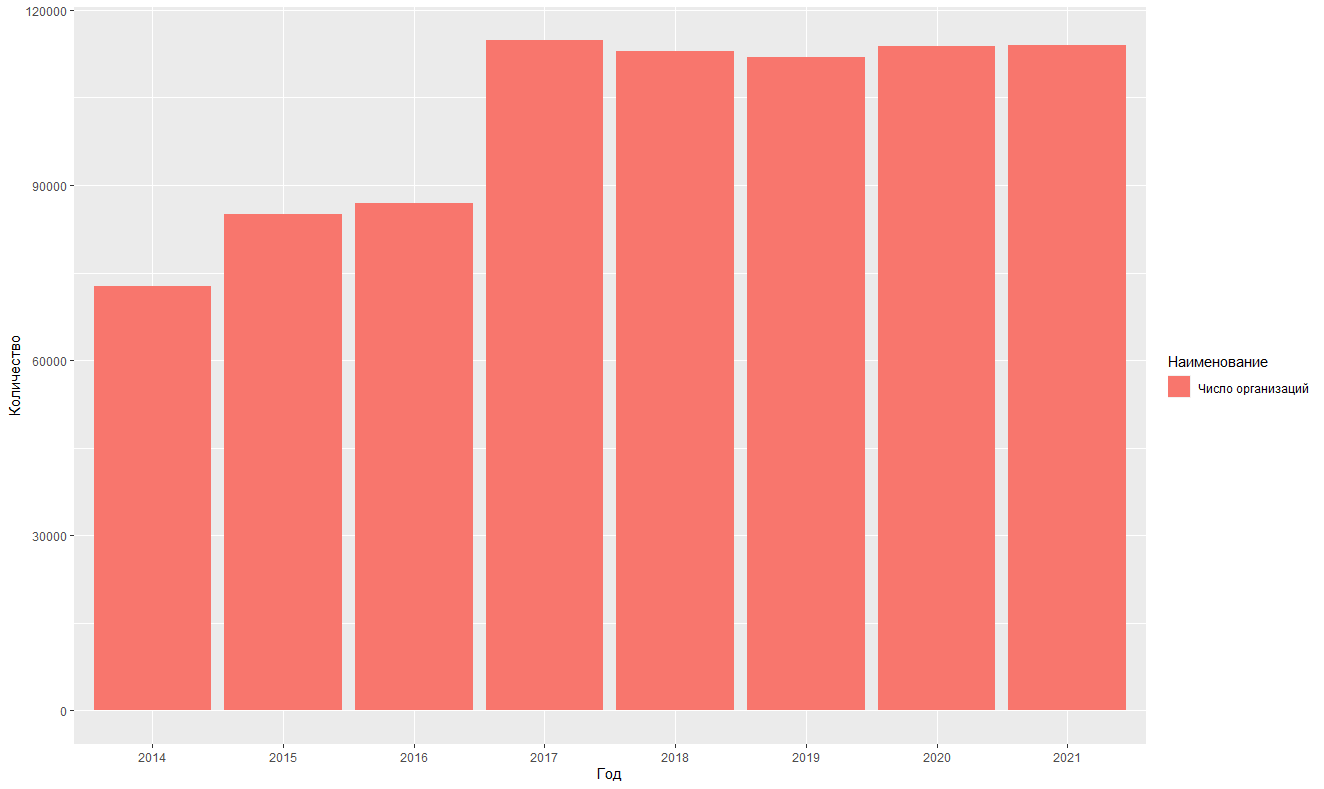


Рисунок 13 – Количество организаций на территории РФ за последние 9 лет.

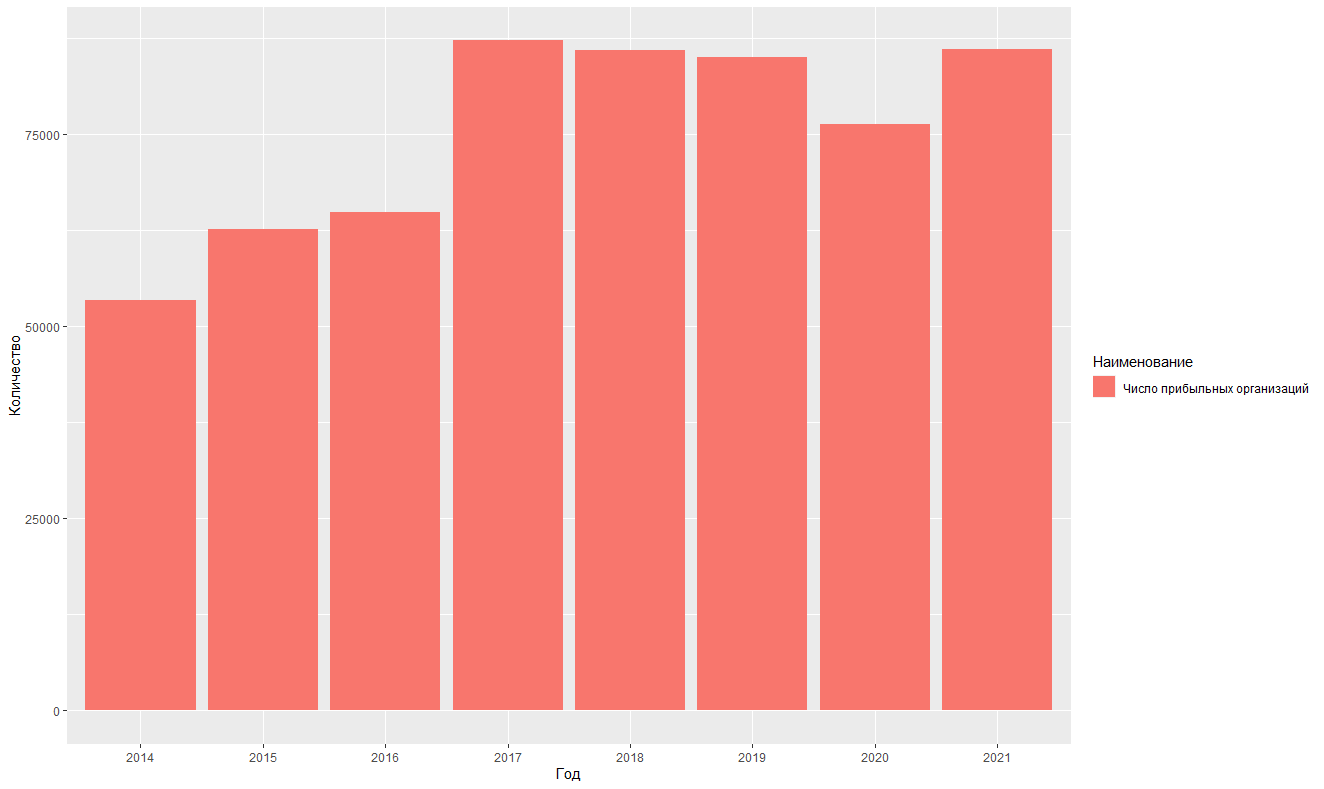


Рисунок 14 – Количество прибыльных организаций на территории РФ за последние 9 лет.

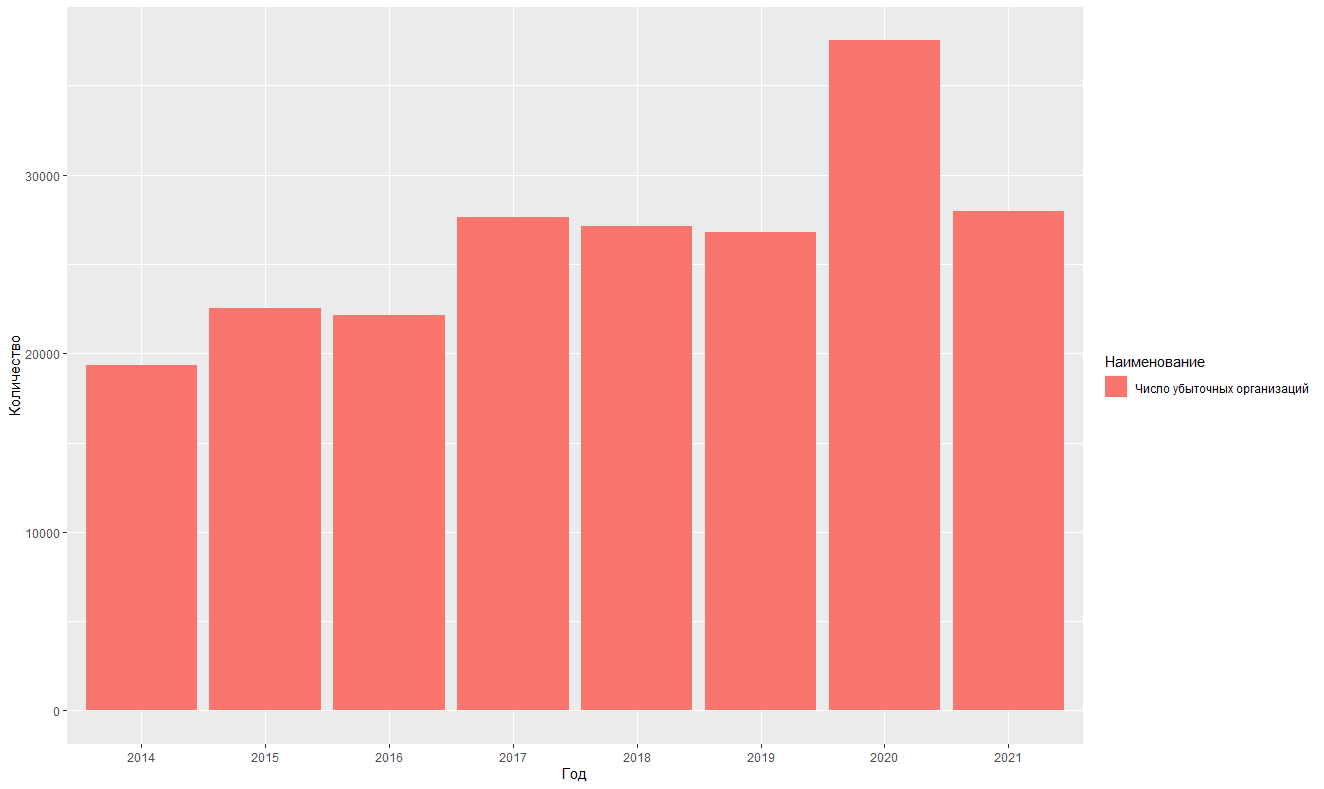


Рисунок 15 – Количество убыточных организаций на территории РФ за последние 9 лет.

При анализе выше описанных графиков становится очевидным влияние пандемии Covid-19 на туристическую отрасль в стране. В 2020 году отмечается резкое снижение числа туристов и путешественников, что является результатом ограничений и сокращения туристической деятельности. Этот упадок в турпотоках и путешествиях внутри страны привел к серьезным вызовам для туристической индустрии, включая снижение выручки и прибыли организаций, что мы рассмотрим далее.

## 4.2 Выручка, прибыль, убыток и НФР туристических организаций

На рисунках 16, 17, 18 и 19 представлены выручка, прибыль, убыток и НФР тур. организаций соответственно.

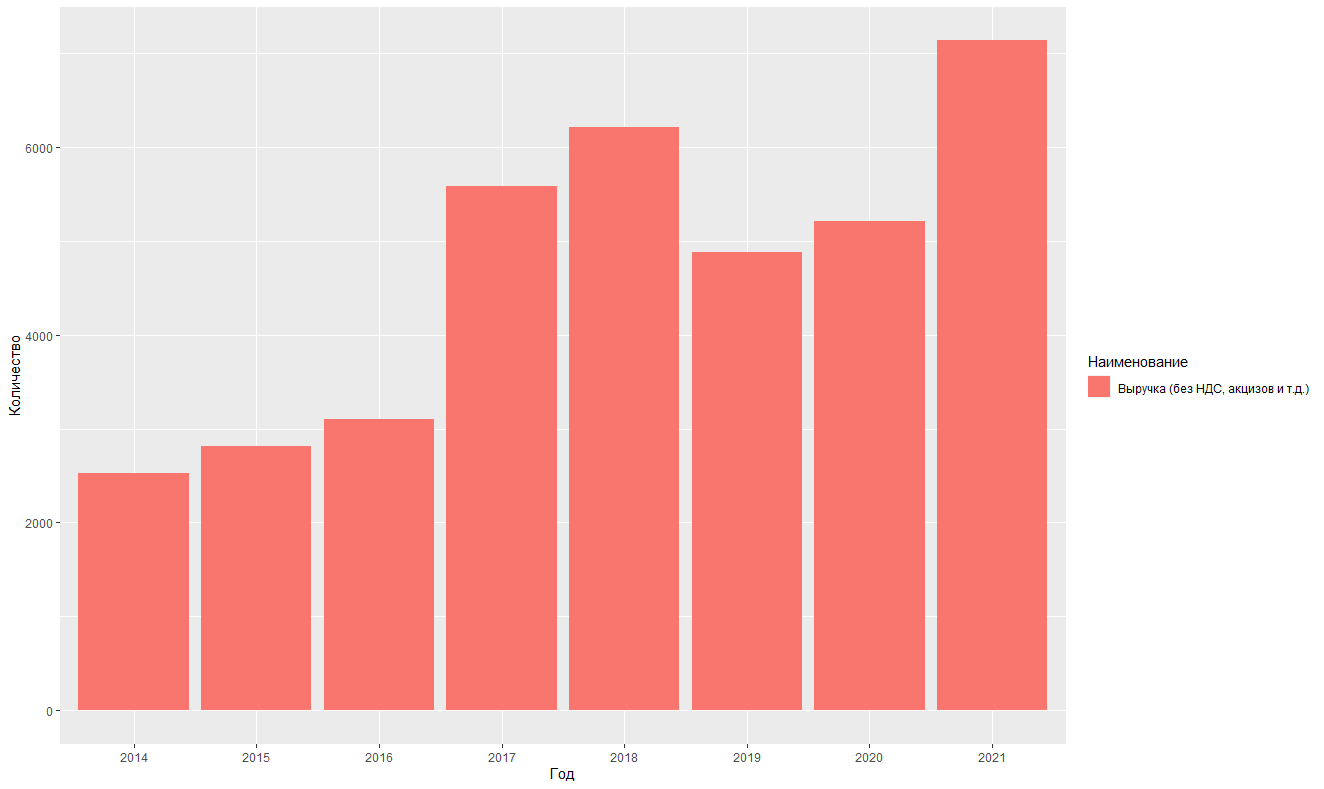


Рисунок 16 – Общая выручка тур. организаций за последние 9 лет.

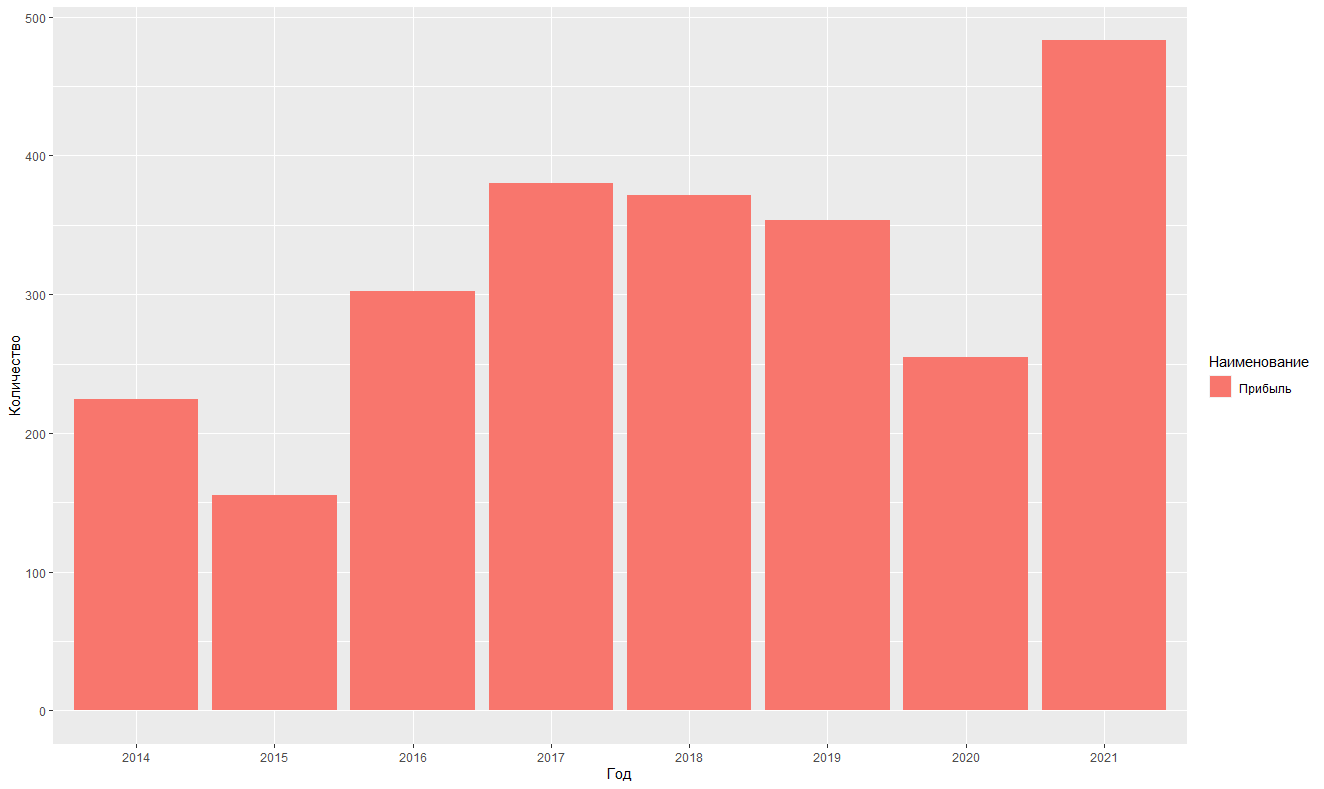


Рисунок 17 – Общая прибыль тур. организаций за последние 9 лет.

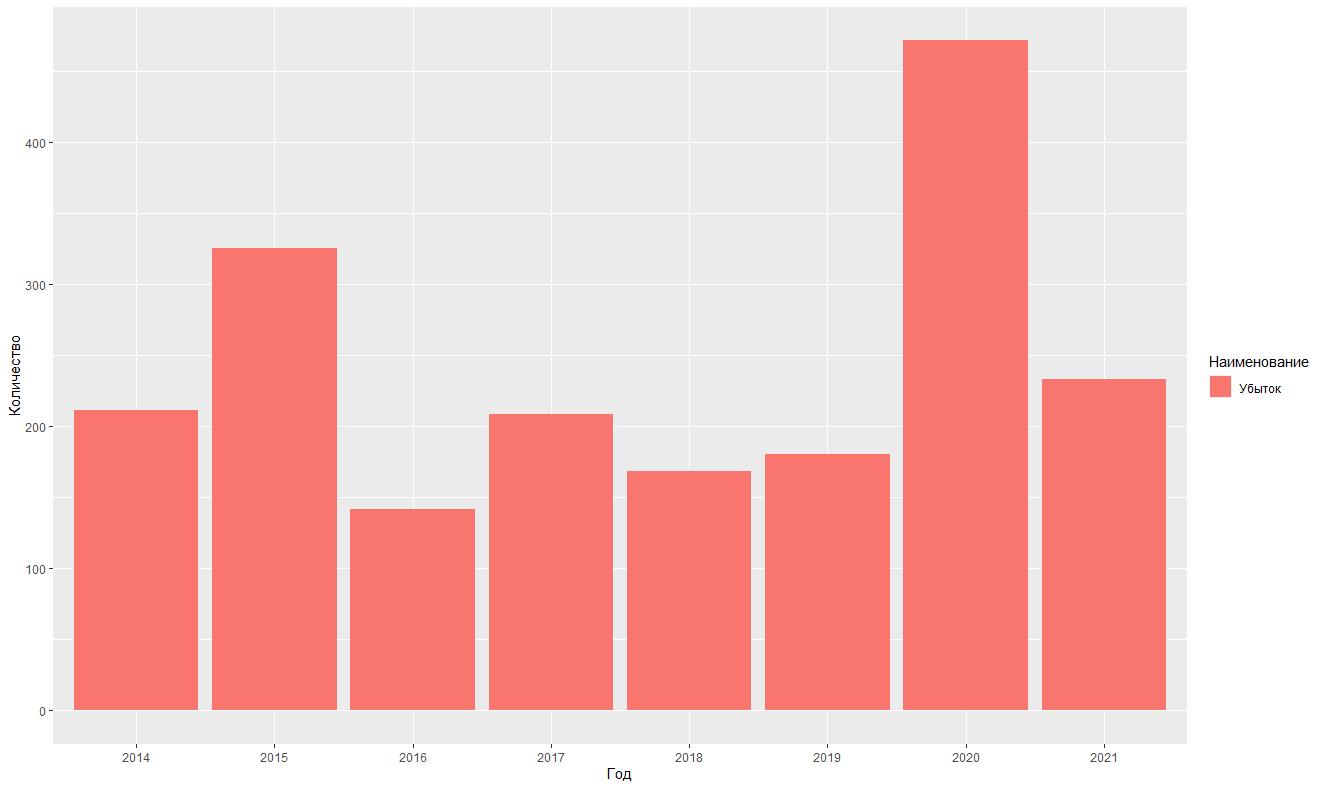


Рисунок 18 – Общий убыток тур. организаций за последние 9 лет.

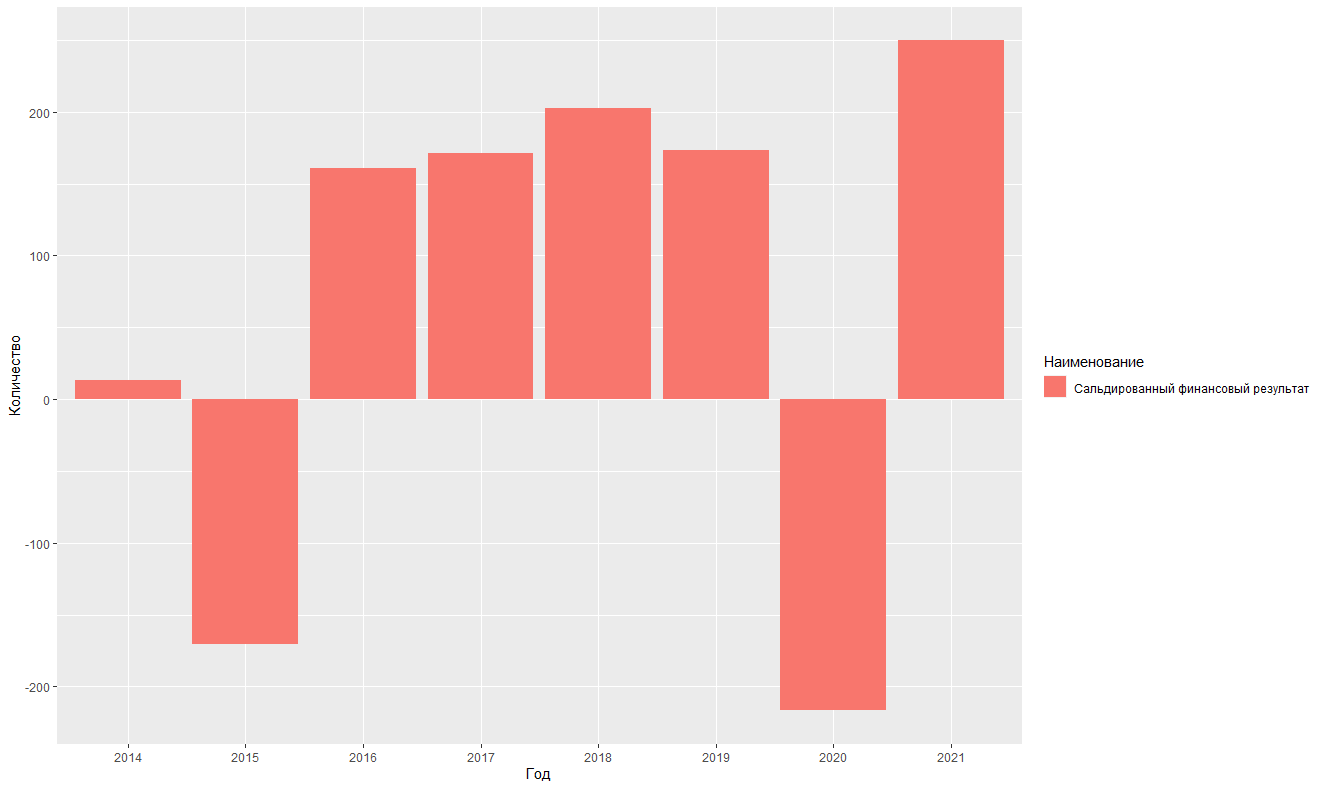


Рисунок 19 – НФР тур. организаций за последние 9 лет.

Анализируя представленные графики, мы можем наблюдать серьезные убытки и снижение выручки в 2022 году, вызванные пандемией. Однако, также стоит отметить интересную тенденцию в 2015 году, когда отмечаются колоссальные убытки со стороны туристических организаций. Вероятно, это связано с серией терактов, произошедших в разных частях мира, что вызвало опасения и неуверенность у путешественников. Эти события оказали значительное влияние на решения о поездках и предпочтения в отдыхе.

Таким образом, как пандемия, так и геополитические события имели заметный эффект на доходность туристических организаций.

## 4.3 Ввод в действие объектов туризма

На рисунке 20 мы можем увидеть ввод в действие объектов туризма.

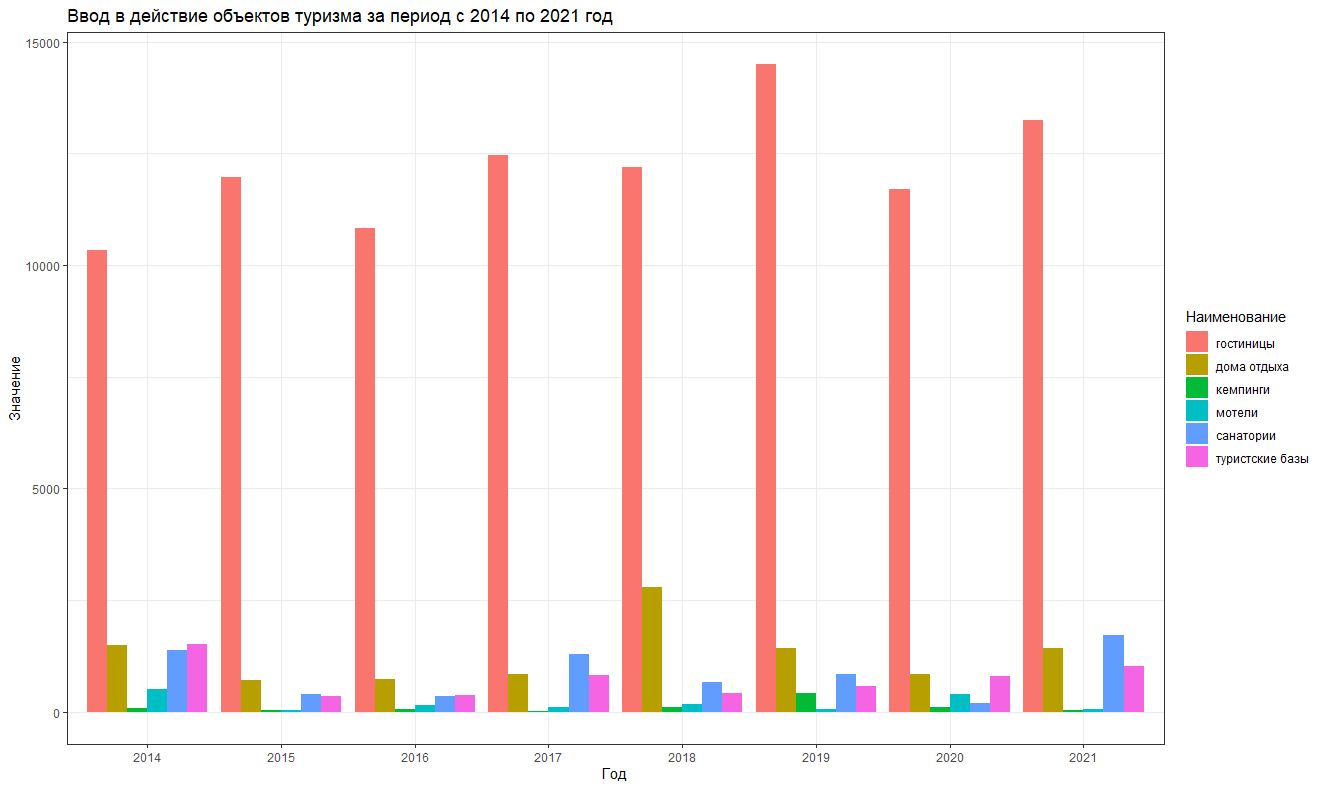


Рисунок 20 - Ввод в действие объектов туризма за период с 2014 по 2021 год.

За промежуток девяти лет наблюдается стабильное открытие гостиниц и отелей в значительном количестве, указывающее на широкую доступность размещения для туристов. Однако, кемпинги и мотели остаются наименее популярными вариантами размещения. В нашей стране данные виды объектов туризма не привлекают массового числа отдыхающих. Санатории и туристические базы, в свою очередь, удерживают свою позицию на постоянном уровне, предлагая определенную альтернативу для туристов, их спрос сохраняется стабильным.

# ЗАКЛЮЧЕНИЕ

В результате проведенного исследования были изучены турпотоки в России, включая въездные и выездные турпоездки, а также путешествия внутри страны.

По результатам анализа графиков были выявлены ведущие роли различных стран в числе приезжих и выезжающих туристов, а также популярные туристические регионы, такие как Московская область, Ленинградская область и Краснодарский край. Также были выделены тенденции упадка в туристической индустрии в связи с пандемией Covid-19 и другими факторами, такими как волны терактов.

В процессе исследования были разобраны различные виды туристических объектов, от гостиниц и отелей до кемпингов и санаториев.

Также был углубленно изучен язык для анализа данных R, его методы построения графиков и алгоритмы кластеризации данных.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. **Конфликтов, О.А.** Анализ данных в туризме и гостиничном бизнесе. — Москва: Издательство "Альпина Паблишер", 2018. — 256 с. — ISBN 978-5-9614-6934-2.
   * 1. **Иванов, С.Н., Романова, О.В.** Статистика туризма. — Москва: Издательство "Финансы и статистика", 2017. — 384 с. — ISBN 978-5-279-04054-6.
     2. Упрощенная формальная модель интеллектуального агента [Электронный ресурс].– URL: <https://studopedia.ru/2_27516_uproshchennaya-formalnaya-model-intellektualnogo-agenta.html>
     3. **Макаров, А.Л., Соколов, В.В.** Теория и практика статистического анализа данных в туризме. — Москва: Издательство "Юрайт", 2019. — 296 с. — ISBN 978-5-534-01918-1.
     4. Статистика туризма в России [Электронный ресурс]. — Росстат (Федеральная служба государственной статистики). — URL: https://rosstat.gov.ru/

# ПРИЛОЖЕНИЕ A

**Листинг программы**

library(readxl)

library(tidyr)

library(ggplot2)

library(dplyr)

library(RColorBrewer)

library(factoextra)

library(dendextend)

library(randomcoloR)

setwd("C:/Users/nagal/OneDrive/GitHub/Coursework\_2023/Данные")

#ВЪЕЗДНЫЕ

inbound\_tours<-read\_excel("Въездные турпоездки.xlsx")

colors <- c("#FF0000", "#00FF00", "#0000FF", "#FFFF00", "#00FFFF", "#FF00FF", "#800000",

"#008000", "#000080", "#808000", "#008080", "#800080", "#FFA07A", "#20B2AA",

"#B0E0E6", "#BA55D3", "#BC8F8F", "#C71585", "#00BFFF", "#F4A460", "#FFFACD",

"#F08080", "#E6E6FA", "#D3D3D3", "#696969", "#800000", "#A9A9A9", "#000000",

"#00FF00", "#800080", "#FF4500", "#1E90FF", "#F0E68C", "#DAA520", "#808080",

"#00CED1", "#FFDAB9", "#FF1493", "#191970", "#BDB76B", "#7FFF00", "#CD853F",

"#FFC0CB", "#D2691E", "#4682B4", "#B0C4DE", "#DDA0DD", "#F5DEB3", "#DEB887",

"#9ACD32", "#FFF0F5", "#FF8C00", "#87CEFA", "#7B68EE", "#FF69B4", "#B22222",

"#F0FFFF", "#00FF7F", "#32CD32", "#FFD700", "#FDF5E6", "#FF00FF", "#ADD8E6",

"#FFFFE0", "#FFA500", "#FFFAFA")

inbound\_tours\_long <- inbound\_tours %>%

gather(key = "Год", value = "number\_of\_travelers", -Страна) %>%

group\_by(Страна, Год) %>%

summarise(Общее\_количество = sum(number\_of\_travelers))

ggplot(inbound\_tours\_long, aes(x = Год, y = Общее\_количество, fill = Страна)) +

geom\_bar(stat = "identity", position = "dodge") +

scale\_y\_continuous(name = "Количество въезжающих туристов", limits = c(0, max(inbound\_tours\_long$Общее\_количество))) +

labs(title = "Количество въезжающих туристов из каждой страны за все года",

x = "Год", y = "Количество въезжающих туристов") +

scale\_fill\_manual(values = colors) +

guides(fill=FALSE)+

theme\_bw()

#Общее количество приезжих по страннам за 9 лет

all\_inbound\_tours<-data.frame(Страна=inbound\_tours$Страна, ОбщееКоличество=rowSums(inbound\_tours[,2:10]))

ggplot(all\_inbound\_tours, aes(x = Страна, y = ОбщееКоличество)) +

geom\_bar(stat = "identity", fill = "navyblue", color = "black") +

labs(x = "", y = "Количество человек",

title = "Общее количество въезжающих туристов по странам за 9 лет") +

scale\_x\_discrete(labels = all\_inbound\_tours$Страна,

limits = all\_inbound\_tours$Страна) +

coord\_flip() +

theme\_bw()

#ВЫЕЗДНЫЕ

field\_tours<-read\_excel("Выездные турпоездки.xlsx")

field\_tours\_long <- field\_tours %>%

gather(key = "Год", value = "number\_of\_travelers", -Страна) %>%

group\_by(Страна, Год) %>%

summarise(Общее\_количество = sum(number\_of\_travelers))

ggplot(field\_tours\_long, aes(x = Год, y = Общее\_количество, fill = Страна)) +

geom\_bar(stat = "identity", position = "dodge") +

scale\_y\_continuous(name = "Количество граждан", limits = c(0, max(field\_tours\_long$Общее\_количество))) +

labs(title = "Количество выездных граждан России в страны за все года",

x = "Год", y = "Количество человек") +

scale\_fill\_manual(values = colors) +

guides(fill=FALSE)+

theme\_bw()

#Общее количество уезжих по страннам за 9 лет

all\_field\_tours <- data.frame(Страна = field\_tours$Страна,

ОбщееКоличество = rowSums(field\_tours[, 2:10]))

ggplot(all\_field\_tours, aes(x = Страна, y = ОбщееКоличество)) +

geom\_bar(stat = "identity", fill = "navyblue", color = "black") +

labs(x = "", y = "Количество человек",

title = "Общее количество выездных граждан по странам за 9 лет") +

scale\_x\_discrete(labels = all\_field\_tours$Страна,

limits = all\_field\_tours$Страна) +

coord\_flip() +

theme\_bw()

#Путешествие в России

In\_Russian<-read\_excel("Внутри России.xlsx")

set.seed(123)

colors <- randomColor(count = 87, luminosity = "bright")

ggplot(In\_Russian, aes(x = Округ, y = ОбщееКоличество / 1000, fill = Округ)) +

geom\_bar(stat = "identity", position = "dodge") +

scale\_y\_continuous(name = "Количество человек (тыс.)", limits = c(0, max(In\_Russian$ОбщееКоличество/1000))) +

labs(title = "Количесво человек, путешествующих по областям в 2022 году") +

theme\_bw() +

theme(legend.position = "bottom",axis.text.x = element\_blank()) +

theme(axis.text.x = element\_text(angle = 45, vjust = 1, hjust = 1, size = 9)) +

scale\_x\_discrete(limits = unique(In\_Russian$Округ), expand = c(0, 0.5))+

guides(fill = FALSE)

#кластеризация

In\_Russian\_clast <- select(In\_Russian, -Округ)

In\_Russian\_clast\_scaled <- scale(In\_Russian\_clast)

hc <- hclust(dist(In\_Russian\_clast\_scaled), method = "ward.D2")

plot(hc, labels = In\_Russian$Округ, main="Дендограмма",ylab="Сходство",xlab="Округа")

rect.hclust(hc, 3, border="red")

abline(h = 1.5, col = "blue", lwd='2')

dend <- as.dendrogram(hc)

dend <- color\_branches(dend, 3)

plot(dend)

groups <- cutree(hc, 3)

In\_Russian[groups==1, 1]

In\_Russian[groups==2, 1]

In\_Russian[groups==3, 1]

#Путешествия по россии за 2022 общее по округам

In\_Russian\_all <- read\_excel("Внутри России общее.xlsx")

colors <- brewer.pal(8, "Dark2")

ggplot(In\_Russian\_all, aes(x = Округ, y = Количество/1000, fill = Округ)) +

geom\_bar(stat = "identity") +

scale\_fill\_manual(values = colors) +

labs(title = "Количесво человек, путешествующих в округах России в 2022 году",

x = "Округ", y = "Количество человек(в тыс.)") +

guides(fill = FALSE)

#по кварталам, по сезонам

season<-read\_excel("По сезонам.xlsx")

season\_long <- season %>%

gather(key = "Квартал", value = "Количество", -Округ) %>%

mutate(Квартал = factor(Квартал, levels = c("1 квартал", "2 квартал", "3 квартал", "4 квартал","Холодные","Теплые")))

season\_long <- season\_long %>%

mutate(Количество = Количество / 1000)

# для кварталов холодных и теплых

season\_cold\_warm <- season\_long %>%

filter(Квартал %in% c("Холодные", "Теплые"))

season\_cold\_warm <- season\_cold\_warm %>% rename(Сезон = Квартал)

ggplot(season\_cold\_warm, aes(x = Округ, y = Количество/1000, fill = Сезон )) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Количество человек, путешествующих по округам в холодные и теплые сезоны в 2022 году",

x = "Округ",

y = "Количество путешественников (тыс.)") +

scale\_fill\_manual(values = c("#A6CEE3", "#FF8000")) +

theme\_bw() +

theme(axis.text.x = element\_text(angle = 45, vjust = 1, hjust = 1, size = 9)) +

scale\_x\_discrete(limits = unique(season\_cold\_warm$Округ), expand = c(0, 0.5))

#по кварталам, по сезонам(общее)

season\_all<-read\_excel("По сезонам общее.xlsx")

season\_long\_all <- season\_all %>%

gather(key = "Квартал", value = "Количество", -Округ) %>%

mutate(Квартал = factor(Квартал, levels = c("1 квартал", "2 квартал", "3 квартал", "4 квартал","Холодные","Теплые")))

season\_long\_all <- season\_long\_all %>%

mutate(Количество = Количество / 1000)

# для кварталов холодных и теплых

season\_23\_all <- season\_long\_all %>%

filter(Квартал %in% c("Холодные", "Теплые"))

season\_23\_all <- season\_23\_all %>% rename(Сезон = Квартал)

ggplot(season\_23\_all, aes(x = Округ, y = Количество/1000, fill = Сезон)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Количество человек, путешествующих по округам в холодные и теплые сезоны в 2022 году",

x = "Округ",

y = "Количество путешественников (тыс.)") +

scale\_fill\_manual(values = c("#56B4E9", "#E69F00")) +

theme\_bw() +

scale\_x\_discrete(limits = unique(season\_23\_all$Округ), expand = c(0, 0.5))

#Туриндустрия

travel\_industry<-read\_excel("Туриндустрия.xlsx")

#Все организации

number\_organ<- rbind(travel\_industry[1:1, ])

number\_organ\_long<-gather(number\_organ,key="Год",value="Количество", -Наименование)

ggplot(number\_organ\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#Прибыльные

number\_organ\_plus<- rbind(travel\_industry[2:2, ])

number\_organ\_plus\_long<-gather(number\_organ\_plus,key="Год",value="Количество", -Наименование)

ggplot(number\_organ\_plus\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#Убыточные

number\_organ\_minus<- rbind(travel\_industry[3:3, ])

number\_organ\_minus\_long<-gather(number\_organ\_minus,key="Год",value="Количество", -Наименование)

ggplot(number\_organ\_minus\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#Выручка

revenue<- rbind(travel\_industry[4:4, ])

revenue\_long<-gather(revenue,key="Год",value="Количество", -Наименование)

ggplot(revenue\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#Прибыль

revenue\_plus<- rbind(travel\_industry[5:5, ])

revenue\_plus\_long<-gather(revenue\_plus,key="Год",value="Количество", -Наименование)

ggplot(revenue\_plus\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#Убыток

revenue\_minus<- rbind(travel\_industry[6:6, ])

revenue\_minus\_long<-gather(revenue\_minus,key="Год",value="Количество", -Наименование)

ggplot(revenue\_minus\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#NFR(прибыль минус убыток)

NFR<- rbind(travel\_industry[7:7, ])

NFR\_long<-gather(NFR,key="Год",value="Количество", -Наименование)

ggplot(NFR\_long, aes(x = Год, y = Количество, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge")

#Ввод в действие объектов туризма с отелями

tourism\_facilities<-travel\_industry[9:14,]

tourism\_facilities\_long <- gather(tourism\_facilities, key = "Год", value = "Значение", -Наименование)

ggplot(tourism\_facilities\_long, aes(x = Год, y = Значение, fill = Наименование)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Ввод в действие объектов туризма за период с 2014 по 2021 год",

x = "Год",

y = "Значение") +

theme\_bw()