**Тестовое задание ЦДП**

**Полученные в ходе анализа результаты**

Общие доходы

Минимальное значение общих доходов граждан за рассматриваемых период (2018-2023) составило 235021.1, в то время как максимальное достигло 4793233417.8. Среднее значение доходов составляет 36291448.3. Если цифры скорректированы с учетом инфляции, то это указывает на общее улучшение финансового положения граждан в рассматриваемый период.

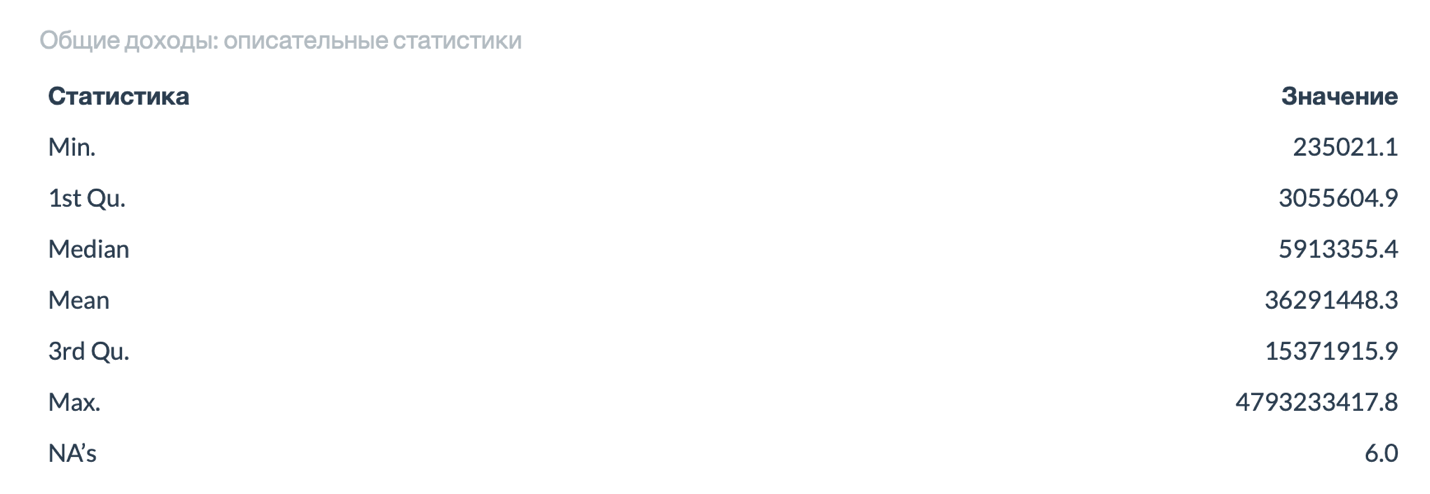
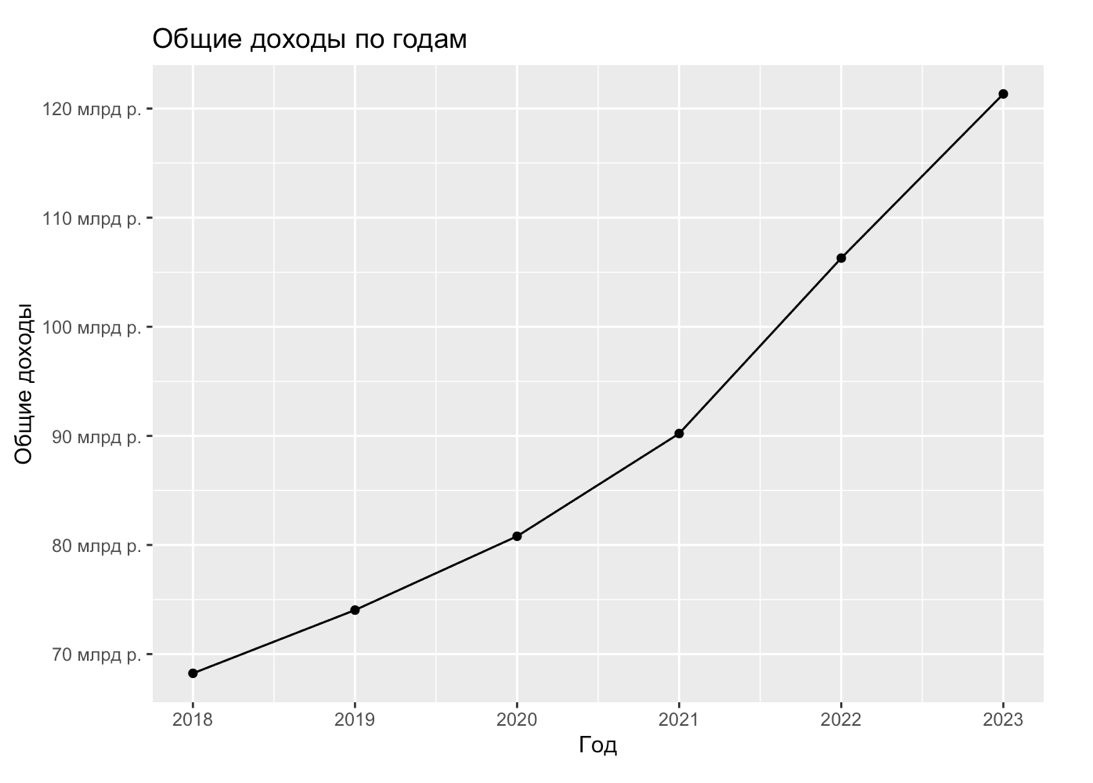
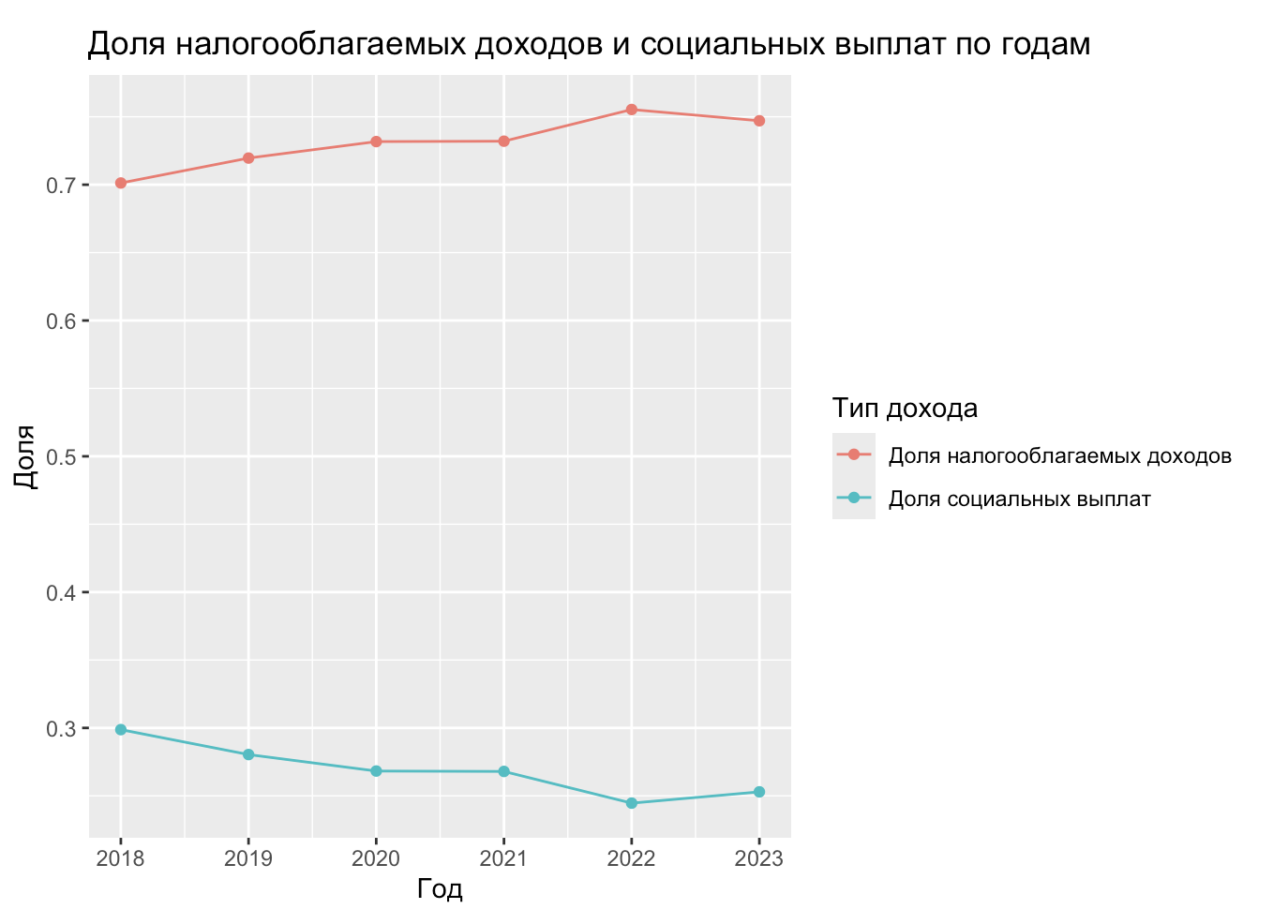


График ниже показывает значительное увеличение общих доходов населения с 2018 по 2023 год.



Доля социальных выплат

Графики ниже демонстрируют, что доля социальных выплат в структуре доходов населения с 2018 по 2023 годами увеличилась. Это может свидетельствовать о росте зависимости населения от государственных пособий и социальных программ (данное предположение нужно проверять дополнительно).



*Примечание: в будущих расчетах также надо будет учесть темп инфляции, который здесь не учитывался ввиду ограниченности времени.*

**Описание функционала содержимого в Rmd файле кода:**

1. Код загружает данные о социальных выплатах и налогооблагаемых доходах из файла Excel. Он извлекает данные с листов, относящихся к годам с 2017 по 2022, и создает новый датафрейм, в котором годы нумеруются с 2018 по 2023.

3. Объединенные данные очищаются от пропусков, и производится переименование столбцов для удобства дальнейшего анализа.

4. Код группирует данные по муниципальным образованиям и подсчитывает количество уникальных кодов ОКТМО для каждого муниципального района. Далее фильтруются только те районы, у которых более одного уникального кода, что позволяет выявить сложные структуры муниципалитетов.

5. Найденный первый код ОКТМО для каждого муниципального района заменяет все последующие коды, что упрощает дальнейший анализ.

6. Код создает новые столбцы в датафрейме, суммируя налогооблагаемые доходы и социальные выплаты, чтобы получить общие доходы для каждого муниципалитета.

7. Создаются графики, которые показывают динамику общих доходов по годам и доли различных типов доходов (налогооблагаемых и социальных выплат) в структуре доходов населения.

8. Код анализирует, как изменились доли социальных выплат и налогооблагаемых доходов в структуре доходов населения в 2023 году по сравнению с 2018 годом, и предлагает различные способы визуализации этих изменений.

9. Создается тепловая карта, показывающая распределение долей социальных выплат по регионам с 2018 по 2023 годы, что помогает визуализировать изменения в зависимости от географического положения. (PS: Тепловая карта нуждается в доработке и представлена только в ознакомительных целях.)

**Краткий дизайн исследования**

Цель исследования заключается в оценке того, как инвестиционные проекты способствуют увеличению налогооблагаемых доходов населения, основываясь на предположении, что такие проекты могут привести к росту занятости, увеличению доходов от предприятий и улучшению инфраструктуры.

В качестве основного метода исследования предлагается использовать метод разностных различий (Difference-in-Differences, DiD). Этот подход позволит сравнить изменения в доходах в муниципалитетах, где реализуются инвестиционные проекты, с теми, где их нет, принимая во внимание изменения, произошедшие до и после внедрения проектов. Также планируется применять регрессионный анализ с фиксированными эффектами, чтобы учесть специфические характеристики муниципалитетов, такие как численность населения и уровень безработицы, которые могут влиять на уровень доходов.

Механизм имплементации инвестиционных проектов окажет значительное влияние на дизайн исследования. Например, отбор муниципалитетов для участия в программе может быть случайным или целенаправленным. В случае случайного отбора можно использовать методы разностных различий, минимизируя влияние сторонних факторов. Если же отбор целенаправленный, то потребуется контролировать различные параметры, которые могут искажать результаты, например, уровень доходов.

Важно также учитывать характеристики муниципалитетов. Если при отборе учитываются демографические или экономические параметры, то их следует включить в модель для снижения искажений. Масштаб инвестиций также играет ключевую роль: при наличии значительных различий в объеме инвестиций их следует включать в модель как предиктор, чтобы оценить, как уровень финансирования связан с ростом доходов населения.

Количество муниципалитетов, участвующих в программе, также важно. Большое количество муниципалитетов повышает статистическую надежность результатов, тогда как малый охват требует применения методов подбора контрольных групп для уменьшения искажений.

Кроме того, необходимо учитывать направление реализуемых проектов, так как различные виды проектов могут по-разному влиять на доходы. Важно учитывать тип и длительность проектов для разделения краткосрочных и долгосрочных эффектов.