

Кластеризация регионов РФ

Kirill Sidorov

2025-02-05

Содержание

Выводы для бизнеса	2
Краткий литературный обзор	2
Описание данных	3
Методология	3
Только данные ВРП 2022 года	4
Данные по рейтингу социально-экономического положения	8
Данные отгрузки, занятости и др.	12

Выводы для бизнеса

Краткий литературный обзор

Вопросом кластеризации регионов страны исследователи задаются постоянно с разной степенью интенсивности. Как правило, кластеризация служит инструментом поиска укрупненных федеральных групп, по которым оценивается воздействие того или иного инструмента политики (policy).

Например, наиболее известными работами в этом жанре являются исследования о степени вклада федеральных и региональных факторов процентного канала трансмиссионного механизма денежно-кредитной политики (ДКП) за период ковид и послесанкционного 2022 года (Скуратова, Зверева, 2024). В указанной работе были выделены факторы региональной неоднородности для формирования укрупненных групп: неравенство по доходам, размер предприятий региона, банковского сектора.

В другой работе (Зверева et al., 2024) исследовалось влияние внутрирегионального неравенства на работу каналов ДКП, где для выделения основных групп регионов использовался метод k-median в зависимости от уровня неравенства в регионе.

В более ранних работах (Айвазян et al. 2016) предлагалось выделить группы регионов РФ по производственному потенциалу: зависимость ВРП от стоимости основных фондов и численности занятых. На основе данных был построен интегральный индикатор специализации регионов (сельское хозяйство, добыча полезных ископаемых и проч.), на основе которого были выделены группы регионов РФ.

Для полноты картины представим работу исследователей других стран. Так, например, в работе сотрудников Федерального Банка Нью-Йорка (Kim-Sherman, Seltzer, 2024) поднимается проблема коррелированности ошибок панельных регрессий в экономике климата (climate economics). На основе подхода к выделению кластеров на основе климатических катастроф авторы показывают, что ущерб стране, испытывающий климатическую катастрофу, получает на 0.45% больше ущерба при увеличении ущерба в соседней стране кластера на 1%.

Описание данных

Для первоначальной работы с данными были выбраны показатели экономического положения регионов: ВРП, оборот торговли, стоимость отгрузок и др. Полный список приведен ниже:

Показатель	Пояснение
GRP_2022	ВРП в постоянных ценах 2016 по субъектам Российской Федерации за 2022 год
employed_2023	Численность и состав рабочей силы в возрасте 15 - 72 лет по регионам РФ
retail_sales_2023	Объем розничной торговли по субъектам РФ
industrial_2023	Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по субъектам Российской Федерации. Годовые данные Промышленное производство, млн руб
mining_2023	Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по субъектам Российской Федерации. Годовые данные Добыча полезных ископаемых, млн руб
manufacturing_2023	Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по субъектам Российской Федерации. Годовые данные Обрабатывающие производства, млн руб
energy_2023	Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по субъектам Российской Федерации. Годовые данные Энергоснабжение, млн руб
water_2023	Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по субъектам Российской Федерации. Годовые данные Водоснабжение, млн руб
credit_2023	Кредиторская задолженность организаций
revenue_2023	Прибыль/убыток до налогообложения по субъектам РФ
overdue_debit_2023	Просроченная дебиторская задолженность организаций
profit_orgs_share_2023	Удельный вес прибыльных организаций
ria_ratings_2023	Рейтинг соц-экон положения регионов РФ, составленный агентством РИА НОВОСТИ

Источник: Росстат, ЕМИСС, РИА НОВОСТИ.

Методология

Для отправной точки кластеризации выбран подход [K-means](#). Данный алгоритм делит данные на отдельные группы одинаковой дисперсии, минимизируя критерий *инерции* или сумму квадратов отклонений кластера, что представимо следующей формулой:

$$\sum_{i=0}^n \min_{\mu_j \in C} (||x_i - \mu_j||^2)$$

Где μ_j - среднее кластера C , x_i - элемент наблюдаемого множества X , $\|*\|$ - выбранная норма (Евклидова, L1 и др.).

Метод имеет свои ограничения, иногда более подходящей моделью может служить c-means (оценивает вероятность вхождения в кластер), а также улучшенный метод k-means++. В научных работах также используется подход k-median, что в условиях смещенных распределений метрик может нивелировать этот недостаток.

Для нашего случая воспользуемся базовым k-means. Для реализации каждый наблюдаемый фактор необходимо стандартизовать, чтобы модель не отдавала предпочтение одному количественно наибольшему признаку.

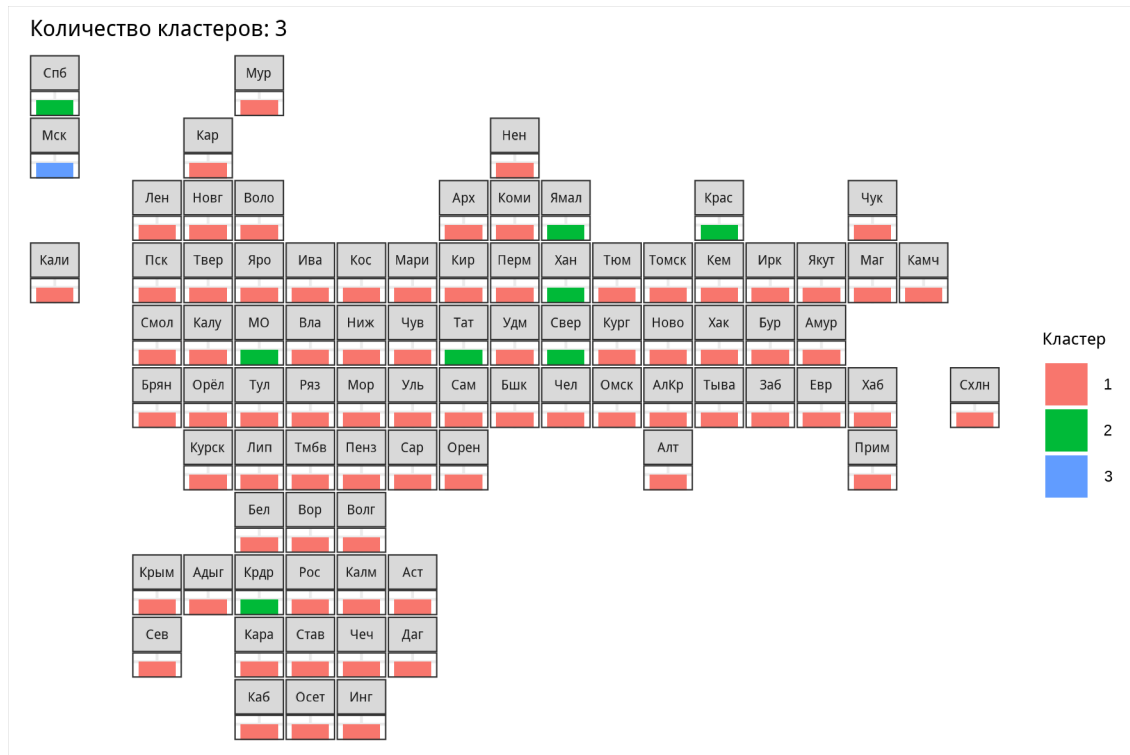
Стандартизация выглядит следующим образом:

$$z = \frac{x_i - \text{mean}(X)}{\text{sd}(X)}$$

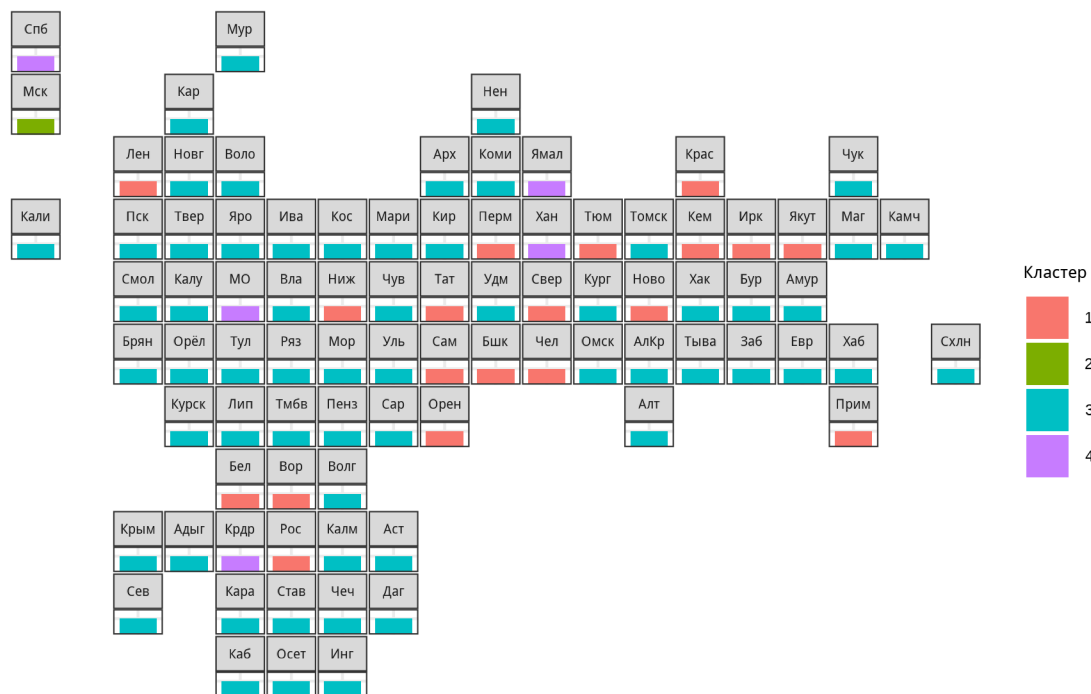
Где $\text{mean}(X)$ - среднее по всей выборке, sd - стандартное отклонение по всей выборке.

Только данные ВРП 2022 года

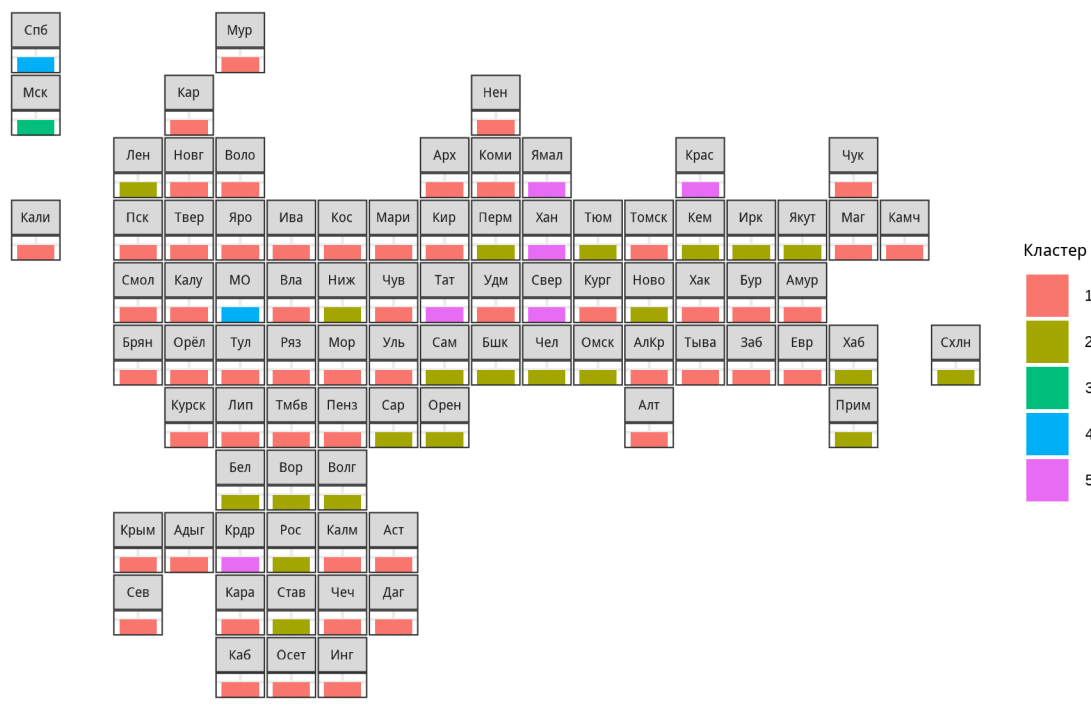
Для начала используем модель только по ВРП, результаты приведены ниже:



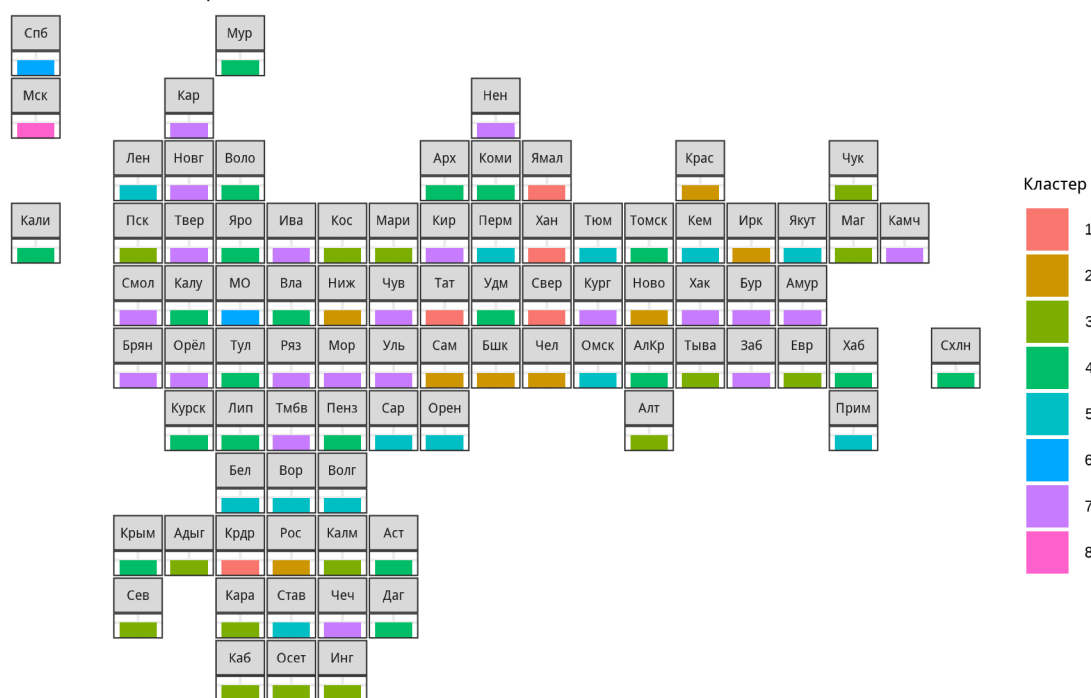
Количество кластеров: 4



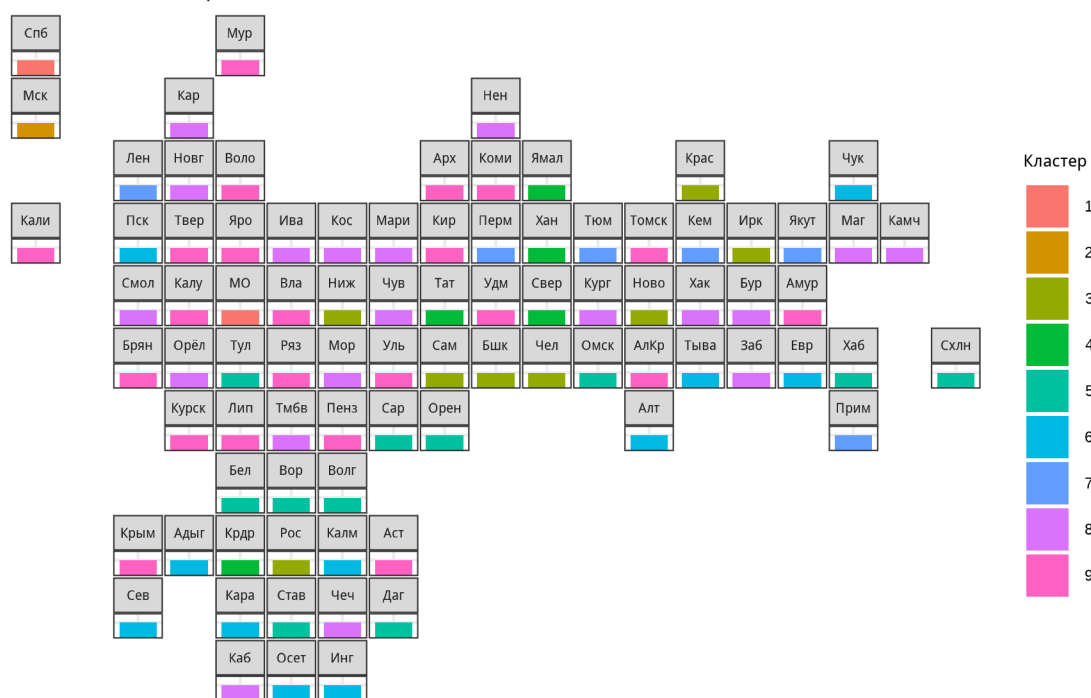
Количество кластеров: 5



Количество кластеров: 8



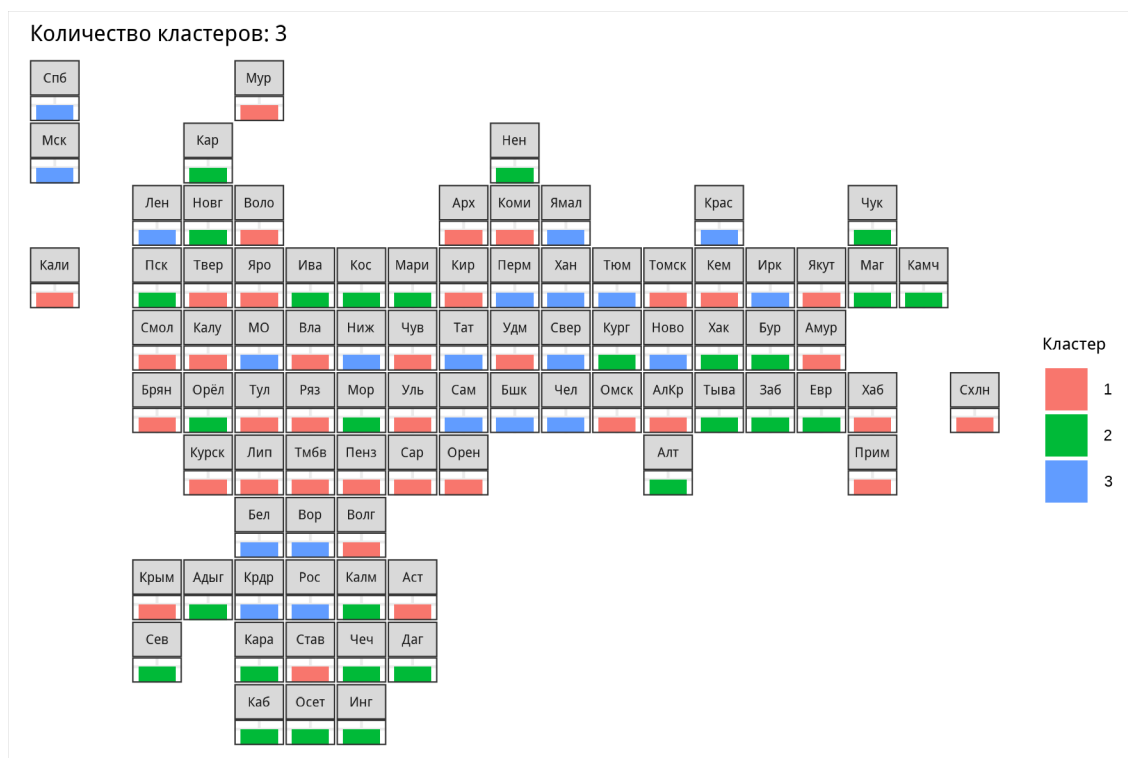
Количество кластеров: 9



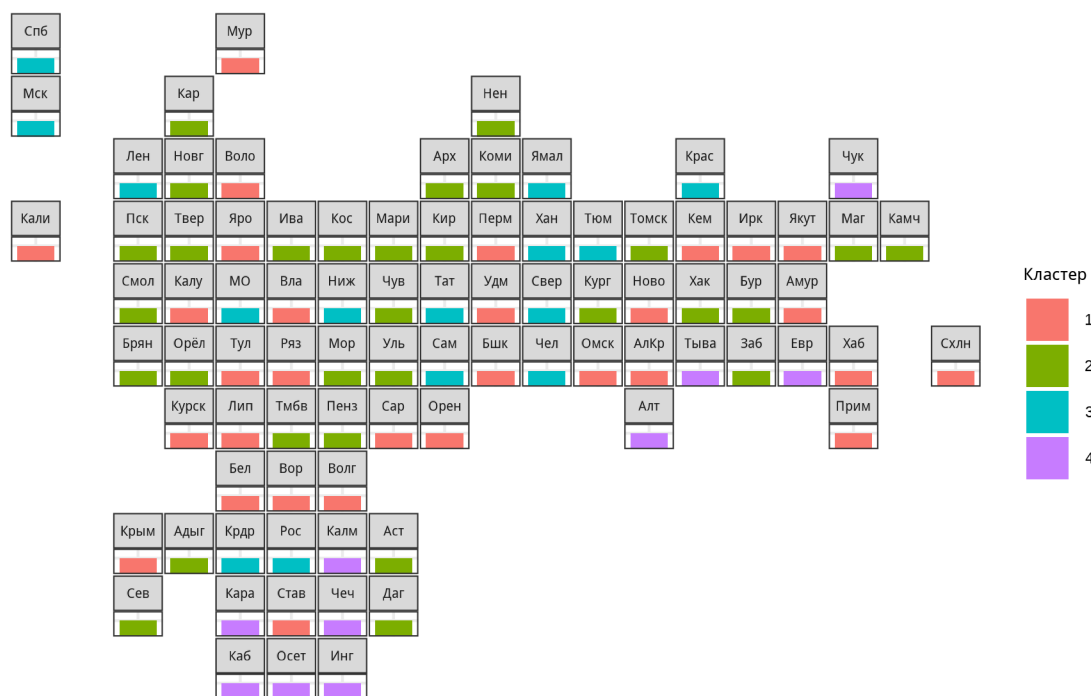
Предварительно выглядит так, что наиболее адекватным разделением кажется количество кластеров в районе - 5-7. Наиболее отличительная черта - город Москва всегда выделен в отдельный кластер. Посмотрим дальше.

Данные по рейтингу социально-экономического положения

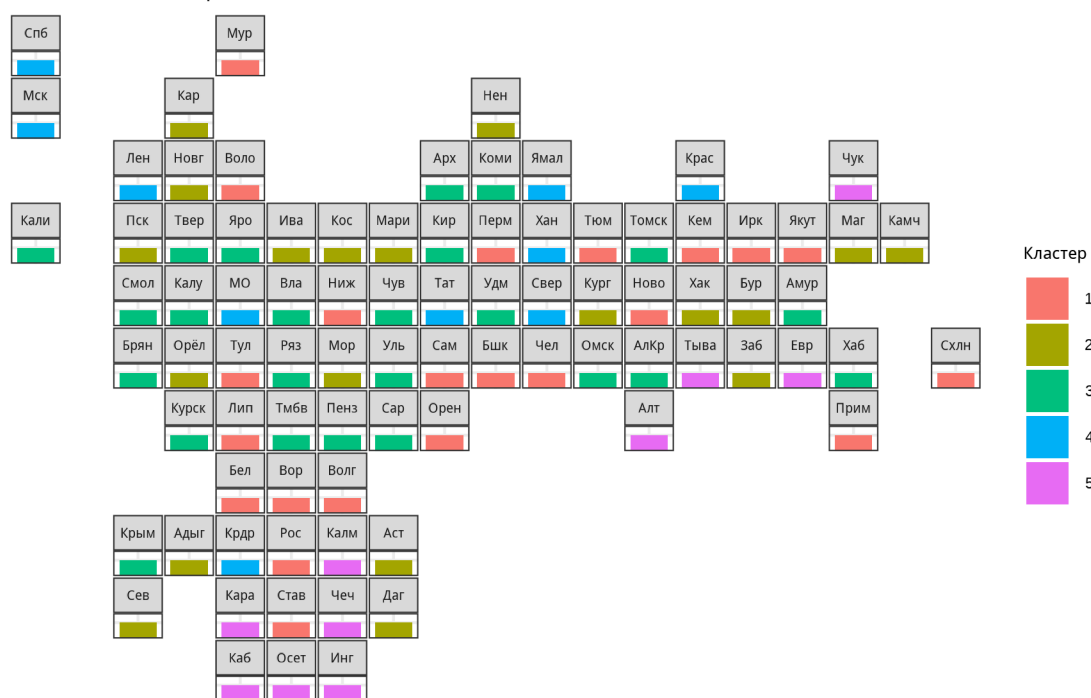
Отдельно решили рассмотреть **рейтинг**, составленный нвоостным агентством “РИА НОВОСТИ”. Рейтинг сочетает в себе интегральный индекс, учитывающий бюджетную сферу, показатели производительности регионов и социальной сферы. Аналогично предыдущей кластеризации, кластеризовать будет от 3 до 9 групп. Результат приведен ниже:



Количество кластеров: 4



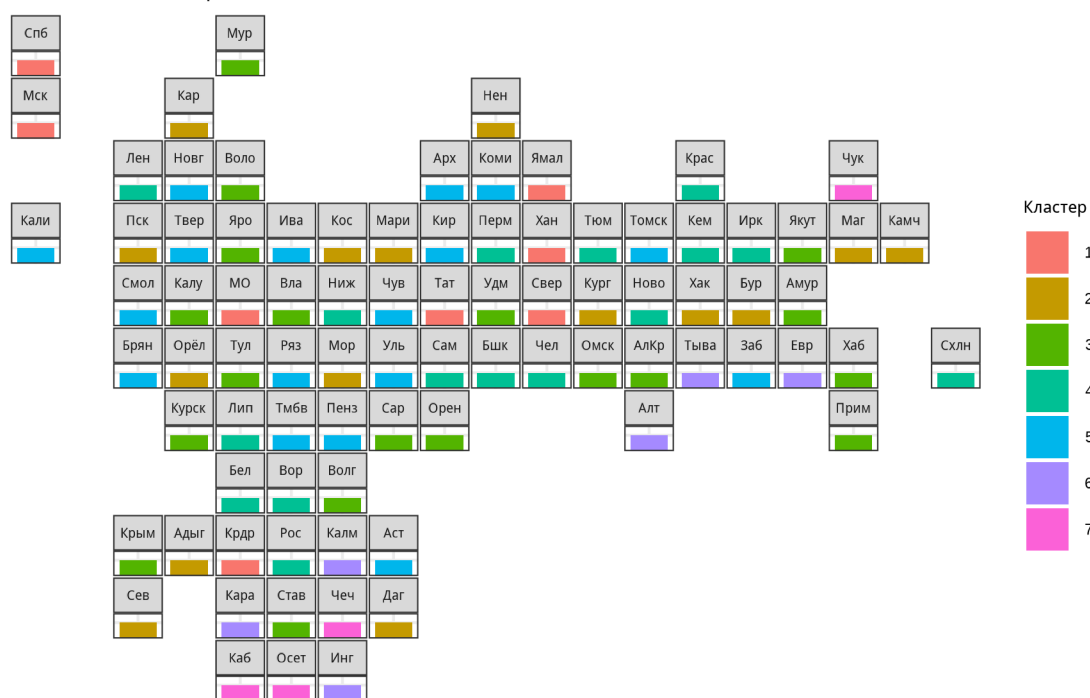
Количество кластеров: 5



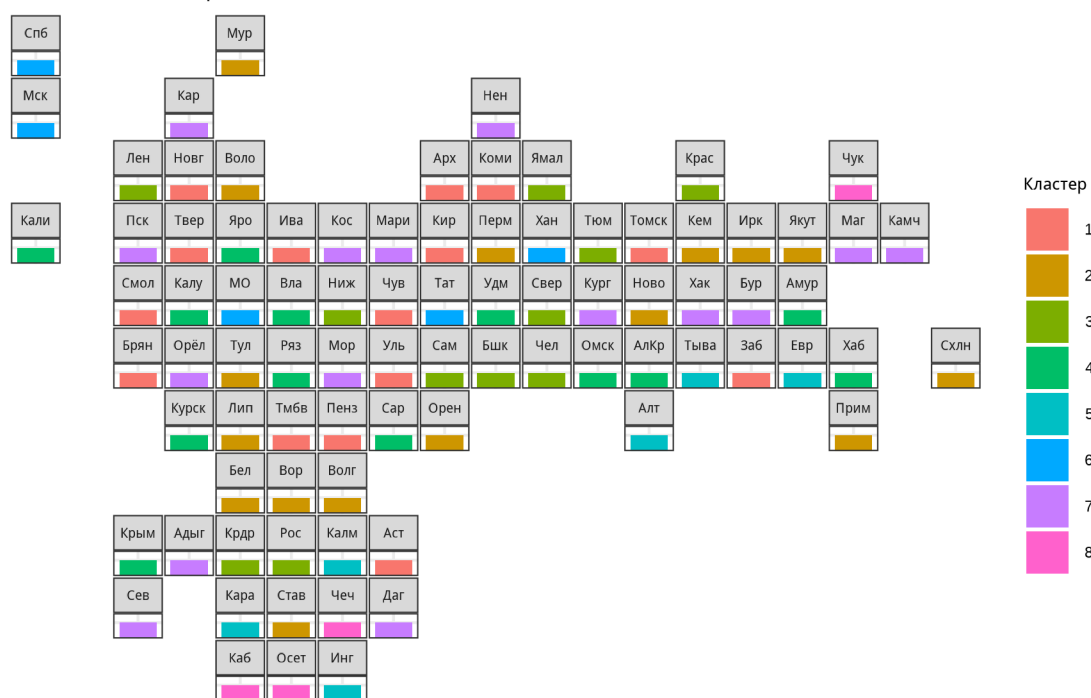
Количество кластеров: 6



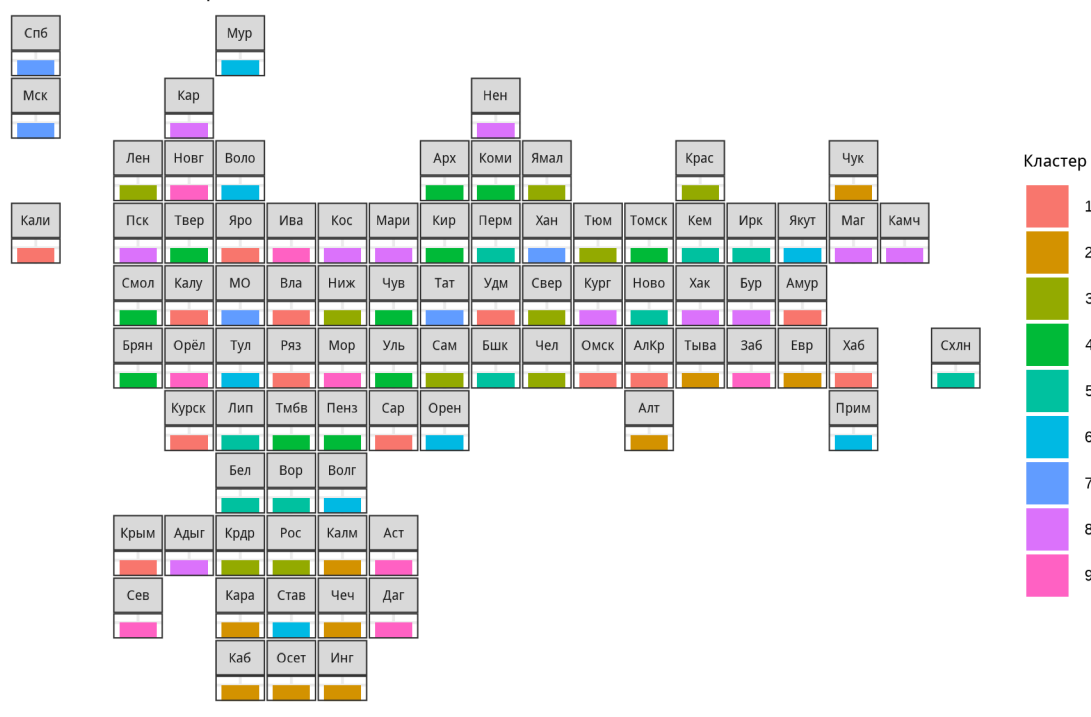
Количество кластеров: 7



Количество кластеров: 8



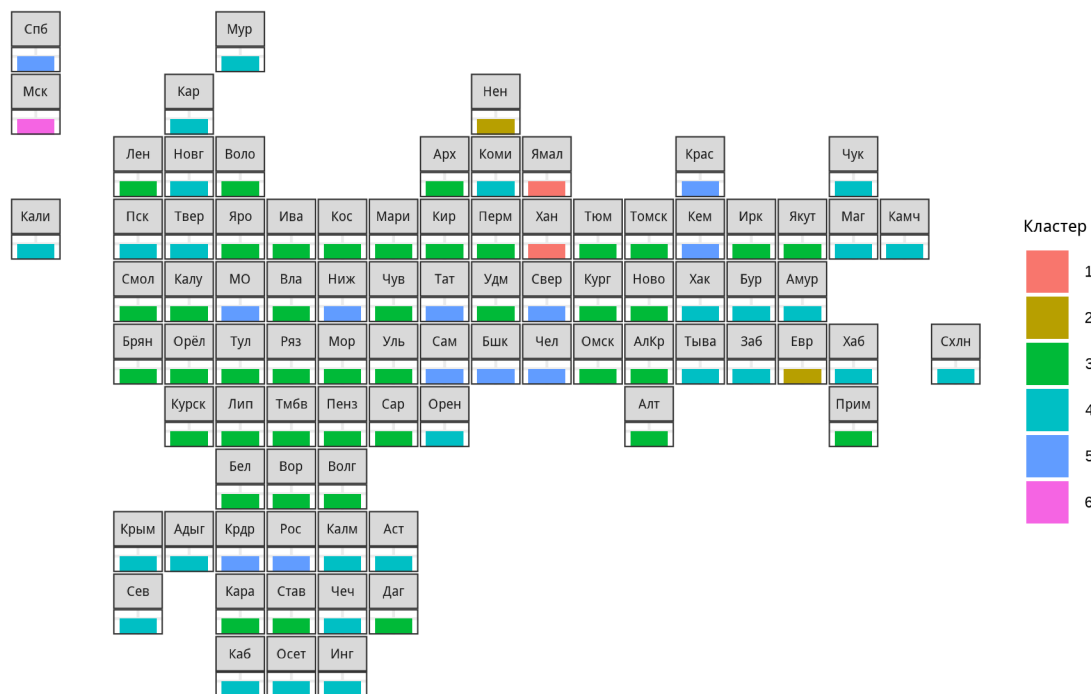
Количество кластеров: 9



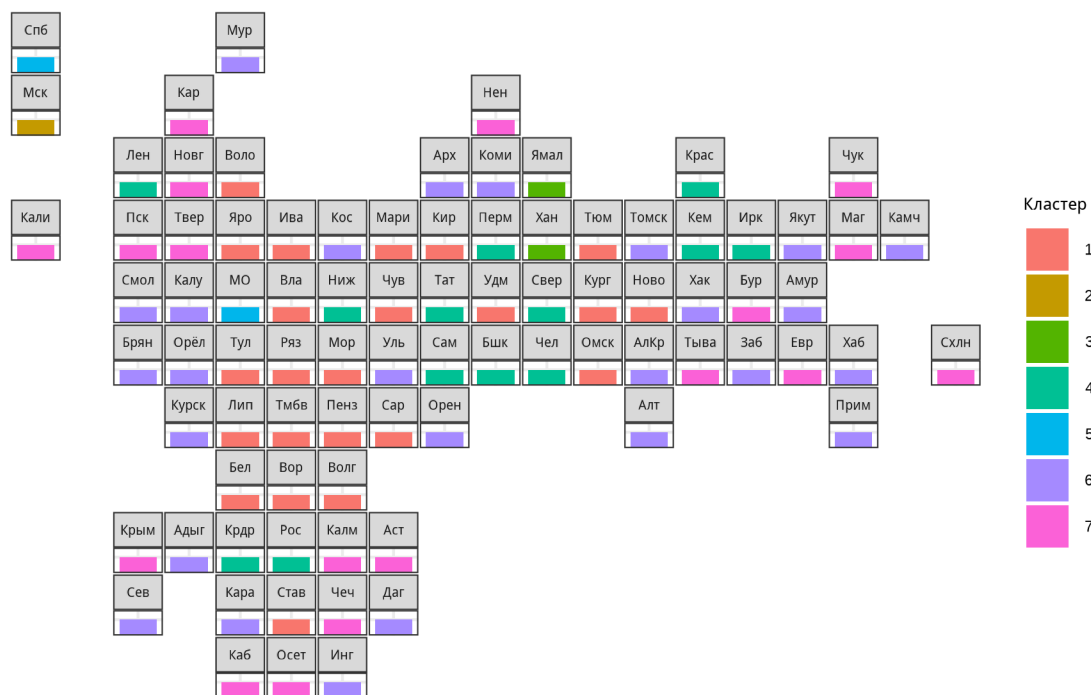
Кластер

- 1
- 2
- 3
- 4

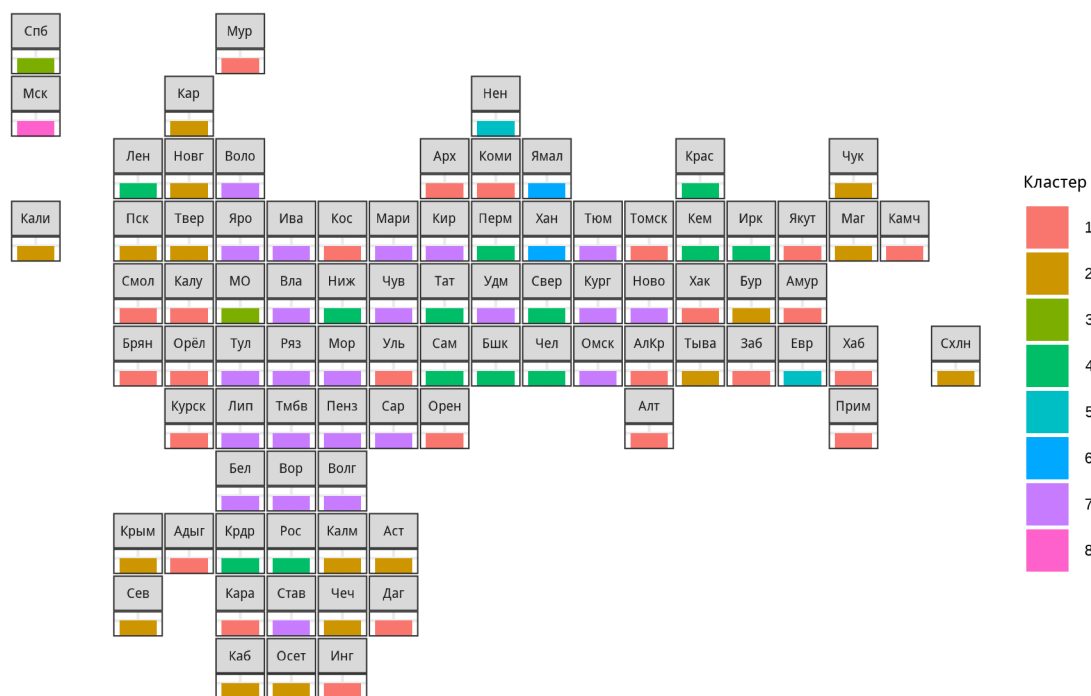
Количество кластеров: 6



Количество кластеров: 7



Количество кластеров: 8



Количество кластеров: 9

