

Жгун Т. В., Проузи Д. К.

**ПРИМЕНЕНИЕ ТЕПЛОВЫХ КАРТ ДЛЯ ВИЗУАЛИЗАЦИИ
ИЗМЕНЕНИЯ КАЧЕСТВА ЖИЗНИ РЕГИОНОВ РФ**

Татьяна Валентиновна Жгун

кандидат физико-математических наук

Tatyana.Zhgun@novsu.ru

Давид Кристофер Проузи

студент

prowsedavid@yandex.ru

Новгородский государственный университет имени Ярослава Мудрого

**THE APPLICATION OF HEATMAPS FOR VISUALIZING CHANGES IN
LIFE QUALITY FOR THE CONSTITUTE ENTITIES OF THE RUSSIAN
FEDERATION**

Tatyana Valentinovna Zhgun

David Christopher Prowse

Novgorod State University a. Yaroslav the Wise

Аннотация. Главной задачей визуализации является облегчение восприятия данных для их последующего анализа, например, при исследовании закономерностей изменения параметров исследуемой системы. Тепловые карты являются средством для визуального извлечения информации, содержащейся в многомерных наборах данных и обеспечения визуальной оценки эффективности функционирования системы, которая характеризуется этим набором данных. В статье рассматривается применение тепловых карт для визуализации набора данных, характеризующих качество жизни населения РФ. Такое представление информации обеспечивает эффективную реализацию обозначенных выше целей.

Abstract. *The main task of visualization is to facilitate the perception of data for its subsequent analysis. For example, data visualization is necessary when studying the patterns of change in the parameters of the system under study. Heat maps are a means for visually extracting information from multidimensional data sets and provide visual evaluation of a systems condition. The article considers the use of heat maps for visualizing a data set characterizing the quality of life of the Russian population. Such visualization of data effectively realizes the above goals.*

Ключевые слова: *Визуализация данных, тепловые карты, качество системы, комплексный индикатор качества жизни*

Keywords: *Data visualization, heatmap, system quality, complex indicator of quality of life*

Введение

Задача визуализации информации в широком смысле слова насчитывает десятки тысяч лет. Наскальные изображения, первые карты местности, а теперь и современная картография, томография, системы распознавание лиц, интернет-навигация — вот только часть диапазона задач, которые объединены общим понятием визуализации. В узком смысле методы визуализации рассматриваются применительно к обработке массивов числовых данных.

Сегодня визуализация данных особенно важна, так как люди теряются в обилии окружающей информации, и на ее восприятие тратится слишком много времени. По статистике, 60 % людей воспринимают быстрее именно визуальную информацию. Информация человеческим мозгом воспринимается гораздо легче, если она представлена в виде графиков, схем или иных графических элементов. Визуализация данных сейчас находится в центре внимания многих аналитиков и разработчиков. У IT-гигантов: *Oracle, Microsoft, IBM* имеются специальные инструменты, разработанные для визуализации данных. Компании *Tableau, Qlik* специализируются исключительно на этой проблематике.

Визуализация данных — это мощный инструмент для убеждения других и донесения нужной информации и может быть использована для различных целей:

1. **Описательная цель.** Используется в повседневной деятельности для информирования аудитории о существующем положении вещей. Например, при составлении отчетов о проделанной работе, для характеристики текущей ситуации.

2. **Иллюстративная цель.** Визуальная информация используется вместо развернутого описания сложного процесса. Например, сложный бизнес-процесс представляется в виде схемы.

3. **Исследовательская цель.** Визуальная информация используется при исследовании закономерностей и тенденций изучаемого явления.

Визуально представленная информация в сравнении с обычным текстом и таблицами привлекает намного больше аудитории, увеличивает вовлечение читателей, быстрее воспринимается, легче запоминается.

Главным аспектом задачи визуализации является ее практическое назначение, а именно, облегчение восприятия данных для их последующего анализа, например, при визуализации данных для получения закономерностей изменения параметров исследуемой системы. В этом смысле визуализация по существу является инструментом системы поддержки принятия решения. В частности, визуализация данных оказывается полезной при формировании представления о состоянии системы при оценке изменения ее качества. Графическое представление параметров системы помогает в формировании интуитивной оценки изменения качества системы и подтверждает (или опровергает) вычисленные формальным способом оценки изменения качества функционирования системы.

Одна из главных целей визуализации — удобное и наглядное сравнение множества показателей за один взгляд. Именно быстрое понимание особенностей данных является главным преимуществом визуализации в сравнении с представлением информации в виде таблицы или текста. Данные должны быть представлены так, чтобы их можно было легко сравнивать. Однако при большом

количестве параметров визуализация становится весьма проблематичной. При этом успех во многом зависит от выбранной формы представления данных. Поэтому усилия современных разработчиков направлены именно на решение визуализации многомерных данных

1 Представления данных являются с помощью тепловых карт

Одним из инструментов представления многомерных данных являются сегодня тепловые карты (*heatmap*). Тепловые карты — это тип визуализации, в которой цвет выступает в качестве дополнительного измерения. Исследователи в разных областях используют тепловые карты, чтобы представить многомерные данные [3].

Простейший пример тепловой карты, знакомый нам с детства — карта региона, на которой цветом показана плотность населения — *рисунок 1*. Чтение информации в таком виде не представляет труда.

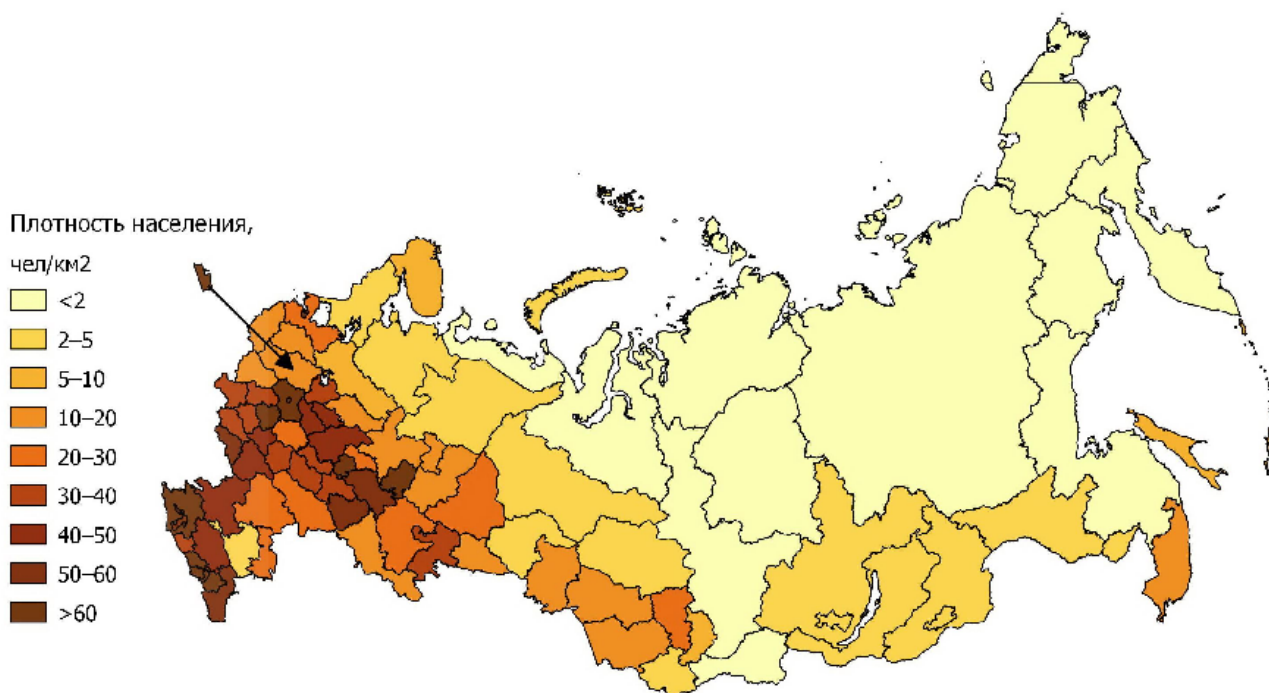


Рисунок 1 — Плотность населения России за 2016 год

Служба такси *Uber* с помощью тепловых карт (рисунок 2) помогает своим водителям определить, где на данный момент находится больше всего потенциальных клиентов. Для этого на карте города красным подсвечиваются зоны с наибольшим количеством заказов такси за последний час.

При представлении матрицы данных с помощью тепловых карт табличное представление информации сохраняется, но числа в ячейках заменяются на заливку этих ячеек цветом по определенному принципу. Если цветное представление нецелесообразно, используются оттенки серого. История тепловых карт приведена в [6].



Рисунок 2 — Тепловая карты службы такси *Uber*

В публикации Wall Street Journal (WSJ). [4] тепловые карты наглядно представляют результаты вакцинации. На *рисунке 3* представлены данные по заболеваемости инфекционными болезнями во всех штатах США. Число инфицированных людей, измеряемое в течение 70 с лишним лет в значительно снизилось после вакцинации — цвет карты изменился на светлый.

Hepatitis A

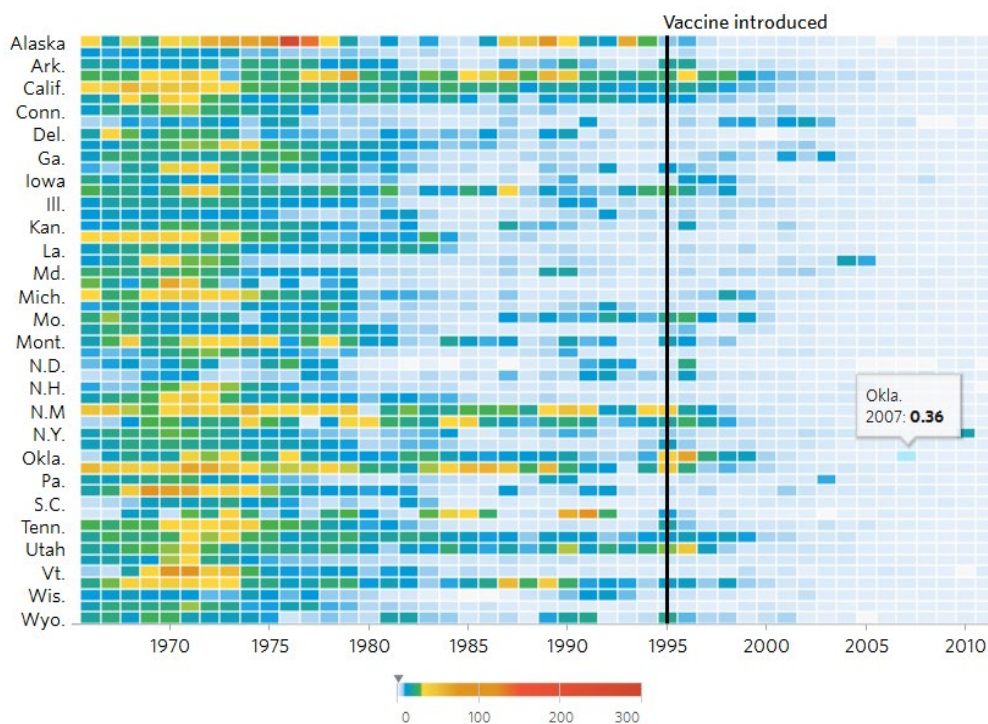


Рисунок 3 — Тепловая карта о заболеваемости гепатитом во всех штатах США

Метод тепловых карт широко используется для визуализации данных в генетике. Тепловая карта — рисунок 4 — позволяет судить о частотах контактов между удаленными геномными элементами [2].

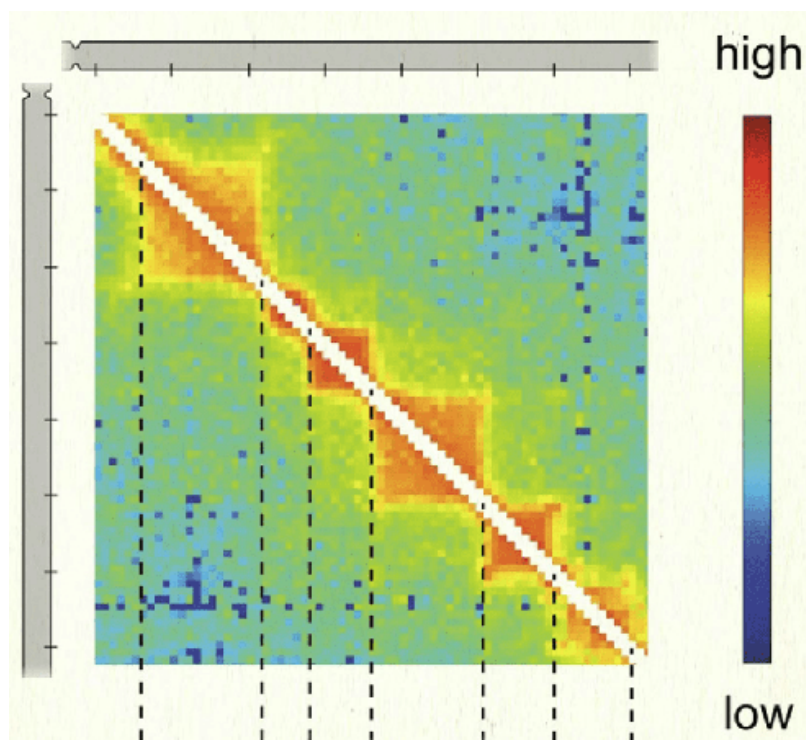


Рисунок 4 — Тепловая карта частотах контактов между удаленными геномными элементами по [2]

2 Применение тепловых карт при анализе качества систем

В рассмотренных выше примерах визуализация рассматривалась для двумерных массивов данных. Визуализация данных более высоких размерностей является значительно более сложной задачей. Очевидно, весьма плодотворной будет применение принципов построения тепловой карты для данных другой природы. В частности, при анализе изменения качества систем, когда нужно рассматривать данные в трех измерениях: объект-показатель-время.

Таблица 1 — Переменные для вычисления интегральных индикаторов качества жизни населения

№ п/п/	Название показателя
1	Отношение ВРП на душу населения к прожиточному минимуму
2	Отношение среднедушевых денежных доходов к прожиточному минимуму
3	Доля численности населения с доходами ниже прожиточного минимума
4	Отношение совокупных доходов 20% самых богатых и 20% самых бедных
5	Обеспеченность населения легковыми автомобилями на 1 000 чел.
6	Доля семей, состоящих на учете на получение жилья
7	Общая площадь жилищного фонда на одного жителя (кв. м/10 чел.)
8	Доля ветхого и аварийного жилья
9	Плотность автомобильных дорог общего пользования (км/10 000 кв. км)
10	Ожидаемая продолжительность жизни при рождении (лет)
11	Число умерших детей в возрасте до 1 года на 1000 чел.
12	Коэффициент естественного прироста (на 1000 чел.)
13	Число умерших от инфекционных, паразитарных болезней и туберкулеза на 100 000 чел.
14	Число умерших от новообразований на 100 000 чел.
15	Число умерших от болезней системы кровообращения на 100 000 чел.
16	Число умерших от болезней органов дыхания на 100 000 чел.
17	Число умерших от болезней органов пищеварения на 100 000 чел.
18	Заболеваемость от травм, отравлений и других внешних причин на 100 000 чел.
19	Число инвалидов на 1 000 чел.
20	Зарегистрировано случаев заболевания врожденными аномалиями на 1 000 чел.
21	Доля специалистов с высшим образованием среди занятых в экономике (%)
22	Приведенная производительность труда (ВРП к среднегодовой численности занятых в экономике, тыс. руб./чел)
23	Количество выпускников высших и средних учебных заведений на 1000 чел.
24	Уровень безработицы (%)
25	Доля работников занятых во вредных и опасных условиях труда в среднегодовой численности занятых в экономике
26	Численность пострадавших на производстве со смертельным исходом или с утратой трудоспособности на 1 раб. день и более на 1 000 работающих
27	Коэффициент миграционного прироста на 10 000 населения

28	Число зарегистрированных умышленных убийств на 100 000 чел.
29	Число фактов умышл. причинения тяжкого вреда здоровью на 100 000 чел.
30	Число зарегистрированных изнасилований на 100 000 чел.
31	Число разбоев, грабежей, краж из квартир граждан в расчете на 100 000 чел.
32	Число зарегистрированных присвоений или растрат в расчете на 100 000 чел.
33	Число состоящих на учете с диагнозом наркомания и токсикомания на 100 000 чел.
34	Число состоящих на учете с диагнозом алкоголизм в расчете на 100 000 чел.
35	Число больных, инфицированных туберкулезом в расчете на 100 000 чел.
36	Смертность от внешних причин (на 100 000 чел.)
37	Число больных психическими расстройствами на 100 000 чел.

Рассмотрим в качестве многомерного набора данных данные государственной статистики, характеризующей качество жизни населения субъектов РФ. Качество жизни за 2010–2017 годы, согласно [1], характеризуют переменные, представленные в таблице 1. Рассматриваем значения 37 переменных для 83 субъектов Российской Федерации при наблюдении за 8 лет. Задачей исследователя является оценка изменения качества жизни за рассматриваемой период с разных точек зрения.

Численные значения величин предварительно приведены в единую шкалу. Значения переменных лежат на интервале $[0, 1]$, при этом большее значение параметра соответствует более высокому качеству системы и изображается более темным оттенком. При последовательном просмотре по годам отчетливо видно, как темнеет общий тон картинке. На рисунке 5 представлены данные для начала наблюдения — 2010 год и для последнего года 2017. Строки соответствуют субъектам РФ, столбцы — значения переменных. Тон стал более темным.

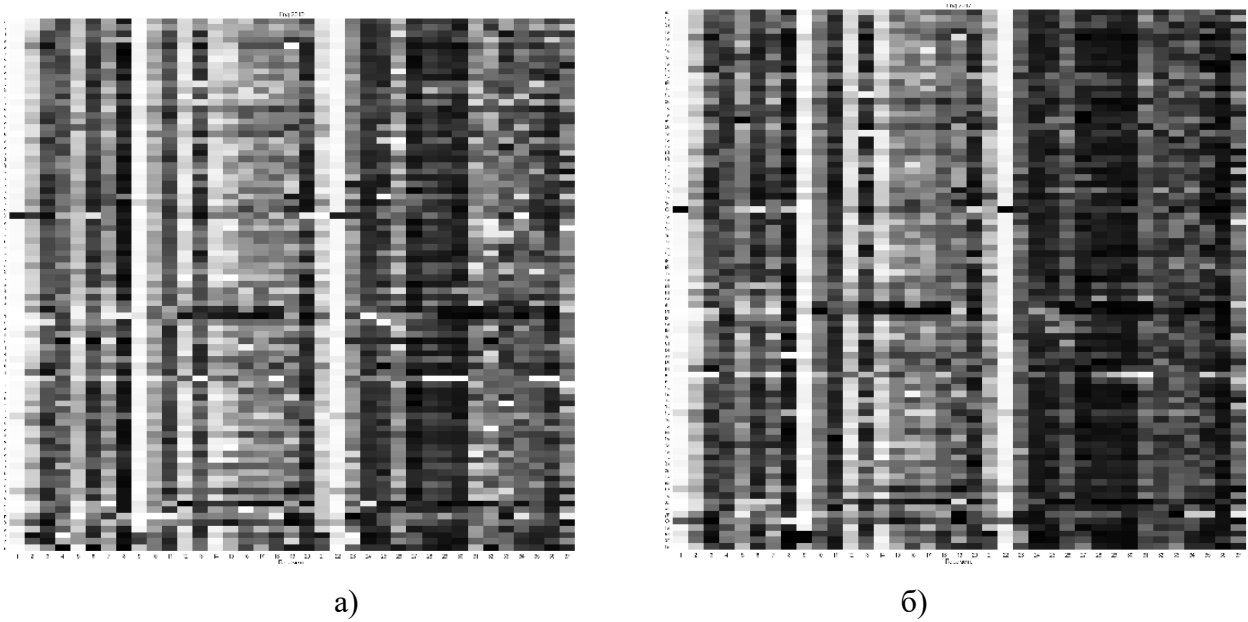


Рисунок 5 — Визуализация переменных, характеризующих качество жизни для субъектов РФ по годам для 2010 (а) и 2017 (б) годов.

Если рассматривать проекцию трехмерного куба данных по переменным, то хорошо видны выбросы данных для некоторых объектов. На рисунке 6 отчетливо видны выбросы в значениях переменной 36 — Смертность от внешних причин (на 100 000 чел.) — в Тыве для 2010 года и в Астраханской области для 2010 и 2013 годов. Резкое улучшение ситуации (потемнение второго и третьего столбца) для 2011 и 2012 годов заставляет задуматься о корректности данных для этих наблюдений.

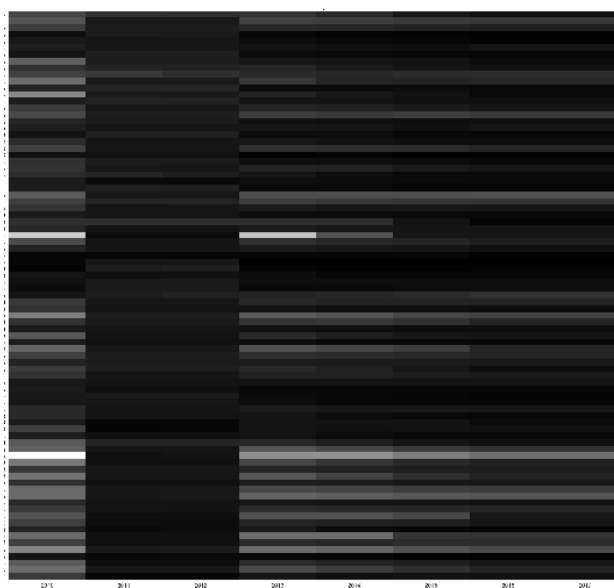


Рисунок 6 — Наличие выбросов для переменной 36. Выброс для Тывы — 2010 год, для Астраханской области — 2010 и 2013 годы

При наблюдении по переменным при изменении фиксированного параметра во времени улучшение ситуации для всех регионов — иллюстрирует рисунок 7.

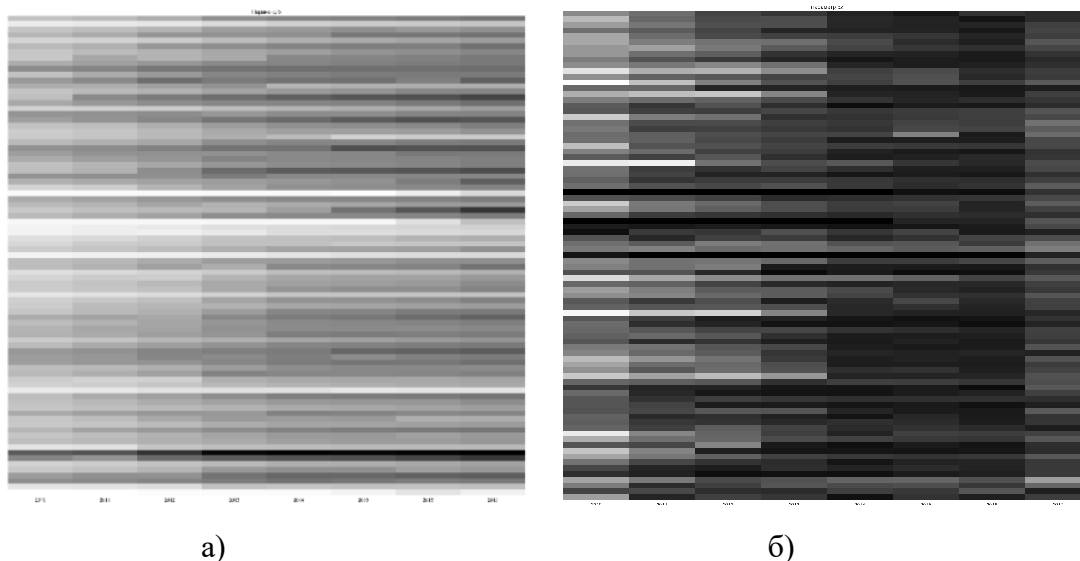


Рисунок 7 — Улучшение наблюдаемых переменных 5 — Обеспеченность населения легковыми автомобилями на 1 000 чел. и 32 — Число зарегистрированных присвоений или растрат в расчете на 100 000 чел.

Для других переменных улучшение со временем не происходит. Отсутствие изменений для переменных 12 и 13 и ухудшение для переменной 17 представлено на рисунке 8.

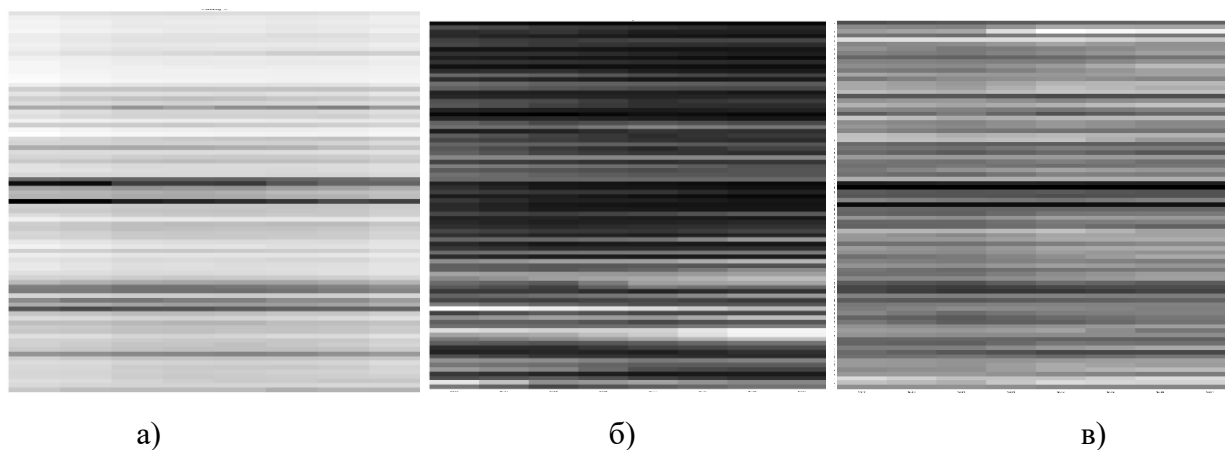


Рисунок 8 — Отсутствие изменений для переменных 12 (а) — Коэффициент естественного прироста, 13 — Коэффициент естественного прироста (б) и ухудшение ситуации для переменной 17 (в). — Число умерших от болезней органов пищеварения

Можно посмотреть на ситуацию для каждого из субъектов РФ. На *Рисунке 9* можно увидеть, что для Рязанской области ситуация неоднозначная. Нельзя

сказать, чтобы она улучшалась по всем наблюдаемым параметрам. Для переменных 16 и 17 мы наблюдаем ухудшение ситуации.

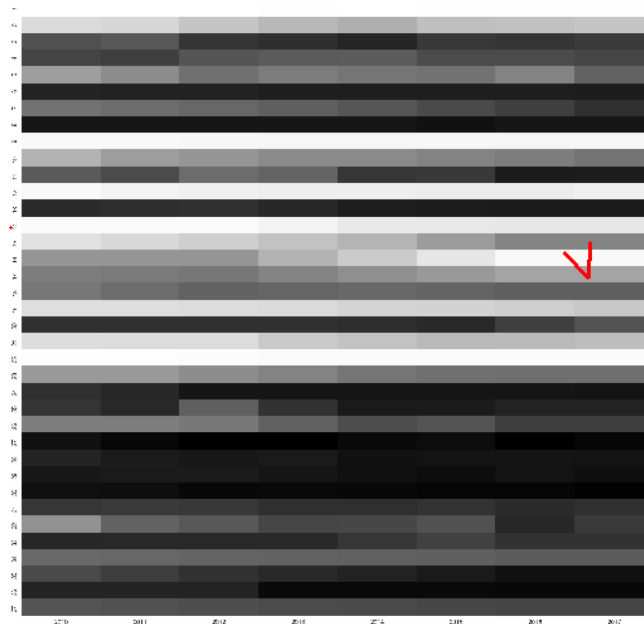


Рисунок 9 — Изменение качества жизни для Рязанской области. Для переменных 16 — Число умерших от болезней органов дыхания 17 — Число умерших от болезней органов пищеварения наблюдается ухудшение (отмечены галочкой).

На рисунке 10 представлена ситуация для Калининградской области. В этом случае мы видим только улучшение рассматриваемых переменных (или не ухудшение).



Рисунок 10 — Изменение качества жизни для Калининградской области.

Заключение

Визуализация оказывает неоценимую помощь исследователям при получении знаний об объекте исследований, особенно если такие исследования связаны с обработкой многомерных данных большого объема. При этом выбор способов визуализации субъективен. Главная цель визуализации — упростить и ускорить восприятие информации, представить данные так, чтобы с ними было легко работать. При графическом исследовании состояния системы при оценке изменения ее качества удобен формат тепловых карт.

Тепловые карты имеют две основные цели: предоставление средства для визуального извлечения информации, содержащейся в многомерных наборах данных и обеспечение визуальной оценки эффективности функционирования системы, которая характеризуется этим набором данных. Применение тепловых карт для визуализации набора данных, характеризующих качество жизни населения РФ, обеспечивает эффективную реализацию обозначенных выше целей.

Список литературы

1. Исакин, М. А. Модификация метода k-средних с неизвестным числом классов / М. А. Исакин. Текст: непосредственный // Прикладная эконометрика. — 2006. — № 4. — С. 62–70.
2. Разин, С. Новый взгляд на геном: не просто цепочка генов, а трехмерная сеть, интегрирующая функциональные домены ядра / С. Разин. Текст: электронный // Биохимия. — 2018. — 28 июня. — URL: <https://biomolecula.ru/articles/novy-vzgliad-na-genom-ne-prosto-tsepochka-genov-a-trekhmernaia-set-integriruiushchaia-funktsionalnye-domeny-iadra>.
3. Романова, И. К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах / И. К. Романова. Текст: непосредственный // Наука и Образование. МГТУ им. Н. Э. Баумана. Электрон. журн. — 2016. — № 03. — С. 133–167. DOI: 10.7463/0316.0834876.

4. DeBold, T. Battling Infectious Diseases in the 20th Century: The Impact of Vaccines / DeBold Tynan, Friedman Dov // Wall Street Journal, Published. 2015. 11 Feb. URL: <http://graphics.wsj.com/infectious-diseases-and-vaccines/>.

5. Jeffery, G. A full-scale security visualization effectiveness measurement and presentation approach / Jeffery Garae, Ryan KL Ko, Mark Apperley // 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) . IEEE pub. 2018. P. 639–650.

6. Wilkinson, L. The History of the Cluster Heat Map / L. Wilkinson, M. Friendly // American Statistician. 2009. Vol. 63, №. 2. P. 179–184.