

Представление данных переписи: визуализация данных

Выбор тем в международных переписях населения¹

Выпущено в июне 2017 года

ВВЕДЕНИЕ

В настоящей технической записке описываются эффективные методы представления и визуализации данных переписей и опросов, осуществляемых национальными статистическими службами (NSO, в соответствии с английским акронимом).

Визуализация данных помогает NSO охватить более широкую аудиторию, эффективнее передавать высокоприоритетную информацию и обнаруживать скрытые закономерности. Наглядные представления данных можно создавать с помощью стандартного офисного программного обеспечения, однако создание иллюстраций профессионального качества требует дополнительных ресурсов. Но в конечном итоге это окупается, поскольку графика значительно повышает эффективность статистических продуктов.

Будучи органами, отвечающими за сбор данных, национальные статистические службы должны уделять приоритетное внимание подготовке эффективных визуальных материалов как для существующих аналитических продуктов, так и в качестве самостоятельных продуктов в целях повышения доступности официальной статистики для пользователей данных, директивных органов и общественности.

Термины

Наглядный материал: группа графических и текстовых элементов, используемых для передачи информации, например, диаграмма, карта, таблица

Визуализация: процесс создания наглядного материала

Настоящая техническая записка завершает документ из двух частей, посвященный *Communicating With Census Data*. Он объясняет важность доведения результатов переписи до сведения широкой аудитории путем подготовки результатов обработки данных, вызывающих интерес и отклик у потребителя.

¹ Настоящая техническая записка является одной из серии «Избранные темы международных переписей населения», в которой рассматриваются вопросы, представляющие интерес для международного статистического сообщества. Бюро переписи населения США помогает странам совершенствовать национальные системы статистики, содействуя устойчивому развитию статистических компетенций.

В первой части технической записки обсуждаются важные вопросы, связанные с визуализацией данных, включая значение визуальных материалов, способы их создания и рекомендации. Во второй части приводятся рекомендуемые методы создания трех видов иллюстраций: таблиц, диаграмм и карт.

СОЗДАНИЕ ИЛЛЮСТРАЦИЙ

Данные можно представлять в различных визуальных форматах и на разных носителях, дополняющих друг друга. Например, таблицы данных имеют большую ценность, но если предоставить только таблицы без диаграмм или карт, пользы от такого предоставления данных будет меньше.

Большинство статических визуальных материалов создаются с помощью стандартного офисного программного обеспечения, но есть и такие, которые требуют более сложных инструментов и навыков.

Визуализация данных

Человек естественным образом формирует **визуальные образы** в своем сознании (Few 2012, p. 65). Во время чтения, сна, выполнения повседневных задач человеческий мозг способен сжимать огромное количество информации, формируя визуальное представление какого-либо явления. Визуализация — это «инструмент глаз и мозга, благодаря которому мы воспринимаем то, что находится за пределами их естественной досягаемости» (Cairo 2013, p. 10).

При **визуализации данных** эта врожденная способность используется, чтобы помещать в контекст большие объемы данных и графики. Согласно определению, «графика на основе данных наглядно отображает измеренные величины путем использования точек, линий, системы координат, чисел, символов, слов, затенения и цвета» (Tufte 1983, p. 9).

Технический прогресс и стремительный рост количества доступных данных привели к расширению доступных возможностей их визуализации (Caio 2011, p. 14). Визуализация **данных переписей и опросов** может быть полезной по нескольким причинам:

- **Демонстрация закономерностей, которые иначе остались бы незамеченными.** Даже когда исходные данные характеризуются большим объемом и уровнем сложности, продуманные визуальные репрезентации позволяют продемонстрировать наиболее важные закономерности. Эти закономерности трудно выявить без помощи иллюстраций.
- **Быстрая передача информации.** Большинство людей неспособны интерпретировать таблицу данных без детального анализа. Визуализация превращает таблицу данных в графический объект, который можно легко интерпретировать.
- **Предоставление убедительных доказательств.** Визуализация данных подкрепляет тезисы автора, облегчая передачу знаний аудитории.

Эксперты в предметной области могут использовать визуальные материалы на этапе разведочного анализа данных (см. раздел «Разведочный анализ данных» в первой части документа). Однако данная техническая записка касается использования визуализации данных для сообщения результатов иной аудитории.

Типы и форматы иллюстраций

Одна из первых задач при подготовке визуальных материалов — определение формата для обнародования или публикации. Требования к **бумажным** и веб-продуктам отличаются. При выборе бумаги необходимо обращать внимание на ее вес, толщину, покрытие и яркость. **Электронные** продукты, как правило, требуют надежного соединения с Интернетом и должны быть доступны для просмотра на различных устройствах (включая персональные компьютеры, смартфоны и планшеты).

Визуальный материал может быть **статическим** или **интерактивным**. Статические визуальные материалы наиболее распространены, так как их создавать проще всего. Многие виды статических иллюстраций можно создавать с помощью стандартного офисного программного обеспечения.

Интерактивные визуальные материалы чаще используются в интернет-продуктах. Правильно реализованные интерактивные иллюстрации могут отображать большой объем информации и дольше удерживать внимание аудитории. Однако они требуют более широких технических возможностей и иногда — привлечения внешних консультантов.

Основное внимание в этой технической записке уделяется статическим изображениям, поскольку их проще использовать в целях национальных статистических служб.

Статические изображения

Аналитические продукты национальных статистических служб предоставляют информацию об основных тенденциях и закономерностях на протяжении определенных периодов времени и в пределах определенной территории. Для большинства аналитических продуктов национальных переписей и опросов основные виды статических изображений — это **таблицы, диаграммы и карты**. Они популярны благодаря их эффективности и простоте использования для передачи информации.

У каждого вида иллюстраций есть свои сильные и слабые стороны. В конечном счете выбор визуального представления — это

компромисс, состоящий в принятии определенных недостатков формата, если они перевешиваются его сильными сторонами. Подробнее об этом — ниже.

В поле 1 описывается процедура создания иллюстрации. Авторы могут поддасться искушению использовать в рамках всего продукта однообразный набор иллюстраций для всех возможных переменных, однако не каждую переменную нужно наглядно иллюстрировать. Следует оценить преимущества и недостатки визуализации данных, чтобы определить, увеличивает ли иллюстрация ценность продукта или отвлекает аудиторию от ключевого сообщения.

Требуемые навыки и программное обеспечение

Поле 1.

Процедура создания иллюстрации

- 1) Какие данные нужно визуализировать?** В качестве показателя для наглядного отображения рекомендуется выбрать главную переменную. Обычно она становится осью X в двумерной системе координат, на которой изображается график. Это могут быть годы проведения переписи, категории ответа и т. д.
- 2) Почему эти данные необходимо визуализировать?** У иллюстрации должно быть конкретное назначение. Выберите интересные закономерности, которые можно показать только с помощью иллюстрации.
- 3) Какой вид иллюстрации выбрать?** Выбор зависит от особенностей данных. Например, если нужно продемонстрировать территориальную закономерность, лучше сделать это с помощью карты. Если же нужно показать динамику закономерности по времени, скорее всего, подойдет график. Для всех видов иллюстраций характерны те или иные компромиссы.
- 4) Каким образом иллюстрация дополняет другие иллюстрации в рамках продукта?** Если иллюстрация не является самостоятельным продуктом, ее необходимо выбирать с учетом наличия других иллюстраций, формы публикации, доступного места и темы продукта.
- 5) Как создать иллюстрацию?** Для создания иллюстраций требуется соответствующее программное обеспечение. Карты обычно создают в геоинформационных системах (GIS, в соответствии с английским акронимом), а диаграммы и таблицы можно строить с помощью офисных приложений.

Источник: U.S. Census Bureau.

Создание визуальных материалов профессионального качества требует навыков в области **графического дизайна, веб-дизайна, программирования, картографии и управления данными**. Кроме того, нужны художественные и аналитические способности.

Многие приложения позволяют создавать привлекательные визуальные материалы профессионального качества. Некоторые популярные инструменты такого рода, как платные, так и с открытым кодом, перечислены на рисунке 1. Следует иметь в виду, что некоторые программные средства требуют углубленного изучения и обширной практики, особенно системы для графического дизайна, разработки веб-сайтов, картографии и анализа данных.

Инструменты визуализации данных в стандартных офисных пакетах позволяют при минимальной дополнительной подготовке создавать иллюстрации на основе данных, например таблицы и диаграммы, подходящие для большинства задач. Такие графические элементы можно варьировать в широких пределах, при этом их качество может достигать профессионального уровня.

Однако для описываемых задач подходят не все доступные возможности подобных приложений. Чтобы обеспечить легкость интерпретации, изображения не должны содержать лишнего. Рекомендации по конкретным видам иллюстраций подробнее рассматриваются ниже.

Рисунок 1.

Программные средства для создания иллюстраций

Для создания иллюстраций некоторых типов требуются специальные приложения.

Категория	Примеры	Функциональные возможности	Уровень сложности
Офисное программное обеспечение	Microsoft Excel, OpenOffice	Создание простых иллюстраций	Низкий
Специализированные средства визуализации	Tableau, WEAVE	Создание более сложных иллюстраций	Средний
Анализ данных	Stata, R	Обработка и анализ данных	Высокий
Графический дизайн	Adobe Illustrator, Inkscape	Создание графики профессионального качества	Высокий
Картография	ArcGIS, QGIS	Изготовление карт, работа с геопространственными данными	Высокий
Веб-разработка	D3, Leaflet	Создание интерактивных веб-иллюстраций	Высокий

Примечание. Многие инструменты, перечисленные в этой таблице, рассчитаны на профессиональное применение и требуют длительного обучения и практической работы для достижения профессионального уровня.
Источник: U.S. Census Bureau.

Любые из этих приложений можно освоить путем обучения. но курсы дают лишь ограниченное представление о теме. Чтобы стать экспертом по визуализации данных, потребуется обширная практика — возможно, на протяжении нескольких лет. Поэтому в некоторых организациях, чтобы не отвлекать специалистов в предметной области на создание графики, стоит рассмотреть возможность найма профессиональных художников или заключения контрактов с поставщиками услуг по подготовке визуальных материалов.

РЕКОМЕНДАЦИИ

Основная цель визуального материала — улучшить понимание. Для достижения этой цели авторы должны следовать приведенным ниже указаниям. В настоящей технической записке применение рекомендаций показано на примерах Бюро переписи населения США.

Использование структурированных данных

Ключевой компонент визуального элемента — лежащие в его основе данные. Как правило, для создания визуального элемента данные должны быть представлены машиночитаемом формате — Comma Separated Values (CSV, соответствии с английским акронимом), eXtensible Markup Language (XML, соответствии с английским акронимом), JavaScript Object Notation (JSON, соответствии с английским акронимом) или в виде реляционной базы данных. Данные в перечисленных форматах являются **структурированными**, что означает возможность легкого считывания их компьютером для анализа. В свою очередь неструктурированные данные, например в формате Portable Document Format (PDF, соответствии с английским акронимом), с помощью программного обеспечения для анализа данных считывать затруднительно. На рисунке 2 показано различие между структурированными и неструктурированными данными.

Согласно устоявшейся практике, национальным статистическим службам следует выпускать результаты обработки данных в структурированном виде, чтобы облегчить их использование для

целей визуализации как в самой службе, так и вне ее. Применение перечисленных форматов способствует глобальным инициативам по повышению прозрачности и доступности правительственных данных, а также увеличению практической ценности национальных статистических данных (United Nations 2013).

Структурированные данные могут размещаться в обычной или перекрестной таблице. Каждая ячейка в обычной таблице соответствует одной переменной, а ячейки в перекрестной таблице — сочетанию нескольких переменных. Визуальный элемент может быть создан с использованием таблицы любого из этих типов в зависимости от цели автора.

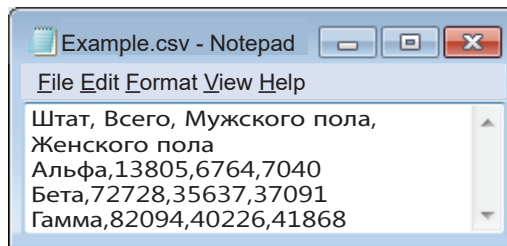
Рисунок 2.

Структурированные и неструктурированные данные

Представленная ниже информационная таблица может показаться структурированной. Однако поскольку она представлена в формате данного документа (PDF), считывание чисел из таблицы с помощью компьютера вызовет трудности. Таким образом, на самом деле таблица является неструктурированной.

Штат	Всего	Мужского пола	Женского пола
Альфа	13 805	6764	7040
Бета	72 728	35 637	37 091
Гамма	82 094	40 226	41 868

На изображении ниже — та же таблица, сохраненная в файле значений, разделенных запятыми (CSV), и открытая в приложении Notepad. Этот формат данных является структурированным: каждый числовой столбец отделен разделителем (в данном случае, запятой), и каждая запись представляет собой отдельную строку.



Источник: U.S. Census Bureau

Наличие контекстных сведений

Для визуализации данных важен **контекст**. Когда объем продукта большой, читатели могут просматривать только иллюстрации, игнорируя текст. Поэтому изображения должны содержать достаточно информации, чтобы аудитория могла ее быстро интерпретировать.

Название — первое, на что обращают внимание; оно должно носить описательный характер. В названии следует четко указать переменные, демонстрируемые с помощью иллюстрации, и охваченный период. В названии также можно передать основную мысль диаграммы.

Приведите **источники**, которые использовались для создания иллюстрации. Сведения об опубликованных отчетах должны присутствовать на самой иллюстрации с перекрестными ссылками на сноски или полный список источников в конце документа. В случае одиночных иллюстраций, предназначенных для плакатов или публикации в Интернете, сведения об источнике должны быть указаны в полном объеме. В сведения об источнике включается имя автора или название организации (например, национальной статистической службы (NSO); мероприятие, отчет и год (например, перепись 2010 Census, Summary File 1, 2011); а также гиперссылка — при наличии.

Также в визуальные материалы следует включать инструкции по **использованию данных**. Применительно к национальным переписям и опросам такие инструкции, как правило, содержат ссылки на информационные документы, содержащие определения переменных величин, погрешности выборки и систематической ошибки (например, погрешности охвата и ответа).

Что касается таблиц данных, обычно для доступа к **структурированному (машиночитаемому) набору данных** предоставляется гиперссылка. Как указывалось выше, предпочтительно, чтобы организации предоставляли доступ к структурированным наборам данных — в целях содействия использованию данных госструктурами, научно-исследовательским сообществом и частным сектором.

Использование графики

Графический дизайн — это искусство. Для создания привлекательных визуальных материалов профессионального уровня нередко требуются услуги графических дизайнеров. Тем не менее, даже пользователи с минимальными художественными способностями могут создавать изображения достаточно высокого качества, эффективно используя графические элементы, рассматриваемые ниже.

При использовании этих графических элементов авторам следует проявлять особое внимание. В частности, некоторый процент населения не может различать отдельные цвета из-за дальтонизма (как правило, оттенки красного и зеленого), поэтому обычно нужно использовать цветовые палитры, учитывающие эту особенность. Цвет также может иметь положительное или отрицательное культурное значение. Следует избегать перегруженности иллюстрации такими элементами, так как это отвлекает зрителя, сводя на нет эффективность изображения.

Пользуйтесь **цветом**, чтобы привлечь внимание к конкретным элементам иллюстрации, приглушайте менее важные элементы. График, к примеру, может содержать несколько линий, одна из которых отражает тенденцию. Если сделать ее красной, а

остальные — светло-серыми, это позволит привлечь внимание к самому важному элементу иллюстрации.

Символы позволяют точнее передать соотношение, чем круговые диаграммы, так как последние вызывают сложности интерпретации. На градуированных символьных картах символы разного размера выражают абсолютное значение (чем больше размер, тем больше значение, и наоборот).

На взаимодействие аудитории с иллюстрацией также влияет выбор **шрифта**. На картах, например, разные шрифты могут использоваться для указания названий природных элементов (водоемов и др.) и социокультурных элементов (названий населенных пунктов и др.). С помощью цвета и начертания шрифта (полуужирный, курсив и т. д.) тоже можно выделять одни элементы и приглушать другие.

Другие примеры графических элементов, которые можно использовать в иллюстрациях, включают ориентацию, форму, длину и ширину линий, размер, кривизну, отметки, обводку, насыщенность, пространственное расположение и движение (Knaflic 2015, p. 105).

ПРИМЕРЫ ИЛЛЮСТРАЦИЙ

Дальнейшая часть данной технической записки посвящена рекомендациям по созданию действенных, привлекательных, выразительных иллюстраций на основе данных переписей и опросов. Рассматриваются иллюстрации трех основных типов: диаграммы, карты и таблицы. Данные рекомендации дополняют приведенные выше общие советы, применимые к иллюстрациям всех видов.

ДИАГРАММЫ

Диаграммы различного уровня сложности удобно использовать для наглядного представления национальной статистики. Для простоты в технической записке описываются диаграммы четырех типов: **столбчатые диаграммы, графики, демографические пирамиды и пиктограммы**.

СТОЛБЧАТЫЕ ДИАГРАММЫ

Столбчатые диаграммы используются для отображения дискретных или категориальных переменных переписи либо опроса за конкретный период или их динамику по времени. Такие диаграммы могут иметь форму одиночных столбцов для конкретной категории, составных диаграмм (когда столбцы размещаются поверх друг друга) и двойных столбчатых диаграмм (для непосредственного сравнения показателей). Столбцы могут быть как вертикальными, так и горизонтальными.

На рисунке 3, например, изображена составная столбчатая диаграмма, на которой общий коэффициент демографической нагрузки разделен на показатели по молодежи и пожилым. Можно видеть, что коэффициент демографической нагрузки пожилых со временем увеличивается по отношению к показателю по молодежи.

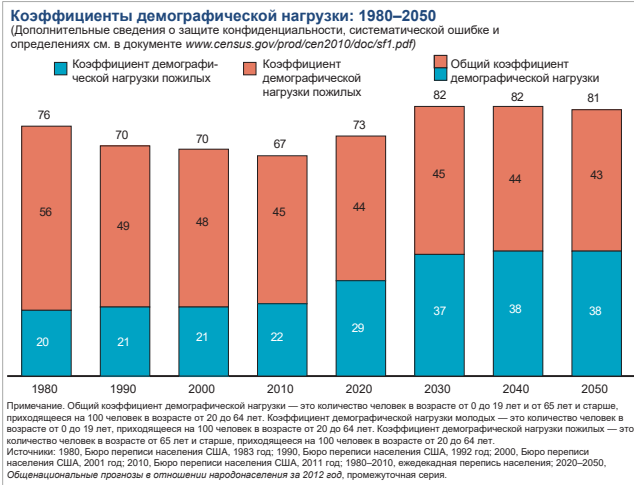
Преимущества: позволяет представить большие наборы данных в простом формате; возможность сопоставить одновременно несколько категорий; доступно для понимания широкой аудитории.

Недостатки: субъективность, зависимость от порядка следования категорий и масштабирования; во избежание перегруженности допустимо отображение лишь ограниченного количества столбцов.

Рисунок 3.

Пример иллюстрации: столбчатая диаграмма

Столбчатые диаграммы удобно использовать для демонстрации количественных отношений. На рисунке приведен пример составной столбчатой диаграммы, в которой одна группа столбцов (коэффициент демографической нагрузки молодежи) помещена поверх другой (коэффициент демографической нагрузки пожилых). Данная диаграмма подтверждает тезис автора о старении населения США, убедительно демонстрируя, что коэффициент демографической нагрузки пожилых вырос по отношению к аналогичному показателю по молодым.



Источник: U.S. Census Bureau, 2014.

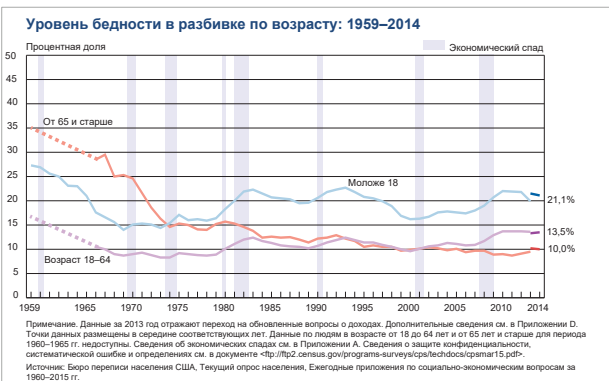
ГРАФИКИ

Графики используются для отражения развития тенденций, присутствующих в результатах последовательных переписей населения или опросов, в зависимости от времени. Например, график на рисунке 4 показывает, как уровень бедности увеличивается и уменьшается со временем, согласно данным переписей и опросов. Наклон линии на графике имеет определенное значение при использовании как временных, так и невременных данных.

Рисунок 4.

Пример иллюстрации: график

Графики удобно использовать для демонстрации тенденций. В данном случае уровни бедности представлены в разбивке по возрастам, чтобы показать, как росли или уменьшались показатели по определенным группам населения. Приведена дополнительная информация — путем указания лет, в которые национальная экономика находилась в состоянии спада. Однако эти сведения помещены на задний план, чтобы не отвлекать от главных результатов. Добавлены примечания, разъясняющие причину появления последовательности данных за период 2013–2014 гг., когда вопросы анкетирования изменились.



Источник: U.S. Census Bureau, 2015.

Преимущества: возможность наблюдать тенденции в зависимости от времени; удобно для отображения нескольких переменных одной категории; возможность объединения с другими наглядными элементами для выделения важных областей тенденции.

Недостатки: ограничение допустимого количества линий во избежание перегруженности (хотя это количество больше, чем в случае столбчатых диаграмм); для четкого отражения динамики тенденции по времени требуется большое количество элементов данных.

ДЕМОГРАФИЧЕСКИЕ ПИРАМИДЫ

Демографические пирамиды (см. пример на рисунке 5) относятся к числу самых эффективных средств наглядного представления структуры населения по результатам переписи или опроса. Например, пирамида, у которой основание (соответствующее населению младшего возраста) шире вершины, относится к стране с иными потребностями, чем страны с пирамидой равномерной ширины. По форме демографической пирамиды аудитория сможет быстро определить траекторию демографического развития страны в предстоящие годы.

Преимущества: наглядное представление структуры населения; возможность увидеть расхождения по возрасту и полу; возможность разбиения по определенным подгруппам, например, по этнической принадлежности, языку или региону.

Недостатки: утрата гранулярности при использовании больших возрастных групп; сложности с добавлением дополнительных переменных.

Пиктограммы

Пиктограммы — это легко узнаваемые аудиторией значки, которые используются для наглядного представления данных. Значки при этом служат для пропорционального отображения количества (например, один значок соответствует 10 000 людям), группы значков используются для отражения общих тенденций. На рисунке 6, например, с помощью значков проиллюстрировано диспропорциональное соотношение количества мужчин и женщин среди военных ветеранов.

Преимущества: понятность для широкой аудитории, возможность быстрого представления общих закономерностей или тенденций; оптимально для публикации в СМИ.

Недостатки: слишком абстрактная репрезентация; как правило, не подходит для профессиональных пользователей данных и аналитических отчетов.

РЕКОМЕНДАЦИИ

Примечание. Примеры применения рекомендаций см. на рисунках 3–6.

Демонстрируйте тенденцию или закономерность: каждая диаграмма выделяет нечто важное в истории, излагаемой автором. Если диаграмма малопонятная или не имеет практической ценности, не включайте ее в продукт.

Данные должны быть однозначно идентифицируемыми: выделяйте каждый элемент данных цветами и границами, чтобы его можно было отличить от других. Например, линии тенденций на рисунке 4 изображены тремя разными цветами, составные столбцы на рисунке 3 имеют разные цвета и отделены границами. Все иллюстрации снабжены расшифровкой условных обозначений,

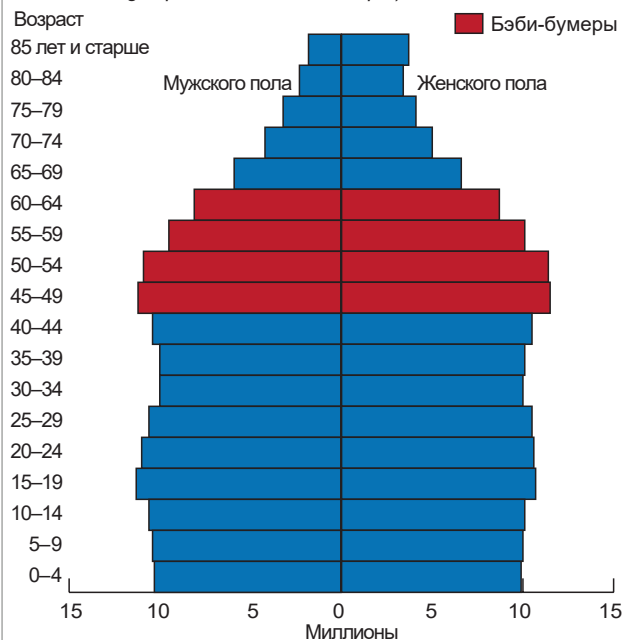
Рисунок 5.

Пример иллюстрации: демографическая пирамида

Демографические пирамиды удобны для наглядного представления распределения населения по возрасту и полу. На данной иллюстрации также выделена группа населения «бэби-бумеры» — люди, родившиеся в период с 1946-го по 1964 год.

Население в разбивке по возрасту и полу, 2010 г.

(Дополнительные сведения о защите конфиденциальности, систематической ошибке и определениях см. в документе www.census.gov/prod/cen2010/doc/sf1.pdf)



Источник: U.S. Census Bureau, 2011; 2010 Census.

Источник: U.S. Census Bureau, 2014.

в которой указано, каким показателям соответствует каждый цвет. На рисунках 3 и 4 также имеются подписи.

Точность осей: значения вдоль осей X (горизонтальной) и Y (вертикальной) должны соответствовать точкам на графиках. В случае динамических рядов точки данных следует размещать с привязкой к периоду времени между соседними точками (см. пример. на рис. 4). Оси, отражающие количественные показатели, желательно начинать с нуля.

Выделяйте основные идеи: подкрепляйте важные тезисы с помощью графических элементов. Например, на рисунке 5 красным цветом выделена одна из возрастных групп, а серым цветом на рисунке 4 обозначена длительность экономического спада в годах.

Учитывайте целевую аудиторию: выбирайте тип диаграммы в соответствии с аудиторией и сообщением. Рисунок 6 доступен для понимания большинству аудиторий. тогда как рисунок 5 больше подходит для экспертов в предметной области, чем для широкой публики. На рисунках 3 и 4 представлены диаграммы стандартного вида, но они могут быть сложны для понимания аудиторией, незнакомой с исходными данными.

Чего следует избегать

3D-графика: пространственная графика обычно неэффективна, отвлекает читателя и может затруднять понимание данных. Не используйте 3D-графику, если в данных отсутствует третье измерение.

Круговые диаграммы: мозг человека испытывает трудности при интерпретации круглых областей и углов. Поэтому круговые диаграммы могут дезориентировать, особенно имеющие больше двух-трех сегментов.

Чрезмерное усложнение: не передавайте слишком много идей одновременно. Диаграмма должна быть нацеленной на передачу главного сообщения и легкой для понимания аудиторией.

Некорректный масштаб: значения масштаба необходимо выбирать корректно, с учетом типа данных. Ограничение масштаба (например, для данных с узким диапазоном значений) может привести к неверной интерпретации или завышению вариативности.

Рисунок 6.

Пример иллюстрации: пиктограмма

Пиктограммы обычно используются в продуктах, рассчитанных на широкую аудиторию. Простые значки, как в данном примере, легко воспринимаются читателями с любым уровнем образования, включая школьников. В данном примере один значок обозначает один миллион ветеранов.

Всего в США 21,8 млн ветеранов.



Источник: U.S. Census Bureau, 2012.

Карты

Карты служат для отображения пространственного распределения данных. Карты позволяют продемонстрировать закономерности, которые нельзя увидеть в таблицах и на диаграммах. Правильно подготовленная карта позволяет подкрепить тезисы и усилить аргументацию автора. Однако неудачно оформленная карта способна запутать аудиторию или вызвать у читателя потерю интереса к продукту.

Карты можно поделить на две широких категории — тематические и справочные. Тематические, как показано на рисунке 7, используются в качестве иллюстрации к определенной теме, например, теме социальной или физической географии. Справочные карты служат для изображения заметных объектов местности, природных особенностей, населенных пунктов и других элементов, используемых для навигации или передачи контекстной информации. В данной технической записке речь идет о тематических картах, так как они чаще всего используются для представления данных переписи или опроса.

ПРЕИМУЩЕСТВА

Территориальные закономерности: возможность продемонстрировать закономерности и несовпадения по регионам, которые сложно увидеть на диаграммах или в таблицах.

Подкрепление тезисов: закономерности, показанные на карте, подкрепляют тезисы автора, предоставляя дополнительные возможности анализа.

Быстрая передача информации: правильно составленная карта позволяет представить колоссальное количество данных в виде понятного и легко интерпретируемого графического изображения.

Побуждение аудитории: карты передают большой объем информации, тем самым привлекая внимание читателей и побуждая их к обсуждению значения демонстрируемых закономерностей.

НЕДОСТАТКИ

Высокие требования к техническим ресурсам:

для создания карт профессионального качества обычно нужны геоинформационные системы (GIS, соответствии с английским акронимом) и программное обеспечение для графического дизайна.

Длительное обучение:

для эффективного использования картографических инструментов требуется длительный период их освоения. Для создания интерактивных веб-карт также нужны навыки разработки веб-сайтов.

Ограниченное количество показателей:

обычно тематическая карта демонстрирует лишь один показатель. С увеличением числа показателей иллюстрация усложняется и эффективность карты снижается, чего могут избежать лишь опытные профессионалы.

РЕКОМЕНДАЦИИ

Примечание. Пример применения данных рекомендаций представлен на рисунке 8.

Указание **групп населения** в условных обозначениях подкрепляет тезисы автора. Для охвата важнейших тем, представленных на карте, идентифицированы конкретные точки изменения тенденции.

Выбрана подходящая **цветовая схема**. Диапазоны чисел выше нуля обозначаются цветами, соответствующими линейному градиенту, а диапазоны чисел ниже нуля — с помощью определенного цвета.

Подписи (почтовые коды штатов) не отвлекают на себя внимание и несут описательный характер. Цвет подписи — черный или белый, в зависимости от затененности фона. Подписи размещены приблизительно в центре многоугольника, но не касаются его границ. В случаях, когда последнее неизбежно, подписи вынесены за пределы штата, при этом сам штат указан линией.

Для некоторых территорий предусмотрены **карты-врезки**. Такие карты позволяют указать дополнительные важные подробности.

Для справочных целей приведен **средний национальный показатель**.

Для уменьшения геометрической сложности геопространственные данные (линии и многоугольники) отображаются в **упрощенном** виде.

ЧЕГО СЛЕДУЕТ ИЗБЕГАТЬ

Слишком большое количество карт: включать в продукт следует лишь карты, подкрепляющие тезисы. Не следует размещать на картах все имеющиеся показатели, особенно те, что отражают известную закономерность (например, распределение населения). Карты должны подкреплять тезисы и демонстрировать закономерности, которые иначе были бы упущены.

Рисунок 7.

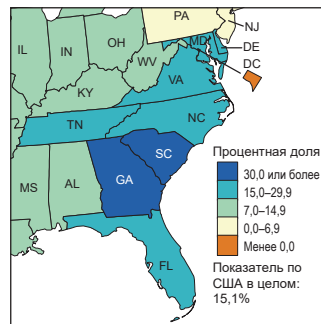
Виды тематических карт, часто используемых в национальной статистике

Картограмма

Области на таких картах окрашены или заштрихованы соразмерно некоторой статистической переменной. Данные обобщаются в рамках заранее определенных территориальных единиц (административных или статистических, например, соответствующих географии переписи).

Оптимальные случаи применения: желательно использовать, когда данные являются стандартизованными (например, коэффициентами), дискретными и равномерно распределенными в рамках четко определенных территориальных единиц.

Рекомендации по дизайну: количество категорий должно быть ограниченным (3–7).



Градуированная символьная карта

Такие карты содержат символы разных размеров для отображения относительного количественного значения. Этот вид иллюстрации используется для координатных данных и данных о местоположении.

Оптимальные случаи применения: лучше использовать, когда вариативность и диапазон данных достаточно велики. Назначение такой карты — показать относительные величины явлений в конкретных местоположениях. Символьную карту также удобно использовать для отображения результатов подсчетов.

Рекомендации по дизайну: не следует использовать для стандартизованных данных, таких как коэффициенты и процентные соотношения.

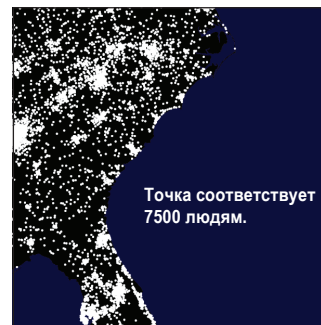


Карта густоты растровых точек

С помощью точек указывается наличие особенности или вхождения; такая карта служит для отображения пространственной закономерности и относительной плотности. Каждая точка может соответствовать одному или нескольким вхождениям.

Оптимальные случаи применения: рекомендуется использовать для результатов подсчетов; возможно отображение нескольких наборов данных (с использованием разных символов или цветов).

Рекомендации по дизайну: для размещения точек на карте требуются дополнительные инструменты (например, средства геокодирования); учитывая трудности восприятия, следует тщательно выбирать характеристики оформления (размер точек, значение, расположение).



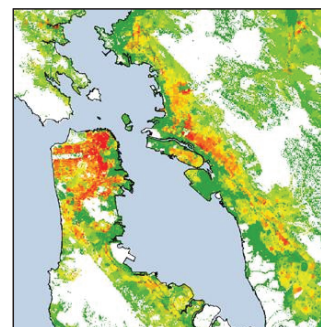
Дазиметрическая карта

Тематическая карта с символами для пространственной классификации подсчетных данных; используется, как альтернатива картограммам.

Оптимальные случаи применения: дазиметрические карты лучше всего использовать в случаях, когда оказывается неверным предположение о равномерном распределении (например, населения). Понимание такой карты улучшается, когда пользователю дополнительно предоставляется описание принципа составления карты. Как и в случае с картограммами, данный метод предназначен для отображения стандартизованных данных (например, коэффициентов).

Рекомендации по дизайну: метод отличается повышенной сложностью и отнимает много времени; размеры областей на карте редко соответствуют границам учетных единиц.

Источники: U.S. Census Bureau, CDC (в соответствии с английским акронимом), and USGS (в соответствии с английским акронимом).



Цвет: в общем случае для категориальных (номинальных) данных используется цветовая схема с несколькими оттенками. Для интервальных данных и соотношений используется один цвет (например, синий). В приведенном примере применяется схема с несколькими оттенками, так как имеется естественная нулевая отметка (есть показатели выше и ниже нуля).

Картографические элементы: как правило, стрелка-указатель севера и масштабный отрезок не нужны. В данном контексте аудитории очевидно, что север находится наверху карты, а для интерпретации закономерностей знание масштаба не требуется. Однако если карта справочная, она должна сопровождаться стрелкой-указателем севера и масштабным отрезком.

Перегруженность: не следует засорять карту ненужной информацией. Тематические карты, как правило, должны отображать только одну переменную. Подписи должны быть простыми. Если карта имеет большое число элементов, количество подписей следует ограничивать.

Диаграммы: обычно диаграммы (например, круговые) поверх данных карты размещать не следует. Диаграммы сводят на нет преимущества карты, затрудняя интерпретацию закономерностей. Вместо размещения диаграмм поверх данных карты можно создать несколько карт, каждая из которых отражает одну переменную.

Неверное деление на категории: выбранные точки изменения должны отражать закономерности, важные для автора. И наоборот, нельзя выбирать точки изменения так, чтобы скрыть нежелательную информацию или создать ложное представление о географической вариативности.

ТАБЛИЦЫ

При представлении данных основное внимание нередко уделяют визуализации в виде диаграмм, карт и др. Но, пожалуй, самая распространенная форма представления данных переписи — таблицы данных (как, например, на рисунке 9), и их оформление требует особого внимания.

ПРЕИМУЩЕСТВА

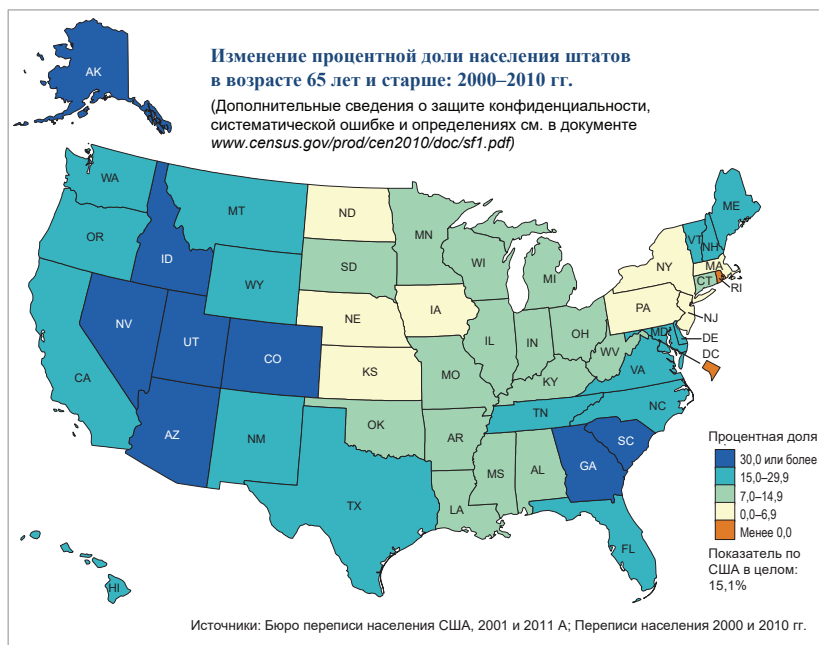
Улучшенная доступность: чтобы не утомлять читателя, на картах и диаграммах отображают лишь ограниченное количество переменных. Таблицы же позволяют отображать множество переменных, уровней географии и периодов времени. В таблицах также может быть более четко указана контекстная информация, например, сведения о погрешности.

Больше возможностей для пользователей: пользователи могут самостоятельно анализировать таблицу данных, при этом никакой элемент данных не получает приоритет по сравнению с остальными. Кроме того, пользователи данных рассчитывают иметь доступ к таблицам данных помимо сводной статистики по результатам их обработки.

Рисунок 8.

Пример иллюстрации: карты

Данный формат карт используется во всех продуктах Бюро переписи населения США, являясь стандартом фирменного стиля ведомства.



Источник: U.S. Census Bureau, 2014.

Эффективность — выше, чем у текстовых описаний: таблицы позволяют отобразить больше данных на единицу площади, чем текст, при этом опытные пользователи легко найдут нужные числа.

Широта применения: таблицы — пожалуй, самая распространенная форма представления данных национальной статистики. Таблицы доступны для понимания практически всем аудиториям.

НЕДОСТАТКИ

Длительность анализа: на интерпретацию данных таблицы у аудитории может уйти немало времени. Чтобы подчеркнуть основные моменты, используется форматирование (полуужирный шрифт, выделение цветом). Также для выделения важных результатов табличную информацию следует дополнять обобщающей статистикой, картами и диаграммами.

Таблицы — плохой адвокат для тезисов автора: они не вызывают мгновенной реакции аудитории. Карты и диаграммы передают сообщение гораздо эффективнее.

РЕКОМЕНДАЦИИ

Примечание. Пример применения рекомендаций приведен на рисунке 9.

Таблица должна иметь **однородный формат**. В частности, между числами должны быть равные промежутки, запятые в столбце должны быть выравнены по вертикали, линии должны иметь одинаковую толщину, все элементы должны быть выравнены относительно друг друга.

Рисунок 9.

Пример иллюстрации: таблицы

Таблицы могут отражать большой объем данных, поэтому при их разработке нужно проявлять особое внимание. Данные примеры демонстрируют возможности использования графических элементов для привлечения интереса аудитории к статистическим данным. Графика сведена к минимуму: например, отсутствуют линии, разделяющие строки, а линии, отделяющие столбцы от заголовков — тонкие.

Население в возрасте 65 лет и старше в разбивке по возрастным категориям: 1900–2050 гг.

(Числа указаны в тысячах. Сведения о защите конфиденциальности, систематической ошибке и определениях см. в документе www.census.gov/prod/cen2010/doc/sf1.pdf)

Источник, год, референтная дата	Общая численность населения	65 лет и старше		85 лет и старше	
		Количество	Процентная доля	Количество	Процентная доля
Перепись					
1900 (1 июня)	75 995	3080	4,1	122	0,2
1910 (15 апреля)	91 972	3 950	4,3	167	0,2
1920 (1 января)	105 711	4 933	4,7	210	0,2
1930 (1 апреля)	122 775	6634	5,4	272	0,2
<Для целей примера таблица приведена частично>					
2000 (1 апреля)	281 422	34 992	12,4	4 240	1,5
2010 (1 апреля)	308 746	40 268	13,0	5493	1,8
Прогноз					
2020 (1 июля)	333 896	55 969	16,8	6693	2,0

<Для целей примера таблица приведена частично>

Примечание. Данные за период с 1900-го по 1950 год не включают Аляску и Гавайи. Источники: 1900 to 1940, and 1960 to 1980, U.S. Bureau of the Census, 1983; 1950, U.S. Bureau of the Census, 1953; 1990, U.S. Bureau of the Census, 1992; 2000, U.S. Census Bureau, 2001; 2010, U.S. Census Bureau, 2011; 2020 to 2050, U.S. Census Bureau, 2012b; 1900 to 2010, decennial census; 2020 to 2050, 2012 National Population Projections, Middle series.

Динамика и процентные доли численности населения в возрасте 65 лет и старше по регионам и штатам: 2000 и 2010 годы

(Сведения о защите конфиденциальности, систематической ошибке и определениях см. в документе www.census.gov/prod/cen2010/doc/sf1.pdf)

Регион и штат	65 лет и старше		Изменение за период 2000–2010 гг.	
	2000	2010	Количество	Процентная доля
	Соединенные Штаты	34 991 753	40 267 984	5 276 231
Северо-восток	7 372 282	7 804 833	432 551	5,9
Коннектикут	470 183	506 559	36 376	7,7
Мэн	183 402	211 080	27 678	15,1
Массачусетс	860 162	902 724	42 562	4,9
Нью-Гэмпшир	147 970	178 268	30 298	20,5
Нью-Джерси	1 113 136	1 185 993	72 857	6,5
Нью-Йорк	2 448 352	2 617 943	169 591	6,9
Пенсильвания	1 919 165	1 959 307	40 142	2,1
Род-Айленд	152 402	151 881	-521	-0,3
Вермонт	77 510	91 078	13 568	17,5

<Для целей примера таблица приведена частично>

Источники: U.S. Census Bureau, 2001, Census 2000 Summary File 1, Table P12, Washington, DC, available at <http://factfinder2.census.gov/>, accessed on February 20, 2012; U.S. Census Bureau, 2011, 2010 Census Summary File 1, Table PCT12, Washington, DC, available at <http://factfinder2.census.gov/>, accessed on February 20, 2012.

Источник: U.S. Census Bureau, 2014.

ЧЕГО СЛЕДУЕТ ИЗБЕГАТЬ

Отвлекающего форматирования: число линий в таблице следует ограничивать, а все необходимые линии должны быть тонкими. Заголовки столбцов должны быть отделены от данных; число выделенных элементов должно быть минимальным.

Перегруженность информацией: при выборе данных для включения в таблицу следует проявлять осмотрительность. Как и любой визуальный элемент, таблица должна поддерживать тезисы автора. Не нужно включать в таблицу все переменные — полные таблицы должны быть доступны для загрузки на сайте NSO или прикреплены к основному документу в качестве приложений.

Бессистемность: каждая таблица должна отображать единый набор сведений в разбивке по годам, географическим регионам и/или по теме. В случае разбивки данных по регионам не следует исключать равные единицы (например, провинции или округа) без объяснения.

ЛИТЕРАТУРА

Cairo, A., *The Functional Art: an Introduction to Information Graphics and Visualization*, New Riders, 2013.

Centers for Disease Control and Prevention (CDC), *Types of Thematic Maps*, available at www.cdc.gov/dhdsdp/maps/gisx/resources/maps4.html, accessed on February 22, 2017.

Few, S., *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, 2nd ed., Analytics Press, 2012.

Knaflig, C.N., *Storytelling with Data: a Data Visualization Guide for Business Professionals*, Wiley, 2015.

Tufte, E.R., *The Visual Display of Quantitative Information*, 1st ed., Graphics Press, 1983.

United Nations, *Fundamental Principles of Official Statistics*, October 28, 2013, available at <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>, accessed on February 22, 2017.

U.S. Census Bureau, "How Do We Know? A Snapshot of Our Nation's Veterans," 2012, available at www.census.gov/library/visualizations/2012/comm/veterans.html, accessed on February 22, 2017.

_____, P23-212, *65+ in the United States: 2010*, U.S. Government Printing Office, Washington, DC, 2014, available at www.census.gov/library/publications/2014/demo/p23-212.html, accessed on February 22, 2017.

_____, Current Population Reports, P60-252, *Income and Poverty in the United States: 2014*, U.S. Government Printing Office, Washington, DC, 2015, available at www.census.gov/library/publications/2015/demo/p60-252.html, accessed on February 22, 2017.

U.S. Geological Survey (USGS), "Dasymetric Mapping: An Alternative Approach to Visually and Statistically Enhancing Population Density", 2016, available at <http://geography.wr.usgs.gov/science/dasymetric/index.htm>, accessed on February 22, 2017.