

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ
ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

ПРИДАТЬ ЗНАЧИМОСТЬ СТАТИСТИЧЕСКИМ ДАННЫМ



Часть 2

Руководство по представлению
статистических данных



**ОРГАНИЗАЦИЯ
ОБЪЕДИНЕННЫХ НАЦИЙ**



ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ
ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

Придать значимость статистическим данным

Часть 2:

Руководство по представлению статистических
данных



ОРГАНИЗАЦИЯ ОБЪЕДИНЕННЫХ НАЦИЙ
Женева, 2009 год

ПРИМЕЧАНИЕ

Принятые обозначения и представленные материалы ни в коей мере не выражают мнения Секретариата Объединенных Наций в отношении правового статуса каких-либо стран, территорий, городов или районов или их руководителей или касающиеся определения пограничных полос и границ.

ECE/CES/STAT/NONE/2009/3

СОДЕРЖАНИЕ

.....

Введение	v
1. Донести свою мысль до читателя	1
2. Визуализация статистических данных	8
3. Таблицы	14
4. Графики	20
5. Карты	36
6. Передовые методы визуализации	49
7. Вопросы доступности	55
8. Справочная и дополнительная литература	60

Введение

Руководства "Придать значимость статистическим данным" были подготовлены на рабочих сессиях Европейской экономической комиссии Организации Объединенных Наций (ЕЭК ООН) по передаче и распространению статистических данных¹ в рамках программы работы Конференции европейских статистиков².

Данные руководства призваны стать практическим инструментом оказания помощи руководителям, статистикам и сотрудникам, отвечающим за связи со средствами массовой информации в статистических организациях, особенно в тех организациях, которые находятся в процессе разработки своих стратегий распространения. Данные руководства содержат рекомендации по использованию текста, таблиц, графиков, карт и других инструментов, призванных сделать сухую статистику живой и актуальной для нестатистиков. Они содержат предложения, руководящие принципы и примеры, но не жесткие правила или шаблоны.

Эффективный информационный бюллетень опирается на определенную комбинацию текста, таблиц и графиков, призванную максимально повысить его способность служить передаточным механизмом различных типов информации. Основное место в публикации "Придать значимость статистическим данным - Часть 1: Как написать историю в цифрах?" (вышедшей в свет в 2006 году) было уделено использованию эффективных методов написания статистических статей. Публикация "Придать значимость статистическим данным - Часть 2: Руководство по представлению статистики" призвана оказать помощь читателям в нахождении наиболее эффективного способа донесения своих мыслей до неспециалистов благодаря использованию наиболее подходящего набора инструментов и методов из имеющегося в настоящее время громадного арсенала коммуникационных технологий.

В настоящем руководстве признается существование многочисленных практических и культурных различий между статистическими организациями, а также возможность использования различных подходов в зависимости от страны.

Участие в подготовке данного руководства приняла группа экспертов в области передачи и распространения статистических данных. Речь идет о следующих экспертах (перечисленных в порядке латинского алфавита):

- Петтери Баер, ЕЭК ООН
- Коллин Блессинг, Управление по энергетической информации Соединенных Штатов
- Эйлин Каппони, Организация экономического сотрудничества и развития
- Жером Кюкье, Организация экономического сотрудничества и развития
- Керри Дафф, Австралийское бюро статистики
- Джон Фландерс, Статистическое управление Канады
- Коллин Фланнери, Бюро переписей Соединенных Штатов

¹ Информация о рабочих сессиях по передаче и распространению статистических данных ЕЭК ООН размещена на вебсайте ЕЭК ООН по следующему адресу: <http://www.unece.org/stats/archive/04.05.e.htm>.

² Информация о Конференции европейских статистиков размещена на вебсайте ЕЭК ООН по следующему адресу: <http://www.unece.org/stats/archive/act.00.e.htm>.

- Джессика Гарднер, ЕЭК ООН
- Мартин Гренье, Статистическое управление Канады
- Армин Гроссенбахер, Швейцарское федеральное статистическое управление
- Дэвид Мардер, Управление национальной статистики Соединенного Королевства
- Кеннет Мейер, Бюро переписей Соединенных Штатов
- Терри Миттон, Организация экономического сотрудничества и развития
- Эрик Сент-Джон, Статистическое управление Канады
- Томас Шульц, Швейцарское федеральное статистическое управление
- Анн-Кристин Вандерс, ЕЭК ООН

Весьма ценный вклад в его подготовку также внесли Мартэн Ляшанс (Статистическое управление Канады), Эндрю Мэйр (Австралийское бюро статистики), Алан Смит (Управление национальной статистики Соединенного Королевства) и Стивен Вейл (ЕЭК ООН).

1. Донести свою мысль до читателя

1.1 Письменное слово

Информационные бюллетени зачастую являются инструментом, с помощью которого статистическая организация сообщает ключевые выводы своих статистических и аналитических программ целевой аудитории, которой, вероятнее всего, является широкая общественность. Текст является основным инструментом пояснения выводов, выделения тенденций и предоставления контекстуальной информации.

В настоящей главе будут изложены многочисленные рекомендации по подготовке "эффективных" информационных бюллетеней или других документов, таких как отчеты или аналитические статьи.

Что делает информационный бюллетень, отчет или аналитическую статью эффективными? Возможно, самый лучший ответ на этот вопрос дает одна из рекомендаций первой публикации "Придать значимость статистическим данным – Часть 1: Как написать статью о цифрах", которая представляет собой исходный набор рекомендаций по донесению своей мысли до читателя. Эффективным является информационный бюллетень, который:

- рассказывает историю о данных;
- представляет интерес для общественности и отвечает на вопрос "Почему аудитория должна захотеть прочитать об этом";
- быстро овладевает вниманием читателя с помощью заголовка или изображения;
- является понятным, интересным и часто увлекательным;
- побуждает других, включая средства массовой информации, использовать статистические данные надлежащим образом для придания дополнительного эффекта тем идеям, которые они сообщают.

Ниже приводятся некоторые рекомендации, призванные помочь вам в начале работы над вашим текстом.

1.2 Целевая аудитория: ваше первое решение

Первым важным решением, которое вы должны принять, является определение аудитории, т.е. ответа на вопрос "Для кого вы пишете?". Проще говоря, аудитория "командует парадом". Это означает, что вы должны давать аудитории то, что она желает получить от вас. Вы должны уметь слушать вашу аудиторию для нахождения и использования правильного стиля, языка и визуального и графического оформления, которые привлекут ее внимание.

Определение аудитории осложняется в настоящее время существованием Интернета. Большинство статистических организаций наделены мандатом на информирование широкой общественности, которую составляют не являющиеся специалистами, но достаточно хорошо образованные обычные люди. В эпоху печатных информационных бюллетеней в качестве основной целевой аудитории рассматривались средства массовой информации, к которым организации обращались с целью доведения своих ключевых выводов до внимания общественности.

Однако сегодня статистические организации сформировали достаточно большой круг своих читателей, с которыми они общаются напрямую, через свои вебсайты, электронную почту и другие формы Интернет-распространения. Это означает, что они общаются с большим числом аудиторий одновременно: общественностью, пользователями данных, банкирами, финансовыми аналитиками, университетскими профессорами, учащимися и т.д., каждая из которых имеет свои собственные потребности в данных.

Мир коммуникаций постоянно эволюционирует. Успешные коммерческие средства массовой информации это знают и постоянно следят, зачастую в режиме реального времени, за тем, какие из их материалов пользуются наибольшим вниманием. Исходя из этого, они направляют свои ресурсы на создание более богатого содержания путем использования таких инструментов, как видео, дополнительные фото или дополнительный анализ, с целью поощрения более активного взаимодействия с каждой из аудиторий.

В любом случае вывод заключается в том, что, перед тем как направить ценные ресурсы на использование того или иного метода коммуникации – нового или традиционного, – необходимо в первую очередь решить, что представляет собой аудитория или заинтересованные стороны, что они хотят от вас и каким образом они хотят это получить.

Если вы хотите обратиться одновременно к нескольким аудиториям, вы должны определить надлежащий метод общения с каждой из них путем доведения до их сведения вашей мысли через соответствующие каналы и с использованием соответствующих методов коммуникации.

Однако во многих случаях нехватка времени и ресурсов будут означать невозможность одновременного охвата всех целевых аудиторий. У вас есть выбор: вы можете установить приоритеты или же, если вы хотите охватить как можно более широкую аудиторию, вы можете максимально ясно определить общую платформу.

Именно так и поступают многие статистические организации. Они ориентируются на широкую общественность, но в то же время предпринимают скоординированные усилия по информированию данной аудитории с использованием журналистов в качестве "проводника". Целевой аудиторией является общественность, однако журналисты являются инструментом общения с этой аудиторией. Эксперты и специалисты также могут извлечь выгоды от данного подхода. ачастую простые и ясные методы, используемые для информирования широкой аудитории, высоко ценятся даже самой высокоспециализированной аудиторией.

1.3 Понимание контекста, в котором вы передаете данные

Передача статистических данных происходит не обособленно. Следовательно, важно, чтобы вы понимали контекст, в котором вы осуществляете их передачу. Механизмы взаимодействия различных аудиторий со средствами массовой информации постоянно меняются. Существуют также явные различия между поколениями с точки зрения технических возможностей и понимания статистики.

При планировании мероприятий по передаче статистических материалов вы должны учитывать четыре конкретные тенденции в использовании онлайн-овых средств распространения информации, с которыми сопряжены как возможности, так и риски:

1. Всемирная паутина во все большей степени становится средой для развлечений. Любые сообщения, которые не привлекают к себе интереса, рискуют остаться без внимания со стороны молодой аудитории.
2. Общество сформировало в отношении потребления информации "снэк-культуру". Аудитории во все большей степени необходимы малые фрагменты информации, которые не требуют много времени для ознакомления;
3. Аудитории, использующие Интернет, стремятся к "достаточности": они находят примерно необходимые им фрагменты информации и останавливаются на этом, не занимаясь дальнейшим поиском более релевантного фрагмента информации.
4. При изучении различных аудиторий и стилей представления следите за тем, чтобы не исключить из охвата важные аудитории в результате повышения привлекательности или упрощения для использования ваших инструментов статистической коммуникации.

Таким образом, возникает вопрос о том, что вы можете сделать для наиболее эффективного использования Интернета. Вы должны использовать наиболее подходящие инструменты языка, структуры и представления с целью донесения ваших мыслей до аудитории. Нижеследующие разделы содержат иллюстрации возможных подходов.

1.4 Стил ь изложения: расскажите историю

В первую очередь найдите историю

Для того чтобы данные были значимыми для широкой аудитории, необходимо найти смысл в цифрах. Слово "история" зачастую пугает людей в статистическом/научном мире, поскольку содержит оттенок вымысла или приукрашивания, что может привести к ошибочному толкованию данных. Данное опасение может быть оправданным, если аналитики не относятся к данным с должной тщательностью и вниманием.

Однако, в свою очередь, отказ от истории может привести к гораздо худшим последствиям. Люди зачастую не доверяют статистике и считают, что их вводят в заблуждение, поскольку они не могут понять данные. Это происходит потому, что мы, производители данных, не можем сделать их значимыми и объяснить их в терминах, понятных простым людям. В отсутствие сюжетной линии бюллетень становится простым изложением цифр.

Статистическая история должна опираться на достаточное знание данных и анализируемого явления. В ином случае она может быть интересной, но на практике не соответствовать действительности. При написании статистической истории мы должны также учитывать Основные принципы официальной статистики³:

1. беспристрастность;
2. профессионализм;
3. метаданные;
4. комментарии при ошибочной интерпретации;
5. различные источники;
6. конфиденциальность;
7. транспарентность;
8. национальная координация;
9. международные стандарты;
10. международное сотрудничество.

Чрезвычайно важно, чтобы статистические организации оставались беспристрастными и обеспечивали конфиденциальность в отношении респондентов и малых подгрупп населения.

В вашем тексте наиболее важные и значимые выводы должны приводиться в контексте кратко- и долгосрочных тенденций. В нем должны анализироваться взаимные связи, причины и следствия в той степени, в которой они могут быть подкреплены фактами. Он должен демонстрировать читателям важность новейшей информации.

Пишите в журналистском стиле

Используйте стиль написания, принятый журналистами: "перевернутая пирамида". Сначала изложите наиболее важные факты, а затем второстепенные вопросы в порядке снижения их значимости. Читатели быстро теряют интерес, вследствие чего наиболее важная информация должна излагаться в начале текста.

Избегайте начинать ваш текст с изложения методики и заканчивать его выводами. Вы можете изложить ключевые положения методологии в примечаниях для читателей, поскольку чем менее сложным является пояснение методики, тем лучше. Вывод должен излагаться во вводном или первом абзаце.

Вводная часть является важнейшим элементом вашего текста. Она должна рассказывать историю о данных. Она должна описывать кратко, ясно и просто сюжетную линию и помещать ее в контекст. Она должна быть посвящена одной мысли или теме и содержать минимум данных.

³ Данные принципы были приняты Статистической комиссией Организации Объединенных Наций в 1994 году. Они подробно описаны на вебсайте ЕЭК ООН по адресу: <http://www.unece.org/stats/archive/docs.fp.e.htm>.

ХОРОШИЙ ПРИМЕР вводного абзаца:

Чистая прибыль нефинансовых компаний в Нидерландах составила во втором квартале 2008 года 19 млрд. евро. Это является наиболее низким уровнем за три года. Прибыль снизилась на 11% по сравнению со вторым кварталом 2007 года. Падение чистой прибыли является результатом двух основных факторов: увеличения процентных выплат - компании платят больше чистых выплат по процентам - и снижением прибыли иностранных филиалов.

Источник: Статистическое управление Нидерландов

Не перегружайте читателя избыточным количеством цифр в тексте, в котором следует использовать только ключевые округленные цифры. Менее важные цифры должны включаться в сопроводительные таблицы. Используйте текст для изложения анализа, трендов и контекста, а не для повторения величин, приводимых в таблицах.

Уделяйте внимание структуре

Структурируйте ваш текст таким образом, чтобы каждый компонент обладал своим собственным значением, а также вносил вклад в общую историю, которую вы рассказываете. Подзаголовки являются одним из эффективных инструментов улучшения организации информационного бюллетеня. Они обеспечивают разбивку на удобные и значимые разделы.

Краткий подзаголовок служит резюме основного вывода подраздела. Он может являться более привлекательным и понятным, если содержит глагол.

ХОРОШИЕ ПРИМЕРЫ подзаголовков:

"Уровни запасов несколько снижаются"

"Рост в области энергетических продуктов ведет к увеличению импорта"

В случае материалов, распространяемых через Интернет, каждый подраздел должен нести самостоятельную смысловую нагрузку, что означает необходимость разъяснения терминов и цитирования источников. Поисковые программы, как правило, ориентируют пользователей на гиперссылки в рамках вебсайтов, а не на домашнюю страницу или другие шлюзы, которые вы создаете для направления посетителей к искомому материалу.

Ваши мысли должны также излагаться многослойно для удовлетворения различных информационных потребностей вашей аудитории. Начинайте каждый подраздел с тематического предложения, которое четко излагает основной вывод подраздела. В последующих абзацах вы можете развить ваш вывод.

Данная структура обеспечит четкий путь к более подробным данным, анализу и технической информации для аудитории, которая проявляет интерес к таким подробностям. Кроме того, в электронных публикациях вы можете всегда использовать гиперссылки для направления читателей к более сложному анализу.

1.5 Язык: четкий, лаконичный и простой

Простой язык является залогом любого успешного сообщения. Старый девиз "Будь прост и краток" сегодня актуален более, чем когда-либо. Это касается не только вашего текста. Данный принцип также следует использовать в отношении таблиц, средств визуализации и графиков, в случае которых наличие многочисленных возможностей показаться умным создает искушение избыточной детализации.

Стремитесь выражаться обыденным языком. Слишком часто мы используем материалы, написанные официальным языком, который является неэффективным и во многих случаях непонятным или неясным.

Использование обыденного языка не означает использование снисходительного тона, банальностей или чрезмерного упрощения, а также отхода от правил грамматики. Обыденный язык обеспечивает четкое и лаконичное изложение мысли. Он используется с учетом потребностей читателя и с выбором правильной интонации.

Обыденный язык быстрее читается и доносит вашу мысль до читателя чаще, проще и понятным образом.

Помните: четкие и простые сообщения **НЕ** означают "оболванивание".

Некоторые рекомендации по четкому изложению:

- используйте короткие предложения;
- стремитесь к соблюдению принципа "одна мысль – одно предложение";
- разбивайте длинные предложения;
- начинайте каждый абзац с наиболее важной мысли;
- стремитесь к краткости абзацев;
- сохраняйте свежесть языка.

В случае глаголов избегайте пассивного залога; используйте активный залог. Пассивные глаголы могут вносить путаницу и вести к усложнению структуры и утрате ясности.

ПЛОХОЙ ПРИМЕР:

"Рост безработицы был вызван экономическим кризисом".

ХОРОШИЙ ПРИМЕР:

"Экономический кризис привел к росту безработицы".

По мере возможности избегайте жаргонизмов. Жаргон является языком, который обладает конкретным значением для определенной группы лиц. Он практически никогда не подходит для использования в отношении широкой аудитории. Используйте самые простые слова, которые отражают вашу мысль и обеспечивают ее понимание читателем. Однако если ваша аудитория состоит полностью из специалистов в конкретной области, жаргон может стать наиболее подходящим языком.

1.6 Оценка воздействия: анализ средств массовой информации

Насколько правильно и разумно ваша аудитория интерпретировала ваши данные? Может ли она использовать анализ для вынесения обоснованных суждений с целью принятия мер по вопросам экономической, экологической и социальной политики?

Вы не узнаете этого, если не будете анализировать то, каким образом ваша аудитория использует ваш информационный бюллетень. Статистическим организациям было бы целесообразно осуществлять мониторинг того, каким образом средства массовой информации и другие аудитории используют их информацию. К числу таких методов относятся:

- создание программы "Голос пользователя" или любых других механизмов сбора откликов пользователей;
- измерение реакции общественности на основе групп пользователей или фокус-групп, состоящих из ключевых пользователей;
- проведение опросов пользователей и консультаций с влиятельными клиентами на разовой основе;
- использование служб изучения рыночной конъюнктуры, если вы нуждаетесь в глубоком понимании того или иного чувствительного вопроса или аудитории, доступ к которой затруднен;
- мониторинг Интернет-трафика для определения того, какие данные и материалы пользуются наиболее высоким спросом;
- анализ поисковых запросов по ключевым словам является мощной, хотя и простой методикой выяснения того, какую информацию ищет ваша аудитория.

Вы можете работать напрямую со средствами массовой информации для обеспечения точного и корректного отражения ими ваших мыслей. Средства массовой информации являются самым простым, дешевым и эффективным инструментом доведения ваших мыслей до сведения широкой аудитории.

К сожалению, некоторые средства массовой информации имеют свои собственные интересы, что может привести к ошибочному использованию, интерпретации и искажению ваших данных. Чрезвычайно важно осуществлять мониторинг за распространением ваших сообщений через средства массовой информации, а также за тоном и качеством охвата. В случае любого явно ошибочного использования вы должны оперативно, последовательно и настойчиво реагировать на такие ситуации.

Теперь, когда мы обсудили основные элементы текста, обратим наше внимание на визуальные элементы, которые могут помочь нам в разъяснении данных.

2. Визуализация статистических данных

2.1 Почему картинка по-прежнему стоит тысячи слов?

Всем хорошо известно, что "картинка стоит тысячи слов". Одним из наилучших способов обеспечения понимания данных является визуализация цифр в качестве изображения. Картинка облегчает выявление или разъяснение трендов, которые в ином случае останутся скрытыми для читателя.

Вы можете визуализировать ваши данные различными способами, начиная с простых гистограмм и кончая более сложными диаграммами рассеяния, тематическими картами и анимированными демографическими пирамидами. В этой области не существует дефицита технической поддержки: по методам визуализации данных были написаны целые книги; существует масса вебсайтов, посвященных этой теме, и широкое разнообразие программного обеспечения и загружаемых программ для решения различных задач.

В настоящей главе приводятся краткий исторический обзор по теме визуализации данных и руководящие принципы разработки эффективных графических презентаций данных.

2.2 Визуализация является неотъемлемой частью статистического бизнес-процесса

Эффективная презентация данных должна стать неотъемлемой частью вашего процесса статистического производства. Поскольку визуализация данных является столь важным элементом распространения информации о статистических трендах и взаимосвязях, она должна являться одним из текущих, а не оставляемых на потом видов деятельности. Визуализация включена в стадию "Распространение" Типовой модели статистического бизнес-процесса⁴, которая разрабатывается Руководящей группой по статистическим метаданным ЕЭК ООН/Евростата/ОЭСР.

Визуальные презентации являются для читателей легко понятными. Практически каждый встречается с графиками и картами в газетах, на телевидении, в Интернете и в книгах. Статистика, представляемая в виде графика или карты, является намного более понятной, чем длинный перечень цифр, при условии, естественно, что визуальные презентации оформлены надлежащим образом.

Презентации должны иллюстрировать тренды и взаимосвязи простым и легко понятным образом. Они являются эффективным способом извлечения информации из базы данных и усвоения ее читателем.

Однако необходимо проявлять осторожность. Некачественная визуализация статистической информации может вводить в заблуждение. Существуют многочисленные способы представления вводящей в заблуждение информации, будь то преднамеренно или, как это происходит в большинстве случаев, непреднамеренно. Необходимо найти баланс между дизайном и функциональностью. Сложная визуализация зачастую препятствует

⁴ <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

коммуникации. Поскольку интерпретирующие графики могут быть сложными, не заставляйте ваших читателей "докапываться" до вашей мысли.

Неправильное понимание и интерпретация могут быть также обусловлены различными культурными традициями. Так, например, цвета могут иметь неодинаковое символическое значение в различных частях мира.

2.3 Историческое влияние одного шотландца на визуализацию данных

История статистических графиков в том виде, в котором мы их понимаем сегодня, насчитывает менее 300 лет. Хотя древние римляне и арабы умели хорошо обращаться с цифрами, они не пользовались гистограммами или кривыми для визуального представления величин.

Рене Декарт (1596-1650 годы), французский энциклопедист и философ, впервые начал использовать прямоугольную систему координат с целью визуализации своих наблюдений. Однако первым начал готовить графические представления статистических данных в том виде, в каком они нам известны сегодня, шотландский инженер-экономист Уильям Плэйфэйр (1759-1823 годы).

Плэйфэйр опубликовал две книги о развитии мировой торговли на рубеже XVIII-XIX веков (1796-1801 годы), в которых он использовал так называемый "графический метод". Он был искусным пропагандистом данного метода, обосновывая преимущества визуальных презентаций с помощью следующих аргументов:

1. Графические презентации помогают упростить сложные взаимосвязи, которые с трудом поддаются наблюдению:

"Выражается надежда, что с помощью этих графиков информация будет восприниматься без усталости и затруднений, связанных с изучением составляющих ее компонентов".

2. Владельцы предприятий, политики и директивные органы нуждаются в кратком визуальном изложении сути информации, поскольку у них нет времени для ее подробного изучения:

"Люди высокого положения или активные бизнесмены могут уделять внимание только общим выводам [...]".

3. Таблицы, особенно объемные таблицы, могут выглядеть скучными и трудными для понимания. Графические презентации облегчают жизнь пользователю:

"[...] таблица цифр позволяет извлечь в течение пяти минут столько же информации, сколько целые дни ее надежного запоминания.

4. Графики притягивают к себе глаз. Плэйфэйр обратил внимание на это, подчеркнув полезность его метода для понимания связей и количеств.
5. Графический метод привлекает и стимулирует ум, помогая ему не только в восприятии, но и в понимании структур и взаимосвязей.

Хотя со времен Плэйфэйра в методах визуализации произошла революция, в этой области по-прежнему существуют широкие возможности для совершенствования и новых разработок (см. главу 6). Однако важно избегать

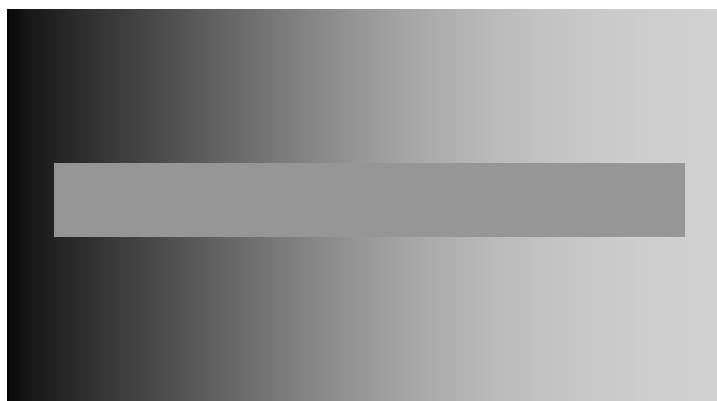
избыточно крикливых презентаций, искажающих содержание. **Напомним, что технология является лишь служанкой.** Не следует добавлять бесполезные замечания и непонятные элементы лишь в силу наличия таких возможностей. Стремитесь к тому, чтобы идея оставалась ясной для читателя.

К числу новых светил в области анализа графических презентаций и человеческого восприятия в настоящее время относятся Жак Бертен, Уильям Кливленд, Ян-Эрик Кристьянсен, Веза Куусева, Ханс Рослинг, Эдвард Тафт и Ховард Уэйнер. Несмотря на разработки этих современных ученых фундаментальные характеристики статистических графиков продолжают опираться на разработки Уильяма Плэйфэйра.

2.4 Основные сведения о человеческом восприятии

Наша способность быстро и легко схватывать визуальную информацию опирается на способность мозга воспринимать упорядоченное и неупорядоченное. В значительной степени механизмы данной способности расположены на подсознательном уровне. Сопоставление происходит еще до того, как мы начнем думать о нем.

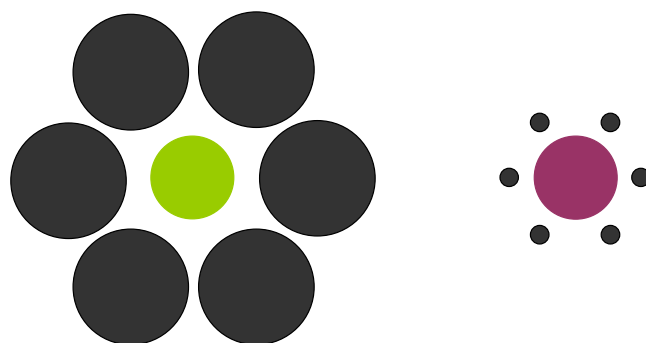
Посмотрим на нижеприведенный рисунок. Какой конец горизонтальной полоски, изображенной в центре, является более темным?



Источник: Helmholtz, H. (1821-1894), "Simultaneous Contrast Illusion", In Wikipedia, The Free Encyclopedia, retrieved in July 2009 from http://en.wikipedia.org/wiki/Optical_illusion.

Полоска имеет одинаковое затемнение на обоих концах. Различия в тоновом цвете искажают восприятие. Вы легко увидите это, закрыв фон полоской бумаги.

Сейчас рассмотрим два следующих изображения. Какой из кругов является бо⁵льшим: тот, который помещен в центр диаграммы слева, или же тот, который помещен в центр диаграммы справа?



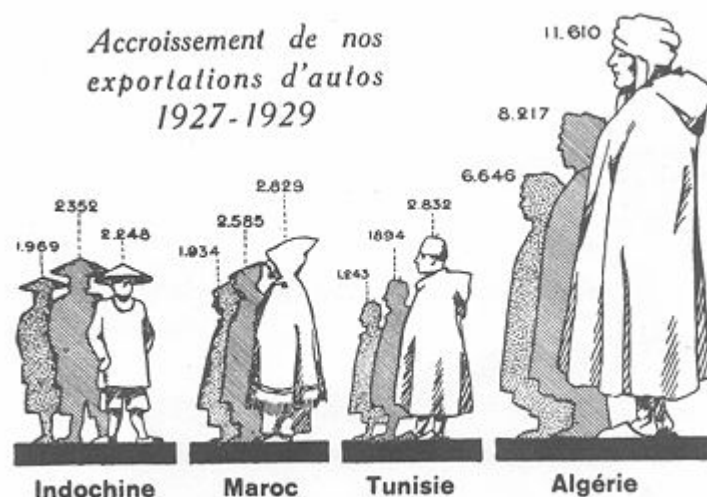
Источник: Ebbinghaus, H. (1850-1909), "The Ebbinghaus Illusion", In Wikipedia, The Free Encyclopedia, retrieved in July 2009 from http://en.wikipedia.org/wiki/Hermann_Ebbinghaus.

Круги в центре каждой диаграммы имеют одинаковый размер. Для проверки вы можете измерить их.

Эти два примера иллюстрируют, каким образом человеческий ум автоматически воспринимает объекты в контексте их окружения. Именно контраст с окружающим фоном обуславливает наши выводы в отношении цвета полоски и размера двух кругов. Исключите фон и вы скорее придете к верному выводу.

Смысл этих упражнений для статистиков заключается в следующем: вы должны проявлять тщательность при подготовке визуальных презентаций статистических наблюдений. Контекст, в котором представляются выводы, может исказить их восприятие пользователем.

Рассмотрим третий пример. Нижеприводимый график призван иллюстрировать рост экспорта автомобилей в Индокитае, Марокко, Тунисе и Алжире в период 1927-1929 годов. Насколько хорошо размер мужских фигур отражает реальное изменение в данных?



Источник: Satet, R. (1932), *Les Graphiques*, Paris. Quoted in Tufte (2001).

При зрительном восприятии данной картинке, наш ум сопоставляет относительный размер каждого объекта. На данном графике данные об экспорте автомобилей в период 1927-1929 годов иллюстрируются высотой и объемом четырех наборов рисунков. Относительные величины искажаются следующими двумя факторами:

1. Используется ошибочный масштаб. Высота фигур в случае Алжира должна быть намного большей, о, что в четыре раза больше, должно быть изображено в четыре раза более объемным или высоким.
2. При сопоставлении фигуры за последний год с фигурами предыдущих лет разница в величинах между годами с трудом поддается определению. Фигура на переднем плане выглядит намного большей по сравнению с другими фигурами. Межгодовые изменения должна иллюстрировать только относительная высота фигур.

2.5 Восприятие также опирается на опыт

Опыт также играет свою роль в восприятии графиков. Необходимо знать свою аудиторию и ее способности, опыт и возможные различия. Не думайте, что они знают то, что знаете вы, либо о статистике, либо о рассматриваемом вопросе.

Статистики должны сопоставлять свой собственный опыт в области выявления трендов в цифрах с возможной неопытностью своих читателей. В качестве профессионалов анализа статистических цифр статистики, по всей видимости, лучше, чем среднестатистический человек, поймут основополагающую идею. Статистики используют визуализацию для пояснения своих выводов, поскольку широкой аудитории необходима дополнительная помощь в понимании статистической информации. Визуальная презентация данных должна облегчать выявление и понимание основных выводов.

2.6 Правильно используйте инструменты

Помощь в визуализации такой информации способны оказать многочисленные технологические инструменты. Сделайте так, чтобы в центре вашего внимания оказались суть и идея графика, а не методология, дизайн или технология графического представления.

Базовое программное обеспечение табулирования предлагает на выбор более 70 стандартных типов графиков, а также дает возможность создавать специальные графики. Такой широкий диапазон вариантов зачастую подталкивает поставщиков информации к использованию всевозможных различных способов визуализации данных просто в силу наличия таких возможностей. Однако результатом этого может во многих случаях стать чрезмерное усложнение или искажение статистической информации.

Этот риск постоянно повышается вследствие непрерывно возрастающего показного великолепия и разнообразия новых инструментов. При использовании этих инструментов вам не следует забывать о том, что технология является лишь средством достижения определенной цели. Мысль, содержащаяся в статистической информации, является намного более важной, чем просто демонстрация того, каким количеством новомодных средств вы умеете пользоваться.

2.7 Памятка по разработке эффективных визуальных презентаций данных

При подготовке визуальных презентаций вы должны учитывать:

- **Целевую группу:** для различных аудиторий (например, деловых или научных кругов и общественности) могут требоваться различные формы презентации.
- **Роль графики в общей презентации:** анализ большого изображения или сосредоточение внимания на ключевых моментах могут потребовать различных типов графического представления.
- **Каким образом и где излагать идею:** длинный подробный анализ или быстрая слайдовая презентация.
- **Контекстуальные вопросы, которые могут исказить понимание:** опытный пользователь данных или новичок.
- **Что является более эффективным решением - текстовой анализ или таблица данных.**
- **Соображения, связанные с доступностью:**
 - Предоставление текстовых альтернатив в отношении нетекстовых элементов, таких как графики и изображения.
 - Не полагайтесь только на цвет. Будет ли по-прежнему понятна презентация, если вы откажетесь от цвета? Обеспечивает ли комбинация цветов достаточный контраст? Не создают ли используемые цвета проблемы для дальтоников (красный/зеленый)?
 - Обеспечьте, чтобы времязависимое содержание могло контролироваться пользователем (например, пометка на паузу анимированных графиков).
- **Непротиворечивость между графическими изображениями данных:** обеспечение того, чтобы элементы визуализации разрабатывались последовательно и с использованием общих правил, по мере возможности (например, голубой цвет для изображения воды на карте).
- **Размер, продолжительность и сложность:** насколько ваша презентация является легкой для понимания? Не является ли она слишком большой для просмотра в ходе одной сессии?
- **Возможность ошибочного толкования:** протестируйте вашу презентацию на коллегам, друзьях или лицах из вашей целевой группы для выяснения того, понимают ли они изложенную в презентации идею.

3. Таблицы

3.1 Почему таблицы имеют важное значение

Хорошие таблицы являются неотъемлемой частью вашего материала, будь то информационный бюллетень, аналитическая статья или исследовательский документ. Эффективное использование таблиц помогает свести к минимуму количество цифровых данных в вашем тексте. Оно также устраняет необходимость обсуждения менее значительных переменных, которые не имеют важного значения для вашего сюжета.

В своей книге, посвященной написанию материалов о цифрах, Миллер (2004 год) изложил следующие руководящие принципы по проектированию эффективных таблиц:

- Облегчите для вашей аудитории нахождение и понимание цифр в рамках ваших таблиц.
- Разрабатывайте формат и разметку ваших таблиц ясным и не разрушающим структуру образом, с тем чтобы внимание читателя было сосредоточено на основных выводах, которые подкрепляются вашими данными, а не на структуре таблицы.

В настоящей главе мы опишем принципы подготовки эффективной таблицы.

3.2 Два типа таблиц

Интерес для вас должны представлять следующие два типа таблиц. В первую очередь, речь идет о малых таблицах, называемых презентационными (или демонстрационными) таблицами. Они могут использоваться для привлечения внимания читателя к основным цифрам в информационных бюллетенях, на вебстраницах или в аналитических публикациях.

Ко второму типу относятся крупные таблицы, называемые справочными таблицами. Они все чаще заменяются интерактивными базами данных, которые позволяют пользователю генерировать в режиме онлайн свои собственные таблицы. Поскольку справочные таблицы являются в большей степени аналитическим инструментом, они в настоящем документе не обсуждаются.

В презентационных таблицах данные должны представляться сжатым, хорошо организованным образом в поддержку сопроводительного анализа. Хорошо сформатированная таблица может содержать большой объем информации, быстро усваиваемой читателем.

Таблицы должны быть автономными, будь то в докладе, статье, публикации или на вебстранице. Каждая таблица должна содержать достаточно метаданных, таких как описательное название и указание источника, для того чтобы они могли копироваться и включаться в другой документ, сохраняя при этом свой смысл. Если вы обеспечите автономность ваших таблиц, то это повысит их шансы быть правильно понятыми как в рамках исходного контекста, так и вне его.

3.3 Краткая памятка по подготовке эффективной таблицы

Для описания приводимых в таблице данных необходимы пять вспомогательных компонентов:

- **Название таблицы** должно давать точное и четкое описание данных. Оно должно отвечать на следующие три вопроса: "Что?", "Где?" и "Когда?". Будьте лаконичны и избегайте использования глаголов.

ХОРОШИЙ ПРИМЕР
"Планируемый возраст выхода на пенсию в разбивке по занятиям, Канада, 2007 год".
Данный текст дает всю необходимую информацию для правильного понимания и использования данных.

- **Заголовки колонок** в верхней части таблицы должны описывать данные, приводимые в каждой колонке таблицы, и предоставлять любые соответствующие метаданные (например, единица измерения, временной период или географический район).
- **Строки** в первой колонке таблицы должны описывать данные, приводимые в каждой строке таблицы.
- **Сноски** внизу таблицы могут содержать любую дополнительную информацию, необходимую для правильного понимания и использования данных (например, определения).
- **Строка источника** внизу таблицы должна указывать источник данных, например организацию, разработавшую данные, и метод сбора данных (например, перепись населения или обследование рабочей силы).

Нижеприводимый рисунок иллюстрирует организацию этих элементов таблицы.

Название таблицы	
	Заголовки колонок
Строки	Данные
Сноски	
Источник	

В презентационной таблице вы должны приводить лишь небольшой подбор ваших данных, наиболее эффективно описывающих ваш вывод, как это иллюстрирует нижеприводимый пример.

ХОРОШИЙ ПРИМЕР презентационной таблицы

Продажи промышленных товаров в Канаде, в разбивке по провинциям и территориям, июнь-июль 2008 года

Скорректированы на сезонные колебания

	Июнь 2008 года ^Г	Июль 2008 года ^Р	Июнь-июль 2008 года
	В млн. долл.		Изменение в % ¹
Канада	52 685	54 105	2.7
Ньюфаундленд и Лабрадор	692	674	-2.5
Остров Принца Эдварда	123	115	-6.1
Нью-Брансуик	1 914	1 872	-2.2
Квебек	13 019	13 280	2.0
Онтарио	23 902	25 015	4.7
Манитоба	1 360	1 445	6.2
Саскачеван	1 079	1 108	2.8
Альберта	6 298	6 316	0.3
Британская Колумбия	3 347	3 306	-1.2
Юкон	3	4	45.5
Северо-западные территории и Нунаву	4	3	-27.4

^Г Пересмотренные

^Р Предварительные

¹ Изменение в процентах рассчитывается на основе данных в тыс. канадских долларов.

Источник: Статистическое управление Канады

Необходимо разместить величины данных таким образом, чтобы пользователь мог легко извлекать ключевую информацию. Пользователи могут счесть удобным просмотр по колонкам или строкам в зависимости от излагаемого вывода. Вы должны учитывать это при принятии решения относительно того, будет ли ваша таблица представлена в горизонтальном или вертикальном формате. Затушевывание строк или создание легких оттенков может также использоваться для стимулирования читателей к горизонтальному, а также вертикальному просмотру. Разрядка и затушевывание могут менять способ прочтения таблицы.

Колонки должны быть равномерно размещены в пространстве, не слишком далеко друг от друга. Таблица должна быть лишь настолько широкой, насколько этого требует содержание данных.

Для обеспечения легкого понимания ваших таблиц вы должны следовать следующим принципам:

- Избегайте излишнего текста.
- Располагайте ваши данные либо в хронологическом порядке временных рядов, либо в соответствии с какой-то стандартной классификацией. В случае длинных рядов данных более целесообразным может являться использование обратного хронологического порядка (т.е. начиная с новейших данных с углублением в прошлое) в некоторых случаях, таких как ежемесячный уровень безработицы.
- Используйте минимум разрядок в десятичной дроби.
- Используйте разделители для тысяч. Использование пробела вместо символа содействует избежанию проблем языкового перевода.
- Выравнивайте цифры по знаку десятичной дроби (или по правой стороне в отсутствие десятичных дробей), с тем чтобы обеспечить ясность относительной величины. Выравнивайте цифры по центру колонки только в том случае, если все они имеют одинаковую величину.

- Не оставляйте пустыми клетки данных. Отсутствующие величины могут обозначаться как "данные отсутствуют" или "не применимо". В любом случае могут использоваться аббревиатуры "ДО" или "НП", которые, однако, требуют расшифровки.

Некоторые из этих руководящих принципов обсуждаются и иллюстрируются в следующем разделе.

При подготовке серии таблиц для публикации или вебсайта вы должны использовать одинаковую компоновку для всех таблиц. Определитесь с объемом информации, которая должна быть представлена в названиях таблиц (что является очевидным, а что нет) и будьте последовательны в использовании сокращений.

3.4 Использование округлений и десятичных дробей

Многие не являющиеся статистиками пользователи сталкиваются с трудностями при проведении различия между цифрами с тремя или более разрядами десятичной дроби. Вы можете помочь им путем округления величин, представляемых в ваших таблицах. Округление может также использоваться в тех случаях, когда данные не имеют высокого уровня точности. В некоторых случаях лишь округленные данные являются надежными и, следовательно, должны приводиться в таблицах. Однако вы должны позаботиться о том, чтобы округление данных не привело к избыточной потере информации.

ХОРОШИЙ ПРИМЕР	ПЛОХОЙ ПРИМЕР
1 320 000	1324567
1 670 000	1673985
1 830 000	1829456

В вышеприведенном примере округленные цифры слева легче поддаются пониманию и запоминанию, чем точные цифры справа. Использование пробела в качестве разделителя тысяч также иллюстрируется в настоящем примере.

Если вам необходимо привести величины с разными разрядами десятичной дроби, вы должны выровнять их по разрядам, а не по правой стороне. В нижеприведимом примере величины в левой колонке легче воспринимаются, чем величины в правой. Данный пример также демонстрирует, что использование одного и того же разряда десятичной дроби во всех величинах является намного более целесообразным.

ХОРОШИЙ ПРИМЕР	ПЛОХОЙ ПРИМЕР
93.2	93.2
1045.0	1045
385.6	385.63

Числовые величины должны быть выровнены по правой стороне. Используя тот же пример, отметим, насколько трудно поддаются считыванию величины, когда цифры выровнены по левой стороне, как это показано ниже.

ХОРОШИЙ ПРИМЕР	ПЛОХОЙ ПРИМЕР
93.2	93.2
1045.0	1045.0
385.6	385.6

3.5 Пример возможностей улучшения таблицы

В качестве иллюстрации эффективности руководящих принципов, изложенных в разделе 3.3, мы приводим пример неудачной таблицы и демонстрируем возможности ее простого усовершенствования.

ПЛОХОЙ ПРИМЕР

Конечное потребление энергии в разбивке по секторам – проценты

	1980	1985	1990	1995	2000	2002	2003
Транспорт	27.81	27.92	28.24	31.12	36.82	39.48	39.13
Бытовой сектор	31.11	33.91	30.41	27.61	24.33	23.71	23.97
Промышленность	31.47	27.21	23.86	22.11	21.41	19.53	18.78
Сельское хозяйство	n/a	n/a	3.51	3.7	3.11	2.91	2.82
Сфера услуг	9.61	10.96	13.98	15.46	14.33	14.37	15.3
Итого	100	100	100	100	100	100	100

Что неправильно в вышеприведенной таблице?

- Неизвестно, к какому географическому району относятся данные.
- Источник данных не указан.
- Величины выровнены по центру, а не по правой стороне.
- Не должны указываться величины с двумя десятичными разрядами (чрезмерная информация).
- Итоговые величины должны иметь тот же десятичный разряд, что и другие величины.
- Аббревиатура "n/a" не поясняется.
- Затушевка серым цветом и линии одного и того же размера между каждой строкой и каждой колонкой не способствует пониманию различий в представленных в таблице данных.
- Таблица не обязательно должна занимать всю ширину страницы.

ХОРОШИЙ ПРИМЕР

Доля секторов в общем потреблении энергии (в %)

Ирландия, 1980–2003 годы

	1980	1985	1990	2000	1995	2002	2003
Транспорт	27.8	27.9	28.2	31.1	36.8	39.5	39.1
Бытовой сектор	31.1	33.9	30.4	27.6	24.3	23.7	24.0
Промышленность	31.5	27.2	23.9	22.1	21.4	19.5	18.8
Сельское хозяйство	n/a ¹	n/a ¹	3.5	3.7	3.1	2.9	2.8
Сфера услуг	9.6	11.0	14.0	15.5	14.4	14.4	15.3
Итого	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹ Данные о потреблении энергии сельскохозяйственным сектором не собирались до 1990 года.

Источник: Департамент государственных предприятий Ирландии.

Каким образом данная таблица была усовершенствована?

- Вся информация, необходимая для понимания данных, предоставляется в заголовке и подзаголовке.
- Указан источник данных.
- Все величины выровнены по правой стороне и указываются с одним десятичным разрядом.
- Сокращение "n/a" поясняется в сноске.
- В таблице использованы только те линии, которые разделяют различные компоненты таблицы (заголовок, данные, сноску и источник); кроме того, удалено ненужное затушевывание.
- Таблица не шире, чем это требуется для показа всех подзаголовков и данных.

4. Графики

4.1 Причины использования графиков

Статистика зачастую лучше поддается пониманию, если она изображена в виде графика, а не таблицы. График служит визуальным представлением статистических данных, в котором данные представлены такими символами, как столбцы или линии. Они являются весьма эффективным инструментом графического представления, поскольку обеспечивают оперативное и удобное изображение данных, облегчают сопоставление и позволяют выявлять тренды и взаимосвязи между данными.

График, как правило, приобретает форму одно- или двухмерного рисунка, такого как гистограмма или линейный график. Хотя существуют также трехмерные графики, они, как правило, рассматриваются в качестве слишком сложных для быстрого понимания.

Графики могут использоваться для иллюстрации трендов в большом объеме данных или изложения ключевого вывода или идеи. Возможность использования графиков следует рассматривать, если вы хотите показать:

- **Сопоставление:** Сколько? Какая позиция является большей или меньшей?
- **Изменения во времени:** Каким образом меняется переменная?
- **Плотность распределения:** Каким образом распределены элементы? Каковы различия?
- **Корреляцию:** Существует ли связь между переменными?
- **Относительную часть целого:** Каким образом определен элемент соотносится с целым?

В настоящей главе мы рассматриваем наиболее широко используемые типы графиков и даем рекомендации по подготовке эффективных графиков.

4.2 Памятка по разработке эффективного графика

Если вы считаете, что график является наиболее удачным способом представления ваших данных, причем неважно, какой тип графика будет использоваться, вы должны соблюдать следующие три принципа:

1. **Определите вашу целевую аудиторию:** Что она знает о данном вопросе?
2. **Определите мысль, которую вы хотите донести до вашего читателя:** О чем говорят данные? Хотите ли вы донести до читателя еще какие-то мысли?
3. **Определите характер вашей мысли:** Хотите ли вы сопоставить данные, продемонстрировать временные тренды или проанализировать взаимосвязи в ваших данных?

Хороший график

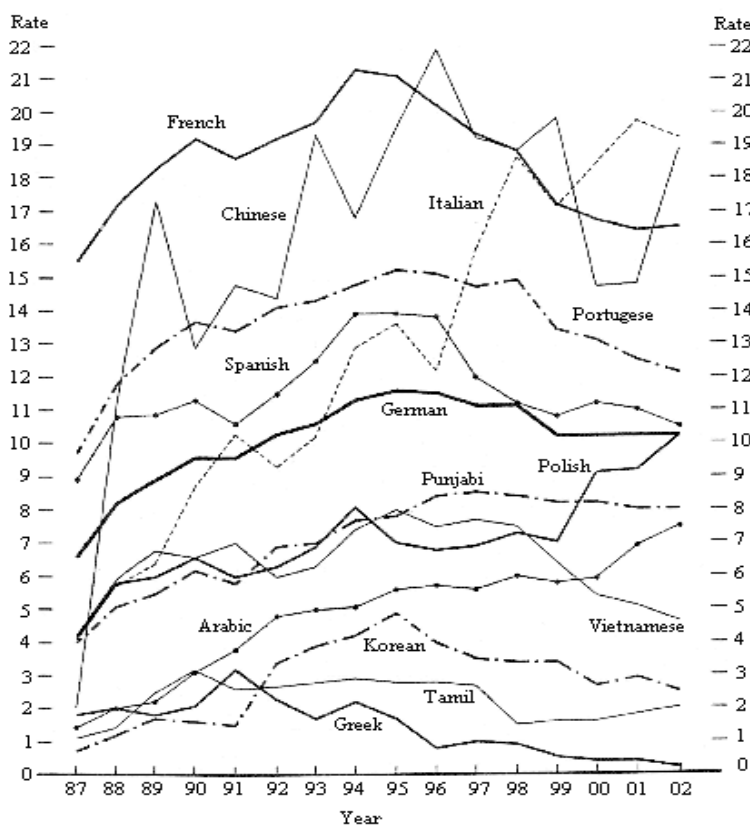
- привлекает внимание читателя;
- просто, четко и точно излагает информацию;
- не вводит в заблуждение;
- обеспечивает сжатое изложение данных (например, одна линия графика вместо нескольких круговых диаграмм);
- облегчает сопоставление данных и выявление трендов и различий;
- иллюстрирует выводы, темы или сюжет в сопроводительном тексте.

4.3 Случаи, когда графики могут являться не самым удачным инструментом

График не всегда является наиболее подходящим инструментом для представления статистической информации. В некоторых случаях текст и/или таблица данных могут служить более эффективными средствами разъяснения данных вашей аудитории и обеспечивать значительную экономию времени и трудозатрат.

ПЛОХОЙ ПРИМЕР линейного графика

Число учащихся, использующих английский язык в качестве второго языка в Западном университете, в разбивке по первому языку, 1987-2002 годы



Источник: Statistics Canada, *Learning Resources: Using graphs*⁵.

⁵ <http://www.statcan.gc.ca/edu/power-pouvoir/ch9/using-utilisation/5214829-eng.htm>

Вы должны еще раз проанализировать целесообразность использования графика в тех случаях, когда ваши данные:

- характеризуются большим разбросом;
- имеют слишком мало значений;
- имеют слишком много значений;
- демонстрируют малую вариацию или отсутствие таковой.

Вам следует избегать любой схожести с вышеприведенным линейным графиком. Данные являются слишком многочисленными, и те выводы, которые аналитик надеялся проиллюстрировать с его помощью, теряются в джунглях линий.

4.4 Выбор соответствующего типа графика

Крайне важно знать, какие типы графиков следует использовать в отношении определенных типов информации. Некоторые графики более, чем другие подходят для иллюстрации определенных типов данных. В настоящем разделе мы излагаем рекомендации в отношении наиболее широко используемых типов графиков: гистограмм и демографических пирамид, линейных графиков, круговых диаграмм и диаграмм рассеяния.

Гистограммы

Гистограмма является простейшим типом графика с точки зрения подготовки и понимания. Она используется для сопоставления частотности или величин по разным категориям или группам.

ХОРОШИЙ ПРИМЕР гистограммы



Источник: Статистическая база данных ЕЭК ООН

Гистограммы могут быть как вертикальными, так и горизонтальными. При горизонтальной ориентации текст воспринимается легче, как об этом свидетельствует вышеприведенный пример. Кроме того, сопоставление различных величин облегчает ранжирование столбцов по размеру от наименьшего до наибольшего, а не в случайном порядке.

Столбцы должны быть намного шире, чем пробелы между ними. Пробелы не должны быть больше 40% от ширины столбцов.

Составная столбиковая диаграмма может использоваться для демонстрации и составления составляющих компонентов итоговых цифр. При использовании данного типа графика следует проявлять осторожность. Они могут создавать трудности с анализом и сопоставлением, если в каждом составном столбике используется слишком много элементов или многие элементы довольно близки по размеру.

ХОРОШИЙ ПРИМЕР составной столбиковой диаграммы

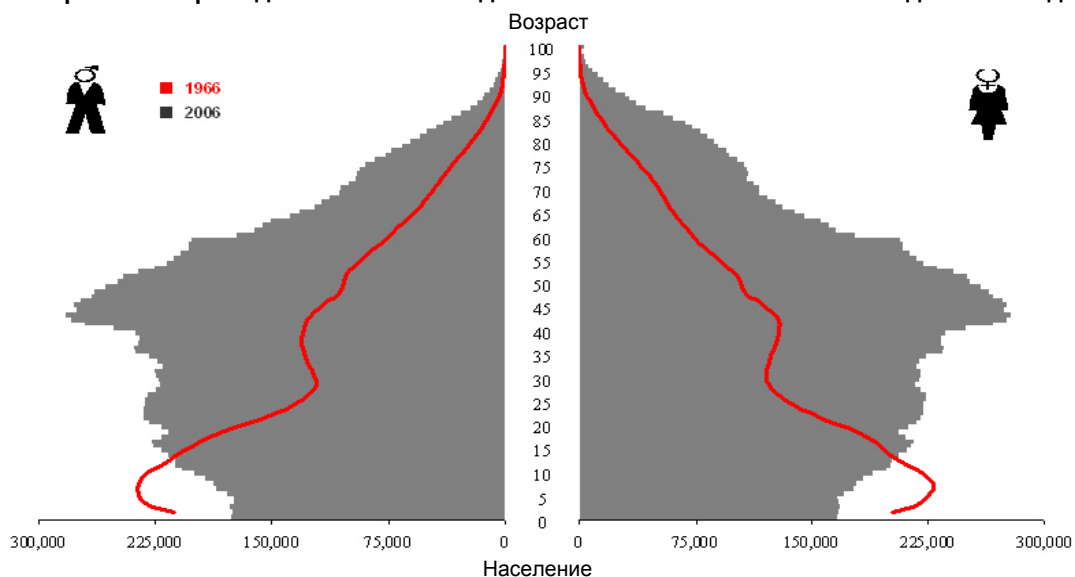


Источник: Статистическая база данных ЕЭК ООН

Демографическая пирамида представляет собой комбинацию двух горизонтальных гистограмм, описывающих возрастную структуру женского и мужского населения страны или региона. Мужчины традиционно указываются слева, а женщины - справа. В случае сопоставления разных демографических пирамид на них, как правило, более целесообразно указывать процент мужчин и женщин в общем населении, а не их численность.

ХОРОШИЙ ПРИМЕР демографической пирамиды

Возрастная пирамида населения Канады по состоянию на 1 июля 1966 года и 2006 года



Источник: Статистическое управление Канады⁶

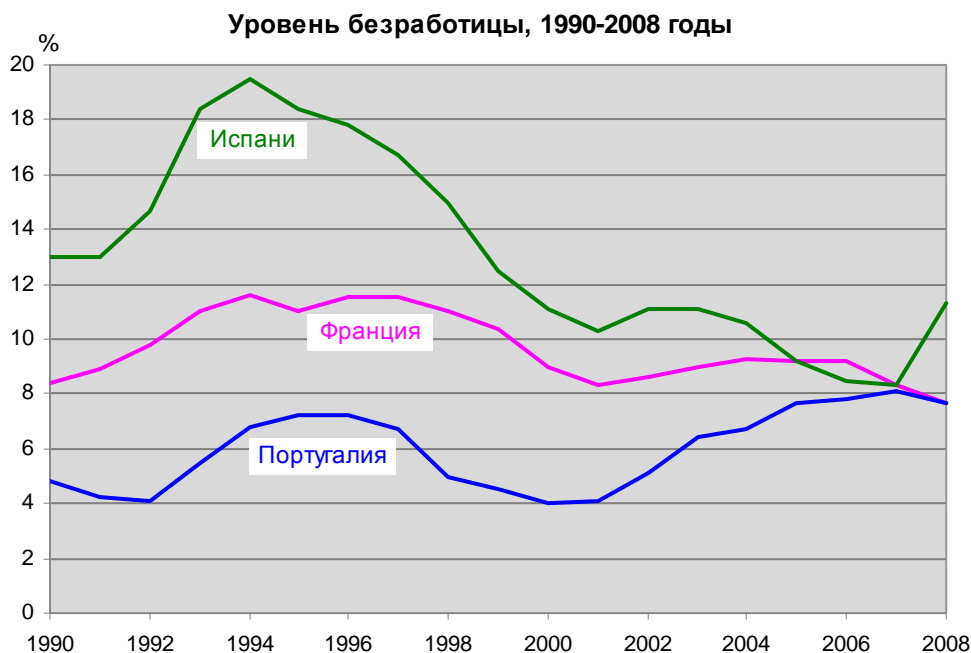
В большинстве европейских стран демографические пирамиды более не имеют форму пирамиды, однако остаются весьма эффективным способом представления большого объема информации о возрастной структуре населения, особенно в тех случаях, когда они являются "анимированными", т.е. изображают изменения во времени.

Линейные графики

Линейный график служит эффективным средством визуализации трендов в данных во времени и, следовательно, наиболее подходящим типом графика для временных⁵ рядов. Существует возможность корректировки параметров графика для более эффективной передачи выводов, однако необходимо проявлять осторожность во избежание искажения данных. Данный вопрос обсуждается и иллюстрируется в разделе 4.6.

⁶ <http://www.statcan.gc.ca/daily-quotidien/061026/figure.htm>

ХОРОШИЙ ПРИМЕР линейного графика



Источник: Статистическая база данных ЕЭК ООН

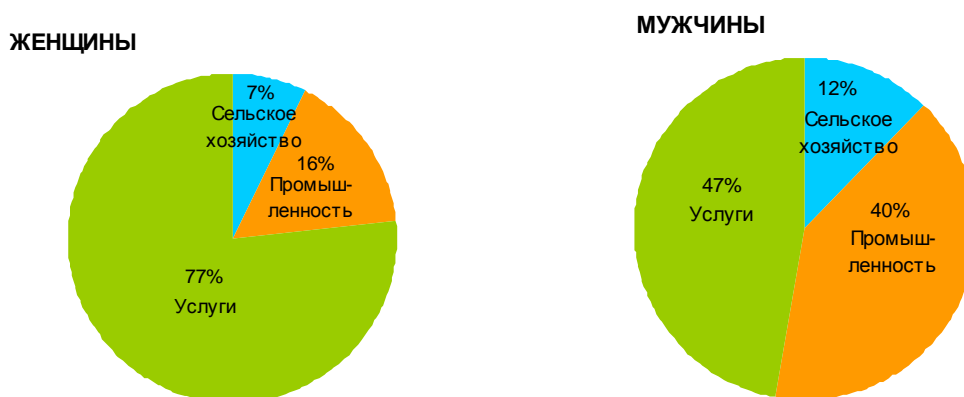
Круговые диаграммы

Круговая диаграмма может использоваться для иллюстрации процентного распределения одной переменной, однако позволяет использовать лишь небольшое число категорий, как правило, не более шести. Использование такого типа графика не рекомендуется многими статистиками, поскольку сопоставление различных сегментов круга может создавать трудности, особенно при сопоставлении данных между различными круговыми диаграммами. Для преодоления данной проблемы в сегментах могут указываться их фактические величины. В некоторых случаях на диаграмме также могут указываться названия категорий, что позволяет отказаться от разъяснения условных обозначений. Как правило, наиболее целесообразно располагать сегменты от малого к большому, а не чередовать малые и крупные сегменты.

В большинстве случаев более целесообразно использовать другие типы графиков (например, гистограммы), однако круговые диаграммы не следует полностью сбрасывать со счетов, поскольку они являются эффективным средством визуализации удельного веса категории в целом. Круговые диаграммы также могут использоваться для представления обзорной информации, примером чему может служить нижеприводимый рисунок.

ХОРОШИЙ ПРИМЕР круговой диаграммы

Занятость в разбивке по основным секторам в Латвии, 2007 год



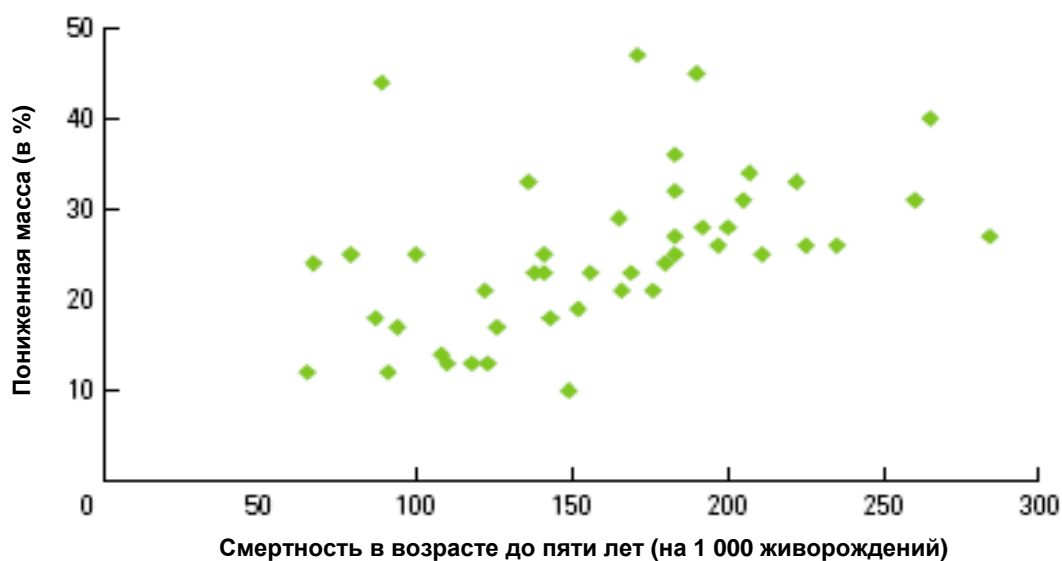
Источник: Статистическая база данных ЕЭК ООН

Диаграммы рассеяния

Диаграмма рассеяния используется для демонстрации взаимосвязи между двумя переменными. Она является наиболее точным способом демонстрации корреляции, как об этом свидетельствует нижеприводимый пример. Однако некоторые аналитики предпочитают использовать гистограммы, поскольку диаграммы рассеяния могут создавать трудности с интерпретацией.

ХОРОШИЙ ПРИМЕР диаграммы рассеяния

Смертность в возрасте до пяти лет и распространенность пониженной массы в африканских странах, расположенных к югу от Сахары, 2003 год



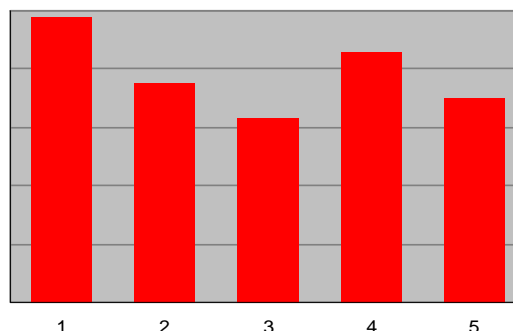
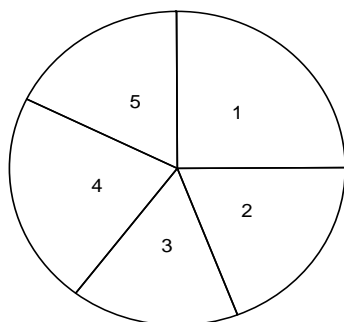
Источник: Jamison et al. (2006) Disease and Mortality in Sub-Saharan Africa, 2nd edition, Washington D.C., The World Bank⁷.

⁷ [http://www.dcp2.org/file/66/Disease and Mortality in SSA.pdf](http://www.dcp2.org/file/66/Disease%20and%20Mortality%20in%20SSA.pdf)

Экспериментирование с различными типами графиков

Какой тип графика вам следует использовать? Вам необходимо сделать такой выбор. Одним из удачных подходов является экспериментирование с различными типами графиков для определения набора инструментов, наиболее подходящего для передачи вашей мысли.

Ниже приводятся иллюстрации двух различных способов графического представления одних и тех же данных. Какой из них является более ясным?



Можете ли вы сказать, какой сегмент круговой диаграммы является наибольшим? Некоторые пользователи считают более трудным сопоставление сегментов, чем столбиков или линий. На круговой диаграмме сегменты 1 и 4 выглядят практически одинаковыми, в то время как разница в их относительном размере является очевидной на гистограмме.

4.5 Составляющие эффективного графика

Компоненты графика

Различные компоненты графика борются между собой за внимание читателя. Чем больше характеристик вы включите, тем труднее вам будет четко донести до читателя свою мысль.

Компоненты графика подразделяются на следующие три категории:

1. **Компоненты данных**, которые изображают данные: столбики, линии, сегменты или точки.
2. **Вспомогательные компоненты**, которые содействуют пониманию данных: название графика, условные обозначения, названия данных, сетка, сноски и источник данных.
3. **Декоративные элементы**, которые не связаны с данными.

Компоненты данных сами по себе никогда не являются самодостаточными. Для обеспечения правильного понимания ваших графиков вам необходимо использовать следующие вспомогательные компоненты:

- **Название графика** должно обеспечивать четкое представление о том, чему посвящен данный график. Оно должно быть кратким и лаконичным. Существует два типа названий:

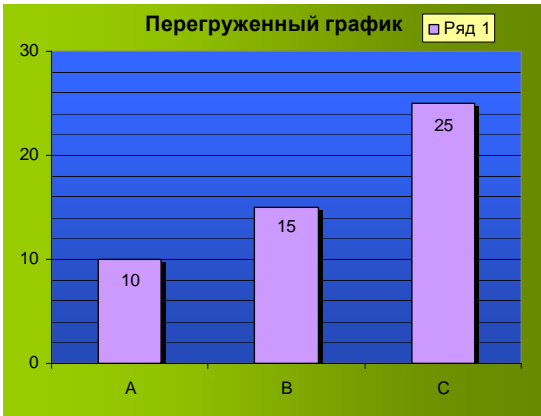
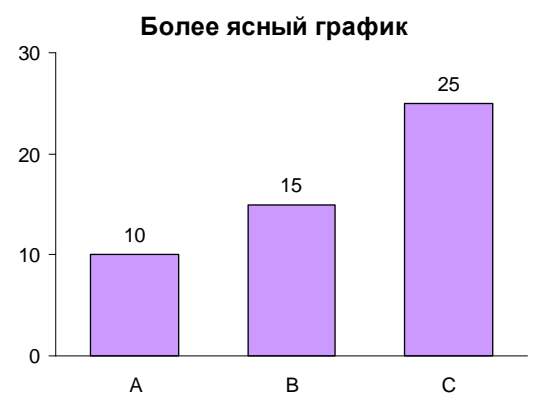
- **Информативное название** дает всю информацию, необходимую для понимания данных. Оно должно обеспечивать ответы на следующие три вопроса: "Что?", "Где?" и "Когда?".
- **Описательное название** представляет собой надпись, которая описывает основную динамику или тренд, иллюстрируемые графиком. Оно в нескольких словах излагает историю, которую иллюстрирует график.
- **Названия осей** должны описывать величины, приводимые в графике. Названия размещаются горизонтально на обеих осях.
- **Обозначения осей** должны указывать единицу измерения данных (например, "в тысячах", "в %", "возраст (в годах)" или "долл."). Не следует указывать обозначения, когда единица измерения является очевидной (например, "годы" в отношении временных рядов).
- **Линии сетки** могут добавляться в гистограммы и линейные графики для облегчения считывания и сопоставления значений данных.
- **Условные обозначения и названия данных** должны пояснять символы, динамику или цвета, используемые для представления данных на графике. Условные обозначения не указываются, когда на графике приводится только один ряд значений. По мере возможности, вы должны использовать названия данных, а не условные обозначения. Названия данных указываются на компоненте данных или вслед за ним (столбцы, сектора, линии) для облегчения их выявления и понимания.
- **Сноски** могут использоваться для представления определений или методологической информации.
- **Источник данных** должен указываться внизу графика.

Данные должны быть в центре внимания

Для максимального повышения эффективности графика данные должны занимать центральное место. Вспомогательные компоненты должны:

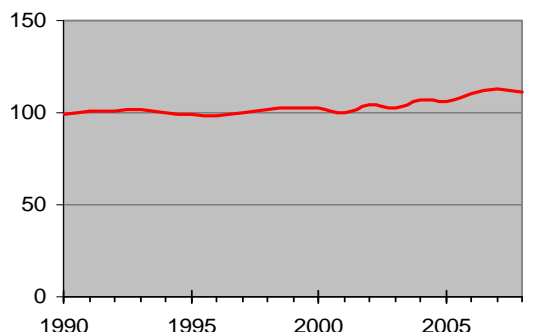
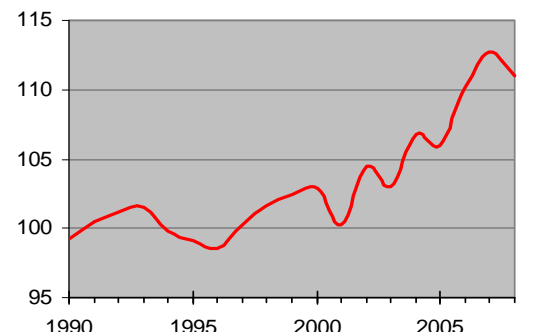
- **Присутствовать только в случае необходимости.** Названия осей, условные обозначения и названия данных будут иметь важное значение для правильного понимания вашего графика или будут отсутствовать вообще в зависимости от характера ваших данных.
- **Быть вынесены в подзаголовок.** Используйте более светлые линии для обозначения осей и линий сетки по сравнению с компонентами данных. Декоративные компоненты не должны отвлекать на себя внимание пользователя.

Компоненты данных также могут вступать в противоречие друг с другом. Чем больше переменных и значений вы захотите указать, тем более трудным будет ясное представление данных. Эффективный график несет в себе четкое визуальное сообщение. Если перед графиком ставятся избыточные задачи, он становится головоломкой, которая потребует излишней работы ума для понимания. В худшем случае он будет просто вводить в заблуждение.

ПЛОХОЙ ПРИМЕР	ХОРОШИЙ ПРИМЕР
<p style="text-align: center;">Перегруженный график</p> 	<p style="text-align: center;">Более ясный график</p> 
<p>Все компоненты обладают максимальной нагрузкой. В результате этого график является перегруженным, что создает трудности для восприятия, даже если в нем приводятся всего лишь три величины.</p>	<p>Настоящий график является намного более простым для понимания. Минимальное использование вспомогательных компонентов обеспечивает уделение главного внимания данным.</p>

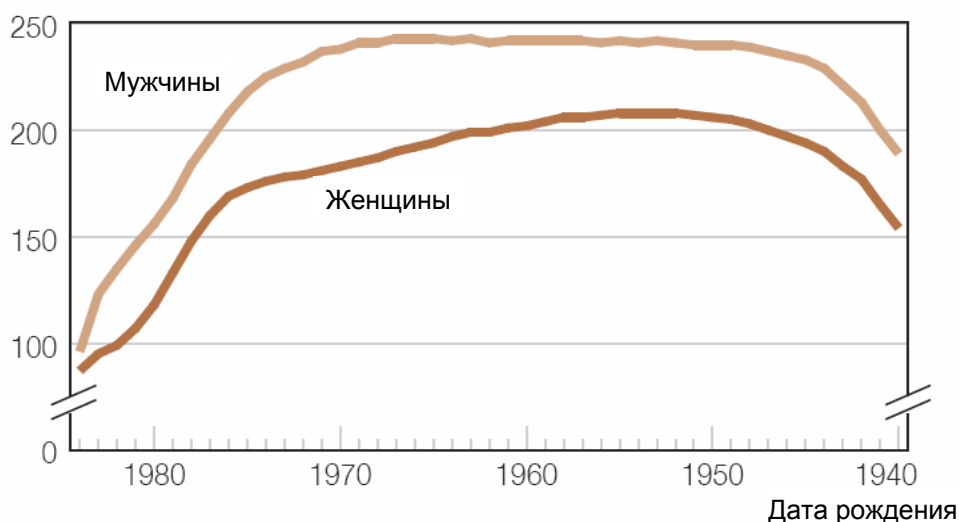
4.6 Корректировка параметров графика

При разработке графика вы можете скорректировать его масштаб для более четкого оформления вашей мысли. На нижеприводимых графиках показаны одни и те же данные, которые иллюстрируются с помощью весьма различных изображений:

	
<p>Данный график демонстрирует общую тенденцию стабильности на протяжении 10 лет, после чего начинается умеренный рост.</p>	<p>Уменьшение масштаба по оси у позволяет подчеркнуть изменения. Первые 10 лет характеризуются малыми изменениями в сторону повышения и понижения, на фоне чего рост в конце периода выглядит намного более значительным.</p>

Целесообразно использовать определенный символ для указания того, что шкала значений начинается не с нуля, как, например, на оси у на правом рисунке вышеприведенного примера. Наилучшим вариантом является начало с 0 с использованием зигзага или разрыва, как это показано на нижеприведенном примере.

ХОРОШИЙ ПРИМЕР графика, на котором ось у не начинается с нуля
Зачитываемый в целях пенсии доход лиц в возрасте 20-64 года в Швеции, 2004 год
Средний доход в тыс. шведских крон



Источник: Statistics Sweden (2006), *Women and Men in Sweden: Facts and figures 2006*⁸.

4.7 Регулирование когнитивной нагрузки ваших графиков

Ваши данные могут нести несколько идей, которые вы хотите подчеркнуть с использованием графика. Графики, как и любой другой элемент публикации, могут иметь "когнитивную нагрузку". Когнитивная нагрузка обычно означает степень трудности для понимания читателем той идеи, которую вы пытаетесь до него донести. График с высокой когнитивной нагрузкой может создавать трудности с пониманием и запоминанием. Его идея может с трудом поддаваться передаче. График с низкой когнитивной нагрузкой будет пониматься сразу же. Его идея будет очевидной. Большинство руководящих принципов в области эффективной разработки графиков рекомендуют использование низкой когнитивной нагрузки.

При разработке графика вы должны регулировать его когнитивную нагрузку. Вы можете ее снизить и обеспечить четкость идеи за счет использования надлежащих правил и форматов. Вы также можете произвольно попытаться повысить его когнитивную нагрузку, если хотите, чтобы ваша аудитория сосредоточила свое внимание на одном из трудноуловимых аспектов ваших данных. Повышение когнитивной нагрузки заставит читателей рассматривать график с иной точки зрения. Ниже приводится пример графика с высокой когнитивной нагрузкой.

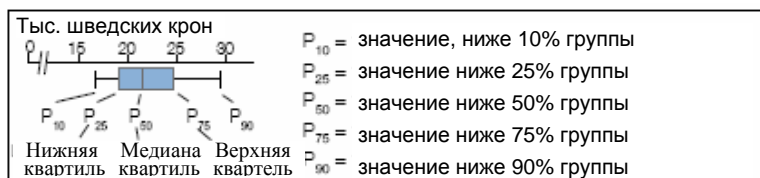
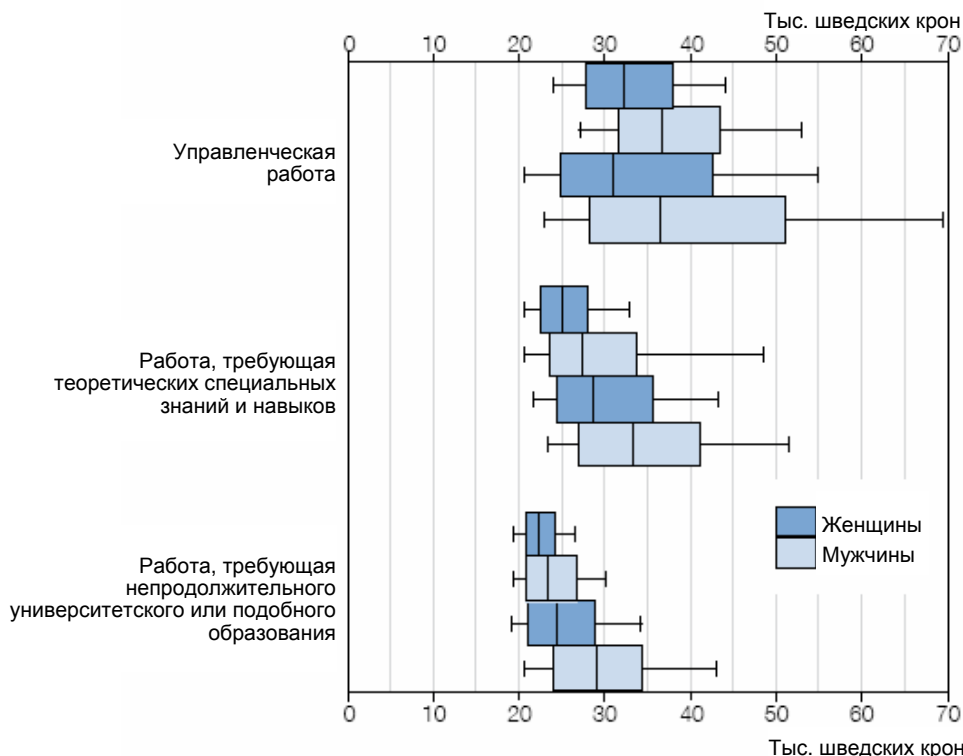
⁸ http://www.scb.se/statistik/_publikationer/LE0202_2006A01_BR_X10ST0602.pdf

ХОРОШИЙ ПРИМЕР графика с высокой когнитивной нагрузкой

Распределение заработной платы в профессиональных группах, которые требуют высшего образования, в Швеции, 2004 год

Ежемесячная зарплата в шведских кронах

Две верхние гистограммы в каждой профессиональной группе описывают мужчин и женщин, занятых в государственном секторе, а две нижние - в частном секторе.



Источник: Statistics Sweden (2008), *Women and Men in Sweden: Facts and figures 2008*⁹.

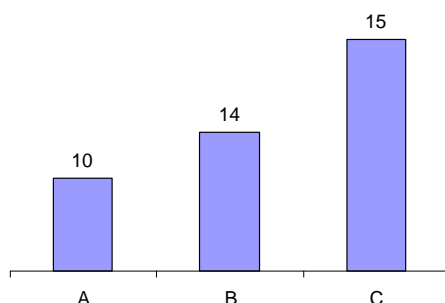
4.8 Рекомендации по улучшению ваших графиков

Будьте точны

Графические объекты должны пропорционально изображать соотношения. График, который изображает данные в качестве объектов непропорционального размера (см. нижеприводимый пример) вводит в заблуждение.

⁹ http://www.scb.se/statistik/publikationer/LE0202_2008A01_BR_X10BR0801ENG.pdf

ПЛОХОЙ ПРИМЕР относительного размера между графическими объектами



Упорядочивайте ваши данные

При использовании гистограмм или круглых диаграмм вы должны выстраивать ваши данные в порядке от наименьшего до наибольшего значения, чтобы облегчить их сопоставление.

ПЛОХОЙ ПРИМЕР	ХОРОШИЙ ПРИМЕР																												
<p>Уровень подростковой рождаемости, 2006 год</p> <table border="1"><thead><tr><th>Страна</th><th>Уровень рождаемости</th></tr></thead><tbody><tr><td>Албания</td><td>13</td></tr><tr><td>Грузия</td><td>36</td></tr><tr><td>Греция</td><td>11</td></tr><tr><td>Венгрия</td><td>20</td></tr><tr><td>Румыния</td><td>35</td></tr><tr><td>Сербия</td><td>23</td></tr></tbody></table>	Страна	Уровень рождаемости	Албания	13	Грузия	36	Греция	11	Венгрия	20	Румыния	35	Сербия	23	<p>Уровень подростковой рождаемости, 2006 год</p> <table border="1"><thead><tr><th>Страна</th><th>Уровень рождаемости</th></tr></thead><tbody><tr><td>Греция</td><td>11</td></tr><tr><td>Албания</td><td>13</td></tr><tr><td>Венгрия</td><td>20</td></tr><tr><td>Сербия</td><td>23</td></tr><tr><td>Румыния</td><td>35</td></tr><tr><td>Грузия</td><td>36</td></tr></tbody></table>	Страна	Уровень рождаемости	Греция	11	Албания	13	Венгрия	20	Сербия	23	Румыния	35	Грузия	36
Страна	Уровень рождаемости																												
Албания	13																												
Грузия	36																												
Греция	11																												
Венгрия	20																												
Румыния	35																												
Сербия	23																												
Страна	Уровень рождаемости																												
Греция	11																												
Албания	13																												
Венгрия	20																												
Сербия	23																												
Румыния	35																												
Грузия	36																												
<p>Данные приводятся в алфавитном порядке (латиница) названий стран. Эти данные создают значительные сложности для сопоставления. В центре внимания оказываются первая и последняя величина, которые не имеют особого значения.</p>	<p>Данные приводятся в восходящем порядке значений, что облегчает их сопоставление. В центре внимания оказываются минимальная и максимальная величины набора данных.</p>																												

Избегайте ошибочных корреляций

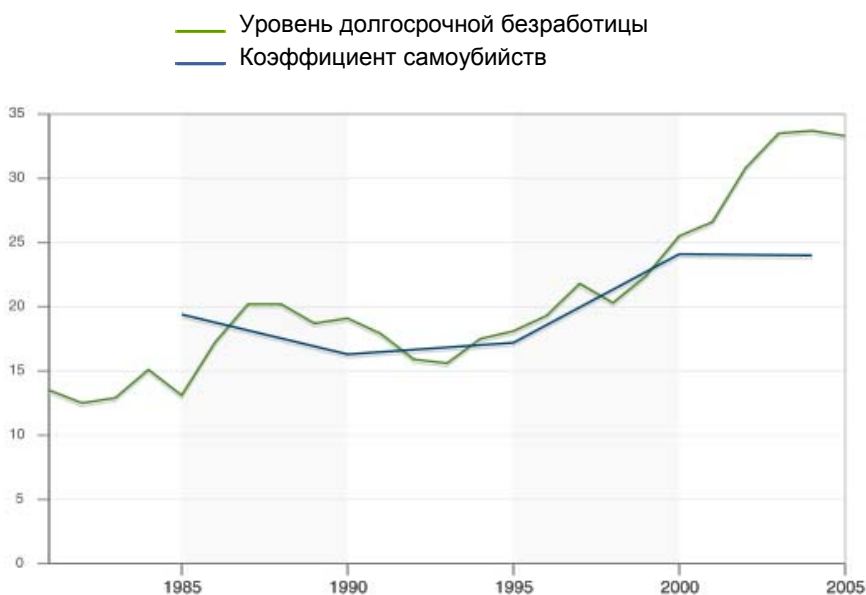
Отложение кривых переменных с использованием различного масштаба на одном и том же графике способно привести к ошибочным выводам. Тот факт, что две кривые имеют схожую динамику, является недостаточным для установления корреляции.

На нижеприводимом графике предпринимается попытка увязать коэффициент самоубийств с уровнем долгосрочной безработицы в Японии. Обе кривые, как представляется, обладают схожей динамикой, однако эти две переменные являются несопоставимыми. Первая описывает число самоубийств **на 100 000** населения, в то время как вторая описывает число безработных в течение 12 месяцев или более в качестве **процента** от общей численности

безработных. Обе переменные обладают значениями от 10 до 35 на протяжении всего рассматриваемого периода, однако небольшая корректировка определения или масштаба приведет к получению совершенно иного графика. Этот график успешно создает впечатление взаимосвязанности двух переменных, однако он не может доказать и не доказывает этого.

ПЛОХОЙ ПРИМЕР: обманчивая корреляция между двумя переменными

Самоубийства и долгосрочная безработица в Японии



Источник: Swivel¹⁰

С осторожностью относитесь к использованию двойной оси y

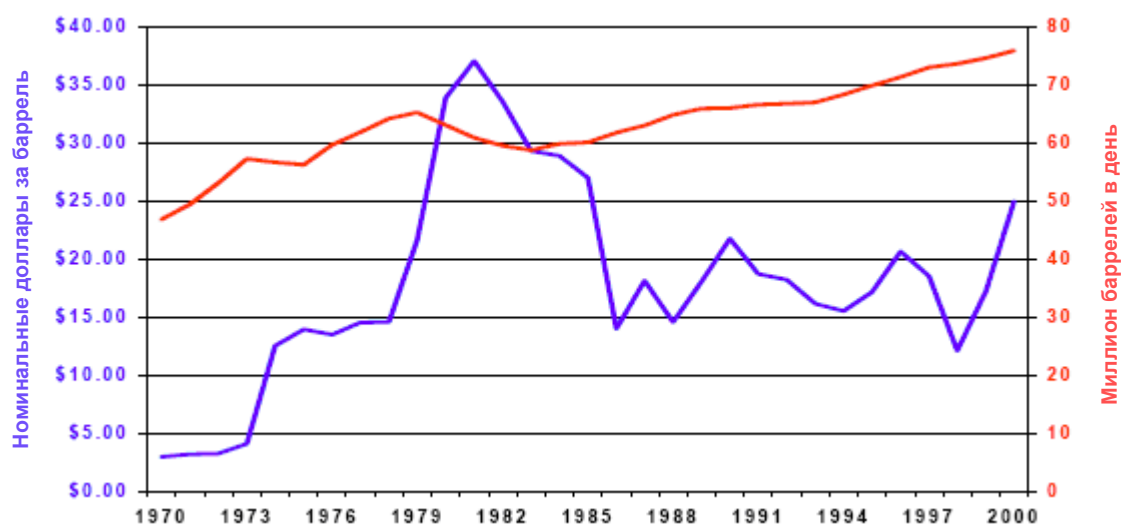
Двойная ось y способна создавать путаницу. Вы можете успешно использовать этот тип графики только в том случае, если вы имеете две различные переменные (например, цены и физический объем). Однако вы должны проявлять исключительную осторожность при использовании обозначений и наносить каждую линию данных одним цветом, как это показано на нижеприведенном примере.

Когда вы имеете дело со схожими единицами в двух различных масштабах на одном и том же графике, вы можете ввести в заблуждение даже экспертов. Линии могут "пересекаться" на графике с двойной осью y, на котором при использовании одинакового масштаба они никогда не будут близко подходить друг к другу.

¹⁰ <http://www.swivel.com/graphs/show/28847825>

ХОРОШИЙ ПРИМЕР графика с двойной осью у

Мировой спрос и мировые цены на нефть



Источник: Blessing et al. (2003), *Cognitive Testing of Statistical Graphs: Methodology and Results*¹¹.

Избегайте использования избыточных элементов графики

Любые графические элементы, которые ничего не сообщают о наборе данных, который вы представляете, затрудняют понимание вашего графика. Это особенно касается использования трехмерных графиков для представления простых наборов данных. Хотя в настоящее время в наличии имеется программное обеспечение для легкой подготовки трехмерных графиков, которые могут некоторым казаться привлекательными, эти графики часто приводят к искажению данных. Вашу мысль будет намного легче понять, если она будет изображена с помощью простого двухмерного графика, приводимого ниже.

ПЛОХОЙ ПРИМЕР



ХОРОШИЙ ПРИМЕР

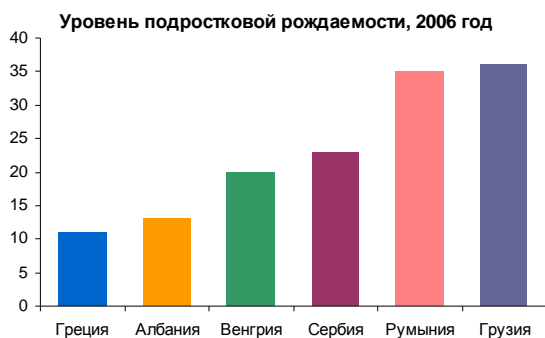


Трехмерные графики редко повышают полезность и часто вводят читателей в заблуждение. Эти изображения обладают глубиной, которая визуально приближает одни части и отдаляет другие. Мозг компенсирует это, считая удаленные объекты бо⁵льшими по размеру. Однако изображение данных, имеющих более двух признаков, с использованием глубины может на практике обеспечить более четкое понимание читателями вашей мысли.

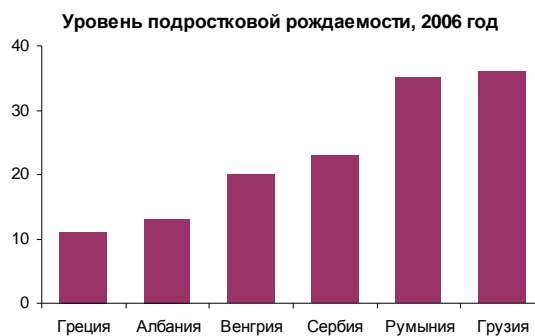
¹¹ <http://www.fcs.m.gov/03papers/BlessingBradsher.pdf>

Как правило, вы должны избегать добавления любых признаков, которые не несут в себе никаких данных. Так, например, использование отличного цвета для каждой величины на гистограмме затруднит его прочтение, как об этом свидетельствует ниже приводимый пример. Вы должны ограничиваться одним цветом.

ПЛОХОЙ ПРИМЕР



ХОРОШИЙ ПРИМЕР



5. Карты

5.1 Почему карта стоит тысячи цифр

Географическая информация является неотъемлемой частью всех статистических данных. Географические районы обладают границами, названиями и несут другую информацию, позволяющую локализовать их на месте и соотнести с ними статистическую информацию. Данная пространственная привязка имеет особое значение в случае данных переписи.

Карты являются наиболее эффективным инструментом визуализации пространственных трендов. При условии их тщательной разработки и оформления они становятся уже не просто декоративным элементом статистической презентации. Они помогают выявлению и пониманию распределений и трендов, которые трудно различимы в таблицах и графиках.

Если "картинка стоит тысячи слов", то "карта стоит тысячи цифр". В наше время зрительных образов карты являются мощным носителем информации. Они служат ценными вспомогательными инструментами принятия решений для экспертов, политиков и широкой общественности и содействуют удовлетворению растущего спроса на информацию во всех слоях общества.

Преимущества карт

Карты являются хорошо составленными, если они легко поддаются пониманию. Они должны помогать пользователю осмыслить с первого взгляда большой объем информации. Они способны резюмировать объемные таблицы данных или длинные и сложные тексты. Когда вы хотите представить статистическую информацию по всем регионам своей страны, вы можете подготовить целый набор графиков или же можете отобразить всю эту информацию на одной карте.

В настоящее время в наличии имеются многочисленные картографические технологии, начиная с географических информационных систем (ГИС), которые обладают широким набором аналитических функций и обеспечивают интеграцию компонентов картографии, и кончая высокопроизводительными картографическими информационными системами (КИС) для профессионалов в области составления карт и атласов в статистических управлениях. В эпоху бумажных публикаций карты зачастую недостаточно использовались статистиками, поскольку не содержали точных цифр. Этот недостаток исчез с появлением интерактивных инструментов картирования, которые позволяют пользователю получать данные "за картой".

Использование карт в статистике

Карты могут быть весьма полезными как при подготовке переписей и обследований, так и при анализе и представлении результатов. Мы должны рассмотреть возможность использования карт, если вы хотите:

- показать географическое местоположение и пространственное распределение ваших данных;
- сопоставить различные районы;
- резюмировать большой объем данных и снизить их сложность;
- четко оформить свою мысль;
- проверить ваши выводы;
- привлечь внимание читателя;
- хранить пространственную информацию в географической информационной системе.

В настоящей главе мы рассматриваем наиболее широко используемые типы карт и даем рекомендации по подготовке эффективных карт.

5.2 Памятка по разработке эффективной карты

Составление карт являет собой смесь искусства, науки и технологии. Это является сложной задачей, поскольку существуют неограниченные возможности с точки зрения организации макета карты.

С развитием технологий Web 2.0 (см. разделы 6.2 и 6.4) в онлайн-доступе появилось большое число интерактивных инструментов картирования, которые позволяют пользователям загружать свои данные и моментально создавать тематические карты. Подготовка тематических карт стала намного более дешевой и быстрой, что однако необязательно означает разработку эффективных карт, точно отражающих вашу мысль.

Для разработки эффективной карты вы должны соблюдать следующие четыре руководящих принципа:

1. **Определите вашу целевую аудиторию:** Каким образом и в каком контексте будет использоваться карта? Существуют ли какие-либо ограничения с точки зрения ее доступности?
2. **Определите мысль, которую вы хотите донести до читателя:** Что изображают данные? Желаете ли вы донести до читателя какие-то еще дополнительные мысли?
3. **Определите характер ваших данных:** Какое число переменных должно быть картировано? Имеется ли временное измерение?
4. **Определите надлежащую методику картирования, цвета и условные обозначения:** Каков характер ваших данных (количественные или качественные, абсолютные или относительные величины)? Существуют ли какие-либо технические ограничения (например, формат или черно-белая палитра)? Каковы правила для цветов или классификаций?

Эффективная карта:

- является простой и легко понимаемой;
- несет явную и объективную мысль;
- обеспечивает точное представление данных и не вводит в заблуждение;
- привлекает внимание читателя к наиболее важной информации;
- хорошо оформлена и является привлекательной;
- подходит для формата выходных данных и вашей аудитории;
- может публиковаться без дополнительных пояснений;
- является доступной для дальтоников.

5.3 В каких случаях составление карты является нецелесообразным

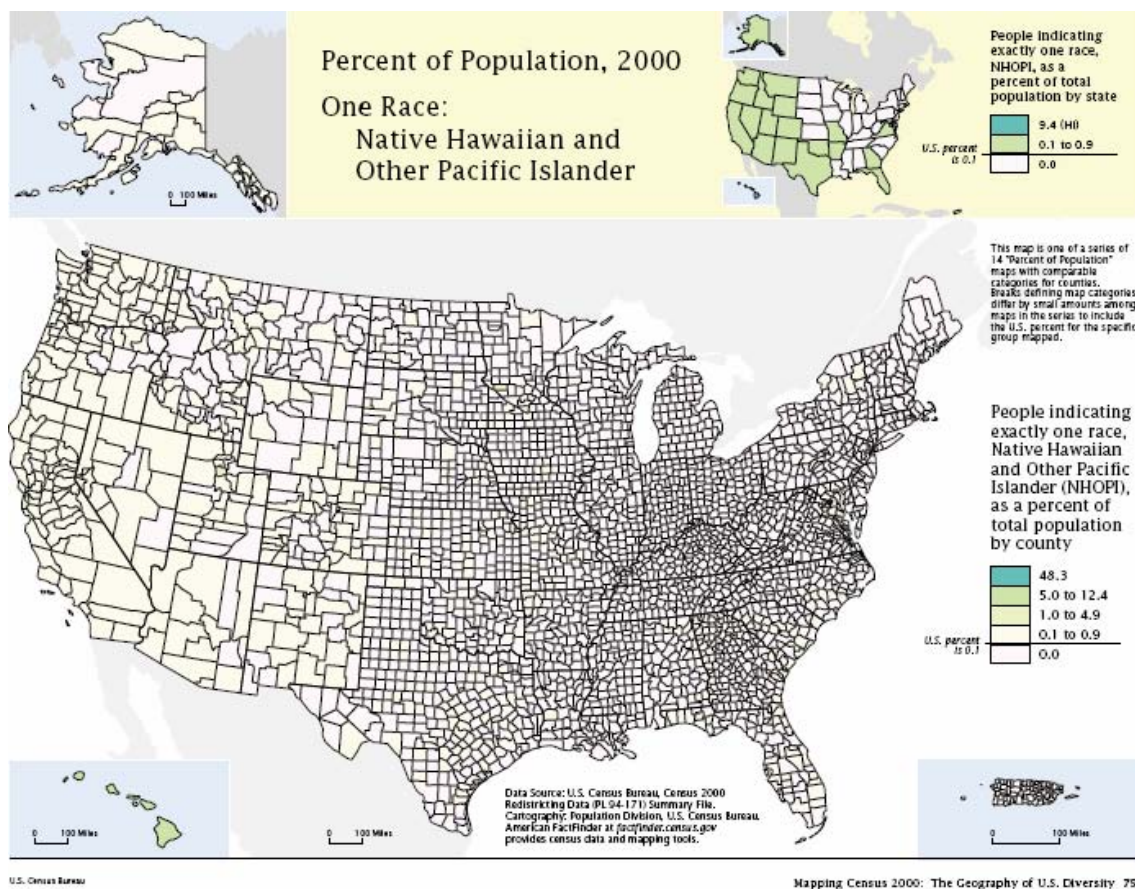
До того как приступить к составлению карты, вы должны рассмотреть вопрос о том, что является наиболее подходящим инструментом для представления ваших данных. Не тратьте ваше время и усилия, если график или таблица данных являются более эффективным способом доведения вашей мысли до читателя.

Не стоит заниматься картированием ваших данных, если:

- данные не имеют географической разбивки;
- данные не характеризуются значительной вариацией;
- ваша целевая аудитория может испытывать трудности с пониманием вашей карты;
- у вас не имеется достаточно места для включения карты, с тем чтобы она могла быть надлежащим образом прочитана и понята.

Нижеприведенный пример служит иллюстрацией того, каким образом усилия на подготовку карты могут не оправдывать себя по той причине, что отсутствует достаточная информация для картирования. Целью данной карты является представление пространственного распределения малой этнической группы в Соединенных Штатах Америки: выходцев с Гавайев и других тихоокеанских островов. Карта выглядит практически пустой, поскольку эта группа составляет менее 1% от общей численности населения в большинстве округов. Те немногочисленные округа, в которых данная этническая группа составляет более 1% населения, найти нелегко.

ПЛОХОЙ ПРИМЕР карты



Источник: Brewer, C.A. and Suchan, T.A., U.S. Census Bureau (2001), *Mapping Census 2000: The Geography of U.S. Diversity*, U.S. Government Printing Office, Washington DC¹².

5.4 Различные типы карт

Карты могут быть классифицированы по масштабу, функциям, дизайну, производственной технологии или же способу их использования в публикации.

Как правило, выделяется два типа карт:

- **Общие (топографические) карты** используются для обеспечения ориентиров в пространстве и указания местоположения набора различных объектов, таких как озера, реки, горы, прибрежная линия, дороги и т.д. Они помогают пользователям определять границы географических районов.
- **Тематические (статистические) карты** используются для изображения пространственного распределения одного или более статистических признаков. Тематическая карта всегда разрабатывается с конкретной целью и призвана отвечать на конкретные вопросы в отношении политических, общественных, культурных, экономических, сельскохозяйственных и природных явлений.

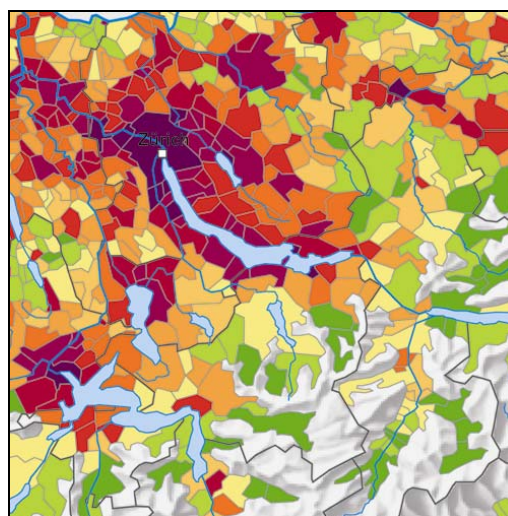
¹² <http://www.census.gov/population/www/cen2000/atlas/index.html>

Пример общей карты



Источник: swisstopo, the Federal Geo-Information center¹³

Пример тематической карты



Источник: Swiss Federal Statistical Office¹⁴

Статические или интерактивные карты

Карты могут быть статическими или интерактивными. Статические карты могут редактироваться пользователем. Интерактивные карты обеспечивают гибкость и дают пользователю возможность менять формат, отбирать и получать данные, анимировать карту и менять тему или акцент на аспектах, представляющих ключевой интерес.

Карты могут опираться на изображение или данные. Если карта опирается только на изображение, она является заранее отформатированной и статической. Когда вы разрабатываете карту на основе данных, вы храните всю информацию (данные, метаданные), которая вам необходима для создания карты в базе данных. Сама по себе карта воспроизводится только по запросу пользователя в Интернет. Благодаря данному подходу пользователь может легко обновлять и менять данные и параметры карты без необходимости повторного ее создания.

Тематические атласы

В своей простейшей форме атлас представляет собой набор карт. Тематические атласы развернуто представляют статистику, которая сопровождается и обогащается ценной информацией в форме текста, графиков и таблиц. Большинство статистических организаций признают их потенциал с точки зрения представления данных и занимаются разработкой популярных атласов переписей или тематических атласов (например, по народонаселению, состоянию здоровья или экономике).

¹³ <http://www.swisstopo.admin.ch/internet/swisstopo/en/home.html>

¹⁴ http://www.bfs.admin.ch/bfs/portal/en/index/regionen/thematische_karten/maps.html

Технологии атласов значительно усовершенствовались в последнее десятилетие. Современные онлайн-атласы и информационные системы (АИС) позволяют пользователю получать данные "за картой", открывать районы, "составлять" свои собственные карты, интегрировать свои собственные данные и общаться с автором карты или учреждением, являющимся его издателем. "За кулисами" разработаны новые производственные процессы, которые облегчают интеграцию различных типов знаний, таких как картография, графический дизайн, анализ данных, написание статей и перевод.

ХОРОШИЙ ПРИМЕР онлайн-тематического атласа



Источник: International Monetary Fund, *IMF Data Mapper*¹⁵.

5.5 Выбор соответствующего типа карты

Рекомендации в отношении графиков также применимы и к картам: чрезвычайно важно знать, какой тип карты соответствует конкретному типу информации. Выбор соответствующего метода картирования зависит от характера данных. Данный раздел содержит руководящие принципы в отношении трех наиболее широко используемых типов тематических карт: картограмм, точечных карт и пропорциональных символьных карт.

¹⁵ <http://www.imf.org/external/datamapper/index.php>

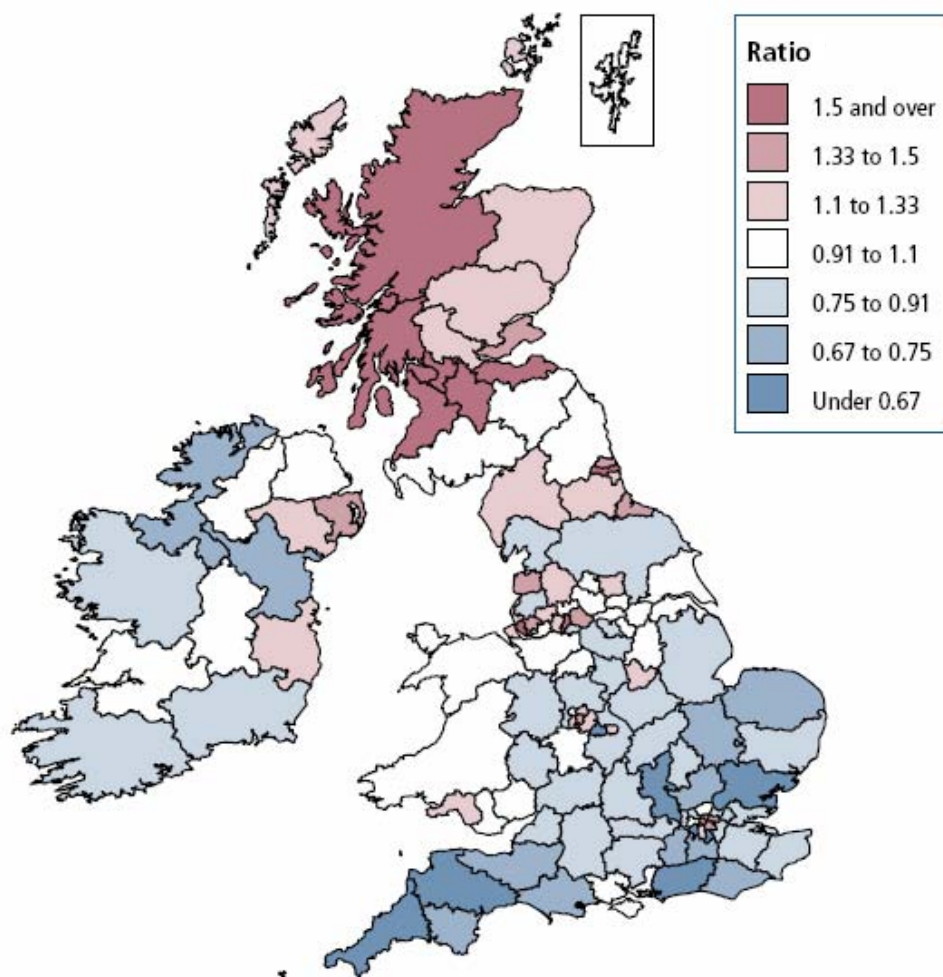
Картограммы

Наиболее широко используемым типом карт является картограмма, на которой территории затушевываются пропорционально величине иллюстрируемой переменной. Этот тип карт является удобным способом визуализации структур в пространстве, как это показывает нижеприведенный пример.

Изобилие картированных данных и удобство в работе с использованием ГИС, к сожалению, привели к частому неправильному использованию технологии картограмм. С помощью этого метода могут составляться только карты соотношений (т.е. соотношений, коэффициентов или плотности). Вы НЕ ДОЛЖНЫ использовать данный метод для представления абсолютных величин, таких как численность населения.

ХОРОШИЙ ПРИМЕР картограммы

Заблеваемость мужчин раком гортани в разбивке по санитарным округам, Соединенное Королевство и Ирландия, 1991-1999 годы



Источник: Office for National Statistics (2005), *Cancer Atlas of the United Kingdom and Ireland 1991-2000*, London¹⁶.

¹⁶ <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14059>

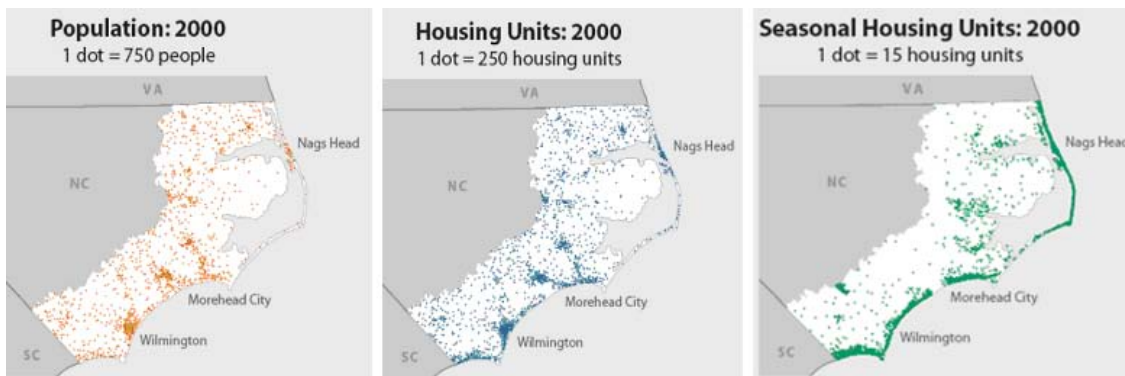
Картограмма опирается на данные, агрегированные по заранее определенным территориям, таким как санитарные округа Соединенного Королевства и Ирландии. Данный тип карты хорошо подходит для обсуждения вопросов здравоохранения применительно к данным санитарным округам. Однако реальные тренды зачастую не согласуются с заранее определенными территориями, что может привести к значительным ошибкам в интерпретации. Таким образом, картограммы следует предпочтительно использовать для изображения явлений, которые отличаются одинаковым распределением в рамках каждой пространственной единицы.

Классификация данных является одним из важных вопросов для составления картограмм. Пространственные тренды, изображаемые на карте, определяются группировкой величин данных. Вы можете разрабатывать различные карты с использованием одной и той же информации путем применения различных классификационных методов, разбивки данных по различным классам или путем выбора различных границ классов. Сегодня существуют многочисленные методы, но не правила, в связи с чем вы должны определить наиболее подходящий метод для классификации ваших данных.

Точечные карты

Точечные карты описывают местоположение и плотность населения или явления с использованием символов. Они позволяют пользователям быстро понять общий размах данных, а также их концентрацию или дисперсию. Каждая точка описывает дискретную величину, как правило, большое число объектов, как об этом свидетельствуют нижеприведенные примеры.

ХОРОШИЙ ПРИМЕР точечных карт



Источник: U.S. Census Bureau, *Census Data and Emergency Preparedness*¹⁷.

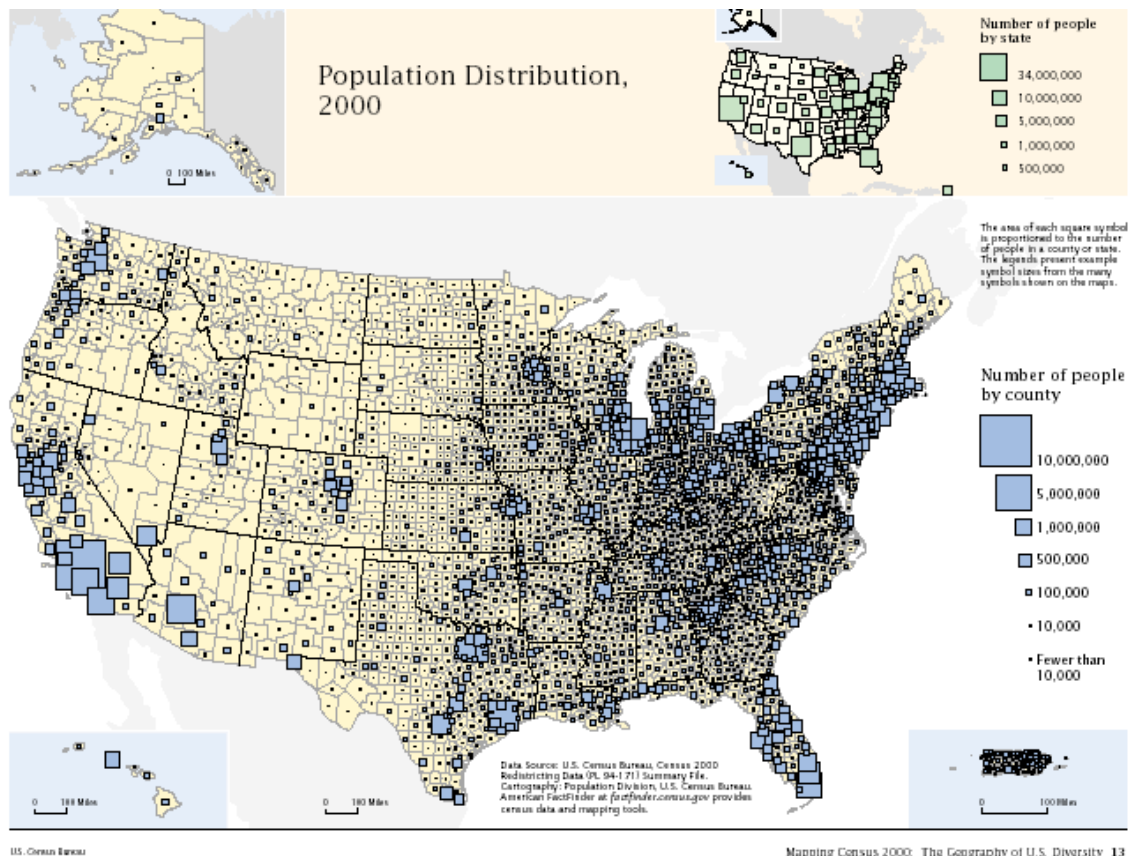
Пропорциональные символьные карты

Пропорциональная (или градуированная) символьная карта используется для изображения символьных величин. Размер символа пропорционален размеру описываемой совокупности или размаху явления. Каждый символ соответствует конкретной точке в рамках пространственной единицы, как правило, центру или столице.

¹⁷ <http://www.census.gov/Press-Release/www/emergencies/>

Наиболее широко используются круги, поскольку они являются компактными и удобными для масштабирования. Однако, как об этом свидетельствует нижеприведенный пример, могут также использоваться и другие геометрические формы, такие как квадраты или треугольники.

ХОРОШИЙ ПРИМЕР пропорциональной символьной карты



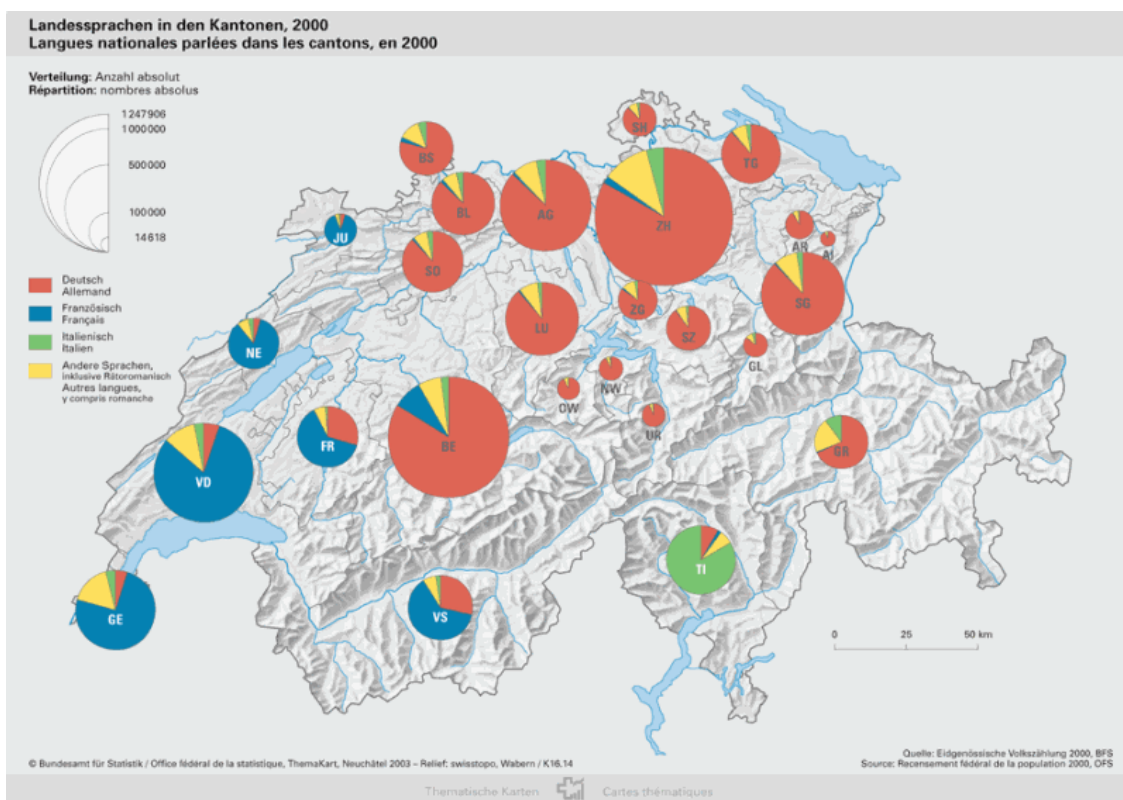
Источник: Brewer, C.A. and Suchan, T.A., U.S. Census Bureau (2001), *Mapping Census 2000: The Geography of U.S. Diversity*, U.S. Government Printing Office, Washington DC¹⁸.

Вышеприведенный пример свидетельствует о том, что для создания эффективной пропорциональной карты могут потребоваться специальные навыки картирования. Когда размер символа больше размера соответствующей пространственной единицы, выявление единицы, к которой относится символ, может быть сопряжено с трудностями. Данная трудность толкования еще более усугубляется наложением большого числа символов, иллюстрацией чего служит настоящая карта распределения населения в Соединенных Штатах Америки.

Методы составления графиков и картирования могут комбинироваться для изображения распределения различных категорий населения на одной и той же карте. В случае сложных символьных карт в качестве символов используются круговые диаграммы или гистограммы. На нижеприведенной карте Швейцарии круговые диаграммы описывают долю населения, говорящего на каждом из основных национальных языков (немецком, французском и итальянском) или на любом другом языке, в 26 кантонах. Размер символа описывает общую численность населения каждого кантона.

¹⁸ <http://www.census.gov/population/www/cen2000/atlas/index.html>.

ХОРОШИЙ ПРИМЕР сложной символьной карты



Источник: Swiss Federal Statistical Office, *Map Gallery Switzerland – Languages and religions*¹⁹.

Данная сложная методика должна использоваться с большой осторожностью, поскольку существует большой риск перегруженности карты и условных обозначений. Такие карты могут составляться только на уровне страны (например, стран Европы) и уровне регионов (например, швейцарских кантонов на вышеприведенной карте). Кроме того, на круговой диаграмме или гистограмме может быть показано только небольшое число категорий (не более пяти).

5.6 Рекомендации по разработке: стремитесь к простоте!

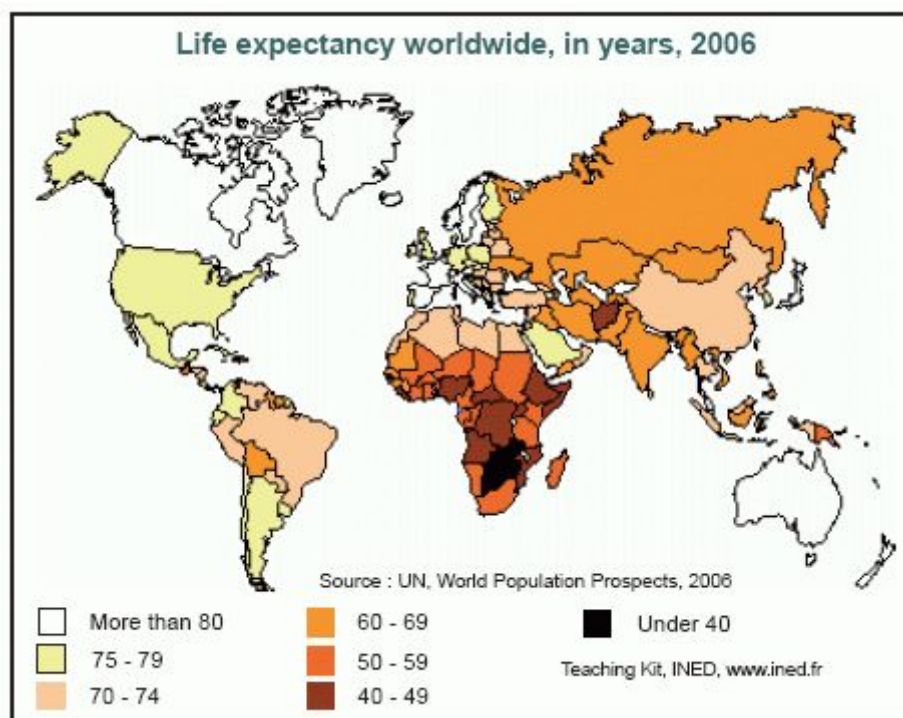
В отношении всех методов визуализации данных наиболее важной рекомендацией является обеспечение максимальной, по возможности, **простоты ваших карт!** Будьте осторожны, чтобы не увести в сторону или не запутать вашу аудиторию путем изображения избыточного числа информации или использования слишком большого числа визуально вступающих друг с другом в конфликт элементов.

Знание своей аудитории имеет ключевое значение. Каковы ее квалификация и знания? Насколько часто она использует карты в своей повседневной жизни? И не может ли ваша карта вызвать у нее раздражение? Вы должны знать больные места вашей целевой аудитории. Карта обладает мощным визуальным воздействием, и некоторые цвета или символы могут иметь отрицательную коннотацию для определенных групп лиц.

¹⁹ http://www.bfs.admin.ch/bfs/portal/en/index/regionen/thematische_karten/maps.html

Вы всегда должны составлять ваши карты таким образом, чтобы они были **автономны по отношению к тексту или таблицам данных**. Так же как и график, карта должна восприниматься автономно без дополнительных ссылок на сопроводительный текст или примечания. После опубликования карта должна поддаваться сканированию или загрузке и использованию вне контекста. Таким образом чрезвычайно важно использовать различные ниже описываемые компоненты карты.

ХОРОШИЙ ПРИМЕР
карты, которая не требует дополнительных пояснений



Источник: Institut National d'Etudes Démographiques (INED).

Компоненты карты

Компоненты карты конкурируют друг с другом в борьбе за внимание читателя. С целью достижения максимального эффекта вашего сообщения необходимо обеспечить, чтобы ваши данные находились в центре макета карты, особенно когда на ней представлена также другая информация (вода, высоты и т.д.). Карта должна занимать 80–85% общего пространства презентации.

Для оказания помощи пользователю в понимании карты необходимы следующие компоненты:

- **Название карты** должно четко формулировать идею, которой посвящена карта. Оно должно быть кратким и сжатым. Для представления более подробной информации могут быть добавлены подзаголовки (например, единицы измерения).
- **Условные обозначения** должны разъяснять все символы, структуры и цвета, используемые для представления данных на карте.
- **Географические единицы**, используемые для представления данных на карте, должны указываться либо в заголовке (или подзаголовке), либо в условных обозначениях.

- Для указания важных или значимых мест или другой информации могут быть добавлены **текстовые метки**.
- С целью оказания помощи пользователям в измерении расстояний и сопоставлении различных карт может быть указан **масштаб карты**.
- Для представления определений и другой методологической информации могут использоваться **сноски**.
- Внизу карты должен указываться **источник данных**.
- Внизу карты в **информации об авторских правах** должен указываться автор, ответственный за ее содержание.

В некоторые карты могут включаться другие компоненты, однако они не носят столь важного характера:

- **Стрелка, показывающая направление истинного меридиана**, необходима в тех случаях, когда карта не ориентирована по истинному меридиану.
- **Широта и долгота** должны указываться только на картах мира или континентов.
- **Обзорная карта расположения объектов** является малой копией базовой карты, которая показывает картированную область в более широком контексте. Она может быть полезна в том случае, если ваша аудитория не знакома с географией района.
- Для улучшения понимания карты могут добавляться **графики**.

Используйте только ключевые слова в названии и условных обозначениях

Хотя карты несут в себе визуальное сообщение, важное значение также имеет сопроводительный текст. Формулировка названия и любых условных обозначений требует к себе тщательного отношения, поскольку она определяет понимание и интерпретацию пользователем вашей карты. Ниже приводятся некоторые рекомендации по использованию текста на картах:

- Будьте точны, сохраняя при этом простоту.
- Используйте только ключевые слова и избегайте повторения одних и тех же слов в названии, условных обозначениях или сносках.
- Используйте нейтральные формулировки.
- Избегайте сокращений и акронимов.
- Используйте шрифт меньшего размера в отношении условных обозначений по сравнению с названием и еще меньший в отношении сносок, однако при этом весь текст должен оставаться удобно читаемым.

Тщательно подходите к определению условных обозначений

Формулировка условных обозначений должна обеспечивать правильное понимание и интерпретацию вашей карты. Каждый тип карты требует отличного типа условных обозначений. В то же время существуют некоторые базовые правила в отношении картограмм и пропорциональных символьных карт:

- Все границы интервалов должны носить недвусмысленный характер: избегайте интервалов, таких как 100-200, 200-300, 300-400.
- Не допускайте разрыва между интервалами: избегайте таких интервалов, как 1,0-1,5; 2,0-2,5; 3,0-3,5.
- Необходимо обозначать районы, в отношении которых отсутствуют данные.
- На карте, описывающей несколько переменных, условные обозначения должны приводиться в нисходящем порядке важности переменных.

Цвет: еще один важный компонент

Цвет является одним из наиболее мощных графических компонентов. Отнеситесь с большой тщательностью к выбору цветов для использования на вашей карте, поскольку они способны оказывать влияние на ее восприятие и вводить в заблуждение.

Выбор цвета на тематической карте зависит от данных и типа карты. Однако при этом следует учитывать три следующих аспекта. Во-первых, вы должны знать все существующие условности, связанные с определенным цветом, а также его возможную положительную или негативную коннотацию.

Затем вы должны удостовериться, что цвета, используемые на вашей карте, обеспечат понимание вашей мысли каждым читателем. Так, например, дальтоники испытывают трудности с проведением различия между определенными цветами. Наиболее широко распространенным случаем является красно-зеленый дальтонизм. Если вы будете использовать красный и зеленый цвета для проведения различия между двумя типами областей, таким как рост и снижение, то дальтоники не смогут провести такое различие. Существует легкое решение: эти люди увидят разницу, если вы замените красный цвет пурпурным.

И наконец, если вы используете относительно небольшое число категорий данных в отношении величин на непрерывной шкале (например, плотность населения), вы должны рассмотреть возможность использования различных оттенков одного и того же цвета, а не различных цветов. Если вы имеете данные в дискретных категориях или с положительными или отрицательными величинами, то более целесообразным является использование различных цветов.

6. Передовые методы визуализации

6.1 Почему визуализация означает нечто большее, чем картинка

Передовые инструменты и методы открывают новые возможности с точки зрения визуализации данных и привлечения к ним интереса пользователей. Составители динамических таблиц, графиков и карт позволяют пользователям манипулировать данными и составлять свои собственные графические изображения. Анимация и видео являются многообещающими форматами, в чем-то схожими с телевидением. Они являются эффективными средствами иллюстрации изменений во времени и содержат вербальные или контекстуальные описания для пояснения значений цифр. Предлагаются также и другие новые типы визуализации, такие как искрографики и облака тегов (см. раздел 6.4), служащие альтернативными методами иллюстрации информации.

Новейшие вебтехнологии и создаваемые ими ожидания, в том числе пользователей, заставляют статистические организации менять свои способы распространения данных. Интернет в настоящее время является двусторонним форумом, на котором пользователи могут обмениваться своими собственными графическими представлениями данных и обсуждать свои выводы. Примерами онлайн-сообществ, которые обсуждают данные и графики и обмениваются ими, являются такие вебсайты, как Many Eyes²⁰, Swivel²¹ и Data Place²².

Разработка новых методов визуализации и более интерактивных вебсайтов, хотя и ведет к повышению гибкости в работе с пользователями, может также создавать проблемы для статистических организаций. Непрерывно возрастает риск случайного или преднамеренного искажения или ошибочной интерпретации пользователями статистических данных с последующим распространением этих искажений и ошибочных толкований среди других пользователей. Исходя из этого, необходимо, чтобы статистические организации разработали четкую политику в отношении методики применения и предоставления новых методов визуализации.

Настоящая глава содержит обзор этих новых инструментов и методов визуализации.

6.2 Динамическая визуализация

С появлением Интернета и развитием технологий Web 2.0²³ пользователи могут интерактивно работать с данными, создавать свои собственные визуальные представления. Многие статистические организации в настоящее время предоставляют доступ к своим базам данных через свои вебсайты, позволяя

²⁰ Many Eyes является вебсайтом, на котором пользователи могут загружать свои данные, создавать графики и графические представления и обсуждать свои выводы (см. <http://manyeyes.alphaworks.ibm.com/manyeyes/>).

²¹ Swivel обладает функциями схожими с Many Eyes (см. www.swivel.com).

²² Data Place является вебсайтом, который предоставляет пользователям статистику о городах и штатах Соединенных Штатов (см. www.dataplace.org/).

²³ "Web 2.0" является термином, описывающим новую волну Интернет-технологий, которые позволяют пользователям выйти за рамки простого доступа к информации в режиме онлайн. Благодаря им пользователи могут добавлять, менять вебконтент или влиять на него. К числу примеров относятся вики, такие как Википедия, блоги и сайты социальных сетей, таких как Facebook или LinkedIn.

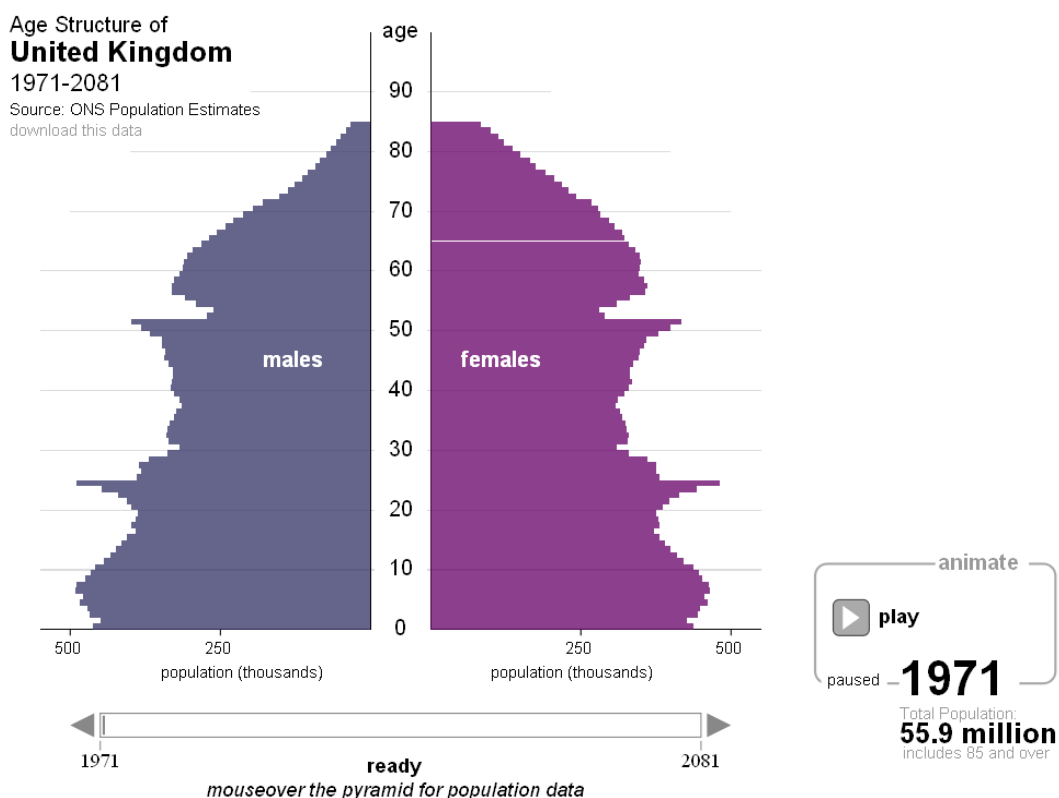
пользователям самостоятельно искать и загружать статистическую информацию. Данные функции во все большей степени дополняются набором инструментов визуализации, которые позволяют пользователям создавать таблицы, графики или карты в режиме онлайн без необходимости загрузки данных и работы в другом приложении.

Существуют определенные опасения в отношении последствий предоставления такого уровня свободы пользователям. Существует риск создания ими бессмысленных графиков или разработки ненадлежащих корреляций. Однако преимущества предоставления пользователям доступа и возможностей работы с данными не подлежат сомнению. Потенциальные проблемы могут быть минимизированы за счет предоставления ключевых метаданных четким и очевидным образом, оказания поддержки менее опытным пользователям и мониторинга и корректировки любых случаев ненадлежащего использования.

6.3 Анимация и видео

Анимация и видео являются двумя новыми важными методами визуализации данных. С учетом популярности телевидения и кино неудивительно, что пользователям нравится идея получения сообщений в формате видеозаписи. Такой формат облегчает изложение идеи путем комбинирования аудио- и текстовых описаний с графическими иллюстрациями для разъяснения смысла цифр.

ХОРОШИЙ ПРИМЕР использования анимации для сообщения статистических данных



Источник: Управление национальной статистикой Соединенного Королевства²⁴.

²⁴ http://www.statistics.gov.uk/populationestimates/svg_pyramid/uk/index.html

Динамические демографические пирамиды, разработанные рядом статистических организаций, в том числе Управлением национальной статистики Соединенного Королевства и Статистическим управлением Канады, являются хорошими примерами комбинирования анимации с интерактивностью в рамках простого интерфейса. Пользователи могут включать щелчком мыши видеозапись, иллюстрирующую изменение демографической пирамиды во времени. Они могут взаимодействовать с графиком путем выбора возрастных групп и подробных цифр и доли в общей численности населения.

Ханс Рослинг, соучредитель Gapminder²⁵, добился большого успеха в использовании анимации для иллюстрации данных в качестве одного из способов распространения статистики. Рослинг приобрел массовую аудиторию путем видеовещания в режиме онлайн, которое становится все более популярной Интернет-функцией. Его доклад на Конференции ТРД²⁶ в 2006 году был просмотрен в режиме онлайн и загружен тысячи раз. В основе популярности этой технологии лежит использование для распространения статистики видеороликов, которые представляют собой короткие видеолекции по таким вопросам, как материнская смертность, глобализация, энергетика и человеческое развитие.

ХОРОШИЙ ПРИМЕР комбинирования анимации и новых вебтехнологий для распространения статистики



Источник: Gapminder²⁷.

²⁵ <http://www.gapminder.org/>

²⁶ ТРД означает: технология, развлечения, дизайн. Речь идет о ежегодной конференции, на которой ведущие специалисты обсуждают текущие проблемы в этих областях. Доклады публикуются на ее вебсайте www.ted.com.





²⁷ <http://www.gapminder.org/video/gap-cast/>

6.4 Web 2.0 и создание сообществ вокруг данных

Такие вебсайты, как Many Eyes, Swivel и Data Place, добавляют новое измерение к визуальным презентациям путем создания онлайн-сообществ вокруг визуализации и обмена данными. Эти вебсайты позволяют пользователям загружать наборы данных и создавать графики для обмена и обсуждения с другими пользователями. Другие приложения, часто называемые "мэшапами" ("mashups"), комбинируют данные или функции из двух или более источников для создания нового сервиса. Примером "мэшапа" служит комбинация статистических и картографических данных для создания географического представления данных²⁸.

Некоторые разработчики официальной статистики проводят эксперименты с целью изучения потенциала этих вебсервисов для общения с более широкой аудиторией. Например, ЕЭК ООН загрузила четыре набора данных в Swivel, что имело пока неоднозначные результаты. В качестве позитивного аспекта следует отметить, что каждый из этих наборов данных был просмотрен более 5 000 раз за последние 18 месяцев, что свидетельствует о том, что к ним обращались многочисленные пользователи. Однако лишь два пользователя оставили свои комментарии, а гиперссылка на вебсайт Статистической базы данных ЕЭК ООН была задействована лишь 10 раз за весь период, что не позволило получить какой-либо информации об этих новых пользователях.

ХОРОШИЙ ПРИМЕР использования новых вебсервисов для охвата более широкой аудитории

Data	Description	Created	Views
 GDP per capita at current prices and PPPs, US\$	Gross Domestic Product per Capita estimates are compiled according to the 1993 version of the System of National Accounts (SNA 1993) for all countries, except Turkey for which the 1968 SNA is still used. Common currency (US\$) estimates are computed...	about 1 year ago by UNECE	6735
 Male Life Expectancy at Birth by Country and Year	Life expectancy is the average number of years of life remaining to be lived by a female or male at birth, based on current age-specific mortality rates. Footnotes: Azerbaijan - 1980: data refer to 1981 Belarus - 1980: data refer to 1981 Bosnia and...	about 1 year ago by UNECE	4358
 Female Life Expectancy at Birth by Country and Year	Life expectancy is the average number of years of life remaining to be lived by a female or male at birth, based on current age-specific mortality rates. Footnotes: Azerbaijan - 1980: data refer to 1981 Belarus - 1980: data refer to 1981 Bosnia...	about 1 year ago by UNECE	4271
 Industrial Production Index (2005 = 100) by Year ** Updated**	An industrial production index is an index covering production in mining, manufacturing and public utilities (electricity, gas and water), but excluding construction. The exact coverage, the weighting system and the methods of calculation vary from...	about 1 year ago by UNECE	3708

Источник: Swivel²⁹.

²⁸ Например, последняя версия программы статистического распространения PC-Axis suite (<http://www.pc-axis.scb.se/>) обеспечивает возможность комбинирования данных из Google Maps и Google Earth.

²⁹ <http://www.swivel.com/users/show/1005968>

Хотя достигнутый на настоящий момент успех выглядит неоднозначно, создание такого типа онлайн-сообщества, несомненно, служит относительно легким способом привлечения новых пользователей и, следовательно, новой областью визуализации данных, заслуживающей внимания.³⁰

6.5 Другие новые методы визуализации

Искрографики

Искрографики представляют собой малые, длиной в слово линейные графики, которые описывают тренды во времени. Они обладают преимуществом визуального изображения большого объема информации и могут размещаться вместе со словами, поясняющими их значение.

Искрографики придумал Тафти (2006 год). На нижеследующем примере показаны искрографики, используемые для иллюстрации колебаний обменного курса евро по отношению к другим валютам. Эти "интенсивные, простые, размером в слово графики" усиливают табличные данные графическим изображением, не занимая много места.

ХОРОШИЙ ПРИМЕР искрографика



Источник: Tufte, E.R. (2006), *Beautiful Evidence*, Cheshire CT, Graphics Press.

Облака тегов

Облако тегов (в некоторых случаях называемые также облаком ключевых слов) служит визуальным изображением частоты использования слова или тега в конкретном тексте или наборе данных. Меняющиеся размеры и цвета иллюстрируют популярность или важность тегов. Их можно часто увидеть на вебсайтах в виде перечня категорий, каждое слово которых обеспечивает переход по гиперссылке с помощью щелчка мыши для получения дополнительной информации по соответствующему слову.

Облака тегов являются полезным способом выявления общих терминов в тексте и построения классификаций ключевых слов. Нижеприведенный пример был создан с использованием текста настоящей главы (на английском языке) и четко иллюстрирует ключевые слова.

³⁰ Дополнительную информацию см. в специальном выпуске *Статистического журнала МАОС*, том 25, № 3-4, 2008 год: "Web 2.0 and Official Statistics", размещенном по адресу: <http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>

ХОРОШИЙ ПРИМЕР облака тегов



Tag Cloud Builder
a project of OCLC Research beta v1.0

Tag Cloud

2006 allow animation clouds **communicate** create **data** development discuss download
dynamic emerging examples gapminder good graph http illustrating information interactive map
meaning new numbers online organizations own place popularity providing sets share sites
source sparklines **statistical** such swivel **tag** techniques time uk **users** using
video **visualization** ways web words www

Создан с использованием Tag Cloud Builder³¹

³¹ <http://tagcloud.oclc.org/tagcloud/TagCloudDemo>

7. Вопросы доступности

Для того чтобы базовая статистическая информация широко и беспрепятственно использовалась, она должна быть универсально доступной. Это означает, что каждый должен иметь возможность просматривать и понимать информацию независимо от используемой технологии или своих физических ограничений.

Для привлечения и охвата большой аудитории информация должна представляться структурированным и хорошо организованным образом и соответствовать установленным и общепризнанным стандартам. В настоящей главе обсуждаются три частично пересекающихся аспекта доступности:

- предоставление надлежащих метаданных для оказания помощи пользователям в понимании данных;
- предоставление данных в различных форматах, в том числе с использованием новых средств, таких как мобильные телефоны или другие переносные устройства;
- обеспечение максимальной доступности информации для лиц с физическими недостатками, зачастую в соответствии с конкретными правовыми или политическими требованиями.

Для того чтобы статистическая информация стала доступна каждому, необходимы дополнительное информирование, работа и ресурсы. Однако принцип равного доступа к базовой статистической информации имеет важное значение. Хорошо продуманная стратегия доступности, несомненно, принесет выгоду всем.

В настоящей главе мы обсуждаем основные компоненты качественной информации, которые должны учитываться при распространении статистики.

7.1 Текст

Текст должен быть кратким, последовательным и хорошо структурированным, с тем чтобы пользователи могли легко находить искомую информацию. Изложение вашего текста в виде четко различимых и подчиненных логике разделов с заголовками и подзаголовками облегчит поиск информации и ее преобразование в другие форматы.

Для эффективного обслуживания всех целевых групп текст должен иметься в наличии в различных форматах, например на шрифте Брайля, в аудиоформате и в наборе сверхкрупным шрифтом. При создании оригинального текста вы должны учитывать возможное преобразование форматов в целях снижения их влияния. Например, включение текстового описания по каждому объекту графической информации облегчит преобразование на шрифт Брайля, в аудиоформат и даже HTML³² для размещения на вебсайте.

Правила доступности, разработанные Международным консорциумом Всемирной паутины (W3C)³³, были призваны обеспечить доступ к информации в электронном формате с использованием адаптивных технологий. К их числу относятся программы для чтения с экрана, системы укрупнения знаков и

³² Язык гипертекстовой разметки.

³³ <http://www.w3.org/>

переносные устройства доступа к данным (PDA), перечень которых может быть продолжен.

7.2 Таблицы

Вы также должны уделять внимание вопросам доступности при использовании таблиц для представления ваших данных. Если данные приводятся без информации, необходимой для их понимания и интерпретации, они будут непригодны для использования или будут вводить в заблуждение.

Рассмотрим, например, таблицу, занимающую несколько страниц. Если заголовки колонок или строк не будут повторяться на следующих страницах, пользователю будет тяжело читать таблицу. Важно повторять заголовки колонок и строк на каждой странице.

ХОРОШИЙ ПРИМЕР ТАБЛИЦЫ, удовлетворяющей требованиям доступности

Поголовье скота в Канаде, 2007-2008 годы

Тип скота	2007		2008	
	Тыс. голов	%	Тыс. голов	%
Крупный рогатый скот	15 885	50,2	15 195	52,0
Свины	14 690	46,4	12 985	44,4
Овцы	1 096	3,5	1 062	3,6
ИТОГО	31 671	100	29 242	100

Источник: Canadian Cattle Statistics August 2008³⁴.

Каждая величина увязана только с определенным типом скота и годом, но также описывает либо количество голов (в тыс.), либо %. Программа считывания с экрана будет способна озвучить всю эту информацию по каждой величине в таблице. Таким образом, пользователи услышат: "В 2007 году процент крупного рогатого скота составляет 50,2".

В электронном формате, таком как HTML, в отношении заголовков колонок и строк могут включаться теги наряду с кратким описанием их содержания, в случае необходимости. Использование тегов является полезным для всех. При загрузке таблиц заголовки колонок и строк будут переноситься напрямую во избежание путаницы. Кроме того, когда структура вашей таблицы является сложной, важно точно описать, каким образом организованы данные, с тем чтобы пользователи могли извлечь максимум информации с минимальными усилиями.

7.3 Графики

При подготовке графиков целесообразно учитывать, что не все пользователи имеют доступ к картинкам. Для представления информации, показанной на вашем графике, может быть создано текстовое описание. Ниже приводится пример из Руководства по магнитной записи Национальной ассоциации Брайля.

³⁴ <http://www.thedairysite.com/articles/1613/canadian-cattle-statistics-august-2008>

На вебсайте использование тега "alt text" (краткого текстового эквивалента) или признака "longdesc" (длинное описание, хранящееся на другой странице и доступное через адрес электронной почты) является подходом, признанным и описанным W3C. Эти описания могут также использоваться для подготовки документов в аудиоформате, на шрифте Брайля или в других форматах.

Вы может иметь такое большое число графиков, что вам будет трудно подготовить вручную их текстовые эквиваленты. В настоящее время реализуется проект по разработке инструментов, которые будут автоматически извлекать информацию из графика. Они также будут способны генерировать дескриптивную страницу одновременно с созданием графика.

ХОРОШИЙ ПРИМЕР графика с текстовым эквивалентом

Взаимосвязь между давлением пара воды и ее температурой



"Взаимосвязь между давлением пара воды и ее температурой". Речь идет о линейном графике, на котором по оси x откладывается температура в градусах Цельсия от нуля до 100 градусов. По оси y откладывается давление в миллиметрах ртутного столба от 0 до 800 миллиметров. Кривая начинается в точке пересечения и поднимается таким образом, что при $x = 25$ градусам, $y =$ примерно 40 миллиметрам. При $x = 50$, $y = 100$. При $x = 75$, $y =$ чуть менее 300. При $x = 100$, $y =$ около 760".

Источник: Accessible Digital Media: Design Guidelines for Electronic Publications, Multimedia and the Web³⁵.

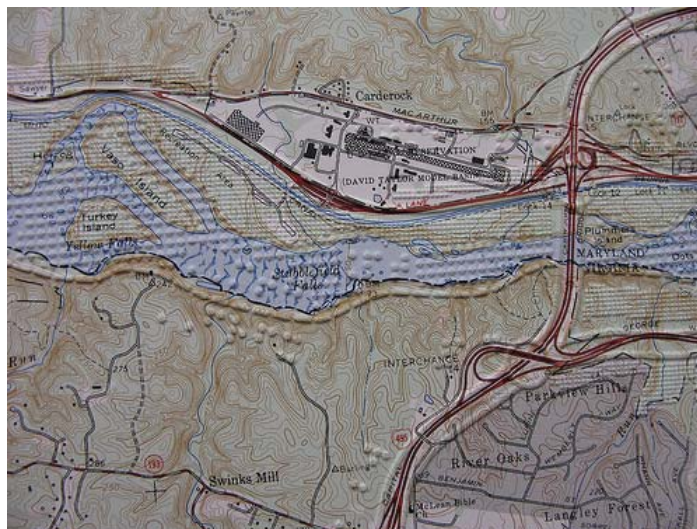
7.4 Карты

Карты создают крупные технические проблемы с точки зрения доступности для лиц с определенными физическими недостатками. Информация передается путем комбинации изображений и цветов, которые практически не совместимы со стандартами доступности. Исходя из этого мы должны предусмотреть возможность подготовки текста в качестве альтернативы представлению информации на карте и/или предоставления доступа к таблицам данных.

³⁵ http://ncam.wgbh.org/publications/adm/guideline_f.html

Еще одним вариантом является предоставление инструмента поиска, который позволит пользователям определять параметры, такие как район, город или улица. Поиск в базе данных может в некоторых случаях иметь следующую форму: "Искать число мужчин в возрасте от 15 до 49 лет, проживающих в конкретном районе и работающих в агропромышленном секторе". Результаты должны генерироваться в форме статистической таблицы, содержащей ту же самую информацию, которая приводится в графическом формате.

ХОРОШИЙ ПРИМЕР карты, доступной для слепых
Карта части Вашингтона (округ Колумбия) в формате Брайля



Источник: US Geological Survey Library, courtesy of Flickr³⁶.

7.5 Метаданные

Статистические организации должны обеспечивать пользователей метаданными, которые им необходимы для понимания данных, в том числе их сильных сторон и ограничений. Эти метаданные должны постоянно обновляться путем включения последних изменений в определения, классификации и методологию.

Следует учитывать следующие рекомендации³⁷ для обеспечения того, чтобы ваши данные были доступны и могли использоваться каждым:

- предоставление доступа к метаданным с использованием различных средств, таких как бумажные публикации, КД-ПЗУ и т.д. Все метаданные должны размещаться в Интернете, поскольку Сеть является наиболее доступной средой для пользователей с точки зрения нахождения большинства обновленных метаданных;
- предоставление метаданных таким образом, который обеспечивал бы удовлетворение потребностей широкого диапазона пользователей с различными требованиями и/или статистическими знаниями;

³⁶ <http://www.flickr.com/photos/98169608@N00/3296197787/>

³⁷ Руководящие принципы по предоставлению и распространению метаданных являются адаптированными принципами, содержащимися в публикации ОЭСР 2007 года "Data and Metadata Reporting and Presentation Handbook", размещенной по адресу <http://www.oecd.org/bookshop?9789264030329>.

- распространение метаданных на бесплатной основе через Интернет, даже если бумажные версии или же описываемая ими статистика предоставляются за плату;
- обеспечение активной увязки метаданных с таблицами и графиками, которые они описывают, и наоборот;
- предоставление метаданных не только на национальном языке, но также, по мере возможности, на одном из широко используемых языков, примером которых может служить английский;
- создание механизма локального поиска на основе свободного текстового поиска;
- создание механизма для обеспечения стабильности URL (унифицированных указателей ресурсов) или обеспечение связи между старыми и новыми URL, которые будут перенаправлять пользователей на новый адрес. Это является одним из ключевых вопросов с учетом важности связей между вебсайтами;
- указание имен контактных лиц или электронных адресов для запроса дополнительной информации о концепциях, определениях и статистических методологиях. В некоторых организациях "контактом" может являться общий корпоративный контакт или справочная служба для всех запросов клиентов.

Некоторые статистические организации представляют метаданные в виде слоев пирамиды. При применении такого подхода детализация метаданных возрастает по мере продвижения пользователя сверху вниз:

- **На верхушке пирамиды:** метаданные, которые имеют основное значение для базового понимания статистики, включая информацию о статусе данных (предварительные или окончательные). Случайные пользователи в большинстве случаев ограничатся этим типом метаданных.
- **Ниже:** пояснительные примечания, содержащие краткие описания статистики (определения, ключевые вопросы, ограничения и т.д.), которые способны оказывать влияние на использование данных. К этому типу метаданных могут обращаться промежуточные пользователи.
- **В основании пирамиды:** самая подробная методологическая информация обычно в форме методологических руководств или схожих документов. Этот тип метаданных, как представляется, будет представлять интерес только для наиболее продвинутых пользователей и тех, кто использует данные для подробных исследований.

8. Справочная и дополнительная литература

Bertin, J. (1981), *Graphics and Graphic Information-processing*, New York, de Gruyter.

Bertin, J. (1983), *Semiology of Graphics: Diagrams Networks Maps*, Madison Wis., University of Wisconsin Press.

Blessing, C., Bradsher-Fredrick, H., Miller, H., Miller, R. and Rutchik, R. (2003), *Cognitive Testing of Statistical Graphs: Methodology and Results*, Washington D.C., U.S. Energy Information Administration.

Bosch ten, O. and Jonge de, E. (2008), "Visualising official statistics", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 103-116, available at:
<http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Cleveland, W.S. and McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods", in *Journal of the American Statistical Association*, 79, pp. 531-554, available at:
<https://secure.cs.uvic.ca/twiki/pub/Research/Chisel/ComputationalAestheticsProject/cleveland.pdf>.

Cleveland, W.S. and McGill, R. (1987) "Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs (with Discussion)", in *Journal of the Royal Statistical Society Series A*, 150, pp. 192-229, available at:
http://www.wjh.harvard.edu/~kwn/Kosslyn_pdfs/1987Cave_JRoyStatSocA_CommentaryClevelandArticle.pdf.

Few, S. (2004), *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Oakland CA, Analytics Press.

Gardner, J. (2008), "Blogs, wikis and official statistics: New perspectives on the use of Web 2.0 by statistical offices", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 81-92, available at:
<http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Harris, R.L. (2000), *Information Graphics*, New York and Oxford, Oxford University Press.

Kennedy, D. (2007), *Research Paper: Data Visualization*, Canberra, Australian Bureau of Statistics, available at:
[http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/7F8E375FC22D26A5CA25731C0022DF1B/\\$File/1211055001_jul%202007.pdf](http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/7F8E375FC22D26A5CA25731C0022DF1B/$File/1211055001_jul%202007.pdf).

Miller, J.E. (2004), *The Chicago Guide to Writing About Numbers*, Chicago, University of Chicago Press.

Playfair, W. (1786), *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Exports, Imports, and General Trade of England, at a Single View*, London.

Playfair, W. (1801), *Statistical Breviary: Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe*, London, Wallis.

Robbins, N.B. (2005), *Creating More Effective Graphs*, Hoboken NJ, John Wiley & Sons.

Schulz, T. (2009), *Guidelines on the Presentation of Statistical Maps*, UNECE Work Session on the Communication and Dissemination of Statistics, Warsaw, available at: <http://www.unece.org/stats/documents/ece/ces/ge.45/2009/crp.1.e.pdf>.

Smith, A. and Rogers, S. (2008), "Web 2.0 and official statistics: The case for a multi-disciplinary approach", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 117-123, available at: <http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Tufte, E.R. (2001), *The Visual Display of Quantitative Information (2nd ed.)*, Cheshire CT, Graphics Press.

Tufte, E.R. (2006), *Beautiful Evidence*, Cheshire CT, Graphics Press.

Wainer, H. (1984), "How to Display Data Badly?", in *The American Statistician*, vol. 38, no. 2, pp. 137-147, available at: http://www.soc.washington.edu/users/bp Pettit/soc504/wainer_display.pdf.

Ware, C. (2004), *Information Visualization: Perception for Design*, San Francisco CA, Morgan Kaufmann Publishers.

Дополнительная полезная информация и дискуссии по вопросам представления статистики могут быть найдены по адресу: <http://blogstats.wordpress.com/>.

ПРИДАТЬ ЗНАЧИМОСТЬ СТАТИСТИЧЕСКИМ ДАННЫМ

Руководство по представлению
статистических данных

Картинка стоит тысячи слов. Тренды в данных зачастую более очевидны, когда цифры представлены в виде графики. Существуют многочисленные способы представления данных, начиная с простых гистограмм, кончая более сложными диаграммами рассеяния, тематическими картами и анимированными графиками.

Настоящее руководство служит практическим инструментом, призванным оказать помощь разработчикам статистических данных в представлении данных четким и значимым образом. Оно содержит рекомендации по подготовке эффективных таблиц, графиков и карт, использованию других форм визуализации с целью "оживления" сухой статистики. В нем также предлагаются решения, помогающие избежать подготовки низкокачественных или вводящих в заблуждение визуальных материалов. Ясные визуальные представления данных, подготовленные с учетом целевой аудитории, будут содействовать расширению использования статистических данных и раскрытию содержащейся в них ценной информации.

06 2007

