

## Кирилл Еременко. Работа с данными в любой сфере

Что общего у аналитика данных и Шерлока Холмса? Как у Netflix получилось создать 100%-ный хит — сериал «Карточный домик»? Ответ кроется в правильном использовании данных. Эта книга — практическое руководство и увлекательное путешествие в науку о данных, независимо от того, хотите ли вы использовать анализ данных в своей профессии, собираетесь ли стать аналитиком данных, или уже работаете в этой области. Эту книгу хорошо дополняет [Джон Форман. Много цифр: Анализ больших данных при помощи Excel](#). продвинутый

Кирилл Еременко. Работа с данными в любой сфере.<sup>1</sup> — М.: Альпина Пабlishер, 2019. — 304 с.



Купить книгу в издательстве [Альпина Пабlishер](#), цифровую книгу в [ЛитРес](#), бумажную книгу в [Ozon](#)

### Глава 1. Определение данных

*Данные* — это любая единица информации. Это побочный продукт любых действий, пронизывающих каждую часть нашей жизни, причем не только в сфере Интернета. Данные — это прошлое, а прошлое — это данные. И аналитики данных могут использовать их для лучшего понимания наших нынешних и будущих действий. Они применяют тот же принцип, что веками использовали историки: мы можем учиться на опыте истории. Мы можем учиться на наших успехах — и на наших ошибках, чтобы улучшить настоящее и будущее. Единственный аспект данных, который в последние годы резко изменился, — наша способность собирать, организовывать, анализировать и визуализировать их в любых контекстах.

*Большие данные* — это название, присвоенное массивам данных со столбцами и строками, которых настолько много, что они не могут быть обработаны обычным аппаратным и программным обеспечением в течение разумного промежутка времени. Ранее данные хранились на бумаге, в архивах, библиотеках и хранилищах. Теперь почти все новые данные, которые мы собираем, хранятся в цифровом формате.

Когда новые технологии облегчили хранение данных, исследователи начали обращать внимание на то, как эти сохраненные данные могут быть использованы на практике. Так мы начали создавать [порядок из хаоса](#).

Сериал «Карточный домик» компании Netflix, впервые доказал индустрии, насколько сильны могут быть данные в управлении производством контента. В преддверии создания «Карточного домика» Netflix собирала данные о своих пользователях. Эти данные были использованы

---

<sup>1</sup> Удивительно, насколько часто издательства при переводе книг изменяют названия в маркетинговых целях. Не берусь судить, насколько лучше будет продаваться книга с выбранным названием, но читая книгу, становится очевидным, что английское название намного лучше отражает содержание: *Confident data skills. Master the fundamentals of working with data and supercharge your career* — что-то типа, Уверенные навыки по работе с данными. Овладейте основами работы с данными и перегрузите свою карьеру.

компанией, чтобы представить нужный фильм нужной аудитории. Исследователи данных в Netflix выявили значительное число подписчиков Netflix, которые наслаждались и работой Кевина Спейси, и серьезными политическими драмами.

### Глава 3. Мышление, необходимое для эффективного анализа данных

В Интернете есть много простых генераторов облака слов, где вы можете поиграть со шрифтами, макетами и цветовыми схемами (облако слов на рис. 1 было сгенерировано на основе анализа заголовков заметок в блоге baguzin.ru с использованием <https://www.wordclouds.com/>). Они выглядят гораздо более привлекательно, чем упорядоченные списки. Обратитесь к ним при подготовке своей следующей презентации; вы удивитесь тому, как легко окажется запустить дискуссию.



Рис. 1. Облако тегов, созданное на основе названия заметок блога baguzin.ru

Новички в какой-либо дисциплине склонны месяц за месяцем изучать теорию вместо того, чтобы настроить свое мышление на применение полученных знаний на практике. В качестве упражнения просто введите в поисковике слова «бесплатные массивы данных»/«free datasets» — и найдете множество сайтов, которые позволяют скачать их CSV-файлы (файлы для хранения табличных данных), готовые для анализа. Мы также предлагаем множество бесплатных массивов данных на <https://www.superdatascience.com/>.

У технологической исследовательской фирмы Gartner есть модель для разделения науки о данных на четыре типа, и, если бизнес-аналитика соответствует первому типу анализа, наука о данных может помочь поставить галочки для трех остальных (рис. 2).

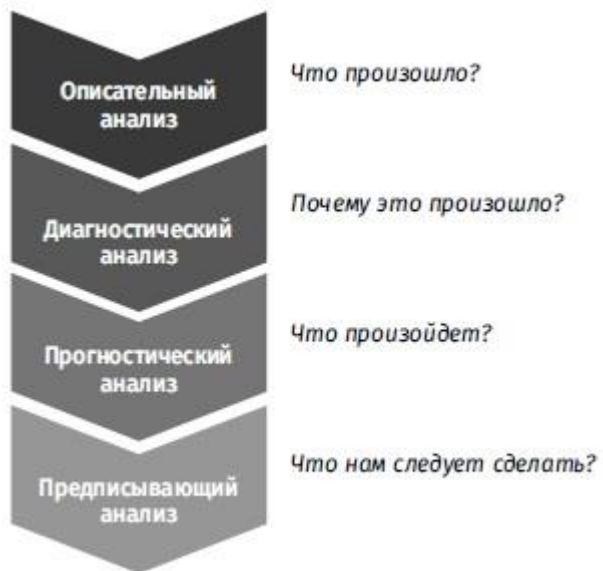


Рис. 2. Шкала аналитических значений

Цепляться за бизнес-анализ может быть заманчиво, когда вы или ваша компания работали таким образом в течение многих лет, но наука о данных предлагает гораздо более впечатляющий набор инструментов. К сожалению, традиционная зависимость бизнес-анализа от Excel может научить вас плохим привычкам. Будьте готовы использовать программу, которая не является Excel. На мой взгляд, одни из лучших программ для анализа массивов данных — R и Python.





Рис. 3. Процесс анализа и обработки данных

#### Глава 4. Сформулируйте вопрос

Глубинный анализ (майнинг) данных — возможно, самая приятная для меня часть процесса в любом проекте. На этом экспериментальном этапе мне нравится помещать данные в <https://www.tableau.com/>, которое умеет их читать и поможет вам создать предварительные наглядные визуализации, такие как легко читаемые таблицы, диаграммы и графики. Это обеспечивает прекрасный задел, который вы можете использовать в качестве фокусирующей линзы, чтобы сформулировать нужные вопросы.

С помощью количественных методов собирают числовую информацию, в то время как посредством качественных — нечисловую. Гораздо проще собрать в массив данных, чем использовать качественные методы, но наше окончательное решение о том, какие данные собирать, не должно основываться на принципе «простота ради простоты». Мы не можем, например, предпочесть качественный подход к сбору сведений о возрасте, потому что возраст — это факт, а не мнение (что бы вы себе ни говорили). Мы будем использовать количественные методы, если нужно подсчитать элементы, или измерить изменения в заработной плате, или узнать больше о демографии потребителей.

Качественные методы связаны с открытыми вопросами, имеющими бесконечное количество ответов. По своей природе они носят исследовательский характер и помогают выявить — но не количественно — тенденции во мнениях, мыслях и чувствах. Мы можем применять этот подход, когда нужно больше контекста, чтобы понять проблему, или когда проблема слишком сложна, чтобы решить ее количественным методом.

Единственное, на чем я настаиваю в начале любого проекта, связанного с наукой о данных, — это письменное подтверждение согласия заинтересованных лиц. По моему опыту участники, в каком бы качестве они ни выступали, в ходе реализации проекта склонны менять свое представление о том, чего они хотят. Такое поведение понятно в случае, когда сам проект имеет расплывчатый характер, но оно способствует разрастанию масштаба, которое может либо вывести вас за рамки исходных параметров, либо полностью убить проект. Поэтому, прежде чем перейти к подготовке данных, получите письменное подтверждение согласия.

Подготовьте людей, с которыми имеете дело, к возможности того, что результаты могут быть не такими, как им хочется. Поясните заранее, что вы не знаете, каковы будут результаты. Заказчикам могут не понравиться итоги. Сразу же дайте понять, что вы добываете факты, а не стремитесь польстить, и надеетесь, что они не будут «убивать гонца», если отчет окажется не особенно благоприятным.

## Глава 5. Подготовка данных

Для того чтобы сделать исходные (сырые) данные пригодными для анализа, их нужно сначала подготовить:

- 1) извлечь данные из исходных источников;
- 2) перевести данные на понятный язык, чтобы они стали доступны в реляционной базе данных;
- 3) загрузить данные в конечный источник.

Этот процесс известен как ETL (Extract — Transform — Load).

CSV-файлы — это самый простой тип необработанных файлов с данными, полностью лишенными какого-либо форматирования, что делает их доступными для различных программ, в которые мы можем их импортировать. В CSV-файлах строки размещаются на новых строках и столбцы разделяются запятыми в каждой строке. Отсюда и аббревиатура — comma separated values (данные, разделенные запятой).

Для серьезных проектов в области науки о данных работать с данными в их первоначальном хранилище запрещено. Иначе вы можете случайно изменить необработанные данные.

Notepad++ — инструмент, которым я пользуюсь, когда хочу посмотреть извлеченные мной данные. Это мощный редактор для просмотра CSV-файлов, и он гораздо удобнее, чем программа «Блокнот», которая стандартно поставляется с Windows (Notepad++ можно загрузить бесплатно: <https://notepad-plus-plus.org/downloads/>).

EditPad Lite — еще одна бесплатная программа для личного использования.

## Глава 6. Анализ данных (часть I)

Основные алгоритмы, используемые в анализе данных, можно разделить на три группы:

- 1) алгоритмы классификации;
- 2) алгоритмы кластеризации;
- 3) алгоритмы обучения с подкреплением.

Мы используем классификацию, когда уже знаем, в какие группы хотим объединить наши данные с помощью анализа, и мы используем кластеризацию, когда не знаем, что это будут за группы. Например, если бы мы хотели провести анализ ответов «да/нет» на вопрос, мы бы использовали алгоритм классификации. А вот если бы мы хотели оценить респондентов одного и того же опроса на основе их возраста и расстояния до ближайшего магазина нашей компании, то использовали кластеризацию.

### Алгоритмы классификации

*Дерево решений* можно визуализировать в виде блок-схемы. Листья этих деревьев дают нам все возможные ответы на все вопросы, которые мы можем задать нашим данным. В мире бизнеса деревья решений можно использовать для классификации групп клиентов. Например, команда разработчиков игр собрала информацию о потенциальном новом подписчике, о среднем времени игры геймеров и их возрасте. Мы создали точечную диаграмму (рис. 4) с большим количеством точек данных на основе возраста (X1) и времени, проведенного за игрой в часах (X2).

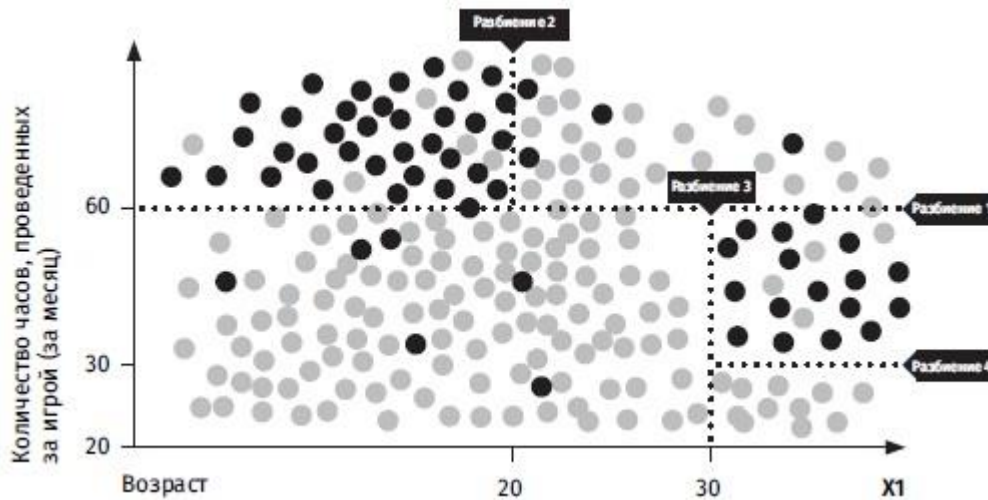


Рис. 4. Корреляционный график отображения разделенных на листья данных о клиентах

Серые точки обозначают пользователей, которые не стали подписчиками; черные — подписавшихся. Если бы мы запустили классификационный алгоритм дерева решений, точечная диаграмма была бы разбита на листья, как определено алгоритмом (рис. 5).

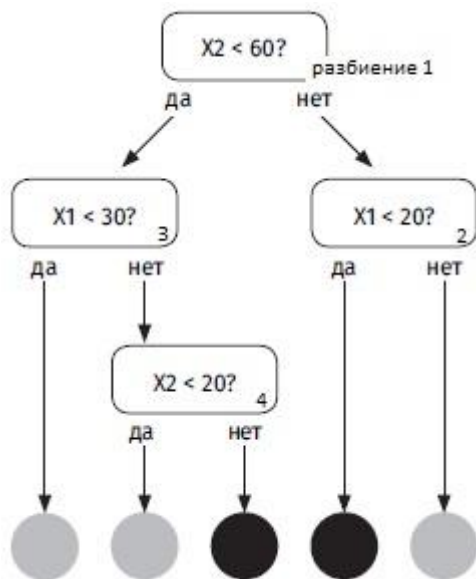


Рис. 5. Блок-схема построения дерева решений

Чем этот алгоритм полезен для нас? Предположим, что нашему новому клиенту 19 лет и за первый месяц он потратил 65 часов игрового времени ( $X_1 = 19$  и  $X_2 = 65$ ). Наш алгоритм уже обнаружил, что статистически значимое число людей в возрасте до 20 лет, которые играют более 60 часов, с высокой степенью вероятности станут постоянными пользователями, а это значит, что мы можем нацелить на этого клиента рекламу, которая побудит его или ее заплатить за подписку.

*Алгоритм случайного леса* вместо одного дерева использует много разных деревьев, чтобы сделать один и тот же прогноз, принимая среднее значение результатов отдельных деревьев. Мы можем применить силу толпы; когда деревья решений используются в совокупности, они могут давать гораздо более обоснованные прогнозы. Подумайте об этом как о процессе голосования — каждое решающее дерево проголосует (сделает прогноз), а затем алгоритм случайного леса возьмет вариант с наибольшим количеством голосов в качестве результата.

*Метод k-ближайших соседей* использует шаблоны в данных для отнесения новых точек данных в той или иной категориях. Для начала выберите число  $k$  — количество ближайших соседей. Измерьте расстояние между новой точкой данных и всеми существующими точками. Подсчитайте количество ближайших  $k$ -соседей, относящихся к различным категориям. Отнесите точку данных к категории с наибольшим количеством соседей (рис. 6).

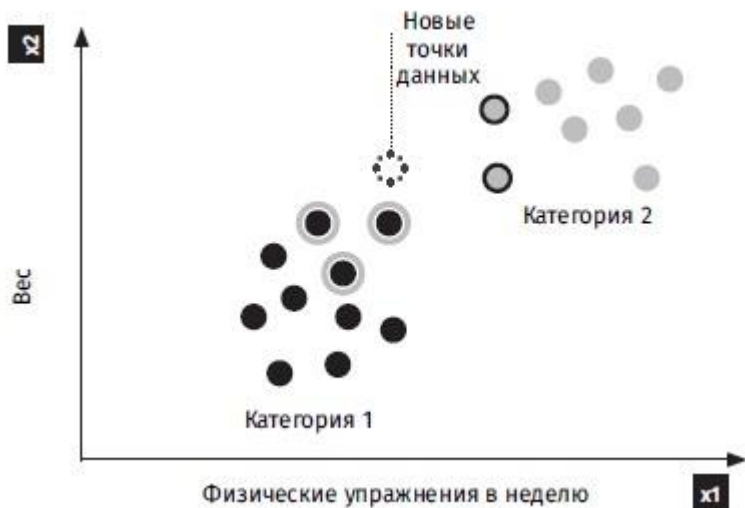


Рис. 6. Алгоритм k ближайших соседей (k = 5)

Наивный байесовский классификатор ([подробнее](#)) основан на сильном, наивном допущении независимости признаков: все характеристики массива данных не зависят друг от друга. На самом деле было бы наивным так полагать, поскольку для многих массивов данных может быть выявлен уровень корреляции содержащихся в них независимых переменных. Несмотря на это наивный алгоритм Байеса хорошо зарекомендовал себя во многих сложных приложениях, например, для обнаружения спама в электронной почте.

Наивный байесовский классификатор использует переменные нашей точки данных, чтобы отнести ее к наиболее подходящему классу. Некий винодел обнаружил, что продолжительность солнечного сияния и количество осадков оказывают влияние на виноградные лозы и повышают шансы на успех вина – категория «Победители». Предположим, что в период созревания урожая выпало 601,98 мм осадков и что на это время пришлось 3543 часа солнечного сияния. Используя эту информацию, мы можем построить график рассеяния для урожая этого года, и наивный классификатор Байеса поможет нам определить, в какую категорию попадет урожай этого года:

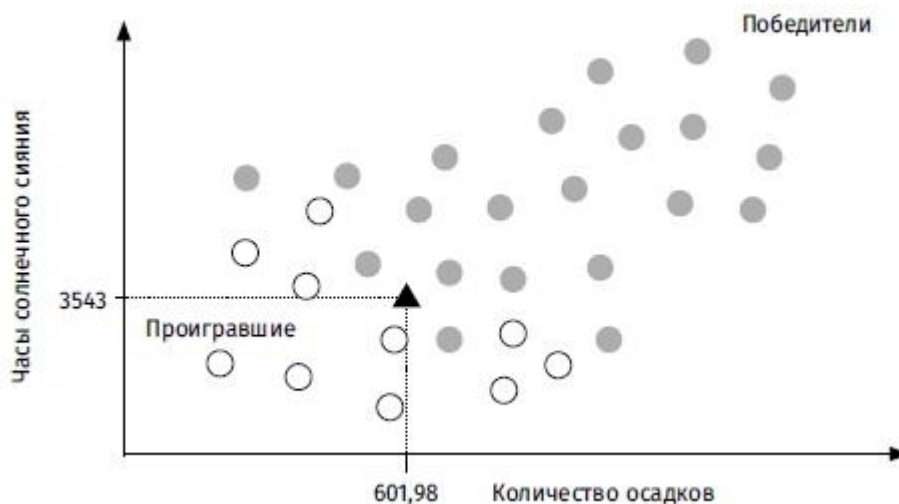


Рис. 7. Качество урожая винограда в зависимости от продолжительности солнечного сияния и количества осадков

Установите априорную вероятность новой точки данных на основе всех существующих точек:

$$P(\text{победитель}) = \frac{\text{Количество победителей}}{\text{Общее количество наблюдений}} = \frac{20}{30} = 0,667.$$

Вычислите вероятности того, что новая точка данных находится в непосредственной близости от области, куда фактически попадает рассматриваемый вариант (рис. 8). Радиус круга мы выбираем произвольно; это параметр, который мы можем настраивать, чтобы влиять на эффективность алгоритма. Вероятность того, что новая точка данных попадет в круг:

$P(X) = \text{Аналогичные наблюдения} / \text{Общее число наблюдений} = 4/30 = 0,133$ .

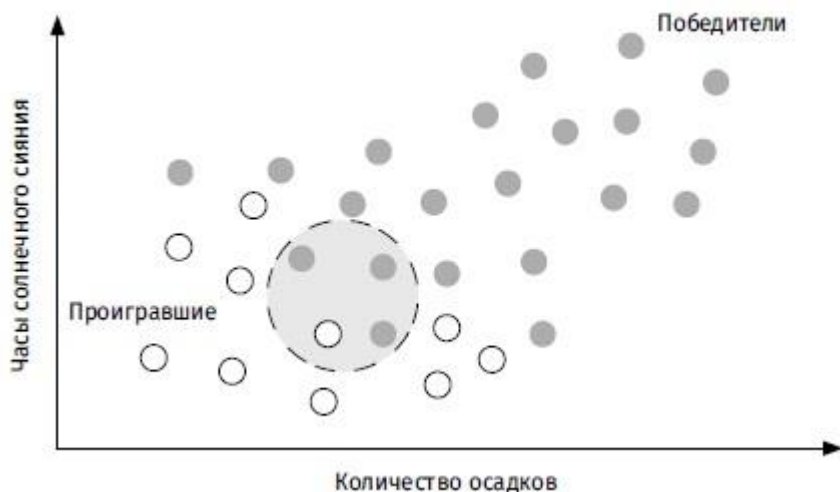


Рис. 8. Расчет вероятности того, что новая точка данных попадет в круг

Вычислите вероятность того, что точка попадет в круг и одновременно будет победителем:

$P(X|\text{победитель}) = \text{Количество победителей в круге} / \text{Общее число победителей} = 3/20 = 0,15$ .

Рассчитайте апостериорную вероятность (этот шаг выполняется с помощью формулы Байеса):

$$P(\text{Победитель}|X) = \frac{P(X|\text{Победитель}) \times P(\text{Победитель})}{P(X)}$$

Вероятностные и детерминированные классификаторы. Метод k ближайших соседей дает однозначный ответ. Этот метод принадлежит к семейству детерминированных алгоритмов классификации. Вероятностные модели, такие как наивный байесовский классификатор, предсказывают распределение вероятностей по всем классам. Затем это распределение можно использовать, чтобы отнести данные к классу.

*Логистическая регрессия* позволяет найти вероятности категориальных переменных. Например, можно построить регрессию для ответов «да» и «нет» в зависимости от возраста респондентов:

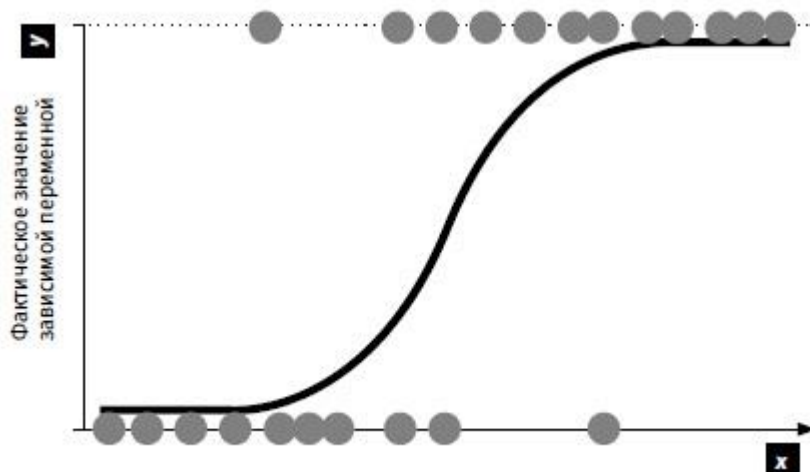


Рис. 9. Линия логистической регрессии

**Кластеризация.** Если вы не знаете, какими могут оказаться группы в результате анализа, следует использовать метод кластеризации. Алгоритм *k-средних* обнаруживает статистически значимые категории или группы в массиве данных. Это идеально подходит тогда, когда у нас есть две или более независимых переменных в массиве данных и мы хотим объединить точки данных в группы с похожими атрибутами (подробнее о методе см. в [книге](#) Джона Формана).

*Иерархическая кластеризация* может выполняться двумя методами: снизу вверх и сверху вниз. В первом случае метод можно представить в виде дендрограммы, на которой отражены этапы, где



в начале каждой точке соответствует свой кластер, а в конце все точки объединены в один кластер:

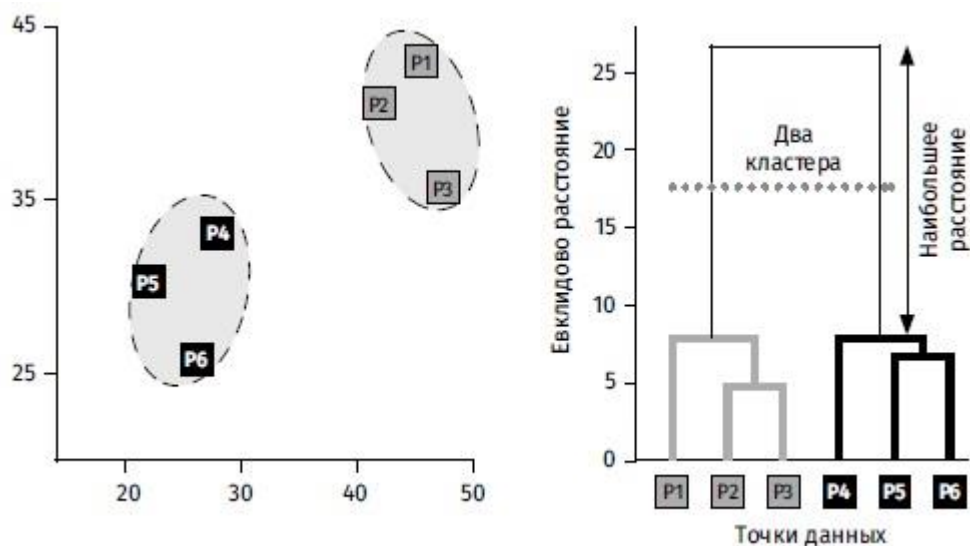


Рис. 10. Разделение на два кластера

Найдите самую длинную вертикальную линию на дендрограмме (справа на рис. 10), установите пороговую линию (горизонтальный пунктир) в точке так, чтобы она пересекала сегмент. Полученное число кластеров оптимально для вашей задачи. В нашем случае это два кластера.

Если вы хотите узнать больше о том, как можно работать с алгоритмами кластеризации посетите [SuperDataScience](http://SuperDataScience), чтобы найти ряд ресурсов, учебных пособий и курсов.

## Глава 7. Анализ данных (часть II)

*Обучение с подкреплением* — это, по сути, форма машинного обучения, которая опирается на концепцию бихевиоризма при обучении искусственного интеллекта и управлении роботами. Наша собственная нервная система является самым впечатляющим из всех алгоритмов обучения с подкреплением.

Допустим у вас есть 5 игровых автоматов – «одноруких бандитов». Вероятность выигрыша установлена казино и составляет: 40%, 25%, 10%, 55% и 75%. Мы не знаем этих вероятностей. Мы лишь знаем, что вероятности различны, и наша задача отыскать автомат с наилучшими шансами на выигрыш, потеряв как можно меньше денег. Вы начинаете в предположении, что шансы всех автоматов 50%, а доверительный интервал простирается от 0 до 100%:

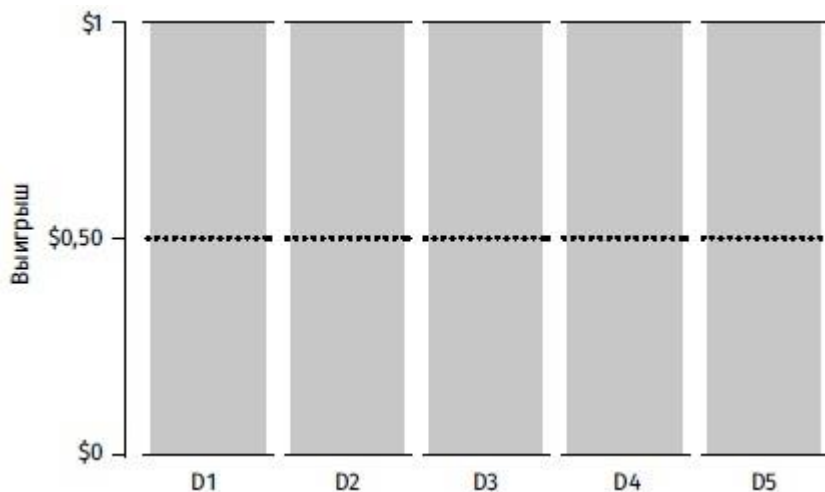


Рис. 11. Перед началом игры

По мере того, как вы играете на том или ином автомате, среднее будет смещаться от 50%, а доверительные границы – сужаться. Причем чем больше сыграно на каком-то автомате игр, тем уже будет доверительный интервал. Например, после ряда игр картина может быть такой:

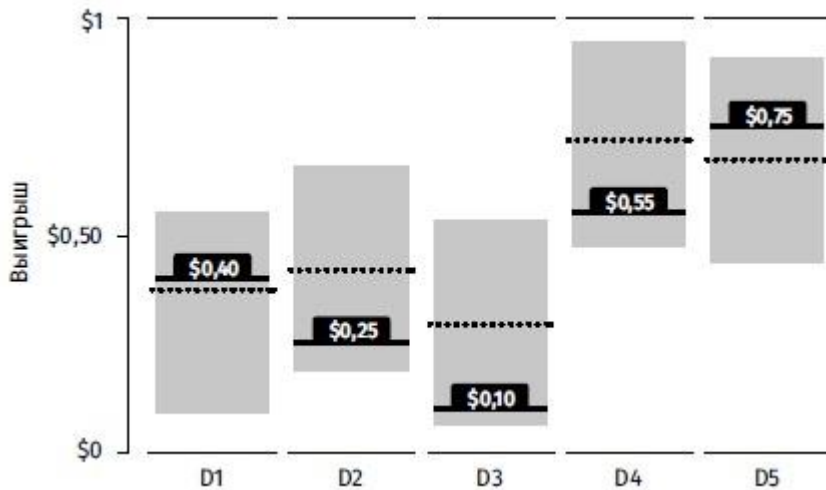


Рис. 12. После 50 игр

Истинные значения не известны, и нанесены для наглядности. Всё что вы узнали после проведенных игр – это пять средних (пунктирные линии) и пять серых прямоугольников – доверительные интервалы. Эти прямоугольники отражают область, в которую значения попадут с некоторой заданной вероятностью (например, 95%). В этот момент наиболее перспективным вам представляется четвертый автомат. Однако сыграв на нем следующие 10 игр, вы видите, что среднее сместилось вниз, а интервал стал уже (так как возросло число игр, сыгранных на нем):

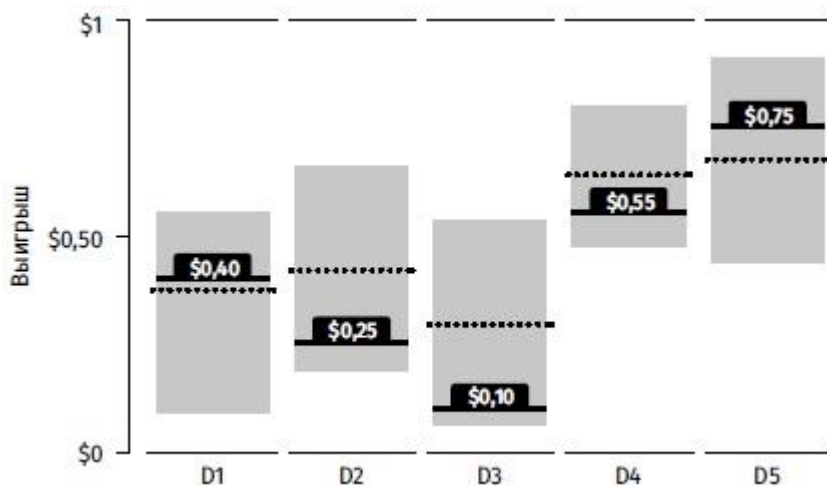


Рис. 13. Ситуация после следующих 10 игр, сыгранных на четвертом автомате

Когда наблюдаемое среднее значение D4 приближается к ожидаемому выигрышу для этой машины, а ее доверительные границы достаточно узки, алгоритм переключается на автомат D5.

Алгоритм *выборки Томпсона* основан не на детерминистском выборе, как обучение с подкреплением, а на вероятностном. Т.е., в любой момент автомат будет выбираться на основе случайного числа с вероятностью нормального распределения:



Рис. 14. Выборка Томпсона для трех одноруких бандитов после нескольких игр

С максимальной вероятностью будет выбран третий автомат. По мере того, как на нем будет играть всё больше туров, его наблюдаемое среднее будет смещаться влево, в сторону истинного значения, и алгоритм переключится на второй автомат.

## Глава 8. Визуализация данных

Визуализация данных — это процесс создания наглядных средств, помогающих людям видеть и понимать информацию. Визуализируя данные, мы представляем их в контексте. Ведь без контекста данные бессмысленны.

Вот мысль, которая, как я знаю, расстроит многих аналитиков данных: в некотором смысле ваш проект не касается ни данных, ни визуальных элементов. В конечном счете речь идет о людях. Нас привлекли к решению бизнес-вопроса, который повлияет на заинтересованных лиц, будь то клиенты или руководители. И если наша информация не будет представлена таким образом, чтобы ее поняли люди, которые могут санкционировать изменения, рекомендованные нашими данными, — тогда все наши усилия окажутся потрачены впустую.

Вдохновляющие идеи для визуализации проектов смотрите на сайте журналиста, работающего с данными, Дэвида Маккэндлесса, [informationisbeautiful.net](http://informationisbeautiful.net), что выводит визуальное повествование на новый уровень.

Есть несколько приемов, которым вы можете научиться, чтобы более эффективно воздействовать на свою аудиторию с помощью визуальных эффектов. Вернитесь к первому шагу процесса анализа и обработки данных: определите вопросы и имейте их в виду, когда начнете отбирать визуальные элементы.

Ограничьте объем текста. Я предпочитаю использовать как можно меньше текста, позволяя визуальным элементам выполнять **большую** часть работы. Используйте два шрифта: один для заголовков, а другой для основного текста, и, если у вас есть соблазн уменьшить размер шрифта для любой информации, спросите себя, не следует ли ее вообще выбросить.

*Цветные схемы.* Многие люди недооценивают важность цвета. Я был одним из них, пока не изучил теорию цвета и не увидел, насколько могу улучшить свои визуализации, просто правильно комбинируя цвета. Следовать принципам использования цвета довольно просто. Лучший метод, на мой взгляд, основан на цветовом колесе:

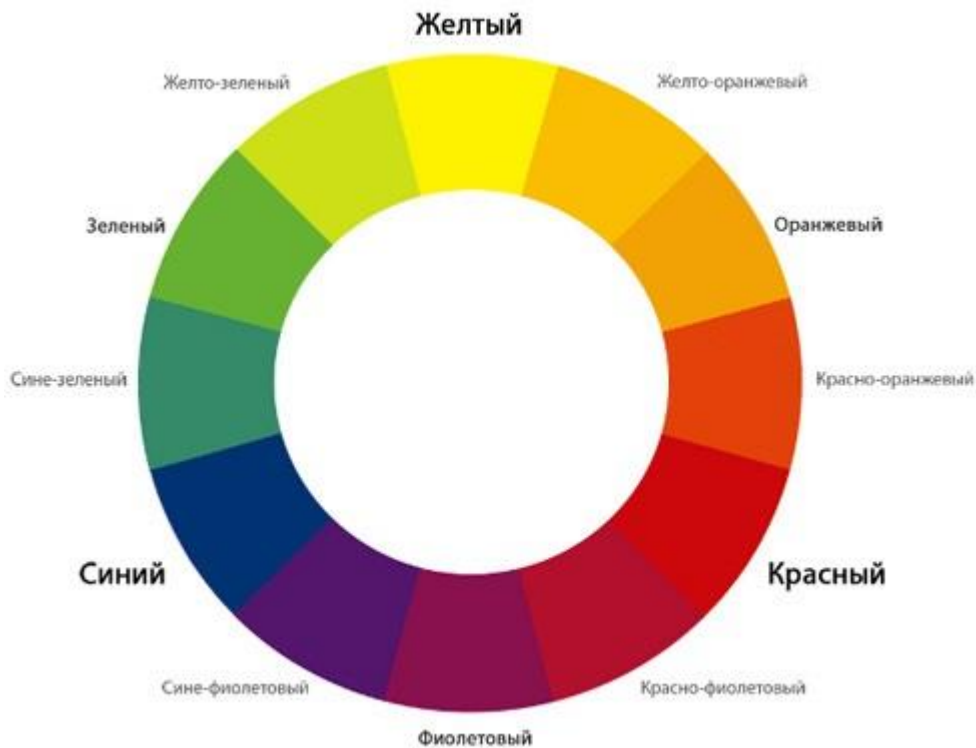


Рис. 15. Цветовое колесо

*Монохромная схема.* После того как вы выбрали один цвет, найдите более светлый или более темный оттенок для различения категорий. *Аналоговая схема.* Выберите один цвет и два прилегающих к нему цвета. Отлично подходит для тепловых карт, а также для иллюстрации постепенных изменений в данных. *Контрастная схема.* Выберите два цвета, диаметрально противоположные друг другу на колесе. Схема эффективна при сравнении двух значений и обычно используются для выделения отдельных категорий. *Расщепленно-контрастная схема.* Идеально подходит, если у вас есть три категории, но вы хотите, чтобы ваша аудитория сосредоточилась на одной, в частности, для того, чтобы выделить тенденции или особенности. *Триада.* Идеально подходит для изображения трех категорий равной значимости. *Квадратическая схема* для четырех категорий равной значимости.



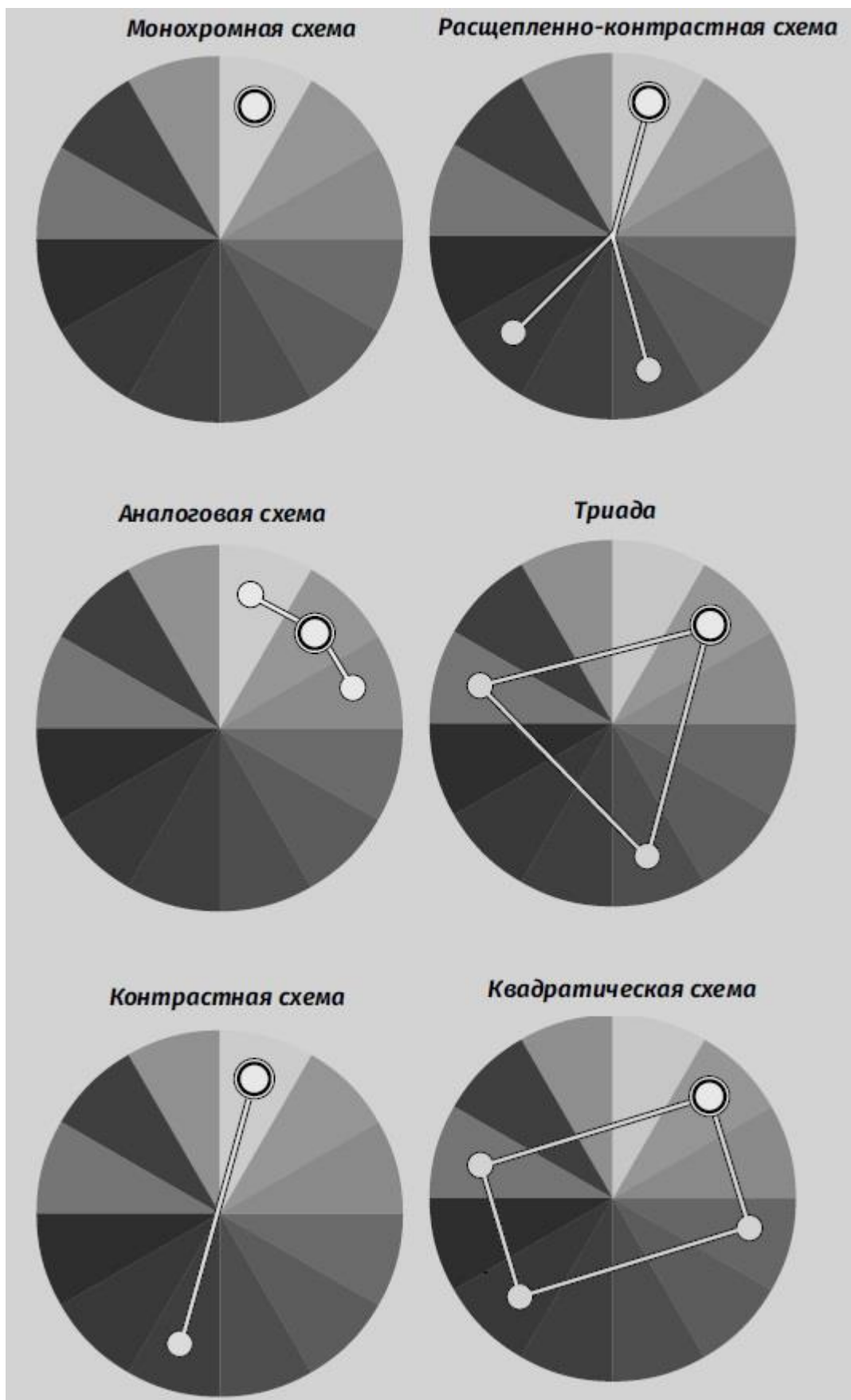


Рис. 16. Цветовые схемы

Не бойтесь сокращений. Нет необходимости включать все числа из массива данных — относительные различия на диаграммах, которые вы используете, должны представлять лишь некоторые из них. Иногда названия столбцов также могут быть удалены в зависимости от того, какую идею вы хотите донести до своих слушателей (см. также [Принцип Эдварда Тафти минимизации количества элементов диаграммы](#)).

Типы графиков: тепловая карта, гистограммы и распределение вероятностей, линейный график, диаграмма с областями, точечная диаграмма, пузырьковая диаграмма, круговая диаграмма, плоское дерево, диаграмма «водопад», диаграмма Сэнки (отображает движение данных, используя размер и направление), шкала Ликерта (сумма ответов на вопросы анкеты).

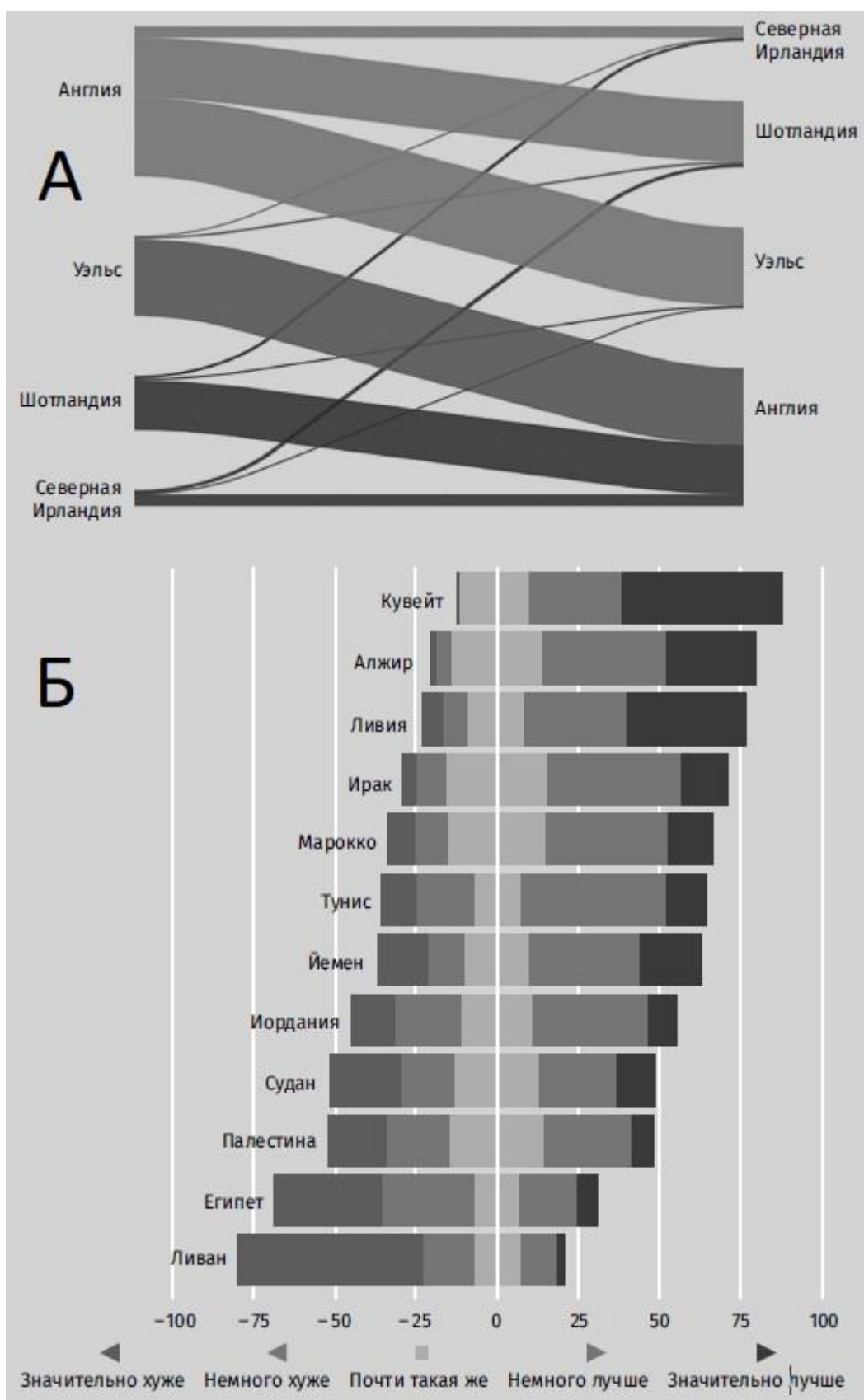


Рис. 17. Диаграмма Сэнки (а) и шкала Ликерта (б) с ответами на вопрос: как вы думаете, какова будет экономическая ситуация в вашей стране в ближайшие годы (3-5 лет) по сравнению с нынешней ситуацией?

## Глава 9. Презентация данных

Когда меня спрашивают, что отличает хорошего аналитика данных от аналитика высшего класса, мой ответ всегда таков: способность передавать идеи. Рассказывание историй является ключевой частью процесса науки о данных. Есть множество книг о том, как сделать эффективную презентацию (см. например, [Мартин Сайкс и др. От слайдов к историям. Пошаговая методика создания убеждающих презентаций](#)). Не существует единого подхода к созданию хорошей презентации. Я разработал свой собственный стиль. Я начинаю каждую презентацию с мозгового штурма по теме проекта. Для меня это означает запись начальной (А) и конечной (Б) точек проекта друг напротив друга на листе бумаги. Пустое пространство между ними представляет мой маршрут, и я хочу показать аудитории, как я прошел путь от А до Б. В этом пространстве я записываю этапы процесса анализа и обработки данных, выделяя любые моменты, которые считаю особенно впечатляющими или показательными.

Процесс обработки и анализа данных имеет логичную, простую в использовании структуру, так почему бы не объяснить сущность этапов и не провести слушателей по ним шаг за шагом? Когда люди чувствуют себя комфортно — ведь им ясно, о чем идет речь, — они запоминают больше информации и будут не только слышать, что вы говорите, но на самом деле слушать вас.

Одно из главных правил в бизнес-маркетинге — как можно раньше показать своей аудитории преимущества того, что вы продаете. Можно либо рассмешить аудиторию, либо задать ей вопрос, который требует поднятия руки.

Позаботьтесь о том, чтобы все в аудитории одинаково хорошо понимали, о чем идет речь. На презентацию придут люди, в разной степени сведущие в вашей теме. Обращайтесь к человеку, у которого, как вам кажется, знаний меньше, чем у остальных.

Популярный спикер Рэнди Олсон в своей основополагающей работе «Хьюстон, у нас есть история?» говорит о пользе рассказывания историй для передачи сложных научных концепций.

Рассказывая о собственном опыте, вы не должны сосредотачиваться на положительных моментах. Никто не захочет слушать историю о том, как богатый становится еще богаче. Поведайте о препятствиях, с которыми вы столкнулись, и о том, как их преодолели.

Не переборщите со слайдами — аудитории нужно время переварить информацию; показывайте не более одного слайда каждые три минуты. Ограничьте количество текста на каждом слайде; если текста будет слишком много, аудитории придется одновременно читать написанное и слушать выступающего — и в конечном итоге они будут делать и то и другое плохо!

Найдите несколько дружелюбных лиц в разных местах зала: если вы будете смотреть то на одних слушателей, то на других, аудитория решит, что вы взаимодействуете со всеми.

## Глава 10. Ваша карьера в науке о данных

В эпоху, когда так много рабочих мест рискуют быть ликвидированы в течение 20 лет, наука о данных должна представлять интерес для всех, кто хочет обеспечить себя гарантированной и интересной работой.

Какого рода должности могут ожидать вас? Бизнес-аналитик, аналитик подготовки данных, аналитик моделирования данных, специалист по данным /расширенной аналитике /практик машинного обучения /старший научный сотрудник по данным, менеджер по анализу и обработке данных.

В докладе Burning Glass Technologies и IBM за 2017 г. рассматривается спрос на специалистов по работе с данными в шести ключевых секторах: профессиональные услуги (в том числе консультации по вопросам управления, юриспруденции и медицины); финансы и страхование; производство; информация; здравоохранение и социальная помощь; розничная торговля.