

## Что такое язык R

R — язык программирования, созданный специально для статистического анализа данных. Его разработали на факультете статистики Оклендского университета для внутреннего использования под влиянием другого подобного языка — S, который был платным и недоступным для широкого круга разработчиков. Поэтому в Окленде решили создать бесплатную альтернативу.

На другие популярные языки R абсолютно не похож. У него свой уникальный синтаксис, функции и принципы работы. У языка R чёткая сфера применения — статистические вычисления, анализ данных и [машинное обучение](#). Его создавали специально для этих задач, и для других он не подходит.

R — не только язык для анализа данных, но и целая рабочая среда, куда уже встроены готовые методы статистического анализа и [инструменты для визуализации](#).

## Зачем изучать R

Язык R — один из самых распространённых в научной среде. Им пользуются математики, биологи, генетики и другие учёные, которым нужно проводить статистические исследования и строить модели. Поэтому язык R нужно изучать тем, кто планирует заниматься научными исследованиями.

Вне научной среды язык программирования R очень востребован. Статистические исследования важны для двух коммерческих специальностей — [аналитиков данных](#) и [специалистов по Data Science](#). Им регулярно нужно проводить математические расчёты на основе выборок данных, и язык R для этого подходит идеально. В сфере анализа данных и машинного обучения без R — никуда.

Кроме того, язык разработки R хорошо интегрируется с другими языками программирования. Например, благодаря библиотеке [Shiny](#) он позволяет встраивать в веб-страницы интерактивные дашборды, графики и таблицы — то есть добавлять элементы сложного статистического анализа в интерфейс сайтов и приложений. Поэтому его полезно будет изучить и тем, кто знает другие языки программирования, чтобы использовать в своей работе.

## Особенности языка программирования R

**Может работать с разными парадигмами программирования, но лучше всего справляется с объектно-ориентированной.** Функции и таблицы для него — объекты, которые взаимодействуют между собой. Это позволяет писать сложные распределённые программы, в которых удобно

по несколько раз использовать одни и те же функции и объекты.

**Интерпретируемый.** Программа на языке R сразу готова к исполнению — её не нужно собирать в исполняемый файл с помощью компилятора. Можно ещё в процессе написания кода проверять его работу по частям.

**Простой по синтаксису.** R в своей основе не содержит сложных конструкций и запутанных функций. Даже типов данных у него всего четыре: символьные, числовые, логические и комплексные. Но эти простые типы и функции языка R можно собирать в сложные конструкции и структуры данных. Это как легко — детали сами по себе простые, а создать из них можно как игрушечную машинку, так и боевой вертолёт.

**Есть удобные интерактивные инструменты.** Для R есть среда разработки [RStudio](#) с подсветкой синтаксиса, навигацией по тексту программы, сортировкой таблиц и отображением графиков в отдельном окне. Есть приложение-блокнот [Jupyter Notebook](#) для создания и обмена программами на R прямо в браузере. И дистрибутив [Anaconda](#) с коллекцией популярных библиотек.

**Доступны тысячи библиотек и расширений.** [Готовые функции](#) для визуализации данных, быстрых статистических операций, распознавания текстов, A/B-тестирования и отдельных научных отраслей.

### **Возможности языка R**

Обычно на R не пишут классические сервисы с интерфейсом, кнопками и текстом, к которым все привыкли. Чаще всего программы на R представляют собой код, выполнение которого выводит какой-то определённый результат или графики. С помощью языка R можно:

- **Очистить и обработать данные.** Например, для обработки массива информации о популярности разных товаров на R можно написать программу, которая сгруппирует данные о покупках по товарам, удалит дубли и подготовит таблицу для дальнейшей аналитики.
- **Провести статистический тест.** Посчитать среднюю продолжительность и увидеть, есть ли статистически значимая разница между несколькими показателями.
- **Объединять данные из разных таблиц.** Взять таблицы разных форматов, собрать из них данные и обрабатывать все как единый файл.
- **Отрисовать интерактивный график.** Распределить данные, отрегулировать параметры.

- **Анализировать регрессионные модели.** Выявить отношения между переменными — например, как доход магазина зависит от разных факторов.

- **Провести другую математическую операцию.** Объединить многомерные массивы, спрогнозировать величину, распознать текст. Для большинства задач есть готовые библиотеки, либо можно написать собственный код, используя широкий математический инструментарий.

### **Что можно написать на R**

Аналитика — не единственное применение языка R. На нём можно писать интерактивные веб-приложения. Для этого есть отдельная библиотека — Shiny. С её помощью можно быстро создать страницу, на которой будут выводиться результаты работы программы, написанной на языке R.

Пример такой программы — [статистика заболеваемости коронавирусом](#).

### **ем R отличается других инструментов анализа данных**

У языка программирования R есть два типа «конкурентов»:

- Инструменты для анализа данных без кода: Power BI, Excel, Google Sheets, Tableau.

- Языки программирования, ориентированные на работу с данными: Python и Julia.

От каждого из этих типов у R свои отличия.

#### **Инструменты для анализа данных**

Эти инструменты гораздо проще в освоении, особенно на базовом уровне — даже в школах детей учат работать с Excel. Они имеют графический интерфейс и позволяют быстро проводить простые операции. Например, для подсчёта среднего значения в Excel достаточно пары кликов.

Однако из простоты вытекает ограниченность. Если какой-то функции в инструменте нет, с этим ничего не поделать — придётся ждать, пока разработчик её добавит. Самостоятельно написать инструмент, нужный для конкретной задачи, не получится. Кроме того, не все инструменты подходят для анализа слишком больших массивов данных, например таблиц величиной в миллионы строк.

#### **Другие языки программирования**

Python, Julia и другие языки, которые используют для анализа данных, считаются более универсальными. На них можно писать полноценные приложения, создавать графические интерфейсы и разрабатывать программы, напрямую не связанные с математикой и статистикой.

При этом из-за своей универсальности они несколько проигрывают R,

который создавали специально для статистического анализа. В других языках меньше готовых аналитических инструментов, и сложные математические операции проводить в них менее удобно. Рассмотрим основные плюсы и минусы языка программирования R.

### ✓ Плюсы

- **Практически неограниченный набор готовых функций для анализа данных.** Если нужно провести какую-то операцию, для неё наверняка уже написана функция или библиотека.
- **Возможность работать с большими данными.** Классическим аналитическим программам таблицы с миллионами строк не по зубам.
- **Разнообразный интерфейс.** Можно работать как в командной строке, так и в пользовательском редакторе с графическим интерфейсом. У инструментов для анализа нет первого, у языков программирования — второго.
- **Бесплатная лицензия.** За большинство программ для анализа данных нужно платить.
- **Работа под любой операционной системой.** Даже непопулярными дистрибутивами Linux.
- **Удобное создание и редактирование графиков.** Можно создавать визуализацию любого типа.
- **Активное сообщество.** Язык R использует и поддерживает много людей, по нему проводят конференции, отвечают на вопросы, записывают видеоуроки.

### ✗ Минусы

- **Узкая специализация.** В отличие от других языков программирования, R используют только для аналитики. Полноценное приложение на нём не напишешь. Поэтому нужен он скорее не разработчикам, а тем, кто занимается анализом данных.
- **Высокий порог входа.** Язык R сложен для изучения, несмотря на простой синтаксис. У него много нюансов, без знания которых работать не получится. Для полноценного программирования на R придётся глубоко погрузиться в математику.
- **Малая популярность в русскоязычном сообществе.** В России Python

намного популярнее. По R сложнее найти вакансии, а обсуждения, уроки и материалы по языку почти все написаны на английском.

### **Где и как учить язык R**

Если специалист хочет заниматься наукой, то изучать язык R лучше в вузе.

Когда интересно развитие в другой области знаний, то, как и в случае с другими языками программирования, его можно выучить на курсах.

Существуют как отдельные программы по языку R, так и курсы анализа данных или Data Science, где R изучают как рабочий инструмент для этих специальностей.

Чтобы язык R действительно приносил пользу, вместе с ним нужно изучить математику: алгоритмы, функции, статистический анализ. На курсах по анализу данных это обычно преподают, а осваивать самостоятельно можно, например, по вузовским учебникам.