

# Знакомство с анализом данных и R

Храмов Д.А.

14.02.2020

# Содержание

- ▶ Чем занимается “Интеллектуальный анализ данных”.
- ▶ Инструмент для работы: пакет R. Что это и почему именно он?
- ▶ Установка R и RStudio. Дополнительные пакеты.
- ▶ Полезные ресурсы.
- ▶ R Markdown — создание отчетов.

# Интеллектуальный анализ данных = Data Mining

**Data Mining** занимается выявлением в данных ранее неизвестных закономерностей, описанием этих закономерностей, а также прогнозированием на их основе.

PRICE	SQFT	AGE	FEATS	NE	CUST	COR	TAX
2050	2650	13	7	1	1	0	1639
2080	2600	-9999		4	1	1	0 1088
2150	2664	6	5	1	1	0	1193
2150	2921	3	6	1	1	0	1635
1999	2580	4	4	1	1	0	1732
1900	2580	4	4	1	0	0	1534
1800	2774	2	4	1	0	0	1765
1560	1920	1	5	1	1	0	1161
1450	2150	-9999		4	1	0	0 -9999
1449	1710	1	3	1	1	0	1010
1375	1837	4	5	1	0	0	1191
1270	1880	8	6	1	0	0	930
1250	2150	15	3	1	0	0	984
1235	1894	14	5	1	1	0	1112
1170	1928	18	8	1	1	0	600
1180	1830	-9999		3	1	0	0 733
1155	1767	16	4	1	0	0	794
1110	1630	15	3	1	0	1	867
1139	1680	17	4	1	0	1	750
...							

Выборка данных по продаже домов в Альбукерке (шт. Нью-Мексико, США).

# Data Mining $\subset$ Data Science

Data Mining (интеллектуальный анализ данных) — часть **науки о данных** (data science), куда входят:

- ▶ Статистический анализ данных (statistics)
- ▶ Машинное обучение (machine learning)
- ▶ Распознавание образов (pattern recognition)
- ▶ Большие данные (big data)
- ▶ Глубокое обучение (deep learning)
- ▶ Анализ текстов (text mining)

Все эти дисциплины тесно связаны друг с другом.

Будем называть наш предмет просто — **Анализ данных**.

# Основные направления

- ▶ Прогнозирование (prediction)
- ▶ Выявление структуры (structure discovery)
- ▶ Выявление взаимоотношений (relationship mining)
- ▶ Ранжирование (ranking)
- ▶ ...

## Прогнозирование

Цель прогнозирования — разработать модель, предсказывающую значение интересующей величины (выходной переменной) по значениям независимых переменных (предикторов).

Прогнозирование предполагает обучение модели на данных. Обучение — настройка параметров модели.

Обучение происходит на данных, где для заданных значений предикторов известны значения выходной переменной (обучающие данные). Научившись, модель позволяет предсказать значение выходной переменной для новых (неразмеченных) данных.

Если выходная переменная принимает непрерывные значения, то мы имеем дело с **регрессией** — поиском зависимости между входными и выходной переменными. Если выходная переменная имеет конечный дискретный набор значений (классов), то говорят о задаче **классификации**.

## Выявление структуры

Алгоритмы выявления структуры пытаются обнаружить в данных структуру, без использования каких-либо априорных представлений о ней.

Наиболее известной группой подобных алгоритмов являются алгоритмы кластеризации.

Кластеризация является логическим продолжением идеи классификации. Но:

- ▶ для классификации нужны обучающие данные, в которых известна связь между предикторами и выходной переменной (меткой класса). Число классов задано.
- ▶ при кластеризации обучающих данных нет, а число классов заранее не известно. Данные — одни предикторы без метки классов.

Результатом кластеризации является разбиение множества объектов на группы близких в каком-либо смысле объектов.

# Выявление взаимоотношений

Цель выявления взаимоотношений — установить взаимосвязи между переменными в наборе данных с бОльшим числом переменных. Например, определить, какие переменные сильнее всего связаны с интересующей или в какой из пар связанных переменных связь сильнее, чем в других.

**В поиске связующих правил** целью является найти правила вида “if-then”, указывающие на то, что если (if) одни переменные принимают какое-то множество значений, то (then) другая переменная будет иметь определенное значение. Например, если в транзакции встретился набор товаров А, то можно сделать вывод, что в этой же транзакции должен появиться набор товаров В.

**Поиск последовательных шаблонов** является дальнейшим развитием поиска связующих правил и предназначен для выявления взаимоотношений между проявлениями последовательных во времени событий (если пользователь приобрел товар А, то впоследствии он приобретет товар В).



# Ранжирование

**Ранжирование** (machine-learned ranking) — задача сортировки набора элементов исходя из их релевантности. Релевантность понимается по отношению к некоторому объекту.

1. В задаче информационного поиска объект — это запрос, элементы — всевозможные документы (ссылки на них), а релевантность — соответствие документа запросу
2. В задаче рекомендаций объект — это пользователь, элементы — тот или иной рекомендуемый контент (товары, видео, музыка), а релевантность — вероятность того, что пользователь воспользуется (купит/лайкнет/просмотрит) данным контентом.

Обучающая выборка состоит из выборки поисковых запросов, подмножества документов, им отвечающим, и оценок релевантности каждого документа запросу.

## Примеры использования

Сферы применения: везде, где имеются какие-либо данные.

Преимущества: скорость, автоматизация, точность, возможность масштабирования.

- ▶ рекомендательные системы в интернет-коммерции;
- ▶ социальные сети — влияние на выборы;
- ▶ поисковые системы — оперативное предсказание эпидемий;
- ▶ массовые онлайн-курсы: прогнозирование оценок, рекомендации студентам;
- ▶ розничные сети: подбор товаров;
- ▶ лечебные учреждения: истории болезней, диагнозы;
- ▶ наблюдения за погодой;
- ▶ банки: кредитный скоринг и рекомендательные системы;
- ▶ технические системы (самолет): управление, предотвращение аварий;
- ▶ ...

# Почему стоит заниматься анализом данных

- ▶ это современно.
- ▶ это востребовано: “Data Scientist: The Sexiest Job of the 21st Century (Harvard Business Review, 10/2012)”
- ▶ это неплохо оплачивается — в среднем 113 тыс. дол. в год против 63.5 тыс. дол. у программистов.

Объем данных в цифровой форме растет гораздо быстрее темпов подготовки специалистов. На рынке труда наблюдается бум спроса на data science-специалистов (пример). Возникают новые отрасли по работе с данными.

# Поиск курсов по data science на Coursera

coursera

Изучить ▾

data science



Для организаций

Показано результатов для запроса "data science": 1056

Фильтр

Язык ▾

Уровень ▾

Навыки ▾

Сотрудничайте ▾



## IBM Data Science

IBM

PROFESSIONAL CERTIFICATE

★★★★☆ 4.5 (26 994)

270K студенты

Beginner



## Машинное обучение и анализ данных

Moscow Institute of Physics and Technology

SPECIALIZATION

★★★★☆ 4.7 (4 956)

67K студенты

Intermediate



## Data Science

Johns Hopkins University

SPECIALIZATION

★★★★☆ 4.4 (31 437)

730K студенты

Beginner



## Математика и Python для анализа данных

Yandex

COURSE

★★★★☆ 4.8 (4 476)

59K студенты

# Кто такой Data scientist?

Это специалист, владеющий тремя группами навыков:

1. IT-грамотность — программирование, придумывание и решение алгоритмических задач, владение софтом;
2. знания математики и статистики;
3. содержательный опыт в какой-то области — понимание запросов своей организации или задач своей отрасли науки.

## Инструментарий: пакет R

R — язык программирования для анализа данных.

Создан в 1993 г. Россом Айхэкой (Ross Ihaka) и Робертом Джентлменом (Robert Gentleman) в Оклендском технологическом университете (Новая Зеландия).

R — свободное ПО, распространяемое с открытым исходным кодом.

CRAN (The Comprehensive R Archive Network) — доступно более 15000 пакетов расширений.

Позволяет реализовать полный цикл работ с данными: импорт данных, предварительную обработку, моделирование, валидацию, визуализацию результатов, подготовку документации.

RStudio — отличная среда разработки. Есть Open Source Edition.

# Другие статистические пакеты

## Проприетарные:

- ▶ SAS
- ▶ SPSS
- ▶ Statistica
- ▶ Minitab
- ▶ Stata

## Свободные:

- ▶ Python (+Pandas, scikit-learn)
- ▶ Julia

# Почему R?

Субъективно, потому что:

- ▶ более низкий порог вхождения, чем у Python;
- ▶ R заточен под статистическую обработку данных и машинное обучение, тогда как Python — язык программирования общего назначения, “дообученный” анализу данных.
- ▶ концепции анализа данных лучше осваивать на R — меньше программистских нюансов.

Но:

- ▶ Deep Learning — R может, но через Python.
- ▶ Big Data — понадобятся специальные инструменты.
- ▶ Знание R+Python — совсем хорошо. Возможно, эту связку заменит Julia.





[\[Home\]](#)

## Download

[CRAN](#)

## R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Developer Pages](#)

[R Blog](#)

## R Foundation

[Foundation](#)

# The R Project for Statistical Computing

## Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

## News

- [R version 3.6.2 \(Dark and Stormy Night\)](#) has been released on 2019-12-12.
- useR! 2020 will take place in St. Louis, Missouri, USA.
- [R version 3.5.3 \(Great Truth\)](#) has been released on 2019-03-11.
- The R Foundation Conference Committee has released a [call for proposals](#) to host useR! 2020 in North America.
- You can now support the R Foundation with a renewable subscription as a [supporting member](#)
- The R Foundation has been awarded the Personality/Organization of the year 2018 award by the professional association of German market and social researchers.



# Установка R: R for Windows

## R for Windows

Subdirectories:

<a href="#">base</a>	Binaries for base distribution. This is what you want to <a href="#">install R for the first time</a> .
<a href="#">contrib</a>	Binaries of contributed CRAN packages (for R $\geq$ 2.13.x; managed by Uwe Ligges). There is also information on <a href="#">third party software</a> available for CRAN Windows services and corresponding environment and make variables.
<a href="#">old.contrib</a>	Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 2.13.x; managed by Uwe Ligges).
<a href="#">Rtools</a>	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

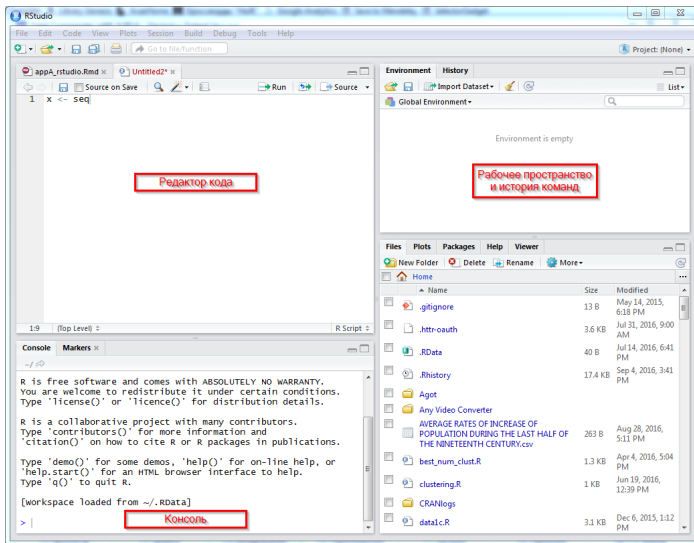
Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

# Установка IDE RStudio

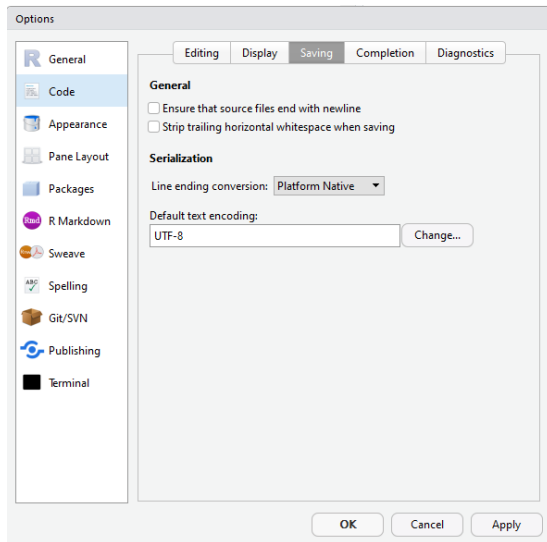
	RStudio Desktop Open Source License <b>Free</b> <a href="#">DOWNLOAD</a> <a href="#">Learn more</a>	RStudio Desktop Commercial License <b>\$995</b> /year <a href="#">BUY</a> <a href="#">Learn more</a>	RStudio Server Open Source License <b>Free</b> <a href="#">DOWNLOAD</a> <a href="#">Learn more</a>	RStudio Server Pro Commercial License <b>\$4,975</b> /year (5 Named Users) <a href="#">BUY</a> <a href="#">Evaluation</a>   <a href="#">Learn more</a>
Integrated Tools for R	✓	✓	✓	✓
Priority Support		✓		✓
Access via Web Browser			✓	✓
Enterprise Security				✓
Project Sharing				✓
Manage Multiple R Sessions & Versions				✓
Admin Dashboard				✓
Load Balancing				✓
Auditing and Monitoring				✓

<https://rstudio.com/products/rstudio/download/>

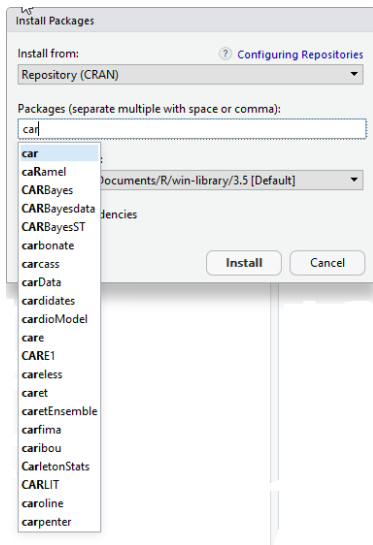
# Окна RStudio



# Настройка кодировки скриптов UTF-8: Tools/General Options...

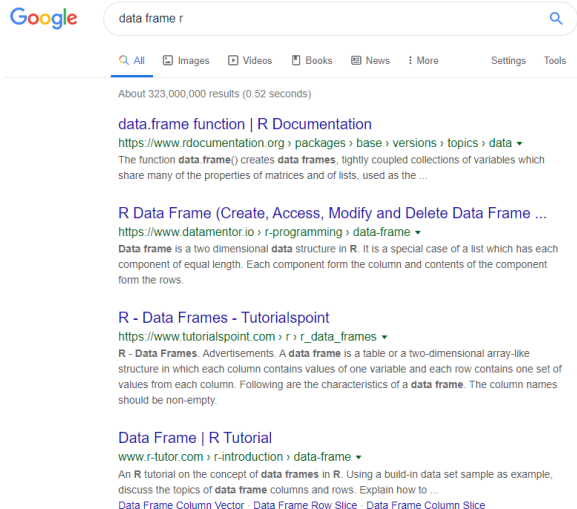


## Установка пакетов: Tools/Install Packages...



- ▶ <https://rdr.io/snippets/> — запуск коротких скриптов.
- ▶ <https://rstudio.cloud/> — R+Rstudio, куча предустановленных пакетов, нужна регистрация.

# Поиск информации по R



The screenshot shows a Google search interface. The search bar contains the text "data frame r". Below the search bar, there are tabs for "All", "Images", "Videos", "Books", "News", and "More". The "All" tab is selected. Below the tabs, it says "About 323,000,000 results (0.52 seconds)". The search results are listed below, each with a title, a URL, and a brief description.

Google

data frame r

All Images Videos Books News More Settings Tools

About 323,000,000 results (0.52 seconds)

**data.frame function | R Documentation**  
<https://www.rdocumentation.org/packages/base/versions/topics/data> ▼  
The function **data.frame()** creates **data frames**, tightly coupled collections of variables which share many of the properties of matrices and of lists, used as the ...

**R Data Frame (Create, Access, Modify and Delete Data Frame ...**  
<https://www.datamentor.io/r-programming/data-frame> ▼  
**Data frame** is a two dimensional **data** structure in **R**. It is a special case of a list which has each component of equal length. Each component form the column and contents of the component form the rows.

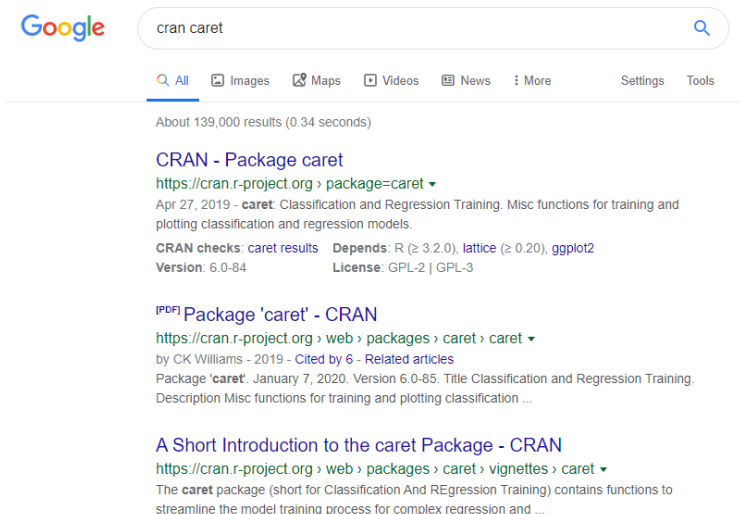
**R - Data Frames - Tutorialspoint**  
[https://www.tutorialspoint.com/r/r\\_data\\_frames](https://www.tutorialspoint.com/r/r_data_frames) ▼  
**R - Data Frames.** Advertisements. A **data frame** is a table or a two-dimensional array-like structure in which each column contains values of one variable and each row contains one set of values from each column. Following are the characteristics of a **data frame**. The column names should be non-empty.

**Data Frame | R Tutorial**  
[www.r-tutor.com/r-introduction/data-frame](http://www.r-tutor.com/r-introduction/data-frame) ▼  
An R tutorial on the concept of **data frames** in **R**. Using a build-in data set sample as example, discuss the topics of **data frame** columns and rows. Explain how to ...  
[Data Frame Column Vector](#) · [Data Frame Row Slice](#) · [Data Frame Column Slice](#)

Добавляем “R” к запросу (data frame)



# Поиск пакета в CRAN



Google

cran caret

Q All Images Maps Videos News More Settings Tools

About 139,000 results (0.34 seconds)

**CRAN - Package caret**  
<https://cran.r-project.org/package=caret> ▼  
Apr 27, 2019 - **caret**: Classification and Regression Training. Misc functions for training and plotting classification and regression models.  
**CRAN checks:** [caret results](#) **Depends:** R (≥ 3.2.0), [lattice](#) (≥ 0.20), [ggplot2](#)  
**Version:** 6.0-84 **License:** GPL-2 | GPL-3

<sup>[PDF]</sup> **Package 'caret' - CRAN**  
<https://cran.r-project.org/web/packages/caret/caret> ▼  
by CK Williams - 2019 - [Cited by 6](#) - [Related articles](#)  
Package '**caret**'. January 7, 2020. Version 6.0-85. Title Classification and Regression Training. Description Misc functions for training and plotting classification ...

**A Short Introduction to the caret Package - CRAN**  
<https://cran.r-project.org/web/packages/caret/vignettes/caret> ▼  
The **caret** package (short for Classification And REgression Training) contains functions to streamline the model training process for complex regression and ...

CRAN + имя (назначение) пакета

## R и анализ данных: литература

- ▶ *Шипунов А. Б., Балдин Е. М. и др.* Наглядная статистика. Используем R!
- ▶ *Шитиков В. К., Мастицкий С. Э.* Классификация, регрессия и другие алгоритмы Data Mining с использованием R
- ▶ *Кабаков Р.* R в действии. Анализ и визуализация данных в программе R.
- ▶ *Мастицкий С. Э.* Визуализация данных с помощью ggplot2.
- ▶ *Clark M.* An Introduction to Machine learning with Application in R
- ▶ *Шолле Ф.* Глубокое обучение на R.
- ▶ *James G., Witten D., Hastie T., Tibshirani R.* An Introduction to Statistical Learning with Applications in R.
- ▶ *Wickham H., Gromlund G.* R for Data Science
- ▶ *Hyndman R. G., Athanasopoulos G.* Forecasting: Principles and Practice
- ▶ *Chang W.* R Graphics Cookbook, 2nd edition
- ▶ *Wickham H.* Advanced R, 2nd edition
- ▶ *Gillespie C., Lovelace R.* Efficient R programming
- ▶ *Sonderegger D. L.* STA 444/5 - Introductory Data Science using R

## R и анализ данных: сайты и видеокурсы

- ▶ <http://www.statmethods.net/> — сайт Р. Кабакова с материалами его книг.
- ▶ <http://r-analytics.blogspot.com/> — сайт С. Э. Мастицкого.
- ▶ *Aust F., Barth M.* papaja: Reproducible APA manuscripts with R Markdown
- ▶ *Аббакумов В.Л.* Анализ данных на R в примерах и задачах: часть 1, часть 2.
- ▶ *Hastie T., Tibshirani R.* In-depth introduction to machine learning

## Другие курсы по машинному обучению

- ▶ *Andrew Ng*. Machine Learning, Stanford University (Octave), 2017
- ▶ *Andrew Ng*. Machine Learning, Stanford University (Octave), 2012. Курсы Эндрю Ына в рекламе не нуждаются.
- ▶ *Воронцов К., Соколов Е.* Введение в машинное обучение (Python)
- ▶ *Кашницкий Ю.* Open Machine Learning Course (Python)

## Markdown: подготовка отчетов

Markdown — язык разметки текстовых документов.

Документ в Markdown должен восприниматься как обычный текст, чтобы инструкции форматирования не бросались в глаза. Вмешательство в текст должно быть минимальным.

Поэтому синтаксис Markdown полностью состоит из символов пунктуации, отобранных так, чтобы они выглядели похоже на то, что обозначают.

Например, звездочки вокруг слова действительно обращают на себя \*внимание\*. Списки в Markdown выглядят так, как и должны выглядеть списки. Даже цитаты выглядят как цитаты из электронной почты.

# Обычный текст

Абзацы в Markdown отделяются от окружающего текста пустой строкой.

"Жесткий" перевод строки выполняется `<br/>` с помощью HTML-элемента `<br/>`.

---

Абзацы в Markdown отделяются от окружающего текста пустой строкой.

“Жесткий” перевод строки выполняется с помощью HTML-элемента `<br/>`.

## Заголовки

# Заголовок 1

## Заголовок 2

#### Заголовок 4

---

# Заголовок 1

## Заголовок 2

#### Заголовок 4

# Шрифт

\*курсив\* и \_курсив\_

\*\*полужирный\*\* и \_\_полужирный\_\_

верхний индекс<sup>2</sup>

~~зачеркнутый~~

---

*курсив* и *курсив*

**полужирный** и **полужирный**

верхний индекс<sup>2</sup>

зачеркнутый



# Списки

- \* маркированный список (символы: \*,+, -)
- \* пункт 2
  - + подпункт 1 (в начале - табуляция)
  - + подпункт 2

1. нумерованный список
2. пункт 2
  - + подпункт 1

- ▶ маркированный список (символы: \*,+, -)
- ▶ пункт 2
  - ▶ подпункт 1 (в начале - табуляция)
  - ▶ подпункт 2

1. нумерованный список
2. пункт 2
  - ▶ подпункт 1

## Ссылки и рисунки

[ссылка] ([www.rstudio.com](http://www.rstudio.com))

! [Рисунок] ([images/rstudio\\_icon.png](#))

---

ССЫЛКА



Рисунок

Непосредственно после текста, к которому нужно сделать сноску, помещаем [Текст сноски.]

---

Непосредственно после текста, к которому нужно сделать сноску, помещаем<sup>1</sup>

---

<sup>1</sup>Текст сноски.

# Таблицы

Заголовок таблицы	Второй столбец
-----	-----
Ячейка 1	Ячейка 2
Ячейка 3	Ячейка 4

---

Заголовок таблицы	Второй столбец
Ячейка 1	Ячейка 2
Ячейка 3	Ячейка 4

## Цитаты

Цитаты оформляются с помощью символа >.

> Donec sit amet nisl. Aliquam semper ipsum sit amet velit.  
> id sem consetetuer libero luctus adipiscing.

*Donec sit amet nisl. Aliquam semper ipsum sit amet velit.*  
*Suspendisse id sem consetetuer libero luctus adipiscing.*

Предварительно отформатированный текст:

...

Donec sit amet nisl. Aliquam semper ipsum sit amet velit.  
Suspendisse id sem consetetuer libero luctus adipiscing.  
...

Donec sit amet nisl. Aliquam semper ipsum sit amet velit.  
Suspendisse id sem consetetuer libero luctus adipiscing.

# Pandoc

Pandoc — универсальный конвертер, преобразует документы из/в markdown, reStructuredText, textile, HTML, DocBook, LaTeX, MediaWiki markup, TWiki markup, OPML, Emacs Org-Mode, Txt2Tags, Microsoft Word docx, LibreOffice ODT, EPUB и Haddock.

Для каждого формата документов в Pandoc существуют шаблоны (template).

Шаблоны могут содержать переменные, их **можно настраивать**.

Значения переменных задаются в:

1. YAML-заголовке документа (метаданные);
2. командной строке при вызове Pandoc.

## Метаданные документа (YAML)

YAML (YAML Ain't Markup Language — “YAML — Не язык разметки”) — формат для описания конфигурационных файлов, похожий на языки разметки (JSON).

Простейший YAML-заголовок для документа HTML:

```
---  
title: "Sample"  
output: html_document  
---
```

Переменные шаблона: `title` и `output`.

Можно задавать пользовательские переменные, но их нужно включать в шаблон.

## Более сложный заголовок

```
---  
title: "Интеллектуальный анализ данных"  
subtitle: "Программа курса"  
author: "Храмов Д.А."  
date: "22.04.2017"  
output:  
  html_document:  
    toc: true  
---
```

Другие форматы вывода (output):

- ▶ word\_document
- ▶ beamer\_presentation
- ▶ pdf\_document
- ▶ ...



# Формулы

Может быть строчной  $i^2 = -1$  или выключной

\$\$

$$\int x^2 dx = \frac{x^3}{3} + C$$

\$\$

Может быть строчной  $i^2 = -1$  или выключной

$$\int x^2 dx = \frac{x^3}{3} + C$$

## Поддержка расширений Markdown

```
---  
title: "Habits"  
author: John Doe  
date: March 22, 2005  
output: md_document  
      variant: markdown+grid_tables  
---
```

# R Markdown

R Markdown (<http://rmarkdown.rstudio.com/>) — пакет R для подготовки документов и их преобразования в разные форматы (Pandoc внутри). Поддерживает создание документов

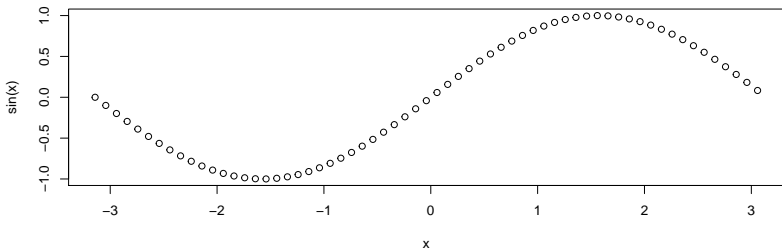
- ▶ HTML
- ▶ PDF
- ▶ MS Word
- ▶ Beamer (слайды в LaTeX)
- ▶ слайды на HTML5
- ▶ документы в стиле Тафти
- ▶ книг (bookdown)
- ▶ информационных панелей (flexdashboard)
- ▶ интерактивных веб-приложений (shiny)
- ▶ научных статей (rticles)
- ▶ веб-сайтов
- ▶ ...

## “Живой” код на R, Python, SQL, ...

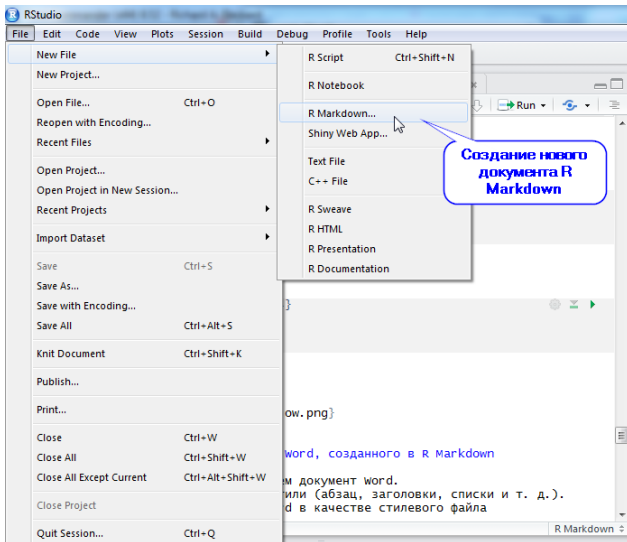
Вместо:

```
```r  
x <- seq(-pi,pi,.1)  
plot(x,sin(x))  
```
```

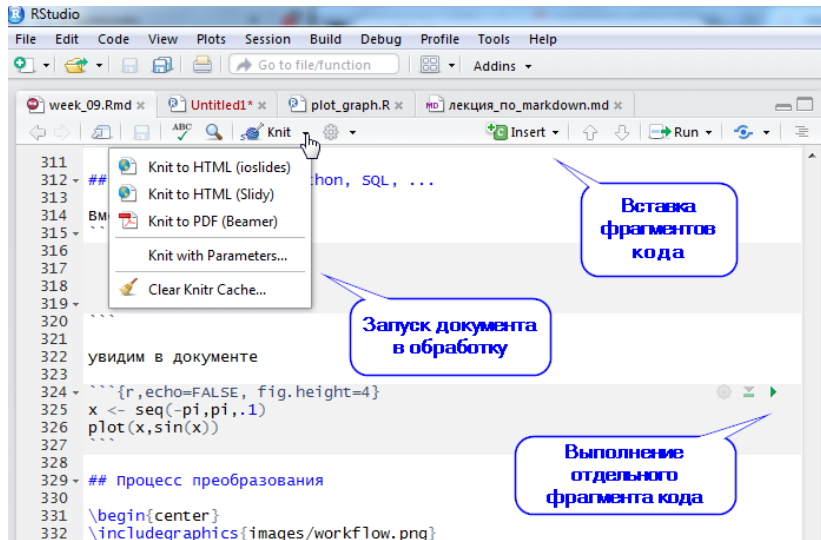
увидим в документе



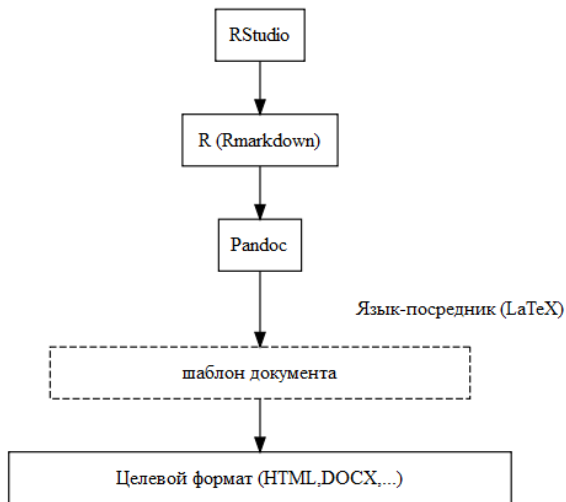
# RStudio: создать файл R Markdown



# RStudio: обработать документ R Markdown



# Процесс преобразования



# Документы в стиле Тафти

## Figures

### Margin Figures

Images and graphics play an integral role in Tufte's work. To place figures or tables in the margin you can use the `fig.margin` knitr chunk option. For example:

```
library(ggplot2)
qplot(Sepal.Length, Petal.Length, data = iris,
      color = Species)
```

Note the use of the `fig.cap` chunk option to provide a figure caption. You can adjust the proportions of figures using the `fig.width` and `fig.height` chunk options. These are specified in inches, and will be automatically scaled down to fit within the handout margin.

### Equations

You can also include  $\text{\LaTeX}$  equations in the margin by explicitly invoking the `marginfigure` environment.

Note the use of the `\caption` command to add additional text below the equation.

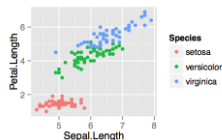


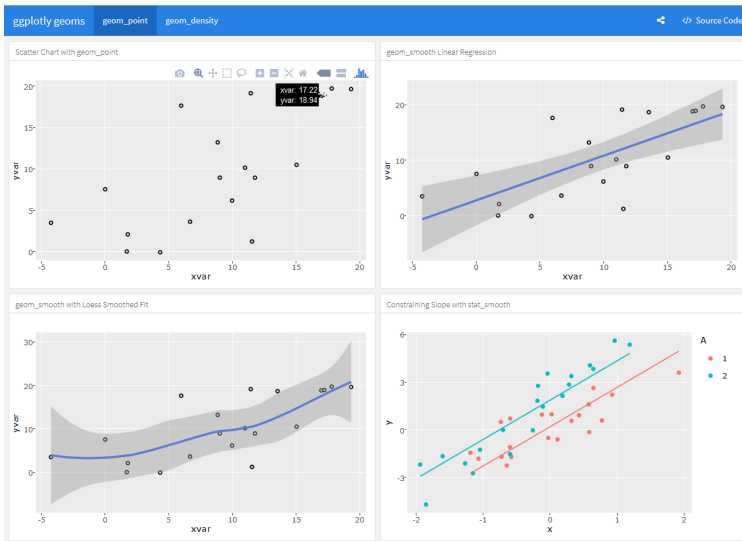
Figure 1: Sepal length vs. petal length, colored by species

$$\frac{d}{dx} \left( \int_0^x f(u) du \right) = f(x).$$

Figure 2: An equation



# Информационная панель (dashboard)



## Настройка стилей документа Word, созданного в R Markdown

1. С помощью R Markdown создаем документ Word.
2. Настраиваем в нем нужные стили (абзац, заголовки, списки и т. д.).
3. Сохраняем этот документ Word в качестве стилевого файла (reference\_docx).
4. Подключаем стилевой файл в документе .Rmd и используем настроенные стили

```
---  
title: "Test Report"  
author: "Your Name"  
date: "28.04.2017"  
output:  
  word_document:  
    reference_docx: word-styles-reference-01.docx  
---
```

# Оформление библиографии в R Markdown

YAML-заголовок документа:

```
---  
title: "Анализ методов даунскейлинга"  
output: html_document:  
bibliography: lst_downscaling.bib  
csl: ieee-with-url.csl  
---
```

- ▶ Название стилей оформления библиографии: [https://www.overleaf.com/learn/latex/Biblatex\\_citation\\_styles](https://www.overleaf.com/learn/latex/Biblatex_citation_styles)
- ▶ Источник стилей (нужно положить в папку с Rmd-файлом) — Official repository for Citation Style Language (CSL) citation styles: <https://github.com/Keniajin/styles>
- ▶ Оформление цитат в R Markdown: [https://crsh.github.io/papaja\\_man/writing.html# citations](https://crsh.github.io/papaja_man/writing.html# citations)

Создать bib-файл можно с помощью `rcrossref::cr_cn()`.

R Notebook – удобный инструмент для создания красивых аналитических отчетов, позволяет хранить вместе код, изображения, комментарии, формулы и графики. Это документ R Markdown “публикабельного” вида с исходником внутри.

Ближайший аналог: Jupyter Notebook (<http://jupyter.org/>).

## Ссылки по R Markdown

- ▶ **R Markdown Cheatsheet:** <https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf>
- ▶ **R Markdown Reference Guide:** <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>
- ▶ **Happy collaboration with Rmd to docx:**  
[http://rmarkdown.rstudio.com/articles\\_docx.html](http://rmarkdown.rstudio.com/articles_docx.html)
- ▶ *Xie Y., Allaire J. J., Golemund G.* R Markdown: The Definitive Guide