

Как быть в курсе современных  
концепций разработки в R

**MEDIA-TEL**  
by devoteam

04 Февраля 2022  
Шутов Илья



Innovative technology consulting for business

## Ключевые составляющие успешного решения

- Четкая постановка задачи.
  - Учет специфики предметной области.
  - Подходящие математические алгоритмы.
  - Хорошие входные данные.
  - Эффективная программная реализация.
  - Доходчивый способ презентации результатов пользователю.
- 
- В контексте R остановимся на вопросах реализации и алгоритмах.

# Источники знаний

- Классические труды
- Специализированные книги по R
- Материалы конференций
- Документация по пакетам, README
- Сайты контрибуторов
- Твиттер
- Персональные блоги
- Курсы

# Классические труды

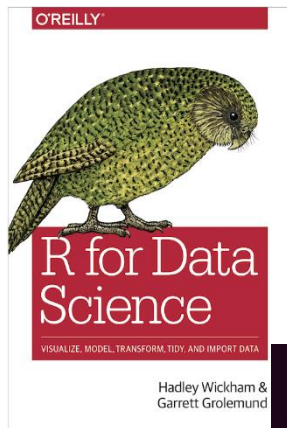
## Алгоритмы



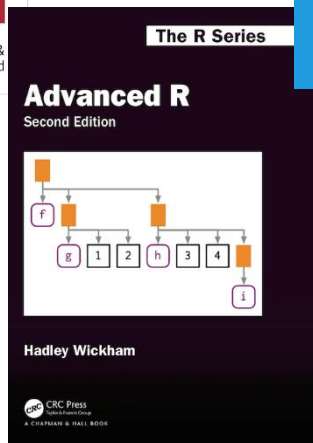
## Разработка



# Специализированные книги по R

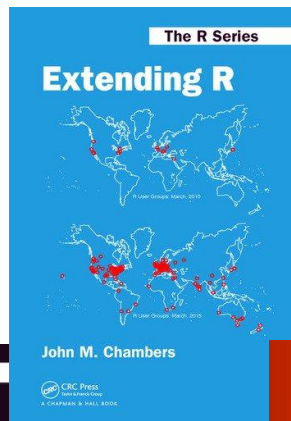


Hadley Wickham & Garrett Golemund



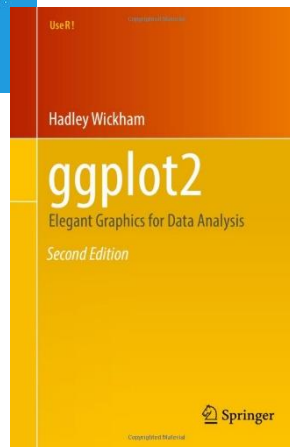
Hadley Wickham

CRC Press  
A CHAPMAN & HALL BOOK



John M. Chambers

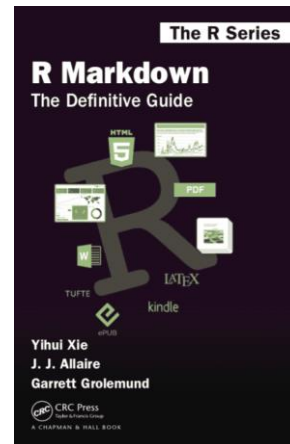
CRC Press  
A CHAPMAN & HALL BOOK



Hadley Wickham

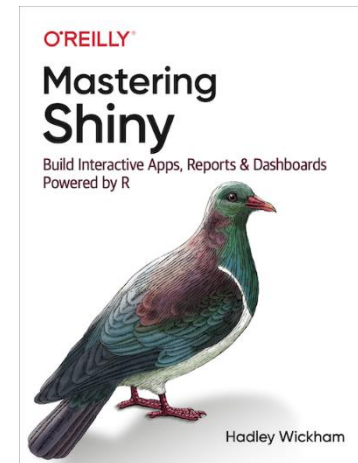
ggplot2  
Elegant Graphics for Data Analysis  
Second Edition

Springer



Yihui Xie  
J. J. Allaire  
Garrett Golemund

CRC Press  
A CHAPMAN & HALL BOOK



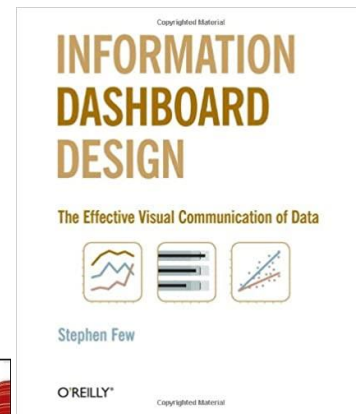
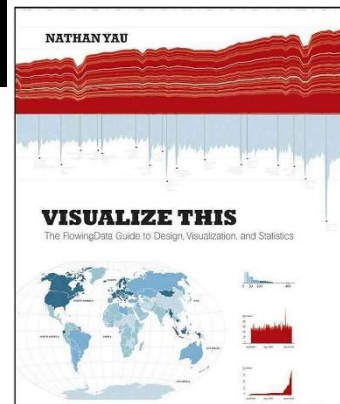
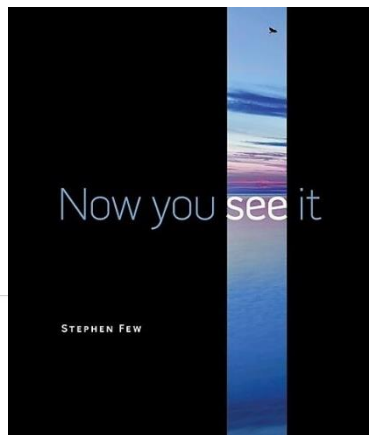
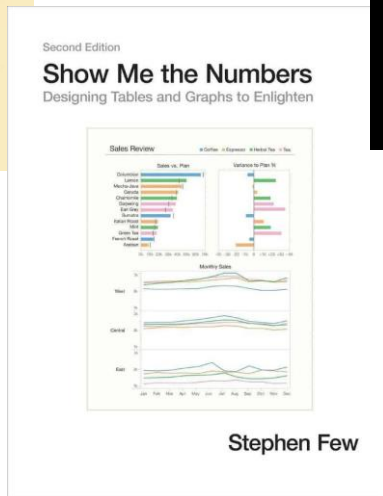
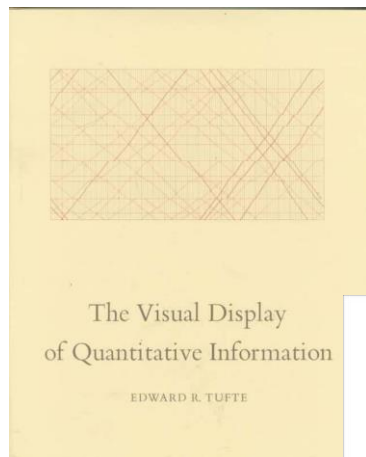
O'REILLY

Mastering  
Shiny

Build Interactive Apps, Reports & Dashboards  
Powered by R

Hadley Wickham

# Книги по визуализации данных



# The R Series



[◀ View All Book Series](#)

## BOOK SERIES

# Chapman & Hall/CRC The R Series

## About the Series

This book series reflects the recent rapid growth in the development and application of R, the programming language and software environment for statistical computing and graphics. R is now widely used in academic research, education, and industry. It is constantly growing, with new versions of the core software released regularly and more than 16,000 packages available. It is difficult for the documentation to keep pace with the expansion of the software, and this vital book series provides a forum for the publication of books covering many aspects of the development and application of R.

The scope of the series is wide, covering three main threads:

- Applications of R to specific disciplines such as biology, epidemiology, genetics, engineering, finance, and the social sciences.
- Using R for the study of topics of statistical methodology, such as linear and mixed modeling, time series, Bayesian methods, and missing data.

[Read More](#)

Series Titles

Authors

62 Series Titles

## Series Editors

### John M. Chambers

*Stanford University Stanford University  
Stanford University, California, USA Stanford  
University, California, USA*

### Torsten Hothorn

*Institut für Statistik, München, Germany*

### Duncan Temple Lang

*University of California, Davis, USA*

### Hadley Wickham

*RStudio, Houston, TX, USA*

## Attention Authors

Interested in proposing a book for a book series?

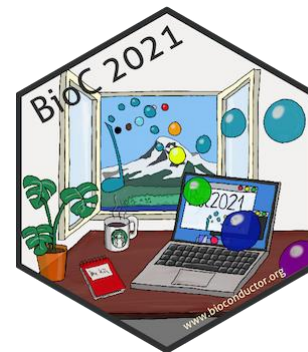
[Learn how to become a Taylor & Francis author »](#)

## Using R

*Get your copy of the latest CRC*

# Материалы конференций


Примеры крупных






# Документация по пакетам / Readme / News

master data.table / NEWS.md Go to file ...

 ben-schwen enable use of `by=.I` (#5235) ✓ Latest commit 8b257b8 on 9 Dec 2021 History

62 contributors  +41

2066 lines (1337 sloc) | 236 KB <> 📄 Raw Blame 🖨 📄 ✎ 🗑

If you are viewing this file on CRAN, please check [latest news on GitHub](#) where the formatting is also better.

Benchmarks are regularly updated: [here](#)

## data.table v1.14.3 (in development)

---

### NEW FEATURES

- `nafill()` now applies `fill=` to the front/back of the vector when `type="locf|nocb"`, #3594. Thanks to @ben519 for the feature request. It also now returns a named object based on the input names. Note that if you are considering joining and then using `nafill(..., type='locf|nocb')` afterwards, please review `roll= / rollends=` which should achieve the same result in one step more efficiently. `nafill()` is for when filling-while-joining (i.e. `roll= / rollends= / nomatch=`) cannot be applied.
- `mean(na.rm=TRUE)` by group is now GForce optimized, #4849. Thanks to the [h2oai/db-benchmark](#) project for spotting this issue. The 1 billion row example in the issue shows 48s reduced to 14s. The optimization also applies to type `integer64` resulting in a difference to the `bit64::mean.integer64` method: `data.table` returns a `double` result whereas `bit64` rounds the mean to the nearest integer.
- `fwrite()` now writes UTF-8 or native csv files by specifying the `encoding=` argument, #1770. Thanks to @shrekta for the request and the PR.
- `data.table()` no longer fills empty vectors with `NA` with warning. Instead a 0-row `data.table` is returned, #3727. Since `data.table()` is used internally by `.()`, this brings the following examples in line with expectations in most cases. Thanks to @shrekta for the suggestion and PR.

# Сайты значимых контрибуторов

## Примеры

Tidyverse Packages Blog Learn Help Contribute



R packages for data science

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

OpenSci About Blog Projects Packages Community Resources Search

### Transforming science through open data, software & reproducibility

We help develop R packages for the sciences via community driven learning, review and maintenance of contributed software in the R ecosystem

Browse Our Packages Our Mission

FREE MASTERCLASS TESTIMONIALS COURSES BLOG CONTACT LOG IN

### Go from zero-experience to full-time data scientist WITHOUT spending years learning data science.

#### R-TRACK 5-Course R-Track System

Business Analysis with R (6 Weeks) | Advanced ML & Time Series (6 Weeks) | Web Apps & Shiny Developer (6 Weeks)

Data Science Foundations (8 Weeks) | Machine Learning & Time Series (10 Weeks) | Web-Application Development (12 Weeks)

Business Science University Track

tidyverts home tidyverts

### Tidy tools for time series.

#### PACKAGES

- tibble: Temporal data frames and tools
- fable: Tidy forecasting
- feats: Feature extraction and statistics

Learn how H2O.ai is responding to LogG

## H2O

The #1 open-source machine learning platform for the enterprise

Download Latest

Features & Benefits How It Works Enterprise Support H2O Automate

Overview FAQ Blog Get Arrow Documentation Subgroups Community API Links

## APACHE ARROW

A cross-language development platform for in-memory analytics

Star 8.5k Follow @ApacheArrow 8.432 followers

Appsilion Home About Data for Good R Shiny Computer Vision Careers Blog Let's Talk

### World-Class Enterprise Shiny Dashboards

LET'S TALK

Futureverse Packages & Docs Blog Usage Statistics Quality Talks Publications Roadmap About

## Futureverse

A Unifying Parallelization Framework in R for Everyone

The future framework makes it easy to parallelize existing R code - often with only a minor change of code. The goal is to lower the barriers so that anyone can safely speed up their existing R code in a worry-free manner. As it is a cross-platform solution that requires no additional setup or technical skills, anyone can be up and running within a few minutes.

The future framework removes common hurdles and protects against pitfalls that follow from adding parallelization. Instead of leaving it to the developers and end-users to be aware of and deal with these problems, they are handled at the core of the highly-validated future ecosystem. Just as with sequential R code, output, messages, warnings, and errors work as expected and can be handled using traditional R techniques - regardless how the code is parallelized.

It is designed so that you as a developer can stay with your favorite coding style, may it be base R or tidyverse. If you like base R apply() there is a corresponding future\_apply() in the future.apply package and if you like tidyverse purrr

## Твиттер

Да-да, все западные персоны там!

Appsilon [[@appsilon](#)]; Wes McKinney [[@wesmckinn](#)]; Daily R Cheatsheets [[@daily\\_r\\_sheets](#)]; easystats [[@easystats4u](#)]; One R Package a Day [[@RLangPackage](#)]; R Function A Day [[@rfunctionaday](#)]; Garrick Aden-Buie [[@grrrrck](#)]; R Consortium [[@RConsortium](#)]; Matt Dancho (Business Science) [[@mdancho84](#)]; Arun Srinivasan [[@arun\\_sriniv](#)]; rdatatable [[@rdatatable](#)]; Mark van der Loo [[@markvdloo](#)]; Miles McBain [[@MilesMcBain](#)]; Joe Cheng [[@jcheng](#)]; The R Foundation [[@\\_R\\_Foundation](#)]; Bioconductor [[@Bioconductor](#)]; Max Kuhn [[@topepos](#)]; b0B Ruđis [[@hrbrmstr](#)]; We are R-Ladies [[@WeAreRLadies](#)]; Colin Fay [[@\\_ColinFay](#)]; Julia Silge [[@juliasilge](#)]; R-Ladies Global [[@RLadiesGlobal](#)]; Thomas Lin Pedersen [[@thomasp85](#)]; Dean Attali [[@daattali](#)]; Jim Hester [[@jimhester\\_](#)]; Dirk Eddelbuettel [[@eddelbuettel](#)]; rOpenSci [[@rOpenSci](#)]; Karl Broman [[@kwbroman](#)]; One R Tip a Day [[@RLangTip](#)]; Yihui Xie [[@xieyihui](#)]; RStudio Tips [[@rstudiotips](#)]; Garrett Golemund [[@StatGarrett](#)]; RStudio Daily NEWS [[@rstudiodaily](#)]; Hadley Wickham [[@hadleywickham](#)]; Jenny Bryan [[@JennyBryan](#)]; David Robinson [[@drob](#)]; Mara Averick [[@dataandme](#)]; Noam Ross [[@noamross](#)]; Rstats [[@rstatstweet](#)]; Shiny Developer Series [[@shinydevseries](#)]; Tom Mock [[@thomas\\_mock](#)]; Daniel Lemire [[@lemire](#)]; R Weekly [[@rweekly\\_org](#)]; H2O.ai [[@h2oai](#)]; Winston Chang [[@winston\\_chang](#)]; Matt Dowle [[@MattDowle](#)]

**coursera**



**codewars**



**ООО «Медиа-тел»**  
**(Devoteam Group)**

123242, Москва, ул.  
Зоологическая, д.2  
Телефон: +7 499 272 76 58  
E-mail: info@media-tel.ru

**51**  
консультант

**2**  
офиса

с  
**2016**