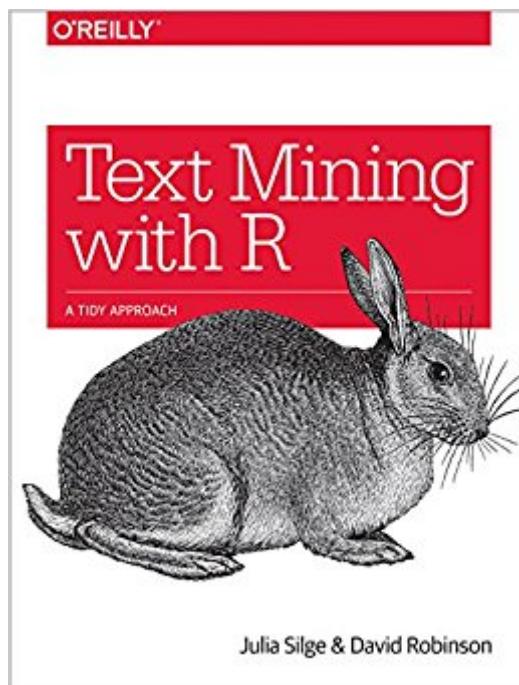


The book was found

# Text Mining With R: A Tidy Approach



## Synopsis

Much of the data available today is unstructured and text-heavy, making it challenging for analysts to apply their usual data wrangling and visualization tools. With this practical book, you'll explore text-mining techniques with tidytext, a package that authors Julia Silge and David Robinson developed using the tidy principles behind R packages like ggraph and dplyr. You'll learn how tidytext and other tidy tools in R can make text analysis easier and more effective. The authors demonstrate how treating text as data frames enables you to manipulate, summarize, and visualize characteristics of text. You'll also learn how to integrate natural language processing (NLP) into effective workflows. Practical code examples and data explorations will help you generate real insights from literature, news, and social media. Learn how to apply the tidy text format to NLP:

- Use sentiment analysis to mine the emotional content of text
- Identify a document's most important terms with frequency measurements
- Explore relationships and connections between words with the ggraph and tidytext packages
- Convert back and forth between R's tidy and non-tidy text formats
- Use topic modeling to classify document collections into natural groups
- Examine case studies that compare Twitter archives, dig into NASA metadata, and analyze thousands of Usenet messages

## Book Information

Paperback: 194 pages

Publisher: O'Reilly Media; 1 edition (July 2, 2017)

Language: English

ISBN-10: 1491981652

ISBN-13: 978-1491981658

Product Dimensions: 6.9 x 0.4 x 9.1 inches

Shipping Weight: 12.6 ounces (View shipping rates and policies)

Average Customer Review: 5.0 out of 5 stars 5 customer reviews

Best Sellers Rank: #28,825 in Books (See Top 100 in Books) #5 in Books > Computers & Technology > Computer Science > AI & Machine Learning > Natural Language Processing #15 in Books > Textbooks > Computer Science > Artificial Intelligence #23 in Books > Computers & Technology > Databases & Big Data > Data Mining

## Customer Reviews

[View larger](#) From the Preface If you work in analytics or data science, like we do, you are familiar with the fact that data is being generated all the time at ever faster rates. (You may even be

a little weary of people pontificating about this fact.) Analysts are often trained to handle tabular or rectangular data that is mostly numeric, but much of the data proliferating today is unstructured and text-heavy. Many of us who work in analytical fields are not trained in even simple interpretation of natural language. We developed the `tidytext` (Silge and Robinson 2016) R package because we were familiar with many methods for data wrangling and visualization, but couldn't easily apply these same methods to text. We found that using tidy data principles can make many text mining tasks easier, more effective, and consistent with tools already in wide use. Treating text as data frames of individual words allows us to manipulate, summarize, and visualize the characteristics of text easily and integrate natural language processing into effective workflows we were already using. This book serves as an introduction of text mining using the `tidytext` package and other tidy tools in R. The functions provided by the `tidytext` package are relatively simple; what is important are the possible applications. Thus, this book provides compelling examples of real text mining problems.

**About this book** This book is focused on practical software examples and data explorations. There are few equations, but a great deal of code. We especially focus on generating real insights from the literature, news, and social media that we analyze. We don't assume any previous knowledge of text mining. Professional linguists and text analysts will likely find our examples elementary, though we are confident they can build on the framework for their own analyses. We do assume that the reader is at least slightly familiar with `dplyr`, `ggplot2`, and the `%>%` `pipe` operator in R, and is interested in applying these tools to text data. For users who don't have this background, we recommend books such as *R for Data Science*. We believe that with a basic background and interest in tidy data, even a user early in their R career can understand and apply our examples.

**Topics this book does not cover**

- Clustering, classification, and prediction
- Word embedding
- More complex tokenization
- Languages other than English

Julia Silge is a data scientist at Stack Overflow; her work involves analyzing complex datasets and communicating about technical topics with diverse audiences. She has a PhD in astrophysics and loves Jane Austen and making beautiful charts. Julia worked in academia and ed tech before moving into data science and discovering the statistical programming language R. David Robinson is a data scientist at Stack Overflow with a PhD in Quantitative and Computational Biology from Princeton University. He enjoys developing open source R packages, including `broom`, `ggridge`, `fuzzyjoin` and `widyr`, as well as blogging about statistics, R, and text mining on his blog, *Variance Explained*.

Small book but crammed with info. I have stuck my nose in it everyday since getting it and have loved every minute. I'm glad I bought the hard copy as I can't get on WiFi at work and since it's such a lightweight book it's very easy to carry. Pro tip: I would recommend brushing up on your tidyverse before diving in. This should help make understanding what the authors are trying to do much easier.

This is a delightful book on practical textual mining. While it is not targeted at someone just starting out in R, anyone with intermediate knowledge will find this a precious gold mine! :-) It details how to compare texts from different authors, how to graph word connections, word correlations, apply clustering to texts, and so on. It's a short book, but it's worth its weight in gold. I'm already reading this a second time!

This book was enormously helpful for a new research project that required significant text analysis. The writing and sample code are very clear and easy to follow. If you've never done text analysis and just want to get started, this is an excellent way to begin. If you have worked with other text analysis methods or packages, I still recommend the book highly as the tidytext / tidyverse approach is sufficiently different that it's useful to be guided through the underlying ideas and code. For intermediate to advanced users, I also recommend the tidytext approach for speed and elegance. In the past, I've played with a variety of other R packages for text analysis and read related academic papers but still hit serious road bumps when code seemed to freeze with large data sets or the methods were not well explained. I found the tidytext package and approach to be impressively fast and reliable, even with a large corpus. Finally, for faculty, I recommend the book for instruction and am planning to use it myself in the undergrad applied stats class I teach in the spring.

Lots of good info for new R users.

Great book: well explained and easy to follow examples. The tidytext methodology is simple to use and elegant.

[Download to continue reading...](#)

Text Mining with R: A Tidy Approach  
Mercury, Mining, and Empire: The Human and Ecological Cost of Colonial Silver Mining in the Andes  
Discovering Knowledge in Data: An Introduction to Data

Mining (Wiley Series on Methods and Applications in Data Mining) Handbook of Biblical Hebrew: An Inductive Approach Based on the Hebrew Text of Esther (An Inductive Approach Based on the Hebrew Text of Esther, 2 Vols. in 1) TIDYING UP: ORGANIZE & DECLUTTER 10 MINUTE CLEANING HABIT:SIMPLE STRESS-FREE HABIT TO SPARK JOY WHEN YOU TIDY UP The Art of Clean Up: Life Made Neat and Tidy Speed Cleaning To A Healthier Life. Speed Cleaning Techniques And Tips: How To Properly and quickly Tidy a Unclean House In Minutes instead Of Hours R for Data Science: Import, Tidy, Transform, Visualize, and Model Data Texting Women: 7 Simple Steps From Text to Sex (Flirty Texts, Texting Girls, How To Text Girls, Art Seduction, How to Seduce a Woman, Funny Text, Pick Up Women, Funny Pick Up Lines, Picking Up Women) How To Approach Women: Fearless Approach, Approach Women And Life With Unstoppable Confidence (how to get a girl) Learn German: Parallel Text - Easy, Funny Stories (German - English) - Bilingual (Learning German with Parallel Text Book 1) Learn German III: Parallel Text - Easy Stories (German - English) Bilingual - Dual Language (Learning German with Parallel Text 3) (German Edition) Learn German II: Parallel Text - Easy Stories (English - German), Dual Language - Bilingual (Learning German with Parallel Text Book 2) Learn German IV: Parallel Text - Easy Stories (English - German) (Learning German with Parallel Text Book 4) Learn Italian III: Parallel Text - Short Stories (Italian - English) (Learn Italian with Parallel Text Book 3) Holy Bible Text Edition NLT: New Living Translation (Text Edition: Full Size) Live Text Field Exp Edition ~ New (live text) Revelations of Divine Love (Short Text and Long Text) Memes: Memes - World's Funniest Text Messages Book 5 (Memes, Funny books, Funny Texts, Funniest books, Funniest text) Eine Nacht in Venedig. Komische Oper ... A Night in Venice. Comic opera ... Vollstndiger Klavierauszug mit Text. Vocal score with English Text. Ger.& Eng

[Contact Us](#)

[DMCA](#)

[Privacy](#)

[FAQ & Help](#)