1. Text won’t be tidy at all stages of an analysis, and it is important to be able to convert back and forth between tidy and nontidy formats:

Chapter 1 outlines the tidy text format and the unnest\_tokens() function. It also introduces the gutenbergr and janeaustenr packages, which provide useful liter‐ ary text datasets that we’ll use throughout this book.

• Chapter 2 shows how to perform sentiment analysis on a tidy text dataset using the sentiments dataset from tidytext and inner\_join() from dplyr.

* Chapter 3 describes the tf-idf statistic (term frequency times inverse document frequency), a quantity used for identifying terms that are especially important to a particular document.
* Chapter 4 introduces n-grams and how to analyze word networks in text using the widyr and ggraph packages.

Text won’t be tidy at all stages of an analysis, and it is important to be able to convert back and forth between tidy and nontidy formats:

* Chapter 5 introduces methods for tidying document-term matrices and Corpus objects from the tm and quanteda packages, as well as for casting tidy text data‐ sets into those formats.
* Chapter 6 explores the concept of topic modeling, and uses the tidy() method to interpret and visualize the output of the topicmodels package.

We conclude with several case studies that bring together multiple tidy text mining approaches we’ve learned:

* Chapter 7 demonstrates an application of a tidy text analysis by analyzing the authors’ own Twitter archives. How do Dave’s and Julia’s tweeting habits com‐ pare?

Spacy R image recognition

**Kapitel 2**

**Sentiment datasæts/lexicons:**

The three general-purpose lexicons are:

* AFINN from Finn Årup Nielsen
* Bing from Bing Liu and collaborators
* NRC from Saif Mohammad and Peter Turney