

Einführung

Vielleicht sind Sie wie ich der Meinung, dass der beste Weg, etwas zu lernen, darin besteht ist, es zu tun. Lesen Sie nicht nur einfach über etwas – praktizieren Sie es. Wenn Sie Bauarbeiter sein wollen, dann bauen Sie. Wenn Sie Autor sein wollen, dann schreiben Sie. Wenn Sie Tischler sein wollen, dann tischlern Sie.

Dieses Buch basiert auf dem Prinzip »Learning by Doing«. Mein Ziel ist, dass Sie Ihre R-Kenntnisse und Fähigkeiten erweitern, in dem Sie R verwenden, um Projekte aus verschiedenen Bereichen zu erstellen. Gleichzeitig lernen Sie etwas über die Bereiche und Wissensgebiete, aus denen die Projekte stammen.

Trotz dieser noblen Absichten kann ein Buch wie dieses in eine Falle stolpern. Es kann schnell zu einem Kochbuch werden: Verwenden Sie dieses Package, verwenden Sie jene Funktion, erstellen Sie eine Grafik – und Simsalabim ist es Zeit, weiterzugehen.

Ein solches Buch wollte ich nicht schreiben. Stattdessen führen sämtliche Kapitel ab Teil 2 (das ist die Stelle, an der es mit den Projekten losgeht) Sie nicht einfach nur durch die Projekte. Stattdessen erhalten Sie zuerst etwas Hintergrundwissen zum Themenbereich und dann arbeiten Sie (in den meisten Kapiteln) ein kleines, übersichtliches Projekt aus diesem Bereich durch, damit Sie erste Erfahrungen sammeln können.

Damit jedoch nicht genug: Am Ende jedes Kapitels finden Sie einen *Projektvorschlag*, der Sie ermutigen soll, das neu erworbene Wissen anzuwenden. Für alle diese Projekte liefere ich Ihnen gerade so viele Informationen, dass Sie direkt loslegen können. (Falls nötig, gehören hierzu auch Tipps über mögliche Stolperfallen.)

Im Verlauf des Buches finden Sie außerdem *Schnelle Projektvorschläge*. Diese basieren auf kleinen Veränderungen an Projekten, die Sie bereits fertiggestellt haben, und stellen zusätzliche Herausforderungen an Ihre zunehmenden Fähigkeiten dar.

Eines noch: Jeder der behandelten Themenbereiche könnte der Ausgangspunkt für ein eigenes Buch sein, und daher kann ich immer nur an der Oberfläche kratzen. Kapitel 17 verweist auf Ressourcen mit weiterführenden Informationen.

Über dieses Buch

Ich habe dieses Buch in sechs Teile aufgeteilt.

Teil 1: Das Handwerkszeug

In diesem Teil dreht sich alles um R und RStudio. Ich erörtere R-Funktionen, Strukturen und Packages. Außerdem zeige ich Ihnen, wie Sie eine Vielzahl von Diagrammtypen erstellen.

Teil 2: Interaktion mit dem Anwender

Die Projekte beginnen in Teil 2, in dem Sie lernen, interaktive Anwendungen zu erstellen, die direkt auf Benutzereingaben reagieren. Ich stelle das Package `shiny` vor, das Webbrowser unterstützt und das Package `shinydashboard`, mit dem Sie Dashboards erstellen können.

Teil 3: Maschinelles Lernen

Dies ist der umfangreichste Teil des Buches. Zuerst erzähle ich Ihnen etwas über das Machine Learning Repository der University of California-Irvine, in dem Sie die Datensätze für die Projekte finden. Außerdem stelle ich Ihnen das Package `rattle` vor, mit dem Sie Anwendungen für das maschinelle Lernen erstellen können. Die Projekte behandeln Entscheidungsbäume, Random Forests, Support Vector Machines, K-Means-Clustering und neuronale Netzwerke.

Teil 4: Große Datensätze

Die beiden Projekte in Teil 4 untersuchen weit größere Datensätze als die, denen Sie in den vorderen Teilen des Buches begegnen. Das erste Projekt ist eine Analyse der Kundensegmentierung von mehr als 300.000 Kunden eines Online-Shops. Eine darauf aufbauende Analyse setzt Techniken des maschinellen Lernens ein.

Im zweiten Projekt analysieren Sie einen Datensatz mit Daten zu mehr als 500.000 Flügen.

Teil 5: Karten und Bilder

Auch in Teil 5 gibt es zwei Projekte. Das erste Projekt stellt die Lage (neben anderen Informationen) der Flughäfen eines der US-Bundesstaaten auf einer Karte dar. Das zweite Projekt zeigt Ihnen, wie Sie eine animierte Grafik und eine statische, nicht animierte Grafik miteinander kombinieren können.

Teil 6: Der Top-Ten-Teil

Das erste Kapitel in Teil 6 liefert Informationen über nützliche Packages, die bei Ihren zukünftigen Projekten hilfreich sein können. Im zweiten Kapitel erfahren Sie, wo Sie mehr über die Themenbereiche dieses Buches lernen können.

Was Sie nicht lesen müssen

In jedem Lehrbuch finden Sie jede Menge Informationen. Da ist auch dieses Buch keine Ausnahme. Ich habe versucht, nur Nützliches in das Buch aufzunehmen. Das ist mir jedoch nicht immer gleich gut gelungen. Wenn Sie also an einem Thema nicht besonders interessiert sind, brauchen Sie die Abschnitte, die mit dem Symbol »Vorsicht Technik!« gekennzeichnet sind, und die grauen Kästen nicht zu lesen.

Törichte Annahmen über den Leser

Bei diesem Buch setze ich Folgendes voraus:

- ✓ Sie kennen sich mit Windows oder dem Mac aus. Ich werde nicht erläutern, wie man auf Elemente zeigt, klickt, etwas auswählt und so weiter.
- ✓ Sie können R und RStudio installieren (in Kapitel 1 zeige ich Ihnen, wie das geht) und damit die Beispiele nachvollziehen. Ich verwende die Windows-Version von RStudio. Sie sollten aber nicht auf Probleme stoßen, wenn Sie einen Mac verwenden.

Symbole, die in diesem Buch verwendet werden

Wie in allen Büchern der Dummies-Reihe finden Sie auch in diesem Buch überall Symbole. Dabei handelt es sich um kleine Bildchen am Seitenrand, anhand derer Sie erkennen können, worum es in dem Abschnitt daneben geht.



Dieses Symbol ist das Zeichen für einen Tipp, einen Hinweis oder eine einfache Lösung, um Ihnen die Arbeit zu erleichtern.



Dieses Symbol ist ein Zeichen für zeitlose Weisheiten, die Sie auch lange, nachdem Sie das Buch gelesen haben, noch brauchen werden.



Achten Sie auf die Informationen neben diesem Symbol. Es weist darauf hin, dass Sie etwas besser nicht tun sollten, wenn Sie Ihre Arbeit nicht zunichtemachen wollen.



Wie bereits im Abschnitt »Was Sie nicht lesen müssen« weiter vorne erwähnt, steht dieses Symbol für Material, das Sie überspringen können, wenn Ihnen der Inhalt zu technisch vorkommt. (Ich habe versucht, dies auf ein Minimum zu beschränken.)

Wie es weitergeht

Sie können mit dem Lesen des Buchs an jeder beliebigen Stelle beginnen, aber hier sind noch ein paar Tipps. Sie möchten die Grundlagen von R und RStudio kennenlernen? Alles, was Sie dazu brauchen, finden Sie in den Kapiteln 1 und 2. Sie möchten lieber mit der Erstellung von Diagrammen beginnen? Dann schlagen Sie Kapitel 3 auf. Alles andere finden Sie im Inhaltsverzeichnis oder im Stichwortverzeichnis.

Wenn Sie ein Buch lieber von vorne nach hinten lesen, dann blättern Sie bitte um ...