

R für Data Science

Daten importieren, bereinigen, umformen
und visualisieren

DAS INHALTS- VERZEICHNIS

» Hier geht's
direkt
zum Buch

Einführung	13
-------------------------	-----------

Teil I Gesamtbild

1 Datenvisualisierung	29
Einführung	29
Erste Schritte	30
Zielsetzung	32
ggplot2-Aufrufe	42
Verteilungen visualisieren	42
Beziehungen visualisieren	46
Diagramme speichern	53
Häufige Probleme	54
Zusammenfassung	55
2 Workflow: Grundlagen	57
Grundlagen der Codierung	57
Kommentare	58
Was macht einen Namen aus?	59
Funktionen aufrufen	60
Zusammenfassung	62
3 Datentransformation	63
Einführung	63
Voraussetzungen	63
Zeilen	66
Spalten	71
Die Pipe	76

Gruppen	78
Fallstudie: Aggregate und Stichprobengröße	85
Zusammenfassung	87
4 Workflow: Programmierstil	89
Namen	90
Leerzeichen	90
Pipes	91
ggplot2	93
Abschnittskommentare	93
Übungen	94
Zusammenfassung	94
5 Datenaufbereitung	95
Einführung	95
Aufbereitete Daten	96
Daten länger machen	99
Daten breiter machen	108
Zusammenfassung	112
6 Workflow: Skripte und Projekte	113
Skripte	113
Projekte	117
Übungen	122
Zusammenfassung	122
7 Daten importieren	125
Einführung	125
Daten aus einer Datei lesen	125
Spaltentypen steuern	132
Daten aus mehreren Dateien einlesen	135
In eine Datei schreiben	136
Dateneingabe	138
Zusammenfassung	139
8 Workflow: Hilfe abrufen	141
Google ist Ihr Freund	141
Ein Reprex erstellen	142
Sich selbst einbringen	144
Zusammenfassung	144

Teil II Visualisieren

9	Datenvisualisierung	147
	Einführung	147
	Ästhetische Zuordnungen	148
	Geometrische Objekte	152
	Facetten	158
	Statistische Transformationen	160
	Positionsanpassungen	164
	Koordinatensysteme	168
	Die geschichtete Grammatik der grafischen Darstellung	170
	Zusammenfassung	171
10	Explorative Datenanalyse	173
	Einführung	173
	Fragen	174
	Variation	175
	Ungewöhnliche Werte	179
	Kovariation	182
	Muster und Modelle	192
	Zusammenfassung	195
11	Kommunikation	197
	Einführung	197
	Beschriftungen	198
	Anmerkungen	200
	Skalen	205
	Themen	220
	Layout	223
	Zusammenfassung	227

Teil III Transformieren

12	Logische Vektoren	231
	Einführung	231
	Vergleiche	232
	Boolesche Algebra	236
	Zusammenfassungen	239
	Bedingte Transformationen	242
	Zusammenfassung	246

13 Zahlen	247
Einführung	247
Zahlen erzeugen	247
Zähler	248
Numerische Transformationen	250
Zahlen in Bereiche aufteilen	256
Allgemeine Transformationen	257
Numerische Zusammenfassungen	261
Zusammenfassung	267
14 Strings	269
Einführung	269
Einen String erzeugen	270
Viele Strings aus Daten erstellen	272
Daten aus Strings extrahieren	275
Buchstaben	281
Nicht englischer Text	283
Zusammenfassung	286
15 Reguläre Ausdrücke	287
Einführung	287
Muster – Grundlagen	288
Wichtige Funktionen	290
Details zu Mustern	295
Mustersteuerung	302
Praxis	304
Reguläre Ausdrücke an anderen Stellen	309
Zusammenfassung	311
16 Faktoren	313
Einführung	313
Faktoren-Basics	313
General Social Survey	315
Faktorreihenfolge ändern	316
Faktorlevels ändern	321
Geordnete Faktoren	323
Zusammenfassung	324
17 Datum und Uhrzeit	325
Einführung	325
Datums-/Zeitwerte erzeugen	326
Datums-/Zeitkomponenten	332

Zeiträume	339
Zeitzone n	343
Zusammenfassung	345
18 Fehlende Werte	347
Einführung	347
Explizit fehlende Werte	347
Implizit fehlende Werte	349
Faktoren und leere Gruppen	352
Zusammenfassung	355
19 Verknüpfungen	357
Einführung	357
Schlüssel	358
Grundlegende Verknüpfungen	363
Wie funktionieren Verknüpfungen?	369
Nicht-Gleichheitsverknüpfungen	375
Zusammenfassung	381

Teil IV Importieren

20 Tabellenkalkulationen	385
Einführung	385
Excel	385
Google Sheets	398
Zusammenfassung	403
21 Datenbanken	405
Einführung	405
Datenbankgrundlagen	406
Mit einer Datenbank verbinden	406
Grundlagen von dbplyr	409
SQL	411
Übersetzung von Funktionen	420
Zusammenfassung	423
22 Arrow	425
Einführung	425
Die Daten erhalten	426
Ein Datenset öffnen	426
Das Parquet-Format	428

Das Paket dplyr mit Arrow verwenden	430
Das Paket dbplyr mit Arrow verwenden	432
Zusammenfassung	433
23 Hierarchische Daten	435
Einführung	435
Listen	436
Verschachtelung beseitigen	440
Fallstudien	444
JSON	452
Zusammenfassung	456
24 Web-Scraping	457
Einführung	457
Ethische und rechtliche Anmerkungen zum Scraping	458
HTML-Grundlagen	460
Daten extrahieren	461
Die richtigen Selektoren finden	466
Alles zusammen	467
Dynamische Sites	472
Zusammenfassung	472

Teil V Programmieren

25 Funktionen	475
Einführung	475
Vektorfunktionen	476
Dataframe-Funktionen	482
Diagrammfunktionen	489
Stil	495
Zusammenfassung	496
26 Iteration	499
Einführung	499
Mehrere Spalten ändern	500
Mehrere Dateien lesen	509
Mehrere Ausgaben speichern	518
Zusammenfassung	523

27 Ein Praxisleitfaden zu Basis-R	525
Einführung	525
Mehrere Elemente mit [auswählen	526
Ein einzelnes Element mit \$ und [[auswählen	530
Familie der apply-Funktionen	533
for-Schleifen	535
Diagramme	537
Zusammenfassung	538

Teil VI Kommunizieren

28 Quarto	541
Einführung	541
Quarto-Grundlagen	542
Visueller Editor	545
Quelltexteditor	547
Codeblöcke	549
Bilder	553
Tabellen	557
Zwischenspeichern (Caching)	558
Fehlersuche	560
YAML-Header	561
Workflow	564
Zusammenfassung	566
29 Quarto-Formate	567
Einführung	567
Ausgabeoptionen	567
Dokumente	568
Präsentationen	569
Interaktivität	569
Websites und Bücher	572
Andere Formate	573
Zusammenfassung	573
Index	575