

TINU SCHNEIDER

Datenanalyse

Martin Schneider
Dipl. Umwelt-Ing. ETH
Gartenstrasse 11
CH-3600 Thun

Telefon +41 (0)31 301 00 03
Mobil +41 (0)76 445 83 01
schneider@tinuschneider.ch
www.tinuschneider.ch

Demo R-Bookdown

Am Beispiel Bevölkerungsentwicklung 1991 – 2045

Verfasst von Tinu Schneider, Scharans

Erstellt am 18. März 2022

Inhalt

Impressum	2
1 Einleitung bookdown	3
1.1 Textauszeichnung	3
1.2 Weitere Möglichkeiten	4
2 Abbildungen	4
3 Tabellen	5
4 Verweise innerhalb des Textes	6
4.1 Referenzen	6
4.2 Daten automatisch in Text	6
5 Fazit	7
Literatur	7
Anhang	8
A Rohdaten	8

Impressum

Dieses Dokument wird mit [bookdown](#) auf Basis der Programmiersprache [R](#) erstellt.

Dazu werden die aktuellen Rohdaten aus mehreren Exceldateien automatisch eingelesen und daraus zusammen mit dem Fliesstext aus der Quelldatei ein HTML- und ein PDF-Dokument erzeugt.

Der Bericht als PDF ist unter diesem Link verfügbar

www.tinuschneider.ch/demo-bookdown/Demo_R-bookdown.pdf

Eine zugehörige PowerPoint-Präsentation findet sich hier

www.tinuschneider.ch/demo-bookdown/demo_bookdown_Praesentation.pptx

Und der Quellcode ist bei GitHub https://github.com/tinu-schneider/Demo_Bookdown

Autor

Tinu Schneider

Linda 3

7412 Scharans

1 Einleitung bookdown

Wissenschaftliche Texte stellen gewisse Anforderungen an Format und Inhalt, wir wollen Fliesstext mit Daten, Grafiken, Tabellen und Verweisen mischen. Eine gute Lösung ist das Textsatzprogramm *LaTeX*, dies ist aber aufwändig zu lernen. Eine passable Lösung ist natürlich Word, hier fehlt aber die Automatisierung, wenn z.B. die Daten geändert haben und insbesondere die Zuverlässigkeit im alltäglichen Gebrauch.

Mit R und bookdown können wir die Datenanalyse und das Erstellen eines Berichtes in einem einzigen Dokument (Quelldatei) verknüpfen und integrieren damit immer die aktuellen Daten, Grafiken und Tabellen in den Bericht.

Die hier verwendeten zwei Datensätze stammen vom Bundesamt für Statistik (BfS) und sind im Anhang [A](#) aufgeführt. Es sind die Daten der Bevölkerung in der Schweiz von 1991–2014 und drei Szenarien zur zukünftigen Entwicklung ab 2015.

Aus der bookdown-Quelldatei können wir eine interaktive HTML-Version und eine ansprechend gesetzte PDF-Version generieren¹.

Wir zeigen hier die wichtigsten Möglichkeiten von bookdown.

1.1 Textauszeichnung

Um Text zu strukturieren verwenden wir die einfache Text-Auszeichnungssprache Markdown

1.1.1 Überschriften

Wir generieren Überschriften mit der Raute #. Eine einzelne Raute steht für eine Hauptüberschrift # Hauptüberschrift H1 und dann weiter bis zum ##### Untertitel H5 mit fünf Rauten.

1.1.2 Formatierungen

Mit Markdown können wir die wichtigsten Formatierungen ohne grossen Aufwand direkt im Text schreiben. ****fett**** wird **fett** und *_italic_* wird *italic*.

Listen werden mit

- Ein Punkt
- Ein nächster Punkt
- Und noch ein Punkt

zu

- Ein Punkt
- Ein nächster Punkt
- Und noch ein Punkt

Bei nummerierten Listen müssen wir uns nicht um die Nummern kümmern, unsere Liste

1. Erstens
1. Zweitens
1. Drittens

wird automatisch korrekt nummeriert:

¹Zur Not ist es sogar möglich, ein Worddokument zu erstellen.

1. Erstens
2. Zweitens
3. Drittens

1.2 Weitere Möglichkeiten

Markdown wird mit R zusammen RMarkdown und erlaubt uns, die Datenanalyse in die gleiche Quelldatei zu integrieren.

Wir können damit

- Tabellen erstellen
- Interaktive Abbildungen aus den aktuellen Daten erstellen
- Querverweise auf Abbildungen, Tabellen, Kapitel definieren
- Fussnoten und externe Links erstellen
- Mathematik-Satz integrieren
- Externe Grafiken einbinden
- Aus der gleichen Quelle PDF, HTML oder Word generieren
- und vieles mehr ...

2 Abbildungen

Wir können in bookdown entweder bestehende Grafiken einbinden (.jpg, .png) oder mit R-Code aus den aktuellen Daten die entsprechenden Grafiken direkt erstellen.

Abbildungen und Tabellen werden mit Legenden in den Bericht eingebunden; die Nummerierung wird intern automatisch erstellt und nachgeführt.

Die Abbildung 2.1 zeigt den Ausschnitt einer externen statischen Grafik.

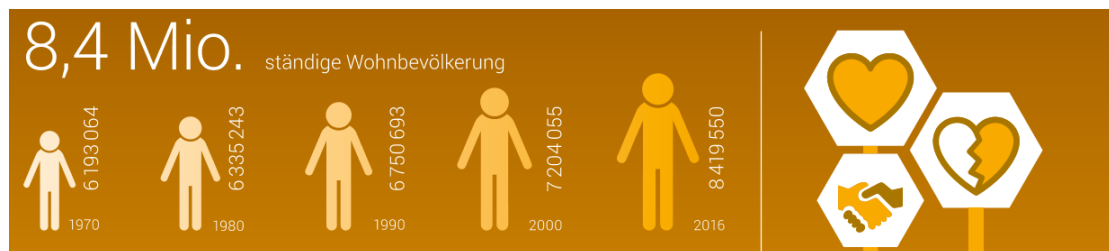


Abbildung 2.1 Die externe Grafik des BfS zeigt verschiedene Zahlen zur Bevölkerung in der Schweiz (Ausschnitt).

Aus den Rohdaten des BfS² können wir mit R eine eigene Grafik mit den Daten zur Entwicklung der Bevölkerung in der Schweiz ab 1991 und den drei Haupt-Szenarien bis 2045 erstellen.

Die Abbildung 2.2 ist in der HTML-Version des Berichtes interaktiv, wenn wir mit der Maus über die Linien fahren, werden die zugehörigen Werte der Linien angezeigt. In der PDF-Version des Berichtes ist die gleiche Abbildung ohne Interaktivität erstellt.

Wenn sich die Rohdaten ändern, wird diese Abbildung automatisch angepasst, da sie immer aus den aktuellen Daten erstellt wird. Wir können das an Änderungen an der Datei der Szenarien zeigen (data/su-d-01.02.04.04.xlsx).

²<https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung.html>

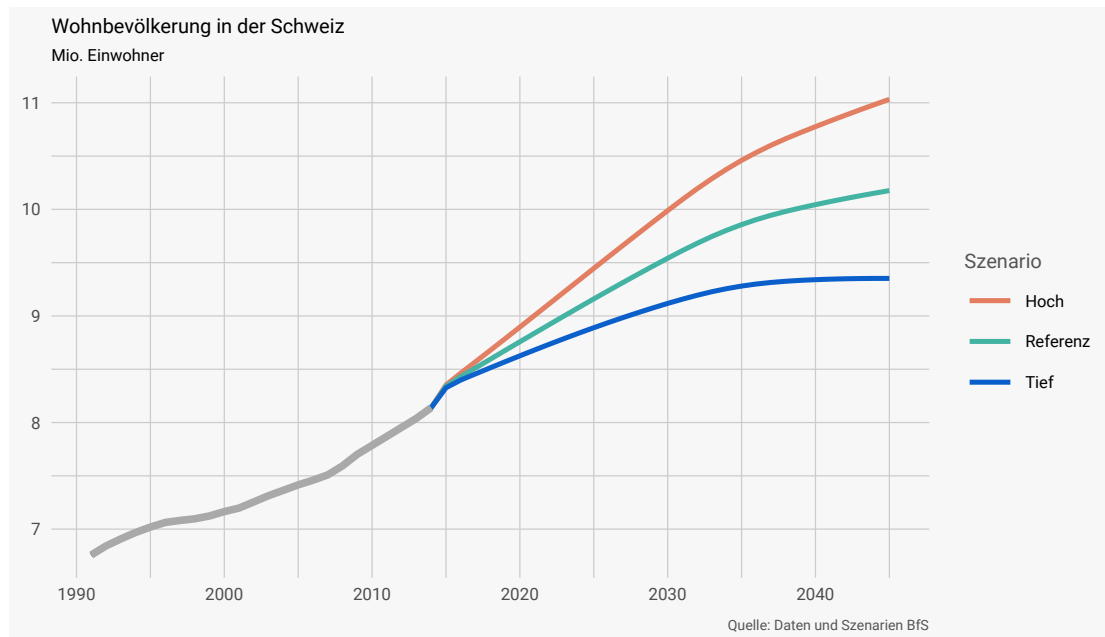


Abbildung 2.2 Bevölkerungsentwicklung in der Schweiz ab 1991 und die drei Szenarien 2015–2045.

3 Tabellen

Die Daten, welche für die Analyse eingelesen wurden, können in beliebiger Form als Tabelle gedruckt werden. Je nach Ausgabemedium (PDF, HTML) wird die Tabelle unterschiedlich formatiert.

Die Tabelle 3.1 zeigt einen Auszug der Szenarien des BfS zur Bevölkerungsentwicklung. Damit die Werte von Jahr zu Jahr einfacher zu vergleichen sind, benutzen wir hier einen Monofont (gleichabständig), so stehen alle Ziffern untereinander.

Tabelle 3.1 Die ersten Zeilen der Szenarien des BfS zur Bevölkerungsentwicklung der Schweiz.

Jahr	Tief	Referenz	Hoch
2015	8'326'041	8'339'505	8'352'840
2016	8'398'845	8'430'975	8'463'462
2017	8'456'550	8'512'664	8'570'246
2018	8'513'716	8'594'378	8'677'937
2019	8'570'241	8'676'062	8'786'438
2020	8'625'977	8'757'648	8'895'631

Im Anhang sind in der Tabelle A.1 die Daten von 1991 bis 2016 der Wohnbevölkerung aufgeführt.

4 Verweise innerhalb des Textes

Als Verweise unterscheiden wir

- Referenzen auf Abbildungen und Tabellen
- Referenzen auf Titel, z.B. Abschnitte, Kapitel
- Zitierung von Literatur

Es ist mit bookdown zudem möglich, einzelne Werte aus den Daten automatisch in den Text zu integrieren. Damit gibt es nie mehr copy & paste - Fehler, weil eine Stelle beim Aktualisieren der Daten vergessen ging.

Als Anker für die Referenz können wir in bookdown ein eigenes Label wählen (z.B. `tabelle_anhang`) und uns anschliessend auf dieses Label beziehen; die zugehörige Nummerierung wird automatisch erstellt.

4.1 Referenzen

4.1.1 Referenzen auf Abbildungen und Tabellen

Wir referenzieren eine Abbildung mit dem Präfix `fig:` und dem Label des entsprechenden R-Codes, beispielsweise wird `\@ref(fig:plot-chart)` zur Referenz [2.2](#).

Tabellen werden analog mit dem Präfix `tab:` erstellt, `\@ref(tab:tabelle)` wird zu [3.1](#).

4.1.2 Referenzen auf Titel

Wir können einem beliebigen Titel im Text ein Label zuweisen, z.B. `{#referenzen}` als Label für den Titel dieses Kapitels. Diesen Titel können wir anschliessend mit `\@ref(referenzen)` referenzieren, das sieht so aus: [4.1](#).

4.1.3 Literatur zitieren

Wir können externe Literatur zitieren, zum Beispiel benutzen wir hier das R-Paket bookdown (Xie 2015). Donald Knuth schrieb schon früh zum Thema 'Literate Programming' (Knuth 1984), um auf lesbaren Code aufmerksam zu machen.

4.2 Daten automatisch in Text

Es ist mit bookdown einfach, auf die im Hintergrund vorhandenen Daten zuzugreifen und Werte direkt in den Fliesstext zu integrieren.

So können wir z.B. den Wert des Referenz-Szenarios für das Jahr 2025 direkt in den Text schreiben lassen: 9'159'869 Personen.

Wenn sich die Daten ändern, wird dieser Wert im Text automatisch aktualisiert.

Hier wird zu diesem Zweck die folgende Funktion aufgerufen:

```
gib_formatierten_wert(daten, "Referenz", 2025)
```

Diese Funktion wurde für diese Analyse geschrieben und liefert aus den Daten einen formatierten Wert eines Szenarios für ein bestimmtes Jahr zurück.

Damit ist es auch möglich, ad-hoc-Berechnungen vorzunehmen. Wenn wir beispielsweise die Differenz zwischen 'Szenario hoch' und 'Szenario tief' im Jahr 2045 wissen möchten, schreiben wir auch das direkt in der Quelldatei.

```
tief <- .gib_wert_von(daten, "Tief", 2045)
hoch <- .gib_wert_von(daten, "Hoch", 2045)
differenz <- hoch - tief
print(differenz)
```

```
## [1] 1680399
```

Die Differenz beträgt also 1'680'399 Personen.

5 Fazit

Wir haben einen hübschen kurzen Bericht erstellt und können ihn als HTML und PDF anschauen.

Dieses Format eignet sich ebenso gut für die Dokumentation grosser Projekte und – wie der Name sagt – natürlich auch für Bücher.

Literatur

- Knuth, Donald E. 1984. "Literate Programming." *The Computer Journal* 27 (2): 97–111.
- Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <http://yihui.name/knitr/>.

A Rohdaten

Tabelle A.1 Die Daten der Bevölkerung der Schweiz 1991–2014. Quelle: BfS.

Jahr	Einwohner
1991	6'757'188
1992	6'842'768
1993	6'907'959
1994	6'968'570
1995	7'019'019
1996	7'062'354
1997	7'081'346
1998	7'096'465
1999	7'123'537
2000	7'164'444
2001	7'197'638
2002	7'255'653
2003	7'313'853
2004	7'364'148
2005	7'415'102
2006	7'459'128
2007	7'508'739
2008	7'593'494
2009	7'701'856
2010	7'785'806
2011	7'870'134
2012	7'954'662
2013	8'039'060
2014	8'139'631

Tabelle A.2 Die Daten der drei Szenarien zur Entwicklung der Bevölkerung 2015–2045. Quelle: BfS

Jahr	Tief	Referenz	Hoch
2015	8'326'041	8'339'505	8'352'840
2016	8'398'845	8'430'975	8'463'462
2017	8'456'550	8'512'664	8'570'246
2018	8'513'716	8'594'378	8'677'937
2019	8'570'241	8'676'062	8'786'438
2020	8'625'977	8'757'648	8'895'631
2021	8'680'953	8'838'977	9'005'343
2022	8'734'863	8'919'985	9'115'416
2023	8'787'752	9'000'580	9'225'672
2024	8'839'334	9'080'605	9'335'947
2025	8'889'569	9'159'869	9'445'971
2026	8'938'415	9'238'284	9'555'651
2027	8'985'655	9'315'740	9'664'842
2028	9'031'350	9'392'119	9'773'387
2029	9'075'374	9'467'377	9'881'154
2030	9'117'736	9'541'471	9'988'076
2031	9'157'408	9'613'379	10'093'101
2032	9'194'360	9'682'011	10'194'161
2033	9'227'705	9'746'393	10'289'137
2034	9'256'334	9'804'417	10'377'970
2035	9'280'322	9'856'971	10'459'605
2036	9'299'644	9'904'140	10'533'939
2037	9'314'383	9'944'908	10'602'007
2038	9'325'545	9'981'164	10'663'671
2039	9'334'128	10'013'978	10'721'003
2040	9'340'319	10'044'330	10'776'045
2041	9'345'189	10'073'195	10'829'714
2042	9'348'734	10'100'717	10'882'095
2043	9'351'012	10'127'066	10'933'283
2044	9'352'092	10'152'155	10'983'389
2045	9'352'082	10'176'124	11'032'481