

# Der Aufbau von Datensätzen

Cornelius Puschmann

## Table of contents

1	Strukturierte vs. unstrukturierte Daten	2
2	Data Frames und strukturierte Daten	2
3	Beobachtungen und Variablen	2
4	Zusammenfassung	3



Figure 1: Wie soll ich mich hier nur zurecht finden?, Bild generiert von Midjourney

Sozialwissenschaftlichen Daten sind Informationen, die zur Beschreibung und Erklärung der sozialen Welt verwendet werden. Sie können aus einer Vielzahl von Quellen stammen, zum Beispiel aus Umfragen, Experimenten, Interviews und Inhaltsanalysen. Kommunikationswissenschaftliche Daten können quantitativ oder qualitativ sein. Quantitative Daten sind in Zahlen ausgedrückt und können verwendet werden, um Muster und (kausale) Zusammenhänge zu identifizieren. Qualitative Daten sind in Text oder (Bewegt)Bildern ausgedrückt und können verwendet werden, um die Bedeutung von Ereignissen und Erfahrungen genauer zu verstehen.

Video

<https://nc.uni-bremen.de/index.php/s/875rxixCks7TKP5/download/%232%20Logik%20von%20Dataframes.mp4>

## 1 Strukturierte vs. unstrukturierte Daten

Unter *strukturierten Datensätzen* verstehen wir digital zusammengestellte Daten, die bestimmte strukturierende Merkmale aufweise. Häufig lassen sich solche strukturierten Daten als Tabellen darstellen und speichern. Ein konkretes Beispiel wäre eine Sammlung von Zeitungsartikeln. Handelt es sich um einzelne Textdateien, zu deren Inhalt man keine weiteren Informationen besitzt, würde man in der Regel von unstrukturierten Daten sprechen. Fügt man aber Titel, Publikationsdatum und Quelle hinzu, könnten wir diese Daten als Tabelle abspeichern und von einem strukturierten Datensatz sprechen.

## 2 Data Frames und strukturierte Daten

Nahezu alle strukturierte Daten lassen sich in R als Data Frames speichern. Dies bietet zahlreiche Vorteile, weil sich Data Frames leicht umformen, refaktorisieren oder aggregieren lassen. So kann man leicht auszählen, wie viele Artikel zur gleichen Quelle gehören, zu welchem Zeitpunkt wie viele Artikel veröffentlicht wurden, oder (bei einer Befragung) wie viele Stunden die StudienteilnehmerInnen im Mittel wöchentlich fernsehen. Häufig lassen sich zunächst unstrukturierte Daten in ein strukturiertes Format überführen, wenn man relevante Metadaten nutzt, beispielsweise Dateinamen.

## 3 Beobachtungen und Variablen

Der Aufbau von Tabellen (bzw. Data Frames) folgt dabei in der Regel einer logischen Struktur. Die Zeilen geben die Anzahl der *Beobachtungen* (oder *Fälle*) an, während die Spalten einzelne *Variablen* (oder *Merkmale*) repräsentieren.

Einheit	Beobachtung (oder Fall)	Variable (oder Merkmal)
<b>Tabellen- Strukturmerkmal</b>	Zeile	Spalte
<b>Beispiele</b>	<ul style="list-style-type: none"> <li>• Respondent in einer Befragung</li> <li>• Zeitungsartikel in einer Inhaltsanalyse</li> <li>• Versuchsdurchlauf in einem Experiment</li> </ul>	<ul style="list-style-type: none"> <li>• Antwort auf eine Frage in einem Fragebogen</li> <li>• Quelle oder Titel eines Zeitungsartikels in einer Inhaltsanalyse</li> <li>• Messwert in einem Experiment</li> </ul>

## 4 Zusammenfassung

Im nächsten Kapitel werden wir uns eingehender mit der Behandlung von tabellenstrukturierten Daten in unterschiedlichen Formaten beschäftigen.