

Einführung in R

Stephan Goerigk

2022-10-13

Contents

Über dieses Skript	7
1 Warum ist R so gut?	9
1.1 Open Source	9
1.2 Vielseitigkeit	9
1.3 R Markdown	9
1.4 Transparenz	10
2 R Materialien	11
2.1 Cheat Sheets	11
2.2 Hilfe und Inspiration online	11
2.3 Andere Bücher	11
3 Installation	13
3.1 Installation von R	13
3.2 Installation von RStudio	14
4 Programmaufbau	15
4.1 Die vier RStudio Fenster	15
4.2 R Packages	17
5 Datenformate	21
5.1 Skalar	21
5.2 Vektor	23
5.3 Matrizen und Dataframes	25
6 Daten erstellen	29
6.1 Manuell	29
6.2 Automatisch	30
6.3 Zufällig	30
7 Daten importieren und speichern	31
7.1 Working Directory	31
7.2 Working Environment	31

7.3	Daten importieren	31
7.4	Daten speichern	31
8	Daten Explorieren	33
8.1	Informationen über R Objekt	33
8.2	Deskriptivstatistiken	33
9	Datenmanipulation	35
9.1	Löschen	35
9.2	Umbenennen	35
9.3	Verändern	35
9.4	Aggreggieren	35
9.5	Transformieren	35
10	Graphiken	37
10.1	ggplot()	38
10.2	geom_histogram()	38
10.3	geom_point()	38
10.4	geom_line() und geom_smooth()	38
10.5	geom_bar()	38
10.6	geom_boxplot()	38
10.7	Facetting	38
10.8	Ästhetische Anpassungen	38
10.9	theme()	38
10.10	Abbildungen kombinieren	38
10.11	Abbildungen exportieren	38
11	Tabellen	39
12	Preprocessing	41
12.1	Fehlende Werte	41
12.2	Ausreißer	41
12.3	Schiefe Daten	41
13	Korrelation	43
13.1	Pearson Korrelation	43
13.2	Spearman Korrelation	43
13.3	Kendall Korrelation	43
13.4	Korrelationmatrix	43
14	Regression	45
14.1	Lineare Regression	45
14.2	Logistische Regression	45
15	Mehrstufige UVs	47
15.1	Omnibus Test	47
15.2	Paarweise Vergleiche	47

<i>CONTENTS</i>	5
16 Mehrere UVs	49
16.1 Multiple Regression	49
16.2 Interaktionsterme	49
17 Manuskripterstellung	51

Über dieses Skript

Liebe Studierende,

dieses Skript soll Sie in die grundlegenden Analysewerkzeuge in R einführen, von der grundlegenden Kodierung und Analyse bis hin zur Datenverarbeitung, dem Plotten und der statistischen Inferenz.

Wenn R Ihre erste Programmiersprache ist, ist das völlig in Ordnung. Wir gehen alles Schritt für Schritt gemeinsam durch. Die Techniken in diesem Buch sind zwar auf die meisten Datenanalyseprobleme anwendbar, da wir jedoch aus der Psychologie kommen, werde ich den Kurs auf die Lösung von Analyseproblemen ausrichten, die in der psychologischen Forschung häufig auftreten.

Chapter 1

Warum ist R so gut?

1.1 Open Source

R ist zu 100 % kostenlos und verfügt daher über eine große Unterstützergemeinschaft. Im Gegensatz zu SPSS, Matlab, Excel und JMP ist R völlig kostenlos und wird es auch immer bleiben. Das hilft nicht nur Ihrem Geldbeutel - es bedeutet auch, dass eine riesige Gemeinschaft von R-Programmierern ständig neue R-Funktionen und -Pakete in einer Geschwindigkeit entwickelt und verbreitet, die alle anderen Pakete in den Schatten stellt. Die Größe der R-Programmiersgemeinschaft ist atemberaubend. Wenn Sie jemals eine Frage dazu haben, wie man etwas in R implementiert, wird eine schnelle Google-Suche Sie praktisch jedes Mal zur Antwort führen.

1.2 Vielseitigkeit

R ist unglaublich vielseitig. Sie können R für alles verwenden, von der Berechnung einfacher zusammenfassender Statistiken über die Durchführung komplexer Simulationen bis hin zur Erstellung großartiger Diagramme. Wenn Sie sich eine analytische Aufgabe vorstellen können, können Sie sie mit ziemlicher Sicherheit in R implementieren.

1.3 R Markdown

Mit RStudio, einem Programm, das Sie beim Schreiben von R-Code unterstützt, können Sie mit RMarkdown einfach und nahtlos R-Code, Analysen, Diagramme und geschriebenen Text zu eleganten Dokumenten an einem Ort kombinieren.

Tatsächlich habe ich dieses gesamte Skript (Text, Formatierung, Diagramme, Code... ja, alles) in RStudio mit R Markdown geschrieben. Mit RStudio müssen Sie sich nicht mehr mit zwei oder drei Programmen herumschlagen, z. B. Excel, Word und SPSS, wo Sie die Hälfte Ihrer Zeit mit dem Kopieren, Einfügen und Formatieren von Daten, Bildern und Tests verbringen, sondern können alles an einem Ort erledigen, so dass nichts mehr falsch gelesen, getippt oder vergessen wird.

1.4 Transparenz

In R durchgeführte Analysen sind transparent, leicht weiterzugeben und reproduzierbar. Wenn Sie einen SPSS-Benutzer fragen, wie er eine bestimmte Analyse durchgeführt hat, wird er sich ggf. nicht daran erinnern, was er vor Monaten oder Jahren tatsächlich getan hat. Wenn Sie einen R-Anwender (der gute Programmiertechniken verwendet) fragen, wie er eine Analyse durchgeführt hat, sollte er Ihnen immer den genauen Code zeigen können, den er verwendet hat. Das bedeutet natürlich nicht, dass er die richtige Analyse verwendet oder sie korrekt interpretiert hat, aber mit dem gesamten Originalcode sollten etwaige Probleme völlig transparent sein! Dies ist eine Grundvoraussetzung für offene, replizierbare Forschung.

Chapter 2

R Materialien

2.1 Cheat Sheets

Im Laufe dieses Buches werden Sie viele neue Funktionen kennenlernen. Wäre es nicht schön, wenn jemand ein Wörterbuch mit vielen gängigen R-Funktionen erstellen würde? Ja, das wäre es, und zum Glück haben einige freundliche R-Programmierer genau das getan. Im Folgenden finden Sie eine Tabelle mit einigen der Funktionen, die ich empfehle. Ich empfehle Ihnen dringend, diese auszudrucken und die Funktionen zu markieren, wenn Sie sie lernen!

Link zum Base R Cheat Sheet

Link zu den R Studio Cheat Sheets

Insbesondere die Cheat Sheets zu ggplot2 und dplyr kann ich Ihnen nur wärmstens ans Herz legen

2.2 Hilfe und Inspiration online

Eine Google Suche nach einem spezifischen R Problem bringt Sie (fast) immer an Ihr Ziel. Häufig findet man gute Antworten in den github Hilfsmaterialien einzelner Pakete, auf den Community Seiten von R Studio und in den Foren von stackoverflow.

2.3 Andere Bücher

Die Inhalte dieser Bücher sind nicht prüfunsrelevant.

Es gibt viele, viele ausgezeichnete Bücher über R. Hier sind zwei, die ich empfehlen kann (von denen eines sogar umsonst ist):

Discovering Statistics with R von Field, Miles and Field

R for Data Science von Garrett Golemund and Hadley Wickham

Chapter 3

Installation

Um R benutzen zu können müssen wir zwei Softwarepakete herunterladen:

- **R**
- **RStudio**

R ist die Programmiersprache, mit der wir arbeiten. R-Studio ist eine Benutzeroberfläche, die uns das Programmieren mit R ungemein erleichtert.

3.1 Installation von R

Um R zu installieren, klicken Sie auf den, Ihrem Betriebssystem entsprechenden, Link und befolgen Sie die Anleitungen:

Operating System	Link
Windows	http://cran.r-project.org/bin/windows/base/
Mac	http://cran.r-project.org/bin/macosx/

Nach dieser Installation haben Sie bereits die volle Funktionalität des Programms. Sie werden jedoch feststellen, dass beinahe alle R-Nutzer RStudio zum programmieren nutzen, da dieses eine leichter nutzbare Oberfläche hat. Tatsächlich müssen Sie nach der Installation von RStudio das R Basisprogramm nie wieder öffnen.

3.2 Installation von RStudio

Bitte installieren Sie dann RStudio - das Programm, über welches wir auf R zugreifen und mit dem wir unsere Skripte schreiben.

Um RStudio zu installieren, klicken Sie auf diesen Link und befolgen Sie die Anleitungen: <http://www.rstudio.com/products/rstudio/download/>

Chapter 4

Programmaufbau

4.1 Die vier RStudio Fenster

Wenn Sie RStudio öffnen, sehen Sie vier Fenster, wie in der folgenden Abbildung dargestellt:

Wenn Sie mögen, können Sie die Reihenfolge der Fenster in den RStudio Einstellungen verändern. Sie können die Fenster auch verstecken (Minimieren/Maximieren Symbol an der oberen rechten Ecke jedes Fensters) oder ihre Größe verändern, indem die sie ihre Grenzbalken anklicken und verschieben.

Lassen Sie uns jetzt schauen, was genau die Funktion jedes der Fenster ist:

4.1.1 Source - Ihr Schreibblock für Code

Im Source Fenster erstellen und bearbeiten Sie “R-Skripte” - Ihre Codesammlungen. Keine Sorge, R-Skripte sind nur Textdateien mit der Erweiterung “.R”. Wenn Sie RStudio öffnen, wird automatisch ein neues unbenanntes Skript gestartet. Bevor Sie mit der Eingabe eines unbenannten R-Skripts beginnen, sollten Sie die Datei immer unter einem neuen Dateinamen speichern (z.B. “Mein_RSript.R”). Wenn Ihr Computer während der Arbeit abstürzt, steht Ihr Code in R zur Verfügung, wenn Sie RStudio erneut öffnen.

Sie werden feststellen, dass R beim Schreiben des Skripts den Code während der Eingabe nicht tatsächlich ausführt. Damit R Ihren Code tatsächlich ausführt, müssen Sie den Code zunächst an die Konsole “senden” (wir werden im nächsten Abschnitt darüber sprechen).

Es gibt viele Möglichkeiten, Ihren Code aus dem Skript an die Konsole zu senden. Die langsamste Methode ist das Kopieren und Einfügen. Schneller geht es, wenn

Sie den Code, den Sie auswerten möchten, markieren und auf die Schaltfläche “Run” oben rechts in der Quelle klicken. Alternativ können Sie auch die Tastenkombination “Command + Return” auf dem Mac oder “Control + Enter” auf dem PC verwenden, um den gesamten markierten Code an die Konsole zu senden.

4.1.2 Konsole - Das Herzstück von R

Die Konsole ist das Herzstück von R. Hier führt R den Code aus. Am Anfang der Konsole sehen Sie das Zeichen “>”. Dies ist eine Eingabeaufforderung (sog. “Prompt”), die Ihnen mitteilt, dass R bereit für neuen Code ist. Sie können direkt nach dem Prompt > Code in die Konsole eingeben und erhalten sofort eine Antwort. Wenn Sie zum Beispiel 2+2 in die Konsole eingeben und die Eingabetaste drücken, werden Sie sehen, dass R sofort eine Ausgabe von 4 liefert.

```
2+2
```

```
## [1] 4
```

Versuchen Sie, 2+2 zu berechnen, indem Sie den Code direkt in die Konsole eingeben - und dann Enter drücken. Sie sollten das Ergebnis [1] 4 sehen. Machen Sie sich keine Gedanken über die [1], dazu kommen wir später.

Geben Sie denselben Code in das Skript ein und senden Sie ihn an die Konsole, indem Sie den Code markieren und auf die Schaltfläche “Run” in der oberen rechten Ecke des Quelltextfensters klicken. Alternativ können Sie auch die Tastenkombination “Command + Return” auf dem Mac oder “Control + Enter” unter Windows verwenden.

Tipp: Wie Sie sehen, können Sie Code entweder über das Skript oder durch direkte Eingabe in die Konsole ausführen. In 99% der Fälle sollten Sie jedoch das Skript und nicht die Konsole verwenden. Der Grund dafür ist ganz einfach: Wenn Sie den Code in die Konsole eingeben, wird er nicht gespeichert (obwohl Sie in Ihrem Befehlsverlauf nachsehen können). Und wenn Sie beim Eingeben von Code in die Konsole einen Fehler machen, müssen Sie alles noch einmal von vorne eingeben. Stattdessen ist es besser, den gesamten Code in das Skript zu schreiben. Wenn Sie bereit sind, einen Code auszuführen, können Sie ihn mit “Run” an die Konsole senden.

4.1.3 Environment/History - Das Gedächtnis von R

In dem Tab “Environment” dieses Bereichs werden die Namen aller Datenobjekte (wie Vektoren, Matrizen und Datenrahmen) angezeigt, die Sie in Ihrer aktuellen R-Session definiert haben. Sie können auch Informationen wie die Anzahl der Spalten und Zeilen in Datensätzen sehen. Der Tab enthält auch einige anklickbare Aktionen wie

Datensatz importieren”, wodurch eine grafische Benutzeroberfläche (GUI) für wichtige Daten in R geöffnet wird.

Der Tab “History” dieses Bereichs zeigt Ihnen einfach eine Sammlung aller Befehle an den Sie zuvor in der Konsole ausgewertet haben. Wenn man mit Skripten arbeitet, schaut man sich diese, allerdings relativ selten an.

Wenn Sie sich mit R besser auskennen, werden Sie das Fenster Environment/History vielleicht nützlich finden. Aber für den Moment können Sie es einfach ignorieren. Wenn Sie Ihren Bildschirm entrümpeln wollen, können Sie das Fenster auch einfach minimieren, indem Sie auf die Schaltfläche Minimieren oben rechts im Fenster klicken.

4.1.4 Files/Plots/Packages/Help/Viewer - Interaktion von R mit Dateien

Die Tabs Files/Plots/Packages/Help/Viewer zeigen Ihnen viele hilfreiche Informationen. Schauen wir uns die einzelnen Registerkarten im Detail an:

1. Files - Der Tab “Files” gibt Ihnen Zugriff auf das Dateiverzeichnis auf Ihrer Festplatte. Dateien, die Sie in Ihrem R Projekt benutzen liegen in der Regel in einem von Ihnen definierten Arbeitsverzeichnis. Wir werden in Kürze ausführlicher über Arbeitsverzeichnisse sprechen.
2. Plots - Das Plots-Panel zeigt (keine große Überraschung) alle Ihre Plots an.
3. Pakete - Zeigt eine Liste aller auf Ihrer Festplatte installierten R-Pakete an und gibt an, ob sie derzeit geladen sind oder nicht. Pakete, die in der aktuellen Sitzung geladen sind, sind markiert, während Pakete, die installiert, aber noch nicht geladen sind, nicht markiert sind. Auf die Pakete gehen wir im nächsten Abschnitt näher ein.
4. Hilfe - Hilfemenü für R-Funktionen. Sie können entweder den Namen einer Funktion in das Suchfenster eingeben oder den Code `?function.name` verwenden, um nach einer Funktion mit dem Namen `function.name` zu suchen:

```
?hist # Wie funktioniert die Histogrammfunktion?  
?t.test # Wie funktioniert der t-Test?
```

4.2 R Packages

Wenn Sie R zum ersten Mal herunterladen und installieren, installieren Sie die Base R Software. Base R enthält die meisten Funktionen, die Sie täglich

verwendet werden, wie `mean()` und `hist()`. Allerdings werden hier nur Funktionen angezeigt, die von den ursprünglichen Autoren der Sprache R geschrieben wurden. Wenn Sie auf Daten und Code zugreifen möchten, die von anderen Personen geschrieben wurden, müssen Sie diese als “Package” installieren. Ein R-Package ist einfach ein Bündel von Funktionen (also bereits geschriebener Code), die in einem übersichtlichen Paket gespeichert sind.

Ein Paket ist wie eine Glühbirne. Um es nutzen zu können, müssen Sie es zunächst in Ihr Haus (d.h. auf Ihren Computer) bestellen, indem Sie es installieren. Wenn Sie ein Paket einmal installiert haben, brauchen Sie es nie wieder zu installieren. Jedes Mal, wenn Sie das Paket tatsächlich verwenden wollen, müssen Sie es jedoch einschalten, indem Sie es laden. Und so geht’s:

4.2.1 R Packages installieren

Ein Paket zu installieren bedeutet einfach, den Paketcode auf Ihren Computer herunterzuladen. Die gängigste Methode ist das Herunterladen aus dem Comprehensive R Archive Network (CRAN).

Um ein neues R-Paket von CRAN zu installieren, können Sie einfach den Code `install.packages("name")` ausführen, wobei “name” der Name des Pakets ist.

Um zum Beispiel das Paket `ggplot2` herunterzuladen, welches wir oft zum Erstellen von Graphen verwenden, geben Sie ein:

```
# install.packages("ggplot2")
```

4.2.2 R Packages laden

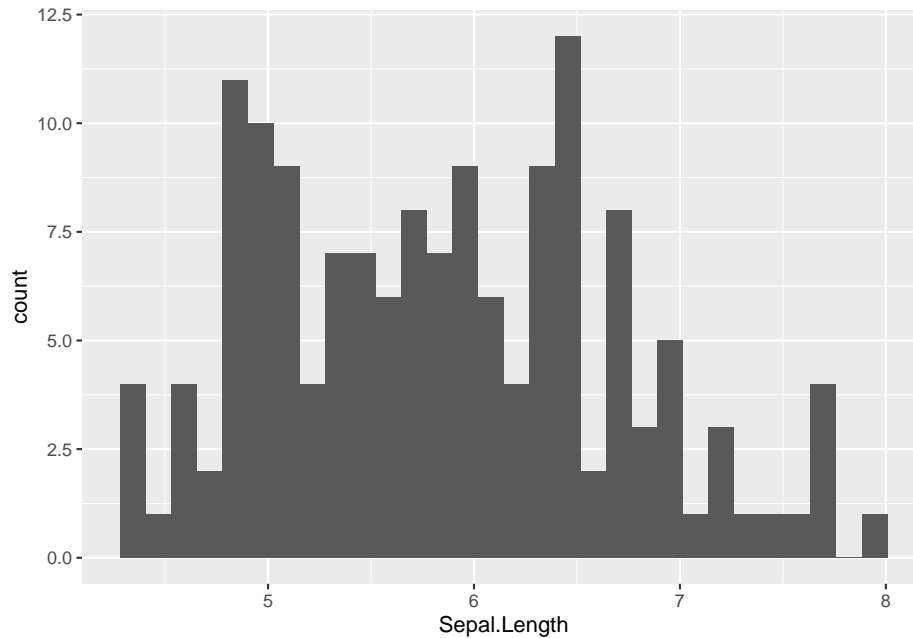
Sobald Sie ein Paket installiert haben, befindet es sich auf Ihrem Computer. Aber nur weil es auf Ihrem Computer ist, bedeutet das nicht, dass R bereit ist, es zu benutzen. Wenn Sie etwas, wie eine Funktion oder einen Datensatz, aus einem Paket verwenden wollen, müssen Sie *immer* zuerst das Paket in Ihrer R-Sitzung *laden*. Genau wie bei einer Glühbirne müssen Sie sie einschalten, um sie zu benutzen!

Um ein Paket zu laden, verwenden Sie die Funktion `library()`. Nachdem wir zum Beispiel das Paket `ggplot2` installiert haben, können wir es mit `library("ggplot2")` laden:

```
# Laden des "ggplot2" Pakets, damit wir es benutzen können!  
# Pakete müssen zu Beginn jeder R Session neu geladen werden!  
library("ggplot2")
```

Jetzt, wo Sie das Paket `ggplot2` geladen haben, können Sie jede seiner Funktionen benutzen (hier die Funktion `ggplot`, um einen Graph zu erstellen)!

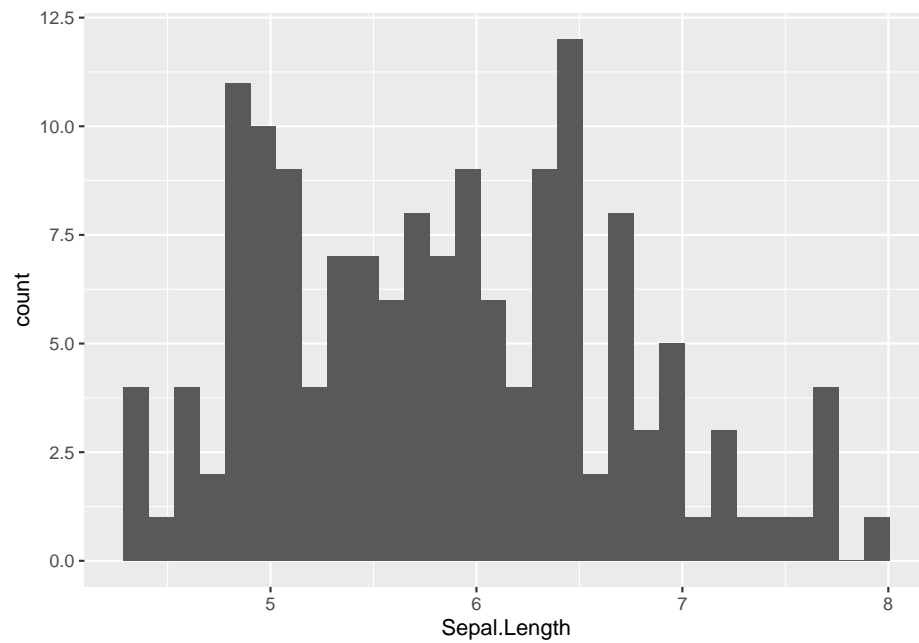
```
ggplot(data = iris, aes(x = Sepal.Length)) +  
  geom_histogram()
```



Pakete müssen zu Beginn jeder R Session neu geladen werden. Deswegen schreiben wir in der Regel ganz an den Anfang unseres Skripts gleich mehrere Zeilen, mit `library()` Befehlen für alle R Pakete, die wir für unsere Analyse benötigen werden.

In R gibt es eine Möglichkeit, ein Paket vorübergehend zu laden, ohne die Funktion `library()` zu verwenden. Um dies zu tun, können Sie einfach die Notation `package::funktion` verwenden. Diese Notation sagt R einfach, dass es das Paket nur für diesen einen Codeabschnitt laden soll. Zum Beispiel könnte ich die Funktion `ggplot` aus dem Paket `ggplot2` wie folgt verwenden:

```
ggplot2::ggplot(data = iris, aes(x = Sepal.Length)) +  
  geom_histogram()
```



Ein Vorteil der Notation “package::function” ist, dass für jeden, der den Code liest, sofort klar ist, welches Paket die Funktion enthält. Ein Nachteil ist jedoch, dass Sie, wenn Sie eine Funktion aus einem Paket häufig verwenden, gezwungen sind, den Paketnamen ständig neu einzugeben. Sie können jede Methode verwenden, die für Sie sinnvoll ist.

Chapter 5

Datenformate

5.1 Skalar

Der einfachste Objekttyp in R ist der **Skalar**. Ein Skalar Objekt ist einfach nur ein einzelner Wert, z.B. eine Zahl oder ein Wort.

Hier sind einige Beispiele für numerische Skalar Objekte:

```
# Examples of numeric scalars
a <- 100
b <- 3 / 100
c <- (a + b) / b
```

Skalare müssen nicht numerisch sein, sondern auch Worte. Wortobjekte heißen in R **characters** (aka strings). In R schreibt man characters immer in Anführungszeichen `"`. Hier sind einige Beispiele für character Skalare:

```
# Beispiele für character Skalare
d <- "Psychologe"
e <- "Zigarre"
f <- "Haben Psychologen wirklich alle Bärte und rauchen Zigarre?"
```

Wie Sie sich vermutlich vorstellen können, behandelt R numerische und character Skalare unterschiedlich. Zum Beispiel lassen sich mit numerischen Skalaren grundlegende arithmetische Operationen durchführen (Addition, Subtraktion, Multiplikation...) – das funktioniert mit character Skalaren nicht. Wenn Sie dennoch probieren numerische Operationen auf character Skalare anzuwenden, bekommen Sie eine Fehlermeldung, so wie diese:

```
a = "1"
b = "2"
a + b
```

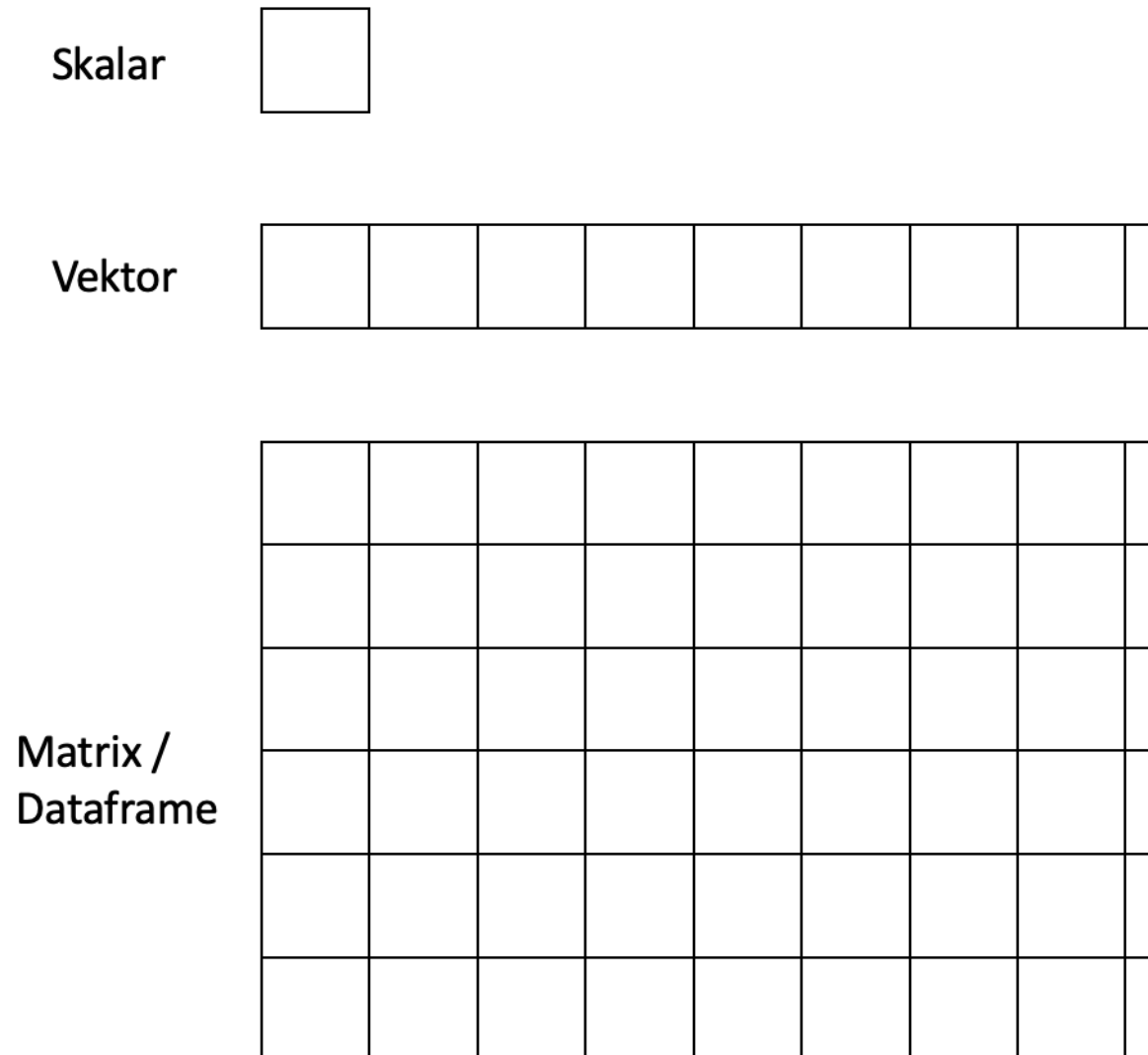


Figure 5.1: Skalar, Vektor, Matrix

“Fehler in $a + b$: nicht-numerisches Argument für binären Operator”

5.2 Vektor

Machen wir weiter mit **Vektoren**. Ein Vektor Objekt ist einfach eine Kombination mehrerer Skalare in einem einzelnen Objekt (z.B. eine Zahlen- oder Wortreihe). Zum Beispiel könnten die Zahlen von 1-10 in einen Vektor mit der Länge 10 kombiniert werden. Oder die Buchstaben des Alphabets könnten in einen Vektor mit der Länge 26 gespeichert werden. Genau wie Skalare, können Vektoren numerisch oder characters sein (Aber nicht beides auf einmal!)

Die einfachste Art einen Vektor zu erstellen ist mit der `c()` Funktion. Das `c` steht für “concatenate”, was auf Englisch so viel heißt wie “zusammenbringen”. Die `c()` Funktion nimmt mehrere Skalare als Input und erstellt einen Vektor, der diese Objekte enthält.

Wenn `man()` benutzt, muss man immer ein **Komma** zwischen die Objekte setzen (Skalare oder Vektoren), die man kombinieren möchte.

Lassen Sie uns die `c()` Funktion nutzen um einen Vektor zu erstellen der `a` heißt und die Zahlen von 1 bis 7 enthält

Let’s use the `c()` function to create a vector called `a` containing the integers from 1 to 5.

```
a = c(1, 2, 3, 4, 5, 6, 7)
# Das Ergebnis ausgeben
a
```

```
## [1] 1 2 3 4 5 6 7
```

Sie können auch character Vektoren erstellen, indem Sie die `c()` auf einzelne character Skalare Funktion anwenden:

```
char.vec = c("Freud", "Wundt", "Bandura", "Watson", "Jung")
# Das Ergebnis ausgeben
char.vec
```

```
## [1] "Freud" "Wundt" "Bandura" "Watson" "Jung"
```

5.2.1 Vektor Typen

Vektoren sind ein zentrales Element von R. Ein Vektor kann Zahlen, Buchstaben oder logische Werte enthalten, aber niemals eine Kombination

Der Vektor ist die Entsprechung der **Variable** und die Skalare, aus denen der Vektor besteht sind die **Merkmalsausprägungen** der Variable.

5.2.2 Faktor Variablen

Wir haben bereits gelernt, wie man einen Vektor aus character Objekten erstellt. Manchmal brauchen wir in R jedoch Variablen, die nicht nur Worte enthalten, sondern dem Programm mitteilen, dass es sich um feste Gruppen oder **Kategorien** handelt. Es geht also nicht nur um eine "Sammlung" von Worten (z.B. Nachnamen von Probanden), sondern um festgelegte Analyseeinheiten. Solche Variablen heißen in R **factor**.

In einer factor Variable ist jeder Kategorie eine Zahl zugeordnet (z.B. 1 = männlich, 2 = weiblich).

Um Faktor Variablen zu erstellen, machen wir einen Vorgang, den man **Kodieren** nennt und das geht so:

Wir haben einen Vektor mit Codes 1 und 2 für männlich und weiblich vorliegen:

```
geschlecht = c(1, 2, 2, 1, 2)
# Das Ergebnis ausgeben
geschlecht
```

```
## [1] 1 2 2 1 2
```

In dieser Form erkennt R diesen Vektor als numerische Variable. Um Sie in einen Faktor umzuwandeln, definieren wir die Zahlen (1 und 2) als **levels** des Faktors und geben dann jedem level einen Namen (**labels**):

```
geschlecht = factor(geschlecht, levels = c(1,2), labels = c("männlich", "weiblich"))
# Das Ergebnis ausgeben
geschlecht
```

```
## [1] männlich weiblich weiblich männlich weiblich
## Levels: männlich weiblich
```

Das Ergebnis ist eine codierte Faktorvariable. Wenn wir Sie uns ausgeben lassen erhalten wir unter den Merkmalsausprägungen eine Liste mit den einzelnen Kategorien (levels) des Faktors.

R wird uns für Faktoren alle Ergebnisse nach der **Reihenfolge** der levels anzeigen. Wenn wir keine Faktorvariable haben, sondern eine character Variable funktioniert die Reihenfolge immer alphabetisch.

5.2.3 Vektor Indizierung

Manchmal möchten wir wieder einen einzelnen Skalar auswählen, der als Teil von einem Vektor gespeichert ist. Diese **Auswahl** eines Einzelements nennt man **Indizierung**. Die Auswahl eines kleineren Objekts aus einem größeren Objekt funktioniert in R immer mit `[]`.

Benötigen wir aus einem Vektor z.B. genau den 3. Skalar, schreiben wir einfach eine 3 in eckige Klammern hinter den Vektor.

```
char.vec = c("Freud", "Wundt", "Bandura", "Watson", "Jung")
# Das Ergebnis ausgeben
char.vec[5]

## [1] "Jung"
```

5.3 Matrizen und Dataframes

In der Psychologie beobachten wir für unsere Studien fast immer mehr als eine Variable. Wir könnten diese alle in einzelnen Vektoren speichern und uns die Objektnamen merken. Z.B.

```
Name = c("Max", "Maja", "Mia", "Moritz", "Markus")
Alter = c(20, 31, 25, 34, 51)
Diagnose = c("Depression", "Zwangsstörung", "Depression", "Soziale Phobie", "Depression")
```

5.3.1 Erstellen von Datenmatrizen

Praktischer ist es, die einzelnen Vektoren in Tabellenform zu speichern, der **Datenmatrix**. In R heißen Datenmatrizen `data.frame`. Wir können die Vektoren folgendermaßen zu einem `data.frame` kombinieren:

```
df = data.frame(Name, Alter, Diagnose)
# Das Ergebnis ausgeben
df
```

```
##      Name Alter      Diagnose
## 1    Max    20    Depression
## 2   Maja    31 Zwangsstörung
## 3    Mia    25    Depression
## 4 Moritz    34 Soziale Phobie
## 5 Markus    51    Depression
```

Wie in jeder Datenmatrix entsprechen die **Zeilen** den einzelnen Personen (Fällen) und die **Spalten** den Variablen.

R bezeichnet Zeilen und Spalten als **rows** und **columns**. Wollen wir z.B. wissen, wie viele Zeilen der `data.frame` hat, können wir `nrow()` benutzen. Für die Anzahl der Spalten nehmen wir `ncol()`:

```
nrow(df)
```

```
## [1] 5
```

```
ncol(df)
```

```
## [1] 3
```

Wenn wir die einzelnen Vektoren nicht bereits vorher definiert haben, können wir auch alles in einem Schritt machen. Das Ergebnis ist das gleiche:

```
df = data.frame("Name" = c("Max", "Maja", "Mia", "Moritz", "Markus"),
               "Alter" = c(20, 31, 25, 34, 51),
               "Diagnose" = c("Depression", "Zwangsstörung", "Depression", "Soziale Phobie", "Depression")
               )
# Das Ergebnis ausgeben
df
```

```
##      Name Alter      Diagnose
## 1    Max    20    Depression
## 2   Maja    31 Zwangsstörung
## 3    Mia    25    Depression
## 4 Moritz    34 Soziale Phobie
## 5 Markus    51    Depression
```

Wollen wir wieder eine einzelne Variable aus dem benutzen, können wir diese über das \$ Zeichen anwählen:

```
df$Alter
```

```
## [1] 20 31 25 34 51
```

5.3.2 Indizierung

Wollen wir aus dem data.frame wieder einzelne Elemente benutzen, nutzen wir wieder die Indizierung. Auch hier brauchen wir die []. Da wir im data.frame Zeilen und Spalten haben, brauchen wir eine Möglichkeit beides auszuwählen, wie ein Cursor der von links nach rechts, bzw. von oben nach unten läuft.

Wir trennen dafür unsere [] mit einem Komma [,]. Alles was **links vom Komma** steht bezieht sich auf Zeilen alles **rechts vom Komma** bezieht sich auf Spalten.

Lassen Sie uns einmal die Zelle in der 1. Zeile (also die 1. Person) und der 3. Variable auswählen:

```
df[1,3]
```

```
## [1] Depression
## Levels: Depression Soziale Phobie Zwangsstörung
```

Lassen wir die Zahl vor dem Komma weg, bekommen wir alle Werte aus der Spalte:

```
df[,3]
```

```
## [1] Depression      Zwangsstörung Depression      Soziale Phobie Depression  
## Levels: Depression Soziale Phobie Zwangsstörung
```

Lassen wir die Zahl nach dem Komma weg, bekommen wir alle Werte aus der Reihe:

```
df[1,]
```

```
##   Name Alter  Diagnose  
## 1  Max    20 Depression
```


Chapter 6

Daten erstellen

6.1 Manuell

Die manuelle Eingabe von Daten erfolgt über die `c()` Funktion. Mit ihrer Hilfe können wir Skalare zu Vektoren verbinden...

```
a = c(1, 2, 4, 6, 1)
```

...mehrere Vektoren aneinanderhängen...

```
a = c(1, 2, 4, 6, 1)
b = c(2, 3)
c = c(a, b)
c
```

```
## [1] 1 2 4 6 1 2 3
```

...und Vektoren gleicher Länge zu `data.frames` kombinieren:

```
daten = data.frame(aufmerksamkeit = c(58, 46, 29, 51),
                   gedaechtnis = c(22, 67, 22, 31))
daten
```

```
##   aufmerksamkeit gedaechtnis
## 1             58           22
## 2             46           67
## 3             29           22
## 4             51           31
```

6.2 Automatisch

Wir haben bereits die `c()` Funktion gelernt.

Die `c()` Funktion ist die einfachste Art einen Vektor zu erstellen, sie ist aber vermutlich auch die umständlichste. Stellen Sie sich zum Beispiel vor, Sie wollen einen Vektor erstellen, der alle Zahlen von 0 bis 100 enthält. Diese Zahlen wollen Sie definitiv nicht alle in die Klammer von `c()` eintippen.

Glücklicherweise hat R viele eingebaute Funktionen, um leicht automatisch numerische Vektoren zu erstellen.

Lassen Sie uns mit dreien davon starten: `a:b`, `seq()`, and `rep()`:

Funktion	Beispiel	Ergebnis
<code>c(a, b, ...)</code>	<code>c(1, 5, 9)</code>	1, 5, 9
<code>a:b</code>	<code>1:5</code>	1, 2, 3, 4, 5
<code>seq(from, to, by, length.out)</code>	<code>seq(from = 0, to = 6, by = 2)</code>	0, 2, 4, 6
<code>rep(x, times, each, length.out)</code>	<code>rep(c(7, 8), times = 2, each = 2)</code>	7, 7, 8, 8, 7, 7, 8, 8

6.3 Zufällig

In R haben Sie die Möglichkeit Daten anhand einer Wahrscheinlichkeitsverteilung zu simulieren.

Wollen wir beispielsweise eine normalverteilte Variable mit zufälligen Werten erstellen, können wir die `rnorm()` Funktion nutzen. Dafür müssen wir lediglich angeben, wie viele Werte wir haben wollen (`n`) und welchen Mittelwert (`mean`) und welche Standardabweichung (`sd`) die Verteilung haben soll:

```
rnorm(n = 20, mean = 0, sd = 1)
```

```
## [1]  1.55841478  0.12349688 -0.81843797  0.49951008 -1.67462373 -0.56036097
## [7] -0.19531351 -1.44516307 -1.58724874 -1.07656746 -1.41580396 -0.30508666
## [13] -1.04031622  0.81096622  1.34244122 -0.06229402 -0.95722602  0.03262094
## [19]  2.39932131  0.76793479
```

Chapter 7

Daten importieren und speichern

7.1 Working Directory

7.2 Working Environment

7.3 Daten importieren

7.4 Daten speichern

Chapter 8

Daten Explorieren

8.1 Informationen über R Objekt

```
#length()  
#str()  
#nrow()  
#ncol()  
#is.na()
```

8.2 Deskriptivstatistiken

8.2.1 Nominale Variablen

8.2.1.1 Absolute Häufigkeiten

8.2.1.2 Relative Häufigkeiten

8.2.2 Numerische Variablen

8.2.2.1 Lagemaße

8.2.2.2 Streuungsmaße

8.2.3 Alles auf einen Blick

8.2.4 Gruppenweise Deskriptivstatistiken

Chapter 9

Datenmanipulation

9.1 Löschen

9.2 Umbenennen

9.3 Verändern

9.4 Aggreggieren

9.5 Transformieren

9.5.1 Von Long nach Wide

9.5.2 Skalieren

Chapter 10

Graphiken

10.1 `ggplot()`

10.2 `geom_histogram()`

10.3 `geom_point()`

10.4 `geom_line()` und `geom_smooth()`

10.5 `geom_bar()`

10.6 `geom_boxplot()`

10.7 Facetting

10.8 Ästhetische Anpassungen

10.8.1 `ggtitle()`

10.8.2 `labs()`

10.8.3 Farben, Formen, Füllungen

10.8.4 Achsen verändern

10.9 `theme()`

10.10 Abbildungen kombinieren

10.11 Abbildungen exportieren

Chapter 11

Tabellen

Chapter 12

Preprocessing

12.1 Fehlende Werte

12.2 Ausreißer

12.3 Schiefe Daten

Chapter 13

Korrelation

13.1 Pearson Korrelation

13.2 Spearman Korrelation

13.3 Kendall Korrelation

13.4 Korrelationmatrix

Chapter 14

Regression

14.1 Lineare Regression

14.2 Logistische Regression

Chapter 15

Mehrstufige UVs

15.1 Omnibus Test

15.2 Paarweise Vergleiche

Chapter 16

Mehrere UVs

16.1 Multiple Regression

16.2 Interaktionsterme

Chapter 17

Manuskripterstellung