

# Studienbrief 3

## Einführung in R

### Wiederholungsfragen Studienführer 2

#### Aufgabe 2.0:

Bearbeiten Sie die restlichen Wiederholungsfragen unter folgenden Link:

[https://fomshinyapps.shinyapps.io/Lernfortschrittskontrolle\\_01/](https://fomshinyapps.shinyapps.io/Lernfortschrittskontrolle_01/)

sowie folgende Wiederholungsfragen:

[https://fomshinyapps.shinyapps.io/Lernfortschrittskontrolle\\_02/](https://fomshinyapps.shinyapps.io/Lernfortschrittskontrolle_02/)

Und, wie war es dieses Mal? Hat es wieder gut geklappt? Großartig!



### Einführung in R (Kapitel 4 im Skript)

Was ist das Ziel dieses Studienbriefes?

#### Lernergebnisse

Was erreichen Sie nach erfolgreichem Abschluss diesen Studienführers:

- Sie können Gründe für den Einsatz von R nennen.
- Sie kennen die Unterschiede zwischen R, RStudio und mosaic.
- Sie können Datensätze laden, einfache Befehle eingeben und wichtige Grundbefehle in R ausführen.



Bearbeiten Sie die Folien 74 bis 79. Sie sehen, wir und Sie nutzen R, weil es ein weitverbreitetes und frei verfügbares Statistikprogramm ist. In Verbindung mit R (das eigentliche Statistikprogramm) wird RStudio, ein ebenfalls frei verfügbarer Editor genutzt. Und innerhalb R wird noch mosaic eingesetzt, ein sogenanntes Paket (*package*). Sie können R beim Studium nutzen, zu Hause oder auch in der Firma (wie es viele Unternehmen auch machen).

**Tipp:** Spätestens jetzt sollten Sie R und die Zusatzprogramme installiert haben oder den Zugang über die RStudio-Cloud eingerichtet haben. Wenn Sie die lokale Version nutzen, denken Sie daran das RStudio-Projekt zu kopieren, wie in den Dokumenten im Online-Campus beschrieben. So ist sichergestellt, dass wir alle das gleiche sehen und die gleichen Dateien etc. zur Verfügung haben.



Wenn Sie bis Folie 82 weiterlesen bekommen Sie einen Eindruck, wie die Befehle in R aufgebaut sind, insbesondere, wenn Sie das Paket mosaic nutzen. Folie 82 zeigt den Aufbau von RStudio, dem Editor, mit dem wir in der nächsten Zeit viel arbeiten werden.

### Aufgabe 3.1:

Warum ist Codieren eine wichtige Fähigkeit?

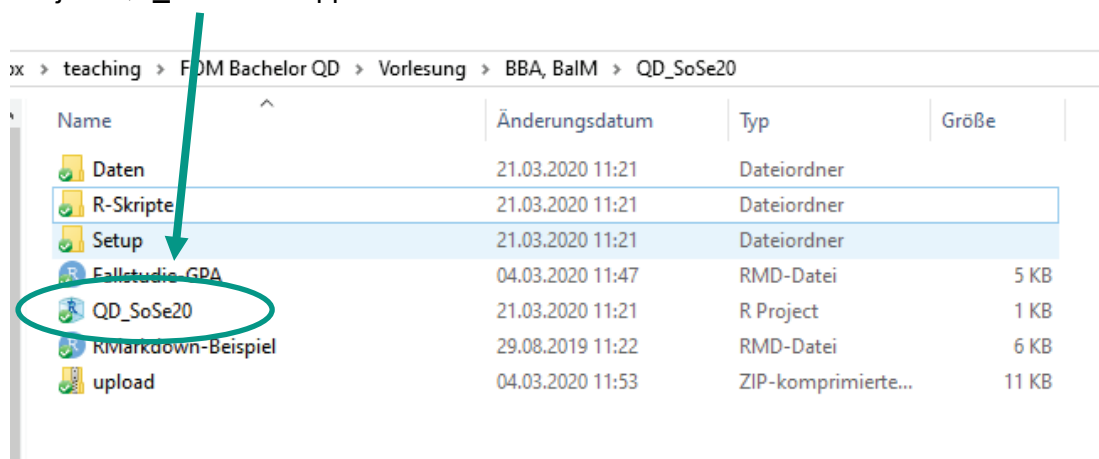


Alles installiert? Oder RStudio-Cloud-Zugang eingerichtet? Dann können wir loslegen.

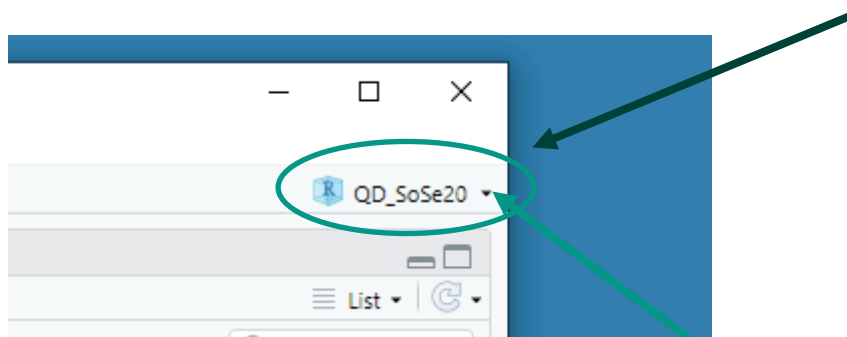
Hier gibt es zwei Möglichkeiten:

### Sie arbeiten lokal

Starten Sie R und RStudio, in dem Sie im Verzeichnis *QD\_SoSe20* auf das R-Projekt *QD\_SoSe20* doppelklicken:



Wenn Sie RStudio schließen und das nächste Mal wieder aufrufen, wird das Projekt automatisch wieder geladen. Sie erkennen das in RStudio oben rechts:

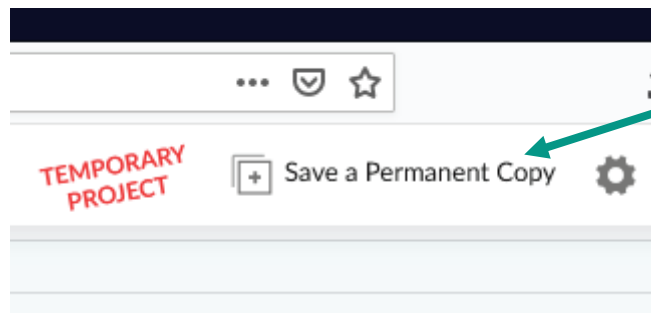


Falls dort nicht *QD\_SoSe20* stehen sollte, wählen Sie das Projekt aus der Dropdown-Liste aus, die geöffnet wird, wenn Sie auf den Pfeil rechts drücken. Und

wenn das auch nichts hilft, dann gehen Sie zum Anfang dieses Abschnitts und starten das Projekt, wie oben erklärt.

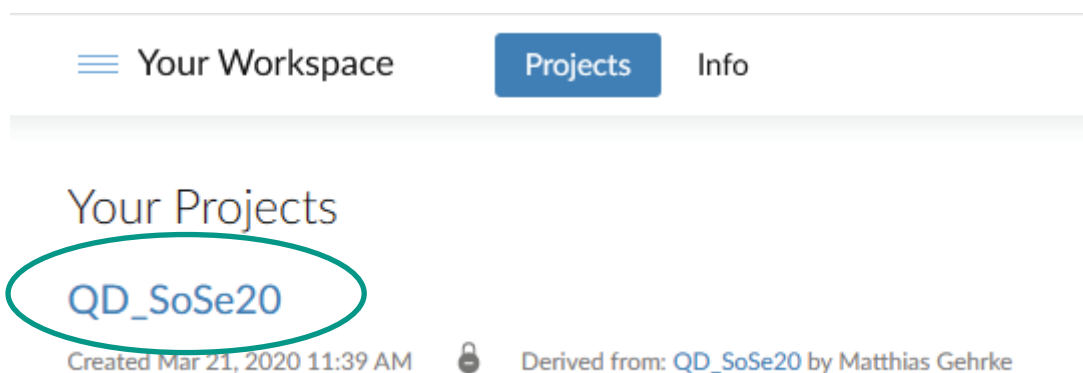
### Sie arbeiten in der RStudio-Cloud

In der RStudio-Cloud Version startet das Projekt automatisch durch Aufruf auf den Link <https://rstudio.cloud/project/1008008>. Wichtig, Sie müssen das Projekt noch als permanenten Kopie in Ihrem eigenen Workspace ablegen. Klicken Sie dazu hier:



Sonst fehlen Ihnen beim nächsten Mal die Sachen, die Sie gemacht haben ☹️.

Beim nächsten Login in RStudio-Cloud wird das Projekt automatisch aufgerufen bzw. Sie bekommen eine Liste, aus der Sie das QD\_SoSe20-Projekt wieder aufrufen können:

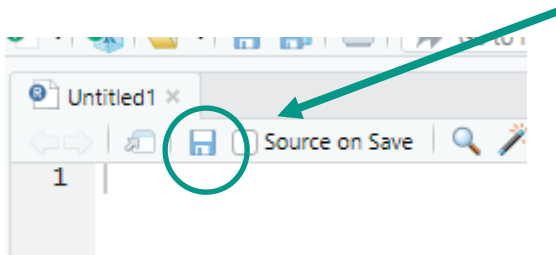


### Verzweigung Ende – jetzt geht es wieder für alle gemeinsam weiter.

Bearbeiten Sie Folien 83 bis ... Auf den Folien stehen die Befehle, die Sie in R eingeben müssen. Wo sollen Sie die eingeben?

Sie könnten alles direkt in der Console eingeben (Fenster unten links, siehe Folie 82), aber da wird nichts abgespeichert. Daher erstellen wir ein neues R-Skript, mit dem wir dann im Fenster oben links arbeiten können (was bei Ihnen aber noch nicht zu sehen ist).

Daher öffnen Sie bitte ein neues R-Skript mit *File – New File – R Skript*. Speichern Sie das Skript ab, in dem Sie auf die kleine blaue Diskette klicken:



Wie Sie die Datei nennen ist eigentlich egal. Aber achten Sie darauf keinen Punkt im Dateinamen zu haben, dann ist es kein R-Skript mehr. Also z. B. `R_200321` (`R_` mit dem Datum des heutigen Tages in der Schreibweise `R_JJMMTT`).

**Tip:** Rechts unten sehen Sie im Tab Files jetzt die neue R-Datei. Hier können Sie immer sehen, welche Dateien Sie angelegt haben, oder auch, welche ich schon vorbereitet hatte (z. B. im Verzeichnis Setup – siehe Installationshinweise).

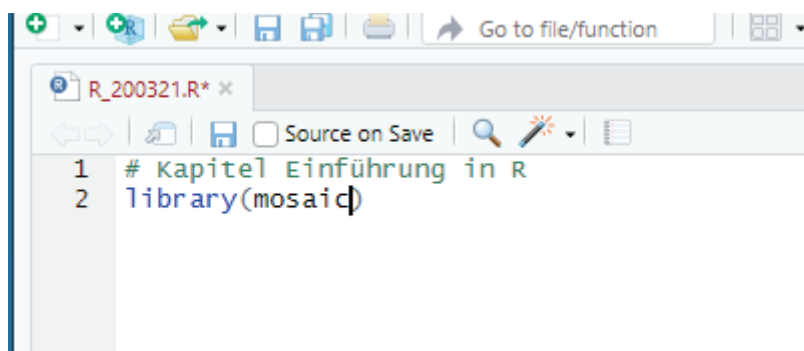


Jetzt haben Sie wie in der Folie 82 oben rechts das Fenster, in dem Sie alle Befehle eingeben. Diese werden dann in der R-Skript-Datei gespeichert.

Der erste Befehl ist bei uns immer `library(mosaic)`. Beim Aufruf kommen viele Meldungen wie auf Folie 83 zu sehen ist. In Wirklichkeit sind Meldung rot, aber es sind keine Fehlermeldungen, sondern nur Hinweise.

Bevor Sie mit eigentlichen Befehlen loslegen, können Sie im R-Skript noch einen Kommentar einfügen, z. B.: `# Kapitel Einführung in R`.

Das Ganze sieht dann so aus:



**Tip:** Sie sehen unterschiedliche Farben. Kommentare werden mit # eingeleitet und in grün angezeigt. Der eigentliche Befehl wird in blau angezeigt.



Bisher haben wir noch keinen Befehl ausgeführt. Gehen Sie mit den Cursor irgendwo in die Zeile, in der `library(...)` steht (bei mir die Zeile 2 – siehe linker Rand) und drücken die Tastenkombination `Ctrl+S` bzw. auf dem Mac `Cmd+S`.

Jetzt wird der Befehl „abgeschickt“ und taucht unten in der Console mitsamt den ganzen Meldungen in Rot auf.

Wenn es tatsächlich eine Fehlermeldung gibt, z. B. so:

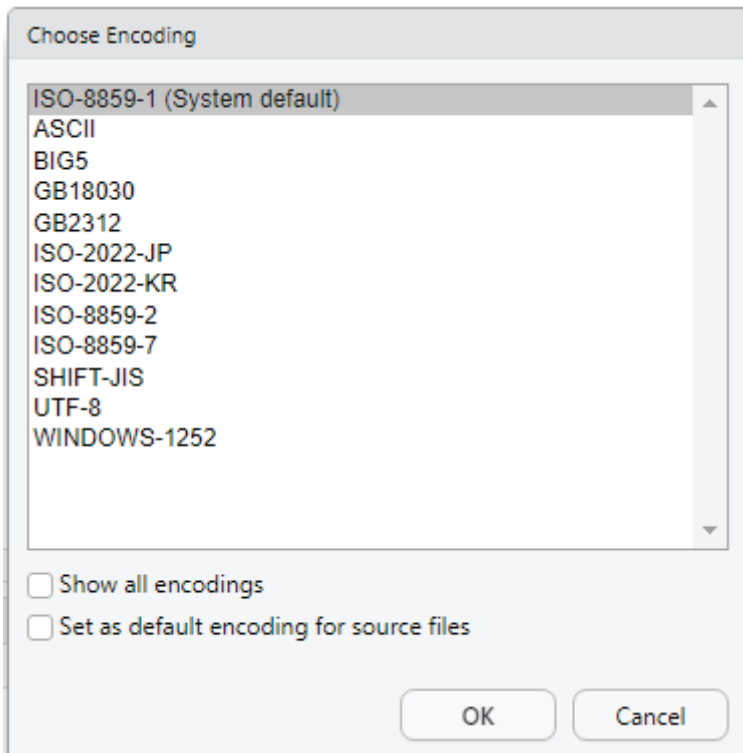
```
> # Kapitel Einführung in R
> library(mosaic)
Fehler in library(mosaic) : es gibt kein Paket namens 'mosaic'
> |
```

dann müssen Sie das Paket `mosaic` noch installieren und danach ausführen.

**Tip:** Einzelne Pakete können Sie mit `install.package("package")` nachinstallieren.



Speichern Sie jetzt das R-Skript noch einmal, indem Sie wieder auf die blaue Diskette drücken. Es kann sein, dass jetzt folgendes Fenster auf geht:

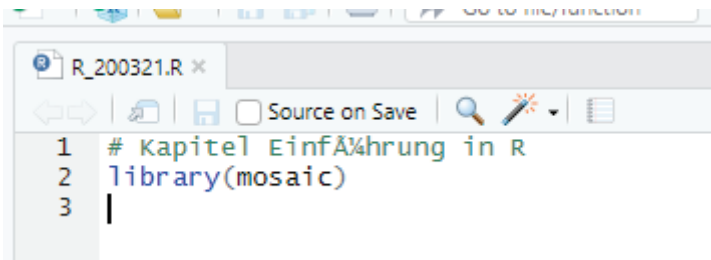


Wählen Sie Ihr *System default* aus (beim Mac ist das UTF-8) und machen das Häkchen bei *Set as default encoding for source files* an. Wenn Sie OK drücken, kommt später dieses Fenster nicht mehr.

### Was ist Encoding?

Zeichensätze sind auf den verschiedenen Computer-Betriebssystemen unterschiedlich codiert. Das betrifft insbesondere Umlaute und Sonderzeichen. Auf Windows-Rechnern sind die Zeichen *ISO-8859-1* codiert, auf Mac *UTF-8* und wie

Sie sehen, gibt es noch wie viel mehr mögliche Codierungen. Normalerweise können Sie das einmal festlegen und dann vergessen. Problematisch wird das nur, wenn Sie z. B. auf einem Windows-Rechner arbeiten, Ihr Kommilitone\*in auf einem Mac und Sie tauschen die R-Skripte aus. Sie lesen z. B. einen Mac-codierten R-Skript ein, dann sieht die erste Zeile so aus:



```

1 # Kapitel Einführung in R
2 library(mosaic)
3 |

```

Immer wenn so etwas passiert, können Sie die Datei mit *File – Reopen with Encoding* neu öffnen, in dem Sie die Codierung auswählen, in der R-Skript abgespeichert ist.

**Tip:** R-Dateien, die ich im Online Campus einstelle oder in den R-Projekten zur Verfügung gestellt habe, sind immer *UTF-8* codiert.



### Weiter geht's

Jetzt bearbeiten Sie bitte eigenständig die Folien 83 bis 98.

**Tip:** Auf Folie 85 (und später noch häufiger) wird die sogenannte Tilde ~ verwendet. Diese verbindet die abhängige Variable mit der (oder den) unabhängigen Variablen. Auf Windows-Rechnern bekommen Sie diese mit der Tastenkombination `Alt + ~`, auf Mac mit `Cmd + ~`.



Auf Folie 91 kommt noch die Zeichenfolge `%>%`, die sogenannte *Pipe* dazu. Die Pipe übergibt das Ergebnis der Funktion, die links steht, an die Funktion, die rechts von der Pipe steht. Auch dafür gibt es ein Kürzel: Windows `Alt + M` bzw. Mac `Cmd + M`.

**Tip:** Befehle, die Sie nur einmal ausführen wollen, die also gar nicht zu eigentlichen Bearbeitung gehören, wie z. B. `install.packages` oder `?select` (Hilfe zur Funktion `select`) führen Sie am besten in der Console aus. Dann werden Sie nicht im R-Skript gespeichert.



Und, hat alles geklappt? Am Anfang ist die Eingabe von Befehlen etwas ungewohnt, aber mit der Zeit bekommen Sie Übung darin – und dann geht es so viel schneller als über Menübefehle.

---

**Aufgabe 3.1:**

Bringen Sie die Variablentypen auf Folie 97 und die Skalenniveaus aus dem Kapitel *Grundlagen quantitativer Datenanalyse* zusammen. Was wird idealerweise mit welchem Variablentype abgespeichert?



**Tip:** Die Lösung finden Sie Ende dieses Studienbriefes. – Aber immer erst versuchen, selbst die Lösung zu finden. Ihr Lernerfolg kommt durch die aktive Auseinandersetzung mit einer Fragestellung!



Schauen Sie sich jetzt die Folien 99 bis 101 an. Vielleicht haben Sie ja auch vorher schon einmal auf die Folien 100 oder 101 gelunzt – sehr gut!

---

**Aufgabe 3.2:**

Nehmen Sie bitte jetzt an der Umfrage auf Folie 102 teil. Bitte beantworten Sie die dort gestellten Fragen ohne zusätzliche Hilfsmittel zu nutzen oder im Internet zu recherchieren. Das ist keine Prüfung, sondern hilft uns, anonyme Daten zu sammeln, wie Sie jetzt mit Statistik zurechtkommen. Am Ende kommt die Umfrage noch einmal und dann geht für Sie hoffentlich schon viel einfacher 😊.



Und Statistik hat ja immer mit Daten zu tun.

Zum Abschluss dieses Studienbriefes schauen Sie sich bitte die Folie 103 an. Hier sehen Sie ein Beispiel eines quantitativen Forschungsprozesses (Sie erinnern sich, dass hatten wir schon am Ende des zweiten Studienbriefes).

---

**Lernpfad: Einführung in R**

- R und RStudio oder RStudio in der Cloud-Version sollten jetzt zuverlässig funktionieren.
- Bitte schauen Sie sich insbesondere die Folien 78 bis 98 im Skript sorgfältig an. Klären Sie erforderlichenfalls Ihnen noch unbekannte Begriffe eigenständig. Wichtig ist, dass Sie selbst mit R und RStudio arbeiten. Übung macht den Meister oder die Meisterin!
- Schauen Sie sich noch folgenden Screencast zum Datenhandling in R an: <https://vimeo.com/398512955> (Passwort: FOMtest). Damit lernen Sie noch mehr Funktionen zum Datenhandling kennen.
- Zur Vertiefung können Sie die Kapitel 3 und 4 aus dem Buch von Sebastian Sauer *Moderne Datenanalyse mit R* durchgehen. Das Buch ist über die FOM-Literaturrecherche frei verfügbar (Link siehe Literaturhinweise im Online Campus).



- Zur Ergänzung bearbeiten Sie bitte die Aufgaben in Kapitel 0 des Übungsheftes.
  - Zur Selbstkontrolle schauen Sie in das Lösungsheft.
- 

## Zusammenfassung

Halten wir fest, was Sie in diesem Kapitel gelernt haben:

- Sie können die ersten Befehle mit R und RStudio ausführen.
- Sie wissen, warum Sie mit R arbeiten.
- Sie kennen wichtige Variablentypen in R und wissen, wann Sie diese einsetzen sollten.

Im nächsten Studienbrief geht es um die Explorative Datenanalyse. Die dazugehörige Sitzung werden wir mit einem Webinar beginnen, so dass Sie alle Fragen stellen können, die Ihnen bis dahin gekommen sind. Viel Spaß weiterhin!



## Lösungshinweise

### Aufgabe 3.1

Metrische Daten werden in der Regel als `numeric` oder `double` abgespeichert, unabhängig davon, ob sie intervall- oder verhältnisskaliert sind. Wenn metrische Daten diskret sind (z. B. Anzahl Studierender) ist `integer` das geeignete Format.

Nominalskalierte Variablen (kategoriale Variablen) werden bei vielen Ausprägungen (Kategorien) als `character` abgespeichert, bei wenigen ist `factor` geeignet. Nach Faktoren kann später gruppiert werden.

Ordinalskalierte Variablen sind eigentlich auch kategoriale Variablen. Auf Folie 97 sehen Sie daher bei `wday` die (selten genutzte) Möglichkeit, diesen als `Ord.factor`, also geordnete Faktor-Variablen abzuspeichern. Häufig werden den Kategorien von ordinalskalierten Variablen aber auch ganzzahlige Werte zugewiesen (siehe Ende der Zeile von `wday`) und die Variable dann als `integer` abgespeichert.

### Hinweise

Dieser Studienbrief wurde von Matthias Gehrke auf Basis des Konzepts von Roswitha Grassl (beide FOM, <https://www.fom.de/>) entwickelt und stehen unter einer Creative Commons Namensnennung – Nicht-kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Lizenz (<https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode.de>).