

Übungsblatt 6: Weitere Features von R Markdown

Wissenschaftliches Schreiben und Arbeiten, WiSe 2023/'24

Hinweis: Für dieses Übungsblatt kannst du alle Aufgaben im selben R-Markdown-Dokument bearbeiten. Bitte nenne es `abgabe_<Vorname>_<Nachname>.Rmd`, wobei du die Platzhalter `<Vorname>` und `<Nachname>` durch deinen Vor- und Nachnamen ersetzen solltest. Am Ende musst du diese .Rmd-Datei, eine daraus kompilierte .pdf-Datei sowie eine .bib-Quellenverzeichnis-Datei abgeben. Es wäre ratsam, dass du das Dokument während des Bearbeitens der Aufgaben immer mal wieder im jeweils aktuellen Zwischenstand kompilierst, um immer wieder sicherzugehen, dass sich bisher noch keine Fehler eingeschlichen haben.

Aufgabe A: Dokument initialisieren

1. Erstelle ein neues R-Markdown-Dokument. In der Präambel solltest du als **Titel** „Übungsblatt 6“ festlegen, als **Autor** deinen Vor- und Nachnamen, als **Datum** das aktuelle Datum und schließlich als **Output-Format** bitte PDF.

Aufgabe B: R-Code-Block nicht ausführen

1. Erstelle in deinem R-Markdown-Dokument einen Abschnitt mit Titel „Aufgabe B“.
2. Erstelle in diesem Abschnitt einen R-Code-Block, der zwar angezeigt, aber **nicht ausgeführt** werden soll. Der R-Code-Block soll nur folgenden Befehl enthalten:

```
print("Dieser print-Befehl wird nicht ausgeführt.")
```

Tipp: Um einen neuen R-Code-Block schnell zu erstellen, kann man in RStudio auch einfach die Tastenkombination Strg+Alt+I benutzen.

Aufgabe C: Tabellen und Grafiken

1. Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe C“.
2. Verschiedene Arten von Tabellen:
 - a. Schreibe als Fließtext im neuen Abschnitt Folgendes auf:
„Beispiel einer durch R-Code erzeugten, tabellenartigen Ausgabe:“
 - b. Erstelle direkt darunter einen R-Code-Block, der ausgeführt, aber **nicht angezeigt** werden soll.
 - c. Schreibe in diesem neuen R-Code-Block einen Befehl auf, der den vorinstallierten Datensatz `airquality` als Datenrahmen in die aktuelle R-Arbeitsumgebung lädt.
 - d. Schreibe anschließend im selben R-Code-Block einen Befehl auf, der die ersten sechs Zeilen des Datenrahmens `airquality` ausgibt.
 - e. Schreibe dann als Fließtext darunter Folgendes auf:
„Beispiel einer einfachen, mit Markdown-Syntax erzeugten Tabelle:“
 - f. Erstelle darunter mithilfe der Markdown-Syntax folgende Tabelle:

Modultitel	CP	Kürzel	Prüfungs-Anmeldefrist
Selbstreflexion und wissenschaftliches Arbeiten	6	LIN-BS-040	08.02.2024

- g. Schreibe danach wiederum als Fließtext Folgendes auf:
„Beispiel einer mit der `knitr::kable`-Funktion erzeugten Tabelle:“
 - h. Erstelle darunter mithilfe der `knitr::kable`-Funktion noch einmal die gleiche Tabelle wie in Schritt f., aber diesmal sollen alle Spalten zentriert (statt linksbündig) formatiert sein. Außerdem soll die Tabelle mit der Caption „Informationen zum Modul LIN-BS-040“ betitelt sein.
 - i. Schreibe anschließend als Fließtext Folgendes auf:
„Ausgabe eines Teils des `airquality`-Datensatzes als formatierte Tabelle mit `knitr::kable`:“
 - j. Gib nun den gleichen Teildatensatz von `airquality` wie in Schritt d. erneut aus, aber diesmal nicht als rohe Konsolen-Ausgabe, sondern als formatierte Tabelle mit `knitr::kable`. Dabei sollen auch Zeilenlabels (eine einfache Nummerierung der Zeilen) in einer zusätzlichen Spalte ganz links mit angezeigt werden.
3. Verschiedene Arten von Grafiken:
- a. Schreibe im Weiteren als Fließtext Folgendes auf:
„Beispiel einer durch R-Code und mithilfe von `ggplot2` erzeugten Histogramm-Grafik:“
 - b. Lass dir darunter von einem unsichtbaren R-Code-Block eine `ggplot2`-Grafik ausgeben, die ein Histogramm der Windstärken-Werte aus der Spalte `Wind` des Datensatzes `airquality` darstellt. Dabei soll jeder Balken des Histogramms zweieinhalb Einheiten breit sein.
Tipp: Du kannst dieses Histogramm in etwa nach der Blaupause des Codes erstellen, der in der [letzten Vorlesung auf Folie 16](#) gezeigt wurde.
 - c. Schreibe hierunter nun als Fließtext Folgendes auf:
„Beispiel einer einfachen, mit Markdown-Syntax importierten Bilddatei-Grafik:“
 - d. Lade dir die Bilddatei `beispiel_bild_vogel.jpg` aus dem Moodle herunter und binde sie dann unter Verwendung der Markdown-Syntax im Dokument ein.
Du kannst sie dabei mit der Caption „Ein zufälliger Vogel“ betiteln.
 - e. Schreibe darunter Folgendes auf:
„Beispiel einer mit `knitr::include_graphics` importierten Bilddatei-Grafik:“
 - f. Erstelle nun einen unsichtbaren R-Code-Block, der die Bilddatei `beispiel_bild_vogel.jpg` erneut einbindet, diesmal allerdings unter Verwendung von `knitr::include_graphics`. Stelle dabei durch geeignete Optionen im Header des R-Code-Blocks sicher, dass die Grafik diesmal zentriert angezeigt wird und nur 40% der Textbreite einnimmt. Setze als Caption der eingebundenen Grafik diesmal „Ein zufälliger Vogel (kleiner)“ fest.

Aufgabe D: Fußnoten, Literaturverzeichnis, Zitieren

1. Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe D“.
2. Erstellung von Fußnoten:
 - a. Schreibe in diesem neuen Abschnitt zunächst folgenden Satz auf:
„Lorem ipsum dolor sit amet, consectetur adipiscing elit.“
 - b. Erstelle eine Fußnote mit einem Verweis zu ihr direkt nach Ende des obigen Satzes. Die Fußnote selbst soll Folgendes enthalten:
„Das ist kein echtes Latein, sondern nur ein beliebter sinnloser Platzhalter-Satz.“
 - c. Schreibe darunter dann diesen weiteren Satz auf:
„Ceterum censeo Carthaginem esse delendam.“
 - d. Erstelle eine zweite Fußnote mit einem Verweis zu ihr direkt nach dem Ende dieses weiteren Satzes. Die Fußnote selbst soll Folgendes enthalten:
„Aber das hier ist ein echter lateinischer Satz.“

3. Literaturverzeichnis und Zitieren:

- Geh auf [Google Scholar](#) und finde dort durch Stichwort-Suche mindestens zwei wissenschaftliche Paper, die du interessant findest.
- Lass dir Zitiervorlagen für die gefundenen Paper im BibTeX-Format herausgeben. Speichere diese dann in einer neuen .bib-Datei ab, der du den Dateinamen `abgabe_<Vorname>_<Nachname>.bib` geben kannst. Platziere diese .bib-Datei in deinem aktuellen R-Arbeitsverzeichnis.
- Füge anschließend eine Option zur Präambel deines Dokuments hinzu, die die Quelldatenbank `abgabe_<Vorname>_<Nachname>.bib` fürs Zitieren verfügbar macht.
- Lade dir die Datei `apa.csl` aus dem Moodle herunter und platziere sie in deinem aktuellen R-Arbeitsverzeichnis.
(In dieser Datei ist der Zitierstil der American Psychological Association, kurz APA, enkodiert.)
Füge dann eine Option zur Präambel deines Dokuments hinzu, die dafür sorgt, dass der Zitierstil aus der Datei `apa.csl` geladen und verwendet wird.
- Verfasse nun in deinem R-Markdown-Dokument einen kurzen Paragraphen, in dem du ganz grob die Kernaussagen der von dir ausgesuchten Paper zusammenfasst. Dabei sollst du mindestens einmal einen Zitierverweis der Form „AUTOR:INNEN (JAHR)“ und mindestens einmal einen Zitierverweis der Form „(AUTOR:INNEN, JAHR)“ verwenden.
- Sorge zuletzt dafür, dass ganz unten in deinem R-Markdown-Code ein Abschnittstitel namens „Literaturverzeichnis“ definiert wird, sodass unmittelbar nach diesem das automatisch generierte Literaturverzeichnis mit deinen zitierten Quellen erscheint.

Aufgabe E: Mathematische Ausdrücke formatieren

- Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe E“.
- Schreibe Folgendes auf, wobei du die enthaltene Formel als mathematischen Ausdruck mit der Syntax des LaTeX-Mathe-Modus formatieren sollst:
„Der Satz des Pythagoras lautet $a^2 + b^2 = c^2$.“
- Schreibe nun Folgendes auf: „Das hier ist der zweite Teil des Fundamentalsatzes der Analysis.“
- Schreibe darunter in einer eigenen Zeile und zentriert folgende mathematische Formel auf:

$$\int_a^b f(x)dx = F(b) - F(a)$$

Tipp: Um das Symbol des bestimmten Integrals mit Grenzen a und b aufzuschreiben, also \int_a^b , kannst du im LaTeX-Mathe-Modus einfach `\int_a^b` tippen.

Aufgabe F: Diverse LaTeX-Befehle in R Markdown verwenden

- Füge gleich zu Beginn des Hauptteils deines R-Markdown-Dokuments (also direkt nach der Präambel) folgende zwei LaTeX-Befehle ein:

```
\tableofcontents  
\newpage
```

Wenn du das Dokument jetzt kompilierst, sollte auf der ersten Seite des PDFs ein automatisches Inhaltsverzeichnis im LaTeX-Stil erscheinen, anschließend ein Seitenumbruch.

- Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe F“.
- Benutze einen passenden LaTeX-Textformatierungs-Befehl, um in diesem neuen Abschnitt folgenden Satz in Kapitälchen-Schriftart aufzuschreiben:

DIESER SATZ ERSCHEINT IN KAPITÄLCHEN-SCHRIFTART.

Aufgabe G: Fehlende Daten (NA)

1. Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe G“.
2. Erstelle einen sichtbaren R-Code-Block, in dem das Minimum und das Maximum der Spalte `Solar.R` (Messwerte solarer Strahlung) aus dem Datensatz `airquality` berechnet wird.
Tipp: Beachte, dass diese Spalte fehlende Werte (NAs) enthält! Daher musst du beim Aufrufen der Funktionen für Minimum und Maximum noch ein zusätzliches Argument angeben, damit die fehlenden Werte ignoriert werden – sonst kommt als Ergebnis einfach nur NA heraus. Weißt du noch, welches?
3. Lass dir, weiterhin im selben R-Code-Block, eine Übersicht des Datensatzes `airquality` mit dem `summary(...)`-Befehl herausgeben.
Du solltest darin erkennen können, welche Wertspalten wie viele NAs enthalten.
4. Entferne aus `airquality` alle Zeilen, die mindestens ein NA enthalten. Speichere den so reduzierten Datensatz in einer neu definierten Variablen namens `airquality_nur_volle_Zeilen`.
5. Lass dir nun mit `summary(...)` eine Übersicht des Datensatzes `airquality_nur_volle_Zeilen` herausgeben.
Stelle sicher, dass jetzt keine NAs mehr in der Übersicht angezeigt werden.

Aber fällt dir noch etwas auf? Die in der Übersicht genannten Mittelwerte (Median, Durchschnitt) haben sich im Vergleich zum originalen, vollen Datensatz `airquality` ebenfalls verändert!

Der Grund ist, dass auch Zeilen gelöscht wurden, denen z.B. nur der Wert für `Ozone` fehlte, die aber durchaus einen Wert für `Solar.R` hatten (oder umgekehrt). Dadurch haben wir auch existierende, wertvolle Informationen aus dem Datensatz entfernt.

Daher ist so ein radikaler Schritt wie hier, nämlich alle teilweise unvollständigen Datenzeilen zu löschen, nur in seltenen Fällen wirklich angebracht!

Aufgabe H: Eine eigene Funktion definieren

1. Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe H“.
2. Erstelle einen sichtbaren R-Code-Block, in dem du eine neue Funktion definierst, die in einer Variablen `log_rund` gespeichert wird.
Die Funktion `log_rund` soll einen Zahlenwert `x` als Argument nehmen. Anschließend soll sie den Logarithmus von `x` berechnen und ihn, auf die 2. Nachkommastelle gerundet, als Ergebnis herausgeben.
Tipp: Um den Logarithmus einer `Zahl` zu berechnen, kann man in R `log(Zahl)` benutzen. Und um eine `Zahl` auf die 2. Nachkommastelle zu runden, kann `round(Zahl, 2)` verwendet werden.
3. Prüfe nun, weiterhin im selben R-Code-Block, ob deine neu definierte Funktion `log_rund` korrekt erkannt wird, indem du sie auf den Zahlenwert 100 anwendest.
Dabei sollte, wenn du alles richtig gemacht hast, als Ergebnis exakt 4.61 herauskommen.

Aufgabe I: Die Hilfsfunktionen `tapply` und `sapply`

1. Erweitere dein R-Markdown-Dokument um einen Abschnitt mit Titel „Aufgabe I“.
2. Erstelle einen sichtbaren R-Code-Block und wende darin wie folgt die `tapply`-Hilfsfunktion auf den Datensatz `iris` an: Berechne, gruppiert nach Kategorie-Labels aus der Spalte `Species`, den Median der Zahlenwerte aus der Spalte `Petal.Width` (Blütenblattbreite).
3. Entferne die Textwert-Spalte `Species` aus dem Datenrahmen `iris` und speichere den so reduzierten Datenrahmen in einer neuen Variablen namens `iris_ohne_Species`.
(Dieser neue Datenrahmen enthält jetzt nur noch vier Zahlenwert-Spalten, nämlich `Sepal.Length`, `Sepal.Width`, `Petal.Length` und `Petal.Width`.)
4. Wende die `sapply`-Hilfsfunktion so auf `iris_ohne_Species` an, dass für *jeden* Zahlenwert aus *jeder* der vier Spalten die Ausgabe deiner eigens definierten Funktion `log_rund` (s. Aufgabe H) berechnet wird. Lass dir die ersten sechs Zeilen der dadurch generierten Ergebnis-Matrix herausgeben.

Aufgabe J: Kursevaluation (freiwillig, anonym)

1. Das offizielle Evaluations-Formular für diesen Kurs, also „Wissenschaftliches Schreiben und Arbeiten“, ist jetzt online: <https://pep.uni-potsdam.de/ce/c4a2fad6/de.html>
Hier hast du die Gelegenheit, komplett anonymisiert zu bewerten, wie dir die Lehre in diesem Kurs gefallen hat bzw. welche Aspekte deiner Meinung nach verbessert werden könnten.
Die Teilnahme an der Evaluation ist freiwillig. Bitte nimm dir aber, wenn du es dir einrichten kannst, die paar Minuten Zeit, um das Formular auszufüllen und so dein ehrliches Feedback abzugeben.

Abgabe:

Lade bis zum 25.01.24 um 23:59 Uhr folgende drei Dateien in der Abgabemaske im Kurs-Moodle hoch:

- abgabe_<Vorname>_<Nachname>.Rmd
- abgabe_<Vorname>_<Nachname>.bib
- abgabe_<Vorname>_<Nachname>.pdf