Einführung in Quarto

Dynamische und reproduzierbare Berichte mit R

Daniela Palleschi

Humboldt-Universität zu Berlin

Mi. den 11.01.2023

Überblick

- Quarto
- Unsere erstes Quarto-Dokument
- Codierung in Quarto
- Plots in Quarto

Heutige Ziele

- lernen, was dynamische Berichte sind
- unser eigenes Quarto-Dokument erstellen
- lernen, wie man ein Quarto-Dokument bearbeitet
- lernen, wie man Code in ein Quarto-Dokument einfügt
- ein Quarto-Dokument in verschiedenen Formaten wiedergeben

Lust auf mehr Praxis?

- wir können mehr lernen/mehr Übung bekommen, indem wir
 - Kap. 29 und 30 in Wickham et al. (2023) (speziell zu Quarto), oder
 - Ch. 2 in Nordmann & DeBruine (2022) (reproduzierbarer Arbeitsablauf im Allgemeinen)
 - Nordmann & DeBruine (2022) verwendet Rmarkdown-Skripte, während wir die nächste Generation verwenden werden: Quarto
 - wir sollten in Quarto immer noch in der Lage sein, genau die gleichen Dinge zu tun, wie sie in Rmarkdown vorgeschlagen werden

Wiederholung

Letzte Woche haben wir gelernt...

- was Datenrahmen sind
- den Unterschied zwischen kategorialen und kontinuierlichen Daten
- wie man Diagramme mit ggplot erstellt
- die richtige Darstellung für unsere Daten auszuwählen

Wiederholung: ggplot()

Sehen Sie sich diesen Code an. Was würde passieren, wenn wir ihn ausführen würden?

```
library(languageR)
library(tidyverse)
df_lexdec <- lexdec

fig_lexdec <- df_lexdec |>
ggplot() +
    aes(x = RT, fill = Class) +
    geom_histogram(position = "identity", alpha = .5) +
    theme_bw()
```

Wiederholung: ggplot()

Welche Darstellung in Abbildung 1 wird durch den folgenden Code erzeugt?

```
1 fig_lexdec1 <-
2   df_lexdec |>
3   ggplot() +
4   aes(x = RT, colour = Class) +
5   geom_density(alpha = .5) +
6   theme_bw()
```

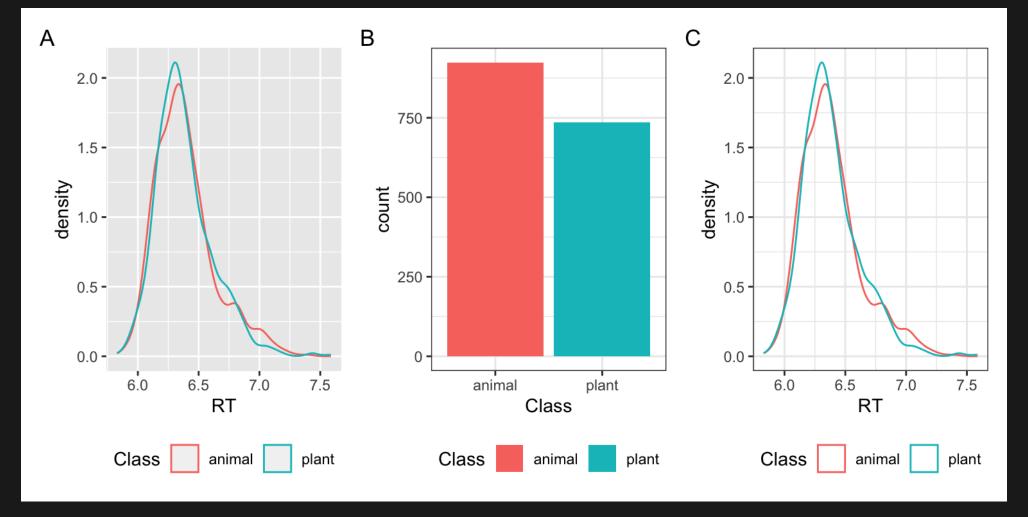


Abbildung 1: Drei aus dem lexdec-Datensatz generierte Diagramme

Set-up

- wir müssen eine LaTeX-Distribution verwenden, um PDF-Dokumente mit Quarto zu erstellen
 - LaTeX ist ein Schriftsatzsystem
 - TinyTex ist eine eigene LaTeX-Distribution, mit der wir PDFs erstellen können.
 - Das Paket tinytex kann uns helfen, TinyTex zu installieren

Installation von LaTeX über tinytex

- Führen Sie den folgenden Code in der Konsole aus
- oder, wenn Sie ihn in einem Skript ausführen wollen, um zu dokumentieren, was Sie getan haben, kommentieren Sie ihn nach der Ausführung aus (d.h. fügen Sie ein # davor)

```
1 # run this in the console
2 install.packages("tinytex")
3 tinytex::install_tinytex()
```

Ordner für Woche 3

- 1. Fügen Sie einen Unterordner mit dem Namen 03-quarto in Notes hinzu
- 2. Gehen Sie zu Moodle und speichern den Materialordner für '03 Einführung in Quarto' in Ihrem moodle Ordner
- 3. Öffnen Sie das Dokument _blatt.html auf Ihren Computer
 - Sehen Sie das Dokument an; Sie können oben rechts auf verschiedene Schaltflächen klicken. Probieren Sie es.

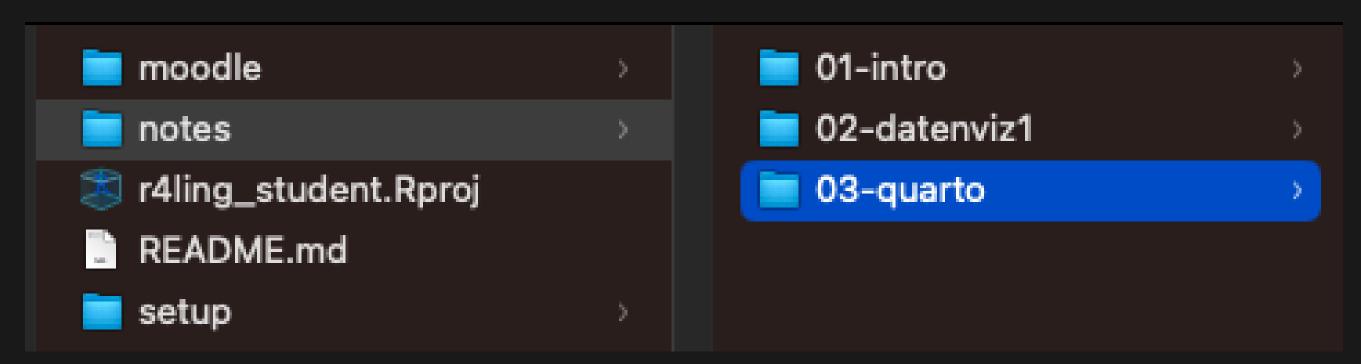


Abbildung 2: Notes folder structure

Quarto

- Quarto ist ein Dateityp, der dynamische Berichte erstellt
- Quarto-Dokumente sehen genauso aus wie ihr Vorgänger, Rmarkdown

Dynamische Berichte

- diejenigen, die Text, Code, Codeausgabe enthalten
- Quarto bietet ein "unified authoring framework" für Data Science, das Ihren Text, Ihren Code und Ihre Code-Ausgabe einschließt (Wickham et al., 2023, Kap 29.1)
- Quarto wurde entwickelt, um auf drei Arten verwendet zu werden:
- 1. Für die Kommunikation mit Entscheidungsträgern, die sich auf die Schlussfolgerungen und nicht auf den Code hinter der Analyse konzentrieren wollen.
- 2. für die Zusammenarbeit mit anderen Datenwissenschaftlern (einschließlich Ihnen in der Zukunft!), die sich sowohl für Ihre Schlussfolgerungen als auch für die Art und Weise interessieren, wie Sie zu ihnen gekommen sind (d. h. für den Code).
- 3. als eine Umgebung, in der Datenwissenschaft betrieben wird, als ein modernes Labornotizbuch, in dem wir nicht nur aufzeichnen können, was wir getan haben, sondern auch unsere Gedankengänge.

R v. Rmarkdown v. Quarto

- R -Dateien enthalten nur (R-)Quellcode
- Rmd dynamische Berichte mit
 - R-Code (und R-Pakete)
- qmd dynamische Berichte (RStudio v2022.07 oder später) mit
 - R-Code (und R-Pakete)
 - Native Unterstützung für Python (und Jupyter-Notebooks)
 - Native Unterstützung für Julia

R v. Rmarkdown v. Quarto



Aufgabe 1: RStudio version

Beispiel 1

- 1. Führen den folgenden Code in der Konsole aus: RStudio. Version () \$version
 - wenn die ausgegebene Version 2022 07 oder höher ist, können Sie Quarto benutzen
 - wenn nicht:
- 2. Aktualisieren Sie RStudio: Help > Check for updates

Markdown

- md-Dateien
- ein Klartext-Editor-Format, das
 - Formatierungselemente hinzufügt, die unabhängig von Gerät und Ausgabeformat sind (PDF, Word-Dokument, html...)
 - leicht zu lesen ist
- Markdown-Dokumente sind das Bindeglied zwischen unserem Quelldokument (qmd) und unserer Ausgabe (z.B. PDF)

Folder structure

- jede qmd sollte (normalerweise) in einem eigenen Ordner sein
 - d.h. es sollten nicht mehrere qmd Dateien im selben Ordner sein
- dies ist nur mein Vorschlag, um die Ordner ordentlich und organisiert zu halten
 - d.h., es gibt keinen technischen Grund dafür (die Dokumente laufen auch dann, wenn sie sich alle im selben Ordner befinden)
- werfen wir einen Blick auf einige meiner früheren und aktuellen Projektordner

Unsere erstes Quarto-Dokument

- letzte Woche haben wir ein R-Skript erstellt, das wir über Moodle eingereicht haben
- wir werden nun unsere erste qmd-Datei erstellen
- von nun an wird dies die Datei sein, die wir in Moodle einreichen (kein R-Skript)

Unsere erstes Quarto-Dokument



Aufgabe 2: erste Quarto

Beispiel 2

- 1. Erstellen Sie in Ihrem R-Projekt-Ordner, in dem ihr Ihre Kursunterlagen/Notizen aufbewahren, einen neuen Ordner für Woche 3
- 2. File > New Document > Quarto Document
 - Geben Sie ihm einen Titel wie "Quarto Woche 3"
 - Deaktivieren Sie die Option "open with Visual Editor".
- 3. Schauen das neue Skript an, um mehr über Quarto zu erfahren.
- 4. Klicken Sie auf die Schaltfläche "Render" am oberen Rand des Dokuments
 - Speichern Sie das Dokument in dem Ordner für Woche 3, den Sie gerade erstellt haben.
 - Was geschiehen? Vergleichen die Ausgabe mit dem Quellcode des Dokuments.
- 5. Gehen Sie zurück zu Ihrem neuen Ordner 03-quarto
 - Was hat sich geändert?

Quarto-Grundlagen

- Quarto-Dokumente (wie Rmarkdown) enthalten drei wichtige Arten von Inhalten:
 - 1. den YAML-Header, der von —— umgeben ist
 - 2. Text mit einer einfachen Formatierung oder Strukturierung wie # Überschrift oder *Kursivschrift*
 - 3. R-Code-Chunk, umgeben von ```{r} ````

```
\``\{r\}
#| code-line-numbers: false
# Dies ist ein Code Chunk
1 + 1
\```
```

[1] 2

YAML

- stand ursprünglich für Yet Another Markup Language
 - wurde aber in YAML Ain't Markup Language umbenannt, um den Zweck der Sprache als datenorientiert und nicht als Dokumentauszeichnung zu betonen (laut Wikipedia)
- enthält alle Metainformationen zu Ihrem Dokument
 - z.B. Titel, Autorenname
- auch Formatierungsinformationen
 - z.B. Typ der Ausgabedatei
- es gibt viele Möglichkeiten der Dokumentformatierung und -anpassung, die wir in diesem Kurs nicht behandeln werden
 - aber ich habe zum Beispiel viele YAML-Formatierungsoptionen im Quellcode meiner Folien

YAML



Aufgabe 3: YAML

Beispiel 3

- 1. Ändern Sie den Titel, wenn Sie das tun möchten.
- 2. Raten Sie, wie man einen "Untertitel" (EN: subtitle) hinzufügen könnte (Hinweis: es ist ähnlich wie beim Hinzufügen eines title)
- 3. Fügen Sie einen Autor hinzu, Autor: "vorname nachname" (siehe Beispiel unten)
- 4. Füge ein Inhaltsverzeichnis hinzu (EN: Table of Contents, toc), indem du format so änderst, dass es wie folgt aussieht:

```
2 title: "Quarto - Woche 3"
  author: "Vorname Nachname
  format:
    html:
      toc: true
```

5. Rendern nun das Dokument. Sehen Sie Ihre Änderungen?

Strukturierung Ihres Dokuments

- wir können unser Dokument strukturieren mit
 - # Überschriften
 - ## Zwischenüberschriften
 - ### Unter-Zwischenüberschriften, usw.

Strukturierung eures Dokuments

```
title: "Quarto - Woche 3"
author: "Vorname Nachname"
format:
  html:
    toc: true
# Überschrift 1
Hier ist ein Text über das Thema, das mit dieser Überschrift verbunden ist.
# Überschrift 2
Hier ist ein weiterer Text zu einem anderen Thema.
## Unterüberschrift 2.1
Dies ist ein Text liber das Unterthema
```

Strukturierung Ihres Dokuments

i Die Bedeutung der Formatierung

Zwischenüberschrift anstelle von ##Zwischenüberschrift, um als Überschrift gelesen zu werden. YAML erfordert außerdem einen sehr präzisen Schriftsatz. Da die Abstände in der YAML (und anderswo) so wichtig sind, möchte ich die Leerzeichen sehen und zählen können. Um dies zu tun, geht in RStudio:

- gehen zu Ihren Globalen Einstellungen (Werkzeuge > Globale Einstellungen)
- unter Code (linke Spalte) > Display (Tab), markieren das Kästchen > Show whitespace character

Strukturierung Ihres Dokuments



Aufgabe 4: Überschriften

Beispiel 4

- 1. Kopiern den obigen Code (Überschriften und Unterüberschriften) und ersetzen den Text in der Quarto-Vorlage.
- 2. Ersetzen die erste Überschrift durch den Titel Quarto
 - Schreiben einen Text, der Quarto beschreibt, unter die Überschrift
- 3. Schreiben eine Unterüberschrift namens YAML
 - Schreiben einen Text, der die YAML-Struktur beschreibt, die wir besprochen haben
- 4. Erstellen eine Unterüberschrift mit dem Namen Quarto-Struktur.
 - Schreiben einige Notizen darüber, wie wir ein Quarto-Dokument strukturieren können (z.B. durch das Erstellen von Überschriften)
- 5. Finden Sie in RStudio die Schaltfläche Outline oben links im qmd Text Editor Fenster
 - Was sehent Sie, wenn Sie darauf klicken?

Textformatierung

• zum Formatieren von Text müssen wir die Markdown-Syntax verwenden

Format	Markdown				Ausgabe
Kursivschrift	Dieser T	ext	ist	*kursiv*	Dieser Text ist <i>kursiv</i>
Fett	Dieser T	ext	ist	**fett**	Dieser Text ist fett
Subskription	Dieser T	ext	ist	~tiefgestellt~	Dieser Text ist _{tiefgestellt}
Hochgestelt	Dieser T	ext	ist	^hochgestellt^	Dieser Text ist hochgestellt

Aufzählungen

- wir können Aufzählungslisten mit Bindestrichen erstellen.
 - Unteraufzählungen müssen eingerückt werden (drückt die Tabulatortaste)
- nummerierte Listen k\u00f6nnen durch einfaches Schreiben einer nummerierten Liste erstellt werden
 - Unteraufzählungen müssen in nummerierten Listen doppelt eingerückt werden

Aufzählungen

```
dies ist ein Aufzählungszeichen
+ dies ist ein Unterpunkt
1. Dies ist ein nummerierter Punkt
a. dies ist ein unternummerierter Punkt (beachtt den doppelten Einzug)
2. dies ist der zweite nummerierte Punkt
```

- dies ist ein Aufzählungszeichen
 - dies ist ein Unterpunkt
- 1. Dies ist ein nummerierter Punkt
 - a. dies ist ein unternummerierter Punkt (beachtt den doppelten Einzug)
- 2. dies ist der zweite nummerierte Punkt

Aufzählungen



Aufgabe 5: Aufzählungen

Beispiel 5

- 1. Fügen Ihrem qmd Dokumententext eine Textformatierung hinzu.
- 2. Fügen eine Aufzählungsliste hinzu
- 3. Fügen eine nummerierte Liste hinzu
- 4. Rendern Sie das Dokument. Hat es geklappt?

Codierung in Quarto

- Der große Vorteil von dynamischen Berichten ist die Integration von Text und Code
- Vorletzte Woche haben wir gelernt, wie man einfache mathematische Berechnungen in R durchführt.
- wie würden wir R-Befehle in ein qmd-Dokument einfügen?
 - Inline-Code (Code, der innerhalb einer Textzeile ausgeführt wird)
 - Code-Chunke (ein Code-Chunk, der nicht in Text enthalten ist)

Code-Chunks

- Code Chunks sind zwischen ```{r} und ``` eingebettet.
- eine schöne Tastenkombination: Cmd-Option-I (Mac) oder Strg-Alt-I (PC)

```
1 ```{r}
2 #/ eval: false
3
4 # Addition
5 4+6
6 ```
```

- ihr könnt den Code in Ihrer RStudio-Sitzung ausführen, indem ihr:
 - auf das kleine grüne Dreieck oben rechts im Chunk klickt
 - die Tastenkombination Cmd/Strg-Enter verwendt, um eine einzelne Code-Zeile auszuführen (je nachdem, worauf der Cursor steht)
 - der Tastenkombination Cmd/Strg-Shift-Enter benutzt, um den gesamten Code-Chunk auszuführen (falls es mehrere Befehle innerhalb eines einzelnen Abschnitts gibt)

Code-Chunks



Aufgabe 6: Code-Chunks

Beispiel 6

- 1. Füge einen Code Chunk zu deiner qmd Datei hinzu
 - Füge einige mathematische Operationen ein (Addition, Subtraktion, etc)
 - Fügt informative Anmerkungen zu Ihrem Code hinzu (z.B. # Addition)
- 2. Füge einen Text unter deinem Code-Chunk hinzu, der beschreibt, was der obige Code erreicht hat.
- 3. Rendern Sie das Dokument. Hat es geklappt?

Code-Chunks

(i) Erinnerung! Überschriften und Code-Anmerkungen

Denken Sie beim Schreiben von Notizen/bei der Bearbeitung von Übungen im Unterricht daran, informative Überschriften/Unterüberschriften zu erstellen! Auf diese Weise wird das Dokument strukturiert und übersichtlich, wenn ihr-in-der-Zukunft (oder ich) darauf zurückblickt.

Überschriften/Zwischenüberschriften strukturieren das gesamte Dokument. Code-Anmerkungen beschreiben, was bestimmte Teile des Codes bewirken (und warum). Beide beginnen mit einem Hashtag + Leerzeichen (#), aber Überschriften stehen außerhalb eines Codeabschnitts, während Codeanmerkungen innerhalb eines Codeabschnitts erscheinen.

Tipp: Klicken Sie auf die Schaltfläche "Outline" oben rechts im Texteditor-Fenster. Was zeigt sie an?

Code-Chunk-Optionen

- wir können die Ausführung von Code-Chunken steuern
- wir wollen nicht immer unseren Code in einem Bericht wiederholen
 - wir können dies in jedem Code-Chunk mit # | echo: true oder false steuern
- wir wollen nicht immer unseren Code in einem Bericht ausführen lassen
 - wir können dies in jedem Code-Chunk mit # | eval: true oder false steuern

Code-Chunk-Optionen

• Dies würde wie folgt aussehen:

```
1 ```{r}
2 #/ eval: true
3
4 # Addition
5 4+6
6 ```
[1] 10
```

- Wichtig ist, dass die Codechunk-Optionen:
 - mit # | beginnen, mit einem Leerzeichen dahinter und keinem Leerzeichen davor
 - direkt unter ```{r} platziert werden

Code-Chunk-Optionen

```
Q Aufgabe 7: c()
```

Beispiel 7

- 1. Erinnern Sie sich, dass wir letzte Woche die Funktion c() (EN: concatenate) gesehen haben, die mehrere Werte kombiniert (z.B. mean (c(3,4,25)) ergibt den Mittelwert von 3,4 und 25)
- 2. In einem Code-Stück: Erstellen sie ein Objekt, das eine Liste von Zahlen enthält (z.B. Objektname <- c(...))
- 3. Berechnen Sie den Mittelwert dieser Zahlen, indem Sie nur den Objektnamen verwendt.
- 4. Speichern Sie den Mittelwert dieser Zahlen als ein Objekt
- 5. Rendern Sie das Dokument und seht sich den Abschnitt mit Ihrem Code-Chunk an.
 - Ändern Sie nun im Quellcode die Chunk-Einstellungen auf echo: false und rendern das Dokument. Was ändert sich?
 - Setzen nun echo: true, aber eval: false. Rendern das Dokument. Was ändert sich?

- Ein großer Vorteil der gerenderten Quarto-Dokumente besteht darin, dass wir unsere Abbildungen zusammen mit den Textbeschreibungen anzeigen können
- Lassen Sie uns versuchen, eine Handlung von letzter Woche in unserem neuen Quarto-Dokument zu reproduzieren

Set-up

• unsere Pakete in einen Codechunk laden: tidyverse, languageR, und ggthemes

```
1 ```{r}
2 # Pakete laden
3 library(tidyverse)
4 library(languageR)
5 library(ggthemes)
6 ```
```

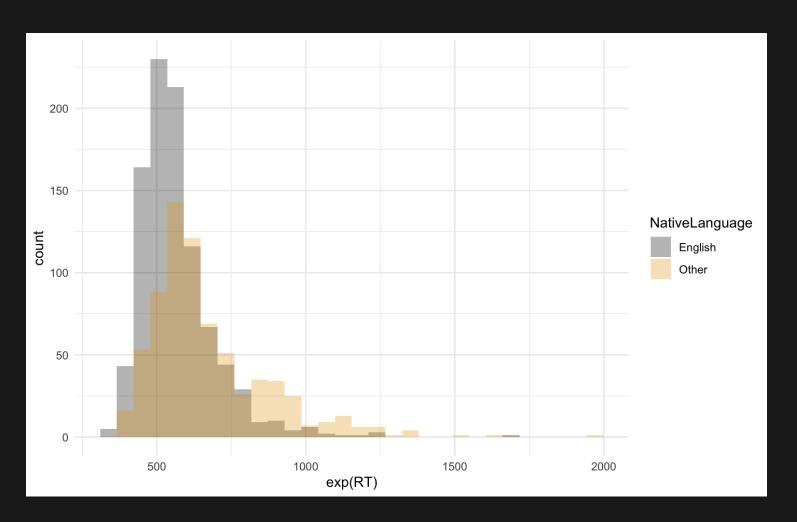
 unsere Daten in einen separaten Codechunk laden (am besten ist es, einen einzigen Codechunk für einen einzigen Zweck zu verwenden)

```
1 ```{r}
2 # Daten laden
3 df_lexdec <- lexdec
4 ```</pre>
```

- Erstellen Sie jetzt einfach einen neuen Codechunk, der einen Code von letzter Woche enthält
- wir speichern es als Objekt mit dem Namen fig_lexdec_hist:

```
## histogram of reaction times by native language
ggplot(data = df_lexdec) +
aes(x = exp(RT), fill = NativeLanguage) + ## set aest
geom_histogram(position = "identity", alpha = 0.3) +
scale_fill_colorblind() + ## make fill colorblind fri
theme_minimal() ## set plot theme
```

Abbildung 3: Histogram of reactiontimes per native language from lexdec



Plots drucken

- Erinnern Sie sich an die letzte Woche: Wenn Sie einen Plot benennen, wird er nur gedruckt, wenn Sie den Namen des Objekts eingeben
- wenn Sie den Plot nicht als Objekt speichern, wird er gedruckt, wenn Sie den Code ausführen, der den Plot erzeugt
- Wenn Sie den Plot als Objekt speichern, wird er nicht gedruckt, wenn Sie den Code ausführen.
 - In diesem Fall müssen Sie den Objektnamen ausführen, um zu sehen, was unter diesem Namen gespeichert ist
 - Dies gilt für alle Arten von Objekten, nicht nur für Diagramme!



Aufgabe 8: Plots in Quarto

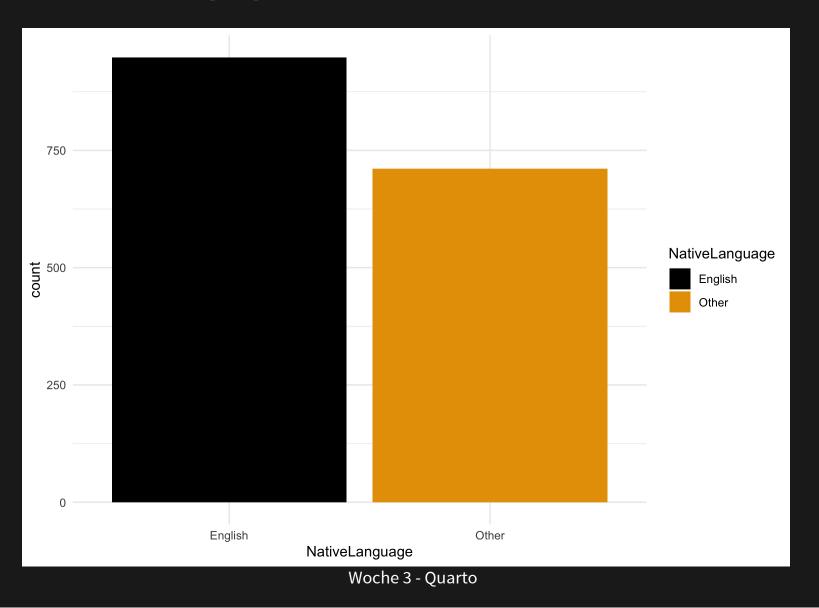
Beispiel 8

- 1. Einen neuen Codeabschnitt erstellen und das Balkendiagramm von letzter Woche erzeugen, aber als Objekt speichern
- 2. In einem separaten Codechunk nur den Objektnamen dieses Diagramms angeben
- 3. Rendern Sie das Dokument, um zu sehen, wo die Abbildung gedruckt wurde.

```
1 fig_lexdec_l1 <-
2     ggplot(data = df_lexdec) +
3     aes(x = NativeLanguage, fill = NativeLanguage) +
4     # add the geom:
5     geom_bar() +
6     scale_fill_colorblind() + # add colourblind colours
7     theme_minimal()

1 fig_lexdec_l1</pre>
```

Abbildung 4: Barplot of observations per native language



- es gibt mehrere Ausgabeformate, die wahrscheinlich nützlichsten sind:
 - html (default)
 - pdf
 - revealjs (Folien)
 - docx

- wenn wir das Dokument rendern:
 - 1. Quarto sendet die qmd-Datei an knitr (ein R-Paket für dynamische Berichte mit R)
 - 2. knitr führt die Code-Chunke aus und erstellt ein neues md Dokument mit Code und Ausgabe
 - 3. die md-Datei wird von pandoc verarbeitet, das md-Dateien in die fertige Datei konvertieren kann, mit vielen Ausgabeformaten

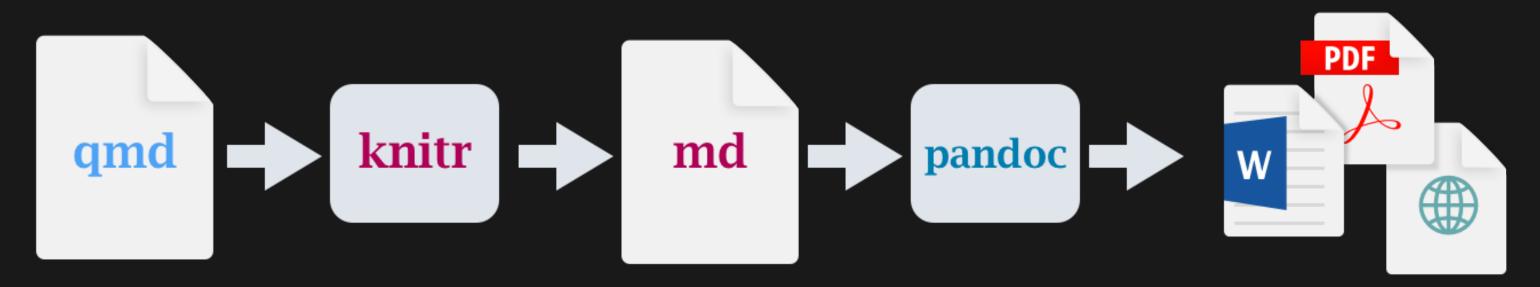


Diagramm des Quarto-Workflows von qmd, zu knitr, zu md, zu pandoc, zur Ausgabe im PDF-, MS Word- oder HTML-Format. (Quelle: Wickham et al. (2023))

02

(i) Andere Verwendungen

Quarto kann für eine Vielzahl von Zwecken verwendet werden, wie z. B.:

- Websites/Blogs
- Notizen machen
- Dokumentieren von allem, was mit Code zu tun hat, um die Reproduzierbarkeit zu verbessern
 - Tipps zum Arbeitsablauf
 - Bearbeitung von csv-Dateien (z. B. Stimuluslisten)



Aufgabe 9: Ausgabeformate

Beispiel 9

- 1. Ersetzt html in der YAML durch revealjs. Rendert das Dokument.
 - Schauen Sie den Ordner für die Notizen dieser Woche an. Welche Dateien seht?
- 2. Setzt nun format auf pdf. Rendert das Dokument.
 - Läuft es?
 - Versuche, pdf durch den Buchstaben lzu ersetzen. R schlägt eine Vervollständigung vor, welche ist es? Wähle sie aus und rendere das Dokument.
- 3. Setzt das Format wieder auf html. Rendert das Dokument.
- 4. Geht zurück zu Ihrem Ordner mit den Notizen dieser Woche. Welche Dateien seht?
 - Ist die Ausgabe von revealjs dort?

Heutige Ziele ****

Heute haben wir...

- gelernt, was dynamische Berichte sind 🔽
- unser eigenes Quarto-Dokument erstellt
- gelernt, wie man ein Quarto-Dokument bearbeitet 🔽
- gelernt, wie man Code in ein Quarto-Dokument einfügt 🔽
- ein Quarto-Dokument in verschiedenen Formaten wiedergebt 🔽

Extra: Reproduzierbarkeit in Quarto

- die Paketversionen mit sessionInfo() ausgeben
 - wenn ich ein neues Dokument beginne, ist eines der ersten Dinge, die ich tue, eine Kopfzeile # Session Info am unteren Ende hinzuzufügen, mit dem folgenden:

1 sessionInfo()



Beispiel 10

• fügt eine "Session Info" Abschnitt am Ende des Dokuments hin

Session Info

Hergestellt mit R version 4.3.0 (2023-04-21) (Already Tomorrow) und RStudioversion 2023.3.0.386 (Cherry Blossom).

```
1 sessionInfo()
R version 4.3.0 (2023-04-21)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Ventura 13.2.1

Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib; LAPACK version 3.11.0

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/c/en_US.UTF-8/en_US.UTF-8
time zone: Europe/Berlin
tzcode source: internal
attached base packages:
```

Literaturverzeichnis

Nordmann, E., & DeBruine, L. (2022). *Applied Data Skills*. Zenodo. https://doi.org/10.5281/zenodo.6365078 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for Data Science* (2. Aufl.).