

# WDDA FS 2025: Leitfaden für Aufgabenserie 2

2025-03-09

## Einleitung

Dieser Leitfaden bietet detaillierte Erklärungen für die Übungen in WDDA FS 2025 Aufgabenserie 2. Für jede Aufgabe werden wir den Denkprozess (und die notwendigen Schritte in R) durchgehen und erklären, wie man zur richtigen Lösung gelangt.

## Aufgabe 1: Häufigkeitsverteilung von transport

Die Variable `transport` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Ermitteln Sie die Häufigkeitsverteilung von `transport`.

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden:

Laden Sie den Datensatz aus der Excel-Datei in R mit der `readxl`-Bibliothek.

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")
```

#### 2. Variable extrahieren:

Wählen Sie die Variable `transport` aus dem Datensatz aus.

#### 3. Häufigkeitsverteilung berechnen:

Verwenden Sie die Funktion `table()`, um die Häufigkeiten der einzelnen Kategorien zu ermitteln.

```
data$transport |> table() |> print()
```

```
##
## Bicycle      Bus      Car      Other      Train      Walk
##           7       35       22         2       149         2
```

#### 4. Ergebnisvergleich:

Vergleichen Sie Ihre Resultate mit den Lösungsvorgaben:

- Bicycle: 7
- Bus: 35
- Car: 22

- Other: 2
- Train: 149
- Walk: 2

## Aufgabe 2: Balkendiagramme für gender, transport und maths

Die Variablen `gender`, `transport` und `maths` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Erstellen Sie geeignete Balkendiagramme, um die Verteilung der Daten für (a) `gender`, (b) `transport` und (c) `maths` darzustellen.

### Schritt-für-Schritt-Erklärung:

#### 1. Daten vorbereiten:

Laden Sie den Datensatz aus der Excel-Datei und extrahieren Sie die Variablen `gender`, `transport` und `maths`.

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")
```

#### 2. Häufigkeiten berechnen:

Nutzen Sie `table()` für jede Variable.

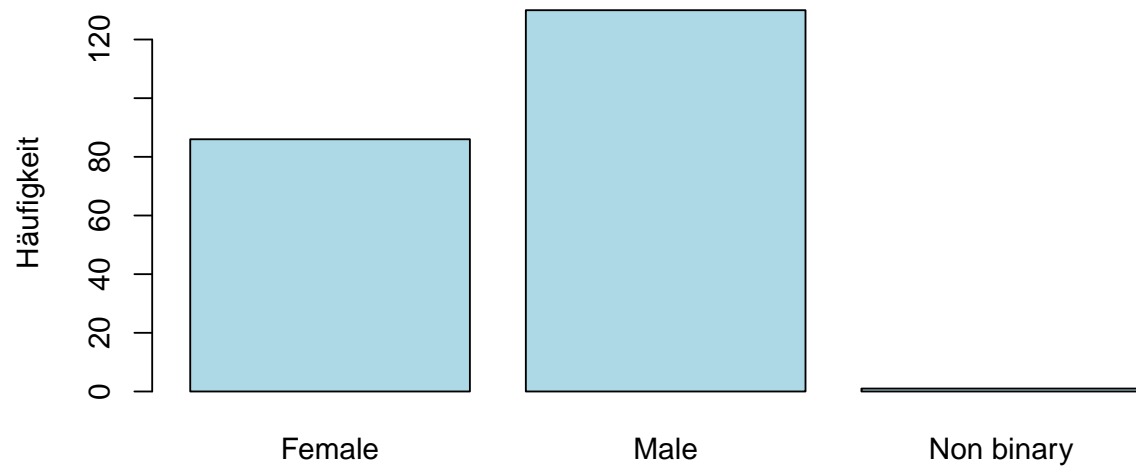
```
gender_freq <- data$gender |> table()
transport_freq <- data$transport |> table()
maths_freq <- data$maths |> table()
```

#### 3. Balkendiagramme erstellen:

Nutzen Sie `barplot()`, um die Häufigkeiten grafisch darzustellen.

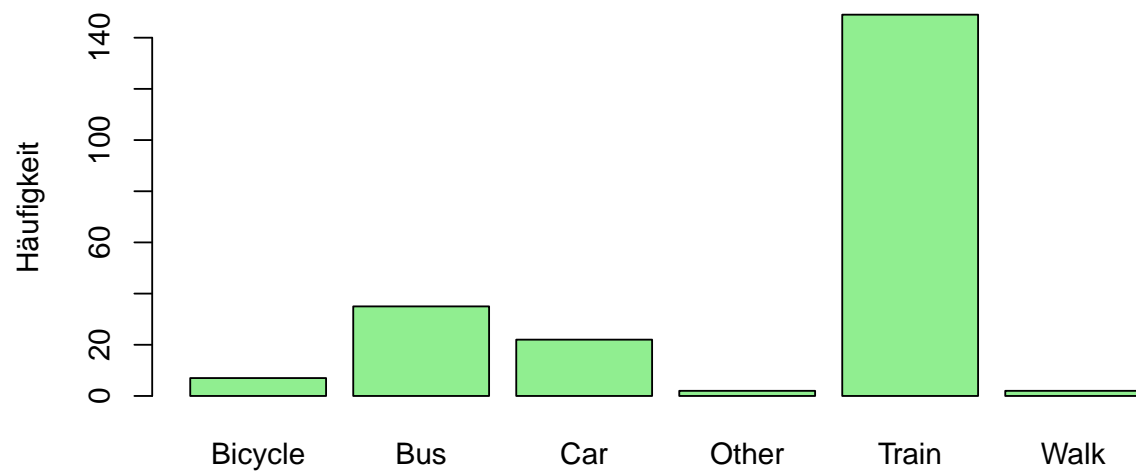
```
# Balkendiagramm für gender
barplot(gender_freq,
  main = "Verteilung von Gender",
  col = "lightblue",
  ylab = "Häufigkeit")
```

**Verteilung von Gender**

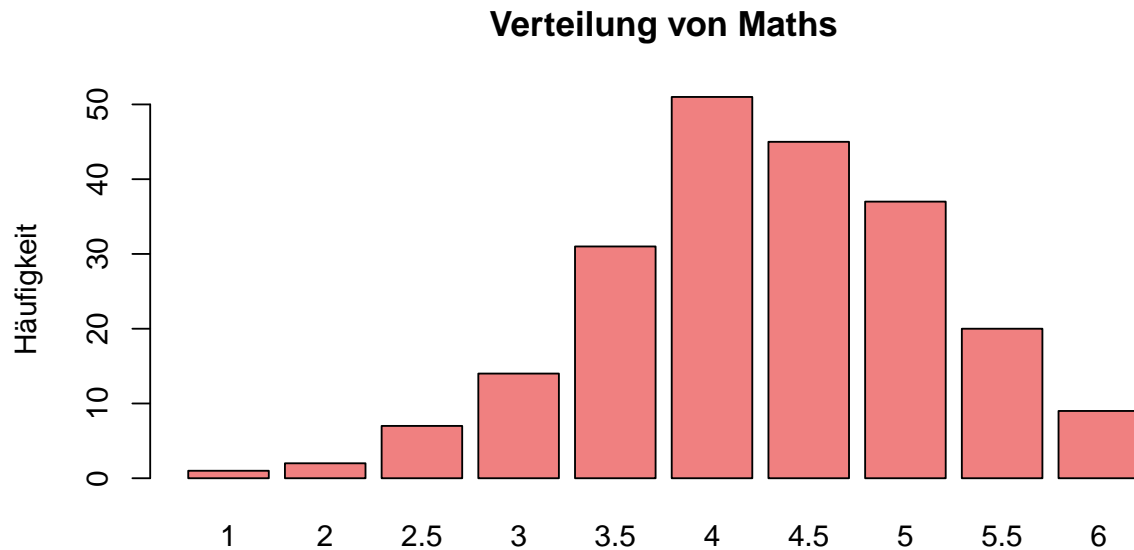


```
# Balkendiagramm für transport
barplot(transport_freq,
  main = "Verteilung von Transport",
  col = "lightgreen",
  ylab = "Häufigkeit")
```

**Verteilung von Transport**



```
# Balkendiagramm für maths
barplot(maths_freq,
        main = "Verteilung von Maths",
        col = "lightcoral",
        ylab = "Häufigkeit")
```



#### 4. Interpretation:

Erläutern Sie, was die Diagramme aussagen und ob die Verteilungen erwartungsgemäss erscheinen.

### Aufgabe 3: Analyse der Diagrammdarstellung der Semester-Noten

#### Aufgabenstellung:

- Erklären Sie, warum das gezeigte Diagramm (Screenshot) eine schlechte Grafik ist.
- Skizzieren Sie, wie das Diagramm stattdessen aussehen soll.

#### Schritt-für-Schritt-Erklärung:

##### 1. Problemerkennung:

- Das Diagramm verwendet eine Darstellungsform (z. B. Linien- oder Punktdiagramm), die für ordinale Daten wie Noten ungeeignet ist.
- Die Ordinalskala der Noten wird nicht adäquat berücksichtigt, was zu Missverständnissen führen kann.

##### 2. Verbesserungsvorschlag:

- Verwenden Sie ein Balkendiagramm, das die Häufigkeiten der einzelnen Noten klar darstellt.

- Alternativ kann ein Boxplot zur Veranschaulichung der Verteilung genutzt werden, wobei bei ordinalen Daten ein Balkendiagramm meist besser geeignet ist.

### 3. Skizze:

Skizzieren Sie ein Diagramm, in dem:

- Die x-Achse die Noten (ordinal) anzeigt.
- Die y-Achse die Anzahl der Schüler pro Note darstellt.
- Klare und nicht verzerrte Balken verwendet werden.

## Aufgabe 4: Kreisdiagramme für eye und eyetext

Die Variablen `eye` und `eyetext` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Erstellen Sie geeignete Kreisdiagramme, um die relative Verteilung der Daten von `eye` und `eyetext` darzustellen.

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und relative Häufigkeiten berechnen:

Laden Sie die Daten und erstellen Sie eine Häufigkeitstabelle, die Sie in relative Häufigkeiten umwandeln.

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

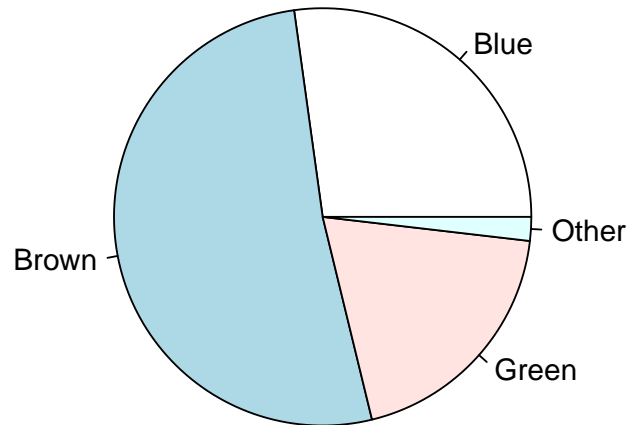
eye_rel <- data$eye |> table() |> prop.table()
```

#### 2. Kreisdiagramm erstellen:

Nutzen Sie `pie()`, um die relativen Anteile darzustellen.

```
pie(eye_rel, main = "Relative Verteilung von Eye")
```

## Relative Verteilung von Eye



### 3. Wiederholung für **eyetext**:

Berechnen Sie die Häufigkeiten für **eyetext** und erstellen Sie ein entsprechendes Kreisdiagramm.

## Aufgabe 5: Problematische Darstellung des Marktanteils bei Steve Jobs

### Aufgabenstellung:

Das Bild zeigt Steve Jobs, der in einer Keynote den Marktanteil seiner Firma Apple präsentiert. Erklären Sie, was das Problem mit der Darstellung ist, und geben Sie eine korrigierte Version an.

### Schritt-für-Schritt-Erklärung:

#### 1. Problemerkennung:

- Die Darstellung leidet unter perspektivischen Verzerrungen: Einige Segmente (z. B. die grüne Scheibe) wirken optisch grösser oder kleiner als ihre tatsächlichen Prozentwerte.
- Dies führt zu Fehlinterpretationen der realen Anteile.

#### 2. Korrigierte Darstellung:

- Nutzen Sie ein einfaches Kreisdiagramm ohne perspektivische Effekte, das die genauen Prozentwerte wiedergibt.

- Alternativ ist ein Balkendiagramm vorzuziehen, bei dem die Längen der Balken direkt die Prozentzahlen widerspiegeln.

### 3. Zusammenfassung:

Die korrigierte Darstellung muss die tatsächlichen Verhältnisse (z. B. 19.5% versus 21.2%) ohne optische Täuschung klar erkennbar machen.

## Aufgabe 6: Häufigkeitsverteilung von height in Bins

### Aufgabenstellung:

Finden Sie die Häufigkeitsverteilung für die Daten `height` mit Bins der Breite 5cm, von 155cm bis 185cm.

### Schritt-für-Schritt-Erklärung:

#### 1. Bins erstellen:

Erzeugen Sie mit `seq()` eine Sequenz von Grenzwerten.

```
bins <- seq(155, 185, by = 5)
```

#### 2. Daten in Intervalle einteilen:

Nutzen Sie `cut()` zur Unterteilung der `height`-Daten.

```
height_bins <- data$height |> cut(breaks = bins, right = TRUE)
```

#### 3. Häufigkeiten berechnen:

Wenden Sie `table()` auf die in Intervalle eingeteilten Daten an.

```
height_freq <- table(height_bins)
print(height_freq)
```

```
## height_bins
## (155,160] (160,165] (165,170] (170,175] (175,180] (180,185]
##          11          22          36          34          41          38
```

#### 4. Ergebnisvergleich:

Vergleichen Sie Ihre Häufigkeiten mit den vorgegebenen Werten.

## Aufgabe 7: Kumulative Häufigkeit für height

### Aufgabenstellung:

Ermitteln Sie die kumulative Häufigkeit für `height` mit denselben Bins wie in Aufgabe 6.

### Schritt-für-Schritt-Erklärung:

#### 1. Kumulative Häufigkeiten berechnen:

Verwenden Sie `cumsum()` auf der Häufigkeitstabelle.

```
height_cum <- height_freq |> cumsum()
print(height_cum)
```

```
## (155,160] (160,165] (165,170] (170,175] (175,180] (180,185]
##          11          33          69          103          144          182
```

## 2. Interpretation:

Erläutern Sie, wie die kumulative Häufigkeit angibt, wie viele Werte bis zu einem bestimmten Intervall erreicht werden.

## Aufgabe 8: Histogramme für foot und reaction1

Die Variablen `foot` und `reaction1` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Erstellen Sie geeignete Histogramme für die Variablen `foot` und `reaction1`.

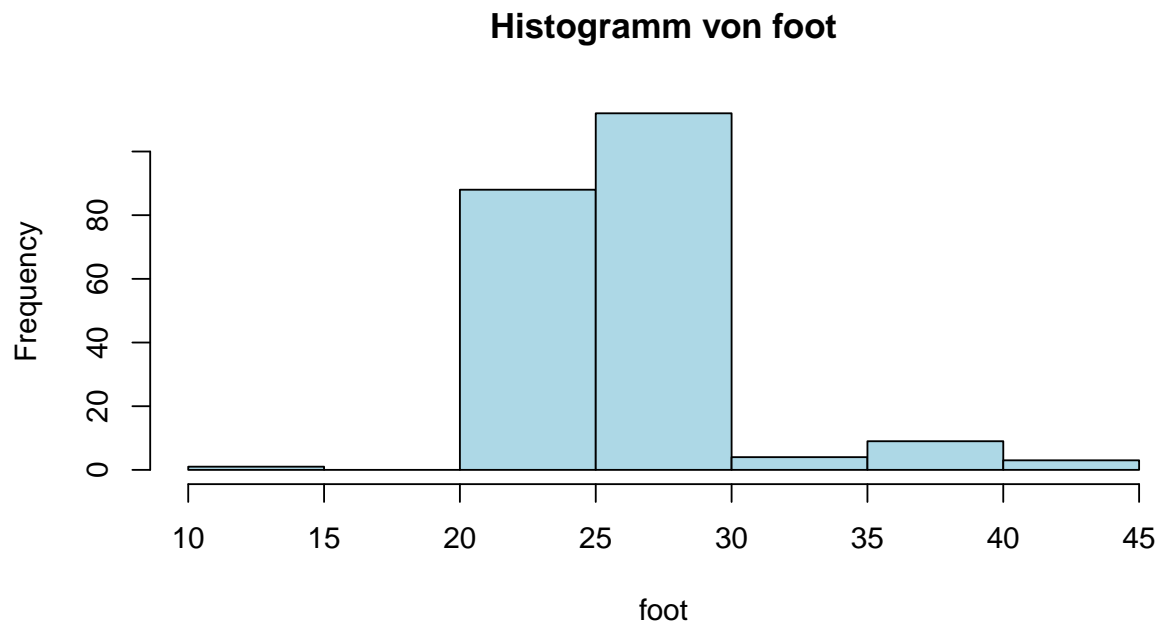
### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und Histogramm für foot erstellen:

Laden Sie die Daten und nutzen Sie die Funktion `hist()`, um die Verteilung darzustellen.

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

# Convert foot to numeric if it's not already
foot_numeric <- as.numeric(data$foot)
hist(foot_numeric, main = "Histogramm von foot", xlab = "foot", col = "lightblue")
```

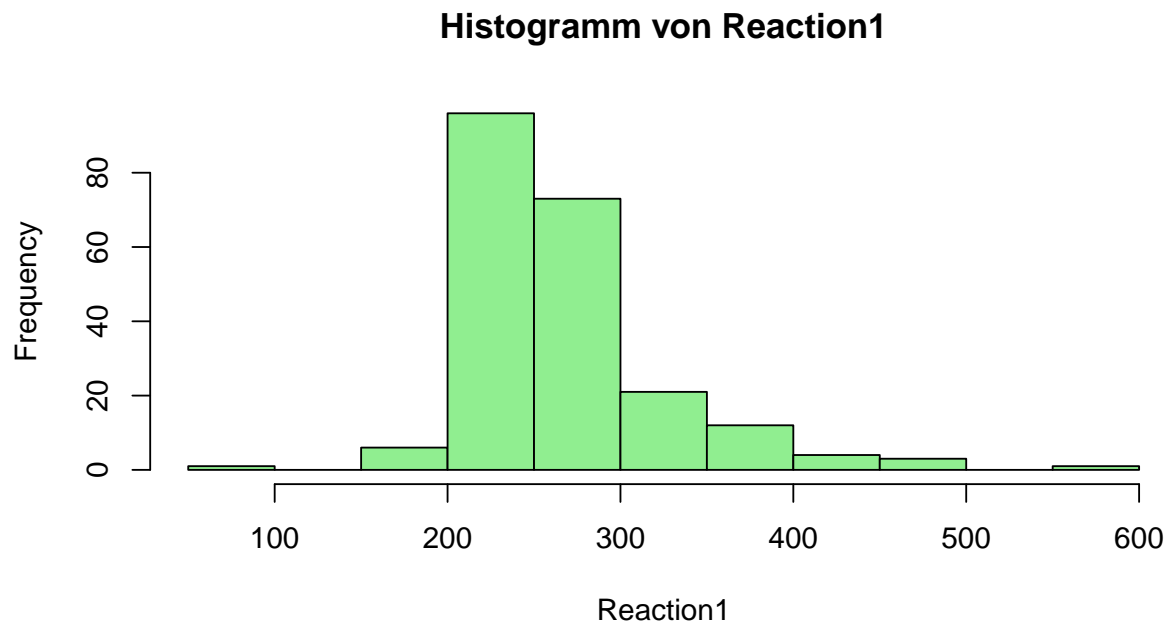




## 2. Histogramm für reaction1:

Erstellen Sie analog ein Histogramm für reaction1.

```
# Convert reaction1 to numeric if it's not already
reaction1_numeric <- as.numeric(data$reaction1)
hist(reaction1_numeric, main = "Histogramm von Reaction1", xlab = "Reaction1", col = "lightgreen")
```



## 3. Analyse:

Kommentieren Sie, ob die Verteilungen symmetrisch oder schief sind.

# Aufgabe 9: Histogramm für height und Geschlechtervergleich

Die Variablen **height** und **gender** finden Sie in WDDA\_02.xlsx im Tabellenblatt BFH.

### Aufgabenstellung:

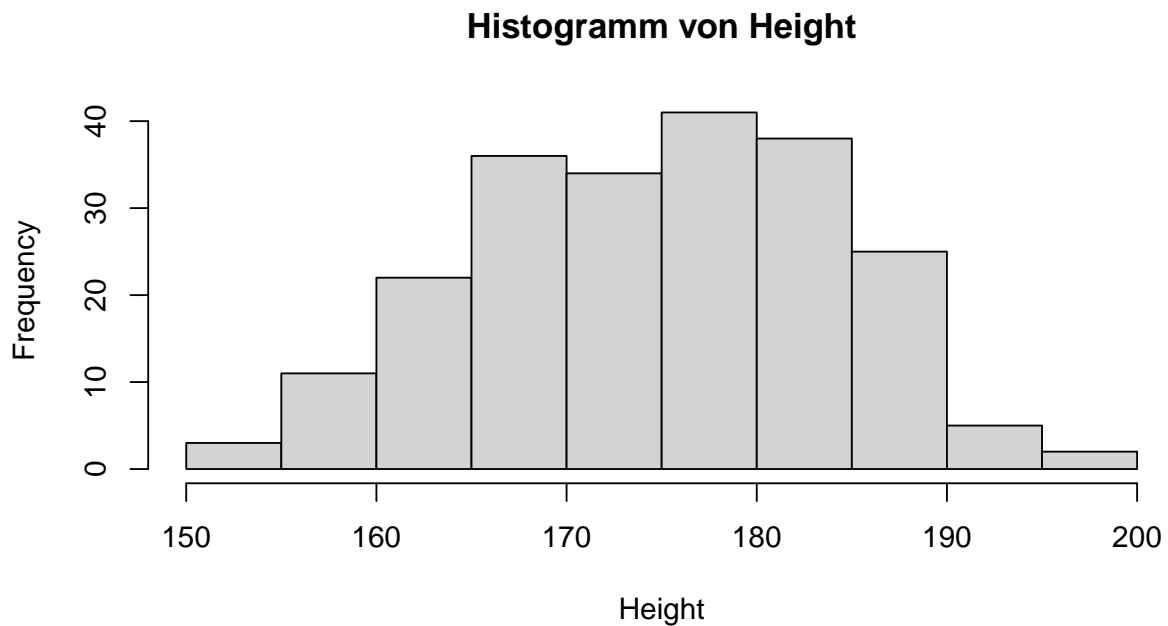
- (a) Zeichnen Sie ein Histogramm der Variable **height**.
- (b) Erstellen Sie zwei Unterlisten von **height** für Männer und Frauen und stellen Sie beide Histogramme auf der gleichen Skala dar.

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und Gesamthistogramm für height erstellen:

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

hist(data$height, main = "Histogramm von Height", xlab = "Height", col = "lightgrey")
```



## 2. Daten nach Geschlecht aufteilen:

Erzeugen Sie zwei Subsets für Männer und Frauen.

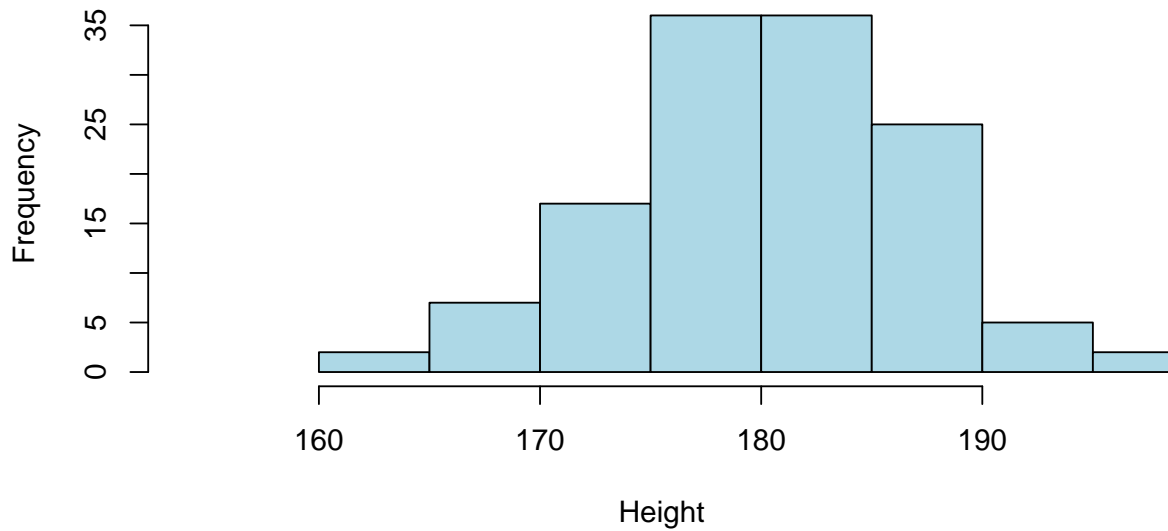
```
height_m <- data$height[data$gender == "Male"]  
height_f <- data$height[data$gender == "Female"]
```

## 3. Histogramme mit einheitlicher Skala:

Bestimmen Sie gemeinsame x- und y-Achsenlimits und plotten Sie beide Histogramme übereinander.

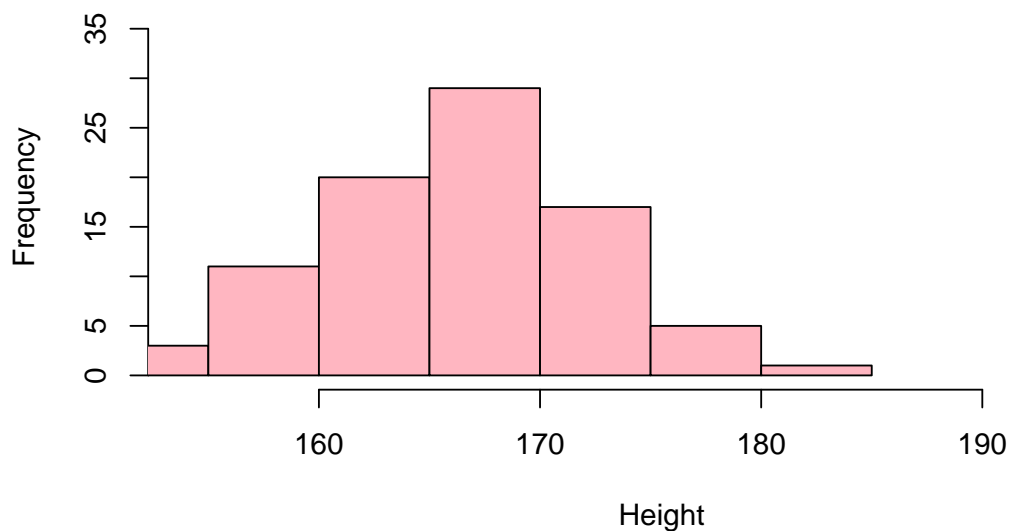
```
x_range <- range(data$height)  
y_max <- max(hist(height_m, plot = FALSE)$counts, hist(height_f, plot = FALSE)$counts)  
  
hist(height_m, main = "Histogramm von Height (Männer)", xlab = "Height",  
      xlim = x_range, ylim = c(0, y_max), col = "lightblue")
```

### Histogramm von Height (Männer)



```
hist(height_f, main = "Histogramm von Height (Frauen)", xlab = "Height",  
      xlim = x_range, ylim = c(0, y_max), col = "lightpink")
```

### Histogramm von Height (Frauen)



#### 4. Diskussion:

Vergleichen Sie die Verteilungen und diskutieren Sie, ob Unterschiede (z.B. Bimodalität) zwischen den Geschlechtern erkennbar sind.

## Aufgabe 10: Histogramme für hair und Geschlechtervergleich

Die Variablen `hair` und `gender` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

- (a) Zeichnen Sie ein passendes Histogramm für `hair`.
- (b) Erstellen Sie zwei Unterlisten von `hair` in Bezug auf das Geschlecht und stellen Sie die beiden Histogramme mit der gleichen Skala dar.

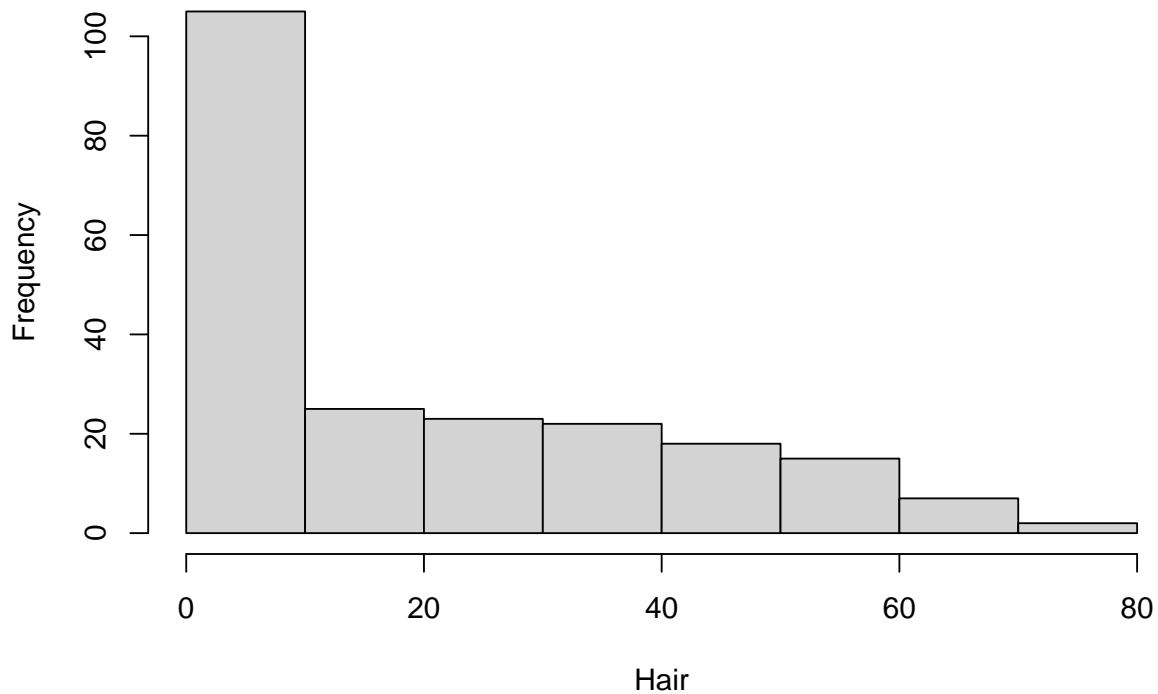
### Schritt-für-Schritt-Erklärung:

1. Daten laden und Histogramm für `hair` erstellen:

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

# Convert hair to numeric if needed
hair_numeric <- as.numeric(data$hair)
hist(hair_numeric, main = "Histogramm von Hair", xlab = "Hair", col = "lightgrey")
```

Histogramm von Hair



2. Aufteilung nach Geschlecht:

Erzeugen Sie Subsets für Männer und Frauen.

```
# Convert hair values by gender to numeric
hair_m <- as.numeric(data$hair[data$gender == "Male"])
hair_f <- as.numeric(data$hair[data$gender == "Female"])
```

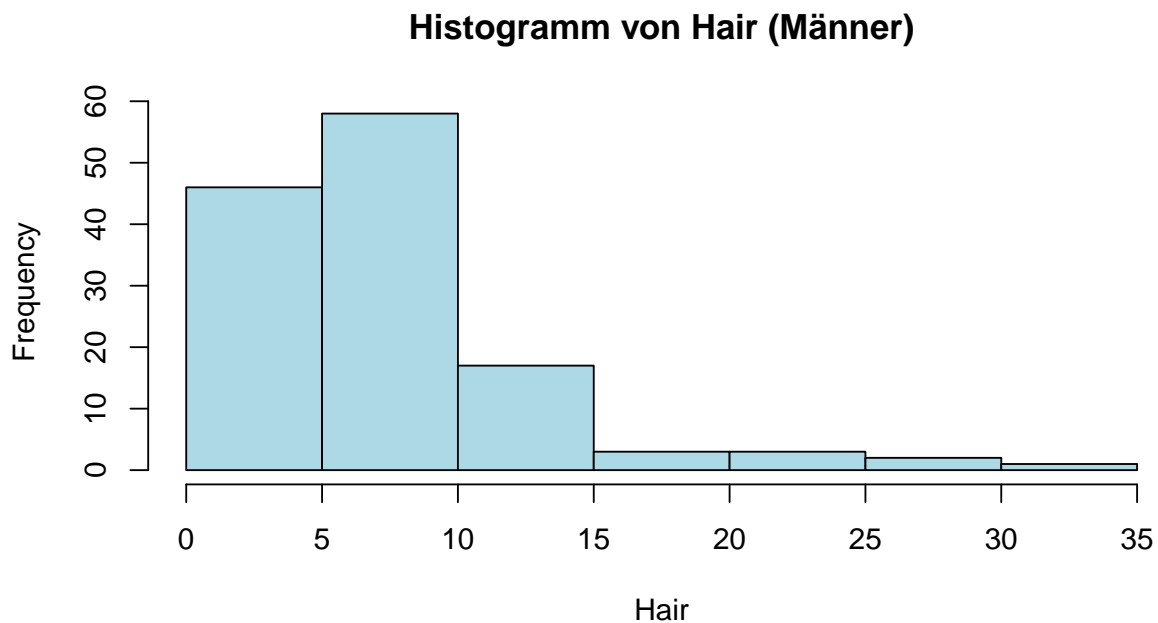
### 3. Histogramme mit einheitlicher Skala:

Bestimmen Sie gemeinsame Achsenlimits und plotten Sie die Histogramme.

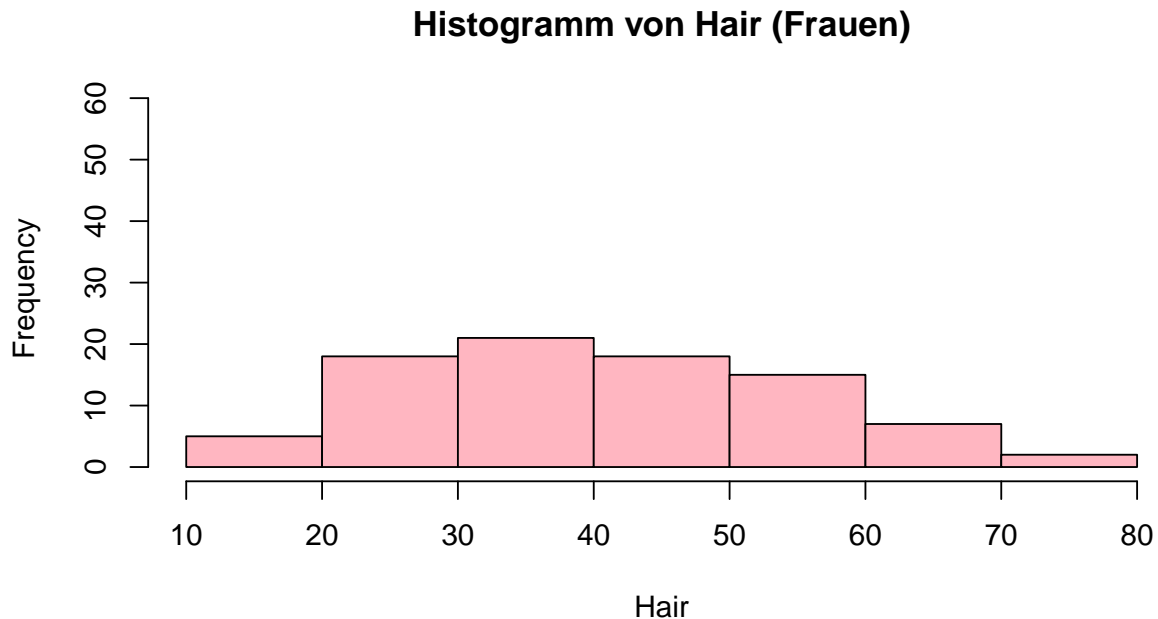
```
x_range_hair <- range(c(hair_m, hair_f), na.rm = TRUE)

# Create histograms with appropriate breaks
hist_m <- hist(hair_m, plot = FALSE)
hist_f <- hist(hair_f, plot = FALSE)
y_max_hair <- max(hist_m$counts, hist_f$counts)

hist(hair_m, main = "Histogramm von Hair (Männer)", xlab = "Hair",
     ylim = c(0, y_max_hair), col = "lightblue")
```



```
hist(hair_f, main = "Histogramm von Hair (Frauen)", xlab = "Hair",
     ylim = c(0, y_max_hair), col = "lightpink")
```



#### 4. Interpretation:

Diskutieren Sie, ob und welche Unterschiede in der Verteilung zwischen den Geschlechtern ersichtlich sind.

## Aufgabe 11: Histogramm für cash und Beurteilung der Schiefe

Die Variable `cash` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

#### Aufgabenstellung:

Erstellen Sie ein passendes Histogramm für `cash` und geben Sie an, ob die Daten schief sind.

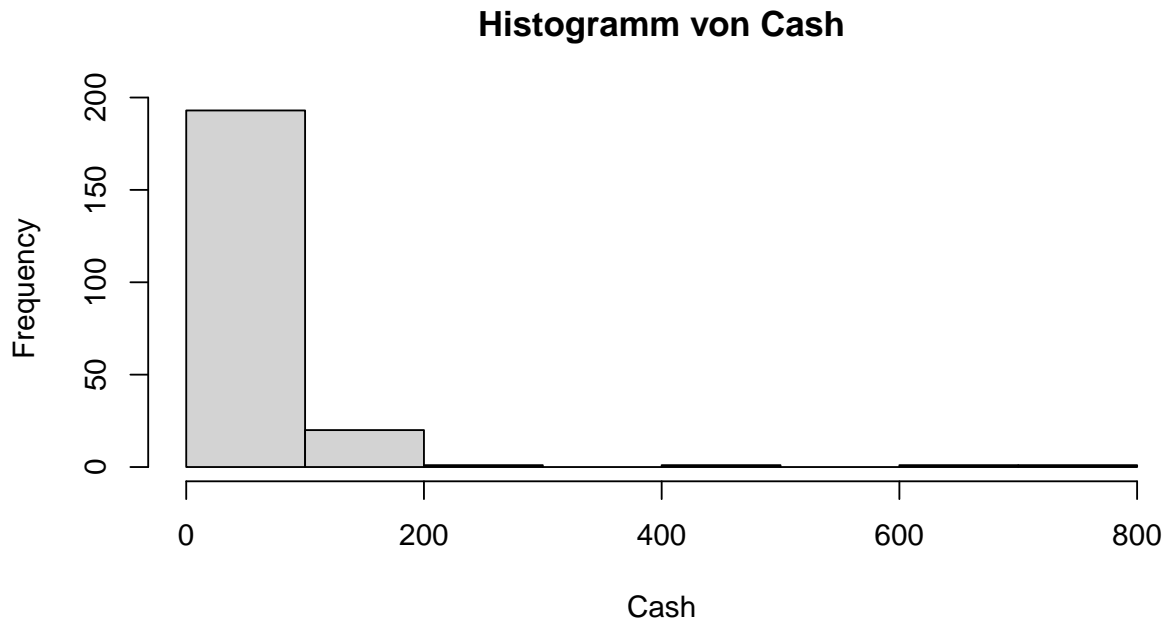
#### Schritt-für-Schritt-Erklärung:

##### 1. Daten laden und Histogramm erstellen:

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

# Convert cash to numeric if needed and handle NA values
cash_numeric <- as.numeric(data$cash)
cash_numeric <- cash_numeric[!is.na(cash_numeric)]

# Use breaks that work with the data range
hist(cash_numeric,
     main = "Histogramm von Cash",
     xlab = "Cash",
     col = "lightgrey",
     breaks = "Sturges") # Using Sturges algorithm for breaks
```



**2. Analyse der Schiefe:**

Beobachten Sie, ob die Verteilung einen langen rechten Schwanz (rechtsschief) aufweist.

Die Lösung weist darauf hin, dass extreme Werte vorhanden sind, welche die Verteilung asymmetrisch machen.

**3. Zusammenfassung:**

Erklären Sie, dass eine rechtsschiefe Verteilung charakterisiert ist durch wenige, aber sehr hohe Ausreisser.

## Aufgabe 12: Analyse der Körpergrösse: Hoehe.m und Hoehe.f

Die Variablen `height` und `gender` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

**Aufgabenstellung:**

Erstellen Sie `Hoehe.m` und `Hoehe.f` für die Grösse der Männer und Frauen. Wie viele Frauen sind grösser als der kleinste Mann?

**Schritt-für-Schritt-Erklärung:**

**1. Daten laden und nach Geschlecht aufteilen:**

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

Hoehe.m <- data$height[data$gender == "Male"]
Hoehe.f <- data$height[data$gender == "Female"]
```

**2. Ermittlung des kleinsten Mannes:**

```
min_m <- Hoehe.m |> min()
```

### 3. Zählen der Frauen, die grösser sind als der kleinste Mann:

```
count_f <- (Hoehe.f > min_m) |> sum()  
print(count_f)
```

```
## [1] 58
```

Es sollten 61 Frauen grösser sein als der kleinste Mann.

## Aufgabe 13: Fusslängenanalyse

Die Variablen `foot`, `height` und `gender` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Erstellen Sie `foot.m` und `foot.f` für die Fusslängen der Männer und Frauen.

Hat der grösste Mann auch die grössten Fusslaengen? Wie sieht es bei den Frauen aus?

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und Fusslängen nach Geschlecht aufteilen:

```
library(readxl)  
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")  
  
foot.m <- data$foot[data$gender == "Male"]  
foot.f <- data$foot[data$gender == "Female"]
```

#### 2. Vergleich:

- Ermitteln Sie die maximale Fusslänge bei Männern mit `max(foot.m)`.
- Bestimmen Sie anhand der Körpergrösse, ob der grösste Mann (mittels `max(Hoehe.m)`) auch die grösste Fusslänge aufweist.
- Wiederholen Sie den Vergleich für Frauen.

#### 3. Interpretation:

Die Lösung zeigt, dass beim grössten Mann die Fusslänge nicht maximal ist, während bei den Frauen der Zusammenhang anders sein kann.

## Aufgabe 14: Mindestgrösse für den Beitritt zur britischen Polizei

### Aufgabenstellung:

Wie viele Personen in den Daten erfüllen die Mindestgrösse (Männer  $\geq 173\text{cm}$ , Frauen  $\geq 163\text{cm}$ ) für die britische Polizei?

### Schritt-für-Schritt-Erklärung:



### 1. Filterung der Daten:

Selektieren Sie Personen, die die entsprechenden Mindestgrösse erfüllen.

```
suitable <- data |>
  subset((gender == "Male" & height >= 173) | (gender == "Female" & height >= 163)) |>
  {\\(df) df$height}()

count_suitable <- suitable |> length()
print(count_suitable)
```

```
## [1] 185
```

### 2. Ergebnis:

Die Lösung nennt 185 Personen.

## Aufgabe 15: Geburtsmonate mit lubridate

Die Variable dob (Geburtsdatum) finden Sie in WDDA\_02.xlsx im Tabellenblatt BFH.

### Aufgabenstellung:

Erstellen Sie die Liste month (mit der library lubridate), welche die Geburtsmonate der Personen enthält. Wie viele Personen wurden im April geboren?

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und Bibliotheken einbinden:

```
library(readxl)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")
```

#### 2. Extraktion der Monate:

Extrahieren Sie die Monate aus dem Geburtsdatum mit der month()-Funktion.

```
months <- data$dob |> month()
```

#### 3. Zählen der im April Geborenen:

Da month() für April den Wert 4 zurückgibt:

```
count_april <- (months == 4) |> sum()
print(count_april)
```

```
## [1] 24
```

Die Lösung gibt 24 Personen an.

## Aufgabe 16: Kontingenztabelle für gender und eye

Die Variablen `gender` und `eye` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Erstellen Sie eine Kontingenztabelle mit den Variablen `gender` und `eye`.

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und Tabelle erstellen:

Laden Sie die Daten und verwenden Sie `table()` zur Erstellung der Kontingenztabelle.

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

kontingenz <- data |> with(table(gender, eye))
print(kontingenz)
```

```
##           eye
## gender      Blue Brown Green Other
## Female      23    47    15     1
## Male        36    64    27     3
## Non binary   0     1     0     0
```

#### 2. Interpretation:

Vergleichen Sie Ihre Ergebnisse mit den Lösungsvorgaben und diskutieren Sie mögliche Zusammenhänge.

## Aufgabe 17: Streudiagramme und Analyse der Zusammenhänge

Die Variablen `height`, `hair`, `foot` und `reaction1` finden Sie in `WDDA_02.xlsx` im Tabellenblatt `BFH`.

### Aufgabenstellung:

Erstellen Sie ein Streudiagramm und kommentieren Sie die Art der Beziehung zwischen den folgenden Variablen:

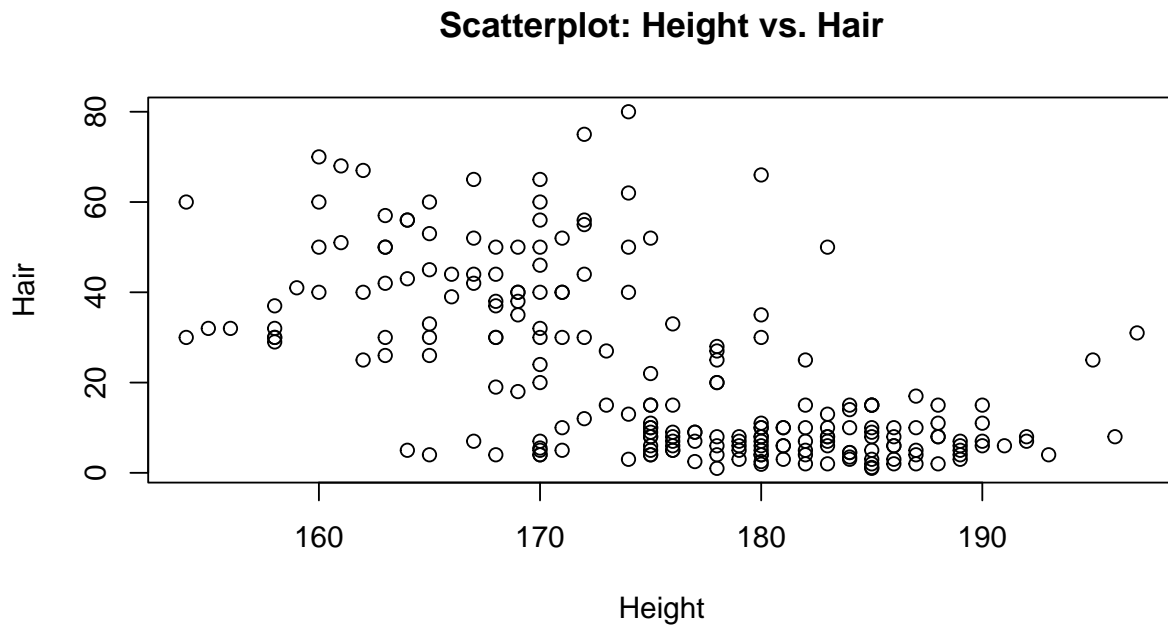
- (a) `height` und `hair`
- (b) `height` und `foot`
- (c) `height` und `reaction1`

### Schritt-für-Schritt-Erklärung:

#### 1. Daten laden und Streudiagramm für height und hair erstellen:

```
library(readxl)
data <- read_excel("../data/WDDA_02.xlsx", sheet = "BFH")

plot(data$height, data$hair, main = "Scatterplot: Height vs. Hair",
      xlab = "Height", ylab = "Hair")
```

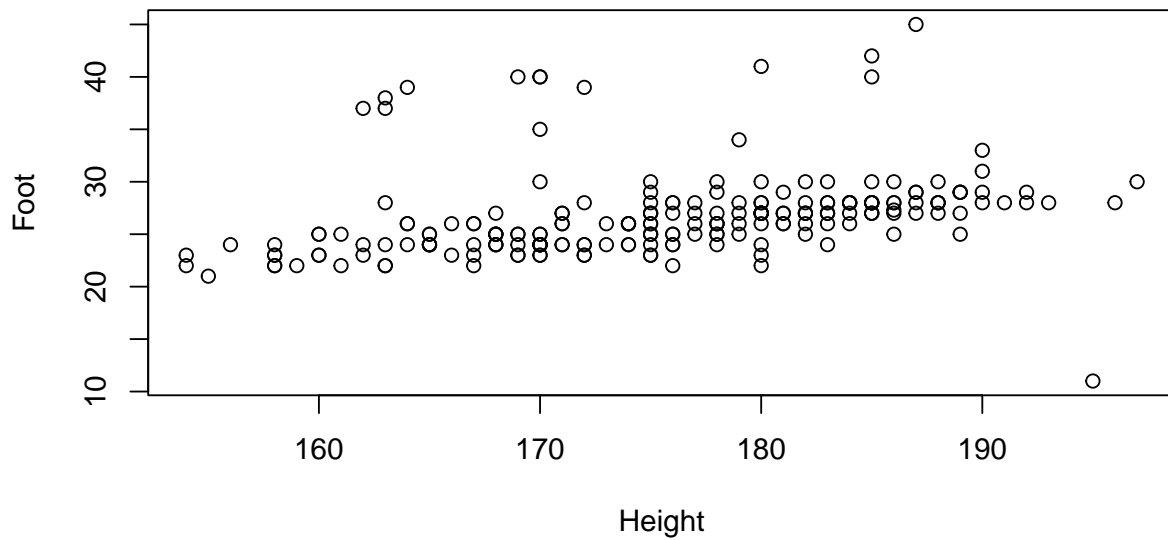


- Interpretieren Sie, ob ein negativer oder positiver Zusammenhang besteht.

## 2. Streudiagramm für height und foot:

```
plot(data$height, data$foot, main = "Scatterplot: Height vs. Foot",
      xlab = "Height", ylab = "Foot")
```

**Scatterplot: Height vs. Foot**

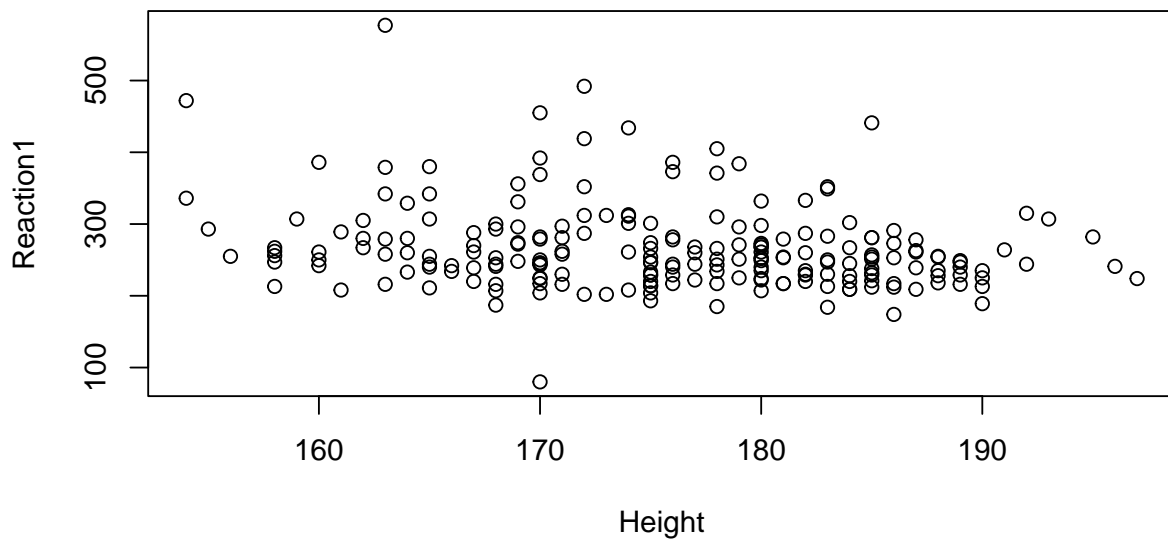


- Es wird ein positiver linearer Zusammenhang erwartet.

3. Streudiagramm für height und reaction1:

```
plot(data$height, data$reaction1, main = "Scatterplot: Height vs. Reaction1",  
      xlab = "Height", ylab = "Reaction1")
```

**Scatterplot: Height vs. Reaction1**



- Diskutieren Sie, dass hier der Zusammenhang weniger eindeutig ist.

#### 4. Kommentar:

Fassen Sie zusammen, welche Zusammenhänge bestehen und welche Interpretationen sich daraus ableiten lassen.

## Aufgabe 18: Analyse der Golfspieler-Umfrage

### Aufgabenstellung:

Analysieren Sie die Umfragedaten von Golfspielern und -spielerinnen hinsichtlich der Wahrnehmung, ob die Grün zu schnell sind.

- Kombinieren Sie die beiden Kontingenztabelle zu einer Gesamtübersicht (Zeilen: “männlich” und “weiblich”; Spalten: “zu schnell” und “ok”).
- Untersuchen Sie bei Spielern mit niedrigem Handicap, welche Gruppe den höheren Prozentsatz an Beschwerden aufweist.
- Untersuchen Sie bei Spielern mit höherem Handicap, welche Gruppe den höheren Prozentsatz an Beschwerden aufweist.
- Diskutieren Sie abschliessend, welche Rückschlüsse auf die Präferenzen in Bezug auf die Geschwindigkeit der Grün gezogen werden können und ob die aggregierten Ergebnisse konsistent mit den gruppenspezifischen Ergebnissen sind.

### Schritt-für-Schritt-Erklärung:

#### 1. Erstellung der Ausgangsdaten:

Erstellen Sie separate Matrizen für Männer und Frauen anhand der vorliegenden Umfragedaten.

```
# Daten für Männer:
# Zeilen: Handicap unter 15 und Handicap 15 und mehr
# Spalten: "Zu schnell" und "Ok"
golfer_maennlich <- matrix(c(10, 40, 25, 25), nrow = 2, byrow = TRUE)
colnames(golfer_maennlich) <- c("Zu schnell", "Ok")
rownames(golfer_maennlich) <- c("Handicap < 15", "Handicap  $\geq$  15")

# Daten für Frauen:
golfer_weiblich <- matrix(c(1, 9, 39, 51), nrow = 2, byrow = TRUE)
colnames(golfer_weiblich) <- c("Zu schnell", "Ok")
rownames(golfer_weiblich) <- c("Handicap < 15", "Handicap  $\geq$  15")
```

#### 2. Kombinierte Kontingenztabelle:

Fassen Sie die Daten in einer Gesamtübersicht zusammen, in der die Zeilen die Geschlechter und die Spalten die Bewertungen darstellen.

#### 3. Berechnung der Prozentsätze:

Ermitteln Sie die Prozentsätze innerhalb der jeweiligen Gruppen.

```
total_maennlich <- sum(golfer_maennlich)
total_weiblich <- sum(golfer_weiblich)
perc_maennlich <- round(golfer_maennlich / total_maennlich * 100, 1)
perc_weiblich <- round(golfer_weiblich / total_weiblich * 100, 1)
print(perc_maennlich)
```

```
##                Zu schnell Ok
## Handicap < 15          10 40
## Handicap  $\geq$  15      25 25
```

```
print(perc_weiblich)
```

```
##                Zu schnell Ok
## Handicap < 15          1  9
## Handicap  $\geq$  15      39 51
```

- Die Lösung zeigt, dass insgesamt Frauen mit ca. 40% einen höheren Anteil an der Aussage „zu schnell“ aufweisen als Männer (35%).
- Bei Spielern mit niedrigem Handicap liegt der höhere Beschwerdeanteil bei den Männern (20%), während bei den Spielern mit höherem Handicap ebenfalls die Männer (50%) den höheren Anteil haben.

#### 4. Diskussion der Inkonsistenzen:

- Diskutieren Sie, dass aggregierte Daten (aus Teil a) einen anderen Eindruck vermitteln können als die getrennte Betrachtung (Teile b und c).
- Erklären Sie, dass Unterschiede in der Handicapverteilung Einfluss auf die aggregierten Ergebnisse haben und warum dies zu scheinbaren Widersprüchen führen kann.