

Lösung: Häufigkeitsverteilung

① Häufigkeitsverteilung der Wartezeiten:

```
# Schritt 1: Spannweite bestimmen
waiting = faithful$waiting
range(waiting)

## [1] 43 96

# Schritt 2: Spannweite runden und aufteilen
breaks = seq(40, 100, by=5)
breaks

## [1] 40 45 50 55 60 65 70 75 80 85 90 95 100

# Schritt 3: Werte in die Intervalle aufteilen
waiting.cut = cut(waiting, breaks, right=FALSE)
```

Lösung: Häufigkeitsverteilung

```
# Schritt 4: Häufigkeiten bestimmen
```

```
waiting.freq = table(waiting.cut)
```

```
# Schritt 5: Ergebnis formatieren
```

```
head(cbind(waiting.freq), 7)
```

```
##           waiting.freq
```

```
## [40, 45)           1
```

```
## [45, 50)          20
```

```
## [50, 55)          32
```

```
## [55, 60)          24
```

```
## [60, 65)          17
```

```
## [65, 70)           9
```

```
## [70, 75)          23
```

Lösung: Häufigkeitsverteilung

- 2 Welches Intervall der Eruptionsdauer enthält die meisten Eruptionen?

```
# Schritt 1: Häufigkeitsverteilung von duration
duration = faithful$eruptions
breaks = seq(1.5, 5.5, by=0.5)
duration.cut = cut(duration, breaks, right=FALSE)
duration.freq = table(duration.cut)

# Schritt 2: Maximum in duration.freq finden
duration.freq.max = max(duration.freq)
duration.freq.max

## [1] 73
```

Lösung: Häufigkeitsverteilung

```
# Schritt 3: Welches Intervall enthält das Maximum?  
x = which(duration.freq == duration.freq.max)  
names(x)  
  
## [1] "[4,4.5)"
```

Das Intervall von 4 bis 4.5 Minuten enthält die meisten Eruptionen.

Lösung: Häufigkeitsverteilung

- 8 Bestimmen Sie die Häufigkeitsverteilung der Eruptionszeiten aus **faithful** mit der Funktion `hist`.

```
# Schritt 1: Histogramm bestimmen und Verteilung speichern
duration = faithful$eruptions
breaks = seq(1.5, 5.5, by=0.5)
h = hist(duration, breaks, right=FALSE, plot=FALSE)
duration.freq = h$counts
duration.freq

## [1] 51 41  5  7 30 73 61  4

# Anmerkung: breaks ist optional
```

Lösung: Häufigkeitsverteilung

```
# Schritt 2: Bezeichnungen für die Intervalle
len = length(h$breaks)
a = h$breaks[1:len-1] # rechte Seiten
b = h$breaks[2:len]   # linke Seiten
labels = paste("[", a, ", ", b, ")", sep="")
# Schritt 3: names von duration.freq bestimmen
# und formatieren
names(duration.freq) = labels
head(cbind(duration.freq), 3)

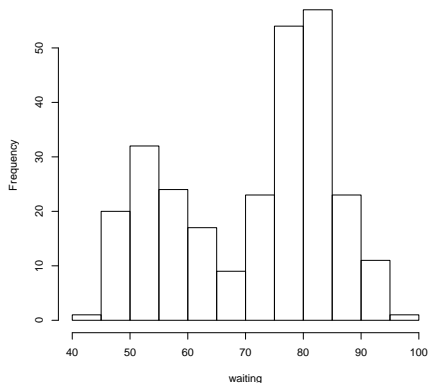
##           duration.freq
## [1.5, 2)             51
## [2, 2.5)             41
## [2.5, 3)              5
```

Lösung: Histogramm

```
waiting = faithful$waiting
```

```
hist(waiting, right=FALSE)
```

Histogram of waiting



Lösung: Relative Häufigkeitsverteilung stetiger Daten

```
waiting = faithful$waiting
breaks = seq(40, 100, by=5)
waiting.cut = cut(waiting, breaks, right=FALSE)
waiting.freq = table(waiting.cut)
waiting.relfreq = waiting.freq/nrow(faithful)
waiting.percentage = 100*waiting.relfreq
old = options(digits=1)
head(cbind(waiting.freq, waiting.percentage), 2)
```

```
##           waiting.freq waiting.percentage
## [40,45)              1              0.4
## [45,50)             20              7.4
```

```
options(old)
```


Lösung: Kumulierte Häufigkeitsverteilung

```
waiting = faithful$waiting
breaks = seq(40, 100, by=5)
waiting.cut = cut(waiting, breaks, right=FALSE)
waiting.freq = table(waiting.cut)
waiting.cumfreq = cumsum(waiting.freq)
head(cbind(waiting.cumfreq), 4)
```

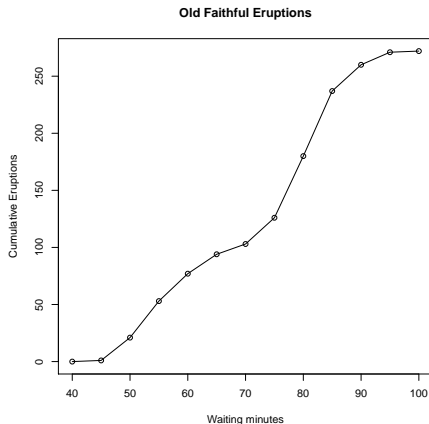
##	waiting.cumfreq
## [40, 45)	1
## [45, 50)	21
## [50, 55)	53
## [55, 60)	77

Lösung: Kumulierte Häufigkeitsverteilungskurve

```
waiting = faithful$waiting
breaks = seq(40, 100, by=5)
waiting.cut = cut(waiting, breaks, right=FALSE)
waiting.freq = table(waiting.cut)
waiting.cumfreq0 = c(0, cumsum(waiting.freq))
```

Lösung: Kumulierte Häufigkeitsverteilungskurve

```
plot(breaks, waiting.cumfreq0, main="Old Faithful Eruptions",  
     xlab="Waiting minutes",   ylab="Cumulative Eruptions")  
lines(breaks, waiting.cumfreq0)
```



Lösung: Kumulierte relative Häufigkeitsverteilung

```
waiting = faithful$waiting
breaks = seq(40, 100, by=5)
waiting.cut = cut(waiting, breaks, right=FALSE)
waiting.freq = table(waiting.cut)
waiting.cumfreq = cumsum(waiting.freq)
waiting.cumrelfreq = waiting.cumfreq/nrow(faithful)
waiting.cumpercent = 100 * waiting.cumrelfreq
old = options(digits=3)
head(cbind(waiting.cumfreq, waiting.cumpercent), 2)

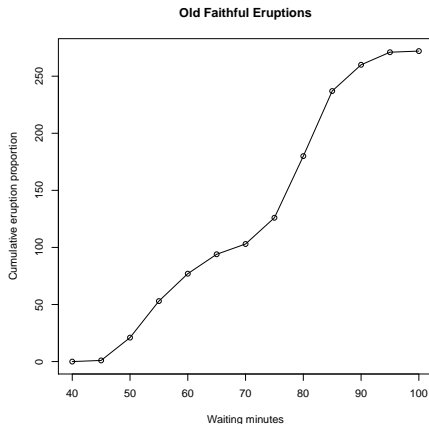
##           waiting.cumfreq waiting.cumpercent
## [40,45)                1             0.368
## [45,50)               21             7.721
```

Lösung: Kumulierte relative Häufigkeitskurve

```
waiting = faithful$waiting
breaks = seq(40, 100, by=5)
waiting.cut = cut(waiting, breaks, right=FALSE)
waiting.freq = table(waiting.cut)
waiting.cumfreq = cumsum(waiting.freq)
waiting.cumrelfreq = waiting.cumfreq/nrow(faithful)
waiting.cumrelfreq0 = c(0, waiting.cumfreq)
```

Lösung: Kumulierte relative Häufigkeitskurve

```
plot(breaks, waiting.cumrelfreq0, main="Old Faithful Eruptions",  
     xlab="Waiting minutes", ylab="Cumulative eruption proportion")  
lines(breaks, waiting.cumrelfreq0)
```



Lösung: Histogramm

```
# Dateipfad bitte anpassen  
load("/Users/michel/Documents/RScripts/Daten_Wachstum.RData")  
# range(Daten_Wachstum$Alter)  
breiten<-seq(15, 51, by=3)  
hist(Daten_Wachstum$Alter, breiten, right=FALSE)
```

