

Tutorium 1

Aufgabe 1

Speichern und einlesen der Daten

```
# Einlesen der Daten  
getwd()
```

```
[1] "C:/Users/marcp/Desktop/Uni/Arbeit/Tut Angewandte Statistik/Tut/Data"
```

```
titanic <- read.csv("titanic.csv")  
attach(titanic)
```

Aufgabe 2

Haben mehr Frauen oder mehr Männer überlebt?

```
# Tabelle mit Rohdaten  
gender_table <- table(Sex, Survived)  
print(gender_table)
```

```
      Survived  
Sex      0    1  
female  81 233  
male   468 109
```

```
# Tabelle nur überlebt
survived_table <- table(Sex[Survived == 1])
print(survived_table)
```

```
female  male
    233   109
```

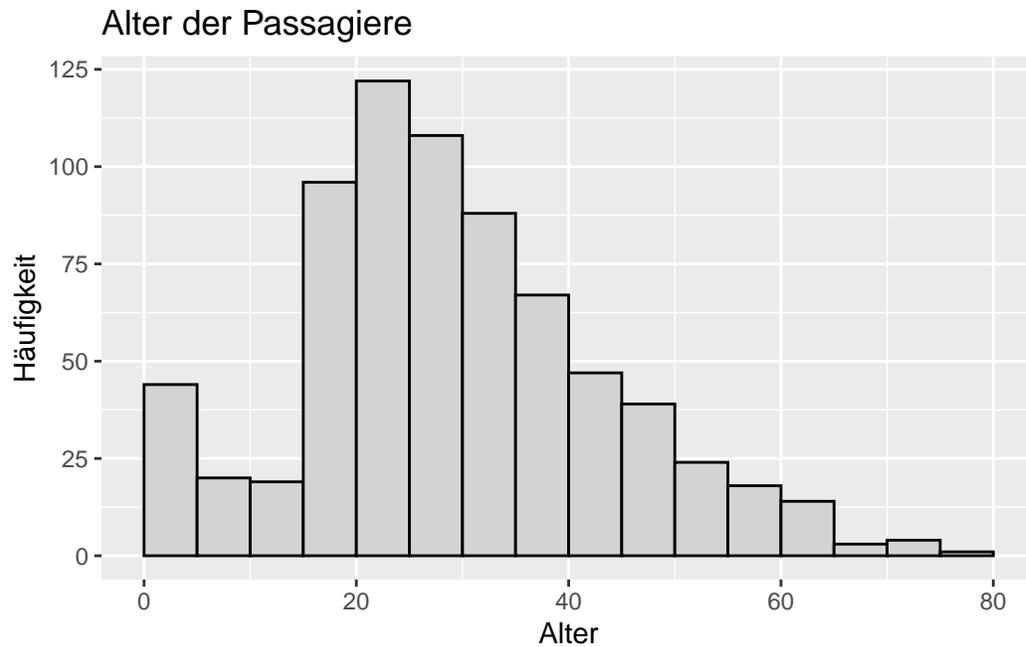
Es haben 233 Frauen und 109 Männer überlebt

Aufgabe 3

Erstellen Sie vier Histogramme mit `ggplot2::geom_histogram` zu dem Alter der Passagiere. Ein Histogramm mit allen Passagieren und drei weitere für jeweils eine Klasse. Welche Unterschiede fallen Ihnen auf?

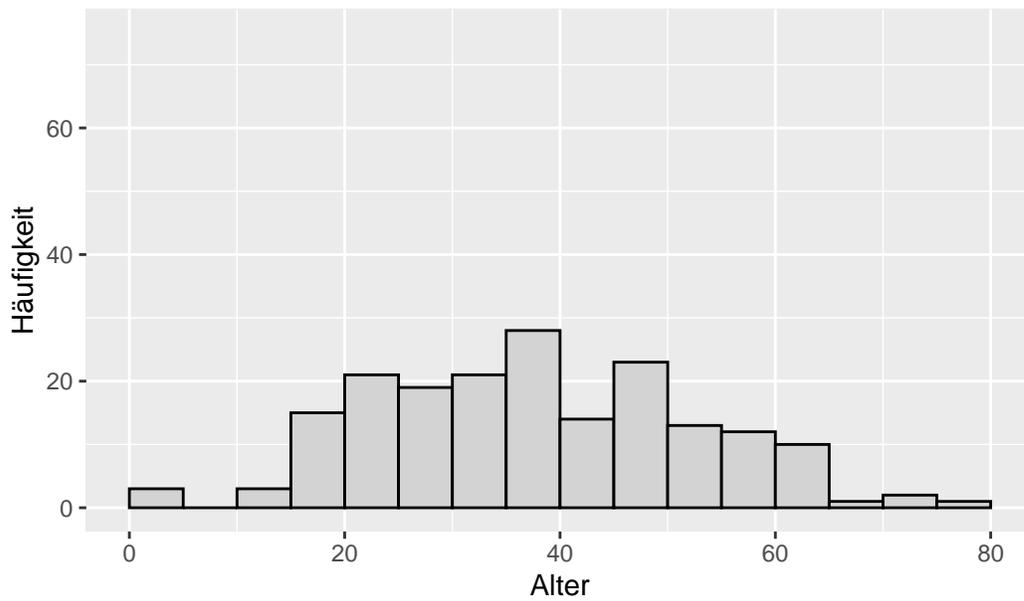
```
library(ggplot2)

ggplot(titanic, aes(x = Age)) + # Datengrundlage
  geom_histogram(breaks = seq(0, 80, 5), fill = "lightgrey", color = "black",
                na.rm = TRUE) + # Eigenschaften
  labs(title = "Alter der Passagiere", x = "Alter", y = "Häufigkeit") # Beschriftung
```



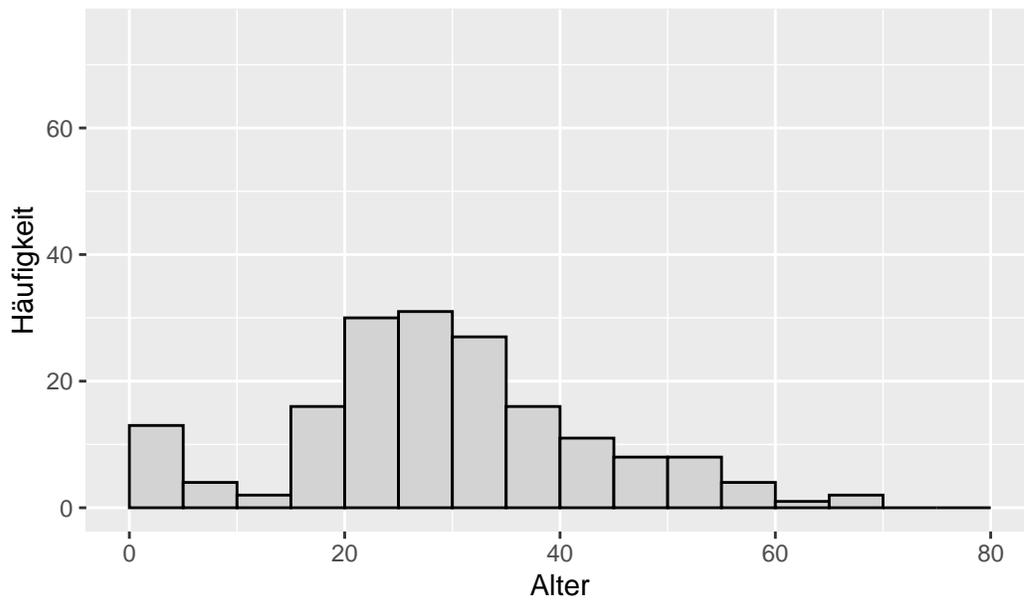
```
frst_class <- data.frame(titanic[Pclass == 1, ])  
# [Pclass == 1, ] -> Daten mit Pclass = 1  
scnd_class <- data.frame(titanic[Pclass == 2, ])  
thrd_class <- data.frame(titanic[Pclass == 3, ])  
  
ggplot(frst_class, aes(x = Age)) +  
  geom_histogram(breaks = seq(0, 80, 5), fill = "lightgrey",  
                color = "black", na.rm = TRUE) +  
  ylim(0,75) + # Häufigkeit auf Y-Achse von 0 bis 75  
  labs(title = "Alter der Passagiere in der 1. Klasse", x = "Alter",  
        y = "Häufigkeit")
```

Alter der Passagiere in der 1. Klasse



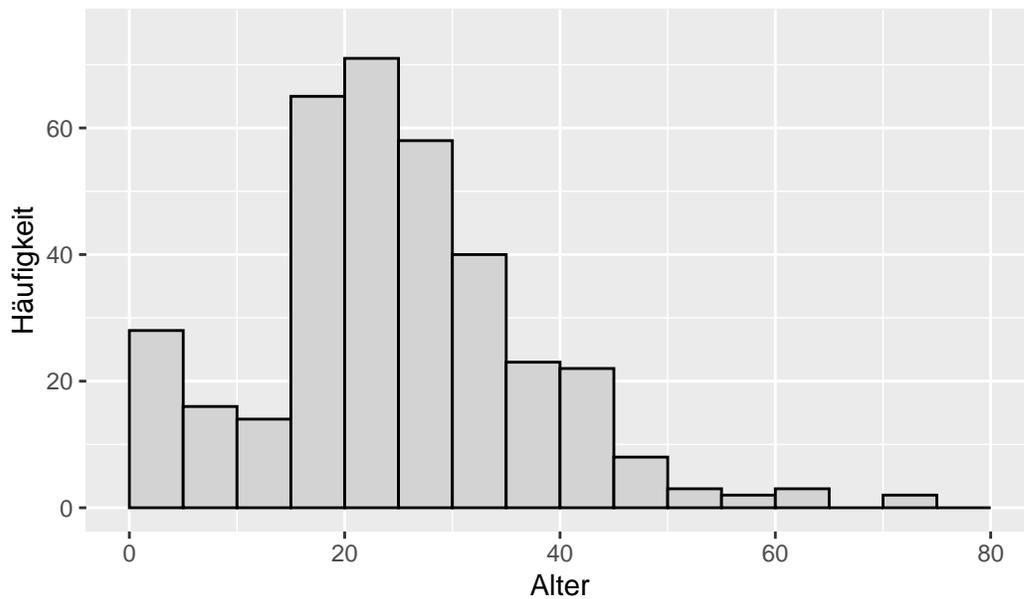
```
ggplot(scnd_class, aes(x = Age)) +  
  geom_histogram(breaks = seq(0, 80, 5), fill = "lightgrey",  
                color = "black", na.rm = TRUE) +  
  ylim(0,75) +  
  labs(title = "Alter der Passagiere in der 2. Klasse", x = "Alter",  
        y = "Häufigkeit")
```

Alter der Passagiere in der 2. Klasse



```
ggplot(thrd_class, aes(x = Age)) +  
  geom_histogram(breaks = seq(0, 80, 5), fill = "lightgrey",  
                color = "black", na.rm = TRUE) +  
  ylim(0,75) +  
  labs(title = "Alter der Passagiere in der 3. Klasse", x = "Alter",  
        y = "Häufigkeit")
```

Alter der Passagiere in der 3. Klasse



1. Passagiere der 1. Klasse im Schnitt älter, als die der anderen Klassen
2. Histogramm der 1. Klasse symmetrisch
3. Histogramm der 2. und 3. Klasse rechtsschief

Aufgabe 4

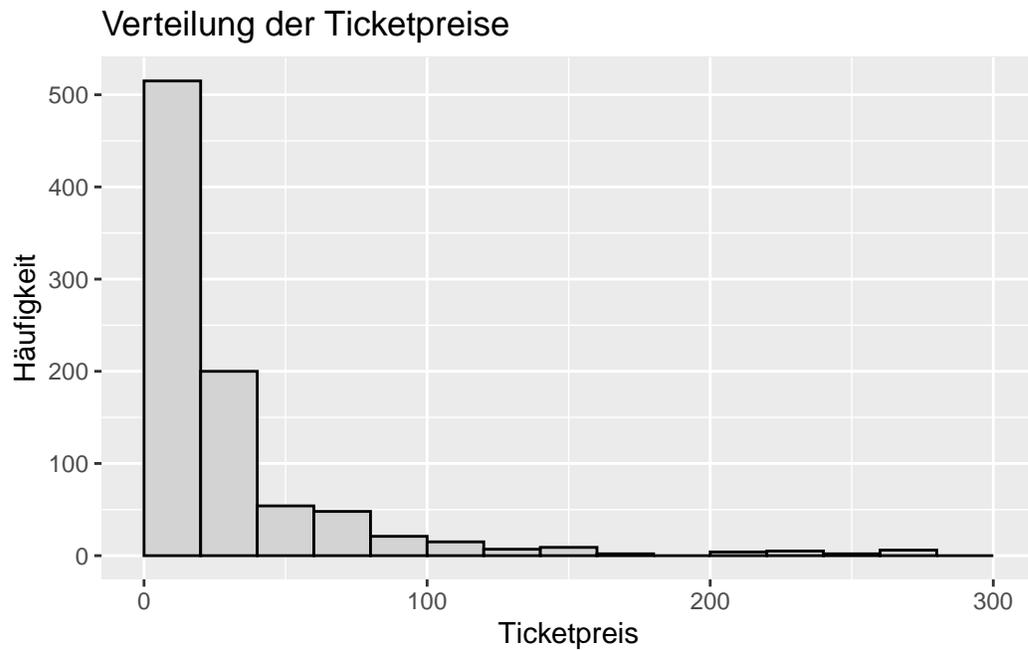
Berechnen Sie die Wahrscheinlichkeit eine zufällige Person auszuwählen, die ein Ticket mit einem Preis von über 100 erworben hatte. Welche Ihnen bekannte Verteilung könnte die Daten gut beschreiben?

```
# Wahrscheinlichkeit berechnen  
print(sum(Fare >= 100) / length(Fare))
```

```
[1] 0.05948373
```

Die Wahrscheinlichkeit eine zufällige Person auszuwählen, die über 100 GE für ihr Ticket gezahlt ist 5,95%

```
ggplot(titanic, aes(x = Fare)) +  
  geom_histogram(breaks = seq(0, 300, 20), fill = "lightgrey",  
                color = "black", na.rm = TRUE) +  
  labs(title = "Verteilung der Ticketpreise", x = "Ticketpreis", y = "Häufigkeit")
```



Das Histogramm erinnert an eine Gamma/Exponential Verteilung