' 1. Variablen'

' 1.1 Numerische kontinuierliche Variablen'

```
# entweder Integerwerte oder kontinuierliche Werte
```

```
zeit = read.table(file.path(pfadu, "zeit.txt"))  
head(zeit)  
class(zeit$Mo)
```

```
asp = read.table(file.path(pfadu, "asp.txt"))  
head(asp)  
class(asp$d)
```

'1.2 Faktoren'

```
# Diese sind die Variablen-Namen im Data-Frame  
names(asp)
```

```
class(asp$Vpn)
```

```
# Ein Faktor hat eine oder mehrere Stufen: diese sind  
# die verschiedenen Kategorien aus denen, der Faktor besteht  
levels(asp$Kons)
```

```
# Die Stufen sieht man auch wenn ein Faktor aufgerufen wird:  
asp$Kons[1:10]
```

'1.3 Funktionen für die Zusammenfassung numerischer Daten'

```
Mittelwert: mean()
```

```
mean(asp$d)
```

```
# Median (oder 50% Quantil). Der mittlere Wert in der sortierten Reihenfolge  
x = c(15, 19, -1, 10, 11, 18, 90000)  
median(x)
```

```
# das gleiche  
quantile(x, .5)
```

```
# Nach Sortierung sieht man, dass 15 die an der mittleren Stelle vorkommt.  
sort(x)  
# -1 10 11 15 18 19 90000
```

```
# Es gibt auch andere Bruchteile von Quantilen. Insbesondere  
# wird von dem interquartilen Bereich Gebrauch gemacht, um die Streueung einzuschätzen  
IQR(x)
```

```
# das gleich  
quantile(x, .75) - quantile(x, .25)
```

```
# Eine andere Messung der Streuungsgröße ist die Standardabweichung
```

```
sd(x)
```

```
y = c(15, 19, -1, 10, 11, 18, 20)
sd(y)
```

'1.4 Boxplot'

```
# Eine Abbildung, von der wir in der Statistik öfters Gebrauch machen
# werden ist der Boxplot. Die dicke Linie ist der Median, der
# Quadrat ist der interquartile Bereich
boxplot(y)
```

```
boxplot(asp$d)
```

```
# Meistens möchte man sehen, ob eine numerische Variable
# von den Faktoren-Stufen beeinflusst wird. Hier ist die Syntax dafür:
# (entspricht: 'd gegeben Kons')
boxplot(d ~ Kons, data = asp)
```

```
# Um gleichzeitig zwei Faktoren abzubilden:
# (entspricht: 'd gegeben Kons gekreuzt mit Bet')
boxplot(d ~ Kons * Bet, data = asp)
```

'1.5 Barchart'

```
# Die Funktion barchart() in library(lattice) soll verwendet werden
# wenn man die Häufigkeiten von Faktoren abbilden will - also wenn
# die Ausgabe von table() abgebildet werden soll. Um zu barchart() zu gelangen,
# muss library(lattice) einmal im Session eingegeben werden - die
library(lattice)
```

```
tab = table(asp$Kons)
tab
barchart(tab)
# oder
barchart(tab, horizontal=F)
```

```
# Gekreuzte Tabellen können auch mit barchart() abgebildet werden:
```

```
tab = table(asp$Kons, asp$Bet)
tab
#      be  un
# k  853  425
# t  450 1164
```

```
barchart(tab, horizontal=F, auto.key=T)
```

```
# oder in der anderen Reihenfolge
tab = table(asp$Bet, asp$Kons)
barchart(tab, horizontal=F)
```

```
# Für mehrere Faktoren, z.B.
# (Hier muss read.delim() statt read.table() verwendet werden,
# weil es Leerstellen in einigen Stufen vorkommen
```

```

kj = read.delim(file.path(pfadu, "kj.txt"))
names(kj)
tab = table(kj$str, kj$age, kj$gender)
barchart(tab, horizontal=F, beside=T, auto.key=T)

# 1.6 Fragen

# Für den Data-Frame dip
dip = read.table(file.path(pfadu, "dip.txt"))

# erstellen Sie einen Boxplot der Dauer
# für die Faktoren V (Vokal) und Vpn (Versuchsperson) zusammen
# Gibt es systematische Dauer-Unterschiede zwischen
# den Diphthongen und/oder zwischen den Versuchspersonen?
head(dip)
boxplot(d ~ V * Vpn, data = dip)
# Es gibt einen Unterschied zwischen den Sprechern
# (68 hat längere Dauern) aber kaum zwischen den Diphthongen.
# Dies ist auch deutlich, wenn man die Faktoren getrennt
# anschaut
boxplot(d ~ V, data = dip)
boxplot(d ~ Vpn, data = dip)

# Für den Data-Frame rating:
rating = read.table(file.path(pfadu, "rating.txt"))
# verwenden Sie boxplots um einzuschätzen, inwiefern Rating
# von der Grammatikalität (Gram) und Sprache (Lang)
# beeinflusst werden.
head(rating)
boxplot(Rating ~ Gram * Lang, data = rating)
# es gibt minimale/keine Unterschiede zwischen den Sprachen:
boxplot(Rating ~ Lang, data = rating)
# aber doch in der Grammatikalität
boxplot(Rating ~ Gram, data = rating)
# (und in beiden Sprachen, wie die erste gekreuzte Abbildung zeigte)

# Für den Data-Frame sz
sz = read.table(file.path(pfadu, "sz.txt"))

# erstellen Sie einen Barchart, um die Häufigkeiten zu zeigen,
# mit denen /s/ oder /z/ von Sprechern aus Bayern und
# Schleswig-Holstein produziert wurden.
library(lattice)
head(sz)
tab = table(sz$Frikativ, sz$Dialekt)
barchart(tab, auto.key=T, horizontal=F)
# Eindeutig produzieren die Bayern proportional mehr /s/
# im Vgl. zu Sprechern aus SH.

```

```
# Für den Data-Frame preasp
preasp = read.table(file.path(pfadu, "preasp.txt"))

# erstellen Sie einen Barchart, um einzuschätzen, inwiefern
# die Verteilung von ±preasp (ob Pre-aspiration vorkam oder nicht)
# von dem davor kommenden Vokal (vtype) beeinflusst wurde.
head(preasp)
tab = table(preasp$vtype, preasp$Pre)
# es gibt übrigens noch eine Syntax dafür:
tab = with(preasp, table(vtype, Pre))
barchart(tab, auto.key=T, horizontal=F)
# Die Proportion [+preasp]:[-preasp] ist ähnlich für /e/ und /o/;
# aber diese Proportionen für /e, o/ sind beide größer als für /a/
# (d.h. für /a/ ist die Proportion [+preasp]:[-preasp]
# ca. 50-50, und für /e, o/ eher 70:30.
```