

Solution to Series 1

1. a) Einen ersten Überblick, wie die Spannweiten von Schreibhand und Nicht-Schreibhand verteilt sind, gewinnt man mit einer Analyse der Kennzahlen:

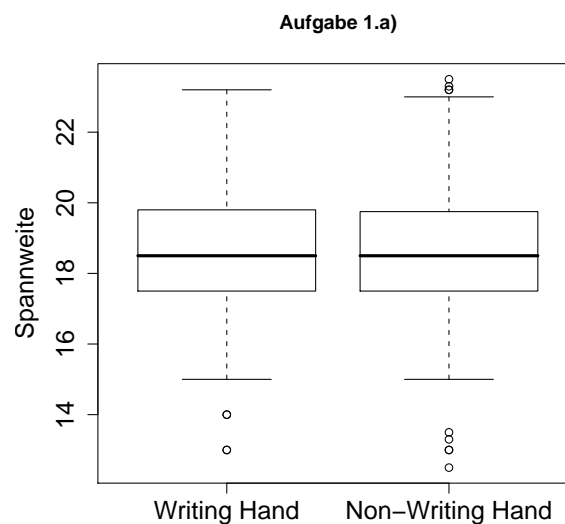
R-Output: (nur teilweise aufgeführt)

```
> library(MASS)
> summary(survey)
```

Wr.Hnd		NW.Hnd	
"Min. :13.00	"	"Min. :12.50	"
"1st Qu.:17.50	"	"1st Qu.:17.50	"
"Median :18.50	"	"Median :18.50	"
"Mean :18.67	"	"Mean :18.58	"
"3rd Qu.:19.80	"	"3rd Qu.:19.73	"
"Max. :23.20	"	"Max. :23.50	"
"NA's : 1.00	"	"NA's : 1.00	"

Die Kennzahlen sind sehr ähnlich und wir können daraus nicht folgern, dass es Unterschiede zwischen den Spannweiten gibt. Viel übersichtlicher ist aber eine grafische Analyse mittels einem Boxplot:

```
> boxplot(survey$Wr.Hnd, survey$NW.Hnd, main="Aufgabe 1.a)",
          ylab="Spannweite", names=c("Writing Hand", "Non-Writing Hand"),
          cex.lab = 1.5, cex.axis = 1.5)
```



Wir folgern, dass es in diesem Datensatz keine Unterschiede zwischen der Spannweite der Schreibhand und der Nicht-Schreibhand gibt.

- b) Die gewünschte Antwort bekommen wir mit dem Befehl

```
> survey[rev(order(survey$Age))[1:2],]
```

	Sex	Wr.Hnd	NW.Hnd	W.Hnd	Fold	Pulse	Clap	Exer	Smoke	Height	M.I	Age
171	Female	16.5	17.0	Right	L on R	NA	Right	Some	Never	168.00	Metric	73.000
154	Male	21.5	21.6	Right	R on L	69	Right	Freq	Never	172.72	Imperial	70.417

oder alternativ mit

```
> survey[rev(order(survey$Age))[1:2], "Smoke"]
```

```
[1] Never Never
```

```
Levels: Heavy Never Occas Regul
```

Die beiden ältesten Studierenden haben das siebzigste Altersjahr bereits hinter sich und sind *Nichtraucher*.

- c) Mit der R-Funktion

```
> pairs(survey)
```

erhalten wir paarweise Plots aller Variablen, die in dieser Studie beobachtet wurden. Wir erkennen daraus z.B., dass das Geschlecht der Studierenden einen Einfluss auf den Puls hat, was durchaus plausibel ist. Der Traininszustand (Variable "Exercise") scheint gemäss dem Plot etwas erstaunlicherweise keinen Einfluss auf die Herzfrequenz zu haben. Weiter scheint es, dass zwischen dem Ruhepuls von Links- und Rechtshändern Unterschiede bestehen. Dies muss aber auf jeden Fall weiter untersucht und zusätzlich hinterfragt werden.

- d) Die erste Zeile kreiert neue Variablen mit den Namen *Agejung* und *Pulsejung*, welche nur aus den Werten der Personen bestehen, welche jünger als 30 Jahre alt sind. Danach wird Puls gegen Alter gezeichnet.

Der Befehl in der zweiten Zeile fittet eine lineare Regressionsgerade. Die ganze Information über das lineare Modell wird in der Variablen *lmobj* abgespeichert. Durch den Befehl *abline* wird die Regressionsgerade in den bestehenden Plot gezeichnet.