

## Lösung zu Kapitel 9: Beispiel 2

Von den 20 besten Nachwuchstennisspielern im Alter von 12 Jahren wurde nach acht Jahren überprüft, welchen Rang sie in der nationalen Rangliste einnehmen.

Das Datenfile `tennis.csv` enthält die beiden Variablen:

`rang12` Ranglistenplatz in der Alterskategorie

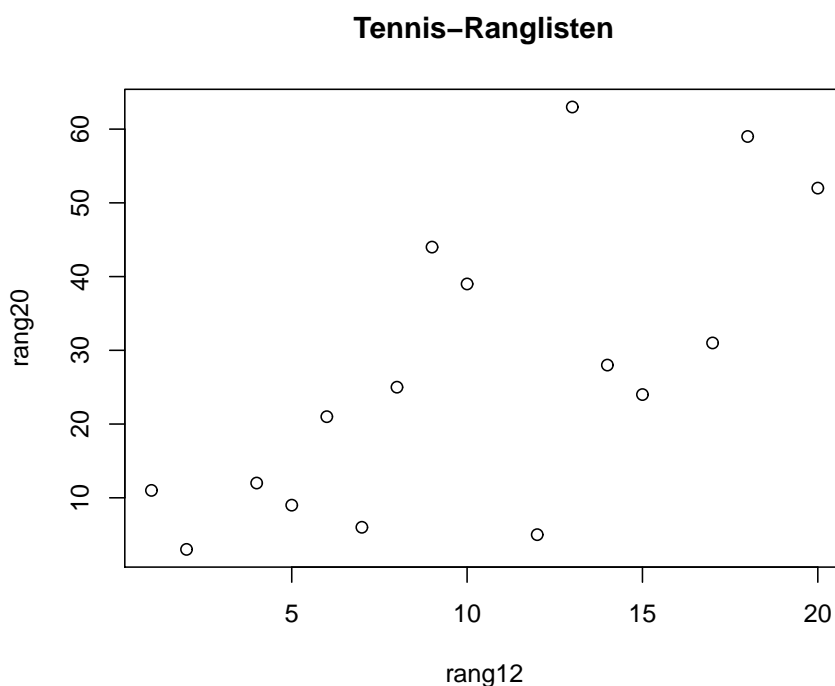
`rang20` nationaler Ranglistenplatz (Missing bedeutet Karriere beendet)

- Gibt es einen Zusammenhang zwischen der Spielstärke mit 12 Jahren und der mit 20 Jahren?

Nach Einlesen der Daten machen wir uns mit einem Streudiagramm ein Bild von den Daten:

**R**

```
> tennis <- read.csv2("tennis.csv", header = T)
> attach(tennis)
> plot(rang12, rang20, main = "Tennis-Ranglisten")
```



**Abbildung 1:** Streudiagramm der Ranglistenplätze von Tennisspielern.

Es ist klar ersichtlich (► Abbildung 1), dass es einen Zusammenhang zwischen den Ranglistenplätzen im Alter von 12 und 20 Jahren gibt.

Die Stärke des Zusammenhangs kann mit einem Korrelationskoeffizienten bestimmt werden. Da hier nur ordinale Angaben (Plätze in einer Rangliste), verwenden wir den Spearman-Korrelationskoeffizienten (statt dem üblichen Pearson-Korrelationskoeffizienten).



```
> cor.test(rang12, rang20, method = "spearman")
```

Spearman's rank correlation rho

```
data: rang12 and rang20
S = 208, p-value = 0.003796
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.6941176
```

Der Korrelationskoeffizient beträgt ca. 0.69, der p-Wert für den Test, ob der Spearman-Korrelationskoeffizient von 0 abweicht, zeigt mit einem Wert von weniger als 0.004 ein signifikantes Ergebnis an. Der Platz in der Rangliste für 12-Jährige Tennisspieler korreliert also stark mit dem Platz in der Gesamtangliste acht Jahre später (zumindest für die besten Jugenspieler).

Zum Schluss lösen wir die Verbindung zum Datensatz wieder auf:



```
> detach(tennis)
```