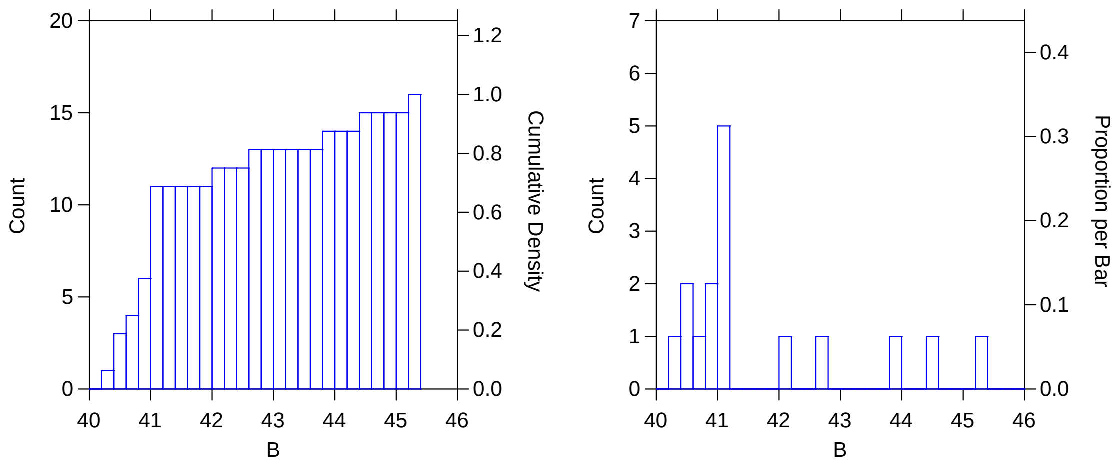
Musterlösung zur Übungsserie 2

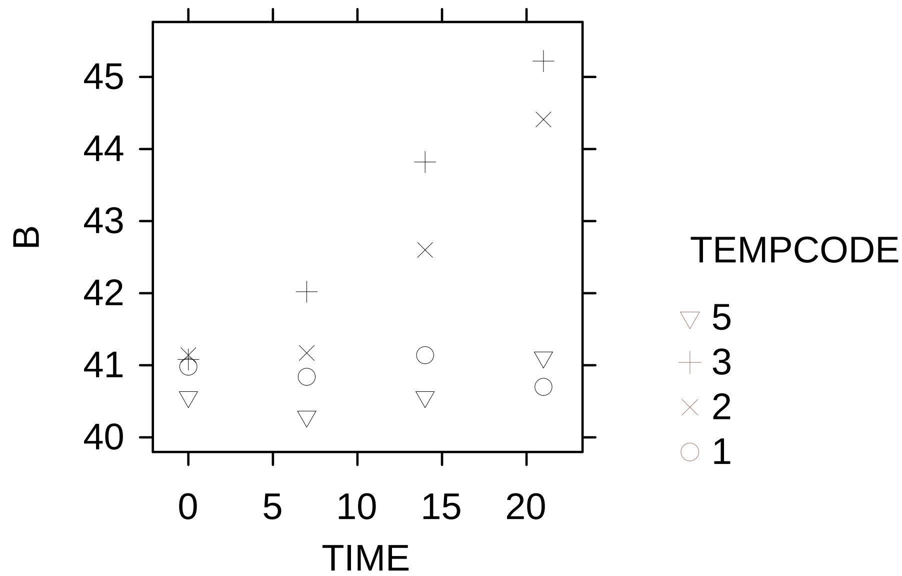
1. **a)** Die gestellten Fragen lassen sich am Besten mit dem kumulativen Häufigkeitsdiagramm beantworten:



25% der Intensitätswerte bleiben unter 40.8, während 20% der Beobachtungen eine Intensität über 43 aufweisen.

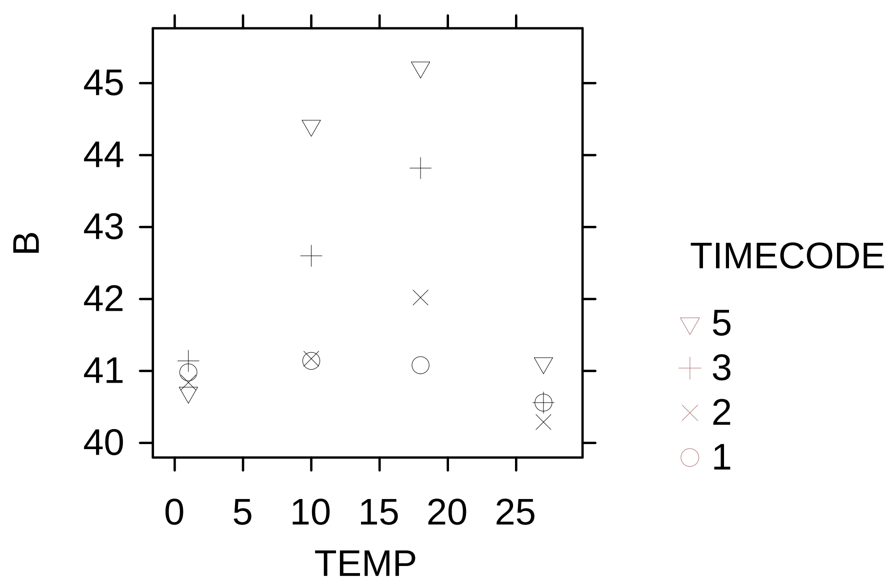
b)
```
Stem and Leaf Plot of variable:   B, N = 16

         Minimum:        40.290
         Lower hinge:    40.770
         Median:         41.125
         Upper hinge:    42.310
         Maximum:        45.220

              40   2
              40 H 55789
              41 M 01111
              41
              42 H 0
              42   6
              43
              43   8
              44   4

      * * * Outside Values * * *
              45   2
```
Als "Outside Value" wird nur der Wert 45.22 angegeben.

c) Die Farbintensität steigt mit der Lagerungsdauer nur bei Äpfeln, die bei 10 oder 18 Grad gelagert wurden.



(Bemerkung zur Kodierung: 1 entspricht der Temperatur von 1 Grad, 2 von 10 Grad, 3 von 18 Grad und 5 von 27 Grad.)

d) Die Farbintensität hängt nur bei Äpfeln, die lange gelagert werden von der Temperatur ab, was auch zu erwarten ist.



(Bemerkung zur Kodierung: 1 entspricht den kürzesten Lagerungsdauer, 5 den längsten Lagerungsdauer.)

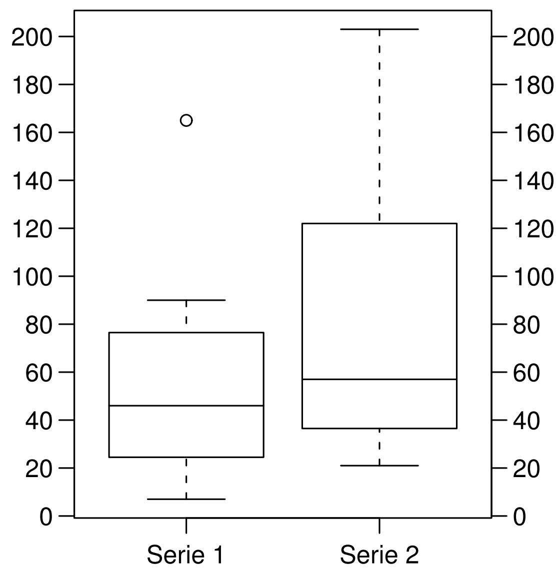
2. **a)** Man findet:

A5: symmetrische Verteilung um 0.6, kleine Streuung.
B1: Lage gegen kleinere Werte hin verschoben, extreme Werten nach oben.
C4: grosse Streuung, symmetrisch, keine extremen Werte.
D3: Lage gegen grössere Werte hin verschoben, extreme Werten nach unten.
E2: symmetrische Verteilung um 0.5, kleine Streuung.

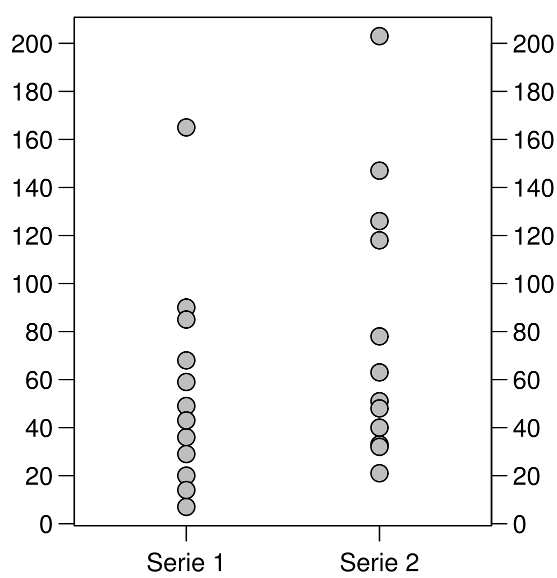
b) Man findet:

A3: kleinste Streuung.
B2: Ausreisser nach oben: Mittelwert und Varianz gross. 1.Quartil bei 2.2.
C4: Kleine Streuung (aber grösser als A), 1.Quartil bei 2.2.
D5: Kleine Streuung (aber grösser als A), 1.Quartil bei 2.5 (grösser als C).
E1: Ausreisser nach unten: kleiner Mittelwert und grosse Varianz. 1.Quartil bei 1.

3. b) Die 2 Box-Plots sehen so aus:



Der Median ist ungefähr gleich in beiden Serien. Weiter scheint es, dass die 2. Serie eine grössere Streuung als die 1. hat (die Kiste ist grösser). Das kann bei Box-Plots mit wenigen Beobachtungen (hier 12) passieren (der Box-Plot kann in solchen Fälle irreführend aussehen). Man sieht nämlich im Streudiagramm, dass die 2 Serien ähnlich sind. Weiter der Wilcoxon-Mann-Whitney Test beibehält die Nullhypothese $H_0 : F_{serie1} = F_{serie2}$: der p-Wert ist 0.2913 (siehe später).
Man darf deshalb die 2 Serien zusammenfassen.



c) Man findet:

Median	=	50
Mittelwert	=	67.71

4. a) Sei X die Anzahl der verunreinigten Einzelproben in einer Sammelprobe. Die Erfolgswahrscheinlichkeit p, dass eine Einzelprobe verunreinigt ist, beträgt 0.02. Unter der Annahme, dass die Einzelproben voneinander unabhängig sind, gilt $X \sim \mathrm{Bin}\,(n = 10, p = 0.02)$.

Die Wahrscheinlichkeit, in der Sammelprobe keine Verunreinigung zu finden, ist gegeben durch

$$\mathrm{P}\,[X = 0] = \binom{10}{0} 0.02^0 \cdot 0.98^{10} = 0.98^{10} = 0.817\,.$$

Anderer Lösungsweg: Jede einzelne Probe ist unabhängig von den anderen Proben mit 98% Wahrscheinlichkeit sauber. Also gilt **(Multiplikationssatz für unabhängige Ereignisse)**

$$\mathrm{P}\,[\text{alle Proben sauber}] = \prod_{i=1}^{10} \mathrm{P}\,[\text{i-te Probe sauber}] = 0.98^{10} = 0.817\,.$$

b) Die Zufallsvariable Y kann nur die Werte 1 oder 11 annehmen, denn:

- falls alle Proben sauber sind, ist man nach einer Untersuchung fertig: $Y{=}1$
- sonst muss man jede Probe einzeln untersuchen (man darf nicht stoppen, wenn man die erste verunreinigte Probe gefunden hat, denn es könnte ja noch mehrere geben!), also $Y{=}11$.

Folglich

$$\begin{cases} \mathrm{P}\,[Y = 1] = & \mathrm{P}\,[\text{alle Proben sauber}] = & 0.817 \\ \mathrm{P}\,[Y = 11] = & 1 - \mathrm{P}\,[Y = 1] = & 0.183 \end{cases}$$

c) Die durchschnittliche Anzahl Analysen pro Sammelprobe ist gegeben durch den Erwartungswert der Zufallsvariablen Y:

$$\begin{aligned} \mathbf{E}\,[Y] &= \sum_{k=0}^{\infty} k\,\mathrm{P}\,[Y = k] = 1 \cdot \mathrm{P}\,[Y = 1] + 11 \cdot \mathrm{P}\,[Y = 11] \\ &= 1 \cdot 0.817 + 11 \cdot 0.183 = 2.83 \end{aligned}$$

„Im Durchschnitt“ spart man also $10 - 2.83 = 7.17 \approx 7$ Untersuchungen ein.