

Lösung zu Kapitel 8: Beispiel 1

Im Datenfile `us-president.csv` ist das Alter der US-Präsidenten bei Amtsantritt angegeben.

- Beschreiben Sie den Datensatz mit Histogramm, Boxplot und Maßzahlen!

Zuerst lesen wir die Daten ein und schauen uns die ersten Fälle an.

R

```
> us_president <- read.csv2("us-president.csv")
> head(us_president)
```

```
  President Age
1 Washington  57
2   J.Adams  61
3  Jefferson  57
4   Madison  58
5   Monroe  58
6  J.Q.Adams  57
```

Jetzt werden die Daten mittels Histogramm grafisch analysiert (s. ► Abbildung 1). Das Alter der Präsidenten bei Amtsantritt sieht hier in etwa normalverteilt aus. Das heißt, es gibt nur sehr wenige sehr junge Präsidenten (bei Amtsantritt um die 40 Jahre) und sehr wenige sehr alte Präsidenten (bei Amtsantritt um die 70 Jahre). Anscheinend gibt es jedoch bisher sehr viel weniger Präsidenten im Alter von 60 Jahren bei Amtsantritt als lt. Normalverteilung zu erwarten wäre. An dieser Stelle ist im Histogramm ein Loch ersichtlich.

R

```
> attach(us_president)
> hist(Age)
```

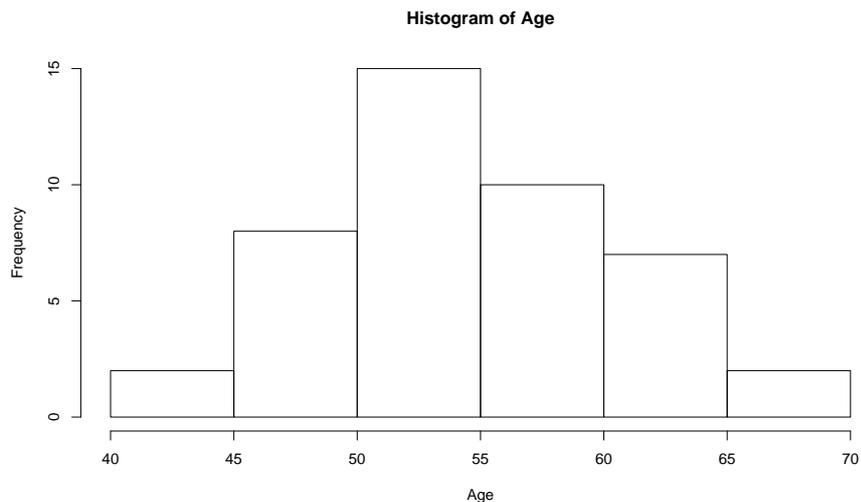


Abbildung 1: Histogramm des Alters der US-Präsidenten bei Amtsantritt

R

```
> boxplot(Age)
```

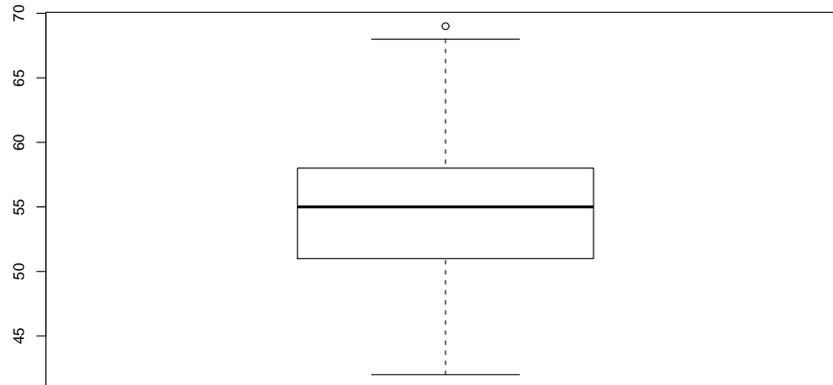


Abbildung 2: Boxplot des Alters der US-Präsidenten bei Amtsantritt

Betrachtet man dazu die deskriptiv Statistiken der Variable `Age` so sieht man, dass der jüngste Präsident bei Amtsantritt 42 Jahre und der älteste 69 Jahre alt war. Im Mittel waren die Präsidenten bisher bei Amtsantritt 54.89 Jahre alt. Das getrimmte Mittel und der Median sind nur unwesentlich kleiner als der Mittelwert. Dies liegt wohl an dem moderaten Ausreisser der im Boxplot (► Abbildung 2) ersichtlich ist. Es liegt eine leicht schiefe Verteilung vor.

R

```
> summary(Age)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 42.00  51.00   55.00  54.89  58.00   69.00
```

R

```
> mean(Age)
```

```
[1] 54.88636
```

R

```
> mean(Age, trim = 0.05)
```

```
[1] 54.825
```

R

```
> median(Age)
```

```
[1] 55
```

R

```
> min(Age)
```

```
[1] 42
```

R

```
> max(Age)
```

```
[1] 69
```

R

```
> quantile(Age, 0.25)
```

```
25%  
51
```

R

```
> quantile(Age, 0.75)
```

```
75%  
58
```

R

```
> sd(Age)
```

```
[1] 6.217814
```

R

```
> var(Age)
```

```
[1] 38.66121
```

R

```
> IQR(Age)
```

```
[1] 7
```

R

```
> range(Age)
```

```
[1] 42 69
```

R

```
> detach(us_president)
```