

Programmieren mit R für Einsteiger

4. Grafiken / 4.6 Verteilungsplots



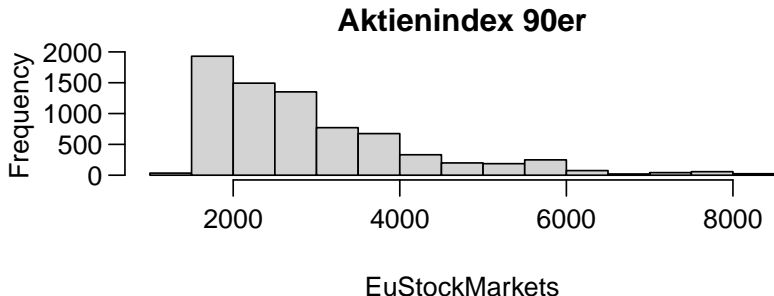
Berry Boessenkool



frei verwenden, zitieren

2022-02-25 11:41

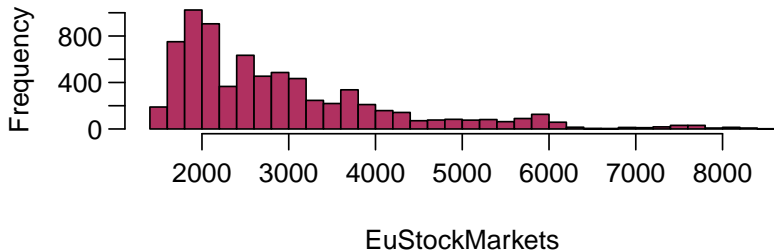
```
h <- hist(EuStockMarkets, las=1, main="Aktienindex 90er")
```



```
str(h)
## List of 6
## $ breaks : int [1:16] 1000 1500 2000 2500 3000 ...
## $ counts : int [1:15] 34 1930 1493 1352 771 ...
## $ density : num [1:15] 9.14e-06 5.19e-04 ...
## $ mids : num [1:15] 1250 1750 2250 2750 3250 ...
## $ xname : chr "EuStockMarkets"
## $ equidist: logi TRUE
```

```
hist(EuStockMarkets, las=1, breaks=50, col="maroon")
```

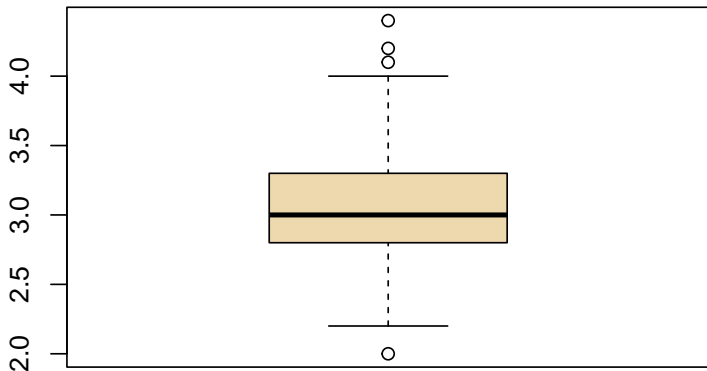
Histogram of EuStockMarkets



```
summary(iris$Sepal.Width)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2.000	2.800	3.000	3.057	3.300	4.400

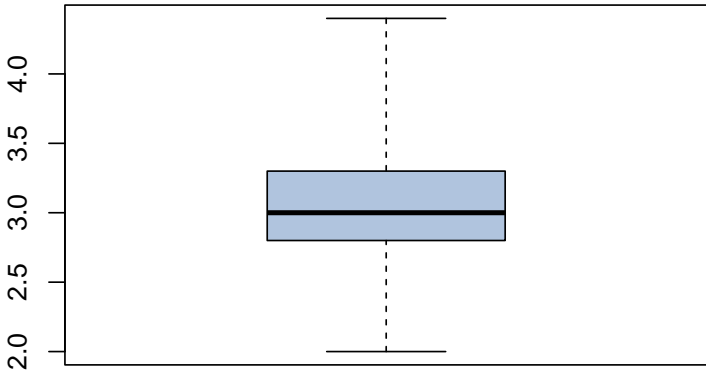
```
boxplot(iris$Sepal.Width, col="wheat2")
```



Punkte sind meist kein 'Outlier' sondern gehören dazu

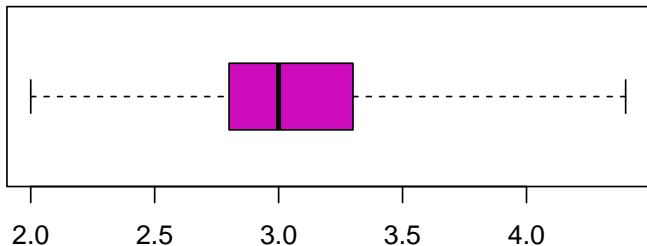
`boxplot()` zeichnet 'Ausreißer' standardmäßig als individuelle Punkte. Die Schwelle dafür, als solche eingestuft zu werden, ist eine willkürliche Distanz außerhalb der IQR Box. (IQR = InterQuartileRange = Spanne zwischen 25% und 75% Quantile der Daten)

```
boxplot(iris$Sepal.Width, range=0, col="lightsteelblue")
```

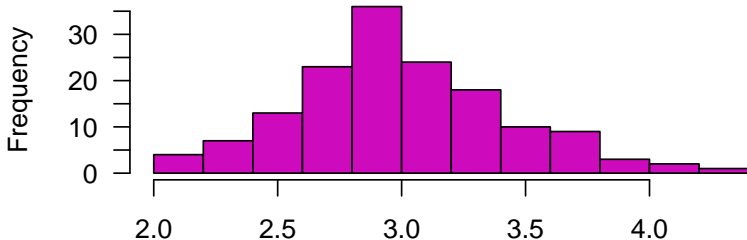


Horizontale Boxplots

```
boxplot(iris$Sepal.Width, horizontal=TRUE, range=0, col=6)
```

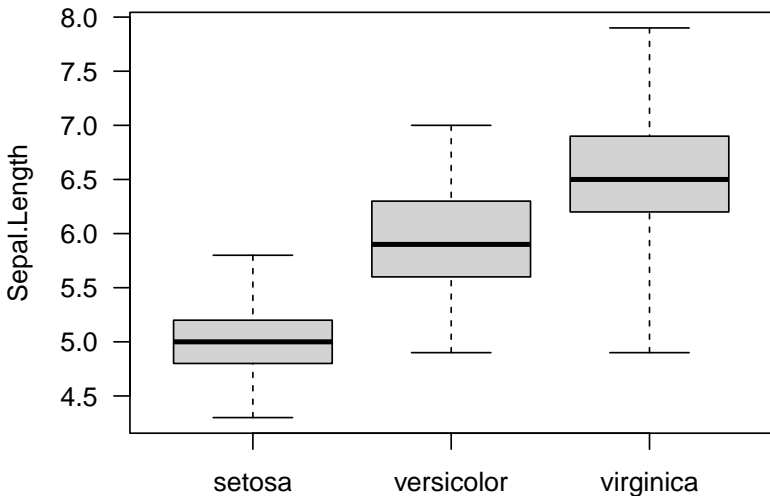


```
hist(iris$Sepal.Width, col=6, main="", las=1)
```



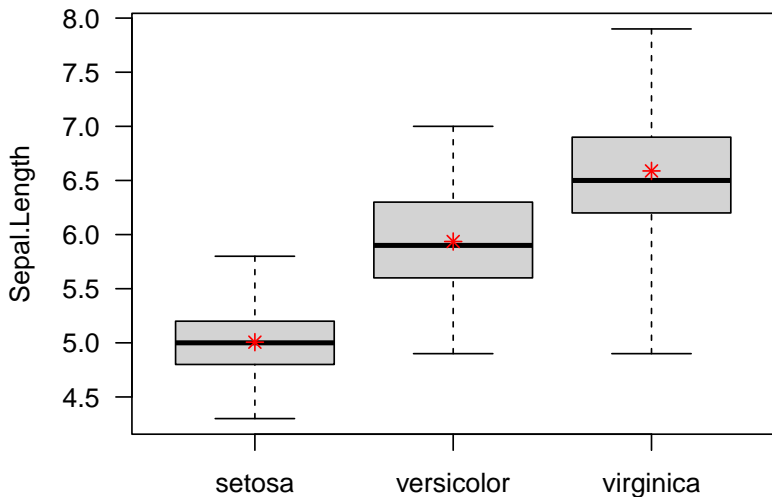
Boxplots mehrerer Spalten

```
boxplot(Sepal.Length ~ Species, data=iris, range=0, las=1)  
# auch manchmal hilfreich (+ einfacher als bei barplots):  
werte ~ gruppe1+gruppe2 # für doppelt gruppierte Daten
```



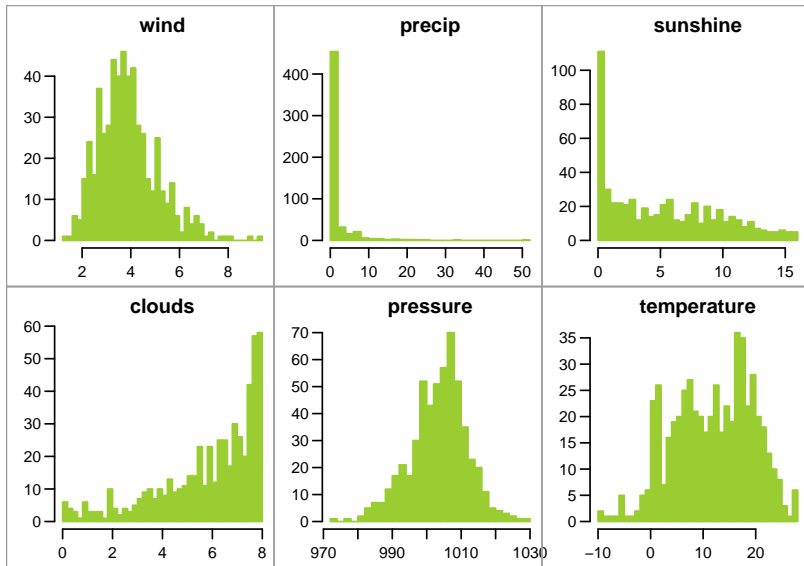
Boxplots mit arithmetischen Mittelwerten

```
boxplot(Sepal.Length ~ Species, data=iris, range=0, las=1)  
averages <- tapply(iris$Sepal.Length, iris$Species, mean)  
points(1:3, averages, pch=8, col="red")
```



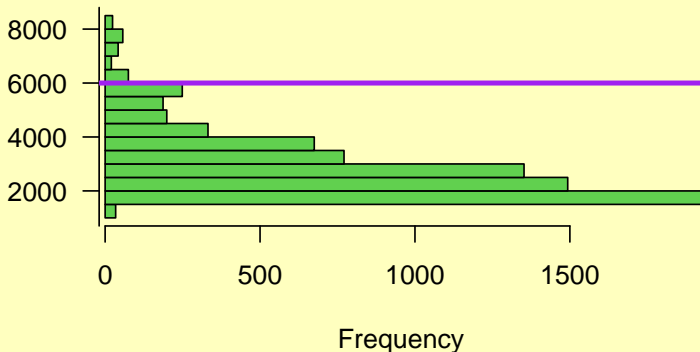
Histogramme und Boxplots: grafische Darstellung der Häufigkeitsverteilung

- ▶ `hist` (`breaks`, `las`, `col`, `main`)
- ▶ `boxplot` (`range=0`, `horizontal`)
- ▶ `boxplot` ($y \sim x$)
- ▶ Boxplots um viele Gruppen zu vergleichen, Histogramme um die tatsächliche Verteilung zu sehen

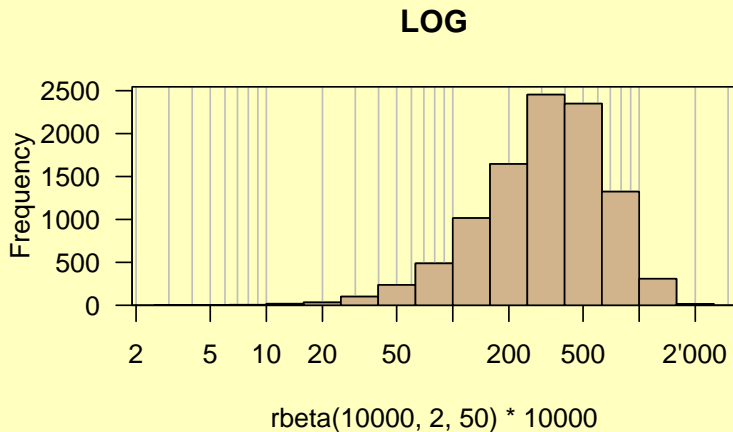


```
hpos <- berryFunctions::horizHist(EuStockMarkets, col=3)  
abline(h=hpos(6000), col="purple", lwd=3)
```

Histogram of EuStockMarkets

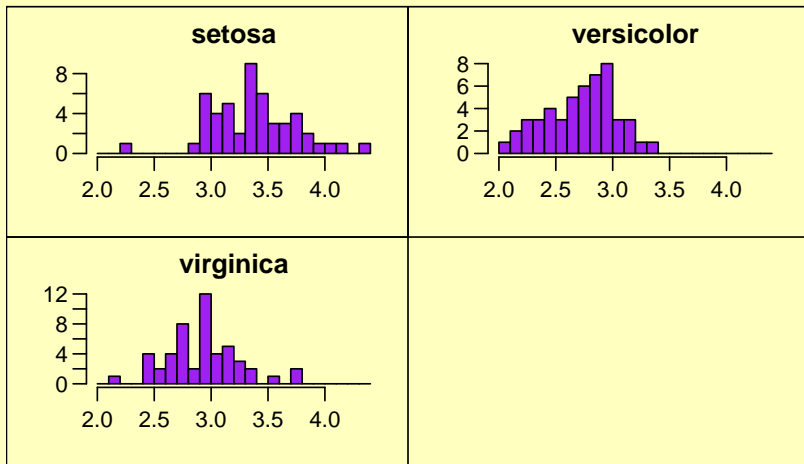


```
berryFunctions::logHist(rbeta(1e4, 2, 50)*1e4, main="LOG")
```

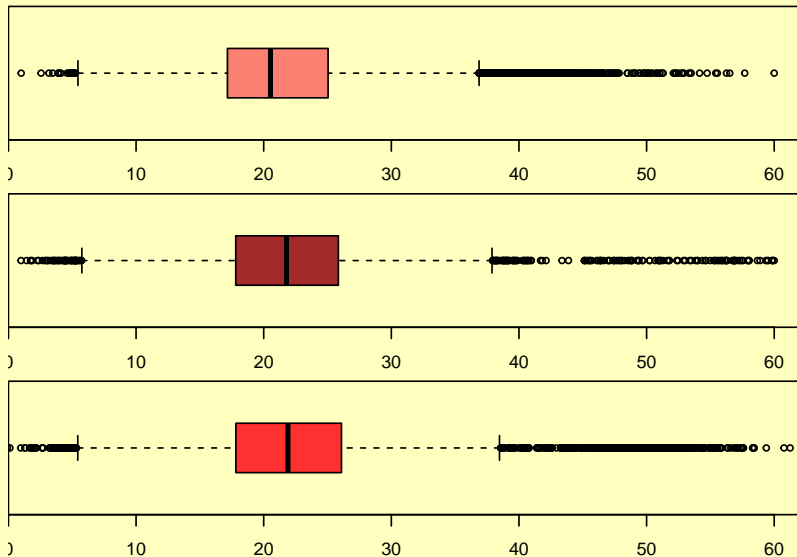


```
berryFunctions::groupHist(iris, "Sepal.Width", "Species",  
                           unit="cm")
```

Histograms of Sepal.Width [cm] in iris, grouped by Species

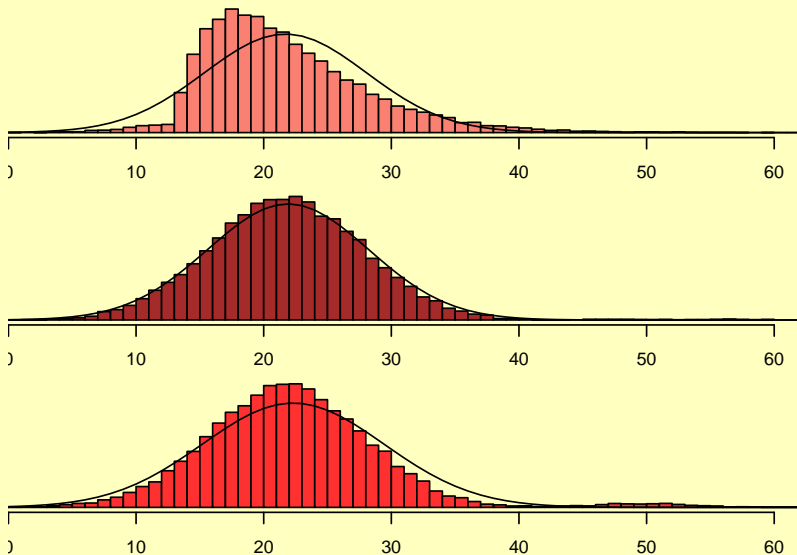


Boxplots sollten nur verglichen werden,



wenn die unterliegende Verteilungen gleich sind!

siehe z.B. [why not to trust statistics](#) und [boxplot R](#)



```
vioplot::vioplot(x,y,z, h=0.5, horizontal=TRUE)
```

