Biostatistik, WS 2013/2014

Deskriptive Statistik

Matthias Birkner

http://www.mathematik.uni-mainz.de/~birkner/Biostatistik1314/

22.11.2013

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

It is easy to lie with statistics.
It is hard to tell the truth without it.

Andrejs Dunkels

Was ist Statistik?

Die Natur ist voller Variabilität.

Wie geht man mit variablen Daten um?

Es gibt eine mathematische Theorie des Zufalls:

die Stochastik.

Idee der Statistik

Variabilität

(Erscheinung der Natur)

durch

Zufall

(mathematische Abstraktion)

modellieren.

Statistik

=

Datenanalyse

mit Hilfe

stochastischer Modelle

Beispiel

Daten aus einer Diplomarbeit aus 2001 am Forschungsinstitut Senckenberg, Frankfurt am Main

Crustaceensektion

Leitung: Dr. Michael Türkay

Charybdis acutidens TÜRKAY 1985

Der Springkrebs

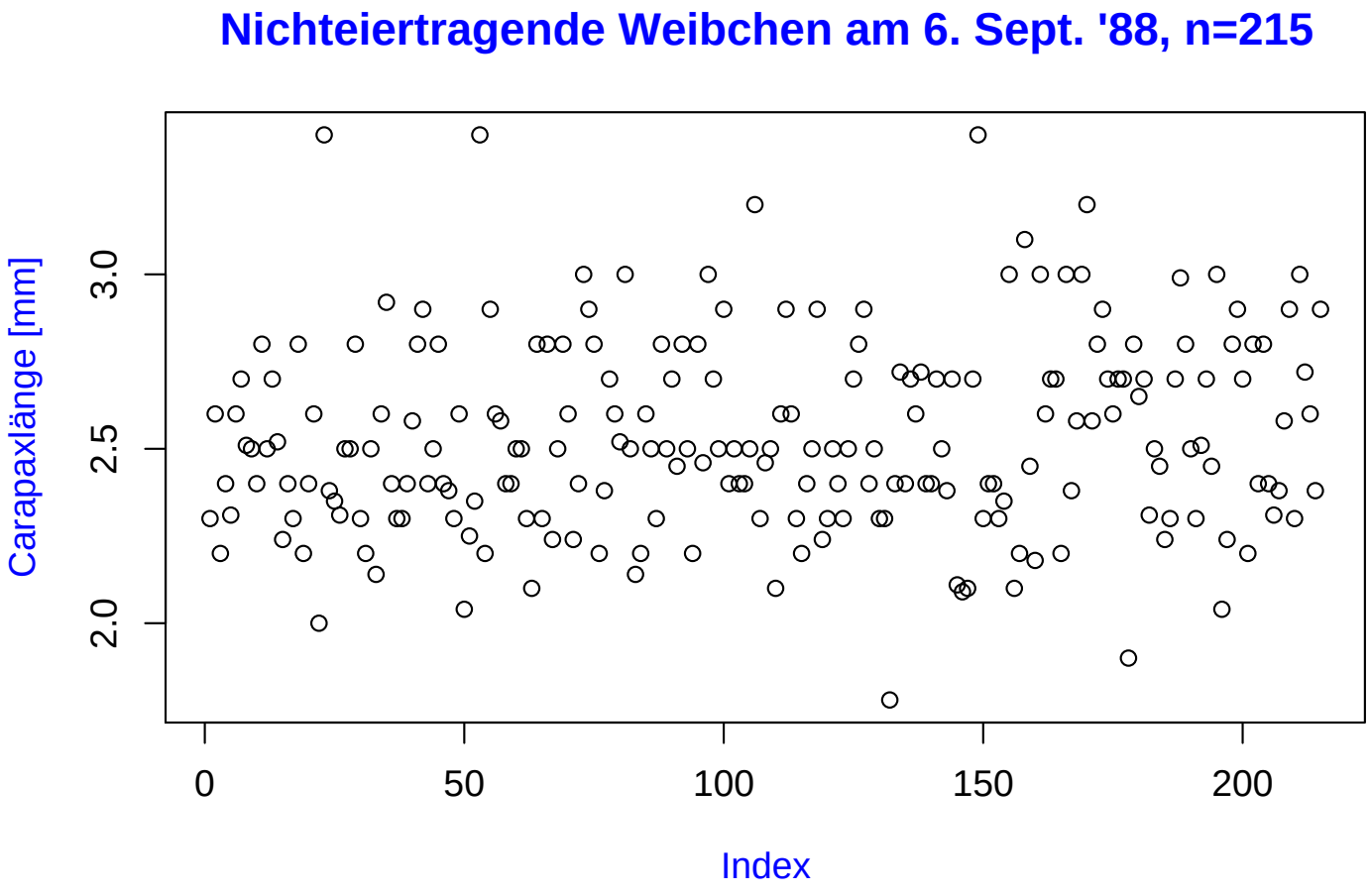
Galathea intermedia

Helgoländer Tiefe Rinne, Fang vom 6.9.1988

Carapaxlänge (mm):

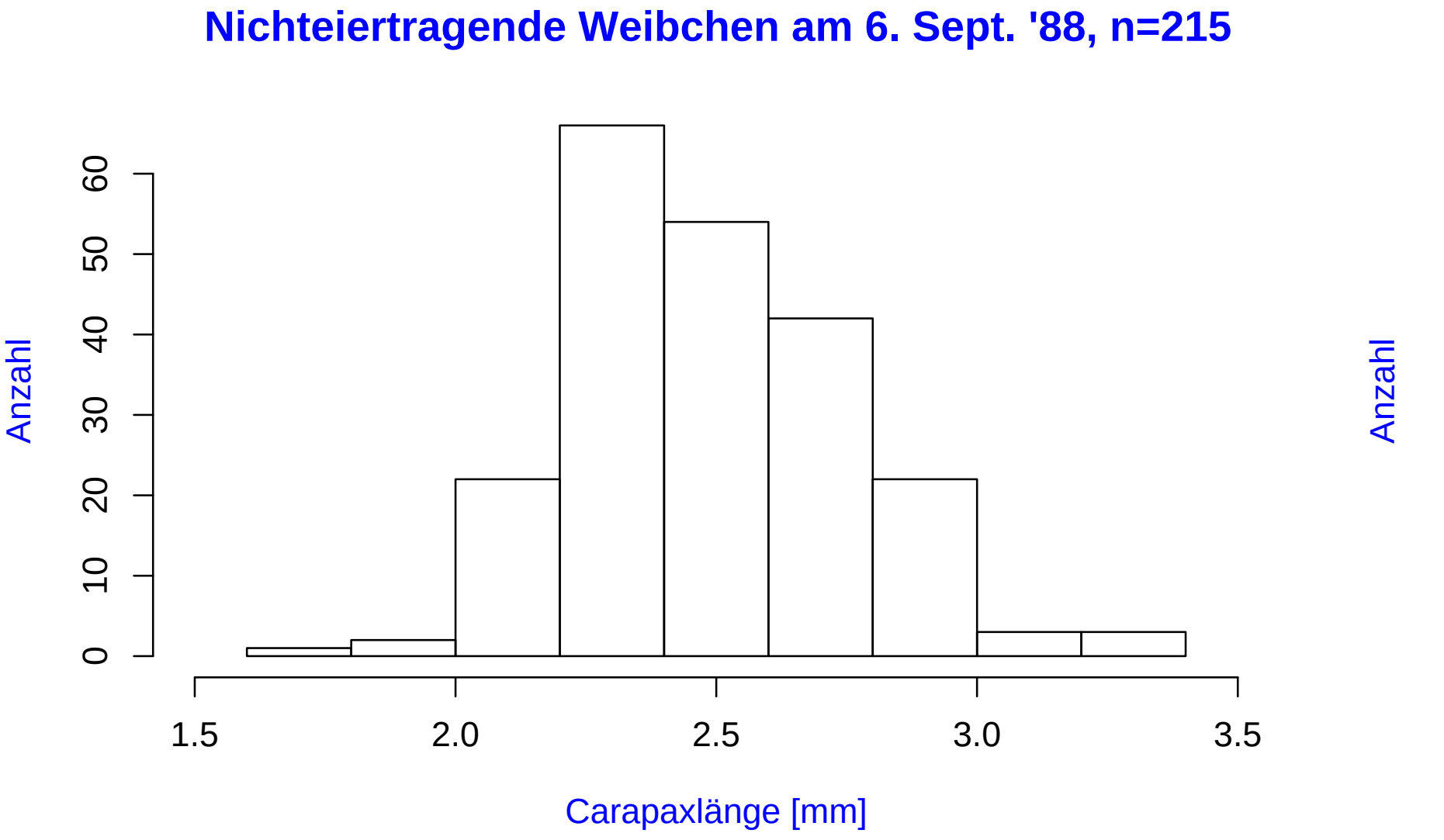
Nichteiertragende Weibchen ($n = 215$)

2,9	3,0	2,9	2,5	2,7	2,9	2,9	3,0
3,0	2,9	3,4	2,8	2,9	2,8	2,8	2,4
2,8	2,5	2,7	3,0	2,9	3,2	3,1	3,0
2,7	2,5	3,0	2,8	2,8	2,8	2,7	3,0
2,6	3,0	2,9	2,8	2,9	2,9	2,3	2,7
2,6	2,7	2,5	.	.	.	.	.

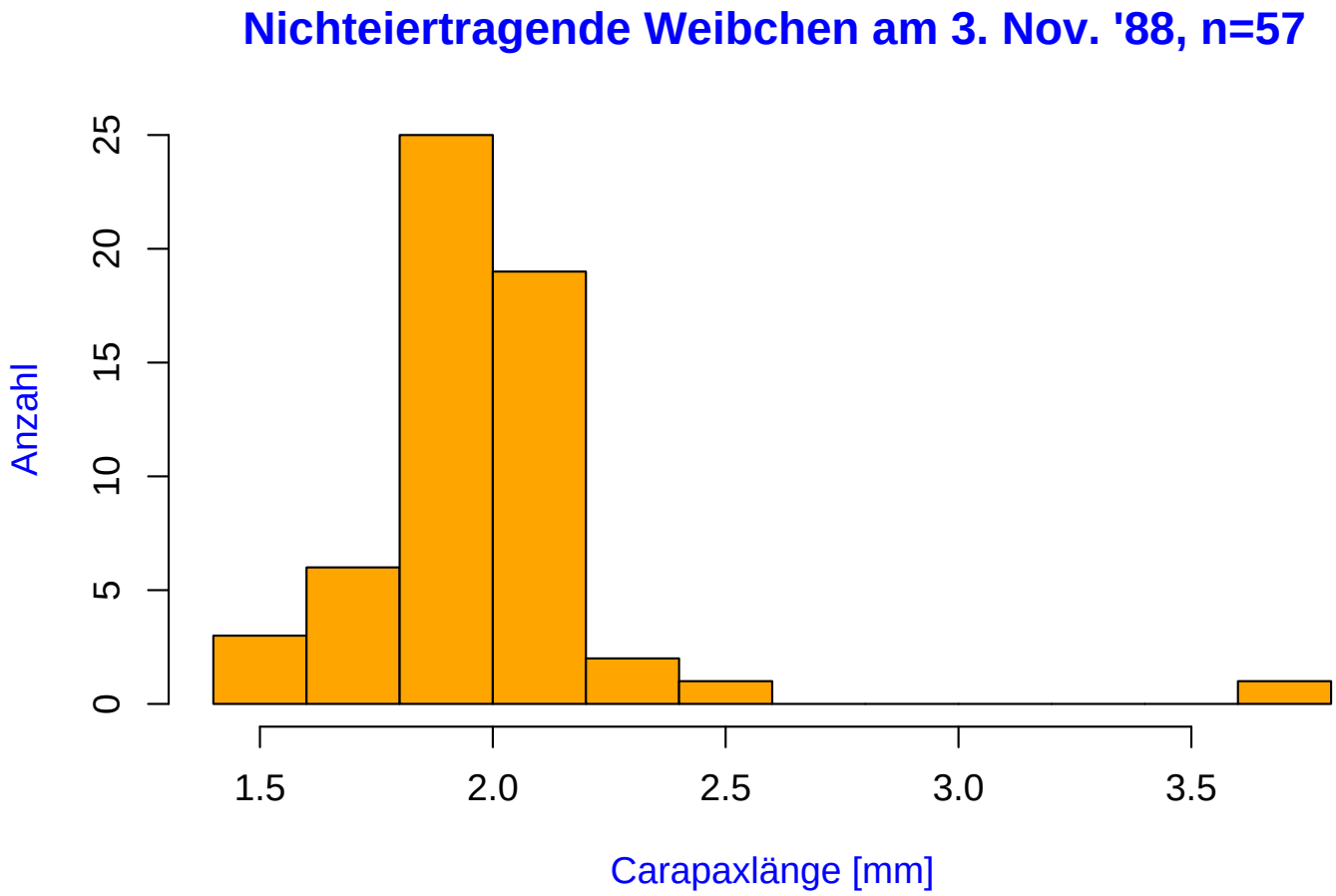

Nichteiertragende Weibchen am 6. Sept. '88, n=215
Carapaxlänge [mm]
2.0
2.5
3.0
0
50
100
150
200
Index

Eine Möglichkeit der graphischen Darstellung:

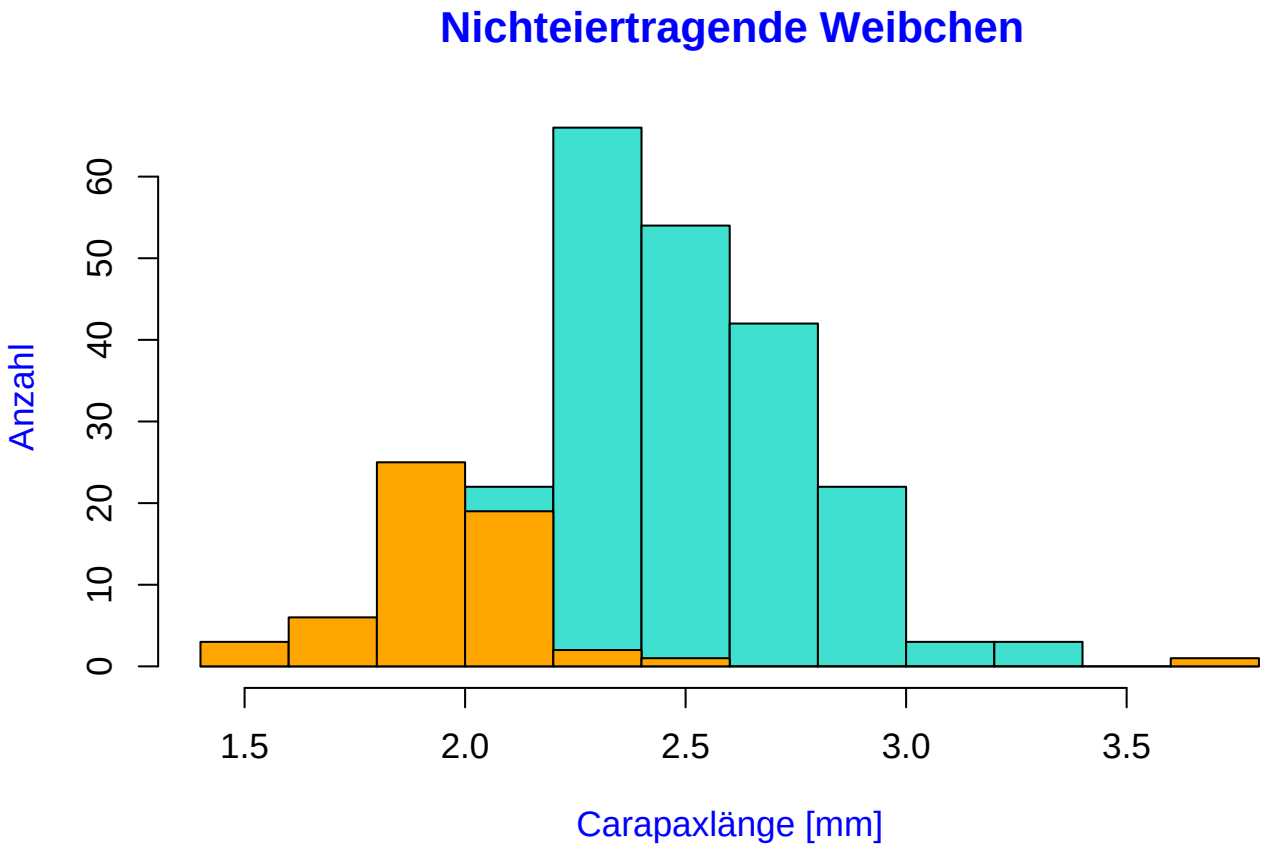
das Histogramm


Nichteiertragende Weibchen am 6. Sept. '88, n=215
Anzahl
0
10
20
30
40
50
60
1.5
2.0
2.5
3.0
3.5
Carapaxlänge [mm]

Analoge Daten zwei Monate später (3.11.88):

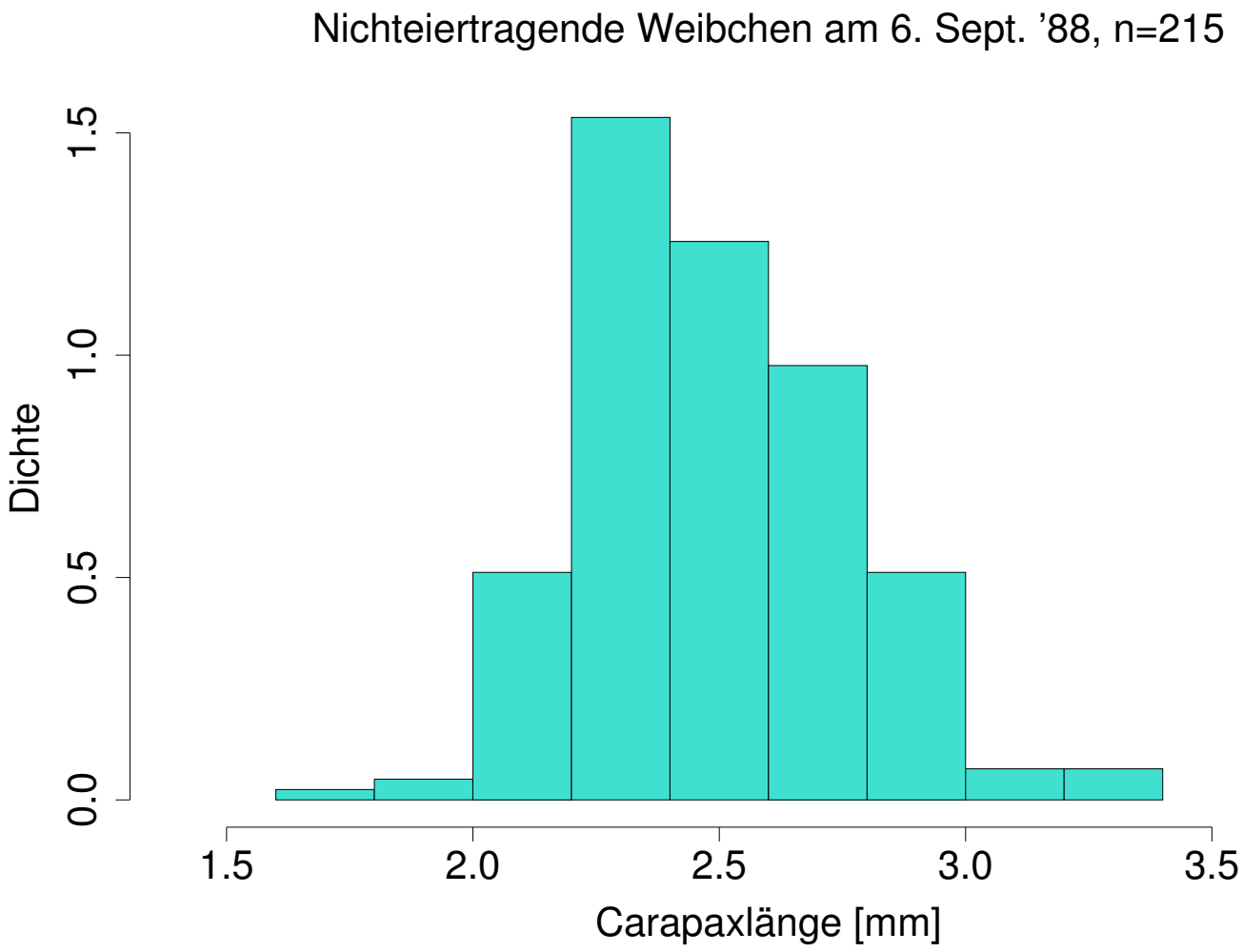

Nichteiertragende Weibchen am 3. Nov. '88, n=57
Anzahl
0
5
10
15
20
25
1.5
2.0
2.5
3.0
3.5
Carapaxlänge [mm]

Vergleich der beiden Verteilungen

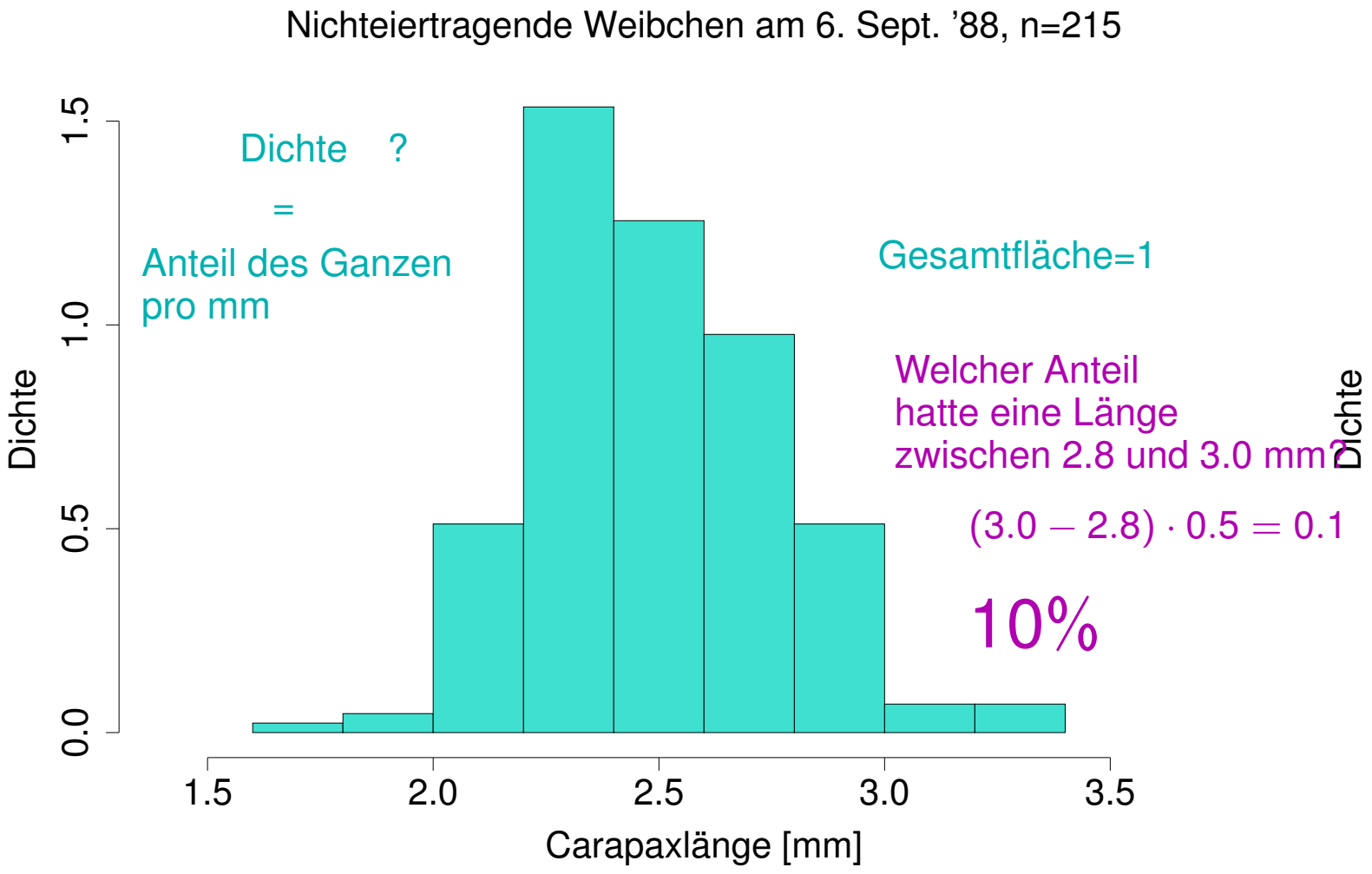


Problem: ungleiche Stichprobenumfänge: 6.Sept: $n = 215$, 3.Nov : $n = 57$

Idee: stauche vertikale Achse so, dass Gesamtfläche = 1.


Nichteiertragende Weibchen am 6. Sept. ’88, n=215
Dichte
0.0
0.5
1.0
1.5
1.5
2.0
2.5
3.0
3.5
Carapaxlänge [mm]

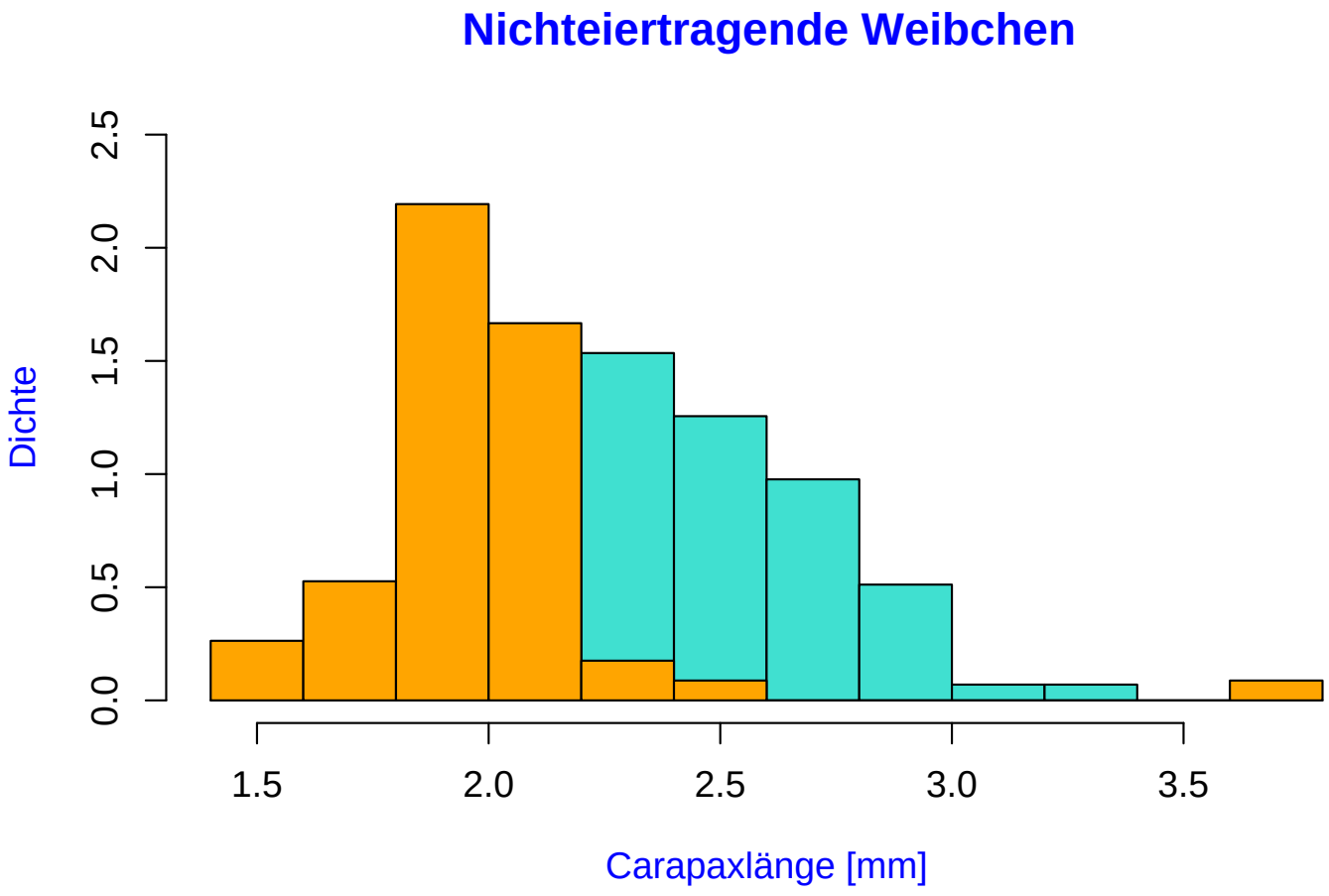
Die neue
vertikale Koordinate
ist jetzt eine
Dichte
(engl. density).

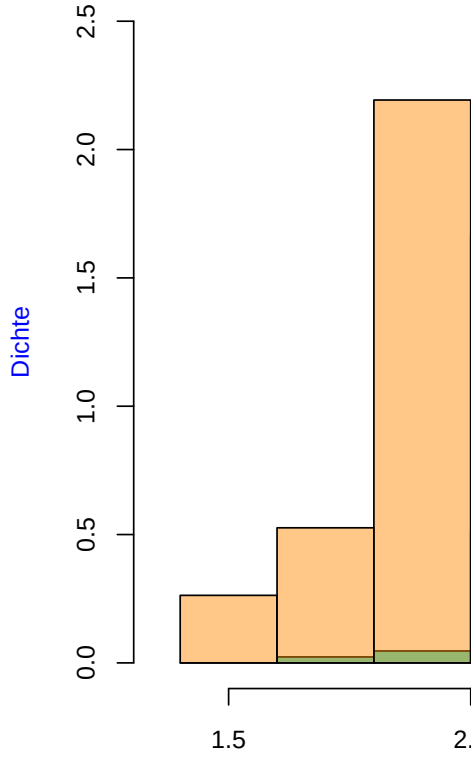

Nichteiertragende Weibchen am 6. Sept. ’88, n=215
Dichte ?
=
Anteil des Ganzen
pro mm
Gesamtfläche=1
Welcher Anteil
hatte eine Länge
zwischen 2.8 und 3.0 mm?
$(3.0 - 2.8) \cdot 0.5 = 0.1$
10%
Dichte
0.0
0.5
1.0
1.5
1.5
2.0
2.5
3.0
3.5
Carapaxlänge [mm]

Die beiden Histogramme sind jetzt vergleichbar

(sie haben dieselbe Gesamtfläche).

Versuche, die Histogramme zusammen zu zeigen:





Vorschlag

Wenn Sie Schauwerbegestalter(in) sind:

Total abgefahrene 3D-Plots können in der Werbung nützlich sein,

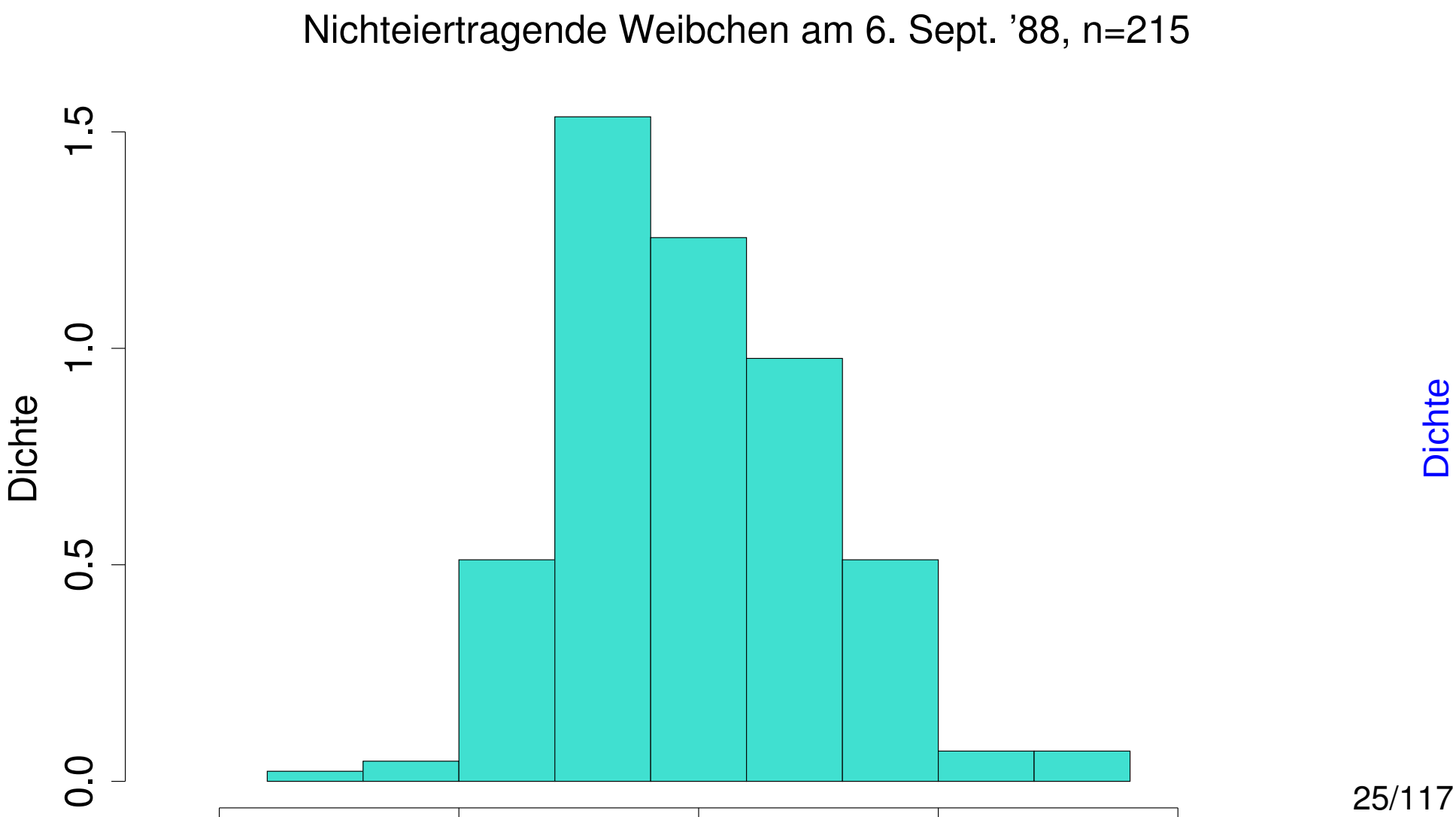
für die Wissenschaft sind einfache und klare 2D-Darstellungen meistens angemessener.

Problem

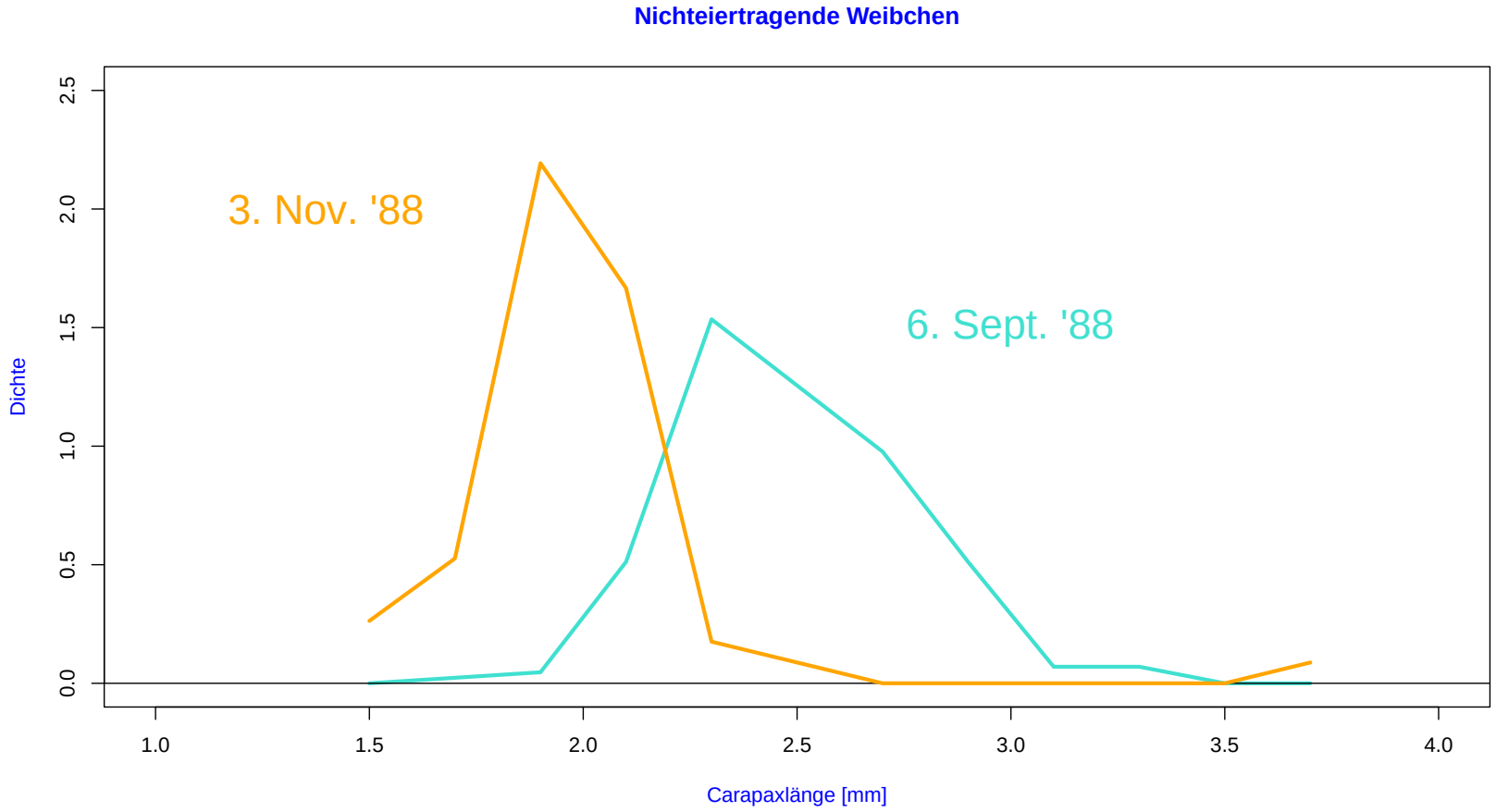
Histogramme kann man nicht ohne weiteres in demselben Graphen darstellen,

weil sie einander überdecken würden.

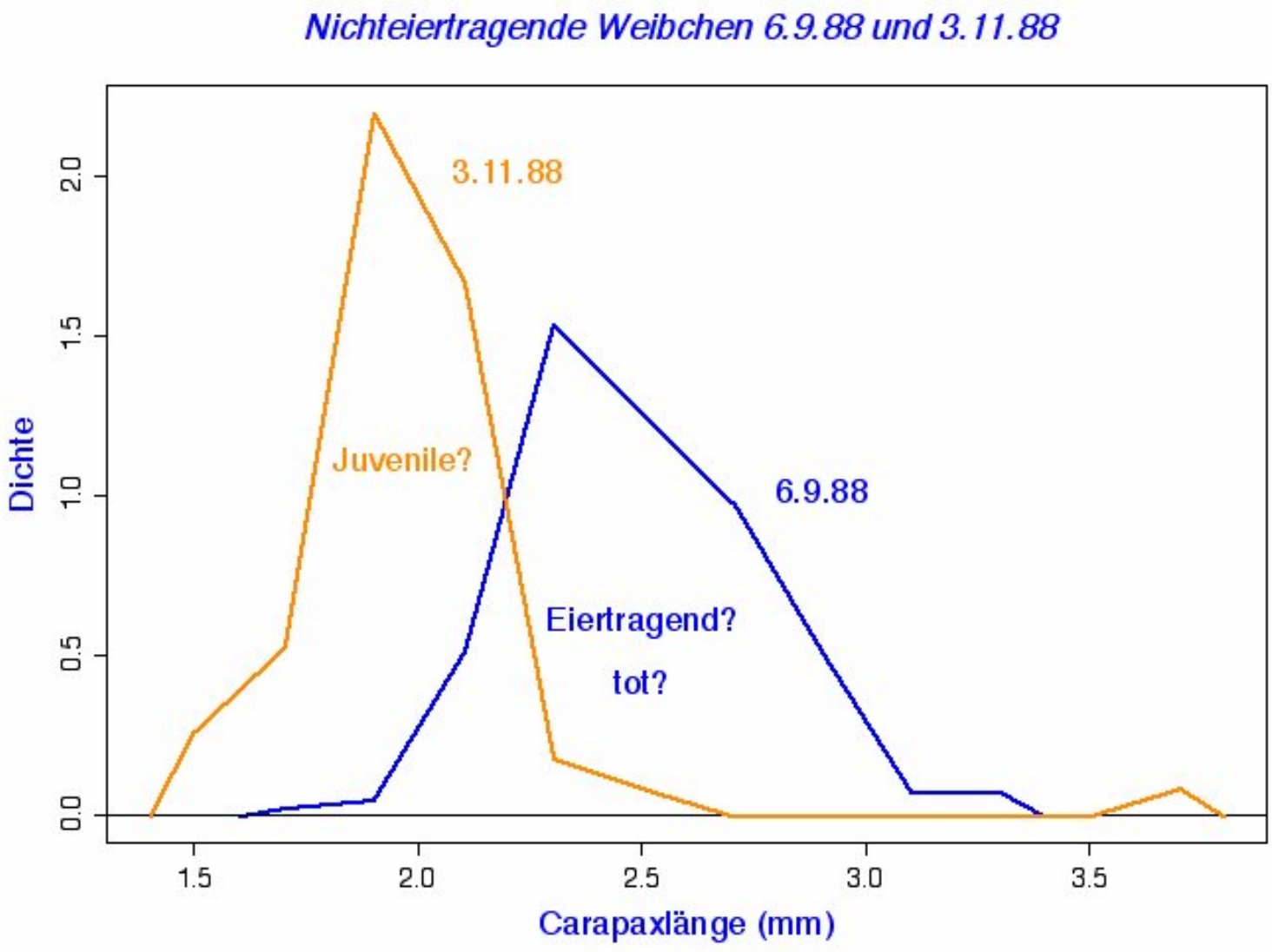
Einfache und klare Lösung: Dichtepolygone



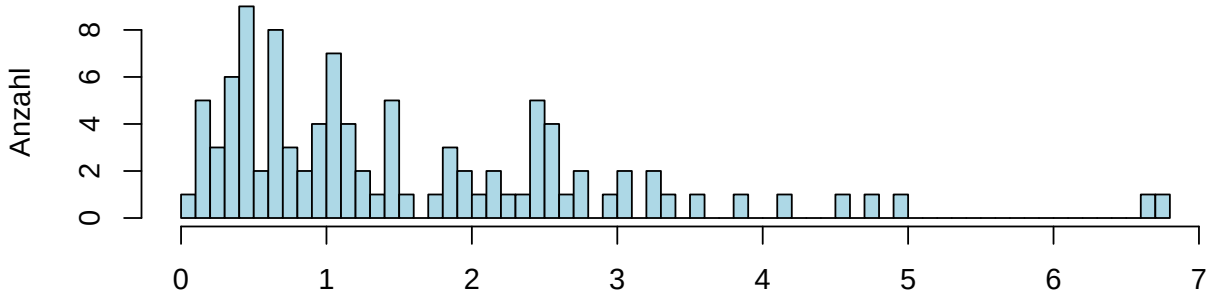
Zwei und mehr Dichtepolygone in einem Plot

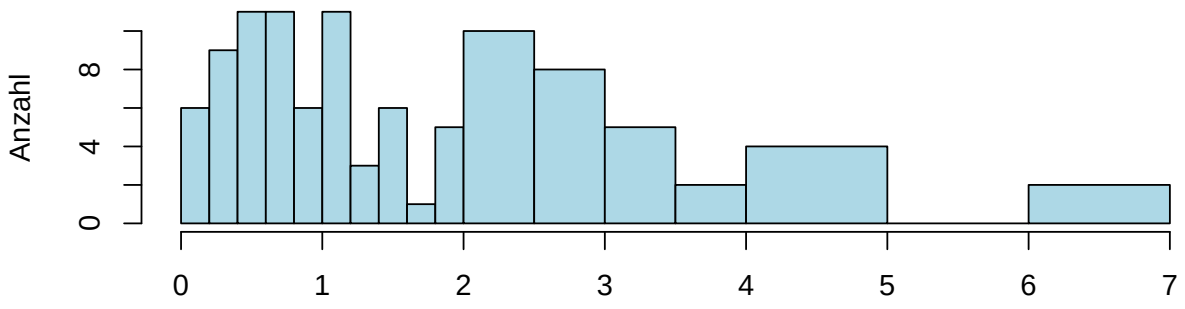


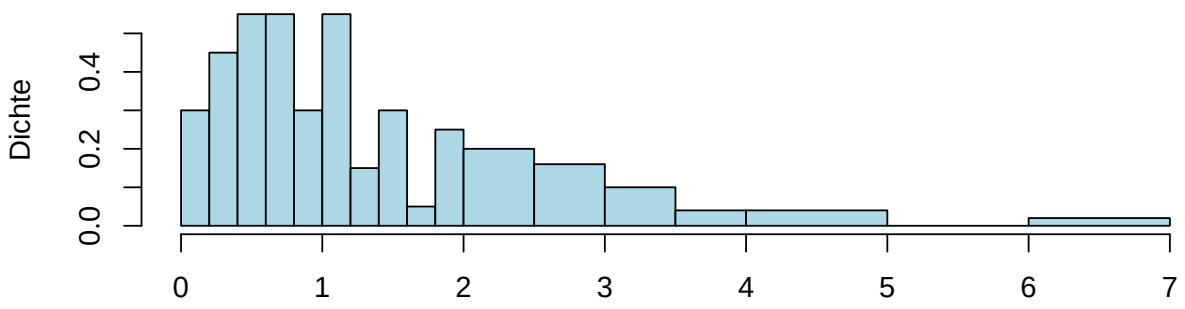
Biologische Interpretation der Verschiebung?


Nichteiertragende Weibchen 6.9.88 und 3.11.88
3.11.88
Juvenile?
6.9.88
Eiertragend?
tot?
Dichte
0.0
0.5
1.0
1.5
2.0
1.5
2.0
2.5
3.0
3.5
Carapaxlänge (mm)

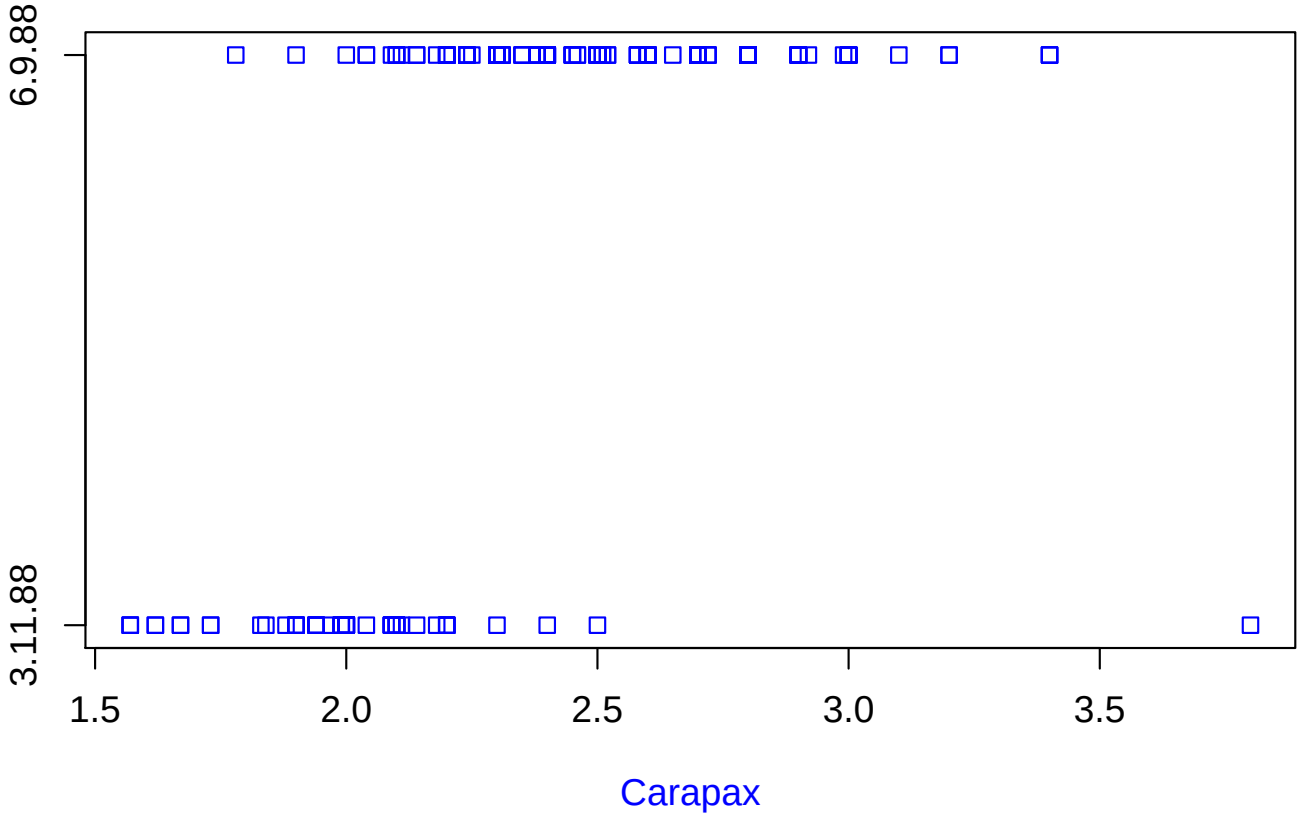
Anzahl vs. Dichte

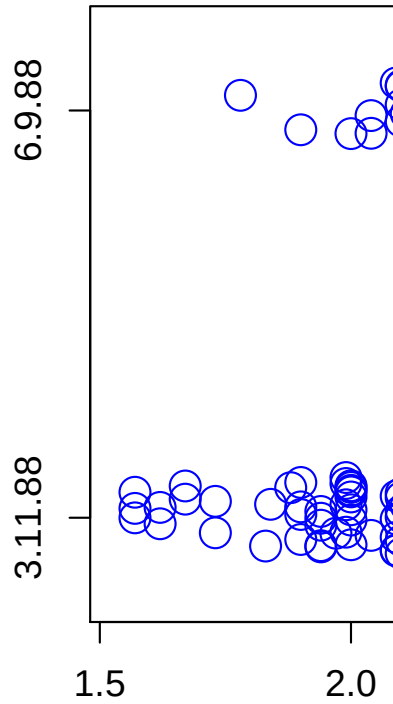


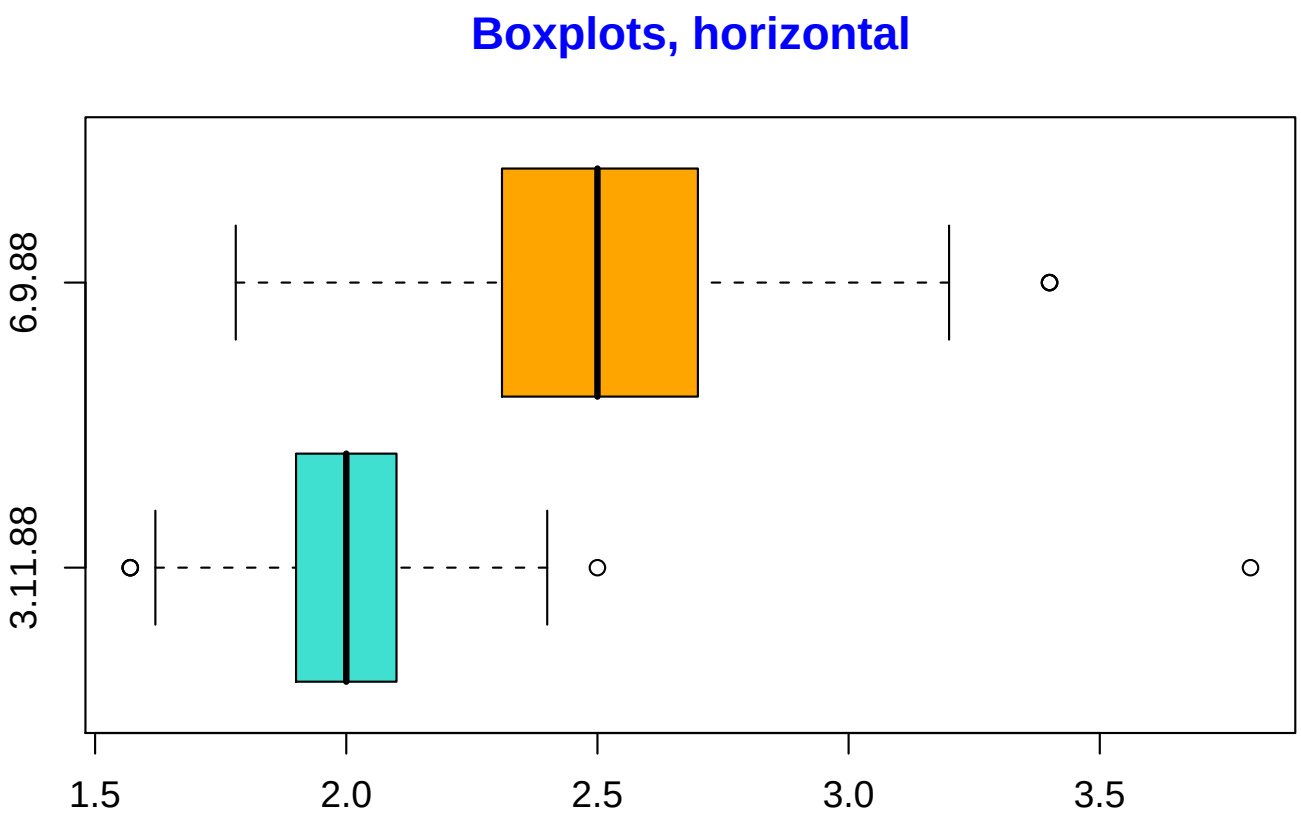


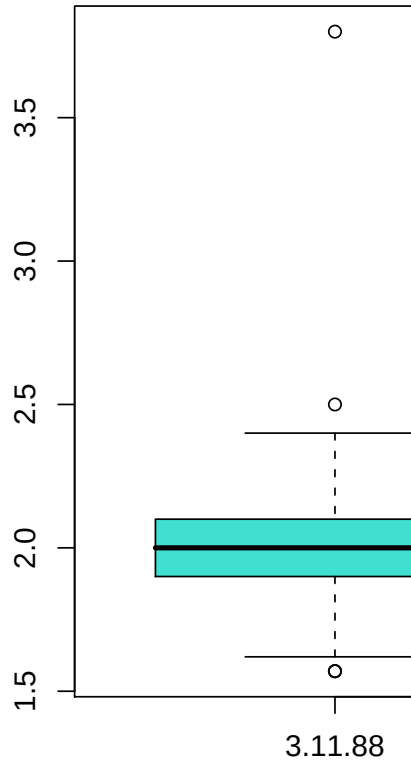


Also: Bei Histogrammen mit ungleichmäßiger Unterteilung immer Dichten verwenden!


6.9.88
3.11.88
1.5
2.0
2.5
3.0
3.5
Carapax


6.9.88
3.11.88
1.5
2.0


Boxplots, horizontal
6.9.88
3.11.88
1.5
2.0
2.5
3.0
3.5


1.5
2.0
2.5
3.0
3.5
3.11.88

Histogramme und Dichtepolygone
geben
ein ausführliches Bild
eines Datensatzes.

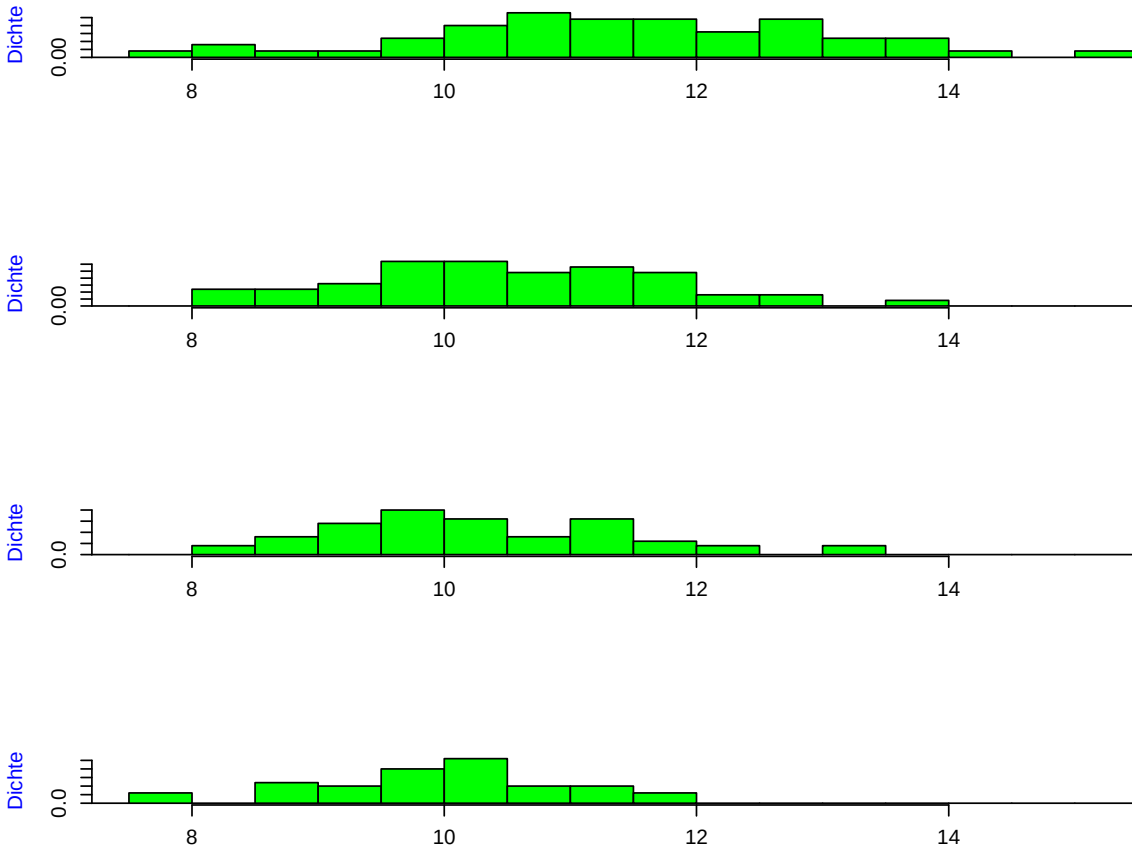
Manchmal zu ausführlich.

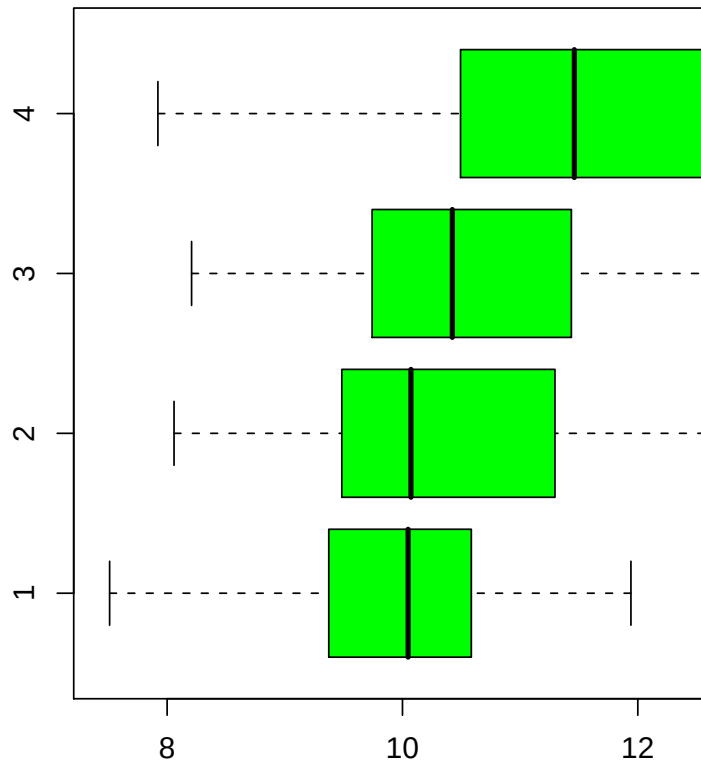
Zu viel Information erschwert den Überblick

Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum Baum

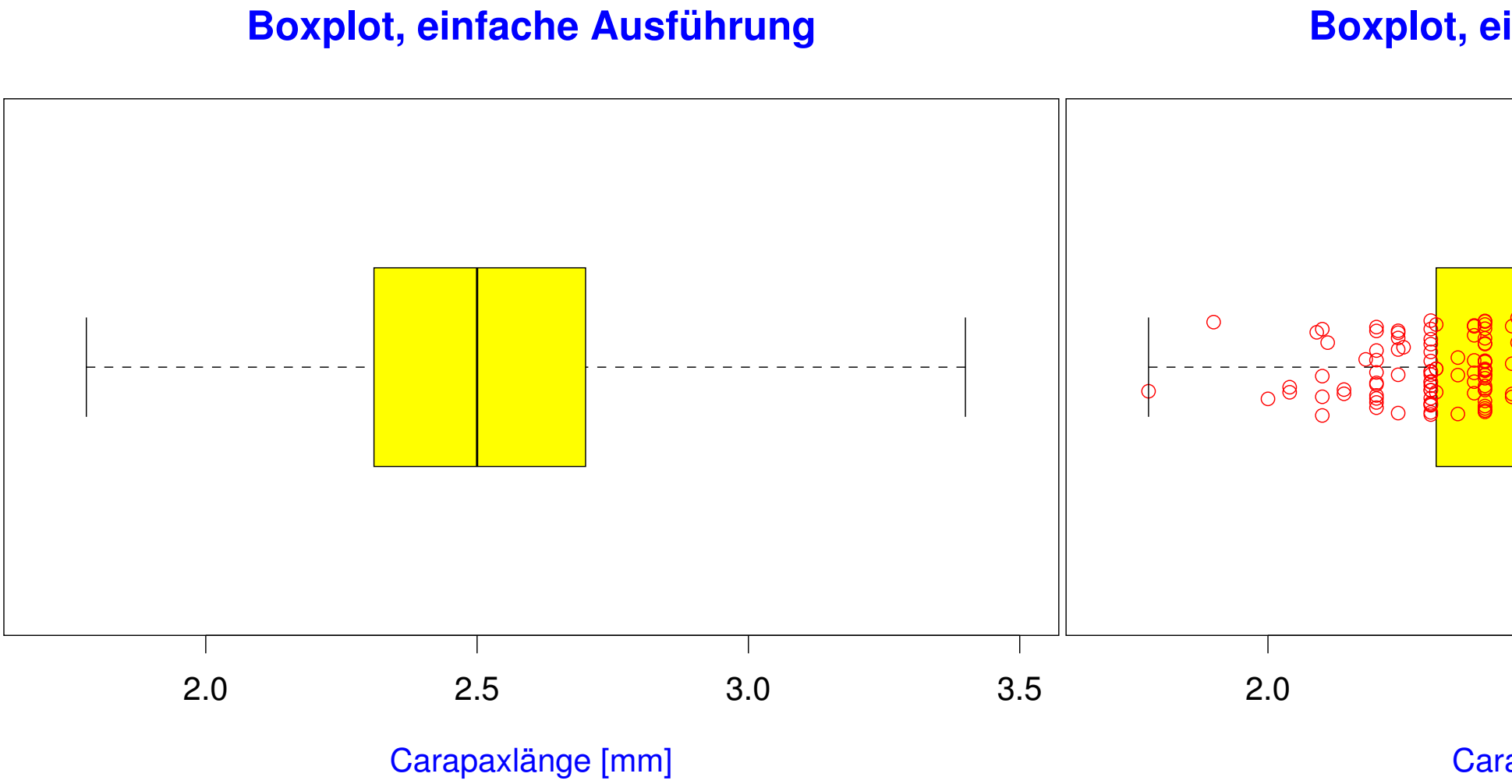
Wald?

Beispiel:
Vergleich von mehreren Gruppen


Dichte
0.00
8
10
12
14
Dichte
0.00
8
10
12
14
Dichte
0.0
8
10
12
14
Dichte
0.0
8
10
12
14


4
3
2
1
8
10
12

Der Boxplot


Boxplot, einfache Ausführung
Boxplot, ei
2.0
2.5
3.0
3.5
2.0
Carapaxlänge [mm]
Cara

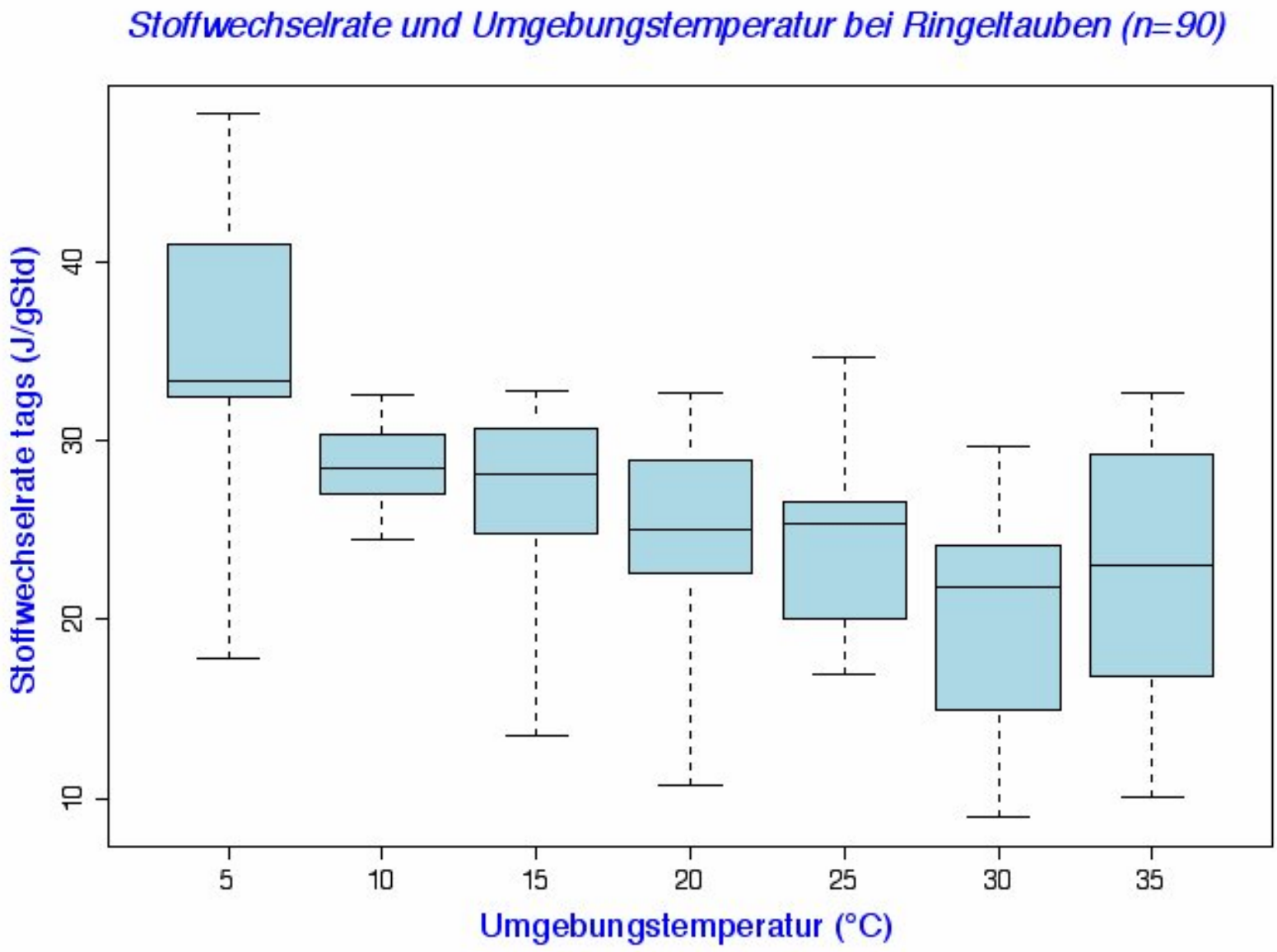
Beispiel:

Die Ringeltaube

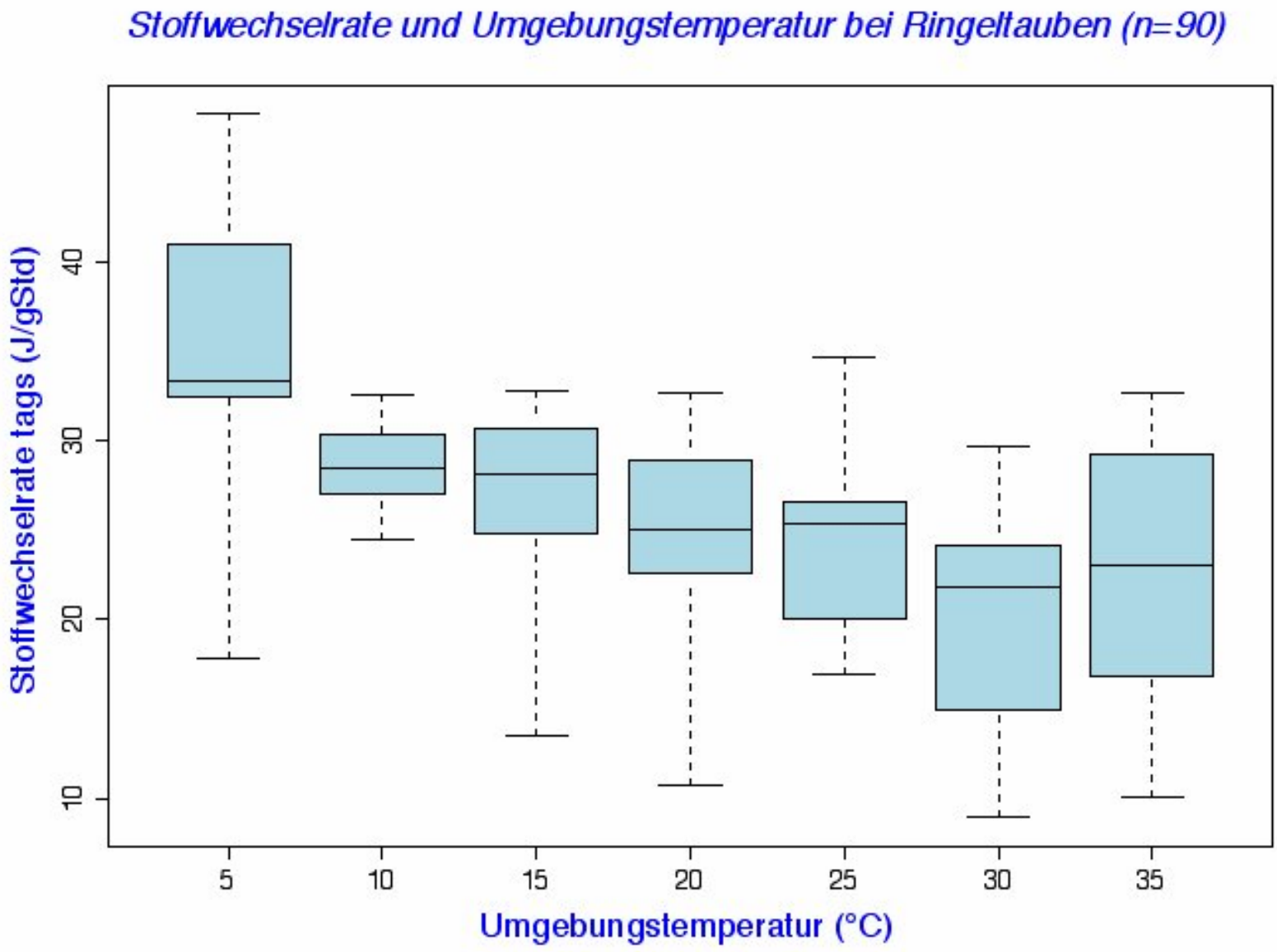
Palumbus palumbus

Wie hängt die Stoffwechselrate bei der Ringeltaube von der Umgebungstemperatur ab?

Daten
aus dem
AK Stoffwechselphysiologie
Prof. Prinzinger
Universität Frankfurt


Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)
Stoffwechselrate tags (J/gStd)
10
20
30
40
5
10
15
20
25
30
35
Umgebungstemperatur (°C)

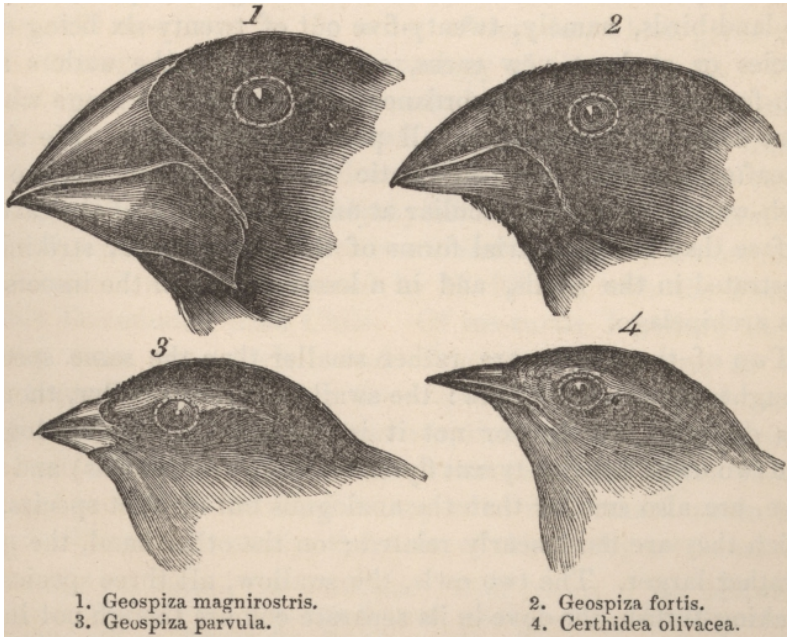
Klar:
Stoffwechselrate
höher
bei
tiefen Temperaturen


Stoffwechselrate und Umgebungstemperatur bei Ringeltauben (n=90)
Stoffwechselrate tags (J/gStd)
10
20
30
40
5
10
15
20
25
30
35
Umgebungstemperatur (°C)

Vermutung:
Bei hohen Temperaturen
nimmt die Stoffwechselrate
wieder zu
(Hitzestress).

Charles Robert Darwin (1809-1882)

Darwin-Finken



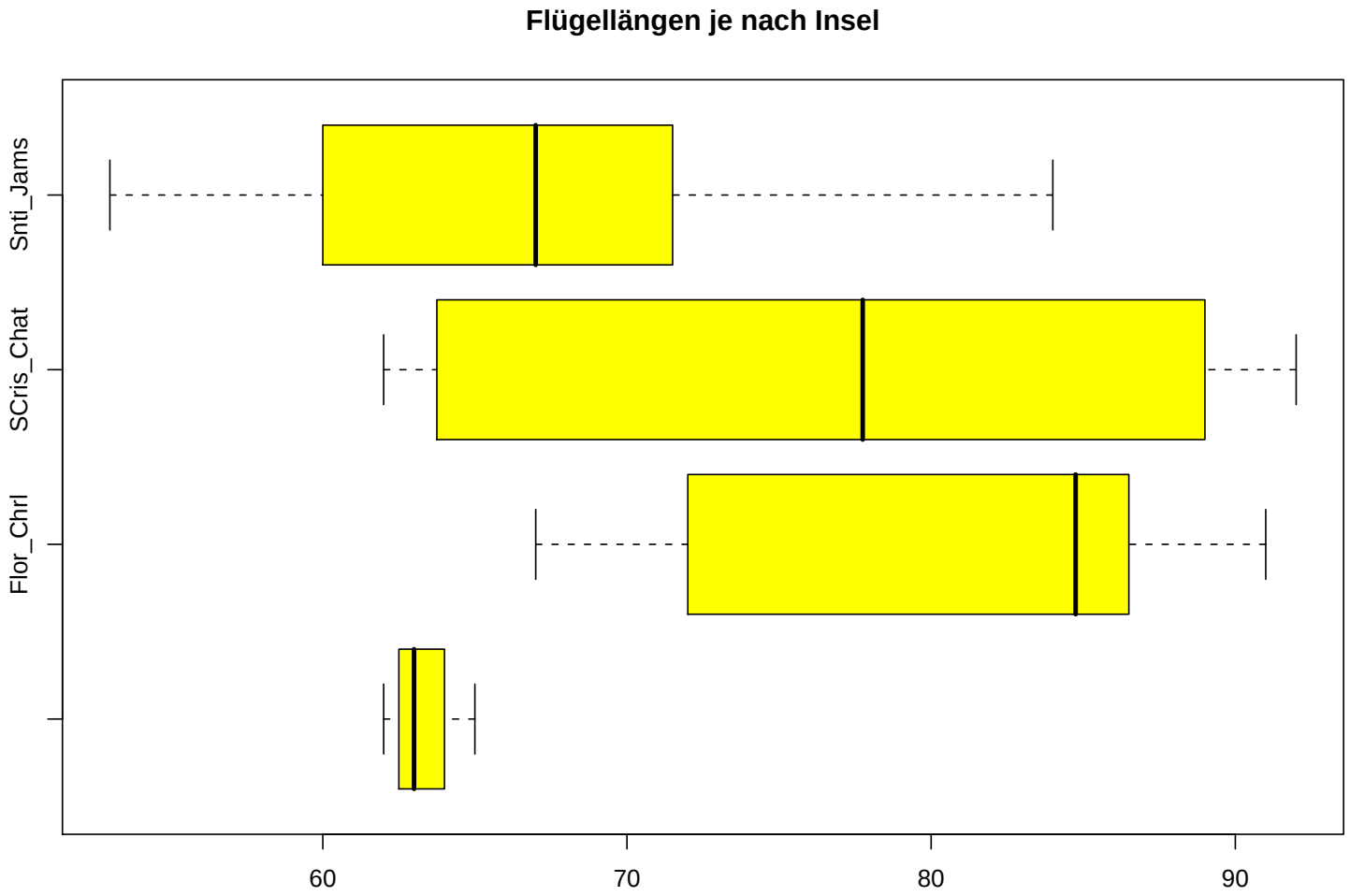
http://darwin-online.org.uk/graphics/Zoology_Illustrations.html

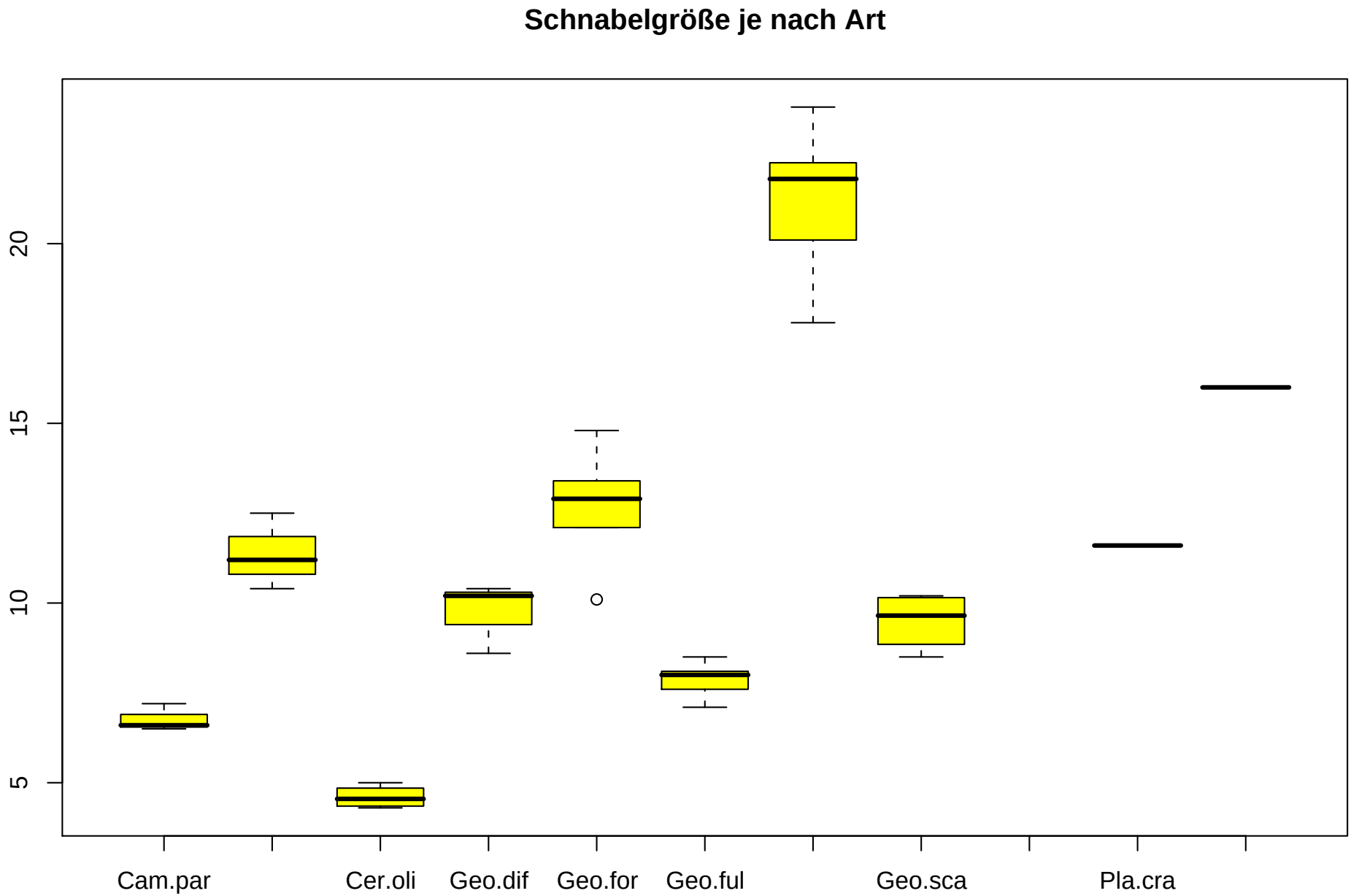
Darwins Finken-Sammlung

Sulloway, F.J. (1982) The Beagle collections of Darwin's Finches (Geospizinae). *Bulletin of the British Museum (Natural History), Zoology series* **43**: 49-94.

- `http://datadryad.org/repo/handle/10255/dryad.154`

Flügellängen der Darwin-Finken




Schnabelgröße je nach Art
5
10
15
20
Cam.par
Cer.oli
Geo.dif
Geo.for
Geo.ful
Geo.sca
Pla.cra

Fazit

1. Histogramme erlauben einen detailierten Blick auf die Daten
2. Dichtepolygone erlauben Vergleiche zwischen vielen Verteilungen
3. Boxplot können große Datenmengen vereinfacht zusammenfassen
4. Bei kleinen Datenmengen eher Stripcharts verwenden
5. Vorsicht mit Tricks wie 3D oder halbtransparenten Farben
6. Jeder Datensatz ist anders; keine Patentrezepte

Es ist oft möglich,
das Wesentliche
an einer Stichprobe

mit ein paar Zahlen
zusammenzufassen.

Wesentlich:

1. Wie groß?

Lageparameter

2. Wie variabel?

Streuungsparameter

Eine Möglichkeit
kennen wir schon
aus dem Boxplot:

Lageparameter

Der Median

Streuungsparameter

Der Quartilabstand $(Q_3 - Q_1)$

Der Median:

die Hälfte der Beobachtungen sind kleiner,
die Hälfte sind größer.

Der Median ist
das 50%-Quantil
der Daten.

Die Quartile

Das erste Quartil, Q_1:
ein Viertel der Beobachtungen
sind kleiner,
drei Viertel sind größer.

Q_1 ist das
25%-Quantil
der Daten.

Die Quartile

Das dritte Quartil, Q_3:
drei Viertel der Beobachtungen
sind kleiner,
ein Viertel sind größer.

Q_3 ist das
75%-Quantil
der Daten.

Am häufigsten werden benutzt:

Lageparameter

Der Mittelwert $\overline{x}$

Streuungsparameter

Die Standardabweichung s

Der Mittelwert

(engl. *mean*)

NOTATION:

Wenn die Beobachtungen

$$x_1, x_2, x_3, \ldots, x_n$$

heißen,

schreibt man oft

$$\overline{x}$$

für den Mittelwert.

DEFINITION:

Mittelwert

=

Der Mittelwert von $x_1, x_2, \ldots, x_n$ als Formel:

$$\begin{aligned} \overline{x} &= (x_1 + x_2 + \cdots + x_n)/n \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i \end{aligned}$$

Beispiel:

$$x_1 = 3,\ x_2 = 0,\ x_3 = 2,\ x_4 = 3,\ x_5 = 1$$

$$\overline{x} = \text{Summe}/\text{Anzahl}$$

$$\overline{x} = (3 + 0 + 2 + 3 + 1)/5$$

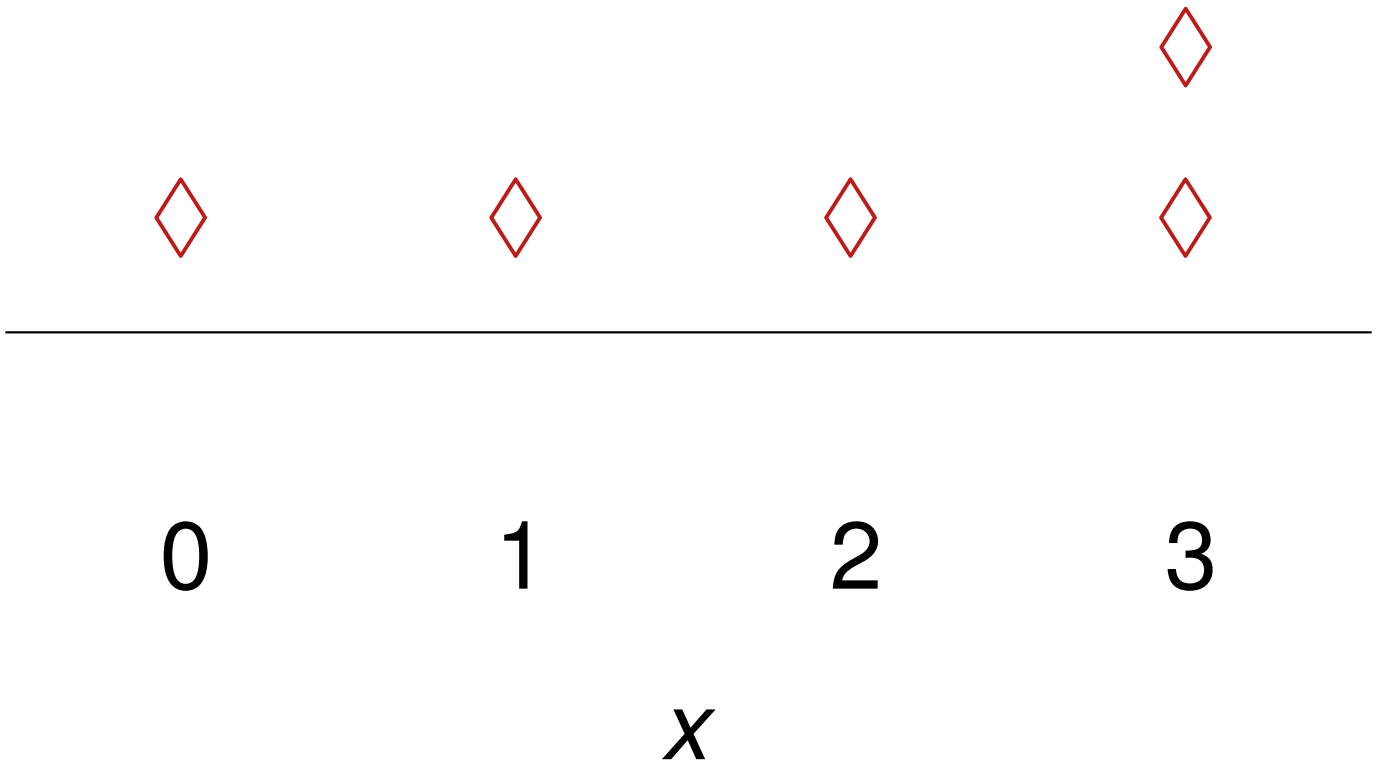
$$\overline{x} = 9/5$$

$$\overline{x} = 1{,}8$$

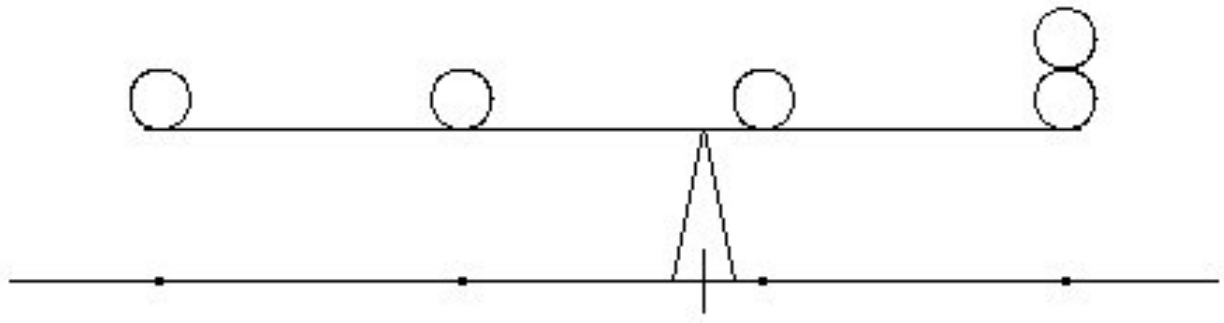
Geometrische Bedeutung des Mittelwerts:

Der Schwerpunkt

Wo muß der Drehpunkt sein, damit die Waage im Gleichgewicht ist?



$m = 1{,}8$?

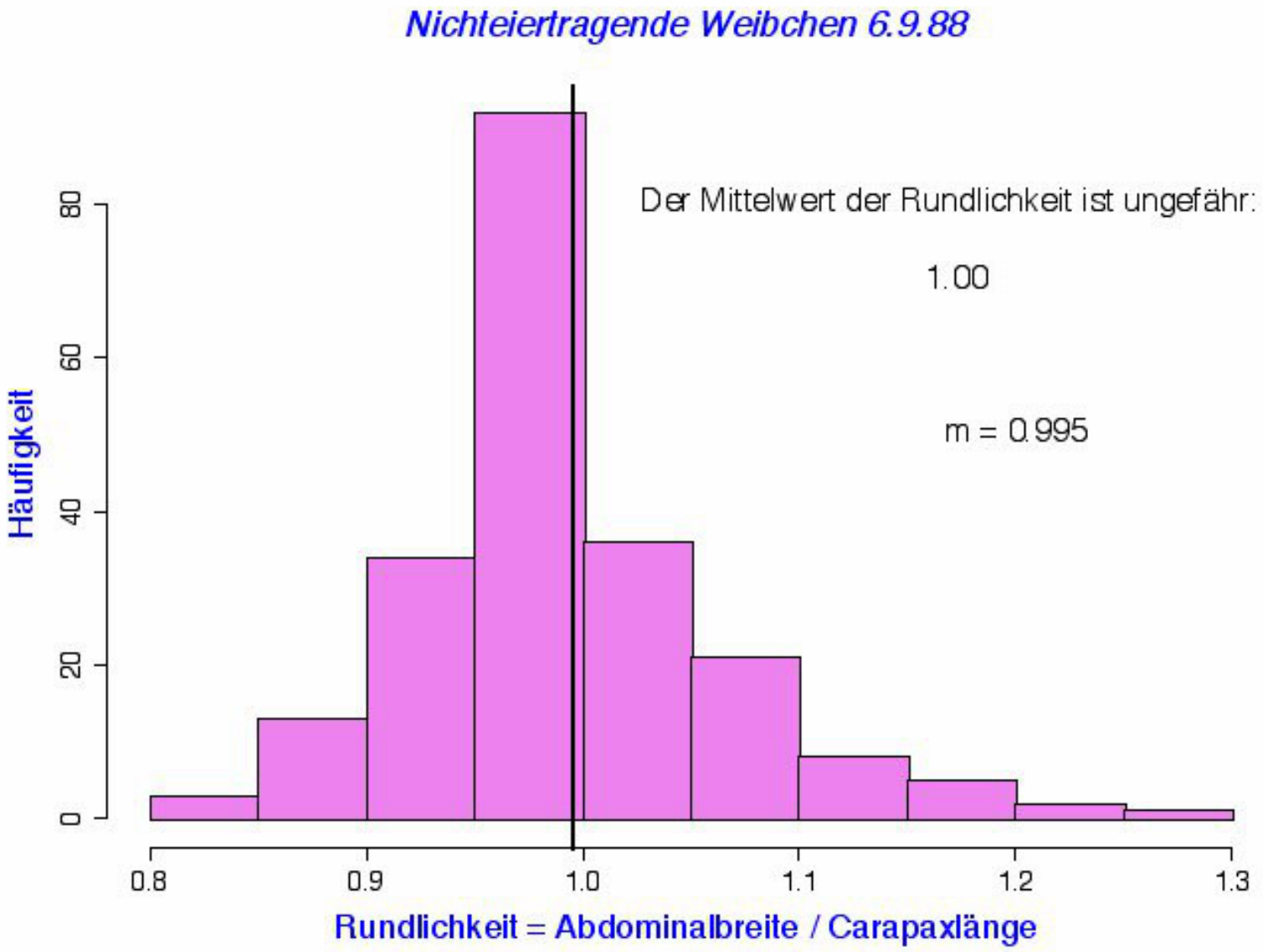


richtig

Beispiel: *Galathea intermedia*

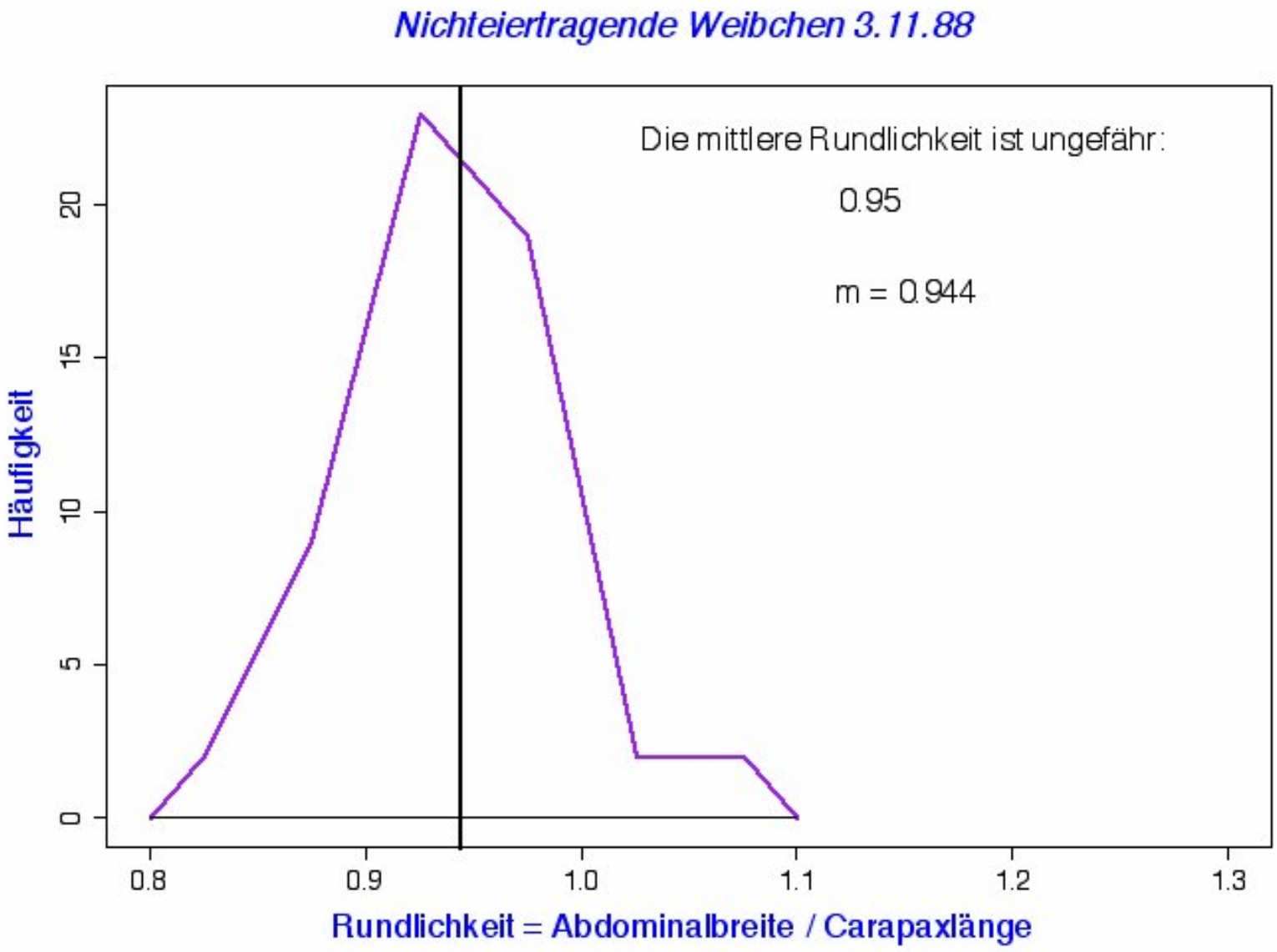
„Rundlichkeit“
:=
Abdominalbreite / Carapaxlänge

Vermutung:
Rundlichkeit nimmt
bei Geschlechtsreife zu


Nichteiertragende Weibchen 6.9.88
Der Mittelwert der Rundlichkeit ist ungefähr:
1.00
m = 0.995
Häufigkeit
0
20
40
60
80
0.8
0.9
1.0
1.1
1.2
1.3
Rundlichkeit = Abdominalbreite / Carapaxlänge

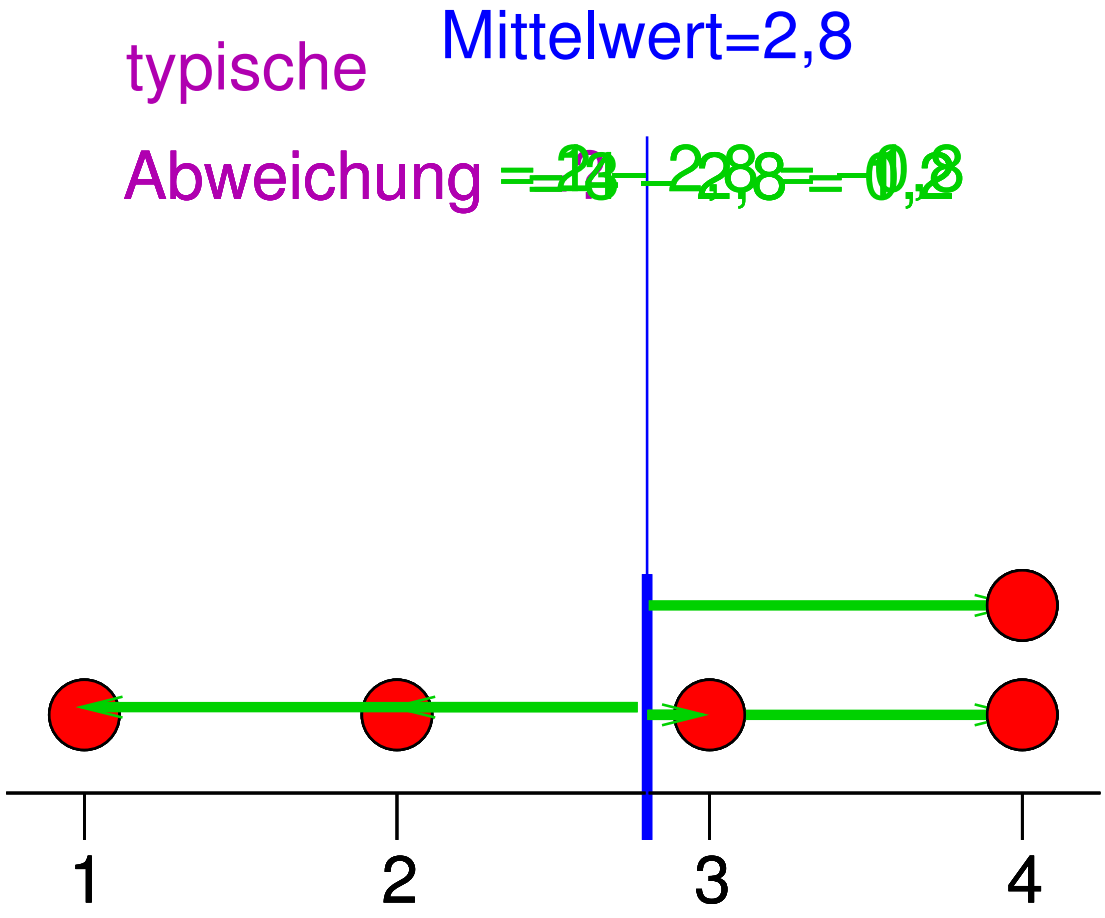
Beispiel:

3.11.88


Nichteiertragende Weibchen 3.11.88
Die mittlere Rundlichkeit ist ungefähr:
0.95
m = 0.944
Häufigkeit
0
5
10
15
20
0.8
0.9
1.0
1.1
1.2
1.3
Rundlichkeit = Abdominalbreite / Carapaxlänge

Die Standardabweichung

Wie weit weicht
eine typische Beobachtung
vom
Mittelwert
ab ?


typische
Mittelwert=2,8
Abweichung
1
2
3
4

Die Standardabweichung σ (“sigma”)
[auch *SD* von engl. *standard deviation*]
ist ein

etwas komisches

gewichtetes Mittel
der Abweichungsbeträge

und zwar

$$\sigma = \sqrt{\text{Summe}(\text{Abweichungen}^2)/n}$$

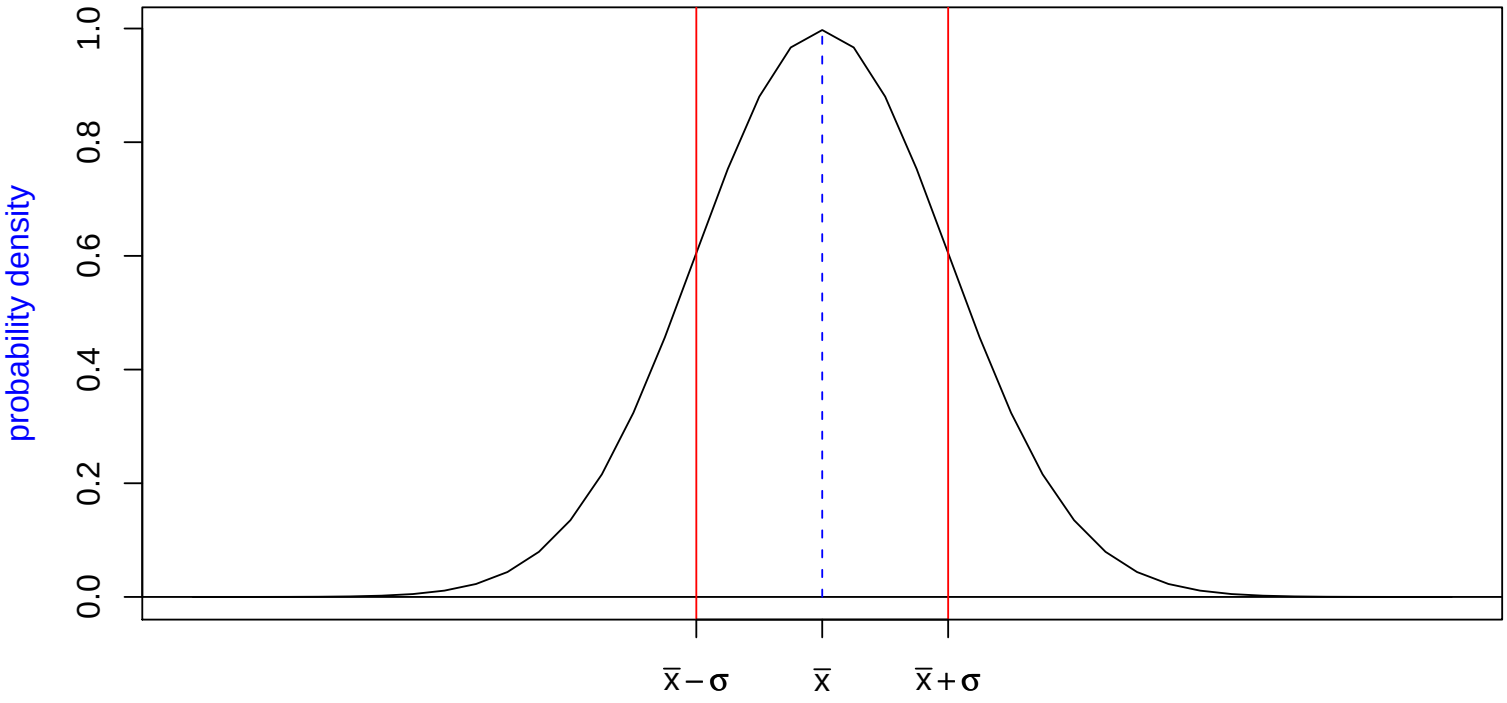
Die Standardabweichung von $x_1, x_2, \ldots, x_n$ als Formel:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

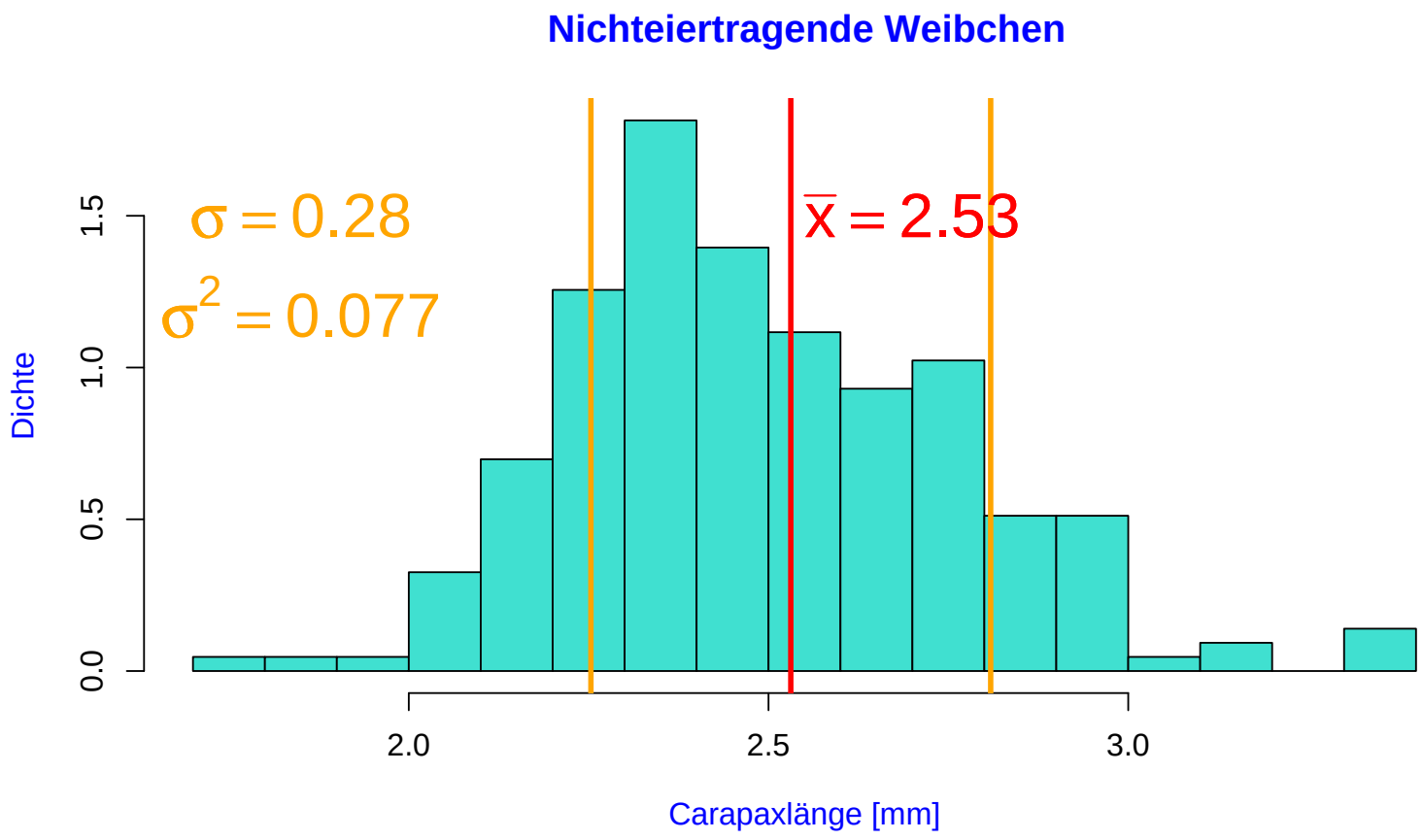
$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ heißt Varianz.

Faustregel für die Standardabweichung

Bei ungefähr glockenförmigen (also eingipfligen und symmetrischen) Verteilungen liegen ca. 2/3 der Verteilung zwischen $\overline{x} - \sigma$ und $\overline{x} + \sigma$.



Standardabweichung der Carapaxlängen nichteiertragender Weibchen vom 6.9.88



Hier liegt der Anteil zwischen $\overline{x} - \sigma$ und $\overline{x} + \sigma$ bei 72%.

Varianz der Carapaxlängen nichteiertragender Weibchen vom 6.9.88

Alle Carapaxlängen im Meer: $\mathcal{X} = (X_1, X_2, \ldots, X_N)$.
Carapaxlängen in unserer Stichprobe: $\mathcal{S} = (S_1, S_2, \ldots, S_{n=215})$
Stichprobenvarianz:

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{215} (S_i - \overline{S})^2 \approx 0{,}0768$$

Können wir 0,0768 als Schätzwert für die Varianz σ_X^2 in der ganzen Population verwenden?
Ja, können wir machen. Allerdings ist σ_S^2 im Durchschnitt um den Faktor $\frac{n-1}{n}$ $(= 214/215 \approx 0,995)$ kleiner als σ_X^2

Varianzbegriffe

Varianz in der Population: $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$

Stichprobenvarianz: $\sigma_S^2 = \frac{1}{n}\sum_{i=1}^{n}(S_i - \overline{S})^2$

korrigierte Stichprobenvarianz:

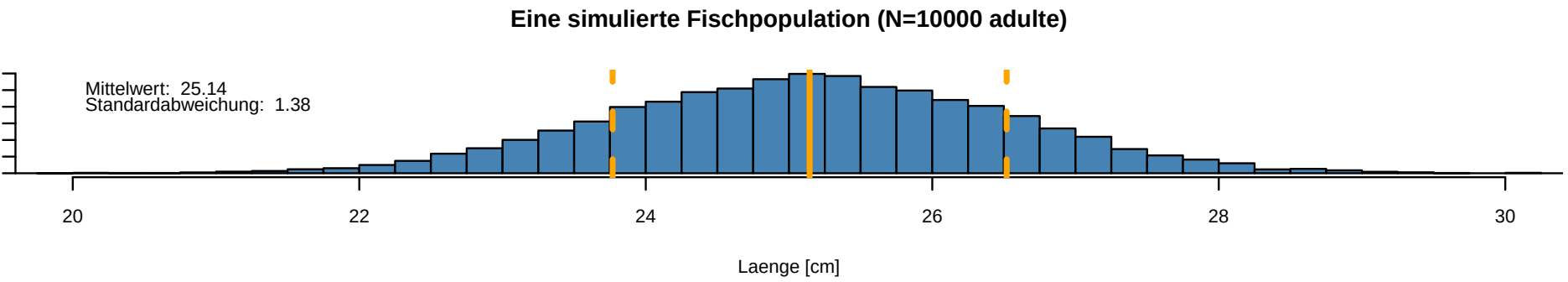
$$\begin{aligned} s^2 &= \frac{n}{n-1}\sigma_S^2 \\ &= \frac{n}{n-1} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n}(S_i - \overline{S})^2 \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^{n}(S_i - \overline{S})^2 \end{aligned}$$

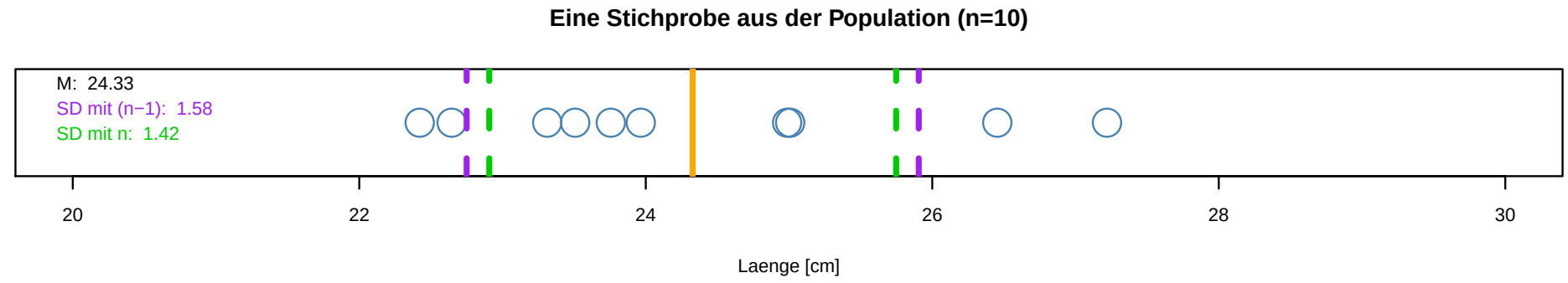
Mit “Standardabweichung von $\mathcal{S}$” ist meistens das korrigierte s gemeint.

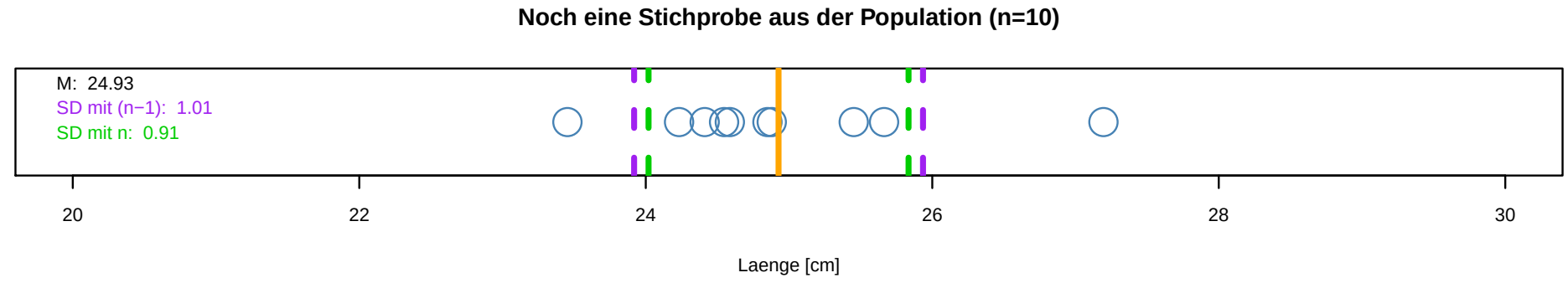
$$\overline{x} = 10/5 = 2$$

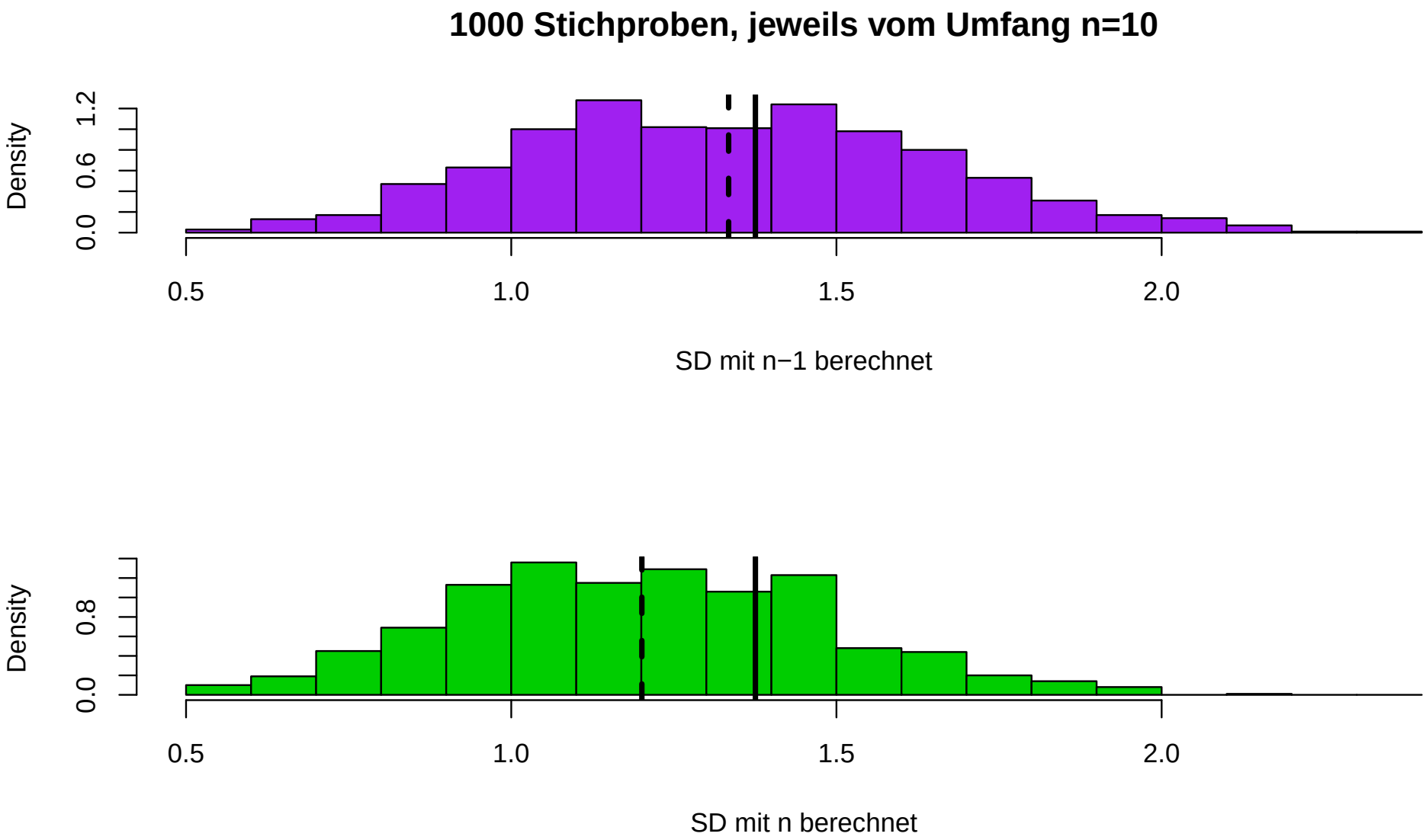
1	1	4	9	1	16

$$\begin{aligned} s^2 &= \text{Summe}\left((x-\overline{x})^2\right)/(n-1) \\ &= 16/(5-1) = 4 \\ s &= 2 \end{aligned}$$


Eine simulierte Fischpopulation (N=10000 adulte)
Mittelwert: 25.14
Standardabweichung: 1.38
20
22
24
26
28
30
Laenge [cm]


Eine Stichprobe aus der Population (n=10)
M: 24.33
SD mit (n−1): 1.58
SD mit n: 1.42
20
22
24
26
28
30
Laenge [cm]


Noch eine Stichprobe aus der Population (n=10)
M: 24.93
SD mit (n−1): 1.01
SD mit n: 0.91
20
22
24
26
28
30
Laenge [cm]


1000 Stichproben, jeweils vom Umfang n=10
Density
0.0
0.6
1.2
0.5
1.0
1.5
2.0
SD mit n−1 berechnet
Density
0.0
0.8
0.5
1.0
1.5
2.0
SD mit n berechnet

σ mit n oder $n-1$ berechnen?

Die Standardabweichung σ eines Zufallsexperiments mit n gleichwahrscheinlichen Ausgängen $x_1, \ldots, x_n$ (z.B. Würfelwurf) ist klar definiert durch

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\overline{x}-x_i)^2}.$$

Wenn es sich bei $x_1, \ldots, x_n$ um eine Stichprobe handelt (wie meistens in der Statistik), sollten Sie die Formel

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\overline{x}-x_i)^2}$$

verwenden.

Mittelwert und Standardabweichung. . .

- charakterisieren die Daten gut, falls deren Verteilung glockenförmig ist
- und müssen andernfalls mit Vorsicht interpretiert werden.

Wir betrachten dazu einige Lehrbuch-Beispiele aus der Ökologie, siehe z.B.

M. Begon, C. R. Townsend, and J. L. Harper.
Ecology: From Individuals to Ecosystems.
Blackell Publishing, 4 edition, 2008.

Im Folgenden verwenden wir zum Teil simulierte Daten, wenn die Originaldaten nicht verfügbar waren.
(Nehmen Sie also nicht alle Datenpunkte wörtlich.)

Bachstelzen fressen Dungfliegen

Räuber

Bachstelze (White Wagtail)
Motacilla alba alba

image (c) by Artur Mikołajewski

Beute

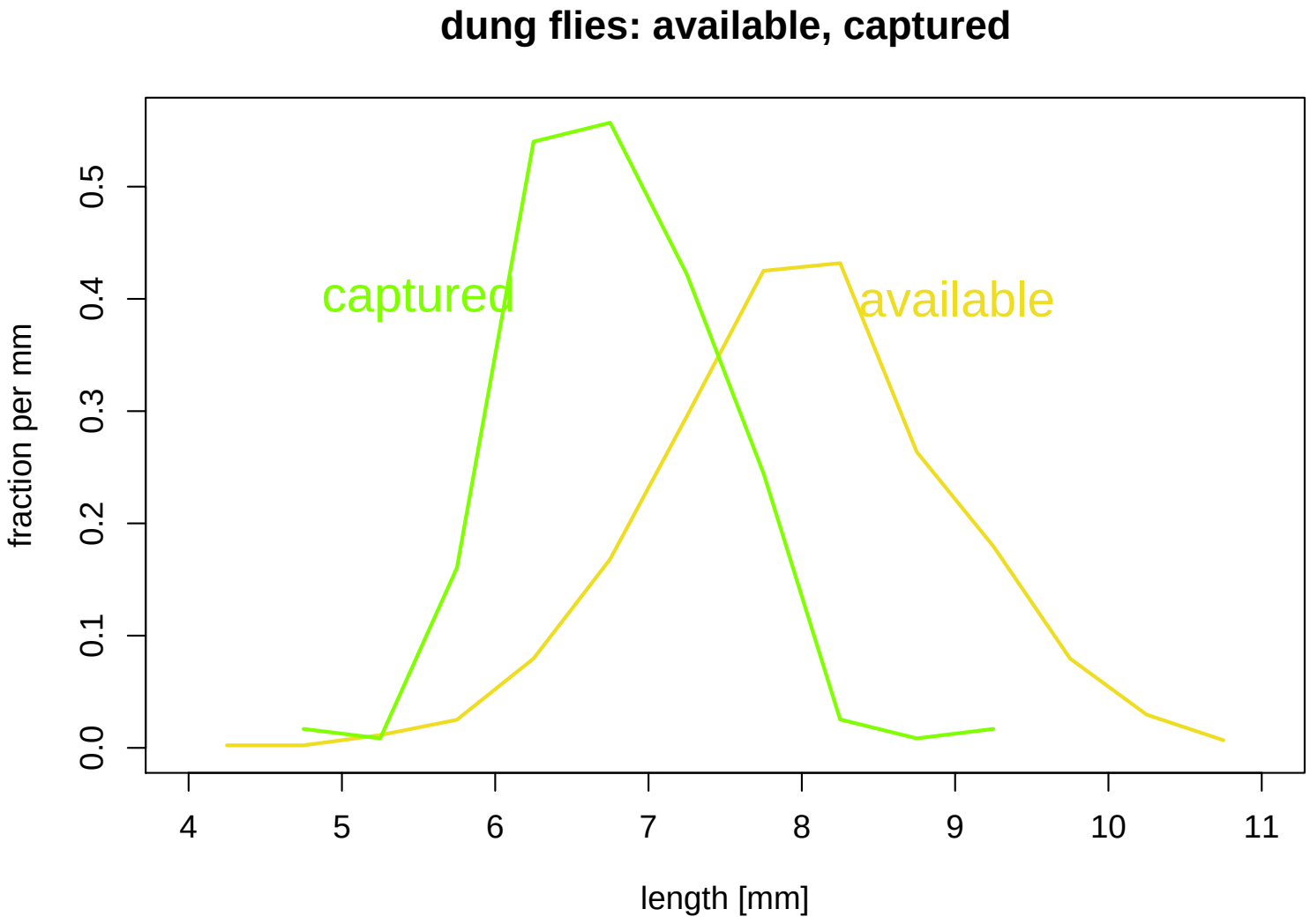
Gelbe Dungfliege
Scatophaga stercoraria

image (c) by Viatour Luc

Vermutung

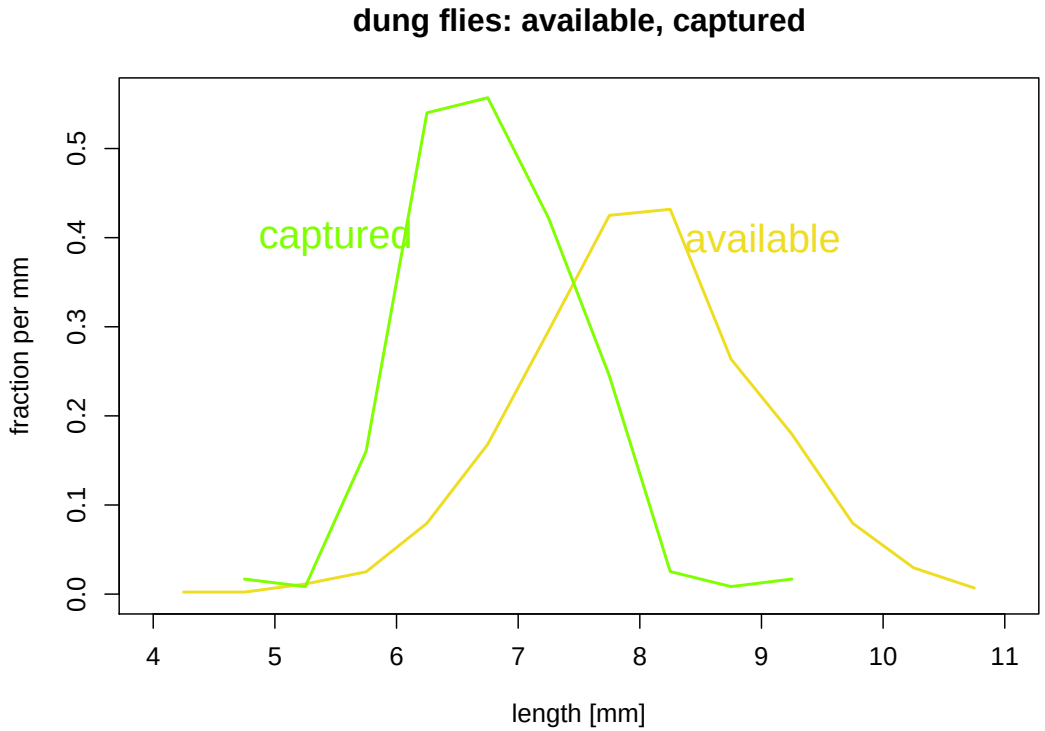
- Die Fliegen sind unterschiedlich groß
- Effizienz für die Bachstelze = Energiegewinn / Zeit zum Fangen und fressen
- Laborexperimente lassen vermuten, dass die Effizienz bei 7mm großen Fliegen maximal ist.

N.B. Davies.
Prey selection and social behaviour in wagtails (Aves: Motacillidae).
J. Anim. Ecol., 46:37–57, 1977.


dung flies: available, captured
captured
available
fraction per mm
0.0
0.1
0.2
0.3
0.4
0.5
4
5
6
7
8
9
10
11
length [mm]

Vergleich der Größenverteilungen

	captured		available
Mittelwert	6.29	$<$	7.99
Standardabweichung	0.69	$<$	0.96



Interpretation

Die Bachstelzen bevorzugen Dungfliegen, die etwa 7mm groß sind.

Hier waren die Verteilungen glockenförmig und es genügten 4 Werte (die beiden Mittelwerte und die beiden Standardabweichungen), um die Daten adäquat zu beschreiben.

Nephila madagascariensis

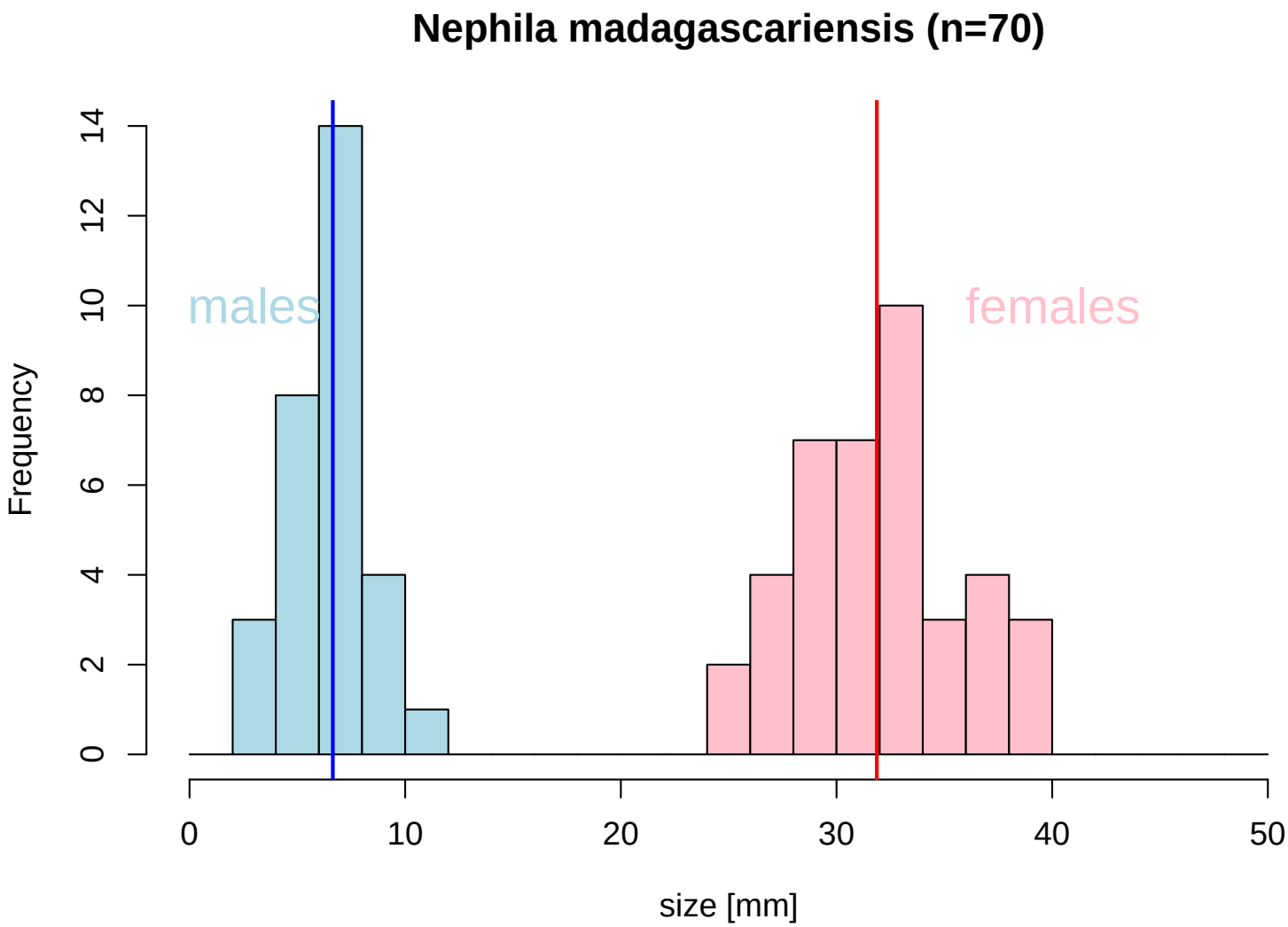
image (c) by Bernard Gagnon

Simulated Data:

Eine Stichprobe von 70 Spinnen

Mittlere Größe: 21,06 mm

Standardabweichung der Größe: 12,94 mm


Nephila madagascariensis (n=70)
males
females
Frequency
0
2
4
6
8
10
12
14
0
10
20
30
40
50
size [mm]

Nephila madagascariensis

image (c) by Arthur Chapman

Fazit des Spinnenbeispiels

Wenn die Daten aus verschiedenen Gruppen zusammengesetzt sind, die sich bezüglich des Merkmals deutlich unterscheiden, kann es sinnvoll sein, Kenngrößen wie den Mittelwert für jede Gruppe einzeln zu berechnen.

Kupfertolerantes Rotes Straußgras

Rotes Straußgras
Agrostis tenuis

image (c) Kristian Peters

Kupfer
Cuprum

Hendrick met de Bles

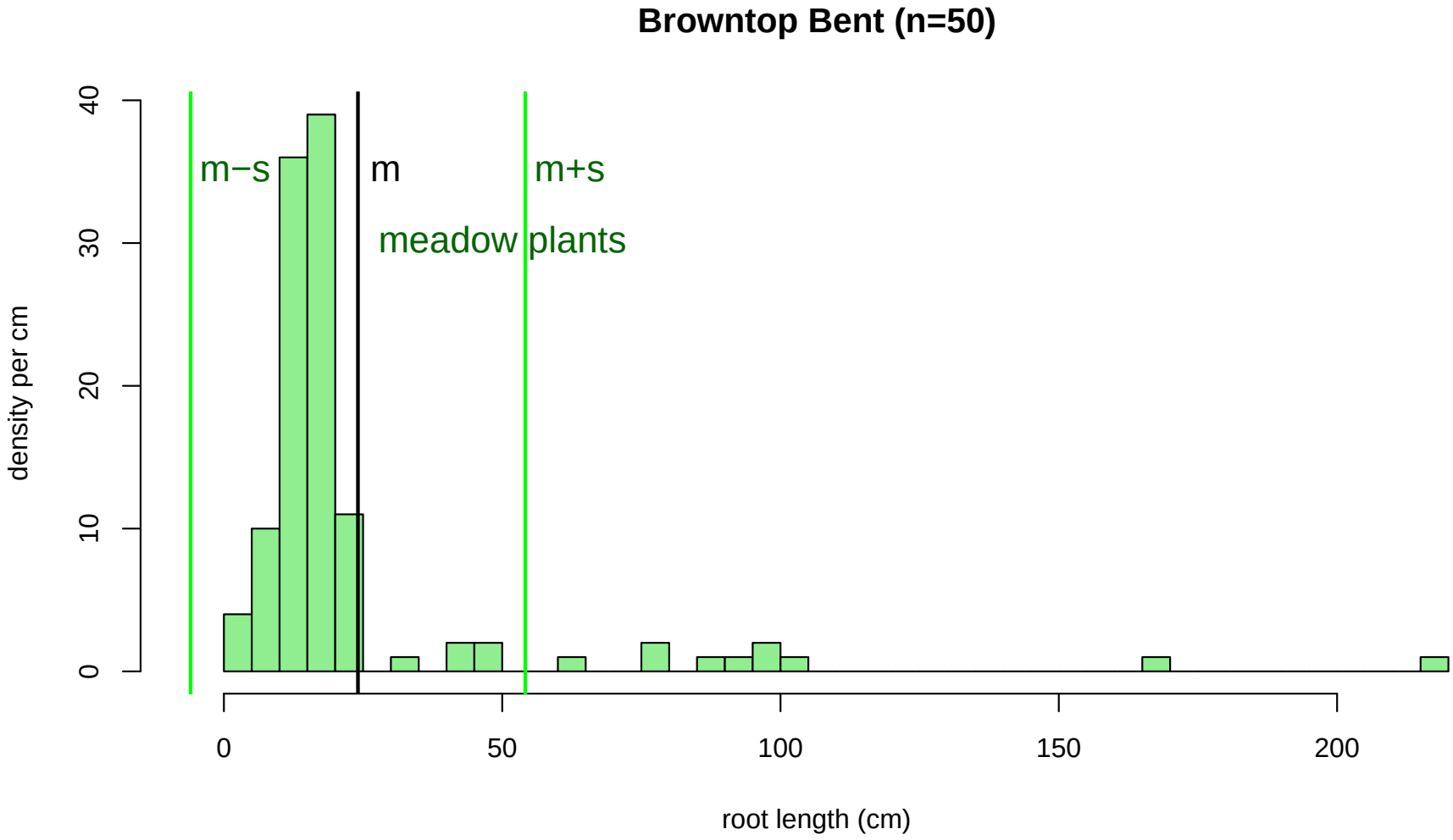

A.D. Bradshaw.
Population Differentiation in *agrostis tenius Sibth*. III. populations in varied environments.
New Phytologist, 59(1):92 – 103, 1960.


T. McNeilly and A.D Bradshaw.
Evolutionary Processes in Populations of Copper Tolerant Agrostis tenuis Sibth.
Evolution, 22:108–118, 1968.

Wir verwenden hier wieder simulierte Daten, da die Originaldaten nicht zur Verfügung stehen.

Anpassung an Kupfer?

- Pflanzen, denen das Kupfer schadet, haben kürzere Wurzeln.
- Die Wurzellängen von Pflanzen aus der Umgebung von Kupferminen wird gemessen.
- Samen von unbelasteten Wiesen werden bei Kupferminen eingesät.
- Die Wurzellängen dieser “Wiesenpflanzen” werden gemessen.

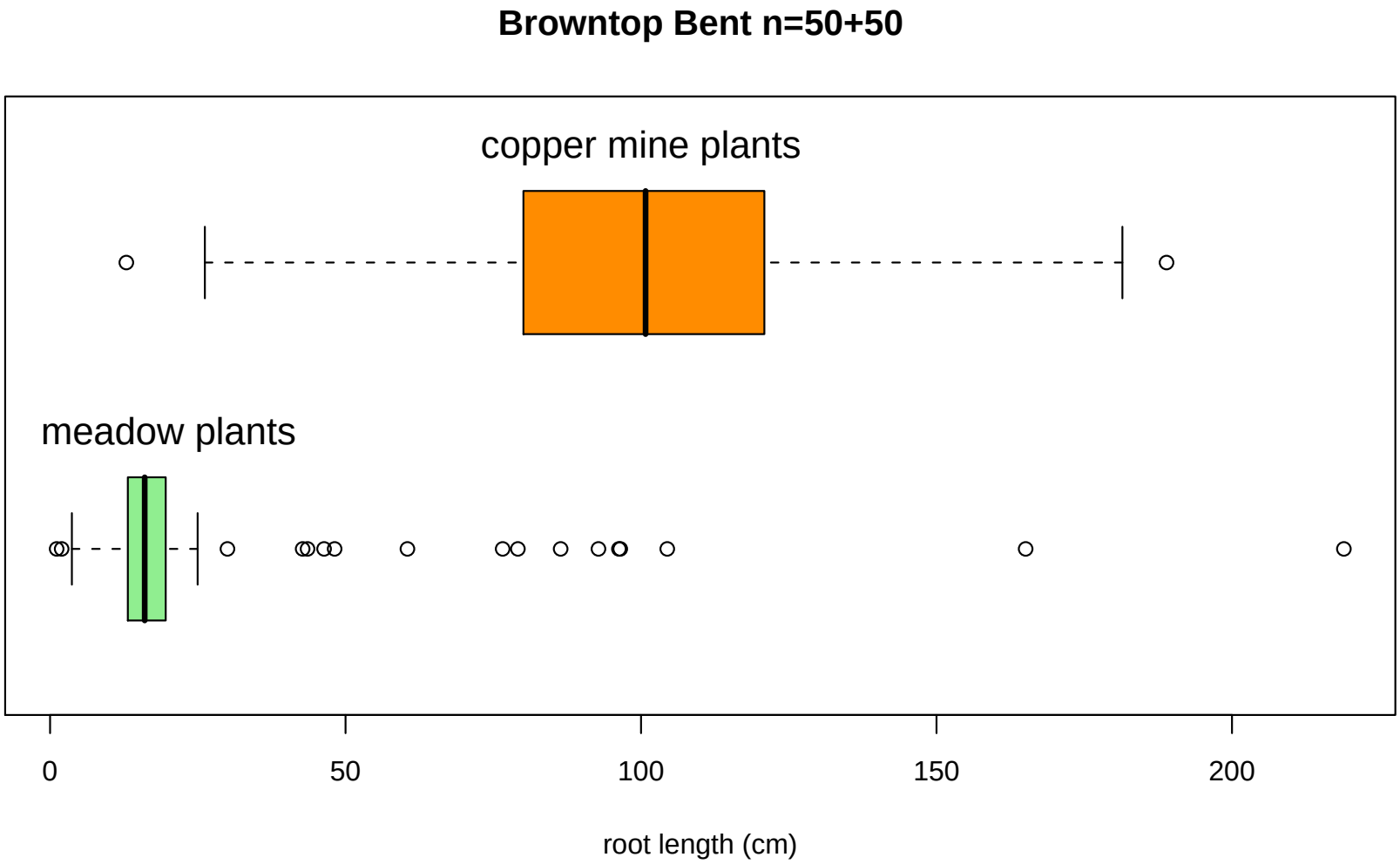


2/3 der Wurzellängen innerhalb [m-sd,m+sd]???? Nein!

Fazit des Straußgras-Beispiels

Manche Verteilungen können nur mit mehr als zwei Variablen angemessen beschrieben werden.

z.B. mit den fünf Werten der Boxplots:
min, Q_1, median, Q_3, max


Browntop Bent n=50+50
copper mine plants
meadow plants
0
50
100
150
200
root length (cm)

Schlussfolgerung

In der Biologie sind viele Datenverteilungen annähernd glockenförmig und können durch den Mittelwert und die Standardabweichung hinreichend beschrieben werden.

Es gibt aber auch Ausnahmen. Also:
Immer die Daten erst mal graphisch untersuchen!

Verlassen sie sich niemals allein auf numerische Kenngrößen!

Was ist R?

- Wir verwenden R in dieser Vorlesung als (sehr mächtigen) „Statistik-Taschenrechner.“

  (R ist eine für die Statistik und für stochastische Simulation entwickelte Programmiersprache, zudem sind viele statistische Standardverfahren bereits in R implementiert oder als Zusatzpaket verfügbar.)
- R hat eine sehr aktive Benutzer- und Entwicklergemeinde (die nahezu alle Bereiche der Statistik und viele Anwendungsbereiche (z.B. Populationsgenetik, Finanzmathematik) überdeckt).
- R ist frei verfügbar unter der GNU general public license, für (nahezu) alle Rechnerarchitekturen erhältlich: http://www.r-project.org/
- R ist auf ZDV-Rechnern installiert.

R installieren, starten, anhalten

Installation: Windows, Mac OS: Binaries von
http://www.r-project.org/ (siehe Link Download, Packages, CRAN dort)
Linux: Für die meisten Distributionen gibt es fertige Pakete
Fragen oder Probleme: In der Übung ansprechen

R starten: Windows, Mac OS: Icon (ggf. aus Menu) anklicken,
Linux/Unix: > R auf einer Konsole
(oder mit ESS aus Emacs heraus, oder aus rstudio, ...).

R beenden: q() (fragt, ob Daten gespeichert werden sollen)

laufende Rechnungen unterbrechen: CTRL-C

Datensatz `x` in R eingeben

```
x <- c(  53,52,41,41,42,58,40,43,42,38,43,49,34,51,45,
         39,41,45,45,39,37,36,42,44,47,43,46,43,43,45,
         42,52,49,44,50,40,47,46,50,50,41,51,41,47,42,
         52,36,46,42,56,39,40,36,42,36,36,47,45,47,49 )
```

(aus Datei einlesen: `x <- scan('Dateiname')`)

Mittelwert (`mean`), Standardabweichung (`sd`), Median, und Quantile

```
mean(x)
sd(x)
median(x)
quantile(x, 0.25, type=1)
quantile(x, 0.75, type=1)
summary(x)
```

Boxplot, Histogramm

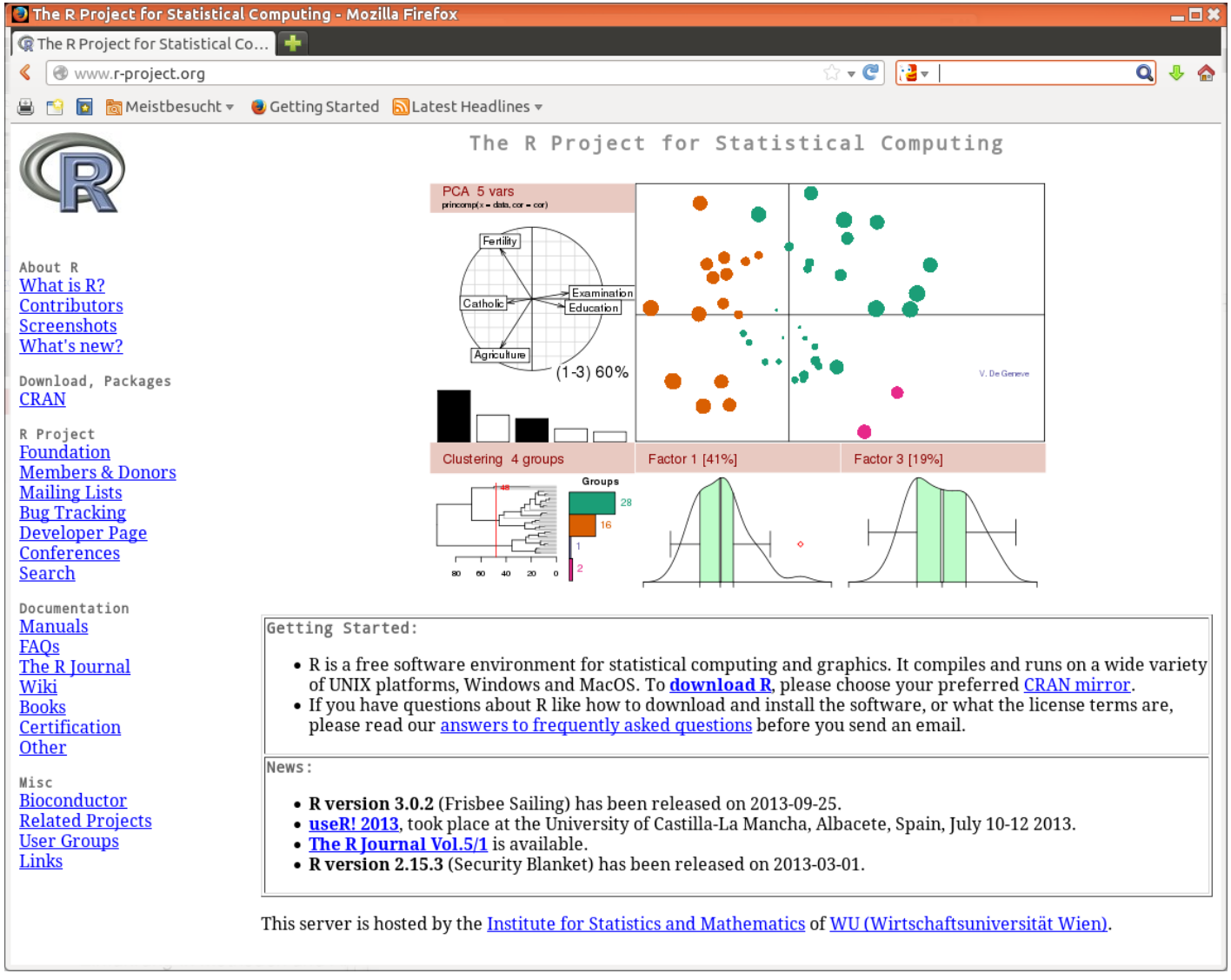
```
boxplot(x)
hist(x)
```

Nur zur Information: Literatur zu R

Wir werden im Zusammenhang der Vorlesung nur wenige R-Befehle verwenden, so dass Sie i.A. keine über die Folien hinausgehende Literatur benötigen werden.

- „Standardreferenz“:

  W.N. Venables et al, *An Introduction to R*,
  http://cran.r-project.org/manuals.html
- Günther Sawitzki, *Einführung in R*,
  http://sintro.r-forge.r-project.org/
- William N. Venables, Brian D. Ripley, *Modern applied statistics with S* („Standardlehrbuch“, UB Lehrbuchsammlung)
- Lothar Sachs and Jürgen Hedderich, *Angewandte Statistik – Methodensammlung mit R* (E-Book, UB)
- Christine Duller, *Einführung in die nichtparametrische Statistik mit SAS und R : ein anwendungsorientiertes Lehr- und Arbeitsbuch* (E-Book, UB)
- Helge Toutenburg, Christian Heumann, *Deskriptive Statistik : Eine Einführung in Methoden und Anwendungen mit R und SPSS* (E-Book, UB)
- Uwe Ligges, *Programmieren mit R* (E-Book, UB)


The R Project for Statistical Computing - Mozilla Firefox
The R Project for Statistical Co...
www.r-project.org
Meistbesucht
Getting Started
Latest Headlines
The R Project for Statistical Computing
About R
What is R?
Contributors
Screenshots
What's new?
Download, Packages
CRAN
R Project
Foundation
Members & Donors
Mailing Lists
Bug Tracking
Developer Page
Conferences
Search
Documentation
Manuals
FAQs
The R Journal
Wiki
Books
Certification
Other
Misc
Bioconductor
Related Projects
User Groups
Links
PCA 5 vars
Fertility
Examination
Education
Catholic
Agriculture
(1-3) 60%
V. De Geneve
Clustering 4 groups
Factor 1 [41%]
Factor 3 [19%]
Groups
Getting Started:
R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.
If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.
News:
R version 3.0.2 (Frisbee Sailing) has been released on 2013-09-25.
useR! 2013, took place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013.
The R Journal Vol.5/1 is available.
R version 2.15.3 (Security Blanket) has been released on 2013-03-01.
This server is hosted by the Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien).