Statistik

für Betriebswirtschaft und internationales Management

Sommersemester 2015

Prof. Dr. Stefan Etschberger
Hochschule Augsburg

Outline

Material zur Vorlesung

Kursmaterial:

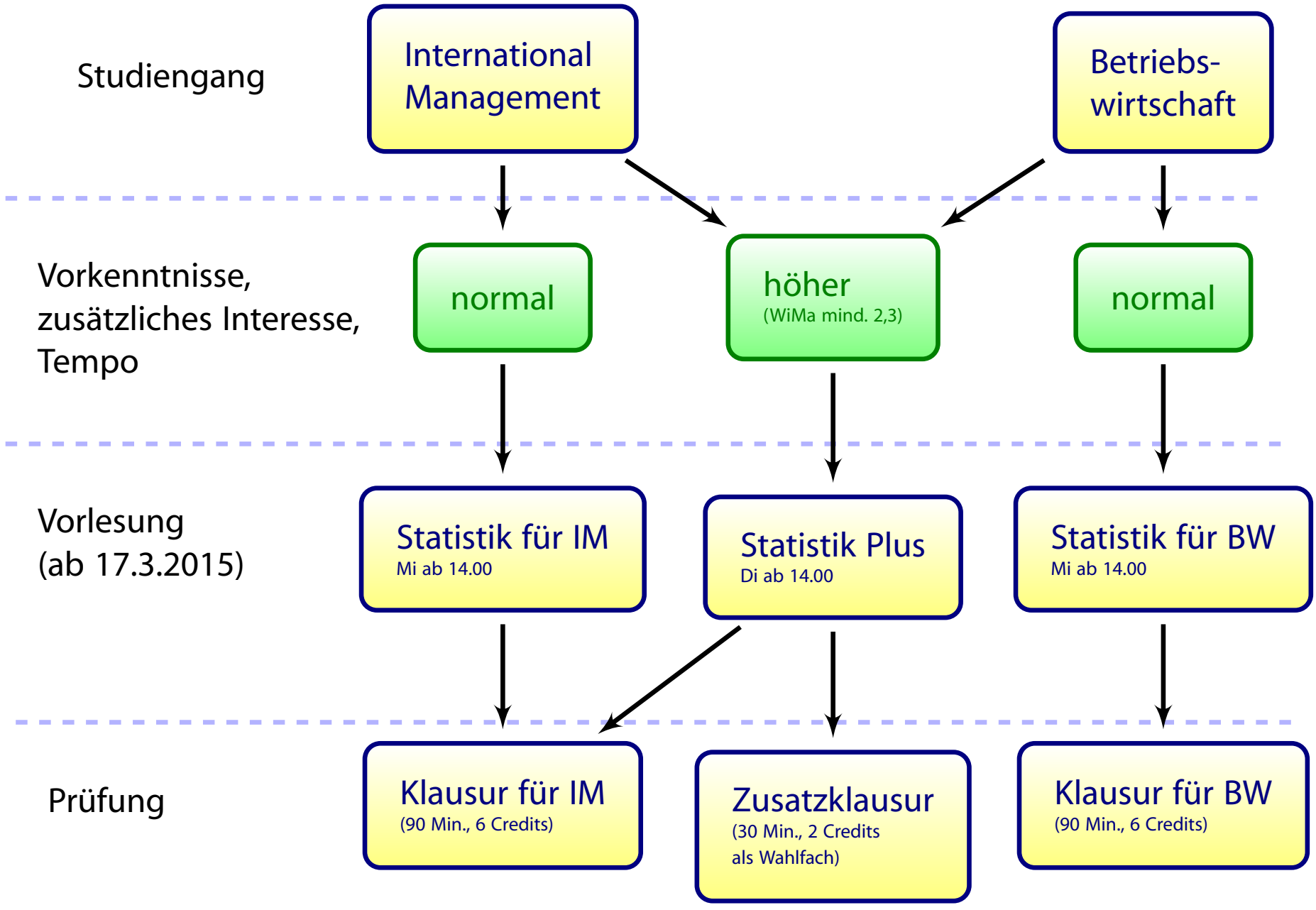
- Aufgabensatz (beinhaltet Aufgaben zu R)
- Handout der Folien
- Alle Folien inklusive Anmerkungen (nach der jeweiligen Vorlesung)
- Beispieldaten
- Alle Auswertungen als **R**-Datei

Literatur:

Bamberg, Günter, Franz Baur und Michael Krapp (2011). **Statistik**. 16. Aufl. München: Oldenbourg Verlag. ISBN: 3486702580.

Dalgaard, Peter (2002). **Introductory Statistics with R**. New York: Springer.

Fahrmeir, Ludwig, Rita Künstler, Iris Pigeot und Gerhard Tutz (2009). **Statistik: Der Weg zur Datenanalyse**. 7. Aufl. Berlin, Heidelberg: Springer. ISBN: 3642019382.


Studiengang
International Management
Betriebs-wirtschaft
Vorkenntnisse, zusätzliches Interesse, Tempo
normal
höher
(WiMa mind. 2,3)
normal
Vorlesung (ab 17.3.2015)
Statistik für IM
Mi ab 14.00
Statistik Plus
Di ab 14.00
Statistik für BW
Mi ab 14.00
Prüfung
Klausur für IM
(90 Min., 6 Credits)
Zusatzklausur
(30 Min., 2 Credits als Wahlfach)
Klausur für BW
(90 Min., 6 Credits)

Prüfung

Klausur:

- Klausur am Ende des Semesters
- Bearbeitungszeit: 90 Minuten
- R ist prüfungsrelevant: Siehe Anmerkungen in Übungsaufgaben!
- Hilfsmittel:
  - Schreibzeug,
  - Taschenrechner, der nicht 70! berechnen kann,
  - ein Blatt (DIN-A4, vorne und hinten beschrieben) mit handgeschriebenen Notizen (keine Kopien oder Ausdrucke),
- Danach (optional): Für Teilnehmer der Statistik-Plus Vorlesung noch eine 30-minütige Teilklausur über zusätzliche Inhalte (2 Wahlfachcredits zusätzlich möglich; Hilfsmittel TR und ein Blatt)

Fragebogen

Umfrage Statistik Sommersemester 2015

Bitte beantworten Sie folgende Fragen vollständig und füllen Sie jeweils *nur eine* beliebige Spalte *leserlich* aus.

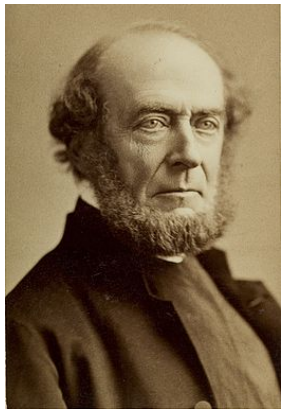
Ihr Alter (in Jahren)										
Ihre Größe (in cm)										
Ihr Geschlecht (m/w)										
Wie alt ist (bzw. wäre) Ihr Vater heute?										
Wie alt ist (bzw. wäre) Ihre Mutter heute?										
Größe Ihres Vaters (cm)										
Größe Ihrer Mutter (cm)										
Wie viele Geschwister haben Sie?										
Wunschfarbe für Ihr nächstes Smartphone; mögliche Auswahl: (si)lber, (sc)hwarz, (w)eiß, (g)elb, (b)lau, (r)ot										
Ausgaben für Ihre mobile Kommunikation (egal wer bezahlt hat) in den vergangenen 12 Monaten inkl. Hardware (Smartphone, Zubehör), Software (Apps), Vertrag										
Wie viele Paar Schuhe besitzen Sie?										
Ausgaben für Ihre Schuhe (egal wer bezahlt hat) in den letzten 12 Monaten										
Ihre Essgewohnheiten; mögliche Auswahl: (vega)n, (vege)tarisch, (p)escetarisch, also vegetarisch und Fische, (c)arnivor, also auch Fleisch										
Ihre Note in der Matheklausur (Note oder (n)icht mitgeschrieben)										
Waren Sie mit Ihrer Leistung in der Matheklausur zufrieden? Antworten: (s)ehr zufrieden, (z)ufrieden, (g)eht so, (u)nzufrieden, (n)icht mitgeschrieben										

Statistik: Table of Contents

Zitate

- **Leonard Henry Courteney (1832-1918)**:

  „ There are three kinds of lies: lies, damned lies and statistics.“

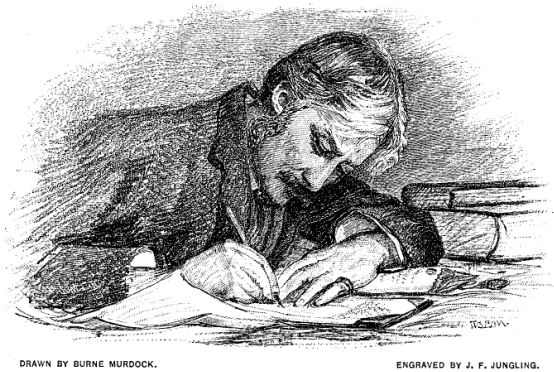


- **Winston Curchill (1874-1965) angeblich**:

  „ Ich glaube nur den Statistiken, die ich selbst gefälscht habe.“

- **Andrew Lang (1844-1912)**:

  „ Wir benutzen die Statistik wie ein Betrunkener einen Laternenpfahl: Vor allem zur Stütze unseres Standpunktes und weniger zum Beleuchten eines Sachverhalts.“



Quellen: Wikimedia Commons

Statistik ist überall

Morgens in Zeitung: Mehr Statistiken als Goethe und Schiller im ganzen Leben gesehen haben:

- Arbeitslosenzahlen wachsen
- Vogelgrippe breitet sich aus
- 78,643% der Deutschen unzufrieden mit Löw
- Bundesbürger verzehrt 5,8 Liter Speiseeis pro Jahr
- Musiker leben länger als andere Leute
- Tennispieler B hat noch nie gegen einen brilletragenden Linkshänder verloren, der jünger ist als er
- in New York schläft man am sichersten im Central Park

Viele dieser Statistiken: Falsch, bewußt manipuliert oder unpassend ausgesucht.

Fehlerquellen:

- Zahlenmanipulation
- irreführende Darstellung der Zahlen
- ungenügendes Wissen

Fehlerhafter Sprachgebrauch: „Richtig“ Fragen

1. Frage:

„Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten?“

Resultat:

- Dafür: 44%
- Dagegen: 20%
- Unentschieden: 36%

2. Frage:

„Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten oder muss man es jedem einzelnen überlassen, ob er in der Gewerkschaft sein will oder nicht?“

Resultat:

- Dafür: 24%
- Dagegen: 70%
- Unentschieden: 6%

Geänderte Definitionen

Laut einem „Bericht zur Bekämpfung des Analphabetismus in Deutschland“:

- Heute gibt es in Deutschland ca. 7 Millionen Analphabeten
- Zu Kaiser Wilhelms Zeiten gab es weniger als 10 000

Was leiten Sie daraus ab?

BILDUNG

7,5 Millionen Deutsche sind Analphabeten

Ein Siebtel der erwerbsfähigen Bevölkerung kann laut einer Studie kaum lesen und schreiben – doppelt so viel wie bisher gedacht. Bildungsministerin Schavan will reagieren [weiter…]

ANALPHABETISMUS

Ein Land verlernt das Lesen

Studenten verstehen abstrakte Texte nicht mehr, ein Schulbuchverlag kürzt Klassiker, Banker besuchen Lesekurse: Viele Deutsche haben keine Lust mehr zu lesen. [weiter…]

ANALPHABETISMUS

Buchstäblich resigniert

Mehr als sieben Millionen Deutsche können kaum lesen und schreiben. Erst jetzt hat die Politik das Problem erkannt. Aber es gibt zu wenig Geld für Kurse. Von M. Spiewak [weiter…]

Quelle: Zeit.de

Definition

Zu Kaiser Wilhelms Zeiten:

„Analphabet ist, wer seinen Namen nicht schreiben kann.“

Definition heute:

„Ein Analphabet ist eine Person, die sich nicht beteiligen kann an all den zielgerichteten Aktivitäten ihrer Gruppe und ihrer Gemeinschaft, bei denen Lesen, Schreiben und Rechnen erforderlich ist und an der weiteren Nutzung dieser Kulturtechniken für ihre weitere Entwicklung und die der Gesellschaft“.

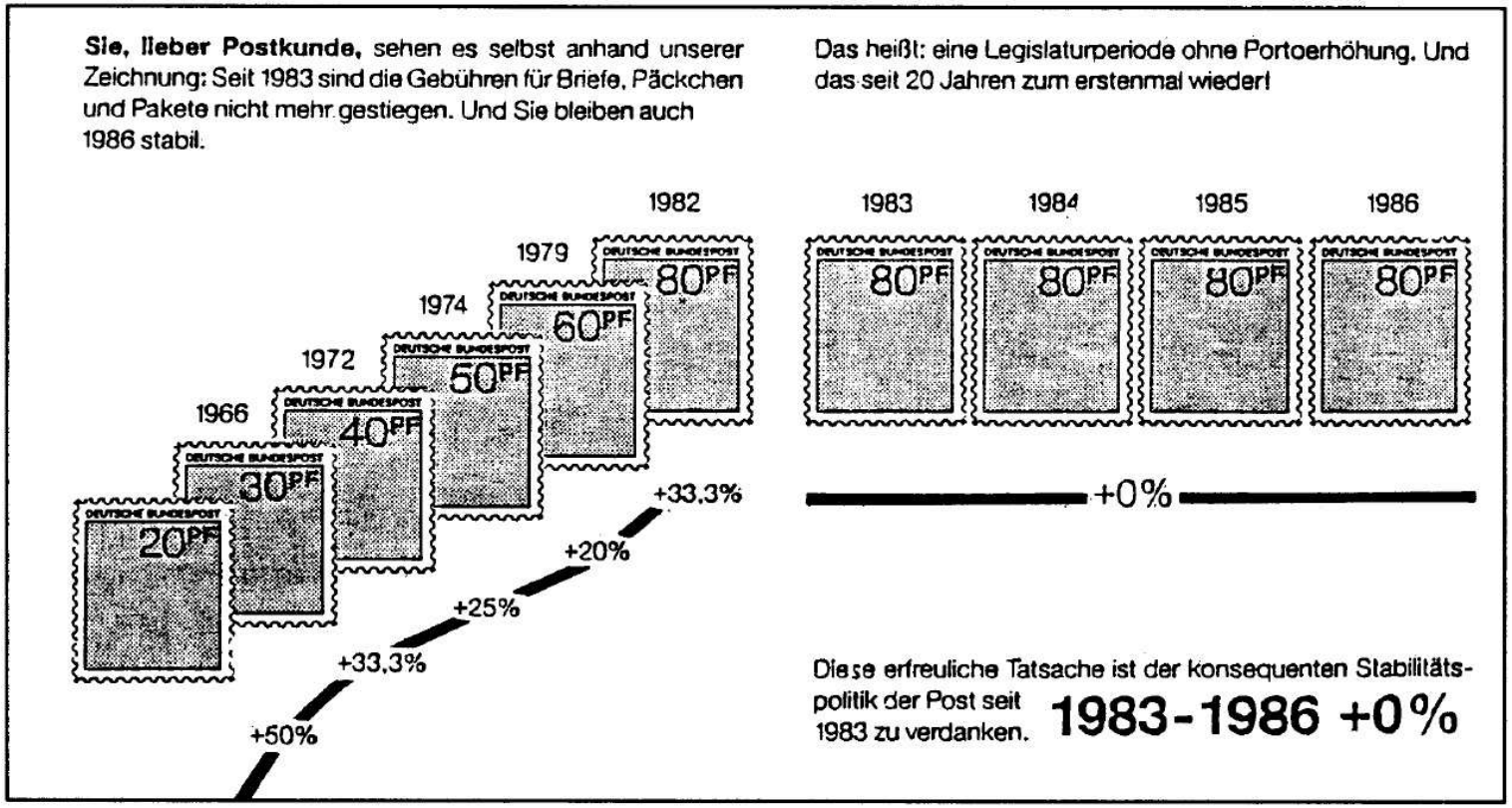
Aussage des Vertriebsleiters:

„Unser Umsatz stieg vor einem Jahr um 1%. Dieses Jahr stieg das Umsatzwachstum um 50%!“

Im Klartext:

- Basisjahr: Umsatz 100
- Dann: Wachstum auf 101
- Dieses Jahr: Wachstum des Wachstums um 50% bedeutet 1,5% Wachstum. Also Umsatz dann 102,5049

Verzerrte Zeitachse

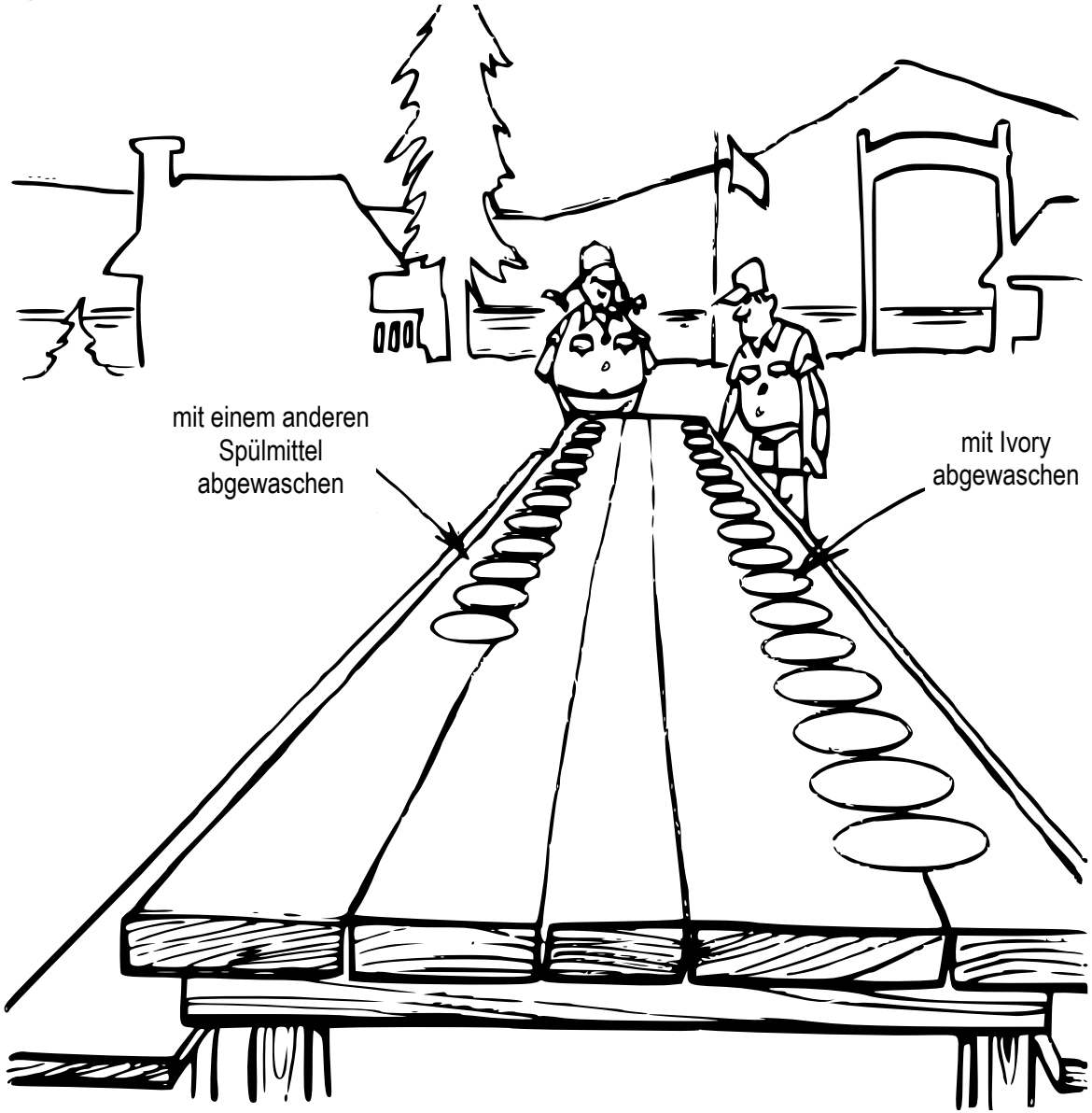


Quelle Kramer (2011)

Grafik aussagekräftig?

Quelle: Bach u. a. (2006)

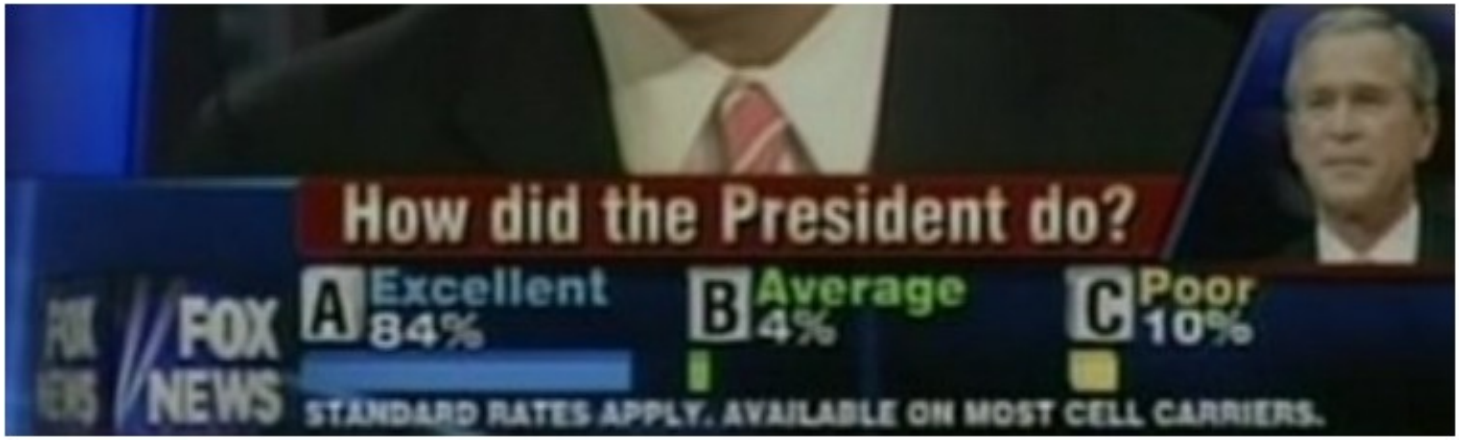
11 zu 15



Falsche Berechnungsbasis

- Ein Einzelhändler bezieht ein Produkt zu 100 € und verkauft es für 200 €. Hat er eine Gewinnspanne von 50% oder 100%?
- Bahn: 9 Tote pro 10 Mio Passagieren je Kilometer
  Flugzeug: 3 Tote pro 10 Mio Passagieren je Kilometer
  Bahn: 7 pro 10 Mio Passagiere je Stunde
  Flugzeug: 24 pro 10 Mio Passagiere je Stunde
- Nur 40 % aller durch Autounfälle Gestorbenen hatten keinen Sicherheitsgut angelegt
  Also: Keinen Gurt anlegen ist sicherer
- Die Hälfte der Todesfälle ereignen sich in Krankenhäusern
  Also: Krankenhäuser sind lebenssgefährlich
- Zwei Drittel aller alkoholabhängigen Personen sind verheiratet
  Also: die Ehe führt zum Alkohol

Fernsehumfragen



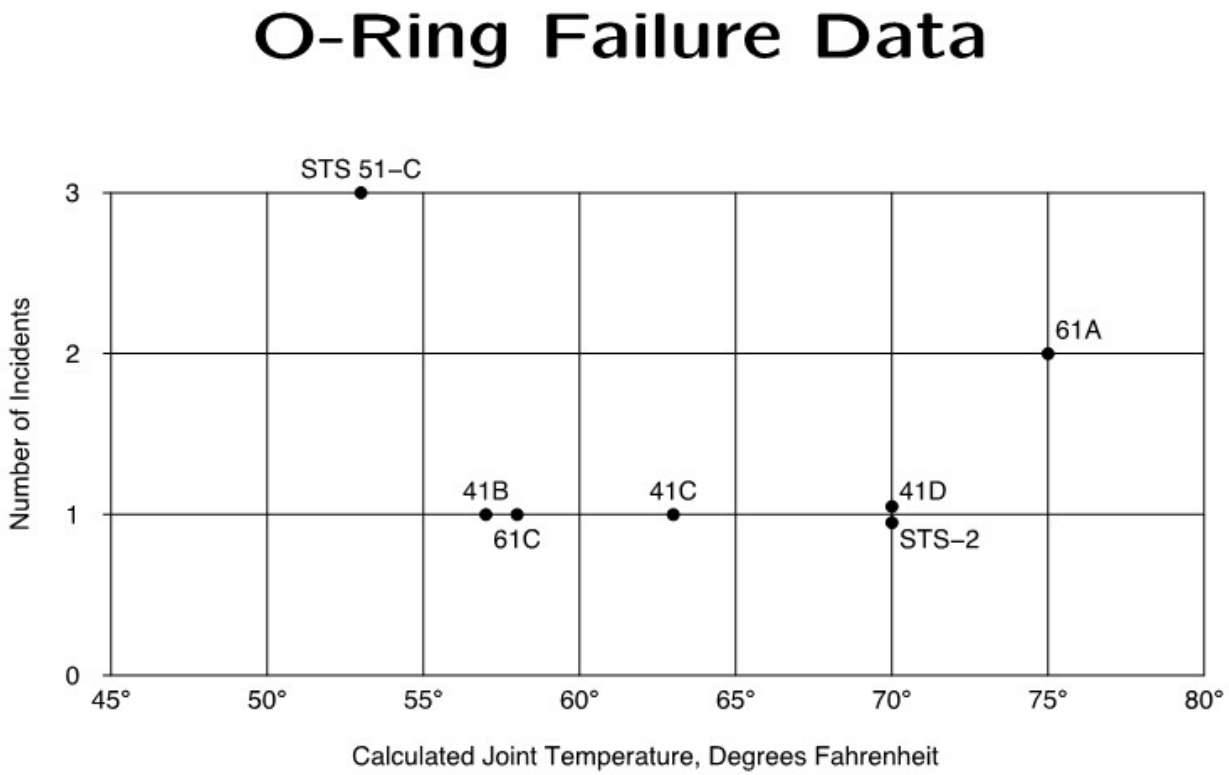
- Kostenpflichtige Telefonabstimmung nach regierungsfreundlichem Bericht im Fernsehen
- In den meisten Umfragen erreichte Bush zu diesem Zeitpunkt nur 30 % Zustimmung

Challenger-Katastrophe

Am 28. Januar 1986, 73 Sekunden nach dem Start der Mission STS-51-L, brach die Raumfähre in etwa 15 Kilometer Höhe auseinander. Dabei starben alle sieben Astronauten. Es war der bis dahin schwerste Unfall in der Raumfahrtgeschichte der USA.

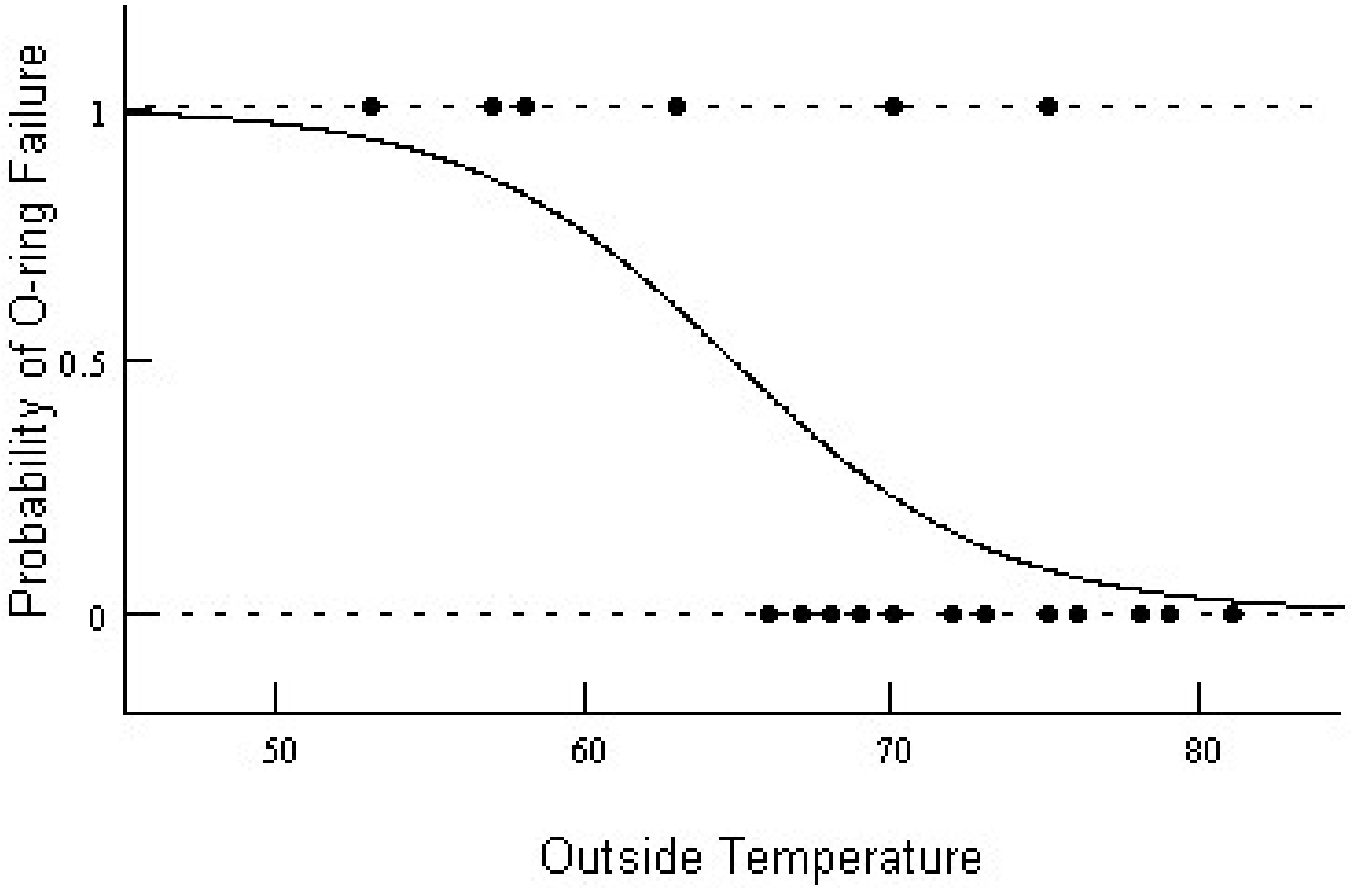
Challenger Unfall 1986

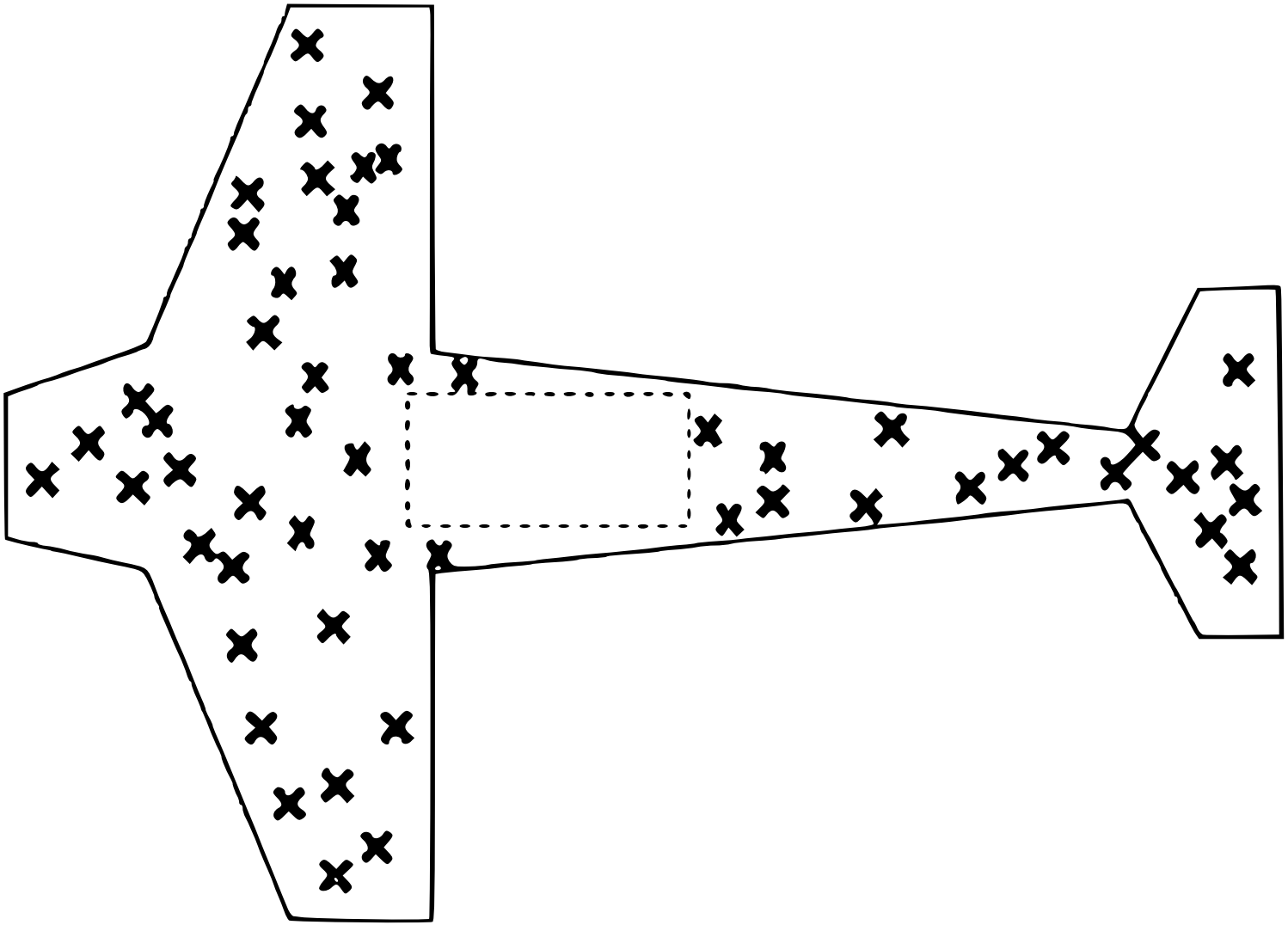
- Grund für Explosion: 2 Gummidichtungsringe waren undicht
- Die Temperatur der Dichtungsringe: Unter 20° F (ca. -6,7° C).
- Probleme mit Dichtungsringen bei Start der vorigen Fähre: Umgebungstemperatur 53° F (ca. 11,7° C).
- Frage: Ist der Dichtungsfehler durch die Umgebungstemperatur zu prognostizieren?



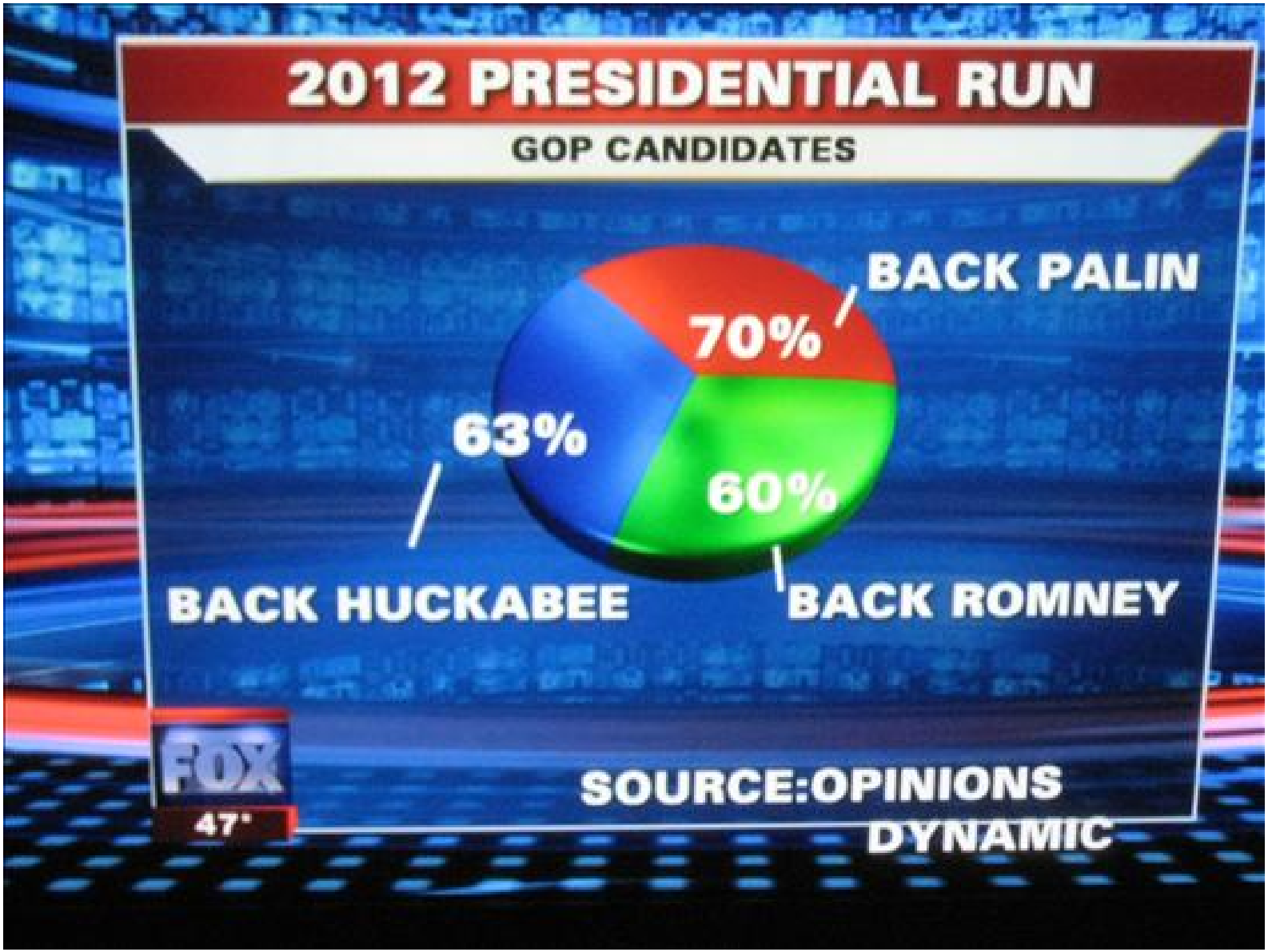
Challenger Unfall 1986

- Fehler in Analyse: Starts ohne Fehler wurden nicht berücksichtigt
- Korrekte Modellierung mittels logistischer Regression liefert:



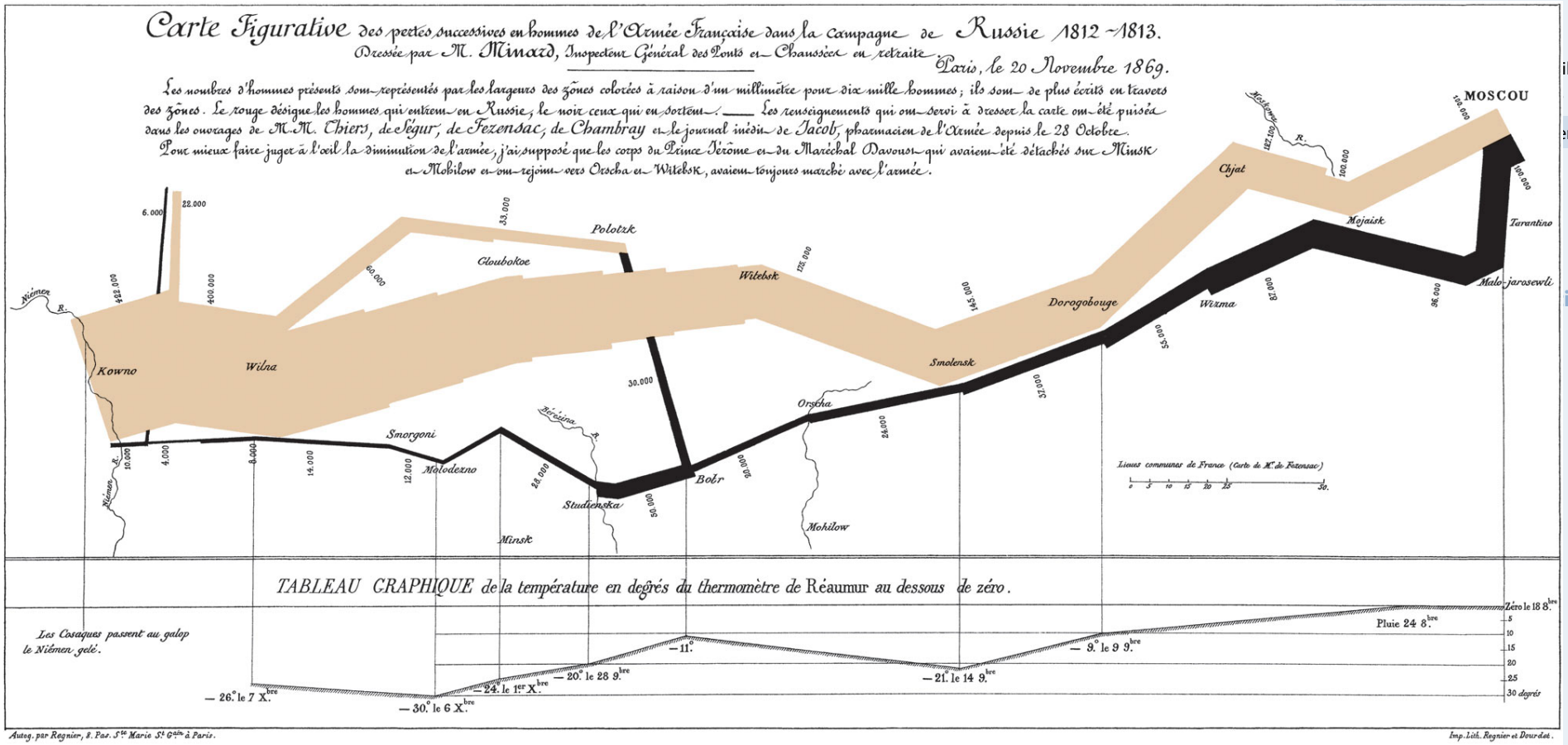


Aussage?



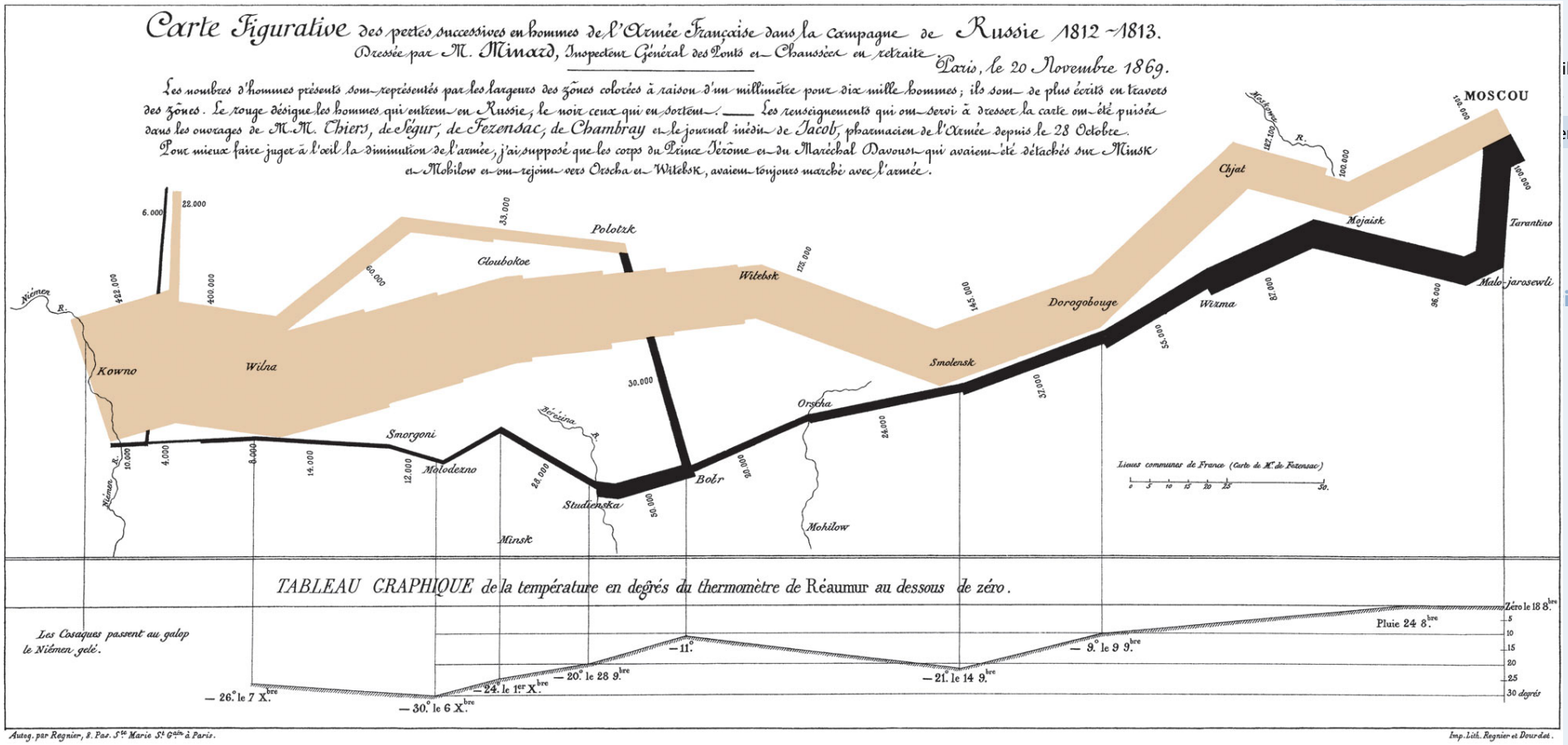
Quelle: Friendly, Michael (2001), Gallery of Data Visualization, `http://www.datavis.ca/gallery/`, Stand November 2014

Minards Grafik von 1869 über Napoleons Rußlandfeldzug



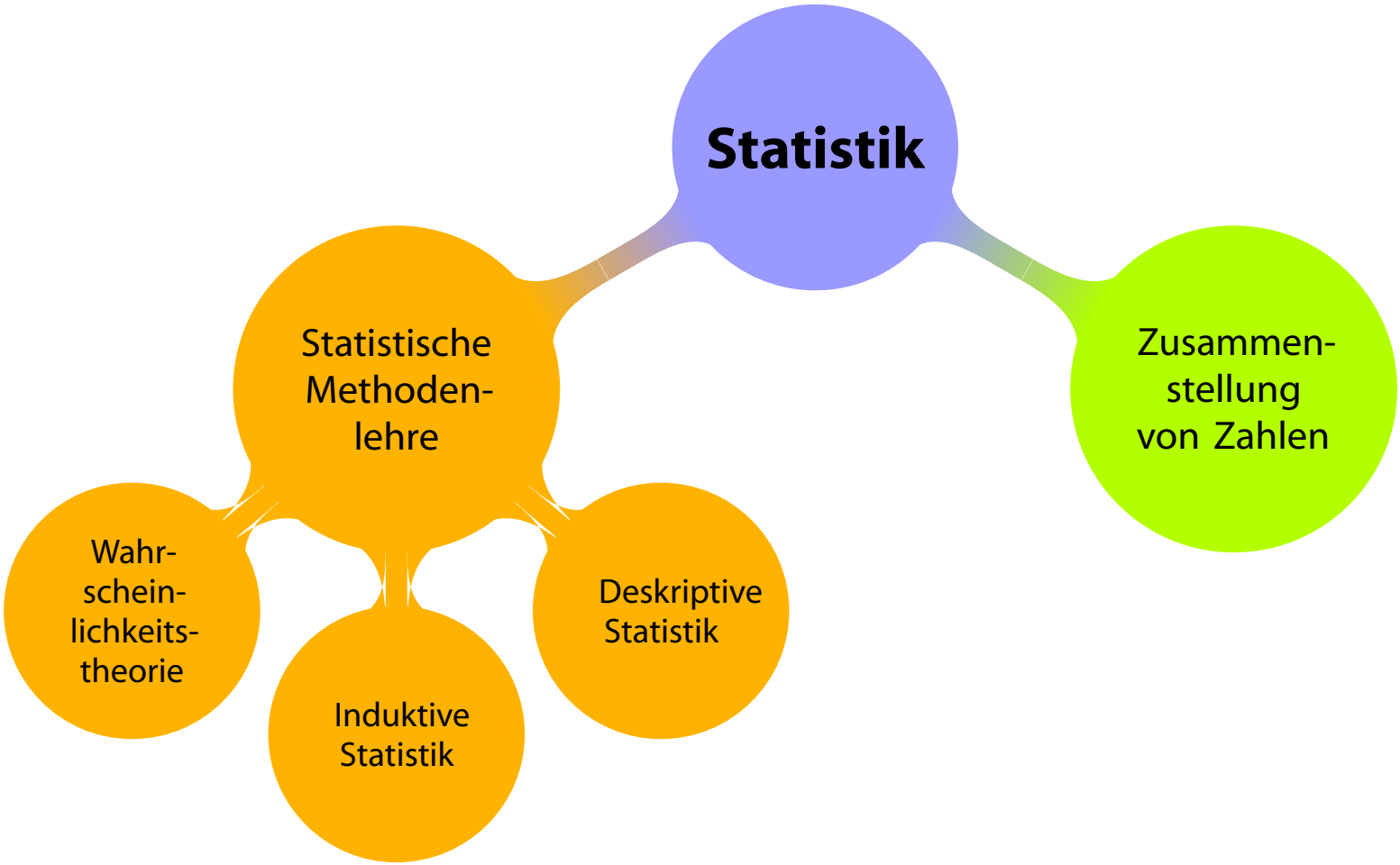
Quelle: Wikimedia Commons, `http://goo.gl/T7ZNme`, Stand November 2014

Minards Grafik von 1869 über Napoleons Rußlandfeldzug



Quelle: Wikimedia Commons, `http://goo.gl/T7ZNme`, Stand November 2014

Bedeutungen des Begriffs Statistik



Beispiel

12 Beschäftigte werden nach der Entfernung zum Arbeitsplatz (in km) befragt.
Antworten: 4, 11, 1, 3, 5, 4, 20, 4, 6, 16, 10, 6

- deskriptiv:
  - Durchschnittliche Entfernung: $7{,}5$
  - Klassenbildung:

Klasse	$[0;5)$	$[5;15)$	$[15;30)$
Häufigkeit	5	5	2

- induktiv:
  - Schätze die mittlere Entfernung **aller** Beschäftigten.
  - Prüfe, ob die mittlere Entfernung geringer als 10 km ist.

Merkmale

Objekt

- Merkmalsträger: Untersuchte statistische Einheit

Variable

- Merkmal: Interessierende Eigenschaft des Merkmalträgers

Werte

- (Merkmals-)Ausprägung: Konkret beobachteter Wert des Merkmals
- Grundgesamtheit: Menge aller relevanten Merkmalsträger
- Typen von Merkmalen:

a) qualitativ – quantitativ
   - qualitativ: z.B. Geschlecht
   - quantitativ: z.B. Schuhgröße
   - Qualitative Merkmale sind quantifizierbar (z.B.: weiblich 1, männlich 0)

Zahlen:

b) diskret – stetig
   - diskret: Abzählbar viele unterschiedliche Ausprägungen
   - stetig: Alle Zwischenwerte realisierbar

Skalenniveaus

Nominalskala:

- Zahlen haben nur Bezeichnungsfunktion
- z.B. Artikelnummern

Ordinalskala:

- zusätzlich Rangbildung möglich
- z.B. Schulnoten
- Differenzen sind aber **nicht** interpretierbar!
  ➠ Addition usw. ist unzulässig.

qualitativ

Kardinalskala: , metrische Skala

- zusätzlich Differenzbildung sinnvoll
- z.B. Gewinn
- Noch feinere Unterscheidung in: Absolutskala, Verhältnisskala, Intervallskala

quantitativ (Zahlen)

Skalendegression und Skalenprogression

Ziel der Skalierung: Gegebene Information angemessen abbilden, möglichst ohne Über- bzw. Unterschätzungen

Es gilt:

- Grundsätzlich können alle Merkmale nominal skaliert werden.
- Grundsätzlich kann jedes metrische Merkmal ordinal skaliert werden.

Das nennt man Skalendegression. Dabei: Informationsverlust

Aber:

- Nominale Merkmale dürfen **nicht** ordinal- oder metrisch skaliert werden.
- Ordinale Merkmale dürfen **nicht** metrisch skaliert werden.

Das nennt nennt man Skalenprogression. Dabei: Interpretation von **mehr Informationen** in die Merkmale, als inhaltlich vertretbar.
(Gefahr der **Fehlinterpretation**)

Was ist R und warum soll man es benutzen?

- R ist ein freies Softwarepaket zu Statistik und Datenanalyse
- R ist sehr mächtig und weit verbreitet in Wissenschaft und Industrie (sogar von mehr Leuten benutzt als z.B. SPSS)
- Ursprung von R: 1993 an der Universität Auckland von Ross Ihaka and Robert Gentleman entwickelt
- Seitdem: Viele Leute haben R verbessert mit tausenden von Paketen für viele Anwendungen
- Nachteil (auf den ersten Blick): Kein point und click tool
- Großer Vorteil (auf den zweiten Blick): Kein point und click tool

```
> summary(diamonds$price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    326     950    2401    3933    5324   18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
>
```

Download: `R-project.org`

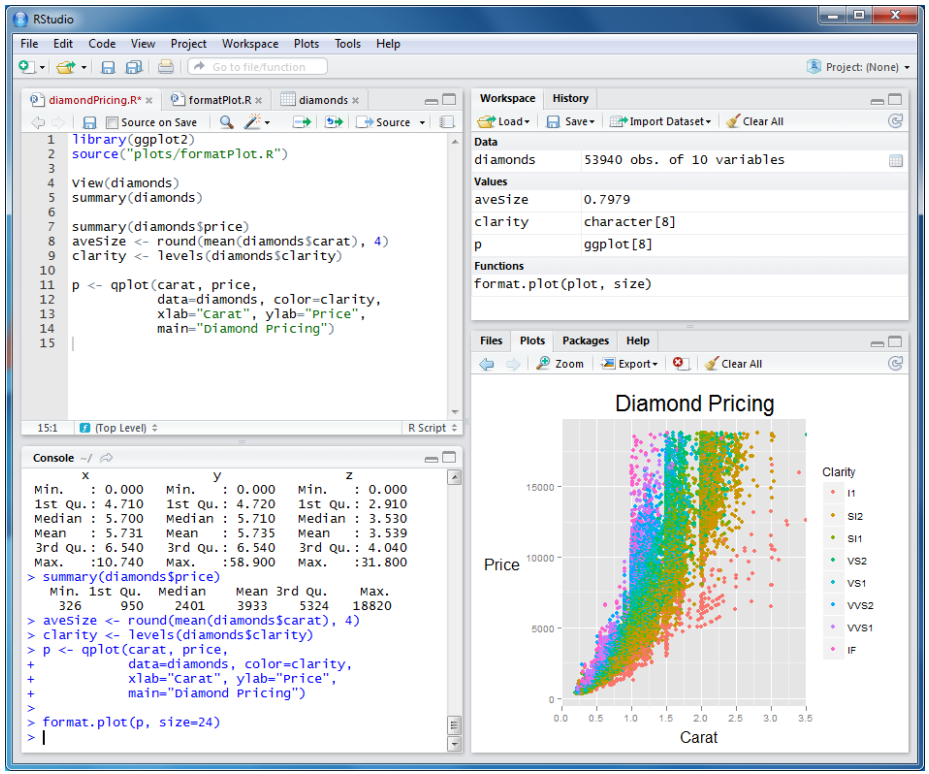
graphics source: `http://goo.gl/W70kms`

Was ist RStudio?

- RStudio ist ein Integrated Development Environment (IDE) um R leichter benutzen zu können.
- Gibt's für OSX, Linux und Windows
- Ist auch frei
- Trotzdem: Sie müssen Kommandos schreiben
- Aber: RStudio unterstützt Sie dabei
- **Download:** `RStudio.com`

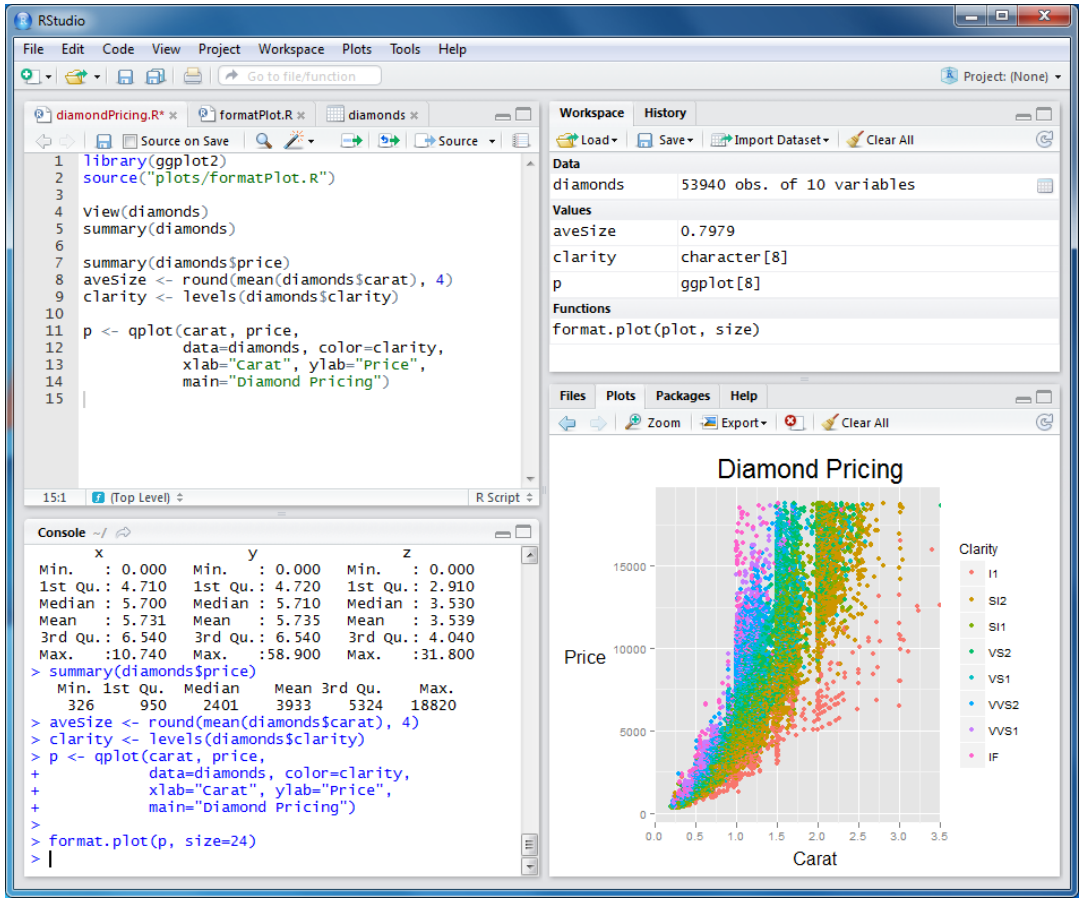
RStudio™

Free & Open-Source IDE for R



RStudio Kennenlernen

- Code
- Console
- Workspace
- History
- Files
- Plots
- Packages
- Help
- Auto-Completion
- Data Import



Daten einlesen und Struktur anschauen

```r
# Arbeitsverzeichnis setzen (alternativ über Menü)
setwd("C:/ste/work/vorlesungen/2014WS_Doktorandenworkshop/2015_01_Statistik_Workshop")

# Daten einlesen aus einer csv-Datei (Excel)
MyData = read.csv2(file="../Daten/Umfrage_HSA_2014_03.csv", header=TRUE)
```

```r
# inspect structure of data
str(MyData)
```

```
## 'data.frame': 205 obs. of  14 variables:
##  $ Alter      : int  21 20 19 20 20 24 20 27 23 21 ...
##  $ Groesse    : int  173 164 172 168 169 185 170 165 184 178 ...
##  $ Geschlecht : Factor w/ 2 levels "Frau","Mann": 1 1 1 1 1 2 1 1 2 2 ...
##  $ AlterV     : int  54 57 49 45 43 54 49 53 52 55 ...
##  $ AlterM     : int  51 57 58 49 42 52 53 53 48 55 ...
##  $ GroesseV   : int  187 172 193 185 182 179 182 175 182 180 ...
##  $ GroesseM   : int  170 166 162 164 164 163 172 165 175 168 ...
##  $ Geschwister: int  1 0 3 3 5 2 2 1 2 1 ...
##  $ Farbe      : Factor w/ 6 levels "blau","gelb",..: 6 6 4 4 6 4 3 6 4 6 ...
##  $ AusgKomm   : num  156 450 240 35.8 450 250 100 300 450 1300 ...
##  $ AnzSchuhe  : int  17 22 15 15 22 8 20 10 3 7 ...
##  $ AusgSchuhe : int  50 500 400 100 450 90 250 200 300 200 ...
##  $ NoteMathe  : num  5 1.7 2.3 2 2 4 NA 4 2.7 2.7 ...
##  $ MatheZufr  : Ord.factor w/ 4 levels "nicht"<"geht so"<..: 1 4 4 4 4 2 1 1 3 3 ...
```


Erste Zeilen der Datentabelle


```
# Erste Zeilen in Datentabelle
head(MyData, 6)
```

```
##   Alter Groesse Geschlecht AlterV AlterM GroesseV GroesseM Geschwister   Farbe AusgKomm AnzSchuhe
## 1    21     173       Frau     54     51      187      170           1   weiss    156.0        17
## 2    20     164       Frau     57     57      172      166           0   weiss    450.0        22
## 3    19     172       Frau     49     58      193      162           3 schwarz    240.0        15
## 4    20     168       Frau     45     49      185      164           3 schwarz     35.8        15
## 5    20     169       Frau     43     42      182      164           5   weiss    450.0        22
## 6    24     185       Mann     54     52      179      163           2 schwarz    250.0         8
##   AusgSchuhe NoteMathe MatheZufr
## 1         50       5.0     nicht
## 2        500       1.7      sehr
## 3        400       2.3      sehr
## 4        100       2.0      sehr
## 5        450       2.0      sehr
## 6         90       4.0   geht so
```

```
# lege MyData als den "Standard"-Datensatz fest
attach(MyData)
```

```
# Wie Viele Objekte gibt's im Datensatz?
nrow(MyData)
```

```
## [1] 205
```

```
# Wie Viele Merkmale?
ncol(MyData)
```

```
## [1] 14
```

Daten kennenlernen


```
# Auswahl spezieller Objekte und Merkmale über [Zeile, Spalte]
MyData[1:3, 2:5]
```

```
##   Groesse Geschlecht AlterV AlterM
## 1     173       Frau     54     51
## 2     164       Frau     57     57
## 3     172       Frau     49     58
```

```
# Auswahl von Objekten über logische Ausdrücke
head(MyData$Geschlecht=="Frau" & MyData$Alter<19, 30)
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [17] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
# Einsetzen in Klammern und Ausgabe von Alter des Studenten, seines Vaters und seiner Mutter
MyData[MyData$Geschlecht=="Frau" & MyData$Alter<19, # Objektauswahl
      c("Alter", "AlterM", "AlterV")  # Welche Merkmale anzeigen?
      ]
```

```
##     Alter AlterM AlterV
## 23     18     50     52
## 44     18     37     43
## 51     18     51     54
## 57     18     53     57
## 74     18     53     49
## 126    18     44     45
## 139    18     51     58
## 185    18     46     48
## 193    18     49     47
```

Daten kennenlernen


```
# Zeige die Männer, die mehr als 1000 Euro für Schuhe
# und Mobilfunk zusammen ausgegeben haben
MyData[MyData$Geschlecht=="Mann" & MyData$AusgSchuhe + MyData$AusgKomm > 1000,
       c("Alter", "Geschwister", "Farbe", "AusgSchuhe", "AusgKomm")]

##     Alter Geschwister   Farbe AusgSchuhe AusgKomm
## 10     21           1   weiss        200     1300
## 15     20           1     rot        400      815
## 26     20           1 schwarz        200     1250
## 40     21           0  silber        300      825
## 87     20           1    blau       1000      350
## 113    25           0 schwarz        280     1200
## 146    24           1 schwarz        300      900
## 177    19           2 schwarz        500      720
## 178    23           1 schwarz        450      630
## 192    20           0 schwarz        400      950
```

Statistik: Table of Contents

Auswertungsmethoden für eindimensionales Datenmaterial

- Merkmal X wird an n Merkmalsträgern beobachtet ➠
  Urliste $(x_1, \ldots, x_n)$
  Im Beispiel: $x_1 = 4,\ x_2 = 11, \ldots, x_{12} = 6$
- Urlisten sind oft unübersichtlich, z.B.:

```
##  [1] 4 5 4 1 5 4 3 4 5 6 6 5 5 4 7 4 6 5 6 4 5 4 7 5 5 6 7 3 7 6 6 7 4 5 4 7 7 5 5 5 5 6 6 4 5 2 5 4
## [49] 7 5
```

- Dann zweckmäßig: Häufigkeitsverteilungen

Ausprägung (sortiert)	a_j	1	2	3	4	5	6	7	$\sum$
absolute Häufigkeit	$h(a_j) = h_j$	1	1	2	12	17	9	8	50
kumulierte abs. H.	$H(a_j) = \sum_{i=1}^{j} h(a_i)$	1	2	4	16	33	42	50	—
relative Häufigkeit	$f(a_j) = h(a_j)/n$	$\frac{1}{50}$	$\frac{1}{50}$	$\frac{2}{50}$	$\frac{12}{50}$	$\frac{17}{50}$	$\frac{9}{50}$	$\frac{8}{50}$	1
kumulierte rel. H.	$F(a_j) = \sum_{i=1}^{j} f(a_i)$	$\frac{1}{50}$	$\frac{2}{50}$	$\frac{4}{50}$	$\frac{16}{50}$	$\frac{33}{50}$	$\frac{42}{50}$	1	—

Beispiel: Alter von Studierenden in Statistik Vorlesung

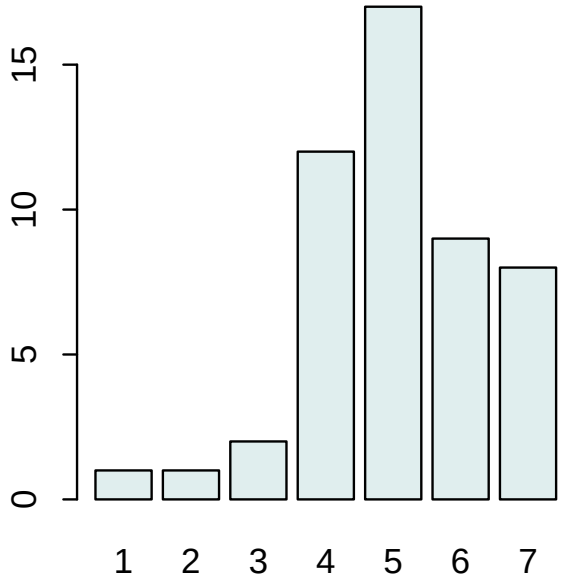
	$h(a_j)$	$H(a_j) = \sum_{i=1}^{j} h(a_i)$	$f(a_j) = \frac{h(a_j)}{n}$	$F(a_j) = \sum_{i=1}^{j} f(a_i)$
18	10	10	0.0488	0.0488
19	27	37	0.1317	0.1805
20	39	76	0.1902	0.3707
21	29	105	0.1415	0.5122
22	26	131	0.1268	0.6390
23	23	154	0.1122	0.7512
24	14	168	0.0683	0.8195
25	6	174	0.0293	0.8488
26	7	181	0.0341	0.8829
27	6	187	0.0293	0.9122
28	7	194	0.0341	0.9463
29	4	198	0.0195	0.9659
31	1	199	0.0049	0.9707
32	3	202	0.0146	0.9854
33	1	203	0.0049	0.9902
34	1	204	0.0049	0.9951
36	1	205	0.0049	1.0000

❶ Balkendiagramm

```
table(x)

## x
##  1  2  3  4  5  6  7
##  1  1  2 12 17  9  8

barplot(table(x), col="azure2")
```



(Höhe proportional zu Häufigkeit)

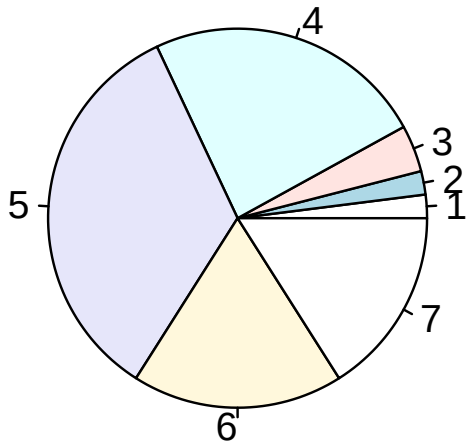
❷ Kreissektorendiagramm

Winkel: $w_j = 360° \cdot f(a_j)$

z.B. $w_1 = 360° \cdot \frac{1}{50} = 7{,}2°$

$w_7 = 360° \cdot \frac{8}{50} = 57{,}6°$

```
pie(table(x))
```

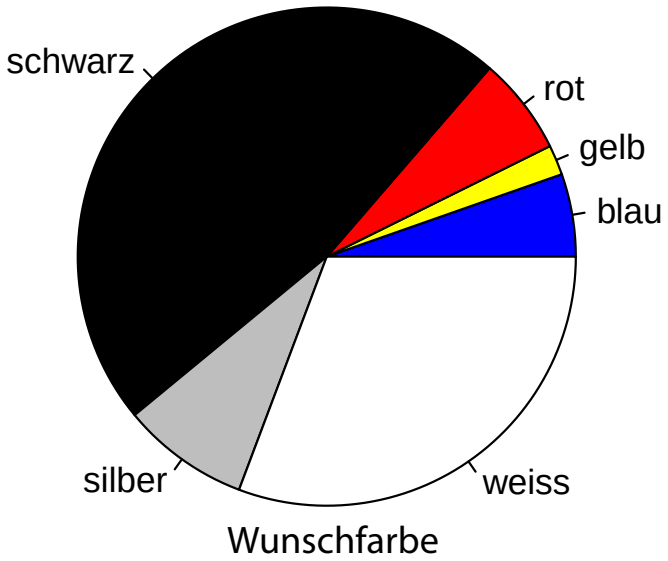


(Fläche proportional zu Häufigkeit)

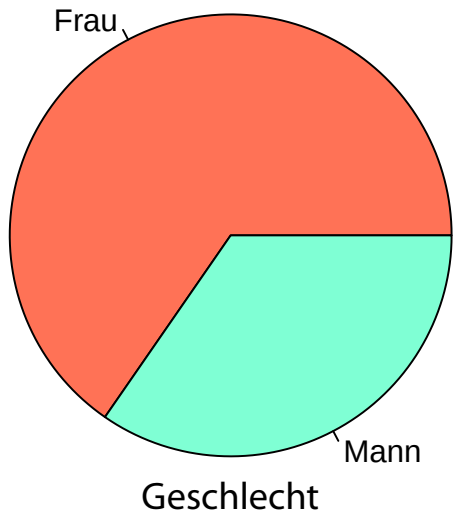
Umfrage: Kuchen

Kreissektorendiagramm

```
pie(table(MyData$Farbe),
    col=c("blue", "yellow", "red",
          "black", "grey", "white"))
```

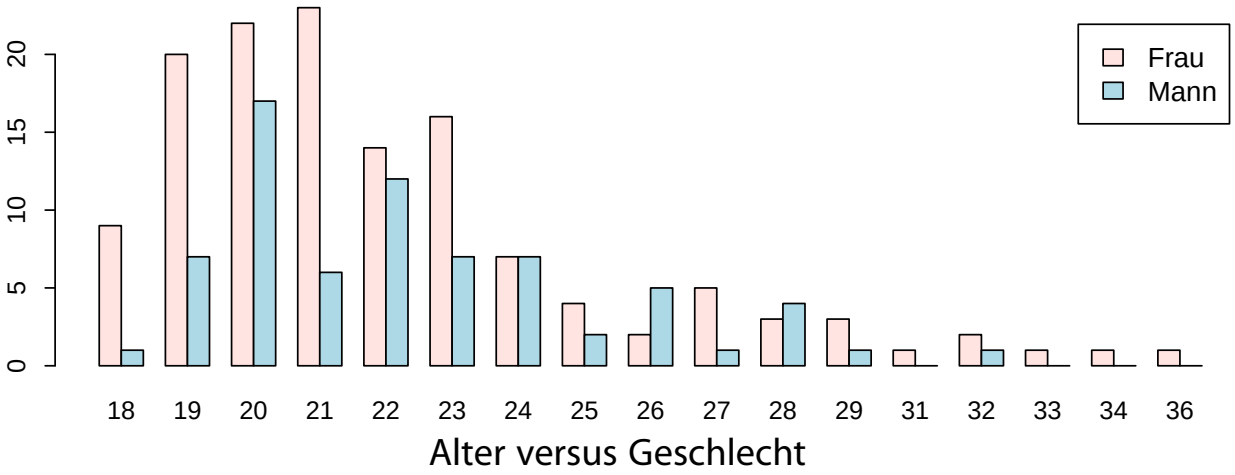



```
pie(table(MyData$Geschlecht),
    col=c("coral1", "aquamarine"))
```

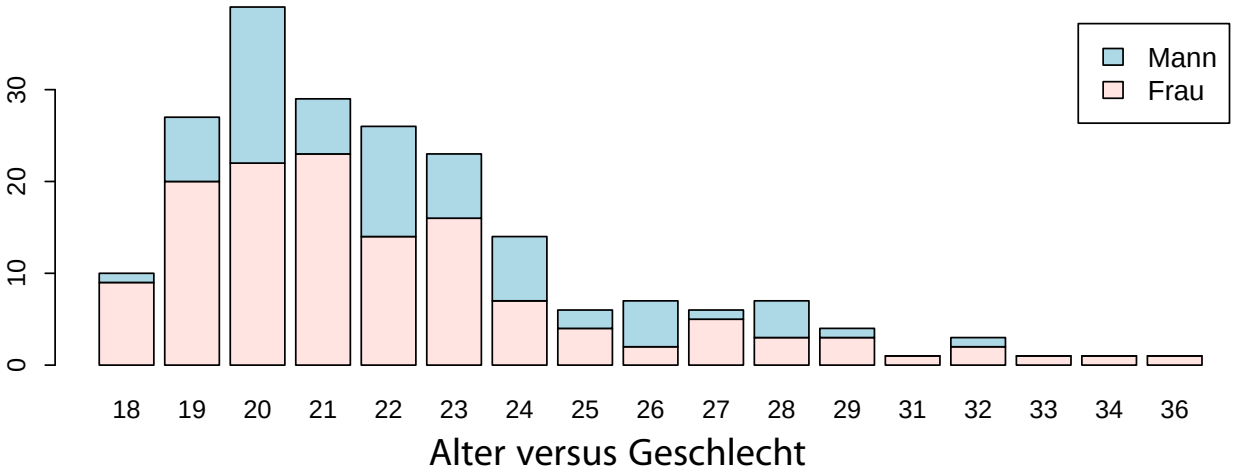


Balkendiagramm, Klassen getrennt oder gestapelt


```
barplot(xtabs(~ Geschlecht + Alter),
        legend=TRUE, beside=TRUE, col=c("mistyrose", "lightblue"))
```




```
barplot(xtabs(~ Geschlecht + Alter),
        legend=TRUE, beside=FALSE, col=c("mistyrose", "lightblue"))
```



❸ Histogramm

- für klassierte Daten
- Fläche proportional zu Häufigkeit:

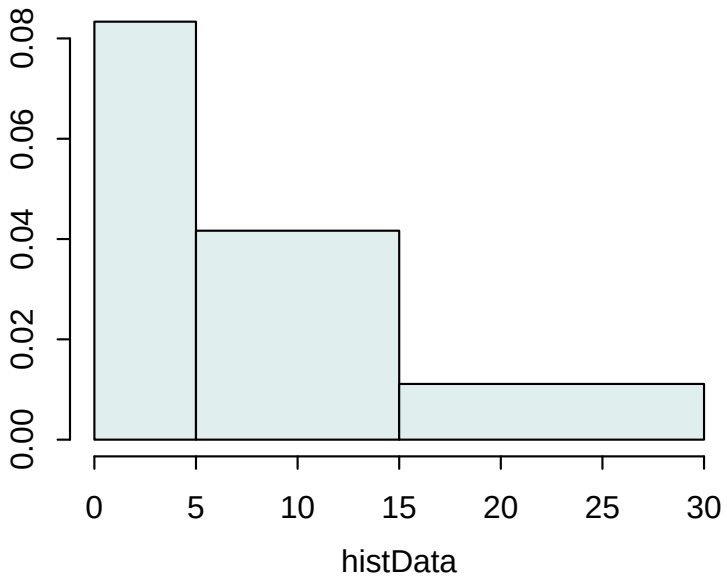
$$\text{Höhe}_j \cdot \text{Breite}_j = c \cdot h(a_j)$$

$$\Rightarrow \quad \text{Höhe}_j = c \cdot \frac{h(a_j)}{\text{Breite}_j}$$

- Im Beispiel mit $c = \frac{1}{12}$:

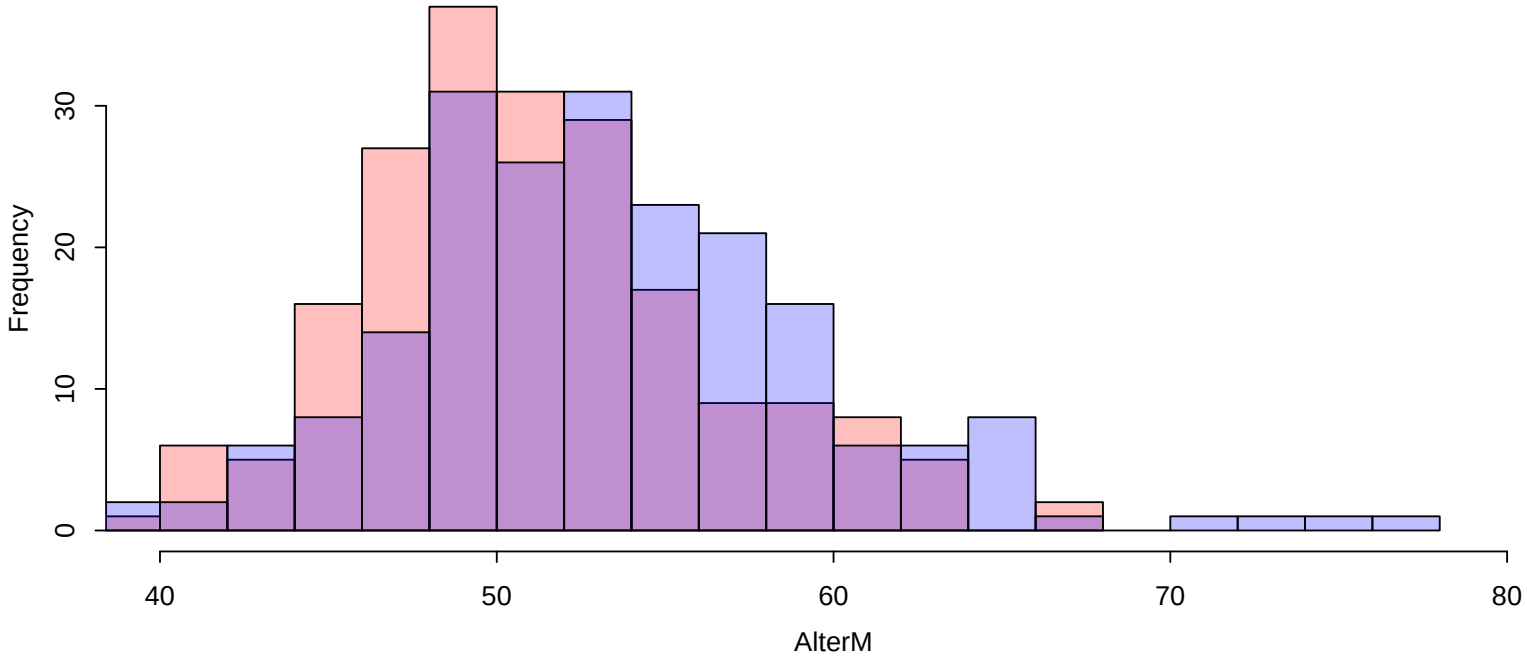
Klasse	$[0;5)$	$[5;15)$	$[15;30]$
$h(a_j)$	5	5	2
Breite_j	5	10	15
Höhe_j	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{90}$

```
histData <- c(0,1,2,3,4,
              5,6,7,10,14,
              15,30)
truehist(histData,
         breaks=c(0, 4.999, 14.999, 30),
         col="azure2", ylab='')
```



Histogramm

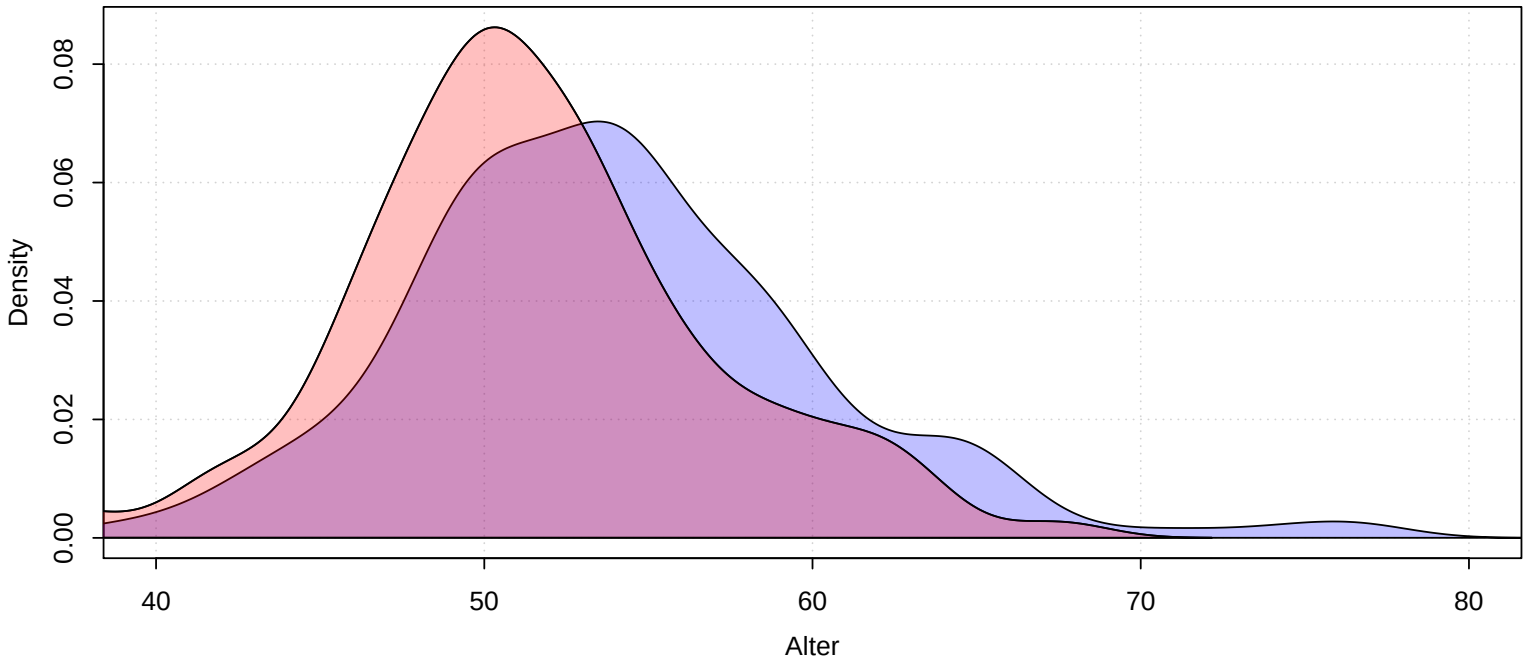
```
plot(hist(AlterM, plot=F, breaks=20),
          col=rgb(1,0,0,1/4), # make red transparent
          main="",
          xlim=c(40,80)) # draw from 40 to 80
plot(hist(AlterV, plot=F, breaks=20),
          col=rgb(0,0,1,1/4),
          add=TRUE)
```



Histogramm: Alter der Väter (blau) und Mütter (rosa)

Dichteplot

```
densMutter = density(AlterM)
densVater = density(AlterV)
plot(densMutter, main="", xlab="Alter",
xlim=c(40,80), # draw from 40 to 80
panel.first=grid()) # draw a grid
polygon(densVater, density=-1, col=rgb(0,0,1,1/4))
polygon(densMutter, density=-1, col=rgb(1,0,0,1/4))
```



Dichteplot: Alter der Väter (blau) und Mütter (rosa)