Statistik

für Betriebswirtschaft, Internationales Management, Wirtschaftsinformatik und Informatik

Sommersemester 2016

Veranstaltungen zur Statistik für BW/IM Sommersemester 2016				
Was?	Wer?	Wann?	Wo?	ab wann?
Vorlesung Statistik	Etschberger	Mi, 14.00-17.00	B2.14	16.03.2016
Vorlesung Statistik PLUS	Etschberger/Jansen	- Blocktermin -	?	?
Übung Statistik	Etschberger	Mi, 17.00-18.30	A1.10	30.03.2016
Übung Statistik	Jansen	Di, 11.30-13.00	W1.06	22.03.2016
Übung Statistik	Jansen	Di, 14.00-15.30	W2.14	22.03.2016
Übung Statistik	Jansen	Mi, 11.30-13.00	W2.11	30.03.2016
Übung Statistik	Jansen	Do, 14.00-15.30	W2.14	31.03.2016
Übung Statistik	Schneller	Do, 14.00-15.30	W3.03	31.03.2016
Übung Statistik	Schneller	Do, 15.30-17.00	W3.03	31.03.2016
Übung Statistik	Wins	Di, 14.00-15.30	J3.19	22.03.2016
Übung Statistik	Wins	Di, 15.30-17.00	J3.13	22.03.2016
Offener Statistikraum	Etschberger/Tutoren	?	?	?
Veranstaltungen für Teilnehmer der WiMa-Klausur im Juli 2016				
Was?	Wer?	Wann?	Wo?	ab wann?
Tutorium Mathematik	Burkart	Do 13.30-15.00	W1.06	07.04.2016
Tutorium Mathematik	Burkart	Do 15.00-16.15	W1.06	07.04.2016
Offener Matheraum	Jansen/Tutoren	?	?	?

	HSA Statistik SS 2016 Sessionlist		
2			
3	Datum	Statistik für IM/BW	Nr.
4	Mittwoch, 16. März 2016	Einführung, R Installation, Rstudio Einführung, Skalen	1
5	Mittwoch, 23. März 2016	univ. deskr. Stat., Quantile, Plots	2
6	Mittwoch, 30. März 2016	Streuung, Konzentrationsmaße	3
7	Mittwoch, 6. April 2016	Kontingenztabellen, Mosaikplots, Korrelation	4
8	Mittwoch, 13. April 2016	Preisindizes, lineare Regression	5
9	Mittwoch, 20. April 2016	Kombinatorik, Wahrscheinlichkeit	6
10	Mittwoch, 27. April 2016	Wahrscheinlichkeit, diskrete Zufallsvariablen	7
11	Mittwoch, 4. Mai 2016	Pyramid	
12	Mittwoch, 11. Mai 2016	Binomial-, Hypergeom.-, Poisson-Verteilung	8
13	Mittwoch, 18. Mai 2016	Stetige ZV, Gleichverteilung	9
14	Mittwoch, 25. Mai 2016	Normalverteilung, Verteilungsparameter	10
15	Mittwoch, 1. Juni 2016	Schätzfunktionen und Punktschätzer	11
16	Mittwoch, 8. Juni 2016	Konfidenzintervalle	12
17	Mittwoch, 15. Juni 2016	Tests	13
18	Mittwoch, 22. Juni 2016	Puffer, WH, Fragen zur Probekl.	14
19	Mittwoch, 29. Juni 2016	AW Prüfungswoche	

Prof. Dr. Stefan Etschberger
Hochschule Augsburg

Outline

Material zur Vorlesung

hs-augsburg.de/~ste

Kursmaterial:

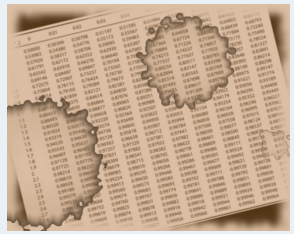
- Aufgabensatz (beinhaltet Aufgaben zu R)
- Handout der Folien
- Alle Folien inklusive Anmerkungen (nach der jeweiligen Vorlesung)
- Beispieldaten
- Alle Auswertungen als **R**-Datei

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.



Prüfung

Klausur:

- Klausur am Ende des Semesters
- Bearbeitungszeit: 90 Minuten
- Erreichbare Punktzahl: 90
- 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~~ 2 Blatt~~,~~ er

Datenbasis

Fragebogen

Umfrage Statistik

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

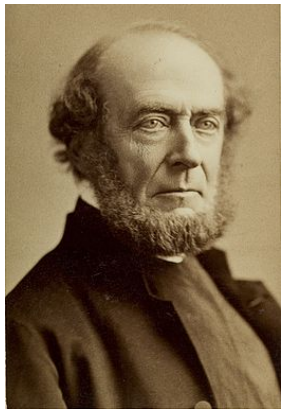
Studiengang									
Ihr Alter (in Jahren)									
Ihre Größe (in cm)									
Ihr Geschlecht (m/w)									
Wie alt ist (bzw. wäre) Ihr Vater heute?									
Wie alt ist (bzw. wäre) Ihre Mutter heute?									
Größe Ihres Vaters (cm)									
Größe Ihrer Mutter (cm)									
Wie viele Geschwister haben 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: (f)ruktarisch, (vegan), (vege)tarisch, (p)escetarisch – also vegetarisch und Fische, (c)arnivor, also auch Fleisch									
Sind Sie Raucher? mögliche Auswahl: (ja), (nein)									
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.“

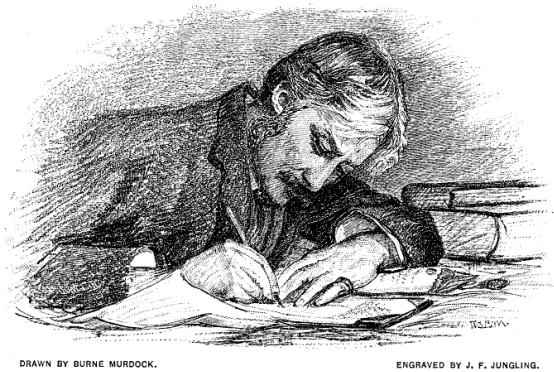


- **Winston Curchill (1874-1965) angeblich**:

  „ 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 14,28%
- 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“.

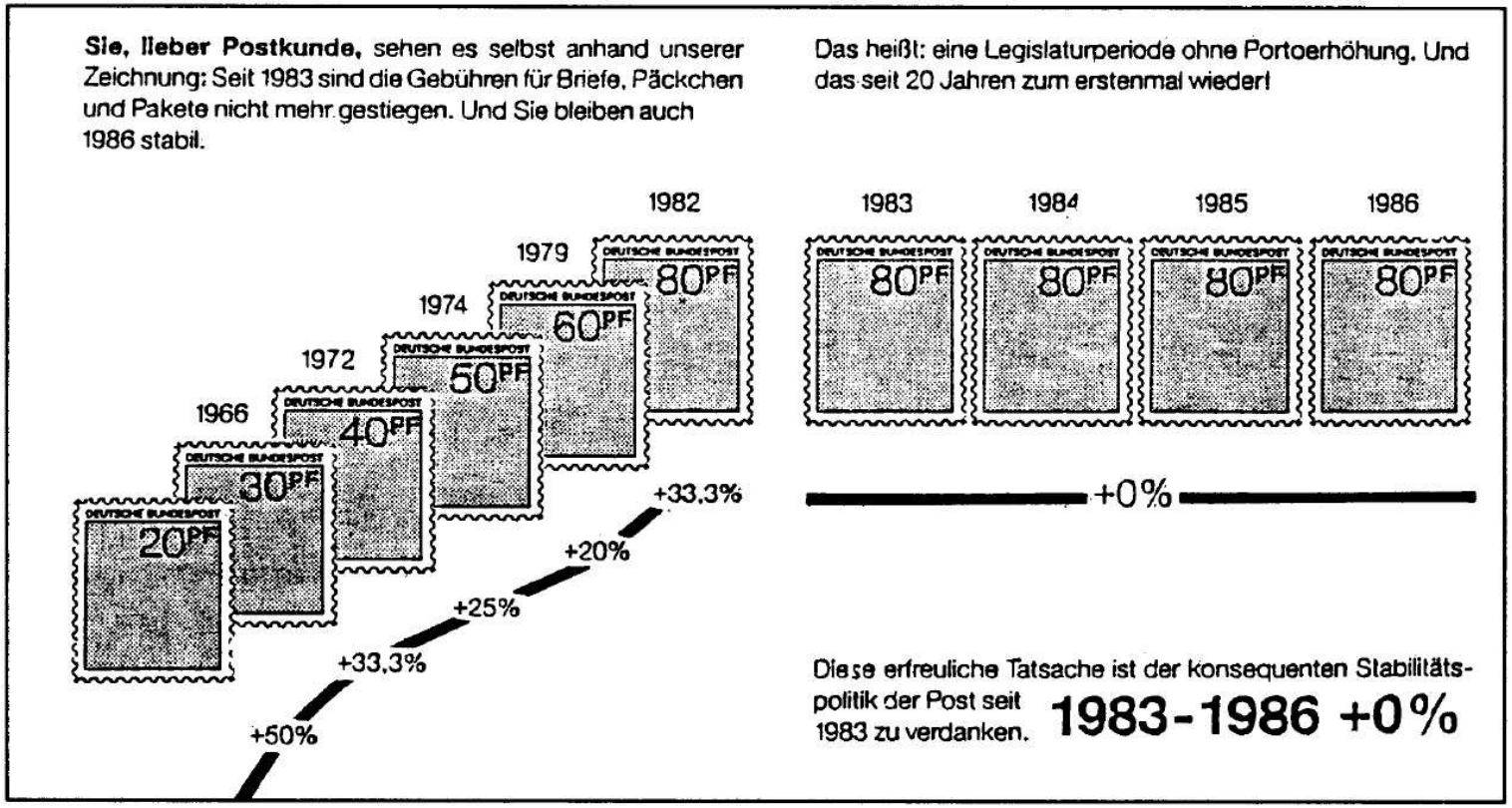
Aussage des Vertriebsleiters:

„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



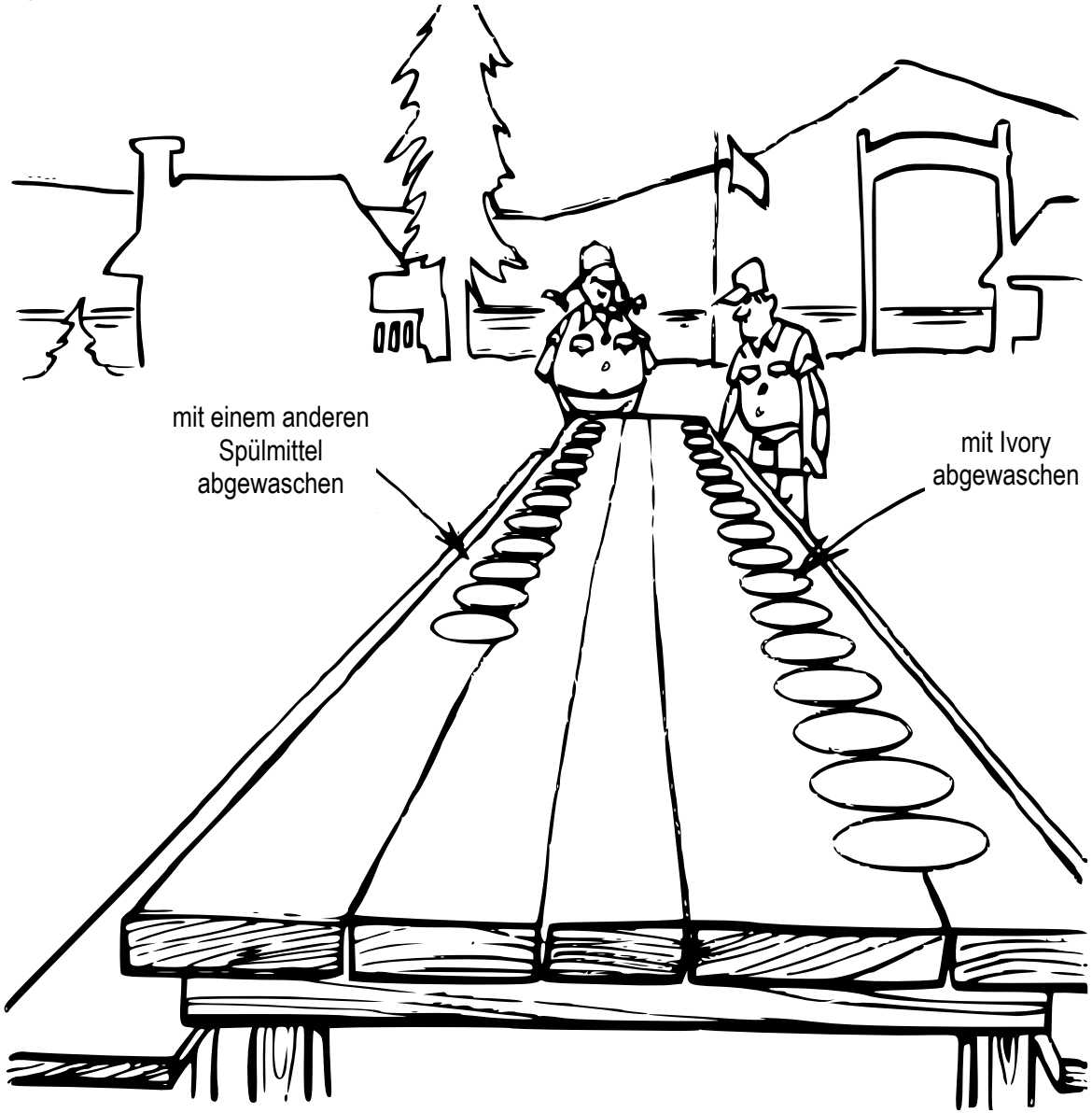
Quelle Kramer (2011)

Grafik aussagekräftig?

Quelle: Bach u. a. (2006)

Größenproportionen

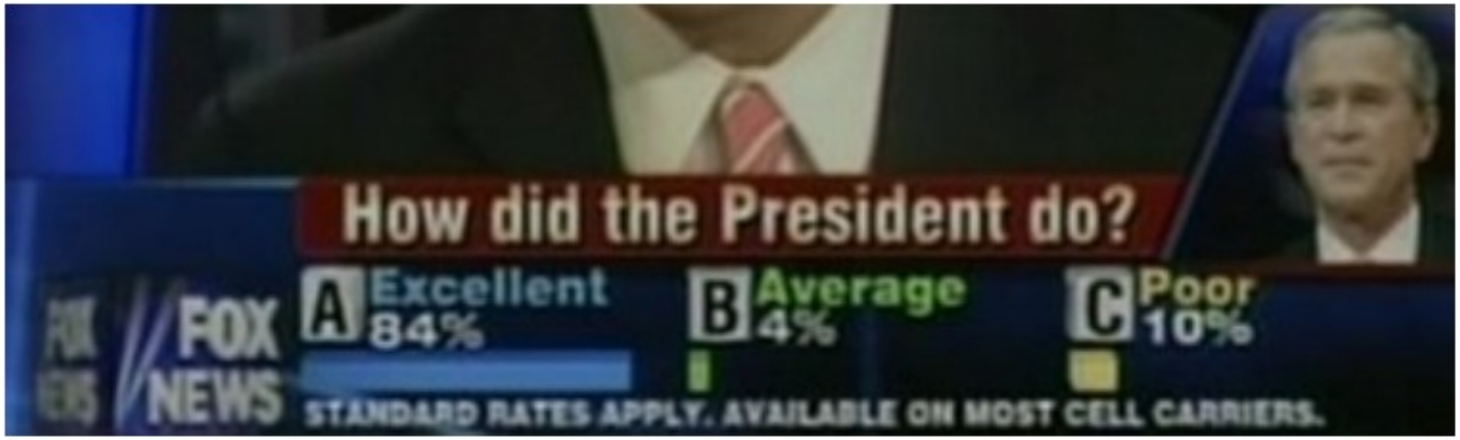
11 zu 15



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



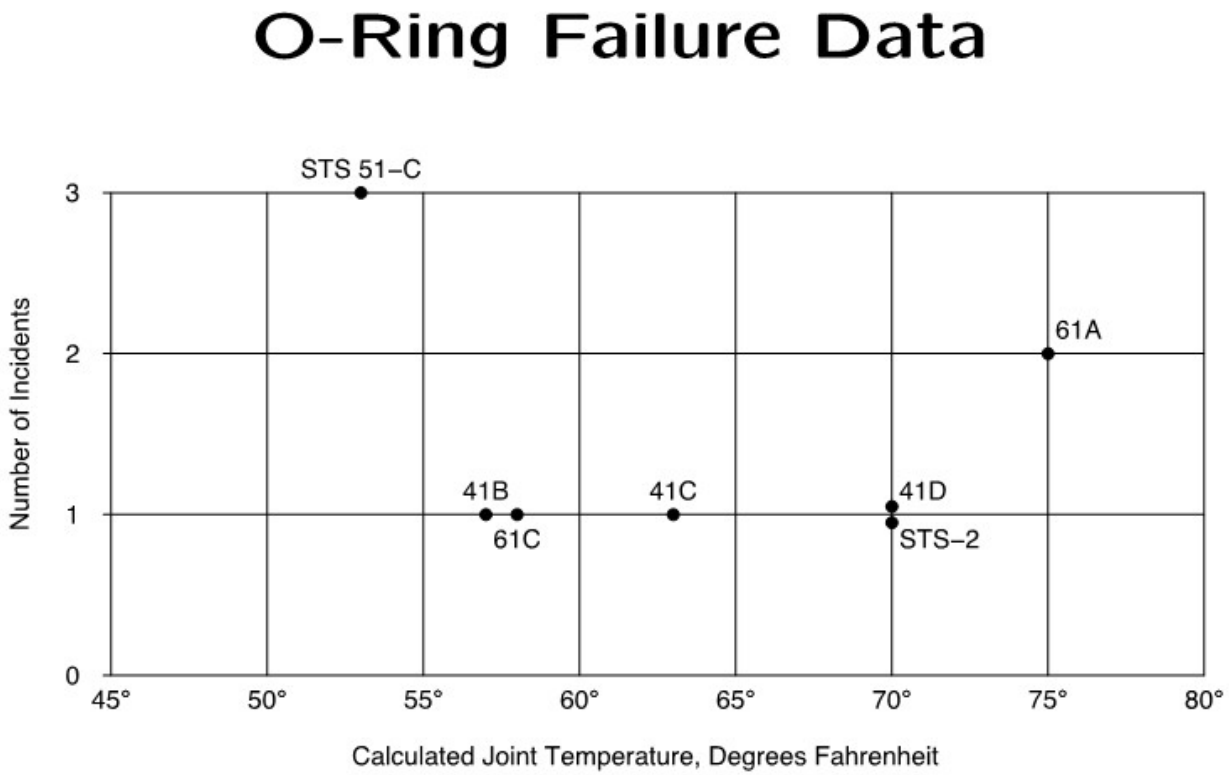
- Kostenpflichtige Telefonabstimmung nach regierungsfreundlichem Bericht im Fernsehen
- In den meisten Umfragen erreichte Bush zu diesem Zeitpunkt nur 30 % Zustimmung

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.

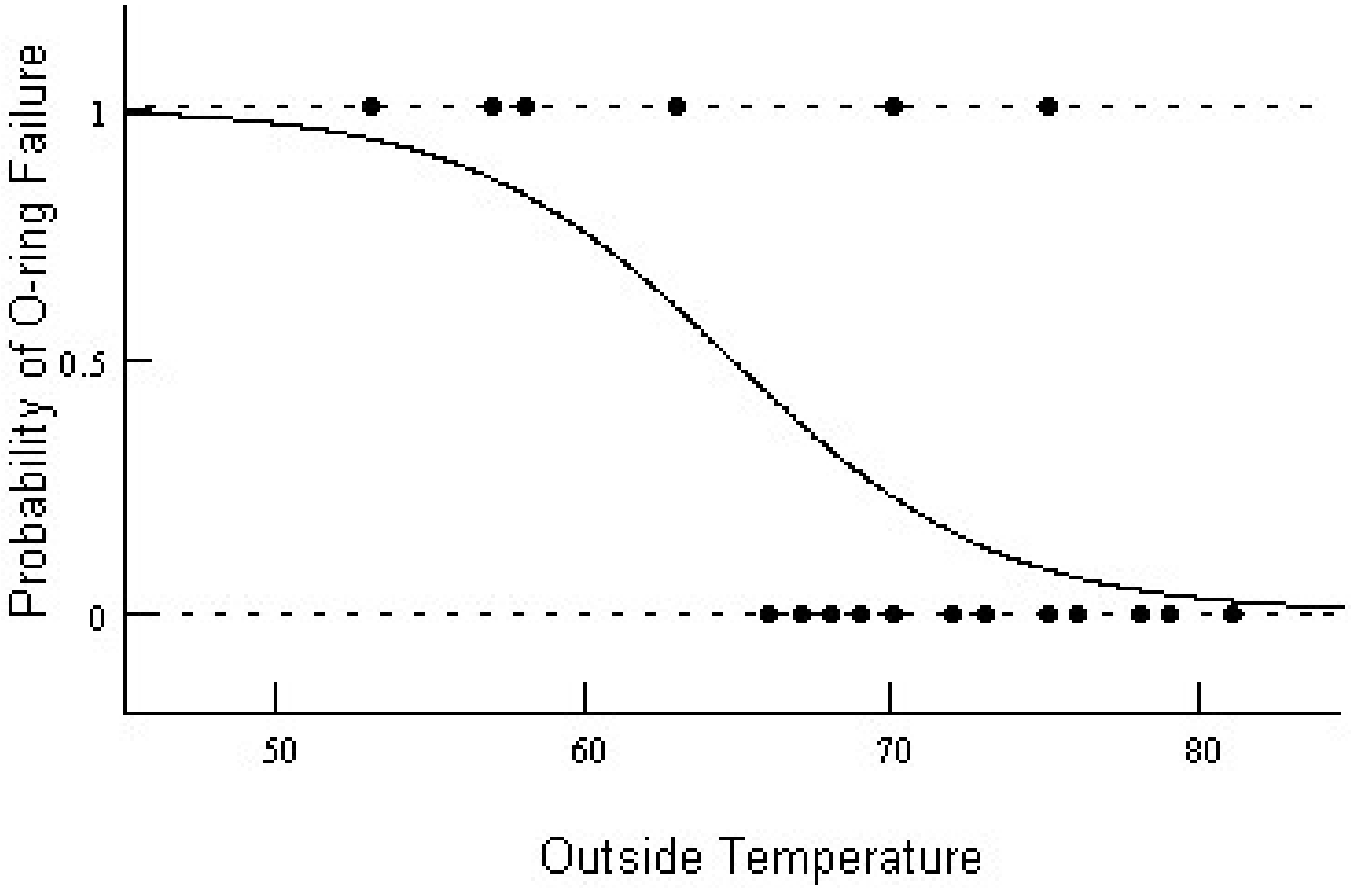
Challenger Unfall 1986

- 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?

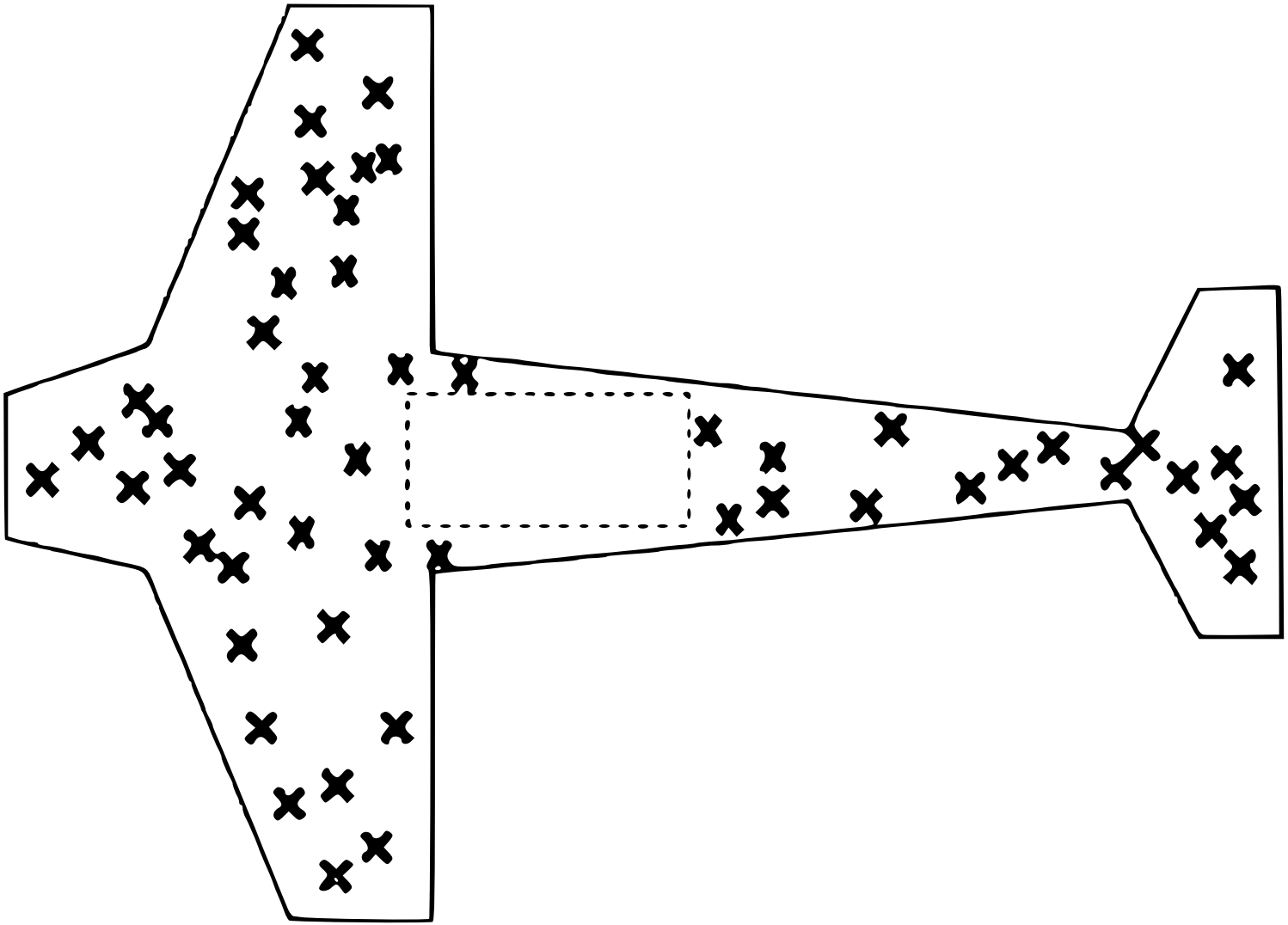


Challenger Unfall 1986

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

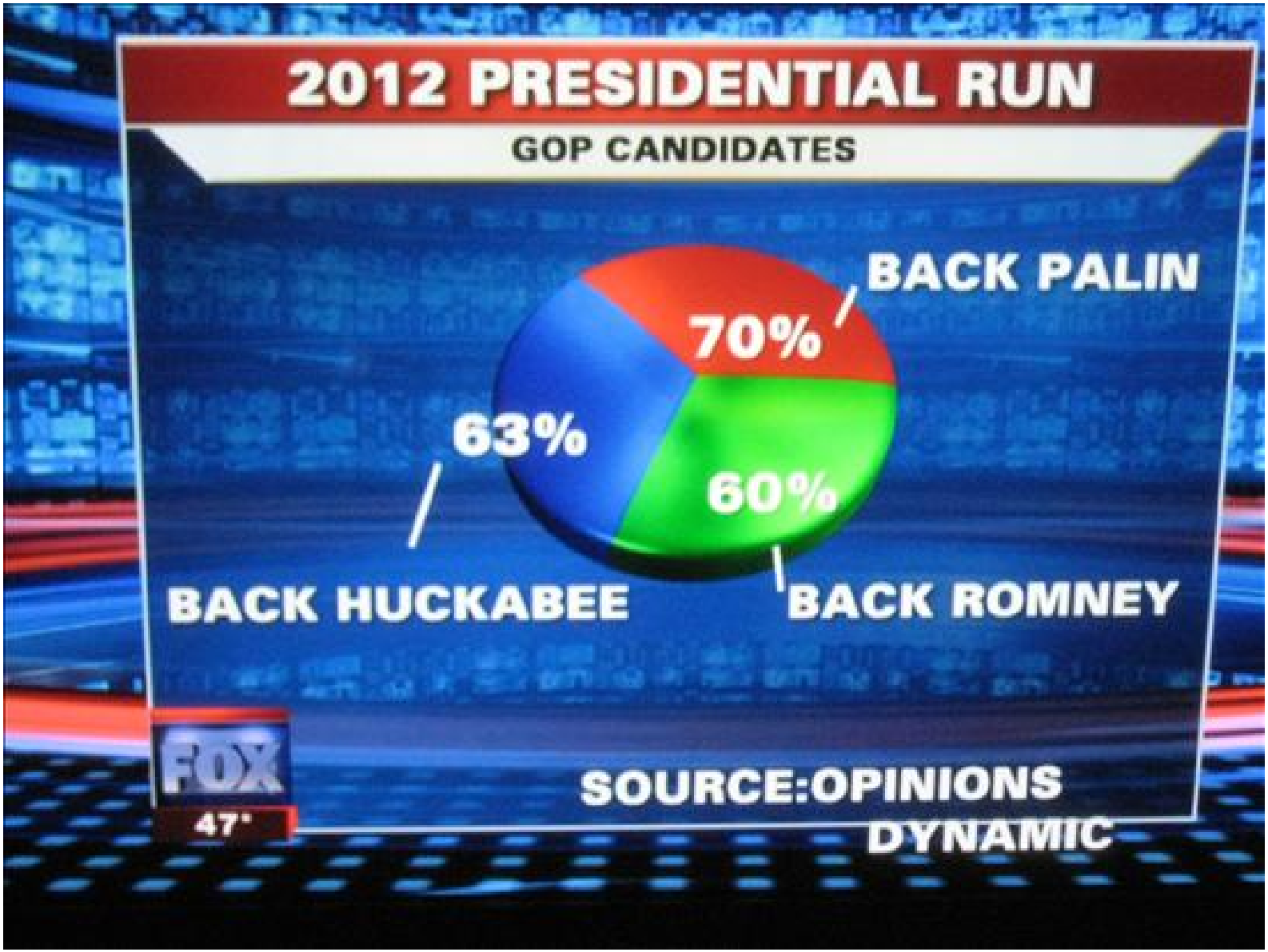


Auswahl der Stichprobe



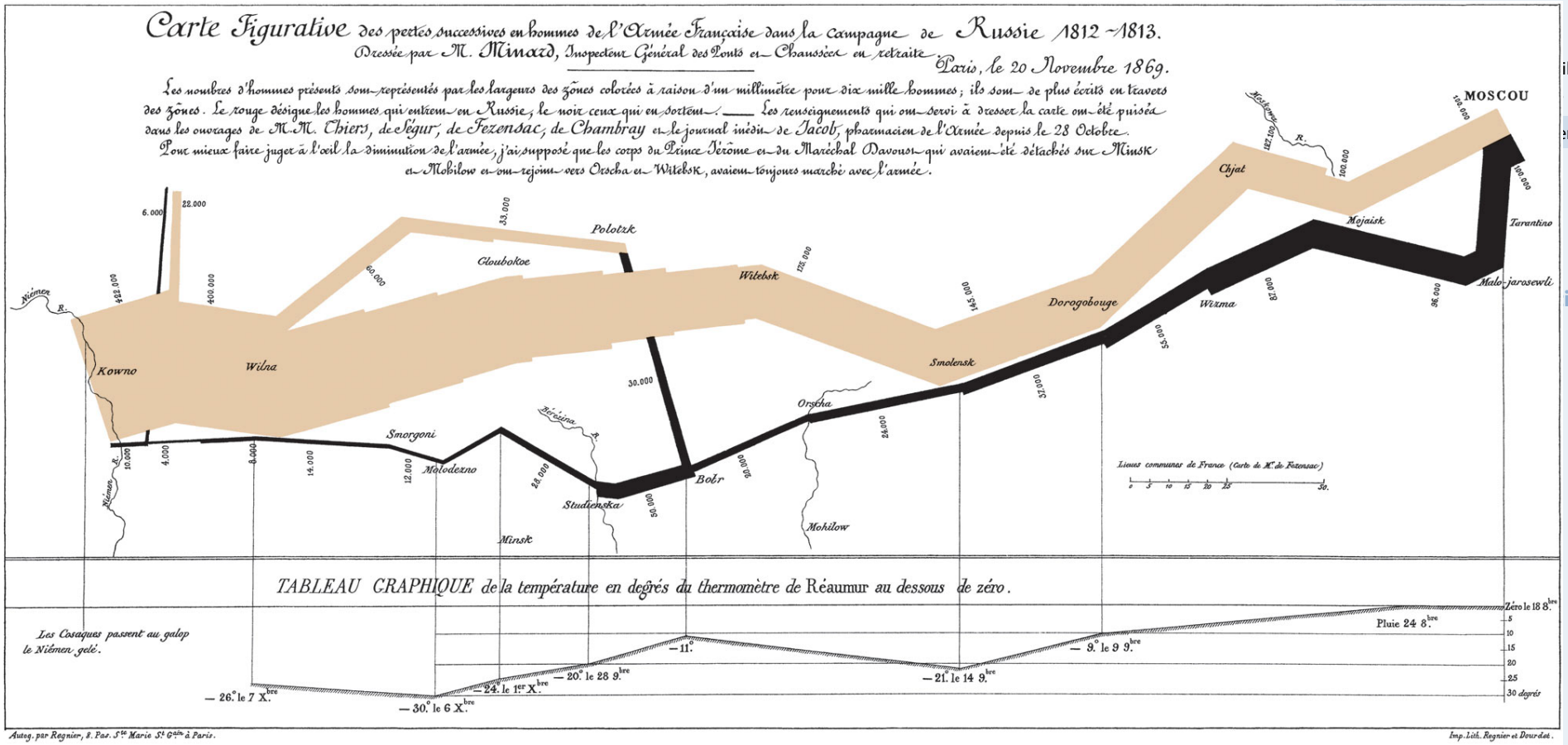
Kandidat für die schlechteste Grafik aller Zeiten

Aussage?



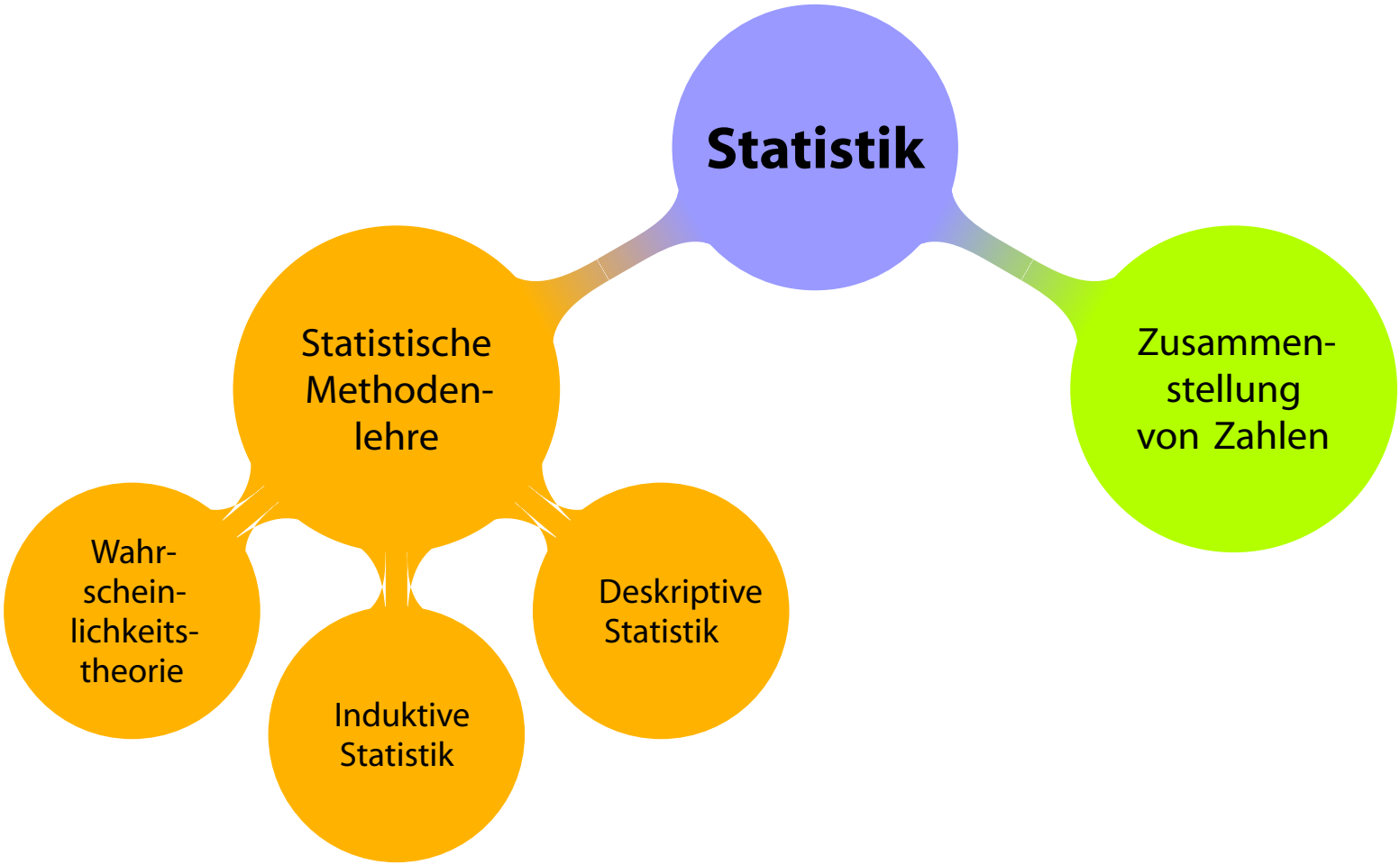
Quelle: Friendly, Michael (2001), Gallery of Data Visualization, `http://www.datavis.ca/gallery/`, Stand November 2014

Minards Grafik von 1869 über Napoleons Rußlandfeldzug



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

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, cases

- Merkmalsträger: Untersuchte statistische Einheit
- Merkmal: Interessierende Eigenschaft des Merkmalträgers
- (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)

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

Nominalskala:

- Zahlen haben nur Bezeichnungsfunktion
- z.B. Artikelnummern

Ordinalskala:

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

metrische Merkmale

Kardinalskala:

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

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

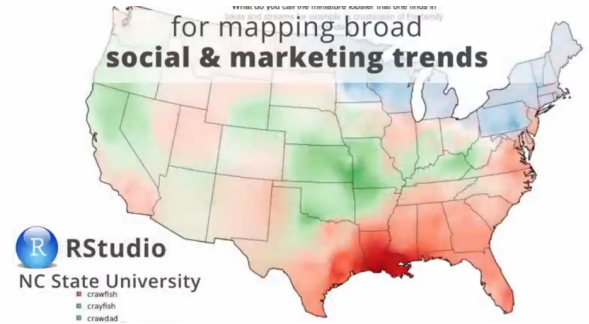
Aber:

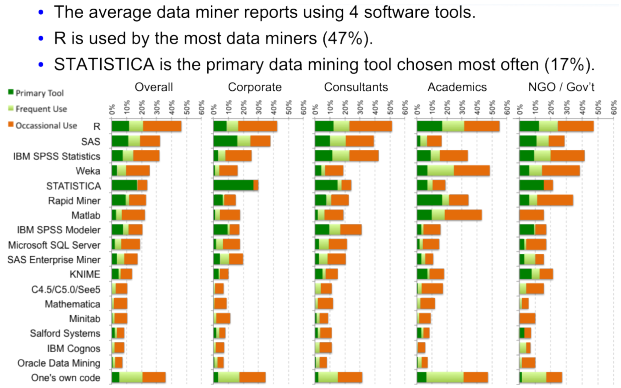
- Nominale Merkmale dürfen **nicht** ordinal- oder metrisch skaliert werden.
- Ordinale Merkmale dürfen **nicht** metrisch skaliert werden.

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

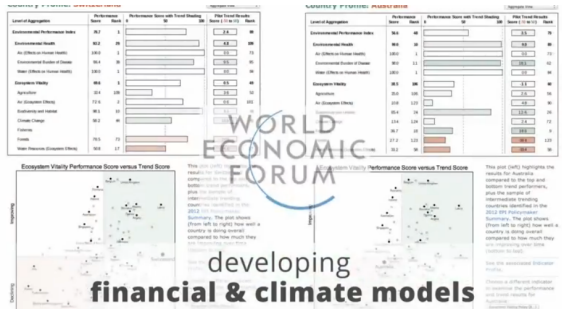
Was ist R und warum soll man es benutzen?

- 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)





source: http://goo.gl/axhGhh



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

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

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

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

```
> 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)
>
```

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

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`

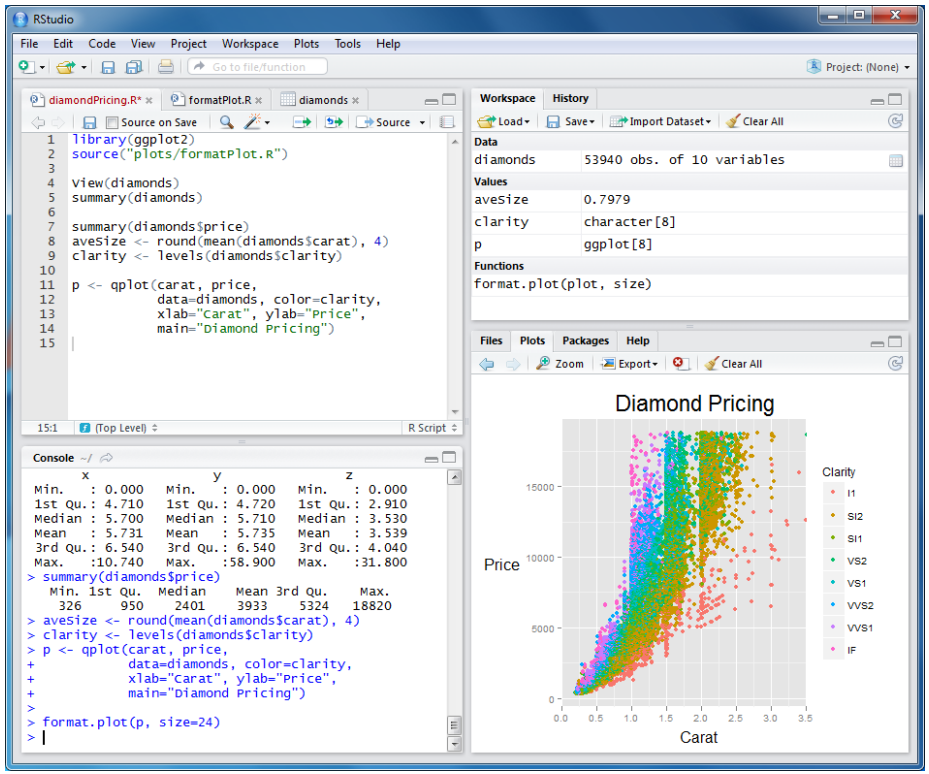
graphics source: `http://goo.gl/W70kms`

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`

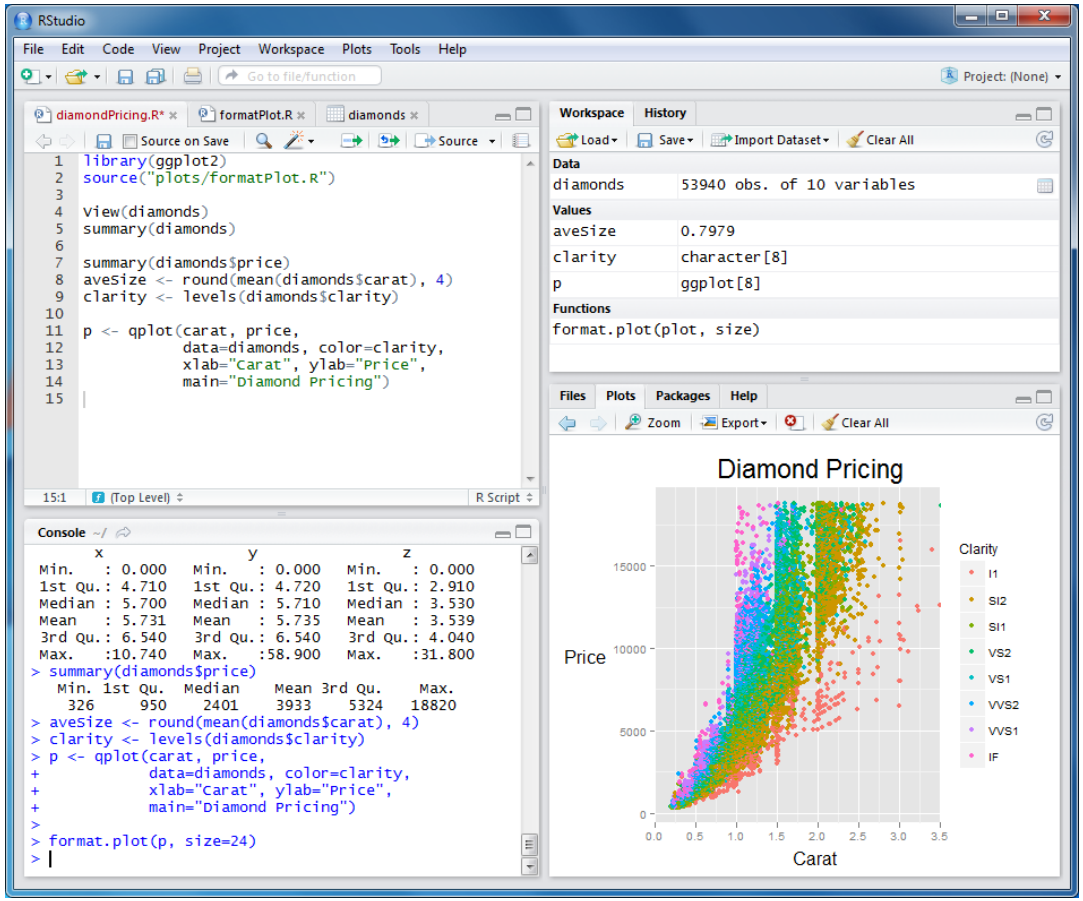
RStudio™

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/2015SS_HSA_Statistik")

# Daten einlesen aus einer csv-Datei (Excel)
MyData = read.csv2(file="../_genericFiles/Daten/Umfrage_HSA_2015_03.csv", header=TRUE)
```

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

```
## 'data.frame': 377 obs. of  17 variables:
##  $ Jahrgang       : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
##  $ Alter          : int  20 25 19 21 25 20 25 20 23 21 ...
##  $ Groesse        : int  174 157 163 185 178 170 165 175 180 161 ...
##  $ Geschlecht     : Factor w/ 2 levels "Frau","Mann": 1 1 1 2 2 1 1 2 2 1 ...
##  $ AlterV         : int  55 54 51 52 60 50 60 52 56 70 ...
##  $ AlterM         : int  53 61 49 50 63 55 60 49 50 55 ...
##  $ GroesseV       : int  187 185 178 183 170 183 185 175 175 180 ...
##  $ GroesseM       : int  169 160 168 165 160 160 170 169 170 165 ...
##  $ Geschwister    : int  3 1 1 4 2 2 4 1 1 2 ...
##  $ Farbe          : Factor w/ 6 levels "blau","gelb",..: 4 6 4 4 1 6 1 6 4 4 ...
##  $ AusgKomm       : num  240 119 270 40 550 ...
##  $ AnzSchuhe      : int  25 30 25 6 5 65 10 7 10 22 ...
##  $ AusgSchuhe     : int  450 300 100 100 80 250 150 400 150 300 ...
##  $ Essgewohnheiten: Factor w/ 5 levels "carnivor","fruktarisch",..: 1 1 1 1 1 1 5 1 1 1 ...
##  $ Raucher        : Factor w/ 2 levels "ja","nein": NA 2 2 2 1 2 2 2 2 1 ...
##  $ NoteMathe      : num  2.3 3.3 1.7 2 4 4 3.3 2.7 3.7 3.3 ...
##  $ MatheZufr      : Ord.factor w/ 4 levels "unzufrieden"<..: 2 2 2 2 2 2 2 2 2 2 ...
```

Erste Zeilen der Datentabelle

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

```
##   Jahrgang Alter Groesse Geschlecht AlterV AlterM GroesseV GroesseM Geschwister   Farbe AusgKomm
## 1     2015    20     174       Frau     55     53      187      169           3 schwarz    240.0
## 2     2015    25     157       Frau     54     61      185      160           1   weiss    119.4
## 3     2015    19     163       Frau     51     49      178      168           1 schwarz    270.0
## 4     2015    21     185       Mann     52     50      183      165           4 schwarz     40.0
## 5     2015    25     178       Mann     60     63      170      160           2    blau    550.0
## 6     2015    20     170       Frau     50     55      183      160           2   weiss    420.0
##   AnzSchuhe AusgSchuhe Essgewohnheiten Raucher NoteMathe MatheZufr
## 1        25        450        carnivor    <NA>       2.3   geht so
## 2        30        300        carnivor    nein       3.3   geht so
## 3        25        100        carnivor    nein       1.7   geht so
## 4         6        100        carnivor    nein       2.0   geht so
## 5         5         80        carnivor      ja       4.0   geht so
## 6        65        250        carnivor    nein       4.0   geht so
```

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

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

```
## [1] 377
```

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

```
## [1] 17
```


Daten kennenlernen


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

##   Alter Groesse Geschlecht AlterV
## 1    20     174       Frau     55
## 2    25     157       Frau     54
## 3    19     163       Frau     51
```

```
# Auswahl von Objekten über logische Ausdrücke
Auswahl = (MyData$Geschlecht=="Mann" & MyData$Alter < 19)
# zeige die ersten Einträge
head(Auswahl, 30)

##  [1] 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

# Ausgabe der Auswahl: Alter, Alter des Vaters und der Mutter
MyData[Auswahl,                        # Objektauswahl
      c("Alter", "AlterM", "AlterV")]  # Welche Merkmale?

##     Alter AlterM AlterV
## 23     18     44     48
## 268    18     46     52
```

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
## 19     20           0   weiss        300      924
## 23     18           1  silber        300     1000
## 33     25           1 schwarz        300     1000
## 42     24           1 schwarz       1000      600
## 81     25           2  silber        200     1900
## 106    21           1 schwarz        200      860
## 121    22           0  silber        300     1100
## 142    20           2 schwarz        290     1570
## 161    19           1 schwarz        600      800
## 168    21           2    blau        600      505
## 179    21           0  silber        300      825
## 211    23           1 schwarz        450      630
## 223    20           1     rot        400      815
## 227    20           1 schwarz        200     1250
## 249    20           1    blau       1000      350
## 256    25           0 schwarz        280     1200
## 272    24           1 schwarz        300      900
## 281    19           2 schwarz        500      720
## 315    21           1   weiss        200     1300
## 353    20           0 schwarz        400      950
```