

Statistik

für Betriebswirtschaft, Internationales Management,
Wirtschaftsinformatik und Informatik

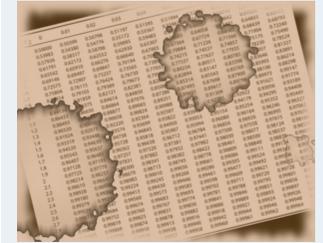
Sommersemester 2017

HSA Statistik SS 2017 Sessionlist		
Datum	Statistik für BW/IM/I/Winf	Nr.
15.03.2017	Einführung Statistik	1
22.03.2017	Differentialrechnung, 2-dim Diff.Rechnung	2
29.03.2017	univ. deskr. Stat., Quantile, Plots	3
05.04.2017	Streuung, Konzentrationsmaße	4
12.04.2017	Kontingenztabellen, Mosaikplots, Korrelation	5
19.04.2017	Preisindizes, lineare Regression	6
26.04.2017	Wahrscheinlichkeitsbegriff	7
03.05.2017	Bedingte Wahrscheinlichkeit, Bayes	8
10.05.2017	diskrete Zufallsvariablen	9
17.05.2017	Stetige ZV, Gleichverteilung	10
24.05.2017	Pyramid	
31.05.2017	Normalverteilung, Verteilungsparameter	11
07.06.2017	Schätzfunktionen und Punktschätzer	12
14.06.2017	Konfidenzintervalle	13
21.06.2017	Wiederholung, Besprechung Probeklausur	14
28.06.2017	Prüfungswoche	15

Mathe PLUS: Interesse?

<https://goo.gl/EfsIDY>

Auswertung nächste Woche in VL:
falls weniger als 7 Interessenten
(mit Prüfungsabsicht) findet Kurs nicht statt



1 Statistik: Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung
R und RStudio

2 Differenzieren 2

Partielle Ableitung
Kurvendiskussion
Optimierung mit
Nebenbedingungen

3 Deskriptive Statistik

Häufigkeiten
Lage und Streuung
Konzentration
Zwei Merkmale
Korrelation
Preisindizes
Lineare Regression

4 Wahrscheinlichkeitstheorie

Kombinatorik
Zufall und Wahrscheinlichkeit
Zufallsvariablen und Verteilungen
Verteilungsparameter

5 Induktive Statistik

Grundlagen
Punkt-Schätzung
Intervall-Schätzung
Signifikanztests

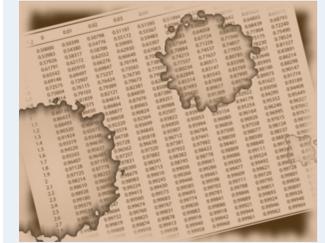
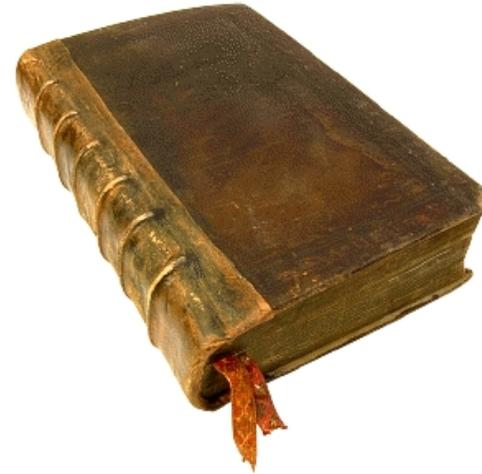
1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
5. Induktive Statistik

Quellen

Tabellen

Kursmaterial:

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



1. Einführung
2. Differenzieren 2
3. Deskriptive Statistik
4. W-Theorie
5. Induktive Statistik

Quellen

Tabellen

Literatur:

-  Bamberg, Günter, Franz Baur und Michael Krapp (2017). **Statistik: Eine Einführung für Wirtschafts- und Sozialwissenschaftler**. 18. voll aktualisierte Auflage. De Gruyter Oldenbourg.
-  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.

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 Ausdrücke),
- ▶ 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)

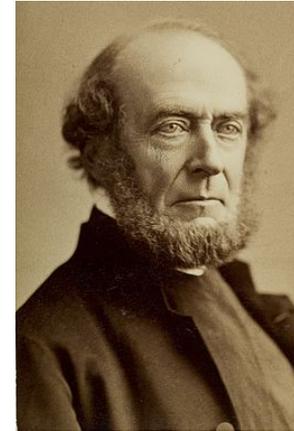
- 1 Statistik: Einführung
- 2 Differenzieren 2
- 3 Deskriptive Statistik
- 4 Wahrscheinlichkeitstheorie
- 5 Induktive Statistik



- 1 Statistik: Einführung
 - Berühmte Leute zur Statistik
 - Wie lügt man mit Statistik?
 - Gute und schlechte Grafiken
 - Begriff Statistik
 - Grundbegriffe der Datenerhebung
 - R und RStudio

► Leonard Henry Courteney (1832-1918):

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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

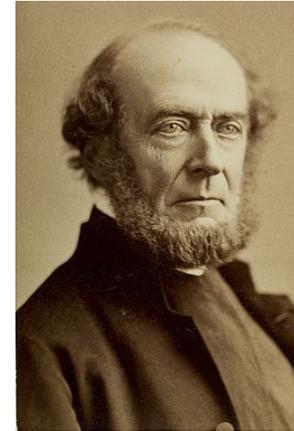
5. Induktive Statistik

Quellen

Tabellen

► **Leonard Henry Courtney
(1832-1918):**

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



► **Winston Churchill (1874-1965)
angeblich:**

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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

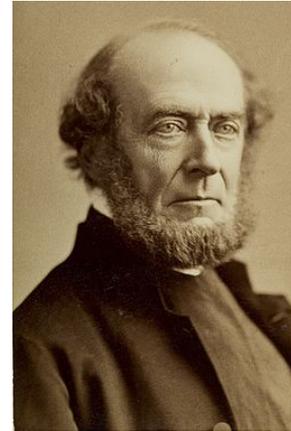
5. Induktive Statistik

Quellen

Tabellen

► **Leonard Henry Courtney (1832-1918):**

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



► **Winston Churchill (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.“



DRAWN BY BURNE MURDOCK.

ENGRAVED BY J. F. JUNGLING.



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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
- ▶ Tennisspieler 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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Frage:

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

Resultat:

- ▶ Dafür:
- ▶ Dagegen:
- ▶ Unentschieden:

Schätzen Sie das Ergebnis!

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:
- ▶ Dagegen:
- ▶ Unentschieden:

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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%

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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

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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Definition

Zu Kaiser Wilhelms Zeiten:

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

Definition heute:

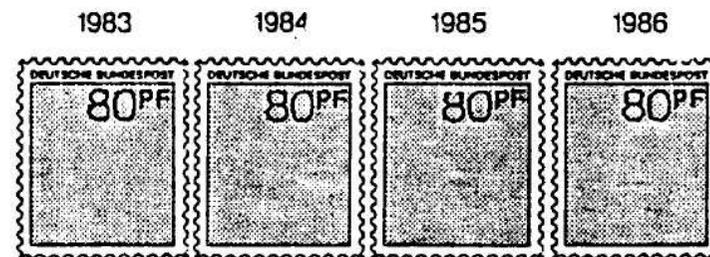
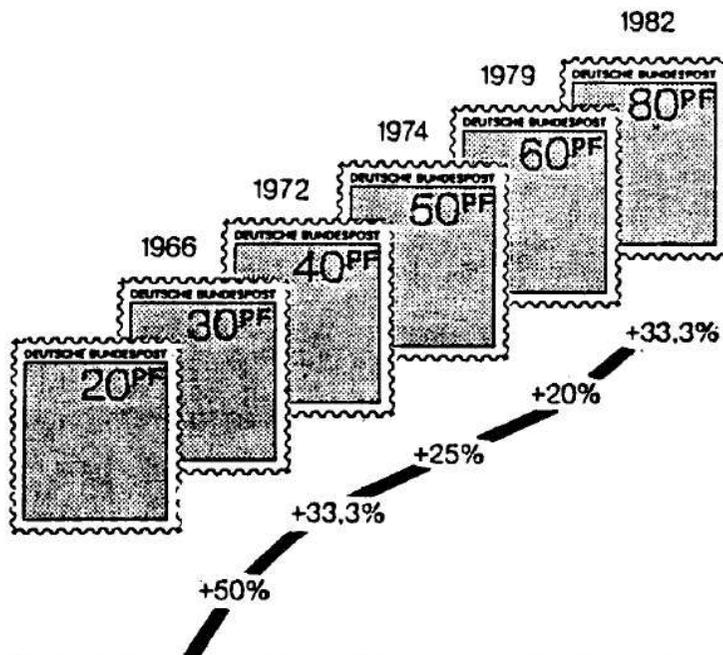
„Ein Analphabet ist eine Person, die sich nicht beteiligen kann an all den zielgerichteten Aktivitäten ihrer Gruppe und ihrer Gemeinschaft, bei denen Lesen, Schreiben und Rechnen erforderlich ist und an der weiteren Nutzung dieser Kulturtechniken für ihre weitere Entwicklung und die der Gesellschaft“.



Seit 1983 stabile Gebühren

Sie, lieber Postkunde, sehen es selbst anhand unserer Zeichnung: Seit 1983 sind die Gebühren für Briefe, Päckchen und Pakete nicht mehr gestiegen. Und Sie bleiben auch 1986 stabil.

Das heißt: eine Legislaturperiode ohne Portoerhöhung. Und das seit 20 Jahren zum erstenmal wieder!



— +0% —

Diese erfreuliche Tatsache ist der konsequenten Stabilitäts-
politik der Post seit **1983-1986 +0%**
1983 zu verdanken.

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

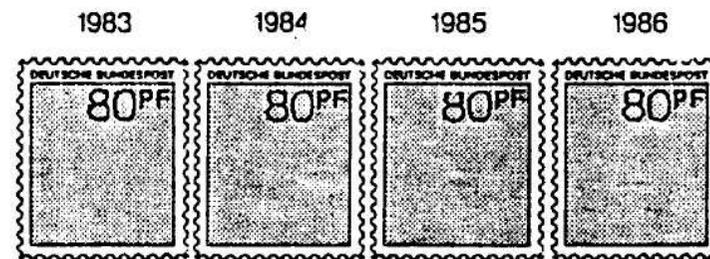
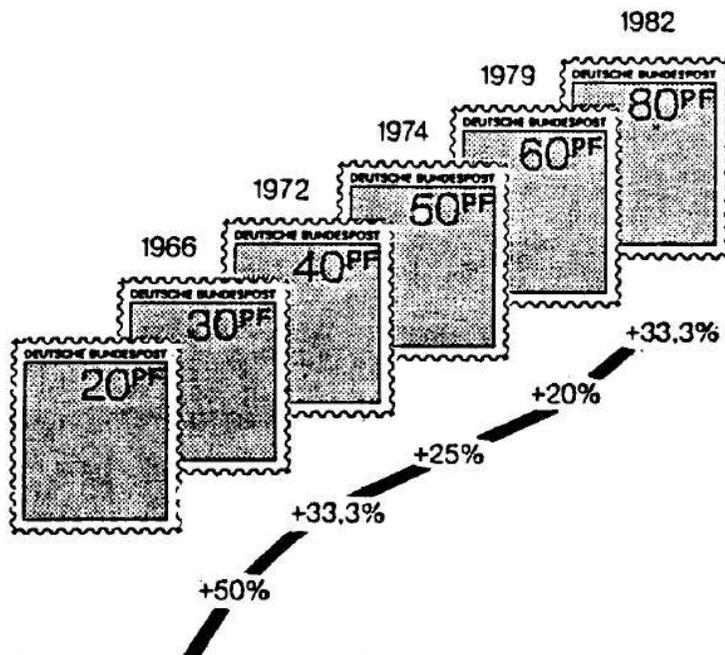
Quelle Kramer, (2011)



Seit 1983 stabile Gebühren

Sie, lieber Postkunde, sehen es selbst anhand unserer Zeichnung: Seit 1983 sind die Gebühren für Briefe, Päckchen und Pakete nicht mehr gestiegen. Und Sie bleiben auch 1986 stabil.

Das heißt: eine Legislaturperiode ohne Portoerhöhung. Und das seit 20 Jahren zum erstenmal wieder!



+0%

Diese erfreuliche Tatsache ist der konsequenten Stabilitäts-politik der Post seit 1983 zu verdanken. **1983-1986 +0%**

Quelle Kramer, (2011)

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



Grafik aussagekräftig?



Quelle: Bach u. a., (2006)

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Grafik aussagekräftig?



Quelle: Bach u. a., (2006)



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

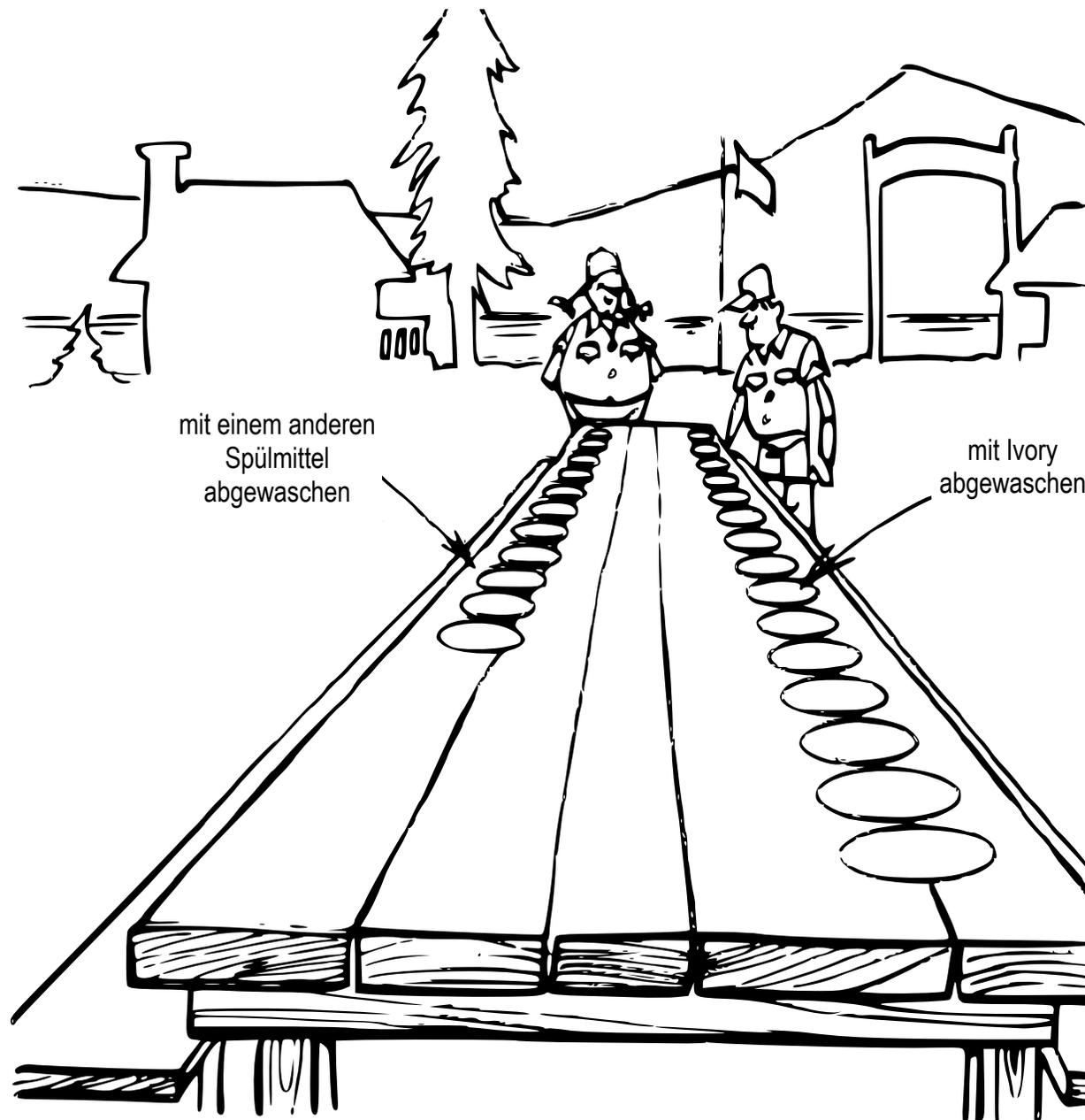
3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

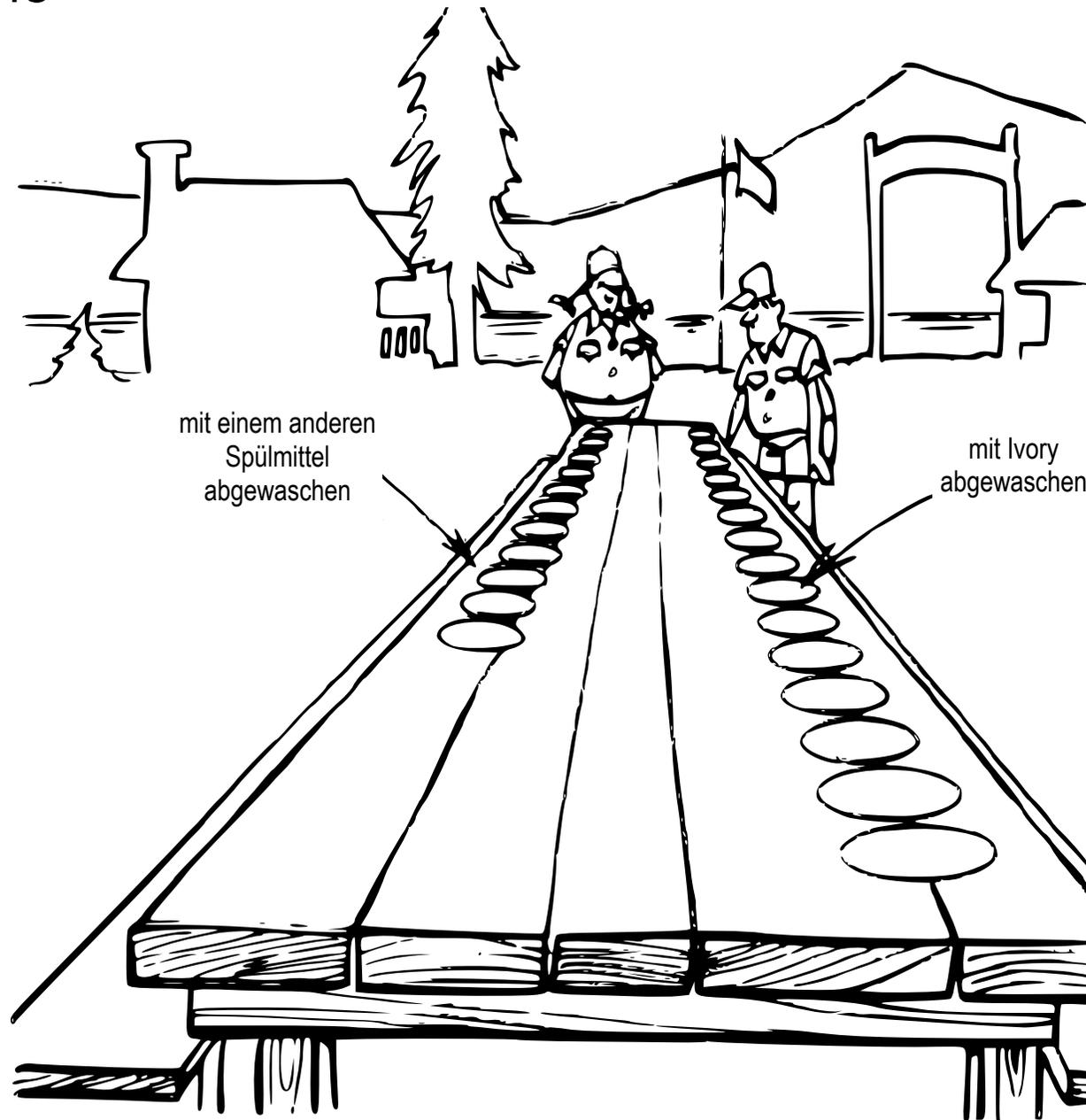
4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

11 zu 15



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

- ▶ Ein Einzelhändler bezieht ein Produkt zu 100 € und verkauft es für 200 €. Hat er eine Gewinnspanne von 50% oder 100%?



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



- ▶ 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

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

- ▶ 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 Sicherheitsgurt angelegt
Also: Keinen Gurt anlegen ist sicherer

- ▶ 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 Sicherheitsgurt angelegt
Also: Keinen Gurt anlegen ist sicherer
- ▶ Die Hälfte der Todesfälle ereignen sich in Krankenhäusern
Also: Krankenhäuser sind lebenssgefährlich



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

- ▶ 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 Sicherheitsgurt 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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

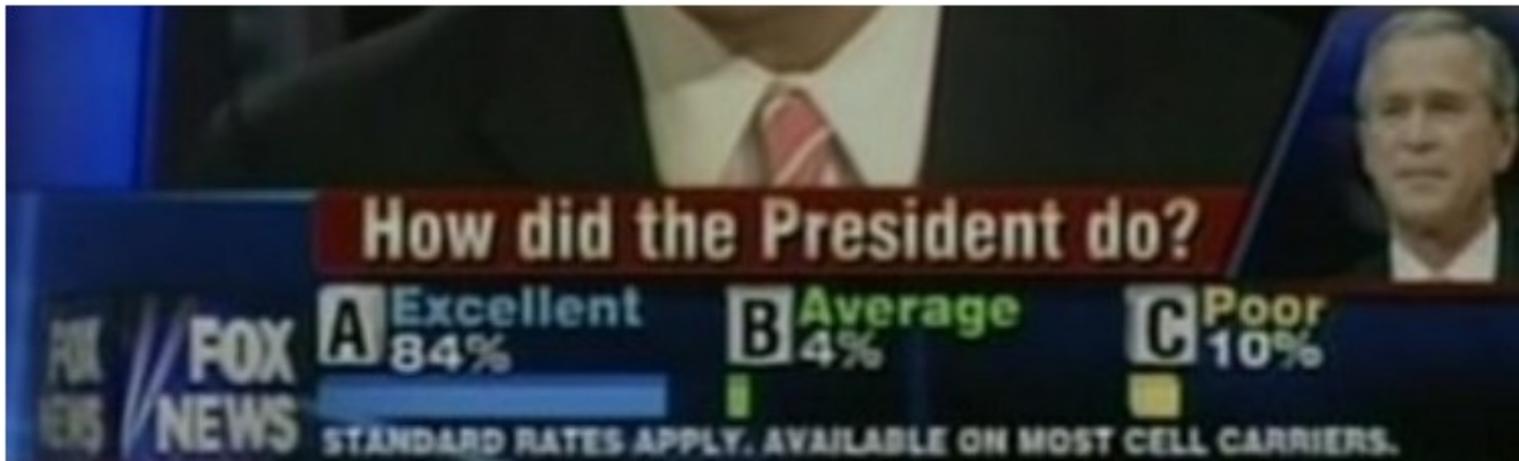
4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Fernsehumfragen



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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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.



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

- ▶ 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?



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

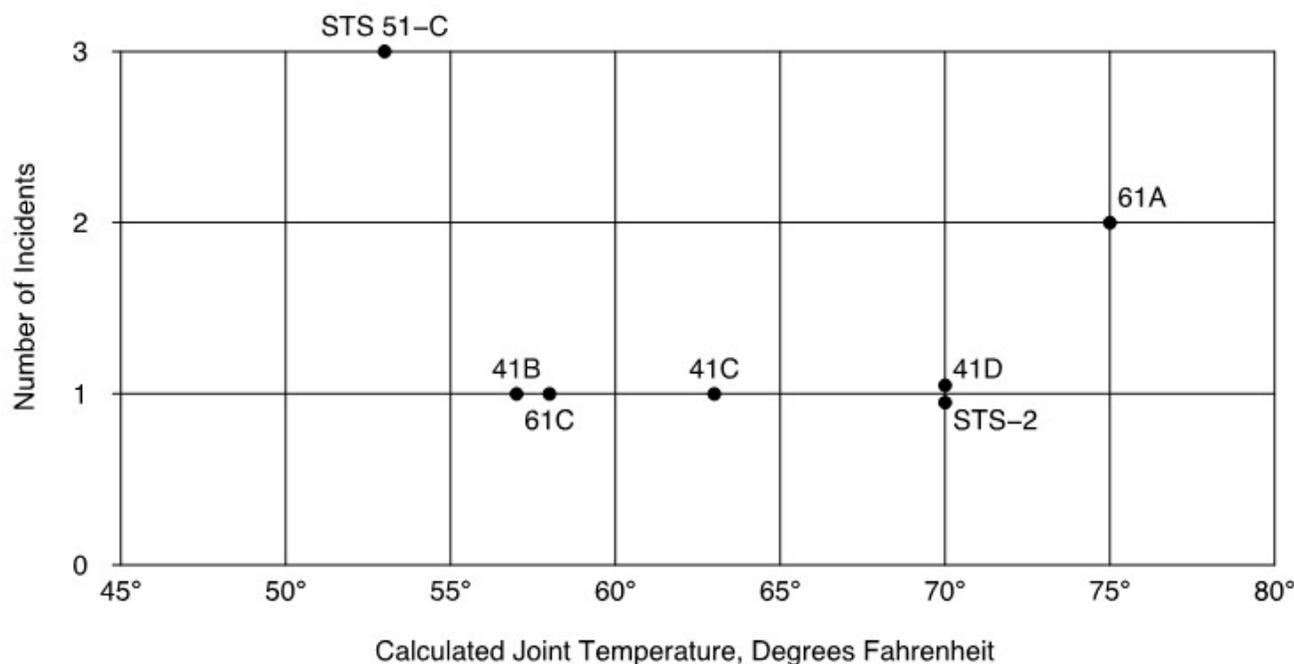
5. Induktive Statistik

Quellen

Tabellen

- ▶ 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?

O-Ring Failure Data



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



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

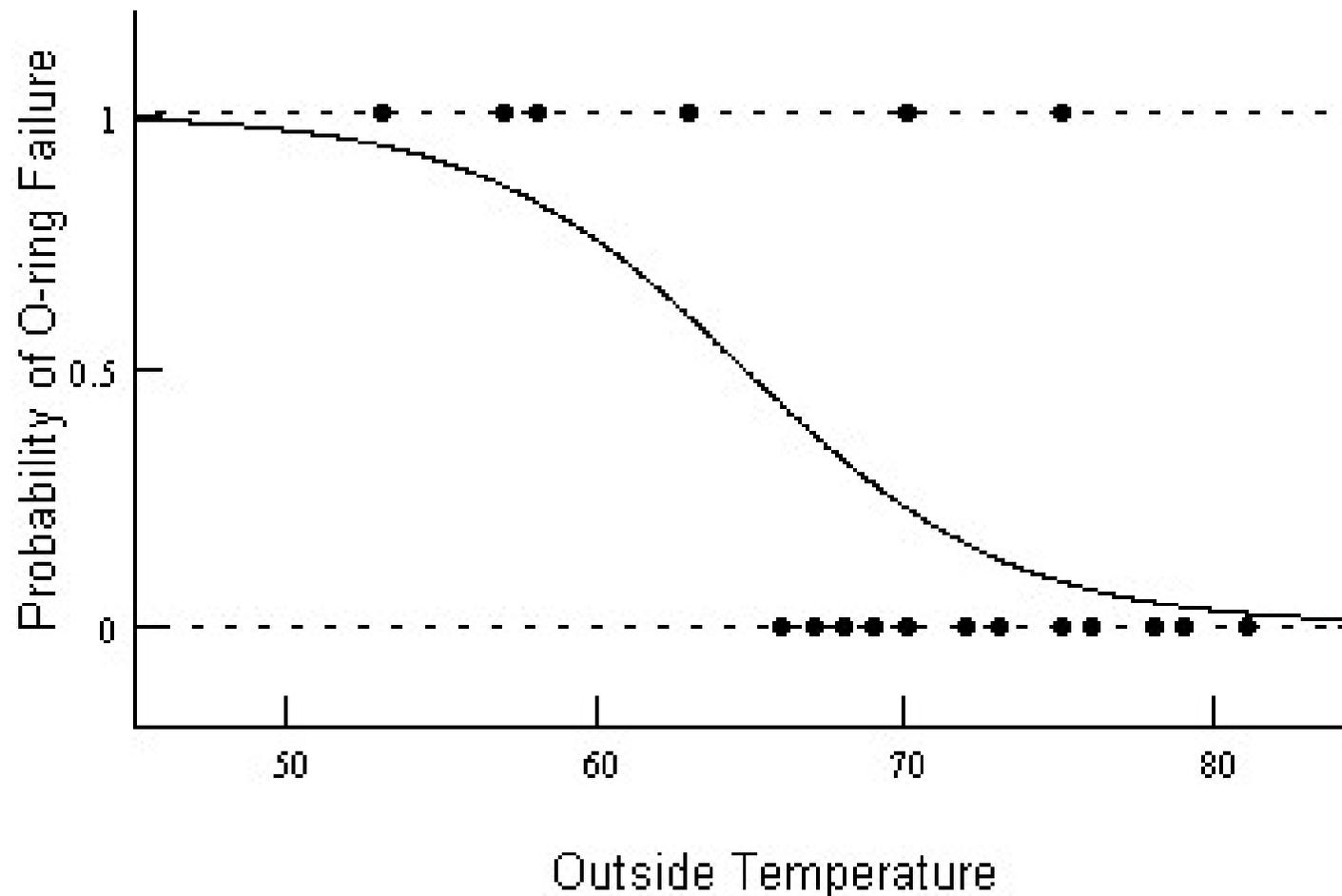
4. W-Theorie

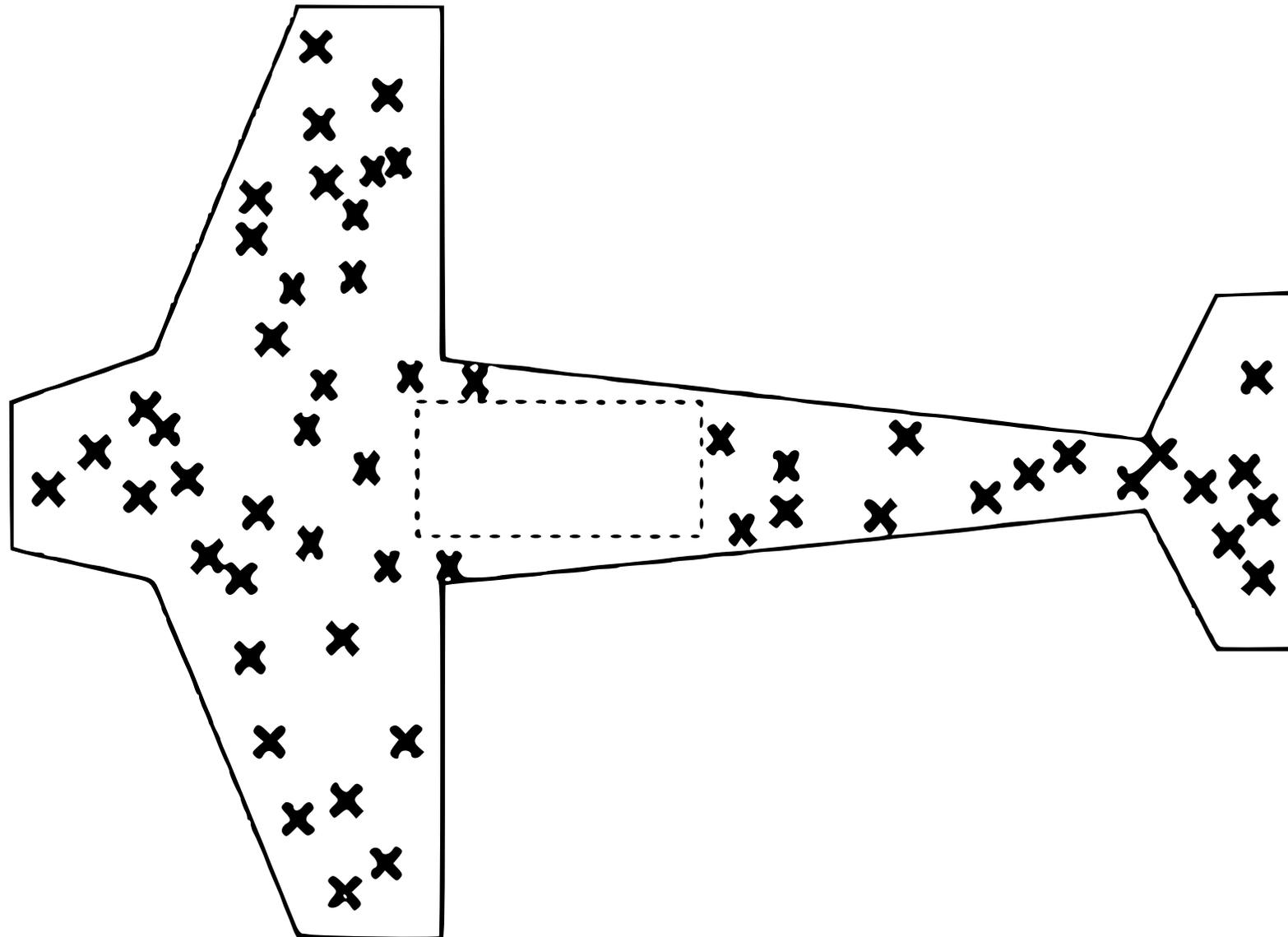
5. Induktive Statistik

Quellen

Tabellen

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





1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

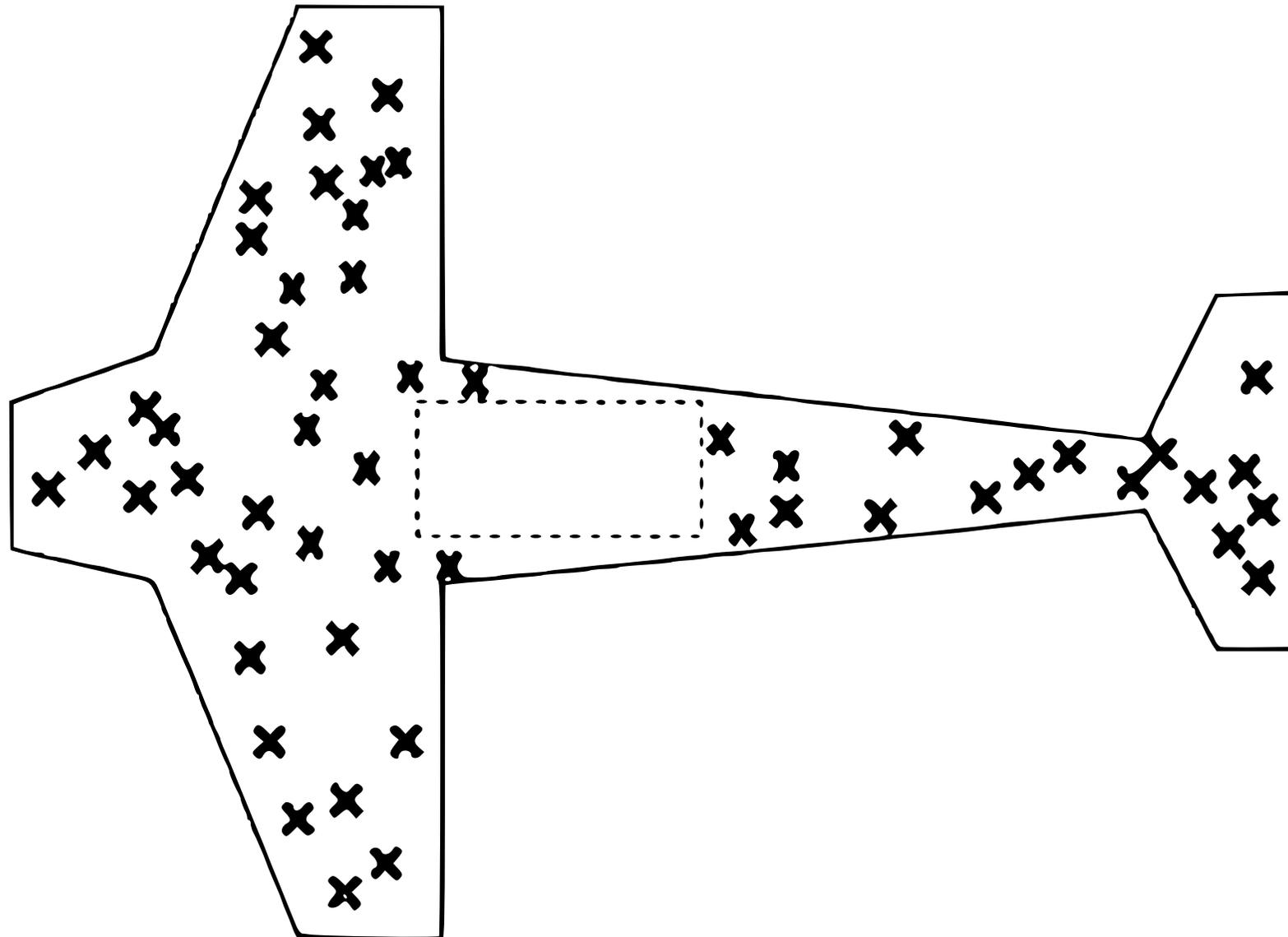
3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

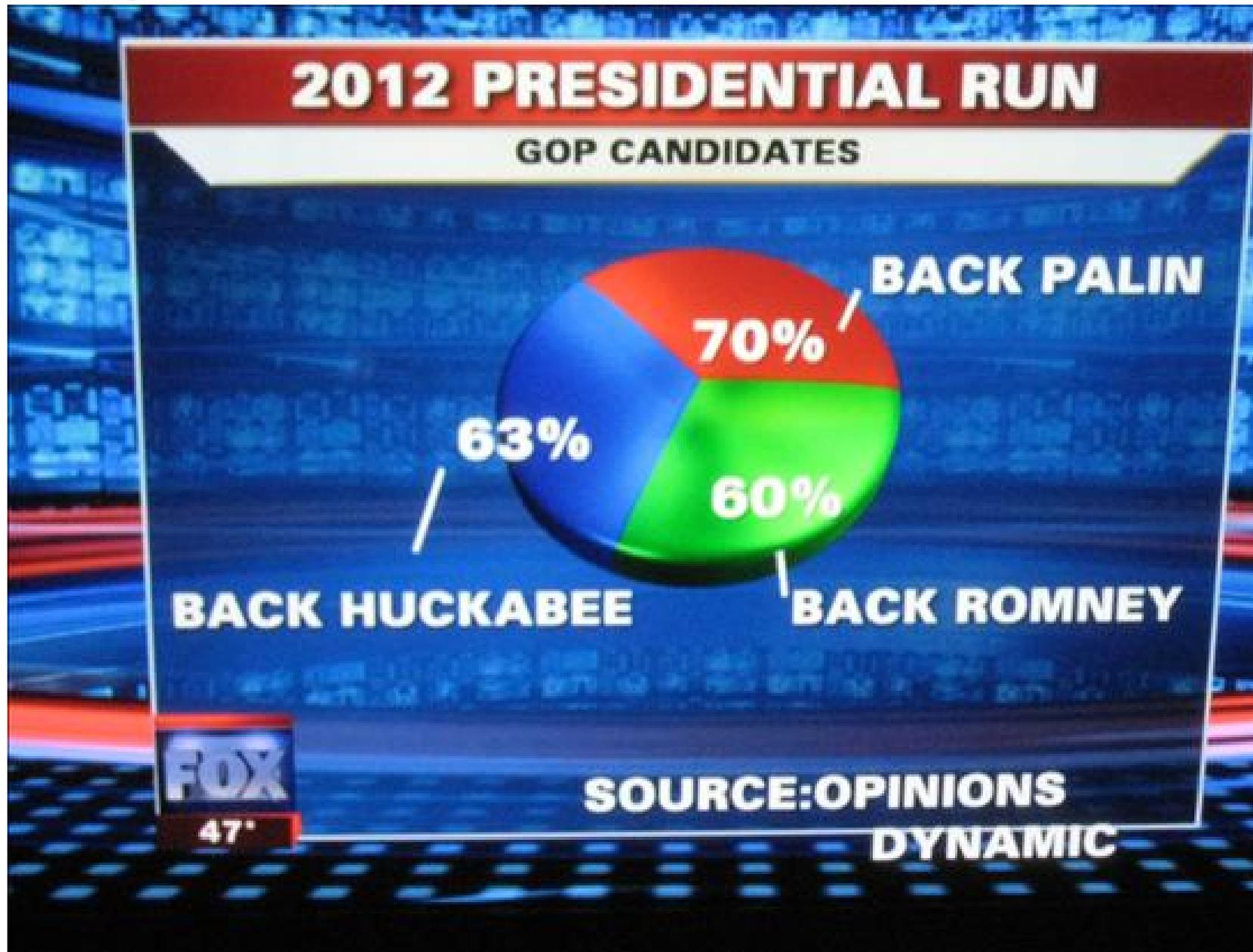
5. Induktive Statistik

Quellen

Tabellen



Aussage?



1. Einführung

- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung
- R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

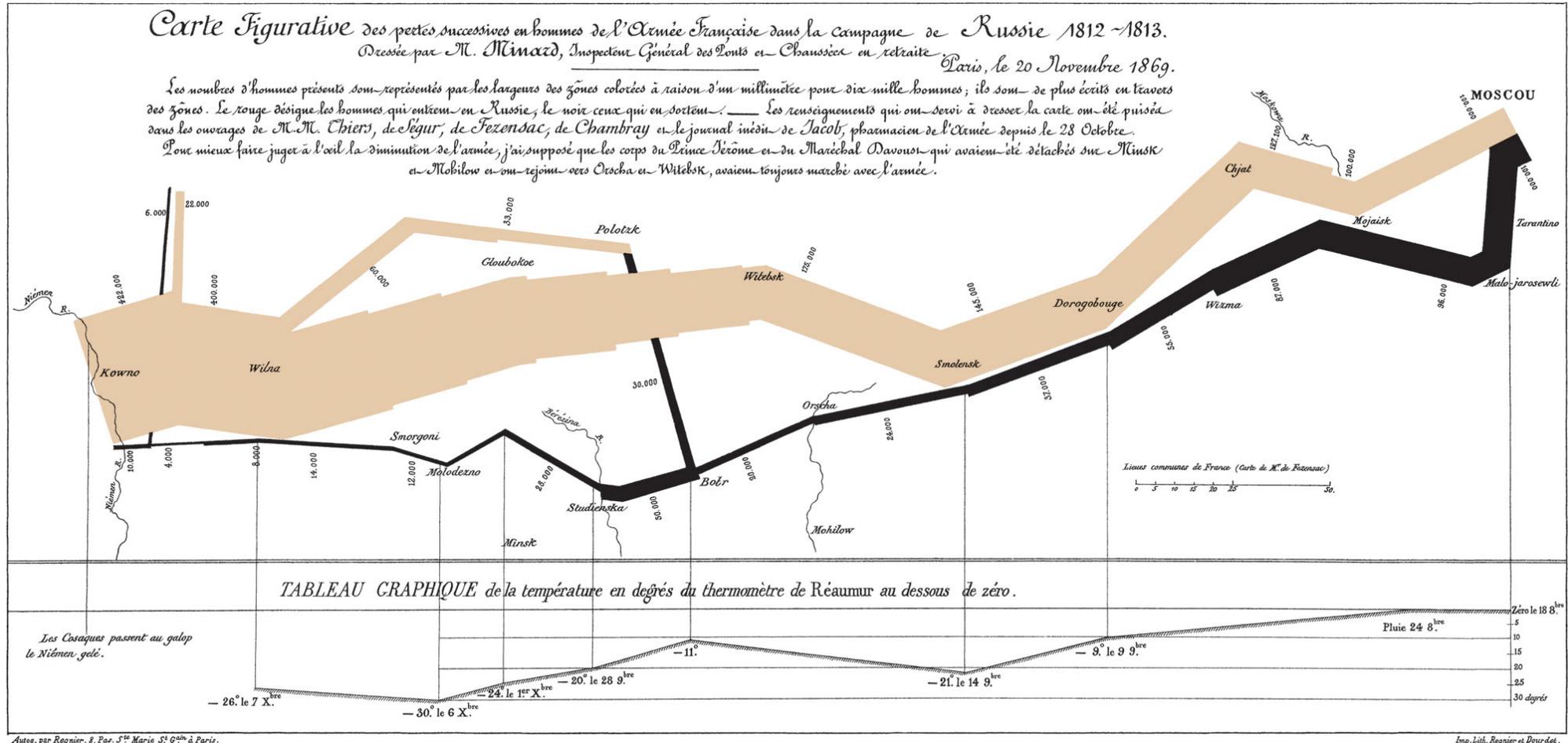
5. Induktive Statistik

Quellen

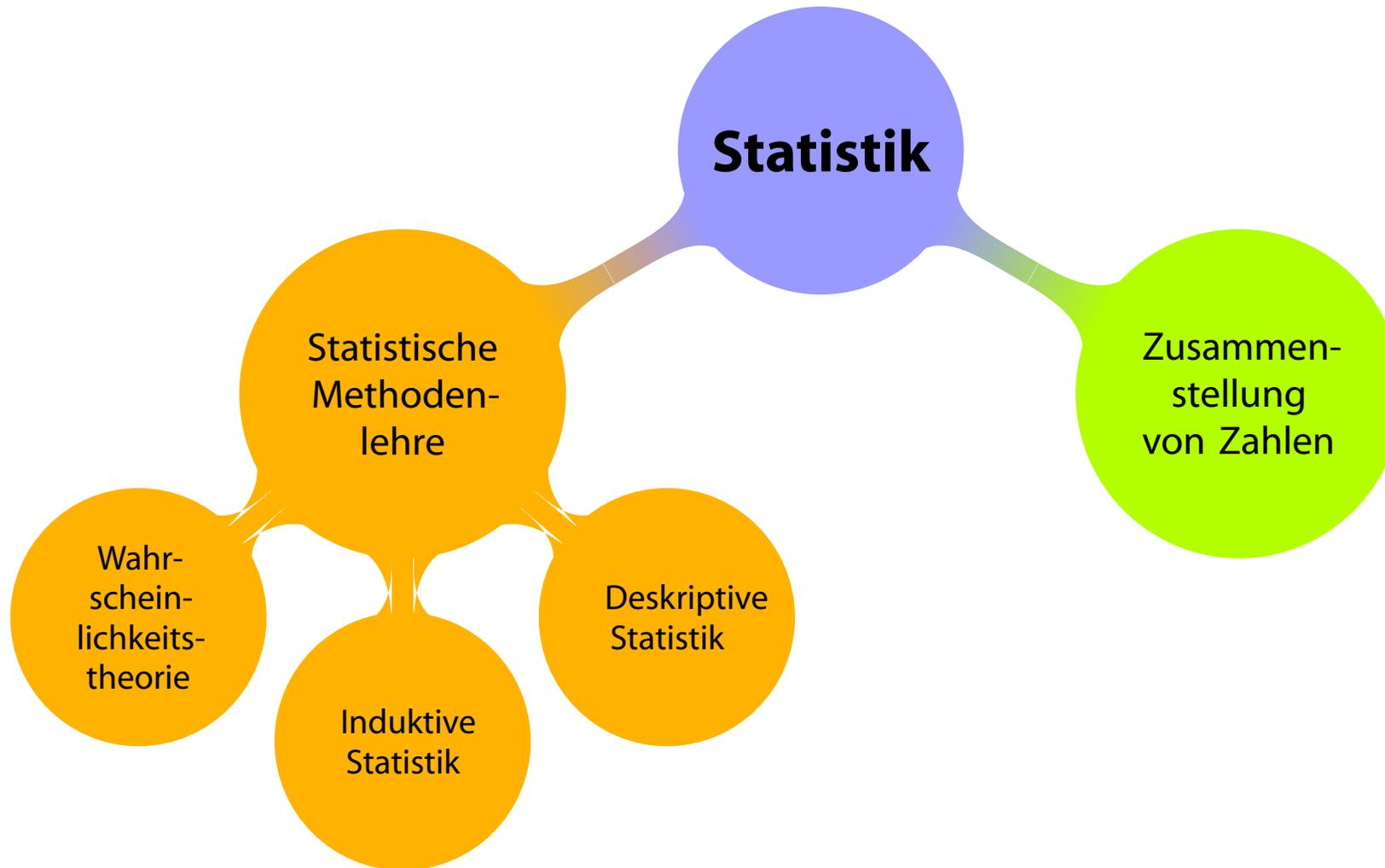
Tabellen



Minards Grafik von 1869 über Napoleons Rußlandfeldzug



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



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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:

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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:

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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.

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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.

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Objekte, cases

- ▶ **Merkmalsträger**: Untersuchte statistische Einheit
- ▶ **Merkmal**: Interessierende Eigenschaft des Merkmalsträgers
- ▶ (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



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Nominalskala:

- ▶ Zahlen haben nur Bezeichnungsfunktion
- ▶ z.B. Artikelnummern

Ordinalskala:

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

Kardinalskala: *metrisch*

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



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



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

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

Was ist R und warum soll man es benutzen?

- ▶ R ist ein **freies** Softwarepaket zu Statistik und Datenanalyse



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

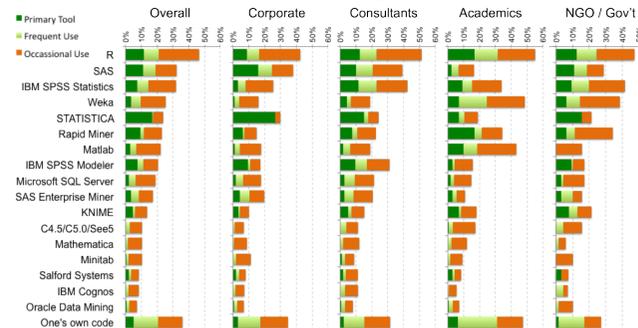
Tabellen

Was ist R und warum soll man es benutzen?

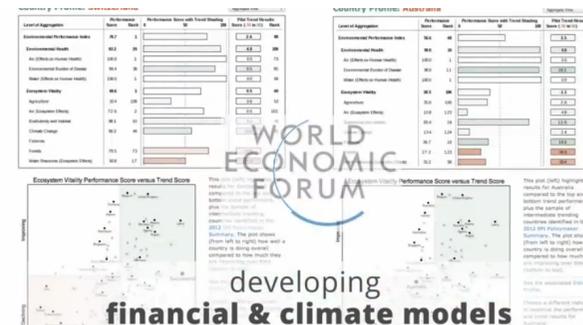
- ▶ R ist ein **freies** Softwarepaket zu Statistik und Datenanalyse
- ▶ R ist sehr mächtig und **weit verbreitet** in Wissenschaft und Industrie (sogar von mehr Leuten benutzt als z.B. SPSS)



- The average data miner reports using 4 software tools.
- R is used by the most data miners (47%).
- STATISTICA is the primary data mining tool chosen most often (17%).



source: <http://goo.gl/axhGhh>



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



1. Einführung

- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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)
> |
```



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

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



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



- ▶ RStudio ist ein **Integrated Development Environment (IDE)** um R leichter benutzen zu können.



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



- ▶ RStudio ist ein **Integrated Development Environment (IDE)** um R leichter benutzen zu können.
- ▶ Gibt's für OSX, Linux und Windows



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



- ▶ RStudio ist ein **Integrated Development Environment (IDE)** um R leichter benutzen zu können.
- ▶ Gibt's für OSX, Linux und Windows
- ▶ Ist auch frei



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

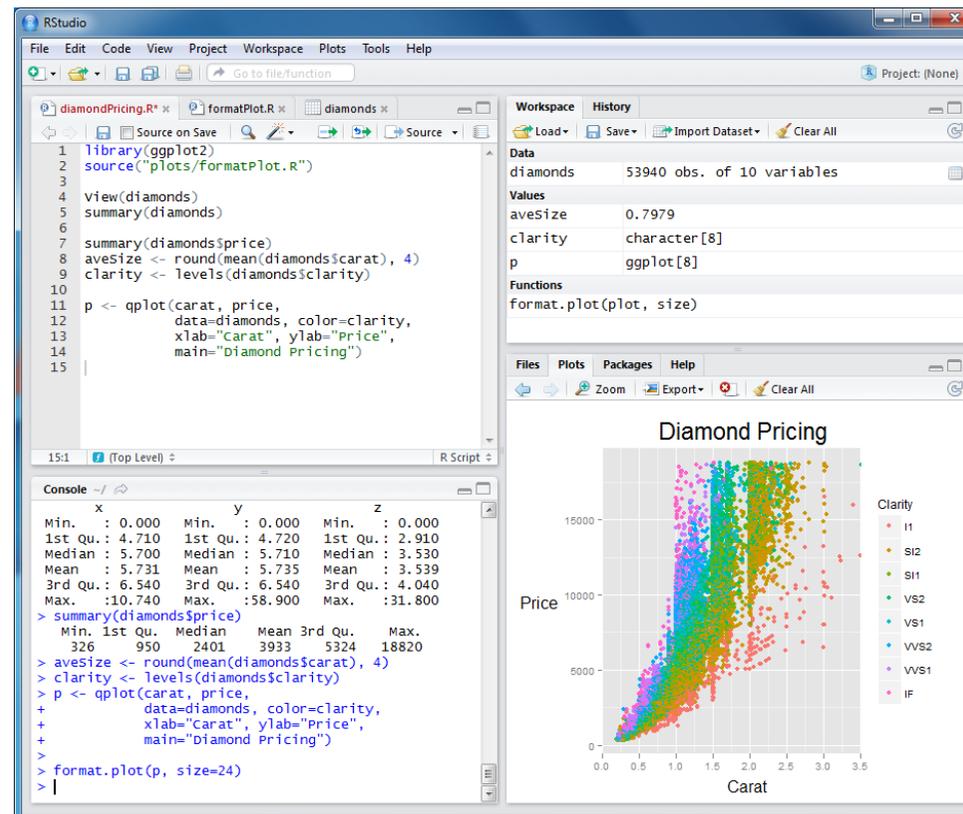
Tabellen

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



Free & Open-Source IDE for R



1. Einführung

- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

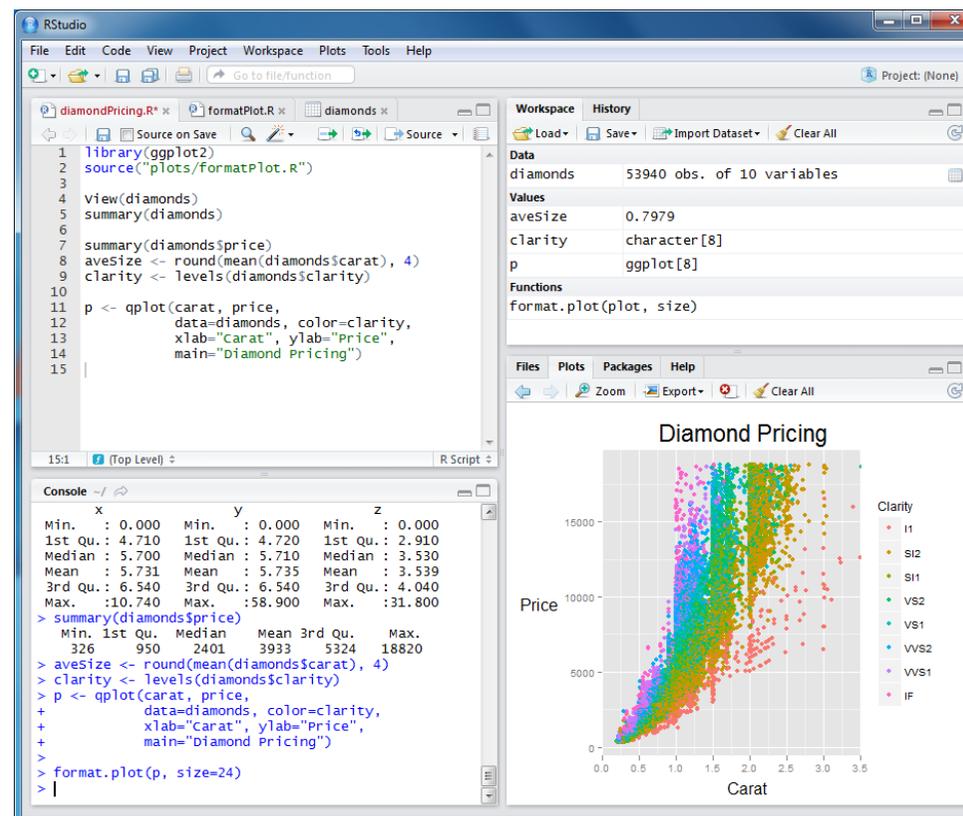
Tabellen



- ▶ 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



Free & Open-Source IDE for R



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

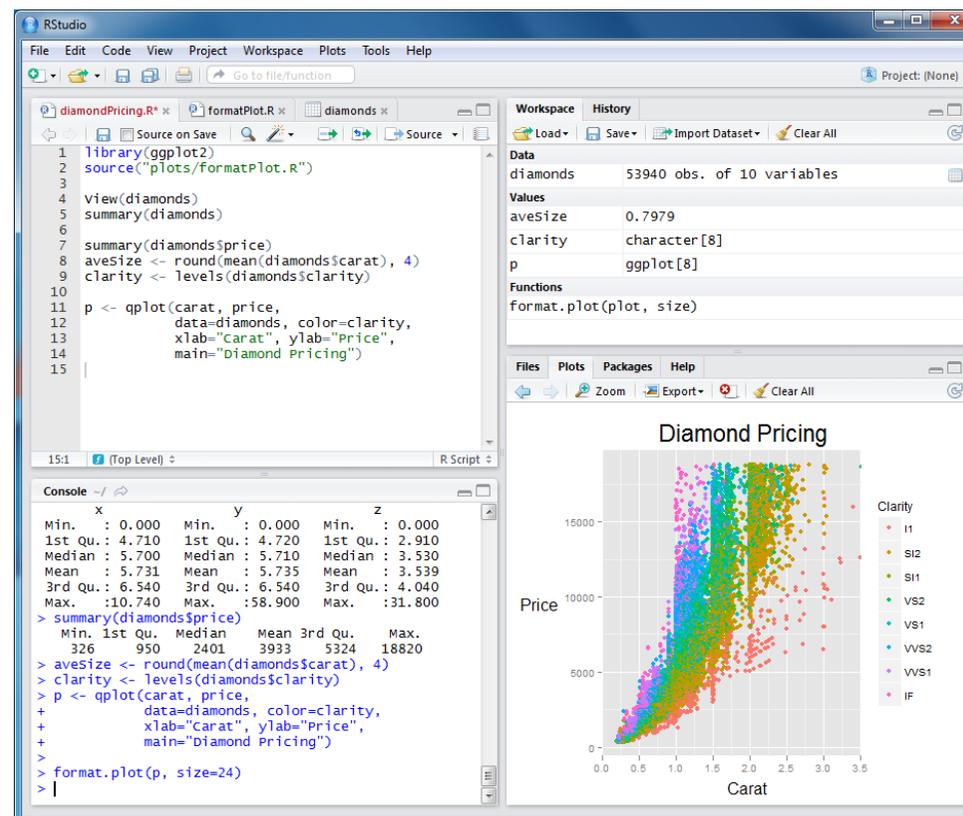
Tabellen

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



Free & Open-Source IDE for R



1. Einführung

- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



RStudio Kennenlernen

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

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for loading the 'diamonds' dataset, summarizing it, and creating a scatter plot of Price vs. Carat, colored by Clarity.
- Console:** Shows the execution of the code, including summary statistics for 'x', 'y', and 'z' (likely carat, price, and clarity), and the execution of the plot command.
- Workspace:** Lists the loaded data (diamonds) and functions (format.plot).
- Plots:** Displays a scatter plot titled 'Diamond Pricing' with 'Price' on the y-axis (0 to 15000) and 'Carat' on the x-axis (0.0 to 3.5). Points are colored by clarity levels: I1, SI2, SI1, VS2, VS1, VVS2, VVS1, and IF.

1. Einführung

- Berühmte Leute zur Statistik
- Wie lügt man mit Statistik?
- Gute und schlechte Grafiken
- Begriff Statistik
- Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



```
# Daten einlesen (hier: Netzwerkkonexion nötig)
MyData = read.csv2("https://goo.gl/xnCLRn")
```

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

```
## 'data.frame': 670 obs. of 18 variables:
## $ Jahrgang : int 2015 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 : num 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 ...
## $ Studiengang : Factor w/ 5 levels "BW","ET","IM",...: NA ...
```

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



```
# 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 Studiengang
## 1         25        450      carnivor    <NA>      2.3 geht so    <NA>
## 2         30        300      carnivor    nein      3.3 geht so    <NA>
## 3         25        100      carnivor    nein      1.7 geht so    <NA>
## 4          6        100      carnivor    nein      2.0 geht so    <NA>
## 5          5         80      carnivor     ja      4.0 geht so    <NA>
## 6         65        250      carnivor    nein      4.0 geht so    <NA>
```

```
# lege MyData als den "Standard"-Datensatz fest
```

```
attach(MyData)
```

```
# Wie Viele Objekte gibt's im Datensatz?
```

```
nrow(MyData)
```

```
## [1] 670
```

```
# Wie Viele Merkmale?
```

```
ncol(MyData)
```

```
## [1] 18
```

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



```
# 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
## [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
## 424   17     46     50
## 456   18     52     55
## 460   18     50     57
## 464   18     40     44
## 479   18     52     44
## 501   18     51     55
## 566   18     52     57
## 620   18     49     58
```

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen

```
# Zeige die Männer, die mehr als 1300 Euro für Schuhe  
# und Mobilfunk zusammen ausgegeben haben  
MyData.Auswahl = MyData[MyData$Geschlecht=="Mann" &  
  MyData$AusgSchuhe + MyData$AusgKomm > 1300,  
  c("Alter", "Geschwister", "Farbe",  
    "AusgSchuhe", "AusgKomm")]
```



```
# ohne NAs
```

```
MyData.Auswahl = na.exclude(MyData.Auswahl)
```

```
MyData.Auswahl
```

##	Alter	Geschwister	Farbe	AusgSchuhe	AusgKomm
## 42	24	1.0	schwarz	1000	600
## 81	25	2.0	silber	200	1900
## 121	22	0.0	silber	300	1100
## 142	20	2.0	schwarz	290	1570
## 161	19	1.0	schwarz	600	800
## 227	20	1.0	schwarz	200	1250
## 249	20	1.0	blau	1000	350
## 256	25	0.0	schwarz	280	1200
## 315	21	1.0	weiss	200	1300
## 353	20	0.0	schwarz	400	950
## 415	26	1.0	blau	600	1850
## 419	21	0.0	schwarz	200	1500
## 492	23	2.0	weiss	160	1800
## 493	26	2.0	schwarz	300	2000
## 494	20	2.0	schwarz	250	1500
## 535	20	2.0	weiss	2500	1500
## 548	26	2.0	schwarz	240	1200
## 562	24	1.0	schwarz	70	4668
## 573	21	1.0	schwarz	300	1200
## 581	19	2.0	silber	500	950
## 582	20	1.0	schwarz	500	1000
## 604	24	1.0	schwarz	150	1340
## 605	21	1.0	silber	600	800
## 615	25	4.5	schwarz	1200	600
## 646	22	1.0	rot	200	2500
## 647	23	1.0	schwarz	200	2000

1. Einführung

Berühmte Leute zur Statistik

Wie lügt man mit Statistik?

Gute und schlechte Grafiken

Begriff Statistik

Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen



```
# Neue Spalte Gesamtausgaben:  
MyData.Auswahl$AusgGesamt = MyData.Auswahl$AusgKomm + MyData.Auswahl$AusgSchuhe  
# sortiert nach Gesamtausgaben  
MyData.Auswahl[order(MyData.Auswahl$AusgGesamt), ]
```

##	Alter	Geschwister	Farbe	AusgSchuhe	AusgKomm	AusgGesamt
## 249	20	1.0	blau	1000	350	1350
## 353	20	0.0	schwarz	400	950	1350
## 121	22	0.0	silber	300	1100	1400
## 161	19	1.0	schwarz	600	800	1400
## 605	21	1.0	silber	600	800	1400
## 548	26	2.0	schwarz	240	1200	1440
## 227	20	1.0	schwarz	200	1250	1450
## 581	19	2.0	silber	500	950	1450
## 256	25	0.0	schwarz	280	1200	1480
## 604	24	1.0	schwarz	150	1340	1490
## 315	21	1.0	weiss	200	1300	1500
## 573	21	1.0	schwarz	300	1200	1500
## 582	20	1.0	schwarz	500	1000	1500
## 42	24	1.0	schwarz	1000	600	1600
## 653	27	2.0	schwarz	700	950	1650
## 419	21	0.0	schwarz	200	1500	1700
## 494	20	2.0	schwarz	250	1500	1750
## 615	25	4.5	schwarz	1200	600	1800
## 142	20	2.0	schwarz	290	1570	1860
## 492	23	2.0	weiss	160	1800	1960
## 663	27	2.0	schwarz	200	1800	2000
## 81	25	2.0	silber	200	1900	2100
## 647	23	1.0	schwarz	200	2000	2200
## 493	26	2.0	schwarz	300	2000	2300
## 415	26	1.0	blau	600	1850	2450

1. Einführung

Berühmte Leute zur Statistik
Wie lügt man mit Statistik?
Gute und schlechte Grafiken
Begriff Statistik
Grundbegriffe der
Datenerhebung

R und RStudio

2. Differenzieren 2

3. Deskriptive Statistik

4. W-Theorie

5. Induktive Statistik

Quellen

Tabellen