

## **Application Performance Management**

# **Datenauswertung & Präsentation**

Michael Faes

# Auswertung & Präsentation

Ziel jeder Performance-Analyse: Hilfestellung für Entscheidungen

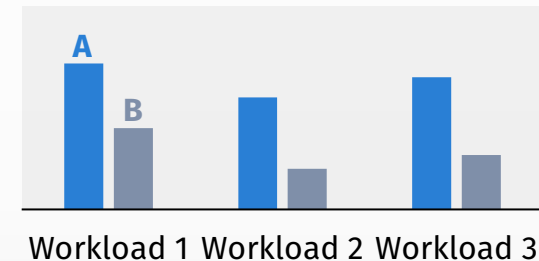
→ Wichtiger Schritt: **Präsentation/Kommunikation der Ergebnisse**

*Mit System A erreicht die App einen Durchsatz von durchschnittlich 120% mehr als mit System B, bei gleichbleibender Auslastung der Ressourcen. Die durchschnittliche Antwortzeit wird um etwa 20% verbessert.*

**Worte**



**Bilder**

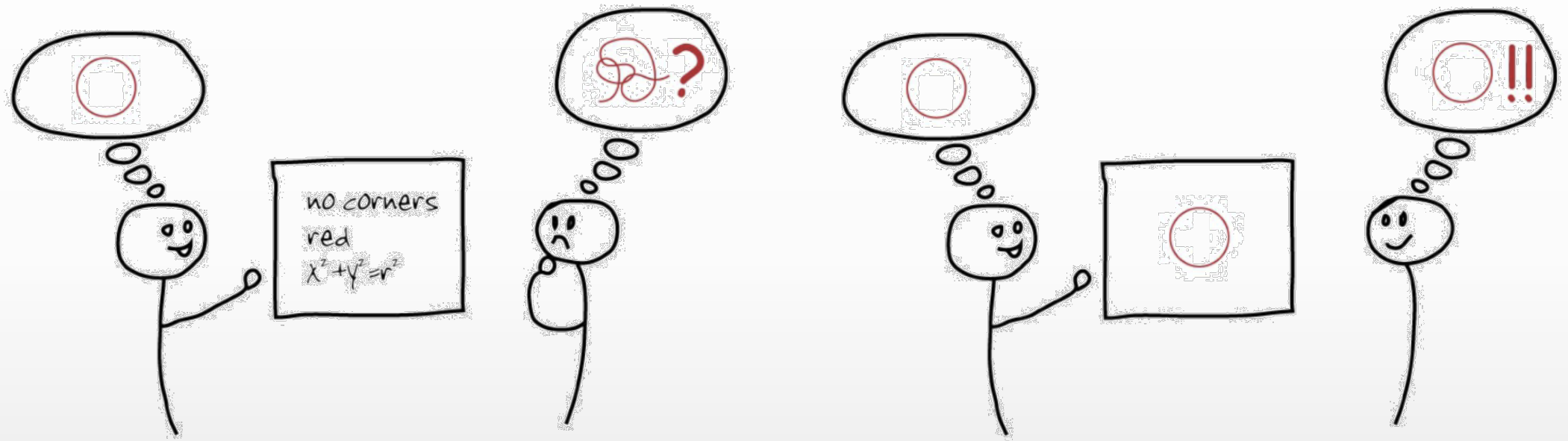


**Diagramme**

Herausforderung: Resultate müssen evtl. jemanden **überzeugen**, der keine Ahnung von Performance hat (oder schon nur Informatik...).

*Egal, wie korrekt/interessant/signifikant die Resultate sind, wenn Entscheidungsträger sie nicht verstehen, sind sie **nichts wert!***

# Ein Bild sagt mehr als 1000 Worte



**Text**

**Bilder, Diagramme, ...**

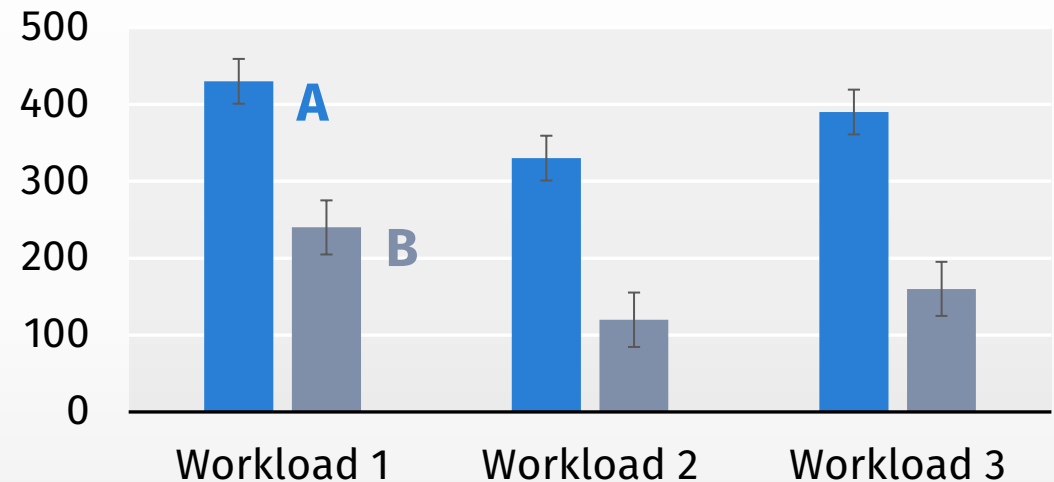
ist offensichtlicher und verständlicher für andere

# Ein Diagramm sagt mehr als 1000 Worte

*Mit System A erreicht die Applikation für den Workload 1 einen Durchsatz von 430 Anfragen/s (A/s), während System B nur 240 A/s erreicht. Für die beiden weiteren Workloads erreicht System A 330 bzw. 390 A/s und System B 120 bzw. 160 A/s. Im Durchschnitt entspricht dies einer Verbesserung von 120%. Die Standardabweichungen sind klein.*

**Text**

Durchsatz  
(Anfragen/s)



**Diagramm**

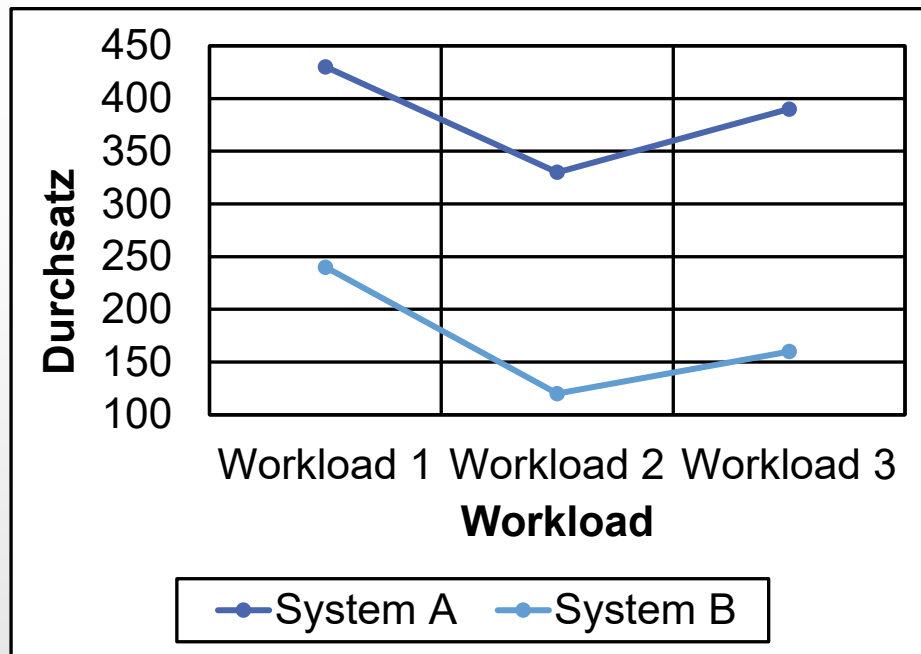
Diagramme sind nicht *immer* besser. Für komplexe Sachverhalte braucht es (zusätzlich) Text. Aber wenn möglich, dann Diagramme.

# Diagramme trivial?

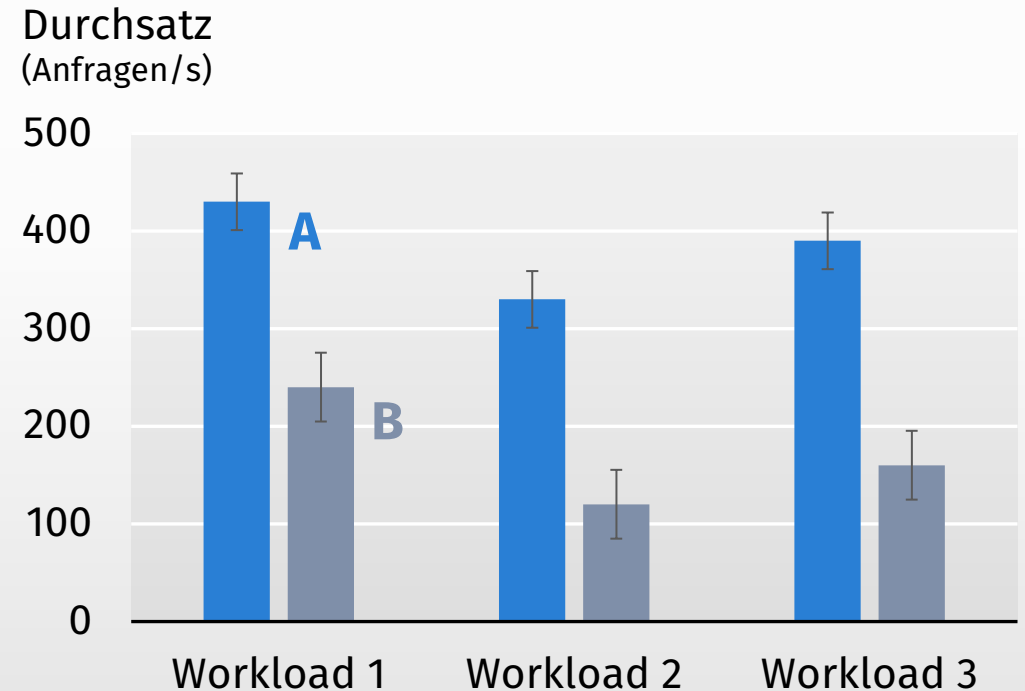
«Excel & Co. erzeugen gute Diagramme doch automatisch.»

Wirklich?

Software nicht einfach machen lassen! Man muss wissen, wie man mit der SW umgeht und die Daten optimal anzeigt!!



schlechtes Diagramm



# Übersicht Woche 4

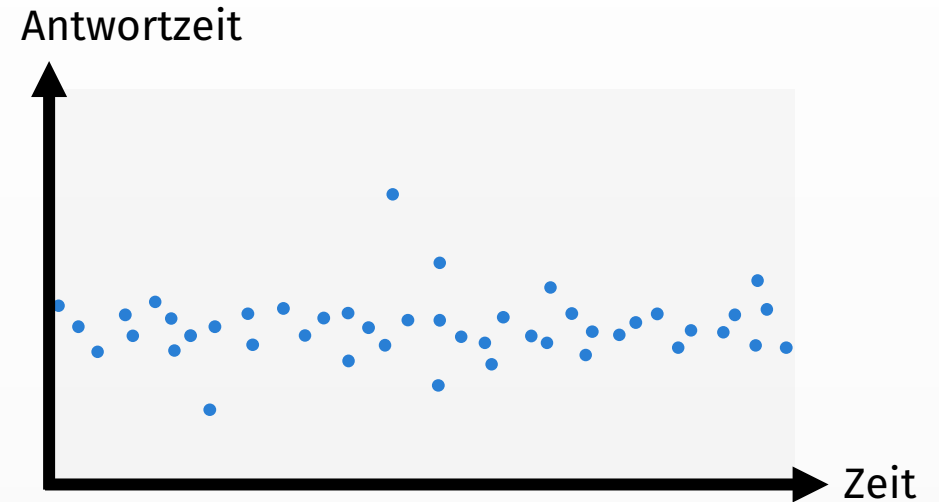
1. Übungsbesprechung
2. Messen & Auswerten: Grundlagen
3. Statistische Auswertung von Performance-Messungen
4. Präsentation von Resultaten (Visuell)
5. Übung: ...

# **Messen & Auswerten: Grundlagen**

# Zufallsvariablen

Rückblick: Jede **Messung produziert leicht unterschiedlichen Wert**, obwohl wir immer «das gleiche» messen...

**Grund:** Unkontrollierbare, *zufällige* Unterschiede in Umgebung.



**Modellierung als *Zufallsvariable*:**

=> Wert ist nicht immer gleich, sondern abhängig vom Zufall!

**Gemessener Wert ist keine fixe Zahl und auch nicht beschreibbar durch Funktion von anderen Variablen. Sondern: abhängig vom Zufall.**

Vereinfachung! Theoretisch (vielleicht) schon eindeutig bestimmbar, aber nicht praktikabel. Deshalb Modellierung als Zufallsvariable.



# Verteilung einer Zufallsvariable

Wahrscheinlichkeit immer zwischen 0 und 1.

Jede Zufallsvariable hat eine (bekannte oder unbekannte) *Verteilung*:



Verteilung ordnet jedem möglichen Wert der Variable eine Mess-Wahrscheinlichkeit  $p$  zu, wobei  $0 \leq p \leq 1$ .

# Normalverteilung

Wenn wir Zufallsvariable (z. B. Antwortzeit) messen, müssen wir im Prinzip nicht 1 Wert, sondern komplette Verteilung bestimmen!

Typische Annahme:

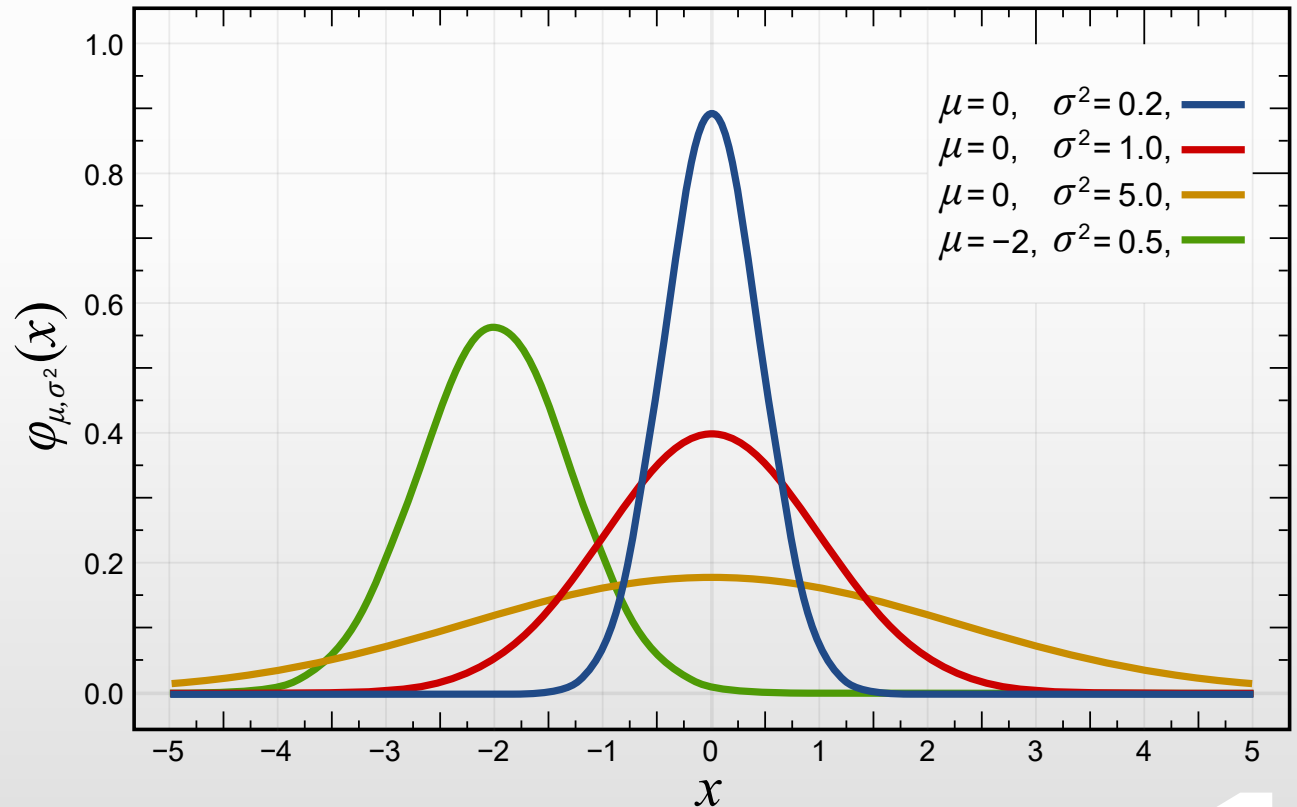
*Normalverteilung*

(Glocken-, Gauskurve)

Parameter: Abweichung, Varianz

Warum eigentlich???

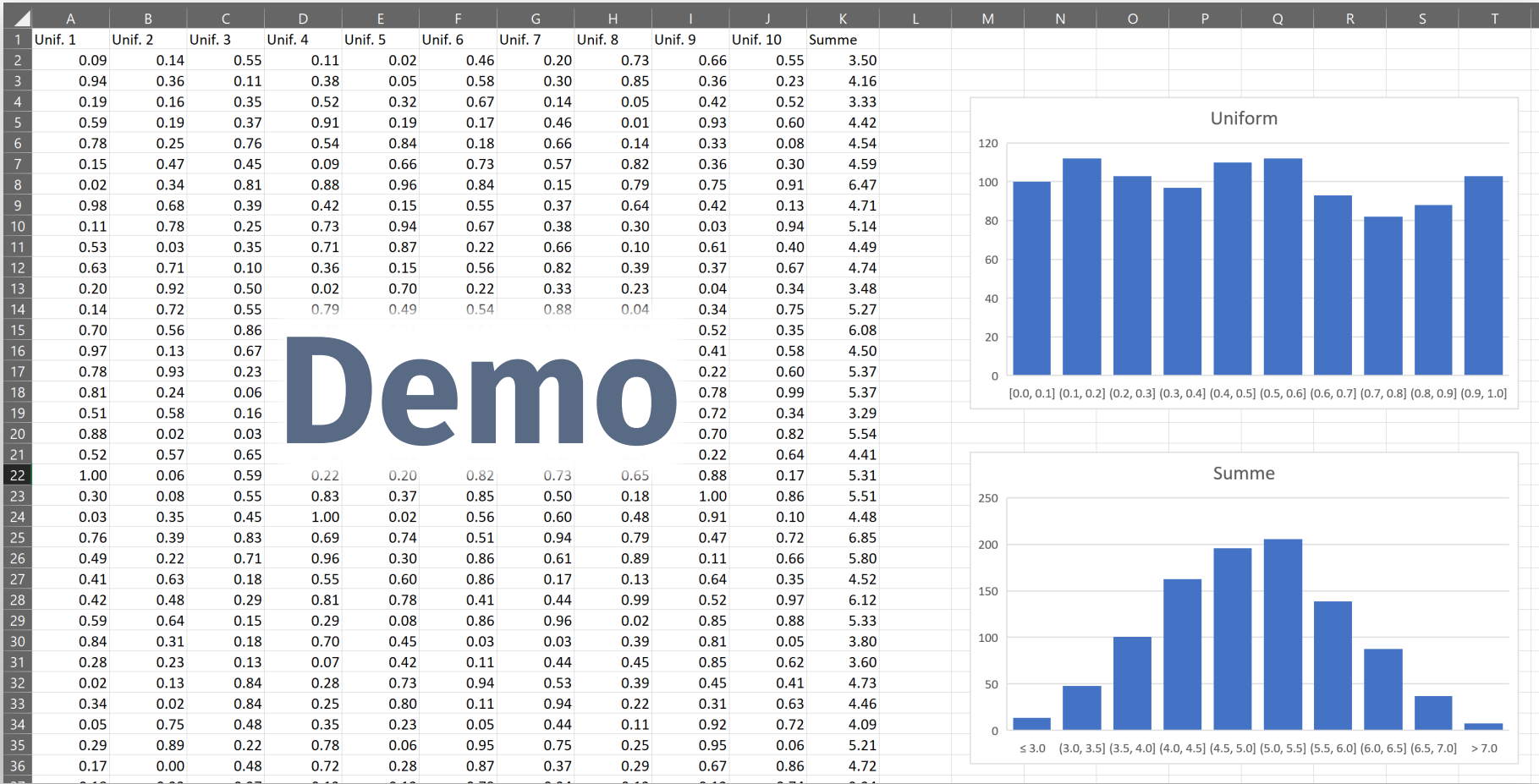
-> symmetrische Abweichungen



# Zentraler Grenzwertsatz

Bernoulli-Verteilung: zwei Werte mit jeweils 0 und 1.

Zentraler Grenzwertsatz: Summe von vielen (unabhängigen) Zufällen, egal mit welcher Verteilung, ergibt Normalverteilung!



Summe von zwei Glockenkurven, ergibt wieder eine Glockenkurve

# Normalverteilung und Performance

Normalverteilung ist oft sinnvolle Annahme für Performance-Messung, da viele unabhängigen Effekte zu Messfehler beitragen.

Einige statistische Verfahren *gehen von Normalverteilung aus*:

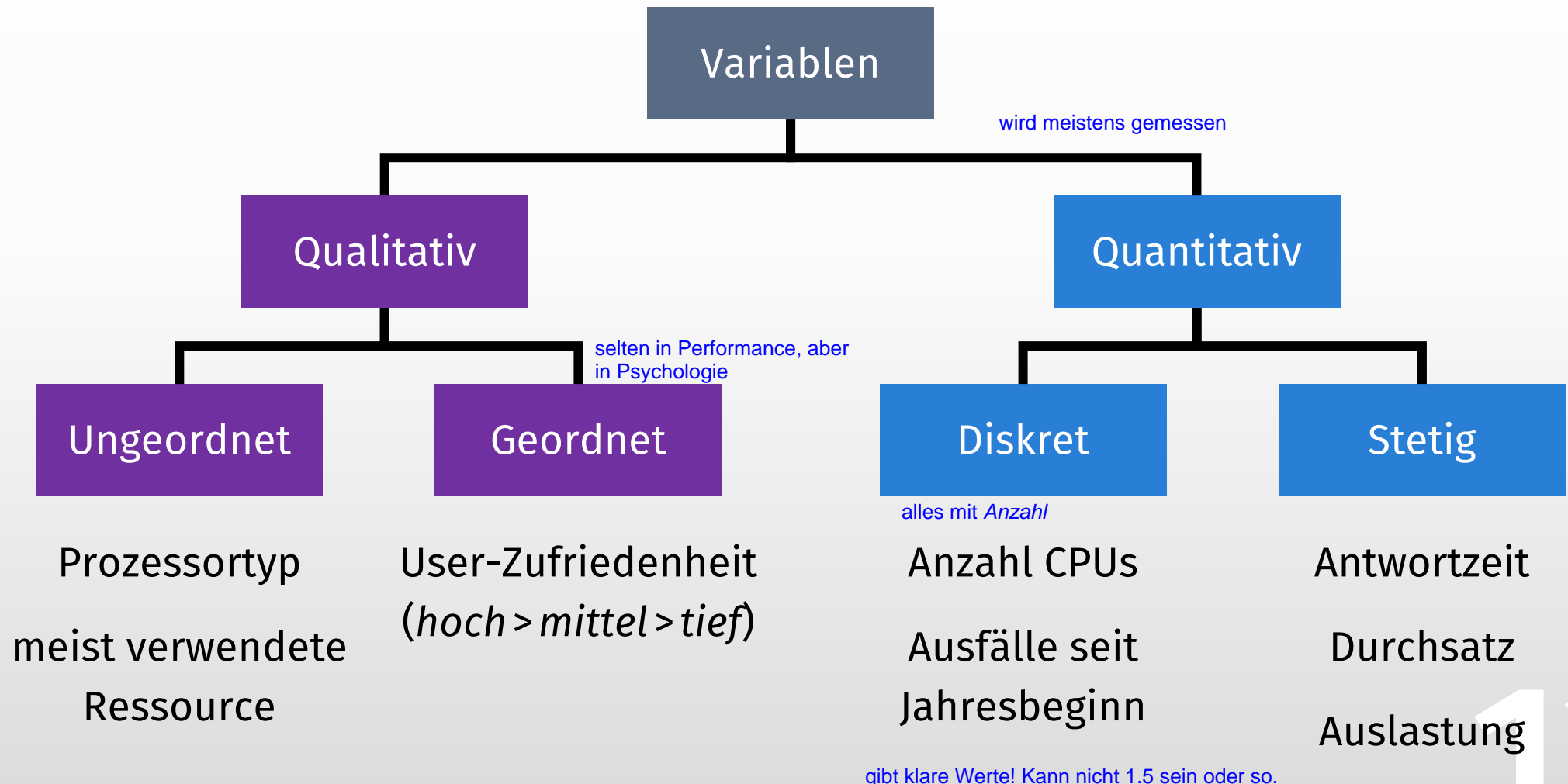
```
Result "ch.fhnw.apm.docfinder.DocFinderBenchmarks.findDocs":  
  3.780 ±(99.9%) 0.494 ops/s [Average]  
  (min, avg, max) = (3.626, 3.780, 3.911), stdev = 0.128  
  CI (99.9%): [3.286, 4.274] (assumes normal distribution)
```

JMH -Tool nimmt an, dass es eine Normalverteilung ist

**Achtung:** Annahme gilt nicht immer! Bevor man statistische Auswertung macht, sollte man **Annahme überprüfen, z. B. visuell.**

# Typen von Variablen

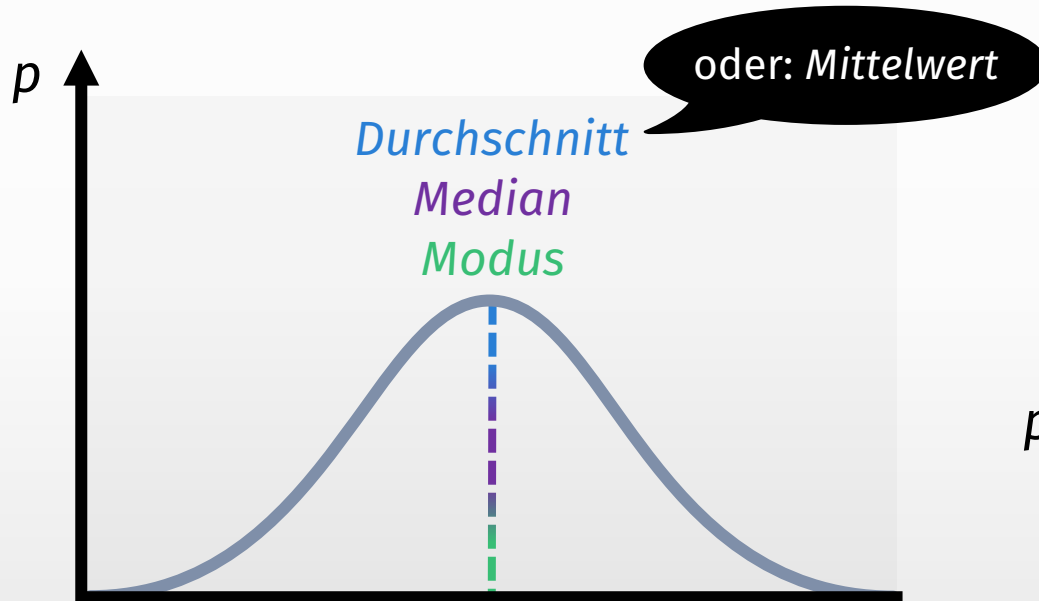
Nicht alle «Messungen» liefern gleiche Art von Variable:



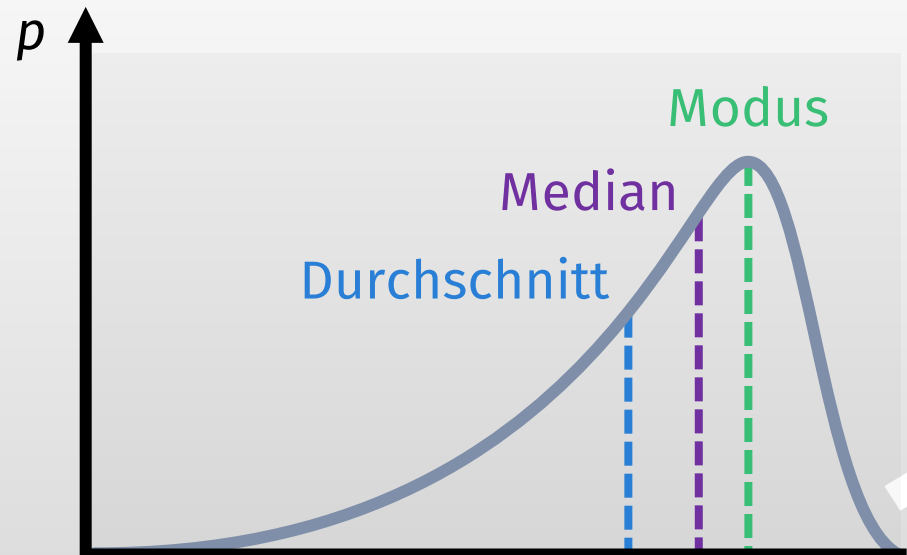
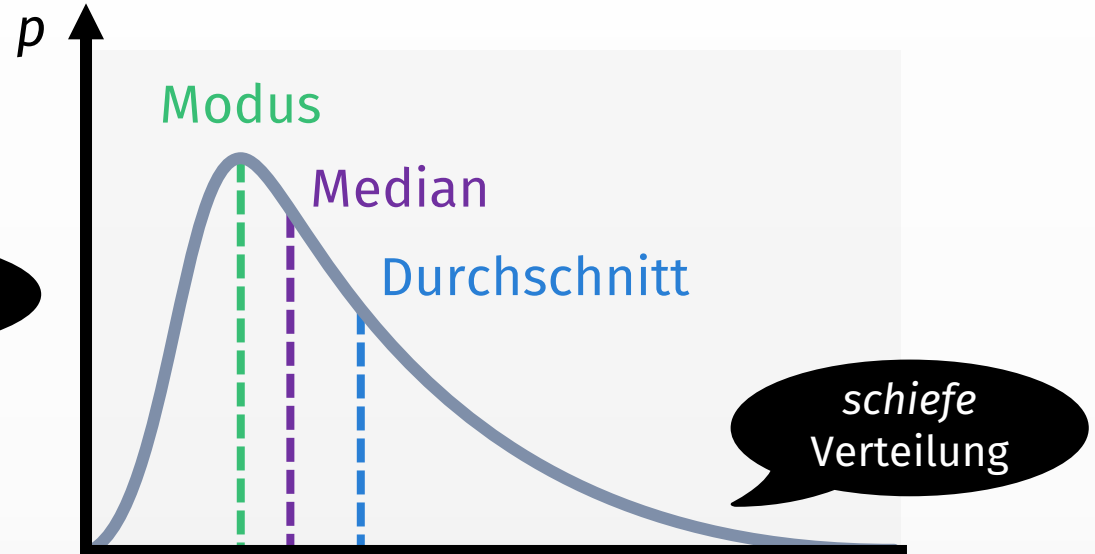
# Durchschnitt, Median & Modus

Modus: wo man am meisten gemessen hat

Einfachste Art, Zufallsvariable  
zusammenzufassen: 1 Wert



**Median:** Wert, der in der Mitte  
aller (sortierten) Messwerten liegt.



# Variablentypen & Auswertung

Nicht alle Variablentypen erlauben alle Arten von Auswertungen!

Typ	Operationen	Eigenschaften	Zusammenfassung
Qualitativ ungeordnet ( <i>Nominalskala</i> ) <small>Bsp: Prozessortyp, Workloads</small>	$=/\neq$	Häufigkeit	Modus
Qualitativ geordnet ( <i>Ordinalskala</i> )	$=/\neq$ $</>$	Häufigkeit Rangfolge	Modus Median
Quantitativ ( <i>Intervalskala</i> )	$=/\neq$ $</>$ $+/-$	Häufigkeit Rangfolge Abstand	Modus Median Durchschnitt
Quantitativ ( <i>Verhältnisskala</i> ) <small>(wenn natürlicher Nullpunkt)</small>	$=/\neq$ $</>$ $+/-$ $\times/\div$	Häufigkeit Rangfolge Abstand natürl. Nullpunkt	Modus Median Durchschnitt Geom. Mittel

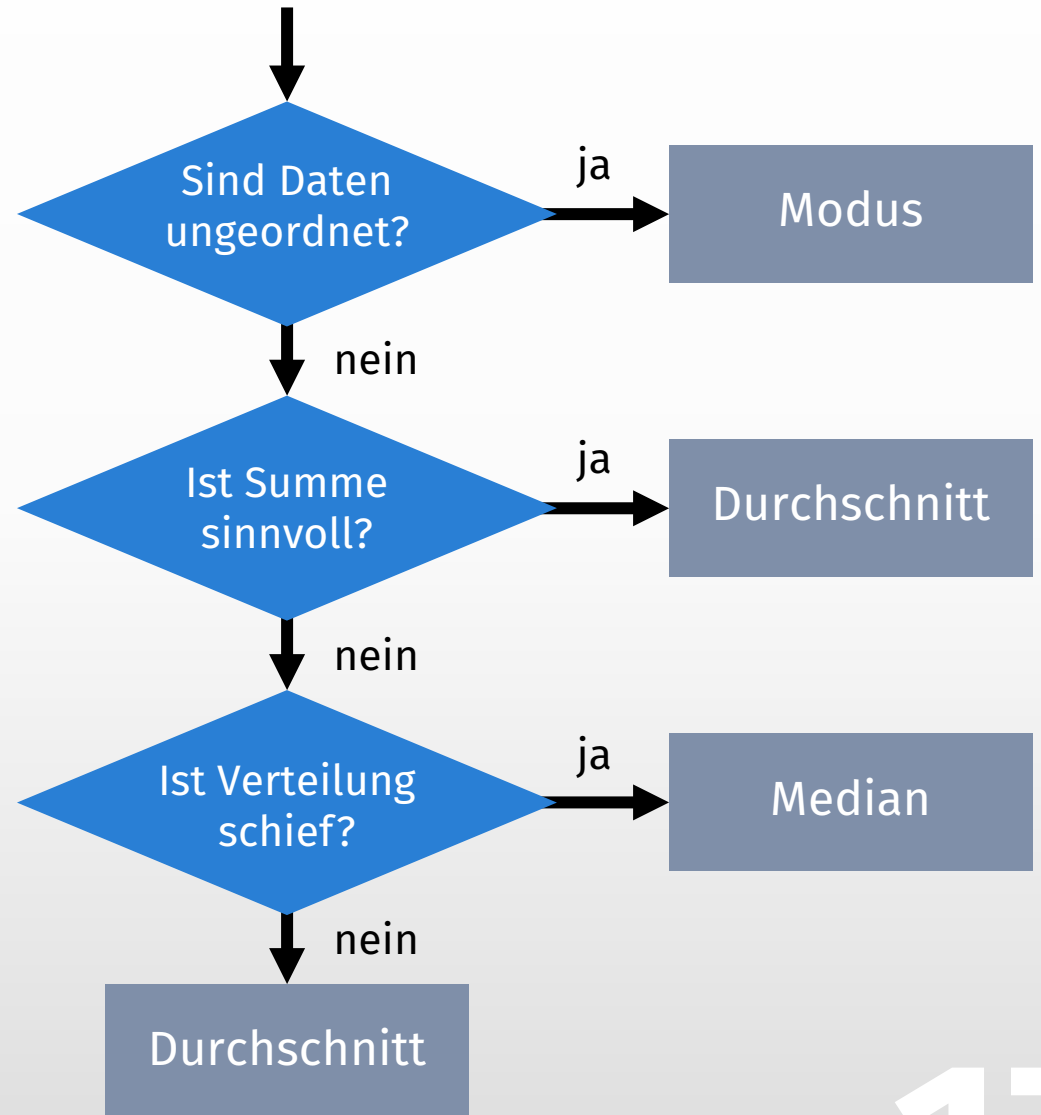
# **Statistische Auswertung von Performance-Messungen**



# Durchschnitt, Median oder Modus?

häufiger Fehler: falsche Zusammenfassungsart

Häufig wird Durchschnitt als Zusammenfassung verwendet, obwohl Median oder Modus sinnvoller wären!



# Durchschnitt & Median: Häufige Fehler

## 1. Durchschnitt/Median von deutlich unterschiedlichen Werten

Auch wenn Verwendung von Durchschnitt/Median für Variable korrekt wäre, macht Verwendung trotzdem nicht immer Sinn.

### Beispiel 1:

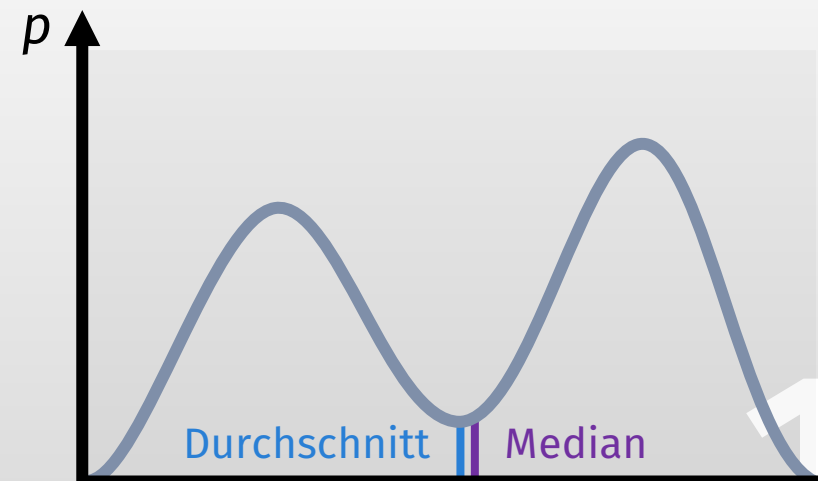
Anfrage	Antwortzeit (ms)
A	520
B	540
C	20

Durchschnitt: 360 ms.

Repräsentiert Daten nicht gut und bringt keinerlei Einsicht.

### Beispiel 2: Durchschnitt/Median für Verteilung mit mehreren Peaks:

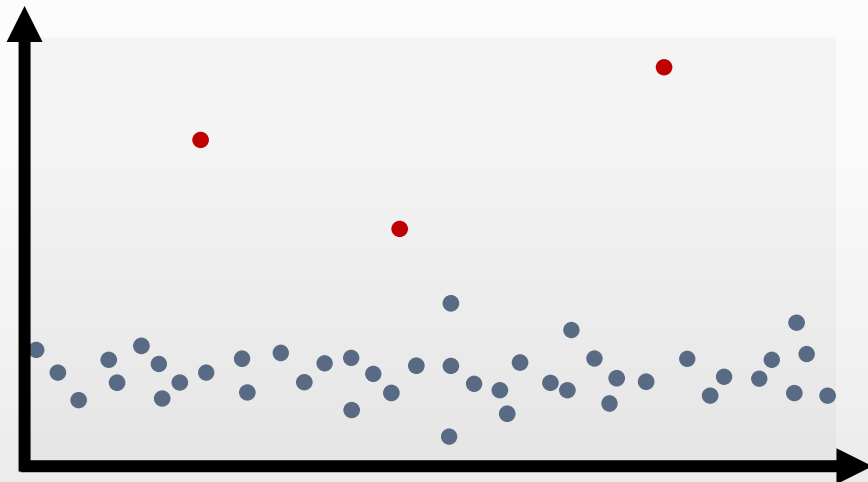
- > Aufteilung in 2 Datensätze (wenn möglich)
- > macht nicht sinn, nur einen Wert zu nehmen / zu reduzieren
- > es ist Bi-Modal



## 2. Durchschnitt ohne Rücksicht auf Ausreisser

Vergleich nur von Durchschnitten kann falsches Bild liefern, vor allem, wenn *Ausreisser* vorhanden sind. (= outliers)

**Beispiel:** Antwortzeit von zwei Systemen:



	System A	System B
	10	5
	9	5
	11	5
	10	4
	10	31
Summe	50	50
Durchschnitt	10	10
Median	10	5

ist nicht abhängig von Ausreisser

3. Durchschnitt von Verhältnissen

Rückblick Woche 1: Ratio Game

**Fehler:** *Nicht die gleiche Basis.*

	Workload 1	Workload 2	Ø
A	200%	50%	125%
B	100%	100%	100%

System	Workload 1	Workload 2
A	20	10
B	10	20

	Workload 1	Workload 2	Ø
A	100%	100%	100%
B	50%	200%	125%

Weiteres Beispiel:

Durchschnittl. CPU-Auslastung

Korrekt:

**Durchschnitt** = 
$$\frac{\text{Summe von Auslastungen}}{\text{Summe von Messdauer}}$$
$$= \frac{45 + 45 + 45 + 45 + 20}{1 + 1 + 1 + 1 + 100} = 21\%$$

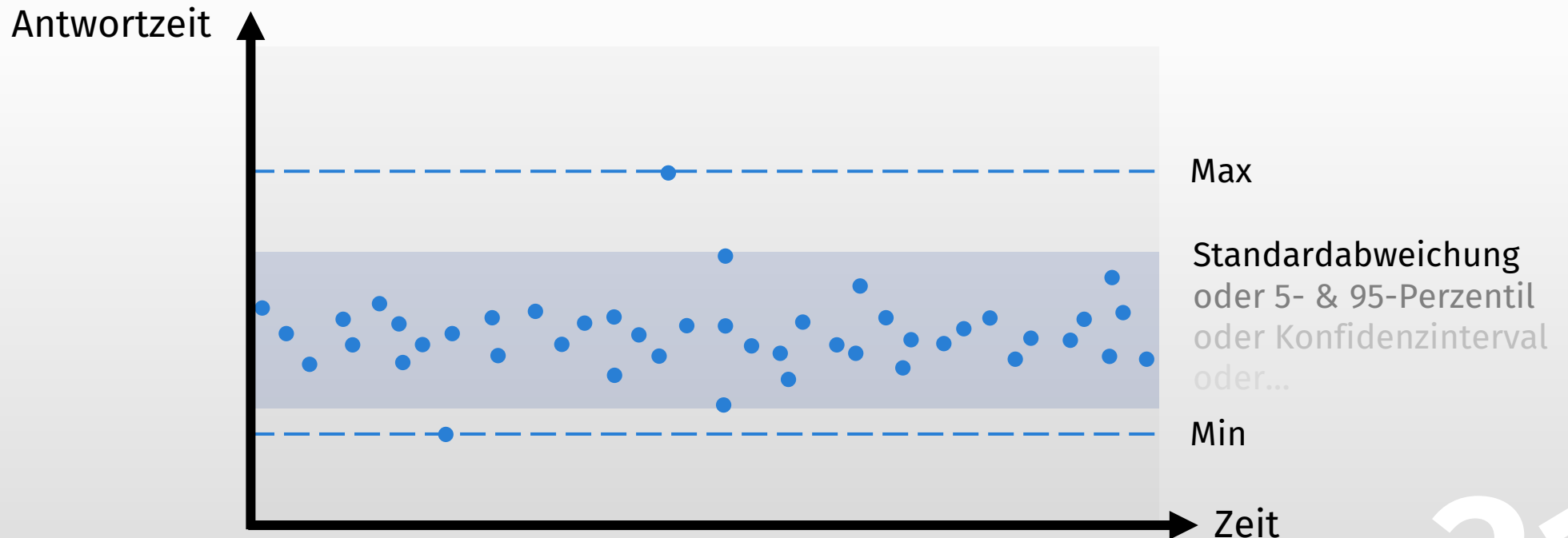
Messdauer (s)	CPU-Ausl.
1	45 %
1	45 %
1	45 %
1	45 %
100	20 %
Summe	200 %
Durchschnitt	<del>40 %</del>

# Streuung

Eine einzige Zahl reicht meist nicht, um Zufallsvariable zu beschreiben.

«Und dann war da noch der Mann, der ertrank, als er einen Fluss durchqueren wollte, der **durchschnittlich** 20 cm tief war.»

-> Durchschnitt ist nicht immer aussagekräftig



# Varianz & Standardabweichung

Einfachste Möglichkeit, Streuung anzugeben: *Minimum* & *Maximum*.

- Oft nicht besonders nützlich, wegen Ausreissern.

**Besser:** *Stichproben-Varianz*  $s^2$ , bzw. *-Standardabweichung*  $s$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{wobei} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

wenn nur ein Wert ist Varianz = 0 Mittelwert wenn nur ein Wert Mittelwert = Messwert

Standardabweichung ist so etwas wie *durchschnittliche Abweichung vom Mittelwert* (aber grössere Abweichungen stärker gewichtet).

**Wichtig:** Mittelwert und Standardabweichung **der Stichprobe!** Sind nur Schätzungen für «wahren» Mittelwert und «wahre» Std.abweichung.

- Qualität der Schätzung hängt von *Grösse der Stichprobe* ab

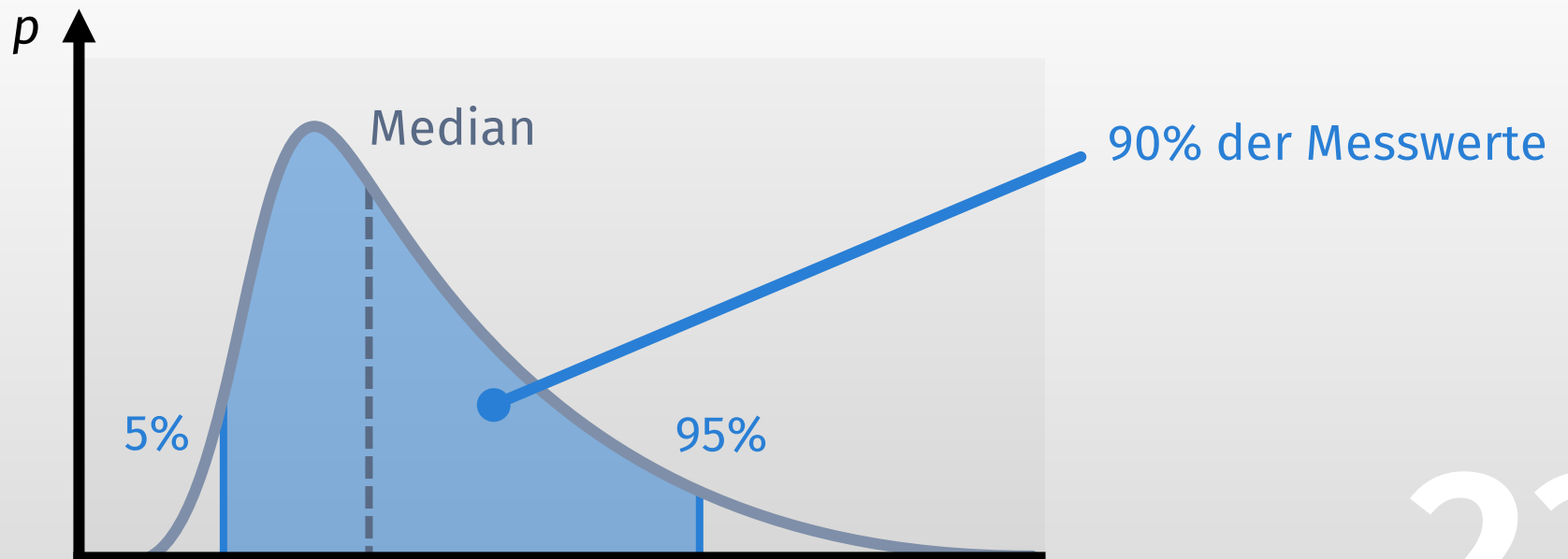
# Perzentile

Varianz und Standardabweichung machen nur Sinn, wenn:

- Variable quantitativ ist (nicht qualitativ)
- Verteilung symmetrisch ist (nicht schief)

Ansonsten besser *Perzentile*, z. B. 5- und 95-Perzentil:

Bsp: Wie lange dauerten die 5 schnellsten Antwortzeiten?



# Konfidenzintervall

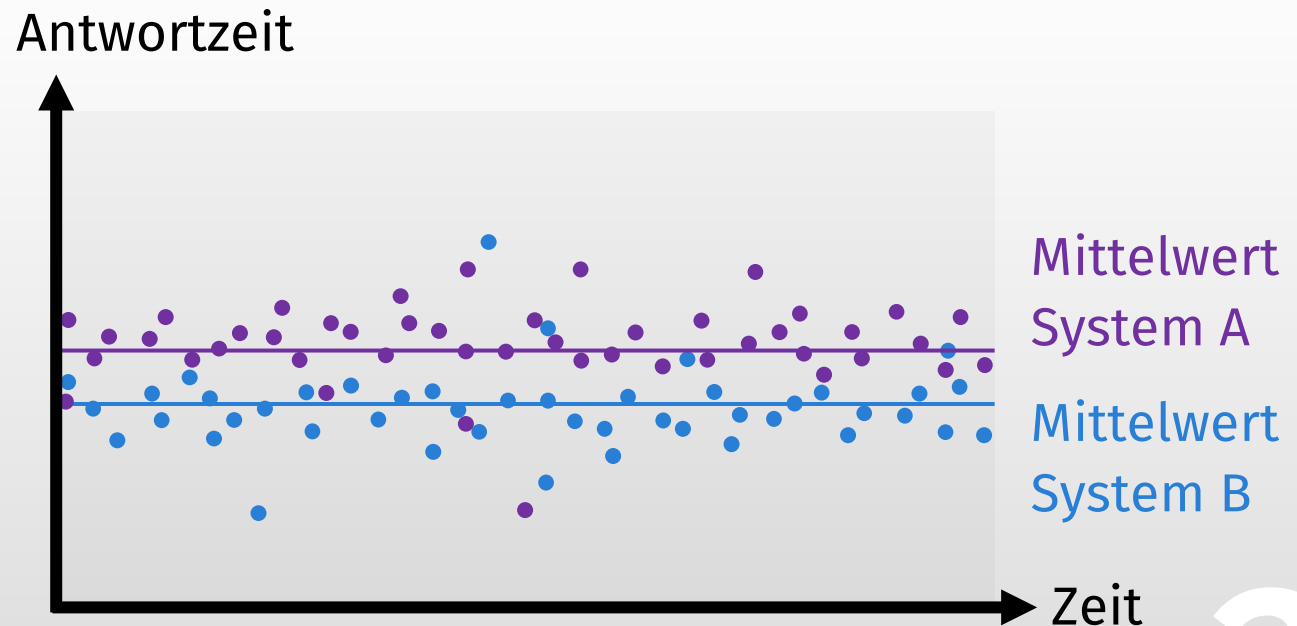
Grundsätzliches Problem: Haben immer nur **endliche Anzahl von Messwerten**. **Unmöglich wahren Mittelwert/Varianz zu bestimmen!**

Mit «genug» grosser Stichprobe können wir relativ sicher sein, «gute» Schätzung zu erhalten. Aber wie viel ist «genug» und wie gut ist «gut»?

**Problem vor allem  
beim Vergleichen  
von Alternativen!**

Ist *System A* wirklich  
besser als *System B*?

-> kann nur mit Mittelwert nicht entschieden werden





Konfidenzintervall des Mittelwerts: Zwei Zahlen, zwischen denen wahrer Mittelwert mit bestimmter Wahrscheinlichkeit liegt, z. B. 95%.  
-> nicht garantiert



Bedeutet konkret: Wenn man Konfidenzintervall 100× mit bestimmter Anzahl Messwerte berechnen würde, wäre wahrer Mittelwert durchschnittlich 5× nicht in Intervall enthalten...

je sicherer ich sein will, dass Mittelwert genau ist, desto grösser ist das Konfidenzintervall.

	A	B	C	D	E	F	G	H	I	J
1	Normalverteilung			Schätzung						
2	$\mu =$	100		n		5	10	20	50	100
3	$\sigma =$	5		Mittel		99.57	100.26	101.02	101.30	100.39
4				Std-Abw.		3.07	2.24	4.18	4.19	4.69
5	97.32			CI	90%	[96.6, 102.5]	[99.0, 101.6]	[99.4, 102.6]	[100.3, 102.3]	[99.6, 101.2]
6	104.14				95%	[95.8, 103.4]	[98.7, 101.9]	[99.1, 103.0]	[100.1, 102.5]	[99.5, 101.3]
7	98.23				99%			[98.3, 103.7]	[99.8, 102.8]	[99.2, 101.6]
8	96.91				99.9%			[97.4, 104.7]	[99.3, 103.2]	[98.9, 101.9]
9	101.23				99.99%			[96.4, 105.6]	[99.0, 103.6]	[98.6, 102.2]
10	101.03									
11	100.15									
12	101.30									

Demo

Genaue Formel für Konfidenzintervall recht kompliziert.

**Entscheidend:** Konfidenzintervall ist abhängig von 3 Grössen:

1. Je grösser die Streuung der Verteilung, desto **grösser** das Konfidenzintervall.
2. Je höher wir *Konfidenzniveau* wählen (95%, 99%, ...), desto **grösser** das Konfidenzintervall
3. Je grösser die Stichprobe, desto **kleiner** das Konfidenzintervall!

Guter Grund für mehr Replikation!

n = 20:

*n = Wiederholungen*

```
Result "findDocs":  
  2.305 ±(99.9%) 0.254 ops/s [Average]  
  (min, avg, max) = (1.679, 2.305, 2.733), stdev = 0.292  
  CI (99.9%): [2.051, 2.559] (assumes normal distribution)
```

n = 50:

```
Result "DocFinderBenchmarks.findDocs":  
  2.321 ±(99.9%) 0.113 ops/s [Average]  
  (min, avg, max) = (1.863, 2.321, 2.623), stdev = 0.228  
  CI (99.9%): [2.208, 2.434] (assumes normal distribution)
```

*Standardabweichung*

# **Präsentation von Resultaten**

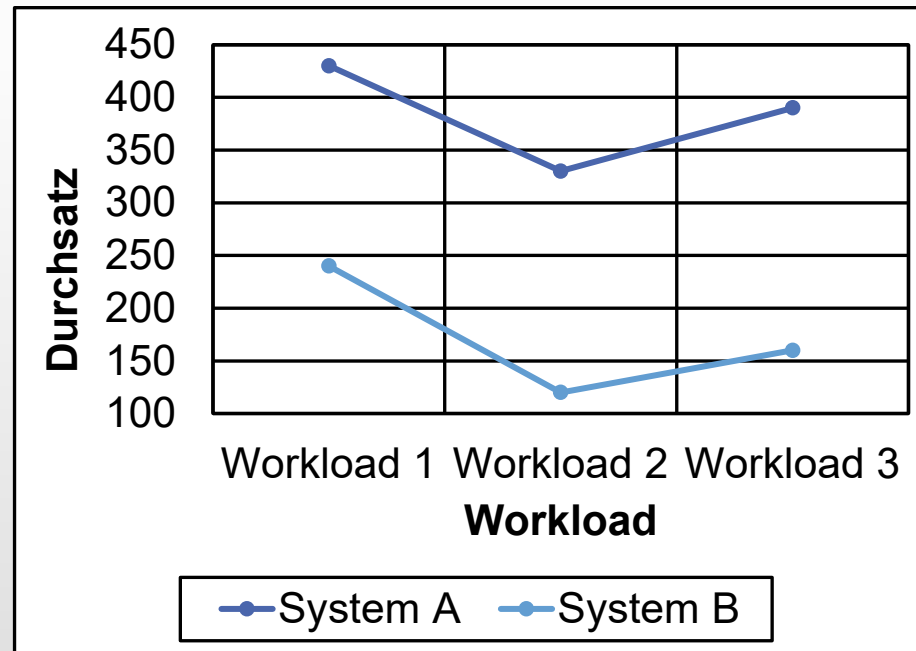
# Diagramme

Statistik ist Grundlage für letzten Schritt in Performance-Analyse:

Präsentation der Resultate: Visualisierung in Form von Diagrammen

**Diskussion:** Welche Probleme hat dieses Diagramm?

- Einheit fehlt
- Was ist der Wert?
- Linien sind falsch
- Achsenbeschriftung



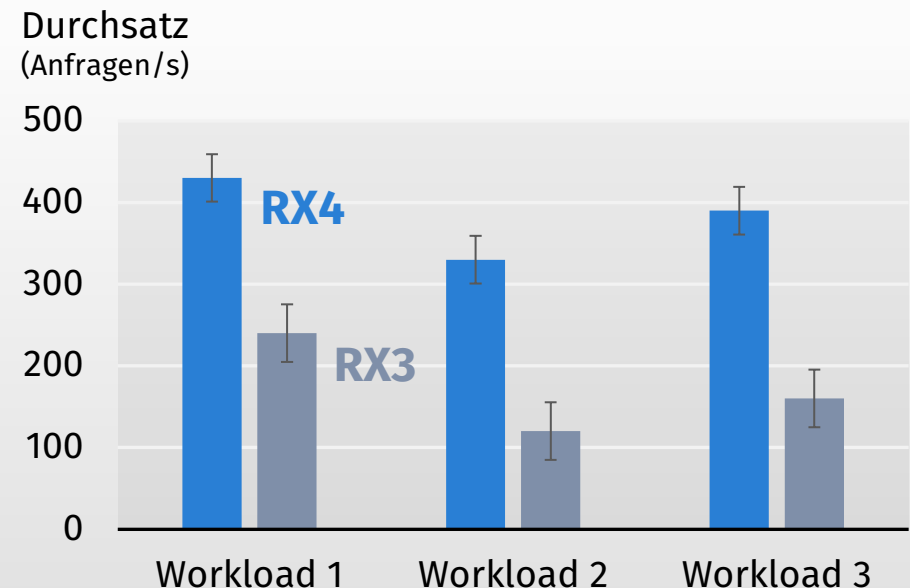
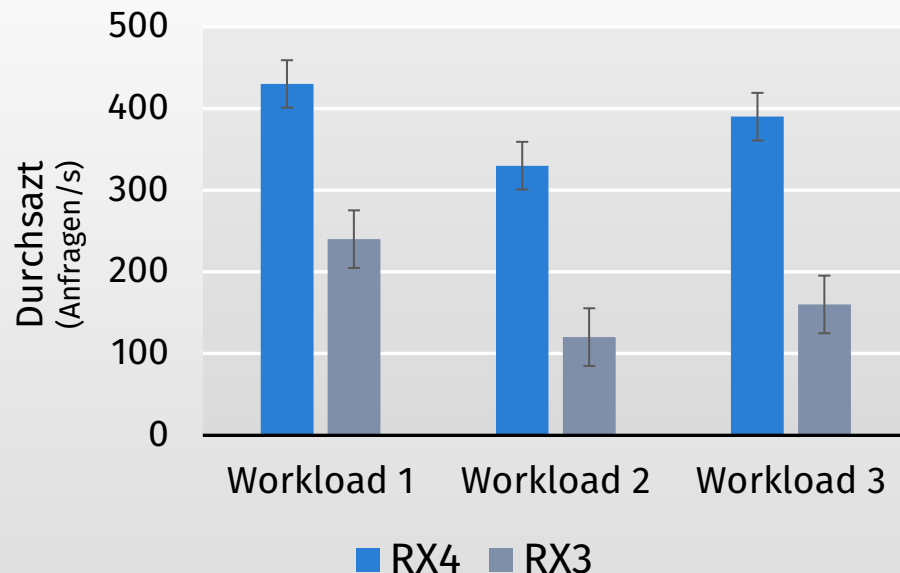
# Richtlinien für gute Diagramme

## 1. Aufwand für den Leser minimieren

Wichtigste «Metrik» für Diagramm: *Wie viel Aufwand ist nötig, dass die Message des Diagramms beim Leser ankommt?*

Daten werden benötigt, um die Message aufzuzeigen

**Beispiele:** Legende vs. direkte Beschriftung, Textausrichtung

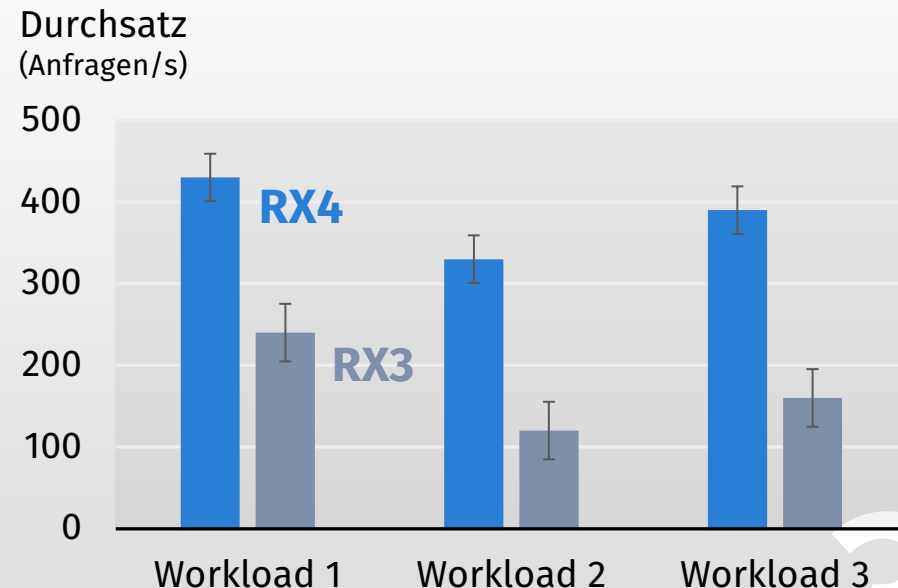
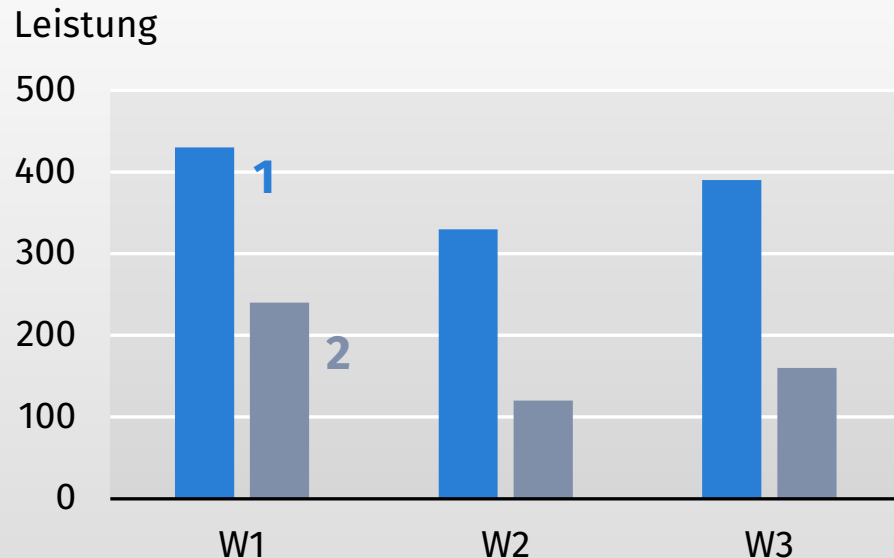


## 2. Aussagekraft maximieren

Diagramm sollte **möglichst selbsterklärend** sein und keine Möglichkeit für Missverständnisse bieten.

### Beispiele

- Stichwörter statt Symbole
- Präzise Beschriftung von Achsen, Einheiten
- Fehlerbalken, z. B. Standardabweichung oder Konfidenzintervall

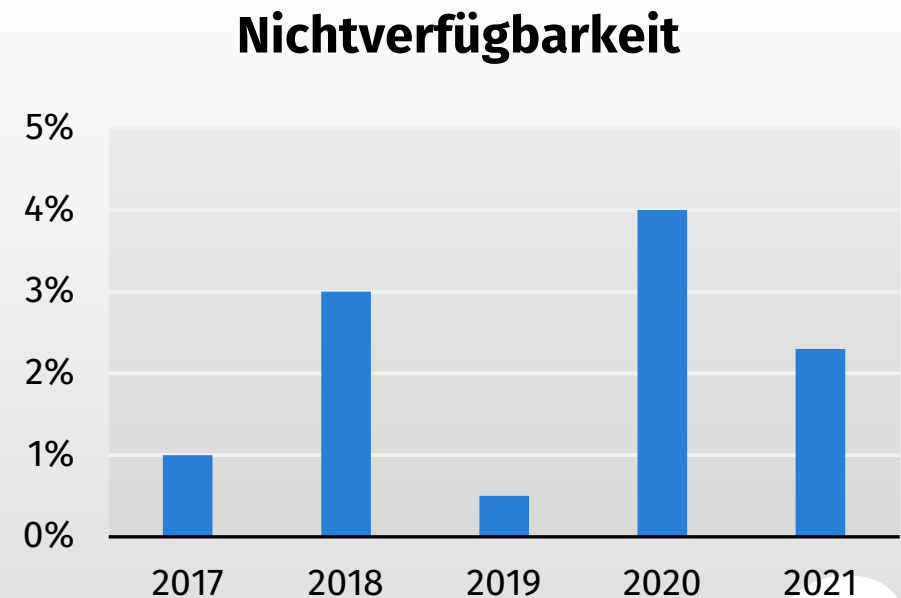
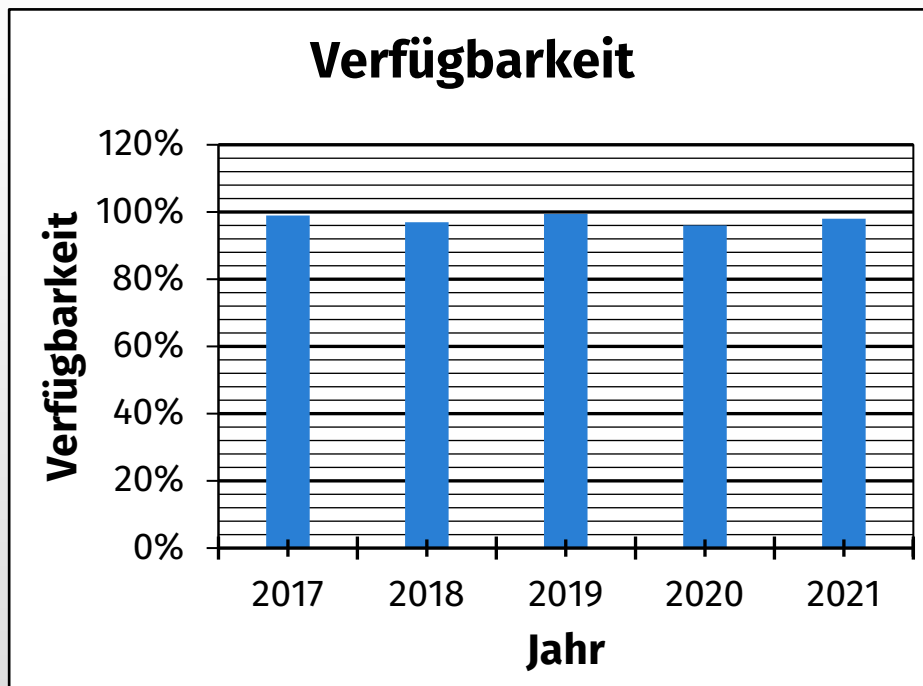


### 3. «Tinte» minimieren

Ziel: *Information-zu-Tinten-Verhältnis* maximieren. Nicht, um Druckertinte zu sparen, sondern um Ablenkung zu minimieren.

#### Beispiele

- Gitterlinien dezent, nur falls nötig; keine Rahmen
- Auswahl der dargestellten Grösse



## 4. Konventionen befolgen

Personen haben Erwartungen an Diagramme. Wenn Diagramm Erwartungen nicht entspricht: Missverständnisse, mehr Aufwand.

### Beispiele

- Ursprung bei (0, 0) und links unten
- Unabhängige Variable auf x-Achse, Antwortvariable auf y-Achse
- Lineare Achsen (Bsp: Achsenabstand sind immer gleich!)

Generell: Abweichungen von Richtlinien sind erlaubt, aber müssen gut begründet sein (und evtl. erläutert werden).

*«Zuerst muss man die Regeln kennen, um sie brechen zu können.»*



**Fragen?**

