

Application Performance Management

Datenauswertung & Präsentation

Michael Faes

Auswertung & Präsentation

Ziel jeder Performance-Analyse: Hilfestellung für Entscheidungen

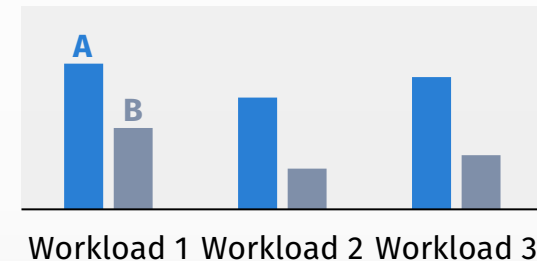
→ Wichtiger Schritt: **Präsentation/Kommunikation der Ergebnisse**

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

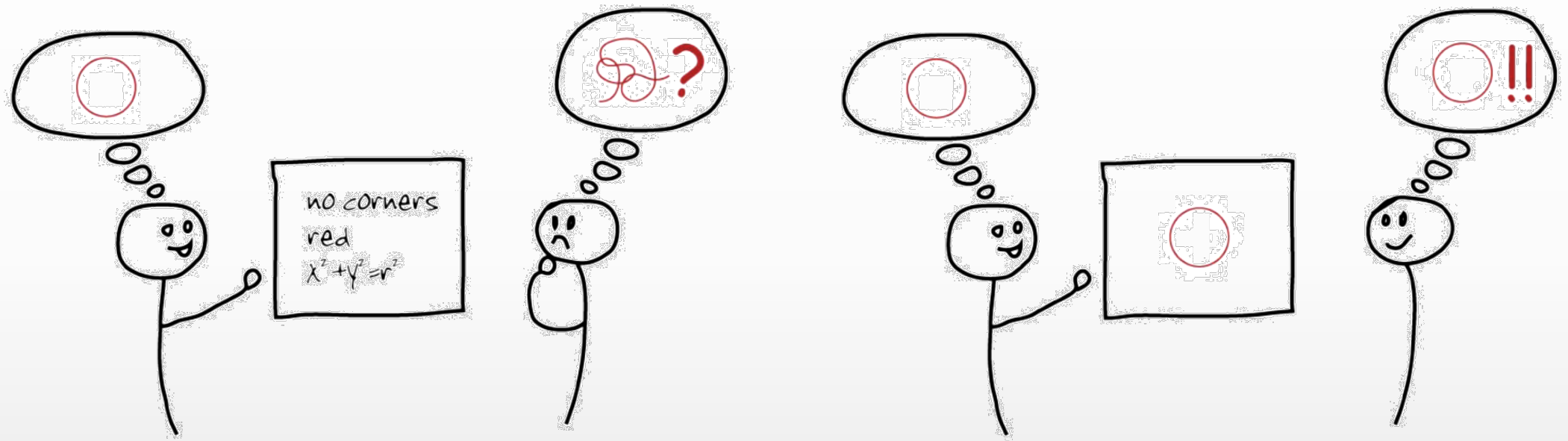


Diagramme

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



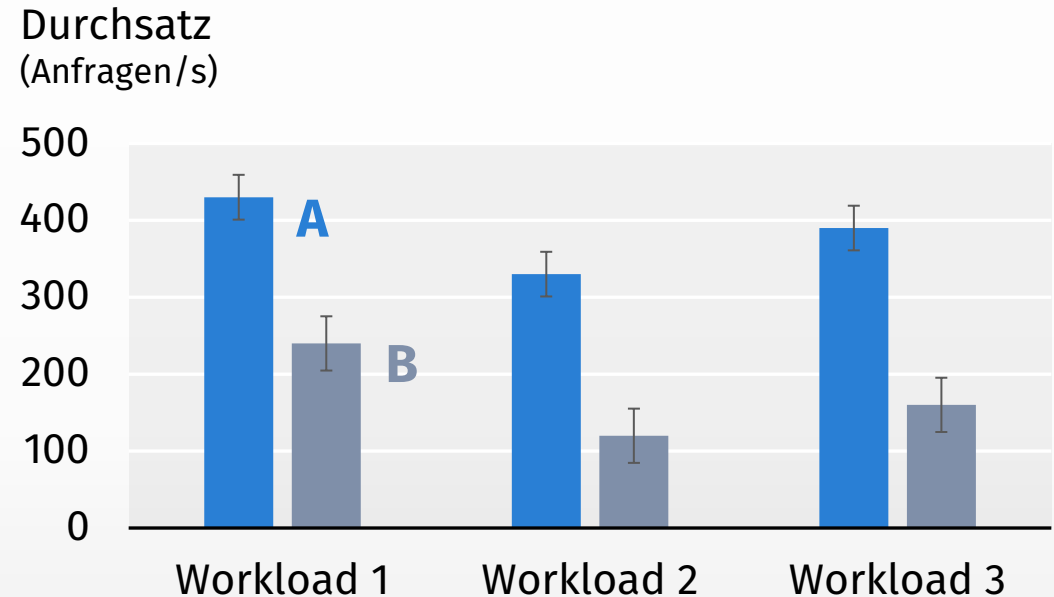
Text

Bilder, Diagramme, ...

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



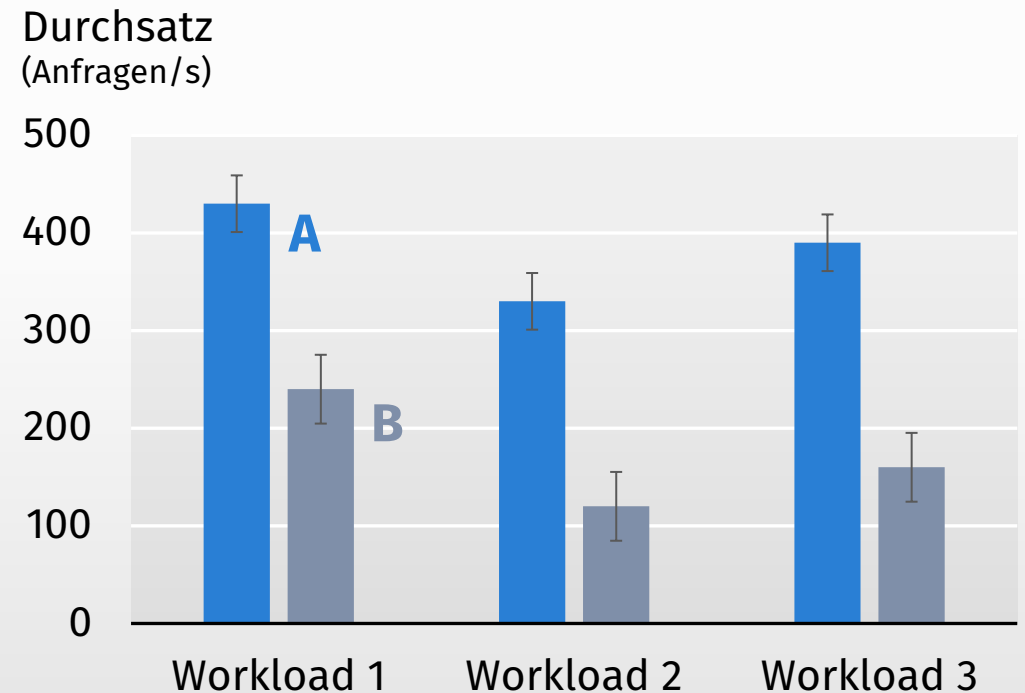
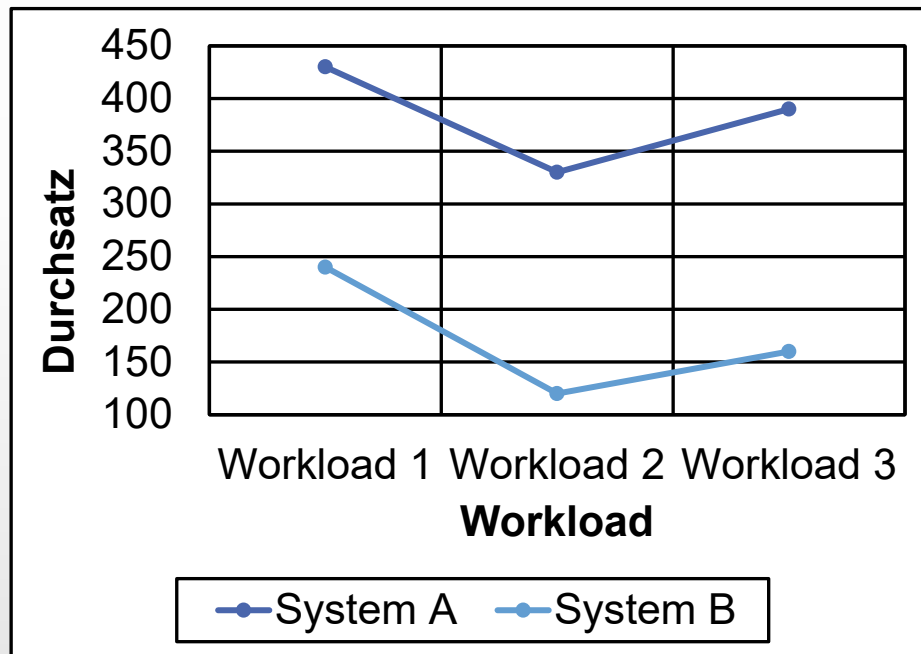
Diagramm

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?



Übersicht Woche 4

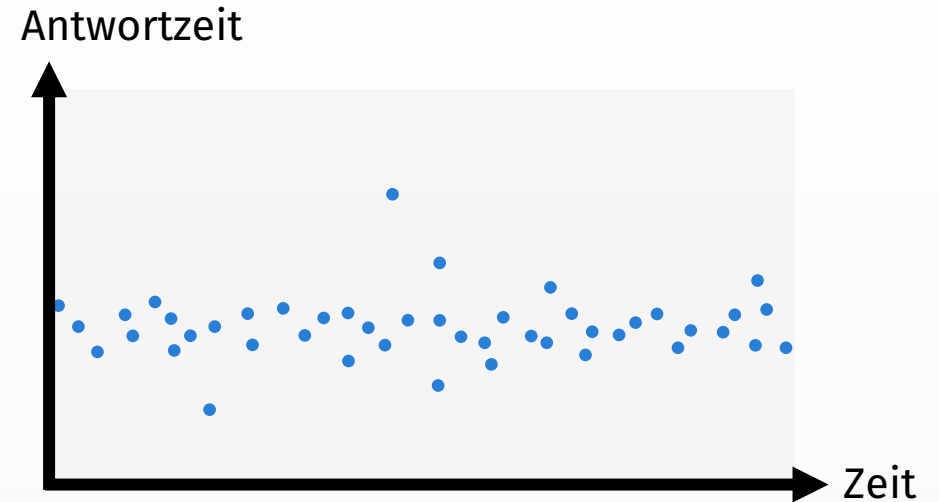
1. Übungsbesprechung
2. Messen & Auswerten: Grundlagen
3. Statistische Auswertung von Performance-Messungen
4. Präsentation von Resultaten
5. Übung: *Präsentation von Performance-Resultaten*

Messen & Auswerten: Grundlagen

Zufallsvariablen

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

Grund: Unkontrollierbare, *zufällige* Unterschiede in Umgebung.



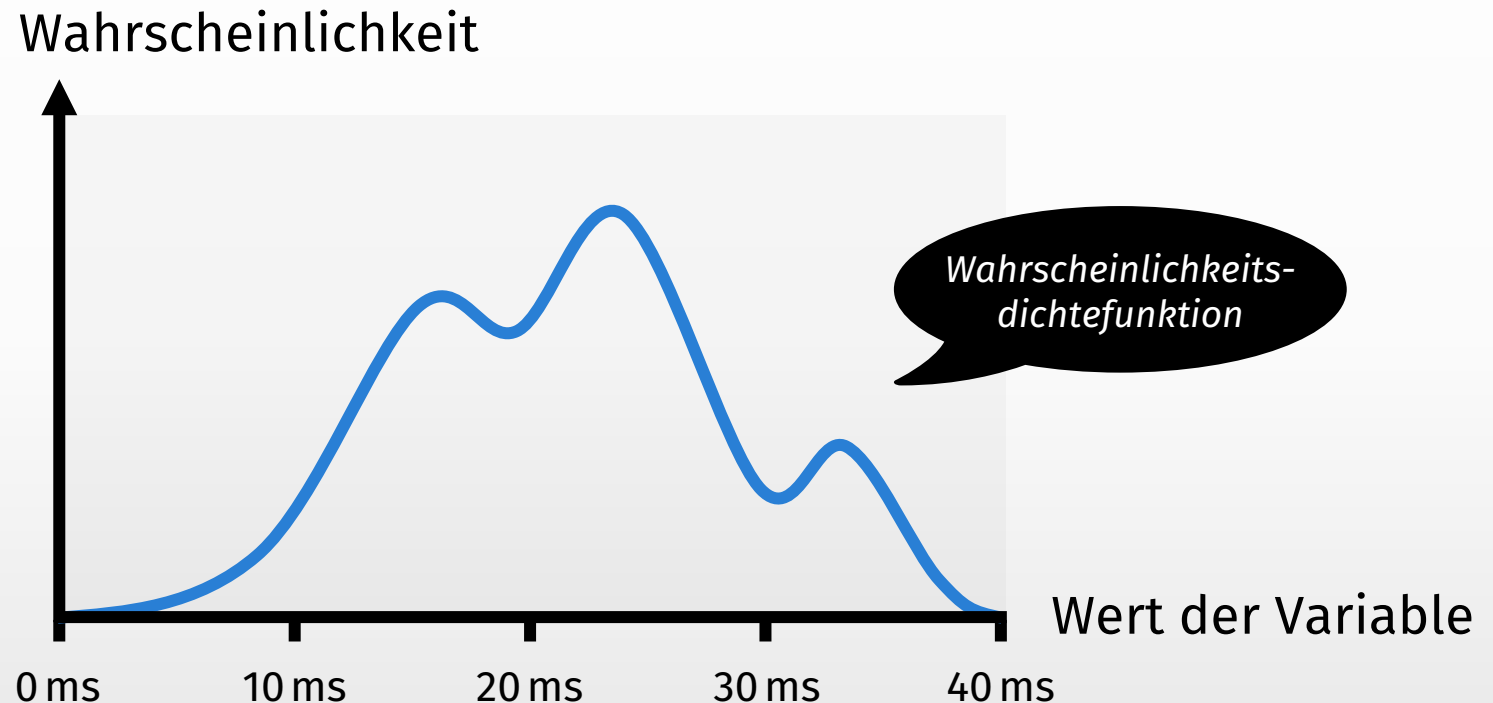
Modellierung als *Zufallsvariable*:

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

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



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

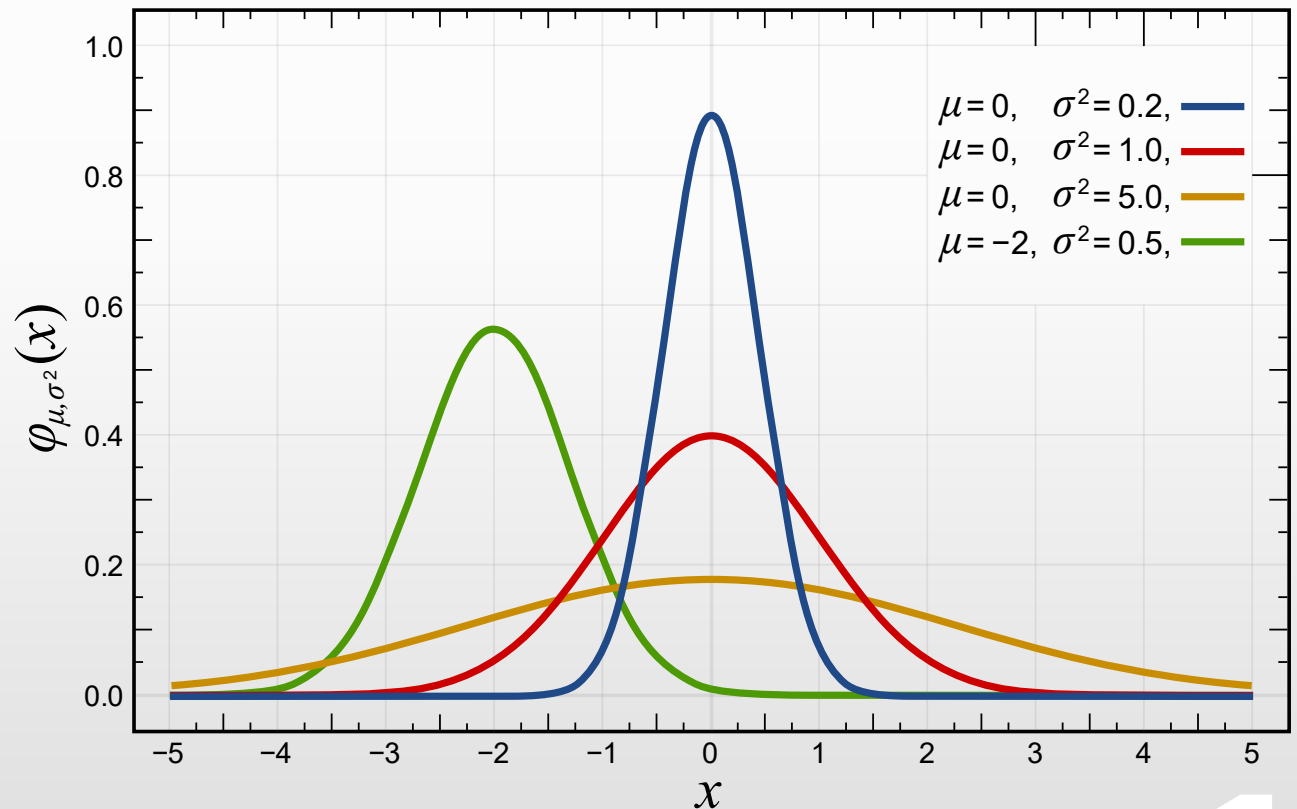
Normalverteilung

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

Typische Annahme:

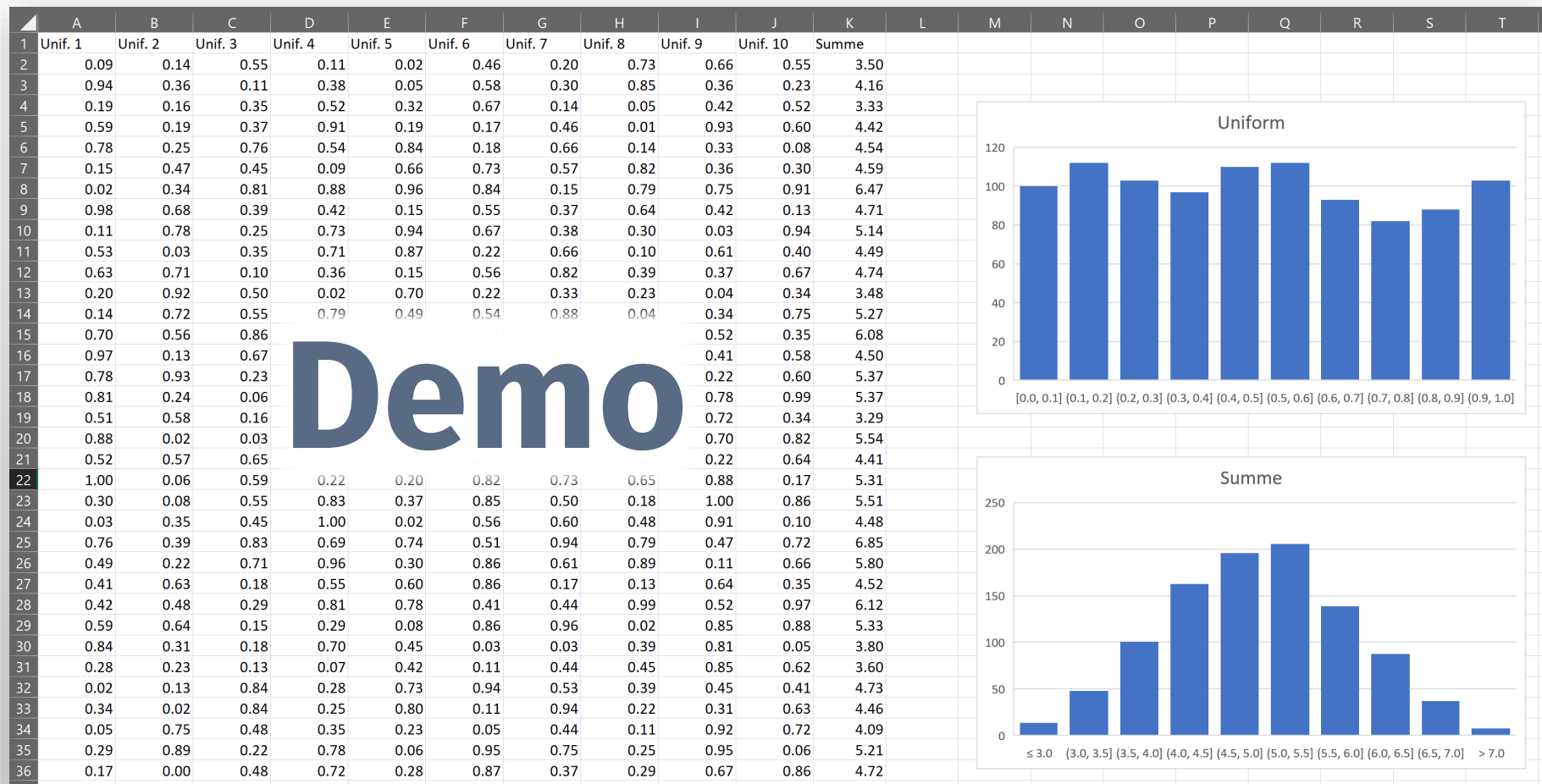
Normalverteilung

Warum eigentlich???



Zentraler Grenzwertsatz

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



Normalverteilung und Performance

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

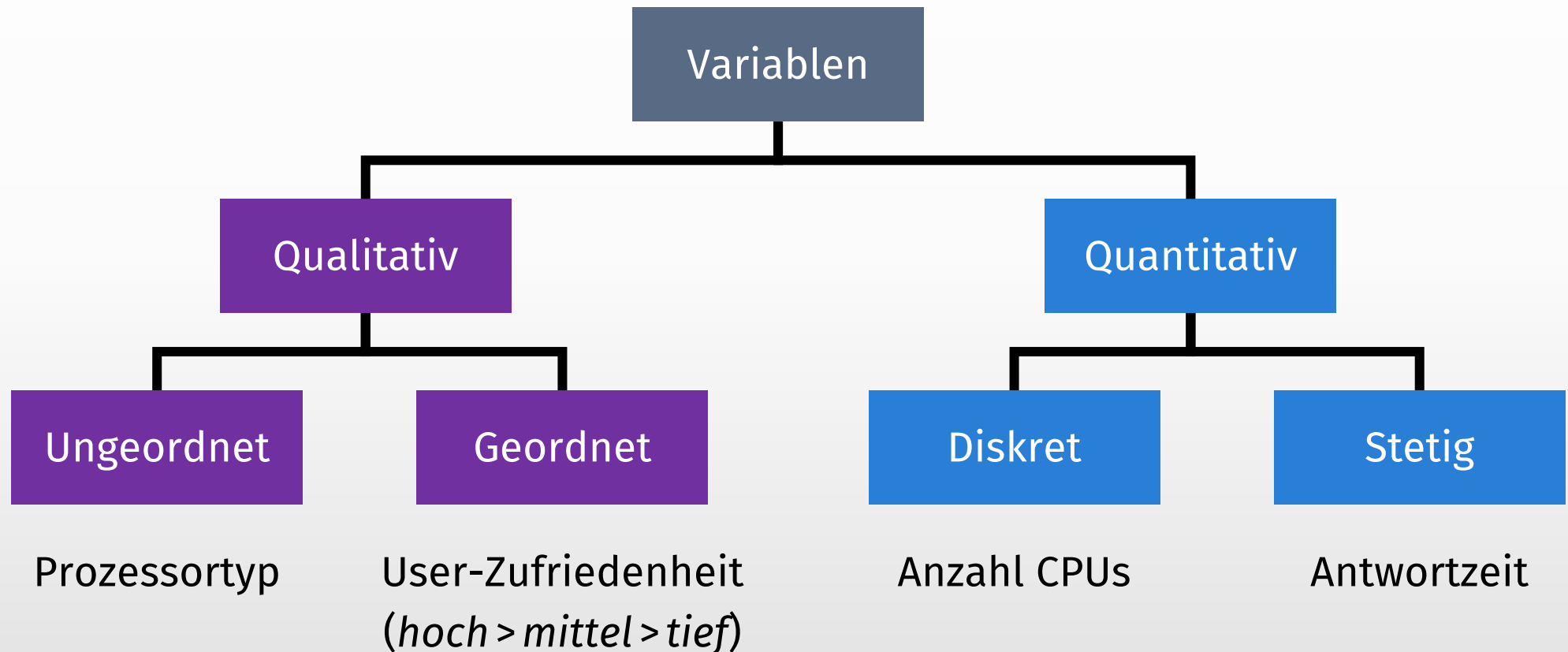
Einige statistische Verfahren *gehen von Normalverteilung aus*:

```
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)
```

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

Typen von Variablen

Nicht alle Messungen/Beobachtungen liefern gleiche Art von Variable:



Übung: Typen von Variablen

Welche Variablentypen liefern folgende Messungen/Beobachtungen?

- a) Ausfälle seit Beginn des Jahres
- b) Durchsatz
- c) häufigste Anfragenart
- d) CPU-Auslastung
- e) Usability-Bewertung (Noten 1 bis 6)
- f) am stärksten ausgelasteter Server
- g) durchschn. Garbage-Collections pro Sekunde
- h) Anzahl *Unique Users* im Februar 2023

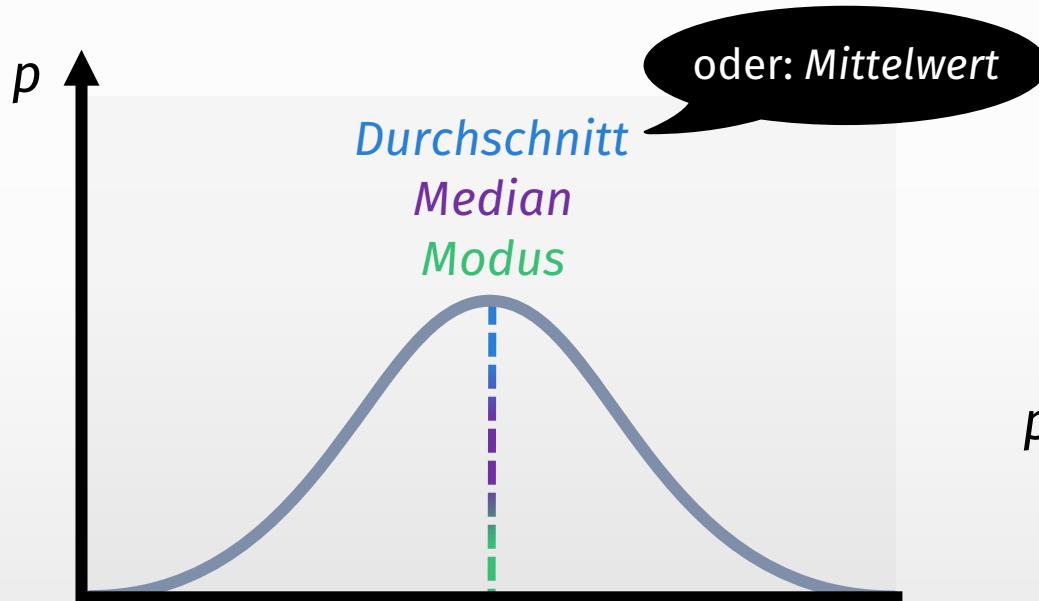
Übung: Typen von Variablen

Welche Variablentypen liefern folgende Messungen/Beobachtungen?

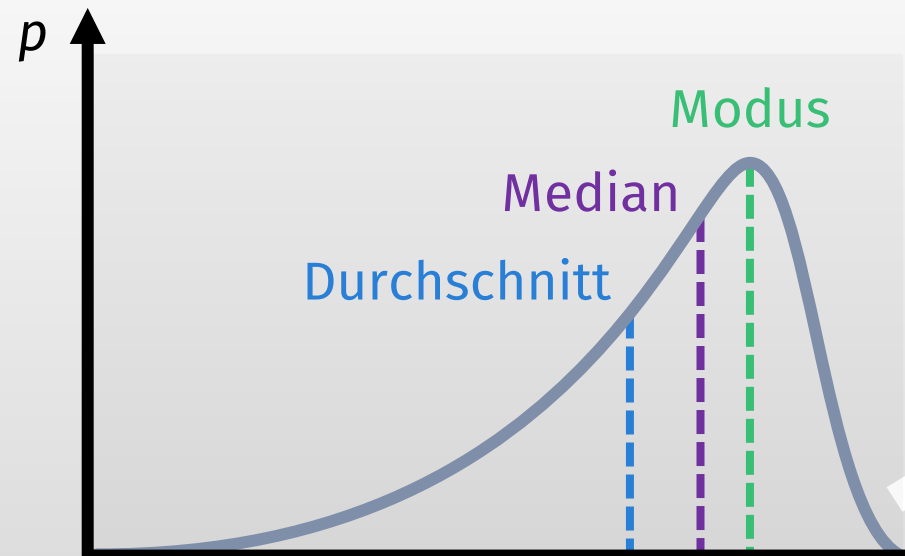
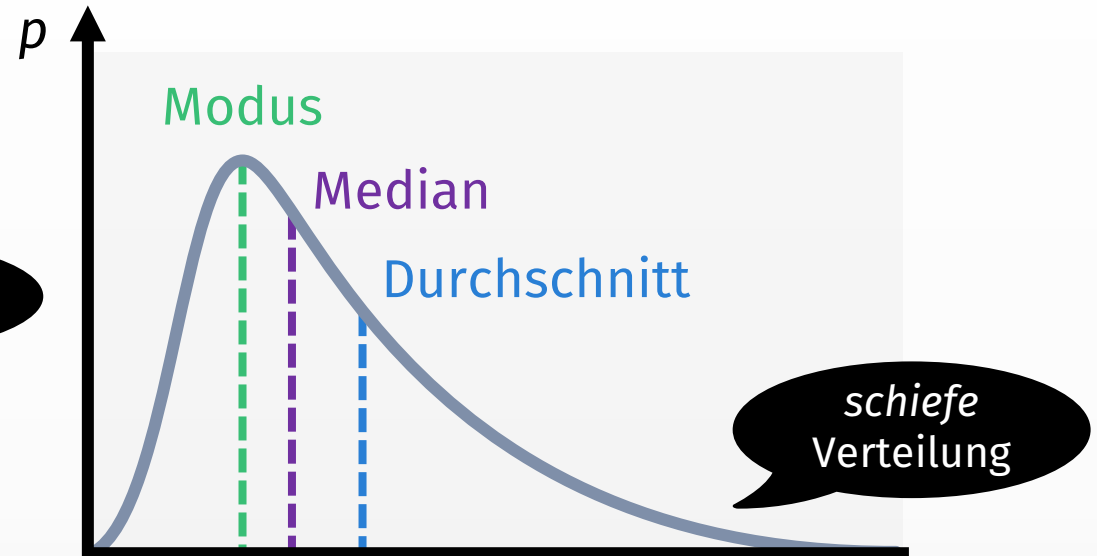
- | | |
|---|------------------------|
| a) Ausfälle seit Beginn des Jahres | Diskret |
| b) Durchsatz | Stetig |
| c) häufigste Anfragenart | Ungeordnet |
| d) CPU-Auslastung | Stetig |
| e) Usability-Bewertung (Noten 1 bis 6) | Geordnet (qualitativ!) |
| f) am stärksten ausgelasteter Server | Ungeordnet |
| g) durchschn. Garbage-Collections pro Sek. | Stetig |
| h) Anzahl <i>Unique Users</i> im Februar 2023 | Diskret |

Durchschnitt, Median & Modus

Einfachste Art, Zufallsvariable
zusammenzufassen: 1 Wert



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



Variablentypen & Auswertung

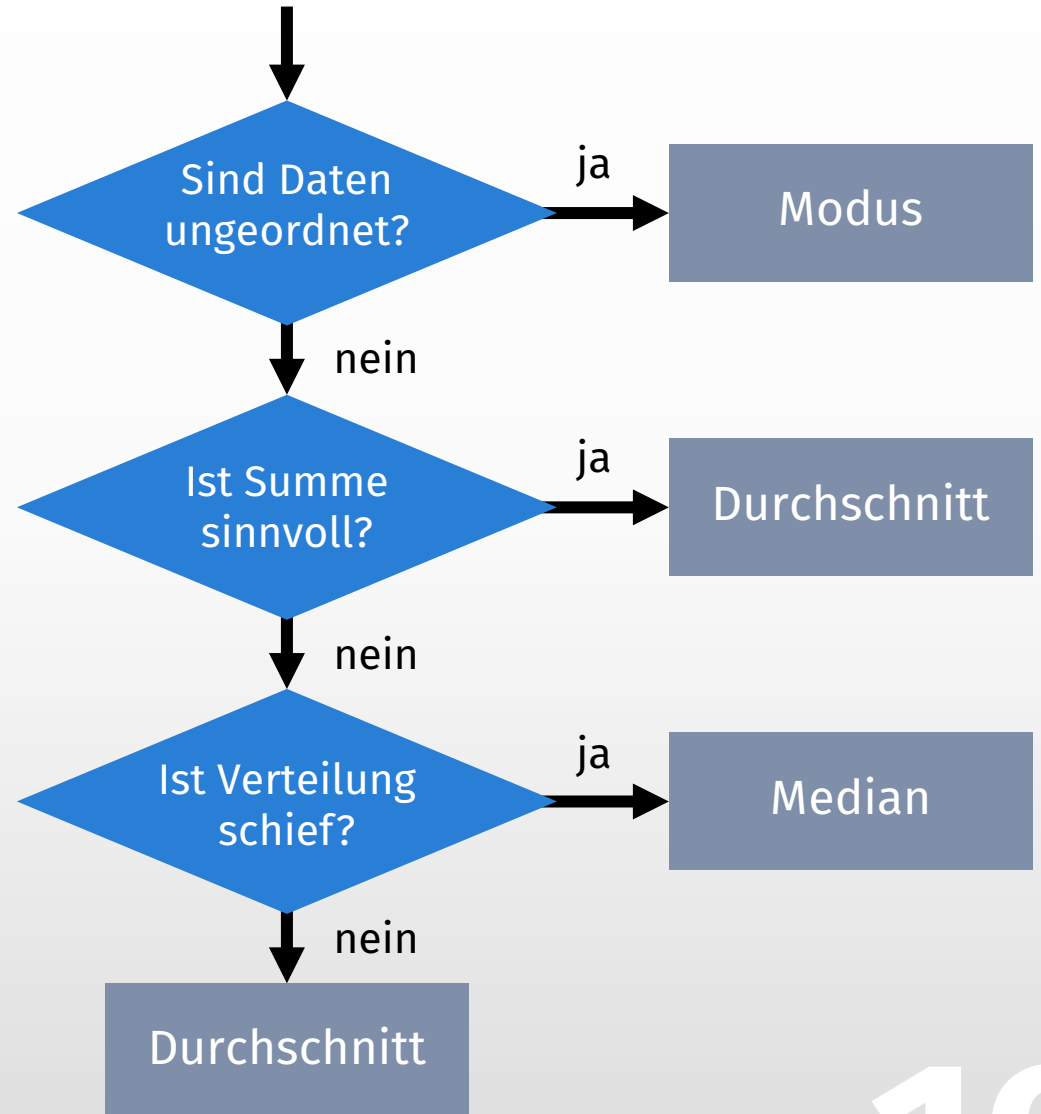
Nicht alle Variablentypen erlauben alle Arten von Auswertungen!

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

Statistische Auswertung von Performance-Messungen

Durchschnitt, Median oder Modus?

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.

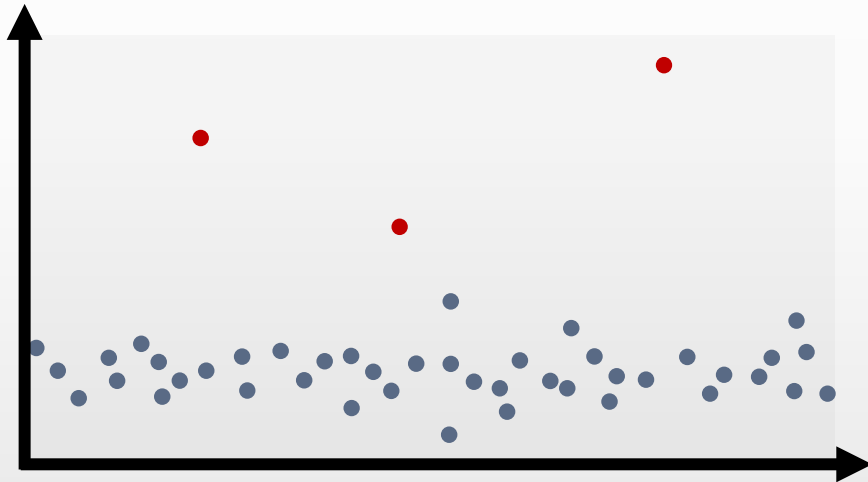
Beispiel 2: Durchschnitt/Median für Verteilung mit mehreren Peaks:



2. Durchschnitt ohne Rücksicht auf Ausreisser

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

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

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%

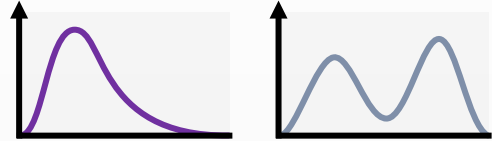
Messdauer (s)	CPU-Ausl.
1	45 %
1	45 %
1	45 %
1	45 %
100	20 %
Summe	200 %
Durchschnitt	40 %

Übung: Durchschnitt, Median oder Modus?

Entscheiden Sie, ob man zum Zusammenfassen folgender Daten am ehesten den Durchschnitt, den Median oder den Modus (oder ...?) verwenden sollte (ohne die genauen Daten zu kennen):

- a) Ausfälle pro Monat (letzte 36 Monate)
- b) letzte 1000 Antwortzeiten für versch. Anfragen
- c) häufigste Anfragenart pro Tag (letzte 30 Tage)
- d) CPU-Auslastung (letzte Stunde)
- e) Usability-Bewertung (Noten 1 bis 6)
- f) am stärksten ausgelasteter Server pro Stunde (letzte 24 h)

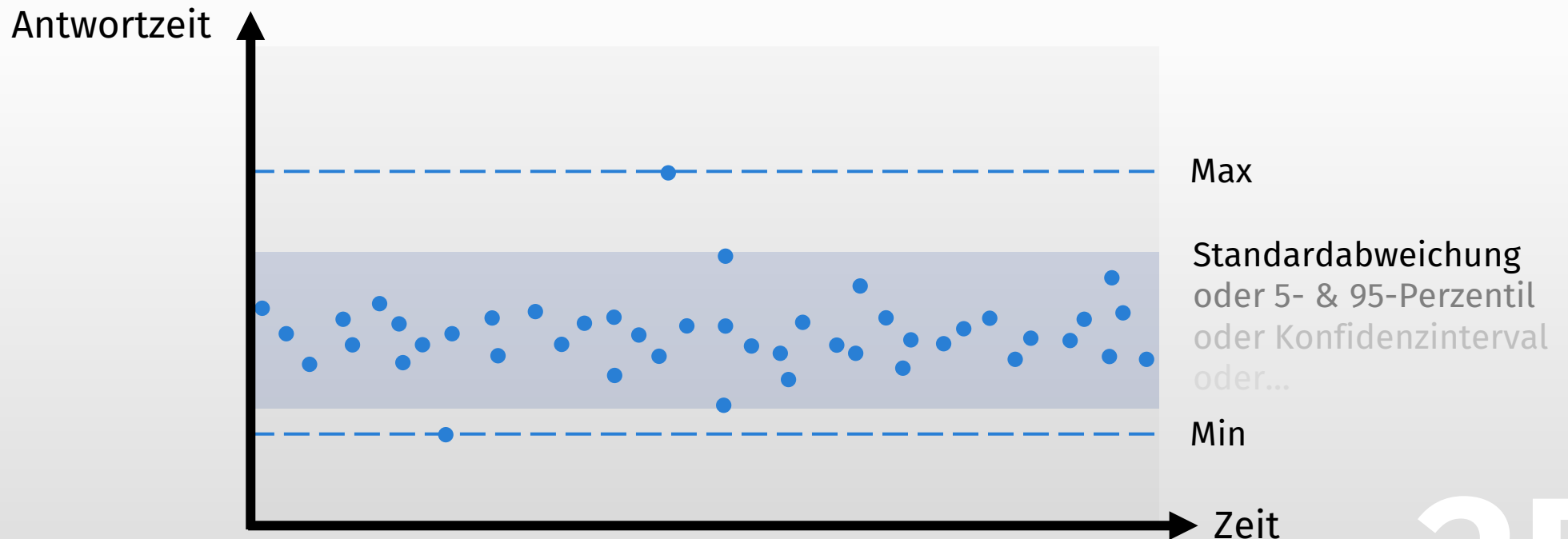
Übung: Durchschnitt, Median oder Modus?

- | | |
|---|--|
| a) Ausfälle pro Monat (letzte 36 Monate) | Durchschnitt
(Summe sinnvoll) |
| b) letzte 1000 Antwortzeiten für versch. Anfragen | Median / keines
 |
| c) häufigste Anfragenart pro Tag (letzte 30 Tage) | Modus |
| d) CPU-Auslastung (letzte Stunde) | Durchschnitt
(Summe sinnvoll) |
| e) Usability-Bewertung (Noten 1 bis 6) | Median
(Summe sinnlos) |
| f) am stärksten ausgelasteter Server pro Stunde | Modus |

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.»



Varianz & Standardabweichung

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

- Oft nicht besonders nützlich, wegen Ausreissern.

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

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

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.

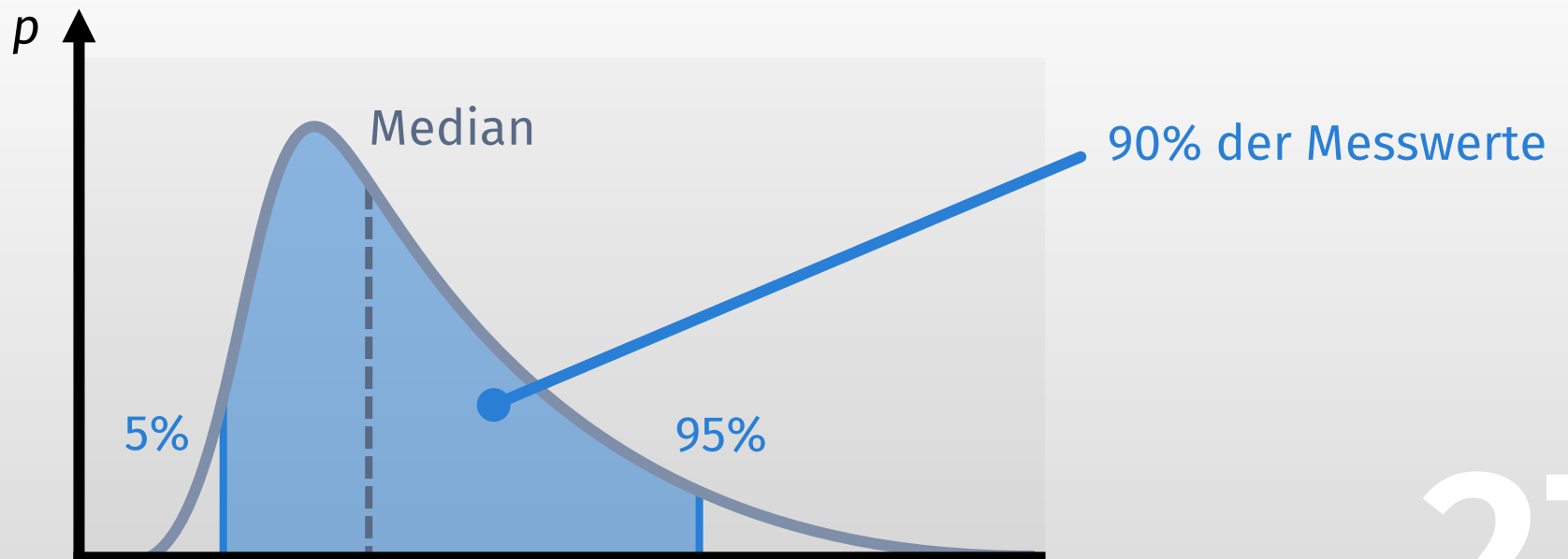
- Qualität der Schätzung hängt von *Grösse der Stichprobe* ab

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:



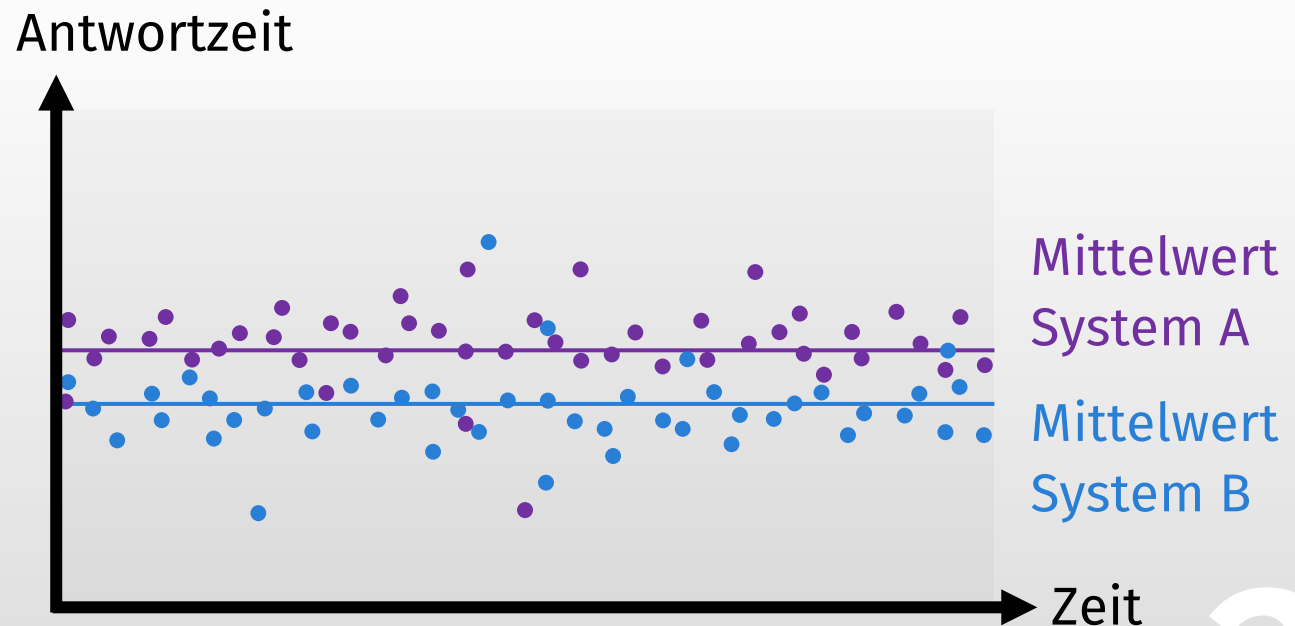
Konfidenzintervall

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*?



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



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

	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:

```
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:

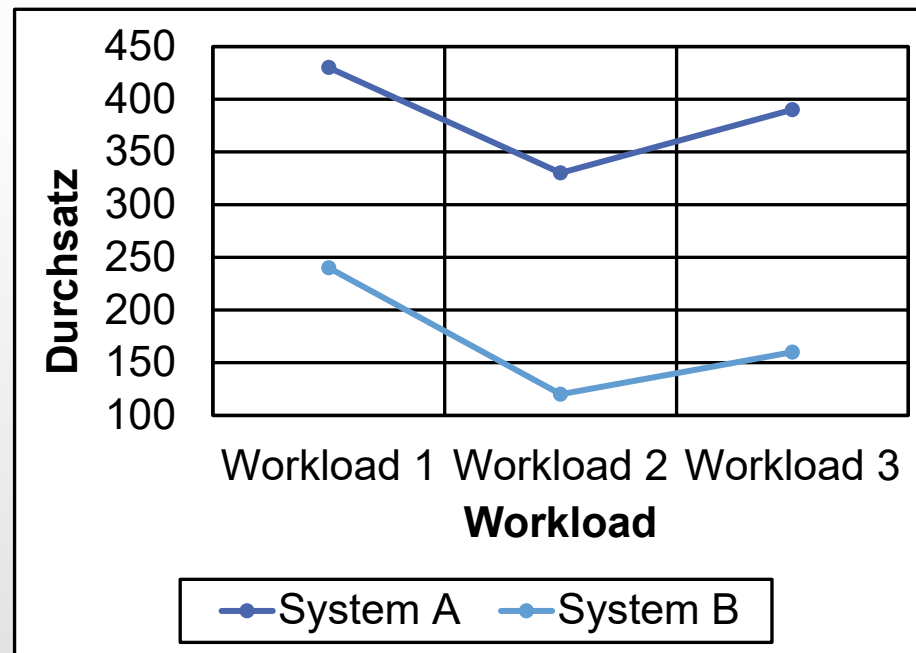
```
Result "DocFinderBenchmarks.findDocs":  
  2.321 ±(99.9%) 0.113 ops/s [Average]  
  (min, avg, max) = (1.863, 2.321, 2.623), stdev = 0.228  
  CI (99.9%): [2.208, 2.434] (assumes normal distribution)
```

Präsentation von Resultaten

Diagramme

Statistik ist Grundlage für letzten Schritt in Performance-Analyse:
Präsentation der Resultate: Visualisierung in Form von Diagrammen

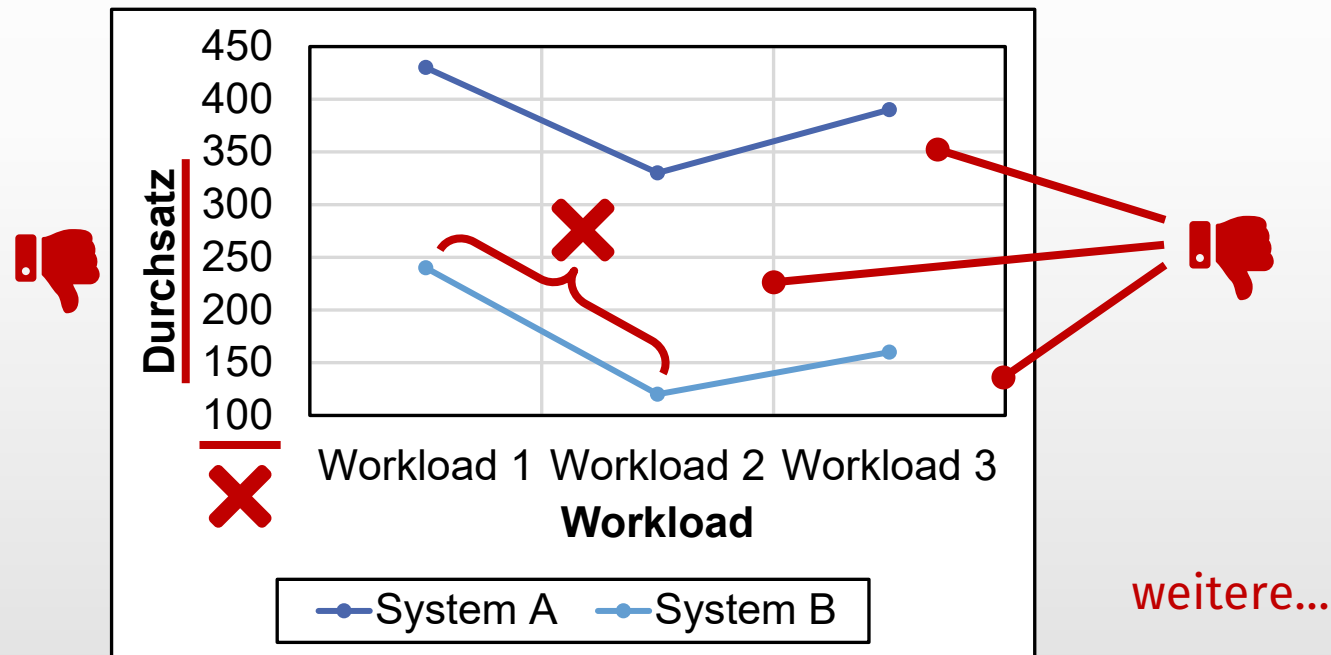
Diskussion: Welche Probleme hat dieses Diagramm?



Diagramme

Statistik ist Grundlage für letzten Schritt in Performance-Analyse:
Präsentation der Resultate: Visualisierung in Form von Diagrammen

Diskussion: Welche Probleme hat dieses Diagramm?



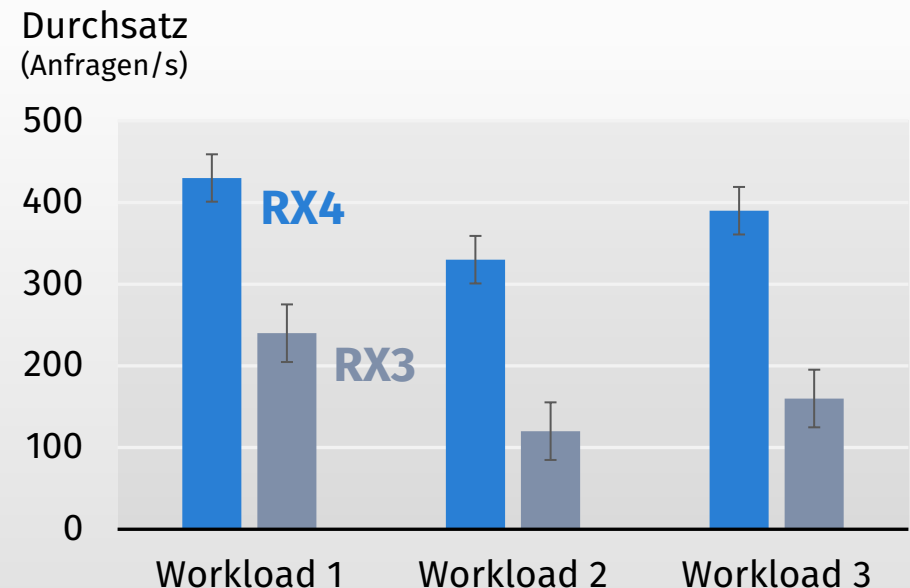
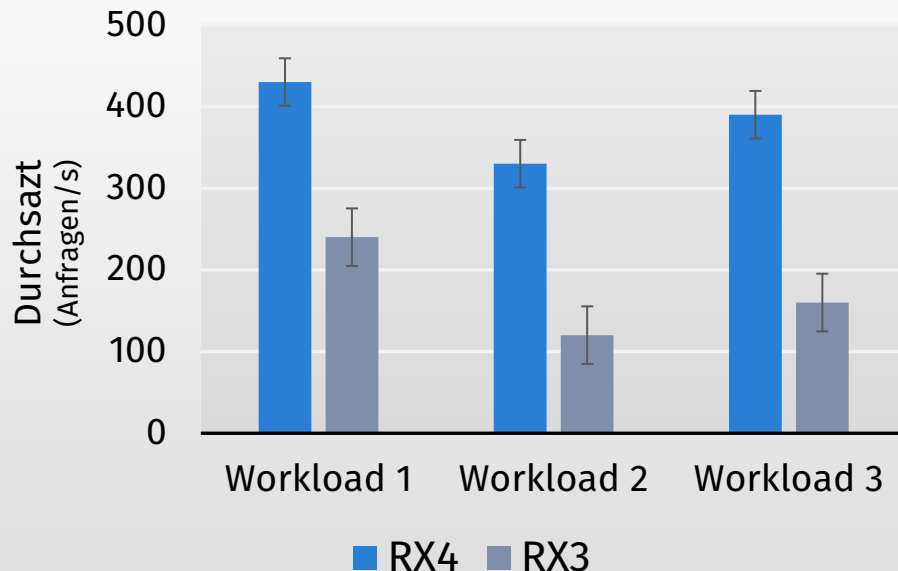
Richtlinien für gute Diagramme

1. Aufwand für den Leser minimieren

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



Beispiele: Legende vs. direkte Beschriftung, Textausrichtung



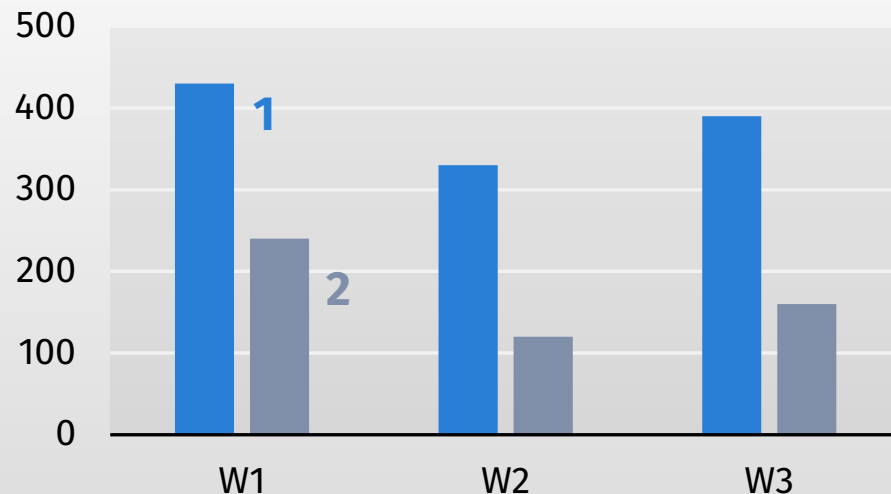
2. Aussagekraft maximieren

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

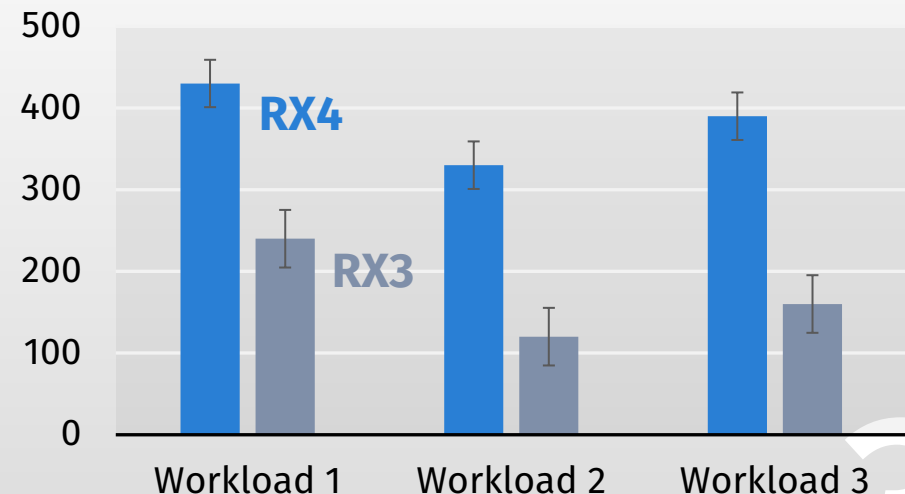
Beispiele

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

Leistung



Durchsatz
(Anfragen/s)

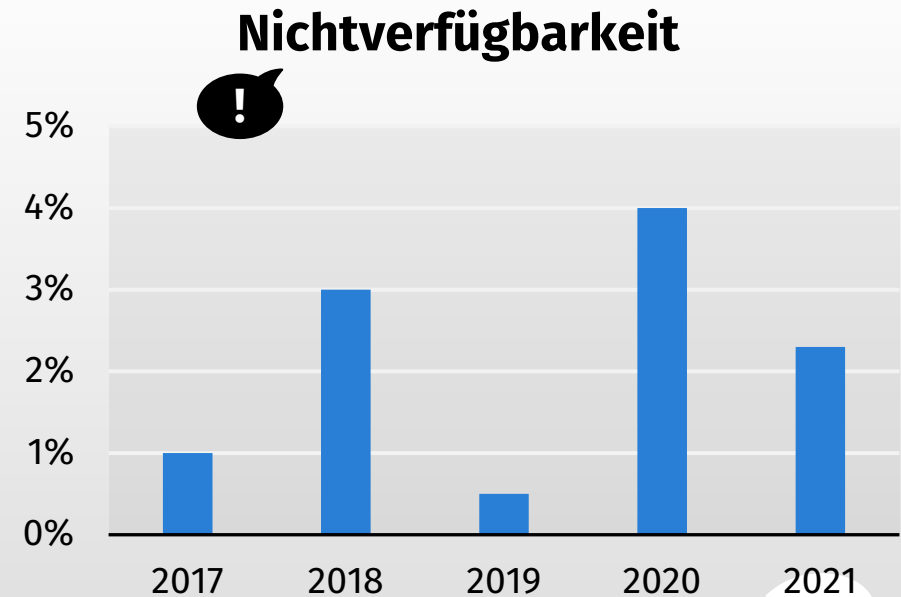
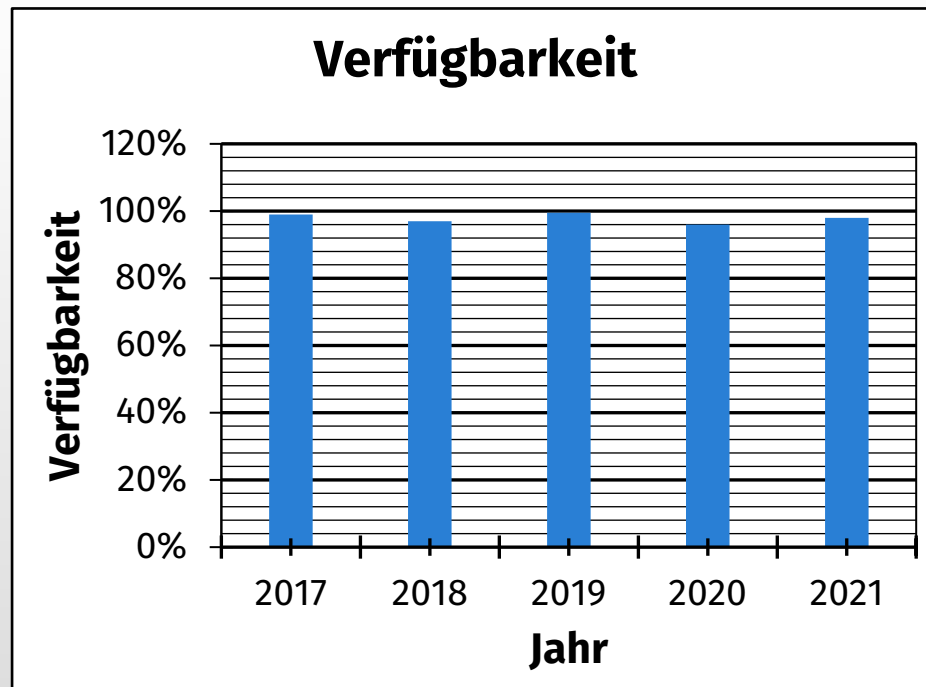


3. «Tinte» minimieren

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

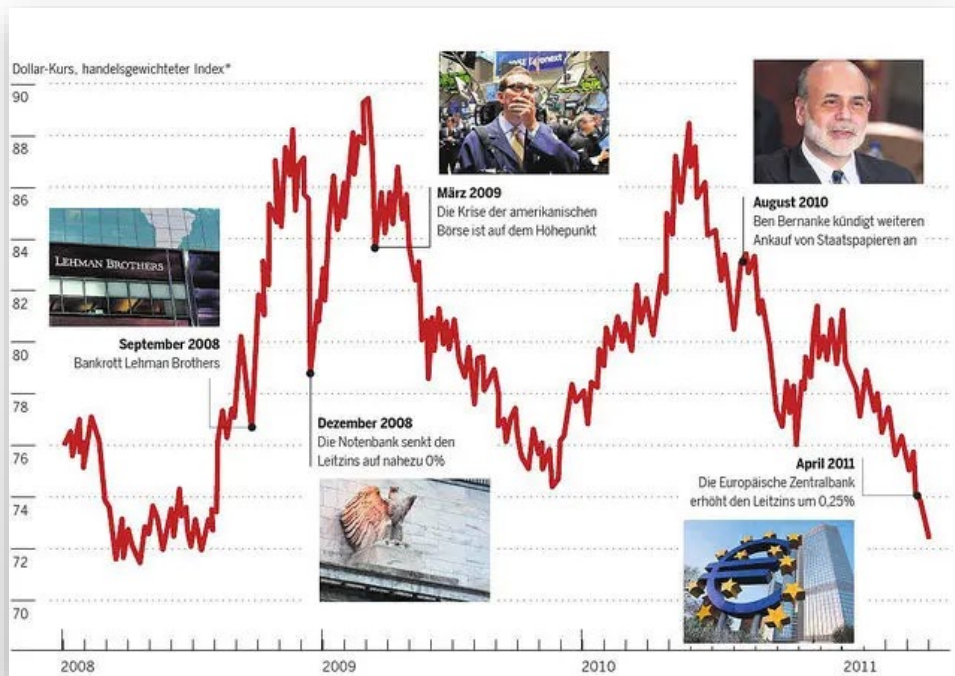
- Ursprung bei (0, 0) und links unten
- Unabhängige Variable auf x-Achse, Antwortvariable auf y-Achse
- Lineare Achsen

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.»

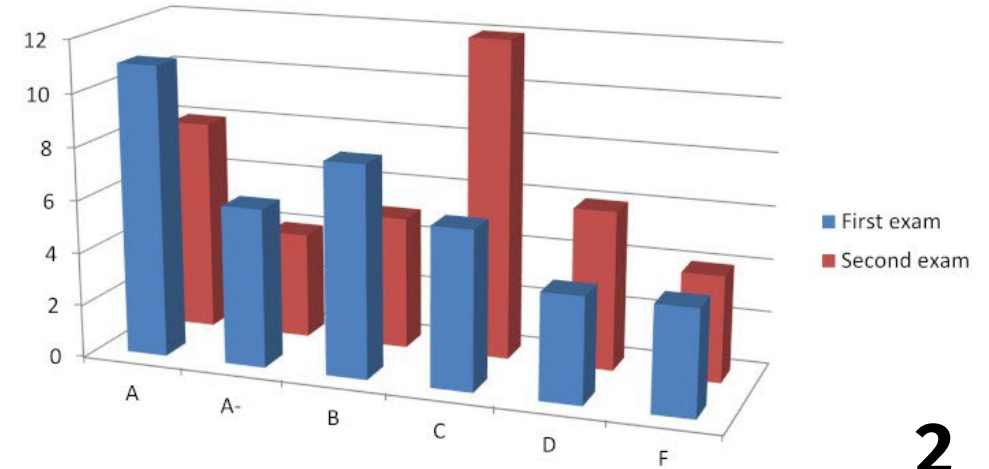
Übung: Diagramme

1

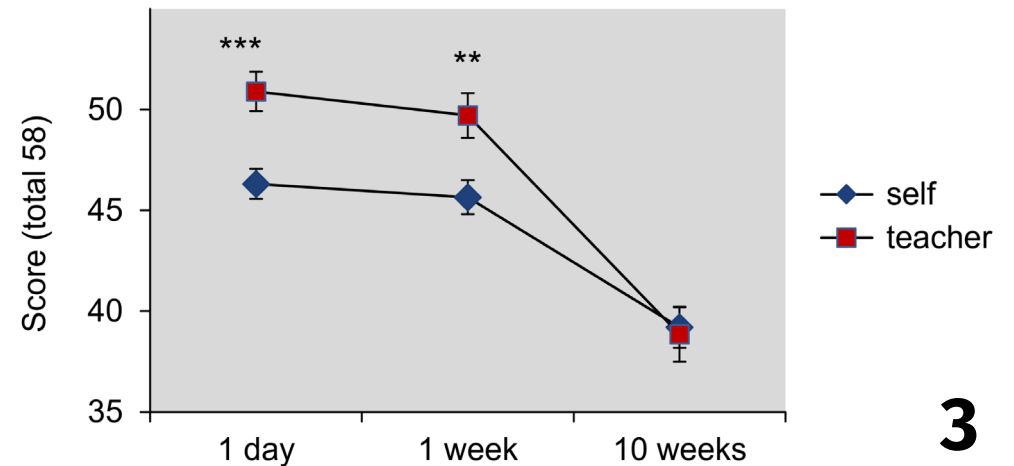


<https://www.tagesanzeiger.ch/schwere-schwindsucht-498113179659>

Grade Distribution



2



3

Fragen?

