

## Datenmanagement & -analyse

### Übung 7 – Explorative Datenanalyse

Dr. Nikolai Stein

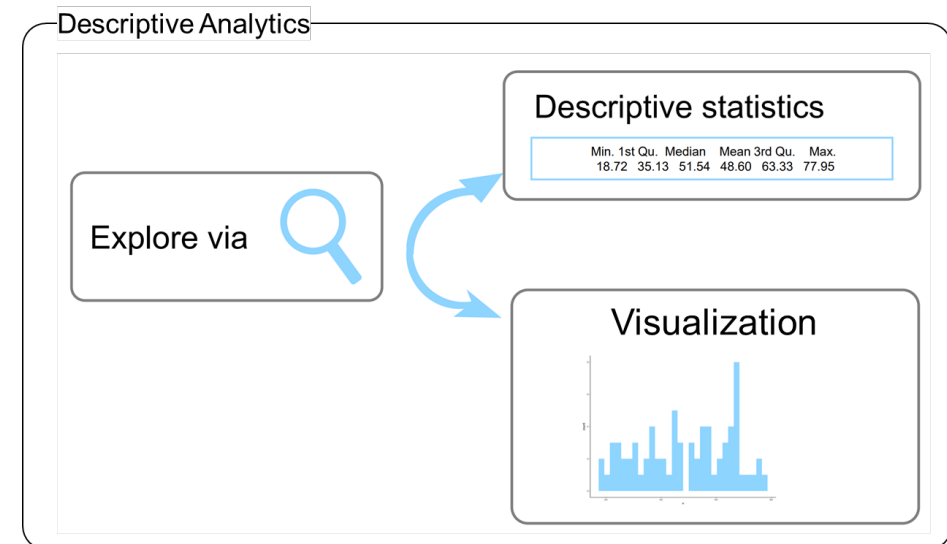
Lehrstuhl für WI & BA

Julius-Maximilians-Universität Würzburg

Sommersemester 2021

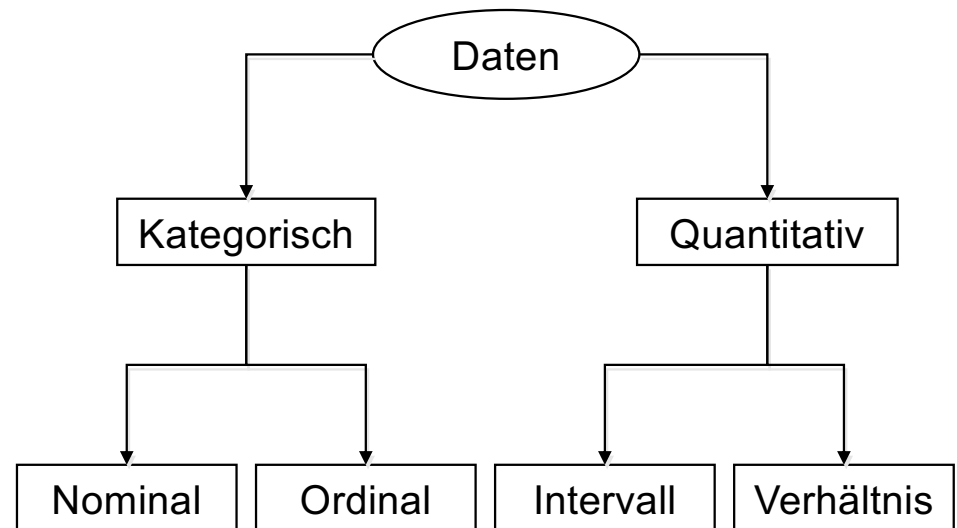


- Einer der wichtigsten und am meisten übersehenen Teile der Statistik ist die Explorative Datenanalyse (EDA)
- Die Ziele der EDA sind:
  - Hypothesen vorschlagen
  - Annahmen zu bewerten
  - Unterstützung bei der Auswahl geeigneter statistischer Verfahren
  - Eine Grundlage für die weitere Datenerhebung zu schaffen
- Außerdem hilft es, ein Gefühl für die Zahlen zu bekommen
  - Leichter, Fehler zu finden
  - Leichter zu erraten, was tatsächlich passiert ist
  - Leichter, ungerade Ausreißer zu finden



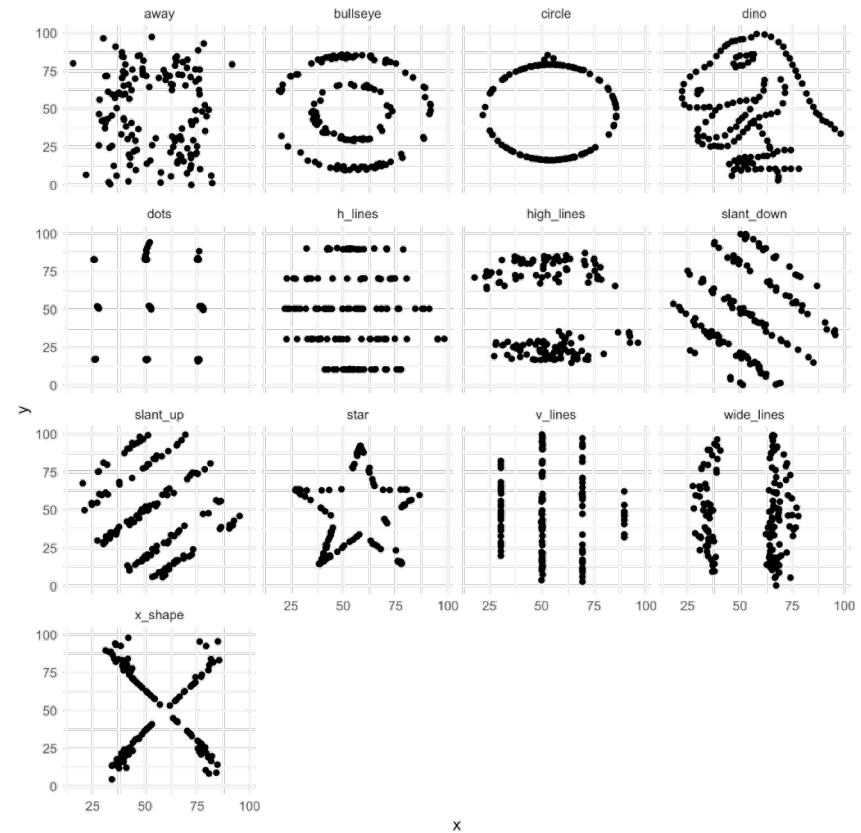
## Kategorische und quantitative Daten

- Kategorische (Qualitative) Daten
  - Bezeichnungen oder Namen, die zur Identifizierung eines Attributs eines jeden Elements verwendet werden
  - Kann entweder numerisch oder nicht numerisch sein
  - Geeignete statistische Analysen sind eher begrenzt
- Quantitative Daten
  - geben an, wie viele oder wie viel
  - Immer numerisch
  - Gewöhnliche arithmetische Operationen sind sinnvoll



## Die Grenzen der Statistik (2)

ID	$N$	$\bar{X}$	$\bar{Y}$	$\sigma_X$	$\sigma_Y$	$R$
1	142	54.3	47.8	16.8	26.9	-0.064
2	142	54.3	47.8	16.8	26.9	-0.069
3	142	54.3	47.8	16.8	26.9	-0.068
4	142	54.3	47.8	16.8	26.9	-0.064
5	142	54.3	47.8	16.8	26.9	-0.060
6	142	54.3	47.8	16.8	26.9	-0.062
7	142	54.3	47.8	16.8	26.9	-0.069
8	142	54.3	47.8	16.8	26.9	-0.069
9	142	54.3	47.8	16.8	26.9	-0.069
10	142	54.3	47.8	16.8	26.9	-0.063
11	142	54.3	47.8	16.8	26.9	-0.069
12	142	54.3	47.8	16.8	26.9	-0.067
13	142	54.3	47.8	16.8	26.9	-0.066



„Transformation des Symbolischen ins Geometrische“  
(McCormick et al., 1987)

"Grafische Exzellenz ist etwas, das dem Betrachter die größte Anzahl von Ideen in der kürzesten Zeit mit der wenigsten Tinte auf dem kleinsten Raum vermittelt."  
(Edward Tufte)

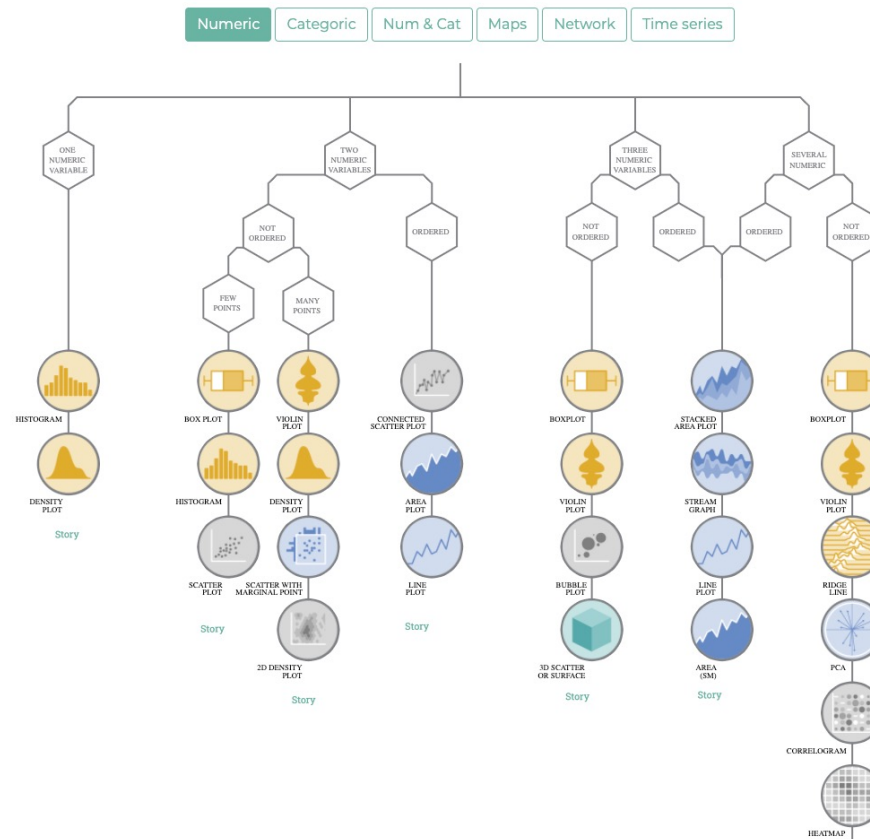
„Die Darstellung von Informationen mit Hilfe von räumlichen oder grafischen Darstellungen, um Vergleiche, Mustererkennung, Erkennung von Veränderungen und andere kognitive Fähigkeiten durch den Einsatz des visuellen Systems zu erleichtern.“

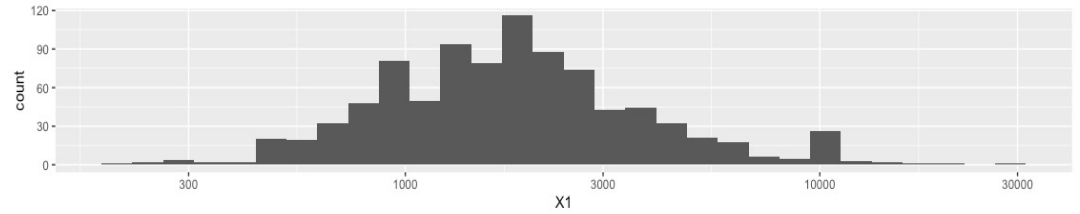
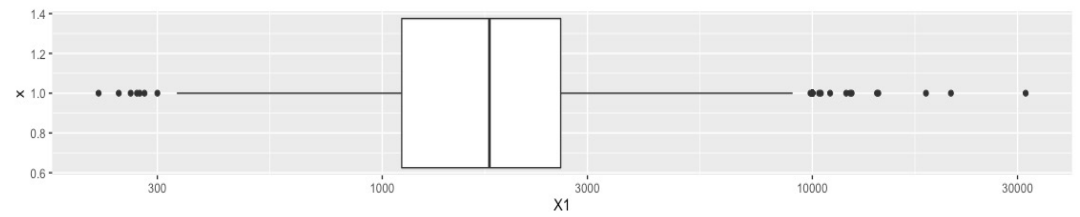
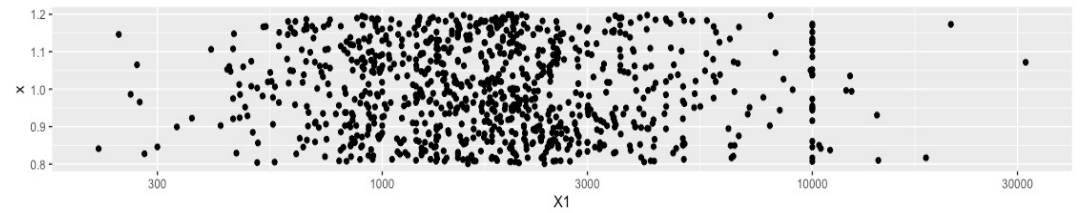
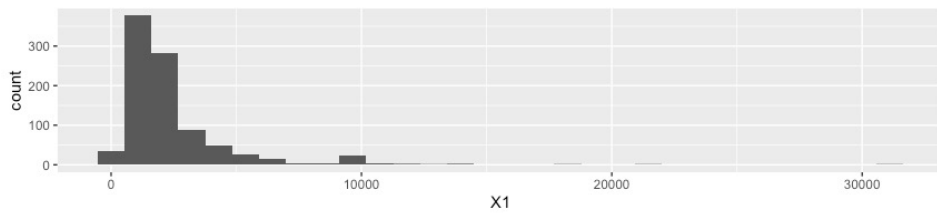
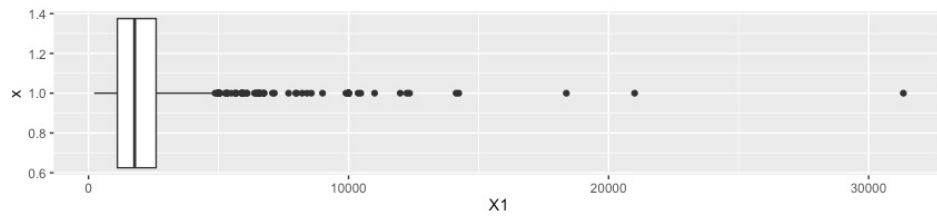
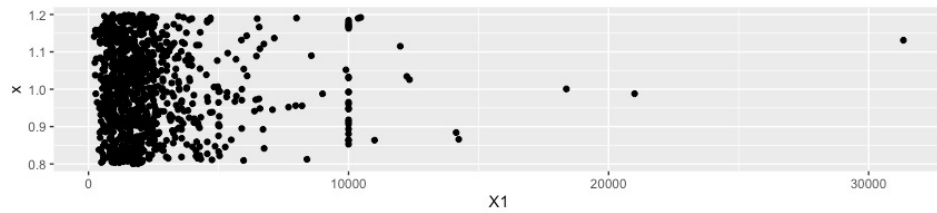
- Erforschen/Berechnen
  - Analysieren
  - Über Informationen nachdenken
  
- Kommunizieren
  - Erläutern
  - Entscheidungen treffen
  - Über Informationen urteilen

# Nützliche Diagramme

- Punktdiagramm
- Jitter-Plot
- Box-Plot
- Histogramm
- Kumulative Verteilungsfunktion

<https://www.data-to-viz.com/#explore>





Quelle: Spiegelhalter, "The Art of Statistics"