

# Einführung

Kai Arzheimer | Vorlesung Forschungsmethoden

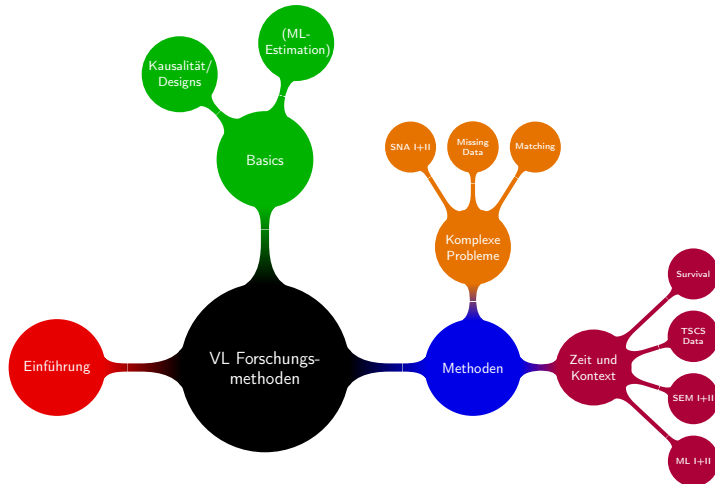
## VL Forschungsmethoden: Für wen?

- ▶ Zielgruppe
  - ▶ **Master Empirische Demokratieforschung**
  - ▶ Interessierte aus anderen Studiengängen
- ▶ Voraussetzungen
  - ▶ Statistik I (baby steps)
  - ▶ Statistik II (Stata, Erweiterungen/Komplikationen  
Regressionsmodell)
- ▶ Erwartungen/Ziele
  - ▶ Aktuelle Analysen aus dem Bereich Demokratieforschung (+ ähnliche Gebiete) verstehen können
  - ▶ Idee, wie eigene empirische Arbeit organisiert/strukturiert werden könnte

## Organisation

- ▶ Master-Vorlesung
  - ▶ *Selbständigkeit und Eigenmotivation*
  - ▶ Verbunden mit Kleingruppe (Dr. Berning)
  - ▶ In beiden Teilen: bei Fragen jederzeit unterbrechen
- ▶ Empfohlene Literatur
  - ▶ King/Keohane/Verba: *Designing Social Inquiry*. Princeton 1994
  - ▶ Brady/Collier: *Rethinking Social Inquiry. Diverse Tools, Shared Standards*. Lanham 2004
  - ▶ Gelman/Hill: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge 2007
  - ▶ Gelegentlich Hinweise zu einzelnen Themen
- ▶ Folien: <https://www.kai-arzheimer.com/vorlesung-forschungsmethoden/>

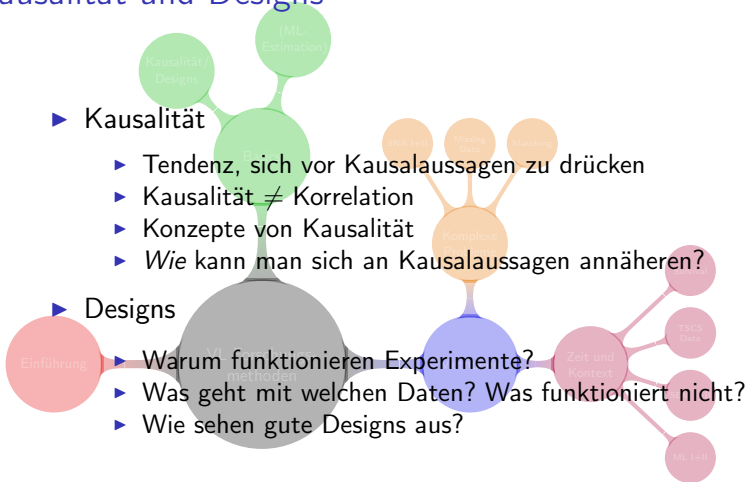
# Übersicht VL Forschungsmethoden



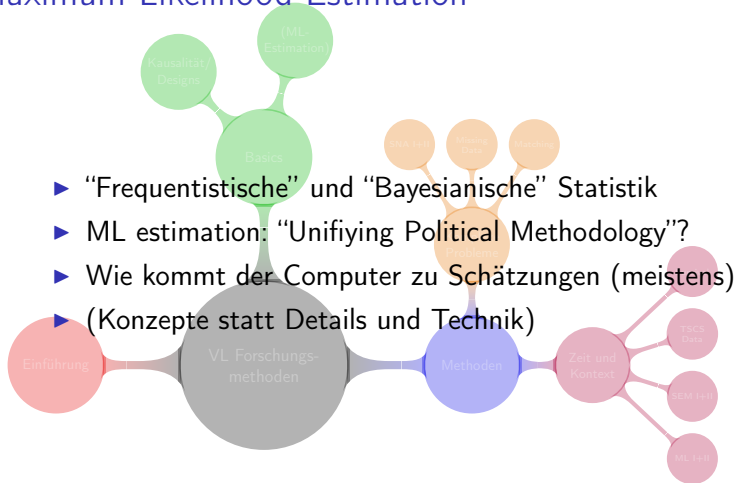
## Plus: zwei Sitzungen am Semesterende

- ▶ Interaktionseffekte
- ▶ Offene Fragestunde zum Semesterende

# Kausalität und Designs



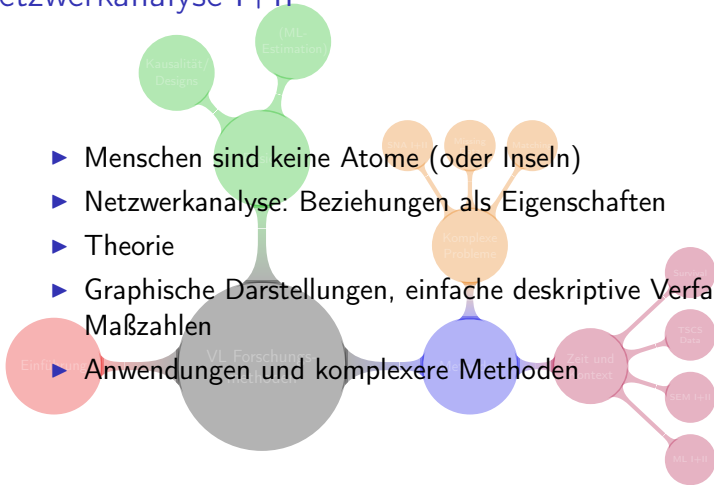
# Maximum Likelihood Estimation



- ▶ “Frequentistische” und “Bayesianische” Statistik
- ▶ ML estimation: “Unifying Political Methodology”?
- ▶ Wie kommt der Computer zu Schätzungen (meistens)
- ▶ (Konzepte statt Details und Technik)

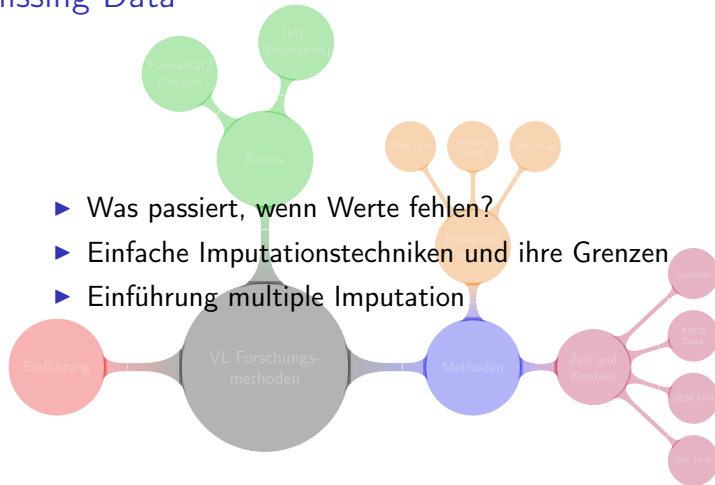
# Netzwerkanalyse I+II

- ▶ Menschen sind keine Atome (oder Inseln)
- ▶ Netzwerkanalyse: Beziehungen als Eigenschaften
- ▶ Theorie
- ▶ Graphische Darstellungen, einfache deskriptive Verfahren und Maßzahlen
- ▶ Anwendungen und komplexere Methoden



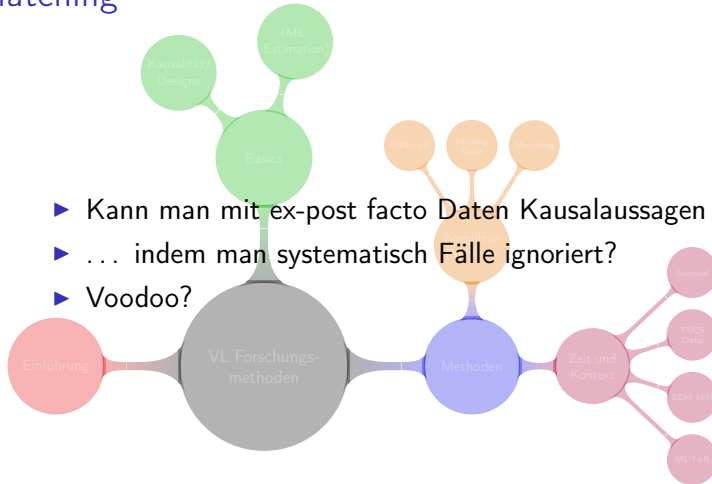


# Missing Data



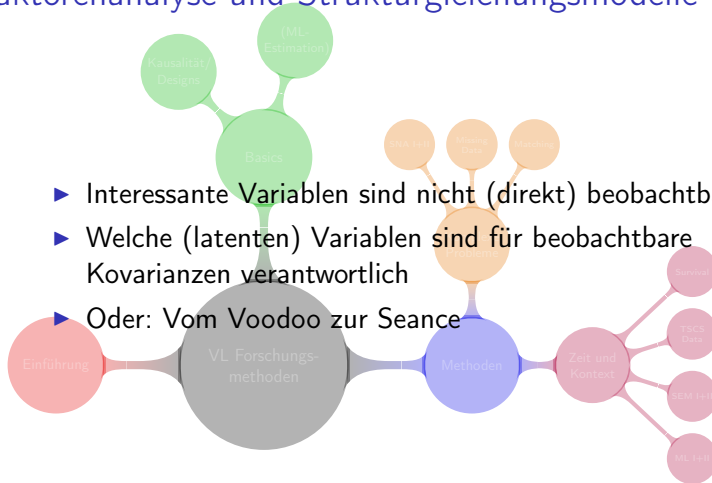
- ▶ Was passiert, wenn Werte fehlen?
- ▶ Einfache Imputationstechniken und ihre Grenzen
- ▶ Einführung multiple Imputation

# Matching



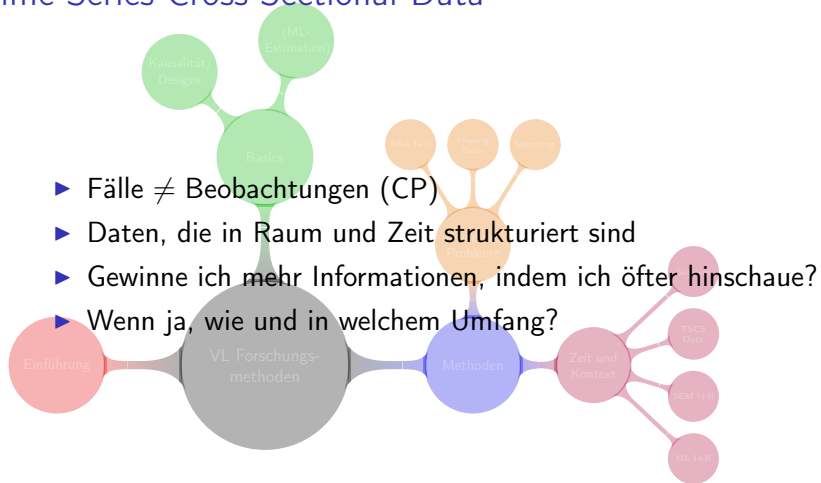
- ▶ Kann man mit ex-post facto Daten Kausalaussagen testen ...
- ▶ ... indem man systematisch Fälle ignoriert?
- ▶ Voodoo?

# Faktorenanalyse und Strukturgleichungsmodelle I+II



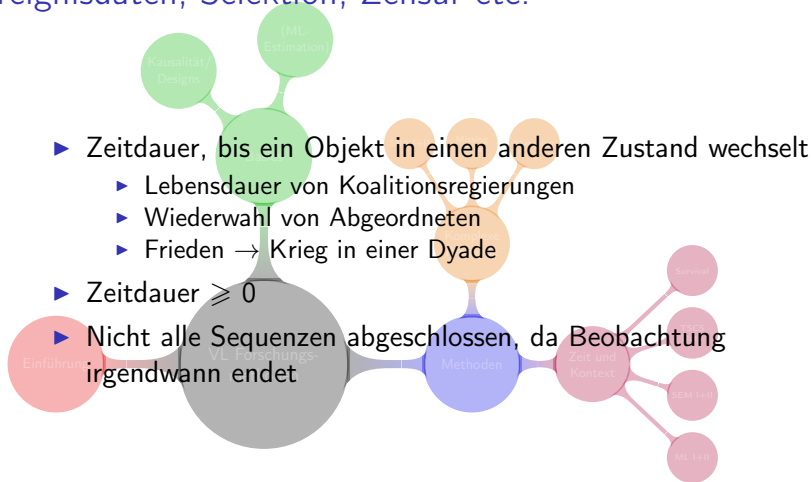
- ▶ Interessante Variablen sind nicht (direkt) beobachtbar
- ▶ Welche (latenten) Variablen sind für beobachtbare Kovarianzen verantwortlich
- ▶ Oder: Vom Voodoo zur Seance

# Time-Series Cross-Sectional Data

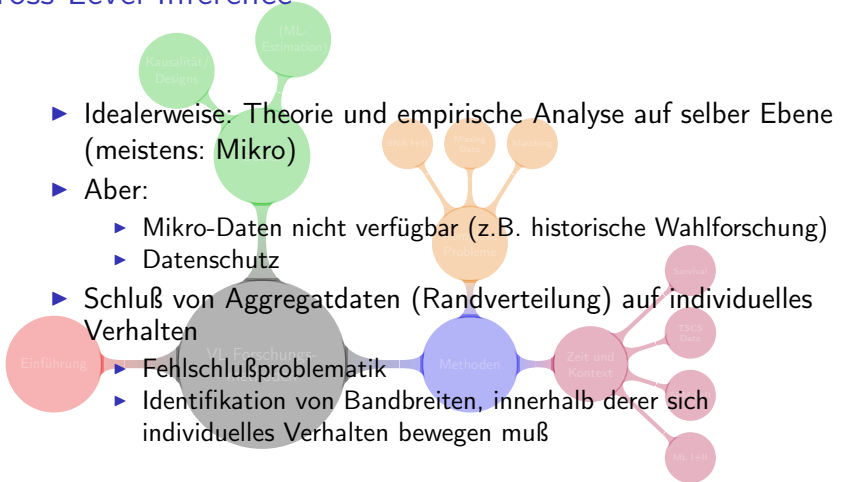


- ▶ Fälle  $\neq$  Beobachtungen (CP)
- ▶ Daten, die in Raum und Zeit strukturiert sind
- ▶ Gewinne ich mehr Informationen, indem ich öfter hinschaue?
- ▶ Wenn ja, wie und in welchem Umfang?

## Ereignisdaten, Selektion, Zensur etc.

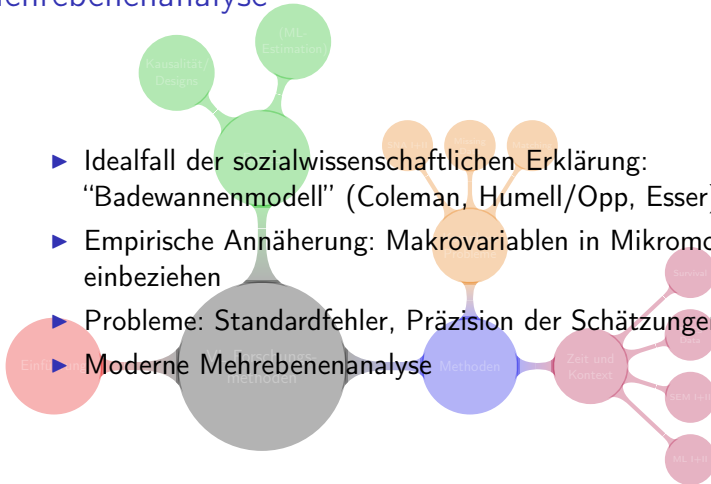


# Cross-Level Inference



# Mehrebenenanalyse

- ▶ Idealfall der sozialwissenschaftlichen Erklärung: “Badewannenmodell” (Coleman, Humell/Opp, Esser)
- ▶ Empirische Annäherung: Makrovariablen in Mikromodell einbeziehen
- ▶ Probleme: Standardfehler, Präzision der Schätzungen
- ▶ Moderne Mehrebenenanalyse



# Was sind Daten?

- ▶ Systematische, quantifizierte Beobachtungen
- ▶ Ergebnis von
  - ▶ “Datengenerierendem Prozeß” (DGP)
    - ▶ Kausale Mechanismen und ihre Konsequenzen
    - ▶ Z.B. Wirtschaftsaufschwung → Sozialisation → individuelle Wertorientierungen → gesellschaftlicher Wandel
    - ▶ Schwachsinniger Name
  - ▶ Unseren Beobachtungen
- ▶ Statistische Analyse
  - ▶ Theoretische Annahmen über DGP
  - ▶ Mathematische Modellierung des DGP
  - ▶ Ableitung prüfbarer Hypothesen
  - ▶ Wie gut passen Daten und Hypothesen zueinander?
  - ▶ Probabilistische Tests



## Welche Probleme gibt es (Auswahl)?

- ▶ Wirklichkeit zu komplex
- ▶ Theorien zu unspezifisch
- ▶ Messungen falsch
- ▶ Tests setzen normalerweise korrekte Modellspezifikation voraus
- ▶ Komplette Validierungsproblematik ...
- ▶ *Komplexe Daten*

## Was bedeutet “komplexe Daten”?

- ▶ Theorien argumentieren auf unterschiedlichen Ebenen
- ▶ Variablen haben nicht nur verschiedene Skalenniveaus
- ▶ Sondern unterscheiden sich in der Logik ihrer Konstruktion

## Wie lassen sich Variablen systematisch organisieren?

Ebene	Mikro		Meso		Makro	...	
Typ	global	→	analytisch				
	relational	→	strukturell				
	kontextuell		←	global	→	analytisch	
				relational	→	strukturell	
				kontextuell	←	global	→
						relational	→
				kontextuell	←		

→: Aggregation

←: Disaggregation

Quelle: Darstellung nach Hox 2002 und Lazarsfeld/Menzel 1961

## Was bedeutet das? I

- ▶ Globale Variablen
  - ▶ Eine Ebene, eine Einheit
  - ▶ Bsp: Alter, Geschlecht, Einstellungen, Wahlsystem, Dezentralisierung ...
- ▶ Analytische Variablen
  - ▶ Zwei Ebenen, mehrere Einheiten
  - ▶ Basieren auf Aggregation (Mittelwerte, Anteile ... ) untergeordneter Einheiten
  - ▶ Z.B. Arbeitslosigkeit

## Was bedeutet das? II

- ▶ Relationale Variablen
  - ▶ Eine Ebene, mehrere Einheiten
  - ▶ Beziehungen *zwischen* Objekten auf einer Ebene
  - ▶ Z.B. Kommunikation zwischen Wählern ("ties": ja/nein; Häufigkeit)
- ▶ Strukturelle Variablen
  - ▶ Zwei Ebenen, mehrere Einheiten
  - ▶ Aggregation relationaler Variablen, z.B. Dichte der ties vgl. mit anderen Einheiten

## Was bedeutet das III: kontextuell

- ▶ Bezieht sich auf (vermutete) Wirkung übergeordneter auf untergeordnete Ebene
- ▶ Kontextuelle Variablen
  - ▶ Zwei Ebenen, mehrere Einheiten
  - ▶ Eigenschaften der übergeordneten Ebene
  - ▶ z.B. alle Wähler demselben Wahlsystem ausgesetzt (global, analytisch, strukturell)

## Zusammenfassung

- ▶ Stand der Forschung erfordert, daß man über das Programm von SM I+II hinausgeht
- ▶ Weites Feld
- ▶ Wichtiger als Detailwissen
  - ▶ Verständnis von Konzepten
  - ▶ Vorstellung davon, was warum (nicht) geht

## Literaturempfehlungen für die nächsten Wochen

- ▶ Kausalität und Designs: King, Keohane & Verba 1994
- ▶ Maximum-Likelihood: King 1998 (Unifying Political Methodology), Eliason 1993 (Maximum Likelihood Estimation: Logic and Practice)
- ▶ Netzwerkanalyse: Ebook von Hanneman/Riddle:  
<http://www.faculty.ucr.edu/~hanneman/nettext/>