

Zwei Wege um Daten zu gewinnen

716408 | Sozialwiss. Methoden – How 2 do Things with Numbers

KMH

SS 22 | KMH (updated: 2022-04-20)



Daten sind Daten - oder?

Daten gewinnen – aber wie?



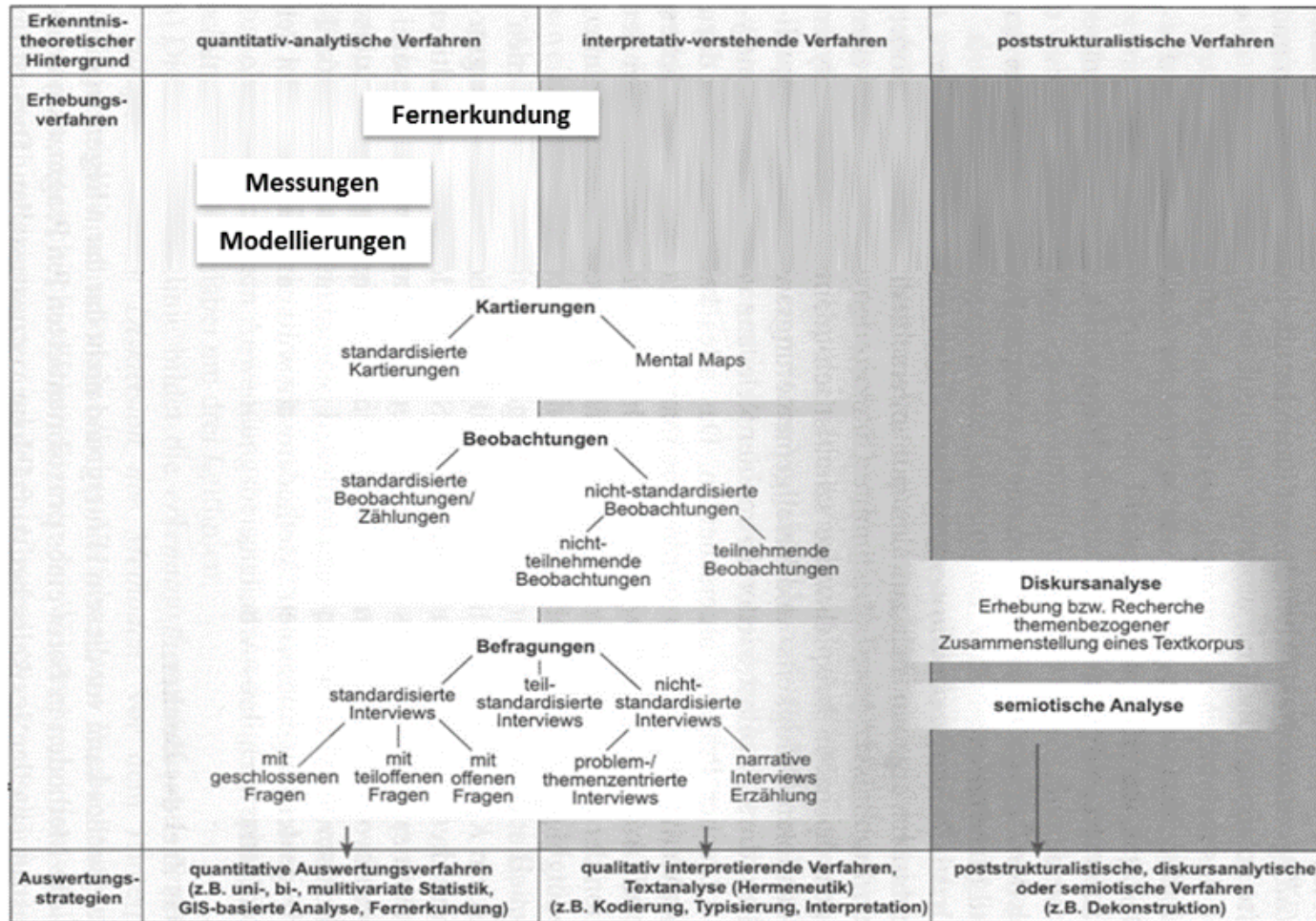
Zwei Wege um Daten zu gewinnen

Messen → Primärdaten	Erschließen → Sekundärdaten
<ul style="list-style-type: none">• Zuvor nicht vorhanden• Durch (eigene) Messung erzeugt• Passgenau zu Untersuchungsfragen• Oftmals aufwändig	<ul style="list-style-type: none">• Aus Primärdaten abgeleitet bzw. ableitbar• Bestenfalls in Repositorien vorgehalten• Durch Klassifikation, Aggregation und/oder Generalisierung erzeugt• Können passgenau zu Untersuchungsfragen sein• Geringerer Aufwand*



Wie entstehen Primärdaten?

Primärerhebung quantitativer Daten



(Überarbeitung von: Reuber & Pfaffenbach 2005:21)

Primärerhebung quantitativer Daten

- **3 zentrale Methoden:**

- **Beobachtung**

systematisches Beobachten & quant. Registrieren „relevanter“ Sachverhalte

- **Befragen**

systematisch gesteuerte Kommunikation zwischen Personen; basierend auf standardisierten Fragen- & Antwortenkatalog

- **Nonreaktive Verfahren**

quant. Erfassen einzelner Merkmale von Interesse in Texten, Bildern & Videos



Wie gewinnt man
sekundärstatistische Daten?

Sekundärstatistische Daten gewinnen

- Einfachster Fall: **Repositorien**
 - Metadaten sichern Auffindbarkeit
 - <https://www.data.gv.at/>
 - https://www.statistik.at/web_de/services/statcube
 - <https://www.govdata.de/>
 - <https://www-genesis.destatis.de/genesis/online>
- Aufwändiger: **Digitalisierung**
 - Konversionsfehler → erhöhter Zeitaufwand zur Qualitätssicherung
 - Erschließen (historischer) Datensets

Und woher kommen Sekundärdaten?

- **Sekundärdaten**

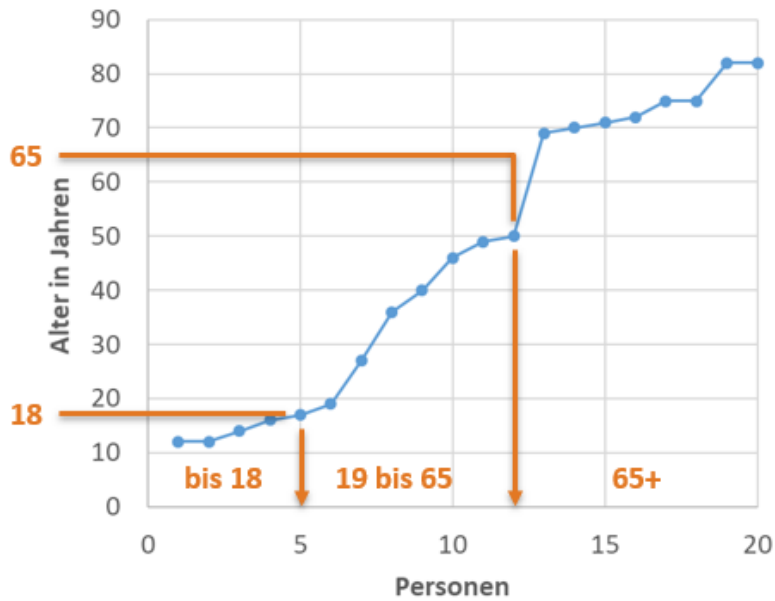
- Meist: durch Verarbeitung aus Primärdaten gewonnen
 - Eigene Erstellung
 - Erstellung durch Dritte

- **Verarbeitung:**

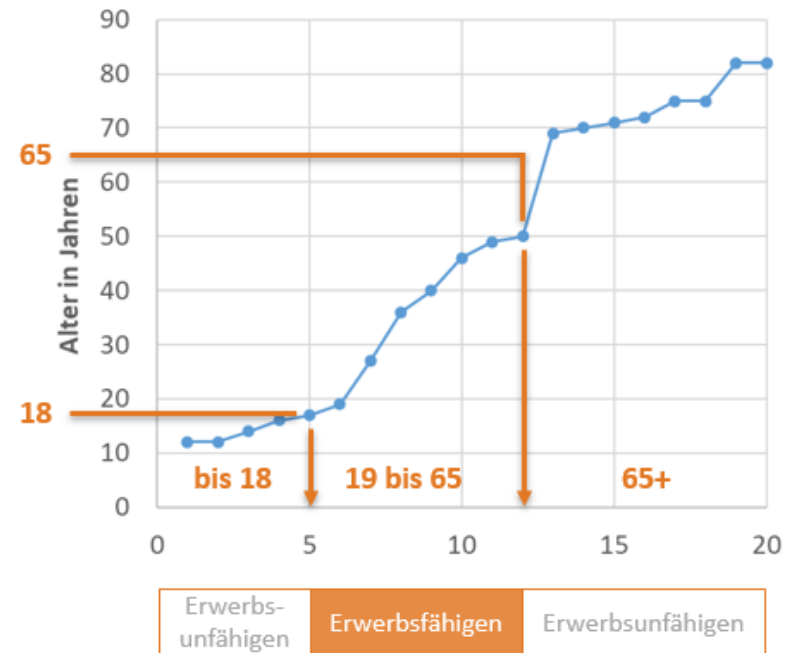
- Klassifikation:
 - Zuordnung zu Klasse von Dingen mit gemeinsamen Eigenschaften
- Aggregation:
 - Gruppierung bestehender Klassen zu neuen (→ Definition einer neuen Klasse)
- Generalisierung:
 - Definition Teilmengen-Beziehung zwischen Elementen verschiedener Klassen

@ Klassifikation, Aggregation & Generalisierung

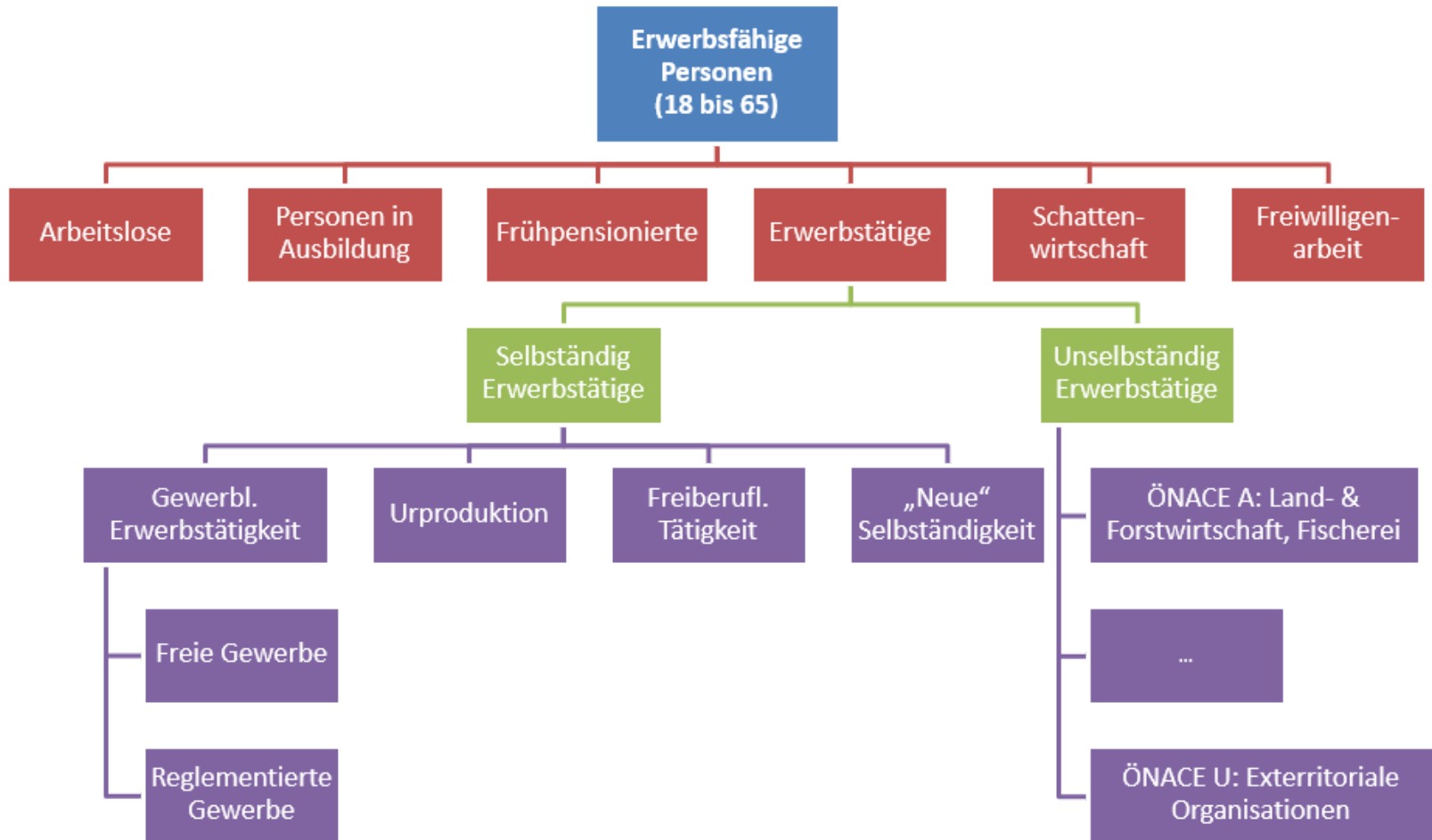
Klassifikation



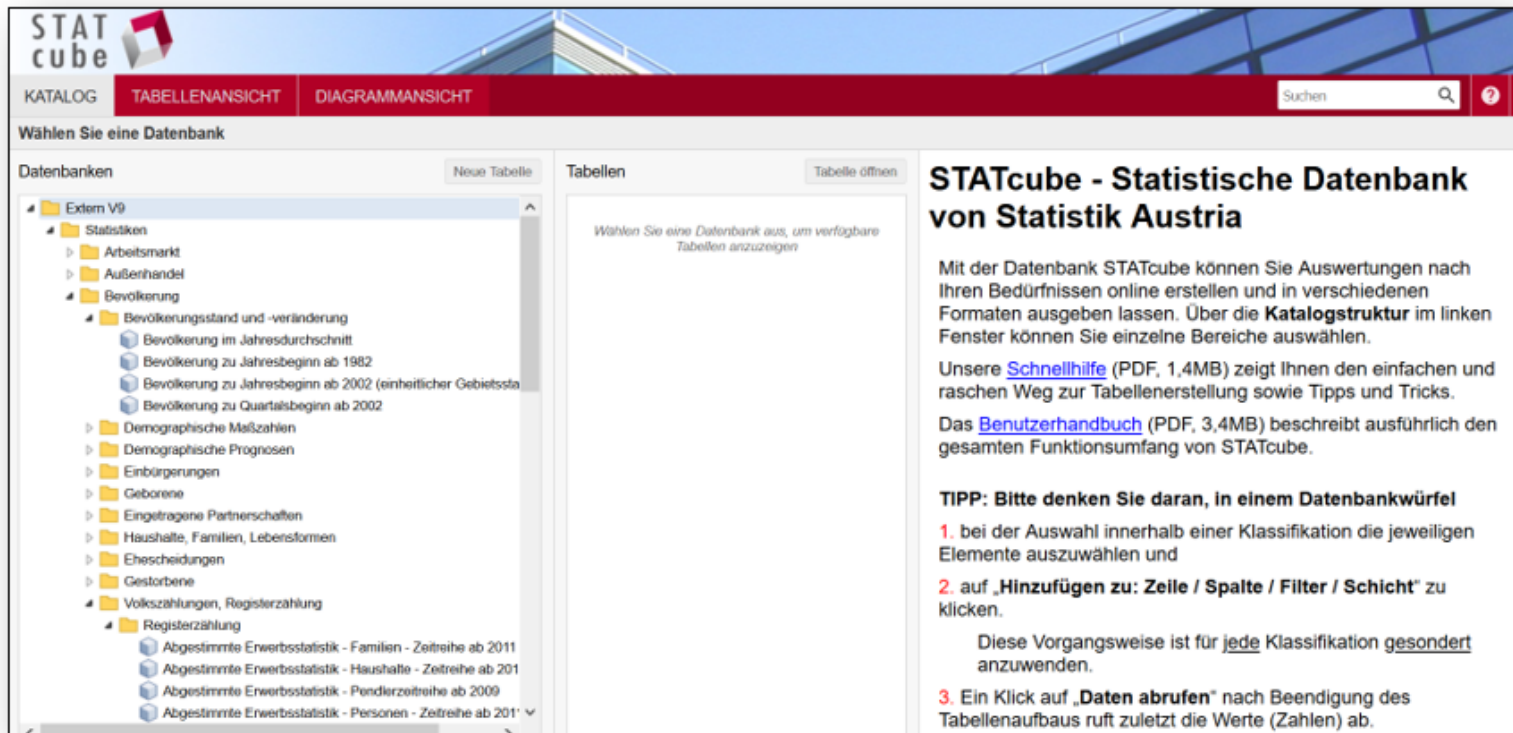
Aggregation



@ Klassifikation, Aggregation & Generalisierung



Woher Sekundärdaten beziehen?



STATcube - Statistische Datenbank von Statistik Austria

Mit der Datenbank STATcube können Sie Auswertungen nach Ihren Bedürfnissen online erstellen und in verschiedenen Formaten ausgeben lassen. Über die **Katalogstruktur** im linken Fenster können Sie einzelne Bereiche auswählen.

Unsere [Schnellhilfe](#) (PDF, 1,4MB) zeigt Ihnen den einfachen und raschen Weg zur Tabellenerstellung sowie Tipps und Tricks.

Das [Benutzerhandbuch](#) (PDF, 3,4MB) beschreibt ausführlich den gesamten Funktionsumfang von STATcube.

TIPP: Bitte denken Sie daran, in einem Datenbankwürfel

1. bei der Auswahl innerhalb einer Klassifikation die jeweiligen Elemente auszuwählen und
2. auf „**Hinzufügen zu: Zeile / Spalte / Filter / Schicht**“ zu klicken.

Diese Vorgangsweise ist für jede Klassifikation gesondert anzuwenden.

3. Ein Klick auf „**Daten abrufen**“ nach Beendigung des Tabellenaufbaus ruft zuletzt die Werte (Zahlen) ab.

<https://statcube.at>

Woher Sekundärdaten beziehen?

The screenshot shows the homepage of the GENESIS-ONLINE database. At the top left is the logo for STATIS (Statistisches Bundesamt). To the right are navigation links for 'Menü', 'Anmeldung', 'Hilfe', and 'English'. A teal banner contains the text 'GENESIS-ONLINE'. Below this, the main heading reads 'Die Datenbank des Statistischen Bundesamtes'. A search bar is labeled 'Datenbank durchsuchen'. Underneath, a list of 'TOP 5 Tabellen' includes: '51000-0007 Aus- und Einfuhr', '61111-0002 Verbraucherpreisindex', '61111-0001 Verbraucherpreisindex Jahre', '61241-0004 Erzeugerpreisindex gewerblicher Produkte', and '12621-0002 Durchschnittliche Lebenserwartung'. The 'Themen' section features six numbered cards: 1. Gebiet, Bevölkerung, Arbeitsmarkt, Wahlen; 2. Bildung, Sozialleistungen, Gesundheit, Recht; 3. Wohnen, Umwelt; 4. Wirtschaftsbereiche; 5. Außenhandel, Unternehmen, Handwerk; 6. Preise, Verdienste, Einkommen und Verbrauch.

<https://www-genesis.destatis.de>

Woher Sekundärdaten beziehen



The screenshot shows the Zenodo website interface. At the top, there is a search bar and navigation links for 'Upload' and 'Communities'. The user profile 'karl-michael.hoefler@uibk.ac.at' is visible. The main content area is titled 'COVID-19 related communities' and features a community for 'Chicago COVID-19 Response'. Below this, there is a section for 'Featured uploads related to COVID-19', which includes a dataset from August 12, 2020, titled 'OpenAIRE Covid-19 publications, datasets, software and projects metadata'. The metadata entry lists numerous authors and provides a description of the dataset's purpose and upload date.

(Planemad, Wikimedia, CC0; GitHub, Wikimedia, CC0; GESIS, Wikimedia, CC0)

Eine Auswahl

- Wikidata: <https://www.wikidata.org>
- gesis: <https://www.gesis.org>
- European Data Portal: <https://data.europa.eu>
- OECD Data: <https://data.oecd.org/>
- UNdata: <https://data.un.org/>
- NASA Socioeconomic Data and Application Center: <https://sedac.ciesin.columbia.edu/>
- Registry of Research Data Repositories: <https://www.re3data.org/>



Was jetzt ... ?

Was jetzt: Primär- oder Sekundärstatische Daten?



(United Nations Photo, CC BY NC ND)

Was jetzt: Primär- oder Sekundärstatische Daten?

Pragmatische Entscheidung:

- Mit welchem **Aufwand** kann ich mir einen Überblick auf das Angebot an Sekundärdaten verschaffen?
- Decken diese Sekundärdaten meine **Untersuchungsfrage & -objekte** ab?
- Sind diese Sekundärdaten für meinen Zweck **nutzbar**?
- Welche **Ressourcen** stehen mir zur Datengewinnung zur Verfügung?
 - Können Sekundärdaten angekauft werden?
- ...

Egal ob Primär- oder Sekundärdaten:

Qualitätskriterien nicht vergessen:**

- **Objektivität (Unabhängigkeit):**
 - Unabhängigkeit eines Messinstruments von Untersucher & Untersuchungssituation
- **Reliabilität (Zuverlässigkeit):**
 - Reproduzierbarkeit von Messergebnissen
- **Validität (Gültigkeit):**
 - Genauigkeit, mit dem ein Test das misst, was er messen soll (= Erkenntnisgegenstand)