

DATA LITERACY ESSENTIALS

HWS 2021/22

Data Literacy Essentials: Crash Course Data Literacy

Dr. Irene Schumm und Dr. Ira Maschmann

[Webseite zu den Data Literacy Essentials](#)

Outline



HWS 2021/22

- Was ist „Data Literacy“?
- Warum ist „Data Literacy“ wichtig?
- Elemente der „Data Literacy“

Was ist Data Literacy?



HWS 2021/22

„Data Literacy ist die Fähigkeit, auf kritische Art und Weise Daten

- zu sammeln,
- zu managen,
- auszuwerten und
- anzuwenden“

Ridsdale et al. (2015)

... und was meinen wir mit „Daten“?

→ Fakten oder Gegebenheiten wie Zahlen, Text, Bild, Audio, Video, die Grundlage einer Auswertung sind bzw. sein können

Qualitative Auswertung

Q2: Was hat Ihnen an Ihrem Studium am besten gefallen?

Sehr gut fand ich die Wahloptionen bei den Kursen sowie die stets ...



Quantitative Auswertung

Jahr	BIP	Arbeitslosenquote
2019	300000	3,1
2020	350000	2,9
2021	...	...



... und was meinen wir mit „Daten“?

→ Fakten oder Gegebenheiten wie Zahlen, Text, Bild, Audio, Video, die Grundlage einer Auswertung sind bzw. sein können

Qualitative Auswertung

Q2: Was hat Ihnen an Ihrem Studium am besten gefallen?

Sehr gut fand ich die Wahloptionen bei den Kursen sowie die stets ...



Quantitative Auswertung

Jahr	BIP	Arbeitslosenquote
2019	300000	3,1
2020	350000	2,9
2021	...	...



Warum ist Data Literacy wichtig?

Datenangebot

„Datafication of life“

Vertrauen in Daten

„In God we trust; all
others must bring data.“

- W. Edwards Deming

→ Data Literacy ist eine essentielle Schlüsselkompetenz

1. Datafication of life: Datenangebot



HWS 2021/22

"It was estimated in 2012 that 90% of the world's data had come into existence within the previous two years (Vessey, et al., 2014)."

Welche Bereiche, in denen „beiläufig“ Daten gesammelt werden, fallen Ihnen ein?

1. Datafication of life: Privatleben

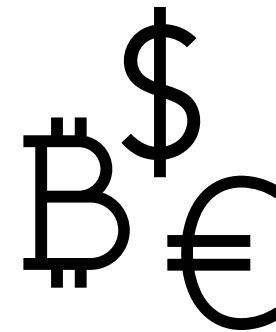
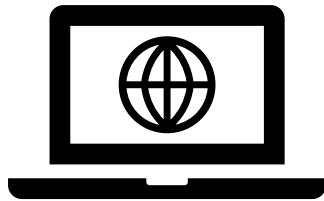


<https://pixabay.com/de/vectors/laufuhr-gps-aktivit%c3%a4t-uhr-armband-1246430/>

1. Datafication of life: Unternehmen



HWS 2021/22



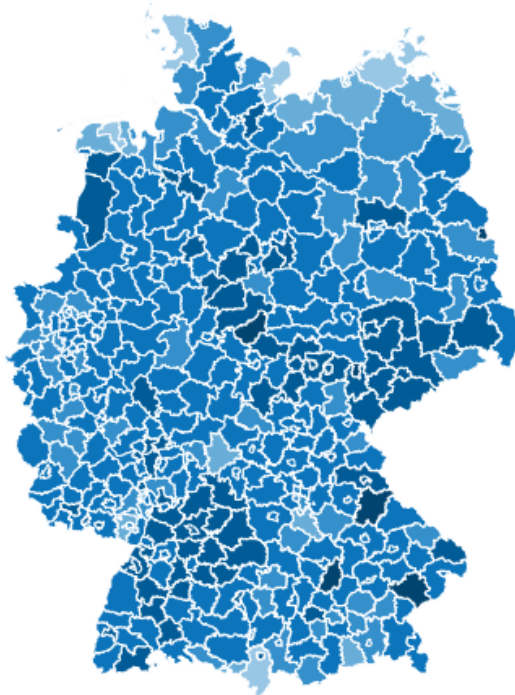
1. Datafication of life: Politik

Abb. 5

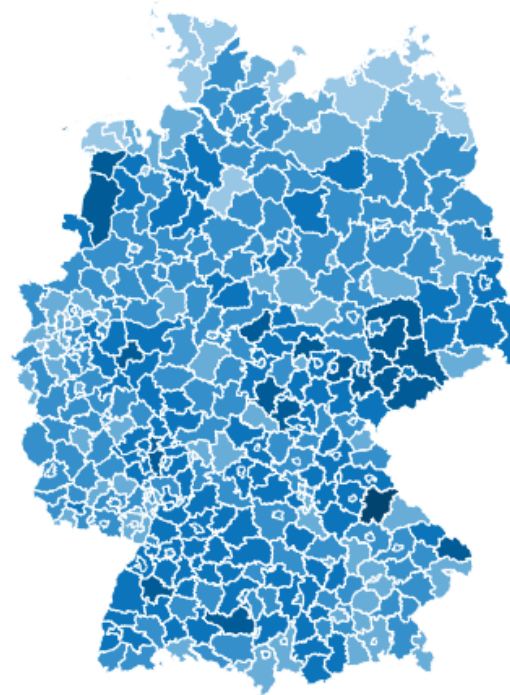
Mobilität im Vergleich

Index KW44/2020 = 100

KW50/2020



KW07/2021



- 100 und mehr
- 95 bis unter 100
- 90 bis unter 95
- 85 bis unter 90
- 80 bis unter 85
- weniger als 80

Quelle: Zachariae (2021); GeoBasis-DE/BKG 2020; Berechnungen des ifo Instituts.

© ifo Institut

2. Vertrauen in Daten: Möglichkeiten

Hypothese: Annahme über einen Zusammenhang

Diese kann **gerichtet** oder **ungerichtet** sein

wird überprüft durch **Daten** und einen **Hypothesentest**

und **abgelehnt** oder **beibehalten** werden

Hypothesen wirken vertrauenswürdiger, wenn sie mit Daten untermauert sind!

Ob ein Wort mit einem Vokal oder einem Konsonanten beginnt, beeinflusst, wie sehr wir es mögen

$K = V$ oder $K > V$

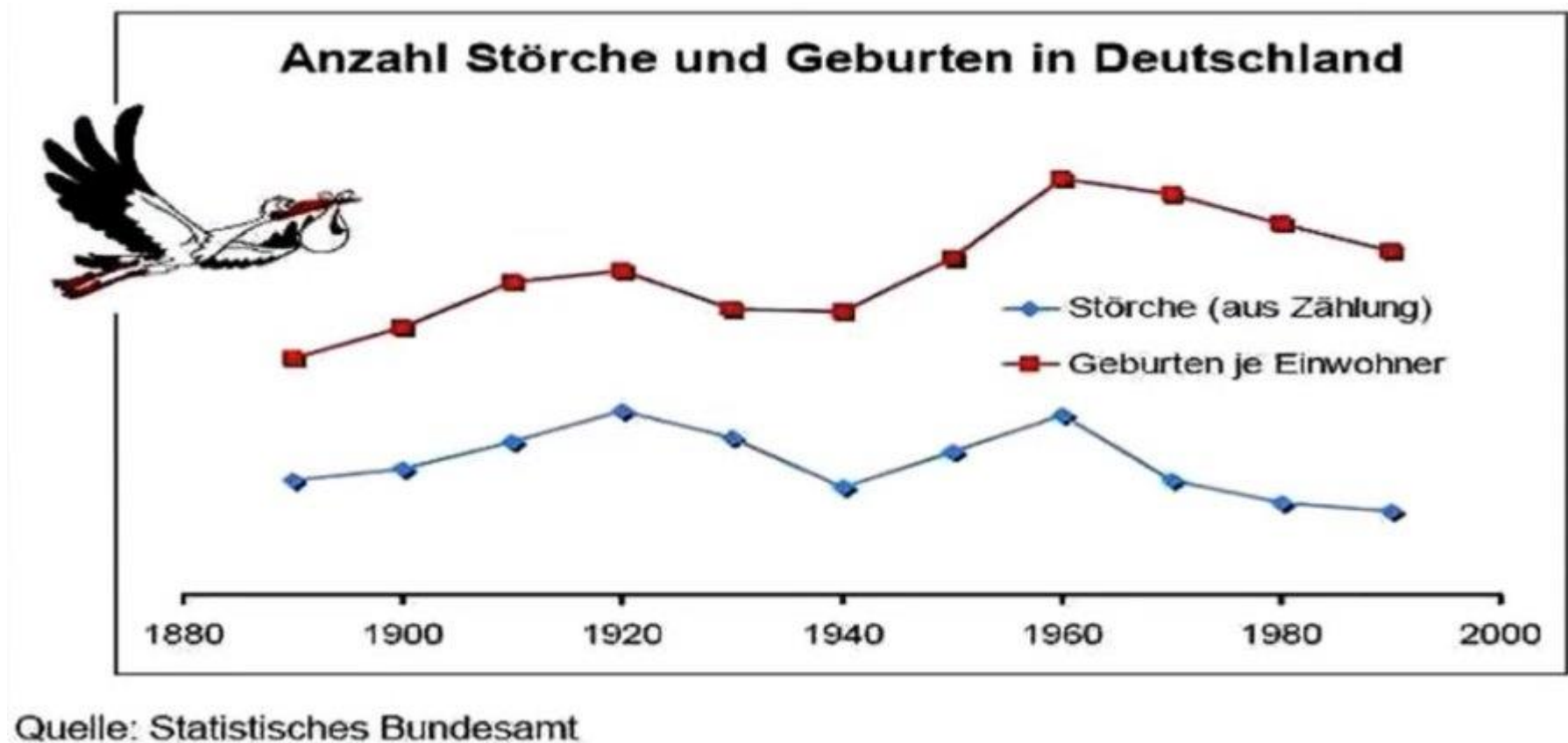
10 Experimente, $N = 924$



$K > V$

(Maschmann, 2019)

2. Vertrauen in Daten: Scheinschlüsse

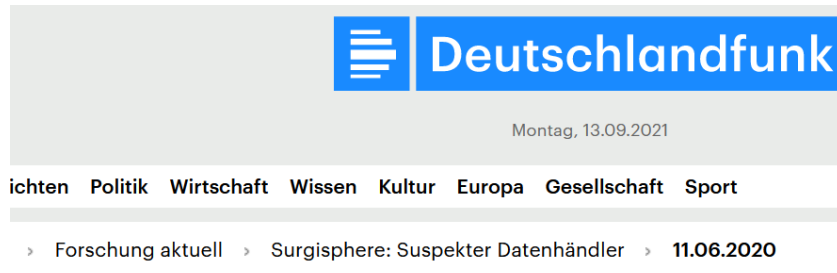


https://twitter.com/sz_associates/status/1123488857985560576?lang=ms

2. Vertrauen in Daten: Manipulation



HWS 2021/22



Forschungsskandal

Surgisphere: Suspekter Datenhändler

Es ist ein ausgewachsener Forschungsskandal: Gleich zwei wichtige Fachjournale haben vergangene Woche jeweils eine viel beachtete COVID-19-Studie zurückgezogen. Die Daten dahinter sind möglicherweise frei erfunden. Sie stammen von der bis vor Kurzem quasi unbekannten US-amerikanischen Firma „Surgisphere“.

„Es fiel mir auf, dass nur vier Autoren für eine derart große Datenmenge verantwortlich gezeichnet haben und dass einer dieser Autoren der Inhaber einer Firma ist, die mir bisher nie etwas gesagt hat: Surgisphere. Das kam mir sehr verdächtig vor.“

Surgisphere-Skandal

17.06.2020, 11:55 Uhr

Wie gefährlich sind gefälschte Daten für die Coronavirus-Forschung?

Zwei Covid-19-Studien wurden wegen windiger Datengrundlage zurückgezogen. Warum hat der „Peer-Review“-Prozess versagt? Und ist er optimierbar? Ein Gastbeitrag. VON ULRICH DIRNAGL

<https://www.tagesspiegel.de/wissen/surgisphere-skandal-wie-gefaehrlich-sind-gefaelschte-daten-fuer-die-coronavirus-forschung/25921646.html>
https://www.deutschlandfunk.de/forschungsskandal-surgisphere-suspekter-datenhaendler.676.de.html?dram:article_id=478441

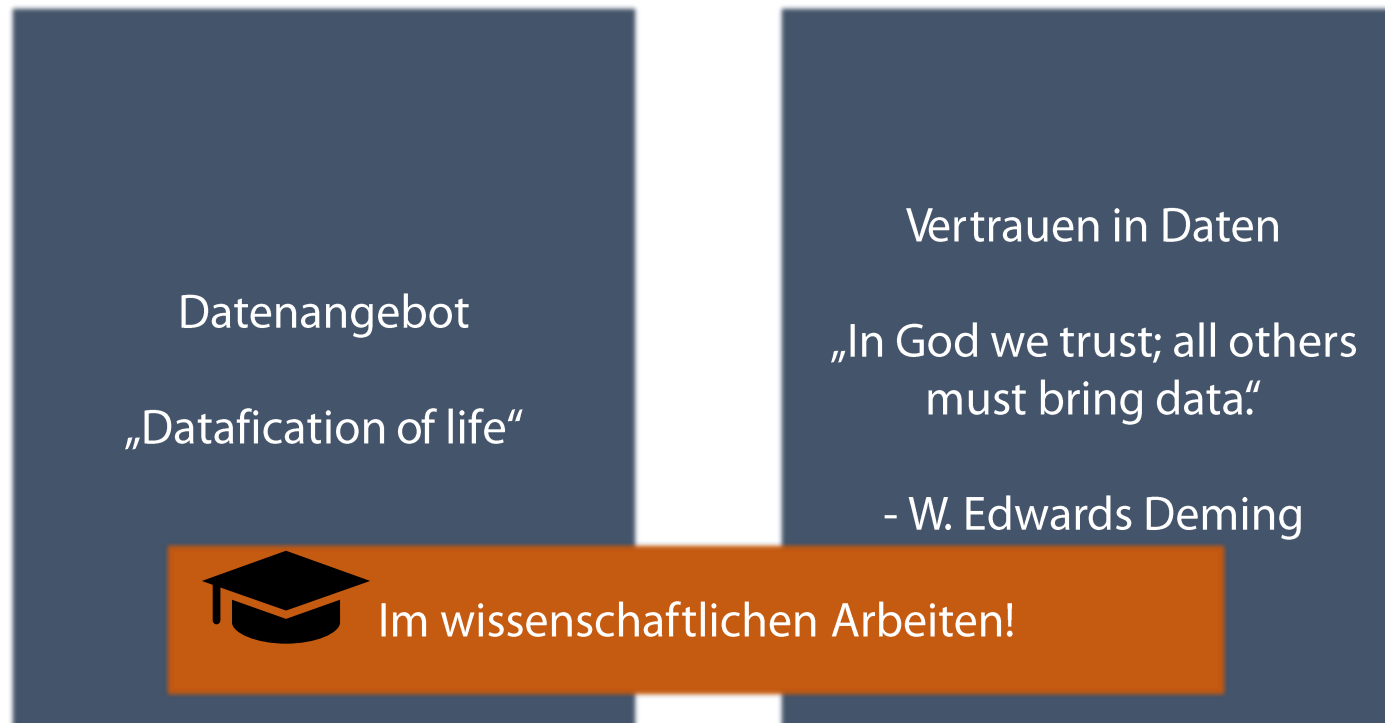
2. Vertrauen in Daten: Verantwortung



HWS 2021/22

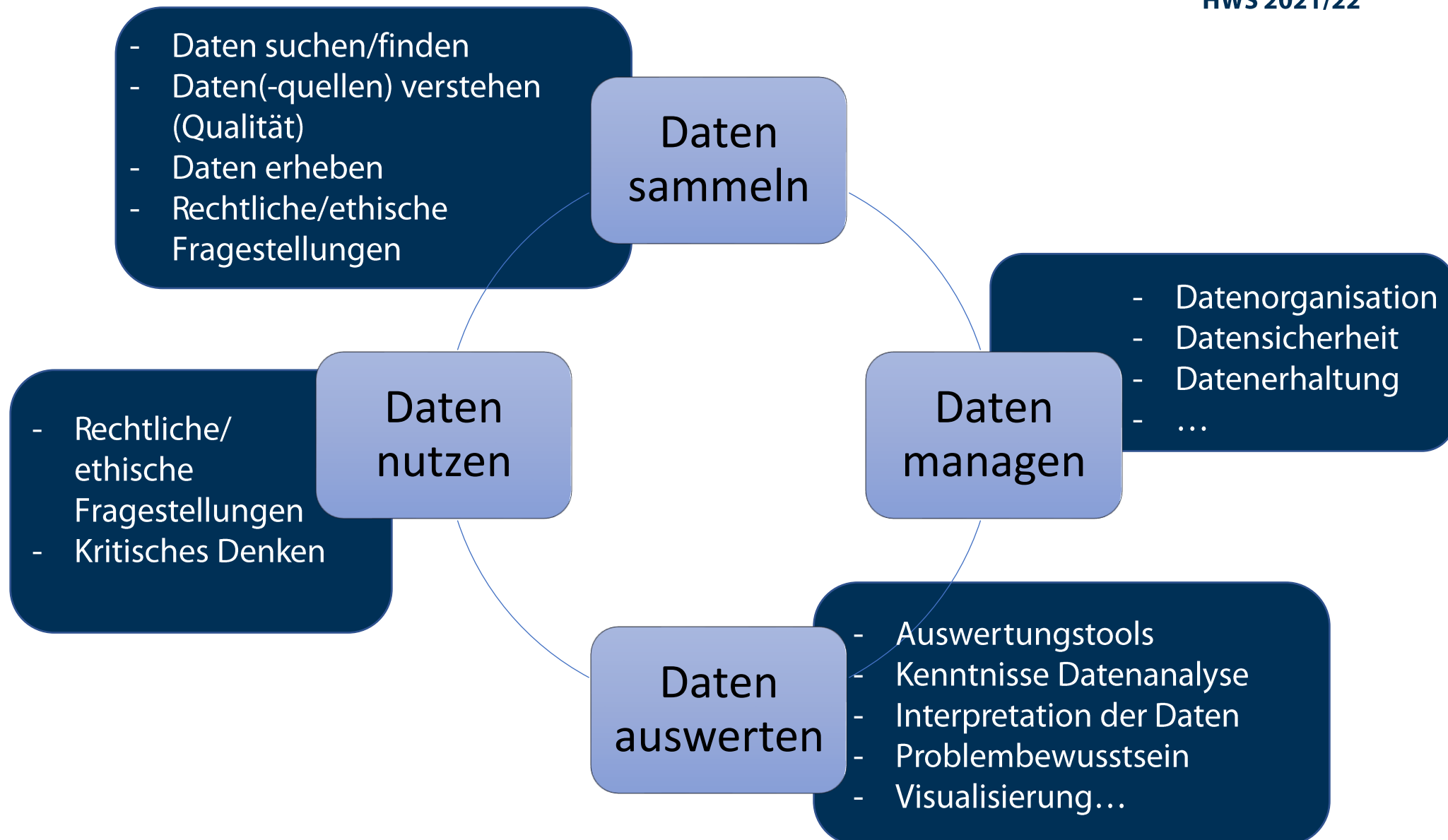
"It is becoming increasingly important for the everyday person to have the **skill to distinguish** between '**good**' and '**bad**' **representations of data** (Twindale, Blake, & Grant, 2013; and Swan, Vahey, Kratcoski, van 't Hooft, Rafanan, & Stanford, 2009). "

Warum ist Data Literacy wichtig?



→ Data Literacy ist eine essentielle Schlüsselkompetenz

Elemente der Data Literacy



- Studentin Lea soll in einem empirischen Praktikum das Thema „Berufseinstieg in Abhängigkeit vom Studienfach“ untersuchen und in Kontext setzen mit aktueller Forschung.
- Neben dem Einführungsgespräch mit ihrem Betreuer, besucht sie zunächst noch einige Kurse in der UB und spricht mit dem Fachreferenten für ihr Fach, um mehr über die Literatur- und Datenzugänge der Universität zu erfahren.



Elemente der Data Literacy

- Daten suchen/finden
- Daten(-quellen) verstehen (Qualität)
- Daten erheben
- Rechtliche/ethische Fragestellungen

Daten sammeln

- Lea hat im DLE-Kurs „Daten recherchieren“ gelernt, wie und wo sie nach Forschungsdaten suchen kann. Zunächst sucht sie nach passenden Datenbanken auf der Webseite [re3data](https://re3data.org/). Dank des DLE-Kurses „Daten verstehen“ kann sie die Aussagekraft gefundener Daten einordnen.
- Durch den DLE-Kurs „Daten erheben“ weiß sie, welche Möglichkeiten der Datenerhebung es gibt und sie kann entscheiden, welche sich für ihre Frage eignen. Sie möchte Daten von LinkedIn erheben.
- Dank des DLE-Kurses „Datenschutz und Urheberrecht bei Forschungsdaten“ kann sie einschätzen, ob sie die Daten einfach so erheben und verwenden darf.

- Rechtliche/
ethische
Fragestellungen

rganisation
herheit
altung

Elemente der Data Literacy



HWS 2021/22

- Daten suchen/finden
- Daten(-quellen) verstehen (Qualität)

Daten

- Leas Studienfreundin sitzt bereits seit einiger Zeit an ihrer Bachelorarbeit. Bei einem Festplattencrash sind Leas Freundin 2 Wochen vor Abgabe der Arbeit wesentliche Dateien abhanden gekommen. In dem DLE-Kurs „Daten organisieren“ hat sie verschiedene Strategien und Möglichkeiten kennengelernt, wie sie ihre Daten systematisch organisieren und sichern kann.

Daten
managen

- Datenorganisation
- Datensicherheit
- Datenerhaltung
- ...

Auswertungstools
Kenntnisse Datenanalyse
Interpretation der Daten
Problembewusstsein
- Visualisierung...

Elemente der Data Literacy

- Daten suchen/finden

- Lea hat in dem DLE-Kurs „Daten analysieren“ zwei Programme kennengelernt: R und Stata. Sie weiß nun, welches Programm zu ihren Anforderungen passt und kann loslegen.
- Aus dem DLE-Kurs „Daten visualisieren“ hat sie mitgenommen, wie sie den Inhalt ihrer Forschungsarbeit sinnvoll darstellen und vermitteln kann.

- Datenorganisation
- Datensicherheit
- Datenerhaltung
- ...

Daten auswerten

- Auswertungstools
- Kenntnisse Datenanalyse
- Interpretation der Daten
- Problembewusstsein
- Visualisierung...

Elemente der Data Literacy

- Daten suchen/finden
- Daten(-quellen) verstehen (Qualität)
- Daten organisieren
- Daten sichern/ethische Fragestellungen



Daten sammeln

Die bisherigen Erkenntnisse sollen über die Erstellung eines multivariaten Gesamtmodells getestet werden. In diesem können – im Gegensatz zu den bivariaten Untersuchungen – alle Interaktionseffekte der einzelnen Variablen Berücksichtigung finden. Dafür wird die Gesamtheit der bisher aufgeführten Variablen im Rahmen einer binär logistischen Regression in einem Schritt in dieses Modell integriert. Insgesamt konnten 1.888 Fälle einbezogen werden.

Für die Beschreibung der Modellgüte wird zuerst das Ergebnis des Omnibus-Tests der Modellkoeffizienten berichtet. Mit einem $p < .05$ zeigt sich, dass die Variablen einen signifikanten Zuwachs bei der Modellanpassung liefern. Das dargestellte Modell hat eine höhere Erklärungskraft als das „Null-Modell“, welches nur die Konstante enthält. Der Hosmer-Lemeshow-Test bestätigt ebenfalls eine vorhandene Modellpassung an die Daten, da er kein signifikantes Ergebnis liefert. Für die einzelnen Variablen im Modell werden – sofern signifikant – die Exp(B), d.h. die entlogarithmierten logit-Koeffizienten als Odds ratios wiedergegeben.⁵ Insgesamt zeigen sich vier signifikante Variablen.

Der erste Effekt ist auf den Hochschultyp bezogen: Haben die ehemaligen Studierenden einen Studiengang an einer Universität abgeschlossen, verringert sich das Chancenverhältnis für die Teilnahme an einer berufsbegleitenden wissenschaftlichen Weiterbildung (Exp(B) = .509, siehe Tabelle 3). Absolventinnen und Absolventen der ehemaligen Fachhochschulen profitieren auf dem Arbeitsmarkt vom starken Praxisbezug ihrer Hochschulausbildung und können schneller in das Berufsleben einsteigen. Ebenso nehmen sie seltener ein weiteres Studium oder eine Promotion auf. Entsprechend scheint diese Gruppe stärker im Berufsleben etabliert zu sein und nutzt die berufsbegleitende wissenschaftliche Weiterbildung als Karriereoption. Ein Einfluss der Einhaltung der Regelstudienzeit lässt sich auch im multivariaten Modell nicht nachweisen. Allerdings zeigt sich – abweichend zu den bivariaten Betrachtungen – als zweiter Effekt ein Einfluss der Fächergruppe: Haben die Absolventinnen und Absolventen ein Fach der Sprach- und Kulturwissenschaften studiert, erhöht dies das Chancenverhältnis der Teilnahme an wissenschaftlicher Weiterbildung.

Zudem zeigen sich auch im multivariaten Modell zwei Effekte der Lebensziele der Absolventinnen und Absolventen. Ist die Karriere- und/oder die Forschungsorientierung bei den Befragten nicht stark ausgeprägt, so sinkt ebenfalls das Chancenverhältnis (Karriereorientierung Exp(B) = .566, Forschungsorientierung Exp(B) = .736), dass die Absolventinnen und Absolventen an einer längerfristigen wissenschaftlichen Weiterbildung neben dem Beruf teilnehmen.

⁵Ein Wert von Eins würde dabei bedeuten, dass durch die Variable keine Veränderung erzeugt wird und somit kein Einfluss auf die Teilnahme an wissenschaftlicher Weiterbildung vorliegt. Höhere Werte beschreiben entsprechend ein erhöhtes Chancenverhältnis für die Teilnahme an wissenschaftlicher Weiterbildung, während kleinere Werte ein verringertes Chancenverhältnis darstellen.

INSGESAMT KÖNNEN 1.888 FÄLLE EINBEZOGEN WERDEN. DIE BISHERIGEN ERKENNTNISSE SOLLEN ÜBER DIE ERSTELLUNG EINES MULTIVARIATEN GESAMTMODELLS GETESTET WERDEN. IN DIESEM KÖNNEN – IM GEGENSATZ ZU DEN BIVARIATEN UNTERSUCHUNGEN – ALLE INTERAKTIONSEFFEKTE DER EINZELNEN VARIABLEN BERÜCKSICHTIGUNG FINDEN. DA FÜR WIRD DIE GESAMTHEIT DER BISHER AUFGEFÜHRTEN VARIABLEN IM RAHMEN EINER BINÄR LOGISTISCHEN REGRESSION IN EINEM SCHRITT IN DIESES MODELL INTEGRIERT. INSGESAMT KONNTEN 1.888 FÄLLE EINBEZOGEN WERDEN.

Tabelle 3: Logistische Regression: Teilnahme an wissenschaftlicher Weiterbildung (n=1.888)

Unabhängige Variablen	Sig.	Exp(B)
Abschluss^a (Ref.: Staatsexamen Medizin)		
Master	,997	,000
Magister	,323	3,715
Diplom	,719	1,599
Staatsexamen Lehramt	,918	,969
Hochschultyp^a (Ref.: Hochschule für angewandte Wissenschaften)		
Sonstige ^b	,010	,509
Fächergruppe^a (Ref.: Sonstige ^b)		
Ingenieurwissenschaften	,260	2,061
Rechts-, Wirtschafts- und Sozialwissenschaften	,126	3,211
Sprach- und Kulturwissenschaften ^a	,050	4,394
Mathematik, Naturwissenschaften	,321	2,228
Medizin	,320	4,349
Abschlussnote	,868	1,037
Regelstudienzeit (Ref.: eingehalten)	,171	,716
Elternschaft (Ref.: Kinder vorhanden)	,066	1,599
Geschlecht (Ref.: Männer)	,847	,953
Alter	,982	1,001
Partnerschaft (Ref.: Partnerschaft gemeinsamer Haushalt)		
Ohne Partner/in	,094	,587
Partnerschaft getrennter Haushalt	,538	1,206

Fortsetzung von Tabelle 3 auf der nächsten Seite

^a - Sig. p < 0,05
Nagelkerke R-Quadrat: 0,092

^bDie Absolventinnen und Absolventen der juristischen Abschlüsse sind in diesem Modell nicht integriert, da sie aufgrund des eigenen Punktesystems bei der Bewertung des Studienabschlusses keine adäquaten Werte für die Variable „Abschlussnote“ aufweisen.

^cAufgrund von geringen Fallzahlen wurden die Fächergruppen der Agrar-, Forst- und Ernährungswissenschaften, der Sport- sowie Kunstwissenschaften zusammengefasst.

Datenorganisation Datensicherheit Datenerhaltung

- Rechtliche/ethische Fragestellungen

Webseite zu den Data Literacy Essentials

This course was partially developed and held by the
BERD – Business, Economic and Related Data initiative.

BERD@BW

funded by



Baden-Württemberg

MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST



BERD@NFDI funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 460037581