

Transkript zum Video Frequenztabellen und Histogramme

aus der Reihe Daten Lesen Lernen ([DaLeLe4all Homepage](#))

Inhalt

Folie 1 – Frequenztabellen und Histogramme	1
Folie 2 – Themen	2
Folie 3 – Lernziele	2
Folie 4 – Thema 1	2
Folie 5 – Warum Frequenztabellen und Histogramme nutzen?	2
Folie 6 – Warum Frequenztabellen und Histogramme nutzen 2	3
Folie 7 – Warum Frequenztabellen und Histogramme nutzen 3	4
Folie 8 – Thema 2	4
Folie 9 – Von Rohdaten zur Frequenztafel	4
Folie 10 – Frequenztafel	5
Folie 11 – Thema 3	5
Folie 12 – Warum Histogramme nutzen?.....	5
Folie 13 – Aufbau von Histogrammen	6
Folie 14 – Aufbau von Histogrammen 2	7
Folie 15 – Aufbau von Histogrammen 3	7
Folie 16 – Zusammenfassung	8
Folie 17 – Vielen Dank für die Aufmerksamkeit	9
Folie 18 – Anhang: Quellen- und Abbildungsverzeichnis	9
Folie 19 – Anhang: Förderung und Copyright	9

Hinweis zur Schreibweise

Im Folgenden werden (sofern vorhanden) hochgestellte Zahlen oder Buchstaben durch \wedge ($A^2 = A^2$) und tiefgestellte Zahlen oder Buchstaben durch $_$ ($a_j = a_j$) markiert.

Folie 1 – Frequenztabellen und Histogramme

Folientext

Datenlesen Lernen: Häufigkeitsstrukturen verstehen – Frequenztabellen und Histogramme. Von Dr. Alexander Silbersdorff, Campus-Institut Data Science, Logo der Georg-August-Universität Göttingen.

Sprechtext

Herzlich willkommen zu diesem Video der Veranstaltungsreihe Daten Lesen Lernen. In diesem Video behandle ich zwei wichtige Darstellungsformen von Häufigkeitsstrukturen in Daten – die Frequenztafel und das Histogramm.

Folie 2 – Themen

Folientext

1. Warum Frequenztabellen und Histogramme nutzen?
2. Frequenztabellen
3. Histogramme

Sprechttext

Dafür werde ich zuerst skizzieren, warum es sinnvoll sein kann, die Komplexität von Daten mithilfe von Frequenztabellen und Histogrammen zu reduzieren. Anschließend werde ich erläutern, wie man Frequenztabellen und Histogramme erstellen und interpretieren kann.

Folie 3 – Lernziele

Folientext

Nach diesem Video können Sie...

- die Bedeutung von visuellen Darstellungen bei Daten erkennen.
- Frequenztabellen für die Histogrammerstellung aufbereiten.
- Histogramme interpretieren.

Sprechttext

Mithilfe dieser Vorlesung werden Sie insbesondere lernen, warum tabellarische Darstellungen und Visualisierungen einen prägnanten Eindruck von Häufigkeitsstrukturen in Daten vermitteln. Des Weiteren werden Sie in der Lage sein, für solche Daten Frequenztabellen zu erstellen und diese für die Erstellung von Histogrammen aufzubereiten. Schlussendlich lernen Sie, Histogrammdarstellungen von Daten zu interpretieren.

Folie 4 – Thema 1

Folientext

1. Warum Frequenztabellen und Histogramme nutzen

Sprechttext

Beginnen wir also mit der Frage, warum tabellarische Darstellungsformen und Visualisierungen im Allgemeinen und Frequenztabellen sowie Histogramme im Besonderen genutzt werden.

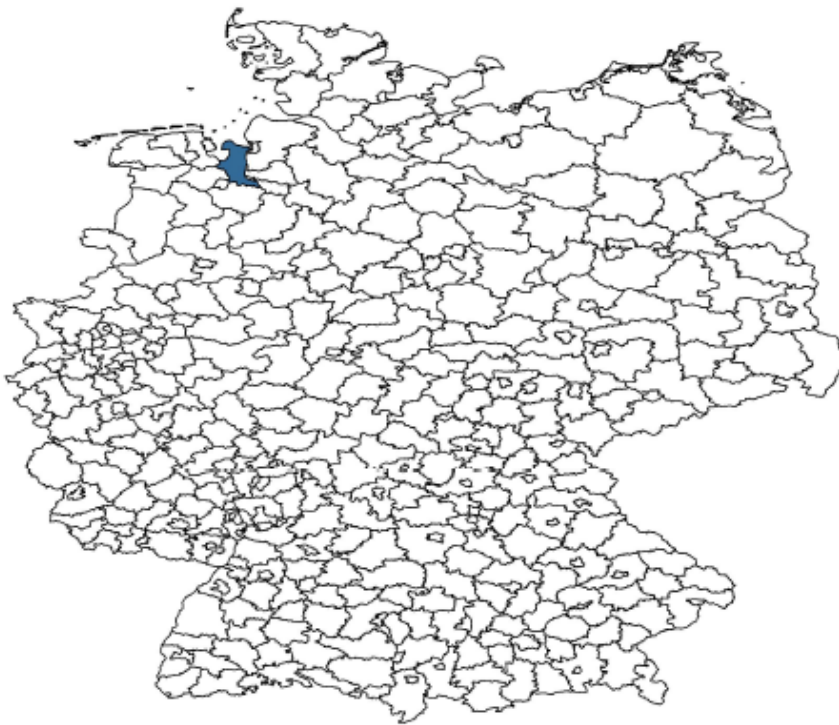
Folie 5 – Warum Frequenztabellen und Histogramme nutzen?

Folientext

Beispieldatensatz

- Landwirtschaftlich nutzbare Flächen im Landkreis Wesermarsch
- Gesamtackerfläche in Hektar (im Folgenden ha)
- Zufällig ausgewählte 50 Flächen
- Abbildung: Landkarte der Gebietseinheiten Deutschlands, Hervorhebung des Landkreises Wesermarsch

Landkreis Wesermarsch



Sprechttext

Bevor ich inhaltlich einsteige möchte ich kurz den Beispieldatensatz skizzieren, mit welchem ich die Anwendung von Frequenztabellen und Histogrammen illustrieren werde. Die Daten stammen aus einem Datensatz über landwirtschaftlich nutzbare Flächen im schönen Landkreis Wesermarsch an der deutschen Nordseeküste. Für diese Flächen besteht Anspruch auf Förderungen im Rahmen der gemeinsamen Agrarpolitik der EU und somit ist die Analyse dieser Flächendaten an verschiedener Stelle von Interesse. Aus diesem Datensatz wurden für die Vorlesung 50 Datenpunkte zufällig ausgewählt, welche die Hektar landwirtschaftlich nutzbarer Fläche wiedergeben und diese Daten werden wir uns nun im weiteren Verlauf genauer anschauen.

Folie 6 – Warum Frequenztabellen und Histogramme nutzen 2

Folientext

- 50 Datenpunkte als Rohdaten
- Abbildung: Tabelle mit 50 Rohdaten

6.88	72.99	80.73	11.31	14.35
2.32	49.76	102.43	9.92	38.54
18.46	6.48	30.89	75.99	9.97
2.14	23.38	29.44	41.16	54.57
37.59	49.94	85.11	74.6	10.01
36.85	2.44	43.51	7.17	70.43
78.29	39.4	8.72	5.21	7.78
40.88	3.91	32.19	60.76	183.26
36.91	22.68	20.55	6.59	42.84
102.97	34.7	9.31	60.45	44.52

Sprechtext

Zum Einstieg sind diese 50 Werte nun in Rohform auf dieser Folie abgebildet. Und wenngleich diese 50 Werte nur ein Teil des kompletten Datensatzes sind, so vermute ich, dass es für einen Großteil von Ihnen problematisch ist, sich anhand dieser 50 Werte direkt ein Bild zu machen wie die Größenordnungen dieser Flächen verteilt sind. Sollte ich Sie also bspw. fragen, was denn typische Größen für landwirtschaftliche Flächen in der Wesermarsch sind, vermute ich, dass nur die wenigsten von Ihnen dies einfach so auf Grundlage der Rohdaten beantworten könnten.

Folie 7 – Warum Frequenztabellen und Histogramme nutzen 3

Folientext

- Fähigkeiten zur Erfassung großer Datenmengen beschränkt
- Zahlenmengen in der Regel komplex
- Reduktion der Komplexität notwendig

Sprechtext

Denn für die meisten von uns gilt zu konstatieren, dass wir schlichtweg nicht die kognitiven Kapazitäten dafür haben mit einem Zahlenwust von 50 Zahlen oder mehr gut umgehen zu können. Manche von Ihnen mögen das aufgrund überdurchschnittlicher Fähigkeiten in diesem Bereich noch hinkriegen, aber spätestens für größere Datenmengen gehen auch die kognitiv fähigsten Zahlenvirtuosen unter uns früher oder später in die Knie. Und gerade in der heutigen Zeit sind große Datensätze mit tausenden von Zahlen oder mehr Usus. Folglich ist es in der Regel notwendig, die Daten so aufzubereiten, dass wir einen Überblick über die Daten und die darin erhaltenen Strukturen bekommen und somit ein Verständnis der Empirie entwickeln, welches wir anschließend nutzen können.

Folie 8 – Thema 2

Folientext

2. Frequenztabellen

Sprechtext

Ein erster Schritt, um dies zu tun, ist es eine Frequenztafel zu nutzen.

Folie 9 – Von Rohdaten zur Frequenztafel

Folientext

- Vereinfachung durch Zusammenfassung in Frequenztabellen
- Abbildung: Tabelle der Rohdaten (siehe [Folie 6](#))
- Abbildung: Frequenztafel

Fläche in ha	Absolute Häufigkeit
$0 \leq x \leq 10$	14
$10 \leq x \leq 20$	4
$20 \leq x \leq 30$	4
$30 \leq x \leq 50$	15
$50 \leq x \leq 75$	6
$75 \leq x \leq 100$	4
$100 \leq x \leq 125$	2
$125 \leq x \leq 150$	0
$150 \leq x \leq 200$	1
Summe	50

Sprechtext

Der Grundgedanke von Frequenztabellen ist es, die Daten zu gruppieren. Diese Gruppierung erfolgt in unserem Beispiel über eine Aufteilung in Bereiche von Flächen in Hektar. Für diese Gruppen werden anschließend die Häufigkeiten – oder hochtrabender gesagt, die Frequenzen – notiert, d. h. es wird notiert wie viele Beobachtungen in die jeweiligen Gruppierungen fallen. Für unseren Beispieldatensatz habe ich neun Gruppen definiert: eine erste Gruppe von 0 bis 10 Hektar, und dann weitere Gruppen von 10 bis 20 Hektar, 20 bis 30 Hektar, 30 bis 50 Hektar, 50 bis 75 Hektar, 75 bis 100 Hektar, 100 bis 125 Hektar, 125 bis 150 Hektar und abschließend 150 bis 200 Hektar.

Anschließend gilt es für die erste Gruppe zu eruieren, wie viele der Werte in die erste Spanne von null bis zehn Hektar fallen. In unserem Beispiel wäre dies 14 Mal der Fall.

In die zweite Spanne von 10 bis 20 fallen vier Werte, und wenn wir dies weiterführen, sehen wir eine Frequenzstruktur, wie sie hier in der rechten Spalte der Tabelle abgebildet ist.

Folie 10 – Frequenztafel

Folientext

- Frequenzdaten geben einen bündigen numerischen Überblick für Häufigkeiten
- Abbildung: Frequenztafel (siehe [Folie 9](#))

Sprechtext

Und somit erhalten wir eine Tabelle mit nun neun Häufigkeitswerten anstelle von 50 Rohwerten und mit diesen neun Häufigkeitswerten lässt sich bereits deutlich besser umgehen als mit jenen 50 Werten aus den Rohdaten. So lässt sich beispielsweise auf einen Blick erkennen, dass mit Häufigkeiten von 14 und 15 die größten Häufigkeiten respektive bei der ersten Gruppe mit Null bis zehn Hektar und der vierten Gruppe mit 30 bis 50 Hektar zu finden sind. Entsprechend vermittelt die Frequenztafel in der Regel einen instantanen Eindruck davon, in welchen Größenkategorien sich die größten Frequenzen in unserem Datensatz wiederfinden lassen. Insofern sind Frequenztabellen ein probates Mittel, um die Komplexität von Daten zu reduzieren und zentrale Eigenschaften der darin enthaltenen Häufigkeitsstrukturen für uns interpretierbar zu machen.

Folie 11 – Thema 3

Folientext

3. Histogramme

Sprechtext

Lassen Sie mich nun auf die Visualisierung dieser Häufigkeitsstrukturen mit Histogrammen eingehen.

Folie 12 – Warum Histogramme nutzen?

Folientext

- Auch Frequenztabellen oft überkomplex
- Intuitiver Umgang mit Visualisierung

Sprechtext

Einsteigen möchte ich mit der Frage, warum in den meisten Kontexten Histogramme anstelle von Frequenztabellen genutzt werden, bzw. warum ich Ihnen empfehlen würde Histogramme statt Frequenztabellen für den ersten Eindruck von Daten zu nutzen. Wenngleich Frequenztabellen zweifelsohne nützlich sind und große Zahlenkolonnen für viele ein gewisses wissenschaftliches Flair mitbringen, bereiten Frequenztabellen oftmals aber auch Probleme. Diese Probleme sind eng verbandelt mit den Problemen der Interpretation der Rohdaten, welche ich eben erläutert habe.

Denn auch für Frequenztabellen gilt, dass die meisten unserer Gehirne von der darin enthaltenen Menge an Zahlen schnell überfordert sind. Denn selbst bei der eben präsentierten Tabelle mit neun Frequenzen steckt ziemlich viel Information in den enthaltenen Zahlen, die es zu verarbeiten gilt. Neben der neun Häufigkeiten selbst gilt es auch noch die Gruppierungen in Form von zwei mal neun Zahlen mit ihren Eigenheiten zu durchblicken. Kurzum – auch Frequenztabellen bedürfen ggf. einer weitergehenden Adaption, damit wir die Datenlage kognitiv besser und schneller greifen können. Und der hier gängigste Weg ist die Visualisierung in Form eines Histogramms. Denn gute Visualisierungen haben den ziemlich grandiosen Vorteil, dass viele Informationen, die sonst umständlich aus Zahlenkonstrukten rausgelesen werden müssen, im Idealfall auf intuitive Weise dargestellt werden.

Folie 13 – Aufbau von Histogrammen

Folientext

- Histogramme visualisieren üblicherweise die relativen anstelle der absoluten Häufigkeiten
- Abbildung: Frequenztable, erweitert um relative Häufigkeit

Fläche in ha	Absolute Häufigkeit	Relative Häufigkeit
$0 \leq x \leq 10$	14	$\frac{14}{50} = 0,28$
$10 \leq x \leq 20$	4	$\frac{4}{50} = 0,08$
$20 \leq x \leq 30$	4	$\frac{4}{50} = 0,08$
$30 \leq x \leq 50$	15	$\frac{15}{50} = 0,30$
$50 \leq x \leq 75$	6	$\frac{6}{50} = 0,12$
$75 \leq x \leq 100$	4	$\frac{4}{50} = 0,08$
$100 \leq x \leq 125$	2	$\frac{2}{50} = 0,04$
$125 \leq x \leq 150$	0	$\frac{0}{50} = 0,00$
$150 \leq x \leq 200$	1	$\frac{1}{50} = 0,02$
Summe	50	1,00

Sprechtext

Für die Erstellung von Histogrammen benötigen wir einige zusätzliche Informationen in unserer Frequenztable. Zuerst wird für Histogramme üblicherweise die relative Häufigkeit anstelle absoluter Häufigkeiten verwendet, d. h. der Anteil an Beobachtungen, die in eine bestimmte Gruppe fallen wird betrachtet. Absolute Häufigkeiten haben den Nachteil, dass wir gegebenenfalls Probleme haben, die Größenordnungen zu überblicken – insbesondere dann, wenn wir mit großen, komplexen und krummen Frequenzen konfrontiert werden. Hingegen ist die Interpretation bei relativen Häufigkeiten in der Regel einfacher, da wir uns hier in dem für die meisten von uns wohlvertrauten Zahlenraum zwischen null und eins bzw. bei einer Darstellung in Prozent zwischen 0 und 100 bewegen. Für die Berechnung der relativen Häufigkeiten teilen wir jetzt die absolute Häufigkeit in jeder Gruppe durch die Summe aller Häufigkeiten, d. h. für die erste Gruppe teilen wir die 14 Beobachtungen durch die Gesamtzahl 50 und erhalten somit eine relative Häufigkeit von 0.28 oder 28%. Für die zweite Gruppe teilen wir die vier Beobachtungen durch die Gesamtzahl 50 und erhalten

somit eine relative Häufigkeit von 0.08 oder 8%. Und dies berechnen wir dann auch für die weiteren sieben Häufigkeiten und haben somit dann die relativen Häufigkeiten, die hier abgebildet sind.

Folie 14 – Aufbau von Histogrammen 2

Folientext

- Auf Grundlage der relativen Häufigkeit wird die Dichte berechnet
- Die Dichte ist gegeben durch: $\text{Dichte} = \frac{\text{Häufigkeit}}{\text{Breite der Klassen}}$
- Abbildung: Frequenztafel, erweitert um Breite und Dichte

Fläche in ha	Breite	Relative Häufigkeit	Dichte
$0 \leq x \leq 10$	10	0,28	$\frac{0.28}{10} = 0,0280$
$10 \leq x \leq 20$	10	0,08	$\frac{0.08}{10} = 0,0080$
$20 \leq x \leq 30$	10	0,08	$\frac{0.08}{10} = 0,0080$
$30 \leq x \leq 50$	20	0,30	$\frac{0.30}{20} = 0,0150$
$50 \leq x \leq 75$	25	0,12	$\frac{0.12}{25} = 0,0048$
$75 \leq x \leq 100$	25	0,08	$\frac{0.08}{25} = 0,0032$
$100 \leq x \leq 125$	25	0,04	$\frac{0.04}{25} = 0,0016$
$125 \leq x \leq 150$	25	0,00	$\frac{0.00}{25} = 0,0000$
$150 \leq x \leq 200$	50	0,02	$\frac{0.02}{50} = 0,0004$

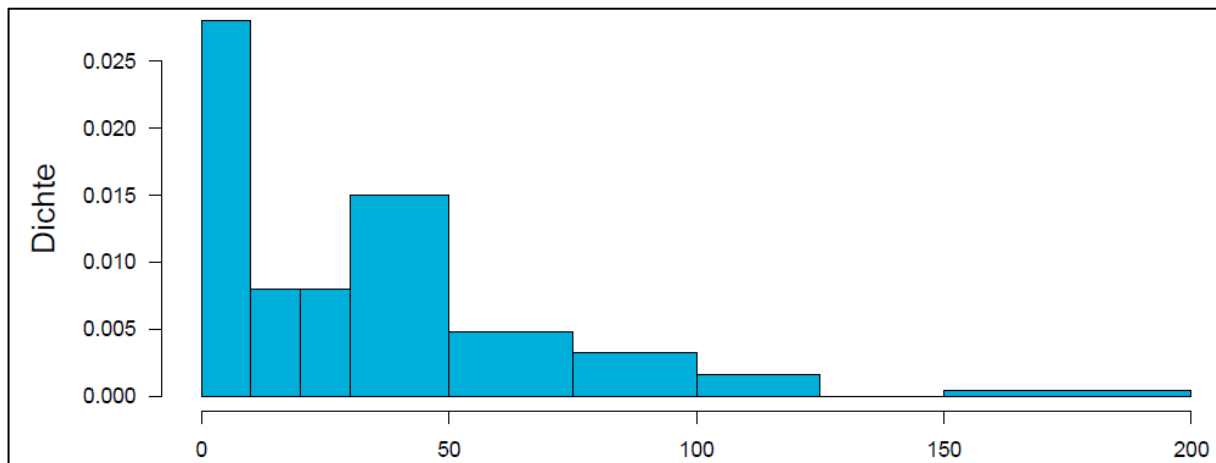
Sprechtext

In einem weiteren Schritt gilt es die sogenannte Dichte zu bestimmen. Diese Dichte berechnet sich aus dem Quotienten der Häufigkeit – üblicherweise die relative Häufigkeit – und der Breite der Klasse auf welcher die Gruppierung der Daten erfolgt ist. In unserem Beispiel ist die Breite der 1. Gruppe 10, während wir in der letzten Gruppe eine Breite von 50 haben. Für die Dichte würden wir entsprechend die relative Häufigkeit der 1. Gruppe – 0,28 durch die Gruppenbreite – 10 – teilen und somit eine Dichte von 0,028 erhalten. Im Fall der letzten Gruppe würden wir die relative Häufigkeit 0,02 durch 50 teilen und somit eine Dichte von 0,0004 erhalten.

Folie 15 – Aufbau von Histogrammen 3

Folientext

- Beim Histogramm sind die Höhen der Rechtecke proportional zur Dichte.
- Die Breite der Rechtecke sind proportional zur Breite der Klasse.
- Ein Histogramm ist somit eine flächenproportionale Darstellung der Häufigkeiten, d.h. die Fläche der Rechtecke ist proportional zu den Häufigkeiten.
- Abbildung: Histogramm



Sprechtext

Mithilfe der so ermittelten neun Dichten können wir nun das Histogramm konstruieren, welches eine graphische Darstellung von Rechtecken ist – üblicherweise in Form senkrechter Balken, wie abgebildet. Die Höhe der Balken wird durch die Dichte bestimmt und deren Breite spiegelt die Breite der Klasse für die jeweilige Gruppe wider. Die zugrundeliegenden Häufigkeiten sind entsprechend nicht aus der Höhe der Balken, sondern aus der Fläche der Balken zu schließen. Hochtrabend ausgedrückt sprechen wir daher bei Histogrammen von einer flächenproportionalen Darstellungsform. Neben verschiedenen anderen Vorteilen erlaubt diese flächenproportionale Darstellungsform eine lokale Interpretation der Daten, wo die Dichte uns die lokale Häufigkeit der Daten an einer beliebigen Stelle wiedergibt. Demnach ist auf einen Blick zu erkennen, dass auf den Intervallen zwischen 0 und 10 und 30 und 50 die Daten relativ reichhaltig verteilt sind, während sie zwischen 20 und 30 vergleichsweise spärlich verteilt sind und ab 50 die Dichte der Daten sukzessive abfällt. Die Fläche der jeweiligen Balken betrachtend ist entsprechend wahrzunehmen, dass sich die meisten Beobachtungen in der 1. und der 4. Gruppierung konzentrieren. Nun könnten wir auch noch weitergehend die einzelnen Balken im Detail analysieren, aber häufig und im Rahmen dieser Vorlesung, geht es bei der Visualisierung der Daten in Histogrammform vorrangig darum, einen groben Eindruck von der Verteilung der Daten zu gewinnen und nicht so sehr darum, die feinen Unterschiede in den Datenstrukturen unter die Lupe zu nehmen. Und insbesondere für einen solchen groben Eindruck erweist sich das Histogramm auf Grundlage seiner Eigenschaften als ein sehr nutzerfreundliches, sehr robustes Werkzeug. Insbesondere ist hervorzuheben, dass es vergleichsweise robust gegenüber abweichenden Spezifikationen der Klassengrößen bzw. der Gruppierungsstruktur ist. Daher ist die visuelle Darstellung von Häufigkeitsstrukturen in Form von Histogrammen häufig ein sehr empfehlenswerter bzw. der empfehlenswerte erste Schritt, wenn Sie einen Eindruck von der Häufigkeitsstruktur in Ihren Daten gewinnen möchten.

Folie 16 – Zusammenfassung

Folientext

In diesem Video haben wir vermittelt...

- wieso eine Vereinfachung von Komplexität in Daten wichtig ist.
- wie Frequenztabellen für eine Komplexitätsreduktion genutzt werden können.
- wieso Histogramme eine bewährte graphische Darstellungsform von Häufigkeitsstrukturen in Daten sind.

Sprechtext

In diesem Video haben wir besprochen, wieso es häufig notwendig ist, Daten in tabellarischer Form oder graphischer Form vereinfacht darzustellen. Anschließend sind wir darauf eingegangen, wie

Frequenztabellen genutzt werden können, um die Häufigkeitsstruktur in Daten darzustellen. Und zuletzt haben wir die Visualisierung dieser Struktur über Histogramme erläutert.

Folie 17 – Vielen Dank für die Aufmerksamkeit

Folientext

Inhalt und Gestaltung

- Dr. Alexander Silbersdorff, Dr. Jennifer Lorenz, Dr. Benjamin Säfken, Sina Ike

Barrierefreiheit und Gestaltung

- Dr. Nina-Kristin Pendzich, Katrin Lux, Thomas Finkbeiner, Susanne Martini, Kristina Schneider

Unterstützung

- Dominik Becker, Nele Wolf, Anna von Plessen

Abbildungen grafischer Logos

- Logo des Sign Lab Göttingen
- Logo des Zentrums für Statistik Göttingen
- Logo des Campus-Institut Data Science Göttingen
- Logo des Projekt Daten Lesen Lernen
- Logo der Firma Yomma

Sprechttext

Damit bin ich am Ende dieses Videos und ich danke Ihnen für Ihre Aufmerksamkeit und danke auch den Unterstützer:innen, die bei der Erstellung dieses Videos mitgewirkt haben. Vielen Dank!

Folie 18 – Anhang: Quellen- und Abbildungsverzeichnis

Folientext

Abbildungs- und Quellenverzeichnis

- Abbildung 1: eigene Grafik
- Abbildung 2: <https://www.piqsels.com/en/public-domain-photo-zwfoh>
- Abbildung 3: eigene Grafik
- Abbildung 4: eigene Grafik
- Abbildung 5: eigene Grafik
- Abbildung 6: eigene Grafik
- Abbildung 7: eigene Grafik

Folie 19 – Anhang: Förderung und Copyright

Folientext

Dieses Video wurde dank Unterstützung des Förderprogramms „Innovative Lehr- und Lernkonzepte: Innovation plus“ des Niedersächsischen Ministeriums für Wissenschaft und Kultur im Studienjahr 2021/22 umgesetzt.

Copyright: Georg-August-Universität Göttingen, www.uni-goettingen.de