

Einführung in die Statistik

Einführung

Dr. Jakob Hoffmann, Economic Geography Group



Ankündigung: Skilager



Kennenlernen



Kennenlernen durch Statistik

Kurzumfrage

Wann bist du heute morgen aufgestanden?

--:-- --

Was ist die PLZ deines aktuellen Wohnortes?

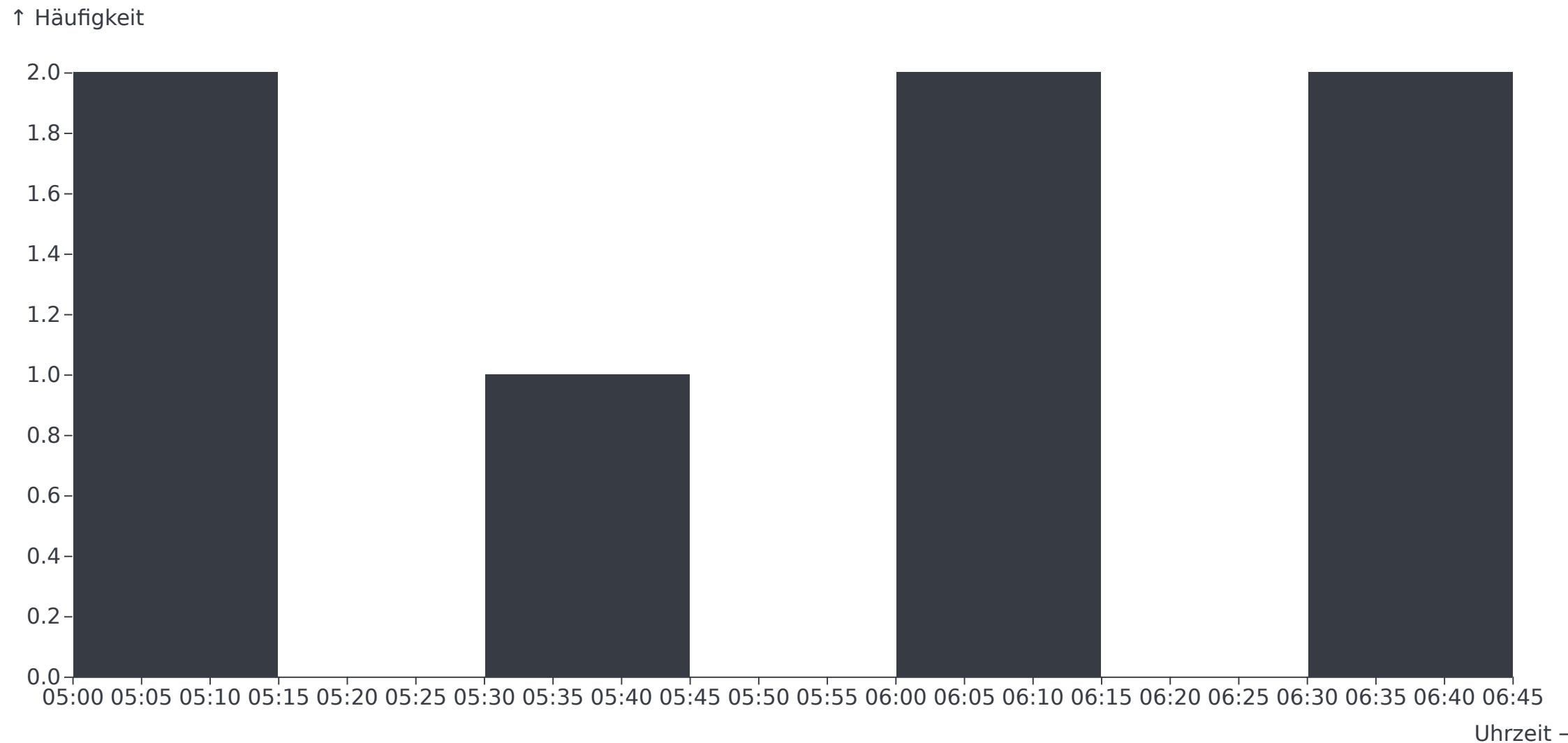
z.B. 80333

Was ist die PLZ deines Wohnortes während des Abiturs?

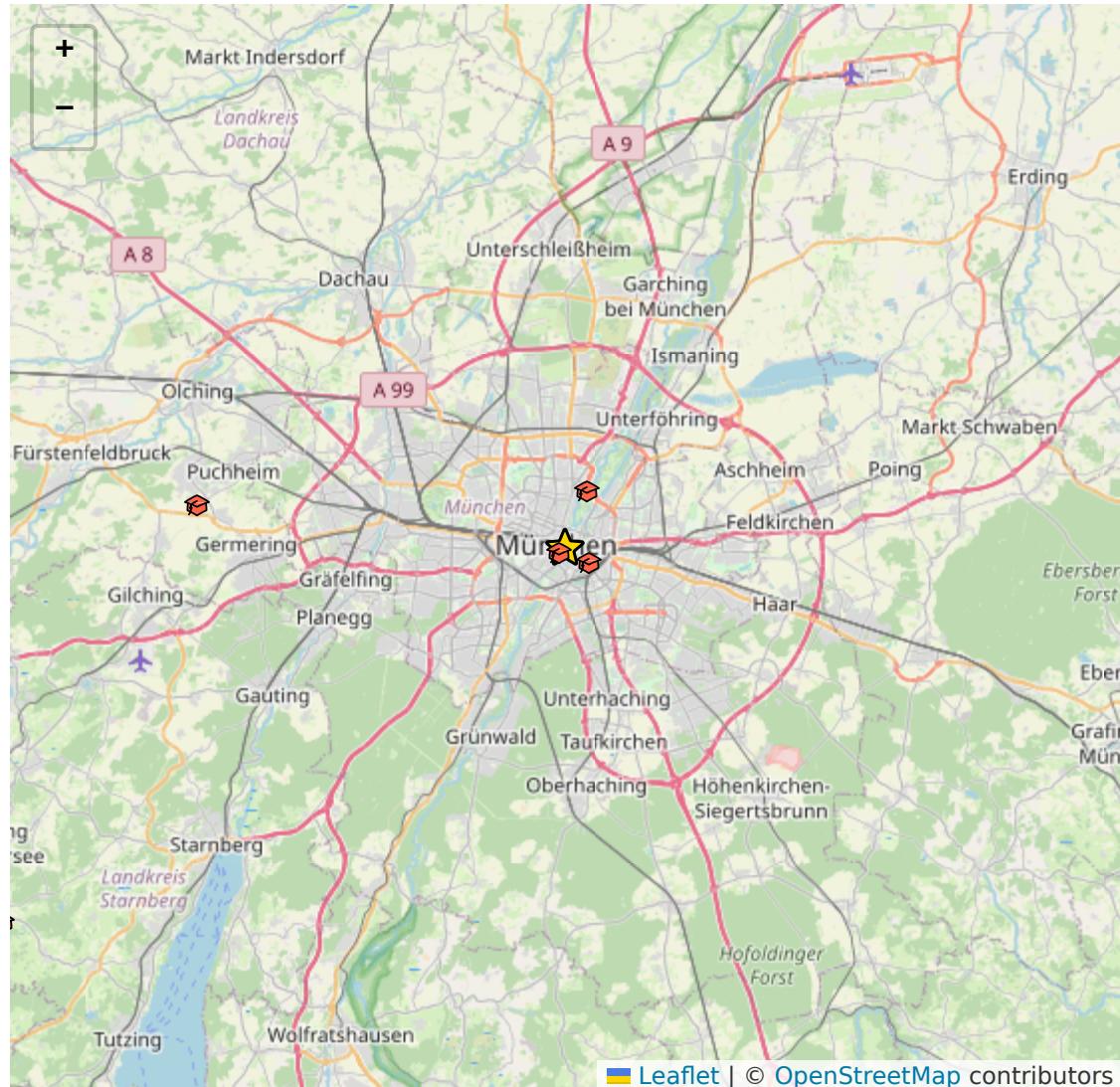
z.B. 40229



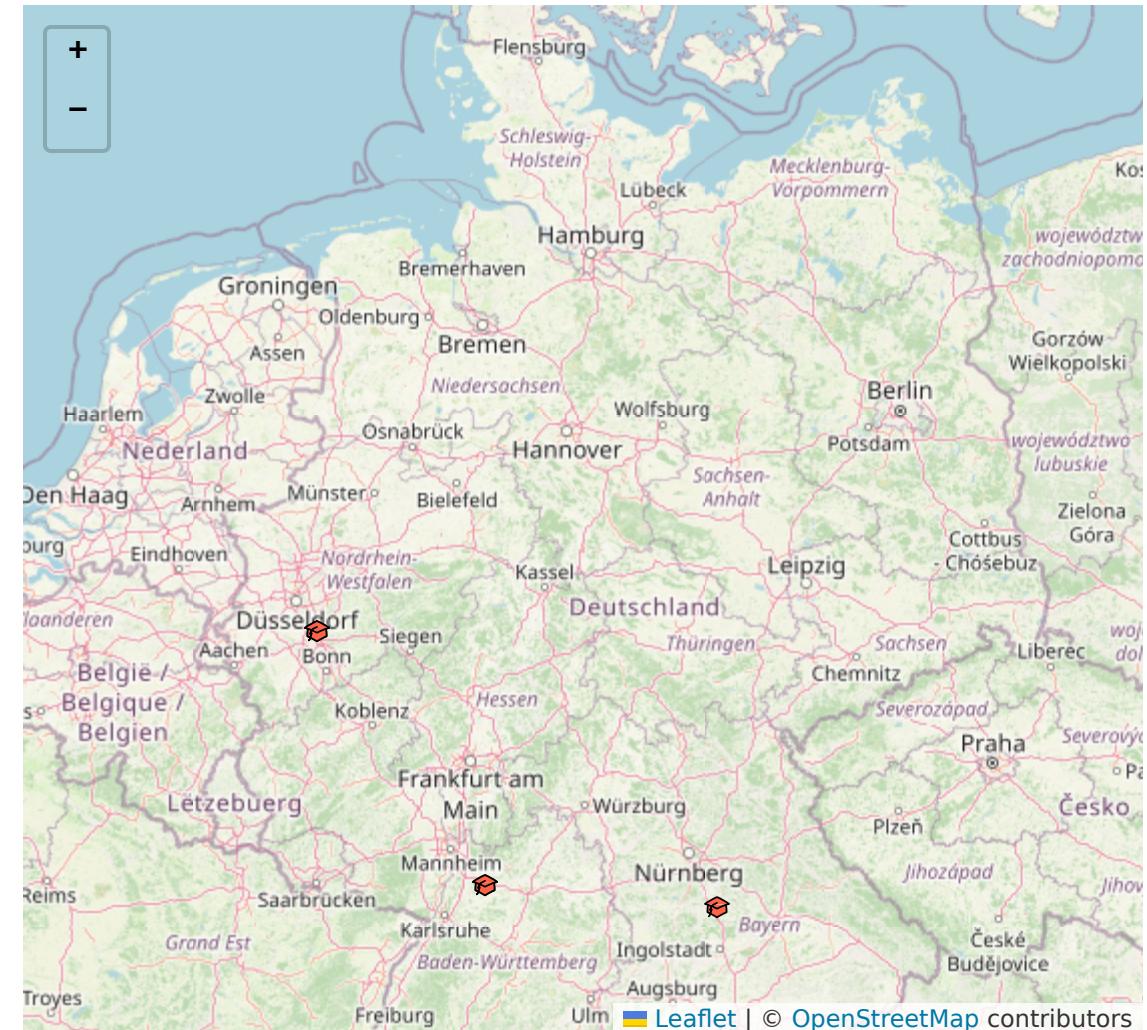
Wann bist du heute morgen aufgestanden?



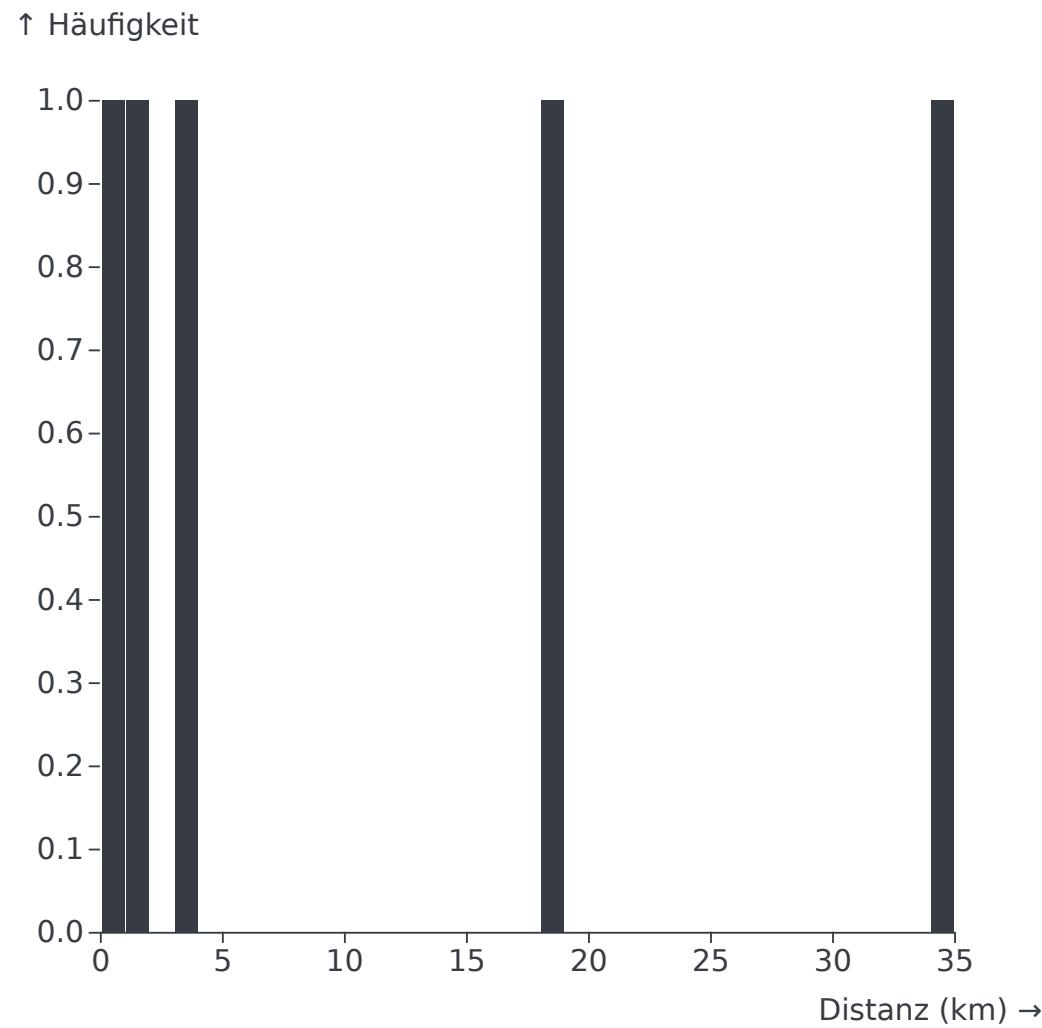
Wohnort aktuell (n = 6)



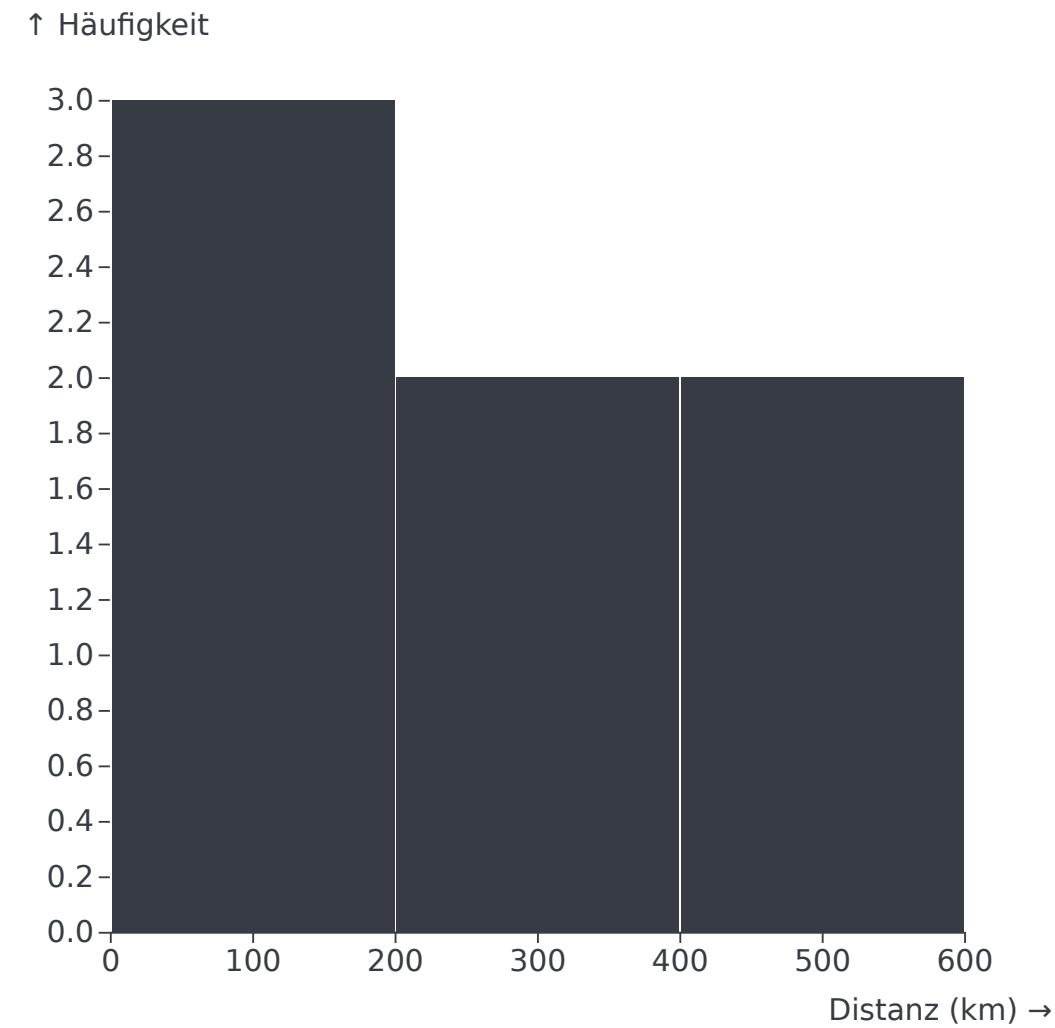
Wohnort während des Abiturs (n = 7)



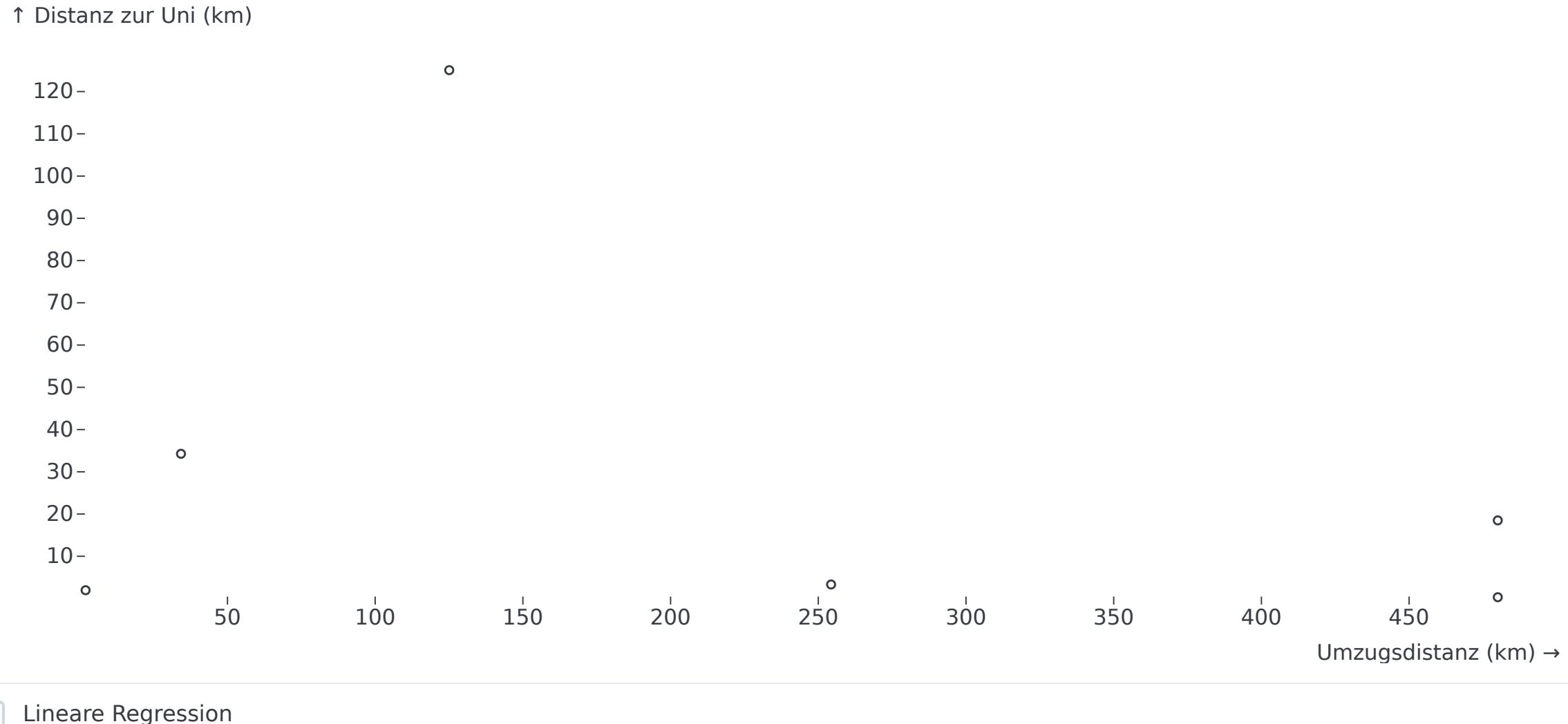
Weg zur Uni (n = 6)



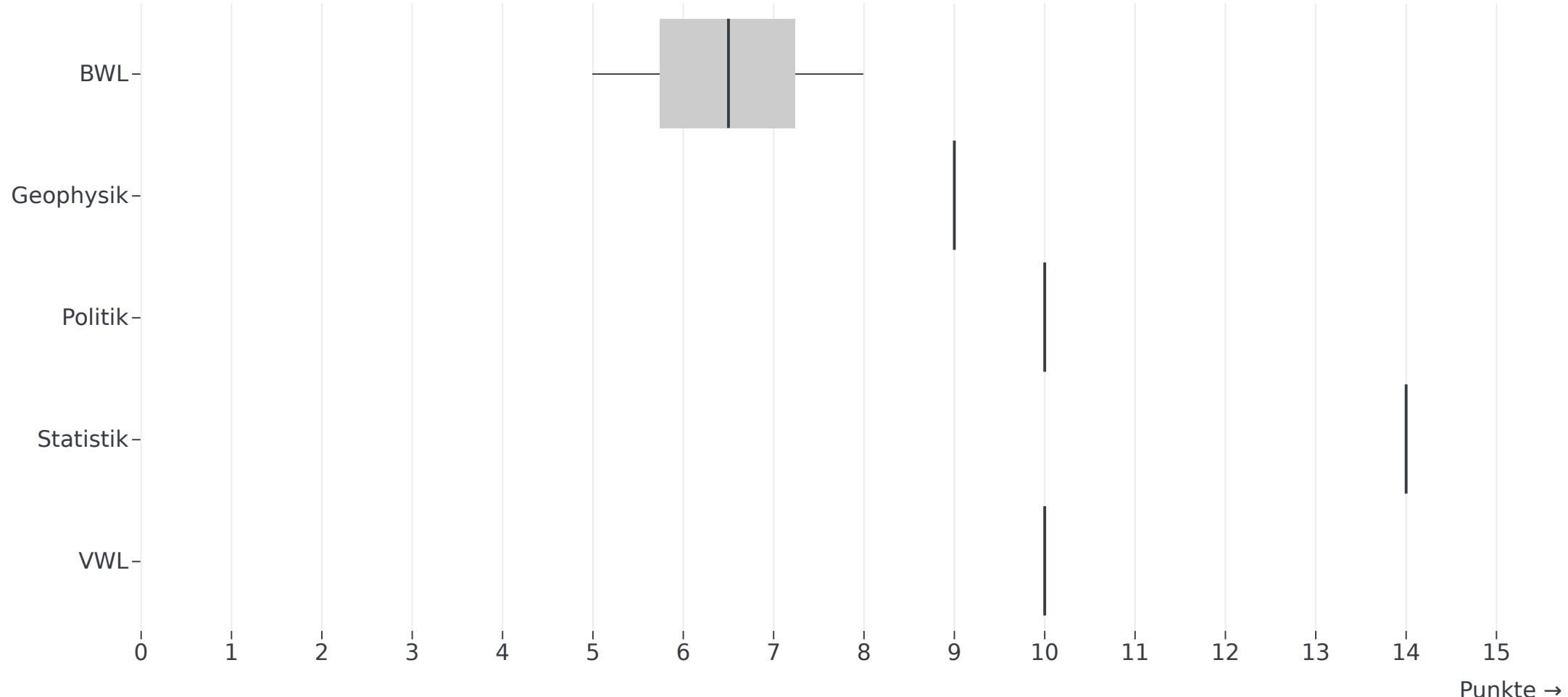
Umzugsdistanz (n = 7)



Wohnen weithergezogene Studierende näher an der Uni?



Wie viele Punkte hattest du im Abitur in Mathe?



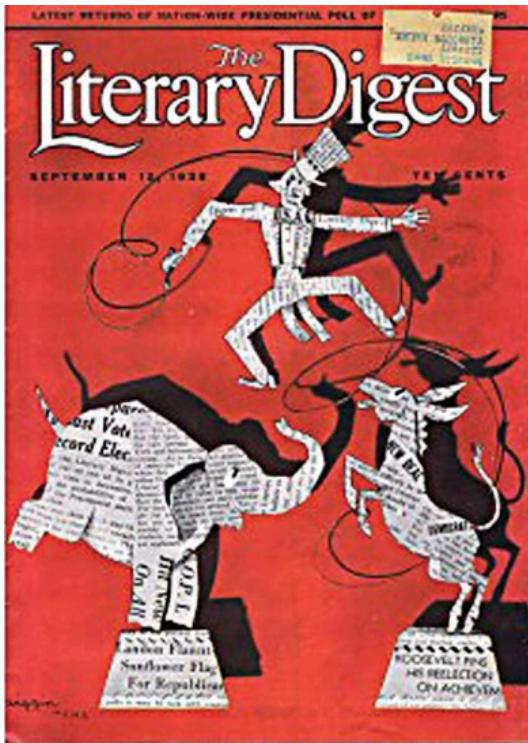
Ziele der Statistik

Ziele der Statistik

- 1. Beschreiben:** Daten veranschaulichen, Informationen extrahieren, auf Basis begrenzter Daten robuste Aussagen über größere Populationen treffen
- 2. Vorhersagen:** Prognosen über unbeobachtete Ausprägungen machen
- 3. Verstehen:** Hypothetische, theoretische Zusammenhänge überprüfen

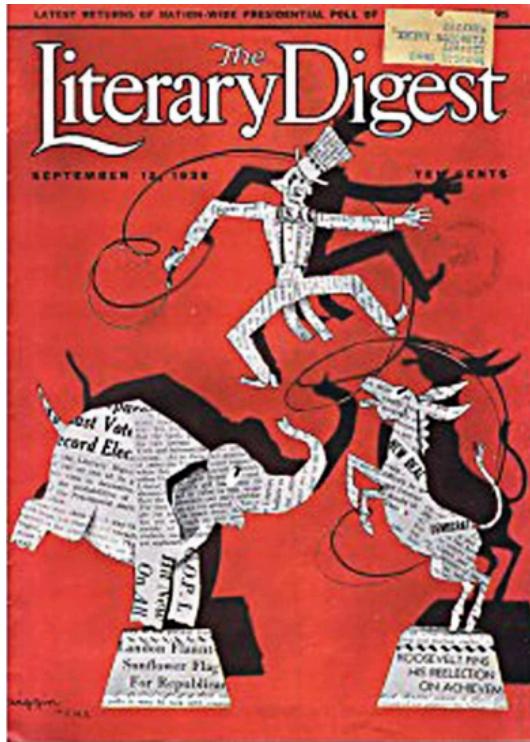


Datenbasierte Kristallkugeln?



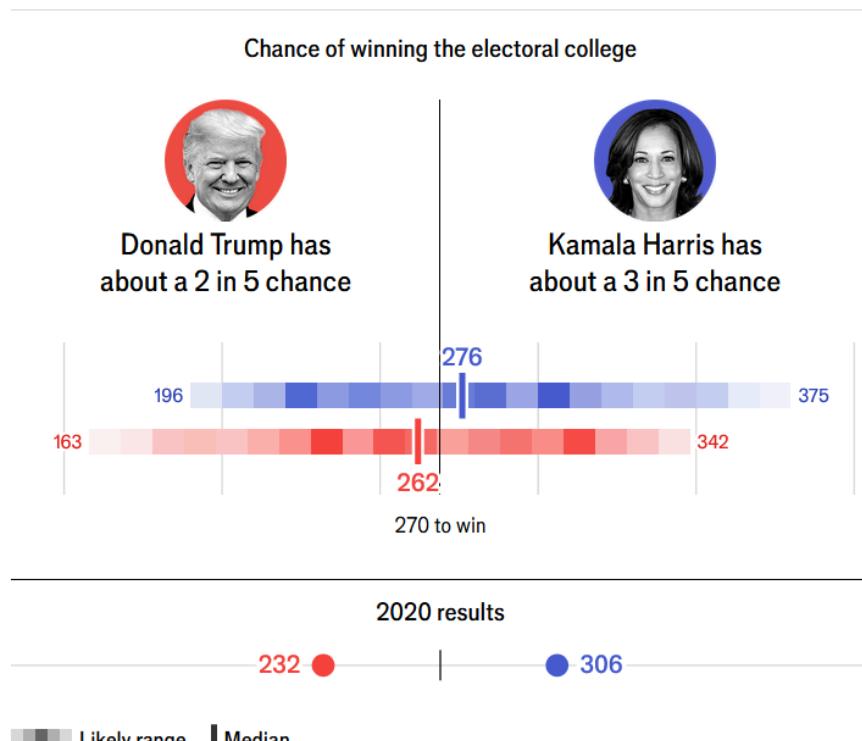
- US-Wahl 1936, Roosevelt vs. Landon
- Wählerbefragung mit **10 mio.** Befragten (ausgewählt aus Leserschaft, Telefonbuch und Automobilregister) und **2.4 mio.** Antworten
- Vorhersage:
 - 43% Roosevelt vs. 57% Landon
- Ergebnis:
 - 62% Roosevelt vs. 38% Landon
- Wie kam es zum Fehler von 38 pp?

Datenbasierte Kristallkugeln?



- **Sampling bias**: Leser des Literary Digest, Automobilbesitzer und Telefonbesitzer waren nicht repräsentativ für die Gesamtbevölkerung
- **Non-response bias**: Gegner Roosevelts haben häufiger an der Befragung teilgenommen als seine Befürworter
- **Frage**: Kann ich auf Basis der vorherigen Befragung auf die Mathenote der gesamten LMU schließen?

Quantifizierung von Unsicherheit



Quelle: Economist

- Der Economist nutzt ein statistisches Modell, welches Wählerbefragungen und andere Prädiktoren kombiniert.
- Aktuell besteht noch große Unsicherheit in der Vorhersage: Ergebnisse von 196 Sitzen (verheerende Niederlage) bis 375 Sitzen (überwältigender Erfolg) für Kamala Harris sind statistisch glaubhaft.
- Statistik dient nicht nur der Extraktion von Informationen aus Daten, sondern auch einer Einschätzung des Grades, zu dem Aussagen durch Daten unterstützt werden.

Quantifizierung von Unsicherheit

Table 1. Likelihood Scale	
Term*	Likelihood of the Outcome
<i>Virtually certain</i>	99-100% probability
<i>Very likely</i>	90-100% probability
<i>Likely</i>	66-100% probability
<i>About as likely as not</i>	33 to 66% probability
<i>Unlikely</i>	0-33% probability
<i>Very unlikely</i>	0-10% probability
<i>Exceptionally unlikely</i>	0-1% probability

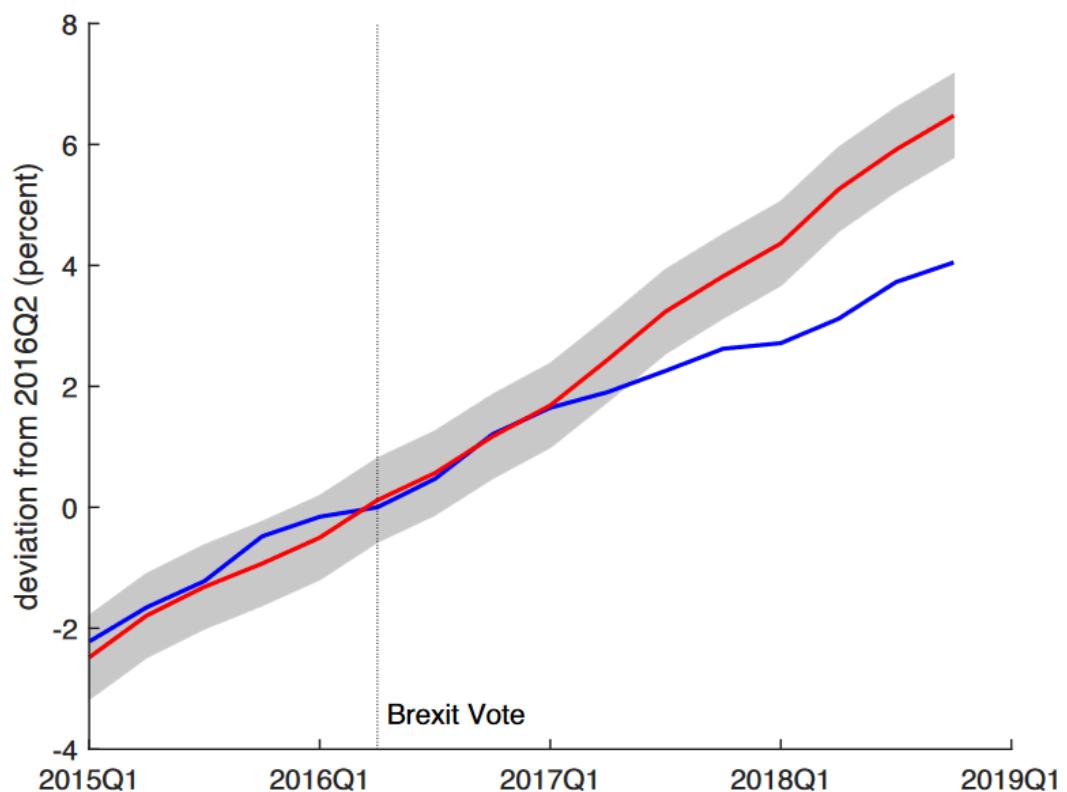
* Additional terms that were used in limited circumstances in the AR4 (*extremely likely* – 95-100% probability, *more likely than not* – >50-100% probability, and *extremely unlikely* – 0-5% probability) may also be used in the AR5 when appropriate.

“It is *virtually certain* that hot extremes (including heatwaves) have become more frequent and more intense across most land regions since the 1950s, while cold extremes (including cold waves) have become less frequent and less severe, with *high confidence* that human-induced climate change is the main driver of these changes. Some recent hot extremes observed over the past decade would have been *extremely unlikely* to occur without human influence on the climate system. Marine heatwaves have approximately doubled in frequency since the 1980s (*high confidence*), and human influence has *very likely* contributed to most of them since at least 2006.”

Quelle: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties



Was wäre wenn...?

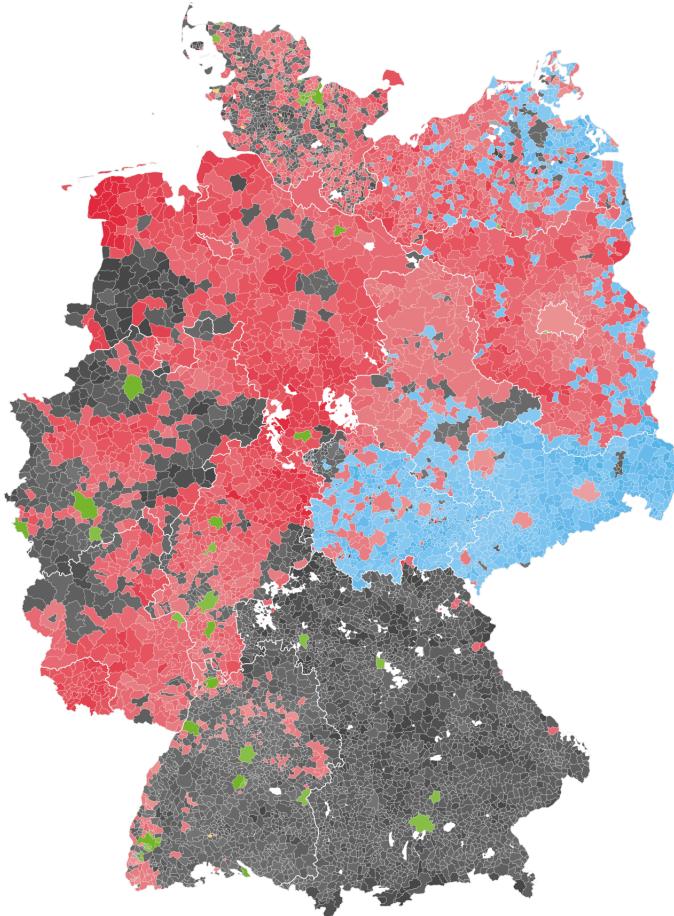


- Frage: Hat der Brexit der britischen Wirtschaft geschadet?
- Synthetic control: Tatsächliche Entwicklung der Wirtschaft (blau) vs. hypothetische Entwicklung ohne Brexit (rot)
- Bis Ende 2018 ist ein geschätzter Outputverlust von 1.7 pp bis 2.5 pp entstanden.

Quelle: Born et al. 2019

Statistik und Geographie

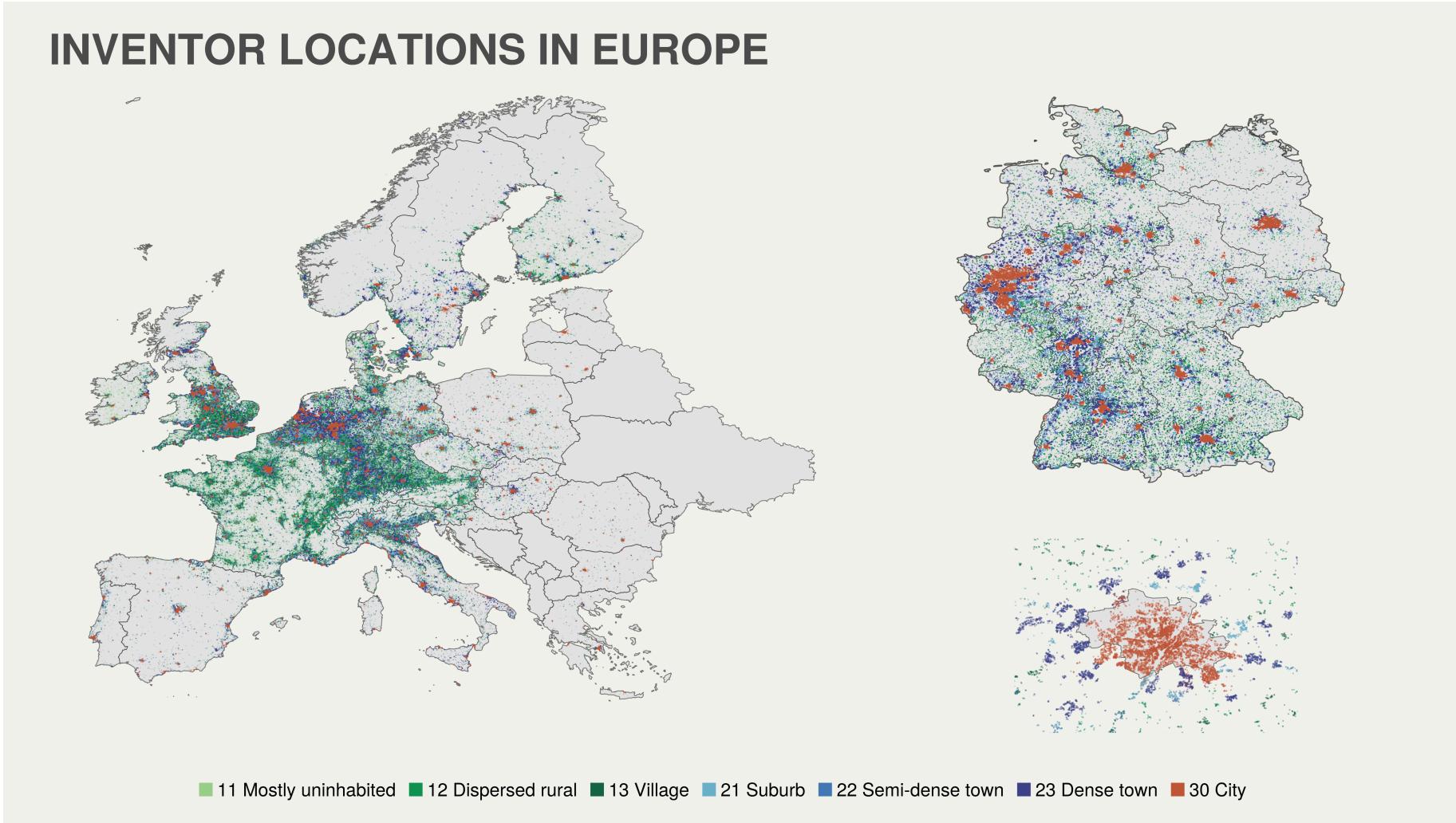
Die Politische Karte Deutschlands



Quelle: Zeit.de

- In Thüringen hat mit der AfD eine vom Verfassungsschutz als rechtsextrem eingestufte Partei die meisten Stimmen erlangt.
- Der Blick auf die Karte suggeriert, dass ein erheblicher Anteil Deutschlands AfD wählt.
- Thüringen beheimatet 2,5% der deutschen Bevölkerung.

Wo erfindet Europa?



Multiskalarität und Aggregation

- TODO: Entwicklung Gesamtdeutschland vs Gewinner / Verlierer
- Mainz
- Spiky map

Räumliche Autokorrelation

Organisatorisches

Was auf euch zu kommt



Lernziele

- Möglichkeiten, Ziele und Grenzen statistischer Analyse
- Interpretation statistischer Ergebnisse und Visualisierungen
- Überblick über Verfahren und deren Anwendungsbereiche



Inhalte

- Statistische Forschungsdesigns
- Daten beschreiben und visualisieren
- Statistische Modelle
- Quantifizierung von Unsicherheit



Diese Vorlesung

- Angewandte (nicht theoretische) Statistik
- Grundlagen deskriptiver und inferentieller Statistik
- Praktische Anwendung mit SPSS in der Übung
- Pause (2 x 40 min)



Übung zur Vorlesung

- Hört nicht auf ältere Studierende; der Besuch der Übung ist nicht optional



Überblick

- Sitzungen / Blöcke



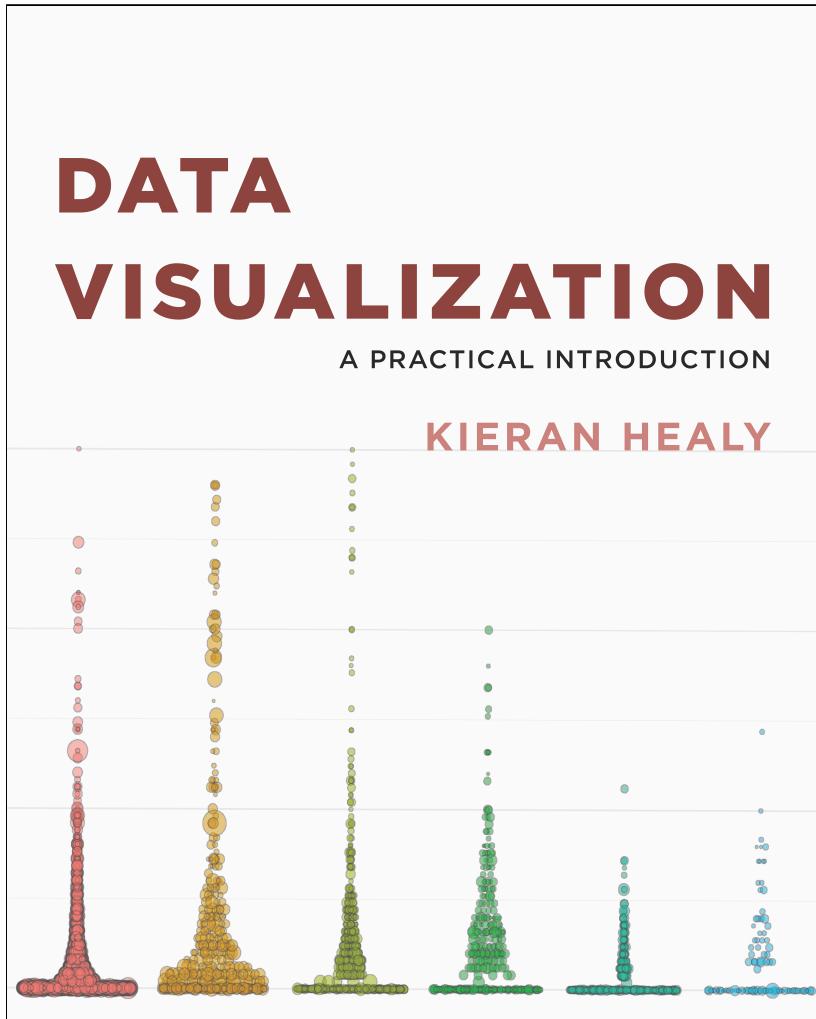
Lernmaterialien



Lehrbücher

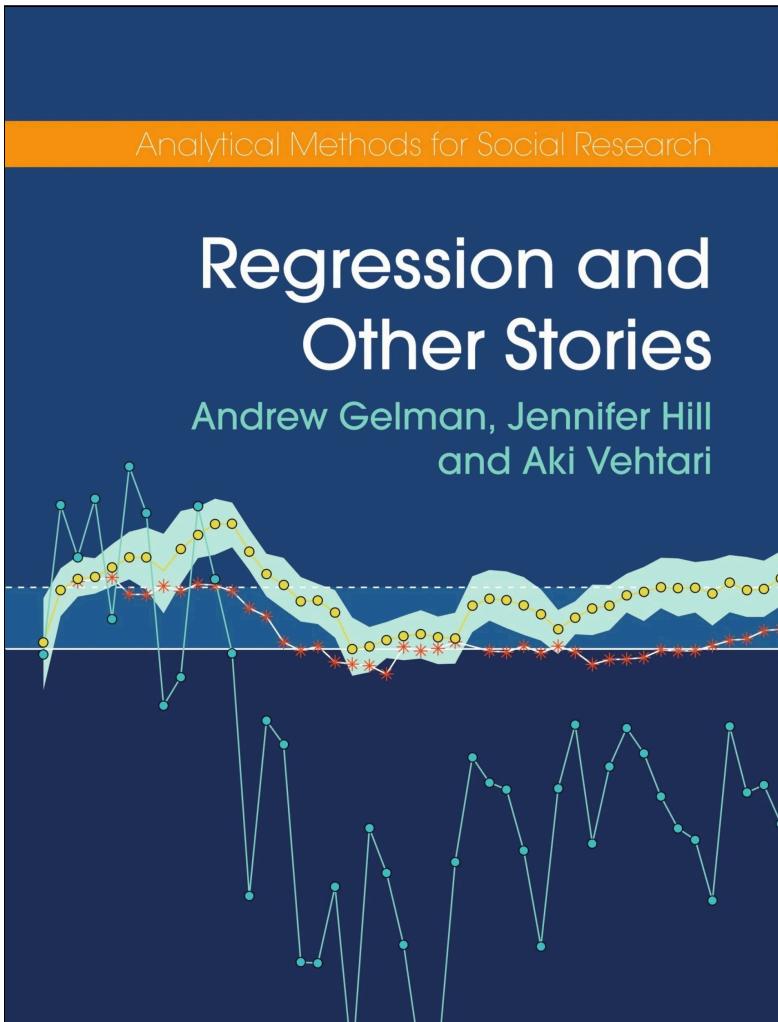


Datenvisualisierung



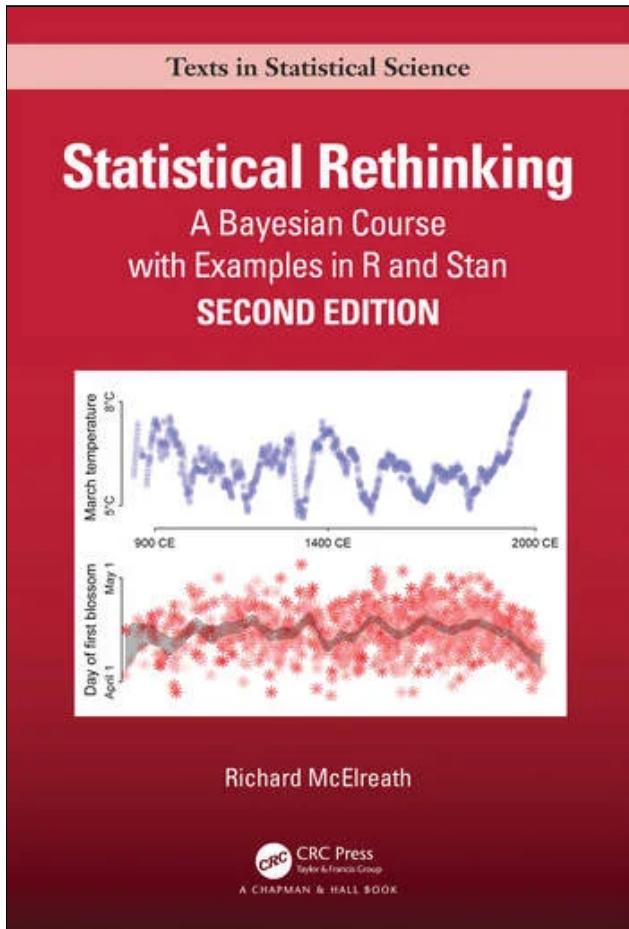
- Einführung in die statistische Visualisierung mit R
- Viele sozialwissenschaftliche Beispiele
- Online verfügbar unter:
<https://socviz.co/>

Regressionsanalyse



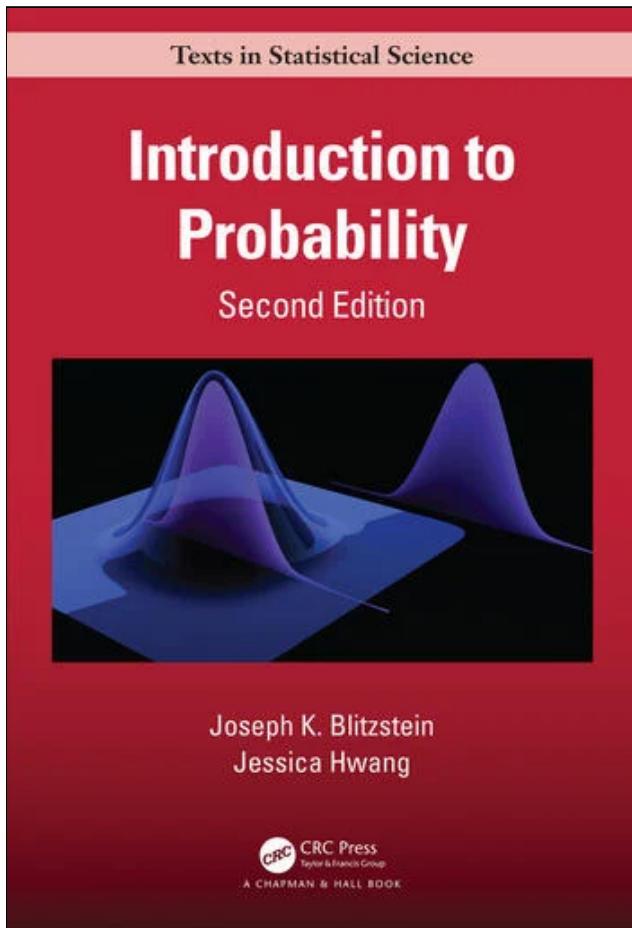
- Umfassende und moderne Einführung in die Regressionsanalyse
- Viele sozial- und politikwissenschaftliche Anwendungsbeispiele in R
- Online verfügbar unter:
<https://avehtari.github.io/ROS-Examples/>

Statistische Modellierung



- Moderne perspektive auf statistische Modellierung
- Bayesianische Statistik, Anwendung in R
- Online verfügbar unter:
<https://xcelab.net/rm/>

Wahrscheinlichkeitstheorie



- Umfassende aber zugängliche Einführung in die Wahrscheinlichkeitstheorie
- Anwendungsbeispiele in R
- Online verfügbar unter: <https://projects.iq.harvard.edu/stat110/home>

Prüfungsfragen



Prüfungsfragen

- Was sind mögliche Fehlerquellen bei der Wahlvorhersage auf Basis von Meinungsumfragen?
- Was sind neben Meinungsumfragen alternative/ergänzende Verfahren zur Vorhersage von Wahlergebnissen?

