**Problem Set 3: Praktische Datenanalyse und Regression**

Mit den Daten aus

Sieberer, U., Saalfeld, T., Ohmura, T., Bergmann, H., & Bailer, S. (2018). Roll-Call Votes in the German Bundestag: A New Dataset, 1949–2013. British Journal of Political Science, 1-9. doi:10.1017/S0007123418000406

Beantworten wir die Fragestellung: Wann weichen Parlamentarier\*Innen im Abstimmungsverhalten von ihrer Fraktion ab? Weiterer theoretischer Hintergrund:

Thomas Zittel, Dominic Nyhuis, Two Faces of Party Unity: Roll-Call Behavior and Vote Explanations in the German Bundestag, Parliamentary Affairs, Volume 72, Issue 2, April 2019, Pages 406–424, <https://doi.org/10.1093/pa/gsx076>

**Aufgaben**

1. Optional: Stellt den Analysedatensatz aus den Rohdaten her. Berechnet den Prozentsatz an abweichenden (weak/strong deviation) Wahlverhalten (vote\_deviate) für alle Parlamentarier\*Innen und verbindet es mit dem Datensatz ihrer Charakteristika. Beschränkt die Analyse auf die 17 Legislaturperiode.
   1. Abstimmungsverhalten - [BTVote Voting Behavior](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/24U1FR)
   2. Charactkeristika der Parlamentarier\*Innen - [BTVote MP Characteristics](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QSFXLQ)

Die Verbindung von zwei unterschiedlichen Datensätzen kann durch die „join“ Befehle durchgeführt werden. Referenz: <https://dplyr.tidyverse.org/reference/join.html>

Aufbereitete Daten für die weitere Datenanalyse:

1. Gibt es ein Verhältnis zwischen abweichendem Wahlverhalten und der Sicherheit des Wahlkreises (elecsafe\_overall)? Stellt dieses Verhältnis graphisch und durch eine Regressionsanalyse dar. Erkläre kurz die Bedeutung der Koeffizienten.
2. Analysiere dieses Verhältnis für Wahlkreissicherheit **und** Fraktionszugehörigkeit. Stell dieses Verhältnis graphisch und durch eine Regressionsanalyse dar. Erkläre kurz die Bedeutung der Koeffizienten.
3. Analysiere dieses Verhältnis für Wahlkreissicherheit **und** Mandatstyp. Stell dieses Verhältnis graphisch und durch eine Regressionsanalyse dar. Erkläre kurz die Bedeutung der Koeffizienten.
4. Welche anderen Kontrollvariablen sollte das Model beinhalten?
   1. Berechne ein eigenes Model, dass alle theoretisch relevanten Variable die in den Daten enthalten sind nutzt. Erkälre kurz, was die Ergebnisse des Models für unsere Fragestellung bedueten.
   2. Diskuttiere kurz welche Variabeln theortisch relevant sind, aber nicht zur Verfügung stehen.