

On weapons and dialects: The initial hypothesis and the final story

Marie-José Kolly | NZZ Visuals
@mjkolly

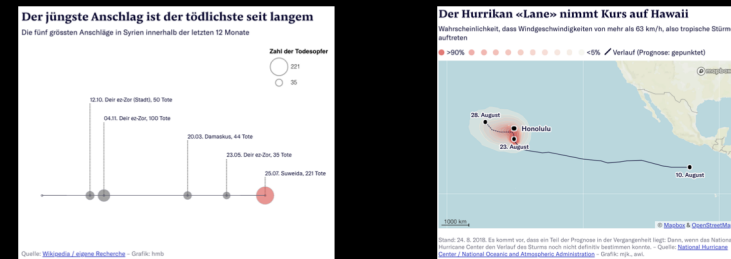
NZZ Visuals, das sind...



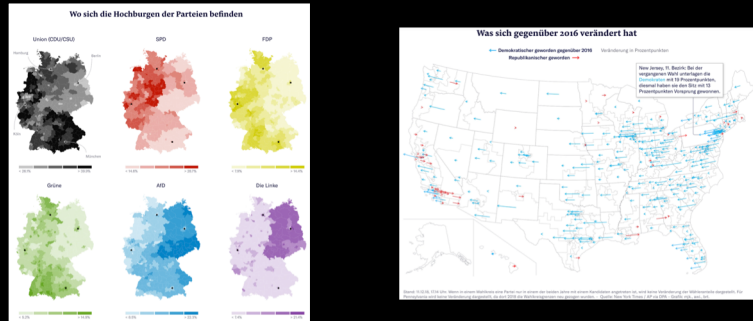
R at the NZZ: Data-driven stories



R at the NZZ: Rapid response



R at the NZZ: Elections



The outcome was unexpected

Case 1: on weapon exports

- No specific anticipation
- Data made a clear point

Case 2: on dialects

- Expected specific effects
- No confirmation

The outcome was unexpected

Case 1: on weapon exports

- No specific anticipation
- Data made a clear point

Case 2: on dialects

- Expected specific effects
- No confirmation

The outcome was unexpected

Case 1: on weapon exports

- No specific anticipation
- Data made a clear point

Case 2: on dialects

- Expected specific effects
- No confirmation

On weapon exports: Motivation

Debate over export regulation.

But how much war material is **already** sold to warring countries?

Hypothesis: No specific anticipation.

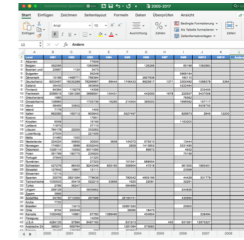
Data on weapon exports

State Sekretariat for Economic Affairs (Seco)

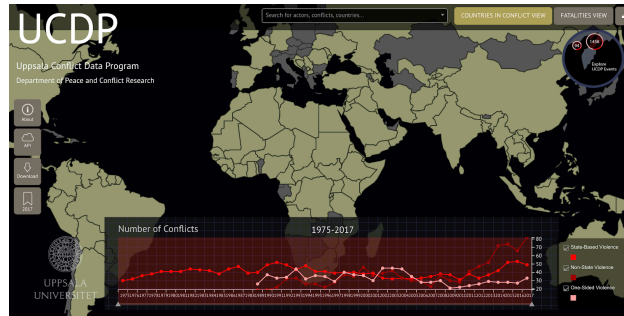
SRF Data

Ausfuhr von Kriegsmaterial (01.01. - 31.12.2017)
 Die Tabelle enthält die Werte der ausgeführten Güter, die nach Artikel 5 des Bundesgesetzes vom 15.12.1998 über die Kriegsmaterial (SRM 141.11) bzw. nach der Liste in Anhang 1 zur Verordnung vom 22.12.1998 über das Kriegsmaterial (SRM 141.11) als Kriegsmaterial definiert sind. Diese zu den Bestimmungsländern im Sinne von Artikel 10 Absatz 4 der Verordnung vom 12.10.2011 über die Statistik über Ausfuhrerlöse (SRM 622) zu addieren, ist das Land, in welchem die Ware verwendet oder vor ihrer endgültigen Wiederverkauf verpackt, verwendet oder sonst wie beabsichtigt wird. Die SRMCO wird jährlich (bis zum vergangenen Bestehen) wieder veröffentlicht. Das Land der letzten Bestimmung, d. h. der Bestimmungsort, ist jeweils vermerkt. Die Bestimmungsorte sind jeweils (unter Berücksichtigung des Landes) angegeben.
 Für weitergehende Informationen zu den Kriegsmaterialexporten (z. B. Bewilligungen, Endbestimmungen) bitte unter www.seco.admin.ch/Abfragen oder unter www.seco.admin.ch/Abfragen oder unter www.seco.admin.ch/Exportkontrollen/Kriegsmaterial (Facebook).

Land	Januar - Dezember 2016 CHF	Januar - Dezember 2017 CHF
Deutschland	80 154 250	117 709 071
Frankreich	17 320 441	11 147 483
Italien	3 314 458	4 306 050
Niederlande	4 202 349	1 148 308
Belgien	2 300 648	2 770 422
Luxemburg	209 610	249 340
Dänemark	3 180 706	7 699 313
Von Kriegsmaterial	17 234 172	4 284 214



Data on conflicts



Data on conflicts

conflict_id	location	side_a	side_a_id	side_b	side_b_id	year	start_date	start_prec	start_prec2	end_date	end_prec
230	Yemen (North Yemen)	Government of Yemen (North Yemen)	323	Ansarullah, AQAP, Forces of Had	1093, BIA, G. Government	2015	15.03.48	4	17.12.09	1	0
230	Yemen (North Yemen)	Government of Yemen (North Yemen)	323	Forces of Had	6214 Government	2016	15.03.48	4	17.12.09	1	0
230	Yemen (North Yemen)	Government of Yemen (North Yemen)	323	Forces of Had	6214 Government	2017	15.03.48	4	17.12.09	1	1
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2000	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2001	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2006	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2007	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2008	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2009	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2010	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2011	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2012	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2013	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2014	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2015	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2016	11.09.01	1	11.09.01	1	0
418	United States of America	Government of United States of America	769	G. Government of Qaida	769	2017	11.09.01	1	11.09.01	1	04.12.17
325	Pakistan	Government of Pakistan	142	BIA	387	2004	18.05.79	1	05.07.11	1	1
325	Pakistan	Government of Pakistan	142	Rajshah, BIA	288, 287	2006	18.05.79	1	26.03.06	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA	287, 287	2007	18.05.79	1	26.03.06	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA	287, 289	2008	18.05.79	1	26.03.06	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA, BIA	287, 289	2009	18.05.79	1	26.03.06	1	20.10.09
325	Pakistan	Government of Pakistan	142	BIA	287	2011	18.05.79	1	05.07.11	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA, BIA	287, 1146, 289	2012	18.05.79	1	05.07.11	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA	287	2013	18.05.79	1	05.07.11	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA	287, 289	2014	18.05.79	1	05.07.11	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA, BIA	287, 1146, 289	2015	18.05.79	1	05.07.11	1	0
325	Pakistan	Government of Pakistan	142	BIA, BIA	287, 1152	2016	18.05.79	1	05.07.11	1	1
325	Pakistan	Government of Pakistan	142	BIA, BIA	287, 1152	2017	18.05.79	1	05.07.11	1	23.10.18

Scripting: Rename countries via countrycode

```

180 # -----
181 ## RENAME COUNTRIES FOR LATER MATCHING THEM WITH CONFLICT DATA ##
182
183 # Get German and English country names to match export data (German) with conflict data (English)
184 codelist_enDe<-data.frame("en"~as.character(codelist$country.name.en), "de"~as.character(codelist$country.name.de))
185 codelist_enDe<-as.character(codelist_enDe)
186
187 # Create duplicate of dataframe as we will rename some countries
188 export_restructured_dup<-export_restructured
189
190 # Which levels of Land are not found in our country vector?
191 export_restructured_dup$Land[which(!export_restructured_dup$Land %in% codelist_enDe)] %>% unique()
192
193 # Check how these should be written
194 codelist_enDe[grep("Poly", codelist_enDe),]
195
196 # Replace variants in variable "Land" to match codelist
197 export_restructured_dup$Land[export_restructured_dup$Land == "Grossbritannien"] <- "Großbritannien"
198 export_restructured_dup$Land[export_restructured_dup$Land == "Hong-Kong"] <- "Hongkong"
199 export_restructured_dup$Land[export_restructured_dup$Land == "Tschechien"] <- "Tschechische Republik"
200 export_restructured_dup$Land[export_restructured_dup$Land == "Tschechische Rep. "] <- "Tschechische Republik"
201 export_restructured_dup$Land[export_restructured_dup$Land == "USA"] <- "Vereinigte Staaten"
202 export_restructured_dup$Land[export_restructured_dup$Land == "Saudi Arabien"] <- "Saudi Arabien"
203 export_restructured_dup$Land[export_restructured_dup$Land == "Heiliger Stuhl"] <- "Vatikanstaat"
204 export_restructured_dup$Land[export_restructured_dup$Land == "Macao"] <- "Macao"
205 export_restructured_dup$Land[export_restructured_dup$Land == "Russland"] <- "Russische Föderation"
206 export_restructured_dup$Land[export_restructured_dup$Land == "Südkorea"] <- "Korea, Republik von"
207 export_restructured_dup$Land[export_restructured_dup$Land == "Brunei"] <- "Brunei Darussalam"
208 export_restructured_dup$Land[export_restructured_dup$Land == "Suriname"] <- "Surinam"
209 export_restructured_dup$Land[export_restructured_dup$Land == "Bosnien-Herzegowina"] <- "Bosnien und Herzegowina"
210 export_restructured_dup$Land[export_restructured_dup$Land == "Kap Verde"] <- "Cabo Verde"
211 export_restructured_dup$Land[export_restructured_dup$Land == "Trinidad Tobago"] <- "Trinidad und Tobago"

```

Scripting: «Reshape» data

```

277 # When there are several actors on one side, duplicate all the according rows and make country name unique (e.g. entries for India, entries for Pakistan, in the case of location="India,
278 # Strategy: create new dataframe with duplicate, replace country name, replace country name in original data frame, rbind
279
280 # Which levels are not matched yet?
281 conflictGov$mainSides[which(!conflictGov$mainSides %in% codelist$country.name.en)] %>% unique()
282 conflictGov$oSides[which(!conflictGov$oSides %in% codelist$country.name.en)] %>% unique()
283
284 codelist_enDe[grep("Mor", codelist_enDe),]
285
286 # India, Pakistan
287 indpak<-conflictGov[conflictGov$mainSides == "India, Pakistan",]
288 indpak$mainSides<- "Pakistan"
289 conflictGov$mainSides[conflictGov$mainSides == "India, Pakistan"]<- "India"
290 conflictGov<-rbind(conflictGov,indpak)
291
292 # Cambodia (Kampuchea), Thailand
293 camtha<-conflictGov[conflictGov$mainSides == "Cambodia (Kampuchea), Thailand",]
294 camtha$mainSides<- "Thailand"
295 conflictGov$mainSides[conflictGov$mainSides == "Cambodia (Kampuchea), Thailand"]<- "Cambodia"
296 conflictGov<-rbind(conflictGov,camtha)
297
298 # Eritrea, Ethiopia
299 eriethe<-conflictGov[conflictGov$mainSides == "Eritrea, Ethiopia",]
300 eriethe$mainSides<- "Ethiopia"
301 conflictGov$mainSides[conflictGov$mainSides == "Eritrea, Ethiopia"]<- "Eritrea"
302 conflictGov<-rbind(conflictGov,eriethe)
303
304 # South Sudan, Sudan

```

Scripting: Merge dataframes

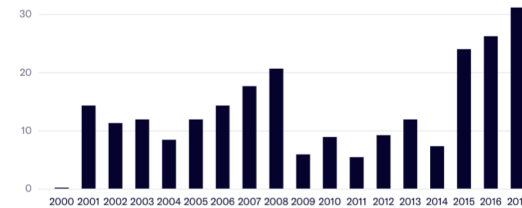
```

464 # Merge the conflict and the weapon export data using the English codelist column
465 conflict_export_merged<-conflict_export %>%
466   left_join(weapon_export_merged, by=c("mainSides"="en"))
467
468 # How many countries were sold weapons in years where they were in conflict?
469 cntries_export_conflict_sometime<-conflict_export_merged %>%
470   filter(!is.na(conflict_export_merged$Land)) %>%
471   filter(value@ $ Jahr == start_year & Jahr == end_year)
472   unique(cntries_export_conflict_sometime$Land)
473
474
475
476 # -----
477 ## RESULTS AND VISUALIZATION ##

```

Kriegsmaterialexporte an Konfliktparteien nahmen jüngst wieder zu

Exporte in Länder, deren Regierung im Exportjahr in einen internen oder internationalen Konflikt verwickelt war, als Anteil an den gesamten Kriegsmaterialexporten, in Prozent



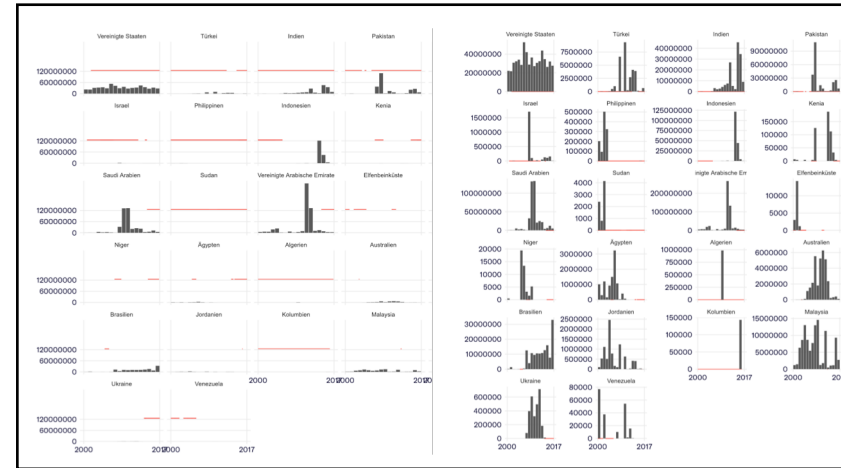
Sämtliche Exportzahlen in diesem Artikel umfassen Kriegsmaterial im Sinne der Schweizer Gesetzgebung: Dual-Use-Güter, die sowohl für zivile als auch für militärische Zwecke verwendet werden können, sowie «besondere militärische Güter», wurden nicht berücksichtigt. – Quellen: [Staatssekretariat für Wirtschaft \(Seco\)](#), [Utopiaa Conflict Data Program](#) – Grafik: mjk. [Daten herunterladen](#)

More scripting: visualization

```

568 # Plot small multiples
569 ggplot(countryXtime_aum_conflict, aes(x=Jahr_aum, y=value))>
570   geom_segment(data=conflict_export_red, aes(x=start_date3, xend=ep_end_date, y=125000000, yend=125000000, col="darkgray"))>
571   geom_bar(stat="identity") +
572   facet_wrap(~Land.neu) +
573   scale_x_date(breaks=as.Date(c("2000-01-01", "2017-12-31")), labels=c("2000", "2017"),
574               limits = as.Date(c("2000-01-01", "2017-12-31")))>
575   scale_y_continuous(breaks=c(0, 60000000, 120000000), labels=c("0", "60000000", "120000000"))>
576   theme_minimal() +
577   theme(axis.text.x=element_text(family="GT America", color="#05032d", size=11),
578         axis.text.y=element_text(family="GT America", color="#05032d", size=11),
579         legend.position="none",
580         axis.title.x=element_blank(),
581         axis.title.y=element_blank(),
582         panel.grid.minor.x = element_blank(),
583         panel.grid.minor.y = element_blank())
584 ggsave("conflict_export.svg", width=12)
585 ggsave("conflict_export.pdf", width=12)
586
587 #Plot small multiples with free Y-scale
588 ggplot(countryXtime_aum_conflict, aes(x=Jahr_aum, y=value))>
589   geom_segment(data=conflict_export_red, aes(x=start_date3, xend=ep_end_date, y=2, yend=2, col="darkgray"))>
590   geom_bar(stat="identity") +
591   facet_wrap(~Land.neu, scales="free.y") +
592   scale_x_date(breaks=as.Date(c("2000-01-01", "2017-12-31")), labels=c("2000", "2017"),
593               limits = as.Date(c("2000-01-01", "2017-12-31")))>
594   theme_minimal() +
595   theme(axis.text.x=element_text(family="GT America", color="#05032d", size=11),

```



R to design



The figure is a screenshot of the Neue Zürcher Zeitung website. The main headline is "Huawei-Finanzchefin verhaftet, Macron kommt - gilt es jaunes - entgegen, Serbien droht Kosovo: der kompakte Nachrichtenüberblick am Donnerstag". Below the headline is a large bar chart showing Huawei's financial performance in Canada. The chart has a y-axis ranging from 0 to 60,000 and an x-axis with labels 2000 and 2017. The article text below the chart reads: "So viele Waffen exportiert die Schweiz bereits heute in Länder, die Krieg führen". The article is dated 6.12.2018 and is written by Stefan Hübner.



We write about methods. And publish code on github.com/nzzdev/st-methods

Die Methodik im Detail

Datenquellen
Das [Uppsala Conflict Data Program](#) (UCDP) der Universität Uppsala und des Peace Research Institute in Oslo [gammelt und kuratiert](#) weltweit Daten zu bewaffneten Konflikten. Seine Datenbank reicht bis ins Jahr 1946 zurück und wird jährlich aktualisiert. Zum Zeitpunkt der Publikation dieses Artikels reichten die Daten bis zum 31. 12. 2017. Das UCDP publiziert seine Forschung regelmäßig in wissenschaftlichen Artikeln, und seine Datenbank dient Forschern, Organisationen der Vereinten Nationen und Regierungen als [Datatank](#).

Daten zu Kriegsmaterialexporten stellt das [Swiss Inventory for Weapons](#) (SIW) nach Güterklassen aufgeschlüsselt zur Verfügung. Vergleichbar sind die Daten ab dem Jahr 2000.

Was ein Konflikt ist
Nach der Definition des UCDP liegt ein bewaffneter Konflikt vor, wenn zwischen zwei Parteien bewaffnete Auseinandersetzungen stattfinden, die die Führung oder das Territorium betreffen und in welchen in einem Kalendertag 25 Personen oder mehr umkommen.

Welche Konflikte wir berücksichtigen
Wir berücksichtigen für diesen Artikel Staaten, deren Regierung gemäss UCDP als *aktive Akteure* an einem Konflikt beteiligt ist. Ausnahmen davon bilden zum einen *Scheidbarkeiten* und die *Wenigsten Arabische Emirate*, die gemäss dem UCDP im Zeitraum zwischen 2000 und 2017 lediglich als *reaktive Akteure* auftraten, also eine Partei unterstützen und nicht als *zentrale* Konfliktpartei galten. Auch in der [Eingangsseite des Datenartikels](#) für [Westsahara](#) wird die saarische Beteiligung am Bürgerkrieg in Jemen nicht unter das *Aktorenkriterium* der Kriegserklärung, da es sich dabei nicht um einen internen bewaffneten Konflikt auf ihrem Territorium handelt. Die NZZ aber betrachtet diese Staaten im Jemen-Konflikt – wie gemäss dem UCDP als *aktive Akteure* die jemenitische Regierung gegen weitere Akteure kämpft – als *zentrale* Konfliktparteien. Ergänzt nach *Scheidbarkeiten* und in die *Emirate* wurden deshalb für den Zeitraum ihrer Beteiligung an diesem Konflikt mit berücksichtigt.

The outcome was unexpected

Case 1: on weapon exports

- No specific anticipation
- Data made a clear point

Case 2: on dialects

- Expected specific effects
- No confirmation

On dialects

Stereotype: «Mischdialekte»

Hypothesis:
More mobility => more «mix»

More linguistic variability in larger places?

Die Methodik im Detail

Sprachwissenschaftliche Methode

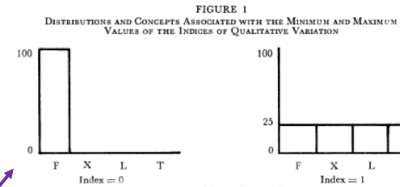
Wo wandelt sich die Sprache, und wie schnell? Um diese Fragen zu beantworten, kennt die Linguistik verschiedene Methoden. Eine von mehreren Möglichkeiten ist es, den Sprach- oder Dialektgebrauch verschiedener Generationen zu vergleichen. **Unterscheidet sich die jüngere Generation von der älteren, kann das ein Hinweis auf Sprachwandel sein.** Das setzt die **Annahme** voraus, dass Individuen ihre Sprechweise weitgehend beibehalten, wenn sie älter werden.

Datenerhebung und Datensatz

Die Daten für die vorliegende Analyse wurden von März 2013 bis September 2017 über die Smartphone Applikation **«Dialekt App»** erhoben (von A. Leemann, M.-J. Kolly, D. Wanitsch, F. Hasler, C. Eisenblatt, I. Werlen). Zahlreiche Schweizerinnen und Schweizer haben dabei je für 16 Dialektwörter angegeben, wie sie sie sagen oder aussprechen, zudem aus welchem Ort ihr Dialekt stammt und wie alt sie sind – und ihre Daten für die Forschung freigegeben. Details zur Applikation und diesem Crowdsourcing, insbesondere zu Vor- und Nachteilen dieser Erhebungsmethode, können **in diesem Artikel** nachgelesen werden.

Der ursprüngliche Datensatz, der von der Applikation erhoben wird, enthält teilweise mehrere 16-Wörter-Datensätze pro Gerät, auf dem die Applikation installiert ist – etwa wenn eine Smartphone-Besitzer sein Gerät an seine Freunde ausgeliehen hat, die dann ihre Daten ebenfalls für die Forschung freigegeben haben. Oder wenn derselbe Smartphone-Besitzer mehrmals Daten übermittelt hat.

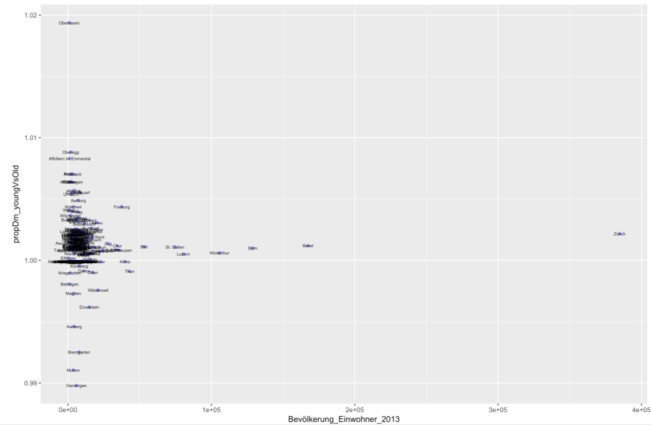
Statistics: measure linguistic variability via qualvar



Deviation from the mode = 0

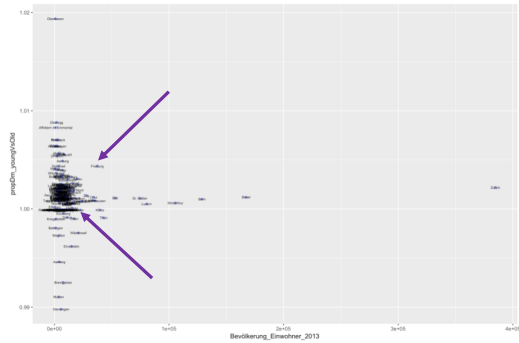
Deviation from the mode = 1

Results

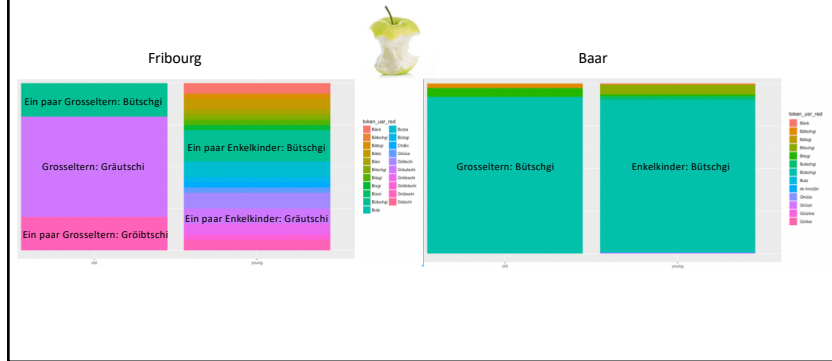


Find an alternative spin

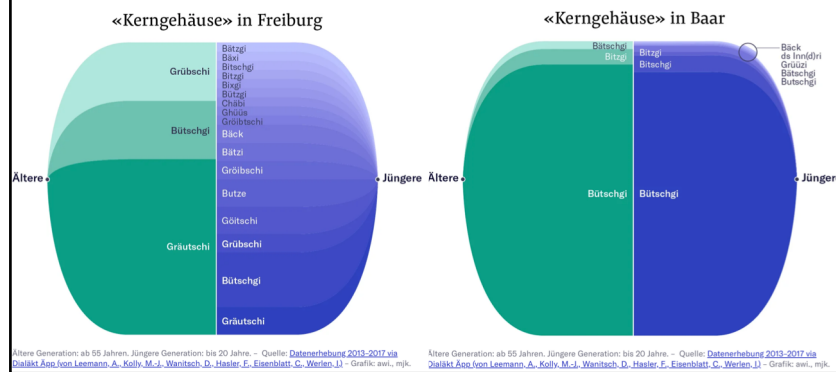
Pick extrema *pars pro toto*



Visualize what differentiates them



Work on design



Do reporter work

Dialekt Hochdeutsch

Nicolas: D Grandmama siit gwüssi Sache, won i nid säge. Zum Bispüu «Patinoire». Däm sägen i «Ilschbaan».

Mutter Valentine: Ja, u Zopf – wiisch u, wie si däm siit?

Nicolas: «Züpfä»?

Valentine: Nei. Aber «Züpfä» isch o as Wort für Zopf. Das siit de Papi. Aber wie säge d Grandmama und i däm?

Nicolas (ruft): «Trütscha»!

Explain linguistic change via protagonists

weiteren Grund dafür zu geben, dass sich die Baarer Generationen nur so wenig unterscheiden. Sie sprechen ähnlich wie die Menschen im Nachbarkanton Zürich. Man versteht sie überall.

Dialekt Hochdeutsch

Tabea: Aber alli, wo so extreemi Dialäkt händ, Walliser oder Bündner, händ mee s Gfüül, si müend sich aapasse. Wenn s anderswo gönd go studiere oder schaffe, dänn sind si immer «anders». Und mir händ immer s Gfüül, mir siget «normaler», und so passt mer sich weniger aa.

Sylvia: Und dur das ich da au d Veränderig weniger starch.

Tabea: Au wil mer öis überall verstaat.

Rosemarie: Also innerhalb vo de Schwilz sett mer enand no verstaat. Usert ebe villecht Freiburger oder so ...

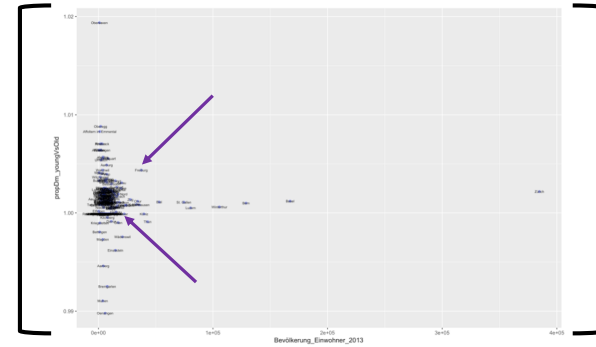
Die Freiburger dagegen fühlen sich weniger gut verstanden:

Dialekt Hochdeutsch

Elisabeth: Üsa Dialäkt isch spezieu, as verstöö na nid ali. U du muesch de würllich a huufe Energie u Chraft haa, dass du de dezue stüsch, zu dim Dialäkt.

Valentine: Wäge de eint oder ander lachet di de haut tüfah uus.

This story is still data-driven!



Dialog zwischen dem Generationen-Nikolas, 7 Jahre und Elisabeth, 60 Jahre, in Freiburg. (Bilder: Anlick Ramp / NZZ)

Wo junge Schweizer noch Grossmamis Dialekt sprechen

Bei den Jungen vermischen sich die Dialekte, heisst es. Verstehen Schweizer Kinder ihre Grosseltern bald nicht mehr? In Freiburg ist das teilweise der Fall, in Baar hingegen hat sich die Sprache kaum verändert. Eine Spurensuche.



Thank you for your attention!

Questions?

Contributors to the article on weapons:
Authors: Marie-José Kolly, Anna Wiederkehr, Michael Surber, Andreas Rüesch, Daniel Steinvorth, Andres Wysling, Patrick Zoll
Participation: David Bauer, tools: Beni Buess, Manuel Roth, Philip Küng.

Contributors to the article on dialects:
Authors: Marie-José Kolly, Alexandra Kohler, Stefanie Hasler (Video), Anna Wiederkehr (Grafik), Annick Ramp (Fotografie).
Participation: David Bauer (Konzept), Markus Stein (Video), Beni Buess und Philip Küng (interaktive Elemente).

Help and feedback by many others.