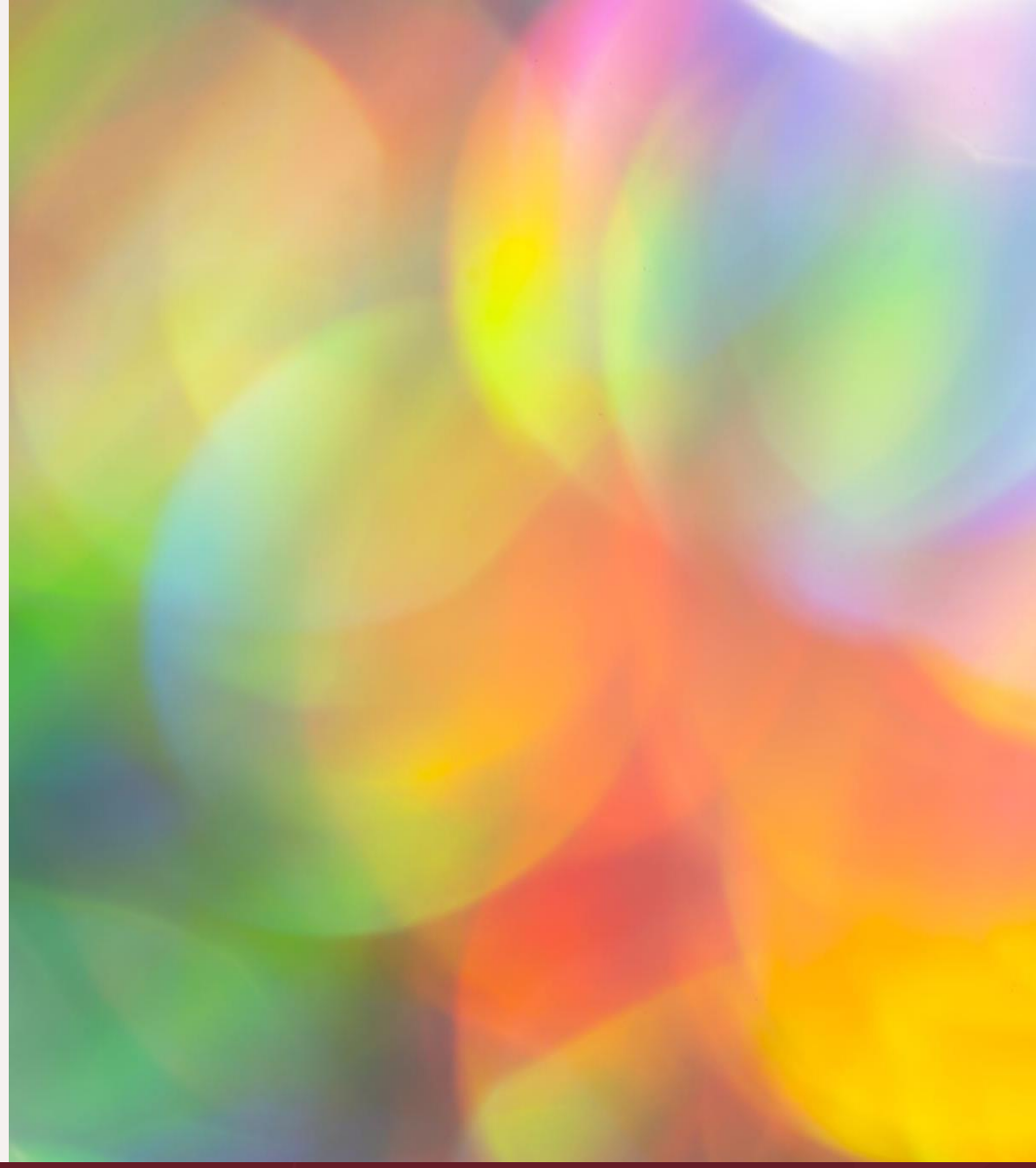
Python interdisziplinär

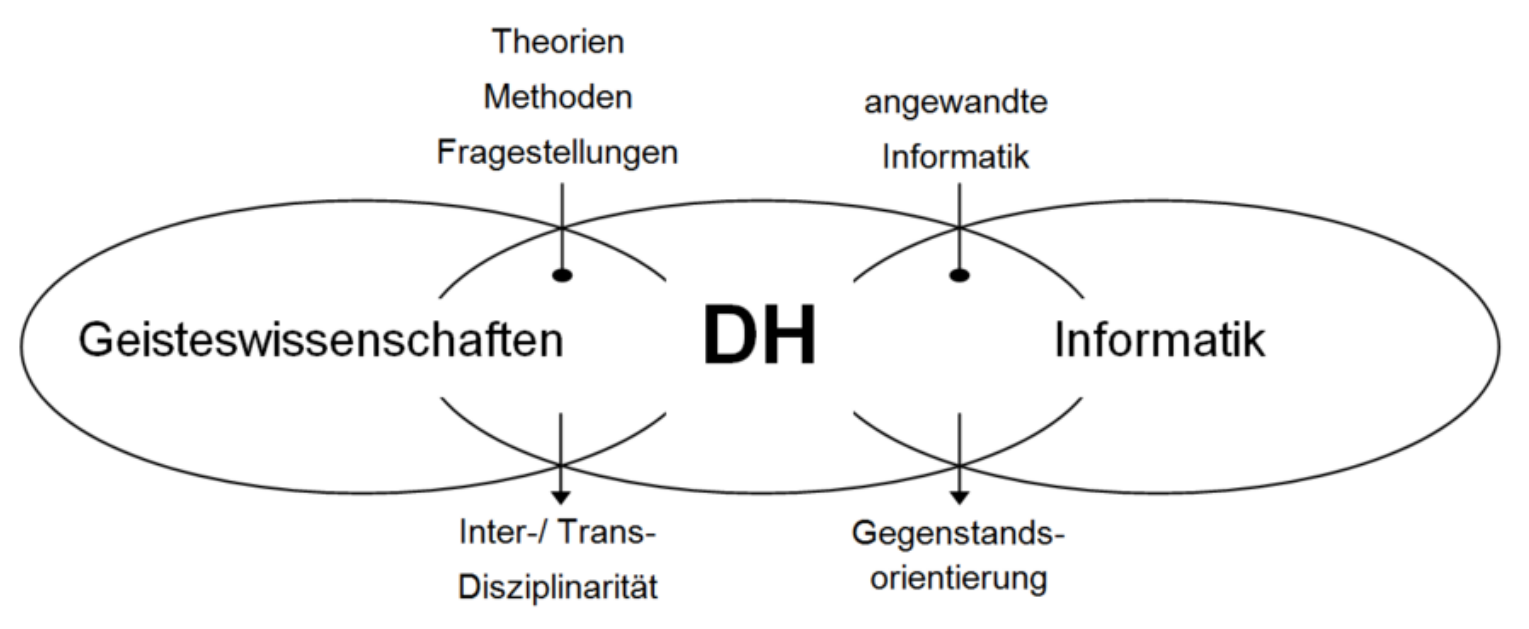
Ein Anwendungsfall zwischen Geoinformatik und Geschichtswissenschaften



Maximilian Hege, B.A.
Python-Barcamp Rügen
Lietzow, 22.09.2024

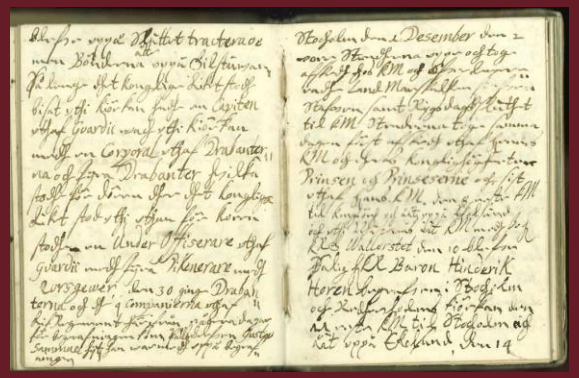
Digital Humanities? Was ist das?

- ganz kurz: die interdisziplinäre Verwendung von Methoden aus der Informatik in den Geisteswissenschaften (in meinem Fall Geschichte)



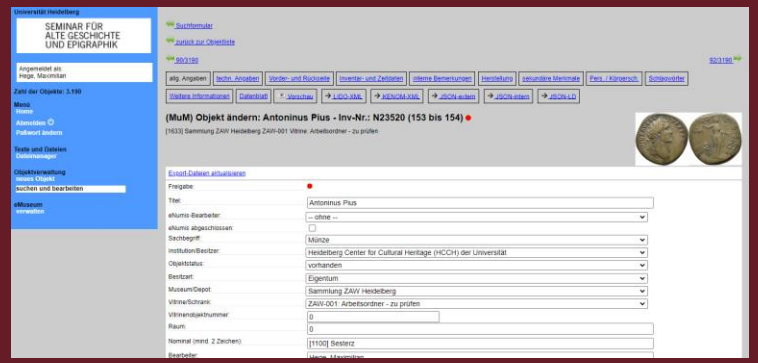
https://www.geschichte.uni-wuppertal.de/de/lehrgebiete/digital-humanities/

Felder der Digital History (Auswahl)



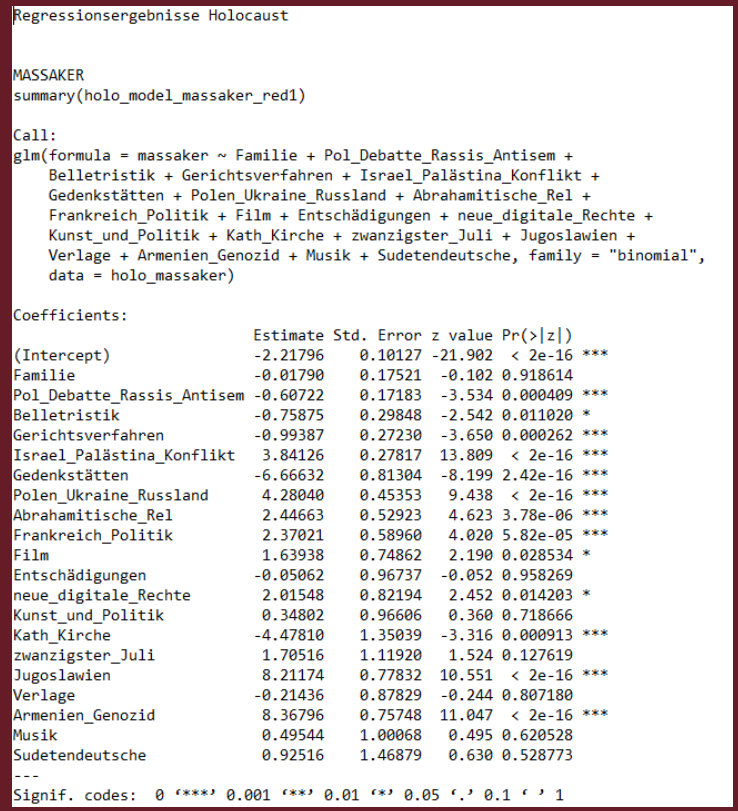
Transkription alter Handschriften

https://www.compgen.de/wp-content/uploads/2019/09/8176120393_64e2626ed9_b.jpg



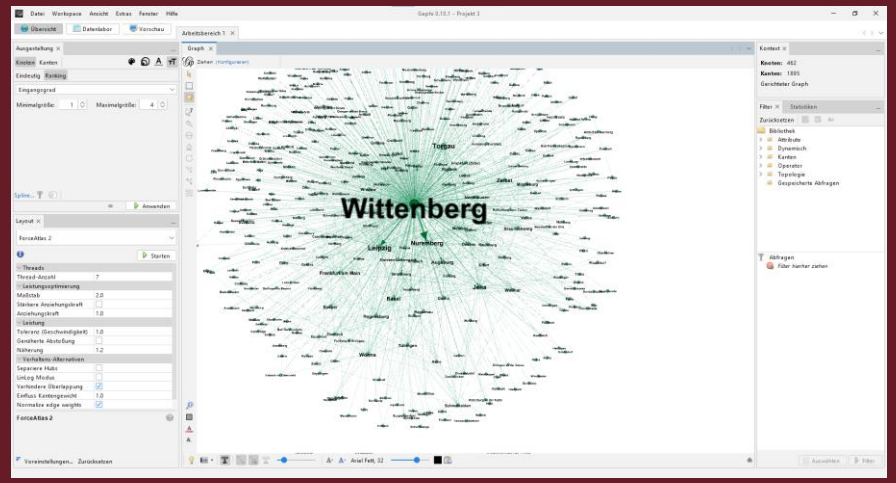
Sammeldatenbanken

https://pecunia.zaw.uni-heidelberg.de/ikmk/home?lang=de



```
Regressionsergebnisse Holocaust


MASSAKER
summary(holo_model_massaker_red1)

Call:
glm(formula = massaker ~ Familie + Pol_Debatte_Rassis_Antisem +
    Belletristik + Gerichtsverfahren + Israel_Palästina_Konflikt +
    Gedenkstätten + Polen_Ukraine_Russland + Abrahamitische_Rel +
    Frankreich_Politik + Film + Entschädigungen + neue_digitale_Rechte +
    Kunst_und_Politik + Kath_Kirche + zwanzigster_Juli + Jugoslawien +
    Verlage + Armenien_Genozid + Musik + Sudetendeutsche, family = "binomial",
    data = holo_massaker)

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -2.21796    0.10127 -21.902  < 2e-16 ***
Familie                   -0.01790    0.17521  -0.102 0.918614
Pol_Debatte_Rassis_Antisem -0.60722    0.17183  -3.534 0.000409 ***
Belletristik              -0.75875    0.29848  -2.542 0.011020 *
Gerichtsverfahren         -0.99387    0.27230  -3.650 0.000262 ***
Israel_Palästina_Konflikt  3.84126    0.27817  13.809  < 2e-16 ***
Gedenkstätten             -6.66632    0.81304  -8.199 2.42e-16 ***
Polen_Ukraine_Russland     4.28040    0.45353   9.438  < 2e-16 ***
Abrahamitische_Rel         2.44663    0.52923   4.623 3.78e-06 ***
Frankreich_Politik         2.37021    0.58960   4.020 5.82e-05 ***
Film                       1.63938    0.74862   2.190 0.028534 *
Entschädigungen           -0.05062    0.96737  -0.052 0.958269
neue_digitale_Rechte       2.01548    0.82194   2.452 0.014203 *
Kunst_und_Politik          0.34802    0.96606   0.360 0.718666
Kath_Kirche               -4.47810    1.35039  -3.316 0.000913 ***
zwanzigster_Juli           1.70516    1.11920   1.524 0.127619
Jugoslawien                8.21174    0.77832  10.551  < 2e-16 ***
Verlage                   -0.21436    0.87829  -0.244 0.807180
Armenien_Genozid           8.36796    0.75748  11.047  < 2e-16 ***
Musik                      0.49544    1.00068   0.495 0.620528
Sudetendeutsche            0.92516    1.46879   0.630 0.528773
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

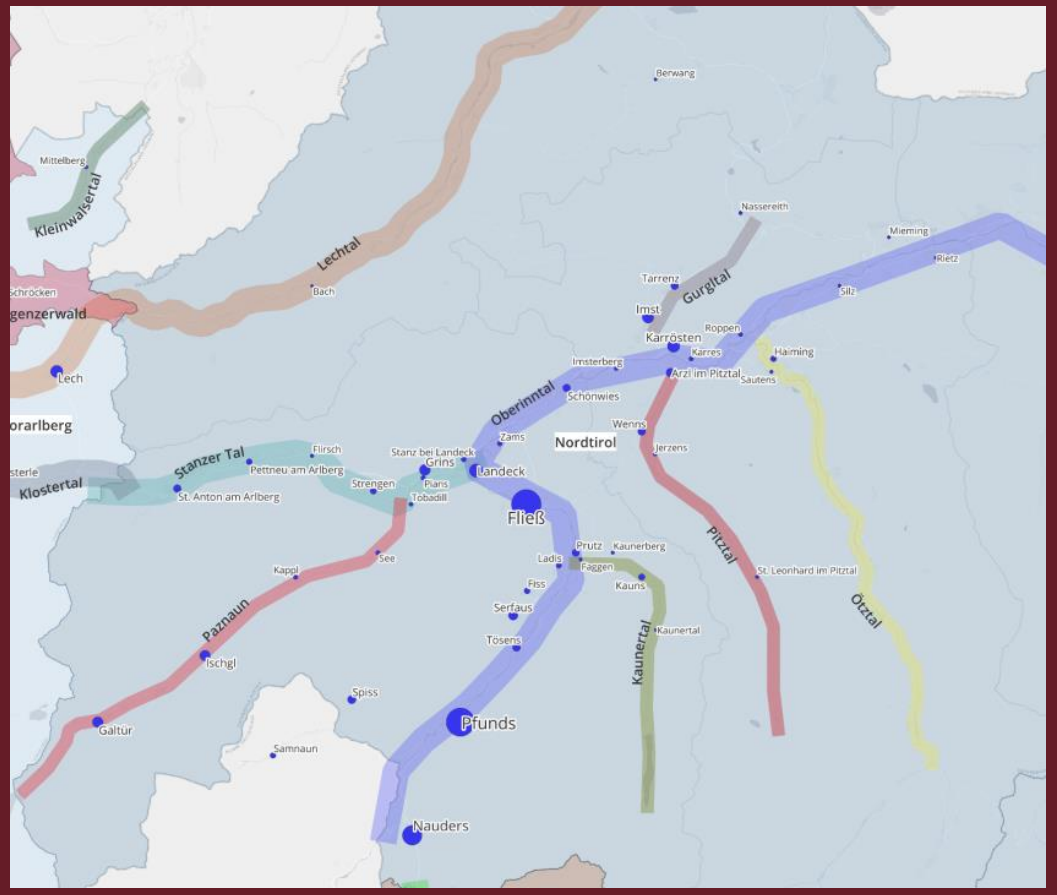
Topic Modeling



Netzwerkanalyse

Anwendungsfall: meine Bachelorarbeit

Die Sozialtopographie der Schwabenkinder. Eine GIS-gestützte Untersuchung jugendlicher Arbeitsmigration von Vorarlberg und Tirol nach Oberschwaben 1812-1938



Was sind Schwabenkinder?

- Schwabenkinder sind eine Migrationsbewegung von Kindern zwischen ~1600 bis ~1945
- Sie gingen vom mittleren Alpenraum nach Oberschwaben im Frühjahr, um sich dort als Hirte oder Dienstbote zu verdingen

https://www.schwabenkinder.eu/media/filer_thumbnails/2012/02/20/karlstrasse_fn1.jpg__800x0_q85.jpg

Die Schwabenkinder-Datenbank

- https://www.schwabenkinder.eu/de/Datenbank/datenbank-suche/
- Enthält 7.040 Schwabengänge von insgesamt 5.425 Kindern
- Basiert auf Dienstbotenverzeichnissen im LK Bodensee und RV
- Problem: keine Arbeit mit Normdaten → keine Auswertung der Daten

Josef Andreas Baldauf
Diensteintrittsalter: 10 Jahre

* 09.05.1886 † nicht bekannt, Junge aus Dalaas, Vorarlberg

DIENSTZEITEN UND DIENSTHERREN	
26.03.1897 bis 30.10.1897	im Alter von 10 Jahren bei Johann Georg Ehinger Dienstbezeichnung: Hirte Dienstort: Praßbergmühle Gemeinde Leupolz
14.08.1899 bis 29.10.1899	im Alter von 13 Jahren bei Josef Geyer Dienstbezeichnung: Hirte Dienstort: Weiler Gemeinde Leupolz
25.03.1900 bis 27.10.1900	im Alter von 13 Jahren bei Andreas Wild Dienstbezeichnung: Hirte Dienstort: Weiler Gemeinde Leupolz
VATER	Franz Josef Baldauf Ökonom
MUTTER	Creszentia Muther
GESCHWISTER	▸ *Maria Benedikta Baldauf*
BEFREIUNG VON DER SCHULE	nicht bekannt nicht bekannt nicht bekannt
QUELLE	OA Leupolz, B 41

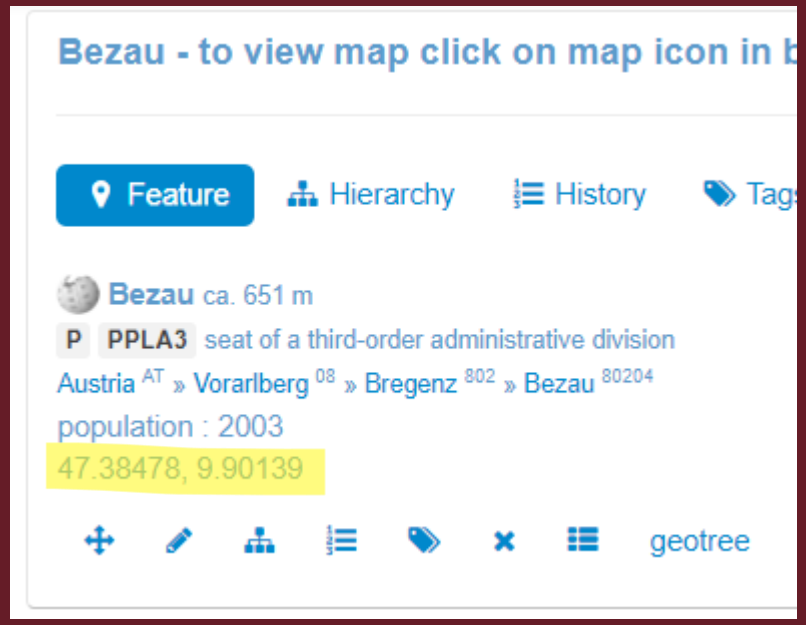
◂ zur Übersicht

Schritt 1: Scraping

- Daten aus der DB scrapen und in eine Excel-Tabelle abspeichern

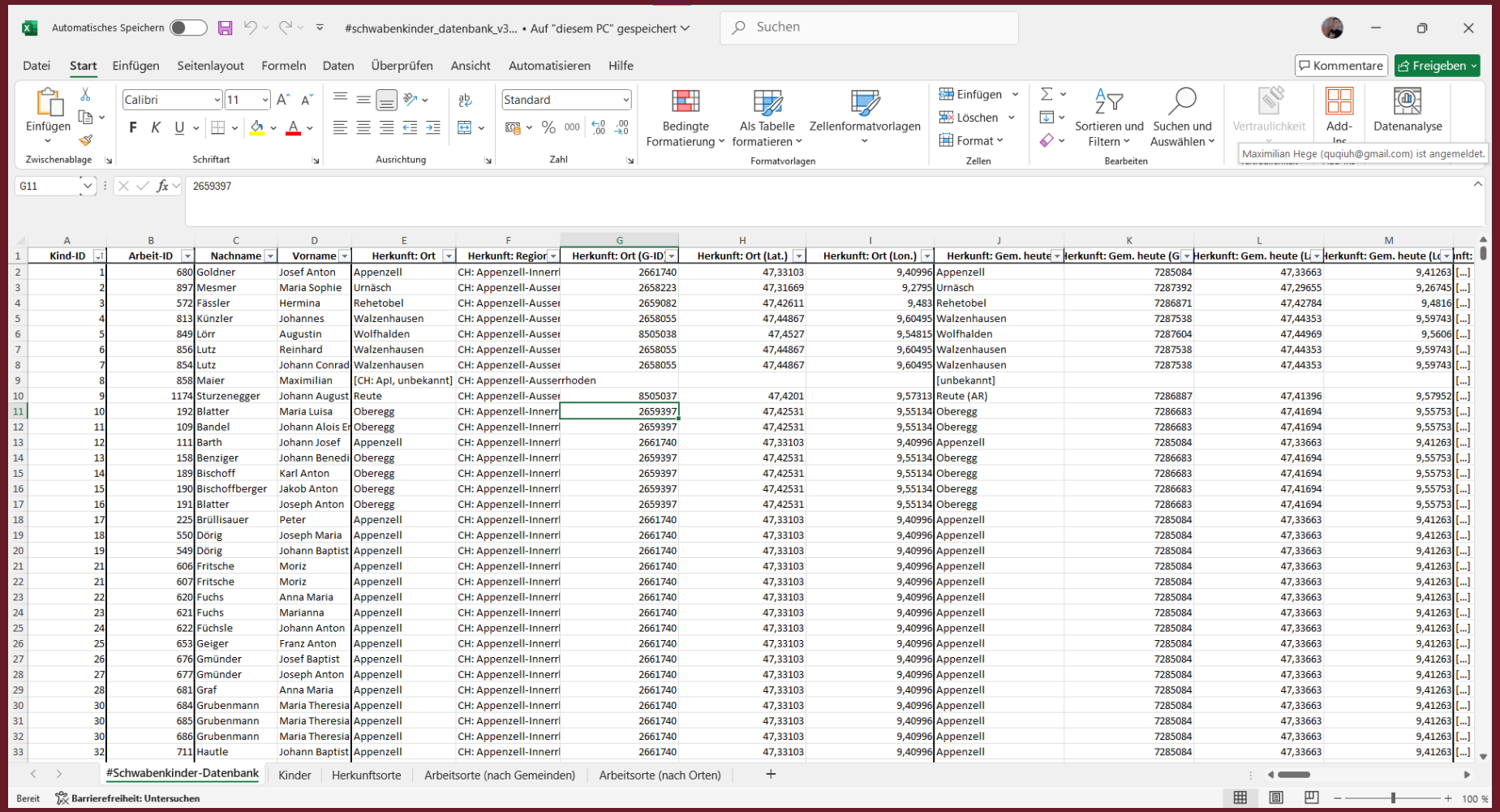
```python
import requests
import os
import pandas as pd

# Schritt 1: Herunterladen der XLS-Dateien
download_folder = "D:\schwabenkinder"
url_template = "https://www.schwabenkinder.eu/de/Datenbank/datenbank-suche/export/xls/{}"

num_files = 7000  # Anzahl der herunterzuladenden Dateien
for i in range(1, num_files + 1):
    url = url_template.format(i)
    filename = os.path.join(download_folder, f"datei_{i}.xls")
    response = requests.get(url)
    if response.status_code == 200:
        with open(filename, "wb") as file:
            file.write(response.content)
        print(f"Datei {i} heruntergeladen.")
    else:
        print(f"Fehler beim Herunterladen der Datei {i}.")

# Schritt 2: Kombinieren der Daten
# Ordnerpfad mit den XLS-Dateien
folder_path = "D:\schwabenkinder"

# Kombinierte Daten initialisieren
combined_data = pd.DataFrame()

# Durch alle Dateien im Ordner iterieren
for filename in os.listdir(folder_path):
    if filename.endswith(".xls") or filename.endswith(".xlsx"):
        file_path = os.path.join(folder_path, filename)
        try:
            df = pd.read_excel(file_path)
            # Dateinamen als neue Spalte hinzufügen
            df.insert(0, "Dateiname", filename)
            combined_data = pd.concat([combined_data, df])
            print(f"Datei {filename} kombiniert.")
        except Exception as e:
            print(f"Fehler beim Kombinieren der Daten aus Datei {filename}: {str(e)}")

# Kombinierte Daten in eine neue Excel-Datei speichern
combined_data.to_excel("kombinierte_daten.xlsx", index=False)
print("Alle Daten kombiniert und in kombinierte_daten.xlsx gespeichert.")
```

Schritt 2: Geoparsing der Ortsnamen (mithilfe der GeoNames-API)




```python
def query_geonames_api(place):

    # Überprüfen, ob Ort bereits abgefragt wurde
    if place in queried_places:
        print(f'API-Anfrage für {place} übersprungen (bereits abgefragt)')
        return None

    endpoint = 'http://api.geonames.org/searchJSON'
    api_key = 'xeilian'

    # API-Anfrage senden
    anfrage = endpoint + "?q=" + place + "&MaxRows=1&featureClass=P&username=" + api_key
    print(anfrage)
    response = requests.get(anfrage)
    print(f'API-Anfrage für {place} gesendet')

    # Ergebnis verarbeiten
    if response.status_code == 200:
        data = response.json()
        if 'geonames' in data and len(data['geonames']) > 0:
            result = data['geonames'][0]
            return {
                'adminname1': result.get('adminName1', ''),
                'lat': result.get('lat', ''),
                'lng': result.get('lng', ''),
                'population': result.get('population', ''),
                'toponym_name': result.get('toponymName', '')
            }

    return None
```

Schritt 3: Datenbereinigung



Kind-ID	Arbeit-ID	Nachname	Vorname	Herkunft: Ort	Herkunft: Region	Herkunft: Ort (G-ID)	Herkunft: Ort (Lat.)	Herkunft: Ort (Lon.)	Herkunft: Gem. heute	Herkunft: Gem. heute (G	Herkunft: Gem. heute (L	Herkunft: Gem. heute (Lo	nft:
1	680	Goldner	Josef Anton	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
2	897	Mesmer	Maria Sophie	Urnäsch	CH: Appenzell-Ausser	2658223	47,31669	9,2795	Urnäsch	7287392	47,29655	9,26745	[...]
3	572	Fässler	Hermina	Rehetobel	CH: Appenzell-Ausser	2659082	47,42611	9,483	Rehetobel	7286871	47,42784	9,4816	[...]
4	813	Künzler	Johannes	Walzenhausen	CH: Appenzell-Ausser	2658055	47,44867	9,60495	Walzenhausen	7287538	47,44353	9,59743	[...]
5	849	Lörr	Augustin	Wolfhalden	CH: Appenzell-Ausser	8505038	47,4527	9,54815	Wolfhalden	7287604	47,44969	9,5606	[...]
6	856	Lutz	Reinhard	Walzenhausen	CH: Appenzell-Ausser	2658055	47,44867	9,60495	Walzenhausen	7287538	47,44353	9,59743	[...]
7	854	Lutz	Johann Conrad	Walzenhausen	CH: Appenzell-Ausser	2658055	47,44867	9,60495	Walzenhausen	7287538	47,44353	9,59743	[...]
8	858	Maier	Maximilian	[CH: API, unbekannt]	CH: Appenzell-Ausserrhoden				[unbekannt]				[...]
9	1174	Sturzenegger	Johann August	Reute	CH: Appenzell-Ausser	8505037	47,4201	9,57313	Reute (AR)	7286887	47,41396	9,57952	[...]
10	192	Blatter	Maria Luisa	Oberegg	CH: Appenzell-Innerr	2659397	47,42531	9,55134	Oberegg	7286683	47,41694	9,55753	[...]
11	109	Bandel	Johann Alois Er	Oberegg	CH: Appenzell-Innerr	2659397	47,42531	9,55134	Oberegg	7286683	47,41694	9,55753	[...]
12	111	Barth	Johann Josef	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
13	158	Benziger	Johann Benedi	Oberegg	CH: Appenzell-Innerr	2659397	47,42531	9,55134	Oberegg	7286683	47,41694	9,55753	[...]
14	189	Bischoff	Karl Anton	Oberegg	CH: Appenzell-Innerr	2659397	47,42531	9,55134	Oberegg	7286683	47,41694	9,55753	[...]
15	190	Bischoffberger	Jakob Anton	Oberegg	CH: Appenzell-Innerr	2659397	47,42531	9,55134	Oberegg	7286683	47,41694	9,55753	[...]
16	191	Blatter	Joseph Anton	Oberegg	CH: Appenzell-Innerr	2659397	47,42531	9,55134	Oberegg	7286683	47,41694	9,55753	[...]
17	225	Brüllisauer	Peter	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
18	550	Dörig	Joseph Maria	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
19	549	Dörig	Johann Baptist	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
21	606	Fritsche	Moriz	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
21	607	Fritsche	Moriz	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
22	620	Fuchs	Anna Maria	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
23	621	Fuchs	Marianna	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
24	622	Füchsle	Johann Anton	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
25	653	Geiger	Franz Anton	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
26	676	Gmünder	Josef Baptist	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
27	677	Gmünder	Joseph Anton	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
28	681	Graf	Anna Maria	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
30	684	Grubenmann	Maria Theresia	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
30	685	Grubenmann	Maria Theresia	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
30	686	Grubenmann	Maria Theresia	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]
32	711	Hautle	Johann Baptist	Appenzell	CH: Appenzell-Innerr	2661740	47,33103	9,40996	Appenzell	7285084	47,33663	9,41263	[...]

Schritt 4: Auswertung

A	B	C
Herkunft: Ort	Anzahl: Kinder	Anzahl: Dienste
[AT: VB, unbekannt]	4	4
[CH: ApI, unbekannt]	1	1
[CH: Grb, unbekannt]	3	3
[CH: Stg, unbekannt]	6	7
[Dial]	1	1
[IT: Süd, unbekannt]	7	7
[St. Ulrich]	1	1
[Wolfach]	1	1
Agums	1	1
Alberschwende	69	93
Almens	5	6
Altach	5	5
Altdorf	1	1
Altenstadt	47	49
Altstätten	32	36
Alvaneu	9	9
Alvaschein	4	4
Andelsbuch	117	177
Andiast	7	10
Angedair	16	23
Appenzell	60	76
Arzl	22	40
Au (SG)	1	1
Au (VB)	131	171
Ausserbraz	1	1
Azmoos-Wartau	1	1
Bach	1	1
Balzers	7	7
Bangs	1	1
Bartholomäberg	27	30
Batschuns	1	1

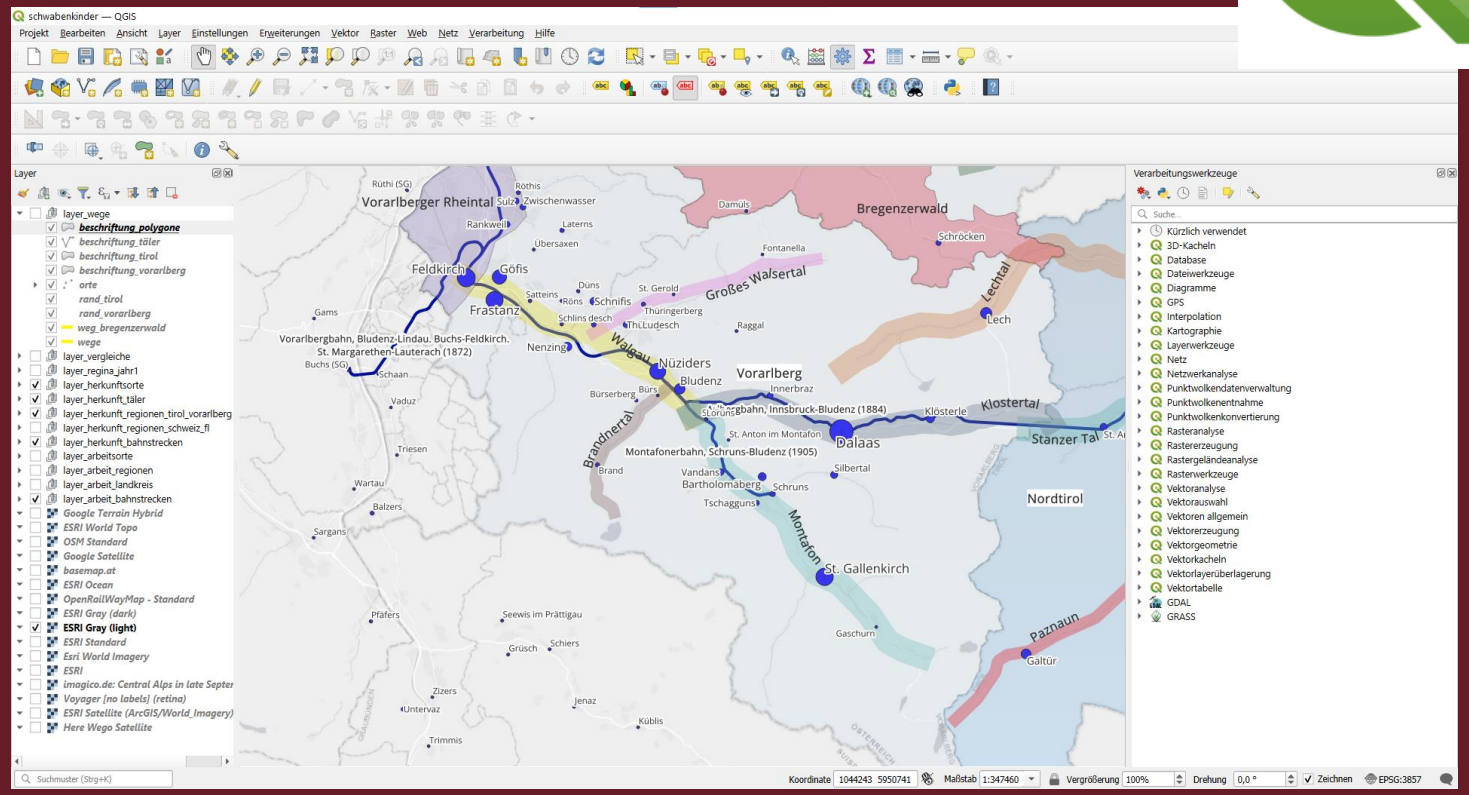
Heimatorte

A	B	C	D	E	F
Arbeit: Gem. alt	VB	Tirol	Südtirol/Trentino	Dienste VB/Tirol Ges.	Dienste DB Ges.
[Allmannshofen?]	1	0	0	1	1
[Bayern?]	1	0	0	1	1
[Heglbach?]	1	0	0	1	1
[Schwaben?]	895	0	0	895	895
[unbekannt]	398	1	1	400	402
[Württemberg?]	24	0	0	24	24
~Kißlegg~	7	0	0	7	15
~Wangen im Allgäu~	10	0	0	10	10
Achberg	34	6	2	42	46
Ailingen	54	67	36	157	229
Altmannshofen	65	13	2	80	92
Amtzell	67	53	10	130	132
Arnach	2	2	4	8	8
Aulendorf	0	5	3	8	14
Berg	0	21	12	33	34
Beuren	6	0	0	6	7
Blitzenreute	17	42	17	76	106
Bodnegg	1	27	3	31	32
Deggenhausen	1	1	2	4	5
Deuchelried	110	6	0	116	118
Diepoldshofen	41	57	2	100	200
Dietmanns	2	2	2	6	6
Ebersbach-Musbach	0	1	0	1	3
Eglofs	161	2	1	164	196
Eichstegen	0	10	3	13	13
Eisenharz	15	0	0	15	26
Emmelhofen	175	17	3	195	347
Eriskirch	1	107	24	132	140
Eschach	24	63	26	113	150
Esenhausen	2	47	6	55	61
Frickingen	0	1	1	2	2

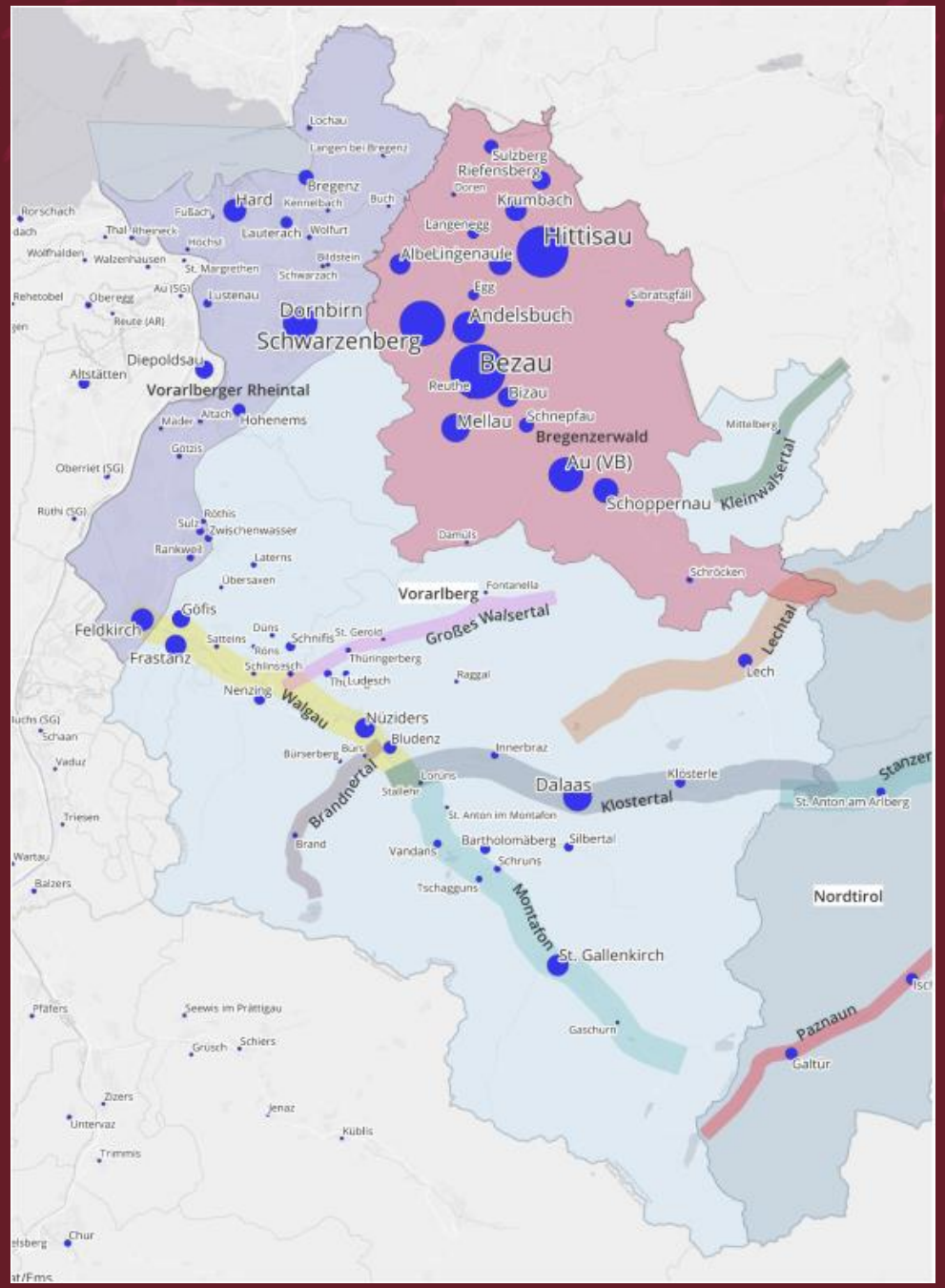
Arbeitsorte (in Abhängigkeit zu der Herkunftsregion)

Schritt 5: GIS (endlich ☺)

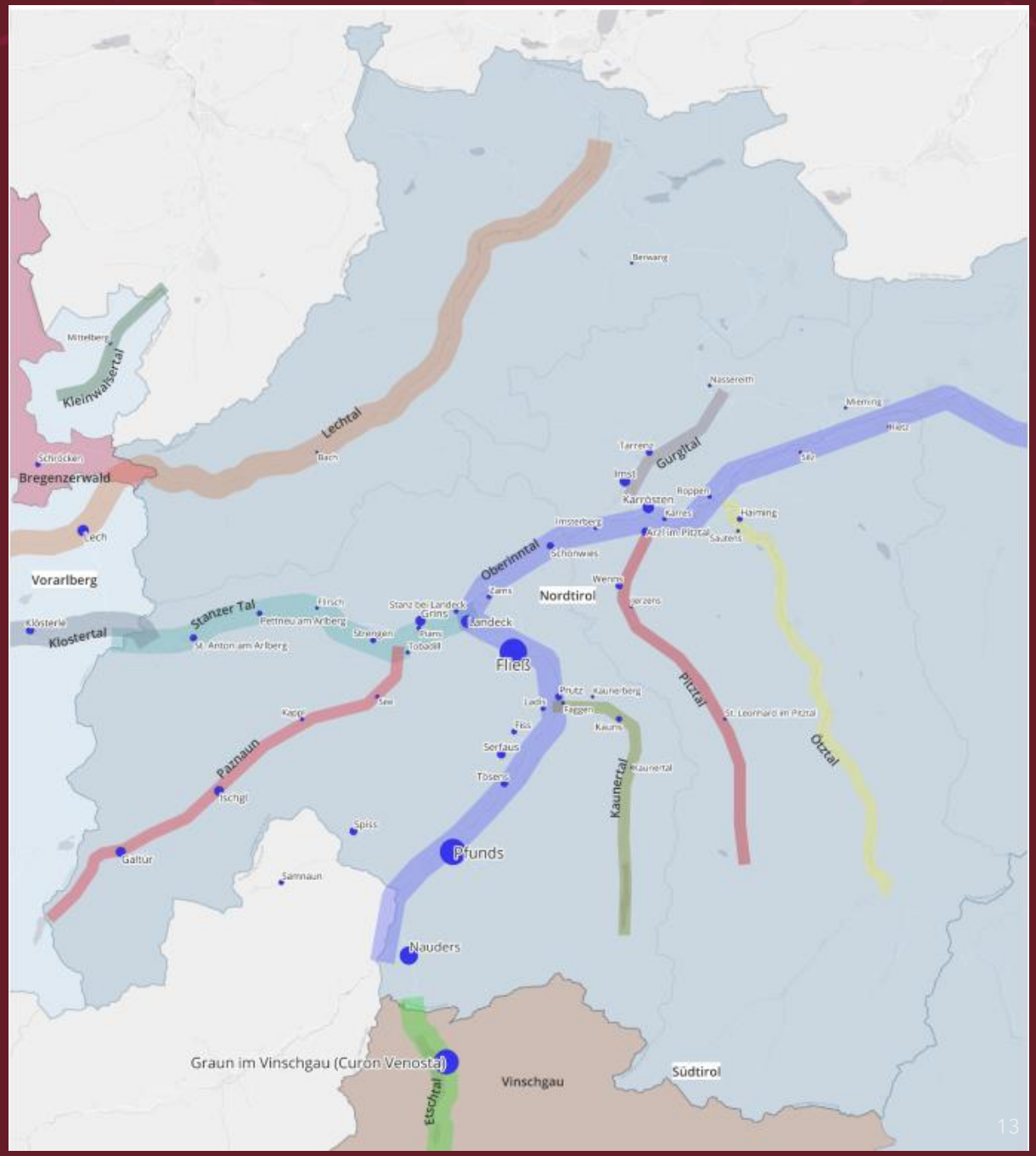
- GIS = **G**eo**I**nformation**S**ysteme



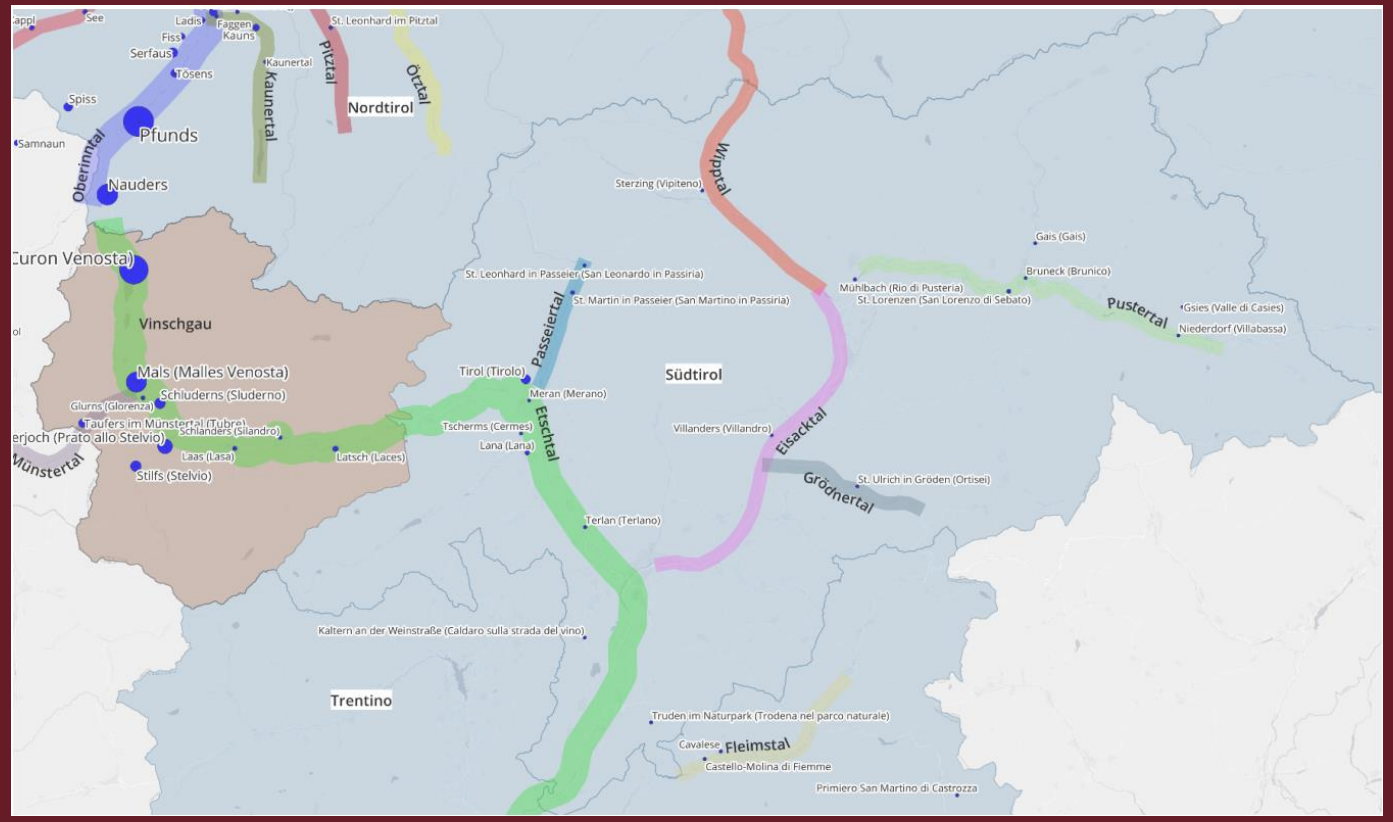
Herkunftsorte
(Vorarlberg)

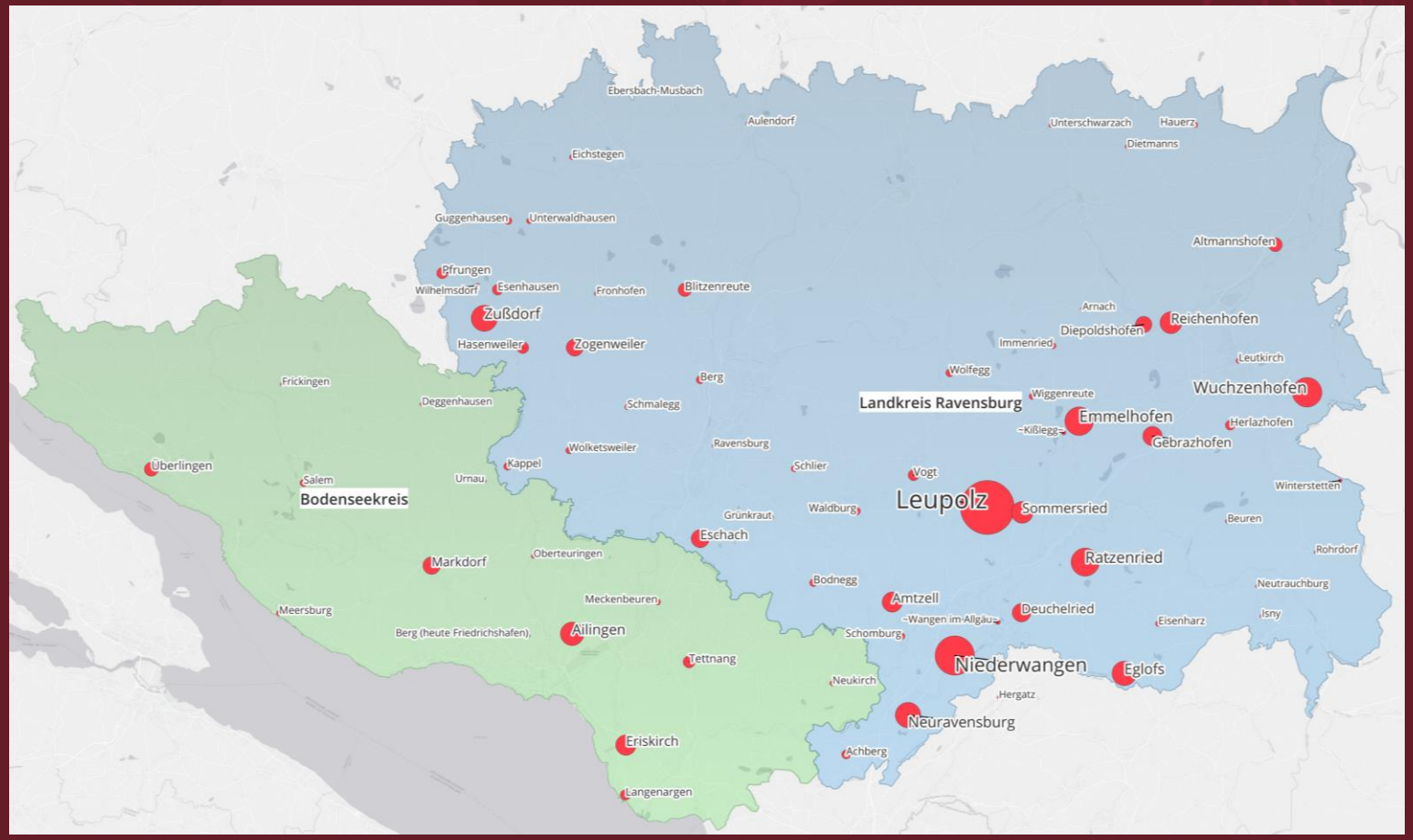


Herkunftsorte
(Nordtirol)



Herkunftsorte
(Südtirol)





Arbeitsorte (nach heutigen Landkreisen)

Arbeitsorte (nach alten Verwaltungs-grenzen)

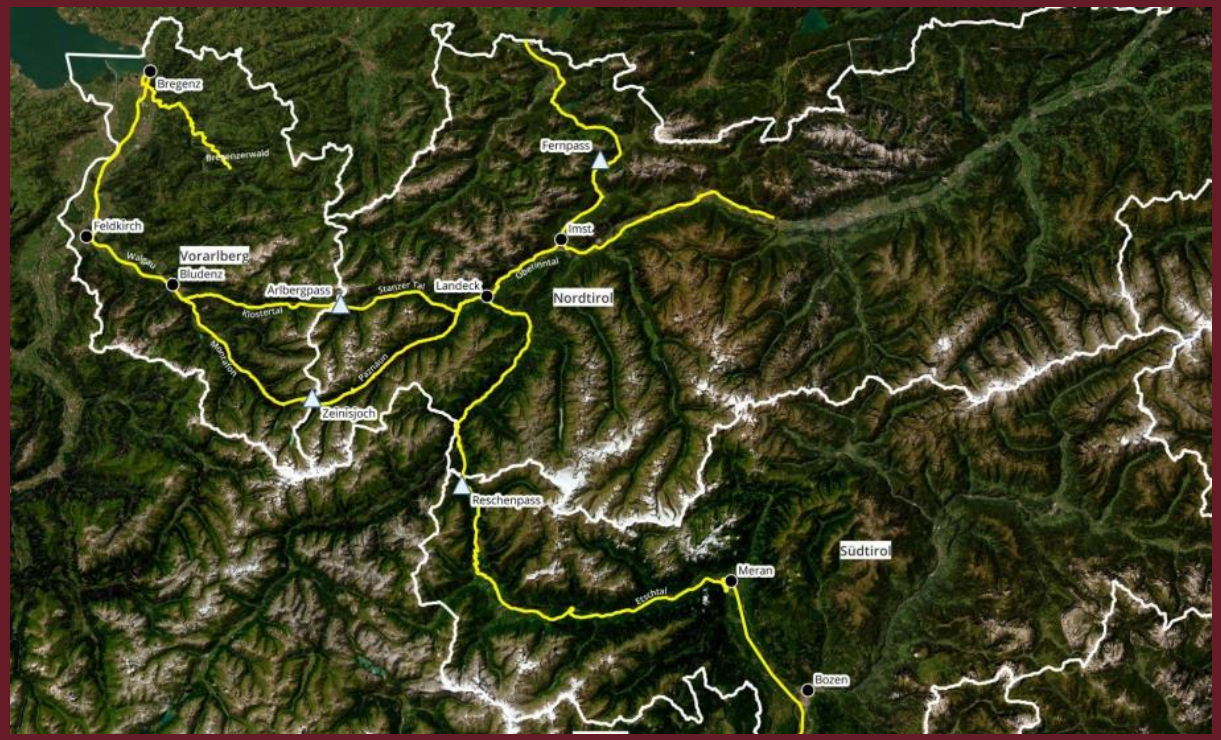
Aber warum? Pushfaktoren

- Dichte Bevölkerung v.a. in VB, weniger in Nordtirol und Südtirol
- Topographie: Vorarlberger Kinder haben eine kurze Strecke. Nordtiroler müssen den Arlberg, Südtiroler auch den Reschenpass überqueren
- Kulturelle Gemeinsamkeiten zw. Vorarlberg und Oberschwaben
- Starke Zersiedelung durch Realteilung: Besitz wird auf alle Kinder verteilt + hohe Geburtenrate führt zu Zersiedelung und Verarmung durch Aufteilung des Vermögens

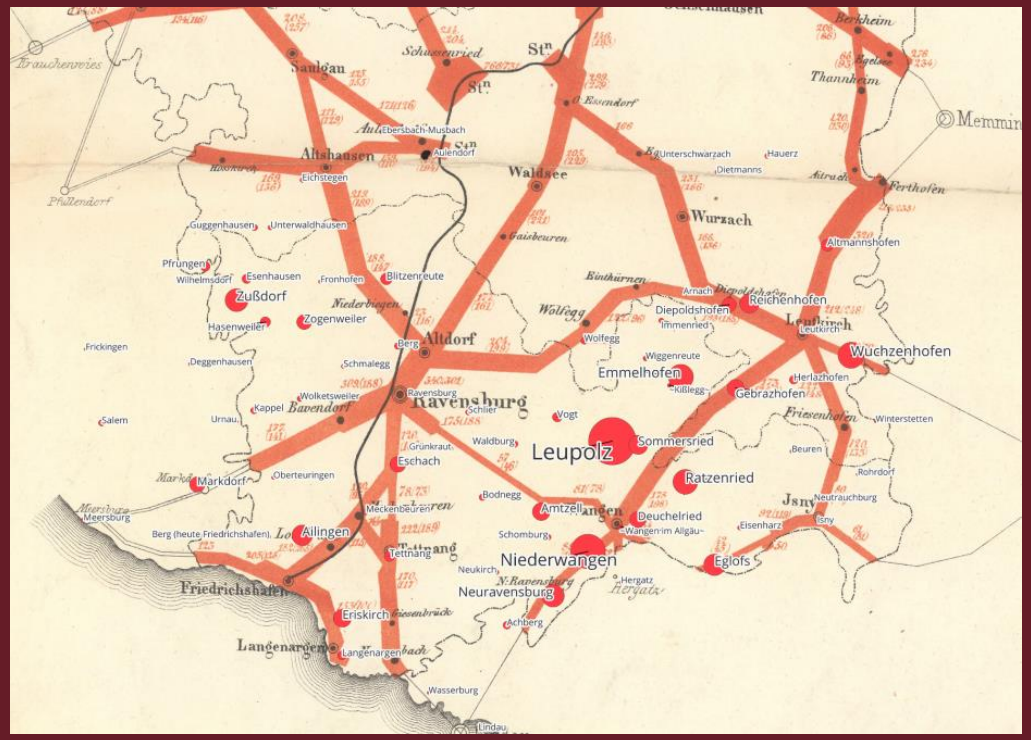
Aber warum? Pullfaktoren

- Oberschwaben hat auch eine hohe Bevölkerungsdichte, jedoch verteilt sie sich viel besser durch die Hügellandschaft (in VB/Tirol Verteilung auf wenige Täler wegen Hochgebirge)
- Hoher agrarischer Bedarf
- Ebenfalls hohe Zersiedelung, ABER: Anerbenrecht: der Erstgeborene bekommt alles, dadurch bleibt das Vermögen an einem Ort konzentriert → relativ reiche bäuerliche Oberschicht

Wege und Infrastruktur: zu Fuß oder per Fuhrwerk



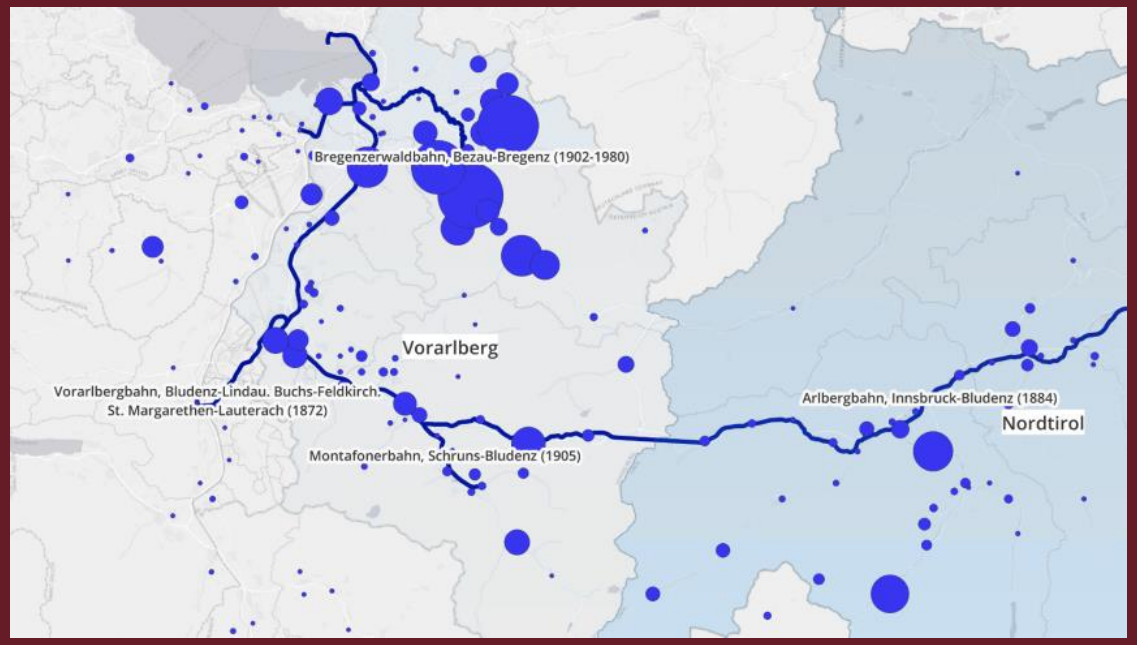
Wanderwege bis zum Bodensee



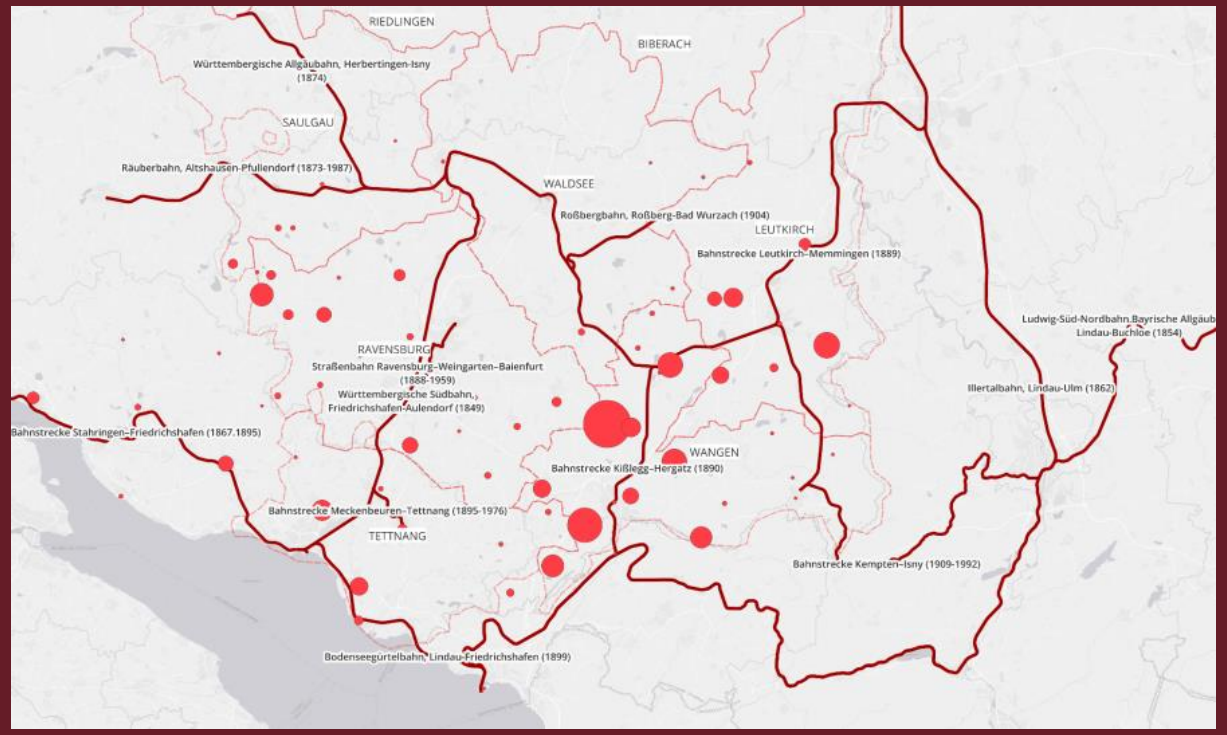
Straßennetz in Oberschwaben

Wege und Infrastruktur: per Zug

- Ab den 1850er-Jahren: Aufbau eines Eisenbahnnetzes in den Herkunfts- und Arbeitsorten

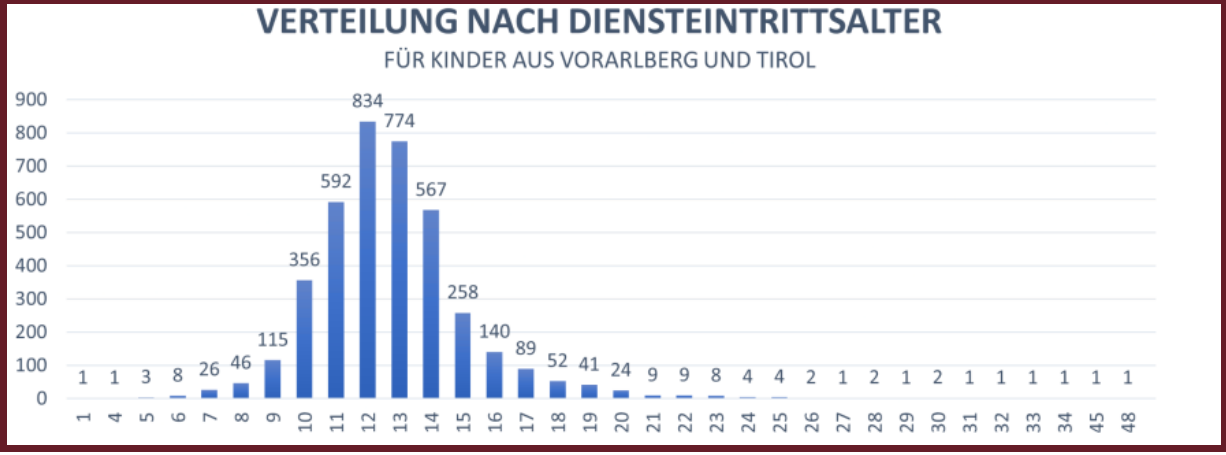


Bahnlinien in den Herkunftsregionen

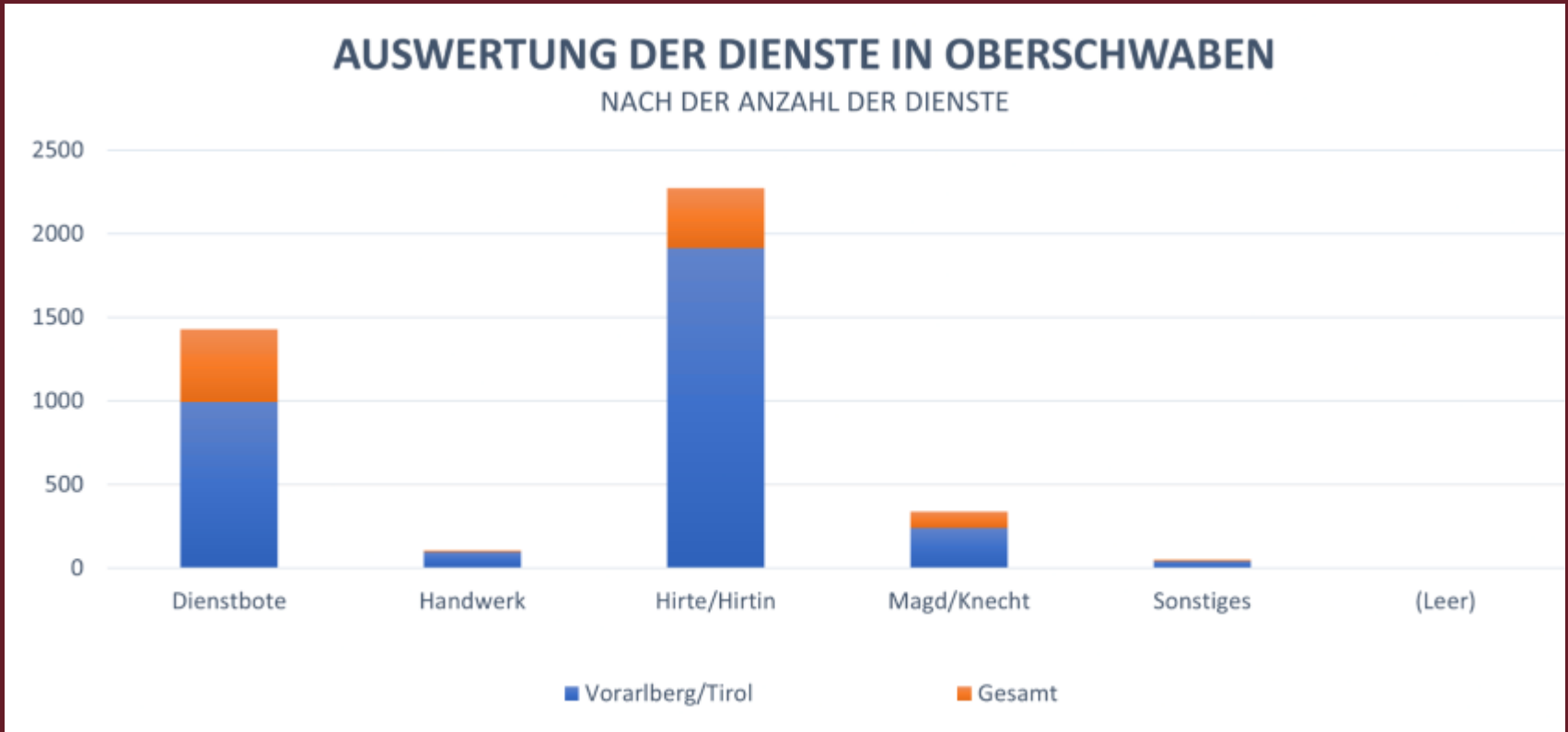


Bahnlinien in den Arbeitsregionen

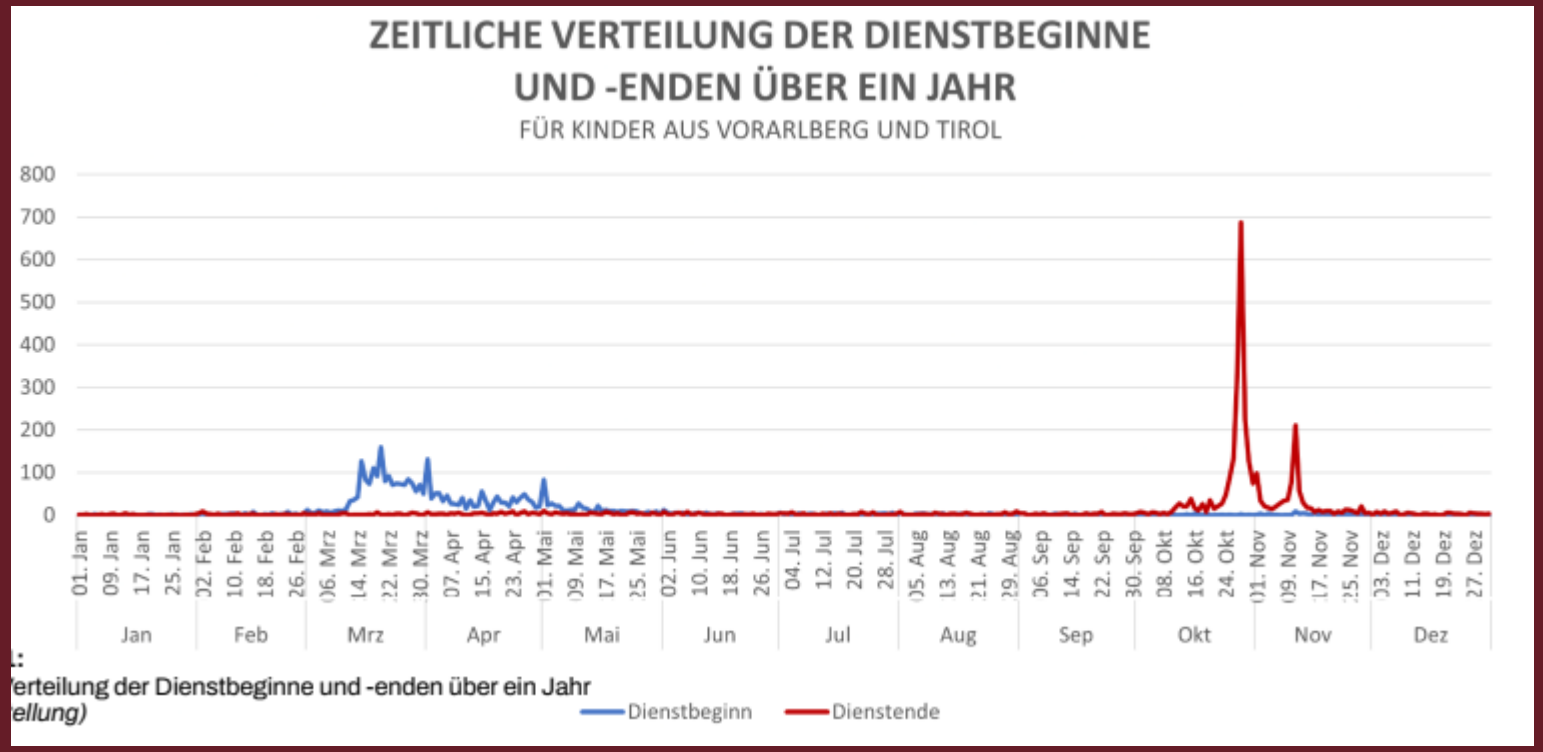
Statistische Erhebungen



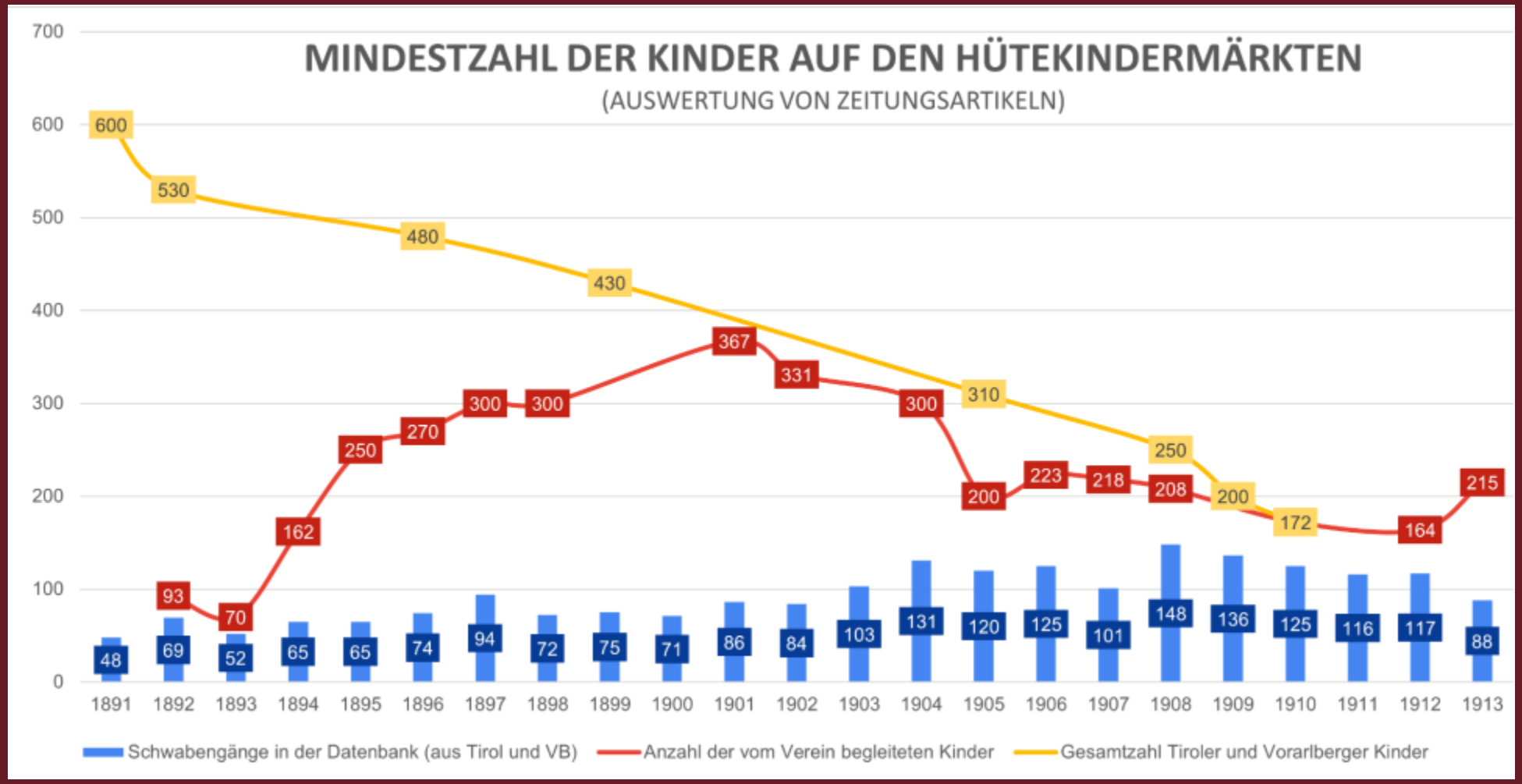
Statistische Erhebungen



Statistische Erhebungen



Statistische Erhebungen



https://github.com/xeilian/schwabenkinder

Vielen Dank für Eure Aufmerksamkeit!