

Berlin Children Health & Development Dashboard

— Cem Kozcuer

Presented at:

2022-01-27

Urban Technology - Prof. Dr. Bießmann
Berliner Hochschule für Technik

Bezirke



LOR-PR

(Lebensweltlich orientierte
Prognoseräume)



Initial idea:

Dashboard with district related census
data about Berlin

Problem: too big districts

example:

Tempelhof-Schöneberg :(

vs.


Tempelhof, Schöneberg-Nord, Schöneberg-Süd :)

2nd idea:

Dashboard with district and health and social service related data about Berlin

Problem:

**too many datasets, inaccessible data
portals**



no url parameter for GSI, too laborious to do manually

formativ System

Aktualität ▼ Titel


Seite 1 ▼ von 241 ▶

Leistungen gemäß
2019/


3

Metadaten

~ 16,000 documents


 **cemkozcu**er added links to webpages where the lists got fetched

4c8957c 20 days ago 4 commits

 dataset_lists

initial commit

20 days ago

 README.md

added links to webpages where the lists got fetched

20 days ago

 README.md



Berlin open data dataset lists

This repository contains CSV files that are lists of the datasets available on the [Berlin open data webpage](#) and the [webpage of the Gesundheits- und Sozialinformationssystem \(GSI\)](#).

The lists contain

- titles
- descriptions (if available)
- downloading urls

of the datasets.

The purpose of this repository is to provide an alternative overview and access to the actual datasets in contrast to the aforementioned data portals.

Please note that the websites of [Berlin open data](#) and the [GSI](#) may change with regard to the maintained datasets and their URLs.

For information regarding the datasets, their metadata and legal use please refer to the aforementioned websites.

Current dataset lists

- [Berlin Open Data - "Kategorie: Sozialleistungen" \(517 datasets\) <- fetched here](#)
- [Berlin Open Data - "Kategorie: Gesundheit" \(508 datasets\) <- fetched here](#)
- [GIS - "Kategorie: Gesundheitsdaten" \(6200 datasets\) <- fetched here](#)
- [GIS - "Kategorie: Sozialdaten" \(11792 datasets\) <- fetched here](#)

Please note that the dataset lists were created and maintained to the best of our knowledge, however do not claim to be complete or error free.

	title	description	url
0	COVID-19 in Berlin, Fallzahlen und Indikatoren...	Tabelleninhalt: COVID-19 E...	https://daten.berlin...
1	COVID-19 in Berlin, Verteilung in den Bezirken...	Tabelleninhalt: die an das...	https://daten.berlin...
2	COVID-19 in Berlin, 7-Tage-Inzidenz nach Alter...	Tabelleninhalt: COVID-19 E...	https://daten.berlin...

https://github.com/cemkozcu/berlin-opendata-dataset-lists

What health and social service related data can I use?

Requirements:

Fine-grained districts

Geodata available

Time series of at least about 10 years



Children Health and Development Dashboard (LOR-PR)

5 Datasets about health and development of children when they start primary school:

- Children with language-deficits
- Children that went longer than 2 years to Kita
- Children that have an own tv at home
- Children that went to U8
- Children that have visuo-motoric disorder

Dataset for 2008

	A	B	C	D	E	F	G	H	I
1	Tabelle A.10:								
2	Anteil der Kinder mit auffälligen Befunden in der Visuomotorik bei der Einschulungsuntersuchung in Berlin 2008 nach Prognoserräumen (LOR-PR)								Methodik
3									
4	Bezirk	Prognoseraum		Anteil	Bezirk	Prognoseraum		Anteil	
5		Nr.	Bezeichnung	in %		Nr.	Bezeichnung	in %	
6	Mitte	0101	Zentrum	12,0	Tempelhof-Schöneberg	0701	Schöneberg-Nord	14,1	
7		0102	Moabit	16,8		0702	Schöneberg-Süd	17,4	
8		0103	Gesundbrunnen	29,7		0703	Friedenau	10,7	
9		0104	Wedding	25,6		0704	Tempelhof	12,8	
10	Friedrichshain-Kreuzberg	0201	Kreuzberg Nord	23,9		0705	Mariendorf	12,4	
11		0202	Kreuzberg Süd	14,8		0706	Marienfelde	22,6	

row 6

Getting the data to work

Problem:

everything in excel :(

Inconsistent (manual) records

Dataset for 2010

	A	B	C	D	E	F	G	H	I
1	Tabelle A.12:								
2	Anteil der Kinder mit auffälligen Testergebnissen in der Visuomotorik bei der Einschulungsuntersuchung in Berlin 2010 nach Prognoserräumen (LOR-PR)								
3									
4									
5									
6	Bezirk	Prognoseraum		Anteil	Bezirk	Prognoseraum		Anteil	
7		Nr.	Bezeichnung	in % ¹		Nr.	Bezeichnung	in % ¹	
8	Mitte	0101	Zentrum	6,7	Tempelhof-Schöneberg	0701	Schöneberg-Nord	10,9	
9		0102	Moabit	16,6		0702	Schöneberg-Süd	5,8	
10		0103	Gesundbrunnen	30,5		0703	Friedenau	6,0	
11		0104	Wedding	24,5		0704	Tempelhof	16,0	
12	Friedrichshain-Kreuzberg	0201	Kreuzberg Nord	23,6		0705	Mariendorf	13,4	
13		0202	Kreuzberg Süd	15,7		0706	Marienfelde	15,0	

row 9

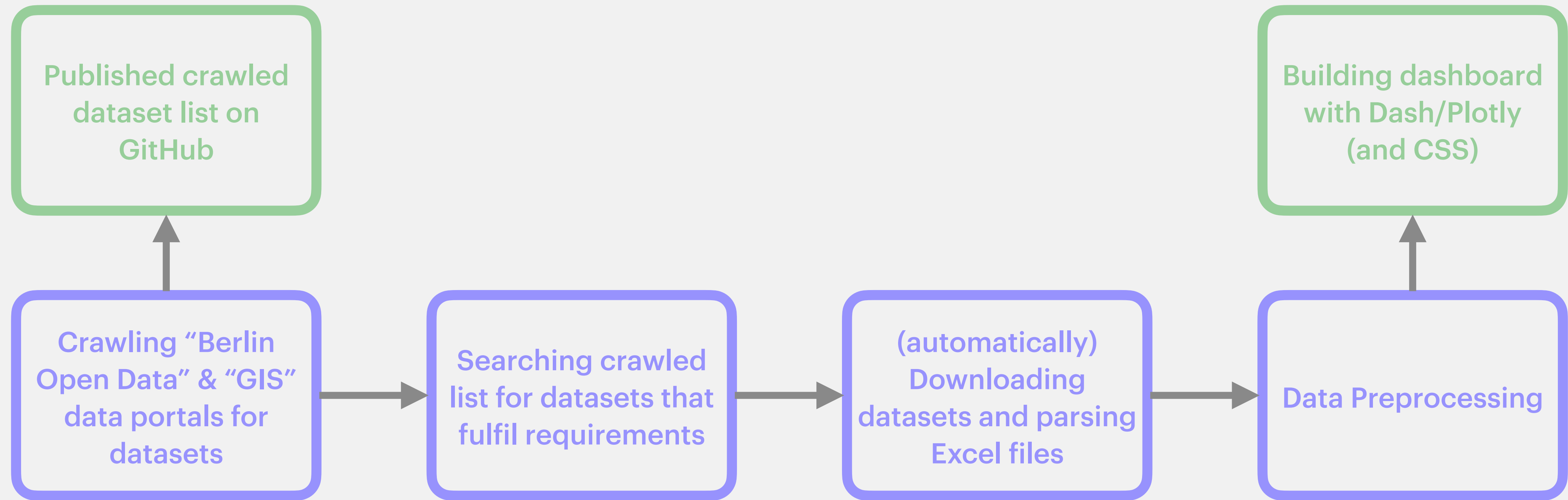
Solution:

Sanity check function

Manual definitions

Preprocessing Data:

- some missing data for some districts -> filled with 0
- merging all years -> outer join
- 0 filling for dimension reduction bad -> replaced with median



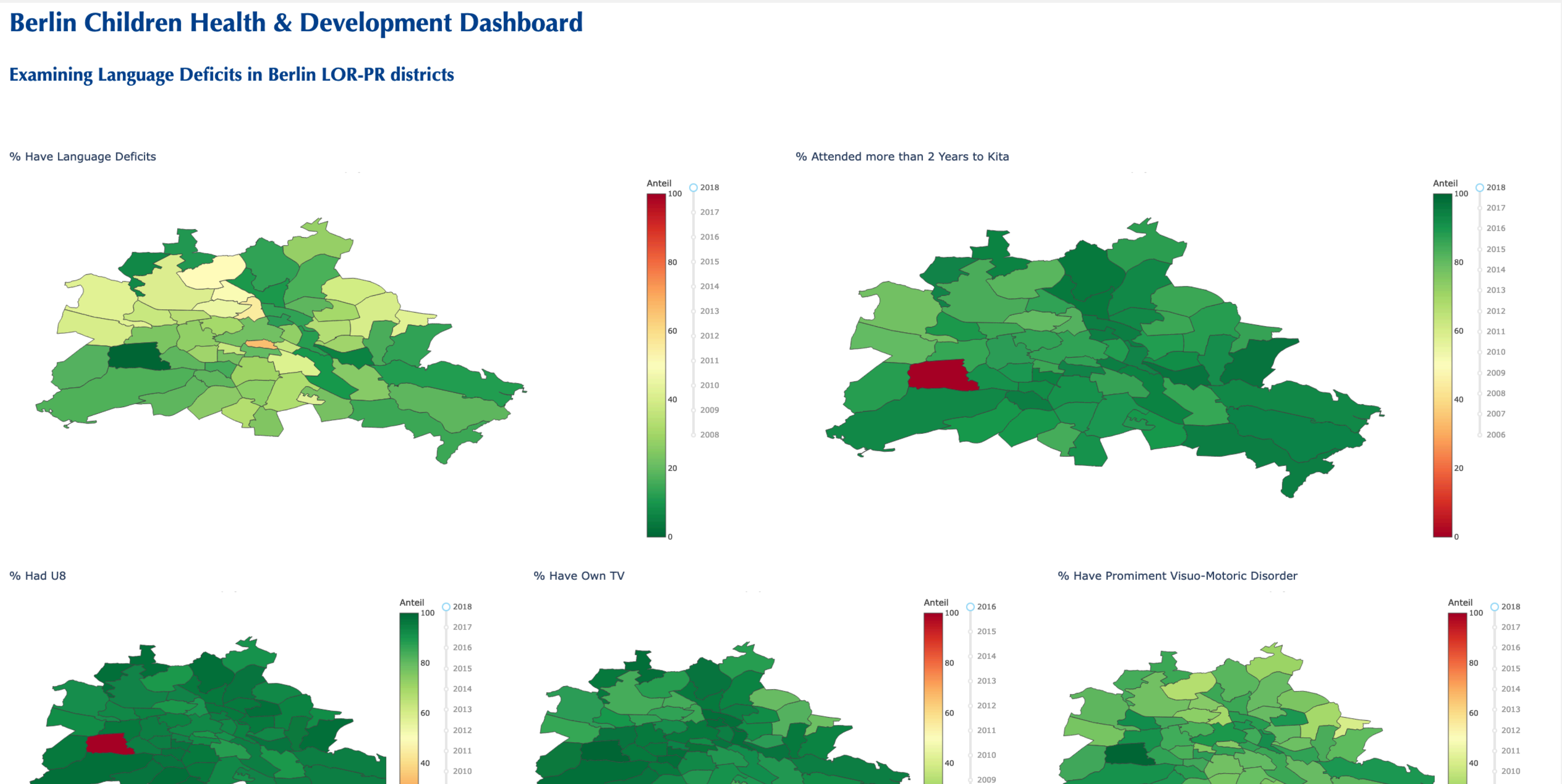
Tech stack:

Pandas

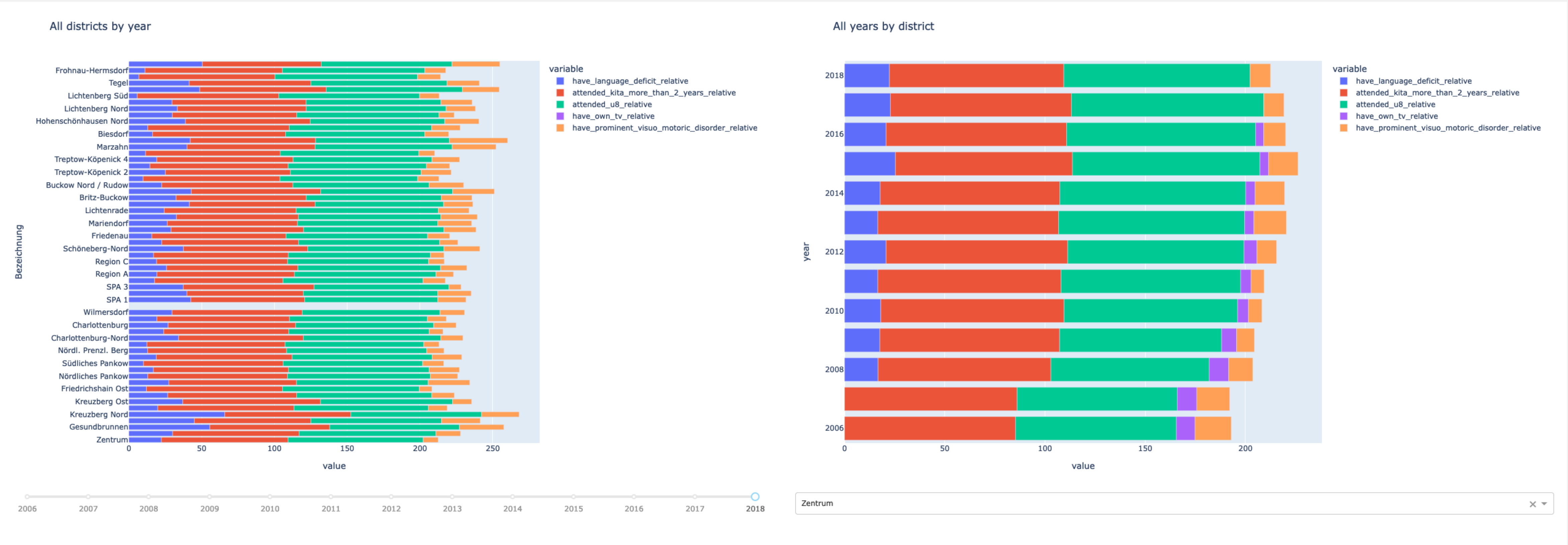
Beatifulsoup

Dash/Plotly

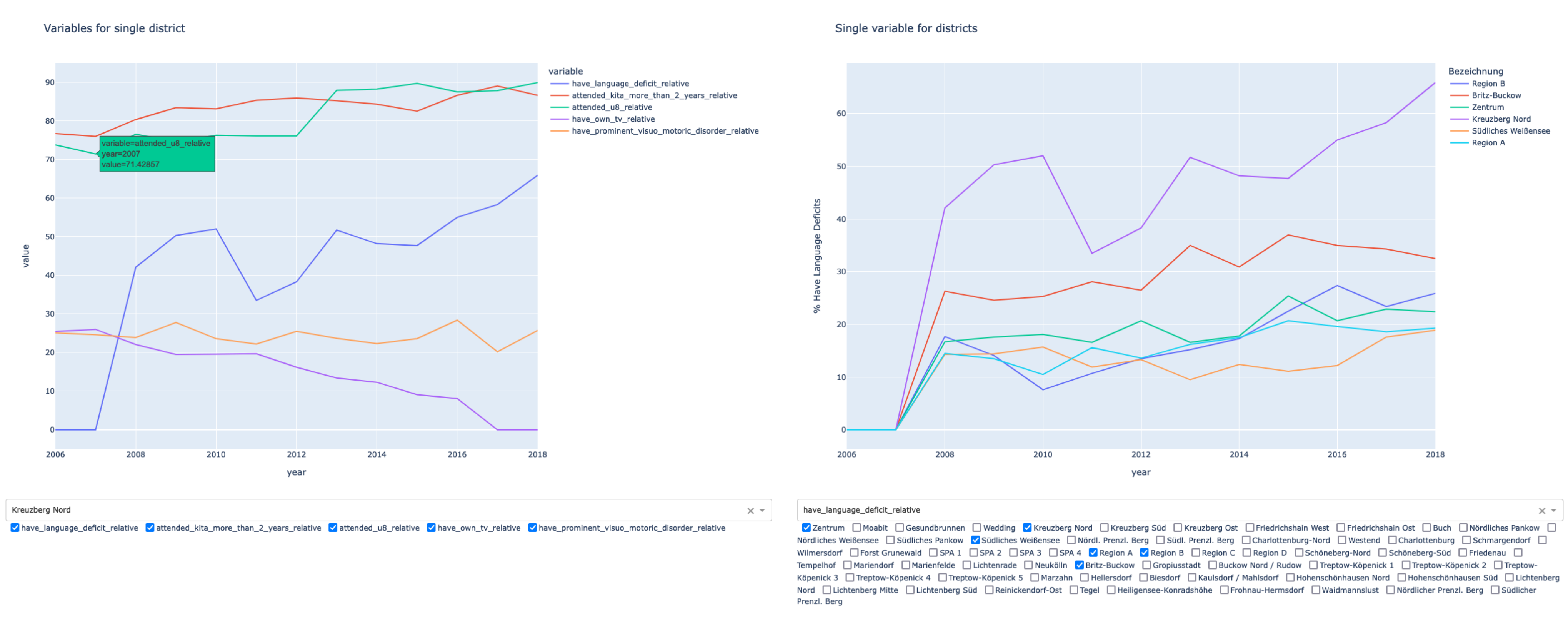
Scikitlearn



Actual Dashboard

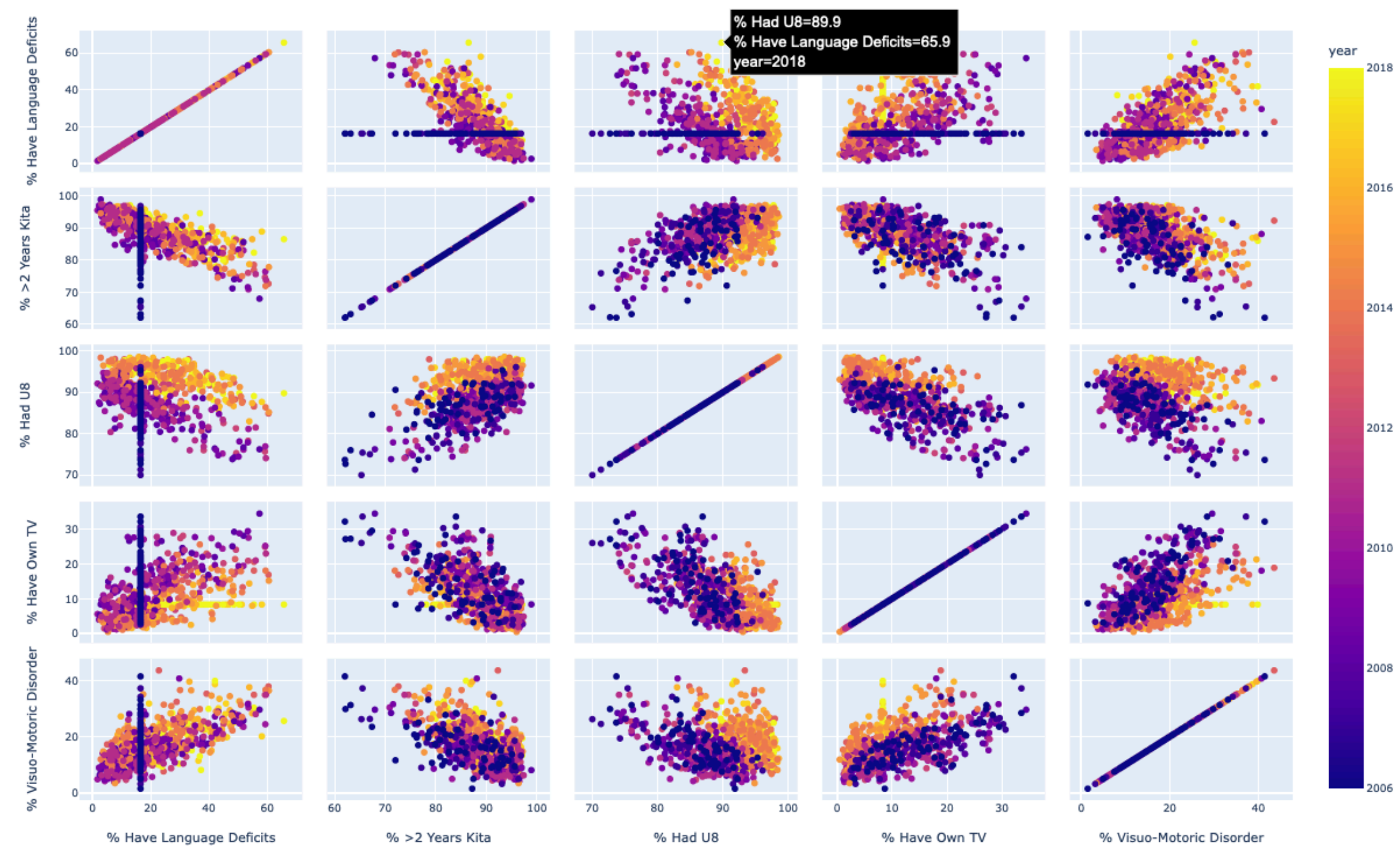


Actual Dashboard

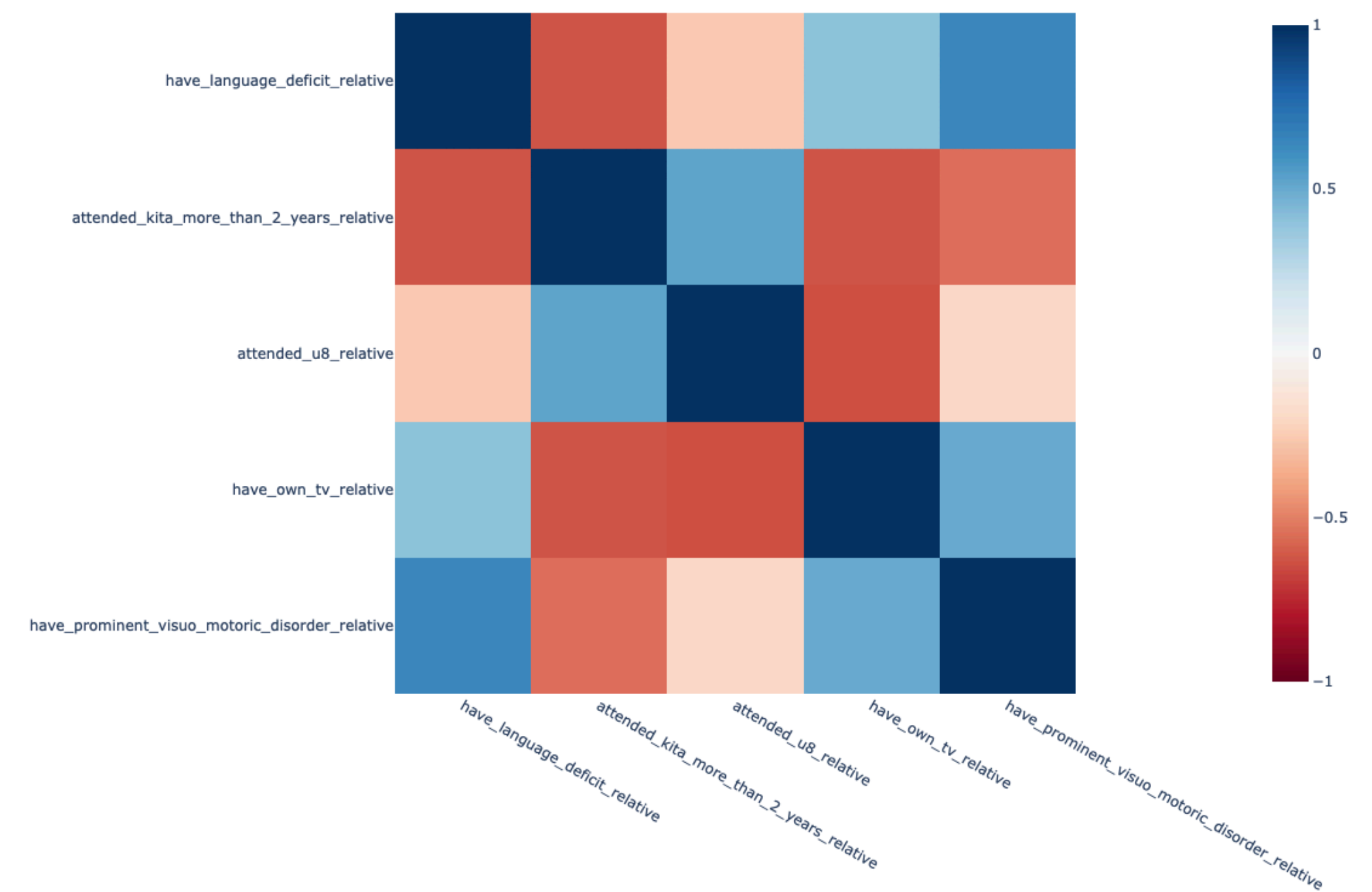


Actual Dashboard

Scatter Matrix of Variables

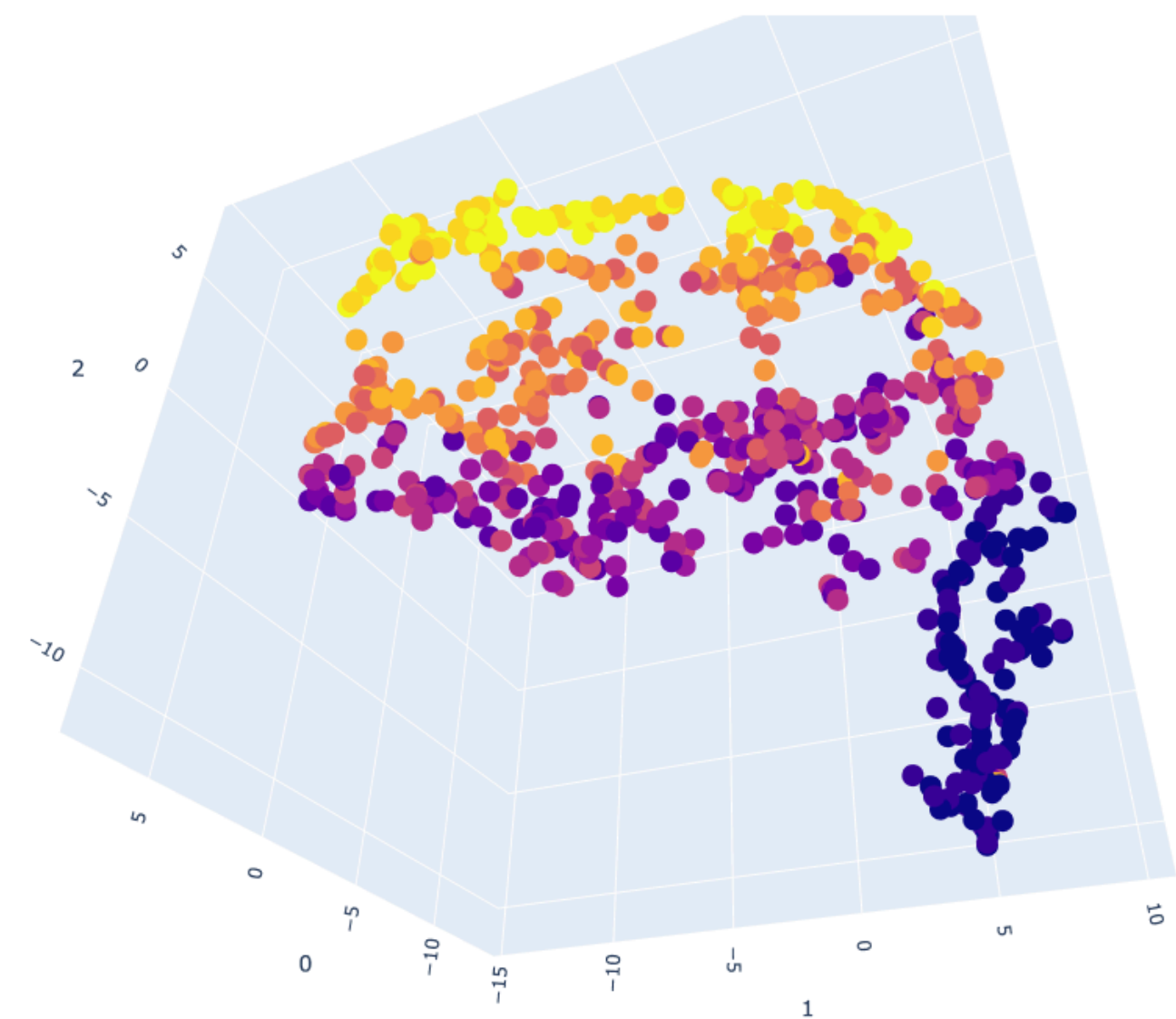


Correlation Map of Variables



Actual Dashboard

t-SNE all data, colored by year



t-SNE plot for one year, colored by district

