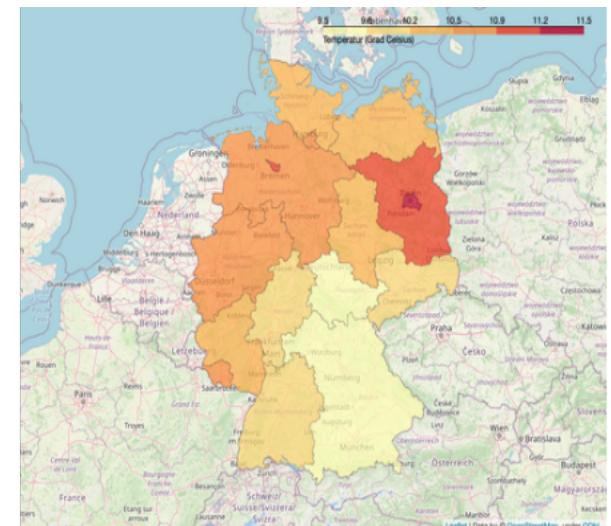
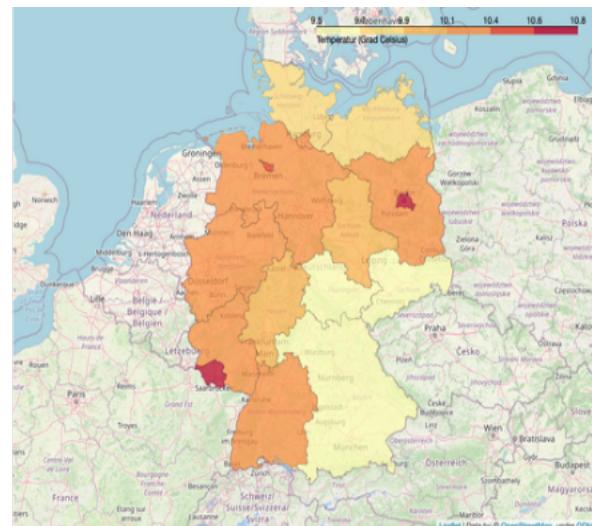
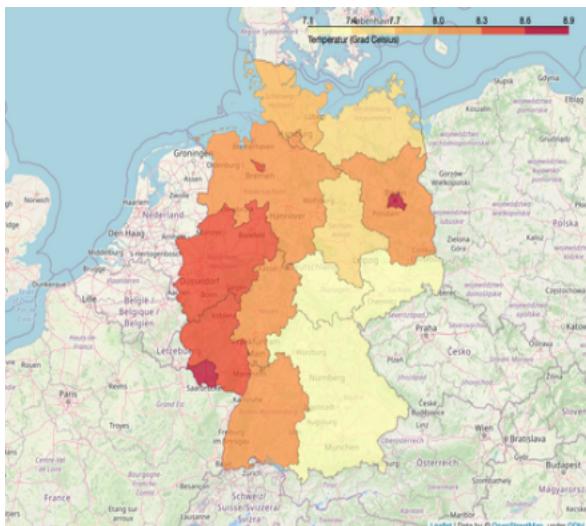


Smart Data Analytics

Analyse der DWD-Lufttemperaturdaten

Zihan Chen, Thassilo Helmold, Gerrit Merz, Robin Schnaitt – Team 4



Let's plan a journey... Das Wichtigste: Wetter!



<https://www.kennzeichenbox.de/magazin/kennzeichenmitnahme/>

Stadt oder Land?

Meer oder Berge?

In welcher Jahreszeit?

Nord oder Süd?

Welches Bundesland?

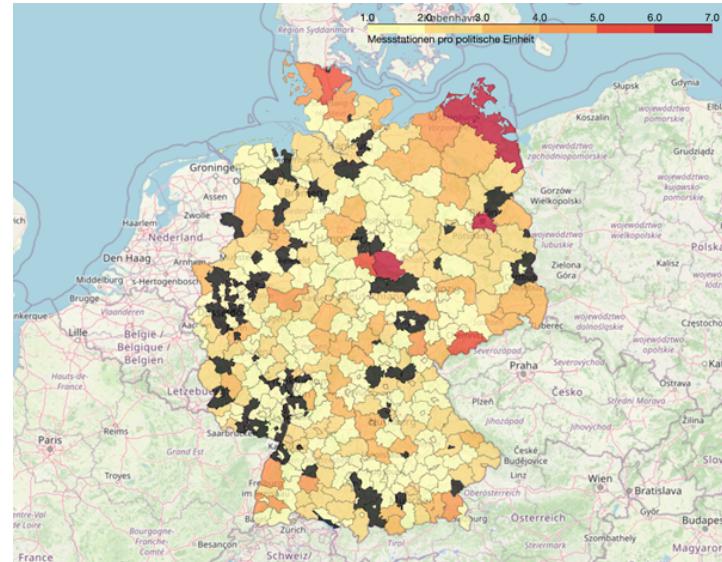
Data Pre-Processing

- Zuordnung von Messstationen zu Landkreisen mittels geopandas.contains() Funktion und OpenDataSoft-Daten

Geo Point	Geo Shape	Id 0	ISO	Name 0	Id 1	Name 1
1 48.1075826288, 9.77425436647	{"type":"Polygon", "coordinates": [[[9.8851	86	DEU	Germany	1	Baden-Württemberg
2 48.940602979, 9.12434425665	{"type":"MultiPolygon", "coordinates": [[[[]	86	DEU	Germany	1	Baden-Württemberg

<https://public.opendatasoft.com/explore/dataset/landkreise-in-germany/table/>

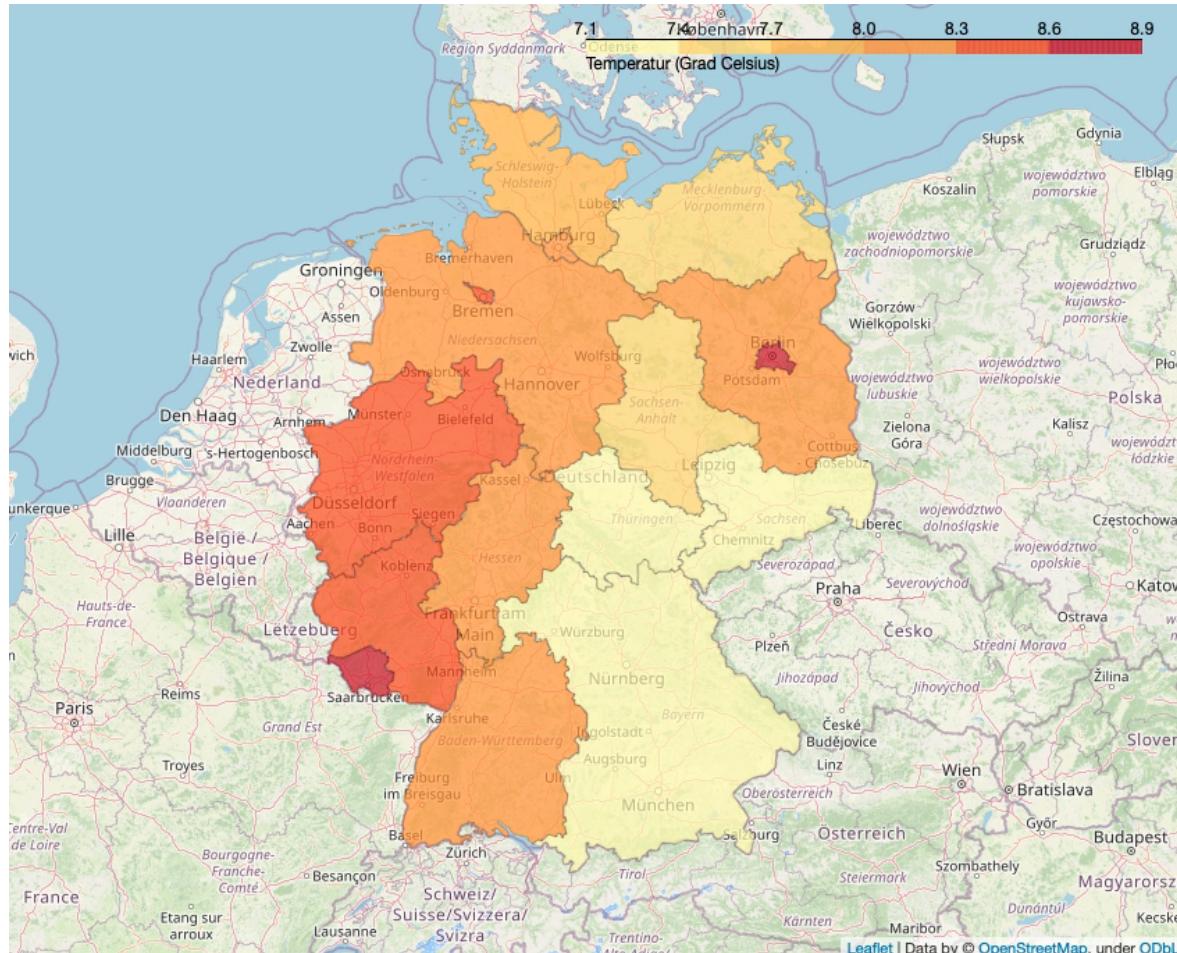
- Folium-Kartenbibliothek mit OpenStreetmap-Daten



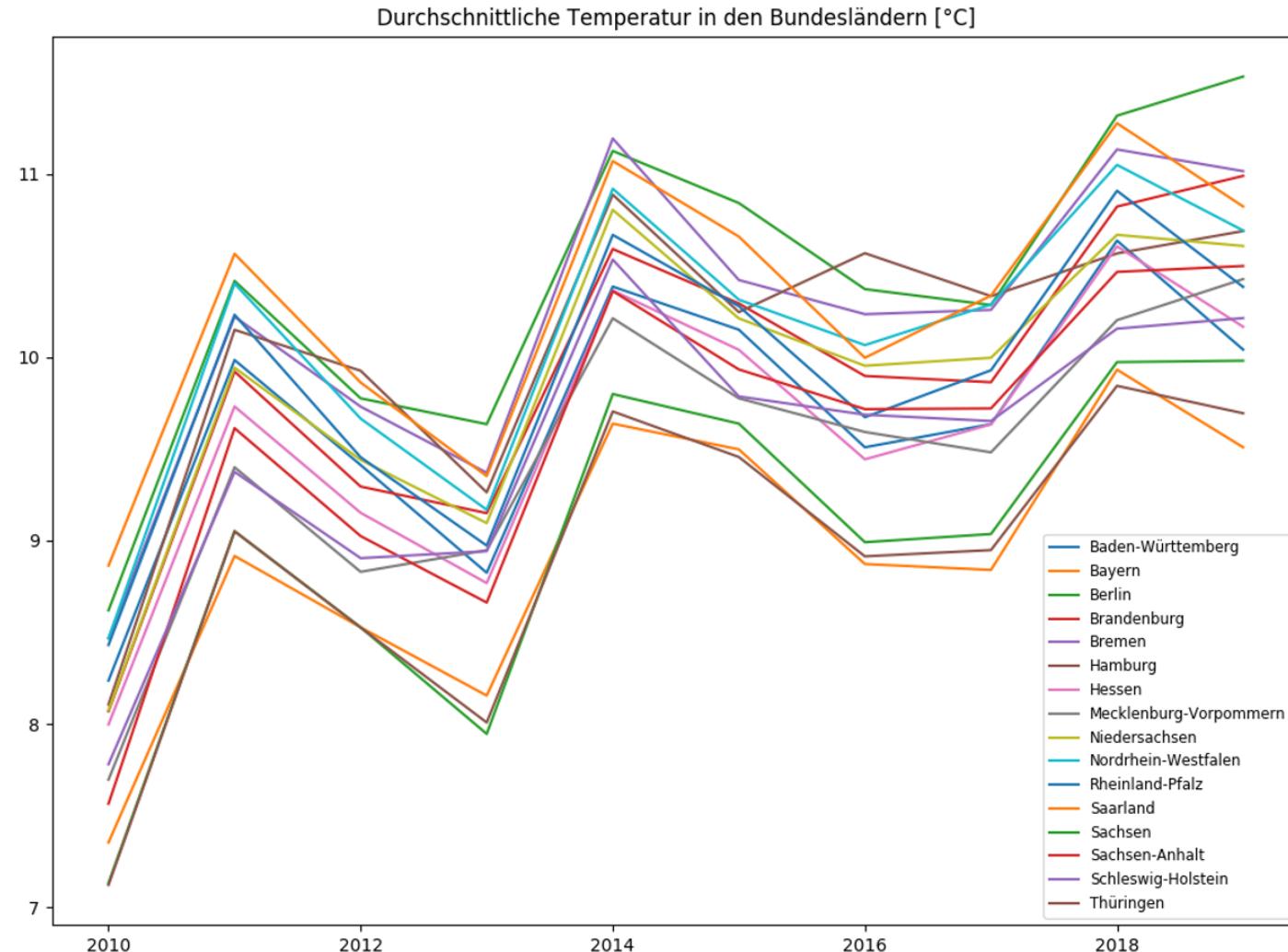
Visualisierung Anzahl der Lufttemperatur-Messstationsanzahl in den Landkreisen.

Data Exploration

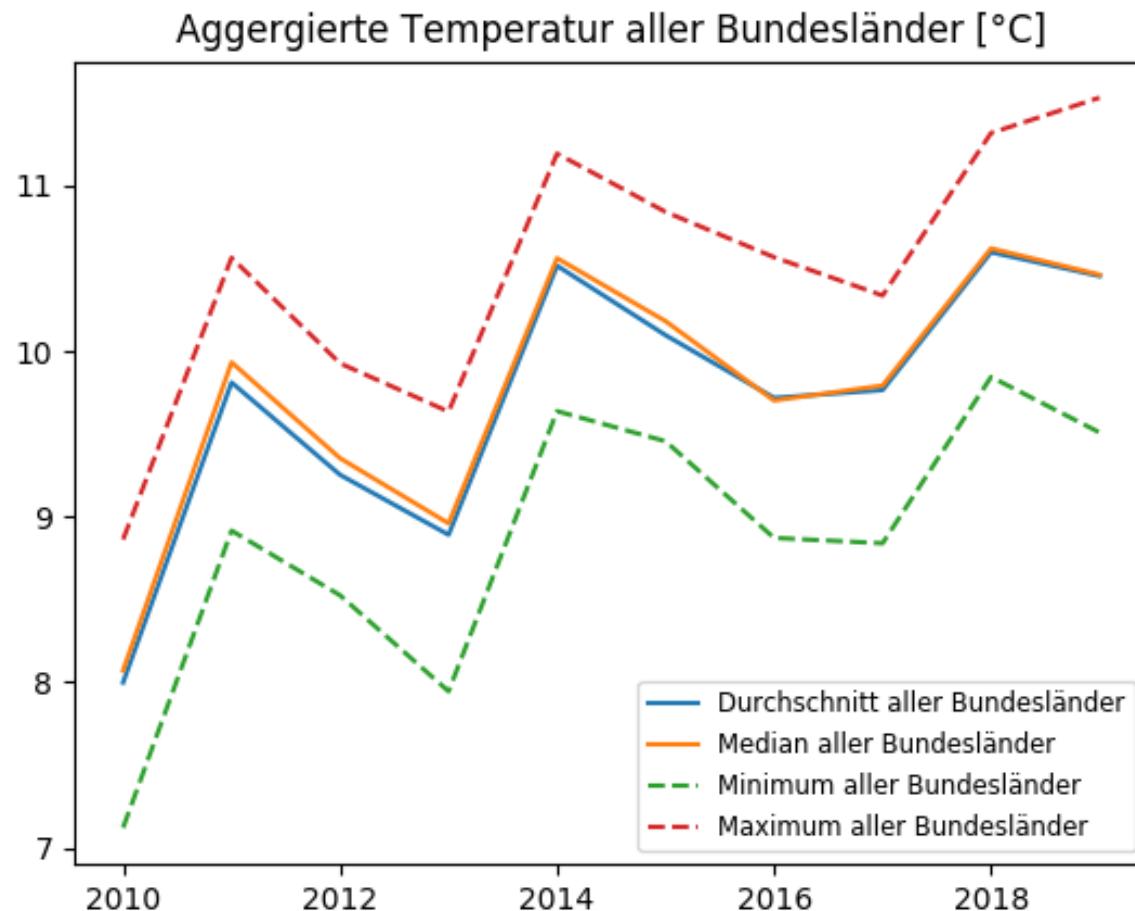
- Jährliche Durchschnittstemperaturen auf Bundeslandebene (2010-2019)
[Video]



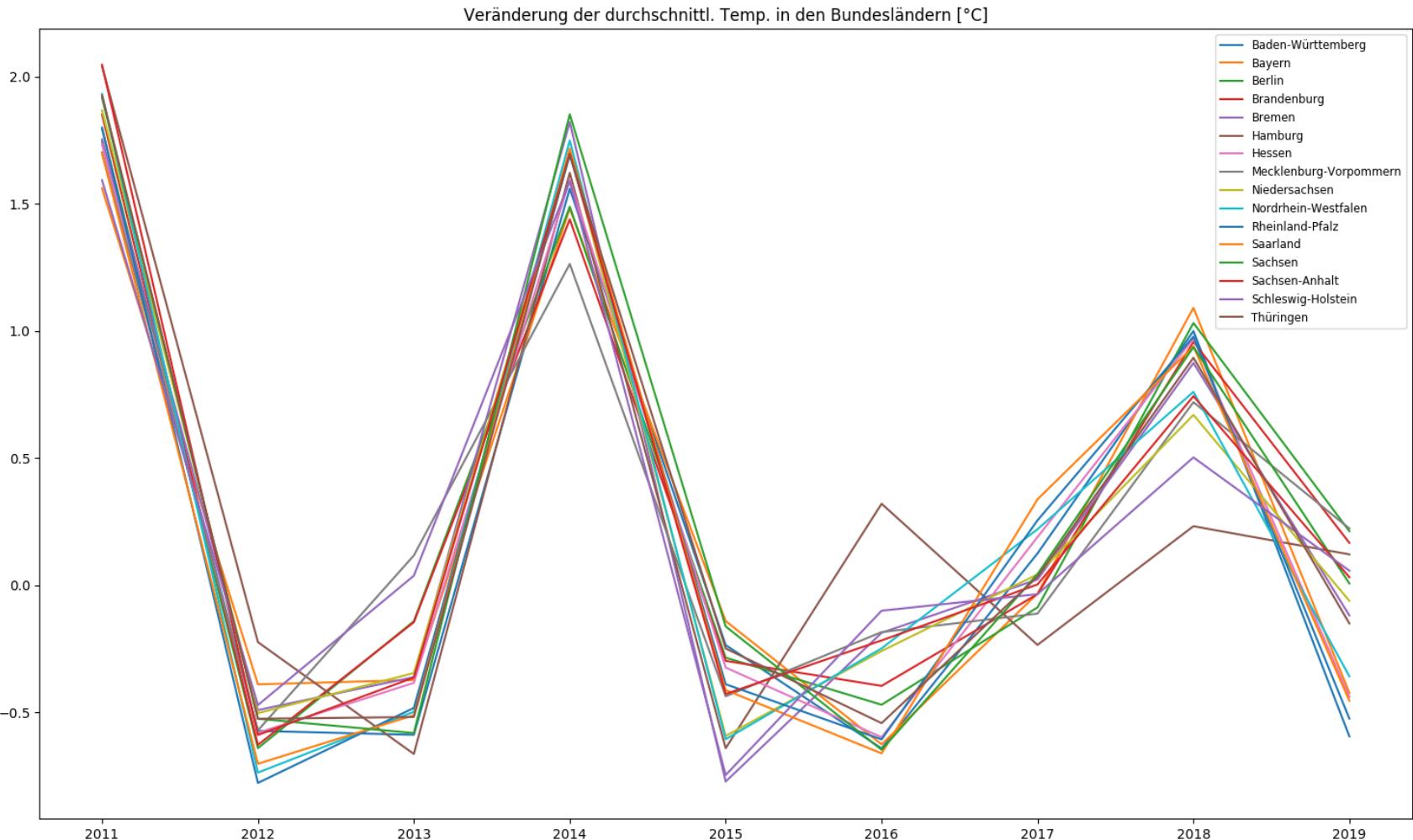
Data Exploration



Statistische Größen

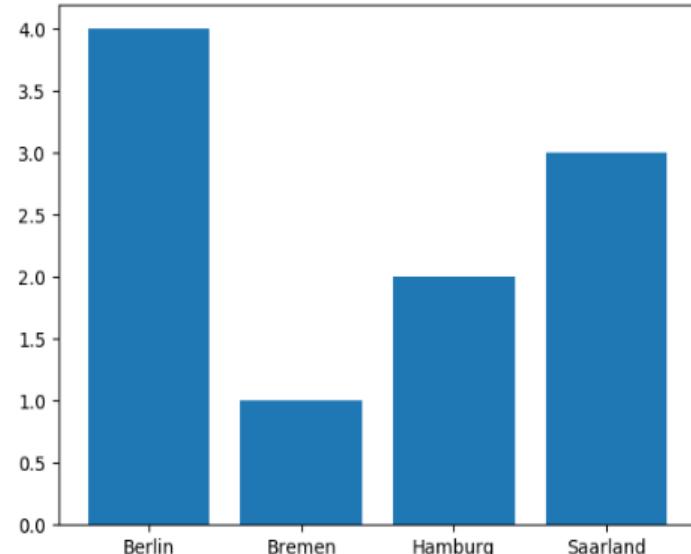


Änderungsrate

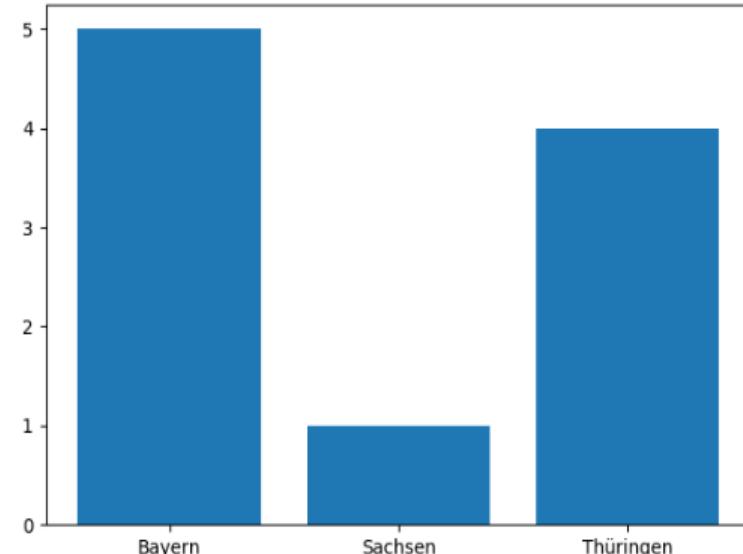


Data Exploration

Ranking der höchsten durchschnittlichen Temperaturen von 2010 bis 2019

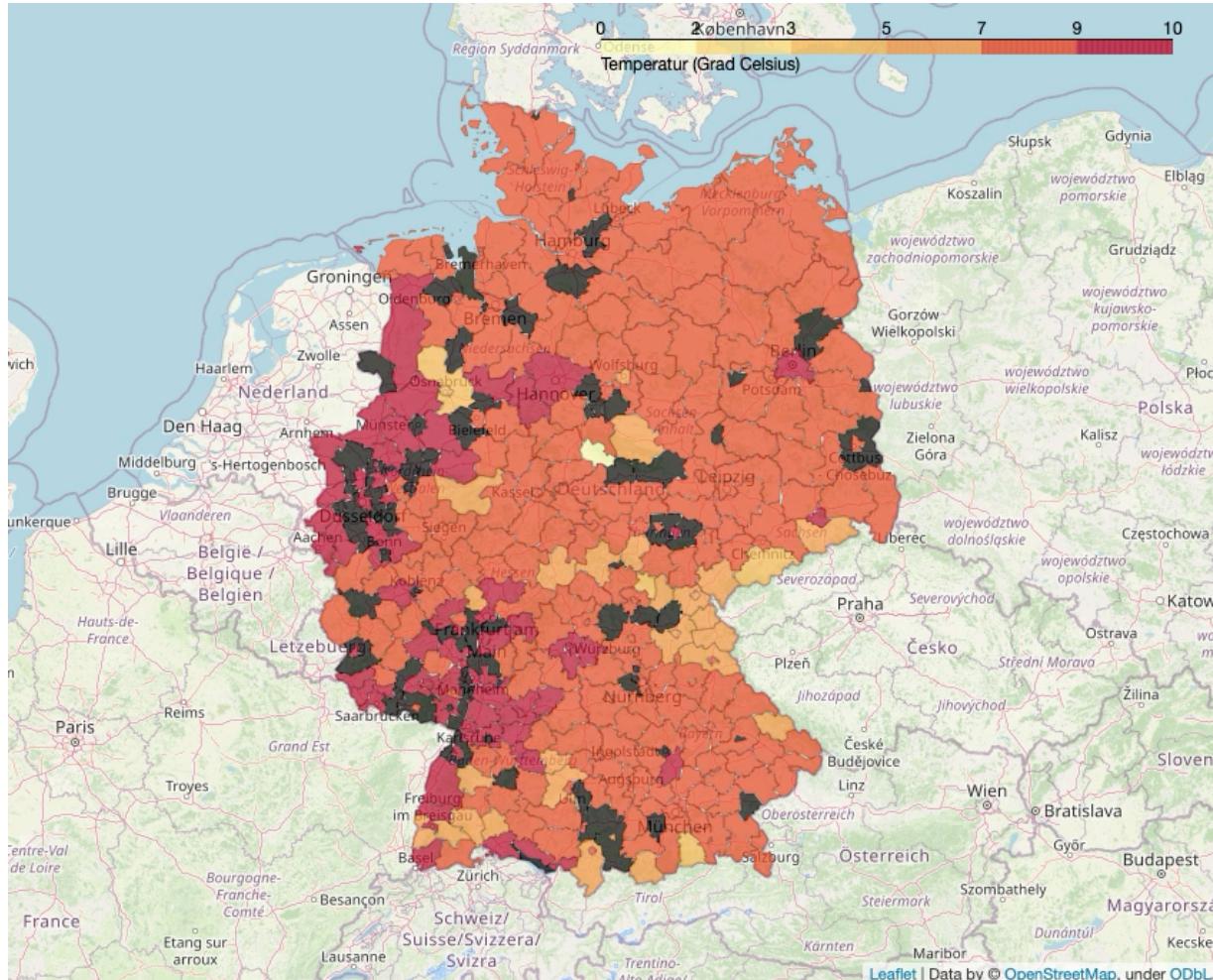


Ranking der niedrigsten durchschnittlichen Temperaturen von 2010 bis 2019



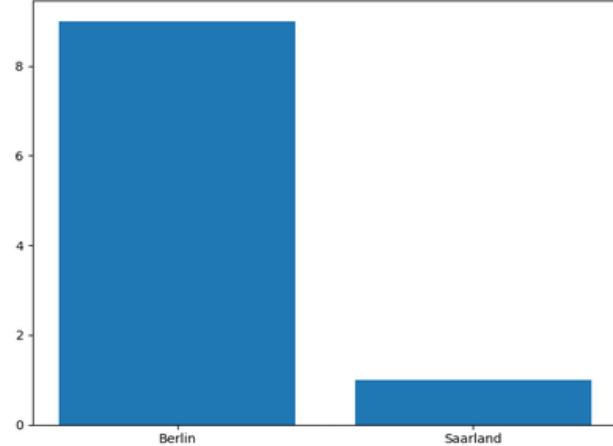
Data Exploration - Qualitative (4)

- Jährliche Durchschnittstemperatur auf Kreisebene (2010-19) [Video]

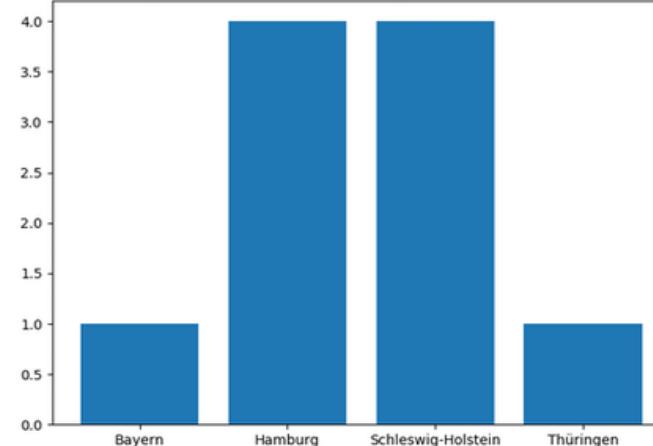


Data Exploration

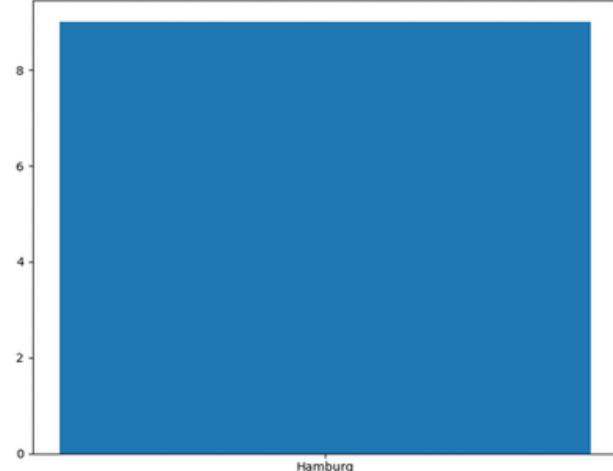
Ranking der höchsten durchschnittlichen Temperaturen von 2010 bis 2019 (Sommer)



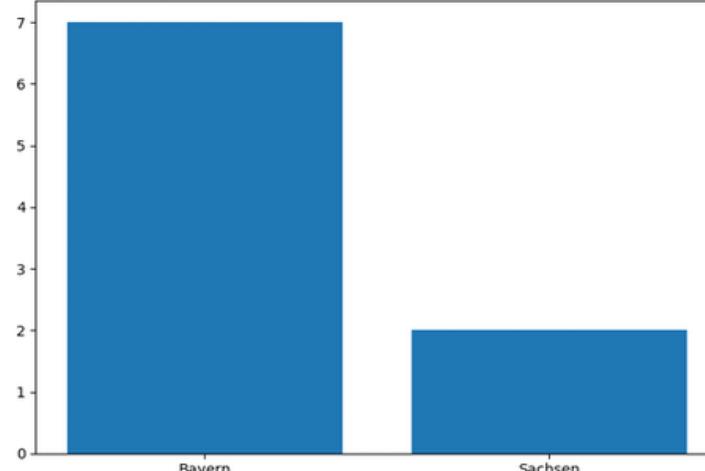
Ranking der niedrigsten durchschnittlichen Temperaturen von 2010 bis 2019 (Sommer)



Ranking der höchsten durchschnittlichen Temperaturen von 2010 bis 2019 (Winter)



Ranking der niedrigsten durchschnittlichen Temperaturen von 2010 bis 2019 (Winter)

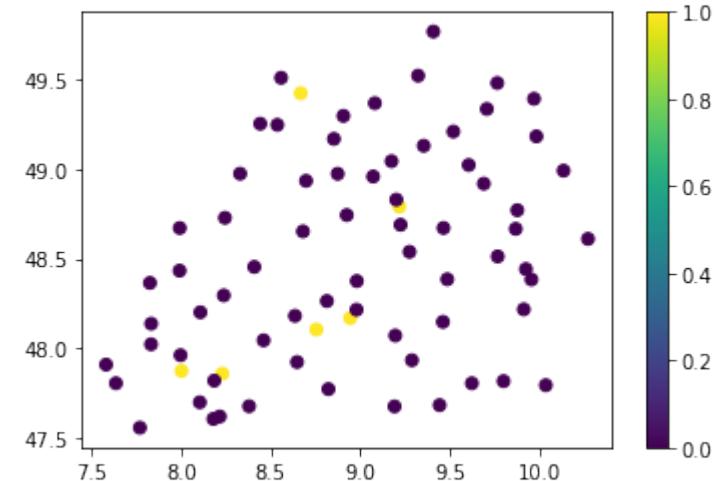
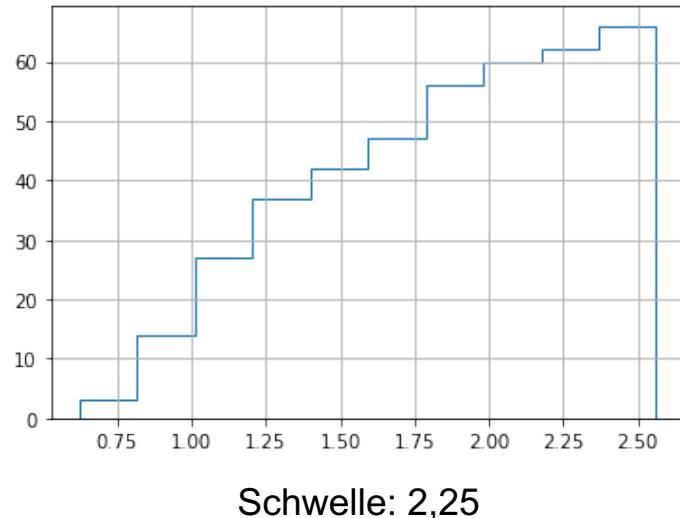


Erklärungsmöglichkeiten

- Hamburg & Schleswig-Holstein: Maritimes Klima
→ Milde Sommer und warme Winter
- Berlin, Saarland, Hamburg, Bremen: Kleine Bundesländer
 - Größere Länder haben es schwerer, ein Ausreißer zu werden, weil sich die Temperaturen über Stationen ausgleichen
 - Andere warme Regionen werden hier nicht gleich gewertet, weil keine politische Trennung
 - Aber: Liegt *nicht* an Stadt vs. Land. Wurde getrennt untersucht, kein signifikanter Unterschied (bundesweit und in BW)
- Bayern: Höhere Lagen führen zu kälterem Winterklima

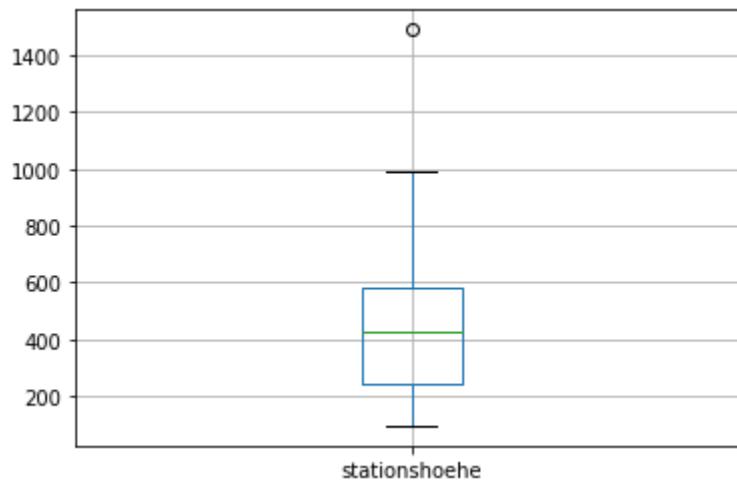
Regionale Betrachtung: Baden-Württemberg

- Suche zunächst nach Stationen, die stark vom Landesdurchschnitt abweichen
- Berechnung der Root-Mean-Squared-Difference zum globalen Mean



Regionale Betrachtung: Baden-Württemberg

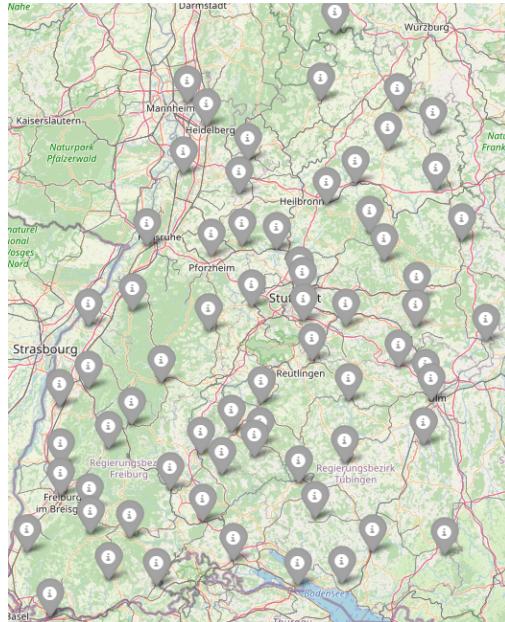
stationshoehe	geobreite	geolaenge	stationsname	bundesland	geom_station	land	total_diff	large_diff
1490.0	8.0038	47.8749	Feldberg/Schwarzwald	Baden-Württemberg	POINT (8.00380 47.87490)	Deutschland	5.794898	True
110.0	8.6676	49.4206	Heidelberg	Baden-Württemberg	POINT (8.66760 49.42060)	Deutschland	2.410743	True
974.0	8.7548	48.1054	Klippenbeck	Baden-Württemberg	POINT (8.75480 48.10540)	Deutschland	2.561289	True
852.0	8.2308	47.8597	Lenzkirch-Ruhbühl	Baden-Württemberg	POINT (8.23080 47.85970)	Deutschland	2.373517	True
900.0	8.9433	48.1694	Meßstetten-Appental	Baden-Württemberg	POINT (8.94330 48.16940)	Deutschland	2.455823	True
224.0	9.2167	48.7896	Stuttgart (Neckartal)	Baden-Württemberg	POINT (9.21670 48.78960)	Deutschland	2.269831	True



Clustering

~~Richtig oder Falsch~~

Nützlich oder weniger nützlich



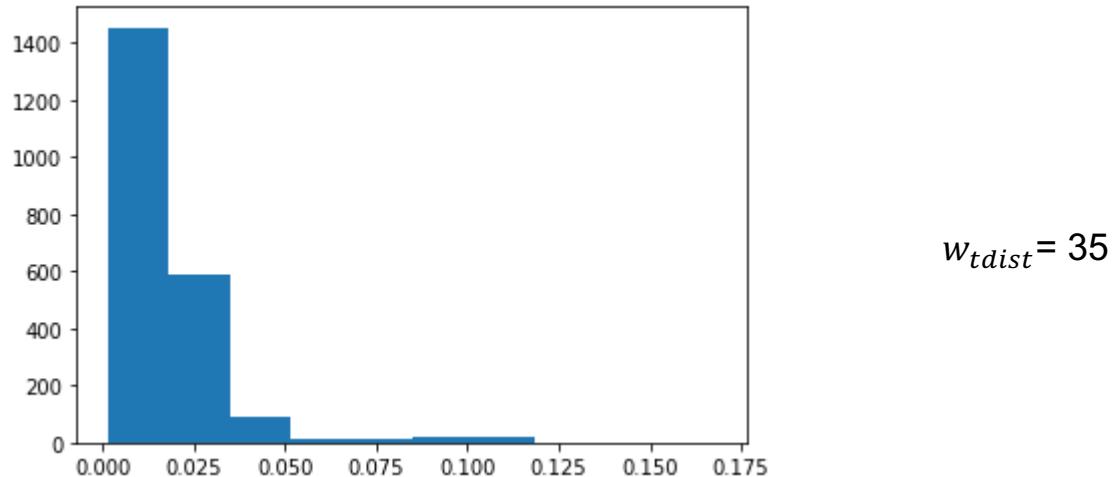
Clustering

- Clustering-Algorithmus: DBSCAN
- Distanz-Metrik:
 - Geografische Distanz
 - „Temperaturdistanz“

$$dist = w_{dist} \cdot geodist + w_{tdist} \cdot tempdist$$

Clustering mit Korrelation

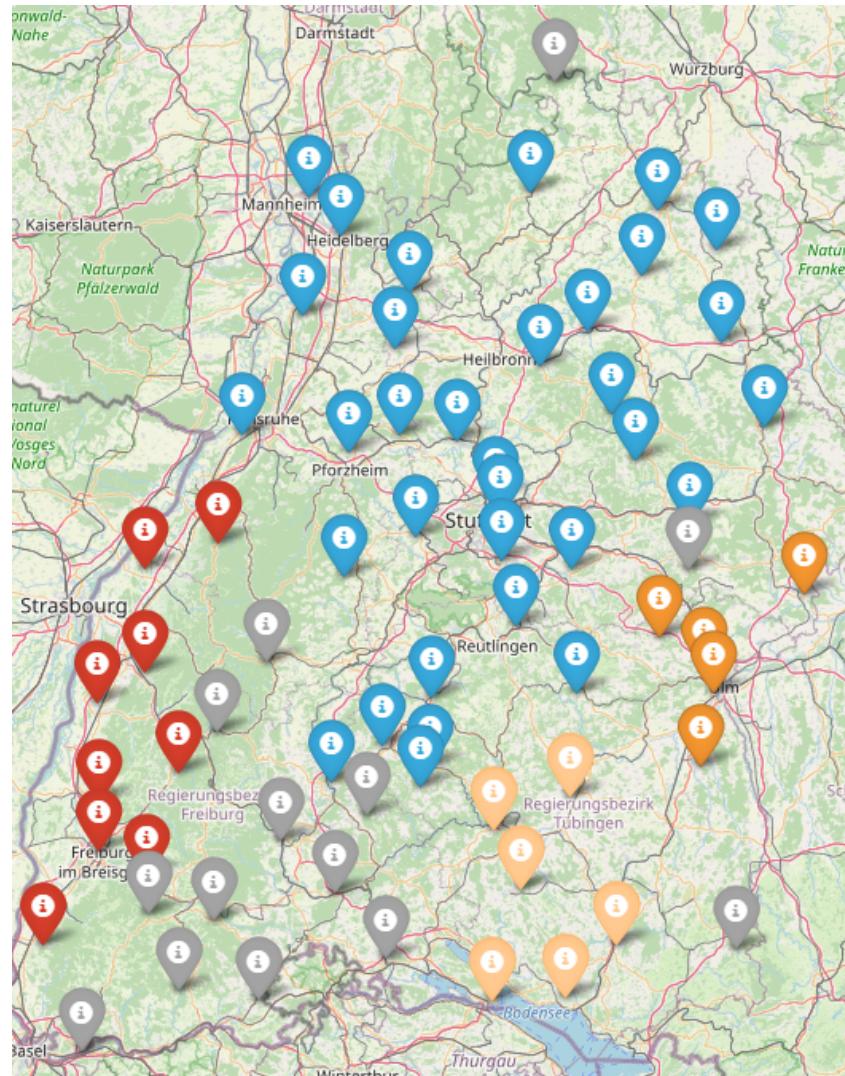
$$tempdist(X, Y) = 1 - \text{corr}(X, Y)$$



- Eps: 0.10 gives 0 clusters and 67 noise
- Eps: 0.20 gives 0 clusters and 67 noise
- Eps: 0.30 gives 0 clusters and 67 noise
- Eps: 0.40 gives 1 clusters and 62 noise
- Eps: 0.50 gives 3 clusters and 43 noise
- Eps: 0.60 gives 4 clusters and 14 noise
- Eps: 0.70 gives 1 clusters and 6 noise
- Eps: 0.80 gives 1 clusters and 3 noise
- Eps: 0.90 gives 1 clusters and 1 noise
- Eps: 1.00 gives 1 clusters and 1 noise

$\varepsilon = 0,6$

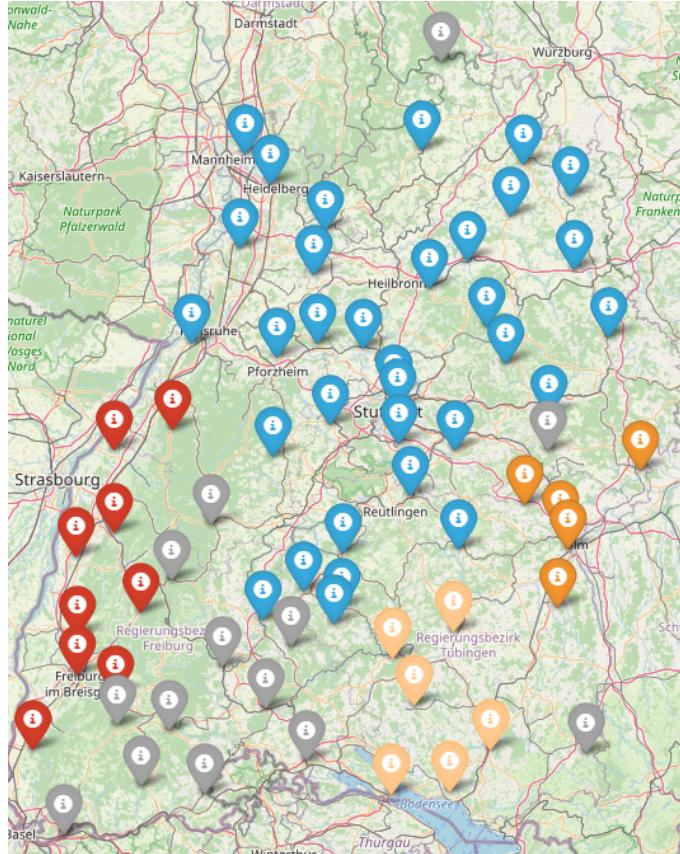
Clustering mit Korrelation



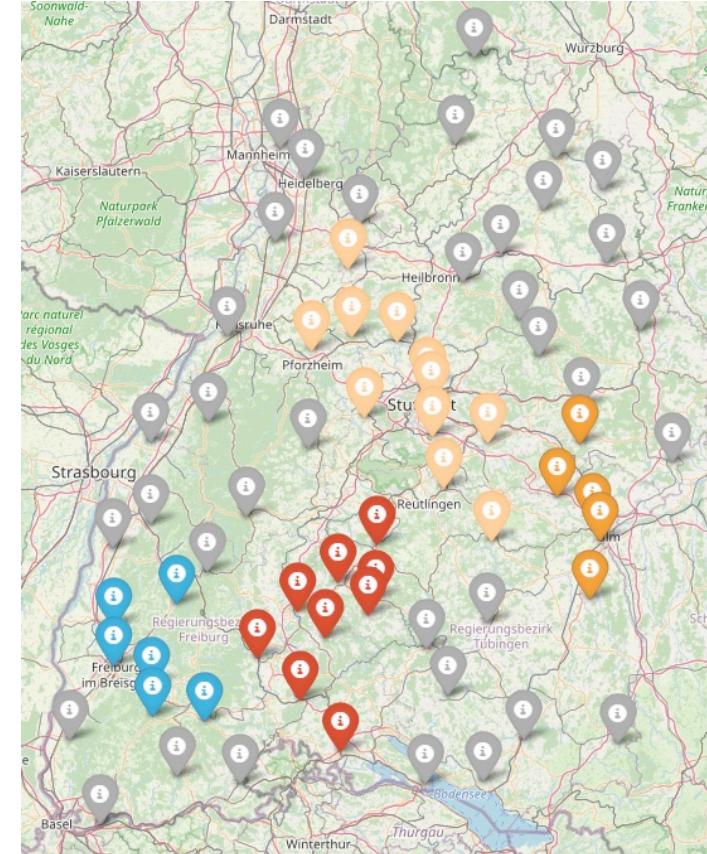
$$w_{tdist} = 35$$

$$\varepsilon = 0,6$$

Clustering mit Korrelation



$$w_{t\text{dist}} = 35$$



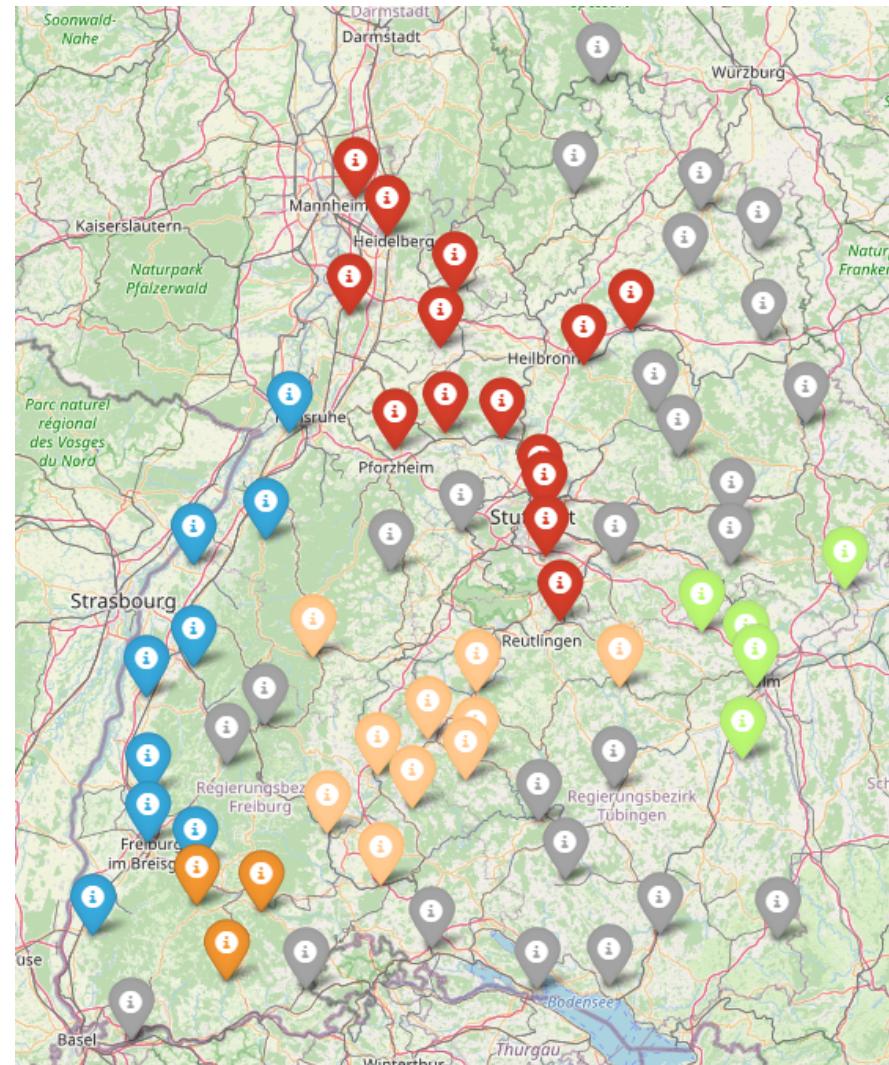
$$w_{t\text{dist}} = 1$$

Clustering mit Cosinus-Distanz

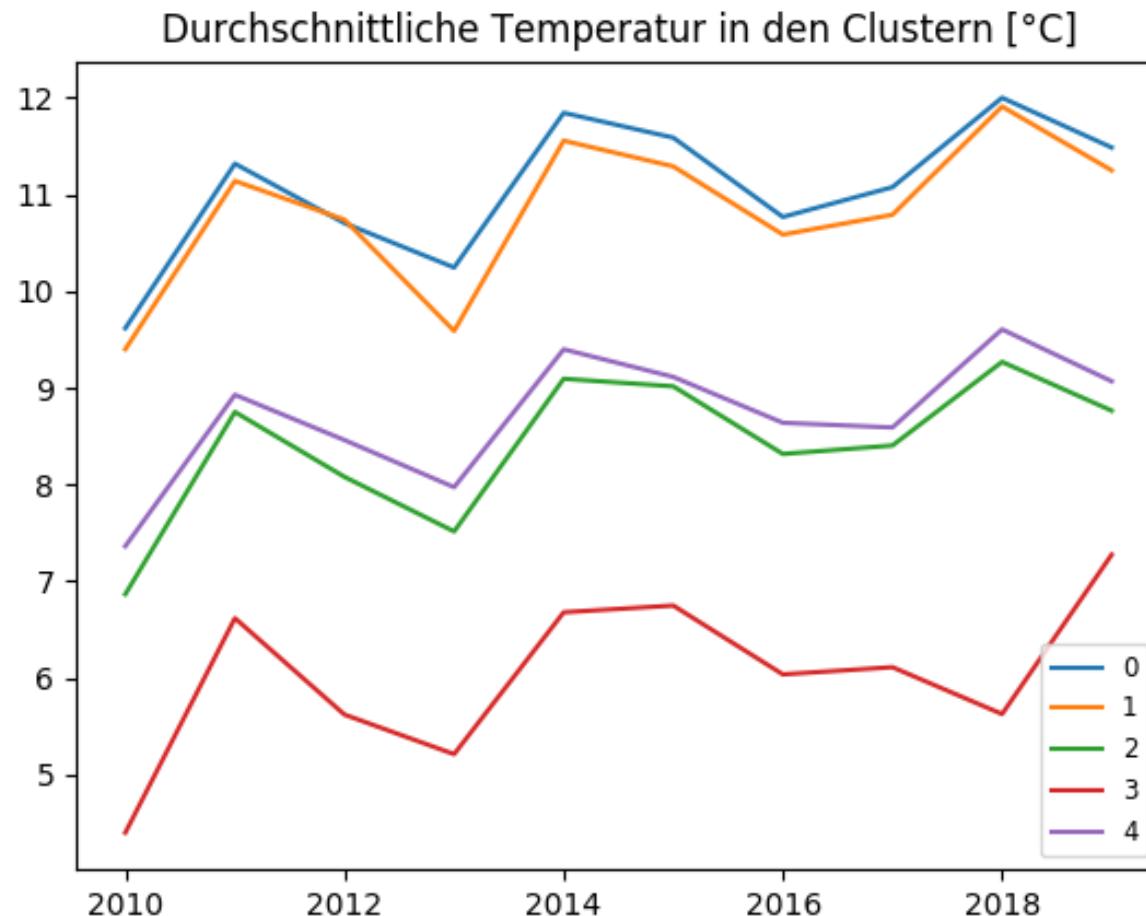
Clustern nach gemeinsamer Abweichung
vom globalen Trend:

Differenz zum globalen Durchschnitt

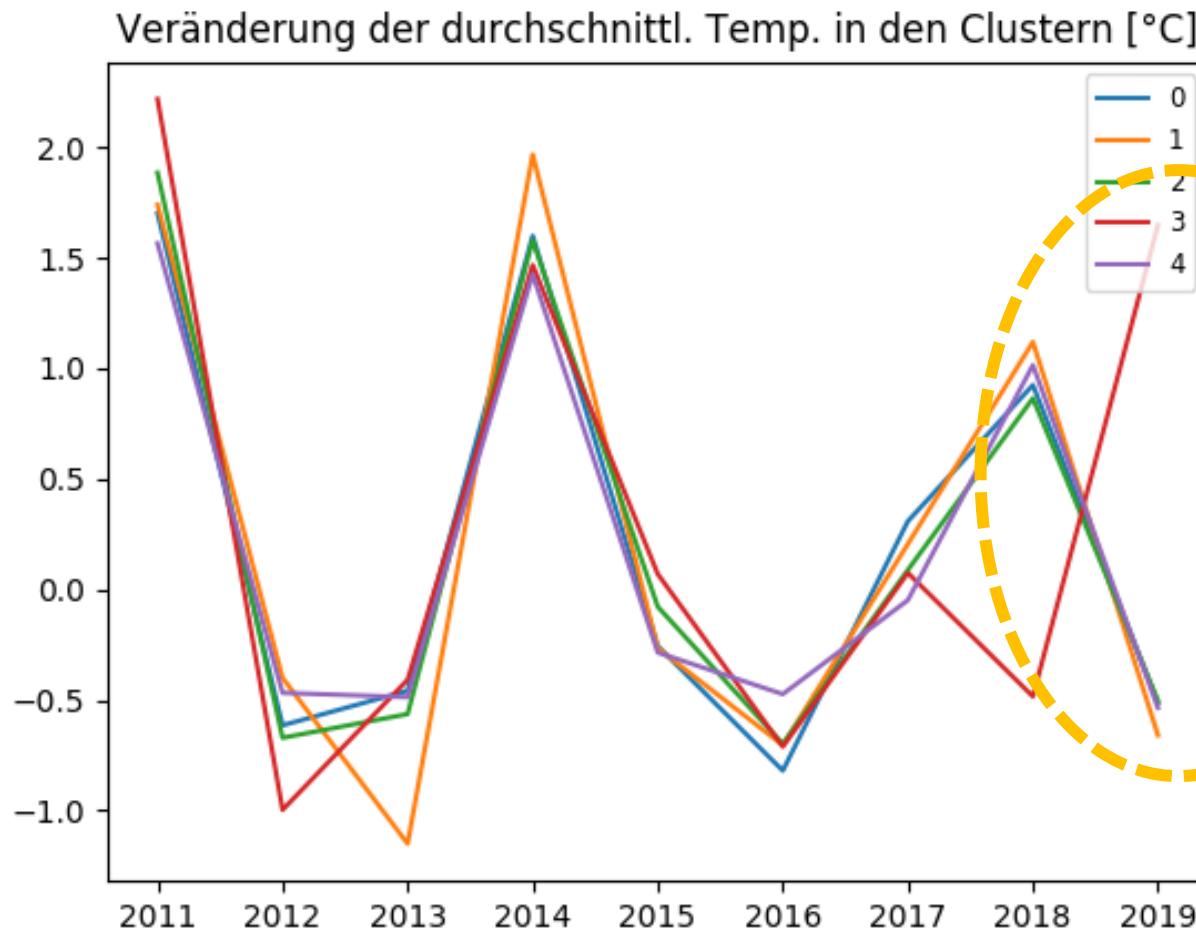
$$\begin{aligned} \text{tempdist} &= \text{cosinedist} \\ &= 1 - \cos(\phi) \end{aligned}$$



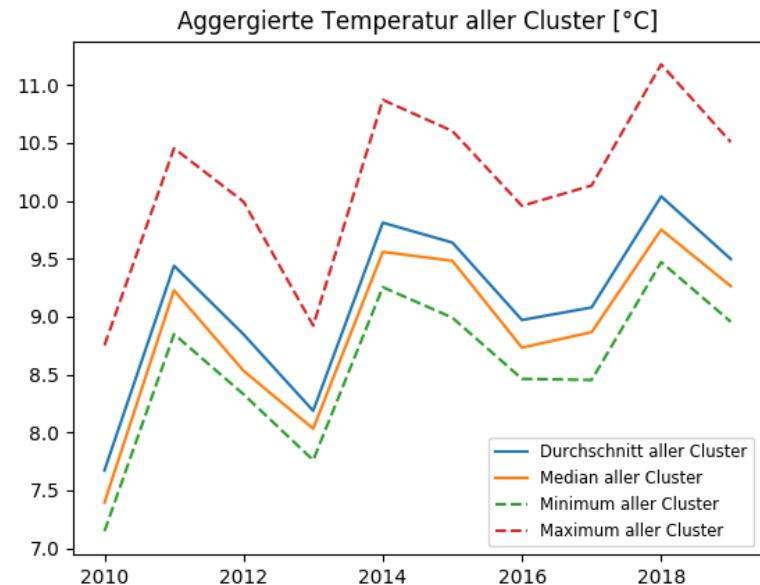
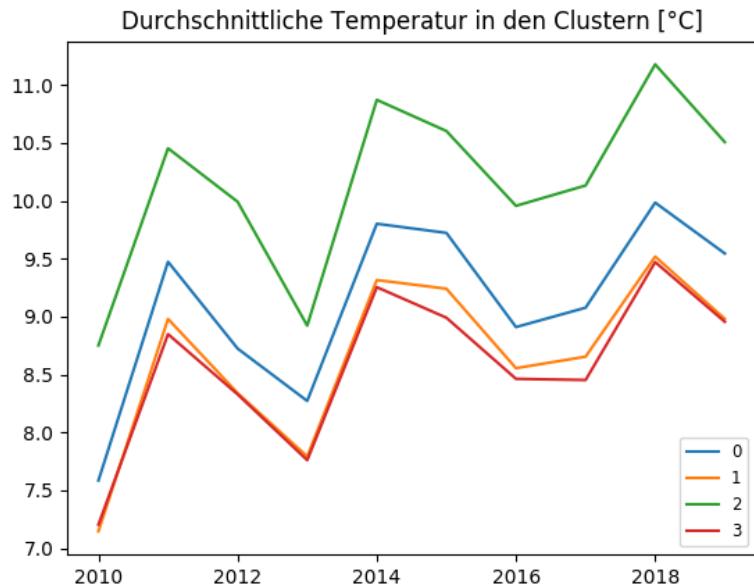
Analyse der Cluster – „Cosine“ (1)



Analyse der Cluster – „Cosine“ (2)

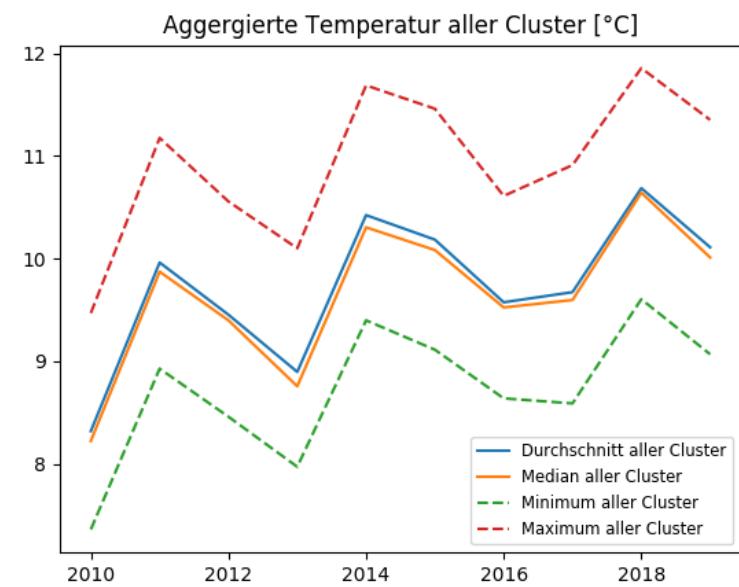
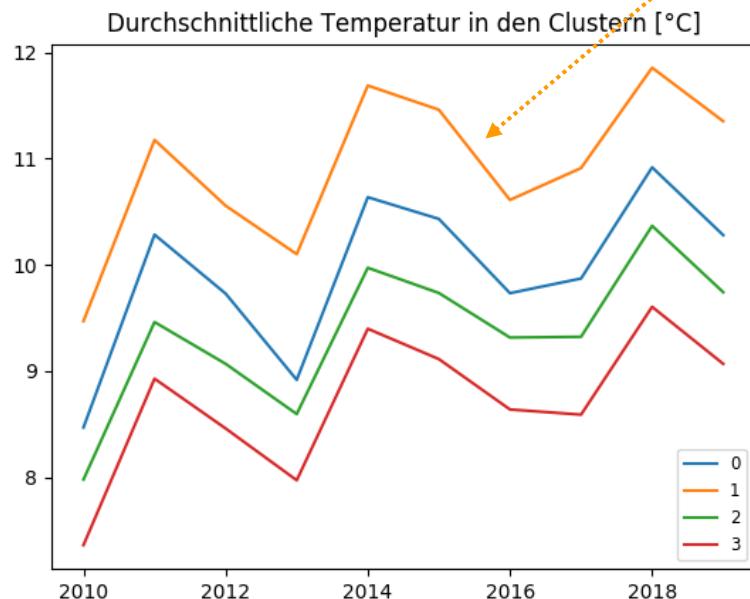


Analyse der Cluster – „Corr-1“

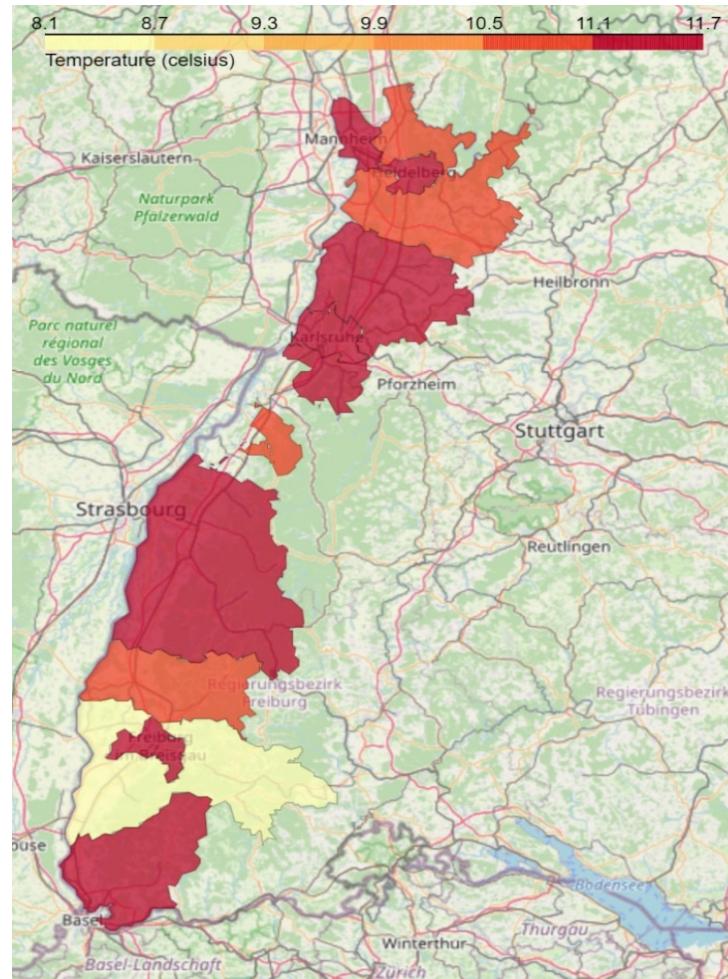
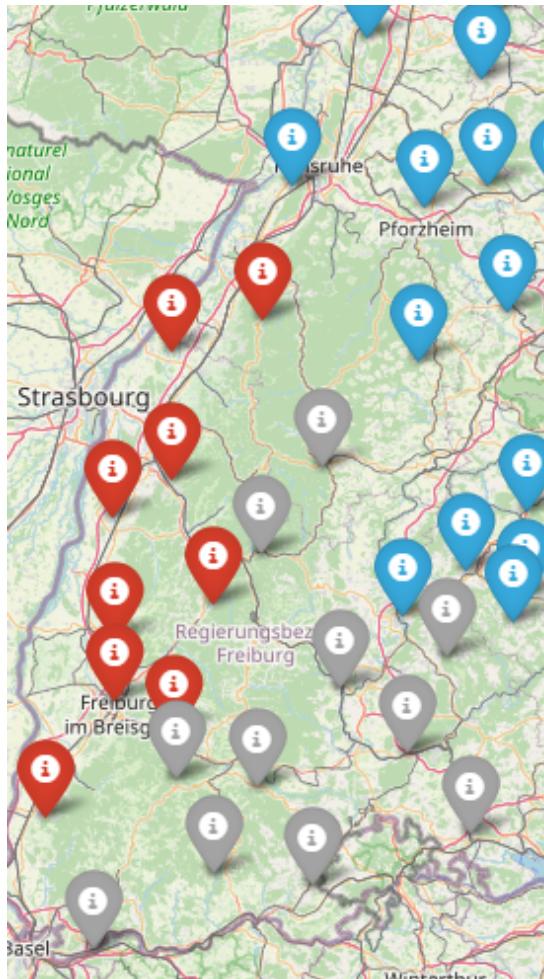


Analyse der Cluster – „Corr-35“

Oberrheinische Tiefebene



Oberrheinische Tiefebene

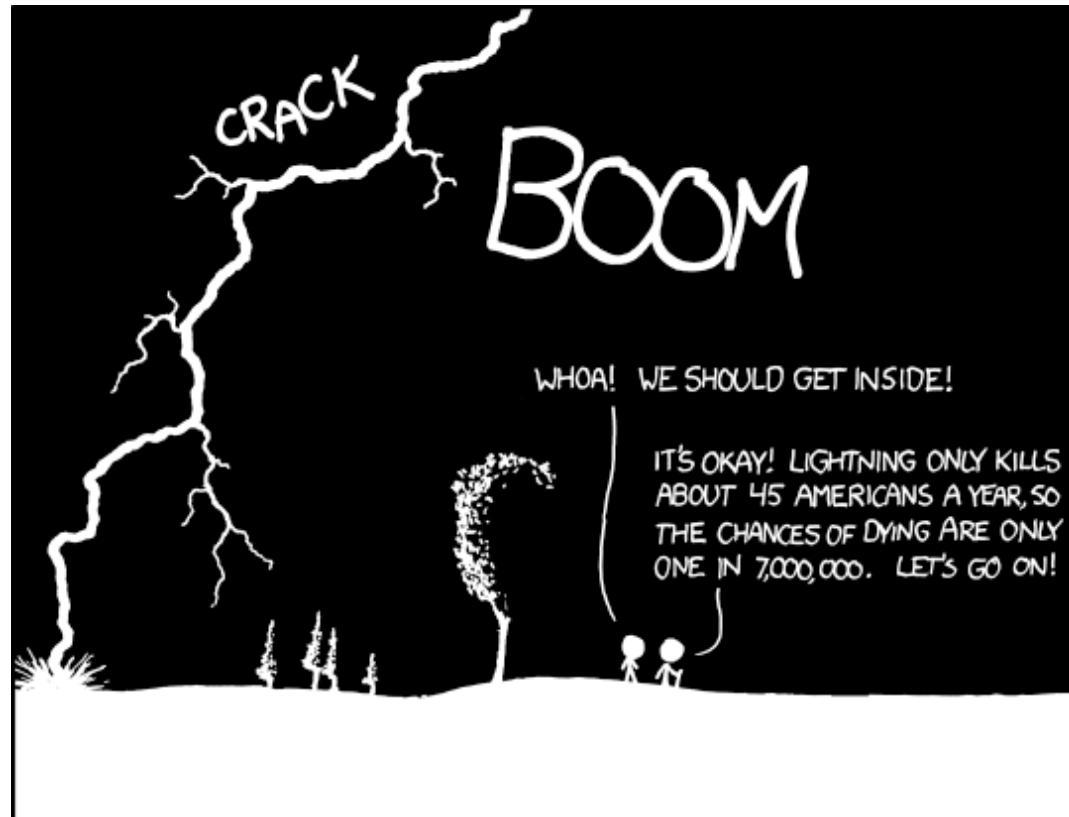


Ausblick

- Lufttemperaturdaten nicht alles – Luftfeuchtigkeit, Niederschlagsmengen, Unwetterhäufigkeit, ...
- Domänenwissen und Datenanalyse verknüpfen



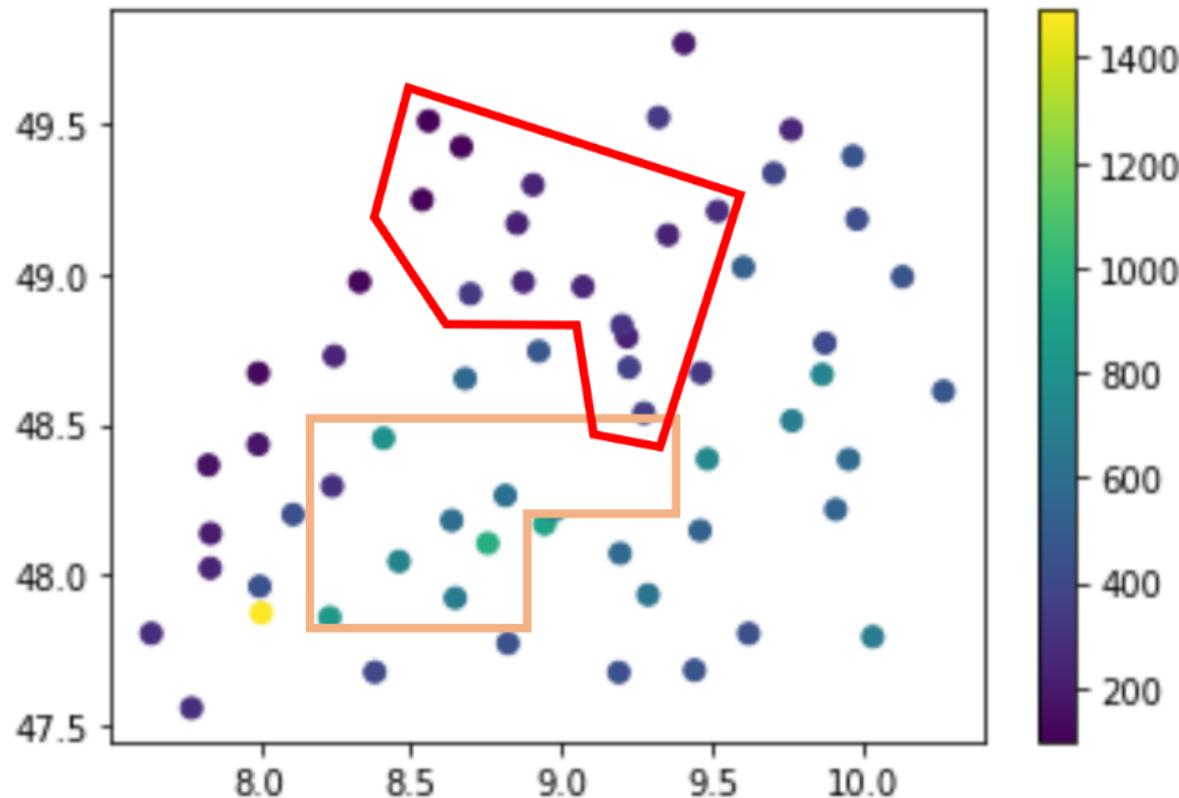
Thank you for your attention!



THE ANNUAL DEATH RATE AMONG PEOPLE
WHO KNOW THAT STATISTIC IS ONE IN SIX.

<https://xkcd.com/795/>

Backup Slides



Backup Slides - Data Provision

- Erste Herausforderung: Daten-Infrastruktur
- Bereitgestelltes Docker-Image (kleine Version), gehostet auf bwCloud-Server
- Abruf der benötigten Daten aus der Datenbank mit Python-Skripts

Backup Slides



Backup Slides - Data Exploration

- Monatliche Durchschnittstemperatur auf Bundeslandebene (2010-19)
[Video]

