Data Engineering

Project Presentation

CONTENT

01 Background

02 Result

03 Challenges

04 Outlook

Background

Background

Transportation is important for your private life

- Every person uses transportation to get from A to B
- There are different ways of transportation
- E.g., public and private

Some areas have better infrastructure than others

- In some areas there are more means of transportation
- Some have more public transportation
- E.g., Metro, Trams and Busses

Predominantly bigger cities have a good infrastructure

- Only big cities have trams and metros
- E.g., Nürnberg

Background

Private vehicles are less eco-friendly

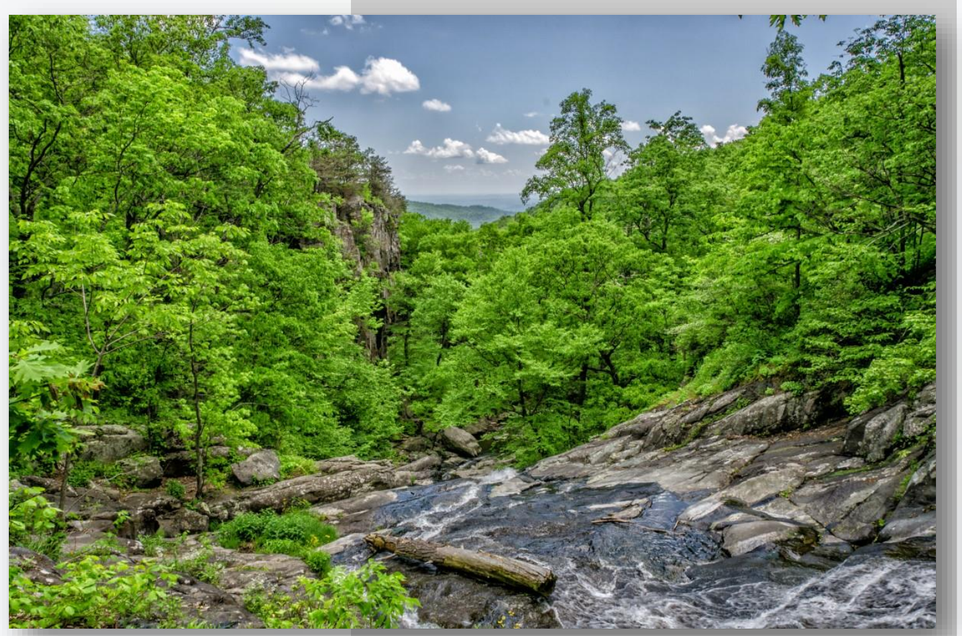
- Cars produce a lot of CO2
- Especially calculated on the transported distance per capita

Using public transport is more eco-friendly

- Bulking transport together produces less CO2 per capita
- Better fuels (e.g., Electricity) can be used

Expanding public transportation in the right regions can contribute to sustainability

- More public transport, means less CO2



Data sources

County Data

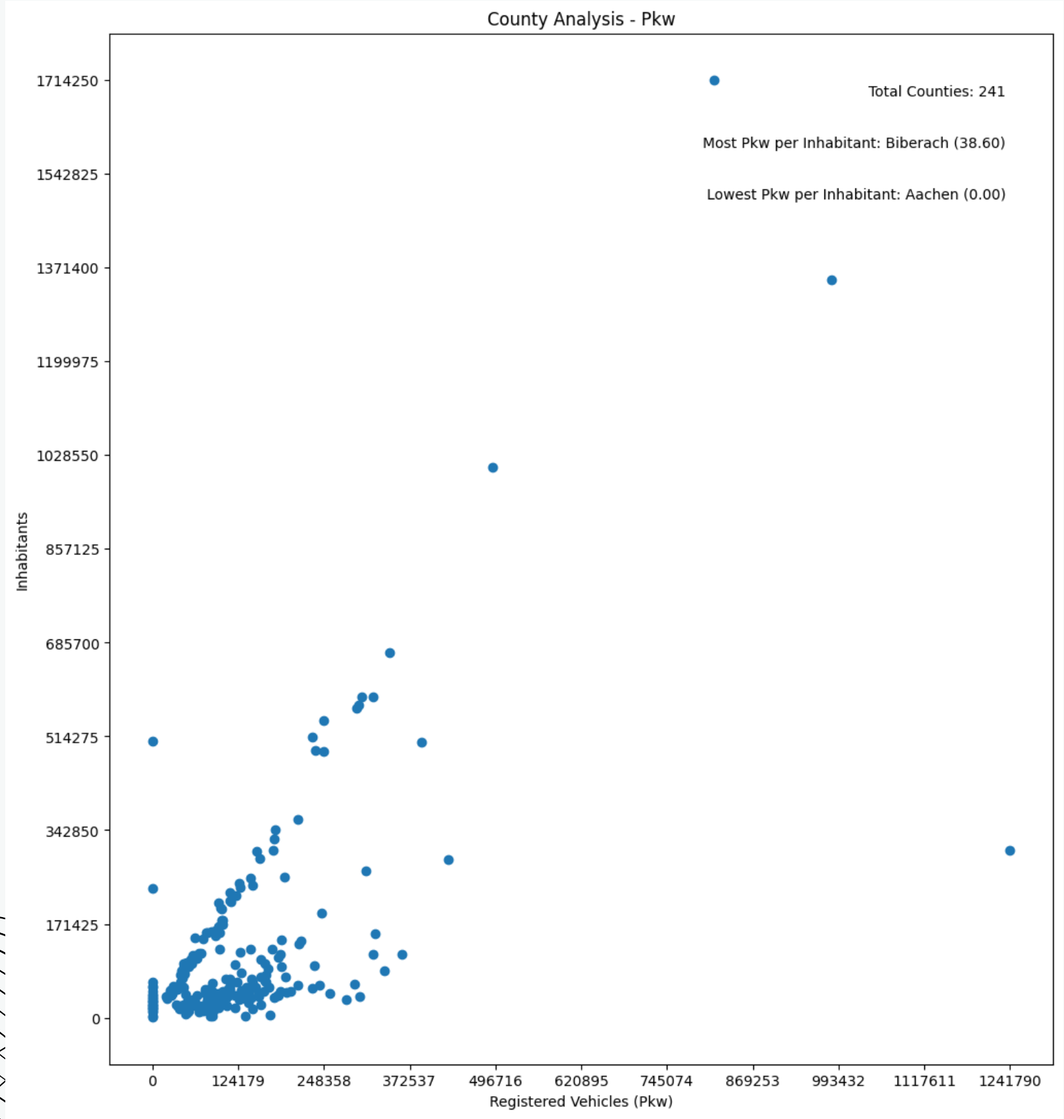
- County name
- Amount of inhabitants
- Other data which did not get used (e.g., ZIP-Code etc.)

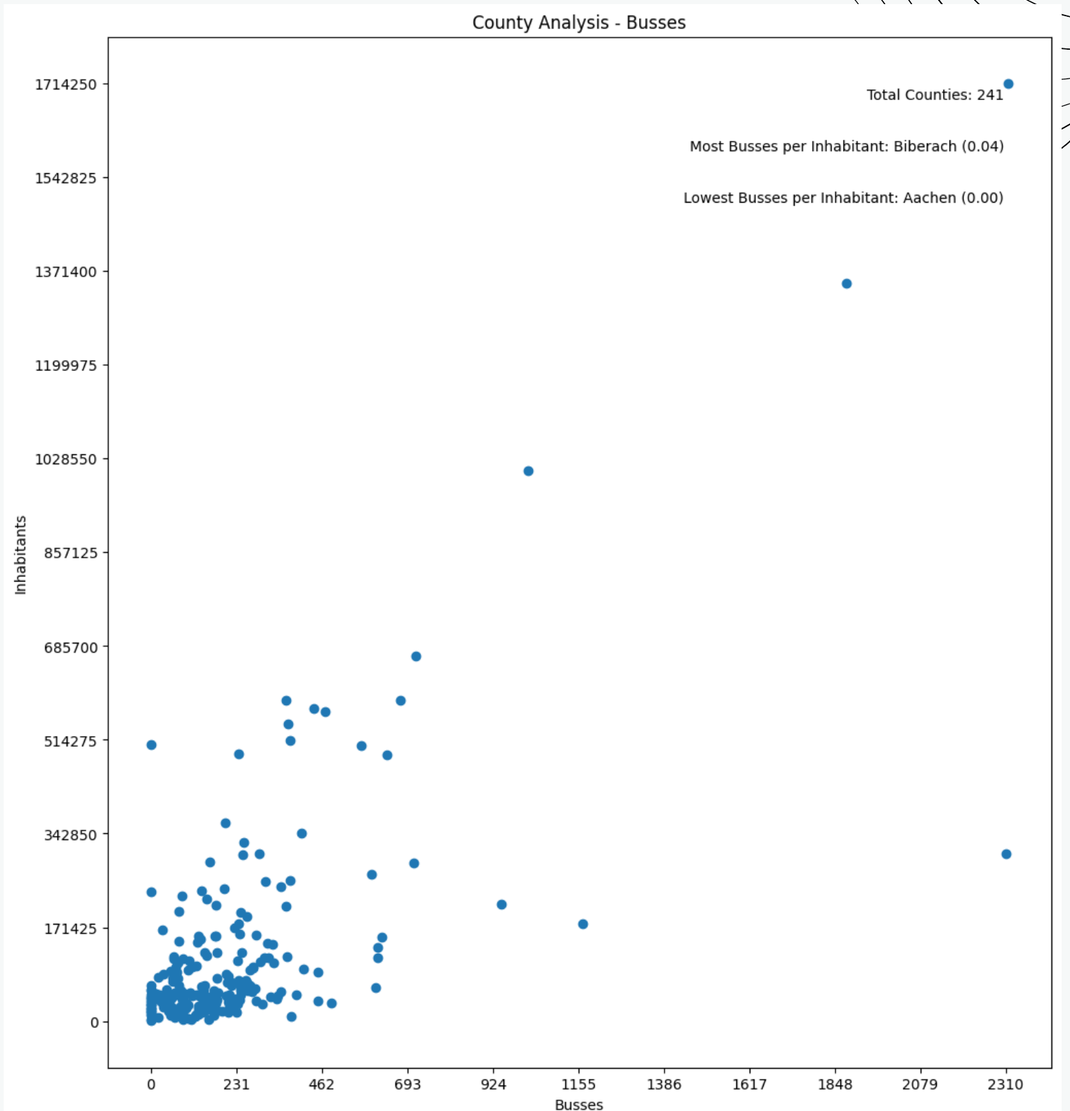
Vehicle Data

- County name
- Registered private cars
- Registered busses
- Other data which did not get used (e.g., Motorbikes etc.)

Results

Results


County Analysis - Pkw
Total Counties: 241
Most Pkw per Inhabitant: Biberach (38.60)
Lowest Pkw per Inhabitant: Aachen (0.00)
Inhabitants
1714250
1542825
1371400
1199975
1028550
857125
685700
514275
342850
171425
0
0
124179
248358
372537
496716
620895
745074
869253
993432
1117611
1241790
Registered Vehicles (Pkw)


County Analysis - Busses
Total Counties: 241
Most Busses per Inhabitant: Biberach (0.04)
Lowest Busses per Inhabitant: Aachen (0.00)
Inhabitants
1714250
1542825
1371400
1199975
1028550
857125
685700
514275
342850
171425
0
0
231
462
693
924
1155
1386
1617
1848
2079
2310
Busses

Results

Top 10 Counties - Vehicles (Pkw) per Inhabitant:

County	Registered Vehicles / Inhabitants
Biberach	38.601714
Heidenheim	34.345295
Steinburg	31.945873
Harburg	31.346890
Oldenburg	12.099098
Altenkirchen (Westerwald)	9.840982
Miltenberg	9.607798
Diepholz	9.025261
Steinfurt	8.400418
Mettmann	7.888302

Correlation between Registered Vehicles (Pkw) & Inhabitants
Correlation coefficient: 0.6695308650759194
P-value: 1.0483485167769133e-32

Top 10 Counties - Buses per Inhabitant:

County	Registered Buses / Inhabitants
Biberach	0.044857
Altenkirchen (Westerwald)	0.043589
Steinburg	0.041459
Heidenheim	0.035088
Harburg	0.018771
Ebersberg	0.015418
Kelheim	0.015148
Steinfurt	0.014587
Miltenberg	0.013180
Diepholz	0.013131

Correlation between Registered Buses & Inhabitants
Correlation coefficient: 0.7132689567282635
P-value: 9.283068344589747e-39

Top 10 Counties - Ratio of Cars to Buses per Inhabitant:

County	Ratio Cars to Buses per Inhabitant
Heidenheim	0.390287
Kusel	0.369557
Harburg	0.307322
Steinburg	0.295795
Biberach	0.245871
Oldenburg	0.186140
Miesbach	0.113131
Neustadt an der Weinstraße	0.106359
Miltenberg	0.080065
Sankt Wendel	0.069892

Correlation of the Cars to Buses per Inhabitant
Correlation coefficient: -0.12070086124386821
P-value: 0.0760205404250345

Challenges

Challenges

Diverse syntax of the county names

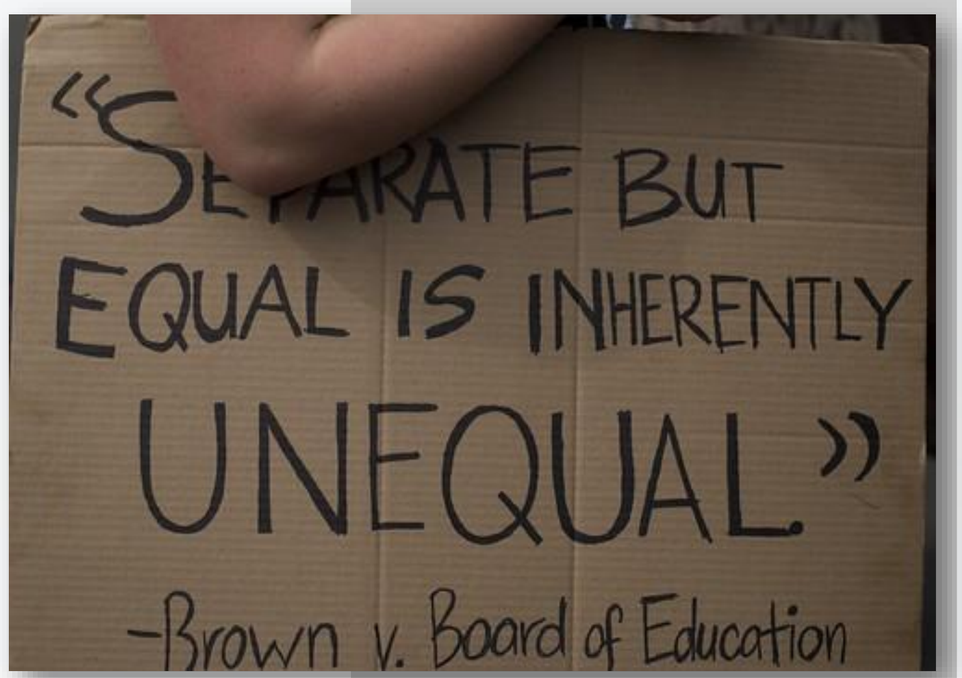
- Some started with the ZIP-Code
- Some had “Kreisfreie Stadt” at the end

Diverse definition of the county itself

- Some based the counties on the ZIP-Codes
- Some based the counties on the German “Landkreise”

The data could not be joined together correctly

- Although an attempt was made to make them equate, they could not be merged 100%



Challenges

This has led to distorted results

- Some counties were left out
- Some counties were cut

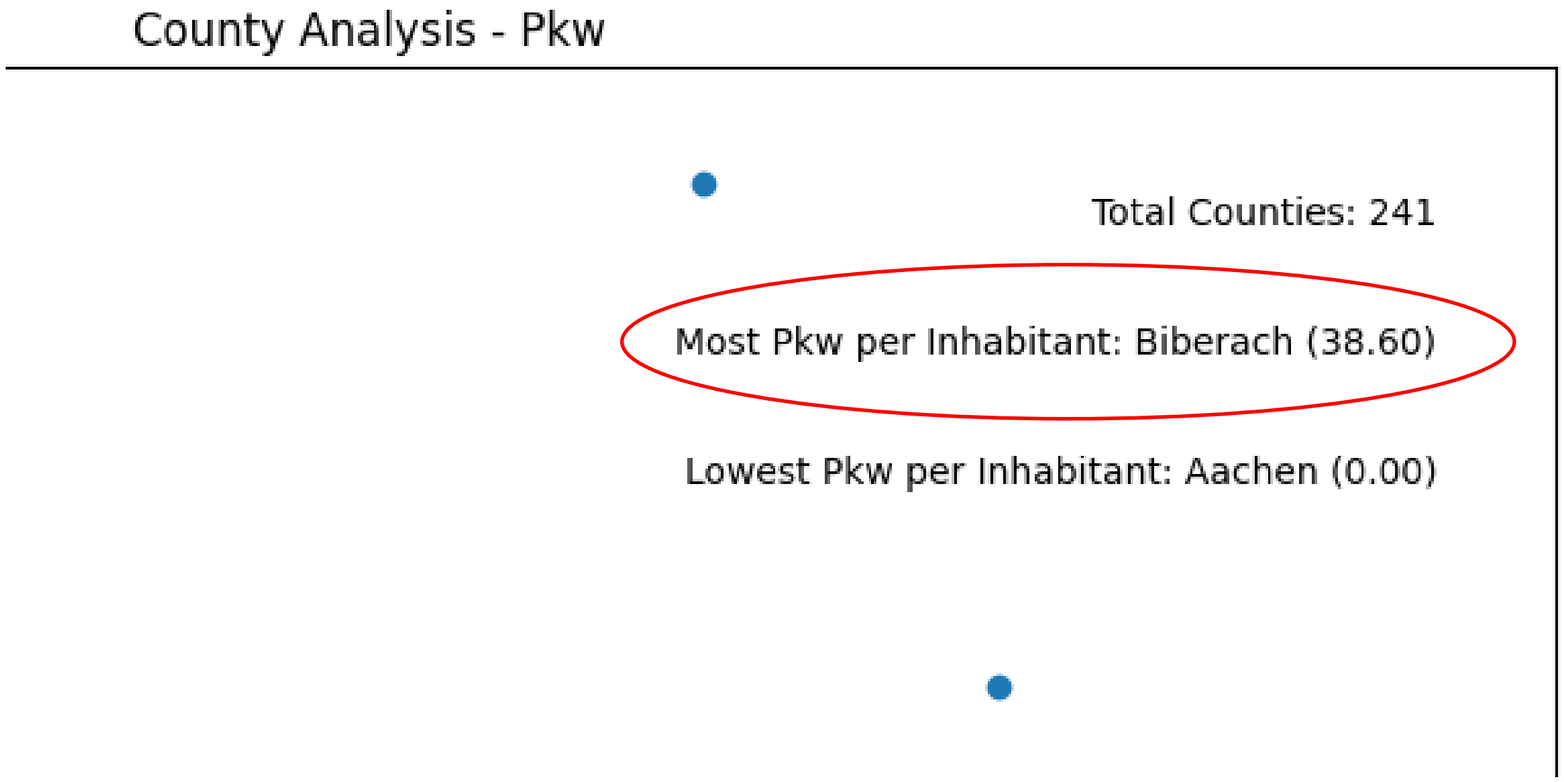
Some counties were not joined although they should

- In one table counties consisted of more rows
- In some cases, they could not be merged

Automatically joining counties with same “core name” was not possible

- Otherwise, some would be joined which should not
- For example, Erlangen and Erlangen-Höchststadt

Results


County Analysis - Pkw
Total Counties: 241
Most Pkw per Inhabitant: Biberach (38.60)
Lowest Pkw per Inhabitant: Aachen (0.00)

Outlook

Implications

Therefore, the whole data was polluted

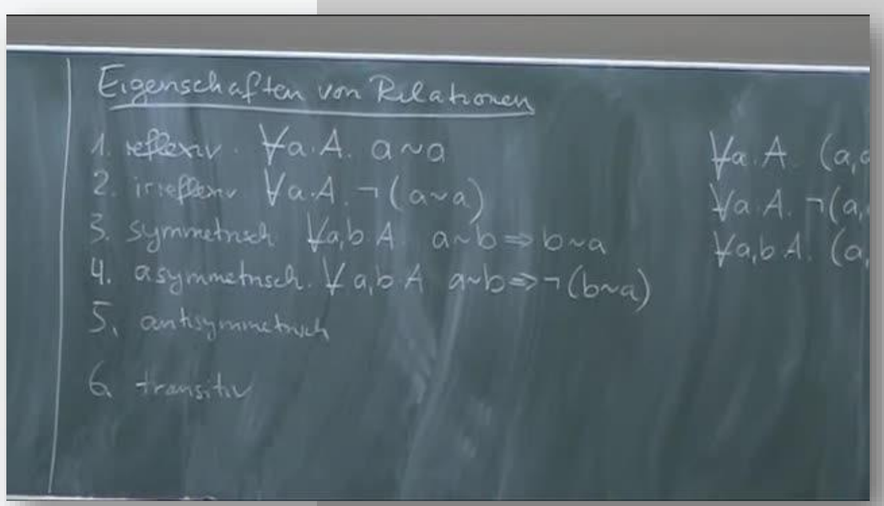
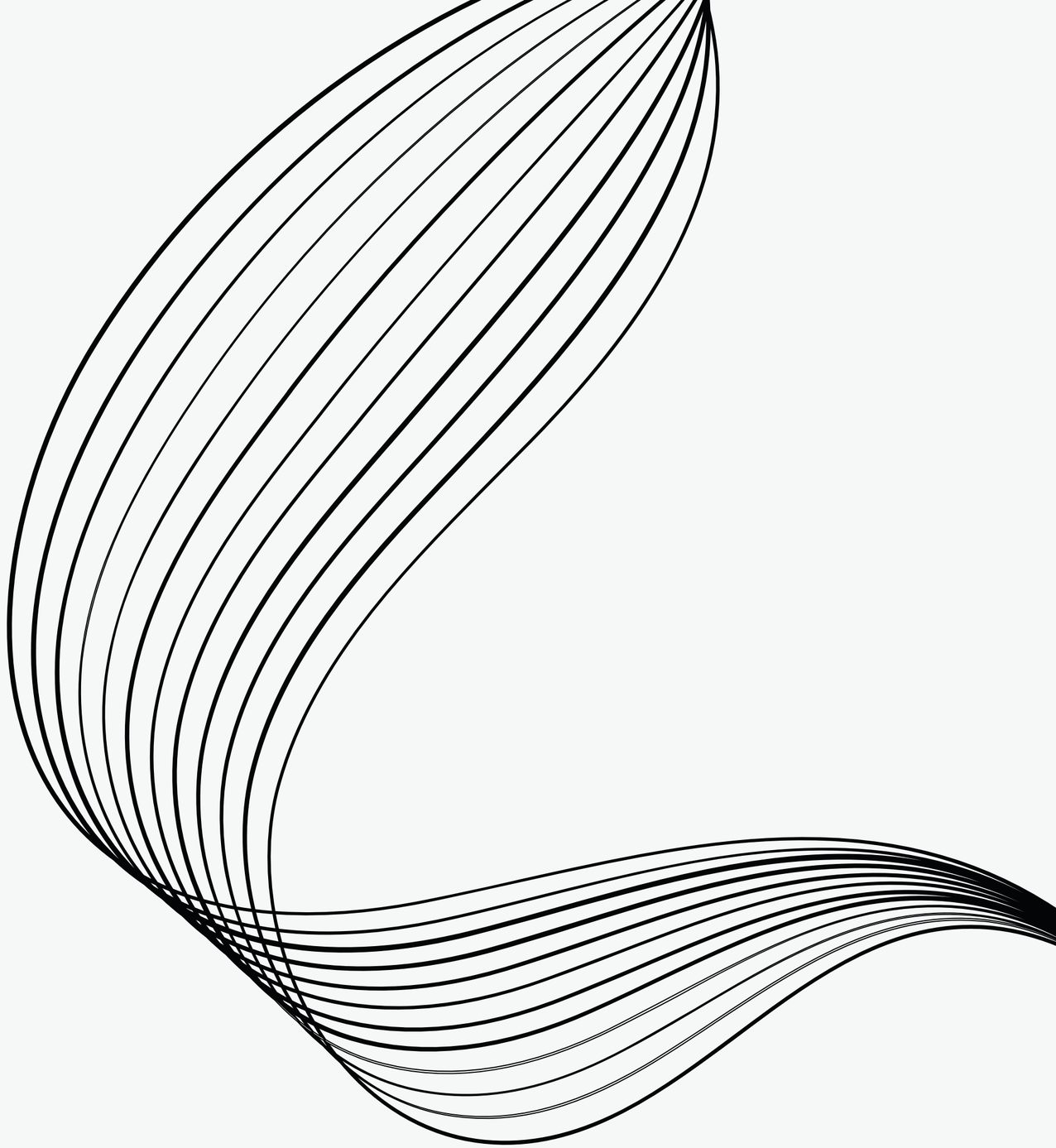
- Hence the data cannot be fully entrusted

All results could not be verified as they also could be polluted

- Despite some of the data could be reliable, the polluted data lead to a dilution of the informative value

Data cleaning would be required to make a statement

- Use the same format and syntax of the county names



THANKS FOR LISTENING