



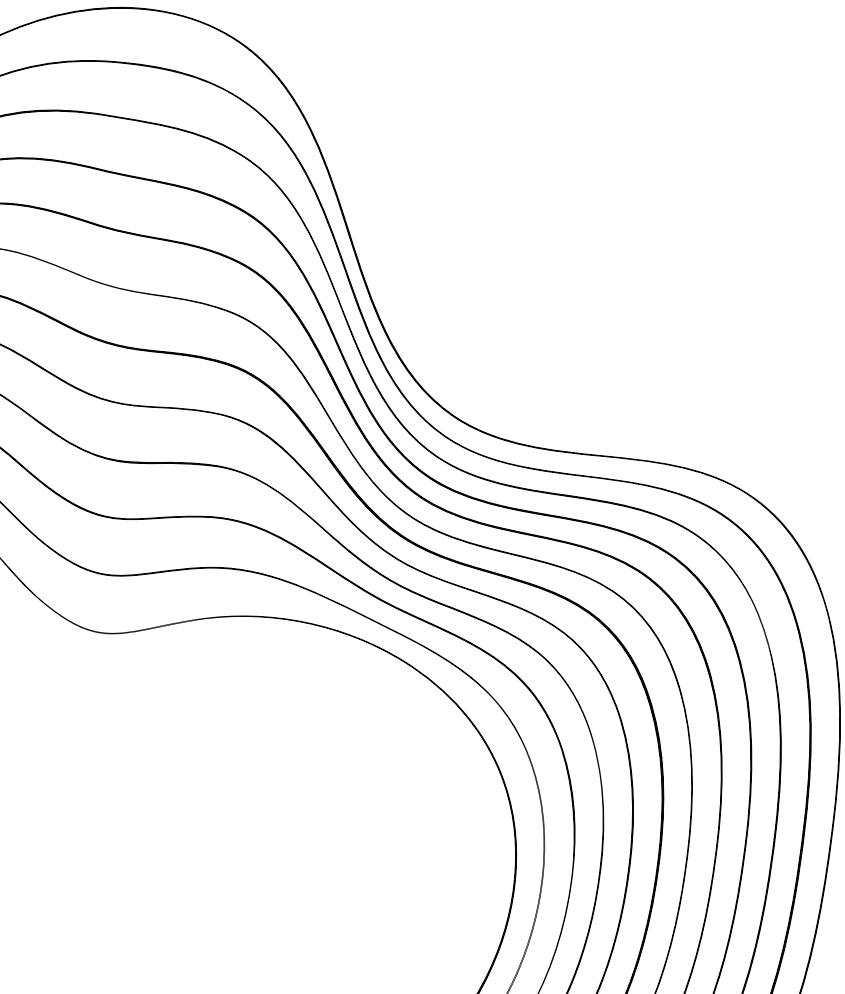
Data Engineering

P r o j e c t P r e s e n t a t i o n





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



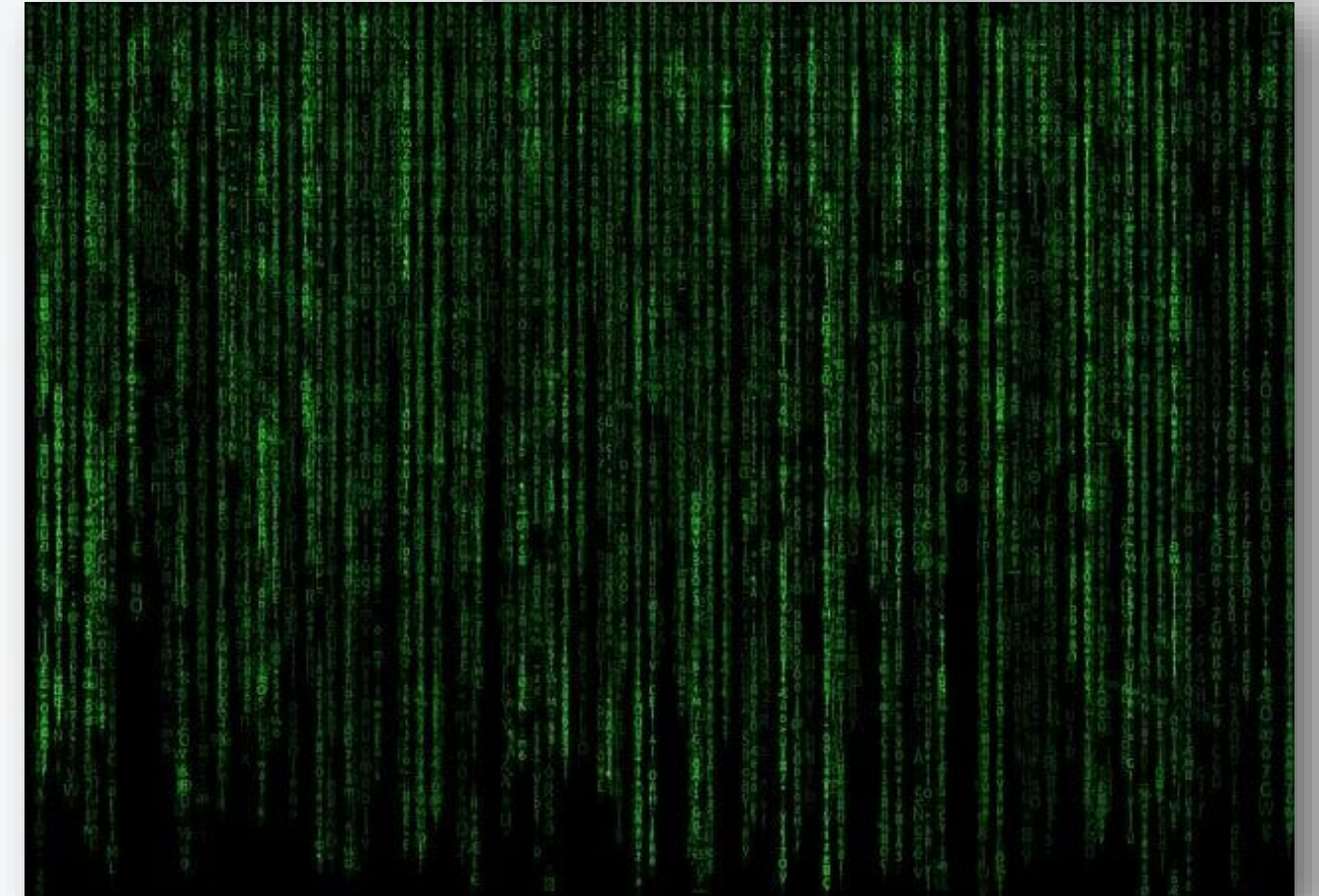
Data sources

County Data

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

Vehicle Data

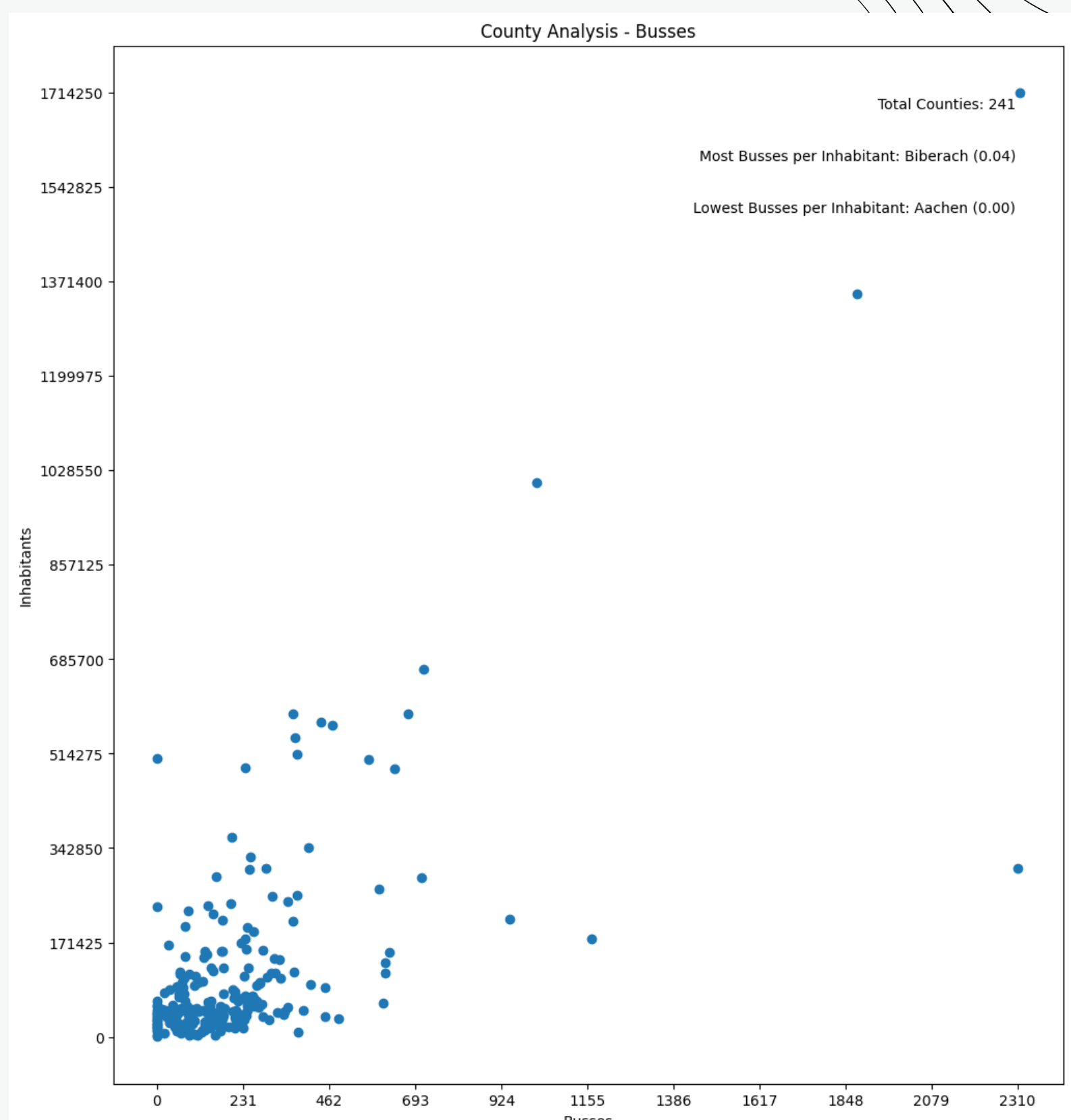
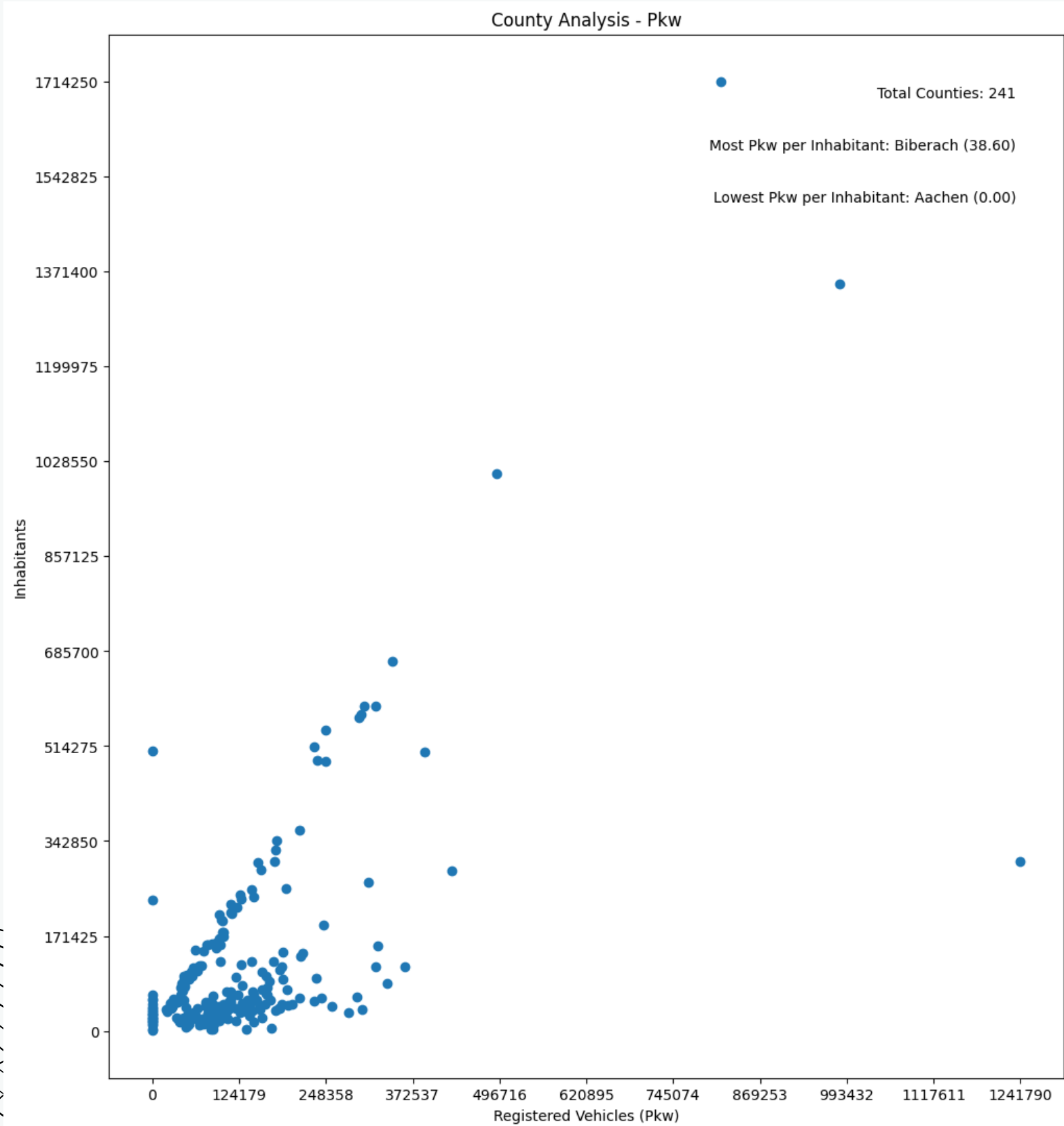
- County name
- Registered private cars
- Registered busses
- Other data which did not get used (e.g., Motorbikes etc.)





Results

Results



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

Top 10 Counties - Buses per Inhabitant:

County	Registered Busses / 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

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

County	Ratio Cars to Buses per Inhabitant
Neustadt an der Weinstraße	5560.666667
Mülheim an der Ruhr	2979.375000
Kusel	2395.100000
Worms	2207.619048
Pinneberg	2079.850575
Lörrach	2041.735294
Speyer	1979.066667
Sankt Wendel	1834.382353
Memmingen	1700.750000
Harburg	1669.990196



Challenges

Challenges

Diverse syntax of the county names

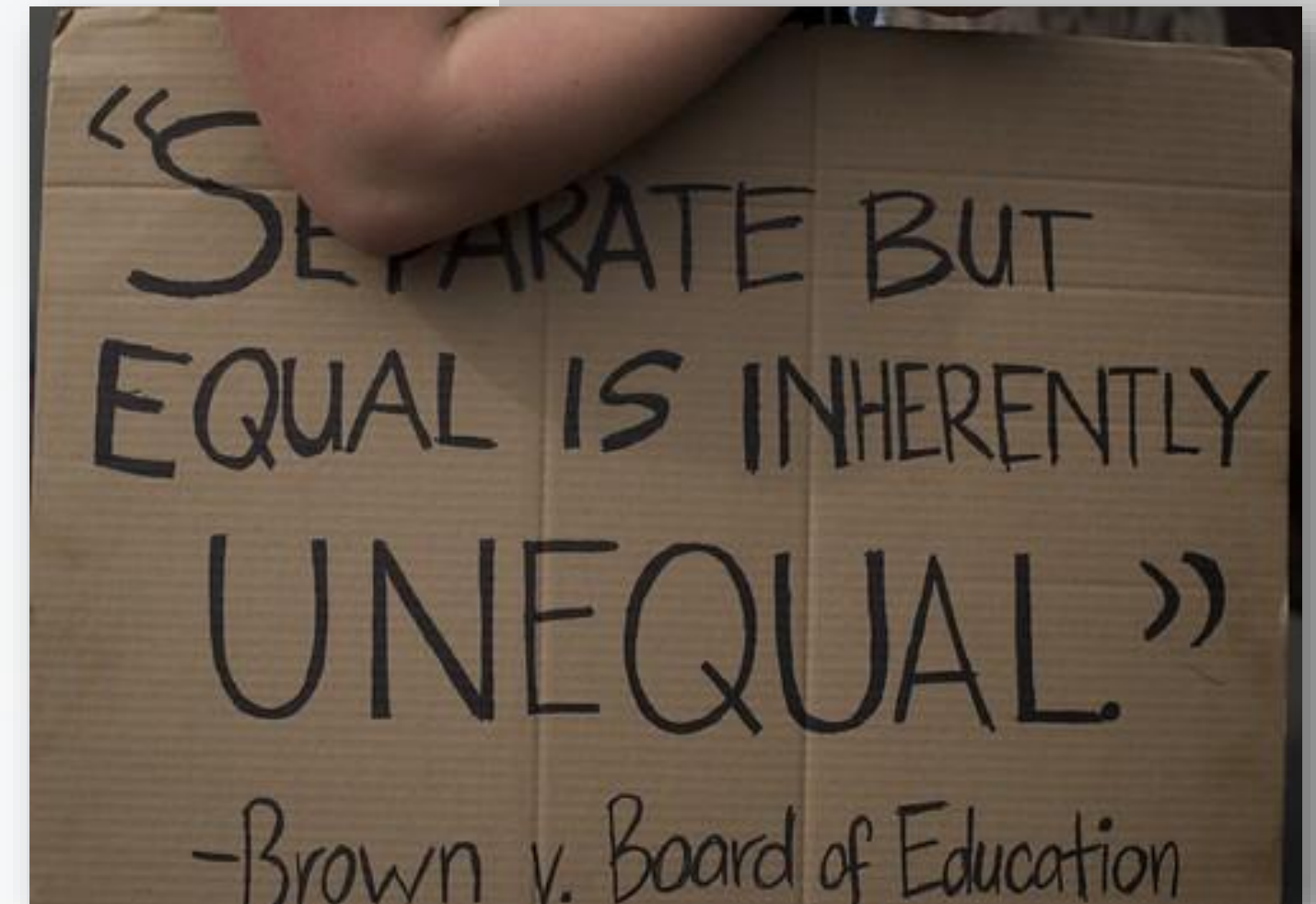
- Some started with the ZIP-Code
- Some had "Kreisfreie Stadt" at the end

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

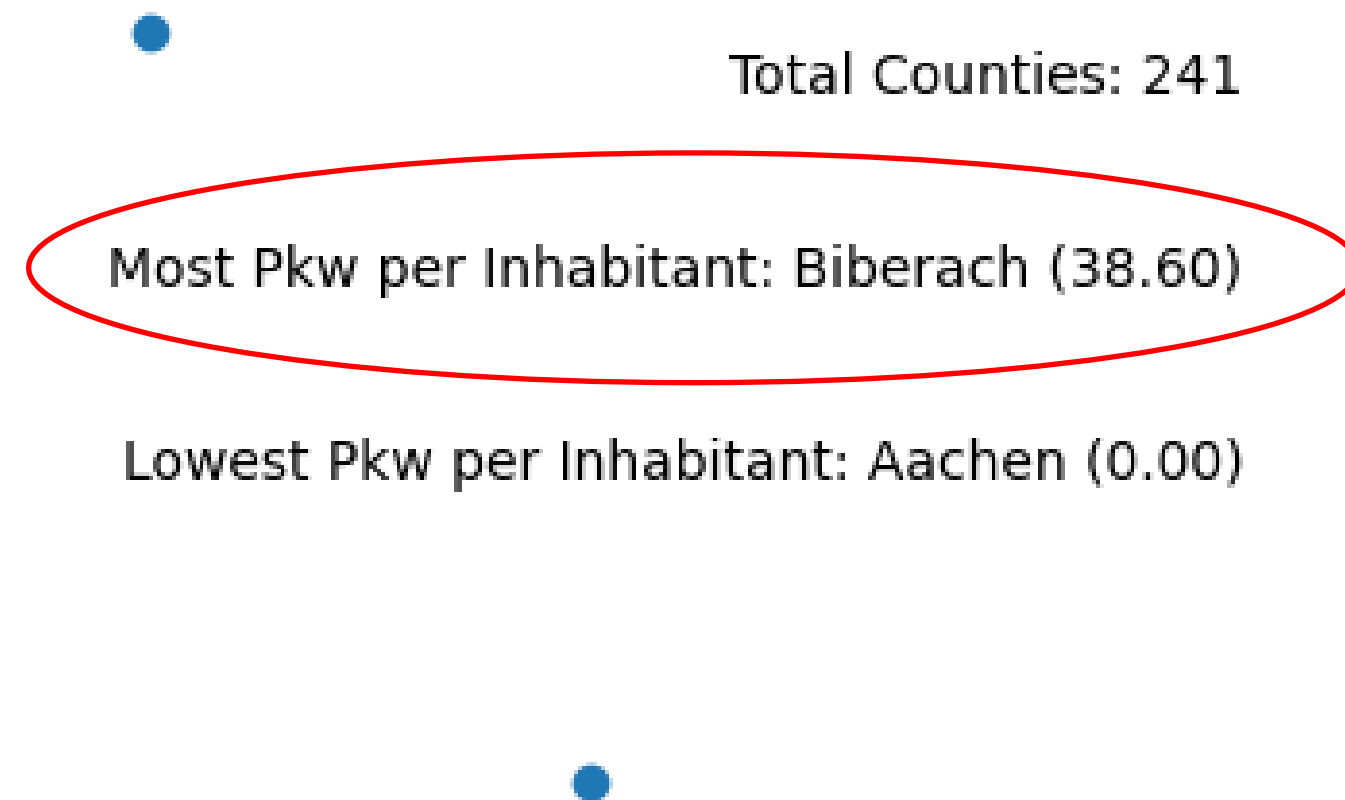
Automatically joining counties with same “core name” was not possible

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



Results

County Analysis - Pkw





Outlook

Implications

Therefore, the whole data was polluted

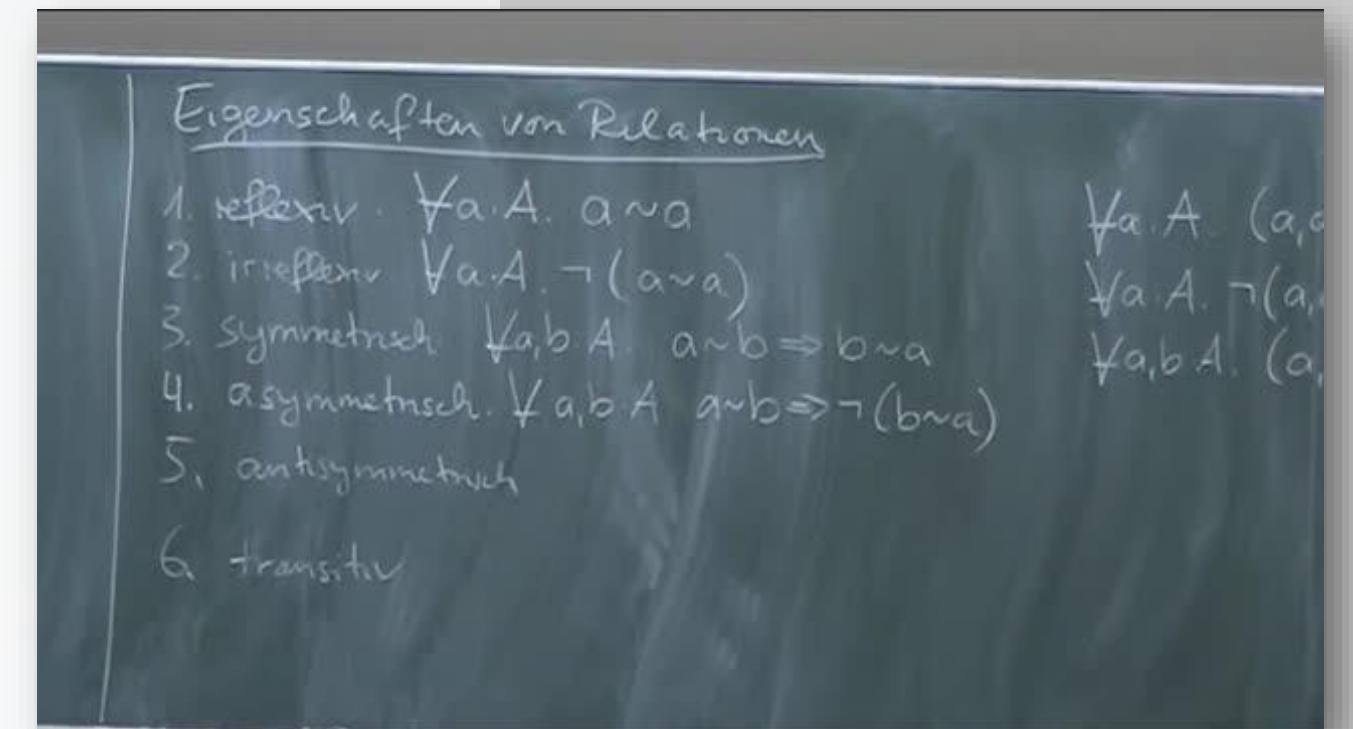
- Hence the data cannot be fully entrusted

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**

