Correlation between number of train stations and number of cars in a county?

Florian Oberndörfer

Methods of **A**dvanced **D**ata **E**ngineering, January 2024

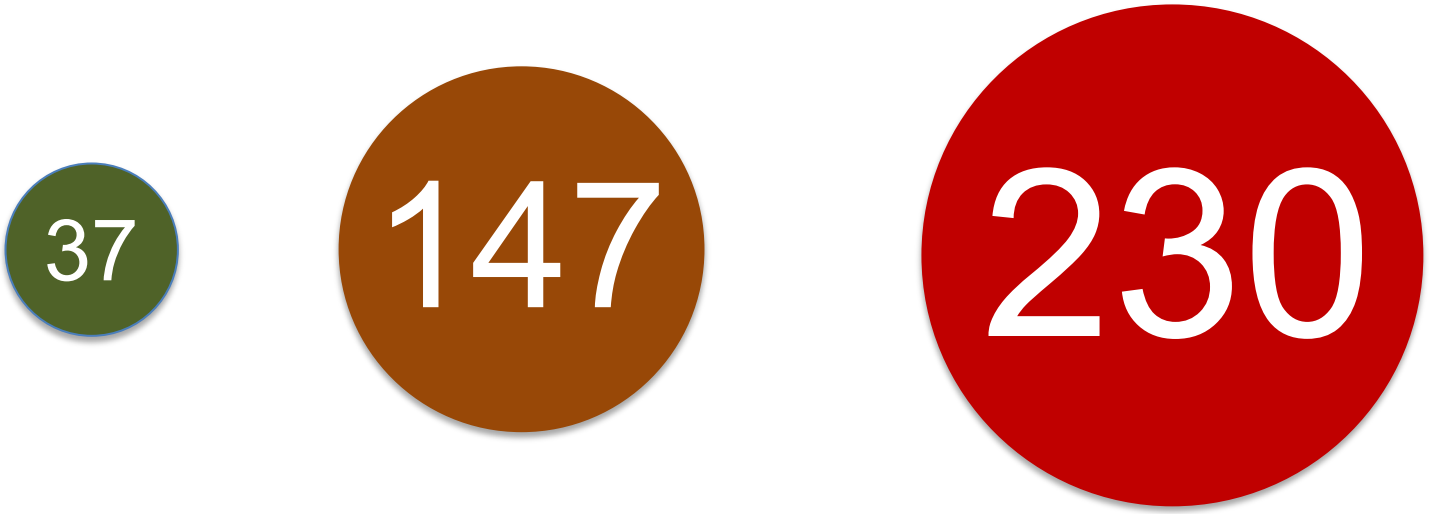
FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Main Question

Is there a correlation between the
number of train stations per square km
and the
number of cars per 1000 residents
in a county?

Motivation

Car: 147g CO_2 per person and km



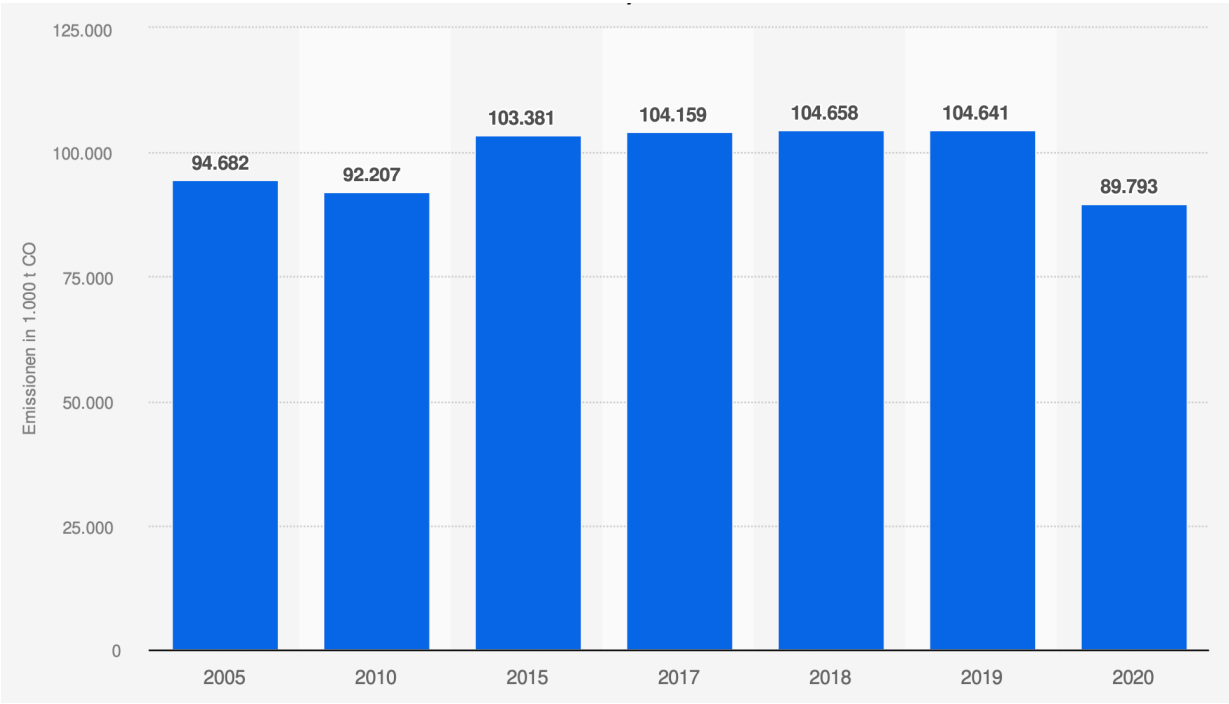
Train: 37g CO_2 per person and km

Plane: 230g CO_2 per person and km

Source: Tagesspiegel

Motivation

CO_2 emissions by private cars in Germany:



Source: Statista

Main Datasources

1. Train stations per zip code

- Source: Deutsche Bahn
- Type: CSV
- License: CC BY 4.0

2. Number of cars per county

- Source: Mobilithek
- Type: CSV
- License: Datenlizenz Deutschland Namensnennung 2.0

3. Residents and area per zip code

- Source: Suche Postleitzahl
- Type: CSV
- License: ODbL

Problem

Train station data only exists per **zip code**

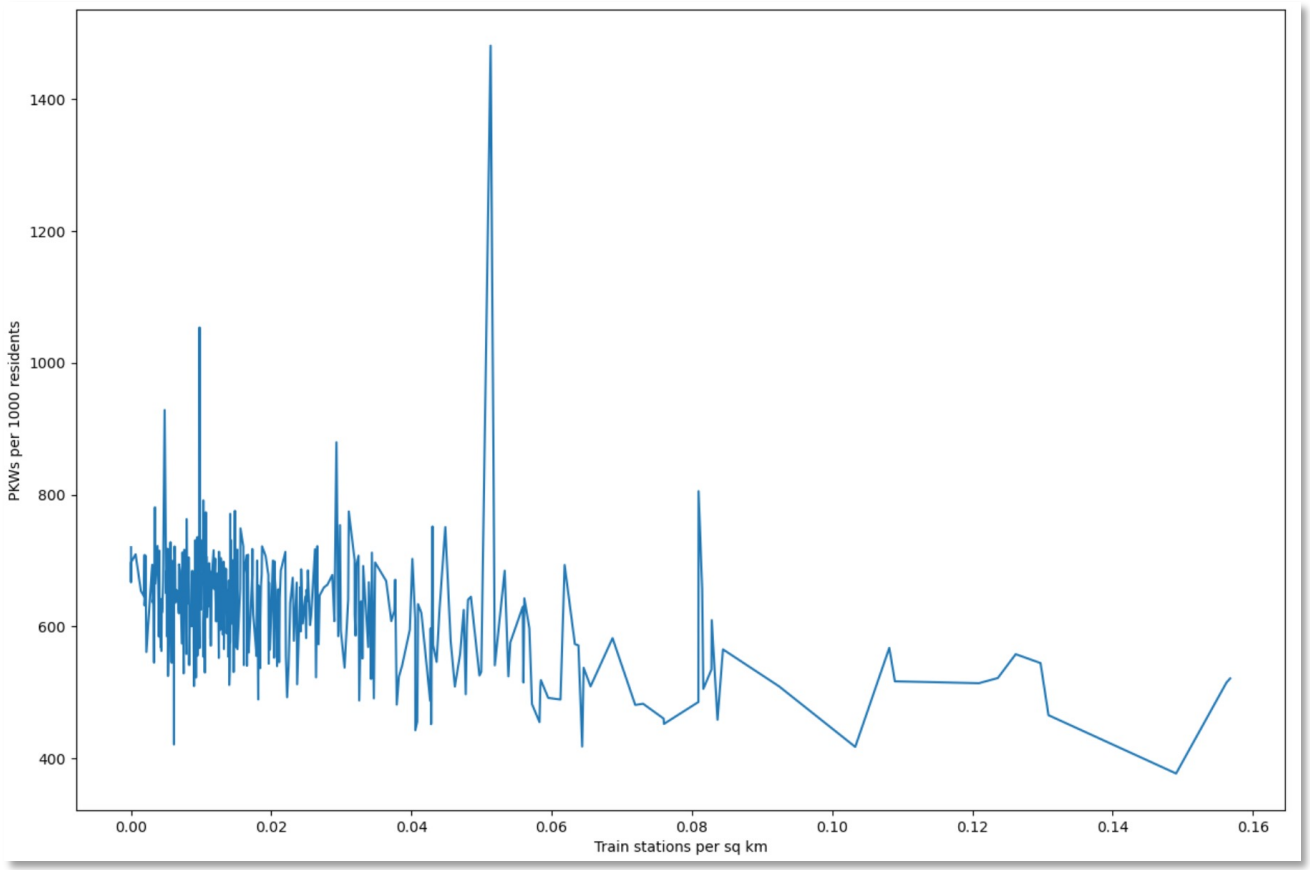
Number of cars data only exists per **county**

Helper Datasource

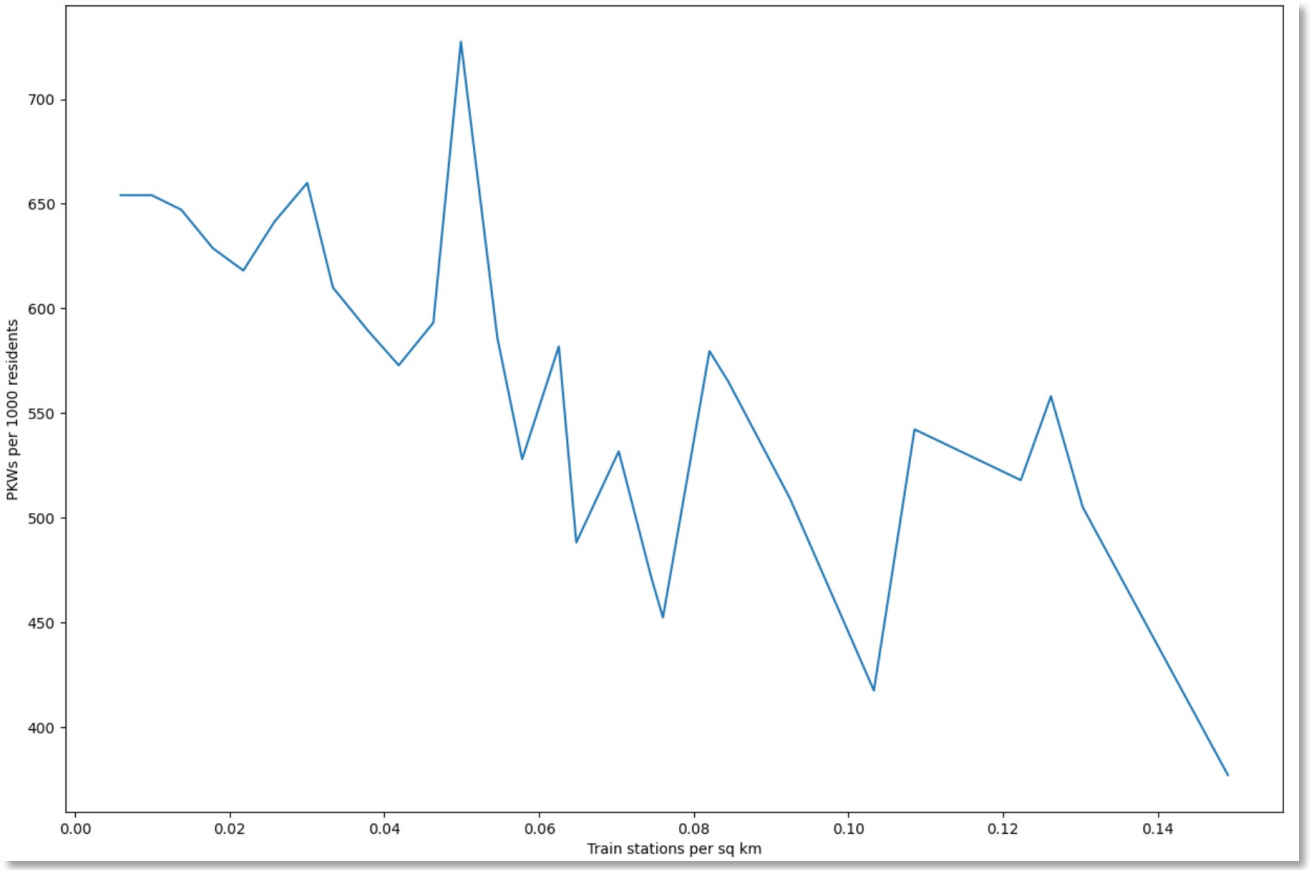
Allocation table for zip code and county

- Source: Suche Postleitzahl
- Type: CSV
- License: ODbL
- Shows the county for every zip code

Diagrams - RAW Data

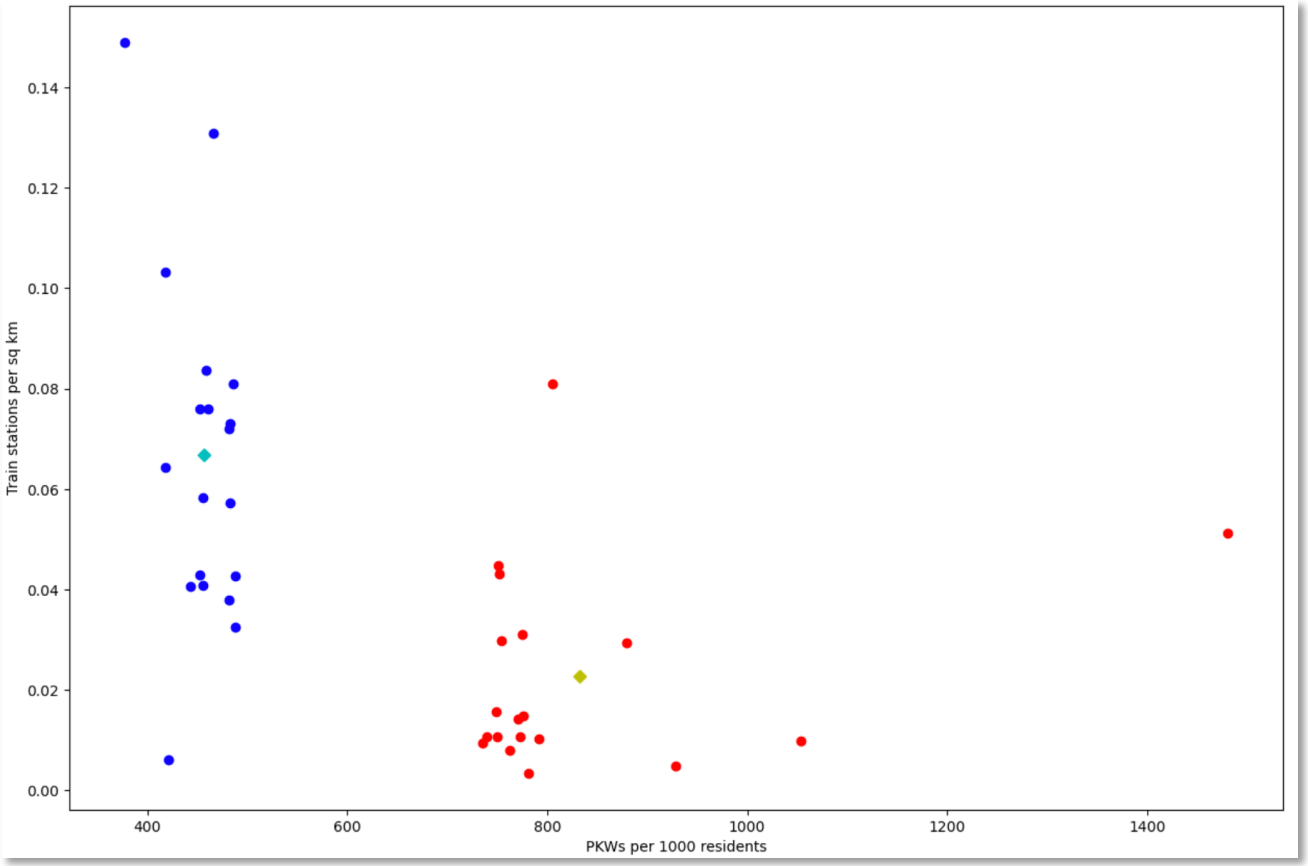


Diagrams - Averaged



Diagrams – Extreme values

Blue: Less 5% of counties regarding cars per 1000 residents
Red: Top 5% of counties regarding cars per 1000 residents



Cyan: Average value of blue points

Yellow: Average value of red points

Conclusion

- Resulting diagrams are not as clearly as hoped
- Tendency, that increasing number of train stations would lead to decreasing number of cars
- Especially in the diagram with the extreme values, we can see a big difference

Methods of Advanced Data Engineering

Correlation between number of cars and number of train stations in a county?

Florian Oberndörfer

florian.oberndoerfer@fau.de

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT