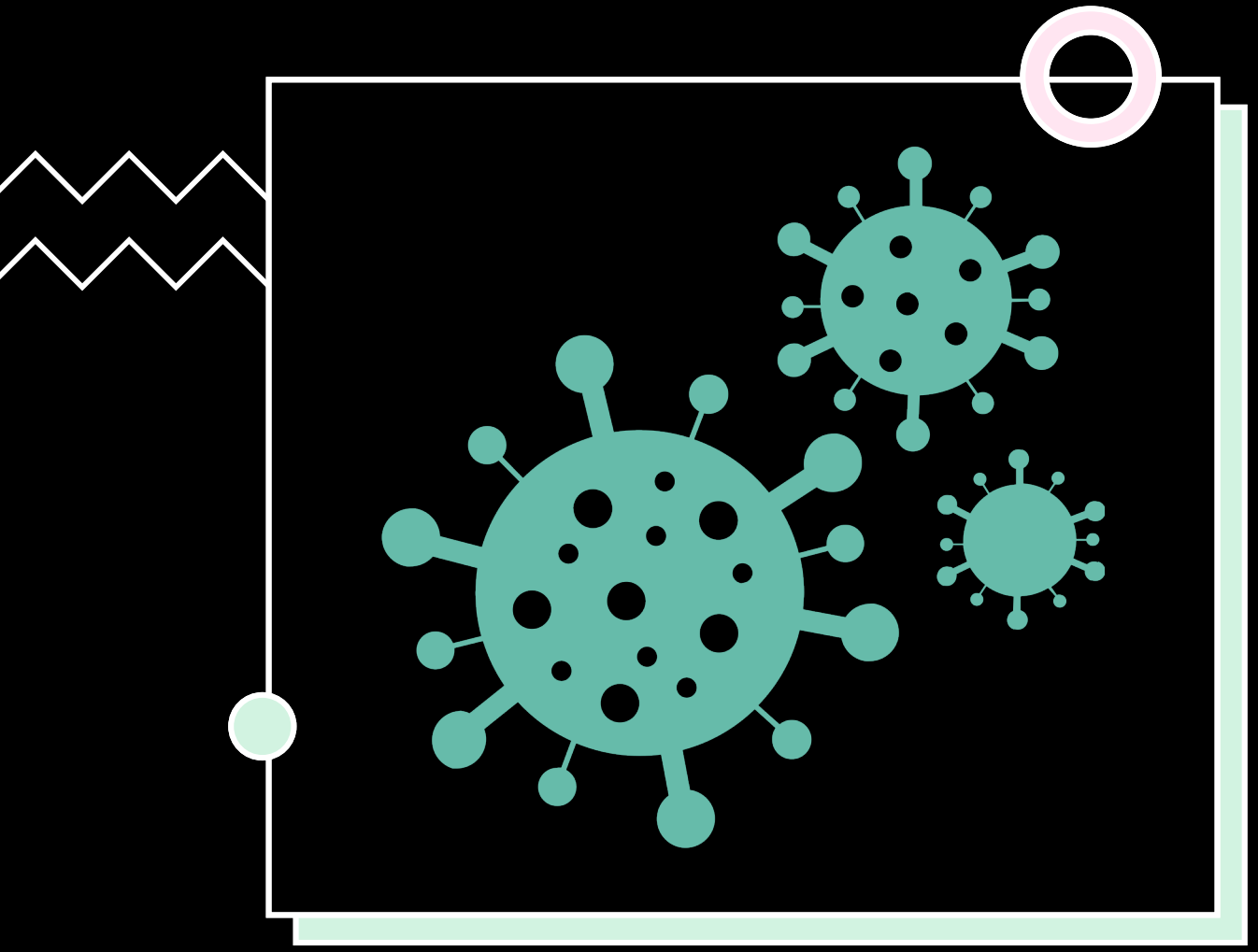


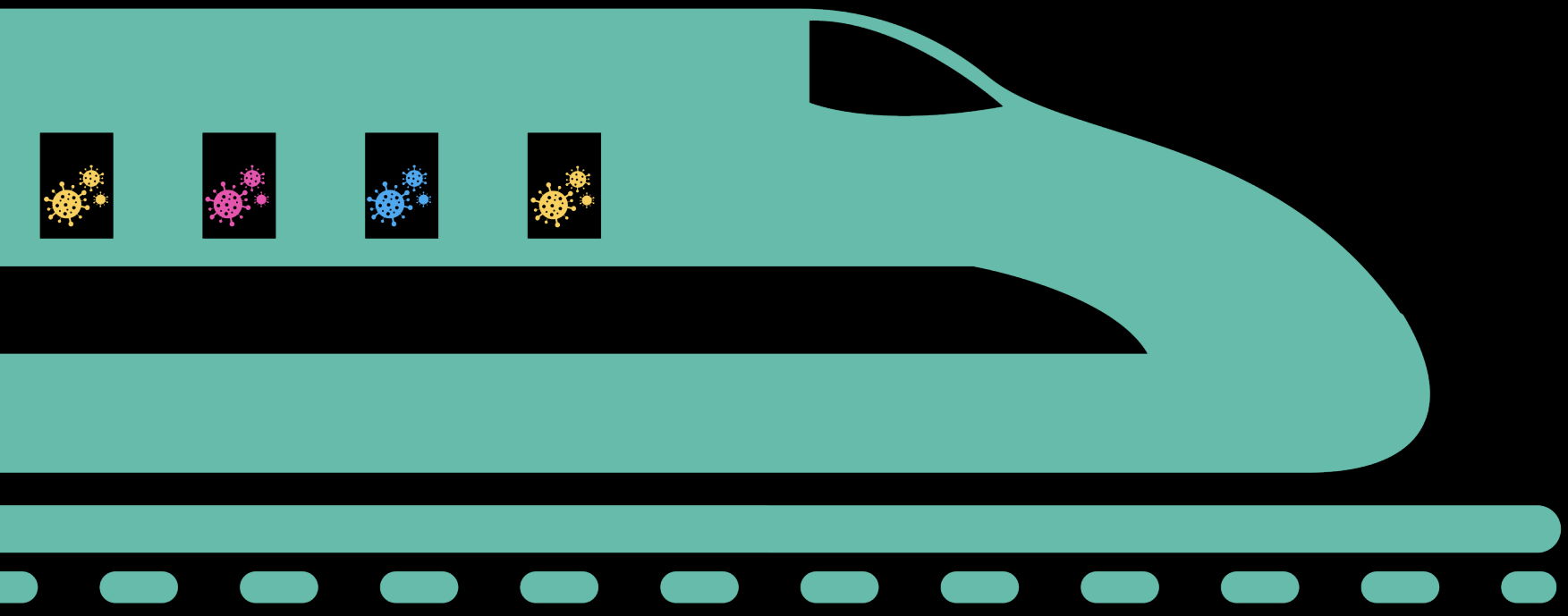
DS II SEMINAR

MARCEL OCHSENDORF

FINAL PRESENTATION

RESEARCH QUESTION

IS IT POSSIBLE TO IDENTIFY COVID SPREADS DUE TO PUBLIC TRANSPORTATION SERVICE?*

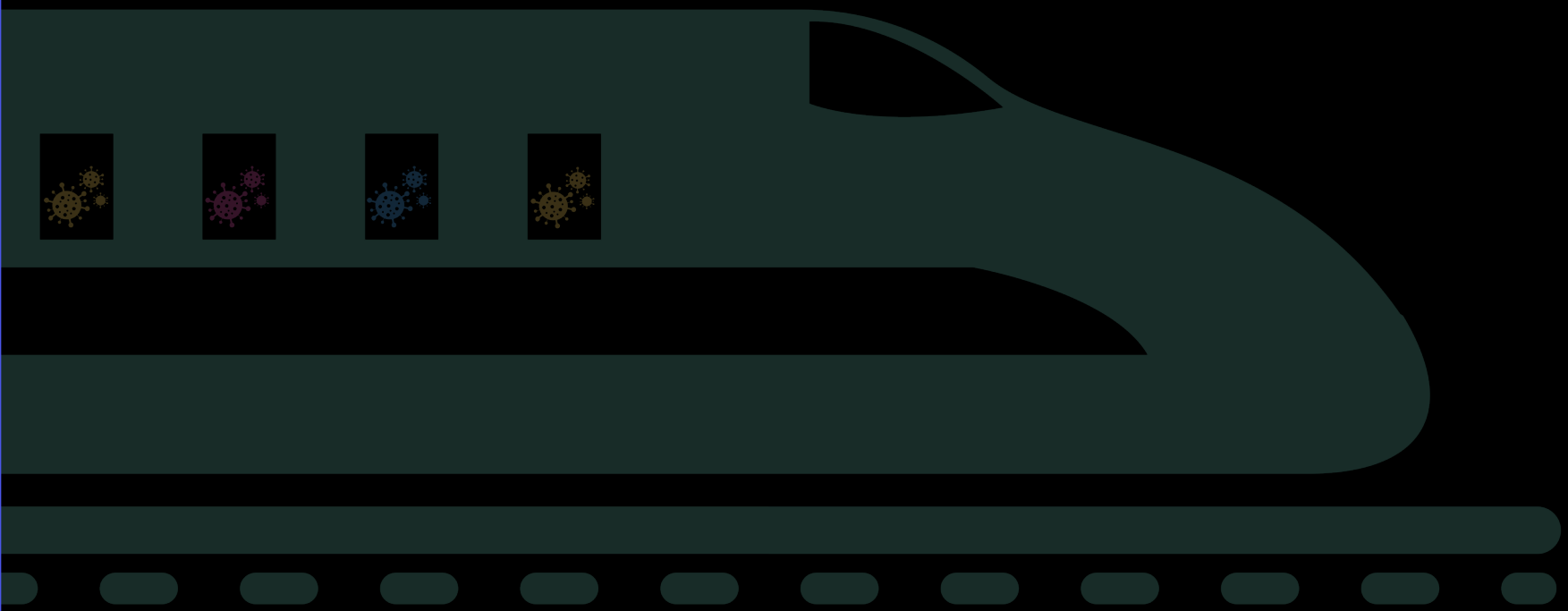


* ASSUMING 100% RAILWAY TRAFFIC [NO CARS, ...]

RESEARCH QUESTION

IS IT POSSIBLE TO IDENTIFY COVID SPREADS DUE TO PUBLIC TRANSPORTATION SERVICE?*

not possible with the current used datasets



* ASSUMING 100% RAILWAY TRAFFIC [NO CARS,...]

TIMELINE

PREPARE DATA [DB | RKI]

BUILD MAINSTATION AND DEPATURE DATASET

FETCH TRAIN DATA FOR SPECIFIED INTERVALS

COMBINE PROCESSED DATA

GENERATE HEATMAPS

BUILD AN INTERACTIVE MAP

INITIAL

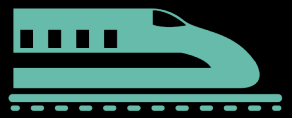
INTERMEDIATE

FINAL

PREPROCESSING

DB API

[RAILWAY STATIONS]



DISTRICTS

[GEOJSON POLYGONS AND AGS]

- STATIONS INSIDE DISTRICTS
- FILTERED STATIONS (DB EVU ONLY)
- GEOJSON WITH STATION FEATURES [POI]
- SEVERAL LOOKUP-TABLES PERFORMANT FOR VISUALISATION
  - [TRAIN LIST, DISTRICTS,]

CONVERTING COORDINATES

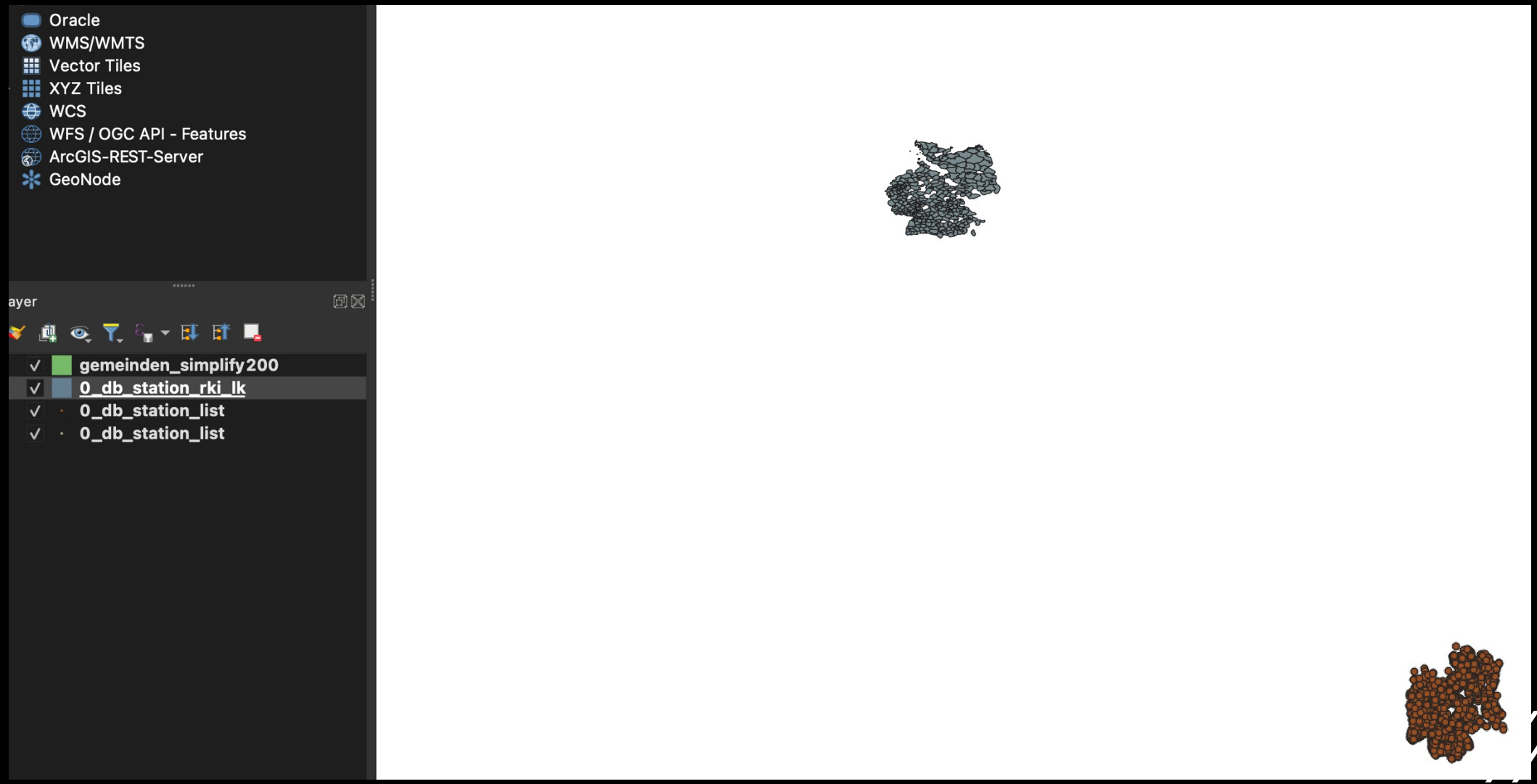
DISTRICT, DB AND RKI DATA ARE USING DIFFERENT COODINATE SYSTEMS

LAYERS ARE NOT MATCHING [OFFSET, SCALE]

REFORMAT THESE WITH QGIS AND GEOPANDAS

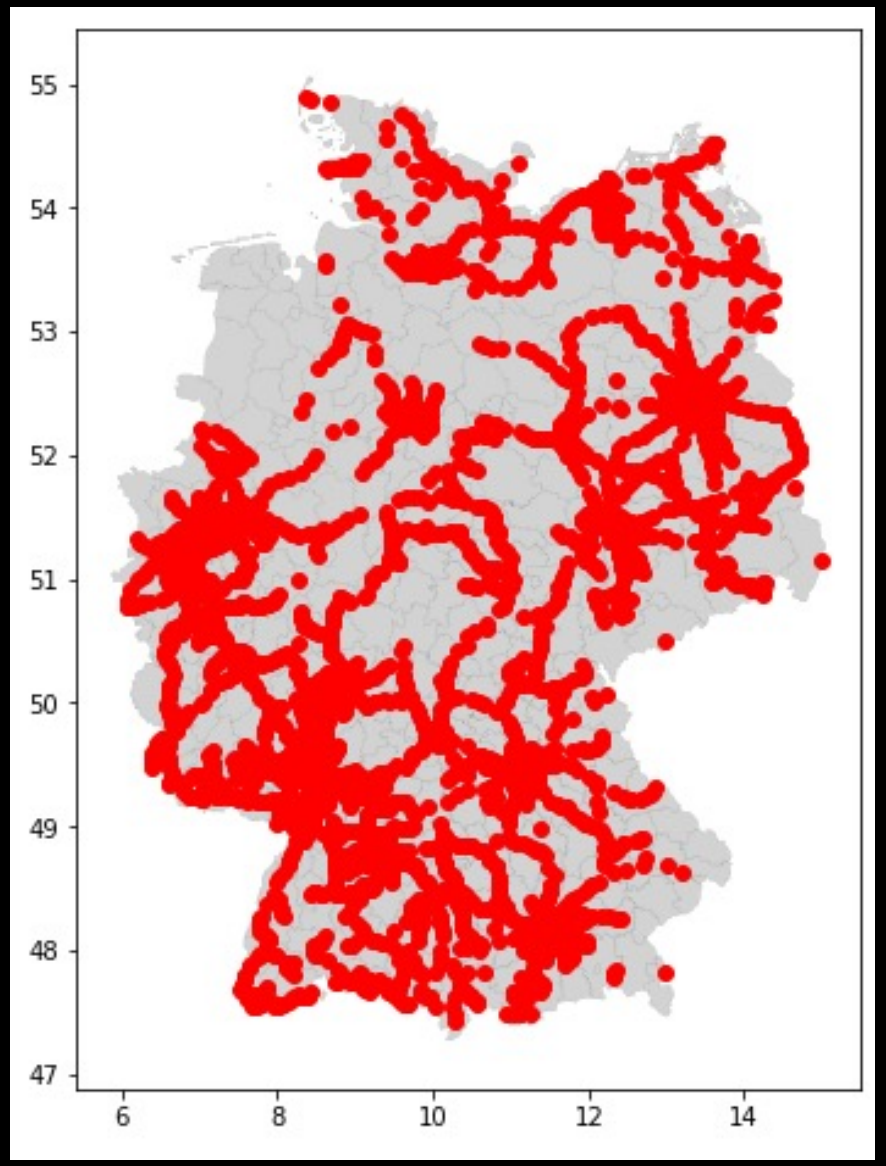
QGIS IS QUITE COMPLEX TO USE

CONVERTING COORDINATES



CUSTOM GEOJSON


```
"type": "FeatureCollection",
"crs": {
    "type": "name",
    "properties": {
        "name": "urn:ogc:def:crs:OGC:1.3:CRS84"
    }
},
"features": [
    {
        "type": "Feature",
        "properties": {
            "station_name": "Ürzig(DB)",
            "station_id": 8005945,
            "featureclass": "Admin-1",
            "nameascii": "Ürzig(DB)",
            "name": "Ürzig(DB)",
            "latitude": 7.004806,
            "longitude": 49.995933,
            "geonameid": -1,
            "note": null
        },
        "geometry": {
            "type": "Point",
            "coordinates": [
                49.995933,
                7.004806
            ]
        }
    },
```

CUSTOM TRAIN DATASET

DB API

[DEPARTURE TABLE]

- 5GB CUSTOM DATASET CREATED FOR ALL DB TRAINS [01.11.2021 – 16.01.2022]
- INCLUDING ARRIVAL | DEPARTURE DELAYS
- USEFUL FOR FURTHER USAGE

- MERGE SMALL DEPARTURE SNIPPETS INTO A LARGE SET
- RECONSTRUCTION OF THE TRAIN DRIVE COURSE

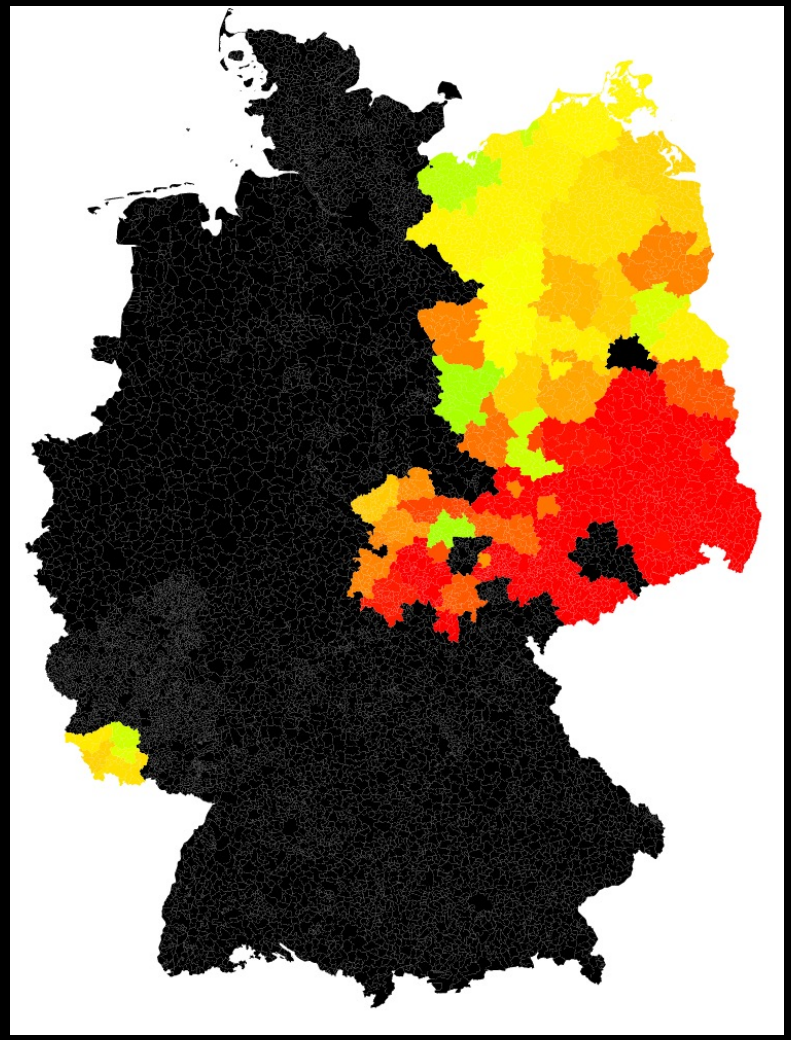
PROCESS RKI DATA

- ONE DOWNLOAD OF DISTRIC DATA AT THE END OF EACH DAY
- INCIDENCE AND AGS_ID USED

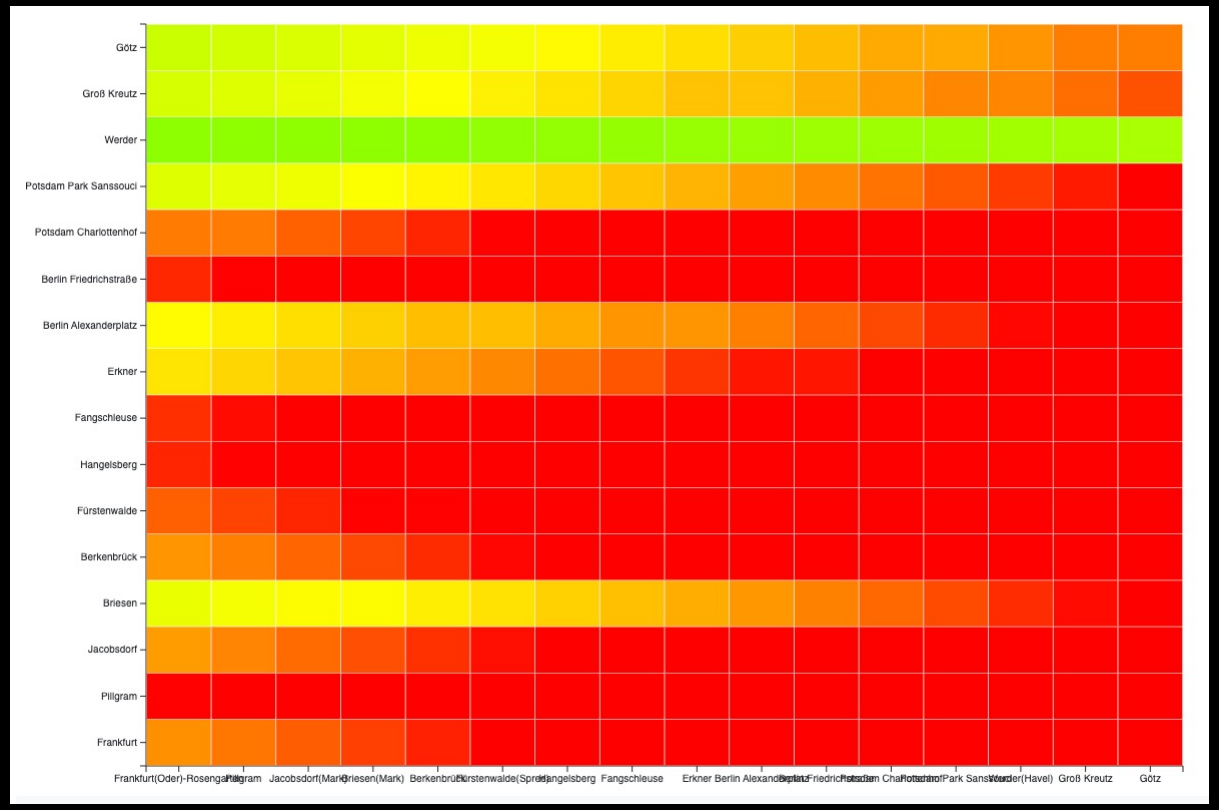
```
rki_ags;population;incidence;cases_100k;cases_week
09462;74048;530.7368193604149;8741.627052722559;393
05711;333509;382.8982126419377;9402.145069548349;1277
09181;121019;626.3479288376205;8873.813202885498;758
09363;42535;420.8299047842953;10429.058422475608;179
09261;73065;567.9874084719086;11261.20577567919;415
05758;250547;467.77650500704453;9017.868902840586;1172
03354;48503;313.3826773601633;4317.258726264355;152
09361;42052;332.92114524873966;9830.685817559212;140
09661;70858;491.12309125292836;9685.568319738068;348
08211;55449;804.3427293549028;9448.321881368463;446
03459;359471;375.2736660259103;7817.876824556083;1349
05314;330579;535.4242102492899;8482.087488920954;1770
04011;566573;1498.4829845403858;9675.716986160654;8490
04012;113557;915.8396223922788;9411.132735102196;1040
09463;40842;364.82052788795846;10712.012144361197;149
05166;298536;504.1268054773964;7394.418093630249;1505
03355;184235;528.1298341791734;4983.309360327842;973
01004;79905;604.467805519054;5920.780927351229;483
03252;148580;301.52106609234085;5386.323865930812;448
09562;112385;557.0138363660631;7373.759843395471;626
05970;275491;326.68943813046525;7680.468690447238;900
03361;137574;830.0987105121607;7632.256094901653;1142
05512;117388;454.04981769857227;8175.452346066038;533
07311;48750;369.2307692307692;10038.97435897436;180
```

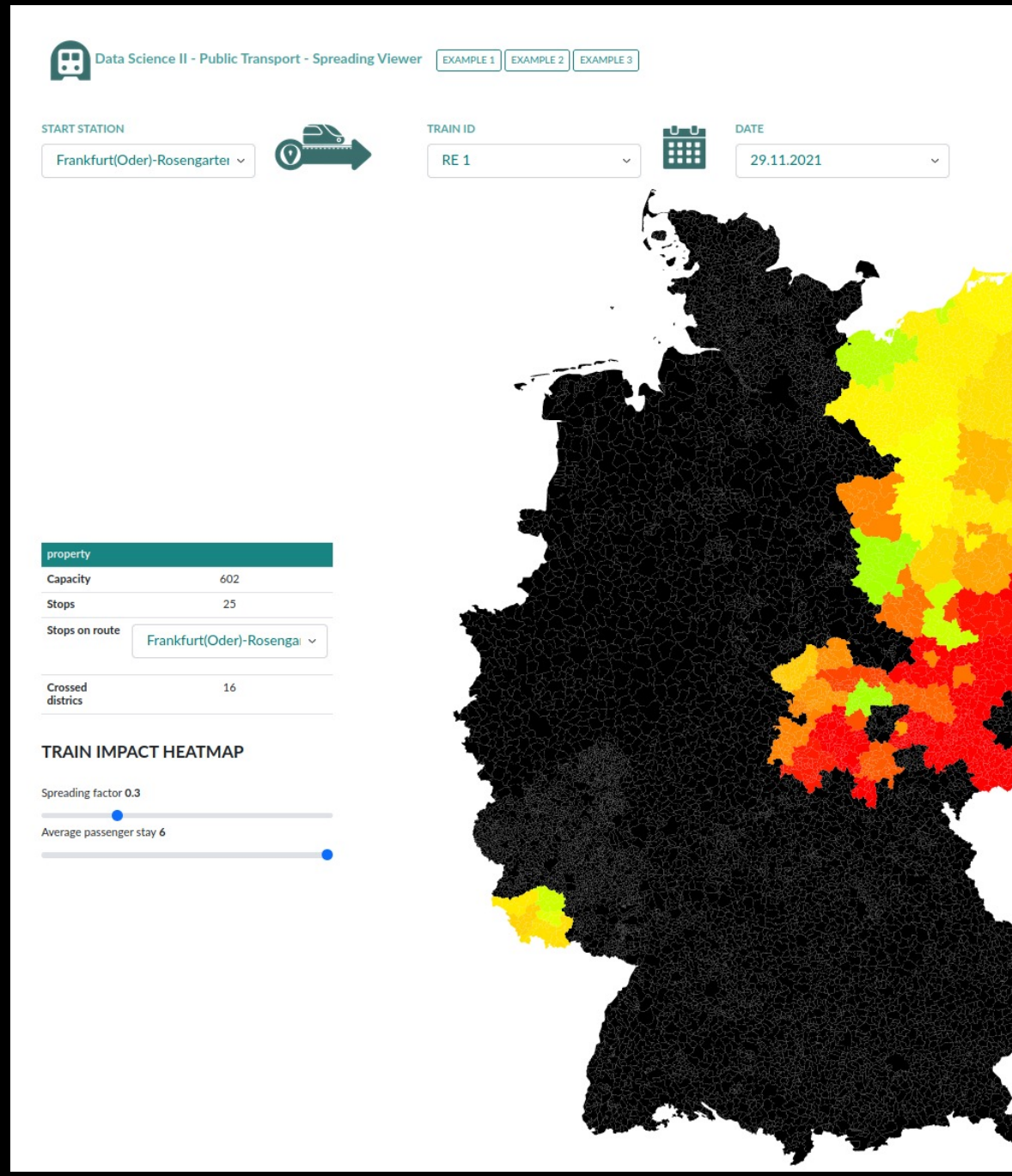
CHECK YOUR DATASETS!

- USED [HTTPS://API.CORONA-ZAHLEN.ORG/DOCS/ENDPOINTS/DISTRICTS.HTML](https://api.corona-zahlen.org/docs/endpoints/districts.html) FOR COVID DISTRICT DATA
- CSV RESPONSES LOOKED GOOD
- SEVERAL DISTRICTS ARE MISSING
- DATA APPARENTLY LOOKEDGOOD
- A SPORADIC TEST OF THIS DATA WAS SUCCESSFUL
- ONLY NOTICED AFTER WORKING VISUALIZATION



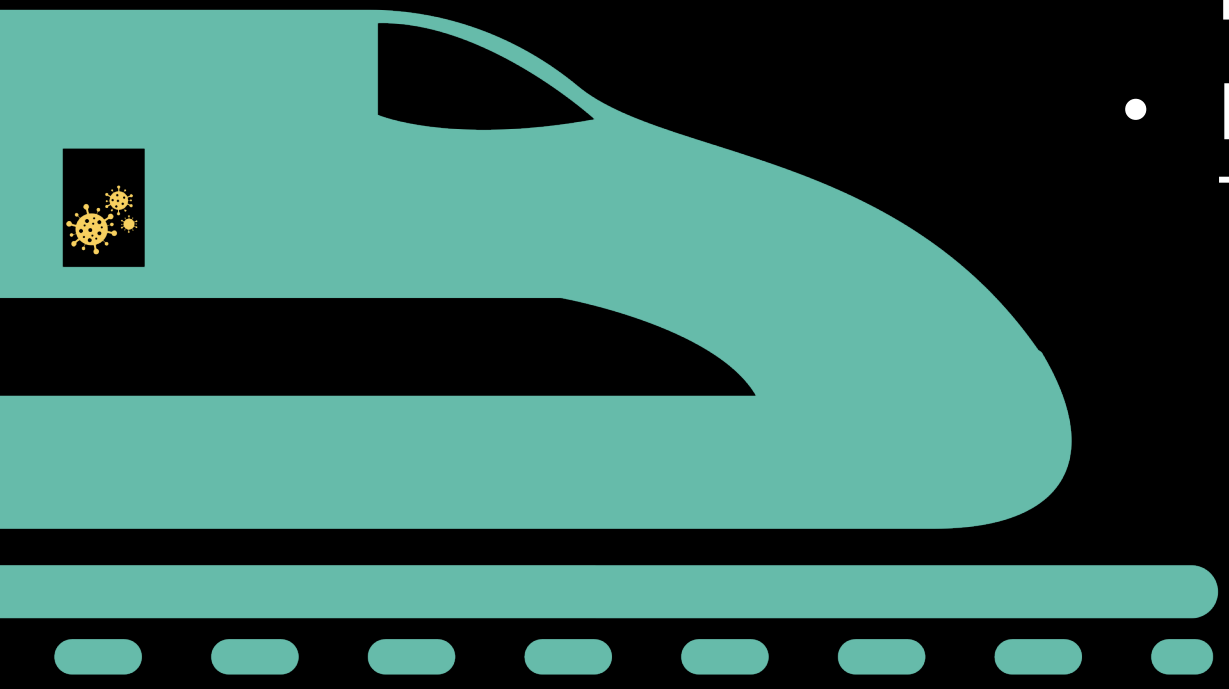
VISUALISATION

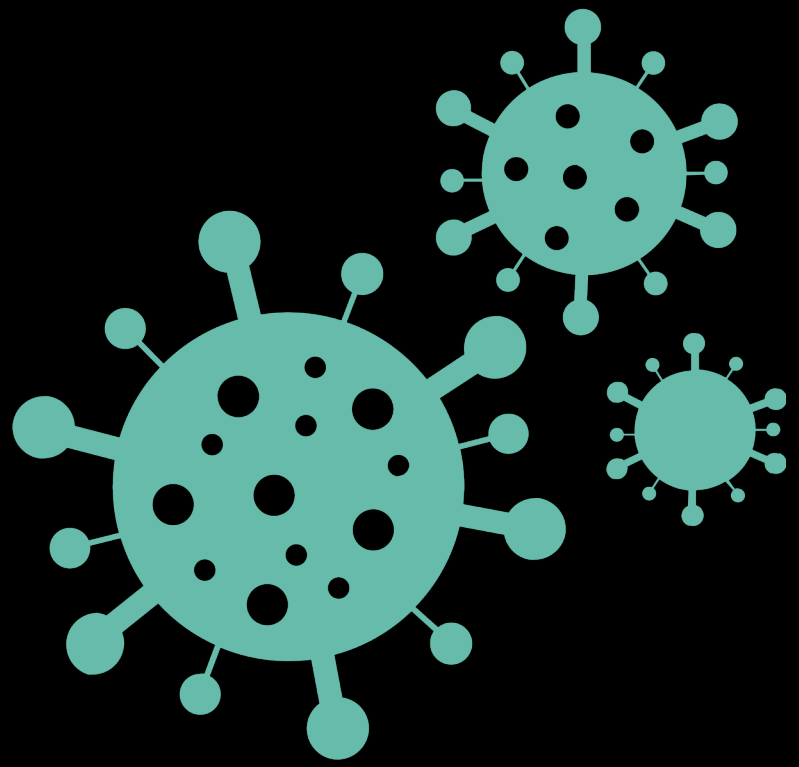




FUTURE WORK

- FIXING COVID DISTRICTDATA
- **ADD PASSENGER COUNTS**
- **INCLUDE LONG DISTANCE TRAINS INSTEAD OF REGIONAL**
- RUN OTHER VISUALISATION TASKS ON DELAY TRAIN DATA





QUESTIONS ?

https://github.com/RBEGamer/DataScienceIISeminar